Amplified-reflection plasmon instabilities in grating-gate plasmonic crystals

Aleksandr S. Petrov* and Dmitry Svintsov
Laboratory of 2D Materials' Optoelectronics, Moscow Institute of Physics and Technology, Dolgoprudny 141700, Russia

Victor Ryzhii
Institute of Ultra High Frequency Semiconductor Electronics, RAS, Moscow 117105, Russia
and Research Institute of Electrical Communication, Tohoku University, Sendai 980-8577, Japan

Michael S. Shur
Department of Electrical, Computer, and System Engineering and Department of Physics, Applied Physics, and Astronomy, Rensselaer Polytechnic Institute, Troy, New York 12180, USA
(Received 24 October 2016; revised manuscript received 16 December 2016; published 4 January 2017)

We identify a possible mechanism of the plasmon instabilities in periodically gated two-dimensional electron systems with a modulated electron density (plasmonic crystals) under direct current. The instability occurs due to the amplified reflection of the small density perturbations from the gated/ungated boundaries under the proper phase-matching conditions between the crystal unit cells. Based on the transfer-matrix formalism, we derive the generic dispersion equation for the traveling plasmons in these structures. Its solution in the hydrodynamic limit shows that the threshold drift velocity for the instability can be tuned below the plasmon phase and carrier saturation velocities, and the plasmon growth rate can exceed the collisional damping rate typical of III–V semiconductors and graphene at room temperature.

DOI: 10.1103/PhysRevB.95.045405

I. INTRODUCTION

The emission of terahertz (THz) radiation from two-dimensional electron systems (2DESs) under direct current has been observed in a large number of experiments, starting from the pioneering work of Tsui, Gornik, and Logan [1]. At the current stage of technology, the emission from these structures remains up to room temperature [2], and its frequency is voltage tunable from 0.5 to 2 THz [3], while the linewidth can be as narrow as ∼40 GHz [4]. It is commonly accepted that the radiation appears as a result of plasmon excitation [1,5–7] in 2DESs and the subsequent coupling of plasmons to the free-space radiation upon interaction with single [7] or multiple [8] metal gates. The periodically gated 2DESs typically demonstrate emissions of higher power and narrower linewidth [4,9,10] compared to the plasmonic transistors with a single gate. Despite these experimental advances, no accepted theory on the mechanism of plasmon self-excitation in grating-gated plasmonic nanostructures exists.

Early works have suggested the excitation of plasmons by hot electrons [11]. However, in the latest experiments [4] the emission starts in a thresholdlike manner, which signifies the occurrence of plasma instability. In the simplest case of dc electron flow in a 2DES parallel to the conducting gate, the dissipative instabilities [12] can develop at a drift velocity equal to the plasmon velocity. A similar estimate of threshold velocity was obtained for amplified transmission of radiation through a periodically gated 2DES with uniform density [5,6,13]. Such high velocity can hardly be achieved in an experiment, particularly due to the choking of electron flow [14].

The onset of terahertz emission in grating-gated 2DESs at low longitudinal electric field (∼1 kV/cm in [10]) motivated the search for the low-threshold plasmon instabilities. In Refs. [15–17] it was supposed that the latter can emerge due to the transit time effects in the high-field domains of 2DESs. However, the transit time effects generally require deviations from the linear relation between the current and electric field, e.g., due to velocity saturation. The voltage drop across each cell of an experimentally relevant grating-gated 2DES [9] is less than 20 mV, and the transit time effects should be suppressed at such voltages.

A new class of plasma instabilities based on amplified plasmon reflection in bounded 2DESs was put forward by Dyakonov and Shur [7]. Their threshold velocity is limited only by the carrier scattering by impurities or phonons and can be made very low in sufficiently clean systems. At the same time, the Dyakonov-Shur (DS) instability relies on essentially asymmetric boundary conditions at the 2DES contacts: the impedance at the drain should be greater than that at the source [18]. Such asymmetry is not present in a weakly biased *isolated* gated cell in an infinite 2DES [Fig. 1(b)].

In the present paper, we theoretically show that the reflection-type plasma instabilities can develop in *periodically* gated 2DES (plasmonic crystals) with a modulated electron density [see Fig. 1(a)]. We find that the traveling waves with quasimomentum not at the edge of the plasmonic Brillouin zone are generally more unstable than purely periodic waves considered in Refs. [6,13]. We show that the drift velocity required for the emergence of unstable modes can be well below the plasmon phase velocity. This is in contrast to the case of "plasmonic boom" instabilities in all-gated plasmonic crystals with a varying carrier density [19] or varying width [20] that occur at "superplasmonic" drift velocities. Remarkably, the proposed mechanism of instability requires neither transit-time nor velocity saturation effects [15,16]. It can thus

*aleksandr.petrov@phystech.edu

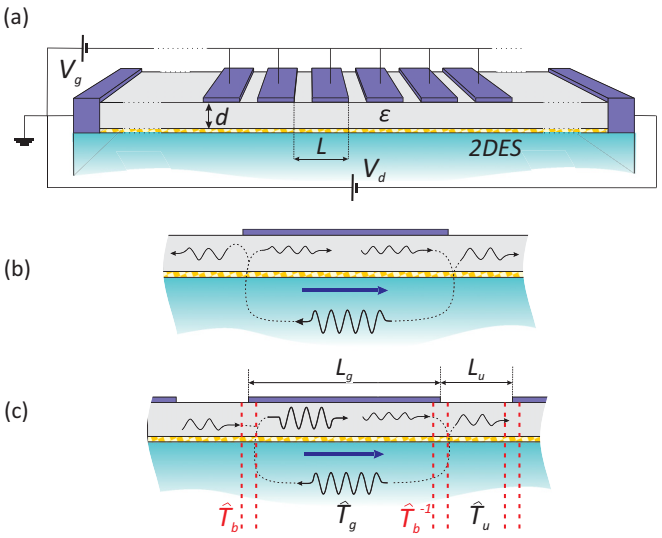

FIG. 1. (a) Schematic view of a field-effect transistor with a two-dimensional conducting channel and periodic grating-gate structure. Schematic view of plasmon reflections from gated/ungated boundaries (b) under an isolated gate in an infinite 2DES and (c) in the multigate structure. The reflection from the right boundary leads to the wave amplification under direct current, while the reflection from the left boundary leads to the wave attenuation in (b). However, this attenuation can be compensated by the wave incident from the neighboring cell in (c). The symbols $\hat{T}_b$, $\hat{T}_g$, $\hat{T}_b^{-1}$, and $\hat{T}_u$ stand for the transfer matrices describing the plasmon propagation across the parts of plasmonic crystal.

be responsible for the plasmon instabilities and THz emission in grating-gate structures with graphene and GaN channels, where the critical field and saturation velocity are very large [21].

Our mechanism of instability can be understood as follows. In a single gated cell of a plasmonic crystal, the downstream plasmon undergoes an amplified Dyakonov-Shur reflection from the gated/ungated boundary [22,23]. In an isolated gated cell, the reflected upstream wave would be attenuated upon the reflection from the opposite boundary. But under proper phase-matching conditions in multigate structures, the fraction of the plasmon energy from the previous cell can compensate the reflection loss of the upstream wave [Fig. 1(c)]. This periodic amplified reflection results in the net instability.

In Sec. II, we derive the generic dispersion equation for the plasmons in a periodically gated 2DES in the presence of the electron drift and discuss the stability of the solutions. In Sec. III, we find the eigenfrequencies and the growth rates of the unstable plasmonic modes in the hydrodynamic limit. Section IV discusses the possible experimental manifestations of this instability and further extensions of the model.

II. PLASMON DISPERSION AND CONDITIONS OF INSTABILITY

To provide a quantitative picture of plasmon instability in the structure shown in Fig. 1(a), we derive the dispersion equation for the traveling waves with a nonzero Bloch phase $\theta = qL$, where q is the quasimomentum and L is the length of the plasmonic crystal cell. According to the microscopic studies of wave reflection at the gated/ungated boundary [24], the net amplitude of the wave can be approximated as a sum of the "fast" downstream and "slow" upstream plasmons in both the gated and ungated sections. This so-called quasioptical approximation provides sufficient accuracy for long ungated sections and/or high frequencies [25]. We denote the plasmon wave vectors as $k_\pm^{\mathrm{g,u}}$, where the plus sign stands for the downstream waves, the minus sign stands for the upstream waves, and the superscripts g and u denote the gated and ungated sections, respectively. The amplitudes of electric potential in these waves are denoted as $\delta\varphi_\pm^{\mathrm{g,u}}$.

Let us compose the vectors $\delta\boldsymbol{\varphi}^{\mathrm{g,u}} = \{\delta\varphi_+^{\mathrm{g,u}}, \delta\varphi_-^{\mathrm{g,u}}\}^{\mathrm{T}}$. In the neighboring cells of the crystal, they can differ only by the factor $e^{i\theta}$. On the other hand, they are related via the transfer matrix of the unit cell $\hat{T}$, $\delta\boldsymbol{\varphi}_N = \hat{T}\delta\boldsymbol{\varphi}_{N+1}$. This leads us to the general dispersion equation [26]

$$\det(\hat{T} - \hat{I}e^{i\theta}) = 0, \tag{1}$$

where $\hat{I}$ is the identity matrix. The transfer matrix of the unit cell is represented as the product of four matrices indicated in Fig. 1(c), $\hat{T}_{\mathrm{u}}$, characterizing the ungated section; the matrix of the wave reflection and transmission at the ungated/gated boundary $\hat{T}_{\mathrm{b}}$; the transfer matrix of the gated part $\hat{T}_{\mathrm{g}}$; and $\hat{T}_{\mathrm{b}}^{-1}$ for another boundary:

$$\hat{T} = \hat{T}_{\mathrm{u}} \cdot \hat{T}_{\mathrm{b}} \cdot \hat{T}_{\mathrm{g}} \cdot \hat{T}_{\mathrm{b}}^{-1}. \tag{2}$$

The transfer matrices of free wave propagation $\hat{T}_{\mathrm{g,u}}$ have the diagonal form

$$\hat{T}_{\mathrm{g,u}} = \begin{pmatrix} e^{ik_+^{\mathrm{g,u}}L_{\mathrm{g,u}}} & 0 \\ 0 & e^{ik_-^{\mathrm{g,u}}L_{\mathrm{g,u}}} \end{pmatrix}, \tag{3}$$

where L_{g} and L_{u} are the lengths of the gated and ungated regions, respectively. Keeping in mind the flow-induced nonreciprocity, we can present the T matrix describing the boundary as [27,28]

$$\hat{T}_b = \frac{1}{t_+}\begin{pmatrix} 1 & -r_- \\ r_+ & t_-t_+ - r_-r_+ \end{pmatrix}, \tag{4}$$

where $r_{+(-)}$ and $t_{+(-)}$ are the reflection and transmission coefficients of the waves incident from the ungated (gated) region. In the absence of drift, $r_+ = -r_-$. In the presence of current, the reflection coefficients r_+ and r_- have different moduli, in particular due to the difference of the forward and backward wave velocities.

Knowing all elements of the T matrices, one readily obtains the dispersion equation for plasmons in a drifting 2DES with a grating gate:

$$\begin{aligned} &\cos\left(\theta + \frac{k_+^{\mathrm{g}} + k_-^{\mathrm{g}}}{2}L_{\mathrm{g}} + \frac{k_+^{\mathrm{u}} + k_-^{\mathrm{u}}}{2}L_{\mathrm{u}}\right) \\ &= \cos\left(\frac{k_+^{\mathrm{g}} - k_-^{\mathrm{g}}}{2}L_{\mathrm{g}}\right)\cos\left(\frac{k_+^{\mathrm{u}} - k_-^{\mathrm{u}}}{2}L_{\mathrm{u}}\right) \\ &\quad - Z\sin\left(\frac{k_+^{\mathrm{g}} - k_-^{\mathrm{g}}}{2}L_{\mathrm{g}}\right)\sin\left(\frac{k_+^{\mathrm{u}} - k_-^{\mathrm{u}}}{2}L_{\mathrm{u}}\right), \end{aligned} \tag{5}$$

where

$$Z = 1 - 2\,r_+r_-/(t_+t_-) \tag{6}$$

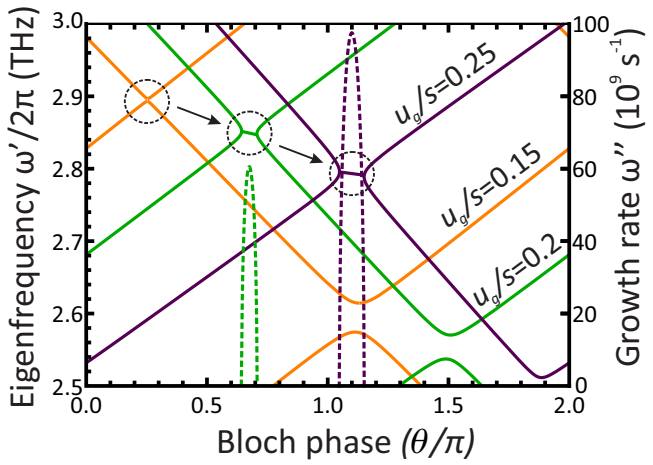

FIG. 2. Dispersion curves $\omega'(\theta)$ (solid lines, left scale) and plasmon growth rates $\omega''(\theta)$ (dashed lines, right scale) calculated for a GaAs-based 2DES with the grating gate at different drift velocities u_g normalized by the plasma-wave velocity s. The instability develops from the stable anticrossing at $u_g/s = 0.15$ through the merging of the dispersion curves ($u_g/s \approx 0.17$) to the unstable anticrossing ($u_g/s = 0.2$). The threshold velocity for instability $u_{th} \approx 0.17\,s$. Dashed circles highlight the plasmonic band gap being transformed into the instability domain with increasing velocity.

is the "modulation depth" factor. Equation (5) is generic, and its functional form does not depend on the transport properties in gated and ungated sections and at the boundaries. In the absence of a drift, it is similar to the plasmon dispersion in a fully gated 2DES with a modulated density [19] or to the photon dispersion in one-dimensional photonic crystals [29] and electron dispersion in the Kronig-Penney potential. All the information about "bulk" carrier transport in Eq. (5) is contained in the frequency-dependent plasmon wave vectors $k^{\mathrm{g}}(\omega)$ and $k^{\mathrm{u}}(\omega)$, while all the information about boundary transport is enclosed in the coefficients r and t.

In what follows, we shall solve Eq. (5) for complex plasmon frequencies $\omega = \omega' + i\omega''$ at real Bloch phase θ. The appearance of roots with a positive imaginary part $\omega'' > 0$ would imply the exponential growth of the plasmon amplitude in time, $\delta\varphi \propto e^{\omega'' t}$, i.e., the plasmon instability. It appears that the instability conditions for the waves described by (5) can be derived in a very general form. As the system approaches the instability with increasing drift velocity, the stable anticrossing of the plasmon bands transforms into an unstable one by passing through the gap shrinkage (Fig. 2) [30]. For a real-valued coefficient Z, this shrinkage occurs when (1) the absolute value of the right-hand side of Eq. (5) equals unity, which is the condition of plasmonic band gap, and (2) the frequency derivative of the right-hand side equals zero, which is the condition of infinite smallness of the gap. In addition, if the frequency dependence of the modulation depth is weak, which will be justified in the next section, these conditions are equivalent to

$$Z = 1, \tag{7}$$

$$\frac{k_+^{\mathrm{g}} - k_-^{\mathrm{g}}}{2} L_{\mathrm{g}} + \frac{k_+^{\mathrm{u}} - k_-^{\mathrm{u}}}{2} L_{\mathrm{u}} = \pi m,\ m \in Z. \tag{8}$$

Physically, the condition $Z = 1$ corresponds to the perfect energy transfer from the wave in the ungated section to the wave in the gated one, $r_+ = 0$.

III. ANALYSIS OF THE DISPERSION RELATION

The usefulness of Eq. (5) stems from the fact that it can be applied to a wide class of two-dimensional systems once their plasmon dispersion relations $k(\omega)$ in the presence of drift are known. The calculation of the reflection and transmission coefficients governing the value of Z can also be done by various methods differing in complexity and accuracy [24,31].

For numerical estimates of the critical velocity and instability growth rates, we use the frequency dependencies of the wave vectors $k_\pm^{\mathrm{g,u}}(\omega)$ obtained within the hydrodynamic model. The latter is justified when carrier-carrier collision frequency exceeds the plasma frequency and the frequency of the carrier collisions with impurities and phonons. Both the theoretical estimates [7,32] and experiments [33,34] support the applicability of the hydrodynamic model up to the terahertz frequencies in III–V 2DESs and graphene [35]. The simultaneous solution of Poisson, Euler, and continuity equations (see Appendix A for details) leads us to

$$k_\pm^{\mathrm{g}} = \frac{\omega}{u_{\mathrm{g}} \pm s}, \tag{9}$$

$$k_\pm^{\mathrm{u}} = \frac{\omega u_{\mathrm{u}} \pm a \mp \sqrt{a^2 \pm 2a\omega u_{\mathrm{u}}}}{u_{\mathrm{u}}^2}, \tag{10}$$

where $u_{\mathrm{g,u}}$ are the carrier drift velocities, $s = \sqrt{eV_g/m^*}$ is the plasma wave velocity in the absence of drift, $a = \pi e^2 n_u/(\varepsilon m^*)$, n_u is the carrier density in the ungated region, V_g is the gate-to-channel bias, ε is the gate dielectric constant, and m^* is the electron effective mass. The latter is taken to be $0.067m_0$, which is appropriate for GaAs [36], a widely used material for high-electron-mobility transistors.

The determination of the reflection and transmission coefficients requires an imposition of the boundary conditions for the electric potential, drift velocity, and carrier density at the gated/ungated interface. If the electron transport obeyed the hydrodynamic equations at the transient regions as well, the boundary conditions for the determination of r and t would represent the continuity of (1) current and (2) carrier energy [19,20]. The latter may be obtained by integrating the Euler equation across the transient region. However, the length of the transient regions is comparable to the screening length in the 2DES and is generally smaller than the collision-limited free path. This makes the ballistic description of the transport at the boundary favorable to the hydrodynamic approach.

Within the ballistic approach, the current across the boundary is calculated as the difference of particle fluxes supplied by the gated and ungated sections. As a result, the variation of current δj becomes a linear function of the quasi-Fermi-level drop across the boundary, $\delta F_{\mathrm{g}} - \delta F_{\mathrm{u}}$ [37], which should be used as the second boundary condition. The explicit form of this relation is presented in Appendix B, Eq. (B11). However, the reflection coefficient r can be determined with sufficient accuracy (see Fig. 6) if we simply require the continuity of electric potential across the gated/ungated boundary,

$\delta\varphi_g = \delta\varphi_u$. This results in the following expressions for the reflection and transmission coefficients:

$$r_+ = -\frac{1-2k_+^g d}{1+2k_+^g d}, \quad r_- = -\frac{k_+^g}{k_-^g}\frac{1+2k_-^g d}{1+2k_+^g d}, \tag{11}$$

$$t_+ = \frac{4k_+^g d}{1+2k_+^g d}, \quad t_- = \frac{1-k_+^g/k_-^g}{1+2k_+^g d}. \tag{12}$$

With the above reflection and transmission coefficients, the dependence $Z(\omega)$ has a smooth minimum with the minimal value below unity at any nonzero flow velocity, which justifies the neglect of $dZ/d\omega$ in the derivation of the instability condition. As seen from Eqs. (6) and (7), the threshold velocity for the instability corresponds to the reflectionless passage of the upstream plasmon from the ungated to the gated sections ($r_+ = 0$). At given frequency ω, this velocity is

$$u_{th} = |s - 2\omega d| \approx s\left|1 - 4\pi\frac{d}{\lambda}\right|, \tag{13}$$

where λ is the plasmon wavelength in the gated section. In typical experiments [38], the plasmon wavelength λ is on the order of hundreds of nanometers, while the gate-to-channel separation d is several tens of nanometers. Despite the fact that the ratio d/λ is usually small, a large prefactor of 4π in Eq. (13) provides an extra order of magnitude to this ratio, and the instability can develop at drift velocities far below the plasmon velocity s.

These findings are substantiated in Fig. 2, which shows the calculated plasmon eigenfrequency $\omega'(\theta)$ and growth rate $\omega''(\theta)$ in GaAs-based 2DES under the grating gate at different drift velocities. The lengths of the gated and ungated sections are $L_g = 0.6\,\mu$m and $L_u = 0.25\,\mu$m, respectively, the gate-to-channel separation is $d = 10$ nm, gate dielectric permittivity $\varepsilon = 12.9$, and the carrier densities are $n_g = 5\times10^{11}\,\mathrm{cm}^{-2}$ and $n_u = 2\times10^{12}\,\mathrm{cm}^{-2}$. The onset of the instability with increasing drift velocity represents a transformation of a stable-type plasmon band anticrossing to the unstable one via passing through the gapless plasmon bands. In the unstable case, the neighboring branches of the dispersion curves merge through the complex plane with the two complex conjugate solutions for each Bloch phase (in Fig. 2, the solutions with $\omega'' < 0$ are not shown). For the parameters in Fig. 2, the unstable mode at 2.9 THz appears at $u_{th} = 0.17$ s, in agreement with Eq. (13).

Above the threshold velocity, the waves are unstable for a finite range of the quasiwave vectors. Generally, these unstable wave vectors lie away from the edges of the Brillouin zone; that is, their Bloch phase $\theta = qL \neq 0, 2\pi, 4\pi$. The reason is that the extrema of the plasmonic bands (being at the edges of the Brillouin zone at zero drift velocity) are shifted away by the Doppler effect. Within the unstable domains, the real part of the plasmon frequency ω' almost does not depend on the quasiwave vector, while its imaginary part ω'' varies abruptly. Expanding the dispersion equation (5) near the band anticrossing at $\theta \approx \theta_{cr}$ and $\omega \approx \omega_{cr}$, we find

$$\omega''^2 = \frac{2|\theta-\theta_{cr}|}{(d\alpha/d\omega)^2}|\tan[\theta_{cr} + \alpha(\omega_{cr})]|, \tag{14}$$

where $\alpha(\omega) = (k_+^g + k_-^g)L_g/2 + (k_+^u + k_-^u)L_u/2$. Equation (14) describes the square-root growth of the imaginary part of the plasmon frequency above the critical Bloch phase observed in Fig. 2.

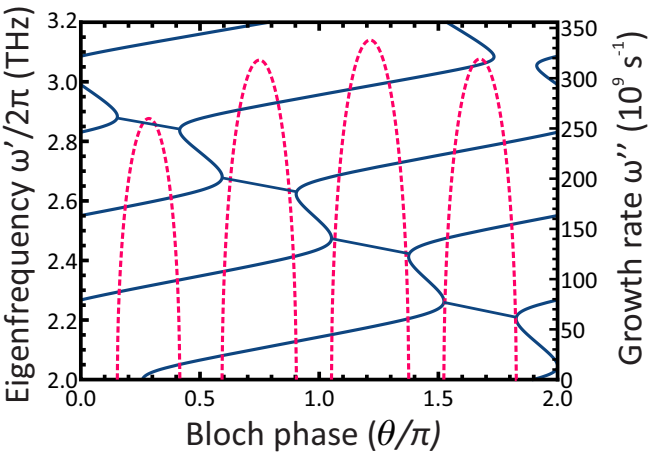

FIG. 3. Dispersion curves $\omega'(\theta)/2\pi$ (solid lines, left scale) and instability growth rates $\omega''(\theta)$ (dashed lines, right scale) at high drift velocity $u_g = 0.59s$ and other parameters as in Fig. 2.

As the drift velocity increases, the branches of plasmon dispersion become denser, as illustrated in Fig. 3. This is explained by the fact that the backward wave vector in the gated region grows as $k_-^g \propto 1/(u_g - s)$, and at large velocities even a slight variation of the frequency results in a strong variation of the right-hand side of dispersion equation (5). One can also observe that the highest plasmon growth rate is achieved at the frequency for which the modulation depth Z has the minimal value. This is seen from an arc-shaped envelope of the dependencies $\omega''(\theta)$ in Fig. 3: the frequency of ~2.4 THz corresponds to the minimum of $Z(\omega)$ and to the maximum of the unstable mode growth rate.

It is remarkable that the maximization conditions for the instability growth rate can be derived analytically. This maximum is achieved in the middle of the unstable domain with respect to the quasimomentum θ, and the wider the instability domain is, the larger the growth rate is. The center of the unstable domain corresponds to zero-frequency derivatives of both the left-hand (lhs) and right-hand (rhs) sides of Eq. (5). Moreover, the instability growth rate can be maximized with respect to all other parameters of the problem (gate length, carrier density, etc.) by requiring the maximum difference of the lhs and rhs of Eq. (5) in the middle of the unstable domain. In this case, the imbalance between the rhs and lhs as a function of real frequencies can be compensated only via the introduction of a sufficiently large imaginary part of the frequency.

The superposition of these requirements leads us to the following conditions for the growth rate maximization:

$$Z \to \min, \tag{15}$$

$$\frac{k_+^g - k_-^g}{2}L_g = \frac{\pi}{2} + \pi n,\ n \in Z, \tag{16}$$

$$\frac{k_+^u - k_-^u}{2}L_u = \frac{\pi}{2} + \pi m,\ m \in Z. \tag{17}$$

The two latter equations can be considered antireflection conditions for coatings represented by gated and ungated regions and, at the same time, the Dyakonov-Shur eigenfrequency conditions for the gated and ungated plasma resonators [7]. This supports our interpretation of the instability as an amplified DS reflection supplemented by the perfect energy transfer between the cells of a plasmonic crystal.

We further notice that the flow-induced corrections to the phases in Eqs. (16) and (17) are order $O(u^2_{g,u})$. When the carrier density in the ungated region exceeds that in the gated region (which corresponds to the range of parameters considered), the phase (17) can be considered flow independent. This leads us to the eigenfrequency providing the highest growth rate (at $m = 0$)

$$\omega'_{\rm res} = \sqrt{\frac{\pi^2 e^2 n_u}{\varepsilon m^* L_u}}. \tag{18}$$

The growth rate $\omega''_{\rm max}$ at this frequency is obtained via separating the real and imaginary parts of Eq. (5) and expanding them with respect to ω'' and u. This leads us to

$$\omega''^2_{\rm max} = \frac{2[1 - Z(\omega'_{\rm res})]}{Z(\omega'_{\rm res})T_1^2 + 4T_1T_2 - (T_1u_1/s)^2}, \tag{19}$$

where $T_1 = sL_g/(s^2 - u_g^2)$ and $T_2 = \omega'_{\rm res}L_u/2a$.

The dependence of the maximum growth rate on the drift velocity and the length of the ungated region is shown in Fig. 4. The gate-to-channel separation equals 20 nm, and the gate length is 0.17 μm. It can be seen that a length of ungated domain L_u^* exists and the corresponding resonant frequency $\omega'^*_{\rm res} = s/2d$ such that the development of instability is thresholdless. At higher drift velocities, the highest value of ω'' is also achieved roughly at that frequency. The regions filled with white in Fig. 4 correspond to the stability of modes with frequency $\omega'_{\rm res}$; the equation for the boundary between stable and unstable domains is readily obtained by substituting $\omega = \omega'_{\rm res}$ into the condition $Z(\omega,u) = 1$.

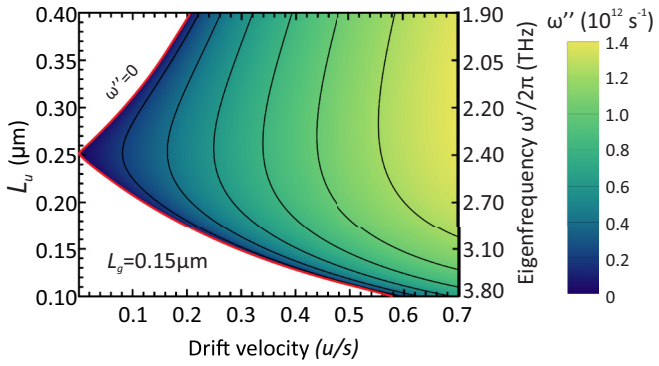

FIG. 4. Color map of the instability growth rate vs drift velocity and the length of the ungated region at the resonant frequency [Eq. (18)]. Gate length $L_g = 0.17$ μm; gate-to-channel separation $d = 20$ nm. The instability at $L_u = 0.25$ μm is thresholdless with respect to the drift velocity [see Eq. (13)] and has the highest growth rate.

IV. DISCUSSION

For a realistic estimate of the instability growth rate one has to include damping due to carrier-phonon and carrier-impurity scattering. This can generally be achieved by adding a friction term in the Euler equation. For high momentum relaxation rates τ_p ($\omega\tau_p \gg 1$) the scattering typically reduces the growth rate calculated within the collisionless model by $1/2\tau_p$ [7]. The compensation of collisional damping by the instability growth in GaAs at 77 K [mobility $\mu = 2 \times 10^5$ cm^2/(Vs)] occurs at $\omega'' = 65 \times 10^9$ s^{-1}. Under optimal conditions, this corresponds to the velocity $u^* = 0.03\,s = 1.8 \times 10^4$ m/s, which is well below the carrier saturation velocity of 3×10^5 m/s. The reduction in the saturation velocity with increasing temperature limits the critical temperature for plasmon instability in GaAs to approximately 200 K. However, in 2DESs based on different III–V semiconductors, the instability is feasible up to room temperature. One example is InAs with room-temperature mobility $\mu = 2.6 \times 10^4$ cm^2/(V s) [39] corresponding to the compensation of damping at the growth rate $\omega'' = 1.5 \times 10^{12}$ s^{-1}. This growth rate can be achieved at $u^* \approx 3.4 \times 10^5$ m/s, which is more than two times smaller than the carrier saturation velocity.

Another aspect of carrier relaxation is the voltage drop along the channel, which may distort the uniform carrier density within a single cell of the plasmonic crystal. From the above estimates for InAs at 300 K, we find this voltage drop to be $\Delta V \approx u^*L_g/\mu = 11$ mV, which is indeed small compared to the gate voltage required to support the given carrier density under the gates $V_g = ms^2/e = 210$ mV. A further increase in drift velocity requires considerably larger voltage drops due to the saturation effects. Particularly, $u = 3 \times 10^5$ m/s is achieved at $\Delta V \approx 40$ mV. While the density uniformity under a single gate is maintained even in high fields, the uniformity over the whole crystal can be supported by applying a gradually changing voltage to the gates with the aid of separative capacitors [19].

The maximum attainable growth rate in Fig. 4 of 1.2×10^{12} s^{-1} corresponds to the momentum relaxation time of 0.4 ps. This time is much shorter than the electron-phonon scattering time in graphene at room temperature, $\tau_p \approx 3$ ps [40]. Hence, one might expect the development of amplified-reflection instabilities in graphene-based transistors with grating gates [41]. We note, however, that the study of plasma instabilities in graphene requires an essential modification of hydrodynamic equations [32,42] and will be left for future work.

For a quantitative comparison of the presented model and experimental data on THz emission in grating-gate structures [4] one needs to consider the reflection of the unstable traveling waves at terminals of plasmonic crystal. The geometrical asymmetry of the plasmonic crystal unit cell (the presence of two gates of unequal length) should also be taken into account. This can be done within the developed transfer matrix formalism, although the resulting dispersion equations are quite cumbersome. The geometrical asymmetry was shown to be crucial for efficient THz detection in plasmonic field-effect transistors [43] and is expected to be beneficial for achieving the low-threshold instabilities [17], although the full theory of the asymmetry effect on the instability has yet

to be developed. Further possible extensions of our model include the renouncement of quasioptical approximation and full electrodynamic treatment of the plasmon reflection at the boundaries, including the excitation of the evanescent waves [24]. Within the same formalism, one can also consider the self-excitation of the edge plasmons traveling along the gated/ungated boundary [22], which might have larger instability growth rates compared to the bulk modes.

In conclusion, we have theoretically demonstrated the instability of direct current flow in grating-gated 2DESs against the excitation of traveling plasmons. The mechanism of instability is associated with the amplified Dyakonov-Shur-type plasmon reflection from gated/ungated boundaries and proper phase matching of plasmons under the neighboring gates. Using the transfer matrix approach, we have derived the generic dispersion relation for plasmonic crystals with flow-induced nonreciprocity. In a particular case of alternating gated and ungated regions of a 2DES, this equation has unstable solutions at flow velocities which can be well below the plasma wave and saturation velocities. The growth rate of predicted instability is of the order of (but not limited to) $1.2 \times 10^{12}\ \mathrm{s}^{-1}$, which makes the instability feasible in InAs-based 2DESs and in graphene at room temperature.

ACKNOWLEDGMENTS

This work was supported by Grant No. 16-19-10557 of the Russian Science Foundation. The authors thank V. Leiman, V. Kachorovskii, and V. Enaldiev for helpful discussions.

APPENDIX A: PLASMON DISPERSION IN A DRIFTING TWO-DIMENSIONAL ELECTRON GAS

In this appendix, we sketch the derivation of Eqs. (9) and (10). The details can be found in Refs. [7,12]. The system of hydrodynamic equations describing one-dimensional transport in semiconductor plasma reads:

$$\partial_t n + \partial_x (nu) = 0, \tag{A1}$$

$$\partial_t u + u\partial_x u = (e/m^*)\partial_x \varphi, \tag{A2}$$

where n is the carrier density, u is the carrier drift velocity, φ is the local electric potential, and ∂_t and ∂_x stand for the partial derivatives with respect to the time and coordinate. Poisson's equation reads

$$\partial_z^2 \varphi + \partial_x^2 \varphi = 4\pi e n \delta(z), \tag{A3}$$

where the z axis is directed normally to the 2DES plane located at $z = 0$. In the gated section, the boundary conditions are $\varphi|_{z=d} = 0$, $\varphi_{z=-\infty} = 0$, where d is the gate-to-channel separation; in the ungated section, the boundary condition is $\varphi_{z\pm\infty} = 0$. Assuming small harmonic perturbations of all quantities, $\varphi = \delta\varphi e^{i(kx-\omega t)}$, $u = u_0 + \delta u e^{i(kx-\omega t)}$, $n = n_0 + \delta n e^{i(kx-\omega t)}$, we find the linearized equations of motion for the electron plasma:

$$-i(\omega - ku_0)\delta n + ikn_0 \delta u = 0, \tag{A4}$$

$$-i(\omega - ku_0)\delta u = ik(e/m^*)\delta\varphi. \tag{A5}$$

The solution of Poisson's equation is presented as

$$-C_{\mathrm{g,u}}\delta\varphi = e\delta n, \tag{A6}$$

where $C_{\mathrm{g,u}}$ are the effective capacitances of the gated and ungated sections, $C_g = \varepsilon/(4\pi d)$ provided $kd \ll 1$, and $C_u = k\varepsilon/(2\pi)$. The criterion of consistency for Eqs. (A4)–(A6) leads us to the dispersion laws (10). We also note that Eqs. (9) and (10) can be obtained by applying a Galilean transform $\omega \to \omega - ku_0$ to the plasmon dispersion law in the 2DES $\omega^2 = (4\pi n_0 e^2 k/\varepsilon m^*)[1 + \coth(kd)]^{-1}$ [44] in the limits $kd \ll 1$ and $kd \gg 1$, respectively.

APPENDIX B: AMPLIFIED REFLECTION AT THE GATED/UNGATED BOUNDARY

In this appendix we establish the reflection coefficient of a plasma wave incident from the gated section to the gated/ungated boundary. An example of the band diagram of the transient region between these sections is shown in Fig. 5; the wave is assumed to be incident from the left. We shall show that plasmon reflection from the boundary can lead to the wave amplification under direct current flow, similar to the reflection from the drain side of the Dyakonov-Shur transistor [7]. It is remarkable that the amplified reflection persists for various regimes of carrier transport across the boundary, including (1) the ballistic transport, in which the free path between electron collisions l_{ee} is larger than the length of the transient region, and (2) the hydrodynamic transport, in which l_{ee} is the shortest length in the system. We also note that the amplified reflection is also described with sufficient accuracy if we require the continuity of electric potential across the boundary.

Hydrodynamic transport across the boundary. If the hydrodynamic equations were valid across the whole plasmonic crystal, the boundary conditions matching the transport in gated and ungated parts of the channel would require the continuity of current,

$$j = nu = \mathrm{const}, \tag{B1}$$

and the continuity of carrier energy [19,20],

$$E = m^* u^2/2 - e\varphi = \mathrm{const}. \tag{B2}$$

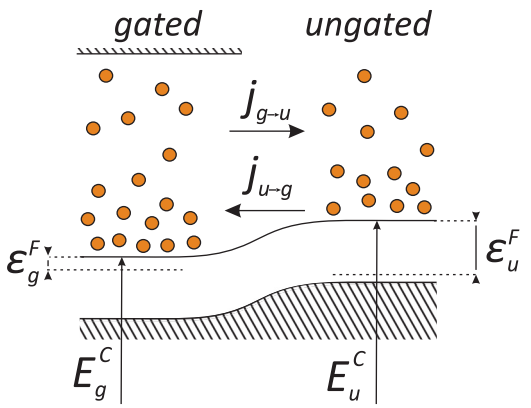

FIG. 5. Schematic band diagram of the junction between the gated (left) and ungated (right) regions. $E^C_{g,u}$ denote the positions of conduction-band bottoms, and $\varepsilon^F_{g,u}$ are the respective Fermi energies.

The latter is nothing but the Bernoulli equation for electron plasma neglecting the pressure term. We present the wave amplitude in the gated section as a sum of incident and reflected waves, $\delta\varphi_g = \delta\varphi_+^g e^{ik_+^g x} + \delta\varphi_-^g e^{ik_-^g x}$. In the ungated section, we retain only the transmitted wave $\delta\varphi_u = \delta\varphi_+^u e^{ik_+^u x}$. Expressing the current and carrier energy via the potential with Eqs. (A4), (A5), and (A6), we arrive at the continuity relations

$$\frac{\delta\varphi_+^g}{2k_+^g d} + \frac{\delta\varphi_-^g}{2k_+^g d} = \delta\varphi_+^u, \tag{B3}$$

$$\frac{\delta\varphi_+^g}{\omega - k_+^g u_g} + \frac{\delta\varphi_-^g}{\omega - k_-^g u_g} = \frac{\delta\varphi_+^u}{\omega - k_+^u u_u}. \tag{B4}$$

The reflection coefficient under the requirement of energy continuity $r_E = \delta\varphi_-^g/\delta\varphi_+^g$ is expressed as

$$r_E = \frac{s+u_0}{s-u_0}\frac{1-2\frac{d\omega}{s}\left(1-\frac{k_+^u u_u}{\omega}\right)}{1+2\frac{d\omega}{s}\left(1-\frac{k_+^u u_u}{\omega}\right)}. \tag{B5}$$

The first fraction (exceeding unity) describes the amplified Dyakonov-Shur reflection [7], while the second one describes possible wave leakage from the gated section.

Ballistic transport across the boundary. In the ballistic limit, the net current across the boundary is the difference of carrier fluxes supplied by the gated ($j_{g\to u}$) and ungated ($j_{u\to g}$) regions, as shown in Fig. 5:

$$j_{g\to u} = \frac{2}{(2\pi\hbar)^2}\int\limits_{p_x>p_{cr}} v_x f(\mathbf{p})\mathrm{d}^2\mathbf{p}, \tag{B6}$$

$$j_{u\to g} = \frac{2}{(2\pi\hbar)^2}\int\limits_{p_x<0} v_x f(\mathbf{p})\mathrm{d}^2\mathbf{p}, \tag{B7}$$

where v and p are the electron velocity and momentum, respectively, $p_{cr} = \sqrt{2m(E_u^c - E_g^c)}$ is the minimal carrier momentum required to overcome the barrier at the boundary, and E_u^c and E_g^c are the positions of the conduction-band bottom in the respective regions. In accordance with the hydrodynamic description in the bulk, we take the distribution function in the local equilibrium form

$$f(\mathbf{p}) = \exp\left[-\frac{(\mathbf{p}-m^*\mathbf{u})^2}{2m^*T}\right], \tag{B8}$$

where $\mathbf{u}$ is the drift velocity and T is the temperature in the energy units. Evaluating the integrals, we find the net current $j = j_{g\to u} + j_{u\to g}$ across the boundary:

$$j = n_g\left[\frac{e^{-\xi_g^2}}{2\sqrt{\pi}}v_T + \frac{\mathrm{erfc}(\xi_g)}{2}u_g\right] + n_u\left[\frac{e^{-\xi_u^2}}{2\sqrt{\pi}}v_T + \frac{\mathrm{erfc}(\xi_u)}{2}u_u\right]. \tag{B9}$$

Here $\xi_g = (v_{cr}-u_g)/v_T$, $\xi_u = u_u/v_T$, $v_T = \sqrt{2T/m^*}$, $v_{cr} = p_{cr}/m^*$, $\mathrm{erfc}(x) = 2/\sqrt{\pi}\int_x^\infty e^{-y^2}dy$ is the complementary error function, and $n_{g,u}$ is the carrier density in the respective region.

Assuming small harmonic perturbations of the quantities E_g^c, E_u^c, n_g, n_u, u_g, and u_u in Eq. (B9), we find the microscopic boundary condition relating the ac current across the boundary δj to the ac variations of density, velocity, and electric potential:

$$\delta j = \delta n_g\left[\frac{e^{-\xi_g^2}}{2\sqrt{\pi}}v_T + \frac{\mathrm{erfc}(\xi_g)}{2}u_g\right] + \delta n_u\left[\frac{e^{-\xi_u^2}}{2\sqrt{\pi}}v_T + \frac{\mathrm{erfc}(\xi_u)}{2}u_u\right] + n_g v_T\frac{e^{-\xi_g^2}}{2\sqrt{\pi}}\frac{\left(\delta E_g^C - \delta E_u^C\right)}{T} + n_g\delta u_g\left[\frac{1}{\sqrt{\pi}}\frac{v_{cr}}{v_T}e^{-\xi_g^2} + \frac{\mathrm{erfc}(\xi_g)}{2}\right] + n_u\delta u_u\frac{\mathrm{erfc}(\xi_u)}{2}. \tag{B10}$$

The current continuity equation (B3) along with Eq. (B10) allows one to obtain the plasmon reflection coefficient r_b under the assumption of ballistic electron transport:

$$r_b = -\frac{\alpha_+ - \beta/(2k_+^g d)}{\alpha_- - \beta/(2k_-^g d)}, \tag{B11}$$

where

$$\alpha_\pm = \left(s^2 \mp s v_{cr} + v_T^2/2\right)e^{-\xi_g^2} - \sqrt{\pi}v_T(s\pm u_g) + (u_g \mp s)v_T\frac{\sqrt{\pi}\,\mathrm{erfc}(\xi_g)}{2}, \tag{B12}$$

$$\beta = v_T^2 k_+^u d e^{-\xi_u^2} + s^2 e^{-\xi_g^2} + \frac{\sqrt{\pi}}{2}\mathrm{erfc}(\xi_u)v_T u_u \times\left[-2k_+^u d + \frac{s^2}{u_g^2}\left(\frac{n_u}{n_g}\right)^3\frac{k_+^u u_u}{\omega - k_+^u u_u}\right]. \tag{B13}$$

Continuity of electric potential across the boundary. Instead of using the cumbersome microscopic condition (B10), one can require the continuity of quasi-Fermi level across the boundary, which is the boundary condition commonly used in

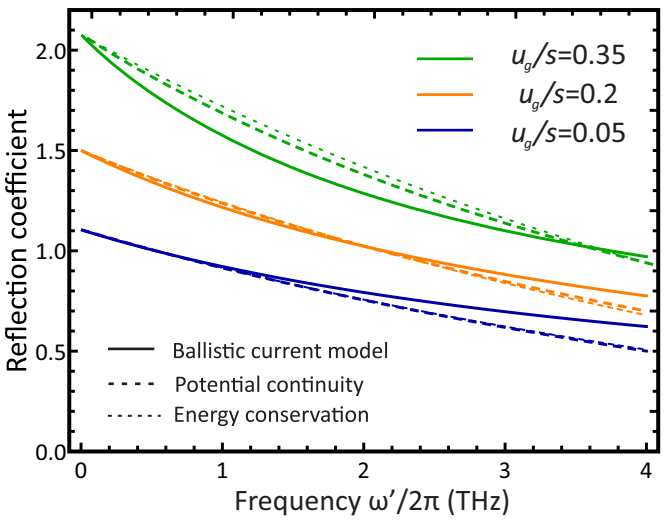

FIG. 6. Calculated plasmon reflection coefficients from the gated/ungated boundary vs frequency at different drift velocities in the gated section (in units of plasma-wave velocity s). Solid lines stand for the reflection coefficients obtained with microscopic calculation of ballistic current at the interface r_b; dashed lines stand for the reflection coefficients obtained by matching of ac potential variations r_φ, and thin dotted lines stand for the coefficients obtained with the energy conservation boundary condition r_E.

modeling transport across the heterojunctions [37]. In addition, the variations of carrier Fermi energy are typically small compared to the variations of electric potential, $\delta\varepsilon_F/e\delta\varphi \approx v_T/s \ll 1$. In such a situation, the continuity of the quasi-Fermi level is equivalent to the continuity of the electric potential. The reflection coefficient for the potential continuity boundary condition r_φ is obtained by combining Eq. (B3) and the potential continuity equation

$$\delta\varphi_+^g + \delta\varphi_-^g = \delta\varphi_+^u. \tag{B14}$$

This leads us to

$$r_\varphi = -\frac{k_-^g}{k_+^g}\frac{1-2k_+^g d}{1-2k_-^g d} = \frac{s+u_g}{s-u_g}\frac{1-\frac{2d\omega}{s+u_g}}{1+\frac{2d\omega}{s-u_g}}, \tag{B15}$$

which is obtained from expression (B11) by setting $\alpha_+ = \alpha_- = \beta = 1$.

The comparison of reflection coefficients r_E, r_b, and r_φ calculated with various boundary conditions is presented in Fig. 6. The discrepancy between all three results is small (less than 10%). This speaks in favor of the fact that the amplified reflection from the gated/ungated interface emerges due to the different dispersion laws of plasmons in gated and ungated domains and is relatively insensitive to the details of transport at the boundary. Due to the close results obtained with various approaches, we use the reflection and transmission coefficients calculated with the potential continuity boundary condition in calculations of plasmon spectra and instabilities.

[1] D. Tsui, E. Gornik, and R. Logan, Solid State Commun. **35**, 875 (1980).
[2] N. Dyakonova, A. El Fatimy, J. Lusakowski, W. Knap, M. I. Dyakonov, M.-A. Poisson, E. Morvan, S. Bollaert, A. Shchepetov, Y. Roelens, C. Gaquiere, D. Theron, and A. Cappy, Appl. Phys. Lett. **88**, 141906 (2006).
[3] A. El Fatimy, N. Dyakonova, Y. Meziani, T. Otsuji, W. Knap, S. Vandenbrouk, K. Madjour, D. Théron, C. Gaquiere, M. Poisson *et al.*, J. Appl. Phys. **107**, 024504 (2010).
[4] T. Otsuji, T. Watanabe, S. A. B. Tombet, A. Satou, W. M. Knap, V. V. Popov, M. Ryzhii, and V. Ryzhii, IEEE Trans. Terahertz Sci. Technol. **3**, 63 (2013).
[5] O. Matov, O. Meshkov, O. Polischuk, and V. Popov, Phys. A (Amsterdam, Neth.) **241**, 409 (1997).
[6] S. A. Mikhailov, Phys. Rev. B **58**, 1517 (1998).
[7] M. Dyakonov and M. Shur, Phys. Rev. Lett. **71**, 2465 (1993).
[8] M. V. Krasheninnikov and A. V. Chaplik, Sov. Phys. JETP **61**, 75 (1985).
[9] T. Otsuji, Y. M. Meziani, M. Hanabe, T. Nishimura, and E. Sano, Solid-State Electron. **51**, 1319 (2007).
[10] T. Otsuji, Y. M. Meziani, T. Nishimura, T. Suemitsu, W. Knap, E. Sano, T. Asano, and V. V. Popov, J. Phys. Condens. Matter **20**, 384206 (2008).
[11] A. Chaplik, Surf. Sci. Rep. **5**, 289 (1985).
[12] M. Krasheninnikov and A. Chaplik, Sov. Phys. JETP **52**, 279 (1980).
[13] K. Kempa, P. Bakshi, H. Xie, and W. L. Schaich, Phys. Rev. B **47**, 4532 (1993).
[14] M. I. Dyakonov and M. S. Shur, Phys. Rev. B **51**, 14341 (1995).
[15] V. Ryzhii, A. Satou, and M. S. Shur, Phys. Status Solidi A **202**, R113 (2005).
[16] V. Ryzhii, A. Satou, M. Ryzhii, T. Otsuji, and M. Shur, J. Phys. Condens. Matter **20**, 384207 (2008).
[17] Y. Koseki, V. Ryzhii, T. Otsuji, V. V. Popov, and A. Satou, Phys. Rev. B **93**, 245408 (2016).
[18] M. V. Cheremisin and G. G. Samsonidze, Semiconductors **33**, 578 (1999).
[19] V. Y. Kachorovskii and M. Shur, Appl. Phys. Lett. **100**, 232108 (2012).
[20] G. R. Aizin, J. Mikalopas, and M. Shur, Phys. Rev. B **93**, 195315 (2016).
[21] I. Meric, M. Y. Han, A. F. Young, B. Ozyilmaz, P. Kim, and K. L. Shepard, Nat. Nanotechnol. **3**, 654 (2008).
[22] M. I. Dyakonov, Semiconductors **42**, 984 (2008).
[23] O. Sydoruk, R. R. A. Syms, and L. Solymar, J. Appl. Phys. **112**, 104512 (2012).
[24] O. Sydoruk, K. Choonee, and G. C. Dyer, IEEE Trans. Terahertz Sci. Technol. **5**, 486 (2015).
[25] M. Karabiyik, R. Sinha, C. Al-Amin, G. C. Dyer, N. Pala, and M. S. Shur, Proc. SPIE **9102**, 91020K (2014).
[26] J. Pendry, J. Mod. Opt. **41**, 209 (1994).
[27] O. Sydoruk, J. B. Wu, A. Mayorov, C. D. Wood, D. K. Mistry, and J. E. Cunningham, Phys. Rev. B **92**, 195304 (2015).
[28] R. E. Collin, *Field Theory of Guided Waves*, 2nd ed. (IEEE Press, New York, 1991), Chap. 3.4.
[29] J. M. Bendickson, J. P. Dowling, and M. Scalora, Phys. Rev. E **53**, 4107 (1996).
[30] L. D. Landau, E. Lifshitz, and L. Pitaevskii, *Course of Theoretical Physics: Physical Kinetics* (Pergamon Press, Oxford, 1993), Chap. 64.
[31] O. Sydoruk, R. Syms, and L. Solymar, Opt. Express **20**, 19618 (2012).
[32] D. Svintsov, V. Vyurkov, S. Yurchenko, T. Otsuji, and V. Ryzhii, J. Appl. Phys. **111**, 083715 (2012).
[33] M. J. M. de Jong and L. W. Molenkamp, Phys. Rev. B **51**, 13389 (1995).
[34] S. Rudin, G. Rupper, A. Gutin, and M. Shur, J. Appl. Phys. **115**, 064503 (2014).
[35] D. A. Bandurin, I. Torre, R. K. Kumar, M. Ben Shalom, A. Tomadin, A. Principi, G. H. Auton, E. Khestanova, K. S. Novoselov, I. V. Grigorieva, L. A. Ponomarenko, A. K. Geim, and M. Polini, Science **351**, 1055 (2016).
[36] M. Levinstein, S. Rumyantsev, and M. Shur, *Handbook Series on Semiconductor Parameters*, Vol. 1, Si, Ge, C, GaAs, GaP, GaSb, InAs, InP, InSb (World Scientific, Singapore, 2012).
[37] K. Horio and H. Yanai, IEEE Trans. Electron Devices **37**, 1093 (1990).
[38] W. Knap, J. Lusakowski, T. Parenty, S. Bollaert, A. Cappy, V. V. Popov, and M. S. Shur, Appl. Phys. Lett. **84**, 2331 (2004).
[39] R. Tsai, M. Barsky, J. B. Boos, B. R. Bennett, J. Lee, N. A. Papanicolaou, R. Magno, C. Namba, P. H. Liu, D. Park, R. Grundbacher, and A. Gutierrez, in *Proceedings of the 25th IEEE*

Gallium Arsenide Integrated Circuit (GaAs IC) Symposium, 2003 (IEEE Press, Piscataway, NJ, 2003), pp. 294–297.

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

[41] P. Olbrich, J. Kamann, M. König, J. Munzert, L. Tutsch, J. Eroms, D. Weiss, M.-H. Liu, L. E. Golub, E. L. Ivchenko, V. V. Popov, D. V. Fateev, K. V. Mashinsky, F. Fromm, T. Seyller, and S. D. Ganichev, Phys. Rev. B **93**, 075422 (2016).

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

[43] V. V. Popov, D. V. Fateev, T. Otsuji, Y. M. Meziani, D. Coquillat, and W. Knap, Appl. Phys. Lett. **99**, 243504 (2011).

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