



Plasmonic drag photocurrent in graphene at extreme nonlocality

Vladimir Silkin  and Dmitry Svintsov 

Laboratory of 2d Materials for Optoelectronics, Moscow Institute of Physics and Technology, Dolgoprudny 141700, Russia



(Received 30 June 2021; revised 27 September 2021; accepted 12 October 2021; published 28 October 2021)

It is commonly assumed that the photocurrent in two-dimensional systems with a centrosymmetric lattice is generated at structural inhomogeneities, such as p - n junctions. Here, we study an alternative mechanism of photocurrent generation associated with the inhomogeneity of the driving electromagnetic field, termed “plasmonic drag.” It is associated with direct momentum transfer from the field to conduction electrons and can be characterized by a nonlocal nonlinear conductivity $\sigma^{(2)}(\mathbf{q}, \omega)$. By constructing a classical kinetic model fully accounting for nonlocality, we show that the nonlinear conductivity is resonantly enhanced for wave phase velocity coinciding with electron Fermi velocity. The enhancement is interpreted as phase locking between electrons and the wave. We discuss a possible experiment where a nonuniform field is created by a propagating graphene plasmon and find an upper limit of the current responsivity vs plasmon velocity. This limit is set by a competition between resonantly growing $\sigma^{(2)}(\mathbf{q}, \omega)$ and the diverging kinetic energy of electrons as the wave velocity approaches the Fermi velocity.

DOI: [10.1103/PhysRevB.104.155438](https://doi.org/10.1103/PhysRevB.104.155438)

I. INTRODUCTION

It is commonly believed that photocurrent generation in two-dimensional (2D) systems without lattice inversion asymmetry occurs at structural inhomogeneities, such as p - n junctions [1–3] and contacts with metals [4,5]. At the same time, the inversion asymmetry of the driving electromagnetic field can lead to the emergence of the photocurrent. The physics beyond such photocurrent is the direct transfer of electromagnetic field momentum to the electrons, termed in the literature the photon drag [6] or the dynamic Hall effect [7,8]. As the optoelectronic structures become deeply sub-wavelength, their electromagnetic response is dominated by near fields. The mechanism of current generation by a field with finite momentum is thus termed plasmonic drag [9–11], implying that momentum is transferred via excitation of 2D plasmons.

Experimental examples of systems supporting drag photocurrent are gate edges [1] and grating couplers without inversion symmetry [12,13] (Fig. 1). It can occur upon tilted illumination of a fully symmetric structure with identical contacts or illumination of one of its contacts which acts as a photon-to-plasmon coupler [14]. Compared to photovoltaic and photothermoelectric effects, the plasmonic drag does not need channel doping nonuniformity. We may therefore speculate that it is the most omnipresent mechanism of photocurrent generation in two dimensions [15].

As the size L of optoelectronic devices shrinks, the characteristic wave number $q \sim L^{-1}$ of electromagnetic near fields (including plasmons) goes up [16]. At large wave vectors, the current-field relations become nonlocal and reflect the dynamics of individual electrons. While the linear nonlocal response of 2D systems has been studied quite well [17–19], the nonlinear processes of photocurrent generation at strong nonlocality have been poorly explored.

Here, we theoretically explore the limits of plasmonic drag in graphene at arbitrarily strong nonlocality. It is realized when the phase velocity of plasma waves v_{ph} approaches the Fermi velocity of electrons v_0 while the plasmon wave vector reaches $q = \omega/v_0$. According to theoretical [20] and experimental [19] works, this is the ultimate (extreme) limit of nonlocality achievable in 2D systems. Our choice of graphene is dictated by recent observations of ultraslow and ultraconfined acoustic graphene plasmons [19,21–23] and their potential for photocurrent harvesting [24].

A number of works dealt with the nonlinear optical properties of graphene [11,12,25–27], but the resulting expressions were analyzed only in the long-wavelength limit. Very recently, Ref. [28] predicted an enhancement of the rectified current in graphene at strong nonlocality via the solution of the time-dependent Schrodinger equation. This approach is applicable to arbitrarily strong fields but does not consider carrier scattering as a limiting factor for any resonant effect. Here, we derive the nonlocal nonlinear conductivity of graphene $\sigma^{(2)}(\mathbf{q}, \omega)$ as the proportionality coefficient between the photocurrent density and squared ac electric field, $\mathbf{j}_2 = \mathbf{n}_q \sigma^{(2)}(\mathbf{q}, \omega) E_1^2$. It possesses a square-root singularity at a phase velocity approaching the Fermi velocity of 2D electrons. We interpret this singularity as phase locking between dragged electrons and the electromagnetic field. Second, we quantify the electromagnetic energy flux S_{pl} carried by 2D plasmons at large q . The ratio of these quantities is the current responsivity measured in photodetection experiments. The energy flux has a singularity counterbalancing that in $\sigma^{(2)}$, which appears due to the large kinetic energy of the charge carriers in the electromagnetic wave. As a result, the photocurrent responsivity has a universal maximum order of $0.25e/E_F$, achieved at $\omega \approx 1.4qv_0$.

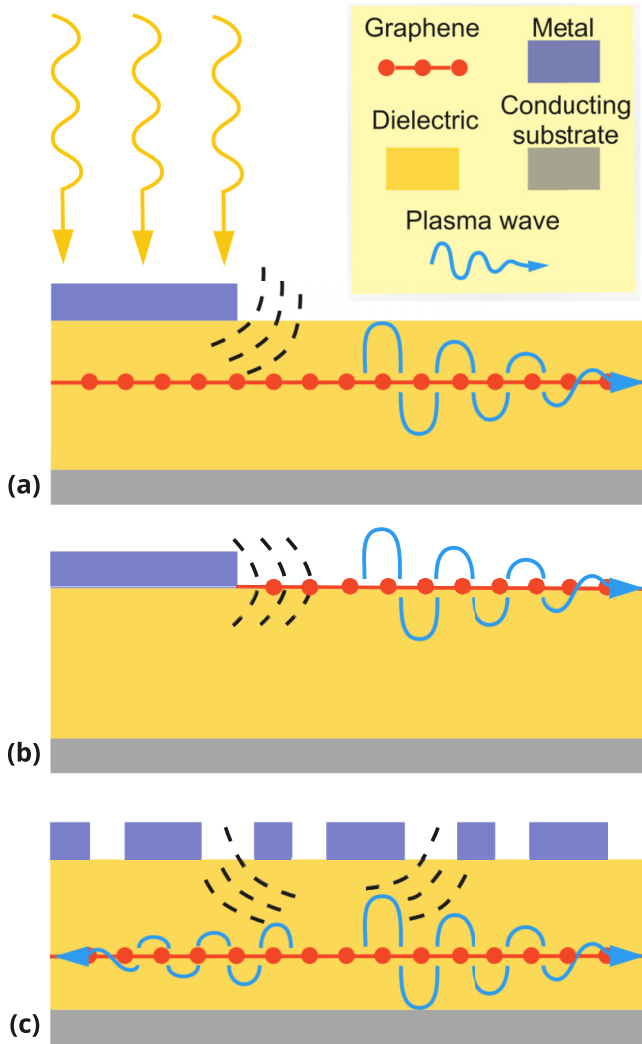


FIG. 1. Possible structures where plasmonic drag can be observed upon diffraction of the incident plane electromagnetic wave: (a) graphene under the edge of a metal gate, (b) graphene contacted by metal, and (c) graphene under the grating lacking inversion symmetry.

II. KINETIC THEORY OF THE PLASMONIC DRAG PHOTOCURRENT

Evaluation of a photocurrent \mathbf{j}_2 under illumination of any setup with a 2D channel (including those shown in Fig. 1) can be factored into three main steps: (1) finding the nonlinear response relations between \mathbf{j}_2 and the electric field in the 2D plane $\mathbf{E}(\mathbf{r}, t)$, (2) solution of the linear-response diffraction problem for a given setup, and (3) substitution of linear-response fields into the nonlinear material relations. At the first stage, the electric field in the 2D plane $\mathbf{E}(\mathbf{r}, t)$ is assumed to be fixed, and nonlinear material relations are derived from classical or quantum kinetic equations [29]. Deriving the relation between the total field $\mathbf{E}(\mathbf{r}, t)$ and the incident field \mathbf{E}_{inc} is, on the contrary, a linear-response, purely electromagnetic problem. Its solution requires knowledge of the linear 2D conductivity $\sigma^{(1)}$ and structure of the dielectric environment. All effects of the self-consistent field are taken into account

at that stage and are decoupled from the evaluation of the nonlinear response.

In further calculations, we shall specify the full field acting on a two-dimensional system of the form $\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_1 \cos(\mathbf{q}\mathbf{r} - \omega t) = \frac{1}{2} \mathbf{E}_1 e^{i(\mathbf{q}\mathbf{r} - \omega t)} + \text{H.c.}$ (the subscript 1 is the order with respect to the electromagnetic field; 0 corresponds to unperturbed quantities). Such an approximation implies that most of the diffracted field is carried by 2D plasmons. This fact is indeed justified by the numerical and analytical solutions of the diffraction problems [13,14,30]. Possible extensions of our approach to arbitrary nonuniform fields will be sketched in Sec. IV. Following the decoupling approach described above, we shall not relate \mathbf{E}_1 to the incident field \mathbf{E}_{inc} . These relations depend on the details of the dielectric environment and can be found elsewhere [13,14,30–32]. The plasmon field is longitudinal, with $\mathbf{E}_1 \parallel \mathbf{q}$.

Our starting point for the evaluation of the photocurrent is the classical kinetic equation for the electron distribution function $f(\mathbf{r}, \mathbf{p}, t)$. The above approach is valid at classical frequencies and wave vectors, $\hbar\omega \ll E_F$ and $q \ll q_F$, where E_F and q_F are the Fermi energy and wave vector, respectively. In that case, interband transitions are blocked, although at higher frequencies they may also contribute to the drag photocurrent [33,34]. We search for the distribution function as a power series of the electromagnetic field, $f = f_0 + f_1 + f_2$, where $f_n \propto E_1^n$. The linear-response part of the distribution function is decomposed as $f_1(\mathbf{r}, \mathbf{p}, t) = \frac{1}{2} f_1(\mathbf{q}, \mathbf{p}, \omega) e^{i(\mathbf{q}\mathbf{r} - \omega t)} + \text{H.c.}$ It obeys the kinetic equation

$$-i(\omega - \mathbf{q}\mathbf{v}_p + i/\tau) f_1(\mathbf{q}, \mathbf{p}, \omega) - e\mathbf{E}_1 \frac{\partial f_0}{\partial \mathbf{p}} = 0, \quad (1)$$

where \mathbf{v}_p is the electron velocity, τ is the momentum relaxation time, f_0 is the equilibrium (Fermi) distribution function, and $e > 0$ is the elementary charge. The nonlinear distribution function $f_2(\mathbf{r}, \mathbf{p}, t)$ will possess harmonics oscillating at zero and double (2ω) frequency. Being interested only in the direct photocurrent, we solve the kinetic equation only for the time-averaged distribution function $\langle f_2(\mathbf{r}, \mathbf{p}, t) \rangle_t \equiv \langle f_2 \rangle$:

$$-e \left\langle \mathbf{E}(\mathbf{r}, t) \frac{\partial f_1(\mathbf{r}, \mathbf{p}, t)}{\partial \mathbf{p}} \right\rangle_t = -\frac{\langle f_2 \rangle}{\tau}. \quad (2)$$

By virtue of the continuity equation, the distribution function $\langle f_2 \rangle$ is automatically uniform in space. We have also neglected the second-order electric field \mathbf{E}_2 in Eq. (2). It does not emerge in short-circuit measurements of the photocurrent. For open-circuit measurements of the photovoltage, it is given by $\mathbf{E}_2 = \mathbf{j}_2 / \sigma_{dc}$, where σ_{dc} is the dc conductivity of graphene.

Substituting the complex representations for the electric field $\mathbf{E}(\mathbf{r}, t)$ and distribution function $f_1(\mathbf{r}, \mathbf{p}, t)$ into (2) and performing the time averaging, we obtain

$$\begin{aligned} \langle f_2 \rangle &= \frac{e\tau}{2} \text{Re} \left\{ \mathbf{E}_1 \frac{\partial f_1(\mathbf{q}, \mathbf{p}, \omega)}{\partial \mathbf{p}} \right\} \\ &= \frac{e^2 E_1^2}{2} \frac{\partial}{\partial \mathbf{p}} \left\{ \frac{\partial f_0 / \partial \mathbf{p}}{(\omega - \mathbf{q}\mathbf{v}_p)^2 + \tau^{-2}} \right\}. \end{aligned} \quad (3)$$

The rectified current is obtained by integration over momentum space $\mathbf{j}_2 = -eg \sum_{\mathbf{p}} \langle f_2 \rangle$, where $g = 4$ is the electron degeneracy factor in graphene. Apparently, the rectified

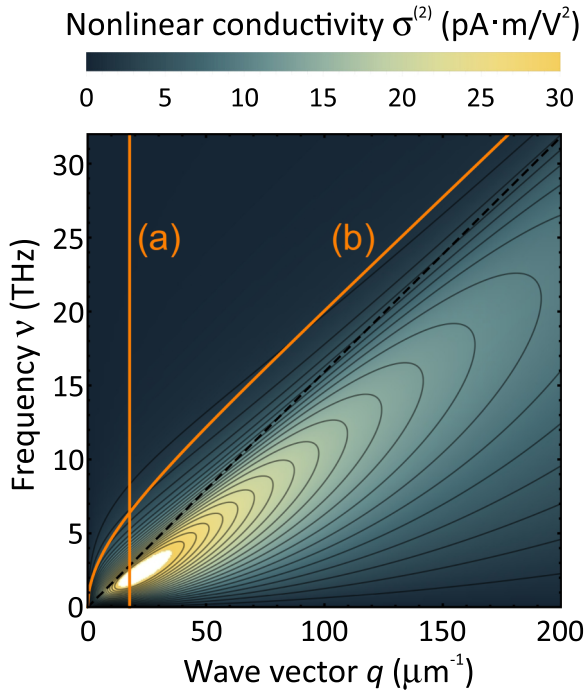


FIG. 2. Color map of nonlinear conductivity $\sigma^{(2)}(q, \omega)$ vs frequency $\nu = \omega/2\pi$ and wave vector q . Vertical line (a) corresponds to the constant wave vector $q = q_F$; line (b) shows the dispersion of graphene plasmons. Momentum relaxation time $\tau = 10^{-13}$ s; carrier density $n = 10^{10}$ cm $^{-2}$.

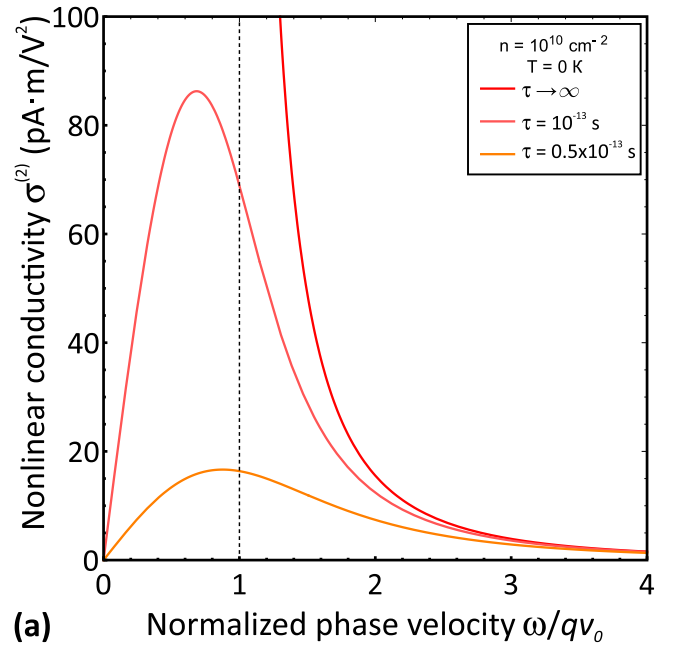
current is proportional to E_1^2 and directed along $\mathbf{n}_q = \mathbf{q}/q$. Introducing the nonlinear conductivity $\mathbf{j}_2 = \sigma^{(2)}(q, \omega)E_1^2\mathbf{n}_q$, we find (see Appendix A for details)

$$\sigma^{(2)}(q, \omega) = \frac{ge^3}{2\pi\hbar^2v_0} \frac{f_0(0)}{q^2} \times \left(\tilde{v}_{ph} + \text{Im} \frac{(\tilde{v}_{ph} - iv_{\text{eff}})\sqrt{(\tilde{v}_{ph} - iv_{\text{eff}})^2 - 1}}{2v_{\text{eff}}} \right), \quad (4)$$

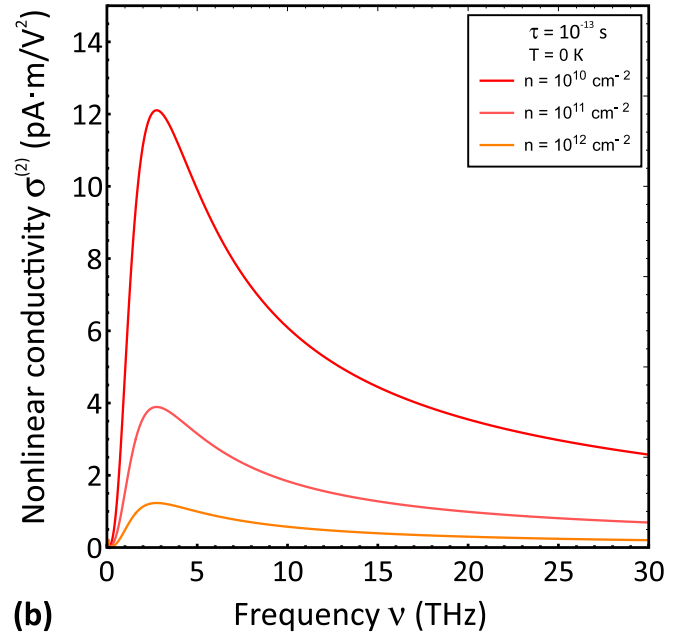
where $\tilde{v}_{ph} = \omega/qv_0$ is the dimensionless phase velocity, $v_{\text{eff}} = (qv_0\tau)^{-1}$ is the effective collision frequency, and $f_0(0)$ is the equilibrium electron distribution function evaluated at zero energy. Accounting for the holes in graphene amounts to the replacement $f_0(0) \rightarrow 2f_0(0) - 1$.

The most remarkable property of nonlinear conductivity (4) is the presence of the square-root singularity as the wave phase velocity ω/q approaches the electron Fermi velocity from either side, as shown in Fig. 2. A detailed inspection shows that the singular contribution to the current comes from electrons moving in phase with the wave, i.e., at angles $\cos\theta \approx \omega/(qv_0)$. We may say that these electrons are trapped in the minima of the harmonic potential induced by the wave and move synchronously with the wave velocity. The growth of $\sigma^{(2)}$ with increasing wave vector q is linear at the initial stage, as dictated by symmetry considerations [8]. Further, it becomes superlinear as q approaches the singularity.

The effect of the finite scattering rate τ^{-1} on nonlinear conductivity is highly nontrivial. It depends critically on the ratio of the phase velocity ω/q and Fermi velocity v_0 , which



(a) Nonlinear conductivity $\sigma^{(2)}$ vs normalized phase velocity ω/qv_0 at constant wave vector $q = q_F = 20$ μm^{-1} .



(b) Nonlinear conductivity $\sigma^{(2)}$ vs frequency ν (THz) at constant wave vector $q = q_F = 20$ μm^{-1} .

FIG. 3. Dependence of the nonlinear conductivity $\sigma^{(2)}$ on (a) the normalized phase velocity at constant wave vector $q = q_F = 20$ μm^{-1} , corresponding to line cut (a) in Fig. 2, and (b) ω and q bound by plasmon dispersion at various carrier densities $n = 10^{10}$, 10^{11} and 10^{12} cm $^{-2}$. Momentum relaxation time $\tau = 10^{-13}$ s.

is illustrated in Fig. 3(a). For slow waves with $\omega/qv_0 < 1$, the nonlinear conductivity $\sigma^{(2)}$ is approximately proportional to the relaxation time. This result is interpreted as follows: the amount of momentum transferred from the field to electrons is proportional to the rate of Landau damping. Being a collisionless process, it does not depend on τ . The current established at a given momentum transfer rate (i.e., at given force density) is inversely proportional to the electron scattering rate.

The situation for fast waves, $\omega/qv_0 > 1$, is different. Such waves cannot induce intraband Landau damping by virtue

of momentum conservation, and the finite scattering rate is required to soften the momentum constraint. At the same time, scattering acts to dissipate the generated current. As a result, $\sigma^{(2)}$ becomes independent of the scattering time for relatively fast phase velocities $\tilde{v}_{ph} \gg 1$. The nonlinear conductivity in this limit is given by Eq. (A9). This τ independence of $\sigma^{(2)}$ breaks down as we approach the singularity at $\omega = qv_0$.

For real 2D plasmons, the wave vector is bound to the frequency via the dispersion relation. Therefore, the nonlinear conductivity $\sigma^{(2)}(q, \omega)$ can be probed only for a restricted set of ω and q . To inspect the accessible values of $\sigma^{(2)}$, we use the dispersion of ungated plasmons in graphene [20]:

$$\omega_{pl}(q) = qv_0 \frac{1 + g\alpha q_F/q}{\sqrt{1 + 2g\alpha q_F/q}}, \quad (5)$$

where $\alpha = e^2/(\kappa\hbar v_0)$ is the effective coupling constant and κ is the background dielectric constant. A distinctive property of dispersion (5) lies in the complete absence of Landau damping. This is guaranteed by the phase velocity of the graphene plasmon, which always exceeds the Fermi velocity v_0 . Therefore, the singularity of nonlinear conductivity at $q = \omega/v_0$ can be *approached* from the side of smaller wave vectors but cannot be addressed exactly.

To study this situation quantitatively, we bind q and ω via Eq. (5) and plot $\sigma^{(2)}$ vs frequency. The result is shown in Fig. 3(b). Once the frequency is small, the nonlinear conductivity decreases. Indeed, the initial part of the plasmon dispersion curve has both low frequency and a small characteristic wave vector (transferred momentum). At very high frequencies, $\sigma^{(2)}$ becomes small again. Although the plasmon dispersion approaches a singular line $\omega = qv_0$ at large ω , the inertia of electrons does not let them keep up with the rapidly oscillating electric field.

The only way to probe the singular nonlinear conductivity is thus to bring the whole plasmon dispersion closer to the singular line $\omega = qv_0$. This can be achieved via a reduction of the carrier density and/or an increase in the background dielectric constant. The result is illustrated in Fig. 3(b), where reducing $n = q_F^2/\pi$ from 10^{12} to 10^{10} cm $^{-2}$ leads to an order of magnitude enhancement of the nonlinear conductivity.

III. LIMITS TO PLASMONIC DRAG RESPONSIVITY

Experimental measurement of the nonlinear conductivity at finite wave vector q is very challenging as the amplitude of the electric field E_1 in the 2D plane is different from the incident field E_{inc} . This difference stems from strong self-consistent field effects which, in fact, are responsible for the launching of the plasmons. In this regard, we note that the predicted resonance in $\sigma^{(2)}$ is conceptually different from numerous ‘‘plasmonic enhancement’’ phenomena occurring via self-consistent fields [25,27,35]. The latter effects should be included in the linear-response fields \mathbf{E}_1 via the solution of the plane-wave diffraction problem.

A more common measurable quantity is the photocurrent responsivity $R_J = j_2/S$, where S is the incident power density. Under perfect matching conditions, the power flow of the incident electromagnetic wave S is fully transformed into the power flow carried by the 2D plasmon S_{pl} . A nearly perfect conversion is attainable with the proper design of the grating

couplers [36]. In simpler systems, such as metal edges, the conversion coefficient is well below unity, $S_{pl}/S \ll 1$ [14,37]. Approximating $S_{pl} \approx S$, we obtain a natural upper bound of the plasmonic drag responsivity.

The power flow density \mathbf{s} carried by the plasmon is the sum of electromagnetic and kinetic contributions [38]:

$$\mathbf{s} = \frac{c}{8\pi} [\mathbf{E}_1 \times \mathbf{H}_1^*] + \frac{1}{4} \frac{\partial \text{Im}\sigma^{(1)}(q, \omega)}{\partial \mathbf{q}} (\mathbf{E}_1, \mathbf{E}_1^*) \delta(z), \quad (6)$$

where the plasmon fields, in the quasistatic limit, are given by

$$E_{1z} = \frac{1}{2} E_1 e^{-|q||z|} e^{iqx} \text{sgnz} + \text{H.c.}, \quad (7)$$

$$H_{1y} = -\frac{\omega\epsilon}{cq} E_{1z} \quad (8)$$

and $\sigma^{(1)}(q, \omega)$ is the linear conductivity of graphene. In the classical limit, it is given by

$$\sigma^{(1)}(q, \omega) = ig \frac{e^2}{\hbar} \frac{E_F}{2\pi\hbar} \frac{\omega}{q^2 v_0^2} \left[\frac{\omega}{\sqrt{\omega^2 - q^2 v_0^2}} - 1 \right]. \quad (9)$$

Integrating the flow density (6) over the vertical coordinate z , we obtain the full power flow $S_{pl} = \int s_x dz$ in the form

$$S_{pl} = \frac{E_1^2}{4} \frac{1}{q} \frac{\partial}{\partial q} [q \sigma^{(1)}(q, \omega)] \Big|_{\omega=\omega_{pl}(q)}. \quad (10)$$

It is now apparent that the plasmon power flow diverges as its phase velocity approaches v_0 . Physically, it comes from a very large contribution of the carrier kinetic energy to the net energy flow. Formally, it comes from differentiating the singular conductivity $\sigma^{(1)} \propto [\omega^2 - q^2 v_0^2]^{-1/2}$; differentiation enhances the strength of the singularity. As a result, the current responsivity $R_J = j_2/S_{pl}$ is bounded from above as the plasmon phase velocity approaches v_0 . The singular growth in $\sigma^{(2)}$ at $\omega/q \rightarrow v_0$ is overwhelmed by a faster growth in S_{pl} . This situation is illustrated in Fig. 4. Figure 4(a) shows R_J as a function of independent q and ω ; naturally, the maximum of R_J lies above the singular line $\omega = qv_0$. Figure 4(b) shows the responsivity evaluated at ω and q bound by dispersion (5); again, this function has a pronounced maximum.

Instructively, it is possible to present the current responsivity in a universal form at $T = 0$:

$$R_J = \frac{e}{E_F} f(\tilde{v}_{ph}), \quad (11)$$

where $f(\tilde{v}_{ph})$ is the dimensionless function of the scaled phase velocity $\tilde{v}_{ph} = \omega/qv_0$ (Appendix B). The function $f(\tilde{v}_{ph})$ reaches a maximum value of 0.243 at $\omega = 1.38qv_0$. Remarkably, the result does not depend on the carrier density and dielectric environment.

The ultimate plasmonic drag responsivity (11) can be compared to that of a perfect photovoltaic cell $R_{pv} < e/\hbar\omega$. Naturally, the drag responsivity is below the photovoltaic limit, as all our calculations were performed in the classical domain $E_F \gg \hbar\omega$. Nevertheless, the maximum responsivities of 2–20 A/W in Fig. 4 are large compared to those of typical graphene photodetectors operating in the terahertz [39,40] and infrared [41] frequency ranges. It should be also noted that plasmonic drag mechanism provides a fast photoresponse

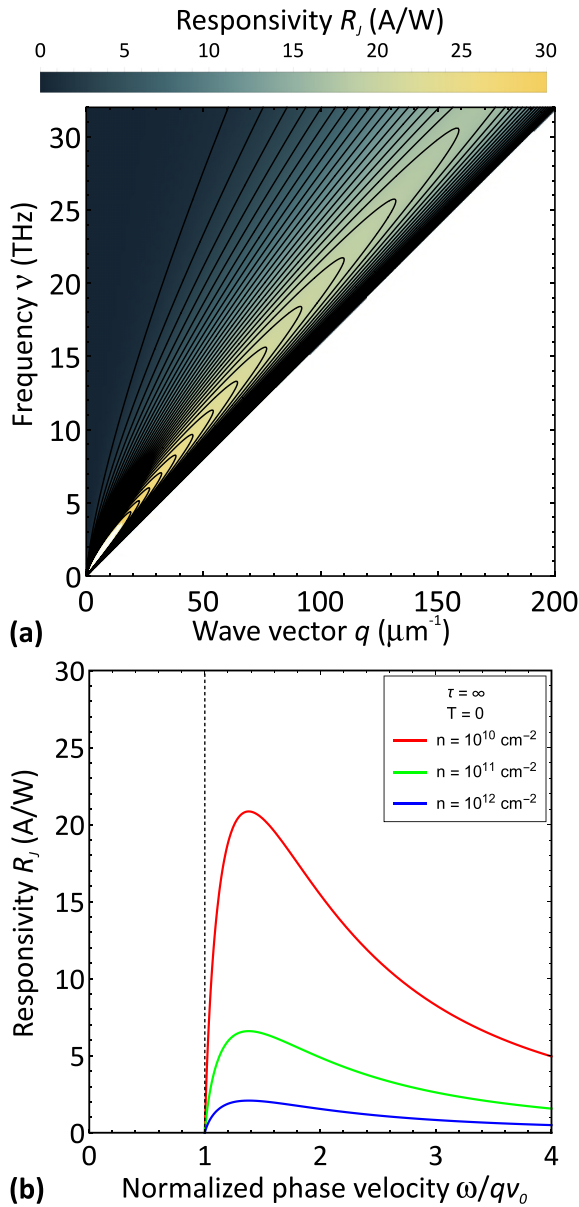


FIG. 4. (a) Dependence of the current responsivity R_J on frequency $\omega/2\pi$ and wave vector q calculated at carrier density $n = 10^{10} \text{ cm}^{-2}$. (b) Plasmonic drag responsivity evaluated at ω and q bound by plasmon dispersion relation at various carrier densities indicated in inset.

[42], as the timescale for the decay of the photocurrent is the momentum relaxation time τ .

All previous calculations carried out at $T = 0$ have indicated that plasmonic drag responsivity benefits from low carrier density. At finite temperature, the density is limited by thermal excitation of the carriers, and the plasmon velocity cannot get very close to v_0 . Accounting for the finite temperature amounts to a simple replacement of the “effective Fermi energy” in the expressions for linear conductivity and plasmon dispersion,

$$E_F \rightarrow kT \ln(1 + e^{E_F/kT})(1 + e^{-E_F/kT}). \quad (12)$$

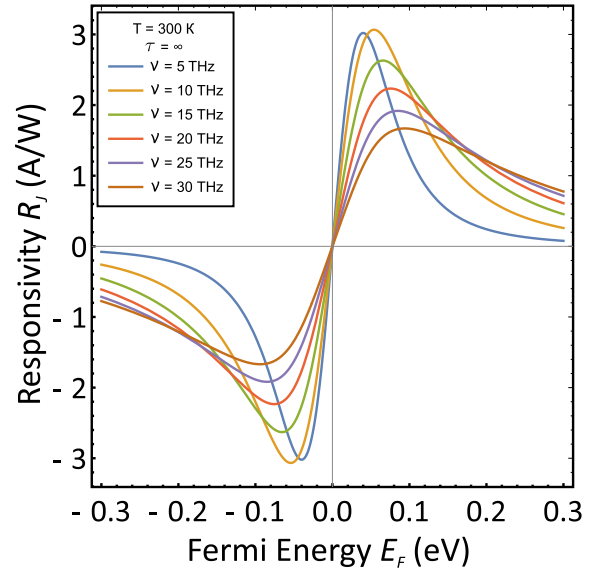


FIG. 5. Plasmonic drag responsivity vs Fermi energy evaluated at different frequencies of incoming radiation. The wave vector is obtained from the plasmon dispersion relation. Temperature $T = 300 \text{ K}$.

It is possible to evaluate R_J at finite temperature, which is done in Fig. 5 at $T = 300 \text{ K}$. The obtained dependence of R_J on the Fermi energy is an antisymmetric function of E_F , with a maximum located at $E_F \sim kT$. The functional dependence of R_J on the frequency is quite peculiar. Namely, the responsivity grows with increasing frequency at large Fermi energies. This growth is associated with an increased plasmon wave vector at higher frequencies and higher average momentum transferred to an electronic system.

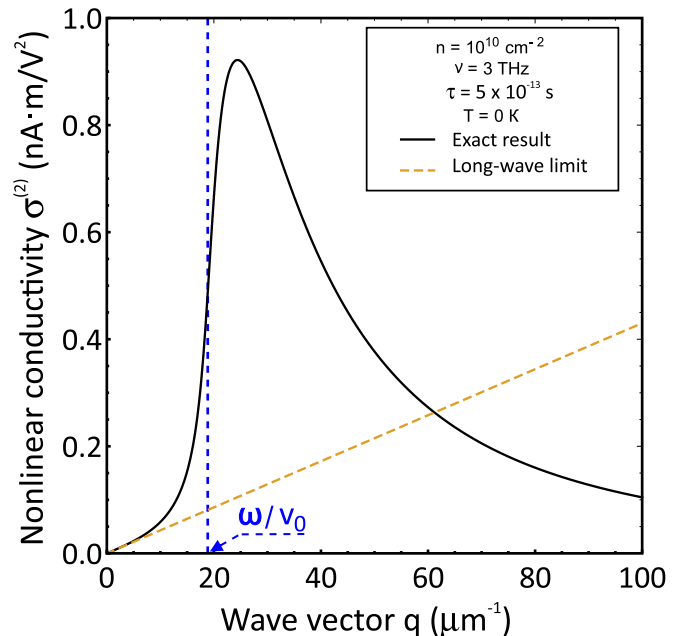


FIG. 6. Comparison of the approximation [(A11)] with the exact expression (A4) for nonlinear conductivity for $\sigma^{(2)}$.

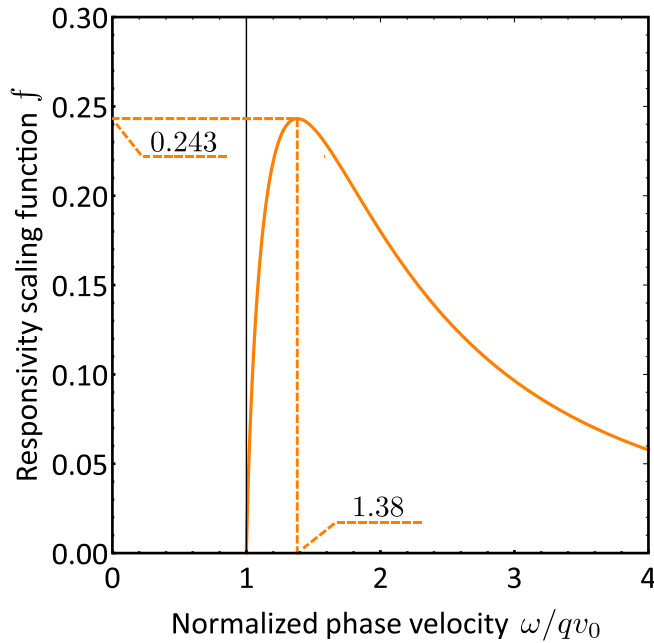


FIG. 7. The dependence of the dimensionless function f governing the universal behavior of plasmonic drag responsivity.

IV. DISCUSSION AND CONCLUSIONS

The above calculations of the plasmonic drag photocurrent were based on a specific spatial profile of the electric field, $\mathbf{E} \propto \cos(\mathbf{q}\mathbf{r} - \omega t)$. Real fields in the vicinity of asymmetric scatterers, such as those shown in Fig. 1, have more complex spatial profiles. In addition to a plasmonic component, they have an evanescent part that is responsible for the lightning-rod effect at the metal contacts. The solution of the kinetic equation for an arbitrary spatial profile of the field $E(x, t) = \frac{1}{2}E_1(x)e^{-i\omega t} + \text{H.c.}$ does not represent a problem if the field is longitudinal and depends only on the x coordinate. In particular, the result for $f_1(x, \mathbf{p}, \omega)$ reads

$$f_1(x, \mathbf{p}, \omega) = e \frac{\partial f_0}{\partial \varepsilon} \int_{-\infty}^x dx' E_1(x') e^{\frac{i\omega - \tau^{-1}}{v_x}(x-x')} \quad (13)$$

for right-moving electrons ($v_x = v_0 \cos \theta > 0$) and

$$f_1(x, \mathbf{p}, \omega) = -e \frac{\partial f_0}{\partial \varepsilon} \int_x^{+\infty} dx' E_1(x') e^{\frac{i\omega - \tau^{-1}}{v_x}(x-x')} \quad (14)$$

for left-moving electrons ($v_x < 0$). In a similar fashion, one can obtain the second-order distribution function $\langle f_2 \rangle$ in a nonuniform field. Indeed, its governing equation has only first-order x derivatives and is analytically solvable for an arbitrary nonuniform field.

The structure of the solutions for an arbitrary electric field enables us to understand better the origins of the singular conductivity at $\omega = qv_0$ (both linear and nonlinear) and to understand its limiting factors. When the electric field represents a purely running wave, $E_1(x') \propto e^{iqx'}$, the phase factor of the field partially compensates the phase factor of the perturbed electrons. This compensation is complete if the wave phase velocity equals the electron velocity. Integration of the nearly constant function over a very large length leads to divergent conductivity. Of course, the scattering of electrons renders

all integrals convergent, the same role is played by the finite extent of the fields in real samples [43].

Experimental manifestations of the plasmonic drag can be present in all graphene-based structures with an asymmetric electromagnetic environment. It can be distinguished from well-known photothermoelectric and photovoltaic effects by the nonvanishing current in the absence of p - n junctions in graphene channels. Particularly, the plasmonic drag current can appear upon illumination of the top gate edge above graphene at zero top gate voltage, i.e., at uniform channel doping [Fig. 1(a)]. It can also appear upon illumination of gratings with an asymmetric unit cell [Fig. 1(c)] at zero voltage applied to the gratings. Illumination of a metal in contact with graphene [Fig. 1(b)] also results in a plasmonic drag current but can be masked by photovoltaic and photothermoelectric effects at the Schottky junction. A salient feature of the plasmonic drag photocurrent is its strong sensitivity to the angle of incidence of the electromagnetic wave. This results from the strong angular dependence of the photon-to-plasmon conversion coefficient [14].

To conclude, we have shown resonant enhancement of the nonlinear conductivity of graphene being a proportionality coefficient between the rectified current and squared electric field. The resonance occurs whenever the wave vector and frequency of the wave satisfy $qv_0 = \omega$. The condition is interpreted as phase locking between surface electromagnetic waves and dragged electrons. The responsivity of the graphene photodetector exploiting this effect of plasmon drag, however, remains finite in the phase-locking condition. This occurs due to a counterbalancing singularity in the power density carried by a plasmon. We expect that similar phase-locking singularities would also appear in a closely related effect of second-harmonic generation [44].

ACKNOWLEDGMENTS

This work was supported by the Grant No. 20-1-3-43-1 of the ‘‘Basis’’ foundation (kinetic theory of photocurrent) and Grant No. 21-79-20225 of the Russian Science Foundation (analysis of detection responsivity). The authors thank M. Glazov and M. Jablan for valuable discussions.

APPENDIX A: CALCULATION OF NONLINEAR CONDUCTIVITY

We calculate $\sigma^{(2)} = j_2/E_1^2$ by integrating the second-order distribution function over momentum space:

$$\mathbf{j}_2 = -\frac{g}{(2\pi\hbar)^2} e \int_{-\infty}^{+\infty} \mathbf{v}_p \langle f_2 \rangle dp_x dp_y. \quad (A1)$$

After revealing the derivatives in $\langle f_2 \rangle$, Eq. (3), we move to the polar coordinates:

$$dp_x dp_y = p dp d\phi, \quad \frac{\partial f_0}{\partial p_x} = \frac{\partial f_0}{\partial p} \cos \phi, \\ \frac{\partial^2 f_0}{\partial p_x^2} = \cos^2 \phi \frac{\partial^2 f_0}{\partial p^2} + \frac{\sin^2 \phi}{p} \frac{\partial f_0}{\partial p}, \quad \frac{\partial v_x}{\partial p_x} = \frac{\sin^2 \phi}{p} v_0. \quad (A2)$$

The integrals over the absolute value of the momentum yield

$$\int_0^{+\infty} \frac{\partial f_0}{\partial p} dp = -f_0(0), \quad \int_0^{+\infty} \frac{\partial^2 f_0}{\partial p^2} p dp = f_0(0). \quad (\text{A3})$$

These identities are valid at arbitrary T . As a result, $\sigma^{(2)}$ is expressed as

$$\sigma^{(2)} = -\frac{g}{(2\pi\hbar)^2} \frac{e^3 f_0(0)}{q^2 v_0} I_\phi\left(\frac{\omega}{qv_0}, \frac{1}{qv_0\tau}\right), \quad (\text{A4})$$

where

$$I_\phi(a, b) = \int_0^{2\pi} \left\{ \frac{\cos\phi(\cos^2\phi - 1/2)}{(\cos\phi - a)^2 + b^2} + \frac{\cos^2\phi \sin^2\phi(\cos\phi - a)}{[(\cos\phi - a)^2 + b^2]^2} \right\} d\phi. \quad (\text{A5})$$

The latter integral can be calculated analytically with the residue theorem. We introduce a complex variable:

$$z = e^{i\phi}, \quad \cos\phi = \frac{1+z^2}{2z}, \quad d\phi = \frac{1}{iz} dz, \quad (\text{A6})$$

and the integration line is now the unit circle $|z| = 1$. Inside of it, the integrand has three poles:

$$\begin{aligned} z_1 &= a - ib - \sqrt{(a - ib)^2 - 1} \quad (\text{second-order pole}), \\ z_2 &= a + ib - \sqrt{(a + ib)^2 - 1} \quad (\text{second-order pole}), \\ z_3 &= 0 \quad (\text{first-order pole}). \end{aligned} \quad (\text{A7})$$

Thus, we can find the residues of the integrand and the exact form for the integral (A5):

$$I_\phi(a, b) = -2\pi \left(a + \text{Im} \frac{(a - ib)\sqrt{(a - ib)^2 - 1}}{2b} \right). \quad (\text{A8})$$

Upon transformations, we have used the fact that the residues at poles z_1 and z_2 are complex conjugate to each

other, so that the whole integral is real. The branch cut of the square-root function $f(z) = \sqrt{z}$ here is chosen to run along the real axis from $-\infty$ to 0.

In the collisionless electron plasma ($\tau \rightarrow \infty$ and $b \rightarrow 0$) the integral (A5) is nonzero only for fast waves, $\omega > qv_0$. It equals

$$I_\phi(\omega > qv_0, b \rightarrow 0) \approx 2\pi \left(\frac{2a^2 - 1}{2\sqrt{a^2 - 1}} - a \right). \quad (\text{A9})$$

For slow waves with $\omega/qv_0 < 1$, the nonlinear conductivity $\sigma^{(2)}$ is approximately proportional to the relaxation time. In this case and in the nearly collisionless electron system ($b \rightarrow 0$), we find

$$I_\phi(\omega < qv_0, b \rightarrow 0) \approx 2\pi \frac{a\sqrt{1 - a^2}}{2b}. \quad (\text{A10})$$

It is possible to expand the nonlinear conductivity (A4) in the long-wavelength limit ($q \rightarrow 0$):

$$\sigma^{(2)}(q, \omega) \Big|_{q \rightarrow 0} = -\frac{g}{(2\pi\hbar)^2} \frac{e^3 f_0(0)}{v_0} \frac{\pi}{4} \frac{v_0^3}{\omega^3} \frac{(\omega\tau)^4}{((\omega\tau)^2 + 1)^2} q. \quad (\text{A11})$$

This dependence is shown in Fig. 6 with a dashed line. The full form of the nonlinear conductivity is shown with a solid line. Apparently, the limiting form coincides with the full one at small q , while near the resonance the full conductivity is much larger.

APPENDIX B: UNIVERSAL RESPONSIVITY FUNCTION

The universal function in the expression for the plasmonic drag responsivity (11) is given by

$$f(a) = \frac{2(a^2 - 1)[a^3 - (a^2 - 1)\sqrt{a^2 - 1}]}{a(a^4 - a^2 - 1)}. \quad (\text{B1})$$

It is readily obtained by combining the linear- and nonlinear-response conductivities. Its plot is shown in Fig. 7.

-
- [1] J. C. W. Song, M. S. Rudner, C. M. Marcus, and L. S. Levitov, Hot carrier transport and photocurrent response in graphene, *Nano Lett.* **11**, 4688 (2011).
- [2] V. M. Muravev and I. V. Kukushkin, Plasmonic detector/spectrometer of subterahertz radiation based on two-dimensional electron system with embedded defect, *Appl. Phys. Lett.* **100**, 082102 (2012).
- [3] K. J. Tielrooij, L. Piatkowski, M. Massicotte, A. Woessner, Q. Ma, Y. Lee, K. S. Myhro, C. N. Lau, P. Jarillo-Herrero, N. F. Van Hulst, and F. H. L. Koppens, Generation of photovoltage in graphene on a femtosecond timescale through efficient carrier heating, *Nat. Nanotechnol.* **10**, 437 (2015).
- [4] X. Cai, A. B. Sushkov, R. J. Suess, M. M. Jadidi, G. S. Jenkins, L. O. Nyakiti, R. L. Myers-Ward, S. Li, J. Yan, D. K. Gaskill, T. E. Murphy, H. D. Drew, and M. S. Fuhrer, Sensitive room-temperature terahertz detection via the photothermoelectric effect in graphene, *Nat. Nanotechnol.* **9**, 814 (2014).
- [5] T. J. Echtermeyer, P. S. Nene, M. Trushin, R. V. Gorbachev, A. L. Eiden, S. Milana, Z. Sun, J. Schliemann, E. Lidorikis, K. S. Novoselov, and A. C. Ferrari, Photothermoelectric and photoelectric contributions to light detection in metal-graphene-metal photodetectors, *Nano Lett.* **14**, 3733 (2014).
- [6] A. F. Gibson, M. F. Kimmitt, and A. C. Walker, Photon drag in germanium, *Appl. Phys. Lett.* **17**, 75 (1970).
- [7] H. E. M. Barlow, The application of the Hall effect in a semiconductor to the measurement of power in an electromagnetic field, *Proc. Inst. Electr. Eng., Part B* **102**, 179 (1955).
- [8] J. Karch, P. Olbrich, M. Schmalzbauer, C. Zoth, C. Brinsteiner, M. Fehrenbacher, U. Wurstbauer, M. M. Glazov, S. A. Tarasenko, E. L. Ivchenko, D. Weiss, J. Eroms, R. Yakimova, S. Lara-Avila, S. Kubatkin, and S. D. Ganichev, Dynamic Hall Effect Driven by Circularly Polarized Light in a Graphene Layer, *Phys. Rev. Lett.* **105**, 227402 (2010).
- [9] A. S. Vengurlekar and T. Ishihara, Surface plasmon enhanced photon drag in metal films, *Appl. Phys. Lett.* **87**, 091118 (2005).

- [10] M. Durach and N. Noginova, On the nature of the plasmon drag effect, *Phys. Rev. B* **93**, 161406(R) (2016).
- [11] V. V. Popov, Terahertz rectification by periodic two-dimensional electron plasma, *Appl. Phys. Lett.* **102**, 253504 (2013).
- [12] 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, Terahertz ratchet effects in graphene with a lateral superlattice, *Phys. Rev. B* **93**, 075422 (2016).
- [13] V. V. Popov, D. V. Fateev, E. L. Ivchenko, and S. D. Ganichev, Noncentrosymmetric plasmon modes and giant terahertz photocurrent in a two-dimensional plasmonic crystal, *Phys. Rev. B* **91**, 235436 (2015).
- [14] E. Nikulin, D. Mylnikov, D. Bandurin, and D. Svintsov, Edge diffraction, plasmon launching, and universal absorption enhancement in two-dimensional junctions, *Phys. Rev. B* **103**, 085306 (2021).
- [15] The plasmonic drag effect is conceptually equivalent to the distributed resistive self-mixing [45] and Dyakonov-Shur [45] rectification. This can be seen by presenting the photocurrent via a linear-response electric field and its gradients; all three mechanisms will be described by identical expressions.
- [16] A. V. Chaplik, Possible crystallization of charge carriers in low-density inversion layers, *Sov. Phys. JETP* **35**, 395 (1972).
- [17] J. Nehls, T. Schmidt, U. Merkt, D. Heitmann, A. G. Norman, and R. A. Stradling, Direct manifestation of the Fermi pressure in a two-dimensional electron system, *Phys. Rev. B* **54**, 7651 (1996).
- [18] D. A. Bandurin, E. Mönch, K. Kapralov, I. Y. Phinney, K. Lindner, S. Liu, J. H. Edgar, I. A. Dmitriev, P. Jarillo-Herrero, D. Svintsov, and S. D. Ganichev, Cyclotron resonance overtones and near-field magnetoabsorption via terahertz Bernstein modes in graphene, [arXiv:2106.02117](https://arxiv.org/abs/2106.02117).
- [19] M. B. Lundeberg, Y. Gao, R. Asgari, C. Tan, B. V. Duppen, M. Autore, P. Alonso-González, A. Woessner, K. Watanabe, T. Taniguchi, R. Hillenbrand, J. Hone, M. Polini, and F. H. L. Koppens, Tuning quantum nonlocal effects in graphene plasmonics, *Science* **357**, 187 (2017).
- [20] V. Ryzhii, Terahertz plasma waves in gated graphene heterostructures, *Jpn. J. Appl. Phys., Part 2* **45**, L923 (2006).
- [21] D. A. Iranzo, S. Nanot, E. J. C. Dias, I. Epstein, C. Peng, D. K. Efetov, M. B. Lundeberg, R. Parret, J. Osmond, J. Y. Hong, J. Kong, D. R. Englund, N. M. R. Peres, and F. H. L. Koppens, Probing the ultimate plasmon confinement limits with a van der Waals heterostructure, *Science* **360**, 291 (2018).
- [22] I. Epstein, D. Alcaraz, Z. Huang, V.-V. Pusapati, J.-P. Hugonin, A. Kumar, X. M. Deputy, T. Khodkov, T. G. Rappoport, J.-Y. Hong, N. M. R. Peres, J. Kong, D. R. Smith, and F. H. L. Koppens, Far-field excitation of single graphene plasmon cavities with ultracompressed mode volumes, *Science* **368**, 1219 (2020).
- [23] A. Bylinkin, E. Titova, V. Mikheev, E. Zhukova, S. Zhukov, M. Belyanchikov, M. Kashchenko, A. Miakonkikh, and D. Svintsov, Tight-Binding Terahertz Plasmons in Chemical-Vapor-Deposited Graphene, *Phys. Rev. Appl.* **11**, 054017 (2019).
- [24] D. A. Bandurin, D. Svintsov, I. Gayduchenko, S. G. Xu, A. Principi, M. Moskotin, I. Tret'yakov, D. Yagodkin, S. Zhukov, T. Taniguchi, K. Watanabe, I. V. Grigorieva, M. Polini, G. N. Goltsman, A. K. Geim, and G. Fedorov, Resonant terahertz detection using graphene plasmons, *Nat. Commun.* **9**, 5392 (2018).
- [25] H. Rostami, M. I. Katsnelson, and M. Polini, Theory of plasmonic effects in nonlinear optics: The case of graphene, *Phys. Rev. B* **95**, 035416 (2017).
- [26] A. Principi, D. Bandurin, H. Rostami, and M. Polini, Pseudo-Euler equations from nonlinear optics: Plasmon-assisted photodetection beyond hydrodynamics, *Phys. Rev. B* **99**, 075410 (2019).
- [27] S. A. Mikhailov, Theory of the giant plasmon-enhanced second-harmonic generation in graphene and semiconductor two-dimensional electron systems, *Phys. Rev. B* **84**, 045432 (2011).
- [28] M. Jablan, Quasiclassical nonlinear plasmon resonance in graphene, *Phys. Rev. B* **101**, 085424 (2020).
- [29] E. L. Ivchenko and G. Pikus, *Superlattices and Other Heterostructures: Symmetry and Optical Phenomena*, Springer Series in Solid-State Sciences Vol. 110 (Springer, Heidelberg, 2012).
- [30] S. A. Mikhailov, Plasma instability and amplification of electromagnetic waves in low-dimensional electron systems, *Phys. Rev. B* **58**, 1517 (1998).
- [31] L. Zhang, X.-L. Fu, and J.-Z. Yang, Excitation of propagating plasmons in semi-infinite graphene layer by free space photons, *Commun. Theor. Phys.* **61**, 751 (2014).
- [32] D. Margetis, M. Maier, and M. Luskun, On the Wiener-Hopf method for surface plasmons: Diffraction from semiinfinite metamaterial sheet, *Stud. Appl. Math.* **139**, 599 (2017).
- [33] A. M. Danishevskii, A. A. Kastalskii, S. M. Ryvkin, and I. D. Yaroshetskii, Dragging of free carriers by photons in direct interband transitions in semiconductors, *Sov. Phys. JETP* **31**, 292 (1970).
- [34] M. V. Entin, L. I. Magarill, and D. L. Shepelyansky, Theory of resonant photon drag in monolayer graphene, *Phys. Rev. B* **81**, 165441 (2010).
- [35] X. Yao, M. Tokman, and A. Belyanin, Efficient Nonlinear Generation of THz Plasmons in Graphene and Topological Insulators, *Phys. Rev. Lett.* **112**, 055501 (2014).
- [36] S. N. Tsvetkova, D. H. Kwon, A. Díaz-Rubio, and S. A. Tret'yakov, Near-perfect conversion of a propagating plane wave into a surface wave using metasurfaces, *Phys. Rev. B* **97**, 115447 (2018).
- [37] E. J. C. Dias and F. J. García De Abajo, Fundamental Limits to the Coupling between Light and 2D Polaritons by Small Scatterers, *ACS Nano* **13**, 5184 (2019).
- [38] L. D. Landau, L. P. Pitaevskii, and E. M. Lifshitz, *Electrodynamics of Continuous Media*, Course of Theoretical Physics Vol. 8 (Elsevier, Oxford, 2013).
- [39] D. A. Bandurin, I. Gayduchenko, Y. Cao, M. Moskotin, A. Principi, I. V. Grigorieva, G. Goltsman, G. Fedorov, and D. Svintsov, Dual origin of room temperature sub-terahertz photoresponse in graphene field effect transistors, *Appl. Phys. Lett.* **112**, 141101 (2018).
- [40] L. Vicarelli, M. S. Vitiello, D. Coquillat, A. Lombardo, A. C. Ferrari, W. Knap, M. Polini, V. Pellegrini, and A. Tredicucci, Graphene field-effect transistors as room-temperature terahertz detectors, *Nat. Mater.* **11**, 865 (2012).
- [41] M. Badioli, A. Woessner, K. J. Tielrooij, S. Nanot, G. Navickaite, T. Stauber, F. J. García De Abajo, and F. H. L.

- Koppens, Phonon-mediated mid-infrared photoresponse of graphene, *Nano Lett.* **14**, 6374 (2014).
- [42] V. M. Muravev, V. V. Solov'ev, A. A. Fortunatov, G. E. Tsydynzhapov, and I. V. Kukushkin, On the response time of plasmonic terahertz detectors, *JETP Lett.* **103**, 792 (2016).
- [43] The spreading of electron velocity directions in 2D graphene further reduces the strength of the phase-locking singularity. For a single electron with constant v_x , the linear distribution function, Eq. (14), is inversely proportional to $\omega - qv_x$. Averaging over angles θ turns this inverse proportionality into the square-root singularity.
- [44] M. M. Glazov, Second harmonic generation in graphene, *JETP Lett.* **93**, 366 (2011).
- [45] M. Sakowicz, M. B. Lifshits, O. A. Klimenko, F. Schuster, D. Coquillat, F. Teppe, and W. Knap, Terahertz responsivity of field effect transistors versus their static channel conductivity and loading effects, *J. Appl. Phys.* **110**, 054512 (2011).