Laser-driven parametric instability and generation of entangled photon-plasmon states in graphene

Mikhail Tokman,[1] Yongrui Wang,[2] Ivan Oladyshkin,[1] A. Ryan Kutayiah,[2] and Alexey Belyanin[2]
[1]*Institute of Applied Physics, Russian Academy of Sciences, Nizhny Novgorod, Russia*
[2]*Department of Physics and Astronomy, Texas A&M University, College Station, Texas 77843, USA*
(Received 28 January 2016; revised manuscript received 30 May 2016; published 14 June 2016)

We show that a strong infrared laser beam obliquely incident on graphene can experience a parametric instability with respect to decay into lower-frequency (idler) photons and THz surface plasmons. The instability is due to a strong in-plane second-order nonlinear response of graphene which originates from its spatial dispersion. The parametric decay leads to efficient generation of THz plasmons and gives rise to quantum entanglement of idler photons and surface plasmon states.

DOI: 10.1103/PhysRevB.93.235422

I. INTRODUCTION

Nonlinear parametric decay of a pump laser photon into two lower-frequency photons (usually called "signal" and "idler") in a nonlinear crystal possessing a second-order nonlinearity is the most popular method of generating entangled photon states [1]. At higher pump intensities the parametric process can experience gain which leads to the instability and exponential amplification of coupled signal and idler fields. Stimulated parametric decay enables optical parametric amplifiers and oscillators as popular tunable sources of long-wavelength radiation from near to far infrared [2]. They typically employ bulk transparent crystals under phase-matching conditions for frequencies and wave vectors of the fields participating in a three-wave mixing interaction:

$$\omega_s = \omega_p - \omega_i, \quad \boldsymbol{k}_s = \boldsymbol{k}_p - \boldsymbol{k}_i, \tag{1}$$

where the subscripts s, p, and i represent signal, pump, and idler, respectively. In view of these requirements, the very idea of parametric amplification supported by just a monolayer of material seems unrealistic. Surprisingly, we find that stimulated parametric decay of laser photons is feasible in 2D systems of massless Dirac electrons.

Any surface has anisotropy between in-plane and out-of-plane excitations, and graphene is no exception. However, the second-order susceptibility $\chi^{(2)}_{ijk}$ associated with this surface anisotropy is very small in graphene [3] and we don't even consider it below. A much stronger nonlinear response is expected when all fields and electron excitations lie in-plane. This is obvious already in the classical free-carrier limit because of an extreme band nonparabolicity [4]. However, graphene is a centrosymmetric system for low-energy in-plane excitations, which should prohibit any second-order response. Nevertheless, a nonzero $\chi^{(2)}$ appears beyond the electric dipole approximation when one includes the dependence of $\chi^{(2)}$ on the in-plane photon wave vectors, i.e., the spatial dispersion. In this case the isotropy of graphene is effectively broken by the wave vector direction. The spatial dispersion effects turn out to be quite large because of a large magnitude of the electron velocity v_F, similarly to spatial dispersion in a hot plasma. Further enhancement of $\chi^{(2)}$ occurs at resonance between the pump frequency and twice the Fermi energy: $\omega_p = 2\epsilon_F/\hbar$. Finally, the efficiency of parametric down-conversion is enhanced when one of the generated fields is not a photon but a surface plasmon mode supported by a massless 2D electron layer. A nonzero value of the nonlocal in-plane $\chi^{(2)}$ and plasmon enhancement of the nonlinear signal were pointed out before for second-harmonic generation [5,6] (which only included intraband transitions in a free-carrier model) and for difference-frequency generation [7]. Here we develop the first theory of the parametric decay in graphene, which includes fully quantum description of the nonlinear response and quantization of all fields. The same formalism can be applied to other systems of massless Dirac fermions, for example, surface states in 3D topological insulators such as Bi_2Se_3.

The schematic of the nonlinear process is shown in Fig. 1. An obliquely incident pump photon decays into an idler photon and a signal plasmon of a much lower frequency $\omega_s = \omega_p - \omega_i \ll \omega_{p,i}$ but a comparable wave vector $q_s \sim q_p$. The second of phase matching conditions in Eqs. (1) is replaced by its in-plane projection $\boldsymbol{q}_s = \boldsymbol{q}_p - \boldsymbol{q}_i$. In addition, the signal frequency should match the real part of surface plasmon dispersion $\omega(q)$ shown in the inset to Fig. 1: $\omega_s = \omega(q_s)$. Note that both positive and negative projections of the idler wave vector q_i are possible, where the positive direction is assumed to the right. In particular, negative values of q_i give access to larger plasmon wave vectors $q_s = q_p - q_i = |q_p| + |q_i|$ and frequencies.

II. QUANTIZED SURFACE PLASMON FIELD IN GRAPHENE

Consider a geometry of Fig. 1, i.e., a 2D layer of massless Dirac electrons in $z = 0$ plane between two media with dielectric constants ϵ_1 and ϵ_2. The plasmon frequency ω_s and in-plane wave vector $\boldsymbol{q}_s$ are related through the dispersion relation for a TM-polarized surface mode [7]:

$$4\pi\chi_s + \frac{\epsilon_1}{p_1} + \frac{\epsilon_2}{p_2} = 0, \tag{2}$$

where $p_{1,2} = \sqrt{q_s^2 - \epsilon_{1,2}\frac{\omega_s^2}{c^2}}$. At the THz frequencies smaller than twice the Fermi energy $2\epsilon_F/\hbar$ one needs only to take into account the intraband contributions to the linear

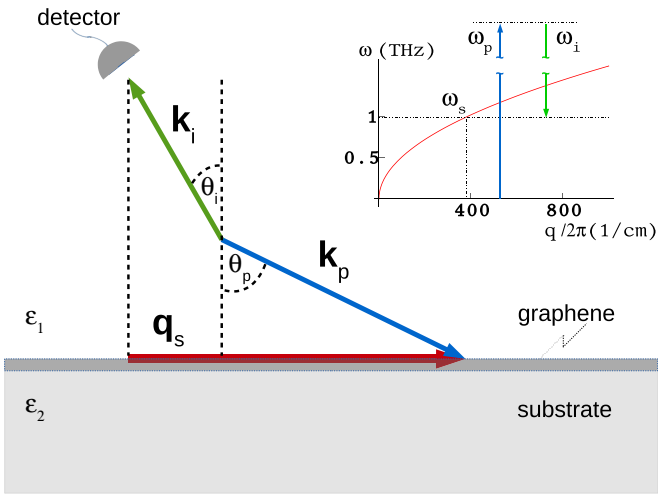

FIG. 1. Schematic of the parametric decay of the pump photon into an idler photon and a surface plasmon, which satisfies conservation of energy and in-plane component of momentum. Inset shows matching of the signal frequency to the real part of surface plasmon frequency obtained by solving Eq. (2).

2D susceptibility $\chi_s(\omega_s,q_s)$, which in the limit of strong degeneracy is given by [7]

$$\chi_s(\omega_s,q_s)=\frac{2e^2E_F}{\pi\hbar^2\omega_s}\frac{(\omega_s+i\gamma)}{(v_Fq_s)^2}\times\left[1-\frac{\omega_s+i\gamma_s}{\omega_s+i\gamma+v_Fq_s}\sqrt{1+\frac{2v_Fq_s}{\omega_s+i\gamma-v_Fq_s}}\right], \tag{3}$$

where E_F is the Fermi energy and γ is the decay rate of the fermion momentum at the surface plasmon frequency.

The z distribution of the field $\boldsymbol{E}_s(z)$ is [7]

$$\boldsymbol{E}_s(z)=\left(\boldsymbol{x}_0\pm z_0\frac{iq_s}{p_{1,2}}\right)E_{s0}e^{\mp p_{1,2}z}, \tag{4}$$

where the upper and lower signs correspond to $z>0, z<0$, respectively.

The quantization for the plasmon field in the limit of weak dissipation $\omega_s\gg\gamma$ consists of two steps. First, a standard quantization procedure is applied neglecting any dissipation [8–11]. Second, Heisenberg equations of motion for the field operators are formulated which include interaction with a dissipative reservoir and the effect of external and nonlinear currents. The first step leads to

$$\hat{\boldsymbol{E}}=\sum_{\boldsymbol{q}_s}\boldsymbol{E}_s(z)\hat{a}_se^{i\boldsymbol{q}_s\boldsymbol{r}_\parallel-i\omega_st}+\text{H.c.}, \tag{5}$$

where $\boldsymbol{r}_\parallel=(x,y)$ and $\hat{a}_s,\hat{a}_s^\dagger$ are annihilation and creation operators of surface plasmon modes. Similar to the case of propagating fields [8,12], the energy of the plasmon field inside a volume V can be written as

$$\hat{H}=\frac{1}{8\pi}\sum_s(\hat{a}_s^\dagger\hat{a}_s+\hat{a}_s\hat{a}_s^\dagger)\int_V\left(\boldsymbol{E}_s^*\frac{\partial(\omega\tilde{\epsilon})}{\partial\omega}\boldsymbol{E}_s+\boldsymbol{B}_s\boldsymbol{B}_s^*\right)d^3r,$$

where $\tilde{\epsilon}$ is the dielectric permittivity tensor.

The normalization constant E_{s0} can be chosen so that the Hamiltonian for the plasmon field takes the standard form: $\hat{H}_s=\sum_{\boldsymbol{q}_s}\hbar\omega_s(\boldsymbol{q}_s)(\hat{a}_s^\dagger\hat{a}_s+1/2)$. This is achieved if we request the following normalization condition, similar to the case of a photon field [12,13]:

$$\hat{H}_s=\int_V\left(\boldsymbol{E}_s^*\frac{\partial(\omega\tilde{\epsilon})}{\partial\omega}\boldsymbol{E}_s+\boldsymbol{B}_s\boldsymbol{B}_s^*\right)d^3r=4\pi\hbar\omega_s. \tag{6}$$

This volume integral is calculated in the Appendices. The result for the normalization constant E_{s0} (in quasielectrostatic approximation) is

$$|E_{s0}|^2=\hbar(\text{Re}(\partial\chi_s/\partial\omega))^{-1}. \tag{7}$$

The effect of dissipation of a plasmon field (within $\omega_s\gg\gamma_s$) and its nonlinear interaction with other fields can be taken into account within the Heisenberg-Langevin approach [8]. For quasimonochromatic wave fields, it is convenient to consider a wave packet of surface plasmon modes with frequencies and wave vectors concentrated in a narrow spectral range $\Delta\omega\ll\omega_s,\Delta q\ll q_s$ near a central component $\propto e^{i\boldsymbol{q}_s\boldsymbol{r}_\parallel-i\omega_st}$ [9–11,14]. Within this approach we introduce the annihilation and creation operators $\hat{a}_s(\boldsymbol{r}_\parallel,t)$ and $\hat{a}_s^\dagger(\boldsymbol{r}_\parallel,t)$ that are slowly varying in time and space relative to ω_s and q_s. Their commutator is equal to the number of quantized modes per unit area $L_x\times L_y=1$ within the spectral interval $\Delta\omega$: $[\hat{a}_s,\hat{a}_s^\dagger]=\frac{\Delta\omega}{2\pi L_yv_s}$, where v_s is the group velocity of a surface plasmon which determines its spectral density of states and L_y is the aperture size of the beam. The commutation relations for the Fourier components of the creation and annihilation operators of the plasmon field envelope $\hat{a}_s=\int d\omega\,\hat{a}_{s\omega}e^{-i(\omega-\omega_s)t}$ and $\hat{a}_s^\dagger=\int d\omega\,\hat{a}_{s\omega}^\dagger e^{i(\omega-\omega_s)t}$ have the form

$$[\hat{a}_{s\omega},\hat{a}_{s\omega'}^\dagger]=\frac{\delta(\omega-\omega')}{2\pi L_yv_s}. \tag{8}$$

Equation (5) for the field operator remains valid for a wave packet after we replace constant operators $\hat{a}_s$ and $\hat{a}_s^\dagger$ with slowly varying operator amplitudes and remove the summation over wave vectors.

Equations for a slowly varying field amplitude of a surface plasmon wave packet can be obtained in the same way as for the propagating optical fields; see, e.g., [9–11]:

$$\frac{\partial\hat{a}_s}{\partial t}+v_s\frac{\partial\hat{a}_s}{\partial x}+\gamma_s\hat{a}_s=\frac{i}{\hbar}\hat{P}_s^{(2)}E_{s0}^*+\hat{F}_s, \tag{9}$$

where $\gamma_s=\hbar^{-1}(\text{Im}[\chi_s])|E_{s0}^2|$, $\hat{F}_s(t)$ is the operator of the Langevin noise, and $\hat{\boldsymbol{P}}_s^{(2)}=\boldsymbol{x}_0\hat{P}_s^{(2)}e^{iq_sx-i\omega_st}+\text{H.c.}$ is the second-order nonlinear component of the polarization operator.

The Langevin noise source ensures a correct expression for the commutator of the plasmon field in the presence of its interaction with a dissipative reservoir. It is convenient to define the properties of the noise source in terms of its spectral components $\hat{F}_s=\int\hat{F}_{s\omega}e^{-i\omega t}\,d\omega$ and $\hat{F}_{s\omega}^\dagger=\hat{F}_{s;-\omega}$. Assuming a dissipative reservoir in thermal equilibrium and adjusting for

the 2D geometry, we can write [8,11]

$$\langle \hat{F}^{\dagger}_{\omega'}(x')\hat{F}_{\omega}(x)\rangle = \frac{\gamma_s N_T(\omega_s)}{\pi L_y}\delta(\omega-\omega')\delta(x-x'),$$
$$\langle \hat{F}_{\omega}(x)\hat{F}^{\dagger}_{\omega'}(x')\rangle = \frac{\gamma_s (N_T(\omega_s)+1)}{\pi L_y}\delta(\omega-\omega')\delta(x-x'), \tag{10}$$

where $\langle\cdots\rangle$ means averaging over both an initial quantum state in the Heisenberg picture and the statistics of the dissipative reservoir, $N_T(\omega_s) = (e^{\hbar\omega_s/(k_BT)}-1)^{-1}$. In the absence of the nonlinear polarization, the solution of Eqs. (9) and (10) in the limit $\gamma_s x/v_s \to \infty$ corresponds to thermal equilibrium:

$$\langle \hat{a}^{\dagger}_s \hat{a}_s\rangle \to \langle \hat{a}^{\dagger}_s \hat{a}_s\rangle_T = \frac{N_T(\omega_s)\Delta\omega}{2\pi L_y v_s}. \tag{11}$$

Equation (11) satisfies a general property of the thermal emission: its power $\approx L_y v_s \hbar\omega_s \langle \hat{a}^{\dagger}_s \hat{a}_s\rangle_T$ received by a matched antenna does not depend on the size and shape of the aperture [15].

III. PARAMETRIC INSTABILITY IN GRAPHENE

A. Nonlinear generation of the plasmon field

Generation of surface plasmons in graphene is possible with both P-polarized (see [7]) and S-polarized optical pumping. The theory is developed in the same way for both polarizations. The difference is that in the case of P-polarized photons only the $\chi^{(2)}_{xxx}$ component of the second-order susceptibility is involved, whereas for S-polarized pumping both xyy and yyx components contribute to the parametric process. S-polarized radiation maximizes the in-plane projection of the electric field. In addition, theory predicts a larger magnitude of the xyy component as compared to the xxx component. Therefore, we will consider only the case of S-polarization, assuming an S-polarized bichromatic pump+idler field incident from the $z>0$ half-space:

$$\hat{\boldsymbol{E}} = \sum_{j=p,i}\hat{\boldsymbol{E}}_j, \quad \hat{\boldsymbol{E}}_j = \boldsymbol{y}_0 E_{j0}\hat{c}_j e^{-ik_j z + iq_j x - i\omega_j t} + \text{H.c.},$$

where the normalization fields $|E_{j0}|^2 = 2\pi\hbar\omega_j/n_1^2$ are defined for a unit quantization volume; $\hat{c}_j$ are Heisenberg operators of slowly varying amplitudes corresponding to a finite spectral width $\Delta\omega$ [9–11,13]. The nonlinear 2D polarization at frequencies $\omega_{p,i,s}$ generated in the graphene plane $z=0$ is given by

$$\hat{\boldsymbol{P}}^{(2)}_s = \boldsymbol{x}_0\chi^{(s,2)}_{xyy}\hat{\mathcal{E}}^{\dagger}_i\hat{\mathcal{E}}_p + \text{H.c.},$$
$$\hat{\boldsymbol{P}}^{(2)}_i = \boldsymbol{y}_0\chi^{(i,2)}_{yyx}E^*_{s0}\hat{a}^{\dagger}_s\hat{\mathcal{E}}_p e^{-iq_s x + i\omega_s t} + \text{H.c.},$$
$$\hat{\boldsymbol{P}}^{(2)}_p = \boldsymbol{y}_0\chi^{(p,2)}_{yyx}E_{s0}\hat{a}_s\hat{\mathcal{E}}_i e^{iq_s x - i\omega_s t} + \text{H.c.}, \tag{12}$$

where $\hat{\mathcal{E}}_{p,i}$ are the $\propto\exp(-i\omega t)$ parts of the electric field operators at the pump and idler frequencies $\omega_{p,i}$ in the graphene plane. The second-order susceptibilities at corresponding frequencies are $\chi^{(s,2)}_{xyy} = \chi^{(2)}_{xyy}(\omega_s = \omega_p - \omega_i)$, $\chi^{(i,2)}_{yyx} = \chi^{(2)}_{yyx}(\omega_i = \omega_p - \omega_s)$, $\chi^{(p,2)}_{yyx} = \chi^{(2)}_{yyx}(\omega_p = \omega_i + \omega_s)$. Index α in $\chi^{(2)}_{\alpha\beta\gamma}(\omega = \omega' \mp \omega'')$ corresponds to the polarization of the field at the mixing frequency ω, and the index β corresponds to the polarization of the field at the larger of the two frequencies ω', ω''.

Now we invoke the boundary conditions connecting the fields on both sides of the graphene layer. Besides the continuity of the electric field, we will use the relationships for magnetic field components:

$$\hat{\mathcal{B}}^{(i,p)}_z(z=+0) = \hat{\mathcal{B}}^{(i,p)}_z(z=-0),$$
$$\hat{\mathcal{B}}^{(i,p)}_x(z=+0) - \hat{\mathcal{B}}^{(i,p)}_x(z=-0) = -4\pi\frac{i\omega_{i,p}\hat{P}^{(2)}_{i,p}}{c},$$

where $\hat{\mathcal{B}}^{(i,p)}_{z,x}$ are operators of the magnetic field components that are related to the electric field operators by standard Maxwell's equations.

Using the nonlinear polarizations and boundary conditions for the fields, Eq. (9) becomes

$$\frac{\partial\hat{a}_s}{\partial t} + v_s\frac{\partial\hat{a}_s}{\partial x} + (\gamma_s - \hat{G})\cdot\hat{a}_s = \hat{J} + \hat{F}_s, \tag{13}$$

where

$$\hat{J} = \Gamma\chi^{(s,2)}_{xyy}\hat{c}^{\dagger}_i\hat{c}_p, \quad \Gamma = i\frac{2\pi\sqrt{\omega_i\omega_p}}{n_1^2}T_i T_p E^*_{s0},$$
$$\hat{G} = |\Gamma|^2\frac{n_1}{c}\left(\frac{\chi^{(s,2)}_{xyy}\chi^{(i,2)*}_{yyx}\hat{c}^{\dagger}_p\hat{c}_p}{T_i\cos\theta_{1i}} - \frac{\chi^{(s,2)}_{xyy}\chi^{(p,2)}_{yyx}\hat{c}^{\dagger}_i\hat{c}_i}{T_p\cos\theta_{1p}}\right).$$

Here $T_{p,i} = 2n_1\cos\theta_{1p,i}/(n_1\cos\theta_{1p,i} + n_2\cos\theta_{2p,i})$ are Fresnel transmission coefficients for S-polarized pump and idler fields with incidence angles $\theta_{1p,i}$ and refraction angles $\theta_{2p,i}$. Equation (13) was derived neglecting the terms of the order $\alpha|\chi^{(2)}|^2$ and $|\chi^{(2)}|^3$ where $\alpha = e^2/\hbar c$.

The terms $\hat{J}$ and $\hat{G}$ in Eq. (13) include all possible three-wave mixing processes. The term $\hat{J}$ describes difference frequency generation of surface plasmons in graphene by a bichromatic quantum field. For classical fields this process has been predicted in [7] and observed in [16]. The operator $\hat{G}$ describes the creation of plasmons by a parametric decay of the pump photons.

When solving operator-valued equations, in addition to the fields incident from $z>0$ one also needs to specify operators of noise fields incident from $z<0$ [13]. This allows one to take into account current fluctuations in a graphene layer caused by zero-point and thermal fluctuations of the field in the region $z<0$. It is easy to show that in this case one should replace $\hat{c}_{p,i} \to \hat{c}_{p,i} + \hat{c}^{(-)}_{p,i}, \hat{c}^{\dagger}_{p,i} \to \hat{c}^{\dagger}_{p,i} + \hat{c}^{\dagger(-)}_{p,i}$ in Eq. (13), where creation and annihilation operators $\hat{c}^{(-)}_{p,i}$ and $\hat{c}^{\dagger(-)}_{p,i}$ correspond to the waves incident on a graphene layer from $z<0$ at angles $\theta_{2(p,i)}$. The corresponding terms can be treated as modification of the Langevin source term in Eq. (13). However, under the condition $\hbar\omega_s \ll k_BT \ll \hbar\omega_p$ the effect of this modification on the plasmon field correlator $\langle\hat{a}^{\dagger}_s\hat{a}_s\rangle$ is negligible as compared to standard Langevin fluctuations given by Eq. (10).

The operator-valued Eq. (13) has a stationary solution given by

$$\hat{a}_s = \exp\left(\frac{\hat{G}-\gamma_s}{v_s}x\right) \times\left\{\hat{a}_s(0)+\int_0^x\left[\exp\left(\frac{\hat{G}-\gamma_s}{v_s}x'\right)\right]^{-1}(\hat{J}+\hat{F}_s)\frac{dx'}{v_s}\right\}, \tag{14}$$

where $\hat{a}_s(0)$ is the corresponding boundary condition; $[\exp(\frac{\hat{G}-\gamma_s}{v_s}x)]^{-1}$ is the operator inverse to $\exp(\frac{\hat{G}-\gamma_s}{v_s}x)$.

Below we will assume that the pump field at frequency ω_p is a coherent classical field, whereas the field at the idler frequency ω_i is present only as a quantum and/or thermal noise.

Whenever the contribution of electromagnetic noise incident from the half-space $z<0$ can be neglected, one can also neglect the term $\hat{J}$ as compared to the Langevin noise term. For a coherent pumping, the operator $\hat{G}$ can be replaced by a c number:

$$G \approx |\Gamma|^2\frac{n_1}{c}\frac{\chi_{xyy}^{(s,2)}\chi_{yyx}^{(i,2)*}\langle\hat{c}_p^\dagger\hat{c}_p\rangle}{T_i\cos\theta_{1i}}. \tag{15}$$

Taking the thermal noise as a boundary condition and taking into account Eqs. (10) and (14) one can get

$$\hat{a}_s^\dagger\hat{a}_s = \exp\left[2\frac{\mathrm{Re}[G]-\gamma_s}{v_s}x\right](\hat{a}_s^\dagger\hat{a}_s)_T \times\left[1+\frac{\gamma_s}{\mathrm{Re}[G]-\gamma_s}\left(1-\exp\left[-2\frac{\mathrm{Re}[G]-\gamma_s}{v_s}x\right]\right)\right], \tag{16}$$

where the operator $(\hat{a}_s^\dagger\hat{a}_s)_T$ corresponds to the thermal field and has an average value given by Eq. (11). Note that there is a $1/v_s$ dependence in the gain factor in Eq. (16) which describes the enhancement in the gain for slowly moving plasmons as compared to photons.

From Eq. (16) one can obtain an important result, namely the criterion for parametric instability:

$$\mathrm{Re}\left(\chi_{xyy}^{(s,2)}\chi_{yyx}^{(i,2)*}\right)>0, \tag{17}$$

$$\mathrm{Re}[G]\approx|\Gamma|^2\frac{\mathrm{Re}\left[\chi_{xyy}^{(s,2)}\chi_{yyx}^{(i,2)*}\right]I_p}{c^2\hbar\omega_p}\frac{n_1^2}{T_i\cos\theta_{1i}}>\gamma_s, \tag{18}$$

where I_p is the incident pump intensity.

B. Coupled oscillators model for the parametric gain

The instability condition Eqs. (17) and (18) can be easily interpreted and understood within the classical model of two parametrically coupled oscillators. Consider a classical pump beam of amplitude E_p and ω_p incident on a nonlinear 2D layer in vacuum. The pump field decays into a surface plasmon field within a unit area $A_s=1$ and an idler photon field at frequency ω_i within a volume of a cylinder of length l oriented at an angle θ_i with respect to the normal to area A_s. In this mean-field zero-dimensional (0D) model one can derive the following coupled differential equations for the complex amplitudes of the plasmon and idler fields:

$$\frac{\partial E_s}{\partial t}+\gamma_s E_s = i\zeta_s E_p E_i^*, \tag{19}$$

$$\frac{\partial E_i^*}{\partial t}+\gamma_i E_i^* = -i\zeta_i^* E_p^* E_s, \tag{20}$$

where

$$\zeta_s=\frac{1}{2}\chi_{xyy}^{(s,2)}\left[\mathrm{Re}\left(\frac{\partial\chi_s}{\partial\omega}\right)\right]^{-1},\quad \zeta_i=\frac{\pi}{l\cos\theta_i}\omega_i\chi_{yyx}^{(i,2)*},$$

$\gamma_i=c/l$ is the effective decay rate of the idler field in the 0D model. Equations (19) and (20) have an exponentially growing solution for both parametrically coupled waves [2] if $\mathrm{Re}(\zeta_s\zeta_i^*)|E_p|^2>\gamma_s\gamma_i$, which coincides with Eqs. (17) and (18) if we assume $n_2=n_1=1$.

C. Spectrum and magnitude of the parametric gain

To calculate the magnitude of the parametric gain we need to substitute the components of the second-order susceptibility tensor. Their derivation is straightforward but cumbersome, so we keep it in the Appendices. Their salient feature is the presence of resonances when one of the three frequencies involved in three-wave mixing is close to $2\epsilon_F=2\hbar v_F k_F$. This is a weaker resonance than the one that exists in coupled quantum wells [17] where $\chi^{(2)}$ would scale as a product of two Lorentzians. Still, it enhances the value of $\chi^{(2)}$ by a factor of ω/γ where γ is the decay rate of the optical polarization. A similar resonance exists in the third-order nonlinear response of graphene [18].

Far from resonance, when $|\omega_p-2v_Fk_F|\gg\gamma$, one can neglect dissipation. In this case all components of the nonlinear susceptibility tensor satisfy symmetry properties

$$\chi_{xyy}^{(s,2)}=\chi_{yyx}^{(i,2)}=\chi_{yyx}^{(p,2)*}, \tag{21}$$

which ensure Manley-Rowe relationships [19,20].

Close to resonance one has to include the imaginary part of the frequency which describes the decay rate of the optical or plasmon polarization. If dissipation is included, Eqs. (21) can be violated. In this case one has to use a more general procedure outlined in the Appendices. From the derivation in the Appendices one can obtain that if the resonance condition is satisfied for the idler photon frequency, $|\omega_i-2v_Fk_F|<\gamma$, then $\chi_{xyy}^{(s,2)}=-\chi_{yyx}^{(i,2)*}$. In this case inequality (17) is violated, i.e., the parametric instability is impossible. At the same time, when the resonance exists for the pump frequency, $|\omega_p-2v_Fk_F|<\gamma$, we obtain

$$\chi_{xyy}^{(s,2)}=\chi_{yyx}^{(i,2)*}\approx\frac{3e^3v_F^2}{8\pi\hbar^2}\frac{q_p}{\omega_i\omega_s^2\gamma}, \tag{22}$$

which satisfies Eq. (17). Therefore, we will assume that the frequency of the pump field is close to $2v_Fk_F$. We independently evaluated the $\chi_{xxx}^{(2)}$ component of the nonlinear susceptibility tensor (see Appendix B) and found that its value is three times lower at resonance as compared to Eq. (22). This translates into a more than an order of magnitude lower parametric gain for a P-polarized pump as compared

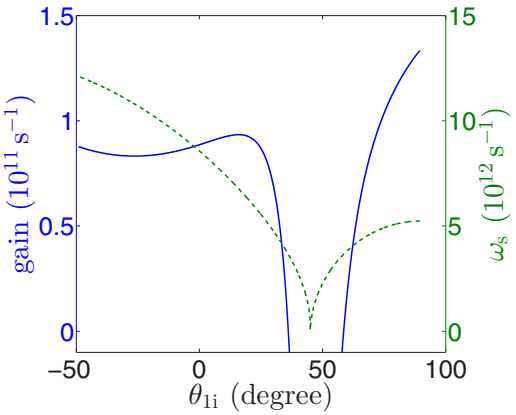

FIG. 2. Gain Re[G] (solid blue line) and the plasmon frequency corresponding to phase matching conditions (green dashed line) as a function of the angle θ_{1i} between the direction of the idler wave vector in medium 1 and the normal.

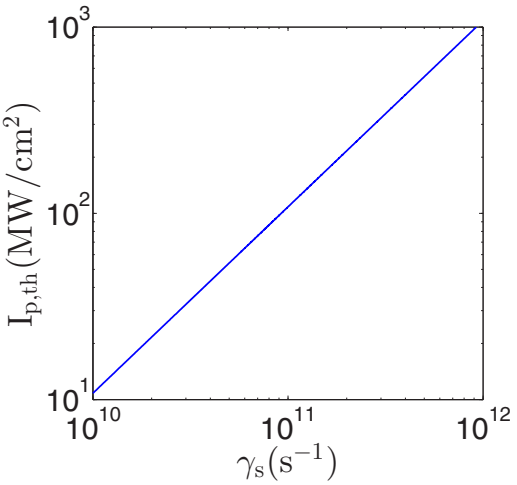

FIG. 3. Pumping intensity I_p needed to reach the parametric instability threshold, Re[G] = γ_s in Eq. (18), as a function of the plasmon decay rate γ_s.

to an S-polarized pump, if one takes into account a smaller in-plane projection of the electric field and the $|\chi^{(2)}|^2$ scaling of the gain. Furthermore, we assume $\omega_{p,i} \gg \omega_s \gg \gamma_s$ and consider strongly degenerate graphene. In addition to resonant enhancement of the nonlinearity, this eliminates interband absorption losses for the plasmons and reduces electron scattering.

Figure 2 shows the gain [left-hand side of Eq. (18)] and the plasmon frequency corresponding to phase matching conditions as a function of the emission angle of the idler field θ_{1i}. Negative angles correspond to negative projections of q_i. For the plot we took $n_1 = 1$ and $n_2 = 2$, $\gamma_{p,i} = 10^{12}$ s^{-1}, the pump beam at a 10-μm wavelength, and incidence angle of $\pi/4$. The pumping intensity was assumed to be $I_p = 100$ MW/cm^2, which is 10 times lower than the intensities in the experiment [16]. The gain is only weakly dependent on the idler emission angle except for a narrow range around $\theta_{1i} = \pi/4$, where $q_p \simeq q_i$ and therefore $q_s, \omega_s \to 0$. In this range the gain becomes negative; however, the approximation $\omega_s \gg \gamma_s$ becomes invalid, so this case requires a separate investigation.

In Fig. 3 we show the pumping intensity I_p needed to reach the parametric instability threshold, Re[G] = γ_s, as a function of plasmon decay rate γ_s, for the same numerical parameters as in Fig. 2 and for the idler emitted at $\theta_{1i} = 20^\circ$. In this case the phase matching condition is satisfied when the plasmon frequency $\omega_s/2\pi$ is equal to 1 THz (see Fig. 2).

The magnitude of the gain can be further increased by non-Bernal stacking of multiple graphene layers, which will reduce the threshold intensity.

Low-energy surface states of a 3D topological insulator Bi_2Se_3 are massless 2D Dirac fermions described by the effective Hamiltonian $H = v_F(\vec{\sigma} \times \vec{p})_z$ [21], where v_F is two times smaller than in graphene. The states have different chirality as compared to those in graphene but the same matrix elements of the interaction Hamiltonian and the same structure of the optical response. Repeating the same derivation, one can show that the parametric gain for a Bi_2Se_3 film (i.e., two uncoupled surfaces) will have a magnitude lower by a factor of $\sim 2^6$ due to a two times lower v_F and two times lower degeneracy.

IV. GENERATED IDLER FIELD FLUX

The outgoing flux of idler photons generated as a result of parametric decay of the pump carries information on the intensity of generated surface plasmon field, the surface nonlinearity, and even the quantum state of plasmons. Therefore, the detection of idler photons is a valuable diagnostic tool, especially in the experiments where the direct detection of surface plasmons is problematic. To calculate the operator of the idler field generated by the nonlinear current $\frac{\partial}{\partial t}\hat{\boldsymbol{P}}_i^{(2)}$, we use Eqs. (12) and standard boundary conditions from the previous section, arriving at

$$\delta\hat{\boldsymbol{E}}_i = \boldsymbol{y}_0 E_{i0}\hat{c}_i^{(2)} e^{ik_i z + iq_i x - i\omega_i t} + \text{H.c.}, \quad \hat{c}_i^{(2)} = \frac{n_1 \Gamma \chi_{yyx}^{(i,2)}}{c\,\cos\theta_{1i}} \hat{a}_s^\dagger \hat{c}_p.$$

If one only needs to know the average flux of the idler photons on the detector of transverse area A_D, $\langle\hat{\Pi}_i^{(2)}\rangle = \langle\frac{cA_D}{n_1}\hat{c}_i^\dagger\hat{c}_i\rangle$, it is enough to calculate the average value of the plasmon quanta $\langle\hat{a}_s^\dagger\hat{a}_s\rangle$ generated from length L_x. Using Eq. (16) we obtain

$$\langle\hat{a}_s^\dagger\hat{a}_s\rangle = \frac{1}{L_x}\int_0^x \langle\hat{a}_s^\dagger\hat{a}_s\rangle dx \approx \langle\hat{a}_s^\dagger\hat{a}_s\rangle_T \frac{e^\Xi - 1}{\Xi},$$
$$\Xi = 2\frac{\text{Re}[G] - \gamma_s}{v_s} L_x. \tag{23}$$

The resulting average flux of the idler photons on the detector is given by

$$\langle\hat{\Pi}_i^{(2)}\rangle = \frac{n_1^2|\Gamma|^2\left|\chi_{yyx}^{(i,2)}\right|^2 I_p L_x \Delta\omega}{2\pi c^2 v_s \hbar\omega_p \cos\theta_{1i}}\left(\frac{e^\Xi - 1}{\Xi}N_T(\omega_s) + 1\right).$$

Close to the parametric instability threshold, when $\frac{e^\Xi - 1}{\Xi} \sim 1$, the idler photon flux is

$$\langle\hat{\Pi}_i^{(2)}\rangle \sim \frac{\Delta\omega\gamma_s L_x}{2\pi v_s} N_T(\omega_s), \tag{24}$$

i.e., it is of the order of the thermal flux at a much lower surface plasmon frequency, $N_T(\omega_s) \gg N_T(\omega_p)$, collected from the length equal to the plasmon decay length, $L_x \sim v_s/\gamma_s$. Far above threshold, both idler and surface plasmon fluxes increase exponentially, $\propto e^{\Xi}$.

V. PLASMON-PHOTON ENTANGLEMENT

The total idler field propagating away from the graphene layer to the detector consists of the reflected and transmitted noise field and the generated parametric field calculated above:

$$\hat{c}_r \approx R_i \hat{c}_i + \sqrt{1-R_i^2}\hat{c}_i^{(-)} + \frac{n_1 \Gamma \chi_{yyx}^{(i,2)}}{c\cos\theta_{1i}} \hat{a}_s^\dagger \hat{c}_p, \qquad (25)$$

where $R_i = \frac{n_1\cos\theta_{1i} - n_2\cos\theta_{2i}}{n_1\cos\theta_{1i} + n_2\cos\theta_{2i}}$ is the Fresnel reflection coefficient for the S-polarized field. Here we neglected absorption in monolayer graphene $\sim\pi\alpha$. It could be easily included by redefining R_i. Calculating quantum-mechanical averages of the quantities quadratic with respect to the reflected field, it is easy to see that Eq. (25) corresponds to an entangled plasmon-photon state (see also [11]).

In general, the calculations of quantum-mechanical averages of any physical quantities are much easier to perform in the Heisenberg picture using Eq. (25) for Heisenberg operators, without converting to the Schrödinger picture. In particular, it is obvious from Eq. (25) that for a given spectrum of the pump field any physical observable for a surface plasmon field can be related to a corresponding observable for the idler field at frequency ω_i. For example, if the pump field spectrum is much narrower than the spectrum of the plasmon fluctuations, then the spectrum of surface plasmons is related to the spectrum of idler photons. However, to demonstrate how the entangled state is formed, we will discuss the Schrödinger picture as well, with certain simplifications. Namely, consider the equation of motion Eq. (9) for the plasmon field in the particular case of single-mode fields, classical pumping, and neglecting dissipation. This means that we take $\mathcal{E}_p$ to be a c number, $\partial/\partial x \to 0, \gamma, \hat{F}_s \to 0$, and $\chi_{xyy}^{(s,2)} = \chi_{yyx}^{(i,2)} = \chi^{(2)}$. We will also assume for simplicity that the media on both sides of the graphene sheet have dielectric constants equal to 1. In this case Eqs. (9) and (12) yield

$$\dot{\hat{a}}_s = \frac{i}{\hbar}\chi^{(2)}\mathcal{E}_p E_{s0}^* E_{i0}^* \hat{c}_i^\dagger, \qquad (26)$$

and the Hermitian-conjugate equation

$$\dot{\hat{a}}_s^\dagger = \frac{i}{\hbar}\chi^{(2)}\mathcal{E}_p^* E_{s0} E_{i0} \hat{c}_i. \qquad (27)$$

Since for single-mode fields the normalizations we chose correspond to $[\hat{a}_s, \hat{a}_s^\dagger] = [\hat{c}_s, \hat{c}_s^\dagger] = 1$, Eqs. (26) and (27) can be interpreted as Heisenberg equations in the interaction representation:

$$\dot{\hat{a}}_s = \frac{i}{\hbar}[\hat{V}, \hat{a}_s], \quad \dot{\hat{a}}_s^\dagger = \frac{i}{\hbar}[\hat{V}, \hat{a}_s^\dagger]. \qquad (28)$$

Here the interaction Hamiltonian is

$$\hat{V} = -\xi \hat{a}_s^\dagger \hat{c}_i^\dagger - \xi^* \hat{a}_s \hat{c}_i, \qquad (29)$$

where $\xi = \chi^{(2)}\mathcal{E}_p E_{s0}^* E_{i0}^*$.

Now let's solve the Schrödinger equation, also in the interaction representation:

$$i\hbar\dot{\Psi} = \hat{V}\Psi. \qquad (30)$$

Starting for simplicity with the initial condition in the form $\Psi(t=0) = |0\rangle_s|0\rangle_i$, the solution to Eq. (30) can be written as

$$\Psi = \sum_{n=0}^{\infty} C_n |n\rangle_s |n\rangle_i, \qquad (31)$$

where the coefficients C_n can be found from equations

$$\dot{C}_0 + \frac{i}{\hbar}\xi^* C_1 = 0,$$

$$\dot{C}_{n\neq 0} - \frac{i}{\hbar}\xi n C_{n-1} + \frac{i}{\hbar}\xi^*(n+1)C_{n+1} = 0, \qquad (32)$$

with initial conditions $C_0(t=0) = 1$, $C_{n\neq 0} = 0$. It is clear that at $t > 0$ the solution (31) is entangled, i.e., it cannot be factorized as $\Psi = (\sum_{n=0}^{\infty} A_n|n\rangle_s)(\sum_{n=0}^{\infty} B_n|n\rangle_i)$. In particular, within the perturbation theory, the solution can be expressed via Bell states $\Phi_\pm = (1/\sqrt{2})(|0\rangle_s|0\rangle_i \pm |1\rangle_s|1\rangle_i)$:

$$\Psi \approx \frac{|0\rangle_s|0\rangle_i + \tilde{C}|1\rangle_s|1\rangle_i}{\sqrt{1+|\tilde{C}|^2}} = \frac{1+\tilde{C}}{\sqrt{1+|\tilde{C}|^2}}\frac{\Phi_+}{\sqrt{2}} + \frac{1-\tilde{C}}{\sqrt{1+|\tilde{C}|^2}}\frac{\Phi_-}{\sqrt{2}}, \qquad (33)$$

where $\tilde{C} = \frac{i}{\hbar}\xi t \ll 1$.

In conclusion, we showed the feasibility of observing both spontaneous and stimulated parametric decay of photons of a strong laser pump obliquely incident on graphene. We calculated the flux of surface plasmons and idler photons generated by parametric decay of the pump, and demonstrated their entanglement. A rigorous quantum theory of the process including quantization of all fields and fluctuations has been developed.

ACKNOWLEDGMENTS

This work has been supported by the Air Force Office for Scientific Research through Grants No. FA9550-15-1-0153 and No. FA9550-14-1-0376. M.T. and I.O. acknowledge support from the Russian Foundation for Basic Research Grant No. 14-22-02034.

APPENDIX A: NORMALIZATION OF THE SURFACE PLASMON FIELD

The normalization constant E_{s0} in Eq. (7) can be obtained by evaluating the normalization integral (6):

$$\int_V \left(\boldsymbol{E}_s^* \frac{\partial(\omega\tilde{\epsilon})}{\partial\omega}\boldsymbol{E}_s + \boldsymbol{B}_s\boldsymbol{B}_s^* \right) d^3r = 4\pi\hbar\omega_s. \qquad (A1)$$

Here $\tilde{\epsilon}$ is the dielectric permittivity tensor; the volume V is formed by a closed cylindrical surface that crosses the (x,y) plane along the boundary of the area $A = 1$.

One can show that when the flux of the complex vector $\boldsymbol{E}_s \times \boldsymbol{B}_s^*$ through the surface of the quantized volume is equal to zero (which is the case, for example, for periodic

boundary conditions or in a resonator), then $\int_V d^3r\, \boldsymbol{B}_s \boldsymbol{B}_s^* = \int_V d^3r\, \boldsymbol{E}_s \tilde{\epsilon} \boldsymbol{E}_s^*$. This condition allows one to transform the volume integral in Eq. (A1) in the following way:

$$\int_V \left(\boldsymbol{E}_s^* \frac{\partial(\omega\tilde{\epsilon})}{\partial\omega} \boldsymbol{E}_s + \boldsymbol{B}_s \boldsymbol{B}_s^* \right) d^3r = \int_V \boldsymbol{E}_s^* \frac{\partial(\omega^2\tilde{\epsilon})}{\omega\partial\omega} \boldsymbol{E}_s d^3r = \int_{-\infty}^{+\infty} \boldsymbol{E}_s^* \frac{\partial(\omega^2\tilde{\epsilon})}{\omega\partial\omega} \boldsymbol{E}_s dz. \quad \text{(A2)}$$

Next, we substitute into Eq. (A2) the z dependence of the dielectric permittivity which follows from the geometry of the system:

$$\epsilon_{xy} = \epsilon_{yx} = \epsilon_{xz} = \epsilon_{zx} = \epsilon_{yz} = \epsilon_{zy} = 0,$$

$$\epsilon_{xx} = \epsilon_{yy} = \begin{cases} \epsilon_1 & \text{for} \quad z > 0, \\ 1 + 4\pi\chi_s\delta(z) & \text{for} \quad z = 0, \\ \epsilon_2 & \text{for} \quad z < 0, \end{cases}$$

$$\epsilon_{zz} = \begin{cases} \epsilon_1 & \text{for} \quad z > 0, \\ 1 & \text{for} \quad z = 0, \\ \epsilon_2 & \text{for} \quad z < 0, \end{cases}$$

and use Eqs. (A1) and (A2) with Eq. (4) to arrive at

$$|E_{s0}|^2 \left[\left(\frac{q_s^2}{p_1^2} - 1 \right) \frac{\epsilon_1}{p_1} + \left(\frac{q_s^2}{p_2^2} - 1 \right) \frac{\epsilon_2}{p_2} + 4\pi\omega_s Re\left(\frac{\partial\chi_s}{\partial\omega} \right) \right] = 4\pi\hbar\omega_s. \quad \text{(A3)}$$

In the quasielectrostatic approximation $p_{1,2}^2 \to q_s^2$ and we obtain Eq. (7), where χ_s is given by Eq. (3).

APPENDIX B: SECOND-ORDER NONLINEAR SUSCEPTIBILITY IN GRAPHENE

Here we provide the general expressions for the components of the second-order susceptibility tensor that are relevant for the parametric three-wave mixing in graphene.

The Hamiltonian of graphene near the Dirac point $\boldsymbol{K}$ is

$$H = v_F \boldsymbol{\sigma} \cdot \hat{\boldsymbol{p}} = v_F \begin{pmatrix} 0 & \hat{p}_x - i\hat{p}_y \\ \hat{p}_x + i\hat{p}_y & 0 \end{pmatrix}, \quad \text{(B1)}$$

where $\hat{\boldsymbol{p}}$ is the momentum operator relative to $\boldsymbol{K}$ and $\boldsymbol{\sigma}$ is a 2D vector of Pauli matrices. The eigenenergies are $\epsilon_\pm(\boldsymbol{k}) = \pm\hbar v_F k$, and eigenstates are

$$\langle \boldsymbol{r}_\parallel | s, \boldsymbol{k} \rangle = \frac{1}{\sqrt{2A}} \exp(i\boldsymbol{k}\cdot\boldsymbol{r}_\parallel) \begin{pmatrix} s \\ e^{i\phi(\boldsymbol{k})} \end{pmatrix}, \quad \text{(B2)}$$

where $s = 1$ for conduction band, $s = -1$ for valence band, A is the area of graphene, and $\phi(\boldsymbol{k})$ is the angle of the wave vector $\boldsymbol{k}$ with the x axis.

The interaction Hamiltonian between graphene and the optical field which has an in-plane component of the electric field can be written as

$$\hat{H}_{int}^{op} = v_F \frac{e}{c} \boldsymbol{\sigma} \cdot \boldsymbol{A} = \frac{e}{c} \hat{\boldsymbol{v}} \cdot \boldsymbol{A}, \quad \text{(B3)}$$

where $\hat{\boldsymbol{v}} = v_F\boldsymbol{\sigma}$ is the velocity operator, and $\boldsymbol{A}$ is the vector potential of the optical field, which is related to the electric field by $\boldsymbol{E} = (-1/c)\partial\boldsymbol{A}/\partial t$. Using this Hamiltonian, the evolution equation for the density matrix is given by

$$i\hbar \frac{\partial}{\partial t} \rho_{mn} = (\epsilon_m - \epsilon_n)\rho_{mn} + \frac{e}{c}(\hat{\boldsymbol{v}} \cdot \boldsymbol{A})_{mn}(\rho_{nn} - \rho_{mm}) + \frac{e}{c} \sum_{l \neq m,n} ((\hat{\boldsymbol{v}} \cdot \boldsymbol{A})_{ml}\rho_{ln} - \rho_{ml}(\boldsymbol{v} \cdot \boldsymbol{A})_{ln}), \quad \text{(B4)}$$

where both linear and nonlinear effects are included. We calculate the field-induced current in second order with respect to the optical field, as a quantum-mechanical average of the current operator $\hat{\boldsymbol{j}} = -ev_F\boldsymbol{\sigma}$ with the density matrix.

We will seek the response at the sum frequency $\omega_1 + \omega_2$ to the bichromatic optical field with in-plane electric fields at frequencies $\omega_{1,2}$ directed along unit vectors $\boldsymbol{\eta}_{1,2}$:

$$\boldsymbol{A} = \boldsymbol{\eta}_1 A(\omega_1) e^{i(\boldsymbol{q}_1\cdot\boldsymbol{r}_\parallel - \omega_1 t)} + \boldsymbol{\eta}_2 A(\omega_2) e^{i(\boldsymbol{q}_2\cdot\boldsymbol{r}_\parallel - \omega_2 t)} + \text{c.c.} \quad \text{(B5)}$$

The result will be applicable to the difference frequency process by choosing either positive or negative frequencies, with the corresponding change in $\boldsymbol{q}$ for a given ω. The second-order density matrix elements at the sum frequency $\omega_1 + \omega_2$ are evaluated to be

$$\rho_{mn}^{(2)}(\omega_1 + \omega_2) = \left(\frac{e}{c} \right)^2 \frac{A(\omega_1)A(\omega_2)}{\hbar(\omega_1 + \omega_2) - (\epsilon_m - \epsilon_n)} \times \sum_{l \neq m,n} ((\hat{\boldsymbol{v}} \cdot \boldsymbol{\eta}_1) e^{i\boldsymbol{q}_1\cdot\boldsymbol{r}_\parallel})_{ml} ((\hat{\boldsymbol{v}} \cdot \boldsymbol{\eta}_2) e^{i\boldsymbol{q}_2\cdot\boldsymbol{r}_\parallel})_{ln} \times \left[\frac{(\rho_{nn} - \rho_{ll})}{\hbar\omega_2 - (\epsilon_l - \epsilon_n)} - \frac{(\rho_{ll} - \rho_{mm})}{\hbar\omega_1 - (\epsilon_m - \epsilon_l)} \right] + \{1 \leftrightarrow 2\}. \quad \text{(B6)}$$

The matrix elements entering the above expression are given by

$$((\hat{\boldsymbol{v}} \cdot \boldsymbol{\eta}) e^{i\boldsymbol{q}\cdot\boldsymbol{r}_\parallel})_{mn} = \tfrac{1}{2} v_F [(\eta_x - i\eta_y) s_m e^{i\phi_n} + (\eta_x + i\eta_y) s_n e^{-i\phi_m}] \delta_{\boldsymbol{k}_m, \boldsymbol{k}_n + \boldsymbol{q}}. \quad \text{(B7)}$$

The average of the corresponding Fourier harmonic of the induced current with the density matrix can be calculated as

$$\boldsymbol{J}^{(2)}(\omega_1 + \omega_2) = -e \sum_{mn} (\hat{\boldsymbol{v}} e^{-i(\boldsymbol{q}_1 + \boldsymbol{q}_2)\cdot\boldsymbol{r}_\parallel})_{nm} \rho_{mn}^{(2)}(\omega_1 + \omega_2). \quad \text{(B8)}$$

Next, we transform from summation to integration over $\boldsymbol{k}$ states, introduce the corresponding occupation numbers $f(s,\boldsymbol{k})$ of the momentum states in each band, apply the momentum conservation in a three-wave mixing process, and

take into account spin and valley degeneracy. The result is

$$\begin{aligned}\boldsymbol{J}^{(2)}(\omega_1+\omega_2) = &-\frac{e^3v_F^2}{8\pi^2c^2\hbar^2}A(\omega_1)A(\omega_2)\sum_{s_m,s_n,s_l}\int d^2\boldsymbol{k}\frac{1}{(\omega_1+\omega_2)-v_F(s_m|\boldsymbol{k}+\boldsymbol{q}_1|-s_n|\boldsymbol{k}-\boldsymbol{q}_2|)}\\ &\times\left[\frac{f(s_n,|\boldsymbol{k}-\boldsymbol{q}_2|)-f(s_l,|\boldsymbol{k}|)}{\omega_2-v_F(s_l|\boldsymbol{k}|-s_n|\boldsymbol{k}-\boldsymbol{q}_2|)}-\frac{f(s_l,|\boldsymbol{k}|)-f(s_m,|\boldsymbol{k}+\boldsymbol{q}_1|)}{\omega_1-v_F(s_m|\boldsymbol{k}+\boldsymbol{q}_1|-s_l|\boldsymbol{k}|)}\right]\\ &\times[(\eta_{1x}-i\eta_{1y})s_me^{i\phi(\boldsymbol{k})}+(\eta_{1x}+i\eta_{1y})s_le^{-i\phi(\boldsymbol{k}+\boldsymbol{q}_1)}][(\eta_{2x}-i\eta_{2y})s_le^{i\phi(\boldsymbol{k}-\boldsymbol{q}_2)}+(\eta_{2x}+i\eta_{2y})s_ne^{-i\phi(\boldsymbol{k})}]\\ &\times[(\hat{x}+i\hat{y})s_me^{-i\phi(\boldsymbol{k}-\boldsymbol{q}_2)}+(\hat{x}-i\hat{y})s_ne^{i\phi(\boldsymbol{k}+\boldsymbol{q}_1)}]+\{1\leftrightarrow 2\}. \end{aligned}\tag{B9}$$

This equation can be integrated numerically for any given geometry of incident fields and electron distribution. We consider the limit of the Fermi distribution with a strong degeneracy, direct all in-plane photon wave vectors along x axis, and expand the integrand in Eq. (B9) in powers of q_1,q_2. The integral over the term of zeroth order in q vanishes, as expected from symmetry. We will keep the terms linear in q. Also we have to evaluate separately the intraband contribution $s_l=s_m=s_n$ and all types of mixed interband-intraband contributions: $s_m=s_n=-s_l$, $s_m=s_l=-s_n$, and $s_n=s_l=-s_m$. Here we give only the component of the second-order nonlinear conductivity tensor which gives the main contribution to the signal for an S-polarized pump:

$$\begin{aligned}\sigma^{(2)}_{xyy}(\omega_1+\omega_2;\omega_1,\omega_2) = &-s(\epsilon_F)\frac{e^3v_F^2}{\pi\hbar^2}\frac{1}{\omega_1^2\omega_2^2(\omega_1+\omega_2)\left(\omega_1^2-4v_F^2k_F^2\right)\left(\omega_2^2-4v_F^2k_F^2\right)\left((\omega_1+\omega_2)^2-4v_F^2k_F^2\right)}\\ &\times\left[4(v_Fk_F)^2\omega_1\omega_2(\omega_1+\omega_2)^2\left(q_1\omega_2^2+q_2\omega_1^2\right)\right.\\ &+4(v_Fk_F)^4\left(q_1\omega_2^4-(6q_1+4q_2)\omega_1\omega_2^3-8(q_1+q_2)\omega_1^2\omega_2^2-(4q_1+6q_2)\omega_1^3\omega_2+q_2\omega_1^4\right)\\ &\left.-16(v_Fk_F)^6(q_1\omega_2(\omega_2-2\omega_1)+q_2\omega_1(\omega_1-2\omega_2))\right]. \end{aligned}\tag{B10}$$

Here $s(\epsilon_F)=\pm1$ depending on whether the Fermi level is in the conduction or valence band.

For completeness, we also give the xxx component of the nonlinear conductivity which determines the second-order response to a P-polarized pump:

$$\begin{aligned}\sigma^{(2)}_{xxx}(\omega_1+\omega_2;\omega_1,\omega_2) = &\,s(\epsilon_F)\frac{e^3v_F^2}{4\pi\hbar^2\omega_1\omega_2}\left(\frac{q_1+q_2}{\omega_1+\omega_2}+\frac{q_1}{\omega_1}+\frac{q_2}{\omega_2}\right)\\ &-s(\epsilon_F)\frac{e^3v_F^2}{4\pi\hbar^2}\frac{1}{\omega_1^2\omega_2^2(\omega_1+\omega_2)\left(\omega_1^2-4v_F^2k_F^2\right)\left(\omega_2^2-4v_F^2k_F^2\right)\left((\omega_1+\omega_2)^2-4v_F^2k_F^2\right)}\\ &\times\left[\omega_1^2\omega_2^2(\omega_1+\omega_2)^2(q_1\omega_2(2\omega_1+\omega_2)+q_2\omega_1(\omega_1+2\omega_2))\right.\\ &-4v_F^2k_F^2\left(\omega_1^2+\omega_1\omega_2+\omega_2^2\right)^2(q_1\omega_2(2\omega_1+\omega_2)+q_2\omega_1(\omega_1+2\omega_2))\\ &\left.+16v_F^4k_F^4\left(q_2\omega_1^4+2(q_1+q_2)\omega_1\omega_2\left(2\omega_1^2+3\omega_1\omega_2+2\omega_2^2\right)+q_1\omega_2^4\right)\right]. \end{aligned}\tag{B11}$$

The result for the difference frequency can be obtained from Eqs. (B10) and (B11) by flipping the sign of ω_2 and q_2. After converting the nonlinear conductivity to the nonlinear susceptibility according to

$$\chi^{(2)}_{ijk}(\omega_1+\omega_2;\omega_1,\omega_2)=\frac{i\sigma^{(2)}_{ijk}(\omega_1+\omega_2;\omega_1,\omega_2)}{\omega_1+\omega_2},$$

one can verify that in the absence of dissipation all components of the nonlinear susceptibility tensor that we calculated satisfy permutation relations originated from symmetry properties; see, e.g., Chap. 2.9 in [19]:

$$\chi^{(2)}_{ijk}(\omega_3=\omega_1+\omega_2)=\chi^{(2)}_{jik}(-\omega_1=-\omega_3+\omega_2)=\chi^{(2)}_{kji}(-\omega_2=-\omega_3+\omega_1),\tag{B12}$$

where in-plane wave vectors have to be permuted together with frequencies.

The second-order response goes to zero when the Fermi energy ϵ_F goes to zero, and is maximized when one of the three frequencies involved in three-wave mixing is close to $2\epsilon_F/\hbar=2v_Fk_F$. Close to resonance with $2\epsilon_F/\hbar$ one has to include the imaginary part of the frequency which comes from the omitted relaxation term $-\gamma\rho_{mn}$ in the density-matrix equations. This amounts to substituting $\omega_1\to\omega_1+i\gamma_1,\omega_2\to\omega_2+i\gamma_2,\omega_1+\omega_2\to\omega_1+\omega_2+i\gamma_3$. Note that if we flip the sign of ω_2 the sign of $+i\gamma_2$ remains the same. Even if dissipation is included we can still use Eqs. (B12) to derive the components of the nonlinear susceptibility tensor from other components. In order to do that, one needs to use Eqs. (B12) in the absence of dissipation and then add imaginary parts of frequencies. Of course the resulting expressions after adding dissipation won't satisfy Eqs. (B12).

[1] P. G. Kwiat, K. Mattle, H. Weinfurter, A. Zeilinger, A. V. Sergienko, and Y. Shih, Phys. Rev. Lett. **75**, 4337 (1995).
[2] Y. R. Shen, *The Principles of Nonlinear Optics* (Wiley, Hoboken, NJ, 2003).
[3] J. J. Dean and H. M. van Driel, Phys. Rev. B **82**, 125411 (2010).
[4] S. A. Mikhailov and K. Ziegler, J. Phys.: Condens. Matter **20**, 384204 (2008).
[5] S. A. Mikhailov, Phys. Rev. B **84**, 045432 (2011).
[6] D. A. Smirnova, I. V. Shadrivov, A. E. Miroshnichenko, A. I. Smirnov, and Y. S. Kivshar, Phys. Rev. B **90**, 035412 (2014).
[7] X. Yao, M. Tokman, and A. Belyanin, Phys. Rev. Lett. **112**, 055501 (2014).
[8] M. O. Scully and M. S. Zubairy, *Quantum Optics* (Cambridge University Press, Cambridge, UK, 1997).
[9] M. D. Lukin, Rev. Mod. Phys. **75**, 457 (2003).
[10] V. Vdovin and M. Tokman, Phys. Rev. A **87**, 012323 (2013).
[11] M. Tokman, X. Yao, and A. Belyanin, Phys. Rev. Lett. **110**, 077404 (2013).
[12] V. M. Fain and Ya. I. Khanin, *Quantum Electronics* (The MIT Press, Cambridge, MA, 1969), Vol. 1.
[13] M. Tokman, Y. Wang, and A. Belyanin, Phys. Rev. B **92**, 075409 (2015).
[14] M. D. Tokman, M. A. Erukhimova, and V. V. Vdovin, Ann. Phys. (NY) **360**, 571 (2015).
[15] G. Bekefi, *Radiation Processes in Plasmas* (Wiley, New York, 1966).
[16] T. J. Constant, S. M. Hornett, D. E. Chang, and E. Hendry, Nat. Phys. **12**, 124 (2016).
[17] M. Belkin, F. Capasso, A. Belyanin, D. L. Sivco, A. Y. Cho, D. C. Oakley, C. J. Vineis, and G. W. Turner, Nat. Photon. **1**, 288 (2007).
[18] S. A. Mikhailov, Phys. Rev. B **90**, 241301(R) (2014).
[19] Yu. A. Il'inskii and L. V. Keldysh, *Electromagnetic Response of Material Media* (Springer, New York, 1994).
[20] M. D. Tokman and M. A Erukhimova, Phys. Rev. E **84**, 056610 (2011).
[21] H. J. Zhang, C. X. Liu, X. L. Qi, X. Dai, Z. Fang, and S.-C. Zhang, Nat. Phys. **5**, 438 (2009).