

Free electron nonlinearities in heavily doped semiconductors plasmonics

Federico De Luca^{1,2}, Michele Ortolani^{3,4} and Cristian Ciraci^{1,*}

¹*Istituto Italiano di Tecnologia, Center for Biomolecular Nanotechnologies, Via Barsanti 14, 73010 Arnesano, Italy*

²*Dipartimento di Matematica e Fisica “E. De Giorgi,” Università del Salento, via Arnesano, 73100 Lecce, Italy*

³*Department of Physics, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy*

⁴*Istituto Italiano di Tecnologia, Center for Life NanoSciences, Viale Regina Elena 291, 00161 Rome, Italy*



(Received 11 December 2020; revised 14 February 2021; accepted 22 February 2021; published 10 March 2021)

Heavily doped semiconductors have emerged as tunable low-loss plasmonic materials at midinfrared frequencies. In this article we investigate nonlinear optical phenomena associated with high concentration of free electrons. We use a hydrodynamic description to study free electron dynamics in heavily doped semiconductors up to third-order terms, which are usually negligible for noble metals. We find that cascaded third-harmonic generation due to second-harmonic signals can be as strong as direct third-harmonic generation contributions even when the second-harmonic generation efficiency is zero. Moreover, we show that when coupled with plasmonic enhancement free electron nonlinearities could be up to two orders of magnitude larger than conventional semiconductor nonlinearities. Our work might open a new route for nonlinear optical integrated devices at midinfrared frequencies.

DOI: [10.1103/PhysRevB.103.115305](https://doi.org/10.1103/PhysRevB.103.115305)

I. INTRODUCTION

Plasmonic nanoantennas emerged as a disruptive technology for concentrating and controlling light at the nanoscale thanks to their unique capabilities of field localization and enhancement [1–6]. Nonetheless, their widespread application has struggled to concretize because of several shortcomings of the main constituent of plasmonic materials: noble metals. The use of noble metals for practical devices has been limited due to, above all, high losses, high reflectivity, and very poor compatibility with standard microelectronics manufacturing processes. In this context, heavily doped semiconductors (i.e., with charge densities $n_0 \sim 10^{19}$ – 10^{20} cm^{−3}) have been introduced as low-loss and tunable alternative materials for plasmonics in the near-infrared (NIR), i.e., $0.8 < \lambda < 2$ μm, and in the midinfrared (MIR), i.e., $2 < \lambda < 20$ μm [7,8], λ being the free-space wavelength. The MIR overlaps with the so-called molecular fingerprint region and is particularly important for sensing applications in many fields from healthcare to security [9]. Moreover, interest in MIR devices is growing in the context of free-space communications since the 3–5-μm and the 8–13-μm regions are less affected by atmospheric turbulence, clouds and fog than the NIR region [10]. Because semiconductors offer the possibility of tuning their optical response through doping or electrical and optical excitation, and some of them are also highly compatible with silicon complementary metal-oxide semiconductor technology, their use for plasmonics may lead to breakthrough technologies [9].

Plasmonic systems owe their properties to surface-plasmon modes—the resonant collective oscillations of free electrons

(FEs)—that appear in materials with a high carrier concentration (i.e., metals and heavily doped semiconductors) as a consequence of the interaction with an external light excitation. Naturally, the local-field enhancement provided by these modes especially affects optical processes whose dependence on the field intensity is nonlinear. Indeed, this is extremely important for nonlinear optical effects since they are in general very weak and normally require high laser intensities and long propagation distances in macroscopic nonlinear crystals to reach operational efficiencies. Although, recently, high-index dielectric resonators have been used to partially overcome these limitations [11–13], plasmonic nanoantennas made of noble metals have been commonly used as local-field amplifier in hybrid systems to boost optical nonlinearity from dielectric material placed in their vicinity [14–27]. The nonlinear response may also originate directly from the plasmonic material itself [28–30] and, in particular, from the dynamics of nonequilibrium FEs [31–36]. For instance, in noble metals, FE plasmonic nonlinearities have been shown to strongly contribute to second-order processes [34,37,38] in the visible, NIR, and experimental measurements in gold nanoparticle arrays have demonstrated second-harmonic generation (SHG) efficiencies comparable to those of nonlinear crystals when normalized to the active volumes [39,40].

Although FE optical nonlinearities have mostly been observed in metals, analogous effects may also occur in heavily doped semiconductors. The dynamic properties of doped semiconductors undergo, in fact, an interesting transition from the size-quantization regime to the classical regime of plasmon oscillations. In the former regime, semiconductor quantum wells have enabled the strongest nonlinear susceptibilities [21]. The resonant nature of electron transitions in quantum wells, however, severely hinders their application in ultrafast devices since it inevitably causes a slower response

*cristian.ciraci@iit.it

and increased decoherence. The plasmonic regime for semiconductors, on the other hand, unlocks a new realm of possible ultrafast nonlinear effects [41,42] and offers several advantages with respect to nonlinear plasmonics in noble metals.

To understand the advantage of using heavily doped semiconductors, it should be noted that within the hydrodynamic description, third-order nonlinear effects are inversely proportional to the square of the equilibrium free-carrier densities, i.e., $\mathbf{P}_{\text{NL}}^{(3)} \propto \frac{1}{n_0^2}$ (as we will show later), where $\mathbf{P}_{\text{NL}}^{(3)}$ is the third-order polarization vector. In gold, for example, the high concentration of free carriers ($n_0 \sim 10^{22} \text{ cm}^{-3}$), leads to very weak contributions to the third-order nonlinear polarization if compared to contributions due to the crystal lattice nonlinear susceptibility $\chi^{(3)}$. On the other hand, doped semiconductors with a plasma frequency in the MIR have a much lower charge density ($n_0 \sim 10^{19} \text{ cm}^{-3}$) than noble metals such that nonlinear terms may grow as much as six orders of magnitude, overcoming by far the lattice nonlinearities. Moreover, semiconductors have an effective electron mass m that can be one order of magnitude smaller than that of noble metals. These characteristics cause also an increase in the nonlinear active volumes, whose measure (normalized to the operation wavelength) can be linked [43,44] to the Fermi velocity $v_F \propto \frac{n_0^{1/3}}{m}$.

The large variety of semiconductors with different effective electron mass and doping levels give access to a wide range of possibilities to optimize and increase intrinsic nonlinear effects. Furthermore, these quantities, unlike in metals, can also be used to control their optical linear properties. Indeed, assuming a Drude model, a doped semiconductor is characterized by a dielectric function $\varepsilon(\omega) = \varepsilon_\infty - \frac{\omega_p^2}{\omega^2 + i\gamma\omega}$, where ε_∞ is the background permittivity, γ is the damping rate, and $\omega_p = \sqrt{\frac{e^2 n_0}{\varepsilon_0 m}}$ is the plasma frequency of the material, being e and ε_0 the elementary charge (absolute value) and the dielectric constant of vacuum (note that hole doping is unsuitable for plasmonics due to inter-valence-band transitions taking place in the MIR). By choosing the semiconductor, i.e., m and the doping level n_0 , is it then possible to tune the dielectric function at will [45]. For example, one could generate an ϵ -near-zero (ENZ) medium to explore unconventional nonlinear optical effects [26,46]. For all these reasons, the study of nonlinear FE dynamics in heavily doped semiconductors seems to represent a very promising test field toward the understanding of the fundamental mechanisms of nonlinearity in condensed matter and, therefore, in the direction of designing innovative optical technologies.

In this article, we focus on the process of third-harmonic generation (THG) from heavily doped semiconductors. We use a hydrodynamic description to study the free carriers dynamics up to third-order terms, usually neglected for noble metals, and consider both cascaded and direct contributions. We first investigate THG efficiency in unpatterned doped semiconductor slabs. Interestingly, we find that cascaded THG due to SHG signals can be as strong as direct THG contributions even when the SHG efficiency is zero, i.e., there is no SHG in the far field. We also suggest a possible experiment to demonstrate such an effect. We then analyze the conditions that allow for the maximum THG efficiency and show that

FE nonlinearities can be up to two orders of magnitude larger than intrinsic lattice nonlinearities. We finally introduce a grating pattern on the semiconductor slab surface that allows to strongly couple with normally incident waves and excite local plasmons to further enhance the conversion efficiency.

II. HYDRODYNAMIC MODEL FOR THG

To model the many-body dynamics of interest, a theoretical framework that describes not only the charge-density distribution and the nonlinear effects, but also nonlocal corrections, is needed. Indeed, because of the fermionic nature of the carriers, FE nonlinearities are intrinsically nonlocal in the sense that the induced currents depend not only on the value of the electric field at a given point but also, through their spatial derivatives, on the value of the fields in the surrounding area. In addition, the formalism should give the possibility to compare FE nonlinearities to those arising from the crystal lattice in order to solve a possible ambiguity in the origin of the nonlinear response. An appropriate representation of nonlinear and nonlocal FE dynamics in heavily doped semiconductors may be performed taking into consideration the quasiclassical formalism of the hydrodynamic model in the limit of Thomas-Fermi approximation [43,47,48]. The hydrodynamic theory has been successfully used to describe the FE dynamics in noble-metal systems and is then extremely relevant to describe the behavior of heavily doped semiconductors. Furthermore, lattice nonlinearities can be easily included into the model by taking into consideration an intrinsic nonlinear susceptibility as shown in the following.

Within the hydrodynamic formalism [49–51], the many-body dynamics of a FE fluid can be described taking into account two macroscopic variables, the charge-density $n(\mathbf{r}, t)$, and the current density $\mathbf{J}(\mathbf{r}, t) = -en\mathbf{v}$, where $\mathbf{v}(\mathbf{r}, t)$ is the electron velocity field. Under the influence of external electric- and magnetic-fields $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{H}(\mathbf{r}, t)$, the system can be described by the following equation:

$$m \left(\frac{\partial}{\partial t} - \frac{\mathbf{J}}{en} \cdot \nabla + \gamma \right) \frac{\mathbf{J}}{en} = +e\mathbf{E} - \frac{\mathbf{J}}{n} \times \mu_0 \mathbf{H} + \nabla \frac{\delta T^{\text{TF}}[n]}{\delta n}, \quad (1)$$

where μ_0 is the magnetic permeability of vacuum and γ is the damping rate. The last term takes into account the nonlocal effects due to the quantum pressure. Here, $T^{\text{TF}}[n]$ is the kinetic-energy functional in the Thomas-Fermi approximation, whose derivative with respect to n is $\frac{\delta T^{\text{TF}}[n]}{\delta n} = c_{\text{TF}} \frac{5}{3} n^{2/3}$ with $c_{\text{TF}} = \frac{\hbar^2}{m} \frac{3}{10} (3\pi^2)^{2/3}$ [52].

Following a perturbative approach, it is possible to write $n(\mathbf{r}, t) = n_0 + n_{\text{ind}}(\mathbf{r}, t)$, where $n_{\text{ind}} = \frac{1}{e} \nabla \cdot \mathbf{P}$ is the induced charge density and $n_{\text{ind}} \ll n_0$. Here $\mathbf{P}(\mathbf{r}, t)$ is the polarization vector, which is related to the current density through its time derivative, i.e., $\dot{\mathbf{P}} = \mathbf{J}$. Note that we assume n_0 to be constant inside the semiconductor and abruptly go to zero outside. This is not necessarily true at the semiconductor surface where small variations of the density might occur. We assume, however, that these effects can be neglected. We also neglect electron spill out and consider hard-wall boundary conditions

[51]. Equation (1) then becomes

$$\ddot{\mathbf{P}} + \gamma \dot{\mathbf{P}} = \frac{n_0 e^2}{m} \mathbf{E} + \beta^2 \nabla (\nabla \cdot \mathbf{P}) + \mathbf{S}_{\text{NL}}^{(2)} + \mathbf{S}_{\text{NL}}^{(3)}, \quad (2)$$

where time derivatives are now expressed in dot notation. This equation is essentially constituted by the FE gas linear model, i.e., the Drude model plus a nonlocal linear term, where $\beta^2 = \frac{10}{9} \frac{c_{\text{FE}}}{m} n_0^{2/3}$, and the second- and third-order nonlinear sources $\mathbf{S}_{\text{NL}}^{(2)}$ and $\mathbf{S}_{\text{NL}}^{(3)}$, respectively, whose expressions are as follows:

$$\begin{aligned} \mathbf{S}_{\text{NL}}^{(2)} = & \frac{e}{m} \mathbf{E} \nabla \cdot \mathbf{P} - \frac{e \mu_0}{m} \dot{\mathbf{P}} \times \mathbf{H} + \frac{1}{e n_0} (\dot{\mathbf{P}} \nabla \cdot \dot{\mathbf{P}} + \dot{\mathbf{P}} \cdot \nabla \dot{\mathbf{P}}) \\ & + \frac{1}{3} \frac{\beta^2}{e n_0} \nabla (\nabla \cdot \mathbf{P})^2, \end{aligned} \quad (3a)$$

$$\begin{aligned} \mathbf{S}_{\text{NL}}^{(3)} = & -\frac{1}{e^2 n_0^2} (\nabla \cdot \mathbf{P} (\dot{\mathbf{P}} \nabla \cdot \dot{\mathbf{P}} + \dot{\mathbf{P}} \cdot \nabla \dot{\mathbf{P}}) + \dot{\mathbf{P}} \cdot \dot{\mathbf{P}} \nabla \nabla \cdot \mathbf{P}) \\ & - \frac{1}{27} \frac{\beta^2}{e^2 n_0^2} \nabla (\nabla \cdot \mathbf{P})^3. \end{aligned} \quad (3b)$$

In deriving these equations, contributions up to the third order have been considered. In particular, we used $n^{-1} \simeq n_0^{-1} (1 - \frac{n_{\text{ind}}}{n_0})$ and $n^{2/3} \simeq n_0^{2/3} [1 + \frac{2}{3} \frac{n_{\text{ind}}}{n_0} - \frac{1}{9} (\frac{n_{\text{ind}}}{n_0})^2]$. As it can be noted, the nonlinear sources include both local and nonlocal terms. In particular, second-order terms include Coulomb, Lorentz, convective, and quantum pressure contributions in that order in Eq. (3a), whereas third-order terms in Eq. (3b) only include convective and pressure contributions. Among all these terms, those proportional to $\nabla \cdot \mathbf{P}$, i.e., to the induced charge density, are surface contributions since they enable to describe nonlinear effects originating from a region very close to the surface of a certain material, where n_{ind} varies abruptly [36,53].

In the most general picture, for a single input field, THG is the process in which a signal of frequency 3ω is generated by the interaction of a field of frequency ω with a nonlinear material. There are many ways, ideally infinite, to generate an output at 3ω . In the most common process, three photons of energy $\hbar\omega$ combine to give a single photon of energy $3\hbar\omega$. In this case, the THG is usually referred to as direct THG. Additionally the THG process can be seen as a two-step process: first, a photon of energy $2\hbar\omega$ is generated by the interaction of two photons of energy $\hbar\omega$; then, this combines to another photon of energy $\hbar\omega$ in order to get a photon of energy $3\hbar\omega = 2\hbar\omega + \hbar\omega$. The THG is then the combination of two second-order nonlinear processes, namely, SHG followed by sum-frequency generation, and the process is commonly defined as cascaded THG. Being a combination of second-order processes, cascaded generation is forbidden in centrosymmetric bulk crystals, nevertheless, as we will show later, it is allowed in heavily doped semiconductors, similar to metals [53–55] where second-order nonlinearities arise from FE dynamics.

In order to describe the above-mentioned process, let us assume time-harmonic dependence of the fields, i.e., $\mathbf{F}(\mathbf{r}, t) = \sum_j \mathbf{F}_j(\mathbf{r}) e^{-i\omega_j t}$ with $\mathbf{F} = \mathbf{E}$, \mathbf{H} , or \mathbf{P} . Equation (2) with

Eq. (3) and Maxwell's equations can be rewritten as a set of equations for each harmonic ω_j as follows:

$$\nabla \times \nabla \times \mathbf{E}_j - \varepsilon \frac{\omega_j^2}{c^2} \mathbf{E}_j - \omega_j^2 \mu_0 (\mathbf{P}_j + \mathbf{P}_{\omega_j}^{\text{NL}}) = 0, \quad (4a)$$

$$\beta^2 \nabla (\nabla \cdot \mathbf{P}_j) + (\omega_j^2 + i\gamma\omega_j) \mathbf{P}_j = -\frac{n_0 e^2}{m} \mathbf{E}_j + \mathbf{S}_{\omega_j}. \quad (4b)$$

In writing these equations we have considered dielectric local contributions from the semiconductor, both linear, through the local permittivity ε , and nonlinear, through the nonlinear polarization $\mathbf{P}_{\omega_j}^{\text{NL}}$. Coupling between different harmonics occurs through the nonlinear contributions $\mathbf{P}_{\omega_j}^{\text{NL}}$ and \mathbf{S}_{ω_j} . For simplicity, we assume that the pump field is not affected by the nonlinear process (undepleted pump approximation), i.e., $\mathbf{P}_{\omega_1}^{\text{NL}} = \mathbf{S}_{\omega_1} = 0$ as we expect harmonic signals to be several orders of magnitude smaller than the pump fields. The system of Eq. (4) reduces then to three sets of one-way coupled equations, one for the fundamental field (FF) ($j = 1$), one for the second-harmonic (SH) frequency ($j = 2$), and one for the third-harmonic (TH) frequency ($j = 3$).

The term that takes into account crystal lattice nonlinearities can be described using a bulk third-order susceptibility $\chi^{(3)}$ as

$$\mathbf{P}_{\omega_3}^{\text{NL}} = \varepsilon_0 \chi^{(3)} (\mathbf{E}_1 \cdot \mathbf{E}_1) \mathbf{E}_1. \quad (5)$$

In writing this equation we have assumed a centrosymmetric material, which also implies $\chi^{(2)} = 0$. The nonlinear source terms due to free charges are as follows:

$$\begin{aligned} \mathbf{S}_{\omega_2} = & -\frac{e}{m} (\mathbf{E}_1 \nabla \cdot \mathbf{P}_1) - i \frac{e \mu_0}{m} \omega_1 \mathbf{P}_1 \times \mathbf{H}_1 \\ & + \frac{\omega_1^2}{e n_0} (\mathbf{P}_1 \nabla \cdot \mathbf{P}_1 + \mathbf{P}_1 \cdot \nabla \mathbf{P}_1) \\ & - \frac{2}{3} \frac{\beta^2}{e n_0} (\nabla \cdot \mathbf{P}_1 \nabla \nabla \cdot \mathbf{P}_1) \end{aligned} \quad (6)$$

for the SHG and $\mathbf{S}_{\omega_3} = \mathbf{S}_{\omega_3}^{(2)} + \mathbf{S}_{\omega_3}^{(3)}$ for the THG with

$$\begin{aligned} \mathbf{S}_{\omega_3}^{(2)} = & -\frac{e}{m} (\mathbf{E}_2 \nabla \cdot \mathbf{P}_1 + \mathbf{E}_1 \nabla \cdot \mathbf{P}_2) \\ & - i \frac{e \mu_0}{m} (\omega_2 \mathbf{P}_2 \times \mathbf{H}_1 + \omega_1 \mathbf{P}_1 \times \mathbf{H}_2) \\ & + \frac{\omega_1 \omega_2}{e n_0} (\mathbf{P}_2 \nabla \cdot \mathbf{P}_1 + \mathbf{P}_2 \cdot \nabla \mathbf{P}_1 + \mathbf{P}_1 \nabla \cdot \mathbf{P}_2 + \mathbf{P}_1 \cdot \nabla \mathbf{P}_2) \\ & - \frac{2}{3} \frac{\beta^2}{e n_0} (\nabla \cdot \mathbf{P}_2 \nabla \nabla \cdot \mathbf{P}_1 + \nabla \cdot \mathbf{P}_1 \nabla \nabla \cdot \mathbf{P}_2), \end{aligned} \quad (7a)$$

$$\begin{aligned} \mathbf{S}_{\omega_3}^{(3)} = & -\frac{\omega_1^2}{e^2 n_0^2} [\nabla \cdot \mathbf{P}_1 (\mathbf{P}_1 \nabla \cdot \mathbf{P}_1 + \mathbf{P}_1 \cdot \nabla \mathbf{P}_1) \\ & + \mathbf{P}_1 \cdot \mathbf{P}_1 \nabla \nabla \cdot \mathbf{P}_1] + \frac{1}{27} \frac{\beta^2}{e^2 n_0^2} \nabla (\nabla \cdot \mathbf{P}_1)^3 \end{aligned} \quad (7b)$$

describing cascaded and direct THG due to FE dynamics, respectively. The hydrodynamic nonlinear sources in the general case of $\mathbf{S}_{\omega_1} \neq 0$ are reported in the Appendix.

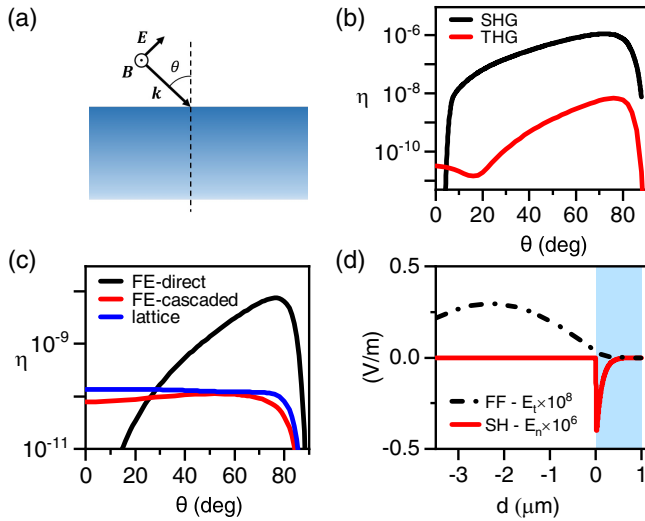


FIG. 1. Hydrodynamic nonlinearities in a highly doped semiconductor slab: (a) a TM plane wave impinging on the slab with an angle of incidence θ is needed to induce a charge oscillation along the finite thickness of the slab; (b) SHG and THG efficiencies η as a function of the incident angle for a InP slab [doping level $n_0 = 5.0 \times 10^{19} \text{ cm}^{-3}$, FF wavelength $\lambda_{\text{FF}} = 10 \mu\text{m}$]; (c) THG efficiencies associated with FE direct to FE cascaded and to lattice nonlinear contributions; (d) nonzero component of the FF and SH fields in the vicinity of the semiconductor surface as a function of the distance d from it. The simulated slab is $1\text{-}\mu\text{m}$ thick.

III. HYDRODYNAMIC NONLINEARITIES IN SEMICONDUCTOR SLABS

A. Cascaded vs direct THG

An initial step toward the understanding of nonlinear conversion in heavily doped semiconductors is to consider a very simple geometry, namely, a slab of a semiconductor. Since a forcing electric field normal to the semiconductor surface is needed to induce a charge oscillation along the finite thickness of the slab, let us consider a transverse magnetic (TM) plane wave impinging on the slab with an angle of incidence θ as depicted in Fig. 1(a).

By solving Eqs. (4)–(7) we are able to calculate SH and TH generated signals. We solved these equations numerically using the finite-element method within a customized frequency-dependent two-dimensional implementation in COMSOL MULTIPHYSICS [56]. Although our model is valid for arbitrary systems, its efficient numerical implementation for three-dimensional systems might require more advanced methods [57]. Our model can be applied to many semiconductors and, in principle, our study may be developed considering a fictitious ideal material. Nevertheless, for the sake of a more realistic investigation, without loss of generality, for our analysis we consider indium phosphide (InP), a direct band-gap III–V semiconductor that has been recently studied as a low loss plasmonic material for the MIR region [58,59] after being introduced among many alternative plasmonic materials few years ago [8]. InP is one of the most common materials in optoelectronic applications thanks to its easy integration and compatibility with the conventional microelectronic foundry and is particularly

interesting for our analysis because of its very small effective electron mass $m = 0.078m_e$ [8], m_e being the electron mass. MIR properties of undoped InP can be described through a constant linear permittivity $\epsilon_\infty = 9.55$ [8] and a nonlinear dispersionless bulk permittivity $\chi^{(3)} = 1.0 \times 10^{-18} \text{ m}^2/\text{V}^2$, a value that corresponds to the highest among most common semiconductors [60]. The damping rate γ is in general dispersive in the MIR [45], although its dispersion is expected not to have effects on the nonlinear generation from the slab. Indeed, here, the important parameter to be considered is the real part of the dielectric function, that does not vary sensibly with γ . Conversely, in the case of a resonant structure, such as the grating described later, γ becomes important because it affects width and amplitude of the resonance and, as a consequence, the FF enhancement. However, what matters in this case is the value of γ around the resonance rather than its dispersion. For all these reasons, we consider in this article a dispersionless $\gamma = 10 \text{ ps}^{-1}$. The pump peak intensity considered is $I_0 = 10 \text{ W}/\mu\text{m}^2$ ($1 \text{ GW}/\text{cm}^2$) in all cases, and it corresponds to an input electric-field $E_0 = 8.7 \times 10^7 \text{ V/m}$, a value that can be obtained with a commercial difference-frequency generation source based on ultrafast laser pulses, focused with a lens objective at the diffraction limit at $\lambda = 8 \mu\text{m}$, and that is also close to the estimated breakdown field of heavily doped InP. The THG efficiency has been obtained by normalizing the power of the generated signal to the input power at the fundamental frequency $\eta = I_G/I_0$, where I_G is the generated intensity. Since we are considering third-order nonlinearities, I_G is expected to be proportional to I_0^3 as a consequence η will scale with I_0^2 . The simulated slab is $1\text{-}\mu\text{m}$ thick.

Figure 1(b) shows SHG and THG efficiencies as a function of the incident angle θ for an InP slab with doping level of $n_0 = 5.0 \times 10^{19} \text{ cm}^{-3}$ excited by a FF oscillating at $\lambda_{\text{FF}} = 10 \mu\text{m}$. As one would expect for a centrosymmetric material, the SHG signal at normal incidence is zero whereas it peaks at larger angles $\theta \simeq 72^\circ$, similar to SHG from noble metals [61,62]. The THG efficiency, on the other hand, gives a nonzero contribution also at normal incidence, whereas increasing for large angles, peaking at $\theta \simeq 77^\circ$. As can be inferred by Eq. (7b), direct FE contributions are zero at normal incidence since $\nabla \cdot \mathbf{P}_1$ is identically zero for $\theta = 0$. Hence, it would be logical to think that the contributions shown in Fig. 1(b) are due to the lattice nonlinearities since also the SHG is zero. In order to clarify this point, in Fig. 1(c) we plot the THG efficiencies associated with each contribution. Surprisingly the FE cascaded THG signal is not zero at normal incidence, moreover, it is comparable to the lattice $\chi^{(3)}$ contribution. How is it possible to obtain a cascaded effect from a signal that is apparently zero? To understand this, we plot in Fig. 1(d) the nonzero component of the FF and SH field in the vicinity of the semiconductor surface. It can be observed indeed that the SH field has a nonzero longitudinal component that follows the profile of the square FF amplitude. This component is induced by the Lorentz term in Eq. (6) inside the semiconductor. However, because the vacuum does not support longitudinal modes, such a component cannot be coupled to the far field. On the other hand, a SH longitudinal component in the semiconductor combines back with the

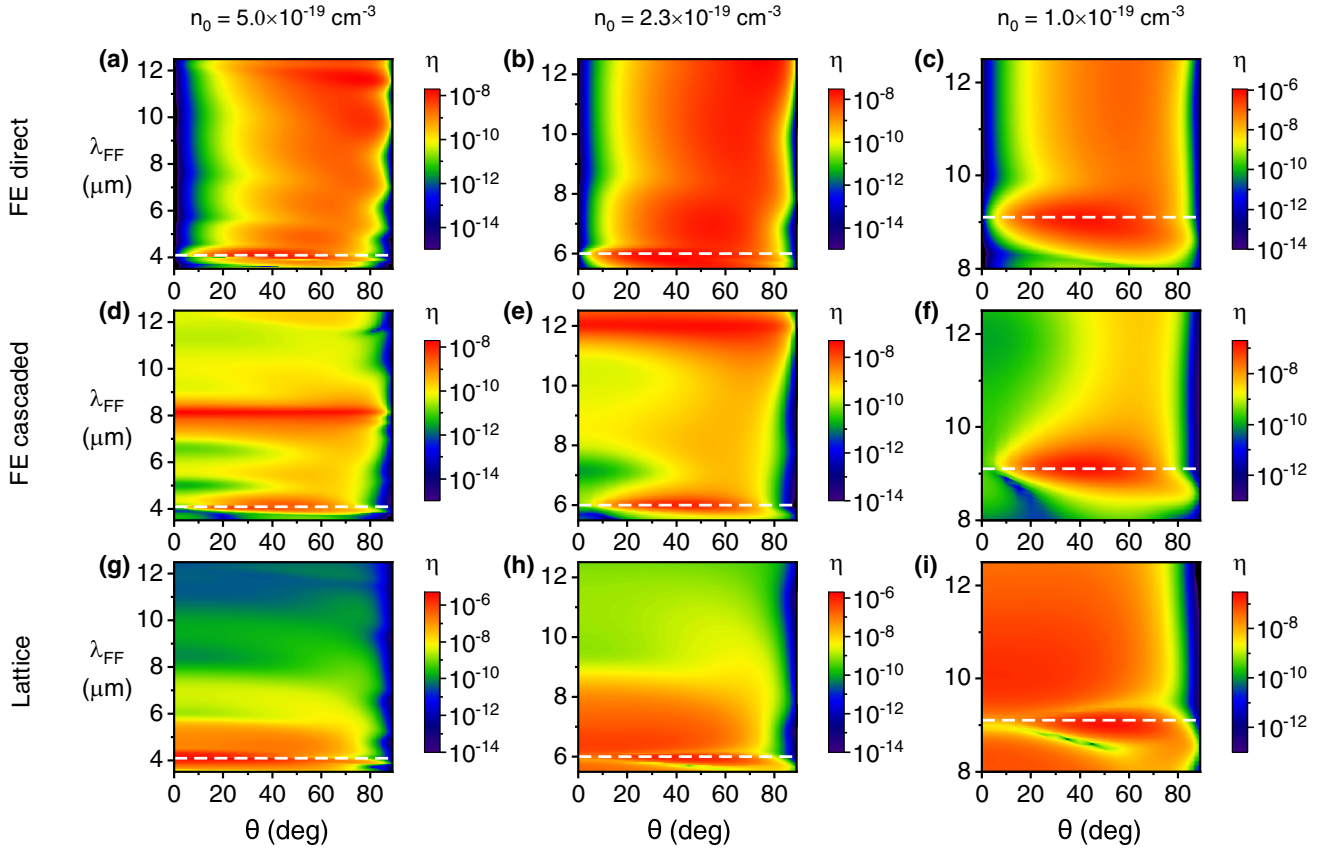


FIG. 2. THG efficiency η of a highly doped InP slab as a function of the incident angle θ and of the wavelength λ_{FF} of the FF, calculated taking into account distinct nonlinear sources: (a)–(c) are obtained considering only FE direct contributions; (d)–(f) refer to the FE cascaded THG; (g)–(i) to the THG originating from background lattice third-order susceptibility $\chi^{(3)}$. In each column a different value (highlighted with a dotted white line) of the screened plasma wavelength $\tilde{\lambda}_p$ (i.e., a different doping level) is considered. In (a), (d), and (g) $\tilde{\lambda}_p \simeq 4$ ($n_0 = 5.0 \times 10^{19} \text{ cm}^{-3}$); in (b), (e), and (h) $\tilde{\lambda}_p \simeq 6 \mu\text{m}$ ($n_0 = 2.3 \times 10^{19} \text{ cm}^{-3}$); and in (c), (f), and (i) $\tilde{\lambda}_p \simeq 9 \mu\text{m}$ ($n_0 = 1.0 \times 10^{19} \text{ cm}^{-3}$).

magnetic field associated with the FF [through the Lorentz term in Eq. (7a)], finally resulting in a transverse electric-field component at the TH. A similar contribution is also given by one of the convective terms. Indeed, a THG cascaded contribution can be generated through an apparently zero SHG.

Although this effect can be easily observed with numerical simulations, it is not that simple to differentiate between cascaded and direct contributions in an experimental setup. Let us anticipate, however, that it is indeed possible to take advantage from this process and experimentally demonstrate this peculiar phenomenon by properly tuning the frequency with respect to the natural ENZ frequency of doped semiconductors as we will show below.

B. Hydrodynamic vs crystalline lattice contributions

Having clarified the nature of cascaded effects, let us now focus on spectral properties of the THG process. For a given material characterized by an effective-mass m and a fixed doping level n_0 , the nature of its optical response depends on the frequency of the external electromagnetic field. That is, the doped semiconductor has a metallic behavior for frequencies below the plasma frequency, whereas it behaves, such as a dielectric for higher frequencies.

In Fig. 2 each map plot reports the simulated reflected THG efficiency η , of a slab of doped semiconductor as function of the FF incident angle θ , and wavelength λ_{FF} , for different conditions. Specifically, each row refers to the THG efficiency obtained considering a distinct nonlinear source: $\mathbf{S}_{\omega_3}^{(3)}$ (FE direct), $\mathbf{S}_{\omega_3}^{(2)}$ (FE cascaded), and $\omega_1^2 \mu_0 \mathbf{P}_{\omega_3}^{\text{NL}}$ (lattice) respectively. Each column refers to a different level of doping, and, as a consequence, to a different screened plasma frequency $\tilde{\omega}_p \simeq \omega_p / \sqrt{\epsilon_\infty}$.

As discussed in the Introduction, a higher n_0 is predicted to correspond to smaller hydrodynamic effects and, thus, to a smaller THG efficiency for a given ratio $\frac{\omega_{\text{FF}}}{\tilde{\omega}_p}$. For instance, if one considers the FE direct contribution, when $\lambda_{\text{FF}} = \tilde{\lambda}_p = \frac{2\pi c}{\tilde{\omega}_p}$, η goes from a peak of the order of 10^{-8} in Fig. 2(a) to a maximum of 10^{-6} in Fig. 2(c), i.e., the efficiency is between 10 and 100 times bigger for a doping five times smaller. Instead, for a given level of doping, an increase in the efficiency is expected when there are the most favorable energetic condition for the enhancement of the fields involved in the process. Peaks of the efficiency are, hence, predicted for $\lambda_{\text{FF}} = \tilde{\lambda}_p$ and $\lambda_{\text{FF}} = 3\tilde{\lambda}_p$ in all cases, namely, when the FF and the generated field are at the ENZ condition, respectively. These features can be clearly seen in all plots of Fig. 2 when the FF is at the ENZ condition. However,

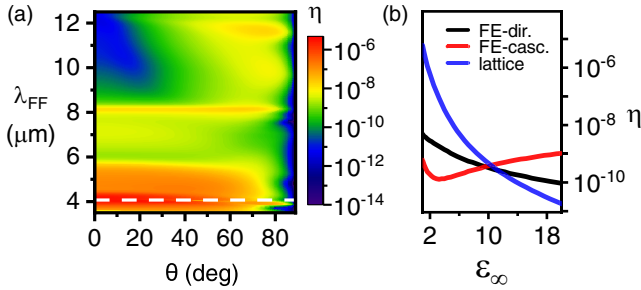


FIG. 3. (a) THG efficiency η of a highly doped InP slab as a function of the incident angle θ and of the wavelength λ_{FF} of the FF, calculated for $\tilde{\lambda}_p \simeq 4$ ($n_0 = 5.0 \times 10^{19} \text{ cm}^{-3}$) taking into account all the nonlinear sources; (b) THG efficiency as function of the background permittivity for the three distinct nonlinear contributions with $n_0 = 5.0 \times 10^{19} \text{ cm}^{-3}$, $m = 0.1m_e$, $\lambda_{FF} = 1.8\tilde{\lambda}_p$, and $\theta = 40^\circ$.

when the generated signal is at the ENZ condition (only possible for the first column for $\lambda_{FF} \simeq 12 \mu\text{m}$), there is a sharp peak only for the FE direct THG [Fig. 2(a)]. In the case of the cascaded process, there is not a clear maximum of η , probably because an enhancement of the generated field is not so effective for the overall efficiency if the other fields involved in the process are not enhanced as well. The same happens for the lattice nonlinearities because of the limited interaction of the input field with the slab in the metallic regime of the material since these contributions originate in the bulk.

A maximum of the efficiency for $\lambda_{FF} = 2\tilde{\lambda}_p$ is by contrast a unique feature expected only for the cascaded THG. It corresponds to the enhancement of the field generated through the SHG when it is at the ENZ condition. The aforementioned peculiarity represents a crucial clue for the experimental demonstration of cascade effects in heavily doped semiconductors, and can be clearly seen in Figs. 2(d) and 2(e) for $\lambda_{FF} \simeq 8$ and $\lambda_{FF} \simeq 12 \mu\text{m}$, respectively.

A very important characteristic that emerges from Fig. 2 is the order of magnitude of the FE THG compared to that arising from lattice nonlinearities. Indeed, although we used a very high value of $\chi^{(3)}$, third-order hydrodynamic contributions can be the most important in the portion of the MIR we considered. This could be expected in the metallic regime of the material far from the screened plasma wavelength where the interaction of the pump with the bulk crystal lattice is hindered. However, for the smaller level of doping the efficiency of THG is higher for the hydrodynamic direct process than for the lattice even at ENZ. This represents a remarkable prediction toward the application of heavily doped semiconductor for nonlinear plasmonics.

C. What to expect from experiments

At this point of the discussion, it is interesting to wonder what kind of results one should expect from an experiment. Let us now consider all the nonlinear sources at once to show what is expected in a more realistic setup. The results are reported in Fig. 3(a) in the case of $\tilde{\lambda}_p \approx 4 \mu\text{m}$ ($n_0 = 5.0 \times 10^{19} \text{ cm}^{-3}$). Unsurprisingly, in Fig. 3(a) one can see all the main characteristics of the three nonlinear sources, which

appear separately in the first column of Fig. 2, combined in a single plot. If our model is accurate enough in describing nonlinearities in heavily doped semiconductors, an interesting way to demonstrate our predictions, taking into consideration Fig. 3(a), seems to look for a very peculiar trend of the THG efficiency at wavelengths close to $2\tilde{\lambda}_p$ and to $3\tilde{\lambda}_p$. Comparing Figs. 3(a) and 2(d), it is clear that the peak at $2\tilde{\lambda}_p$ can only be caused by a cascaded generation. A similar reasoning seems to make sense for the peak at $3\tilde{\lambda}_p$ when considering the FE direct THG.

D. The influence of the semiconductor background permittivity

Hitherto, we considered a well-defined material, however, it is interesting to study how the hydrodynamic nonlinear sources are influenced by the material physical properties other than the doping level n_0 , even if the actual physical parameter cannot be varied in the experiments as is the case for the background permittivity ϵ_∞ . In Fig. 3(b) we show the THG efficiency η as function of the ϵ_∞ for the three THG contributions with all other variables unchanged. The trend of hydrodynamic THG efficiencies in Fig. 3(b) might be intuitively understood thinking to the interaction of light with the slab from the point of view of Snell's law. Indeed, the refractive index increases with the square root of ϵ_∞ . This also defines the angle of refraction of a light inside a material (the higher the refractive index, the smaller the angle of refraction). Therefore, for a given angle of incidence, a higher ϵ_∞ causes the transmitted FF at the interface air/semiconductor to be closer to normal incidence when propagating inside the slab. However, we know from the previous paragraphs that surface contributions are hindered at normal incidence, so, the FE direct THG efficiency due only to surface contributions [see Eq. (7b)] becomes smaller when increasing ϵ_∞ . Conversely, the Lorentz term is maximum at normal incidence, and it is then favored for higher ϵ_∞ . As a consequence, taking into account Eq. (7a), the cascaded THG efficiency has a very peculiar trend caused by a balance between surface and Lorentz contributions. Indeed, in Fig. 3(b), the FE cascaded contribution curve has the same decreasing trend as the FE direct contribution one for very small values of ϵ_∞ , i.e., far from normal refraction where surface contributions are dominant with respect to the Lorentz term. It reaches a minimum where the contributions considered are of the same order and finally grows steadily with ϵ_∞ when the Lorentz contribution becomes the most relevant (closer to normal refraction). Instead, for background lattice nonlinearities, η is higher for lower ϵ_∞ simply because of the higher transparency of the material. As a consequence, one should expect relatively higher hydrodynamic nonlinear contributions when considering a material with a lower background permittivity, such as the conductive indium tin oxide [48].

IV. A RESONANT PATTERN

The analysis developed so far was useful to understand some fundamental properties of hydrodynamic nonlinearities. From the point of view of technology, however, it is perhaps more interesting to apply our model to a nanopatterned semiconductor slab with a geometrical resonance in

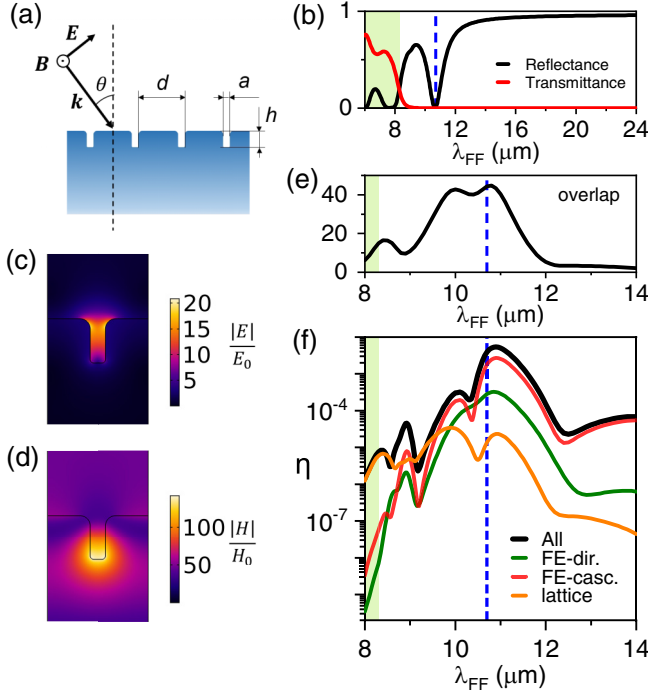


FIG. 4. Optical properties of a InP grooves array for THG: (a) schematics of the geometry; (b) linear spectra of the structure at normal incidence when $n_0 = 1.2 \times 10^{19} \text{ cm}^{-3}$ ($\tilde{\lambda}_p = 8.3 \mu\text{m}$), $d = 900$, $a = 150$, and $h = 400 \text{ nm}$, the thickness of the slab on which the grooves are carved is $H = 3 \mu\text{m}$; (c) and (d) electric- and magnetic-field enhancements in correspondence of the resonance; (e) overlap integral as a function of the FF wavelength λ_{FF} ; (f) THG efficiency as a function of λ_{FF} in the proximity of the resonance for the three distinct nonlinear contributions, FE direct, FE cascaded, lattice, and for their combination (all).

the MIR. To this aim, we take now into consideration an infinite array of parallel subwavelength grooves (a grating) as depicted in Fig. 4(a). This system supports plasmonic resonances and allows to couple virtually all incident energy into the active material [63–65] at normal incidence. In Fig. 4(b) we report the simulated linear spectrum of the pattern we designed in order to have a resonance in the MIR for a TM-polarized excitation. Once the material properties have been fixed, the electromagnetic characteristics of the pattern can be designed as function of the array period d and of the dimension of the grooves, the aperture a , and the height h [see Fig. 4(a)]. The material considered is again InP and its doping level is $n_0 = 1.2 \times 10^{19} \text{ cm}^{-3}$ ($\tilde{\lambda}_p = 8.3 \mu\text{m}$). The pattern produces a resonance around $\lambda_{\text{res}} = 10.6 \mu\text{m}$ where the reflectance is almost zero [see Fig. 4(b)]. In Figs. 4(c) and 4(d) we report the electric- and magnetic-field enhancements in correspondence of the resonance. The fields are confined and enhanced into the grooves. This means that approximately all the pump energy impinging on the structure is coupled inside the material and localized in a much smaller volume.

In Fig. 4(f), we report the reflected THG efficiency η at normal incidence of the grooves array as function of the wavelength of the FF λ_{FF} , considering the same processes analyzed for the slab: FE direct, FE cascaded, lattice, and

all nonlinear sources combined. The pump peak intensity of the TM plane wave impinging on the structure is the same as before in the case of the slab, $I_0 = 10 \text{ W}/\mu\text{m}^2$ ($1 \text{ GW}/\text{cm}^2$). A very important feature is the possibility to get FE direct THG even at normal incidence thanks to the plasmonic behavior. As it can be noted, all the curves as expected reach a peak approaching the resonance. Their trend is influenced by the overlap between the fundamental and the generated fields, estimated calculating the overlap integral for the bulk contribution $\int \chi^{(3)} |\mathbf{E}_\omega|^2 \mathbf{E}_\omega \cdot \mathbf{E}_{3\omega}^* dV$ and reported in Fig. 4(e) (note that the exact overlap integral should take into account the specific expressions for all nonlinear sources). The peak efficiency of FE contributions is several orders of magnitude higher than in the case of the slab, and it always surpasses that of intrinsic lattice nonlinearities. There, the maximum value was around 10^{-6} for the direct process and smaller for the cascaded, instead, here we get a value larger than 10^{-4} for the first process and bigger than 10^{-3} for the cascaded. Consequently, for the grooves array the efficiency is, at least, 10^2 times larger for the FE direct THG and more than 10^4 times higher for the cascaded THG with the same pump intensity; whereas the THG due to the lattice increases just by a factor of 10. Therefore, although we have considered a material with very high third-order nonlinear susceptibility $\chi^{(3)}$, we predict that FE contributions will overcome by far conventional semiconductor nonlinearities in THG processes in the MIR.

V. CONCLUSIONS AND PERSPECTIVES

We have used a hydrodynamic model to describe FE nonlinear optical dynamics in heavily doped semiconductors. In particular, by deriving nonlinear hydrodynamic terms up to the third order, we have investigated FE contributions to the THG process. Contrary to noble metals at optical frequencies, the heavily doped semiconductor TH response is predicted to be strongly driven by hydrodynamic nonlinearities. We have found, in fact, that such nonlinearities might generate THG signals up to two orders of magnitude larger than conventional semiconductor nonlinearities when combined with plasmonic field enhancement. Moreover, we have showed that cascaded contributions, which are very often neglected, might be extremely important, even when one of the constituent signals appear to be negligible. Although in most realistic experiments it would be impossible to differentiate cascaded effects from conventional nonlinear signals, we have proposed a simple method to experimentally verify our findings. Field-effect gated devices and optical generation of free carriers both characteristics of semiconductors could provide a route for tunable nonlinear optical systems on semiconductor-based photonic integrated circuit platforms. We believe that these results could open new avenues for integrated nonlinear optics at MIR frequencies and beyond.

APPENDIX: HYDRODYNAMIC NONLINEAR SOURCES

In Sec. II, we derived the system of equations Eq. (4) and, assuming $\mathbf{S}_{\omega_1} = 0$, we reported in Eqs. (6) and (7) only the nonlinear sources \mathbf{S}_{ω_2} and \mathbf{S}_{ω_3} . However, in the most

general case $\mathbf{S}_{\omega_1} \neq 0$. It means that, for high input intensities, the nonlinear field produced may be converted back to the

FF. Taking into account Eq. (3) this may happen through difference-frequency generation of the second order,

$$\begin{aligned} \mathbf{S}_{\omega_1=\omega_2-\omega_1}^{(2)} = & -\frac{e}{m}(\mathbf{E}_1^* \nabla \cdot \mathbf{P}_2 + \mathbf{E}_2 \nabla \cdot \mathbf{P}_1^*) - i\frac{e\mu_0}{m}(\omega_2 \mathbf{P}_2 \times \mathbf{H}_1^* - \omega_1 \mathbf{P}_1^* \times \mathbf{H}_2) \\ & + \frac{\omega_1 \omega_2}{en_0}(\mathbf{P}_2 \nabla \cdot \mathbf{P}_1^* + \mathbf{P}_2 \cdot \nabla \mathbf{P}_1^* + \mathbf{P}_1^* \nabla \cdot \mathbf{P}_2 + \mathbf{P}_1^* \cdot \nabla \mathbf{P}_2) \\ & - \frac{2}{3} \frac{\beta^2}{en_0}(\nabla \cdot \mathbf{P}_2 \nabla \nabla \cdot \mathbf{P}_1^* + \nabla \cdot \mathbf{P}_1^* \nabla \nabla \cdot \mathbf{P}_2), \end{aligned} \quad (\text{A1})$$

through difference-frequency generation of the third order,

$$\begin{aligned} \mathbf{S}_{\omega_1=\omega_3-\omega_1-\omega_1}^{(3)} = & -\frac{\omega_1^2}{e^2 n_0^2}[(\nabla \cdot \mathbf{P}_3)(\mathbf{P}_1^* \nabla \cdot \mathbf{P}_1^* + \mathbf{P}_1^* \cdot \nabla \mathbf{P}_1^*) + \mathbf{P}_1^* \cdot \mathbf{P}_1^* \nabla (\nabla \cdot \mathbf{P}_3)] \\ & + \frac{\omega_1 \omega_3}{e^2 n_0^2}[(\nabla \cdot \mathbf{P}_1^*)(\mathbf{P}_3 \nabla \cdot \mathbf{P}_1^* + \mathbf{P}_3 \cdot \nabla \mathbf{P}_1^* + \mathbf{P}_1^* \nabla \cdot \mathbf{P}_3 + \mathbf{P}_1^* \cdot \nabla \mathbf{P}_3) + 2\mathbf{P}_3 \cdot \mathbf{P}_1^* \nabla (\nabla \cdot \mathbf{P}_1^*)] \\ & + \frac{1}{9} \frac{\beta^2}{e^2 n_0^2} \nabla [(\nabla \cdot \mathbf{P}_1^*)^2 (\nabla \cdot \mathbf{P}_3)], \end{aligned} \quad (\text{A2})$$

eventually, third-order effects may also occur as Kerr-type nonlinearities,

$$\begin{aligned} \mathbf{S}_{\omega_1=\omega_1-\omega_1+\omega_1}^{(3)} = & -\frac{\omega_1^2}{e^2 n_0^2}[(\nabla \cdot \mathbf{P}_1^*)(\mathbf{P}_1 \cdot \nabla \mathbf{P}_1) + \mathbf{P}_1 \cdot \mathbf{P}_1 \nabla (\nabla \cdot \mathbf{P}_1^*)] \\ & + \frac{\omega_1^2}{e^2 n_0^2}[(\nabla \cdot \mathbf{P}_1)(\mathbf{P}_1 \cdot \nabla \mathbf{P}_1^* + \mathbf{P}_1^* \nabla \cdot \mathbf{P}_1 + \mathbf{P}_1^* \cdot \nabla \mathbf{P}_1) + 2|\mathbf{P}_1|^2 \nabla (\nabla \cdot \mathbf{P}_1)] \\ & + \frac{1}{9} \frac{\beta^2}{e^2 n_0^2} \nabla [|\nabla \cdot \mathbf{P}_1|^2 (\nabla \cdot \mathbf{P}_1)], \end{aligned} \quad (\text{A3})$$

where all the variables and the constants have been introduced in Secs. I and II. As described in Sec. II, in deriving

these equations we assumed time-harmonic dependence of the fields.

-
- [1] J. J. Baumberg, J. Aizpurua, M. H. Mikkelsen, and D. R. Smith, Extreme nanophotonics from ultrathin metallic gaps, *Nature Mater.* **18**, 668 (2019).
 - [2] O. S. Ojambati, R. Chikkaraddy, W. D. Deacon, M. Horton, D. Kos, V. A. Turek, U. F. Keyser, and J. J. Baumberg, Quantum electrodynamics at room temperature coupling a single vibrating molecule with a plasmonic nanocavity, *Nat. Commun.* **10**, 1049 (2019).
 - [3] J. Huang, A. J. Traverso, G. Yang, and M. H. Mikkelsen, Real-time tunable strong coupling: From individual nanocavities to metasurfaces, *ACS Photonics* **6**, 838 (2019).
 - [4] W. Chen, S. Zhang, M. Kang, W. Liu, Z. Ou, Y. Li, Y. Zhang, Z. Guan, and H. Xu, Probing the limits of plasmonic enhancement using a two-dimensional atomic crystal probe, *Light: Sci. Appl.* **7**, 56 (2018).
 - [5] A. I. Fernández-Domínguez, S. I. Bozhevolnyi, and N. A. Mortensen, Plasmon-enhanced generation of nonclassical light, *ACS Photonics* **5**, 3447 (2018).
 - [6] A. F. Koenderink, Single-photon nanoantennas, *ACS Photonics* **4**, 710 (2017).
 - [7] A. Boltasseva and H. A. Atwater, Low-loss plasmonic metamaterials, *Science* **331**, 290 (2011).
 - [8] G. V. Naik, V. M. Shalae, and A. Boltasseva, Alternative plasmonic materials: Beyond gold and silver, *Adv. Mater.* **25**, 3264 (2013).
 - [9] T. Taliercio and P. Biagioni, Semiconductor infrared plasmonics, *Nanophotonics* **8**, 949 (2019).
 - [10] Y. Su, W. Wang, X. Hu, H. Hu, X. Huang, Y. Wang, J. Si, X. Xie, B. Han, H. Feng, Q. Hao, G. Zhu, T. Duan, and W. Zhao, 10 Gbps DPSK transmission over free-space link in the mid-infrared, *Opt. Express* **26**, 34515 (2018).
 - [11] K. Koshelev, S. Kruk, E. Melik-Gaykazyan, J.-H. Choi, A. Bogdanov, H.-G. Park, and Y. Kivshar, Subwavelength dielectric resonators for nonlinear nanophotonics, *Science* **367**, 288 (2020).
 - [12] V. F. Gili, L. Carletti, A. Locatelli, D. Rocco, M. Finazzi, L. Ghirardini, I. Favero, C. Gomez, A. Lemaître, M. Celebrano, C. De Angelis, and G. Leo, Monolithic AlGaAs second-harmonic nanoantennas, *Opt. Express* **24**, 15965 (2016).
 - [13] L. Carletti, G. Marino, L. Ghirardini, V. F. Gili, D. Rocco, I. Favero, A. Locatelli, A. V. Zayats, M. Celebrano, M. Finazzi, G. Leo, C. de Angelis, and D. N. Neshev, Nonlinear goniometry by second-harmonic generation in AlGaAs nanoantennas, *ACS Photonics* **5**, 4386 (2018).

- [14] Q. Shen, W. Jin, G. Yang, A. W. Rodriguez, and M. H. Mikkelsen, Active control of multiple, simultaneous nonlinear optical processes in plasmonic nanogap cavities, *ACS Photonics* **7**, 901 (2020).
- [15] R. Sarma, D. de Ceglia, N. Nookala, M. A. Vincenti, S. Campione, O. Wolf, M. Scalora, M. B. Sinclair, M. A. Belkin, and I. Brener, Broadband and efficient second-harmonic generation from a hybrid dielectric metasurface/semiconductor quantum-well structure, *ACS Photonics* **6**, 1458 (2019).
- [16] Y. Zeng, H. Qian, M. J. Rozin, Z. Liu, and A. R. Tao, Enhanced second harmonic generation in double-resonance colloidal metasurfaces, *Adv. Funct. Mater.* **28**, 1803019 (2018).
- [17] A. R. Echarri, J. D. Cox, R. Yu, and F. J. G. de Abajo, Enhancement of nonlinear optical phenomena by localized resonances, *ACS Photonics* **5**, 1521 (2018).
- [18] Q. Shen, T. B. Hoang, G. Yang, V. D. Wheeler, and M. H. Mikkelsen, Probing the origin of highly-efficient third-harmonic generation in plasmonic nanogaps, *Opt. Express* **26**, 20718 (2018).
- [19] S. Guddala and S. A. Ramakrishna, Optical limiting by nonlinear tuning of resonance in metamaterial absorbers, *Opt. Lett.* **41**, 5150 (2016).
- [20] T. Shibamura, G. Grinblat, P. Albella, and S. A. Maier, Efficient third harmonic generation from metal – Dielectric hybrid nanoantennas, *Nano Lett.* **17**, 2647 (2017).
- [21] J. Lee, M. Tymchenko, C. Argyropoulos, P.-Y. Chen, F. Lu, F. Demmerle, G. Boehm, M.-C. Amann, A. Alù, and M. A. Belkin, Giant nonlinear response from plasmonic metasurfaces coupled to intersubband transitions, *Nature (London)* **511**, 65 (2014).
- [22] F. Wang, A. B. F. Martinson, and H. Harutyunyan, Efficient nonlinear metasurface based on nonplanar plasmonic nanocavities, *ACS Photonics* **4**, 1188 (2017).
- [23] A. Noor, A. R. Damodaran, I.-H. Lee, S. A. Maier, S.-H. Oh, and C. Ciraci, Mode-matching enhancement of second-harmonic generation with plasmonic nanopatch antennas, *ACS Photonics* **7**, 3333 (2020).
- [24] M. Hentschel, B. Metzger, B. Knabe, K. Buse, and H. Giessen, Linear and nonlinear optical properties of hybrid metallic–dielectric plasmonic nanoantennas, *Beilstein J. Nanotechnol.* **7**, 111 (2016).
- [25] E. Barakat, M.-P. Bernal, and F. I. Baida, Theoretical analysis of enhanced nonlinear conversion from metallo-dielectric nanostructures, *Opt. Express* **20**, 16258 (2012).
- [26] J. Deng, Y. Tang, S. Chen, K. Li, A. V. Zayats, and G. Li, Giant enhancement of second-order nonlinearity of epsilon-near-zero medium by a plasmonic metasurface, *Nano Lett.* **20**, 5421 (2020).
- [27] M. P. Nielsen, X. Shi, P. Dichtl, S. A. Maier, and R. F. Oulton, Giant nonlinear response at a plasmonic nanofocus drives efficient four-wave mixing, *Science* **358**, 1179 (2017).
- [28] H. Harutyunyan, G. Volpe, R. Quidant, and L. Novotny, Enhancing the Nonlinear Optical Response Using Multifrequency Gold-Nanowire Antennas, *Phys. Rev. Lett.* **108**, 217403 (2012).
- [29] K. D. Ko, A. Kumar, K. H. Fung, R. Ambekar, G. L. Liu, N. X. Fang, and K. C. Toussaint, Nonlinear optical response from arrays of Au bowtie nanoantennas, *Nano Lett.* **11**, 61 (2011).
- [30] R. Czaplicki, A. Kiviniemi, M. J. Huttunen, X. Zang, T. Stolt, I. Vartiainen, J. Butet, M. Kuittinen, O. J. F. Martin, and M. Kauranen, Less is more: Enhancement of second-harmonic generation from metasurfaces by reduced nanoparticle density, *Nano Lett.* **18**, 7709 (2018).
- [31] A. V. Krasavin, P. Ginzburg, and A. V. Zayats, Free-electron optical nonlinearities in plasmonic nanostructures: A review of the hydrodynamic description, *Laser Photon. Rev.* **12**, 1700082 (2018).
- [32] A. Chizmeshya and E. Zaremba, Second-harmonic generation at metal surfaces using an extended Thomas Fermi von Weizsacker theory, *Phys. Rev. B* **37**, 2805 (1988).
- [33] N. Crouseilles, P.-A. Hervieux, and G. Manfredi, Quantum hydrodynamic model for the nonlinear electron dynamics in thin metal films, *Phys. Rev. B* **78**, 155412 (2008).
- [34] M. Scalora, M. A. Vincenti, D. de Ceglia, V. Roppo, M. Centini, N. Akozbek, and M. J. Bloemer, Second- and third-harmonic generation in metal-based structures, *Phys. Rev. A* **82**, 043828 (2010).
- [35] Y. Pavlyukh, J. Berakdar, and W. Hubner, Semi – classical approximation for second-harmonic generation in nanoparticles, *New J. Phys.* **14**, 093044 (2012).
- [36] C. Ciraci, E. Poutrina, M. Scalora, and D. R. Smith, Origin of second-harmonic generation enhancement in optical split-ring resonators, *Phys. Rev. B* **85**, 201403(R) (2012).
- [37] A. Capretti, C. Forestiere, L. Dal Negro, and G. Miano, Full-wave analytical solution of second-harmonic generation in metal nanospheres, *Plasmonics* **9**, 151 (2013).
- [38] B. Wells, A. Y. Bykov, G. Marino, M. E. Nasir, A. V. Zayats, and V. A. Podolskiy, Structural second-order nonlinearity in plasmonic metamaterials, *Optica* **5**, 1502 (2018).
- [39] M. W. Klein, M. Wegener, N. Feth, and S. Linden, Experiments on second- and third-harmonic generation from magnetic metamaterials, *Opt. Express* **15**, 5238 (2007).
- [40] R. Czaplicki, H. Husu, R. Siikanen, J. Mäkitalo, M. Kauranen, J. Laukkanen, J. Lehtolahti, and M. Kuittinen, Enhancement of Second-Harmonic Generation from Metal Nanoparticles by Passive Elements, *Phys. Rev. Lett.* **110**, 093902 (2013).
- [41] M. P. Fischer, A. Riede, K. Gallacher, J. Frigerio, G. Pellegrini, M. Ortolani, D. J. Paul, G. Isella, A. Leitenstorfer, P. Biagioni, and D. Brida, Plasmonic mid-infrared third harmonic generation in germanium nanoantennas, *Light: Sci. Appl.* **7**, 106 (2018).
- [42] M. Wagner, A. S. McLeod, S. J. Maddox, Z. Fei, M. Liu, R. D. Averitt, M. M. Fogler, S. R. Bank, F. Keilmann, and D. N. Basov, Ultrafast dynamics of surface plasmons in insulators by time-resolved infrared nanospectroscopy, *Nano Lett.* **14**, 4529 (2014).
- [43] J. R. Maack, N. A. Mortensen, and M. Wubs, Size-dependent nonlocal effects in plasmonic semiconductor particles, *Europhys. Lett.* **119**, 17003 (2017).
- [44] E. J. C. Dias, D. A. Iranzo, P. A. D. Gonçalves, Y. Hajati, Y. V. Bludov, A.-P. Jauho, N. A. Mortensen, F. H. L. Koppens, and N. M. R. Peres, Probing nonlocal effects in metals with graphene plasmons, *Phys. Rev. B* **97**, 245405 (2018).
- [45] J. Frigerio, A. Ballabio, G. Isella, E. Sakat, G. Pellegrini, P. Biagioni, M. Bollani, E. Napolitani, C. Manganello, M. Virgilio, A. Grupp, M. P. Fischer, D. Brida, K. Gallacher, D. J. Paul, L. Baldassarre, P. Calvani, V. Gilierti, A. Nucara, and M. Ortolani, Tunability of the dielectric function of heavily doped germanium thin films for mid-infrared plasmonics, *Phys. Rev. B* **94**, 085202 (2016).

- [46] O. Reshef, I. De Leon, M. Z. Alam, and R. W. Boyd, Nonlinear optical effects in epsilon-near-zero media, *Nat. Rev. Mater.* **4**, 535 (2019).
- [47] M. Scalora, J. Trull, D. de Ceglia, M. A. Vincenti, N. Akozbek, Z. Coppens, L. Rodríguez-Suné, and C. Cojocaru, Electrodynamics of conductive oxides: Intensity-dependent anisotropy, reconstruction of the effective dielectric constant, and harmonic generation, *Phys. Rev. A* **101**, 053828 (2020).
- [48] L. Rodríguez-Suné, M. Scalora, A. S. Johnson, C. Cojocaru, N. Akozbek, Z. J. Coppens, D. Perez-Salinas, S. Wall, and J. Trull, Study of second and third harmonic generation from an indium tin oxide nanolayer: Influence of nonlocal effects and hot electrons, *APL Photonics* **5**, 010801 (2020).
- [49] W. Yan, Hydrodynamic theory for quantum plasmonics: Linear-response dynamics of the inhomogeneous electron gas, *Phys. Rev. B* **91**, 115416 (2015).
- [50] D. de Ceglia, M. Scalora, M. A. Vincenti, S. Campione, K. Kelley, E. L. Runnerstrom, J.-P. Maria, G. A. Keeler, and T. S. Luk, Viscoelastic optical nonlocality of low-loss epsilon-near-zero nanofilms, *Sci. Rep.* **8**, 9335 (2018).
- [51] S. Raza, S. I. Bozhevolnyi, M. Wubs, and N. A. Mortensen, Nonlocal optical response in metallic nanostructures, *J. Phys.: Condens. Matter* **27**, 183204 (2015).
- [52] C. Ciraci and F. Della Sala, Quantum hydrodynamic theory for plasmonics: Impact of the electron density tail, *Phys. Rev. B* **93**, 205405 (2016).
- [53] F. De Luca and C. Ciraci, Difference-frequency generation in plasmonic nanostructures: A parameter-free hydrodynamic description, *J. Opt. Soc. Am. B* **36**, 1979 (2019).
- [54] M. Celebrano, X. Wu, M. Baselli, S. G. mann, P. Biagioni, A. Locatelli, C. de Angelis, G. Cerullo, R. Osellame, B. Hecht, L. Duò, F. Ciccacci, and M. Finazzi, Mode matching in multiresonant plasmonic nanoantennas for enhanced second harmonic generation, *Nat. Nanotechnol.* **10**, 412 (2015).
- [55] M. Celebrano, A. Locatelli, L. Ghirardini, G. Pellegrini, P. Biagioni, A. Zilli, X. Wu, S. Grossmann, L. Carletti, C. De Angelis, L. Duò, B. Hecht, and M. Finazzi, Evidence of cascaded third-harmonic generation in noncentrosymmetric gold nanoantennas, *Nano Lett.* **19**, 7013 (2019).
- [56] COMSOL MULTIPHYSICS, www.comsol.com.
- [57] F. Vidal-Codina, N. C. Nguyen, C. Ciraci, S. H. Oh, and J. Peraire, A nested hybridizable discontinuous Galerkin method for computing second-harmonic generation in three-dimensional metallic nanostructures, *J. Comput. Phys.* **429**, 110000 (2021).
- [58] M. E. A. Panah, O. Takayama, S. V. Morozov, K. E. Kudryavtsev, E. S. Semenova, and A. V. Lavrinenko, Highly doped inp as a low loss plasmonic material for mid-ir region, *Opt. Express* **24**, 29077 (2016).
- [59] M. E. A. Panah, L. Han, K. Norrman, N. Pryds, A. Nadtochiy, A. Zhukov, A. V. Lavrinenko, and E. S. Semenova, Mid-ir optical properties of silicon doped inp, *Opt. Mater. Express* **7**, 2260 (2017).
- [60] R. W. Boyd, *Nonlinear Optics* (Academic, San Diego, 2006).
- [61] K. A. O'Donnell and R. Torre, Characterization of the second-harmonic response of a silver-air interface, *New J. Phys.* **7**, 154 (2005).
- [62] L. Rodríguez-Suné, J. Trull, C. Cojocaru, N. Akozbek, D. de Ceglia, M. A. Vincenti, and M. Scalora, Harmonic generation from gold nanolayers: bound and hot electron contributions to nonlinear dispersion, *Opt. Express* **29**, 8581 (2021).
- [63] J. Le Perche, P. Quémerais, A. Barbara, and T. López-Ríos, Why Metallic Surfaces with Grooves a Few Nanometers Deep and Wide May Strongly Absorb Visible Light, *Phys. Rev. Lett.* **100**, 066408 (2008).
- [64] A. Pitelet, N. Schmitt, D. Loukrezis, C. Scheid, H. De Gerssem, C. Ciraci, E. Centeno, and A. Moreau, Influence of spatial dispersion on surface plasmons, nanoparticles, and grating couplers, *J. Opt. Soc. Am. B* **36**, 2989 (2019).
- [65] M. Dechaux, P. H. Tichit, C. Ciraci, J. Benedicto, R. Pollès, E. Centeno, D. R. Smith, and A. Moreau, Influence of spatial dispersion in metals on the optical response of deeply subwavelength slit arrays, *Phys. Rev. B* **93**, 045413 (2016).