Self-excitation of surface plasmon polaritons

V. G. Bordo*

NanoSyd, Mads Clausen Institute, Syddansk Universitet, Alsion 2, DK-6400 Sønderborg, Denmark

(Received 21 January 2015; revised manuscript received 29 February 2016; published 18 April 2016)

The novel effect of self-excitation of surface plasmons (SESP) in a plasmonic nanocavity is predicted, and its theory is developed from first principles. It is assumed that the cavity is formed by a nanogap between two metals and contains polarizable inclusions. Basing on the dyadic Green's function of the structure, the equations for the field in the cavity are investigated. It is shown that under certain conditions the field becomes unstable that leads to its self-excitation. The threshold criterion for self-excitation as well as the frequency of self-oscillation are derived in an analytical form. The SESP effect is explained in terms of a positive feedback for the polarization of inclusions provided by the field reflected from the cavity walls. These findings suggest a principally new avenue to surface plasmon generation which does not employ stimulated emission and is different from SPASER or plasmon laser.

DOI: 10.1103/PhysRevB.93.155421

I. INTRODUCTION

Surfaces of metals and their interfaces with dielectrics support electromagnetic excitations known as surface plasmon polaritons (SPPs) or briefly surface plasmons (SPs). Their field is localized near the boundary in a narrow region of a thickness which is less than the optical radiation wavelength. This remarkable property has been used in plenty of metal nanostructures to generate, control, and manipulate electromagnetic fields on a nanoscale that is the subject matter of plasmonics [1]. This field holds promise for developing diverse optical and optoelectronic nanodevices for subwavelength waveguiding [2], light energy concentration [3], ultra-sensitive sensing [4], high-resolution microscopy [5], ultra-fast computations, and many other applications.

One of the most serious obstacles preventing from wide utilization of SPPs in photonic circuits is high metallic losses which lead to short SPP propagation lengths in the optical spectral range. Moreover, the SPP mode losses increase with the mode confinement that blocks the further miniaturization of plasmonic devices. As a solution to this problem, it has been suggested to introduce optical gain (i.e., a population inversion) in the dielectric material bordering the metal surface to reduce SPP propagation losses [6]. This effect has been demonstrated experimentally for simple plasmonic waveguides [7,8]. It was also proposed that in such systems Surface Plasmon Amplification by Stimulated Emission of Radiation (SPASER) can occur if the gain medium undergoes a strong enough pumping [9,10]. This mechanism can be implemented in plasmonic metamaterials which are considered as being promising in the field of active nanoplasmonics [11].

However, currently available gain materials put limits for possible spaser characteristics, such as spasing threshold and output intensity, and for its operation conditions [12]. Moreover, for the tightly confining subwavelength plasmonic waveguides with electrically pumped semiconductor gain medium the current densities required to achieve SPP amplification and spasing are unsustainably high [13,14]. These limitations can be weakened to some extent in a three-level scheme of pumping with one of the transitions coupled to a plasmon mode [15]. In such a case, spasing can be achieved without population inversion on the spasing transition. This scheme nevertheless requires pumping at the other two transitions one of which should be coherently driven.

In this paper, we propose a principally new avenue to SPP generation. It is based on the utilization of a feedback mechanism which exists for an oscillating dipole moment in close vicinity (within a wavelength) from a reflecting surface. Such a dipole is driven by its own reflected field that modifies essentially its dynamics [16]. In particular, its damping constant is renormalized in accordance with the imaginary part of the reflected field at the dipole position. In the case where the latter quantity is negative the reflected field performs a positive work on the dipole oscillations thus providing a positive feedback. This effect can occur, in principle, in any subwavelength structures, for example, between two parallel surfaces [17,18], inside or nearby a cylindrical waveguide [19], or inside a cylindrical cavity [20]. If an ensemble of dipoles is located nearby a surface their mutual interaction mediated by the reflected field can cause a collective instability which is accompanied by the ensemble polarization amplification. It was predicted that such *a polarization loop gain* can lead to lasing without inversion in specially designed nanocavities [21].

In the case of a dipole in/at a metal cavity the dipole field launches SPP modes [22] which act back on the dipole polarization thus establishing a feedback loop. The considered mechanism is an essentially classical effect. It does not imply stimulated emission in the structure and hence it does not require a population inversion in the dipole. In this sense, the predicted here effect of Self-Excitation of Surface Plasmons (SESP) is principally different from SPASER or plasmon laser [23].

The proposed principle has much in common with the other processes which are known in literature as self-excited oscillations or self-oscillations [24]. All such phenomena are closed loop processes which employ a positive feedback. Their point of departure is a small initial deviation or field (a seed) leading to an unstable initial exponential increase of the output. This increase is usually limited by a certain mechanism, for example by saturation. Lasers and spasers are

*bordo@mci.sdu.dk

another examples of such systems where a seed is provided by the initial number of spontaneously emitted quanta.

In this paper, we consider the SESP effect in a metal-insulator-metal (MIM) cavity. Basing on the dyadic Green's function of this structure we derive a criterion for self-excitation of SPPs as well as an expression for the frequency of self-oscillation. The paper is organized as follows. In Sec. II, we introduce a simple model which facilitates understanding of the mechanism discussed in the subsequent sections. In Sec. III, a detailed theory of the SESP effect is developed and the criterion for the self-excitation of SPPs is formulated. Some numerical results are given in Sec. IV. The main results of the paper are summarized in Sec. V.

II. SIMPLE MODEL

Before considering the detailed theory of the SESP effect, let us turn to a simple model which will help to understand the underlying mechanism. Let us assume that a harmonically oscillating dipole p located nearby a reflecting surface is excited by an external field $E\exp(-i\omega t)$ of the frequency ω (for simplicity, we consider a one-dimensional problem). The equation of motion of the dipole reads as

$$\ddot{p} + 2\gamma\dot{p} + \omega_0^2 p = \frac{e^2}{m}(Ee^{-i\omega t} + E_R), \tag{1}$$

where ω_0 is the frequency of free dipole oscillations, γ is the damping constant, e and m are the charge and the effective mass of the dipole, respectively, and E_R is the electric field reflected from the surface at the dipole position. The reflected field is proportional to the dipole moment and can be written as

$$E_R = Fp \tag{2}$$

with F being the proportionality coefficient.

Let us note that Eq. (2) expresses a trivial fact that the solution of linear Maxwell's equations is proportional to the source magnitude which is given by p [cf. Eq. (25) below]. The coefficient F here has a sense of the reflected contribution to the field susceptibility, $\bar{F}^R(\mathbf{r},\mathbf{r}';\omega)$, which is discussed in Section III, taken at the dipole position $\mathbf{r} = \mathbf{r}' = \mathbf{r}_0$. The explicit form of the field susceptibility for a dipole located above an arbitrary flat surface is given in Ref. [25].

Now, combining Eqs. (1) and (2), we come to the equation

$$\ddot{p} + 2\gamma\dot{p} + \bar{\omega}_0^2 p = \frac{e^2}{m}Ee^{-i\omega t}, \tag{3}$$

where we have introduced the renormalized frequency $\bar{\omega}_0$ so that

$$\bar{\omega}_0^2 = \omega_0^2 - \frac{e^2}{m}F = \omega_0^2 - \frac{e^2}{m}\mathrm{Re}F - i\frac{e^2}{m}\mathrm{Im}F. \tag{4}$$

Applying the Laplace transform to Eq. (3) with a zero initial condition, one obtains that the Laplace-transformed quantity $\bar{p}(s)$ has three poles at $s = -i\omega$ and $s = s_\pm$ with

$$s_\pm = -\gamma \pm i\sqrt{\bar{\omega}_0^2 - \gamma^2}. \tag{5}$$

In the vicinity of the resonance, where $|\omega - \omega_0| \sim \gamma$, the contribution to $p(t)$ which comes from the pole $s = s_+$ has an order of magnitude γ/ω_0 times smaller than that from the other two poles and can be neglected. The term originating from the pole $s = i\omega$ describes forced vibrations, while the one originating from the pole $s = s_-$ corresponds to damped free oscillations.

Assuming a reasonable approximation that $\gamma^2, (e^2/m)|F| \ll \omega_0^2$, one finds

$$\mathrm{Re}(s_-) \approx -\gamma - \frac{e^2}{2m\omega_0}\mathrm{Im}F. \tag{6}$$

Equation (6) describes the modification of the damping constant in the time dependence $p(t) \sim \exp(s_- t)$ which is due to the presence of the reflecting surface. It can lead to either an increase or a decrease of the damping depending on whether $\mathrm{Im}F > 0$ or $\mathrm{Im}F < 0$. Let us note that such an approach has been proven to be an efficient model which describes quantum electrodynamics of an atomic system in the vicinity of a reflecting surface [16]. It correctly describes the lifetimes of emitting atoms and molecules near a metal mirror.

Let us consider now an ensemble of M identical oscillating dipoles nearby a reflecting surface. Their equations of motion are given by

$$\ddot{p}_i + 2\gamma\dot{p}_i + \omega_0^2 p_i = \frac{e^2}{m}\left(Ee^{-i\omega t} + \sum_{j=1}^{M} F_{ij}p_j\right), \quad i = 1,2,\ldots,M, \tag{7}$$

where the quantity F_{ij} describes the reflected field of the jth dipole at the position of the ith dipole. For the sake of simplicity, we neglect here the dipole-dipole interactions which are not relevant to the feedback mechanism considered below and assume that the external field amplitude E does not change significantly across the ensemble of dipoles.

Let us assume that the quantities F_{ij} are factorized as

$$F_{ij} = g_i h_j \tag{8}$$

(this corresponds to a degenerate kernel in the integral equation discussed in Sec. III A). Then one comes to the equation

$$\ddot{\mathcal{P}} + 2\gamma\dot{\mathcal{P}} + \tilde{\omega}_0^2\mathcal{P} = \frac{e^2}{m}\mathcal{H}Ee^{-i\omega t}, \tag{9}$$

with

$$\tilde{\omega}_0^2 = \omega_0^2 - \frac{e^2}{m}\mathcal{F}, \tag{10}$$

$$\mathcal{P} = \sum_{i=1}^{M} h_i p_i, \tag{11}$$

$$\mathcal{H} = \sum_{i=1}^{M} h_i \tag{12}$$

and

$$\mathcal{F} = \sum_{i=1}^{M} F_{ii}, \tag{13}$$

which is equivalent to Eq. (3).

One obtains from here that the quantity $\mathcal{P}$, which is a linear combination of the dipole moments, has a component

which depends on time as $\exp(\tilde{s}_- t)$ with $\mathrm{Re}(\tilde{s}_-) \approx -\gamma - (e^2/2m\omega_0)\mathrm{Im}\mathcal{F}$. Let us note that in this case the quantity $\mathcal{F}$ is proportional to the number of dipoles in the system, M. If $\mathrm{Im}\mathcal{F}$ is negative, then for a large enough number M the quantity $\mathrm{Re}(\tilde{s}_-)$ becomes positive, that signifies an exponential increase of the dipole moments with time, i.e., their self-excitation.

In the examples considered above, the quantity $\mathrm{Im}F$ determines a phase shift between the dipole p and its own reflected field. If $\mathrm{Im}F < 0$, such a phase shift leads to a negative contribution to the damping constant. For a sufficiently large ensemble of dipoles, a negative value of $\mathrm{Im}\mathcal{F}$ is equivalent to 'negative damping' which is known to be accompanied by self-excitation [24].

As is well known, a dipole oscillating above a conducting surface excites SPPs at the surface [26]. Accordingly, a self-excited ensemble of dipoles will generate at a metal surface sustained SPPs whose amplitude will exponentially increase with time.

III. THEORY OF THE SESP EFFECT

Let us consider a subwavelength gap between two identical metals with the dielectric function $\epsilon_m(\omega)$ which is enclosed from all sides with perfectly conducting plates [27] [see Fig. 1(a)]. Let us assume that the gap (host) material has the dielectric function ϵ_h and contains polarizable inclusions which have a resonance at the frequency ω_0. Let the gap thickness be d and its two other extensions be L_x and L_y along the x and y axes, respectively. Let us assume that such a cavity undergoes excitation by an external electromagnetic field which is given by $\mathbf{E}_i(\mathbf{r})\exp(-i\omega t)$ inside the gap with ω being its frequency. We shall then seek the field in the cavity in the form $\tilde{\mathbf{E}}(\mathbf{r},t)\exp(-i\omega t)$, where a tilde denotes an amplitude which varies in time much slower than $\exp(\pm i\omega t)$.

The external field polarizes the inclusions which in their turn act as sources of the scattered electromagnetic field in the cavity. A part of this scattered field is represented by SPP modes of the cavity. On the other hand, the scattered field is reflected by the cavity walls and acts back on the polarization of inclusions thus providing a feedback mechanism [Fig. 1(b)]. This effect can be described in terms of the field susceptibility tensor [25], $\bar{\mathbf{F}}(\mathbf{r},\mathbf{r}';\omega)$, which relates the electric field at the point $\mathbf{r}$ generated by a classical dipole, oscillating at frequency ω, with the dipole moment itself, located at $\mathbf{r}'$. It can be decomposed into direct, $\bar{\mathbf{F}}^0$, and reflected, $\bar{\mathbf{F}}^R$, contributions, the latter one being originating from the dipole field reflected from the cavity walls.

A. Integral equation for the cavity field

To derive the equation for the scattered electric field in the gap, we start with the equation for the Hertz vector potential, $\mathbf{\Pi}$, associated with a distribution of electric dipoles (Gaussian units) [28]:

$$\nabla^2\mathbf{\Pi} - \frac{\epsilon_h}{c^2}\frac{\partial^2\mathbf{\Pi}}{\partial t^2} = -\frac{4\pi}{\epsilon_h}\mathbf{P}, \tag{14}$$

where $\mathbf{P}$ is the overall polarization of the dipoles and we have taken into account that the magnetic permeability $\mu = 1$. The

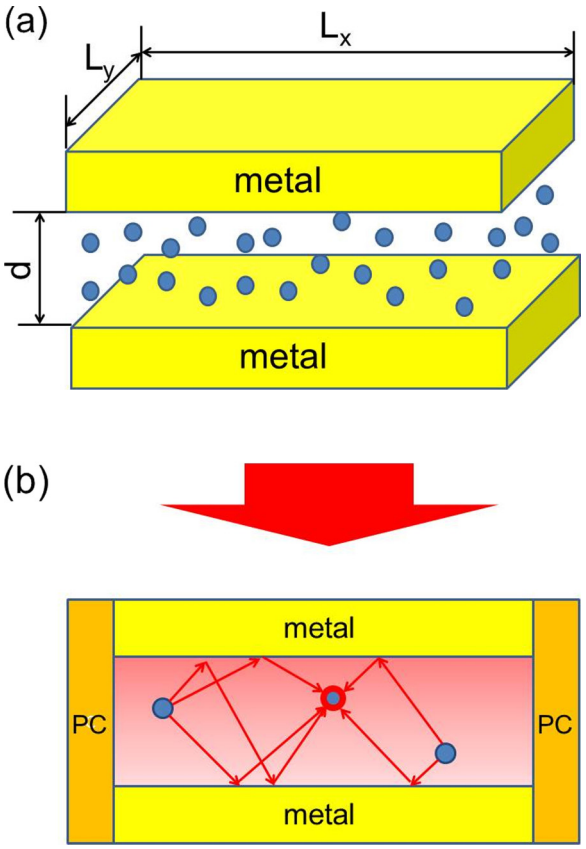

FIG. 1. (a) Geometry of the MIM cavity under consideration. Inclusions in the gap material are shown by dark blue circles. The cavity is enclosed by perfectly conducting walls (not shown). (b) Scheme of the cavity excitation. The large red arrow represents an external field which penetrates into the gap. The red thin arrows show some of the possible paths of the field scattered by the inclusions and reflected by the cavity walls. In the case of a positive feedback, the resulting field which each inclusion undergoes is enhanced. This leads in turn to a larger scattered field that establishes loop gain.

electric field amplitude is expressed in terms of $\mathbf{\Pi}$ as follows

$$\mathbf{E}_{\mathrm{sca}} = \nabla(\nabla\cdot\mathbf{\Pi}) - \frac{\epsilon_h}{c^2}\frac{\partial^2\mathbf{\Pi}}{\partial t^2}. \tag{15}$$

Introducing the Fourier transforms in time,

$$\mathbf{\Pi}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\pi(\Omega)e^{-i\Omega t}d\Omega, \tag{16}$$

$$\mathbf{E}_{\mathrm{sca}}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{e}(\Omega)e^{-i\Omega t}d\Omega, \tag{17}$$

and

$$\mathbf{P}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathbf{p}(\Omega)e^{-i\Omega t}d\Omega, \tag{18}$$

one arrives to *exact* equations

$$\nabla^2\pi(\Omega) + \frac{\Omega^2}{c^2}\epsilon_h\pi(\Omega) = -\frac{4\pi}{\epsilon_h}\mathbf{p}(\Omega) \tag{19}$$

and

$$\mathbf{e}(\Omega) = \nabla(\nabla\cdot\pi(\Omega)) + \frac{\Omega^2}{c^2}\epsilon_h\pi(\Omega). \tag{20}$$

The solution of the inhomogeneous wave equation, Eq. (19), can be expressed in terms of the dyadic Green's function, $\bar{G}(\mathbf{r},\mathbf{r}';\Omega)$, as follows [20]

$$\pi(\mathbf{r};\Omega) = \frac{4\pi}{\epsilon_h}\int \bar{G}(\mathbf{r},\mathbf{r}';\Omega)\mathbf{p}(\mathbf{r}';\Omega)d\mathbf{r}'. \tag{21}$$

Now substituting Eq. (21) into Eq. (20) one finds for the electric field

$$\mathbf{e}(\mathbf{r};\Omega) = \frac{4\pi}{\epsilon_h}\int\left[\nabla(\nabla\cdot\bar{G}(\mathbf{r},\mathbf{r}';\Omega)) + \frac{\Omega^2}{c^2}\epsilon_h\bar{G}(\mathbf{r},\mathbf{r}';\Omega)\right]\mathbf{p}(\mathbf{r}';\Omega)d\mathbf{r}', \tag{22}$$

where the nabla operators act on the variable $\mathbf{r}$.

Making in Eq. (22) the substitution

$$\mathbf{p}(\mathbf{r};\Omega) = \mathbf{p}_0\delta(\mathbf{r}-\mathbf{r}_0), \tag{23}$$

which corresponds to a point dipole source of magnitude $\mathbf{p}_0$ located at $\mathbf{r}_0$, one obtains

$$\mathbf{e}(\mathbf{r};\Omega) = \frac{4\pi}{\epsilon_h}\left[\nabla(\nabla\cdot\bar{G}(\mathbf{r},\mathbf{r}_0;\Omega)) + \frac{\Omega^2}{c^2}\epsilon_h\bar{G}(\mathbf{r},\mathbf{r}_0;\Omega)\right]\mathbf{p}_0. \tag{24}$$

Alternatively, the same quantity can be written in terms of the field susceptibility tensor, $\bar{\mathbf{F}}(\mathbf{r},\mathbf{r}_0;\Omega)$, as follows [25]

$$\mathbf{e}(\mathbf{r};\Omega) = \bar{\mathbf{F}}(\mathbf{r},\mathbf{r}_0;\Omega)\mathbf{p}_0. \tag{25}$$

From here one concludes that

$$\bar{\mathbf{F}}(\mathbf{r},\mathbf{r}_0;\Omega) = \frac{4\pi}{\epsilon_h}\left[\nabla(\nabla\cdot\bar{G}(\mathbf{r},\mathbf{r}_0;\Omega)) + \frac{\Omega^2}{c^2}\epsilon_h\bar{G}(\mathbf{r},\mathbf{r}_0;\Omega)\right]. \tag{26}$$

This allows one to rewrite Eq. (22) in the form

$$\mathbf{e}(\mathbf{r};\Omega) = \int\bar{\mathbf{F}}(\mathbf{r},\mathbf{r}';\Omega)\mathbf{p}(\mathbf{r}';\Omega)d\mathbf{r}'. \tag{27}$$

On the other hand, one can consider an alternative approach to the electric field amplitude. Making in Eq. (14) the substitutions

$$\mathbf{P}(t) = \tilde{\mathbf{P}}(t)e^{-i\omega t} \tag{28}$$

and

$$\mathbf{\Pi}(t) = \tilde{\mathbf{\Pi}}(t)e^{-i\omega t}, \tag{29}$$

where a tilde denotes an amplitude varying much slower than $e^{\pm i\omega t}$, one obtains

$$\nabla^2\tilde{\mathbf{\Pi}} + \frac{\omega^2}{c^2}\epsilon_h\tilde{\mathbf{\Pi}} - \frac{\epsilon_h}{c^2}\left(\frac{\partial^2\tilde{\mathbf{\Pi}}}{\partial t^2} - 2i\omega\frac{\partial\tilde{\mathbf{\Pi}}}{\partial t}\right) = -\frac{4\pi}{\epsilon_h}\tilde{\mathbf{P}}. \tag{30}$$

Taking into account the inequalities

$$|\partial\tilde{\mathbf{\Pi}}/\partial t| \ll \omega|\tilde{\mathbf{\Pi}}| \tag{31}$$

and

$$|\partial^2\tilde{\mathbf{\Pi}}/\partial t^2| \ll \omega^2|\tilde{\mathbf{\Pi}}|, \tag{32}$$

one comes to an *approximate* equation

$$\nabla^2\tilde{\mathbf{\Pi}} + \frac{\omega^2}{c^2}\epsilon_h\tilde{\mathbf{\Pi}} \approx -\frac{4\pi}{\epsilon_h}\tilde{\mathbf{P}}. \tag{33}$$

One obtains from here the approximations for the Hertz vector

$$\tilde{\mathbf{\Pi}}(\mathbf{r},t) \approx \frac{4\pi}{\epsilon_h}\int\bar{G}(\mathbf{r},\mathbf{r}';\omega)\tilde{\mathbf{P}}(\mathbf{r}',t)d\mathbf{r}' \tag{34}$$

and for the scattered field

$$\tilde{\mathbf{E}}_{\rm sca}(\mathbf{r},t) \approx \int\bar{\mathbf{F}}(\mathbf{r},\mathbf{r}';\omega)\tilde{\mathbf{P}}(\mathbf{r}',t)d\mathbf{r}'. \tag{35}$$

Now the slowly varying amplitude of the total electric field in the cavity takes the form

$$\tilde{\mathbf{E}}(\mathbf{r},t) \approx \mathbf{E}_i(\mathbf{r}) + \int\bar{\mathbf{F}}(\mathbf{r},\mathbf{r}';\omega)\tilde{\mathbf{P}}(\mathbf{r}',t)d\mathbf{r}'. \tag{36}$$

Here the first term on the right-hand side represents the external field amplitude, whereas the second term describes the overall contribution of the field scattered by the inclusions. The polarization of inclusions under the integral depends, in its turn, on the total field in the cavity, $\tilde{\mathbf{E}}(\mathbf{r},t)$. Therefore, Eq. (36) is an integral equation with respect to the electric field in the cavity.

B. Direct and reflected contributions

The field susceptibility tensor can be decomposed into direct and reflected contributions,

$$\bar{\mathbf{F}}(\mathbf{r},\mathbf{r}';\omega) = \bar{\mathbf{F}}^0(\mathbf{r},\mathbf{r}';\omega) + \bar{\mathbf{F}}^R(\mathbf{r},\mathbf{r}';\omega), \tag{37}$$

which can in turn be represented as the Fourier integrals

$$\bar{\mathbf{F}}^\sigma(\mathbf{r},\mathbf{r}';\omega) = \frac{1}{(2\pi)^2}\int\bar{\mathcal{F}}^\sigma(z,z';\omega,\kappa)e^{i\kappa_x(x-x')}e^{i\kappa_y(y-y')}d\kappa_x d\kappa_y, \tag{38}$$

where $\sigma = 0$ and $\sigma = R$ distinguish between different contributions. The Fourier transform $\bar{\mathcal{F}}^R$ is given by [18]

$$\begin{aligned}\bar{\mathcal{F}}^R(z,z';\omega,\kappa) &= 2\pi i\frac{\tilde{\omega}^2}{W_g}[\hat{s}\hat{s}(C^-_{ss}(z')e^{-iW_g(z-z')} + C^+_{ss}(z')e^{iW_g(z-z')}) \\ &+ \hat{p}_+\hat{p}_-C_{p+p-}(z')e^{iW_g(z-z')} + \hat{p}_-\hat{p}_+C_{p-p+}(z')e^{-iW_g(z-z')} \\ &+ \hat{p}_+\hat{p}_+C_{p+p+}e^{iW_g(z-z')} + \hat{p}_-\hat{p}_-C_{p-p-}e^{-iW_g(z-z')}],\end{aligned} \tag{39}$$

where we have introduced the following notations:

$$C^-_{ss}(z') = \frac{R_s e^{iW_g(d-2z')} + R_s^2 e^{2iW_g d}}{1 - R_s^2 e^{2iW_g d}}, \tag{40}$$

$$C^+_{ss}(z') = \frac{R_s e^{iW_g(d+2z')} + R_s^2 e^{2iW_g d}}{1 - R_s^2 e^{2iW_g d}}, \tag{41}$$

$$C_{p+p-}(z') = \frac{R_p e^{iW_g(d+2z')}}{1 - R_p^2 e^{2iW_g d}}, \tag{42}$$

$$C_{p-p+}(z') = \frac{R_p e^{iW_g(d-2z')}}{1 - R_p^2 e^{2iW_g d}}, \tag{43}$$

$$C_{p-p-} = C_{p+p+} = \frac{R_p^2 e^{2iW_g d}}{1 - R_p^2 e^{2iW_g d}}, \tag{44}$$

with $\tilde{\omega} = \omega/c = 2\pi/\lambda$, c being the speed of light in vacuum and λ being the wavelength,

$$\kappa = \sqrt{\kappa_x^2 + \kappa_y^2}, \tag{45}$$

$$\hat{\vec{s}} = \hat{\vec{\kappa}} \times \hat{\vec{z}}, \tag{46}$$

$$\hat{\vec{p}}_{\pm} = (\tilde{\omega}\sqrt{\epsilon_h})^{-1}(\kappa\hat{\vec{z}} \mp W_g\hat{\vec{\kappa}}), \tag{47}$$

$$W_g = (\tilde{\omega}^2\epsilon_h - \kappa^2)^{1/2}. \tag{48}$$

Here R_s and R_p are the Fresnel reflection coefficients for s- and p-polarized light at a metal-gap interface, respectively, and a hat above a symbol notates a unit vector. The quantity $\bar{\mathcal{F}}^R(z,z';\omega,\kappa)$, Eq. (39), has the poles, originating from the normal modes (both waveguide and SPP modes) of the metal-gap-metal structure, which satisfy the equations

$$R_{s,p}e^{iW_g d} = \pm 1 \tag{49}$$

where the upper sign corresponds to odd modes and the lower sign corresponds to even modes. In what follows, we shall restrict ourselves by the consideration of the gap SPP of the MIM waveguide [29]. Such a SPP has a transverse magnetic (TM) polarization [p-polarization in Eq. (39)] and exists for all values of the gap thickness d.

The quantity $\bar{\mathbf{F}}^0(\mathbf{r},\mathbf{r}';\omega)$ describes the Lorentz local field of the inclusions embedded into the host medium. Its explicit form can be obtained by applying the Clausius-Mossotti relation to a composite medium containing inclusions (Gaussian units) [30,31]:

$$\tilde{\mathbf{E}}(\mathbf{r}) = \mathbf{E}_i(\mathbf{r}) + \frac{4\pi}{3\epsilon_h}\tilde{\mathbf{P}}(\mathbf{r}). \tag{50}$$

Comparing this expression with the contribution originating from $\bar{\mathbf{F}}^0$ in (36) one obtains

$$F^0_{ij}(\mathbf{r},\mathbf{r}';\omega) = \frac{4\pi}{3\epsilon_h}\delta_{ij}\delta(\mathbf{r}-\mathbf{r}') = \frac{4\pi}{3\epsilon_h}\delta_{ij}\delta(z-z') \times \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i\kappa_x(x-x')}e^{i\kappa_y(y-y')}d\kappa_x d\kappa_y \tag{51}$$

and consequently

$$\mathcal{F}^0_{ij}(z,z';\omega,\kappa) = \frac{4\pi}{3\epsilon_h}\delta_{ij}\delta(z-z'). \tag{52}$$

C. Polarization of inclusions

The polarization of inclusions, $\mathbf{P}(\mathbf{r},t)$, which enters the equation for the field in the cavity, Eq. (36), can be described in the framework of the harmonic oscillator model. It obeys the equation

$$\frac{d^2\mathbf{P}}{dt^2} + 2\gamma\frac{d\mathbf{P}}{dt} + \omega_0^2\mathbf{P} = a\tilde{\mathbf{E}}(t)e^{-i\omega t}, \tag{53}$$

where γ is the relaxation constant. The coefficient a characterizes the coupling between the inclusions and the electric field and depends on the model which is used to describe an inclusion (see Appendix A for the detail). To find the evolution of the cavity field, one has to solve Eq. (53) jointly with the integral equation (36).

Both the electric fields and the polarization can be expanded in the Fourier series over the intervals $0 \leqslant x \leqslant L_x$ and $0 \leqslant y \leqslant L_y$ as follows

$$\mathbf{E}_i(\mathbf{r}) = \sum_{m,n=1}^{\infty}\mathbf{e}^i_{mn}(z)\sin\left(\frac{\pi m}{L_x}x\right)\sin\left(\frac{\pi n}{L_y}y\right), \tag{54}$$

$$\tilde{\mathbf{E}}(\mathbf{r},t) = \sum_{m,n=1}^{\infty}\mathbf{e}_{mn}(z,t)\sin\left(\frac{\pi m}{L_x}x\right)\sin\left(\frac{\pi n}{L_y}y\right) \tag{55}$$

and

$$\tilde{\mathbf{P}}(\mathbf{r},t) = \sum_{m,n=1}^{\infty}\mathbf{p}_{mn}(z,t)\sin\left(\frac{\pi m}{L_x}x\right)\sin\left(\frac{\pi n}{L_y}y\right). \tag{56}$$

Here the set of integers m and n distinguishes between different Fabry-Pérot modes of the plasmonic cavity.

The substitution of the Fourier series into Eqs. (36) and (53) on the assumption that $L_x, L_y \gg 2d$ leads to the set of equations for the mode $\{mn\}$ electric field and polarization:

$$\mathbf{e}_{mn}(z,t) \approx \mathbf{e}^i_{mn}(z) + \int_{-d/2}^{d/2}\bar{\mathcal{F}}(z,z';\omega,\kappa_{mn})\mathbf{p}_{mn}(z',t)dz', \tag{57}$$

$$\frac{d\mathbf{p}_{mn}(z,t)}{dt} + (\gamma - i\Delta)\mathbf{p}_{mn}(z,t) = i\beta\mathbf{e}_{mn}(z,t), \tag{58}$$

where $\bar{\mathcal{F}} = \bar{\mathcal{F}}^0 + \bar{\mathcal{F}}^R$, $|\Delta| = |\omega - \omega_0| \ll \omega_0$, $\beta = a/(2\omega_0)$,

$$\kappa_{mn} = \pi\sqrt{\left(\frac{m}{L_x}\right)^2 + \left(\frac{n}{L_y}\right)^2} \tag{59}$$

is the absolute value of the mode wave vector, and we have neglected the terms of the order of γ/ω_0.

D. Time evolution of the mode field

To investigate the temporal behavior of the mode field one has to jointly solve Eqs. (57) and (58). Assuming that there is no initial polarization of inclusions and performing the Laplace transform of those equations in time one comes to the integral equation

$$\left[1 - \frac{4\pi}{3\epsilon_h}\chi(s)\right]\tilde{\mathbf{e}}_{mn}(z,s) = \frac{\mathbf{e}^i_{mn}(z)}{s} + \chi(s)\int_{-d/2}^{d/2}\bar{\mathcal{F}}^R(z,z';\omega,\kappa_{mn})\tilde{\mathbf{e}}_{mn}(z',s)dz', \tag{60}$$

where

$$\tilde{\mathbf{e}}_{mn}(z,s) = \int_0^{\infty}\mathbf{e}_{mn}(z,t)e^{-st}dt \tag{61}$$

is the Laplace transformed mode field amplitude with $s = \sigma + i\Omega$,

$$\chi(s) = \frac{i\beta}{s + \gamma - i\Delta} \tag{62}$$

is the Laplace transform of the linear susceptibility of the inclusions in the resonant approximation which is assumed to be isotropic.

The kernels in the integral equation (60) can be written as follows

$$\mathcal{F}^R_{jk}(z,z') = \mathcal{H}^-_{jk}(z')e^{-iW_g z} + \mathcal{H}^+_{jk}(z')e^{iW_g z}, \tag{63}$$

where the explicit form of the functions $\mathcal{H}^{\pm}_{jk}(z)$ is given in Appendix B. It has a form of a sum of two terms in which the dependencies on the z and z' coordinates are factorized. An integral equation with such a degenerate kernel can be reduced to a set of linear algebraic equations [32]. This set of equations can be written in a vector form as follows

$$\hat{N}\vec{\mathcal{E}}(s) = \frac{1}{s}\vec{\mathcal{E}}_i, \tag{64}$$

where

$$\hat{N} = \left[1 - \frac{4\pi}{3\epsilon_h}\chi(s)\right]\hat{I} - \chi(s)\hat{M}, \tag{65}$$

$$\vec{\mathcal{E}}(s) = \begin{pmatrix} \mathcal{E}^-_\kappa(s) \\ \mathcal{E}^+_\kappa(s) \\ \mathcal{E}^-_z(s) \\ \mathcal{E}^+_z(s) \end{pmatrix}, \tag{66}$$

$$\vec{\mathcal{E}}_i = \begin{pmatrix} \mathcal{E}^{i,-}_\kappa \\ \mathcal{E}^{i,+}_\kappa \\ \mathcal{E}^{i,-}_z \\ \mathcal{E}^{i,+}_z \end{pmatrix} \tag{67}$$

with

$$\mathcal{E}^{\pm}_j(s) = \sum_k \int_{-d/2}^{d/2} \mathcal{H}^{\pm}_{jk}(z)\tilde{e}_k(z,s)dz \tag{68}$$

and

$$\mathcal{E}^{i,\pm}_j = \sum_k \int_{-d/2}^{d/2} \mathcal{H}^{\pm}_{jk}(z)e^i_k(z)dz. \tag{69}$$

$\hat{I}$ is the unit 4×4 matrix, and we have omitted everywhere the subscripts mn for the sake of brevity. Here the matrix $\hat{M}$ has the following form:

$$\hat{M} = \begin{pmatrix} \mathcal{H}^{--}_{\kappa\kappa} & \mathcal{H}^{+-}_{\kappa\kappa} & \mathcal{H}^{--}_{\kappa z} & \mathcal{H}^{+-}_{\kappa z} \\ \mathcal{H}^{-+}_{\kappa\kappa} & \mathcal{H}^{++}_{\kappa\kappa} & \mathcal{H}^{-+}_{\kappa z} & \mathcal{H}^{++}_{\kappa z} \\ \mathcal{H}^{--}_{z\kappa} & \mathcal{H}^{+-}_{z\kappa} & \mathcal{H}^{--}_{zz} & \mathcal{H}^{+-}_{zz} \\ \mathcal{H}^{-+}_{z\kappa} & \mathcal{H}^{++}_{z\kappa} & \mathcal{H}^{-+}_{zz} & \mathcal{H}^{++}_{zz} \end{pmatrix} \tag{70}$$

with

$$\mathcal{H}^{\pm\alpha}_{jk} = \int_{-d/2}^{d/2} e^{\pm iW_g z}\mathcal{H}^{\alpha}_{jk}(z)dz, \quad \alpha = \pm. \tag{71}$$

If the solution of the vector equation (64) is known, the solution of the original integral equation (60) is found as follows (we omit the mn subscript)

$$\tilde{e}_j(z,s) = \left[1 - \frac{4\pi}{3\epsilon_h}\chi(s)\right]^{-1} \times \left\{\frac{e^i_j(z)}{s} + \chi(s)[\mathcal{E}^-_j(s)e^{-iW_g z} + \mathcal{E}^+_j(s)e^{iW_g z}]\right\}. \tag{72}$$

The temporal behavior of the mode field components is determined by the poles s_k of the functions $\tilde{e}_j(z,s)$ in the complex plane of s. Each pole results in a contribution to the mode field which is proportional to $\exp(s_k t)$. If at least one of the poles is located in the right half of the s plane, the mode field will exponentially increase with time.

As it follows from Eqs. (64) and (72), the quantities $\tilde{e}_j(s)$ have a pole at $s = 0$, which describes forced polarization oscillations $\sim\exp(-i\omega t)$. The other pole, which originates from the factors $[1 - (4\pi/3\epsilon_h)\chi(s)]^{-1}$ and $\chi(s)$, can be approximated as $s_0 \approx -\gamma + i\Delta$ that leads to damped free oscillations [33].

The other contributions come from the poles of the functions $\mathcal{E}^{\pm}_j(s)$. They are given by the zeros of the determinant of the matrix $\hat{N}$ which enters the left-hand side part of Eq. (64). The straightforward calculations give

$$\det\hat{N} = \left[1 - \frac{4\pi}{3\epsilon_h}\chi(s)\right]^3[1 - \chi(s)\mathcal{F}'_0] \tag{73}$$

with

$$\mathcal{F}_0 = \frac{4\pi i d R_p e^{iW_g d}}{1 - R_p^2 e^{2iW_g d}}\left[W_g\left(R_p e^{iW_g d} - \frac{\sin W_g d}{W_g d}\right) + \frac{\kappa^2}{W_g}\left(R_p e^{iW_g d} + \frac{\sin W_g d}{W_g d}\right)\right] \tag{74}$$

and $\mathcal{F}'_0 = \mathcal{F}_0 + (4\pi/3\epsilon_h)$. The first factor in (73) gives the pole s_0 which is considered above, while the second factor provides one more pole which can be approximated as

$$s_1 \approx -\gamma + i\Delta + i\beta\mathcal{F}'_0 \tag{75}$$

and corresponds to a resonant contribution to $e_j(z,t)$.

E. Self-excitation criterion

The term originating from the pole s_1, Eq. (75), describes the mode field evolution which is relevant to the feedback provided by the reflections from the cavity walls. This contribution has a temporal dependence $\sim\exp(s_1 t)$. The real part of s_1 may be either negative or positive, depending on the sign of $\mathrm{Im}(\mathcal{F}_0)$. The condition $\mathrm{Re}(s_1) > 0$ specifies the criterion for the mode self-excitation. It can be written in the form

$$\mathrm{Re}(\chi_r\mathcal{F}_0) > 1, \tag{76}$$

where we have introduced the linear susceptibility of inclusions at the resonance $\Delta = 0$, $\chi_r = i\beta/\gamma$, and we have taken into account that the quantities β and ϵ_h are real.

On the other hand, $\mathrm{Im}(s_0) \equiv \Omega_0$, gives the frequency of the mode self-oscillation, ω_{so}:

$$\omega_{so} = \omega - \Omega_0 = \omega_0 - \gamma\,\mathrm{Im}(\chi_r\mathcal{F}'_0), \tag{77}$$

which differs from both ω and ω_0 that represents the frequency pulling effect.

Let us note that the criterion (76) can be obtained from Poynting's theorem (see Appendix C). It can be interpreted as a condition that the average rate of doing work by the electromagnetic field in the cavity on polarizing the inclusions exceeds the rate of the energy dissipation in them.

An additional rule, which should hold for a nonzero self-excited mode field, follows from Eq. (57). The exciting field should have a nonzero Fourier component $\mathbf{e}^i_{mn}$, which in turn should have a nonzero projection onto the plane containing both the z axis and the mode wave vector, $\vec{\kappa}_{mn} = (\pi m/L_x, \pi n/L_y)$ (we imply a TM polarization of the mode).

In the self-oscillation criterion (76), the quantity $\mathcal{F}_0$ is determined completely by the properties of the cavity, whereas the parameter χ_r depends solely on the properties of the inclusions. Up to this point, we have not specified anywhere the physical origin of the polarizable inclusions. They can be represented by impurity atoms, ions, molecules, quantum dots (QDs), or metal nanoparticles (NPs). For a two-level model of inclusions, that can be applied for atomic systems or QDs, χ_r is found as $i\mu^2 N/(\hbar\gamma_\perp)$ with μ the root mean square of the dipole transition matrix element, $\gamma_\perp$ the transverse relaxation rate and N the number density of inclusions which are supposed to be in the ground state (see Appendix A). If the gap material is represented by a nanocomposite which contains spherical metal NPs, one can estimate χ_r as $3i\epsilon_h f\omega_{sp}/(4\pi\Gamma)$ with f the volume fraction of NPs, ω_{sp} the localized surface plasmon

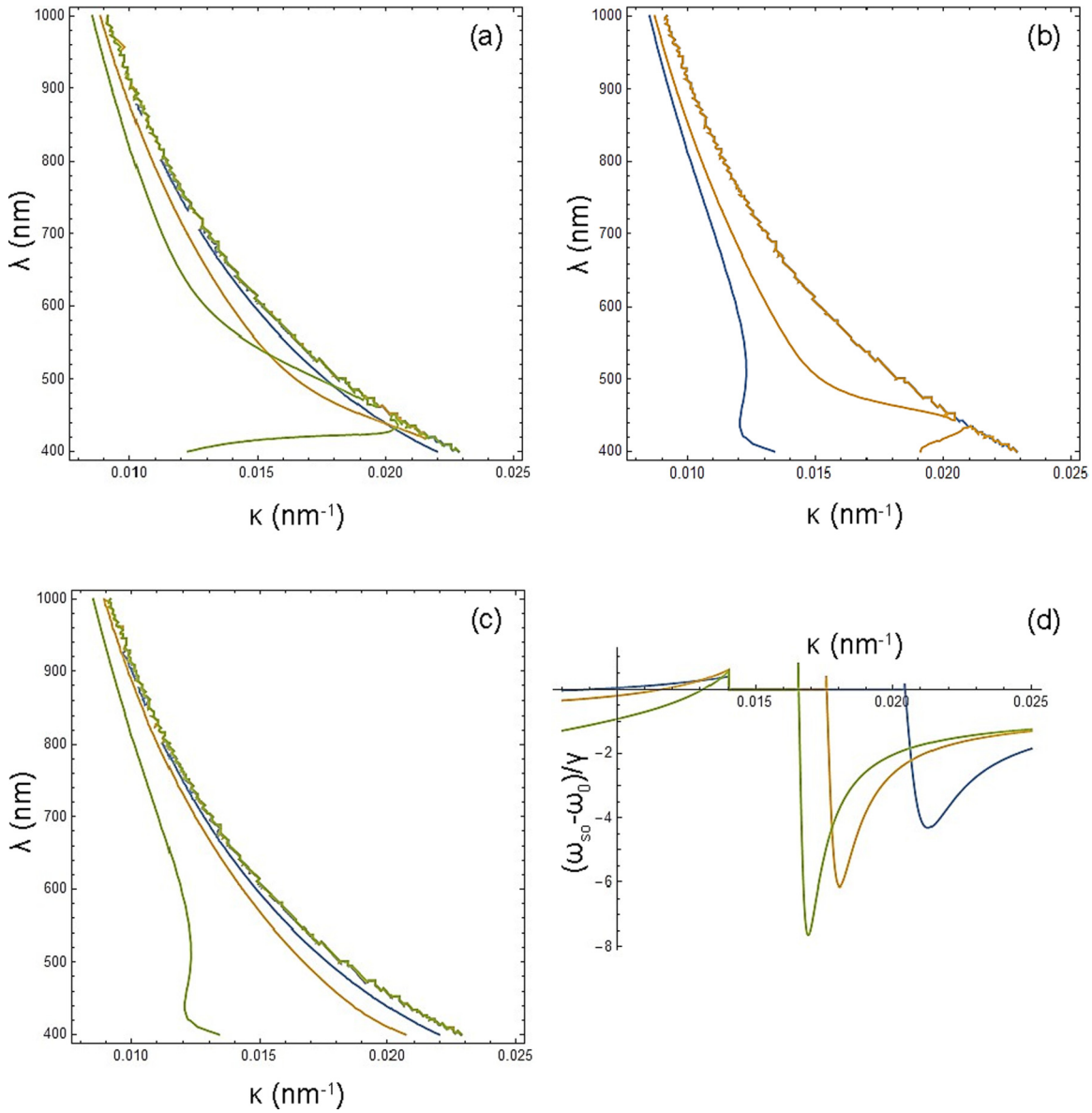

FIG. 2. Results of calculations for the MIM cavity composed of a glass slab ($\epsilon_h = 1.45^2$) containing Ag NPs ($f = 0.02$) between two metal surfaces. (a)–(c) Islands of instability. (a) A slab of different thickness is enclosed between Ag surfaces: $d = 50$ nm (blue curve), $d = 100$ nm (brown curve), and $d = 150$ nm (green curve). $\Gamma = 4.99 \times 10^{14}$ s^{-1}. (b) A slab of thickness $d = 50$ nm is enclosed between different metals: Ag (blue curve) and Au (brown curve). $\Gamma = 1.72 \times 10^{14}$ s^{-1}. (c) A slab of thickness $d = 50$ nm containing NPs with different Γ is enclosed between two Ag surfaces (the corresponding NP radius is given in brackets): $\Gamma = 4.99 \times 10^{14}$ s^{-1} ($R = 3$ nm) (blue curve), $\Gamma = 3.12 \times 10^{14}$ s^{-1} ($R = 5$ nm) (brown curve), and $\Gamma = 1.72 \times 10^{14}$ s^{-1} ($R = 10$ nm) (green curve). (d) Frequency pulling of the mode self-oscillation for $\lambda = 650$ nm as a function of κ. A slab of different thickness is enclosed between Ag surfaces: $d = 50$ nm (blue curve), $d = 100$ nm (brown curve), and $d = 150$ nm (green curve). $\Gamma = 4.99 \times 10^{14}$ s^{-1}. The discontinuities in the plots correspond to the regions where the series for the field susceptibility given in Ref. [18] do not converge.

polariton (LSPP) frequency, and Γ the electron collision rate in the NP material (see Appendix A).

IV. NUMERICAL RESULTS

We illustrate the general theory developed above by some numerical calculations. We assume that the gap is filled with glass which contains spherical Ag NPs. We adopt the Drude model for the dielectric function of metals, both in the cladding and in NPs, $\epsilon_m(\omega) = \epsilon_\infty - \omega_p^2/[\omega(\omega + i\Gamma)]$, where ω_p is the plasma frequency, Γ is the relaxation rate, and ϵ_∞ is the offset which takes into account the interband transitions [31]. For NPs the quantity Γ includes the size effect contribution and can be written as $\Gamma = \Gamma_\infty + bv_F/R$, where Γ_∞ is the damping constant for an unbounded metal, $b \approx 1$, v_F is the Fermi velocity and R characterizes the size of the metal particle [31]. Using the parameters given in Ref. [31] for silver, one obtains $\lambda_{sp} \equiv 2\pi c/\omega_{sp} \approx 410$ nm.

Figures 2(a)–2(c) show the contour plots $\mathrm{Re}(\chi_r \mathcal{F}_0) = 1$ in the plane κ-λ. The chosen values of the NP sizes and volume fraction correspond to those which can be obtained for Ag NPs embedded in a glass matrix [34,35]. The areas embraced by these contours correspond to the values of κ and λ for which the mode field is unstable and can be self-excited. They can thus be called the 'islands of instability.' Here λ is the wavelength of the exciting field whereas κ is defined by Eq. (59) and can be varied by changing the lengths L_x and L_y. The kinks at $\lambda \approx 410$ nm originate from the resonance with LSPPs in metal NPs. All islands have a common upper-right border which is described by the equation $\lambda = 2\pi\sqrt{\epsilon_h}/\kappa$ or, equivalently, $W_g = 0$. The vicinity of this curve is favorable for a feedback provided by the surface reflections: It corresponds to a grazing incidence of the waves, scattered by NPs, onto the cladding surfaces for which $R_p \approx -1$ and hence the losses in the metal cladding are minimized. As one can see, the area of the islands increases with the increase of d. It increases also with the decreasing of Γ (i.e., with the increasing of the NP size). Figure 2(b) illustrates how the shape of the islands depends on the metal cladding. The behavior of the frequency pulling of self-oscillation, $(\omega_{so} - \omega_0)/\gamma$, is shown in Fig. 2(d). The parts of this dependence which are outside the instability regions depicted in Fig. 2(a) correspond to decaying self-oscillations. This effect depends noticeably on the gap thickness, but depends only slightly on the cladding metal (not shown).

V. CONCLUSION

In conclusion, we have developed here the theory of SPP self-excitation in the MIM structure from first principles. We have derived an analytical criterion for the SPP self-excitation, Eq. (76), and an expression for the frequency pulling effect, Eq. (77). For a given cavity, the self-excitation threshold condition is imposed on the inclusions linear susceptibility, χ_r, rather than on the population inversion as in conventional lasers or spasers. The spectral interval where the SESP effect can occur spans over the visible and infrared regions and can be engineered by a proper design of the MIM cavity. An additional opportunity for tuning the resonances of the linear susceptibility χ can be provided by a choice of the metal composition, size, and shape of NPs as inclusions in the gap material [36]. For example, for a rod-shaped NP the LSPP resonance splits into two, one of which can be shifted to regions where the optical losses are lower [37].

The approach developed above predicts an exponential growth of the mode field with time. This is valid, however, only upon condition that the saturation effect is negligible, i.e., as long as the nonlinear terms in Eq. (53), which describe the evolution of the polarization, can be omitted. For metal nanoparticles in a nanocomposite this effect becomes essential at very high field intensities of the order of 50 GW/cm^2 [38].

Let us note that the theory developed here is formulated in completely classical terms. On the other hand, it has much in common with the semiclassical laser (or spaser) theory where stimulated emission plays key role. This similarity is, however, not accidental. According to papers [39,40] stimulated emission can be described classically if one introduces an appropriate phase shift between the dipole moment and the driving force in Eq. (53). Usually this effect disappears for an unperturbed (nonexcited) system due to the ensemble average over phases. However in our case all inclusion dipoles have the same phase with respect to their own reflected field in virtue of the condition $d \ll \lambda = 2\pi c/\omega$.

The general character of the polarization loop gain discussed here suggests that the SESP effect is not specific for the considered structure and can be observed in other subwavelength hybrid plasmonic structures. Its experimental verification will open up a prospect for the development of a new generation of nanoscale active plasmonic devices which do not require a powerful pumping and hence will not be highly energy consuming.

APPENDIX A: COEFFICIENT a IN EQ. (53)

The polarization of inclusions is described by Eq. (53), where the coupling coefficient a depends on the model of inclusions. We distinguish between two general cases: (i) the inclusions are represented by two-level systems which model impurity atoms, ions, molecules or quantum dots and (ii) the inclusions are noninteracting spherical metal nanoparticles.

(i) *Two-level systems.* In this case the equation for the polarization can be deduced from the optical Bloch equations [41]. In particular, for an isotropic distribution of anisotropic molecules

$$a = \frac{2\omega_0}{\hbar}\overline{|\mu|^2}N, \tag{A1}$$

where $\overline{|\mu|^2}$ is the average square of the transition dipole moment, N is the number density of molecules, and we have assumed that the population of the excited state is negligible.

(ii) *Metal nanoparticles.* In this case an explicit form of the coefficient a can be derived as follows. We require that Eq. (53) would give a quasistatic result for the polarizability of a spherical particle (Gaussian units) [28],

$$\alpha(\omega) = \epsilon_h \frac{\epsilon_m(\omega) - \epsilon_h}{\epsilon_m(\omega) + 2\epsilon_h} R^3, \tag{A2}$$

where R is the sphere radius and

$$\epsilon_m(\omega) = \epsilon_\infty - \frac{\omega_p^2}{\omega(\omega + i\Gamma)} \tag{A3}$$

is the dielectric function of the nanoparticle material with ω_p the plasma frequency, Γ the relaxation constant, and ϵ_∞ the offset which takes into account the interband transitions [31]. Using the inequalities $\Gamma,\omega \ll \omega_p$ Eq. (A2) can be reduced to the form

$$\alpha(\omega) \approx \epsilon_h \frac{\omega_{sp}^2}{\omega_{sp}^2 - \omega^2 - i\Gamma\omega} R^3, \tag{A4}$$

where $\omega_{sp} = \omega_p/\sqrt{\epsilon_\infty + 2\epsilon_h}$ is the frequency of the localized surface plasmon polariton supported by a spherical metal particle.

Assuming a monochromatic field ($\tilde{\mathbf{E}} = const$) and making the substitution $\mathbf{P} = \tilde{\mathbf{P}}e^{-i\omega t}$ in Eq. (53) one obtains

$$\tilde{\mathbf{P}} = \frac{a}{\omega_0^2 - \omega^2 - 2i\gamma\omega}\tilde{\mathbf{E}} \equiv N\alpha(\omega)\tilde{\mathbf{E}}, \tag{A5}$$

where N is the number density of nanoparticles. Now comparing Eqs. (A4) and (A5) one obtains $\omega_0 = \omega_{sp}$, $2\gamma = \Gamma$ and $a = N\epsilon_h\omega_{sp}^2 R^3$. The latter quantity can be alternatively written as $a = (3/4\pi) f \epsilon_h \omega_{sp}^2$ with f being the volume fraction of nanoparticles.

Figure 3 shows the comparison of $\alpha(\omega)$ calculated with the use of Eq. (A2) with its approximation given by Eq. (A4) in the spectral range under consideration. It confirms that Eq. (A4) gives a reasonable estimate for $\alpha(\omega)$.

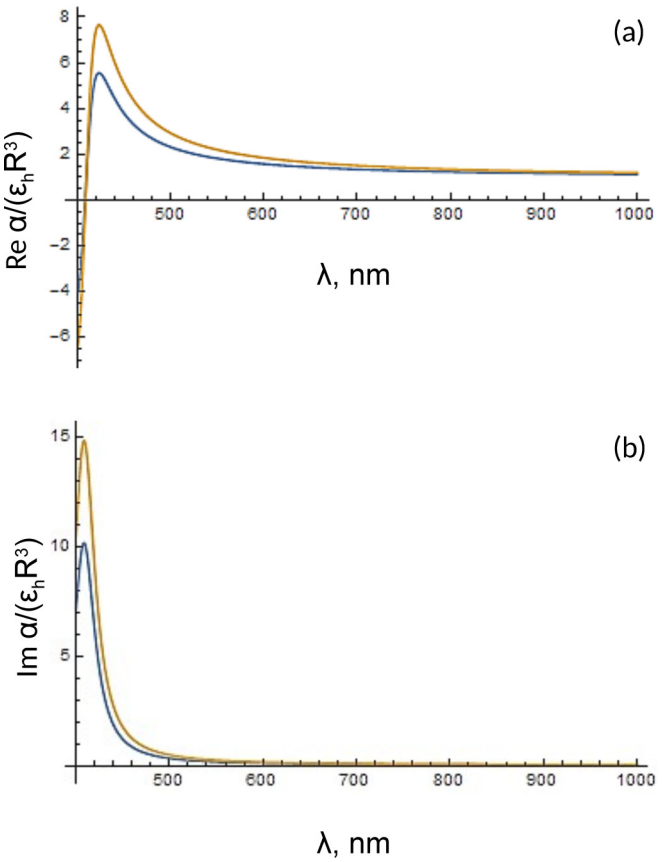

FIG. 3. Comparison of the real (a) and imaginary (b) parts of $\alpha(\omega)$ (blue curves) with their approximations (brown curves) for $\omega_p = 14.0 \times 10^{15}$ s^{-1}, $\Gamma = 3.12 \times 10^{14}$ s^{-1}, $\epsilon_\infty = 5$, and $\epsilon_h = 1.45^2$.

APPENDIX B: FUNCTIONS $\mathcal{H}_{jk}^{\pm}(z)$ IN EQ. (63)

The functions $\mathcal{H}_{jk}^{\pm}(z)$ which enter Eq. (63) have the following form:

$$\mathcal{H}_{\kappa\kappa}^{-}(z) = W_g H_0[R_p e^{iW_g(d+z)} - e^{-iW_g z}], \tag{B1}$$

$$\mathcal{H}_{\kappa\kappa}^{+}(z) = W_g H_0[R_p e^{iW_g(d-z)} - e^{iW_g z}], \tag{B2}$$

$$\mathcal{H}_{\kappa z}^{-}(z) = \kappa H_0[R_p e^{iW_g(d+z)} + e^{-iW_g z}], \tag{B3}$$

$$\mathcal{H}_{\kappa z}^{+}(z) = -\kappa H_0[R_p e^{iW_g(d-z)} + e^{iW_g z}], \tag{B4}$$

$$\mathcal{H}_{z\kappa}^{-}(z) = \kappa H_0[R_p e^{iW_g(d+z)} - e^{-iW_g z}], \tag{B5}$$

$$\mathcal{H}_{z\kappa}^{+}(z) = -\kappa H_0[R_p e^{iW_g(d-z)} - e^{iW_g z}], \tag{B6}$$

$$\mathcal{H}_{zz}^{-}(z) = \frac{\kappa^2}{W_g} H_0[R_p e^{iW_g(d+z)} + e^{-iW_g z}], \tag{B7}$$

$$\mathcal{H}_{zz}^{+}(z) = \frac{\kappa^2}{W_g} H_0[R_p e^{iW_g(d-z)} + e^{iW_g z}], \tag{B8}$$

where

$$H_0 = \frac{2\pi i}{\sqrt{\epsilon_h}} \frac{R_p e^{iW_g d}}{1 - R_p^2 e^{2iW_g d}}. \tag{B9}$$

APPENDIX C: POYNTING'S THEOREM

1. Electromagnetic energy balance

Let us consider the implementation of Poynting's theorem which expresses conservation of energy for the system under consideration. It can be written in the following form (Gaussian units) [28]:

$$\frac{c}{4\pi}\int_S (\mathbf{E}\times\mathbf{H})\mathbf{n}da + \int_V \mathbf{E}\mathbf{J}dv = -\frac{1}{4\pi}\int_V \left(\mathbf{E}\frac{\partial\mathbf{D}}{\partial t} + \mathbf{H}\frac{\partial\mathbf{B}}{\partial t}\right)dv, \tag{C1}$$

where we have used conventional notations. The right-hand side part of this equation represents the rate of change of the electromagnetic energy W_{em} contained in the volume V. The first term in the left-hand side is the flux of electromagnetic energy across the surface S which embraces the volume V while the second term expresses the rate of doing work on driving the current $\mathbf{J}$ by the electromagnetic field. Poynting's theorem is an identity which follows from Maxwell's equations and which is valid for any time dependence of the involved quantities.

Let us apply Poynting's theorem to the gap between two metals. Equation (C1) should be fulfilled for the total field in the volume which can be represented as

$$\mathbf{E} = \mathbf{E}_i + \mathbf{E}_{\text{sca}}, \tag{C2}$$

$$\mathbf{H} = \mathbf{H}_i + \mathbf{H}_{\text{sca}}, \tag{C3}$$

where $\mathbf{E}_i$ and $\mathbf{H}_i$ are the electric and magnetic field amplitudes, respectively, of the "incident" field which would exist in the gap without inclusions and $\mathbf{E}_{\text{sca}}$ and $\mathbf{H}_{\text{sca}}$ are the ones of the field scattered by the inclusions. In such an approach the inclusions are considered as the sources of the scattered field

which are characterized by the displacement current

$$\mathbf{J} = \frac{\partial \mathbf{P}}{\partial t} \tag{C4}$$

with $\mathbf{P}$ being the polarization of the inclusions. Under such a definition of $\mathbf{J}$, the quantity $\mathbf{D}$ in the right-hand side of Eq. (C1) has a sense of the electric displacement field associated with the polarization of the host material in the gap (i.e., without inclusions), $\mathbf{D} = \epsilon_h \mathbf{E}$. Then Maxwell's equations for the "incident" and scattered fields are given by

$$\mathrm{rot}\mathbf{E}_i + \frac{1}{c}\frac{\partial \mathbf{H}_i}{\partial t} = 0, \tag{C5}$$

$$\mathrm{rot}\mathbf{H}_i - \frac{\epsilon_h}{c}\frac{\partial \mathbf{E}_i}{\partial t} = 0 \tag{C6}$$

and

$$\mathrm{rot}\mathbf{E}_{\mathrm{sca}} + \frac{1}{c}\frac{\partial \mathbf{H}_{\mathrm{sca}}}{\partial t} = 0, \tag{C7}$$

$$\mathrm{rot}\mathbf{H}_{\mathrm{sca}} - \frac{\epsilon_h}{c}\frac{\partial \mathbf{E}_{\mathrm{sca}}}{\partial t} = \frac{4\pi}{c}\mathbf{J}, \tag{C8}$$

respectively.

The energy flux across the surface S represented by the first term in the left-hand side of Eq. (C1) can be accordingly written in the form

$$Q = Q_i + Q_{\mathrm{sca}} + Q_{\mathrm{int}} \tag{C9}$$

with

$$Q_i = \frac{c}{4\pi}\int_S (\mathbf{E}_i \times \mathbf{H}_i)\mathbf{n}da, \tag{C10}$$

$$Q_{\mathrm{sca}} = \frac{c}{4\pi}\int_S (\mathbf{E}_{\mathrm{sca}} \times \mathbf{H}_{\mathrm{sca}})\mathbf{n}da \tag{C11}$$

and

$$Q_{\mathrm{int}} = \frac{c}{4\pi}\int_S (\mathbf{E}_{\mathrm{sca}} \times \mathbf{H}_i)\mathbf{n}da + \frac{c}{4\pi}\int_S (\mathbf{E}_i \times \mathbf{H}_{\mathrm{sca}})\mathbf{n}da. \tag{C12}$$

Let us note that the latter quantity originates from interference between the "incident" and the scattered field. Equations (C10)–(C12) can be reduced to integrals over the volume V using the identity

$$\int_S (\mathbf{A} \times \mathbf{B})\mathbf{n}da = \int_V \mathrm{div}(\mathbf{A} \times \mathbf{B})dv, \tag{C13}$$

with $\mathbf{A}$ and $\mathbf{B}$ being arbitrary vectors. The right-hand side of Eq. (C13) can in turn be transformed with the use of the identity

$$\mathrm{div}(\mathbf{A} \times \mathbf{B}) = \mathbf{B}\mathrm{rot}\mathbf{A} - \mathbf{A}\mathrm{rot}\mathbf{B}. \tag{C14}$$

The fields everywhere in the equations above should be taken in a real form, i.e.,

$$\mathbf{E}(t) = \tfrac{1}{2}[\tilde{\mathbf{E}}(t)e^{-i\omega t} + \tilde{\mathbf{E}}^*(t)e^{i\omega t}], \tag{C15}$$

etc., where a tilde denotes a slowly varying in time amplitude. Substituting the fields in the form (C15) into the expressions (C10)–(C12), carrying out an average over a time interval much larger than $2\pi/\omega$ and taking into account Maxwell's equations (C5)–(C8), one obtains

$$\langle Q_i \rangle = 0, \tag{C16}$$

$$\langle Q_{\mathrm{sca}} \rangle = \frac{\omega}{2}\int_V \mathrm{Im}(\tilde{\mathbf{E}}_{\mathrm{sca}}\tilde{\mathbf{P}}^*)dv \tag{C17}$$

and

$$\langle Q_{\mathrm{int}} \rangle = \frac{\omega}{2}\int_V \mathrm{Im}(\tilde{\mathbf{E}}_i\tilde{\mathbf{P}}^*)dv, \tag{C18}$$

where the angular brackets denote averaging over time. Here we have taken into account that ϵ_h is real and we have neglected the time derivatives of the slowly varying amplitudes. Analogously, the rate of doing work by the electromagnetic field which is given by the second term in the left-hand side of Eq. (C1) averaged over time is obtained in the form

$$\langle A \rangle = -\frac{\omega}{2}\int_V \mathrm{Im}(\tilde{\mathbf{E}}\tilde{\mathbf{P}}^*)dv. \tag{C19}$$

The quantities (C16)–(C19) being combined together give

$$\langle Q_{\mathrm{int}} \rangle + \langle Q_{\mathrm{sca}} \rangle + \langle A \rangle = 0, \tag{C20}$$

that expresses the conservation of the total electromagnetic energy in the system as it is *a priori* expected.

2. Total energy balance

The latter quantity, Eq. (C19), can be represented in a different way. Following Loudon [42] we can write

$$\mathbf{E}\dot{\mathbf{P}} = \frac{1}{2a}\frac{d}{dt}\left(\dot{\mathbf{P}}^2 + \omega_0^2\mathbf{P}^2\right) + \frac{2\gamma}{a}\dot{\mathbf{P}}^2, \tag{C21}$$

where we have used the equation for the polarization $\mathbf{P}$. The first term on the right-hand side of this equation can be regarded as the time derivative of the mechanical (kinetic plus potential) energy density, W_m, of the inclusions [43]. The rate corresponding to this term in Poynting's equation (C1), being moved to the right-hand side, complements the rate of change of the electromagnetic energy to that of the total (electromagnetic plus mechanical) energy. The remaining term, $(2\gamma/a)\dot{\mathbf{P}}^2$, gives the rate of the energy dissipation, Q_{dis}, in the inclusions due to the damping mechanism [42]. Carrying out the time averaging as before, one obtains

$$\langle Q_{\mathrm{dis}} \rangle = \frac{\gamma\omega^2}{a}\int_V |\tilde{\mathbf{P}}|^2 dv. \tag{C22}$$

Now the total energy balance in the gap is given by the equation

$$\langle Q_{\mathrm{int}} \rangle + \langle Q_{\mathrm{sca}} \rangle + \langle Q_{\mathrm{dis}} \rangle = -\left\langle \frac{dW_m}{dt} \right\rangle, \tag{C23}$$

where we have taken into account that $\langle dW_{em}/dt \rangle = 0$.

The left-hand side of this equation has the form

$$\langle Q_{\rm int}\rangle + \langle Q_{\rm sca}\rangle + \langle Q_{\rm dis}\rangle = \frac{\omega}{2}\int_V \left\{ {\rm Im}[(\tilde{\mathbf{E}}_i + \tilde{\mathbf{E}}_{\rm sca})\tilde{\mathbf{P}}^*] + \frac{2\gamma\omega}{a}|\tilde{\mathbf{P}}|^2 \right\} dv$$
$$= \frac{\omega}{2}\frac{L_x L_y}{4}\sum_{m,n=1}^{\infty}\int_{-d/2}^{d/2}\sum_{i,j}\left\{ {\rm Im}\left[\left(e^i_{mn,i}(z) + \int_{-d/2}^{d/2}\mathcal{F}^R_{ij}(z,z';\omega,\kappa_{mn})p_{mn,j}(z',t)dz' \right) p^*_{mn,i}(z,t) \right] \right.$$
$$\left. + \frac{2\gamma\omega}{a}\delta_{ij}p^*_{mn,i}(z)p_{mn,j}(z) \right\} dz, \tag{C24}$$

where we have taken into account the explicit form of $\mathcal{F}^0_{ij}$, Eq. (52). The time average on the right-hand side can be represented in an analogous form:

$$-\left\langle \frac{dW_m}{dt}\right\rangle = -\frac{\omega^2+\omega_0^2}{2a}\int_V {\rm Re}(\dot{\tilde{\mathbf{P}}}\tilde{\mathbf{P}}^*)dv = -\frac{\omega^2+\omega_0^2}{2a}\frac{L_x L_y}{4}\sum_{m,n=1}^{\infty}\int_{-d/2}^{d/2}\sum_i {\rm Re}[\dot{p}_{mn,i}(z,t)p^*_{mn,i}(z,t)]dz, \tag{C25}$$

where we have kept the lowest nonvanishing terms with respect to $\dot{\tilde{\mathbf{P}}}(t)$.

As it follows from here, Eq. (C23) is equivalent to the real part of the following one:

$$\sum_{m,n=1}^{\infty}\sum_i \left[-\frac{\omega^2+\omega_0^2}{2a}\dot{p}_{mn,i}(z,t) + i\frac{\omega}{2}e^i_{mn,i}(z) + i\frac{\omega}{2}\int_{-d/2}^{d/2}\sum_j \mathcal{F}^R_{ij}(z,z';\omega,\kappa_{mn})p_{mn,j}(z',t)dz' - \frac{\gamma\omega^2}{a}p_{mn,i}(z,t) \right] p^*_{mn,i}(z,t) = 0. \tag{C26}$$

Let us assume that only a single mode $\{mn\}$ is excited, i.e., $\mathbf{p}_{mn}(z,t) \neq 0$ for a single $\{mn\}$ pair. Then, taking into account that $\omega \approx \omega_0$ and $a = 2\omega_0\beta$, one comes to the equation

$$\dot{\mathbf{p}}_{mn}(z,t) + \gamma\mathbf{p}_{mn}(z,t) = i\beta\left[\mathbf{e}^i_{mn}(z) + \int_{-d/2}^{d/2}\bar{\mathcal{F}}^R(z,z';\omega,\kappa_{mn})\mathbf{p}_{mn}(z',t)dz' \right], \tag{C27}$$

which is equivalent to Eqs. (57) and (58) with $\Delta \approx 0$. The criterion of self-excitation for the solution of Eq. (C27) is the same as before, Eq. (76).

The increasing of the polarization (and hence the mechanical energy), associated with the mode $\{mn\}$, with time implies that the left-hand side in Eq. (C23) is negative. Therefore the criterion of self-excitation is equivalent to the condition that

$$-\langle Q_{\rm int}\rangle - \langle Q_{\rm sca}\rangle = \langle A\rangle > \langle Q_{\rm dis}\rangle, \tag{C28}$$

where we have taken into account Eq. (C20). In other words, the average rate of doing work by the electromagnetic field in the cavity on polarizing the inclusions exceeds the rate of the energy dissipation in them.

[1] S. A. Maier, *Plasmonics: Fundamentals and Applications*, (Springer, New York, 2007).
[2] D. K. Gramotnev and S. I. Bozhevolnyi, Nat. Photonics **4**, 83 (2010).
[3] J. A. Schuller, E. S. Barnard, W. Cai, Y. C. Jun, J. S. White, and M. L. Brongersma, Nat. Mater. **9**, 193 (2010).
[4] A. V. Kabashin, P. Evans, S. Pastkovsky, W. Hendren, G. A. Wurtz, R. Atkinson, R. Pollard, V. A. Podolskiy, and A. V. Zayats, Nat. Mater. **8**, 867 (2009).
[5] A. Hartschuh, E. J. Sánchez, X. S. Xie, and L. Novotny, Phys. Rev. Lett. **90**, 095503 (2003).
[6] A. N. Sudarkin and P. A. Demkovich, Sov. Phys. - Tech. Phys. **34**, 764 (1989).
[7] M. Ambati, S. H. Nam, E. Ulin-Avila, D. A. Genov, G. Bartal, and X. Zhang, Nano Lett. **8**, 3998 (2008).
[8] J. Grandidier, G. Colas des Francs, S. Massenot, A. Bouhelier, L. Markey, J.-C. Weeber, C. Finot, and A. Dereux, Nano Lett. **9**, 2935 (2009).
[9] D. J. Bergman and M. I. Stockman, Phys. Rev. Lett. **90**, 027402 (2003).
[10] M. I. Stockman, Opt. Express **19**, 22029 (2011).
[11] O. Hess, J. B. Pendry, S. A. Mayer, R. F. Oulton, J. M. Hamm, and K. L. Tsakmakidis, Nat. Mater. **11**, 573 (2012).
[12] K. Leosson, J. Nanophotonics **6**, 061801 (2012).
[13] J. B. Khurgin and G. Sun, Appl. Phys. Lett. **100**, 011105 (2012).
[14] J. B. Khurgin and G. Sun, Opt. Express **20**, 15309 (2012).
[15] K. E. Dorfman, P. K. Jha, D. V. Voronine, P. Genevet, F. Capasso, and M. O. Scully, Phys. Rev. Lett. **111**, 043601 (2013).
[16] R. R. Chance, A. Prock, and R. Silbey, J. Chem. Phys. **60**, 2744 (1974).
[17] S. M. Dutra and P. L. Knight, Phys. Rev. A **53**, 3587 (1996).
[18] H. Nha and W. Jhe, Phys. Rev. A **54**, 3505 (1996).
[19] H. Nha and W. Jhe, Phys. Rev. A **56**, 2213 (1997).
[20] V. Bordo, J. Opt. Soc. Am. B **29**, 1799 (2012).
[21] V. G. Bordo, Phys. Rev. A **88**, 013803 (2013).
[22] G. L. Yip, IEEE Trans. Microwave Theory Tech. **18**, 1033 (1970).
[23] R.-M. Ma, R. F. Oulton, V. J. Sorger, and X. Zhang, Laser Photonics Rev. **7**, 1 (2013).
[24] A. Jenkins, Phys. Rep. **525**, 167 (2013).

[25] J. M. Wylie and J. E. Sipe, Phys. Rev. A **30**, 1185 (1984).
[26] A. Sommerfeld, *Partial Differential Equations in Physics*, Lectures on Theoretical Physics, Vol. VI (Academic Press, New York, London, 1964), Secs. 32, 33.
[27] This assumption does not restrict the generality of the consideration, however it considerably simplifies the relevant mathematics.
[28] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941).
[29] S. I. Bozhevolnyi and T. Søndergaard, Opt. Express **15**, 10869 (2007).
[30] M. Born and E. Wolf, *Principles of Optics* (Pergamon Press, Oxford, 1970).
[31] W. Cai and V. Shalaev, *Optical Metamaterials* (Springer, New York, 2010).
[32] A. D. Polyanin and A. V. Manzhirov, *Handbook of Integral Equations* (CRC Press, Washington, 1998).
[33] Here the poles are given for a slowly varying field amplitude. To obtain the overall time dependence, one has to multiply this amplitude by $\exp(-i\omega t)$.
[34] M. C. Mathpal, P. Kumar, S. Kumar, A. K. Tripathi, M. K. Singh, J. Prakash, and A. Agarwal, RCS Adv. **5**, 12555 (2015).
[35] A. Stalmashonak, G. Seifert, and A. Abdolvand, *Ultra-Short Pulsed Laser Engineered Metal-Glass Nanocomposites* (Springer, Heidelberg, 2013).
[36] K.-S. Lee and M. A. El-Sayed, J. Phys. Chem. B **110**, 19220 (2006).
[37] M. Pelton, J. Aizpurua, and G. Bryant, Laser & Photon. Rev. **2**, 136 (2008).
[38] O. Plaksin, Y. Takeda, H. Amekura, N. Kishimoto, and S. Plaksin, J. Appl. Phys. **103**, 114302 (2008).
[39] M. Cray, M.-L. Shih, and P. W. Milonni, Am. J. Phys. **50**, 1016 (1982).
[40] B. Fain and P. W. Milonni, J. Opt. Soc. Am. B **4**, 78 (1987).
[41] R. H. Pantell and H. E. Puthoff, *Fundamentals of Quantum Electronics* (John Wiley & Sons, New York, 1969).
[42] R. Loudon, J. Phys. A: Gen. Phys. **3**, 233 (1970).
[43] J. D. Jackson, *Classical Electrodynamics* (John Wiley & Sons, New York, 1998).