Theory of nonretarded ballistic surface plasma waves in metal films

Hai-Yao Deng*
Department of Physics and Astronomy, University of Exeter, EX4 4QL Exeter, United Kingdom
(Received 6 February 2017; published 31 March 2017)

We present a theory of surface plasma waves in metal films with arbitrary electronic collision rate τ. Both *tangential* and *normal* modes are investigated. A universal self-amplification channel for these waves is established as a result of the unique interplay between ballistic electronic motions and boundary effects. The channel is shown to be protected by a general principle and its properties independent of τ. The effects of film thickness and surface roughness are also calculated. Experimental implications, such as Ferrel radiation, are discussed.

DOI: 10.1103/PhysRevB.95.125442

I. INTRODUCTION

Surface plasma waves (SPWs) [1–3] are fascinating to a wide spectrum of scientists not only for their fundamental physical properties [4–6] but also their promising potential [7] in a myriad of applications, including microscopy [8], sensing [9], and nano-optics [10–15] as well as information processing [16]. As charge density waves highly localized about the interface between a metal and a dielectric, SPWs strongly interact and form a bound entity with light that might render an atomic resolution of molecular dynamics [17]. Nowadays SPWs are pivotal in nano-optics.

The standard theory of SPWs was delivered shortly after the pioneering work [1] by Ritchie in 1957 and has since been comprehensively discoursed in many textbooks and review articles [3,4,11,18]. In this theory, the electrical properties of a metal are prescribed with a dielectric function ϵ. To analytically treat ϵ, the simple Drude model or the slightly more involved hydrodynamic model is often invoked [4,19–24]. For either model to be valid, electronic collisions in the metal must be sufficiently frequent so that the electronic mean free path, $l_0 = v_F\tau$, where v_F is the Fermi velocity and τ the thermal charge relaxation time, is much shorter than the SPW wavelength or the typical length of the system [25–27]. The general case with arbitrary τ, especially the collision-less limit, where $\tau \to \infty$, defies these models and has yet to be entertained. Other models based on *ab initio* quantum mechanical computations [4] are helpful in understanding the complexity of real materials but fall short in providing an intuitive and systematic picture of SPWs underpinned by electrons experiencing less frequent collisions.

The purpose of this paper is to furnish a comprehensive theory for SPWs of ballistically moving electrons. Ballistic SPWs are not only interesting in themselves but could have ramified applications in plasmonics and other arenas. Recently [28,29], we considered ballistic SPWs in semi-infinite metals. We showed that such waves are intrinsically unstable and possess a universal self-amplification channel that exists irrespective of the value of τ. This result was initially established by examining the charge dynamics [28] in the system and later corroborated by an energy conversion analysis [29] in the waves. In the present work, we study ballistic SPWs in metal films, which possess two surfaces and are experimentally more realistic and interesting.

In the next section, we specify the system under consideration and state our main results. Some preliminary remarks are made on their experimental implications. In Sec. III, the theory in support of the results is systematically presented, followed by a complementary energy conversion analysis in Sec. IV. We discuss the results and conclude the paper in Sec. V. Some calculations of technical interest are displayed in the appendices A, B, and C.

II. RESULTS

A. System

We consider ballistic SPWs in a metal film surrounded by vacuum. By the so-called jellium model [26,27], the metal is described as a free electron gas embedded in a static background of homogeneously distributed positive charges. This description is valid if the length scale in question is much longer than the microscopic lattice constant and interband transitions are negligible. The kinetic energy of electrons is $\varepsilon(\mathbf{v}) = \frac{1}{2}m\mathbf{v}^2$, where m and $\mathbf{v}$ denote the mass and velocity of the electrons, respectively. The film resides in the region $0 \leqslant z \leqslant d$ with two surfaces located at $z = 0$ and $z = d$, respectively. The surfaces are treated as geometric planes of a hard wall type and they strictly prevent electrons from leaking out of the metal. To simplify our analysis, the surfaces are assumed with identical properties so that the system is symmetric about the midplane $z = d/2$. To avoid quantum size effects, we assume $d \gg \hbar/mv_F$, where $\hbar$ is the reduced Planck constant. Throughout we write $\mathbf{x} = (\mathbf{r}, z)$ and reserve $\mathbf{r} = (x, y)$ for planar components while we let t be the time. We neglect retardation effects in total [20,30].

B. Results

With two surfaces, a film possesses two branches of SPWs, which at large d degrade into those for two semi-infinite metals. Reflection symmetry about the midplane requires the corresponding charge densities to bear a definite sign under the reflection. The branch whose charge density is invariant under the reflection is called symmetric while the one whose charge density changes sign under reflection is called antisymmetric. In the literature, the symmetric and antisymmetric SPWs are also designated as *tangential* and *normal* oscillations, respectively. Profiles of the charge densities for symmetric

*h.deng@exeter.ac.uk

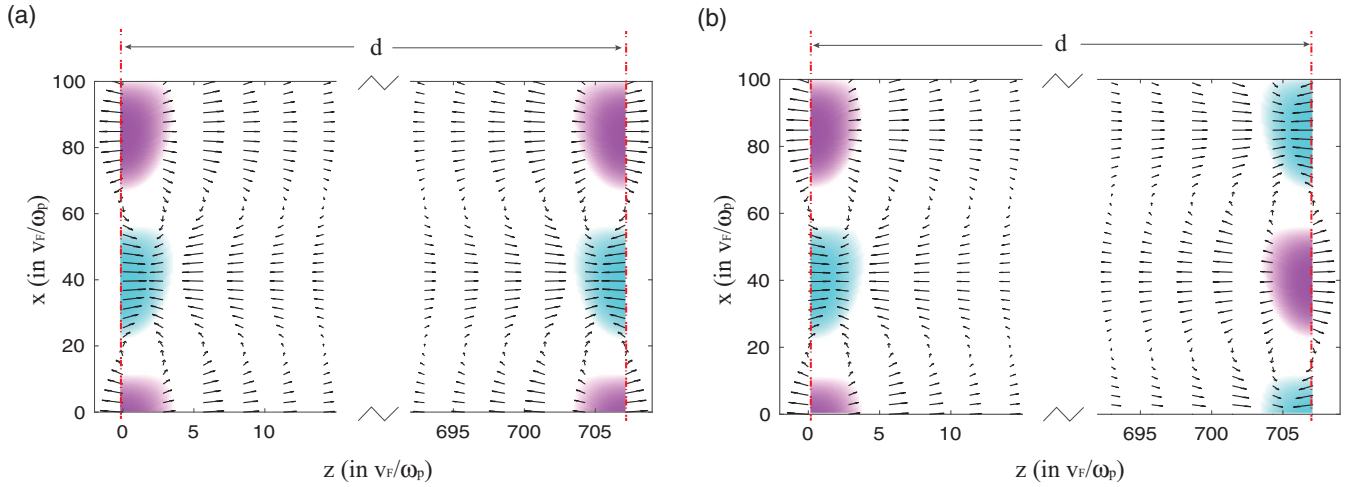

FIG. 1. Snapshots of the charge density (color) and electric field (arrows) of SPWs supported in a metal film in the region $z \in [0,d]$. $k/k_s = 0.1$, $d = 500/k_s$, and $p = 1$, with $k_s = (\omega_p/\sqrt{2})/v_F$. The symmetric mode $\rho_+(z)$ and antisymmetric mode $\rho_-(z)$ are displayed in panels (a) and (b), respectively.

and antisymmetric SPWs are mapped in Figs. 1(a) and 1(b), respectively, together with the electric field accompanying them.

We find that the SPW frequency $\omega_s^{\pm}$ is significantly (as much as 30%) higher than $\omega_{s0}^{\pm} = (\omega_p/\sqrt{2})\sqrt{1 \pm e^{-kd}}$ which would be obtained by the hydrodynamic/Drude theory. Here the plus (minus) sign is affixed and refers to symmetric (antisymmetric) modes, ω_p denotes the characteristic plasma frequency of the metal, and k is the SPW wave number. The dependences of $\omega_s^{\pm}$ on k, d, and surface scattering, the effects of which could be summarized in the *Fuchs* parameter p in the simplest possible scattering picture, are displayed in the upper panels of Figs. 2(a), 2(b), and 2(c), respectively. The great contrast between $\omega_s^{\pm}$ and $\omega_{s0}^{\pm}$ would be ideal for experimentally verifying our theory. Unfortunately, in the most commonly experimented materials, such as noble metals, due to pronounced interband transitions there is no simple relation between ω_p and $\omega_{s0}^{\pm}$.

More interestingly, we reveal a universal self-amplification channel for SPWs irrespective of their symmetry. Namely, we find that the net amplification rate of SPWs can be generally written as $\gamma^{\pm} = \gamma_0^{\pm} - \tau^{-1}$, where γ_0 is warranted to be nonnegative by a general principle and independent of τ. In the conventional theory, $\gamma_0^{\pm}$ vanishes identically and amplification would be impossible without extrinsic energy supply [31–38]. The dependences of $\gamma_0^{\pm}$ on k, d, and p are shown in the lower panels of Figs. 2(a), 2(b), and 2(c), respectively, where we observe that (1) $\gamma_0^{\pm}$ is generally a sizable fraction (as much as ~10%) of ω_p, (2) it increases as k increases, i.e., higher amplification obtains for shorter wavelengths, and (3) it increases as p increases, i.e., smooth surfaces produce higher amplification than rough surfaces. We also see that γ_0^{+} is more sensitive to film thickness than γ_0^{-}.

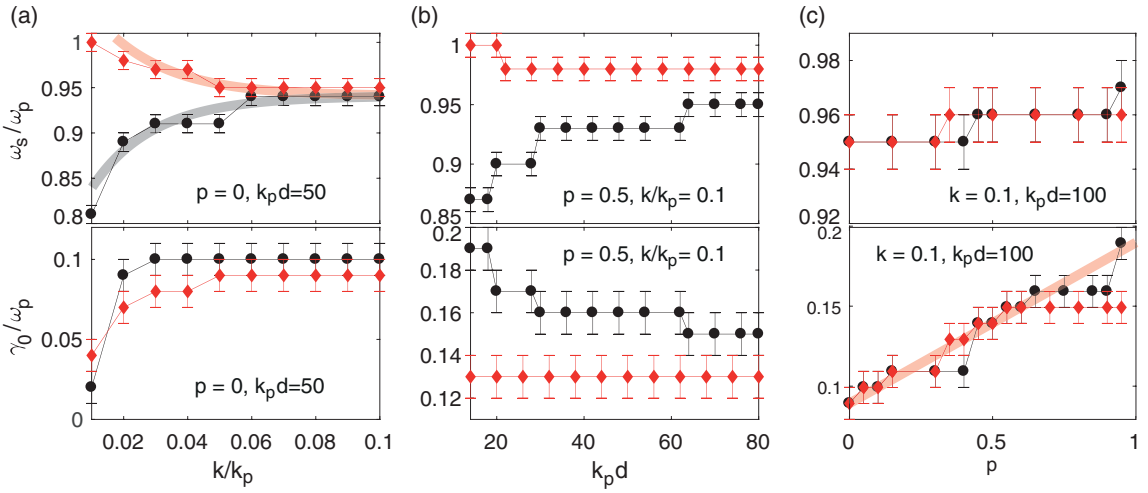

FIG. 2. Plot of the SPW frequency ω_s and self-amplification rate γ_0 versus wave number k, film thickness d, and surface roughness parameter p. Circles and diamonds are designated for symmetric and antisymmetric modes, respectively. ω_s and γ_0 are obtained by numerically solving Eq. (39), with Landau damping automatically included. $k_p = \omega_p/v_F$. The cutoff $q_c = 1.5k_p$ has been used. The error bar is ± 0.01, corresponding to the grid resolution of $\bar{\omega}$ in the complex frequency plane used in our numerical method. In the upper panel of (a), the thick lines are given by $\propto \sqrt{1-(1 \mp e^{-kd})(1+p)/4}$ with $p = 0$. In the lower panel of (c), the thick line is $\sim 0.1 \times (1+p)$.

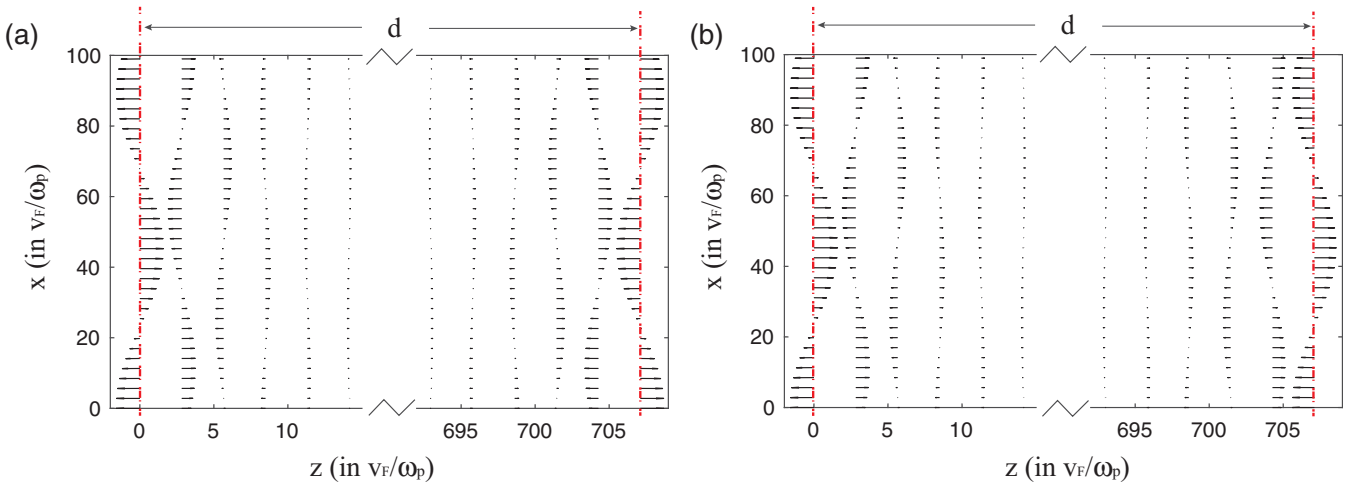

FIG. 3. Snapshots of the surface-ballistic current density for (a) symmetric and (b) antisymmetric modes, with the same parameters as in Fig. 2. Note that the currents are directed largely normal to the surface, unlike diffusive currents, which are largely normal to the local electric field.

Additionally, we show that the electrical current density in the system can be split into two disparate components, which we call $\mathbf{J}_D$ and $\mathbf{J}_B$, respectively. An example of their profiles is exhibited in Figs. 3(a) and 3(b), respectively, for the symmetric and antisymmetric modes. What critically sets them apart rests with their distinct relations with the electric field $\mathbf{E}$ present in the system. $\mathbf{J}_D$ responds to $\mathbf{E}$ as if the system had no surfaces and is therefore primarily a bulk property. As such, it can also be satisfactorily captured by the hydrodynamic/Drude model. For this reason, we designate it a diffusive component, regardless of the value of τ. On the contrary, $\mathbf{J}_B$ represents genuine surface effects and would totally disappear were the surfaces absent. In particular, it synthesizes the effects ensuing from the fact that the system is not translationally invariant along the direction normal to the surfaces. These effects are completely beyond the hydrodynamic/Drude model but well within the scope of Boltzmann's approach, which is employed in our theory to be expounded in the next section. We thus designate $\mathbf{J}_B$ as a surface-ballistic component.

Finally, we find that the self-amplification channel is a direct consequence of $\mathbf{J}_B$. Indeed, were it not for $\mathbf{J}_B$, SPWs would behave in accord with the hydrodynamic/Drude model. This is already clear from the orientations of $\mathbf{J}_{D/B}$ relative to $\mathbf{E}$. As seen in Fig. 3, $\mathbf{J}_D$ points at right angles with $\mathbf{E}$ almost locally, whereas $\mathbf{J}_B$ flows normal to the surface paying little regard to $\mathbf{E}$. Therefore, $\mathbf{E}$ does no work on $\mathbf{J}_D$ on average while, as shown in Sec. IV, it does a negative amount of work on $\mathbf{J}_B$, thereby imparting energy from the electrons to SPWs and destabilizing the Fermi sea.

C. Remarks

Experimentally verifying the self-amplification channel and the theory in general would be of considerable interest, as it would drastically change the way we conceive and utilize SPWs and renew our interest in surface science in a broad sense. The self-amplification channel could manifest itself for instance in the temperature dependence of various spectra, e.g., electron loss spectra. We discuss this aspect in Sec. V. Here we mainly concern ourselves with the experimental implications of the surface-ballistic current $\mathbf{J}_B$.

Being an integral part of the electrical responses of metals, $\mathbf{J}_B$ is expected to play a role in virtually every phenomena where surface is not negligible. Examples include electron energy losses, reflectance, and van der Waals forces. Unlike $\mathbf{J}_D$, which does not reflect surface scattering effects, $\mathbf{J}_B$ is surface specific via the *Fuchs* parameter. Moreover, they differ in phase by $\sim\pi/2$. To be specific, let us consider the Ferrel radiation [2]. Ferrel predicted that antisymmetric SPWs in thin films would radiate in a characteristic pattern. Some experiments even claimed to have observed this radiation [39–41]. Ferrel considered only $\mathbf{J}_D$. Following him, we find that including $\mathbf{J}_B$ could boost the radiation power by a factor $\sim 1+(3/2\pi)^2(1+2p)^2$. Though a crude estimate, it does imply that surface properties could be utilized to tune the radiation. In this paper, we focus on the fundamental theory of ballistic SPWs. A systematic treatment of Ferrel radiation will be published elsewhere.

As aforementioned, a major obstacle in experimentally studying the theory lies with interband transitions, which have been neglected in our theory. A detailed discussion of their effects is presented in Sec. V.

III. THEORY

This section is devoted to a thorough exposition of the theory. We begin with a discussion of the equation of continuity in the presence of surfaces. Thence we proceed to Boltzmann's approach and analyze how to handle surface effects in this approach. The electronic distribution functions, obtained by solving Boltzmann's equation, are discussed in detail. The electrical current densities are then calculated and the exact equation of motion for the charge density is established. Solutions to the equation are discussed and the properties of SPWs are analyzed. Various limits are presented and connections are made with the hydrodynamic/Drude models.

A. Equation of continuity

The starting point of our theory is the equation of continuity, $(\partial_t + 1/\tau)\rho(\mathbf{x},t) + \partial_{\mathbf{x}} \cdot \mathbf{j}(\mathbf{x},t) = 0$, which relates the charge density $\rho(\mathbf{x},t)$ and the current density $\mathbf{j}(\mathbf{x},t)$ in a universal manner. Here $\mathbf{j}(\mathbf{x},t)$ arises in the presence of an electric field $\mathbf{E}(\mathbf{x},t)$ and the damping term $-\rho(\mathbf{x},t)/\tau$ is included to account for the thermal currents due to electronic collisions that would drive the system toward thermodynamic equilibrium. In the jellium model, $\rho(\mathbf{x},t)$ appears when the electron density is perturbed away from its equilibrium value n_0.

As the surfaces strictly prevent electrons from escaping the metal, we may write $\mathbf{j}(\mathbf{x},t) = [\Theta(z) - \Theta(z-d)]\mathbf{J}(\mathbf{x},t)$, where $\Theta(z)$ is the Heaviside step function. In doing this, we have embodied the surfaces as hard walls and considered the fact that $\mathbf{J}(\mathbf{x},t)$ may not vanish even in the immediate neighborhood of the surfaces—as is obviously the case with the Drude model. With this prescription, the equation of continuity can be rewritten

$$\left(\partial_t + \frac{1}{\tau}\right)\rho(\mathbf{x},t) + \partial_{\mathbf{x}} \cdot \mathbf{J}(\mathbf{x},t) = S(\mathbf{x},t), \tag{1}$$

where the effective source term

$$S(\mathbf{x},t) = J_z(\mathbf{x}_d,t)\delta(z-d) - J_z(\mathbf{x}_0,t)\delta(z), \tag{2}$$

results directly from the presence of the surfaces. Here $\mathbf{x}_0 = (\mathbf{r},0)$ and $\mathbf{x}_d = (\mathbf{r},d)$ denote points on the surface at $z = 0$ and those on that at $z = d$, respectively. Physically, $S(\mathbf{x},t)$ corresponds to the scenario that charges must pile up on the surfaces if they do not come to a halt before they reach them.

Without loss of generality we seek fields in this form: $\rho(\mathbf{x},t) = \mathrm{Re}\,[\rho(z)e^{i(kx-\omega t)}]$ and $\mathbf{J}(\mathbf{x},t) = \mathrm{Re}\,[\mathbf{J}(z)e^{i(kx-\omega t)}]$. Similarly, for the electric field $\mathbf{E}(\mathbf{x},t) = \mathrm{Re}\,[\mathbf{E}(z)e^{i(kx-\omega t)}]$ and the electrostatic potential $\phi(\mathbf{x},t) = \mathrm{Re}\,[\phi(z)e^{i(kx-\omega t)}]$. In these expressions, Re/Im takes the real/imaginary part of a quantity, $k \geqslant 0$ is a wave number, and ω is the eigenfrequency to be determined. Equation (1) becomes

$$-i\bar{\omega}\rho(z) + \nabla \cdot \mathbf{J}(z) = S(z), \tag{3}$$

where $\bar{\omega} = \omega + i/\tau$, $\nabla = (ik, \partial_y, \partial_z)$, and

$$S(z) = J_z(d)\delta(z-d) - J_z(0)\delta(z). \tag{4}$$

Equation (3) will serve as the equation of motion for $\rho(z)$ when supplemented with additional relations to be formulated between $\mathbf{J}(z)$ and $\rho(z)$ in what follows.

B. The law of electrostatics

If the SPW phase velocity is much smaller than the speed of light c in vacuum, i.e., $k > k_0$, where k_0 is the wave number of light at the SPW frequency, the system will be in the nonretarded regime [30], and we can relate $\phi(\mathbf{x},t)$ and $\rho(\mathbf{x},t)$ by the laws of electrostatics. Without external charges, we have [30]

$$\phi(z) = \frac{2\pi}{k}\int dz' e^{-k|z-z'|}\rho(z').$$

Instead of $\rho(z)$, we directly work with its Fourier components. Generically, we may write

$$\rho(z) = \sum_{n=0}^{\infty} \rho_n \cos(q_n z), \quad q_n = \frac{\pi n}{d}.$$

The components are given by

$$\rho_n = \frac{1}{d_n}\int_0^d dz\rho(z)\cos(q_n z), \quad d_n = \frac{d}{2-\delta_{n,0}}, \tag{5}$$

where $\delta_{m,n}$ denotes the Kroneker symbol.

As the surfaces of the film are assumed identical, the system is invariant under reflection about its midplane. This symmetry makes it useful to write $\rho(z)$ as a superposition of a symmetric mode $\rho_+(z)$ and an antisymmetric mode $\rho_-(z)$. Namely,

$$\rho(z) = \rho_+(z) + \rho_-(z),$$

where $\rho_+(z)$ includes all the terms with even n whereas $\rho_-(z)$ those with odd n. As such, $\rho_+(0) = \rho_+(d)$ and $\rho_-(0) = -\rho_-(d)$. Due to the symmetry $\rho_+(z)$ and $\rho_-(z)$ will be shown to be strictly decoupled. We impose on q_n a cutoff q_c of the order of a reciprocal lattice constant; otherwise, the jellium model would cease to be valid. Obviously, $q_c \sim k_F$, where k_F is the Fermi wave number of the electrons in the metal.

In terms of ρ_n, we can rewrite

$$\phi(z) = \sum_{n=0}^{\infty} \frac{2\pi\rho_n}{k^2+q_n^2}[2\cos(q_n z) - e^{-kz} - (-1)^n e^{-k(d-z)}]. \tag{6}$$

The electric field, $\mathbf{E}(z) = -\nabla\phi(z)$, can then be obtained straightforwardly. In equation (6) the exponentials, e^{-kz} and $e^{-k(d-z)}$, would all vanish if the surfaces were sent to infinity. We may then write $\phi(z) = \phi^{bulk}(z) + \phi^{surface}(z)$, where $\phi^{surface}(z)$ includes the contributions from all the exponentials while $\phi^{bulk}(z)$ contains the remaining contributions. Accordingly, $\mathbf{E}(z) = \mathbf{E}^{bulk}(z) + \mathbf{E}^{surface}(z)$. Such a partition proves useful in analyzing surface specific effects.

C. Electronic distribution function

The electric field $\mathbf{E}(\mathbf{x},t)$ drives an electrical current $\mathbf{J}(\mathbf{x},t)$. We employ Boltzmann's equation, which is valid as long as interband transitions are negligible, to calculate this current. Including the transitions in our formalism is straightforward but will be skipped here. Surfaces scatter electrons. On the microscopic level, one can in principle introduce a surface potential $\phi_s(\mathbf{x})$ in Boltzmann's equation to produce such scattering. The corresponding surface field $\mathbf{E}_s(\mathbf{x}) = -\partial_{\mathbf{x}}\phi_s(\mathbf{x})$ should be peaked on the surfaces and may have an infinitesimal spread complying with the hard-wall picture of surfaces. However, as $\phi_s(\mathbf{x})$ can hardly be known and varies from one sample to another, this method is impractical and futile.

Alternatively surface scattering effects can be dealt with using boundary conditions. This is possible because $\mathbf{E}_s(\mathbf{x})$ acts only within the immediate neighborhoods of the surfaces. In the bulk of the sample, the electronic distribution function $f(\mathbf{x},\mathbf{v},t)$ sought as solutions to Boltzmann's equation can be specified up to some parameters, which summarize the effects of—while without actually knowing—$\phi_s(\mathbf{x})$. With translational symmetry along the surfaces, only one such

parameter, i.e., the so-called *Fuchs* parameter p, is needed in the simplest model. Physically, p measures the probability that an electron is bounced back when impinging upon the surface. We write $f(\mathbf{x},\mathbf{v},t) = f_0(\varepsilon(\mathbf{v})) + g(\mathbf{x},\mathbf{v},t)$, where $f_0(\varepsilon)$ denotes the Fermi-Dirac distribution and $g(\mathbf{x},\mathbf{v},t)$ represents the nonequilibrium part due to the presence of $\mathbf{E}(\mathbf{x},t)$. The current density can then be calculated by $\mathbf{J}(\mathbf{x},t) = (m/2\pi\hbar)^2 \int d^3\mathbf{v} e\mathbf{v} g(\mathbf{x},\mathbf{v},t)$, where e denotes the charge of an electron. It is worth pointing out that, as $g(\mathbf{x},\mathbf{v},t)$ is a distribution for the bulk, the actual charge density is not given by $\tilde{\rho}(\mathbf{x},t) = (m/2\pi\hbar)^2 \int d^3\mathbf{v} e g(\mathbf{x},\mathbf{v},t)$, i.e., $\rho(\mathbf{x},t) \neq \tilde{\rho}(\mathbf{x},t)$. Actually, $\tilde{\rho}(\mathbf{x},t)$ satisfies $(\partial_t + 1/\tau)\tilde{\rho}(\mathbf{x},t) + \partial_{\mathbf{x}} \cdot \mathbf{J}(\mathbf{x},t) = 0$ rather than Eq. (1). By comparison, one sees that what is missing from $\tilde{\rho}(\mathbf{x},t)$ is the charges localized on the surface.

As before we write $g(\mathbf{x},\mathbf{v},t) = \mathrm{Re}\,[g(\mathbf{v},z)e^{i(kx-\omega t)}]$. For linear responses, Boltzmann's equation can be written

$$\frac{\partial g(\mathbf{v},z)}{\partial z} + \lambda^{-1} g(\mathbf{v},z) + e f_0'(\varepsilon)\frac{\mathbf{v}\cdot\mathbf{E}(z)}{v_z} = 0, \tag{7}$$

where $\lambda = i v_z/\tilde{\omega}$ with $\tilde{\omega} = \bar{\omega} - k v_x$ and $f_0'(\varepsilon) = \partial f_0/\partial\varepsilon(\mathbf{v})$. In this equation, the velocity $\mathbf{v}$ is more of a parameter than an argument and can be used to tag electron beams. It is straightforward to solve the equation under appropriate boundary conditions (see Appendix B). We divide $g(\mathbf{v},z)$ into a bulk and a surface term, i.e.,

$$g(\mathbf{v},z) = g_{bulk}(\mathbf{v},z) + g_{surface}(\mathbf{v},z),$$

where the bulk term would exist even in the absence of surfaces whereas the surface term would not. Using Eq. (6) for $\mathbf{E}(z)$, we obtain

$$g_{bulk}(\mathbf{v},z) = -e f_0' \sum_{n=-\infty}^{\infty} \frac{2\pi\rho_n}{k^2+q_n^2}\frac{k v_x + q_n v_z}{\bar{\omega} - (k v_x + q_n v_z)} e^{i q_n z}, \tag{8}$$

where we have defined $\rho_{n<0} := \rho_{-n}$. For large d equation (8) converges to the distribution function of a boundless system for either the symmetric mode or the antisymmetric mode. It is notable that $g_{bulk}(\mathbf{v},z)$ bears a single form for all electrons regardless of their velocities.

As for $g_{surface}(\mathbf{v},z)$, we find it with a subtle structure: It can be written as a sum of two contributions, one of which, $g_{D,surface}(\mathbf{v},z)$, has a single form for all electrons irrespective of their velocities while the other, $g_{B,surface}(\mathbf{v},z)$, does not. Explicitly, we find

$$g_{D,surface}(\mathbf{v},z) = g^{(1)}_{D,surface}(\mathbf{v},z) + g^{(2)}_{D,surface}(\mathbf{v},z),$$

where

$$g^{(1)}_{D,surface}(\mathbf{v},z) = -e f_0' \sum_{n=0}^{\infty} \frac{2\pi\rho_n}{k^2+q_n^2}\frac{k(v_z - i v_x)}{k v_z + i\tilde{\omega}} e^{-kz}, \tag{9}$$

and

$$g^{(2)}_{D,surface}(\mathbf{v},z) = -e f_0' \sum_{n=0}^{\infty} \frac{2\pi\rho_n (-1)^n}{k^2+q_n^2}\frac{k(v_z + i v_x)}{k v_z - i\tilde{\omega}} e^{-k(d-z)} \tag{10}$$

originate from the surfaces at $z = 0$ and $z = d$, respectively.

We may combine $g_{bulk}(\mathbf{v})$ and $g_{D,surface}(\mathbf{v},z)$ in a single term,

$$g_D(\mathbf{v},z) = g_{bulk}(\mathbf{v},z) + g_{D,surface}(\mathbf{v},z),$$

in order to separate them from

$$g_B(\mathbf{v},z) := g_{B,surface}(\mathbf{v},z).$$

The subscripts, D and B, refer to 'diffusive' and 'surface-ballistic,' respectively. In so doing, we have decomposed

$$g(\mathbf{v},z) = g_D(\mathbf{v},z) + g_B(\mathbf{v},z)$$

in a diffusive and a surface-ballistic component. It is underlined that $g_B(\mathbf{v},z)$ arises only when the surfaces are present. For boundless systems without surfaces, it does not exist even if the electronic motions are totally ballistic, i.e., $\tau \to \infty$. In other words, $g_B(\mathbf{v},z)$ represents genuine surface effects. It may be interpreted as a contribution from electrons which experience the electric field only on the surfaces and propagate freely in the body. Its expressions are given in what follows.

Electrons in the film can bounce back and forth between its surfaces. Each bounce gives a factor $p_1 p_2 e^{2i\tilde{\omega}d/|v_z|}$, whose magnitude is generally smaller than unity (see Appendix B). Here p_1 and p_2 are the *Fuchs* parameters for the surfaces at $z = 0$ and $z = d$, respectively. Consequently, we neglect multiple bounces, which allows us to write

$$g_{B,surface}(\mathbf{v},z) = g^{(1)}_{B,surface}(\mathbf{v},z) + g^{(2)}_{B,surface}(\mathbf{v},z),$$

where $g^{(1)}_{B,surface}(\mathbf{v},z)$ and $g^{(2)}_{B,surface}(\mathbf{v},z)$ originate from the surfaces at $z = 0$ and $z = d$, respectively. They are given by

$$g^{(1)}_{B,surface}(\mathbf{v},z) = \Theta(v_z) e^{i\frac{\tilde{\omega}z}{v_z}}\left[g^{(1)}_{B,emg}(\mathbf{v}) + p_1 g^{(1)}_{B,ref}(\mathbf{v})\right],$$
$$g^{(2)}_{B,surface}(\mathbf{v},z) = \Theta(-v_z) e^{i\frac{\tilde{\omega}(z-d)}{v_z}}\left[g^{(2)}_{B,emg}(\mathbf{v}) + p_2 g^{(2)}_{B,ref}(\mathbf{v})\right],$$

where $g^{(1/2)}_{B,emg}(\mathbf{v},z)$ is contributed by electrons that directly emerge from the surface at $z = 0/d$ while $g^{(1/2)}_{B,ref}(\mathbf{v},z)$ by reflected electrons and hence proportional to $p_{1/2}$. In what follows we take $p_1 = p_2 = p$. The expressions of $g^{(1/2)}_{B,emg/ref}(\mathbf{v})$ are involved but with a recognizable structure:

$$g^{(1/2)}_{B,emg}(\mathbf{v}) = e f_0' \sum_{n=0}^{\infty} \frac{2\pi\rho_n \langle 1|(-1)^n\rangle}{k^2+q_n^2}\left[\frac{2(\tilde{\omega}k v_x) + q_n^2 v_z^2}{\tilde{\omega}^2 - q_n^2 v_z^2} + \frac{k(v_z \mp i v_x)}{k v_z \pm i\tilde{\omega}} + (-1)^n e^{-kd}\frac{k(v_z \pm i v_x)}{k v_z \mp i\tilde{\omega}}\right], \tag{11}$$

where the symbol $\langle 1|(-1)^n\rangle$ returns 1 and $(-1)^n$ for $g^{(1)}_{B,emg}$ and $g^{(2)}_{B,emg}$, respectively. In addition, we have

$$g^{(1)}_{B,ref}(\mathbf{v}) = e f_0' \sum_{n=0}^{\infty} \frac{2\pi\rho_n}{k^2+q_n^2}\left[\left(e^{i\frac{\tilde{\omega}d}{v_z}}(-1)^n - 1\right)\frac{2(\tilde{\omega}k v_x - q_n^2 v_z^2)}{\tilde{\omega}^2 - q_n^2 v_z^2} + \left(1 - e^{\left(\frac{i\tilde{\omega}}{v_z} - k\right)d}\right)\frac{k(v_z - i v_x)}{k v_z - i\tilde{\omega}} + (-1)^n\left(e^{-kd} - e^{i\frac{\tilde{\omega}d}{v_z}}\right)\frac{k(v_z + i v_x)}{k v_z + i\tilde{\omega}}\right], \tag{12}$$

and

$$g^{(2)}_{B,ref}(\mathbf{v}) = ef_0' \sum_{n=0}^{\infty} \frac{2\pi\rho_n(-1)^n}{k^2+q_n^2}\Big[\big(e^{-i\frac{\tilde{\omega}d}{v_z}}(-1)^n - 1\big)\frac{2\big(\tilde{\omega}kv_x - q_n^2v_z^2\big)}{\tilde{\omega}^2 - q_n^2v_z^2} + \big(1 - e^{-\left(\frac{i\tilde{\omega}}{v_z}+k\right)d}\big)\frac{k(v_z + iv_x)}{kv_z + i\tilde{\omega}}$$
$$+ (-1)^n\big(e^{-kd} - e^{-i\frac{\tilde{\omega}d}{v_z}}\big)\frac{k(v_z - iv_x)}{kv_z - i\tilde{\omega}}\Big], \tag{13}$$

1. *Positiveness of Im($\bar{\omega}$)*

What sets $g_B(\mathbf{v},z)$ apart from its diffusive counterpart rests with its disparate z dependence. Let us take the contribution originating from the surface at $z=0$ for example. Here $g^{(1)}_{B,surface}(\mathbf{v},z) \propto e^{i\tilde{\omega}z/v_z} \propto e^{-\mathrm{Im}(\bar{\omega})z/v_z}$, where $v_z \geqslant 0$. Unless $\mathrm{Im}(\bar{\omega}) \geqslant 0$, this expression would diverge for small v_z. As such, we may conclude that $\mathrm{Im}(\bar{\omega}) \geqslant 0$, a result to be confirmed in what follows by specific calculations. In Appendix B, we frame this result as a consequence of the causality principle: Out-going electrons are determined by in-coming ones, not otherwise.

D. Current densities

We are now prepared to discuss the behaviors of the current density, which is written $\mathbf{J}(z) = \mathbf{J}_D(z) + \mathbf{J}_B(z)$, where

$$\mathbf{J}_{D/B}(z) = (m/2\pi\hbar)^3 \int d^3\mathbf{v} e\mathbf{v} g_{D/B}(\mathbf{v},z)$$

is the diffusive/surface-ballistic component of $\mathbf{J}(z)$. The equation of motion for $\rho(z)$ follows upon inserting $\mathbf{J}(z)$ in Eq. (3). In our calculations, the zero temperature is assumed whenever a concrete form of $f_0(\varepsilon)$ is required, though generalization to finite temperatures is straightforward.

1. *Diffusive current density*

Since $g_D(\mathbf{v},z)$ consists of a bulk and a surface component, we accordingly write $\mathbf{J}_D(z) = \mathbf{J}_{bulk}(z) + \mathbf{J}_{D,surface}(z)$, where $\mathbf{J}_{bulk}(z)$ and $\mathbf{J}_{D,surface}(z)$ arise from $g_{bulk}(\mathbf{v},z)$ and $g_{D,surface}(\mathbf{v},z)$, respectively. By straightforward manipulation, one may show that $\mathbf{J}_{D,surface}(z) \propto \mathbf{E}^{surface}(z)$. To the lowest order in kv_F/ω_p, where $\omega_p = \sqrt{4\pi n_0 e^2/m}$ is the characteristic plasma frequency of the metal, we have

$$\mathbf{J}_{D,surface} = \frac{i}{\bar{\omega}}\frac{\omega_p^2}{4\pi}\mathbf{E}^{surface}(z), \tag{14}$$

where the prefactor heading $\mathbf{E}^{surface}(z)$ is recognized as the Drude conductivity. In addition, we find

$$\mathbf{J}_{bulk}(z) = \frac{i}{\bar{\omega}}\frac{\omega_p^2}{4\pi}\mathbf{E}^{bulk}(z) + \mathbf{J}'(z), \tag{15}$$

where

$$\mathbf{J}'(z) = \sum_{n=-\infty}^{\infty} \frac{2\pi\rho_n e^{iq_nz}}{k^2+q_n^2}\mathbf{F}(k,q_n;\bar{\omega}) \tag{16}$$

signifies nonlocal electrical responses that would engender dispersive plasma waves. In the expression

$$\mathbf{F}(k,q;\bar{\omega}) = \left(\frac{m}{2\pi\hbar}\right)^3 \int d^3\mathbf{v}(-e^2f_0')\mathbf{v}\sum_{l=2}^{\infty}\left(\frac{kv_x + qv_z}{\bar{\omega}}\right)^l, \tag{17}$$

only terms with odd l contribute in the series. Note that the normal component of $\mathbf{J}'(z)$ vanishes identically at all surfaces, i.e., $J_z'(0) = J_z'(d) = 0$.

Piecing everything together we obtain

$$\mathbf{J}_D(z) = \frac{i}{\bar{\omega}}\frac{\omega_p^2}{4\pi}\mathbf{E}(z) + \mathbf{J}'(z).$$

As in the hydrodynamic/Drude model, which is valid only for diffusive electronic motions, the relation between $\mathbf{J}_D(z)$ and $\mathbf{E}(z)$ assumes the form of a generalized Ohm's law. This is why we consider $\mathbf{J}_D(z)$ a diffusive component, irrespective of the value of τ. Its divergence is easily found to be

$$\nabla\cdot\mathbf{J}_D(z) = \frac{i}{\bar{\omega}}\sum_{n=0}^{\infty}\Omega^2(k,q_n;\bar{\omega})\rho_n\cos(q_nz), \tag{18}$$

where, with $\mathbf{k} := (k,q)$,

$$\Omega^2(k,q;\bar{\omega}) = \omega_p^2 + \frac{4\pi\bar{\omega}\mathbf{k}\cdot\mathbf{F}(k,q;\bar{\omega})}{\mathbf{k}\cdot\mathbf{k}}. \tag{19}$$

Fourier transforming Eq. (18) yields

$$\frac{1}{d_n}\int_0^d dz\cos(q_nz)\nabla\cdot\mathbf{J}_D(z) = \frac{i}{\bar{\omega}}\Omega^2(k,q_n;\bar{\omega})\rho_n. \tag{20}$$

We will show that $\Omega(k,q;\bar{\omega})$ is intimately related to the properties of bulk plasma waves. As expected, $\Omega(k,q;\bar{\omega})$ only depends on the length of $\mathbf{k}$, not its direction. This becomes evident by writing $kv_x + qv_z = \mathbf{k}\cdot\mathbf{v}$ in Eq. (17). The first nonvanishing contribution to $\Omega(k,q;\bar{\omega})$ comes from the term $l=1$ in the series in $\mathbf{F}(k,q;\bar{\omega})$. Retaining only this term, we get

$$\Omega^2(k,q;\bar{\omega}) \approx \omega_p^2\left[1 + \frac{3}{5}\frac{(k^2+q^2)v_F^2}{\bar{\omega}^2}\right]. \tag{21}$$

Upon replacing $\bar{\omega}$ with ω_p, one immediately revisits the dispersion relation for bulk waves, which could also be reached through the hydrodynamic model. In the Drude model, the dispersion is totally neglected.

It is noted that $\Omega(k,q;\bar{\omega})$ generally possesses an imaginary part. In case $\mathrm{Im}(\bar{\omega})$ is vanishingly small, the imaginary part arises from a pole, located at $\bar{\omega} = kv_x + qv_z$, in the integrand in Eq. (17), giving rise to Landau damping in bulk waves and SPWs. In our numerical computation of $\bar{\omega}$, Landau damping will be automatically included.

2. Surface-ballistic current density

Separating the contributions of emerging electrons from that of reflected electrons, we write $\mathbf{J}_B(z) = \mathbf{J}_{B,emg}(z) + p\mathbf{J}_{B,ref}(z)$. Explicitly, we find

$$\begin{aligned}\mathbf{J}_{B,emg/ref}(z) &= \left(\frac{m}{2\pi\hbar}\right)^3 \int d^3\mathbf{v} e\mathbf{v}[\Theta(v_z)e^{i\frac{\tilde{\omega}z}{v_z}} g^{(1)}_{B,emg/ref}(\mathbf{v}) + \Theta(-v_z)e^{i\frac{\tilde{\omega}(z-d)}{v_z}} g^{(2)}_{B,emg/ref}(\mathbf{v})] \\ &=: \left(\frac{m}{2\pi\hbar}\right)^3 \int d^3\mathbf{v}\mathbf{J}_{B,emg/ref}(\mathbf{v},z), \end{aligned} \tag{22}$$

where we have defined $\mathbf{J}_{B,emg/ref}(\mathbf{v},z)$ as the contribution from the beam of electrons with velocity $\mathbf{v}$. Using the expressions of $g^{(1/2)}_{B,emg/ref}(\mathbf{v})$ given by f [Eqs. (11)–(13)], we can rewrite it

$$J_{B,emg/ref,x|z}(\mathbf{v},z) = \Theta(v_z)e^2 f_0' \sum_{n=0}^{\infty} \frac{2\pi\rho_n}{k^2+q_n^2} L_{emg/ref}\left(v_x,v_z,k,q_n,\bar{\omega},(-1)^n\right) v_{x|z}(e^{i\frac{\tilde{\omega}z}{v_z}} + (+|-)(-1)^n e^{i\frac{\tilde{\omega}(d-z)}{v_z}}), \tag{23}$$

where, with $s = \pm 1$,

$$L_{emg}(v_x,v_z,k,q,\bar{\omega},s) = \frac{2\left(q^2v_z^2 + \tilde{\omega}kv_x\right)}{\tilde{\omega}^2 - q^2v_z^2} + \frac{k(v_z - iv_x)}{kv_z + i\tilde{\omega}} + se^{-kd}\frac{k(v_z+iv_x)}{kv_z - i\tilde{\omega}}, \tag{24}$$

$$L_{ref}(v_x,v_z,k,q,\bar{\omega},s) = \frac{2\left(q^2v_z^2 - \tilde{\omega}kv_x\right)}{\tilde{\omega}^2 - q^2v_z^2}(1 - se^{i\frac{\tilde{\omega}d}{v_z}}) + \frac{k(v_z-iv_x)}{kv_z - i\tilde{\omega}}(1 - e^{(i\frac{\tilde{\omega}}{v_z}-k)d}) + se^{-kd}\frac{k(v_z+iv_x)}{kv_z+i\tilde{\omega}}(1 - e^{(i\frac{\tilde{\omega}}{v_z}+k)d}). \tag{25}$$

In the limit $d \to \infty$, all the exponentials in $L_{emg/ref}$ vanish and we would recover the result for semi-infinite metals; $\mathbf{J}_{B,emg/ref}(z)$ could then be written as a sum of that for two semi-infinite metals. As expected, the surfaces of the film are decoupled in this limit. For thin films, Eq. (23) implies that $\mathbf{J}_{B,emg/ref}(\mathbf{v},z)$ mainly runs along the surface for symmetric modes while normal to it for antisymmetric modes.

The divergence of $\mathbf{J}_B(z)$ can be easily obtained. In the first place we have

$$\nabla\cdot\mathbf{J}_{B,emg/ref}(z) = i\bar{\omega}\left(\frac{m}{2\pi\hbar}\right)^3 \int d^3\mathbf{v}\Theta(v_z)e^2 f_0' \sum_{n=0}^{\infty} \frac{2\pi\rho_n}{k^2+q_n^2} L_{emg/ref}(v_x,v_z,k,q_n,\bar{\omega},(-1)^n)\left(e^{i\frac{\tilde{\omega}z}{v_z}} + (-1)^n e^{i\frac{\tilde{\omega}(d-z)}{v_z}}\right), \tag{26}$$

whose Fourier transform is

$$\frac{1}{d_m}\int_0^d dz\cos(q_mz)\nabla\cdot\mathbf{J}_{B,emg/ref}(z) = \frac{i}{\bar{\omega}}\sum_{n=0}^{\infty}\mathcal{M}_{emg/ref,mn}\rho_n, \tag{27}$$

with

$$\mathcal{M}_{emg/ref,mn} = \frac{\Gamma_{mn}}{d_m}\frac{2\pi\bar{\omega}^2}{k^2+q_n^2}\left(\frac{m}{2\pi\hbar}\right)^3 \int d^3\mathbf{v}\Theta(v_z)e^2f_0'\frac{i\tilde{\omega}v_z}{\tilde{\omega}^2 - q_m^2v_z^2}\left(1 - (-1)^n e^{i\frac{\tilde{\omega}d}{v_z}}\right) L_{emg/ref}(v_x,v_z,k,q_n,\bar{\omega},(-1)^n). \tag{28}$$

Here $\Gamma_{mn} = 1 + (-1)^{m+n}$, which would vanish identically unless m and n have the same parity. It follows that

$$\frac{1}{d_m}\int_0^d dz\cos(q_mz)\nabla\cdot\mathbf{J}_B(z) = \frac{i}{\bar{\omega}}\sum_{n=0}^{\infty}\mathcal{M}_{mn}\rho_n, \quad \mathcal{M}_{mn} = \mathcal{M}_{emg,mn} + p\mathcal{M}_{ref,mn}. \tag{29}$$

We can write $\mathcal{M} = \mathcal{M}^+ \bigoplus \mathcal{M}^-$, where $\mathcal{M}^{\pm} = \mathcal{M}^{\pm}_{emg} + p\mathcal{M}^{\pm}_{ref}$ operates on the space of $\rho_{\pm}(z)$, with

$$\mathcal{M}^{\pm}_{emg/ref,mn} = \frac{1}{d_m}\frac{4\pi\bar{\omega}^2}{k^2+q_n^2}\left(\frac{m}{2\pi\hbar}\right)^3 \int d^3\mathbf{v}\Theta(v_z)e^2f_0'\frac{i\tilde{\omega}v_z}{\tilde{\omega}^2 - q_m^2v_z^2}\left(1 \mp e^{i\frac{\tilde{\omega}d}{v_z}}\right) L_{emg/ref}(v_x,v_z,k,q_n,\bar{\omega},\pm 1). \tag{30}$$

In Appendix C, we show that $\mathcal{M}^{\pm}$ is of the order of kv_F/ω_p.

E. Equation of motion and SPW solutions

1. Symmetric and antisymmetric modes

We proceed to transform Eq. (3) into the equation of motion for $\rho(z)$. In the first place let us show that $\rho_+(z)$ and $\rho_-(z)$ are strictly decoupled. As is clear from preceding subsections, $\nabla\cdot\mathbf{J}(z)$ and hence the entire left hand side of Eq. (3) are block diagonal with respect to the subspaces, respectively, spanned by $\rho_+(z)$ and $\rho_-(z)$. We can prove that $S(z)$ disconnects the subspaces as well. To this end, we Fourier transform $S(z)$ in Eq. (4) to obtain

$$S_m = \frac{1}{d_m}\int_0^d dz\cos(q_mz)S(z) = \frac{1}{d_m}[J_z(d)(-1)^m - J_z(0)].$$

Linearly depending on $\rho(z) = \rho_+(z) + \rho_-(z)$, $J_z(z)$ can be split as $J_z(z) = J_z^+(z) + J_z^-(z)$, where $J_z^{+/-}(z)$ denotes the contributions from $\rho_{+/-}(z)$. From their expressions given in

preceding sessions, we easily deduce that

$$J_z^{+/-}(0) \pm J_z^{+/-}(d) = 0, \tag{31}$$

by which we rewrite

$$S_m = -\frac{1}{d_m}[J_z^+(0)(1+(-1)^m) + J_z^-(0)(1-(-1)^m)]. \tag{32}$$

This equation allows us to organize S_m in the form of a column vector $\mathcal{S} = \mathcal{S}_+ \bigoplus \mathcal{S}_-$, where $\mathcal{S}_{+,l} = S_{2l}$ contains all the elements $m = 2l$ with $l = 0,1,...$, while $\mathcal{S}_{-,l} = S_{2l+1}$ contains all the elements $m = 2l+1$. As such, the symmetric and antisymmetric modes belong to different sectors and are strictly decoupled. We can write

$$\mathcal{S}_{+/-} = -\frac{4J_z^{+/-}(0)}{d}\mathbb{E}_{+/-}, \tag{33}$$

where $\mathbb{E}_{+,l} = 1 - \delta_{l,0}/2$ and $\mathbb{E}_{-,l} = 1$.

2. *Equation of motion*

The equation of motion is obtained by Fourier transforming Eq. (3) and using Eqs. (20), (27), and (30) as well as (33). We find

$$[\mathcal{H}^{+/-}(\bar{\omega}) - \bar{\omega}^2\mathbb{I}]\rho^{+/-} = i\bar{\omega}J_z^{+/-}(0)\frac{4}{d}\mathbb{E}_{+/-}, \tag{34}$$

where the matrix reads

$$\mathcal{H}_{ll'}^{+/-}(\bar{\omega}) = \delta_{l,l'}\Omega^2(k,q_l^{+/-};\bar{\omega}) + \mathcal{M}_{ll'}^{+/-}.$$

Here the column vectors are defined by $\rho_l^+ = \rho_{2l}$ and $\rho_l^- = \rho_{2l+1}$. We can rewrite

$$J_z^{+/-}(0) = -\frac{i}{\bar{\omega}}\frac{d}{4}\sum_{l=0}^{\infty}\mathcal{G}_l^{+/-}\rho_l^{+/-} = -\frac{i}{\bar{\omega}}\frac{d}{4}\mathcal{G}^{+/-}\rho^{+/-}, \tag{35}$$

where $\mathcal{G}^{+/-} = \mathcal{G}_D^{+/-} + \mathcal{G}_B^{+/-}$ is a row vector. We have

$$\mathcal{G}_l^{+/-} = \frac{4\pi G^{+/-}(k,q_l^{+/-};\bar{\omega})}{k^2 + (q_l^{+/-})^2}, \tag{36}$$

where $G^{+/-}(k,q;\bar{\omega}) = G_D^{+/-}(k) + G_B^{+/-}(k,q;\bar{\omega})$, with

$$G_D^{+/-}(k) = \frac{2}{d}\frac{\omega_p^2}{4\pi}k(1 \mp e^{-kd}),$$

which is comparable to the counterpart for semi-infinite metals, and

$$\begin{aligned} &G_B^{+/-}(k,q;\bar{\omega}) \\ &= i\bar{\omega}\frac{2}{d}\left(\frac{m}{2\pi\hbar}\right)^3\int d^3\mathbf{v}\Theta(v_z)(-e^2f_0')\left(\pm e^{i\frac{\bar{\omega}d}{v_z}} - 1\right) \\ &\quad \times v_z L(v_x,v_z,k,q,\bar{\omega},\pm 1), \end{aligned} \tag{37}$$

where $L(v_x,v_z,k,q,\bar{\omega},s) = L_{emg}(v_x,v_z,k,q,\bar{\omega},s) + pL_{ref}(v_x,v_z,k,q,\bar{\omega},s)$, with $L_{emg/ref}$ given by Eqs. (24) and (25).

3. *SPWs as localized solutions*

Two types of solutions exist to Eq. (34), depending on whether $J_z^{+/-}(0)$ vanishes or not. SPWs are described by solutions with $J_z^{+/-}(0) \neq 0$. These solutions represent localized surface waves, for which the equation can be directly solved. We obtain

$$1 = \mathcal{G}^{+/-}[\mathcal{H}^{+/-}(\bar{\omega}) - \bar{\omega}^2\mathbb{I}]^{-1}\mathbb{E}_{+/-}, \tag{38}$$

which involves no approximations.

Let us write the solution as $\bar{\omega} = \omega_s + i\gamma_0$ and hence the SPW eigenfrequency is given by $\omega = \omega_s + i\gamma$ with $\gamma = \gamma_0 - 1/\tau$. One can show that $\omega_s + i\gamma_0$ always occurs with $-\omega_s + i\gamma_0$, in accord with the fact that $\rho(\mathbf{x},t)$ is real valued. We shall take $\omega_s \geqslant 0$ for definiteness.

Dropping $\mathcal{M}^{+/-}$ as an approximation, the equation becomes

$$1 = \sum_{l=0}^{\infty}\frac{4\pi G^{+/-}(k,q_l^{+/-};\bar{\omega})}{k^2 + (q_l^{+/-})^2}\frac{(1-\delta_{l,0}/2)|1}{\Omega^2(k,q_l^{+/-};\bar{\omega}) - \bar{\omega}^2}. \tag{39}$$

In addition, we have

$$\rho_l^{+/-} = \frac{i\bar{\omega}J_z^{+/-}(0)}{\Omega^2(k,q_l^{+/-};\bar{\omega}) - \bar{\omega}^2}\frac{4}{d}[(1-\delta_{l,0}/2)|1]. \tag{40}$$

Notably, τ is not explicitly involved in any of the above equations, implying that the value of $\bar{\omega}$ does not depend on τ.

F. Approximate and numerical solutions

1. *Hydrodynamic/Drude limits*

The hydrodynamic model is attained when the surface-ballistic effects, synthesized in the quantity $G_B^{+/-}(k,q;\bar{\omega})$, are ignored in total and the bulk plasma wave dispersion is taken as given by Eq. (21), i.e., $\Omega(k,q;\bar{\omega}) \approx \omega_p^2 + (3/5)(k^2+q^2)v_F^2$. In the Drude model, the dispersion is also ignored. In both models, $\bar{\omega}$ is real valued and $\mathrm{Im}(\omega) = -1/\tau$. Solving Eq. (39) without $G_B^{+/-}(k,q;\bar{\omega})$, for large d we obtain $\omega_s^{+/-} = \omega_{s0}^{+/-}$, with $\omega_{s0}^{+/-} = (\omega_p/\sqrt{2})\sqrt{1 \pm e^{-kd}}$ for the symmetric/antisymmetric modes of SPWs. Note that the bulk wave frequency always lies above the SPW frequency and hence the factor $1/(\Omega^2(k,q;\omega_p) - \bar{\omega}^2)$ never develops a pole near ω_s: SPWs cannot decay via bulk waves.

2. *Approximate solutions*

We can solve (39) approximately. To the lowest order in γ_0/ω_s, we may determine ω_s by approximating the real part of (39) as follows

$$1 \approx \sum_{l=0}^{\infty}\frac{4\pi\,\mathrm{Re}[G^{+/-}(k,q_l^{+/-};\omega_s)]}{k^2 + (q_l^{+/-})^2}\frac{(1-\delta_{l,0}/2)|1}{\Omega^2(k,q_l^{+/-};\omega_s) - \omega_s^2}. \tag{41}$$

The as-obtained ω_s is then substituted in the imaginary part of Eq. (39) to get γ_0. We find

$$\frac{\gamma_0}{\omega_s} \approx -\frac{1}{2}\frac{\sum_{l=0}^{\infty}\frac{4\pi}{k^2+(q_l^{+/-})^2}\frac{(1-\delta_{l,0}/2)|1}{\Omega^2(k,q_l^{+/-};\omega_s)-\omega_s^2}\mathrm{Im}[G^{+/-}(k,q_l^{+/-};\omega_s)]}{\sum_{l=0}^{\infty}\frac{4\pi}{k^2+(q_l^{+/-})^2}\frac{(1-\delta_{l,0}/2)|1}{\Omega^2(k,q_l^{+/-};\omega_s)-\omega_s^2}\mathrm{Re}[G^{+/-}(k,q_l^{+/-};\omega_s)]\frac{\omega_s^2}{\Omega^2(k,q_l^{+/-};\omega_s)-\omega_s^2}}, \tag{42}$$

which can be brought into a rather simple form if we take $\Omega(k,q;\omega_s)\approx\omega_p$ and $\omega_s^2/\omega_p^2\sim 1/2$ for $kd\gg 1$. We get

$$\frac{\gamma_0}{\omega_s} \approx -\frac{1}{2}\frac{\sum_{l=0}^{\infty}\frac{(1-\delta_{l,0}/2)|1}{k^2+(q_l^{+/-})^2}\mathrm{Im}[G^{+/-}(k,q_l^{+/-};\omega_s)]}{\sum_{l=0}^{\infty}\frac{(1-\delta_{l,0}/2)|1}{k^2+(q_l^{+/-})^2}\mathrm{Re}[G^{+/-}(k,q_l^{+/-};\omega_s)]} = \frac{1}{2}\frac{\mathrm{Re}[J_z^{+/-}(0)]}{\mathrm{Im}[J_z^{+/-}(0)]}, \tag{43}$$

with $J_z^{+/-}(0)$ evaluated by Eq. (35) with ω_s in place of $\bar{\omega}$. This relation can also be established by an energy analysis, see Sec. IV. By virtue of the relation that $\mathrm{Im}[G^{+/-}(k,q;\omega_s)]+\mathrm{Im}[G^{+/-}(k,q;-\omega_s)]=0$, the same $\mathrm{Im}(\bar{\omega})$ exists for $-\omega_s$, as anticipated from the fact that charge density waves are real-valued waves.

To make progress, we need to evaluate $G_B^{+/-}(k,q;\bar{\omega})$. Writing the integration in Eq. (37) in spherical coordinates and performing it over the magnitude of $\mathbf{v}$, we arrive at

$$G_B^{+/-}(k,q;\bar{\omega}) = -i\frac{2}{d}\frac{\omega_p^2}{4\pi}\frac{3\bar{\omega}}{4\pi v_F}\int_0^{2\pi}d\varphi\int_0^{\pi/2}d\theta\sin\theta\cos(\theta) \times\tilde{L}(v_F\sin\theta\cos\varphi, v_F\cos\theta, k, q, \bar{\omega}, \pm), \tag{44}$$

where we have written $\mathbf{v}=v(\sin\theta\cos\varphi,\sin\theta\sin\varphi,\cos\theta)$ and $\tilde{L}(v_x,v_z,k,q,\bar{\omega},s)=(1\mp e^{i\tilde{\omega}d/v_z})L(v_x,v_z,k,q,\bar{\omega},s)$. We expand all factors other than $e^{i\tilde{\omega}d/v_z}$ in $L(v_x,v_z,k,q,\bar{\omega},s)$ into a series of $kv_F/\bar{\omega}$ and retain only the leading term. We find

$$\begin{aligned}L(v_x,v_z,k,q,\bar{\omega},s) \approx{}& \frac{2q^2v_z^2}{\bar{\omega}^2-q^2v_z^2}(1+p-spe^{i\tilde{\omega}d/v_z})\\ &+(1-se^{-kd})\left[\frac{kv_x}{\bar{\omega}}(1-p+pe^{i\tilde{\omega}d/v_z})\right.\\ &\left.+\frac{kv_z}{i\bar{\omega}}(1-p-pe^{i\tilde{\omega}d/v_z})\right].\end{aligned} \tag{45}$$

Upon being formally integrated over φ, the integral in Eq. (44) ends up in this form, $\int_0^1 dr(L_0(r)+e^{ir_0/r}L_1(r)+e^{2ir_0/r}L_2(r))$, where only the dependence on $r=\cos\theta$ is explicitly noted down in the integrand and $r_0=\bar{\omega}d/v_F\gg 1$. As $e^{ir_0/r}$ and $e^{2ir_0/r}$ are rapidly oscillating functions whereas $L_{1,2,3}(r)$ are slowly varying functions, $L_0(r)$ is the dominant contribution to the integral. We neglect other contributions and obtain

$$\begin{aligned}G_B^{+/-}(k,q;\bar{\omega}) \approx{}& -i\frac{2}{d}\frac{\omega_p^2}{4\pi}\frac{3\bar{\omega}}{v_F}(1+p)\int_0^1 dr\frac{q^2v_F^2r^3}{\bar{\omega}^2-q^2v_F^2r^2}\\ &-\frac{2}{d}\frac{\omega_p^2}{4\pi}k(1\mp e^{-kd})\frac{1-p}{2}.\end{aligned} \tag{46}$$

This expression explicitly shows that $\mathrm{Im}[G_B^{+}(k,q;\omega_s)]=\mathrm{Im}[G_B^{-}(k,q;\omega_s)]<0$, leading to $\gamma_0>0$ by virtue of Eq. (42).

It follows that $\mathrm{Re}[G^{+/-}(k,q;\omega_s)]=(2/d)(\omega_p^2/4\pi)k(1\mp e^{-kd})(1+p)/2$. Substituting this in Eq. (41) and converting the sum therein into an integral for large d, we get $\omega_s^{+/-}/\omega_p=\sqrt{1-(1\pm e^{-kd})(1+p)/4}$. See that $\omega_s^{+/-}$ depends on surface properties via the parameter p. Only for $p=1$ would the conventional value, $\omega_p/\sqrt{2}$, be recovered. For $p=0$ and at large kd, $\omega_s=(\sqrt{3}/2)\omega_p$ is slightly larger than the former. It is notable that, ω_s^+ remains finite even for $k=0$, in distinct contrast with the Drude model. The reason is simple: In the Drude model no electric field could exist in the metal for symmetric modes at $k=0$, while in our theory, due to a spatial spread of charge density, the electric field does not vanish. The same conclusion applies to semi-infinite metals.

To estimate γ_0 by Eq. (42), we take in Eq. (46) $\int_0^1 dr\frac{q^2v_F^2r^3}{\bar{\omega}^2-q^2v_F^2r^2}\approx(qv_F/2\bar{\omega})^2$ for simplicity. Thus, $\mathrm{Im}[G_B^{+/-}(k,q;\omega_s)]\approx-(2/d)(\omega_p^2/4\pi)(3/4)(v_F/\omega_s)(1+p)q^2$, which is then plugged in Eq. (42) to produce $\gamma_0\sim(3/2\pi)(\omega_p/\sqrt{2})/(1\mp e^{-kd})\approx 0.35\omega_p$. In obtaining this expression, we have put $\sum_l\frac{(q_l^{+/-})^2}{(q_l^{+/-})^2+k^2}\approx q_cd/2\pi$ with $q_cv_F\sim\omega_p/\sqrt{2}$. Landau damping has been excluded here, as the approximation only takes the real part of $\Omega(k,q;\omega_s)$.

3. *Numerical solutions*

We can also accurately solve Eq. (39) numerically. The results are displayed in Figs. 1(a), 1(b), and 1(c). A comparison with the approximate solution is not direct, because the approximate solution has excluded while the numerical solution has automatically taken care of Landau damping. It is stressed that the numerical solutions do not depend on the value of q_c, provided it is large enough—in excess of k_s.

IV. ENERGY CONVERSION WITH SURFACES

In this section, we show that the surface plays a critical role in the energy conversion of bounded systems. While it might be straightforward to handle this issue if the surface potential $\phi_s(\mathbf{x},t)$ is exactly known, it is less clear otherwise. Here we derive from Eq. (1) a generic equation that governs the evolution of the electrostatic potential energy, denoted by

$$E_p(t)=(1/2)\int d^3\mathbf{x}\rho(\mathbf{x},t)\phi(\mathbf{x},t)$$

of the system, dispensing with the need to know $\phi_s(\mathbf{x},t)$. We then use it to furnish another proof of Eq. (43). For this purpose, we multiply Eq. (1) by $\phi(\mathbf{x},t)$ and integrate it over space to

obtain

$$\left(\partial_t + \frac{2}{\tau}\right)E_p(t) = -P^{(1)}(t) - P^{(2)}(t), \tag{47}$$

where $P^{(1)}(t) = \int d^3\mathbf{x}\mathbf{J}(\mathbf{x},t)\cdot\mathbf{E}(\mathbf{x},t)$ is no more than the work done by the electric field on the electrons per unit time and

$$\begin{aligned} P^{(2)}(t) &= \frac{1}{2}\int d^3\mathbf{x}\mathcal{J}(\mathbf{x},t)E_z(\mathbf{x},t), \\ \mathcal{J}(\mathbf{x},t) &= J_z(\mathbf{x}_0,t)\Theta(z) - J_z(\mathbf{x}_d,t)\Theta(z-d). \end{aligned} \tag{48}$$

It is evident that $P^{(2)}(t)$ signifies the work done by the surface on the electrons per unit time: Electrons impinging toward the surface may lose their momentum. As far as we are concerned, this term and its consequences have hitherto not been discussed in existing work. We can translate Eq. (47) into the following, see Appendix A for details,

$$\begin{aligned} &\gamma_0\int dz(\mathrm{Re}[\rho(z)]\mathrm{Re}[\phi(z)] + \mathrm{Re}\to\mathrm{Im}) \\ &= -\frac{1}{2}\left\{\int dz(\mathrm{Re}[\mathbf{J}(z)]\cdot\mathrm{Re}[\mathbf{E}(z)] + \mathrm{Re}\to\mathrm{Im}) \right. \\ &\quad \left. + \frac{1}{2}\int dz(\mathrm{Re}[\mathcal{J}(z)]\mathrm{Re}[E_z(z)] + \mathrm{Re}\to\mathrm{Im})\right\}, \end{aligned} \tag{49}$$

where the integral is extended over the metal. If the phase of $\rho(z)$ is global, i.e., independent of z, Eq. (49) holds valid even without the abbreviated term.

Now we show how Eq. (43) can also be reached from Eq. (49). Neglecting Landau damping, by Eq. (40) we can show that $\rho(z)$ has a global phase. We can then ignore in this equation the terms abbreviated as $\mathrm{Re}\to\mathrm{Im}$ without affecting the results. To the zeroth order in γ_0, it is obvious that $\mathrm{Re}[\mathbf{J}_D(z)]\cdot\mathrm{Re}[\mathbf{E}(z)] = 0$ and $\mathrm{Re}[\mathcal{J}_D(z)]\mathrm{Re}[E_z(z)] = 0$, i.e., diffusive currents do not bear net work from the electric field. As for the surface-ballistic currents, note that $\mathbf{J}_B(\mathbf{v},z)$ contains the rapidly oscillating factor $e^{i\bar{\omega}z/v_z}$, which suppresses the term $\int dz\mathrm{Re}[\mathbf{J}_B(z)]\cdot\mathrm{Re}[\mathbf{E}(z)]$ by the factor kv_F/ω_p, echoing the fact that $\mathcal{M}^{+/-}$ can be neglected in Eq. (34). As such, we have

$$\begin{aligned} \gamma_0 &\approx -\frac{1}{2}\frac{\frac{1}{2}\int dz\mathrm{Re}[\mathcal{J}_B(z)]\mathrm{Re}[E_z(z)]}{\int dz\mathrm{Re}[\rho(z)]\mathrm{Re}[\phi(z)]} \\ &= -\frac{1}{4}\frac{\mathrm{Re}\big[J_{B,z}(0)\big]\mathrm{Re}[\phi(0)] - \mathrm{Re}\big[J_{B,z}(d)\big]\mathrm{Re}[\phi(d)]}{\int dz\mathrm{Re}[\rho(z)]\mathrm{Re}[\phi(z)]} \\ &= -\frac{1}{2}\frac{\mathrm{Re}[J_z(0)]\mathrm{Re}[\phi(0)]}{\int dz\mathrm{Re}[\rho(z)]\mathrm{Re}[\phi(z)]}, \end{aligned}$$

where in the last equality, we have used the fact that, for either symmetric or antisymmetric modes $J_z(d)\phi(d) + J_z(0)\phi(0) = 0$ and that $\mathrm{Re}[J_z(0)] = \mathrm{Re}[J_{B,z}(0)]$. To evaluate the denominator, we utilize the equation of motion in real space. It can be easily obtained from Eq. (3) with $\nabla\cdot\mathbf{J}(z) \approx \nabla\cdot\mathbf{J}_D(z) \approx (i/\bar{\omega})\omega_p^2\rho(z)$. We find

$$\begin{aligned} \rho(z) &\approx \frac{1}{i}\frac{\bar{\omega}}{\omega_p^2 - \bar{\omega}^2}[J_z(d)\delta(z-d) - J_z(0)\delta(z)], \\ \Longrightarrow \mathrm{Re}[\rho(z)] &\approx \frac{\omega_s}{\omega_p^2 - \omega_s^2}(\mathrm{Im}[J_z(d)]\delta(z-d) \\ &\quad - \mathrm{Im}[J_z(0)]\delta(z)), \end{aligned} \tag{50}$$

As a result, $\int dz\mathrm{Re}[\rho(z)]\mathrm{Re}[\phi(z)] \approx -\frac{\omega_s}{\omega_p^2-\omega_s^2}\mathrm{Im}[J_z(0)]\mathrm{Re}[\phi(0)]$. By substitution, we immediately recover Eq. (43).

V. DISCUSSIONS AND CONCLUSIONS

Thus, on the basis of Boltzmann's equation, we have established a rigorous theory for SPWs in metal films with arbitrary electronic collision rate $1/\tau$. As a key consequence of the theory, we find that there exists a self-amplification channel for SPWs, which would cause the latter to spontaneously amplify at a rate γ_0 if not for electronic collisions. Surprisingly, the value of γ_0 turns out to be independent of τ. The presence of this channel is guaranteed by the causality principle. Whether the system could actually amplify or not depends on the competition between γ_0 and $1/\tau$. If $\gamma_0 > 1/\tau$, SPWs will amplify and the system will become unstable. In our theory, the nonequilibrium deviation $g(\mathbf{v},z)$ refers to the Fermi-Dirac distribution $f_0(\varepsilon)$; as such, the instability is one of the Fermi sea. Needless to say, the instability will be terminated once the system deviates far enough from the Fermi sea and settles in a stable state. Clarifying the nature of the destination state is a subject of crucial importance for future study.

One central feature of our theory is the classification of current densities into a diffusive component $\mathbf{J}_D(z)$ and a surface-ballistic component $\mathbf{J}_B(z)$. This classification is not based on the value of τ but according to whether the component obeys the (generalized) Ohm's law or not. Apart from this, these components are also discriminated in other ways. Firstly, they are controlled by different length scales. As it largely follows the local electric field $\mathbf{E}(z)$, the characteristic length associated with $\mathbf{J}_D(z)$ is k^{-1}. On the other hand, the length for $\mathbf{J}_B(z)$ is v_F/γ_0, because of simple z dependence. Secondly, they are oriented disparately. $\mathbf{J}_D(z)$ is largely oriented normal to $\mathbf{E}(z)$ locally whereas $\mathbf{J}_B(z)$ is normal to the surfaces—especially for p close to unity. Considering energy conversion, this explains why $\mathbf{J}_D(z)$ does not destabilize the Fermi sea but $\mathbf{J}_B(z)$ does. Thirdly, $\mathbf{J}_D(z)$ is a bulk property and exists regardless of the surface. On the contrary, $\mathbf{J}_B(z)$ reflects true surface effects and it would disappear without surfaces.

Although our theory applies at finite temperature, our calculation of $\bar{\omega}$ is done only at zero temperature, i.e., we have taken $f_0(\varepsilon)$ to be a step distribution. Clarifying the temperature dependence of $\mathrm{Im}(\bar{\omega})$ is important for experimental studies of the present theory, because the net amplification/damping rate $\gamma = \gamma_0 - 1/\tau$ can be directly measured. Arguably, γ_0 could bear a different temperature dependence than $1/\tau$. In sufficiently pure samples, in which the residual resistivity is small enough, there might exist a critical temperature T^*, above which $\gamma < 0$ while below it $\gamma > 0$. In other words, T^* marks the transition of the system from the Fermi sea to a more stable state.

Another problem that needs to be addressed in the future for experimental studies is concerned with the effects of interband transitions. In the most experimented materials, such as silver and gold, these transitions are known to have dramatic effects. They not only open a loss channel due to interband absorption but also significantly shift the SPW frequency. Including them in our formalism consists of a simple generalization: In addition to $\mathbf{J}_D(z)$ and $\mathbf{J}_B(z)$, the

total current density $\mathbf{J}(z)$ must now also have a component $\mathbf{J}_{int}(z)$ accounting for interband transitions. The equation of motion is obtained by substituting $\mathbf{J}(z)$ in Eq. (3). One may write $J_{int,\mu}(z) = \sum_\nu \int dz' \sigma_{\mu\nu}(z,z';\omega) E_\nu(z')$, where $\mu,\nu = x,y,z$ and the interband conductivity $\sigma_{\mu\nu}$ can in principle be calculated using Greenwood-Kubo formula. In practice, calculating $\sigma_{\mu\nu}$ could be a formidable task even for the imaginably simplest surfaces. Nevertheless, one may argue that $\mathbf{J}_{int}(z)$ primarily affects the properties of bulk waves, namely, $\Omega(k,q;\bar{\omega})$. The causality principle should still protect the amplification channel, though the value of $\bar{\omega}$ may depend on τ. A systematic analysis will be presented elsewhere.

To conclude, we have presented a theory for SPWs in metal films taking into account the unique interplay between ballistic electronic motions and boundary effects, from which it emerges a universal self-amplification channel for these waves. It is expected that the study will bear far-reaching practical and fundamental consequences, which are to be explored in the future. We hope that the work could stimulate more effort on this subject.

ACKNOWLEDGMENTS

The author is grateful to K. Wakabayashi for some help with numerical computation. He also thanks E. Mariani for enormous support.

APPENDIX A: MORE ABOUT EQS. (47)–(49)

The not-so-obvious step in proving Eq. (49) is to show that

$$\partial_t E_p(t) = \int d^3\mathbf{x}\phi(\mathbf{x},t)\partial_t\rho(\mathbf{x},t). \tag{A1}$$

For this purpose, we write $\phi(\mathbf{x},t) = \mathrm{Re}[e^{-i\omega t}\phi(\mathbf{x})]$ and $\rho(\mathbf{x},t) = \mathrm{Re}[e^{-i\omega t}\rho(\mathbf{x})]$, where $\phi(\mathbf{x}) = e^{ikx}\phi(z)$ and $\rho(\mathbf{x}) = e^{ikx}\rho(z)$. Moreover, we put $\omega = \omega_s + i\gamma$, $\phi(\mathbf{x}) = \phi'(\mathbf{x}) + i\phi''(\mathbf{x})$ and similarly for other complex quantities. By substitution, we find

$$\begin{aligned}&\partial_t E_p(t) - \int d^3\mathbf{x}\phi(\mathbf{x},t)\partial_t\rho(\mathbf{x},t)\\ &\quad= \frac{e^{2\gamma t}}{2}\int d^3\mathbf{x}[\phi''(\mathbf{x})\rho'(\mathbf{x}) - \phi'(\mathbf{x})\rho''(\mathbf{x})].\end{aligned} \tag{A2}$$

However, $\int d^3\mathbf{x}\phi'(\mathbf{x})\rho''(\mathbf{x}) = \int d^3\mathbf{x}\phi''(\mathbf{x})\rho'(\mathbf{x}) = 0$. Actually, we have

$$\begin{aligned}&\int d^3\mathbf{x}\phi'(\mathbf{x})\rho''(\mathbf{x})\\ &= \int dz\int d^2\mathbf{r}(\phi'(z)\rho''(z)\cos^2 kx - \phi''(z)\rho'(z)\sin^2 kx)\\ &\propto \int dz(\phi'(z)\rho''(z) - \phi''(z)\rho'(z))\\ &\propto \int dz\int dz'[\rho''(z)e^{-k|z-z'|}\rho'(z') - \rho'(z)e^{-k|z-z'|}\rho''(z')]\\ &= 0,\end{aligned} \tag{A3}$$

thus completing the proof.

Let us suppose $\rho(z)$ has a global phase, i.e., $\rho(z) = cn(z)$, where c is a complex constant and $n(z)$ is real valued. One can show that Eq. (49) can be turned into an equation that involves only $n(z)$, wherein c plays no role. In other words, Eq. (49) can be evaluated by simply pretending c [and $\rho(z)$] to be real. The proof is evident considering the linear relations between $\rho(z)$ and $\mathbf{J}(z)$ and that between $\rho(z)$ and $\phi(z)$ as well as that between $\rho(z)$ and $\mathbf{E}(z)$.

APPENDIX B: ELECTRONIC DISTRIBUTION FUNCTIONS

The general solution to Eq. (7) is given by

$$g(\mathbf{v},z) = e^{i\frac{\tilde{\omega}z}{v_z}}\left(C(\mathbf{v}) - \frac{e\partial_{\mathbf{v}}f_0}{mv_z}\cdot\int_0^z dz' e^{-i\frac{\tilde{\omega}z'}{v_z}}\mathbf{E}(z')\right), \tag{B1}$$

where $C(\mathbf{v})$ is an arbitrary integration constant to be determined by boundary conditions. Let p_1 and p_2 be the *Fuchs* parameters for the (uniform) surfaces at $z = 0$ and $z = d$, respectively. The boundary condition at $z = 0$ is taken that $g((v_x,v_y,v_z > 0),z = 0) = p_1 g((v_x,v_y,-v_z),z = 0)$ while that at $z = d$ assumes $g((v_x,v_y,v_z < 0),z = 0) = p_2 g((v_x,v_y,-v_z),z = 0)$, both evaluated at $E_z(z) = 0$. After some algebra, one finds

$$g(\mathbf{v},z) = e^{i\frac{\tilde{\omega}z}{v_z}}\begin{cases}\frac{1}{p_1p_2 - e^{-\frac{2d\tilde{\omega}}{v_z}}}\int_0^d dz'\frac{e\mathbf{E}(z')\cdot\partial_{\mathbf{v}}f_0}{mv_z}\left(e^{-i\frac{(2d+z')\tilde{\omega}}{v_z}} + p_1e^{-i\frac{(2d-z')\tilde{\omega}}{v_z}}\right) + \int_z^d dz'\frac{e\mathbf{E}(z')\cdot\partial_{\mathbf{v}}f_0}{mv_z}e^{-i\frac{\tilde{\omega}z'}{v_z}}, & \text{for } v_z \geqslant 0,\\ \frac{-1}{p_1p_2 - e^{\frac{2d\tilde{\omega}}{v_z}}}\int_0^d dz'\frac{e\mathbf{E}(z')\cdot\partial_{\mathbf{v}}f_0}{mv_z}\left(e^{i\frac{(2d-z')\tilde{\omega}}{v_z}} + p_1e^{i\frac{\tilde{\omega}z'}{v_z}}\right) - \int_0^z dz'\frac{e\mathbf{E}(z')\cdot\partial_{\mathbf{v}}f_0}{mv_z}e^{-i\frac{\tilde{\omega}z'}{v_z}}, & \text{for } v_z < 0.\end{cases} \tag{B2}$$

The electronic distribution functions presented in the main text in Sec. III are obtained by approximating $(1 - p_1p_2e^{i\frac{2d\tilde{\omega}}{v_z}})^{-1} \approx 1$ for $v_z \geqslant 0$ and $(1 - p_1p_2e^{-i\frac{2d\tilde{\omega}}{v_z}})^{-1} \approx 1$ for $v_z < 0$ in this equation.

1. Causality principle

It should be pointed out that, in applying the boundary conditions, we have implicitly assumed $\mathrm{Im}(\tilde{\omega}) \geqslant 0$; otherwise, we would find unphysical solutions that violate the principle of causality, which states that the number of outgoing electrons is determined by the number of incoming electrons, not otherwise. It is easy to show that, had we assumed $\mathrm{Im}(\tilde{\omega}) < 0$, we would have found the opposite: the number of reflected electrons would be fixed while the number of incident electrons would go to infinity as $p_{1/2} \to 0$.

APPENDIX C: THE MATRIX $\mathcal{M}^{+/-}$

In the first place, we show that $\mathcal{M}^{+/-}/\omega_p^2 \propto ikv_F/\bar{\omega} + ...$, where the ellipsis stands for higher order terms in $kv_F/\bar{\omega}$. We take the symmetric modes for illustration, as the reasoning can be replicated for the antisymmetric modes as well. Writing $\int d^3\mathbf{v}\Theta(v_z) = \int_0^{2\pi} d\varphi \int_0^{\pi/2} d\theta \sin\theta \int_0^{\infty} dv^2(v/2)$ and integrating over v, we find

$$\frac{\mathcal{M}^+_{l,l'}}{\omega_p^2} = i\left(1-\frac{\delta_{l,0}}{2}\right)\frac{3}{2\pi kd}\int_0^{2\pi} d\varphi \int_0^{\pi/2} d\theta \sin\theta \frac{\bar{\omega}^2\cos\theta - \bar{\omega}kv_F\sin\theta\cos\theta\cos\varphi}{(\bar{\omega}-kv_F\sin\theta\cos\varphi)^2 - (q_l^+)^2v_F^2\cos^2\theta}\frac{k\bar{\omega}/v_F}{k^2+(q_{l'}^+)^2}$$
$$\times\left(e^{i\left(\frac{\bar{\omega}}{v_F\cos\theta}-k\tan\theta\cos\varphi\right)d}-1\right)L(v_F\sin\theta\cos\varphi, v_F\cos\theta, k, q_{l'}^+, \bar{\omega}, +1). \tag{C1}$$

To the lowest order in $kv_F/\bar{\omega}$, we only need to retain $L^{(0)}$ in the expansion $L^{sym} = \sum_{m=0}^{\infty} L^{(m)}(kv_F/\bar{\omega})^m$. Thus,

$$L(v_F\sin\theta\cos\varphi, v_F\cos\theta, k, q, \bar{\omega}, +1) \approx 2\frac{q^2v_F^2\cos^2\theta}{\bar{\omega}^2 - q^2v_F^2\cos^2\theta}\left(1+p-pe^{i\left(\frac{\bar{\omega}}{v_F\cos\theta}-k\tan\theta\cos\varphi\right)d}\right). \tag{C2}$$

Substituting this back in (C1) and approximating

$$\frac{\bar{\omega}^2\cos\theta - \bar{\omega}kv_F\sin\theta\cos\theta\cos\varphi}{(\bar{\omega}-kv_F\sin\theta\cos\varphi)^2 - q^2v_F^2\cos^2\theta} \approx \frac{\bar{\omega}^2\cos\theta}{\bar{\omega}^2 - q^2v_F^2\cos^2\theta}, \quad \frac{q^2}{q^2+k^2} \approx 1, \tag{C3}$$

we arrive at

$$\frac{\mathcal{M}^+_{l,l'}}{\omega_p^2} = i\left(1-\frac{\delta_{l,0}}{2}\right)\frac{3}{\pi kd}\left(\frac{kv_F}{\bar{\omega}}\right)\int_0^{2\pi} d\varphi \int_0^{\pi/2} d\theta \sin\theta \frac{\cos^3\theta}{1-(q_l^+v_F/\bar{\omega})^2\cos^2\theta}\frac{\mathcal{L}(\theta,\varphi)}{1-(q_{l'}^+v_F/\bar{\omega})^2\cos^2\theta}, \tag{C4}$$

where

$$\mathcal{L}(\theta,\varphi) = \left(e^{i\left(\frac{\bar{\omega}}{v_F\cos\theta}-k\tan\theta\cos\varphi\right)d}-1\right)\times\left(1+p-pe^{i\left(\frac{\bar{\omega}}{v_F\cos\theta}-k\tan\theta\cos\varphi\right)d}\right). \tag{C5}$$

Clearly, we have $\mathcal{M}^+/\omega_p^2 \sim kv_F/\bar{\omega}$, as stated.

We may proceed further if we take

$$\frac{3}{\pi}\int_0^{2\pi} d\varphi \int_0^{\pi/2} d\theta \sin\theta \frac{\cos^3\theta}{1-(q_l^+v_F/\bar{\omega})^2\cos^2\theta}\frac{\mathcal{L}(\theta,\varphi)}{1-(q_{l'}^+v_F/\bar{\omega})^2\cos^2\theta} \approx \frac{3}{\pi}\int_0^{2\pi} d\varphi \int_0^{\pi/2} d\theta \sin\theta\cos^3\theta\mathcal{L}(\theta,\varphi) \sim -1, \tag{C6}$$

from which it follows that $\mathcal{M}^+_{l,l'} \approx M_0 = -i\omega_p^2(1/kd)(kv_F/\bar{\omega})$, which is a constant. Therefore, $\mathcal{M}^+_{l,l'} \approx M_0\mathbb{Z}_{l,l'}$, where $\mathbb{Z}_{l,l'} = 1$ constitutes a unity matrix. We write, with $\mathcal{W}^+_{l,l'} = \delta_{l,l'}\Omega(k, q_l^+;\bar{\omega})$,

$$[(\mathcal{W}^+)^2 - \bar{\omega}^2\mathbb{I} + \mathcal{M}^+]^{-1} = U^{-1}[(\tilde{\mathcal{W}}^+)^2 - \bar{\omega}^2\mathbb{I} + \tilde{\mathcal{M}}^+]^{-1}U, \tag{C7}$$

where U is a similarity transformation that brings $\mathcal{M}^+$ and hence $\mathbb{Z}$ to a diagonal form. We have used a tilde to indicate the transformed matrices, e.g., we write $\tilde{\mathbb{Z}} = U\mathbb{Z}U^{-1}$. See that $\tilde{\mathbb{Z}}$ has only one nonvanishing element, whose value amounts to the dimension N_c of the matrix. Let it be the l_0th element. Then $\tilde{\mathbb{Z}}_{l,l'} = N_c\delta_{l,l_0}\delta_{l',l_0}$. Obviously, $N_c = q_cd/2\pi \sim (\bar{\omega}/kv_F)(kd/2\pi)$. As such, $M_0 \sim 1/N_c$ and $\tilde{\mathcal{M}}^+_{l,l'} \sim -i(\omega_p^2/2\pi)\delta_{l,l_0}\delta_{l',l_0}$. Introducing $\tilde{\mathcal{G}}^+ = \mathcal{G}^+U^{-1}$ and $\tilde{\mathbb{E}}_+ = U\mathbb{E}_+$, we can rewrite the equation of motion for the symmetric modes as

$$1 = \tilde{\mathcal{G}}^+[(\tilde{\mathcal{W}}^+)^2 - \bar{\omega}^2\mathbb{I} + \tilde{\mathcal{M}}^+]^{-1}\tilde{\mathbb{E}}_+. \tag{C8}$$

Taking $\mathcal{W}^+ \approx \omega_p\mathbb{I}$ and hence $\tilde{\mathcal{W}}^+ \approx \omega_p\mathbb{I}$, this equation becomes

$$1 = \sum_l \tilde{\mathcal{G}}_l^+\frac{1}{\omega_p^2 - \bar{\omega}^2 + \tilde{\mathcal{M}}^+_{l,l}}\tilde{\mathbb{E}}_{+,l} = \sum_l \mathcal{G}_l^+\frac{1}{\omega_p^2-\bar{\omega}^2}\mathbb{E}_{+,l} + \left[\tilde{\mathcal{G}}_{l_0}^+\frac{1}{\omega_p^2(1+i/2\pi)-\bar{\omega}^2}\tilde{\mathbb{E}}_{+,l_0} - \mathcal{G}_{l_0}^+\frac{1}{\omega_p^2-\bar{\omega}^2}\mathbb{E}_{+,l_0}\right]$$
$$\approx \sum_l \mathcal{G}_l^+\frac{1}{\omega_p^2-\bar{\omega}^2}\mathbb{E}_{+,l}.$$

The term in the square bracket makes only a contribution of the order of $\sim 1/N_c$ and can be neglected for large N_c.

[1] R. H. Ritchie, Phys. Rev. **106**, 874 (1957).
[2] R. A. Ferrell, Phys. Rev. **111**, 1214 (1958).
[3] H. Raether, *Surface Plasmons on Smooth and Rough Surfaces and on Gratings* (Springer, Berlin, Heidelberg, 1988).
[4] J. M. Pitarke, V. M. Silkin, E. V. Chulkov, and P. M. Echenique, Rep. Prog. Phys. **70**, 1 (2007).
[5] P. J. Feibelman, Prog. Surf. Sci. **12**, 287 (1982).
[6] P. M. Echenique and J. B. Pendry, J. Phys. C **8**, 2936 (1975).

[7] Interview with J. Krenn, Nat. Photonics **6**, 714 (2012).
[8] B. Rothenhäusler and K. Wolfgang, Nature (London) **332**, 615 (1988).
[9] X. D. Hoa, A. G. Kirk, and M. Tabrizian, Biosens. Bioelectron. **23**, 151 (2007).
[10] A. V. Zayats, I. S. Igor, and A. A. Maradudin, Phys. Rep. **408**, 131 (2005).
[11] S. A. Maier *Plasmonics: Fundamentals and Applications* (Springer Science and Business Media, New York, 2007).
[12] E. Ozbay, Science **311**, 189 (2006).
[13] M. L. Brongersma and P. G. Kik, *Surface Plasmon Nanophotonics* (Springer, New York, 2007).
[14] W. L. Barnes, A. Dereux, and T. W. Ebbesen, Nature (London) **424**, 824 (2003); W. L. Barnes, J. Opt. A **8**, S87 (2006).
[15] T. W. Ebbesen, C. Genet, and S. I. Bozhevolnyi, Phys. Today **61**, 44 (2008).
[16] M. S. Tame, K. R. McEnergy, S. K. Ozdemir, J. Lee, S. A. Maier, and M. S. Kim, Nat. Phys. **9**, 329 (2013).
[17] S. Nie and Steven R. Emory, Science **275**, 1102 (1997).
[18] D. Sarid and W. Cgallener, *Modern Introduction to Surface Plasmons: Theory, Mathematic Modeling and Applications* (Cambridge University Press, Cambridge, 2010).
[19] J. Harris, Phys. Rev. B **4**, 1022 (1971).
[20] A. L. Fetter, Ann. Physics **81**, 367 (1973); Phys. Rev. B **33**, 3717 (1986).
[21] Y. Luo, A. I. Fernandez-Dominguez, A. Wiener, S. A. Maier, and J. B. Pendry, Phys. Rev. Lett. **111**, 093901 (2013).
[22] Y. Luo, R. Zhao, and J. B. Pendry, Proc. Natl. Acad. Sci. U.S.A. **111**, 18422 (2014).
[23] J. B. Pendry, Y. Luo, and R. Zhao, Science **348**, 521 (2015)
[24] O. Schnitzer, V. Giannini, S. A. Maier, and R. V. Craster, Proc. R. Soc. A **472**, 20160258 (2016).
[25] J. M. Ziman, *Electrons and Phonons: the Theory of Transport Phenomena in Solids* (Oxford University Press, Clarendon, 2001).
[26] D. Pines, *Elementary Excitations in Solids* (W. A. Benjamin, New York, 1963).
[27] A. A. Abrikosov, *Fundamentals of the Theory of Metals* (Elsiver Science Publishers B. V., North-Holland, 1988).
[28] H.-Y. Deng, K. Wakabayashi, and C.-H. Lam, Phys. Rev. B **95**, 045428 (2017).
[29] H.-Y. Deng, arXiv:1606.06239.
[30] H.-Y. Deng and K. Wakabayashi, Phys. Rev. B **92**, 045434 (2015).
[31] D. J. Bergman and M. I. Stockman, Phys. Rev. Lett. **90**, 027402 (2003).
[32] J. Seidel, S. Grafstrom, and L. Eng, Phys. Rev. Lett. **94**, 177401 (2005).
[33] I. De Leon and P. Berini, Phys. Rev. B **78**, 161401(R) (2008).
[34] I. De Leon and P. Berini, Nat. Photonics **4**, 382 (2010).
[35] D. Y. Fedyanin and A. Y. Arsenin, Opt. Express **19**, 12524 (2011).
[36] P. Berini and I. De Leon, Nat. Photonics **6**, 16 (2012).
[37] D. Y. Fedyanin, A. V. Arsenin, and A. V. Zayats, Nano Lett. **12**, 2459 (2012).
[38] S. Kéna-Cohen, P. N. Stavrinou, D. D. C. Bradley, and S. A. Maier, Nano Lett. **13**, 1323 (2013).
[39] R. W. Brown, P. Wessel, and E. P. Trounson, Phys. Rev. Lett. **5**, 472 (1960).
[40] E. T. Arakawa, R. J. Herickhoff, and R. D. Birkhoff, Phys. Rev. Lett. **12**, 319 (1964).
[41] J. F. Donohue and E. Y. Wang, J. Appl. Phys. **59**, 3137 (1986); **62**, 1313 (1987).