Role of disorder in plasmon-assisted near-field heat transfer between two-dimensional metals

Jonathan L. Wise,[*] Denis M. Basko, and Frank W. J. Hekking[†]

Université Grenoble Alpes and CNRS, LPMMC, 25 rue des Martyrs, 38042 Grenoble, France

We perform a theoretical study of the near-field heat transfer between two atomically thin metallic layers, isolated galvanically but coupled by the Coulomb interaction, within the framework of fluctuational electrodynamics in the Coulomb limit. We clarify the role of disorder and spatial dispersion, and identify several distinct regimes of the heat transfer. We find that the plasmon contribution to the heat current is suppressed in both the clean and diffusive limits, but dominates in a parametrically wide crossover region at sufficiently high temperatures.

I. INTRODUCTION

In metals, heat is most efficiently transported by electrons, as manifested by the Wiedemann-Franz law. In the absence of a galvanic contact between two nearby metals, heat can be transferred by electromagnetic fluctuations [1–5]. In the near field, the heat flux can be strongly enhanced compared to the Planckian radiative one, as was observed in many experiments (see reviews [6–8] as well as Refs. [9–15] for more recent experiments). Depending on the circumstances, such contactless heat transfer can be useful or harmful.

The standard framework for the theoretical description of this heat transfer is fluctuational electrodynamics [2,16]. It provides a very intuitive picture: the transferred heat can be viewed as energy dissipated in one part of the structure by fluctuating electromagnetic fields, which are created by fluctuating charges and currents in the other part of the structure. According to the fluctuation-dissipation theorem, both dissipation and fluctuations are determined by the material response functions (such as optical conductivity), which thus serve as the input to this phenomenological theory. The resulting heat current is then entirely determined by the system geometry and the model used for the material response. For the latter, models with different levels of sophistication have been used in the literature. They range from assuming a local frequency-dependent conductivity [2], to including nonlocal effects via spatial dispersion [17] or via surface contributions to the response [3,18]. From the literature, it is often difficult to decide which ingredients are important in which situation.

In particular, in several works dedicated to different materials in the near-field regime, the important role played by collective plasmon excitations has been pointed out [19–27]. However, Ref. [28] concluded that surface plasmons were unimportant for the heat transfer between two bulk semi-infinite metals, and in Ref. [29] no plasmon contribution was reported for heat current between two thin metallic layers in the clean and diffusive limits.

To clarify this issue, we revisit the problem of heat transfer between two thin parallel metallic layers, assuming the separation between them to be small enough, so that retardation effects can be neglected and the transfer is dominated by the Coulomb interaction (the conditions for this are discussed in Sec. II B). Including disorder and spatial dispersion within the framework of fluctuational electrodynamics in the Coulomb limit, we identify different regimes of interlayer separation and temperature where different physical ingredients are important. We find that the plasmon contribution to the heat current is suppressed in both the clean and diffusive limits, but dominates in a parametrically wide crossover region at sufficiently high temperatures. We also discuss the applicability of the effective circuit approach [30] to heat transfer between metals.

The paper is organized as follows. In Sec. II we specify our model for the system, discuss the main regimes of heat transfer, present the resulting expressions for the heat current, and discuss their sensitivity to the specific model chosen here. We give the detailed derivation of the results in Sec. III, and verify them numerically in Sec. IV. In Sec. V, we interpret our results in terms of circuit theory, and in Sec. VI we discuss some relevant experiments. In the Appendix, we give a derivation of the density response function for a two-dimensional (2D) metal with short-range impurities, interpolating between clean and diffusive limits.

II. MODEL AND SUMMARY OF RESULTS

A. Model

We consider two thin parallel metallic layers, labeled 1 and 2, separated by a distance d, and kept at different temperatures T_1 and T_2. We assume to be in the limit $d \ll c/T_{1,2}$, where c is the speed of light, and we use units where Planck and Boltzmann constants are set to 1. In this limit, heat transfer is dominated by the Coulomb interaction. Our analysis is based

---

*jonathan.wise@lpmmc.cnrs.fr
†Deceased.

on the standard expression for the heat current per unit area [23,24,29,31,32],

$$J(T_1, T_2) = \int_0^\infty \frac{d\omega}{\pi} \int \frac{d^2\mathbf{q}}{(2\pi)^2} \, \omega[\mathcal{N}_1(\omega) - \mathcal{N}_2(\omega)]\mathcal{T}(\mathbf{q}, \omega),$$
(1)

$$\mathcal{T}(\mathbf{q}, \omega) = 2|V_{12}(\mathbf{q}, \omega)|^2 \, \text{Im} \, \Pi_1(\mathbf{q}, \omega) \, \text{Im} \, \Pi_2(\mathbf{q}, \omega),$$
(2)

often called the Caroli formula, in analogy to a similar expression for electron current across a tunnel junction [33]. In Eq. (1), $\mathcal{N}_{1,2}(\omega) = 1/[\exp(\omega/T_{1,2}) - 1]$ are the Bose distribution functions in the two layers. In Eq. (2),

$$V_{12}(\mathbf{q}, \omega) = \frac{v_\mathbf{q} e^{-qd}}{(1 - v_\mathbf{q}\Pi_1)(1 - v_\mathbf{q}\Pi_2) - v_\mathbf{q}^2\Pi_1\Pi_2 e^{-2qd}}$$
(3)

is the Coulomb interaction between the layers, screened in the random-phase approximation, with $v_\mathbf{q} = 2\pi e^2/q$ being the bare 2D Coulomb potential and $e < 0$ the electron charge. Finally, $\Pi_1(\mathbf{q}, \omega)$ and $\Pi_2(\mathbf{q}, \omega)$ are the susceptibilities determining the linear response $\delta\rho_i(\mathbf{r}, t) = \Pi_i(\mathbf{q}, \omega) \, e\varphi_{\mathbf{q},\omega} \, e^{i\mathbf{qr}-i\omega t}$ of the 2D electron density ρ_i in the corresponding layer $i = 1, 2$ to a perturbing electrostatic potential $\varphi_{\mathbf{q},\omega} \, e^{i\mathbf{qr}-i\omega t}$ applied to that layer. The density response function is related to the in-plane longitudinal optical conductivity, $\sigma(\mathbf{q}, \omega) = (i\omega/q^2)e^2\Pi(\mathbf{q}, \omega)$. Equation (1) can be derived from the Coulomb limit of fluctuational electrodynamics [23], from nonequilibrium Green's function formalism [31], or from the kinetic equation [29].

The density response functions encode all material characteristics of the two layers. We model each layer as a degenerate isotropic 2D electron gas with short-range impurities. Such a system can be characterized by three parameters: (i) ν, the electronic density of states per unit area at the Fermi level including both spin projections, whose energy dependence is neglected; (ii) v_F, the Fermi velocity; and (iii) τ, the elastic scattering time. The Fermi momentum p_F is assumed to be the largest momentum scale, $p_F \gg q$, ω/v_F, T/v_F, $1/(v_F\tau)$. Under these assumptions, the density response function of each layer is temperature independent and given by [34]

$$\Pi(\mathbf{q}, \omega) = -\nu\left[1 + \frac{i\omega\tau}{\sqrt{(1 - i\omega\tau)^2 + (v_F q\tau)^2} - 1}\right],$$
(4)

for an arbitrary relation between ω, $v_F q$, and τ. Equation (4) interpolates between two well-known expressions corresponding to the clean limit ($\tau \to \infty$) and the diffusive limit ($\omega, v_F q \ll 1/\tau$):

$$\Pi(\mathbf{q}, \omega) = -\nu\left[1 + \frac{i\omega}{\sqrt{(v_F q)^2 - \omega^2}}\right] \quad \text{(clean)}, \quad (5a)$$

$$\Pi(\mathbf{q}, \omega) = -\frac{\nu D q^2}{D q^2 - i\omega} \quad \text{(diffusive)}, \quad (5b)$$

where $D = v_F^2\tau/2$ is the diffusion coefficient. Equation (4) corresponds to a 2D analog of Mermin's prescription [35], which was recently used to model the response in monolayer graphene [20,36]. In the Appendix we give a derivation of Eq. (4) based on the Boltzmann equation for electrons scattering on impurities. In spite of several assumptions underlying Eq. (4), its simplicity enables one to obtain an important
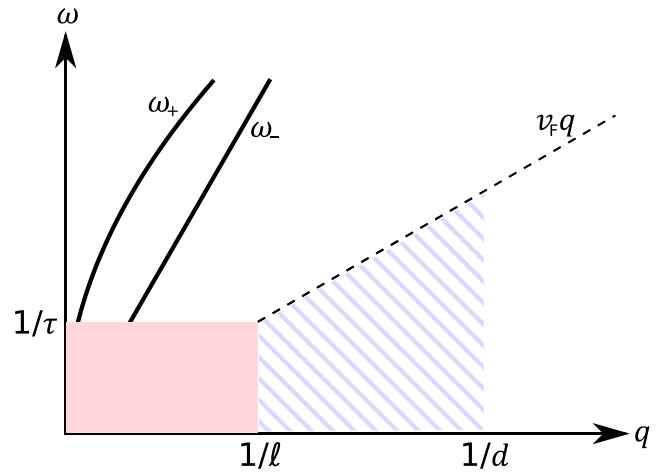

FIG. 1. A schematic picture showing the behavior of $\mathcal{T}(q, \omega)$ [Eq. (2)] in the (q, ω) plane for $d \ll \ell$. For $q \gg 1/d$ (to the right of the hatched area), $\mathcal{T}(q, \omega)$ is suppressed by the factor e^{-qd}. In the hatched area, $\mathcal{T}(q, \omega)$ is well approximated by the clean limit expression (5a), while the shaded area $q\ell \ll 1$, $\omega\tau \ll 1$ corresponds to the diffusive limit (5b). In the white region above the shaded and hatched areas ($\omega \gg 1/\tau$, $\omega > v_F q$) the integrand is small except, maybe, in the vicinity of the symmetric and antisymmetric plasmon dispersions (upper and lower solid lines, respectively) where $|V_{12}|^2$ is peaked.

insight into the interplay between the spatial dispersion [that is, the **q** dependence of the conductivity $\sigma(\mathbf{q}, \omega)$] and the impurity scattering. For simplicity we also assume the two metals to be identical.

The three independent material parameters, introduced above, define two important length scales: (i) the 2D screening length $1/\kappa = (2\pi e^2\nu)^{-1}$, and (ii) the mean free path $\ell = v_F\tau$. Typically, the screening length is very short (on the atomic scale), so we assume $\kappa\ell \gg 1$ and $\kappa d \gg 1$. The mean free path can vary from several nanometers to several microns, and may be larger or smaller than the separation d between the two layers.

B. Summary of results

In the isotropic model, formulated above, Eq. (1) represents a two-dimensional integral over q and ω. This integral can be rather straightforwardly evaluated numerically, but much better insight into relevant physics is obtained by studying different asymptotics of the integral in various parameter regimes. The latter approach was adopted in Ref. [29] using the two limiting expressions (5a) and (5b). It turns out, however, that these expressions miss the plasmon contribution.

We show schematically the behavior of $\mathcal{T}(q, \omega)$ [Eq. (2)] in the (q, ω) plane for $d \ll \ell$ in Fig. 1. At large $\omega \gtrsim \max\{T_1, T_2\}$, the integrand in Eq. (1) is suppressed by the Bose function at ω, and this cutoff may be positioned anywhere in Fig. 1, depending on the temperatures. At large $q \gtrsim 1/d$, the integrand is suppressed by e^{-qd} in the numerator of Eq. (3); Fig. 1 corresponds to the case $d \ll \ell$, but for larger d the spatial cutoff may shift to the diffusive shaded area.

The strongly coupled plasmon modes (in the case of identical layers, symmetric and antisymmetric, denoted by "$\pm$") manifest themselves as poles of $V_{12}(q, \omega)$. In the clean limit, $\tau \to \infty$, the plasmon frequencies are real and given by (for $q \ll \kappa$)

$$\omega_{\pm} = v_F \sqrt{\kappa q \frac{1 \pm e^{-qd}}{2}}. \tag{6}$$

At finite τ, but such that $\omega_{\pm}\tau \gg 1$, the poles acquire a small imaginary part, so $|V_{12}(q, \omega)|^2$ is peaked around the dispersions (6). At $\omega_{\pm}\tau \lesssim 1$, when the diffusive expression (5b) applies, the plasmons are overdamped and do not produce a separate contribution to the integral. In the strictly clean limit, $\tau \to \infty$, their contribution vanishes as well, since for $\omega > v_F q$ the integrand vanishes because $\text{Im}\, \Pi(q, \omega > v_F q) = 0$.

For temperature-independent $\Pi(\mathbf{q}, \omega)$, Eq. (1) naturally splits into the difference $J(T_1, T_2) = J(T_1) - J(T_2)$. A detailed analysis of different asymptotic regimes for the integral in Eq. (1) (given in the next section) results in several asymptotic expressions for $J(T)$:

$$J_{\text{lc}}(T) = \frac{\pi^2}{60} \frac{T^4}{v_F^2 (\kappa d)^2} \ln \frac{v_F}{Td}, \tag{7a}$$

$$J_{\text{hc}}(T) = \frac{\pi^2}{900} \frac{v_F}{d^3} \frac{T}{(\kappa d)^2}, \tag{7b}$$

$$J_{\text{lp}}(T) = \frac{\zeta(3)}{4\pi} \frac{T^3}{D\kappa d}, \tag{7c}$$

$$J_{\text{hp}}(T) = \frac{T}{16\pi \tau d^2} \mathcal{L}(\ell^2 \kappa / 4d), \tag{7d}$$

$$J_{\text{ld}}(T) = \frac{\zeta(3)}{8\pi} \frac{T^3}{D\kappa d}, \tag{7e}$$

$$J_{\text{hd}}(T) = \frac{1}{16\pi} \frac{D\kappa}{d^3} T. \tag{7f}$$

Subscripts "l" and "h" denote low and high temperature, while "c," "p," and "d" stand for clean, plasmonic, and diffusive, respectively. Here $\zeta(x)$ is the Riemann zeta function, and $\mathcal{L}(x)$ is a slow logarithmic function defined in Eq. (16). For moderate values of $\ln(\ell^2 \kappa / 4d) < 10$, it can be approximated with 10% precision as

$$\mathcal{L}(x) \approx \frac{4 \ln^3 x}{1 + (\ln x)/\ln(1 + \ln x)}. \tag{8}$$

Figure 2 schematically shows the domains of validity for expressions (7) in the $(1/d, T)$ plane. The clean and diffusive regimes were also identified in Ref. [29].

Each of the above regimes is characterized by typical scales of q and ω, which dominate the integral in Eq. (1). Namely, $q \sim 1/d$ in all high-temperature cases, $T/(v_F \sqrt{\kappa d})$ in the low-temperature plasmonic case, and $\sqrt{T/D\kappa d}$ in the low-temperature diffusive case. In the low-temperature clean case, the momentum integral is logarithmic, determined by the whole interval $T/v_F < q < 1/d$. In order for the results to be valid for some real sample with a finite in-plane size, this size should be much larger than the corresponding convergence scale $1/q$ specified here. The convergence scale of the frequency integral is $\omega \sim T$ in all low-temperature cases, while
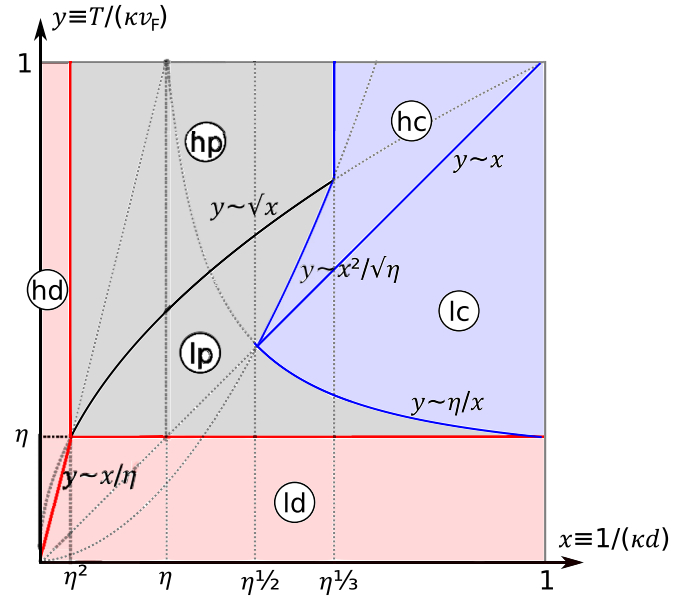

FIG. 2. The regions in which clean (blue), plasmonic (grey), and diffusive (red) contributions are dominant in the heat current and the domains of validity for asymptotic expressions (7) in the parameter plane $(1/d, T)$, shown in the dimensionless variables $x \equiv 1/(\kappa d)$, $y \equiv T/(\kappa v_F)$. The boundaries between the regimes are governed by a single dimensionless material parameter $\eta = 1/(\kappa \ell) \ll 1$.

in the high-temperature cases it is $v_F \sqrt{\kappa/d}$ in the plasmonic case, $D\kappa/d$ in the diffusive case, and v_F/d in the clean case.

The specified q and ω scales also determine the condition for the validity of the Coulomb limit, which is $q \gg \omega/c$. In both clean regimes this condition reduces to $v_F \ll c$, which is always true. In both plasmonic regimes, the condition $q \gg \omega/c$ translates into a condition on the antisymmetric plasmon velocity, $v_F \sqrt{\kappa d} \ll c$. In the high-temperature diffusive regime, the condition is $v_F \kappa \ell \ll c$. Finally, in the low-temperature diffusive regime, the condition reads

$$\frac{Td}{c} \ll \frac{c}{v_F} \frac{1}{\kappa \ell}. \tag{9}$$

Note that none of these cases corresponds to comparing the separation d to the thermal wavelength c/T.

Plasmon contributions dominate in a parametrically wide region of the parameter space. Crucially, this behavior is captured by neither of the limiting expressions (5a) or (5b), but by the leading term of the small q expansion of the full expression (4). This is equivalent to using the Drude expression for the optical conductivity, $\sigma(\omega) = e^2 \nu D/(1 - i\omega\tau)$, that is, neglecting the spatial dispersion. We find that the spatial dispersion can also be neglected to describe the diffusive contribution. It becomes important only in the clean region where $J(T)$ is dominated by the hatched area in Fig. 1. In all three high-temperature regions, one can approximate the Bose distribution as $\mathcal{N}(\omega) \approx T/\omega$, so the density fluctuations can be treated classically and the resulting $J(T) \propto T$.

Expressions (7) were obtained for two identical metallic layers. If they are different, but the material parameters ν, v_F,

and τ are of the same order, our expressions are still valid as order-of-magnitude estimates. In particular, this applies to the plasmon contribution: the plasmons remain strongly hybridized even when the bare plasmonic dispersions of each layer do not match exactly. The case when the two materials are strongly different is beyond the scope of the present paper.

C. Generality of the results

The above results were derived for a very specific model of a 2D metal, described in Sec. II A. We now discuss how sensitive these results are to the details of the model.

One assumption used in derivation of Eq. (4) is that the electron scattering is dominated by short-range impurities. If it is relaxed, Eq. (4) is not valid quantitatively (see the Appendix). Note, however, that in the clean regimes (lc, hc), the result is not sensitive to the electron scattering at all, so expressions (7a) and (7b) remain quantitatively valid. In the other four regimes (ld, hd, lp, hp, and crossovers between them) it is sufficient to use only the local Drude conductivity $\sigma(\omega)$, which is very general. Thus, all information about the electron scattering, needed to describe these regimes, is encoded in the momentum relaxation time τ, and expressions (7c)–(7f) remain valid for any disorder (weak enough not to induce Anderson localization effects).

If the electron momentum relaxation is due to some inelastic scattering mechanism (such as electron-phonon), the relaxation time acquires a temperature dependence. Then, the screened Coulomb interaction depends on both temperatures through the respective polarization operators, so the heat current can no longer be written as $J(T_1, T_2) = J(T_1) - J(T_2)$. Even if the layers are made of the same material, when kept at different temperatures, they should be treated as different because their inelastic scattering times are different. If the two temperatures are of the same order, our expressions can still be used as order-of-magnitude estimates; the situation when they are strongly different is beyond the scope of this paper.

Another strong assumption behind Eq. (4) is the strictly 2D character of the electron motion, valid for atomically thin materials. For a metallic slab whose thickness h exceeds a few Fermi wavelengths, several electronic transverse modes contribute to the density response, making Eq. (4) invalid even in the clean limit. However, in electrodynamics, the conditions for a material slab to be described as an infinitely thin layer with *some* density response function $\Pi(\mathbf{q}, \omega)$ are much weaker. Namely, (i) the normal component of the electric field should not penetrate inside the slab, since the description in terms of a 2D density response function implies that the electrons respond only to the in-plane component, and (ii) the in-plane component should be approximately constant over the slab thickness. The first condition is usually satisfied in metals at frequencies below the bulk plasma frequency ω_p (typically, several eV), when the normal electric field component is screened on the length scale of the bulk screening radius (typically, on the atomic scale). The second condition requires the layer thickness h to be smaller compared to both $1/q$ and the skin depth $\delta(\omega)$ in the corresponding regime of the skin effect [37]. The relevant values of q and ω are

determined by the convergence scales of the corresponding integrals [as specified after Eqs. (7)]. Note that in all cases the 2D response function must satisfy $\Pi(q \ll 1/h, \omega = 0) = -\nu$, where $\nu = \nu_{3D} h$ is determined by the bulk density of states per unit volume, ν_{3D}. This defines $\kappa = 2\pi e^2 \nu$, as before. Since thicker layers imply larger ν and κ, the convergence scale of q may become quite small in some of the discussed regimes, so the Coulomb limit condition $q \gg \omega/c$ may become violated for h not small enough.

III. DERIVATION OF THE RESULTS

For identical layers, Eq. (2) can be written in a slightly simpler form:

$$\frac{\mathcal{T}}{2} = \left| \frac{\mathrm{Im}\, \Pi\, v_q e^{-qd}}{[1 - \Pi v_q (1 - e^{-qd})][1 - \Pi v_q (1 + e^{-qd})]} \right|^2. \quad (10)$$

Below, we analyze the contributions from the three regions of the (q, ω) plane in Fig. 1: the vicinity of the plasmon dispersions, the diffusive region, and the clean region. In each region, we separate two temperature regimes. We see that in the high-temperature regimes, the integral (1) is dominated by a certain frequency scale, different in each regime, but always determined by $q \sim 1/d$ and being much lower than the temperature. Then, the thermal cutoff plays no role, and one can approximate the Bose distribution $\mathcal{N}(\omega) \approx T/\omega$ [that is why $J(T) \propto T$ in Eqs. (7b), (7d), and (7f)]. In the low-temperature regimes, it is the e^{-qd} cutoff which is ineffective, so the frequency integral is determined by $\omega \sim T$, while the q integral converges on a scale much smaller than $1/d$. Then one can approximate $e^{-qd} \approx 1$ in the numerator of Eq. (10) and $1 + e^{-qd} \approx 2$, $1 - e^{-qd} \approx qd$ in the denominator.

A. Plasmon contribution

As discussed earlier, the plasmon contribution comes from the region $\omega > 1/\tau, v_F q$, since otherwise the plasmons are overdamped. At such frequencies one can expand $\Pi(q, \omega)$ to the leading order in q and approximate

$$\Pi v_q = \frac{v_F^2 \kappa q}{2\omega(\omega + i/\tau)}, \quad (11)$$

which corresponds to neglecting the spatial dispersion in the conductivity which takes the Drude form, $\sigma(\omega) = e^2 \nu D/(1 - i\omega\tau)$. (Generally, the spatial dispersion can be neglected when $q \ll \max\{\sqrt{\omega/D}, \omega/v_F\}$.) The integral is then dominated by the vicinities of the two plasmon dispersions, where one of the factors in the denominator of Eq. (10) is small. The plasmon contribution exists only if the plasmon frequencies $\omega_\pm$ at $q \sim 1/d$ (when the spatial cutoff becomes effective) exceed $1/\tau$. This gives a condition $d \ll \kappa \ell^2$ (the vertical line $x = \eta^2$ in Fig. 2).

Let us first consider the temperature interval $1/\tau \ll T \ll v_F \sqrt{\kappa/d}$ (the two inequalities are consistent when $d \ll \kappa \ell^2$), where the thermal cutoff plays first leaving the spatial cutoff e^{-qd} ineffective. Then one can expand e^{-qd} and perform the q integration first, approximating the integrand by a Lorentzian in the vicinity of each pole. The remaining frequency integral

can be calculated exactly:

$$J_{\text{lp}}(T) = \int_0^\infty \frac{\omega \, d\omega/\pi}{e^{\omega/T} - 1} \frac{\omega}{4\tau v_F^2 \kappa^2} \left(\frac{\omega^2}{v_F^2} + \frac{\kappa}{d} \right)$$
$$= \frac{T^3}{2\pi \tau v_F^2 \kappa d} \left[\zeta(3) + 12 \, \zeta(5) \frac{T^2 d}{v_F^2 \kappa} \right]. \quad (12)$$

Here $\zeta(x)$ is the Riemann zeta function, and the first (second) term in the square brackets comes from the antisymmetric (symmetric) plasmons, respectively. In the considered region $1/\tau \ll T \ll v_F \sqrt{\kappa/d}$ the symmetric contribution is always small compared to the antisymmetric one.

At high temperatures, $T \gg v_F \sqrt{\kappa/d}$, the integral is determined by the spatial factor e^{-qd}, while the thermal cutoff is ineffective so the Bose distribution $\mathcal{N}(\omega) \approx T/\omega$. Taking expression (11), Eq. (10) can also be written in the form suitable for integration over ω:

$$\mathcal{T}(q, \omega) = \frac{\omega^2}{2\tau^2} \frac{(\omega_+^2 - \omega_-^2)^2}{|\omega(\omega + i/\tau) - \omega_+^2|^2 |\omega(\omega + i/\tau) - \omega_-^2|^2}. \quad (13)$$

Note that one cannot just do two separate Lorentzian integrals because the separation $\omega_+^2 - \omega_-^2$ becomes exponentially small at $q \gtrsim 1/d$. Fortunately, the ω integral can be calculated exactly:

$$\int_0^\infty \frac{d\omega}{\pi} \mathcal{T}(q, \omega) = \frac{1}{2\tau} \frac{(\omega_+^2 - \omega_-^2)^2}{(\omega_+^2 - \omega_-^2)^2 + 2(\omega_+^2 + \omega_-^2)/\tau^2}. \quad (14)$$

Then the q integral reads as

$$J_{\text{hp}}(T) = \frac{T}{\tau} \int_0^\infty \frac{q \, dq}{4\pi} \frac{\ell^2 \kappa q}{\ell^2 \kappa q + 2 e^{2qd}}, \quad (15)$$

leading to Eq. (7d) with the function $\mathcal{L}(x)$ defined as

$$\mathcal{L}(x) \equiv \int_0^\infty \frac{u^2 \, du}{u + e^u/x}. \quad (16)$$

B. Diffusive contribution

Let us focus on the contribution from the shaded region in Fig. 1: $q \ll 1/\ell$, $\omega \ll 1/\tau$. Then one can use expression (5b) for $\Pi(q, \omega)$, and since we are interested in $q \lesssim 1/d \ll \kappa$, we have $Dq^2 \ll D\kappa q(1 \pm e^{-qd})/2$, which is again equivalent to neglecting the spatial dispersion in the conductivity. Thus, we can write

$$\mathcal{T} \approx \frac{2\omega^2 (D\kappa q)^2 e^{-2qd}}{[\omega^2 + (D\kappa q)^2 (1 + e^{-qd})^2][\omega^2 + (D\kappa q)^2 (1 - e^{-qd})^2]}. \quad (17)$$

At low frequencies, when the momentum integral should converge on some scale $q \ll 1/d$, the two factors in the denominator are strongly different, and it is the second factor that determines the convergence scale $q \sim \sqrt{\omega/(D\kappa d)}$. When $\omega \sim T \ll D\kappa/d$ this scale is indeed much smaller than $1/d$, so we expand e^{-qd}, integrate over q, then over ω, and arrive at Eq. (7e). Moreover, the convergence scale $\sqrt{T/(D\kappa d)} \ll 1/\ell$ provided that $T \ll \kappa d/\tau$. Thus, since we always assume

$\kappa d \gg 1$, Eq. (7e) is valid for the diffusive contribution everywhere below the horizontal line $y \sim \eta$ ($T \sim 1/\tau$) in Fig. 2.

For $T \gg D\kappa/d$, the q integral is dominated by $q \sim 1/d$. Then the typical frequency scale of Eq. (17) is $\omega \sim D\kappa/d$, so for $T \gg D\kappa/d$ the thermal cutoff is ineffective. Then we approximate $\mathcal{N}(\omega) \approx T/\omega$, straightforwardly integrate over ω, then over q, and arrive at Eq. (7f). Note that Eq. (7f) is valid even at $T \gg 1/\tau$ provided that the convergence scale $\omega \sim D\kappa/d \ll 1/\tau$, that is, to the left of the vertical line $x = \eta^2$ in Fig. 2. As we have seen, to the right of this line the plasmon contribution becomes important.

C. Clean contribution

For $T \gg 1/\tau$, $1/d \gg 1/\ell$, one should take into account the contribution from the hatched area in Fig. 1. Here one can take the limit $\tau \to \infty$ and use expression (5a) for $\Pi(q, \omega)$. Then, in the integration region $\omega < v_F q$, $\text{Re} \, \Pi(q, \omega) v_q = -\kappa/q$, so one can neglect unity in both factors in the denominator of Eq. (10) and write

$$\mathcal{T} \approx \frac{2}{v_q^2} \left| \frac{\text{Im} \, \Pi}{\Pi^2} \right|^2 \frac{e^{-2qd}}{(1 - e^{-2qd})^2} = \frac{\omega^2 (v_F^2 q^2 - \omega^2)}{2 v_F^4 \kappa^2 q^2 \sinh^2 qd}. \quad (18)$$

For $T \gg v_F/d$ we approximate $\mathcal{N}(\omega) \approx T/\omega$, straightforwardly integrate over ω between 0 and $v_F q$, then integrate over q, and arrive at Eq. (7b). For $T \ll v_F/d$, in most of the integration region we have $\omega \sim T \ll v_F q$, so the upper limit $\omega < v_F q$ is not important except for the narrow region $q \sim T/v_F$ which determines the lower cutoff of the logarithmic q integral:

$$J_{\text{lc}}(T) = \frac{1}{4\pi^2} \int_0^\infty \frac{\omega^3 \, d\omega}{e^{\omega/T} - 1} \int_{\sim T/v_F}^\infty \frac{q \, dq}{\sinh^2 qd}, \quad (19)$$

leading to Eq. (7a).

In the region $T \gg 1/\tau$, $1/d \gg 1/\ell$, the clean contribution and the plasmon contribution both exist and should be added up, since they come from two distinct regions in the (q, ω) plane. Thus, to determine the dominant asymptotics, one can combine the four expressions (7a)–(7d) as $J = \max\{\min\{J_{\text{lp}}, J_{\text{hp}}\}, \min\{J_{\text{lc}}, J_{\text{hc}}\}\}$, which results in the complicated shape of the boundary between the clean and plasmonic regions in Fig. 1.

IV. NUMERICAL VERIFICATION

In order to illustrate the various behaviors and crossovers indicated in Fig. 2, and to verify our asymptotic expressions (7), we evaluate numerically the integral in Eq. (1) using the full response function (4).

To describe clean and plasmonic regimes, we take parameters typical for doped graphene. If the temperature is well below the Fermi energy ϵ_F (counted from the Dirac point), the electrons in graphene can be viewed as a conventional 2D electron gas with the density of states per unit area $\nu = 2|\epsilon_F|/(\pi v_F^2)$, including the valley and spin degeneracies. In Fig. 3, we plot $J(T)$ for two sets of parameters corresponding to $1/(\kappa d) \approx \eta^{1/3} = 0.034$, and $1/(\kappa d) \approx \eta^{1/2}$ [we remind that $\eta \equiv 1/(\kappa \ell)$].

To study the diffusive crossover for realistic materials, we introduce dielectric screening. Equation (3), written for
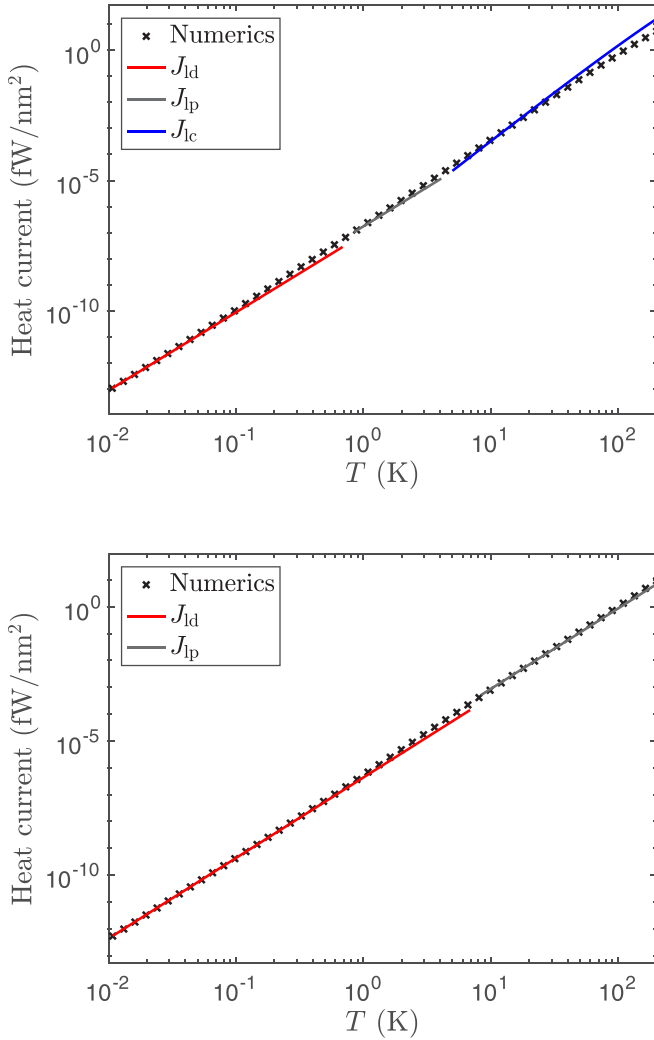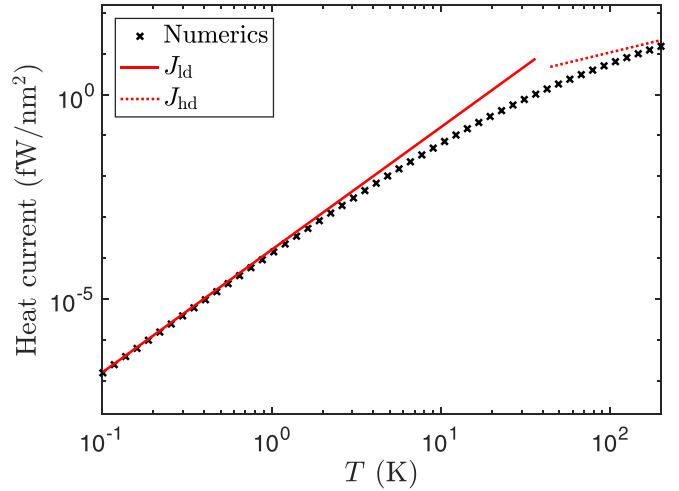
FIG. 4. Power per unit area as a function of temperature, $J(T)$, for $\nu = 2.1$ eV^{-1} nm^{-2}, $v_F = 0.94 \times 10^5$ m/s, and $\ell = 2$ μm, $d = 100$ nm, characteristic of two WSe$_2$ monolayers with Fermi energy $\epsilon_F = 50$ meV, embedded in boron nitride with $\epsilon_\parallel = 7$, $\epsilon_\perp = 5$ [38] (the 2D screening length $\kappa^{-1} = 0.3$ nm, the Fermi momentum $p_F = 1.6$ nm^{-1}). The black crosses show the numerical integration results, while the solid and dashed red lines represent expressions (7e) and (7f), respectively.

In Fig. 4 we show the crossover between low- and high-temperature diffusive asymptotics (7e) and (7f) for two hole-doped tungsten diselenide monolayers embedded in boron nitride. The valence band of WSe$_2$ is parabolic with the hole effective mass m_h being about half of the free electron mass. The spin degeneracy is lifted by a strong spin-orbit coupling, so only valley degeneracy remains, and the density of states per unit area is $\nu = m_h/\pi$. We take $\epsilon_F = 50$ meV and a very short $\ell = 2$ nm, still consistent with $\kappa\ell \gg 1$, $p_F\ell \gg 1$. The taken separation $d = 100$ nm corresponds to $1/(\kappa d) = 0.006$, well below $\eta^2 = 0.1$, and hence to the diffusive region.

We are not showing the high-temperature clean and plasmonic regimes; for realistic material parameters, they correspond to temperatures so high that the assumptions behind our model (degenerate Fermi gas, near-field Coulomb regime) are no longer valid. However, we checked numerically the validity of the asymptotic expressions (7b) and (7d) for $J_{hc}(T)$ and $J_{hp}(T)$.

V. COMPARISON TO CIRCUIT THEORY

Often, complicated structures can be described in terms of effective electric circuits made of lumped elements (capacitors, inductors, and resistors). In this approach, all details of the structure's geometry are hidden inside the effective circuit parameters, resulting in a much simpler description (provided that such reduction is valid). The theory of heat transfer in electric circuits was developed in Ref. [30]. In the circuit analog of the fluctuational electrodynamics, dissipative circuit elements represent thermal baths and provide thermal voltage fluctuations (Johnson-Nyquist noise), while reactive elements mediate the electromagnetic interactions, resulting in energy exchange between the baths. We check now whether such a

FIG. 3. Power per unit area as a function of temperature, $J(T)$, for $\nu = 0.29$ eV^{-1} nm^{-2}, $v_F = 10^6$ m/s, and $\ell = 10$ μm, $d = 10$ nm (upper panel), $\ell = 1$ μm, $d = 20$ nm (lower panel), characteristic of two graphene monolayers with Fermi energy $\epsilon_F = 0.2$ eV, suspended in vacuum (the 2D screening length $\kappa^{-1} = 0.4$ nm, the Fermi momentum $p_F = 0.3$ nm^{-1}). The black crosses show numerical results, while the red, black, and blue solid lines represent Eqs. (7e), (7c), and (7a), respectively.

metallic layers surrounded by vacuum, can be generalized to the situation when the layers are embedded in a dielectric medium. This generalization is particularly simple when the medium is characterized by a uniaxially anisotropic dielectric constant, $\varepsilon_\parallel$ in the plane parallel to the layers, and $\varepsilon_\perp$ along the z direction, perpendicular to the layers. The solution of the Poisson equation in such a medium gives the 2D Coulomb potential at a distance z from a charged layer:

$$v_{\mathbf{q}}(z) = \frac{2\pi e^2}{\sqrt{\varepsilon_\parallel \varepsilon_\perp}\, q}\, e^{-\sqrt{\varepsilon_\parallel/\varepsilon_\perp}\, q|z|}. \tag{20}$$

Thus, in all expressions (7) it is sufficient to rescale

$$\kappa \to \frac{\kappa}{\sqrt{\varepsilon_\parallel \varepsilon_\perp}}, \quad d \to \sqrt{\varepsilon_\parallel/\varepsilon_\perp}\, d. \tag{21}$$
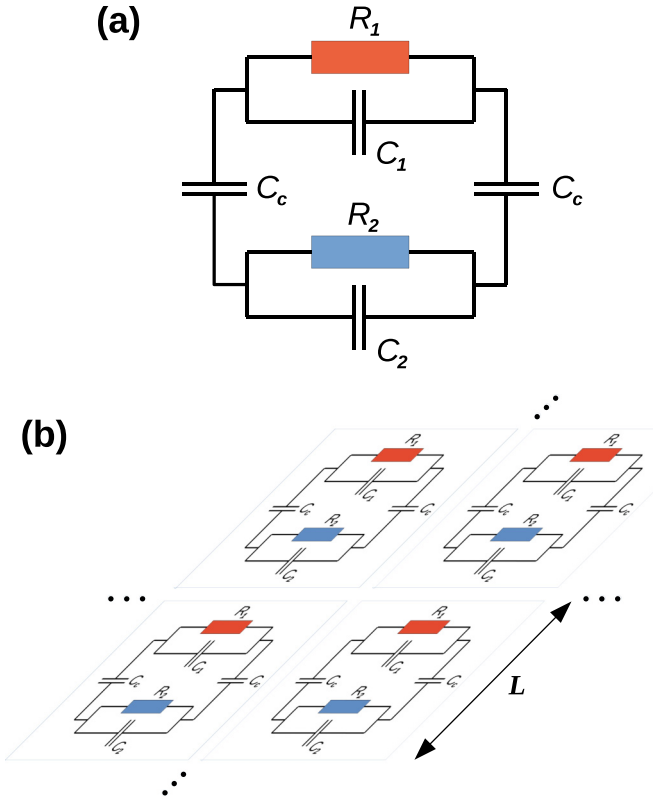
(a)

(b)

FIG. 5. Circuit representation of two parallel metallic layers. (a) The elementary RC circuit described by Eq. (23). (b) The 2D system represented as a tiling of the elementary circuits, each one corresponding to a region of size L.

circuit approach can be applied to the heat transfer between two metallic layers.

Let us focus on the diffusive regime. As we have seen, the dynamics of density excitations is overdamped in this regime, so it is natural to consider a circuit made of resistors and capacitors only, such as shown in Fig. 5(a). Indeed, the electronic excitations in each layer constitute a dissipative bath analogous to a resistor. To mimic charge oscillations within each layer, the resistor should be shunted by a capacitor. The Coulomb interaction between the layers resembles that between the plates of a capacitor, so the two RC contours are connected by two coupling capacitors C_c.

For the circuit in Fig. 5(a), the power transferred from the first to the second resistor is given by [30]

$$P(T_1, T_2) = \int_0^\infty \frac{d\omega}{\pi} \omega [\mathcal{N}_1(\omega) - \mathcal{N}_2(\omega)] \mathcal{T}(\omega), \quad (22)$$

$$\mathcal{T}(\omega) = \frac{2 \operatorname{Re} Z_1(\omega) \operatorname{Re} Z_2(\omega)}{|Z_1(\omega) + Z_2(\omega) - 2/(i\omega C_c)|^2}, \quad (23)$$

where $Z_{1,2}(\omega) = (1/R_{1,2} - i\omega C_{1,2})^{-1}$ is the impedance of each RC contour. As before, for simplicity we assume the two subsystems to be identical: $R_1 = R_2 = R$, $C_1 = C_2 = C$. Writing the transmission as

$$\mathcal{T}(\omega) = \frac{\omega^2 (RC_c)^2}{2[1 + \omega^2 (RC)^2][1 + \omega^2 R^2 (C + C_c)^2]}, \quad (24)$$

we obtain the following asymptotic expressions for the transferred power $P(T_1, T_2) = P(T_1) - P(T_2)$:

$$P(T) = \frac{C_c^2}{4(C + C_c)(2C + C_c)} \frac{T}{RC}, \quad T \gg \frac{1}{RC}, \quad (25a)$$

$$P(T) = \frac{\pi}{12} T^2, \quad \frac{1}{R(C + C_c)} \ll T \ll \frac{1}{RC}, \quad (25b)$$

$$P(T) = \frac{\pi^3}{30} T^4 (RC_c)^2, \quad T \ll \frac{1}{R(C + C_c)}. \quad (25c)$$

The intermediate "universal" regime, where the power depends only on the temperature but not on the circuit, is present only when $C \ll C_c$.

To relate results (25) to those of Sec. III, it is important to realize that while Eqs. (25) give the full transferred power, Eqs. (7) give the power *per unit area*. To make a meaningful comparison we must therefore invoke a length scale L, such that the infinite sample can be divided into squares of size L. Then Eqs. (25) describe the power transferred in each square, and the contributions of different squares can be added up independently, as schematically shown in Fig. 5(b). Thus, the relevant length scale should be associated with the typical convergence scale of the q integral in Eq. (1).

In the diffusive regime, it is natural to associate the resistance R to the resistance per square $1/\sigma$ of each metallic layer, $R \sim 1/\sigma = 2\pi/(\kappa D)$. The coupling capacitance is associated to the geometric capacitance between the two layers, $C_c \sim L^2/d$, where the in-plane length scale $L \gtrsim d$ must be invoked because the capacitance is proportional to the area. The capacitance C should be associated to an intrinsic property of each layer, such that RC corresponds to the characteristic relaxation time of charge density modulations. In a 2D system, this time depends on the in-plane length scale L and is given by L/σ. This gives $C \sim L$. Recalling the convergence scales of the q integral in Sec. III B, we associate $L \sim \sqrt{D\kappa d/T}$ and $L \sim d$ in the low- and high-temperature diffusive regimes, respectively. Then Eqs. (25c) and (25a) match Eqs. (7e) and (7f) at low and high temperatures, respectively. Expression (25b) does not correspond to any parametric region because at $T \gg D\kappa/d$ the two capacitances become of the same order.

To summarize, while the proposed effective RC circuit does capture the qualitative picture of the heat transfer in the diffusive regime, one cannot completely disregard the 2D geometry of the system. This geometry manifests itself in the appearance of the length scale L, which must supplement the circuit picture in order to reproduce the temperature dependence of the heat current. Moreover, this length scale is temperature dependent, so one cannot represent a given system by a given array of elementary circuits in the whole temperature range. This strongly limits the usefulness of the circuit analogy. For this reason, we do not consider the other two regimes (plasmonic and clean), whose modeling would require a more complicated circuit; this would have to include also inductors in order to mimic the physics of weakly damped density excitations.

VI. DISCUSSION OF EXPERIMENTS

Let us now discuss some relevant experiments from the viewpoint of the theory presented in this paper. In the lit-

erature we are aware of, experiments in the planar geometry are not very numerous; apparently, a tip-to-plane or sphere-to-plane configurations are easier to realize in a controllable way.

In a recent experiment [14], two monolayer graphene sheets were placed on insulating silicon (dielectric constant $\varepsilon = 11.7$) and separated by a vacuum gap of width $d = 430$ nm. The experiment was performed around room temperature, and the linear thermal conductance per unit area, $dJ/dT = 30$ W m^{-2} K^{-1}, was measured. Importantly, the doping level was sufficiently high, $\epsilon_F = 0.27$ eV, so the peculiarities of the Dirac spectrum can be safely ignored and the electrons can be treated as the usual 2D electron gas with the 2D screening length $\kappa^{-1} = 2.8$ Å. The single-layer plasmon frequency at $q = 1/d$ is $v_F\sqrt{\kappa/(\varepsilon+1)d} \sim 3 \times 10^{13}$ s^{-1}, slightly below room temperature ($T = 300$ K corresponds to $\omega = 3.93 \times 10^{13}$ s^{-1}) and above $1/\tau = 10^{13}$ s^{-1} [14]. These conditions correspond to the crossover between the low- and high-temperature plasmon regimes. The plasmon velocity is about 20 times smaller than the speed of light, so the heat transfer is well within the Coulomb limit.

To give a theoretical estimate for the thermal conductance, one has to account for the strong dielectric contrast (vacuum vs silicon) between the two sides of each graphene layer. This amounts to replacing $v_{\mathbf{q}}\Pi \to v_{\mathbf{q}}\Pi + (1-\varepsilon)/2$ in the denominator of Eq. (3). In the high-temperature plasmon regime, one should multiply $2e^{2qd}$ in the denominator of Eq. (15) by

$$\frac{[(\varepsilon+1)^2 - (\varepsilon-1)^2 e^{-2qd}][(\varepsilon+1) - (\varepsilon-1)e^{-2qd}]}{8},$$

so ε enters only inside the logarithmic function $\mathcal{L}$. Setting $\mathcal{L} = 1$ gives $dJ/dT = 11$ W m^{-2} K^{-1}, which agrees by an order of magnitude with the experimental value.

Several works have been dedicated to transfer between finite-thickness layers of conventional metals, to be reviewed below. In each case, one should decide whether these layers should correspond to the 2D or bulk limit, according to the discussion in Sec. II C. In all cases, the separation d between the sheets is larger than their thickness h, so the criterion $h \ll 1/q$ is always fulfilled. Thus, the main issue is the comparison between the layer thickness and the skin depth.

In the pioneering study in Ref. [39], heat transfer between two $h = 100$-nm-thick chromium plates was measured at room temperature, probing separation dependence of heat flux over scales $d = 1–8$ μm. Reference [39] provides no information on electron scattering, but if we take a typical value for the mean free path $\ell_{3D} = 10$ nm and use the values $\nu_{3D} = 0.06$ eV^{-1} Å^{-3} [40], $v_F = 0.5 \times 10^6$ m/s [41], we obtain the normal skin depth $\delta = 50$ nm at frequency $\omega = 3.93 \times 10^{13}$ s^{-1}, corresponding to $T = 300$ K. Thus, such plates cannot be treated as two dimensional.

In Ref. [42], heat transfer was studied for two tungsten layers $h = 150$ nm thick on alumina substrate, over a wide range of separations $d = 1–300$ μm, with the cold layer held at 5 K, while the hot layer temperature was varied in the range 10–40 K. Specifically, at small separations, the observed temperature and distance dependence was $J(T) \propto T^{1.5}/d^{2.5}$. The measured dc conductivity of the material, corresponding to $4\pi\sigma_{3D} = 0.6 \times 10^{18}$ s^{-1}, was constant in the temperature

range 4–77 K, identifying the static disorder as the dominant source of electron scattering with $\tau = 6$ fs and $\ell_{3D} = 6$ nm ($v_F = 10^6$ m/s from Ref. [43]). This gives the normal skin depth corresponding to $T = 40$ K as $\delta = 240$ nm, and even longer at lower temperatures. Thus, one could expect this experiment to correspond to the low-temperature diffusive 2D regime, whose conditions can be written as $T \ll 1/\tau$, $T \ll 4\pi\sigma_{3D}h/d$. However, if we check the Coulomb limit condition (9) estimating $\nu_{3D} = 10^{47}$ J^{-1} m^{-3} = 0.016 eV^{-1} Å^{-3} from $\omega_p^2 = 4\pi\nu_{3D}v_F^2/3$ with $\omega_p = 0.97 \times 10^{16}$ s^{-1} [42], we obtain $\kappa\ell = 1.3 \times 10^5$, and the two sides of the inequality become comparable only for the smallest $d = 1$ μm and $T = 10$ K. Thus, the whole experimental curve corresponds to the regime when retardation is important, so it is natural that the experimental result is not consistent with $J_{\mathrm{ld}}(T) \propto T^3/d$.

Linear response thermal conductance $dJ(T)/dT$ at room temperature was measured in Ref. [10] for two gold layers $h = 100$ nm thick with separations 60 nm $< d <$ 10 μm. At small separations a dependence $1/d^\alpha$ with α between 2 and 3 was observed. These results were successfully interpreted using the Drude model parameters $\omega_p = 0.6 \times 10^{16}$ s^{-1} and $\tau = 6$ fs [44], which give the skin depth $\delta = 50$ nm. Estimating ν_{3D} from ω_p and $v_F = 1.4 \times 10^6$ m/s [43], we again obtain that inequality (9) is not satisfied, although for the smallest separation the two sides become comparable.

Finally, a recent preprint [15] reports a series of measurements of heat transfer between two aluminum films on silicon, separated by a vacuum gap $d = 215$ nm. The film thicknesses h were varied in the interval 13–79 nm. The experiment was performed around room temperature, and for the thinnest films with $h = 13$ nm the linear thermal conductance per unit area, $dJ/dT = 60$ W m^{-2} K^{-1}, was measured. For thicker layers, the thermal conductance was slightly smaller (about 50 W m^{-2} K^{-1} for $d = 79$ nm). No information on electron scattering was given in Ref. [15], but the 13-nm-thick sample was thinner than the smallest possible skin depth $c/\omega_p = 30$ nm (with $\omega_p = 0.97 \times 10^{16}$ s^{-1} [44]), so one would expect it to be in the 2D limit. From $\nu_{3D} = 1.45 \times 10^{47}$ J^{-1} m^{-3} = 0.023 eV^{-1} Å^{-3} [45], we obtain the 2D screening length $1/\kappa = 3.6 \times 10^{-3}$ Å. If we take a typical value $\ell = 10$ nm, we nominally find the system to be in the low-temperature diffusive limit. Still, with $v_F = 1.6 \times 10^6$ m/s [43], inequality (9) is not satisfied, with the left-hand side a few times larger than the right-hand side.

To summarize, metallic layers studied in experiments may have thicknesses a few times larger or a few times smaller than the skin depth, but the inequality is never very strong. More stringent is condition (9), which enables one to neglect retardation effects and use the Coulomb limit. For typical metals, it requires temperatures below 100 K, while h and d should be in the range of tens of nanometers.

VII. CONCLUSIONS

We have studied the problem of heat transfer between two thin parallel metallic layers, mediated by the Coulomb interaction. Using a simple model for a 2D electron gas subject to scattering on short-range impurities, we described the crossover between clean and diffusive limits and showed that strongly coupled surface plasmons dominate the heat

transfer in a parametrically wide region at sufficiently high temperatures, but their contribution is suppressed in both the clean and diffusive limits. We also clarified the role of the spatial dispersion in the optical conductivity, which turns out to be important only in the clean limit. In all other regimes, the effect of disorder is correctly captured by the relaxation time in the Drude conductivity.

We have shown that in the diffusive limit, the heat transfer is quantitatively similar to that in an effective *RC* circuit. However, for this analogy to be meaningful, one must specify a length scale. This length scale should correspond to the size of regions within the infinite 2D sample where the transfer occurs independently. In other words, each region can be described by a separate circuit, and contributions from different regions add up. This length scale must be determined from the microscopic theory and turns out to be temperature dependent. This greatly limits the usefulness of the circuit analogy, especially when the two temperatures are strongly different.

Comparing the theory, developed here, to experimental data, available in the literature, we found that the theory qualitatively agrees with the experiment on heat transfer between two graphene sheets, which falls into the crossover between the low- and high-temperature plasmonic regimes. For finite-thickness layers of conventional metals, we find that, in all experiments that we are aware of, the parameters do not fulfill the stringent condition to be in the Coulomb limit when retardation can be neglected; still, this limit is realistic for nanometer-sized structures at low temperatures.

ACKNOWLEDGMENTS

APPENDIX: DENSITY RESPONSE FUNCTION FROM THE BOLTZMANN EQUATION

We assume the 2D electron gas to be in the good metallic regime, when the mean free path ℓ is much larger than the Fermi wavelength, so we may neglect localization effects. In this limit, and for perturbations smooth on the scale of the Fermi wavelength, the electron dynamics can be described by the semiclassical Boltzmann kinetic equation [46]. The electron distribution function $f_{\mathbf{p}}(\mathbf{r}, t)$ is assumed to depend on the 2D momentum $\mathbf{p}$ and the 2D position $\mathbf{r}$, while the dynamics in the third dimension is assumed to be completely frozen by a tight confinement. Then the kinetic equation reads

$$\frac{\partial f_{\mathbf{p}}}{\partial t} + \mathbf{v}_{\mathbf{p}} \cdot \frac{\partial f_{\mathbf{p}}}{\partial \mathbf{r}} + \mathbf{F} \cdot \frac{\partial f_{\mathbf{p}}}{\partial \mathbf{p}} = \text{St}[f], \quad (A1)$$

where $\mathbf{F}$ is an externally applied force, and $\mathbf{v}_{\mathbf{p}} = \partial \epsilon_{\mathbf{p}}/\partial \mathbf{p}$ is the electron group velocity determined by the energy dispersion $\epsilon_{\mathbf{p}}$. The collision integral on the right-hand side, written in the Born approximation,

$$\text{St}[f] = 2\pi n_{\text{i}} \int \frac{d^2 \mathbf{p}'}{(2\pi)^2} |U(\mathbf{p} - \mathbf{p}')|^2 \delta(\epsilon_{\mathbf{p}} - \epsilon_{\mathbf{p}'})(f_{\mathbf{p}'} - f_{\mathbf{p}}), \quad (A2)$$

is determined by the impurity concentration n_{i} and the Fourier transform of the impurity potential $U(\mathbf{p} - \mathbf{p}')$. Beyond the Born approximation, $|U(\mathbf{p} - \mathbf{p}')|^2$ should be replaced by the exact scattering amplitude, properly normalized.

In the absence of perturbations, the electrons are assumed to have the Fermi-Dirac distribution determined by the Fermi energy ϵ_F and the temperature T:

$$f_{\mathbf{p}}^{\text{eq}} = \frac{1}{e^{(\epsilon_{\mathbf{p}} - \epsilon_F)/T} + 1}. \quad (A3)$$

If a perturbing electrostatic potential $\varphi_{\mathbf{q},\omega} e^{i\mathbf{q}\mathbf{r} - i\omega t} + \text{c.c.}$ is applied to the 2D system (again, we neglect its dependence on the third coordinate), it enters Eq. (A1) via the associated electrostatic force $\mathbf{F} = -ie\mathbf{q}\varphi_{\mathbf{q},\omega} e^{i\mathbf{q}\mathbf{r} - i\omega t} + \text{c.c.}$ To the linear order in the perturbation, the distribution function can be sought in the form

$$f_{\mathbf{p}}(\mathbf{r}, t) = f_{\mathbf{p}}^{\text{eq}} + e\varphi_{\mathbf{q},\omega} g_{\mathbf{p}} e^{i\mathbf{q}\mathbf{r} - i\omega t} + \text{c.c.}, \quad (A4)$$

where $g_{\mathbf{p}}$ is position and time independent and satisfies the following linear integral equation:

$$-i\omega g_{\mathbf{p}} + i\mathbf{q}\mathbf{v}_{\mathbf{p}} g_{\mathbf{p}} - i\mathbf{q}\mathbf{v}_{\mathbf{p}} \frac{\partial f^{\text{eq}}}{\partial \epsilon_{\mathbf{p}}}$$
$$= 2\pi n_{\text{i}} \int \frac{d^2 \mathbf{p}'}{(2\pi)^2} |U(\mathbf{p} - \mathbf{p}')|^2 \delta(\epsilon_{\mathbf{p}} - \epsilon_{\mathbf{p}'})(g_{\mathbf{p}'} - g_{\mathbf{p}}). \quad (A5)$$

An explicit solution of this equation can be found only in the case of short-range impurities when $U(\mathbf{p} - \mathbf{p}')$ does not depend on momentum. In this case the collision integral reduces to the relaxation time approximation:

$$\text{St}[f] = \frac{\overline{f_{\mathbf{p}}} - f_{\mathbf{p}}}{\tau}, \quad (A6)$$

where the overbar denotes the average over the momentum directions on a constant energy surface,

$$\overline{f_{\mathbf{p}}} \equiv \frac{\int d^2 \mathbf{p}' f_{\mathbf{p}'} \delta(\epsilon_{\mathbf{p}'} - \epsilon_{\mathbf{p}})}{\int d^2 \mathbf{p}' \delta(\epsilon_{\mathbf{p}'} - \epsilon_{\mathbf{p}})}, \quad (A7)$$

and the relaxation time and the density of states per unit area are given by

$$\frac{1}{\tau} = \pi \nu n_{\text{i}} |U|^2, \quad \nu = 2 \int \frac{d^2 \mathbf{p}'}{(2\pi)^2} \delta(\epsilon_{\mathbf{p}} - \epsilon_{\mathbf{p}'}) \quad (A8)$$

(the factor of 2 in front of the integral takes into account two spin projections). Then, Eq. (A5) gives

$$g_{\mathbf{p}} = \frac{i\mathbf{q}\mathbf{v}_{\mathbf{p}}(\partial f^{\text{eq}}/\partial \epsilon_{\mathbf{p}}) + (1/\tau)\overline{g_{\mathbf{p}}}}{i\mathbf{q}\mathbf{v}_{\mathbf{p}} - i\omega + 1/\tau}. \quad (A9)$$

Averaging both sides over the momentum directions, one obtains a closed equation for $\overline{g_{\mathbf{p}}}$ and readily finds

$$\overline{g_{\mathbf{p}}} = \frac{1 + (i\omega\tau - 1)S}{1 - S} \frac{\partial f^{\text{eq}}}{\partial \epsilon_{\mathbf{p}}}, \quad (A10)$$

where S stands for the following angular average:

$$S \equiv \overline{(1 - i\omega\tau + i\mathbf{q}\mathbf{v_p}\tau)^{-1}}$$

$$= \int_0^{2\pi} \frac{d\phi}{2\pi} \frac{1}{1 - i\omega\tau + iv_F q \cos\phi}$$

$$= \frac{1}{\sqrt{(1 - i\omega\tau)^2 + (v_F q \tau)^2}}. \tag{A11}$$

The last two lines were written under the assumption of an isotropic dispersion $\epsilon_\mathbf{p}$. Finally, since the electron density is given by

$$\rho(\mathbf{r}, t) = 2 \int \frac{d^2\mathbf{p}}{(2\pi)^2} f_\mathbf{p}(\mathbf{r}, t) \tag{A12}$$

(again, the factor of 2 takes care of the spin multiplicity), the density response function can be found as

$$\Pi(\mathbf{q}, \omega) = 2 \int \frac{d^2\mathbf{p}}{(2\pi)^2} g_\mathbf{p}. \tag{A13}$$

Collecting all factors and assuming that $-\partial f^{\mathrm{eq}}/\partial\epsilon_\mathbf{p}$ is a narrow peak around the Fermi energy of width $\sim T$, so that the energy dependence of ν and v_F can be neglected, we arrive at Eq. (4).

If $U(\mathbf{p} - \mathbf{p}')$ is momentum dependent, no closed solution can be obtained even in the simplest of isotropic scattering when the scattering amplitude depends only on the difference $\phi - \phi'$ of the polar angles ϕ, ϕ' associated to the 2D vectors $\mathbf{p}, \mathbf{p}'$. Indeed, in this case the solution can be sought as a sum over polar harmonics, $g_\mathbf{p} = \sum_m g_m e^{im\phi}$, which are the eigenfunctions of the collision integral. Different harmonics do not separate because of the second term in Eq. (A5), which is responsible for the spatial dispersion of the conductivity $\sigma(\mathbf{q}, \omega)$. Only when the spatial dispersion is neglected, the Drude conductivity $\sigma(\omega)$ can be written in terms of the transport relaxation time τ_1, determined by the first eigenvalue of the collision integral, $-1/\tau_1$. Otherwise, the result contains all eigenvalues $-1/\tau_{m>0}$. Still, qualitatively, it is τ_1 that determines the relevant time scale: in the limit of Eq. (A6) all $\tau_{m>0} = \tau$, while in the opposite limit of small-angle scattering τ_m quickly grows with m, so high harmonics are suppressed and the result is determined by the first few τ_m's.

[1] S. M. Rytov, *Theory of Electric Fluctuations and Thermal Radiation* (Air Force Cambridge Research Center, Bedford, MA, 1953).

[2] D. Polder and M. Van Hove, Phys. Rev. B **4**, 3303 (1971).

[3] M. L. Levin, V. G. Polevoi, and S. M. Rytov, Zh. Eksp. Teor. Fiz. **79**, 2087 (1980) [Sov. Phys. JETP **52**, 1054 (1980)].

[4] J. J. Loomis and H. J. Maris, Phys. Rev. B **50**, 18517 (1994).

[5] J. B. Pendry, J. Phys.: Condens. Matter **11**, 6621 (1999).

[6] K. Joulain, J.-P. Mulet, F. Marquier, R. Carminati, and J.-J. Greffet, Surf. Sci. Rep. **57**, 59 (2005).

[7] A. I. Volokitin and B. N. J. Persson, Rev. Mod. Phys. **79**, 1291 (2007).

[8] B. Song, A. Fiorino, E. Meyhofer, and P. Reddy, AIP Adv. **5**, 053503 (2015).

[9] K. Kim, B. Song, V. Fernández-Hurtado, W. Lee, W. Jeong, L. Cui, D. Thompson, J. Feist, M. T. H. Reid, F. J. García-Vidal, J. C. Cuevas, E. Meyhofer, and P. Reddy, Nature **528**, 387 (2015).

[10] B. Song, D. Thompson, A. Fiorino, Y. Ganjeh, P. Reddy, and E. Meyhofer, Nat. Nanotechnol. **11**, 509 (2016).

[11] R. St-Gelais, L. Zhu, S. Fan, and M. Lipson, Nat. Nanotechnol. **11**, 515 (2016).

[12] K. Kloppstech, N. Könne, S.-A. Biehs, A. W. Rodriguez, L. Worbes, D. Hellmann, and A. Kittel, Nat. Commun. **8**, 14475 (2017).

[13] L. Cui, W. Jeong, V. Fernández-Hurtado, J. Feist, F. J. García-Vidal, J. C. Cuevas, E. Meyhofer, and P. Reddy, Nat. Commun. **8**, 14479 (2017).

[14] J. Yang, W. Du, Y. Su, Y. Fu, S. Gong, S. He, and Y. Ma, Nat. Commun. **9**, 4033 (2018).

[15] P. Sabbaghi, L. Long, X. Ying, L. Lambert, S. Taylor, C. Messner, and L. Wang, arXiv:1907.09638.

[16] S. M. Rytov, Y. A. Kravtsov, and V. I. Tatarskii, *Principles of Statistical Radiophysics* (Springer-Verlag, Berlin, 1989).

[17] P.-O. Chapuis, S. Volz, C. Henkel, K. Joulain, and J.-J. Greffet, Phys. Rev. B **77**, 035431 (2008).

[18] A. I. Volokitin and B. N. J. Persson, Phys. Rev. B **63**, 205404 (2001).

[19] E. Rousseau, M. Laroche, and J.-J. Greffet, Appl. Phys. Lett. **95**, 231913 (2009).

[20] V. B. Svetovoy, P. J. van Zwol, and J. Chevrier, Phys. Rev. B **85**, 155418 (2012).

[21] O. Ilic, M. Jablan, J. D. Joannopoulos, I. Celanovic, H. Buljan, and M. Soljačić, Phys. Rev. B **85**, 155422 (2012).

[22] P. Rodriguez-López, W.-K. Tse, and D. A. R. Dalvit, J. Phys.: Condens. Matter **27**, 214019 (2015).

[23] R. Yu, A. Manjavacas, and F. J. García de Abajo, Nat. Commun. **8**, 2 (2017).

[24] J.-H. Jiang and J.-S. Wang, Phys. Rev. B **96**, 155437 (2017).

[25] B. Zhao, B. Guizal, Z. M. Zhang, S. Fan, and M. Antezza, Phys. Rev. B **95**, 245437 (2017).

[26] W. B. Zhang, C. Y. Zhao, and B. X. Wang, Phys. Rev. B **100**, 075425 (2019).

[27] L. Wang, M. Bie, W. Cai, L. Ge, Z. Ji, Y. Jia, K. Gong, X. Zhang, J. Wang, and J. Xu, Opt. Express **27**, 36790 (2019).

[28] G. D. Mahan, Phys. Rev. B **95**, 115427 (2017).

[29] A. Kamenev, arXiv:1811.10187.

[30] L. M. A. Pascal, H. Courtois, and F. W. J. Hekking, Phys. Rev. B **83**, 125113 (2011).

[31] Z.-Q. Zhang, J.-T. Lü, and J.-S. Wang, Phys. Rev. B **97**, 195450 (2018).

[32] J.-S. Wang, Z.-Q. Zhang, and J.-T. Lü, Phys. Rev. E **98**, 012118 (2018).

[33] C. Caroli, R. Combescot, P. Nozieres, and D. Saint-James, J. Phys. C **4**, 916 (1971).

[34] G. Zala, B. N. Narozhny, and I. L. Aleiner, Phys. Rev. B **64**, 214204 (2001).

[35] N. D. Mermin, Phys. Rev. B **1**, 2362 (1970).

[36] M. Jablan, H. Buljan, and M. Soljačić, Phys. Rev. B **80**, 245435 (2009).

[37] E. M. Lifshitz and L. P. Pitaevskii, *Physical Kinetics* (Butterworth-Heinemann, Oxford, 1981). We remind here of the main regimes for the skin effect. If the bulk mean free path $\ell_{3D} \ll c/\omega_p$, where ω_p is the bulk plasma frequency, the skin depth δ can be found from the local 3D Drude conductivity $\sigma_{3D}(\omega) = (4\pi)^{-1}\omega_p^2\tau/(1 - i\omega\tau)$: $\delta \sim (c/\omega_p)/\sqrt{\omega\tau}$ for $\omega \ll 1/\tau$ and $\delta \sim c/\omega_p$ for $\omega \gg 1/\tau$. For a sufficiently clean metal with $\ell_{3D} \gg c/\omega_p$, the above expressions for δ work for $\omega \ll (c/\omega_p\ell_{3D})^2/\tau$ and $\omega \gg (v_F/c)\omega_p$, respectively, while in the interval $(c/\omega_p\ell_{3D})^2/\tau \ll \omega \ll (v_F/c)\omega_p$ one has the anomalous skin effect with nonlocal response and $\delta \sim (c/\omega_p)^{2/3}(v_F/\omega)^{1/3}$.

[38] R. Geick, C. H. Perry, and G. Rupprecht, Phys. Rev. **146**, 543 (1966).

[39] C. M. Hargreaves, Phys. Lett. A **30**, 491 (1969); more precise measurements were described in the Ph.D. thesis of C. M. Hargreaves, University of Leiden, 1973, reproduced in Ref. [8].

[40] J. Rath and J. Callaway, Phys. Rev. B **8**, 5398 (1973).

[41] S. K. Sinha, S. H. Liu, L. D. Muhlestein, and N. Wakabayashi, Phys. Rev. Lett. **23**, 311 (1969).

[42] T. Kralik, P. Hanzelka, M. Zobac, V. Musilova, T. Fort, and M. Horak, Phys. Rev. Lett. **109**, 224302 (2012).

[43] D. Gall, J. Appl. Phys. **119**, 085101 (2016).

[44] M. A. Ordal, R. J. Bell, R. W. Alexander, L. L. Long, and M. R. Querry, Appl. Opt. **24**, 4493 (1985).

[45] N. A. Court, A. J. Ferguson, R. Lutchyn, and R. G. Clark, Phys. Rev. B **77**, 100501(R) (2008).

[46] A. A. Abrikosov, *Fundamentals of the Theory of Metals* (Dover, New York, 1988).