Adiabatic Amplification of Plasmons and Demons in 2D Systems

Zhiyuan Sun,[1] D. N. Basov,[1,2] and M. M. Fogler[1]

[1]*Department of Physics, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093*
[2]*Department of Physics, Columbia University, 538 West 120th Street, New York, New York 10027*

We theoretically investigate charged collective modes in a two-dimensional conductor with hot electrons where the instantaneous mode frequencies gradually increase or decrease with time. We show that the loss compensation or even amplification of the modes may occur. We apply our theory to two types of collective modes in graphene, the plasmons and the energy waves, which can be probed in optical pump-probe experiments.

Introduction.—Plasmons in metals, semiconductors, and other solid-state systems have been a topic of intensive research for over half a century [1]. Plasmonics has found a number of technological applications in chemical sensing, light manipulation, and information processing. Photoexcitation by ultrashort laser pulses [2–4] is one of the methods to generate plasmons. When the pulsed excitation is of high enough power, it can modify material properties of either the plasmonic medium or its electromagnetic environment, which is the principle underlying the emerging field of *active* plasmonics [2,5–7]. For example, photoexcitation-induced population inversion may permit plasmon loss compensation or amplification [5,6]. More often, plasmon lifetime remains quite short, e.g., tens of femtoseconds (fs) in noble metals, which is an obstacle to applications. In experiments using ultrafast optical pulses, the plasmon frequency changes with time as the system relaxes back to equilibrium. However, because of high damping, it has been customary to treat plasmonic response of the system as quasistationary during the plasmon lifetime.

Recently, graphene has emerged as a new plasmonic medium distinguished by record-high tunability and confinement [8,9]. Combating damping remains a challenge; however, plasmon quality factors as high as $Q \sim 30$ have been demonstrated [4,10] for graphene encapsulated in hexagonal boron nitride. A new scientific frontier in graphene plasmonics is nonlinear [11,12] and nonequilibrium dynamics probed in ultrafast optical experiments [13,14]. Plasmon amplification through stimulated emission [15,16] has been proposed theoretically and plasmon switching by optical pumping has been demonstrated experimentally [4].

These encouraging developments motivate us to study the regime where the plasmon lifetime is comparable or longer than the characteristic relaxation time in a material. Although this regime may or may not be realizable in graphene, we consider this as a theoretical possibility. Previously, collective modes in media undergoing adiabatic evolution have been discussed in theoretical astrophysics [17], plasma physics [18], and general relativity [19]. In this Letter, we apply similar ideas to solid-state materials, which are better suited for controlled experiments. Our key finding is that loss compensation or even amplification can be a *natural* outcome of the transient plasmon dynamics. Additionally, we show that the same concept applies to the energy wave in graphene [20,21], which is a collective mode similar to acoustic plasmons (or "demons" [1]) in metals and semiconductors [22–24] and also to "cosmic sound" in the early universe [25].

Qualitative picture.—To model a nonequilibrium system under intense photoexcitation we assume that its electron temperature T is much larger than the lattice temperature T_l. Such a hot-electron state typically forms in metals and semiconductors a few tens of fs after optical pumping. This rapid thermalization (that is, relaxation of the electron distribution to the Fermi-Dirac form with the temperature T) is due to strong interactions of electrons with each other and with optical phonons. Subsequently, T gradually decreases toward T_l at a much slower "cooling" rate measured in picoseconds (ps), predominantly due to emission of acoustic phonons. If the plasmon dispersion depends on T, plasmons propagating in this transient state would have a slowly changing frequency, i.e., the plasmons would be chirped. We will show that such an adiabatic change of the plasmon frequency could induce adiabatic amplification of the plasmon amplitude.

Adiabatic change of parameters has been previously considered in the context of plasmon-polariton focusing in tapered waveguides [26]. As plasmon approaches the narrow end of the waveguide, its group velocity decreases and its electric field increases. In this situation the change of parameters occurs in space. The mechanism we study relies instead on having parameters changing in time. To explain our key idea let us treat the plasmon as a harmonic oscillator with the equation of motion

$$(\partial_t^2 + \gamma(t)\partial_t + \omega^2(t))X = 0 \qquad (1)$$

for its canonical coordinate $X(t)$ (e.g., charge density). Here $\gamma(t)$ is the damping rate and $\omega(t)$ is the instantaneous mode frequency. Suppose $\omega(t)$ changes monotonically with the decay rate $\kappa \equiv -\partial_t \ln \omega$, which is slow enough, $\omega \gg \kappa$, then the Wentzel-Kramers-Brillouin (WKB) approximation to the solution of Eq. (1) is valid:

$$X(t) = A(t)e^{-iS(t)}, \qquad (2)$$

$$A(t) = \frac{1}{\sqrt{\omega(t)}} \exp\left(-\frac{1}{2}\int_0^t \gamma(t_0)dt_0\right), \qquad (3)$$

$$S(t) = \int_0^t \omega(t_0)dt_0. \qquad (4)$$

If both γ and κ are constant, the time-dependent plasmon amplitude has the form

$$A(t) = e^{\frac{1}{2}(\kappa-\gamma)t}. \qquad (5)$$

Clearly, the frequency decay rate κ competes with the damping rate γ. If the condition $\kappa > \gamma$ is met, then the oscillation amplitude increases with time, as shown in Figs. 1(a) and 1(b).

Although the adiabatic principle appears simple and straightforward, its application to actual solid-state systems may require sorting out some important details. In the remainder of this Letter we do so on the examples of two types of collective modes: the plasmons and the energy waves in graphene.

Plasmons in two-dimensional (2D) materials.—2D materials are very promising for active plasmonics because they are not affected by a finite penetration length of optical beams and are much more tunable than bulk metals. It is well known [27–29] that such plasmons have a characteristic square-root dispersion with momentum (Fig. 2), $\omega_q = \sqrt{(2/\epsilon)Dq}$ where ϵ is the permittivity of the
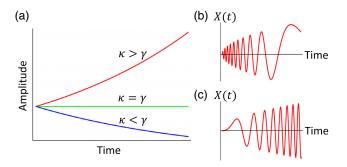
environment and D is the Drude weight (see below). Our goal is to show that the time dependence of D may give rise to adiabatic amplification of plasmons.

If the system has the spatial translational symmetry, different momenta are decoupled. For a given $\mathbf{q}$, in the linear-response regime, the plasmon dynamics is determined by the electrical conductivity operator with the kernel $\sigma_q(t, t_0)$. Consider the following model for the conductivity kernel:

$$\sigma_q(t, t_0) = \frac{1}{\pi}D(t)e^{-\Gamma(t-t_0)}\theta(t - t_0), \qquad (6)$$

$$D(t) = D(0)e^{-2\int_0^t \kappa(t')dt'}. \qquad (7)$$

This model is motivated by a popular physical picture (see, for example, Ref. [35]) where the current damping occurs because the "density of photoexcited carriers" decays with the rate 2κ and because, additionally, these carriers experience momentum relaxation with the rate Γ; see [36] for further discussion.

Let us focus for now on the case of undoped graphene where the Drude weight $D(t)$ is proportional to the electron temperature [30] $D(t) = 2\ln 2(e^2/\hbar^2)T(t)$ and where the underdamped plasmons exist at frequencies $\tau_{ee}^{-1} \ll \omega \ll T/\hbar$. The lower limit is set by the electron-electron scattering rate τ_{ee}^{-1}; the upper limit is imposed by the Landau damping due to the interband transitions; see Fig. 2. In particular, the dimensionless Landau damping rate of the thermal plasmons is given by [34] $\Gamma/\omega = (\pi/16\ln 2)(\hbar\omega/T)^2$, which is small if $\hbar\omega \ll T$. Note also that the assumption of scalar D can be justified if κ and γ are much smaller than the electron-electron

(a)

(b) $X(t)$

(c) $X(t)$

FIG. 1.   (a) A schematic showing the amplitude of a plasmon as a function of time t for different relations between the mode frequency decay rate κ and the damping rate γ. (b) The canonical coordinate X as a function of t in the $\kappa > \gamma$ case. The amplitude grows as the frequency drops (a down-chirp). (c) $X(t)$ for the case where amplification occurs while frequency increases (an up-chirp), as in the tunneling process sketched in Fig. 3(a) below.
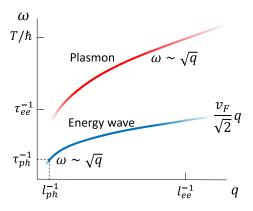

FIG. 2.   Dispersion of the plasmon [30] and the energy wave [20,21,31] in a weakly doped graphene with hot electrons (schematically). The plasmon (energy wave) exists at ω above (below) τ_{ee}^{-1}; otherwise, it is overdamped, as indicated by the fainting ends of the curves. For $T \gg \mu$ and $\alpha \ll 1$, $\tau_{ee}^{-1} = a\alpha^2 T$ with $\alpha = e^2/\epsilon\hbar v_F$ and $a \sim 4$ [21,32–34]. The plasmon is also overdamped at $\omega \gtrsim T/\hbar$, while the energy wave is damped by electron-phonon and disorder scattering characterized by the rate τ_{ph}^{-1}.

relaxation rate τ_{ee}^{-1} so that an isotropic electron distribution (in the absence of a probe) is maintained.

It is straightforward to show that the equation of motion for the plasmon has the same form as Eq. (1) with X equal to ρ_q, the Fourier harmonic of the charge density, and with the dissipation rate equal to

$$\gamma = 2\kappa + \Gamma. \tag{8}$$

Unfortunately, the condition $\kappa > \gamma$ seems impossible to satisfy since $\Gamma > 0$ and $\kappa > 0$. In other words, the amplification of down-chirped plasmons cannot occur due to the plasmon damping rate being larger than the frequency decay rate; see also the Supplemental Material [37].

Suppose, however, that the Drude weight is *growing*, $\kappa < 0$. In this case (an up-chirped plasmon) the criterion for amplification $\kappa < -\Gamma$ can be met if the growth rate is fast enough; see Fig. 2(c). Under what conditions can this scenario be realized? One possibility is to leverage the dependence of the Drude weight on the carrier density or effective mass, which is another common attribute of ultrafast pump-probe experiments [2,7]. We speculate that the plasmon amplification may be possible by exploiting tunneling in a vertical semiconductor-insulator-graphene heterostructure; see Fig. 3(a). The semiconductor could be, e.g., a transition-metal dichalcogenide and the insluator could be hexagonal boron nitride (hBN), as in recent experiments [41]. With a suitable bias voltage applied, the initial state with a lower electrochemical potential in graphene can be maintained as the insulator band gap would prevent electron tunneling in any direction. However, once they are heated to energies close or above the insulator's band edge, the electrons in the semiconductor layer would tunnel to graphene. (This is similar to a hot-electron doping effect [42] whereas in [41]
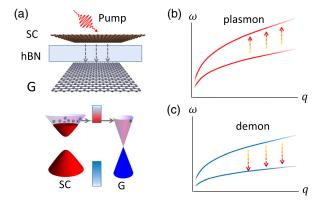

FIG. 3. (a) Sketch of the heterostructure made of parallel ultrathin layers of semiconductor (SC), insulator, and graphene (G) where plasmon amplification can occur when hot electrons tunnel from the semiconductor to graphene. (b) A qualitative change of the plasmon dispersion during the tunneling. (c) A qualitative change of the demon dispersion with increasing temperature.

the tunneling was in the opposite direction.) For tunneling to be rapid the insulator must be thin, which implies that the charges and current in the two layers would also be coupled electromagnetically. Therefore, the plasmons are the modes of the combined system. If the effective carrier mass in the semiconductor is larger than that in graphene, then the initial Drude weight is low but as a result of tunneling, the combined Drude weight of the carriers in the system (and hence, the electric current) would increase. The upper limit for the amplification factor can be estimated by completely neglecting the damping, $\Gamma \to 0$, in the expression for the charge density amplitude

$$\rho_q(t) \propto e^{-\frac{1}{2}(\kappa+\Gamma)t}, \qquad \kappa < 0. \tag{9}$$

The amplification is proportional to the square root of the plasmon frequency, or the fourth root of the Drude weight. If the increase of the latter comes from the decrease of the effective mass by, say, a factor of 2, then plasmon amplification by as much as $\sim 20\%$ may be possible. For more elaborate estimates, the carrier dynamics beyond the simple Drude approximation would need to be included in the model (see, for example, Ref. [43] and the theory references cited therein).

The tunneling time of hot electrons across ultrathin hBN layers can be as short as 7 fs [44], which would correspond to κ perhaps as high as several tens of ps^{-1}. In comparison, the damping rate in hBN-encapsulated graphene was found to be $\Gamma \sim 2$ and 20 ps^{-1} before and after the optical pump, respectively [4]. Hence, fulfilling the condition $\kappa > \gamma$ may be feasible. Since the semiconductor would partially absorb the pump pulse, graphene may remain relatively cool, which may help reduce the plasmon damping due to electron-phonon scattering [10].

Experimental investigation of the frequency, amplitude, and spatial interference patterns of the amplified plasmons as a function of time may be possible by far- and near-field pump-probe optical techniques [2,4,9].

Energy wave (demon) in graphene.—Our second example of a collective mode that may exhibit adiabatic amplification is the energy wave in graphene. This mode is predicted [20,21,31] to exist in the hydrodynamic regime of frequencies that are lower than the electron-electron collision rate τ_{ee}^{-1}; see Fig. 2. In this regime, only collective variables immune to interparticle collisions, i.e, the zero modes of the collision integral, are important: the local temperature $T(\mathbf{r})$, chemical potential $\mu(\mathbf{r})$, and drift velocity $\mathbf{u}(\mathbf{r})$. Their dynamics is described by a set of hydrodynamic equations [21,33,36]. The energy wave is the propagating longitudinal mode resulting from this set of equations. Consider a weakly doped graphene, $\mu \ll T$. The dispersion relation of the energy wave, neglecting dissipation, is

$$\omega_q = v_F \sqrt{\frac{1}{2}q^2 + \frac{4\pi}{3}\frac{e^2}{\epsilon}\frac{n^2}{n_E}q}, \qquad (10)$$

where n is the average electron density and $n_E \equiv \langle \varepsilon \rangle$ is the average kinetic energy density (relative to the zero-doping, zero-temperature state). The latter behaves as [36] $n_E \propto T^3$ in the regime we consider, $T \gg |\mu|$. For $q \gg (e^2/\epsilon)(n^2/n_E) \equiv q_c$, the dispersion of Eq. (10) approaches $\omega = (1/\sqrt{2})v_F q$. This collective mode is neutral because electrons and holes oscillate in phase. It is similar to acoustic plasmons observed in semiconductors [23,24]. Incidentally, the plasmons in Ref. [45] were referred to as acoustic because their dispersion was changed from the square-root to a linear one due to screening by a nearby gate. This is unlike the original meaning of the term acoustic plasmon (or "demon") introduced for a system where the screening is by electrons from a different band of the same material [1].

For $q \ll q_c$, the second term in the square root of Eq. (10) dominates, so $\omega_q \propto \sqrt{q}$. In this case the energy wave is no longer neutral: it involves both energy and charge density oscillations. However, it is different from the plasmon. First, the energy wave is in the hydrodynamic regime $\omega \ll \tau_{ee}^{-1}$ while the plasmon is in the high frequency regime $\omega \gg \tau_{ee}^{-1}$. (In practice, the range of admissible q is also limited from below by the inverse mean-free path l_{ph} due to electron-phonon and disorder scattering; see Fig. 2.) Second, the frequency of the plasmon increases with electron temperature T [similar to what is shown in Fig. 3(b)] while that of the demons decreases [Fig. 3(c)]. We will focus on the small q region of the energy wave where its frequency can be efficiently controlled by T. For $\mu \sim 40$ meV ≈ 500 K and $T = 3000$ K, which is the regime probed in a recent experiment [4], the wavelength corresponding to momentum q_c is about 1 μm. The change of temperature causes the change of the energy density n_E, which in turn affects the frequency of the energy wave through Eq. (10). We assume that this change is adiabatic, in other words, that the decay rate $\kappa = -\frac{1}{2}\partial_t \ln n_E$ is a small parameter. Keeping only the leading terms in the hydrodynamic equations, we get the WKB solutions for the charge density n_q and the energy density [36]

$$n_q(t) \propto [n_E(t)]^{-1/4} e^{-iS(t)}, \qquad (11)$$

$$n_{Eq}(t) \propto [n_E(t)]^{3/4} e^{-iS(t)}, \qquad (12)$$

with $S(t)$ given by Eq. (4). Therefore, if we want to increase the energy density oscillations, we need to *increase* the average energy density n_E; in other words, we need to heat up graphene. This can be done using, for example, a moderate intensity laser source that heats the sample faster than the characteristic time τ_{ph} of electron scattering by phonons and disorder. According to Eq. (12), the naive upper bound for the amplification factor (neglecting any damping) is $(T/T_l)^{9/4}$ where T is the electron temperature after the photoexcitation. The amplification is only possible if graphene is slightly doped, in which case the energy mode is not purely neutral. Hence, it can also be probed by optical pump-probe spectroscopy, at THz frequencies. Alternatively, it may be possible to exploit coupling of this mode to phonons and probe it by inelastic light scattering, similar to acoustic plasmons in semiconductors [23,24].

Three-temperature state.—A solid-state system exhibiting adiabatic amplification of collective modes would have another interesting nonequilibrium property. It would have not two but three different effective temperatures. In addition to the lattice temperature T_l and the electron temperature T, it would also have the collective mode temperature T_m. The temperature T_m characterizes the modes created by random thermal fluctuations rather than those induced by an external probe pulse. In the absence of damping, $\Gamma = 0$, the time evolution of T_m can be deduced from the principle of entropy conservation in an adiabatic process. The entropy of an ensemble of identical harmonic oscillators depends only on the ratio of temperature and their mode frequency. The damping introduces an additional factor $e^{-\Gamma t}$. Therefore, we expect $(T_m(t)/T_m(0)) \approx (\omega_q(t)/\omega_q(0))e^{-\Gamma t}$. [Here we assume that T_m is still much larger than the final equilibrium temperature $T_m(t=\infty) = T_l$.] When the hot-electron state is just created, $T_m = T_m(0)$ and $T = T(0)$ should presumably be of the same order. Thereafter, they would diverge from one another. For example, for graphene plasmons we find $T_m(t) \sim \omega_q(t)e^{-\Gamma t} \sim [T(t)]^{1/2}e^{-\Gamma t}$. For graphene energy wave, the same argument yields $T_m(t) \sim \omega_q(t)e^{-\Gamma t} \sim [T(t)]^{-3/2}e^{-\Gamma t}$.

In summary, we proposed the concept of adiabatic amplification of chirped collective modes in nonequilibrium systems under photoexcitation and suggested two possible routes for its experimental realization in 2D materials. Although we focused on systems with hot electrons, the concept of adiabatic amplification is also applicable to systems with "cold" electrons, for example, superconducting films [46,47], where plasmon damping can be even smaller than in graphene.

[1] D. Pines, Can. J. Phys. **34**, 1379 (1956).
[2] K. F. MacDonald, Z. L. Sámson, M. I. Stockman, and N. I. Zheludev, Nat. Photonics **3**, 55 (2009).
[3] H. Baida, D. Mongin, D. Christofilos, G. Bachelier, A. Crut, P. Maioli, N. Del Fatti, and F. Vallée, Phys. Rev. Lett. **107**, 057402 (2011).

[4] G. X. Ni, L. Wang, M. D. Goldflam, M. Wagner, Z. Fei, A. S. McLeod, M. K. Liu, F. Keilmann, B. Özyilmaz, A. H. Castro Neto, J. Hone, M. M. Fogler, and D. N. Basov, Nat. Photonics **10**, 244 (2016).

[5] O. Hess, J. B. Pendry, S. A. Maier, R. F. Oulton, J. M. Hamm, and K. L. Tsakmakidis, Nat. Mater. **11**, 573 (2012).

[6] M. I. Stockman, in Active Plasmonics and Tuneable Plasmonic Metamaterials, edited by A. V. Zayats and S. A. Maier (Wiley, New York, 2013), Vol. 32, pp. 1–39.

[7] H. R. Seren, J. Zhang, G. R. Keiser, S. J. Maddox, X. Zhao, K. Fan, S. R. Bank, X. Zhang, and R. D. Averitt, Light Sci. Appl. **5**, e16078 (2016).

[8] J. Chen, M. Badioli, P. Alonso-González, S. Thongrattanasiri, F. Huth, J. Osmond, M. Spasenović, A. Centeno, A. Pesquera, P. Godignon, A. Z. Elorza, N. Camara, F. J. García de Abajo, R. Hillenbrand, and F. H. L. Koppens, Nature (London) **487**, 77 (2012).

[9] Z. Fei, A. S. Rodin, G. O. Andreev, W. Bao, A. S. McLeod, M. Wagner, L. M. Zhang, Z. Zhao, M. Thiemens, G. Dominguez, M. M. Fogler, A. H. Castro Neto, C. N. Lau, F. Keilmann, and D. N. Basov, Nature (London) **487**, 82 (2012).

[10] A. Woessner, M. B. Lundeberg, Y. Gao, A. Principi, P. Alonso-González, M. Carrega, K. Watanabe, T. Taniguchi, G. Vignale, M. Polini, J. Hone, R. Hillenbrand, and F. H. L. Koppens, Nat. Mater. **14**, 421 (2015).

[11] J. D. Cox and F. Javier García de Abajo, Nat. Commun. **5**, 5725 (2014).

[12] S. A. Mikhailov, Phys. Rev. B **90**, 241301 (2014).

[13] M. Wagner, Z. Fei, A. S. McLeod, A. S. Rodin, W. Bao, E. G. Iwinski, Z. Zhao, M. Goldflam, M. Liu, G. Dominguez, M. Thiemens, M. M. Fogler, A. H. Castro Neto, C. N. Lau, S. Amarie, F. Keilmann, and D. N. Basov, Nano Lett. **14**, 894 (2014).

[14] D. Brida, A. Tomadin, C. Manzoni, Y. J. Kim, A. Lombardo, S. Milana, R. R. Nair, K. S. Novoselov, A. C. Ferrari, G. Cerullo, and M. Polini, Nat. Commun. **4**, 1987 (2013).

[15] F. Rana, IEEE Trans. Nanotechnol. **7**, 91 (2008).

[16] V. Apalkov and M. I. Stockman, Light Sci. Appl. **3**, e191 (2014).

[17] R. J. Morton, A. W. Hood, and R. Erdélyi, Astron. Astrophys. **512**, A23 (2010).

[18] P. F. Schmit, I. Y. Dodin, and N. J. Fisch, Phys. Rev. Lett. **105**, 175003 (2010).

[19] L. P. Grishchuk, Sov. Phys. JETP **67**, 825 (1975).

[20] T. V. Phan, J. C. W. Song, and L. S. Levitov, arXiv:1306.4972.

[21] U. Briskot, M. Schütt, I. V. Gornyi, M. Titov, B. N. Narozhny, and A. D. Mirlin, Phys. Rev. B **92**, 115426 (2015).

[22] J. Ruvalds, Adv. Phys. **30**, 677 (1981).

[23] A. Pinczuk, J. Shah, and P. A. Wolff, Phys. Rev. Lett. **47**, 1487 (1981).

[24] P. Padmanabhan, S. M. Young, M. Henstridge, S. Bhowmick, P. K. Bhattacharya, and R. Merlin, Phys. Rev. Lett. **113**, 027402 (2014).

[25] R. A. Sunyaev and Y. B. Zeldovich, Astrophys. Space Sci. **7**, 3 (1970).

[26] M. I. Stockman, Phys. Rev. Lett. **93**, 137404 (2004).

[27] A. N. Grigorenko, M. Polini, and K. S. Novoselov, Nat. Photonics **6**, 749 (2012).

[28] D. Basov, M. Fogler, A. Lanzara, F. Wang, and Y. Zhang, Rev. Mod. Phys. **86**, 959 (2014).

[29] F. J. García de Abajo, ACS Photonics **1**, 135 (2014).

[30] O. Vafek, Phys. Rev. Lett. **97**, 266406 (2006).

[31] D. Svintsov, V. Vyurkov, S. Yurchenko, T. Otsuji, and V. Ryzhii, J. Appl. Phys. **111**, 083715 (2012).

[32] A. B. Kashuba, Phys. Rev. B **78**, 085415 (2008).

[33] M. Müller, L. Fritz, and S. Sachdev, Phys. Rev. B **78**, 115406 (2008).

[34] M. Schütt, P. M. Ostrovsky, I. V. Gornyi, and A. D. Mirlin, Phys. Rev. B **83**, 155441 (2011).

[35] J. Orenstein and J. S. Dodge, Phys. Rev. B **92**, 134507 (2015).

[36] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.117.076805 for technical details.

[37] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.117.076805, which includes Refs. [38–40], for a derivation of the conductivity kernel and the energy wave dispersion.

[38] M. Müller and S. Sachdev, Phys. Rev. B **78**, 115419 (2008).

[39] M. Müller, J. Schmalian, and L. Fritz, Phys. Rev. Lett. **103**, 025301 (2009).

[40] B. N. Narozhny, I. V. Gornyi, M. Titov, M. Schütt, and A. D. Mirlin, Phys. Rev. B **91**, 035414 (2015).

[41] M. Massicotte, P. Schmidt, F. Vialla, K. Watanabe, T. Taniguchi, K. J. Tielrooij, and F. H. L. Koppens, Nat Commun **7**, 12174 (2016).

[42] Z. Fang, Y. Wang, Z. Liu, A. Schlather, P. M. Ajayan, F. H. L. Koppens, P. Nordlander, and N. J. Halas, ACS Nano **6**, 10222 (2012).

[43] R. Huber, F. Tauser, A. Brodschelm, M. Bichler, G. Abstreiter, and A. Leitenstorfer, Nature (London) **414**, 286 (2001).

[44] Q. Ma, T. I. Andersen, N. L. Nair, N. M. Gabor, M. Massicotte, C. H. Lui, A. F. Young, W. Fang, K. Watanabe, T. Taniguchi, J. Kong, N. Gedik, F. H. L. Koppens, and P. Jarillo-Herrero, Nat. Phys. **12**, 455 (2016).

[45] P. Alonso-González, A. Y. Nikitin, Y. Gao, A. Woessner, M. B. Lundeberg, A. Principi, N. Forcellini, W. Yan, S. Velez, A. J. Huber, K. Watanabe, T. Taniguchi, L. E. Hueso, M. Polini, J. Hone, F. H. L. Koppens, and R. Hillenbrand, arXiv:1601.05753.

[46] O. Buisson, P. Xavier, and J. Richard, Phys. Rev. Lett. **73**, 3153 (1994).

[47] H. T. Stinson, J. S. Wu, B. Y. Jiang, Z. Fei, A. S. Rodin, B. C. Chapler, A. S. McLeod, A. Castro Neto, Y. S. Lee, M. M. Fogler, and D. N. Basov, Phys. Rev. B **90**, 014502 (2014).