

Tuning quantum non-local effects in graphene plasmonics

Mark B. Lundberg,¹ Yuanda Gao,² Reza Asgari,^{3,4} Cheng Tan,² Ben Van Duppen,⁵ Marta Autore,⁶ Pablo Alonso-González,^{6,7} Achim Woessner,¹ Kenji Watanabe,⁸ Takashi Taniguchi,⁸ Rainer Hillenbrand,^{9,10} James Hone,² Marco Polini,^{11,*} and Frank H. L. Koppens^{1,12,†}

¹*ICFO — Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain*

²*Department of Mechanical Engineering, Columbia University, New York, NY 10027, USA*

³*School of Physics, Institute for Research in Fundamental Sciences (IPM), 19395-5531, Tehran, Iran*

⁴*School of Nano Science, Institute for Research in Fundamental Sciences (IPM), 19395-5531, Tehran, Iran*

⁵*Department of Physics, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerp, Belgium*

⁶*CIC nanoGUNE, E-20018, Donostia-San Sebastián, Spain*

⁷*Departamento de Física, Universidad de Oviedo, Oviedo 33007, Spain*

⁸*National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan*

⁹*CIC nanoGUNE and EHU/UPV, E-20018, Donostia-San Sebastián, Spain*

¹⁰*IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain*

¹¹*Istituto Italiano di Tecnologia, Graphene Labs, Via Morego 30, I-16163 Genova, Italy*

¹²*ICREA — Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain.*

The response of an electron system to electromagnetic fields with sharp spatial variations is strongly dependent on quantum electronic properties, even in ambient conditions, but difficult to access experimentally. We use propagating graphene plasmons, together with an engineered dielectric-metallic environment, to probe the graphene electron liquid and unveil its detailed electronic response at short wavelengths. The near-field imaging experiments reveal a parameter-free match with the full theoretical quantum description of the massless Dirac electron gas, in which we identify three types of quantum effects as keys to understanding the experimental response of graphene to short-ranged terahertz electric fields. The first type is of single-particle nature and is related to shape deformations of the Fermi surface during a plasmon oscillation. The second and third types are a many-body effect controlled by the inertia and compressibility of the interacting electron liquid in graphene. We demonstrate how, in principle, our experimental approach can determine the full spatiotemporal response of an electron system.

The quantum physics of electron systems involves complex short-distance interactions and motions that depend sensitively on electron correlations and Fermi surface deformations.^{1,2} These are often considered irrelevant in optical and transport measurements, which probe the response to electrical fields with long length scales. When free electron systems are driven by electric fields varying rapidly in both time and space, however, the response pattern in dynamical current reveals these complex short-range effects. This aspect of electron response, known as non-locality or spatial dispersion in conductivity, arises due to the internal spreading of energy via the moving electrons. Even in ambient conditions (as Fermi liquid parameters depend on temperature weakly^{1,2}), the spatial dispersion in an electron system retains a detailed connection to Fermi-surface and electron-electron correlation effects, and hence it provides a unique window into quantum theories of electron systems without requiring extremes of low temperature or high magnetic field. Unfortunately, these quantum regimes cannot be accessed by standard optical and transport probes.

Plasmons—electric waves resulting from an inertial electron conductivity combined with electric restoring

forces—can act as a carrier of the spatiotemporal electric fields necessary to probe non-locality. All systems exhibit non-local effects for plasmon wavelengths approaching the electronic Fermi wavelength λ_F , which has been confirmed in experimental studies of metals and semiconductor two-dimensional (2D) electron gases.³⁻⁵ Such experiments have however led to challenges in quantitative interpretation, due to strong interactions that go beyond standard (e.g. random phase approximation) theoretical treatments,^{3,4} and possible complications by edge effects and tunneling.⁵⁻⁹

In graphene, it is possible to access a different, velocity-based, type of non-locality due to the ability to tune its plasmon phase velocity to low values, close to its Fermi velocity of $v_F \approx c/300$, where c is the speed of light in vacuum. Here, we exploit heterostructures of high-quality graphene, hexagonal Boron Nitride (h-BN), and nearby metal¹⁰ to confine the plasmons vertically down to 5 nm (see insets of Fig. 1A), and as a consequence slow down the propagation velocity, as illustrated in Fig. 1A. In addition, by tuning to relatively low carrier density, as accessible in high-quality graphene, we can further slow plasmons down to about $c/250$, approaching the Fermi velocity, such that non-local effects become very significant. We probe propagating plasmons by exploiting the near-field optical scattering scanning probe technique^{11,12} for frequencies as low as a few THz¹⁰ to avoid interband plasmon losses. Importantly, our tech-

* Marco.Polini@iit.it

† frank.koppens@icfo.eu

nique allows a very accurate determination of the plasmon wavelength, independent of edge effects, which enables us to experimentally determine the non-local dynamical conductivity of graphene, $\sigma(\omega, q)$, a function of angular frequency ω and wavevector q .

To illustrate the tunability of the non-local effects, we consider the effect of the environment on the graphene plasmon properties. In general, graphene plasmons occur for ω, q values that satisfy the self-oscillation condition (see Methods)

$$\sigma(\omega, q) - \frac{i\omega}{q^2} C(\omega, q) = 0, \quad (1)$$

where $C(\omega, q)$ is the dynamical capacitance of the environment around the graphene.

For graphene–dielectric plasmons, the small capacitance values ($C = 2\varepsilon q$, for permittivity ε) mean that non-local regimes are only accessible for short plasmon wavelengths on the order of the Fermi wavelength.^{13–16} In such experiments the high observed phase velocities, typically more than twice v_F , have given negligible non-local effects.¹⁷ The addition of a metal film at distance d from the graphene increases capacitance up to $C \approx \varepsilon/d$. As a consequence of Eq. (1) and as shown in Fig. 1B this drives the plasmon to larger values of q , and thus smaller velocity. As this capacitive effect also morphs the plasmon from the graphene–dielectric unscreened dispersion $\omega \propto \sqrt{q}$ to a linear dispersion $\omega \propto q$, we take the plasmon phase velocity $v_p = \omega/q$ as the key parameter characterizing plasmons in this system.^{10,18}

The graphene–metal system has two tuning parameters: the separation d fixed at fabrication, which controls capacitance C , and the graphene sheet carrier density n_s , which gives *in-situ* control of conductivity σ . A birds-eye view of this combined tunability is shown in Fig. 1C, which shows that the combination of the two tuning parameters n_s and d allows the plasmon velocity v_p to be lowered to small values near the Fermi velocity $v_F \approx 1.0 \times 10^6$ m/s. Figure 1C also shows how the local approximation begins to break down in this regime (with $> 10\%$ error for $v_p/v_F < 2$) as non-local effects become significant, due to reasons that will be elaborated below.

Our experimental realization of this concept is depicted in Fig. 2. We encapsulated graphene in the h-BN dielectric, a combination which has shown the highest-quality graphene plasmons to date,¹⁷ and placed it on top of an AuPd metal layer. Three devices of this form were created, with distinct graphene–metal separation d controlled by the chosen thickness of the bottom h-BN layer (further fabrication details in Methods), and contacted with Au electrodes.¹⁹ In order to visualize plasmons and characterize their propagation, we used a scattering-type near-field optical microscope in photocurrent mode.^{10,20,21} We operated the microscope in the Terahertz (THz) frequency range (we chose a laser frequency of $\frac{\omega}{2\pi} = 3.11$ THz), as this allowed us to probe down to the lowest plasmon velocities while respecting the minimum-wavelength limitations of the near-field

probe.²² In order to detect the photocurrent, we left a narrow split in the AuPd metal layer (Fig. 2A, inset) so that a graphene p-n junction could be formed by applying distinct gate voltages V_L and V_R . As a second purpose, this sharp split also served as a launching edge for graphene plasmons.

Plasmons appear experimentally as characteristic interference fringes in the dependence of photocurrent on tip position (Fig. 2B), due to tip–plasmon interference that modulates the absorbed power.^{10,21} Besides the edge-reflection fringes examined in earlier works, we also observed fringes associated with plasmons launching from the split in the AuPd gate, particularly in the thinner- d devices. In either case, the fringes allowed a determination of the plasmon wavelength $\lambda = \frac{2\pi}{q}$, by fitting to the photocurrent with an appropriately subtracted background (Fig. 2B; see Methods for fitting details). This, in combination with the known excitation frequency directly yields the plasmon phase velocity v_p .

The main results are shown in Fig. 3. In each of the three devices, we extracted the plasmon phase velocity from many scanning photocurrent maps, each taken with a different gate voltage. The data have been collated into a common form by converting gate voltage to carrier density n_s (see Methods), allowing a direct comparison with theory. Qualitatively and consistent with the map in Fig. 1C, the smallest plasmon velocities are seen for the smallest n_s and d . In Fig. 3 we compare the experimental v_p values to two theories: the local approximation theory (dashed curve) shows a significant discrepancy with the data, whereas the full non-local theory (shaded curve) shows excellent agreement without any fitting parameters. Note that the local approximation predicts plasmon velocities falling below v_F for the 5.5 nm device, in contrast to the full theory which is forbidden from this region (for reasons explained below). We now proceed to describe this full non-local theory.

Figure 4A schematically depicts the three layers of our non-local theory, based on dominant effects known from electron liquid theory.^{1,2} Including all non-local corrections, the conductivity takes the following convenient form (for frequency and wavevector below Fermi values, as in this experiment)

$$\sigma(\omega, q) = \frac{e^2 v_F \sqrt{|n_s|}}{\sqrt{\pi} \hbar} \frac{i}{\omega} f\left(\frac{v_F q}{\omega}\right), \quad (2)$$

where $f(z)$ is a dimensionless function that describes the non-local response,

$$f(z) = \frac{2}{z^2} \left(\frac{1}{(1 - z^2)^{-1/2} - 1} + \delta \right)^{-1}. \quad (3)$$

Using this functional form we can gradually introduce the different layers of non-local response, which are plotted in Fig. 4B. Note that the local approximation consists of ignoring the q -dependence (which amounts to setting $f(z) = 1$), yielding a Drude response $\sigma \propto i/\omega$ from Eq. (2).

The first layer of non-local response is to consider the response of non-interacting electrons^{13–16,18} (via random phase approximation—RPA), which is the case of Eq. (3) with $\delta = 0$. In the RPA, conductivity σ increases with q , due to the change in Fermi surface deformations. This is closely related to Landau–Bohm–Gross dispersion^{23,24} in classical plasma physics: some of the electrons, those travelling with a velocity that nearly matches v_p , can interact longer with each passing wavefront and thereby provide enhanced response (Fig. 4A). Classically, this non-local dispersion would come along with Landau damping due to fast thermal electrons that fully match the plasma velocity and dissipate energy; this does not occur in a quantum degenerate system due to the narrowly-distributed electron velocity (the Fermi velocity), which instead yields a divergent intraband contribution to the conductivity as $q \rightarrow \omega/v_F$ (i.e., $z \rightarrow 1$), and no Landau damping before this point. This divergence results in the non-local plasmon velocity never falling below the Fermi velocity (as can be seen in Fig. 3), in striking contrast to the prediction of a local approximation.

The second and third layers of our non-local theory involve microscopic electron-electron interactions (many-body effects). We have calculated these many-body corrections fully consistently, including the realistic screening by capacitance $C(\omega, q)$ (see Methods). The major many-body effect is renormalization of band structure, which in graphene amounts to an increase in Fermi velocity (Fig. 4B). While the value $v_F = 1.0 \times 10^6$ m/s is nominally assumed in graphene plasmon studies, the Fermi velocity actually varies logarithmically with carrier density, from its bare value of 0.85×10^6 m/s up to as much as 3.0×10^6 m/s for very low carrier densities.^{25–29} Since our experiments enter a regime of relatively low densities, it is crucial to include this n_s -dependent velocity renormalization. The secondary many-body effect has to do with electron liquid correlations, which produce a Pauli-Coulomb hole^{1,2} around each reference electron (Fig. 4C). We capture this by including a local field factor $G(\omega, q)$, via $\sigma^{-1} = \sigma_{\text{RPA}}^{-1} + q^2 G/(i\omega C)$, that forces consistency between the dynamic response and the isothermal compressibility.² As described in the Methods, in our experimental regime this ultimately introduces a factor $\delta = 1 - \frac{\kappa_0}{\kappa}$ into Eq. (3), where κ_0 is the RPA compressibility, and κ is the proper isothermal compressibility^{15,27,30}.

Figure 4B shows how we can isolate the graphene conductivity function $\sigma(\omega, q)$ to directly observe the non-locality. This is possible due to Eq. (1), which implies that a determination of the plasmon wavevector, $q_p = q_p(\omega)$, produces a measurement of the dynamical conductivity at that wavevector: $\text{Im} \sigma(q_p, \omega) = (\omega/q_p^2) C(q_p, \omega)$. We are able to exactly calculate $C(\omega, q)$ from Maxwell equations (see Methods), and hence this approach of introducing variable d (causing variable C and variable q_p) allows us to map out $\text{Im} \sigma(q_p)$ as shown in Fig. 4. Each device (differing in d) thus provides a distinct probe of the functional dependence of the conductivity on wavevector q under otherwise-identical parameters (ω, n_s). It can be

seen in Fig. 4B that our data are only matched by theory after taking into account all three layers of quantum corrections, and that the measured conductivity of graphene shows significant departures from the local theory (a horizontal line).

The recipe set forth in this work can be transferred to probe other electron systems with exotic physical properties. Not only does this technique reveal the collective excitation (plasmon), but we have also shown how one may isolate the electronic response from its environment, quantitatively mapping out the underlying response function (non-local conductivity) as a function of wavelength. This kind of spatial spectroscopy forms a valuable counterpart to the traditional temporal (frequency) spectroscopy, and the marriage of these two approaches into a precision spatiotemporal spectroscopy—a full determination of $\sigma(\omega, q)$ —would provide an unprecedented window into electron physics. This may allow a greatly enriched understanding of electron correlation physics such as those underlying fractional quantum hall effects (e.g., magneto-rotons^{31–33}) and the binding mechanism in superconductors,³⁴ as well as probing the non-locality of Fermi-surface deformations in unusual band structures (e.g., Weyl fermions^{35–39}).

ACKNOWLEDGMENTS

F.H.L.K., M.P., and R.H. acknowledge support by the European Union Seventh Framework Programme under grant agreement no. 696656 Graphene Flagship. M.P. acknowledges support by Fondazione Istituto Italiano di Tecnologia. F.H.L.K. acknowledges financial support from the European Union Seventh Framework Programme under the ERC starting grant (307806, CarbonLight) and project GRASP (FP7-ICT-2013-613024-GRASP). F.H.L.K. acknowledges support from the Spanish Ministry of Economy and Competitiveness, through the “Severo Ochoa” Programme for Centres of Excellence in R&D (SEV-2015-0522), support by Fundacio Cellex Barcelona, CERCA Programme / Generalitat de Catalunya, the Mineco grants Ramón y Cajal (RYC-2012-12281), Plan Nacional (FIS2013-47161-P and FIS2014-59639-JIN), and support from the Government of Catalonia through the SGR grant (2014-SGR-1535).

P.A.-G. and R.H. acknowledge support from the European Union through ERC starting grant (TERATOMO grant no. 258461) and the Spanish Ministry of Economy and Competitiveness (national project MAT2012-36580). Y.G., C.T., and J.H. acknowledge support from the US Office of Naval Research N00014-13-1-0662. C.T. was supported under contract FA9550-11-C-0028 and awarded by the Department of Defense, Air Force Office of Scientific Research, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a. This research used resources of the Center for Functional Nanomaterials, which is a U.S. DOE Office of Science Facility, at Brookhaven National Laboratory under Contract No. DE-SC0012704. B.v.D. acknowledges support from the Flemish Science Foundation (FWO-VI) by a post-doctoral Fellowship. P.A.-G. acknowledges financial sup-

port from the national project FIS2014-60195-JIN and the ERC starting grant 715496, 2DNANOPTICA. This work used open source software (www.python.org, www.matplotlib.org, www.povray.org).

ing scattering-type scanning near-field optical microscope systems such as the ones used in this study. All other authors declare no competing financial interests.

COMPETING FINANCIAL INTERESTS

R.H. is co-founder of Neaspec GmbH, a company produc-

-
- ¹ D. Pines, P. Nozières, *The Theory of Quantum Liquids* (W.A. Benjamin, 1966).
- ² G. Giuliani, G. Vignale, *Quantum Theory of the Electron Liquid* (Cambridge University Press, 2005).
- ³ A. vom Felde, J. Sprösser-Prou, J. Fink, *Phys. Rev. B* **40**, 10181 (1989).
- ⁴ C. F. Hirjibehedin, A. Pinczuk, B. S. Dennis, L. N. Pfeiffer, K. W. West, *Phys. Rev. B* **65**, 161309 (2002).
- ⁵ C. Ciraci, *et al.*, *Science* **337**, 1072 (2012).
- ⁶ J. A. Scholl, A. L. Koh, J. A. Dionne, *Nature* **483**, 421 (2012).
- ⁷ M. S. Tame, *et al.*, *Nat. Phys.* **9**, 329 (2013).
- ⁸ X. Li, D. Xiao, Z. Zhang, *New J. Phys.* **15**, 023011 (2013).
- ⁹ W. Zhu, *et al.*, *Nat. Commun.* **7**, 11495 (2016).
- ¹⁰ P. Alonso-González, *et al.*, *Nat. Nano.* **12**, 31 (2017).
- ¹¹ Z. Fei, *et al.*, *Nature* **487**, 82–85 (2012).
- ¹² J. Chen, *et al.*, *Nature* **487**, 77 (2012).
- ¹³ B. Wunsch, T. Stauber, F. Sols, F. Guinea, *New J. Phys.* **8**, 318 (2006).
- ¹⁴ L. A. Falkovsky, A. A. Varlamov, *Eur. Phys. J. B* **56**, 281 (2007).
- ¹⁵ E. H. Hwang, B. Y.-K. Hu, S. Das Sarma, *Phys. Rev. Lett.* **99**, 226801 (2007).
- ¹⁶ A. Principi, M. Polini, G. Vignale, *Phys. Rev. B* **80**, 075418 (2009).
- ¹⁷ A. Woessner, *et al.*, *Nat. Mat.* **14**, 421 (2015).
- ¹⁸ V. Ryzhii, *Jpn. J. Appl. Phys.* **45**, L923 (2006).
- ¹⁹ L. Wang, *et al.*, *Science* **342**, 614 (2013).
- ²⁰ A. Woessner, *et al.*, *Nat. Commun.* **7**, 10783 (2016).
- ²¹ M. B. Lundeberg, *et al.*, *Nat. Mat.* **16**, 204 (2016).
- ²² Z. Fei, *et al.*, *Nano Lett.* **11**, 4701–4705 (2011).
- ²³ L. Landau, *J. Phys. USSR* **10**, 45 (1946).
- ²⁴ D. Bohm, E. P. Gross, *Phys. Rev.* **75**, 1851 (1949).
- ²⁵ S. H. Abedinpour, *et al.*, *Phys. Rev. B* **84**, 045429 (2011).
- ²⁶ D. C. Elias, *et al.*, *Nat. Phys.* **7**, 701 (2011).
- ²⁷ M. Orlita, *et al.*, *New J. Phys.* **14**, 095008 (2012).
- ²⁸ S. Das Sarma, E. H. Hwang, *Phys. Rev. B* **87**, 045425 (2013).
- ²⁹ L. S. Levitov, A. V. Shtyk, M. V. Feigelman, *Phys. Rev. B* **88**, 235403 (2013).
- ³⁰ G. L. Yu, *et al.*, *Proc. Natl. Acad. Sci.* **110**, 3282 (2013).
- ³¹ S. M. Girvin, A. H. MacDonald, P. M. Platzman, *Phys. Rev. B* **33**, 2481 (1986).
- ³² I. V. Kukushkin, J. H. Smet, V. W. Scarola, V. Umansky, K. von Klitzing, *Science* **324**, 1044 (2009).
- ³³ F. D. M. Haldane, *Phys. Rev. Lett.* **107**, 116801 (2011).
- ³⁴ D. A. Kirzhnits, E. G. Maksimov, D. I. Khomskii, *J. Low Temp. Phys.* **10**, 79 (1973).
- ³⁵ S.-Y. Xu, *et al.*, *Science* **349**, 613 (2015).
- ³⁶ J. Hofmann, S. Das Sarma, *Phys. Rev. B* **93**, 241402 (2016).
- ³⁷ F. M. D. Pellegrino, M. I. Katsnelson, M. Polini, *Phys. Rev. B* **92**, 201407 (2015).
- ³⁸ A. A. Soluyanov, *et al.*, *Nature* **527**, 495 (2015).
- ³⁹ J. C. Song, M. S. Rudner, *arXiv preprint arXiv:1702.00022* (2017).

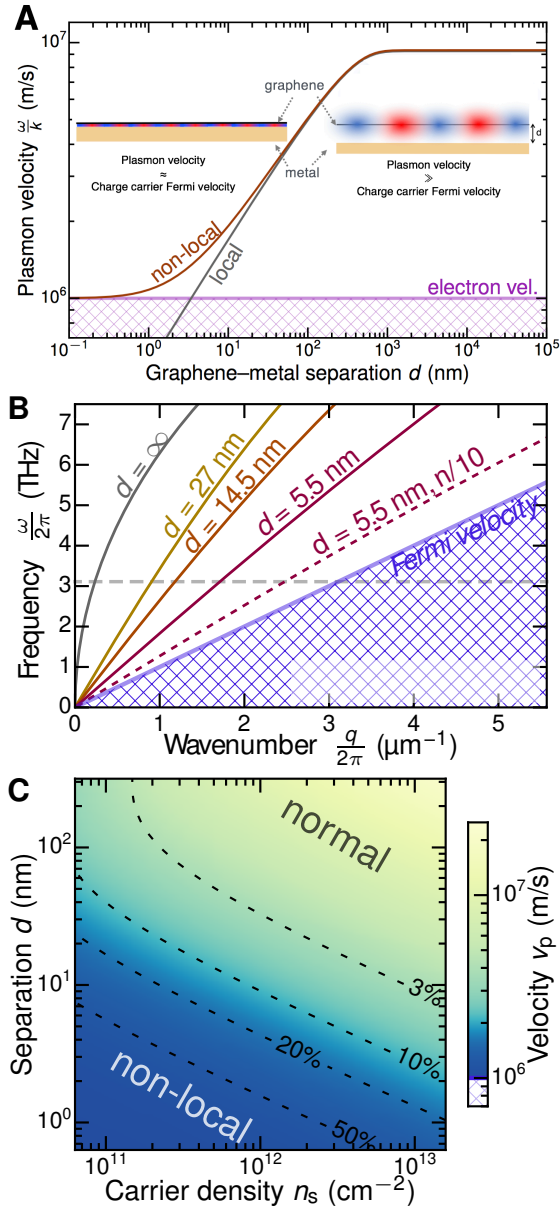


FIG. 1. **Concept.** (A) Effect on graphene plasmon velocity from changing separation d in the graphene-metal system. Insets show plasmon electric fields for large or small separation d . (B) Frequency-wavenumber dispersion of plasmon at various d ; the solid lines are all computed with equal carrier density $n_s = 10^{12} \text{ cm}^{-2}$, whereas the dashed line shows the smallest- d case with a factor 10 lower carrier density $n_s = 10^{11} \text{ cm}^{-2}$. Horizontal dashed gray line indicates frequency for which the experiment has been performed. (C) Plasmon velocity dependence on d and carrier density n_s . Contours indicate discrepancy between local and non-local plasmon models.

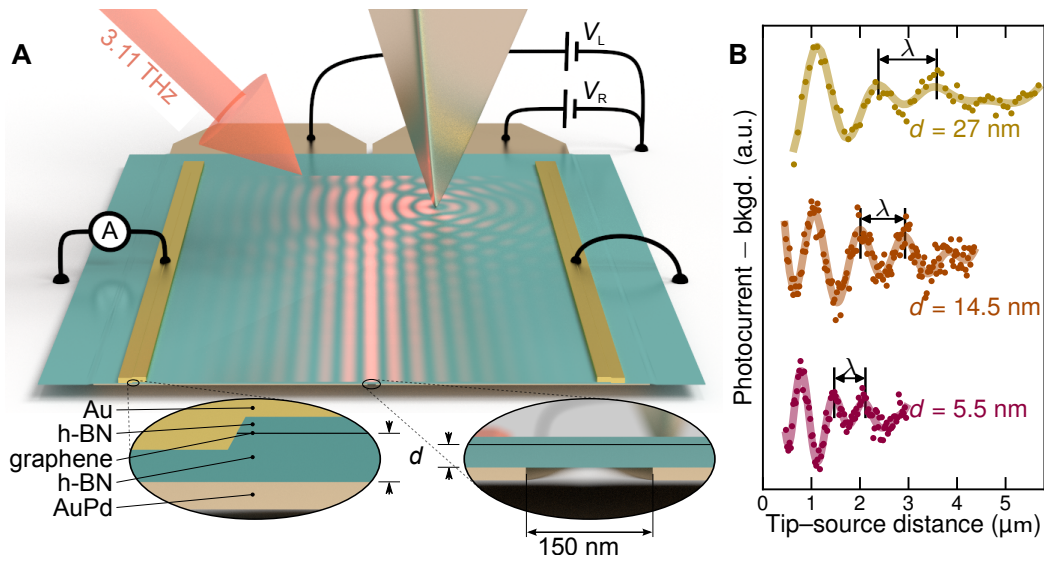


FIG. 2. **Experiment.** (A) A metallized tip (inverted pyramid) scans over a graphene sheet that has been encapsulated in h-BN and placed on a split metallic film. Terahertz laser light illuminates the entire device, launching plasmons (orange arrows) at the tip and split. Gate voltages V_L and V_R control the electron density and the junction photocurrent sensitivity. (B) Photocurrent traces in three different devices, each at $n_s = 1 \times 10^{12} \text{ cm}^{-2}$, showing interference fringes used to extract the plasmon wavelength λ (and hence velocity $v_p = \lambda \frac{\omega}{2\pi}$) via the indicated fits.

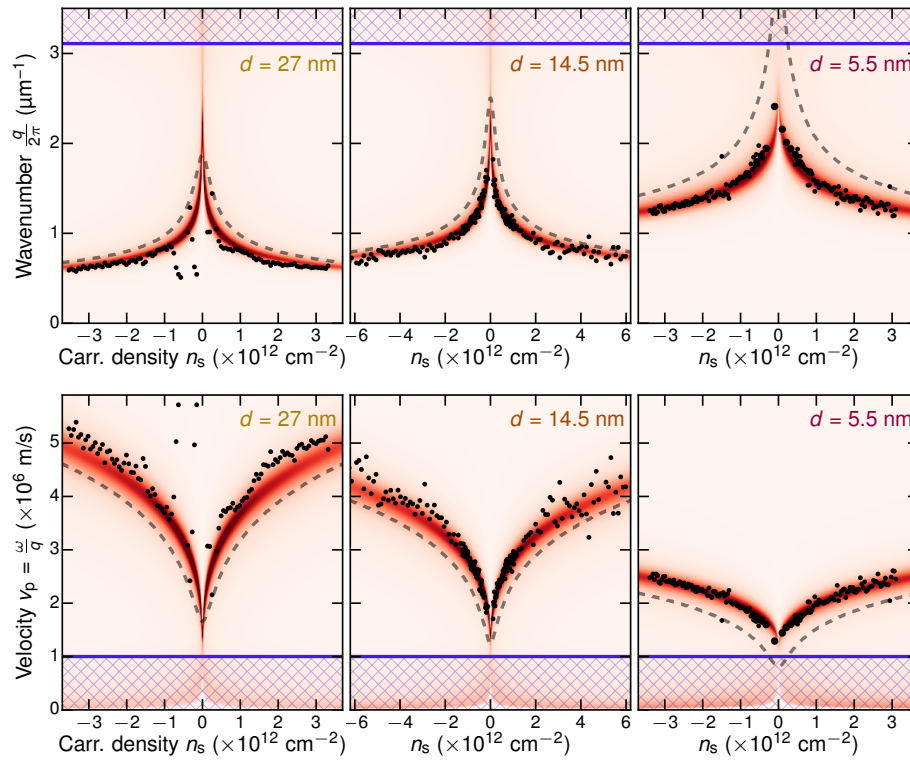


FIG. 3. **Tunable non-local effects.** Extracted plasmon wavelength dependence on carrier density n_s , for three devices of differing separation d , show parameter-free agreement with the full theory (color map) and significant deviation from a local-response theory (dashed line). The red color map indicates the inverse of the left-hand-side of Eq. (1), so that plasmons appear as a red peak, the width being associated with propagation distance. The hatched region below the solid line indicates phase velocities below v_F . The upper and lower rows show the same data, plotted differently.

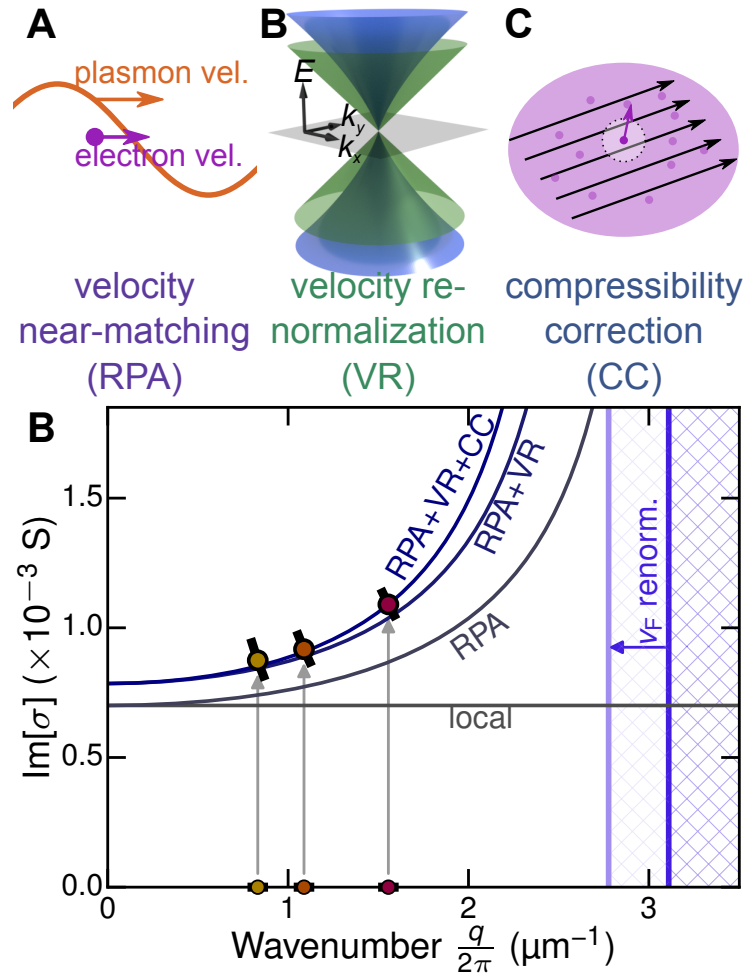


FIG. 4. **Non-local conductivity of graphene.** (A,B,C) Schematic representations of the three main mechanisms governing graphene response beyond the local approximation. (D) Experimentally-extracted $\sigma(\omega, q)$ at $n_s = 1 \times 10^{12} \text{ cm}^{-2}$, compared to theoretical approximations for the interacting electron system in graphene: random phase approximation (RPA), with added velocity renormalization (RPA+VR), and then with compressibility correction (RPA+VR+CC); local RPA appears as a horizontal line.