LETTER

Fundamental limits to graphene plasmonics

G. X. Ni[1,2], A. S. McLeod[1,2], Z. Sun[2], L. Wang[3], L. Xiong[1,2], K. W. Post[2], S. S. Sunku[1,4], B.-Y. Jiang[2], J. Hone[3], C. R. Dean[1], M. M. Fogler[2] & D. N. Basov[1,2]*

Plasmon polaritons are hybrid excitations of light and mobile electrons that can confine the energy of long-wavelength radiation at the nanoscale. Plasmon polaritons may enable many enigmatic quantum effects, including lasing[1], topological protection[2,3] and dipole-forbidden absorption[4]. A necessary condition for realizing such phenomena is a long plasmonic lifetime, which is notoriously difficult to achieve for highly confined modes[5]. Plasmon polaritons in graphene—hybrids of Dirac quasiparticles and infrared photons—provide a platform for exploring light–matter interaction at the nanoscale[6,7]. However, plasmonic dissipation in graphene is substantial[8] and its fundamental limits remain undetermined. Here we use nanometre-scale infrared imaging to investigate propagating plasmon polaritons in high-mobility encapsulated graphene at cryogenic temperatures. In this regime, the propagation of plasmon polaritons is primarily restricted by the dielectric losses of the encapsulated layers, with a minor contribution from electron–phonon interactions. At liquid-nitrogen temperatures, the intrinsic plasmonic propagation length can exceed 10 micrometres, or 50 plasmonic wavelengths, thus setting a record for highly confined and tunable polariton modes. Our nanoscale imaging results reveal the physics of plasmonic dissipation and will be instrumental in mitigating such losses in heterostructure engineering applications.

Here we investigate plasmon polariton propagation and dissipation using near-field infrared microscopy, enabling a direct visualization of polaritonic standing waves on the surface of graphene and other van der Waals materials[6,7,9,10]. Both the wavelength of plasmon polariton waves, λ_p, and their travel range, which quantifies their dissipation, can be readily confirmed from nanoscale infrared images (Figs. 1, 2). These two observables permit direct extraction of the response functions of a plasmonic medium, including the complex conductivity $\sigma(\omega) = \sigma' + i\sigma''$, as we detail below. The response functions encode information on both interactions and scattering processes in an electron liquid. Thus, near-field examination of plasmon polariton images facilitates inquiry into fundamental electronic phenomena of the media supporting these polaritons. As customary in solid-state physics, scattering processes of diverse origin can be distinguished by investigating their dependence on control parameters—in particular, temperature. This motivates a plasmonic imaging investigation of graphene at cryogenic temperatures with the help of a newly developed apparatus[11]. We examined a high-mobility micro-device constructed from a graphene monolayer encapsulated between two thin slabs of hexagonal boron nitride (hBN), with a graphene carrier density that is tunable through a silicon back gate (Fig. 1a, b). The encapsulation preserved the graphene in its pristine form[12]. Our cryo-imaging approach to recording plasmon polariton standing waves, together with a theoretical model free of adjustable parameters, has enabled us to identify the physics governing plasmon propagation in high-mobility graphene.

Nanoscale imaging of plasmon polaritons is achieved using the following experimental protocol. The metallized tip of an atomic force microscope (AFM) probe is illuminated by a focused laser field with an infrared frequency of ω. The tip functions as an optical antenna that intensifies the incident electric field at the apex. When the tip

is brought near the sample, this concentrated field can excite polaritonic modes with a wavelength of $\lambda_p(\omega)$ that is much shorter than the free-space wavelength $\lambda_{IR} = 2\pi/\omega$ of the incident light[13]. The optical antenna functionality of the AFM tip also facilitates scattering of local near fields back into the far field, where they are registered by conventional optical detection. Proper demodulation methods allow one to isolate the back-scattered light component that is associated with the local electric field confined to the area under the tip, thus achieving a spatial resolution of the order of 10 nm (ref. [14]). In such a near-field imaging experiment, plasmon polaritons manifest as a periodic modulation (fringes) of the observed near-field signal as a function of the tip position. Similarly to imaging experiments at room temperature[8], fringes of two distinct periodicities, λ_p and $\lambda_p/2$, appear in Figs. 1 and 2. The $\lambda_p/2$-period fringes are due to plasmon polaritons that are emitted by the tip and complete a round-trip between the tip and reflection from the sample edges. The λ_p-period fringes are produced by the interference between tip-launched plasmon polaritons and the evanescent component of the reflected ones[15]. Alternatively, the λ_p-period fringes
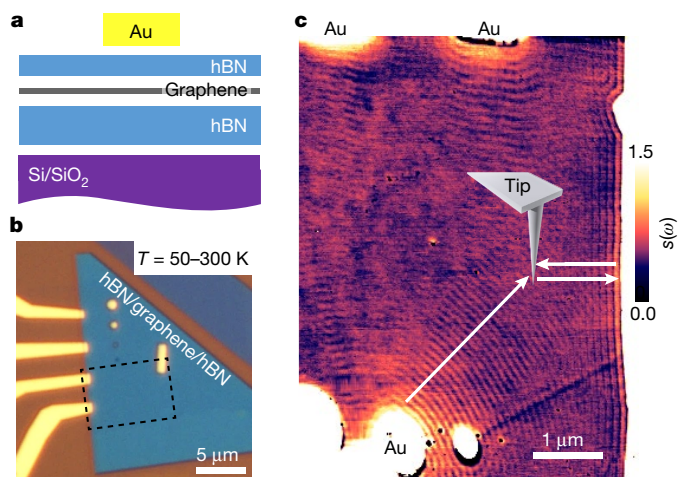

Fig. 1 | Nanoscale infrared imaging of surface plasmons in Au/hBN/graphene/hBN encapsulated structures at cryogenic temperature. **a**, Sketch of the layered Au/hBN/graphene/hBN/SiO₂/Si heterostructure. The lithographically defined gold (Au) microstructures on top of the hBN capping layer serve as effective plasmonic launchers. **b**, Optical image of the device. The black dashed rectangle marks the area shown in **c**. The entire field of view is within the diffraction-limited spot of our infrared laser, which operates at an energy of 886 cm⁻¹ (110 meV) or $\lambda_{IR} = 11.28\,\mu m$. **c**, Nanoscale infrared image of plasmonic interference fringes from the encapsulated graphene monolayer, expressed by the normalized scattering amplitude s acquired at a back-gate voltage of 97 V and a temperature of $T = 60$ K. The arrows represent the propagation direction of the plasmon waves. These experiments simultaneously visualize the local electric field associated with interference from plasmon polaritons emitted by the near-field probe and reflected by sample edges, as well as from polaritons emitted by the gold microstructures (labelled as Au).

[1]Department of Physics, Columbia University, New York, NY, USA. [2]Department of Physics, University of California, San Diego, La Jolla, CA, USA. [3]Department of Mechanical Engineering, Columbia University, New York, NY, USA. [4]Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, USA. [5]These authors contributed equally: G. X. Ni, A. S. McLeod. *e-mail: db3056@columbia.edu
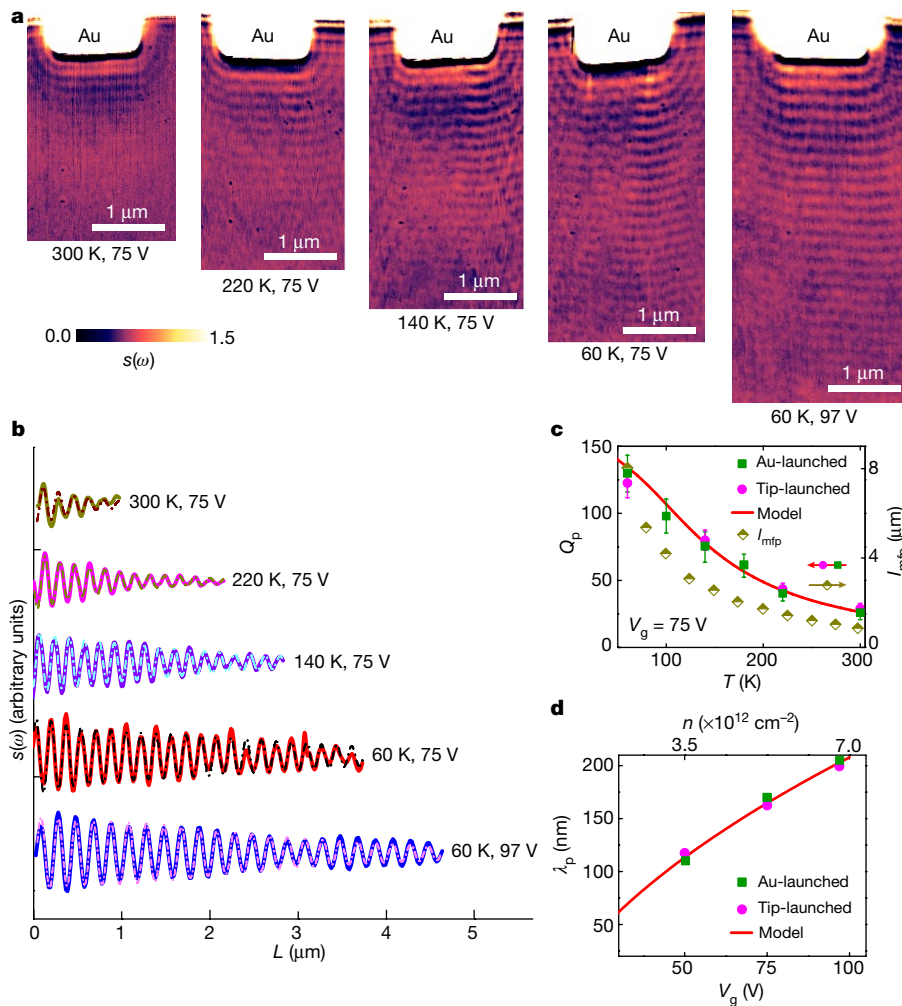
Fig. 2 | Temperature- and gate-dependent trends in surface plasmon propagation in graphene. a, Nanoscale infrared images of the normalized scattering amplitude s acquired at sequential sample temperatures and gate voltages. A gold electrode (labelled Au at the top of the images) functions as an antenna that emits graphene plasmons. **b**, Line profiles of plasmonic interference fringes propagating from left to right, as a function of the distance L from the gold launcher. The variable attenuation of the propagating waveform (solid curves) is characteristic of the temperature-dependent plasmon scattering rate. Dash-dotted lines represent the results of numerical simulations performed to identify the temperature dependence of the complex plasmon wavevector and associated scattering rate (see main text and Methods). **c**, The temperature dependence of the quality factor Q_p for both λ_p- and $\lambda_p/2$-period plasmon waves obtained from nanoscale infrared images taken at $V_g = 75$ V. The red solid curve is the theoretical model detailed in Methods. The electronic mean free path $l_{mfp}(T)$ is also plotted for comparison. The error bars define the 95% confidence intervals. **d**, The voltage (V_g) and carrier density (n) dependence of the plasmon wavelength, obtained from nanoscale infrared data for both λ_p- and $\lambda_p/2$-period plasmonic patterns. The red solid curve shows a $\lambda_p \propto \sqrt{V_g}$ dependence, in accord with previous results[9,17].

can arise from plasmon polaritons launched by emitters other than the tip that are present in the device[8]. The efficiency of the plasmon polariton emission depends on the dielectric polarizability, shape and size of the emitter. Specifically, graphene edges are weak emitters, but elongated metallic objects placed on graphene can act as efficient ones[16]. In our structures, the gold contacts on top of the hBN-encapsulating layer act as fixed plasmonic antennas (Figs. 1, 2).

We begin with a survey of a large-area ($6\,\mu m \times 8\,\mu m$) image of plasmon polariton standing waves obtained at $T = 60$ K (Fig. 1c) with an infrared laser operating at $\lambda_{IR} = 11.28\,\mu m$ and at a back-gate voltage of $V_g = 97$ V. Here we display raw data in the form of the amplitude s of the scattered field, normalized by the amplitude detected at the gold contacts, whose optical response provides a convenient reference that is independent from temperature and carrier density (see Methods for details). The most prominent aspect of the image in Fig. 1c is that the entire field of view is filled with plasmonic interference fringes. The $\lambda_p/2$-period fringes are most evident on the right-hand side of the field of view, close to a natural boundary of the graphene monolayer, whereas λ_p-period fringes dominate in the vicinity of gold antennas. Figure 1c reveals that plasmon polaritons are highly confined, with $\lambda_{IR}/\lambda_p > 60$; however, they travel over several micrometres, far exceeding previous benchmark results at room temperature[8,15].

The nanoscale infrared images and corresponding line profiles in Fig. 2 attest to a dramatic reduction of plasmonic losses as the temperature of the graphene is reduced. Figure 2a shows detailed raster scans obtained in the vicinity of one of the gold antennas at $V_g = 75$ V and at temperatures between 60 K and 300 K. When the gate voltage increases above $V_g = 50$ V, plasmonic fringes emerge out of a featureless background, as illustrated in a voltage-dependent scan shown in Extended

Data Fig. 3. At room temperature, plasmon polariton propagation is restricted to distances of less than $1\,\mu m$. As the temperature is lowered, we observe a systematic increase of both the overall travel range and the number of detectable fringes. At $T = 60$ K plasmonic oscillations extend beyond $4\,\mu m$ at $V_g = 75$ V and beyond $5\,\mu m$ at $V_g = 97$ V (Fig. 2a, b). These overall trends are in accord with the temperature dependence of the electronic mean free path, l_{mfp}, in high-mobility devices prepared under identical protocols[12]. In fact, even the magnitudes of the plasmon polariton travel range and of l_{mfp} appear to be on par with each other: l_{mfp} increases from $1\,\mu m$ at $T = 300$ K up to $8\,\mu m$ at 60 K (Fig. 2c). The latter value exceeds the entire field of view in Fig. 2a and is consistent with the notion of ballistic electronic transport. The novelty of the images in Fig. 2 is that they manifest the real space behaviour of graphene plasmon polaritons supported by ballistic electrons.

To quantify the quasiparticle dynamics underlying the propagation of plasmon polaritons in graphene, we present an analysis based on the complex momentum of the polariton, $q_p = q_p' + iq_p''$. The real part, q_p', defines the plasmon polariton wavelength through $\lambda_p = 2\pi/q_p'$, whereas the imaginary part, q_p'', quantifies dissipation and the quality factor $Q_p = q_p'/q_p''$ (Table 1). The complex momentum of plasmon polaritons, and thus both λ_p and Q_p, are governed by the optical conductivity of graphene (see Methods section 'Plasmons near a graphene edge') or, alternatively, can be obtained through numerical analysis of plasmonic fringes (Methods section 'Plasmonic damping rate analysis'). The evolution of λ_p with the gate voltage (or carrier density) is presented in Fig. 2d and is in complete accord with the results of earlier studies at room temperature[8–10,15,17]. A noteworthy aspect of our data is the rapid growth of the quality factor at low temperatures (Fig. 2c).

Table 1 | Comparison of graphene plasmon parameters with those of metals (Ag) and doped semiconductors (n-doped InSb and CdO)

	Confinement ratio	Quality factor	Lifetime (fs)
Definition	λ_{IR}/λ_p	$Q_p = q'_p/q''_p$	$\tau = 2Q_p/\omega$
Graphene (experimental, $T = 60$ K)	66	130	1,600
Graphene (intrinsic, $T = 60$ K)	66	970	12,000
Ag[a,b] ($T = 10$ K)	~1	36	14
n-InSb[c], n-CdO[a] ($T = 300$ K)	<10	37	270

Parameters relevant for graphene entries: frequency, $\omega/(2\pi c) = 886$ cm^{-1}; gate voltage, $V_g = 75$ V; plasmon wavelength, $\lambda_p = 170$ nm; and dielectric loss of the substrate, $\gamma_z = 5.9$ cm^{-1} (see Methods for details).
[a]Data from ref. [29].
[b]Data from ref. [30].
[c]Data from ref. [31].

This temperature dependence tracks the variation of the electronic mean free path. We therefore surmise that the evolution of $Q_p(T)$ reflects the same fundamental physics as that governing electronic transport in graphene, but probed at finite frequency ω. The highest quality factor detected in our plasmonic imaging experiments is $Q_p = 130$.

A detailed quantitative analysis of finite-frequency dissipation of Dirac quasiparticles requires that intrinsic dissipation in graphene is distinguished from extrinsic losses associated with the dielectric environment of graphene. We outline this procedure in Methods section 'Plasmonic damping rate analysis'. In Fig. 3a we show the plasmonic scattering rate $\gamma_p(T) = \omega\sigma'/\sigma''$, which quantifies intrinsic dissipation originating from within the graphene monolayer. Inspection of these data reveals a dramatic drop in the intrinsic plasmonic scattering rate from about 20 cm^{-1} at room temperature to <2.0 cm^{-1} at $T = 60$ K (we note the frequency unit conversion rule: 1 cm$^{-1} = 30$ GHz $= 1.43$ K). The corresponding direct current (d.c.) scattering rate measured for a similar encapsulated device is also plotted in Fig. 3a, showing the same overall trend in temperature dependence. Another important finding reported here is that the plasmonic scattering rate exceeds the transport scattering rate by a factor of two at room temperature. Notably, the same physical model (blue and red lines in Fig. 3a), which is free of adjustable parameters, yields a consistent description of the dissipation encountered by Dirac quasiparticles in both d.c. transport and at a finite frequency in the form of collective plasmon polaritons.

We now outline the key elements that contribute to d.c. and plasmonic scattering in our model. A substantial contribution to plasmon polariton losses is from acoustic phonons, as reported in refs [18,19]. It is practical to discriminate between two distinct loss channels associated with acoustic phonons, which in the related literature are commonly

referred to as the deformation potential effect and the pseudo-magnetic field effect. The deformation potential effect shifts the energy of Dirac bands (Fig. 3d), whereas the pseudo-magnetic field effect displaces the overall momentum of Dirac cones in opposite directions for the two valleys of the Brillouin zone at K and K′ (see Fig. 3d and Methods section 'Plasmonic damping rate analysis'). First-principle calculations show that the deformation potential channel is vanishingly small[20,21], leaving the pseudo-magnetic field component as the dominant acoustic scatterer. However, the pseudo-magnetic field effect yields a conventional linear temperature dependence of the quasiparticle scattering rate that is nearly indistinguishable for d.c. scattering and plasmonic scattering at $\omega = 886$ cm^{-1} (orange and blue lines in Fig. 3b). This finding prompted a search for scattering mechanisms with a pronounced frequency dependence. Such a contribution emerges from inter-valley electron–phonon scattering, with scattering rate γ_K (black and red lines in Fig. 3b). The matrix element used to numerically evaluate γ_K requires the knowledge of both the energy of the zone-boundary optical phonon of graphene, ω_K, and the electron–phonon coupling constant, β_K, for which we used the results of prior first-principles calculations[20] (Methods section 'Plasmonic damping rate analysis'). The physical mechanism behind the frequency dependence of γ_K is the thermally activated nature of phonon-mediated inter-valley scattering, which yields a scattering rate proportional to $\exp[-(\omega_K - \omega)/T]$, where $\omega_K = 1,200$ cm^{-1}. An additional small contribution to plasmon polariton damping comes from electron–electron scattering, with scattering rate γ_{ee} (magenta line in Fig. 3b). The net result is that three added channels—the pseudo-magnetic field effect, inter-valley scattering and electron–electron scattering—reproduce the observed temperature dependence of both d.c. and plasmonic scattering measured for ultrahigh-mobility encapsulated graphene. We stress that our scattering rate analysis involves accurate evaluation of the temperature-dependent losses in hBN microcrystals that are nominally identical to those used in the actual graphene structures (Methods section 'The optical constants of hBN'). This procedure enabled the accurate decomposition of various scattering mechanisms and thus enhanced the reliability of our findings.

Our experimental results, together with our unified parameter-free theoretical analysis, account for quasiparticle dynamics in ultra-high-mobility graphene, even at finite frequencies ω relevant for graphene plasmonics. Our experimental values of the quality factor Q_p and the lifetime τ of plasmons in graphene, extracted from cryogenic imaging, are higher than those of any other materials used for infrared plasmonics (Table 1). Although limited by acoustic-phonon scattering, the intrinsic propagation length $l_p = 1/(2q''_p)$ of plasmon polaritons in our high-mobility structures can exceed 13 µm at low temperatures (the intrinsic quality factor is $Q_p > 970$; Table 1). In this regime, transport
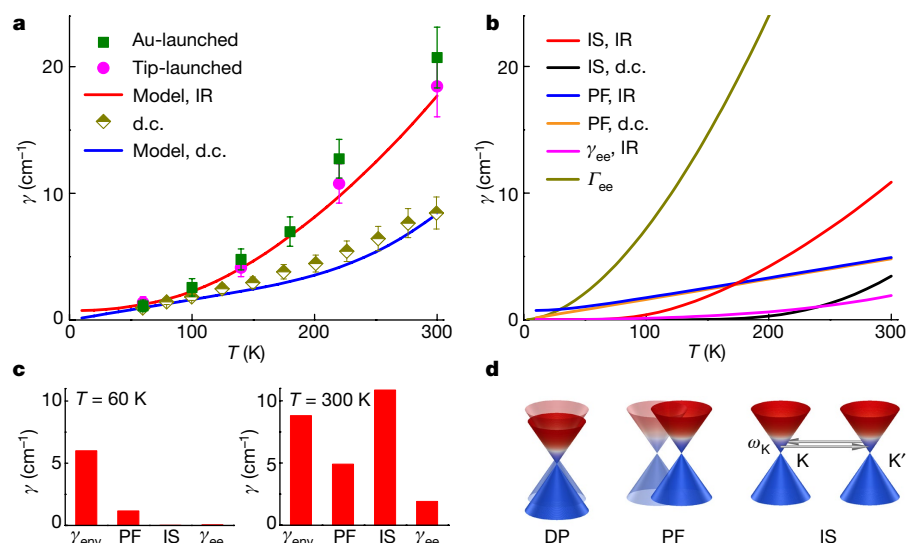
Fig. 3 | Plasmonic and electronic transport in high-mobility graphene. a, Temperature dependence of the plasmonic scattering rate (squares and circles) and of the d.c. scattering rate (diamonds). Solid lines display the results of our parameter-free modelling. The error bars represent 95% confidence intervals. **b**, Key contributions to plasmonic and electronic scattering and their temperature dependences. **c**, Histograms displaying the contribution of different scattering channels at $T = 60$ K (left) and 300 K (right). **d**, Schematics of transformations of graphene electronic structure associated with the deformation potential (DP), pseudo-magnetic field (PF) and inter-valley scattering (IS). The arrows in denote the possible scattering path across the K and K′ valleys.

of plasmon polaritons is dominated by the dielectric environment and can be limited by the physical dimensions of macroscopic devices. Although losses associated with the encapsulating hBN layers and SiO₂ are extrinsic to graphene, dielectrics are practically unavoidable in actual high-mobility graphene devices. At ambient conditions, losses from (1) the dielectric layers, (2) acoustic-phonon scattering and (3) inter-valley scattering all sum together with unequal contributions. The dielectric losses can be reduced by 20% by using a thicker hBN bottom layer to increase the distance between the lossier SiO₂ and graphene. However, the other two mechanisms are more difficult to circumvent. We stress that the contribution of electron–electron scattering to finite-frequency plasmon damping is small, even though the electron–electron collision rate Γ_{ee} is comparable to the electron–phonon scattering rate (see Fig. 3b and Methods section 'Plasmonic damping rate analysis'). In general, electronic interactions can influence the conductivity only if the Galilean invariance of the system is broken, which is a weak effect in our high-carrier-density graphene at low temperature[22]. Our data imply that the contribution of the electron–electron interaction to plasmonic responses is more prominent at higher temperatures, at lower carrier densities or in samples of special geometry. However, a major change in the electrodynamic response should occur at frequencies below the electron–electron collision rate Γ_{ee}. In that regime (called the hydrodynamic regime), which has been recently studied in d.c. transport experiments[23,24], the plasmons are expected to become overdamped by another type of collective modes: energy waves or 'demon' modes[22,25]. These modes could be studied using terahertz near-field microscopy combined with photocurrent detection[26]. Finally, we note that the extrinsic plasmonic de-phasing is surprisingly weak in many macroscopic areas of Fig. 1c on which we focused our analysis, whereas its presence is obvious in other areas. These remaining extrinsic limitations on plasmon polariton propagation, including the possible role of sample boundaries[27], will be a subject of future nanoscale imaging studies. Nevertheless, this first report of plasmonic nanoscale imaging studies of encapsulated graphene at cryogenic temperatures demonstrates potential for the exploration of plasmonic switching and nonlinear phenomena in ultrahigh-mobility graphene at the liquid-nitrogen temperatures that are typically used for mid-infrared technologies. Besides imposing fundamental limits on plasmon damping, the electron–phonon interaction probed here may have a role in other many-body phenomena, such as the recently discovered superconductivity in twisted bilayer graphene[28].

Online content

Any Methods, including any statements of data availability and Nature Research reporting summaries, along with any additional references and Source Data files, are available in the online version of the paper at https://doi.org/10.1038/s41586-018-0136-9.

1. Lu, Y. J. et al. Plasmonic nanolaser using epitaxially grown silver film. *Science* **337**, 450–453 (2012).
2. Jin, D. et al. Infrared topological plasmons in graphene. *Phys. Rev. Lett.* **118**, 245301 (2017); erratum **119**, 019901 (2017).
3. Karzig, T., Bardyn, C. E., Lindner, N. H. & Refael, G. Topological polaritons. *Phys. Rev. X* **5**, 031001 (2015).
4. Rivera, N., Kaminer, I., Zhen, B., Joannopoulos, J. D. & Soljacic, M. Shrinking light to allow forbidden transitions on the atomic scale. *Science* **353**, 263–269 (2016).
5. Boltasseva, A. & Atwater, H. A. Low-loss plasmonics metamaterials. *Science* **331**, 290–291 (2011).
6. Basov, D. N., Fogler, M. M. & Garcia de Abajo, F. J. Polaritons in van der Waals materials. *Science* **354**, aag1992 (2016).
7. Low, T. et al. Polaritons in layered two-dimensional materials. *Nat. Mater.* **16**, 182–194 (2017).
8. Woessner, A. et al. Highly confined low-loss plasmons in graphene–boron nitride heterostructures. *Nat. Mater.* **14**, 421–425 (2015).
9. Fei, Z. et al. Gate-tuning of graphene plasmons revealed by infrared nano-imaging. *Nature* **487**, 82–85 (2012).
10. Chen, J. et al. Optical nano-imaging of gate-tunable graphene plasmons. *Nature* **487**, 77–81 (2012).
11. McLeod, A. S. et al. Nanotextured phase coexistence in the correlated insulator V₂O₃. *Nat. Phys.* **13**, 80–86 (2017).
12. Wang, L. et al. One-dimensional electrical contact to a two-dimensional material. *Science* **342**, 614–617 (2013).
13. Novotny, L. & Hecht, B. *Principles of Nano-Optics* (Cambridge Univ. Press, Cambridge, 2006).
14. Atkin, J. H., Berweger, S., Jones, A. C. & Raschke, M. B. Nano-optical imaging and spectroscopy of order, phases, and domains in complex solids. *Adv. Phys.* **61**, 745–842 (2012).
15. Ni, G. X. et al. Ultrafast optical switching of infrared plasmon polaritons in high-mobility graphene. *Nat. Photon.* **10**, 244–247 (2016).
16. Alonso-Gonzalez, P. et al. Controlling graphene plasmons with resonant metal antennas and spatial conductivity patterns. *Science* **344**, 1369–1373 (2014).
17. Ni, G. X. et al. Plasmons in graphene moiré superlattices. *Nat. Mater.* **14**, 1217–1222 (2015).
18. Principi, A. et al. Plasmon losses due to electron-phonon scattering: the case of graphene encapsulated in hexagonal Boron Nitride. *Phys. Rev. B* **90**, 165408 (2014).
19. Hwang, E. & Das Sarma, S. Acoustic phonon scattering limited carrier mobility in two-dimensional extrinsic graphene. *Phys. Rev. B* **77**, 115449 (2008).
20. Park, C. H. et al. Electron-phonon interactions and the intrinsic electrical resistivity of graphene. *Nano Lett.* **14**, 1113–1119 (2014).
21. Sohier, T., Calandra, M. & Mauri, F. Density-functional calculation of static screening in two-dimensional materials: the long-wavelength dielectric function of graphene. *Phys. Rev. B* **91**, 165428 (2015).
22. Sun, Z., Basov, D. N. & Fogler, M. M. Universal linear and nonlinear electrodynamics of a Dirac fluid. *Proc. Natl Acad. Sci. USA* **115**, 3285–3289 (2018).
23. Bandurin, D. A. et al. Negative local resistance caused by viscous electron backflow in graphene. *Science* **351**, 1055–1058 (2016).
24. Kumar, R. K. et al. Superballistic flow of viscous electron fluid through graphene constrictions. *Nat. Phys.* **13**, 1182 (2017).
25. Phan, T. V., Song, J. C. W. & Levitov, L. S. Ballistic heat transfer and energy waves in an electron system. Preprint at https://arxiv.org/abs/1306.4972 (2013).
26. Lundeberg, M. et al. Tuning quantum non-local effects in graphene plasmonics. *Science* **357**, 187–191 (2017).
27. Halbertal, D. et al. Imaging resonant dissipation from individual atomic defects in graphene. *Science* **358**, 1303–1306 (2017).
28. Cao, Y. et al. Unconventional superconductivity in magic-angle graphene superlattices. *Nature* **556**, 43–50 (2018).
29. Caldwell, J. D. et al. Low-loss, infrared and terahertz nanophotonics using surface phonon polaritons. *Nanophotonics* **4**, 44–68 (2014).
30. Sun, L. et al. Enhancement of plamonic performance in epitaxial silver at low temperature. *Sci. Rep.* **7**, 8917 (2017).
31. Chochol, J. et al. Plasmonic behavior of III–V semiconductor in far-infrared and terahertz range. *J. Eur. Opt. Soc.-Rapid* **13**, 13 (2017).

Author contributions G.X.N. and A.S.M. performed the nanoscale infrared measurements and characterizations. Z.S., L.X., B.-Y.J. and M.M.F. provided theoretical calculations. L.W., J.H. and C.R.D. designed and created the device structures. G.X.N., L.X., K.W.P. and S.S.S. performed far-field optical spectroscopy measurements and characterizations. D.N.B. supervised the project. G.X.N., Z.S., M.M.F. and D.N.B. co-wrote the manuscript with input from all co-authors.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0136-9.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to D.N.B.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Device fabrication and characterization. Our high-quality hBN/graphene/hBN devices were fabricated using a polymer-free multilevel stacking dry-transfer process to form the encapsulated graphene structure[12]. Briefly, a very thin (7 nm) hBN crystal was used to pick up single-layer graphene and another hBN crystal in sequence. Bubble- and wrinkle-free stacks were then transferred onto SiO$_2$/Si substrates. For the etching process, only a small portion of the graphene edge was exposed for edge-contact deposition (Extended Data Fig. 1). Thus, most of the graphene specimen remained intact. This technique is essential for maintaining graphene microcrystals free of contamination, bubbles and wrinkles. Consequently, the charge neutrality point of graphene in these samples is very close to zero gate voltage at ambient conditions, as confirmed from our near-field gating experiments (Extended Data Fig. 1). After multi-stacking transfer, metal contacts were fabricated at the exposed graphene edge. We note that graphene was entirely encapsulated within the hBN flakes.

Electrical contacts to graphene were provided by a one-dimensional metallic edge of deposited gold, permitting completely polymer-free assembly of the device layers[12]. Together with a silicon back gate fabricated under 285 nm of thermal oxide, these contacts enabled in situ tuning of the graphene carrier density to values as high as 7×10^{12} cm^{-2} for electrons and 5×10^{12} cm^{-2} for holes during the imaging experiment. During our nanoscale infrared imaging experiments, these metallic contacts were also used as effective emitters of graphene plasmons through their efficient in-coupling (scattering) of incident infrared light into polaritons, as reported previously[16] and as detailed below. Our device includes a thin, 7-nm-thick top layer of hBN to facilitate nanoscale infrared imaging of the embedded graphene monolayer owing to the finite penetration depth of electric near-fields and subsurface detection capabilities of nanoscale infrared imaging[8,15]. The hBN is optically transparent at our imaging energy of 886 cm^{-1} (110 meV), ensuring that our nanoscale infrared i plasmons show oscillations with a single p mages strictly document the plasmonic optical response of graphene.

The AFM topography image of the device is shown in Extended Data Fig. 1a. Extended Data Fig. 1b and c shows room-temperature near-field images obtained at gate voltages $V_g = 0$ V and 50 V, respectively. In Extended Data Fig. 1b, the near-field shows practically no contrast resulting from the presence of graphene crystal. This confirms that the encapsulation by hBN kept graphene nearly charge-neutral at ambient conditions. By contrast, Extended Data Fig. 1c clearly shows a rectangular graphene shape due to infrared contrast arising from the overall conductivity enhancement under gating. Moreover, multiple plasmonic fringes along the Au antenna and graphene edges can be detected, which are discussed in the main text and also in 'Plasmons near a graphene edge' below.

Cryogenic infrared nanoscale imaging experiments. The nanoscale infrared imaging experiments were performed using a home-built scattering-type scanning near-field optical microscope (SNOM) operating with variable sample temperatures of 20–450 K. All measurements were conducted under ultrahigh vacuum conditions at a pressure lower than 10^{-9} mbar to prevent sample-surface contamination. Further technical details of the microscope are documented in ref. [11]. The cryogenic scattering-type SNOM is equipped with continuous-wave mid-infrared quantum cascade lasers (DRS Daylight Solutions) and CO$_2$ lasers (Access Laser). The scattering-type SNOM is based on an AFM, which in the present experiments operated in non-contact mode using cantilevered metallic AFM probes with a tip apex radius of ~25 nm and tapping frequencies of ~270 kHz. A pseudo-heterodyne interferometric detection module is implemented to extract both the scattering amplitude s and the phase of the near-field signal; here, we discuss only the former. To suppress background contributions to the back-scattered near-field signal, we demodulated the detected signal at the third harmonic of the probe tapping frequency.

Plasmons near a graphene edge. In the main text, we focused on plasmons launched by the Au pad. In this section, we present temperature-dependent data for plasmons launched by the tip of the scattering-type SNOM near the graphene edges. Extended Data Fig. 2a shows raster-scanned images of the edge regions at a gate voltage of $V_g = 75$ V, taken over a temperature range of $T = 60–300$ K. At $T = 300$ K, plasmons propagate to about 1 μm from the edge of the sample, consistent with the earlier experiments[8,15]. As the temperature is lowered, both the range and the number of detectable plasmonic fringes increase. At $T = 60$ K, the number of plasmonic oscillations almost doubles compared to the room temperature. This behaviour is consistent with what is observed for plasmons launched by the Au pad (Fig. 2).

Whereas images of the Au-launched plasmons show oscillations with a single period, λ_p, the tip-launched plasmons near the edge exhibit two distinct periodicities, λ_p and $\lambda_p/2$, as a function of the tip–edge distance L'. According to previous studies[8,15], the $\lambda_p/2$ period is due to the standing wave produced by plasmons making a round-trip from the tip to the edge and back. In addition to the standing wave, the total electric field in the system contains a component that does not oscillate, but instead decays as a power-law of the distance x from the edge.

This part of the total field, which exists because of the long-range nature of the Coulomb interaction, creates the λ_p-period oscillations as a function of the tip–edge separation L' in the images.

To model the scattering-type SNOM images shown in Extended Data Fig. 2a quantitatively, we used an electromagnetic solver described previously[9,32]. This software uses a boundary-element method to solve the integral equation of the convolution type

$$\hat{\epsilon}\Phi = \Phi_{\text{tip}} \tag{1}$$

for the quasi-static scalar potential $\Phi(x, y)$ induced in graphene in response to the external perturbation, which comprises the incident field plus the potential $\Phi_{\text{tip}}(x, y)$ created by the charges on the tip (the hat denotes an integral operator). The potential Φ_{tip} is not known a priori; it depends on Φ, on the incident field and the tip geometry in a self-consistent manner (see refs [9,32]). In our model, the tip is approximated by a metallic spheroid[9,15] and the incident field is taken to be uniform.

The kernel of equation (1) has the physical meaning of the two-dimensional dielectric function of graphene, which has the well known momentum-space representation

$$\epsilon(q, \omega) = 1 - \frac{|q|}{q_p(\omega)}$$

where

$$q_p(\omega) = q_p'(\omega) + iq_p''(\omega) = \frac{i\kappa\omega}{2\pi\sigma(\omega)} \quad \text{and} \quad \sigma(\omega) = \sigma'(\omega) + i\sigma''(\omega) \tag{2}$$

are the complex plasmon wavenumber and the sheet conductivity of graphene, respectively, and κ is the effective dielectric function of the graphene environment; see 'Plasmonic damping rate analysis'. We note that the plasmon wavelength λ_p and the plasmon quality factor Q are given by

$$\lambda_p = \frac{2\pi}{q_p'} \quad \text{and} \quad Q = \frac{q_p'}{q_p''} \tag{3}$$

Our software code solves equation (1) and the equations relating Φ_{tip} and Φ numerically and subsequently computes the observable scattering-type SNOM signal as a function of the tip–edge distance.

To fit the experimental data with these simulations, we adopted the Drude model of the conductivity

$$\sigma(\omega) = \frac{i}{\pi} \frac{D}{\omega + i\gamma(\omega)} \tag{4}$$

where the Drude weight

$$D = \frac{2e^2}{\hbar^2} T \ln\left[2\cosh\left(\frac{\mu}{2T}\right)\right] \simeq \frac{e^2}{\hbar^2} |\mu| \tag{5}$$

and the scattering rate γ are adjustable parameters. In equation (5), $\hbar$ is the Planck constant and μ is the chemical potential. Our best fits of the computed line profiles to the experimental data are shown in Extended Data Fig. 2b. The physical origin of γ is discussed in Methods section 'Plasmonic damping rate analysis'.

Gate dependence of the plasmon fringes. Extended Data Fig. 3 illustrates the gating-induced carrier-density dependence of nanoscale infrared images at cryogenic temperature. These data are presented in a form of two-dimensional map of the near-field amplitude s as a function of the gate voltage and the distance L from the Au launcher. These data were acquired by repeated scanning along the same linear path for a set of increasing gate voltages. The direction of the scan was normal to the Au launcher. The data shown in Extended Data Fig. 3 encode information about the variation of λ_p and γ with electrostatic doping. For example, the period of the plasmon fringes, which is equal to λ_p, is seen to monotonically increase with V_g. The procedure used for the quantitative determination of λ_p and γ is detailed in 'Simulations of plasmons launched by the Au pad'.

The optical constants of hBN. The optical constants of hBN in infrared frequencies are governed by its optical phonon modes. To the best of our knowledge, there are no previously published data on the hBN phonon parameters at low temperatures. We have obtained this information by measuring the reflectance $R(\omega)$ of exfoliated hBN micro-crystals on SiO$_2$/Si substrates. Because of the limited sample area, these experiments were performed using a diffraction-limited infrared microscope, which imposed a low-frequency cut-off for these far-field reflectance measurements. These experiments were done under high vacuum conditions. Extended Data Fig. 4 shows the optical image of the hBN flake under study.

The flake thickness is ~17 nm, which is close to the thickness of the bottom hBN layer in our hBN/graphene/hBN device structures. Extended Data Fig. 5 shows the $R(\omega)$ value of hBN/SiO$_2$/Si multilayer structures at different temperatures normalized to that of a gold reference. As one can see, $R(\omega)$ is dominated by two resonance peaks, one at ~1,070 cm^{-1} due to SiO$_2$ phonons and the other at ~1,370 cm^{-1} due to hBN in-plane phonons. Extraction of the phonon parameters was done with the help of the RefFIT package[33], which fits $R(\omega)$ spectra using the multi-layer Fresnel reflection formulas, with every layer modelled with Lorentzian oscillators. The response of amorphous SiO$_2$ on a Si substrate is complex and, according to earlier studies, can be described with multiple Lorentzians as

$$\epsilon_2(\omega) = \epsilon_\infty \left(1 + \sum_j \frac{\omega_{pj}^2}{\omega_{tj}^2 - \omega^2 - i\gamma_j\omega} \right) \quad (6)$$

where ω_{tj} is the transverse optical phonon frequency of the phonon species and ϵ_∞ is the high-frequency dielectric constant. The parameters ω_{tj}, ω_{pj} and γ_j for the room-temperature case are given in Extended Data Table 1. The corresponding fits to the reflectance data are shown by the red dashed lines in Extended Data Fig. 5. For the in- and out-of-plane permittivities of hBN we found it sufficient to use the single-Lorentzian model

$$\epsilon_x(\omega) = \epsilon_{\infty x} \left(1 + \frac{\omega_{px}^2}{\omega_{tx}^2 - \omega^2 - i\gamma_x\omega} \right)$$

$$\epsilon_z(\omega) = \epsilon_{\infty z} \left(1 + \frac{\omega_{pz}^2}{\omega_{tz}^2 - \omega^2 - i\gamma_z\omega} \right)$$

Extended Data Fig. 6 shows the obtained linewidth γ_x and frequency ω_{tx} of the in-plane phonon. In particular, the linewidth is found to vary linearly with T, decreasing from ~9 cm^{-1} at room temperature to ~5 cm^{-1} at cryogenic temperatures. This temperature dependence has been included in our theoretical estimation of the dielectric losses (Methods section 'Plasmonic damping rate analysis'). Unfortunately, sample thickness prohibits the evaluation of the out-of-plane permittivity of hBN from these $R(\omega)$ results. Instead, we employed the Fourier-transform infrared nanospectroscopy (nano-FTIR) technique to determine ϵ_z (see Methods section 'Plasmonic damping rate analysis' and Extended Data Table 2).

Simulations of plasmons launched by the Au pad. The electromagnetic solver described in 'Gate dependence of the plasmon fringes' gives an accurate representation of the total near-field signal near the edge of graphene. However, it is more difficult to implement the same approach for a system that contains three conductors, each of a nontrivial shape: a graphene micro-crystal, an elongated tip and a box-like Au pad. Fortunately, this is not necessary. For the purpose of extracting the plasmon wavelength λ_p and the quality factor Q, we only need to consider the long-range plasmonic field launched by the pad. We assume that in the interior of the graphene flake, the tip acts as a passive detector and that the near-field signal that it registers is directly proportional to the normal component E^z of the electric field underneath the tip. The first assumption is reasonable if the pad is a more efficient plasmon launcher than the tip. This is corroborated by the fact that the plasmonic fringes are dominated by simple λ_p-period oscillations instead of 'double peaks', as in Extended Data Fig. 2. The second assumption is not strictly true, but it is very common in the literature[8,9,32]. On the basis of these considerations, we included only the Au pad and graphene in our model.

The simulation was performed using a publicly available software tool[34] that implements the boundary-element method to solve electrodynamic problems involving nanoparticles. In our simulations the pad's dimensions were 1,000 nm × 500 nm × 60 nm. The last two values (the width and thickness) matched the physical ones. However, the first dimension (the length of the pad in the x direction) was shorter than that of the actual device because we found it computationally prohibitive to simulate such a many-micrometre-long object on our computer (a desktop with a memory of 16 GB and a clock speed of 2.7 GHz). We believe that this reduction in the length of the pad in our model is still reasonable because the pad's length controls mostly the antenna effect (that is, the overall magnitude of the field at its leading edge), whereas we want to reproduce the profile of the field distribution—not its absolute magnitude. We also rounded the pad's corners using a curvature radius of 60 nm to obtain a smaller boundary-element method mesh (see Extended Data Fig. 7a) and speed up the computation.

The permittivity of Au in our model was taken to be $\epsilon_{Au} = (1.0 + 1.0i) \times 10^3$ but the results were insensitive to this parameter as long as it was large enough by absolute value. We modelled graphene as an infinite planar slab of thickness $d_g = 0.1$ nm and three-dimensional permittivity $\epsilon_g = -\coth(q_p d_g/2)$, which is related to the two-dimensional dielectric function introduced in Methods section 'Plasmons near a graphene edge' as $\epsilon_g \approx 2\epsilon(q, \omega)/(q d_g)$. This relation ensures

that the slab supports propagating waves of momentum q_p. The results for the quantity of interest, which is the electric field just above the graphene, were practically independent of the arbitrary parameter d_g as long as d_g was small enough. The hBN substrate was not explicitly included in the simulation, but its effect was accounted for implicitly by using the proper value of q_p. The external field was taken to be uniform and linearly polarized in the y direction, Extended Data Fig. 7a. Using this model, we numerically calculated the field component E^z within a 4,000-nm^2 square region just above the graphene layer. An example of such a two-dimensional map is shown in the false-colour image of Extended Data Fig. 7b for the simulation parameters $\lambda_p = 170$ nm and $Q = 130$, which are the ones that provide the best fit to the experimental data obtained at $V_g = 75$ V and $T = 60$ K. The qualitative features are similar to those seen in Fig. 2. Namely, the plasmon oscillations display a generic wave pattern that starts as a plane wave near the Au launcher, then turns into a Fresnel diffraction pattern further away and afterwards gradually transforms into a decaying cylindrical wave modulated by the Fraunhofer diffraction features; see Extended Data Fig. 7b. To avoid complications in the fitting procedure by these plasmonic diffraction effects and to suppress the influence of statistical noise in the experimental data, we averaged multiple line traces from both the simulation and the experimental two-dimensional maps over the strip $x = -300$ nm to $x = 300$ nm. This averaging gave us smooth line traces suitable for fitting. Our best fit for the above values of V_g and T is illustrated in Extended Data Fig. 7c. The red trace is the data and the blue trace is the calculated, strip-averaged E^z. The same fit can be also found in Fig. 2, together with those for the remaining data. To align the phases of the two curves in Extended Data Fig. 7c, we used a phase shift θ. However, this additional adjustable parameter does not affect our main goal of extracting Q, which we achieved by analysing the decay of the absolute value of E^z, that is, the envelope of the oscillations. The fit in Extended Data Fig. 7c is clearly good, as shown in Fig. 2. Hence, in our subsequent analysis we relied on the deduced Q, as described in 'Electron scattering by the acoustic phonons of graphene'.

Plasmonic damping rate analysis. *Losses in the dielectric environment of grapheme.* According to equations (2) and (3), the reciprocal of the plasmonic quality factor Q can be split into two additive components

$$Q^{-1} = \frac{q_p''}{q_p'} \approx \frac{\sigma'}{\sigma''} + \frac{\kappa''}{\kappa'}$$

In this section we discuss the second term, which represents losses in the dielectric environment of graphene. To evaluate such losses that are inherent to our encapsulated structures, we need to find the effective permittivity $\kappa = \kappa'(q, \omega) + \kappa''(q, \omega)$. This requires solving Maxwell's equations for the potential distribution in our multilayer structure (Extended Data Fig. 8a), which is created in response to a periodically modulated charge density of $\sim e^{iq\cdot r}$ introduced in the graphene plane, where r is the two-dimensional coordinate on the graphene plane. In the quasi-static limit of large momenta, $q \gg \omega/c$, the solution of this electrodynamic problem is straightforward, so we present only the final result

$$\kappa(q, \omega) = \sqrt{\epsilon_x}\sqrt{\epsilon_z} \frac{1 - l_1 l_2}{(1 + l_1)(1 + l_2)} \quad (7)$$

where

$$l_j = r_j e^{2ik_z d_j}, \quad k_z = iq\frac{\sqrt{\epsilon_x}}{\sqrt{\epsilon_z}} \quad \text{and} \quad r_j = \frac{\epsilon_z k_z - i\epsilon_j q}{\epsilon_z k_z + i\epsilon_j q} \quad (8)$$

It is instructive to consider two limiting cases. If $d_1, d_2 \to \infty$, then $e^{2ik_z d_j} \to 0$, so the effective permittivity $\kappa_{hBN} = \sqrt{\epsilon_x}\sqrt{\epsilon_z}$ of bulk hBN is obtained, as expected. In the opposite limit, $d_1, d_2 \to 0$, we find $\kappa = (\epsilon_1 + \epsilon_2)/2$, that is, the effective permittivity is equal to the average of the permittivities of the semi-infinite top and bottom layers (vacuum and SiO$_2$), whereas the permittivity of hBN does not enter the calculation. Our device had $d_1 = 7$ nm and $d_2 = 14$ nm (Extended Data Fig. 8a). In this intermediate case, it is imperative to properly account for contributions from both hBN and SiO$_2$ losses.

As mentioned in 'The optical constants of hBN', the temperature dependence of the damping constant γ_z of the hBN c-axis permittivity is not experimentally attainable (where the c axis is the high-symmetry axis perpendicular to the a–b plane). Owing to the thermal occupation factors involved, both γ_x and γ_z are expected to monotonically decrease with temperature. We thus assume that γ_z has the same temperature dependence as γ_x, that is,

$$\gamma_z(T) = \gamma_x(T)\frac{\gamma_z(300 \text{ K})}{\gamma_x(300 \text{ K})} \quad (9)$$

To determine $\gamma_z(300\,\mathrm{K})$, we performed nano-FTIR experiments for an hBN crystal taken away from the sample edges (Extended Data Fig. 8b) from which we obtained $\gamma_z(300\,\mathrm{K}) \approx 3.4\,\mathrm{cm}^{-1}$ can be obtained via numerical modelling, in agreement with previous work[35]. For the plasmon wavelength $\lambda_\mathrm{p} = 170\,\mathrm{nm}$, the final estimate of the environmental dielectric loss

$$\gamma_\mathrm{env}(\omega) \equiv \omega \frac{\kappa''(\omega)}{\kappa'(\omega)}$$

calculated with the help of equations (7)–(9), is plotted in Extended Data Fig. 8c for $\gamma_z(300\,\mathrm{K}) = 3.4\,\mathrm{cm}^{-1}$. Clearly, $\gamma_\mathrm{env}(T)$ is well approximated by a linear function.

Electron scattering by the acoustic phonons of grapheme. The scattering rate γ was defined in equation (4) as one of the two parameters (along with the Drude weight, D, in equation (5)) that define the graphene conductivity $\sigma = \sigma' + \sigma''$. In this section, we discuss theoretical estimates of this scattering rate, which is generally temperature-dependent. In this discussion, the conductivity appears as a primary quantity and γ as a derived one. In the case of interest, $\gamma \ll \omega$, it is given by

$$\gamma(\omega) = \omega \frac{\sigma'}{\sigma''} = \frac{\pi}{D} \omega^2 \sigma'(\omega) \tag{10}$$

In principle, the conductivity and the scattering rate depend on both the frequency, ω, and the momentum, q. The characteristic scale of the momentum dependence is ω/v_F, where v_F is the Fermi velocity. The plasmon momentum, $q_\mathrm{p} \ll \omega/v_\mathrm{F}$, is much smaller. Therefore, to calculate the plasmonic losses it is sufficient to compute the zero-momentum scattering rate $\gamma(\omega) \equiv \gamma(q = 0, \omega)$, which is an easier task. Noting that the power P absorbed per unit area of the system in the presence of an in-plane electric field $Ee^{-i\omega t} + E^*e^{i\omega t}$ is equal to $P = 2\sigma'(\omega)|E|^2$, the desired scattering rate can be expressed as

$$\gamma(\omega) = \frac{\pi\omega^2}{2D} \frac{P}{|E|^2} = \frac{\omega^2}{e^2 v_\mathrm{F}^2} \frac{\partial\mu}{\partial n} \frac{P}{|E|^2} \tag{11}$$

We begin with the acoustic phonon contribution to γ, which we denote γ_A. The velocities of the longitudinal and transverse acoustic phonons in graphene, $v_1 = 2.2 \times 10^6\,\mathrm{cm\,s}^{-1}$ and $v_\mathrm{t} = 1.4 \times 10^6\,\mathrm{cm\,s}^{-1}$, are much smaller than the Fermi velocity, $v_\mathrm{F} \approx 1.0 \times 10^8\,\mathrm{cm\,s}^{-1}$. This implies that the perturbations created by acoustic phonons are quasi-static for electrons. In conventional metals and semiconductors, such perturbations are described in terms of the deformation potential. However, in graphene, it has become common to refer only to a part of the total perturbation as the deformation potential, namely, the part that is proportional to the diagonal component of the strain. The remaining part is referred to as the gauge field or the pseudo-magnetic field[36]. The physical picture that goes with this terminology is that the deformation potential shifts the Dirac cones of graphene in energy, whereas the pseudo-magnetic field moves them in the momentum space, in opposite directions for the two valleys. The issue of the relative importance of deformation potential and pseudo-magnetic field remains unsettled. Although the deformation potential is absent[37] or very small[38] in a naive tight-binding model, it is nevertheless often assumed to be the dominant contribution[18,19,39,40]. However, recent first-principles calculations[20,41] found that the deformation potential is indeed negligible. We assume this to be the case and consider pseudo-magnetic field scattering only. This permits us to neglect electron–electron interactions because the pseudo-magnetic field, unlike the deformation potential, does not perturb the electron density.

To find the phonon-induced correction to the power absorption P, we use the Fermi golden rule (see Extended Data Fig. 9)

$$
\begin{aligned}
P = \ & 2\pi\omega \left|\frac{eA}{c}\right|^2 \int \frac{N\mathrm{d}^2k}{(2\pi)^2} \frac{\mathrm{d}^2q}{(2\pi)^2} \\
& \times \sum_{\zeta,\zeta_\gamma} \sum_{S,S'} \sum_{\nu=l,t} \zeta\zeta_\gamma f(\varepsilon_{Sk})[1 - f(\varepsilon_{S'k'})] \big[n_\mathrm{B}\big(\zeta\hbar\omega_{\nu q}\big) + 1\big] \\
& \times \delta\big(\varepsilon_{S'k'} - \varepsilon_{Sk} + \zeta\hbar\omega_{\nu q} - \zeta_\gamma\hbar\omega\big) \\
& \times \left|\sum_{S''} \left(\frac{W^\nu_{S'k';S''k} v^x_{S''k;Sk}}{\varepsilon_{Sk} - \varepsilon_{S''k} + \zeta_\gamma\hbar\omega} - \frac{v^x_{S'k';S''k} W^\nu_{S''k';Sk}}{\varepsilon_{S'k'} - \varepsilon_{S''k'} + \zeta_\gamma\hbar\omega}\right)\right|^2
\end{aligned}
\tag{12}
$$

Here $A = cE/(i\omega)$ is the in-plane component of the vector potential; c is the speed of light; $N = 4$ is the spin–valley degeneracy; $\zeta = 1$ ($\zeta_\gamma = -1$) represents the emission of a phonon (photon) and $\zeta = -1$ ($\zeta_\gamma = 1$) represents the absorption of a phonon (photon); $\varepsilon_{Sk} = S\hbar v_\mathrm{F}|k|$ is the energy of an electron with momentum k and band index $S = \pm 1$; $\omega_{\nu q} = v_\nu|q|$ is the frequency of the phonon of type ν and momentum q; $f(\varepsilon)$ and $n_\mathrm{B}(\varepsilon)$ are the Fermi and Bose distribution functions, respectively; and $k' = k - q$. The matrix elements $W^\nu_{Sk;S'k'} = \psi^\dagger_{Sk} W^\nu \psi_{S'k'}$ of the

electron–phonon coupling and the electron velocity $v^x_{Sk;S'k} = \psi^\dagger_{Sk} v^x \psi_{S'k}$ are calculated from[41]

$$
\begin{aligned}
W^\nu &= i\beta_\mathrm{A} \sqrt{\frac{\hbar q^2}{\mu_\mathrm{s}\omega_{\nu q}}} \begin{pmatrix} 0 & i^{s_\nu}e^{2i\theta_q} \\ i^{-s_\nu}e^{-2i\theta_q} & 0 \end{pmatrix} \\
\nu^x &= \nu_\mathrm{F} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\
\psi_{Sk} &= \frac{1}{\sqrt{2}} \begin{pmatrix} e^{-i\theta_k} \\ S \end{pmatrix}
\end{aligned}
\tag{13}
$$

where $\beta_\mathrm{A} = 5.0\,\mathrm{eV}$ is the pseudo-magnetic field coupling constant, $\mu_\mathrm{s} = 7.6 \times 10^{-8}\,\mathrm{g\,cm}^{-2}$ is the graphene mass density, s_ν is equal to 0 (1) for $\nu = l$ ($\nu = t$) phonons and θ_q is the polar angle of q, with the x axis being along a zigzag direction.

Numerical evaluation of $\gamma_\mathrm{A}(\omega, T)$ using equations (11)–(13) for $\omega = 886\,\mathrm{cm}^{-1}$ and for the d.c. limit $\omega \to 0$ gives the results shown in Extended Data Fig. 10a by the green and blue lines, respectively. These lines are almost straight and nearly coincide with one another in the range of our experiments, $T \geq 60\,\mathrm{K}$. For the d.c. scattering rate $\gamma_\mathrm{A}(\omega = 0, T)$, such a linear behaviour with respect to T is expected[18–20,37–39,41]. γ_A can be derived analytically by assuming $|\mu| \gg \omega, T$ and $T \gg T_\mathrm{BG}$, where $T_\mathrm{BG} = 2|\mu|v_l/v_\mathrm{F}$ is the Bloch–Grüneisen temperature. This set of inequalities permits one to neglect $\omega_{\nu q}$ inside the δ-function and to replace $f(\varepsilon)$ by a step-function and $n_\mathrm{B}(\varepsilon)$ by T/ε. After some algebra, we find

$$
\begin{aligned}
\gamma_\mathrm{A}(0) &= \frac{1}{\hbar^3} \frac{\beta_\mathrm{A}^2}{\mu_\mathrm{s}v_\mathrm{F}^2} |\mu| \left(\frac{1}{v_1^2} + \frac{1}{v_\mathrm{t}^2}\right) T \\
\gamma_\mathrm{A}(\omega) &\approx \left(1 + \frac{1}{12} \frac{\omega^2}{\mu^2}\right) \gamma_\mathrm{A}(0)
\end{aligned}
\tag{14}
$$

The first equation in equation (14) is in agreement with prior work[20,38–41], as is also the common observation that the linear temperature dependence sets in already at $T \approx 0.2 T_\mathrm{BG} \approx (10\,\mathrm{K})\sqrt{n[10^{12}\,\mathrm{cm}^{-2}]} \approx 25\,\mathrm{K}$; see Extended Data Fig. 10. The second equation in equation (14) is our new result. Importantly, it explains why the infrared scattering rate $\gamma_\mathrm{A}(\omega, T)$ at the frequency of interest, $\omega = 886\,\mathrm{cm}^{-1} \approx 0.4|\mu|$, is hardly different from the d.c. rate $\gamma_\mathrm{A}(0, T)$ in Extended Data Fig. 10a.

The weak frequency dependence given by our formula is inconsistent with the rapid increase of $\gamma_\mathrm{A}(\omega, T)$ with ω obtained in a previous theoretical work[18] that studied electron–phonon scattering by the deformation potential mechanism. By examining the formulas given in ref. [18], we traced the difference between the deformation potential and pseudo-magnetic field cases to a strongly ω-dependent electron screening. This screening influences the deformation potential in electron–phonon coupling but plays no role in the pseudo-magnetic field channel we consider here.

At temperatures $T \lesssim 0.2 T_\mathrm{BG}$ the assumptions used in deriving equation (14) do not hold. Such temperatures have not been reached in our experiments, but we can discuss the predictions of our theory. In this ultimate low-temperature regime, the d.c. and alternating current (a.c.) scattering rates would no longer have the same scaling with temperature. The d.c. rate (Extended Data Fig. 10a, blue line) should vanish, whereas the a.c. one (Extended Data Fig. 10a, green line) should approach a non-zero value as $T \to 0$. The physical reason for this difference is that the scattering rate $\gamma_\mathrm{A}(\omega, 0)$ at absolute zero is solely due to emission of acoustic phonons, which is possible only at a finite ω.

The red and black solid lines in Extended Data Fig. 10a represent the electron–phonon inter-valley scattering rate. This process, as well as the total electron–phonon scattering rates plotted in Extended Data Fig. 10b and c, are discussed in the next section.

Inter-valley electron-phonon scattering. Inter-valley electron–phonon scattering is dominated by transverse optical phonons at the K-point of the Brillouin zone, the so-called A_1' phonons[42]. The matrix element involved in this scattering is given by[41]

$$W^\mathrm{K}_{SkS'k'} = i\beta_\mathrm{K} \sqrt{\frac{2\hbar}{\mu_\mathrm{s}\omega_\mathrm{K}}} \begin{pmatrix} e^{i\theta_k} \\ s \end{pmatrix}^\dagger \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} \begin{pmatrix} e^{-i\theta_{k'}} \\ s' \end{pmatrix} \tag{15}$$

where β_K is the electron–A_1'-phonon coupling constant. According to previous calculations[20,42] of the corresponding d.c. scattering rate $\gamma_\mathrm{K}(0)$, A_1'-phonon scattering is activated at $T \approx 0.15\hbar\omega_\mathrm{K} \approx 250\,\mathrm{K}$, and at room temperature it is responsible for a large fraction of the total $\gamma(0) \approx \gamma_\mathrm{A}(0) + \gamma_\mathrm{K}(0)$. Here we show that inter-valley scattering has an even larger contribution $\gamma_\mathrm{K}(\omega)$ to the infrared scattering rate and that it begins at a temperature as low as 100 K. This scattering mechanism can therefore naturally explain the observed super-linear temperature increase of the plasmon damping seen in Extended Data Fig. 10.

Within a naive tight-binding model[20,37], one finds $\beta_K = 4\beta_A/b = 14\,\text{eV Å}^{-1}$, where $b = 1.42\,\text{Å}$ is the carbon–carbon bond length. However, virtual electron–phonon scattering, such as electron screening and vertex corrections, are important for the inter-valley processes. These many-body effects cause a red shift of the A_1'-phonon frequency ω_K (Kohn anomaly[42]) and an enhancement[41–43] of the electron–phonon coupling constant β_K. Both effects are difficult to compute accurately, and so we treat ω_K and β_K as adjustable parameters.

To calculate γ_K we use again the Fermi golden rule, equation (12), this time with the matrix element from equation (15). For high doping, $|\mu| \gg \omega$, we can keep only the intraband terms $S'' = S' = S$, which are the dominant ones. After some algebra, we arrive at

$$\gamma_K(\omega) = \frac{3}{2}\frac{\beta_K^2|\mu|}{\mu_s\omega_K\omega\hbar^3 v_F^2}\sum_{\zeta,\zeta_\gamma}\zeta\zeta_\gamma\int_0^\infty x(|\mu|x - \zeta\hbar\omega_K + \zeta_\gamma\hbar\omega)$$
$$\times f(\mu x)[1 - f(|\mu|x - \zeta\hbar\omega_K + \zeta_\gamma\hbar\omega)][1 + n_B(\zeta\hbar\omega_K)]dx$$

This integral can be readily evaluated numerically[34]. At low temperatures the growth of $\gamma_K(\omega) \approx \exp[(\omega_K - \omega)/T]$ is roughly exponential because it is dictated by the thermal tails of the occupation factors of the phonons and electrons. This consideration makes it clear why the infrared scattering rate has an earlier onset in T than the d.c. one. In Extended Data Fig. 10c we show the results of our calculations for the total rate $\gamma = \gamma_A + \gamma_K$. The circles (squares) represent the plasmon damping rates extracted from the experiment, with the environmental loss rate chosen as in Extended Data Fig. 8c. For both d.c. and infrared cases, shown by the red and blue curves, respectively, good agreement between the theory and the experimental data is achieved for $\beta_K = 14\,\text{eV Å}^{-1}$ and $\omega_K = 1{,}200\,\text{cm}^{-1}$. Note that these values are basically the same as the previous first-principles calculations results[20,41,42], where $\beta_K = 13.9\,\text{eV Å}^{-1}$ and $\omega_K = 1{,}209\,\text{cm}^{-1}$. Other parameters that we used were $\mu = 0.27\,\text{eV}$ and $\beta_A = 5.0\,\text{eV}$.

Electron–electron scattering. In this section we estimate the plasmon damping due to electron–electron interactions using an original method. For previous work on this subject see, for example, ref. [43]. In general, electron interactions can influence the conductivity only if the Galilean invariance of the system is broken[44], which is a weak effect in doped graphene at low temperatures.

We begin by noting that the electron–electron collision rate[44]

$$\Gamma_{ee} = \frac{\pi}{4}\frac{T^2}{|\mu|}\ln\left(\frac{2|\mu|}{T}\right) \tag{16}$$

is comparable to the phonon-induced scattering rate; see Extended Data Figs. 10 and 11. However, such collisions conserve the total momentum and thus approximately conserve the total current as well. In fact, there exists a nonequilibrium, current-carrying state that is immune to electron–electron collisions, namely, a Fermi sea with a Doppler-shifted energy spectrum $\varepsilon_{Sk} \rightarrow \varepsilon_{Sk} - \hbar u^x k^x$. This state corresponds to a uniform hydrodynamic flow with velocity u^x. In comparison, the Fermi distribution of noninteracting electrons subject to an electric field E is typically shifted in momentum as $\hbar k^x \rightarrow \hbar k^x - \Delta p^x$ with $\Delta p^x = eE\gamma^{-1}$. At low temperatures, this latter state is similar, but not identical, to a hydrodynamic flow with velocity $u^x = \Delta p^x/m$, where $m = \hbar k_F/v_F$ is the effective mass and k_F is the Fermi momentum. We can call these two types of perturbed distribution functions the velocity mode and the momentum mode, respectively. The Galilean invariance is indeed broken because m depends on the electron energy. However, this dependence is weak; therefore, we can anticipate that the damping rate entering the conductivity is much smaller than the electron collision rate Γ_{ee}.

To proceed, we suppose that both momentum-conserving and nonconserving processes are present, with rates Γ_{ee} and Γ_d, respectively. As shown in recent work[22], the form of the conductivity can be obtained by assuming that the perturbation of the electron distribution function is a linear combination of the velocity and momentum modes. This ansatz leads to the two-term Drude formula

$$\sigma(\omega) = \frac{i}{\pi}\frac{D_h}{\omega + i\Gamma_d} + \frac{i}{\pi}\frac{D - D_h}{\omega + i\Gamma_d + i\Gamma_{ee}}$$

(see also ref. [45]) in lieu of the conventional single-term one, equation (4). We can still define the effective electron–electron scattering rate similarly to equation (10) to obtain

$$\gamma_{ee} \equiv \frac{\pi}{D}\omega^2\sigma'(\omega) - \Gamma_d \approx \frac{D - D_h}{D}\Gamma_{ee}, \quad \omega \gg \Gamma_{ee}, \Gamma_d \tag{17}$$
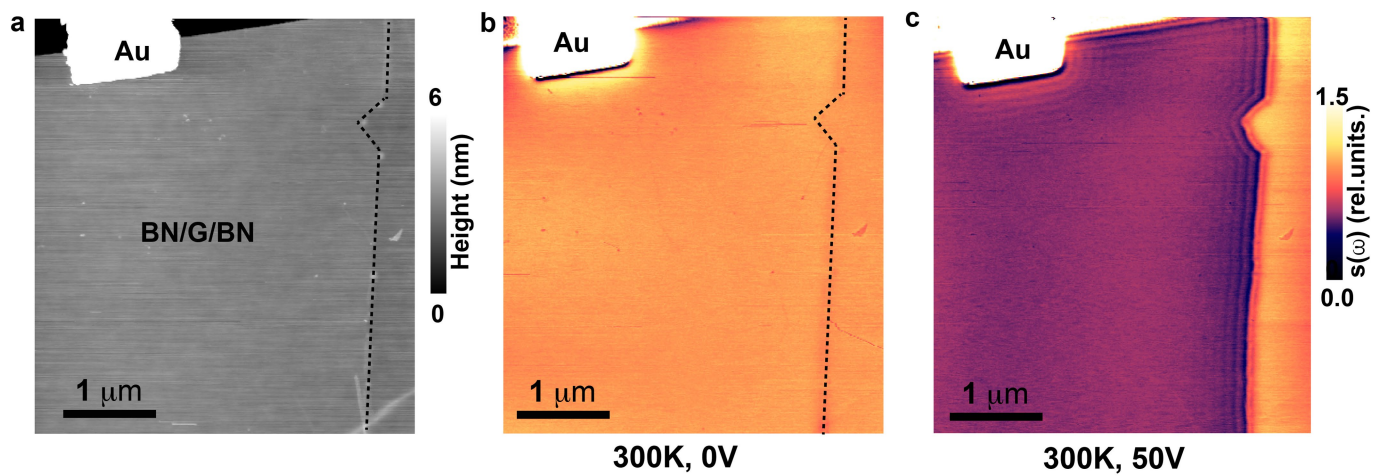
Here $D = \int_{-\infty}^\infty \sigma'(\omega)d\omega$ is the total optical weight given by equation (4) and $D_h = \pi e^2 n/m_h$ is the hydrodynamic Drude weight. For weak interactions, the corresponding hydrodynamic mass m_h is given by[22,46]

$$m_h = \frac{3}{2}\frac{1}{nv_F^2}\sum_{S=\pm 1}\left\{\int_{-\infty}^\infty \frac{Nd^2k}{(2\pi)^2}[f(\varepsilon_{Sk}) - \Theta(-S)]\varepsilon_{Sk}\right\}$$
$$= \frac{3}{\pi}\frac{T^3}{\hbar^2 v_F^4 n}\left[\frac{\pi^2}{3}\frac{\mu}{T} + \frac{1}{3}\frac{\mu^3}{T^3} - 4\,\text{Li}_3(-e^{\mu/T})\right] \tag{18}$$

where Θ is the unit step function and $\text{Li}_n(z)$ is the polylogarithm function of order n and argument z. The crucial point is that at the temperatures of interest, $T \ll |\mu|$, the two Drude weights, D and D_h, are nearly equal. As a result, the plasmon damping rate γ_{ee} due to electron interactions is indeed much smaller than the electron collision rate Γ_{ee}; see equation (17) and Extended Data Fig. 11.

Data availability. The data that support the findings of this study are available from the corresponding author on reasonable request.

32. Fei, Z. et al. Infrared nanoscopy of Dirac plasmons at the graphene–SiO$_2$ interface. *Nano Lett.* **11**, 4701–4705 (2011).
33. Kuzmenko, A. B. RefFIT: software to fit optical spectra, v1.2.95.
34. Waxenegger, J., Trügler, A. & Hohenester, U. Plasmonics simulations with the MNPBEM toolbox: consideration of substrates and layer structures. *Comput. Phys. Commun.* **193**, 138–150 (2015).
35. Dai, S. et al. Tunable phonon polaritons in atomically thin van der Waals crystals of boron nitride. *Science* **343**, 1125–1129 (2014).
36. Amorim, B. et al. Novel effects of strains in graphene and other 2D materials. *Phys. Rep.* **617**, 1–54 (2016).
37. Pietronero, L., Strässler, S., Zeller, H. R. & Rice, M. J. Electrical conductivity of a graphite layer. *Phys. Rev. B* **22**, 904–910 (1980).
38. Woods, L. M. & Mahan, G. D. Electron–phonon effects in graphene and armchair (10,10) single-wall carbon nanotubes. *Phys. Rev. B* **61**, 10651 (2000).
39. von Oppen, F., Guinea, F. & Mariani, E. Synthetic electric fields and phonon damping in carbon nanotubes and graphene. *Phys. Rev. B* **80**, 075420 (2009).
40. Mariani, E. & von Oppen, F. Temperature-dependent resistivity of suspended graphene. *Phys. Rev. B* **82**, 195403 (2010).
41. Sohier, T. et al. Phonon-limited resistivity of graphene by first-principles calculations: electron–phonon interactions, strain-induced gauge field, and Boltzmann equation. *Phys. Rev. B* **90**, 125414 (2014).
42. Piscanec, S., Lazzeri, M., Mauri, F., Ferrari, A. C. & Robertson, J. Kohn anomalies and electron–phonon interactions in graphite. *Phys. Rev. Lett.* **93**, 185503 (2004).
43. Basko, D. M. & Aleiner, I. L. Interplay of Coulomb and electron–phonon interactions in graphene. *Phys. Rev. B* **77**, 041409 (2008).
44. Giuliani, G. & Vignale, G. *Quantum Theory of the Electron Liquid* (Cambridge Univ. Press, Cambridge, 2005).
45. Maslov, D. L. & Chubukov, A. V. Optical response of correlated electron systems. *Rep. Prog. Phys.* **80**, 026503 (2017).
46. Sun, Z., Basov, D. N. & Fogler, M. M. Adiabatic amplification of plasmons and demons in 2D systems. *Phys. Rev. Lett.* **117**, 076805 (2016).

a Au

BN/G/BN

Height (nm)

6

0

1 μm

b Au

1 μm

300K, 0V

c Au

s(ω) (rel.units.)

1.5

0.0

1 μm

300K, 50V

Extended Data Fig. 1 | AFM topography and $s(\omega)$ image of the hBN/ graphene/hBN encapsulated device. a, AFM topography image. The black dashed line marks the physical edge of graphene. **b, c**, Room- temperature $s(\omega)$ images obtained at $V_g = 0$ V and $V_g = 50$ V, respectively. G, graphene.

a

b
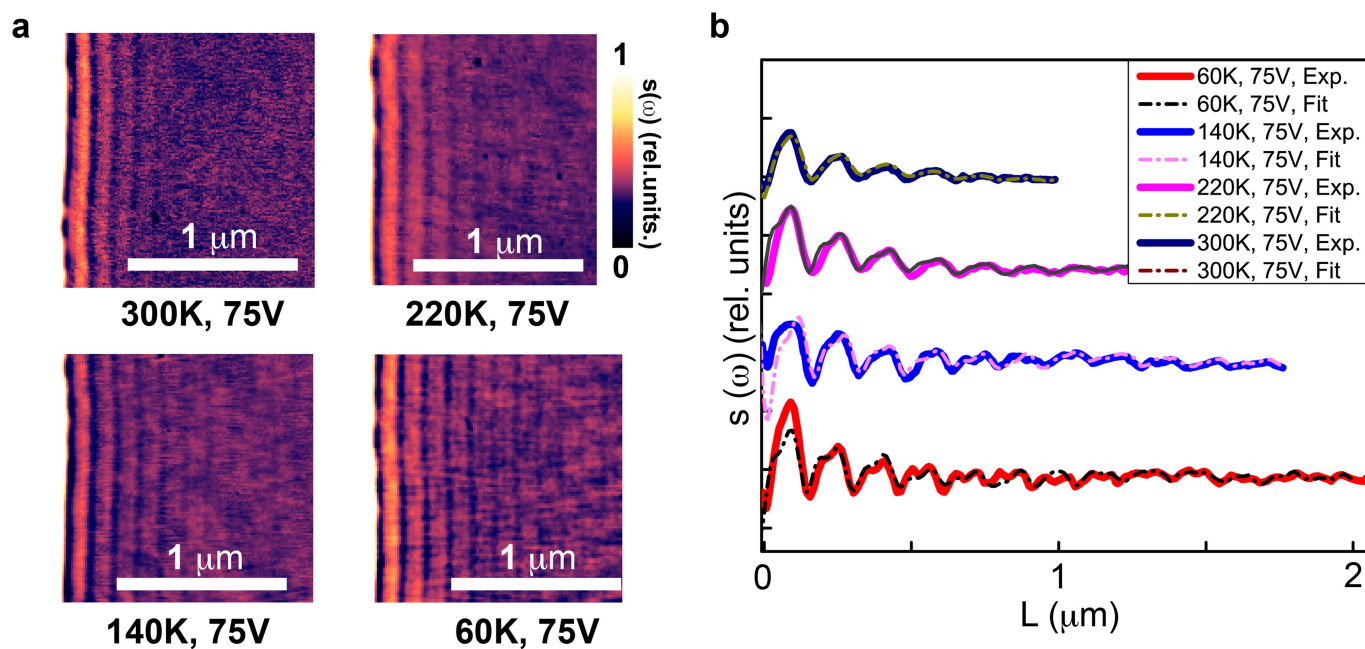
Extended Data Fig. 2 | Plasmonic fringes near a graphene edge at different temperatures. a, Maps of the near-field amplitude s measured at several temperatures for fixed $V_g = 75$ V and $\omega = 886$ cm^{-1}. **b**, Line profiles obtained by averaging over the vertical coordinates in each map. The solid lines are the data and the dash-dotted lines are the simulations.

a

b

c
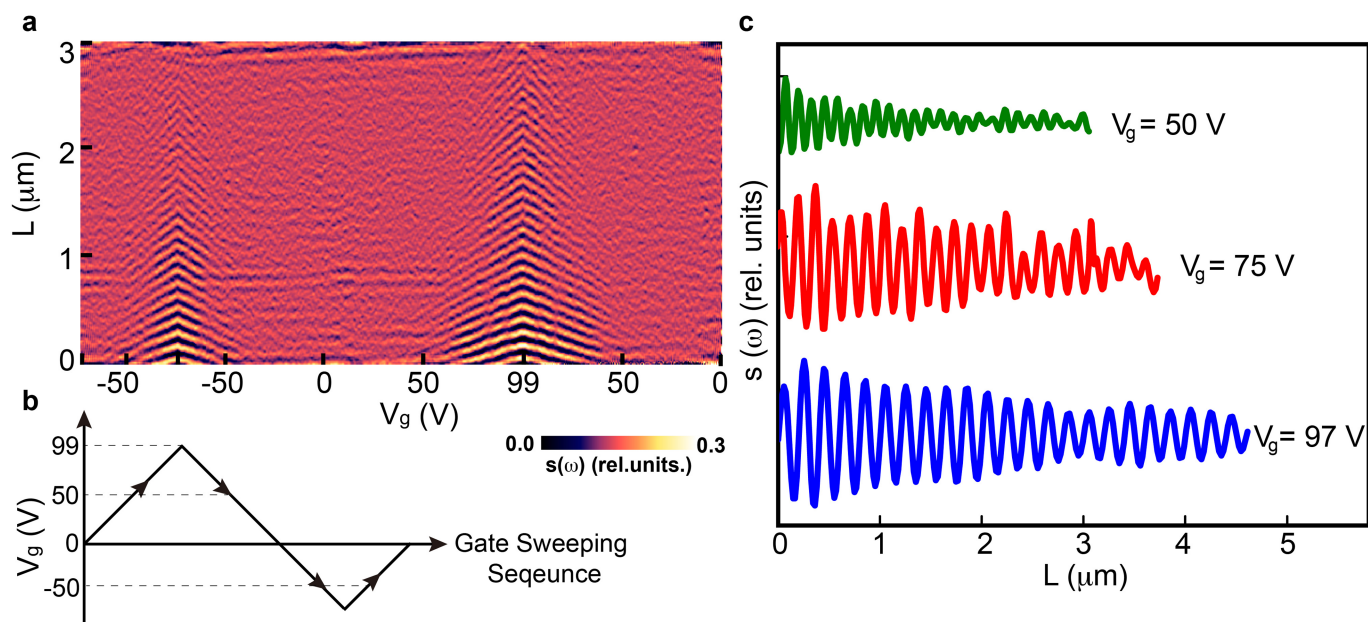
Extended Data Fig. 3 | Gate dependence of plasmon propagation at cryogenic temperature. a, Near-field amplitude s as a function of the gate voltage and the distance L from the Au launcher at a fixed frequency of $\omega = 886\,cm^{-1}$. **b**, Illustration of the gate sweeping sequence. **c**, Line profiles (averaged over ~200 nm perpendicular to the propagation direction) of plasmonic interference fringes at different gate voltages.
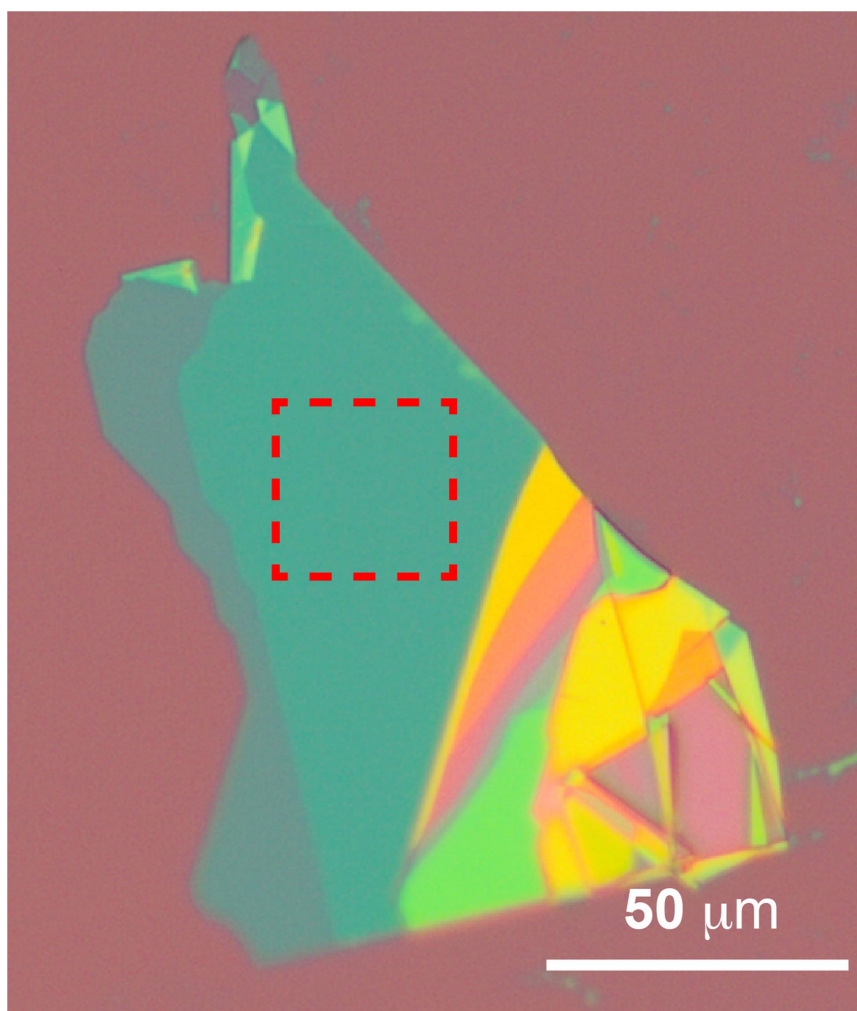
Extended Data Fig. 4 | Optical image of an hBN sample on a SiO$_2$/Si substrate. The dashed red square marks the approximate location of the region used in the reflectance measurements. The hBN thickness is 17 nm.
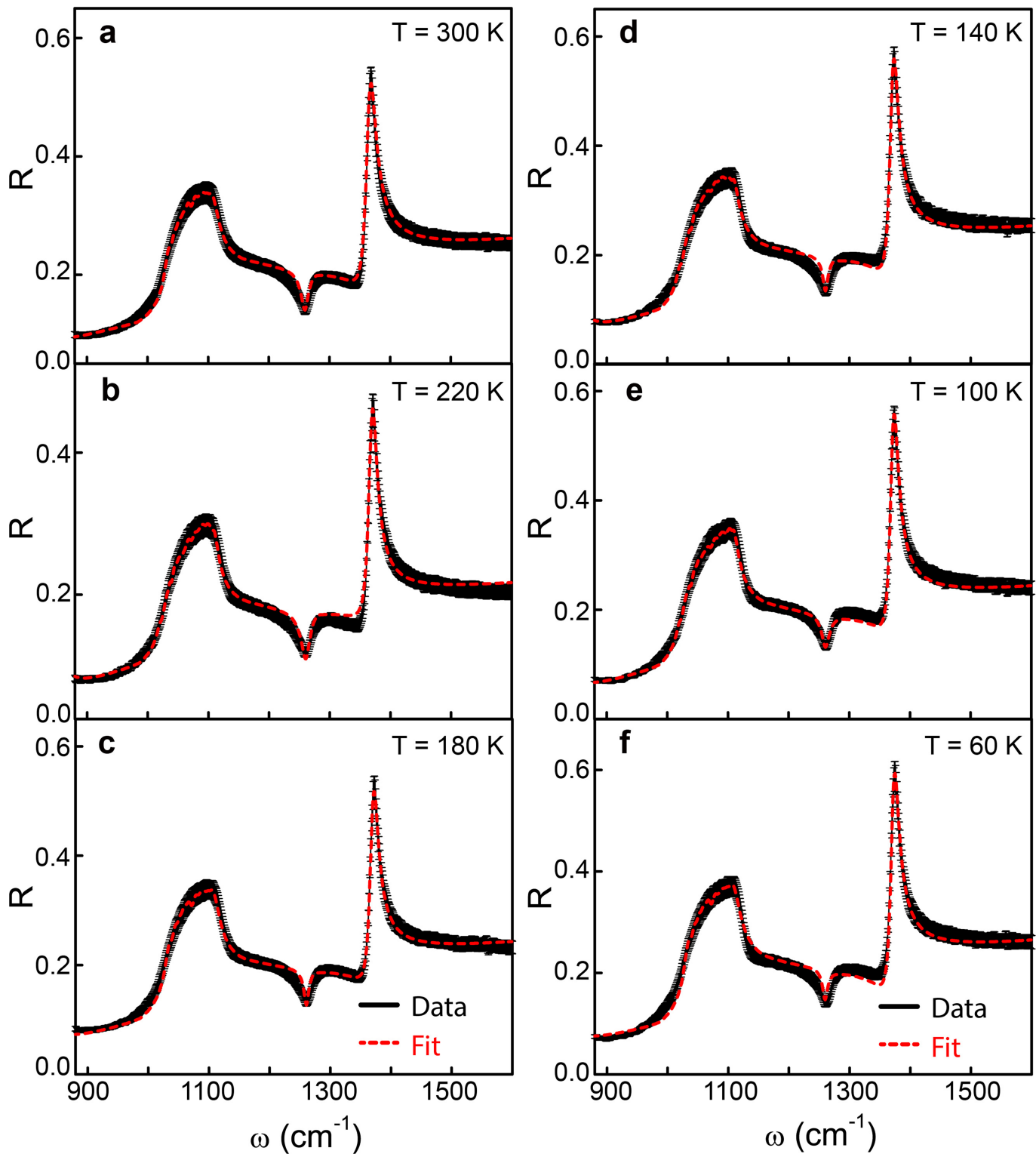
Extended Data Fig. 5 | Reflectance spectra of the hBN/SiO₂/Si structure at _T_=60–300 K. a–f, The black points are the experimental data and the red dashed lines are the fits. The sharp resonance at 1,370 cm⁻¹ is due to the in-plane optical phonon of hBN, and the broad peak at around 1,070 cm⁻¹ is due to the optical phonon of SiO₂.

a

b
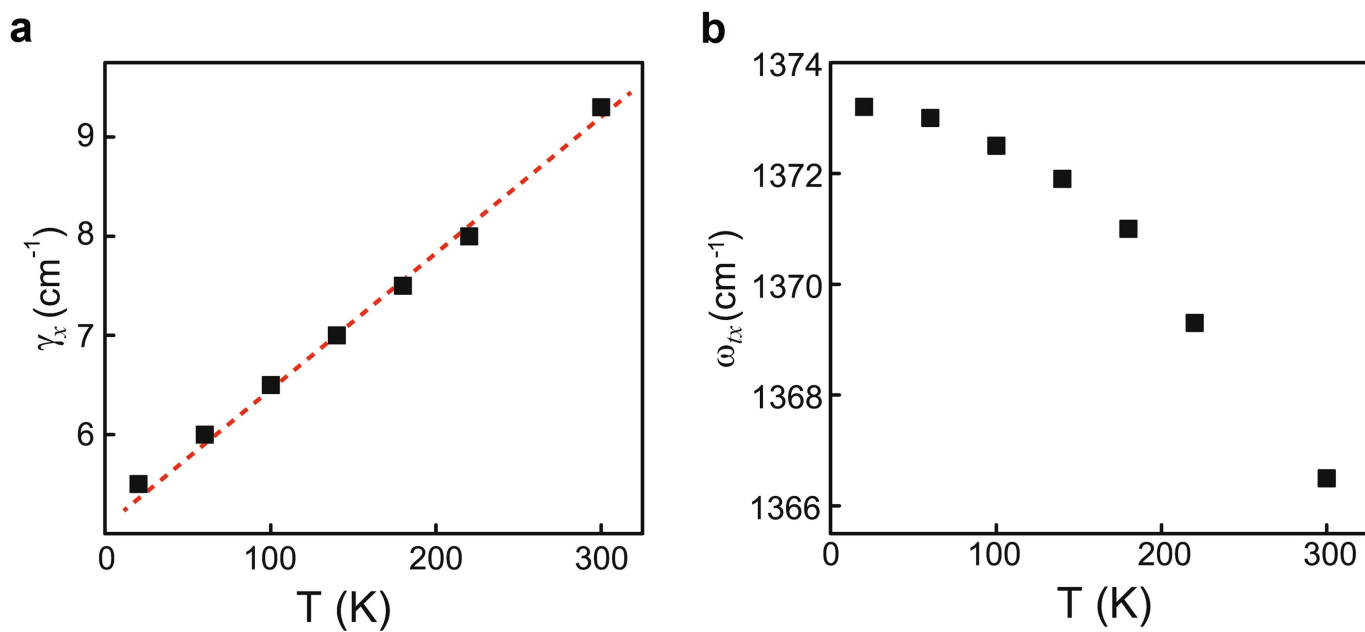
Extended Data Fig. 6 | The hBN in-plane phonon parameters versus the temperature. a, The fitted phonon linewidth γ_x. The squares are the data and the dashed line is a guide for the eye. **b**, The fitted frequency ω_{tx}.

a

b

c
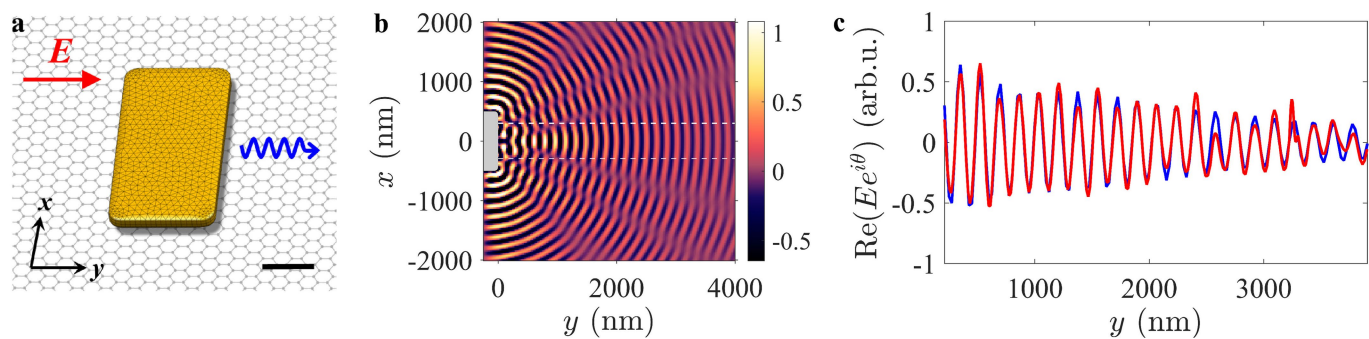
Extended Data Fig. 7 | Plasmonic oscillations in the vicinity of the Au launching pads. a, Illustration of the Au pad (gold, with the triangular mesh used in the simulation) on graphene (honeycomb lattice). The red arrow depicts the direction of the external field. The blue arrow symbolizes launched plasmons. Scale bar, 200 nm. **b**, An example of the simulated electric field distribution (E^z, imaginary part) just above the graphene layer. The grey rectangle represents the right half of the Au pad.

The simulation parameters are $\lambda_p = 170$ nm and $Q = 130$. **c**, Comparison of the theoretical fit (blue) and the experimental data (red) for $V_g = 75$ V and $T = 60$ K. Both the theoretical and experimental traces are obtained by averaging multiple line cuts inside the 600-nm-wide strip indicated by the white dashed lines in **b**. The phase shift $\theta = 112°$ (see text) was used to align the oscillations.

a

Vacuum ϵ_1

hBN ϵ_x, ϵ_z G $d_1 = 7$ nm

hBN ϵ_x, ϵ_z $d_2 = 14$ nm

SiO$_2$ ϵ_2
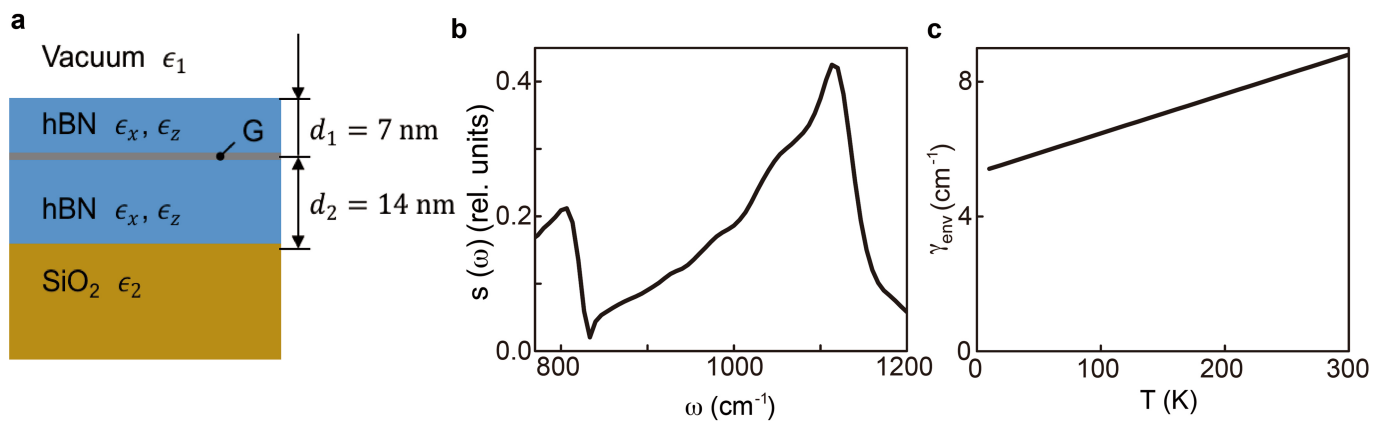
Extended Data Fig. 8 | Device structure and dielectric losses.
a, Schematic of the device, showing the notations for the permittivities and thicknesses of the layers. **b**, Nano-FTIR spectrum $s(\omega)$ obtained with the hBN crystal taken away from the sample edges. **c**, The contribution γ_{env} of the dielectric environment to the plasmon linewidth as a function of temperature at a frequency of $\omega = 886$ cm^{-1}. The hBN c-axis damping constant is $\gamma_z(300\,\mathrm{K}) = 3.4$ cm^{-1}, which is consistent with previous results[35].
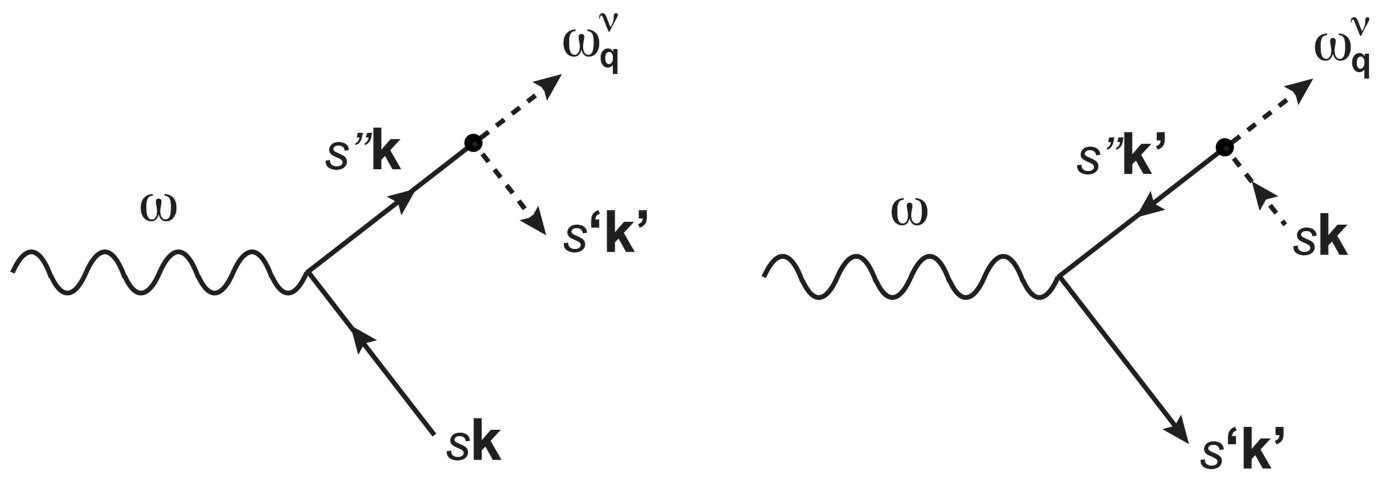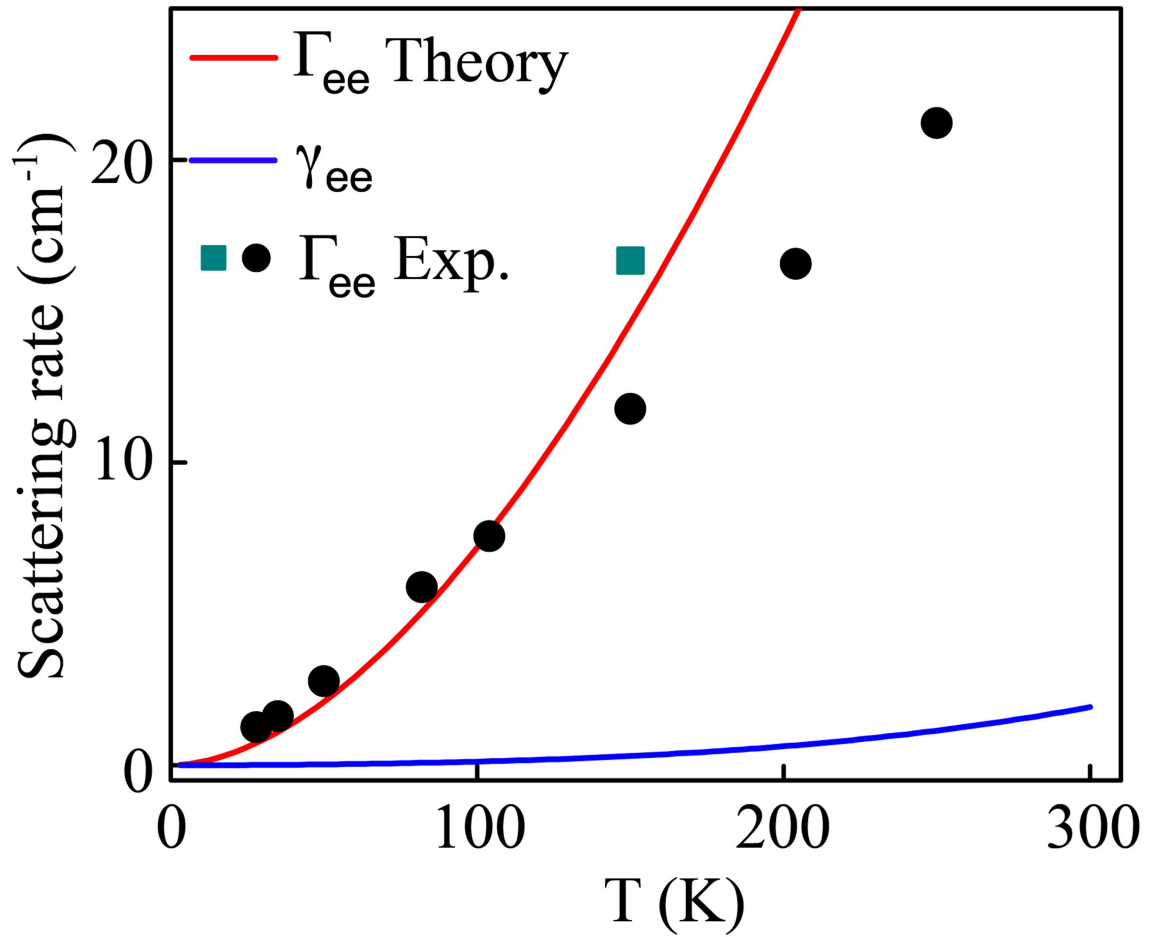
Extended Data Fig. 9 | Electron–phonon scattering processes. Diagrams of the scattering processes included in equation (12). The wavy, straight and dashed lines represent photons, electrons and phonons, respectively.
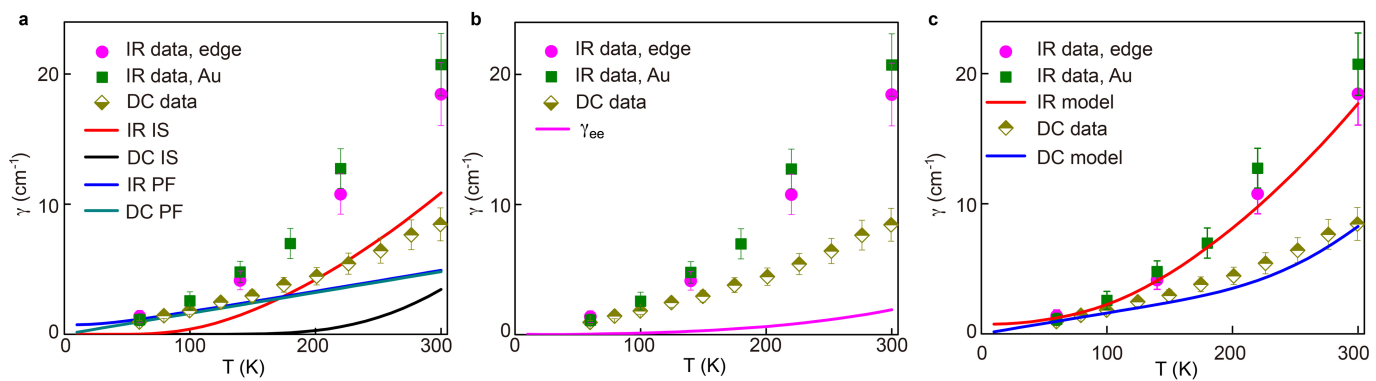
Extended Data Fig. 10 | Electron–phonon scattering rate as a function of temperature. a–c, Temperature dependence of the plasmonic scattering rate and of the d.c. scattering rate. Solid lines in **a** and **b** display the results of parameter-free modelling for electron–phonon scattering contributions (**a**) and electron–electron scattering contributions (**b**). Solid lines in **c** display the results of the sum in **a** and **b**, as discussed in the main text.

Extended Data Fig. 11 | Plasmon damping rate due to electron–electron scattering as a function of temperature. The blue solid curve is the plasmon damping rate γ_{ee} due to electron–electron interactions, computed from equations (17) and (18) for a Fermi energy of $\varepsilon_F \equiv \mu(T=0) = 0.27$ eV. The red solid curve is the electron collision rate Γ_{ee} from equation (16). The squares and circles are Γ_{ee} values extracted from recent d.c. transport studies[23,24] at a different carrier density, $n \approx 10^{12}$ cm^{-2}.

Extended Data Table 1 | Phonon oscillator parameters for SiO$_2$

ω_{tj} (cm^{-1})	ω_{pj} (cm^{-1})	γ_j (cm^{-1})
1108.5	179.8	15.1
1104	219.7	12.6
1098	203.4	12.76
1092	330.8	19.6
1086	392.4	19
1077	355	13.76
1067	344	13.2
1059	386.9	16
1052	221	9.7
1045	302.8	13.2
1035	237.7	13
1030	37.68	16.9

The fitting parameters of equation (6) for $T = 300$ K.

Extended Data Table 2 | Phonon oscillator parameters for hBN

Axis, a	$\varepsilon_{\infty a}$	ω_{ta} (cm^{-1})	ω_{pa} (cm^{-1})	γ_a (cm^{-1})
t	4.90	Extended Data Fig. 6b	869	Extended Data Fig. 6a
z	2.95	760	346	Extended Data Fig. 8b

The single-Lorentzian fitting parameters of equation (6) for $T = 300$ K. See also Extended Data Figs. 6, 8.