

In the format provided by the authors and unedited.

Out-of-plane heat transfer in van der Waals stacks through electron-hyperbolic phonon coupling

Klaas-Jan Tielrooij ^{1*}, Niels C. H. Hesp ¹, Alessandro Principi^{2,3}, Mark B. Lundeberg¹, Eva A. A. Pogna⁴, Luca Banszerus⁵, Zoltán Mics⁶, Mathieu Massicotte¹, Peter Schmidt ¹, Diana Davydovskaya¹, David G. Purdie⁷, Ilya Goykhman⁷, Giancarlo Soavi⁷, Antonio Lombardo⁸, Kenji Watanabe ⁸, Takashi Taniguchi⁸, Mischa Bonn⁶, Dmitry Turchinovich ^{6,11}, Christoph Stampfer⁵, Andrea C. Ferrari⁷, Giulio Cerullo⁴, Marco Polini⁹ and Frank H. L. Koppens ^{1,10*}

¹ICFO – Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, Castelldefels (Barcelona), Spain. ²Radboud University, Institute for Molecules and Materials, Nijmegen, The Netherlands. ³School of Physics & Astronomy, University of Manchester, Manchester, UK. ⁴IFN-CNR, Dipartimento di Fisica, Politecnico di Milano, Milano, Italy. ⁵JARA-FIT and 2nd Institute of Physics, RWTH Aachen University, Aachen, Germany. ⁶Max Planck Institute for Polymer Research, Mainz, Germany. ⁷Cambridge Graphene Centre, University of Cambridge, Cambridge, UK. ⁸National Institute for Material Science, Tsukuba, Japan. ⁹Istituto Italiano di Tecnologia, Graphene Labs, Genova, Italy. ¹⁰ICREA – Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain. Present address: ¹¹Fakultät für Physik, Universität Duisburg-Essen, Duisburg, Germany. Klaas-Jan Tielrooij and Niels Hesp are equally contributing authors. *e-mail: klaas-jan.tielrooij@icfo.eu; frank.koppens@icfo.eu

Out-of-plane heat transfer in van der Waals stacks through electron-hyperbolic phonon coupling

Klaas-Jan Tielrooij*, Niels C. H. Hesp, Alessandro Principi, Mark B. Lundeberg, Eva A. A. Pogna, Luca Banszerus, Zoltán Mics, Mathieu Massicotte, Peter Schmidt, Diana Davydovskaya, David G. Purdie, Ilya Goykhman, Giancarlo Soavi, Antonio Lombardo, Kenji Watanabe, Takashi Taniguchi, Mischa Bonn, Dmitry Turchinovich, Christoph Stampfer, Andrea C. Ferrari, Giulio Cerullo, Marco Polini, Frank H. L. Koppens*

* Correspondence to: klaas-jan.tielrooij@icfo.eu, frank.koppens@icfo.eu

This PDF file includes:

Supplementary Section 1-2

Supplementary Figures 1-10

Supplementary Section 1: Cooling channels for hot carriers in graphene

Here we assess cooling channels for a hot-carrier distribution in graphene that compete with cooling through near-field coupling to hyperbolic hBN phonon polaritons. We mainly discuss supercollision and normal collision scattering to graphene acoustic phonons and optical graphene phonon cooling, while also briefly discussing flexural phonons, Wiedemann-Franz cooling and hot-carrier tunneling.

Cooling to graphene acoustic phonons

Electron-acoustic phonon deformation potential — Since coupling to graphene acoustic phonons (either through disorder-assisted supercollisions or through normal collisions) depends heavily on the electron-phonon deformation potential D , we first discuss its value. The value of D is reasonably well established, with transport measurements on ultra-clean, hBN-encapsulated devices giving ~ 18 – 20 eV, assuming phonon-limited momentum scattering[S1, S2]. The cooling dynamics of SiO₂-supported graphene are consistent with $D = 12$ – 18 eV, assuming disorder-assisted cooling[S3]. We use transport measurements to determine the deformation potential for our device. Momentum-non-conserving collisions can occur because of: *i*) long-range scattering mechanisms, *ii*) short-range scattering mechanisms, and *iii*) electron-phonon scattering[S4]. If momentum scattering is solely determined by electron-phonon interaction (process *iii*), we can use the measured graphene mobility μ at a given carrier density n to obtain the deformation potential following Refs. [S5, S6]: $\mu = \frac{4\hbar v_F^2 e \rho v_s^2}{\pi D^2 n k_B T}$, where v_F is the Fermi velocity, ρ the mass density, v_s the sound velocity and \hbar , e and k_B the reduced Planck constant, electron charge and Boltzmann constant, respectively. Inserting relevant numbers gives $D = 35$ eV. However, since the mean free path for low carrier concentrations scales linearly with n (see Fig. 2a), we know that long-range scattering (process *i*) also plays a role [S4] and therefore the value for D is an *upper* limit. Thus, we find that our transport data are consistent with $D < 35$ eV.

Supercollision cooling — The supercollision cooling mechanism relies on the presence of short-range scatterers (long-range scattering mechanisms give infinite cooling time) [S7] and gives a cooling time that scales with $\frac{k_F \ell}{D^2}$, where k_F is the Fermi momentum, ℓ the mean free path (limited by short-range scattering), and D the electron-phonon deformation potential

[S3]. We now proceed to calculate the deformation potential that would be necessary to reproduce the observed photovoltage dip, if supercollision scattering would dominate, where we closely follow Ref. [S3] in calculating the experimental photovoltage dip ΔV_{PTE} . Since this model requires that $T_{\text{el}} \ll T_{\text{F}}$, we apply this analysis to our data at $n = 1.7 \times 10^{12}/\text{cm}^2$, where $T_{\text{F}} = \frac{E_{\text{F}}}{k_{\text{B}}} \approx 1800$ K. We numerically solve the energy dissipation rate $C_n \frac{dT_{\text{e}}}{dt} = -A(T_{\text{e}}^3 - T_{\text{L}}^3)$ with the heat capacity (for non-neutral graphene) $C_n = \alpha T_{\text{el}}$ and $\frac{A}{\alpha} = 0.47 \frac{1}{k_{\text{F}} \ell} \frac{D^2}{\rho v_{\text{s}}^2} \frac{E_{\text{F}}}{(\hbar v_{\text{F}})^2} \frac{k_{\text{B}}}{\hbar}$ to obtain the temperature dynamics $T_{\text{e}}(t)$. The photovoltage V_{PTE} at a delay time Δt between two pulses follows from $V_{\text{PTE}} = \int_0^{\Delta t} V(t, T_1) dt + \int_{\Delta t}^{\infty} V(t - \Delta t, T_2) dt$. Here we use the instantaneous photo-thermoelectric voltage $V(t, T) = BT(T - T_{\text{L}})$, where B is a proportionality constant related to the Seebeck effect. The initial hot-electron temperature after the first (second) laser pulse T_1 (T_2) is given by $\sqrt{T_{\text{in}}^2 + T_{\text{add}}^2}$, with T_{in} the temperature before arrival of the laser pulse, and T_{add} the temperature equivalent to the added pulse energy. We obtain a photovoltage dip by repeating this calculation while varying Δt , and fit the data to extract the deformation potential, finding $D = 65$ eV (see **Supplementary Figure S5a**). We note that using Ref. [S7] to relate $\frac{A}{\alpha}$ to D would give a D that is a factor $\sqrt{8}$ higher. Thus, we find that in order to reproduce the observed cooling dynamics, the deformation potential would have to be an unrealistically high $D > 65$ eV. This value is a *lower* bound, because the analysis assumes that all scatterers that lead to the mean free path also contribute to supercollision scattering, whereas actually only a fraction – the short-range scatterers – contribute [S7]. Since transport measurements indicate $D < 35$ eV and cooling dynamics indicate $D > 65$ eV, we conclude from this quantitative analysis that supercollision is likely not the dominant cooling mechanism in hBN-encapsulated graphene. This is corroborated by the observed trend that cooling becomes faster with increasing carrier density – exactly opposite to the trend that is measured for supercollision-dominated devices where faster cooling is observed around the Dirac point [S3, S8]. Finally, noise thermometry studies on hBN-encapsulated graphene [S9] and experimental-theoretical terahertz spectroscopy studies on multilayer epitaxial graphene and CVD graphene on different substrates [S10] both led to the conclusion that supercollision cooling does not explain the experimental results.

Normal collision cooling — Cooling to acoustic graphene phonons without disorder-assisted scattering is generally believed to be slow, with typically nanosecond timescales [S11]. However, in the regime where the electron temperature exceeds the lattice and Fermi temperatures, cooling occurs significantly faster. In this regime, cooling is governed by non-

exponential cooling dynamics, according to [S11] $T_e(\Delta t) = \frac{T_e(0)}{\sqrt{\Delta t/\tau_0+1}}$, with $\tau_0 = \frac{424}{D^2[T_e(0)]^2}$, where D is in eV, T_e in meV and τ_0 in μs . We compare these cooling dynamics with our experimental data in Fig. **S5b**, showing that our experimental decay is significantly faster. For the calculated dynamics we use $D = 35$ eV, the maximum value that is consistent with our transport data. Using lower (more realistic) values for D , cooling would be even slower. We also compare the calculated dynamics with exponential decay with a cooling time of 28 ps, which corresponds to the fastest (initial) decay of these cooling dynamics. This is roughly one order of magnitude slower than the experimentally observed dynamics. From this quantitative analysis we conclude that cooling to acoustic graphene phonons is likely not the dominant cooling mechanism. Qualitatively, we observe exponential decay dynamics (at least above a lattice temperature of 200 K) instead of the predicted non-exponential cooling. Furthermore, cooling through normal collisions with graphene acoustic phonons in the overheating regime is predicted to be independent of lattice temperature, whereas we observe clearly slower cooling at lower lattice temperatures.

Cooling to graphene optical phonons

There is general consensus in the literature that cooling to *graphene optical phonons* occurs on a rapid (sub-picosecond) timescale [S12, S13], but mainly for carriers with high energy, on the order of the optical phonon energy of ~ 0.2 eV [S14]. Therefore, one would expect this process to mainly play a role at sufficiently high fluences. On the other hand, a recent study on *non-encapsulated graphene devices* showed that even at moderate fluences (similar to the ones used in our study) cooling by coupling to optical phonons plays a role for the observed picosecond decay dynamics of the THz photoconductivity, as measured by optical pump – THz probe spectroscopy (the same technique that we used in Supp. Fig. **S4b**). This prompts us to study in more detail if cooling to optical phonons could explain our experimental results. Comparing our data and the results in Ref. [S10], we see that some trends are similar for cooling to graphene optical phonons and for cooling to hBN hyperbolic phonons. In particular, both cooling mechanisms lead to slower cooling for decreasing Fermi energies and for increasing fluence. However, there are also a number of distinct differences: *i*) We observe significantly slower cooling for lower substrate temperatures (for all examined equilibrium Fermi energies, $E_F = 0.04 - 0.17$ eV), e.g. a twofold increase in cooling time by

decreasing T_L from 300 K to 200 K, whereas optical phonon cooling predicts that the cooling dynamics are independent of substrate temperature down to 50 K or less (for $E_F = 0.3$ eV) [S10]; *ii*) The optical phonon cooling model predicts either bi-exponential cooling dynamics for $E_F = 0.3$ eV, or rather slow cooling with a time constant of 5-7 ps at $E_F \approx 0$ eV (both at room temperature) [S10], which is not consistent with our observations at room temperature, where we see exponential decay with a decay constant well below 4 ps. Furthermore, we point out that a recent noise thermometry study also showed that cooling through optical phonons plays a minor role for hBN-encapsulated graphene [S15]. Finally, we observe a significant effect of reducing the thickness of the hBN encapsulant (see **Supplementary Figure S6**), which indicates that the observed picosecond cooling dynamics are related to the encapsulation material. Thus, whereas a fraction of the energy of hot electrons is likely lost on a sub-picosecond timescale to optical phonons, the main cooling channel giving rise to the observed picosecond cooling dynamics corresponds to hyperbolic phonon cooling.

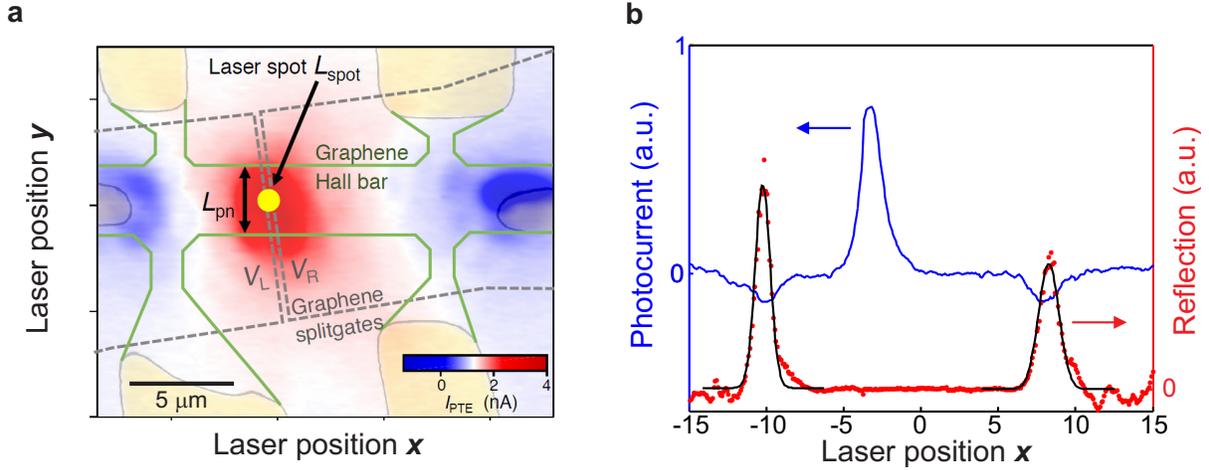
Other cooling channels

Concerning *flexural phonons*, we expect that these will give an even smaller contribution than normal collision scattering with graphene acoustic phonons, because they are quenched by the encapsulation. Even taking into account the elasticity of the substrate, the main branch of the flexural modes is predicted to be (i) gapped and (ii) strongly damped [S16]. *Wiedemann-Franz cooling* refers to lateral heat spreading out of the spot that is excited by the incident light. To assess the relevance of this process, we calculate the cooling length (at room temperature), which is given by $\zeta = \sqrt{\frac{\kappa}{\Gamma}}$ [S17]. Here, Γ is the interfacial heat conductivity and κ is given by the Wiedemann-Franz law: $\kappa = \frac{\pi^2 k_B^2 T \sigma}{e^2}$, where σ is the graphene electrical conductivity and e the electron charge. We find a cooling length of $\zeta \sim 1 \mu\text{m}$, which is smaller than our spot size of $L_{\text{spot}} \sim 2 \mu\text{m}$ (and much smaller than the device size), and thus this lateral heat spreading process is irrelevant for our results at room temperature. We note that at lower temperatures, the conductivity increases and the interfacial heat conductivity Γ decreases, leading to longer cooling lengths. Therefore, at lower lattice temperatures, we do not exclude that Wiedemann-Franz cooling plays a role (see Fig. S9). Finally, we mention that *cooling by tunneling of hot carriers through the hBN slab* is irrelevant for our devices, since the hBN slab is between the graphene and the bottom

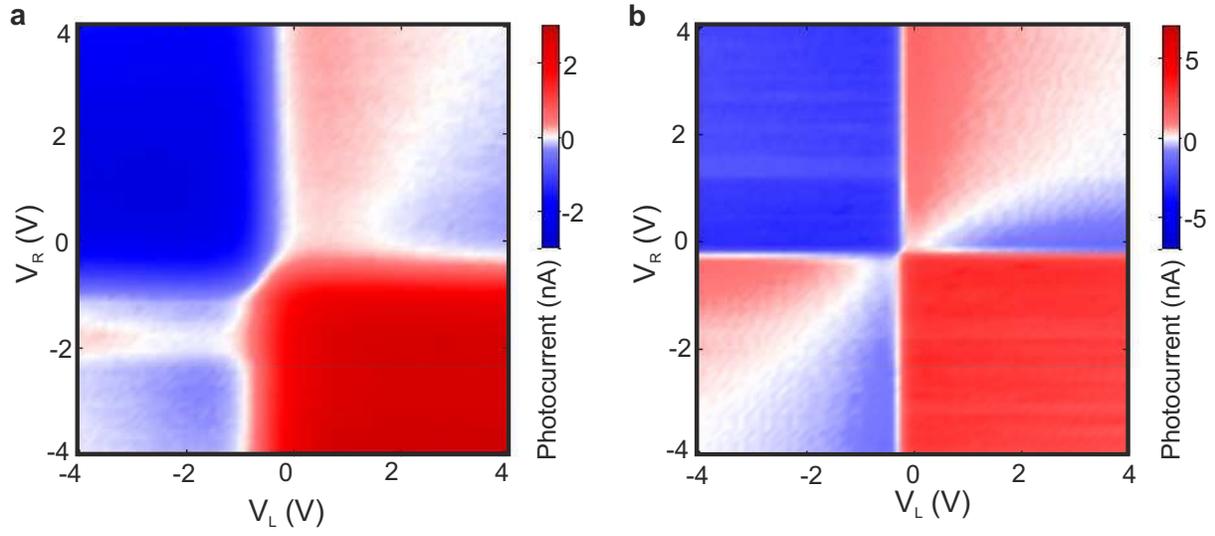
gate has a thickness of 70 nm, whereas tunnelling only plays a role when the two graphene sheets are separated by hBN of just a few layers thick [S18].

Supplementary Section 2: Fabrication of thin hBN encapsulated sample

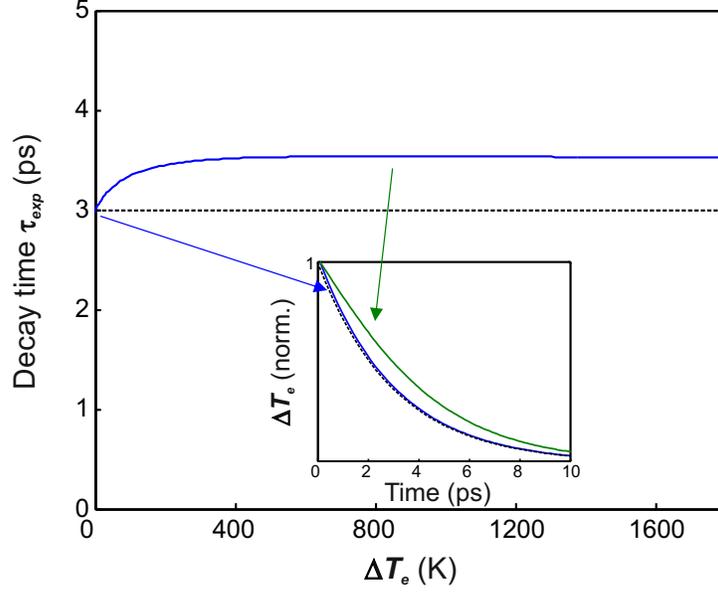
Hexagonal BN and graphene flakes are first produced by micro-mechanical cleavage from bulk crystals onto silicon substrates coated with 285nm silicon dioxide (SiO_2). Suitable hBN and single layer graphene flakes for the encapsulation are then identified by bright and dark field optical microscopy[S21, S22] and optical contrast measurements[S22] and Raman spectroscopy[S23–S25]. The target flakes are then picked up and assembled into the desired heterostructure using a hot-pick up technique, similar to Ref.S26, to minimize contamination containing blisters at the hBN/graphene interfaces. From atomic force microscopy (AFM) characterization (see **Supplementary Figure S6d**) we measure a step change in height $\sim 2\text{nm}$ and $\sim 3\text{nm}$ for the bottom and top hBN respectively. We note that part of this contribution (up to $\sim 1\text{nm}$) may arise from a water layer present on the surface of the hBN [S26], so that we estimate up to 7 layers in hBN encapsulants.



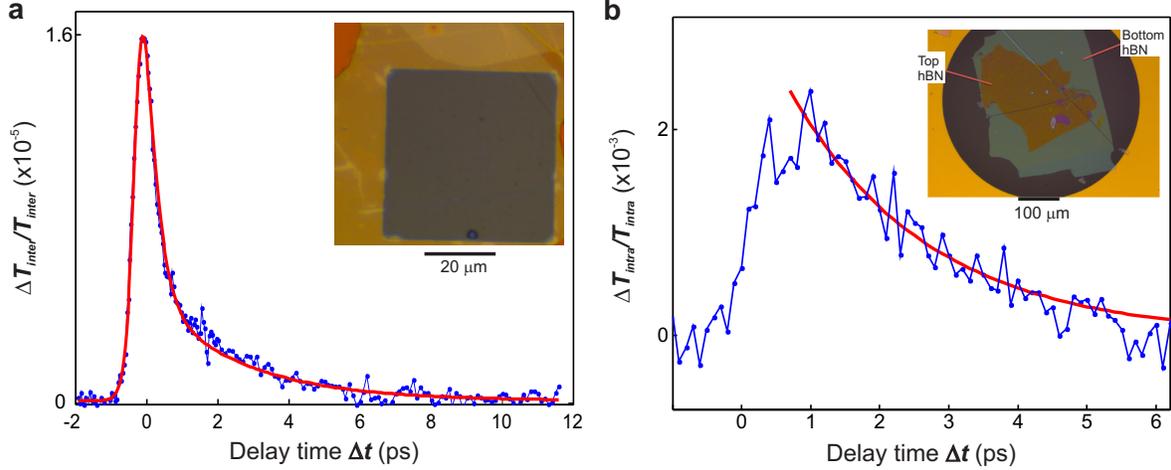
Supplementary Figure S1: Photocurrent and reflection data. a) Spatial image, where the laser focus is scanned over the device, while photocurrent and laser reflection are simultaneously measured (at room temperature). The reflection image indicates the metal contacts (yellow) and we observe negative photocurrent at the metal-graphene interfaces and positive photocurrent at the pn -junction. b) Photovoltage and spatial-derivative reflection line traces as indicated in panel b, with Gaussian fits to the reflection image to determine the spot size. The spatial extent of the photocurrent is similar to the spot size for these room temperature measurements. This shows that the cooling length is relatively short, indicating that on a micron length-scale lateral heat transport is still slower than hot-carrier cooling (at room temperature).



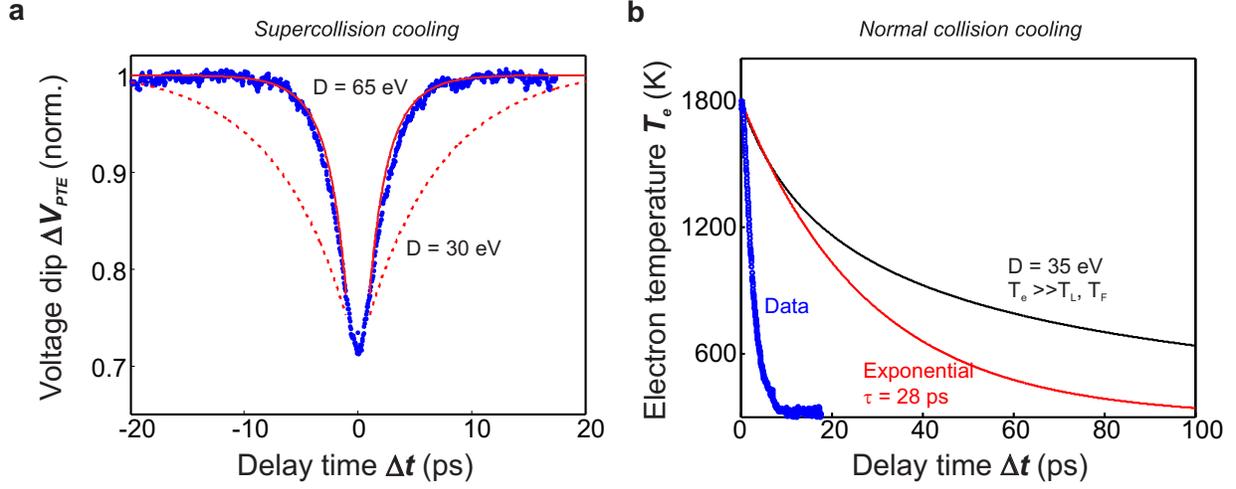
Supplementary Figure S2: Characteristic sixfold pattern for photo-thermoelectric effect. Photovoltage at the interface between the two graphene regions, for single-pulse excitation as a function of voltage on the two graphene split gates V_L and V_R at room temperature (a) and 30 K (b). The sixfold patterns indicate photo-thermoelectric photovoltage generation, which scales with the light-induced increase in carrier temperature [S17, S19].



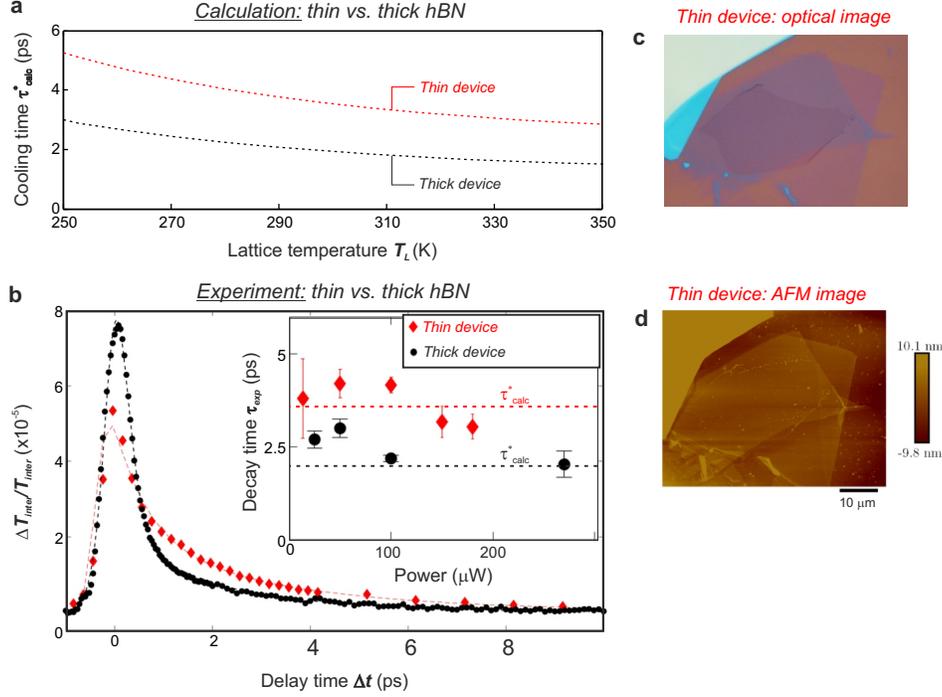
Supplementary Figure S3: Effect of nonlinearity on measured dynamics. To study the relation between the observed dynamics of the photovoltage dip and the actual cooling dynamics, we calculate the photovoltage dip dynamics for a given exponential cooling time of 3 ps. To this purpose, we integrate the generated photocurrent as a function of real time, after an ultrafast laser pulse heats up the carriers to a certain temperature that is related to the incident power and a second pulse heats up the carriers to a temperature that depends on the residual heat in the carrier system (following Ref. [S3]). The higher the power, the higher the peak temperature and the stronger the nonlinearity (due to the temperature-dependent electronic heat capacity). We find that the photocurrent dip as a function of delay time between the two ultrafast pulses gives an overestimation of the cooling dynamics by $<20\%$. The inset shows the 'real' cooling dynamics (black dashed line), the photovoltage dynamics for low power (blue line) and for a laser power that corresponds to $T_e \approx 1000$ K (green line).



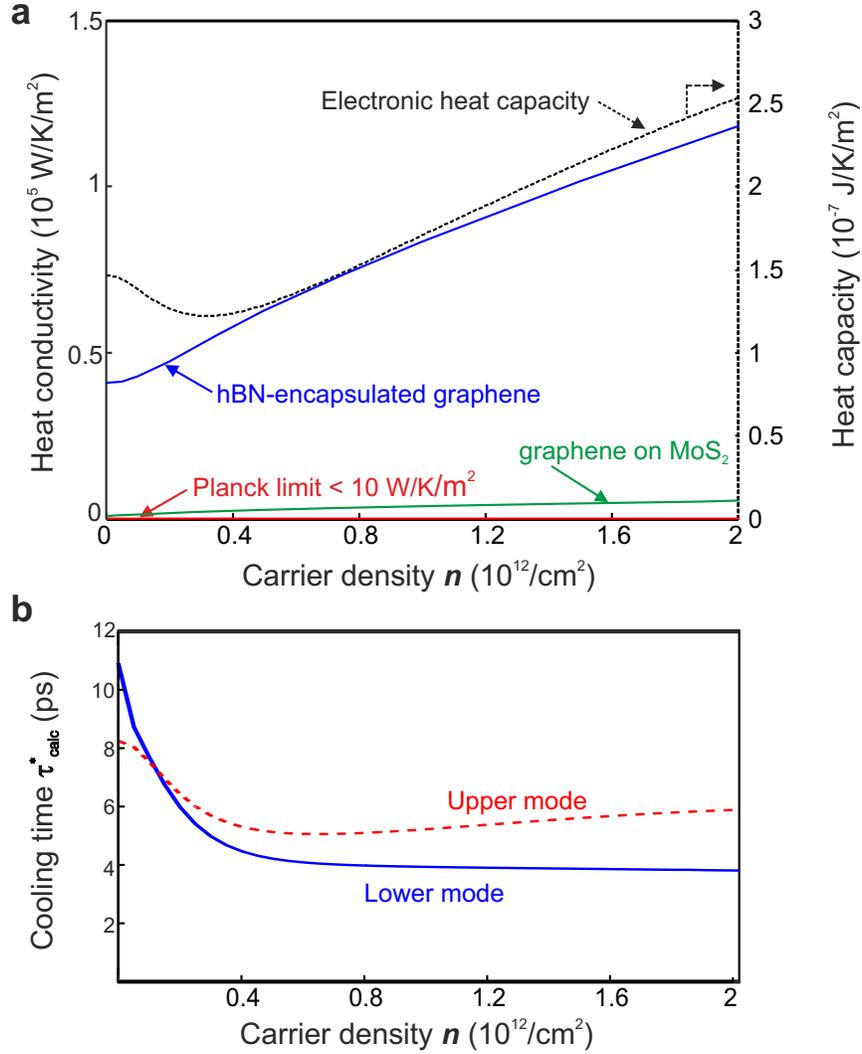
Supplementary Figure S4: Verification of RT decay time using pump-probe spectroscopies. **a)** Ultrafast optical pump – optical probe spectroscopy on hBN-encapsulated CVD graphene, prepared as in Ref. [S20] and fully covering a $50 \mu\text{m}$ square aperture on a transparent SiO_2 substrate (see inset). The ultrashort pump pulses, with a wavelength of 780 nm , 100 fs duration, and incident fluence of $\sim 10 \mu\text{J}/\text{cm}^2$, creates a non-equilibrium distribution that quickly thermalizes through electron-electron scattering [S13]. This modified, hot-carrier distribution affects available interband transitions through Pauli blocking, which we probe using ultrashort pulses at $1.3 \mu\text{m}$. The differential transmission $\Delta T_{\text{inter}}/T_{\text{inter}}$ presents a bi-exponential decay, with the slower component representing the main hot-electron cooling channel with a time scale of $\sim 2.5 \text{ ps}$. The fast initial decay is likely related to the super-linear relation between changes in the electron temperature and changes in the differential transmission. This fast decay component is not seen in time-resolved photocurrent and optical pump – terahertz probe measurements since these techniques both have a sub-linear relation between changes in the electron temperature and changes in the observed signal (see also Supp. Fig. **S8**). **b)** The pump-induced change in transmitted terahertz (THz) signal as a function of pump-probe delay time Δt using an hBN-encapsulated CVD graphene sandwich with dimensions $>200 \mu\text{m}$, prepared as in Ref. [S20] and transferred inside a $500 \mu\text{m}$ round aperture (see inset). The loosely focused pump light, with a wavelength of 800 nm and pulse energy density of $\sim 20 \mu\text{J}/\text{cm}^2$ creates hot carriers and a modified carrier distribution. Terahertz light with frequency $0.4 - 1.2 \text{ THz}$ subsequently probes the effect of the modified distribution on intraband transitions, i.e. the electronic response. The pump-probe dynamics of the differential THz transmission $\Delta T_{\text{intra}}/T_{\text{intra}}$ represent hot-carrier cooling (for carriers with energy below the optical phonon energy) and give a timescale of $\sim 2.2 \text{ ps}$.



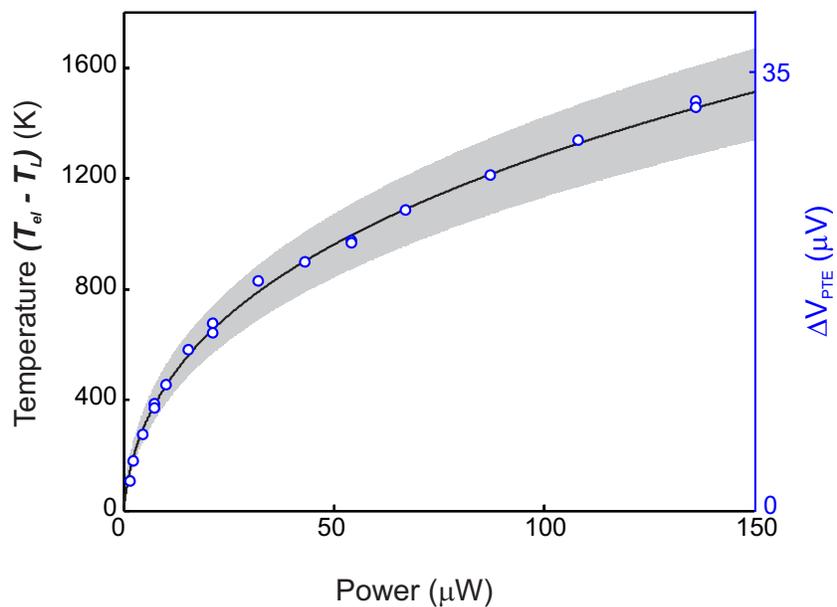
Supplementary Figure S5: Comparing data with alternative cooling mechanisms. The photovoltage dip dynamics ($T_L = 300$ K and $n = 1.7 \times 10^{12}/\text{cm}^2$, compared with the dynamics according to the supercollision cooling model with a deformation potential of 65 eV (solid red line) and 30 eV (dashed red line). The model takes into account that we are in the strong heating regime with $T_e = 1500$ K (see Methods).



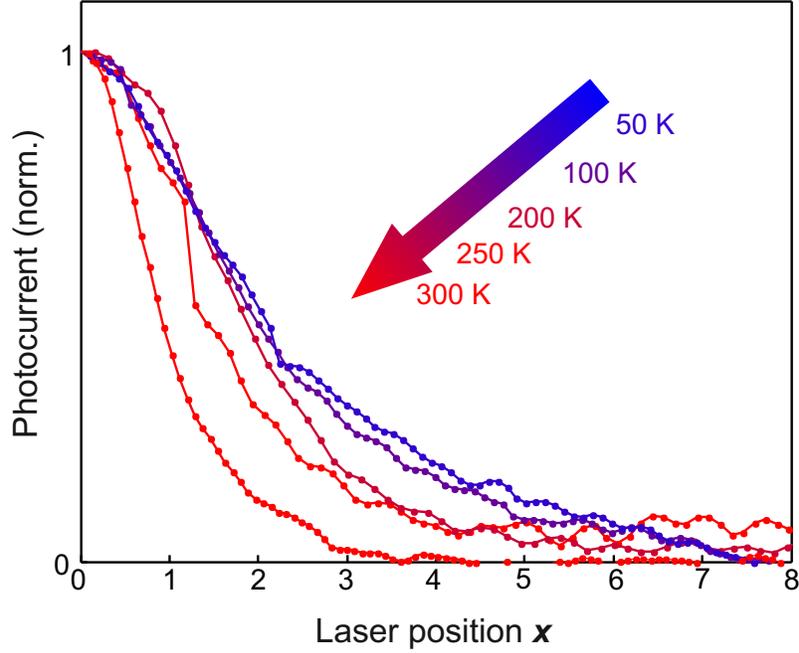
Supplementary Figure S6: Dynamics for different hBN thicknesses. **a)** Calculated (near-equilibrium) cooling time according to the hyperbolic cooling model τ_{calc}^* for a "thick" device with 30 nm bottom and 30 nm top hBN encapsulant (black, dashed line) and a "thin" device with 2 nm top and 3 nm bottom hBN encapsulant (red, dashed line). Clearly, cooling is slower in the device with very thin encapsulation. **b)** Differential transmission as a function of pump-probe delay time for graphene/hBN devices with different encapsulant thickness, as in panel **b**. For both devices, the probe beam was around 1300-1400 nm. The wavelength of the pump beam was 1550 nm (785 nm) for the thin (thick) device and the power 180 (100) μW . We compare measurements taken at slightly different powers, because due to the difference in focus spot size, these powers result in similar initial carrier temperatures for both measurements (hence the similar size of the peak differential transmission right after time zero). We point out that the pump wavelength is not relevant for the picosecond cooling dynamics, since within tens of femtoseconds carrier heating occurs and any memory of the initial energy of the photoexcited e-h pairs is lost. The inset of panel **b** gives the experimental decay times, which are the slower time constants from bi-exponential fits, corresponding to carrier cooling. We compare these experimental decay times with the calculated cooling times (dashed lines). Panels **c)** and **d)** give optical and AFM images of the thin device. We note that both devices have similar Fermi energy as extracted from Raman measurements (not shown), with the "thick" ("thin") device having a G peak at 1584 (1583) cm^{-1} and a 2D peak at 2686 (2685) cm^{-1} .



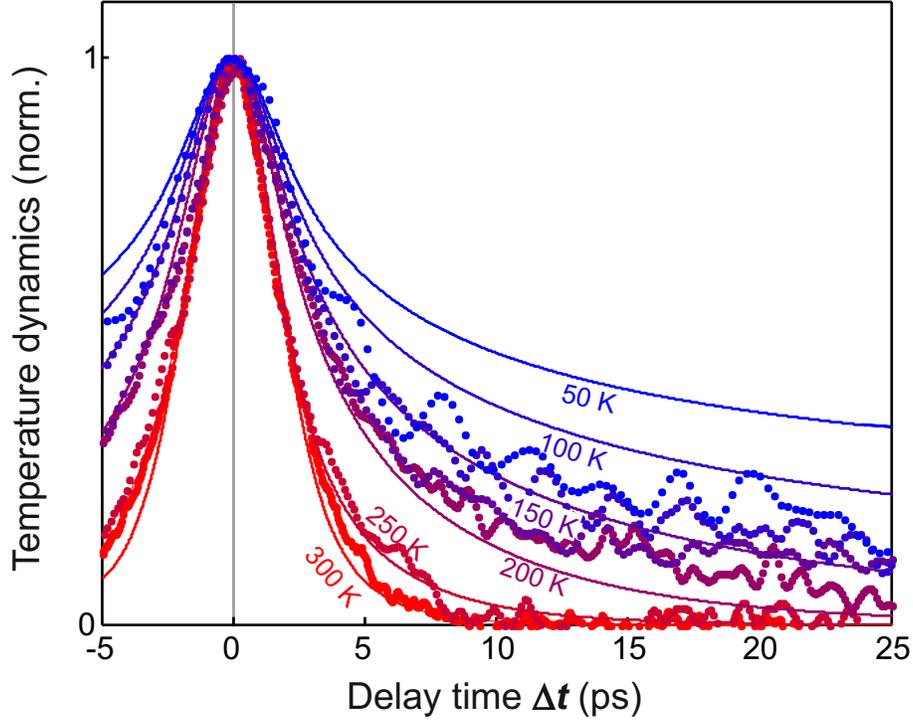
Supplementary Figure S7: Interfacial heat conductivity, heat capacity and contributions of the hyperbolic modes. **a)** Since the hyperbolic cooling time τ_{calc}^* is determined by the ratio of the interfacial heat conductivity Γ and the electronic heat capacity C_n , we plot these separate entities as a function of carrier density n . We also show the interfacial heat conductivity for MoS₂ hyperbolic phonons, and for black-body radiation of graphene in vacuum, which is orders of magnitude lower than for hBN or MoS₂. **b)** The separate contributions to the cooling time from the two distinct hyperbolic phonon polariton modes in hBN: the 'Lower' mode at 100 meV and the 'Upper' mode at 180 meV.



Supplementary Figure S8: Estimation of hot carrier temperature. The photovoltage as a function of laser power (right vertical axis) and the extracted laser-induced temperature increase (left vertical axis) for undoped graphene, using the analysis described in the Methods section. The grey shaded area gives the 95% confidence interval for the extracted electron temperature.



Supplementary Figure S9: Spatial extent of photocurrent. Spatial line traces of photocurrent vs. laser position for five different lattice temperatures. Position 0 corresponds to the pn -junction. At lower temperatures, the spatial extent of the photocurrent increases. This is caused by increased mechanical vibrations of the sample (due to the compressor of our helium closed cycle cryostat) and by a longer cooling length (due to slower hyperbolic hot-carrier cooling and a longer mean free path). This means that lateral heat diffusion out of the laser spot leads to an additional cooling channel, i.e. Wiedemann-Franz cooling. As a result of the larger photocurrent spot size the initial temperature is significantly lower for lower lattice temperature.



Supplementary Figure S10: Cooling dynamics at high doping. Comparison of the complete cooling dynamics as measured (data points) and calculated (solid line) for graphene with $n = 1.7$ and $2 \times 10^{12}/\text{cm}^2$, respectively, and varying lattice temperature. We use an initial temperature increase of $\Delta T = 300\text{--}500$ K for all temperatures, except 1250 K for $T_L = 300$ K, the same as we did for undoped graphene. We see deviations between data and model starting from $T_L = 150$ K, most likely caused by normal collision with graphene acoustic phonons [S11] and scattering with optical phonons [S10] becoming more relevant cooling pathways.

Supplementary references

- [S1] Dean, C.R. et al. Boron nitride substrates for high-quality graphene electronics. *Nature Nanotech* **5**, 722 (2010)
- [S2] Wang, L. et al. One-Dimensional Electrical Contact to a Two-Dimensional Material. *Science* **342**, 614-617 (2013)
- [S3] Graham, M.W., Shi, S.-F., Ralph, D.C., Park, J. & McEuen, P.L. Photocurrent measurements of supercollision cooling in graphene. *Nature Phys.* **9**, 103-108 (2013)
- [S4] Das Sarma, S., Adam, S., Hwang, E. H. & Rossi, E. Electronic transport in two-dimensional graphene. *Rev. Mod. Phys.* **83**, 407 (2011)
- [S5] Hwang, E. H. & Das Sarma, S. Acoustic phonon scattering limited carrier mobility in two-dimensional extrinsic graphene. *Phys. Rev. B* **77**, 1 (2008)
- [S6] Principi, A. et al. Plasmon losses due to electron-phonon scattering: The case of graphene encapsulated in hexagonal boron nitride. *Phys. Rev. B* **90**, 14 (2014)
- [S7] Song, J.C.W., Reizer, M.Y. & Levitov, L.S. Disorder-Assisted Electron-Phonon Scattering and Cooling Pathways in Graphene. *Phys. Rev. Lett.* **109**, 106602 (2012).
- [S8] Graham, M.W. et al. Transient Absorption and Photocurrent Microscopy Show That Hot Electron Supercollisions Describe the Rate-Limiting Relaxation Step in Graphene. *Nano Lett.* **13** 54975502 (2013)
- [S9] Crossno, J. et al. Development of high frequency and wide bandwidth Johnson noise thermometry. *App. Phys. Lett.* **106**, 023121 (2015)
- [S10] Mihnev, M.T. et al. Microscopic origins of the terahertz carrier relaxation and cooling dynamics in graphene. *Nature Comm.* **7**, 11617 (2016)
- [S11] Bistritzer, R. & MacDonald, A. H. Electronic cooling in graphene. *Phys. Rev. Lett.* **102**, 206410 (2009).
- [S12] Kampfrath, T. et al. Strongly coupled optical phonons in the ultrafast dynamics of the electronic energy and current relaxation in graphite. *Phys. Rev. Lett.* **95**, 187403 (2005).
- [S13] D. Brida et al., Ultrafast collinear scattering and carrier multiplication in graphene. *Nature Comm.* **4**, 1987 (2013)

- [S14] Viljas, J.K. & Heikkilä, T.T. Electron-phonon heat transfer in monolayer and bilayer graphene. *Phys. Rev. B* **81**, 245404 (2010)
- [S15] Yang, W. et al. Hyperbolic cooling of a graphene on hBN transistor in the Zener-Klein regime. *Arxiv:1702.02829* (2017)
- [S16] Amorim, B and Guinea, F. Flexural mode of graphene on a substrate. *Phys. Rev. B* **88**, 115418 (2013)
- [S17] Song, J.C.W. et al. Hot Carrier Transport and Photocurrent Response in Graphene. *Nano Lett.* **11**, 4688-4692 (2011)
- [S18] Britnell et al. Electron Tunneling through Ultrathin Boron Nitride Crystalline Barriers. *Nano Lett.* **12**, 1707 (2012)
- [S19] Gabor N.M. et al. Hot Carrier-Assisted Intrinsic Photoresponse in Graphene. *Science* **334**, 648-652 (2011)
- [S20] Banszerus, L. et al. Ultrahigh-mobility graphene devices from chemical vapor deposition on reusable copper. *Science Adv.* **1**, e1500222 (2015)
- [S21] R. V. Gorbachev. Hunting for monolayer boron nitride: Optical and raman signatures. *Small* **7**, 465 (2011).
- [S22] Casiraghi, C. et al. Rayleigh imaging of graphene and graphene layers. *Nano Letters* **7**, 2711 (2007).
- [S23] Ferrari, A.C. et al., The Raman Spectrum of Graphene and Graphene Layers *Phys. Rev. Lett.* **97**, 187401 (2006).
- [S24] Ferrari, A.C., Basko, D.M. Raman spectroscopy as a versatile tool for studying the properties of graphene. *Nat. Nanotechnol.* **8**, 235 (2013).
- [S25] Arenal, R. et al. Raman Spectroscopy of Single-Wall Boron Nitride Nanotubes. *Nano Lett.* **6**, 1812 (2006)
- [S26] F. Pizzocchero et al. The hot pick-up technique for batch assembly of van der Waals heterostructures. *Nat. Commun.* **7**, 11894 (2016).