# 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] Mark B. Lundeberg,[1] Eva A. A. Pogna,[3] Luca Banszerus,[4] Zoltán Mics,[5] Mathieu Massicotte,[1] Peter Schmidt,[1] Diana Davydovskaya,[1] David G. Purdie,[6] Ilya Goykhman,[6] Giancarlo Soavi,[6] Antonio Lombardo,[6] Kenji Watanabe,[7] Takashi Taniguchi,[7] Mischa Bonn,[5] Dmitry Turchinovich,[5] Christoph Stampfer,[4] Andrea C. Ferrari,[6] Giulio Cerullo,[3] Marco Polini,[8] and Frank H. L. Koppens[1],[9],[*]

[1]*ICFO - Institut de Ciències Fotòniques, The Barcelona Institute of Science and Technology, Castelldefels (Barcelona) 08860, Spain*
[2]*Radboud University, Institute for Molecules and Materials, NL-6525 AJ Nijmegen, the Netherlands*
[3]*IFN-CNR, Dipartimento di Fisica, Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy*
[4]*JARA-FIT and 2nd Institute of Physics, RWTH Aachen University, 52074 Aachen, Germany*
[5]*Max Planck Institute for Polymer Research, Ackermannweg 10, Mainz 55128, Germany*
[6]*Cambridge Graphene Centre, University of Cambridge, Cambridge CB3 OFA, UK*
[7]*National Institute for Material Science, 1-1 Namiki, Tsukuba 305-0044, Japan*
[8]*Istituto Italiano di Tecnologia, Graphene Labs, Via Morego 30, I-16163 Genova, Italy*
[9]*ICREA - Institució Catalana de Reçerca i Estudis Avancats, 08010 Barcelona, Spain*

**Van der Waals heterostructures have emerged as promising building blocks that offer access to new physics, novel device functionalities, and superior electrical and optoelectronic properties [1–7]. Applications such as thermal management, photodetection, light emission, data communication, high-speed electronics and light harvesting [8–16] require a thorough understanding of (nanoscale) heat flow. Here, using time-resolved photocurrent measurements we identify an efficient out-of-plane energy transfer channel, where charge carriers in graphene couple to hyperbolic phonon polaritons [17–19] in the encapsulating layered material. This hyperbolic cooling is particularly efficient, giving picosecond cooling times, for hexagonal BN, where the high-momentum hyperbolic phonon polaritons enable efficient near-field energy transfer. We study this heat transfer mechanism through distinct control knobs to vary carrier density and lattice temperature, and find excellent agreement with theory without any adjustable parameters. These insights may lead to the ability to control heat flow in van der Waals heterostructures.**

Owing to its large in-plane thermal conductivity, graphene has been suggested as material for the thermal management of nanoscale devices [8]. At the same time, graphene is well-known for its ability to convert incident light into electrical heat, i.e. hot electrons that can be used to generate photocurrent, with applications in photodetection, data communication and light harvesting [10, 20, 21]. Understanding, and ultimately controlling, heat flow in graphene-van der Waals heterostructures is therefore of paramount importance. For example, a short cooling time of graphene hot carriers is advantageous for thermal management and for high switching rates of photodetectors (PDs) for data communication, whereas a long cooling time is favorable for photodetection sensitivity [10, 20, 21]. Of particular relevance are heterostructure devices that contain high-quality graphene encapsulated by layered materials, such as hexagonal BN (hBN) and $MoS_2$, which have the potential to crucially improve the performance of electronic and optoelectronic devices [1, 2].

A number of cooling pathways for graphene carriers have been proposed, involving among others strongly coupled optical phonons [22–24], acoustic phonons [25–28], substrate phonons [29] and plasmons [30] (see also Appendix 1). Here, using several experimental approaches, we show that cooling in graphene encapsulated by hBN is governed by out-of-plane coupling of graphene electrons to special polar phonon modes that occur in layered materials (LMs): hyperbolic phonon polaritons, where $\epsilon_{xx}\epsilon_{zz} < 0$, with $\epsilon_{xx}$ and $\epsilon_{zz}$ the permittivity parallel and perpendicular to the LM plane. Owing to this property, these materials carry deep sub-wavelength, ray-like optical phonon polaritons. For hBN, within the two Reststrahlen bands a large number of hyperbolic phonon modes exist with high momenta, far outside the light cone. The most notable modes occur at an energy of $\sim$100 meV and $\sim$180 meV [17], such that energy overlap with the graphene hot-carrier distribution is substantial. The unusual hyperbolic character gives rise to a very high density of optical states, and thus large thermal energy densities [17, 18], thereby providing a potentially efficient cooling pathway for hot carriers in graphene. By near-field coupling between graphene and hBN, efficient energy transfer from hot carriers to hyperbolic phonon polaritons is possible [31]. Here we show that the measured

---

*Electronic address: Correspondence: klaas-jan.tielrooij@icfo.eu, frank.koppens@icfo.eu
‡ Equal contribution.*

carrier dynamics of hBN-encapsulated graphene can be explained by this hyperbolic cooling process, as illustrated in **Fig. 1a**.
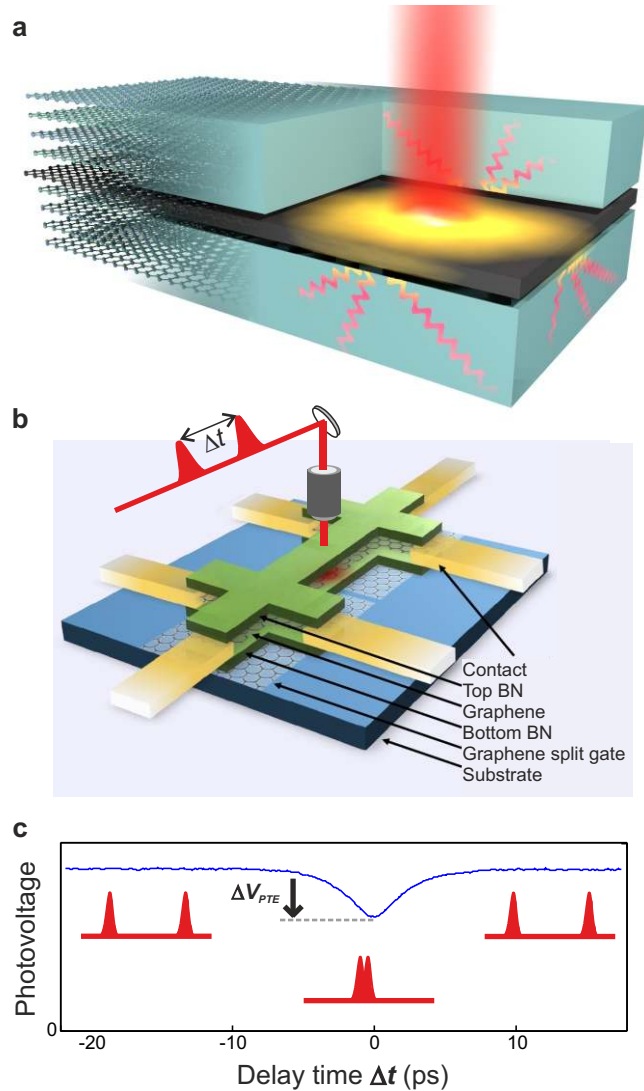


**Fig. 1**: **Hot carrier cooling in hBN-encapsulated graphene. a)** Schematic representation of out-of-plane heat transfer in hBN-encapsulated graphene: highly efficient near-field emission from graphene hot carriers into directional hyperbolic phonon polaritons of the encapsulating material. **b)** Schematic drawing of the hBN-encapsulated graphene device with graphene split gates. By applying a different voltage to the two split gates we create a *pn*-junction in the encapsulated graphene layer, which we illuminate with two ultrafast pulses that arrive with a variable time delay $\Delta t$. **c)** The resultant photo-thermoelectric photovoltage measured between two contacts shows a dip when the two ultrafast pulses overlap. This is due to the nonlinear relation between hot carrier temperature and incident light intensity. The dynamics of the photovoltage dip $\Delta V_{\mathrm{PTE}}$ as a function of time delay $\Delta t$ therefore correspond to the cooling dynamics of the hot carriers at the *pn*-junction.

We use ultrafast time-resolved photocurrent measure-

ments and show the results obtained from one hBN-encapsulated, Hall-bar shaped, exfoliated graphene device (see **Fig. 1b**). The device contains chemical vapor deposited (CVD) graphene split gates underneath the bottom hBN (thickness: 70 nm), in order to generate a *pn*-junction in the middle of the device. A second hBN-encapsulated device with metallic split gates gave fully consistent results. With 800 nm light incident on the *pn*-junction, a photovoltage is generated by the photo-thermoelectric (PTE) effect [20, 21] (see the photocurrent and reflection data in **Fig. 5** and the characteristic PTE sixfold pattern in **Fig. 6**). By varying the delay $\Delta t$ between two sub-picosecond pulses we extract the carrier dynamics from the photovoltage signal $\Delta V_{\mathrm{PTE}}(\Delta t)$ (see **Fig. 1c**). Since the photo-thermoelectric voltage scales with the light-induced increase in carrier temperature, the decay dynamics of $\Delta V_{\mathrm{PTE}}(\Delta t)$ closely mimic the cooling dynamics of the hot electron system (see **Fig. 7** for an analysis of the extent to which the experimentally observed decay dynamics correspond to the underlying cooling dynamics). In the case of exponential decay dynamics, which we observe above a lattice temperature of ~200 K, we extract an experimental cooling time scale $\tau_{\mathrm{exp}}$ by describing the decay dynamics with $\Delta V_{\mathrm{PTE}}(\Delta t) \propto e^{-\Delta t/\tau_{\mathrm{exp}}}$.

First, we electrically characterize our device using four-probe measurements (see **Fig. 2a**) and find a mean free path of $k_{\mathrm{F}}\ell = 80\text{–}100$ at a carrier density of $n = 1.7 \times 10^{12}/\mathrm{cm}^2$, corresponding to a mobility of 25,000–30,000 cm$^2$/Vs and momentum scattering time of 340–440 fs. As expected, this is much higher than SiO$_2$-supported devices with typically $k_{\mathrm{F}}\ell < 10$, mobility $< 5,000$ cm$^2$/Vs and momentum scattering time $< 100$ fs (see e.g. Ref. [27]). In such devices, carrier cooling is typically ascribed to disorder-assisted cooling to graphene acoustic phonons [26–28].

To study hot-carrier cooling in our high-mobility, encapsulated devices we examine $\Delta V_{\mathrm{PTE}}(\Delta t)$ while varying graphene's most characteristic parameter, the carrier density $n$. In particular, we apply a gate voltage $V_{\mathrm{L}} = +V - V_{\mathrm{D}}$ to the left split gate and $V_{\mathrm{R}} = -V - V_{\mathrm{D}}$ to the right gate, such that there is always either a *pn*-junction or an *np*-junction, with equal electron and hole densities in the two graphene regions ($V_{\mathrm{D}}$ is the gate voltage that corresponds to the Dirac point). The incident laser fluence is typically 5 – 40 $\mu$J/cm$^2$. The data show that cooling becomes faster upon increasing the carrier density (see **Fig. 2b**). We also vary the lattice temperature $T_{\mathrm{L}}$ and observe faster decay for increasing lattice temperature (see **Fig. 2c**, taken at $n \approx 10^{12}/\mathrm{cm}^2$). At low temperatures (below ~200 K), cooling is non-exponential, whereas at room temperature we observe exponential decay of the photovoltage dip, with a timescale of $\tau_{\mathrm{exp}} \approx 2.5$ ps (for $n = 1.7 \times 10^{12}/\mathrm{cm}^2$). We independently verify this cooling time using two alternative measurement techniques that are both sensitive to electron cooling dynamics in different ways (see **Fig. 8**). Firstly, using ultrafast optical pump – optical probe

spectroscopy, which probes interband transitions [24], we find $\tau_{\exp}$= 2.55 ps for the decay of the absorption photobleaching. Secondly, using optical pump – terahertz probe spectroscopy, which probes intraband transitions [13], we obtain $\tau_{\exp}$ = 2.2 ps for the decay of the photoconductivity. For these two experiments we use similar excitation conditions (an incident pulse fluence of 8–20 $\mu$J/cm$^2$) as in the photocurrent measurements, and the measurements are performed on two separate devices consisting of large-area, high-quality hBN-encapsulated CVD graphene as in Ref. [32]. Thus, all three techniques consistently yield similar cooling times for hBN-encapsulated graphene.

We compare these observations with different cooling mechanisms, and note that the data are qualitatively and quantitatively inconsistent with in-plane cooling by scattering with graphene acoustic phonons through normal [25] or disorder-assisted collisions [26–28] (see Appendix 1 and **Fig. 9**). In-plane cooling can also occur by scattering to graphene optical phonons, typically occurring on a sub-picosecond timescale [22, 24]. Reference [23] reports that for *non-encapsulated* graphene this channel gives picosecond decay dynamics of the THz photoconductivity. However these results deviate in several ways from our observations for hBN-encapsulated graphene (see Appendix 1). For example, we observe a twofold increase in cooling time by decreasing $T_L$ from 300 K to 200 K, whereas optical phonon cooling gives cooling dynamics that are independent of substrate temperature (for $E_F = 0.3$ eV), see Ref. [23]. This indicates that, besides a fraction of hot graphene electrons possibly cooling through optical phonons, a different cooling mechanism plays an important role. Since we observe a striking effect of the hBN crystal slab thickness on the cooling dynamics (see **Fig. 10** and Appendix 2), we propose an out-of-plane cooling mechanism where hot carriers in graphene lose their energy to remote polar phonons in the encapsulating layered material. The energy transfer can be understood from fluctuation electrodynamics, where any process that dissipates energy in the form of heat, has a reverse process that is driven by thermal fluctuations and thus becomes stronger at higher temperatures [33, 34]. An example of such paired processes is light absorption and blackbody radiation, which means that graphene emits thermal noise due to the dissipative real part (indicated by $\mathcal{R}$) of the frequency- and momentum-dependent optical sheet conductivity $\sigma(\omega, k)$. This thermal noise is efficiently absorbed by hBN, hBN being a lossy-polarizable material, leading to an energy transfer rate [31]:

$$\mathcal{Q} = \iiint_{-\infty}^{\infty} \frac{d\omega dk_x dk_y}{(2\pi)^3} [n_B(\omega, T_e) - n_B(\omega, T_L)] M(\omega, k) \quad, \tag{1}$$

where $n_B(\omega, T) = \frac{\hbar|\omega|}{e^{\hbar|\omega|/k_B T} - 1}$ and $M(\omega, k) = 4\frac{\mathcal{R}\{Y(\omega,k)\}\mathcal{R}\{\sigma(\omega,k)\}}{|Y(\omega,k)+\sigma(\omega,k)|^2}$ ($k_B$ is Boltzmann's constant). Heat transfer from hot graphene electrons to hBN hyperbolic phonon polaritons is thus governed by the Bose factor
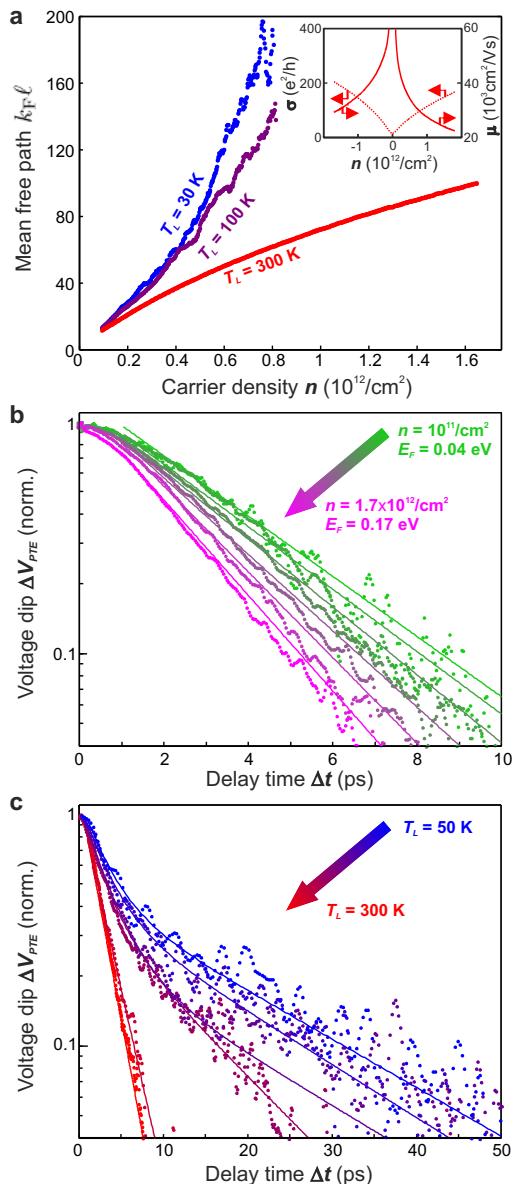


**Fig. 2**: **The effect of doping and lattice temperature. a)** From four-probe transport measurements, we extract the dimensionless mean free path parameter $k_F\ell$ as a function of carrier density for three different lattice temperatures. These data show that carrier scattering in our hBN-encapsulated graphene is much less efficient than in SiO$_2$-supported graphene where $k_F\ell$ is typically 10 or less, and where disorder-assisted supercollisions constitute the dominant cooling mechanism [26–28]. The inset shows the conductivity and mobility vs. gate-induced carrier density. By comparing four-probe and two-probe measurements we find that each contact has a contact resistance of 5 –8 k$\Omega$. **b-c)** The dynamics of the photovoltage dip $\Delta V_{\mathrm{PTE}}$, showing that hot-carrier cooling becomes faster upon increasing the carrier density **(b)** and upon increasing the lattice temperature **(c)**. Panel **b** is measured at a lattice temperature of $T_L = 300$ K and panel **c** at a carrier density of $n \approx 1.7 \times 10^{12}$/cm$^2$. The solid lines are bi-exponential guides to the eye.

$[n_B(\omega, T_e) - n_B(\omega, T_L)]$, which describes the energy disequilibrium between hot graphene carriers and the cold hBN phonon system, and the impedance matching function $M(\omega, k)$, which is nonzero when the surface admittance $Y(\omega, k)$ has a nonzero real part. We calculate $Y(\omega, k)$ as in Ref. [35] (see Methods for details) and find that nearby the hyperbolic hBN phonon frequencies $Y(\omega, k)$ is real over a large $k$-space area and relatively wide frequency band. Since $\sigma(\omega, k)$, calculated using the Random Phase Approximation [36], also has a significant real part, this leads to an impedance matching function $M$ approaching unity. Due to this near-field coupling to hyperbolic modes, the heat conductivity exceeds Planck's law for blackbody radiation by orders of magnitude (see **Fig. 11**). The reason for this is that in vacuum, the $k$-space for blackbody radiation is limited to $k < \omega/c$ (with $c$ the speed of light), whereas this restriction is lifted in the near-field interaction with hBN hyperbolic phonon polaritons. This super-Planckian coupling to hyperbolic hBN phonons thus provides a highly efficient cooling channel for hot carriers in graphene. Cooling to hyperbolic modes also occurs in materials such as $MoS_2$, although there it is not as efficient as for hBN (see **Fig. 11**).

To compare our hyperbolic cooling theory with the experimental data, we examine the calculated energy transfer rate $\mathcal{Q}$. Here, solving Eq. (1) gives cooling dynamics with a cooling time scale

$$\tau_{\mathrm{calc}}(T_e, T_L) = C_n \frac{T_e - T_L}{\mathcal{Q}} \quad , \qquad (2)$$

where $C_n$ is the electronic heat capacity of graphene at constant $n$ [31]. In the limit of weak heating, where $T_e$ approaches $T_L$, we obtain exponential decay with the near-equilibrium timescale $\tau_{\mathrm{calc}}(T_e \to T_L) = \tau_{\mathrm{calc}}^* = C_n \left(\frac{\partial \mathcal{Q}}{\partial T_e}\big|_{T_e = T_L}\right)^{-1}$, where $\frac{\partial \mathcal{Q}}{\partial T_e}\big|_{T_e = T_L}$ is the interfacial heat conductivity $\Gamma$. We compare the calculated near-equilibrium cooling time $\tau_{\mathrm{calc}}^*$ with the measured exponential decay time $\tau_{\mathrm{exp}}$ (see **Fig. 3**), although strictly speaking these experimental timescales correspond to the strong heating regime, where $T_e \gg T_L$. The reason for this is that our technique is not sensitive enough at low incident powers, which means that experimentally we cannot directly access the near-equilibrium cooling time $\tau_{\mathrm{calc}}^*$. Nevertheless, we find that our hyperbolic hBN cooling model semi-quantitatively reproduces the experimentally observed trends for the entire range of investigated carrier densities (up to $1.6 \times 10^{12}/\mathrm{cm}^2$) and lattice temperatures (200 – 300 K). In particular, cooling slows down for lower lattice temperatures, which we attribute mainly to the smaller Bose factor in Eq. (1) and thus smaller energy transfer rate $\mathcal{Q}$. The calculations also reproduce the observation of a longer cooling time around the Dirac point. This is the result of the energy transfer rate $\mathcal{Q}$ smoothly decreasing towards zero carrier density, while the electronic heat capacity $C_n$ flattens around the Dirac point towards its neutral graphene value (see **Fig. 11**). At higher carrier densities ($n > 10^{12}/\mathrm{cm}^2$) the in-
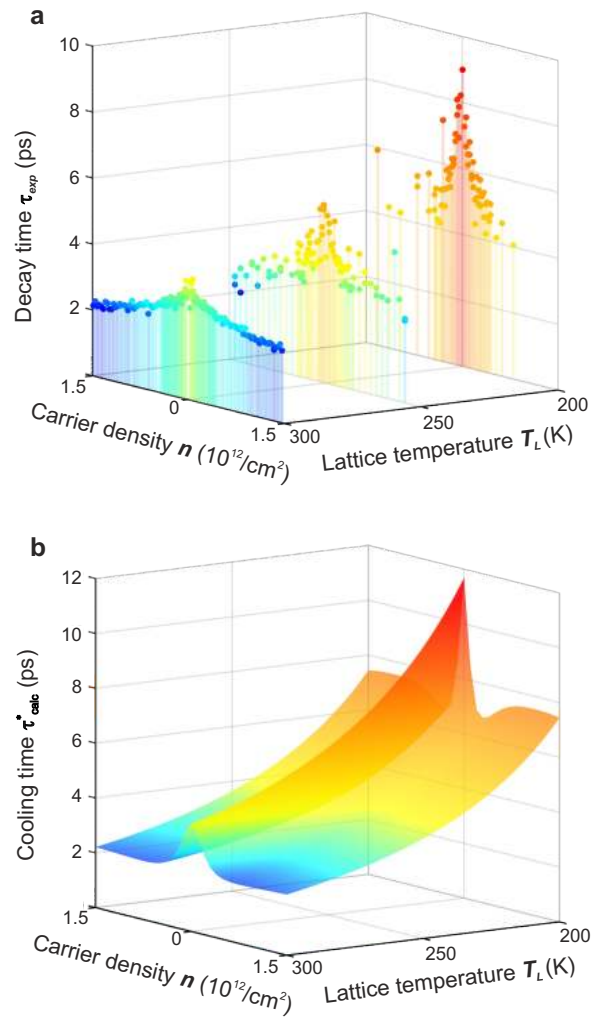
**Fig. 3**: **Qualitative comparison with hyperbolic hBN cooling.** Comparison between the experimental decay time $\tau_{\mathrm{exp}}$ extracted from the dynamics of the photovoltage dip (**a**) and the predicted near-equilibrium cooling time $\tau_{\mathrm{calc}}^*$ for super-Planckian cooling to hyperbolic hBN phonons (**b**). Both in experiment and theory, we vary the carrier density and lattice temperature.

creasing energy transfer rate is compensated by the increasing heat capacity, leading to a weak dependence of cooling time on carrier density. Cooling to hyperbolic modes was also observed in a noise thermometry study [37]. We further note that the two distinct hyperbolic modes contribute almost equally to the overall cooling time, with the lower-energy mode slightly dominant (see **Fig. 11**). This can be attributed to more energy overlap in the Bose factor. We note that the hyperbolic cooling model reproduces the observed slower decay dynamics for encapsulation with very thin hBN flakes (see **Fig. 10**). In this case, cooling is slower because overall there is a lower density of hyperbolic modes to couple to.

To make a more quantitative comparison, we take into account that the measurements are typically done in the

strong heating regime, where $T_e \gg T_L$. We first measure the cooling dynamics for increasing laser power and estimate the electron temperatures that correspond to each power from the characteristic power-dependent photoresponse (see **Fig. 12** and Methods), thus obtaining the exponential decay time $\tau_{\mathrm{exp}}$ vs. $T_e$ (see **Fig. 4a**). In the experimentally accessible regime ($\Delta T_e = T_e - T_L > 200$ K) the $\tau_{\mathrm{exp}}$ increases with increasing $T_e$ (measured for $n = 1.2 \times 10^{12}/\mathrm{cm}^2$). We compare this with the calculated cooling time $\tau_{\mathrm{calc}}$, which describes the "instantaneous" cooling time at a certain $T_e$, and find quantitative agreement *without any adjustable parameters*. The reason for the increasing cooling time with increasing $T_e$ can be seen from Eq. (2) and by noting that $\mathcal{Q}/(T_e - T_L)$ scales roughly linearly with $(T_e - T_L)$. At the same time, the electronic heat capacity $C_n$, increases more than linearly with increasing $T_e$ for large $T_e$. This leads to a net increase in the cooling time with increasing $T_e$. The agreement between experiment and calculation prompts us to make a more quantitative comparison for varying carrier density $n$. **Fig. 4b** shows $\tau_{\mathrm{exp}}$ for three different laser powers ($P = 21$, 47 and 94 $\mu$W), corresponding to three different initial hot electron temperatures ($\Delta T_e = 650$, 950 and 1250 K), together with the calculated near-equilibrium cooling time $\tau^*_{\mathrm{calc}}$. The comparison between experimental results in the strong heating regime ($\tau_{\mathrm{exp}}$) and theoretical results close to equilibrium ($\tau^*_{\mathrm{calc}}$) is justified in **Fig. 4a**, where we show that the cooling time at $T_e \sim 1000$ K is similar to $\tau^*_{\mathrm{calc}}$. Again we find agreement between experiment and theory *without any adjustable parameters*.

Finally, we compare the calculated time-domain cooling dynamics in the strong heating regime directly with the measured photocurrent dynamics. The calculated cooling dynamics describe cooling from an initial $T_e$ down to $T_L$, as constructed from the electron temperature-dependent cooling time $\tau_{\mathrm{calc}}$ in Eq. (2). We compare these calculated cooling dynamics directly with the measured photovoltage dynamics $\Delta V_{\mathrm{PTE}}(\Delta t)$ for six different lattice temperatures, at a low carrier density of $n = 0.06 \times 10^{12}/\mathrm{cm}^2$ (see **Fig. 4c**). Using the initial temperature (at the *pn*-junction) as a fit parameter, we find that the hyperbolic cooling model can describe the experimental data very well. At room temperature, the initial temperature is similar to the one we calculate (see **Fig. 12**), whereas we find a lower initial temperature at lower lattice temperatures. This is due to the increased mechanical vibrations of the sample and lateral heat diffusion out of the laser spot at lower $T_L$, which both lead to a larger photocurrent spot size and thus a lower effective initial $T_e$ at the *pn*-junction (see **Fig. 13**). For a lattice temperature of 50 K, we furthermore see an increasing discrepancy for a delay time >15 ps, which most likely indicates that hBN-phonons alone do not account for the complete cooling dynamics. Similarly, we find that at a higher carrier density of $n = 1.7 \times 10^{12}/\mathrm{cm}^2$ the hyperbolic cooling model starts deviating from the experimental data already for $T_L < 200$ K (see **Fig. 14**).
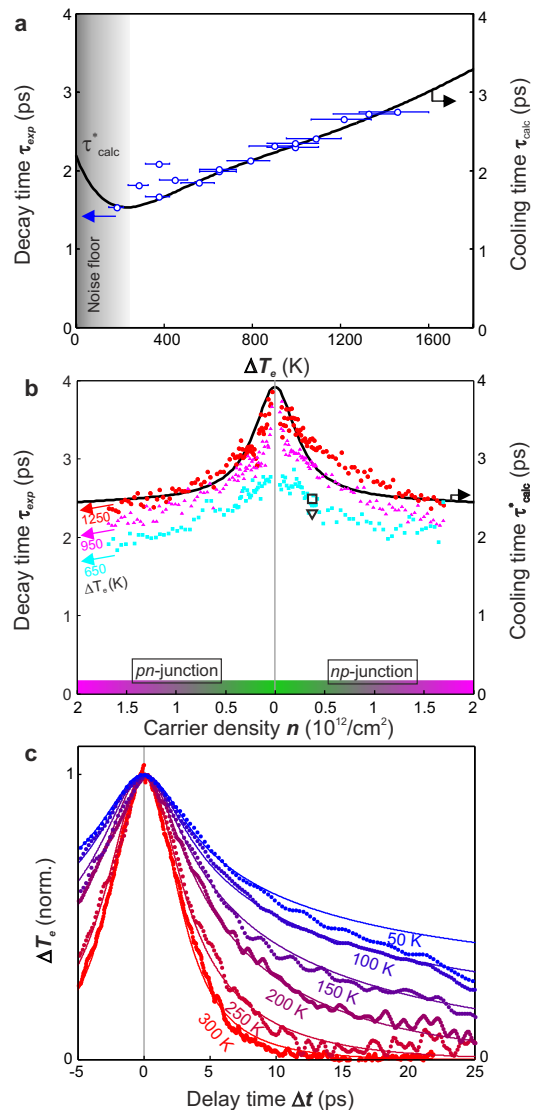


**Fig. 4**: **Quantitative comparison with hyperbolic hBN cooling. a)** Comparison between measured decay time $\tau_{\mathrm{exp}}$ (blue symbols and left vertical axis) and calculated cooling time $\tau_{\mathrm{calc}}$ (solid black line and right vertical axis) as a function of electron temperature *without any adjustable parameters*. **b)** The measured decay time $\tau_{\mathrm{exp}}$ (symbols and left vertical axis) at room temperature as a function of carrier density for three different laser powers, compared with the near-equilibrium cooling time $\tau^*_{\mathrm{calc}}$ (solid line and right vertical axis) according to super-Planckian cooling to hyperbolic hBN phonons *without any adjustable parameters*. The open square (triangle) shows the obtained decay time from time-resolved optical (terahertz) spectroscopy for $\Delta T_e = 600 \pm 200$ K. **c)** Comparison of the complete cooling dynamics as measured (data points) and calculated (solid line) for $n = 0.06 \times 10^{12}/\mathrm{cm}^2$ for varying lattice temperature. The calculated dynamics are convoluted with a Gaussian function representing the experimental time resolution. We find an initial temperature increase of $\Delta T = 300$–500 K for all temperatures, except 1250 K for $T_L = 300$ K. The lower $\Delta T$ is related to the larger effective photocurrent spot size.

Most likely, at sufficiently low lattice temperature and sufficiently high carrier density, cooling through optical phonons [22, 23] and momentum-conserving cooling to acoustic graphene phonons [25, 38, 39] become the dominant channels.

In summary, we addressed the issue of out-of-plane heat transfer in the device architecture of hBN-encapsulated graphene. Combining experiments with microscopic theoretical calculations, we showed that the dominant cooling channel is the one in which heat transfer from hot carriers to the hBN polar substrate occurs via near-field coupling to hyperbolic phonon modes. This efficient mechanism explains the observation of a lower-than-expected cooling time in clean hBN-encapsulated graphene. We note that the observation of slower cooling for thin hBN encapsulation could have important implications for photodetection applications. In addition, we predict that significantly slower cooling can be achieved using alternative layered dielectrics, such as MoS$_2$. Furthermore, the near-field coupling we studied in this work between hot graphene carriers and hyperbolic hBN phonons may pave the way to novel approaches in fields such as nanophotonics, ultrahigh resolution microscopy and nanoscale thermal management.

## Methods

*Extraction of the initial hot carrier temperature* — The measured photovoltage $V_{\mathrm{PTE}}$ scales with the *time-averaged* increase in electron temperature $\langle T_{\mathrm{e}} - T_{\mathrm{L}} \rangle$. In the case of cooling length smaller than the laser spot size (which is the case at room temperature) the relevant heat equation gives a simple linear scaling between peak and average increase in $T_{\mathrm{e}}$, governed by the interfacial heat conductivity $\Gamma$. We assume that $\Gamma$ and the Seebeck coefficient are constant with power [40], so that the photovoltage as a function of power directly represents the peak $T_{\mathrm{e}}$ increase as a function of power: $V_{\mathrm{PTE}} = a(T_{\mathrm{e}} - T_{\mathrm{L}})$. For undoped graphene, the peak $T_{\mathrm{e}}$ after illumination with laser power $P$ is $T_{\mathrm{e}} = \sqrt[3]{T_{\mathrm{L}}^3 + bP}$, where $b$ is a constant that depends on laser repetition rate $f$, absorption coefficient $\eta_{\mathrm{abs}}$, heating efficiency $\eta_{\mathrm{heat}}$ and spot size $L_{\mathrm{spot}}$. The cube root comes from the $T^2$-scaling of the electronic heat capacity for un-doped graphene (we apply this analysis to a low carrier density of $0.06 \times 10^{12}/\mathrm{cm}^2$). By fitting $V_{\mathrm{PTE}}$ as a function of $P$, we extract the constants $a$ and $b$, which allows us to recover the power-dependent peak $T_{\mathrm{e}}$ (see **Fig. 11**). We verify the obtained constant

$b$, and find good agreement when using an absorption of $\eta_{\mathrm{abs}} = 1\%$ (due to the layered dielectrics), a heating efficiency of $\eta_{\mathrm{heat}} = 80\%$ and a spot size of $L_{\mathrm{spot}} \sim 2\ \mu\mathrm{m}$.

*Numerical calculations* — Theoretical curves are obtained by numerically integrating Eq. (1) [31]. The non-local conductivity is connected to the density-density response function of graphene, $\chi_{nn}(k,\omega)$, by the formula $\sigma(\omega,k) = ie^2\omega\chi_{nn}(k,\omega)/k^2$. Within the random-phase approximation, we approximate $\chi_{nn}(k,\omega)$ with its non-interacting expression, which has been given in Refs. [41–44]. Neglecting retardation effects, the surface admittance is connected to the screened Coulomb interaction between two electrons in graphene, $V(k,\omega)$, by the formula $Y(\omega,k) = -ie^2\omega V^{-1}(k,\omega)/k^2$. The expression for $V(k,\omega)$ is given, for graphene embedded into slabs of hBN and in the presence of non-hyperbolic dielectrics and metallic gates, in Ref. [45].

## References

[1] Geim, A.K. & Grigorieva, I.V. Van der Waals heterostructures *Nature* **499**, 419-425 (2013)

[2] Dean, C.R. et al. Boron nitride substrates for high-quality graphene electronics. *Nature Nanotech* **5**, 722 (2010)

[3] Mayorov, A.S. et al. Micrometer-scale ballistic transport

in encapsulated graphene at room temperature. *Nano Lett.* **11**, 2396 (2011)

[4] Britnell, L et al. Electron tunneling through ultrathin boron nitride crystalline barriers. *Nano Lett.* **12**, 1707 (2012)

[5] Dean, C.R. et al. Hofstadter's butterfly and the fractal quantum Hall effect in moiré superlattices. *Nature* **497**, 598 (2013)

[6] Ferrari, A.C. et al., Science and technology roadmap for graphene, related two-dimensional crystals, and hybrid systems. *Nanoscale* **7**, 4598 (2015)

[7] Wang, L. et al. One-Dimensional Electrical Contact to a Two-Dimensional Material. *Science* **342**, 614-617 (2013)

[8] Balandin, A.A. et al. Superior Thermal Conductivity of Single-Layer Graphene. *Nano Lett.* **8**, 902–907 (2008)

[9] Britnell, L. et al. Strong Light-Matter Interactions in Heterostructures of Atomically Thin Films. *Science* **340** 1311–1314 (2013)

[10] Koppens, F.H.L. et al. Photodetectors based on graphene, other two-dimensional materials and hybrid systems. *Nature Nanotech.* **9**, 780793 (2014)

[11] Lopez-Sanchez, O. et al. Light Generation and Harvesting in a van der Waals Heterostructure. *ACS Nano* **8**, 3042 (2014)

[12] Bonaccorso, F. et al. Graphene, related two-dimensional crystals, and hybrid systems for energy conversion and storage. *Science* **347**, 1246501 (2015)

[13] Mics, Z et al. Thermodynamic picture of ultrafast charge transport in graphene. *Nature Comm.* **6**, 7655 (2015)

[14] Kim, Y.D. et al. Bright visible light emission from graphene. *Nature Nanotech.* **10**, 676-681 (2015)

[15] Massicotte, M. et al. Picosecond photoresponse in van der Waals heterostructures. *Nature Nanotech.* **11**, 42–46 (2016)

[16] Bonaccorso, F., Sun, Z., Hasan, T., Ferrari, A.C. Graphene photonics and optoelectronics. *Nat. Photonics* **4**, 611 (2010)

[17] Caldwell, J.D. et al. Sub-diffractional volume-confined polaritons in the natural hyperbolic material hexagonal boron nitride. *Nature Comm.* **5**, 5221 (2014)

[18] Dai, S. et al. Tunable phonon polaritons in atomically thin van der Waals crystals of boron nitride. *Science* **343** 1125 (2014)

[19] Basov, D.N., Fogler, M.M. & Garcia de Abajo, F.J. Polaritons in van der Waals materials. *Science* **354**, 1992 (2016)

[20] Gabor N.M. et al. Hot Carrier-Assisted Intrinsic Photoresponse in Graphene. *Science* **334**, 648-652 (2011)

[21] Song, J.C.W. et al. Hot Carrier Transport and Photocurrent Response in Graphene. *Nano Lett.* **11**, 4688-4692 (2011)

[22] 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)

[23] Mihnev, M.T. et al. Microscopic origins of the terahertz carrier relaxation and cooling dynamics in graphene. *Nature Comm.* **7**, 11617 (2016)

[24] D. Brida et al., Ultrafast collinear scattering and carrier multiplication in graphene. *Nature Comm.* **4**, 1987 (2013)

[25] Bistritzer, R. & MacDonald, A. H. Electronic cooling in graphene. *Phys. Rev. Lett.* **102**, 206410 (2009)

[26] 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)

[27] 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)

[28] Betz, A.C. et al. supercollision cooling in undoped graphene. *Nature Phys.* **9**, 109-112 (2013)

[29] Low, T. et al. Cooling of photoexcited carriers in graphene by internal and substrate phonons. *Phys. Rev. B* **86**, 045413 (2012)

[30] Hamm, J.M. et al. Nonequilibrium plasmon emission drives ultrafast carrier relaxation dynamics in photoexcited graphene. *Phys. Rev. B* **93**, 041408 (2016)

[31] Principi, A. et al. Super-Planckian electron cooling in a van der Waals stack. *Phys. Rev. Lett.* **118**, 126804 (2017)

[32] Banszerus, L. et al. Ultrahigh-mobility graphene devices from chemical vapor deposition on reusable copper. *Science Adv.* **1**, e1500222 (2015)

[33] Nyquist, H. Thermal Agitation of Electric Charge in Conductors. *Phys. Rev.* **32**, 110113 (1928)

[34] Mihnev, M.T. et al. Electronic cooling via interlayer Coulomb coupling in multilayer epitaxial graphene. *Nature Comm.* **6**, 8105 (2015)

[35] Tomadin, A, et al. Accessing phonon polaritons in hyperbolic crystals by angle-resolved photoemission spectroscopy. *Phys. Rev. Lett.* **115**, 087401 (2015)

[36] Giuliani, G.F. & Vignale G. Quantum Theory of the Electron Liquid (Cambridge University Press, Cambridge, 2005)

[37] Yang, W. et al. Hyperbolic cooling of a graphene on hBN transistor in the Zener-Klein regime. *Nature Nanotech. in press* (2017)

[38] Ma, Q. et al. Competing Channels for Hot-Electron Cooling in Graphene. *Phys. Rev. Lett.* **112**, 247401 (2014)

[39] Jadidi, M.M. et al. Infrared Nonlinear Photomixing Spectroscopy of Graphene Thermal Relaxation. *Phys. Rev. Lett.* **117**, 257401 (2016)

[40] Hwang, E.J., Rossi, E. & Das Sarma, S. Theory of thermopower in two-dimensional graphene. *Phys. Rev. B.* **80**, 235415 (2009)

[41] Wunsch, B. et al. Dynamical polarization of graphene at finite doping. *New J. Phys.* **8**, 318 (2006)

[42] Hwang, E.H. & Das Sarma, S. Dielectric function, screening, and plasmons in two-dimensional graphene. *Phys. Rev. B* **75**, 205418 (2007)

[43] Principi, A., Polini, M. & Vignale, G. Linear response of doped graphene sheets to vector potentials. *Phys. Rev. B* **80**, 075418 (2009)

[44] Kotov, V.N. et al. Electron-Electron Interactions in Graphene: Current Status and Perspectives. *Rev. Mod. Phys.* **84**, 1067 (2012)

[45] Alonso-González, P. et al. Acoustic terahertz graphene plasmons revealed by photocurrent nanoscopy *Nature Nanotech.* **2**, 3135 (2017)

[46] Das Sarma, S., Adam, S., Hwang, E. H. & Rossi, E. Electronic transport in two-dimensional graphene. *Rev. Mod. Phys.* **83**, 407 (2011)

[47] Hwang, E. H. & Das Sarma, S. Acoustic phonon scattering limited carrier mobility in two-dimensional extrinsic graphene. *Phys. Rev. B* **77**, 115449 (2008)

[48] 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)

[49] Graham, M.W. et al. Transient Absorption and Photocurrent Microscopy Show That Hot Electron Super-

collisions Describe the Rate-Limiting Relaxation Step in Graphene. *Nano Lett.* **13** 54975502 (2013)

[50] Crossno, J. et al. Development of high frequency and wide bandwidth Johnson noise thermometry. *App. Phys. Lett.* **106**, 023121 (2015)

[51] Viljas, J.K. & Heikkilä, T.T. Electron-phonon heat transfer in monolayer and bilayer graphene. *Phys. Rev. B* **81**, 245404 (2010)

[52] Amorim, B and Guinea, F. Flexural mode of graphene on a substrate. *Phys. Rev. B* **88**, 115418 (2013)

[53] R. V. Gorbachev. Hunting for monolayer boron nitride: Optical and raman signatures. *Small* **7**, 465 (2011)

[54] Casiraghi, C. et al. Rayleigh imaging of graphene and graphene layers. *Nano Letters* **7**, 2711 (2007)

[55] Ferrari, A.C. et al., The Raman Spectrum of Graphene and Graphene Layers. *Phys. Rev. Lett.* **97**, 187401 (2006)

[56] Ferrari, A.C., Basko, D.M. Raman spectroscopy as a versatile tool for studying the properties of graphene. *Nat. Nanotechnol.* **8**, 235 (2013)

[57] Arenal, R. et al. Raman Spectroscopy of Single-Wall Boron Nitride Nanotubes. *Nano Lett.* **6**, 1812 (2006)

[58] F. Pizzocchero et al. The hot pick-up technique for batch assembly of van der Waals heterostructures. *Nat. Commun.* **7**, 11894 (2016)

## Appendix 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 ultraclean, hBN-encapsulated devices giving ∼18–20 eV, assuming phonon-limited momentum scattering [2, 7]. The cooling dynamics of $SiO_2$-supported graphene are consistent with $D = 12$–18 eV, assuming disorder-assisted cooling [27]. 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 [46]. If momentum scattering is solely determined by electron-phonon interaction (process *iii*), we can use the measured graphene mobility $\mu$ at a given carrier density $n$ to obtain the deformation potential following Refs. [47, 48]: $\mu = \frac{4\hbar v_F^2 e\rho v_s^2}{\pi D^2 n k_B T}$, where $v_F$ is the Fermi velocity, $\rho$ 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 [46] 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) [26] and gives a cooling time that scales with $\frac{k_F \ell}{D^2}$, where $k_F$ is the Fermi momentum, $\ell$ the mean free path (limited by short-range scattering), and $D$ the electron-phonon deformation potential [27]. 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. [27] in calculating the experimental photovoltage dip $\Delta V_{\mathrm{PTE}}$. Since this model requires that $T_{\mathrm{el}} \ll T_F$, we apply this analysis to our data at $n = 1.7 \times 10^{12}/\mathrm{cm}^2$, where $T_F = \frac{E_F}{k_B} \approx 1800$ K. We numerically solve the energy dissipation rate $C_n \frac{dT_e}{dt} = -A(T_e^3 - T_L^3)$ with the heat capacity (for non-neutral graphene) $C_n = \alpha T_{\mathrm{el}}$ and $\frac{A}{\alpha} = 0.47 \frac{1}{k_F \ell} \frac{D^2}{\rho v_s^2} \frac{E_F}{(\hbar v_F)^2} \frac{k_B}{\hbar}$ to obtain the temperature dynamics $T_e(t)$. The photovoltage $V_{\mathrm{PTE}}$ at a delay time $\Delta t$ between two pulses follows from $V_{\mathrm{PTE}} = \int_0^{\Delta t} V(t, T_1)dt + \int_{\Delta t}^{\infty} V(t - \Delta t, T_2)dt$. Here we use the instantaneous photo-thermoelectric voltage $V(t, T) = BT(T - T_L)$, 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 $\sqrt{T_{\mathrm{in}}^2 + T_{\mathrm{add}}^2}$, with $T_{\mathrm{in}}$ the temperature before arrival of the laser pulse, and $T_{\mathrm{add}}$ the temperature equivalent to the added pulse energy. We obtain a photovoltage dip by repeating this calculation while varying $\Delta t$, and fit the data to extract the deformation potential, finding $D = 65$ eV (see **Fig. 9a**). We note that using Ref. [26] 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 [26]. 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 [27, 49]. Finally, noise thermometry studies on hBN-encapsulated graphene [50] and experimental-theoretical terahertz spectroscopy studies on multilayer epitaxial graphene and CVD graphene on different substrates[23] 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 [25]. 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 [25] $T_e(\Delta t) = \frac{T_e(0)}{\sqrt{\Delta t/\tau_0 + 1}}$, with $\tau_0 = \frac{424}{D^2[T_e(0)]^2}$, where $D$ is in eV, $T_e$ in meV and $\tau_0$ in $\mu$s. We compare these cooling dynamics with our experimental data in **Fig. 9b**, 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 [22, 24], but mainly for carriers with high energy, on the order of the optical phonon energy of $\sim$0.2 eV [51]. 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 **Fig. 8b**). 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. [23], 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) [23]; *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) [23], 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 [37]. Finally, we observe a significant effect of reducing the thickness of the hBN encapsulant (see **Fig. 10**), 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 [52]. *Wiedemann-Franz cooling* refers to lateral heat spreading out of the spot that is exited by the incident light. To assess the relevance of this process, we calculate the cooling length (at room temperature), which is given by $\zeta = \sqrt{\frac{\kappa}{\Gamma}}$ [21]. Here, $\Gamma$ is the interfacial heat conductivity and $\kappa$ is given by the Wiedemann-Franz law: $\kappa = \frac{\pi^2 k_{\mathrm{B}}^2 T \sigma}{e^2}$, where $\sigma$ is the graphene electrical conductivity and $e$ the electron charge. We find a cooling length of $\zeta \sim 1~\mu$m, which is smaller than our spot size of $L_{\mathrm{spot}} \sim 2~\mu$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 $\Gamma$ decreases, leading to longer cooling lengths. Therefore, at lower lattice temperatures, we do not exclude that Wiedeman-Franz cooling plays a role (see **Fig. 13**). Finally, we mention that *cooling by tunneling of hot carriers through the hBN slab* is irrelevant for our devices, since the hBN slab in 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 [4].

## Appendix 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 [53, 54] and optical contrast measurements [54] and Raman spectroscopy [55–57]. The target flakes are then picked up and assembled into the desired heterostructure using a hot-pick up technique, similar to Ref. [58], to minimize contamination containing blisters at the hBN/graphene interfaces. From atomic force microscopy (AFM) characterization (see **Fig. 10d**) we measure a step change in height$\sim$ 2nm and $\sim$ 3nm for the bottom and top hBN respectively. We note that part of this contribution (up to $\sim$ 1nm) may arise from a water layer present on the surface of the hBN [58], so that we estimate up to 7 layers in hBN encapsulants.
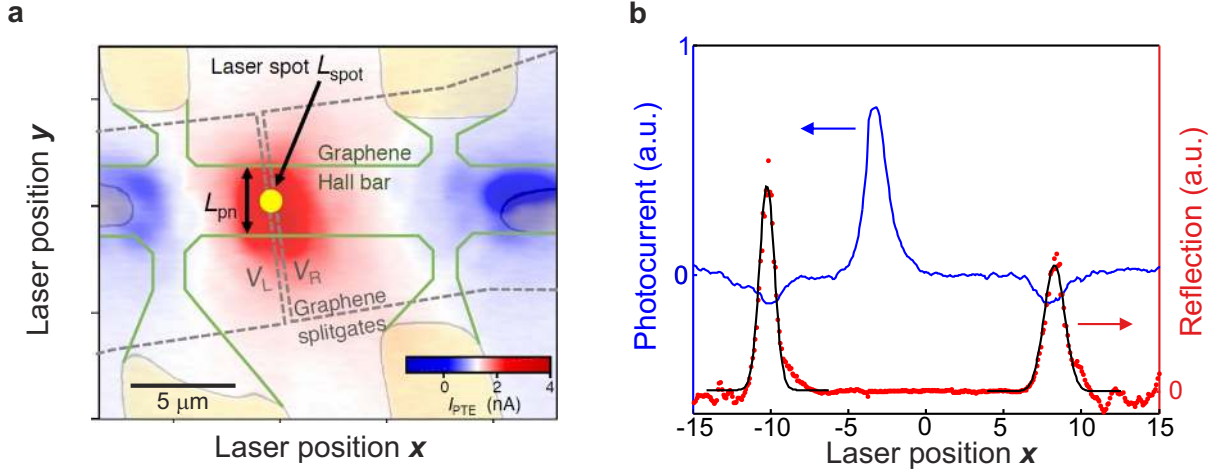
**Appendix 3: Additional figures**



**Fig. 5**: **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).
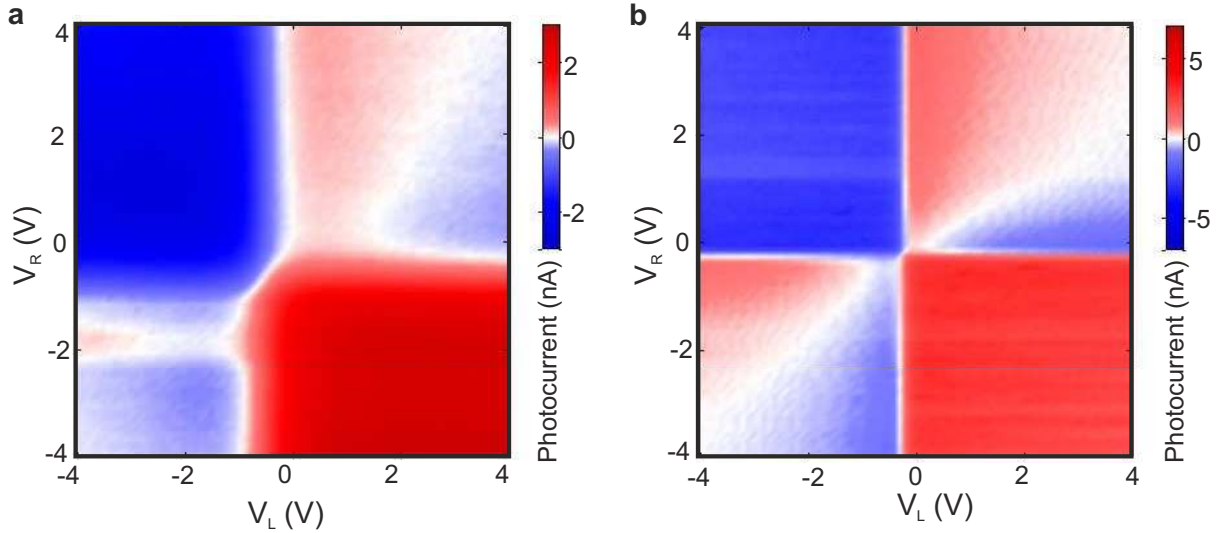


**Fig. 6**: **Characteristic sixfold pattern for photo-thermoelectric effect.** Photovoltage at the interface between the two graphene regions, for single-pulse excitation as a function of voltage on the two graphene split gates $V_L$ and $V_R$ at room temperature **(a)** and 30 K **(b)**. The sixfold patterns indicate photo-thermoelectric photovoltage generation, which scales with the light-induced increase in carrier temperature [20, 21].
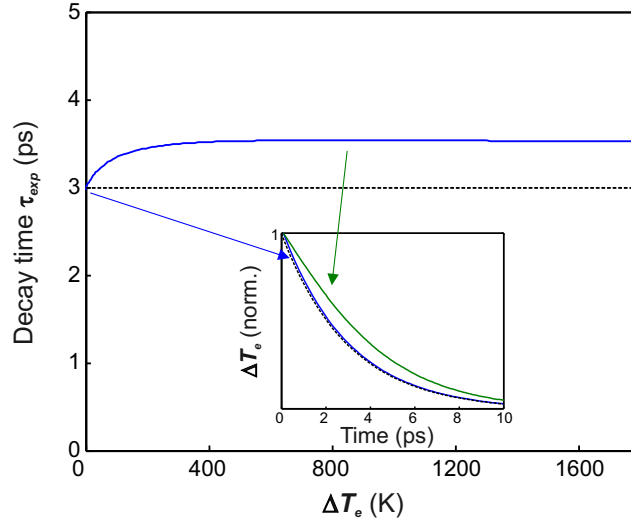
**Fig. 7**: **Effect of nonlinearity on measured dynamics.** To study the relation between the observed dynamics of the photovoltage dip and the actual cooling dynamics, we calculate the photovoltage dip dynamics for a given exponential cooling time of 3 ps. To this purpose, we integrate the generated 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. [27]). 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).
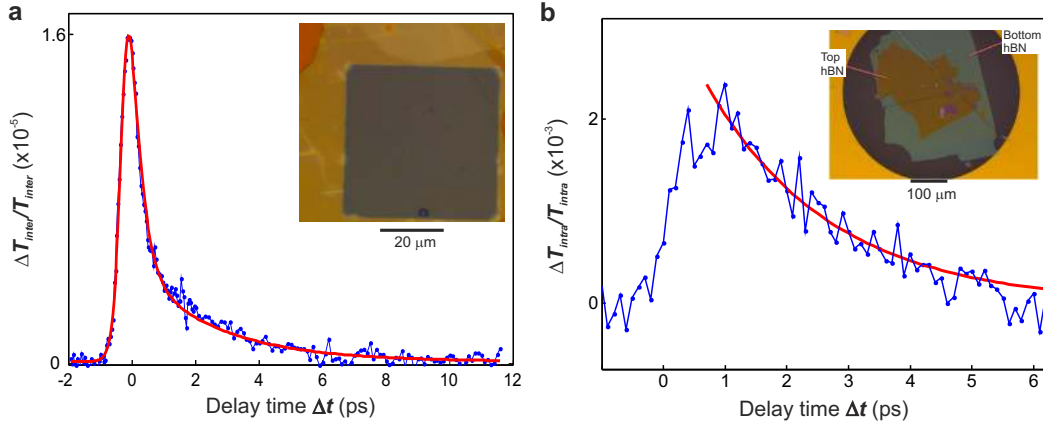


**Fig. 8**: **Verification of RT decay time using pump-probe spectroscopies. a)** Ultrafast optical pump – optical probe spectroscopy on hBN-encapsulated CVD graphene, prepared as in Ref. [32] and fully covering a 50 $\mu$m square aperture on a transparent SiO$_2$ substrate (see inset). The ultrashort pump pulses, with a wavelength of 780 nm, 100 fs-duration, and incident fluence of $\sim$10 $\mu$J/cm$^2$, creates a non-equilibrium distribution that quickly thermalizes through electron-electron scattering [24]. This modified, hot-carrier distribution affects available interband transitions through Pauli blocking, which we probe using ultrashort pulses at 1.3 $\mu$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 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 **Fig. 12**). **b)** The pump-induced change in transmitted terahertz (THz) signal as a function of pump-probe delay time $\Delta t$ using an hBN-encapsulated CVD graphene sandwich with dimensions >200 $\mu$m, prepared as in Ref. [32] and transferred inside a 500 $\mu$m round aperture (see inset). The loosely focused pump light, with a wavelength of 800 nm and pulse energy density of $\sim$ 20 $\mu$J/cm$^2$ creates hot carriers and a modified carrier distribution. Terahertz light with frequency 0.4 – 1.2 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$2.2 ps.
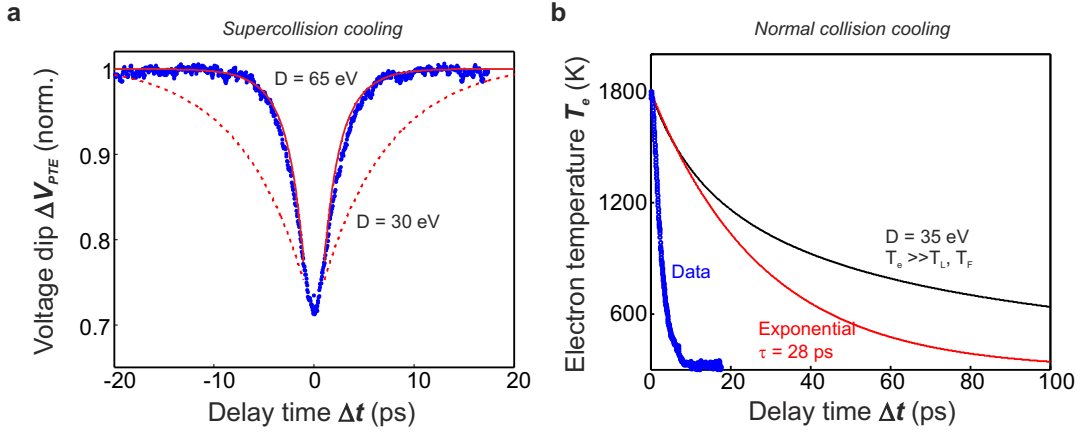
**Fig. 9**: **Comparing data with alternative cooling mechanisms.** The photovoltage dip dynamics ($T_L = 300$ K and $n = 1.7 \times 10^{12}/cm^2$, compared with the dynamics according to the supercollision cooling model with a deformation potential of 65 eV (solid red line) and 30 eV (dashed red line). The model takes into account that we are in the strong heating regime with $T_e = 1500$ K (see Methods).
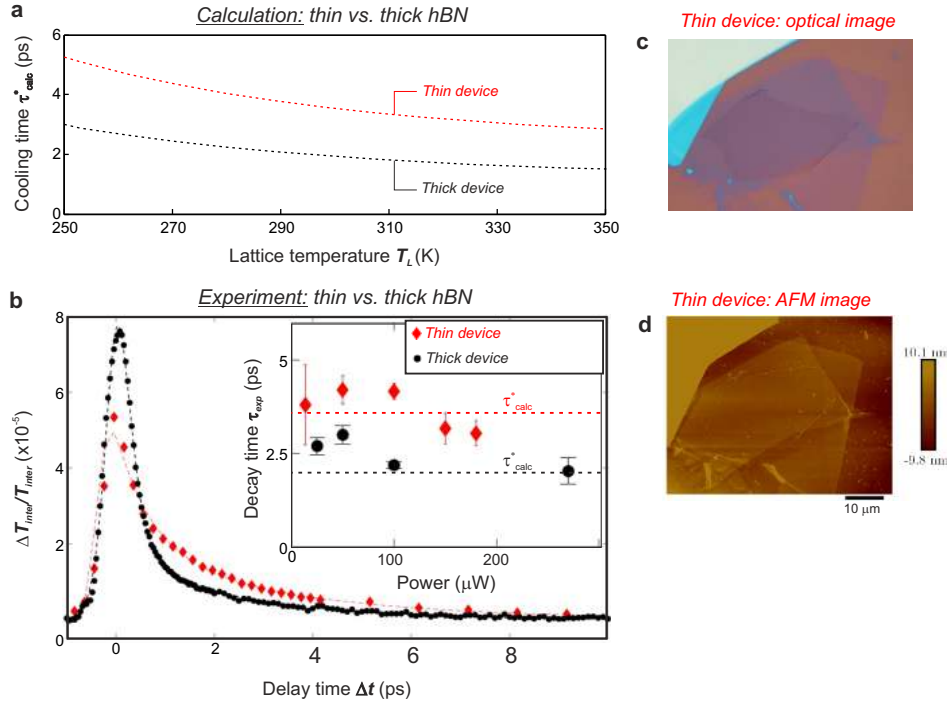


**Fig. 10**: **Dynamics for different hBN thicknesses. a)** Calculated (near-equilibrium) cooling time according to the hyperbolic cooling model $\tau^*_{calc}$ for a "thick" device with 30 nm bottom and 30 nm top hBN encapsulant (black, dashed line) and a "thin" device with 2 nm top and 3 nm bottom 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) $\mu$W. 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$^{-1}$.
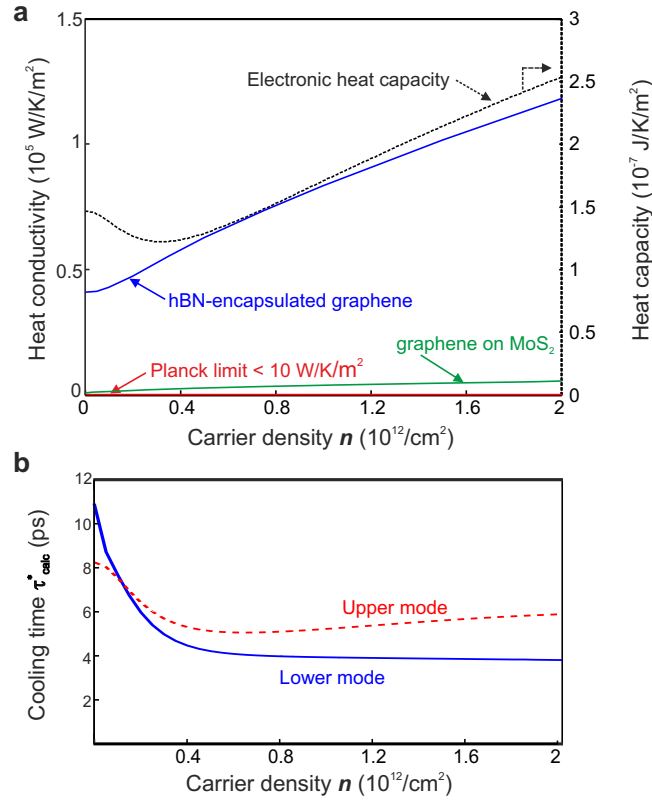
**Fig. 11**: **Interfacial heat conductivity, heat capacity and contributions of the hyperbolic modes. a)** Since the hyperbolic cooling time $\tau^*_{\text{calc}}$ is determined by the ratio of the interfacial heat conductivity $\Gamma$ and the electronic heat capacity $C_n$, we plot these separate entities as a function of carrier density $n$. We also show the interfacial heat conductivity for $\text{MoS}_2$ hyperbolic phonons, and for black-body radiation of graphene in vacuum, which is orders of magnitude lower than for hBN or $\text{MoS}_2$. **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.
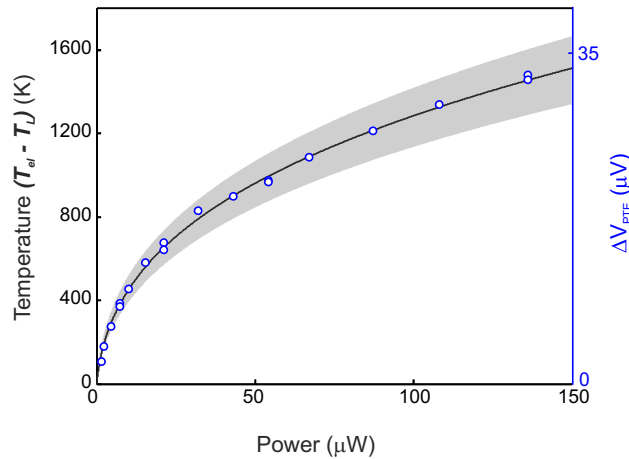


**Fig. 12**: **Estimation of hot carrier temperature.** The photovoltage as a function of laser power (right vertical axis) and the extracted laser-induced temperature increase (left vertical axis) for undoped graphene, using the analysis described in the Methods section. The grey shaded area gives the 95% confidence interval for the extracted electron temperature.
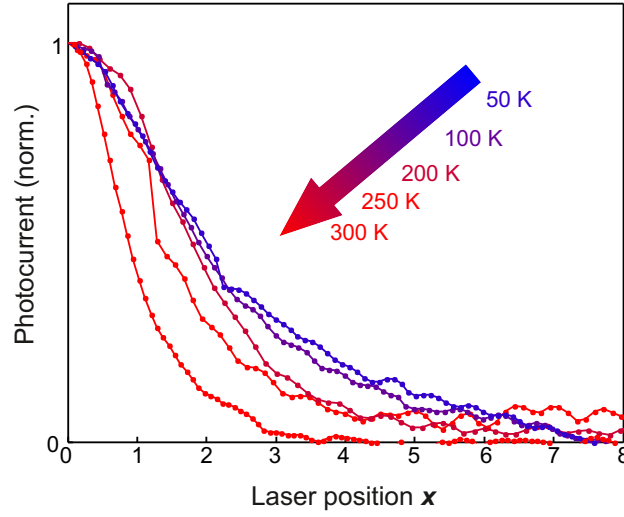
**Fig. 13**: **Spatial extent of photocurrent.** Spatial line traces of photocurrent vs. laser position for five different lattice temperatures. Position 0 corresponds to the $pn-$junction. At lower temperatures, the spatial extent of the photocurrent increases. This is caused by increased mechanical 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.
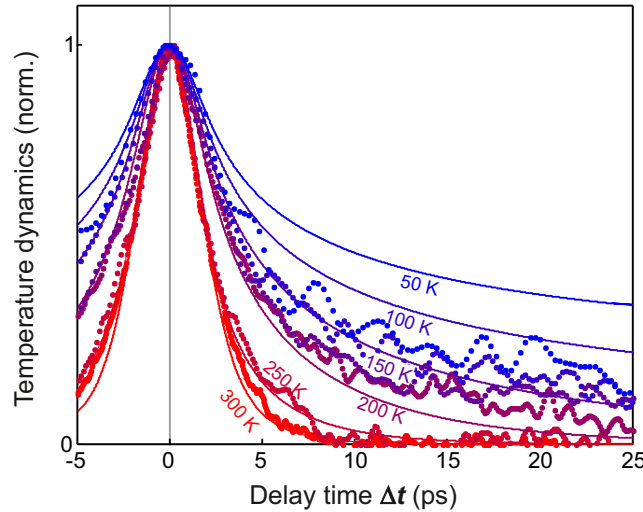


**Fig. 14**: **Cooling dynamics at high doping.** Comparison of the complete cooling dynamics as measured (data points) and calculated (solid line) for graphene with $n = 1.7$ and $2 \times 10^{12}/cm^2$, respectively, and varying lattice temperature. We use an initial temperature increase of $\Delta T = 300$–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 [25] and scattering with optical phonons [23] becoming more relevant cooling pathways.