MIT Open Access Articles

Energy flows in graphene: hot carrier dynamics and cooling

Citation: Song, Justin C W, and Leonid S Levitov. "Energy Flows in Graphene: Hot Carrier Dynamics and Cooling." J. Phys.: Condens. Matter 27, no. 16 (April 2, 2015): 164201.

As Published: http://dx.doi.org/10.1088/0953-8984/27/16/164201

Publisher: IOP Publishing

Persistent URL: http://hdl.handle.net/1721.1/98223

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Massachusetts Institute of Technology

Energy Flows in Graphene: Hot Carrier Dynamics and Cooling

Justin C. W. Song[1,2] and Leonid S. Levitov[1]

[1] *Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA and*
[2] *School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*

Long lifetimes of hot carriers can lead to qualitatively new types of responses in materials. The magnitude and time scales for these responses reflect the mechanisms governing energy flows. We examine the microscopics of two processes which are key for energy transport, focusing on the unusual behavior arising due to graphene's unique combination of material properties. One is hot carrier generation in its photoexcitation dynamics, where hot carriers multiply through an Auger type carrier-carrier scattering cascade. The hot-carrier generation manifests itself through elevated electronic temperatures which can be accessed in a variety of ways, in particular optical conductivity measurements. Another process of high interest is electron-lattice cooling. We survey different cooling pathways and discuss the cooling bottleneck arising for the momentum-conserving electron-phonon scattering pathway. We show how this bottleneck can be relieved by higher-order collisions - supercollisions - and examine the variety of supercollision processes that can occur in graphene.

I. INTRODUCTION

Every so often we encounter things that neatly combine beauty and utility. In a fairy tale a much-admired flower—"the sacred lotus of Hindostan"—turns out to be a common flower from the kitchen-garden, an artichoke.[1] Likewise, the subject of energy flows in materials is aesthetically pleasing and, at the same time, harbors practical opportunities. Understanding energy transport mechanisms is of keen interest for designing new approaches to handle, convert, and utilize energy in a bid to address key technological challenges. One area of high current interest which may benefit from this research is IT hardware, in particular finding ways to circumvent the saturation of operating frequencies in integrated electronics due to the large amounts of power dissipated in microprocessors.[2] Another such area is the development of efficient solar cells, where the relation with hot carriers stems from the Shockley-Queisser limit that sets an upper bound for conversion efficiencies in single-junction solar cells.[3] Currently high expectations are pinned on two-dimensional (2D) materials,[4] such as graphene and the atomically thin dichalcogenides, which possess a number of potentially useful properties.

Energy-related phenomena in graphene span a wide range of energies, from optical down to THz. The high optical activity of 2D materials, which can absorb an order of magnitude more sunlight than Si layers of similar thickness,[5] is of interest for optoelectronics research.[6] Additionally, the 2D structure renders electronic states fully exposed, allowing carriers and also heat to be extracted via a vertical transfer process (e.g. in a sandwich-type structure[7]). Another unique aspect of graphene is the ease with which electrons can be pushed out of thermal equilibrium with the lattice. In such a hot-carrier regime, system states with elevated electronic temperatures different from that of the lattice can be fairly long-lived, resulting in electron energy transport decoupled from that of the lattice. The strong thermoelectric response of graphene[8,9] generates strong coupling between

energy modes and charge modes, giving rise to a range of novel transport and optoelectronic phenomena.[10–17] Strikingly, hot-carrier effects in graphene exist in a wide range of technologically relevant temperatures including room temperature. Combined with its fast electrical response, this makes graphene an attractive material for high speed and gate tunable manipulation of energy flows on the nanoscale.

The hot carrier regime in graphene mainly stems from anomalously slow electron-lattice relaxation.[18,19] Strong carbon-carbon bonds, which give graphene's lattice its rigidity, also result in a high optical phonon frequency, $\omega_0 = 200\,\mathrm{meV}$. The large value of ω_0 suppresses the optical phonon contribution to electron-lattice relaxation below a few hundred kelvin. At the same time, the generally weak scattering between electrons and long-wavelength acoustic phonons is further constrained by the large mismatch in Fermi velocity and sound velocity, $v/s \approx 100$. As a result, once the electrons are heated up they stay out of thermal equilibrium with the lattice over long times, proliferating over extended spatial lengthscales.[18,19] Hot carriers have been studied in a variety of other systems including semiconductors like GaAs,[20–22] and metals.[23] However in these other materials, hot carriers only exist at very low temperatures or under intense pumping. In contrast, hot carriers in graphene can exist even at room temperature and under weak driving.[10]

The mechanisms responsible for the generation and cooling of hot carriers, which we discuss below, are characterized by vastly different time scales. For the carrier-carrier scattering processes occurring in the cascade triggered by photoexcitation, the times can be as fast as tens of femtoseconds.[24,25] These fast times determine the branching ratio for the electron and phonon scattering pathways, controlling the energy part captured by the electronic system when energy is pumped into graphene.[13] The resulting hot-carrier state is manifested through an elevated electronic temperature.[10,12,13,24–27] Once a hot carrier distribution is established, slower processes (up to hundreds of picoseconds), such as phonon emission, determine how long the carriers can stay

hot. Both processes, fast scattering and slow cooling, contribute to the performance of optoelectronic devices.[10,13,26]

Graphene's large interaction parameter, $\alpha \approx 2.2$, its gapless spectrum and tunable carrier density make it an ideal venue to probe the role of Auger type processes in the energy relaxation of high energy photoexcited carriers cascading down to lower energies. The large difference in the time scales for hot carrier generation and cooling allows us to treat these processes separately. Notably, as discussed below, graphene affords convenient knobs for controling both generation and cooling rates. These knobs provide means to manipulate the energy flows, underpining current attempts to exploit graphene as a future energy material.[10,13]

A topically relevant context for studying energy flows in graphene is photoexcitation dynamics,[13,24,25,28–38] which is sensitive to both the fast scattering processes of hot carrier generation and the slow electron-lattice cooling processes. Photoexcitation has been the main technique for probing hot carriers in graphene[10,12,13,24–26] yielding high electronic temperatures under intense irradiation.[34]

In this article, we pedagogically lay out how energy flows through graphene electrons. To this end, we utilize the lens of hot carriers as a transparent way of addressing electronic energy flow in graphene. As we will see, this allows for a unified and multi-timescale treatment of photoexcitation that spans three orders of magnitude. In doing so, we survey relevant concepts introduced in the field, as well as supplement gaps with new results. Many of the processes described herein work in synergy with each other and make graphene a favorable candidate for new optoelectronic devices. Indeed, utilizing some of the concepts reviewed in this article for efficient optoelectronics, including energy harvesting devices and photodetectors,[10,15,16] has become an emerging field. Here, we do not describe these engineering efforts in detail but instead focus on the physics of the fundamental processes.

The article is organized in two parts, discussing energy relaxation pathways at high and low energy (Sec. II and Sec. III, respectively). As illustrated in Fig. 1, this delineation also appropriately characterizes the time scales involved in the photoexcited carriers relaxation kinetics.

In Sec. II, we consider energy relaxation of (photoexcited) carriers at high energy. Focusing on the case of *doped* graphene, we detail the competition between carrier-carrier scattering (within a single band) and optical phonon emission. As we argue, the relatively high carrier density of doped graphene and large phase space available for carrier-carrier scattering allow it to dominate over optical phonon emission in the relaxation of high energy carriers. This manifests in a thermalized electron state that is characterized by an electronic temperature which is elevated above the lattice temperature. In this hot carrier regime, the carrier densities within each band remain unchanged; energy from the high energy electrons is captured by the ambient carriers as electronic heat.

In Sec. III, we deal with thermal equilibration of the residual energy captured by ambient carriers. The relevant processes - electron-lattice cooling - occur at low energies, relaxing electrons close to the Fermi surface; we discuss these processes near equilibrium, giving estimates of the cooling power for different pathways. The generally low efficiency of these pathways leads to an interesting competition problem. As we will see, while first-order processes such as single optical and acoustic phonon emission are ineffective, there are other processes – *supercollisions* – that can relieve the cooling bottleneck. We discuss disorder-assisted supercollisions,[39] recently seen by a number of experimental groups,[26,27,40,41] as well as other forms of supercollisions and their physical manifestations. These include *few-body* scattering off ripples, flexural phonons, and pairs of counterpropagating phonons.

II. HOT CARRIER GENERATION AND THE PHOTOEXCITATION CASCADE

We begin by examining the photoexcitation cascade in graphene with emphasis on how energy flows and gets partitioned between different degrees of freedom. As we will see, the generation of hot carriers is an important process contributing to energy relaxation after initial photoexcitation. Here we will concentrate on the experimentally relevant case of doped graphene.

Photoexcitation in doped graphene proceeds as depicted in Fig. 1: photons with energy $hf > 2\mu$ create high-energy electron-hole pairs which form an outlying distribution of carriers high above the Fermi level (small red peak in Fig. 1 left panel); μ is the chemical potential.[24,25,31] A similar peak is formed by holes in the negative-energy band. The high-energy carriers then relax, losing energy to phonons or scattering with ambient carriers.[13,24,25,28–31,33–38,43] The amount of energy captured by the electronic system in this fast thermalization process depends on the competition between the rates for these pathways, which we will discuss below. After thermalization, a hot carrier distribution is formed with an elevated electronic temperature (Fig. 1 middle panel). This hot carrier distribution subsequently cools down by the emission of acoustic phonons to the lattice.[18,19,26,27,39,40] In this section, we discuss the thermalization process (hot carrier generation). In the next section we discuss how the hot carrier distribution cools with the lattice.

A central question in the thermalization cascade is how the energy of photoexcited carriers is partitioned between the electron and lattice degrees of freedom. There are two primary types of energy relaxation pathways: (i) carrier-carrier scattering[13,30,35,38,44] and (ii) optical phonon emission.[28,29,36] Only the processes of type (i) help to produce hot carriers, since the energy lost to the lattice degrees of freedom has very little impact on
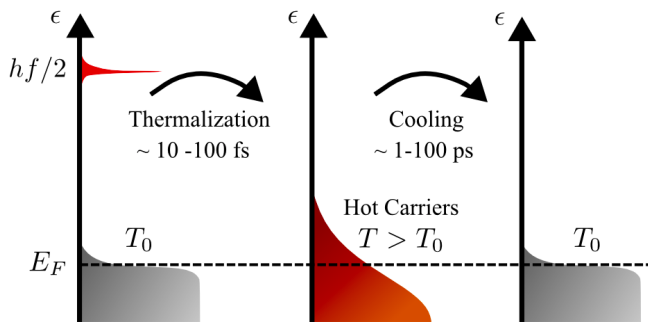
FIG. 1. Main stages of energy relaxation of photoexcited carriers. In the first stage, the carriers cascade from energy $\epsilon = hf/2$ down to low energies, losing energy via Auger processes and phonon emission. These processes lead to fast thermalization over timescales on the order tens to hundreds of femtoseconds, producing a relatively long-lived hot carrier distribution (middle panel). In the second stage, electron-lattice cooling mediated by acoustic phonons takes place over longer time scales (several to a hundred picoseconds) relaxing the hot carrier distribution back to equilibrium ($T = T_0$, right panel).

electron temperature (the lattice specific heat can be $10^3 - 10^4$ times larger than electron contribution, hence the energy transferred directly to electron is much more effective in producing hot carriers). The rates for these pathways analyzed in recent literature were found to be quite fast (tens to hundreds of fs), lying in the same ballpark for processes of type (i) and (ii).[44] The competition of these pathways determines the amount of energy that the ambient distribution captures from the initially photoexcited high-energy carriers.

Type (i) processes, which are key for various graphene-based energy harvesting proposals, have recently received a lot of attention.[13,30,35,38,44] Microscopically, they can be understood as Auger type processes in which carrier-carrier scattering between photoexcited high energy carriers scatter with low energy ambient carriers, see Fig. 2. The technological promise of graphene energy harvesting largely relies on the high effectiveness of these processes in passing energy among the high- and low-energy carriers. To avoid confusion, we differentiate the Auger processes into two distinct classes namely: a) intra-band carrier-carrier scattering (also called Impact Excitation, Auger Heating, see Fig. 2a) and b) inter-band carrier-carrier scattering (also called Impact Ionization, Carrier Multiplication, see Fig. 2b). Importantly, while intra-band carrier-carrier scattering (Fig. 2a) conserves the number of carriers in each band, inter-band carrier-carrier scattering (Fig. 2b) does not. However, as discussed below, inter-band carrier-carrier scattering is blocked by kinematic constraints arising due to energy and momentum conservation, rendering intra-band scattering the dominant mechanism for hot carrier production (Fig. 1 middle panel). In addition to scattering processes a) and b), there also exist electron-hole recom-

bination processes arising from carrier-carrier scattering (Auger recombination). These processes were analyzed in Refs. 30, 35, 38, and 42; here we will concentrate on a) and b).

We begin with the Hamiltonian for graphene $N = 4$ species of massless Dirac carriers,

$$\mathcal{H} = \sum_{\mathbf{k},i} \psi_{\mathbf{k},i}^\dagger H_0 \psi_{\mathbf{k},i} + \mathcal{H}_{\text{el}-\text{el}} + \mathcal{H}_{\text{el}-\text{ph}} + \mathcal{H}_{\text{ph}}, \quad (1)$$

$$\mathcal{H}_{\text{el}-\text{el}} = \frac{1}{2} \sum_{\mathbf{q},\mathbf{k},\mathbf{k}',i,j} V(\mathbf{q}) \psi_{\mathbf{k}+\mathbf{q},i}^\dagger \psi_{\mathbf{k}'-\mathbf{q},j}^\dagger \psi_{\mathbf{k}',j} \psi_{\mathbf{k},i}.$$

Here $\psi_{\mathbf{k},i}$ $\psi_{\mathbf{k},i}^\dagger$ describe two-component (pseudo)spin states, $i, j = 1...N$ label valley and spin degrees of freedom. The term $H_0 = v\sigma \cdot \mathbf{k}$ describes graphene's Dirac spectrum, $V(\mathbf{q}) = 2\pi e^2/|\mathbf{q}|\kappa$ is the Coulomb interaction, and κ is the bare dielectric constant. The terms $\mathcal{H}_{\text{ph}}$ and $\mathcal{H}_{\text{el}-\text{ph}}$ describing phonons in graphene and the electron phonon interaction will be specified below. In the above Hamiltonian, Eq.(1), we ignored several effects some of which are small and some of which will be introduced later as needed. The small effects are intervalley scattering terms in the carrier-carrier interaction and Umklapp scattering. The former is small because it originates from short range Hubbard-type interactions, which are smaller than the long range $V(\mathbf{q})$ interaction. Umklapp scattering in graphene can only occur due to three-particle collisions. This is so because, due to graphene lattice symmetry, vectors $\mathbf{K}$, $\mathbf{K}'$ and $\mathbf{K} - \mathbf{K}'$ are 3 times smaller than the reciprocal lattice vectors. In the single-particle Hamiltonian H_0 we ignored effects such as coupling to disorder and trigonal warping of the linear Dirac spectrum. Each of these effects will be discussed in due time.

A. Carrier-Carrier Scattering: Kinematic Constraints

We first consider Auger-type processes depicted in Fig. 2. Both these processes involves a photo-excited carrier with high energy and momentum, $\epsilon_{\mathbf{k_1}} \gg \mu$, $|\mathbf{k_1}| \gg k_F$, which is scattered to a lower energy state of momentum $\mathbf{k}_1'$ with recoil momentum $\mathbf{q} = \mathbf{k_1} - \mathbf{k}_1'$ passed to an electron in the Fermi sea. Here k_F is the Fermi wavevector. The latter particle-hole pair excitation process is depicted by a transition from $\mathbf{k_2}$ to $\mathbf{k}_2'$ in Fig. 2. The transition rate for this process, evaluated by the Golden Rule approach, takes the form[44,45]

$$W_{\mathbf{k}_1',\mathbf{k_1}} = \frac{2\pi N}{\hbar} \sum_{\mathbf{q},\mathbf{k_2},\mathbf{k}_2'} f_{\mathbf{k_2}}(1 - f_{\mathbf{k}_2'}) F_{\mathbf{k_2},\mathbf{k}_2'} |\tilde{V}_{\mathbf{q}}|^2 \quad (2)$$

$$\times \delta_{\mathbf{k}_1',\mathbf{k_1}+\mathbf{q}} \delta_{\mathbf{k}_2',\mathbf{k_2}-\mathbf{q}} \delta(\epsilon_{\mathbf{k}_1'} - \epsilon_{\mathbf{k_1}} + \epsilon_{\mathbf{k}_2'} - \epsilon_{\mathbf{k_2}}).$$

Here $f_{\mathbf{k}}$ is a Fermi function, and $F_{\mathbf{k},\mathbf{k}'} = |\langle \mathbf{k}'s'|\mathbf{k}s\rangle|^2$ is the coherence factor ($s, s' = \pm$ label states in the electron and hole Dirac cones). The effective Coulomb interaction $\tilde{V}$ which mediates scattering between the photo-excited

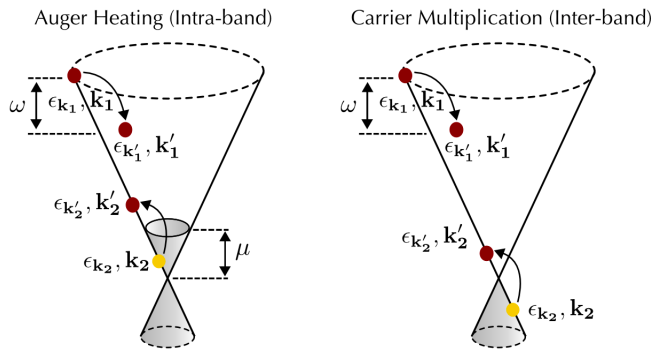**Auger Heating (Intra-band)**   **Carrier Multiplication (Inter-band)**

FIG. 2. Types of carrier-carrier scattering in graphene: a) Intra-band processes, and b) Inter-band processes. Processes a) may change electron temperature (hot carrier generation) but cannot change the number of carriers in each band. Carrier multiplication may only occur due to processes b) which change the carrier number in a band. The transition rate, described by Eq.(2), obeys kinematic constraints due to energy and momentum conservation. Because of these constraints, processes b) can only occur when transitions are collinear, $\omega = v|\mathbf{q}|$, resulting in a severely constrained phase space effectively blocking b) for two-body collisions (see text).

carrier and the carriers in the Fermi sea is taken in the form

$$\tilde{V}_{\mathbf{q}} = \frac{V_{\mathbf{q}}^0}{\varepsilon(\omega, \mathbf{q})}, \quad \varepsilon(\omega, \mathbf{q}) = 1 - V_{\mathbf{q}}^0 \Pi(\mathbf{q}, \omega). \qquad (3)$$

Here $V_{\mathbf{q}}^0 = 2\pi e^2/|\mathbf{q}|\kappa$, and the "permittivity" $\varepsilon(\omega, \mathbf{q})$ accounts for dynamical screening. This random-phase approximation (RPA) model uses the polarization operator $\Pi(\mathbf{q}, \omega) = N \sum_{\mathbf{k},s,s'} F_{\mathbf{k},\mathbf{k}+\mathbf{q};ss'} \frac{f(\epsilon_{\mathbf{k},s}) - f(\epsilon_{\mathbf{k}+\mathbf{q},s'})}{\omega + \epsilon_{\mathbf{k},s} - \epsilon_{\mathbf{k}+\mathbf{q},s'} + i0}$, with the band indices $\{s, s'\} = \pm$. This includes both intra- $(s = s')$ and inter- $(s \neq s')$ band contributions.[46]

Importantly, the electronic transitions in a massless Dirac band governed by the Hamiltonian (1) are subject to *kinematic constraints*.[30,47] These constraints arise due to the combined effect of linear dispersion in two Dirac cones, $E_{\pm}(\mathbf{p}) = \pm v|\mathbf{p}|$, and the momentum conserving character of carrier scattering. Here we analyze the simplest case of a two-body collision. Each of the two particles participating in a collision can make transitions between states in the upper and lower Dirac cones which we denote by $+$ and $-$. Two kinds of transitions can be distinguished: intra-band transitions $(+ \to + \text{ or } - \to -)$, see eg. Fig. 2a yellow to red circles, and inter-band transitions $(+ \to - \text{ or } - \to +)$, see eg. Fig. 2b yellow to red circles. Since momentum change in any transition satisfies $||\mathbf{k}| - |\mathbf{k}'|| < |\mathbf{k} - \mathbf{k}'| < |\mathbf{k}| + |\mathbf{k}'|$, intra-band transitions can only occur when the energy and momentum change are related by $|\Delta\epsilon| \leq v|\Delta k|$, whereas inter-band transitions are possible only when $|\Delta\epsilon| \geq v|\Delta k|$. Here $\mathbf{k}$ and $\mathbf{k}'$ denote initial and final momentum for a single transition; $\mathbf{k} \to \mathbf{k}'$ can refer to either of the transitions $\mathbf{k}_1 \to \mathbf{k}'_1$ or $\mathbf{k}_2 \to \mathbf{k}'_2$ as shown in Fig. 2.

For Eq.(2) to give a non-vanishing result, the tran-

sitions $\mathbf{k}_1 \to \mathbf{k}'_1$, $\mathbf{k}_2 \to \mathbf{k}'_2$ must occur in like pairs, i.e. both transitions are intra-band or both are inter-band. Since the transition $\mathbf{k}_1 \to \mathbf{k}'_1$ is restricted to be within a single band, the transition $\mathbf{k}_2 \to \mathbf{k}'_2$ must also be intra-band. While inter-band carrier-carrier scattering in Fig. 2b can technically occur when the energy and momentum exchanged are collinear $\omega = v|\mathbf{q}|$, the vanishing phase space for these transitions effectively block inter-band carrier-carrier scattering. As a result, intra-band carrier-carrier scattering in Fig. 2a are expected to play the dominant role in Auger-type relaxation in doped graphene.[13,24,25,44]

Kinematical blocking of inter-band processes can in principle be relieved by three-body (or, higher-order) collisions. Indeed, since the constraints arise from simultaneous momentum and energy conservation, relaxation of momentum conservation for example via coupling to disorder or high order processes in which multiple pairs are created at the same time provide a viable route to unblocking inter-band carrier-carrier scattering. Such processes may become important at high excitation power, however we expect the effect of such processes to be weak in the low excitation power regime discussed below.

Finding pathways in which interband processes are allowed has been the subject of recent research; several authors have investigated inter-band carrier-carrier scattering both theoretically and experimentally.[24,25,35,38,43,48] Early theoretical work in Ref. 35 and 38 that simulated the early thermalization dynamics of photoexcited carriers through a density matrix formalism and Bloch equations suggested that inter-band processes were possible. In particular, Ref. 35 and 38 predicted that Carrier Multiplication events dominated over Auger Recombination particularly under high excitation power; this in part spurred much of the current interest in Auger processes in graphene.[6] Later, other theories suggested that inter-band carrier-carrier scattering events are allowed when electron lifetime effects,[48] or when trigonal warping was taken into account.[49] On the experimental end, Angle Resolved Photoemission Spectroscopy (ARPES)[24,25] and optical pump THz probe[43] experiments have been used to search for carrier multiplication events. However, differences in techniques and results have left the current status of interaction mediated relaxation in undoped graphene hotly contested.[24,25,43] Since the intraband process in Fig. 2a vanishes in undoped graphene, a fuller understanding of the high excitation power regime, how higher order collisions affect interband scattering, and how these Auger processes compete with phonon emission is needed for a complete picture of the photoexcitation cascade in undoped graphene.

Below, we focus on the technologically relevant case of doped graphene. In this case, intra-band carrier scattering is both kinematically allowed and can provide fast relaxation of photoexcited carriers.[44] Importantly, the intra-band scattering processes allow the ambient carrier distribution to capture the energy of photoexcited carriers as electronic heat, elevating the electronic tem-

perature and giving rise to the hot carrier regime (middle panel, Fig. 1). In doped graphene, energy captured via hot carriers becomes the most relevant quantity for evaluating the efficiency of graphene optoelectronics since strong thermoelectricity in graphene[8,9] allow hot carriers to drive optoelectronic circuits, and dominates its photocurrent response.[10]

B. Intra-band Carrier-Carrier Scattering

The energy relaxation rate of a photo-excited carrier via intra-band carrier-carrier scattering is[44]

$$\mathcal{J}_{\text{el}} = \sum_{\mathbf{k_1'}}(\epsilon_{\mathbf{k_1'}} - \epsilon_{\mathbf{k_1}})W_{\mathbf{k_1'},\mathbf{k_1}}(1 - f_{\mathbf{k_1'}})F_{\mathbf{k_1},\mathbf{k_1'}} \qquad (4)$$

As shown in Fig. 2a, the typical energy of an excited pair is much smaller than the photo-excitation energy $\epsilon_{\mathbf{k_1}}$. As a result, it is convenient to factorize the transition rate through the spectrum of secondary pair excitations using the secondary pair susceptibility, $\chi''(\mathbf{q},\omega) = N\sum_{\mathbf{k}} F_{\mathbf{k},\mathbf{k+q}}(f_{\mathbf{k}} - f_{\mathbf{k+q}})\delta(\epsilon_{\mathbf{k+q}} - \epsilon_{\mathbf{k}} - \omega) = -\frac{1}{\pi}\text{Im}\Pi_+(\mathbf{q},\omega)$; for the intra-band process discussed below $\Pi_+(\mathbf{q},\omega)$ refers to polarization from intra-band scattering only. Following this standard procedure, we obtain

$$\mathcal{J}_{\text{el}}(\epsilon_{\mathbf{k_1}})= \int_{-\infty}^{\infty}d\omega\,\omega P(\omega), \qquad (5)$$

$$P(\omega)= A\sum_{\mathbf{q}}|\tilde{V}_{\mathbf{q}}|^2 F_{\mathbf{k_1},\mathbf{k_1'}}\chi''(\mathbf{q},\omega)\delta(\epsilon_{\mathbf{k_1'}} - \epsilon_{\mathbf{k_1}} + \omega) \quad (6)$$

where $A = \frac{2\pi}{\hbar}[N(\omega)+1)][1-f(\epsilon_{\mathbf{k}} - \omega)]$ and $\mathbf{k_1'} = \mathbf{k_1} - \mathbf{q}$, and ω denotes the amount of energy transferred in each scattering event, and $P(\omega)$ is the transition probability for scattering.

Interestingly, $P(\omega)$ depends very weakly on the initial energy of the original photoexcited pair, since $\epsilon_{\mathbf{k}}$ only enters in the Fermi function in A. Further, numerically evaluating Eq.(6), Ref. 44 found a non-monotonic $P(\omega)$ peaking close to $\omega \approx \mu$, and decaying rapidly for $\omega \gg \mu$.[44] As a result, for energies $\epsilon_{\mathbf{k_1}} \gg \mu$, the energy relaxation of a high energy carrier dominated by carrier-carrier scattering proceeds in steps of μ.

Large values of $\mathcal{J}_{\text{el}}$, as large as several eV/ps, have been estimated for typical dopings in graphene.[44] The efficiency of intraband carrier-carrier scattering noted above can be linked to the large values of μ in graphene. Indeed, the relation between efficiency and μ can be clarified by simple dimensional analysis. The weak dependence of $P(\omega)$ on initial carrier energy (see above) means that it depends on ω essentially via the dimensionless parameter $x = \omega/\mu$. As a result, $\mathcal{J}_{\text{el}}$ can be described by

$$\mathcal{J}_{\text{el}}(\epsilon) = \frac{\mu^2}{\hbar}\int_0^{\epsilon/\mu}x\tilde{P}(x)dx, \qquad (7)$$

where $\tilde{P}(x) = \hbar P(\Delta\epsilon)$ is dimensionless. Efficient $\mathcal{J}_{\text{el}}$ arises from the fast $\Gamma \sim \mu/(2\pi\hbar) \approx 20\,\text{ps}^{-1}$ (for typical

doping of $\mu = 0.1\,\text{eV}$) allowed from unitarity; the large density of carriers available within the band in doped graphene provides a large phase space for fast intraband carrier-carrier scattering.

C. Optical Phonon Emission

An alternative channel for energy relaxation of photo-excited carriers occurs through the emission of optical phonons and gives an energy relaxation rate of $\mathcal{J}_{\text{ph}}$. The transition rate of this process[45] can be described by Fermi's golden rule

$$W_{\mathbf{k'},\mathbf{k}}^{\text{el}-\text{ph}} = \frac{2\pi N}{\hbar}\sum_{\mathbf{q}}|M(\mathbf{k'},\mathbf{k})|^2 \qquad (8)$$
$$\delta(\Delta\epsilon_{\mathbf{k'},\mathbf{k}} + \omega_{\mathbf{q}})\delta_{\mathbf{k'},\mathbf{k+q}}(N(\omega_{\mathbf{q}}) + 1),$$

where $\Delta\epsilon_{\mathbf{k'},\mathbf{k}} = \epsilon_{\mathbf{k'}} - \epsilon_{\mathbf{k}}$, $\omega_{\mathbf{q}} = \omega_0 = 200\,\text{meV}$ is the optical phonon dispersion relation, and $N(\omega_{\mathbf{q}})$ is a Bose function. Here $\mathbf{k}$ is the initial momentum of the photo-excited electron, $\mathbf{k'}$ is the momentum it gets scattered into, and $\mathbf{q}$ is the momentum of the optical phonon. The electron-phonon matrix element $M(\mathbf{k'},\mathbf{k})$ is[18,19,50]

$$|M(\mathbf{k'},\mathbf{k})|^2 = g_0^2 F_{\mathbf{k},\mathbf{k'}}, \quad g_0 = \frac{2\hbar^2 v}{\sqrt{\rho\omega_0 a^4}}, \qquad (9)$$

where $F_{\mathbf{k},\mathbf{k'}}$ is the coherence factor for graphene, g_0 is the electron-optical phonon coupling constant[86], ρ is graphene's mass density, and a is the distance between nearest neighbor carbon atoms. In the same fashion as above, the energy-loss rate of the photo-excited carrier at energy ϵ due to the emission of an optical phonon can be evaluated as $\mathcal{J}_{\text{ph}}(\epsilon) = \sum_{\mathbf{k'}} W_{\mathbf{k'},\mathbf{k}}^{\text{el}-\text{ph}}(\epsilon_k' - \epsilon)[1 - f(\epsilon_{\mathbf{k'}})]$. This yields the rate[44]

$$\mathcal{J}_{\text{ph}}(\epsilon) = \frac{\pi N}{\hbar}\omega_0 g_0^2[1 - f(\epsilon - \omega_0)](N(\omega_0) + 1)\nu(\epsilon - \omega_0), \tag{10}$$

where $\nu(\epsilon) = \epsilon/(2\pi v^2\hbar^2)$ is the electron density of states in graphene. $\mathcal{J}_{\text{ph}}(\epsilon)$ varies linearly with the photo-excited carrier energy $\epsilon > \omega_0 + \mu$ and vanishes for $\epsilon < \omega_0 + \mu$; here we have set $T = 0$ for clarity (a good approximation for $k_{\text{B}}T \ll \mu,\omega_0$). Because the electron-optical phonon coupling is a constant, this result is to be expected from the increased phase space to scatter into at higher photo-excited carrier energy.

To get an order of magnitude estimate of the energy relaxation rate, we use values $(N(\omega_0) + 1) \approx 1$ and $1 - f(\epsilon - \omega_0) \approx \Theta(\epsilon - \omega_0 - \mu)$. This gives

$$\mathcal{J}_{\text{ph}}(\epsilon) \approx \frac{\epsilon - \omega_0}{\tau_0}\Theta(\epsilon - \omega_0 - \mu), \quad \tau_0 = \frac{2v^2\hbar^3}{N\omega_0 g_0^2} \qquad (11)$$

Using $\rho = 7.6 \times 10^{-11}\,\text{kg}\,\text{cm}^{-2}$, we find $\tau_0 \approx 350\,\text{fs}$.

D. Auger Processes vs. Phonon Emission: Branching Ratio

Comparing $\mathcal{J}_{\mathrm{el}}$ from intraband carrier-carrier scattering (in Sec. II B) with the energy relaxation rate arising from the emission of optical phonons, $\mathcal{J}_{\mathrm{ph}}$ (in Sec. II C), yields a branching ratio $\mathcal{J}_{\mathrm{el}}/\mathcal{J}_{\mathrm{ph}} > 1$ for typical dopings.[44] Indeed, large branching ratios $\mathcal{J}_{\mathrm{el}}/\mathcal{J}_{\mathrm{ph}} \sim 4$ have been inferred experimentally.[13] Furthermore, while $\mathcal{J}_{\mathrm{ph}}$ does not depend on carrier density, $\mathcal{J}$ does. As a result, gate voltage can be used to tune the branching ratio $\mathcal{J}_{\mathrm{el}}/\mathcal{J}_{\mathrm{ph}}$.[44]

The amount of heat absorbed, ΔQ_{el}, by the electronic system from the cascade of a single photoexcited carrier at $\epsilon = hf/2$ is

$$\Delta Q_{\mathrm{el}} = \int_0^{t_0} \mathcal{J}_{\mathrm{el}} dt = \int_\mu^{hf/2} \frac{d\epsilon}{1 + \mathcal{J}_{\mathrm{ph}}/\mathcal{J}_{\mathrm{el}}}. \qquad (12)$$

For large enough $\mu \gtrsim 0.1$, $\mathcal{J}_{\mathrm{el}}/\mathcal{J}_{\mathrm{ph}} > 1$. As a result, a large fraction of the energy from the photoexcited carriers is absorbed as heat in the electronic system. This precipitates an increase in the electronic temperature, T_{el}, and generates hot carriers.

E. Hot Carriers for Energy Harvesting

The ability to drive optoelectronic circuits from an elevated T_{el},[10,11] means that the energy captured by fast carrier-carrier scattering can be used to increase the efficiencies of optoelectronic devices. Indeed, one proposal in which efficiencies may be gained is using graphene's unique hot carrier generating characteristics for "hot carrier solar cells".[51] In these types of solar cells, the extraction of hot carriers can be used to drive a nanoscale heat engine to reach efficiencies as high as $60\,\%$, far larger than that imposed by the more traditional Shockley-Queisser limit for single-junction photovoltaic cells.[3]

There are two material requirements for such solar cells: i) efficient thermalization of high energy photoexcited electrons by ambient carriers so that photon energy can be captured as electronic heat, and ii) slow electron-lattice cooling so that the heat captured in (i) is not quickly lost to the lattice.[51] Indeed, the inefficient thermalization in conventional solar cells (such as those made out of Silicon) is one of the largest contributors to the low efficiencies of photovoltaic cells limited by the Shockley-Queisser limit.[3]

The combination of fast intraband Auger processes described in this section, slow electron-lattice cooling[18,19] that can be achieved in high quality graphene, and graphene's two-dimensional nature which enable vertical extraction of hot carriers make it an interesting candidate material for a new paradigm of solar cells based on hot carriers.[51] Vertical extraction is a front runner among the currently discussed proposals for extracting hot carriers to drive optoelectronic circuits;[7,10,15,16] utilizing graphene in energy harvesting is a topic of active research.

F. Measuring Hot Carrier Temperature

Measuring the electronic temperature is an ideal way of experimentally tracking energy flows in graphene. Since fast carrier-carrier scattering around the Fermi surface (tens of femtoseconds) allows for an electronic temperature to be established quickly[24,25,31] (see also above discussion on intra-band carrier-carrier scattering), the electronic temperature can be used to characterize both the short timescale thermalization and longer timescale cooling. Indeed, tracking the temperature dynamics of hot carriers[26] provides a sensitive probe of the cooling mechanisms that will be discussed in the next section. There are a variety of ways of measuring electronic temperature in graphene, including photocurrent,[12,26] noise-thermometry,[40] transient absorption,[27] angle resolved photoemission,[24] and THz conductivity.[13,52–58] Recently, the last two have been used in pump-probe type experiments[13,24,25] to probe the initial thermalization dynamics of the photoexcitation cascade (Fig. 1, left panel). Photoconductivity in pump-probe type experiments have recently received wide attention because of their ability to probe a wide variety of dynamical processes that can occur on short time scales.[59] Here, we will discuss THz conductivity and how effects from the dynamics of the electronic temperature alone lead to changes in the measured THz conductivity.

As we will see, the optical conductivity depends directly on the energy-dependent carrier distribution, $n_{\mathbf{k}}$, and gives a different value according to how far away from equilibrium the carrier distribution is. Analysis of a kinetic equation using the relaxation time approximation produces the conductivity

$$\sigma(\omega) = -\sum_{\mathbf{k}} \frac{\tau_{\mathrm{tr}}(\epsilon_{\mathbf{k}})}{1 + [\omega \tau_{\mathrm{tr}}(\epsilon_{\mathbf{k}})]^2} e^2 \mathbf{v_k} \cdot \nabla_{\mathbf{k}} n_{\mathbf{k}} \qquad (13)$$

where $\tau_{\mathrm{tr}}(\mathbf{k})$ is the transport scattering time, $\mathbf{v_k} = \partial \epsilon_{\mathbf{k}}/\partial \mathbf{k}$ is the group velocity, and $n_{\mathbf{k}} = (1 + e^{\beta(\epsilon_{\mathbf{k}} - \mu)})^{-1}$ depends on the electronic temperature through $\beta = 1/k_{\mathrm{B}} T_{\mathrm{el}}$. Here we have shown only the real part of the conductivity readily obtained in THz photoconductivity experiments;[13,52–58] the imaginary part can also be similarly obtained. In pump-probe type experiments, the pump-induced change in conductivity (also called photoconductivity) provides information about how far the carrier distribution is pushed out of equilibrium. As such, we will be most interested in the temperature dependent changes in conductivity $\Delta\sigma_\omega = \sigma_\omega(T_{\mathrm{el},1}) - \sigma_\omega(T_{\mathrm{el},0})$. Here $T_{\mathrm{el},0}$ denotes the initial electronic temperature and $T_{\mathrm{el},1}$ denotes the electronic temperature after pump.

Two regimes naturally arise in graphene: (i) the degenerate limit, $\mu \gg k_{\mathrm{B}} T_{\mathrm{el}}$, in doped graphene, and (ii)

the non-degenerate limit, $\mu < k_B T_{el}$, in graphene close to charge neutrality. Since carrier density in graphene can be tuned by gate, these two regimes can be easily accessed.

In the degenerate limit, $\mu \gg k_B T_{el}$, a Sommerfeld expansion can be employed in the analysis of Eq.(13). This yields an estimate for $\Delta\sigma_\omega$ as[13]

$$\Delta\sigma_\omega \approx \left(k_B^2 T_{el,1}^2 - k_B^2 T_{el,0}^2\right)\frac{\pi^2}{6}N\nu(E_F)\frac{\partial^2 F(\epsilon)}{\partial\epsilon^2}\bigg|_{\epsilon=E_F}$$

$$F(\epsilon) = e^2 v^2 \frac{\tau_{tr}(\epsilon)}{1+\omega^2[\tau_{tr}(\epsilon)]^2} \tag{14}$$

where E_F is the Fermi energy, and $N = 4$ is the valley/spin degeneracy in graphene. Here the carrier density has been fixed constant by accounting for changes in chemical potential as a function of temperature $\mu \approx E_F - \frac{\pi^2}{6}k_B^2 T^2/E_F$; temperature dependence of the transport time, $\tau_{tr}(\epsilon)$, has also been neglected and can be included in a more elaborate analysis. Both of these assumptions are valid when $\mu \gg k_B T_{el}$.

For graphene, the transport time can be modeled as $\tau_{tr} = a\epsilon$;[60–62] here a is a sample dependent constant. As a result, $\frac{\partial^2 F(\epsilon)}{\partial\epsilon^2} < 0$ for small frequencies $\omega \sim$ THz giving a *negative* $\Delta\sigma_\omega < 0$ when electronic system gets hotter $T_{el,1} > T_{el,0}$. Here, an estimate of sample dependent a can be extracted from the dc conductivity via the einstein relation, $\sigma_{dc} = e^2 v^2 \nu(\mu)\tau(\mu)/2$. We note parenthetically, that since the heat capacity in the degenerate limit goes as $C_{el} \propto T$, the absolute change in optical conductivity $|\Delta\sigma_\omega| \propto \Delta Q_{el}$ making the optical conductivity a good probe of the amount of heat injected into the electronic system. In the context of probing the photoexcitation cascade, $|\Delta\sigma_\omega|$ is sensitive to the amount of heat captured by the ambient carriers.[13] Pump induced negative photoconductivity in doped graphene has recently been observed experimentally.[13,54–56]

In contrast, $\Delta\sigma_\omega > 0$ in the non-degenerate limit ($\mu < k_B T_{el}$). To illustrate this, we use $\tau_{tr} = a\epsilon$ as above and fix $\mu = 0$. For $\omega\tau_{tr}(\epsilon = k_B T) \ll 1$, we estimate

$$\Delta\sigma_\omega \approx \frac{Ne^2 a\pi}{12\hbar^2}k_B^2\left(T_{el,1}^2 - T_{el,0}^2\right) \tag{15}$$

which is *positive* as the temperature of the electronic system gets hotter $T_{el,1} > T_{el,0}$. Interestingly, $\Delta\sigma$ in this case is no longer directly proportional to the amount of heat pumped in the the system (since $C_{el} \propto T^2$ in the non-degenerate limit). In recent experiments, the different signs of the two regimes have been observed in dynamical measurements involving an optical pump THz probe of a single gate tunable graphene sample.[57,58] These have been described in terms of metallic behavior (in the degenerate limit) and semiconducting-like behavior (non-degenerate limit) respectively.[57,58] A related gate tunable photoconductivity sign change was also observed in steady state photoconductivity measured in biased graphene.[63]

While we only describe the effect of an increase in electron temperature on $\Delta\sigma$, a number of other scattering channels can also affect $\Delta\sigma$.[13,57,58,64] However, many of the qualitative features, such as photoconductivity sign, remain unchanged. For example, using an energy independent τ_{tr}, Ref. 57 and 58 also found a positive $\Delta\sigma$ in the non-degenerate limit; positive $\Delta\sigma$ in this limit has been attributed to a change in the Drude weight.[57,58] Additionally, optical phonons in both graphene[13,64] and the substrate[64] emitted in the photoexcitation cascade can also scatter with ambient carriers to change the optical conductivity measured.

III. HOT CARRIER COOLING AND SUPERCOLLISIONS

The mechanisms responsible for cooling hot carriers in graphene control how long the electronic system stays out of thermal equilibrium with the lattice. The time scales over which the electron system is hotter than the lattice determines the time scales (or length scales) over which hot carrier effects can be observed.[11] For instance, large and spatially extended photocurrent response in graphene p-n junctions has been attributed to long hot carrier cooling lengths.[10] As a result, hot carrier cooling mechanisms play a critical role in determining the magnitude and qualitative behavior of graphene response. Below we detail what controls cooling in graphene.

Hot carrier cooling proceeds via the emission of phonons through the electron-phonon interaction $\mathcal{H}_{el-ph}$ in Eq.(1). The electron-phonon interaction in graphene arises from the usual deformation potential coupling,

$$\mathcal{H}_{el-ph} = \sum_{\mathbf{q}} g\sqrt{\omega_{\mathbf{q}}}\left(b_{\mathbf{q}} + b_{-\mathbf{q}}^\dagger\right)n_{\mathbf{q}}, \quad g = D/\sqrt{2\rho s^2}, \tag{16}$$

where D is the deformation potential, ρ is the mass density of the graphene sheet, and s is the speed of sound in graphene. In this section, we use $D = 20\,\text{eV}$ throughout.[18,19] As discussed earlier, the high energy of optical phonons in graphene quench their contribution to electron-phonon cooling in graphene. While cooling through substrate phonons can also occur,[65,66] here we will only consider cooling mediated through the exchange of acoustic phonons in graphene.

While $\mathcal{H}_{el-ph}$ is central to electronic cooling, as we will see disorder plays a surprising role. The effect of the disorder potential can be modeled as a sum of impurity potentials and a gauge field originating from strain that couples to electron velocity,[67–70]

$$\mathcal{H}_{dis} = \sum_{\mathbf{r},\alpha} \psi_\alpha^\dagger(\mathbf{r})\left[\sum_i V(\mathbf{r}-\mathbf{r}_i) + v\sigma\cdot\mathbf{A}(\mathbf{r})\right]\psi_\alpha(\mathbf{r}). \tag{17}$$

The second type of disorder is peculiar for graphene, where it can arise due to disorder-induced ripples on the graphene sheet.[68–72] The random vector potential $\mathbf{A}(\mathbf{r})$ depends on the microscopic ripple profile.

We begin to analyze electronic cooling (middle panel Fig. 1) by considering electron-phonon scattering rate described by Fermi's Golden Rule,

$$W_{\mathbf{k}',\mathbf{k}} = \frac{2\pi}{\hbar} \sum_{\mathbf{q}} \left[|M_+|^2 N_{\omega_{\mathbf{q}}} \delta(\epsilon_{\mathbf{k}'} - \epsilon_{\mathbf{k}} - \omega_{\mathbf{q}}) \right. \tag{18}$$

$$\left. + |M_-|^2 (N_{\omega_{\mathbf{q}}} + 1)\delta(\epsilon_{\mathbf{k}'} - \epsilon_{\mathbf{k}} + \omega_{\mathbf{q}}) \right],$$

where $\mathbf{q}$ is phonon momentum, and $N_{\omega_{\mathbf{q}}} = 1/(e^{\beta\omega_{\mathbf{q}}} - 1)$ is the Bose distribution. In the absence of disorder, the matrix elements are derived from the deformation potential, $M_{\pm}^{(0)} = g\sqrt{\omega_{\mathbf{q}}}\delta_{\mathbf{k}'-\mathbf{k}\mp\mathbf{q}}$, where the delta function enforces momentum conservation. This gives the cooling power

$$\mathcal{J} = \sum_{\mathbf{k},\mathbf{k}',i} W_{\mathbf{k}',\mathbf{k}}(\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}'})f(\epsilon_{\mathbf{k}})\left[1 - f(\epsilon_{\mathbf{k}'})\right] \tag{19}$$

where $f(\epsilon) = 1/(e^{\beta(\epsilon-\mu)} + 1)$ are Fermi functions, $W_{\mathbf{k}',\mathbf{k}}$ is the transition probability, and $\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}'}$ is the energy exchanged in each scattering event. Here we note that since thermalization (through carrier-carrier scattering, see Sec. II B) occurs far faster than the cooling time scales described here, a separation of time scales allows us to treat the electron distribution as a Fermi distribution with a higher temperature, T_{el}, than the lattice, T_{ph}.

The small Fermi surfaces in graphene allow easy access to two distinct regimes of electron-phonon scattering, namely $T > T_{\text{BG}}$, and $T < T_{\text{BG}}$. Here $T_{\text{BG}} = s\hbar k_{\text{F}}$ is the Bloch Grüneissen temperature, which are tens of kelvin for typical densities in graphene; k_{F} is the Fermi wavevector. In the more familiar $T < T_{\text{BG}}$ limit, electron-phonon scattering is constrained by the small available phonon phase space set by the low temperature;[73] this is analogous to electron-phonon scattering in metals.[74] The Bloch-Grüneissen regime exhibits suppressed cooling powers that decrease as temperature is decreased, giving $\mathcal{J} \propto (T_{\text{el}}^{\delta} - T_{\text{ph}}^{\delta})$ where $\delta = 4$.[75,76] Combined with a small electronic heat capacity, $T < T_{\text{BG}}$ feature long cooling times that vary between nanoseconds to microseconds in the several Kelvin to tens of milliKelvin range.[75–78] The extremely small cooling powers achievable at ultracold temperatures allow for increased sensitivity for graphene bolometers.[77]

Surprisingly, long cooling times from the emission of acoustic phonons can also occur in the regime $T > T_{\text{BG}}$ that can be as long as a few nanoseconds.[18,19] This appears because the mismatch of Fermi and sound velocities, $v/s \approx 100$, together with the conservation of energy and momentum severely constrain the available phase space. As a result, only very long wavelength acoustic phonons (i.e. weakly coupled) are emitted. Indeed, the phonons exchanged have energy $\omega_{\mathbf{q}} \approx s\hbar k_{\text{F}} = k_{\text{B}}T_{\text{BG}}$, a small fraction of $k_{\text{B}}T$ for practically interesting temperatures (see Fig. 3a). This yields a cooling power that is suppressed[18,19]

$$\mathcal{J}_0 = B(T_{\text{el}} - T_{\text{ph}}), \quad B = \pi N\hbar g^2 [\nu(\mu)]^2 k_{\text{F}}^2 s^2 k_{\text{B}}. \tag{20}$$

where $N(\omega_{\mathbf{q}}) \approx k_{\text{B}}T/\omega_{\mathbf{q}}$. As a result, cooling times can be as long as several nanoseconds even at room
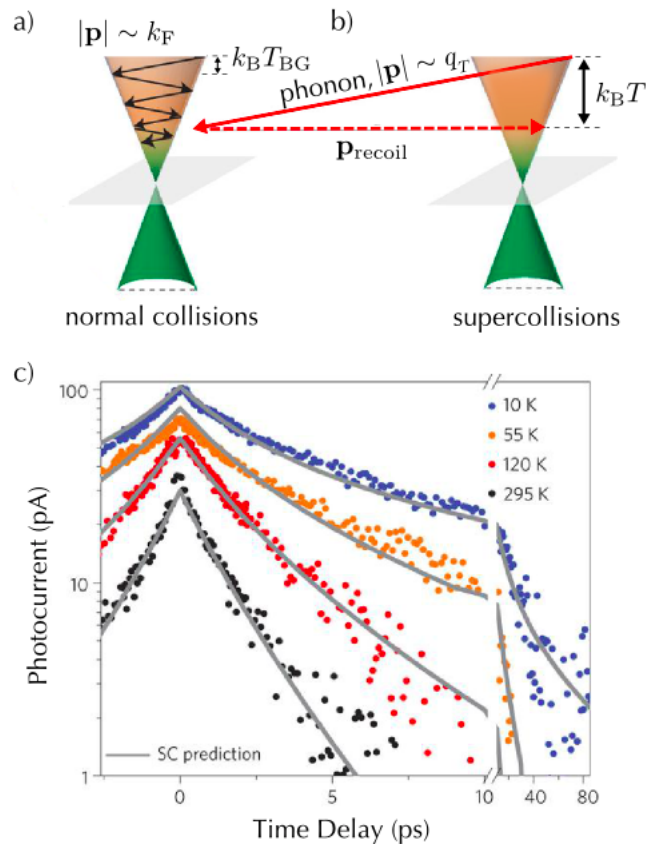

FIG. 3. a,b) Kinematics of electron cooling in graphene showing (a) the normal momentum conserving process (takes many steps), and (b) the supercollision process (takes few steps). In (a), phonon momenta are constrained by the Fermi surface so that they carry off energy $\sim k_{\text{B}}T_{\text{BG}} \approx$ several meV (for typical carrier densities, see for example Ref. 26 and 40) In (b), phonon momenta are totally unconstrained for supercollisions, with the recoil momentum ($|\mathbf{p}_{\text{recoil}}| \approx q_T$) transferred to the lattice via disorder scattering. The energy scales in this illustration have been exaggerated to illustrate the kinematics of phonon emission; hot electrons are thermal and close to the Fermi surface. c) Electron temperature dynamics extracted from a photocurrent experiment[26] (see text) showing two distinct regimes of cooling expected from supercollisions: i) an initial power law cooling dynamics (see Eq. 28), and ii) later exponential regime (see Eq. 29); photocurrent is proportional to the electronic temperature. Gray lines denote predicted dynamics from supercollisions [Eq. (29)]. Illustration in (a,b) and experimental data (c) adapted from Ref. 26.

temperature, and the cooling rate exhibits an anomalous temperature dependence increasing with decreasing temperature.[18,19]

The cooling power of electrons in $T > T_{\text{BG}}$ can be dramatically enhanced by supercollision processes that can relieve the bottleneck from the momentum conserving normal process. This is most easily demonstrated by electron-phonon scattering in the presence of disorder (see Fig. 3b for an illustration). Since disorder breaks

translational invariance, emitted (absorbed) phonon momenta in Eq.(18) can take on unconstrained values, $|\mathbf{q}| \lesssim q_T$. As a result, thermal phonons are exchanged in supercollision electron-phonon scattering processes boosting the exchange of energy between electron and lattice systems.

At low disorder concentration, we can describe supercollision electron-phonon scattering by dressing the electron-phonon vertex with multiple scattering on a single impurity. This gives an expression for the transition matrix elements $M_\pm$ which is exact in the impurity potential:

$$M_\pm = \langle \mathbf{k}'|M_\pm^{(0)}G\hat{T} + \hat{T}GM_\pm^{(0)} + \hat{T}GM_\pm^{(0)}G\hat{T}|\mathbf{k}\rangle, \quad (21)$$

where $G(\mathbf{p}) = \frac{1}{\epsilon - H_0(\mathbf{p}) + \mu}$ is the electron Green's function, $\hat{T}$ is the T-matrix (scattering operator) for a single impurity. Scattering off an impurity relieves the momentum conservation that originally constrained the phase space

for single acoustic phonon scattering allowing higher energy thermal phonons to be emitted. The supercollision transition matrix element $M_\pm$ are smaller than that of single acoustic phonon emission $M_\pm^{(0)}$ resulting in a smaller frequency of scattering.[39] However, the ability to emit thermal phonons more than makes up for the decreased frequency in scattering, thereby boosting cooling powers. Indeed, supercollisions dominate electron-lattice cooling over a wide range of technological relevant temperatures, see Table I.

In the following, we lay out a variety of supercollision scattering processes and calculate their cooling powers. We begin first with supercollision cooling that arises from scattering off short-range impurity scatterers which will serve as a conceptually simple illustration of supercollisions; other supercollision processes follow from similar reasoning (see Table I and discussion below) and yield qualitatively similar cooling powers.

Supercollision Process	Cooling Power (Linearized)	Crossover Temperature, T_*	Equation
Short-Range Disorder + 1 Acoustic Phonon	$T^2\Delta T$	$\approx$ few T_{BG}	Eq.(24)
Ripple Disorder + 1 Acoustic Phonon	$T^2\Delta T$	$\approx$ few T_{BG}	Eq.(34)
Coulomb Disorder + 1 Acoustic Phonon	vanishes	-	Eq.(31)
Two Acoustic Phonons: Anharmonic Coupling	$T^5\Delta T$	$\approx 20\,T_{BG}$	Eq.(39)
Two Acoustic Phonons: Deep Off-Shell Process	vanishes at leading order	-	Eq.(41)
Two Flexural Phonons	$T^2\Delta T$	$\approx 10\,T_{BG}$	Eq.(44)

Table I: Various supercollision processes in monolayer graphene and their linearized cooling power for $T > T_{BG}$. T_* indicates the crossover temperatures above which the supercollision process dominates over momentum conserving single acoustic phonon emission, $\mathcal{J}_0$.

A. Short Range Disorder-Assisted Supercollisions

Here we consider scattering by short-range impurities (the first term in Eq.(17)) first discussed in Ref. 39. Short-range disorder can be modeled by a delta function potential

$$V(\mathbf{r} - \mathbf{r}_j) = u\delta(\mathbf{r} - \mathbf{r}_j)(\hat{1} \pm \sigma_z)/2, \quad (22)$$

where the plus (minus) sign corresponds to the impurity positions on the A (B) sites of the carbon lattice, and u is the strength of the short-range scatterer. Focussing only on contributions first order in u, we can approximate $\hat{T}_{\mathbf{p}',\mathbf{p}} = \frac{1}{2}u(\hat{1} \pm \sigma_z)e^{i(\mathbf{p}'-\mathbf{p})\mathbf{r}_j} + O(u^2)$ in Eq.(21).

As we shall see, the main contribution to cooling will arise from phonons with momenta of order q_T. Thus we anticipate that the virtual electron states in the transition matrix element Eq.(21), described by the Green's functions $G(\mathbf{p})$, are characterized by large momenta $|\mathbf{p}| \sim q_T$ which are much greater than $\mathbf{k}$, $\mathbf{k}'$. In this case, for the off-mass-shell virtual states such that $v|\mathbf{p}| \gg \mu, k_B T$, we can approximate $G(\mathbf{p}) \approx -\frac{1}{H_0(\mathbf{p})}$. This form of $G(\mathbf{p})$ describes deep off-shell virtual states, i.e. virtual states

far from mass shell. The stiffness of the electron dispersion, $v \gg s$, along with the estimate $|\mathbf{p}| \sim q_T$, makes it an accurate approximation for all virtual states not too close to the Fermi surface. In this limit, as we now show, drastic simplifications occur because of particle-hole symmetry $H_0(-\mathbf{p}) = -H_0(\mathbf{p})$.

Concentrating on contributions first order in u, we evaluate the first two terms in Eq.(21). This gives the commutator of $H_0^{-1}(\mathbf{q})$ and $\pm\sigma_z$ arising because the virtual electron states in the first and second term have momenta $\mathbf{p} \approx -\mathbf{q}$ and $\mathbf{p} \approx +\mathbf{q}$ (see above). We obtain

$$M_\pm = \frac{\pm iug\sqrt{\omega_\mathbf{q}}}{\hbar v|\mathbf{q}|^2}\langle \mathbf{k}'|(\sigma \times \mathbf{q})_z|\mathbf{k}\rangle e^{i(\mathbf{p}'-\mathbf{p}\pm\mathbf{q})\mathbf{r}_j}, \quad (23)$$

with the phase factor describing the dependence on the impurity position. We note that the finite commutator, $[\sigma_z, \sigma \cdot \mathbf{q}] = 2i(\hat{\mathbf{z}} \times \mathbf{q}) \cdot \sigma$ arises because of the matrix/sublattice structure of the impurity potential used. As we will see later, this structure is critical; for impurities that have an identity structure in pseudospin space, M vanishes. In the following, we proceed to evaluate the energy-loss power in Eq.(19).

Integrating Eq.(19) using Eq.(23) in the same way as

detailed above, and summing over A(B) impurities after squaring $|M|$,[61] the supercollision cooling power for $\mu \gg k_B T$ is[39]

$$\mathcal{J}_{\text{dis}} = A_{\text{dis}} \left(T_{\text{el}}^3 - T_{\text{ph}}^3 \right), \qquad A_{\text{dis}} = 9.62 \frac{g^2 \nu^2(\mu) k_B^3}{\hbar k_F \ell}, \tag{24}$$

where the mean free path $k_F \ell = 2\hbar^2 v^2/(u^2 n_0)$.[61]

It is useful to compare the supercollision cooling power in Eq.(24) with the momentum conserving channel in Eq.(20). The enhancement factor of disorder-assisted cooling over the momentum conserving pathways depends on both disorder and temperature:

$$\frac{\mathcal{J}_{\text{dis}}}{\mathcal{J}_0} = \frac{0.77}{k_F \ell} \frac{T_{\text{el}}^2 + T_{\text{el}} T_{\text{ph}} + T_{\text{ph}}^2}{T_{\text{BG}}^2}, \tag{25}$$

where we used the value of g described above. At room temperature, $T_{\text{el(ph)}} \sim 300\,\text{K}$, and taking $\mu = 100\,\text{meV}$ ($n \sim 10^{12}\,\text{cm}^{-2}$) we find $T_{\text{el(ph)}}/T_{\text{BG}} \approx 50$. For $k_F \ell = 20$, the enhancement factor $\mathcal{J}/\mathcal{J}_0$ can be as large as 100 times. For small $\Delta T = T_{\text{el}} - T_{\text{ph}}$, we find that $\mathcal{J}_{\text{dis}}$ dominates over $\mathcal{J}_0$ at temperatures $T > T_* = \sqrt{\frac{B}{3A}} = \left(\frac{\pi}{6\zeta(3)} k_F \ell \right)^{1/2} T_{\text{BG}}$. Taking $k_F \ell = 20$ for a rough estimate, we see that the disorder assisted cooling channel dominates for $T \gtrsim 3 T_{\text{BG}}$. The crossover temperature can be controlled by gate voltage, since $T_{\text{BG}} \propto \sqrt{n}$. For typical carrier densities n this gives a crossover temperature T_* of a few tens of kelvin (see Table I).

Supercollision cooling in Eq.(24) is not only large, but it also exhibits qualitatively different features from the normal momentum conservation process that include, a cooling rate that decreases with decreasing temperature and two-stage cooling dynamics (discussed below), as well as a linear dependence on carrier density, see Eq.(24). In contrast, normal momentum-conserving electron-phonon processes (single acoustic phonon emission) depends quadratically on carrier density.

B.   Temperature Dynamics

The temperature dynamics of hot carrier cooling can be described by accounting for both disorder-assisted supercollisions and momentum-conserving cooling, yielding hot carrier relaxation dynamics

$$d\mathcal{Q}/dt = -\mathcal{J}_{\text{dis}} - \mathcal{J}_0, \tag{26}$$

where $\mathcal{Q}$ is the electron energy density. For small $\Delta T \ll T_{\text{el}}, T_{\text{ph}}$, and taking $\mathcal{Q} = C\Delta T$, with $C = \alpha T$ the heat capacity of the degenerate electronic system ($\alpha = \frac{\pi^2}{3} N\nu(\mu)k_B^2$), the cooling times describing relaxation to equilibrium, $\Delta T_{\text{el}}(t) = e^{-(t-t_0)/\tau}\Delta T_{\text{el},0}$, exhibit a nonmonotonic T dependence:

$$\frac{1}{\tau} = \frac{3A}{\alpha}T + \frac{B}{\alpha T}. \tag{27}$$

The cooling time increases with T at $T < T_*$ and decreases at $T > T_*$, reaching maximal value at $T = T_*$. This non-monotonic temperature dependence in τ provides a clear experimental signature of the competition between different cooling pathways (see Sec. III C).

On the other hand for $T_{\text{el}} \gg T_{\text{ph}}$, the energy-relaxation power is $d\mathcal{Q}/dt \approx AT_{\text{el}}^3$. This yields a $1/(t - t_0)$ dynamics:[39]

$$T_{\text{el}}(t) = \frac{T_{\text{el},0}}{1 + (A/\alpha)(t - t_0)T_{\text{el},0}}. \tag{28}$$

For large initial electron temperatures $T_{\text{el}} \gg T_{\text{ph}}$, both the dynamics for $T_{\text{el}} \gg T_{\text{ph}}$ and $T_{\text{el}} \approx T_{\text{ph}}$ can be accessed at early and long times respectively, resulting in a distinct two-stage cooling dynamics. The full dynamics can be found by directly solving Eq.(26), yielding[39]

$$-\frac{2}{\tau}(t - t_0) = F(T_{\text{el}}(t)/T_{\text{ph}}) - F(T_{\text{el},0}/T_{\text{ph}}), \tag{29}$$

where $F(x) = 2\sqrt{3}\arctan[(1+2x)/\sqrt{3}] - \ln[(x^3-1)/(x-1)^3]$ ($\mathcal{J}_0$ has been suppressed only becoming important for $T \lesssim T_*$ and at long times). This solution is characterized by only two parameters: the initial dimensionless temperature $x_0 = T_{\text{el},0}/T_{\text{ph}}$ and the time constant $\tau = 3T_{\text{ph}}A/\alpha$ allowing a clear way of comparing with experiments that probe the carrier dynamics of graphene.

C.   Experimental Observation of Supercollisions

Observing the cooling mechanisms in graphene have been the subject of a number of recent experiments.[26,27,40,41,53,79,80] Indeed by accessing various aspects of electronic cooling, these have found that supercollisions dominates electron-lattice energy relaxation in graphene over a wide range of temperatures. The methods employed can be broadly classified as i) dynamical measurements of temperature dynamics,[26,27,53] and ii) steady-state measurements of cooling power.[40,41,79,80] We briefly review a few of these experiments.

Dynamical measurements of hot carrier temperature dynamics provide a powerful method in identifying cooling channels in graphene.[26,27,53] Ref. 53 tracked the cooling dynamics of hot carriers in an optical pump-THz probe experiment and found that cooling times increased as lattice temperature was decreased. This is consistent with supercollisions described above but contrasts with normal collisions where the opposite behavior is expected (see Eq. 20).

In a similar vein, Ref. 26 extracted the cooling dynamics of graphene under pulsed illumination in a pump-probe setup,[26] using photocurrent from the photothermoelectric effect as a thermometer of the electrons in graphene.[10,26,27] In these measurements, Ref. 26 found a distinct two-regime cooling dynamics as shown in Fig. 3c. When pumped intensely with pulsed irradiation, the electronic temperature decayed via an initial power law
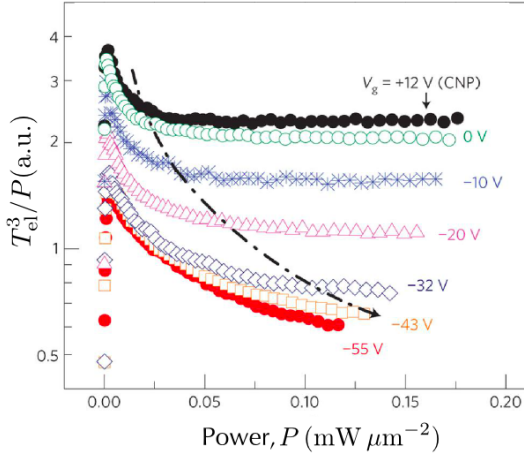
FIG. 4. Experimental observation of supercollisions (see also Fig. 3c for a dynamical measurement). Steady state electronic temperature follows a cubic power law measured through noise thermometry in Ref. 40, as shown by the flat plateau regions [see text, and $T_{\mathrm{el}} \gg T_{\mathrm{ph}}$ regime in Eq. (24)]. Note that the crossover temperature to a supercollision dominated regime (plateau) is controlled by gate potential via the density dependence of $T_{\mathrm{BG}} = s\hbar k_{\mathrm{F}}$ [see Eq. (25)].

regime, and a later exponential regime matching the predictions of Eq. (28), (29) above (see theoretical prediction in gray line coinciding with experimental data in Fig. 3c). Further by controlling the initial excited electron temperature, Ref. 26 was able to vary the power-law regime time constant confirming the $T_{\mathrm{el},0}$ scaling predicted in Eq. (28).

Ref. 40 adopted an alternative steady state approach in which they measured electronic temperature through noise thermometry. By Joule heating the electronic system, and subsequently extracting its (steady state) electronic temperature, Ref. 40 were able to establish that for lattice temperatures larger than T_{BG}, the electronic temperature scaled as $T_{\mathrm{el}} \propto P^{1/3}$ in agreement with Eq. (24) (plateau in Fig. 4); here P is the power pumped into the system. Ref. 26 also found the same scaling when they used a continuous wave illumination. Further, they varied the onset of the plateau region via gate voltage; the smaller the size of the Fermi surface the earlier the onset of supercollisions [see Eq. (25)]. Additionally, they extracted a linear density dependence of the electron-lattice cooling power [see $A_{\mathrm{dis}} \propto [\nu(\mu)]^2$ in Eq. (24)]; this contrasts with the quadratic density dependence expected from normal collisions [Eq. (20)].

Another interesting aspect of electronic cooling in graphene is the non-monotonic temperature dependence of cooling times, Eq. (27), that arises from the competition between normal process (dominating for $T < T_*$) and supercollisions (dominating for $T > T_*$). In Ref. 41, a steady state photothermoelectric photocurrent technique was employed that tracked both the electronic temperature and cooling lengths in high quality gated

graphene devices. They observed a non-monotonic temperature dependence of cooling times [Eq. (27)] peaking at an intermediate temperature T_*. Further, Ref. 41 demonstrated control of the peak temperature T_* via gate voltage and disorder (through an annealing technique).[41]

D. Coulomb Impurities

A large variety of disorder types (short-range, long-range Coulomb scatterers, ripples) can exist in graphene; it is important (on both the technological and fundamental levels) to consider what kind of disorder is necessary for supercollisions. An important type of disorder are Coulomb scatterers, which are believed to dominate the electronic transport characteristics of graphene.[62]

To analyze the effect of Coulomb impurities, we proceed in the same way as detailed above with the transition matrix element for supercollision scattering described by Eq.(21). Using the Born approximation, we write $\hat{T}_{\mathbf{p}',\mathbf{p}} = \frac{2\pi Z e^2}{\kappa |\mathbf{p}'-\mathbf{p}|}$, where Z is the impurity charge and κ is the dielectric constant. We take the unscreened Coulomb potential since the main contribution to cooling come from phonons with momenta $q_T \gg k_{\mathrm{F}}$. Additionally, since Coulomb scatterers are long-ranged, they do not differentiate between sublattices; any differences arising from having Coulomb impurities on different sites can be absorbed into the scattering from short-range scatterers (see above).

To evaluate the matrix element, we begin with the first two terms of Eq.(21). Following the analysis above, the sum of the first two terms of Eq.(21) evaluates approximately as the commutator between $G(\mathbf{p})$ and $\hat{T}$. Since $\hat{T}$ for Coulomb scatterers is an identity matrix in graphene's pseudospin space, this commutator *vanishes* and the sum of the first two terms evaluates to zero. This cancellation originates from the equal (but opposite in sign) contributions of positive and negative energy virtual states. The physical origin of such "destructive interference effect" can be linked to the particle-hole symmetry of graphene's electronic spectrum, $H_0(-\mathbf{p}) = -H_0(\mathbf{p})$ (see Sec. III G for a full discussion).

We proceed to consider the last term in Eq.(21). Noting that the dominant contributors to cooling arise from phonons with momenta q_T we estimate

$$M_{\pm} = \frac{g\sqrt{\omega_{\mathbf{q}}}}{\kappa^2} \sum_{\mathbf{p}} \langle \mathbf{k}' | \frac{(2\pi Z e^2)^2}{|\mathbf{p}-\mathbf{q}||\mathbf{p}|} H_0^{-1}(\mathbf{p}-\mathbf{q}) H_0^{-1}(\mathbf{p}) | \mathbf{k} \rangle. \tag{30}$$

We can evaluate this by rewriting $H_0^{-1}(\mathbf{p})/|\mathbf{p}| = \sigma \cdot \nabla_{\mathbf{p}} \frac{1}{|\mathbf{p}|}$. Performing a 2D Fourier transform (to the conjugate variable $\mathbf{r}$) we obtain

$$M_{\pm} = c \int d^2\mathbf{r} \langle \mathbf{k}' | \frac{(\sigma \cdot \mathbf{r})^2}{|\mathbf{r}|^2} e^{i\mathbf{q}\cdot\mathbf{r}} | \mathbf{k} \rangle = c(2\pi)^2 \delta(\mathbf{q}) \langle \mathbf{k}' | \mathbf{k} \rangle, \tag{31}$$

where $c = g\sqrt{\omega_{\mathbf{q}}}(2\pi Z e^2)^2/(\kappa^2 v^2 \hbar^2)$. Since the contributions we are interested in are $|\mathbf{q}| \gg k_{\mathrm{F}}$, the matrix

element *vanishes*. Hence, we conclude that Coulomb scatterers do not contribute to the supercollision cooling channel. This observation means that, while Coulomb scatterers are central for the electronic transport characteristics of graphene, they contribute very little to the energy relaxation pathways of hot carriers. Coulomb scatterers are therefore not as important for graphene optoelectronic properties as they are for transport properties.

E. Ripple-assisted Supercollisions

Another type of disorder arises from ripples in graphene. Ripples can originate in a number of ways. These include: (i) *intrinsic* ripples which occur in pristine and suspended graphene, and (ii) *extrinsic* undulations caused by corrugations of an underlying substrate. Both of them cause electrons to be scattered off random strain in the graphene sheet. This type of disorder is described by the vector potential term in Eq.(17). Evaluating the electron-phonon matrix element, Eq.(21), we find a nonzero contribution at lowest order in $\mathbf{A_q} = \int d^2 r e^{-i\mathbf{q}\mathbf{r}} \mathbf{A(r)}$. The estimate proceeds in the same way as for short range disorder, with $G(\epsilon, \mathbf{q}) \approx 1/H_0(\mathbf{q})$ describing deep off-shell states. The matrix structure $\sigma \cdot \mathbf{A(r)}$ generates the commutator $[\sigma \cdot \mathbf{A_q}, H_0^{-1}(\mathbf{q})] = \frac{-1}{\hbar v |\mathbf{q}|} 2i\sigma_z \tilde{\mathbf{q}} \cdot \mathbf{A_q}$, giving

$$M_\pm(\mathbf{A_q}) = \langle \mathbf{k'} | \frac{2g\sqrt{\omega_q} i \sigma_z}{\hbar |\mathbf{q}|} \tilde{\mathbf{q}} \cdot \mathbf{A_q} | \mathbf{k} \rangle. \qquad (32)$$

Evaluating the transition rate $W_{\mathbf{k'},\mathbf{k}}$ and plugging it into Eq.(19), we obtain the cooling power

$$\mathcal{J}_{\text{ripple}} = b \sum_{\mathbf{q}} \omega_q (\tilde{N}_{\omega_{\mathbf{q}}} - N_{\omega_{\mathbf{q}}}) \langle |\tilde{\mathbf{q}} \cdot \mathbf{A_q}|^2 \rangle_{\text{dis}}, \qquad (33)$$

where $b = 4\pi N g^2 s^2 [\nu(\mu)]^2 / \hbar$ and averaging over the ripple ensemble is denoted by $\langle ... \rangle_{\text{dis}}$.

We note that the pair correlator $\chi(\mathbf{q}) = \langle A_{\mathbf{q}} A_{-\mathbf{q}} \rangle_{\text{dis}}$ decays at $|\mathbf{q}| \approx q_0 = 2\pi/R$, where R is the characteristic radius of a ripple.[68] Here we assume that the radius R is much shorter than the phonon wavelength for temperatures of interest (indeed, Ref. 71 estimates $R = 4$ nm, whereas at room temperature $\lambda_T = 2\pi/q_T \approx 17$ nm). Since at $q_T \ll q_0$ the integral over $\mathbf{q}$ in the general expression (33) is limited solely by the Bose functions $N_{\omega_{\mathbf{q}}}, \tilde{N}_{\omega_{\mathbf{q}}}$, the correlator $\chi(q)$ is approximately given by its value at $|\mathbf{q}| \ll q_0$. This allows us to evaluate the integral over the Bose functions treating $\chi(q)$ as a constant. The correlator $\chi(\mathbf{q})$ was analyzed in Ref. 68. For the roughness exponent $2H \approx 1$ corresponding to graphene on a substrate,[72] Ref. 68 obtains $\chi(|\mathbf{q}| \ll q_0) \approx (\hbar/a)^2 Z^4/R^2$ where Z is the characteristic out of plane displacement in the graphene sheet caused by ripples and a is the interatomic distance. This yields a cooling power

$$\mathcal{J}_{\text{ripple}} = A_{\text{ripple}} \left(T_{\text{el}}^3 - T_{\text{ph}}^3 \right), \qquad (34)$$

where A_{ripple} is

$$A_{\text{ripple}} = \frac{2\zeta(3) k_{\text{B}}^3 b}{2\pi \hbar^2 s^2} \int \frac{d\theta_q}{2\pi} \langle |\tilde{\mathbf{q}} \cdot \mathbf{A}_{|\mathbf{q}| \ll q_0}|^2 \rangle_{\text{dis}} = \frac{\zeta(3) b Z^4 k_{\text{B}}^3}{2\pi s^2 a^2 R^2}, \qquad (35)$$

where we accounted for a factor of $1/2$ due to the angular integral. Comparing this result with the contribution of normal collisions in Eq.(20), we obtain a crossover temperature

$$T_* = \eta^{-1/2} T_{\text{BG}}, \quad \eta = \frac{6\zeta(3)}{\pi} \frac{Z^4}{R^2 a^2}. \qquad (36)$$

Taking the values reported in Ref. 71, $Z = 0.6$ nm and $R = 4$ nm, we find $\eta \approx 0.9$, which gives $T_* \approx T_{\text{BG}}$. For a doping level of $\mu = 100$ meV this is $T_* = 12$ K. For weaker disorder, we anticipate T_* to be a few times T_{BG}. Recently, shot-noise thermometry measurements in Ref. 80 measured cooling in suspended graphene samples and attributed it to ripple-assisted supercollisions.

F. Two-Acoustic Phonon Emission: Anharmonic Coupling

Next, we consider two-phonon scattering which can utilize the supercollision phase space enhancement in much the same way as the disorder-assisted channel described above; two-phonon processes considered in this section (and Sec. III G, III H) are intrinsic processes which will also occur in pristine graphene (in the absence of disorder). This process is described by an anharmonic coupling[81,82]

$$\mathcal{H}_{\text{el}-2\text{ph}} = D' \sum_{\mathbf{q_1},\mathbf{q_2}} n_{-\mathbf{q_1}-\mathbf{q_2}} u_{\mathbf{q_1}} u_{\mathbf{q_2}}, \qquad (37)$$

where $u_{\mathbf{q}} = \sqrt{\omega_{\mathbf{q}}/2\rho s^2}(b_{\mathbf{q}} + b_{-\mathbf{q}}^\dagger)$, and D' is an energy of order unity in atomic units. The terms of the form $b_{\mathbf{q_1}}^\dagger b_{\mathbf{q_2}}$ describe Compton scattering of phonons by electrons. These processes have a negligible contribution to the cooling power and can be neglected.

Since $\mathbf{k}$, $\mathbf{k'}$ are on the Fermi surface, and the phonon momenta are large, $q_T \gg k_F$, the two phonons *are nearly counterpropagating* (see kinematics in Fig. 5). We can thus set $\mathbf{q_1} \approx -\mathbf{q_2} = \mathbf{q}$. This simplifies the matrix element $M = D'\omega_q/(2\rho s^2)$ yielding a transition probability

$$W_{\mathbf{k'},\mathbf{k}} = d \sum_{\mathbf{q}} |\langle \mathbf{k'} | \mathbf{k} \rangle|^2 \omega_{\mathbf{q}}^2 \left[N_{\omega_{\mathbf{q}}}^2 \delta_+ + (N_{\omega_{\mathbf{q}}} + 1)^2 \delta_- \right], \quad (38)$$

where $d = \frac{2\pi N D'^2}{\hbar(2\rho s^2)^2}$, $\delta_\pm = \delta(\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k'}} \pm 2\omega_{\mathbf{q}})$. Here the coherence factor is $|\langle \mathbf{k'} | \mathbf{k} \rangle|^2 = [1 \pm \cos(\theta_{\mathbf{k}} - \theta_{\mathbf{k'}})]/2$, with the plus (minus) sign for intra-band (inter-band) processes.

The two-phonon cooling power can be obtained from the transition probability by using Eq.(19). We find

$$\mathcal{J}_{2\text{ph}} = [\nu(\mu)]^2 d \int \frac{d^2 q}{(2\pi)^2} 4\omega_q^4 \mathcal{G}(T_{\text{el}}, T_{\text{ph}}), \qquad (39)$$
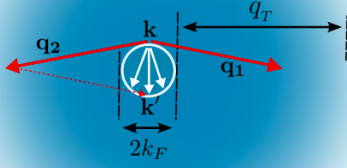
FIG. 5. Kinematics of two-phonon supercollisions wherein two nearly counterpropagating phonons, $\mathbf{q}_1$ and $\mathbf{q}_2$, are emitted. Here $\mathbf{k}, \mathbf{k}'$ denote the initial and final momenta of the electron, the blurred blue circle denotes the thermal distribution of phonons, and the white circle denotes the Fermi surface. The typical ratio of blue and white circle sizes is large $q_T/k_F = vk_BT/(s\mu) \approx 25$ (taking $T = 300\,\mathrm{K}$, and typical $\mu = 100\,\mathrm{meV}$) making $\mathbf{q}_1 \approx -\mathbf{q}_2$. Phonons can be acoustic phonons (see Sec. III F, III G) or flexural phonons (see Sec. III H).

where we defined

$$\mathcal{G}(T_{\mathrm{el}}, T_{\mathrm{ph}}) = (N_{\omega_{\mathbf{q}}} + 1)^2 \tilde{N}_{2\omega_{\mathbf{q}}} - N_{\omega_{\mathbf{q}}}^2 (\tilde{N}_{2\omega_{\mathbf{q}}} + 1). \quad (40)$$

We note that $\mathcal{G}$ vanishes when $T_{\mathrm{el}} = T_{\mathrm{ph}}$, satisfying detailed balance. Expanding in small $\Delta T = T_{\mathrm{el}} - T_{\mathrm{ph}}$ and integrating over $\mathbf{q}$ we obtain the ratio $\mathcal{J}_{2\mathrm{ph}}/\mathcal{J}_0 = (D'/D)^2(T/T_0)^5$, where $k_B T_0 = 0.86\,(\hbar^2 \rho s^4 k_B^2 T_{\mathrm{BG}}^2)^{1/5}$. Taking D' of order unity in atomic units, $D' \sim e^2/a \sim 10\,\mathrm{eV}$, we obtain a crossover temperature which for typical dopings is 5-10 times larger than our T_* estimate for disorder assisted processes (see Table I). This means that the two-phonon processes can dominate only in very clean systems. The contribution of multi-phonon processes to energy transport has been considered in Ref.[83].

G. Two Phonon Emission: Deep Off-Shell Processes

In this section we consider a two-phonon process in which two counterpropagating phonons are emitted (or absorbed) through a virtual state ("off shell" process). Each electron-phonon scattering vertex in this two-phonon process is described by the standard deformation coupling, see Eq.(16). As we now show, analyzing the transition amplitude for this process yields two deep off-shell processes that cancel at leading order. We can see

this by examining the matrix element for this process,

$$M = \langle \mathbf{k}' | M_{\mathbf{q_1}}^{(0)} G(\epsilon, \mathbf{q}_2) M_{\mathbf{q_2}}^{(0)} + M_{\mathbf{q_2}}^{(0)} G(\epsilon, \mathbf{q}_1) M_{\mathbf{q_1}}^{(0)} | \mathbf{k} \rangle. \quad (41)$$

Here $G(\epsilon, \mathbf{q}_1) = \frac{1}{\epsilon - H_0(\mathbf{k}-\mathbf{q_1})+\mu}$, $G(\epsilon, \mathbf{q}_2) = \frac{1}{\epsilon - H_0(\mathbf{k}-\mathbf{q_2})+\mu}$ where μ is Fermi energy. Since $|\mathbf{q_1}|, |\mathbf{q_2}| \approx q_T \gg k_F$, we can neglect $\mathbf{k}, \mathbf{k}'$ and the two phonons scattered counterpropagate, i.e. $\mathbf{q_1} \approx -\mathbf{q_2} = \mathbf{q}$ (see Fig. 5).

It is instructive to consider the undoped system first, $\mu = 0$. In this case, ignoring $\epsilon \sim kT$ compared to $H(q) \sim vq_T$, we have $G^{q_1} \approx -1/H_0(\mathbf{q})$ and $G^{q_2} \approx -1/H_0(-\mathbf{q})$. We also note that the vertex $M^{(0)}$ is an identity in graphene's pseudospin space and it commutes with H_0. Since $H_0(-\mathbf{q}) = -H_0(\mathbf{q})$, the matrix element and the resulting transition probability vanishes. This, in turn, means that the leading contribution to cooling power for this process vanishes.

Importantly, the above argument is not limited to the undoped case, since $vq_T \gg \mu$ for typical carrier densities and temperatures. Hence, even when the system is doped away from neutrality, the off-shell contribution is still negligible, see discussion below. On a conceptual level, the cancellation between the two terms in Eq.(41) may appear counterintuitive since the positive and negative contribution arise from the states above and below the Fermi level. We note in that regard, that the cancellation is made possible because no occupation factors are associated with $G(\epsilon, \mathbf{q}_1)$ and $G(\epsilon, \mathbf{q}_2)$ since they describe deep off-shell states as above. As a result, the nearly counterpropagating phonons allow the positive and negative energy contributions in the virtual states to cancel each other viz. particle-hole symmetry.

In a similar vein, we can consider Compton-like processes in which phonons are absorbed and then re-emitted. The transition amplitude for such a process, similar to our discussion of two-phonon emission and absorption, is given by a sum of two contributions, one describing phonon absorption followed by phonon emission and the other describing phonon emission followed by phonon absorption. In contrast with the two-phonon emission and absorption discussed above, phonon Compton scattering is dominated by processes in which the absorbed and emitted phonon have momenta that co-propagate. In complete analogy with our discussion above, the symmetry $H_0(-\mathbf{q}) = -H_0(\mathbf{q})$ results in the transition matrix element vanishing at leading order. Interference between the term describing absorption and subsequent emission of a phonon and the term describing emission and subsequent absorption of a phonon resembles the QED treatment of Compton scattering.

While the leading order contributions vanish for these processes, in each case there exists a residual contribution. Indeed, restoring finite ϵ to the Greens functions gives a non-vanishing sum

$$G(\epsilon, \mathbf{q}_1) + G(\epsilon, \mathbf{q}_2) \approx \frac{2(\epsilon + \mu)}{[\epsilon - H_0(\mathbf{q}) + \mu][\epsilon + H_0(\mathbf{q}) + \mu]}, \quad (42)$$

where we have imposed $\mathbf{q_1} \approx -\mathbf{q_2} = \mathbf{q}$ for counterpropagating phonons as before.

We can follow the analysis used above to evaluate the cooling power for this residual contribution. In doing so, we note that the main contributions to supercollision cooling arise from momenta $|\mathbf{q}| \approx q_T$. As a result, $\epsilon, \mu \ll vq_T$ in the degenerate limit ($\mu \gg k_B T$) allows us to simplify the matrix element so that $G(\epsilon, \mathbf{q_1}) + G(\epsilon, \mathbf{q_2}) \approx -2\mu/[vq_T]^2$ in Eq.(42). This yields an estimate for the residual cooling power as

$$\mathcal{J}_{2ph}^{residual} = C\Big(T_{el}^2 - T_{ph}^2\Big), \quad C = \frac{N[\nu(\mu)]^2\pi^2 s^2 g^4 \mu^2 k_B^2}{3\hbar v^4}. \tag{43}$$

Interestingly, $\mathcal{J}_{2ph}^{residual}$ scales the same way with density as single acoustic phonon cooling, $\mathcal{J}_0$. Comparing $\mathcal{J}_{2ph}^{residual}$ above to $\mathcal{J}_0$ in Eq.(20), by linearizing in small ΔT, we estimate that $\mathcal{J}_{2ph}^{residual}$ wins over $\mathcal{J}_0$ for $T > T_* \approx 5500$ K. Here, we have used values for the electron-phonon coupling, g, detailed above. We therefore conclude that the residual contribution is irrelevant.

H. Flexural Phonon Supercollisions

Finally, we analyze cooling in free standing graphene in the absence of disorder. In this case, an important contribution arises due to flexural phonons,[84,85] which contribute to the deformation tensor via $u_{ij} = 1/2(\partial_i u_j + \partial_j u_i + \partial_i h \partial_j h)$, with $\mathbf{u}$ and h the in-plane and out-of-plane displacements. Flexural modes have quadratic dispersion $\tilde{\omega}_\mathbf{q} = \kappa_{flex}|\mathbf{q}|^2$ with $\kappa_{flex} \approx 4.6 \times 10^{-7}$ m^{2}s^{-1}.[84,85] Electron-phonon coupling is described by the same deformation potential as above, Eq.(16).

The processes involving pairs of nearly counterpropagating flexural phonons are analyzed as follows. Using the momentum representation, $h_\mathbf{q} = \sqrt{\hbar/2\rho\tilde{\omega}_q}(b_\mathbf{q} + b_{-\mathbf{q}}^\dagger)$, we consider the emission/absorption of two flexural phonons with momenta $\mathbf{q_1}$ and $\mathbf{q_2}$. For $T \gg T_{BG}^{flex} = \hbar\kappa_{flex}k_F^2$ (for typical densities, T_{BG}^{flex} is well below 1 K), we can set $\mathbf{q_1} \approx -\mathbf{q_2} = \mathbf{q}$ so that the matrix elements in a transition rate are $M = \frac{D\hbar}{4\rho\kappa_{flex}}$.[84,85] Integrating to obtain the cooling power yields[39]

$$\mathcal{J}_{flex} = \sum_q (2\hbar\tilde{\omega}_q)^2 \left[(N_{\omega_\mathbf{q}} + 1)^2 N_{2\omega_\mathbf{q}}^{el} - N_{\omega_\mathbf{q}}^2 (N_{2\omega_\mathbf{q}}^{el} + 1)\right]$$

where $\sum_q ... = \frac{\pi N D^2 \hbar^2}{16\rho^2 \kappa_{flex}^2}\big[\nu(\mu)\big]^2 \int \frac{d^2q}{(2\pi)^2}...$, and the expression vanishes for $T_{el} = T_{ph}$.

For easy comparison, $\mathcal{J}_{flex}$ can be linearized in $\Delta T = T_{el} - T_{ph}$ giving

$$\mathcal{J}_{flex} = A_{flex}T^2\Delta T, \qquad A_{flex} = 0.12\frac{N D^2 \nu^2(\mu) k_B^3}{\rho^2 \kappa_{flex}^3}, \tag{44}$$

which scales with T the same way as Eq.(24). Flexural phonons dominate over the one-phonon contribution at

$$T > T_*^{flex} = \sqrt{\frac{B}{A_{flex}}} = \left(\frac{\pi\rho\kappa_{flex}^3}{0.24\hbar s^2}\right)^{1/2} T_{BG} \approx 10\,T_{BG}. \tag{45}$$

The value T_*^{flex} is considerably larger than T_* for disorder-assisted cooling estimated above (see Table I). A comparison with Eq. (24) yields $\mathcal{J}_{flex}/\mathcal{J}_{dis} \approx k_F\ell/200$ which is small for typical $k_F\ell$. Thus the contribution Eq. (44) is relatively weak under realistic conditions. For graphene on substrate this contribution is further diminished as flexural modes get pinned, gapped, and stiffened by the substrate.

IV. CONCLUSION

As we have seen above, graphene provides a natural setting to realize and explore long-lived hot carriers. Arguably, hot carriers is one of the most exciting and promising new developments in graphene research that relies on the unique properties of this material. However, on a practical implementation level, it remains to be seen whether graphene hot carriers are a pack of dogs that can bark. This will happen as more new phenomena triggered by the presence of hot carriers are uncovered. Hot carriers have been proposed as a means to enhance optoelectronic efficiency,[11,13,51] and manipulate energy flows in new ways.[14,17] One can envision several other avenues where hot carrier research may leave a mark. We anticipate new transport phenomena mediated by long-range energy currents along the lines of Refs.14 and 17, in particular new kinds of photoresponse. Another interesting direction is energy harvesting in 2D crystals, such as graphene. One approach which is now actively discussed is vertical extraction of photogenerated hot carriers in field-effect or sandwich-type structures.[7] It is also of interest to explore various knobs by which the key properties of hot carriers such as cooling times can be altered and controlled.[39] Given the wide range of lattice temperatures for which hot carriers can be realized, extending up to room temperature, we have many reasons to expect that this research will eventually transcend into the practical realm.

---

[1] H. C. Andersen, The Gardener and the Manor, (1872).
[2] E. Pop, Nano Research, **3** 147 (2010).
[3] W. Shockley and H. J. Queisser, J. App. Phys., **32**, 510 (1961).
[4] A. K. Geim and I. V. Grigorieva, Nature, **499** 419 (2013).
[5] M. Bernardi, M. Palummo, and J. C. Grossman, Nano Letters, **13** 3664 (2013).
[6] Bonaccorso, F. and Sun, Z. and Hasan, T. and Ferrari, A. C., Nature Photonics, **4** 611 (2010).
[7] L. Britnell, *et al.* Science, **335** 947-950 (2012).
[8] Y. M. Zuev, W. Chang, and P. Kim, Phys. Rev. Lett. **102**, 096807 (2009).
[9] P. Wei, W. Bao, Y. Pu, C. N. Lau, and J. Shi, Phys. Rev. Lett. **102**, 166808 (2009).
[10] N. M. Gabor, *et. al.*, Science, **334**, 648 (2011).
[11] J. C. W. Song, M. S. Rudner, C. M. Marcus, L. S. Levitov, Nano Lett., **11**, 4688 (2011).
[12] D. Sun, *et. al.*, Nat. Nano., **7** 114 (2012).
[13] K.J. Tielrooij, *et al.*, Nat. Phys., **9** 248 (2013).
[14] J Renard, M Studer, and J. A. Folk, Phys. Rev. Lett., **112** 116601 (2014).
[15] J. Yan, *et al.*, Nature Nano., **7** 472 (2012).
[16] P. Herring, *et. al*, NanoLett. **14** 901 (2014).
[17] J. C. W. Song, and L. S. Levitov, Phys. Rev. Lett. **109**, 236602 (2012).
[18] R. Bistritzer, A. H. MacDonald, Phys. Rev. Lett., **102** 206410 (2009).
[19] W.-K. Tse, S. Das Sarma, Phys. Rev. B, **79**, 235406 (2009).
[20] J. Shah, A. Pinczuk, A. C. Gossard, and W. Wiegmann, Phys. Rev. Lett, **54**, 2045 (1985).
[21] D. R. Leadley, R. J. Nicholas, J. J. Harris, and C. T. Foxon, Semicond. Sci. Technol., **4**, 879 (1989).
[22] J. Lutz, *et. al.*, Semicond. Sci. Technol., **8**, 399 (1993).
[23] M. L. Roukes, *et al.*, Phys. Rev. Lett. **55**, 422 (1985)
[24] I. Gierz. *et al.*, Nat. Mat., **12** 1119(2013).
[25] J.C. Johannsen, *et al.*, Phys. Rev. Lett., **111** 027403 (2013).
[26] M. W. Graham, S.-F. Shi, D. C. Ralph, J. Park, and P. L. McEuen, Nat. Phys., **9**, 103 (2013).
[27] M. W. Graham, *et. al.*, NanoLett., **13**, 5497 (2013).
[28] T. Kampfrath L. Perfetti, F. Schapper, C. Frischkorn, and M. Wolf, Phys. Rev. Lett. **95**, 187403 (2005).
[29] S. Butscher , F. Milde, M. Hirtschulz, E. Malić, and A. Knorr, App. Phys. Lett. 91, 203103 (2007).
[30] F. Rana, Phys. Rev. B, **76**, 155431 (2007).
[31] P. A. George *et al.*, NanoLett. 8, 4248 (2008).
[32] F. T. Vasko, and V. Ryzhii, Phys. Rev. B **77**, 195433 (2008).
[33] M. Breusing, C. Ropers, and T. Elsaesser, Phys. Rev. Lett. **102**, 086809 (2009).
[34] C. H. Lui, K. F. Mak, J. Shan, and T. F. Heinz, Phys. Rev. Lett., **105** 127404 (2010).
[35] T. Winzer, A. Knorr, and E. Malić, NanoLett., **10** 4839 (2010).
[36] R. Kim, V. Perebeinos, P. Avouris, Phys. Rev. B, **84** 075449 (2011).
[37] S. Winnerl, *et al.* Phys. Rev. Lett., **107** 237401 (2011).
[38] T. Winzer and E. Malić, Phys. Rev. B, **85** 241404 (2012).
[39] J. C. W. Song, M. Y. Reizer, and L. S. Levitov, Phys. Rev. Lett., **109** 106602 (2012).
[40] A.C. Betz, *et. al.*, Nat. Phys., **9** 109 (2013).
[41] Q. Ma, *et al.*, Phys. Rev. Lett., **112** 247401 (2014).
[42] F. Rana, *et al*, Phys. Rev. B **79** 115447 (2009).
[43] D. Brida, *et al.*, Nat. Comm., **4** 1987 (2013).
[44] J.C.W. Song, K.J. Tielrooij, F.H.L. Koppens, and L.S. Levitov, Phys. Rev. B, **87** 155429 (2013).
[45] W.-K. Tse, E. H. Hwang, and S. Das Sarma, App. Phys. Lett., **93**, 023128 (2008).
[46] E. H. Hwang, and S. Das Sarma, Phys. Rev. B, **75** 205418 (2007).
[47] M. S. Foster and I. L. Aleiner, Phys. Rev. B, **79** 085415 (2009).
[48] A. Tomadin, D. Brida, G. Cerullo, A. C. Ferrari, and M. Polini, Phys. Rev. B, **88** 035430 (2013).
[49] D. M. Basko, Phys. Rev. B, **87** 165437 (2013).
[50] T. Ando, J. Phys. Soc. Jpn., **75** 124701 (2006).
[51] R. T. Ross and A. J. Nozik, J. App. Phys., **53** 3813 (1982).
[52] H. Choi, *et al.*, Appl. Phys. Lett. **94** 172102 (2009).
[53] J. H. Strait, *et. al.*, *Nano Lett.*, **11**, 4902 (2011).
[54] H. Y. Hwang, *et al.*, J Phys. Chem. B, **117**, 15819 (2013)
[55] A. J. Frenzel, *et. al.*, Appl. Phys. Lett., **102** 113111 (2013).
[56] G. Jnawali, Y. Rao, H. Yan, and T. F. Heinz, NanoLett., **13** 524 (2013).
[57] S.-F Shi *et al.*, NanoLett., **14** 1578 (2014).
[58] A. J. Frenzel, C. H. Lui, Y. C. Shin, J. Kong, and N. Gedik, Phys. Rev. Lett. **113**, 056602 (2014).
[59] R. Ulbricht *et al.*, Rev. Mod. Phys. **83**, 543-586 (2011).
[60] K. Nomura, and A. H. MacDonald, Phys. Rev. Lett. **96**, 256602 (2006).
[61] T. Ando, J. Phys. Soc. Jpn. **75**, 074716 (2006).
[62] S. Das Sarma, S. Adam, E. H. Hwang, and E. Rossi, Rev. Mod. Phys. **83**, 407 (2011).
[63] M. Freitag, T. Low, F. Xia, and P. Avouris, Nat. Photonics, **7** 53 (2013).
[64] V. Perebeinos, and P. Avouris, Physical Review B, **81** 195442 (2010).
[65] S. Fratini, and F. Guinea, Phys. Rev. B, **77**, 195415 (2008).
[66] A.S. Price, S.M. Hornett, A.V. Shytov, E. Hendry, and D.W. Horsell, Phys. Rev. B, **85** (2012).
[67] A. F. Morpurgo, and F. Guinea, Phys. Rev. Lett. **97**, 196804 (2006).
[68] M. I. Katsnelson, and A. Geim, Phil. Trans. R. Soc. A **366**, 195 (2008).
[69] E.-A. Kim, A.H. Castro Neto, Europhys. Lett., **84** 57007 (2008).
[70] M.A.H. Vozmediano, M.I. Katsnelson, and F. Guinea, Phys. Rep. **496**, 109 (2010).
[71] M. Lundeberg, J. Folk Phys. Rev. Lett. **105** 146804 (2010).
[72] M. Ishigami, *et. al.*, NanoLett. **7**, 1644 (2007).
[73] E.H. Hwang, and S. Das Sarma, Phys. Rev. B, **77** 115449 (2008).
[74] J. Ziman, Principles of the Theory of Solids. (Cambridge University Press, 1972).

[75] S. S. Kubakaddi, Phys. Rev. B, **79** 075417 (2009).

[76] J. K. Viljas and T. T. Heikkilä, Phys. Rev. B, **81** 245404 (2010).

[77] K. C. Fong, and K. C. Schwab, Phys. Rev. X, **2** 031006 (2012).

[78] A.C. Betz, *et al.*, Phys. Rev. Lett. **109** 056805 (2012).

[79] R. Somphonsane, *et al.*, NanoLett. **13** 4305 (2013)

[80] A. Laitinen, *et. al.*, NanoLett., **14** 3009 (2014).

[81] N. Bonini, M. Lazerri, N. Marzari, F. Mauri, Phys. Rev. Lett., **99**, 176802 (2007).

[82] S. Pisana, *et. al.*, Nat. Mat., **6**, 198 (2007).

[83] P. Virtanen, arXiv: 1312.3833 (2013).

[84] H. Ochoa, E. V. Castro, M. I. Katsnelson, F. Guinea, Phys. Rev. B, **83**, 235416 (2011).

[85] E. V. Castro, *et. al.*, Phys. Rev. Lett., **105**, 266601 (2010).

[86] We use this opportunity to correct a factor of $\sqrt{2}$ in the denominator of g_0 in Eq. 16 of Ref. 44. Accounting for the new coupling constant, the estimate of the scattering rate yields $\tau_0 \approx 350\,\mathrm{fs}$ (see also Eq. 11 in this text).