Ultrafast electronic heat dissipation through surface-to-bulk Coulomb coupling in quantum materials

Alessandro Principi[1,*] and Klaas-Jan Tielrooij[2]

[1]*Department of Physics and Astronomy, University of Manchester, Manchester M13 9PL, United Kingdom*

[2]*Catalan Institute of Nanoscience and Nanotechnology (ICN2), BIST & CSIC, Campus UAB, 08193 Bellaterra, Barcelona, Spain*

(Received 18 June 2022; revised 13 August 2022; accepted 2 September 2022; published 16 September 2022)

The timescale of electronic cooling is an important parameter controlling the performance of devices based on quantum materials for optoelectronic, thermoelectric, and thermal management applications. In most conventional materials, cooling proceeds via the emission of phonons, a process that can bottleneck the carrier relaxation dynamics, thus degrading the device performance. Here we present the theory of near-field radiative heat transfer that occurs when a two-dimensional electron system is coupled via the nonretarded Coulomb interaction to a three-dimensional bulk that can behave as a very efficient electronic heat sink. We apply our theory to study the cooling dynamics of surface states of three-dimensional topological insulators and of graphene in proximity to small-gap bulk materials. The "Coulomb cooling" we introduce is alternative to the conventional phonon-mediated cooling, can be very efficient, and can dominate the cooling dynamics under certain circumstances. We show that this cooling mechanism can lead to a sub-picosecond timescale, significantly faster than the cooling dynamics normally observed in Dirac materials.

DOI: 10.1103/PhysRevB.106.115422

I. INTRODUCTION

Two-dimensional (2D) Dirac materials [1–4] have been extensively studied in the past two decades for the diverse range of intriguing properties they harbor, which could in turn enable a wealth of novel practical applications [5–8]. This is the case of graphene [1,9–11], arguably the most-studied Dirac material, which continues to attract significant interest because of its potential for electronic and optoelectronic applications [11]. However, graphene is just an instance of a broad family of systems which includes the 2D surface states of three-dimensional (3D) topological insulator [2,3,12]. Their inverted bulk band structure allows states to localize at the surface [5,13–16]. In ideal topological insulators, surface states do not hybridize with bulk ones, and are topologically protected against any perturbation that preserves the symmetries of the bulk [2,3,12].

Thanks to the coupling between kinetic momentum and spin, the electronic surface states of a 3D topological insulator present a physics potentially richer than graphene, and promise application to a diverse range of fields, including spintronics [6,7], optoelectronics, and photonics [17,18]. In a recent experimental work [19], the cooling dynamics of surface electrons of bismuth and antimony chalcogenides was studied with pump-probe techniques. Surprisingly, electronic heat relaxation faster than that of bulk carriers was observed. The observed bulk and surface heat-decay rates differ by about an order of magnitude, while the environment experienced by their electrons is the same. Interestingly, a decay rate of a few hundred femtoseconds was obtained under rather strong photoexcitation with a fluence on the order of 100 μJ/cm^2. In comparison, when graphene is excited with a similar fluence, cooling is rather slow—several picoseconds—due to a phonon bottleneck effect [20].

So why does the cooling of the electrons of surface states of topological insulators occur so much faster than their bulk counterpart or than graphene? In graphene, the cooling of photoexcited electrons is ultimately limited by the emission of intrinsic optical and acoustic phonons of the material [21–23] or of the encapsulant or substrate [24–26]. In topological insulators, cooling is also typically ascribed to phonon emission on a picosecond timescale [27]. Such phonon emission mechanisms fail to account for the starkly different relaxation dynamics of surface and bulk electrons of topological insulators, unless one would postulate different electron-phonon couplings or a reduced phase space for phonon emission by surface states.

In this paper, we explore a different mechanism that does not require fine-tuning material parameters, as it relies on electronic systems exclusively. This alternative mechanism is based on the notion that, contrary to electrons in conventional graphene devices, the surface electrons of topological insulators are in close proximity to a macroscopic bulk whose particle-hole excitations can occur at energies comparable to the (surface) thermal excitations. Thanks to the near-field radiative coupling between surface and bulk, the latter can act as a heat sink. Being macroscopic, the bulk can absorb large amounts of heat and therefore efficiently cool down the surface electrons. In this picture, heat is dissipated into particle-hole excitations of the bulk via nonretarded Coulomb interactions (see Fig. 1). Thus we term this mechanism "Coulomb cooling."

*alessandro.principi@manchester.ac.uk

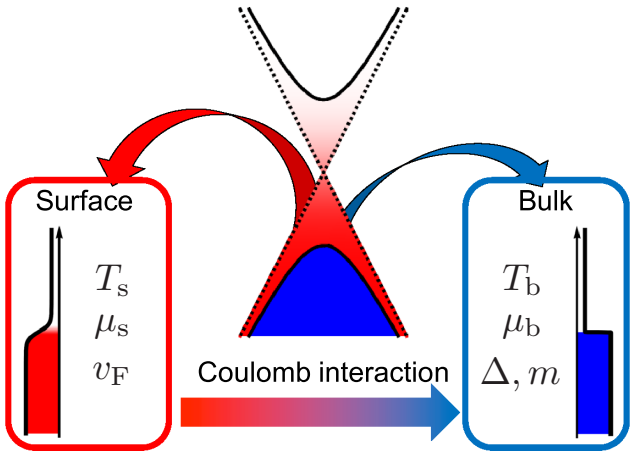

FIG. 1. Schematics of Coulomb cooling of hot surface Dirac fermions into bulk states. Dirac fermions have a constant Fermi velocity v_F, and they exhibit a hot thermal distribution characterized by temperature and chemical potential T_s and μ_s, respectively. Similarly, bulk states are described in terms of the effective mass m and band gap 2Δ. The Fermi distribution of bulk states has chemical potential μ_b (set to zero in later calculations) and temperature $T_b < T_s$.

We note that for atomic-scale heterogeneous systems with massless electrons, such as graphene interfaced with other 2D materials, e.g., hexagonal boron nitride or WS_2, or 3D dielectric substrates, e.g., SiO_2, different relaxation processes have been described. This includes cooling of hot electrons in graphene via polar [24] and hyperbolic [25,26] substrate phonons, as well as thermionic emission [28] and direct charge transfer [29]. In bare monolayer graphene suspended over a hole, hot carrier cooling occurs via optical phonon emission, followed by coupling to acoustic phonons [23]. These mechanisms thus involve either electron-phonon coupling or charge transfer. The mechanism we describe here for 2D-3D surface-bulk TIs and graphene-TI systems does not involve electron-phonon coupling nor charge transfer. Rather it relies purely on Coulomb interaction, which so far has only been explored in multilayer graphene [30].

In the remainder of the paper, we consider a system of 2D massless Dirac fermions in proximity to a 3D gapped bulk. The two are coupled only electrostatically via nonretarded Coulomb interactions. This minimal model of a topological insulator equivalently describes a graphene sheet in proximity to a small-gap material. For both cases, using the kinetic equation for surface electrons, we derive an expression for their cooling rate. This is controlled by the convolution of surface and bulk particle-hole excitation spectra. The higher their overlap, the larger the amount of heat transferred per unit time. We show that the cooling rate reaches a maximum for surface temperatures close to half the bulk-band gap. This is interpreted as a resonance between surface electronic transitions, whose typical energy is the thermal one, and interband particle-hole bulk excitations across the band gap. We find a timescale of a few hundred femtoseconds for topological insulators, and even a few tens of femtoseconds for graphene on a small-bandgap semiconductor.

II. THE MODEL

We model both the surface states of the topological insulator and the electrons in graphene as a gas of massless Dirac fermions [1–3,9–12], i.e.,

$$\mathcal{H}_s = \hbar v_F \sum_{\boldsymbol{k},\alpha,\beta} \hat{\psi}^\dagger_{\boldsymbol{k},\alpha,s} \boldsymbol{k}\cdot\boldsymbol{\sigma}_{\alpha,\beta}\hat{\psi}_{\boldsymbol{k},\beta,s}, \tag{1}$$

where v_F is the Fermi velocity and $\boldsymbol{\sigma}$ is a vector of Pauli matrices. These act in the real spin space for topological insulators, and in pseudospin (sublattice) space for graphene. In Eq. (1), $\hat{\psi}^\dagger_{\boldsymbol{k},\alpha,s}$ ($\hat{\psi}_{\boldsymbol{k},\alpha,s}$) creates (destroys) a surface particle of two-dimensional wave vector $\boldsymbol{k}$ (momentum $\hbar\boldsymbol{k}$) and (pseudo)spin projection α. The band energy is $\varepsilon^{(s)}_{\boldsymbol{k},\lambda} = \lambda\hbar v_F|\boldsymbol{k}|$, where $\lambda = +$ ($\lambda = -$) denotes the surface conduction (valence) band. [The Hamiltonian (1) is obtained from the usual one for topological insulators, which features a cross product between momentum and spin operators [2], by redefining the spin quantization axes.] Equation (1) is in form identical for graphene and topological insulators. However, their electrons exhibit different degeneracies N_f. While the surface states of topological insulators are helical [2], thus yielding $N_f = 1$, in graphene they feature a full spin-valley degeneracy [1]. Thus, for graphene, the number of fermion flavors is $N_f = 4$.

We model the bulk as a two-band system confined into the half-space $z > 0$ and described by the Hamiltonian

$$\mathcal{H}_b = \sum_{\boldsymbol{k},\eta} \varepsilon^{(b)}_{\boldsymbol{k},\eta}\hat{\psi}^\dagger_{\boldsymbol{k},\eta,b}\hat{\psi}_{\boldsymbol{k},\eta,b}. \tag{2}$$

Here, $\varepsilon^{(b)}_{\boldsymbol{k},\eta} = \eta[\Delta + \hbar^2|\boldsymbol{k}|^2/(2m)]$, where 2Δ is the bulk-band gap and m is the bulk-band mass (assumed to be the same for both valence and conduction bands), while $\hat{\psi}^\dagger_{\boldsymbol{k},\eta,b}$ ($\hat{\psi}_{\boldsymbol{k},\eta,b}$) creates (destroys) a bulk particle of three-dimensional wave vector $\boldsymbol{k} = (\boldsymbol{k}_\parallel, k_z)$, in band η. Here, $\boldsymbol{k}_\parallel$ and k_z are the wave vectors parallel and perpendicular to the surface $z = 0$, $\eta = +$ denotes the conduction band, while $\eta = -$ stands for the valence band. For simplicity, we assume that bulk bands are spherically symmetric and have no spin structure. These approximations do not affect the temperature dynamics on a qualitative level. Assuming specular reflection at the interface $z = 0$, the bulk eigenstates acquire the form of standing waves: $\Psi_{\boldsymbol{k},\eta,b}(\boldsymbol{r}, z) = \sqrt{2/V}e^{i\boldsymbol{k}_\parallel\cdot\boldsymbol{r}}\sin(k_z z)$, where $\boldsymbol{r}$ is a vector along the surface.

We assume that bulk and surface electrons are coupled electrostatically by long-range instantaneous Coulomb interactions [31], whose Hamiltonian is

$$\mathcal{H}_{sb} = \frac{1}{2}\int_0^\infty dz \sum_{\boldsymbol{q}} V_{sb}(\boldsymbol{q}, z)\hat{n}_{\boldsymbol{q},s}\hat{n}_{-\boldsymbol{q},b}(z). \tag{3}$$

Here, $\boldsymbol{q}$ is a two-dimensional wave vector along the surface of the topological insulator, while $\hat{n}_{\boldsymbol{q},s}$ and $\hat{n}_{-\boldsymbol{q},b}(z)$ are the 2D-Fourier transforms of the surface and bulk density operators. The precise form of the interaction $V_{sb}(\boldsymbol{q}, z)$ is determined by the solution of the associated Poisson (electrostatic) problem, as we proceed to show.

The interaction between surface and bulk electrons

In this section we derive the Coulomb interaction between surface and bulk electrons. To do so, we consider the electrostatic problem of a single charge added to a conducting 2D sheet (the surface states located at $z = 0$) placed on top of the topological-insulator bulk which fills the half-space $z > 0$. The half-space $z < 0$ is instead empty. The single charge is added as a plane wave of wave vector $\boldsymbol{q}$ in the surface sheet. In response to this added charge, induced bulk and surface densities are generated. These are named $n_{\rm b}(\boldsymbol{r}, z)$ and $n_{\rm s}(\boldsymbol{r})$, respectively. The resulting Poisson equation (in Gaussian units) is

$$\boldsymbol{\nabla} \cdot [\epsilon(z)\boldsymbol{\nabla}\phi(\boldsymbol{r}, z)] = -4\pi e^2 n(\boldsymbol{r}, z), \tag{4}$$

where $n(\boldsymbol{r}, z) = n_{\rm b}(\boldsymbol{r}, z)\Theta(z) + n_{\rm s}(\boldsymbol{r})\delta(z) + e^{-i\boldsymbol{q}\cdot\boldsymbol{r}}\delta(z)$ is the total electron density, and $\epsilon(z) = \epsilon_{\rm b}\Theta(z) + \epsilon_{\rm vac}\Theta(-z)$ is the dielectric function. Here, $\epsilon_{\rm b}$ is the relative dielectric constant of the topological insulator, which accounts for the screening due to filled bands, while $\epsilon_{\rm vac} = 1$ [$\Theta(z)$ is the Heaviside step function]. The Fourier transform of Eq. (4) in the direction parallel to the topological-insulator surface (or the graphene sheet) yields

$$\partial_z[\epsilon(z)\partial_z\phi_{\boldsymbol{q}}(z)] - q^2\epsilon(z)\phi_{\boldsymbol{q}}(z) = -4\pi e^2 n_{\boldsymbol{q}}(z), \tag{5}$$

where $n_{\boldsymbol{q}}(z) = n_{\boldsymbol{q},{\rm b}}(z)\Theta(z) + n_{\boldsymbol{q},{\rm s}}\delta(z) + \delta(z)$. Within linear response, we write $n_{\boldsymbol{q},{\rm s}} = \chi_{\rm s}(\boldsymbol{q}, \omega)\phi_{\boldsymbol{q}}(0)$, where $\chi_{\rm s}(\boldsymbol{q}, \omega)$ is the density-density response function (polarizability) of the surface states, and

$$n_{\boldsymbol{q},{\rm b}}(z) = \int_0^\infty dz'\,\chi_{\rm b}(\boldsymbol{q}, \omega, z, z')\phi_{\boldsymbol{q}}(z') \simeq \bar{\chi}_{\rm b}(\boldsymbol{q}, \omega)\phi_{\boldsymbol{q}}(z). \tag{6}$$

Here, $\chi_{\rm b}(\boldsymbol{q}, \omega, z, z') \simeq \bar{\chi}_{\rm b}(\boldsymbol{q}, \omega)\delta(z - z')$ is the density-density response function [32] (polarizability) of bulk states treated within a semilocal approximation. This approximation allows us to reduce the integro-differential equation to a set of differential ones. Behind it, there is the assumption that the potential and induced charge density are sufficiently localized in the z direction. The result in Eq. (9) is consistent with our assumption and justifies *a posteriori* its use. We do not need to introduce any approximation in the direction parallel to the surface since we can simply take a Fourier transform and reduce equations to algebraic ones. However, even doing so we find [see below Eq. (10)] that the potential approaches a constant at small q. This means that the potential decays fast at large enough distances. Thus, the potential and induced density are localized in space.

Equations of this section contain the response of the doped bulk, embodied by $\chi_{\rm b}(\boldsymbol{q}, \omega, z, z')$ and $\bar{\chi}_{\rm b}(\boldsymbol{q}, \omega)$, for completeness. However, we leave this function unspecified since, as we show below, it plays no role in the undoped regime when the bulk thermal energy ($k_{\rm B}T_{\rm b}$) is much smaller than the bulk-band gap (2Δ). Using these expressions, the Poisson Eq. (5) can be split into the two half-spaces as

$$\begin{cases} \epsilon_{\rm b}\left(\partial_z^2 - q^2\right)\phi_{\boldsymbol{q}}(z) = -4\pi e^2\bar{\chi}_{\rm b}(\boldsymbol{q}, \omega)\phi_{\boldsymbol{q}}(z), & \text{if } z > 0, \\ \left(\partial_z^2 - q^2\right)\phi_{\boldsymbol{q}}(z) = 0, & \text{if } z < 0. \end{cases} \tag{7}$$

The boundary conditions, at the surface $z = 0$ and at $z \to \pm\infty$, are $\phi_{\boldsymbol{q}}(0^+) = \phi_{\boldsymbol{q}}(0^-)$, $\phi_{\boldsymbol{q}}(z \to \pm\infty) = 0$, and

$$\epsilon_{\rm b}\partial_z\phi_{\boldsymbol{q}}(z)|_{z\to 0^+} - \partial_z\phi_{\boldsymbol{q}}(z)|_{z\to 0^-} = -4\pi e^2(n_{\boldsymbol{q},{\rm s}} + 1). \tag{8}$$

Introducing the bulk Thomas-Fermi wave vector [32] $q_{\rm TF}^2(q, \omega) = -4\pi e^2\bar{\chi}_{\rm b}(\boldsymbol{q}, \omega)/\epsilon_{\rm b}$, the solution of Eqs. (7) with the boundary conditions above is

$$\begin{cases} \phi_{\boldsymbol{q}}(z) = \bar{\phi}_{\boldsymbol{q}}e^{-\sqrt{q^2+q_{\rm TF}^2(q,\omega)}z}, & \text{if } z > 0, \\ \phi_{\boldsymbol{q}}(z) = \bar{\phi}_{\boldsymbol{q}}e^{qz}, & \text{if } z < 0. \end{cases} \tag{9}$$

Equation (9) shows that both the potential and the induced density (which is linearly related to the potential) are very localized in the z direction. This result is consistent with our initial semilocal assumption and serves as an *a posteriori* justification of its use. Using Eq. (8), we get

$$\bar{\phi}_{\boldsymbol{q}} = \frac{4\pi e^2}{\epsilon\sqrt{q^2 + q_{\rm TF}^2(q, \omega)} + q - 4\pi e^2\chi_s(\boldsymbol{q}, \omega)}. \tag{10}$$

In the limit of zero frequency and small wave vectors, $\chi_{\rm s}(\boldsymbol{q}, \omega)$ becomes the densities of states of surface electrons. Equation (9) is also applicable to the situation in which a graphene layer is separated by a layer of vacuum (or air) of thickness d from the topological insulator surface. The interaction between electrons in the two systems is therefore weighted with the exponential e^{-qd}.

III. THE COOLING RATE—GENERAL THEORY

We consider the kinetic equation for electrons at the surface of the topological insulator, which interact with those in the bulk via Coulomb interactions. The surface and bulk electrons are described by the Fermi distribution functions $f_{\boldsymbol{k},\lambda}^{({\rm s})}$ and $f_{\tilde{\boldsymbol{k}},\eta}^{({\rm b})}$ at the temperatures $T_{\rm s}$ and $T_{\rm b}$ and chemical potentials $\mu_{\rm s}$ and $\mu_{\rm b}$, respectively. As in Sec. II, $\boldsymbol{k}$ and $\tilde{\boldsymbol{k}}$ are their two- and three-dimensional wave vectors, respectively. We recall that $\lambda = \pm 1$ is used to denote the two surface bands (together forming a Dirac cone), while $\eta = \pm 1$ is used for the bulk conduction and valence bands. For the calculation of the cooling time, we assume that no driving field is present and that material parameters (including temperatures and chemical potentials) are isotropic. However, since the two populations are at different uniform chemical potentials and temperatures, we can study the time evolution of their distribution functions. The kinetic equation satisfied by $f_{\boldsymbol{k},\lambda}^{({\rm s})}$ is [25]

$$\partial_t f_{\boldsymbol{k},\lambda}^{({\rm s})} = -\mathcal{I}_{\boldsymbol{k},\lambda}^{({\rm sb})}, \tag{11}$$

where $\mathcal{I}_{\boldsymbol{k},\lambda}^{({\rm sb})}$ is the electron-electron collision integral between surface and bulk electrons [resulting from the interaction in Eq. (3)]. This conserves their numbers separately but allows for the exchange of energy between them. We will specify the collision integral in the following subsection. First, however, we will derive the general expression for the cooling rate.

To obtain the cooling rate, we multiply Eq. (11) by the energy of the surface state, $\varepsilon_{\boldsymbol{k},\lambda}^{({\rm s})}$, and sum over all wave vectors

$\boldsymbol{k}$ and all values of the surface-band index λ. The left-hand side of the so-obtained equation yields the time derivative of the energy stored in surface states, $\partial_t E_\mathrm{s}$. The latter is rewritten as $\partial_t E_\mathrm{s} = C_\mathrm{s}\partial_t T_\mathrm{s}$, where the heat capacity of surface states is defined as [25]

$$C_\mathrm{s} = \sum_\lambda \int \frac{d^2\boldsymbol{k}}{(2\pi)^2} \xi^{(\mathrm{s})}_{\boldsymbol{k},\lambda} \left(-\frac{\partial f^{(\mathrm{s})}_{\boldsymbol{k},\lambda}}{\partial \xi^{(\mathrm{s})}_{\boldsymbol{k},\lambda}}\right)\left[\frac{\xi^{(\mathrm{s})}_{\boldsymbol{k},\lambda}}{T_\mathrm{s}} + \frac{\partial \mu_\mathrm{s}}{\partial T_\mathrm{s}}\right]. \quad (12)$$

Here, $\xi^{(\mathrm{s})}_{\boldsymbol{k},\lambda} = \varepsilon^{(\mathrm{s})}_{\boldsymbol{k},\lambda} - \mu_s$. The derivative of the surface chemical potential $\mu_\mathrm{s} \equiv \mu_\mathrm{s}(T_\mathrm{s})$ with respect to temperature is obtained by imposing the conservation of the surface electron density [25]. Defining the power dissipated into bulk states as $\mathcal{Q} = \sum_{\boldsymbol{k},\lambda} \mathcal{I}^{(\mathrm{sb})}_{\boldsymbol{k},\lambda} \varepsilon^{(\mathrm{s})}_{\boldsymbol{k},\lambda} \equiv \gamma C_\mathrm{s}(T_\mathrm{s} - T_\mathrm{b})$, Eq. (11) yields the cooling rate [25]

$$\gamma = \frac{\mathcal{Q}}{(T_\mathrm{s} - T_\mathrm{b})C_\mathrm{s}}. \quad (13)$$

In the following we first define the collision integral and then calculate $\mathcal{Q}$, and thus γ.

IV. THE COOLING RATE—SURFACE-TO-BULK POWER DISSIPATION

To calculate the power dissipated into bulk states, we first have to obtain the collision integral due to the interaction of Eq. (3). Within the Fermi-golden-rule approximation, the collision integral on the right-hand side of Eq. (11) reads

$$\mathcal{I}^{(\mathrm{sb})}_{\boldsymbol{k},\lambda} = 2\frac{2\pi}{\hbar \mathcal{A}^2 L_z} \sum_{\boldsymbol{k}',\tilde{\boldsymbol{k}}',\boldsymbol{q}} \sum_{\lambda'} \sum_{\eta,\eta'} \int_{-\infty}^{\infty} d\omega V^2_{\boldsymbol{q},\boldsymbol{k}',\tilde{\boldsymbol{k}}'} F_{\boldsymbol{k},\lambda;\boldsymbol{k}+\boldsymbol{q},\lambda'} \delta\big(\varepsilon^{(\mathrm{s})}_{\boldsymbol{k},\lambda} - \varepsilon^{(\mathrm{s})}_{\boldsymbol{k}+\boldsymbol{q},\lambda'} + \omega\big)\delta\big(\varepsilon^{(\mathrm{b})}_{\boldsymbol{k}',\eta} - \varepsilon^{(\mathrm{b})}_{\tilde{\boldsymbol{k}}',\eta'} - \omega\big)\delta(\boldsymbol{k}'_\parallel - \tilde{\boldsymbol{k}}'_\parallel - \boldsymbol{q})$$
$$\times \big[f^{(\mathrm{s})}_{\boldsymbol{k},\lambda} f^{(\mathrm{b})}_{\boldsymbol{k}',\eta}\big(1 - f^{(\mathrm{s})}_{\boldsymbol{k}+\boldsymbol{q},\lambda'}\big)\big(1 - f^{(\mathrm{b})}_{\tilde{\boldsymbol{k}}',\eta'}\big) - \big(1 - f^{(\mathrm{s})}_{\boldsymbol{k},\lambda}\big)\big(1 - f^{(\mathrm{b})}_{\boldsymbol{k}',\eta}\big) f^{(\mathrm{s})}_{\boldsymbol{k}+\boldsymbol{q},\lambda'} f^{(\mathrm{b})}_{\tilde{\boldsymbol{k}}',\eta'}\big], \quad (14)$$

where the factor 2 upfront accounts for the spin degeneracy of bulk states, ω and $\boldsymbol{q}$ are the transferred energy and momentum parallel to the surface, $\boldsymbol{k}'_\parallel$ and $\tilde{\boldsymbol{k}}'_\parallel$ are the components of the three-dimensional momenta $\boldsymbol{k}'$ and $\tilde{\boldsymbol{k}}'$ parallel to the surface, $\mathcal{A}$ is the surface area of the topological insulator, and L_z its extension in the third dimension. In this equation, $F_{\boldsymbol{k},\lambda;\boldsymbol{k}+\boldsymbol{q},\lambda'}$ is the squared matrix element of the surface-electron density operator between incoming and outgoing scattering states. Since bulk states are standing waves, the z components of the three-dimensional momenta $\boldsymbol{k}'$ and $\tilde{\boldsymbol{k}}'$ are taken to be positive. The matrix element of the screened Coulomb interaction $V_{\boldsymbol{q},\boldsymbol{k}',\tilde{\boldsymbol{k}}'}$ is obtained by integrating it over the incoming and outgoing scattering states, i.e.,

$$V^2_{\boldsymbol{q},\boldsymbol{k}',\tilde{\boldsymbol{k}}'} = 4\left|\int_0^\infty dz \sin(k'_z z)\sin(\tilde{k}'_z z) e^{-\sqrt{q^2+q^2_{\mathrm{TF}}(q,\omega)}z}\right|^2 |\bar{\phi}_{\boldsymbol{q}}|^2. \quad (15)$$

This matrix element is manipulated in Appendix A to give

$$V^2_{\boldsymbol{q},\boldsymbol{k}',\tilde{\boldsymbol{k}}'} \simeq |\bar{\phi}_{\boldsymbol{q}}|^2 \frac{\pi^2}{4\sqrt{q^2+q^2_{\mathrm{TF}}(q,\omega)}} \delta(k'_z - \tilde{k}'_z)$$
$$\equiv V^2_{\boldsymbol{q}} \delta(k'_z - \tilde{k}'_z). \quad (16)$$

To obtain this expression, we have evaluated the integral on the right-hand side of Eq. (15) and approximated it in the limit of small q and $q_{\mathrm{TF}}(q,\omega)$ (the latter is justified when the bulk is undoped and $k_\mathrm{B}T_\mathrm{b} \ll 2\Delta$). In this case, the weight of the integral is located around the lines $\tilde{k}'_z = \pm k'_z$. Thus, we replaced Lorenzians of width $\sqrt{q^2+q^2_{\mathrm{TF}}(q,\omega)}$ with δ functions. Finally, we used that $\tilde{k}'_z$ and k'_z must be taken to be positive.

Putting Eq. (16) back into Eq. (14), after some lengthy but straightforward algebra we get

$$\mathcal{Q} = \frac{4}{\pi\hbar}\int \frac{d^2\boldsymbol{q}}{(2\pi)^2} V^2_{\boldsymbol{q}} \int_0^\infty d\omega\omega[n^{(\mathrm{s})}(\omega) - n^{(\mathrm{b})}(\omega)]$$
$$\times \Im m\chi_\mathrm{b}(\boldsymbol{q},\omega)\Im m\chi_\mathrm{s}(\boldsymbol{q},\omega), \quad (17)$$

where $n^{(\mathrm{s/b})}(\omega) = [e^{\hbar\omega/(k_\mathrm{B}T_{\mathrm{s/b}})} - 1]^{-1}$. In Eq. (17),

$$\Im m\chi_\mathrm{s}(\boldsymbol{q},\omega) = -\frac{\pi}{\mathcal{A}} \sum_{\boldsymbol{k},\lambda,\lambda'} \big(f^{(\mathrm{s})}_{\boldsymbol{k},\lambda} - f^{(\mathrm{s})}_{\boldsymbol{k}+\boldsymbol{q},\lambda'}\big) F_{\boldsymbol{k},\lambda;\boldsymbol{k}+\boldsymbol{q},\lambda'}$$
$$\times \delta\big(\varepsilon^{(\mathrm{s})}_{\boldsymbol{k},\lambda} - \varepsilon^{(\mathrm{s})}_{\boldsymbol{k}+\boldsymbol{q},\lambda'} + \omega\big) \quad (18)$$

and

$$\Im m\chi_\mathrm{b}(\boldsymbol{q},\omega) = -\frac{\pi}{\mathcal{A}L_z} \sum_{\boldsymbol{k}',\eta,\eta'} \big(f^{(\mathrm{b})}_{\boldsymbol{k}',\eta} - f^{(\mathrm{b})}_{\boldsymbol{k}'-\boldsymbol{q},\eta'}\big)$$
$$\times \delta\big(\varepsilon^{(\mathrm{b})}_{\boldsymbol{k}',\eta} - \varepsilon^{(\mathrm{b})}_{\boldsymbol{k}'-\boldsymbol{q},\eta'} - \omega\big) \quad (19)$$

are the imaginary parts of the density-density response functions of surface [33–35] and bulk [32] states, respectively. The function $\Im m\chi_\mathrm{b}(\boldsymbol{q},\omega)$, not to be confused with the function $\bar{\chi}_\mathrm{b}(\boldsymbol{q},\omega)$ introduced in Sec. II A, is calculated in Appendix B.

The cooling rate γ is obtained by inserting $\mathcal{Q}$ in Eq. (17) back into the definition (13), and reads

$$\gamma = -\frac{4}{\pi\hbar C_s}\int \frac{d^2\boldsymbol{q}}{(2\pi)^2} V^2_{\boldsymbol{q}} \int_0^\infty d\omega\omega \frac{n^{(\mathrm{s})}(\omega) - n^{(\mathrm{b})}(\omega)}{T_s - T_\mathrm{b}}$$
$$\times \Im m\chi_\mathrm{b}(\boldsymbol{q},\omega)\Im m\chi_\mathrm{s}(\boldsymbol{q},\omega). \quad (20)$$

This equation is the key result of this paper. We stress that Eq. (20) describes the exchange of heat from surface to bulk electrons and vice versa. This is at the origin of the difference in Bose factors in Eq. (20).

V. RESULTS

We now show our results for the cooling time of electrons in the surface state of a topological insulator and in graphene in proximity to a small-gap material. We find that the bulk behaves as an efficient heat sink for electrons, yielding sub-picosecond cooling times.

We start with the surface states of the topological insulator. We numerically evaluate Eq. (13), with the heat capacity of surface states and power lost to bulk states given by Eqs. (12) and (17), respectively. For the sake of definiteness, in the numerical calculations, we set the surface Fermi velocity [36] $v_F = 0.5\times10^6$ m/s, while the bulk is left undoped, i.e., its electron density is $n_b = 0$ (which translates into $\mu_b = 0$). The bulk temperature is set to $T_b = 300$ K, which is much smaller than the bulk-band gap [37] $2\Delta = 200$ meV. Under these conditions, the bulk bands are nearly unpopulated and we can thus safely take $q_{TF}(q,\omega) = 0$ in Eqs. (10) and (15). Finally, the bulk electron mass is set to [38,39] $m = 0.21m_e$, where $m_e = 9.1\times10^{-31}$ kg is the bare electron mass, while the dielectric constant of the (undoped) topological insulator is taken to be [40] $\epsilon_b = 10$. We would like to point out that the topological insulators that we describe have no charge carriers in the bulk, at low temperature, which corresponds to a Fermi energy inside the gap. Most common growth methods of binary topological insulators, such as Bi_2Se_3 and Sn_2Te_3 result in a Fermi energy either in the valance or conduction band. However, there are well-known methods to produce topological insulators with Fermi energy in the gap, for example by using appropriate substrate materials [41] or using mixtures with controlled stoichiometry such as $Bi_{1.4}Sb_{0.6}Te_{1.51}Se_{1.49}$ (BSTS) [19]. The effect of bulk thermal excitations is included in our theory via the bulk response function. Furthermore, since the typical energies of excitations here are of the order of $k_BT_s \sim 10$–100 meV (in particular, they can be interband transitions), the dielectric constant should be taken to be the one at intermediate frequencies, not the zero-frequency one.

In Fig. 2 we show the cooling rate for surface-to-bulk Coulomb cooling in topological insulators (with Fermi energy in the bulk band gap) calculated from Eq. (13). In Fig. 2(a) we present our numerical results for three different values of the surface electronic density and as a function of temperature T_s. The three chosen densities as $n_s = 0\,\mathrm{cm}^{-2}$, i.e., an undoped system with Fermi energy at the surface Dirac crossing, $n_s = 5\times10^{11}\,\mathrm{cm}^{-2}$ and $n_s = 10^{12}\,\mathrm{cm}^{-2}$. the latter corresponds to a Fermi energy close to the bottom of the bulk conduction band. We find that the cooling rate depends only weakly on surface carrier concentration, while it depends quite strongly on temperature. In particular, it decreases rapidly at low temperatures (i.e., for the surface temperature approaching the bulk one). Curiously, we find that the cooling rate exhibits a maximum at a temperature approximately equal to half the bulk-band gap.

This behavior can be understood by noting that surface electrons can be very hot. As shown in Fig. 2(a), their temperature can be of thousands of Kelvins. At these temperatures their Fermi distribution is very broadened. Thus, they can undergo both intra- and inter-band transitions (where the bands involved here are the surface ones). At the same time bulk electrons are kept at room temperature. Thus, they experience both intraband excitations within the valence and conduction bands and interband ones from the valence to the conduction band. For a surface-electron temperature of the order of the bulk-band gap, excitations of surface states occurring with energies of the order of k_BT_s become resonant with the bulk interband excitations. The weight of such excitations grows sharply for energies larger than 2Δ. The latter fact is seen in the plots of the bulk density-density response (the bulk absorption spectrum) given in Appendix B. However, the cooling rate must fall off rapidly at low temperatures, because interband processes become heavily suppressed. This effect is clearly visible in Fig. 2(a).

We stress that, for different values of the bulk-band gap, the cooling rate is expected to follow similar functional form as shown in Fig. 2(a). The main difference being that the peak

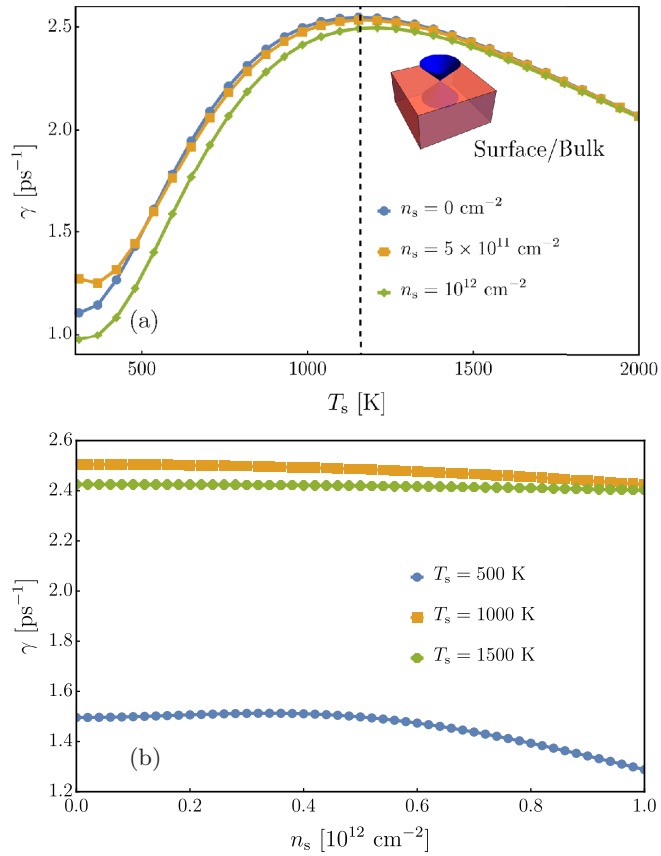

FIG. 2. (a) The cooling rate for Dirac-like surface states ($v_F = 0.5\times10^6$ m/s) in proximity to an undoped 3D bulk ($n_b = 0$, which translates into $\mu_b = 0$) kept at room temperature ($T_b = 300$ K, much smaller than the gap energy: $k_BT_b \ll 2\Delta$), calculated from Eq. (13) and plotted as a function of temperature. Curves exhibit a maximum at a surface electron temperature corresponding to half the bulk-band gap, $\Delta = 100$ meV (dashed line). (b) The cooling rate is very weakly dependent on surface carrier density. The parameter used in these calculations are given at the beginning of Sec. V and are recalled here for convenience. The bulk electron mass is set to [38,39] $m = 0.21m_e$, where $m_e = 9.1\times10^{-31}$ kg is the bare electron mass, while the dielectric constant of the (undoped) topological insulator is taken to be [40] $\epsilon_b = 10$.

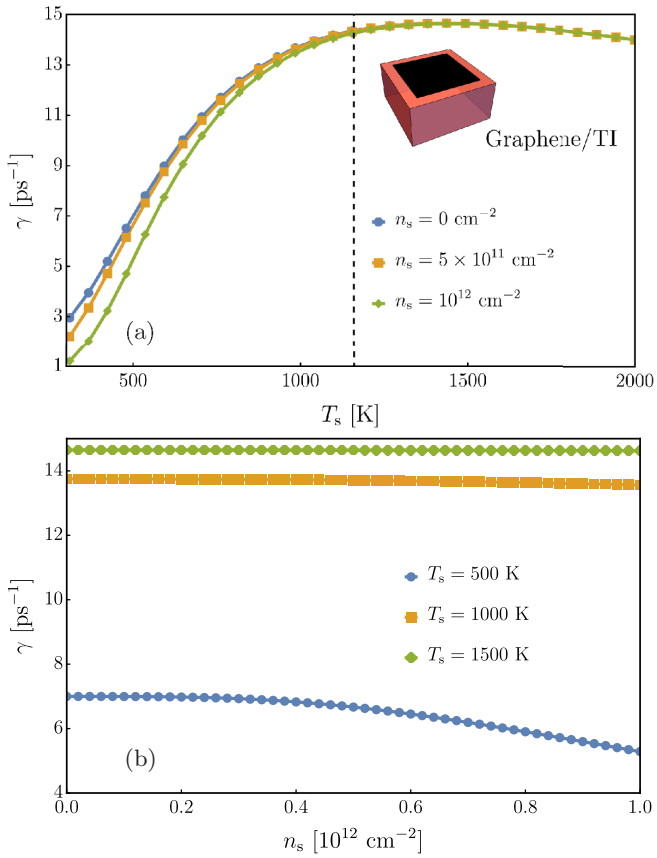

FIG. 3. The cooling rate for Dirac-like states in graphene ($v_F = 10^6$ m/s) in proximity to an undoped 3D bulk identical to that of Fig. 2, calculated from Eq. (13). (a) As a function of temperature, it exhibits a maximum at a surface electron temperature corresponding to half the bulk-band gap Δ (dashed line). (b) The cooling rate is very weakly dependent on surface carrier density. The parameters used in these plots are given in Sec. V and Fig. 2.

occurs at different temperatures of the order of the bulk-band gap. Thus for smaller (larger) gaps, the peak temperature will be lower (higher).

In Fig. 2(b) we show the cooling rate as a function of the surface density n_s and for three values of the temperature, $T_s = 500$, 1000, and 1500 K. We see that the numerical results are practically independent of carrier density at the highest temperatures, and only weakly dependent at the lowest one. This can be understood as a consequence of the complete smearing of the surface occupation function at temperatures much larger than the Fermi energy. Only at the lowest temperatures and highest densities achievable in our model we start observing some deviation from perfect flatness.

In Fig. 3 we show the cooling rate for a graphene sheet in direct contact with a small-gap bulk material as a function of temperature [in Fig. 3(a)] and carrier density [in panel (b)]. For the bulk material, we use the same 3D topological insulator used above, i.e., the 3D parameters are taken to be the same. However, for the "surface" states (now a graphene sheet), the number of fermion flavors is set to [1] $N_f = 4$, while the Fermi velocity is doubled, i.e., we use [1] $v_F = 10^6$ m/s. Comparing Figs. 2 and 3, we see that graphene would exhibit dynamics approximately four-five times faster than the topological-insulator surface states. This is surprising because, thanks to the doubling of the Fermi velocity and quadrupling of the number of fermion flavors, the density of states of undoped graphene and of the surface states studied above are (accidentally) identical. This in turn implies that, in the undoped limit, they also exhibit the same heat capacity.

The reason for the enhanced cooling rate is instead to be found in the typical energies of particle-hole excitations, which are different in the two systems. Due to the linear energy dispersion, the typical energy exchanged by surface and bulk states during a collision is $\omega \simeq \hbar v_F q$. This in turn implies that, for a given value of momentum $\hbar q$, the energy exchanged between graphene and bulk electrons is twice the energy exchanged between surface and bulk states. Thus, the cooling dynamics proceeds at a faster pace in graphene, even if the number of interactions per unit time is the same as in a topological insulator, a fact that is reflected in the larger cooling rate.

If a spacer is introduced between the graphene and bulk material, then the interaction decreases exponentially as e^{-qd}, as shown at the end of Sec. II A. The wave vector q is of the order of the graphene Fermi wave vector. Thus, for the carrier densities and (high) temperatures explored in this paper, the interaction is expected to be suppressed for graphene-bulk distances exceeding 10 nm. We therefore expect the cooling time to grow significantly for larger spacer thicknesses.

VI. DISCUSSION AND CONCLUSIONS

In this paper we have developed the theory of the cooling dynamics of electrons in the surface states of a topological insulator, as well as in a graphene sheet, coupled to bulk states via nonretarded Coulomb interactions. The aim has been to explore the cooling capabilities of all-electronic surface-bulk coupling in these systems. For this reason, we have employed a simplified model: we have treated the surface states (and graphene electrons) as massless Dirac fermions characterized by a Dirac-like energy dispersion. We have neglected corrections due to, e.g., trigonal warping, which are expected to be only minor. We have considered a fully gapped bulk with particle-hole symmetric parabolic-energy bands of equal masses. We have neglected all possible couplings between surface and bulk states (such as impurity or phonon-mediated hopping) with the exception of a density-density interaction of the Coulomb type. The latter conserves the number of electrons in surface and bulk states separately, but allows the exchange of energy between them. We have thus studied how such near-field radiative coupling can efficiently transfer energy between the surface and the bulk. We have thus explored an alternative mechanism to phonon-mediated cooling, which can be very efficient and could become dominant in some circumstances.

In fact, we find that Coulomb cooling can out-compete other cooling mechanisms. We find cooling rates of the order of few inverse picoseconds for the surface states of topological insulators, which agrees with a recent result reported in the literature [19]. More surprisingly, we find that the cooling time of graphene in contact with a narrow-band insulating material

can be as short as a tenth of picosecond. This is significantly faster than graphene cooling via optical, acoustic and substrate phonons, which typically occurs on a picosecond timescale [21–26].

A true comparison between different cooling pathways would require a significantly more detailed work to describe the impact of disorder, phonons, etc., on the cooling dynamics. All these mechanism are also potentially strongly dependent on material characteristics, and therefore it would be hard to derive general and universal trends. On the contrary, our model depends on few, experimentally available parameters (e.g., the surface Fermi velocity, the bulk-band mass, the undoped-material dielectric constant). Thus, it could be employed to study universal trends in electronic cooling of surfaces of topological insulators.

We further observe that, although we have not treated phonon cooling in this paper, the interaction we describe could be thought as the result of phonon emission and re-absorption by the electrons at the surface and in the bulk, respectively. Such phonon-number-conserving processes could be easily incorporated in the theory and would result in a further enhancement of cooling rates. Finally, we stress that the theory also applies, with minor modifications, to graphene in proximity to narrow-gap materials, as we have shown above. This fact, which broadens the applicability of the present theory, has also important practical implications. Graphene is in fact one of the most studied materials for optoelectronic applications. The slowing down of cooling dynamics at high powers limits however its potential, for example, to applications such as higher harmonic generation [19]. The fact that heat dissipation could be made more efficient via near-field coupling to narrow-gap materials offers a novel way to overcome the limitation intrinsic to current graphene devices.

ACKNOWLEDGMENTS

A.P. acknowledges support from the European Commission under the EU Horizon 2020 MSCA-RISE-2019 programme (Project No. 873028 HYDROTRONICS) and of the Leverhulme Trust under the Grant No. RPG-2019-363. K.J.T. acknowledges funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 804349 (ERC StG CUHL).

APPENDIX A: MANIPULATION OF THE SURFACE-BULK INTERACTION

The integration of Eq. (15) yields

$$V^2_{\boldsymbol{q},\boldsymbol{k}',\tilde{\boldsymbol{k}}'} = \left|\bar{\phi}_{\boldsymbol{q}}\right|^2 \left\{ \frac{\sqrt{q^2+q^2_{\mathrm{TF}}(q,\omega)}}{\left[q^2+q^2_{\mathrm{TF}}(q,\omega)\right]+(k'_z-\tilde{k}'_z)^2} - \frac{\sqrt{q^2+q^2_{\mathrm{TF}}(q,\omega)}}{\left[q^2+q^2_{\mathrm{TF}}(q,\omega)\right]+(k'_z+\tilde{k}'_z)^2} \right\}^2 . \tag{A1}$$

We now observe that, for small q and $q^2_{\mathrm{TF}}(q,\omega)$, the regime of interest for the processes we are describing, the two terms on the right-hand side of Eq. (A1) are sharply peaked around $\tilde{k}'_z = k'_z$ and $\tilde{k}'_z = -k'_z$, respectively. Thus, we approximate them with two δ functions, making sure that their total integral (over the variable $\tilde{k}'_z$) remains unchanged. We thus obtain

$$\begin{aligned} V^2_{\boldsymbol{q},\boldsymbol{k}',\tilde{\boldsymbol{k}}'} &\simeq \left[\frac{4\pi e^2}{(\epsilon_b+1)q-4\pi e^2\chi_s(\boldsymbol{q},\omega)}\right]^2 \frac{\pi^2}{4\sqrt{q^2+q^2_{\mathrm{TF}}(q,\omega)}}\left[\delta(k'_z-\tilde{k}'_z)+\delta(k'_z+\tilde{k}'_z)\right] \\ &\simeq \left[\frac{4\pi e^2}{(\epsilon_b+1)q-4\pi e^2\chi_s(\boldsymbol{q},\omega)}\right]^2 \frac{\pi^2}{4\sqrt{q^2+q^2_{\mathrm{TF}}(q,\omega)}}\delta(k'_z-\tilde{k}'_z). \end{aligned} \tag{A2}$$

In the last line we noticed that k'_z and $\tilde{k}'_z$ must be taken as positive, since the bulk wave functions describe standing waves (negative wave vectors correspond to the same wave function).

APPENDIX B: THE DENSITY-DENSITY RESPONSE OF BULK ELECTRONS

The imaginary part of the density-density function in Eq. (19) is given by

$$\Im m\chi_b(\boldsymbol{q},\omega) = -\pi \sum_{\eta,\eta'} \int \frac{d^3\boldsymbol{k}}{(2\pi)^3}[f(\eta\varepsilon_{\boldsymbol{k}}+\eta\Delta-\mu_{\mathrm{b}})-f(\eta\varepsilon_{\boldsymbol{k}}+\eta\Delta+\omega-\mu_{\mathrm{b}})]\delta(\omega+\eta\varepsilon_{\boldsymbol{k}}+\eta\Delta-\eta'\varepsilon_{\boldsymbol{k}+\boldsymbol{q}}-\eta'\Delta). \tag{B1}$$

Here we rewrote $\varepsilon_{\boldsymbol{k},\eta} = \eta\Delta+\eta\varepsilon_{\boldsymbol{k}}$, introducing $\varepsilon_{\boldsymbol{k}} = k^2/(2m)$. We also defined $f(\xi) = [e^{\xi/(k_{\mathrm{B}}T_{\mathrm{b}})}+1]^{-1}$ and introduced the chemical potential of bulk bands μ_{b}. In what follows, we assume $\omega > 0$ and analyze separately the intraband ($\eta' = \eta$) and interband ($\eta' \neq \eta$) contributions to Eq. (B1).

1. Intraband term

In this case, $\eta' = \eta$. Hence,

$$\Im m\chi_b^{(\mathrm{intra})}(\boldsymbol{q},\omega) = -\frac{1}{4\pi}\sum_\eta \int_0^\infty dk k^2 \int_0^\pi d\theta \sin\theta [f(\eta\varepsilon_{\boldsymbol{k}} + \eta\Delta - \mu_\mathrm{b}) - f(\eta\varepsilon_{\boldsymbol{k}} + \eta\Delta + \omega - \mu_\mathrm{b})]\delta(\omega + \eta\varepsilon_{\boldsymbol{k}} - \eta\varepsilon_{\boldsymbol{k}+\boldsymbol{q}}). \tag{B2}$$

The δ function implies that

$$\omega + \eta\frac{k^2}{2m} - \eta\frac{k^2+q^2+2kq\cos(\theta_0)}{2m} = 0 \quad \Rightarrow \quad \cos(\theta_0) = \frac{m}{kq}\left(\frac{q^2}{2m} - \eta\omega\right). \tag{B3}$$

Solutions exist for

$$-\frac{kq}{m} \leqslant \frac{q^2}{2m} - \eta\omega \leqslant \frac{kq}{m} \quad \Rightarrow \quad k \geqslant \frac{m}{q}\left|\frac{q^2}{2m} - \eta\omega\right| \equiv k_0. \tag{B4}$$

Therefore,

$$\begin{aligned}
\Im m\chi_b^{(\mathrm{intra})}(\boldsymbol{q},\omega) &= -\frac{1}{4\pi}\sum_\eta \int_0^\infty dk k^2 [f(\eta\varepsilon_{\boldsymbol{k}} + \eta\Delta - \mu_\mathrm{b}) - f(\eta\varepsilon_{\boldsymbol{k}} + \eta\Delta + \omega - \mu_\mathrm{b})]\frac{m}{kq}\int_0^\pi d\theta\delta(\theta - \theta_0) \\
&= -\frac{m}{4\pi q}\sum_\eta \int_{k_0}^\infty dk k [f(\eta\varepsilon_{\boldsymbol{k}} + \eta\Delta - \mu_\mathrm{b}) - f(\eta\varepsilon_{\boldsymbol{k}} + \eta\Delta + \omega - \mu_\mathrm{b})] \\
&= -\frac{m^2}{4\pi q}\sum_\eta \int_{k_0^2/(2m)}^\infty d\varepsilon [f(\eta\varepsilon + \eta\Delta - \mu_\mathrm{b}) - f(\eta\varepsilon + \eta\Delta + \omega - \mu_\mathrm{b})] \\
&= -\frac{m^2}{4\pi q}\sum_\eta \eta \int_{k_0^2/(2m)}^\infty d\varepsilon [f(\varepsilon + \Delta - \eta\mu_\mathrm{b}) - f(\varepsilon + \Delta + \eta\omega - \eta\mu_\mathrm{b})] \\
&= -\frac{m^2 k_\mathrm{B} T}{4\pi q}\sum_\eta \eta\left\{\ln\left[1 + \exp\left(\frac{\eta\mu_\mathrm{b} - \Delta}{k_\mathrm{B}T_\mathrm{b}} - \frac{k_0^2}{2mk_\mathrm{B}T_\mathrm{b}}\right)\right] - \ln\left[1 + \exp\left(\frac{\eta\mu_\mathrm{b} - \eta\omega - \Delta}{k_\mathrm{B}T_\mathrm{b}} - \frac{k_0^2}{2mk_\mathrm{B}T_\mathrm{b}}\right)\right]\right\}.
\end{aligned} \tag{B5}$$

2. Interband term

In this case, $\eta' = -\eta$. Hence,

$$\begin{aligned}
\Im m\chi_b^{(\mathrm{inter})}(\boldsymbol{q},\omega) = -\frac{1}{4\pi}\sum_\eta \int_0^\infty dk k^2 \int_0^\pi d\theta \sin\theta [f(\eta\varepsilon_{\boldsymbol{k}} + \eta\Delta - \mu_\mathrm{b}) - f(\eta\varepsilon_{\boldsymbol{k}} + \eta\Delta + \omega - \mu_\mathrm{b})] \\
\times \delta(\omega + 2\eta\Delta + \eta\varepsilon_{\boldsymbol{k}} + \eta\varepsilon_{\boldsymbol{k}+\boldsymbol{q}}).
\end{aligned} \tag{B6}$$

It is clear that, for the δ function not to vanish, it must be $\eta = -1$. This in turn implies that

$$\omega - 2\Delta - \frac{k^2}{2m} - \frac{k^2+q^2+2kq\cos(\theta_0)}{2m} = 0 \quad \Rightarrow \quad \cos(\theta_0) = \frac{m}{kq}\left(\omega - 2\Delta - \frac{2k^2+q^2}{2m}\right). \tag{B7}$$

Solutions exist for

$$-\frac{kq}{m} \leqslant \omega - 2\Delta - \frac{2k^2+q^2}{2m} \leqslant \frac{kq}{m} \quad \Rightarrow \quad \begin{cases} k^2 - kq + 2m\Delta - m\omega + \dfrac{q^2}{2} \leqslant 0 \\ k^2 + kq + 2m\Delta - m\omega + \dfrac{q^2}{2} \geqslant 0 \end{cases} \quad \Rightarrow \quad |k_-| < k < k_+,$$

where

$$k_\pm = \frac{q \pm \sqrt{4(m\omega - 2m\Delta) - 2q^2}}{2}, \quad \omega > 2\Delta + \frac{q^2}{2m}. \tag{B8}$$

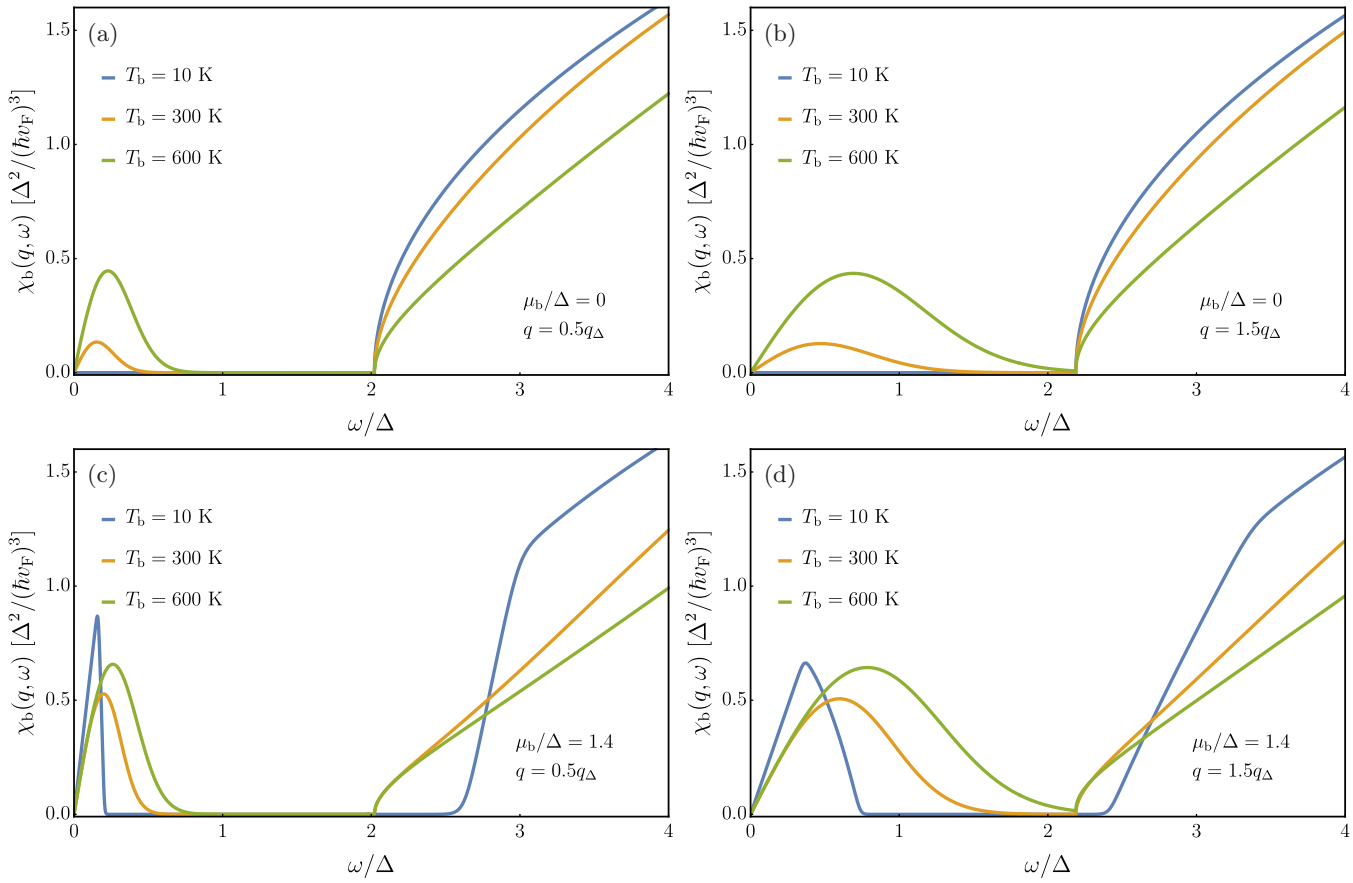

FIG. 4. The density-density response function of bulk electrons in units of $\Delta^2/(\hbar v_F)^3$ ($v_F = 0.5\times10^6$ m/s), for a fixed value of the wave vector q and as a function of energy ω (in units of the half-gap Δ). In each panel we show three curves, one for each temperature ($T =$ 10, 300, and 600 K, respectively). (a) Here $q = 0.5q_\Delta$ and the chemical potential is $\mu_b = 0$ (undoped system). (b) Same as in panel (a), but for $q = 1.5q_\Delta$. (c) Here $q = 0.5q_\Delta$ and the chemical potential is $\mu_b = 1.4\Delta$ (n-doped system). (d) Same as in panel (c), but for $q = 1.5q_\Delta$. The parameters used in these plots are the same as those used in Sec. V, except that $\Delta = 50$ meV. We also defined $q_\Delta = \Delta/(\hbar v_F)$.

Therefore,

$$\begin{aligned}
\Im m\chi_b^{(\text{inter})}(\boldsymbol{q},\omega) &= -\frac{1}{4\pi}\int_0^\infty dk k^2[f(-\varepsilon_{\boldsymbol{k}}-\Delta-\mu_b)-f(-\varepsilon_{\boldsymbol{k}}-\Delta+\omega-\mu_b)]\frac{m}{kq}\int_0^\pi d\theta\delta(\theta-\theta_0)\\
&= -\frac{m}{4\pi q}\int_{|k_-|}^{k_+} dk k[f(-\varepsilon_{\boldsymbol{k}}-\Delta-\mu_b)-f(-\varepsilon_{\boldsymbol{k}}-\Delta+\omega-\mu_b)]\\
&= \frac{m^2}{4\pi q}\int_{k_-^2/(2m)}^{k_+^2/(2m)} d\varepsilon[f(\varepsilon+\Delta+\mu_b)-f(\varepsilon+\Delta-\omega+\mu_b)]\\
&= \frac{m^2 k_B T}{4\pi q}\left\{\ln\left[1+\exp\left(-\frac{\mu_b+\Delta}{k_B T_b}-\frac{k_-^2}{2mk_B T_b}\right)\right]-\ln\left[1+\exp\left(-\frac{\mu_b+\Delta}{k_B T_b}-\frac{k_+^2}{2mk_B T_b}\right)\right]\right.\\
&\quad\left.-\ln\left[1+\exp\left(-\frac{\mu_b-\omega+\Delta}{k_B T_b}-\frac{k_-^2}{2mk_B T_b}\right)\right]+\ln\left[1+\exp\left(-\frac{\mu_b-\omega+\Delta}{k_B T_b}-\frac{k_+^2}{2mk_B T_b}\right)\right]\right\}. \quad \text{(B9)}
\end{aligned}$$

3. Results

Figure 4 shows results for the imaginary part of the bulk density-density response function, $\Im m\chi_b(\boldsymbol{q},\omega)$, where its intra- and interband parts are given in Eqs. (B5) and (B9), respectively. The parameters used in these plots are those given in Sec. V, with the exception of the half-gap which has been set to $\Delta = 50$ meV for convenience reasons. Different panels correspond to different values of the chemical potential and of the wave vector q [in units of $q_\Delta = \Delta/(\hbar v_F)$] used. In each plot the density-density response function of bulk electrons is shows in units of $\Delta^2/(\hbar v_F)^3$, where $v_F = 0.5\times10^6$ m/s, as a function of energy ω. The latter is in units of the half-gap Δ. In each panel we show three curves, one for each temperature, $T = 10$, 300, and 600 K, respectively.

[1] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, The electronic properties of graphene, Rev. Mod. Phys. **81**, 109 (2009).

[2] M. Z. Hasan and C. L. Kane, Colloquium: Topological insulators, Rev. Mod. Phys. **82**, 3045 (2010).

[3] M. Z. Hasan and J. E. Moore, Three-dimensional topological insulators, Annu. Rev. Condens. Matter Phys. **2**, 55 (2011).

[4] T. Wehling, A. Black-Schaffer, and A. Balatsky, Dirac materials, Adv. Phys. **63**, 1 (2014).

[5] L. Fu, C. L. Kane, and E. J. Mele, Topological Insulators in Three Dimensions, Phys. Rev. Lett. **98**, 106803 (2007).

[6] A. A. Burkov and D. G. Hawthorn, Spin and Charge Transport on the Surface of a Topological Insulator, Phys. Rev. Lett. **105**, 066802 (2010).

[7] D. Pesin and A. H. MacDonald, Spintronics and pseudospintronics in graphene and topological insulators, Nat. Mater. **11**, 409 (2012).

[8] S. Lupi and A. Molle, Emerging Dirac materials for THz plasmonics, Appl. Mater. Today **20**, 100732 (2020).

[9] V. N. Kotov, B. Uchoa, V. M. Pereira, F. Guinea, and A. H. Castro Neto, Electron-electron interactions in graphene: Current status and perspectives, Rev. Mod. Phys. **84**, 1067 (2012).

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

[11] M. Massicotte, G. Soavi, A. Principi, and K.-J. Tielrooij, Hot carriers in graphene—fundamentals and applications, Nanoscale **13**, 8376 (2021).

[12] X.-L. Qi and S.-C. Zhang, Topological insulators and superconductors, Rev. Mod. Phys. **83**, 1057 (2011).

[13] C. L. Kane and E. J. Mele, Z_2 Topological Order and the Quantum Spin Hall Effect, Phys. Rev. Lett. **95**, 146802 (2005).

[14] L. Fu and C. L. Kane, Topological insulators with inversion symmetry, Phys. Rev. B **76**, 045302 (2007).

[15] J. C. Y. Teo, L. Fu, and C. L. Kane, Surface states and topological invariants in three-dimensional topological insulators: Application to $Bi_{1-x}Sb_x$, Phys. Rev. B **78**, 045426 (2008).

[16] B. Yan and S.-C. Zhang, Topological materials, Rep. Prog. Phys. **75**, 096501 (2012).

[17] A. Politano, L. Viti, and M. S. Vitiello, Optoelectronic devices, plasmonics, and photonics with topological insulators, APL Mater. **5**, 035504 (2017).

[18] A. Pandey, R. Yadav, M. Kaur, P. Singh, A. Gupta, and S. Husale, High performing flexible optoelectronic devices using thin films of topological insulator, Sci. Rep. **11**, 832 (2021).

[19] S. Kovalev, K. J. Tielrooij, J. C. Deinert, I. Ilyakov, N. Awari, M. Chen, A. Ponomaryov, M. Bawatna, T. V. A. G. de Oliveira, L. M. Eng, K. A. Kuznetsov, D. A. Safronenkov, G. K. Kitaeva, P. I. Kuznetsov, H. A. Hafez, D. Turchinovich, and M. Gensch, Terahertz signatures of ultrafast Dirac fermion relaxation at the surface of topological insulators, npj Quantum Mater. **6**, 84 (2021).

[20] J.-C. Deinert, D. Alcaraz Iranzo, R. Pérez, X. Jia, H. A. Hafez, I. Ilyakov, N. Awari, M. Chen, M. Bawatna, A. N. Ponomaryov, S. Germanskiy, M. Bonn, F. H. L. Koppens, D. Turchinovich, M. Gensch, S. Kovalev, and K.-J. Tielrooij, Grating-graphene metamaterial as a platform for terahertz nonlinear photonics, ACS Nano **15**, 1145 (2021).

[21] R. Bistritzer and A. H. MacDonald, Electronic Cooling in Graphene, Phys. Rev. Lett. **102**, 206410 (2009).

[22] J. C. W. Song, M. Y. Reizer, and L. S. Levitov, Disorder-Assisted Electron-Phonon Scattering and Cooling Pathways in Graphene, Phys. Rev. Lett. **109**, 106602 (2012).

[23] E. A. A. Pogna, X. Jia, A. Principi, A. Block, L. Banszerus, J. Zhang, X. Liu, T. Sohier, S. Forti, K. Soundarapandian, B. Terrés, J. D. Mehew, C. Trovatello, C. Coletti, F. H. L. Koppens, M. Bonn, H. I. Wang, N. van Hulst, M. J. Verstraete, H. Peng *et al.*, Hot-carrier cooling in high-quality graphene is intrinsically limited by optical phonons, ACS Nano **15**, 11285 (2021).

[24] T. Low, V. Perebeinos, R. Kim, M. Freitag, and P. Avouris, Cooling of photoexcited carriers in graphene by internal and substrate phonons, Phys. Rev. B **86**, 045413 (2012).

[25] A. Principi, M. B. Lundeberg, N. C. H. Hesp, K.-J. Tielrooij, F. H. L. Koppens, and M. Polini, Super-Planckian Electron Cooling in a Van Der Waals Stack, Phys. Rev. Lett. **118**, 126804 (2017).

[26] K.-J. Tielrooij, N. C. H. Hesp, A. Principi, M. B. Lundeberg, E. A. A. Pogna, L. Banszerus, Z. Mics, M. Massicotte, P. Schmidt, D. Davydovskaya, D. G. Purdie, I. Goykhman, G. Soavi, A. Lombardo, K. Watanabe, T. Taniguchi, M. Bonn, D. Turchinovich, C. Stampfer, A. C. Ferrari *et al.*, Out-of-plane heat transfer in van der Waals stacks through electron–hyperbolic phonon coupling, Nat. Nanotechnol. **13**, 41 (2018).

[27] Y. H. Wang, D. Hsieh, E. J. Sie, H. Steinberg, D. R. Gardner, Y. S. Lee, P. Jarillo-Herrero, and N. Gedik, Measurement of Intrinsic Dirac Fermion Cooling on the Surface of the Topological Insulator Bi_2Se_3 Using Time-Resolved and Angle-Resolved Photoemission Spectroscopy, Phys. Rev. Lett. **109**, 127401 (2012).

[28] M. Massicotte, P. Schmidt, F. Vialla, K. Watanabe, T. Taniguchi, K. J. Tielrooij, and F. H. L. Koppens, Photo-thermionic effect in vertical graphene heterostructures, Nat. Commun. **7**, 12174 (2016).

[29] S. Fu, I. du Fossé, X. Jia, J. Xu, X. Yu, H. Zhang, W. Zheng, S. Krasel, Z. Chen, Z. M. Wang, K.-J. Tielrooij, M. Bonn, A. J. Houtepen, and H. I. Wang, Long-lived charge separation following pump-wavelength-dependent ultrafast charge transfer in graphene/WS_2 heterostructures, Sci. Adv. **7**, eabd9061 (2021).

[30] M. T. Mihnev, J. R. Tolsma, C. J. Divin, D. Sun, R. Asgari, M. Polini, C. Berger, W. A. de Heer, A. H. MacDonald, and T. B. Norris, Electronic cooling via interlayer coulomb coupling in multilayer epitaxial graphene, Nat. Commun. **6**, 8105 (2015).

[31] A. Principi, E. van Loon, M. Polini, and M. I. Katsnelson, Confining graphene plasmons to the ultimate limit, Phys. Rev. B **98**, 035427 (2018).

[32] G. F. Giuliani and G. Vignale, *Quantum Theory of the Electron Liquid* (Cambridge University Press, Cambridge, UK, 2005).

[33] B. Wunsch, T. Stauber, F. Sols, and F. Guinea, Dynamical polarization of graphene at finite doping, New J. Phys. **8**, 318 (2006).

[34] Y. Barlas, T. Pereg-Barnea, M. Polini, R. Asgari, and A. H. MacDonald, Chirality and Correlations in Graphene, Phys. Rev. Lett. **98**, 236601 (2007).

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

[36] C. S. Tang, B. Xia, X. Zou, S. Chen, H.-W. Ou, L. Wang, A. Rusydi, J.-X. Zhu, and E. E. M. Chia, Terahertz conductivity

of topological surface states in $Bi_{1.5}Sb_{0.5}Te_{1.8}Se_{1.2}$, Sci. Rep. **3**, 3513 (2013).

[37] T. Arakane, T. Sato, S. Souma, K. Kosaka, K. Nakayama, M. Komatsu, T. Takahashi, Z. Ren, K. Segawa, and Y. Ando, Tunable Dirac cone in the topological insulator $Bi_{2-x}Sb_xTe_{3-y}Se_y$, Nat. Commun. **3**, 636 (2012).

[38] M. Orlita, B. A. Piot, G. Martinez, N. K. S. Kumar, C. Faugeras, M. Potemski, C. Michel, E. M. Hankiewicz, T. Brauner, i. c. v. Drašar, S. Schreyeck, S. Grauer, K. Brunner, C. Gould, C. Brüne, and L. W. Molenkamp, Magneto-Optics of Massive Dirac Fermions in Bulk Bi_2Se_3, Phys. Rev. Lett. **114**, 186401 (2015).

[39] M. Lang, L. He, F. Xiu, X. Yu, J. Tang, Y. Wang, X. Kou, W. Jiang, A. V. Fedorov, and K. L. Wang, Revelation of topological surface states in Bi_2Se_3 thin films by *in situ* Al passivation, ACS Nano **6**, 295 (2012).

[40] K. Kuznetsov, P. Kuznetsov, A. Frolov, S. Kovalev, I. Ilyakov, A. Ezhov, and G. Kitaeva, Bulk and surface terahertz conductivity of $Bi_{2-x}Sb_xTe_{3-y}Se_y$ topological insulators, Optic. Eng. **60**, 1 (2021).

[41] F. Bonell, M. G. Cuxart, K. Song, R. Robles, P. Ordejón, S. Roche, A. Mugarza, and S. O. Valenzuela, Growth of twin-free and low-doped topological insulators on $BaF_2(111)$, Crystal Growth Design **17**, 4655 (2017).