Multiple negative differential conductance regions and inelastic phonon assisted tunneling in graphene-hBN-graphene structures

B. Amorim,[1, 2, *] R. M. Ribeiro,[1] and N. M. R. Peres[1]

[1]*International Iberian Nanotechnology Laboratory (INL), 4715-330 Braga, Portugal*
[2]*Center of Physics and Department of Physics, Universidade do Minho, 4710-057 Braga, Portugal*

In this paper we study in detail the effect of the rotational alignment between a hexagonal boron nitride (hBN) slab and the graphene layers in the vertical current of a a graphene-hBN-graphene device. We show how for small rotational angles, the transference of momentum by the hBN crystal lattice leads to multiple peaks in the I-V curve of the device, giving origin to multiple regions displaying negative differential conductance. We also study the effect of scattering by phonons in the vertical current an see how the opening up of inelastic tunneling events allowed by spontaneous emission of optical phonons leads to sharp peaks in the second derivative of the current.

CONTENTS

I. INTRODUCTION

Being able to tailor the properties of materials at will, aiming at unveiling new physics and achieving never though before properties, is the main goal of condensed matter physics and materials science. However, the degree of manipulation we can undertake using conventional materials is somewhat limited. In the last ten years, the advent of two-dimensional materials[1,2] opened new avenues waiting for being explored. One of the less explored avenue is the one opened by van der Waals (vdW) hybrid structures[3], new systems formed by stacking layers of two-dimensional crystals on top of each other, have emerged as a new approach for manipulating and tailoring material properties at will[4,5]. Among the various possible combinations of two dimensional crystals, graphene - semiconductor/insulator - graphene vdW structures, with semiconducting transition metal dichalcogenide (STMDC) or hexagonal boron nitride (hBN) as the semiconductor/insulator, have emerged as some of the most promising from the point of view of applications. The possibility of controlling electrostatically the effective barrier height presented by the insulator/semiconductor to the vertical flow of electrons between the two graphene layers with a gate voltage has enabled the operation these devices as transistors[6-8], with ON/OFF ratios as high as 10^6 being possible in graphene-WS$_2$-graphene devices[8]. It was also shown that graphene-STMDC-graphene devices can operate as photodectectors with high quantum efficiencies and fast response times[9-11]. Due to the extreme high quality and atomically sharp interfaces[12] between different layers in vdW structures , lattice mismatch and relative alignment between consecutive layers play a fundamental role in determining the electronic coupling between different layers of the vdW structure, ultimately determining its electronic and optical properties. Lattice misalignment between different layers has been known to lead to the formation of Moir patterns in rotated graphite layers[13]. The effect of lattice misalignment and mismatch has been extensively studied in the context of twisted graphene bilayers and graphene-on-hBN structures. It was shown theoretically and experimentally, that misalignment in a
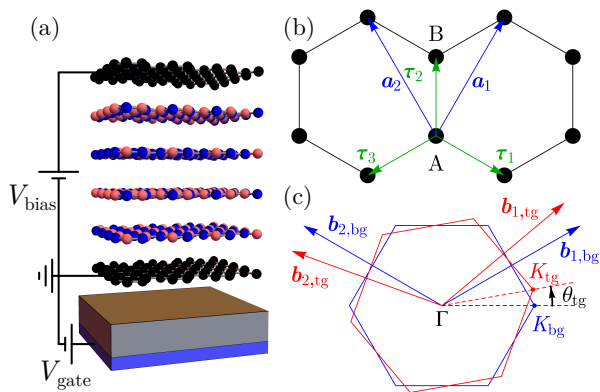
Figure 1. (a) Schematic of a typical graphene-hBN-graphene vdW structure with four boron nitride layers, with applied gate, V_{gate}, and bias, V_{bias}, voltages. (b) Representation of crystalline structure shared by a graphene/boron nitride monolayer, showing the lattice basis, $\{\boldsymbol{a}_1, \boldsymbol{a}_2\}$, the nearest neighbour vectors $\boldsymbol{\tau}_i$, $i = 1, 2, 3$, and the sublattice A/B sites. (c) Representation of the 1st Brillouin zone of the rotated bottom and top graphene layers, showing the K points of both layers and the reciprocal lattice basis vectors $\{\boldsymbol{b}_{1,\text{bg,tg}}, \boldsymbol{b}_{2,\text{bg/tg}}\}$.

graphene bilayer leads to a renormalization of graphene's Fermi velocity[14,15]. It was also found out that mismatch and misalignment controls the formation of mini Dirac cones in the band structure of graphene - hBN structures systems[16–21].. The dependence of the vertical current in vdW structures on the rotation between different layers was first studied in Ref. 22 in the context of twisted bilayer graphene, where it was found that the current is extremely sensitive to the twist angle. Although this dependence was not at first completely appreciated, it was soon understood and verified[23,24] that the misalignment between the graphene layers in graphene-hBN-graphene structures can lead to the occurrence of negative differential conductance (NDC) regions, with the I-V curve displaying peaks whose dependence on the bias voltage depends on the rotation angle between the graphene layers. More recently, the effect of misalignment on the vertical current in devices formed by two graphene bilayers[25–27] and by one graphene monolayer and a graphene bilayer separated by hBN has also been studied.[28] Scattering by phonons can lead to incoherent phonon assisted tunneling between the graphene layers. This effect has been first theoretically studied for vdW structures for twisted graphene bilayers[29]. More recently, effects of phonon assisted scattering on vertical transport have been experimentally detected in graphene-hBN-graphite[30] and graphene-hBN-graphene structures[31] and have been proposed as a possible way to probe the phonon spectrum of vdW structures.

In this paper we describe the vertical current in graphene-hBN-graphene devices with misaligned layers, and for small twist angles, properly taking into account momentum conservation rules, within the non-

equilibrium Green's function framework and using a tight-binding based continuous Hamiltonian. We show that the present theory reduces to the ones used in Refs. 23 and 24. By taking into account processes involving transference of momentum by the hBN crystal lattice to the tunneling electrons, we find that the vertical current depends sensitively on the relative alignment between the graphene layers and the hBN slab and that by carefully controlling this alignment, it is possible to obtain several peaks in the I-V curve of the device, followed by regions of NDC, a possibility that has not been considered previously. We also find out that the structure of graphene wavefunctions manifests itself in the vertical current, suppressing some of the current peaks that would be expected with considerations based only on electronic dispersion relations. We study the effect of resonant disorder in the graphene layers in the vertical current, treated within the self-consistent Born approximation (SCBA) which correctly describes the proportionality of the transport lifetime with the energy[32]. We finally study how phonons and disorder give origin to non-coherent current between the two graphene layers, deriving an expression for it.

The paper is organized as follows. In Sec. II we describe the theoretical framework we employed in this work: in subsection II A, we present the Hamiltonian used to model the graphene-hBN-graphene device and in subsection II B we present the fundamental equations used to treat transport within the non-equilibrium Green's functions formalism. In Sec. III we discuss the coherent tunneling flowing through a pristine device taking into account the lattice mismatch and misalignment between graphene layers and the hBN slab. The consequences of treating graphene as part of the device or as external contacts are discussed and the role of the momentum transferred to the tunneling electrons by the hBN lattice is analyzed in detail. The effect of an in-plane magnetic field in the current is also discussed. In Sec. IV, the effects of disorder and phonon scattering into the vertical current are studied and a expression for the phonon/disorder assisted current to lowest order in perturbation theory is derived. Finally, in Sec. V we conclude. Technical details and longer derivations are include as Appendices.

II. FORMALISM

We want to study the vertical current flowing through a device formed by two graphene layers (bottom, bg, and top, tg) separated by a few layers, $\mathcal{N}$, of hBN. The distance between the two graphene layers is given by d. We assume that the top graphene layer and the hBN slab are rotated with respect to the bottom graphene layer by an angle of θ_{tg} and θ_{hBN}, respectively. We assume that layers forming the hBN slab are perfectly aligned with an AA$'$ stacking[33,34] (consecutive honeycomb lattices are perfectly aligned, with each boron/nitrogen atom of one

layer directly on top of the nitrogen/atom of the next layer). A bias voltage, V_{bias}, can be applied between the top and bottom graphene layers, which will induce a current between the two. The doping of the graphene layers can be controlled by application of a gate voltage to the bottom graphene. A schematic of the typical device structure is shown in Fig. 1.

A. Model Hamiltonian

We model the graphene - hBN - graphene system with the following Hamiltonian

$$
\begin{aligned}
H =& H_{\text{bg}} + H_{\text{tg}} + H_{\text{hBN}} + \\
& + \left(T_{\text{hBN,bg}} + T_{\text{hBN,tg}} + \text{h.c.} \right),
\end{aligned} \tag{1}
$$

where $H_{\text{bg/tg}}$ is the Hamiltonian describing the isolated bottom/top graphene layer and $T_{\text{hBN,bg/tg}} = T_{\text{bg/tg,hBN}}^{\dagger}$ describes the hopping of electrons from the bottom/top graphene layer to the hBN slab. The current between the two graphene layers will be dominated by low energy states. Focusing on the states close to the $\boldsymbol{K}_{\text{bg}}$ and $\boldsymbol{K}'_{\text{bg}} = -\boldsymbol{K}_{\text{bg}}$ points of the bottom graphene layer, we write the Hamiltonian of the bottom graphene layer in sublattice basis and in term of Bloch states as the massless Dirac Hamiltonian

$$
\begin{aligned}
H_{\text{bg}} = \sum_{\boldsymbol{k},\tau} \boldsymbol{c}_{\boldsymbol{k}_\tau,\text{bg}}^{\dagger} \cdot \\
\begin{bmatrix} V_{\text{bg}} & \tau v_F \hbar \left| \boldsymbol{k} \right| e^{-\tau i \theta_{\boldsymbol{k},\text{bg}}} \\ \tau v_F \hbar \left| \boldsymbol{k} \right| e^{-\tau i \theta_{\boldsymbol{k},\text{bg}}.} & V_{\text{bg}} \end{bmatrix} \cdot \boldsymbol{c}_{\boldsymbol{k}_\tau,\text{bg}},
\end{aligned} \tag{2}
$$

where v_F is graphene's Fermi velocity, V_{bg} is a on-site potential induced by the applied bias and gate voltages, $\boldsymbol{c}_{\boldsymbol{k}_\tau,\text{bg}}^{\dagger} = \begin{bmatrix} c_{\boldsymbol{k}_\tau,A,\text{bg}}^{\dagger} & c_{\boldsymbol{k}_\tau,B,\text{bg}}^{\dagger} \end{bmatrix}$ is the electron creation operator for states localized in the A/B sublattice, in the $\tau \boldsymbol{K}_{\text{bg}}$ valley ($\tau = \pm 1$), with momentum $\tau \boldsymbol{K}_{\text{bg}} + \boldsymbol{k}$ (measured from the Brillouin zone center) and $\theta_{\boldsymbol{k},\text{bg}}$ is the angle formed between $\boldsymbol{k}$ and $\boldsymbol{K}_{\text{bg}}$. We choose the zero of energy to lie at the Fermi level of the bottom graphene layer, in which case we have $V_{\text{bg}} = -\epsilon_{\text{F,bg}}$, where $\epsilon_{\text{F,bg}}$ is the Fermi energy of the bottom graphene layer measured from its Dirac point. The Hamiltonian in Eq. (2) is diagonalized by the eigenstates $|\boldsymbol{k},\tau,\lambda\rangle_{\text{bg}} = \left[1, \lambda \tau e^{-i\tau\theta_{\boldsymbol{k},\text{bg}}} \right]^{\dagger}/\sqrt{2}$ with corresponding dispersion relation $\epsilon_{\boldsymbol{k},\lambda} = \lambda v_F \hbar \left| \boldsymbol{k} \right|$, with $\lambda = \pm 1$ for electrons in the conduction/valence band. Since we will be interested in studying the vertical current to lowest order in the graphene-hBN coupling, we neglect the effect of the periodic potential generated by the hBN slab in dispersion relation of graphene electronic states[16–20,35,36]. Using the same reference frame in momentum space as in Eq. (2), the Hamiltonian describing the top graphene layer in the

Dirac cone approximation reads

$$
\begin{aligned}
H_{\text{tg}} = \sum_{\boldsymbol{k},\tau} \boldsymbol{c}_{\boldsymbol{k}_\tau,\text{tg}}^{\dagger} \cdot \\
\begin{bmatrix} V_{\text{tg}} & \tau v_F \hbar \left| \boldsymbol{k}' \right| e^{-i\tau\theta_{\boldsymbol{k}',\text{tg}}} \\ \tau v_F \hbar \left| \boldsymbol{k}' \right| e^{i\tau\theta_{\boldsymbol{k}',\text{tg}}} & V_{\text{tg}} \end{bmatrix} \cdot \boldsymbol{c}_{\boldsymbol{k}_\tau,\text{tg}},
\end{aligned} \tag{3}
$$

where $\boldsymbol{k}' = \boldsymbol{k} + \tau \Delta \boldsymbol{K}_{\text{b,t}}$, is measured from the $\tau \boldsymbol{K}_{\text{tg}}$ point of the top graphene layer, with $\boldsymbol{K}_{\text{tg}} = \mathcal{R}(\theta_{\text{tg}}) \cdot \boldsymbol{K}_{\text{bg}}$ ($\mathcal{R}(\theta)$ a rotation matrix), $\Delta \boldsymbol{K}_{\text{b,t}} = \boldsymbol{K}_{\text{bg}} - \boldsymbol{K}_{\text{tg}}$ is the displacement between the Dirac points of the two rotated graphene layers and $\boldsymbol{k}$ is measured with respect to the $\tau \boldsymbol{K}_{\text{bg}}$ Dirac point of the bottom graphene graphene layer. $\theta_{\boldsymbol{k}',\text{tg}}$ is the angle between $\boldsymbol{k}'$ and $\boldsymbol{K}_{\text{tg}}$ and V_{tg} is an on-site potential, due to the applied bias and gate voltages, and is given by $V_{\text{tg}} = -\epsilon_{\text{F,tg}} - eV_{\text{bias}}$, with $\epsilon_{\text{F,tg}}$ the Fermi level of the top graphene layer measured from its Dirac point and $e > 0$ the fundamental electronic charge. The remaining symbols in Eq. (3) are similarly defined to the ones in Eq. (2). Due to the large band gap of boron nitride, we ignore its momentum dependence, writing the hBN slab Hamiltonian as

$$
\begin{aligned}
H_{\text{hBN}} = & \\
& \sum_{\ell=1}^{\mathcal{N}} \sum_{\boldsymbol{k},\tau} \boldsymbol{c}_{\boldsymbol{k}_\tau,\ell,\text{hBN}}^{\dagger} \cdot \begin{bmatrix} E_{\text{B}} + V_\ell & 0 \\ 0 & E_{\text{N}} + V_\ell \end{bmatrix} \cdot \boldsymbol{c}_{\boldsymbol{k}_\tau,\ell,\text{hBN}} \\
& + \sum_{\ell=1}^{\mathcal{N}-1} \sum_{\boldsymbol{k},\tau} \boldsymbol{c}_{\boldsymbol{k}_\tau,\ell+1,\text{hBN}}^{\dagger} \cdot \begin{bmatrix} 0 & -t_\perp \\ -t_\perp & 0 \end{bmatrix} \cdot \boldsymbol{c}_{\boldsymbol{k}_\tau,\ell,\text{hBN}} + \text{h.c},
\end{aligned} \tag{4}
$$

where $\boldsymbol{c}_{\boldsymbol{k}_\tau,\ell,\text{hBN}}^{\dagger} = \begin{bmatrix} c_{\boldsymbol{k}_\tau,\text{B}\ell,\text{hBN}}^{\dagger} & c_{\boldsymbol{k}_\tau,\text{N}\ell,\text{hBN}}^{\dagger} \end{bmatrix}$ creates an electron in layer $\ell = 1,...,\mathcal{N}$ of the hBN slab, in the boron (B)/nitrogen (N) site, τ specifies the valley, E_{B} and E_{N} are, respectively, the on-site energies of boron and nitride sites measured from the Dirac point of graphene, $t_\perp$ is the nearest neighbour interlayer hoping and V_ℓ is a potential induced by the applied voltages. Due to the large energy offset between graphene and hBN sites, the charge accumulated in the hBN layers will be negligible. In this case a simple electrostatic calculation (see Appendix A) gives us $V_\ell = -\epsilon_{\text{F,bg}} - (\epsilon_{\text{F,tg}} + eV_{\text{bias}}) \ell/(\mathcal{N}+1)$. For two rotated crystal layers, Bloch states from different layers can only be coupled provided momentum is conserved modulo any combination of reciprocal lattice vectors of both layers[22,37], in a so called generalized *Umklapp* process. Focusing on low energy states and considering only the three most relevant processes, the coupling between the graphene layers and the hBN slab is described by (see Appendix B)

$$
T_{\text{hBN,X}} = \sum_{\boldsymbol{k},\tau} \sum_{n=0}^{2} \boldsymbol{c}_{\boldsymbol{k}_\tau + \tau \boldsymbol{g}_n^{\text{X,hBN}}, \ell_{\text{X}}, \text{hBN}}^{\dagger} \cdot \boldsymbol{R}_{\frac{2\pi}{3}}^n \cdot \hat{\boldsymbol{T}} \cdot \boldsymbol{R}_{-\frac{2\pi}{3}}^n \cdot \boldsymbol{c}_{\boldsymbol{k}_\tau,\text{X}}, \tag{5}
$$

where $\boldsymbol{c}_{\boldsymbol{k}_\tau,\text{X}}$ is an annihilation operator of an electron in the X = bg/tg graphene layer with momentum $\boldsymbol{k}_\tau$ mea-

sured from the $\tau \boldsymbol{K}_X$ point, $\boldsymbol{c}^{\dagger}_{\boldsymbol{k}_\tau + \tau \boldsymbol{g}_n^X, \ell_X, \mathrm{hBN}}$ is a creation operator of an electron state in the $\ell_X = 1/\mathcal{N}$ layer of the hBN slab, with momentum $\boldsymbol{k}_\tau + \tau \boldsymbol{g}_n^X$ measured from $\tau \boldsymbol{K}_X$, and with the matrices $\hat{\boldsymbol{T}}$ and $\boldsymbol{R}_\theta$ defined as

$$\boldsymbol{R}_\theta = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{bmatrix}, \tag{6}$$

$$\hat{\boldsymbol{T}} = \begin{bmatrix} t_{\mathrm{B,C}} & t_{\mathrm{B,C}} \\ t_{\mathrm{N,C}} & t_{\mathrm{N,C}} \end{bmatrix}, \tag{7}$$

where $t_{\mathrm{B,C}}$ ($t_{\mathrm{N,C}}$) is the hoping parameter between a carbon site and boron (nitrogen) site and the vectors $\boldsymbol{g}_n^{\mathrm{X,hBN}}$ are given by (see Appendix B)

$$\boldsymbol{g}_0^{\mathrm{X,hBN}} = 0, \tag{8}$$
$$\boldsymbol{g}_1^{\mathrm{X,hBN}} = \boldsymbol{b}_{2,\mathrm{X}} - \boldsymbol{b}_{2,\mathrm{hBN}}, \tag{9}$$
$$\boldsymbol{g}_2^{\mathrm{X,hBN}} = -\boldsymbol{b}_{1,\mathrm{X}} + \boldsymbol{b}_{1,\mathrm{hBN}}, \tag{10}$$

where $\boldsymbol{b}_{i,\mathrm{X}}$ and $\boldsymbol{b}_{i,\mathrm{hBN}}$ ($i = 1, 2$) are, respectively, the reciprocal lattice vectors of the bottom/top graphene layer and of the hBN slab (see Fig. 1). Notice that if the boron nitride slab is formed by an even number of layers, one must replace $\boldsymbol{R}_{\frac{2\pi}{3}}^n \rightarrow \sigma_x \cdot \boldsymbol{R}_{\frac{2\pi}{3}}^n \cdot \sigma_x$ for $T_{\mathrm{hBN,tg}}$, since boron and nitrogen atoms switch positions in consecutive layers of hBN. Different reciprocal lattice vectors are related to each other by $\boldsymbol{b}_{i,\mathrm{tg}} = \mathcal{R}(\theta_{\mathrm{tg}}) \cdot \boldsymbol{b}_{i,\mathrm{bg}}$ and $\boldsymbol{b}_{i,\mathrm{hBN}} = (a_{\mathrm{g}}/a_{\mathrm{hBN}}) \mathcal{R}(\theta_{\mathrm{hBN}}) \cdot \boldsymbol{b}_{i,\mathrm{bg}}$, where $a_{\mathrm{g}}/a_{\mathrm{hBN}}$ is a scale factor, with a_{g} (a_{hBN}) the lattice parameter of graphene (hBN). Hamiltonians of the form of Eq. (5) have previously been used to study twisted graphene bilayers[14,38–40] and graphene-on-hBN structures[35,36,41]. Considering the three processes coupling the bottom graphene with hBN and the three processes connecting hBN to the top graphene layer, described by Eq. (5), there are nine hBN mediated processes coupling the bottom graphene layer to the top one[24]. These nine processes couple a state from the bottom graphene layer with momentum $\boldsymbol{k}$ (measured from $\tau \boldsymbol{K}_{\mathrm{bg}}$) to states of the top graphene layer with momentum $\boldsymbol{k} + \tau \boldsymbol{\mathcal{Q}}_{n,m}$ (measured from $\tau \boldsymbol{K}_{\mathrm{tg}}$) with (see Appendix B)

$$\boldsymbol{\mathcal{Q}}_{n,m} = \Delta \boldsymbol{K}_{\mathrm{b,t}} + \boldsymbol{g}_n^{\mathrm{bg,hBN}} - \boldsymbol{g}_m^{\mathrm{tg,hBN}}, \; n, m = 0, 1, 2. \tag{11}$$

The processes with $n \neq m$ involve transfer of momentum by the hBN lattice, while processes with $n = m$ do not. At zero magnetic field, the overall three-fold rotational invariance of the graphene-hBN-graphene structure implies that these nine processes can be organized in three groups of three, with processes in the same group being related by $2\pi/3$ rotation and therefore giving the same contribution to the vertical current. The three groups are

$$\begin{aligned} &\{\boldsymbol{\mathcal{Q}}_{0,0}, \, \boldsymbol{\mathcal{Q}}_{1,1}, \, \boldsymbol{\mathcal{Q}}_{2,2}\}, \\ &\{\boldsymbol{\mathcal{Q}}_{0,1}, \, \boldsymbol{\mathcal{Q}}_{1,2}, \, \boldsymbol{\mathcal{Q}}_{2,0}\}, \\ &\{\boldsymbol{\mathcal{Q}}_{0,2}, \, \boldsymbol{\mathcal{Q}}_{1,0}, \, \boldsymbol{\mathcal{Q}}_{2,1}\}, \end{aligned} \tag{12}$$

with length of the vectors in each group being the same. For small rotation angles and lattice mismatch, $\delta = a_{\mathrm{hBN}}/a_{\mathrm{g}} - 1$, we have[42]

$$\frac{|\boldsymbol{\mathcal{Q}}_{0,0}|^2}{K_{\mathrm{g}}^2} \simeq \theta_{\mathrm{tg}}^2, \tag{13}$$

$$\frac{|\boldsymbol{\mathcal{Q}}_{0,1}|^2}{K_{\mathrm{g}}^2} \simeq \theta_{\mathrm{tg}}^2 + 3\left(\theta_{\mathrm{hBN}}^2 + \delta^2 - \theta_{\mathrm{tg}}\theta_{\mathrm{hBN}}\right) + \sqrt{3}\delta\theta_{\mathrm{tg}} \tag{14}$$

$$\frac{|\boldsymbol{\mathcal{Q}}_{0,2}|^2}{K_{\mathrm{g}}^2} \simeq \theta_{\mathrm{tg}}^2 + 3\left(\theta_{\mathrm{hBN}}^2 + \delta^2 - \theta_{\mathrm{tg}}\theta_{\mathrm{hBN}}\right) - \sqrt{3}\delta\theta_{\mathrm{tg}} \tag{15}$$

with $K_{\mathrm{g}} = 4\pi/(3a_{\mathrm{g}})$ the length of $\boldsymbol{K}_{\mathrm{bg/tg}}$.

B. Current evaluation

The standard approach to transport in a mesoscopic device assumes that the device is attached to external contacts that are in a thermal equilibrium state with well defined chemical potentials. This is only an approximation as once a current starts flowing through the system, the contacts will also be in a non-equilibrium state[43,44]. The problem of computing the current flowing through a mesoscopic device is then analogous to the problem of computing the water flux through a thin pipe that is connecting two large vessels with different water levels[45]. Once water starts flowing through the pipe, the water levels in each vessel are no longer constants, however, on short time scales, assuming that the water levels in the vessels are constant is a reasonable approximation, provided that these are wide enough with respect to the pipe. In the same way, within short time scales compared to the depletion time of an external battery, it is a reasonable approximation to assume that the external contacts have well defined, constant chemical potentials. Using the non-equilibrium Green's function technique one can then show that in a mesoscopic device that is attached to two non-interacting[46] contacts, bottom and top, in thermal equilibrium state described, respectively, by the Fermi-Dirac distribution functions $f_{\mathrm{b}}(\omega) = \left[e^{\beta(\omega - \mu_{\mathrm{b}})} + 1\right]^{-1}$ and $f_{\mathrm{t}}(\omega) = \left[e^{\beta(\omega - \mu_{\mathrm{t}})} + 1\right]^{-1}$ with $\mu_{\mathrm{b(t)}}$ the chemical potential of the bottom (top) contact, the current flowing from the bottom to the top contact is given by[47] (using a compact notation where a capital bold face symbols represent matrix elements evaluated in some one-particle electron basis and omitting the frequency argument of the different quantities)

$$\begin{aligned} I_{\mathrm{b}\rightarrow\mathrm{t}} = {} & \frac{e}{\hbar} \int \frac{d\omega}{2\pi} f_{\mathrm{b}}(\omega) \mathrm{Tr}\left[\boldsymbol{\Gamma}_{\mathrm{b}} \cdot \boldsymbol{A}\right] \\ & + \frac{e}{\hbar} \int \frac{d\omega}{2\pi} i \mathrm{Tr}\left[\boldsymbol{\Gamma}_{\mathrm{b}} \cdot \boldsymbol{G}^{<}\right], \end{aligned} \tag{16}$$

with the spectral function of the central mesoscopic device given by

$$\begin{aligned} \boldsymbol{A} &= i\left(\boldsymbol{G}^R - \boldsymbol{G}^A\right) \\ &= i\left(\boldsymbol{G}^> - \boldsymbol{G}^<\right), \end{aligned} \tag{17}$$

where $G^{R/A/</>}$ is the retarded/advanced/lesser/greater Green's function of the central device (which takes into account coupling to the external contacts) and $\mathbf{\Gamma}_{\mathrm{b(t)}}$ is a level width function due to the bottom (top) contact. The level width function is the density of states of the contacts weighted by the their coupling to the central device: $\mathbf{\Gamma}_{\mathrm{b(t)}} = 2\pi\boldsymbol{\tau}_{\mathrm{b(t)}} \cdot \delta\left(\omega - \boldsymbol{H}_{\mathrm{b(t)}}\right) \cdot \boldsymbol{\tau}^{\dagger}_{\mathrm{b(t)}}$, with $\boldsymbol{H}_{\mathrm{b(t)}}$ the Hamiltonian describing the bottom (top) contact and $\boldsymbol{\tau}_{\mathrm{b(t)}}$ describing the coupling between the central device and the contact. The second equality in Eq. (17) is true by the very definition of the different Green's functions. A property that will latter be useful is[48]

$$\begin{aligned} \boldsymbol{A} &= \boldsymbol{G}^R \cdot \boldsymbol{\Gamma} \cdot \boldsymbol{G}^A \\ &= \boldsymbol{G}^A \cdot \boldsymbol{\Gamma} \cdot \boldsymbol{G}^R, \end{aligned} \qquad (18)$$

where the decay rate matrix is defined as $\boldsymbol{\Gamma} = -i\left(\left[\boldsymbol{G}^R\right]^{-1} - \left[\boldsymbol{G}^A\right]^{-1}\right)$. This result can be obtained by writing

$$\begin{aligned} \boldsymbol{G}^R - \boldsymbol{G}^A &= \boldsymbol{G}^R \cdot \left(\left[\boldsymbol{G}^A\right]^{-1} - \left[\boldsymbol{G}^R\right]^{-1}\right) \cdot \boldsymbol{G}^A \\ &= \boldsymbol{G}^A \cdot \left(\left[\boldsymbol{G}^A\right]^{-1} - \left[\boldsymbol{G}^R\right]^{-1}\right) \cdot \boldsymbol{G}^R. \end{aligned} \qquad (19)$$

Using the Dyson equation for the retarded/advanced Green's function, $\left[\boldsymbol{G}^{R/A}\right]^{-1} = \left[\boldsymbol{G}^{0,R/A}\right]^{-1} - \boldsymbol{\Sigma}^{R/A}$, with $\boldsymbol{G}^0$ indicating the bare Green's function (in the absence of interactions and coupling to external contacts), and nothing that $\left[\boldsymbol{G}^{0,R}\right]^{-1}$ and $\left[\boldsymbol{G}^{0,A}\right]^{-1}$ only differ by an infinitesimal constant that is taken to zero, the decay rate matrix can be written as

$$\begin{aligned} \boldsymbol{\Gamma} &= i\left[\boldsymbol{\Sigma}^R - \boldsymbol{\Sigma}^A\right] \\ &= i\left[\boldsymbol{\Sigma}^> - \boldsymbol{\Sigma}^<\right], \end{aligned} \qquad (20)$$

where the last identity is inherited from the second equality in Eq. (17). The lesser/greater Green's functions obey the Keldysh equation[47]

$$\boldsymbol{G}^{\lessgtr} = \boldsymbol{G}^R \cdot \boldsymbol{\Sigma}^{\lessgtr} \cdot \boldsymbol{G}^A, \qquad (21)$$

where the lesser/greater self-energy can be split into contributions from the contacts and interactions as

$$\begin{aligned} \boldsymbol{\Sigma}^< &= if_{\mathrm{b}}\boldsymbol{\Gamma}_{\mathrm{b}} + if_{\mathrm{t}}\boldsymbol{\Gamma}_{\mathrm{t}} + \boldsymbol{\Sigma}^<_{\mathrm{int}}, &(22) \\ \boldsymbol{\Sigma}^> &= -i\left(1 - f_{\mathrm{b}}\right)\boldsymbol{\Gamma}_{\mathrm{b}} - i\left(1 - f_{\mathrm{t}}\right)\boldsymbol{\Gamma}_{\mathrm{t}} + \boldsymbol{\Sigma}^>_{\mathrm{int}}, &(23) \end{aligned}$$

with $\boldsymbol{\Sigma}^{\lessgtr}_{\mathrm{int}}$ the contribution from interactions. In the same manner, the decay rate Eq. (20) can be split into a contribution from external contacts and interactions

$$\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_{\mathrm{b}} + \boldsymbol{\Gamma}_{\mathrm{t}} + \boldsymbol{\Gamma}_{\mathrm{int}}. \qquad (24)$$

Using Eqs. (18) and (21)-(24) in Eq. (16), the total current can then be written as a sum of coherent and incoherent contributions

$$I_{\mathrm{b}\to\mathrm{t}} = I^{\mathrm{coh}}_{\mathrm{b}\to\mathrm{t}} + I^{\mathrm{incoh}}_{\mathrm{b}\to\mathrm{t}}, \qquad (25)$$

with the coherent contribution being given by the Landauer formula

$$I^{\mathrm{coh}}_{\mathrm{b}\to\mathrm{t}} = \frac{e}{\hbar} \int \frac{d\omega}{2\pi} \left(f_{\mathrm{b}} - f_{\mathrm{t}}\right) \mathcal{T}, \qquad (26)$$

with the transmission function $\mathcal{T}$ given by

$$\mathcal{T} = \mathrm{Tr}\left[\boldsymbol{\Gamma}_{\mathrm{b}} \cdot \boldsymbol{G}^R \cdot \boldsymbol{\Gamma}_{\mathrm{t}} \cdot \boldsymbol{G}^A\right] \qquad (27)$$

and the incoherent contribution, which describes sequential tunneling processes and plays the same role as vertex corrections in the Kubo formalism for linear response, being given by

$$\begin{aligned} I^{\mathrm{incoh}}_{\mathrm{b}\to\mathrm{t}} &= \frac{e}{\hbar} \int \frac{d\omega}{2\pi} if_{\mathrm{b}}\mathrm{Tr}\left[\boldsymbol{\Gamma}_{\mathrm{b}} \cdot \boldsymbol{G}^A \cdot \boldsymbol{\Sigma}^>_{\mathrm{int}} \cdot \boldsymbol{G}^R\right] \\ &+ \frac{e}{\hbar} \int \frac{d\omega}{2\pi} i\left(1 - f_{\mathrm{b}}\right) \mathrm{Tr}\left[\boldsymbol{\Gamma}_{\mathrm{b}} \cdot \boldsymbol{G}^R \cdot \boldsymbol{\Sigma}^<_{\mathrm{int}} \cdot \boldsymbol{G}^A\right]. \end{aligned} \qquad (28)$$

In the following sections we will use these general formalism together with the model Hamiltonian from Sec. II A to evaluate the vertical current in graphene-hBN-graphene structures.

III. CURRENT IN THE NON-INTERACTING, PRISTINE LIMIT

A. General discussion

When applying the non-equilibrium Green's function formalism to a graphene-hBN-graphene device with metal contacts, one is faced with the issue of how to make the separation between central mesoscopic region, and the external contacts which are in thermal equilibrium. Two natural approaches exist: (A) describing the graphene layers as part of the external contacts, and (B) describing the graphene layers as part of the central mesoscopic device. In all theoretical analytic works to date, the graphene layers were assumed to be part of the external contacts[23,24,27,49,50], being in equilibrium. However, due to the low density of states of graphene, it seems more natural to consider graphene as part of the mesoscopic device. We will start from approach (B) and see that under certain approximations, it reduces to approach (A). We will first consider the non-interacting, pristine case. In approach (B), $\boldsymbol{G}^{R/A}(\omega)$ in Eq. (27) is the Green's function of the graphene-hBN-graphene device. We are interested in the matrix elements of $\boldsymbol{G}^{R/A}$ that connect the bottom and the top contacts. Due to the block diagonal structure of the Hamiltonian (1) (there is no direct coupling between the two graphene layers), these can generally be written as

$$\left[\boldsymbol{G}^R\right]_{\mathrm{b,t}} = \boldsymbol{G}^{0,R}_{\mathrm{bg}} \cdot \boldsymbol{\mathcal{T}}_{\mathrm{bg,tg}} \cdot \boldsymbol{G}^{0,R}_{\mathrm{tg}}, \qquad (29)$$

$$\left[\boldsymbol{G}^A\right]_{\mathrm{t,b}} = \boldsymbol{G}^{0,A}_{\mathrm{tg}} \cdot \boldsymbol{\mathcal{T}}_{\mathrm{tg,bg}} \cdot \boldsymbol{G}^{0,A}_{\mathrm{bg}}, \qquad (30)$$

where $\boldsymbol{G}^{0,R/A}_{\mathrm{bg/tg}}$ are the Green's function of the bottom/top graphene layer in the absence of graphene-hBN coupling
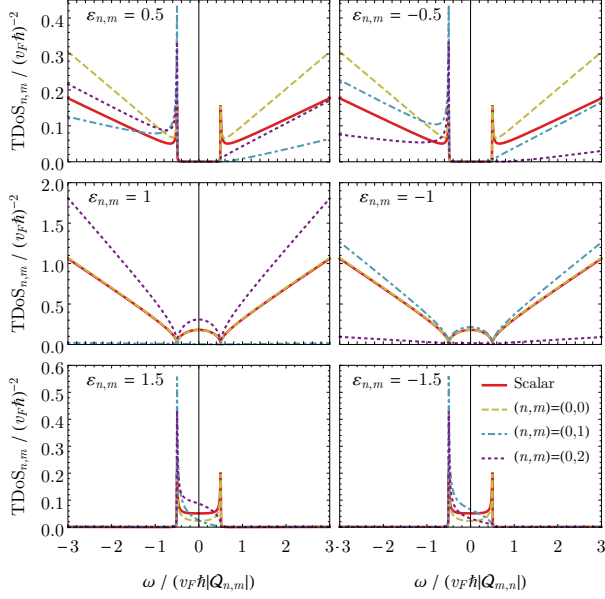
Figure 2. Plot of the quantity $\mathrm{TDoS}_{m,n}(\omega + \varepsilon_{m,n} v_F \hbar |\boldsymbol{Q}_{n,m}|/2, \omega - \varepsilon_{n,m} v_F \hbar |\boldsymbol{Q}_{n,m}|/2)$ for different values of $\varepsilon_{m,n}$ as a function of the energy at zero magnetic field and for rotation angles of $\theta_{\mathrm{TG}} = 1$ and $\theta_{\mathrm{hBN}} = 1.5$. The solid red line shows the the tunneling density of states if the wavefunction overlap factors $\Upsilon^{B/T,n}_{\boldsymbol{k},\lambda}$ in Eq. (40) are set to one. A constant broadening factor of $\gamma = 2.5 \times v_F \hbar |\boldsymbol{Q}_{n,m}| \times 10^{-3}$ was used in all plots.

(but tacking into account coupling to the external contacts) where we have defined the hBN mediated tunneling amplitudes

$$\boldsymbol{\mathcal{T}}_{\mathrm{bg,tg}} = \boldsymbol{T}_{\mathrm{bg,hBN}} \cdot \boldsymbol{G}^R_{\mathrm{hBN}} \cdot \boldsymbol{T}_{\mathrm{hBN,tg}}, \qquad (31)$$

$$\boldsymbol{\mathcal{T}}_{\mathrm{tg,bg}} = \boldsymbol{T}_{\mathrm{tg,hBN}} \cdot \boldsymbol{G}^A_{\mathrm{hBN}} \cdot \boldsymbol{T}_{\mathrm{hBN,bg}}, \qquad (32)$$

with $\boldsymbol{G}^{R/A}_{\mathrm{hBN}}$ the Green's function of the hBN slab, which in general takes into account its coupling to the graphene layers. Therefore, the transmission function Eq. (27) can be written as

$$\mathcal{T} = \mathrm{Tr}\left[\boldsymbol{G}^{0,A}_{\mathrm{bg}} \cdot \boldsymbol{\Gamma}_{\mathrm{b}} \cdot \boldsymbol{G}^{0,R}_{\mathrm{bg}} \boldsymbol{\mathcal{T}}_{\mathrm{bg,tg}} \cdot \right.$$
$$\left. \cdot \boldsymbol{G}^{0,R}_{\mathrm{tg}} \cdot \boldsymbol{\Gamma}_{\mathrm{t}} \cdot \boldsymbol{G}^{0,A}_{\mathrm{tg}} \cdot \boldsymbol{\mathcal{T}}_{\mathrm{tg,bg}} \right]. \qquad (33)$$

If we now use Eq. (18), we can write the spectral function of the bottom graphene layer taking into account the coupling to the bottom metallic contact, $\boldsymbol{A}^0_{\mathrm{bg}} = i\left(\boldsymbol{G}^{0,R}_{\mathrm{bg}} - \boldsymbol{G}^{0,A}_{\mathrm{bg}}\right)$, as $\boldsymbol{A}^0_{\mathrm{bg}} = \boldsymbol{G}^{0,R}_{\mathrm{bg}} \cdot \boldsymbol{\Gamma}_{\mathrm{b}} \cdot \boldsymbol{G}^{0,A}_{\mathrm{bg}} = \boldsymbol{G}^{0,A}_{\mathrm{bg}} \cdot \boldsymbol{\Gamma}_{\mathrm{b}} \cdot \boldsymbol{G}^{0,R}_{\mathrm{bg}}$ and similarly for the top graphene layer. As such, the transmission function can be written as

$$\mathcal{T} = \mathrm{Tr}\left[\boldsymbol{\mathcal{T}}_{\mathrm{bg,tg}} \cdot \boldsymbol{A}^0_{\mathrm{tg}} \cdot \boldsymbol{\mathcal{T}}_{\mathrm{tg,bg}} \cdot \boldsymbol{A}^0_{\mathrm{bg}} \right]. \qquad (34)$$

Eq. (34) is the result that would be directly obtained, if we followed approach (A) instead, in which case the level width functions are given by $\boldsymbol{\Gamma}_{\mathrm{b}} = \boldsymbol{T}_{\mathrm{hBN,bg}} \cdot \boldsymbol{A}^0_{\mathrm{bg}} \cdot \boldsymbol{T}_{\mathrm{hBN,bg}}$

and $\boldsymbol{\Gamma}_{\mathrm{t}} = \boldsymbol{T}_{\mathrm{hBN,tg}} \cdot \boldsymbol{A}^0_{\mathrm{tg}} \cdot \boldsymbol{T}_{\mathrm{hBN,tg}}$. As such we have proved that in the non-interacting case both approaches (A) and (B) coincide. We will leave the discussion for tunneling in the presence disorder and electron-phonon interactions to the next section.

In order to make analytic progress, we will employ the wide-band limit for the metallic contacts, neglecting any frequency dependence of $\boldsymbol{\Gamma}_{\mathrm{b/t}}$, and assume that the contacts couple equally to all graphene states, not spoiling translation invariance. We expect that this last approximation works well for cases where the metallic contacts are deposited on a small region of the graphene sample. Within these approximations, the only effect of the metallic contacts is to introduce a broadening factor of $\gamma_{\mathrm{bg/tg}} = \Gamma_{\mathrm{b/t}}/2$ in the Green's function of the bottom/top graphene layer. We will now write $\mathcal{T}$ for a graphene-hBN-graphene device more explicitly. Using the graphene-hBN coupling Hamiltonian Eq. (5), the transmission function can be written using the Bloch momentum basis as (writing explicitly the frequency argument)

$$\mathcal{T}(\omega) = \sum_{\substack{\boldsymbol{k},\lambda,\lambda' \\ n,m,\tau}} \left| {}_{\mathrm{tg}}\langle \boldsymbol{k} + \tau\boldsymbol{Q}_{n,m}, \tau, \lambda' | \boldsymbol{\mathcal{T}}_{\mathrm{tg,bg}}(\omega) | \boldsymbol{k}, \tau, \lambda \rangle_{\mathrm{bg}} \right|^2$$
$$\times A^0_{\mathrm{tg},\boldsymbol{k}+\tau\boldsymbol{Q}_{n,m},\tau,\lambda'}(\omega_{\mathrm{tg}}) A^0_{\mathrm{bg},\boldsymbol{k},\tau,\lambda}(\omega_{\mathrm{bg}}) \qquad (35)$$

with the sum on n, m going from 0 to 2 and where $\omega_{\mathrm{bg}} = \omega + \epsilon_{\mathrm{F,bg}}$ and $\omega_{\mathrm{tg}} = \omega + \epsilon_{\mathrm{F,tg}} + eV_{\mathrm{bias}}$ are measured from the position of the Dirac point in the bottom and top graphene layers, respectively. The effective tunneling probability can be written as

$$\left| {}_{\mathrm{tg}}\langle \boldsymbol{k} + \boldsymbol{Q}_{n,m}, \tau, \lambda' | \boldsymbol{\mathcal{T}}_{\mathrm{tg,bg}}(\omega) | \boldsymbol{k}, \tau, \lambda \rangle_{\mathrm{bg}} \right|^2 =$$
$$= \Upsilon^{\mathrm{bg},n}_{\boldsymbol{k},\tau,\lambda} \Upsilon^{\mathrm{tg},m}_{\boldsymbol{k}+\tau\boldsymbol{Q}_{n,m},\tau,\lambda} |\mathcal{T}_{n,m}(\omega)|^2, \quad (36)$$

where $\Upsilon^{\mathrm{bg/tg},n}_{\boldsymbol{k},\tau,\lambda} = 1 + \tau\lambda\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{K}}_{\mathrm{bg/tg},n}$, with $\boldsymbol{K}_{\mathrm{bg/tg},n} = \mathcal{R}\left(n2\pi/3\right) \cdot \boldsymbol{K}_{\mathrm{bg/tg}}$, are graphene wavefunction overlap factors and

$$\mathcal{T}_{n,m}(\omega) =$$
$$= \frac{1}{2}\mathrm{tr}\left\{ \hat{\boldsymbol{T}}^\dagger \cdot \boldsymbol{R}^m_{-p\frac{2\pi}{3}} \cdot \left[\boldsymbol{G}^A_{\mathrm{hBN}}(\omega)\right]_{\mathcal{N},1} \cdot \boldsymbol{R}^n_{\frac{2\pi}{3}} \cdot \hat{\boldsymbol{T}} \right\}, \quad (37)$$

with the trace being performed over the sublattice degrees of freedom. Neglecting the frequency dependence of $\boldsymbol{G}^{R/A}_{\mathrm{hBN}}$ and to lowest order in $t_\perp$ we can write

$$|\mathcal{T}_{n,m}|^2 \simeq \left(\frac{t_\perp^2}{E_{\mathrm{B}} E_{\mathrm{N}}}\right)^{\mathcal{N}-1} \times$$
$$\times \begin{cases} 4\frac{t^2_{\mathrm{B,C}} t^2_{\mathrm{N,C}}}{E_{\mathrm{B}} E_{\mathrm{N}}} \cos^2\left(\frac{\pi}{3}(n-m)\right) & , \mathcal{N} \text{ is even} \\ \frac{t^4_{\mathrm{B,C}}}{E^2_{\mathrm{B}}} + \frac{t^4_{\mathrm{N,C}}}{E^2_{\mathrm{N}}} + 2\frac{t^2_{\mathrm{B,C}} t^2_{\mathrm{N,C}}}{E_{\mathrm{B}} E_{\mathrm{N}}} \cos\left(\frac{2\pi}{3}(n-m)\right) & , \mathcal{N} \text{ is odd} \end{cases}$$
$$(38)$$

Notice, that in Eq. (35) both valleys give the same contribution, which can be seen by making a simultaneous

change $\tau \to -\tau$ and $\boldsymbol{k} \to -\boldsymbol{k}$. The transmission function can then be written as

$$\mathcal{T}(\omega) = A g_s g_v \sum_{n,m=0}^{3} |\mathcal{T}_{n,m}|^2 \, \text{TDoS}_{n,m}(\omega_{\text{bg}}, \omega_{\text{tg}}) \quad (39)$$

where A is the area of the device, $g_s = g_v = 2$ are the spin and valley degeneracies and we have defined the tunneling density of states as

$$\text{TDoS}_{n,m}(\omega_{\text{bg}}, \omega_{\text{tg}}) = \sum_{\lambda,\lambda'=\pm 1} \int \frac{d^2\boldsymbol{k}}{(2\pi)^2} \Upsilon^{\text{bg},n}_{\boldsymbol{k},\lambda} \Upsilon^{\text{tg},m}_{\boldsymbol{k}+\boldsymbol{\mathcal{Q}}_{n,m},\lambda'} \times$$
$$\times A^0_{\text{bg},\boldsymbol{k},\lambda}(\omega_{\text{bg}}) A^0_{\text{tg},\boldsymbol{k}+\boldsymbol{\mathcal{Q}}_{n,m},\lambda'}(\omega_{\text{tg}}),$$
$$(40)$$

which only depends of the graphene's dispersion relation and wavefunctions (for simplicity we have dropped the valley indice τ). In the limit of infinite lifetime for graphene electrons, the spectral functions reduce to δ-functions and it is possible to provide an analytic expression for $\text{TDoS}_{n,m}(\omega_{\text{bg}}, \omega_{\text{tg}})$. In the presence of a finite, momentum independent, lifetime, it is still possible to find an approximate analytic expression to Eq. (40). These analytic expressions lead to a significant speed up in the evaluation of the current and are presented in Appendix C. The presence of the spectral functions for the bottom and top graphene layers leads to conservation of energy and momentum in the tunneling process between the two graphene layers.

B. Results

The tunneling in a graphene-hBN-graphene structure is controlled both by energy-momentum conservation and by Pauli's exclusion principle. The constrains imposed by energy-momentum conservation can be understood considering that the Dirac cones of the bottom and top graphene layers are shifted in energy by a value of $\epsilon_{\text{F,tg}} + eV_{\text{bias}} - \epsilon_{\text{F,bg}}$ and in momentum by a value of $|\boldsymbol{\mathcal{Q}}_{n,m}|$, see Fig. 3. The intersection of the shifted cones allows one to visualize the states which respect energy-momentum conservation[24]. Whenever the bias voltage is tuned such that

$$\epsilon_{\text{F,tg}} + eV_{\text{bias}} - \epsilon_{\text{F,bg}} = v_F \hbar |\boldsymbol{\mathcal{Q}}_{n,m}|, \quad (41)$$

there is a complete overlap of the Dirac cones and a maximum in the current occurs. The information regarding energy-momentum conservation for an electron tunneling between the two graphene layers is encoded in the the $\text{TDoS}_{n,m}$ function. In Fig. 2 we plot the quantity $\text{TDoS}_{n,m}(\omega - \varepsilon_{n,m} v_F \hbar |\boldsymbol{\mathcal{Q}}_{n,m}|/2, \omega + \varepsilon_{n,m} v_F \hbar |\boldsymbol{\mathcal{Q}}_{n,m}|/2)$, for different values of $\varepsilon_{n,m} = (\epsilon_{\text{F,tg}} + eV_{\text{bias}} - \epsilon_{\text{F,bg}})/(v_F \hbar |\boldsymbol{\mathcal{Q}}_{n,m}|)$. For $\varepsilon^2_{n,m} < 1$, the tunneling is due to intraband processes (from the conduction/valence band of the bottom graphene into the conduction/valence band of the top graphene), going to
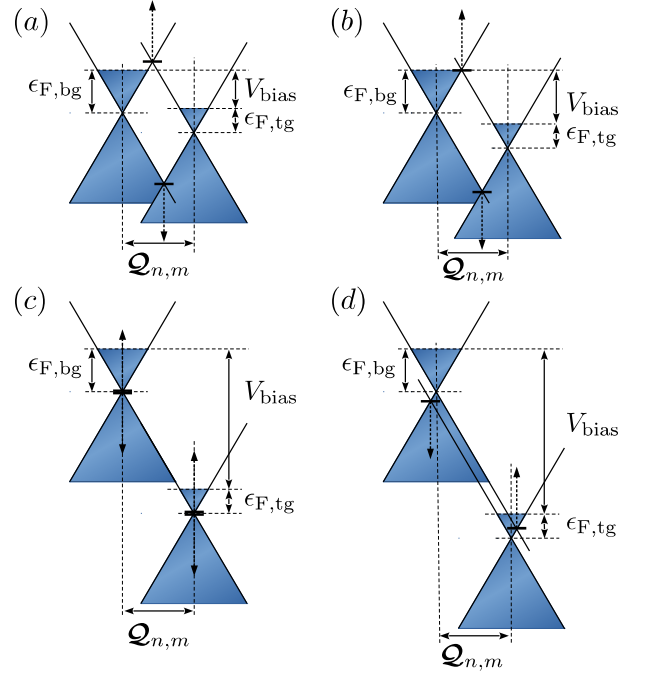

Figure 3. Band diagram representing the constrains imposed by energy-momentum conservation and Pauli's exclusion principle in the vertical current of a graphene-hBN-graphene device. The two cones represent the dispersion relation for electrons of the bottom and top graphene layers. The shadowed blue regions represent the occupation of electronic states in both graphene layers. Energy-momentum conservation is only satisfied when the two shifted Dirac cones intersect and the energy windows where this occurs are represented by the dashed arrows. The following cases are represented: (a) Only intraband are possible, $\varepsilon_{n,m} < 1$, these are however Pauli blocked or there are no states available, therefore in the low temperature limit, no vertical current flows. (b) Threshold bias voltage above which intraband processes which satisfy energy-momentum conservation appear in the energy window where tunneling is allowed by the electronic occupation factors. (c) Condition which corresponds to the occurrence of a peak in the current, when $\varepsilon_{n,m} = 1$, when both intraband and interband (conduction-to-valence and valence-to-conduction) processes are allowed. (d) If one further increases the bias voltage, only interband tunneling, $\varepsilon_{n,m} > 1$, becomes possible and the current diminishes.

zero in the pristine limit for $\omega^2 < (v_F \hbar)^2 |\boldsymbol{\mathcal{Q}}_{n,m}|^2/4$. For $\varepsilon^2_{n,m} > 1$, the tunneling is due to interband processes (from the conduction/valence band of the bottom graphene layer to the valence/conduction band of the top graphene layer), being zero in the pristine limit for $\omega^2 > (v_F \hbar)^2 |\boldsymbol{\mathcal{Q}}_{n,m}|^2/4$. For $\varepsilon^2_{n,m} = 1$, $\text{TDoS}_{n,m}(\omega - \varepsilon_{n,m} v_F \hbar |\boldsymbol{\mathcal{Q}}_{n,m}|/2, \omega + \varepsilon_{n,m} v_F \hbar |\boldsymbol{\mathcal{Q}}_{n,m}|/2)$ diverges in the pristine limit for any value of ω. This divergence in $\text{TDoS}_{n,m}$ leads to a divergence in the vertical current[23,24], which is made finite with the introduction of a finite electronic lifetime. Since for different processes (n,m) with different $|\boldsymbol{\mathcal{Q}}_{n,m}|$ there is a different effective separation in momentum between the Dirac cones of the
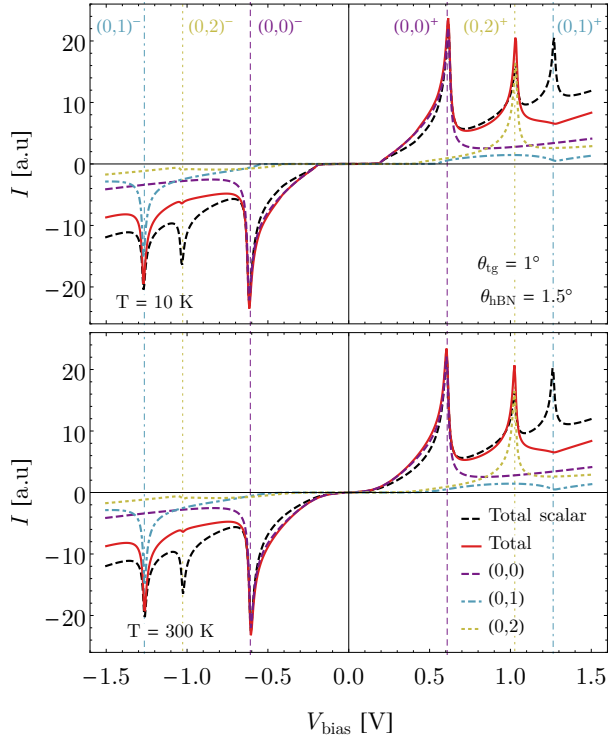
Figure 4. I-V curves for vertical current in a graphene-hBN-graphene device with 4 layers of hBN for rotation angles of $\theta_{\rm tg} = 1$ and $\theta_{\rm hBN} = 1.5$ at gate voltage $V_{\rm gate} = 0$ for two different temperatures. The solid red line indicates the current due to all the 9 processes coupling both graphene layers, for graphene electrons, while the dashed black lines represents the total current for scalar electrons (by setting the wavefunction factors $\Upsilon_{\boldsymbol{k},\lambda}^{\rm bg/tg,n}$ to 1). The remaining lines represent the contributions to the current arising from processes involving different $\boldsymbol{\mathcal{Q}}_{n,m}$ (taking into account the relations imposed by 3-fold rotational invariance, Eq. 12) The dashed vertical lines labeled by $(n,m)^{\pm}$ mark the bias voltages when the condition $\epsilon_{\rm F,tg} + eV_{\rm bias} - \epsilon_{\rm F,bg} = \pm v_F \hbar |\boldsymbol{\mathcal{Q}}_{n,m}|$ is satisfied. Notice that while for scalar electrons all the expected peaks in the current are present, for Dirac electrons some of them are absent. Is is due to the suppression by the $\Upsilon_{\boldsymbol{k},\lambda}^{\rm bg/tg,n}$ factors. A constant broadening factor of $\gamma = 2.5$ meV was used.

Figure 5. I-V curves at constant $V_{\rm gate} = 0$ in a graphene-hBN-graphene device with 4 layers of hBN, at $V_{\rm gate} = 0$ and $T = 300$ K, for different rotation angles between the top and bottom graphene layers, and the hBN slab and the bottom graphene layer. The black dashed line marks the bias voltage when $\varepsilon_{0,0} = \pm 1$ (a condition that is independent of $\theta_{\rm hBN}$). The remaining vertical lines mark the bias voltages when $\varepsilon_{n,m} = \pm 1$ for $n \neq m$ for different values $\theta_{\rm hBN}$ (the color and type of line match the ones used in the plots).

bottom and top graphene layers, one expects the occurrence of multiple peaks in the I-V curve, followed by subsequent regions of negative differential conductance. This is indeed the case as shown in Fig. (4). Based only on energy-momentum conservation, one would expect the occurrence of three peaks in the I-V curve for positive bias voltage and another three for negative bias (notice that according to the discussion of Sec. II A from the nine processes coupling the two graphene layers, only three are independent). This is indeed the case as shown in Fig. (4). However, the computed curve only displays two peaks, with the peaks corresponding to the situations when and $\varepsilon_{0,2} = -1$ being absent. The reason for the suppression of these peaks is due to the spinorial structure of graphene electronic wavefunctions, via the overlap
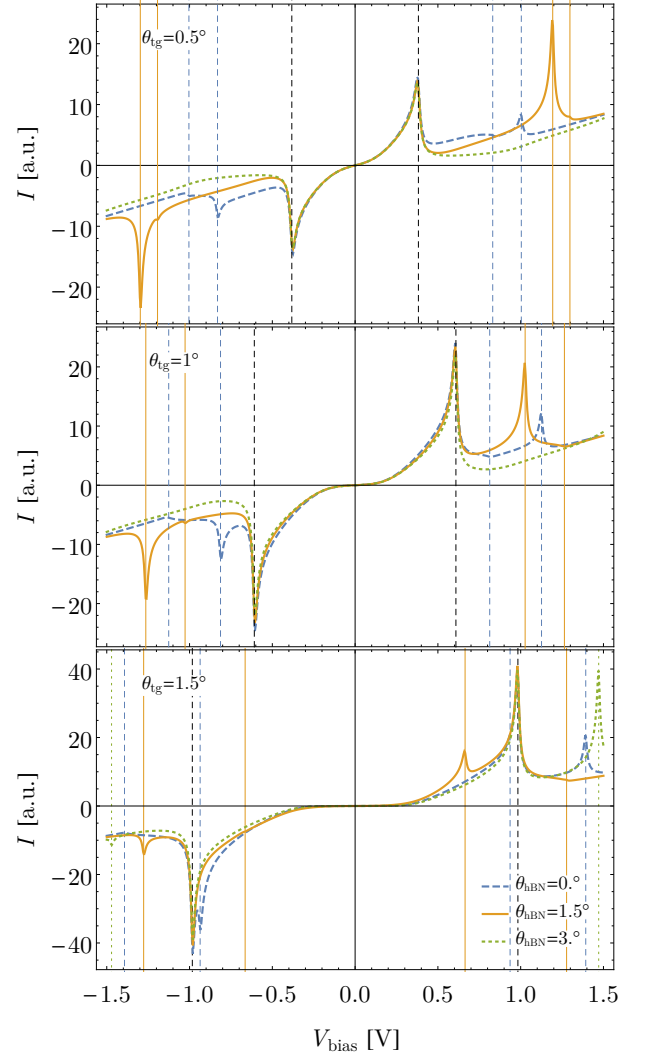
factors $\Upsilon_{\boldsymbol{k},\lambda}^{\rm bg/tg,n}$, that appear in Eq. (40). These overlap factors can severely suppress the value of $\text{TDoS}_{n,m}$ close to $\varepsilon_{n,m} = \pm 1$ and consequently of the height of the peaks in the I-V curve. This is shown in Fig. 2 , where it is shown a considerable suppression of $\text{TDoS}_{n,m}$ for $\varepsilon_{0,1} = 1$ and $\varepsilon_{0,2} = -1$. The effect of the overlap factors can also be seen in Fig. (4), where it is also shown the current that would be obtained, if the electronic wavefunction of graphene where scalars, i.e. by setting $\Upsilon_{\boldsymbol{k},\lambda}^{\rm bg/tg,n} = 1$ in (40) (see Eq C20 in Appendix C), displaying the three peaks expected by kinematic considerations. While the occurrence of NDC in graphene-hBN-graphene has al-
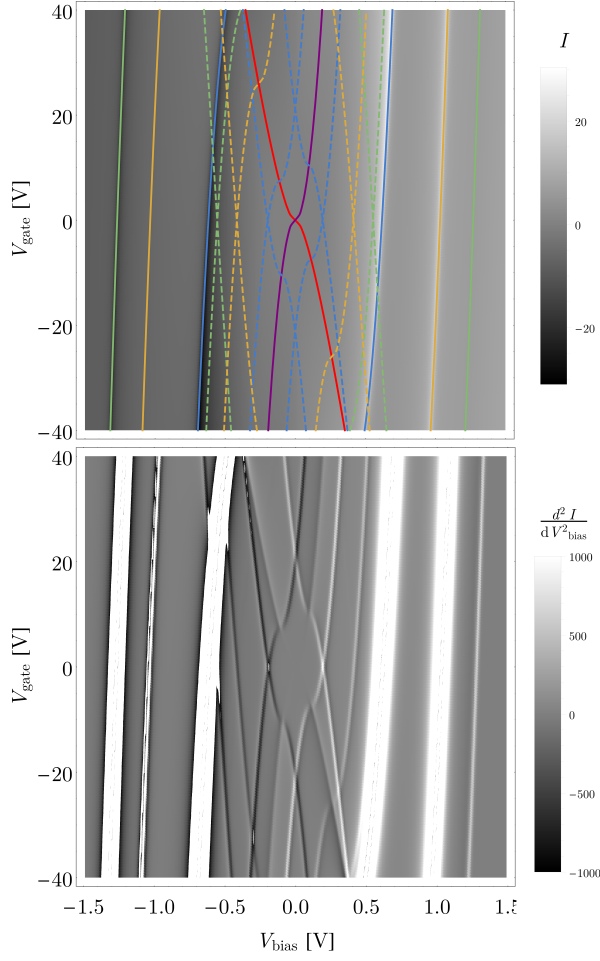
Figure 6. Density plot of current, I, and its second derivative with respect to the applied bias voltage, d^2I/dV_{bias}^2, as a function of the applied bias and gate voltages at $T = 10$ K. In the current plot, it is also shown the lines defined by the following conditions: $\epsilon_{\text{F,bg}} = 0$ and $\epsilon_{\text{F,tg}} = 0$, represented by the solid lines in red and purple; $\epsilon_{\text{F,tg}} + eV_{\text{bias}} - \epsilon_{\text{F,bg}} = \pm v_F \hbar |\mathcal{Q}_{0,m}|$ (Eq. (41)) for $m = 0$, 1 and 2, represented by the solid lines in blue, green and yellow, respectively; $\epsilon_{\text{F,tg}} \pm eV_{\text{bias}} + \epsilon_{\text{F,bg}} = \pm \frac{1}{2} v_F \hbar |\mathcal{Q}_{0,m}|$ for $m = 0$, 1 and 2, (Eq. (42)) represented by the dashed lines in blue, green and yellow. Notice now the guides lines shown in the current plot match perfectly the sharp features shown in the d^2I/dV_{bias}^2 plot. Also, the peaks expected to occur through channels $(0,1)^+$ and $(0,2)^-$ are absent. A constant broadening factor of $\gamma = 2.5$ meV was assumed for both layers.

ready been experimentally observed[23], the occurrence of multiple NDC regions has not. This might be due to the fact that the position of the current peaks depends very sensitively in the rotation angles θ_{tg} and θ_{hBN}. This is exemplified in Fig. 5, where the computed I-V curves for several rotation angles are shown. As shown, for a fixed angle of $\theta_{\text{tg}} = 1°$, changing θ_{hBN} from 1.5° to 3° moves the additional peaks in the current due to the transfer of momentum by the hBN crystal lattice from a bias voltage of ~ 1 V to bias voltages > 1.5 V. Tunneling processes
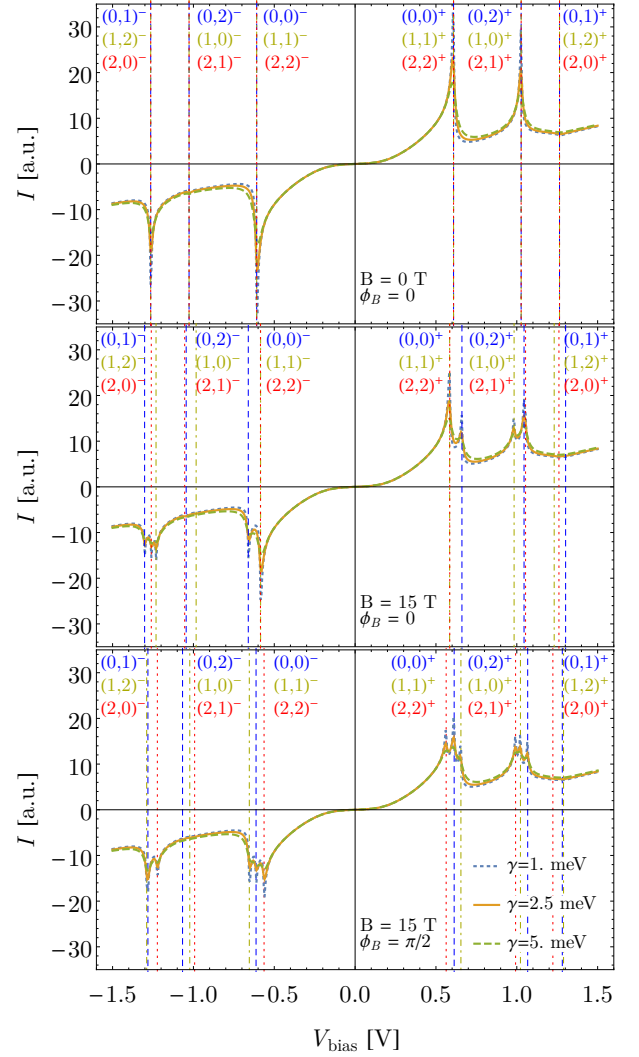

Figure 7. I-V curves for a graphene-hBN-graphene device with 4 layers of hBN, with rotation angles of $\theta_{\text{tg}} = 1$ and $\theta_{\text{hBN}} = 1.5$ at constant $V_{\text{gate}} = 0$ and $T = 300$ K, for different values and orientation of the in-plane magnetic field and electronic broadening factor. The vertical lines, labeled by $(n,m)^\pm$ mark the bias voltages for which $\varepsilon_{n,m} = \pm 1$. Notice how the applied magnetic field leads to a splitting of the peaks that occur at zero magnetic field. As the broadening factor is increased, the peaks become less resolved.

which satisfy energy-momentum conservation, can only contribute to the current if these lie in an energy window between the zero of energy and the bias voltage, as presented in Fig. 3. The condition for which processes allowed by energy-momentum conservation become allowed by the occupation factors occurs in the limit of zero temperature when (see Fig. 3.(b))

$$\epsilon_{\text{F,tg}} \pm eV_{\text{bias}} + \epsilon_{\text{F,bg}} = \pm \frac{1}{2} v_F \hbar |\mathcal{Q}_{n,m}|. \qquad (42)$$

This explains the occurrence of the plateau with nearly zero current seen at low temperature in Fig. 4, and gives origin to the features in the d^2I/dV_{bias}^2 as a function of
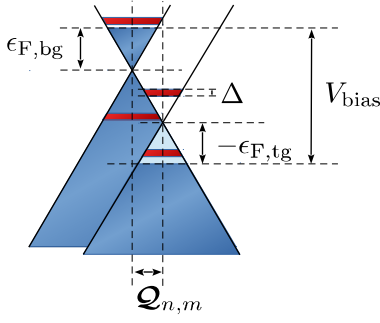
Figure 8. Diagram representing the possible effect of the reconstruction of the graphene Dirac spectrum, due to the presence of hBN, in the vertical current of a graphene-hBN-graphene device when $\varepsilon_{n,m} = \pm 1$. The red bars represent the position in energy of the regions, of width Δ, where graphene's spectrum reconstruction is significant. Provided $eV_{\text{bias}}/\Delta \gg 1$, the peaks in the current will still be present.

applied bias and gate voltages as seen in the density plot of Fig. 6. At higher temperatures, all these sharp features tend to vanish, as the Fermi-Dirac occupation factors become smoother.

By applying an in-plane magnetic field, the three-fold rotational invariance of the graphene-hBN-graphene structure is broken, and therefore, the processes corresponding to the different groups in (12) will contribute differently to the current, and one expects that each peak in the I-V curve will split into three. An in-plane magnetic field of the form $\boldsymbol{B} = B\left(\cos\phi_B, \sin\phi_B, 0\right)$ can be described by the vector potential $\boldsymbol{A} = Bz\left(\sin\phi_B, -\cos\phi_B\right)$. Neglecting the momentum dependence of H_{hBN} the effect of the in-plane magnetic field reduces to an additional transference of momentum to the tunneling electrons,, which is encoded in a shift in the $\boldsymbol{\mathcal{Q}}_{n,m}$ vectors[23,24,51–53]

$$\boldsymbol{\mathcal{Q}}_{n,m} \to \boldsymbol{\mathcal{Q}}_{n,m} + \frac{eBd}{\hbar}\left(\sin\phi_B, -\cos\phi_B\right). \qquad (43)$$

The splitting of the peaks in the I-V curve is shown in Fig. (7), where it is also shown the effect of an increasing electronic broadening factor.

Finally, we comment the possible effect of the hBN in the electronic structure of graphene. It is known that the potential modulation with the periodicity of the Moir pattern formed by graphene on top of hBN can lead to a reconstruction of the density of states of graphene at energies $\sim \pm v_F\hbar\left|\boldsymbol{g}_{1/2}^{\text{g,hBN}}\right|/2$ measured from the original Dirac cone, where $\left|\boldsymbol{g}_{1/2}^{\text{g,hBN}}\right| \simeq 4\pi\sqrt{\delta^2 + \theta_{\text{g,hBN}}^2}/(\sqrt{3}a_{\text{g}})$ is the wavevector of the Moir pattern reciprocal lattice[16–21] with $\theta_{\text{g,hBN}}$ is the rotation angle between the graphene layer and the hBN slab. We have disregarded such effects in our discussion. As we have seen in Fig. (5), the additional peaks in the current enabled by the transference of momentum by the hBN lattice, only appear for reasonable values of the bias voltage for small twist an-

gles between the graphene layers and hBN slab. It is precisely in this case that that the reconstruction of the graphene dispersion relations becomes important at low energy. The effect of this reconstruction should impact not only the peaks that involve transference of momentum by the hBN lattice ($n \neq m$), but also the ones that do not ($n = m$). In this situation one can question the validity of the results from these section. However, we argue that the possible reconstruction of the graphene dispersion relations, should not affect in a profound way the occurrence of peaks and NDC in the I-V curves of graphene-hBN-graphene devices. The energy width, Δ, where the reconstruction of the linear dispersion relation of graphene is significant is of the order of the tens of meV[17,54], while the total energy window of states that contribute to the current is, at low temperatures, of the width of $\sim eV_{\text{bias}}$. Provided the condition $eV_{\text{bias}}/\Delta \gg 1$ is satisfied (see Fig. (8)), we expect that the effect of the dispersion relation reconstruction is negligible, and apart from a possible reduction of the height of the peaks, should not affect the current in any drastic way.

IV. INCOHERENT CURRENT: PHONON AND DISORDER ASSISTED TUNNELING

A. General discussion

We will now study, in a unified way, the effect of phonons and disorder in the current of a graphene-hBN-graphene device. We consider a generic electron-phonon interaction described by the Hamiltonian

$$H_{\text{e-ph}} = \boldsymbol{c}^\dagger \cdot \boldsymbol{M}_\zeta \cdot \boldsymbol{c}\phi_\zeta, \qquad (44)$$

where $\phi_\zeta = \left(a_\zeta + a_\zeta^\dagger\right)/\sqrt{2}$ is the phonon field operator, with $a_\zeta^\dagger$ the creation operator for a phonon mode ζ, $\boldsymbol{M}_\zeta$ is an electron-phonon coupling matrix and $\boldsymbol{c}^\dagger$ is row vector of electronic creation operators in an arbitrary basis. For this electron-phonon interaction, the Fock (or sunset)[55] contribution to the lesser/greater self-energy is given by (from now on we write all frequency arguments explicitly)

$$\boldsymbol{\Sigma}_{\text{e-ph}}^{\lessgtr}(\omega) = i\sum_\zeta \int \frac{d\nu}{2\pi}\boldsymbol{M}_\zeta \cdot \boldsymbol{G}^{\lessgtr}(\omega-\nu) \cdot \boldsymbol{M}_\zeta^\dagger D_\zeta^{\lessgtr}(\nu), \ (45)$$

where $D_\zeta^{\lessgtr}(\nu)$ is the lesser/greater Green's function for the phonon field operator, which, assuming the phonons are in thermal equilibrium, are given by

$$D_\zeta^{\lessgtr}(\nu) = \mp i2\pi b(\pm\nu)2\omega_\zeta\text{sgn}(\nu)\delta\left(\nu^2 - \omega_\zeta^2\right), \qquad (46)$$

where $b(\nu) = \left(e^{\beta\nu} - 1\right)^{-1}$ is the Bose-Einstein distribution function, which satisfies $1+b(\nu) = -b(-\nu)$, and ω_ζ is phonon frequency of mode ζ. Therefore, the self-energy

due to electron-phonon interaction reads

$$\boldsymbol{\Sigma}_{\text{e-ph}}^{\lessgtr}(\omega) = \sum_{\zeta,s=\pm 1} (\pm sb(\pm s\omega_\zeta)) \, \boldsymbol{M}_\zeta \cdot \boldsymbol{G}^{\lessgtr}(\omega - s\omega_\zeta) \cdot \boldsymbol{M}_\zeta^\dagger. \tag{47}$$

We point out that this self-energy can also describe elastic scattering by impurities by drop the summation over s, take $\omega_\zeta \to 0$ and set $\pm sb(\pm s\omega_\zeta) \to 1$, in which case the quantity $\boldsymbol{M}_\zeta \boldsymbol{M}_\zeta^\dagger$ is to be interpreted as the disorder correlator. With this in mind, the following discussion applies both to inelastic scattering by phonons and elastic scattering by impurities. Combining Eq. 47 with Eqs. (21) and (22), it is possible to write to lowest order in the electron-phonon interaction

$$\boldsymbol{\Sigma}_{\text{e-ph}}^<(\omega) \simeq \sum_{\zeta,s=\pm 1} if_b(\omega - s\omega_\zeta)\, sb(s\omega_\zeta) \times$$
$$\times \boldsymbol{M}_\zeta \cdot \boldsymbol{G}^R(\omega - s\omega_\zeta) \cdot \boldsymbol{\Gamma}_b(\omega - s\omega_\zeta) \cdot \boldsymbol{G}^A(\omega - s\omega_\zeta) \cdot \boldsymbol{M}_\zeta^\dagger.$$
$$+ \sum_{\zeta,s=\pm 1} if_t(\omega - s\omega_\zeta)\, sb(s\omega_\zeta) \times$$
$$\times \boldsymbol{M}_\zeta \cdot \boldsymbol{G}^R(\omega - s\omega_\zeta) \cdot \boldsymbol{\Gamma}_t(\omega - s\omega_\zeta) \cdot \boldsymbol{G}^A(\omega - s\omega_\zeta) \cdot \boldsymbol{M}_\zeta^\dagger, \tag{48}$$

with $\boldsymbol{\Sigma}_{\text{e-ph}}^>(\omega)$ obtained by replacing $f_{b/t} \to 1 - f_{b/t}$ and $b(s\omega_\zeta) \to b(-s\omega_\zeta)$. Inserting this expression in Eq. (28), we obtain the lowest order contribution to the

non-coherent current

$$I_{b\to t}^{\text{incoh,1-ph}} =$$
$$= \frac{e}{\hbar} \sum_{\zeta,s} \int \frac{d\omega}{2\pi} f_b(\omega) \left[1 - f_t(\omega - s\omega_\zeta)\right] (-sb(-s\omega_\zeta)) \, \mathcal{T}_{b,t}^{(\zeta,s)}(\omega)$$
$$- \frac{e}{\hbar} \sum_{\zeta,s} \int \frac{d\omega}{2\pi} \left(1 - f_b(\omega)\right) f_t(\omega - s\omega_\zeta) \, sb(s\omega_\zeta) \mathcal{T}_{b,t}^{(\zeta,s)}(\omega)$$
$$+ \frac{e}{\hbar} \sum_{\zeta,s} \int \frac{d\omega}{2\pi} f_b(\omega) \left[1 - f_b(\omega - s\omega_\zeta)\right] (-sb(-s\omega_\zeta)) \, \mathcal{T}_{b,b}^{(\zeta,s)}(\omega)$$
$$- \frac{e}{\hbar} \sum_{\zeta,s} \int \frac{d\omega}{2\pi} \left(1 - f_b(\omega)\right) f_b(\omega - s\omega_\zeta) \, sb(s\omega_\zeta) \mathcal{T}_{b,b}^{(\zeta,s)}(\omega), \tag{49}$$

where the 1-phonon (disorder) assisted transmission function is given by

$$\mathcal{T}_{\ell,\ell'}^{(\zeta,s)}(\omega) = \text{Tr} \left[\boldsymbol{\Gamma}_\ell(\omega) \cdot \boldsymbol{G}^R(\omega) \cdot \boldsymbol{M}_\zeta \cdot \boldsymbol{G}^R(\omega - s\omega_\zeta) \cdot \right.$$
$$\left. \cdot \, \boldsymbol{\Gamma}_{\ell'}(\omega - s\omega_\zeta) \cdot \boldsymbol{G}^A(\omega - s\omega_\zeta) \cdot \boldsymbol{M}_\zeta^\dagger \cdot \boldsymbol{G}^A(\omega) \right]. \tag{50}$$

It is easy to check that

$$f_b(\omega) \left[1 - f_b(\omega - s\omega_\zeta)\right] (-sb(-s\omega_\zeta)) =$$
$$= \left(1 - f_b(\omega)\right) f_b(\omega - s\omega_\zeta) \, sb(s\omega_\zeta) \tag{51}$$

and as such the last two terms of Eq. (49) cancel each other. This cancellation is required since in a steady state no charge accumulation can occur in the device and therefore, the current flowing from the top to the bottom contact should satisfy $I_{t\to b} = -I_{t\to b}$. As such, terms that involve only the occupation factor of one the contacts must cancel at the end of any calculation. Processes assisted by a greater number of phonons can also be obtained. Higher order corrections to Eq. (48) can be obtained by iterating Eq. (47) using Eqs. (21) and (22). Just as for the lowest order case, contributions involving only occupation factors from one of the contacts cancel each other. Therefore, the contribution to the incoherent current assisted by n phonons can be written as

$$I_{b\to t}^{\text{incoh, }n\text{-ph}} =$$
$$= \frac{e}{\hbar} \sum_{\zeta_1,s_1,\ldots,\zeta_n,s_n} \int \frac{d\omega}{2\pi} f_b(\omega) \left(1 - f_t(\omega + s_1\omega_{\zeta_1} + \ldots + s_n\omega_{\zeta_n})\right) \left(s_1 b(s_1\omega_{\zeta_1})\right) \ldots \left(s_n b(s_n\omega_{\zeta_n})\right) \mathcal{T}_{b\to t}^{(\zeta_1,s_1)\ldots(\zeta_n,s_n)}(\omega)$$
$$- \frac{e}{\hbar} \sum_{\zeta_1,s_1,\ldots,\zeta_n,s_n} \int \frac{d\omega}{2\pi} f_t(\omega) \left(1 - f_b(\omega + s_1\omega_{\zeta_1} + \ldots + s_n\omega_{\zeta_n})\right) \left(s_1 b(s_1\omega_{\zeta_1})\right) \ldots \left(s_n b(s_n\omega_{\zeta_n})\right) \mathcal{T}_{t\to b}^{(\zeta_1,s_1)\ldots(\zeta_n,s_n)}(\omega), \tag{52}$$

where we have defined the n-phonon assisted transmission functions

$$\mathcal{T}_{\text{b}\to\text{t}}^{(\zeta_1,s_1)....(\zeta_n,s_n)}(\omega) = \text{Tr}\left[\mathbf{\Gamma}_{\text{b}}(\omega)\cdot\boldsymbol{G}^R(\omega)\cdot\boldsymbol{M}_{\zeta_1}\cdot\boldsymbol{G}^R(\omega+s_1\omega_{\zeta_1})...\cdot\boldsymbol{M}_{\zeta_n}\cdot\boldsymbol{G}^R(\omega+s_1\omega_{\zeta_1}+...+s_n\omega_{\zeta_n})\right.$$
$$\left.\mathbf{\Gamma}_{\text{t}}(\omega+s_1\omega_{\zeta_1}+...+s_n\omega_{\zeta_n})\cdot\boldsymbol{G}^A(\omega+s_1\omega_{\zeta_1}+...+s_n\omega_{\zeta_n})\cdot\boldsymbol{M}_{\zeta_n}^\dagger\cdot...\cdot\boldsymbol{G}^A(\omega+s_1\omega_{\zeta_1})\cdot\boldsymbol{M}_{\zeta_1}^\dagger\boldsymbol{G}^A(\omega)\right], \quad (53)$$

$$\mathcal{T}_{\text{t}\to\text{b}}^{(\zeta_1,s_1)....(\zeta_n,s_n)}(\omega) = \text{Tr}\left[\mathbf{\Gamma}_{\text{t}}(\omega)\cdot\boldsymbol{G}^A(\omega)\cdot\boldsymbol{M}_{\zeta_1}^\dagger\cdot\boldsymbol{G}^A(\omega+s_1\omega_{\zeta_1})\cdot...\cdot\boldsymbol{M}_{\zeta_n}^\dagger\boldsymbol{G}^A(\omega+s_1\omega_{\zeta_1}+...+s_n\omega_{\zeta_n}).\right.$$
$$\left.\mathbf{\Gamma}_{\text{b}}(\omega+s_1\omega_{\zeta_1}+...+s_n\omega_{\zeta_n})\cdot\boldsymbol{G}^R(\omega+s_1\omega_{\zeta_1}+...+s_n\omega_{\zeta_n})\cdot\boldsymbol{M}_{\zeta_n}\cdot...\cdot\boldsymbol{G}^R(\omega+s_1\omega_{\zeta_1})\cdot\boldsymbol{M}_{\zeta_1}\cdot\boldsymbol{G}^R(\omega)\right]. \quad (54)$$

Notice that with respect to Eq. (49), we have made a change of $s_i \to -s_i$ in the first line and made a shift in the frequency variable $\omega \to \omega + s_1\omega_{\zeta_1} + ... + s_n\omega_{\zeta_n}$ in the second line of Eq. (52). Eqs. (52), (53) and (54) have a very simple interpretation. The first/second line of Eq. (52) can be understood has the probability of an electron being injected from the bottom/top contact being collected by the top/bottom contact, while being scattering by n phonons during the contact to contact trip, with $s_i = \pm 1$ representing a phonon absorption/emission process. We will now use this general formalism to study the effect of phonon scattering in vertical transport in a graphene-hBN-graphene device. We will analyze separately the effect of scattering by graphene and hBN phonons.

1. Scattering by phonons in the graphene layers

We know return to the issue of the consequences of considering graphene as part of the external contacts or part of the central mesoscopic region. We will first study the effect of multiple scatterings of electrons in the graphene layers by phonons (or impurities). We will first focus on scattering by phonons in the top graphene layer, with scattering in the bottom layer being treated in the same way. Using Eq. (53), the tunneling amplitude assisted by n phonon scattering events in the the top graphene layer can be written to lowest order in the graphene-hBN coupling as

$$\mathcal{T}_{\text{b}\to\text{t}}^{(\zeta_1,s_1)....(\zeta_n,s_n)}(\omega) \simeq \text{Tr}\left[\boldsymbol{\mathcal{T}}_{\text{tg,bg}}(\omega)\cdot\mathbf{\Gamma}_{\text{b}}(\omega)\cdot\boldsymbol{G}_{\text{bg}}^R(\omega)\cdot\boldsymbol{\mathcal{T}}_{\text{bg,tg}}(\omega)\cdot\right.$$
$$\cdot\boldsymbol{G}_{\text{tg}}^R(\omega)\cdot\boldsymbol{M}_{\zeta_1}\cdot\boldsymbol{G}_{\text{tg}}^R(\omega+s_1\omega_{\zeta_1})...\cdot\boldsymbol{M}_{\zeta_n}\cdot\boldsymbol{G}_{\text{tg}}^R(\omega+s_1\omega_{\zeta_1}+...+s_n\omega_{\zeta_n})\cdot\mathbf{\Gamma}_{\text{t}}(\omega+s_1\omega_{\zeta_1}+...+s_n\omega_{\zeta_n})\cdot$$
$$\left.\cdot\boldsymbol{G}_{\text{tg}}^A(\omega+s_1\omega_{\zeta_1}+...+s_n\omega_{\zeta_n})\cdot\boldsymbol{M}_{\zeta_n}^\dagger\cdot....\cdot\boldsymbol{G}_{\text{tg}}^A(\omega+s_1\omega_{\zeta_1})\cdot\boldsymbol{M}_{\zeta_1}^\dagger\boldsymbol{G}_{\text{tg}}^A(\omega)\right], \quad (55)$$

and similarly for $\mathcal{T}_{\text{t}\to\text{b}}^{(\zeta_1,s_1)....(\zeta_n,s_n)}(\omega)$. These contributions correspond to multiple scatterings of an electron before leaving the top graphene layer. Summing up all the contributions of the form of Eq. (55), together with the contribution from the coherent current, we obtain

$$I_{\text{b}\to\text{t}} = \sum_{n=0}^{\infty} I_{\text{b}\to\text{t}}^{\text{incoh, }n-\text{ph}}$$
$$= \frac{e}{\hbar}\sum_{n=0}^{\infty}\int\frac{d\omega}{2\pi}f_{\text{b}}(\omega)\left(1-f_{\text{t}}(\omega)\right)\times$$
$$\times\text{Tr}\left[\boldsymbol{A}_{\text{bg}}(\omega)\cdot\boldsymbol{\mathcal{T}}_{\text{bg,tg}}(\omega)\cdot\boldsymbol{A}_{\text{tg}}^{(n)}(\omega)\cdot\boldsymbol{\mathcal{T}}_{\text{tg,bg}}(\omega)\right]$$
$$- \frac{e}{\hbar}\sum_{n=0}^{\infty}\int\frac{d\omega}{2\pi}f_{\text{t}}(\omega)\left(1-f_{\text{b}}(\omega)\right)\times$$
$$\times\text{Tr}\left[\boldsymbol{A}_{\text{tg}}^{(n)}(\omega)\cdot\boldsymbol{\mathcal{T}}_{\text{tg,bg}}(\omega)\cdot\boldsymbol{A}_{\text{bg}}(\omega)\cdot\boldsymbol{\mathcal{T}}_{\text{bg,tg}}(\omega)\right]$$
$$(56)$$

where we have written $\boldsymbol{A}_{\text{bg}}(\omega) = \boldsymbol{G}_{\text{bg}}^A(\omega)\cdot\mathbf{\Gamma}_{\text{b}}(\omega)\cdot\boldsymbol{G}_{\text{bg}}^R(\omega)$, since we are considering only scattering in the top graphene layer. It can be checked that the different terms $\boldsymbol{A}_{\text{tg}}^{(n)}(\omega)$ obey the following recursion relation

$$\boldsymbol{A}_{\text{tg}}^{(0)}(\omega) = \boldsymbol{G}^R(\omega)\cdot\mathbf{\Gamma}_{\text{t}}(\omega)\cdot\boldsymbol{G}^A(\omega), \; n=0 \quad (57)$$
$$\boldsymbol{A}_{\text{tg}}^{(n)}(\omega) = \sum_{s_n,\zeta_n}\frac{\left[1-f_{\text{t}}\left(\omega-s_n\omega_{\zeta_n}\right)\right]\left[-s_n b(-s_n\omega_{\zeta_n})\right]}{1-f_{\text{t}}(\omega)}\times$$
$$\times\boldsymbol{G}_{\text{tg}}^R(\omega)\cdot\boldsymbol{M}_{\zeta_n}\cdot\boldsymbol{A}_{\text{tg}}^{(n-1)}(\omega-s_n\omega_{\zeta_n})\cdot$$
$$\cdot\boldsymbol{M}_{\zeta_n}^\dagger\cdot\boldsymbol{G}_{\text{tg}}^A(\omega), \; n>0. \quad (58)$$

This can be compared with the spectral function of the top graphene layer. Assuming that the top graphene layer is in near equilibrium with the top contact, then the spectral function can be written as

$$\boldsymbol{A}_{\text{tg}}(\omega) \simeq \boldsymbol{G}^R(\omega)\cdot\left(\mathbf{\Gamma}_{\text{t}}(\omega)+\mathbf{\Gamma}_{\text{e-ph,tg}}(\omega)\right)\cdot\boldsymbol{G}^A(\omega), \quad (59)$$
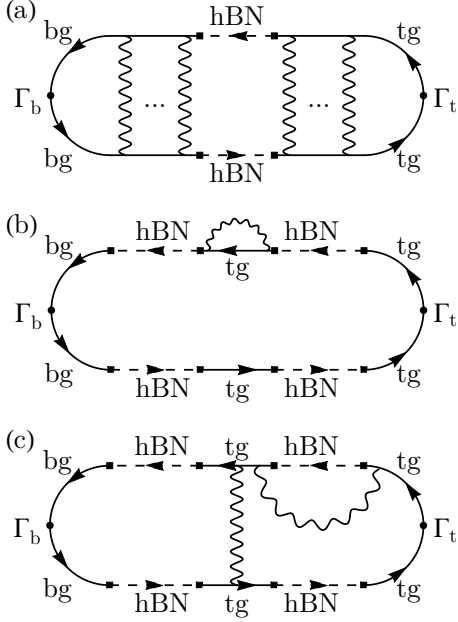
Figure 9. Diagrammatic representation of contributions to the current involving phonon scattering. The dots represent the level width functions due to the bottom and top external metallic contacts, the squares represent the graphene-hBN coupling, solid lines represent graphene electronic propagators and dashed lines represent hBN propagators. The wiggly lines represent phonon propagators. (a) Ladder diagrams that are resumed in Eq. (63). To lowest order in the graphene-hBN coupling these are all the contributions due to electron-phonon interaction in the graphene layers. (b) Diagram contributing to the current in higher order in the graphene-hBN coupling, including the renormalization of the top graphene layer Green's function by phonons. This kind of diagram can be captured in Eq. (63), provided the effect of coupling to the graphene layers is included into $\boldsymbol{G}_{\mathrm{hBN}}$. (c) Higher order diagrams in the graphene-hBN coupling, including electron-phonon interaction in the graphene layers, that is not included in Eq. (63) and more generically cannot be captured when evaluating the current following approach (A).

where, under the approximation that the top graphene is in equilibrium with the top contact, the decay rate due to electron-phonon interaction can be written as

$$\boldsymbol{\Gamma}_{\mathrm{e\text{-}ph,tg}}(\omega) \simeq \sum_{s,\zeta} s \left(1 - f_{\mathrm{t}}(\omega - s\omega_\zeta) + b(s\omega_\zeta)\right) \times$$
$$\times \, \boldsymbol{M}_\zeta \cdot \boldsymbol{A}_{\mathrm{tg}}(\omega - s\omega_\zeta) \cdot \boldsymbol{M}_\zeta^\dagger. \quad (60)$$

It is easy to check that the equilibrium occupation functions satisfy the equality

$$\frac{[1 - f(\omega - s\omega_\zeta)][-sb(-s\omega_\zeta)]}{1 - f(\omega)} =$$
$$= s\left(1 - f(\omega - s\omega_\zeta) + b(s\omega_\zeta)\right). \quad (61)$$

Therefore, by inserting Eq. (60) into Eq. (59) and iterating the equation, we obtain

$$\boldsymbol{A}_{\mathrm{tg}}(\omega) \simeq \sum_{n=0}^{\infty} \boldsymbol{A}_{\mathrm{tg}}^{(n)}(\omega), \quad (62)$$

with the different terms $\boldsymbol{A}_{\mathrm{tg}}^{(n)}(\omega)$ coincide with Eqs. (57) and (58), and the $\simeq$ means we are making the approximation that the top graphene layer is in near equilibrium with the top contact. Therefore, the sum of all incoherent scattering processes occurring before the electron leaves the graphene layer and the coherent contribution reproduces the spectral function of graphene taking into account electron-phonon interaction / disorder. The same is true for scattering in the bottom graphene layer. Notice that in Eqs. (53) and (54) retarded/advanced Green's functions appear to the right/left of $\boldsymbol{\Gamma}_{\mathrm{b}}$. Nevertheless, by using Eq. (18), the previous calculation can also be applied for scattering in the bottom graphene layer. We have thus arrived to an important conclusion: the expression

$$I_{\mathrm{b}\to\mathrm{t}} = \frac{e}{\hbar} \int \frac{d\omega}{2\pi} \left(f_{\mathrm{b}}(\omega) - f_{\mathrm{t}}(\omega)\right) \times$$
$$\times \, \mathrm{Tr}\left[\mathcal{T}_{\mathrm{bg,tg}}(\omega) \cdot \boldsymbol{A}_{\mathrm{tg}}(\omega) \cdot \mathcal{T}_{\mathrm{tg,bg}} \cdot \boldsymbol{A}_{\mathrm{bg}}(\omega)\right], \quad (63)$$

which would be the one obtained if we employed approach (A), actually already includes the effect of multiple non-coherent scattering processes in the graphene layers, provided $\boldsymbol{A}_{\mathrm{tg/bg}}(\omega)$ are replaced with the respective expressions in the presence of phonon/disorder scattering. We also point out that in the case of elastic scattering due to disorder in the graphene layers, the result from Eq. (63) can be obtained by performing disorder averages of Eq. (33), see Appendix F. To lowest order in the graphene-hBN coupling, Eq. (63) actually includes all the possible scattering processes of an electron in the graphene layers. Including the effects of graphene into the Green's function of hBN that appears in $\mathcal{T}_{\mathrm{bg,tg}}(\omega)$, Eq. (63) includes only a subclass of all possible contributions due to electron-phonon interaction, see Fig. 9. Therefore, we conclude that approaches (A) and (B) coincide to lowest order in the graphene-hBN coupling and to higher order in this coupling, approach (A) can correctly capture a class of all the possible electron phonon scatterings.

2. Scattering by phonons in the hBN slab

We will now discuss the effects of scattering by phonons/disorder in the hBN slab. We will restrict ourselves to the case of tunneling assisted by one phonon. We write the electron phonon interaction in a Bloch state basis as

$$H_{\mathrm{e\text{-}ph,hBN}} = \frac{1}{\sqrt{N}} \sum_{\boldsymbol{k},\boldsymbol{q}} \boldsymbol{c}_{\boldsymbol{k}+\boldsymbol{q},\mathrm{hBN}}^\dagger \cdot \boldsymbol{M}_{\boldsymbol{q},\zeta} \cdot \boldsymbol{c}_{\boldsymbol{k},\mathrm{hBN}} \phi_{\boldsymbol{q},\zeta}, \quad (64)$$

where $c^{\dagger}_{\boldsymbol{k},\text{hBN}} = \left[c^{\dagger}_{\boldsymbol{k},\text{B},1,\text{hBN}}, c^{\dagger}_{\boldsymbol{k},\text{N},1,\text{hBN}}, ..., c^{\dagger}_{\boldsymbol{k},\text{B},\mathcal{N},\text{hBN}}, c^{\dagger}_{\boldsymbol{k},\text{N},\mathcal{N},\text{hBN}} \right]$ and $\phi_{\boldsymbol{q},\zeta} = \left(a_{\boldsymbol{q},\zeta} + a^{\dagger}_{-\boldsymbol{q},\zeta} \right)/\sqrt{2}$ is the phonon field operator and N is the number of unit cells in the hBN slab. For small rotation angles between the different layers and assuming only scattering by phonons close to the Γ or K points of hBN, such that only states close to the Dirac points of each layer are involved, using Eq. 52, we can write the 1-phonon assisted tunneling current to lowest order in the graphene-hBN coupling as

$$
\begin{aligned}
I_{\text{b}\to\text{t}}^{\text{incoh, 1-ph}} = {} & \frac{e}{\hbar N} \sum_{\substack{\boldsymbol{k},\lambda,\lambda' \\ n,m}} \sum_{\boldsymbol{q},\zeta,s} \int \frac{d\omega}{2\pi} f_{\text{b}}(\omega)\left(1 - f_{\text{t}}\left(\omega + s\omega_{\zeta}\right)\right)\left(sb(s\omega_{\zeta})\right) \left| \mathcal{T}_{n,m,\boldsymbol{k},\boldsymbol{q}}^{(\zeta,s)1\text{-ph}}(\omega) \right|^2 \times \\
& \times \Upsilon_{\boldsymbol{k},\lambda}^{\text{bg},n} \Upsilon_{\boldsymbol{k}+\boldsymbol{\mathcal{Q}}_{n,m}-\boldsymbol{q},\lambda'}^{\text{tg},m} A_{\text{bg},\boldsymbol{k},\lambda}(\omega_{\text{bg}}) A_{\text{tg},\boldsymbol{k}+\boldsymbol{\mathcal{Q}}_{n,m}-\boldsymbol{q},\lambda'}(\omega_{\text{tg}} + s\omega_{\zeta}) \\
& - \frac{e}{\hbar N} \sum_{\substack{\boldsymbol{k},\lambda,\lambda' \\ n,m}} \sum_{\boldsymbol{q},\zeta,s} \int \frac{d\omega}{2\pi} f_{\text{t}}\left(\omega + s\omega_{\zeta}\right)\left(1 - f_{\text{b}}(\omega)\right)\left(-sb(-s\omega_{\zeta})\right) \left| \mathcal{T}_{n,m,\boldsymbol{k},\boldsymbol{q}}^{(\zeta,s)1\text{-ph}}(\omega) \right|^2 \times \\
& \times \Upsilon_{\boldsymbol{k},\lambda}^{\text{bg},n} \Upsilon_{\boldsymbol{k}+\boldsymbol{\mathcal{Q}}_{n,m}-\boldsymbol{q},\lambda'}^{\text{tg},m} A_{\text{bg},\boldsymbol{k},\lambda}(\omega_{\text{bg}}) A_{\text{tg},\boldsymbol{k}+\boldsymbol{\mathcal{Q}}_{n,m}-\boldsymbol{q},\lambda'}(\omega_{\text{tg}} + s\omega_{\zeta}),
\end{aligned}
$$

where we have introduced the phonon assisted tunneling amplitude between the graphene layers

$$
\mathcal{T}_{n,m,\boldsymbol{k},\boldsymbol{q}}^{(\zeta,s)1\text{-ph}}(\omega) = \frac{1}{2}\text{tr}\left\{ \hat{\boldsymbol{T}}^{\dagger} \cdot \boldsymbol{R}^m_{-p\frac{2\pi}{3}} \cdot \left[\boldsymbol{G}^A_{\text{hBN},\boldsymbol{k}+\boldsymbol{g}_n^{\text{bg}}-\boldsymbol{q}}(\omega + s\omega_{\zeta}) \cdot \boldsymbol{M}^{\dagger}_{\boldsymbol{q},\zeta} \cdot \boldsymbol{G}^A_{\text{hBN},\boldsymbol{k}+\boldsymbol{g}_n^{\text{bg}}}(\omega) \right]_{\mathcal{N},1} \cdot \boldsymbol{R}^n_{\frac{2\pi}{3}} \cdot \hat{\boldsymbol{T}} \right\}.
$$

Neglecting the momentum and frequency dependence of $\boldsymbol{G}^A_{\text{hBN}}$ and assuming dispersionless phonons, one can make a shift in the momentum variable $\boldsymbol{q} \to \boldsymbol{k} - \boldsymbol{k}' + \boldsymbol{\mathcal{Q}}_{n,m}$, such that the summation over $\boldsymbol{k}$ and $\boldsymbol{k}'$ factors and we can write

$$
\begin{aligned}
I_{\text{b}\to\text{t}}^{\text{incoh, 1-ph}} = {} & A A_{\text{cell}} g_s g_v \frac{e}{\hbar} \sum_{n,m} \sum_{\zeta,s} \int \frac{d\omega}{2\pi} \left[f_{\text{b}}(\omega)\left(1 - f_{\text{t}}(\omega + s\omega_{\zeta})\right) sb(s\omega_{\zeta}) - f_{\text{t}}(\omega + s\omega_{\zeta})\left(1 - f_{\text{b}}(\omega)\right) s\left(1 + b(\omega_{\zeta})\right) \right] \times \\
& \times \left| \mathcal{T}_{n,m}^{(\zeta)1\text{-ph}} \right|^2 \text{DoS}_{\text{bg}}(\omega_{\text{bg}}) \text{DoS}_{\text{tg}}(\omega_{\text{tg}} + s\omega_{\zeta}),
\end{aligned} \tag{65}
$$

, where A_{cell} is the area of the unit cell of hBN and graphene's density of states per spin and valley is given by

$$
\text{DoS}(\omega) = \frac{1}{V} \sum_{\boldsymbol{k},\lambda} A_{\boldsymbol{k},\lambda}(\omega) = \frac{|\omega|}{(v_F \hbar)^2}, \tag{66}
$$

where the last equality is valid for for non-interacting electrons in pristine graphene. A similar expression to Eq. (65), which included only processes involving spontaneous emission of phonons (equivalent to assuming that the phonons are at zero temperature), was recently presented without derivation and used in Ref. 31 to model vertical current in graphene-hBN-graphene devices. In the case of elastic scattering by disorder with short range correlation, Eq. (65) becomes,

$$
\begin{aligned}
I_{\text{b}\to\text{t}}^{\text{incoh, 1-dis}} = {} & A A_{\text{cell}} g_s g_v \frac{e}{\hbar} \sum_{\substack{\zeta,s \\ n,m}} \int \frac{d\omega}{2\pi} \left[f_{\text{b}}(\omega) - f_{\text{t}}(\omega) \right] \times \\
& \times \left| \mathcal{T}_{n,m}^{1\text{-dis}} \right|^2 \text{DoS}_{\text{bg}}(\omega_{\text{bg}}) \text{DoS}_{\text{tg}}(\omega_{\text{tg}}),
\end{aligned} \tag{67}
$$

with $\mathcal{T}_{n,m}^{1\text{-dis}}$ a disorder assisted tunneling amplitude. Although an expression of the form of Eq. 67 was previously used to model vertical current in graphene-hBN-graphene devices[6,7], we emphasize that Eq. 67 only describes processes where there is a complete degradation of in-plane momentum conservation, something that has been previously pointed out in Refs. 24 and 27. The complete degradation of momentum conservation only occurs for scattering by dispersionless phonons or for disorder with short distance correlation.

As an example we consider, scattering by optical out-of-plane breathing modes close to the Γ point, with non-zero components of polarization vector given by

$$
\begin{aligned}
\boldsymbol{\xi}^z_{\text{ZB},a,\ell} &= \left(\xi^z_{\zeta,\text{B},1}, \xi^z_{\zeta,\text{N},1}, \xi^z_{\zeta,\text{B},2}, \xi^z_{\zeta,\text{N},2}, ... \right) \\
&= \sqrt{\frac{\mu_{\text{BN}}}{\mathcal{N}}} \left(\frac{1}{\sqrt{m_{\text{B}}}}, \frac{1}{\sqrt{m_{\text{N}}}}, \frac{-1}{\sqrt{m_{\text{B}}}}, \frac{-1}{\sqrt{m_{\text{N}}}}, ... \right),
\end{aligned} \tag{68}
$$

where $\mu_{\text{BN}}^{-1} = m_{\text{B}}^{-1} + m_{\text{N}}^{-1}$ is the reduced mass of the hBN phonon mode. We assume that electron-phonon coupling for this mode can be described as a local change in the value of the interlayer hoping parameter in Hamiltonian 4. Considering electrons due close to the K point and phonons close to the Γ point, we derive an electron-phonon Hamiltonian of the form of Eq. 64, with a mo-

mentum independent coupling constant which reads

$$\boldsymbol{M}_{\text{ZB}}^{\text{hBN}} = \frac{g_{\text{ZB}}^{\text{hBN}}}{\sqrt{\mathcal{N}}} \begin{bmatrix} \mathbf{0} & \boldsymbol{\sigma}_x & & \\ \boldsymbol{\sigma}_x & \mathbf{0} & -\boldsymbol{\sigma}_x & \\ & -\boldsymbol{\sigma}_x & \mathbf{0} & \ddots \\ & & \ddots & \ddots \end{bmatrix}, \quad (69)$$

with the electron-phonon coupling constant given by

$$g_{\text{ZB}}^{\text{hBN}} = -\frac{\partial \log t_\perp}{\partial \log c_{\text{BN}}} \frac{t_\perp}{c_{\text{BN}}} \sqrt{\frac{\hbar}{\mu_{\text{BN}} \omega_{\text{ZB}}^{\text{hBN}}}}, \quad (70)$$

where $-\partial \log t_\perp / \partial \log c_{\text{BN}} \simeq 3$ describes the change of the interlayer hopping, t, with the interlayer distance, c_{BN}, and $\omega_{\text{ZB}}^{\text{hBN}}$ is the out-of-plane breathing phonon frequency: For this electron-phonon interaction we obtain to lowest order in $t_\perp$ and neglecting the frequency and momentum dependence

$$\left| \mathcal{T}_{n,m}^{(\text{ZB})1\text{-ph}} \right|^2 \simeq \frac{(\mathcal{N}-1)^2}{\mathcal{N}} \left| \frac{g_{\text{ZB}}^{\text{hBN}}}{t_\perp} \right|^2 |\mathcal{T}_{n,m}|^2,$$

with $|\mathcal{T}_{n,m}|^2$ given by Eq. (38).

B. Results

In Fig. 10 we show the vertical current as a function of bias voltage taking into account the effect of scattering of graphene electrons by resonant scatterers (treated within the SCBA, see Appendix D) and in-plane graphene electrons (see Appendix E). For comparison we also show current computed used a constant relaxation time. The main difference between modeling electron scattering with a constant relaxation rate or considering scattering by resonant scatters, is that for resonant scatters the electron decay rate has a strong dependence in energy, behaving as ω^{-1}. Therefore, for higher bias voltages (when the graphene Fermi levels are higher), the electron lifetime is larger. This is manifest in Fig. 10, where it is seen that when assuming a constant relaxation rate the second peak in the I-V current is considerably smaller than the first one, while for resonant scatterers both peaks are roughly the same height. Inclusion of phonons, makes again the peak at higher bias voltage smaller due to the fact that the decay rate due to scattering with graphene in-plane optical phonons increases with frequency. Also notice that inclusion of resonant disorder and phonons leads to a small splitting of the peaks in the I-V current. This splitting is due to real part of the self-energy due to both resonant scatterers and phonons. Apart from increasing graphene electron's decay rate and as such providing an additional broadening of peaks in the I-V current, phonons do not play a relevant role for the high bias I-V characteristics of a graphene-hBN-graphene device. This changes if one focus on small bias. At very low temperature, the spontaneous emission of optical phonons becomes possible
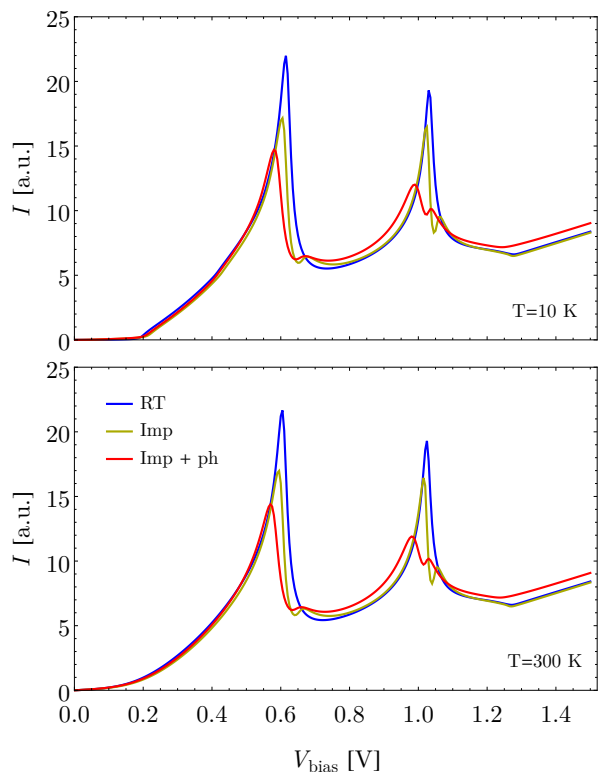

Figure 10. I-V curves at constant $V_{\text{gate}} = 0$ in a graphene-hBN-graphene device with rotations angles $\theta_{\text{tg}} = 1°$ and $\theta_{\text{hBN}} = 1.5°$, considering different sources of scattering in the graphene layers: (RT) constant relaxation time of $\gamma = 3$ meV; (Imp) scattering by resonant scatterers treated within the SCBA with an impurity concentration of $n_{\text{imp}} = 10^{-4}$ impurities per graphene unit cell; (Imp+RT) scattering by resonant scatterers and graphene in-plane optical phonons also with $n_{\text{imp}} = 10^{-4}$.

whenever $V_{\text{bias}} > \omega_{O\text{ph}}$, where $\omega_{O\text{ph}}$ is the optical phonon frequency, opening up new tunneling channels for electrons. Although for small electron-phonon coupling, this phonon assisted contribution to the current is small the opening up of a new tunneling channel can be observed in the derivatives of the current with respect to the bias, as can be seen in Fig. 11. The features in d^2I/dV_{bias}^2 are only significant at low temperature, being smoothed out at higher temperatures due to the smearing of the Fermi occupation factors in graphene. We point out however, that the features due to phonons are a small contribution can be overridden due to features in the coherent current induced by the rotation between different layers (shown in Fig. 6), even if we treat the phonons as dispersionless leading to a complete degradation of electron momentum conservation. We also note in passing, that tunneling assisted by emission of multiple phonons is also possible (see Eqs. (52)-(54)) which would open up new scattering channels when $n\omega_{O\text{ph}} > V_{\text{bias}}$, where n is the number of phonons. These would lead to additional peaks in d^2I/dV_{bias}^2 but would be instead suppressed by higher powers of the electron-phonon coupling.
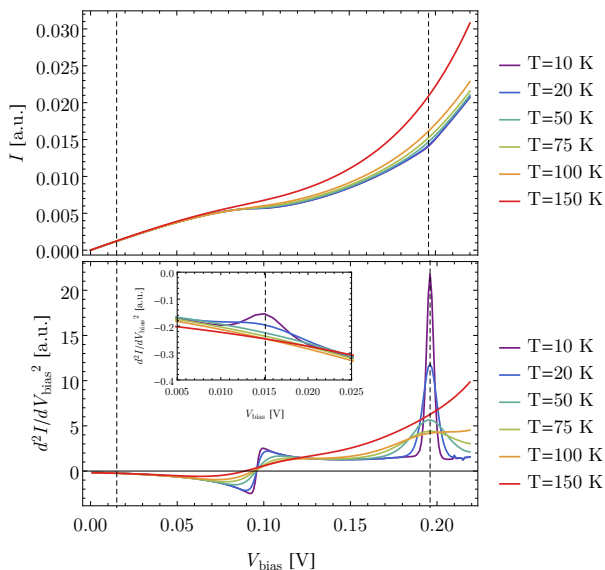
Figure 11. I-V curve and d^2I/dV_{bias}^2 as a function of bias voltages at a constant $V_{\text{gate}} = 10$ V for different temperatures and for rotation angles $\theta_{\text{tg}} = 2°$ and $\theta_{\text{hBN}} = 3°$, including effects of scattering by out-of-plane breathing phonons of hBN, $\omega_{\text{ZB}}^{\text{hBN}} = 15$ meV[56], and of the in-plane graphene phonons, $\omega_{\Gamma\text{O}}^{\text{g}} = 196$ meV[57] (represented by the vertical dashed lines). Processes involving spontaneous emission of phonons open up new tunneling channels that appear as peaks in d^2I/dV_{bias}^2 at low temperature. The inset zooms in the small peak due to the hBN out-of-planes breathing phonon. We point out that the feature that occurs around $V_{\text{bias}} \sim 0.1$ V is not due to phonons, but due to the tunneling density of states structure.

V. CONCLUSIONS

This works provides another example of the extreme sensitivity of the properties of vdW structures to the rotational alignment of the different constitutive layers. We have seen how this additional degree of freedom can be exploited in order to create devices displaying multiple regions of negative differential conductance. The development of devices that display multiple NDC regions is relevant for the development of multivalued logic devices[58,59], which showcases another possible application of vdW structures. We have studied in detail the effect of the rotational alignment between the boron nitride slab and the graphene layers in the vertical current of a graphene-hBN-graphene vdW structure for small rotational misalignment, which have so far not been observed[23]. We have seen now the transference of momentum by the hBN crystalline structure to the tunneling electrons gives origin to additional peaks in the I-V characteristics of this device, followed by regions of negative differential conductance. These additional peaks are however extremely sensitive to the rotation angle between the graphene layers and the hBN slab, and rotational angles as small as 3° can already push these additional peaks to bias voltages higher than 1.5 V. There-

fore, the observation of multiple NDC in graphene-hBN-graphene devices requires a control of the rotational angle between the different layers with a precision of $\lesssim 1°$, something which is within experimental reach[21,23,60]. We expect that the possible reconstruction of graphene spectrum due to the periodic potential induced by hBN for small rotational angles should not affect in a qualitative way the occurrence of multiple NDC regions in graphene-hBN-graphene devices, provided the applied bias voltage is much larger than the width of the region where the spectrum reconstruction is significant. However, a more quantitative treatment of these effects is required.

We have also analyzed the effect of treating graphene as being the source and drain contacts of the graphene-hBN-graphene device, or by treating them as part of the device and taking the source and drain as being external metallic contacts. We have seen that, provided the metallic contacts do not significantly spoil translation invariance of graphene (as expected if the contact is deposited only over a small region of the graphene layer), and in the non-interacting case, both approaches are equivalent. In the presence of interactions both approaches are equivalent to lowest order in the graphene-hBN coupling.

Finally, we have studied, in a unified way, the effect of scattering by disorder and phonon scattering in the vertical current of graphene-hBN-graphene devices. Starting from a NEGF formalism we derived the contribution to the current due to phonon (or disorder) assisted tunneling processes. We have seen now scattering by short range disorder or dispersionless phonons leads to a complete degradation of electron momentum conservation in the graphene-to-graphene tunneling process and how spontaneous emission of phonons at lower temperature appear as sharp features in the derivatives of the current with respect to the bias voltage at the energy of the phonons. These features can however be hidden by features due to the rotational alignment between the different layers. We have focused on the effect of graphene in-plane optical phonons and hBN optical out-of-plane breathing phonons. We have not considered the effect of vibrations at the graphene-hBN interface, as these would require the description of phonons in incommensurate structures something which will be focus of future work.
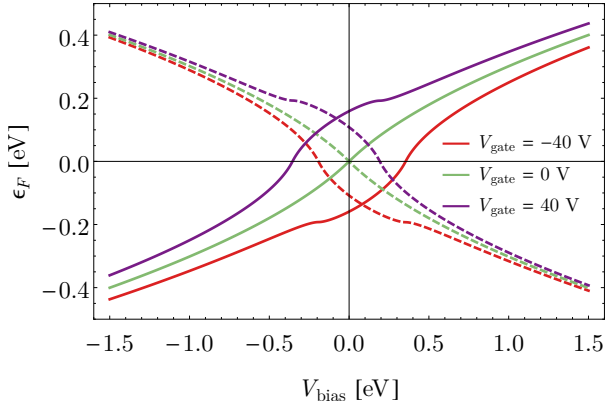
Figure 12. Computed Fermi levels for the bottom and top graphene layers as a function of bias voltage for different gate voltages obtained by solving Eqs. (A9) and (A10). We assume that the following parameters $d_{SiO_2} = 285$ nm, $d_{hBN} = 40$ nm for the thickness of the back gate dielectric, with out-of-plane dielectric constants $\bar{\epsilon}_{SiO_2} = 3.9$ and $\bar{\epsilon}_{hBN} = 5.09$[33]. We assumed that the distance between the two graphene layers are separated by 4 monolayers of hBN, which corresponds to a distance between the graphene layers of $d \simeq 1.6$ nm.

Appendix A: Thomas-Fermi modeling of electrostatic doping

We wish to model the charging of an a graphene-hBN-graphene device by application of a gate, V_{gate}, and bias, V_{bias}, voltages. The graphene-hBN-graphene structure is formed by $\mathcal{N}$ hBN monolayers, sandwiched between two graphene layers. The graphene-hBN-graphene structure is on top of a dielectric spacer (typically hBN/SiO$_2$) separating the structure from a back gate, typically a highly doped Si layer. We treat each layer forming the graphene-hBN-graphene structure as a 2D film with a two dimensional charge density given by ρ_ℓ, $\ell = -1, ..., \mathcal{N}+1$, where $\ell = -1$ indexes the Si layer, $\ell = 0$ and $\ell = \mathcal{N} + 1$ are, respectively, the bottom and top graphene layers and $\ell = 1, ..., \mathcal{N}$ index the layers of hBN slab. Layers $\ell - 1$ and ℓ are separated by a distance d_ℓ and we assume that this is filled with a dielectric with relative constant along the z direction given by given by $\bar{\epsilon}_\ell$. Applying Gauss's law around each plate, and assuming charge neutrality, $\sum_{\ell=-1}^{\mathcal{N}} \rho_\ell = 0$, we obtain

$$\bar{\epsilon}_0 E_0 = \rho_{-1}/\epsilon_0. \tag{A1}$$

$$\bar{\epsilon}_{\ell+1} E_{\ell+1} - \bar{\epsilon}_\ell E_\ell = \rho_\ell/\epsilon_0, \quad \ell = 0, ..., \mathcal{N}, \tag{A2}$$

$$-\bar{\epsilon}_{\mathcal{N}+1} E_{\mathcal{N}+1} = \rho_{\mathcal{N}+1}/\epsilon_0, \tag{A3}$$

where E_ℓ is the electric field along the z direction, between layers $\ell - 1$ and ℓ, and ϵ_0 is vacuum's permittivity. From these equations we can write

$$\bar{\epsilon}_\ell E_\ell = \frac{1}{\epsilon_0} \sum_{k=-1}^{\ell-1} \rho_k, \quad \ell = 0, ..., \mathcal{N}+1, \tag{A4}$$

and the stored electrostatic energy is given by

$$U_{EM} = \sum_{\ell=0}^{\mathcal{N}+1} \frac{1}{2} \epsilon_0 d_\ell \bar{\epsilon}_\ell E_\ell^2$$

$$= \frac{1}{2} \sum_{\ell,\ell'=0}^{\mathcal{N}+1} \rho_\ell \left(\sum_{k=0}^{\min(\ell,\ell')} \frac{d_k}{\epsilon_0 \bar{\epsilon}_k} \right) \rho_{\ell'}. \tag{A5}$$

where we have used the charge neutrality condition in order to eliminate the charge in the Si gate ρ_{-1}. This is nothing more than the Hartree energy for a layered material. We split the charge density of each layer into a contribution from charge carriers and another from charged impurities, $\rho_\ell = -en_\ell + en_\ell^{imp}$, where n_ℓ is the charge carrier concentration ($n_\ell > 0$ for electron doping) and n_ℓ^{imp} is the concentration of charged impurities ($n_\ell^{imp} > 0$ for positively charged impurities). Including the effects of a gate voltage, V_{gate}, applied between the $\ell = -1$ and the $\ell = 0$ layers and a bias voltage between the $\ell = \mathcal{N}+1$ and the $\ell = 0$ layers, we obtain a Thomas-Fermi functional

$$\Phi = \frac{1}{2} \sum_{\ell,\ell'=1}^{\mathcal{N}+1} n_\ell \left(\sum_{k=0}^{\min(\ell,\ell')} \frac{e^2 d_k}{\epsilon_0 \bar{\epsilon}_k} \right) n_{\ell'} - \sum_{\ell=0}^{\mathcal{N}+1} n_\ell e\phi_\ell^{imp}$$

$$- eV_{gate} \sum_{\ell=0}^{\mathcal{N}+1} n_\ell + eV_{bias} \sum_{\ell=0}^{\mathcal{N}+1} \frac{\ell}{\mathcal{N}+1} n_\ell, \tag{A6}$$

where

$$e\phi_\ell^{imp} = \sum_{\ell'=0}^{\mathcal{N}+1} \sum_{k=1}^{\min(\ell,\ell')} \frac{e^2 d_k}{\epsilon_0 \bar{\epsilon}_k} n_{\ell'}^{imp}, \tag{A7}$$

is the potential created by the charged impurities. The Hartree potential felt by electrons in layer ℓ is then given by

$$V_\ell^H = -\frac{\partial \Phi}{\partial n_\ell}$$

$$= V_{gate} - V_{bias} \frac{\ell}{\mathcal{N}+1} n_\ell + e\phi_\ell^{imp}$$

$$- \sum_{\ell'=0}^{\mathcal{N}+1} \left(\sum_{k=0}^{\min(\ell,\ell')} \frac{e^2 d_k}{\epsilon_0 \bar{\epsilon}_k} \right) n_{\ell'}, \quad \ell = 0, ..., \mathcal{N}+1 \tag{A8}$$

Now, we assume that the vertical current flowing between the two graphene layers is small enough, such that we can assume that these are in a near equilibrium state. Furthermore, we employ the Thomas-Fermi approximation, in which the local Fermi level for each layer is given by $\epsilon_{F,\ell} = V_\ell^H$, where $\epsilon_{F,\ell}$ is a function of the local carrier density. This together with Eq. (A8) becomes a system of non-linear equations in the carrier density / local Fermi level.

It can be checked, that due to the large band gap of hBN, most charge density will be accumulated in the graphene layers. As such we approximate $n_\ell = 0$, for

$\ell = 1, ..., \mathcal{N}$ and therefore the $\mathcal{N}+2$ equations are reduced to two

$$\epsilon_{\mathrm{F,bg}} = eV_{\mathrm{gate}} - (n_{\mathrm{bg}} + n_{\mathrm{tg}})\, C_{\mathrm{t}}^{-1} + e\phi_{\mathrm{bg}}^{\mathrm{imp}} \tag{A9}$$

$$\epsilon_{\mathrm{F,tg}} = eV_{\mathrm{gate}} - eV_{\mathrm{bias}} - n_{\mathrm{tg}}C_{\mathrm{t}}^{-1} - n_{\mathrm{bg}}C_{\mathrm{b}}^{-1} + e\phi_{\mathrm{tg}}^{\mathrm{imp}} \tag{A10}$$

where the capacitances are given by (taking into account the series capacitances of a hBN/SiO$_2$ spacer with d_{hBN} the hBN thickness and $d_{\mathrm{SiO_2}}$ the SiO$_2$ thickness)

$$C_{\mathrm{b}}^{-1} = \frac{e^2 d_0}{\epsilon_0 \bar{\epsilon}_0} = \frac{e^2 d_{\mathrm{SiO_2}}}{\epsilon_0 \bar{\epsilon}_{\mathrm{SiO_2}}} + \frac{e^2 d_{\mathrm{hBN}}}{\epsilon_0 \bar{\epsilon}_{\mathrm{hBN}}}$$

$$C_{\mathrm{t}}^{-1} = \sum_{\ell=0}^{\mathcal{N}+1} \frac{e^2 d_\ell}{\epsilon_0 \bar{\epsilon}_\ell} \tag{A11}$$

$$= \frac{e^2 d_{\mathrm{SiO_2}}}{\epsilon_0 \bar{\epsilon}_{\mathrm{SiO_2}}} + \frac{e^2 d_{\mathrm{hBN}}}{\epsilon_0 \bar{\epsilon}_{\mathrm{hBN}}} + \frac{e^2 d}{\epsilon_0 \bar{\epsilon}_{\mathrm{hBN}}}, \tag{A12}$$

and d is the distance between the two graphene layers. The terms $e\phi_{\mathrm{bg/tg}}^{\mathrm{imp}}$ are the potentials induced by the charged impurities in the bottom/top graphene layer that can be tuned to account for intrinsic doping of the graphene layers (acting as an offset in the measurement of V_{gate} and V_{bias}). We finally point out that in the case where the hBN layers have no charge carrier, then the Hartree potential within the hBN slab is given from Eq. (A8) in terms of $\epsilon_{\mathrm{F,bg/}tg}$ as

$$V_\ell^H = \epsilon_{\mathrm{F,tg}} - e\phi_{\mathrm{tg}}^{\mathrm{imp}} + e\phi_\ell^{\mathrm{imp}}$$
$$- \frac{\ell}{\mathcal{N}+1}\left(\epsilon_{\mathrm{F,tg}} + eV_{\mathrm{bias}} - \epsilon_{\mathrm{F,bg}} - e\phi_{\mathrm{tg}}^{\mathrm{imp}} + e\phi_{\mathrm{bg}}^{\mathrm{imp}}\right) \tag{A13}$$

which in the absence of impurities reduces to the expression given in Sec. II A. The solutions of Eqs. (A9)-(A10) for a particular device are shown in Fig. 12.

Appendix B: Interlayer hopping Hamiltonian between non-commensurate layers

We describe the graphene-boron nitride coupling using the general theory of coupling between non-commensurate layers of Refs. 22 and 37. We wish to describe the coupling between two 2D crystals, labeled as ℓ and ℓ', with Bravais lattices spanned by $\{\boldsymbol{a}_{1,\ell}, \boldsymbol{a}_{2,\ell}\}$ and $\{\boldsymbol{a}_{1,\ell'}, \boldsymbol{a}_{2,\ell'}\}$, respectively. In a tight-binding representation the interlayer hopping between layers ℓ and ℓ' can be written as

$$T_{\ell,\ell'} = -\sum_{n,a,m,b} t\left(\mathbf{R}_{n,a,\ell}, \mathbf{R}_{m,b,\ell'}\right) c_{n,a,\ell}^\dagger c_{m,b,\ell'}, \tag{B1}$$

where the indices n, m run over Bravais lattice sites, a, b run over orbitals/sublattice sites, $c_{n,\alpha,\ell}^\dagger$ creates an electron state in layer ℓ at position $\mathbf{R}_{n,a,\ell} = n_1 \boldsymbol{a}_{1,\ell} + n_2 \boldsymbol{a}_{2,\ell} + \boldsymbol{\tau}_{a,\ell}$, and orbital/sublattice a, with $\boldsymbol{\tau}_{a,\ell}$ a sublattice vector, and $t\left(\mathbf{R}_{n,a,\ell}, \mathbf{R}_{m,b,\ell'}\right)$ are hopping terms. Assuming that the hopping $t\left(\mathbf{R}_{n,a,\ell}, \mathbf{R}_{m,b,\ell'}\right)$ only depends on

$\mathbf{R}_{n,a,\ell} - \mathbf{R}_{m,b,\ell'}$ it is possible to write it in Fourier components as[37]

$$t\left(\mathbf{R}_{n,a,\ell}, \mathbf{R}_{m,b,\ell'}\right) = \sqrt{A_{\mathrm{cell},\ell} A_{\mathrm{cell},\ell'}} \times$$
$$\times \int \frac{d^2\boldsymbol{q}}{(2\pi)^2} t_{a,b}^{\ell,\ell'}(\boldsymbol{q})\, e^{i\boldsymbol{q}\cdot\left(\mathbf{R}_{n,a,\ell} - \mathbf{R}_{m,b,\ell'}\right)},$$

where $A_{\mathrm{cell},\ell/\ell'}$ is the area of the unit cell of layer ℓ/ℓ'. If we express, $c_{n,a,\ell}^\dagger$ and $c_{m,b,\ell'}$ in a Bloch basis

$$c_{n,a,\ell}^\dagger = \frac{1}{\sqrt{N_\ell}} \sum_{\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot\mathbf{R}_{n,a,\ell}} c_{\boldsymbol{k},a,\ell}^\dagger, \tag{B2}$$

$$c_{m,b,\ell'}^\dagger = \frac{1}{\sqrt{N_{\ell'}}} \sum_{\boldsymbol{k}} e^{-i\boldsymbol{k}\cdot\mathbf{R}_{m,b,\ell'}} c_{\boldsymbol{k},b,\ell'}^\dagger, \tag{B3}$$

where $N_{\ell/\ell'}$ is the number of unit cells in layer ℓ/ℓ', such that $N_\ell A_{\mathrm{cell},\ell} = N_{\ell'} A_{\mathrm{cell},\ell'}$, the interlayer Hamiltonian becomes

$$T_{\ell,\ell'} = -\sum_{\substack{\boldsymbol{k},\mathbf{G}_{n,\ell} \\ \boldsymbol{k}',\mathbf{G}_{m,\ell'}}} e^{i\boldsymbol{\tau}_{a,\ell}\cdot\mathbf{G}_{n,\ell}} t_{a,b}^{\ell,\ell'}(\boldsymbol{k}+\mathbf{G}_{n,\ell}) \times$$
$$\times e^{-i\boldsymbol{\tau}_{b,\ell'}\cdot\mathbf{G}_{m,\ell'}} c_{\boldsymbol{k},a,\ell}^\dagger c_{\boldsymbol{k}',b,\ell'} \delta_{\boldsymbol{k}+\mathbf{G}_{n,\ell},\boldsymbol{k}'+\mathbf{G}_{m,\ell'}} \tag{B4}$$

where $\mathbf{G}_{n,\ell/\ell'}$ are reciprocal lattice vectors of the 2D crystal ℓ/ℓ'. The Kronecker-δ imposes that in a interlayer hoping process, momentum is conserved modulo any combination of reciprocal lattice vectors of both layers. In general, $t_{a,b}^{\ell,\ell'}(\boldsymbol{q})$ will decay for large values of $|\boldsymbol{q}|$, and therefore only the processes with smallest $|\boldsymbol{k}+\mathbf{G}_{n,\ell}|$ need be considered.

We now specialize to the case where ℓ' is a graphene layer and ℓ is a boron nitride layer. The graphene unit cell contains two carbon atoms in the unit cell, A and B, while boron nitride contains one boron atom, B, and one nitrogen atom, N, in the unit cell, see Fig. 1. We will focus on low energy states, which lie close to the Dirac points, $\pm \boldsymbol{K}_{\mathrm{g}}$, of the graphene layer. Considering only the three most relevant processes coupling the graphene and boron nitride layers, we must consider processes involving $\mathbf{G}_{n,\mathrm{g}} = \mathbf{0}, \boldsymbol{b}_{2,\mathrm{g}}, -\boldsymbol{b}_{1,g}$ and $\mathbf{G}_{n,\mathrm{hBN}} = \mathbf{0}, \boldsymbol{b}_{2,\mathrm{hBN}}, -\boldsymbol{b}_{1,\mathrm{hBN}}$ for states close to the $\boldsymbol{K}_{\mathrm{g}}$ point and processes involving $\mathbf{G}_{n,\mathrm{g}} = \mathbf{0}, -\boldsymbol{b}_{2,\mathrm{g}}, \boldsymbol{b}_{1,\mathrm{g}}$ and $\mathbf{G}_{n,\mathrm{hBN}} = \mathbf{0}, -\boldsymbol{b}_{2,\mathrm{hBN}}, \boldsymbol{b}_{1,\mathrm{hBN}}$ for states close to the $-\boldsymbol{K}_{\mathrm{g}}$ point. It is also assumed that the momentum dependence of $t_{a,b}^{\ell,\ell'}(\boldsymbol{k})$ is weak such that we can approximate $t_{a,b}^{\ell,\ell'}(\boldsymbol{k}+\boldsymbol{K}_\pm+\mathbf{G}_{n,\mathrm{g}}) \simeq t_{a,b}^{\ell,\ell'}(\boldsymbol{K})$, setting : $t_{\mathrm{B,A}}^{\mathrm{hBN,g}}(\boldsymbol{K}) = t_{\mathrm{B,B}}^{\mathrm{hBN,g}}(\boldsymbol{K}) = t_{\mathrm{B,C}}$ and $t_{\mathrm{N,A}}^{\mathrm{hBN,g}}(\boldsymbol{K}) = t_{\mathrm{N,B}}^{\mathrm{hBN,g}}(\boldsymbol{K}) = t_{\mathrm{N,C}}$.

In order to describe the coupling between the bottom and top graphene layers to a slab formed by $\mathcal{N}$ hBN monolayers, we notice that the products of unit cell basis vectors and reciprocal lattice vectors that appears in Eq. (B4) can be written for the bottom graphene layer as $\boldsymbol{\tau}_{\mathrm{hBN,B1}} \cdot \mathbf{G}_{n,\mathrm{hBN}} = \boldsymbol{\tau}_{\mathrm{bg,A}} \cdot \mathbf{G}_{n,\mathrm{bg}} = 0$ and $\boldsymbol{\tau}_{\mathrm{hBN,N1}} \cdot \mathbf{G}_{n,\mathrm{hBN}} = \boldsymbol{\tau}_{\mathrm{bg,B}} \cdot \mathbf{G}_{n,\mathrm{bg}} = \pm n2\pi/3$ (for states close to $\pm \boldsymbol{K}_{\mathrm{g}}$ point). For the coupling between the top

graphene layer and the $\mathcal{N}^{\text{th}}$ hBN layer, one must consider separately the cases when the hBN slab is formed by and even or odd number of layers. For an odd number of layers, in the $\mathcal{N}$th layer the boron and nitrogen atoms occupy the same positions as in the 1st layer and therefore we still have $\boldsymbol{\tau}_{\text{hBN,B}\mathcal{N}} \cdot \mathbf{G}_{n,\text{hBN}} = \boldsymbol{\tau}_{\text{tg,A}} \cdot \mathbf{G}_{n,\text{tg}} = 0$ and $\boldsymbol{\tau}_{\text{hBN,N}\mathcal{N}} \cdot \mathbf{G}_{n,\text{hBN}} = \boldsymbol{\tau}_{\text{tg,B}} \cdot \mathbf{G}_{n,\text{tg}} = \pm n2\pi/3$. If we have an even number of hBN layers, then in the $\mathcal{N}^{\text{th}}$ layer, the boron an nitrogen atoms switch positions compared to the 1st layer, and one obtains instead $\boldsymbol{\tau}_{\text{hBN,N}\mathcal{N}} \cdot \mathbf{G}_{n,\text{hBN}} = \boldsymbol{\tau}_{\text{tg,A}} \cdot \mathbf{G}_{n,\text{tg}} = 0$ and $\boldsymbol{\tau}_{\text{hBN,B}\mathcal{N}} \cdot \mathbf{G}_{n,\text{hBN}} = \boldsymbol{\tau}_{\text{tg,B}} \cdot \mathbf{G}_{n,\text{tg}} = \pm n2\pi/3 \mp n2\pi/3$. With these approximations, one obtains Eq. (5) of the main text.

Appendix C: Analytic expression for the tunneling density of states

In this appendix we provide an analytic expression for Eq. (40). First, we notice that Eq. (40) can be written in a the graphene sublattice basis as

$$\text{TDoS}_{n,m}(\omega_{\text{bg}}, \omega_{\text{tg}}) =$$
$$= \int \frac{d^2\boldsymbol{k}}{(2\pi)^2} \text{tr} \left[\boldsymbol{R}^n_{-\frac{2\pi}{3}} \cdot \boldsymbol{A}_{\text{bg},\boldsymbol{k}}(\omega_{\text{bg}}) \cdot \boldsymbol{R}^n_{\frac{2\pi}{3}} \cdot \right.$$
$$\left. \cdot \boldsymbol{J} \cdot \boldsymbol{R}^m_{-\frac{2\pi}{3}} \cdot \boldsymbol{A}_{\text{tg},\boldsymbol{k}+\boldsymbol{\mathcal{Q}}_{n,m},T}(\omega_{\text{tg}}) \cdot \boldsymbol{R}^m_{\frac{2\pi}{3}} \cdot \boldsymbol{J} \right], \quad \text{(C1)}$$

where tr $\{...\}$ is the trace over graphene sublattice indices, $\boldsymbol{J}$ is a 2×2 matrix of ones, and we have written the spectral function in the sublattice basis as

$$\boldsymbol{A}_{\text{bg/tg},\boldsymbol{k}}(\omega) = i \left[\boldsymbol{G}_{\boldsymbol{k},\theta_{\text{bg/tg}}} \left(\omega^+_{\text{bg/tg}} \right) - \boldsymbol{G}_{\boldsymbol{k},\theta_{\text{bg/tg}}} \left(\omega^-_{\text{bg/tg}} \right) \right],$$
$$\text{(C2)}$$

where the graphene retarded/advanced electron Green´s function in the sublattice space is given by

$$\boldsymbol{G}_{\boldsymbol{k},\theta} \left(\omega^\pm_{\text{bg/tg}} \right) = \frac{\omega^\pm_{\text{bg/tg}} \text{Id} + v_F \hbar \boldsymbol{k} \cdot \boldsymbol{\sigma}_\theta}{\left(\omega^\pm_{\text{bg/tg}} \right)^2 - (v_F \hbar)^2 |\boldsymbol{k}|^2}, \quad \text{(C3)}$$

with $\boldsymbol{\sigma}_\theta = (\cos\theta\sigma_x - \sin\theta\sigma_y, \sin\theta\sigma_x + \cos\theta\sigma_y)$. In the limit of an infinite electron lifetime, we have $\omega^\pm_{\text{bg/tg}} = \omega_{\text{bg/tg}} \pm i0^+$. In the presence of perturbations that induce a momentum independent self-energy that is diagonal in the sublattice basis (such as short range diagonal disorder or scattering by in-plane optical phonons), we make the replacement $\omega^\pm_{\text{bg/tg}} \to \omega_{\text{bg/tg}} - \text{Re}\Sigma_{\text{bg/tg}}(\omega_{\text{bg/tg}}) \pm i\gamma_{\text{bg/tg}}(\omega_{\text{bg/tg}})$, where $\gamma_{\text{bg/tg}}(\omega_{\text{bg/tg}}) = -\text{Im}\Sigma^R_{\text{bg/tg}}(\omega_{\text{bg/tg}})$ is the broadening factor. In the presence of the external metallic contacts and disorder/phonon scattering, we obtain $\gamma_{\text{bg/tg}} = (\Gamma_{\text{b/t}} + \Gamma_{\text{e-ph,bg/tg}})/2$. In terms of Green's functions, and noticing that the matrices $\boldsymbol{R}^n_{\pm\frac{2\pi}{3}}$ perform a rotation of the electronic Green's functions, $\text{TDoS}_{n,m}(\omega)$ can be written as

$$\text{TDoS}_{n,m}(\omega_{\text{bg}}, \omega_{\text{tg}}) = i^2 \sum_{s,s'=\pm1} \int \frac{d^2\boldsymbol{k}}{(2\pi)^2} ss' \times$$
$$\times \text{tr} \left[\boldsymbol{G}_{\boldsymbol{k},\theta_{\text{bg}}+n\frac{2\pi}{3}} \left(\omega^s_{\text{bg}} \right) \cdot \boldsymbol{J} \cdot \right.$$
$$\left. \cdot \boldsymbol{G}_{\boldsymbol{k}+\boldsymbol{\mathcal{Q}}_{n,m},\theta_{\text{tg}}+m\frac{2\pi}{3}} \left(\omega^{s'}_{\text{tg}} \right) \cdot \boldsymbol{J} \right] \quad \text{(C4)}$$

Performing the trace over the sublattice degrees of freedom we get

$$\text{TDoS}_{n,m}(\omega_{\text{bg}}, \omega_{\text{tg}}) = i^2 \sum_{s,s'=\pm1} \int \frac{d^2\boldsymbol{k}}{(2\pi)^2} ss' \times$$
$$\times \frac{2 \left(\omega^s_{\text{bg}} + v_F \hbar \boldsymbol{k} \cdot \hat{\boldsymbol{K}}_{\text{bg},n} \right)}{\left(\omega^s_{\text{bg}} \right)^2 - (v_F \hbar)^2 |\boldsymbol{k}|^2}$$
$$\times \frac{2 \left(\omega^{s'}_{\text{tg}} + v_F \hbar (\boldsymbol{k} + \boldsymbol{\mathcal{Q}}_{n,m}) \cdot \hat{\boldsymbol{K}}_{\text{tg},m} \right)}{\left(\omega^{s'}_{\text{tg}} \right)^2 - (v_F \hbar)^2 |\boldsymbol{k} + \boldsymbol{\mathcal{Q}}_{n,m}|^2} \quad \text{(C5)}$$

The advantage of this form, with respect to Eq. (40), is that Eq. (C5) is analytic in $\boldsymbol{k}$ and as such, contour integration methods can be used to compute the integrals. In order to make analytic progress, in the first term of the previous expression we take the limit $\gamma_{\text{bg}} \to 0$, such that $\omega^{s'}_{\text{bg}} \to \omega_{\text{bg}} = \omega + \epsilon_{\text{F,bg}}$ and

$$i \sum_{s=\pm1} s \frac{\omega^s_{\text{bg}} + v_F \hbar \boldsymbol{k} \cdot \hat{\boldsymbol{K}}_{\text{bg},n}}{\left(\omega^s_{\text{bg}} \right)^2 - (v_F \hbar)^2 |\boldsymbol{k}|^2} \to$$
$$\to 2\pi \frac{\omega_{\text{bg}} + v_F \hbar \boldsymbol{k} \cdot \hat{\boldsymbol{K}}_{\text{bg},n}}{2v_F \hbar |\boldsymbol{k}|} \times$$
$$\times \sum_{s=\pm1} s\delta \left(\omega_{\text{bg}} - sv_F\hbar|\boldsymbol{k}| \right) \quad \text{(C6)}$$

We use the δ-function to perform the integration over $|\boldsymbol{k}|$, obtaining

$$\text{TDoS}_{n,m}(\omega_{\text{bg}}, \omega_{\text{tg}}) \simeq$$
$$\simeq i \frac{\omega_{\text{bg}}}{(v_F\hbar)^2} \int \frac{d\theta_{\boldsymbol{k}}}{2\pi} \left(\frac{\omega_{\text{bg}} + v_F\hbar\boldsymbol{k} \cdot \hat{\boldsymbol{K}}_{\text{bg},n}}{v_F\hbar|\boldsymbol{k}|} \right) \Bigg|_{|\boldsymbol{k}|=\frac{|\omega_{\text{bg}}|}{v_F\hbar}} \times$$
$$\times \sum_{s'=\pm1} s' \frac{2 \left(\omega^{s'}_{\text{tg}} + v_F\hbar(\boldsymbol{k} + \boldsymbol{\mathcal{Q}}_{n,m}) \cdot \hat{\boldsymbol{K}}_{\text{tg},m} \right)}{\left(\omega^{s'}_{\text{tg}} \right)^2 - (v_F\hbar)^2 |\boldsymbol{k} + \boldsymbol{\mathcal{Q}}_{n,m}|^2}. \quad \text{(C7)}$$

The remaining integration over the angular variable $\theta_{\boldsymbol{k}}$ can be performed using contour integration methods. Performing a change of variables $z = e^{i\theta_{\boldsymbol{k}}}$ such that

$$\cos\theta_{\boldsymbol{k}} = \frac{z + z^{-1}}{2}, \quad \text{(C8)}$$
$$\sin\theta_{\boldsymbol{k}} = \frac{z - z^{-1}}{2i}, \quad \text{(C9)}$$

Eq. (C7) can be written as an integral over the z variable around the unit circle in the complex plane

$$\text{TDoS}_{n,m}(\omega_{\text{bg}},\omega_{\text{tg}}) \simeq i\frac{\omega_{\text{bg}}}{(v_F\hbar)^2}\oint_{|z|=1}\frac{dz}{2\pi i}\frac{1}{z}\left(\frac{\omega_{\text{bg}}+v_F\hbar\,|\boldsymbol{k}|\left(\frac{z+z^{-1}}{2}\hat{K}^x_{\text{bg},n}+\frac{z-z^{-1}}{2i}\hat{K}^y_{\text{bg},n}\right)}{v_F\hbar\,|\boldsymbol{k}|}\right)\Bigg|_{|\boldsymbol{k}|=\frac{|\omega_{\text{bg}}|}{v_F\hbar}}\times$$

$$\times\sum_{s'=\pm1}s'\frac{2\left(\omega^{s'}_{\text{tg}}+v_F\hbar\,|\boldsymbol{k}|\left(\frac{z+z^{-1}}{2}\hat{K}^x_{\text{tg},m}+\frac{z-z^{-1}}{2i}\hat{K}^y_{\text{tg},m}\right)+v_F\hbar\boldsymbol{\mathcal{Q}}_{n,m}\cdot\hat{\boldsymbol{K}}_{\text{bg},m}\right)}{\left(\omega^{s'}_{\text{tg}}\right)^2-|\boldsymbol{k}|^2-|\boldsymbol{\mathcal{Q}}_{n,m}|^2-2v_F\hbar\,|\boldsymbol{k}|\,|\boldsymbol{\mathcal{Q}}_{n,m}|\left(\frac{z+z^{-1}}{2}\cos\theta_{\boldsymbol{\mathcal{Q}}_{n,m}}+\frac{z-z^{-1}}{2i}\sin\theta_{\boldsymbol{\mathcal{Q}}_{n,m}}\right)}\,,\quad\text{(C10)}$$

with $\theta_{\boldsymbol{\mathcal{Q}}_{n,m}}$ the angle of the vector $\boldsymbol{\mathcal{Q}}_{n,m}$ with the reference x axis. The integrand has a double pole at $z=0$ and two simple poles at $z=e^{i\theta_{\boldsymbol{\mathcal{Q}}_{m,n}}}w^s_{\gtrless,\text{tg}}$, with

$$w^s_{\gtrless,\text{tg}}=\mathcal{C}^s_{\text{tg}}\mp i\mathcal{S}^s_{\text{tg}},\tag{C11}$$

$$\mathcal{C}^s_{\text{tg}}=\frac{(\omega_{\text{tg}}+si\gamma_{\text{tg}})^2-(v_F\hbar)^2\left(|\boldsymbol{\mathcal{Q}}_{n,m}|^2+|\boldsymbol{k}|^2\right)}{2(v_F\hbar)^2\,|\boldsymbol{\mathcal{Q}}_{n,m}|\,|\boldsymbol{k}|}\,,\tag{C12}$$

$$\mathcal{S}^s_{\text{tg}}=\text{sgn}\left(\omega^2_{\text{tg}}-\gamma^2_{\text{tg}}-(v_F\hbar)^2\left(|\boldsymbol{\mathcal{Q}}_{n,m}|^2+|\boldsymbol{k}|^2\right)\right)\times$$
$$\times i\sqrt{\left(\mathcal{C}^s_{\text{tg}}\right)^2-1},\tag{C13}$$

defined such that $|w_{<,\text{tg}}|<1$ and $w_{>,\text{tg}}=w^{-1}_{<,\text{tg}}$. The contour integration around the unit circle can be performed analytically collecting the residues at $z=e^{i\theta_{\pm\boldsymbol{\mathcal{Q}}_{n,m}}}w^s_{<,\text{tg}}$ and $z=0$. Notice that we have made the approximation $\gamma_{\text{bg}}\to0$. In general, both γ_{bg} and γ_{tg} will be non-zero. The simplest way to that this into account is to symmetrize Eq. (C5) with respect to the bottom and the top graphene layer an then taking the limit $\gamma_{\text{bg}}\to0$ in the first term and $\gamma_{\text{tg}}\to0$ in the second. The final symmetrized result is given by

$$\text{TDoS}_{n,m}(\omega_{\text{bg}},\omega_{\text{tg}})\simeq\frac{\omega_{\text{bg}}}{(v_F\hbar)^3\,|\boldsymbol{\mathcal{Q}}_{n,m}|}\times$$

$$\times\Bigg[\frac{-1}{\mathcal{S}^+_{\text{tg}}}\left(\frac{\omega^+_{\text{tg}}+|\omega_{\text{bg}}|\left(\mathcal{C}^+_{\text{tg}}X^{\text{tg}}_{n,m}+\mathcal{S}^+_{\text{tg}}Y^{\text{tg}}_{n,m}\right)+v_F\hbar\boldsymbol{\mathcal{Q}}_{n,m}\cdot\hat{\boldsymbol{K}}_{\text{tg},m}}{|\omega_{\text{bg}}|}\right)\left(\frac{\omega_{\text{bg}}+|\omega_{\text{bg}}|\left(\mathcal{C}^+_{\text{tg}}X^{\text{bg}}_{n,m}+\mathcal{S}^+_{\text{tg}}Y^{\text{bg}}_{n,m}\right)}{|\omega_{\text{bg}}|}\right)$$

$$+\frac{1}{\mathcal{S}^-_{\text{tg}}}\left(\frac{\omega^-_{\text{tg}}+|\omega_{\text{bg}}|\left(\mathcal{C}^-_{\text{tg}}X^{\text{tg}}_{n,m}+\mathcal{S}^-_{\text{tg}}Y^{\text{tg}}_{n,m}\right)+v_F\hbar\boldsymbol{\mathcal{Q}}_{n,m}\cdot\hat{\boldsymbol{K}}_{\text{tg},m}}{|\omega_{\text{bg}}|}\right)\left(\frac{\omega_{\text{bg}}+|\omega_{\text{bg}}|\left(\mathcal{C}^-_{\text{tg}}X^{\text{bg}}_{n,m}+\mathcal{S}^-_{\text{tg}}Y^{\text{bg}}_{n,m}\right)}{|\omega_{\text{bg}}|}\right)$$

$$+\frac{2\gamma_{\text{tg}}\left(X^{\text{bg}}_{n,m}+iY^{\text{bg}}_{n,m}\right)}{v_F\hbar\,|\boldsymbol{\mathcal{Q}}_{n,m}|}\left(\frac{v_F\hbar\,|\boldsymbol{\mathcal{Q}}_{n,m}|+\omega_{\text{tg}}\left(X^{\text{tg}}_{n,m}+iY^{\text{tg}}_{n,m}\right)}{|\omega_{\text{bg}}|}\right)\Bigg]\Bigg|_{|\boldsymbol{k}|=|\omega_{\text{bg}}|/(v_F\hbar)}$$

$$+\frac{1}{2}\frac{\omega_{\text{tg}}}{(v_F\hbar)^3\,|\boldsymbol{\mathcal{Q}}_{n,m}|}\times$$

$$\times\Bigg[\frac{-1}{\mathcal{S}^+_{\text{bg}}}\left(\frac{\omega^+_{\text{bg}}-|\omega_{\text{tg}}|\left(\mathcal{C}^+_{\text{bg}}X^{\text{bg}}_{n,m}+\mathcal{S}^+_{\text{bg}}Y^{\text{bg}}_{n,m}\right)-v_F\hbar\boldsymbol{\mathcal{Q}}_{n,m}\cdot\hat{\boldsymbol{K}}_{\text{bg},n}}{|\omega_{\text{tg}}|}\right)\left(\frac{\omega_{\text{tg}}-|\omega_{\text{tg}}|\left(\mathcal{C}^+_{\text{bg}}X^{\text{tg}}_{n,m}+\mathcal{S}^+_{\text{bg}}Y^{\text{tg}}_{n,m}\right)}{|\omega_{\text{tg}}|}\right)$$

$$+\frac{1}{\mathcal{S}^-_{\text{bg}}}\left(\frac{\omega^-_{\text{bg}}-|\omega_{\text{tg}}|\left(\mathcal{C}^-_{\text{bg}}X^{\text{bg}}_{n,m}+\mathcal{S}^-_{\text{bg}}Y^{\text{bg}}_{n,m}\right)-v_F\hbar\boldsymbol{\mathcal{Q}}_{n,m}\cdot\hat{\boldsymbol{K}}_{\text{bg},n}}{|\omega_{\text{tg}}|}\right)\left(\frac{\omega_{\text{tg}}-|\omega_{\text{tg}}|\left(\mathcal{C}^-_{\text{bg}}X^{\text{tg}}_{n,m}+\mathcal{S}^-_{\text{bg}}Y^{\text{tg}}_{n,m}\right)}{|\omega_{\text{tg}}|}\right)$$

$$+\frac{2\gamma_{\text{bg}}\left(-X^{\text{tg}}_{n,m}-iY^{\text{tg}}_{n,m}\right)}{v_F\hbar\,|\boldsymbol{\mathcal{Q}}_{n,m}|}\left(\frac{v_F\hbar\,|\boldsymbol{\mathcal{Q}}_{n,m}|-\omega_{\text{bg}}\left(X^{\text{bg}}_{n,m}+iY^{\text{bg}}_{n,m}\right)}{|\omega_{\text{tg}}|}\right)\Bigg]\Bigg|_{|\boldsymbol{k}|=|\omega_{\text{tg}}|/(v_F\hbar)}\,,\quad\text{(C14)}$$

where we have introduced the quantities

$$\begin{aligned}X^{\text{bg}}_{n,m}&=\hat{\boldsymbol{\mathcal{Q}}}_{n,m}\cdot\hat{\boldsymbol{K}}_{\text{bg},n}, & Y^{\text{bg}}_{n,m}&=\hat{\boldsymbol{\mathcal{Q}}}_{n,m}\times\hat{\boldsymbol{K}}_{\text{bg},n},\\ X^{\text{tg}}_{n,m}&=\hat{\boldsymbol{\mathcal{Q}}}_{n,m}\cdot\hat{\boldsymbol{K}}_{\text{tg},m}, & Y^{\text{tg}}_{n,m}&=\hat{\boldsymbol{\mathcal{Q}}}_{n,m}\times\hat{\boldsymbol{K}}_{\text{tg},m},\end{aligned}\tag{C15}$$

and the quantities $\mathcal{C}^\pm_{\text{tg}}$ and $\mathcal{S}^\pm_{\text{tg}}$ given by Eqs. (C12) and (C13) with the replacements $\omega_{\text{tg}}\to\omega_{\text{bg}}$ and $\gamma_{\text{tg}}\to\gamma_{\text{bg}}$. It

was checked that Eq. (C14) provides a very good approximation to the numeric evaluation of Eq. (C5) when both γ_{bg} and γ_{tg} are non-zero, if the broadening function for each layer is assumed to the the sum of the broadening factors of both layers, i.e., performing the replacement $\gamma_{\text{bg}}, \gamma_{\text{tg}} \to \gamma_{\text{bg}} + \gamma_{\text{tg}}$.

In the limit of infinite electron lifetime in both layers $\gamma_{\text{bg/tg}} \to 0$, we obtain

$$\mathcal{C}_{\text{tg}}^s = \frac{\omega_{\text{tg}}^2 - \omega_{\text{bg}}^2 - (v_F\hbar)^2 |\boldsymbol{\mathcal{Q}}_{n,m}|^2}{2\,(v_F\hbar)\,|\boldsymbol{\mathcal{Q}}_{n,m}|\,|\omega_{\text{bg}}|}, \qquad \text{(C16)}$$

$$\mathcal{S}_{\text{tg}}^s = -s\,\text{sgn}(\omega_{\text{tg}})\sqrt{1 - \left(\mathcal{C}_{\text{tg}}^s\right)^2}, \qquad \text{(C17)}$$

$$\text{TDoS}_{n,m}(\omega_{\text{bg}}, \omega_{\text{tg}}) = \frac{\omega_{\text{tg}}}{(v_F\hbar)^3 |\boldsymbol{\mathcal{Q}}_{n,m}|} \times$$
$$\times \left[\frac{-1}{\mathcal{S}_{\text{tg}}^+} \left(\frac{\omega_{\text{tg}} + |\omega_{\text{bg}}|\left(\mathcal{C}_{\text{tg}}^+ X_{n,m}^{\text{tg}} + \mathcal{S}_{\text{tg}}^+ Y_{n,m}^{\text{tg}}\right) + v_F\hbar\boldsymbol{\mathcal{Q}}_{n,m}\cdot\hat{\boldsymbol{K}}_{\text{tg},m}}{|\omega_{\text{bg}}|} \right) \left(\frac{\omega_{\text{bg}} + |\omega_{\text{bg}}|\left(\mathcal{C}_{\text{tg}}^+ X_{n,m}^{\text{bg}} + \mathcal{S}_{\text{tg}}^+ Y_{n,m}^{\text{bg}}\right)}{|\omega_{\text{bg}}|} \right) \right.$$
$$\left. + \frac{1}{\mathcal{S}_{\text{tg}}^-} \left(\frac{\omega_{\text{tg}} + |\omega_{\text{bg}}|\left(\mathcal{C}_{\text{tg}}^- X_{n,m}^{\text{tg}} + \mathcal{S}_{\text{tg}}^- Y_{n,m}^{\text{tg}}\right) + v_F\hbar\boldsymbol{\mathcal{Q}}_{n,m}\cdot\hat{\boldsymbol{K}}_{\text{tg},m}}{|\omega_{\text{bg}}|} \right) \left(\frac{\omega_{\text{bg}} + |\omega_{\text{bg}}|\left(\mathcal{C}_{\text{tg}}^- X_{n,m}^{\text{bg}} + \mathcal{S}_{\text{tg}}^- Y_{n,m}^{\text{bg}}\right)}{|\omega_{\text{bg}}|} \right) \right]. \quad \text{(C18)}$$

We notice that, in this limit, $\text{TDoS}_{n,m}(\omega_{\text{bg}}, \omega_{\text{tg}})$ is only non-zero when $4\,(v_F\hbar)^2 |\boldsymbol{\mathcal{Q}}_{n,m}|^2 \omega_{\text{bg}}^2 > \left(\omega_{\text{tg}}^2 - \omega_{\text{bg}}^2 - (v_F\hbar)^2 |\boldsymbol{\mathcal{Q}}_{n,m}|^2\right)^2$.

We finally study how the spinorial character of graphene's wavefunction manifests in the form of $\text{TDoS}_{n,m}(\omega_{\text{bg}}, \omega_{\text{tg}})$. If we set the wavefunction overlap factors $\Upsilon_{\boldsymbol{k},\lambda}^{\text{bg/tg},n}$ to 1 in Eq. (40), then instead of Eq. (C5) we would obtain

$$\text{TDoS}_{n,m}^{\text{scalar}}(\omega_{\text{bg}}, \omega_{\text{tg}}) = i^2 \sum_{s,s'=\pm 1} \int \frac{d^2\boldsymbol{k}}{(2\pi)^2} ss' \times$$
$$\times \frac{2\omega_{\text{bg}}^s}{\left(\omega_{\text{bg}}^s\right)^2 - (v_F\hbar)^2 |\boldsymbol{k}|^2}$$
$$\times \frac{2\omega_{\text{tg}}^{s'}}{\left(\omega_{\text{tg}}^{s'}\right)^2 - (v_F\hbar)^2 |\boldsymbol{k} + \boldsymbol{\mathcal{Q}}_{n,m}|^2}. \quad \text{(C19)}$$

In order to evaluate $\text{TDoS}_{n,m}^{\text{scalar}}(\omega_{\text{bg}}, \omega_{\text{tg}})$, we proceed as previously. the only difference is that when performing the integration over the unit circle in the complex variable z, there is no double pole at $z = 0$, and the contour integration only collects the contribution from $z = e^{i\theta \pm \boldsymbol{\mathcal{Q}}_{m,n}} w_{<,\text{tg/bg}}^s$. Symmetrizing the result, this

and $\mathcal{S}_{\text{bg}}^s/\mathcal{C}_{\text{bg}}^s$ are obtained by replacing $\omega_{\text{bg}} \leftrightarrow \omega_{\text{tg}}$. and $\text{TDoS}_{n,m}(\omega_{\text{bg}}, \omega_{\text{tg}})$ simplifies to

leads to

$$\text{TDoS}_{n,m}^{\text{scalar}}(\omega_{\text{bg}}, \omega_{\text{tg}}) = \frac{1}{(v_F\hbar)^3 |\boldsymbol{\mathcal{Q}}_{n,m}|} \times$$
$$\times \frac{1}{2} \left[\left(\frac{\omega_{\text{tg}}^-}{\mathcal{S}_{\text{tg}}^-} - \frac{\omega_{\text{tg}}^+}{\mathcal{S}_{\text{tg}}^+} \right) + \left(\frac{\omega_{\text{bg}}^-}{\mathcal{S}_{\text{bg}}^-} - \frac{\omega_{\text{bg}}^+}{\mathcal{S}_{\text{bg}}^+} \right) \right]. \quad \text{(C20)}$$

Appendix D: Resonant impurities within the SCBA

We consider the effect of resonant impurities, such as vacancies, in the properties of graphene. We focus on this kind of impurities due to the possibility for analytical progress and due to the fact that this model for impurities correctly predicts a transport lifetime in graphene that depends on the Fermi energy as $\tau_{\text{tr}}(\epsilon_F) \propto \epsilon_F$[32]. Resonances due to short range disorder cannot be taken into account by treating then within a Gaussian approximation. A way to overcome this limitation is to employ the T-matrix, which properly takes into account multiple scatterings by the same impurity in the limit of low impurity concentration. Using the T-matrix within the non-crossing approximation, the self-consistent Born approximation (SCBA) for the Green's function of an isolated graphene layer reads

$$\overline{\boldsymbol{G}_{\boldsymbol{k}}^{0,R}}(\omega) = \boldsymbol{G}_{\boldsymbol{k}}^{0,R}(\omega) + \boldsymbol{G}_{\boldsymbol{k}}^{0,R}(\omega)\cdot\boldsymbol{\Sigma}_{\text{imp}}^R(\omega)\cdot\overline{\boldsymbol{G}_{\boldsymbol{k}}^{0,R}}(\omega), \quad \text{(D1)}$$

where matrices the have indices in the sublattice space, a bar denotes disorder averaging and $\boldsymbol{\Sigma}_{\text{imp}}^R(\omega) = $

$n_{\text{imp}}\boldsymbol{T}^R(\omega)$ is the impurity self-energy, where n_{imp} is the impurity concentration (number of impurities by graphene unit cell) and $\boldsymbol{T}^R(\omega)$ is the T-matrix for a single δ-like impurity with strength u. For an impurity potential diagonal in the sublattice basis, the T-matrix is also diagonal with equal components, given by

$$T^R(\omega) = \frac{u}{1 - u\overline{G_1^R}(\omega)} \underset{u \to \infty}{=} -\frac{1}{\overline{G_1^R}(\omega)}, \quad (D2)$$

where we have taken the limit $u \to \infty$ in order to describe vacancies and defined

$$\overline{G_1^R}(\omega) = \int \frac{d^2\boldsymbol{k}}{(2\pi)^2} \left[\overline{G_{\boldsymbol{k}}^{0,R}}(\omega)\right]_A^A \quad (D3)$$

In the Dirac cone approximation, graphene Green's function in the sublattice basis and taking into account a finite electron life (induce by the metallic contact) is given by

$$\left[G_{\boldsymbol{k}}^{0,R}(\omega)\right]_b^a = \frac{1}{\omega - \lambda v_F \hbar |\boldsymbol{k}| + i\gamma_{\text{c}} - \Sigma^R(\omega)} \times$$
$$\times \frac{1}{2}\left[\delta_b^a + \lambda \frac{\boldsymbol{k}}{|\boldsymbol{k}|} \cdot \boldsymbol{\sigma}_b^a\right], \quad (D4)$$

with a, b indices running over the A, B sublattice sites and γ_{c} is the lifetime induced by the metallic contacts, $\gamma_{\text{c}} = \Gamma_{\text{bg/tg}}/2$ (assuming the metallic contacts couple equally to all graphene states and do not spoil translational invariance of graphene). For resonant impurities, the self-energy is momentum independent. Writing it as $\Sigma_{\text{imp}}^R(\omega) = \Sigma_{\text{imp}}(\omega) - i\gamma_{\text{imp}}(\omega)$, we can evaluate $G_1^R(\omega)$ analytically obtaining

$$\overline{G_1^R}(\omega) = \frac{g_1\left(\omega - \Sigma_{\text{imp}}(\omega), \gamma_{\text{imp}}(\omega) + \gamma_{\text{c}}\right)}{4\pi(v_F\hbar)^2}$$
$$- i\frac{g_2\left(\omega - \Sigma_{\text{imp}}(\omega), \gamma_{\text{imp}}(\omega) + \gamma_{\text{c}}\right)}{4\pi(v_F\hbar)^2}, \quad (D5)$$

where the functions g_1 and g_2 are given by

$$g_1(\omega, \eta) = -\frac{\omega}{2}\left[\log\left(\frac{(\Lambda_E - \omega)^2 + \eta^2}{\omega^2 + \eta^2}\right) + (\omega \to -\omega)\right]$$
$$+ \eta\left[\arctan\left(\frac{\Lambda_E - \omega}{\eta}\right) + \arctan\left(\frac{\omega}{\eta}\right) - (\omega \to -\omega)\right]$$
$$(D6)$$

$$g_2(\omega, \eta) = \frac{\eta}{2}\left[\log\left(\frac{(\Lambda_E - \omega)^2 + \eta^2}{\omega^2 + \eta^2}\right) + (\omega \to -\omega)\right]$$
$$+ \omega\left[\arctan\left(\frac{\Lambda_E - \omega}{\eta}\right) + \arctan\left(\frac{\omega}{\eta}\right) - (\omega \to -\omega)\right]$$
$$(D7)$$

with $\Lambda_E \simeq v_F\hbar\left(4\pi/\left(\sqrt{3}a_{\text{g}}^2\right)\right)^{1/2}$ a high energy cutoff. In terms of g_1 and g_2 the self-energy is given by

$$\Sigma_{\text{imp}}(\omega) = -\lambda_{\text{imp}}\frac{g_1(\omega', \gamma')}{g_1^2(\omega', \gamma') + g_2^2(\omega', \gamma')}, \quad (D8)$$

$$\gamma_{\text{imp}}(\omega) = \lambda_{\text{imp}}\frac{g_2(\omega', \gamma')}{g_1^2(\omega', \gamma') + g_2^2(\omega', \gamma')}. \quad (D9)$$
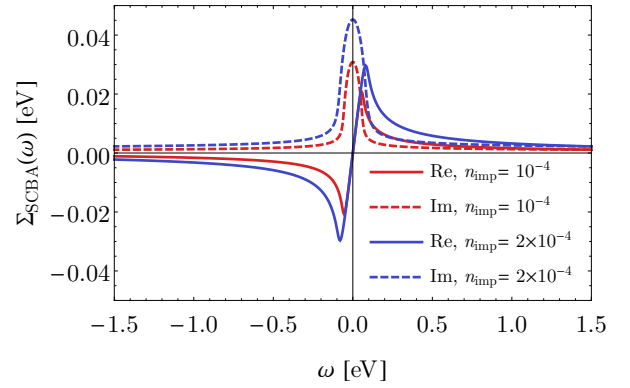

Figure 13. Real and (minus) imaginary parts of the retarded self-energy for graphene electrons due to resonant impurities treated within the SCBA, for two different impurity concentrations (number of impurities per graphene unit cell).

where we have defined $\omega' = \omega - \Sigma_{\text{imp}}(\omega)$, $\gamma' = \gamma_{\text{c}} + \gamma_{\text{imp}}(\omega)$ and $\lambda_{\text{imp}} = 4\pi(v_F\hbar)^2 n_{\text{imp}}$ is a constant characterizing the scattering by resonant disorder. Eqs. (D6)-(D9) form a set of equations that can be easily solved. The solution for self-energy is shown in Fig. 13.

Appendix E: Graphene electron self-energy due to in-plane optical phonons

Electron-phonon interaction in graphene can be modeled by starting from a nearest neighbour tight-binding Hamiltonian for the electrons and assuming that the lattice distortions due to phonons lead to a modulation of hopping integrals[61]. For graphene longitudinal and transverse in-plane phonons close to the Γ point and electrons close to the $\boldsymbol{K}$ point the obtained electron-phonon interaction Hamiltonian is given by

$$H_{\text{g,e-ph}} = \frac{g_{\Gamma\text{O}}^{\text{g}}}{\sqrt{N}} \sum_{\substack{\boldsymbol{k},\boldsymbol{q} \\ \zeta = LO,TO}} \boldsymbol{c}_{\boldsymbol{k}+\boldsymbol{q},\text{g}}^\dagger (\vec{\sigma} \times \vec{\epsilon}_{\boldsymbol{q},\zeta})\, \boldsymbol{c}_{\boldsymbol{k},\text{g}}\phi_{\boldsymbol{q},\zeta}, \quad (E1)$$

wheres

$$g_{\Gamma\text{O}}^{\text{g}} = \frac{3}{2}\left(-\frac{d\log t}{d\log a_{\text{CC}}}\right)\frac{t}{a_{\text{CC}}}\sqrt{\frac{\hbar}{\mu_{\text{g}}\omega_{\Gamma\text{O}}^{\text{g}}}}, \quad (E2)$$

is the electron-phonon coupling constant, with $-d\log t/d\log a_{\text{CC}} \simeq 3$ describing the change in the nearest neighbour hopping, t, with the distance, a_{CC}; $\mu_{\text{g}} = m_{\text{C}}/2$ is the reduced mass of the phonon mode, with m_{C} the carbon atom mass; and $\omega_{\Gamma\text{O}}^{\text{g}}$ is the phonon dispersion for the longitudinal/transverse in-plane optical phonon mode (which are degenerate at Γ and and assume we approximate them as dispersionless). The polarization vectors for the longitudinal and transverse mode can be written as $\vec{\epsilon}_{\boldsymbol{q},LO} = (1,0)$ and $\vec{\epsilon}_{\boldsymbol{q},TO} = (0,1)$. With these approximations we obtain the momentum
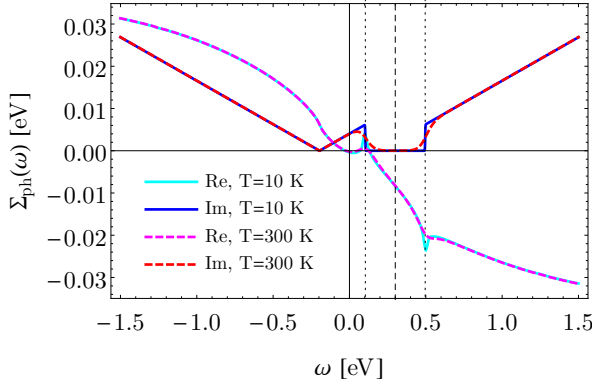
Figure 14. Real and (minus) imaginary parts of the self-energy for graphene electrons due to scattering by in-plane optical phonons for two different temperatures for doped graphene with $\epsilon_F = 0.3$ eV. The zero of energy corresponds to the Dirac point. The dashed vertical line marks $\omega = \epsilon_F$ and the dotted lines mark $\omega = \epsilon_F \pm \omega_{\Gamma O}^g$.

independent electron-phonon interaction matrices

$$M_{LO}^g = -g_{\Gamma O}^g \sigma_y, \quad (E3)$$

$$M_{TO}^g = g_{\Gamma O}^g \sigma_x. \quad (E4)$$

Assuming the graphene layer is in thermal equilibrium and to lowest order in the electron-phonon interaction, the self energy is diagonal in sublattice space and given by

$$\Sigma_{ph}^R(\omega) = A_{cell} \left(g_{\Gamma O}^g\right)^2 \times$$
$$\times \sum_{\lambda,s} \int \frac{d^2\boldsymbol{q}}{(2\pi)^2} s \frac{1 + b(s\omega_{\Gamma O}^g) - f(\epsilon_{\boldsymbol{q},\lambda} - \epsilon_F)}{\omega - \epsilon_{\boldsymbol{q},\lambda} - s\omega_{\Gamma O}^g + i0^+}. \quad (E5)$$

The imaginary part can be computed for pristine graphene at finite temperature as

$$- \mathrm{Im}\Sigma_{ph}^R(\omega) =$$
$$= (g_{\Gamma O}^g)^2 \left[1 + b(\omega_{\Gamma O}^g) - f\left(\omega - \omega_{\Gamma O}^g - \epsilon_F\right)\right] \times$$
$$\times \frac{A_{cell} |\omega - \omega_{\Gamma O}^g|}{2\left(v_F \hbar\right)^2} +$$
$$+ (g_{\Gamma O}^g)^2 \left[b(\omega_{\Gamma O}^g) - f\left(\omega + \omega_{\Gamma O}^g - \epsilon_F\right)\right] \times$$
$$\times \frac{A_{cell} |\omega + \omega_{\Gamma O}^g|}{2\left(v_F \hbar\right)^2}, \quad (E6)$$

where ω and the Fermi energy, ϵ_F, are both measured from the Dirac cone. From this, the real part can be efficiently obtained using the Kramers-Kronig relation

$$\mathrm{Re}\Sigma_{ph}^R(\omega) = -\int \frac{d\nu}{\pi} \frac{\mathrm{Im}\Sigma_{ph}^R(\omega - \nu) - \mathrm{Im}\Sigma_{ph}^R(\omega + \nu)}{\nu}. \quad (E7)$$

The computed self-energy is shown in Fig. 14.

Appendix F: Vertex corrections for resonant impurities

In this Appendix, we provide an alternative derivation of Eq. (63), for the vertical current in a graphene-hBN-graphene device taking into account disorder in the graphene layers, employing approach (B). Instead of describing disorder as an interaction, we will start from Eq. (33) and perform disorder averages of it. Just as in Appendix. D we will consider scattering by resonant disorder. This model will both serve as a concrete example for elastic scattering of the general results present in Sec. IV regarding the equivalences of approaches (A) and (B) and will also show the formal equivalence between the contributions to the current arising from Eq. (28) and vertex corrections. Just as in Sec. III we will assume for simplicity that the external metallic contacts couple to all graphene states and that graphene electronic states are still well describe by Bloch states. With these approximations, we write

$$\boldsymbol{\Gamma}_{b/t} = \Gamma_{b/t} \boldsymbol{I}_{bg/tg}. \quad (F1)$$

Performing an averaging of Eq. (33) with respect to disorder in the bottom and top graphene layers, assuming that these are uncorrelated, and to lowest order in the graphene-hBN coupling we obtain

$$\overline{\mathcal{T}} = \Gamma_b \Gamma_t \mathrm{Tr} \left[\overline{\boldsymbol{G}_{bg}^{0,A} \cdot \boldsymbol{I}_{bg} \cdot \boldsymbol{G}_{bg}^{0,R}} \cdot \boldsymbol{\mathcal{T}}_{bg,tg} \cdot \right. \quad (F2)$$

$$\left. \cdot \overline{\boldsymbol{G}_{tg}^{0,R} \cdot \boldsymbol{I}_{tg} \cdot \boldsymbol{G}_{tg}^{0,A}} \cdot \boldsymbol{\mathcal{T}}_{tg,bg} \right]. \quad (F3)$$

The disorder averaged product of Green's functions is not just the product of average Green's function, as the averaging procedure establishes correlations between the two functions. From now on, we will employing a notation where an upper indice represents an out-going electronic state and a lower indice represents an incoming state, with repeated indices being summed over. With this convention, the average of the product of two Green's functions, in sublattice space, can be written as (suppressing the frequency argument and the bg/tg indice)

$$\overline{\left[G_{\boldsymbol{k}}^{0,A}\right]^a_{b} \delta^b_{c} \left[G_{\boldsymbol{k}}^{R,0}\right]^c_{d}} = \left[\overline{G_{\boldsymbol{k}}^{0,A}}\right]^a_{b} \delta^b_{c} \left[\overline{G_{\boldsymbol{k}}^{R,0}}\right]^c_{d} +$$
$$+ \left[\overline{G_{\boldsymbol{k}}^{0,A}}\right]^a_{a'} \Lambda^{a'c'}_{b'd'}{}'\left[G_2^{AR}\right]^{b'c}_{bc'} \delta^b_{c} \left[\overline{G_{\boldsymbol{k}}^{0,R}}\right]^{d'}_{d}, \quad (F4)$$

where the second term are vertex corrections, we have define the quantity

$$\left[G_2^{AR}\right]^{ac}_{bd} = \int \frac{d^2\boldsymbol{p}}{(2\pi)^2} \left[\overline{G_{\boldsymbol{p}}^A}\right]^a_{b} \left[\overline{G_{\boldsymbol{p}}^R}\right]^c_{d}, \quad (F5)$$

and Λ^{ac}_{bd} is a 4-point function, which obeys a Bethe-Salpeter equation (see Fig. 15)

$$\Lambda^{ac}_{bd} = U^{ac}_{bd} + U^{a'c}_{bd'} \left[G_2^{AR}\right]^{b'd'}_{a'c'} \Lambda^{ac'}_{b'd}, \quad (F6)$$
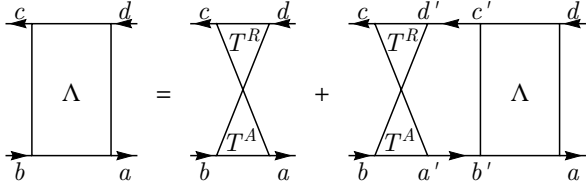
Figure 15. Diagrammatic representation of the Bethe-Salpeter equation, Eq. (F6)

where $U^a{}_b{}^c{}_d$ is an irreducible 4-point function, which within the T-matrix and non-crossing approximation for resonant impurities is given by

$$U^a{}_b{}^c{}_d = n_{\text{imp}} \left| T^R_{\text{imp}} \right|^2 \delta^a{}_b \delta^c{}_d. \qquad (\text{F7})$$

The quantity $\left[G_2^{AR} \right]^a{}_b{}^c{}_d$ can be evaluated analytically yielding

$$\left[G_2^{AR} \right]^a{}_b{}^c{}_d = L_1 \left(\omega', \gamma' \right) \delta^a{}_b \delta^c{}_d + L_2 \left(\omega', \gamma' \right) \frac{1}{2} \boldsymbol{\sigma}^a{}_b \cdot \boldsymbol{\sigma}^c{}_d, \qquad (\text{F8})$$

where

$$L_1 \left(\omega, \eta \right) = \frac{1}{8\pi \left(v_F \hbar \right)^2} \left(\frac{1}{\eta} g_2(\omega, \eta) + \frac{1}{\omega} g_1(\omega, \eta) \right), \quad (\text{F9})$$

$$L_2 \left(\omega, \eta \right) = \frac{1}{8\pi \left(v_F \hbar \right)^2} \left(\frac{1}{\eta} g_2(\omega, \eta) - \frac{1}{\omega} g_1(\omega, \eta) \right) (\text{F10})$$

with the functions g_1 and g_2 defined by Eqs. (D6), (D7) and where we have written $\omega' = \omega - \Sigma_{\text{imp}}$ and $\gamma' = \gamma_{\text{imp}} + \gamma_{\text{c}}$ as in Appendix. D. The Bethe-Salpeter equation for $\Lambda^a{}_b{}^c{}_d$ is now a simple problem of linear algebra. Solving Eq. (F6), yields the non-zero components of $\Lambda^a{}_b{}^c{}_d$ in the sublattice basis

$$\Lambda^A{}_A{}^A{}_A = \Lambda^B{}_B{}^B{}_B = \frac{n_{\text{imp}} \left| T^R \right|^2 \left(1 - L_1 n_{\text{imp}} \left| T^R \right|^2 \right)}{\left[1 - \left(L_1 - L_2 \right) n_{\text{imp}} \left| T^R \right|^2 \right] \left[1 - \left(L_1 + L_2 \right) n_{\text{imp}} \left| T^R \right|^2 \right]}, \qquad (\text{F11})$$

$$\Lambda^A{}_A{}^B{}_B = \Lambda^B{}_B{}^A{}_A = \frac{n_{\text{imp}} \left| T^R \right|^2}{1 - L_1 n_{\text{imp}} \left| T^R \right|^2}, \qquad (\text{F12})$$

$$\Lambda^A{}_B{}^B{}_A = \Lambda^B{}_A{}^A{}_B = \frac{L_2 n_{\text{imp}}^2 \left| T^R \right|^4}{\left[1 - \left(L_1 - L_2 \right) n_{\text{imp}} \left| T^R \right|^2 \right] \left[1 - \left(L_1 + L_2 \right) n_{\text{imp}} \left| T^R \right|^2 \right]}. \qquad (\text{F13})$$

, where we have omitted the frequency arguments of $L_{1/2}$. Using the fact that $\left[G_2^{AR} \right]^{b'}{}_b{}^b{}_{c'} = \left(L_1(\omega) + L_2(\omega) \right) \delta^{b'}{}_{c'}$, the vertex correction contribution in Eq. (F4) can be written as

$$\Lambda^{a'}{}_{b'}{}^{c'}{}_{d'} \left[G_2^{AR} \right]^{b'}{}_b{}^c{}_{c'} =$$

$$= \frac{n_{\text{imp}} \left| T^R \right|^2 \left(L_1 + L_2 \right)}{1 - \left(L_1 + L_2 \right) n_{\text{imp}} \left| T^R \right|^2} \delta^c{}_b \quad (\text{F14})$$

Expressing T^R and $L_{1/2}$ in terms of g_1 and g_2, and using Eqs. (D9) it can be seen that the quantity $\left(L_1 + L_2 \right) n_{\text{imp}} \left| T^R \right|^2$ can be written as the ratio

$$\left(L_1 + L_2 \right) n_{\text{imp}} \left| T^R \right|^2 = \frac{\gamma_{\text{imp}}}{\gamma_{\text{imp}} + \gamma_{\text{c}}}. \qquad (\text{F15})$$

Therefore, Eq. (F4) can be written as

$$\overline{\left[G_{\boldsymbol{k}}^{0,A} \right]^a{}_b \left[G_{\boldsymbol{k}}^{R,0} \right]^b{}_d} = \left[\overline{G_{\boldsymbol{k}}^{0,A}} \right]^a{}_b \left[\overline{G_{\boldsymbol{k}}^{R,0}} \right]^b{}_d$$
$$+ \frac{\gamma_{\text{imp}}}{\gamma_{\text{c}}} \left[\overline{G_{\boldsymbol{k}}^{0,A}} \right]^a{}_b \left[\overline{G_{\boldsymbol{k}}^{R,0}} \right]^b{}_d, \quad (\text{F16})$$

Therefore, the product of a retarded and an advanced Green function is related to the spectral function as

$$\left[\overline{G_{\boldsymbol{k}}^{0,A}}(\omega) \right]^a{}_b \left[\overline{G_{\boldsymbol{k}}^{R,0}}(\omega) \right]^b{}_d = \frac{1}{\gamma_{\text{imp}} + \gamma_{\text{c}}} \left[\overline{A_{\boldsymbol{k}}^0}(\omega) \right]^a{}_b, \qquad (\text{F17})$$

and therefore, the contributions from vertex corrections (incoherent contributions) due to impurities adds to the contribution coming from the product of two average Green's functions (coherent contribution), in such a way that Eq. (F2) reduces to Eq. (34) of the main text. This result is a particular case of the more general discussion of Sec. IV A 1, which is not limited to elastic scattering.

* Electronic address: amorim.bac@gmail.com
[1] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, Science **306**, 666 (2004), http://www.sciencemag.org/content/306/5696/666.full.pdf.
[2] K. S. Novoselov, D. Jiang, F. Schedin, T. J. Booth, V. V. Khotkevich, S. V. Morozov, and A. K. Geim, Proceedings of the National Academy of Sciences of the United States of America **102**, 10451 (2005), http://www.pnas.org/content/102/30/10451.full.pdf.
[3] Ponomarenko L. A., Geim A. K., Zhukov A. A., Jalil R., Morozov S. V., Novoselov K. S., Grigorieva I. V., Hill E. H., Cheianov V. V., Fal/'ko V. I., Watanabe K., Taniguchi T., and Gorbachev R. V., Nat Phys **7**, 958 (2011), 10.1038/nphys2114.
[4] K. S. Novoselov and A. H. C. Neto, Physica Scripta **2012**, 014006 (2012).
[5] Geim A. K. and Grigorieva I. V., 10.1038/nature12385 10.1038/nature12385.
[6] L. Britnell, R. V. Gorbachev, R. Jalil, B. D. Belle, F. Schedin, A. Mishchenko, T. Georgiou, M. I. Katsnelson, L. Eaves, S. V. Morozov, N. M. R. Peres, J. Leist, A. K. Geim, K. S. Novoselov, and L. A. Ponomarenko, Science **335**, 947 (2012), http://www.sciencemag.org/content/335/6071/947.full.pdf.
[7] L. Britnell, R. V. Gorbachev, R. Jalil, B. D. Belle, F. Schedin, M. I. Katsnelson, L. Eaves, S. V. Morozov, A. S. Mayorov, N. M. R. Peres, A. H. C. Neto, J. Leist, A. K. Geim, L. A. Ponomarenko, and K. S. Novoselov, Nano Letters **12**, 1707 (2012), pMID: 22380756, http://dx.doi.org/10.1021/nl3002205.
[8] Georgiou Thanasis, Jalil Rashid, Belle Branson D., Britnell Liam, Gorbachev Roman V., Morozov Sergey V., Kim Yong-Jin, Gholinia Ali, Haigh Sarah J., Makarovsky Oleg, Eaves Laurence, Ponomarenko Leonid A., Geim Andre K., Novoselov Kostya S., and Mishchenko Artem, Nat Nano **8**, 100 (2013).
[9] L. Britnell, R. M. Ribeiro, A. Eckmann, R. Jalil, B. D. Belle, A. Mishchenko, Y.-J. Kim, R. V. Gorbachev, T. Georgiou, S. V. Morozov, A. N. Grigorenko, A. K. Geim, C. Casiraghi, A. H. C. Neto, and K. S. Novoselov, Science **340**, 1311 (2013), http://www.sciencemag.org/content/340/6138/1311.full.pdf.
[10] Yu Woo Jong, Liu Yuan, Zhou Hailong, Yin Anxiang, Li Zheng, Huang Yu, and Duan Xiangfeng, Nat Nano **8**, 952 (2013).
[11] M. Massicotte, P. Schmidt, F. Vialla, K. G. Schädler, A. Reserbat-Plantey, K. Watanabe, T. Taniguchi, K.-J. Tielrooij, and F. H. L. Koppens, Nat Nano **11**, 42 (2016).
[12] S. J. Haigh, A. Gholinia, R. Jalil, S. Romani, L. Britnell, D. C. Elias, K. S. Novoselov, L. A. Ponomarenko, A. K. Geim, and R. Gorbachev, Nat Mater **11**, 764 (2012).
[13] W.-T. Pong and C. Durkan, Journal of Physics D: Applied Physics **38**, R329 (2005).
[14] J. M. B. Lopes dos Santos, N. M. R. Peres, and A. H. Castro Neto, Phys. Rev. Lett. **99**, 256802 (2007).
[15] A. Luican, G. Li, A. Reina, J. Kong, R. R. Nair, K. S. Novoselov, A. K. Geim, and E. Y. Andrei, Phys. Rev. Lett. **106**, 126802 (2011).
[16] C.-H. Park, L. Yang, Y.-W. Son, M. L. Cohen, and S. G. Louie, Phys. Rev. Lett. **101**, 126804 (2008).

[17] M. Yankowitz, J. Xue, D. Cormode, J. D. Sanchez-Yamagishi, K. Watanabe, T. Taniguchi, P. Jarillo-Herrero, P. Jacquod, and B. J. LeRoy, Nat Phys **8**, 382 (2012).
[18] C. Ortix, L. Yang, and J. van den Brink, Phys. Rev. B **86**, 081405 (2012).
[19] J. R. Wallbank, A. A. Patel, M. Mucha-Kruczyński, A. K. Geim, and V. I. Fal'ko, Phys. Rev. B **87**, 245408 (2013).
[20] M. Mucha-Kruczyński, J. R. Wallbank, and V. I. Fal'ko, Phys. Rev. B **88**, 205418 (2013).
[21] L. A. Ponomarenko, R. V. Gorbachev, G. L. Yu, D. C. Elias, R. Jalil, A. A. Patel, A. Mishchenko, A. S. Mayorov, C. R. Woods, J. R. Wallbank, M. Mucha-Kruczynski, B. A. Piot, M. Potemski, I. V. Grigorieva, K. S. Novoselov, F. Guinea, V. I. Fal/'ko, and A. K. Geim, Nature **497**, 594 (2013).
[22] R. Bistritzer and A. H. MacDonald, Phys. Rev. B **81**, 245412 (2010).
[23] A. Mishchenko, J. Tu, Y. Cao, R. V. Gorbachev, J. R. Wallbank, M. T. Greenaway, V. E. Morozov, S. V. Morozov, M. J. Zhu, S. L. Wong, F. Withers, C. R. Woods, Y.-J. Kim, K. Watanabe, T. Taniguchi, E. E. Vdovin, O. Makarovsky, T. M. Fromhold, V. I. Fal'ko, G. A. K., L. Eaves, and K. S. Novoselov, Nat Nano **9**, 808 (2014).
[24] L. Brey, Phys. Rev. Applied **2**, 014003 (2014).
[25] B. Fallahazad, K. Lee, S. Kang, J. Xue, S. Larentis, C. Corbet, K. Kim, H. C. P. Movva, T. Taniguchi, K. Watanabe, L. F. Register, S. K. Banerjee, and E. Tutuc, Nano Letters **15**, 428 (2015), pMID: 25436861, http://dx.doi.org/10.1021/nl503756y.
[26] S. Kang, B. Fallahazad, K. Lee, H. Movva, K. Kim, C. Corbet, T. Taniguchi, K. Watanabe, L. Colombo, L. Register, E. Tutuc, and S. Banerjee, Electron Device Letters, IEEE **36**, 405 (2015).
[27] S. C. de la Barrera and R. M. Feenstra, Applied Physics Letters **106**, 093115 (2015), http://dx.doi.org/10.1063/1.4914324.
[28] T. L. M. Lane, J. R. Wallbank, and V. I. Fal'ko, Applied Physics Letters **107**, 203506 (2015), http://dx.doi.org/10.1063/1.4935988.
[29] V. Perebeinos, J. Tersoff, and P. Avouris, Phys. Rev. Lett. **109**, 236604 (2012).
[30] S. Jung, M. Park, J. Park, T.-Y. Jeong, K. Kim, Ho-Jong Watanabe, T. Taniguchi, D. H. Ha, C. Hwang, and Y.-S. Kim, Scientific Reports **5** (2015), 10.1038/srep16642.
[31] E. E. Vdovin, A. Mishchenko, M. T. Greenaway, M. J. Zhu, D. Ghazaryan, A. Misra, Y. Cao, S. V. Morozov, O. Makarovsky, A. Patanè, G. J. Slotman, M. I. Katsnelson, A. K. Geim, K. S. Novoselov, and L. Eaves, ArXiv e-prints (2015), arXiv:1512.02143 [cond-mat.mes-hall].
[32] N. M. R. Peres, Rev. Mod. Phys. **82**, 2673 (2010).
[33] R. Geick, C. H. Perry, and G. Rupprecht, Phys. Rev. **146**, 543 (1966).
[34] R. M. Ribeiro and N. M. R. Peres, Phys. Rev. B **83**, 235312 (2011).
[35] M. Kindermann, B. Uchoa, and D. L. Miller, Phys. Rev. B **86**, 115415 (2012).
[36] J. Jung, A. Raoux, Z. Qiao, and A. H. MacDonald, Phys. Rev. B **89**, 205414 (2014).
[37] M. Koshino, New Journal of Physics **17**, 015014 (2015).
[38] R. Bistritzer and A. H. MacDonald, Proceedings of the National Academy of Sciences of the United States of America

108, 12233 (2011).

[39] J. M. B. Lopes dos Santos, N. M. R. Peres, and A. H. Castro Neto, Phys. Rev. B **86**, 155449 (2012).

[40] P. Moon and M. Koshino, Phys. Rev. B **87**, 205404 (2013).

[41] P. Moon and M. Koshino, Phys. Rev. B **90**, 155406 (2014).

[42] In Ref. 24, the processes corresponding to $\boldsymbol{\mathcal{Q}}_{0,1}$ and $\boldsymbol{\mathcal{Q}}_{0,2}$ where identified as being equivalent, with $|\boldsymbol{\mathcal{Q}}_{0,1}| = |\boldsymbol{\mathcal{Q}}_{0,2}|$. This was likely caused by first expanding $\boldsymbol{\mathcal{Q}}_{n,m}$ to linear order in θ_{tg}, θ_{hBN} and δ, and only then evaluating $|\boldsymbol{\mathcal{Q}}_{n,m}|$, loosing in the process terms involving the product $\delta\theta_{\mathrm{tg}}$ in $|\boldsymbol{\mathcal{Q}}_{n,m}|^2$, which lift the equivalence between the processes associated with $\boldsymbol{\mathcal{Q}}_{0,1}$ and $\boldsymbol{\mathcal{Q}}_{0,2}$.

[43] W. R. Frensley, Rev. Mod. Phys. **62**, 745 (1990).

[44] P. Vogl and T. Kubis, Journal of Computational Electronics **9**, 237 (2010).

[45] S. D. Stone and A. Szafer, IBM J. Res. Dev. **32**, 384 (1988).

[46] In mesoscopic transport, the problem of computing the current that is flowing through the device is reduced to a problem only involving degrees of freedom in the mesoscopic region by integrating out the external contacts. This is only done exactly provided the contacts are non-interacting.

[47] H. Haug and A.-P. Jauho, *Quantum Kinetics in Transport and Optics of Semiconductors*, 2nd ed., Springer Series in Solid-State Sciences, Vol. 123 (Springer-Verlag Berlin, Heidelberg, 2008).

[48] S. Datta, *Electronic Transport in Mesoscopic Systems*, Cambridge Studies in Semiconductor Physics and Microelectronic Engineering (Cambridge University Press, 1997).

[49] R. M. Feenstra, D. Jena, and G. Gu, Journal of Applied Physics **111**, 043711 (2012), http://dx.doi.org/10.1063/1.3686639.

[50] F. T. Vasko, Phys. Rev. B **87**, 075424 (2013).

[51] L. Brey, G. Platero, and C. Tejedor, Phys. Rev. B **38**, 9649 (1988).

[52] R. K. Hayden, D. K. Maude, L. Eaves, E. C. Valadares, M. Henini, F. W. Sheard, O. H. Hughes, J. C. Portal, and L. Cury, Phys. Rev. Lett. **66**, 1749 (1991).

[53] V. I. Fal'ko and S. V. Meshkov, Semiconductor Science and Technology **6**, 196 (1991).

[54] A. M. DaSilva, J. Jung, S. Adam, and A. H. MacDonald, Phys. Rev. B **92**, 155406 (2015).

[55] We point out that the Hartree (or tadpole) self-energy is local in time and as such does not give origin to lesser/greater self-energy terms, contributing only to the retarded/advanced self-energies.

[56] K. H. Michel and B. Verberck, Phys. Rev. B **83**, 115328 (2011).

[57] L. Wirtz and A. Rubio, Solid State Communications **131**, 141 (2004).

[58] S. Sen, F. Capasso, A. Y. Cho, and D. Sivco, IEEE Transactions on Electron Devices **34**, 2185 (1987).

[59] S. Sen, F. Capasso, A. Y. Cho, and D. L. Sivco, IEEE Electron Device Letters **9**, 533 (1988).

[60] C. R. Woods, L. Britnell, A. Eckmann, R. S. Ma, J. C. Lu, H. M. Guo, X. Lin, G. L. Yu, Y. Cao, R. V. Gorbachev, A. V. Kretinin, J. Park, L. A. Ponomarenko, M. I. Katsnelson, Y. N. Gornostyrev, K. Watanabe, T. Taniguchi, C. Casiraghi, H.-J. Gao, A. K. Geim, and K. S. Novoselov, Nat Phys **10**, 451 (2014).

[61] H. Suzuura and T. Ando, Phys. Rev. B **65**, 235412 (2002).