Multiple negative differential conductance regions and inelastic phonon assisted tunneling in graphene/*h*-BN/graphene structures

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

[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*

(Received 21 March 2016; published 2 June 2016)

In this paper we study in detail the effect of the rotational alignment between a hexagonal boron nitride (h-BN) slab and the graphene layers in the vertical current of a a graphene/h-BN/graphene device. We show how for small rotational angles, the transference of momentum by the h-BN 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 and 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.

DOI: 10.1103/PhysRevB.93.235403

I. INTRODUCTION

Being able to tailor the properties of materials at will, aiming at unveiling new physics, and achieving never thought 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 10 years, the advent of two-dimensional materials [1,2] opened new avenues waiting to be explored. One of the less explored avenues 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, which 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 (h-BN) as the semiconductor insulator, have appeared as some of the most promising configurations 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 of these devices as transistors [6–8], with ON/OFF ratios as high as 10^6; this magnitude of the ON/OFF ratio is obtained in graphene/WS_2/graphene devices [8]. It was also shown that graphene/STMDC/graphene devices can operate as photodetectors 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, the lattice mismatch and the relative alignment between consecutive layers play a fundamental role in determining the electronic coupling between different layers of the vdW structure. Ultimately, these crystallographic properties determine electronic and optical properties of the devices.

Lattice misalignment between different layers has been known to lead to the formation of Moiré patterns in rotated graphite layers [13]. The effect of lattice misalignment and lattice mismatch has been extensively studied in the context of twisted graphene bilayers and in graphene on h-BN structures. It was shown both theoretically and experimentally that the misalignment in graphene bilayers leads to a renormalization of graphene's Fermi velocity [14,15]. It was also found out that the mismatch and the misalignment control the formation of mini Dirac cones in the band structure of graphene/h-BN structures [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. In this reference was found that the current is extremely sensitive to the twisting angle. Although this dependence was not at first completely appreciated, it was soon understood and experimentally verified [23,24] that the misalignment between the graphene layers in graphene/h-BN/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. Systems displaying NDC have application in the design of high-frequency devices. Graphene/h-BN/graphene devices displaying NDC have been proposed as frequency multipliers [25]. This kind of system can also be used to implement radio-frequency oscillators and other devices operating in the MHz region; indeed these have already been achieved [23]. Theoretical analysis on the dependence of the oscillator frequency on the parameters of the graphene/h-BN/graphene device and of the external circuit indicate that oscillation frequencies up to hundreds of GHz could be achieved [26]. More recently, the effect of misalignment on the vertical current in devices formed by two graphene bilayers [27–29] and by one graphene monolayer and a graphene bilayer separated by h-BN has also been studied [30].

Scattering by phonons can lead to incoherent phonon assisted tunneling between the graphene layers. This effect has been first theoretically studied for vdW structures in twisted graphene bilayers [31]. More recently, effects of phonon assisted scattering on vertical transport have been experimentally detected in graphene/h-BN/graphite [32] and graphene/h-BN/graphene structures [33], and it has been proposed as a possible way to probe the phonon spectrum of vdW structures.

*amorim.bac@gmail.com

 ©2016 American Physical Society

In this paper, we describe the vertical current in graphene/h-BN/graphene devices with misaligned layers, and for small twist angles, properly taking into account momentum conservation rules. Our approach to the transport problem employs the nonequilibrium Green's function formalism and describes the graphene/h-BN/graphene system with a tight-binding-based continuous Hamiltonian. To lowest order in the graphene/h-BN coupling and neglecting the momentum and frequency dependence of the h-BN Green's function, the present formulation reduces to the theories used in Refs. [23,24]. By taking into account processes involving transference of momentum by the h-BN crystal lattice to the tunneling electrons, we find that the vertical current depends sensitively on the relative alignment between the graphene layers and the h-BN slab. 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. Differently from what is discussed in Ref. [24], the multipeak structure occurs at zero magnetic field. Application of an in-plane magnetic field further splits each peak into three. We also find out that the structure of graphene wave functions 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). This approach correctly describes the proportionality of the transport lifetime with the energy [34]. We finally study how phonons and disorder give origin to noncoherent current between the two graphene layers, deriving an expression for this quantity.

The paper is organized as follows. In Sec. II, we describe the theoretical framework we employed in this work: in Sec. II A, we present the Hamiltonian used to model the graphene/h-BN/graphene device, and in Sec. II B we present the fundamental equations used to treat transport within the nonequilibrium 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 h-BN 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 h-BN 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. This paper has some rather technical parts, namely, Secs. III A and IV A. The reader more interested in the results and in a physical discussion of them is invited to skip these subsections and go directly to Secs. III B and IV B. Longer derivations are included as Appendixes.

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 h-BN. The distance between the two graphene layers is given by d. We assume that the top graphene layer and the h-BN slab are rotated with respect to the bottom graphene layer by an angle of $\theta_{\rm tg}$ and $\theta_{h\text{-BN}}$, respectively. We assume that layers forming the h-BN slab are perfectly aligned with an AA' stacking [35,36] (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 $V_{\rm bias}$ can be applied between the top and bottom graphene layers inducing 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 a typical device structure is shown in Fig. 1.

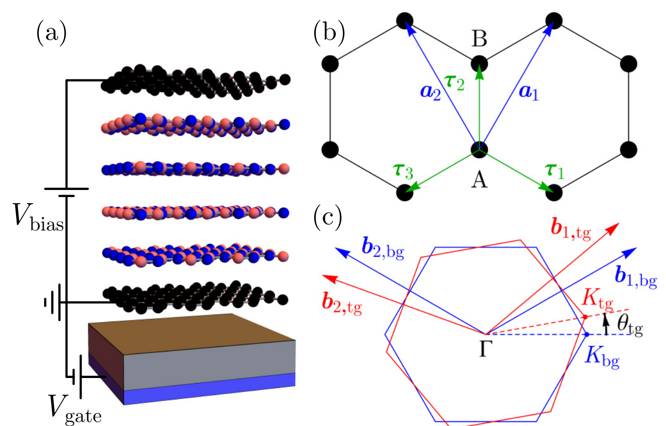

FIG. 1. (a) Schematic of a typical graphene/h-BN/graphene vdW structure with four boron-nitride layers, with applied gate $V_{\rm gate}$ and bias $V_{\rm bias}$ voltages. (b) Representation of crystalline structure shared by a graphene/boron-nitride monolayer, showing the lattice basis $\{\boldsymbol{a}_1,\boldsymbol{a}_2\}$, the nearest-neighbor vectors $\boldsymbol{\tau}_i$, $i=1,2,3$, and the sublattice A/B sites. (c) Representation of the first Brillouin zone of the rotated bottom and top graphene layers, showing the K points of both layers and the reciprocal lattice basis vectors $\{\boldsymbol{b}_{1,\rm bg,tg},\boldsymbol{b}_{2,\rm bg/tg}\}$.

A. Model Hamiltonian

We model the graphene/h-BN/graphene system with the following Hamiltonian:

$$H = H_{\rm bg} + H_{\rm tg} + H_{h\text{-BN}} + (T_{h\text{-BN},\rm bg} + T_{h\text{-BN},\rm tg} + \text{H.c.}), \quad (1)$$

where $H_{\rm bg/tg}$ is the Hamiltonian describing the isolated bottom/top graphene layer and $T_{h\text{-BN},\rm bg/tg} = T^{\dagger}_{\rm bg/tg,\mathit{h}\text{-BN}}$ describes the hopping of electrons from the bottom/top graphene layer to the h-BN slab. The current between the two graphene layers will be dominated by low-energy states. Focusing on the states close to the $\boldsymbol{K}_{\rm bg}$ and $\boldsymbol{K}'_{\rm bg} = -\boldsymbol{K}_{\rm bg}$ 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 a massless Dirac Hamiltonian

$$H_{\rm bg} = \sum_{\boldsymbol{k},\tau} \boldsymbol{c}^{\dagger}_{\boldsymbol{k}_\tau,\rm bg} \cdot \begin{bmatrix} V_{\rm bg} & \tau v_F \hbar |\boldsymbol{k}| e^{-\tau i\theta_{\boldsymbol{k},\rm bg}} \\ \tau v_F \hbar |\boldsymbol{k}| e^{-\tau i\theta_{\boldsymbol{k},\rm bg}} & V_{\rm bg} \end{bmatrix} \cdot \boldsymbol{c}_{\boldsymbol{k}_\tau,\rm bg}, \quad (2)$$

where v_F is graphene's Fermi velocity, $V_{\rm bg}$ is an onsite potential induced by the applied bias and gate voltages, $\boldsymbol{c}^{\dagger}_{\boldsymbol{k}_\tau,{\rm bg}} = [c^{\dagger}_{\boldsymbol{k}_\tau,A,{\rm bg}} \quad c^{\dagger}_{\boldsymbol{k}_\tau,B,{\rm bg}}]$ is the electron creation operator for states localized in the A/B sublattice, in the $\tau \boldsymbol{K}_{\rm bg}$ valley ($\tau = \pm 1$), with momentum $\tau \boldsymbol{K}_{\rm bg} + \boldsymbol{k}$ (measured from the Brillouin zone center), and $\theta_{\boldsymbol{k},{\rm bg}}$ is the angle formed between $\boldsymbol{k}$ and $\boldsymbol{K}_{\rm bg}$. We choose the zero of energy to lie at the Fermi level of the bottom graphene layer, in which case we have $V_{\rm bg} = -\epsilon_{\rm F,bg}$, where $\epsilon_{\rm F,bg}$ is the Fermi energy of the bottom graphene layer measured from its Dirac point. The Hamiltonian in Eq. (2) is diagonalized by the eigenstates $|\boldsymbol{k},\tau,\lambda\rangle_{\rm bg} = [1, \lambda\tau e^{-i\tau\theta_{\boldsymbol{k},{\rm bg}}}]^{\dagger}/\sqrt{2}$ with the corresponding dispersion relation given by $\epsilon_{\boldsymbol{k},\lambda} = \lambda v_F \hbar |\boldsymbol{k}|$, and with $\lambda = \pm 1$ for electrons in the conduction/valence band. 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 as

$$H_{\rm tg} = \sum_{\boldsymbol{k},\tau} \boldsymbol{c}^{\dagger}_{\boldsymbol{k}_\tau,{\rm tg}} \cdot \begin{bmatrix} V_{\rm tg} & \tau v_F \hbar |\boldsymbol{k}'| e^{-i\tau\theta_{\boldsymbol{k}',{\rm tg}}} \\ \tau v_F \hbar |\boldsymbol{k}'| e^{i\tau\theta_{\boldsymbol{k}',{\rm tg}}} & V_{\rm tg} \end{bmatrix} \cdot \boldsymbol{c}_{\boldsymbol{k}_\tau,{\rm tg}}, \quad (3)$$

where $\boldsymbol{k}' = \boldsymbol{k} + \tau \Delta \boldsymbol{K}_{\rm b,t}$ is measured from the $\tau \boldsymbol{K}_{\rm tg}$ point of the top graphene layer, with $\boldsymbol{K}_{\rm tg} = \mathcal{R}(\theta_{\rm tg}) \cdot \boldsymbol{K}_{\rm bg}$ [$\mathcal{R}(\theta)$ a rotation matrix], $\Delta \boldsymbol{K}_{\rm b,t} = \boldsymbol{K}_{\rm bg} - \boldsymbol{K}_{\rm tg}$ is the displacement between the Dirac points of the two rotated graphene layers, and $\boldsymbol{k}$ is measured with respect to the $\tau \boldsymbol{K}_{\rm bg}$ Dirac point of the bottom graphene graphene layer. $\theta_{\boldsymbol{k}',{\rm tg}}$ is the angle between $\boldsymbol{k}'$ and $\boldsymbol{K}_{\rm tg}$ and $V_{\rm tg}$ is an onsite potential, due to the applied bias and gate voltages, and is given by $V_{\rm tg} = -\epsilon_{\rm F,tg} - eV_{\rm bias}$, with $\epsilon_{\rm F,tg}$ the Fermi level of the top graphene layer measured from its Dirac point and $e > 0$ the elementary electronic charge. The remaining symbols in Eq. (3) are similarly defined to the ones in Eq. (2). Hamiltonians (2) and (3) do not take into account the effect of the Moiré potential due to lattice mismatch and misorientation between the graphene and the h-BN layers [16–20,37,38]. We will later discuss the possible effect of Moiré physics to our results (see Sec. III B). Due to the large band gap of boron nitride, we ignore its momentum dependence, writing the h-BN slab Hamiltonian as

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

where $\boldsymbol{c}^{\dagger}_{\boldsymbol{k}_\tau,\ell,h\text{-BN}} = [c^{\dagger}_{\boldsymbol{k}_\tau,{\rm B}\ell,h\text{-BN}} \quad c^{\dagger}_{\boldsymbol{k}_\tau,{\rm N}\ell,h\text{-BN}}]$ creates an electron in layer $\ell = 1,\ldots,\mathcal{N}$ of the h-BN slab, in the boron (B)/nitrogen (N) site, τ specifies the valley, $E_{\rm B}$ and $E_{\rm N}$ are, respectively, the onsite energies of boron and nitride sites measured from the Dirac point of graphene, and $t_\perp$ is the nearest-neighbor interlayer hopping and V_ℓ is a potential induced by the applied voltages. Due to the large energy offset between graphene and h-BN sites, the charge accumulated in the h-BN layers will be negligible. In this case, a simple electrostatic calculation (see Appendix A) gives us $V_\ell = -\epsilon_{\rm F,bg} - (\epsilon_{\rm F,tg} + eV_{\rm bias})\ell/(\mathcal{N}+1)$. 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 [22,39], in a so-called generalized *umklapp* process. Focusing on the low-energy states and considering only the three most relevant processes, the coupling between the graphene layers and the h-BN slab is described by (see Appendix B)

$$T_{h\text{-BN},{\rm X}} = \sum_{\boldsymbol{k},\tau} \sum_{n=0}^{2} \boldsymbol{c}^{\dagger}_{\boldsymbol{k}_\tau+\tau \boldsymbol{g}_n^{{\rm X},h\text{-BN}},\ell_{\rm X},h\text{-BN}} \cdot \boldsymbol{R}^n_{\frac{2\pi}{3}} \cdot \hat{\boldsymbol{T}} \cdot \boldsymbol{R}^n_{-\frac{2\pi}{3}} \cdot \boldsymbol{c}_{\boldsymbol{k}_\tau,{\rm X}}, \quad (5)$$

where $\boldsymbol{c}_{\boldsymbol{k}_\tau,{\rm X}}$ is an annihilation operator of an electron in the X = bg/tg graphene layer with momentum $\boldsymbol{k}_\tau$ measured from the $\tau \boldsymbol{K}_{\rm X}$ point, $\boldsymbol{c}^{\dagger}_{\boldsymbol{k}_\tau+\tau \boldsymbol{g}_n^{\rm X},\ell_{\rm X},h\text{-BN}}$ is a creation operator of an electron state in the $\ell_{\rm X} = 1/\mathcal{N}$ layer of the h-BN slab, with momentum $\boldsymbol{k}_\tau + \tau \boldsymbol{g}_n^{\rm X}$ measured from $\tau \boldsymbol{K}_{\rm X}$, and with the matrices $\hat{\boldsymbol{T}}$ and $\boldsymbol{R}_\theta$ defined as

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

$$\hat{\boldsymbol{T}} = \begin{bmatrix} t_{\rm B,C} & t_{\rm B,C} \\ t_{\rm N,C} & t_{\rm N,C} \end{bmatrix}, \quad (7)$$

where $t_{\rm B,C}$ ($t_{\rm N,C}$) is the hopping parameter between a carbon site and boron (nitrogen) site and the vectors $\boldsymbol{g}_n^{{\rm X},h\text{-BN}}$ are given by (see Appendix B)

$$\boldsymbol{g}_0^{{\rm X},h\text{-BN}} = 0, \quad (8)$$

$$\boldsymbol{g}_1^{{\rm X},h\text{-BN}} = \boldsymbol{b}_{2,{\rm X}} - \boldsymbol{b}_{2,h\text{-BN}}, \quad (9)$$

$$\boldsymbol{g}_2^{{\rm X},h\text{-BN}} = -\boldsymbol{b}_{1,{\rm X}} + \boldsymbol{b}_{1,h\text{-BN}}, \quad (10)$$

where $\boldsymbol{b}_{i,{\rm X}}$ and $\boldsymbol{b}_{i,h\text{-BN}}$ ($i = 1,2$) are, respectively, the reciprocal lattice vectors of the bottom/top graphene layer and of the h-BN slab (see Fig. 1). Notice that if the boron-nitride slab is formed by an even number of layers, one must replace $\boldsymbol{R}^n_{\frac{2\pi}{3}} \to \sigma_x \cdot \boldsymbol{R}^n_{\frac{2\pi}{3}} \cdot \sigma_x$ for $T_{h\text{-BN},{\rm tg}}$ since boron and nitrogen atoms switch positions in consecutive layers of h-BN. Different reciprocal lattice vectors are related to each other by $\boldsymbol{b}_{i,{\rm tg}} = \mathcal{R}(\theta_{\rm tg}) \cdot \boldsymbol{b}_{i,{\rm bg}}$ and $\boldsymbol{b}_{i,h\text{-BN}} = (a_{\rm g}/a_{h\text{-BN}})\mathcal{R}(\theta_{h\text{-BN}}) \cdot \boldsymbol{b}_{i,{\rm bg}}$, where $a_{\rm g}/a_{h\text{-BN}}$ is a scaling factor, with $a_{\rm g}$ ($a_{h\text{-BN}}$) the lattice parameter of graphene (h-BN). Hamiltonians of the form of Eq. (5) have previously been used to study twisted graphene bilayers [14,40–42] and graphene on h-BN structures [37,38,43]. Considering the three processes coupling the bottom graphene with h-BN and the three processes connecting h-BN to the top graphene layer, described by Eq. (5), there are nine h-BN 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 $\tau \boldsymbol{K}_{\rm bg}$) to states of the top graphene layer with momentum $\boldsymbol{k} + \tau \boldsymbol{\mathcal{Q}}_{n,m}$ (measured from $\tau \boldsymbol{K}_{\rm tg}$) with (see Appendix B)

$$\boldsymbol{\mathcal{Q}}_{n,m} = \Delta \boldsymbol{K}_{\rm b,t} + \boldsymbol{g}_n^{{\rm bg},h\text{-BN}} - \boldsymbol{g}_m^{{\rm tg},h\text{-BN}}, \quad n,m = 0,1,2. \quad (11)$$

The processes with $n \neq m$ involve transfer of momentum by the h-BN lattice, while processes with $n = m$ do not. At zero magnetic field, the overall threefold rotational invariance of the graphene/h-BN/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

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

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

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

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

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

with $K_{\text{g}} = 4\pi/(3a_{\text{g}})$ the length of $\boldsymbol{K}_{\text{bg/tg}}$. For numerical purposes we use $a_{\text{g}} \simeq 2.46$ Å, $a_{h\text{-BN}} \simeq 2.50$ Å, such that $\delta \simeq 1.8\%$, $v_F \simeq 10^6\ \text{ms}^{-1}$, $t_\perp \simeq 0.32$ eV [36], $E_{\text{B}} \simeq 3.33$ eV, $E_{\text{N}} \simeq -1.49$ eV, $t_{\text{B,C}} \simeq 0.432$ eV, and $t_{\text{N,C}} \simeq 0.29$ eV [38].

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 nonequilibrium state [44,45]. 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 [46]. Once water starts flowing through the pipe, the water levels in each vessel are no longer constant, 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 nonequilibrium Green's function technique, one can find an expression for the current of a mesoscopic device connected to two noninteracting[2] contacts, bottom and top, in a thermal equilibrium state described, respectively, by the Fermi-Dirac distribution functions $f_{\text{b}}(\omega) = [e^{\beta(\omega-\mu_{\text{b}})} + 1]^{-1}$ and $f_{\text{t}}(\omega) = [e^{\beta(\omega-\mu_{\text{t}})} + 1]^{-1}$ with $\mu_{\text{b (t)}}$ the chemical potential of the bottom (top) contact. In these conditions, the current flowing from the bottom to the top contact is given by [47] (using a compact notation where capital boldface symbols represent matrix elements evaluated in some one-particle electron basis and omitting the frequency argument of the different quantities)

$$I_{\text{b}\to\text{t}} = \frac{e}{\hbar}\int\frac{d\omega}{2\pi} f_{\text{b}}(\omega)\text{Tr}[\boldsymbol{\Gamma}_{\text{b}}\cdot\boldsymbol{A}] + \frac{e}{\hbar}\int\frac{d\omega}{2\pi} i\text{Tr}[\boldsymbol{\Gamma}_{\text{b}}\cdot\boldsymbol{G}^<], \tag{16}$$

with the spectral function of the central mesoscopic device given by

$$\boldsymbol{A} = i(\boldsymbol{G}^R - \boldsymbol{G}^A) = i(\boldsymbol{G}^> - \boldsymbol{G}^<), \tag{17}$$

where $\boldsymbol{G}^{R/A/</>}$ is the retarded/advanced/lesser/greater Green's function of the central device (which takes into account coupling to the external contacts) and $\boldsymbol{\Gamma}_{\text{b (t)}}$ 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: $\boldsymbol{\Gamma}_{\text{b (t)}} = 2\pi\boldsymbol{\tau}_{\text{b (t)}}\cdot\delta(\omega - \boldsymbol{H}_{\text{b (t)}})\cdot\boldsymbol{\tau}_{\text{b (t)}}^\dagger$, with $\boldsymbol{H}_{\text{b (t)}}$ the Hamiltonian describing the bottom (top) contact and $\boldsymbol{\tau}_{\text{b (t)}}$ 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 later be useful is [48]

$$\boldsymbol{A} = \boldsymbol{G}^R\cdot\boldsymbol{\Gamma}\cdot\boldsymbol{G}^A = \boldsymbol{G}^A\cdot\boldsymbol{\Gamma}\cdot\boldsymbol{G}^R, \tag{18}$$

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

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

Using the Dyson equation for the retarded/advanced Green's function $[\boldsymbol{G}^{R/A}]^{-1} = [\boldsymbol{G}^{0,R/A}]^{-1} - \boldsymbol{\Sigma}^{R/A}$, with $\boldsymbol{G}^0$ indicating the bare Green's function (in the absence of interactions and coupling to external contacts), and noting that $[\boldsymbol{G}^{0,R}]^{-1}$ and $[\boldsymbol{G}^{0,A}]^{-1}$ only differ by an infinitesimal constant that is taken to zero, the decay rate matrix can be written as

$$\boldsymbol{\Gamma} = i[\boldsymbol{\Sigma}^R - \boldsymbol{\Sigma}^A] = i[\boldsymbol{\Sigma}^> - \boldsymbol{\Sigma}^<], \tag{20}$$

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

$$\boldsymbol{G}^{\lessgtr} = \boldsymbol{G}^R\cdot\boldsymbol{\Sigma}^{\lessgtr}\cdot\boldsymbol{G}^A, \tag{21}$$

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

[2]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 noninteracting.

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

$$\mathbf{\Sigma}^{<} = i f_{\mathrm{b}} \mathbf{\Gamma}_{\mathrm{b}} + i f_{\mathrm{t}} \mathbf{\Gamma}_{\mathrm{t}} + \mathbf{\Sigma}^{<}_{\mathrm{int}}, \tag{22}$$

$$\mathbf{\Sigma}^{>} = -i(1-f_{\mathrm{b}})\mathbf{\Gamma}_{\mathrm{b}} - i(1-f_{\mathrm{t}})\mathbf{\Gamma}_{\mathrm{t}} + \mathbf{\Sigma}^{>}_{\mathrm{int}}, \tag{23}$$

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

$$\mathbf{\Gamma} = \mathbf{\Gamma}_{\mathrm{b}} + \mathbf{\Gamma}_{\mathrm{t}} + \mathbf{\Gamma}_{\mathrm{int}}. \tag{24}$$

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

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

with the coherent contribution being given by the Landauer formula

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

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

$$\mathcal{T} = \mathrm{Tr}[\mathbf{\Gamma}_{\mathrm{b}} \cdot \boldsymbol{G}^{R} \cdot \mathbf{\Gamma}_{\mathrm{t}} \cdot \boldsymbol{G}^{A}] \tag{27}$$

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

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

In the following sections, we will use this general formalism together with the model Hamiltonian from Sec. II A to evaluate the vertical current in graphene/h-BN/graphene structures.

III. CURRENT IN THE NONINTERACTING, PRISTINE LIMIT

A. General discussion

When applying the nonequilibrium Green's function formalism to a graphene/h-BN/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 works to date, the graphene layers were assumed to be part of the external contacts [23,24,29,49,50], therefore 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 noninteracting, pristine case. In approach (B), $\boldsymbol{G}^{R/A}(\omega)$ in Eq. (27) is the Green's function of the graphene/h-BN/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

$$[\boldsymbol{G}^{R}]_{\mathrm{b,t}} = \boldsymbol{G}^{0,R}_{\mathrm{bg}} \cdot \boldsymbol{\mathcal{T}}_{\mathrm{bg,tg}} \cdot \boldsymbol{G}^{0,R}_{\mathrm{tg}}, \tag{29}$$

$$[\boldsymbol{G}^{A}]_{\mathrm{t,b}} = \boldsymbol{G}^{0,A}_{\mathrm{tg}} \cdot \boldsymbol{\mathcal{T}}_{\mathrm{tg,bg}} \cdot \boldsymbol{G}^{0,A}_{\mathrm{bg}}, \tag{30}$$

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/h-BN coupling (but taking into account the coupling to the external contacts) where we have defined the h-BN mediated tunneling amplitudes

$$\boldsymbol{\mathcal{T}}_{\mathrm{bg,tg}} = \boldsymbol{T}_{\mathrm{bg},h\text{-BN}} \cdot \boldsymbol{G}^{R}_{h\text{-BN}} \cdot \boldsymbol{T}_{h\text{-BN,tg}}, \tag{31}$$

$$\boldsymbol{\mathcal{T}}_{\mathrm{tg,bg}} = \boldsymbol{T}_{\mathrm{tg},h\text{-BN}} \cdot \boldsymbol{G}^{A}_{h\text{-BN}} \cdot \boldsymbol{T}_{h\text{-BN,bg}}, \tag{32}$$

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

$$\mathcal{T} = \mathrm{Tr}\big[\boldsymbol{G}^{0,A}_{\mathrm{bg}} \cdot \mathbf{\Gamma}_{\mathrm{b}} \cdot \boldsymbol{G}^{0,R}_{\mathrm{bg}} \boldsymbol{\mathcal{T}}_{\mathrm{bg,tg}} \cdot \boldsymbol{G}^{0,R}_{\mathrm{tg}} \cdot \mathbf{\Gamma}_{\mathrm{t}} \cdot \boldsymbol{G}^{0,A}_{\mathrm{tg}} \cdot \boldsymbol{\mathcal{T}}_{\mathrm{tg,bg}}\big]. \tag{33}$$

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 $\boldsymbol{A}^{0}_{\mathrm{bg}} = i(\boldsymbol{G}^{0,R}_{\mathrm{bg}} - \boldsymbol{G}^{0,A}_{\mathrm{bg}})$, as $\boldsymbol{A}^{0}_{\mathrm{bg}} = \boldsymbol{G}^{0,R}_{\mathrm{bg}} \cdot \mathbf{\Gamma}_{\mathrm{b}} \cdot \boldsymbol{G}^{0,A}_{\mathrm{bg}} = \boldsymbol{G}^{0,A}_{\mathrm{bg}} \cdot \mathbf{\Gamma}_{\mathrm{b}} \cdot \boldsymbol{G}^{0,R}_{\mathrm{bg}}$ and similarly for the top graphene layer. As such, the transmission function can be written as

$$\mathcal{T} = \mathrm{Tr}\big[\boldsymbol{\mathcal{T}}_{\mathrm{bg,tg}} \cdot \boldsymbol{A}^{0}_{\mathrm{tg}} \cdot \boldsymbol{\mathcal{T}}_{\mathrm{tg,bg}} \cdot \boldsymbol{A}^{0}_{\mathrm{bg}}\big]. \tag{34}$$

Equation (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 $\mathbf{\Gamma}_{\mathrm{b}} = \boldsymbol{T}_{h\text{-BN,bg}} \cdot \boldsymbol{A}^{0}_{\mathrm{bg}} \cdot \boldsymbol{T}_{h\text{-BN,bg}}$ and $\mathbf{\Gamma}_{\mathrm{t}} = \boldsymbol{T}_{h\text{-BN,tg}} \cdot \boldsymbol{A}^{0}_{\mathrm{tg}} \cdot \boldsymbol{T}_{h\text{-BN,tg}}$. As such, we have proved that in the noninteracting 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 analytical progress, we will employ the wide-band limit for the metallic contacts, neglecting any frequency dependence of $\mathbf{\Gamma}_{\mathrm{b/t}}$, 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, which is the case in the experimental devices. 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/h-BN/graphene device more explicitly. Using the graphene/h-BN coupling Hamiltonian (5), the transmission function can be written using the Bloch momentum basis as (writing explicitly the frequency argument)

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

with the sum on n,m going from 0 to 2 and where $\omega_{\rm bg} = \omega + \epsilon_{\rm F,bg}$ and $\omega_{\rm tg} = \omega + \epsilon_{\rm F,tg} + eV_{\rm bias}$ 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

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

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

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

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

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

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

$$\mathcal{T}(\omega) = Ag_sg_v\sum_{n,m=0}^{3}|\mathcal{T}_{n,m}|^2{\rm TDoS}_{n,m}(\omega_{\rm bg},\omega_{\rm tg}), \tag{39}$$

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

$${\rm TDoS}_{n,m}(\omega_{\rm bg},\omega_{\rm tg}) = \sum_{\lambda,\lambda'=\pm1}\int\frac{d^2\boldsymbol{k}}{(2\pi)^2}\Upsilon^{{\rm bg},n}_{\boldsymbol{k},\lambda}\Upsilon^{{\rm tg},m}_{\boldsymbol{k}+\boldsymbol{\mathcal{Q}}_{n,m},\lambda'} \times A^0_{{\rm bg},\boldsymbol{k},\lambda}(\omega_{\rm bg})A^0_{{\rm tg},\boldsymbol{k}+\boldsymbol{\mathcal{Q}}_{n,m},\lambda'}(\omega_{\rm tg}), \tag{40}$$

which only depends on the graphene's dispersion relation and wave functions (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 analytical expression for ${\rm TDoS}_{n,m}(\omega_{\rm bg},\omega_{\rm tg})$. In the presence of a finite, momentum-independent, lifetime, it is still possible to find an approximate analytical expression to Eq. (40). These analytical expressions lead to a significant speedup 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/h-BN/graphene structure is controlled both by energy-momentum conservation and by Pauli's exclusion principle. The constraints 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 $\epsilon_{\rm F,tg} + eV_{\rm bias} - \epsilon_{\rm F,bg}$ and in momentum by a value of $|\boldsymbol{\mathcal{Q}}_{n,m}|$ (see Fig. 2). 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

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

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 tunneling density of

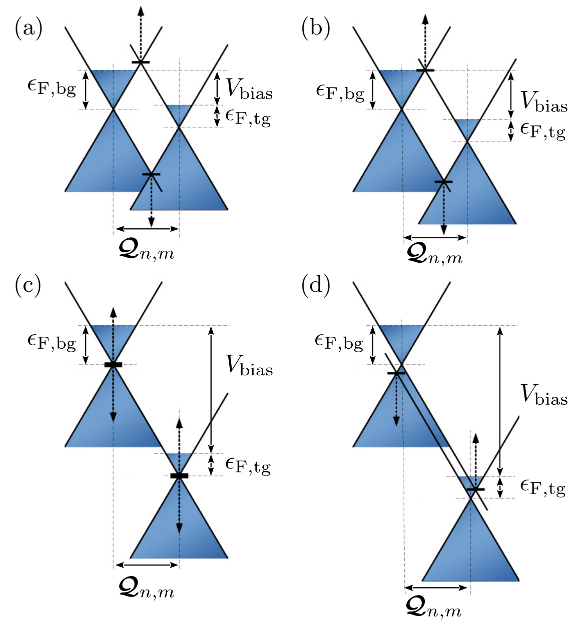

FIG. 2. Band diagram representing the constraints imposed by energy-momentum conservation and Pauli's exclusion principle in the vertical current of a graphene/h-BN/graphene device. The two cones represent the dispersion relation for electrons of the bottom and top graphene layers. The shadowed blue 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 intrabands 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) The 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.

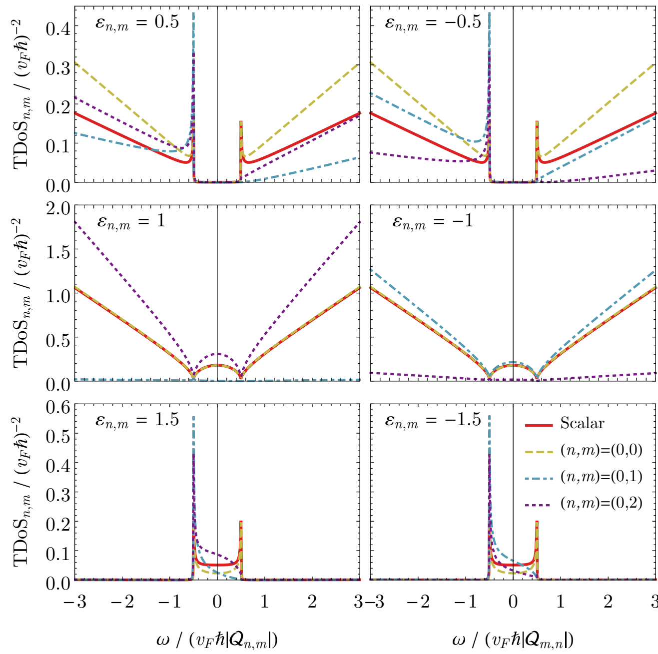

FIG. 3. Plot of the quantity $\mathrm{TDoS}_{m,n}(\omega + \varepsilon_{m,n} v_F \hbar |\mathcal{Q}_{n,m}|/2, \omega - \varepsilon_{n,m} v_F \hbar |\mathcal{Q}_{n,m}|/2)$ for different values of $\varepsilon_{m,n}$ as a function of the energy at zero magnetic field and for rotation angles of $\theta_{\mathrm{TG}} = 1^\circ$ and $\theta_{h\text{-BN}} = 1.5^\circ$. The solid red line shows the tunneling density of states if the wave-function overlap factors $\Upsilon^{B/T,n}_{\boldsymbol{k},\lambda}$ in Eq. (40) are set to one. A constant broadening factor of $\gamma = 2.5 \times v_F \hbar |\mathcal{Q}_{n,m}| \times 10^{-3}$ was used in all plots.

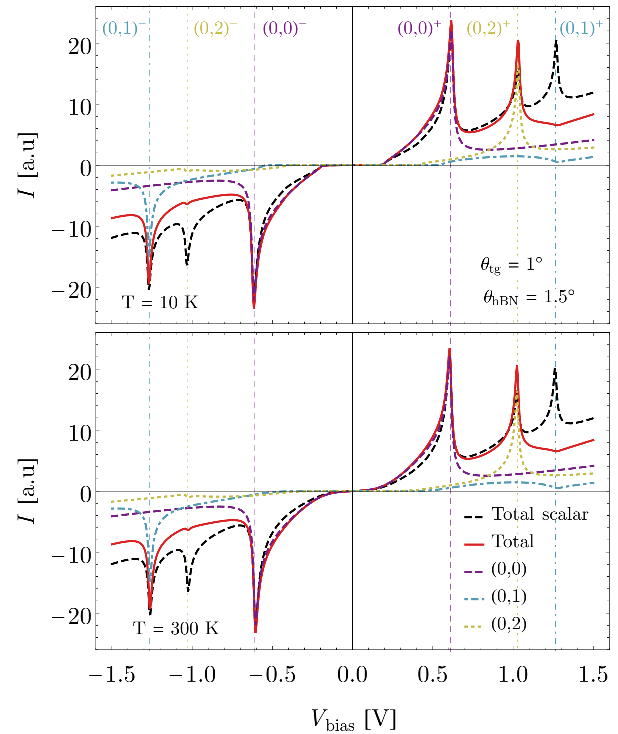

FIG. 4. I-V curves for vertical current in a graphene/h-BN/graphene device with four layers of h-BN for rotation angles of $\theta_{\mathrm{tg}} = 1^\circ$ and $\theta_{h\text{-BN}} = 1.5^\circ$ at gate voltage $V_{\mathrm{gate}} = 0$ for two different temperatures. The solid red line indicates the current due to all the nine processes coupling both graphene layers, for graphene electrons, while the dashed black lines represents the total current for scalar electrons (by setting the wave-function factors $\Upsilon^{\mathrm{bg/tg},n}_{\boldsymbol{k},\lambda}$ to 1). The remaining lines represent the contributions to the current arising from processes involving different $\mathcal{Q}_{n,m}$ [taking into account the relations imposed by threefold rotational invariance, Eq. (12)]. The dashed vertical lines labeled by $(n,m)^\pm$ mark the bias voltages when the condition $\epsilon_{\mathrm{F,tg}} + eV_{\mathrm{bias}} - \epsilon_{\mathrm{F,bg}} = \pm v_F \hbar |\mathcal{Q}_{n,m}|$ 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 $\Upsilon^{\mathrm{bg/tg},n}_{\boldsymbol{k},\lambda}$ factors. A constant broadening factor of $\gamma = 2.5$ meV was used.

states $\mathrm{TDoS}_{n,m}$. In Fig. 3, we plot the quantity $\mathrm{TDoS}_{n,m}(\omega - \varepsilon_{n,m} v_F \hbar |\mathcal{Q}_{n,m}|/2, \omega + \varepsilon_{n,m} v_F \hbar |\mathcal{Q}_{n,m}|/2)$, for different values of $\varepsilon_{n,m} = (\epsilon_{\mathrm{F,tg}} + eV_{\mathrm{bias}} - \epsilon_{\mathrm{F,bg}})/(v_F \hbar |\mathcal{Q}_{n,m}|)$. 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 zero in the pristine limit for $\omega^2 < (v_F \hbar)^2 |\mathcal{Q}_{n,m}|^2/4$. 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 $\omega^2 > (v_F \hbar)^2 |\mathcal{Q}_{n,m}|^2/4$. For $\varepsilon^2_{n,m} = 1$, $\mathrm{TDoS}_{n,m}(\omega - \varepsilon_{n,m} v_F \hbar |\mathcal{Q}_{n,m}|/2, \omega + \varepsilon_{n,m} v_F \hbar |\mathcal{Q}_{n,m}|/2)$ diverges in the pristine limit for any value of ω. This divergence in $\mathrm{TDoS}_{n,m}$ 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 $|\mathcal{Q}_{n,m}|$ there is a different effective separation in momentum between the Dirac cones of the 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 those 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 wave functions, via the overlap factors $\Upsilon^{\mathrm{bg/tg},n}_{\boldsymbol{k},\lambda}$, that appear in Eq. (40). These overlap factors can severely suppress the value of $\mathrm{TDoS}_{n,m}$ 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. 3, where a considerable suppression of $\mathrm{TDoS}_{n,m}$ for $\varepsilon_{0,1} = 1$ and $\varepsilon_{0,2} = -1$ is seen. 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 wave function of graphene where scalars, i.e., by setting $\Upsilon^{\mathrm{bg/tg},n}_{\boldsymbol{k},\lambda} = 1$ in (40) [see Eq. (C20) in Appendix C], displaying the three peaks expected by kinematic considerations. While the occurrence of NDC in graphene/h-BN/graphene has already 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

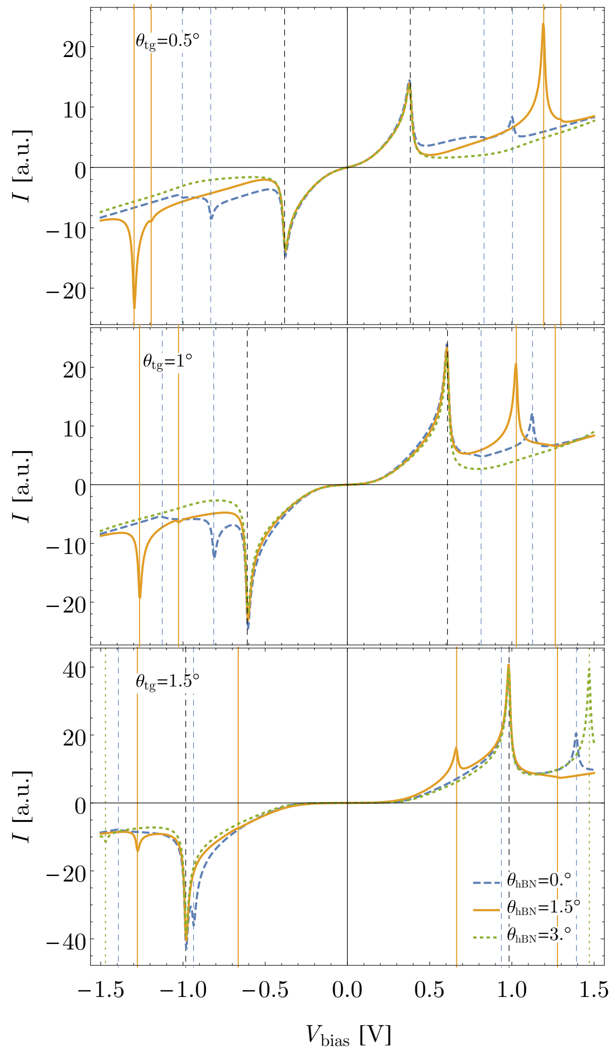

FIG. 5. I-V curves at constant $V_{\text{gate}} = 0$ in a graphene/h-BN/graphene device with four layers of h-BN, at $V_{\text{gate}} = 0$ and $T = 300$ K, for different rotation angles between the top and bottom graphene layers, and the h-BN slab and the bottom graphene layer. The black dashed line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

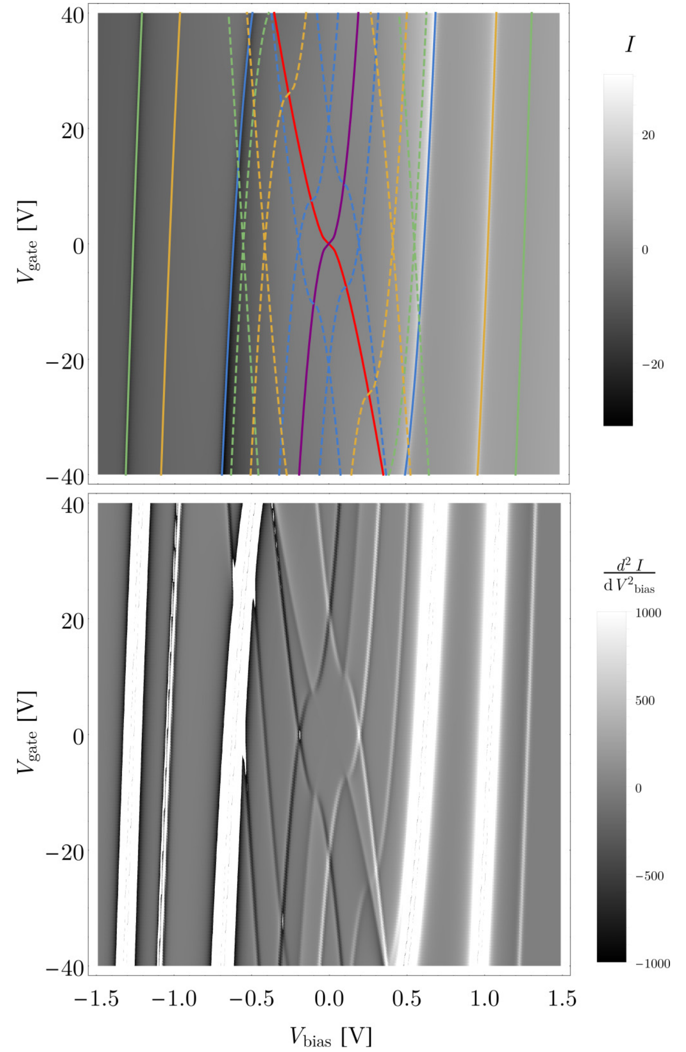

FIG. 6. Density plot of current I and its second derivative with respect to the applied bias voltage d^2I/dV_{bias}^2, as a function of the applied bias and gate voltages at $T = 10$ K. In the current plot, it is also shown the lines defined by the following conditions: $\epsilon_{\text{F,bg}} = 0$ and $\epsilon_{\text{F,tg}} = 0$, represented by the solid lines in red and purple; $\epsilon_{\text{F,tg}} + eV_{\text{bias}} - \epsilon_{\text{F,bg}} = \pm v_F \hbar |\boldsymbol{Q}_{0,m}|$ [Eq. (41)] for $m = 0$, 1, and 2, represented by the solid lines in blue, green, and yellow, respectively; $\epsilon_{\text{F,tg}} \pm eV_{\text{bias}} + \epsilon_{\text{F,bg}} = \pm \frac{1}{2} v_F \hbar |\boldsymbol{Q}_{0,m}|$ for $m = 0$, 1, and 2 [Eq. (42)] represented by the dashed lines in blue, green, and yellow. Notice now the guide 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.

θ_{tg} and $\theta_{h\text{-BN}}$. 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 $\theta_{\text{tg}} = 1^\circ$, changing $\theta_{h\text{-BN}}$ from 1.5° to 3° moves the additional peaks in the current due to the transfer of momentum by the h-BN crystal lattice from a bias voltage of $\sim$1 V to bias voltages >1.5 V. Tunneling processes 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. 2. 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. 2(b)]

$$\epsilon_{\text{F,tg}} \pm eV_{\text{bias}} + \epsilon_{\text{F,bg}} = \pm \tfrac{1}{2} v_F \hbar |\boldsymbol{Q}_{n,m}|. \tag{42}$$

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 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 a smooth function of the energy.

By applying an in-plane magnetic field, the threefold rotational invariance of the graphene/h-BN/graphene structure is broken and, therefore, the processes corresponding to the different groups in (12) will contribute differently to the

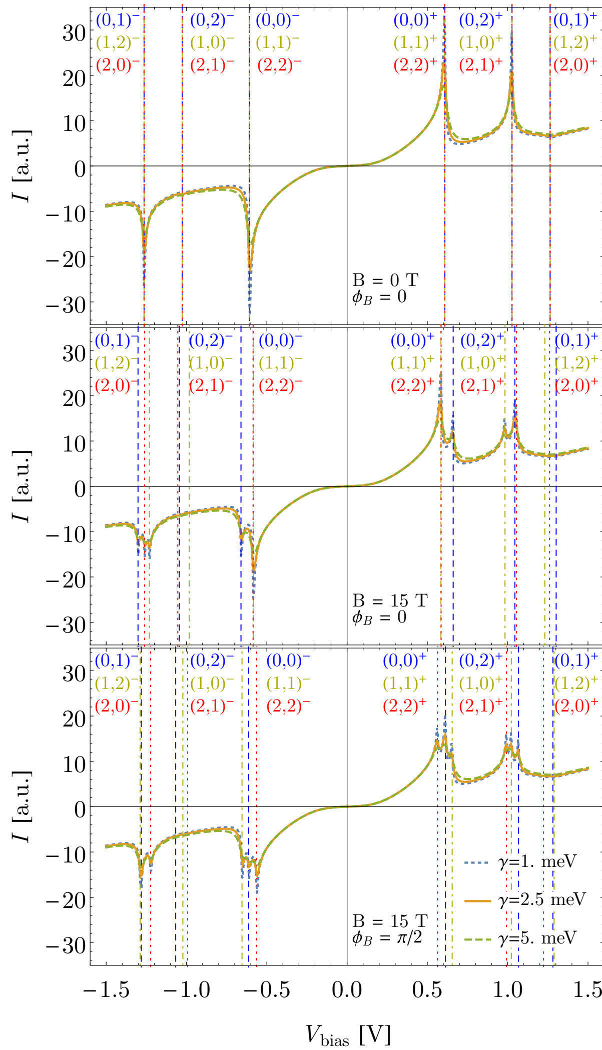

FIG. 7. I-V curves for a graphene/h-BN/graphene device with four layers of h-BN, with rotation angles of $\theta_{\text{tg}} = 1^\circ$ and $\theta_{\text{hBN}} = 1.5^\circ$ at constant $V_{\text{gate}} = 0$ and $T = 300$ K, for different values and orientation of the in-plane magnetic field and electronic 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.

current, and one expects that each peak in the I-V curve will split into three. An in-plane magnetic field of the form $\boldsymbol{B} = B(\cos\phi_B, \sin\phi_B, 0)$ can be described by the vector potential $\boldsymbol{A} = Bz(\sin\phi_B, -\cos\phi_B)$. Neglecting the momentum dependence of $H_{h\text{-BN}}$ 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 $\mathcal{Q}_{n,m}$ vectors [23,24,51–53]

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

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

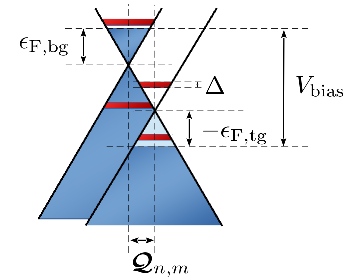

FIG. 8. Diagram representing the possible effect of the reconstruction of the graphene Dirac spectrum, due to the presence of h-BN, in the vertical current of a graphene/h-BN/graphene device when $\varepsilon_{n,m} = \pm 1$. The red bars represent the position in energy of the regions, of width Δ, where graphene's spectrum reconstruction is significant. Provided $eV_{\text{bias}}/\Delta \gg 1$, the peaks in the current will still be present.

Finally, we comment on the possible effect of the h-BN in the electronic structure of graphene. It is known that the potential modulation with the periodicity of the Moiré pattern formed by graphene on top of h-BN can lead to a reconstruction of the density of states of graphene at energies of the order of $\pm v_F \hbar |\boldsymbol{g}_{1/2}^{\text{g},h\text{-BN}}|/2$ measured from the original Dirac cone, where $|\boldsymbol{g}_{1/2}^{\text{g},h\text{-BN}}| \simeq 4\pi\sqrt{\delta^2 + \theta_{\text{g},h\text{-BN}}{}^2}/(\sqrt{3}a_{\text{g}})$ is the wave vector of the Moiré pattern reciprocal lattice [16–21] and with $\theta_{\text{g},h\text{-BN}}$ is the rotation angle between the graphene layer and the h-BN 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 h-BN lattice, only appear for reasonable values of the bias voltage for small twist angles between the graphene layers and h-BN slab. It is precisely in this case 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 h-BN 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 this 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/h-BN/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 or few hundreds 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/h-BN/graphene

device. We consider a generic electron-phonon interaction described by the Hamiltonian

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

where $\phi_{\zeta} = (a_{\zeta} + a_{\zeta}^{\dagger})/\sqrt{2}$ 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)[3] contribution to the lesser/greater self-energy is given by (from now on we write all frequency arguments explicitly)

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

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

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

where $b(\nu) = (e^{\beta\nu}-1)^{-1}$ 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 as

$$\boldsymbol{\Sigma}_{\text{e-ph}}^{\lessgtr}(\omega) = \sum_{\zeta,s=\pm1}[\pm sb(\pm s\omega_{\zeta})]\boldsymbol{M}_{\zeta}\cdot\boldsymbol{G}^{\lessgtr}(\omega-s\omega_{\zeta})\cdot\boldsymbol{M}_{\zeta}^{\dagger}. \tag{47}$$

We point out that this self-energy can also describe elastic scattering by impurities by dropping the summation over s, taking $\omega_{\zeta}\to 0$, and setting $\pm sb(\pm s\omega_{\zeta})\to 1$, 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

$$\begin{aligned}\boldsymbol{\Sigma}_{\text{e-ph}}^{<}(\omega) \simeq &\sum_{\zeta,s=\pm1} if_{\text{b}}(\omega-s\omega_{\zeta})sb(s\omega_{\zeta})\boldsymbol{M}_{\zeta}\cdot\boldsymbol{G}^{R}(\omega-s\omega_{\zeta})\cdot\boldsymbol{\Gamma}_{\text{b}}(\omega-s\omega_{\zeta})\cdot\boldsymbol{G}^{A}(\omega-s\omega_{\zeta})\cdot\boldsymbol{M}_{\zeta}^{\dagger}\\ &+\sum_{\zeta,s=\pm1} if_{\text{t}}(\omega-s\omega_{\zeta})sb(s\omega_{\zeta})\boldsymbol{M}_{\zeta}\cdot\boldsymbol{G}^{R}(\omega-s\omega_{\zeta})\cdot\boldsymbol{\Gamma}_{\text{t}}(\omega-s\omega_{\zeta})\cdot\boldsymbol{G}^{A}(\omega-s\omega_{\zeta})\cdot\boldsymbol{M}_{\zeta}^{\dagger},\end{aligned} \tag{48}$$

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

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

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

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

It is easy to check that

$$f_{\text{b}}(\omega)[1-f_{\text{b}}(\omega-s\omega_{\zeta})][-sb(-s\omega_{\zeta})] = [1-f_{\text{b}}(\omega)]f_{\text{b}}(\omega-s\omega_{\zeta})sb(s\omega_{\zeta}) \tag{51}$$

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 $I_{\text{t}\to\text{b}} = -I_{\text{t}\to\text{b}}$. As such, terms that involve only the occupation factor of one the contacts must cancel at the end of any calculation. A similar approach to this one has been previously used to study phonon assisted tunneling in metal/insulator/metal devices in Ref. [55]. Processes assisted by a greater number of phonons can also be obtained. Higher-order corrections to Eq. (48) can be derived by iterating Eq. (47) using Eqs. (21) and (22). 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

$$\begin{aligned}I_{\text{b}\to\text{t}}^{\text{incoh, }n\text{-ph}} = &\frac{e}{\hbar}\sum_{\zeta_1,s_1,\dots,\zeta_n,s_n}\int\frac{d\omega}{2\pi}f_{\text{b}}(\omega)\big[1-f_{\text{t}}\big(\omega+s_1\omega_{\zeta_1}+\cdots+s_n\omega_{\zeta_n}\big)\big]\big[s_1b\big(s_1\omega_{\zeta_1}\big)\big]\cdots\big[s_nb\big(s_n\omega_{\zeta_n}\big)\big]\mathcal{T}_{\text{b}\to\text{t}}^{(\zeta_1,s_1)\dots(\zeta_n,s_n)}(\omega)\\ &-\frac{e}{\hbar}\sum_{\zeta_1,s_1,\dots,\zeta_n,s_n}\int\frac{d\omega}{2\pi}f_{\text{t}}(\omega)\big[1-f_{\text{b}}\big(\omega+s_1\omega_{\zeta_1}+\cdots+s_n\omega_{\zeta_n}\big)\big]\big[s_1b\big(s_1\omega_{\zeta_1}\big)\big]\cdots\big[s_nb\big(s_n\omega_{\zeta_n}\big)\big]\mathcal{T}_{\text{t}\to\text{b}}^{(\zeta_1,s_1)\dots(\zeta_n,s_n)}(\omega),\end{aligned} \tag{52}$$

[3]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.

where we have defined the n-phonon assisted transmission functions

$$\mathcal{T}_{\rm b\to t}^{(\zeta_1,s_1)\dots(\zeta_n,s_n)}(\omega)=\mathrm{Tr}\big[\mathbf{\Gamma}_{\rm b}(\omega)\cdot\boldsymbol{G}^R(\omega)\cdot\boldsymbol{M}_{\zeta_1}\cdot\boldsymbol{G}^R\big(\omega+s_1\omega_{\zeta_1}\big)\dots\cdot\boldsymbol{M}_{\zeta_n}\cdot\boldsymbol{G}^R\big(\omega+s_1\omega_{\zeta_1}+\dots+s_n\omega_{\zeta_n}\big)\cdot\mathbf{\Gamma}_{\rm t}\big(\omega+s_1\omega_{\zeta_1}+\dots+s_n\omega_{\zeta_n}\big)\cdot\boldsymbol{G}^A\big(\omega+s_1\omega_{\zeta_1}+\dots+s_n\omega_{\zeta_n}\big)\cdot\boldsymbol{M}^\dagger_{\zeta_n}\cdot\dots\cdot\boldsymbol{G}^A\big(\omega+s_1\omega_{\zeta_1}\big)\cdot\boldsymbol{M}^\dagger_{\zeta_1}\boldsymbol{G}^A(\omega)\big], \tag{53}$$

$$\mathcal{T}_{\rm t\to b}^{(\zeta_1,s_1)\dots(\zeta_n,s_n)}(\omega)=\mathrm{Tr}\big[\mathbf{\Gamma}_{\rm t}(\omega)\cdot\boldsymbol{G}^A(\omega)\cdot\boldsymbol{M}^\dagger_{\zeta_1}\cdot\boldsymbol{G}^A\big(\omega+s_1\omega_{\zeta_1}\big)\cdot\dots\cdot\boldsymbol{M}^\dagger_{\zeta_n}\boldsymbol{G}^A\big(\omega+s_1\omega_{\zeta_1}+\dots+s_n\omega_{\zeta_n}\big)\cdot\mathbf{\Gamma}_{\rm b}\big(\omega+s_1\omega_{\zeta_1}+\dots+s_n\omega_{\zeta_n}\big)\cdot\boldsymbol{G}^R\big(\omega+s_1\omega_{\zeta_1}+\dots+s_n\omega_{\zeta_n}\big)\cdot\boldsymbol{M}_{\zeta_n}\cdot\dots\cdot\boldsymbol{G}^R\big(\omega+s_1\omega_{\zeta_1}\big)\cdot\boldsymbol{M}_{\zeta_1}\cdot\boldsymbol{G}^R(\omega)\big]. \tag{54}$$

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 $\omega\to\omega+s_1\omega_{\zeta_1}+\dots+s_n\omega_{\zeta_n}$ in the second line of Eq. (52). Equations (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 scattered 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/h-BN/graphene device. We will analyze separately the effect of scattering by graphene and h-BN phonons.

1. Scattering by phonons in the graphene layers

We now return to the issue of the consequences of considering graphene as part of the external contacts or part of the central mesoscopic region. For simplicity, 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/h-BN coupling as

$$\mathcal{T}_{\rm b\to t}^{(\zeta_1,s_1)\dots(\zeta_n,s_n)}(\omega)\simeq\mathrm{Tr}\big[\boldsymbol{\mathcal{T}}_{\rm tg,bg}(\omega)\cdot\mathbf{\Gamma}_{\rm b}(\omega)\cdot\boldsymbol{G}^R_{\rm bg}(\omega)\cdot\boldsymbol{\mathcal{T}}_{\rm bg,tg}(\omega)\cdot\boldsymbol{G}^R_{\rm tg}(\omega)\cdot\boldsymbol{M}_{\zeta_1}\cdot\boldsymbol{G}^R_{\rm tg}\big(\omega+s_1\omega_{\zeta_1}\big)\dots\cdot\boldsymbol{M}_{\zeta_n}\cdot\boldsymbol{G}^R_{\rm tg}\big(\omega+s_1\omega_{\zeta_1}+\dots+s_n\omega_{\zeta_n}\big)\cdot\mathbf{\Gamma}_{\rm t}\big(\omega+s_1\omega_{\zeta_1}+\dots+s_n\omega_{\zeta_n}\big)\cdot\boldsymbol{G}^A_{\rm tg}\big(\omega+s_1\omega_{\zeta_1}+\dots+s_n\omega_{\zeta_n}\big)\cdot\boldsymbol{M}^\dagger_{\zeta_n}\cdot\dots\cdot\boldsymbol{G}^A_{\rm tg}\big(\omega+s_1\omega_{\zeta_1}\big)\cdot\boldsymbol{M}^\dagger_{\zeta_1}\boldsymbol{G}^A_{\rm tg}(\omega)\big], \tag{55}$$

and similarly for $\mathcal{T}_{\rm t\to b}^{(\zeta_1,s_1)\dots(\zeta_n,s_n)}(\omega)$. 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

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

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

$$\boldsymbol{A}^{(0)}_{\rm tg}(\omega)=\boldsymbol{G}^R(\omega)\cdot\mathbf{\Gamma}_{\rm t}(\omega)\cdot\boldsymbol{G}^A(\omega),\quad n=0 \tag{57}$$

$$\begin{aligned}\boldsymbol{A}^{(n)}_{\rm tg}(\omega)=\sum_{s_n,\zeta_n}&\frac{\big[1-f_{\rm t}\big(\omega-s_n\omega_{\zeta_n}\big)\big]\big[-s_nb\big(-s_n\omega_{\zeta_n}\big)\big]}{1-f_{\rm t}(\omega)}\\&\times\boldsymbol{G}^R_{\rm tg}(\omega)\cdot\boldsymbol{M}_{\zeta_n}\cdot\boldsymbol{A}^{(n-1)}_{\rm tg}\big(\omega-s_n\omega_{\zeta_n}\big)\\&\cdot\boldsymbol{M}^\dagger_{\zeta_n}\cdot\boldsymbol{G}^A_{\rm tg}(\omega),\quad n>0.\end{aligned} \tag{58}$$

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

$$\boldsymbol{A}_{\rm tg}(\omega)\simeq\boldsymbol{G}^R(\omega)\cdot[\mathbf{\Gamma}_{\rm t}(\omega)+\mathbf{\Gamma}_{\text{e-ph,tg}}(\omega)]\cdot\boldsymbol{G}^A(\omega), \tag{59}$$

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

$$\mathbf{\Gamma}_{\text{e-ph,tg}}(\omega)\simeq\sum_{s,\zeta}s[1-f_{\rm t}(\omega-s\omega_\zeta)+b(s\omega_\zeta)]\times\boldsymbol{M}_\zeta\cdot\boldsymbol{A}_{\rm tg}(\omega-s\omega_\zeta)\cdot\boldsymbol{M}^\dagger_\zeta. \tag{60}$$

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

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

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

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

with the different terms $\boldsymbol{A}_{\rm tg}^{(n)}(\omega)$ coinciding with Eqs. (57) and (58), and the $\sim eq$ 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 $\boldsymbol{\Gamma}_{\rm b}$. Nevertheless, by using Eq. (18), the previous calculation can also be applied for scattering in the bottom graphene layer. We have thus arrived at an important conclusion: the expression

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

which would be the one obtained if we employed approach (A), actually already includes the effect of multiple noncoherent scattering processes in the graphene layers, provided $\boldsymbol{A}_{\rm tg/bg}(\omega)$ 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/h-BN 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 h-BN that appears in $\boldsymbol{T}_{\rm bg,tg}(\omega)$, 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/h-BN 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 h-BN slab*

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

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

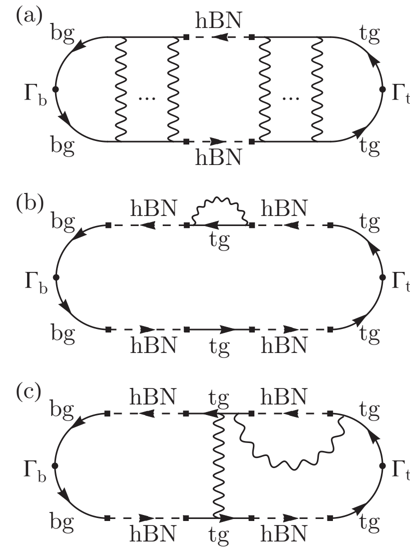

FIG. 9. Diagrammatic representation of contributions to the current involving phonon scattering. The dots represent the level width functions due to the bottom and top external metallic contacts, the squares represent the graphene/h-BN coupling, solid lines represent graphene electronic propagators, and dashed lines represent h-BN propagators. The wiggly lines represent phonon propagators. (a) Ladder diagrams that are resummed in Eq. (63). To lowest order in the graphene/h-BN 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/h-BN 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 $\boldsymbol{G}_{h\text{-BN}}$. (c) Higher-order diagrams in the graphene/h-BN 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

$$\boldsymbol{c}^\dagger_{\boldsymbol{k},h\text{-BN}} = [c^\dagger_{\boldsymbol{k},\text{B},1,h\text{-BN}}, c^\dagger_{\boldsymbol{k},\text{N},1,h\text{-BN}}, \ldots, c^\dagger_{\boldsymbol{k},\text{B},\mathcal{N},h\text{-BN}}, c^\dagger_{\boldsymbol{k},\text{N},\mathcal{N},h\text{-BN}}] \quad (65)$$

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

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

$$\times [1 - f_{\rm b}(\omega)][-sb(-s\omega_\zeta)]\left|\mathcal{T}^{(\zeta,s)\text{1-ph}}_{n,m,\boldsymbol{k},\boldsymbol{q}}(\omega)\right|^2$$
$$\times \Upsilon^{\text{bg},n}_{\boldsymbol{k},\lambda}\Upsilon^{\text{tg},m}_{\boldsymbol{k}+\boldsymbol{\mathcal{Q}}_{n,m}-\boldsymbol{q},\lambda'}A_{\text{bg},\boldsymbol{k},\lambda}(\omega_{\rm bg})$$
$$\times A_{\text{tg},\boldsymbol{k}+\boldsymbol{\mathcal{Q}}_{n,m}-\boldsymbol{q},\lambda'}(\omega_{\rm tg} + s\omega_\zeta),$$

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

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

Neglecting the momentum and frequency dependence of $\boldsymbol{G}^A_{h\text{-BN}}$ 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

$$I^{\text{incoh, 1-ph}}_{\rm b\to t} = AA_{\rm cell}g_s g_v \frac{e}{\hbar}\sum_{n,m}\sum_{\zeta,s} \times \int \frac{d\omega}{2\pi}\{f_{\rm b}(\omega)[1 - f_{\rm t}(\omega + s\omega_\zeta)]sb(s\omega_\zeta) - f_{\rm t}(\omega + s\omega_\zeta)[1 - f_{\rm b}(\omega)]s[1 + b(\omega_\zeta)]\} \times \left|\mathcal{T}^{(\zeta)\text{1-ph}}_{n,m}\right|^2 \text{DoS}_{\rm bg}(\omega_{\rm bg})\text{DoS}_{\rm tg}(\omega_{\rm tg} + s\omega_\zeta), \tag{66}$$

where $A_{\rm cell}$ is the area of the unit cell of h-BN and graphene's density of states per spin and valley is given by

$$\text{DoS}(\omega) = \frac{1}{V}\sum_{\boldsymbol{k},\lambda} A_{\boldsymbol{k},\lambda}(\omega) = \frac{|\omega|}{(v_F\hbar)^2}, \tag{67}$$

where the last equality is valid for noninteracting electrons in pristine graphene. A similar expression to Eq. (66), 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. [33] to model vertical current in graphene/h-BN/graphene devices. In the case of elastic scattering by disorder with short-range correlation, Eq. (66) becomes

$$I^{\text{incoh, 1-dis}}_{\rm b\to t} = AA_{\rm cell}g_s g_v \frac{e}{\hbar}\sum_{\substack{\zeta,s\\ n,m}}\int \frac{d\omega}{2\pi}[f_{\rm b}(\omega) - f_{\rm t}(\omega)] \times \left|\mathcal{T}^{\text{1-dis}}_{n,m}\right|^2 \text{DoS}_{\rm bg}(\omega_{\rm bg})\text{DoS}_{\rm tg}(\omega_{\rm tg}), \tag{68}$$

with $\mathcal{T}^{\text{1-dis}}_{n,m}$ a disorder assisted tunneling amplitude. Although an expression of the form of Eq. (68) was previously used to model vertical current in graphene/h-BN/graphene devices [6,7], we emphasize that Eq. (68) only describes processes where there is a complete degradation of in-plane momentum conservation (as dispersionless phonons or short-range impurities can carry an arbitrary momentum in a scattering event), something that has been previously pointed out in Refs. [24,29]. 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 nonzero components of polarization vector given by

$$\boldsymbol{\xi}^z_{\text{ZB},a,\ell} = \left(\xi^z_{\zeta,\text{B},1}, \xi^z_{\zeta,\text{N},1}, \xi^z_{\zeta,\text{B},2}, \xi^z_{\zeta,\text{N},2}, \dots\right) = \sqrt{\frac{\mu_{\rm BN}}{\mathcal{N}}}\left(\frac{1}{\sqrt{m_{\rm B}}}, \frac{1}{\sqrt{m_{\rm N}}}, \frac{-1}{\sqrt{m_{\rm B}}}, \frac{-1}{\sqrt{m_{\rm N}}}, \dots\right), \tag{69}$$

where $\mu_{\rm BN}^{-1} = m_{\rm B}^{-1} + m_{\rm N}^{-1}$ is the reduced mass of the h-BN phonon mode. We assume that electron-phonon coupling for this mode can be described as a local change in the value of the interlayer hopping parameter in Hamiltonian (4). Considering electrons 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 momentum-independent coupling constant which reads as

$$\boldsymbol{M}^{h\text{-BN}}_{\rm ZB} = \frac{g^{h\text{-BN}}_{\rm ZB}}{\sqrt{\mathcal{N}}}\begin{bmatrix} \boldsymbol{0} & \boldsymbol{\sigma}_x & & \\ \boldsymbol{\sigma}_x & \boldsymbol{0} & -\boldsymbol{\sigma}_x & \\ & -\boldsymbol{\sigma}_x & \boldsymbol{0} & \ddots \\ & & \ddots & \ddots \end{bmatrix}, \tag{70}$$

with the electron-phonon coupling constant given by

$$g^{h\text{-BN}}_{\rm ZB} = -\frac{\partial \,\ln t_\perp}{\partial \,\ln c_{\rm BN}}\frac{t_\perp}{c_{\rm BN}}\sqrt{\frac{\hbar}{\mu_{\rm BN}\omega^{h\text{-BN}}_{\rm ZB}}}, \tag{71}$$

where $-\partial \,\ln t_\perp/\partial \,\ln c_{\rm BN}$ describes the change of the interlayer hopping $t_\perp$, with the interlayer distance, $c_{\rm BN} \simeq 3.3$ Å, and for p orbitals is estimated as $\sim eq3$ [56,57], $\omega^{h\text{-BN}}_{\rm ZB}$ 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

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

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 phonons (see Appendix E). For comparison, we also show the 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

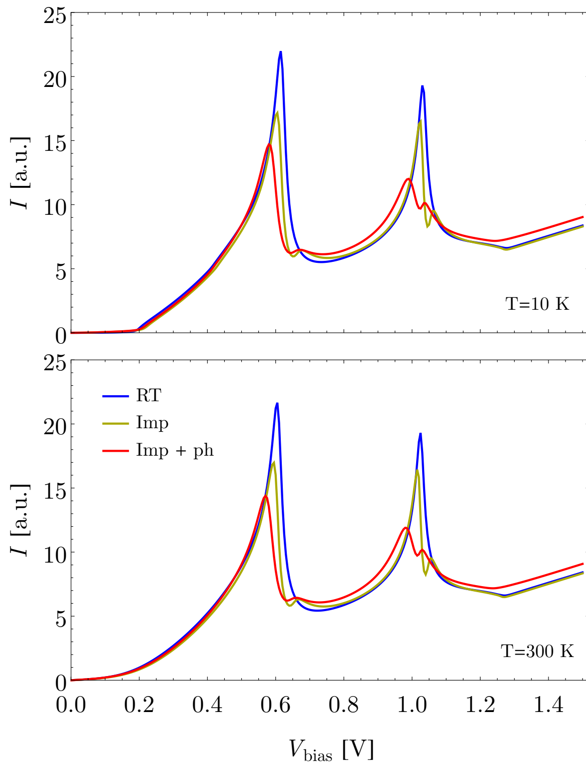

FIG. 10. I-V curves at constant $V_{\text{gate}} = 0$ in a graphene/h-BN/graphene device with rotations angles $\theta_{\text{tg}} = 1^{\circ}$ and $\theta_{h\text{-BN}} = 1.5^{\circ}$, considering different sources of scattering in the graphene layers: (RT) constant relaxation time of $\gamma = 3$ meV; (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 $n_{\text{imp}} = 10^{-4}$.

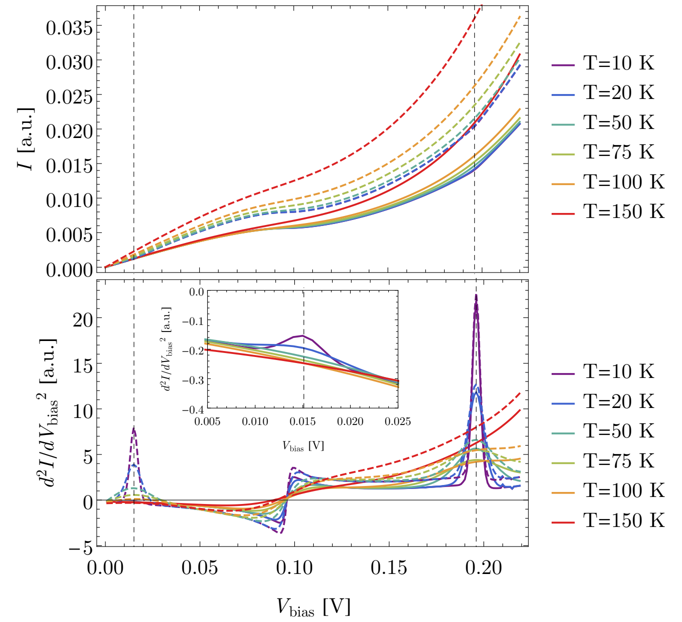

FIG. 11. I-V curve and d^2I/dV_{bias}^2 as a function of bias voltages at a constant $V_{\text{gate}} = 10$ V for different temperatures and for rotation angles $\theta_{\text{tg}} = 2^{\circ}$ and $\theta_{h\text{-BN}} = 3^{\circ}$, including effects of scattering by out-of-plane breathing phonons of h-BN, $\omega_{\text{ZB}}^{h\text{-BN}} = 15$ meV [58], and of the in-plane graphene phonons, $\omega_{\Gamma\text{O}}^{\text{g}} = 196$ meV [59] (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 solid lines are the curves computed with the electron-phonon coupling parameters estimated using Eqs. (71) and (E2): $g_{\text{ZB}}^{h\text{-BN}} \simeq 0.06$ eV and $g_{\Gamma\text{O}}^{\text{g}} \simeq 0.6$ eV for h-BN and graphene, respectively. The dashed lines show the curves computed with a $g_{\text{ZB}}^{h\text{-BN}}$ that is 10 times larger. The inset zooms in the small peak due to the h-BN out-of-plane breathing phonon. 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.

part of the self-energy, both that of resonant scatterers and that of 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/h-BN/graphene device. This changes if one focuses on small bias. At very low temperature, the spontaneous emission of optical phonons becomes possible 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, where we show the current and d^2I/dV_{bias}^2 taking into account the effect of scattering by in-plane optical graphene phonons and out-of-plane boron-nitride breathing phonons. 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. Notice that the electron-phonon couplings estimated using Eqs. (71) and (E2) yield a peak in d^2I/dV_{bias}^2 due to the h-BN breathing mode that is nearly two orders of magnitude larger than the one due to the graphene in-plane optical phonons, while in experimental situations [32,33] all the peaks seem to be of the same order of magnitude. This might either be due to Eq. (71) being an oversimplified estimation of the electron-phonon coupling for the h-BN breathing mode, or due to the possibility of the peak seen in experiments at $\sim$0.15 V also having contributions from a graphene/h-BN out-of-plane interfacial phonon which we are not considering and might have a larger electron-phonon coupling. We also point out that the features due to phonons are a small contribution that 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.

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 [60,61], 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/h-BN/graphene vdW structure for small rotational misalignment, which have so far not been observed [23]. We have seen how the transference of momentum, by the h-BN 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 h-BN slab, and rotational angles as small as 3° can already push these additional peaks to bias voltages higher than 1.5 V. Therefore, the observation of multiple NDC in graphene/h-BN/graphene devices requires a control of the rotational angle between the different layers with a precision of $\lesssim 1^\circ$, something which is within experimental reach [21,23,62].

We have also analyzed the effect of treating graphene as being the source and drain contacts of the graphene/h-BN/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 noninteracting case, both approaches are equivalent. In the presence of interactions, both approaches are equivalent to lowest order in the graphene/h-BN coupling.

Finally, we have studied, in a unified way, the effect of scattering by disorder and phonon scattering in the vertical current of graphene/h-BN/graphene devices. Starting from a nonequilibrium Green's function (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 temperatures appears as sharp features in the derivatives of the current with respect to the bias voltage at the energy of the phonons. This signature of the phonon assisted tunneling 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 h-BN optical out-of-plane breathing phonons. We have not considered the effect of vibrations at the graphene/h-BN interface, as these would require the description of phonons in incommensurate structures, something which will be the focus of future work.

As a side note, we expect that the possible reconstruction of graphene spectrum due to the periodic potential induced by h-BN for small rotational angles should not affect in a qualitative way the occurrence of multiple NDC regions in graphene/h-BN/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 and will be the goal of a future work.

ACKNOWLEDGMENTS

B. Amorim acknowledges financial support from Fundação para a Ciência e a Tecnologia (Portugal), through Grant No. SFRH/BD/78987/2011. R. M. Ribeiro and N. M. R. Peres acknowledge funding from the European Commission through the project "Graphene Driven Revolutions in ICT and Beyond", Reference No. 696656, and from Fundação para a Ciência e a Tecnologia (Portugal) in the framework of the Strategic Funding UID/FIS/04650/2013.

APPENDIX A: THOMAS-FERMI MODELING OF ELECTROSTATIC DOPING

We wish to model the charging of a graphene/h-BN/graphene device by application of a gate V_{gate} and bias V_{bias} voltage. The graphene/h-BN/graphene structure is formed by $\mathcal{N}$ h-BN monolayers, sandwiched between two graphene layers. The graphene/h-BN/graphene structure is on top of a dielectric spacer (typically h-BN/SiO_2) separating the structure from a back gate, typically a highly doped Si layer. We treat each layer forming the graphene/h-BN/graphene structure as a 2D film with a two-dimensional charge density given by ρ_ℓ, $\ell = -1, \ldots, \mathcal{N}+1$, where $\ell = -1$ indexes the Si layer, $\ell = 0$ and $\ell = \mathcal{N}+1$ are, respectively, the bottom and top graphene layers, and $\ell = 1, \ldots, \mathcal{N}$ index the layers of h-BN 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 $\bar{\epsilon}_\ell$. Applying Gauss's law around each plate, and assuming charge neutrality $\sum_{\ell=-1}^{\mathcal{N}} \rho_\ell = 0$, we obtain

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

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

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

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

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

and the stored electrostatic energy is given by

$$\begin{aligned} U_{EM} &= \sum_{\ell=0}^{\mathcal{N}+1} \frac{1}{2} \epsilon_0 d_\ell \bar{\epsilon}_\ell E_\ell^2 \\ &= \frac{1}{2} \sum_{\ell,\ell'=0}^{\mathcal{N}+1} \rho_\ell \left(\sum_{k=0}^{\min(\ell,\ell')} \frac{d_k}{\epsilon_0 \bar{\epsilon}_k} \right) \rho_{\ell'}, \end{aligned} \tag{A5}$$

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 $\rho_\ell = -en_\ell + en_\ell^{\text{imp}}$, where n_ℓ is the charge carrier concentration ($n_\ell > 0$ for electron doping) and n_ℓ^{imp} is the concentration of charged impurities ($n_\ell^{\text{imp}} > 0$ 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

$$\Phi = \frac{1}{2}\sum_{\ell,\ell'=1}^{\mathcal{N}+1} n_\ell \left(\sum_{k=0}^{\min(\ell,\ell')} \frac{e^2 d_k}{\epsilon_0 \bar{\epsilon}_k}\right) n_{\ell'} - \sum_{\ell=0}^{\mathcal{N}+1} n_\ell e\phi_\ell^{\text{imp}} - eV_{\text{gate}}\sum_{\ell=0}^{\mathcal{N}+1} n_\ell + eV_{\text{bias}}\sum_{\ell=0}^{\mathcal{N}+1}\frac{\ell}{\mathcal{N}+1} n_\ell, \tag{A6}$$

where

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

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

$$V_\ell^H = -\frac{\partial \Phi}{\partial n_\ell} = V_{\text{gate}} - V_{\text{bias}}\frac{\ell}{\mathcal{N}+1} n_\ell + e\phi_\ell^{\text{imp}} - \sum_{\ell'=0}^{\mathcal{N}+1}\left(\sum_{k=0}^{\min(\ell,\ell')}\frac{e^2 d_k}{\epsilon_0\bar{\epsilon}_k}\right) n_{\ell'}, \quad \ell = 0,\ldots,\mathcal{N}+1. \tag{A8}$$

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 $\epsilon_{\text{F},\ell} = V_\ell^H$, where $\epsilon_{\text{F},\ell}$ is a function of the local carrier density. This together with Eq. (A8) becomes a system of nonlinear equations in the carrier density/local Fermi level.

It can be checked that due to the large band gap of h-BN, most charge density will be accumulated in the graphene layers. As such, we approximate $n_\ell = 0$, for $\ell = 1,\ldots,\mathcal{N}$, and therefore the $\mathcal{N}+2$ equations are reduced to two

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

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

where the capacitances are given by (taking into account the series capacitances of a h-BN/SiO_2 spacer with $d_{h\text{-BN}}$ the h-BN thickness and d_{SiO_2} the SiO_2 thickness)

$$C_{\text{b}}^{-1} = \frac{e^2 d_0}{\epsilon_0\bar{\epsilon}_0} = \frac{e^2 d_{SiO_2}}{\epsilon_0\bar{\epsilon}_{SiO_2}} + \frac{e^2 d_{h\text{-BN}}}{\epsilon_0\bar{\epsilon}_{h\text{-BN}}},$$

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

$$= \frac{e^2 d_{SiO_2}}{\epsilon_0\bar{\epsilon}_{SiO_2}} + \frac{e^2 d_{h\text{-BN}}}{\epsilon_0\bar{\epsilon}_{h\text{-BN}}} + \frac{e^2 d}{\epsilon_0\bar{\epsilon}_{h\text{-BN}}}, \tag{A12}$$

and d is the distance between the two graphene layers. The terms $e\phi_{\text{bg/tg}}^{\text{imp}}$ 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 h-BN layers have no charge carrier, then the Hartree potential within the h-BN slab is given from Eq. (A8) in terms of $\epsilon_{\text{F,bg/tg}}$ as

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

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

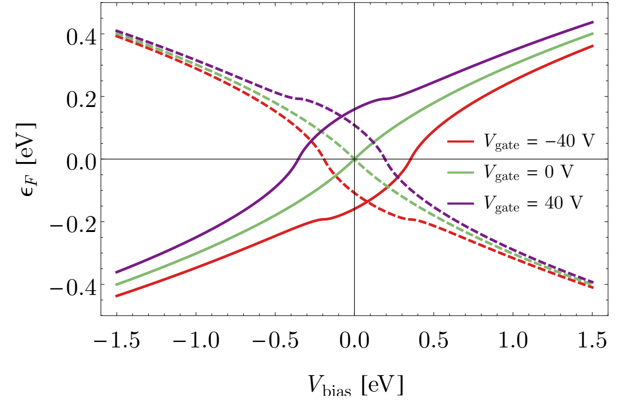

FIG. 12. Computed Fermi levels for the bottom and top graphene layers as a function of bias voltage for different gate voltages obtained by solving Eqs. (A9) and (A10). We assume that the following parameters $d_{SiO_2} = 285$ nm, $d_{h\text{-BN}} = 40$ nm for the thickness of the back-gate dielectric, with out-of-plane dielectric constants $\bar{\epsilon}_{SiO_2} = 3.9$ and $\bar{\epsilon}_{h\text{-BN}} = 5.09$ [35]. We assumed that the distance between the two graphene layers is separated by four monolayers of h-BN, which corresponds to a distance between the graphene layers of $d \simeq 1.6$ nm.

APPENDIX B: INTERLAYER HOPPING HAMILTONIAN BETWEEN NONCOMMENSURATE LAYERS

We describe the graphene/boron nitride coupling using the general theory of coupling between noncommensurate layers of Refs. [22,39]. 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

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

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 $t(\mathbf{R}_{n,a,\ell},\mathbf{R}_{m,b,\ell'})$ are hopping terms. Assuming that the hopping $t(\mathbf{R}_{n,a,\ell},\mathbf{R}_{m,b,\ell'})$ only depends on $\mathbf{R}_{n,a,\ell} - \mathbf{R}_{m,b,\ell'}$, it is possible

to write it in Fourier components as [39]

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

where $A_{\text{cell},\ell/\ell'}$ 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

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

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

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

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

where $\mathbf{G}_{n,\ell/\ell'}$ are reciprocal lattice vectors of the 2D crystal ℓ/ℓ'. The Kronecker δ imposes that in an interlayer hopping 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 $\pm\boldsymbol{K}_\text{g}$ of the graphene layer. Considering only the three most relevant processes coupling the graphene and boron-nitride layers, we must consider processes involving $\mathbf{G}_{n,\text{g}} = \mathbf{0},\boldsymbol{b}_{2,\text{g}},-\boldsymbol{b}_{1,g}$ and $\mathbf{G}_{n,\text{h-BN}} = \mathbf{0},\boldsymbol{b}_{2,h\text{-BN}},-\boldsymbol{b}_{1,h\text{-BN}}$ for states close to the $\boldsymbol{K}_\text{g}$ point and processes involving $\mathbf{G}_{n,\text{g}} = \mathbf{0},-\boldsymbol{b}_{2,\text{g}},\boldsymbol{b}_{1,\text{g}}$ and $\mathbf{G}_{n,h\text{-BN}} = \mathbf{0},-\boldsymbol{b}_{2,h\text{-BN}},\boldsymbol{b}_{1,h\text{-BN}}$ for states close to the $-\boldsymbol{K}_\text{g}$ point. It is also assumed that the momentum dependence of $t_{a,b}^{\ell,\ell'}(\boldsymbol{k})$ is weak such that we can approximate $t_{a,b}^{\ell,\ell'}(\boldsymbol{k}+\boldsymbol{K}_\pm+\mathbf{G}_{n,\text{g}}) \simeq t_{a,b}^{\ell,\ell'}(\boldsymbol{K})$, setting $t_{\text{B,A}}^{h\text{-BN,g}}(\boldsymbol{K}) = t_{\text{B,B}}^{h\text{-BN,g}}(\boldsymbol{K}) = t_{\text{B,C}}$ and $t_{\text{N,A}}^{h\text{-BN,g}}(\boldsymbol{K}) = t_{\text{N,B}}^{h\text{-BN,g}}(\boldsymbol{K}) = t_{\text{N,C}}$.

In order to describe the coupling between the bottom and top graphene layers to a slab formed by $\mathcal{N}$ h-BN monolayers, we notice that the products of unit-cell basis vectors and reciprocal lattice vectors that appear in Eq. (B4) can be written for the bottom graphene layer as $\boldsymbol{\tau}_{h\text{-BN,B1}}\cdot\mathbf{G}_{n,h\text{-BN}} = \boldsymbol{\tau}_{\text{bg,A}}\cdot\mathbf{G}_{n,\text{bg}} = 0$ and $\boldsymbol{\tau}_{h\text{-BN,N1}}\cdot\mathbf{G}_{n,h\text{-BN}} = \boldsymbol{\tau}_{\text{bg,B}}\cdot\mathbf{G}_{n,\text{bg}} = \pm n2\pi/3$ (for states close to $\pm\boldsymbol{K}_\text{g}$ point). For the coupling between the top graphene layer and the $\mathcal{N}$th$h$-BN layer, one must consider separately the cases when the h-BN slab is formed by an 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 first layer and therefore we still have $\boldsymbol{\tau}_{h\text{-BN,B}\mathcal{N}}\cdot\mathbf{G}_{n,h\text{-BN}} = \boldsymbol{\tau}_{\text{tg,A}}\cdot\mathbf{G}_{n,\text{tg}} = 0$ and $\boldsymbol{\tau}_{h\text{-BN,N}\mathcal{N}}\cdot\mathbf{G}_{n,h\text{-BN}} = \boldsymbol{\tau}_{\text{tg,B}}\cdot\mathbf{G}_{n,\text{tg}} = \pm n2\pi/3$. If we have an even number of h-BN layers, then in the $\mathcal{N}$th layer, the boron and nitrogen atoms switch positions compared to the first layer, and one obtains instead $\boldsymbol{\tau}_{h\text{-BN,N}\mathcal{N}}\cdot\mathbf{G}_{n,h\text{-BN}} = \boldsymbol{\tau}_{\text{tg,A}}\cdot\mathbf{G}_{n,\text{tg}} = 0$ and $\boldsymbol{\tau}_{h\text{-BN,B}\mathcal{N}}\cdot\mathbf{G}_{n,h\text{-BN}} = \boldsymbol{\tau}_{\text{tg,B}}\cdot\mathbf{G}_{n,\text{tg}} = \pm n2\pi/3 \mp n2\pi/3$. 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 the graphene sublattice basis as

$$\begin{aligned}\text{TDoS}_{n,m}(\omega_\text{bg},\omega_\text{tg}) &= \int \frac{d^2\boldsymbol{k}}{(2\pi)^2} \text{tr}\big[\boldsymbol{R}_{-\frac{2\pi}{3}}^n \cdot \boldsymbol{A}_{\text{bg},\boldsymbol{k}}(\omega_\text{bg}) \cdot \boldsymbol{R}_{\frac{2\pi}{3}}^n \\ &\quad \cdot \boldsymbol{J} \cdot \boldsymbol{R}_{-\frac{2\pi}{3}}^m \cdot \boldsymbol{A}_{\text{tg},\boldsymbol{k}+\mathcal{Q}_{n,m},T}(\omega_\text{tg}) \cdot \boldsymbol{R}_{\frac{2\pi}{3}}^m \cdot \boldsymbol{J}\big], \quad \text{(C1)}\end{aligned}$$

where tr$\{\ldots\}$ 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

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

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

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

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 $\omega_{\text{bg/tg}}^\pm = \omega_{\text{bg/tg}} \pm i0^+$. 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 $\omega_{\text{bg/tg}}^\pm \to \omega_{\text{bg/tg}} - \text{Re}\Sigma_{\text{bg/tg}}(\omega_{\text{bg/tg}}) \pm i\gamma_{\text{bg/tg}}(\omega_{\text{bg/tg}})$, where $\gamma_{\text{bg/tg}}(\omega_{\text{bg/tg}}) = -\text{Im}\Sigma_{\text{bg/tg}}^R(\omega_{\text{bg/tg}})$ 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 $\boldsymbol{R}_{\pm\frac{2\pi}{3}}^n$ perform a rotation of the electronic Green's functions, $\text{TDoS}_{n,m}(\omega)$ can be written as

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

Performing the trace over the sublattice degrees of freedom we get

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

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 $\gamma_{\rm bg} \to 0$, such that $\omega_{\rm bg}^{s'} \to \omega_{\rm bg} = \omega + \epsilon_{\rm F,bg}$ and

$$i\sum_{s=\pm1} s\frac{\omega_{\rm bg}^{s} + v_F\hbar\boldsymbol{k}\cdot\hat{\boldsymbol{K}}_{{\rm bg},n}}{(\omega_{\rm bg}^{s})^2 - (v_F\hbar)^2|\boldsymbol{k}|^2} \to 2\pi\frac{\omega_{\rm bg} + v_F\hbar\boldsymbol{k}\cdot\hat{\boldsymbol{K}}_{{\rm bg},n}}{2v_F\hbar|\boldsymbol{k}|}\sum_{s=\pm1} s\delta(\omega_{\rm bg} - sv_F\hbar|\boldsymbol{k}|) \tag{C6}$$

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

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

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

$$\cos\theta_{\boldsymbol{k}} = \frac{z+z^{-1}}{2}, \tag{C8}$$

$$\sin\theta_{\boldsymbol{k}} = \frac{z-z^{-1}}{2i}, \tag{C9}$$

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

$$\begin{aligned}\text{TDoS}_{n,m}(\omega_{\rm bg},\omega_{\rm tg}) \simeq\ & i\frac{\omega_{\rm bg}}{(v_F\hbar)^2}\oint_{|z|=1}\frac{dz}{2\pi i}\frac{1}{z}\left(\frac{\omega_{\rm bg} + v_F\hbar|\boldsymbol{k}|\big(\frac{z+z^{-1}}{2}\hat{K}^x_{{\rm bg},n} + \frac{z-z^{-1}}{2i}\hat{K}^y_{{\rm bg},n}\big)}{v_F\hbar|\boldsymbol{k}|}\right)\Bigg|_{|\boldsymbol{k}|=\frac{|\omega_{\rm bg}|}{v_F\hbar}}\\ &\times\sum_{s'=\pm1} s'\frac{2\big[\omega_{\rm tg}^{s'} + v_F\hbar|\boldsymbol{k}|\big(\frac{z+z^{-1}}{2}\hat{K}^x_{{\rm tg},m} + \frac{z-z^{-1}}{2i}\hat{K}^y_{{\rm tg},m}\big) + v_F\hbar\boldsymbol{\mathcal{Q}}_{n,m}\cdot\hat{\boldsymbol{K}}_{{\rm bg},m}\big]}{(\omega_{\rm tg}^{s'})^2 - |\boldsymbol{k}|^2 - |\boldsymbol{\mathcal{Q}}_{n,m}|^2 - 2v_F\hbar|\boldsymbol{k}||\boldsymbol{\mathcal{Q}}_{n,m}|\big(\frac{z+z^{-1}}{2}\cos\theta_{\boldsymbol{\mathcal{Q}}_{n,m}} + \frac{z-z^{-1}}{2i}\sin\theta_{\boldsymbol{\mathcal{Q}}_{n,m}}\big)},\end{aligned} \tag{C10}$$

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 $z = e^{i\theta_{\mathcal{Q}_{n,m}}}w^s_{\gtrless,{\rm tg}}$, with

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

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

$$\mathcal{S}^s_{\rm tg} = \text{sgn}\big[\omega_{\rm tg}^2 - \gamma_{\rm tg}^2 - (v_F\hbar)^2(|\boldsymbol{\mathcal{Q}}_{n,m}|^2 + |\boldsymbol{k}|^2)\big]i\sqrt{(\mathcal{C}^s_{\rm tg})^2 - 1}, \tag{C13}$$

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

$$\begin{aligned}\text{TDoS}_{n,m}(\omega_{\rm bg},\omega_{\rm tg}) \simeq\ & \frac{\omega_{\rm bg}}{(v_F\hbar)^3|\boldsymbol{\mathcal{Q}}_{n,m}|}\\ &\times\Bigg[\frac{-1}{\mathcal{S}^+_{\rm tg}}\left(\frac{\omega^+_{\rm tg} + |\omega_{\rm bg}|\big(\mathcal{C}^+_{\rm tg}X^{\rm tg}_{n,m} + \mathcal{S}^+_{\rm tg}Y^{\rm tg}_{n,m}\big) + v_F\hbar\boldsymbol{\mathcal{Q}}_{n,m}\cdot\hat{\boldsymbol{K}}_{{\rm tg},m}}{|\omega_{\rm bg}|}\right)\left(\frac{\omega_{\rm bg} + |\omega_{\rm bg}|\big(\mathcal{C}^+_{\rm tg}X^{\rm bg}_{n,m} + \mathcal{S}^+_{\rm tg}Y^{\rm bg}_{n,m}\big)}{|\omega_{\rm bg}|}\right)\\ &+\frac{1}{\mathcal{S}^-_{\rm tg}}\left(\frac{\omega^-_{\rm tg} + |\omega_{\rm bg}|\big(\mathcal{C}^-_{\rm tg}X^{\rm tg}_{n,m} + \mathcal{S}^-_{\rm tg}Y^{\rm tg}_{n,m}\big) + v_F\hbar\boldsymbol{\mathcal{Q}}_{n,m}\cdot\hat{\boldsymbol{K}}_{{\rm tg},m}}{|\omega_{\rm bg}|}\right)\left(\frac{\omega_{\rm bg} + |\omega_{\rm bg}|\big(\mathcal{C}^-_{\rm tg}X^{\rm bg}_{n,m} + \mathcal{S}^-_{\rm tg}Y^{\rm bg}_{n,m}\big)}{|\omega_{\rm bg}|}\right)\\ &+\frac{2\gamma_{\rm tg}\big(X^{\rm bg}_{n,m} + iY^{\rm bg}_{n,m}\big)}{v_F\hbar|\boldsymbol{\mathcal{Q}}_{n,m}|}\left(\frac{v_F\hbar|\boldsymbol{\mathcal{Q}}_{n,m}| + \omega_{\rm tg}\big(X^{\rm tg}_{n,m} + iY^{\rm tg}_{n,m}\big)}{|\omega_{\rm bg}|}\right)\Bigg]\Bigg|_{|\boldsymbol{k}|=|\omega_{\rm bg}|/(v_F\hbar)} + \frac{1}{2}\frac{\omega_{\rm tg}}{(v_F\hbar)^3|\boldsymbol{\mathcal{Q}}_{n,m}|}\\ &\times\Bigg[\frac{-1}{\mathcal{S}^+_{\rm bg}}\left(\frac{\omega^+_{\rm bg} - |\omega_{\rm tg}|\big(\mathcal{C}^+_{\rm bg}X^{\rm bg}_{n,m} + \mathcal{S}^+_{\rm bg}Y^{\rm bg}_{n,m}\big) - v_F\hbar\boldsymbol{\mathcal{Q}}_{n,m}\cdot\hat{\boldsymbol{K}}_{{\rm bg},n}}{|\omega_{\rm tg}|}\right)\left(\frac{\omega_{\rm tg} - |\omega_{\rm tg}|\big(\mathcal{C}^+_{\rm bg}X^{\rm tg}_{n,m} + \mathcal{S}^+_{\rm bg}Y^{\rm tg}_{n,m}\big)}{|\omega_{\rm tg}|}\right)\\ &+\frac{1}{\mathcal{S}^-_{\rm bg}}\left(\frac{\omega^-_{\rm bg} - |\omega_{\rm tg}|\big(\mathcal{C}^-_{\rm bg}X^{\rm bg}_{n,m} + \mathcal{S}^-_{\rm bg}Y^{\rm bg}_{n,m}\big) - v_F\hbar\boldsymbol{\mathcal{Q}}_{n,m}\cdot\hat{\boldsymbol{K}}_{{\rm bg},n}}{|\omega_{\rm tg}|}\right)\left(\frac{\omega_{\rm tg} - |\omega_{\rm tg}|\big(\mathcal{C}^-_{\rm bg}X^{\rm tg}_{n,m} + \mathcal{S}^-_{\rm bg}Y^{\rm tg}_{n,m}\big)}{|\omega_{\rm tg}|}\right)\end{aligned}$$

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

where we have introduced the quantities

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

and the quantities $\mathcal{C}^{\pm}_{\rm tg}$ and $\mathcal{S}^{\pm}_{\rm tg}$ given by Eqs. (C12) and (C13) with the replacements $\omega_{\rm tg}\to\omega_{\rm bg}$ and $\gamma_{\rm tg}\to\gamma_{\rm bg}$. It was checked that Eq. (C14) provides a very good approximation to the numeric evaluation of Eq. (C5) when both $\gamma_{\rm bg}$ and $\gamma_{\rm tg}$ are nonzero, if the broadening function for each layer is assumed to be the sum of the broadening factors of both layers, i.e., performing the replacement $\gamma_{\rm bg},\gamma_{\rm tg}\to\gamma_{\rm bg}+\gamma_{\rm tg}$.

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

$$\mathcal{C}^s_{\rm tg}=\frac{\omega^2_{\rm tg}-\omega^2_{\rm bg}-(v_F\hbar)^2|\mathcal{Q}_{n,m}|^2}{2(v_F\hbar)|\mathcal{Q}_{n,m}||\omega_{\rm bg}|}, \tag{C16}$$

$$\mathcal{S}^s_{\rm tg}=-s\,{\rm sgn}(\omega_{\rm tg})\sqrt{1-\left(\mathcal{C}^s_{\rm tg}\right)^2}, \tag{C17}$$

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

$$\begin{aligned} {\rm TDoS}_{n,m}(\omega_{\rm bg},\omega_{\rm tg})=&\frac{\omega_{\rm tg}}{(v_F\hbar)^3|\mathcal{Q}_{n,m}|}\\ &\times\Bigg[\frac{-1}{\mathcal{S}^+_{\rm tg}}\left(\frac{\omega_{\rm tg}+|\omega_{\rm bg}|\left(\mathcal{C}^+_{\rm tg}X^{\rm tg}_{n,m}+\mathcal{S}^+_{\rm tg}Y^{\rm tg}_{n,m}\right)+v_F\hbar\mathcal{Q}_{n,m}\cdot\hat{\boldsymbol{K}}_{{\rm tg},m}}{|\omega_{\rm bg}|}\right)\left(\frac{\omega_{\rm bg}+|\omega_{\rm bg}|\left(\mathcal{C}^+_{\rm tg}X^{\rm bg}_{n,m}+\mathcal{S}^+_{\rm tg}Y^{\rm bg}_{n,m}\right)}{|\omega_{\rm bg}|}\right)\\ &+\frac{1}{\mathcal{S}^-_{\rm tg}}\left(\frac{\omega_{\rm tg}+|\omega_{\rm bg}|\left(\mathcal{C}^-_{\rm tg}X^{\rm tg}_{n,m}+\mathcal{S}^-_{\rm tg}Y^{\rm tg}_{n,m}\right)+v_F\hbar\mathcal{Q}_{n,m}\cdot\hat{\boldsymbol{K}}_{{\rm tg},m}}{|\omega_{\rm bg}|}\right)\left(\frac{\omega_{\rm bg}+|\omega_{\rm bg}|\left(\mathcal{C}^-_{\rm tg}X^{\rm bg}_{n,m}+\mathcal{S}^-_{\rm tg}Y^{\rm bg}_{n,m}\right)}{|\omega_{\rm bg}|}\right)\Bigg]. \end{aligned} \tag{C18}$$

We notice that, in this limit, ${\rm TDoS}_{n,m}(\omega_{\rm bg},\omega_{\rm tg})$ is only nonzero when $4(v_F\hbar)^2|\mathcal{Q}_{n,m}|^2\omega^2_{\rm bg}>(\omega^2_{\rm tg}-\omega^2_{\rm bg}-(v_F\hbar)^2|\mathcal{Q}_{n,m}|^2)^2$.

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

$${\rm TDoS}^{\rm scalar}_{n,m}(\omega_{\rm bg},\omega_{\rm tg})=i^2\sum_{s,s'=\pm1}\int\frac{d^2\boldsymbol{k}}{(2\pi)^2}ss'\frac{2\omega^s_{\rm bg}}{\left(\omega^s_{\rm bg}\right)^2-(v_F\hbar)^2|\boldsymbol{k}|^2}\frac{2\omega^{s'}_{\rm tg}}{\left(\omega^{s'}_{\rm tg}\right)^2-(v_F\hbar)^2|\boldsymbol{k}+\mathcal{Q}_{n,m}|^2}. \tag{C19}$$

In order to evaluate ${\rm TDoS}^{\rm scalar}_{n,m}(\omega_{\rm bg},\omega_{\rm tg})$, 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 $z=e^{i\theta_{\pm\mathcal{Q}_{m,n}}}w^s_{<,{\rm tg/bg}}$. Symmetrizing the result, this leads to

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

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 these kinds 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 $\tau_{\rm tr}(\epsilon_F)\propto\epsilon_F$ [34]. Resonances due to short-range disorder cannot be taken into account by treating them 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 noncrossing approximation, the self-consistent Born approximation (SCBA) for the Green's function of an isolated graphene layer reads as

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

where matrices have indices in the sublattice space, a bar denotes disorder averaging, and $\boldsymbol{\Sigma}^R_{\rm imp}(\omega)=n_{\rm imp}\boldsymbol{T}^R(\omega)$ is the impurity self-energy, where $n_{\rm imp}$ 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

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

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

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

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

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

with a,b indices running over the A,B sublattice sites and γ_c is the lifetime induced by the metallic contacts $\gamma_c = \Gamma_{\mathrm{bg/tg}}/2$ (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 $\Sigma^R_{\mathrm{imp}}(\omega) = \Sigma_{\mathrm{imp}}(\omega) - i\gamma_{\mathrm{imp}}(\omega)$, we can evaluate $G_1^R(\omega)$ analytically, obtaining

$$\overline{G_1^R}(\omega) = \frac{g_1[\omega - \Sigma_{\mathrm{imp}}(\omega), \gamma_{\mathrm{imp}}(\omega) + \gamma_c]}{4\pi(v_F\hbar)^2} - i\frac{g_2[\omega - \Sigma_{\mathrm{imp}}(\omega), \gamma_{\mathrm{imp}}(\omega) + \gamma_c]}{4\pi(v_F\hbar)^2}, \tag{D5}$$

where the functions g_1 and g_2 are given by

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

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

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

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

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

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

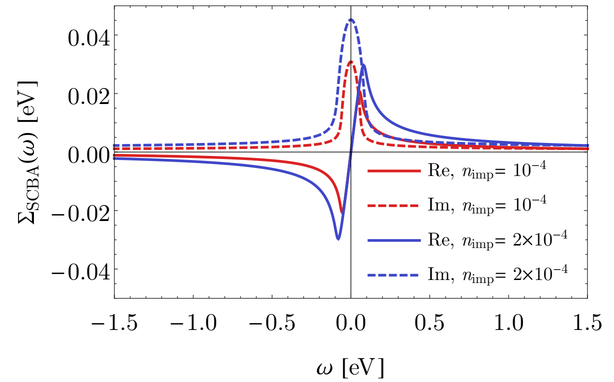

FIG. 13. Real and (minus) imaginary parts of the retarded self-energy for graphene electrons due to resonant impurities treated within the SCBA, for two different impurity concentrations (number of impurities per graphene unit cell).

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-neighbor tight-binding Hamiltonian for the electrons and assuming that the lattice distortions due to phonons lead to a modulation of hopping integrals [63]. 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

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

whereas

$$g^{\mathrm{g}}_{\Gamma\mathrm{O}} = \frac{3}{2}\left(-\frac{d\,\ln t}{d\,\ln a_{\mathrm{CC}}}\right)\frac{t}{a_{\mathrm{CC}}}\sqrt{\frac{\hbar}{\mu_{\mathrm{g}}\omega^{\mathrm{g}}_{\Gamma\mathrm{O}}}} \tag{E2}$$

is the electron-phonon coupling constant, with $-d\,\ln t/d\,\ln a_{\mathrm{CC}} \simeq 3$ [56,57] describing the change in the nearest-neighbor hopping t, with the distance $a_{\mathrm{CC}} = a_{\mathrm{g}}/\sqrt{3}$; $\mu_{\mathrm{g}} = m_{\mathrm{C}}/2$ is the reduced mass of the phonon mode, with m_{C} the carbon atom mass; and $\omega^{\mathrm{g}}_{\Gamma\mathrm{O}}$ is the phonon dispersion for the longitudinal/transverse in-plane optical phonon mode (which are degenerate at Γ and assume we approximate them as dispersionless). The polarization vectors for the longitudinal and transverse modes can be written as $\vec{\epsilon}_{\boldsymbol{q},\mathrm{LO}} = (1,0)$ and $\vec{\epsilon}_{\boldsymbol{q},\mathrm{TO}} = (0,1)$. With these approximations we obtain the momentum-independent electron-phonon interaction matrices

$$\boldsymbol{M}^{\mathrm{g}}_{\mathrm{LO}} = -g^{\mathrm{g}}_{\Gamma\mathrm{O}}\sigma_y, \tag{E3}$$

$$\boldsymbol{M}^{\mathrm{g}}_{\mathrm{TO}} = g^{\mathrm{g}}_{\Gamma\mathrm{O}}\sigma_x. \tag{E4}$$

Assuming the graphene layer is in thermal equilibrium and to lowest order in the electron-phonon interaction, the self-energy

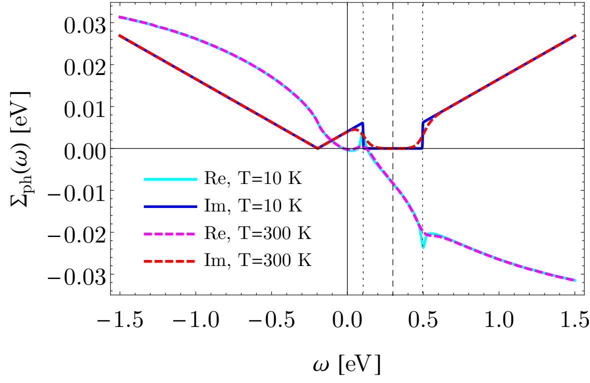

FIG. 14. Real and (minus) imaginary parts of the self-energy for graphene electrons due to scattering by in-plane optical phonons for two different temperatures for doped graphene with $\epsilon_{\rm F} = 0.3$ eV. The zero of energy corresponds to the Dirac point. The dashed vertical line marks $\omega = \epsilon_{\rm F}$ and the dotted lines mark $\omega = \epsilon_{\rm F} \pm \omega^{\rm g}_{\Gamma{\rm O}}$.

is diagonal in sublattice space and given by

$$\Sigma^R_{\rm ph}(\omega) = A_{\rm cell}\big(g^{\rm g}_{\Gamma{\rm O}}\big)^2 \times \sum_{\lambda,s}\int \frac{d^2\boldsymbol{q}}{(2\pi)^2} s\, \frac{1 + b\big(s\omega^{\rm g}_{\Gamma{\rm O}}\big) - f(\epsilon_{\boldsymbol{q},\lambda} - \epsilon_{\rm F})}{\omega - \epsilon_{\boldsymbol{q},\lambda} - s\omega^{\rm g}_{\Gamma{\rm O}} + i0^+}. \tag{E5}$$

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

$$-\,{\rm Im}\Sigma^R_{\rm ph}(\omega) = \big(g^{\rm g}_{\Gamma{\rm O}}\big)^2\big[1 + b\big(\omega^{\rm g}_{\Gamma{\rm O}}\big) - f\big(\omega - \omega^{\rm g}_{\Gamma{\rm O}} - \epsilon_{\rm F}\big)\big] \times \frac{A_{\rm cell}\big|\omega - \omega^{\rm g}_{\Gamma{\rm O}}\big|}{2(v_F\hbar)^2} + \big(g^{\rm g}_{\Gamma{\rm O}}\big)^2\big[b\big(\omega^{\rm g}_{\Gamma{\rm O}}\big) - f\big(\omega + \omega^{\rm g}_{\Gamma{\rm O}} - \epsilon_{\rm F}\big)\big]\frac{A_{\rm cell}\big|\omega + \omega^{\rm g}_{\Gamma{\rm O}}\big|}{2(v_F\hbar)^2}, \tag{E6}$$

where ω and the Fermi energy $\epsilon_{\rm F}$ are both measured from the Dirac cone. From this, the real part can be efficiently obtained using the Kramers-Kronig relation

$${\rm Re}\Sigma^R_{\rm ph}(\omega) = -\int \frac{d\nu}{\pi}\,\frac{{\rm Im}\Sigma^R_{\rm ph}(\omega - \nu) - {\rm Im}\Sigma^R_{\rm ph}(\omega + \nu)}{\nu}. \tag{E7}$$

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/h-BN/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

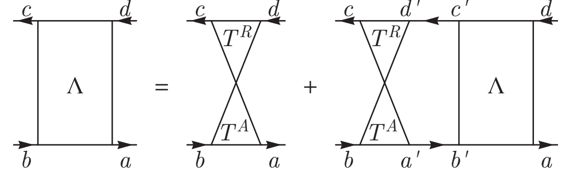

FIG. 15. Diagrammatic representation of the Bethe-Salpeter equation (F6).

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 described by Bloch states. With these approximations, we write

$$\boldsymbol{\Gamma}_{\rm b/t} = \Gamma_{\rm b/t}\boldsymbol{I}_{\rm bg/tg}. \tag{F1}$$

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/h-BN coupling, we obtain

$$\overline{\mathcal{T}} = \Gamma_{\rm b}\Gamma_{\rm t}{\rm Tr}\big[\overline{\boldsymbol{G}^{0,A}_{\rm bg}\cdot\boldsymbol{I}_{\rm bg}\cdot\boldsymbol{G}^{0,R}_{\rm bg}}\cdot\boldsymbol{\mathcal{T}}_{\rm bg,tg} \tag{F2}$$

$$\cdot\,\overline{\boldsymbol{G}^{0,R}_{\rm tg}\cdot\boldsymbol{I}_{\rm tg}\cdot\boldsymbol{G}^{0,A}_{\rm tg}}\cdot\boldsymbol{\mathcal{T}}_{\rm tg,bg}\big]. \tag{F3}$$

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 employ a notation where an upper indice represents an outgoing 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 indices)

$$\overline{\big[G^{0,A}_{\boldsymbol{k}}\big]^a{}_b\delta^b{}_c\big[G^{R,0}_{\boldsymbol{k}}\big]^c{}_d} = \big[\overline{G^{0,A}_{\boldsymbol{k}}}\big]^a{}_b\delta^b{}_c\big[\overline{G^{R,0}_{\boldsymbol{k}}}\big]^c{}_d + \big[\overline{G^{0,A}_{\boldsymbol{k}}}\big]^a{}_{a'}\Lambda^{a'}{}_{b'}{}^{c'}{}_{d'}\big[G^{AR}_2\big]^{b'}{}_b{}^c{}_{c'}\delta^b{}_c \times \big[\overline{G^{0,R}_{\boldsymbol{k}}}\big]^{d'}{}_d, \tag{F4}$$

where the second terms are vertex corrections, we define the quantity

$$\big[G^{AR}_2\big]^a{}_b{}^c{}_d = \int \frac{d^2\boldsymbol{p}}{(2\pi)^2}\big[\overline{G^A_{\boldsymbol{p}}}\big]^a{}_b\big[\overline{G^R_{\boldsymbol{p}}}\big]^c{}_d, \tag{F5}$$

and $\Lambda^a{}_b{}^c{}_d$ is a four-point function, which obeys a Bethe-Salpeter equation (see Fig. 15)

$$\Lambda^a{}_b{}^c{}_d = U^a{}_b{}^c{}_d + U^{a'}{}_b{}^c{}_{d'}\big[G^{AR}_2\big]^{b'}{}_{a'}{}^{d'}{}_{c'}\Lambda^a{}_{b'}{}^{c'}{}_d, \tag{F6}$$

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

$$U^a{}_b{}^c{}_d = n_{\rm imp}\big|T^R_{\rm imp}\big|^2\delta^a{}_b\delta^c{}_d. \tag{F7}$$

The quantity $\big[G^{AR}_2\big]^a{}_b{}^c{}_d$ can be evaluated analytically yielding

$$\big[G^{AR}_2\big]^a{}_b{}^c{}_d = L_1(\omega',\gamma')\delta^a{}_b\delta^c{}_d + L_2(\omega',\gamma')\tfrac{1}{2}\boldsymbol{\sigma}^a{}_b\cdot\boldsymbol{\sigma}^c{}_d, \tag{F8}$$

where

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

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

with the functions g_1 and g_2 defined by Eqs. (D6) and (D7) and where we have written $\omega' = \omega - \Sigma_{\text{imp}}$ and $\gamma' = \gamma_{\text{imp}} + \gamma_c$ 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 nonzero components of $\Lambda^{a}{}_{b}{}^{c}{}_{d}$ in the sublattice basis

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

$$\Lambda^{A}{}_{A}{}^{B}{}_{B} = \Lambda^{B}{}_{B}{}^{A}{}_{A} = \frac{n_{\text{imp}}|T^R|^2}{1-L_1 n_{\text{imp}}|T^R|^2}, \quad \text{(F12)}$$

$$\Lambda^{A}{}_{B}{}^{B}{}_{A} = \Lambda^{B}{}_{A}{}^{A}{}_{B} = \frac{L_2 n_{\text{imp}}^2|T^R|^4}{[1-(L_1-L_2)n_{\text{imp}}|T^R|^2][1-(L_1+L_2)n_{\text{imp}}|T^R|^2]}, \quad \text{(F13)}$$

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

$$\Lambda^{a'}{}_{b'}{}^{c'}{}_{d'}[G_2^{AR}]^{b'}{}_{b}{}^{c}{}_{c'} = \frac{n_{\text{imp}}|T^R|^2(L_1+L_2)}{1-(L_1+L_2)n_{\text{imp}}|T^R|^2}\delta_b^c. \quad \text{(F14)}$$

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 $(L_1+L_2)n_{\text{imp}}|T^R|^2$ can be written as the ratio

$$(L_1+L_2)n_{\text{imp}}|T^R|^2 = \frac{\gamma_{\text{imp}}}{\gamma_{\text{imp}}+\gamma_c}. \quad \text{(F15)}$$

Therefore, Eq. (F4) can be written as

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

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

$$\overline{[G_{\mathbf{k}}^{0,A}(\omega)]^a{}_b[G_{\mathbf{k}}^{R,0}(\omega)]^b{}_d} = \frac{1}{\gamma_{\text{imp}}+\gamma_c}[\overline{A_{\mathbf{k}}^0}(\omega)]^a{}_b \quad \text{(F17)}$$

and, therefore, the contributions from vertex corrections (incoherent contributions) due to impurities add 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.

[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).

[2] K. S. Novoselov, D. Jiang, F. Schedin, T. J. Booth, V. V. Khotkevich, S. V. Morozov, and A. K. Geim, Proc. Natl. Acad. Sci. USA **102**, 10451 (2005).

[3] L. A. Ponomarenko, A. K. Geim, A. A. Zhukov, R. Jalil, S. V. Morozov, K. S. Novoselov, I. V. Grigorieva, E. H. Hill, V. V. Cheianov, V. I. Fal'ko, K. Watanabe, T. Taniguchi, and R. V. Gorbachev, Nat. Phys. **7**, 958 (2011).

[4] K. S. Novoselov and A. H. C. Neto, Phys. Scr. **T146**, 014006 (2012).

[5] A. K. Geim and I. V. Grigorieva, Nature (London) **499**, 419 (2013).

[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).

[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 Lett. **12**, 1707 (2012).

[8] T. Georgiou, R. Jalil, D. Belle Branson, L. Britnell, R. V. Gorbachev, S. V. Morozov, Y.-J. Kim, A. Gholinia, S. J. Haigh, O. Makarovsky, L. Eaves, L. A. Ponomarenko, A. K. Geim, K. S. Novoselov, and A. Mishchenko, Nat. Nanotechnol. **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).

[10] Yu Woo Jong, Liu Yuan, Zhou Hailong, Yin Anxiang, Li Zheng, Huang Yu, and Duan Xiangfeng, Nat. Nanotechnol. **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. Nanotechnol. **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, J. Phys. D: Appl. Phys. **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 (London) **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. Nanotechnol. **9**, 808 (2014).

[24] L. Brey, Phys. Rev. Appl. **2**, 014003 (2014).

[25] P. Zhao, R. M. Feenstra, G. Gu, and D. Jena, IEEE Trans. Electron Devices **60**, 951 (2013).

[26] J. Gaskell, L. Eaves, K. S. Novoselov, A. Mishchenko, A. K. Geim, T. M. Fromhold, and M. T. Greenaway, Appl. Phys. Lett. **107**, 103105 (2015).

[27] 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 Lett. **15**, 428 (2015).

[28] 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, IEEE Electron Device Lett. **36**, 405 (2015).

[29] S. C. de la Barrera and R. M. Feenstra, Appl. Phys. Lett. **106**, 093115 (2015).

[30] T. L. M. Lane, J. R. Wallbank, and V. I. Fal'ko, Appl. Phys. Lett. **107**, 203506 (2015).

[31] V. Perebeinos, J. Tersoff, and P. Avouris, Phys. Rev. Lett. **109**, 236604 (2012).

[32] 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, Sci. Rep. **5**, 16642 (2015).

[33] E. E. Vdovin, A. Mishchenko, M. T. Greenaway, M. J. Zhu, D. Ghazaryan, A. Misra, Y. Cao, S. V. Morozov, O. Makarovsky, T. M. Fromhold, A. Patanè, G. J. Slotman, M. I. Katsnelson, A. K. Geim, K. S. Novoselov, and L. Eaves, Phys. Rev. Lett. **116**, 186603 (2016).

[34] N. M. R. Peres, Rev. Mod. Phys. **82**, 2673 (2010).

[35] R. Geick, C. H. Perry, and G. Rupprecht, Phys. Rev. **146**, 543 (1966).

[36] R. M. Ribeiro and N. M. R. Peres, Phys. Rev. B **83**, 235312 (2011).

[37] M. Kindermann, B. Uchoa, and D. L. Miller, Phys. Rev. B **86**, 115415 (2012).

[38] J. Jung, A. Raoux, Z. Qiao, and A. H. MacDonald, Phys. Rev. B **89**, 205414 (2014).

[39] M. Koshino, New J. Phys. **17**, 015014 (2015).

[40] R. Bistritzer and A. H. MacDonald, Proc. Natl. Acad. Sci. USA **108**, 12233 (2011).

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

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

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

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

[45] P. Vogl and T. Kubis, J. Comput. Electron. **9**, 237 (2010).

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

[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, Berlin, 2008).

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

[49] R. M. Feenstra, D. Jena, and G. Gu, J. Appl. Phys. **111**, 043711 (2012).

[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, Semicond. Sci. Technol. **6**, 196 (1991).

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

[55] C. Caroli, R. Combescot, P. Nozieres, and D. Saint-James, J. Phys. C: Solid State Phys. **5**, 21 (1972).

[56] J. M. Wills and W. A. Harrison, Phys. Rev. B **28**, 4363 (1983).

[57] W. Harrison, *Elementary Electronic Structure*, Elementary Electronic Structure (World Scientific, Singapore, 1999), Chap. 16.

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

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

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

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

[62] 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).

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