Quantum coherence and its dephasing in the giant spin Hall effect and nonlocal voltage generated by magnetotransport through multiterminal graphene bars

Chien-Liang Chen,[1,2] Ching-Ray Chang,[2,*] and Branislav K. Nikolić[1,†]
[1]*Department of Physics and Astronomy, University of Delaware, Newark, Delaware 19716-2570, USA*
[2]*Department of Physics, National Taiwan University, Taipei 10617, Taiwan*
(Received 29 November 2011; revised manuscript received 16 January 2012; published 9 April 2012)

Motivated by the recent experimental observation [Abanin *et al.*, Science **332**, 328 (2011)] of nonlocality in magnetotransport near the Dirac point in six-terminal graphene Hall bars, for a wide range of temperatures and magnetic fields, we develop a nonequilibrium Green's function theory of this phenomenon. In the quantum-coherent regime and strong magnetic field, we find large Zeeman-splitting-driven spin Hall (SH) conductance in four-terminal bars, where the SH current is pure only at the Dirac point (DP). In six-terminal Hall bars, this leads to the nonlocal voltage at a remote location due to direct and inverse SH effect operating at the same time in different parts of the device. The "momentum-relaxing" dephasing reduces their values at the DP by two orders of magnitude while concurrently washing out any features away from the DP. Our theory is based on the Meir-Wingreen formula for spin-resolved charge currents with dephasing introduced via phenomenological many-body self-energies, which is then linearized for multiterminal geometries to extract currents and voltages. This provides a generalization of the multiprobe Landauer-Büttiker formula without employing traditional Büttiker voltage probes to introduce dephasing.

DOI: 10.1103/PhysRevB.85.155414 PACS number(s): 72.80.Vp, 72.15.Gd, 72.10.Bg

I. INTRODUCTION

The recent experiments[1] on magnetotransport near the Dirac point (DP) in graphene have unveiled yet another exotic electronic property of this one-atom-thick carbon allotrope[2] which involves nonlocality and quantum mechanics while manifesting even at room temperature. The traditional observation of nonlocality, where current is injected through a pair of terminals and voltage is measured between another pair of terminals at some *remote* location, requires two-dimensional (2D) systems placed in a strong external magnetic field to generate the integer quantum Hall effect[3] (QHE) or spin-orbit coupling[4] (SOC) that can give rise to mesoscopic[5,6] or quantum[3,7] spin Hall effects (SHEs). In the former case, nonlocality is due to transport through chiral edge states. In the latter case, injected longitudinal charge current generates transverse spin Hall current, which is then detected in the remote part of the device via the inverse-SHE-induced voltages[6,8] on the proviso that spins can survive dephasing between two locations.

On the other hand, nonlocal voltage was observed in Ref. 1 even in weak magnetic fields $B \simeq 1$ T and at room temperature $T = 300$ K, which is outside of the integer QHE regime. Also, high-mobility graphene samples were supported by substrate made of atomically flat hexagonal boron nitride, which rules out Rashba SOC,[4] introduced by charge impurities from the substrate[9] or lattice distortion by adatoms,[10] which would be responsible for the mesoscopic SHE scenario.[5]

It turns out that SHE in the absence of SOC has a simple intuitive explanation based on the classical Newtonian dynamics of massless Dirac fermions. The classical Hamiltonian of low-energy quasiparticles close to the Dirac point is given by $H^{\pm}(\mathbf{p}) = \pm v_F\sqrt{p_x^2 + p_y^2}$, which in the weak external magnetic field $\mathbf{B} = \nabla \times \mathbf{A}$ becomes $H^{\pm}(\mathbf{p}) = \pm v_F\sqrt{(p_x - eA_x)^2 + (p_y - eA_y)^2}$. The classical velocity is then given by $v^{\pm}_{x,y} = \partial H^{\pm}/\partial p_{x,y} = \pm v_F \Pi_{x,y}/\sqrt{\mathbf{\Pi}^2}$, where $\mathbf{\Pi} = \mathbf{p} - e\mathbf{A}$, and the corresponding acceleration is

$$\mathbf{a}^{\pm} = \frac{d\mathbf{v}}{dt} = \pm\frac{ev_F\mathbf{v}^{\pm} \times \mathbf{B}}{\sqrt{\mathbf{\Pi}^2}} = \frac{ev_F^2\mathbf{v}^{\pm} \times \mathbf{B}}{E^{\pm}}. \quad (1)$$

Thus, the quasiparticles with energy E^+ above the DP (or below with energy E^-) moving in a weak (i.e., nonquantizing) perpendicular magnetic field will experience a transverse force which deflects them to the left (right). Furthermore, when $E^{\pm}$ is very close to the DP, such deflecting force will be very large.

Although the Zeeman splitting Δ_Z in 2D electron gases (2DEGs) is typically small in a weak external magnetic field,[4] it can play an essential role in graphene for $k_BT < \Delta_Z$ by shifting the Dirac cones for opposite spins to induce two types of carriers, illustrated in the lower inset in Fig. 1. The quasiparticles with energy E^+ are spin-up polarized, while those with energy E^- are spin-down polarized. These two effects, classical for charge and quantum for spin, conspire to generate transverse spin current in response to longitudinal charge current, as illustrated in Fig. 1. Such phenomenology is similar to SHE in multiterminal 2DEG devices,[5,6] even though no SOC is involved to provide the deflecting force of opposite direction for opposite spins.[11]

These simple arguments for the existence of the Zeeman-splitting-driven SHE (ZSHE) in graphene can be converted into a quasiclassical transport theory based on the Boltzmann equation.[16] However, quasiclassical theory is valid in the high-T and weak-B regime, while experiments[1] have observed increasingly more profound nonlocality in the low-T and/or high-B regime, so that a unified theory is called for that can cover such wide range of parameters. For example, such theory should explain the nonlocal voltage in a strong (quantizing) external magnetic field but at intermediate temperatures where the edge-state transport mechanism is removed.

Here we develop a *fully* quantum transport theory of ZSHE in four-terminal graphene bars, illustrated by the device within the dashed box in Fig. 1, as well as the nonlocal voltage

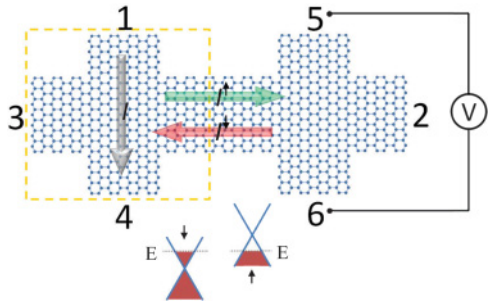

FIG. 1. (Color online) Schematic view of the six-terminal graphene Hall bar, modeled on the tight-binding lattice with a single π orbital per site, which is employed to investigate nonlocal voltage between leads 5 and 6 due to current injected between leads 1 and 4. The dashed box on the left marks the four-terminal bar used in the analysis of Zeeman-splitting-driven SHE, where current injected between leads 1 and 4 induces spin current in leads 2 and 3. The active region of the bar consists of a graphene nanoribbon with armchair edges and a portion of semi-infinite leads modeled as GNRs with zigzag edges. For simplicity, external magnetic field or many-body interactions responsible for dephasing are present only within the illustrated active region.

induced by the combination of direct and inverse ZSHE in six-terminal Hall bars shown in Fig. 1. Our approach is based on the nonequilibrium Green's function formalism[12] (NEGF), which allows one to express spin-resolved charge currents in the device terminals via the Meir-Wingreen formula.[12,13] The dephasing is introduced via phenomenological[14] many-body self-energies which take into account simultaneous phase and momentum relaxation. We then linearize the Meir-Wingreen formula to establish the connection between currents and voltages in the terminals, thereby offering a generalization of the Landauer-Büttiker formula[15] for phase-coherent transport in multiterminal devices to situations where dephasing is present within the active region of the device.

Our theory intrinsically accounts for the contributions of both electrons and holes, which is crucial to describe transport near the DP,[16] and it can also handle arbitrary scattering processes (in contrast to the Boltzmann equation, which breaks down[18] close to the DP). Our central results, summarized in Figs. 2–4, interpolate smoothly between the phase-coherent transport regime at low-T in the quantizing external magnetic field and the semiclassical transport regime where dephasing by many-body interactions destroys features found at low-T while leaving peaks (of reduced magnitude, however) in the SH conductance and nonlocal voltage around the DP, in accord with experimental observations.[1]

This paper is organized as follows. The model Hamiltonian for the multiterminal graphene Hall bars in Fig. 1 in the external magnetic field is introduced in Sec. II. In Sec. III, we discuss how to introduce dephasing via phenomenological many-body self-energies of NEGF formalism and extract currents and voltages in the device terminals in the linear-response transport regime. This approach is applied in Sec. IV to ZSHE in the four- and six-terminal graphene Hall bars from Fig. 1. We conclude in Sec. V.

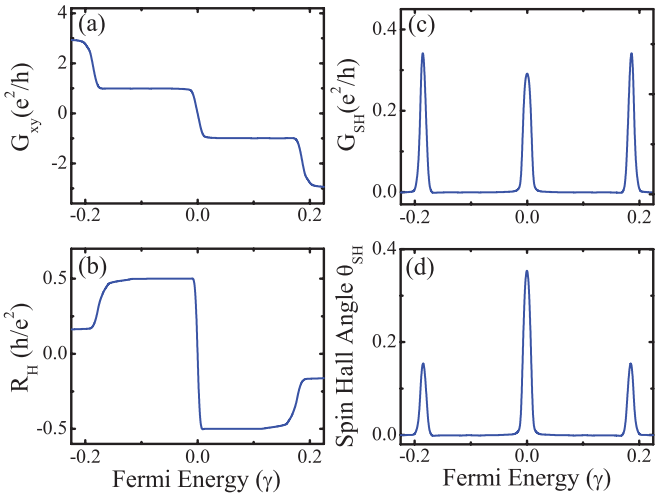

FIG. 2. (Color online) The charge and spin transport quantities in the *four-terminal* graphene Hall bar: (a) charge Hall conductance $G_{xy} = I_2/(V_1 - V_4)$, (b) charge Hall resistance $R_H = (V_3 - V_2)/I_1$, (c) spin Hall conductance $G_{SH} = I_2^S/(V_1 - V_4)$, and (d) spin Hall angle $\theta_{SH} = I_2^S/I_1$. The width of AGNR channel is $W/\ell_B = 3.42$ in the units of the magnetic length ℓ_B, and a small momentum-relaxing dephasing $d_m = 0.04\gamma$ is introduced into the active region shown in Fig. 1.

II. THE DEVICE HAMILTONIAN

Close to the DP, graphene can be described by the tight-binding Hamiltonian with a single π orbital per site,

$$\hat{H} = \sum_{\mathbf{n}} (\varepsilon_{\mathbf{n}} + g\mu_B \sigma B)\hat{c}^\dagger_{\mathbf{n}\sigma}\hat{c}_{\mathbf{n}\sigma} - \gamma \sum_{\langle \mathbf{nm}\rangle,\sigma} e^{i\phi_{\mathbf{nm}}}\hat{c}^\dagger_{\mathbf{n}\sigma}\hat{c}_{\mathbf{m}\sigma}. \quad (2)$$

Here $\varepsilon_{\mathbf{n}}$ is the on-site energy, $\sigma = +1$ for spin-up electrons and $\sigma = -1$ for spin-down electrons, so that Zeeman splitting

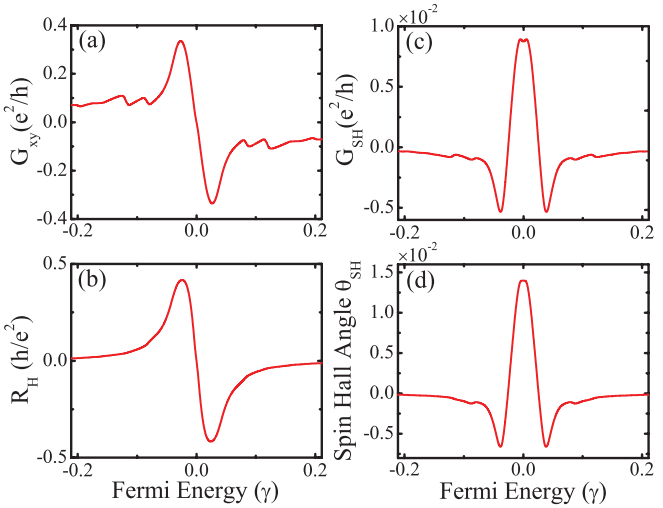

FIG. 3. (Color online) The charge and spin transport quantities in the *four-terminal* graphene Hall bar: (a) charge Hall conductance G_{xy}, (b) charge Hall resistance R_H, (c) spin Hall conductance G_{SH}, and (d) spin Hall angle θ_{SH}. The width of the AGNR channel is $W/\ell_B = 1.53$ in the units of the magnetic length ℓ_B, and large momentum-relaxing dephasing $d_m = 0.4\gamma$ is introduced into the active region shown in Fig. 1.

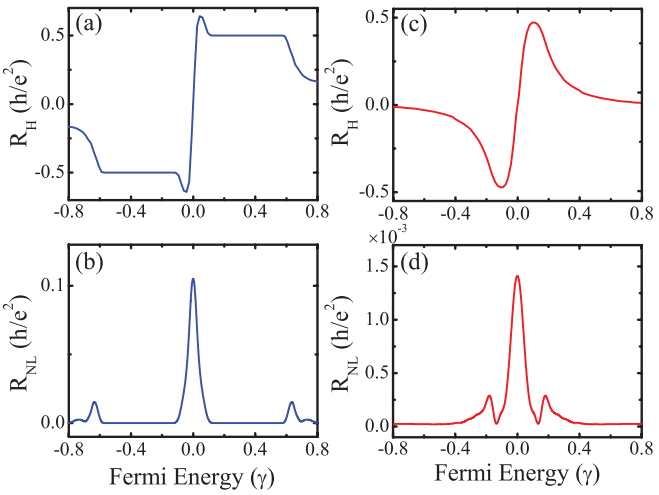

FIG. 4. (Color online) (a) and (c) Charge Hall resistance $R_H = (V_1 - V_4)/I_3$ and (b) and (d) nonlocal resistance $R_{NL} = (V_5 - V_6)/I_1$ as the central quantity measured in the recent experiments[1] on *six-terminal* graphene Hall bars. The quantum coherence is retained in (a) and (b), where only a small momentum-relaxing dephasing $d_m = 0.02\gamma$ is present in the active region of the bar, while much larger dephasing $d_m = 0.5\gamma$ is used for (c) and (d). The width of the AGNR channel in Fig. 1 is $W/\ell_B = 3.42$ in (a) and (b) and $W/\ell_B = 1.53$ in (c) and (d), in the units of the magnetic length ℓ_B.

is given by $\Delta_Z = 2g\mu_B B$ with $g = 2.0$, $\hat{c}^\dagger_{\mathbf{n}\sigma}$ ($\hat{c}_{\mathbf{n}\sigma}$) creates (annihilates) electrons with spin σ in the π orbital located on site $\mathbf{n}$, and γ is the nearest-neighbor hopping parameter.

The active region of the Hall bar in Fig. 1 consists of a graphene nanoribbon with armchair edges (AGNR) and a portion of semi-infinite ideal leads modeled as graphene nanoribbons with zigzag edges (ZGNR). The electronic structure and density of states of AGNR composed of $3p + 2$ dimer lines resembles[17] (if we assume that only the nearest-neighbor hopping γ is nonzero) those of large-area graphene employed experimentally. Although ZGNRs are insulating at very low temperatures due to one-dimensional spin-polarized edge states coupled across the width of the nanoribbon, such unusual magnetic ordering and the corresponding band gap is easily destroyed[19,20] above $T \gtrsim 10$ K, so we employ them as a model for metallic leads.

The external magnetic field enters through the phase factor $\phi_{\mathbf{nm}} = \frac{2\pi}{\phi_0}\int_{\mathbf{n}}^{\mathbf{m}} \mathbf{A}\cdot d\mathbf{s}$, where the vector potential $\mathbf{A} = (By, 0, 0)$ is chosen in the Landau gauge and $\phi_0 = h/e$ is the flux quantum. The weak vs. strong magnetic field is tuned using the ratio W/ℓ_B, where W is the width of the AGNR channel of the bar in Fig. 1. All graphene bars studied in Figs. 2–4 are placed in the quantizing external magnetic field, $W/\ell_B > 1$.

III. NEGF-BASED MODEL FOR DEPHASING IN QUANTUM TRANSPORT THROUGH MULTITERMINAL DEVICES

While the integer QHE and quantum SHE have introduced the intricate physics of Hall conductivity viewed as a topological invariant,[3] the operational description of these effects used to analyze transport measurements[6,7] is typically based on the multiprobe Landauer-Büttiker formula[15]

$$I_\alpha = \frac{e^2}{h}\sum_\beta T_{\alpha\beta}(V_\alpha - V_\beta), \tag{3}$$

written here assuming zero temperature. They relate total charge current I_α in lead α to voltages V_β in all other leads via the matrix of transmission coefficients $T_{\alpha\beta}$. These formulas are valid when phase coherence is retained in the active region of the device, while phase breaking events are assumed to be taking place only in the reservoirs to which the leads are attached at infinity and where electrons are equilibrated to acquire the Fermi-Dirac distribution $f_\alpha(E) = f(E - eV_\alpha)$.

To take into account dephasing effects phenomenologically, Büttiker introduced[21] an elegant concept of voltage probes attached to the active region where no net current flows through them. Thus, for every electron that enters the probe and is absorbed by its reservoir another one has to come out which is not coherent with the one going in. For example, to apply this method to the graphene Hall bar in Fig. 1, one has to attach one-dimensional leads to each site[22] of the honeycomb lattice. This is equivalent to adding a complex energy $-i\eta$ to $\varepsilon_{\mathbf{n}}$ in the Hamiltonian Eq. (2) (parameter η is related to the dephasing time $\eta = \hbar/2\tau_\phi$). In addition, one has to solve Eq. (3) by imposing that current through extra 1D leads is zero, thereby ensuring conservation of the total charge current.[22]

However, besides washing out quantum-coherence-generated fluctuations in $T_{\alpha\beta}$, Büttiker voltage probes are also introducing additional scattering (i.e., reduction of $T_{\alpha\beta}$) in an uncontrolled fashion.[14] The NEGF formalism[12] provides a rigorous prescription for including any dephasing process to any order by starting from a microscopic Hamiltonian and by constructing interacting self-energies due to electron-electron,[23] electron-phonon,[24] or electron-spin[25] interactions. Although the NEGF formalism is virtually the only fully quantitative quantum transport approach capable of scaling to large systems,[26] the self-consistent computation of self-energies by starting from some microscopic many-body Hamiltonian is at present prohibitively expensive for devices containing a realistic number of atoms. Thus, to include dephasing processes in the device in Fig. 1 containing a few thousand carbon atoms, we adopt a phenomenological model of Ref. 14 that is comparable to Büttiker voltage probes in conceptual and numerical simplicity and yet allows one the flexibility of adjusting the degree of phase and momentum relaxation independently.

The two fundamental objects[12] of the NEGF formalism are the retarded $G^{r,\sigma\sigma'}_{\mathbf{nm}}(t,t') = -i\Theta(t-t')\langle\{\hat{c}_{\mathbf{n}\sigma}(t), \hat{c}^\dagger_{\mathbf{m}\sigma'}(t')\}\rangle$ and the lesser $G^{<,\sigma\sigma'}_{\mathbf{nm}}(t,t') = i\langle\hat{c}^\dagger_{\mathbf{n}\sigma'}(t')\hat{c}_{\mathbf{m}\sigma}(t)\rangle$ GFs, which describe the density of available quantum states and how electrons occupy those states, respectively. Here $\langle\cdots\rangle$ denotes the nonequilibrium statistical average.[12] In stationary problems, $\hat{G}^r$ and $\hat{G}^<$ depend only on the time difference $t - t'$ or energy E after Fourier transformation. Their matrix representations in the basis of local orbitals, such as the π ones in Eq. (2) which define the matrix representation $\mathbf{H}$ of the corresponding

Hamiltonian, satisfy the following equations:

$$\mathbf{G}^r(E) = \left[E - \mathbf{H} - \sum_\alpha \mathbf{\Sigma}^r_\alpha(E - eV_\alpha) - \mathbf{\Sigma}^r_{\rm int}(E)\right]^{-1}, \tag{4}$$

$$\mathbf{G}^<(E) = \mathbf{G}^r(E)\left[\sum_\alpha \mathbf{\Sigma}^<_\alpha(E) + \mathbf{\Sigma}^<_{\rm int}(E)\right]\mathbf{G}^a(E). \tag{5}$$

Here $\mathbf{\Sigma}^r_\alpha(E)$ is the retarded self-energy determining the escape rates for electrons to exit into the attached leads, $\mathbf{\Sigma}^<_\alpha(E) = -f_\alpha[\mathbf{\Sigma}^r_\alpha(E - eV_\alpha) - \mathbf{\Sigma}^a_\alpha(E - eV_\alpha)]$ is the corresponding lesser self-energy matrix due to the coupling to the leads, and advanced quantities are defined by $\mathbf{O}^a = [\mathbf{O}^r]^\dagger$. In the "momentum-conserving" model of dephasing, the interacting self-energies are given[14] by

$$\mathbf{\Sigma}^r_{\rm int}(E) = d_p\mathbf{G}^r(E), \tag{6}$$

$$\mathbf{\Sigma}^<_{\rm int}(E) = d_p\mathbf{G}^<(E), \tag{7}$$

while in the "momentum-relaxing" model

$$\mathbf{\Sigma}^r_{\rm int}(E) = \mathcal{D}[d_m\mathbf{G}^r(E)], \tag{8}$$

$$\mathbf{\Sigma}^<_{\rm int}(E) = \mathcal{D}[d_m\mathbf{G}^<(E)]. \tag{9}$$

The operator $\mathcal{D}[\cdots]$ selects the diagonal elements of the matrix on which it acts while setting to zero all the off-diagonal elements. Any linear combination of these two choices can be used to adjust the phase and momentum relaxation lengths independently. When computed self-consistently together with $\mathbf{G}^r(E)$ and $\mathbf{G}^<(E)$, both of these choices for $\mathbf{\Sigma}^r_{\rm int}(E)$ and $\mathbf{\Sigma}^<_{\rm int}(E)$ ensure the conservation[23] of charge current, $\sum_\alpha I_\alpha = 0$.

The spin-resolved charge current in lead α is given by the Meir-Wingreen formula,[12,13]

$$I^\sigma_\alpha = \frac{e}{h}\int dE\,{\rm Tr}\left[\mathbf{\Sigma}^{<,\sigma\sigma}_\alpha(E)\mathbf{G}^{>,\sigma\sigma}(E) - \mathbf{\Sigma}^{>,\sigma\sigma}_\alpha(E)\mathbf{G}^{<,\sigma\sigma}(E)\right], \tag{10}$$

which assumes that interactions responsible for dephasing are localized within the active region of the graphene Hall bar (a generalization to situations where interactions are spread throughout the system, including the leads and the lead-sample interfaces, can be found in Ref. 27). The total charge current in lead α is $I_\alpha = I^\uparrow_\alpha + I^\downarrow_\alpha$, and the total spin current is $I^S_\alpha = I^\uparrow_\alpha - I^\downarrow_\alpha$. The first term in Eq. (10) gives the current flowing from lead α toward the active region [because it is proportional to $\mathbf{G}^>(E)$, which describes the empty states in the active region], while the second term gives the current flowing in the opposite direction [because it is proportional to $\mathbf{G}^<(E)$, which describes the occupied states in the active region]. Likewise, the self-energies $\mathbf{\Sigma}^{\lessgtr}(E)$ are proportional to the occupied lead states and the empty lead states, respectively.

While Eq. (10) is valid both in the linear and nonlinear transport regimes, in its original form it is not useful for the analysis of currents and voltages in multiterminal devices. That is, instead of voltages hidden in the self-energy and GF matrices, one would like to recast Eq. (10) into a form similar to Eq. (3), so that such equations can be easily inverted to get voltages measured between the terminals for known currents injected into the device. For this purpose, we expand all quantities in Eqs. (4) and (5) to linear order in V_α: $\mathbf{G}^r(E) \approx \mathbf{G}^r_0(E) + \mathbf{G}^r_0(E)[\sum_\alpha \mathbf{\Sigma}^r_\alpha(E - eV_\alpha) - \mathbf{\Sigma}^r_\alpha(E)]\mathbf{G}^r_0(E)$, $\mathbf{\Sigma}^r_\alpha(E - eV_\alpha) \approx \mathbf{\Sigma}^r_\alpha(E) - eV_\alpha\partial\mathbf{\Sigma}^r_\alpha(E)/\partial E$, and $f_\alpha(E) \approx f(E) - eV_\alpha\partial f/\partial E$. Here

$$\mathbf{G}^r_0(E) = \left[E - \mathbf{H} - \sum_\alpha \mathbf{\Sigma}^r_\alpha(E) - \mathbf{\Sigma}^r_{\rm int}(E)\right]^{-1} \tag{11}$$

is the retarded GF in equilibrium ($V_\alpha = {\rm const}$).

For both momentum-conserving and momentum-relaxing dephasing (or their linear combination) one has to solve for $\mathbf{G}^r_0(E)$ and $\mathbf{\Sigma}^r_{\rm int}(E)$ using a self-consistent loop where the initial guess is

$$\mathbf{G}^r_{0,{\rm in}}(E) = \left[E - \mathbf{H} - \sum_\alpha \mathbf{\Sigma}^r_\alpha(E)\right]^{-1}. \tag{12}$$

Then

$$\mathbf{G}^r_{0,{\rm out}}(E) = \left[E - \mathbf{H} - \sum_\alpha \mathbf{\Sigma}^r_\alpha(E) - d_p\mathbf{G}^r_{0,{\rm in}}(E)\right]^{-1} \tag{13}$$

in the case of momentum-conserving dephasing or

$$\mathbf{G}^r_{0,{\rm out}}(E) = \left\{E - \mathbf{H} - \sum_\alpha \mathbf{\Sigma}^r_\alpha(E) - \mathcal{D}\left[d_m\mathbf{G}^r_{0,{\rm in}}(E)\right]\right\}^{-1} \tag{14}$$

in the case of momentum-relaxing dephasing is used as the input $\mathbf{G}^r_{0,{\rm in}}(E)$ for the next iteration. We assume that the loop has converged when $||\mathbf{G}^r_{0,{\rm out}}(E) - \mathbf{G}^r_{0,{\rm in}}(E)|| < 10^{-4}$.

Using the converged $\mathbf{G}^r_0(E)$ matrix, the next step is to compute $\mathbf{G}^<_0(E)$, which proceeds differently for momentum-conserving and momentum-relaxing dephasing while yielding the same form for the generalization of Eq. (3) in the zero temperature limit:

$$I_\alpha = \frac{e^2}{h}\sum_\beta \left(T^{\rm coh}_{\alpha\beta} + T^{\rm incoh}_{\alpha\beta}\right)(V_\alpha - V_\beta). \tag{15}$$

Here the "coherent" transmission function is

$$T^{\rm coh}_{\alpha\beta}(E) = {\rm Tr}\,\{\mathbf{\Gamma}_\alpha(E)\mathbf{G}^r_0(E)\mathbf{\Gamma}_\beta(E)\mathbf{G}^a_0(E)\}, \tag{16}$$

while the "incoherent" contribution is given by

$$T^{\rm incoh}_{\alpha\beta}(E) = {\rm Tr}\,\{\mathbf{\Gamma}_\alpha(E)\mathbf{G}^r_0(E)\mathbf{\Gamma}^d_\beta(E)\mathbf{G}^a_0(E)\}, \tag{17}$$

where both are evaluated at $E = E_F$ in Eq. (15). Note that although the expression for $T^{\rm coh}_{\alpha\beta}$ in Eq. (16) resembles the transmission function[12] for phase-coherent transport of a single particle exhibiting elastic scattering only (which is a usual feature when reexpressing[28] the Meir-Wingreen formula), it actually takes into account the many-body interaction effects through $\mathbf{G}^r_0(E)$ in Eq. (11), which includes $\mathbf{\Sigma}^r_{\rm int}(E)$.

In the case of momentum-conserving dephasing, the matrix $\mathbf{\Gamma}^d_\beta$ in Eq. (17) is obtained from

$$\left[\mathbf{G}^r_0\right]^{-1}\mathbf{\Gamma}^d_\beta - d_p\mathbf{\Gamma}^d_\beta\mathbf{G}^a_0 - d_p\mathbf{\Gamma}_\alpha\mathbf{G}^a_0 = \mathbf{0}. \tag{18}$$

This is recognized as the Sylvester equation[29] of matrix algebra, $\mathbf{AX}+\mathbf{XB}+\mathbf{C}=\mathbf{0}$, where we identify the unknown matrix as $\mathbf{X}=\mathbf{\Gamma}^d_\beta$, while the known coefficients are $\mathbf{A}=[\mathbf{G}^r_0]^{-1}$, $\mathbf{B}=-d_p\mathbf{G}^a_0$, and $\mathbf{C}=-d_p\mathbf{\Gamma}_\alpha\mathbf{G}^a_0$.

In the case of momentum-relaxing dephasing, the diagonal elements of the matrix $\mathbf{\Gamma}^d_\beta$ in Eq. (17) are obtained from

$$\left[\Gamma^d_\beta\right]_{jj}=d_m\sum_v[\mathbf{Q}]_{jv}\left[\mathbf{G}^r_0\mathbf{\Gamma}_\beta\mathbf{G}^a_0\right]_{vv}, \tag{19}$$

using $\mathbf{Q}=[1-d_m\mathbf{P}]^{-1}$ and $[\mathbf{P}]_{jv}=[\mathbf{G}^r_0]_{jv}[\mathbf{G}^a_0]_{vj}$.[30] Here the notation $[\mathbf{M}]_{jv}$ denotes the matrix element of $\mathbf{M}$.

IV. APPLICATION OF MOMENTUM-RELAXING MODEL OF DEPHASING TO ZSHE IN GRAPHENE

Since we find that the momentum-conserving model of dephasing cannot reproduce the experimental results of Ref. 1, we focus on momentum-relaxing model. An interested reader can find a detailed comparison of momentum-conserving, momentum-relaxing, and traditional Büttiker voltage probe[21,22] phenomenological methods to introduce dephasing in quantum transport in Ref. 14 for a simple example of disordered wire attached to two ideal semi-infinite leads.

The momentum-relaxing model we employ in this section accounts for the *local* simultaneous phase and momentum relaxation. Also, it can be physically interpreted as a highly simplified version (valid in the high-temperature limit) of the so-called self-consistent Born approximation[12,24] for electron-phonon interaction. We note that the momentum-relaxing model has been employed before to study dephasing effects in the integer QHE,[30] where the phenomenological dephasing length is often invoked[31] to account for electron-electron and electron-phonon scattering without delving into the microscopic details of such interactions.

For phase-coherent transport in multiterminal devices, which is described by Eq. (3), the recently developed algorithms[32,33] that exploit the sparse nature of the Hamiltonian matrix of an open system, $\mathbf{H}_{\rm open}+\sum_\alpha\mathbf{\Sigma}^r_\alpha(E)$, make it possible to find specific submatrices of the inverse matrix $E-\mathbf{H}_{\rm open}$ (i.e., the retarded GF), which enter into the computation of $T^{\rm coh}_{\alpha\beta}(E)$ in Eq. (16), even for large active regions of the size $\sim$1 μm encountered in experiments. However, in the presence of dephasing $\mathbf{H}_{\rm open}+\sum_\alpha\mathbf{\Sigma}^r_\alpha(E)+\mathbf{\Sigma}^r_{\rm int}(E)$ is a dense matrix which then requires brute force (multithreaded or parallelized) computation of *all* elements of $\mathbf{G}^r_0(E)$, which is prohibitively expensive for multiterminal Hall bars of the size employed in the experiments of Ref. 1.

Therefore, since we have to perform such a computation on a grid of energy points, we select a much smaller size for the active region of the device in Fig. 1: $W\simeq 2.7$ nm for four-terminal graphene Hall bars and $W\simeq 2.0$ nm for six-terminal bars. Given the very small device size in our simulation, we have to apply unrealistically large external magnetic fields in order to bring the device into the quantizing regime $W/\ell_B>1$. Nevertheless, the important parameter for comparing our results with experiments is not the absolute value of W or B but their ratio W/ℓ_B, which is indicated in the captions for Figs. 2–4

A. Four-terminal graphene Hall bars

In the analysis of four-terminal bars, voltage $V/2$ is applied to lead 1 and $-V/2$ to lead 4, while voltages on leads 2 and 3 are set to zero. Figure 2 shows that in the quantizing external magnetic field $W/\ell_B>1$, where $\ell_B=\sqrt{\hbar/|eB|}$ is the magnetic length in graphene, the four-terminal bar generates large spin Hall conductance $G_{\rm SH}=(I^\uparrow_2-I^\downarrow_2)/(V_1-V_4)$ and the corresponding SH angle $\theta_{\rm SH}=G_{\rm SH}/G_L$, with $G_L=I_1/(V_1-V_4)$ being the longitudinal charge conductance. The spin current $I^S_2=I^\uparrow_2-I^\downarrow_2$ in the ZSHE is carried by spins polarized along the z axis orthogonal to the plane of graphene.

The value of $G_{\rm SH}$ is comparable to the one predicted[34] for the Rashba SO-coupled four-terminal 2DEGs of the size of the spin precession length (on which a spin-up state precesses to a spin-down state). Unlike mesoscopic SHE[5] in 2DEGs where Rashba SOC induces both the transverse spin deflection[11] and spin dephasing, which compete against each other in the processes of generating *pure* (not accompanied by any net charge flux) spin current, in the ZSHE transverse spin current is pure only at the DP [where charge current $I_2=I^\uparrow_2+I^\downarrow_2$ becomes zero in Figs. 2(a) and 3(a)], and spin precession is *absent*. This might be advantageous for spintronic applications since spin dephasing is evaded, as demonstrated by the experimental detection of nonlocal voltage even at distances $\sim$10 μm away from the device region where SH current was induced.[1] We note that for a very strong magnetic field, as could be achieved in ferromagnetic graphene, the $G_{\rm SH}$ peaks in Fig. 2(c) would become quantized[35] as a realization of quantum SHE[3] in the absence of SOC.

The introduction of dephasing processes into the four-terminal bars, which relax *both*[14] the phase and the momentum of quasiparticles propagating through the active region, destroys the quantization of the charge Hall conductance $G_{xy}=I_2/(V_1-V_4)$ or charge Hall resistance $R_{\rm H}$ and underlying chiral edge states, as demonstrated by the transition from Figs. 2(a) to 3(a) for G_{xy} and from Figs. 2(b) to 3(b) for $R_{\rm H}$. The charge Hall resistance in four-terminal bars is defined as $R_{\rm H}=(V_3-V_2)/I_1$ for the measuring setup where current I_1 is injected into lead 1 and voltages V_3 and V_2 develop as the response to it. The SH conductance and SH angle are concurrently reduced by two orders of magnitude, which are values similar to those found in quasiclassical approaches[16] in the temperature range $T=200$–300 K.

B. Six-terminal graphene Hall bars

In the analysis of six-terminal Hall bars, charge current I_1 is injected through lead 1 and current $-I_1$ flows through lead 4 while $I_\alpha\equiv 0$ in all other leads. We then compute voltages which develop in the leads $\alpha=2,3,5,6$ labeled in Fig. 1 in response to injected current I_1. Figure 4(b) shows peaks in the nonlocal resistance, defined as $R_{\rm NL}=(V_5-V_6)/I_1$, within the phase-coherent transport regime which closely resemble the DP and side peaks observed experimentally in strong (quantizing) external magnetic field.[1] The transition of $R_{\rm NL}$ from Figs. 4(b) to 4(d) shows how dephasing removes both side peaks while leaving the nonlocal voltage around the DP which is two orders of magnitude smaller than in the

phase-coherent regime. The Hall resistance in the six-terminal bar, $R_H = (V_1 - V_4)/I_3$ defined for current injected I_3 and voltages measured between leads 1 and 4 (for $I_1 = I_4 = 0$), changes smoothly from Figs. 4(a) to 4(c) as dephasing in increased, where the curve in Fig. 4(c) looks exactly the same as those observed experimentally[16] for $T = 250$ K and $B = 1$–12 T.

V. CONCLUSIONS

In conclusion, we have developed a fully quantum transport theory of recently observed[1] nonlocal voltage in magnetotransport near the DP in graphene Hall bars which provides a unified picture of this phenomenon and the underlying Zeeman-splitting-driven spin Hall effect from the quantum-coherent transport regime at low temperatures to a semiclassical transport regime at higher temperatures. At the same time, this approach makes it possible to take into account arbitrary strength of the magnetic field or scattering processes by short- and long-range static impurity potential near the DP where the quasiclassical Boltzmann equation breaks down[18] even for transport at high temperatures. Our theory starts from the NEGF-based Meir-Wingreen formula for spin-resolved charge currents in the attached leads, including phenomenological many-body self-energies that take into account relaxation of both the phase and the momentum of Dirac fermions in the active region of the device, which is then linearized to provide the connection between current and voltages in different leads. This can be viewed as a generalization of the Landauer-Büttiker formula for phase-coherent transport in multiterminal device geometries where the usually employed additional Büttiker voltage probes to introduce dephasing, which also generate additional scattering in an uncontrolled fashion, are avoided.

ACKNOWLEDGMENTS

We thank K. Ensslin for illuminating discussions. C.-L.C. and B.K.N. were supported by DOE Grant No. DE-FG02-07ER46374, and C.-R.C. was supported by Republic of China National Science Council Grant No. NSC 98-2112-M-002-012-MY3.

*crchang@phys.ntu.edu.tw
†bnikolic@udel.edu

[1] D. A. Abanin *et al.*, Science **332**, 328 (2011).
[2] A. K. Geim, Science **324**, 1530 (2009).
[3] X.-L. Qi and S.-C. Zhang, Rev. Mod. Phys. **83**, 1057 (2011).
[4] R. Winkler, *Spin-Orbit Coupling Effects in Two-Dimensional Electron and Hole Systems* (Springer, Berlin, 2003).
[5] B. K. Nikolić, L. P. Zârbo, and S. Souma, Phys. Rev. B **72**, 075361 (2005).
[6] C. Brüne *et al.*, Nat. Phys. **6**, 448 (2010).
[7] A. Roth *et al.*, Science **325**, 294 (2009).
[8] E. M. Hankiewicz, L. W. Molenkamp, T. Jungwirth, and J. Sinova, Phys. Rev. B **70**, 241301 (2004).
[9] C. Ertler, S. Konschuh, M. Gmitra, and J. Fabian, Phys. Rev. B **80**, 041405(R) (2009).
[10] A. H. Castro Neto and F. Guinea, Phys. Rev. Lett. **103**, 026804 (2009).
[11] B. K. Nikolić, L. P. Zârbo, and S. Welack, Phys. Rev. B **72**, 075335 (2005).
[12] H. Haug and A.-P. Jauho, *Quantum Kinetics in Transport and Optics of Semiconductors* (Springer, Berlin, 2007).
[13] Y. Meir and N. S. Wingreen, Phys. Rev. Lett. **68**, 2512 (1992).
[14] R. Golizadeh-Mojarad and S. Datta, Phys. Rev. B **75**, 081301(R) (2007).
[15] M. Büttiker, Phys. Rev. Lett. **57**, 1761 (1986).
[16] D. A. Abanin, R. V. Gorbachev, K. S. Novoselov, A. K. Geim, and L. S. Levitov, Phys. Rev. Lett. **107**, 096601 (2011).
[17] A. Cresti, N. Nemec, B. Biel, G. Niebler, F. Triozon, G. Cuniberti, and S. Roche, Nano Res. **1**, 361 (2008).
[18] J. W. Kłos and I. V. Zozoulenko, Phys. Rev. B **82**, 081414 (2010); A. Ferreira, J. Viana-Gomes, J. Nilsson, E. R. Mucciolo, N. M. R. Peres, and A. H. Castro Neto, *ibid.* **83**, 165402 (2011).
[19] O. V. Yazyev and M. I. Katsnelson, Phys. Rev. Lett. **100**, 047209 (2008).
[20] J. Kunstmann, C. Özdoğan, A. Quandt, and H. Fehske, Phys. Rev. B **83**, 045414 (2011).
[21] M. Büttiker, Phys. Rev. B **32**, 1846 (1985).
[22] G. Metalidis and P. Bruno, Phys. Rev. B **73**, 113308 (2006); Y. Xing, Q.-F. Sun, and J. Wang, *ibid.* **77**, 115346 (2008); J. Maassen, F. Zahid, and H. Guo, *ibid.* **80**, 125423 (2009).
[23] K. S. Thygesen and A. Rubio, Phys. Rev. B **77**, 115333 (2008).
[24] T. Frederiksen, M. Paulsson, M. Brandbyge, and A.-P. Jauho, Phys. Rev. B **75**, 205413 (2007).
[25] A. Hurley, N. Baadji, and S. Sanvito, Phys. Rev. B **84**, 035427 (2011).
[26] D. A. Areshkin and B. K. Nikolić, Phys. Rev. B **81**, 155450 (2010).
[27] H. Ness and L. K. Dash, Phys. Rev. B **84**, 235428 (2011).
[28] H. Ness, L. K. Dash, and R. W. Godby, Phys. Rev. B **82**, 085426 (2010).
[29] H. Golub and C. F. van Van Loan, *Matrix Computations* (Johns Hopkins University Press, Baltimore, 1996).
[30] A. Cresti and G. P. Parravicini, Phys. Rev. B **78**, 115313 (2008).
[31] L. P. Pryadko and A. Auerbach, Phys. Rev. Lett. **82**, 1253 (1999).
[32] K. Kazymyrenko and X. Waintal, Phys. Rev. B **77**, 115119 (2008).
[33] M. Wimmer and K. Richter, J. Comput. Phys. **228**, 8548 (2009).
[34] B. K. Nikolić and R. L. Dragomirova, Semicond. Sci. Technol. **24**, 064006 (2009).
[35] Q.-F. Sun and X. C. Xie, Phys. Rev. Lett. **104**, 066805 (2010).