Resonant tunneling and intrinsic bistability in twisted graphene structures

J. F. Rodriguez-Nieva,1 M. S. Dresselhaus,1,2 and L. S. Levitov1

1Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
2Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

(Received 16 June 2015; revised manuscript received 11 July 2016; published 15 August 2016)

We predict that vertical transport in heterostructures formed by twisted graphene layers can exhibit a unique bistability mechanism. Intrinsically bistable $I$-$V$ characteristics arise from resonant tunneling and interlayer charge coupling, enabling multiple stable states in the sequential tunneling regime. We consider a simple trilayer architecture, with the outer layers acting as the source and drain and the middle layer floating. Under bias, the middle layer can be either resonant or nonresonant with the source and drain layers. The bistability is controlled by geometric device parameters easily tunable in experiments. The nanoscale architecture can enable uniquely fast switching times.

DOI: 10.1103/PhysRevB.94.085412

I. INTRODUCTION

Nanoscale systems that can switch between distinct macroscopic states upon variation of some control parameter are in high demand in diverse areas of nanoscience research. Bistable electronic systems which exhibit fast switching are of interest for applications such as low-power memory and logic [1]. Recently, new realizations of intrinsically bistable systems have been discovered, both in graphene [2–6] and in other systems [7–9]. In particular, van der Waals heterostructures comprising graphene layers sandwiched between insulating hexagonal boron nitride (hBN) layers afford electronic environments with tailored band structures and transport characteristics [10]. It was demonstrated that introducing a twist between adjacent graphene layers in such heterostructures can result in a resonant behavior of the tunneling current and nonmonotonic $I$-$V$ characteristics [11]. It is therefore tempting to exploit twisted graphene multilayer structures as a platform for bistable and hysteretic nanoscale systems.

Here we predict intrinsic bistability and hysteretic $I$-$V$ characteristics for vertical transport in heterostructures formed by graphene monolayers separated by hBN barriers in a twisted arrangement similar to that described in Ref. [11]. Essential for our bistability mechanism are resonances originating from momentum-conserving tunneling between linearly dispersing Dirac bands [12] and occurring when the bands are aligned [11] [see Figs. 1(b) and 1(c)]. Bistability arises due to current-induced charge accumulation producing an interlayer bias that tunes the interband tunneling in and out of resonance.

Below we focus on the simplest case of two-step sequential tunneling in a device comprising three graphene monolayers. Such trilayer architecture, pictured in Fig. 1(a), with the top and bottom layers acting as a source and drain and the middle layer electrically decoupled (floating), is similar to previously studied double-barrier quantum-well (QW) structures [13]. However, our bistability mechanism, originating from resonant tunneling between Dirac bands in graphene layers, is distinct from that in the QW structures [13]. In our case, multiple stable states arise because the decoupled layer can, for a fixed external bias, be either in a resonant (low-resistance) or a nonresonant (high-resistance) state. This behavior is illustrated in Fig. 1(c).

\[ |\mathbf{q}_A| = (8\pi/3a_0) \sin(\theta/2), \quad \Delta = \hbar v_F |\mathbf{q}_A|, \quad (1) \]

FIG. 1. (a) Trilayer graphene heterostructure schematics, with layers labeled 1 to 3. Here $I_{12}$ and $d_{12}$ are the interlayer currents and distances. (b) Band structure of twisted graphene layers 1 (blue) and 2 (red). The twist angle $\theta$ defines a characteristic energy $\Delta$ [Eq. (1)] and three superlattice wave vectors $\mathbf{q}_{A,B,C}$ [Eq. (12)]. (c) Bistable $I$-$V$ characteristics. The resonant and nonresonant bistable states are illustrated in the top left inset (details are discussed in Fig. 3). The procedure for finding bistable solutions is illustrated in the bottom right inset [see Eq. (9) and accompanying discussion].

The bistability is governed by geometric parameters, the twist angle $\theta$ and the interlayer distances $d_{ij}$, which are easily tunable in experiments. The twist angle controls the Dirac cones’ displacement in the two layers and the energy at which the cones intersect [see Fig. 1(b)].
where $\nu_F \approx 10^6$ m/s is the carrier velocity and $a_0 \approx 2.46$ Å is the graphene lattice constant. The distances $d_{ij}$, marked in Fig. 1(a), determine the interlayer tunnel conductance values $G_{ij} \sim e^{2(d_{ij}-d_{0})/\lambda}$, where $\lambda$ is the Wentzel-Kramers-Brillouin (WKB) length governing the tunneling amplitude dependence on barrier width. In what follows we will use the conductance ratio

$$Z = G_{12}/G_{23} \sim e^{2(d_{12}-d_{23})/\lambda},$$

where $G_{ij}$ denotes the conductance between the corresponding layers.

The quantities $\theta$ and $d_{ij}$ can be controlled with a large degree of precision. The twist angle $\theta$ can be tuned within $1^\circ$ during fabrication [11], whereas $d_{ij}$ can be varied by adding monolayers of dielectric materials, such as hBN or MoS$_2$. Since typical values $\lambda = h/(2m_eW)^{1/2} \sim 2$ Å, estimated for the tunneling barrier height $W \sim 1$ eV and the effective electron mass $m_e \sim 10^{-30}$ kg, are comparable to the hBN or MoS$_2$ monolayer thickness, variation in $d_{ij}$ results in a fairly gradual change of $Z$.

One appealing aspect of this system is the short interlayer transport length of a nanometer scale, which can allow high operation speeds and fast switching times. This is evident from an estimate for the $RC$ time, $t_{RC} = \kappa/4\pi gd \sim 100$ ns, where $\kappa \sim 1$ is the dielectric constant, $d \sim 1$ nm is the interlayer separation, and $g \sim 10^{-7} \Omega^{-1}\mu$m$^{-2}$ is the interlayer conductance per unit area. The combination of geometric tunability and small transport lengths is not present in previously studied graphene-based bistable systems, such as graphene flash memories [2,3] or graphene resistive memories [4–6]. Small thicknesses can also enable large packing densities.

The steep electronic dispersion in graphene makes the bistable-state properties distinct from those in QW systems. In our case, the bistability is controlled by the resonances arising due to band alignment. The corresponding bias value, which scales as a power law of the energy $\Delta$ given in Eq. (1), can be as large as $\delta V \sim 100$–500 mV (see discussion below). In QW systems, instead, the bias range where bistability occurs is mainly controlled by the amount of charge $n_{QW}$ that can be stored in a quantum well, $\delta V \approx e$,$n_{QW}/C$, where $C$ is the interlayer capacitance [14]. Typical carrier densities in the “charged” and “uncharged” states of a bistable QW system, assessed by magnetic oscillation measurements [15], are on the order of $n_{QW} \sim 10^{11}$/cm$^2$ and $n_{QW} \sim 0$, respectively. These carrier densities yield typical values $\delta V \sim 50$ mV in double-barrier quantum wells with a width of tens of nanometers ($C \sim 0.1$–1 mF). Such values can be as much as an order of magnitude smaller than the above estimate predicts for the graphene case.

II. SEQUENTIAL TUNNELING MODEL

Vertical transport in our trilayer architecture can be described by a simple sequential model. The model validity relies on the interlayer tunnel coupling being weak such that the interlayer charge transfer is slow compared to the intralayer electron relaxation. Indeed, the values $\tau_{RC}$, estimated above, are much longer than typical thermalization times in graphene, $\tau_0 \sim 10$ ps [16]. The $RC$ times, however, are sufficiently fast to be competitive with the speeds of existing switching devices [1].

The interlayer transport mechanism is mainly governed by the twist angle $\theta$, which defines the $K$-point displacement $q_A$ between graphene lattices in adjacent layers, and the interlayer bias. Under bias, the value $\left| q_A \right|$ given in Eq. (1) determines the range of momenta and energies for which momentum-conserving tunneling is allowed. Large values of $\left| q_A \right|$ hinder resonant tunneling given that phonon and defect scattering are necessary to supply the large momentum mismatch between layers. Momentum-nonconserving transport can also occur if the top/bottom layer is made of a different material so that there is a large mismatch between unit cells with respect to that of graphene. For small $\left| q_A \right|$, on the other hand, momentum-conserving tunneling is possible for moderately small values of bias.

In our two-step sequential tunneling model, we treat transport between layers 1 and 2 as momentum conserving. The second step, between layers 2 and 3, is assumed to be momentum nonconserving and described by Ohm’s law. The latter assumption allows us to simplify our discussion and focus on the essential aspects of bistability. In addition, we also assume that the contact resistances are sufficiently small so that all the potential drop occurs predominantly between the graphene layers.

Turning to a systematic development of the model, the low-energy Hamiltonian $H$ describing coherent transport between a pair of twisted graphene monolayers has contributions $H = H_1 + H_2 + H_{12}$. Here $H_{1,2}$ are the free-particle terms describing massless Dirac particles in each graphene layer, and $H_{12}$ describes the interlayer tunnel coupling [17–19]. The free particle terms are

$$H_1 = \sum_{k} \psi_{1,k}^{\dagger} \left[ \hbar \nu_F \sigma \cdot (k + q_A/2) - \mu_1 \right] \psi_{1,k},$$

$$H_2 = \sum_{k} \psi_{2,k}^{\dagger} \left[ \hbar \nu_F \sigma \cdot (k - q_A/2) - \mu_2 \right] \psi_{2,k},$$

where $\mu_{1,2}$ are the Fermi energies measured relative to the Dirac point. For a small twist angle $\theta$, the large-wave-number processes that couple different valleys can be neglected. In this case, it is sufficient to account for a single Dirac cone in each layer [see Eq. (3)]. We adopt this approximation below.

The tunneling coupling can be modeled as a local, periodic function of position [17]:

$$T_{12} = \sum_{k,G} \sum_{k,G} \psi_{1,k}^{\dagger} T_{G,k} \psi_{2,k+G} + H.c.$$ (4)

The periodicity of the interlayer coupling, quantified by the $G$ wave vectors, is determined by the hexagonal superlattice unit cell that is formed by the twisted graphene layers (see Fig. 2). For small $\theta$, only the longest-wavelength contributions are relevant for tunneling. Referred from the Dirac point of layer 1, such long-wavelength components are given by $q_A$, $q_B = q_A - G_1$ and $q_C = q_A - G_2$ (see Fig. 2), where $G_{1,2}$ are the reciprocal vectors of the superlattice Brillouin zone, which is smaller than the graphene Brillouin zone by a factor $\sin^2(\theta)$. While the higher-$q$ harmonics of the interlayer hopping potential spatial modulation also contribute to tunneling, it can be shown that their contributions vanish.
and can be decomposed into Fourier components conserving tunneling coupling has the periodicity of the superlattice dominated by the smallest wave vectors $q_{n,m}$ leads to the tunneling Hamiltonian due to the built-up interlayer electrostatic potential rapidly on the reciprocal lattice vector $G_{1,2}$ scale [17,18]. This leads to the tunneling Hamiltonian composed of only three Fourier components. In this expression for $T_{12}$, the $k$ vectors are relative to the Dirac point of each layer, i.e., $k - q_{A}/2 \to k$ in layer 1 and $k + q_{A}/2 \to k$ in layer 2.

Parenthetically, the lattice of the dielectric material separating the graphene layers can produce slowly varying spatial modulation of the tunneling transition amplitude $T$ in Eq. (5), giving rise to the effects resembling those due to a twist angle $\theta$. This would be the case when the dielectric and graphene are nearly lattice matched as, e.g., in highly oriented hBN-graphene structures, which have a small lattice mismatch of about 1.8% (a detailed discussion of these effects can be found in Ref. [20]). Such effects, if present, would alter the values $q_{A,B,C}$ but otherwise not change our discussion in an essential way.

Under an interlayer bias potential $V_{12}$, the tunneling current $I_{12}$ is

$$I_{12} = \frac{eN}{\hbar} \sum_{k,s,j} |T_{j}^{s'}(k)|^{2} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} A_{1,s}(k,\omega) \times A_{2,s}(k + q_{j}) \langle f_{1}(\omega) - f_{2}(\omega) \rangle, \quad (6)$$

where $s$ ($s'$) refers to the electron (+) and hole (−) bands of layer 1 (2) and $N = 4$ is the spin and valley degeneracy. The functions $f_{1}(\omega) = 1/[e^{\beta(\omega - \mu_{1})} + 1]$ are the Fermi distribution functions for each layer, with $\beta = 1/k_{B}T$ being the inverse thermal energy and $\mu_{1}$ being the Fermi energies. The function $A_{i,s}$ is the spectral function of layer $i$ and band $s$. The energy for the quantities in layer 2 is offset by $\tilde{\omega} = \omega + e\Phi_{12}$ due to the built-up interlayer electrostatic potential $\Phi_{12}$ [see Eq. (3)] between layers 1 and 2. Because of capacitance effects, the interlayer electrostatic and chemical potentials are related by

$$eV_{12} = \mu_{1} - \mu_{2} - e\Phi_{12}, \quad (7)$$

where $\mu_{1}$ and $\Phi_{12}$ are implicit functions of $V_{12}$.

This procedure to obtain the $I$-$V$ response is shown graphically in the inset of Fig. 1(c). The straight line describes transport between layers 2 and 3, which is assumed to follow Ohm’s law, $I_{23} = G_{23}V_{23}$, where $G_{23}$ and $V_{23}$ are the interlayer conductance and interlayer bias potential between layers 2 and 3, respectively.

### III. Electrostatic Feedback

In order to include the electrostatic feedback, Eq. (9) needs to be complemented with further electrostatic considerations that relate the variables $V_{ij}$, $\Phi_{ij}$, and $\mu_{j}$. It is important to note that all variables can be determined once the carrier densities in each layer, $n_{1}, n_{2}$, and $n_{3}$, are known. Indeed, assuming that there is no external gate, the neutrality condition relates the charge densities in the different regions of the device as

$$n_{1} + n_{2} + n_{3} = 0. \quad (10)$$

Furthermore, the application of an external bias potential $V$ fixes the Fermi level difference between layer 1 and layer 3 as

$$eV = \mu_{1} - \mu_{3} + \frac{4\pi e^{2}}{\kappa}(n_{1}d_{13} + n_{2}d_{23}). \quad (11)$$

Here $d_{ij}$ is the interlayer distance between layer $i$ and layer $j$, $\kappa$ is the dielectric constant of the barrier material, and $\mu_{j} = \text{sgn}(n_{j})\hbar v_{F}\sqrt{\pi n_{j}}$. In Eq. (11), we implicitly assume that all layers are undoped at $V = 0$. Equations (9)–(11) then form a closed set of equations from which $n_{1}, n_{2}$, and $n_{3}$ can be obtained. The remaining variables, $V_{ij}$, $\Phi_{ij}$, are functions of $n_{i}$. In particular, the electrostatic potentials are $\Phi_{12} = -4\pi e^{2}d_{12}n_{1}/\kappa$ and $\Phi_{23} = 4\pi e^{2}(n_{1}d_{13} + n_{2}d_{23})/\kappa$, whereas the interlayer bias potentials are $V_{12} = \mu_{1} - (\mu_{2} + \Phi_{12})$ and $V_{23} = (\mu_{2} + \Phi_{12}) - (\mu_{3} + \Phi_{23})$.

For simplicity, here we fix the Fermi energies in Eq. (6) to a constant value $\mu_{1} = \mu$. This is equivalent to turning off capacitance effects. In this case, $V_{ij} = \Phi_{ij}$ (see Fig. 3). This approximation is valid in the regime $4e^{2}d_{ij}\Delta/\kappa(\hbar v_{F})^{2} \approx 15d_{ij}|\Delta|eV/\kappa \gg 1$. In this regime, minimal changes in carrier concentration induce large interlayer electrostatic
in the plane: when $\Delta = 0$, these regions, indicated with black dashed lines contributing to the resonant tunneling current for fixed $V$. These regions, indicated with black dashed lines at the intersection of the twisted Dirac cones, form conical surfaces in the $k$ plane: when $eV_{12} < \Delta$, the lines form hyperbolic curves, and when $eV_{12} > \Delta$, the lines form ellipsoidal curves. When $eV_{12} = \Delta$, a van Hove singularity in the tunneling density of states is obtained. As shown in (d), the nonresonant (high-resistance) bistable state ($eV_{12} > \Delta$) can be Pauli blocked by adjusting the doping. Doping thus affords a way to tune the current ratio between bistable branches in Fig. 1(a).

In this work it is assumed that the Dirac cones are aligned at $V = 0$, and that capacitance effects are neglected. Layers are labeled 1–3 as in Fig. 1(a). (a) $eV_{12} < \Delta$; (b) $eV_{12} = \Delta$; (c) $eV_{12} > \Delta; \mu \ll \Delta$; and (d) $eV_{12} > \Delta; \mu > \Delta$.

potentials. The more realistic scenario which includes quantum capacitance effects [21], such that $\mu_{1,2}$ vary with $V_{12}$, is here considered in Appendix A. However, this more realistic picture only introduces small corrections to the tunneling current without major consequences to our bistability discussion.

IV. MODEL PARAMETERS

In order to solve Eq. (9), we need to specify the matrix elements $T_j$ in Eq. (8). A simple and explicit model for $T_j$ and the wave vectors $q_j$ is provided by Ref. [17]:

$$T_j = i \left( e^{i \varphi_j} \begin{pmatrix} 1 \\ e^{-i \varphi_j} \end{pmatrix} \right), \quad q_j = \frac{\Delta}{\hbar v_F} (\sin \varphi_j, -\cos \varphi_j),$$  \hspace{1cm} (12)

with $\varphi_A = 0$, $\varphi_B = 2\pi/3$, $\varphi_C = 4\pi/3$. This representation is obtained for small twisting angles after performing a $\sigma$ rotation of phase space in layer 2 (see details in Ref. [17]). It is also implicit in Eq. (12) that the top and bottom graphene lattices have a common lattice point [17]: a rigid horizontal translation between lattices adds an additional overall phase to the matrix $T_j$ [18]. We stress, however, that relative phases in $T_j$ do not alter in any significant way the physics of tunneling in Eq. (6). Furthermore, while the interlayer hopping amplitude $t$ is sensitive to several parameters, e.g., twist angle [19] and the choice of dielectric material [20], its order of magnitude is mainly governed by the wave function overlap between the graphene layers. Such dependence will be described below within the WKB approximation. Equations (3) and (12) are expected to be accurate for twist angles $\theta \lesssim 10^\circ$ and energies of 1 eV [19].

For an estimate below we use the value $\theta = 2^\circ$. This defines an energy scale $\Delta = 0.37$ eV. Furthermore, we take a Lorentzian spectral function in Eq. (6) for both layers, $A_j(\omega, \omega_0) = 2\Gamma/(\omega - \omega_0 + i\Gamma)$ with the linewidth $\Gamma$~10 meV. A finite linewidth $\Gamma$ is necessary to have a finite value of the peak current when $eV_{12} = \Delta$ (see Fig. 3). The temperature and Fermi level of the system were taken to be $T = 0$ and $\mu = 0$, respectively. With reference to Eq. (6), we define the interlayer conductance

$$G_{12} = S g_{12}, \quad g_{12} = 2\pi N \frac{|t|^2}{(\hbar v_F)^2} \frac{e^2}{h},$$  \hspace{1cm} (13)

where $S$ is the surface area of the device. The value of $g_{12}$ is sensitive to the twist angle and the stacked dielectric material, if any, via the parameter $t$. Here we use $g_{12} = 10^{-7} \Omega^{-1} \mu m^{-2}$.

Similar values of $g_{12}$ were measured in resonant tunneling devices which contained four layers of BN in between the graphene layers [12]. For $Z$, we consider a value of $Z = G_{12}/G_{23} = 0.2$.

V. BISTABLE I-V CHARACTERISTICS

The bistable $I-V$ characteristics are shown in Fig. 1(c). For a sufficiently large bias, $eV \gtrsim \Delta$, the current branches into two stable states. The low-resistance branch in Fig. 1(c) corresponds to two layers at resonance (i.e., $eV_{12} \approx \Delta$), whereas the high-resistance branch corresponds to a nonresonant state (i.e., $eV_{12} > \Delta$). We note that a third solution is also possible, indicated by a dashed line in the $I-V$ response [see Fig. 1(c)]. This solution, however, is unstable given that a small perturbation in $\delta V_{12}$ will push the system away from this state.

The bistable bias range can be estimated as $\delta V \approx (I_{12}^{(pk)} - I_{12}^{(vi)})/G_{23}$, where $I_{12}^{(pk)}$ and $I_{12}^{(vi)}$ are the peak interlayer current and valley interlayer current (see inset of Fig. 1(c)). To estimate $I_{12}^{(pk)}$ and $I_{12}^{(vi)}$, we first note that the tunneling matrix element $T_{j'}(k)$ varies, upon integration in $k$ space, in the range $0 \leq |T_{j'}(k)| \leq 2t$, taking typical values $|T_{j'}(k)| \approx t$. Thus, it is a good approximation to take band- and wave-vector-independent phase factors $|T_{j'}(k)| = T$. Furthermore, in the typical scenario the model parameters satisfy $\Gamma(-10 \text{meV}) \ll \Delta(-0.1-1 \text{eV})$. With this in mind, the integration of Eq. (6) allows an analytical expression to result in terms of line integrals in conical surfaces (see Fig. 3 and the discussion in Appendix B). Using $\mu_{1,2} = 0$ and $V_{12} = \Phi_{12}$, we find that the nonresonant interlayer current takes the simple form

$$I_{12}(x) = \frac{x^2 - 1/2}{\sqrt{2(x^2 - 1)}}, \quad I_{12}^{(vi)} = \frac{3\sqrt{2}F^2 G_{12}\Delta}{4} e.$$  \hspace{1cm} (14)

Here $x = eV_{12}/\Delta \gtrsim 1$ and $I_{12}^{(vi)}$ is the valley current obtained at $x = \sqrt{3/2}$. When $eV_{12}/\Delta = 1$, however, the current is at resonance and reaches a maximum value which is sensitive to $\Gamma$. To leading order in $\Gamma$, we obtain (see Appendix B)

$$I_{12}^{(pk)} / I_{12}^{(vi)} = \pi \sqrt{\Delta/2\Gamma},$$  \hspace{1cm} (15)
where $\Gamma$, in general, depends on the amount and type of disorder and/or temperature. Equations (14) and (15) yield $e^2 V/\Delta \approx 3 \sqrt{2} \hbar^2 Z [\pi \sqrt{\Delta / 2T} - 1]/4$. Importantly, very small values of $Z$ ($G_{23} \gg G_{12}$) make the bistable bias range negligibly small, whereas large values of $Z$ ($G_{23} \ll G_{12}$) would push the onset of the bistability region to very large bias potentials. Optimally, values of $Z \sim 1$ and very small $\Gamma$ would make the bistability effect more prominent.

Achieving a large current ratio between bistable states is desirable for applications; this facilitates the reading process in a bistable device. From Eqs. (14) and (15), it is shown that the current ratio between the bistable branches is controlled by the parameter $Z\sqrt{\Delta / \Gamma}$. For realistic values of disorder, this ratio can be in the 1–20 ballpark. It is interesting to note that these already high values can be boosted by means of Pauli blocking. As shown in Figs. 3(c) and 3(d), for sufficiently heavily doped samples, the nonresonant bistable state (but not the resonant one) is Pauli blocked. The degree of the electrical current ratio between stable states. These large carrier-density differentials can be used as a smoking gun of intrinsic bistability. These electrostatic considerations are discussed next.

Although electrostatic doping of the graphene layers is not essential for the physics that we describe, it is a convenient feature of bistability. In particular, for a fixed external bias potential, each bistable state exhibits different carrier concentrations. Thus, any in-plane measurement, such as conductance or magnetotransport, will be able to distinguish two distinct bistable states. Indeed, from the inset of Fig. 1(c) we see that the interlayer bias potential for each bistable state differs by an amount $\delta V_{12} \sim \Delta \pi e^2 / (2\hbar)$ (see also discussion in the Appendix A). Taking into account the capacitance of the layers, the induced carrier difference between both states is approximately $\delta n \sim \kappa \Delta / 4 \pi e^2 d_{12}$ (here the quantum capacitance is not included). Using $\theta = 2^\circ$, $\kappa = 1$, and $d_{12} = 1$ nm, we obtain a carrier-density difference $\delta n \sim 10^3$ cm$^{-2}$ between stable states. These large carrier-density differentials can be used as a smoking gun of intrinsic bistability.

VI. OTHER GRAPHENE-BASED BISTABLE SYSTEMS

Although we considered here for simplicity a two-step sequential tunneling structure where only one pair of layers can be resonant, similar ideas apply to more complex structures. Interesting examples include a two-step resonant-resonant structure, opening the possibility for tristability or multi-step “cascade” devices.

Finally, we also expect bistable $I$-$V$ characteristics in twisted graphene trilayers in the absence of any dielectric material. Indeed, incommensurability between graphene lattices already suppresses interlayer hybridization, regardless of the layer separation being a fraction of a nanometer, thus enabling the sequential tunneling regime [18]. Furthermore, the massless Dirac spectrum, and thus Eq. (3) and the subsequent transport model, remains valid but with a modified Fermi velocity [17]. We stress, however, that stacked dielectric materials have two important advantages: (i) they enable tuning the interlayer coupling, and (ii) they facilitate the interlayer potential buildup in order to achieve a resonant behavior.

VII. SUMMARY

In summary, graphene-based van der Waals heterostructures afford a new platform to realize devices with tunable $I$-$V$ characteristics, in particular those with intrinsically bistable and hysteretic behavior. System parameters required to realize the bistable behavior are readily accessible in on-going experiments. The atomic-scale interlayer distances can result in a fast response and large packing densities, making these heterostructures appealing for a variety of applications.

ACKNOWLEDGMENTS

We thank A. D. Liao for useful discussions, and acknowledge support from National Science Foundation Grant No. DMR-1004147 (J.F.R.-N. and M.S.D.), from the STC Center for Integrated Quantum Materials, NSF Grant No. DMR-1231319, and from the Center for Excitonics, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award No. DE-SC0001088 (L.S.L.).

APPENDIX A: CAPACITANCE EFFECTS

In the main text, we fixed the Fermi energy $\mu$, the different graphene layers to some constant value. A more refined model of the $I$-$V$ response should, however, include quantum capacitance effects so that Fermi energy is allowed to vary with $V$. Although the features of bistability are not significantly modified by such corrections, as shown below, carrier-density differentials between bistable states are a smoking gun of intrinsic bistability. These electrostatic considerations are discussed next.

Here we numerically solve Eqs. (6)–(11), assuming a thin device separated by dielectric barriers of thickness $d_{12} = d_{23} = 1.4$ nm (e.g., four layers of hBN) and dielectric constant $\kappa = 5$. The procedure to solve the $I$-$V$ response self-consistently is shown in Fig. 4(a), where $n_1$ and $n_2$ are taken as independent variables [$n_3$ is obtained from Eq. (10)] and $\delta I = I_{12} - I_{23}$ in Eq. (6) is numerically calculated (color map). For fixed $V$, indicated with dotted isolines in Fig. 4(a), the self-consistent solutions to the equilibrium equations are given by the pair $(n_1, n_2)$ such that $\delta I = 0$.

The resulting $I$-$V$ response is shown in Fig. 4(b). Importantly, the $I$-$V$ characteristics are qualitatively similar to those obtained in the main text by neglecting quantum capacitance effects. Furthermore, by inspection of the $n_1$ and $n_2$ axes in Fig. 4(a), we see that the difference in carrier
concentration $\delta n$ between each bistable state is on the order of $\delta n \sim 10^{11} - 10^{12}$ cm$^{-2}$. These carrier-concentration differences can easily be detected by lateral transport measurements and may act as clear fingerprints of intrinsic bistability.

**APPENDIX B: ANALYTIC EXPRESSIONS FOR THE PEAK AND VALLEY RESONANT TUNNELING CURRENT**

We derive here Eqs. (14) and (15) of the main text, under the assumption that the phase factors $T_{ss}^{\mu}$ in Eq. (8) are independent of the wave vector and band index, i.e., $|T_{ss}^{\mu}(k)| = \bar{T}$. Under this assumption, $I_{12}$ depends only on the modulus of $q_j$ but not on its direction, and $\sum |T_{ss}^{\mu}(k)|^2 = 3\bar{T}^2$. Given that $\Gamma \ll \Delta$, when $e\Phi_{12} > \Delta$ (nonresonant state), we can set $\Gamma \to 0$ and thus take $A_{1,2}(k,\omega) = \pi \delta(\omega - s h \nu_{F}|k|)$. The two $\delta$ functions appearing in Eq. (6) can then be integrated in $k$-space, resulting in a one-dimensional integral along the contour of an ellipse:

$$\sum_{ss'} \int \frac{d\mathbf{k}}{2\pi^2} \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \frac{d\omega}{2\pi} A_{1,ss}(k,\omega) A_{2,ss'}(\mathbf{k} + \mathbf{q},\omega)$$

$$= \frac{\delta_+ - \delta_-}{16 \pi^2 (h\nu_F)^2} \int_{\Phi_{12}} \delta F\left(\frac{\omega(\omega - \Delta)^{1/2}}{\Gamma\Delta} + O(\Gamma/\Delta)\right).$$

(B1)

Here we denote $\tilde{\omega}$ by $\omega = e\Phi_{12}$. In addition, the limits of integration on $\omega$ are given by $\omega_1 = e\Phi_{12} + \mu_2$ and $\omega_2 = \mu_1$, whereas the limits of integration on $\phi$ are

$$\phi_1 = \begin{cases} \frac{\pi}{2}, & x_i < 1, \\ -\pi/2, & x_i > 1, \\ 0, & x_i = 1. \end{cases} \quad \frac{x_i - 1}{x_i + 1} = \frac{\mu_1 + e\Phi_{12}}{\Delta}.$$

(B2)

In obtaining Eq. (B1), we parametrized $k$ space using coordinates $k_x = k_x \sin \phi/2$ and $k_y = \sqrt{k_x^2 - q^2} \cos \phi/2$, with $q$ conveniently aligned in the $x$ direction. The integration over $k_x$ absorbs the first $\delta$ function, setting $k_x = e\Phi_{12}/h\nu_F$. Integration over $\omega$ absorbs the second $\delta$ function, fixing the limits of integration $\omega_1$, $2\pi$, and $B(\omega_2)$. In this case, the integral for the tunneling current yields

$$\int_{\Phi_{12}} \frac{d\omega}{2\pi} A_{1,ss}(k,\omega) A_{2,ss'}(\mathbf{k} + \mathbf{q},\omega)$$

$$= \frac{2}{\Gamma\Delta} \int_{\Phi_{12}} \frac{d\omega}{2\pi} \int_{\Phi_{12}} \frac{d\omega}{2\pi} \left[ \int_{\Phi_{12}} \left(\frac{\omega(\omega - \Delta)^{1/2}}{\Gamma\Delta} + O(\Gamma/\Delta)\right) \right].$$

(B3)

In obtaining Eq. (B3), we transformed the integral of the spectral functions into a dimensionless integral of the form $I_{\text{res}}(\epsilon) = \int d^2 x |f(x)^2 + e|g(x)^2 + e^{-1}|^{-1}$. The functions $f$ and $g$ satisfy $f(0) = g(0) = 0$ and have a null Jacobian $\det[\partial_x f, \partial_x g] = 0$ (here $\epsilon = \Gamma/\Delta$). It can be shown that $I_{\text{res}} \propto \epsilon^{-1/2}$ when $\epsilon \ll 1$. An expansion to leading order in powers of $\epsilon$ gives Eqs. (B3). Setting $\mu_1 = 0$ in Eq. (B3), the peak current $I_{12}^{(ps)}$ in Eq. (15) of the main text is obtained.