Gate-versus-defect-induced voltage drop and negative differential resistance in vertical graphene heterostructures

Tae Hyung Kim1,2, Juho Lee1,2, Ryong-Gyu Lee1 and Yong-Hoon Kim3✉

To enable the computer-aided design of vertically stacked two-dimensional (2D) van der Waals (vdW) heterostructure devices, we here introduce a non-equilibrium first-principles simulation method based on the multi-space constrained-search density functional formalism. Applying it to graphene/few-layer hBN/graphene field-effect transistors, we show that the negative differential resistance (NDR) characteristics can be produced not only from the gating-induced mismatch between two graphene Dirac cones in energy-momentum space but from the bias-dependent energetic shift of defect levels. Specifically, for a carbon atom substituted for a nitrogen atom (C\textsubscript{N}) within inner hBN layers, the increase of bias voltage is found to induce a self-consistent electron filling of in-gap C\textsubscript{N} states, which in turn changes voltage drop profiles and produces symmetric NDR characteristics. With the C\textsubscript{N} placed on outer hBN layers, however, the pinning of C\textsubscript{N} states to nearby graphene significantly modifies device characteristics, demonstrating the critical impact of atomic details for 2D vdW devices.

ARTICLE

INTRODUCTION

The recent development of van der Waals (vdW) heterostructures based on diverse combinations of two-dimensional (2D) crystals opened up possibilities to explore novel physical phenomena as well as device applications1,2. Here, much effort has been devoted to identify device architectures that maximally utilize unique properties emerging from hybrid forms of 2D vdW materials in non-equilibrium conditions3–5. For example, the junction configuration with few-layer 2D semiconductors or insulators sandwiched between graphene electrodes represents a promising platform from which phenomena such as the chiral quantum states6,7, giant tunneling magnetoresistance8,9, and negative differential resistance (NDR)7,10–12 can be explored.

For the continued development of novel physics and advanced devices based on atomically thin 2D vdW heterostructures, it will be necessary to properly understand how atomic-scale details affect the heterojunction electronic structures and quantum transport properties under finite-bias non-equilibrium conditions. However, the standard first-principle approach combining density functional theory (DFT) and non-equilibrium Green’s function (NEGF) formalism within the Landauer framework13 has intrinsic shortcomings in simulating 2D vdW heterojunctions under finite-bias conditions (Fig. 1a). Specifically, taking as an example the junction configuration in which a few-layer hexagonal boron nitride (hBN) channel is sandwiched by graphene electrodes, the DFT-NEGF formalism requiring electrodes to be repeated semi-ininitely along the transport direction forces one to replace single-layer graphene electrodes with infinite-layer graphite counterparts (Fig. 1b) or to adopt a hBN nanoribbon channel model in combination with edged graphene electrodes (Fig. 1c). Both approaches then suffer from important shortcomings in that graphene and graphite electrodes will behave differently in the former case14–16 and that it will be difficult to avoid edge-shape and width dependences of hBN nanoribbons in the latter case17,18.

Accordingly, most theoretical studies currently resort to semi-classical approaches such as the Bardeen transfer Hamiltonian formalism19,20 to handle the laterally periodic 2D device geometry (Fig. 1d), eliminating the possibility of interpreting and predicting effects that involve atomistic details in an ab initio manner. In this work, extending the recently developed multi-space constrained-search DFT (MS-DFT) formalism27–29, we analyze the non-equilibrium electronic structure and quantum transport properties of vertical graphene/few-layer hBN/graphene field-effect transistors (FETs) at the atomic level. The MS-DFT formalism is established by invoking a microcanonical viewpoint or employing finite-sized metallic electrodes, in contrast to the semi-innite metallic electrodes from the grand canonical picture of DFT-NEGF. We first explain how the MS-DFT formalism can be straightforwardly extended such that for 2D vdW heterojunctions incorporating graphene and other 2D electrodes quantum transport properties including gating effects can be treated within the true vertical 2D FET geometry (Fig. 1d, left panel). Next, applying MS-DFT to graphene/hBN/graphene FETs, we analyze the gate-induced NDR current density-bias voltage \( J \sim V \) characteristic in the pristine hBN channel case and quantify the hBN layer number-dependent quantum capacitance effect of graphene. We then demonstrate how a point defect introduced into the hBN channel critically affects finite-bias junction electronic structure and device operations. Specifically, we consider a carbon atom substituted for a nitrogen atom (C\textsubscript{N}) in an inner hBN layer (Fig. 1d, right panel), and show that it can lead to an NDR signal even in the absence of gating effects. Its origins will be explained in terms of the self-consistent charging and upshifting of in-gap C\textsubscript{N} defect levels with the increase of bias voltage, which is accompanied by the change in the behavior of the spatial electrostatic potential or “voltage” drop profile. On the other, for the C\textsubscript{N} defect introduced into an outer interfacial hBN layer, we find that the NDR peak disappears and the current density decreases by an order of magnitude. We will show that these drastic changes result from the pinning of in-gap C\textsubscript{N} levels to the neighboring graphene states, which translates into a fixed spatial voltage drop behavior. These non-equilibrium atomic details that unravel the mechanism of hitherto
undiscussed defect-induced NDR \( J-V_g \) characteristics hint at outstanding opportunities in the understanding and design of 2D vdW heterojunction devices.

RESULTS

First-principles approach for the simulation of 2D vdW FETs

We previously developed the MS-DFT formalism and demonstrated for molecular junctions that it produces practically equivalent results to those from DFT-NEGF\(^{28,29}\). Within MS-DFT, a microcanonical picture is adopted and a finite-bias quantum transport process is mapped to a drain-to-source excitation process. Here, we explain how the MS-DFT approach can be extended such that (i) the gate voltage \( V_g \) can be included in addition to the bias voltage \( V_d \), and (ii) transmissions at finite \( V_d \) and/or \( V_g \) can be calculated without recovering the Landauer picture or introducing semi-infinite electrodes.

Regarding (i), we reiterate that within the MS-DFT approach the finite \( V_g \) was embodied as a constraint \( eV_g = \mu_L - \mu_R \) for the total-energy minimization. Similarly, after the left electrode L/channel C/right-electrode R/gate G partitioning (see also Methods), the gate bias \( V_g \) can be straightforwardly implemented as an additional constraint of \( -eV_g = \mu_L - (\mu_L + \mu_R)/2 \), where \( \mu_R \) stands for the electrochemical potential of \( G \), which precedes the \( V_g \) constraint. For the practical implementation of this scheme, as schematically shown in Fig. 1d, we adopted an additional Au monolayer as the bottom gate electrode and extended the spatial tracing and occupation constraining of Kohn-Sham states to the gate electrode region. A vacuum space of 10 Å was introduced between the graphene/hBN/graphene junction and the Au electrode to ensure that the gate leakage current is completely suppressed. This corresponds to the 3.9 nm equivalent oxide thickness (EOT) that is defined as

\[
\text{EOT} = \left( \frac{e_{SO_2}}{e_{gate}} \right) f_{\text{gate}},
\]

where \( e_{SO_2} \) and \( e_{gate} \) are the dielectric constants of SiO\(_2\) \( (e_{SO_2} = 3.9 e_0) \) and vacuum \( (e_{gate} = e_0) \), respectively, and \( f_{\text{gate}} \) is the vacuum thickness (1 nm).

To achieve (ii), after obtaining a non-equilibrium electronic structure within MS-DFT calculations, we invoked as a post-processing step the matrix Green’s function formalism\(^{31,32}\) and calculated finite-bias transmission coefficients according to\(^{13}\)

\[
T(E;V_d,V_g) = Tr[a_L M a_R M^T].
\]

Here, \( a_L(R) \) is the spectral function in the L (R) contact and \( M = x_G^2 Gx_R \), where \( x_G^2 \) is the L-C (R-C) electrode-channel coupling matrix (for details, see Supplementary Note). In computing \( a_L(R) \), since electrode unit cells semi-infinitely repeated along the transport direction do not exist anymore for the present vertical 2D vdW heterojunction model (Fig. 1d), we replaced the surface Green’s function \( g_{L(R)} \) with the region L (R) retarded Green’s function \( G \) calculated from the junction model itself (orange boxes in Fig. 1d). In doing so, we introduce a constant broadening factor (~0.025 eV), which originally enters into the Bardeen transfer Hamiltonian approach\(^{19}\), it now properly accommodates the impact of coupling between the channel and electrodes and their atomistic details\(^{13}\). Once the transmission functions were obtained, the current density-bias voltage \( J(V_d,V_g) \) characteristic under \( V_g \) was calculated using the Landauer–Büttiker formula\(^{13}\)

\[
J(V_d,V_g) = \frac{2e}{h} \int_{\mu_L}^{\mu_U} T(E;V_d,V_g) [f(E - \mu_R) - f(E - \mu_L)] dE,
\]
Gate-induced asymmetric NDR characteristics. a The $V_g = 0$ V current density-bias voltage ($J - V_b$) characteristic of the left graphene (LG)/5L hBN/right graphene (RG) junction (left panel) and a schematic of the tunneling-based electron transport mechanism (right panel). Inset: The $J - V_b$ curves for the hBN 2—5L cases are shown on a logarithmic scale. b The $V_g = 0$ V $\Delta\\rho_{\text{H}}$ (upper) and $\Delta J$ (lower) distributions of the LG/3L hBN/RG junction at $V_b = 1.0$ V. Red, purple, and blue down triangles indicate the positions of LG, hBN, and RG layers, respectively. c The $V_g = 0$ V projected bands of the LG/3L hBN/RG junction at $V_b = 1.0$ V. The circle size is in proportion to the weight of wavefunctions projected to the LG (red) and RG (blue) electrodes. Purple down triangles indicate the projections nearby hBN bands. d Gate-dependent $J - V_b$ curves of the LG/3L hBN/RG junction at $V_g = +2.5$ V (blue), $V_g = -0.5$ V (green), and $V_g = -2.5$ V (red). e The $V_g = -2.5$ V $\Delta\\rho_{\text{H}}$ profiles of the LG/3L hBN/RG junction at $V_b = -1.0$ V (left) and $V_b = +1.0$ V (right). Yellow, red, purple, and blue down triangles indicate the locations of the Au, LG, hBN, and RG layers, respectively. f The $V_g = -2.5$ V projected bands (left panels) and $T(E)$ spectra (right panels) at $V_b = -1.0$ V, $-0.7$ V (NDR valley), $-0.5$ V (NDR peak), $0.0$ V, $+0.7$ V. The sizes of circles quantify the strength of orbital contributions of different 2D layers to the LG (red) and RG (blue) electrodes.

where $f(E - \mu) = 1/(1 + \exp((E - \mu)/k_B T))$ is the Fermi-Dirac distribution function.

We implemented these MS-DFT functionalities within the SIESTA package, which has been extensively employed for the DFT-NEGF program development. For the benchmark, because the laterally infinite vertical 2D vdW FET models shown in Fig. 1d can be handled with existing DFT-NEGF codes, we instead considered the graphene/pentalayer (5L) hBN/graphene junction sandwiched by Pt electrodes and graphite/few-layer hBN/graphite junction (Fig. 1b) models that can be also handled within DFT-NEGF. In addition to checking the computational convergence trend, we confirmed the close agreement between MS-DFT and DFT-NEGF transmission data (see Supplementary Figs. 1 and 2).

Gate-induced asymmetric NDR $J-V_b$ curves and linear background currents

We first applied MS-DFT to 2D vdW heterojunctions composed of few-layer pristine hBN channels sandwiched between graphene electrodes, a prototype configuration for the experimental realization of vertical 2D vdW tunneling FETs and the main focus of this work. We then set $V_g = 0$ V and calculated the finite-bias electronic and quantum transport properties. Varying the number of hBN layers $N_{\text{hBN}}$ from $N_{\text{hBN}} = 2$ to $N_{\text{hBN}} = 5$, as shown in the left panel of Fig. 2a, we obtain the linear $J-V_b$ characteristics and an exponential current decrease with increasing $N_{\text{hBN}}$ with a decay constant of 0.718 Å⁻¹. These results are overall in agreement with the mechanism of quantum tunneling in biased graphene/2D semiconductor or insulator/graphene vdW heterojunctions, which has been extensively discussed in previous reports. The constraints of energy and momentum conservation, as schematically shown in the right panel of Fig. 2a, tunneling between two shifted graphene Dirac cones can occur for a single ring of $k$ points. This ring is located at $E = (\mu_k + \mu_L)/2$ under $V_g = 0$ V, and the current will linearly increase as the circumference of the ring linearly expands with $V_g$.

The top panel of Fig. 2b shows for the trilayer (3L) hBN case the plane-averaged bias-induced modification of the Hartree electrostatic potential difference at $V_b = 1.0$ V, calculated...
The corresponding exchange-correlation potential variations are negligible and will not be explicitly discussed below (for details, see Supplementary Fig. 3). In nanoscale junctions, as will be discussed shortly, the behaviors of the quasi-Fermi level drop and the electrostatic potential drop are in general different\cite{28, 39}, and below we will address the latter as the "voltage drop". We also show in the bottom panel of Fig. 2b the corresponding plane-averaged bias-induced electron density rearrangement or the Landauer resistivity dipole \( \Delta \rho (r) = \rho (r; V_b, V_g) - \rho (r; 0, 0) \) calculated according to\cite{27, 28}:

\[
\Delta \rho (r) = \mathcal{P}(r; V_b, V_g) - \mathcal{P}(r; 0, 0).
\]  

For the LG/3L hBN/RG junction, the LG-RG \( \Delta V_H \) offset calculated at \( V_b = 1.0 \) V was 0.52 eV, which is smaller than the applied electrochemical potential difference \( \mu_L - \mu_R = eV_b = 1.0 \) eV. Specifically, the mismatch between the variation of the electrochemical potential and that of the electrostatic potential is a manifestation of the graphene quantum capacitance\cite{42, 43}. To quantify the hBN layer number-dependent quantum capacitance effects, we additionally measured the \( V_b = 1.0 \) V \( \Delta V_H \) values for the tetralayer (4L) and 5L cases and obtained 0.56 and 0.62 eV, respectively (for the 5L hBN case data, see Supplementary Fig. 4).

To further analyze the corresponding non-equilibrium junction electronic structure, we visualize the finite-\( V_b \) band structures of the LG/3L hBN/RG junction projected onto the LG and RG and show them in Fig. 2c. Note that the possibility to straightforwardly analyze finite-\( V_b \) electronic structures using the standard DFT analysis methods is an important strength of the MS-DFT formalism. Then, in addition to the 0.52 eV offset of the LG and RG Dirac cones, we identify as a notable feature that the bands of

\[ J = 10^4 \mu V^2 \]

\[ V_b [V] \]

\[ \Delta V_H (r) = \mathcal{V}(r; V_b, V_g) - \mathcal{V}(r; 0, 0) \]
several adjacent hBN layers (marked by purple down triangles) are projected onto the LG (RG) band.

Next, forming the LG/3L hBN/RG FET model by introducing a bottom gate electrode (gate electrode placed on the LG side), we studied the effects of gating on graphene-based vertical 2D vdW FETs. In the $J - V_b$ characteristics calculated for $V_g = +2.5$ V, $V_g = -0.5$ V, and $V_g = -2.5$ V shown in Fig. 2d, we find in good agreement with previous experimental data\textsuperscript{10} pronounced asymmetric NDR signals superimposed over linear background currents. Partial penetration of an external field is another general characteristic of the quantum capacitance of 2D electron gas\textsuperscript{44}, which results in nontrivial voltage drop profiles. The $\Delta V_{fi}$ curves obtained under $V_g = +2.5$ V (Fig. 2e, top panels) show that a voltage drop between LG and RG of $\Delta V_{fi} = +0.11$ eV already appears at $V_b = 0$ V (see Supplementary Fig. 5) and it decreases (increases) to $-0.42$ eV ($+0.84$ eV) at $V_b = -1.0$ V ($+1.0$ V). This asymmetric $J - V_b$ characteristics can be understood by observing the corresponding $\Delta V$ plots (Fig. 2e, bottom panels): Unlike in the $V_g = +1.0$ V case, the application of $V_g = +1.0$ V with the same polarity as $V_g = -2.5$ V results in the depletion of LG electrons in screening the gate electric field and accordingly a reduced capacity to support the electrochemical potential difference of $V_b = -1.0$ V.

As shown in Fig. 2f, projecting $V_g = -2.5$ V band structures to LG and RG at varying $V_b$ additionally reveals the detailed operation mechanisms of vertical graphene-based vdW tunneling FETs that exhibit NDR features (for the $V_g = +2.5$ V 3L hBN case and the $V_g = -2.5$ V 4L hBN case, see Supplementary Figs. 6 and 7, respectively). As is evident in the $V_b = 0$ V case (Fig. 2f, fourth left panel), the application of $V_g = -2.5$ V induces the hole doping of both LG and RG, and given the bottom gate device geometry, the hole doping level is higher for LG and the misalignment of the LG and RG Dirac cones results in marginal transmission spectra (Fig. 2f, fourth right panel). Then, when $V_b$ increases positively (negatively), the misalignment of LG and RG Dirac cones increases (decreases). Particularly, when $V_b$ increases negatively, the two Dirac cones are gradually aligned until $V_b = -0.5$ V (Fig. 2f, third left panel) and become misaligned again at $V_b < -0.5$ V (Fig. 2f, first and second left panels). The corresponding $T(E; V_g, V_b)$ spectra then show a strong resonant tunneling behavior at $V_b = -0.5$ V (Fig. 2f, third right panel) and a subsequent order of magnitude decrease at $V_b = -0.7$ V (Fig. 2f, second right panel), leading to a nonlinear $J - V_b$ characteristic with a sharp NDR peak (Fig. 2d, red curve). The peak-to-valley ratio (PVR) is about 2, which is in good agreement with the experimentally obtained values (up to $\approx 4$)\textsuperscript{10}. On the other hand, in the $V_b > 0$ V regime, the offset between LG and RG Dirac cones further increases (Fig. 2f, fifth left panel). The widening of the bias window for a nearly fixed $T(E; V_g, V_b)$ spectrum increases the electron density without an NDR signal (but with a bump that occurs when $V_b$ crosses the transmission peak; Fig. 2f, fifth right panel), producing the asymmetric NDR $J - V_b$ characteristics shown in Fig. 2d.

**Symmetric NDR $J-V_b$ curves from hBN point defects**

It should be noted that the $J - V_b$ curves calculated above exhibit asymmetric features such that NDR peaks appear at positive (negative) $V_b$ under positive (negative) $V_g$. However, the symmetric (i.e., appearing in both positive and negative $V_g$ regimes) NDR $J - V_b$ curve that has also been observed in experiment\textsuperscript{10}, cannot be explained in terms of the gating effect. We now incorporate atomic defects which inevitably occur in typical experimental conditions\textsuperscript{45} and show that it provides a hitherto undiscovered mechanism that produces symmetric NDR $J - V_b$ characteristics. Here, we will set $V_g = 0$ V to make it clear that the identified NDR mechanisms are not related with gating effects. In addition, as a representative system, we will focus on the 3L hBN case with a defect in the middle hBN layer. As the point defect, we will adopt a substitutional carbon atom introduced on a nitrogen site ($C_{\text{N}}$), which is known as the dominant atomic defect type for hBN\textsuperscript{10}. Then, as shown in Fig. 3a, we indeed find in the calculated $J - V_b$ characteristics NDR features symmetric with respect to $V_b$. Compared with the $V_g$-induced NDR $J - V_b$ curves shown in Fig. 2d, we find that the NDR peak appears in a lower $V_b$ regime ($0.15$ V) and the PVR is also smaller ($1.42$), whose details are overall consistent with experimental trends\textsuperscript{10}.

In addition to the NDR $J - V_b$ characteristic symmetric with respect to $V_b$, we also point out that as shown in Fig. 3b the $C_{\text{N}}$ defect induces nonlinear voltage drop $\Delta V_{fi}$ profile that nontrivially varies with increasing $V_b$. Generally, we first note that, unlike in the pristine hBN case (Fig. 2b, e), the incorporation of $C_{\text{N}}$ defect results in nonuniformly distributed $\Delta V_{fi}$ profiles. It can be observed that at $V_b = 0.05$ V $\Delta V_{fi}$ drops more abruptly on the left graphene-hBN interface side (Fig. 3b, left panel). However, as $V_b$ increases to $0.4$ V, the region where $\Delta V_{fi}$ more abruptly drops moves to the lower electrochemical-potential right hBN-graphene interface side (Fig. 3b, right panel).

To fully understand the different electrostatic potential drop profiles at low and high $V_b$ regimes and its correlation with the NDR quantum transport characteristics, we analyzed as shown in Fig. 3c the electronic bands projected onto LG, hBN, and RG (left, center, and right panels, respectively). At zero-bias equilibrium state, as detailed in Supplementary Fig. 8b\textsuperscript{46,47}, the strong hybridization between $C_{\text{N}}$ and LG/RG states across one pristine hBN layer allows the charge transfer from LG/RG to $C_{\text{N}}$, making both LG and RG electrodes initially p-doped. Next, applying a small bias $V_g = 0.05$ V, we find that in-gap $C_{\text{N}}$ defect levels are energetically positioned at the lower electrochemical potential $\mu_L$ (Fig. 3c, left panel). The energetic proximity of $C_{\text{N}}$ defect states to $\mu_L$ then translates into their stronger coupling to RG states than to LG counterparts, explaining the more abrupt voltage drop at the LG-hBN interface side\textsuperscript{28}. However, with increasing $V_b$, the defect state is pulled upward to the middle of the bias window at $V_b = 0.15$ V (Fig. 3c, center panel) and eventually to right below $\mu_L$ at $V_b = 0.4$ V (Fig. 3c, right panel). Namely, we find that with increasing $V_b$ electrons tunnel from LG across the left hBN layer into in-gap $C_{\text{N}}$ states located in the middle hBN layer. Upon the continued electron filling, energetically upshifting $C_{\text{N}}$ defect levels eventually approach $\mu_L$ and eventually couple more strongly to LG states than RG states, explaining the more abrupt voltage drop at the hBN-RG interface side $V_b = 0.4$ V (Fig. 3b, right panel).

We now explain how the self-consistent charging of in-gap $C_{\text{N}}$ states induces the NDR $J - V_b$ feature shown in Fig. 3a. In Fig. 3d, the projected density of states (DOS) and transmission spectra corresponding to the band structures of Fig. 3c are presented (left and right panels, respectively). We first note that the $C_{\text{N}}$ defect-originated sharp DOS and $T(E; V_b)$ peaks are initially located at $\mu_L$ at $V_b = 0.05$ V (Fig. 3d, left panel), and with increasing $V_b$ energetically upshift toward $\mu_L$ (Fig. 3b, center and right panels). At this point, we remind that the $V_g$-induced charging of $C_{\text{N}}$ defect states indicates the strong hybridization between in-gap $C_{\text{N}}$ and graphene states as shown from the initial p-doping of LG/RG in equilibrium (Supplementary Fig. 8b), and the strength of $T(E; V_b)$ peaks quantifies the degree of quantum hybridizations. The energetic movement of the DOS and $T(E; V_b)$ peaks from the $\mu_L$ boundary into the middle of the bias window increases the current density until $V_b = 0.15$ V (NDR peak) (Fig. 3d, first and second right panels, respectively; see insets for zoom-in views). However, as $C_{\text{N}}$-originated in-gap defect levels continue to move upward with increasing $V_b$ and approach $\mu_L$ around $V_b = 0.4$ V (Fig. 3d, third left panel), it crosses the RG Dirac point where RG states are deficient and accordingly the transmission peak is reduced both in terms of height and width (NDR valley) (Fig. 3d, lower right panel).
Quantitatively, between the NDR peak ($V_b = 0.05$ V) and valley ($V_b = 0.4$ V), the TE; $V_b(\text{peak height})$ decreases from 0.215 to 0.152 and the full width at half maximum ($>10^{-3}$) decreases from 0.03 to 0.015 eV (insets of Fig. 3d, second and third right panels).

We emphasize that the above-described non-equilibrium device electronic structures and quantum transport properties are significantly modified when the location of the CN defect is different, signifying the importance of atomic details in the operation of 2D vdW electronic devices. Specifically, when a CN defect is introduced into the left interfacial hBN layer (rather than the middle hBN layer as discussed in Fig. 3), we find that currents decrease by an order of magnitude, $J - V_b$ characteristics become asymmetric, and importantly symmetric NDR features disappear (Fig. 4a, see also Supplementary Fig. 9). Analyses of non-equilibrium junction electronic structures show that in this case in-gap CN defect levels are strongly hybridized with LG states and pinned below the electrochemical potential $\mu_L$ (Fig. 4c, d), resulting in voltage drop profiles in which voltage always drops dominantly at the hBN-RG side (Fig. 4b). Bias-independent voltage drop profiles then nullify the above-described mechanism of defect-induced symmetric NDR $J - V_b$ characteristics, explaining the disappearance of NDR peaks. Moreover, as in the middle-hBN-layer CN case (see Supplementary Fig. 8b), charge transfers to the CN located in the left hBN layer induce at the equilibrium condition ($V_b = 0$ V) the strong and weak p-type doping of LG and RG electrodes, respectively (Supplementary Fig. 8c). Accordingly, unlike in the positively increasing $V_b$ case, the negatively increasing $V_b$ can align the single rings of k points in the LG and RG Dirac cones within the bias window and produce larger currents (see Supplementary Fig. 9).

**NDR characteristics from gating vs quantum hybridization**

We close by providing mechanistic insights into the NDR characteristics originating from the gating-dependent tunneling and the quantum hybridization, whose differences are schematically summarized in Fig. 5a, b, respectively. The quantum-hybridization NDR is a general device concept involving the bias-induced breaking of an initially delocalized quantum-hybridized state, and we previously predicted that an NDR with PVR >10 can be achieved using atomically thin vdW nanowires.
Fig. 5  Schematics of the NDR mechanisms based on the gating and quantum hybridization. a Negatively gated ($V_b = -2.5$ V) pristine 3L hBN case at the NDR peak ($V_b = -0.5$ V, left panel) and NDR valley ($V_b = -0.7$ V, right panel). b Defective 3L hBN case (central CN defect) at the NDR peak ($V_b = 0.15$ V, left panel) and NDR valley ($V_b = 0.4$ V, right panel). The red and blue solid lines indicate the LG and RG Dirac cones, respectively, and green arrows represent the electron flow. The chemical potentials of LG ($\mu_L$) and RG ($\mu_R$) that define the bias window with the voltage $V_b$ are indicated by the red and blue horizontal dotted lines. In (b), pink (cyan) dashed lines represent the hybridization of LG (RG) states with the CN defect level (black solid line). The local DOS for the pristine 3L hBN case at $V_b = -0.5$ V (left panel) and $V_b = -0.7$ V (right panel), and for the defective 3L hBN case at $V_b = 0.15$ V (left panel) and $V_b = 0.4$ V (right panel). The local DOS were obtained from the energy levels corresponding to transmission peaks (marked as ①, ②) in the third and second panels of Fig. 2f, respectively, for the pristine 3L hBN case, and as ①, ② in the second and third panels of Fig. 3c, respectively, for the defective 3L hBN case. Energy windows of $[-0.05 \pm 0.05$ eV] around the transmission peaks and the isosurface value of $1 \times 10^{-5}$ states Å$^{-4}$ were adopted for the pristine (defective) 3L hBN case.

with near-Ohmic lateral homojunction contacts. To reinforce the point defect-mediated quantum-hybridization NDR mechanism in graphene-based vertical 2D vdW heterojunctions, we visualize in Fig. 5c, d the NDR peak (left panels) and valley (right panels) local DOS around transmission peaks obtained from the pristine and CN defect-hBN cases, respectively. It can be noted that in the pristine hBN case (Fig. 5c) the in-phase (out-of-phase) local DOS are mainly located at the LG and RG electrode regions at the NDR peak (valley), reaffirming the tunneling mechanism. On the other hand, in the CN defect case (Fig. 5d), the local DOS have significant weights within the defective hBN region. Furthermore, the NDR peak (valley) is generated from the high (low) level of wavefunction delocalization, as evidenced by the presence (absence) of local DOS weight on the RG region at positive $V_b$ regimes.

**DISCUSSION**

To summarize, based on the recently developed MS-DFT formalism, we established an ab initio approach that enables the faithful modeling and simulation of 2D vdW FETs under finite-voltage operating conditions. The extended MS-DFT formalism then allowed in-depth atomistic analyses of two different NDR mechanisms that can emerge from graphene-based vertical 2D vdW FETs, for which theoretical studies have been so far mostly carried out at the semiclassical level employing the Bardeen transfer Hamiltonian formalism. Our explicit non-equilibrium first-principles calculations first quantified the finite gate/bias-voltage-dependent energy-momentum mismatches between two graphene Dirac cones, which are affected by the graphene quantum capacitance and produce asymmetric NDR $J-V_b$ characteristics. While this standard NDR mechanism concerns graphene electrodes, MS-DFT calculations revealed that atomic defects present within the hBN channel can also provide a hitherto undiscovered NDR mechanism. Specifically, we observed that for the CN defect placed in the middle hBN layer its in-gap defect levels are self-consistently charged and upshifted with increasing $V_b$. Corresponding to bias-dependent changes of voltage drop profiles, the upshift of in-gap CN levels then reduced (enhanced) the coupling between CN and lower (higher) electrochemical-potential graphene states and produced symmetric NDR $J-V_b$ curves. On the other hand, with the CN defect placed at an outer interfacial hBN layer, in-gap CN levels were pinned to adjacent graphene states or voltage drop profiles became bias-independent. These electronic structure characteristics then led to an asymmetric $J-V_b$ curve in which NDR peaks are absent and currents are an order of magnitude smaller, signifying the critical importance of atomic-level details in the operation of 2D vdW devices. Our findings fill in the missing link in the present understanding of finite gate- and/or bias-voltage operations of vertical 2D vdW FETs, and also point towards future directions of computer-aided design of atomically thin 2D nanodevices.

**METHODS**

DFT and MS-DFT calculations for graphene-based vertical vdW heterojunction FETs (Fig. 1a) were carried out using the SIESTA package. The local density approximation, Troullier-Martins-type norm-conserving pseudopotentials, numerical atomic orbital basis sets of double-$\zeta$ plus polarization quality defined with the confinement energy of 100 meV, and 300 Ry of real-space mesh grid cutoff energy were adopted. For the pristine graphene/pristine hBN/graphene junction cases, we used the graphene and hBN unit cell models with numbers of hBN layers from two to five and a 130 x 130 x 1 Monkhorst-Pack k-point grid sampling. Details of the k-point grid convergence test in terms of the finite-bias transmission $T(E; V_b = 1.0$ V) of a graphite/hBN/graphite junction model are described in Supplementary Fig. 1c. For the graphene/defective hBN/graphene junction case, we employed a 5 x 5 supercell graphene/hBN/graphene model and introduced a single CN atomic defect into the middle or interfacial hBN layer, which translates into a defect density of 7.63 x 10$^{13}$ cm$^{-2}$. After performing a convergence test again in terms of the finite-bias transmission $T(E; V_b = 0.6$ V) (see Supplementary Fig. 10),
we employed a 30 × 30 × 1 Monkhorst-Pack k-point grid. To avoid interactions of the slabs with their periodic images, a thick vacuum space of more than 15 Å was included. Equilibrium junction electronic structures and quantum transport properties were analyzed following refs 6,47, and the results are presented in Supplementary Fig. 8.

To perform finite-bias voltage $V_{b}$ MS-DFT calculations,7–9 we classified the top graphene, central h-BN, and bottom graphene as the left electrode (L), channel (C), and right electrode (R), respectively, and traced the spatial origins of each Kohn-Sham wavefunction to L or C or R. We then introduced within MS-DFT the electrochemical potentials of the L and R electrodes, and $V_{b}$ as constraints for total energy minimizations.

**DATA AVAILABILITY**

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

**CODE AVAILABILITY**

The related codes are available from the corresponding author on reasonable request.

Received: 27 August 2021; Accepted: 16 February 2022; Published online: 25 March 2022.
AUTHOR CONTRIBUTIONS
Y.-H.K. developed the theoretical framework and oversaw the project. T.H.K., J.L., and R.-G.L. implemented the method and carried out calculations. All authors analyzed the computational results, and Y.-H.K. wrote the manuscript with significant input from T.H.K. and J.L.

COMPETING INTERESTS
The authors declare no competing interests.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-022-00731-9.

Correspondence and requests for materials should be addressed to Yong-Hoon Kim.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022