Hot-Electron Transistors for Terahertz Operation Based on Two-Dimensional Crystal Heterostructures

Byoung Don Kong,[*] Zhenghe Jin, and Ki Wook Kim[†]

Department of Electrical and Computer Engineering, North Carolina State University,
Raleigh, North Carolina 27695-7911, USA

(Received 27 June 2014; revised manuscript received 1 October 2014; published 10 November 2014)

This work illustrates the feasibility of ultra-high-frequency operation in a vertical three-terminal electronic device by exploiting the advantages of two-dimensional (2D) crystal heterostructures. The proposed device utilizes a gapped 2D material as the tunnel barrier between a graphene base and a metallic emitter, while the Schottky contact with an *n*-type substrate forms the base-collector junction. The ultrathin active region formed by the 2D insulating and semimetallic layers ensure the required subnanometer-scale control in the thickness and the lateral uniformity, overcoming the major limitations of conventional 3D materials. Atomistic simulations based on the coupled density functional theory and nonequilibrium Green's function method, in combination with an equivalent device model, clearly reveal well-defined transistor characteristics with a good current drive. The analysis also points out the prominence of emitter material selection for superior performance. With proper optimization, the device is capable of reaching the intrinsic cutoff frequencies over a few THz even under realistic constraints, indicating a technological pathway beyond the current limit.

DOI: 10.1103/PhysRevApplied.2.054006

I. INTRODUCTION

Vertical tunnel structures have inherent advantages in the high-frequency performance due to the short transit time. With the experimentally observed characteristic, tunneling times of the order of 10^{-15} s [1], the operating frequencies as high as 30 THz have been demonstrated in the metal-oxide-metal diodes for instance [2,3]. One of the primary obstacles in realizing a practical device, albeit the decades of concerted effort, lies in the challenges of fabricating uniform, ultrathin tunnel barriers with subnanometer-scale accuracy. As the tunneling depends exponentially on the barrier thickness, even a small variation could cause current crowding locally in a single device, not to mention the large disparity in the device-to-device characteristics. Further, unpredictable and uncontrollable interface states at the boundaries between conventional three-dimensional (3D) materials pose additional challenges.

The possibility of using a 2D crystal semiconductor as an atomically thin tunnel barrier was explored earlier [4], including a theoretical investigation of a structure analogous to the metal-insulator-metal diode [5]. The 2D crystals, also known as van der Waals crystals, have strong chemical bonding only between the atoms on the same layer, while the interactions between the layers are via the weak van der Waals force. Consequently, layer-by-layer exfoliation or epitaxial growth of dissimilar materials appears possible, making them an ideal candidate for applications requiring atomistic control in the thickness as well as the lateral uniformity. For instance, a number of 2D crystals including hexagonal boron nitride (*h*-BN) and transition metal dichalcogenides can be seamlessly integrated with graphene in a system without causing defects, interfacial scattering centers, or degradation of material properties. With sizable band gaps, they provide perfectly uniform 2D insulators, while graphene can play the role of a highly conductive channel or electrode. At the same time, it has been demonstrated experimentally that various metallic materials can be also incorporated into the layered system without significantly altering the characteristics of 2D crystals.

By exploiting the unique properties of this material system, we propose and theoretically examine a 2D crystal heterostructure that may be capable of information processing in the THz frequencies. The envisioned device structure is outlined in Sec. II, followed by a comprehensive description of the theoretical treatment in Sec. III. More specifically, an atomistic model that combines the first-principles density functional theory (DFT) calculation with a nonequilibrium Green's function (NEGF) method on the same basis is adopted to accurately account for the essential physical details in the subnanometer scale. The calculation results are analyzed in Sec. IV, illustrating the significant potential impact of the proposed concept for realizing active nonlinear devices in a frequency range that is beyond the reach of conventional semiconductor counterparts.

II. ENABLING DEVICE CONCEPT

One device structure that can fully utilize the advantages described above for ultra-high-speed operation is a 2D

---

[*] Present address: U.S. Naval Research Laboratory, 4555 Overlook Avenue, Washington, D.C. 20375, USA.

[†] kwk@ncsu.edu

crystal version of a hot-electron transistor (2D HET). As shown in Fig. 1(a), the specific design adopts the graphene–2D-insulator–metal combination for the base-emitter (BE) junction, whereas an *n*-type semiconductor forms the collector inducing a Schottky barrier at the interface with the semimetallic graphene base (of a few monolayers thick). Among potential choices for the atomically thin barrier, *h*-BN is probably the closest to an ideal 2D insulator with all of the atoms arranged on the same plane. While the band gap of approximately 4–5 eV (or higher) may limit the magnitude of the tunneling current [6], it is not only chemically stable but also exhibits good affinity with graphene. Alternatively, the use of MoS_2 or the like could benefit from generally lower barrier heights (because of much smaller band gaps) but at the expense of more complex interfacial structures (due to multiple atomic planes in a unit cell and mismatch with graphene) [5]. Both possibilities are expected to produce superior performance over the structures relying on conventional tunnel oxides [7–9]. An additional advantage of 2D crystal insulators is that they do not induce the conductance degradation in the grapheme base unlike the graphene-oxide interface [10]. In the case of the emitter electrode, a metal with a less reactive surface characteristic is preferred in order to maintain the 2D crystal quality. In this regard, materials such as Cu, Au, and Ag (that tend to form physisorbed interfaces) are promising candidates. Their compatibility is further illustrated by the reported synthesis of 2D crystals utilizing some of the very metals as the substrate [11–13].

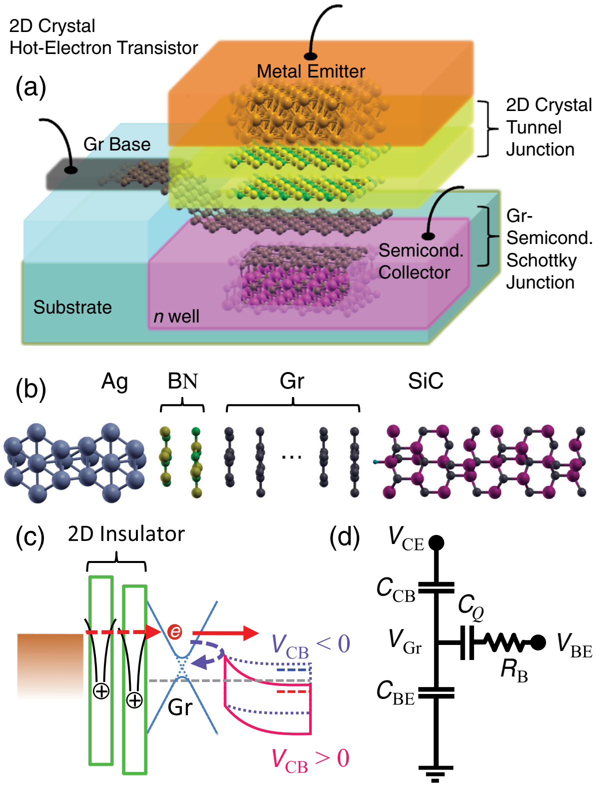

FIG. 1. (a) Two-dimensional HET design with a possible layout for fabrication. (b) Cross-sectional view of the 2D HET layers using the Ag/*h*-BN/Gr/SiC combination as an example. The dangling bonds at the SiC surface are terminated by the hydrogen atoms. (c) Schematic illustration of band diagram and electron transport across the heterogeneous interfaces. (d) Equivalent circuit model of the 2D HET device. The Schottky junction capacitance C_{CB} and the quantum capacitance C_Q vary as a function of the bias.

As for the collector-base (CB) junction, SiC could be one realistic choice for the collector when considering successful, wafer-size epitaxial growth of graphene on this material [14]. Application of a 3D semiconductor here in place of another 2D insulator obviates the need to dope the graphene base very heavily for a sufficient current drive. Rather, the desired control of the junction properties can be readily achieved through the doping in the 3D semiconductor. Accordingly, the choice of emitter electrode material and collector doping determines the overall band alignment in the 2D HET, avoiding severe altering of the carbon lattice in the base region; the device performance can be optimized primarily by adjusting these two factors. Note that the application of a Schottky junction at the CB barrier has already been associated with the performance enhancement in the conventional 3D HETs [15].

III. MODELING APPROACH

To evaluate the proposed device performance with the required atomistic details, we take advantage of the NEGF method combined with the first-principles Hamiltonian based on DFT. The DFT is applied first to characterize the electronic structures accurately, which is followed by the NEGF implementation to analyze carrier transport on the same basis. The SIESTA and TRANSIESTA packages [16] are used with the pseudopotentials in the local density approximation. Hence, the intricate specifics of the studied structure, such as the interfacial microstructures or states and crystal orientations, are fully accounted for. The conventional treatments based on the band-edge concept and/or the effective mass are simply inadequate to describe the complexities of the present device, where the bulklike periodicity is completely absent in the direction of carrier transport with differing crystalline morphism in each layer.

In the actual computation, the BE and CB junctions are treated separately due to the size of the combined calculation; the performance of the entire device is estimated subsequently by utilizing a formalism that couples two characteristics in a systematic, self-consistent manner, as described later in the discussion. All of the considered heterostructure geometries are first optimized by minimizing the interatomic forces. In the case of the BE junction, the *h*-BN and graphene layers are built on the (111) surface of Cu, Ag, or Au crystals with the 1 × 1 (Cu) or 2 × 2 (Ag, Au) construction, respectively [17]. The lattice constants of each layer are well matched; Cu has almost the same lattice constant with h-BN/graphene, while Ag and Au experience negligible strain (less than 2% for both). To mimic a 3D electrode, a sufficient thickness of metal is used for the emitter (nine atomic layers for Cu and six atomic layers for Ag and Au). On the other hand, the h-BN tunnel barrier is assumed to be only two monolayers thick. Both h-BN and graphene take the form of AB stacking. When another 2D insulator is used in place of h-BN, the construction of the BE junction can be achieved by similar considerations. Concerning the CB junction, n-doped 4H-SiC with two 4H periods represents the substrate, followed by the AB-stacked graphene that is arranged on the Si phase. Consistent with epitaxially grown films, the first carbon layer forms a buffer layer with non-sp^2 bonding. The dangling bonds of the top Si atoms are terminated with hydrogen, emulating the surface-annealing treatment in the fabrication process [18]. In the simulations of both BE and CB junctions, the base is chosen relatively thick (i.e., four monolayers) to ensure graphenelike characteristics at least in the middle part of the region, despite the presence of heterointerfaces at both sides. This is also necessary to avoid potential mismatches when superimposing two separately obtained structures into one. The $12 \times 12 \times 1$ Monkhorst-Pack grids are used for all DFT simulations with the pseudopotentials in the local density approximation. Moreover, the double-ζ plus polarization basis sets are applied for B, C, and N, while the single-ζ counterparts are assumed for Cu, Ag, and Au. The atomic configurations of the BE and CB interfaces are illustrated in Fig. 1(b).

Handling a Schottky contact in the DFT scheme requires an extra effort to introduce the effect of impurity atoms in the SiC substrate. While the interface between an intrinsic semiconductor and graphene forms a Schottky contact without doping that can also function as a rectifier, the current level is too low for a practical device. Accordingly, a substantial level of doping appears necessary. We adopt a virtual crystal approximation that has been tested successfully in Si p-n diodes [19]. More specifically, the pseudopotential of a virtual atom that is a hybridization between the silicon and dopant atoms (in this study, phosphorus) with a proper weight is generated by the atomic pseudopotentials engine (APE) [20] and used in place of Si to emulate the n-doping effect on the SiC substrate. We assume the concentration of 1×10^{19} cm^{-3} to reduce the Schottky barrier height sufficiently. The effects of doping can be verified by examining the location of Fermi energy as a function of "impurity" concentration.

With the DFT-determined physical geometry, the transport analysis is approached primarily by the atomistic NEGF formalism that is coupled self-consistently with the DFT Hamiltonian via the Poisson equation. Because of the computational requirement of the three-terminal device indicated earlier, we also apply an equivalent device model as shown in Fig. 1(d) to supplement the calculation. Solving the entire structure in the combined DFT-NEGF scheme for full self-consistency, while ideal, is very difficult to achieve at the moment. The adopted approach provides an adequate alternative as the space-charge characteristics (particularly, the base charges) are not significantly affected by dynamic transport of the carriers injected from the emitter at the specified terminal biases. Not only is electron transport in the base essentially ballistic considering its atomic thickness, those crossing the CB junction are also removed swiftly by the collector electrode due to the favorable electric field. In this picture, the 2D HET under a steady-state condition can be described by a capacitive network. Along with the parallel-plate capacitance (C_{BE}) and the Schottky junction capacitance (C_{CB}) for the BE and CB, another necessary component is the quantum capacitance C_Q that takes into account the interaction between the graphene base channel and the base contact [21]. To induce the charge inside the graphene channel, the Fermi energy must deviate from its reference point. The corresponding potential difference within the base can be interpreted as a capacitive component whose magnitude becomes comparable to those of other contributions in the 2D electron-gas system. For the massless Dirac fermions, the quantum capacitance is given as [22]

$$C_Q = \frac{2q^2 k_B T}{\pi(\hbar v_F)^2} \ln\left[2\left(1 + \cosh\frac{qV_{\mathrm{ch}}}{k_B T}\right)\right], \tag{1}$$

where V_{ch} $(= V_{\mathrm{BE}} - V_{\mathrm{Gr}})$ is the channel potential, q, k_B, $\hbar$, and v_F are the electron charge, Boltzmann constant, Plank's constant, and Fermi velocity, respectively, and T is the temperature. While C_{BE} is a constant, C_{CB} and C_Q vary as the bias changes. Particularly, the dependence of C_{CB} on V_{BE} and V_{CE} (deduced from the characteristics of the graphene/n-SiC Schottky junction) can be considered a feedback mechanism for self-consistency as in the hybrid-π model of the conventional bipolar junction transistor (BJT). In the hybrid-π model, the influence of the collector bias change is usually described by a feedback capacitance between the base and the collector that leads to the Miller effect and Miller capacitance [23]. A steady-state solution for V_{Gr} is found from the nonlinear equations of the capacitive network, determining the space charge in the graphene base for the given V_{BE} and V_{CE}. This reduces the problem to modeling of the electron transport along the emitter-collector electrodes [i.e., the direction shown in Fig. 1(b)] and, subsequently, the collector current as described below.

Specifically, the transmission probabilities at the BE and CB junctions are calculated, respectively, by the self-consistent DFT-NEGF Hamiltonian in accord with the obtained node voltages. Then the results from each junction are combined formally to estimate the performance of the 2D HET. Given the constraints, the treatment offers a practicable solution as the dynamic coupling between the second-nearest electrodes are negligible. For instance, the emitter-collector interaction is effectively shielded by the semimetallic graphene base as well as the emitter tunnel barrier and the collector depletion region. In the implementation, the symmetric double-electrode approach is adopted for the base contact following the work of Büttiker [24] to avoid the fictitious resistance arising from phase breaking. Then the scattering matrix formalism gives the effective transmission probability between the emitter and the collector as [25]

$$T_{\mathrm{eff}} = \frac{T_{\mathrm{BE}} T_{\mathrm{CB}}}{1 - R_{\mathrm{BE}} R_{\mathrm{CB}}}, \tag{2}$$

where T_{BE} (R_{BE}) and T_{CB} (R_{CB}) are the self-consistent DFT-NEGF-obtained transmission (reflection) probabilities at the BE and CB junctions, respectively. Finally, the collector current can be calculated readily by the Landauer-Büttiker treatment; i.e., $I = \frac{2e}{h}\int [f_{\mathrm{E}}(E) - f_{\mathrm{C}}(E)] T_{\mathrm{eff}}(E) dE$, where f_{E} and f_{C} represent the electron distribution at the emitter and collector electrodes. Because of the ballistic assumption in the ultrathin graphene layer, the steady-state base current is considered negligible or zero once the base charge is established at a given bias condition (see, also, the above discussion on C_Q that accounts for base charging). This is analogous to the design requirement of conventional BJTs or HETs, where the collector current must be nearly equal to the emitter current in the steady state (i.e., small base current). On the other hand, if electrons suffer substantial energy relaxation (via scattering) during the base transport, the resulting base current must be considered explicitly. While the present model can actually include this contribution, as the formulation is designed for a multiterminal device, additional investigation is needed for vertical transport of "hot electrons" in few-layer graphene.

IV. RESULTS AND DISCUSSION

Figure 2 summarizes the calculated current-voltage (I-V) characteristics. In addition to the structures with the bilayer h-BN barrier (i.e., Ag/h-BN, Au/h-BN, Cu/h-BN), a case of bilayer $\mathrm{MoS_2}$ as the 2D insulator is also considered for comparison with the Cu emitter (i.e., Cu/$\mathrm{MoS_2}$). The impact of different material combinations is evident from the disparate magnitudes in the collector current. For instance, the HET with the Ag or Au emitter shows a current level of approximately a few $\mathrm{mA/cm^2}$ [Figs. 2(a) and 2(b)], whereas this number drops by 4 to 5 orders of magnitude to less than 0.1 $\mu\mathrm{A/cm^2}$ when Cu is used instead [Fig. 2(c)]. Needless to say, a proper choice of the injecting contact is very crucial.

The origin of this large variation can be found from the energy dispersion and crystal symmetry (mis)match of two constituent materials. Because of the extra degree of freedom, the E-k dispersion of a 3D material forms broader bands than those of a 2D crystal counterpart. However, unlike the initial expectation, our analysis suggests that the extent of detailed alignment still plays an important role even in the tunneling between the 3D and 2D energy bands. In the case of the Cu(111) surface, the lattice constant is well matched with that of graphene or h-BN. As such, the first Brillouin zone in each layer is well aligned at the BE junction. However, the respective locations of the mobile carrier in the momentum space are apart with little overlap. More specifically, graphene has the band extrema at the K point while the 2D equivalent counterpart of Cu projected on the (111) surface finds them around the zone center (see, for example, Fig. 1 of Ref. [5]). The resulting mismatch in the momentum states significantly reduces the tunneling transition probabilities between the Cu and graphene layers. On the other hand, Ag and Au show a much different picture when interfaced with graphene even though their bulk band structures are similar to those of Cu. As illustrated in Fig. 3(a), our calculation strongly suggests that the electrons (more precisely, the electronic states near the Fermi level) in both the Ag (or Au) contact and the graphene channel at the BE junction are comparatively well aligned in the momentum space of the corresponding $\sqrt{3} \times \sqrt{3}\ R30^\circ$ structure. The close proximity of two states that are normally apart in the bulk band structures can be explained by the so-called zone-folding effect. Figure 3(c) schematically elucidates this aspect, where the symmetry point M of bulk Ag crystal is projected near the K point of graphene in the Brillouin zone of the supercell. In fact, this overlap with graphene electron states becomes much more apparent, as indicated in Fig. 3(b), once a thick bulklike metal is considered for the emitter; the

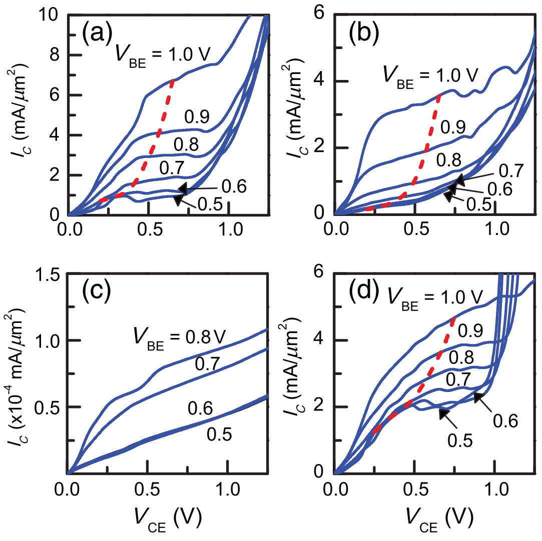

FIG. 2. Calculated current-voltage characteristics of 2D HET. The considered device structures are (a) Ag/h-BN/Gr/SiC, (b) Au/h-BN/Gr/SiC, (c) Cu/h-BN/Gr/SiC, and (d) Cu/$\mathrm{MoS_2}$/Gr/SiC.

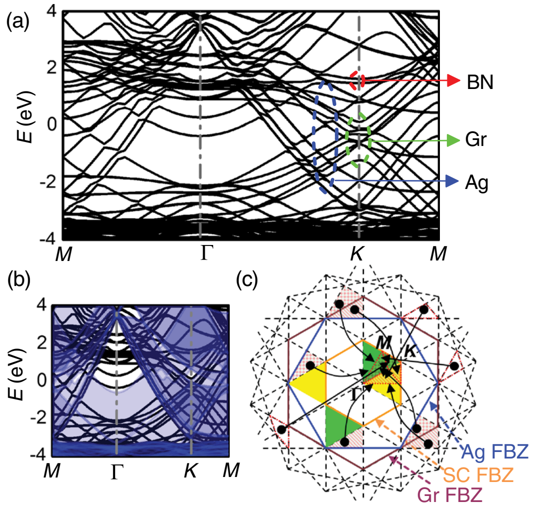

FIG. 3. (a) Band structures of the Ag/*h*-BN/Gr heterojunction drawn along the supercell symmetry lines. The respective origins of the electronic states are specified for the arrowed regions. (b) Projected bulk band structure of Ag(111) superimposed on the interface band structure. The shaded area represents the projected Ag(111) bulk band structure. A much wider range of graphene energy bands near the Dirac point overlaps with the Ag energy bands compared with the Cu case (see Fig. 1 of Ref. [5]). In (a) and (b), the Fermi level is used as the reference for all energy scales, i.e., $E_F = 0$. (c) First Brillouin zones (FBZ) of Ag, graphene, and supercell along with the illustration of the zone folding following the supercell symmetry.

current adoption of six atomic layers for the Ag (Au) contact is simply for computational convenience. Accordingly, an added advantage of the Ag/Gr or Au/Gr combination (where Gr stands for graphene) is that the tunneling current could be rather insensitive to the possible misorientation between the emitter and the base layers. This is a potentially critical point in realistic devices where the loss of interlayer registry appears likely.

One interesting consequence of the findings described above is that the same type of disparity can occur at the metal-graphene Ohmic contact without the tunnel barrier. Assuming that the metal surface is clean and crystalline, the contact resistance of the Cu/Gr interface could be higher than those of Ag/Gr or Au/Gr due to the momentum mismatch between the conduction electrons. Hence, Ag or Au may be preferred as the contact material. This observation is consistent with recent measurements showing an order of magnitude difference in the contact resistance between the Au and Cu electrodes [26,27], Note, also, that the low current level of the Cu/Gr system can actually be useful in applications requiring low leakage. For instance, Cu may prove to be a better choice for the gate electrode in a graphene field-effect transistor. When the gate-dielectric thickness decreases and the tunnel leakage becomes an issue, the Cu electrode can provide an additional "effective" barrier through the momentum-space misalignment with the graphene electron states. In the case when a metal is chemisorbed and interfacial states are formed, the consideration based on the momentum conservation no longer applies.

Along with the proper alignment in the momentum space, another crucial factor for optimizing the device performance is the tunnel barrier height. As mentioned earlier, MoS_2 can be a promising alternative to *h*-BN in this regard, whose barrier height is approximately 1.6 eV in the bilayer form. The result in Fig. 2(d) clearly illustrates the potential promise of the lower barrier height. Even with the momentum mismatch, the current level of the Cu HET becomes comparable to those of Ag HET and Au HET by substituting bilayer MoS_2 for *h*-BN as the emitter tunnel barrier. Likewise, a similar increase of current can be expected for Ag HET and Au HET when the barrier material is switched to another 2D dielectric with a lower energy gap. One caution is that the lower barrier must not be at the expense of band alignment between the electrons in the emitter and the graphene channel.

A systematic investigation of the *I*-*V* characteristics shown in Fig. 2 suggests three distinctive operation regimes for the proposed 2D HETs, namely, linear, saturation, and breakdown. At a small collector bias V_{CE}, the collector current I_{C} goes up linearly with the bias as the increased voltage is applied primarily to the CB junction, thus, lowering the CB barrier. While the potential drop between the graphene channel and the emitter [defined as V_{Gr} according to Fig. 1(d)] is also affected to be precise, its variation appears to be small once V_{BE} is fixed. Hence, the electron injection from the emitter is nearly constant. The linear increase in I_{C} comes from the improved transmission across the CB junction through the electrostatically reduced Schottky barrier [see, for example, $V_{\mathrm{CB}} < 0$ in Fig. 1(c)]. As V_{CE} approaches to V_{Gr}, the device reaches the onset of the saturation regime. In this case, the CB Schottky diode is essentially at zero bias as $V_{\mathrm{Gr}} \approx V_{\mathrm{CE}}$ that is marked by the dashed lines in Fig. 2. Further increase of V_{CE} beyond the point does not affect the current level significantly since the CB junction no longer functions as an effective energy filter; the electrons injected from the BE side are essentially all collected by the collector electrode. The small increase in I_{C} can be attributed to the minor adjustment in V_{Gr} as well as the Schottky effect at the reverse bias. Once the CB barrier becomes completely transparent to the injected electrons (amounting to a short-circuit condition), all of the applied bias appears at the BE diode and the current increases abruptly. The device is now in the breakdown regime.

It is also revealing to examine the device operation from the viewpoint of charge control. In the linear regime when $V_{\mathrm{BE}} > 0$ and $V_{\mathrm{Gr}} > V_{\mathrm{CE}} > 0$, the graphene base is positively biased or charged. Then, an equal amount of opposite charges are induced in the other parts of the device (i.e., the emitter side and/or the collector space-charge region). As the collector voltage goes up, the thickness of the positively charged collector space-charge region increases (i.e., a larger depletion region of the n-doped SiC). Likewise, the charges in the graphene base decrease as the difference between V_{Gr} and V_{BE} becomes smaller. At the onset of saturation, V_{Gr} is essentially the same as V_{CE}. A further increase in V_{CE} sees the continued widening and shrinking of the space-charge region and the base charge, respectively, but the rate of change with respect to the bias slows. This feature, which corresponds to the feedback effect mentioned earlier, follows the square-root dependence of the space-charge region thickness on the applied bias in a uniformly doped substrate. One discrepancy from the current understanding is that the saturation behavior in the Au HET occurs at substantially smaller voltages than the expected $V_{\mathrm{Gr}} = V_{\mathrm{CE}}$ line unlike the other structures. The reason for the deviation is unclear at the moment and may require additional investigation.

Our analysis thus far indicates that the operating principles of 2D HETs are close to those of junction field-effect transistors (JFETs) rather than BJTs, even though the HET concept originated from the latter. The observation is based on a couple of findings. First of all, the injected carrier distribution is governed by the BE junction bias, whereas the CB junction simply works like a filter for the injected electrons as mentioned above. This is similar to the JFET operation, where the gate bias determines the width of the channel, and the drain bias (in reference to the gate bias) controls the extent of the depletion barrier (thus, the electron flow) near the end of the channel. The second point is the weak nonlinearity of the BE junction that is much smaller than the corresponding property in BJTs. The I-V characteristics at the BE junction of a BJT is an exponential function of applied bias, while it is of polynomial order at the metal-insulator-graphene diode of a 2D HET.

To elucidate the device characteristics in greater detail, we calculate the transconductance g_m as well. The results shown in Fig. 4(a) for $V_{\mathrm{BE}} = 0.8$ V suggest that Ag HET and Au HET provide better performance. In fact, both structures indicate a generally larger g_m than the Cu HET with the h-BN or the $\mathrm{MoS_2}$ barrier (even when the latter has a comparable current level). Clearly, a proper choice of the emitter metal enhances not only the absolute magnitude of tunneling carrier injection but also the nonlinear dependence on the bias voltage. This point can be further confirmed by comparing the nonlinearity parameters evaluated at the corresponding BE junctions. The nonlinearity defined in a normalized form as $\frac{d^2I}{dV^2}/\frac{dI}{dV}$ is estimated by fitting the I-V curves to a third-order polynomial $I = a + bV + cV^2 + dV^3$. As expected, the results plotted in Fig. 4(b) associate the Ag and Au emitters with a larger nonlinearity. Another point to note is the comparison between the cases of h-BN and $\mathrm{MoS_2}$ barriers in the Cu HETs. The observed enhancement of the latter ($\mathrm{MoS_2}$) over the former (h-BN) can be attributed to its larger width since a thicker barrier generally yields larger higher-order terms. Since a similar trend can be anticipated in the structures with the Ag or Au emitter, it is possible to achieve an even higher current drive and nonlinearity simultaneously than those in Fig. 4 if a suitable 2D semiconductor with a lower gap can be used for the tunnel barrier.

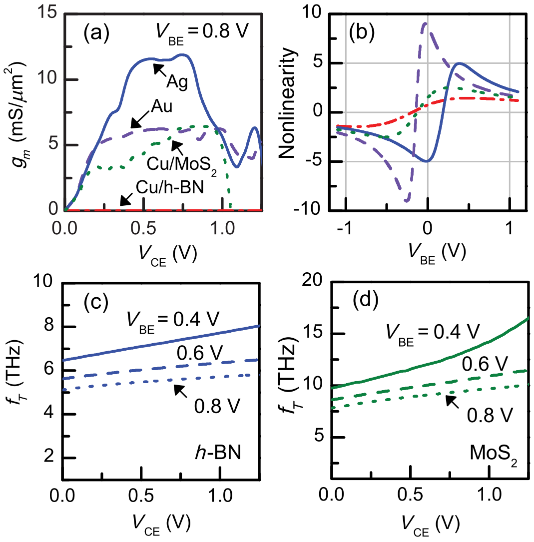

FIG. 4. (a) Calculated transconductance of 2D HETs as a function of collector bias, while V_{BE} is fixed at 0.8 V. (b) Comparison of nonlinearity $[\equiv (d^2I/dV^2)/(dI/dV)]$ at the base-emitter junction for different emitter electrode and insulating barrier choices. The line types and colors are identical to those in (a) to denote different cases. (c), (d) Calculated base switching cutoff frequency when bilayer h-BN or bilayer $\mathrm{MoS_2}$ is used as the base-emitter tunnel barrier, respectively.

Utilizing the characteristics obtained above, we can extend the investigation to the high-frequency performance of the 2D HETs. A complexity here is that the small-signal expressions commonly used for unity current gain and maximum oscillation frequencies in such devices as BJTs [28] may not be applicable as a number of the underlying assumptions in the models are no longer valid (e.g., diffusive vs ballistic transport). Following the analysis on conventional ballistic HETs [29], we instead estimate the base switching cutoff frequency $f_T = \frac{1}{2\pi R_{\mathrm{B}} \mathrm{C_{eff}}}$, where R_{B} and $\mathrm{C_{eff}}$ represent the base series resistance and net effective capacitance of the common emitter configuration from the equivalent device model in Fig. 1(d). This essentially corresponds to the base charging delay, which is the measure of how fast the base potential can be manipulated. In the three cases with the same barrier material (i.e., Ag/h-BN/Gr/n-SiC, Au/h-BN/Gr/n-SiC, and Cu/h-BN/Gr/n-SiC), the obtained values for f_T are identical since the considered capacitive components are unaffected by the choice of the emitter metal. For instance, Cu, Ag, and Au, with their high conductivities, can all be treated as an ideal metal in the estimate of the parallel-plate capacitance C_{BE} (i.e., no distinction). While this insensitivity may be due to the limitation of the model, it is, nevertheless, a good approximation as the amount of charge injection from the emitter via tunneling is inherently minute compared to the total charge components in the system. Assuming $R_\mathrm{B} = 100\ \Omega$ and the effective device cross section of $0.01\ \mu\mathrm{m}^2$, Figs. 4(c) and 4(d) provide f_T of the HETs with bilayer h-BN and bilayer MoS_2, respectively, for the selected bias points.

A realistic range of R_B can be gauged by taking advantage of related information in the literature. From the recent measurements of graphene FETs grown epitaxially on SiC, it has been shown that the sheet resistance of approximately $400\ \Omega/\mathrm{sq}$ is practically realizable along with the contact resistance less than $100\ \Omega\ \mu\mathrm{m}$ [30,31]. Needless to say, the actual base resistance also depends on the specific geometry and can be optimized further by adopting the design practices frequently used in conventional high-frequency applications such as the high aspect ratio of the device cross section and the double-sided base contact [32]. For instance, the 10:1 aspect ratio of the effective area (i.e., $0.01\ \mu\mathrm{m}^2\ \approx 0.32 \times 0.032\ \mu\mathrm{m}^2$) can reduce R_B to about $180\ \Omega$ particularly when the graphene base layer (of $0.32\ \mu\mathrm{m}$ wide) contacts with the base electrodes on both sides (leading to the effective width of $0.64\ \mu\mathrm{m}$). If a steeper ratio such as 20:1 is assumed, the combined impact of the contact and base spreading resistances can be further brought down to around $100\ \Omega$—a practicable value adopted in this study for an approximate estimate.

The results in Figs. 4(c) and 4(d) show that the cutoff frequency varies roughly from 5 to 8 THz and from 7 to 15 THz. The observed increase in f_T with V_CE can be explained by the decrease of C_CB and C_Q. Similarly, the upward trend as V_BE becomes smaller may be attributed mainly to the reduction in C_Q. From the comparison of Figs. 4(c) and 4(d), it becomes clear that the dominant factor in the switching frequency limit is C_BE that is inversely proportional to the BE barrier width. Consequently, there is much to be gained by adopting a 2D insulator of lower barrier height in place of h-BN or even MoS_2. With additional optimization, anticipating a realistic device with f_T well over 1 THz, thus, appears feasible even when making room for practical, fabrication-related uncertainties. One interesting point is that the influence of C_Q in the graphene base is, in fact, rather favorable in the high-frequency operations. The quantum capacitance, often ignored in conventional devices due to its large value, has a magnitude in graphene that is comparable to other capacitive components [22]. The resulting reduction in the net capacitance (via a small C_Q) works toward the improved switching frequency limit in the 2D HETs.

A number of design criteria become apparent from the analysis. For a sufficient current drive, ultrathin tunnel barriers with uniform thicknesses are simply a must. Similarly, the thin base region and the CB barrier control become crucial for a high gain in order to prevent the energy loss of injected "hot" electrons as well as the reflection at the CB junction. To achieve the high-frequency operation, on the other hand, additional conditions need to be addressed. While a short transit distance (from the emitter to the collector contact) determines the intrinsic frequency limit, the low series resistance becomes critical since, in reality, the RC time constant curbs the device operation well before the ultimate upper bound. In this regard, the proposed device possesses the desired traits of a THz device with a good current drive and high gain. Apart from the atomic scale base and tunnel insulator (i.e., short transit time and low energy loss), the Schottky barrier at the CB junction ensures an improved gain by lowering the electron reflection at the interface. At the same time, the highly conductive metallic emitter and base minimize the possible series resistance. In comparison, our analysis also highlights the less-than-desirable features of earlier attempts based on 3D materials such as III-V heterostructures or oxide barriers. For instance, the typical device dimensions of approximately 10 nm used in the conventional tunnel structures are significantly larger than what is required (e.g., $\lesssim 1$ nm) [29,33,34]. This shortcoming undoubtedly affects the device performance negatively from multiple aspects, which is further compounded by much higher rates of intrinsic carrier scattering or energy loss in 3D semiconductors (than the graphene-based 2D systems) [35] as well as those of extrinsic origin such as defects and surface roughness. The series resistance also tends to be greater in the 3D semiconductor nanostructures due to the doping limitations. Likewise, the recent attempts to take advantage of graphene as the 2D base [7–9] have met with limited success due to the continued application of 3D tunnel oxides. The presence of a high tunnel barrier at the CB side (instead of the Schottky junction) is particularly counterproductive for a high gain considering the strong scattering or reflection at the interface.

Finally, it is worth noting that the proposed 2D HET structure also offers other potential advantages over the corresponding conventional devices including structural simplicity circumventing numerous nanofabrication issues. For instance, the device dimension used in the f_T estimation is not very demanding and can be achieved by state-of-the-art photolithography technology. In addition, the epitaxial growth of graphene on a SiC substrate is already well established; the synthesis of h-BN and other 2D dielectric materials on a catalytic metal surface is an active research subject. Hence, mechanical transfer of 2D dielectric or metal/2D dielectric on a graphene-SiC system is clearly a realizable strategy particularly when combined with a proper choice of metallic emitter. Direct growth of 2D insulating layers on graphene can be an alternative option.

V. CONCLUSION

The proposed 2D HET holds significant promise for operation in the THz frequencies. Our theoretical analysis illustrates a pathway to additional enhancements (such as proper selection of materials and related design optimization) that can potentially enable reliable performance in the frequency range inaccessible thus far with conventional device structures. Considering the wide variety of relevant 2D crystals as well as compatible metals and semiconductors [36], fabrication of the structures described in this study or their derivatives appears very feasible. More broadly, tunnel devices based on atomically thin 2D materials are expected to offer unique opportunities in ultra-high-speed applications, clearly warranting further investigation both theoretically and experimentally.

ACKNOWLEDGMENTS

This work was supported in part by FAME (Function Accelerated nanoMaterial Engineering Center) [one of six centers of STARnet (Semiconductor Technology Advanced Research Network), a SRC (Semiconductor Research Corporation) program sponsored by MARCO (Microelectronics Advanced Research Corporation) and DARPA] and the U.S. Office of Naval Research.

[1] S. K. Sekatskii and V. S. Letokhov, Electron tunneling time measurement by field-emission microscopy, Phys. Rev. B **64**, 233311 (2001).

[2] I. Wilke, W. Herrmann, and F. K. Kneubühl, Integrated nanostrip dipole antennas for coherent 30 THz infrared radiation, Appl. Phys. B **58**, 87 (1994).

[3] B. Tiwari, J. A. Bean, G. Szakmány, G. H. Bernstein, P. Fay, and W. Porod, Controlled etching and regrowth of tunnel oxide for antenna-coupled metal-oxide-metal diodes, J. Vac. Sci. Technol. B **27**, 2153 (2009).

[4] L. Britnell, R. V. Gorbachev, R. Jalil, B. D. Belle, F. Schedin, A. Mishchenko, T. Georgiou, M. I. Katsnelson, L. Eaves, S. V. Morozov, N. M. R. Peres, J. Leist, A. K. Geim, K. S. Novoselov, and L. A. Ponomarenko, Field-effect tunneling transistor based on vertical graphene heterostructures, Science **335**, 947 (2012).

[5] B. D. Kong, C. Zeng, D. K. Gaskill, K. L. Wang, and K. W. Kim, Two dimensional crystal tunneling devices for THz operation, Appl. Phys. Lett. **101**, 263112 (2012).

[6] A. Catellani, M. Posternak, A. Baldereschi, and A. J. Freeman, Bulk and surface electronic structure of hexagonal boron nitride, Phys. Rev. B **36**, 6105 (1987).

[7] S. Vaziri, G. Lupina, C. Henkel, A. D. Smith, M. Östling, J. Dabrowski, G. Lippert, W. Mehr, and M. C. Lemme, A graphene-based hot electron transistor, Nano Lett. **13**, 1435 (2013).

[8] C. Zeng, E. B. Song, M. Wang, S. Lee, C. M. Torres, J. Tang, B. H. Weiller, and K. L. Wang, Vertical graphene-base hot-electron transistor, Nano Lett. **13**, 2370 (2013).

[9] W. J. Yu, Z. Li, H. Zhou, Y. Chen, Y. Wang, Y. Huang, and X. Duan, Vertically stacked multi-heterostructures of layered materials for logic transistors and complementary inverters, Nat. Mater. **12**, 246 (2013).

[10] J.-H. Chen, C. Jang, S. Xiao, M. Ishigami, and M. S. Fuhrer, Intrinsic and extrinsic performance limits of graphene devices on SiO_2, Nat. Nanotechnol. **3**, 206 (2008).

[11] A. Nagashima, H. Itoh, T. Ichinokawa, C. Oshima, and S. Otani, Change in the electronic states of graphite overlayers depending on thickness, Phys. Rev. B **50**, 4756 (1994).

[12] A. Nagashima, N. Tejima, Y. Gamou, T. Kawai, and C. Oshima, Electronic structure of monolayer hexagonal boron nitride physisorbed on metal surfaces, Phys. Rev. Lett. **75**, 3918 (1995).

[13] P. Vogt, P. De Padova, C. Quaresima, J. Avila, E. Frantzeskakis, M. C. Asensio, A. Resta, B. Ealet, and G. Le Lay, Silicene: Compelling experimental evidence for graphenelike two-dimensional silicon, Phys. Rev. Lett. **108**, 155501 (2012).

[14] J. L. Tedesco, B. L. VanMil, R. L. Myers-Ward, J. M. McCrate, S. A. Kitt, P. M. Campbell, G. G. Jernigan, J. C. Culbertson, C. R. Eddy, Jr., and D. K. Gaskill, Hall effect mobility of epitaxial graphene grown on silicon carbide, Appl. Phys. Lett. **95**, 122102 (2009).

[15] J. P. Spratt, R. F. Schwarz, and W. M. Kane, Hot electrons in metal films: Injection and collection, Phys. Rev. Lett. **6**, 341 (1961).

[16] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, The SIESTA method for ab initio order-N materials simulation, J. Phys. Condens. Matter **14**, 2745 (2002).

[17] R. Mao, B. D. Kong, C. Gong, S. Xu, T. Jayasekera, K. Cho, and K. W. Kim, First-principles calculation of thermal transport in metal/graphene systems, Phys. Rev. B **87**, 165410 (2013).

[18] T. Jayasekera, B. D. Kong, K. W. Kim, and M. Buongiorno Nardelli, Band engineering and magnetic doping of epitaxial graphene on SiC(0001), Phys. Rev. Lett. **104**, 146801 (2010).

[19] O. Sinai and L. Kronik, Simulated doping of Si from first principles using pseudoatoms, Phys. Rev. B **87**, 235305 (2013).

[20] M. J. Oliveira and F. Nogueira, Generating relativistic pseudo-potentials with explicit incorporation of semi-core states using APE, the Atomic Pseudo-potentials Engine, Comput. Phys. Commun. **178**, 524 (2008).

[21] S. Luryi, Quantum capacitance devices, Appl. Phys. Lett. **52**, 501 (1988).

[22] H. Xu, Z. Zhang, Z. Wang, S. Wang, X. Liang, and L.-M. Peng, Quantum capacitance limited vertical scaling of graphene field-effect transistor, ACS Nano **5**, 2340 (2011).

[23] See, for example, D. A. Neamen, *Semiconductor Physics and Devices*, 3rd ed. (McGraw Hill, New York, 2003), p. 420.

[24] M. Büttiker, Role of quantum coherence in series resistors, Phys. Rev. B **33**, 3020 (1986).

[25] S. Datta, *Electronic Transport in Mesoscopic Systems* (Cambridge University Press, Cambridge, England, 1995).

[26] J. A. Robinson, M. LaBella, M. Zhu, M. Hollander, R. Kasarda, Z. Hughes, K. Trumbull, R. Cavalero, and D. Snyder, Contacting graphene, Appl. Phys. Lett. **98**, 053103 (2011).

[27] T. Druga, M. Wenderoth, F. Lüpke, and R. G. Ulbrich, Graphene-metal contact resistivity on semi-insulating 6H-SiC(0001) measured with Kelvin probe force microscopy, Appl. Phys. Lett. **103**, 051601 (2013).

[28] See, for example, S. M. Sze, *Physics of Semiconductor Devices*, 3rd ed. (Wiley, Hoboken, NJ, 1981), p. 263.

[29] M. Heiblum and M. V. Fischetti, Ballistic hot-electron transistors, IBM J. Res. Dev. **34**, 530 (1990).

[30] J. S. Moon, H.-C. Seo, F. Stratan, M. Antcliffe, A. Schmitz, R. S. Ross, A. A. Kiselev, V. D. Wheeler, L. O. Nyakiti, D. K. Gaskill, K. M. Lee, and P. M. Asbeck, Lateral graphene heterostructure field-effect transistor, IEEE Electron Device Lett. **34**, 1190 (2013).

[31] Z. Guo, R. Dong, P. S. Chakraborty, N. Lourenco, J. Palmer, Y. Hu, M. Ruan, J. Hankinson, J. Kunc, J. D. Cressler, C. Berger, and W. A. de Heer, Record maximum oscillation frequency in C-face epitaxial graphene transistors, Nano Lett. **13**, 942 (2013).

[32] M. J. W. Rodwell, M. Le, and B. Brar, InP bipolar ICs: Scaling roadmaps, frequency limits, manufacturable technologies, Proc. IEEE **96**, 271 (2008).

[33] J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, A. L. Hutchinson, and A. Y. Cho, Quantum cascade laser, Science **264**, 553 (1994).

[34] M. Feng, N. Holonyak, Jr., and R. Chan, Quantum-well-base heterojunction bipolar light-emitting transistor, Appl. Phys. Lett. **84**, 1952 (2004).

[35] See, for instance, K. M. Borysenko, J. T. Mullen, E. A. Barry, S. Paul, Y. G. Semenov, J. M. Zavada, M. Buongiorno Nardelli, and K. W. Kim, First principles analysis of electron-phonon interactions in graphene, Phys. Rev. B **81**, 121412(R) (2010).

[36] C. Gong, H. Zhang, W. Wang, L. Colombo, R. M. Wallace, and K. Cho, Band alignment of two-dimensional transition metal dichalcogenides: Application in tunnel field effect transistors, Appl. Phys. Lett. **103**, 053513 (2013).