Sub-10 nm Gate Length Graphene Transistors: Operating at Terahertz Frequencies with Current Saturation

Jiaxin Zheng1,2,3, Lu Wang3, Ruge Quhe1,2, Qihang Liu1, Hong Li1, Dapeng Yu1, Wai-Ning Mei2, Junjie Shi1, Zhengxiang Gao1 & Jing Lu1

1State Key Laboratory of Mesoscopic Physics and Department of Physics, Peking University, Beijing 100871, P. R. China, 2Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, P. R. China, 3Department of Physics, University of Nebraska at Omaha, Omaha, Nebraska 68182-0266.

Radio-frequency application of graphene transistors is attracting much recent attention due to the high carrier mobility of graphene. The measured intrinsic cut-off frequency ($f_T$) of graphene transistor generally increases with the reduced gate length ($L_{\text{gate}}$) till $L_{\text{gate}} = 40$ nm, and the maximum measured $f_T$ has reached 300 GHz. Using ab initio quantum transport simulation, we reveal for the first time that $f_T$ of a graphene transistor still increases with the reduced $L_{\text{gate}}$ when $L_{\text{gate}}$ scales down to a few nm and reaches astonishing a few tens of THz. We observe a clear drain current saturation when a band gap is opened in graphene, with the maximum intrinsic voltage gain increased by a factor of 20. Our simulation strongly suggests it is possible to design a graphene transistor with an extraordinary high $f_T$ and drain current saturation by continuously shortening $L_{\text{gate}}$ and opening a band gap.

Graphene is of particular interest for ultrahigh speed electronics due to its high carrier mobility and saturation velocity. Because of its zero band gap, graphene field effect transistors (FETs) have a low on/off current ratio, which limits its application in logic devices. There is still a lack of reliable techniques to open a sizable gap without degrading the electronic properties of graphene. However, a large on/off ratio is not necessary for radio frequency (r. f.) electronic applications, which are the core elements in wireless communication devices. Developing the high-performance r. f. graphene transistors is attracting enormous recent attention. In order to facilitate high-performance r. f. applications, FETs should respond quickly to the gate voltage ($V_g$), which requires short gates. Extensive gate length ($L_{\text{gate}}$) scaling work of graphene FETs has been experimentally pursued, with the gate length scaling down to 40 nm. The intrinsic cut-off frequency ($f_T$) represents how fast the channel current is modulated by the gate and is one of the most important figure-of-merit for evaluating the performance of r. f. devices. All these experimental investigations reveal that $f_T$ of graphene transistors generally increases with the decreasing $L_{\text{gate}}$. The measured maximum $f_T$ is 300 GHz for a graphene FET with $L_{\text{gate}} = 144$ nm based on exfoliated graphene. The maximum $f_T$ estimated from the static measurement is 1.4 THz from a graphene FET with $L_{\text{gate}} = 45$ nm fabricated by a self-aligned approach. Therefore, as far as $f_T$ is concerned, graphene FETs have significantly outperformed conventional silicon metal-oxide-semiconductor field effect transistors (MOSFETs) with a highest measured $f_T$ of 485 GHz at $L_{\text{gate}} = 29$ nm and III-V high-electron-mobility transistors (HEMTs) with a highest measured $f_T$ of 660 GHz at $L_{\text{gate}} = 20$ nm for GaAs.

However, the $f_T$ value does not always increase with the reduced $L_{\text{gate}}$ in a transistor when $L_{\text{gate}}$ approaches the size limit. For example, $f_T$ of GaAs metamorphic HEMT peaks at $L_{\text{gate}} = 20$ nm and then it slightly decreases with the reduced $L_{\text{gate}}$. One fundamental issue arises naturally: is there a saturation of $f_T$ with the reduced $L_{\text{gate}}$ in graphene FETs? If such a saturation is absent, a higher intrinsic cut-off frequency, even up to tens of terahertz (THz), can be obtained via continuously shortening $L_{\text{gate}}$ to a few nm in graphene FETs. Besides $f_T$, the maximum oscillation frequency ($f_{\text{max}}$) and intrinsic voltage gain ($A_v$) are the other two important figure-of-merits evaluating the r. f. performance of devices. To obtain a high $f_{\text{max}}$ and $A_v$, a drain current saturation is required. Although a drain current saturation has been reported in some monolayer graphene (MLG) FET devices due to phonon scatter limited velocity saturation and in a dual-gated bilayer graphene (BLG) FET device due to an electrical-field-induced band gap opening, all the existing short-channel graphene FETs with $L_{\text{gate}}$ below 300 nm suffer...
from a lack of drain current saturation\textsuperscript{12,14}, which lead to a small output conductance and a poor $f_{\text{max}}$ and $A_v$. Another fundamental issue of graphene FETs is: is there an effective method to induce a drain current saturation in sub-10 nm scale?

In this article, we investigate theoretically for the first time the performance limit of top-gated graphene FETs with the gate length scaling down from 9.86 to 0.91 nm using the density functional theory (DFT) coupled with the nonequilibrium Green’s function (NEGF) method. We demonstrate that switching effects remain in such short gate graphene transistors. Remarkably, $f_T$ still increases with the decreasing $L_{\text{gate}}$, with values of 3.4–21 THz. The absence of a saturation in $f_T$ with the reduced $L_{\text{gate}}$ clearly shows that $f_T$ of graphene transistors can be continuously improved by shortening $L_{\text{gate}}$. In order to create a drain current saturation, we design two schemes: one is a dual-gated FET made of a MLG sandwiched between two hexagonal boron nitride (h-BN) layers and the other is a dual-gated FET made of a pure BLG. A significant current saturation has been observed in both FETs under a proper vertical electrical field, and the maximum voltage gain has been increased up to 2.3 and 1.5, respectively, a factor of 20 and 13 higher than that of pure MLG FETs. Experimentally, sub-10 nm gate length carbon nanotube transistors have been recently manufactured via top-down approach\textsuperscript{20}. It has been pointed out by Duan \textit{et al}. that a sub-10 nm gate length graphene FET can be realized by using the self-aligned approach with a sub-10 nm nanowire as the top gate\textsuperscript{9}. Finally, we envision an alternative method to fabricate a sub-10 nm graphene transistor, namely, using a sub-10 nm diameter boron nitride (BN) or carbon nanotube as the top gate. We expect our work can inspire experimentalists to further push graphene transistors to performance limit by continuously shortening $L_{\text{gate}}$ and opening a band gap.

**Results**

In an actual graphene FET, metal is always used to contact graphene as electrodes. In our schematic model of a MLG FET presented at Fig. 1, we design the graphene channel contacted underneath two aluminum (Al) electrodes (source and drain). The distance between Al contact

---

**Figure 1** (a-c) Schematic model of a top-gated pure MLG FET (a), dual-gated BN/MLG/BN sandwich FET (b), and dual-gated BLG FET (c). The two-dimensional MLG and BLG channels are contacted underneath two aluminum (Al) electrodes. Gray ball: C; light pink ball: Al; blue ball: B; heavy pink ball: B.
and graphene layer is 0.34 nm, according to the previous work\(^{11,22}\). We further set the thickness of the dielectric region to be \(d_1 = 1.4 \text{ nm}\), and the dielectric constant to be \(\varepsilon_r = 3.9\), which models after SiO\(_2\).

As the gate of FETs scales down to very short lengths, like sub-10 nm region, the transport mechanism turns from channel-dominated diffusive regime to the contact-dominated ballistic regime, so the contact effect between the metal electrode and the channel becomes very important for the device transport properties\(^{30,31}\). We first study the contact effect between the aluminum (Al) electrode and graphene. In Fig. 2(a) and (b), we show the projected density of states (PDOS) on the carbon (C) atoms of graphene in the left/right lead \((g_{L/R}(E))\) and in the channel \((g_{ch}(E))\), respectively, for a graphene FET with \(L_{\text{gate}} = 5.6 \text{ nm}\) under zero drain-source voltage and zero gate voltage. The inset in Fig. 2(a) is the density of states (DOS) of the pure MLG graphene, with the Dirac point exactly at the Fermi level \((E_F)\). We find that the Dirac points of graphene in the left/right lead \((D_{L/R})\) and in the channel \((D_{ch})\) are shifted to \(D_{L/R} = E_F - 0.9 \text{ eV}\) and \(D_{ch} = E_F - 0.25 \text{ eV}\), respectively, indicative of heavy \(n\)-doping of graphene in the contact part and light \(n\)-doping in the channel part by Al contact. This contact effect is also reflected on the transmission spectrum of this device. As shown in Fig. 2(c), there are two transmission minima in the transmission spectrum of this graphene FET but only one minimum in the transmission spectrum of the corresponding MLG FET without being contacted underneath Al electrodes (See the inset in Fig. 2(c)). The transmission coefficient of the device \(T(E)\) behaves like three resistors composed of the channel and the two electrodes in series\(^{24,26}\):

\[
T(E) \propto \frac{g_{ch}(E)g_{ch}(E)g_{ch}(E)}{g_{al}(E)g_{al}(E)g_{al}(E)}
\]

The two minima in the transmission spectrum originate from the \(D_{L/R}\) in \(g_{L/R}(E)\) and \(D_{ch}\) in \(g_{ch}(E)\), respectively. Consequently, there is a lack of symmetry in the transmission spectrum with respect to both \(D_{L/R}\) and \(D_{ch}\), causing the asymmetry in electron and hole conductance. The existence of two minima in the transmission spectrum and electron-hole conductance asymmetry agrees well with both previous theoretical and experimental works\(^{20,23-29}\).

Fig. 3(a) shows the transmission spectra of this device under \(V_g = 6.0, -3.0,\) and \(-5.0 \text{ V}\) with a fixed bias voltage of \(V_{ds} = 0.5 \text{ V}\). There are three transmission minima \(D_1, D_2\), and \(D_3\) in each transmission spectrum because \(D_1\) and \(D_2\) are separated by the bias voltage. \(D_1\) and \(D_2\) are unchanged by \(V_g\). However, \(D_3\) is shifted by \(V_g\) first it is at \(E_F\) under \(V_g = -5 \text{ V}\), then shifted to \(E_F - 0.25 \text{ eV}\) under \(V_g = -3 \text{ V}\), and finally shifted to \(E_F - 0.9 \text{ eV}\) under \(V_g = 6 \text{ V}\). This shift of \(D_{ch}\) with \(V_g\) in the transmission spectrum stems from the shift of \(D_{ch}\) with \(V_g\) in \(g_{ch}(E)\), as shown in Fig. 3(b). As \(D_{ch}\) is at \(E_F(V_g = -5 \text{ V})\), the \(T(E)\) values within the bias window are small, and the current is minimized (off-state). As \(D_{ch}\) is moved gradually out from the bias window, the \(T(E)\) values within the bias window increase, the currents increase gradually, and the FET is finally switched from the off-state \((V_g = -5 \text{ V})\) to the on-state \((V_g = 6 \text{ V})\). The on/off current ratio of this device is 13.3.

We calculate the transmission eigenvalues of all the \(k\)-points in the 1D Brillouin zone at \(E_F\). For a given energy \(E\), the transmission coefficient \(T(E)\) is given by

\[
T(E) = \frac{\sum_{E_n} T_{mn}(E,k) \Delta k}{2\pi} = \frac{1}{2\pi} \sum_{n} \lambda_n(E,k)
\]

where \(\lambda_n\) are the eigenvalues of the transmission matrix \(T\). In Fig. S1 of Supplementary Information, we show the transmission coefficients \(T(E_F,k)\) as a function of \(k\) of the off-state \((V_g = -5.0 \text{ V},\) black curve\) and the on-state \((V_g = 6.0 \text{ V},\) red curve\) in the 1D Brillouin zone. The width and height of the \(T(E_F,k)\) spectrum for the on-state are much larger than those for the off-state, and therefore the total transmission factor \(T(E_F)\) of the on-state is larger than that of the off-state. To visualize the difference between the off- and on-state of this graphene FET, we choose the \((2\pi/5a,0)\) 0 point of the \(k\)-space at \(E_F\) and show its transmission eigenchannel in the off-state (left panel) and on-state (right panel) in Fig. 3(c). The largest transmission eigenvalue at this point of the off-state is \(\lambda_1 = 1.1 \times 10^{-4}\), and the rest are nearly zero; correspondingly, the incoming wave function is almost completely scattered and unable to reach the other lead. In contrast, the largest transmission eigenvalue at this point of the on-state is \(\lambda_1 = 0.97\), and the rest are also nearly zero; consequently, the incoming wave function is scattered little and most of the incoming wave reaches the other lead.

We show the transfer characteristics of the devices with \(L_{\text{gate}}\) scaling down from 9.86 to 0.91 nm in Fig. 4(a), at a fixed drain-source voltage of \(V_{ds} = 0.5 \text{ V}\). As expected, all of them show an \(n\)-type feature. The on-currents, which are defined as the current at \(V_g = 6.0 \text{ V}\), remain almost constant for all the \(L_{\text{gate}}\). By contrast, the leakage currents in the off-state increase with the decreasing \(L_{\text{gate}}\). As shown in Fig. 4(b), the on/off current ratio drops significantly from 15.8 at \(L_{\text{gate}} = 9.86 \text{ nm}\) to 3.35 at \(L_{\text{gate}} = 0.91 \text{ nm}\), and this scaling trend agrees with the experimental results for short gate graphene FETs \((L_{\text{gate}} = 5.6 \text{ nm} - 50 \text{ nm})\)\(^{11,22}\). Such a scaling behavior is attributed to the increasing off-state leakage current with the decreased \(L_{\text{gate}}\) as the gate scales down to very short lengths, the off-state average PDOS on each C atom of the channel graphene at \(D_{ch}\) increases \((\text{Fig. S2, Supplementary Information})\), and the channel graphene behaves more like metal and away from semi-metal. Our calculated on/off ratio of MLG FETs is larger than the previous.

Figure 2 | Contact effect between the graphene channel and metal electrodes for a 5.6-nm gate length MLG FET under zero drain-source voltage and zero gate voltage. (a) Projected density of states on carbon atoms of graphene in the left/right lead (a) and in the channel (b). (c) Transmission spectrum of this device. \(D_{L/R}\) and \(D_{ch}\) denote the Dirac point of the graphene in the left/right electrode and the channel graphene, respectively. Inset in (a): density of states for the pure MLG; inset in (c): transmission spectrum of MLG without being contacted underneath metal electrodes under zero drain-source voltage and zero gate voltage.
The method to fabricate an ultrathin dielectric (sub-10 nm) will improve the gate current ratio and transconductance significantly because the gate oxide thickness is remarkable when L<sub>ox</sub> can be seen in the transfer characteristics shown in Fig. 4(a). The shift of the on/off current ratios are about 3.5–6 for p-type graphene FETs with experimental data (For example, in the work of Wu et al.

Figure 3 | Switching effect for a 5.6-nm gate length MLG FET with a drain-source voltage of V<sub>ds</sub> = 0.5 V. (a) Transmission spectra of the off-state (V<sub>g</sub> = −5.0 V), on-state (V<sub>g</sub> = 6.0 V), and an intermediate state (V<sub>g</sub> = −3.0 V). The dashed vertical line indicates the bias window. D<sub>ch1</sub>, D<sub>ch2</sub>, and D<sub>ch3</sub> denote the Dirac point of the graphene in the left/right electrode; D<sub>g1</sub>, D<sub>g2</sub>, and D<sub>g3</sub> denote the Dirac point of the channel graphene under V<sub>g</sub> = −5.0, −3.0, and 6.0 V, respectively. (b) Projected density of states on carbon atoms in the channel under different gate voltages. The isovalue is 0.6 a.u.

To get f<sub>T</sub>, we need to calculate another key parameter: the intrinsic gate capacitance C<sub>g</sub> as shown in Fig. 4(d), C<sub>g</sub> decreases approximately linearly with L<sub>gate</sub> because C<sub>g</sub> is related to L<sub>gate</sub> through the following equation:

\[ C_g = \varepsilon_0 \varepsilon_r W_{gate} L_{gate} \rho_{ox} \]

where \( \varepsilon_0 \) is the dielectric constant of vacuum, \( \varepsilon_r \) is the relative dielectric constant of the gate dielectric, W<sub>gate</sub> is the gate width, and \( \rho_{ox} \) is the gate dielectric thickness. Actually, besides C<sub>g</sub> transistors also include another kind of capacitance named parasitic capacitance (C<sub>par</sub>), which exists between the gate and source (drain) electrode and is an important parameter to determine the extrinsic cut-off frequency (f<sub>T, ex</sub>):

\[ f_{T, ex} = \frac{g_m}{C_g + C_{par}} \]

Ideally, one can make f<sub>T, ex</sub> close to f<sub>T</sub> through reducing the parasitic capacitance, such as by redesigning the device layout and reducing the total parasitic series resistance between the device’s source (drain) electrode and the gate to increase the current density.

Fig. 4(e) shows L<sub>gate</sub> dependence of f<sub>T</sub> based on the calculated g<sub>m</sub> and C<sub>g</sub>. f<sub>T</sub> increases monotonically from 3.4 to 21 THz as L<sub>gate</sub> scales down from 9.86 to 0.91 nm, because C<sub>g</sub> reduces faster than g<sub>m</sub> with the reduced L<sub>gate</sub>. We notice that f<sub>T</sub> exhibits a 1/L<sub>gate</sub> dependence as L<sub>gate</sub> scales from 9.86 to 2.19 nm, which agrees well with the scaling trend reported experimentally by Wu et al. with L<sub>gate</sub> from 550 to 40 nm.<sup>4</sup> The 1/L<sub>gate</sub> scaling trend suggests that the transport of our devices is in the contact-dominated ballistic regime and the electric field along the channel is dominated by the value of the contact resistance at the device’s source (drain) as shown in Fig. 4(f). This 1/L<sub>gate</sub> dependence is also usually observed for long-channel conventional Si and III-V FETs.<sup>4</sup> As L<sub>gate</sub> scales below 2.19 nm, the scaling trend deviates from the 1/L<sub>gate</sub> dependence, due to the dramatic drop of g<sub>m</sub>. The product f<sub>T</sub>L<sub>gate</sub> as L<sub>gate</sub> scales down is expected to be linearly proportional to g<sub>m</sub> for devices with the same parameters and dimensions. In Fig. 4(f), we show the plot of f<sub>T</sub>L<sub>gate</sub> against g<sub>m</sub> and it exhibits the expected linear dependence.

We summarize the gate length scaling works of the intrinsic cut-off frequency for MLG FETs in Fig. 5. An increase trend of f<sub>T</sub> with the reduced L<sub>gate</sub> is available in all works. Almost all the data can be
roughly fitted by the uniform $f_T = a/L_{	ext{gate}}$ relation ($a = 38600$ GHz nm) except that the data based on CVD grown graphene has a different coefficient ($a = 8000$ GHz nm). It is noteworthy that in the 40–100 nm region, the theoretical $f_T$ data obtained by Guo et al. using a self-consistent ballistic quantum transport simulation approach with the NEGF formalism agrees well with those reported by Duan et al. with a self-aligned approach. This agreement shows the reliability of the quantum transport simulation in predicting $f_T$ of graphene FETs, and the condition will be discussed in the next section.

We study the output characteristics of a graphene FET with $L_{	ext{gate}} = 6.4$ nm and show them in Fig. 6(a). The source-drain ballistic current increases linearly with the applied bias voltage in the checked bias region. In most experimental graphene FET devices, the absence of drain current saturation degrades the maximum oscillation frequency ($f_{\text{max}}$) and the intrinsic voltage gain ($A_v$), which represent how fast power transmission is modulated by the gate and the amplification factor of an input signal, respectively. $f_{\text{max}}$ is defined as the frequency at which the power gain becomes unity, and a typical approximation for it is

$$f_{\text{max}} = \frac{f_T}{2 \sqrt{g_d (R_g + R_{\text{ds}}) + 2 \pi f_T R_g C_g}}$$

$g_d$ is the output conductance, $R_g$ the gate resistance and $R_{\text{ds}}$ the drain-source resistance. $A_v$ is defined as $A_v = g_{\text{m}}/g_d$.

The reason of the absence of current saturation in graphene FET devices is attributed to zero band gap of graphene, high electrical resistance between the device’s source and drain electrode and its gate, and the short channel effect. One way to get current saturation is to open a band gap in graphene. Both theoretical and experimental works show that a vertical external electric field can induce a tunable band gap up to 0.25 eV for BLG without degrading the electronic properties of graphene, and a current saturation in BLG FETs ($L_{\text{gate}} = 4 \sim 9$ μm) is indeed observed and reproduced in a simulation ($L_{\text{gate}} = 40$ nm) within the tight-binding Hamiltonian.
and NEGF formalism as a result of the band gap opening. Our previous theoretical work predicts that a vertical external electric field can induce a tunable band gap up to 0.34 eV for a MLG properly sandwiched between two h-BN single layers without degrading the electronic properties of graphene. Such a BN/MLG/BN sandwich structure has been prepared experimentally recently.

The schematic model of a dual-gated graphene BN/MLG/BN sandwich FET with $L_{\text{gate}} = 6.4$ nm and a dual gated BLG FET with $L_{\text{gate}} = 9.6$ nm are presented in Fig. 1 (b) and (c), and the two individual gates allow us to create a larger band gap and tune channel's conductance individually. The device performance of MLG and BLG FET with the same $L_{\text{gate}}$ is similar if the band gap of BLG is not opened. The vertical electrical field applied to sandwich structure is obtained as. The corresponding total gate voltage is $V_g = V_t + V_b$, reflecting the total doping level. In Fig. 6(b), we compare the simulated output characteristics of a pure MLG FET and a BN/MLG/BN sandwich FET (under a vertical external electric field of $-1$ V/A) with the same $L_{\text{gate}}$. A significant current saturation appears for the sandwich FET at $V_g = 2.0, 2.4$ V (a positive gate voltage has no such effect), and its output conductance $g_{ds}$ is lowered at most by a factor of 80 compared with the pure MLG FET. The calculated $C_g$ of this sandwich FET is nearly intact and $g_{ms}$ is a quarter of those of the pure MLG FET (its transfer characteristics are shown in Fig. S3(a)), and the $f_T$ of this sandwich FET is 1.5 THz (it is degraded by a factor of 4 compared with a value of 6.35 THz for the pure MLG FET with the same $L_{\text{gate}}$). The maximum $A_v$ of this

Figure 5 | Gate length scaling of the intrinsic cut-off frequency for different graphene FETs. The experimental values are based on the epitaxially grown graphene, exfoliated graphene, self-aligned nanowire gate, and CVD grown graphene, respectively. Data are roughly fitted by the curve showing a $1/L_{\text{gate}}$ dependence.

Figure 6 | (a) $I_{ds}-V_{ds}$ output characteristics for the top-gated MLG FETs at variable gate voltages for $L_{\text{gate}} = 6.4$ nm. (b) $I_{ds}-V_{ds}$ output characteristics for a pure MLG FET and a BN/MLG/BN sandwich FET with the same $L_{\text{gate}} = 6.4$ nm under a vertical electrical field of $E = -1$ V/A. (c) $I_{ds}-V_{ds}$ output characteristics for a pure MLG FET and a BLG FET with the same $L_{\text{gate}} = 9.6$ nm under a vertical electrical field of $E = 3$ V/mm. (d) Transmission spectra of the BN/MLG/BN sandwich FET with $L_{\text{gate}} = 6.4$ nm at $V_{ds} = 0.2$ V (red curve) and 0.25 V (blue curve) under a vertical external electric field of $-1$ V/A and a gate voltage of $-1.6$ V. The dashed vertical lines indicate the bias window.
device is therefore increased to 2.3, a factor of 20 higher than that of the pure MLG FET (the experimental enhancement factor and simulation enhancement factor of $A_v$ for BLG FETs ($L_{gate} = 4-9 \mu m$ experimentally and 40 nm theoretically) by a vertical external electric field are 6 and 10, respectively$^{19}$. The $f_{\text{max}}$ is increased by a factor of 2 if we assume that $W_{gate}$ is 1 $\mu m$, $R_{d}$, 980 $\Omega$, and $R_{g}$, 47 $\Omega$$^{26}$. We also observe a significant current saturation in the output characteristics of the BLG FET at $V_{ds} = -4.0 \, V$ under a vertical external electric field of 3 $V/\mu m$ (Fig. 6(c)), and its output conductance $g_{ds}$ is lowered at most by a factor of 40 compared with the pure MLG FET with the same $L_{gate}$. The calculated $C_{g}$ of this BLG FET is nearly intact with the pure MLG FET with the same $L_{gate}$, and $g_{mn}$ is one third of those of the pure MLG FET (its transfer characteristics are shown in Fig. S3(b)), the $f_{\tau}$ is 1.1 THz (it is degraded by a factor of 3 compared with a value of 3.4 THz for the pure MLG with the same $L_{gate}$), the maximum $A_v$ is increased to 1.5, and the $f_{\text{max}}$ is increased by a factor of 1.6. Therefore it is possible to design a sub-10 nm graphene FET that can operate at extraordinary high $f_{\tau}$ and with greatly improved $A_v$, by introducing a band gap in graphene. The gain factor in $g_{ds}$ and $A_v$ is one order of magnitude larger than the degradation one in $f_{\tau}$. Besides the sandwich and bilayer scheme, adsorption of Li on MLG and formation of LiC$_6$ structure$^{14}$ is able to open a larger band gap of about 0.4 eV, and thus LiC$_6$ monolayer is also a proper candidate of the channel of MLG FET with a current saturation. Since MLG grown on insulating SiC substrate already has a band gap of 0.26 eV$^{44}$, we can alternatively use a single-gated MLG grown on SiC as the channel to create a drain current saturation although a simulation of such a device is a challenge due to the difficulty to reproduce a band gap in graphene.

The mechanism of the current saturation in the BN/MLG/BN sandwich FET and the BLG FET can be both attributed to the band gap induced by the vertical external electric field. Taking the BN/MLG/BN sandwich FET for example, the band gap improves the r. f. performance of short-channel graphene transistors$^{19}$ and very small access resistance due to the very small gaps (0.32 nm) between the gated graphene channel and source/drain contacts. An ultrathin dielectric implies a better gate modulation on the channel current$^{31}$. If we use sub-10 nm diameter semiconducting nanowires, BN nanotubes, or carbon nanotubes as the top-gate to fabricate sub-10 nm graphene transistors, the dielectric region is the depletion layer, whose thickness can be controlled by doping concentration and is obviously smaller than the nanowire/nanotube diameters. Therefore, the ultra-thin dielectric (a few nm) can be realized in such fabrication schemes. If we use sub-10 nm diameter metallic nanowires or carbon nanotubes as the local gate and a few h-BN layers (sub-10 nm thick) as dielectrics, the ultra-thin dielectric can also be realized.

A substantial access resistance due to the significant gaps between source/drain and gate electrodes (a large portion of the graphene channel in the gate area is not gated) limits the achievable transconductance and has adverse impact on short channel devices$^{9,47,48}$. Experimentally, the access resistance has been significantly reduced with a self-alignment approach$^{9,10,47}$, through which the edges of the source, drain, and gate electrodes are automatically and precisely positioned so that no overlap and significant gaps exist between them. This process improves the transconductance and the drain current density significantly$^{9,10,47}$, and a slight current saturation was even observed$^{35}$. As a result, this self-alignment approach finally improves the r. f. performance of short-channel graphene transistors. For example, the $f_{\tau}$ values estimated from the static measurement of graphene FETs by Duan et al. in the 40–100 nm gate region have

**Figure 7** Schematic model of a MLG and (5, 0) boron nitride nanotube junction (a) and its $I$-$V_{\text{bias}}$ output characteristic curve (b).
reached 700 ~ 1400 GHz as a result of ultrathin dielectric and ultrasmall access resistance. Such a $f_T$-$L_{\text{gate}}$ relation agrees well with the simulation reported by Guo et al. within a model also with an ultrathin dielectric (16 nm) and an ultrasmall access resistance due to the ultrasmall distance between the source/drain and gate electrodes.$^{6,11}$

In our model, we assume that electron transport is in the ballistic region. Another issue we need to discuss here is the probable roles of electron-phonon scattering. For carbon nanotubes, some works on this issue reveal that the measured or calculated high-field electron mean free path (MFP) is about 10 nm or longer due to optical phonon scattering and the low-field MFP 1600 nm due to acoustic phonon scattering$^{46-48}$. The experimental and theoretical work by Dai et al. also points out that transport through very short (~10 nm) nanotubes is free of significant acoustic and optical phonon scattering and thus ballistic and quasi-ballistic at the low and high field limit, respectively.$^{45}$ For graphene, the measured electron MFP is from tens to hundreds of nanometers on rugged SiO$_2$ substrate, depending on the carrier concentration$^{2}$ and several micrometers in suspended ultraclean membranes$^{25}$ or in graphene encapsulated between inert and ultraflat h-BN layers at room temperature$^{26,27}$. The channel of our graphene transistor is sub-10 nm, which is much smaller than electron MFP of graphene on smooth h-BN substrate or even on rugged SiO$_2$ substrate in high carrier concentration, so the transport mechanism should be safely in the ballistic regime if a smooth substrate is used, and the Landau- Büttiker transport formalism is applicable.

In summary, we have investigated the switching effect and r. f. performance of the sub-10 nm gate length graphene FETs with Al electrodes by employing ab initio quantum transport simulation for the first time. We find that switching effects remain in these ultrashort gate graphene transistors. The intrinsic cutoff frequency increases monotonically from 3.4 to 21 THz, which are a few to tens of times larger than the experimental maximum values of the competitive r. f. FETs$^{5}$, when the gate length scales down from 9.86 to 0.91 nm. A significant current saturation can be created in sub-10 nm graphene FETs by introducing a band gap to graphene. We expect that our theoretical work can stimulate experimental fabrication of sub-10 nm gate length graphene FETs operating at an intrinsic cutoff frequency exceeding the available best experimental values and with current saturation.

**Methods**

The current of graphene FETs under a finite drain-source voltage ($V_{ds}$) and gate voltage ($V_g$) is computed using the Landauer- Büttiker formalism

$$I(V_{ds},V_g) = \frac{2e}{h} \int_{-\infty}^{5} \left[ T(E,V_{ds},V_g) J_f(E) - J_{ff}(E) \right] dE$$

where $J_{ff}$ is the Fermi-Dirac distribution function for the left (L)/right (R) electrode, $J_f$ is the electrochemical potential of the left (L)/right (R) electrode, and $T(E,V_{ds},V_g)$ is the transmission coefficient. $T(E)$ is calculated by using the Fisher and Lee relationship$^{5}$.

Then, the gate effect is calculated by solving the Poisson and Kohn-Sham equations self-consistently with the fixed boundary condition$^{6}$. Namely, we first calculate the Hartree potential $V^h$ from the Poisson equation: $\nabla^2 V^h = -4\pi n(r)$, where the initial electron density $n(r)$ is computed from the DFT and NEGF methods under no gate voltage. Afterward the new Hartree potential is used to calculate a new electron density by solving the Kohn-Sham equation. These procedures are iterated until the desired numerical accuracy is reached, which are presented out by using the well developed ATK 1.2 package$^{28}$. Single-zeta (SZ) basis set is used, the real-space mesh cutoff is 150 Ry, and the temperature is set at 300 K. The local-density-approximation (LDA) is employed for the exchange–correlation functional. The electronic structures of electrodes and central region are calculated with a Monkhorst-Pack$^{29}$ $50 \times 1 \times 100$ and $50 \times 1 \times 1$ k-point grid, respectively.

The intrinsic cut-off frequency $f_r$ is defined as the frequency at which the current gain becomes unity. We calculate $f_r$ using the flowing equation$^{30,31}$:

$$f_r = \frac{2\pi}{\tau_r}$$

where $\tau_r$ is the intrinsic gate capacitance $C_g$ and the transconductance $g_m$, which are computed from

$$g_m = \frac{\partial I}{\partial V_g}$$



1. Novoselov, K. S. et al. Two-dimensional gas of massless Dirac fermions in graphene. Nature 438, 197–200 (2005).
2. Zhang, Y. B., Tan, Y. W., Stormer, H. L. & Kim, P. Experimental observation of the quantum Hall effect and Berry’s phase in graphene. Nature 438, 201–204 (2005).
3. Berger, C. et al. Electronic confinement and coherence in patterned epitaxial graphene. Science 312, 1191–1196 (2006).
4. Geim, A. K. & Novoselov, K. S. The rise of graphene. Nat. Mater. 6, 183–191 (2007).
5. Avouris, P. Graphene: electronic and photonic properties and devices. Nano Lett. 10, 4285–4294 (2010).
6. Schwierz, F. Graphene transistors. Nat. Nanotechnol. 5, 487–496 (2010).
7. Lin, Y. M. et al. Operation of graphene transistors at gigahertz frequencies. Nano Lett. 9, 422–426 (2008).
8. Meric, I. et al. Current saturation in zero-bandgap, top-gated graphene field-effect transistors. Nanoscale, 3, 654–659 (2008).
9. Liao, L. et al. Sub-100 nm channel length graphene transistors. Nano Lett. 10, 3952–3956 (2010).
10. Liao, L. et al. High-speed graphene transistors with a self-aligned nanowire gate. Nature 467, 305–308 (2010).
11. Lin, Y. M. et al. 100-GHz transistors from wafer-scale epitaxial graphene. Science 327, 662 (2010).
12. Bai, J. W. et al. Top-gated carbon vapor deposition grown graphene transistors with current saturation. Nano Lett. 11, 2555–2559 (2011).
13. Meric, I. et al. Channel length scaling in graphene field-effect transistors studied with pulsed current-voltage measurements. Nano Lett. 11, 1093–1097 (2011).
14. Wu, Y. Q. et al. High-frequency, scaled graphene transistors on diamond-like carbon. Nature 472, 74–78 (2011).
15. Kim, K., Choi, J. Y., Kim, T., Cho, S. H. & Chung, H. J. A role for graphene in silicon-based semiconductor devices. Nature 479, 338–344 (2011).
16. Meric, I. et al. High-frequency performance of graphene field effect transistors with saturating IV-characteristics. Presented at the Electron Devices Meeting (IEDM), IEEE, (2011) doi: 10.1109/IEDM.2011.6131472.
17. Lee, S. et al. Tech. Dig. IEDM, 255–258 (IEEE, 2007).
18. Barreiro, A., Lazzeri, M., Moser, J., Mauri, F. & Bachtold, A. Transport properties of graphene in the high-current limit. Phys. Rev. Lett. 103, 076801 (2009).
19. Szafraniek, B. N., Fiori, G., Schall, D., Neumaijer, D. & Kurr, H. Current saturation and voltage gain in bilayer graphene field effect transistors. Nano Lett. 12, 1324–1328 (2012).
20. Franklin, A. D. et al. Sub-10 nm carbon nanotube transistor. Nano Lett. 12, 758–762 (2012).
21. Giovannetti, G. et al. Doping graphene with metal contacts. Phys. Rev. Lett. 101, 026803 (2008).
22. Khomyakov, P. A. et al. First-principles study of the interaction and charge transfer between graphene and metals. Phys. Rev. B 79, 195425 (2009).
23. Wu, Y. Q. et al. Quantum behavior of graphene transistors near the scaling limit. Nano Lett. 12, 1417–1423 (2012).
24. Farmer, D. B. et al. Chemical doping and electron-hole conduction asymmetry in graphene devices. Nano Lett. 9, 388–392 (2009).
25. Barraza-Lopez, S., Vanevic, M., Kindermann, M. & Chou, M. Y. Effects of metallic contacts on electron transport through graphene. Phys. Rev. Lett. 104, 076807 (2010).
26. Ni, Z. Y. et al. Tunable bandgap in silicene and germanene. Nano Lett. 12, 113–118 (2012).
27. A Zyliman, B. et al. Electronic transport and quantum hall effect in bipolar graphene p-n junctions. Phys. Rev. Lett. 99, 166804 (2007).
28. Huard, B. et al. Transport measurements across a tunable potential barrier in graphene. Phys. Rev. Lett. 98, 236803 (2007).
29. Williams, J. R., DeCarlo, L. & Marcus, C. M. Quantum hall effect in a gate-controlled p-n junction of graphene. Science 317, 638–641 (2007).
30. Huard, B., Stander, N., Sulzpio, J. A. & Goldhaber Gordon, D. Evidence of the role of contacts on the observed electron-hole asymmetry in graphene. Phys. Rev. B 78, 121402 (2008).
