Controllable p-n Junction Formation in Monolayer Graphene Using Electrostatic Substrate Engineering

Hsin-Ying Chiu,* Vasili Perebeinos, Yu-Ming Lin, and Phaedon Avouris*
IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598, United States

ABSTRACT We investigate electric transport in graphene on SiO2 in the high field limit and report on the formation of p-n junctions. Previously, doping of graphene has been achieved by using multiple electrostatic gates, or charge transfer from adsorbants. Here we demonstrate a novel approach to create p-n junctions by changing the local electrostatic potential in the vicinity of one of the contacts without the use of extra gates. The approach is based on the electronic modification not of the graphene but of the substrate and produces a well-behaved, sharp junction whose position and height can be controlled.

KEYWORDS Graphene, transistor, p-n junction, high-field transport, trap charges, substrate

While the low field carrier transport in graphene has already been the focus of numerous studies,1–6 the behavior of graphene devices in the high field limit has received much less attention.7–12 Thus, while it was shown that graphene itself can withstand current densities approaching 109 A/cm2,13,14 the behavior of graphene devices at high fields still remains largely unexplored. However, for nanoelectronic applications the interest is in the high field transport regime since devices are usually operated at or near the current saturation limit. Currently, little is known about graphene under those conditions. In this paper, we study the high field regime and show that while graphene itself is stable, there are crucial changes induced in the gate dielectrics that strongly affect the transport in graphene channel. This effect is enhanced in the atomic-layer thin body of the graphene channel. Thus, conclusions about the behavior of devices at high bias need to be carefully examined. Most importantly, we show that the electrostatic modifications of the substrate can be exploited to create an abrupt graphene p-n junction, which offers opportunities extending beyond CMOS technology, such as planar electron Veselago lenses.15 The process is controllable and we create these junctions with desired lengths and barrier heights without resorting to the use of multiple gates,7,16–23 or charge transfer from adsorbates,24–26 as is currently the case.

To explore high bias regime, we fabricated four-terminal, back-gated transistors from exfoliated graphene on an undoped Si substrate covered with 90 nm thick SiO2. As shown in Figure 1b, the graphene channel is 1 µm wide and 6 µm long and contacted by titanium source (S) and drain (D) electrodes.27 Two voltage probes, A and B, were placed along the channel separating it into three segments (denoted as SA, AB, and BD) with equal lengths of 2 µm. We first consider the case of p-type (hole) transport in the channel by choosing negative drain (V_D < 0) and gate voltages (V_G - V_CNP < 0, where V_CNP is the charge neutrality point, also referred to as the Dirac point). Overnight thermal annealing under vacuum minimizes the hysteretic behavior, the ob-

FIGURE 1. Device electrical characteristic and configuration. (a) The channel conductivity (σ) versus gate voltage (V_G) curve of a pristine graphene transistor at source-drain bias (V_D) as 10 mV. The four-probe carrier mobility is about 3500 cm2/(V s) in both electron and hole branches. The false-colored SEM image with 2 µm scale bar is shown in the inset. (b) Schematic of device configuration. One 6 µm long and 1 µm wide graphene layer is connected with S (source) and D (drain) contacts, while A and B terminals are voltage probes.
served clean "V" shape characteristic and a nearly zero $V_{\text{CNP}}$ in Figure 1a signify negligible unintentional doping$^{28}$ including water molecules$^{29}$ and good channel uniformity in the initial device. A doping profile along the channel is then created by applying a large drain bias, up to $-10\ \text{V}$, to electrically stress the devices. We have systematically varied the extent of the electrical stress by applying different gate voltages. The amount of doping induced by the high-field stress can subsequently be inferred from the change of $V_{\text{CNP}}$ in an $I$-$V$ gate sweep performed at low bias. Therefore, in this work, each high-bias measurement is followed by a low-bias gate sweep as a monitoring routine.

Our experimental procedure involves to first sweep $V_D$ up to $-10\ \text{V}$, as $V_G$ is varied from $V_{\text{CNP}}$ ($\sim 0\ \text{V}$ for this device) toward more negative gate voltages up to $V_G = -30\ \text{V}$ in steps of $-5\ \text{V}$. Figure 2a shows a series of the $I_D$-$V_D$ output characteristics of the graphene device. A clear hysteresis in the $V_D$ sweep is observed when $|V_G - V_{\text{CNP}}|$ is increased above $15\ \text{V}$ (green curve). Interestingly, this hysteresis is accompanied by a small plateaulike feature in the corresponding conductance, $G = I_D/V_D$, versus gate voltage curves in the subsequent low-bias measurement (see Figure 2b).

The hysteresis in the output characteristics becomes more pronounced as the device is stressed at larger $|V_G - V_{\text{CNP}}|$ biases, and the plateau in the $G$-$V_G$ curves eventually becomes a second dip to the left of the original Dirac point.

Using the four-probe configuration, the development of these features is monitored for each segment simultaneously. From the output characteristics and $G$-$V_G$ curves shown in Figure 2c,d, we find that only segment BD exhibits similar dip features, while those of the other two segments SA and AB do not show any discernible variation as a result of the stress (insets of Figure 2a–d). This indicates that the observed new features result from changes in the graphene channel in the vicinity of the drain contact. The effect of a possible asymmetrical initial contact resistance is ruled out and verified by swapping the polarity of the bias at the source/drain contacts. We also note that the voltage drop in the AB segment decreases while the voltage drop in segment DB increases; the overall performance of the device is eventually limited by this feature as shown in the inset of Figure 2a,c.

The two dips observed in the $G$-$V_G$ curves in both Figure 2b,d indicate the creation of a doped region near the drain contact as a result of the stressing at high drain biases in contrast to the initial, nearly pristine, graphene channel.
Furthermore, the position of the second dip depends on the gate voltage $V_G$ used during the stressing, while the original Dirac point stays unaffected. In Figure 3b, we plot the gate voltages corresponding to these two conductance minima as a function of the applied $V_G$ during the stress. The unvarying position of the main dip corresponds to the original, unaffected graphene domain in the channel, while the progressive evolution of the second dip demonstrates the correlation between the stress condition and the doping of the channel and the control that the stress voltage provides. We tested twelve devices with the same device geometry and the same result was observed. We thus conclude that the two dips (Dirac points) are topologically well-defined and can be controlled by the applied stress voltage $V_G$. The position of the second Dirac point relative to the initial one indicates the generation of positive charges close to the drain contact. The observation of these two distinct and well-defined dips indicates that the graphene is unlikely to be structurally damaged during stressing. This conclusion is further supported by the fact that the double Dirac feature can be eliminated by overnight thermal annealing in vacuum at 475 K and the original characteristics recovered with $V_{CNP} \approx 0$ V. The other possibility of field-induced accumulation associated with mobile charges (such as alkali atom contaminants) is also not likely, due to the absence of observable hysteresis in the original device and, most importantly, the fact that the application of stress voltages with the opposite polarity does not affect the double dip structure. Therefore, the above observations and the known properties of insulators under stress strongly suggest that trapped charge in the silicon oxide substrate is the source of the local doping under high bias conditions. The electron affinity energy and the band gap of silicon oxide are 0.95 and 8.9 eV, respectively, while the graphene work function is $\sim 4.5$ eV. Therefore, the barriers for an electron or hole to be injected into bulk traps in SiO$_2$ are $\sim 3.6$ and $\sim 4.3$ eV, respectively. These energies are too high to account for our observations. In addition, the leakage current into the back-gate remains constant and in the few picoamps range throughout the entire drain voltage sweep. Thus, we attribute the positive charges to trapped charge sites with energies within the SiO$_2$ bandgap and in the vicinity of the graphene/oxide interface.

To investigate whether there is charge carrier type preference for trapping in these sites, we performed similar stressing studies under $n$-type (electron) transport conditions (i.e., $V_D > 0$, $V_G - V_{CNP} > 0$) in other graphene devices with identical layouts. Contrary to the case of $p$-type transport, no new dip was observed even at higher current densities, except for a slight hysteresis in the output characteristics that is likely due to joule-heating and the modification of surface chemistry after current annealing. Therefore, under similar stress conditions, hole carriers are much easier to be injected into the traps than electrons, which indicates that the processed silicon oxide surface tends to be donorlike (that is, positively charged when empty and neutral when filled with an electron).

The two Dirac points seen in the gate voltage sweep are a signature of the formation of a $p$-$n$ junction in the graphene channel. To gain further insight, we use a step-potential model to investigate this process. A schematic of the $p$-$n$ junction band diagram is shown in Figure 3a.

**FIGURE 3.** A step-potential model of graphene $p$-$n$ junction and its analysis. (a) Schematic band diagram of a $p$-$n$ junction along the graphene channel. Region I indicates the main unaffected region while Region II represents the interfacial trap charges present near the interface. $L_0$ and $L$ are the full channel length and doped channel length, respectively, and $\Delta \phi$ is the potential step barrier between Region I and II. The red dashed line is a guide line for the chemical potential ($E_F$) varying with the gate voltage. (b) The relationship between $L$ and dip positions vs stress $V_G$. The constant main dip position is shown in red. Both the second developed dip (green curve) and the length $L$ (blue curve) vary almost linearly with stress condition. (c) and (d) show transfer characteristics at the stress condition $V_G = -20$ V. The experimental data are shown in black and the modeled results are shown in red for different lengths $L_0 = 6 \mu$m and $L_0 = 2 \mu$m in (c,d), respectively. As the gate varies, the graphene channel exhibit different junction ($p$-$p$, $p$-$n$ and $n$-$n$) configurations.
two regions are determined by the electrostatic potential as regions I and II as (independent, with the gate affecting the carrier numbers in resistance measurements in segments SD and BD of length \( L \)).

The total channel resistance \( R_T \) can be written as (see Supporting Information)

\[
R_T = R_C + \frac{A}{\sqrt{n_0^2 + n[V_C^*]^2}} + \frac{B}{\sqrt{n_0^2 + n_2[V_C^*]^2}}
\]

where \( R_C \) is the contact resistance, assumed to be gate independent, with the gate affecting the carrier numbers in regions I and II as \( (n_0^2 + n[V_C^*]^2)^{1/2} \). The parameters \( A \) and \( B \) are related to the mobilities \( \mu_1 \) and \( \mu_2 \), and residual electron–hole puddle densities \( n_0 \) and \( n_2 \) by \( A = (L_0 - L)/(e W \mu_1) \) and \( B = L/(e W \mu_2) \). The carrier densities in these two regions are determined by the electrostatic potential as \( en_1 = V_0^* C_{ox} \) and \( en_2 = C_{ox} V_0^* + en_0 \), where \( V_0^* = V_G - V_{CNP} \) and \( n_0 \) is the trap charge density. The latter can be inferred from the measured gate voltage difference \( \Delta V \) between the two resistance maxima as \( en_0 = C_{ox} \Delta V \). From the low-bias resistance measurements in segments SD and BD of length \( L_0 = 6 \) and 2 \( \mu \)m correspondingly, and the length \( L \) of Region II can be deduced from the following relation \( A_1/A_2 = (6 - L)/(2 - L) \), which is insensitive to the mobility value \( \mu_1 \).

The relationship between length \( L \) and the stress condition \( V_G \) is shown in Figure 3b. Figure 3c,d shows the measured resistance of segments SD and BD after device stressing at \( V_G = -20 \) V and the corresponding fit obtained from the step potential model described by eq 1. The excellent agreement between the measurements and the modeled curves confirms the abruptness of the p-n junction formed. These results demonstrate a novel, straightforward scheme to create a p-n junction in graphene controllable by the stress conditions, that is, \( V_G \). The long-term stability of the junction as a function of the environmental conditions will be the subject of a future study. To determine the threshold conditions required for the formation of interfacial trap charges, a systematic stress study was performed on another graphene device (Device 2) with identical geometry. Different from the previous stress procedures, the stress was applied at a constant gate voltage \( V_G \) while varying the drain voltage successively from \(-7 \) to \(-10 \) V. The evolution of the device characteristics at two stress gate voltages, \( V_G = -16 \) and \(-26 \) V, is shown in Figure 4a,b, respectively. By performing an analysis of the data using the described method, \( L \) and \( \Delta V \) as a function of stress \( V_G \) at a sequence of different \( V_G \) are shown in Figure 4c,d, respectively. We note that both the hysteresis and the distinct features in output and \( G-V_G \) curves occur at \( V_D = -8.5 \) V and \( V_G = -16 \) V. (When \( |V_D| \) is below 8.5 V, we cannot distinguish the voltage difference between the two peaks). From the four-terminal \( I-V \) curve and the known contact resistance for this device \(~1.9 \) k\( \Omega \) \( \mu \)m in the p-branch, the average onset field (not peak field) is estimated to be \(~2 \) V/\( \mu \)m across the BD segment.
To understand the dependence of the spatial position of the trapped charges as a function of the bias conditions, we consider a simple model based on the current continuity equation

\[ j = n_{\text{tot}}(x)v(x) \]  

where \( j \) is a measured current density, \( v(x) \) is drift velocity of the carriers, and \( n_{\text{tot}}(x) = (n_0^2 + n(x))^{1/2} \) is the carrier density, where we use a typical electron–hole puddle density of \( n_0 = 5 \times 10^{11} \text{ cm}^{-2} \). The carrier concentration \( n(x) \) in the channel is determined by the electrostatic gate and drain biases and the trap carrier density according to

\[ n(x) = -C_{\text{ox}}(V_G - V(x))/e - n_{\text{it}} \]

\[ n(x) = (eV(x) - E_F(x))(eV(x) - E_F(x))/\pi\hbar^2 v_F^2 \]

where the self-consistent potential \( V(x) \) is found by assuming a linear drop of the Fermi energy \( E_F(x) \) from source to drain reduced by the voltage drop at the contacts of about 30%, as determined by our four-probe measurements. Since the carrier density \( n(x) \) decreases as the position moves toward the drain electrode, it is expected that the velocity \( v(x) \) will become higher near the drain contact. A hotter carrier distribution at the drain side yields an enhanced population of the interfacial trap states in the silicon oxide. In particular, these high energy carriers are expected in the region where the traps have already been produced by the previous \( I-V_D \) stress sweep. To quantify the energy of the hot carrier distribution, we assume that at a finite bias the positively moving carriers have higher energy than the negatively moving carriers by an energy \( \Delta E \). The drift velocity is related to \( \Delta E \) according to

\[ \Delta E = \frac{\pi}{2} \hbar v(x)\sqrt{n(x)} \]

In Figure 5a, we estimate the energy \( \Delta E \) as a function of position along the channel using the experimentally extracted values of current density of Device 1 at different stress condition \( V_G \) with fixed \( V_D = -10 \text{ V} \). The transfer characteristics of a 200 nm long graphene device under the bias stress \( V_D = -2.5 \text{ V} \) at various \( V_G \) conditions. There is no sign of second dip, indicating that the channel is uniformly doped. (d) The main dip as a function of the stress voltage in the 200 nm long device, showing a slope of \(-0.5\).
As shown in Figure 4c, the length of the doped Region II can be extended to be as long as 1 μm. This indicates that for a shorter channel device, the entire graphene channel could be doped, causing a smooth shift of the Dirac point instead of the occurrence of two resistance maxima. To verify this uniform doping effect in a short channel device, a two-terminal graphene device with 200 nm channel length was fabricated. Following similar procedures for creating an n-doped junction, we find that the transfer characteristics remain dominated by only a single Dirac point, which shifts toward negative voltages, as shown in Figure 5c. The Dirac position depends linearly on the stress gate voltage, as shown in Figure 5d. This result indicates that we achieved uniform charging of the interfacial traps along the channel in a short-channel device.

In conclusion, we have shown that a sharp p-n junction can be produced in monolayer graphene device structures through the electrostatic modification of the gate insulator surface. Furthermore, we demonstrated the ability of controlling both the length and the potential difference of the doped region along the graphene transistor channel. This novel approach to produce a well-defined p-n junction without extra gates or adsorbates opens up new opportunities for studying quantum transport in graphene devices and the results provide new insight into the phenomena that can take place at the high field transport limit, also demonstrating the need to clearly distinguish between substrate-induced effects and the intrinsic high-field transport behavior of graphene.

Acknowledgment. We thank B. Ek, J. Bucchignano, and G. Wright for expert technical assistance. We also thank James Stathis, Marcus Freitag, Marc Bockrath, Shu-Jen Han, Damon Farmer, Ching-Tzu Chen, Wenjuan Zhu, and Yanqing Wu for helpful discussions. This work was supported by DARPA under Contract FA8650-08-C-7838 through the CERA program. The views, opinions, and/or findings concerning this work are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the Department of Defense.

Supporting Information Available. Details of sample fabrication, electrical measurements, data analysis method, and additional fitting model with e-h puddles. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(15) Cheianov, V. V.; Fal’ko, V.; Alshuler, B. L. Science 2007, 317 (5816), 1252.
(27) Two types of metal composition have been used in our study and show the similar transport behaviors. One is the stack of Ti/Au (20 nm/30 nm) and the other is Ti/Pd/Au (0.5 nm/20 nm/30 nm).
(28) The estimated residual concentration for this device is n0 = ~5 × 10^{11} cm^{-2}.
(38) For voltage probing, terminal A and B are set at either the p-branch or e-branch.
(39) We calculate the contact resistance in the following way: R_{contact} = \frac{V_{oc}}{I_{oc}}, where V_{oc} is the voltage between each terminal and R_{oc} is the resistance between two different data sets from the same run: one for whole channel, segment S1, and second for segment BD.
Supplementary Information

Controllable $p$-$n$ junction formation in monolayer graphene using electrostatic substrate engineering

Hsin-Ying Chiu*, Vasili Perebeinos, Yu-Ming Lin, Phaedon Avouris*

IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598, USA

To whom correspondence should be addressed. E-mail: (H.-Y.C) hchiu@us.ibm.com; (P.A) avouris@us.ibm.com

1. Sample Fabrication

Devices are prepared by mechanically exfoliating Kish graphite (Toshiba Ceramics) on undoped Si substrate caped with 90 nm of thermal grown oxide. Metal contacts in graphene devices were fabricated by standard e-beam lithography and lift-off. Typically, two types of metal composition have been used in our study and show the similar transport behaviors. One is the stack of Ti/Au (20 nm/30 nm) and the other is Ti/Pd/Au (0.5 nm/20 nm/30 nm). In the second case, Ti is used for the adhesion layer. Thereafter, the channel is formed by etching the undesirable graphene area using oxygen plasma by reactive-ion etching (Unaxis RIE). Prior to any electrical measurement, overnight thermal annealing at 475K is performed in vacuum (~ $10^{-6}$ torr). This process is used to remove the gas absorbed and water molecules on the graphene surface and also to discharge any interfacial traps at the graphene-oxide interface.

2. Electrical Measurements

All measurements are completed in a LakeShore probe station at room temperature with the semiconductor device analyzer (Agilent Technologies B1500A). In the beginning of measurements, there is only slight hysteresis in the transfer characteristics ($I$-$V_G$ curve), and the mobility of these as-prepared device is estimated to be 3500 cm$^2$/Vs from the four-terminal measurement. Characteristic of Device I was shown in Fig. 1(a) and its estimated residual concentration $n_0$ is $\sim 3.9 \times 10^{11}$ cm$^{-2}$. The carrier density of graphene can be tuned by the back-gate electrostatic potential. For the 90 nm-thick silicon oxide with a dielectric constant of 3.9, the carrier density is given by $n \sim 2.4 \times 10^{11} (V_G - V_{CNP})$ cm$^{-2}$.

3. Data Analysis Method

Four-terminal measurements provide both voltage and current information for each terminal. A schematic diagram is shown in Fig 1(b). For voltage probing, terminal A and B are set at $I_{A,B} = 0$ A and the voltage reading during the measurement is monitored. Thus,
we can extract data between each terminal and compile two different data sets from the same run; one for whole channel, segment SD, and second for segment BD.

Our analysis is based on equation (1) and the corresponding schematic potential profile is shown in Fig 3(a). In the fit we identify the two peak positions as fixed parameters ($V_{CNP}$ and $\Delta V$). The remaining parameters are obtained from the fitting of the different length data sets. $n_{01}$ and $n_{02}$ are global parameters for the pristine and doped regions in both data sets. Note that in fitting the resistance $R_C$ in the BD region a value twice as small as that used in Eq. (1) to fit the SD channel, was obtained, as expected.

4. Fitting Model with $e$-$h$ Puddles

To account for the electron-hole puddle density near the Dirac point in fitting the experimental resistance a simple analytical model has been proposed in Ref.[37]:

$$ R = \frac{L}{W e \mu \sqrt{n_0^2 + n[V_C]^2}} \quad \text{(S1)} $$

where $L$ and $W$ are the length and the width of the channel respectively, $\mu$ is a mobility, $n_0$ is the electron-hole puddle concentration, and $n$ is gate dependent charge carrier concentration. While Eq. (S1) has been found to work well in fitting the experimental data by several research groups, its extension to the presence of the two Dirac points (Eq. (1) of the main text) has not been tested yet. Here we show that Eq. (S1) can be derived from a microscopic model and it is extension to the presence of the two Dirac points in the channel (Eq. (1) of the main text) reproduces well by the same model.

We assume that there are multiple parallel channels contributing to the current and each of them can be characterized by its own value of the Dirac point. The energy distribution of the Dirac points is assumed to be Gaussian:

$$ F(E) = \frac{1}{t \sqrt{\pi}} \exp \left( - \left( \frac{E}{t} \right)^2 \right) \quad \text{(S2)} $$

where $t$ is the width of the distribution. The electron $n_e$ and hole $n_h$ carrier concentrations can readily be calculated as:

$$ n_e = n_e + n_h = \int_{-\infty}^{\infty} dE F(E) n(E) = \frac{1}{\pi \hbar^2 v_F^2} \left( \frac{t^2}{2} + E_f^2 \right) \quad \text{(S3)} $$

$$ n_e = \int_{E_F}^{\infty} dE F(E) n(E) $$
where \( n(E) = \frac{(E - E_F)^2}{\pi \hbar^2 v_F^2} \) and \( E_F \) is the averaged Fermi level. In Fig. 1S (a), we show that the total carrier density \( n_t \) dependence on the gate voltage determined by the charge carrier density \( e(n_e - n_h) = C_G V_G \) nearly coincides with a simple analytical form
\[
n = \sqrt{n_0^2 + C_G^2 v_F^2 / e^2}, \quad \text{when} \quad n_0 = \frac{t^2}{(2\pi \hbar^2 v_F^2)}.
\]

In the presence of two regions with an averaged Dirac position in the second region being displaced by \( \Delta \phi \), the energy distribution function is given by:
\[
F_2(E) = \frac{1}{t\sqrt{\pi}} \exp \left[ -\left( \frac{E + \Delta \phi}{t} \right)^2 \right] \quad (S4)
\]

In principle, the electron-hole carrier density in the second region can be different, but we keep it the same for simplicity. The conductivity of the second region similarly to Eq. (3) is determined by the carrier concentrations:
\[
n_{v2} = n_{e2} + n_{h2} = \int_{-\infty}^{\infty} dE F_2(E) n(E) = \frac{1}{\pi \hbar^2 v_F^2} \left[ \frac{t^2}{2} + (E_F + \Delta \phi)^2 \right] \quad (S5)
\]
\[
n_{e2} = \int_{E_F}^{\infty} dE F_2(E) n(E)
\]

The total resistance of the sequential channels is given proportional to:
\[
\frac{1}{N} = \frac{L - L_0}{L n_{i1}} + \frac{L_0}{L n_{i2}} \quad (S6)
\]

In Fig. 1S (b), we show that the net carrier density \( N \) dependence on the gate voltage determined nearly coincides with a simple analytical form used in Eq. (1) of the main text.
Figure S1. Model calculations of the total carrier density contributing to the resistivity. (a), The blue curve shows calculations using Eq. (3), and the red one using a simple analytical form Eq. (1S) for $n_0 = 5 \times 10^{11} \text{ cm}^{-2}$. (b) The blue curve shows calculations using Eq. (6S), and the red one using a simple analytical form Eq. (1) from the main text for $n_0 = 5 \times 10^{11} \text{ cm}^{-2}$ and $\Delta \phi = 20 \text{ V}$. 