Field-Effect Tunneling Transistor Based on Vertical Graphene Heterostructures


1School of Physics & Astronomy, University of Manchester, Manchester M13 9PL, UK. 2Manchester Centre for Mesoscience & Nanotechnology, University of Manchester, Manchester M13 9PL, UK. 3Institute for Molecules and Materials, Radboud University of Nijmegen, 6525 AJ Nijmegen, Netherlands. 4School of Physics & Astronomy, University of Nottingham, Nottingham NG7 2RD, UK. 5Institute for Microelectronics Technology, 142432 Chernogolovka, Russia. 6Departamento de Física, Universidade do Minho, P-4710-057, Braga, Portugal. 7Momentive Performance Materials, 22557 West Lunn Road, Strongsville, OH 44070, USA.

*To whom correspondence should be addressed. E-mail: kostya@manchester.ac.uk

An obstacle to the use of graphene as an alternative to silicon electronics has been the absence of an energy gap between its conduction and valence bands, which makes it difficult to achieve low power dissipation in the OFF state. We report a bipolar field-effect transistor that exploits the low density of states in graphene and its one atomic layer thickness. Our prototype devices are graphene heterostructures with atomically thin boron nitride or molybdenum disulfide acting as a vertical transport barrier. They exhibit room temperature switching ratios of ≈50 and ≈10,000 respectively. Such devices have potential for high frequency operation and large-scale integration.

The performance of graphene-based field effect transistors (FETs) has been hampered by graphene’s metallic conductivity at the neutrality point (NP) and the unimpeded electron transport through potential barriers caused by Klein tunneling, which limit the achievable ON-OFF switching ratios to ~10 and those achieved so far at room temperature to <10 (1–7). These low ratios are sufficient for individual high-frequency transistors and analog electronics (4–7) but they present a fundamental problem for any realistic prospect of graphene-based integrated circuits (1–7). A possible solution is to open a band gap in graphene, for example by using bilayer graphene (8, 9), nanoribbons (10, 11), quantum dots (11) or chemical derivatives (12) but it has proven difficult to achieve high ON-OFF ratios without degrading graphene’s electronic quality.

We report an alternative graphene transistor architecture, namely a field-effect transistor based on quantum tunneling (13–17) from a graphene electrode through a thin insulating barrier (in our case, hexagonal boron nitride (hBN) or molybdenum disulfide of ~1 nm thickness). The operation of the device relies on the voltage tunability of the tunneling density of states (DOS) in graphene and of the effective height Δ of the tunnel barrier adjacent to the graphene electrode. To illustrate the proposed concept we concentrate on graphene-hBN-graphene devices [An alternative barrier material (MoS2) is discussed in Supporting online material (18)].

The structure and operational principle of our FET are shown in Fig. 1. For convenience of characterization, both source and drain electrodes were made from graphene layers in the multiterminal Hall bar geometry (18). This device configuration allowed us to measure not only the tunnel current-voltage curves (I-V) but also to characterize the response of the graphene electrodes, thus providing additional information about the transistor operation. The core graphene-hBN-graphene structure was encapsulated in hBN to allow higher quality of the graphene electrodes (19, 20). To fabricate the device shown in Fig. 1A, we first used the standard cleavage technique (24) to prepare relatively thick hBN crystals on top of an oxidized Si wafer (300 nm of SiO2), which acted as a gate electrode (Fig. 1; fig. S1). The crystals served as a high-quality atomically-flat substrate and a bottom encapsulation layer (19). Monolayer graphene (Gr0) was then transferred onto a selected hBN crystal (20-50 nm thick) using a dry transfer procedure (19, 25). After deposition of metal contacts (5nm Ti/50nm Au) and etching to form a multiterminal Hall bar mesa, the structure was annealed at 350°C in forming gas. A few-atom-thick hBN crystal was identified (26) and transferred on top of Gr0 by using the same procedures. This hBN layer served as the tunnel barrier. The whole process of positioning, annealing, and defining a Hall bar was repeated to make the second (top) graphene electrode (Gr1). Finally, a thick hBN crystal encapsulated the entire multilayer structure (Fig. 1A; fig. S1). Further details of our multistep fabrication procedures can be found in refs. (18, 25). We tested devices with tunnel barriers having thickness d from 1 to 30 hBN layers (18). To illustrate
the basic principle of the tunneling FETs, we focus on the data obtained from four devices with a tunnel barrier made of 4-7 layers and discuss the changes observed for other d.

When a gate voltage $V_g$ was applied between the Si substrate and the bottom graphene layer (GrB), the carrier concentrations $n_B$ and $n_T$ in both bottom and top electrodes increased because of the weak screening by monolayer graphene (21), as shown schematically in Fig. 1C. The increase of the Fermi energy $E_F$ in the graphene layers could lead to a reduction in $\Delta$ for electrons tunneling predominantly at this energy (18). In addition, the effective height also decreased relative to the NP because the electric field penetrating through GrB altered the shape of the barrier (decreased relative to the NP because the electric field at this energy ($23$) contributions could dominate changes in the increasingly greater fraction of the gate-induced electric field being screened out by GrB ($28$). In addition, the effective height also decreased relative to the NP because the electric field penetrating through GrB altered the shape of the barrier (22, 23). Furthermore, the increase in the tunneling DoS as $E_F$ moved away from the NP (21) led to an increase in the tunnel current $I$. Depending on parameters, any of the above three contributions could dominate changes in $I$ with varying $V_g$.

We emphasize that the use of graphene in this device architecture is critical because this exploits graphene’s low DOS which, for a given change in $V_g$, led to a much greater increase in $E_F$ compared to a conventional two-dimensional gas with parabolic dispersion [cf. (13–17)]. This difference translated into much greater changes of both $\Delta$ and tunneling DOS.

Figure 2A shows the behavior of in-plane resistivity $\rho$ for the GrB and GrT layers as a function of $V_g$. The curves indicate little residual doping for encapsulated graphene ($\approx 0$ and $<10^{11}$ cm$^{-2}$ for GrB and GrT, respectively). In both layers, $\rho$ strongly depended on $V_g$ showing that GrB did not screen out the electric field induced by the Si-gate electrode. The screening efficiency was quantified by Hall effect measurements (Fig. 2B-D) which showed that the gate induced approximately the same amount of charge in both layers at low concentrations; that is, there was little screening if $n_B$ was small. As the concentration in GrB increased, the $n_B(V_g)$ and $n_T(V_g)$ dependences became super- and sub-linear, respectively (Fig. 2B&C), because of the increase in $n_B$, which led to an increasingly greater fraction of the gate-induced electric field being screened out by GrB (18). Hence more electrons accumulated in the bottom graphene electrode and fewer reached the top electrode. The total charge accumulated in both layers varied linearly in $V_g$ (Fig. 2D), as expected. We could describe the observed redistribution of the charge between the two graphene layers in terms of the corresponding sequential circuit including the quantum capacitance (13, 27) of the graphene layers (fig. S2). Note that, for a parabolic band, the ratio between $n_B$ and $n_T$ would be independent on $V_g$ and, therefore, the electric field penetrating into the tunnel barrier would be substantially reduced even in the limit of zero $n_B$ (13).

A bias voltage $V_h$ applied between GrB and GrT gave rise to a tunnel current through the thin hBN barrier that scaled with device area. Figure 3A shows I–V characteristics for one of our devices at various $V_g$. First, we consider the case of zero $V_h$. At low $V_g$, $I$ was linear in bias, yielding a tunnel resistivity $\rho_T = V/I = 100$ GΩ·μm$^2$ for this hBN thickness. At higher voltages ($V_h$ above $\approx 0.1V$), $I$ grew more rapidly. The I–V curves could be described (inset in Fig. 3A; fig. S3) by the standard quantum-tunneling formulae (22, 23) assuming energy conservation but no momentum conservation at the mismatched graphene-hBN interface (28).

As shown below, we could distinguish experimentally between electron and hole tunneling and found that the tunneling was due to holes. This result is in agreement with a recent theory for the graphene-hBN interface (29), which reports a separation between the Dirac point in graphene and the top of the hBN valence band of $\approx 1.5eV$ whereas the conduction band is $\approx 4eV$ away from the Dirac point. The fit to our data with $\Delta = 1.5eV$ yielded a tunneling mass $m = 0.5 m_0$ ($m_0$ is the free electron mass), in agreement with the effective mass for holes in hBN (30). Furthermore, our analysis indicated that $I$ varied mainly with the change in the tunneling DOS, whereas the change in tunneling probability with applied bias was a secondary (albeit important) effect (18). For our atomically-thin hBN barriers with relatively low $\rho_T$, we were not in a regime of exponential sensitivity to changes in $\Delta E_F(V_h)$.

We demonstrate transistor operation in Fig. 3A, which plots the influence of gate voltage on $I$. $V_g$ substantially enhanced the tunnel current, and the changes were strongest at low bias. The field effect was rather gradual for all gate voltages up to $\pm 50V$, a limit set by the electrical breakdown of our SiO$_2$ gate dielectric (typically $\approx 60$ V). This response is quantified in Fig. 3B, which plots the low-bias tunneling conductivity $\sigma = I/V_h$ as a function of $V_g$. The influence of $V_g$ was highly asymmetric: $\sigma$ changed by a factor of $\approx 20$ for negative $V_g$ (holes) and by a factor of 6 for positive $V_g$ (electrons). We observed changes up to $\approx 50$ for hole tunneling in other devices and always the same asymmetry (18; fig. S4). Also, the ON-OFF ratios showed little change between room and liquid-helium temperatures, as expected for $\Delta \gg$ thermal energy.

To analyze the observed behavior of $\sigma(V_g)$, we modeled the zero-bias conductivity by using the relation $\sigma = \sigma_{\rm DOS}\delta(V_g) = \sigma_{\rm DOS}(V_g) = \sum_i \sigma_{\rm DOS}(V_g) / i$. Where the indices refer to the two graphene layers and $\sigma_{\rm DOS}$ is the transmission coefficient through the hBN barrier (22, 23). The resulting curve shown in Fig. 3B accounts qualitatively for the main features in the measured data, using self-consistently the same tunneling parameters $m$ and $\Delta$ given above. At $V_g$ near zero, corresponding to tunneling from states near the NP, the tunneling DOS in both graphene layers was small and nonzero, and was the result of residual doping, disorder and temperature effects (18). The application of a gate voltage of
either polarity led to a higher DOS and, therefore, higher $\sigma^\dagger$. The gradual increase in $\sigma^\dagger(V_g)$ for both polarities in Fig. 3B was therefore caused by to the increasing DOS. However, $V_g$ also affected the transmission coefficient. Because of the shift of $E_F$ with changing $V_g$, the effective barrier height $\Delta$ decreased for one sign of charge carriers and increased for the other (Fig. 1B), which explains the asymmetry in both experimental and calculated $\sigma^\dagger(V_g)$ in Fig. 3B in terms of the change in $T(V_g)$. For our devices, the effect of $V_g$ on $T(V_g)$ was relatively weak (non-exponential) and comparable with the effect caused by changes in the tunneling DOS. The sign of the asymmetry infers that the hBN barrier height was lower for holes than for electrons, in agreement with the graphene-hBN band structure calculations (29). The weaker dependence of $I$ on $V_g$ at high bias can also be understood in terms of the more gradual increase in the tunneling DOS and in $E_F$ at high doping ($V_g$ = 0.5V correspond to $n_B$ $\approx$ 10$^{11}$ cm$^{-2}$).

Our results and analysis suggest that higher ON-OFF ratios could be achieved by using either higher $V_g$ or making devices with larger $d$, so that the tunneling depends exponentially on bias and is controlled by the barrier height rather than the DOS. The former route is limited by the electrical breakdown of dielectrics at ~1V/nm ($V_g$ = 300V for our SiO$_2$ thickness). By extrapolating the analysis shown in Fig. 3B to such voltages, we found that ON-OFF ratios >10$^4$ would be possible for our 4-7 layer devices if SiO$_2$ of highest quality were used. However, it would still require unrealistically large $V_g$ to enter the regime where $E_F$ becomes comparable with $\Delta$ and changes in $\sigma^\dagger(V_g)$ are exponentially fast. Therefore, we explored the alternative option and investigated devices with both thinner and thicker hBN barriers. For 1 to 3 hBN layers, zero-bias $\sigma^\dagger$ increased exponentially with decreasing number of layers, consistent with quantum tunneling, and we observed a weaker influence of $V_g$ on $I$, as expected for the more conductive regime. On the other hand, the thicker hBN barriers were prone to electrical breakdown. Nonetheless, for a few devices with $d$ $\approx$ 6 to 9 nm, we were able to measure a tunnel current without breakdown. A current >10pA appeared at biases of several volts and increased exponentially with $V_g$. The thicker devices’ I-V characteristics could be fitted using the same hole-tunneling parameters used above, thus indicating quantum tunneling rather than an onset of electrical breakdown. Unfortunately, no substantial changes (i.e., exceeding 50%) in the tunnel current could be induced by $V_g$. This insensitivity to gate voltage remains to be understood but was probably caused by charge traps that screened the influence of the gate.

An alternative method to achieve an exponential dependence of the tunneling current on gate voltage would be to use a barrier dielectric with a smaller $\Delta$, that would be comparable with typical $E_F$ realizable in graphene. One of such candidate materials is MoS$_2$, which has a band gap of about 1.3eV and can be obtained in a mono- or few-layers state similar to hBN and graphene (24). Our first hBN-graphene-MoS$_2$-graphene-hBN devices demonstrate ON-OFF ratio close to 10.000 (fig. S5), which is sufficient for certain types of logic circuits.

We conclude that our tunneling devices offer a viable route for high speed graphene-based analog electronics. The ON-OFF ratios already exceed those demonstrated for planar graphene FETs at room temperature by a factor of 10 (3–7). The transit time for the tunneling electrons through the nanometer-thick barriers is expected to be extremely fast (a few fs) (13–17) and exceeds the electron transit time in submicron planar FETs. It should also be possible to decrease the lateral size of the tunneling FETs down to the 10 nm scale, a requirement for integrated circuits. Furthermore, there appears to be no fundamental limitation to further enhancement the ON-OFF ratios by optimizing the architecture and by using higher-quality gate dielectrics and, in particular, lower tunnel barriers ($\Delta$ $\leq$ maximum achievable $E_F$).

References and Notes

6. S. J. Han et al., High-frequency graphene voltage amplifier. Nano Lett. 11, 3690 (2011).
18. Supporting material on *Science* Online.
35. Gmelin, Gmelin handbook of inorganic and organometallic chemistry (SpringerVerlag, Berlin, 1995).

**Acknowledgments:** Supported by the European Research Council, European Comission FP7, Engineering and Physical Research Council (UK), the Royal Society, U.S. Office of Naval Research, U.S. Air Force Office of Scientific Research, and the Körber Foundation. A. M. acknowledges support from the Swiss National Science Foundation.

**Supporting Online Material**

www.sciencemag.org/cgi/content/full/1218461/DC1

SOM Text
Figs. S1 to S5

References (31–35)

27 December 2011; accepted 23 January 2012
Published online 2 February 2012, 10.1126/science.1218461

---

**Fig. 1.** Graphene field-effect tunneling transistor. (A) Schematic structure of our experimental devices. In the most basic version of the FET, only one graphene electrode (GrB) is essential and the outside electrode can be made from a metal. (B) The corresponding band structure with no gate voltage applied. (C) The same band structure for a finite gate voltage $V_g$ and zero bias $V_b$. (D) Both $V_g$ and $V_b$ are finite. The cones illustrate graphene’s Dirac-like spectrum and, for simplicity, we consider the tunnel barrier for electrons.

**Fig. 2.** Graphene as a tunneling electrode. (A) Resistivities of the source and drain graphene layers as a function of $V_g$. (B-D) Carrier concentrations in the two layers induced by gate voltage, which were calculated from the measured Hall resistivities $\rho_{xy}$ using the standard expression $n = B/e\rho_{xy}$, where $B$ is the magnetic field and $e$ the electron charge. Close to the NP, the spikes appear (shown by dotted curves) because the above expression is not valid in the inhomogeneous regime of electron-hole puddles. The shown device has a 4-layer hBN barrier.

**Fig. 3.** Tunneling characteristics for a graphene-hBN device with $6 \pm 1$ layers of hBN as the tunnel barrier. (A) I-Vs for different $V_g$ (in 10V steps). Note, that due to finite doping, the minimum tunneling conductivity is achieved at $V_g = 3V$. The inset compares the experimental I-V at $V_g = 5V$ (red curve) with theory (dark) which takes into account the linear DoS in the two graphene layers and assumes no momentum conservation. Further examples of experimental curves and their fitting can be found in Supporting online material (I8), (B) Zero-bias conductivity as a function of $V_g$. The symbols are experimental data, and the solid curve is our modeling.
The curve is slightly shifted with respect to zero $V_h$ because of remnant chemical doping. In all the calculations, we assumed the hole tunneling with $m = 0.5m_0$ and $\Delta \approx 1.5$ eV (29, 30) and used $d$ as measured by atomic force microscopy. Both $I$ and $\sigma$ are normalized per tunnel area, which was typically 10 to 100 $\mu$m$^2$ for the studied devices. Temperature: 240 K.