Properties of Self-Aligned Short-Channel Graphene Field-Effect Transistors Based on Boron-Nitride-Dielectric Encapsulation and Edge Contacts

Tarun Chari, Inanc Meric, Cory Dean, and Kenneth Shepard, Fellow, IEEE

Abstract—We present the characterization of ballistic graphene field-effect transistors (GFETs) with an effective oxide thickness of 3.5 nm. Graphene channels are fully encapsulated within hexagonal boron nitride, and self-aligned contacts are formed to the edge of the single-layer graphene. Devices of channel lengths ($L_G$) down to 67 nm are fabricated, and a virtual-source transport model is used to model the resulting current–voltage characteristics. The mobility and source-injection velocity as a function of $L_G$ yields a mean-free-path, ballistic velocity, and effective mobility of 850 nm, $9.3 \times 10^7$ cm/s, and 13700 cm$^2$/Vs, respectively, which are among the highest velocities and mobilities reported for GFETs. Despite these best-in-class attributes, these devices achieve transconductance ($g_m$) and output conductance ($g_d$) of only 600 and 300 $\mu$S/\mu$m, respectively, due to the fundamental limitations of graphene's quantum capacitance and zero-bandgap. $g_m$ values, which are less than those of comparable ballistic silicon devices, benefit from the high ballistic velocity in graphene but are degraded by an effective gate capacitance reduced by the quantum capacitance. The $g_d$ values, which limit the effective power gain achievable in these devices, are significantly worse than comparable silicon devices due to the properties of the zero-bandgap graphene channel.

Index Terms—Graphene field-effect transistor (GFET), heterostructures, quantum capacitance, virtual source (VS).

I. INTRODUCTION

GRAPHENE, the single-atom thick, hexagonally lattice carbon allotrope, has drawn significant interest as an electronic channel material for field-effect transistors (FETs) on both rigid [1]–[4] and flexible substrates [5]–[8] due to its high low-field carrier mobility ($>100\,000$ cm$^2$/Vs at room temperature [9], [10]), atomically thin form, and exceptional mechanical flexibility (1-TPa Young's modulus [11]).

Manuscript received June 18, 2015; accepted September 21, 2015. This work was supported by Semiconductor Research Corporation within the Nanoelectronics Research Initiative through the INDEX Center. The review of this paper was arranged by Editor M. J. Kumar. (Corresponding author: Kenneth Shepard.)

T. Chari and K. Shepard are with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA (e-mail: tchari@ee.columbia.edu; shepard@ee.columbia.edu).

I. Meric is with Intel Corporation, Santa Clara, CA 95052-8119 USA (e-mail: inancmeric@gmail.com).

C. Dean is with the Department of Physics, Columbia University, New York, NY 10027 USA (e-mail: corydean@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TED.2015.2482823

Graphene, however, does not possess a bandgap, creating the possibility for ambipolar conduction in the channel under appropriate bias conditions. The lack of a bandgap also makes it difficult to turn the device OFF, resulting in low peak ON-current-to-OFF-current ratios (at best, less than 100 [12]) and making these devices largely impractical for digital applications. Instead, efforts have been focused on exploiting graphene FETs (GFETs) for analog and RF applications which require power amplification and take the advantage of device transconductance ($g_m$) and output conductance ($g_d$). A number of alternate 2-D materials for devices have been proposed, including materials that have a bandgap, such as MoS$_2$ [13]; bilayer devices, including graphene bilayer devices [14]; and heterostructure devices that include graphene combined with wider bandgap materials [15]. None of these alternate materials has yet achieved the mobilities of boron nitride (BN)-encapsulated single-layer graphene.

GFET devices have traditionally been challenged by poor dielectric interfaces and poor ohmic contacts. Graphene's electrical transport properties are strongly affected by anything in contact with its surface; among the highest reported mobilities are those of graphene suspended in vacuum [10]. Carrier mobilities and transport velocities are negatively impacted by impurities adsorbed on exposed graphene surfaces by charge-trapping states at graphene–dielectric interfaces. GFETs have typically been fabricated on silicon dioxide substrates with a top-gate dielectric thickness greater than 5 nm, made from high-$\kappa$ materials, such as HfO$_2$ or Al$_2$O$_3$ [grown by atomic layer deposition (ALD)]. These devices exhibit low-field mobilities that are typically less than 1000–3000 cm$^2$/Vs [1], [16], [17] and significant hysteresis with changes in gate voltage due to trapped charge from the substrate and gate oxide [18], [19]. Contacts are conventionally formed by depositing metal over an area of the graphene surface to form the source and drain. These approaches yield a contact resistance that is typically 1 k$\Omega$-$\mu$m and no better than 200 $\Omega$-$\mu$m [20]–[23]. Device current–voltage ($I$–$V$) characteristics for GFETs, particularly at short-channel lengths ($L_G$), have generally not shown strong saturating characteristics, due in part to these large contact resistances combined with the relatively poor electrostatics of the large oxide thickness ($t_{ox}$).

The use of hexagonal BN (h-BN) for dielectric encapsulation of graphene allows low-field mobilities in excess...
of 100,000 cm²/Vs to be achieved at room temperature in these heterostructures [9], comparable with the best mobilities achieved in suspended samples [10]. Like graphene, h-BN is a hexagonally latticed crystal, exhibiting strong in-plane bonding. Consequently, there are no out-of-plane dangling bonds, making it an ideal substrate and gate dielectric for graphene [24]. Furthermore, its layered structure permits the use of extremely thin h-BN to serve as the gate dielectric, permitting small effective oxide thickness (EOT). Though attempts have been made to use a graphene/h-BN heterostructure as an FET, the graphene in these structures has still been exposed to impurities during fabrication, leading to less than optimal device characteristics [25], [26].

In this paper, we present the first short-channel self-aligned GFETs (<100 nm) based on full h-BN encapsulation of the graphene sample [9]. Source and drain contacts are formed by making edge contact to the graphene; this is also the first time that this contact technology has been used in a short-channel FET structures. Electrical contact to the edge of graphene yields remarkably low contact resistances, 100 and 200 Ω·μm for electrons and holes, respectively, close to the minimum possible as determined by Landauer–Buttiker considerations. There have been reports of areal contacts achieving comparable or slightly better contact resistivity on epitaxial graphene on SiC substrates [27], but the quality of epitaxial graphene is generally lower than that achieved with graphene grown by other means (such as chemical vapor deposition on copper) or mechanically exfoliated from bulk crystal. This asymmetry in contact resistance is opposite to that observed for conventional top contacts, where the contact resistance for holes is generally lower than that for electrons. Consequently, most GFETs are biased as p-type devices, while the devices presented here are biased as n-type transistors.

II. FABRICATION

To assemble the encapsulated h-BN/graphene/h-BN heterostructures, a van der Waals mechanical transfer technique is employed [9]. First, graphene and h-BN crystals are exfoliated on 285-nm SiO₂ on Si substrates. The h-BN is characterized by atomic force microscopy (AFM) to verify the thickness of the crystals. Next, a thin h-BN is selected to serve as the gate dielectric and then lifted from the SiO₂ surface using a polymer handle. This h-BN crystal is then used to pickup the monolayer graphene, which is then transferred onto a relatively thicker (>15 nm) h-BN crystal that serves as the device substrate. The assembled heterostructure is vacuum annealed at 350 °C for 15 min at <2×10⁻⁷ mbar to remove polymer residue left during the transfer process [Fig. 1(a)]. Characterization by AFM yields a pristine, bubble-free heterostructure.

In most GFET structures, an additional ungated, intrinsic graphene region is present; this is a source of substantial additional contact resistance. In our case, edge contacts are combined using a self-aligned process that effectively eliminates these spacer regions. Starting from the encapsulated graphene heterostructure, transistor fabrication proceeds, as shown in Fig. 1. We use e-beam lithography to define the local top gates using a bilayer resist consisting of 200 nm of a copolymer below 100 nm of poly (methyl methacrylate) (PMMA). The gate metal stack is 2 nm of Cr, 20 nm of Pd, and 100 nm of Au [Fig. 1(b)]. Following the deposition of the gate metal, we anneal the device at 350 °C for 15 min at <2×10⁻⁷ mbar vacuum to remove any PMMA residue (as measured by AFM). Next, we deposit 10 nm of HfO₂ by ALD to electrically isolate the gate [Fig. 1(c)]. Because h-BN is atomically smooth, it is resistant to the nucleation of HfO₂; HfO₂, therefore, grows everywhere except on the surface of h-BN.

This selective growth of HfO₂ is highlighted in Fig. 1(g) and (h), in which an exfoliated 29-nm-thick h-BN crystal on SiO₂ is shown before and after the growth of 15 nm of HfO₂ by ALD. The step height change of 17 nm between the two images indicates that the HfO₂ has been deposited on the SiO₂ and the edge of the h-BN but not on the interior of the crystal.

Following the HfO₂ deposition, we plasmaetch the heterostructure (4-sccm O₂ and 40-sccm CHF₃ plasma under 60-W dc power for 15 s) to expose the edge of the graphene [Fig. 1(d)]. E-beam lithography is then used to define the source–drain contacts. Then, we evaporate 1.5-/10-nm Cr/Au to make electrical contact with the graphene edge [Fig. 1(e)]. Finally, the thick (100 nm) gold leads and probe pads are evaporated to provide low resistance connections to the devices. Fig. 1(f) shows a false color image of the device structure with green representing the local gates and red representing the source–drain contacts.

III. ELECTRICAL CHARACTERIZATION AND MODELING

The I-V characteristics of five encapsulated GFETs with L₉ of 67, 103, 122, 178, and 1240 nm, as determined by scanning electron microscopy (SEM), are measured [Fig. 2(a)]. These devices are all fabricated from the same heterostructure, sharing the same 3.5-nm-thick top h-BN crystal as the gate dielectric (EOT of ~3.5 nm [28] given a dielectric constant for h-BN which closely matches that of silicon dioxide). The oxide capacitance (Cₒₓ), therefore, is estimated to be 10 fF/μm², about three times less than the Cₒₓ for silicon devices at the 65-nm technology. The choice of h-BN...
thickness is based on the limitations of fabrication. Building this heterostructure using <5-nm-thick top h-BN crystals is extremely challenging because the crystal tends to develop physical defects (such as cracks) during the transfer process. For optimum FET performance, a 1–2-nm h-BN crystal would be used, instead of the 3.5 nm used here. The substrate (or bottom) h-BN crystal is also shared among the devices and is 15-nm thick. The contact resistance for these devices is about ∼200 Ω-μm from the metal-graphene edge contact. About ∼200 Ω (measured from a comparable device) is added by the leads connecting to the device. For the devices with $L_G$ below 200 nm, there is little difference in the output characteristics, which suggests ballistic transport. The nonsaturating current characteristics in these ballistic devices are due to the ambipolar nature of the graphene channel [16]. The minimum $g_{ds}$ and maximum $g_m$ values of these short-channel devices are 300 and 600 $μS/μm$ [Fig. 2(b)], respectively, while the long-channel device (1.2 $μm$) has a minimum $g_{ds}$ value of 70 $μS/μm$ and maximum $g_m$ value of 500 $μS/μm$.

To understand these characteristics in more detail, it is useful to fit these to a physically motivated compact model. While gradual channel approximation models have been used quite successfully to model the long-channel GFETs [16], [29], they are a less useful starting point for the devices primarily characterized by ballistic transport [30]. A virtual source (VS) model is valid for both long-channel drift-diffusion transport and short-channel ballistic transport [31]. In this approach, the current is modeled as

$$I_D = \frac{Q_{\text{Sat}}}{W} (V_{\text{DS}}/V_{\text{Sat}}) v_{x0} F_{\text{Sat}}$$

where $Q_{\text{Sat}}$ is the device width, and $Q_{\text{Sat}}, Q_{\text{Sat}}$, and $Q_{\text{Sat}}$ are the areal electron and hole charge densities, respectively, at the virtual drain and source [31], [32]. $v_{x0}$ is the carrier injection velocity and $F_{\text{Sat}}$ is an empirical function used to transition between the linear and saturated transport regimes [31], [33]. $β$ is a fitting parameter which is $2 \pm 0.1$ for all devices. $V_{\text{DS}}$ is the intrinsic drain–source voltage of the GFET, which removes the effects of source and drain contact resistance. $E_c$ and $E_q$ are the local energy potentials at the VS and drain. $C_Q$ is the quantum capacitance and $C_{Q1}$ is a constant $8.426 \text{fF}/\mu\text{m}^2$ [34].

IV. ANALYSIS

The resulting VS model is used to perform a least-squares fit to the experimentally measured data [Fig. 2(a)], determining both the mobility [$μ$, Fig. 2(e)] and the carrier injection velocity [$v_{x0}$, Fig. 2(d)] as a function of $L_G$. From this parameter extraction, we find that the mobility decreases with decreasing $L_G$ due to the onset of ballistic transport. The mobility can be estimated by

$$\frac{1}{μ} = \frac{1}{μ_{\text{eff}}} + \frac{1}{μ_{\text{eff}}} L_G$$

where $μ_{\text{eff}}$ is the mean-free-path mobility and $μ_{\text{eff}}$ is the effective mobility for a device dominated by drift-diffusion transport with $L_G$ much larger than $L_{\text{eff}}$ [31], [35], [36]. This model is derived from the Landauer formulation [37] and applies in the region of linear response, matching other approaches commonly employed to characterize the mean-free-path from measured conductivity [9]. Fig. 3(a) shows $1/μ$ as a function
of $1/L_G$. By employing a linear fit to the data, we determine a $\lambda_{mfp}$ of 848 nm and a $\mu_{eff}$ of 13 734 cm$^2$/Vs. The source-injection velocity is similarly given by

$$\frac{1}{v_0} = \frac{2\alpha}{v_{bal} L_G} + \frac{1}{v_{bal}}$$

where $v_{bal}$ is the ballistic velocity and $\alpha$ is a dimensionless constant of proportionality [26], [30], [31]. Here, we assume that the mean-free-path determined by the length dependence of the mobility in the linear regime is equal to that determined by the velocity dependence in the saturated regime, which is a common assumption in these models [26], [30], [31]. Fig. 3(b) shows $1/v_0$ as a function of $L_G$, which determined a $v_{bal}$ of 9.3 $\times$ 10$^7$ cm/s, very close to the Fermi velocity of graphene, $1 \pm 0.05 \times 10^8$ cm/s [38].

The extracted $\lambda_{mfp}$ indicates that all the devices measured in Fig. 2 are ballistic except for the 1.24-$\mu$m channel-length device. To the best of our knowledge, these GFETs are the first to demonstrate essentially the ballistic transport at channel lengths in excess of 0.5 $\mu$m. The effective mobility is also among the largest reported for a GFET due to the high-quality h-BN/graphene heterostructures. However, these values of mean-free-path and effective mobility are lower than what has been reported for h-BN/graphene heterostructures under low bias [32]. The thinner BN gate dielectrics employed here, as well as the additional processing required to produce the more complex FET structures, are probably resulting in some reduction in encapsulated graphene channel quality. The ballistic velocity achieved in these devices, limited by the Fermi velocity in graphene, is more than five times the band-structure-limited ballistic velocities observed in CMOS devices [39].

Despite the higher carrier velocity and mobility, the $g_m$ and $g_{ds}$ values of these devices are significantly lower than those of silicon devices at comparable channel lengths, due to the effects of quantum capacitance, $C_Q$, and ambipolar channel conduction. The $g_m$ value is proportional to the effective gate capacitance ($C_{eq}$), which for a metal–insulator–graphene (MIG) capacitor [Fig. 4(a)] is the series combination of the electrostatic capacitance ($C_{ox}$) and the quantum capacitance, $C_{eq} = C_{ox} C_Q/(C_{ox} + C_Q)$ [Fig. 4(b)]. Fig. 4(b) shows the ratio $C_{eq}$ to $C_{ox}$ as a function of h-BN thickness ($t_{ox}$) for different fixed-gate biases, indicating that the quantum capacitance prevents $C_{eq}$ from being reduced below $C_Q$. $C_Q$ is on the order of 1 $\mu$F/cm$^2$ and halves $C_{eq}$ from the electrostatic $C_{ox}$ value at a $t_{ox}$ of 1 nm. In contrast, silicon CMOS exhibits a $C_Q$ of $\sim$134 $\mu$F/cm$^2$, 16 times larger than in graphene [40], giving $C_{eq} \sim C_{ox}$ for silicon for any achievable EOT. By using high-$k$ gate dielectrics, others have been able to reduce the EOT below that achieved in this paper. For example, graphene transistors exhibiting a $g_m$ of 1.2 mS/µm have been demonstrated using 4-nm HfO$_2$ [41]. However, silicon devices, with significantly lower carrier velocities and mobilities exhibit $g_m > 2$ mS/µm [42] for channel lengths below 65 nm because the silicon devices are not limited by quantum capacitance.

Another fundamental limitation is the zero bandgap of graphene, which limits the achievable $g_{ds}$. When $V'_{DS} > V'_{GS}-V_{Dirac}$, the drain begins injecting a substantial hole current, giving rise to the second linear regime in the GFET $I$–$V$ characteristic, and increasing $g_{ds}$ [16]. Here, $V'_{DS}$ is the intrinsic source–drain bias, $V'_{GS}$ is the intrinsic gate–source bias, both ignoring the effects of contact resistance, and $V_{Dirac}$ is the Dirac voltage. As a result, the minimum $g_{ds}$ value occurs in a narrow regime of drain biases in which the electron current has saturated but hole densities are not yet substantial. The $g_{ds}$ values for the short-channel GFETs considered here are on the order of 300 $\mu$S/µm, more than three times what is typical for 65-nm silicon devices. In addition, since silicon devices are unipolar, the region of drain biases for which this $g_{ds}$ value is valid is much larger than for GFETs.

**V. Conclusion**

This paper demonstrates the ballistic short-channel GFETs with EOT of less than 3.5 nm, which exhibit among the highest achieved mobilities and carrier velocities. Ballistic behavior is observed with mean-free-paths in excess of 800 nm, even at high biases, opening up intriguing possibilities for devices based on ballistic transport over longer length scales. For example, devices based on electron focusing can be achieved...
at room temperature in graphene [43]. Nevertheless, as an FET channel material, despite carrier velocities that are more than a factor of four greater than those achievable in silicon, the ambipolar conduction and quantum capacitance of graphene are the fundamental limitations of the graphene itself, producing \( g_m \) and \( g_d \) characteristics that are substantially worse than in comparable silicon CMOS devices.

REFERENCES


