ARTICLE

Scaling of High-Field Transport and Localized Heating in Graphene Transistors

Myung-Ho Bae,[†,‡] Sharnali Islam,[†,‡] Vincent E. Dorgan,[†,‡] and Eric Pop[†,‡,§,*]

[†]Micro & Nanotechnology Lab, University of Illinois, Urbana—Champaign, Illinois 61801, United States, [‡]Department of Electrical & Computer Engineering, University of Illinois, Urbana—Champaign, Illinois 61801, United States, and [§]Beckman Institute, University of Illinois, Urbana—Champaign, Illinois 61801, United States

ABSTRACT We use infrared thermal imaging and electrothermal simulations to find that localized Joule heating in graphene field-effect transistors on SiO_2 is primarily governed by device electrostatics. Hot spots become more localized (*i.e.*, sharper) as the underlying oxide thickness is reduced, such that the average and peak device temperatures scale differently, with significant long-term reliability implications. The average temperature is proportional to oxide thickness, but the peak temperature is minimized at an oxide thickness of ∼90 nm due to competing electrostatic and thermal effects. We also find that careful comparison of high-field transport models with thermal imaging can be used to shed light on velocity saturation effects. The results shed light on optimizing heat dissipation and reliability of graphene devices and interconnects.

KEYWORDS: graphene transistor · scaling · high-field transport · saturation velocity · Joule heating · thermal imaging

With its high mobility and high thermal conductivity,[1−4] graphene has garnered much attention as a material for applications such as high-frequency electronics[5] and optoelectronics.[6] Since intrinsic graphene has no band gap, ambipolar transport[7−13] can be readily observed in graphene field-effect transistors (GFETs); that is, both electrons and holes can contribute to conduction along the channel. In addition, no proper carrier depletion region can be achieved in a two-dimensional graphene channel, unlike, for example, in unipolar (n- or p-type) silicon transistors. Instead, during ambipolar conduction the electron and hole populations "meet" at a charge neutral point (CNP) along the GFET channel, under certain bias conditions.[7−13]

Recently, several studies have found that a temperature maximum (hot spot) forms at the position of minimum charge density and maximum electric field along the GFET channel.[8−10] In ambipolar transport the CNP corresponds to the minimum charge density and the thermal hot spot marks the location of the CNP. Combining thermal imaging with electrical measurements and simulations provides valuable information for understanding transport physics in GFETs. However, until now, the hot spot observed in GFETs has been quite broad (>15 μm), making it challenging to fine-tune transport models or to understand the physical reason behind this broadening, *e.g.*, imaging limitations, electrostatics, or simple heat diffusion. In addition, more precise spatial heating information is desirable to understand the long-term reliability of graphene electronics.

RESULTS AND DISCUSSION

In this work we elucidate the high-field hot spot formation in ambipolar GFETs and find that the primary physics behind it is electrostatic in nature. We also examine the role of two simple velocity saturation models[7,14] on high-field transport and dissipation in GFETs, and describe in comprehensive detail our self-consistent electrothermal simulation approach.

Through infrared (IR) thermal imaging of functioning GFETs we show that more spatially confined (sharper) hot spots are formed in devices on thinner (∼100 nm) SiO_2 layers *versus* previous work[8−10] on 300 nm oxides. The measured device current and temperature profiles are in excellent agreement with our simulations, which include electrostatic, thermal, and velocity saturation effects. Once this model is calibrated, we then investigate the hot spot scaling with the SiO_2 substrate thickness over a wide range of practical values. Interestingly, we find that during ambipolar operation the *average* channel temperature scales with oxide thickness as expected, but the *peak* temperature is minimized at an oxide thickness of ∼90 nm, due to competing electrostatic and thermal effects. The results provide novel insight into high-field transport and dissipation in

* Address correspondence to epop@illinois.edu.

Received for review June 17, 2011 and accepted August 26, 2011.

Published online September 13, 2011
10.1021/nn202239y

© 2011 American Chemical Society

graphene devices and suggest that sharply peaked temperatures can have an impact on long-term device reliability[15,16] and must be carefully considered in future device designs.

Electrical Characterization and Electrostatics. The device geometry is shown in Figure 1a, with device fabrication and infrared thermal imaging being described in the Methods section. Figure 1b displays measured graphene resistance (symbols) *versus* back-gate voltage ($V_G \approx V_{GD} \approx V_{GS}$) at small V_{SD} = 20 mV. The peak resistance is at $V_{GD} = V_0 = 5.2$ V, also known as the Dirac voltage. V_0 corresponds to the Fermi level in the graphene sheet crossing the average Dirac point of the X-shaped electronic band structure[8,17] and to zero net charge density in the graphene channel ($n - p = 0$). Nevertheless, we note that zero *net* charge density does not imply a lack of free carriers, as there are equal numbers of electron and hole "puddles" contributing to the nonzero conductivity at the Dirac point ($n = p \neq 0$). This puddle density is caused by charged inhomogeneity due to impurities[18] in the SiO_2 or on the graphene and to thermally excited carriers[19] that form a nonhomogeneous charge and potential landscape[14,17] across the graphene device at the Dirac voltage. At higher (lower) gate voltages with respect to V_0, the majority carriers become electrons (holes), respectively,[14] and the charge inhomogeneity is smoothed out.

On the basis of an analytic electrostatic model that rigorously takes into account the above phenomena,[14] we fit the resistance data as shown by the dashed curve in Figure 1b with a low-field mobility μ_0 = 3700 $cm^2\ V^{-1}\ s^{-1}$ and a puddle density $n_{pd} = 3.5 \times 10^{11}$ cm^{-2}. This fitting also considers the varying contact resistance as a function of gate voltage,[15,20,21] including the role of the finite transfer length, L_T, the distance over which $1/e$ of the current transfers between the graphene and the overlapping metal electrode. The contact resistance is defined by[15]

$$R_C = \frac{1}{W}\frac{\rho_C}{L_T}\coth\left(\frac{L_C}{L_T}\right) \quad (1)$$

where ρ_C is the metal−graphene contact resistivity per unit area, L_C is the length of the metal electrode that overlaps with the graphene, $L_T = (\rho_C/R_S)^{1/2}$ is the current transfer length,

$$R_S = [q\mu_0(n+p)]^{-1} \quad (2)$$

is the graphene sheet resistance, and q is the elementary charge. The electron and hole density per unit area (n and p) are defined by the gate voltage, temperature, and puddle density as given by[14]

$$n, p \approx \frac{1}{2}[\pm n_{cv} + \sqrt{n_{cv}^2 + 4n_0^2}] \quad (3)$$

where the lower (upper) sign corresponds to electrons (holes), $n_{cv} = C_{ox}(V_0 - V_G)/q$, $C_{ox} = \varepsilon_{ox}/t_{ox}$ is the capacitance per unit area (quantum capacitance can be neglected here), ε_{ox} is the dielectric constant of SiO_2, $n_0 = [(n_{pd}/2)^2 + n_{th}^2]^{1/2}$, and $n_{th} = (\pi/6)(k_BT/\hbar v_F)^2$ is the thermal carrier density, with Fermi velocity $v_F \approx 10^8$ cm/s (the complete derivations are given in ref 14).

The solid curve in Figure 1b displays the R_C, which changes with gate voltage, with the single fitting parameter ρ_C = 500 Ω μm^2 for the graphene−metal interface, about a factor of 3 larger than in ref 15. The contact resistance per device width approaches $R_CW \approx$ 1000 Ω μm at large gate voltage, with a transfer length of the order $L_T \approx 0.5$ μm. The total device resistance R (symbols and dashed lines) in Figure 1b includes

$$R = \left(\frac{L}{W}\right)R_S + 2R_C + R_{series} \quad (4)$$

where R_{series} = 600 Ω is the total series resistance of the Pd metal wires contacting our device (Pd resistivity independently measured, $\rho_{Pd} \approx 14$ $\mu\Omega$ cm). For simplicity, in this study we assume a constant mobility that is equal for electrons and holes, although there are indications that the mobility decreases at higher charge densities, as noted by our previous work.[14] However, this does not alter our conclusions and the excellent agreement between experiment and simulation below, since all "hot spot" phenomena take place at relatively low charge density.

Figure 1c displays current *versus* drain−source voltage ($I_D - V_{SD}$) measurements up to relatively high field (symbols) and our simulations (lines) at various back-gate voltages, V_{GD}. We note that the transport is diffusive both at high field and at low field in our devices. At high field, velocity saturation[14] occurs at fields $F > 1$ V/μm, which corresponds to scattering rates[22,23] $1/\tau \approx$ 50 ps^{-1} and a mean free path $l_{HF} \approx v_F/\tau \approx 20$ nm. Taking $v_{sat} \approx 3 \times 10^7$ cm/s at $F \approx 3$ V/μm (refs 14, 22), the high-field mobility is on the order of $v_{sat}/F \approx$ 1000 $cm^2\ V^{-1}\ s^{-1}$. As the low-field mobility is only about a factor of 4 higher in our samples, the low-field mean free path is on the order of $l_{LF} \approx 80$ nm, in accordance with previous estimates made in ref 8. Thus, both the low-field and high-field mean free path of electrons and holes in our samples are significantly smaller than the device dimensions (several micrometers), and diffusive transport is predominant in these samples.

At high lateral field and under diffusive transport conditions, the electrostatic potential varies significantly along the channel.[8] The electrostatic potential at the drain is set by V_{GD} (Figure 1c), while that at the source is

$$V_{GS} = V_{GD} + V_{DS} = V_{GD} - V_{SD} \quad (5)$$

For instance, with V_{SD} decreasing from zero, at V_{GD} = −2 V and $V_{SD} \approx -7.2$ V, V_{GS} is near V_0 = 5.2 V and the Dirac point (CNP) is in the channel exactly at the edge of the source. This is seen as a change in curvature of

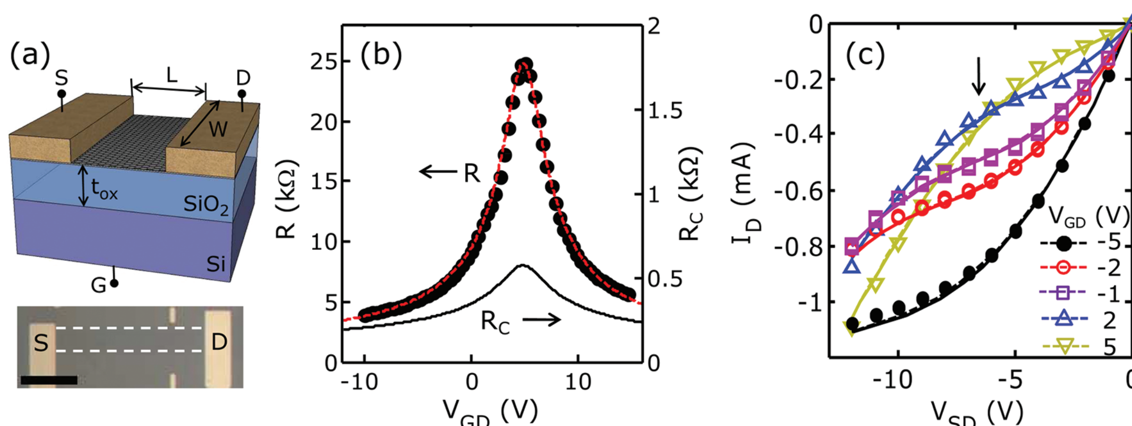

Figure 1. (a) Schematic of GFET (top) and optical image of fabricated device (bottom). Device dimensions are L = 28.8 μm, W = 5 μm, t_{ox} = 100 nm; scale bar is 10 μm. (b) Resistance *vs* back-gate voltage, experimental data (points) and model fit (lines). The fitted contact resistance R_C is also shown, a function of gate voltage (see model and text). (c) Drain current *vs* source−drain voltage at various back-gate voltages; measured data (points) and simulations (lines). The two nearly overlapping families of lines (solid and dashed) are simulations with the two velocity saturation models (see text).

the ambipolar "S"-shaped I_D-V_{SD} plots, marked by an arrow in Figure 1c. The channel resistance now decreases as the source−drain voltage drops below $V_{SD} <$ −7.2 V because the electron density at the source increases. The other, primarily unipolar, operating regimes have been described in detail in ref 8.

Thermal Characterization in Ambipolar Conduction. We now consider the power dissipation through the Joule self-heating effect[24] along the graphene channel and focus specifically on the ambipolar conduction mode described above. As the chemical potential changes drastically, neither the electric field nor the carrier density is uniform along the channel under high-field conditions. But, because carrier movement along the GFET is unidirectional (from source to drain), the current density J must be continuous, where

$$J = \frac{I_D}{W} = q(n+p)v_d \quad (6)$$

is proportional to the local carrier density ($n + p$) and the drift velocity (v_d) at every point along the channel. Thus, regions of high carrier density have low drift velocity, and *vice versa*. The highest field ($F \propto v_d$, see eq 8) and highest localized power dissipation ($p \approx JF$) will be at the region corresponding to the minimum carrier density,[8] which is where one expects the hot spot to be localized. In particular, in the ambipolar conduction state the minimum carrier density spot matches the CNP, which is now located within the GFET channel.

To examine this point, we measured the temperature along the graphene channel with fixed $V_{SD} = -12$ V and at various gate−drain voltages, V_{GD}, as shown in Figure 2a. At $V_{GD} = -5$ V ($\ll V_0$) the drain is heavily hole-doped, but $V_{GS} = +7$ V, so the region near the source is lightly electron-doped (keeping in mind that $V_0 = 5.2$ V for this device). Thus, the CNP is located very close to the source and so is the hot spot, as can be seen in the upper panel of Figure 2a. As we increase V_{GD} as marked in the figure, V_{GS} continues to increase according to eq 5, reaching $V_{GS} = +16$ V ($\gg V_0$) in the bottom panel of Figure 2a. At this point, the source is heavily electron-doped and the drain is lightly hole-doped, very close to the CNP ($V_{GD} = 4$ V $< V_0 = 5.2$ V). Thus, during the entire imaging sequence shown in Figure 2a the GFET is operating in the ambipolar transport regime, but changing the gate voltage gradually alters the relative electron and hole concentrations, moving the hot spot (location of CNP) from near the source to near the drain. This experimental trace of the CNP also provides an excellent tool for checking the validity of electronic *and* thermal transport models under such inhomogeneous carrier density along the channel.

To complement the thermal imaging along the GFET (x-direction), Figure 2b and c show a top view of the hot spot at $V_{GD} = -2$ V and a thermal cross-section of the GFET along the dashed line (y-direction) as indicated. We note that the width of the GFET here is only slightly larger than the IR resolution (see Methods), and thus the cross-section view should be used only for qualitative inspection. By comparison, higher resolution scanning Joule expansion microscopy (SJEM)[15] has revealed a uniform transverse temperature profile with slightly cooler edges from heat sinking and higher carrier density due to fringing heat and electric field effects.

High-Field Electrothermal Model. Our graphene device simulation approach was partially described in previous publications,[8,14,15] and here we briefly review a few more salient features. The model is qualitatively similar to other approaches;[11−13] however it is the only one (to our knowledge) to self-consistently include the thermal effects during high-field transport. The current continuity equation is given by eq 6 and must be satisfied at every point along the GFET channel. The

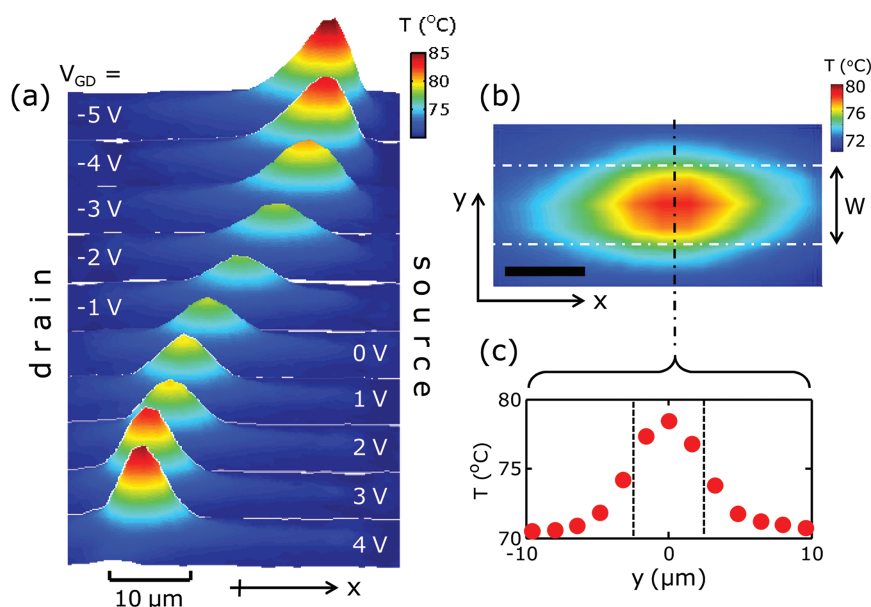

Figure 2. (a) Infrared mapping of temperature profiles along the GFET on t_{ox} = 100 nm, showing markedly "sharper" hot spot formation compared to previous work on t_{ox} = 300 nm (refs 8, 9). The bias conditions are $V_{SD} = -12$ V (last data point in Figure 1c) and changing gate voltage as labeled. The hot spot moves from source to drain, marking the location of minimum charge density and maximum electric field, following the device electrostatics (see text). (b) Top view of the hot spot at $V_{GD} = -2$ V, showing symmetric temperature distribution in the transverse (*y*-direction) as expected. Scale bar is 5 μm. (c) Temperature profile along the cross-section in (b); dashed lines mark the width (*W*) of the device.

charge density at high fields along the channel is computed from the discussion above, however using

$$n_{cv} = \frac{C_{ox}}{q}[V_0 - V_{GD} + V(x)] \quad (7)$$

where the local potential $V(x)$ along the GFET channel is obtained from the Poisson equation.[8] The field along the GFET is then $F(x) = -dV(x)/dx$, which is used to compute the local power dissipation as stated above. The low-field and high-field transport regimes are bridged through the dependence of drift velocity on electric field,[14]

$$v_d(F) = \frac{\mu_0 F}{[1 + (\mu_0 F/v_{sat})^{\gamma}]^{1/\gamma}} \quad (8)$$

where $1 \leq \gamma \leq 2$ is a fitting parameter. For completeness, we also include the effects of heating due to current crowding and cooling or heating (depending on direction of current flow) due to Peltier effects at the graphene–metal contacts.[15] Finally, we obtain the temperature distribution along the GFET through the heat equation[8]

$$A\frac{\partial}{\partial x}\left(k\frac{\partial T}{\partial x}\right) + p' - g(T - T_0) = 0 \quad (9)$$

where $p' = I_D F(x)$ is the Joule heating rate per unit length, $A = Wt_g$ is the graphene cross-section area (monolayer "thickness" t_g = 0.34 nm), and $g \approx 1/[L(\mathscr{R}_B + \mathscr{R}_{ox} + \mathscr{R}_{Si})]$ is the thermal conductance to substrate per unit length, where the 3 terms are the graphene–SiO_2 boundary resistance, the oxide resistance, and the silicon substrate thermal resistance, respectively, as in ref 14. The graphene thermal conductivity k = 1000 Wm^{-1} K^{-1} here, higher than in ref 4 to account for some lateral heat flow along the polymethyl methacrylate (PMMA) top layer (see Methods); however the results are not sensitive to k, as most heat is dissipated into the underlying SiO_2, as in ref 8. We note that the approach adopted here automatically accounts for heat dissipation into the contacts,[15] but that this is a negligible fraction of the total input power, which is predominantly dispersed into the underlying SiO_2 in such large devices.[14] By contrast, another recent study[25] has shown that in short, sub-0.3 μm GFETs a substantial portion of the heat is dissipated to the metallic contacts.

To obtain the current as a function of voltage, the above equations are solved iteratively and self-consistently, until changes in carrier density converge to less than 1% and the temperature converges to within less than 0.01 K between iterations. Figure 1c shows that the simulation results (lines) are in excellent agreement with the experimentally measured I_D–V_{SD} data. All data were stable and reproducible during measurements, partly enabled by protection offered by the top PMMA layer (see Methods) and partly from limiting the maximum voltages applied.[26]

To better understand high-field transport, we considered two recent models for the drift velocity saturation (v_{sat}), as shown in Figure 3. In one case, Meric

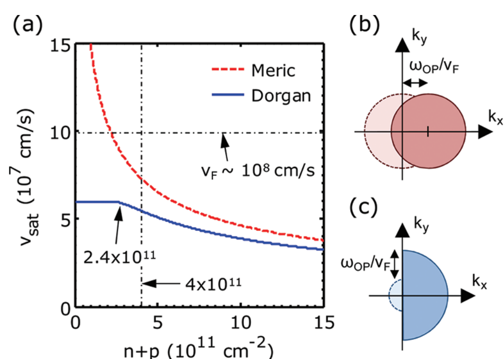

Figure 3. (a) High-field saturation velocity models *vs* carrier density.[7,14,28] At low density, here $<2.4 \times 10^{11}$ cm^{-2}, the Dorgan[14] model reaches a constant value ($\sim 2v_F/\pi \approx 6.3 \times 10^7$ cm/s, slightly lower here at $\sim$70 °C, see eq 12), whereas the Meric[7] model can diverge. However, due to temperature effects and puddle charge, the carrier density in our device is always $>4 \times 10^{11}$ cm^{-2} during operation, as marked by an arrow. Thus, in the device simulated here either model can be applied, as in Figures 1 and 4. (b, c) Schematic assumptions of carrier distribution at high field used to derive the closed-form v_{sat} expressions in the (b) Meric[7] and (c) Dorgan[14] models.

et al.[7,27] have suggested

$$v_{sat} = \frac{\omega_{OP}}{\sqrt{\pi(n+p)}} \quad (10)$$

where $\hbar\omega_{OP}$ is the dominant optical phonon (OP) energy for carrier energy relaxation. This is an approximation based on a shifted Fermi disk in the limit of $T = 0$ K (see Figure 3b and supplement of ref 7) and is generally applicable at "large" carrier density ($n + p \gg n_0$). On the other hand, following initial work by Barreiro and co-workers,[28] Dorgan *et al.*[14] have proposed

$$v_{sat} = \frac{2}{\pi}\frac{\omega_{OP}}{\sqrt{\pi(n+p)}}\sqrt{1 - \frac{{\omega_{OP}}^2}{4\pi(n+p){v_F}^2}}\,\frac{1}{N_{OP}+1},\ n+p \geq n^* \quad (11)$$

$$v_{sat} = \frac{2}{\pi}\frac{v_F}{N_{OP}+1},\ n+p < n^* \quad (12)$$

where $n^* = (\omega_{OP}/v_F)^2/2\pi$, $N_{OP} = 1/[\exp(\hbar\omega_{OP}/k_BT) - 1]$ is the phonon occupation, and k_B is the Boltzmann constant. These expressions are based on a steady-state population in which carriers contributing to current flow occupy states up to an energy $\hbar\omega_{OP}$ higher than carriers moving against the net current[28] (Figure 3c). Note that both models suggest v_{sat} decreases approximately as the inverse square root of the carrier density, and in both models $\hbar\omega_{OP}$ is treated as a fitting parameter. However, v_{sat} in the Meric model is derived in the limit $T = 0$ K and can approach infinity as the carrier density tends to zero. The Dorgan model includes a semiempirical temperature dependence[14] and approaches a constant at low carrier density, $v_{max} \approx (2/\pi)v_F \approx 6.3 \times 10^7$ cm/s (closer to $\sim 6 \times 10^7$ cm/s at 70 °C when the temperature dependence is taken into account, as in eq 12 and Figure 3a).

Consistent with the previous studies[7,14,27] we choose $\hbar\omega_{OP}$ = 59 meV (γ = 1.3 in eq 8) and 81 meV (γ = 1.5) for the Meric and Dorgan models, respectively. These are consistent with the SiO_2 surface phonon energy and with a combination between the SiO_2 phonon and graphene optical phonon energy, respectively. The phonon energy fitting parameters were chosen so as to yield virtually indistinguishable characteristics in Figure 1c. We plot v_{sat} from the two models as a function of total carrier density ($n + p$) in Figure 3a, showing the expected behavior as described above. With our present parameters, the Dorgan model reaches a constant below charge densities $n + p < n^* = 2.4 \times 10^{11}$ cm^{-2}. However, we note that the minimum charge density achieved during all simulations in this work was $\sim 4 \times 10^{11}$ cm^{-2} due to puddle charge and thermally excited carriers. In addition, the maximum longitudinal fields[26] were $\sim$0.9 V/μm (see Figure 4), and thus complete velocity saturation was never fully reached (see, *e.g.*, Figure 3 of ref 14). This explains that relatively good agreement can be attained between either model and our data in Figure 1c, within the present conditions. (Future work on shorter devices at higher electric fields[26] will be needed to elucidate the role of saturation velocity at low carrier density.)

Comparison of Simulation with Data. With the parameters discussed above, Figure 4 shows carrier densities and temperature profiles at the last drain bias point (V_{SD} = −12 V) for three representative gate voltages, V_{GD} = −2, −1, and 2 V. Once again, excellent agreement is found between simulation results obtained with the two different v_{sat} models (solid curves) and the experimental temperature profiles (symbols).[29] The position of the CNP for each V_{GD} can be visualized by comparing Figure 4a−c with Figure 4d−f as the crossing point of electron and hole carrier density profiles and that of the hot spot. We also plot the corresponding electric fields in Figure 4g−i, where the position of the maximum field matches that of the hot spot. The CNP clearly moves from source to drain when the gate voltage changes, as visualized in Figure 2 and previously explained in qualitative terms. We note that the profile of the hot spot with 100 nm underlying oxide thickness (Figures 2 and 4 here) is much better defined and "sharper" than what was previously observed on 300 nm oxide.[8,9]

Comparing the simulations obtained with the two v_{sat} models, we note that the carrier density profiles are nearly identical in Figure 4a−c. However, the lower v_{sat} (at a given carrier density) of the Dorgan model[14] yields slightly higher electric fields and higher hot spot temperatures, as shown in Figure 4d−i (also see the insets). The temperature difference here is up to $\sim$1 °C between the two models, or $\sim$5% of the total temperature change, although the applied power is the same between the separate simulations. We note that since velocity saturation is never fully reached in the

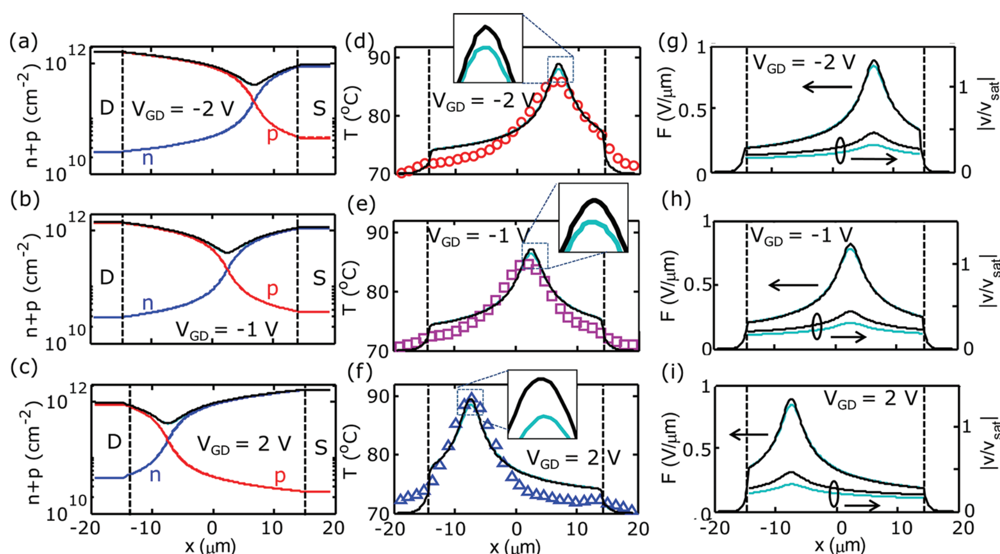

Figure 4. Simulation of carrier density, temperature, and electric field along the GFET at various gate voltages from Figures 1 and 2, with $V_{SD} = -12$ V and same total power dissipation. (a–c) Electron, hole, and total carrier density. (d–f) Simulated (lines) and measured (symbols) temperature profiles. The insets show the difference between the two saturation velocity models (Figure 3), with the Dorgan[14] model providing slightly higher temperatures due to lower saturation velocity. (g–i) Corresponding electric field and $|v/v_{sat}|$ profiles along the channel under the same bias conditions. Comparing the simulations shows the thermal hot spot corresponds to the location of lowest carrier density and highest electric field, *i.e.*, its electrostatic nature.

present simulation (and measurement) conditions, the differences in computed temperature and electric field are more subtle than the apparent difference between the two v_{sat} models in Figure 3 would imply. Nevertheless, the disparities are more apparent if we inspect how "close" to saturation the transport becomes, *i.e.*, the ratio $|v/v_{sat}|$ at each point along the channel, as plotted in Figure 4g–i. In this case, the Dorgan model (upper black curves) yields transport closer to the saturation condition, given that its v_{sat} is typically lower. Following eq 8, this also implies higher local electric fields, thus higher local power dissipation and temperature.

The simulation results in Figure 4 suggest that while the IR microscopy used here provides significant insight into high-field transport in graphene, it is not quite sufficient to distinguish with certainty the drift velocity saturation behavior. Nevertheless, we believe the principle of the approach is sound. In other words, thermal measurements of high-field transport in GFETs at conditions of higher fields (>1 V/μm) and lower carrier densities (<5.5×10^{11} cm^{-2}) through a tool such as Raman spectroscopy[9,30] should resolve with more accuracy the drift saturation behavior, providing significantly more insight than electrical measurements alone.

Scaling of Heating with Oxide Thickness. Having established good agreement between our experimental data, numerical simulations, and qualitative understanding, we now seek to extend our knowledge of ambipolar transport in graphene and test the physical mechanisms defining the hot spot. Thus, we simulate device behavior and temperature profiles with various underlying SiO_2 thickness (t_{ox}) during ambipolar transport as shown in Figure 5. Here, all calculations are performed with total power $P = 9.25$ mW, corresponding to the experimentally applied bias conditions at $V_{GD} = -1$ V with $t_{ox} = 100$ nm (Figure 4e). This is an important consideration for an appropriate comparison, since thinner (thicker) oxides are expected to lead to lower (higher) average channel temperature. Moreover, to compare the hot spot between the various cases, we aligned the positions of the CNP for all t_{ox} values by changing V_{GD} and I_D while keeping the total power constant, as shown in Figure 5a.

We also plot the electric field (F) profiles in Figure 5b. Then, based on Figure 5a, we plot the relationship between hot spot width and t_{ox} in Figure 5c (circles), showing a linear scaling between the two. Here, the size of the hot spot is defined as the full width at half the temperature between the peak and the "shoulder" near the contacts. We also plot the width of the electric field profile (solid curve) vs t_{ox}, showing essentially the same scaling as the hot spot. The experimentally measured widths of the hot spots are shown in Figure 5c as triangles for $t_{ox} = 100$ nm from Figure 4e and for $t_{ox} = 300$ nm from ref 8, respectively. While the scaling is similar to that predicted by our simulations, the slight discrepancy is most likely due to finite resolution of the IR microscope. By comparison, averaging the simulation results with a ~2 μm wide broadening function yields the solid circle in Figure 5c, which is closer to the experimental data for $t_{ox} = 100$ nm. For the $t_{ox} = 300$ nm case, the solid square is from a simulation in ref 8, also showing improved agreement when the particular parameters of this device are used.

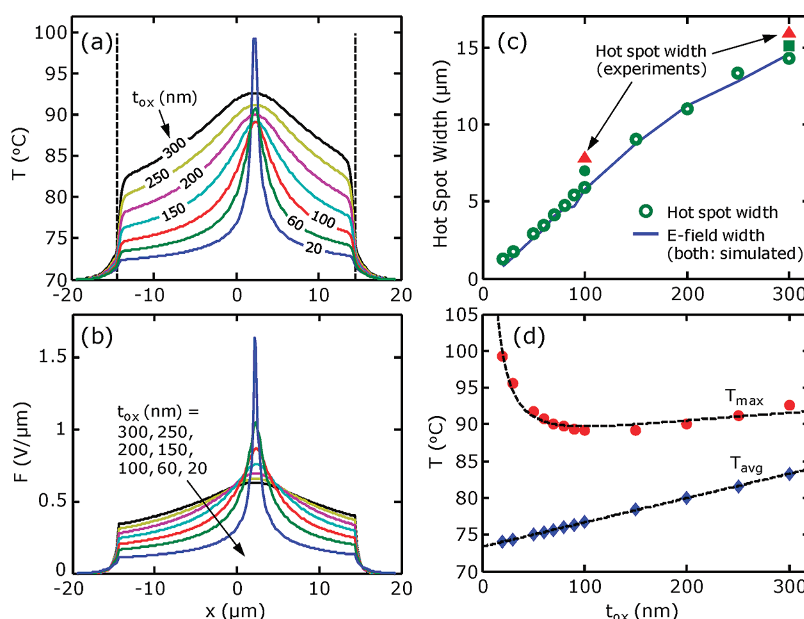

Figure 5. Scaling of GFET hot spot and electric field as a function of underlying SiO_2 thickness. (a) Calculated temperature profiles along device with power input 9.25 mW, corresponding to Figure 4e. (b) Calculated electric field profiles under the same conditions. (c) Scaling of hot spot width (symbols) and electric field width (lines) with t_{ox}. Triangles are experimental data for GFETs on t_{ox} = 100 nm (this work) and 300 nm (ref 8). Circles are calculated widths of the hot spot (see text). (d) Scaling of maximum (T_{max}) and average GFET temperature (T_{avg}) with t_{ox} from (a). Dashed lines are analytic fits. The average temperature T_{avg} does not approach T_0 (=70 °C here) in the limit $t_{ox} \rightarrow 0$ due to the combined effect of the graphene–SiO_2 and silicon substrate thermal resistance ($\mathcal{R}_B + \mathcal{R}_{Si}$), which are independent of t_{ox} (see text after eq 9 and ref 14).

As the oxide thickness is scaled down from $t_{ox} = 300$ to 20 nm, we find that both the *average* channel temperature (Figure 5d) and the width of the hot spot decrease (Figure 5c); that is, the hot spot becomes "sharper". The former occurs because the thermal resistance of the SiO_2 is lowered, and the latter is due to increasing capacitive coupling between the backgate and the charge carriers in the channel. We note that the average channel temperature in Figure 5d does not reach the base temperature (here, $T_0 = 70$ °C) even in the limit of vanishing t_{ox} due to the remaining thermal resistance of the silicon substrate and of the graphene–SiO_2 boundary. To understand this, we note that the average thermal resistance of the device can be estimated as[14] $\mathcal{R}_{th} \approx \mathcal{R}_{ox} + \mathcal{R}_B + \mathcal{R}_{Si}$, where $\mathcal{R}_{ox} \approx t_{ox}/(k_{ox}LW)$ is the thermal resistance of the SiO_2, which scales with t_{ox}, but the second and third terms are the graphene–SiO_2 boundary thermal resistance[14,31] and the spreading thermal resistance[14,24] of the silicon substrate, which are independent of the oxide thickness.

Interestingly, Figure 5d indicates that the *peak* temperature of the hot spot (T_{max}) begins to *increase* when t_{ox} is scaled below ~90 nm, despite a lower average temperature in the channel. This trend occurs because the Joule heating effect induced by the high electric field at the CNP overcomes the cooling effect of the lowered oxide thickness at $t_{ox} \approx 90$ nm. To gain more insight into this observation, we return to the temperature and electric field profiles along the graphene channel in Figure 5a and 5b. We note that the temperature qualitatively follows the electric field profile, and the source of the hot spot is clearly electrostatic in nature. In addition, this finding suggests that one should consider the formation of highly localized hot spots in future devices that would have thinner underlying oxide layers. While a thinner t_{ox} does lead to a lower average temperature, the *peak* temperature is actually increased due to electrostatic effects. This effect is expected to be the same in top-gated as in bottom-gated graphene devices, because the electrostatic effects are controlled by the gate, whereas heat flow is limited by the underlying oxide. The local temperature increase and highly localized electric field at the hot spot could lead to long-term oxide reliability issues,[16] which must be accounted for.

CONCLUSION

In summary, we have examined the physical mechanisms behind high-field hot spot formation in graphene transistors on SiO_2 and found them to be electrostatic in nature. Using self-consistent electrothermal simulations and infrared thermal imaging, we established that the maximum temperature of a graphene device in high-field operation is sensitive to the peak electric field and carrier saturation velocity. We have also confirmed that the average temperature of a functioning GFET

scales proportionally with the thickness of the supporting SiO_2, as expected. However, the maximum temperature of such GFETs can be minimized for a given insulator thickness (here ~90 nm for SiO_2) due to competing electrostatic and heat sinking effects. These results suggest a route for the optimization of graphene substrates for proper heat dissipation and highlight existing trade-offs for practical device reliability.

METHODS

We prepared exfoliated monolayer graphene devices on t_{ox} = 100 nm thermally grown SiO_2 on highly doped Si substrates, which also serve as the back-gate. The graphene layer number was confirmed by optical microscopy and Raman spectroscopy.[32,33] Source and drain contacts are patterned by electron-beam (e-beam) lithography and deposited with Cr (0.5 nm)/Pd (40 nm) by thermal evaporation at 5×10^{-7} Torr base pressure. After lift-off, a rectangular graphene shape (length L = 28.8 μm, width W = 5 μm) is defined by e-beam lithography and oxygen plasma etching, as shown in Figure 1. Finally, a ~70 nm PMMA layer is spun over the substrate, to protect the graphene from spurious doping or ambient moisture during the measurements.[8] Thermal imaging is performed using a QFI InfraScope II IR microscope with 15× objective, spatial resolution of 2.8 μm, pixel size of 1.6 μm, and temperature resolution of ~0.1 °C after calibration.[34] All thermal IR measurements were performed in air at a base temperature T_0 = 70 °C, as needed for optimal IR detector sensitivity.[8,34]

Acknowledgment. We thank Z.-Y. Ong for several useful discussions and F. Lian for providing the device schematic in Figure 1. This work was supported in part by the Office of Naval Research (ONR), the National Science Foundation (NSF) CAREER award, and the Air Force Young Investigator Program (YIP). V.E.D. acknowledges support through a NSF Graduate Research Fellowship.

REFERENCES AND NOTES

1. Geim, A. K.; Kim, P. Carbon Wonderland. *Sci. Am.* **2008**, *298*, 90–97.
2. Bolotin, K. I.; Sikes, K. J.; Hone, J.; Stormer, H. L.; Kim, P. Temperature-Dependent Transport in Suspended Graphene. *Phys. Rev. Lett.* **2008**, *101*, 096802.
3. Balandin, A. A. Thermal Properties of Graphene and Nanostructured Carbon Materials. *Nat. Mater.* **2011**, *10*, 569–581.
4. Seol, J. H.; Jo, I.; Moore, A. L.; Lindsay, L.; Aitken, Z. H.; Pettes, M. T.; Li, X.; Yao, Z.; Huang, R.; Broido, D.; Mingo, N.; Ruoff, R. S.; Shi, L. Two-Dimensional Phonon Transport in Supported Graphene. *Science* **2010**, *328*, 213–216.
5. Lin, Y. M.; Dimitrakopoulos, C.; Jenkins, K. A.; Farmer, D. B.; Chiu, H. Y.; Grill, A.; Avouris, P. 100-GHz Transistors from Wafer-Scale Epitaxial Graphene. *Science* **2010**, *327*, 662.
6. Mueller, T.; Xia, F.; Avouris, P. Graphene Photodetectors for High-Speed Optical Communications. *Nat. Photon* **2010**, *4*, 297–301.
7. Meric, I.; Han, M. Y.; Young, A. F.; Ozyilmaz, B.; Kim, P.; Shepard, K. L. Current Saturation in Zero-Bandgap, Top-Gated Graphene Field-Effect Transistors. *Nat. Nanotechnol.* **2008**, *3*, 654–659.
8. Bae, M.-H.; Ong, Z.-Y.; Estrada, D.; Pop, E. Imaging, Simulation, and Electrostatic Control of Power Dissipation in Graphene Devices. *Nano Lett.* **2010**, *10*, 4787–4793.
9. Freitag, M.; Chiu, H.-Y.; Steiner, M.; Perebeinos, V.; Avouris, P. Thermal Infrared Emission from Biased Graphene. *Nat. Nanotechnol.* **2010**, *5*, 497–501.
10. Jo, I.; Hsu, I. K.; Lee, Y. J.; Sadeghi, M. M.; Kim, S.; Cronin, S.; Tutuc, E.; Banerjee, S. K.; Yao, Z.; Shi, L. Low-Frequency Acoustic Phonon Temperature Distribution in Electrically Biased Graphene. *Nano Lett.* **2011**, *11*, 85–90.
11. Thiele, S. A.; Schaefer, J. A.; Schwierz, F. Modeling of Graphene Metal-Oxide-Semiconductor Field-Effect Transistors with Gapless Large-Area Graphene Channels. *J. Appl. Phys.* **2010**, *107*, 094505–8.
12. Champlain, J. G. A First Principles Theoretical Examination of Graphene-Based Field Effect Transistors. *J. Appl. Phys.* **2011**, *109*, 084515–19.
13. Wang, H.; Hsu, A.; Kong, J.; Antoniadis, D. A.; Palacios, T. Compact Virtual-Source Current-Voltage Model for Top-and Back-Gated Graphene Field-Effect Transistors. *IEEE Trans. Electron Devices* **2011**, *58*, 1523–1533.
14. Dorgan, V. E.; Bae, M.-H.; Pop, E. Mobility and Saturation Velocity in Graphene on SiO_2. *Appl. Phys. Lett.* **2010**, *97*, 082112.
15. Grosse, K. L.; Bae, M.-H.; Lian, F.; Pop, E.; King, W. P. Nanoscale Joule Heating, Peltier Cooling and Current Crowding at Graphene-Metal Contacts. *Nat. Nanotechnol.* **2011**, *6*, 287.
16. Schroder, D. K.; Babcock, J. A. Negative Bias Temperature Instability: Road to Cross in Deep Submicron Silicon Semiconductor Manufacturing. *J. Appl. Phys.* **2003**, *94*, 1–18.
17. Zhu, W. J.; Perebeinos, V.; Freitag, M.; Avouris, P. Carrier Scattering, Mobilities, and Electrostatic Potential in Monolayer, Bilayer, and Trilayer Graphene. *Phys. Rev. B* **2009**, *80*, 235402.
18. Martin, J.; Akerman, N.; Ulbricht, G.; Lohmann, T.; Smet, J. H.; von Klitzing, K.; Yacoby, A. Observation of Electron-Hole Puddles in Graphene Using a Scanning Single-Electron Transistor. *Nat. Phys.* **2008**, *4*, 144–148.
19. Fang, T.; Konar, A.; Xing, H.; Jena, D. Carrier Statistics and Quantum Capacitance of Graphene Sheets and Ribbons. *Appl. Phys. Lett.* **2007**, *91*, 092109–3.
20. Nagashio, K.; Nishimura, T.; Kita, K.; Toriumi, A. Contact Resistivity and Current Flow Path at Metal/Graphene Contact. *Appl. Phys. Lett.* **2010**, *97*, 143514.
21. Xia, F.; Perebeinos, V.; Lin, Y.-M.; Wu, Y.; Avouris, P. The Origins and Limits of Metal-Graphene Junction Resistance. *Nat. Nanotechnol.* **2011**, *6*, 179–184.
22. Fang, T.; Konar, A.; Xing, H.; Jena, D. High Field Transport in 2D Graphene. (cond/mat pre-print) **2010**, arXiv:1008.1161v1.
23. Li, X.; Barry, E. A.; Zavada, J. M.; Nardelli, M. B.; Kim, K. W. Surface Polar Phonon Dominated Electron Transport in Graphene. *Appl. Phys. Lett.* **2010**, *97*, 232105.
24. Pop, E. Energy Dissipation and Transport in Nanoscale Devices. *Nano Res.* **2010**, *3*, 147–169.
25. Liao, A. D.; Wu, J. Z.; Wang, X.; Tahy, K.; Jena, D.; Dai, H.; Pop, E. Thermally Limited Current Carrying Ability of Graphene Nanoribbons. *Phys. Rev. Lett.* **2011**, *106*, 256801.
26. In this work, as in ref 8, we limited the maximum lateral fields to $F < \sim 0.9$ V/μm and the maximum temperature rise to $\Delta T < \sim 100$ K ($T_{max} \sim 170$ °C) in order to obtain reproducible device behavior between subsequent IR and electrical measurements.
27. Meric, I.; Dean, C. R.; Young, A. F.; Baklitskaya, N.; Tremblay, N. J.; Nuckolls, C.; Kim, P.; Shepard, K. L. Channel Length Scaling in Graphene Field-Effect Transistors Studied with Pulsed Current−Voltage Measurements. *Nano Lett.* **2011**, *11*, 1093–1097.
28. Barreiro, A.; Lazzeri, M.; Moser, J.; Mauri, F.; Bachtold, A. Transport Properties of Graphene in the High-Current Limit. *Phys. Rev. Lett.* **2009**, *103*, 076601.
29. A slight discrepancy between simulated and measured temperature is sometimes found near the metal contacts, as previously noted in ref 8. The discrepancy is partly attributed to inhomogeneous charge storage in SiO_2 after persistent high-field device operation and partly to the IR

imaging approach itself with the signal being a more complex convolution from several surfaces (including the metal) near the contacts. More careful, ultra-high-resolution AFM-based thermal imaging at lower bias (ref 15) found near-perfect agreement between our simulations and the temperature profile near the contacts.
30. Tsai, C.-L.; Liao, A.; Pop, E.; Shim, M. Electrical Power Dissipation in Semiconducting Carbon Nanotubes on Single Crystal Quartz and Amorphous SiO_2. *Appl. Phys. Lett.* **2011**, *99*, 053120.
31. Koh, Y. K.; Bae, M.-H.; Cahill, D. G.; Pop, E. Heat Conduction across Monolayer and Few-Layer Graphenes. *Nano Lett.* **2010**, *10*, 4363–4368.
32. Wang, Z.; Chun, I. S.; Li, X.; Ong, Z.-Y.; Pop, E.; Millet, L.; Gillette, M.; Popescu, G. Topography and Refractometry of Nanostructures Using Spatial Light Interference Microscopy. *Opt. Lett.* **2010**, *35*, 208–210.
33. Koh, Y. K.; Bae, M.-H.; Cahill, D. G.; Pop, E. Reliably Counting Atomic Planes of Few-Layer Graphene (n > 4). *ACS Nano* **2011**, *5*, 269–274.
34. QFI Corp., Infrascope II User Manual. www.quantumfocus.com, 2005.