Role of Joule Heating on Current Saturation and Transient Behavior of Graphene Transistors

Sharnali Islam, Zuanyi Li, Vincent E. Dorgan, Myung-Ho Bae, and Eric Pop, *Senior Member, IEEE*

***Abstract*—We use simulations to examine current saturation in sub-micron graphene transistors on SiO_2/Si. We find that self-heating is partly responsible for current saturation (lower output conductance) but degrades current densities above 1 mA/μm by up to 15%. Heating effects are reduced if the supporting insulator is thinned or, in shorter channel devices, by partial heat sinking at the contacts. The transient behavior of such devices has thermal time constants of ∼30–300 ns, which is dominated by the thickness of the supporting insulator and that of the device capping layers (a behavior also expected in ultrathin-body SOI transistors). The results shed important physical insight into the high-field and transient behavior of graphene transistors.**

***Index Terms*—Current saturation, graphene field-effect transistor (FET), scaling, self-heating, thermal transient.**

I. Introduction

GRAPHENE has attracted much interest for transistor applications, initially due to its high carrier mobility, i.e., $\sim 10^4$ $cm^2 \cdot V^{-1} \cdot s^{-1}$ [1]. Recent work has also found drift velocity saturation at high field in graphene, at values several times higher than in silicon [2]. However, velocity saturation alone is not sufficient for current saturation because the carrier density can continue to increase with drain voltage in a zero-band-gap material, where the channel cannot be fully pinched off. Current saturation is important for low output conductance g_o and amplifier gain [1], [3], and in practice, it has been partly achieved through a combination of velocity saturation and electrostatic charge control [4], [5]. At the same time, high-field transport in graphene is also influenced by self-heating [2], [6], as revealed by recent infrared and Raman thermal imaging [7]–[10].

In this letter, we examine the effect of self-heating on current saturation in sub-micron graphene-on-insulator (GOI) transistors through electro-thermal device simulations. We consider the role of the buried oxide thickness $t_{\rm box}$ under the graphene and of the device length L. We also observe that practical graphene devices can be operated in transient (pulsed) mode

Manuscript received November 5, 2012; accepted November 21, 2012. Date of publication January 9, 2013; date of current version January 23, 2013. This work was supported in part by the Office of Naval Research, by the Nanotechnology Research Initiative (NRI), and by a National Science Foundation CAREER Award. The review of this letter was arranged by Editor L. Selmi.

S. Islam, V. E. Dorgan, and E. Pop are with the Department of Electrical and Computer Engineering and the Micro and Nanotechnology Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: epop@illinois.edu).

Z. Li is with the Micro and Nanotechnology Laboratory and the Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

M.-H. Bae was with the Department of Electrical and Computer Engineering and the Micro and Nanotechnology Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA. He is now with the Korea Research Institute of Standards and Science, Daejeon 305-340, Republic of Korea.

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

Digital Object Identifier 10.1109/LED.2012.2230393

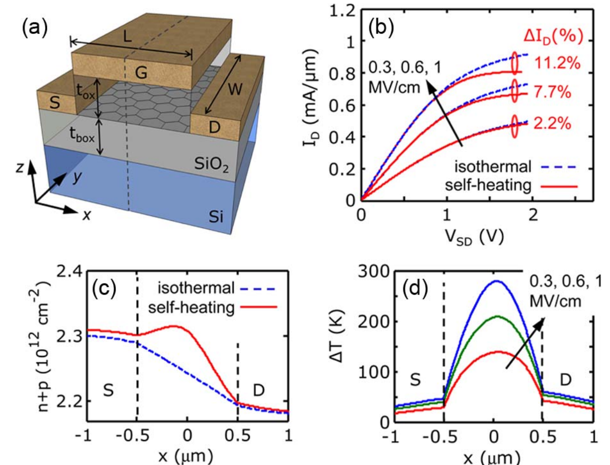

Fig. 1. (a) Schematic of a simulated graphene device on SiO_2/Si substrate (image courtesy of F. Lian). (b) Current saturation with self-heating (solid lines) vs. isothermal simulations (dashed lines) at three vertical fields ($= V_{GS}/t_{\rm ox}$). (c) Carrier density along the channel at 1 MV/cm vertical field with and without self-heating. (d) Temperature profiles at $V_{SD} = 2$ V in (b), including self-heating. The device has $L = W = 1$ μm and $t_{\rm box} = 300$ nm.

and calculate their thermal time constants, i.e., the timescales over which the device temperature ramps up or cools down after electrical switching.

II. Current Saturation

The schematic of a typical GOI transistor is shown in Fig. 1(a). Our simulations are based on the drift–diffusion approach, calculating carrier densities, electric field, drift velocity, potential, and temperature along the channel and contacts self-consistently. The simulator was extensively tuned against experimental data [8], [10], including contact effects [11]. The metal–graphene contact resistance per unit area used here is $\rho_C = 111$ $\Omega \cdot \mu m^2$, which is near the low end of the range for typical Pd–graphene or Au–graphene contacts [11]. The Dirac voltage of simulated devices is $V_0 = 0$ V, and the background temperature is $T_0 =$ 293 K. Other parameters are as in [2], including compact models of mobility and velocity saturation dependence on carrier density and temperature. Since carrier mean free paths in typical GOI transistors are in the 20 to 80 nm range [8], [10], the model is most reliable for devices greater than ∼0.1 μm.

We first investigate self-heating and current saturation in a device with channel length and width $L = W = 1$ μm. Fig. 1(b) shows the computed current-voltage (I_D–V_{SD}) of this device with $t_{\rm box} = 300$ nm SiO_2 and vertical electric fields of 0.3, 0.6, and 1.0 MV/cm, respectively. The dashed lines represent current without self-heating ($T = T_0$), whereas the solid lines show current degradation when Joule heating is self-consistently taken into account. Thus, simulations suggest that self-heating is at least partly responsible for the current

0741-3106/$31.00 © 2013 IEEE

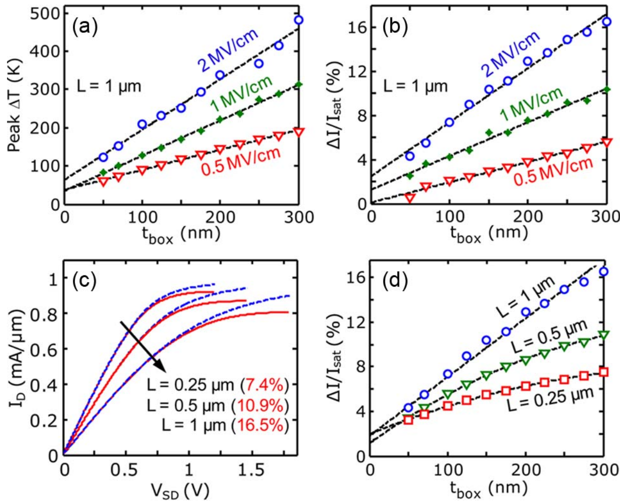

Fig. 2. (a) Calculated peak ΔT and (b) self-heating effect on saturation current as a function of $t_{\rm box}$ for three vertical fields, at channel length $L = 1$ μm. Dashed lines are linear fits. (c) Current–voltage simulations with (solid lines) and without self-heating (dashed lines) for devices of $L = 0.25, 0.5$, and 1 μm, on $t_{\rm box} = 300$ nm and vertical field of 2 MV/cm. (d) Self-heating effect on saturation current as a function of $t_{\rm box}$ for the same three channel lengths and vertical field. Dashed lines show less degradation and sublinear dependence on $t_{\rm box}$ for sub-0.5-μm channel lengths due to heat sinking effect of contacts.

saturation observed in recent experiments on devices of comparable size and bias [4], [5].

Fig. 1(c) shows the total carrier density at the highest voltage and current bias point in Fig. 1(b), with and without self-heating. Interestingly, because graphene is a gapless material, we find that significant self-heating during operation can alter the *majority* carrier concentration by thermal carrier generation [2]. Thus, self-heating at high field influences not only the current saturation but also the internal carrier distributions. Fig. 1(d) displays the temperature profiles corresponding to the maximum bias points for the three cases in Fig. 1(b). We note that sustained temperature rises $\Delta T > 200$ K have been linked with graphene device instability in experimental studies [2], [8].

We now study the peak temperature rise ΔT and the percentage of saturation current degradation $\Delta I/I_{\rm sat}$ as we reduce $t_{\rm box}$ from 300 to 50 nm. For all $t_{\rm box}$, the peak ΔT and $I_{\rm sat}$ are taken at the same $V_{SD} = 2$ V for vertical fields of 0.5, 1, and 2 MV/cm. Fig. 2(a) shows that the peak ΔT of devices with channel length $L = 1$ μm proportionally scales with $t_{\rm box}$, as expected. However, we note that, even in the limit of $t_{\rm box} \rightarrow 0$ (graphene device directly on substrate, which is similar to graphene on SiC), the temperature rise is nonzero due to the thermal resistance of the graphene–substrate interface and that of heat spreading into the substrate itself [2], [10]. Fig. 2(b) shows $\Delta I/I_{\rm sat}$ due to self-heating as a function of $t_{\rm box}$. As a simple guideline, a $\sim$5% degradation in $I_{\rm sat}$ corresponds to $\Delta T \sim 170$ K above room temperature. For current density near $\sim$1 mA/μm, as for the top curve in Fig. 1(b) on $t_{\rm box} = 300$ nm, the current degradation due to Joule heating can be > 10%, and for higher current densities, the self-heating effect is proportionally larger. This can be partly compensated by reducing $t_{\rm box}$ and L, as described here and in the subsequent discussion. We note that elevated temperatures not only decrease device performance but also have profound effects on long-term device and dielectric reliability [12].

We next explore the effect of Joule heating while scaling the channel length from 1 to 0.25 μm. Fig. 2(c) shows current–voltage curves computed with and without self-heating, indicating that the self-heating effect is less in shorter channel devices. Fig. 2(d) also plots $\Delta I/I_{\rm sat}$ due to self-heating versus $t_{\rm box}$ for the same channel lengths, at the same drain output conductance $g_o = \partial I_D/\partial V_{SD}$. Less current degradation at shorter channel lengths is explained by an enhanced role of heat dissipation "laterally" to the contacts in addition to "vertically" through the oxide. This was also recently observed in experimental work on sub-0.5-μm graphene nanoribbons (GNRs) [13], which noted that heat dissipation into the contacts begins to play a role when the device length is $\leq \sim$3 times the thermal healing length (L_H). The thermal healing length is a measure of the lateral heat diffusion along the graphene, $L_H \approx 0.2$ μm in graphene on 300 nm thick SiO_2 and approximately half in GNRs, which have lower thermal conductivity [8], [13]. Increased heat loss to the contacts is also seen as a sublinear rise in current degradation, as shown in Fig. 2(d), for the shorter devices. Our present model numerically accounts for heat spreading into the substrate and the contacts [10]; however, this can be also treated to a good approximation analytically as in [13].

From a practical point of view, our simulations suggest that thermal effects are always significant in modern graphene devices [4], [5] at current densities > 0.5 mA/μm or lateral fields > 1 V/μm. To avoid this, devices could be built on substrates with thinner $t_{\rm box}$ or higher thermal conductivity (e.g., sapphire). However, some amount of self-heating can lead to better current saturation (lower g_o), but not carefully considering such effects can cause long-term device instability [12].

III. Thermal Transient

While the preceding section focused on the effects of self-heating on DC characteristics, this section explores the transient device behavior. We perform finite-element (FE) simulations, as shown in Fig. 3(a) and (b), which show the temperature of one-half of the cross section marked by a dashed line in Fig. 1(a). Isothermal boundary conditions ($T = T_0$) are applied 10 μm away from the device at the bottom and right edges of the Si substrate, and other boundaries are adiabatic. Thermal boundary resistance is accounted for at graphene interfaces with SiO_2 [2], [10]. We used temperature-dependent values for the thermal conductivity and heat capacity of the oxide [14], although the effect was small, < 5%.

An input power of 0.5 mW is initially applied to the graphene channel and then turned off after 2.5 μs. Fig. 3(a) and (b) corresponds to temperature distributions at the end of the heating pulse in a device without and with a capping layer (assumed to be SiO_2), respectively. This can be roughly understood as a typical device in a laboratory setup, versus one that is integrated in a package. The temperature transient of the channel midpoint is shown in the inset in Fig. 3(c) for a capping layer $t_{\rm cap} = 200$ nm and $t_{\rm box} = 250$ nm. The thermal time constant τ is obtained by fitting the temperature decay as $T(t) = T_0 + T_1 e^{-t/\tau} + T_2(1 + t/\tau_0)^{-b}$, where $T_0 = 293$ K is the base temperature, $T_0 + T_1 + T_2$ is the steady-state peak temperature, and the third term is used to fit the long tail of the temperature decay due to the (small) residual heating transient of the Si substrate. Typical values for b are in the range of 0.5–2.5, and typical values for τ_0 are from tens to hundreds of nanoseconds. In the "no cap" case, T_2 is less than 30% T_1, but it becomes comparable with T_1 in cases "with cap."

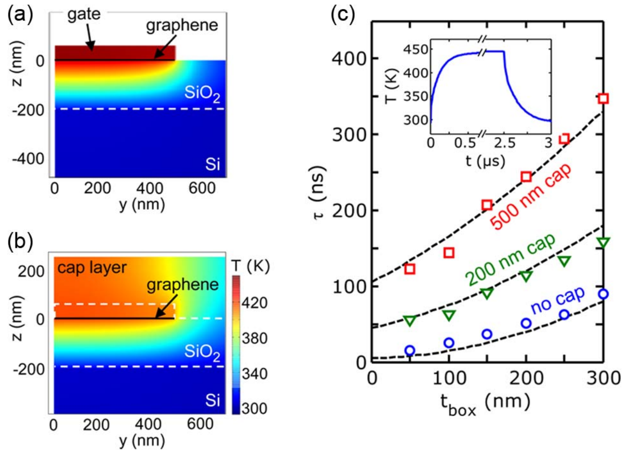

Fig. 3. Cross section of graphene device temperature (a) with no capping layer and (b) with a 200-nm SiO_2 cap layer, 2.5 μs after a 0.5-mW input pulse. (c) Calculated thermal time constants of graphene devices as a function of t_{box} (○) without a capping layer, (▽) with a 200-nm cap layer, and (□) with a 500-nm cap layer. Dashed lines are fits with (1). The inset shows the temperature transient for $t_{\mathrm{cap}} = 200$ nm and $t_{\mathrm{box}} = 250$ nm.

The symbols in Fig. 3(c) summarize the calculated thermal time constant of the graphene device, as the t_{box} is scaled, for devices with a capping layer of 200 and 500 nm and without ("no cap"). We can understand the scaling of the thermal time constant through a simple analytic model, which includes each region as a lumped thermal resistance R_{th} and thermal capacitance C_{th}. The thermal time constant τ is the sum of contributions $(\Sigma R_{\mathrm{th}} C_{\mathrm{th}})$ from the relevant regions, i.e.,

$$\tau \approx f_1 \frac{C_V}{k_{\mathrm{ox}}} t_{\mathrm{box}}^2 + \left[f_2 \frac{C_V}{k_{\mathrm{ox}}} t_{\mathrm{cap}} + \frac{C_{\mathrm{Vm}}}{k_m} t_m \right] (t_{\mathrm{box}} + t_{\mathrm{eq}}) \quad (1)$$

where $C_V = 1.76\ \mathrm{MJ \cdot K^{-1} \cdot m^{-3}}$ and $C_{\mathrm{Vm}} = 2.88\ \mathrm{MJ \cdot K^{-1} \cdot m^{-3}}$ are the heat capacities of the oxide and metal gate [15], respectively; and t_m and $k_m (= 40\ \mathrm{W \cdot m^{-1} \cdot K^{-1}}$ for Pd) are the thickness and the thermal conductivity of the metal gate, respectively. The geometrical prefactors $f_1 \sim 0.6$ and $f_2 \sim 0.8$ represent the fractions of the total temperature drop in the bottom oxide and in the top capping layer, respectively. The last term $t_{\mathrm{eq}} \approx$ 200 nm accounts for the thermal equivalent of transient cooling in the Si substrate (the limit $t_{\mathrm{box}} \rightarrow 0$), which is consistent with previous studies on bulk CMOS devices [16]. We note that the aforementioned analytic model could also be applied to other devices based on atomically thin materials such as MoS_2 or to ultrathin-body silicon-on-insulator (SOI) transistors.

The model of (1) is plotted with dashed lines in Fig. 3(c), which is in good agreement with our FE simulations (symbols). The FE results are realistic within 10%–20% accuracy, depending on the simulated domain size and the choice of 3-D versus 2-D simulations (the main tradeoff being CPU time); however, the main physical trends persist. These results suggest that thermal time constants follow an approximately quadratic dependence on t_{box}, which contributes to both the thermal resistance and the thermal capacitance of the device.

The capping layer and the metal gate do contribute to the term in (1) that is linear in t_{box} but do not aid in "cooling" the device otherwise. Thus, a thicker gate or capping layer only adds "thermal ballast" and can increase the thermal time constant. Interestingly, due to its thinness, the graphene layer itself does not influence the thermal transient of the device, which is dominated by the heating of the surrounding materials. This is a unique aspect of devices based on graphene (or other 2-D monolayer materials such as MoS_2) versus that of older SOI technology, where a substantial thickness of the Si "body" retains nonnegligible heat capacity and thermal resistance [17], [18].

To conclude, we have found that Joule heating during operation is partly responsible for current saturation and degradation observed in graphene device experiments. Self-heating is reduced with thinner dielectrics, and for sub-0.5-μm channel lengths the contacts begin to play a role in heat sinking. The thermal time constants of GOI devices are in the range between 30–300 ns and are strongly dependent on the materials surrounding the channel. Thermal transients are much slower than electrical transients (~1–10 ps), which is consistent with previous work on SOI technology [17], [18]. This implies that graphene devices are slow to heat up or cool down after electrical switching, and for instance, pulsed operation on timescales shorter than the thermal time constant can benefit from reduced self-heating compared with DC operating modes.

References

[1] F. Schwierz, "Graphene transistors," *Nat. Nanotechnol.*, vol. 5, no. 7, pp. 487–496, Jul. 2010.

[2] V. E. Dorgan, M.-H. Bae, and E. Pop, "Mobility and saturation velocity in graphene on SiO_2," *Appl. Phys. Lett.*, vol. 97, no. 8, pp. 082112-1–082112-3, Aug. 2010.

[3] S. Das and J. Appenzeller, "On the importance of bandgap formation in graphene for analog device applications," *IEEE Trans. Nanotechnol.*, vol. 10, no. 5, pp. 1093–1098, Sep. 2011.

[4] J. Bai, L. Liao, H. Zhou, R. Cheng, L. Liu, Y. Huang, and X. Duan, "Top-gated chemical vapor deposition grown graphene transistors with current saturation," *Nano Lett.*, vol. 11, no. 6, pp. 2555–2559, Jun. 2011.

[5] I. Meric, C. R. Dean, A. F. Young, N. Baklitskaya, N. J. Tremblay, C. Nuckolls, P. Kim, and K. L. Shepard, "Channel length scaling in graphene field-effect transistors studied with pulsed current–voltage measurements," *Nano Lett.*, vol. 11, no. 3, pp. 1093–1097, Mar. 2011.

[6] V. Perebeinos and P. Avouris, "Inelastic scattering and current saturation in graphene," *Phys. Rev. B*, vol. 81, no. 19, pp. 195442-1–195442-8, May 2010.

[7] D.-H. Chae, B. Krauss, K. von Klitzing, and J. H. Smet, "Hot phonons in an electrically biased graphene constriction," *Nano Lett.*, vol. 10, no. 2, pp. 466–471, Feb. 2010.

[8] M.-H. Bae, Z. Y. Ong, D. Estrada, and E. Pop, "Imaging, simulation, and electrostatic control of power dissipation in graphene devices," *Nano Lett.*, vol. 10, no. 12, pp. 4787–4793, Jun. 2010.

[9] M. Freitag, H. Y. Chiu, M. Steiner, V. Perebeinos, and P. Avouris, "Thermal infrared emission from biased graphene," *Nat. Nanotechnol.*, vol. 5, no. 7, pp. 497–501, Jul. 2010.

[10] M.-H. Bae, S. Islam, V. E. Dorgan, and E. Pop, "Scaling of high-field transport and localized heating in graphene transistors," *ACS Nano*, vol. 5, no. 10, pp. 7936–7944, Sep. 2011.

[11] K. L. Grosse, M. H. Bae, F. Lian, E. Pop, and W. P. King, "Nanoscale Joule heating, Peltier cooling and current crowding at graphene–metal contacts," *Nat. Nanotechnol.*, vol. 6, no. 5, pp. 287–290, May 2011.

[12] D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," *J. Appl. Phys.*, vol. 94, no. 1, pp. 1–18, Jul. 2003.

[13] A. D. Liao, J. Z. Wu, X. Wang, K. Tahy, D. Jena, H. Dai, and E. Pop, "Thermally limited current carrying ability of graphene nanoribbons," *Phys. Rev. Lett.*, vol. 106, no. 25, pp. 256801-1–256801-4, Jun. 2011.

[14] D. G. Cahill, "Thermal conductivity measurement from 30 to 750 K: The 3 ω method," *Rev. Sci. Instrum.*, vol. 61, no. 2, pp. 802–808, Feb. 1990.

[15] F. P. Incropera, D. P. DeWitt, T. L. Bergman, and A. S. Lavine, *Fundamentals of Heat and Mass Transfer*. New York: Wiley, 2007, p. 931.

[16] D. Takacs and J. Trager, "Temperature increase by self-heating in VLSI CMOS," in *Proc. ESSDERC*, 1987, pp. 729–732.

[17] S. Polonsky and K. A. Jenkins, "Time-resolved measurements of self-heating in SOI and strained-silicon MOSFETs using photon emission microscopy," *IEEE Electron Device Lett.*, vol. 25, no. 4, pp. 208–210, Apr. 2004.

[18] A. L. Caviglia and A. A. Iliadis, "Linear dynamic self-heating in SOI MOSFETs," *IEEE Electron Device Lett.*, vol. 14, no. 3, pp. 133–135, Mar. 1993.