System Level Analysis and Benchmarking of Graphene Interconnects for Low-Power Applications

Vachan Kumar[1], Ramy Nashed[1], Kevin Brenner[2], Romeil Sandhu[2], and Azad Naeemi[1],
[1]School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA
[2]Harper Laboratories, LLC, Atlanta, GA
(Invited paper)

***Abstract*—Stochastic wiring distribution models are used to predict the improvement in energy obtained by replacing a few or all copper metal levels with graphene nanoribbons (GNRs) in a low-power digital circuit. The models developed here also estimate the degradation in the performance by replacing a few or all copper metal levels with GNRs. Replacing a few local copper interconnect levels with GNRs is expected to reduce the energy consumed by local interconnects, without severely degrading the performance of longer global interconnects. The hybrid GNR+copper interconnect is shown to perform worse compared to the all GNR interconnect, if the length of the GNR segment is greater than a critical value. For a logic circuit with $30k$ gates, it is shown that the hybrid interconnect offers a 30 to 40% decrease in energy and a $4\times$ decrease in maximum frequency, whereas the all GNR interconnect offers a 50 to 60% decrease in energy and a $7\times$ decrease in maximum frequency. Further, the impact of edge doping on the resistance per unit length of graphene is analyzed.**

***Index Terms*—Graphene, nanoribbons, interconnects, low-power, digital circuits, edge doping**

I. Introduction

Over the last five decades, transistor scaling [1], [2] has driven the tremendous gains seen in the performance and power of integrated circuits. While transistor performance improves with scaling, interconnect performance degrades due to an increase in the resistance per unit length. This results in a significant degradation in interconnect performance with technology scaling [3]. Additionally, with the interconnect dimensions scaling below the mean free path of electrons in copper ($40nm$), the resistivity of copper rises sharply due to size effects [4], [5]. As a result, the narrow local interconnects suffer due to a sharp rise in resistance with scaling. Additionally, it was shown that for a microprocessor, interconnects consumed 51% of the total power and the local interconnects consumed 47% of the interconnect power [6]. In addition to performance and power challenges, copper interconnects also suffer from reliability issues like electromigration. As a result, the semiconductor industry is in search of novel materials to solve the on-chip interconnects issues. Graphene, which is a single sheet of carbon atoms arranged in a honeycomb lattice structure, was recently shown to exist in a stable state in nature[7], [8]. High mean free path of electrons, smaller capacitance, and higher current carrying capacity make graphene an interesting candidate for replacing copper as the on-chip interconnect material [9], [10].

Although two dimensional graphene suspended in air has been shown to have superior mean free path of $1\mu m$ [11], the mean free path reduces to below $100nm$ when graphene is placed on SiO_2 substrate. The mean free path degrades further when the two dimensional graphene is patterned into thin strips called graphene nanoribbons (GNRs). As a result, the effective resistance per unit length of currently available GNRs can be high compared to that of copper. However, the capacitance of single-layer GNRs is roughly $2.5\times$ smaller compared to that of copper. As a result, graphene can be useful for very low-power applications, where the energy consumed is more important compared to performance. In this study, we explore the use of graphene nanoribbons for very low-power applications in advanced technology nodes. The impact of edge doping on the resistance of GNRs [12] is estimated using the models presented in [13], [14], and the edge doping that minimizes the resistance is computed. For a logic circuit with $30k$ gates, the stochastic wiring models developed in [15] are used to predict the performance and energy of GNR interconnects. Additionally, the impact of replacing a few local copper interconnect levels with GNR to form a hybrid GNR+copper interconnect is studied.

The paper is divided into five sections. Section II deals with the impact of edge doping on the resistance of GNRs. Section III briefly explains the system level models and interconnect architectures used for the comparison of copper and GNRs. The repeater insertion algorithm and the limitations of the hybrid GNR+copper interconnect are also explained in section III. The comparison of the three interconnect architectures in terms of energy and maximum frequency is presented in section IV. The important conclusions are summarized in section V.

II. Impact of Edge Doping on Graphene Resistance

Two dimensional graphene suspended in air has been experimentally shown to possess superior transport properties like mean free path and mobility [11]. However, when graphene is placed on a substrate, the mean free path drops significantly due to surface polar phonons and charged impurities at the interface [14]. Thus, the mean free path of graphene is strongly dependent on the quality of the substrate, and is roughly $100nm$ on silicon dioxide. Further, when the graphene sheets are patterned into thin graphene nanoribbons, the scattering at the edges results in a decrease in the mean free path. The

978-1-4799-5545-9/14/$31.00

impact of edge scattering is more pronounced at advanced technology nodes with smaller wire widths. Edge doping of GNRs, as shown in Fig. 1, improves the resistance by increasing the Fermi level and hence the number of channels available for conduction [12], [13]. The dependence of the per unit length (p.u.l) resistance of GNRs on the doping concentration for a 7.5nm wide wire is shown in Fig. 2.

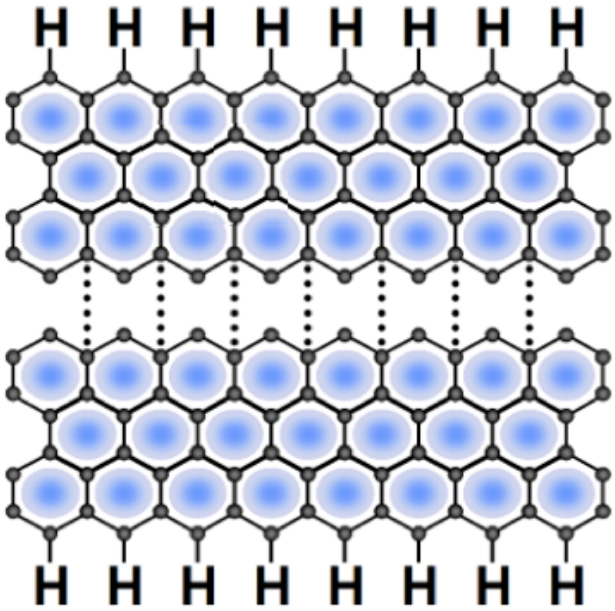

Fig. 1. Edge doping of graphene with hydrogen [12]. The H-passivation at the edge results in sp^2 hybridization.

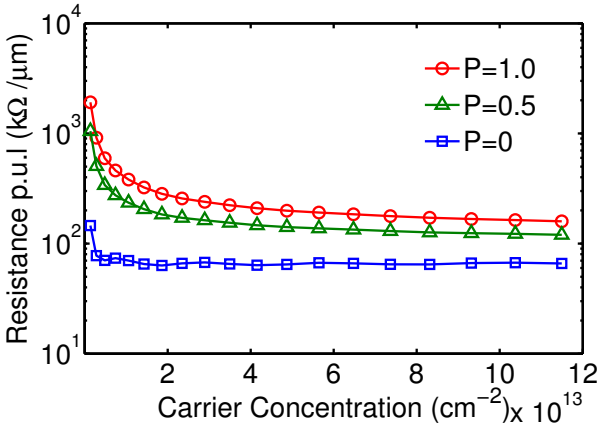

Fig. 2. Resistance per unit length of a GNR interconnect (width=7.5nm) as a function of doping concentration for different values of backscattering probabilities at the GNR edges.

At smaller doping concentrations, the p.u.l resistance decreases significantly with an increase in doping. However, beyond a certain doping concentration, the scattering due to phonons dominates and the mean free path decreases with an increase in doping. Thus, the increase in the number of conduction channels is nullified by the decrease in the effective mean free path. As a result, the p.u.l resistance saturates or increases slightly with an increase in doping beyond a certain doping concentration. In this study, the doping concentration at which the per unit length resistance is 2% higher compared to the saturated value is defined as the optimal doping concentration. The optimal carrier concentration as a function of the ITRS technology node [16] (referred for minimum width) is shown in Fig. 3. At higher edge scattering probabilities ($P = 0.5$ and $P = 1$), the impact of the phonons on the mean free path is smaller; hence, the optimal doping concentration is higher for GNRs with rough edges. At $P = 0$, the impact of scattering due to phonons is high; hence, the resistance per unit length saturates at smaller values of carrier concentration. However, at very small widths, the number of conduction channels is so small that the resistance per unit length saturates at higher values of carrier concentration.

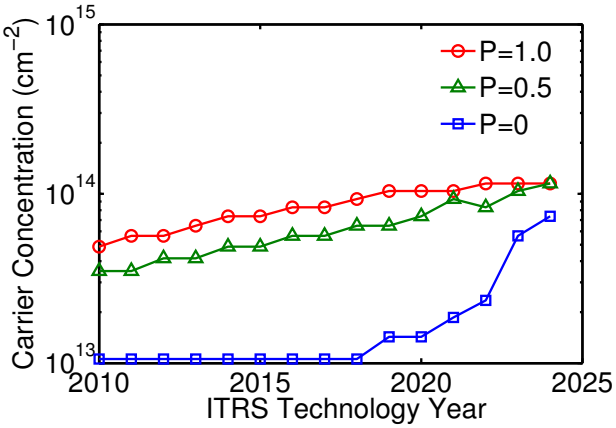

Fig. 3. Optimal carrier concentration to minimize the resistance per unit length.

III. System Level Modeling

The system level model for estimating the impact of using GNR interconnects on the performance and energy of low-power circuits is developed in this section. The low-power circuit is assumed to be a simple core with $30k$ gates. The wiring distribution inside the core is assumed to be given by the stochastic wiring distribution models presented in [15]. The wiring distribution in the core as a function of wire length is shown in Fig. 4. From the wiring distribution, it is clear that the number of short wires is significantly higher compared to the longer wires. As a result, even if GNRs replace copper wires at the local interconnect level, a significant saving in energy is possible. The interconnect architectures used for the comparison of performance and energy, and the repeater insertion algorithm are described in the subsections below.

A. Structures for Comparison

The interconnect architectures used for comparison of copper wires and GNRs at the system level are shown in Fig. 5.

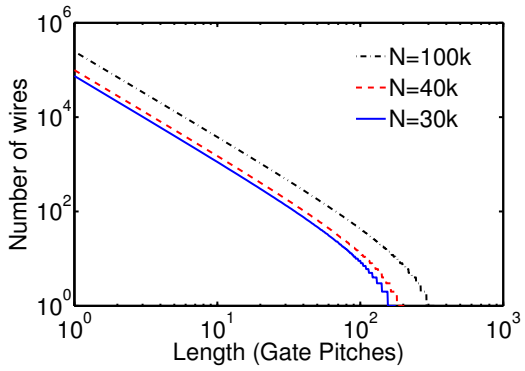

Fig. 4. Stochastic wiring distribution model as a function of the number of logic gates [15].

1) **All copper:** This is the conventional baseline interconnect architecture for benchmarking the performance and energy of the other 2 interconnect architectures. The circuit model in Fig. 5 (a) includes a driver resistance, diffusion capacitance assumed to be the same as the gate capacitance, the distributed RC network for the copper interconnect and the load capacitance of an identical cell. The delay of the circuit is given by the Elmore delay model

$$t_{cu,1}(h, L_c) = 0.69 R_0 C_0 + 0.69 \frac{R_0 c_c L_c}{h} + 0.69\left(\frac{R_0}{h} + r_c L_c\right) C_0 h + 0.38 r_c c_c L_c^2 \quad (1)$$

where $t_{cu,1}$ is the delay of the circuit shown in Fig. 5(a), R_0 is the resistance of a minimum size CMOS driver obtained from ITRS [16], C_0 is the capacitance of a minimum size CMOS driver obtained from ITRS [16], c_c is the capacitance per unit length of copper, r_c is the resistance per unit length of copper, and L_c is the length of the copper interconnect. The total capacitance of the circuit is given by (2) below.

$$C_{cu,1}(h, L_c) = 2C_0 h + c_c L_c \quad (2)$$

2) **Hybrid:** The hybrid model is used for a case where a few lower interconnect levels use GNRs, and the upper interconnect levels use copper. In this case, since the lower interconnect levels use GNRs, short wires are typically routed entirely in GNRs. Although it is preferable to route longer wires entirely using copper (because of lower resistance), a short segment of GNR is typically needed to connect the transistors to the upper metal layers. As a result, longer wires typically can be modeled with an interconnect shown in Fig. 5 (b). The length of the GNR segment can be critical in determining the performance and energy of the hybrid interconnect. This is because the high resistance of the GNR segment and the high capacitance of the copper segment can dominate the delay of this hybrid interconnect. The circuit model for the interconnect consists of the driver resistance, the total lumped resistance (R_T) including the contact and quantum resistances, the distributed RC network for the GNR segments, the distributed RC network for the copper segment and the load capacitance of an identical cell. The delay and total capacitance of the circuit are given by (3) and (4). In these equations, r_g is the resistance per unit length of the GNR interconnect, c_g is the capacitance per unit length of the GNR interconnect, and L_g is the length of the GNR interconnect.

$$\begin{aligned} t_{hyb,1}(h, L_g, L_c) = {} & 0.69 R_0 C_0 + 0.69\left(\frac{R_0}{h} + R_T\right) c_g L_g \\ & + 0.69\left(\frac{R_0}{h} + 2R_T + r_g L_g\right) c_c L_c \\ & + 0.69\left(\frac{R_0}{h} + 3R_T + r_g L_g + r_c L_c\right) c_g L_g \\ & + 0.69\left(\frac{R_0}{h} + 4R_T + 2r_g L_g + r_c L_c\right) C_0 h \\ & + 0.76 r_g c_g L_c^2 + 0.38 r_c c_c L_c^2 \end{aligned} \quad (3)$$

$$C_{hyb,1}(h, L_g, L_c) = 2C_0 h + 2c_g L_g + c_c L_c \quad (4)$$

3) **All GNR:** This interconnect is shown in Fig. 5 (c). The circuit model for the interconnect consists of the driver resistance, the total lumped resistance (R_T) including the contact and quantum resistances, the distributed RC network for the GNR interconnect and the load capacitance of an identical cell. The delay and total capacitance of the circuit are given by (5) and (6).

$$\begin{aligned} t_{gnr,1}(h, L_g) = {} & 0.69 R_0 C_0 + 0.69\left(\frac{R_0}{h} + R_T\right) c_g L_g \\ & + 0.69\left(\frac{R_0}{h} + 2 * R_T + r_g L_g\right) c_c L_c) \\ & + 0.38 r_g c_g L_c^2 \end{aligned} \quad (5)$$

$$C_{gnr,1}(h, L_g) = 2C_0 h + c_g L_g \quad (6)$$

B. Repeater Insertion

Since the delay of long interconnects increases quadratically with the length of the interconnect, repeaters are typically added to break down the interconnect into smaller segments and make the delay linearly dependent on the interconnect length. The delay and energy of the all copper architecture with repeater insertion is given by

$$t_{cu}(h, k, L_{tot}) = k t_{cu,1}\left(h, \frac{L_{tot}}{k}\right) \quad (7)$$

$$E_{cu}(h, k, L_{tot}) = \frac{k}{2} C_{cu,1}\left(h, \frac{L_{tot}}{k}\right) V_{dd}^2 \quad (8)$$

where L_{tot} is the total length of the interconnect and k is the number of repeaters. The delay and energy of the all

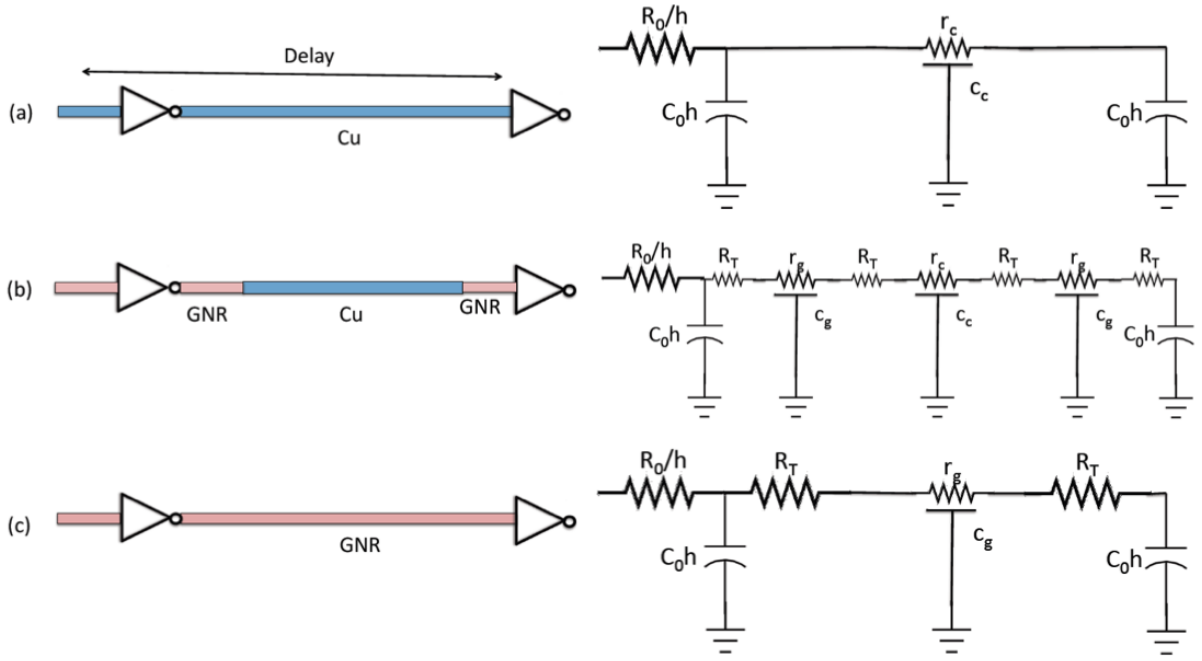

Fig. 5. The interconnect architectures used and the corresponding circuit models for the comparison of delay and energy. (a) The baseline interconnect for comparison, with the entire signal routed using copper (b) A hybrid interconnect with routing in both GNR and copper layers. (c) An interconnect with the entire signal routed in GNR. In the above architectures, the driver resistance ($\frac{R_0}{h}$) and capacitance ($C_0 h$), the receiver capacitance ($C_0 h$), and the contact resistance (R_T) are modeled as lumped circuit elements, whereas the interconnects are modeled as distributed RC networks.

GNR architecture with repeaters is given by equations similar to (9) and (10) above. The delay and energy of the hybrid architecture with repeaters is given by

$$t_{hyb}(h,k,L_g,L_{tot}) = k t_{hyb,1}\left(h, L_g, \frac{L_{tot}}{k} - 2L_g\right) \quad (9)$$

$$E_{hyb}(h,k,L_g,L_{tot}) = \frac{k}{2} C_{hyb,1}\left(h, L_g, \frac{L_{tot}}{k} - 2L_g\right) V_{dd}^2 \quad (10)$$

where L_g is the maximum allowed length for routing in GNR. The optimal size and the number of repeaters depends on the driver resistance and capacitance, and the p.u.l resistance and capacitance of the interconnect [17]. Since the hybrid and GNR interconnects have circuit models (Fig. 5 (b) and (c)) that are different compared to the typical copper models (Fig. 5 (a)), it is necessary to optimize the repeater insertion separately for each of these cases. In this study, an optimal repeater insertion algorithm that minimizes the energy delay product is used. In addition, since the delay is weakly dependent on the size and the number of repeaters close to the optimal point, a sub-optimal repeater insertion can be used. This sub-optimal repeater insertion results in a smaller energy and area, for a small penalty in the delay.

Since the intrinsic RC product of the copper interconnect is small, the sub-optimal repeater insertion for the all copper interconnect results in a small number of larger size repeaters. On the other hand, the sub-optimal repeater insertion results in a large number of smaller size repeaters for the all GNR interconnect due to its large intrinsic RC product. However, the sub-optimal repeater insertion for the hybrid interconnect is very strongly dependent on the maximum allowed length of the GNR segment L_g. The total capacitance of the hybrid architecture with repeaters is given by (11).

$$\begin{aligned} C_{tot} &= 2C_0 hk + 2c_g L_g k + c_c(L_{tot} - 2L_g k) \\ &= c_c L_{tot} + 2k\left(C_0 h + (c_g - c_c)L_g\right) \end{aligned} \quad (11)$$

From (11), it is clear that if $C_0 h + (c_g - c_c)L_g < 0$, repeater insertion increases the capacitance; hence, for $L_g > L_{g,crit}(= \frac{C_0 h}{c_c - c_g})$, the sub-optimal repeater insertion for the hybrid architecture results in a routing structure very similar to the all GNR architecture. However, if $L_g < L_{g,crit}(= \frac{C_0 h}{c_c - c_g})$, shown in Fig.6, the hybrid interconnect results in an energy lower compared to the all copper interconnect, but higher compared to the all GNR interconnect. If the GNR length is greater than the critical length $L_{g,crit}$, the degradation due to the high resistance of the GNR segment and the high capacitance of the copper segment forces the hybrid interconnect to use a large number of smaller sized repeaters. As a result, if the GNR segment length is greater than the critical length, the hybrid interconnect is almost identical to the all GNR interconnect. Thus, to ensure that the hybrid interconnect and the all GNR interconnect are different, the length of the GNR segments in the hybrid interconnect should be lower than the critical length.

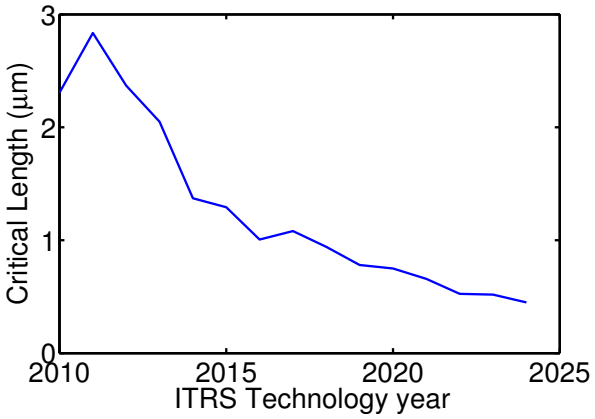

Fig. 6. Maximum length of routing in GNR layers to ensure that the optimal repeater insertion for the hybrid interconnect does not result in all GNR routing.

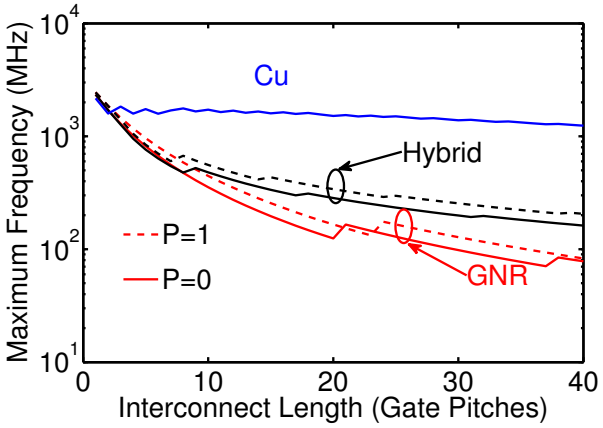

Fig. 7. Maximum frequency as a function of the length of the interconnect in a gate dominated critical path with a logic depth of 40.

IV. Results

The circuit models presented in the previous section are used to estimate the performance and the energy consumption of the three interconnect architectures. For a core with $30k$ gates, we assume that the critical path has a logic depth of 40 gates and is gate-dominated. Gate-dominated paths typically have shorter interconnects and a major portion of the clock cycle is dedicated to logic gate delays, rather than interconnect delays. The maximum frequency of the core as a function of the interconnect length in the critical path is given by (12) and shown in Fig. 7.

$$F_{max} = \frac{1}{N_{crit} t_{cu,hyb,gnr}} \tag{12}$$

where N_{crit} is the logic depth of the critical path. At short interconnect lengths, since the driver resistance and the interconnect capacitance dominate the delay, the all GNR interconnect performs better compared to the all copper interconnect. However, as the interconnect length increases, the intrinsic RC delay of the GNR interconnect dominates; hence, the all copper interconnect performs better. The hybrid interconnect has a performance somewhere in between the all copper and the all GNR interconnects. Since the doping concentration is optimized for each edge scattering probability, the maximum frequency of the hybrid and all GNR interconnects does not have a very strong dependence on the edge scattering probability. The maximum frequency as a function of the ITRS technology year is shown in Fig. 8. At lower technology nodes, the resistance of copper degrades significantly due to size effects; hence, the relative performance of the all GNR and hybrid interconnects compared to copper improves at advanced technology nodes. Further, due to improvements in the driver capacitance, the maximum frequencies improve with scaling.

In a core with $30k$ gates, the total energy consumed if the output of every single gate is switched simultaneously is shown in Fig. 9. The hybrid interconnect results in a 30 to 40% smaller energy compared to the all copper interconnect. Similarly, the all GNR interconnect results in a 50 to 60% smaller

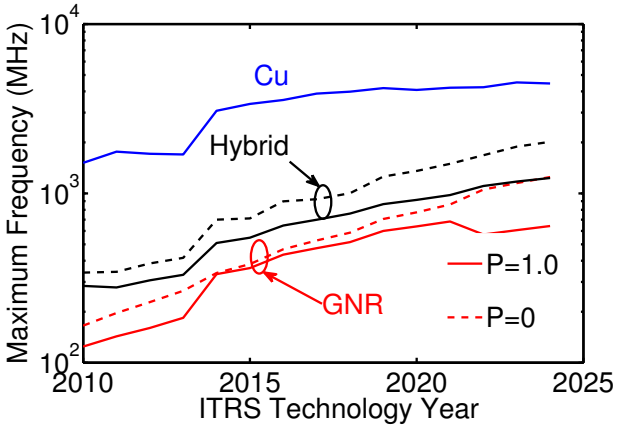

Fig. 8. Maximum frequency as a function of ITRS technology year, assuming a gate dominated critical path with a logic depth of 40 and an interconnect length of 20 gate pitches.

energy compared to the all copper interconnect. However, due to the use of a large number of smaller sized repeaters, the all GNR interconnect uses a significant number of repeaters, as shown in Fig. 10. The number of repeaters used by the hybrid interconnect and the all copper interconnect is approximately $20\times$ smaller compared to that used by the GNR interconnects with rough edges.

V. Conclusions

In this paper, system level models have been developed to estimate the performance and energy dissipation of low-power circuits using GNR interconnects. To reduce the impact of edge scattering on the performance of GNR interconnects at advanced technology nodes, the carrier concentration was optimized for minimum resistance per unit length. Further, interconnect architectures using only GNRs, as well as hybrid interconnects using both GNR and copper, are benchmarked against conventional interconnects using only copper. Replacing a few local copper interconnect levels with GNRs is expected to reduce the energy consumed by local interconnects,

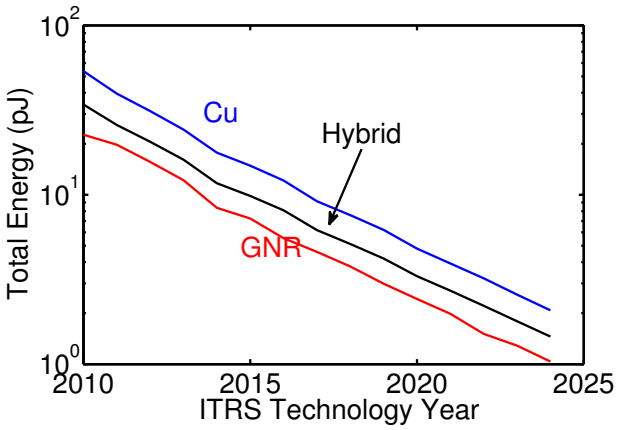

Fig. 9. Total energy consumed by the circuit for the 3 interconnect architectures: all copper, hybrid and all GNR.

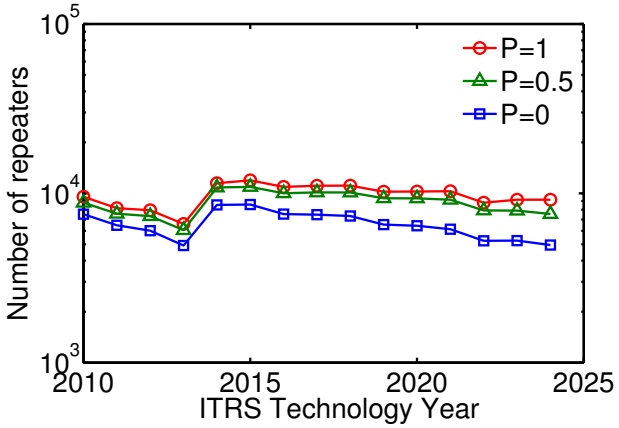

Fig. 10. Total number of repeaters used for routing the all GNR interconnect. The number of repeaters used for the other 2 interconnect architectures is small compared to the all GNR interconnect.

without severely degrading the performance of longer global interconnects. It is shown that if GNRs used in the hybrid interconnect exceed a certain critical length, the hybrid interconnect performs worse compared to the all GNR interconnect. On the other hand, if the length of GNRs in the hybrid interconnect is smaller than the critical length, the hybrid interconnect shows a 30 to 40% improvement in energy, while operating at a maximum frequency $4\times$ smaller compared to the all copper interconnect. The all GNR interconnect shows a 50 to 60% improvement in energy, while operating at a maximum frequency $7\times$ smaller compared to the all copper interconnect.

Acknowledgements

This work was sponsored by the National Science Foundation (NSF) CAREER Award and Harper Laboratories, LLC.

References

[1] R. Dennard, F. Gaensslen, H.-N. Yu, V. LEO RIDEOVT, E. Bassous, and A. R. Leblanc, "Design of ion-implanted mosfet's with very small physical dimensions," *Solid-State Circuits Society Newsletter, IEEE*, vol. 12, no. 1, pp. 38–50, 2007.

[2] G. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, no. 8, April 1965.

[3] M. Bohr, "Interconnect scaling-the real limiter to high performance ulsi," in *Electron Devices Meeting, 1995. IEDM '95., International*, 1995, pp. 241–244.

[4] G. G. Lopez, "The impact of interconnect process variations and size effects for gigascale integration," Ph.D. dissertation, School of Electrical and Computer Engineering, Georgia Institute of Technology, 2009.

[5] W. Steinhoegl, G. Schindler, G. Steinlesberger, M. Traving, and M. Engelhardt, "Scaling laws for the resistivity increase of sub-100 nm interconnects," in *Simulation of Semiconductor Processes and Devices, 2003. SISPAD 2003. International Conference on*, 2003, pp. 27–30.

[6] N. Magen, A. Kolondy, U. Weiser, and N. Shamir, "Interconnect-power dissipation in a microprocessor," *SLIP*, 2004.

[7] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, "Electric field effect in atomically thin carbon films," *Science*, vol. 306, no. 5696, pp. 666–669, 2004. [Online]. Available: http://www.sciencemag.org/content/306/5696/666.abstract

[8] C. Berger, Z. Song, X. Li, X. Wu, N. Brown, C. Naud, D. Mayou, T. Li, J. Hass, A. N. Marchenkov, E. H. Conrad, P. N. First, and W. A. de heer, "Electronic confinement and coherence in patterned epitaxial graphene," *Science*, vol. 312, May 2006.

[9] K.-J. Lee, H. Park, J. Kong, and A. Chandrakasan, "Demonstration of a subthreshold fpga using monolithically integrated graphene interconnects," *Electron Devices, IEEE Transactions on*, vol. 60, no. 1, pp. 383–390, 2012.

[10] A. Naeemi and J. D. Meindl, "Compact physics-based circuit models for graphene nanoribbon interconnects," *IEEE Transactions on Electron Devices*, vol. 56, no. 9, September 2009.

[11] K. Bolotin, K. Sikes, Z. Jiang, M. Klima, G. Fudenberg, J. Hone, P. Kim, and H. Stormer, "Ultrahigh electron mobility in suspended graphene," *Solid State Communications*, vol. 146, no. 9â"10, pp. 351 – 355, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038109808001178

[12] K. Brenner, Y. Yang, and R. Murali, "Edge doping of graphene sheets," *Carbon*, vol. 50, no. 2, pp. 637–645, 2012.

[13] V. Kumar, S. Rakheja, and A. Naeemi, "Performance and energy-per-bit modeling of multilayer graphene nanoribbon conductors," *Electron Devices, IEEE Transactions on*, vol. 59, no. 10, pp. 2753–2761, Oct. 2012.

[14] S. Rakheja, V. Kumar, and A. Naeemi, "Evaluation of the potential performance of graphene nanoribbons as on-chip interconnects," *Proceedings of the IEEE*, 2013.

[15] J. A. Davis, V. K. De, and J. D. Meindl, "A stochastic wire-length distribution for gigascale integration (gsi)–part i: Derivation and validation," *IEEE Transactions on Electron Devices*, vol. 45, no. 3, pp. 580–589, 1998.

[16] "International technology roadmap for semiconductors," online, http://www.itrs.net/Links/2010ITRS/Home2010.htm.

[17] H. B. Bakoglu, *Circuits, Interconnections and Packaging for VLSI*, 1st ed. Springer, 1990.