Carbon Nanotube Interconnects: Implications for Performance, Power Dissipation and Thermal Management

Navin Srivastava, Rajiv V. Joshi* and Kaustav Banerjee

Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106
*IBM T. J. Watson Research Center, Yorktown Heights, NY 10598

Abstract

This paper presents a comprehensive evaluation of carbon nanotube bundle interconnects from all aspects critical to VLSI circuits - performance, power dissipation and reliability - while taking into account practical limitations of the technology. A novel delay model for CNT bundle interconnects has been developed, using which it is shown that CNT bundles can significantly improve the performance of long global interconnects with minimal additional power dissipation (for maximum metallic CNT density). While it is well known that CNT bundle interconnects can carry much higher current densities than copper, their impact on back-end thermal management and interconnect temperature rise is presented here for the first time. It is shown that the use of CNT bundle vias integrated with copper interconnects can improve copper interconnect lifetime by two orders of magnitude and also reduce optimal global interconnect delay by as much as 30%.

Introduction

Metallic carbon nanotubes (CNTs), with their excellent thermal and mechanical properties, have been suggested for use as interconnects in future VLSI designs mainly because of their high current carrying capacity (in excess of 10^9 A/cm^2) [1] which can alleviate electromigration problems that plague metal interconnects. Due to the high fundamental resistance ($R_F \approx 6.5\ K\Omega$) of the one-dimensional conducting system of an isolated single-walled carbon nanotube (SWCNT) [2], a bundle/rope consisting of several nanotubes (see **Fig. 1**) is desirable for interconnect applications. While research into the reliable fabrication and process integration of CNT bundle interconnects progresses [3, 4, 5], there is a need to identify the domains on a chip where these interconnects can replace prevalent (Cu) technology and the advantages they can offer.

Previous works comparing CNT interconnects to Cu have limited applicability and are even contradictory. In [6] it is suggested that CNT interconnects do not offer any performance benefit over copper. However, the analysis in [6] is not practical as it assumes a flat array of CNTs forming an interconnect (**Fig. 1(a)**). On the other hand, [7, 8, 9] suggest that CNT bundle interconnects have superior performance compared to Cu but the assumptions in these works are unrealistic. They do not consider the density of nanotubes in a CNT bundle nor do they analytically model the equivalent circuit parameters of CNT bundle interconnects. Realistic drivers/loads are not considered and the imperfect metal-nanotube contact resistance (often so high as to overshadow the intrinsic resistance [10]) is completely ignored. Moreover, while [7] avoids the calculation of CNT bundle capacitance by unjustifiably assuming that the capacitance is the same as that for copper interconnects, [9] does not explain how the same interconnect analysis program can be used to extract capacitance for copper interconnects as well as for CNT bundles. Finally, [9] concludes that a flat array of metallic CNTs performs better than a Cu interconnect, a result that directly contradicts [6]. All these works evaluate CNT interconnects from a performance perspective only. Although [11] demonstrates the reliability limitations of Cu vias that CNT bundles can overcome, their impact on back-end thermal profile and interconnect reliability has never been demonstrated before. The work presented here fills all these gaps in the existing literature. A comprehensive evaluation of CNT bundle interconnects vis-à-vis Cu is performed and their impact on all aspects of VLSI circuits - performance, power dissipation and reliability – is quantified, while accounting for practical limitations of the technology.

CNT Bundle Interconnect Delay Modeling

Fig. 2 depicts the interconnect structure and the equivalent circuit [12] for an isolated single-walled carbon nanotube (SWCNT) of length less than mean free path of electrons in a CNT. Due to spin degeneracy and sub-lattice degeneracy of electrons in graphene, each nanotube has four conducting channels in parallel. Hence, the conductance of an isolated ballistic single-walled CNT (SWCNT) assuming perfect contacts, given by the two-terminal Landauer-Buttiker formula, is $4e^2/h = 155\ \mu S$, which yields a resistance $R_F = 6.45\ K\Omega$ [2]. Although the current through a CNT saturates at high electrical fields, the voltage bias across an interconnect is low, and in this case, CNTs demonstrate excellent ohmic behavior. Imperfect metal-to-nanotube contacts at each of the two ends of the nanotube give rise to an additional resistance (R_{ct}, typically about $100\ K\Omega$ [9]) in series with R_F. R_F and R_{ct} are divided equally between the contacts at the two ends of the nanotube. For lengths $L > \lambda_{CNT}$, the nanotube resistance is $(h/4e^2)L/\lambda_{CNT}$ [13], which is equivalent to the resistance $R_F = h/4e^2$ appearing once every mean-free-path length. Hence, for $L > \lambda_{CNT}$, the nanotube resistance is a distributed scattering resistance per unit length ($r = (h/4e^2)/\lambda_{CNT}$) as shown in **Fig. 5**. The inductance (L_M) and capacitance (electrostatic C_E and quantum C_Q) per unit length for an isolated SWCNT are derived in [12]. The additional kinetic inductance (L_K) [12] has been excluded in this work, in the light of experimental evidence of potential drop appearing along the nanotube length [14]. Furthermore, the high frequency characteristics of carbon nanotubes reported in [15] show that the *large inductive effects expected due to L_K are not observed* even at high frequencies up to 10 GHz. Hence the inclusion of L_K, as done in [6, 7, 8], can lead to large errors in delay calculation.

The equivalent circuit parameters for a CNT *bundle* are shown in Eqs. 3-8 (**Fig. 2**). It is assumed that the n_{CNT} metallic nanotubes (Eqs. 1, 2 in **Fig. 1**) forming a bundle carry current independent of each other (as a large tunneling resistance ~MΩ exists between adjacent CNTs [16]). The presence of semi-conducting CNTs (which do not contribute to current conduction) and low packing density of metallic CNTs can be accounted for by considering a "sparsely populated" bundle (**Fig. 1(c)**). Electrostatic coupling capacitance between CNTs forming a bundle does not come into play as the CNTs are assumed to carry simultaneous and identical currents. Hence the electrostatic capacitance arises mainly from the interaction between each CNT near the edge of the bundle and the neighboring interconnects (assumed to be at ground potential). The expression for C_E (Eq. (8), **Fig. 2**) is obtained empirically by using an electromagnetic field solver. Details of this capacitance model can be found in [17].

Fig. 3(a) shows that a dense CNT bundle local interconnect (with ideal metal-nanotube contacts) has resistance much lower than that of a Cu interconnect of identical dimensions. With typical imperfect metal-nanotube contacts, resistance is higher than that of a Cu interconnect. However, for long interconnect lengths (global wires), the degrading impact of imperfect contacts diminishes, because R_{ct} is a constant resistance unlike the scattering resistance which increases linearly with length. Hence long CNT bundle interconnects will have smaller resistance than their Cu counterparts. Since CNTs are cylindrical, the CNTS at the edge of a bundle (which contribute to C_E) have larger surface area exposed to the surrounding interconnects than the corresponding surface area for a Cu interconnect with straight edges. Hence, the electrostatic capacitance of such a CNT bundle is expected to be larger than that of a Cu interconnect of equivalent dimensions. **Fig. 3(b)** shows that the capacitance of a dense CNT bundle interconnect is nearly twice that of Cu at all technology nodes. Using this delay model, along with driver parasitics and interconnect dimensions as predicted by the ITRS'04, the performance of CNT bundle and Cu interconnects of identical dimensions is compared in the following section.

Implications for Performance and Power Dissipation

Global interconnect delay is one of the top interconnect challenges facing the semiconductor industry [18]. While nominal

0-7803-9269-8/05/$20.00 (c) 2005 IEEE

gate delay and local interconnect delay traditionally decreases with technology scaling, global interconnect delay increases (**Fig. 4(a)**). However, due to the increasing resistivity of small dimension local vias and contacts, even local interconnect delay can increase as technology scales beyond 45 nm [11]. At the local interconnect level, delay is largely impacted by interconnect capacitance because of the large driver resistances and small load capacitances. Hence, CNT bundle *local* interconnects have larger delay than Cu (**Fig. 4(b)**) due to their larger capacitance. However, delay of long length (global) interconnects is largely impacted by interconnect resistance since large drivers are used to drive these interconnects. Hence, CNT bundle interconnects can reduce intermediate and global interconnect delay by as much as 80% (**Fig. 4(c, d)**) due to their lower resistance, in spite of imperfect metal-nanotube contacts.

Global interconnects are often designed by inserting buffers (repeaters) to drive signals faster [19]. Classical buffer insertion is done by minimizing delay per unit length (τ/l) (Equation (9) in **Fig. 5**). The expression for τ/l with CNT bundle interconnects is shown in Equation (10). **Fig. 6** shows that the optimal delay per unit length $(\tau/l)_{opt}$ with optimally buffered CNT bundle interconnects is lower than that with Cu (40% less for $\lambda_{CNT} = 1\ \mu m$ and as much as 80% less for $\lambda_{CNT} = 10\ \mu m$, at 22 nm node) and decreases as technology scales (inset). **Fig. 7** shows that for all technology nodes $(\tau/l)_{opt}$ with optimally buffered CNT bundle global interconnects deteriorates rapidly when $\lambda<\lambda_0$ (as in the presence of defects in the CNTs) while significant improvement in $(\tau/l)_{opt}$ can be achieved with $\lambda>\lambda_0$. The improvement in $(\tau/l)_{opt}$ slows down as λ becomes large (inset **Fig. 7**). For the buffered global interconnect with optimal delay, repeater power dissipation per unit length with CNT bundle interconnects is comparable to that with Cu for maximum metallic CNT density (**Fig. 8**). In other words, global interconnect delay can be reduced considerably by using densely packed CNT bundle interconnects without incurring additional large power dissipation. **Fig. 9** shows that even when repeater size (s) and inter-repeater length (l) are sub-optimal by a factor of 2, τ/l is less than 25% higher than $(\tau/l)_{opt}$. Hence, large power savings can be achieved for a small delay penalty (> 20% power saving for 5% delay penalty at 45 nm node) using power-optimal buffer insertion [19] (**Fig. 10**). The % saving in power increases as technology scales (consistent with the trend for Cu [19]) but is smaller than that with Cu (inset **Fig. 10**).

Implications for Thermal Management and Reliability

At nanometer scale dimensions, increasing Cu interconnect resistivity (due to enhanced surface and grain boundary scattering) in addition to increasing current density (J) [18] results in higher self-heating of interconnects. Moreover, low-k dielectrics with inherently lower thermal conductivity ($K_{th,ILD}<0.4\ W/mK$) make heat conduction from interconnect layers to the heat sink difficult. Hence, even though vias and interconnects which have higher thermal conductivity ($K_{th,Cu}=385\ W/mK$) improve the effective thermal conductivity of the back-end, metal temperature (T_m) rises significantly above the junction temperature especially at the topmost interconnect layers (**Fig. 11(a)**). All these factors adversely affect electromigration lifetime of Cu interconnects which depends quadratically on J and exponentially on T_m.

Estimations based on measured thermal conductivity (K_{th}) of mats of SWCNT bundles, combined with observations from electrical conductivity experiments, predict K_{th} for an SWCNT bundle in the range *1750-5800 W/mK* [20] at room temperature. This high value of $K_{th,CNT}$ is in the direction along the length of nanotubes (since thermal conductivity in CNT bundles is anisotropic [21] - see **Fig. 11(b)**). Hence vias composed of CNT bundles will serve as more effective heat conduits than Cu vias. **Fig. 12** shows that even when CNT bundles are used *only as vias* integrated with Cu interconnects, maximum interconnect temperature rise is much smaller. On-chip vias are prime candidates for the use of CNT bundles [3, 4] as the current carrying limits of small dimension copper interconnects are severely limited [11]. **Fig. 13** compares the resistance of local vias composed of CNT bundles to those made of Cu. It is found that, at all technology nodes, the resistance of CNT bundle vias is nearly equal to that of Cu vias for very high density CNT bundles and for low imperfect metal-nanotube contact resistance. At the 22 nm node, resistance of local vias with CNT bundles remains larger than that of Cu vias even with ideal metal-nanotube contacts and for the maximum possible metallic CNT density. **Fig. 14** shows the performance and reliability impact of using CNT bundles *only as vias* integrated with Cu interconnects. Since Cu resistivity increases with temperature, τ/l for Cu global interconnects at maximum temperature is about 40% higher than $(\tau/l)_{opt}$ calculated at room temperature while it is only about 10% higher if CNT bundle vias are used (**Fig. 14(a)**). **Fig. 14(b)** shows that the lower interconnect temperatures with CNT bundle vias can lead to two orders of magnitude improvement in the electromigration mean-time-to-failure of Cu global interconnects. Finally, **Fig. 15** summarizes a thermal management strategy based on the use of CNT bundles to replace Cu at different locations in the interconnect stack. 3D electrothermal simulation results for the back-end thermal profile employing Cu interconnects, and the variation in this profile as CNT bundle vias are introduced at different locations, are shown. It is found that CNT bundle vias employed at the intermediate level yield the maximum benefit in terms of back-end thermal management as they lead to the largest reduction in interconnect temperatures with least disruption in the Cu interconnect stack. Thus, even while it remains difficult to achieve CNT bundle vias with lower resistance than Cu (Fig. 13), dummy vias composed of CNT bundles can very effectively be used as *thermal vias* to reduce interconnect temperatures.

Conclusions

In conclusion, CNT bundle interconnects have been evaluated against Cu interconnects from performance, power dissipation and thermal management/reliability perspectives. The equivalent circuit parameters for a *CNT bundle* interconnect are explicitly calculated and it is found that CNT bundles can significantly improve the performance of long global interconnects by as much as 80% with minimal additional power dissipation (for maximum metallic CNT density). Moreover, power-optimal repeater insertion methodology can be applied to CNT bundle interconnects (just as with Cu) to save power with a small delay penalty. Most importantly, it is shown that CNT bundle vias can greatly reduce interconnect temperature rise and thus, when integrated with Cu interconnects, tremendously improve Cu interconnect performance (about 30%) and lifetime (by at least two orders of magnitude) due to lower temperatures. The advantage of CNT bundle vias in controlling the back-end temperature, as shown in this work, will also have significant implications for emerging technologies such as 3-D ICs where thermal management is a big concern.

Acknowledgements

This work at UCSB is being supported through a grant from NIST (Award No. 70NANB5H1215). The authors would like to thank Dr. Franz Kreupl of Infineon Technologies, Munich, Germany, for providing useful technical feedback.

References

[1] B. Q. Wei, et al., *App. Phy. Lett.*, Vol. 79, No. 8, pp. 1172-1174, 2001.
[2] P. L. McEuen, et al., *IEEE Trans. Nano.*, Vol. 1, No. 1, pp. 78-85, 2002.
[3] F. Kreupl, et al., *Microelectronic Engineering*, 64, pp. 399-408, 2002.
[4] J. Li, et al., *App. Phy. Lett.*, Vol. 82, No. 15, pp. 2491-2493, 2003.
[5] M. Nihei, et al., *Intl. Interconnect Tech. Conf.*, 2004, pp. 251-253.
[6] A. Raychowdhury and K. Roy, *ICCAD*, 2004, pp. 237-240.
[7] A. Naeemi et al., *IEDM*, 2004, pp. 699-702.
[8] A. Naeemi et al., *EDL*, Vol. 26, No. 2, pp. 84-86, 2005.
[9] A. Naeemi and J. D. Meindl, *EDL*, Vol. 26, No. 8, pp. 544-546, 2005.
[10] Th. Hunger et al., *Physical Review B*, Vol. 69, 195406, 2004.
[11] N. Srivastava and K. Banerjee, *Proc. VMIC*, 2004, pp. 393-398.
[12] P. J. Burke, *IEEE Trans. Nano.*, Vol. 1, No. 3, pp. 129-144, 2002.
[13] S. Datta, *Electronic Transport in Mesoscopic Systems*, Cambridge Univ. Press, 1995.
[14] J.-Y. Park, et al., *Nano Letters*, Vol. 4, No. 3, pp. 517-520, 2004.
[15] Z. Yu and P. J. Burke, *Nano Letters*, Vol. 5, No. 7, pp. 1403-1406, 2005.
[16] H. Stahl et al., *Phys. Rev. Lett.*, Vol. 85, No. 24, pp. 5186-5189, 2000.
[17] N. Srivastava and K. Banerjee, *ICCAD*, 2005, in press.
[18] Intl. Tech. Roadmap for Semiconductors (ITRS), 2004, http://public.itrs.net
[19] K. Banerjee and A. Mehrotra, *TED*, Vol. 49, No. 11, pp. 2001-2007, 2002.
[20] J. Hone et al., *Physical Review B*, Vol. 59, No. 4, R2514, 1999.
[21] J. Hone, et al, *App. Phy. Lett.*, Vol. 77, No. 5, pp. 666-668, 2000.

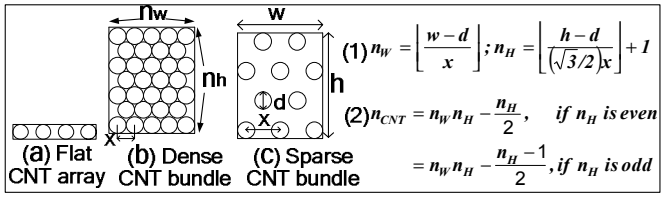

Fig 1: Cross-section schematic of CNTs forming **(a)** flat array, **(b)** dense bundle (x=d) and **(c)** sparse bundle (x>d). Eqs. (1, 2): No. of metallic CNTs (n_{CNT}) in bundle, assuming uniform distribution. n_W and n_H : number of metallic CNTs along interconnect width and height respectively.

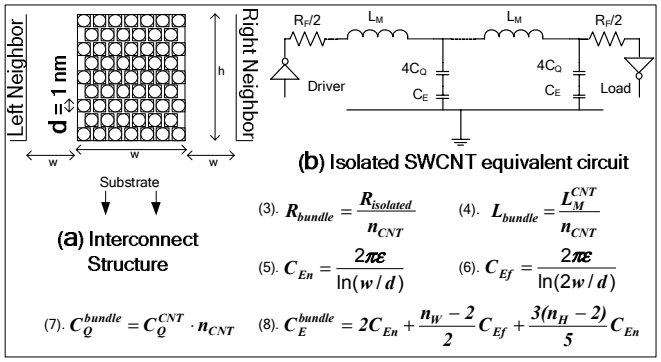

Fig 2: (a) Interconnect structure and **(b)** Equivalent circuit [12] for isolated SWCNT. Equations 3-8: equivalent circuit parameters for a CNT bundle interconnect. C_{En} and C_{Ef} are parallel plate capacitances of isolated CNT with respect to near and far neighboring interconnects respectively [17].

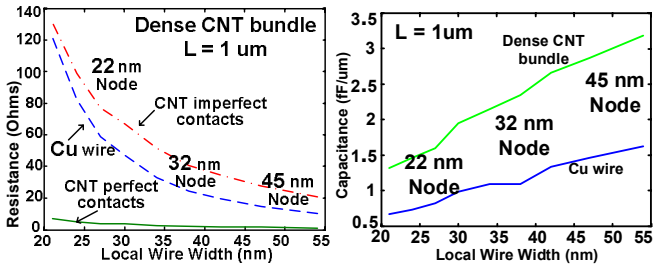

Fig 3: (a) Resistance and **(b)** Capacitance of 1µm long local interconnect using densely packed CNT bundles compared with Cu at different technology nodes (assuming λ_{CNT} = *1 µm* [3]).

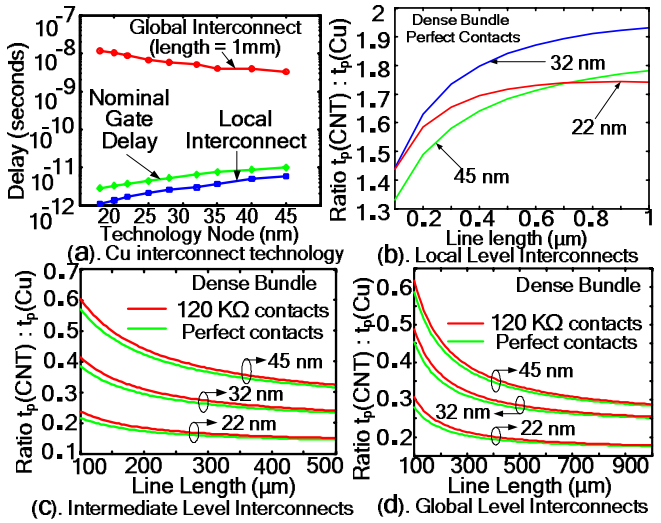

Fig 4: (a) Typical delay for global and local Cu interconnects compared to nominal gate delay [18], as technology scales. **(b-d)** Ratio of CNT bundle interconnect delay (assuming λ_{CNT} = *1 µm* [3]) to that of Cu interconnect of same dimensions at **(b)** local, **(c)** intermediate and **(d)** global levels.

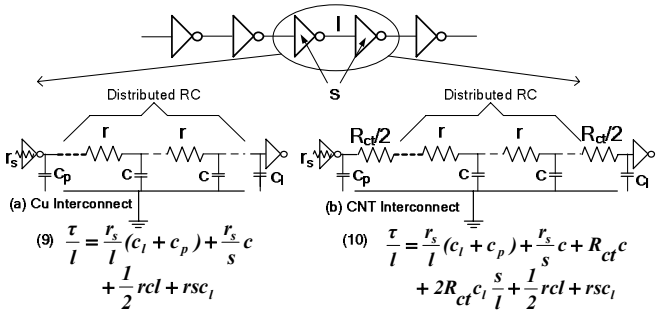

Fig 5: Delay per unit length for optimally buffered global interconnects made of **(a)** Cu and **(b)** CNT bundle. For CNT interconnects, c includes C_E and C_Q (Fig. 2(b)) and r is per unit length resistance $(h/4e^2)/\lambda_{CNT}$.

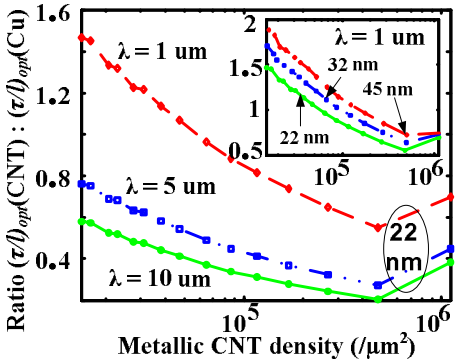

Fig 6: Ratio of optimal delay per unit length for buffered CNT bundle global interconnect to that of Cu, as a function of density of metallic CNTs in the bundle: for different mean free paths (λ) at 22 nm node (main fig.) and at different technology nodes (inset).

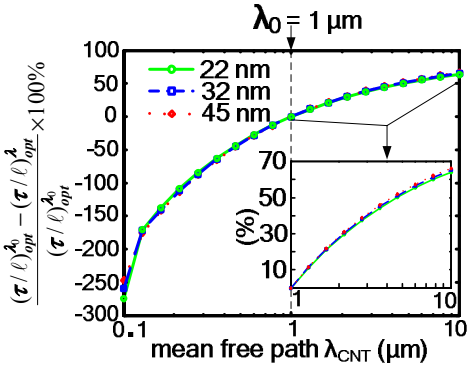

Fig 7: Percent improvement in optimal delay per unit length $(\tau/l)_{opt}$ for buffered CNT bundle global interconnect compared to $(\tau/l)_{opt}$ for the typical value of λ_0 = *1 µm*, as λ_{CNT} is varied. Maximum metallic CNT density (dense bundle) is assumed.

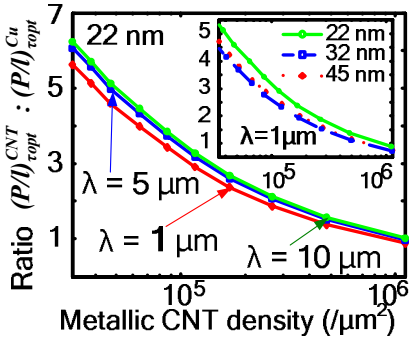

Fig 8: Ratio of repeater power dissipation per unit length for optimally buffered CNT-bundle global interconnect to that of Cu, as a function of metallic CNT density and for different mean free path lengths (λ_{CNT}). Inset: same ratio evaluated at different technology nodes for λ_{CNT} = *1 um*.

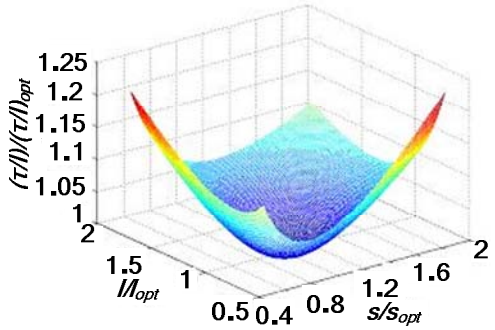

Fig 9: Delay per unit length in sub-optimally buffered CNT bundle global interconnect normalized to optimal delay per unit length, as a function of sub-optimal repeater size (s) and inter-repeater separation (l).

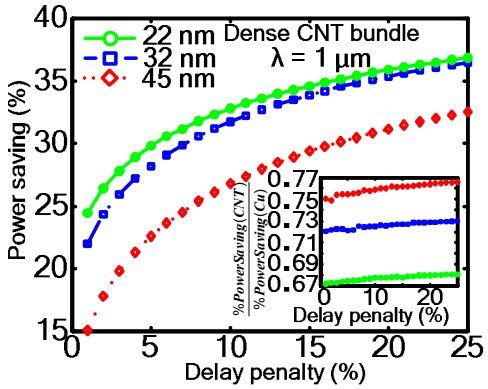

Fig. 10: Percent saving in repeater power dissipation per unit length for "power-optimal" buffer inserted CNT-bundle global interconnect as a function of delay penalty (% of $(\tau/l)_{opt}$). Inset: Ratio of % power saving with CNT bundle to that with Cu.

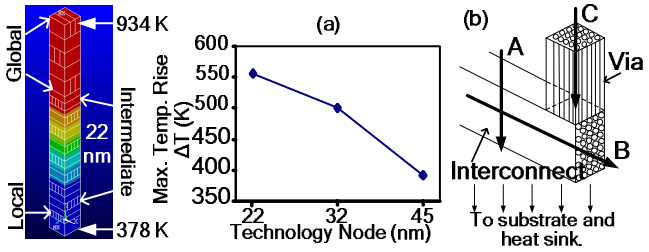

Fig 11: (a) Maximum Cu interconnect temperature rise from 3-D finite element electrothermal simulations, with ITRS prescribed geometries and current at all layers, layer-dependent via densities, duty ratios obtained from SPICE simulations and junction temperature of 378 K. Temperature contour plot for the multi-layer interconnect stack at the 22 nm node is shown on the left. **(b)** Schematic showing heat conduction paths in interconnects and vias. For CNT bundles $K_{th,CNT}$ > 1750 W/mK along 'B' and 'C' and 20 times lower along 'A' (anisotropic $K_{th,CNT}$ [21]).

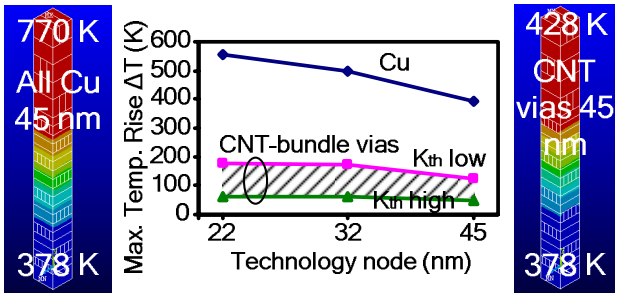

Fig. 12: Maximum interconnect temperature rise for Cu interconnects and vias vs CNT bundle vias integrated with Cu interconnects. For CNT bundles, the shaded region shows the range 1750 W/mK < K_{th} < 5800 W/mK [20]. Reference (substrate) temperature = 378 K.

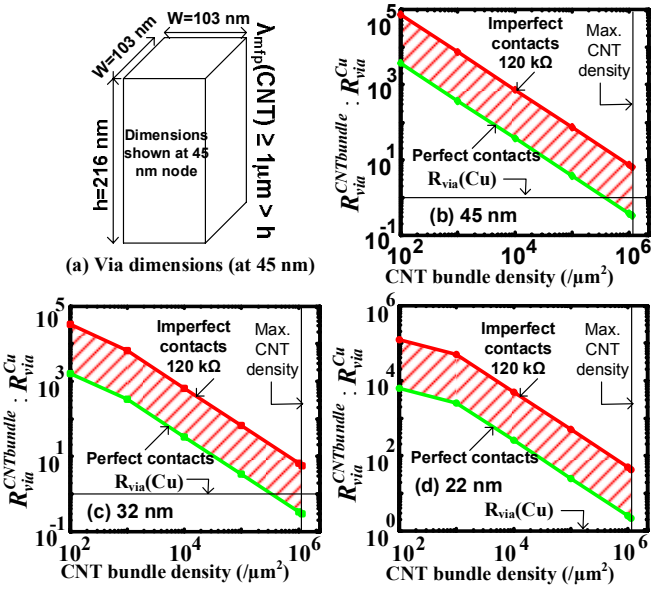

Fig 13: (a) Via dimensions at 45 nm node (via height < λ_{CNT}). **(b-d)** Ratio of CNT bundle via resistance to Cu via resistance as a function of CNT bundle density, at **(b)** 45 nm, **(c)** 32 nm, **(d)** 22 nm nodes. Shaded region: imperfect metal-nanotube contact resistance $0<R_{ct}<120\ K\Omega$ [10].

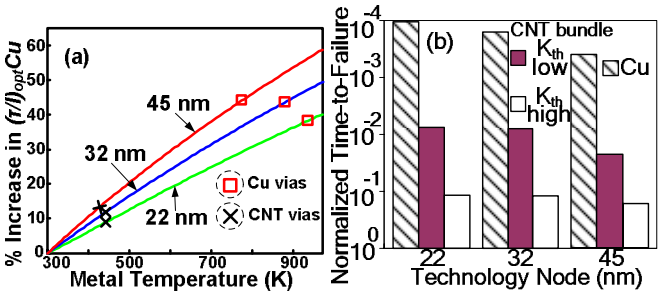

Fig 14: (a) Increase in $(\tau/l)_{opt}$ for Cu global interconnects calculated at different metal temperatures as a % of $(\tau/l)_{opt}$ at room temperature. Squares correspond to *actual interconnect temperature* with Cu vias. Crosses - *actual interconnect temperature* with CNT vias (K_{th} = 5800 W/mK). **(b)** Cu interconnect electromigration (EM) lifetime as a result of high interconnect temperature normalized to EM lifetime at reference temperature (378 K), when vias are composed of Cu and CNT bundles respectively.

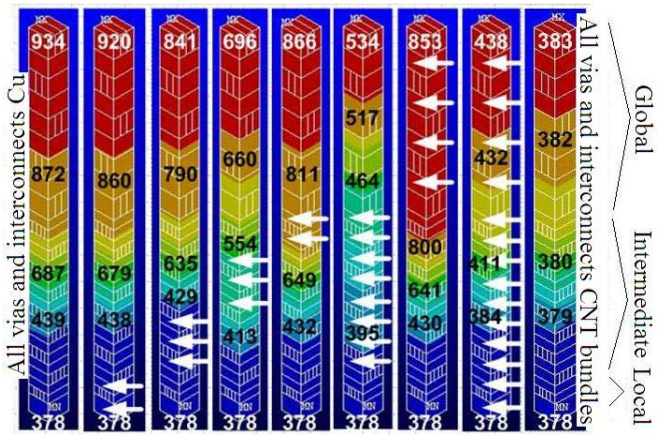

Fig 15: Interconnect temperature contour plots showing interconnect temperature (K) at different layers from 3D finite element electrothermal simulations (22 nm node). Leftmost: all interconnects and vias are Cu. Rightmost: all interconnects and vias are CNT bundles. In remaining images, each arrow points to a particular interconnect level where the vias are assumed to be composed of CNT bundles, while all other vias and all interconnects are made of Cu.