System Level Performance Analysis of Carbon Nanotube Global Interconnects for Emerging Chip Multiprocessors

Sudeep Pasricha, Fadi Kurdahi, Nikil Dutt
Center for Embedded Computer Systems
University of California at Irvine, Irvine, CA 92697
{sudeep, kurdahi, dutt}@cecs.uci.edu

***Abstract* – Although carbon nanotubes (CNTs) have been widely proposed as interconnect fabrics for future ultra deep submicron (UDSM) technologies, there is a lack of system-level performance analysis using these interconnects. In this paper, we investigate the performance of four CNT alternatives that may replace conventional copper (Cu) interconnects at the global interconnect level – *(i)* single walled CNTs (SWCNTs), *(ii)* SWCNT bundles, *(iii)* multi-walled CNTs (MWCNTs) and *(iv)* bundles of mixed SWCNTs/MWCNTs. Detailed RLC equivalent circuit models for conventional CNT interconnects are described and used to calculate propagation delays. These models are then incorporated into a system-level environment to estimate the impact of using CNT global interconnects on the overall performance of several multi-core chip multiprocessor (CMP) applications. Our results indicate that MWCNTs can provide the most significant performance speedup among the CNT alternatives (up to 1.9×) over Cu global interconnects. With further improvements in CNT fabrication technology, it is shown that mixed SWCNT/MWCNT bundles and SWCNT bundles can also become viable global interconnect alternatives.**

***Index Terms* – carbon nanotube, interconnect, CMP, UDSM**

I. INTRODUCTION

With the advent of giga- and tera-scale electronic systems, interconnect design has become a fundamental roadblock in realizing emerging high performance multi-core chip multiprocessors (CMPs) that have tens to hundreds of components integrated on a single chip [1]. Interconnects used in CMPs can be classified into two categories: *(i)* local interconnects, that are used for short distance communication, and have a delay of less than a clock cycle, and *(ii)* global interconnects, that are used for long distance communication to distribute data, clock, power supply and ground across the chip, and have a delay spanning multiple clock cycles. According to the International Roadmap for Semiconductors (ITRS) [4], global interconnect performance has become one of the semiconductor industry's topmost challenges. Conventional copper (Cu) global interconnects have become increasingly susceptible to electromigration at high current densities (>10^6 A/cm^2) leading to considerable degradation in reliability [2]. Additionally, as interconnect dimensions are scaled down, rising crosstalk coupling noise and parasitic resistivity due to electron-surface and grain-boundary scatterings cause global interconnect delay to increase rapidly [3]. There is thus a critical need to investigate innovative global interconnect alternatives to Cu for future ultra deep submicron (UDSM) technologies.

Carbon nanotubes (CNTs) have recently been studied as a possible replacement for Cu interconnects in future technologies, because of their remarkable conductive, mechanical and thermal properties [5]-[8]. Depending on the direction in which they are rolled (called *chirality*), CNTs can behave either as a semiconductor or a conductor. Conducting (or metallic) CNTs possess many extraordinary properties that make them promising candidates for interconnects in UDSM technologies. Due to their covalently bonded structure, they are highly resistant to electromigration and other sources of physical breakdown [5]. They can support very high current densities with very little performance degradation. For instance, it was shown in [15] that the current carrying capacity of CNTs did not degrade even after 350 hours at current densities of ~10^{10} A/cm^2 at 250 ºC. CNTs possess high thermal conductivity in the range of 1700–3000 W/m·K [16]. They also have much better conductivity properties than Cu owing to longer electron mean free path lengths (MFP) in the micrometer range, compared to nanometer range MFP lengths for Cu [7].

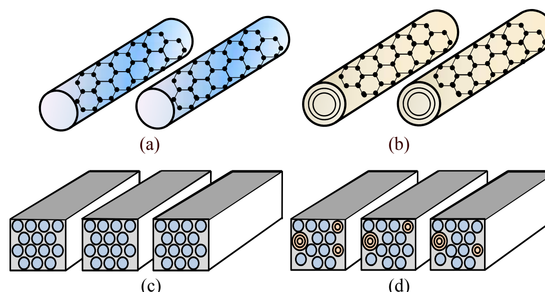


Figure 1. Carbon nanotube (CNT) global interconnect alternatives: (a) single-walled CNT (SWCNT), (b) multi-walled CNT (MWCNT), (c) SWCNT bundle, (d) mixed SWCNT/ MWCNT bundle

While there has been a lot of interest in CNT-based interconnects of late, there are unfortunately few studies that compare CNT interconnects with Cu at the system-level. Indeed, with the exploding number of cores in multi-core CMP applications, it becomes critical for system designers to evaluate the impact of deploying emerging interconnect fabrics, and comparing them with traditional Cu interconnects. Such an evaluation not only enables system designers to

978-1-4244-2553-2/08/$20.00 ©2008 IEEE

understand potential benefits and limitations of CNT interconnects, but also get a true insight into realistic gains that can be achieved by switching to CNT interconnects for future CMP applications. This paper addresses such a need.

CNTs can be broadly classified as single-walled carbon nanotubes (SWCNTs) [9]-[11] and multi-walled carbon nanotubes (MWCNTs) [12]-[13]. SWCNTs consist of a single sheet of graphene rolled into a cylindrical tube, with a diameter in the nanometer range (Fig. 1 (a)). MWCNTs consist of two or more SWCNTs concentrically wrapped around each other, with diameters ranging from a few to several hundred nanometers (Fig. 1 (b)). In addition to SWCNTs and MWCNTs, bundles of CNTs are being evaluated because of their superior conductivity properties. A SWCNT bundle (Fig. 1(c)) consists of several SWCNTs packed together in parallel [14], whereas a mixed SWCNT/MWCNT bundle (Fig. 1(d)) consists of a combination of SWCNTs and MWCNTs packed together in parallel [13].

In this paper, we present a system-level performance analysis of conventional Cu global interconnects with the four CNT global interconnect alternatives shown in Fig. 1 – single walled CNTs (SWCNTs), multi-walled CNTs (MWCNTs), SWCNT bundles, and mixed SWCNT/MWCNT bundles. Our paper makes two novel contributions: *(i) investigating the performance of the recently proposed mixed SWCNT/ MWCNT bundles,* and *(ii) analyzing the overall performance impact of using CNT-based global interconnects at the system level, for emerging multi-core CMP applications.* We describe detailed RLC equivalent circuit models for CNT interconnects and use them to calculate propagation delays. The CNT equivalent circuit models capture the statistical distribution of metallic nanotubes while accurately incorporating recent experimental and theoretical results on inductance, crosstalk capacitance and ohmic resistance. These models are then incorporated into a system-level environment to estimate the impact of using CNT global interconnects on the overall performance of several chip multi-processor (CMP) applications with diverse data traffic profiles. Our results indicate that MWCNTs provide the most significant performance speedup among the CNT alternatives (up to 1.9×) over Cu global interconnects. With further improvements in CNT fabrication technology, it is shown that mixed SWCNT/MWCNT bundles and SWCNT bundles can also become viable global interconnect alternatives.

II. Related Work

The study of the properties of CNTs has received immense interest in the last few years. Researchers have developed circuit models for various CNT interconnect alternatives and compared their performance with Cu interconnects. RLC circuit models have been developed for isolated SWCNTs [9]-[11], SWCNT bundles [8][12][14] and MWCNTs [12]-[13]. Recently, conductance and inductance models for mixed SWCNT/MWCNT bundles were also introduced [13][40]. Other studies have presented some interesting discussions on the impact of process variations on CNT performance [18] and the possibility of CNTs replacing Cu interconnects in future FPGA fabrics [19]. All of these studies have shown that CNT interconnects can be a viable alternative to Cu interconnects in future technologies.

To date however, very little work has been done to understand the impact of CNT global interconnect architectures for emerging CMP applications. Early work in [9] compared the delay of a SWCNT global interconnect with that of a Cu global interconnect, and concluded that SWCNTs are not able to outperform Cu interconnects. Recently, [17] described a dual-walled CNT global bus interconnect and presented experiments to show that it outperforms SWCNT and MWCNT global bus interconnects, but no comparison with Cu interconnects is presented. Our work extends the direction of [17] by presenting a comprehensive comparison between Cu global interconnects and several CNT global interconnect alternatives – SWCNTs, SWCNT bundles, MWCNTs, and mixed SWCNT/MWCNT bundles. To the best of our knowledge, ours is the first work that analyzes the system-level performance impact of various CNT global interconnects, for multi-core CMP applications.

In the next section (Section III), we describe equivalent RLC circuit models for various CNT interconnects. These circuit models are then used to calculate interconnect propagation delays and drive our system-level performance analysis in Section IV.

III. CNT Interconnect Circuit Models

In this section, we present an overview of equivalent RLC circuit models derived from literature for the different CNT interconnects.

A. SWCNT

Unlike Cu, the resistance of a SWCNT is not simply a linear function of its length. Instead, the SWCNT resistance is constant up to a certain length, called *mean free path* length (λ) and then increases linearly thereafter. The constant (quantum) resistance of a SWCNT can be determined using the Landauer-Buttker formula, $R_Q=h/4e^2$ [5] and is about 6.45kΩ. For longer lengths, SWCNT resistance has been shown to depend on its length and bias voltage [11]. For bias voltage less than the critical bias (< 0.16V for global wires [22]), the SWCNT resistance is:

$$R_{SWCNT} = R = \frac{h}{4e^2}, \quad l \leq \lambda$$
$$= \frac{h}{4e^2}\left(\frac{l}{\lambda}\right), \quad l > \lambda \qquad (1)$$

where h is Planck's constant, e is the charge of an electron, l is the SWCNT length and λ is the mean free path (MFP) length, given by $\lambda = v_F d_t/\alpha T$, ($\alpha$ is the total scattering rate, v_F is the Fermi velocity of graphene, d_t is the SWCNT diameter and T is the temperature, assumed to be room temperature $T=300K$, with a more detailed temperature analysis presented in [23]). Ideally, λ is in the order of several µm. However, a rigorous analysis has shown that practically, λ is in the order of 1 µm [10][22].

For SWCNT interconnects, three types of capacitance should be considered: an electrostatic capacitance with the

ground (C_{EG}), a coupling capacitance with the adjacent SWCNT (C_{EC}), and an intrinsic quantum capacitance (C_Q) [6][24]. The electrostatic capacitance per unit length between a SWCNT and the ground plane is given by:

$$C_{EG} = \frac{2\pi\epsilon}{cosh^{-1}\left(\frac{2H}{d_t}\right)} \tag{2}$$

where ε is the permittivity of the dielectric and H is the distance between the SWCNT and ground plane. For a typical value of H/d_t, $C_{EG} \approx$ 50aF/µm [6]. Similarly, the electrostatic capacitance per unit length between two parallel SWCNTs is given by:

$$C_{EC} = \frac{\pi\epsilon}{cosh^{-1}\left(\frac{s}{d_t}\right)} \tag{3}$$

where s is the inter-SWCNT spacing and d_t is the SWCNT diameter. The SWCNT quantum capacitance per unit length is given by:

$$C_Q = \frac{2e^2}{hv_F} \tag{4}$$

where the value of $C_Q \approx 100$ aF/µm. SWCNTs have been shown to consist of four co-propagating quantum channels, and therefore the effective SWCNT quantum capacitance is $4C_Q$ [6]. As per the analysis in [24], $4C_Q$ is in series with a parallel combination of C_{EG} and C_{EC}. Because electrostatic coupling between non-adjacent SWCNTs is very weak compared to coupling between adjacent SWCNTs, the coupling capacitance between nonadjacent SWCNTs can be neglected in parallel global interconnects with more than two lines [11][17].

A SWCNT also has two types of inductance associated with it – kinetic and magnetic. The kinetic inductance (L_K) is due to charge carrier inertia, since electrons do not instantaneously react to an applied electric field. The series inductance which represents this phenomenon is given by [6]:

$$L_K = \frac{h}{2e^2 v_F} \tag{5}$$

Since a nanotube has four co-propagating quantum channels, the effective value of kinetic inductance in the equivalent circuit is $L_K/4$ [6][17]. Kinetic inductance ($L_K/4 \approx$ 4nH/µm) typically dominates magnetic inductance (~ 1.2pH/µm) in SWCNTs [6].

B. SWCNT Bundles

A SWCNT bundle consists of several individual SWCNTs in parallel. An important parameter associated with a SWCNT bundle is its metallic density (P_m) which refers to the probability that a SWCNT in the bundle is metallic (i.e. conducting). The value of $P_m \approx 1/3$ [1] with today's best fabrication techniques, which implies that only 1/3 of the SWCNTs in a bundle are conducting. The effective spacing x between SWCNTs in a bundle is given by $x = d_t/\sqrt{P_m}$ [8][22], where d_t is the diameter of a SWCNT in the bundle. The number of conducting SWCNTs in a bundle (n_{CNT}) of width w and height h is then given as:

$$n_w = \frac{w - d_t}{x}, n_h = \frac{h - d_t}{(\sqrt{3}/2)x} + 1 \tag{6}$$

$$\begin{aligned} n_{CNT} &= n_w n_h - \frac{n_h}{2} && if\ n_h\ is\ even \\ &= n_w n_h - \frac{n_h - 1}{2} && if\ n_h\ is\ odd \end{aligned}$$

where n_w and n_h are the number of SWCNTs in a row and along the height of the bundle, respectively.

The resistance of a SWCNT bundle is simply the parallel combination of n_{CNT} metallic SWCNTs [8][22].

$$R_{SWCNT\ bundle} = \frac{R_{SWCNT}}{n_{CNT}} \tag{7}$$

As far as SWCNT bundle capacitance is concerned, the electrostatic capacitance to the ground and coupling capacitances arise mainly from the SWCNTs lying at the edges of the SWCNT bundle. The coupling and electrostatic capacitances to the ground of a highly metallic SWCNT bundle (P_m=*1*) have been analyzed extensively in [10] (using the field solver RAPHAEL) and [22] (using the 3D field solver FastCap), and found to be equal to the respective capacitances of a Cu wire with the same cross-sectional dimensions. For more realistic SWCNT metallic densities (P_m=*1/3*), capacitance was shown to decrease very slowly. Even for as few as 4 SWCNTs at the corners, capacitance was found to be only 20% smaller (mainly due to fringing effect) than for a densely packed bundle [10]. It is therefore safe to assume that the SWCNT bundle electrostatic capacitance to ground and coupling capacitances are the same as that of a Cu wire with the same cross-sectional dimensions for realistic metallic densities. The effective quantum capacitance of a SWCNT bundle is further reduced in a bundle and found to be negligible compared to its electrostatic counterparts [22].

The kinetic inductance of a SWCNT bundle is the parallel combination of individual SWCNT kinetic inductances. For a SWCNT bundle with wire dimensions corresponding to the 22-nm technology node and a SWCNT diameter of 1nm, the kinetic inductance is ~6pH/µm. The magnetic inductance remains relatively constant at ~1.6pH/µm with wire dimensions. The mutual inductance between SWCNTs in a bundle is accounted for using the partial element equivalent circuit (PEEC) model [22][25]. The total SWCNT bundle inductance is then given by:

$$L_{SWCNT\ bundle} = \left(\frac{L_k}{4n_{CNT}} + L_m\right).l \tag{8}$$

C. MWCNT

A MWCNT consists of two or more concentric SWCNTs. Since the basic building blocks of a MWCNT are essentially SWCNTs of varying diameters, many of the properties of SWCNTs hold for MWCNTs. The number of shells (N_s) in a MWCNT is diameter dependent, and given by:

$$N_s = 1 + \frac{D_{outer} - D_{inner}}{2\delta} \tag{9}$$

where δ=*0.34nm* (van der Waals distance) is the spacing between adjacent concentric shells, and D_{outer} and D_{inner} are the maximum and minimum shell diameters. The ratio of D_{outer}/D_{inner} has been observed to vary from 0.35 to 0.8 [12].

The approximate number of conduction channels per shell for a MWCNT is given by [13]:

$$N_{chan/shell}(d) \approx \begin{cases} (ad+b)P_m, & d > 6\,nm \\ 2P_m, & d < 6\,nm \end{cases} \tag{10}$$

where a = 0.1836 nm^{-1}, b = 1.275, d is the shell diameter and P_m = 1/3 (similar to a SWCNT bundle). Then the resistance for the ith SWCNT shell with diameter d_i is given by:

$$R_{SWCNTi}(d_i, l) = \frac{R_{SWCNT}}{N_{chan/shell}(d_i)} \quad (11)$$

Each shell has its own d_i, λ and $N_{chan/shell}$ that can be derived from D_{outer}. The total MWCNT resistance (R_{MWCNT}) is a parallel combination of the resistances of all the concentric SWCNTs:

$$(R_{MWCNT}(D_{outer}, l))^{-1} = \sum_{N_s}(R_{SWCNTi}(d_i, l))^{-1} \quad (12)$$

where R_{SWCNTi} is the resistance of the ith concentric SWCNT. Additionally, an inter-shell resistance (R_i ≈10 kΩ/μm) must be considered to account for the inter-shell tunnel transport phenomenon [17].

The metallic shell in a MWCNT constitutes an effective electrostatic shield for its inner shells [26]. Thus the capacitance between an internal shell and ground, and between non-adjacent shells can be safely neglected. The electrostatic capacitance per unit length between the outermost nanotube in a MWCNT and the ground is given by (2), where d_t is the diameter of the outermost shell. Similarly, the coupling capacitance between the outer shells of adjacent MWCNTs is given by (3) and the quantum capacitance is given by (4). The electrostatic coupling capacitance between adjacent shells in a MWCNT (C_{ESC}) is derived from a conventional metallic coaxial configuration [27]:

$$C_{ESC} = \frac{2\pi\epsilon}{ln(d_1/d_2)} = \frac{2\pi\epsilon}{ln(d_1/(d_1 - 2\delta))} \quad (13)$$

The quantum capacitance ($4C_Q$) of an external shell is coupled with a parallel combination of the three capacitances C_{ESC}, C_{EG} and C_{EC} in the equivalent RLC circuit.

Finally, the total MWCNT inductance is given by a relation similar to that for a SWCNT bundle (8):

$$L_{MWCNT} = \left(\frac{L_k}{2.\sum_{N_s} N_{chan/shell}(d_i)} + L_m\right).l \quad (14)$$

where L_k is obtained from (5) and $N_{chan/shell}(d_i)$ from (10). L_m is calculated using the equivalent conductivity method [40].

D. Mixed SWCNT/MWCNT Bundles

A mixed SWCNT/MWCNT bundle consists of SWCNTs with a diameter d and MWCNTs with various diameters $D_{inner} \le d_i \le D_{outer}$. It has been shown [2][13] that the outer diameters follow a normal (Gaussian) distribution.

The total resistance for a mixed SWCNT/MWCNT bundle is given by [13]:

$$R_{mixed\ bundle} = \left(\int \frac{N(D_{outer})\,\partial D_{outer}}{R_{MWCNT}(D_{outer}, l)}\right)^{-1} \quad (15)$$

where $R_{MWCNT}(D_{outer}, l)$ is obtained from (12) and $N(D_{outer})$ is the tube count according to D_{outer}'s, with a normal (Gaussian) distribution, a mean diameter $\overline{D_{outer}}$, and a standard deviation $\sigma_{D_{outer}}$ [2][13].

For N_{bundle} CNTs in the bundle, the tube count for a given D_{outer} is:

$$N(D_{outer}) = \frac{N_{bundle}}{\sigma_{D_{outer}}\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{D_{outer} - \overline{D_{outer}}}{\sigma_{D_{outer}}}\right)^2\right] \quad (16)$$

This relation can be used to derive a distribution curve for the tube count. The resulting curve and the corresponding MWCNT resistance curve can be used to determine total resistance of the mixed bundle from (15). This resistance formulation has been validated with experimental results from [2] in [13].

The capacitive characteristics of a mixed SWCNT/ MWCNT bundle have been shown to predominantly be determined by the cross sectional dimensions of a bundle [1]. Therefore, similar to the case of the SWCNT bundle, we can assume that the mixed bundle electrostatic capacitance to ground and coupling capacitances are the same as that of a Cu wire with the same cross-sectional dimensions for practical metallic densities (P_m=1/3). The quantum capacitance is similarly negligible compared to its electrostatic counterparts [22].

Finally, the total kinetic inductance of a mixed bundle is the parallel inductance value of all the conduction channels in the bundle, similar to (8) and (14). Magnetic inductance L_m is calculated using the equivalent conductivity method [40].

IV. EXPERIMENTS

Having presented the equivalent circuit models for CNT interconnects in the previous section, we now use these models to evaluate and compare their propagation delays with Cu. Subsequently, the performance of CNT and Cu global interconnects is compared for several multi-core CMP applications. Finally, we investigate how CNT global interconnects will perform if advances in fabrication technology lead to an increase in mean free path (MFP) length and metallic densities in CNTs.

A. CNT vs. Cu Global Interconnect Delay Comparison

In our first experiment we compare the global wire delay of Cu and CNT interconnect alternatives across the 45-22 nm process technology nodes. We use the equivalent circuit models described in the previous section to determine wire delay for CNTs. The equivalent circuit model for Cu wires is obtained from [28]-[30], and used to derive the wire delay for Cu. We consider optimal repeater sizing and insertion for both Cu and CNT wires, using the formulations presented in [31]. The node driver resistance, load capacitance and process parameters are obtained from ITRS specifications [4]. Since global wire width is typically much larger than minimum wire width (W_{min}) to improve delay and bandwidth characteristics, we consider an aspect ratio of 1 and use a global wire width value of $5W_{min}$ which is shown to optimize the power-delay product for copper wires [32]. This wire width is in the shallow RLC region, where the difference between RC and RLC model latencies is only 10% [10]. For comparison purposes, we use the same width and aspect ratios for the SWCNT bundles and mixed SWCNT/MWCNT bundles. The MWCNT outer diameter is assumed to be equal to the wire width, and the ratio between outer to inner diameters is assumed to be 0.5. For the mixed SWCNT/MWCNT bundle, we assume $\overline{D_{outer}}$ = 4.2nm and $\sigma_{D_{outer}}$ = 1.25nm [2]. The

SWCNT diameter is assumed to be 1nm. The spacing between adjacent wires is assumed to be the same as the wire width ($5W_{min}$), including for the SWCNT case. The CNT mean free path λ is assumed to be 1µm and metallic density P_m is assumed to be 1/3, while Cu mean free path is assumed to be 40 nm. All of these values are practically achievable by using current fabrication techniques.

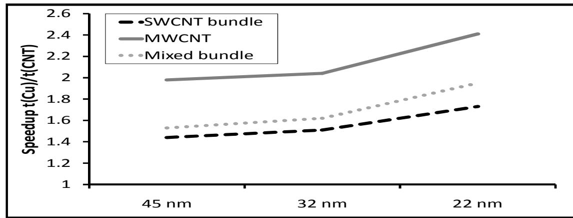


Figure 2. Global Interconnect Delay Comparison between copper and various CNT alternatives (45-22 nm)

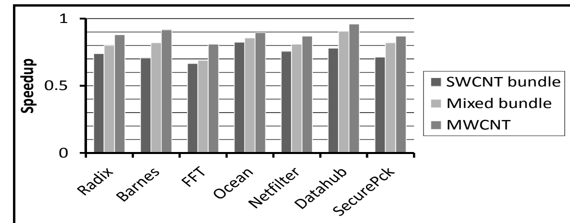
TABLE I. CMP APPLICATION IMPLEMENTATION CHARACTERISTICS

CMP applications	Description	cores	prog. processors	clusters	global links
Radix	Integer radix sort	18	4	3	10
Barnes	Evolution of galaxies	26	6	4	18
FFT	FFT kernel	28	6	4	12
Ocean	Ocean movements	35	10	5	24
Netfilter	Packet processing and forwarding	49	22	6	32
Datahub		68	26	8	34
SecurePck		94	30	8	28

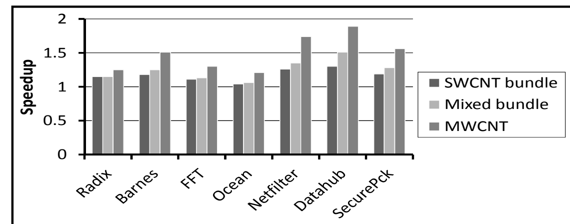
Fig. 2 shows the ratio of propagation delay of copper *t(Cu)* and the propagation delay of CNTs *t(CNT)*, for global wires (> 1mm in length) across the 45-22 nm process technology nodes. The single SWCNT has been omitted from the analysis and the figure because of its much higher propagation delay (almost 100× more) compared to Cu. This large SWCNT propagation delay is due to its very high resistance, because of its extremely small cross-section area. Consequently single SWCNTs are unsuitable as global interconnects. They will however most likely replace Cu at the local interconnect level because of their much lower lateral capacitance which improves latency for short distances [10]. It can be seen from the figure that all the other CNT alternatives have lower interconnect delays compared to Cu. The MWCNT wire in particular has a much lower delay due to its large concentric tubes having greater mean free path lengths which improves conductivity significantly over Cu. A mixed SWCNT/ MWCNT bundle has many large MWCNTs with several shells and more conduction channels than SWCNT bundles of the same dimensions. This makes mixed bundles better conductors than SWCNT bundles.

B. *CNT Global Interconnect Bus Performance Analysis*

Next, we select several multi-core CMP applications to analyze the overall performance impact of using CNT-based global interconnect buses at the system-level. These applications are selected from the well known SPLASH-2 benchmark suite [33] (*Barnes*, *Ocean*, *FFT*, *Radix*). We also select applications from the networking domains (proprietary benchmarks *Netfilter*, *Datahub* and *SecurePck*). The applications have been parallelized and implemented on multiple cores. The CMP applications described above were modeled in SystemC [34] at the transaction-based bus cycle accurate (T-BCA) modeling abstraction that enables fast communication architecture exploration [35]-[36]. The entire chip is assumed to be partitioned into clusters (or islands) of computation cores. Each cluster consists of tightly coupled cores (processors, memories etc.) optimized for dedicated tasks (e.g. packet encryption, image processing etc.) and interconnected via local bus-based Cu links that support high data bandwidths. Global pipelined links (shared bus or point-to-point) are used to connect computation clusters with each other, and facilitate inter-cluster data transfers. Two clusters can be interconnected by multiple global links if higher inter-cluster bandwidth needs to be supported. Table I summarizes the implementation details of the CMP applications, such as number of cores (including memories, peripherals, and processors), programmable processors, computation clusters and inter-cluster global links on the chip.



(a)



(b)

Figure 3. Performance speedup when using CNT-based global interconnect buses instead of conventional Cu global interconnect buses, assuming (a) imperfect metal-CNT contacts, (b) perfect metal-CNT contacts

Our analysis is performed for the 22-nm process technology node (as predicted for the 2016 node of ITRS 2005 [4]) and the interconnect fabric is clocked at frequencies ranging from 1 – 10 GHz, to support data transfer rates in the hundreds of Gb/s range for future high performance systems [18]. The die size is assumed to be 2cm×2cm. A high level simulated annealing floorplanner based on sequence pair representation (PARQUET [38]) is used to create an early layout of the CMP application on the die, and Manhattan distance based wire routing estimates are used to determine wire lengths. For the global interconnect buses, in addition to repeaters, latches are inserted based on wire length (obtained from routing estimates), wire delay and clock frequency of the bus, to pipeline the interconnect and ensure correct operation [39]. For instance, a global Cu wire of length 10 mm has a projected signal delay of 2 ns in the 22-nm technology node, for a $5W_{min}$ wire width. To support a frequency of 10 GHz

(clock period of 0.1 ns), approximately 20 latches need to be inserted to ensure correct multi-cycle operation.

Fig. 3 (a)-(b) show the performance improvement (speedup in application execution time) when using MWCNT, SWCNT bundle, and mixed SWCNT/MWCNT bundle global interconnect buses, instead of Cu global interconnect buses, for the 22-nm technology node, with the interconnect fabric clocked at 10 GHz. Experiments with the interconnect fabric clocked at lower frequencies within the 1 – 10 GHz range yielded speedup results within 5% of the results shown in the figures (assuming computation core frequencies are scaled down by the same factor as the interconnect fabric), and hence the other results are not presented for brevity.

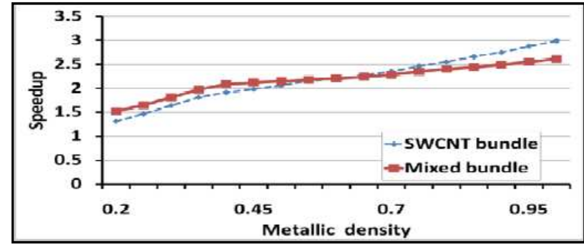
Fig. 3 (a) shows the results for the case when imperfect metal-CNT contacts are assumed. A high metal-CNT contact resistance of 100 KΩ is added to account for the excessive electron scattering at imperfect metal-CNT junctions. It can be seen that the CNT global interconnect buses perform worse than Cu global interconnect buses, and this is primarily an artifact of using poor metal-CNT contacts, typically constructed using Gold or Palladium or Rhodium.

Many recent studies [20]-[21] with state-of-the-art fabrication techniques have however managed to reduce this contact resistance down to a very small value (a few hundred Ω). This has dramatically improved the viability of successful integration of CNTs with CMOS technology. Fig. 3 (b) shows the performance improvement under the assumption of perfect contacts with negligible contact resistance. The CNT global interconnect buses achieve speedup over Cu global interconnect buses because of their lower interconnect delays, which leads to more widely spaced and hence fewer pipeline latches on global inter-cluster buses. This results in global data transfers taking fewer clock cycles and improved CNT performance. Among the applications, *Ocean* can be seen to have a much smaller speedup because it has fewer inter-cluster data transfers, which reduces the advantage of having faster CNT-based buses. On the other hand, the speedup for *Radix* and *FFT* is lower than other applications due to the smaller length of their global interconnect buses, which reduces the impact of faster CNT-based buses on overall performance. Overall, the results indicate that MWCNT global buses provide the maximum speedup (up to 1.9×). SWCNT bundles and mixed SWCNT/MWCNT bundles achieve speedups of up to 1.3× and 1.5× respectively.

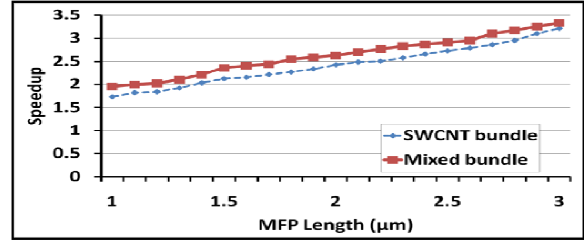
At first glance, these modest performance improvements over Cu interconnects may appear to be not so significant as to justify a migration to CNT global interconnects. However in the next section we present results that indicate more significant CNT performance gains with inevitable improvements in CNT fabrication technology.

C. Impact of MFP and CNT Metallic Density

Our system-level experiments above assumed practical values for CNT mean free path or MFP (λ=1µm) and metallic density (P_m=1/3) while calculating performance gains. The modest CNT speedup obtained above can be improved if breakthroughs in fabrication technology in the future allow for longer MFP lengths (λ) and greater metallic density (P_m) for CNT bundles.



(a)



(b)

Figure 4. SWCNT bundle and mixed bundle delay speedup over copper for varying (a) metallic densities P_m (x-axis), (b) MFP length λ (x-axis)

Fig. 4 (a)-(b) show the interconnect delay speedup over Cu for SWCNT bundles and mixed SWCNT/MWCNT bundles, with increasing values of P_m and λ, respectively. In Fig 4 (a), P_m values range from 0.3 (practical) to 1 (ideal), for a constant $\lambda=1\mu m$. It can be seen that there is a crossover point beyond which SWCNT bundle performance improves upon mixed SWCNT/MWCNT bundle performance. This can be explained as follows. As the metallic density increases, the mixed bundle has fewer metallic nanotubes than the SWCNT bundle due to large MWCNTs which prevent a tight packing. The SWCNT bundle has more tightly packed and consequently more metallic tubes. This leads to improved conductivity and performance for the SWCNT bundles. In Fig. 4 (b), a comparison is shown for λ varying from 1µm (practical) to about 3µm (ideal), with a constant P_m=*1/3*.

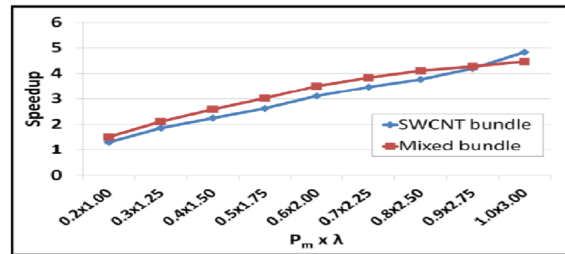


Figure 5. Speedup over Cu for different combinations of metallic density and MFP, for SWCNT bundle and mixed SWCNT/MWCNT bundle. X-axis shows P_m x λ, with λ in µm.

Fig. 5 shows the combined performance improvement for various configurations of P_m and λ lengths for the SWCNT bundle and mixed SWCNT/MWCNT bundles. It is clear from the results that increasing MFP and metallic densities can lead to a substantial improvement in delay (and hence performance) speedup for CNT bundle interconnects – as

much as 4.9× for SWCNT bundles and up to 4.4× for mixed SWCNT/MWCNT bundles. Improving CNT fabrication technology in order to achieve such performance gains is an area of tremendous research activity today [41]-[42].

V. Conclusion And Future Work

In this paper, we presented a preliminary system-level comparative analysis of the performance impact of using CNT-based global interconnects over Cu for multi-core CMP applications. The experimental results indicate that while SWCNTs are not as suitable for global interconnect buses due to their large ohmic resistance and very high delays, global MWCNT buses can provide performance speedups of up to 1.9× for CMP applications. Global interconnect buses implemented with SWCNT bundles and mixed SWCNT/ MWCNT bundles also lead to performance gains over copper global buses of up to 1.3× and 1.5× respectively. These gains can be further improved if the CNT mean free path (MFP) lengths and metallic densities are increased with advances in fabrication technology that are actively being explored today. Ultimately, while many manufacturing and technological factors will contribute to the realization of CNT global interconnects, our experimental results indicate that MWCNT and SWCNT/mixed bundle based global buses have the properties to potentially be viable replacements for Cu global interconnects in future high performance CMP applications, as process technology scales. Our future work will analyze trade-offs between CNT interconnect power, performance and bandwidth density, and explore the reliability of CNT interconnects.

Acknowledgments

This research was partially supported by grants from NSF (CCF-0702797).

References

[1] S. Pasricha, N. Dutt "On-chip Communication Architectures" Morgan Kauffman/Elsevier, Systems On Silicon Series, Apr 2008
[2] J. Li, et al., "Bottom-up approach for carbon nanotube interconnects," Appl. Phys. Lett., Vol. 82, Iss. 15, pp. 2491-2493, Apr. 2003.
[3] S. M. Rossnagel, T. S. Kuan, "Alteration of Cu Conductivity in the Size Effect Regime," JVST, Vol. 22, Iss. 1, pp. 240-247, Jan. 2004.
[4] ITRS, International Technology Roadmap for Semiconductors, 2005.
[5] M. S. Dresselhaus, et al, "Carbon nanotubes: synthesis, structure, properties, and applications", Springer, 2001.
[6] P. J. Burke, "Luttinger liquid theory as a model of the gigahertz electrical properties of carbon nanotubes," Trans. NANO, Vol. 1, Iss. 3, pp. 129-144, Sep. 2002.
[7] F. Kreupl, et al., "Carbon nanotubes for interconnect applications," IEDM Tech. Dig., pp. 683- 686, Dec. 2004.
[8] N. Srivastava, K. Banerjee, "Performance Analysis of Carbon Nanotube Interconnects for VLSI Applications," Proc. ICCAD, pp. 383-390, 2005.
[9] A. Raychowdhury, K. Roy, "A circuit model for carbon nanotube interconnects: comparative study with Cu interconnects for scaled technologies", Proc. ICCAD, pp. 237-240, 2004.
[10] A. Naeemi, J.D. Meindel, "Design and Performance Modeling for Single-Walled Carbon Nanotubes as Local, Semiglobal, and Global Interconnects in Gigascale Integrated Systems", Trans EDL, Vol. 54, Iss. 1, pp. 26-37, Jan 2007.
[11] A. Raychowdhury, K. Roy, "Modeling of Metallic Carbon Nanotube Interconnects for Circuit Simulation and a Comparison with Cu Interconnects for Scaled Technologies," IEEE TCAD, Vol. 25, No. 12, pp. 3042-3044, Jan 2006.
[12] A. Naeemi et al., "Performance Modeling and Optimization for Single- and Multi-Wall Carbon Nanotube Interconnects," Proc. DAC, pp. 568-573, 2007
[13] S. Haruehanroengra, W. Wang, "Analyzing Conductance of Mixed Carbon-Nanotube Bundles for Interconnect Applications", IEEE EDL, Vol. 28, No. 8, pp. 756-759, Aug 2007.
[14] A. Nieuwoudt, M. Mondal, Y. Massoud, "Predicting the Performance and Reliability of Carbon Nanotube Bundles for On-Chip Interconnect," IEEE ASPDAC, pp. 708-713, 2007.
[15] B. Q. Wei, R. Vajtai, P. M. Ajayan, "Reliability and current carrying capacity of carbon nanotubes," Appl. Phys. Lett., Vol. 79, No. 8, pp. 1172-1174, 2001.
[16] J. Hone, et al, "Thermal conductivity of single-walled carbon nanotubes," Phys. Rev., B, Vol. 59, No. 4, 1999.
[17] D. Rossi, et al., "Modeling Crosstalk Effects in CNT Bus Architectures", IEEE Trans. NANO, Vol. 6, Iss. 2, pp. 133-145, Mar 2007
[18] A. Nieuwoudt, Y. Massoud, "Assessing the Implications of Process Variations on Future Carbon Nanotube Bundle Interconnect Solutions," Proc. ISQED, pp. 119-126, 2007.
[19] S. Eachempati, et al., "Assessing Carbon Nanotube Bundle Interconnect for Future FPGA Architectures," Proc. DATE, pp. 1-6, 2007.
[20] S. Sato et al., "Novel approach to fabricate carbon nanotube via interconnects using size-controlled catalyst nanoparticles," Proc. IITC, pp. 230-232, 2006.
[21] O. Hjortstam, et al., "Can we achieve ultra-low resistivity in carbon nanotube-based metal composites?" App. Phy MSP, Vol. 78, No. 8, pp. 1175-1179, May 2004.
[22] H. Cho, et al., "Modeling of the performance of carbon nanotube bundle, cu/low-k and optical on-chip global interconnects", Proc. SLIP, pp. 81-88, 2007.
[23] A. Naeemi, J. D. Meindl, "Physical modeling of temperature coefficient of resistance for single- and multiwall carbon nanotube interconnects," IEEE EDL, Vol.28, Iss. 2, pp. 135-138, Feb 2007.
[24] P. J. Burke, "Quantitative theory of nanowire and nanotube antenna performance," ArXiv Cond. Matter E-Prints, pp. 1-14, Aug. 2004.
[25] M. W. Beattie, L. T. Pileggi, "Inductance 101: Modeling and Extraction," Proc. DAC, pp. 323-328, 2001.
[26] P. G. Collins, P. Avouris, "Multishell conduction in multiwalled carbon nanotubes," App. Phy. MSP, Vol. 74, No. 3, pp. 329-332, Mar. 2002.
[27] E. R. Dobbs, "Basic Electromagnetism" Chapman, Hall 1993.
[28] A. F. Mayadas, M. Shatzkes, "Electrical-Resistivity Model for Polycrystalline Films: the Case of Arbitrary Reflection at External Surfaces," Phys. Review B, Vol. 1, Iss. 4, pp. 1382-1389, 1970.
[29] E. H. Sondheimer, "The mean free path of electrons in metals," Adv. Physics, vol. 1, no. 1, pp. 1-42, 1952.
[30] C. P. Yue, S. S. Wong, "Physical Modeling of Spiral Inductors on Silicon," Trans. EDL, Vol. 47, Iss. 3, pp. 560-568, 2000.
[31] Y. I. Ismail, et al. "Effects of Inductance on the Propagation Delay and Repeater Insertion in VLSI Circuits" Proc. DAC, pp.721-724, 1999.
[32] S.C. Woo et al. "The SPLASH-2 programs: Characterization and methodological considerations", Proc. ISCAS, pp. 24-36, 1995.
[33] M. A. El-Moursy et al, "Optimum Wire Sizing of RLC Interconnect with Repeaters," Proc. GLSVLSI, pp. 27-32, 2003.
[34] SystemC initiative. http://www.systemc.org.
[35] S. Pasricha, N. Dutt, M. Ben-Romdhane, "Extending the Transaction Level Modeling Approach for Fast Communication Architecture Exploration", Proc. DAC, pp. 113-118, 2004.
[36] S. Pasricha, N. Dutt, M. Ben-Romdhane, "Fast Exploration of Bus-based Communication Architectures at the CCATB Abstraction", ACM TECS, Vol. 7, Iss. 2, pp. 22:1-22:32, Feb 2008
[37] AMBA AXI Specification http://www.arm.com/armtech/AXI.
[38] S. N. Adya, I. L. Markov, "Fixed-outline Floorplanning: Enabling Hierarchical Design", IEEE Trans TVLSI, Vol. 11, No. 6, pp. 1120-1135, Dec. 2003.
[39] V. Nookala, et al., "Designing optimized pipelined global interconnects: Algorithms and methodology impact," Proc. ISCAS, pp. 1-4, 2005.
[40] W. Wang, et al. "Inductance of mixed carbon nanotube bundles," Micro & Nano Letters, IET , Vol.2, No.2, pp.35-39, June 2007
[41] Z. Liu et al., "Densification of Carbon Nanotube Bundles for Interconnect Application", Proc. IITC, pp. 201-203, 2007
[42] L. Zhu et al., "Assembling Carbon Nanotube Bundles Using Transfer Process for Fine-Pitch Electrical Interconnect Applications", Proc. EETC, pp. 1981-1985, 2007.