Assessing Carbon Nanotube Bundle Interconnect for Future FPGA Architectures

Soumya Eachempati†, Arthur Nieuwoudt‡, Aman Gayasen§, N.Vijaykrishnan†, Yehia Massoud‡

†Dept. of Computer Science and Engineering
Pennsylvania State University, USA
{eachempa, vijay}@cse.psu.edu

‡Dept. of Electrical and Computer Engineering
Rice University, USA
{abnieu, massoud}@rice.edu

§Synopsys Inc., USA
gayasen@synopsys.com

Abstract

Field Programmable Gate Arrays (FPGAs) are important hardware platforms in various applications due to increasing design complexity and mask costs. However, as CMOS process technology continues to scale, standard copper interconnect will become a major bottleneck for FPGA performance. In this paper, we propose utilizing bundles of single-walled carbon nanotubes (SWCNT) as wires in the FPGA interconnect fabric and compare their performance to standard copper interconnect in future process technologies. To leverage the performance advantages of nanotube-based interconnect, we explore several important aspects of the FPGA routing architecture including the segmentation distribution and the internal population of the wires. The results demonstrate that FPGAs utilizing SWCNT bundle interconnect can achieve a 19% improvement in average area delay product over the best performing architecture for standard copper interconnect in 22 nm process technology.

1. Introduction

Given the increasing demand for Field Programmable Gate Arrays (FPGAs) as hardware platforms in reconfigurable applications, the development of FPGAs with greater performance has been a central goal of the industry. The industry leaders, Xilinx and Altera, have recently leveraged 65nm technology in their newest product offerings and have added several hardware blocks to the FPGA fabric to increase performance. For the foreseeable future, the industry will continue to exploit technology scaling and innovative architectural enhancements to make FPGAs an attractive hardware platform for many applications.

While the reconfigurable nature of FPGAs makes them an attractive hardware solution, they can suffer from large overheads (area, delay, and power) because of their programmable interconnect fabric. The area overhead is mainly attributed to the programmable switches in the fabric. Furthermore, the on-chip communication network in an FPGA typically consists of multiple routing segments and switches, which can result in larger delay compared to ASICs. The impact of interconnect performance will continue to increase as process technology scales. The International Technology Roadmap for Semiconductors (ITRS) predicts traditional copper interconnect will be a major bottleneck when feature sizes become smaller than 45nm [8]. Given the long-term scaling problems associated with traditional copper wires, alternative interconnect technologies and their architectural implications for FPGAs in future process technology must be explored.

Single-walled carbon nanotubes (SWCNT) have been proposed as a possible replacement for on-chip copper interconnect due to their large conductivity and current carrying capabilities [17]. SWCNTs are rolled graphitic sheets that can either be metallic or semiconducting depending on their chirality [28]. Due to their covalently bonded structure, CNTs are extremely resistant to electromigration and other sources of physical breakdown [2]. To reduce the impact of the large contact resistance of an individual SWCNT, bundles of SWCNTs in parallel, depicted in Figure 1, have been proposed and physically demonstrated as a possible interconnect medium [10]. SWCNT bundles can provide substantially lower resistance than copper wires, especially for intermediate and global interconnect applications [16, 21, 20]. Given the extensive research efforts dedicated to the fabrication and process integration of nanotube-
In this paper, we propose utilizing SWCNT bundles as wires in the FPGA interconnect fabric and compare their performance to standard copper interconnect in future process technology. We utilize nanotube-based interconnect for the longer inter-CLB wires and standard copper interconnect for the shorter intra-CLB wires since SWCNT bundles are predicted to be more beneficial for intermediate and global interconnect applications. We explore different segment lengths and switch populations in the interconnect fabric to optimize the performance of the CNT-FPGA architecture and compare it with the performance of an FPGA utilizing scaled copper interconnect in terms of delay and area. The results demonstrate that FPGAs that utilize SWCNT bundle interconnect can achieve a 19% improvement in average area-delay product over the best performing architecture for standard copper interconnect in 22 nm process technology.

2. FPGA Architectures

In this study, we focus on the island-style FPGA architecture (see Figure 2), which has been utilized in FPGAs from both Xilinx and Altera. The FPGA consists of logic blocks (CLBs and LBs used interchangeably) and programmable interconnect resources. The CLBs contain Look-Up-Tables (LUTs) and Flip-Flops (FFs) connected through fast local interconnect. The CLBs access the interconnect fabric through connection blocks (CBs), and the inter-CLB wires are interconnected through switch blocks (SBs). These interconnect resources typically consume approximately 70% of the FPGA area and constitute the major portion of critical path delay for most designs. Furthermore, the interconnect consumes the majority of FPGA power [25, 26]. With technology scaling, the wire delay becomes more significant.

Given the increasing impact of interconnect on FPGA performance as process technology scales, programmable architectures using non-lithographic technologies will be required. Dehon, Goldstein, and Tour have proposed programmable architectures based on self-assembled nano-structures [5, 7, 27]. Goldstein constructed a crossbar-based device by aligning nano-wires in two planes at right angles. The crosspoints contained molecules that provided programmable logic as well as interconnections. The nanostructure suffered from severe signal-degradation since there was no way to restore the signal using only two terminal devices. Dehon overcame this problem by using SiNW based FETs to restore the signals and proposed a PLA architecture based on the nano-structure, but the logic functionality was limited to OR (and inversion). Tour proposed replacing the FPGA's logic blocks with nanocells and connecting them using metal wires. However, Tour’s architecture does not address the performance bottleneck imposed by metal interconnect in future FPGAs. Another recent study explored the use of nano-wires concluding that a material more conductive than NiSi is required to improve the delay in the FPGA interconnect fabric over standard copper wires. Since SWCNT bundles will potentially have less resistance than copper wires as technology scales, they may be a good choice for FPGA interconnect.

3. Carbon Nanotube Interconnect

3.1 Modeling SWCNT Bundles

To evaluate SWCNT bundles for interconnect in future FPGA applications, we utilize the circuit model in [16] that can characterize their behavior in a scalable manner. Each SWCNT has lumped ballistic ($R_b \approx 6.5\ k\Omega$) and contact ($R_c$) resistances that have fixed values regardless of the length of the nanotube bundle ($l_b$). The con-
tact resistance is due to imperfect SWCNT-metal contacts, which are typically constructed using Gold, Palladium, or Rhodium [9]. The nanotubes also have a distributed ohmic resistance \( (R_o) \), which is determined by \( l_b \) and mean free path of acoustic-phonon scattering \( (\lambda_{ap}) \) in the individual nanotubes. Since the individual SWCNTs have a maximum saturation current \( (I_s) \), the overall resistance also depends on the applied bias voltage \( (R_{hb} = V_{bias}/I_s) \) [29]. The total resistance of a nanotube bundle \( (R_b) \) is \( R_t/n_b \) where \( R_t \) is the resistance of an individual nanotube and \( n_b \) is the total number of nanotubes in the bundle. We utilize the diameter-dependent resistance model presented in [21] to characterize the resistance associated with SWCNT bundles.

The diameter of the individual nanotubes in the bundle plays an important role in determining the resistance of SWCNT bundle interconnect. For ohmic resistance, \( \lambda_{ap} \) is proportional to the individual nanotube’s diameter \( (d_t) \). A proportionality constant \( (C_\lambda = \lambda_{ap}/d_t) \) based on the empirical data from [23] captures this diameter dependence [21]. Recent studies have revealed that the contact resistance of a SWCNT greatly increases when \( d_t < 2.0 \) nm [21, 9]. In addition, the saturation current \( (I_s) \) decreases substantially to approximately 5 \( \mu \)A for \( d_t = 1 \) nm [9], which increases the overall impact of \( R_{hb} \). In contrast, as \( d_t \) decreases, \( n_b \) increases, which decreases the overall ohmic and contact resistances for the bundle.

For nanotube bundle interconnect with relatively short length values, which are common in many FPGA applications, \( R_c \) and \( R_{hb} \) have a large impact on \( R_b \) [19, 21]. Therefore, a trade-off exists between minimizing increases in \( R_c \) and \( R_{hb} \) for individual nanotubes due to decreases in \( d_t \) and increasing \( n_b \) by reducing \( d_t \). We exploit the aforementioned trade-off to locate the optimal \( d_t \) value to minimize \( R_b \) for a particular \( l_b \) value. Figure 3 displays the resistance of an SWCNT bundle for optimal and minimum \( d_t \) versus \( l_b \) based on the moderate technology assumption set described in Section 3.2. For short \( l_b \) values \( (l_b < 4 \) \( \mu \)m \), \( R_c \) and \( R_{hb} \) dominate \( R_o \), and therefore, the optimal \( d_t \) value is the smallest diameter for which \( R_c \) and \( I_o \) are approximately equal to their nominal values. For intermediate \( l_b \) values \( (4 \) \( \mu \)m \( < l_b < 30 \) \( \mu \)m \), the optimal \( d_t \) value decreases to increase \( n_b \) since the \( R_o \) becomes a larger portion of \( R_b \) relative to \( R_c \) and \( R_{hb} \). For large \( l_b \) values \( (l_b > 30 \) \( \mu \)m \), the optimal \( d_t \) value is the minimum allowed value due to technology constraints since \( R_o > R_c + R_{hb} \).

The capacitance of a nanotube bundle consists of both a quantum capacitance and an electrostatic capacitance [4], which can be modeled using the techniques presented in [16]. The inductance of an SWCNT bundle can be modeled using the scalable inductance model presented in [22]. While inductance can have a relatively large impact for global interconnect applications [15], for the wire lengths and low operating frequencies utilized in FPGA applications, the inductance will have a relatively small impact on overall interconnect performance [6]. Using the circuit model, we can evaluate the performance advantages of SWCNT bundle interconnect over standard copper technology for future FPGA applications.

### 3.2 Technology and Manufacturing Challenges for SWCNT Bundles

SWCNT manufacturing technology will play a crucial role in determining the viability and performance of nanotube bundles in future integrated circuits [12]. SWCNT bundles have metallic nanotubes that are randomly distributed within the bundle. With no special separation techniques, the metallic nanotubes are distributed with probability \( P_m = 1/3 \) since approximately one-third of possible SWCNT chiralities are metallic [28]. However, techniques such as alternating current (AC) dielectrophoresis [24] and ion-exchange chromatography [13] have the potential to increase the proportion of metallic nanotubes. The increased

---

### Table 1. SWCNT technology assumptions.

<table>
<thead>
<tr>
<th>Technology Assumption</th>
<th>( P_m )</th>
<th>( R_c + R_e ) (k( \Omega ))</th>
<th>( C_\lambda )</th>
<th>( I_o ) (( \mu )A)</th>
<th>Minimum ( d_t ) (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conservative</td>
<td>0.33</td>
<td>200</td>
<td>849</td>
<td>20</td>
<td>1.0</td>
</tr>
<tr>
<td>Moderate</td>
<td>0.60</td>
<td>100</td>
<td>889</td>
<td>25</td>
<td>0.8</td>
</tr>
<tr>
<td>Aggressive</td>
<td>0.90</td>
<td>6.5</td>
<td>1333</td>
<td>30</td>
<td>0.4</td>
</tr>
</tbody>
</table>

---

![Figure 3. Resistance of an SWCNT bundle for optimal and minimum \( d_t \) values versus \( l_b \) for (a) total and ohmic resistance, (b) contact and high bias resistance, and (c) the optimal \( d_t \) value.](image-url)
lumped resistance due to imperfect metal contacts to the nanotubes is significant for local and intermediate interconnect applications [21]. However, as nanotube fabrication and bonding techniques have been improved, the additional resistance due to imperfect contacts has been significantly reduced and in several experimental cases has approached 0 (i.e. $R_i + R_c = R_t$) [17, 11]. The average number of defects per unit length can impact both $\lambda_{ap}$ and $I_o$, which can effect both $R_e$ and $R_{hb}$ [19]. In [23], the experimental mean free path-to-nanotube diameter proportionality constant was reported to be $C_\lambda = \lambda/d_i \approx 1600 \text{ nm}/1.8 \text{ nm} = 889$. $I_o$ values of 15 to 30 $\mu A$ per nanotube have been reported [29]. Bundles of nanotubes with $d_i$ values as small as 0.4 nm have been reported [1], but diameters on the order of 0.8 to 1.0 nm have demonstrated both mechanical stability and good electrical conductivity [14]. Based on current and predicted SWCNT fabrication technology, we utilize the conservative, moderate, and aggressive technology assumptions listed in Table 1 to evaluate the performance of SWCNT interconnect in future FPGA applications.

Another important manufacturing and design consideration for SWCNT interconnect is the increased contact resistance due to fanout. In the absence of possible nanotube bundle junction formation techniques [18], which may be difficult to realize for the interior nanotubes in an SWCNT bundle, accommodating fanout will require multiple metal contacts as depicted Figure 1. By adding additional metal contacts to the nanotube bundle interconnect, the minimum distance for which nanotubes have a resistance advantage over copper increases due to the larger contact resistance. Therefore, the length required for SWCNT interconnect with fanout to have lower resistance is a multiple of the length required for one segment.

4. Exploring the influence of CNT on FPGA Routing Architecture

The routing architecture in an FPGA is tailored to support both long and short connections and provide sufficient flexibility to connect different logic blocks. Consequently, FPGA interconnect in a channel is composed of segments that are of different lengths. For example, the segments can span a single logic block (single length segment), two logic blocks (two length segment), six logic blocks (hex length segment), or span the entire row of logic blocks (long lines). Figure 4(a) provides a typical distribution of wire-lengths in a contemporary Virtex II FPGA. Wire length segment distributions chosen for an architecture are influenced by the underlying wire performance. Consequently, when adopting SWCNT-based interconnect in FPGAs, we need to explore how the segmentation distribution would change.

The level of routing flexibility in the interconnect fabric is determined by the switch box (SB) and connection box (CB) configurations. The switch box population fraction, displayed in Figure 4(b) for the segmentation experiments performed in Section 4.2, is the ratio of the actual number of switch boxes to the maximum number of switch boxes. Figure 5 illustrates an example of how the population fraction of CB and SB in wiring segments is evaluated. In the CNT-based architecture, the additional connectivity imposes challenges since the contact resistance of the connections to the switch and connection boxes is significant. Consequently, the switch and connection box configurations for a SWCNT-based interconnect fabric have different constraints than traditional copper-based architectures.

4.1 Experimental Setup

In order to explore the impact of the wire segmentation and SB/CB topologies, we performed simulations using the 22nm technology node. The logic blocks and switches were modeled to be in scaled CMOS, the diffusion capacitance was obtained through HSPICE simulations, and the process parameters were obtained from ITRS [8]. The SWCNT bundle interconnect model presented in Section 3 was used to characterize SWCNT-based wires. We also selected the best bundle configurations based on the length of the wire segment and the number of contacts due to the SB/CB configuration. The selection was guided by our results shown in Figure 3. Our simulations were performed for the three different technology assumptions listed in Table 1. Using 20 MCNC benchmarks and VPR, [3] we evaluated the area and delay performance of the various architectural options in the design space.

4.2 Segmentation

We simulated 65 different segmentation options, displayed in Figure 6, for both Cu and SWCNT-based architectures. For these cases, the CB/SB configuration is shown in Figure 4(b). Figure 7 shows the area-delay product (ADP) for the different configurations over 20 benchmarks for the moderate technology assumption in 22 nm technology. The results indicate that larger benefit can be obtained from longer length wires (double and hex) for an
SWCNT-based architecture. There is an average improvement of 12.8% and 13.8% over Cu-based architectures in area-delay product (ADP) and critical path delay (CPD) respectively, for moderate technology assumption. Similar trends were obtained for the other technology assumptions and a maximum improvement of 17.5% in critical path delay and 19.0% in area delay product was obtained for the aggressive technology. The benchmarks that experienced the least performance improvement ('diffeq', 'frisc', and 'tseng') utilized a significant number of short interconnect segments, which have a higher per unit length resistance due to the SWCNT-metal contact resistance. In contrast, an improvement of as high as 54% was seen for the 'bigkey' benchmark, which maps relatively well to the long wires in the nanotube-based FPGA architecture. Based on the distribution of wire segment lengths for the best performing SWCNT-based FPGA architecture, displayed in Table 2, the SWCNT-based architectures prefer longer wire segments than Cu-based architectures because of the significant nanotube contact resistance. The percentage of hex segments in the best configuration for CNT is significantly larger than for best Cu solution. Also, notice a slight reduction in fraction of long lines, which do not benefit as much in SWCNT interconnect since they are limited by number of connections that they can provide due to the contact resistance.

### 4.3 Switch Block and Connection Block Population

In addition to exploring the performance impact of wire length segmentation, we also explored the impact of switch block and connection block population fractions. As the SB and CB population fractions are increased, the contact resistance of SWCNT wires increases. Thus, longer length segments will be required to compensate for the increase in contact resistance. We simulate all feasible combinations of SB and CB population fractions for a segment length in an architecture where every wire segment is of the same length. We considered wire segments of length 2 to 6. A total of 33 architectures obtained by varying the fraction of CBs and SBs in the range of 0.5-1, guided by routability, were simulated.

We observed that a smaller number of CB/SB connections is better for the same segment length due to smaller effective contact resistance. However, some of the benchmarks cannot route because of decreased routing flexibility. Consequently, we observed that longer length wire segments based on CNT wires with moderate connectivity provide a good solution that satisfies both flexibility and performance. The longer CNT wire segments help balance the increased contact resistance due to increased CB/SB connections with decreased ohmic resistance. The best combination of wire segments, CB and SB configurations for CNT achieved a 8% improvement in critical path delay and a 14% improvement in ADP over the best configuration of these parameters for a Cu wire. Figure 8(b) shows the improvements with the best combination of segment lengths and CB/CB configurations of CNT over that of Cu averaged over all benchmarks.

We also observed that as the CNT technology assumption were varied from conservative to aggressive, the best
In this paper, we proposed a nanotube-based interconnect solution for future FPGA architectures. We explored several important aspects of the FPGA routing architecture. The results demonstrate that FPGAs that utilize SWCNT bundle interconnect can achieve a 19% improvement in average area delay product over standard copper interconnect in 22 nm process technology. In the future, we will explore the impact of power consumption, cross-talk, noise and process variations on future FPGA architectures that leverage CNT-based interconnect solutions. While many factors will ultimately determine the realization of SWCNT bundle interconnect, the results indicate that nanotube-based solutions will potentially offer substantial performance improvement for future FPGA architectures.

### References


