University of New Mexico
UNM Digital Repository

Electrical and Computer Engineering ETDs | Engineering ETDs

7-2-2012

Carbon nanotubes as interconnect for next generation network on chip

Manoj Kumar Ramalingam Rajasekaran

Follow this and additional works at: https://digitalrepository.unm.edu/ece_etds

Recommended Citation
Ramalingam Rajasekaran, Manoj Kumar. "Carbon nanotubes as interconnect for next generation network on chip." (2012). https://digitalrepository.unm.edu/ece_etds/214

This Thesis is brought to you for free and open access by the Engineering ETDs at UNM Digital Repository. It has been accepted for inclusion in Electrical and Computer Engineering ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.

Manoj Kumar Ramalingam Rajasekaran
Candidate

Electrical and Computer Engineering
Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Dr.Payman Zarkesh-Ha Chairperson

Dr.Charles B.Fleddermann

Dr.Sanjay Krishna

Dr.Steven C. Suddarth

CARBON NANOTUBES AS INTERCONNECT FOR NEXT GENERATION NETWORK ON CHIP

By

Manoj Kumar Ramalingam Rajasekaran

B E., Electronics and Communication,
Anna University, 2008

THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Electrical Engineering

The University of New Mexico

Albuquerque, New Mexico

May 2012

Dedication

To my father Mr.Rajasekaran and mother Mrs.Maheswari, my motivation personified, for their never ending support, encouragement, love and sacrifice, without them I would be nothing. They have bestowed all their wisdom upon me. To my Sisters Nithya, Sneha and brother Jagadeesh for their prayers and affection.

To my best buddies Shantaram, Dewan, Natesan, Kumanoor Karthik, Baskar, Vignesh, Manickam, Kavin, Nallathambi, Baskar(Trichy), Gnanavel, Giri, Nanda, Udhaya, Karthik and Senthil for their constant support. To Thiyagarajan Sir and My friend's father for their timely help before my travel to US. To my friend Theresan who paid me the application fees for Universities in US and also for the VISA interview, for his encouragement. To the Pastor of Garfield Gospel, to Andrew Anna for his time and support in my job search, to George, Paul & John's family.

To my Albuquerque friends, VLSI design group. To George Bezerra and Shilpa for helping with my thesis. To the 4th grade students of Alvarado elementary for their 20$. To Ali, Ashwin, Pankaj and Rakesh for giving valuable feedback. To Sreenivasan Rama Sreedharan and BalaMurugan for dedicating their time for my job search.

Acknowledgement

I would like to thank Prof. Payman Zarkesh-Ha, my advisor and thesis chair, for his guidance, support and encouragement during my Master's program and also for being an outstanding teacher in the classrooms.

I thank Prof. Sanjay Krishna, Prof. Charles Fleddermann and Dr. Steve Suddarth for accepting to be a part of the committee in my thesis defense.

I thank Prof. Kenkre of Physics department for employing me as a student worker for the first month of my study at UNM. Very big thanks to Leora, Jennifer, Daniel, David and Michelle for hiring me at the Center for Academic Program Support and providing me tuition assistance throughout my study at UNM. Thanks to Lauren and Tech Team guys.

I would like to thank Mr.Maurice Thompson of Engineering Student Services for employing me as a Mathematics teacher in summer 2010 and kindly helping out in the future semesters. I thank Prof. Charles Fleddermann for his support during the beginning of my program.

CARBON NANOTUBES AS INTERCONNECTS FOR NEXT GENERATION NETWORK ON CHIP

By
Manoj Kumar Ramalingam Rajasekaran

B.E in Electronics and Communication Engineering, Anna University, 2008
M.S.in Electrical Engineering, University of New Mexico, 2012

Abstract

Multi-core processors provide better performance when compared with their single-core equivalent. Recently, Networks-on-Chip (NoC) have been emerged as a communication methodology for multi core chips. Network-on-Chip uses packet based communication for establishing a communication path between multiple cores connected via interconnects. Clock frequency, energy consumption and chip size are largely determined by these interconnects. According to the International Technology Roadmap for Semiconductors (ITRS), in the next five years up to 80% of microprocessor power will be consumed by interconnects. In the sub 100nm scaling range, interconnect behavior limits the performance and correctness of VLSI systems. The performance of copper interconnects tend to get reduced in the sub 100nm range and hence we need to examine for other interconnect

options. Single Wall Carbon Nanotubes exhibit better performance in sub 100nm processing technology due to their very large current carrying capacity and large electron mean free paths.

This work suggests using Single Wall Carbon Nanotubes (SWCNT) as interconnects for Networks-on-Chip as they consume less energy and gives more throughput and bandwidth when compared with traditional Copper wires.

Contents

Dedication ... iii

Acknowledgement ... iv

Preface ... xii

Thesis Organization ... xii

Contribution of this Thesis ... xiv

Chapter 1 ... 1

Introduction ... 1

1.1 Multicore Processors ... 1

1.2 Challenges and Communication Demands ... 5

1.3 Network-on-Chip ... 6

1.4 On-Chip Network Blocks ... 9

1.5 Operation of Network-on-Chip ... 11

1.6 Advantages and Dis-Advantages of Network on Chip ... 14

Chapter 2 ... 17

CNT for Interconnects ... 17

2.1 Interconnects in VLSI ... 17

2.2 Issues Concerning Interconnects ... 19

2.3 Limitations of Copper Interconnects ... 20

2.4 Carbon Nanotubes ... 22

2.5 Types of Carbon Nanotubes ... 24

2.6 General Properties of Carbon Nanotubes ... 25

Chapter 3 ... 29

CNT in Network-on-Chip ... 29

3.1 Carbon Nanotubes (CNT) Interconnect ... 29

3.2 Delay Calculation for CNT and Copper: 33

3.3 Calculations for a Network on Chip.... 36

3.4 Bandwidth Calculation for CNT in NOC: 37

3.5 Total System Throughput 40

Chapter 4.... 43

Energy Consumption of Network on Chip.... 43

4.1 Communication Probability Distribution.... 43

4.2 Rent's Rule Traffic Patterns 44

4.3 Modeling Energy Consumption.... 47

4.4 Orion-Network on Chip Simulator.... 49

Chapter 5.... 52

Hybrid NoC 52

5.1 Challenges and Potential for Carbon Nanotube Applications 52

5.2 Fabrication of Nanotubes 54

5.3 Hybrid NoC Architecture.... 57

5.4 Throughput and Energy Consumption Analysis.... 58

5.5 A Hybrid NoC with Copper and Carbon nanotubes.... 60

Chapter 6.... 62

Conclusion and Future Work.... 62

REFERENCES 64

List of Figures

Figure 1 Single core processor 2

Figure 2 Multicore Processor 3

Figure 3 Network on Chip 9

Figure 4 A 5X5 NoC 12

Figure 5 Interconnect dimensions 18

Figure 6 ITRS Roadmap showing copper resistivity[7] 20

Figure 7 Conductivity comparison between Copper and SWCNT[40] 22

Figure 8 Carbon Nanotube 23

Figure 9 SWCNT and MWCNT 24

Figure 10 3-d Models of SWCNT types 27

Figure 11 Equivalent Circuit Model for Ideally Contacted SWCNT [14] 30

Figure 12 Configuration of Copper and CNT Interconnects [16] 31

Figure 13 Equivalent circuit model for delay characteristics 33

Figure 14 Delay comparison between copper and CNT 34

Figure 15 Delay Comparison for Cu, Bundles and MWCNT [20][21] 35

Figure 16 4 X 4 NOC with Mesh Topology *37*

Figure 17 Intercore Architecture *38*

Figure 18 Channels determining bisection bandwidth 39

Figure 19 Aggregate Bandwidth of 4 X4 NOC *39*

Figure 20 Channels determining Bisection Bandwidth for 5 X 5 NoC *41*

Figure 21 Throughput Comparison of 4 X 4 NOC *42*

Figure 22 CPD for a 4X4 NOC *47*

Figure 23 Energy Consumption Comparison for 4 x 4 NoC *51*

Figure 24 Conventional NoC and Hybrid NoC *57*

Figure 25 Hybrid NoC with Copper and CNT *60*

List of Tables

Table 1.1 Single core vs Multi core……………………………………………………………………………………………….5

Table 3.1 Capacitance per µm for Copper and SWCNT………………………………………………………………………32

Table 3.2 Wire Parameters for CNT and Copper…………………………………………………………………………………39

Table 3.3 Parameters for throughput calculation ………………………………………………………………………………..42

Table 4.1 Parameters for calculating Energy Consumption…………………………………………………………………51

Table 5.1 Summary of results………61

Preface

Thesis Organization

This thesis is organized in the following manner. Chapter 1 gives an introduction to multi core architectures and their potential advantages. It also covers the basic mechanism to establish communication between the cores and gives a detailed description and need for Network-on-Chip in a System-on-Chip. Chapter 2 discusses about the role of interconnects in VLSI technology and introduces carbon nanotubes as an interconnect for future VLSI circuits. It explains the limitations of using copper interconnects in deep nanometer regime, introduces carbon nanotubes and types of carbon nanotubes and their physical and electrical properties are discussed in detail. Chapter 3 discusses the modeling of interconnects and compares the delay of single wall carbon nanotubes with copper. It analyzes the performance of carbon nanotubes in a Network on Chip. Chapter 4 talks about energy consumption calculation for Network on Chip and compares the energy consumption of Network on Chip with carbon nanotubes and copper. Chapter 5 talks about the bottlenecks involved in implementing carbon nanotubes and introduces a hybrid NoC which uses dedicated buses for local communication and a Network-on-Chip for

global communication. It also proposes a hybrid NoC with copper as local interconnects and carbon nanotubes as a global interconnect. Throughput and Energy consumption are predicted for a Hybrid NoC. Chapter 6 concludes the thesis work and explains what can be implemented in the future.

Contribution of this Thesis

This thesis has analyzed energy consumption, throughput and bandwidth of a Network on Chip by replacing Copper interconnects with single walled carbon nanotubes. Proposed a Hybrid Network on Chip with Copper wires for local interconnects and carbon nanotubes for global interconnects. Predicted the energy consumption and throughput of the Hybrid Network on Chip.

Chapter 1

Introduction

1.1 Multicore Processors

Since the invention of computers there have been tremendous improvements done to the processor and memory storage. Computers started with single core processors and, as personal computers have become more prevalent and more applications have been designed for them, the end-user has seen the need for a faster, more capable system to keep up. Speedup has been achieved by increasing clock speeds and, more recently, adding multiple processing cores to the same chip. Multicore architectures comprise of multiple cores that reside on a single chip and are interconnected using an on-chip packet based or bus based network. At its simplest, multi-core is a design in which a single physical processor contains the core logic of more than one processor. The combined pressures from ever-increasing power consumption and the diminishing returns in performance of uniprocessor architectures have led to the advent of multi-core chips [3]. With a growing number of transistors available at each new technology generation, coupled with a reduction in design complexity enabled by the modular design of multi-core chips, this multi-core

wave looks set to stay, in both general-purpose computing chips as well as application-specific SoCs. This multi core wave may lead to hundreds and even thousands of cores integrated on a single chip. Increasing transistor counts will lead to greater system integration for multiprocessor systems-on-chip (MPSoCs)[3] . The following figure show a basic configuration of a microprocessor system with a single core.

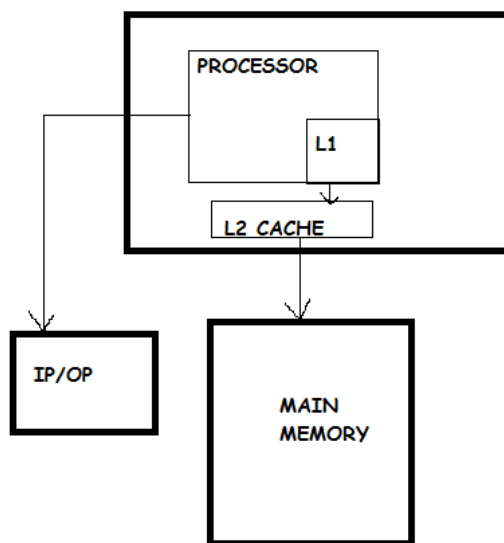

Figure 1 Single core processor

The Level 1 cache is closest to the processor and this is very fast memory used to store data frequently used by the processor. Level 2 cache is just off-chip, slower than L1 cache, but still much faster than the main memory; L2 is

larger than L1 cache and used for the same purpose. Main memory is very large and slower than cache and is used for example to store a file currently being edited in Microsoft Word. Most systems have between 1GB to 4GB of main memory compared to approximately of 32KB of L1 and 2MB of L2 cache [1]. Finally, when data isn't located in cache or main memory the system must retrieve it from the hard disk, which takes much more time than reading from the main system.

The multi-core design puts several processor cores and packages them as a single physical processor. The goal of this design is to enable a system to run more tasks simultaneously and thereby achieve greater overall system performance.

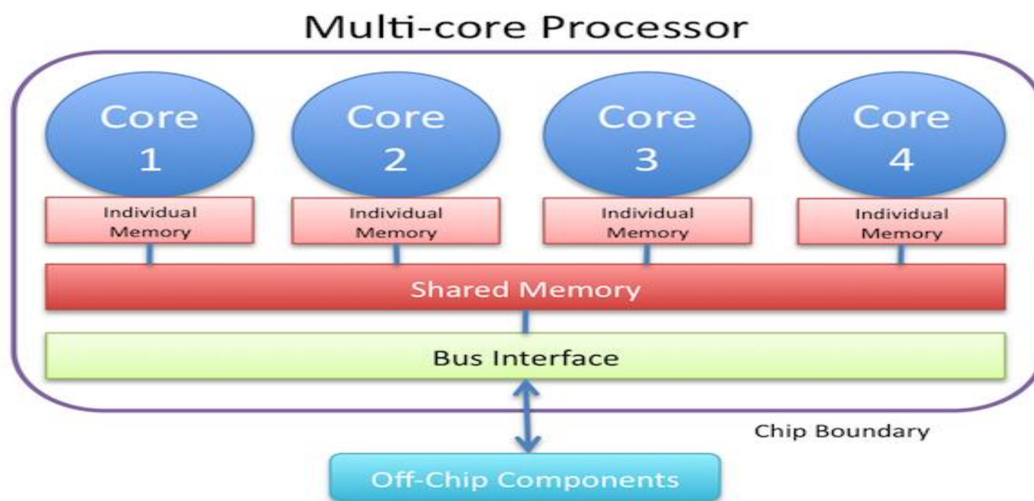

Figure 2 Multicore Processor

In a multi-core design each core has its own execution pipeline and each core has the resources required to run without blocking resources needed by the other software threads. The multi-core design enables two or more cores to run at somewhat slower speeds and at much lower temperatures. The combined throughput of these cores delivers processing power greater than the maximum available today on single core processors and at a much lower level of power consumption. If we set two cores side-by-side, one can see that a method of communication between the cores, and to main memory, is necessary. This is usually accomplished either by using a single communication bus or by using an interconnection network. Multicore processors seem to answer the deficiencies of single core processors, by increasing bandwidth while decreasing power consumption.

Table below shows a comparison of single core and multi core processors (8 cores) used by Packaging Research Center at Georgia Tech. The values shown below are for 45nm technology.

Parameter	Single Core	Multi Core
Vdd	1.0 V	1.0 V
I/O pins	1280	3000(Estimated)
Bandwidth	125GByte/s	1 TeraByte/S
Power	429.78W	107.9W

Table 1.1: Single core vs. Multi core [38]

1.2 Challenges and Communication Demands

Achieving future performance gains will rely on removing the communication bottleneck between the processors and the memory components that feed these bandwidth hungry many core designs. Increasingly, efficient communication between execution units or cores will become a key factor in improving the performance of many core chips. Having multiple cores on a single chip gives rise to some problems and challenges. Power and temperature management are two concerns that can increase with the addition of multiple cores. If two cores were placed on a single chip without any modification, the chip would, in theory, consume twice as much power and generate a large amount of heat. To

account for this each design runs the multiple cores at a lower frequency to reduce power consumption. There is also a question about interconnects which type of interconnect is best suited for multicore processors? We can use either a bus-based approach or an interconnection network. As the number of on-chip cores increases, a scalable and high-bandwidth communication fabric to connect them becomes critically important. As a result, packet switched on-chip networks are rapidly replacing buses and crossbars to emerge as the pervasive communication fabric in many-core chips. Such on-chip networks have routers at every node, connected to neighbors via short local on-chip wiring, while multiplexing multiple communication flows over these interconnects to provide scalability and high bandwidth.

1.3 Network-on-Chip

Several acronyms have emerged as on-chip network research has gained momentum. They are NoC (Network-on-Chip), OCIN (On-Chip Interconnection Network) and OCN (On Chip Network). They all mean the same thing and in this report we are going to use the name Network on Chip. A Network-on-Chip, as a subset of a broader class of interconnection networks, can be viewed as a programmable system that facilitates the transporting of data between nodes [5]. An

on-chip network can be viewed as a system because it integrates many components including channels, buffers, switches and control. With a small number of nodes, dedicated ad hoc wiring can be used to interconnect them. However, the use of dedicated wires is problematic as we increase the number of components on-chip: The amount of wiring required to directly connect every component will become prohibitive.

Designs with low core counts can leverage buses and crossbars, which are considered the simplest variants of on-chip networks. In both traditional multiprocessor systems and newer multi-core architectures, bus-based systems scale only to a modest number of processors. This limited scalability is because bus traffic quickly reached saturation as more cores are added to the bus, so it is hard to attain a high bandwidth. The power required to drive a long bus with many cores tapping onto it is also exorbitant. In addition, a centralized arbiter adds arbitration latency as core counts increase. To address these problems, sophisticated bus designs incorporate segmentation, distributed arbitration, split transactions and increasingly resemble switched on-chip networks. Crossbars address the bandwidth problem of buses, and have

been used for on-chip interconnects for a small number of nodes. However, crossbars scale poorly for a large number of cores; requiring a large area of footprint and consuming high power. In response, hierarchical crossbars, where cores are clustered into nodes and several levels of smaller crossbars provide the interconnection, are used. These sophisticated crossbars resemble multi-hop on-chip networks where each hop comprises small crossbars. On-chip networks are an attractive alternative to buses and crossbars for several reasons. First and foremost, networks represent a scalable solution to on-chip communication, due to their ability to supply scalable bandwidth at low area and power overheads that correlate sub-linearly with the number of nodes. Second, on-chip networks are very efficient in their use of wiring, multiplexing different communication flows on the same links allowing for higher bandwidth. Finally, on-chip networks with regular topologies have local, short interconnects that are fixed in length and can be optimized and built modularly using regular repetitive structures, easing the burden of verification.

1.4 On-Chip Network Blocks

The design of an on-chip network can be broken down into its various building blocks: its topology, routing, flow control, router microarchitecture and design, and link architecture [5]. The following figure shows the structure of a basic Network on Chip.

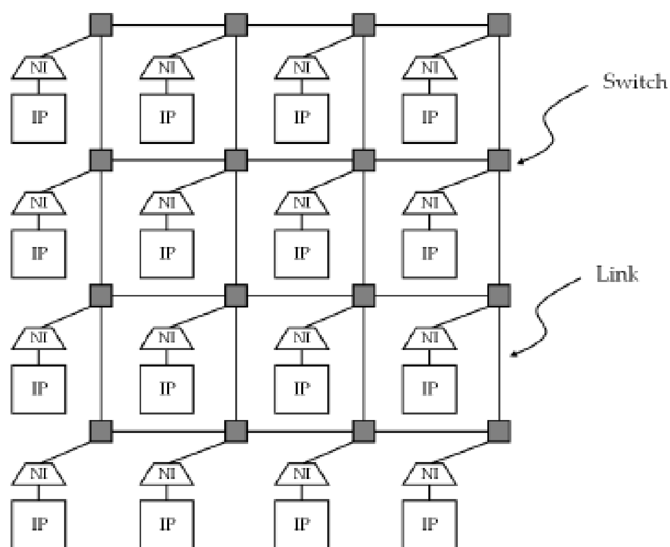

Figure 3 Network on Chip

Topology: A Network on Chip is composed of channels and router nodes. The network topology determines the physical layout and connections between nodes and channels in the network.

Routing: For a given topology, the routing algorithm determines the path through the network that a message will take to reach its destination. A routing algorithm's

ability to balance traffic has a direct impact on the throughput and performance of the network.

Flow control: Flow control determines how resources are allocated to messages as they travel through the network. The flow control mechanism is responsible for allocating and de-allocating buffers and channel bandwidth to waiting packets. A packet is a subdivision of a message. Resources can be allocated to packets in their entirety; however, this requires very large buffer resources making it impractical on chip. Most commonly, on-chip networks handle flow control at the flit level. Flit is a flow control unit which is a subdivision of a packet. Buffers and channel bandwidth are allocated on the smaller granularity of flits rather than whole packets; as a result, routers can be designed with smaller buffers.

Router Microarchitecture: Generic router microarchitecture is comprised of input buffers, router state, router logic, allocators and a crossbar (or switch). An allocator performs a matching between resources and requesters, i.e., and allocator assigns the former to the later. Router functionality is often pipelined to improve throughput. Delay through each router in the on-chip network is the primary contributor to communication latency. As a result,

significant research effort has been spent reducing router pipeline stages and improving throughput.

Link Architecture: All on-chip network prototypes have used conventional full-swing logic and wires. Wires use repeaters to improve signal reach. Interconnects play an important role in on-chip network as the network performance depends highly on the behavior of these interconnects. Latency, Energy Consumption and Throughput of the network depends on the interconnection fabric.

1.5 Operation of Network-on-Chip

Rather than being statistically wired from source to destination, data is injected as packets into a complete network of wires, switches, and routers, and it is the network that dynamically decides how and when to route these packets through its segments. This is the reliable approach. Systems-on-a-Chip (SoCs) typically refers to chips that are tailored to a specific application or domain area, which are designed quickly through the composition of IP blocks. IP blocks include processor cores, memories, memory controllers, I/O interfaces, fixed-function units, etc. The on-chip interconnect integrates all the various IP blocks on a SoC. This on-chip interconnect is usually referred to as a Network-on-a-chip (NoC). The on-chip

network topology determines the physical layout and connections between nodes and channels in the network. The topology determines the number of hops (or routers) a message must traverse as well as the interconnect length between the hops, thus influencing network latency significantly. Many different NoC topologies for multiprocessor systems have been proposed and studied by researchers. For instance in [3], Balfour and Dally published a comprehensive analysis of several possible NoC topologies, such as mesh, torus, fat tree, concentrated mesh and a regular mesh topology. In this work we are going to consider our NoC to be a regular Mesh topology. A 5X5 NoC with mesh topology is shown below.

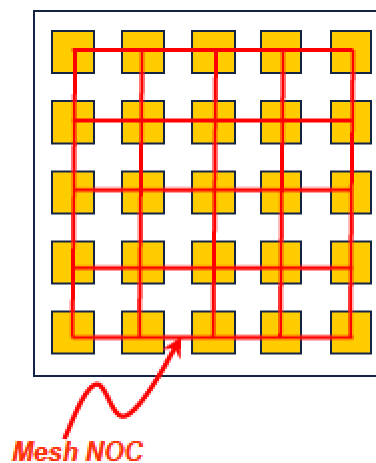

Figure 4 A 5X5 NoC

The above NoC connects a total of 25 cores. Each core is also called nodes. Each node contains a processor and

private level 1 and level 2 caches. The processor to network interface and the router serve as the gateway between the local tile and other on-chip components. NoC is often implemented as packet based communication. When a message is injected into the network, it is first segmented into packets, which are then divided into fixed length flits or flow control units. The packet will consist of a head flit that contains the destination address, body flits and a tail flit that indicates the end of a packet. These packets are interleaved on the links, thus improving link utilization. To transfer data from one core to another through NOC fabric, first it is packetized, then sent to the transmitting router, passed through the network wiring channel, delivered to the receiving router, and finally depacketized [24]

Switches route and buffer messages between resources. Each switch is connected to four neighboring switches through input and output channels that transport packets between nodes. Buffers are storage elements implemented within the nodes, such as registers or memories, and allow packets to be held temporarily at the nodes.

1.6 Advantages and Dis-Advantages of Network on Chip

A Network on Chip has the following advantages,

1. *Predictable electrical parameters enable high-performance circuits*. Unstructured wires have parasitic capacitances and crosstalk noise that are difficult to predict. As a result, in order to ensure reliability, very conservative circuits must be used to drive and receive these wires, leading to excessive power consumption. The well structured and predictable wires of a NoC allow for aggressive circuit techniques, which can reduce power dissipation by a factor of ten and increase wire propagation by three times, while also improving band-width.
2. *Universal interface facilitates reuse of components*. By introducing a universal interface for IP and the network, components can be reused in many systems, thus reducing complexity and simplifying circuit implementation.
3. *Design and testing are facilitated*. Since the system is modular and components are reused, design and testing of entire systems is mostly concerned with optimization of a regular, generic communication medium with predictable parameters. CAD issues involved in the design of dedicated,

customized circuits in specific components, such as wiring routing, are avoided.

4. *Duty factor of the wires is improved*. In traditional chip designs, individual signals must travel as fast as possible to their specific destination, leading to an excessive number of dedicated global wires which are active only 10% of the time, in average. The aggregated flux of information in general-purpose NoCs can provide wire duty factors close to 100%.

5. *Enable the use of fault-tolerant strategies*. With technology scaling and decrease in the voltage usage wires become more susceptible to noise and faults. Eventually, it will be impossible to completely avoid such errors (called upsets) in communication, and the system must be able to deal with them. NoC architecture can implement error-identification/error-correction protocols that make the system tolerant to faults.

6. *Wire pipelining*. Globally asynchronous protocols allow for wire pipelining, thus increasing bandwidth and making communication independent of latency.

7. *Scalability*. The NoC architecture is scalable; the aggregated bandwidth increases with network size.

The dis-advantages of a Network on Chip are latency and the resources spent in packetizing and depacketizing of the messages

Chapter 2

CNT for Interconnects

2.1 Interconnects in VLSI

“Interconnect is everything” was an often-used expression starting in the mid-1990s to characterize the importance of interconnect in deep submicron technologies. The purpose of interconnect is to establish communication between two points. Interconnects are used to connect components on a VLSI chip, connect chips on a multichip module and connect multichip modules on a system board. While device sizes were shrinking with each technology generation, multilevel metal structures rose higher and higher above the surface of the silicon and soon began to dominate the landscape of the integrated circuit. Wiring of chip devices takes place through various conductors produced during processing. In the sub 100nm scaling regime, interconnect behavior limits the performance and correctness of VLSI systems. The wiring in today’s integrated circuits forms a complex geometry that introduces capacitive, resistive, and inductive parasitics. These have multiple effects on the circuit’s behavior. They can cause an increase in propagation delay, impact on energy dissipation and the power distribution, and introduce extra noise sources, which affects the

reliability of the circuit. As the device size scales down the impact of interconnect in the VLSI circuits became even more significant. It controls all of these important electrical characteristics on the chip, even though the tiny devices controlled the actual logic functions. The dimensions associated with a cross section of interconnect are shown below,

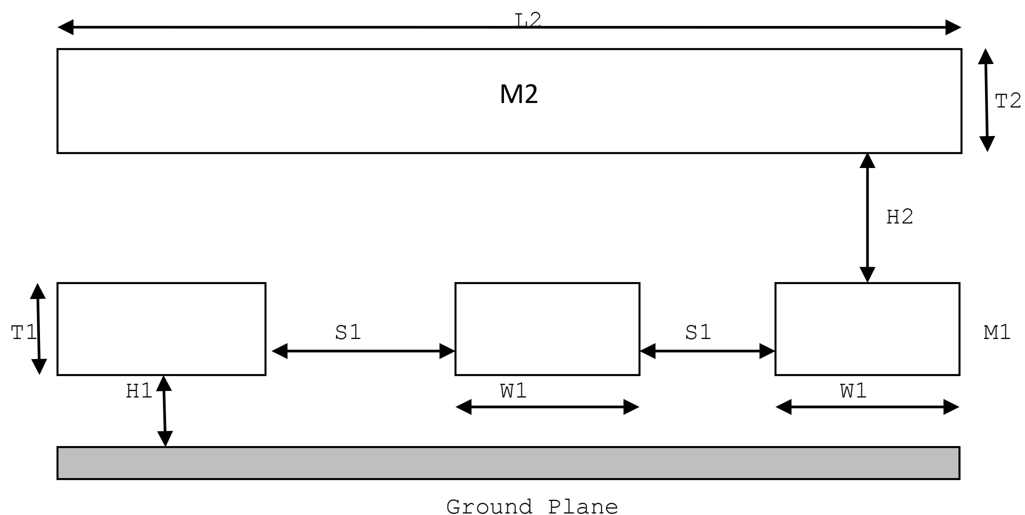

Figure 5 Interconnect dimensions

The rectangular wires are W wide and T thick, as shown. The separation between layers is a distance H (insulator thickness), and from other conductors is a distance S. The length of the wire, L is shown only for Metal 2. The vertical dimensions H1, T1, H2 and T2 are determined during the fabrication process. The designer has no control over these quantities. On the other hand the horizontal

dimensions W1, S1, L1, and L2. are all under the control of the designer.

2.2 Issues Concerning Interconnects

There are number of reasons to emphasize the importance of interconnect. First, as the width W of wires is decreased, the resistance increases. This increase in wire resistance causes RC delay to increase. The spacing S between wires has been decreasing to the point where the coupling between wires is significant. The resulting capacitive coupling introduces additional delay and noise effects that can cause failure in the design, requiring respins of silicon in order to fix the problem. The overall term for all these problems is signal integrity. Recent issues of inductance in wires have been included in the growing list of signal integrity problems.

Electromigration is another issue concerning interconnects. It occurs due to the hampering of crystal sources when there is high current flowing through the interconnects. It is a long term reliability issue and it mostly affects uni-directional nets causing a void or an open circuit.

2.3 Limitations of Copper Interconnects

In the past copper wire replaced aluminum wires due to the low resistance of copper wires when compared with aluminum wires and also the resistance to electromigration was much higher in copper when compared with aluminum, now copper wires are going through similar problems due to the increasing resistivity and as a result, wire delay is becoming serious concern especially when the processing technology approaching the sub nanometer regime. From the report of International Technology Roadmap for Semiconductors (ITRS) plotted below we find that copper resistivity for future technologies is increasing at a very fast rate[7].

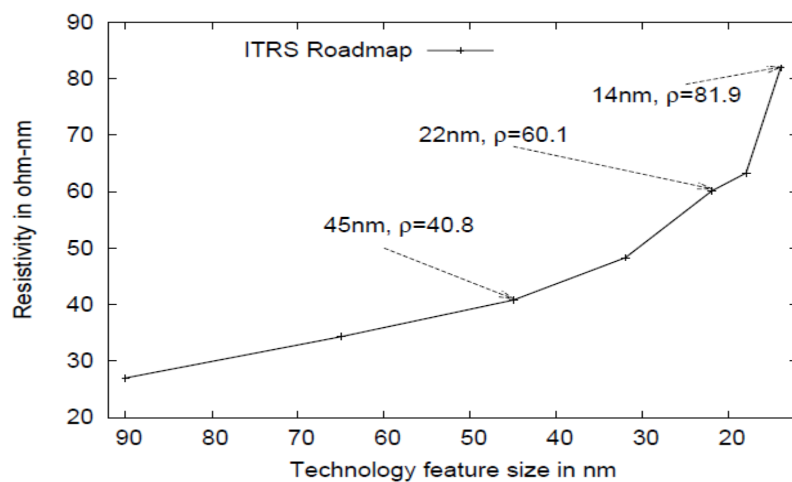

Figure 6 ITRS Roadmap showing copper resistivity[7]

We find that the increase in resistivity is not much when we move from 90nm to 32nm technology node, but as we reduce the feature size further from 32nm, we see a sharp resistivity increase due to scattering effects. For example, increase in resistivity is about 22% as we transition from 18nm to 14nm feature size compared to about 14% when we move from 45nm to 3nm technology node [7].

Besides increasing resistivity, the wire width is also shrinking with newer technologies. That further increases the overall resistance, since resistance of a wire is inversely proportional to the wire width. Therefore, even though the wire length is getting smaller, but decreasing cross section area and increasing resistivity resulting in higher interconnect delay, which is not a good sign as we always look to obtain higher speed. This might lead into serious architectural considerations in multicore systems. Copper interconnects does not scale with the data rate because of frequency dependent loss [8]. In [40] a comparison has been done between conductivity of Copper and SWCNT bundles for various lengths. One third of the nanotubes are assumed to be metallic, temperature is 100 degree Celsius and SWCNTs are 0.34nm apart. The plot is reproduced from [40],

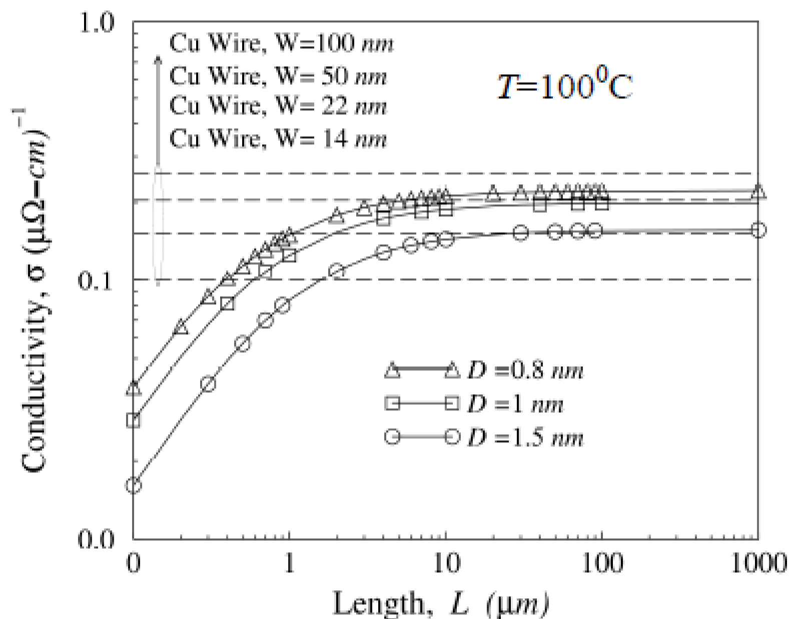

Figure 7 Conductivity comparison between Copper and SWCNT[40]

Hence as we look deeper into the circuit level issues, we find that as we slowly move from deep sub-micron technology to nanotechnology, the traditional copper wires will not be able to keep up, and we will need to look beyond conventional materials for interconnect design.

2.4 Carbon Nanotubes

Carbon Nanotubes popularly called by the acronym CNT was discovered by Sumio Lijima in 1991 [9]. The meaning of the name Carbon Nanotube is described below.

C: Carbon atoms hexagonally arranged.

N: Nanometers are the typical diameter dimensions.

T: Tube for guiding electrons from place to place, just like water pipe guides water. The following figure shows a Carbon nanotube,

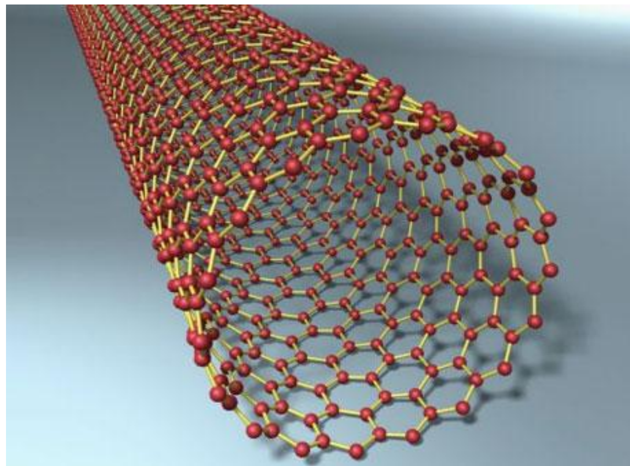

Figure 8 Carbon Nanotube

Nanotubes are composed of sp2 bonds, similar to those observed in graphite and they naturally align themselves into ropes held together by Van der Waals forces. These are allotropes of carbon with a cylindrical nanostructure. Nanotubes are members of the fullerene structural family which also includes the spherical buckyballs, and the ends of a nanotube may be capped with a hemisphere of the buckyball structure. Carbon Nanotubes (CNTs) are cylindrical carbon molecules with novel properties (outstanding mechanical, electrical, thermal and chemical properties: 100 times stronger than steel, best field

emission emitters, can maintain current density of more than 10^{10} A/cm2) [10-11].

2.5 Types of Carbon Nanotubes

CNTs are of two types namely, single walled carbon nanotubes (SWCNTs) and multi walled carbon nanotubes (MWCNTs). SWCNTs were discovered in 1993 and most of these have a diameter close to 1 nm, with a tube length that may be many thousands of times larger and up to order of centimeters. The structure of a SWCNT can be conceptualized by wrapping a one-atom-thick layer of graphite (or graphene) into a seamless cylinder. The way the graphene sheets wraps can be represented by a pair of indices (n,m) called the chiral vector. The relationship between n and m defines three categories of CNTs viz: arm chair, zigzag and chiral. The following picture shows Single Walled and Multi Walled Carbon nanotubes.

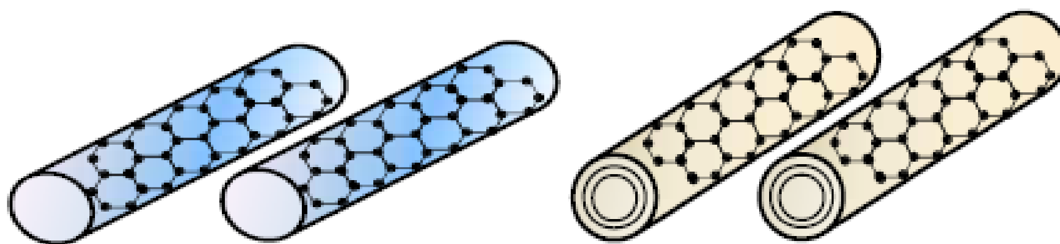

Figure 9 SWCNT and MWCNT

SWCNTs exhibit important electrical properties that are not shown by MWCNTs. The most basic building block of these systems is the electric wire and these are excellent conductors.

MWCNTs consist of multiple layers of graphite rolled in on them to form a tube shape with an interlayer spacing of 3.4 Å. The outer diameter of MWCNTs may range from 1 to 50nm while the inner diameter is usually of several nanometers. Two models are used to describe the structures of MWCNTs such as the Russian Doll model where the sheets of graphite are arranged in concentric cylinders and the Parchment model where a single sheet of graphite is rolled in around itself, resembling a scroll of parchment or a rolled up newspaper.

2.6 General Properties of Carbon Nanotubes

The electronic transport in metallic SWCNTs and MWCNTs occur ballistically over long lengths owing to their nearly one dimensional electronic structure. This enables nanotubes to carry high currents with negligible heating. It is reported that the MWCNTs can carry high current densities up to 10^9 to 10^{10} A/cm^2 and can conduct current without any measurable change in their resistance or morphology for extended times up to 250 degree Celsius

[9][10][11]. The electrical and electronic properties of nanotubes are affected by bending and twisting. They are the strongest and stiffest materials yet discovered in terms of tensile strength and elastic modulus respectively. This strength results from the covalent sp^2 bonds formed between the individual carbon atoms. Standard single walled carbon nanotubes can withstand a pressure up to 24GPa without deformation [10], [11]. Depending on the direction in which the carbon sheet is rolled up i.e. chirality they exhibit metallic or semiconductor properties. Due to lack of chirality any bundle of CNT consists of both metallic and semiconducting nanotubes. A Carbon Nanotube is an extremely versatile material; it is one of the strongest materials, yet highly elastic, highly conducting, small in size, but stable, and quite robust in most chemically harsh environments and it is hard to think of another material that can compete with nanotubes in versatility.

Electrical Properties

The unique electrical Properties of Carbon Nanotubes are, to a large extent, derived from their 1d character and the peculiar electronic structure of graphite. Because of the symmetry and unique electronic structure of graphene, the structure of a nanotube strongly affects its electrical

properties. For a given (n,m) nanotube, if n=m, the nanotube is metallic; if n-m is a multiple of 3, then the nanotube is semiconducting with a very small band gap, otherwise the nanotube is a moderate semiconductor. Thus all armchair (n=m) nanotubes are metallic, and nanotubes (6,4), (9,1) etc. are semiconducting(refer fig).

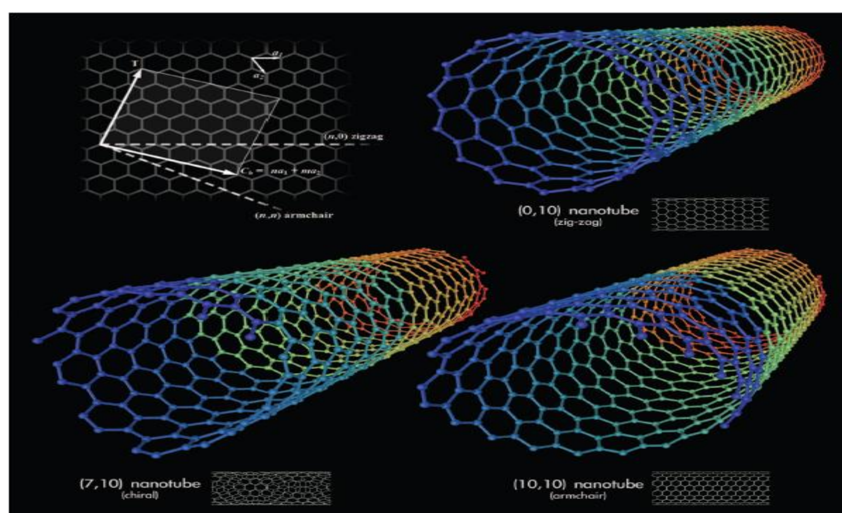

Figure 10 3-d Models of SWCNT types

Mechanical Properties

The carbon nanotubes are expected to have high stiffness and axial strength as a result of the carbon-carbon sp^2 bonding. Nanotubes are stiffest known fiber, with a measured Young's modulus of 1.3 TPa. They have an expected elongation to failure of 20-30% which combined with the stiffness, projects to a tensile strength well above 100 GPa, by far the highest known. For comparison, the Young's

modulus of high-strength steel is around 200 GPa, and its tensile strength is 1-2 GPa.

Thermal Properties

Prior to CNT, diamond was the best thermal conductor. CNT have now been shown to have a thermal conductivity at least twice that of diamond. CNT have the unique property of feeling cold to the touch, like metal, on the sides with the tube ends exposed, but similar to wood on the other sides. The specific heat and thermal conductivity of carbon nanotube systems are determined primarily by phonons. The measurements yield linear specific heat and thermal conductivity above 1 K and below room temperature while a $T^{0.62}$ behavior of the specific heat was observed below 1 K. The linear temperature dependence can be explained with the linear k-vector dependence of the frequency of the longitudinal and twist acoustic phonons. The specific behavior of the specific heat below 1K can be attributed to the transverse acoustic phonons with quadratic k dependence. The measurements of the thermoelectric power (TEP) of nanotube systems give direct information for the type of carriers and conductivity mechanisms.

Chapter 3

CNT in Network-on-Chip

3.1 Carbon Nanotubes (CNT) Interconnect

Carbon Nanotubes are viewed as a potential replacement for copper wires due to its desirable properties such as high thermal conductivity, thermal stability, and large current carrying capacity. Metallic Single Walled Carbon Nanotubes also have resistance to electromigration which is a very good property when compared to copper interconnects. As electromigration causes long term reliability issues Metallic SWCNT's are better option for gigascale interconnection. A Carbon Nanotube is very close to an ideal quantum wire in which electrons can be moved in one dimension only. The phase space for scattering in nanotubes is therefore very limited; electrons can be scattered only backward. The mean free path in high quality carbon nanotubes is therefore in the micrometer range [13].The one dimensional nature of nanotubes however causes a high quantum resistance. Therefore, nanotubes are connected in parallel which helps to lower the overall resistance and inductance. Previous works have showed that to outperform copper interconnects, bundles of densely packed nanotubes should be used to lower resistance and make the signal

travel time small compared to the RC charge-up time. There have been several models for Carbon Nanotubes. The circuit model used for an isolated single CNT is generally accepted [14]. The equivalent circuit model for an ideally contacted CNT isolated above a ground plane is shown below.

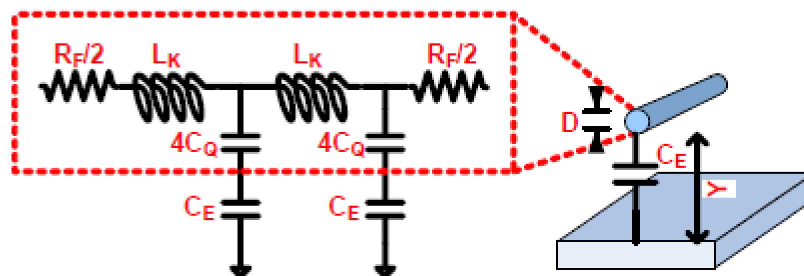

Figure 11 Equivalent Circuit Model for Ideally Contacted SWCNT [14]

The parameters for the circuit model are R_F, the resistance of CNT, L is the Length, L_k is the kinetic inductance, y is the distance between CNT and ground plane and d is the CNT diameter, C_Q and C_E are quantum and electrostatic capacitance respectively.

The fundamental resistance associated with a single CNT scales linearly with the length for nanotubes longer than the mean free path. Meanwhile, the resistance associated with a bundle of CNTs is determined by the size and number of CNTs in the bundle. Taking into account contact resistance, and the effective fraction of contacted CNTs, the effective resistivity is calculated as [15].

$$\rho = \frac{d^2}{k}\left(\frac{h^2}{4.e^2.L_o} + \frac{R_{cont}}{L}\right)$$ for L>=Lo. . . . . (3.1)

$$\rho = \frac{d^2}{k.L}\left(\frac{h^2}{4.e^2.} + R_{cont}\right)$$ for L<Lo. . . . . (3.2)

Where h corresponds to Planck's constant, e to electronic charge, Lo is the mean free path which is usually in the range of micrometers, d is the nanotube diameter and k is the fraction of contacted metallic CNTs in the bundle. The effective resistivity of CNTs is calculated based on assuming Lo, d and k as 1μm, 1nm and 0.33 respectively.

Calculating capacitance of CNT has been done in the works in [16] and [17]. The cross section of copper interconnects, monolayer nanotube interconnects above a thick dielectric layer are shown below,

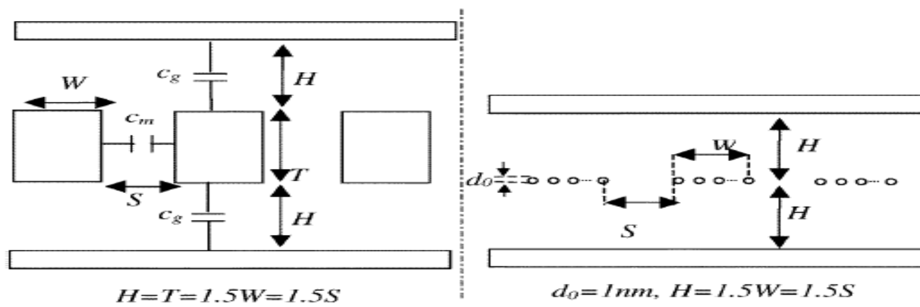

Figure 12 Configuration of Copper and CNT Interconnects [16]

The width and spacing of interconnects in both the cases are assumed to be equal to the minimum feature size. The aspect ratio of copper wires is assumed to be 1.5 times the wire width to avoid electromigration and also accommodate thickness variations due to CMP. The dielectric thicknesses are assumed to be 1.5 times the wire width. Per unit length values of capacitance to ground and capacitance between adjacent interconnects for each case are calculated using Raphael [16]. The results are tabulated below.

	Copper Wires	SWCNT
Capacitance to ground, cg	**35.6 aF/µm**	**27.2 aF/µm**
Line to Line Capacitance, cm	**38.6 aF/µm**	**9.9 aF/µm**
Average Capacitance, 2cm+2cg	**148.5 aF/µm**	**74.5 aF/µm**

Table 3.1: Capacitance per µm for Copper and SWCNT [16]

As interconnect length increases scaling comes into play. The optimal wiring width is defined as the width at which the bandwidth per unit width reciprocal latency product is maximized. It can be seen that the average capacitance per unit length for a monolayer SWCNT above a thick dielectric layer is almost half of that of Copper wires. The above mentioned advantages of monolayer nanotube interconnect in terms of capacitance values remains constant at various

generations of technology as long as the cross sectional dimensions scale proportionally with technology.

3.2 Delay Calculation for CNT and Copper:

As the resistivity of copper increases very rapidly as the processing technology scales down, there have been steady increases in the resistance of the same. The resistance and capacitance of CNT and Copper in 22nm node for various lengths are found [17 [23]. From the capacitance and resistance values obtained we calculated the RC delay of copper and CNT for 22nm node and for various lengths. The delay is calculated as the function of driver resistance, driver capacitance, wire resistance, wire capacitance, and also load capacitance. The equivalent circuit model used to calculate delay is given below,

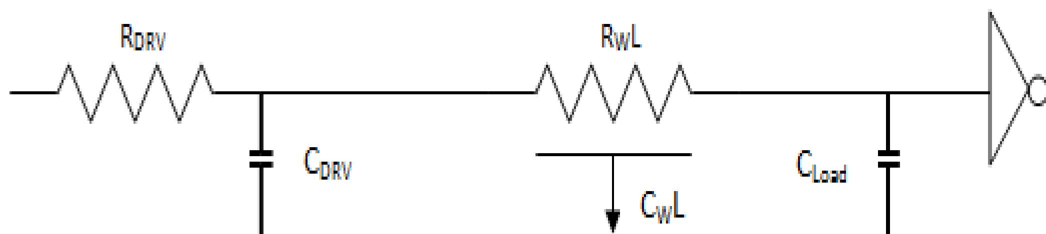

Figure 13 Equivalent circuit model for delay characteristics

The delay is calculated by [15],

$$\tau = R_{DRV}(C_{DRV} + C_{Load)} + 0.4R_W.C_W.L^2 + (R_{DRV}.C_W + R_W.C_{Load}).L \quad . \quad . \quad (3.3)$$

The plot of delay comparison between copper and carbon nanotube are given below,

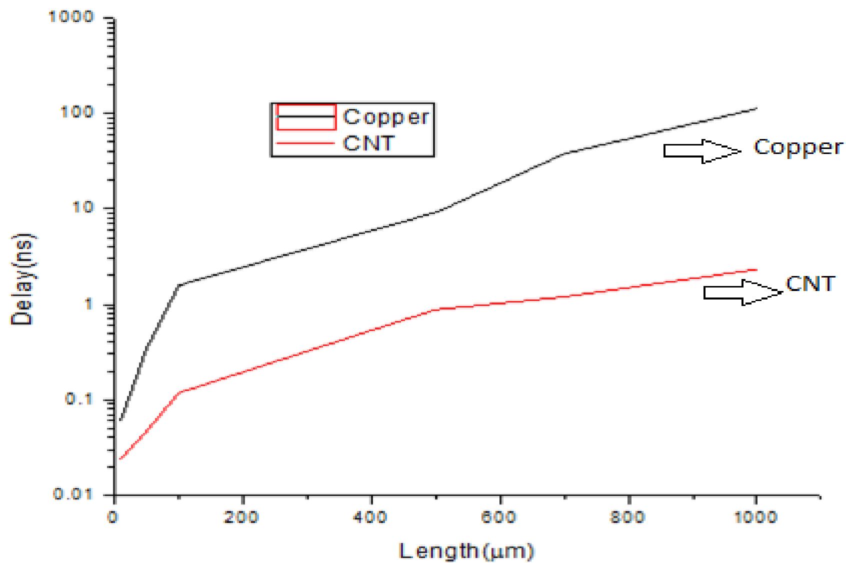

Figure 14 Delay comparison between copper and CNT

The delay calculations shows that in the local interconnect level of short lengths the improvement of delay in SWCNT is 3X times when compared to Copper. Whereas in global interconnect level of higher lengths SWCNT completely outperforms copper. Recently Ashok Srivastava, YaoXu in their paper titled Carbon Nanotubes for Next-Generation Interconnects [20] [21] performed similar delay calculation by connecting Copper, SWCNT and MWCNT between two inverter pairs. They performed delay calculations in 22nm node for

various interconnect lengths and for bundle efficiency varying between 1 and 0.33. The simulations were done through cadence spectre and their result is shown below,

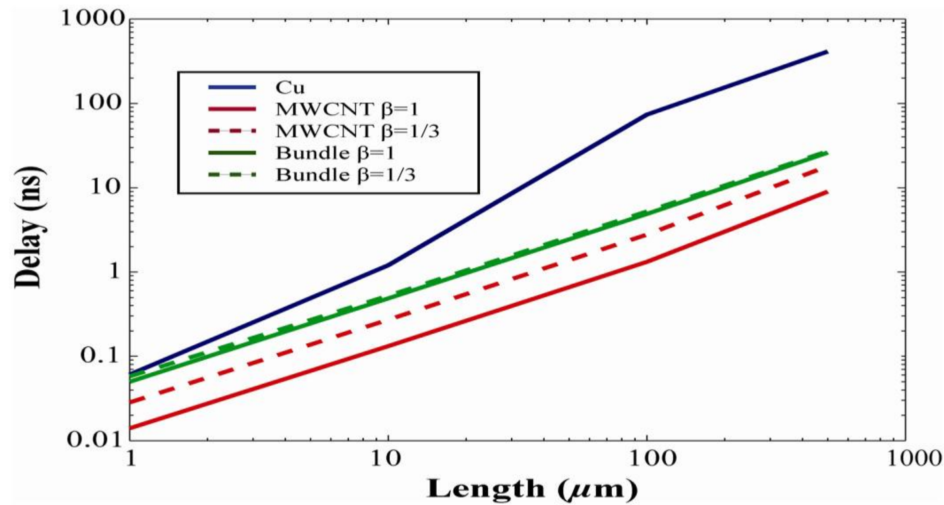

Figure 15 Delay Comparison for Cu, Bundles and MWCNT [20][21]

The results in the above plot match with the results of our design. Here Single Walled Bundles with an efficiency of 0.33(similar consideration of our design) completely outperforms copper interconnects at global level, thus confirming the earlier results mentioned in this thesis. Recalling the fact that we used Single Walled Bundled Carbon Nanotubes with efficiency of 0.33 meaning 1 out of 3 nanotubes used in the bundle are considered to be metallic. The above plot shows that if the efficiency is more in Single Walled Nanotubes then the delay will be even further reduced. But research is still going on in achieving higher

efficiency for Single Walled Nanotubes. We considered the diameter of the nanotube to be 1nm. It is known that if the diameter of the CNT is increased then the capacitance will be further reduced thereby reducing the overall delay. For long interconnects buffers should be inserted to reduce latency. However these performances depends on the number of metallic tubes present in the bundle and level of interconnect i.e. local, intermediate and global, since resistance decreases with increase in number of tubes while capacitance increases at the same time.

3.3 Calculations for a Network on Chip

This section explores on bandwidth and throughput calculations performed on a Network on Chip model with two different interconnects, in our case Single Walled Carbon Nanotube bundles and Copper. For our case study, we analyzed bandwidth, throughput and energy consumption in a 4 X 4 NOC.

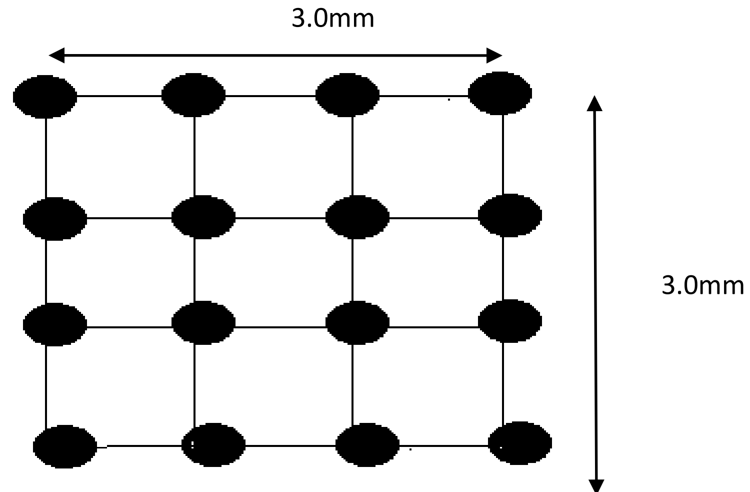

Figure 16 4 X 4 NOC with Mesh Topology

The 4 X 4 NOC considered for our calculations has a die dimension of 3.0mm X 3.0mm. We now evaluate the performance characteristics like bandwidth and throughput for the NOC.

3.4 Bandwidth Calculation for CNT in NOC:

To calculate bandwidth we should know the number of available channels for the nearest routers to communicate. As the length and width of the die is considered to be 3.0 mm and 3.0 mm and there are 4 cores in the length we can consider the length of the nearest neighbor or the length between two routers for the calculation of bandwidth. As shown in the following figure, the inter core communication between two routers are shown,

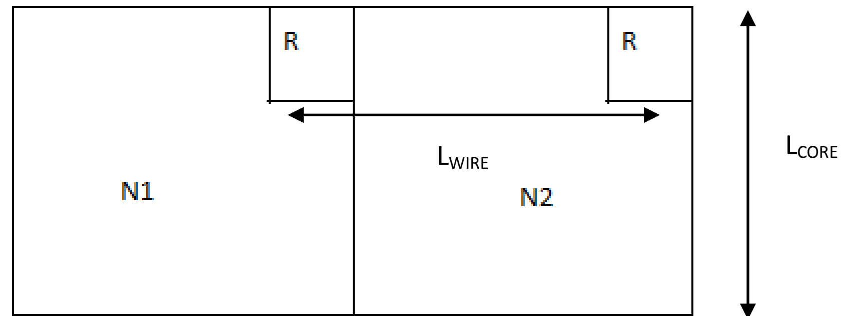

Figure 17 Intercore Architecture

We assume that the interconnect length connecting two different routers (L_{WIRE}) is equal to the length of a core edge (L_{CORE}). The maximum number of available routing channels between the two adjacent cores is determined by the interconnect pitch P, and the core edge length. The number of available routing channels between two cores is given as follows,

$$N_{ch} = L_{core}/\text{Interconnect Pitch} \quad (3.4)$$

The Interconnect pitch for 22nm for global interconnects is 66nm (ITRS). The number of channels for Copper and CNT for the two neighboring nodes is calculated and the aggregate bandwidth of the Network on Chip is calculated from the number of available channels and the bandwidth per wire. The following figure shows the bisection of the channel,

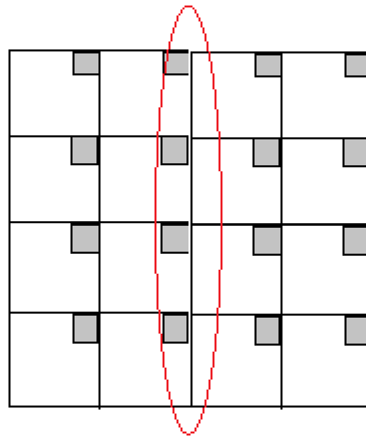

Figure 18 Channels determining bisection bandwidth

Length of wire	700 (μm)
Width	110nm
Spacing	110nm
Thickness	330nm
Aspect Ratio	3

Table 3.2: Wire Parameters for CNT and Copper

The following figure shows the aggregate bandwidth of the 4 X 4 NOC for CNT vs Copper is plotted below,

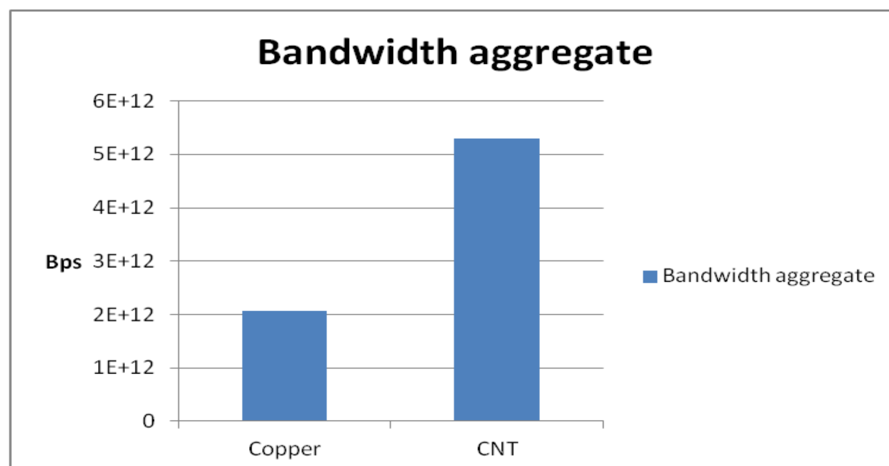

Figure 19 Aggregate Bandwidth of 4 X4 NOC

For 22nm technology the number of available routing channels between two adjacent channels changes for CNT and Copper. Due to the interconnect pitch the number of available routing channels were low when compared to Copper but still due to the high bandwidth per wire, SWCNT showed good improvement in the overall bandwidth of the NOC. The aggregate bandwidth was calculated by multiplying the number of channels in both the direction between the cores and the bandwidth per wire. Carbon Nanotubes showed (2x) times aggregate bandwidth than Copper.

3.5 Total System Throughput

While the aggregate bandwidth gives a measure of the total capacity of all the links in the system, the true measure of a system is given by the maximum throughput the links can sustain without any bottlenecks. In the case of a mesh network, the maximum dataflow occurs at the bisection of the system. This is the path used by 50% of data generated by each core, when we assume uniform traffic i.e. each core sends a message to every other core with equal probability.

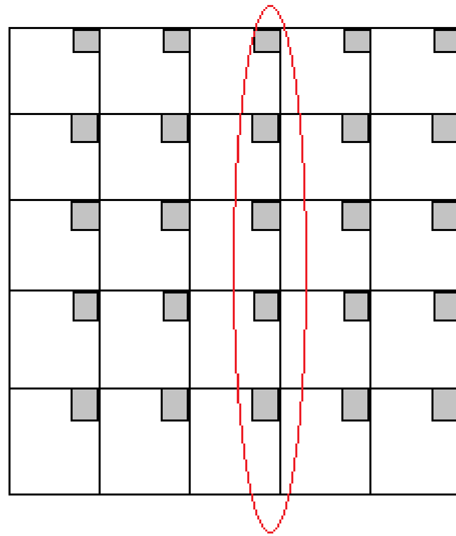

Figure 20 Channels determining Bisection Bandwidth for 5 X 5 NoC

To avoid any bottlenecks, the total bandwidth available at this bisection must be able to support the total traffic generated by the cores. Hence the bisection bandwidth determines the total maximum throughput and is given by,

$$BW_{bisec} = BW_{agg}\sqrt{N} \quad \text{. . . . .} \quad \textbf{(3.5)}$$

Where BW_{agg} is the aggregate bandwidth between two cores and N is the total number of cores on the die. For a fixed system area the number of cores in the system will determine the core edge dimensions and inter-core channel length, which in turn determines the aggregate core-core bandwidth achievable. The throughput of the studied NOC under consideration has been calculated and the following figure shows the throughput comparison between the NOC

using CNT and Copper Wires. Parameters for throughput calculation are given in the following table,

No. of Cores	16
Die Size	3.0mm X 3.0mm
Clock Frequency	1 GHz

Table 3.3: Parameters for throughput calculation

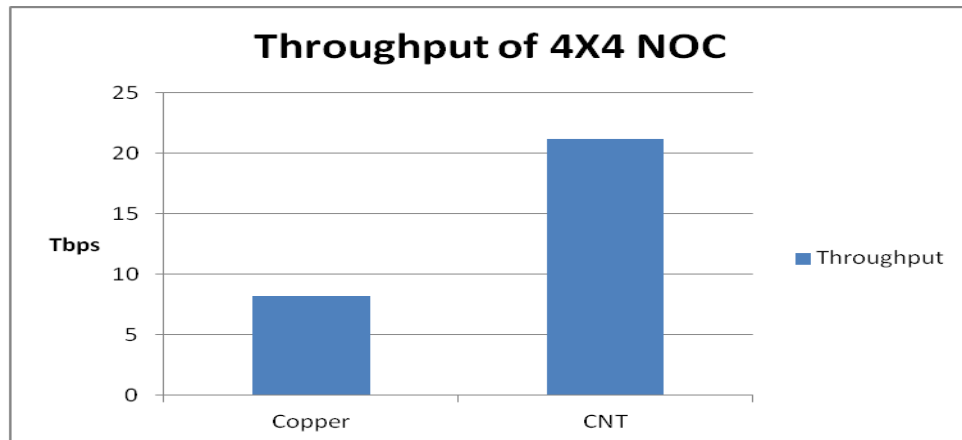

Figure 21 Throughput Comparison of 4 X 4 NOC

If the number of cores in the multi core system increases then the throughput of the system increases. Because the total throughput of the system depends on the bisection bandwidth which inturn depends on the bandwidth aggregate and the number of cores in the bisection.

Chapter 4

Energy Consumption of Network on Chip

4.1 Communication Probability Distribution

Using derivations based on Rent’s rule, the wire length distribution of a VLSI circuit can be estimated from its Rent’s exponent and coefficient, p and k[25]. This distribution is relevant to VLSI design and implementation because it is related to many properties of the System, such as chip area, signal delay, power consumption, and wire routability. In system-on-chip similar information is provided by the Communication Probability Distribution(CPD) of applications. This CPD is used to model communication locality and energy consumption in NOC. In Rent’s rule based traffic generator, the probability of communication between cores is derived direclty, which results in CPDs displaying high locality. The CPD describes the probability that packets will travel a certain distance in the Network on Chip for a given traffic pattern. Since current NoCs use 30 to 40% of the power budget [26] [27]. This distribution is directly related to the energy consumption of an application, because the larger the distance traveled by packets, the more energy is used. It is desirable that the

distance travelled by packets to be as small as possible in order to minimize this cost.

4.2 Rent’s Rule Traffic Patterns

In VLSI, Rent’s rule emerges naturally from circuit placement, in which connections are made as local as possible to minimize wire footprint, power and latency [28]. E.F Rent of IBM published two internal memoranda in 1960 that contained the log plots of “number of pins” versus “number of circuits” in a logic design[39]. These data tend to form a straight line in a log-log plot and yield the following relationship,

$$N_p = K_p . N_g^{\beta} \quad . . . . (4.1)$$

Similar constraints apply to the communication among processors in multi and many core systems. Algorithms used for mapping parallel applications onto cores aim at producing optimized layouts that minimize communication distances.

Greenfield et al [29] argue that, analogous to circuit placement in VLSI, Rent’s rule will naturally arise in multi and many core chips from this optimization process. They extended the concept of conncection locality in

circuits to communication locality among cores, proposing a bandwidth based version of Rent's rule,

$$B = bN^p \quad \text{.........(4.2)}$$

Where B is the bandwidth sent or received by a cluster of N network nodes, b is the average bandwidth per node and Rent's exponent is p which lies between 0 and 1. In recent work, Heirman et al. [30] showed that many parallel applications indeed follow Rent's rule. For generating Rent's rule traffic pattern a formula is generated for finding the probability of wire connecting two terminals with manhattan distance d.

The probability of a wire connecting two terminals with manhattan distance d is given by [25],

$$P(d) = \frac{1}{4d}\left[(1 + d(d-1))^p - (d(d-1))^p + (d(d+1))^p - (1 + d(d+1))^p\right] \quad \text{.........(4.3)}$$

We use the above equation to define the probability of communication between two processors, where d corresponds to the number of hops in the shortest path between source and destination.

The traffic is generated using the above equation and then the resulting CPD is measured. The formula for the CPD of

synthetic Rent's rule traffic can be derived from the above equation and is given by[31];

$$CPD(d) = \Gamma P(d) \cdot \sum_{i=1}^{2\sqrt{N}-2} (\sqrt{N} - i)(\sqrt{N} + i - d) \quad \text{.......................................} (4.4)$$

For $0 < \left(\sqrt{N} + i - d\right) \leq \sqrt{N}$

Where Γ is the normalization coefficient such that

$\sum_{d=1}^{2\sqrt{N}-2} CPD(d) = 1$.

An advantage of this method is the ability to generate traffic pattern with arbitrary Rent's exponents. Because the Rent's exponent is related to communication locality and complexity of applications, it is possible to study the NOC under several application scenarios by varying a single parameter in the model. The following figure shows the CPD produced by the generator on an 8X8 mesh network. In our calculations for Energy consumption N=25 and the maximum value of d is 6.

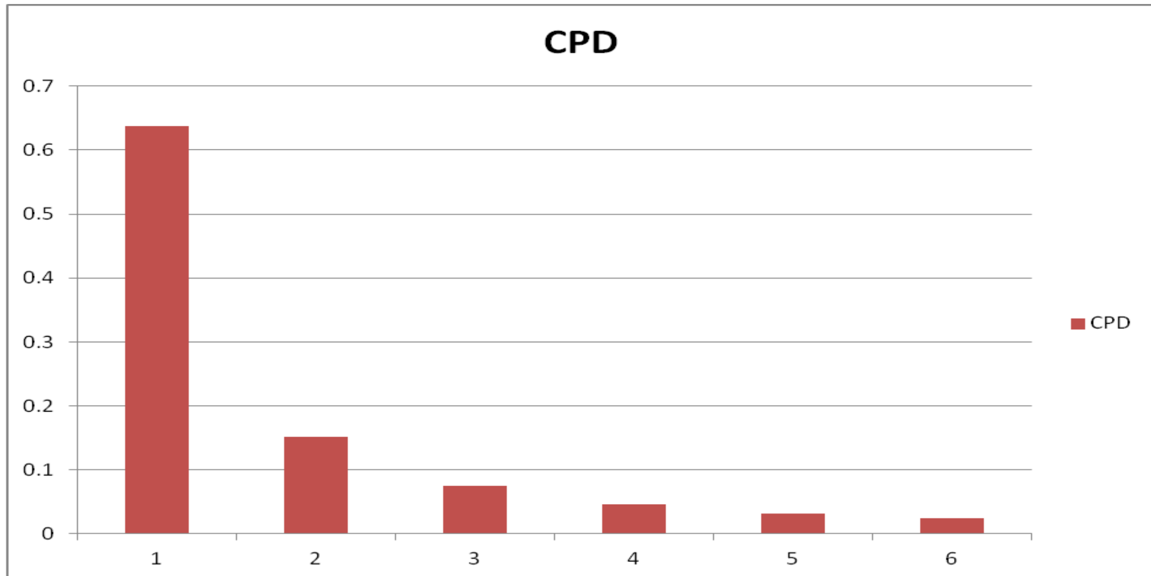

Figure 22 CPD for a 4X4 NOC

The above plot is the communication probability distribution for a 4 X 4 NoC. We see that almost 65% of the communication between local neighbors which has a d of 1 meaning 1 hop distance.

4.3 Modeling Energy Consumption

It can be computationally expensive to analyze NoC energy consumption using simulations, especially with application driven workloads or large system sizes. In this section a simple model for predicting energy consumption based on the CPD, which does not require computer simulations. This model is intended for direct networks in which the length of the wires is the same for every hop, such as mesh and

folded torus, but it could be easily extended to other topologies.

The average energy of a flit traversing a path of length d in the network is given by [31],

$$E_{flit}(d) = d \cdot E_{link} + (d+1) \cdot E_{router} \quad (4.5)$$

Where E_{link} and E_{router} are the energy consumed by the flit when traversing a link and a router, respectively, and d is given by the number of hops traversed in the path. The total energy consumed by the application is obtained by first summing E_{flits} over all communication distances and weighted by the probability of a packet travelling that distance. This value is then multiplied by the number of flits per packet and the total number of packets.

$$E_{total} = N_{packets} \cdot N_{flits} \cdot \sum_{d=1}^{max} E_{flit}(d) \cdot CPD(d) \quad (4.6)$$

In the equation to find the Energy of the Flit, The Energy of the Link (E_{link}) and the Energy of the router (E_{router}) are obtained from architecture level power model called Orion.

This model’s ability to predict energy usage for Rentian traffic based on a single application parameter could significantly simplify and speedup NOC energy analysis.

A potential limitation of this method is the assumption that the energy used for communication is proportional to the distance traveled by packets. This is approximately true for most networks on chip and is commonly used in the literature as simplification step. However, contention in the network could lead to extra dynamic and static energy that are not accounted for the model.

4.4 Orion-Network on Chip Simulator

Orion is a power performance interconnection network simulator that is capable of providing detailed power characteristics, in addition to performance characteristics, to enable rapid power performance trade-offs at the architecture level [32]. Orion does both power modeling and area modeling. It has files to specify micro architectural parameters and technology parameters. For modeling the 4 X 4 NOC with CNT and Copper we chose the flit size to be 64 bits, packets have 5 flits each. The router configuration was specified. The capacitance values for CNT and Copper were specified in the SIM_link.c file. The link power and router power are obtained by ./orion_link l 1 command, where l specifies the Link Length in µm. The command is explained in detail as follows,

Command: orion_link <link_length> <load>

link_length: the length between two routers is in µm. Load is in the range of (0,1). Here load indicates the probability at which flits traverse the links. Note that this returns link power and area for links connected to the input port part of the router. It doesn't account for the links at the output ports because we assume that the output link power can be calculated at the router to which it is connected. The values of E_{flit} and E_{Router} are obtained and then from equation (4.6) the energy consumption of the NOC is calculated. The procedure to check the values for Carbon Nanotubes in Orion is rewriting the codes and input files given to Orion. The input files to change the microarchitectural parameters and to change the interconnect parameters are then given to Orion and then running make command updates the router area, link and router energy consumption.

The following plot is the comparison of energy Consumption between carbon nanotubes and Copper interconnects. The parameters for the following plot is given in the following table,

No. of Cores	16
Die Size	3.0mm X 3.0mm
Clock Frequency	1 GHz
Flit Size	64 bits
Packet Size	5 Flits
Number of Packets	10,000
Rents exponent	0.75

Table 4.1: Parameters for calculating Energy Consumption

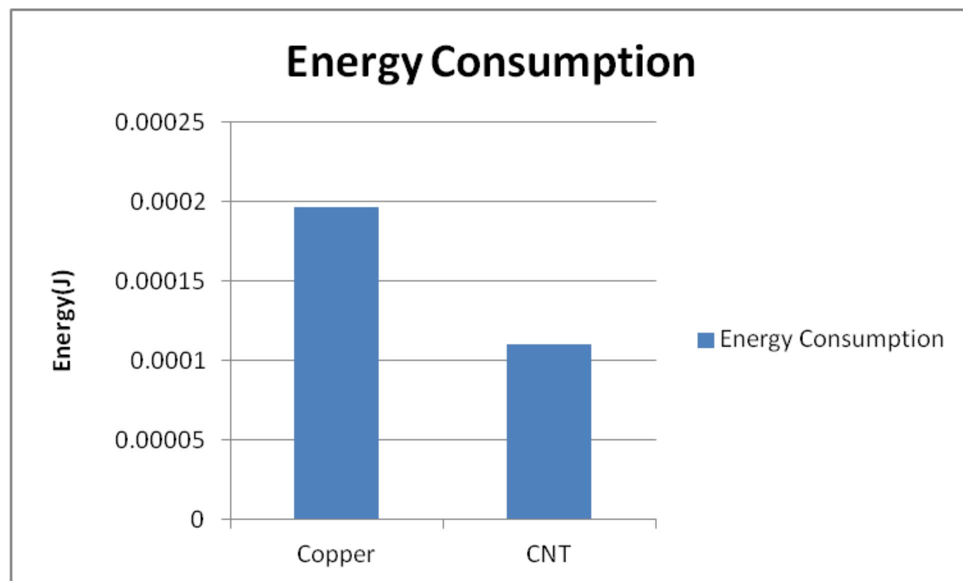

Figure 23 Energy Consumption Comparison for 4 x 4 NoC

We see that the energy consumption for a 4 X 4 NoC implemented with carbon nanotubes as interconnects is 2x less than a 4 X 4 NoC implemented with copper. Hence carbon nanotubes when used as an interconnect for Network-on-Chip provides 2.7x improvement in throughput and 2x reduction in energy consumption.

Chapter 5

Hybrid NoC

5.1 Challenges and Potential for Carbon Nanotube Applications

Carbon Nanotubes have come a long way since their discovery in 1991. The structures that were first reported were MWNTs with a range of diameters and lengths. These were essentially the distant relatives of the highly defective carbon nanofibers grown via catalytic chemical vapor deposition [32].

1. There are general challenges that dace the development of nanotubes into functional devices and structures. First of all, the growth mechanism of nanotubes, similar to that of fullerenes, has a challenging issue of controllability especially controlling the diameter of the nanotubes. Especially for electronic applications, which rely on the electronic structure of nanotubes, this inability to select the size and helicity of nanotubes during growth remains a drawback.

2. There is no controllable way, as of yet, of making connections between nanotubes. Some recent reports, however, suggest the possibility of constructing these

interconnected structures by electron irradiation and by template mediated growth and manipulation.

3. For bulk applications, such as fillers in composites, where the atomic structure has a much smaller impact on the resulting properties, the quantities of nanotubes that can be manufactured still falls short of what industry would need. There are no available techniques that can produce nanotubes of reasonable purity and quality in kilogram quantities. The market price of nanotubes is also too high presently.

4. Much more challenging issue of device manufacturability is the control of chirality. As a result, the tubes are a mixture of metal and semiconductors. In CVD, the general location for tube growth can be controlled by patterning the catalyst material, but the number of tubes and their orientation relative to the substrate are still not well defined [33-35].

5.2 Fabrication of Nanotubes

Carbon Nanotubes fabrication has been going on for more than a decade and interest in the production and fabrication increased worldwide due to its possible technological applications. Making a carbon nanotube is not as easy as picking up a graphene sheet and rolling it up. We don't have any tools small enough to do that. Instead, we have to grow them like a plant [35]. These Carbon Nanotubes can be building blocks for microscopic transistors and similar electronic devices in the future. One of the methods of growing nanotube is called chemical vapor deposition. A carbon source like methane is heated and catalyst particles usually iron or nickel acts like seeds from which nanotubes grow. In Cornell Nanoscale Facility (CNF) carbon nanotubes are grown using chemical vapor deposition [35]. Iron catalyst particles are put only at the places where nanotubes are to be created. Then a carbon containing gas like methane is flown over them inside a hot furnace. The hot carbon binds to the catalyst particles, and a carbon nanotube extrudes. Then electrical contacts are made to the nanotubes allowing us to determine if a nanotube is metallic or semiconducting [35]. The other method is electric-arc discharge method [36]. An electric

arc is an electrical breakdown of a gas which produces an ongoing plasma discharge, similar to the instant spark, resulting from a current flowing through normally nonconductive materials such as air. The arc occurs in the gas filled space between two conductive electrodes and it results in very high temperature, capable of melting or vaporizing just about anything. So this process takes place like this: 1)a current is run through an anode, or a positively charged piece of carbon, 2) then this current jumps through a certain type of plasma material to a cathode, or a negatively charged piece of carbon, where there is an evaporation and deposition of carbon particles in through the plasma, 3)Finally an outer hard shell region made of decomposed graphite is formed and an inner core region with loosely packed columns which consist of straight, stiff multishell carbon nanotubes and closed polyhedral particles also known as carbon nanotube particles. To obtain single shell carbon nanotubes, a catalyst must be added to the evaporated carbon. This catalyst is commonly a metal such as cobalt, nickel, or a mixture of certain other metals. This metal catalyst along with graphite powder is added in a hole drilled through the anode contact. During the arc-discharge, web-like

structures are formed around the cooler parts of the electrodes. Within these structures, bundles of 10-100 single shell nanotubes are formed. This particular method is normally very inefficient, but the use of nickel-yttrium catalyst has improved the efficiency and overall production of single shell nanotubes [36].

There are two important fabrication issues to solve regarding this new technology: 1. To grow nanotubes of useful lengths and 2. To assemble them in the form of transistor like junctions. One of the most used forms of growing nanotubes today is a technique called Chemical Vapour Depostition. In one of the CVD variants, Carbon Nanotubes grow in a steam of gases blown across catalysts on silicon wafers [37].

5.3 Hybrid NoC Architecture

A hybrid network-on-chip (HNOC) has been proposed in [24] which uses standard NoC topology for packet-based global interconnections along with local buses for nearest neighbor communications as shown in the following figure,

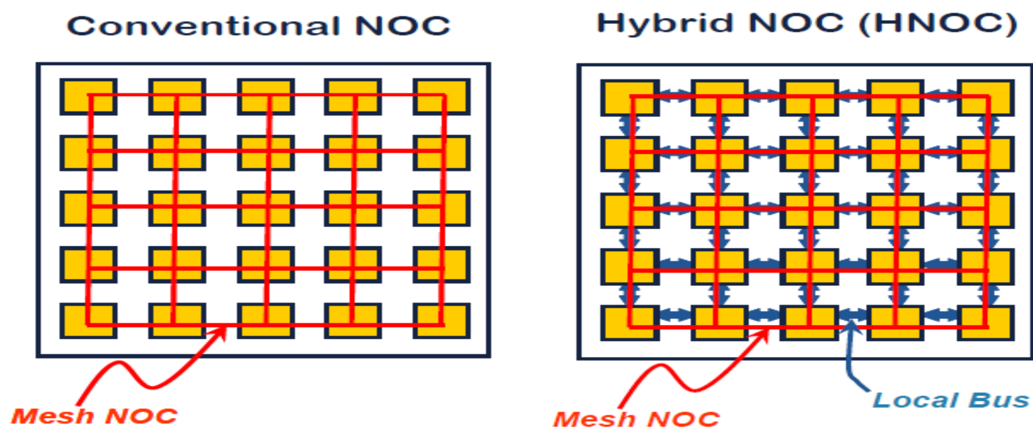

Figure 24 Conventional NoC and Hybrid NoC

Unlike the hybrid optical/electrical NoC architecture that is purely packet based, the HNoC uses local buses to transmit data directly to the nearest neighbors in a parallel fashion, which eliminates the need for serializer, router, and deserializer. Moreover, since the local bus interconnects are short, they inherently exhibit lower loss and therefore can provide higher bandwidth and consume less power. Local bus interconnects are direct connections between neighboring cores dedicated for direct data exchange without any packetizing overhead. The energy consumption and latency for short distance communication

through local buses are therefore much smaller than those through NoC fabric.

5.4 Throughput and Energy Consumption Analysis

The communication probability distribution of a 4 X 4 NoC presented in the previous chapter gives us the value of probability of communication between nearest neighbors. If 64% of the communication can be moved to local buses, the NoC will be responsible only for 36% of the traffic. Therefore the throughput of the NoC can potentially improve by 2.6x for the maximum injection rate. In general, the rate of throughput improvement is determined by[24],

$$\frac{HNOC_{Throughput}}{NOC_{Throughput}} \approx \frac{1}{1 - CPD(1)} \quad \text{.....................(5.1)}$$

The improvement rate is calculated assuming p=0.75, and it has been calculated for an ideal case where that the local buses impose no overhead, and the above equation presents an ultimate HNoC benefit without making too many assumptions, which may be application or design dependent. In practice, however depending on the design and application, the local bus overhead will impact the performance of HNoC.

Similarly, energy consumption can be reduced by introducing local buses. Again, consider the 4 X 4 array of multiprocessor system with HNoC, The CPD shown in figure can be used to compute the energy reduction rate in HNoC. In general for the same throughput the energy reduction rate in an array of N X N processors is approximated by [31]

$$\frac{NOC_{Energy}}{HNOC_{Energy}} \approx \frac{\sum_{d=1}^{2N-2} d \cdot CPD(d)}{\sum_{d-2}^{2N-2} d \cdot CPD(d)} \qquad \text{..................(5.2)}$$

Using the NoC energy model presented in [31] and assuming that the power consumption of routers is dominant, the energy consumption of the HNoC against the conventional NoC for the same throughput can potentially be reduced by factor of 1.6x. Similar to throughput improvement analysis, the energy consumption model presented here is for an ideal case, where it is assumed that the local buses impose no overhead. In practice, however, depending on the design and application, the local bus power overhead will impact the power consumption of HNoC.

5.5 A Hybrid NoC with Copper and Carbon nanotubes

In the previous section we saw about Hybrid NoC which utilizes the Network on chip for the global connection and bus connection for nearest neighbors. As we saw in the previous section about the bottle necks in fabricating carbon nanotubes we propose a modified hybrid NoC with copper for local bus connections and carbon nanotubes for global connection.

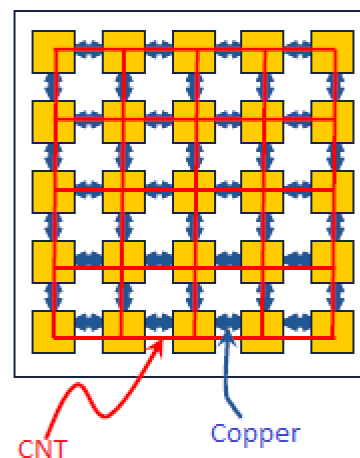

Figure 25 Hybrid NoC with Copper and CNT

Even though carbon nanotubes exhibit better performance when compared to copper wires, the fabrication challenges and cost of production prohibits the research community to utilize nanotubes in existing VLSI chips. As the nearest neighbors are communicated more in a Network on Chip, we propose a hybrid NoC with copper for local bus and carbon

nanotubes for global connection. The proposed hybrid NoC with both copper and carbon nanotubes as interconnects will offer better throughput and less energy consumption when compared with conventional NoCs. The throughput and energy consumption are predicted from the calculations made for conventional NoC using Carbon nanotubes and Hybrid NoC with copper as interconnects for both local bus and global connections. Equations 5.1 and 5.2 are used for calculating the throughput and energy consumption of Hybrid NoC with copper and carbon nanotubes as interconnects. The results are tabulated below,

NoC Parameters	CNT NOC	Hybrid NoC with Copper	Hybrid NoC with Copper/CNT
Throughput Improvement	2.7x	2.6x	6x
Energy consumption reduction	2x	1.6x	3x

Table 5.1: Summary of Results

Chapter 6

Conclusion and Future Work

As technology scales down the interconnects affect the delay and bandwidth of a multi core system to a large extent. Thus Single Walled Carbon Nanotubes in bundles seem to be a potential candidate in nanoscale regimes when compared to traditional copper wires. When Implemented in a Network on Chip Single Walled Carbon Nanotubes showed approximately 2.6x more throughput and 2.0x less energy consumption than traditional copper wires. When implementing single walled carbon nanotubes in a hybrid NoC the throughput improvement is 6x and the energy consumption reduction is by a factor of 3x when compared to conventional NoC. Hence Single Walled Carbon Nanotubes can be used as interconnects in nanoscale regime. However for successful implementation of single walled carbon nanotubes as interconnects in network on chip fabrication methodologies for them have to be improved where the efficiency of metallic tubes are increased in a bundle. Optical interconnects on the other hand also seems to be an option for replacing copper in nanoscale regime. They show signs of higher bandwidth and low latency when compared to Copper interconnects, however they possess serious

integration issues. Hence fresh ideas are needed to take advantage of the novel properties of nanotubes and optical interconnects.

REFERENCES

[1]David Geer, “Chip Makers turn into Multicore Processors”, IEEE Computer Society, May 2005.

[2] “Digital Integrated Circuits”, J.M Rabaey

[3] James Balfour and William J.Dally, “Design Tradeoffs for Tiled CMP On-Chip Networks, “Proceedings of the 20th ACM International Conference on Supercomputing (ICS), June 2006.

[4] http://www.csa.com/discoveryguides/multicore/review.pdf

[5] “Principles and Practices of Interconnection Networks” by William James & Brian Towels.

[6]http://www.smdp.iitkgp.ernet.in/PDF%5CLowpowerPDF%5CIEP_DB.pdf

[7] Banit Agarwal, Navin Srivastava, Frederic T Chong, Kaustav, “ Nano-enhanced Architectures: Using Carbon Nanotube Interconnects in Cache Design.

[8]http://www.altera.com/literature/wp/wp-01161-optical-fpga.pdf.

[9] Jaldappagari Seetharamappa, Shivaraj Yellapa, "Carbon Nanotubes: Next Generation of Electronic Materials, Electro chemical Society Interface".

[10] http://www.unidym.com/files/whitepaper 1430.pdf.

[11] http://www.pa.msu.edu/cmp/csc/ntproperties/.

[12] Azad Naeemi, James Meindl, "Carbon Nanotube Interconnects", Invited Talk, ISPD 2007.

[13] P.L.McEuen, M.S.Fuhrer, and H Park, "Single Walled Carbon Nanotube Electronics,"IEEE Trans.Nanotechnology, Mar2002.

[14] P.J.Burke, Nanotechnology, IEEE Trans on vol. 1, no.3, pp129-144, 2002.

[15] Fred Chen, Ajay Joshi, Vladimir Stojanovic, Anantha Chandrakasan, "Scaling and Evaluation of Carbon Nanotube Interconnects for VLSI Applications".

[16] Azad Naeemi, J.D.Meindl, "Monolayer Nanotube Interconnects: Promising Candidates for Short Local Interconnects", IEEE 2005.

[17] Dipen Patel, Yong Bin Kim, "Carbon Nanotube Bundle Interconnect: Performance Evaluation, Optimum Repeater Size and Insertion for Global Wire".

[18] International Technology Roadmap for Semiconductors, http://www.itrs.net/.

[19] Azad Naeemi, J.D.Meindl, "Design and Performance Modeling of Single-Walled Carbon Nanotubes as Local, Semi global, and Global Interconnects in Gigascale Integrated Systems".

[20] Ashok Srivatsava, Yao Xu and Ashwani Sharma, "Carbon Nanotubes for Next-Generation Interconnects", SPIE 2010.

[21] Ashok Srivatsava, Yao Xu, and Ashwani K.Sharma, "Carbon Nanotubes for next generation very large scale integration interconnects", 2010 Society of Photo-Optical Instrumentation Engineers.

[22] Mayank Kumar Rai and Sankar Sarkar, "Carbon Nanotube as VLSI Interconnect".

[23] Debaprasad Das, Hafizur Rahaman, "Crosstalk Analysis in Carbon Nanotube Interconnects and Its impact on Gate Oxide Reliability", IEEE 2010.

[24] Payman Zarkesh-Ha, George B.P.Bezerra, Stephanie Forrest, Melanie Moses, “Hybrid Network on Chip (HNoC): Local Buses with a Global Mesh Architecture”, SLIP 2010.

[25] 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(3):580-589, 1998.

[26] Y.Hostoke, S.Vangal, A.Singh, N.Borkar, and S.Borkar. A 5-GHz mesh interconnect for a teraflops processor. IEEE MICR, 27(5):51-61, 2007.

[27] M.B.Taylor, J.Kim, J.Miller, D.Wentzlaff, F.Ghodrat, B.Greenwald, H.Hoffman, P.Johnson, J-W.Lee, and W.Lee. The Raw microprocessor: A computational fabric for software circuits and general purpose programs. IEEE MICRO, 22(PART 2):25-35, 2002.

[28] P. Christie and D. Stroobandt. The interpretation and application of rent’s rule. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 8(6):639–648,2000.

[29] D. Greenfield, A. Banerjee, J.-G. Lee, and S. Moore.Implications of Rent’s rule for NoC design and its

fault-tolerance. In Proceedings of the First International Symposium on Networks-on-Chip (NOCS'07), 2007.

[30] W. Heirman, J. Dambre, D. Stroobandt, and J. Campenhout. Rent's rule and parallel programs: Characterizing network traffic behavior. In Proceedings of the 2008 International Workshop on System Level Interconnect Prediction, SLIP'08, 2008.

[31] George B.P Bezerra, Stephanie Forrest, Melanie Moses, AL Davis, Payman Zarkesh-Ha, "Modeling NOC Traffic Locality and Energy Consumption with Rent's Communication Probability Distribution", SLIP 2010.

[32] Hang Sheng Wang, Xinping Zhu, Li-ShiUan Peh, Sharad Malik, "Orion: A Power Performance Simulator for Interconnection Network".

[33] Carbon nanotubes: Synthesis, structure, properties, and applications by M.S.Dresselhaus, G.Dresselhaus.

[34] Paul McEuen, Michael S.Fuhrer, and Hongkun Park, "Single Walled Carbon Nanotube Electronics", IEEE, 2002.

[35]www.research.cornell.edu/kic/events/journalists2007/pdf/carbon_nanotubes_graphene.pdf.

[36] www.iac.uta.edu/mntv/pdf/carbonnanotubefabrication.pdf

[37] http://www1.uprh.edu/nsfnue/Nanotubes/fabrication.pdf

[38] P.Muthana, M.Swaminathan, "Packaging of Multiprocessors: Tradeoffs and Potential Solutions", Electronic technology and components conference, 2005.

[39] H.B.Bakoglu, "Circuits, Interconnections, and Packaging for VLSI".