Chapter 10

SUMMARY AND CONCLUSION
Chapter 10

SUMMARY AND CONCLUSION

10.1 Introduction

Field programmable gate arrays require considerable hardware overhead to offer programmability thereby, making them less power efficient than custom ASICs for implementing same logic circuit. The large number of transistors in FPGA chips suggests that the power trends associated with scaling may impact FPGAs more severely than custom ASICs. Power management in FPGA will be mandatory at very deep submicron technology node to ensure correct functionality, reliability and to reduce packaging costs. Furthermore, low power consumption is needed if FPGAs are to be a viable alternative to ASICs in low-power portable applications.

The aim of this thesis is to investigate the low power architecture for the basic building blocks of FPGA and techniques that can be used for reducing the power consumption of these blocks.

This chapter is organized into four sections. The first section presents the summary of the work presented in each chapter. The second section provides the conclusions drawn from the results obtained in each chapter. The third section summarizes the achievement and the last section outlines area for future research.

10.2 Summary

With the scaling of technology, digital systems have grown immensely complex. However, with increasing circuit complexity, the cost and design cycles of custom VLSI designs have increased significantly. FPGAs offer an efficient and cost effective option for implementing digital systems for medium to low volume production. Digital system designers therefore now get the advantages of low time-to-market of the programmable logic in addition to almost ASIC-like logic density. Commercial FPGAs have on-chip memory blocks and DSP resources, apart from the programmable logic making them more attractive for implementing complete systems on chip.

In very deep submicron technology, leakage power has emerged as a key design challenge because leakage power increases with small geometries. For the FPGAs to continue to retain its semiconductor market and competitive advantages over the high
performance custom VLSI designs, the FPGA industry must adopt new techniques for leakage power reduction.

This thesis has proposed the operation of FPGA blocks in subthreshold regime. Such kind of FPGAs can be used even in ultra low power portable applications like hearing aids, pacemakers etc.

This thesis also proposed low power high speed carbon nanotube field effect transistor based SRAM cell. Due to high-K dielectric and high intrinsic carrier mobility in carbon nanotubes, the CNFET has lower leakage and higher speed compared with CMOS transistor. Since FPGA consists of a large number of SRAM cells therefore, total leakage power consumption will be considerably reduced and there will be a significant improvement in speed if all CMOS based SRAM cells are replaced by CNFET cells.

Power dissipation in CMOS circuits can be classified as either dynamic or static. Dynamic power consumption is due to the logic transitions at circuit nodes whereas static power is dissipated even when a circuit is in the idle state. Historically, dynamic power has dominated power consumption in CMOS circuits; however, technology scaling trends have resulted in leakage power becoming an increasing component of total power. Chapter 2 provides a general and block specific low power techniques for reducing leakage and switching power in CMOS circuits.

The recent architecture and breakdown of power consumption in FPGA is described in detail in chapter 3, and it has been observed that interconnect accounts for the bulk of FPGA's static and dynamic power consumption. Various approaches for reducing FPGA power have been proposed in the literature including approaches for leakage reduction. A significant improvement in power efficiency has to be achieved to make FPGAs viable in portable domain applications.

Chapter 4 proposed the subthreshold regime operation of FPGA building blocks. The subthreshold logic can be easily implemented and derived from traditional existing circuits by lowering the supply voltage to less than the threshold voltage. The building blocks of reconfigurable hardware such as 4-input LUT and one bit full adder cell is implemented and their performance such as delay, power and PDP are estimated in the subthreshold region. The sensitivities of the two schemes against process parameter variations are also explored. It has been found that DTCMOS shows superior robustness against temperature and process variations. Further, an 8T subthreshold SRAM cell is implemented in MOS and DTCMOS schemes and it has been found that
DTCMOS cell outperforms in read SNM and write delay (speed) with minimal penalty in leakage consumption for the selected supply voltage.

Technology scaling of the bulk silicon transistor over the last three decades has not only produced ultra high performance digital circuits but has also sustained Moore's Law. However, ramifications of "short channel effects" such as exponential increase in leakage current and large parameter variations have created challenges in design and testing of bulk integrated circuits. Due to very high intrinsic carrier mobility of carbon nanotubes (CNTs), the carbon nanotube field effect transistors (CNFETs) have caught the attention of device/circuit and system engineers worldwide. Chapter 5 reviews the literature on the architecture and modeling of CNFET. This chapter also compares the performance of bulk and CNFET transistor based 5-stage ring oscillator and FO4 benchmark circuits. This chapter also explores the performance of a CNFET based FPGA 'Basic Logic Element (BLE)’ block and it is found that the CNFET based BLE outperforms in switching power and speed compared to the same implemented in bulk CMOS technology.

The ability of FPGA to implement a variety of circuits on a single chip always results in the under utilization of some logic and interconnect resources. These unused transistors leak power in the absence of switching activity. The interconnect fabric of an FPGA consumes a major portion of the total leakage power. Further as we move to smaller nodes, leakage will ultimately dominate the total power consumption. Chapter 6 focuses on the reduction of leakage power in the interconnect switch matrix multiplexers by ensuring that minimum leakage vector should be applied to all these multiplexers. The analyses of multiplexers have been carried out with varying sizes, topologies and transistor sizing at different temperatures and supply voltages at a deep submicron 22nm technology node. The minimum leakage state heavily depends on the relative magnitude of the subthreshold and the gate leakage currents. Therefore, different low leakage vectors are selected for a minimum and optimum sized multiplexers such as keeping all input lines of multiplexers at logic ‘1’ and inputs to inverters at logic ‘0’ or keeping all the inputs of the multiplexers and inverters at logic ‘1’ will provide a significant reduction in leakage for all unused interconnects without any kind of penalty.

Due to aggressive scaling, secondary effects and process variations, the power consumption and performance of CMOS SRAM cell worsens in deep submicron technology. It has therefore, become difficult to design low power, high speed, robust,
and compact SRAM cells in deep submicron technology. The carbon nanotube based field effect transistor (CNFET) technology with reduced process variation, better gate controllability, high thermal stability and high drive current is a promising alternative to the bulk CMOS. Chapter 7 explores a low leakage, high speed and robust CNFET based 6T-SRAM cell and compares its performance with that of conventional CMOS based cell. All the simulations are carried out at 32nm technology node with equal threshold voltage for CNFET and CMOS transistors. Due to inherent characteristics of CNFET such as good gate controllability, drive current and immunity to short channel effect, the CNFET cell outperforms to CMOS cell in terms of leakage power saving, write margin, speed and read SNM. As the FPGA consists of a large number of configurable SRAM cells, the implementation of low leakage SRAM cell with CNFET technology will greatly contribute to the reduction of overall leakage consumption of FPGAs.

The International Technology Roadmap for Semiconductors (ITRS) predicts that the traditional copper interconnects will be a major bottleneck when feature sizes become smaller than 45nm. This is due to steep rise in parasitic resistance of copper, which not only increases the interconnect delay but also limits their current carrying capability. In order to alleviate such problems, alternative interconnect technologies and their architectural implications for FPGAs in future process technologies must be explored.

CNTs have recently been proposed as a possible replacement for metal interconnects in future technologies. Due to their long mean free paths (MFP), high current carrying capability and high thermal conductivity, CNTs are expected to be a very good alternative material for future FPGA interconnects. Chapter 8 describes the different categories of CNTs such as Single-Wall Carbon Nanotubes (SWCNTs), Multi-Wall Carbon Nanotubes (MWCNTs) and Mixed CNTs. Because of their extremely desirable properties such as high mechanical and thermal stability, high thermal conductivity and large current carrying capacity, CNT bundle based interconnects promises to be good alternative as future FPGA interconnect. However the high resistance (of the order of 6.45 KΩ) associated with an isolated CNT necessitates the use of a bundle of CNTs. Moreover, due to the lack of control on chirality, any bundle of CNTs consists of metallic as well as semiconducting nanotubes. Almost all experimental results have demonstrated that a realistic nanotube bundle contains a mixed bundle of single-walled and multi-walled CNTs (SWCNTs and
MWCNTs. Therefore, more emphasis is given to mixed CNTs and it has been found that FPGAs which are implemented by mixed CNT interconnects outperforms in delay and energy consumption compared to that of traditional Cu interconnects. The chapter 8 also provides important guidelines for selection of vital parameters of mixed CNT bundles so as to optimize the resistance, capacitance and inductance of mixed CNT bundle interconnects. Moreover, exhaustive simulations have been carried out for different interconnect lengths with both CMOS and CNFET drivers for copper and CNT interconnects. It has been found that most FPGA interconnect resources implemented with CNFET drivers and CNT interconnects provide best performance than traditional CMOS and copper interconnects.

The basic switching element in most of the FPGAs interconnect are NMOS pass transistors or multiplexers which suffer from threshold voltage drop that causes high DC power dissipation in level restoring buffers. To eliminate this problem, recent architecture from Xilinx and Altera uses tri-state buffers. However this approach has significant area and power consumption overhead. Chapter 9 suggests some methods other than replacing pass transistors by tri-state buffers for reducing power consumption. The proposed method uses a novel configuration of dynamic threshold MOS (DTMOS) transistor. DTMOS based switches overcome the above disadvantage at a minimal increase in area. A new augmented DTMOS biasing scheme is proposed which provides a fixed body bias and improves the level of high input of level restoring buffer which reduces the DC power consumption. The main advantage of DTMOS over conventional MOS is its higher drive current at lower bias levels. A simulation is performed on the realistic multiplexer based interconnect resources of FPGA such as Double, Hex and Long at a transistor level driving a Copper wire of length varying from 20um to 100um. It has been found that the DTMOS based interconnect resources outperform in speed and PDP compared to conventional MOSFET.

A final complication with DTMOS based interconnect resources is the process complexity and area penalty. The area overhead of the proposed interconnect will be very less because the extra needed transistor for DTMOS based switches is easily shared among all multiplexer based interconnects and the augmented transistor is of minimum size. The above disadvantage can be compensated for higher driving capability, higher operating frequency and low energy consumption of DTMOS circuit over the conventional MOS.
10.3 Conclusions

This thesis investigates low power architecture for the basic building blocks of FPGA and techniques for reducing both switching and leakage power consumption for these blocks. The breakdown of power consumption in FPGA is well-studied in Chapter 3 and it can be concluded from the said chapter that the interconnect accounts for major portion of static and dynamic power consumption.

Chapter 4 implements the building blocks of reconfigurable hardware such as 4-input LUT and one bit full adder cell and compared their performance such as delay, power and PDP in the subthreshold region for MOS and DTCMOS schemes. It can be concluded from Chapter 4 that DTCMOS has superior robustness against temperature and process variations besides having lower delay than that of CMOS for above blocks at 22nm node. Chapter 4 also implements 8 transistors SRAM cell in CMOS and DTCMOS schemes and it has been found that DTCMOS based 8 transistors SRAM cell provides up to 15% and 23% improvement in read SNM and speed at a supply voltage of 200mV. Whereas in standby mode due to different voltage at the cell nodes, the body biasing results in more leakage. Hence, DTCMOS based cell consumes 8% more leakage than that of CMOS cell.

Chapter 5 reviews the literature on the architecture and modeling of CNFET. This chapter also proposed a CNFET based FPGA building block such as BLE. Due to very high intrinsic carrier mobility of carbon nanotubes, it is found that the CNFET based BLE outperforms in speed compared with bulk CMOS technology. Similarly due to very small capacitance of CNT, the CNFET based BLE consumes very small switching power compared to bulk CMOS based LUT.

Chapter 6 proposed input vector control technique of leakage power reduction in the interconnect switch matrix multiplexers. It has been observed that most of the routing multiplexers remain unused in a typical FPGA implemented design. These multiplexers contribute significantly to the leakage power consumption. There is tremendous scope of reducing leakage power in these unused multiplexers provided these multiplexers are fed with the least leakage input vector at all times. Chapter 6 explores minimum leakage vectors for different multiplexer types and sizes at different temperature and supply voltages at a deep submicron 22nm technology node. It can be concluded from Chapter 6 that the proposed selection of input vectors for Hybrid 16:1 multiplexer for optimum size transistors reduces leakage by up to 23% and 16% at 85°C for supply voltages of 0.8V and 0.6V respectively.
Chapter 7 proposed a CNFET based 6 transistor SRAM cell. Due to inherent characteristics of CNFET such as good gate controllability, drive current, immunity to short channel effect and due to very high intrinsic carrier mobility, the CNFET cell outperforms in leakage power consumption, write margin, speed and read SNM as compared to a CMOS cell. It can be concluded from Chapter 7 that CNFET cell is 1.84X faster in speed and provides 21% improvement in read SNM. Furthermore due to absence of dangling bonds in CNFET and because of high K (HfO$_2$) electrolyte gating, the leakage power consumption of CNFET cell is 84% and 40% less than that of CMOS cell at operating temperatures of 27°C and 80°C respectively.

The steep rise in parasitic resistance of copper in deep submicron not only increases the interconnect delay but also limits its current carrying capability. Due to their long mean free paths (MFP), high current carrying capability and high thermal conductivity, CNTs are expected to be a very good alternative material for future FPGA interconnects. Chapter 8 proposed a mixed CNT bundle as interconnects for FPGAs. Double, Hex and Long interconnect resources of FPGA are implemented by CNT and Cu wires with CNFET and CMOS drivers respectively. Due to lower values of extracted R and C components of CNT, the CNT bundle interconnects based FPGA resources outperform the traditional interconnects in terms of speed, power, energy and energy-delay-product (EDP). It can be concluded from Chapter 8, that the Hex interconnect resource implemented by CNFET-CNT has 41% less EDP than the traditional Hex. Similarly larger length of Long interconnects provides more advantage of speed, energy and EDP for CNT interconnects. Therefore, the Long interconnect resource implemented by CNFET-CNT has 44% less EDP than the traditional Long interconnect resource.

Most of FPGAs interconnect has NMOS pass transistor based multiplexer as a switching element which suffer from a weak-'1' and causes high DC power dissipation in level restoring buffers. Chapter 9 proposed use of novel configurations of dynamic threshold MOS (DTMOS) which overcome this disadvantage at a minimal increase in area. By using this novel DTMOS, a realistic multiplexer based interconnect resources of FPGA such as Double, Hex and Long are implemented. These interconnect resources drives a copper wire of length 20um to 100um. Due to high driving current of DTMOS, it has been found that the DTMOS based interconnect resources outperform in speed and PDP compared with the conventional MOS interconnects. It is concluded
from Chapter 9 that at VDD=0.6V, compared to conventional MOS (Conv-MOS), the proposed SVT scheme such as SVT-Double, SVT-Hex and SVT-Long provide an 18%, 17% and 10% improvement in PDP respectively. The area overhead of the proposed DTMOS based interconnect will be minimum because the extra needed augmented transistor is shared among many multiplexer trees and is of minimum size.

10.4 Achievements

- Subthreshold DTCMOS based schemes for implementing FPGA building blocks such as 4-input LUT, 1-Bit full adder cell and 8 transistor SRAM cell are proposed which have better power performance and higher speed than blocks implemented with conventional MOS.
- A CNFET based FPGA building block such as BLE is proposed. Due to very high carrier mobility and low capacitance of carbon nanotubes, the CNFET based BLE outperforms in speed and power consumption compared to BLE implemented in bulk CMOS technology.
- An input vector control scheme for reducing the leakage of unused FPGA multiplexer based interconnect is proposed which reduces significantly leakage power of interconnects without any kind of penalty.
- A CNFET based SRAM cell is proposed due to inherent characteristics of CNFET such as good gate controllability, drive current and immunity to short channel effect, the CNFET cell outperforms the CMOS cell in terms of leakage power saving, write margin, speed and read SNM.
- A mixed CNT bundle based interconnects for FPGAs is proposed and the performance is compared with conventional Cu interconnects. Due to lower values of extracted R and C components of mixed CNT, the mixed CNT bundle interconnects based FPGA resources outperform in delay, power, energy and energy-delay-product (EDP). A combination of CNFET driver and CNT interconnect gives best performance in terms of key parameters.
- A novel configuration of dynamic threshold MOS (DTMOS) scheme for FPGA multiplexer based interconnect is proposed. In DTMOS, the gate potential changes the threshold voltage and produces a strong 1 which reduces the DC power dissipation in level restoring buffer. By using this novel DTMOS, a realistic multiplexer based interconnect resources of FPGA such as Double, Hex and Long are implemented. Due to high driving current of DTMOS, the
Areas for Future Research

10.5 Areas for Future Research

- Subthreshold FPGA
  A subthreshold FPGA design faces a combination of subthreshold circuit challenges and problem inherent in FPGA architectures. Three major challenges stand out for the subthreshold FPGA design. These are process variation, long length of interconnect and memory. Variation threatens to disrupt any subthreshold design. FPGAs typically dissipate 60%-70% of their power in the interconnect network (e.g., wires, buffers, connection boxes and routing switches), 10%-20% in the clock network and 5%-20% in logic. This breakdown indicates that a focus on clocking and interconnect are necessary. These are the two areas on which very little previous subthreshold FPGAs work has been focused. The clock network for an FPGA extends across the entire fabric and drives all of the registers in the design. This large distributed network can consume a significant amount of power. Furthermore, driving the large capacitive load of the clock network with the buffers operating in the subthreshold region presents a significant problem. Variation in the buffer can lead to substantial differences in the drive strength of buffers leading potentially large clock skew across the FPGA fabric. Hence some efficient method to reduce skew is important for viable subthreshold FPGA design. One extension of this work may be the implementation of interconnects by Mixed CNT bundle instead of Copper and all the logics, switch boxes and connection boxes by CNFET instead of CMOS.

- Nanowire Based FPGAs
  The [ITRS-07] has stated that it will be a difficult challenge to progress CMOS technology beyond the 22 nm technology generation. This challenge has stimulated the next-generation devices most likely based on non-planar structures such as double-gate FETs and fin-FETs. However, these technologies rely on an enhancement of individual device performance (such as increased mobility, lower leakage current or higher drive current) and do not solve the issues of the size and density limitations. To directly address these challenges, novel nanoscale transistor channel materials, such as semiconducting nanowires (NWs) has found very good scope to replace the
conventional CMOS. These materials are attractive because they have very narrow
diameters and have no density limitations since they are not fabricated using
conventional lithography techniques.

Due to their crystalline structure, smooth surfaces and the ability to produce
radial and axial nanowire they can reduce scattering, which results in higher carrier
mobility. Semiconducting nanowires of a variety of materials can be grown with
controlled diameters down to 3nm. By material selection or doping, one can engineer
the electrical properties of the nanowires (e.g., P-type, N-type). With nano-imprint, Dip
Pen Nano-lithography and self-assembly technologies it is possible to get sets of
parallel wires which can be used in FPGA routing. These wires could be made using a
single crystal of metal-silicide (NiSi nano-wires). At the nanoscale, one can use single-
molecule switches that exhibit reversible switching behavior. These molecules self-assemble at the cross-points of nanowires, and can be switched between ON and OFF states by the application of a voltage bias.

The future scope of this work is to implement the FPGA routing interconnects with
nanowires and the C-Box, S-Box and logic blocks by single-molecule switches. It is
expected that such nanoscale FPGAs architecture will provide the best performance
with the least area. Building successful nanoscale devices require synergy between the
Design and Process Engineers and the Chemists.

- Sublithography FPGAs

One of the most important challenges for scaling feature sizes is the cost of the
fabrication process. Sublithographic techniques may offer an economical alternative to
costly lithographic feature size scaling. These new sublithographic technologies with
10nm full pitch semiconductor and metal nanowires may enable tera-scale system
integration. In addition, nanowires also provide very high interconnect density. Sub-
lithographic electronic devices may also work as reconfigurable molecules (switching
elements).

Switchable molecules can be assembled and placed one or a few molecules
under each junction in a crossed array to provide programmable junctions. With these
building blocks, one can build a diode junction by crossing P-doped and N-doped
nanowire. Then field-effects are used to control conduction in semiconducting
nanowires to implement switchable crosspoints or memory bits. Using the axial
variation of doping or material combination, a single nanowire can have regions which are gateable and other regions which are not gateable. These devices are sufficient to build programmable memory points and field-effect based inverting and restoring logic gates which can be incorporated in the future nanoscale FPGAs.

The CNFET Technology is assumed to be perfect in this work for simulation purposes. The work can be refined further by assuming limitations of existing CNT technology like imperfect Chirality control, misaligned CNTs and presence of metallic CNTs etc.