N^3ASICS: DESIGNING NANOFABRICS WITH FINE-GRAINED CMOS INTEGRATION

A Thesis Presented

by

PAVAN PANCHAPAKESHAN

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN ELECTRICAL AND COMPUTER ENGINEERING

February 2012

Electrical and Computer Engineering

© Copyright by Pavan Panchapakeshan 2012

All Rights Reserved

N^3ASICS: DESIGNING NANOFABRICS WITH FINE-GRAINED CMOS INTEGRATION

A Thesis Presented

by

PAVAN PANCHAPAKESHAN

Approved as to style and content by:

Csaba Andras Moritz, Chair

Israel Koren, Member

C. Mani Krishna, Member

C.V. Hollot, Department Chair
Electrical and Computer Engineering

ACKNOWLEDGMENTS

I take this opportunity to thank all the people who have helped me with my thesis. First and foremost, I thank my advisor, Prof. Csaba Andras Moritz for his supervision, inspiration, encouragement and support. I would also like to extend my gratitude to the committee members, Prof. Israel Koren and Prof. Mani Krishna for their time, advice and suggestions throughout this project. I am indebted to my lab mates for providing a healthy environment to learn. I am especially grateful to my lab mate, Pritish Narayanan for mentoring and guiding me throughout the course of my thesis. I would like to thank all my friends for making my stay at Amherst a memorable one. Last but not the least, I would like to thank God and my family members, for their constant support, motivation and for helping me through difficult times.

ABSTRACT

N^3ASICS: DESIGNING NANOFABRICS WITH FINE-GRAINED CMOS INTEGRATION

FEBRUARY 2012

PAVAN PANCHAPAKESHAN

B.E, VISHVESHWARIAH TECHNOLOGICAL UNIVERSITY

M.S.E.C.E., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Csaba Andras Moritz

Nanoscale-computing fabrics based on novel materials such as semiconductor nanowires, carbon nanotubes, graphene, etc. have been proposed in recent years. These fabrics employ unconventional manufacturing techniques like Nano-imprint lithography or Super-lattice Nanowire Pattern Transfer to produce ultra-dense nano-structures. However, one key challenge that has received limited attention is the interfacing of unconventional/self-assembly based approaches with conventional CMOS manufacturing to build integrated systems.

We propose a novel nanofabric approach that mixes unconventional nanomanufacturing with CMOS manufacturing flow and design rules to build a reliable nanowire-CMOS 3-D integrated fabric called N^3ASICs with no new manufacturing constraints. In N^3ASICs active devices are formed on a dense semiconductor nanowire array and standard area distributed pins/vias, metal interconnects route signals in 3D.

The proposed N^3ASICs fabric is fully described and thoroughly evaluated at all design levels. Novel nanowire based devices are envisioned and characterized based on

3D physics modeling. Overall N^3ASICs fabric design, associated circuits, interconnection approach, and a layer-by-layer assembly sequence for the fabric are introduced. System level metrics such as power, performance, and density for a nanoprocessor design built using N^3ASICs were evaluated and compared against a functionally equivalent CMOS design. We show that the N^3ASICs version of the processor is 3X denser and 5X more power efficient for a comparable performance than the 16-nm scaled CMOS version without any new/unknown-manufacturing requirement.

Systematic yield implications due to mask overlay misalignment have been evaluated. A partitioning approach to build complex circuits has been studied.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENTS iv

ABSTRACT v

LIST OF TABLES ix

LIST OF FIGURES x

CHAPTER

1. INTRODUCTION AND MOTIVATION 1

2. PHYSICAL FABRIC VISION 4

   2.1 Approaches to Build a Nano-CMOS Hybrid Fabric 4
   2.2 Physical Fabric Vision 5
   2.3 Chapter Summary 7

3. N^3ASICS FABRIC 8

   3.1 Introduction 8
   3.2 CMOS Design Rules Applied to the Fabric 10
   3.3 Assembly Sequence 11
   3.4 Chapter Summary 15

4. N^3ASICS DEVICES AND CIRCUITS 16

   4.1 Introduction 16
   4.2 Device Structure 16
   4.3 Device Simulations 19
   4.4 Behavioral Model Creation for Circuit Simulation in HSPICE 23
   4.5 Circuit Style and Evaluations 23
   4.6 Chapter Summary 27

5. SYSTEM LEVEL EVALUATION **29**

5.1 WISP-0 29
5.2 CMOS Baseline WISP-0 29
5.3 N^3ASICs WISP-0 31
5.4 N^3ASICs-16 and 16nm CMOS Comparison 33
5.5 Chapter Summary 36

6. LOGIC PARTITIONING STUDY **37**

6.1 Introduction 37
6.2 Case study 38

6.2.1 Two bit adder 38

6.2.1.1 Two-level implementation of a 2bit full adder 38
6.2.1.2 With partitioning 39
6.2.1.3 Comparison of the two approaches 39

6.2.2 (7,3)-Counter 40

6.3 Study of Clocking schemes for partitioning approach 41
6.4 Chapter Summary 45

7. IMPACT OF MASK OVERLAY **46**

7.1 Introduction 46
7.2 Alignment and Mask overlay 46
7.3 Mask Overlay simulation 49
7.4 Chapter Summary 50

8. CONCLUSION **52**

BIBLIOGRAPHY **53**

LIST OF TABLES

Table **Page**

4.1 Devices Simulation Parameters 20

4.2 Devices Simulation output 21

5.1 Key system level metrics for WISP-0 34

6.1 Comparison of different metrics of the two approaches for a 2bit adder 39

6.2 Comparison of different metrics for the two approaches for a (7,3) counter 40

6.3 Comparison of 4-phase and 6-phase clocking schemes for (7,3) counter 43

6.4 Comparison of 4-phase and 6-phase clocking schemes for 3bit addder 45

LIST OF FIGURES

Figure **Page**

2.1 Nanowires and alignment markers in the same mold for NIL technique 6

3.1 Nano-CMOS integrated N^3ASICs fabric 8

3.2 N^3ASICs input-output organization 9

3.3 CMOS Design rules applied to N^3ASICs 10

3.4 Patterned Nanowires 12

3.5 Creation of Lithographic contacts and dynamic control rails 12

3.6 Metal gate deposition step 13

3.7 Metal 1 vias and interconnects 14

3.8 Metal 2 interconnects to route across logic planes 14

4.1 Integrated Device-fabric exploration methodology 17

4.2 3D structure of N^3ASICs device (2C-xnwFET) 18

4.3 I_{DS} vs V_{DS} with varying V_{GS} for 2C-xnwFET 21

4.4 I_{DS} vs V_{GS} with varying V_{DS} for 2C-xnwFET 22

4.5 Gate capacitance vs V_{GS} for 2C-xnwFET 22

4.6 Schematic diagram of a sample circuit to illustrate how 2 stages of N^3ASICs are connected 24

4.7 Four phase clocking scheme 25

4.8 N^3ASICs 1 bit full adder top view 26

4.9 Cross sectional view of a cross point in N^{3}ASICs 26

4.10 Simulation waveforms of N^{3}ASICs One bit full adder 27

5.1 WISP-0 Nanoprocessor layout 30

5.2 Methodology for performance characterization of 16nm static CMOS baseline 32

5.3 A N^{3}ASICs tile. Area calculation example 32

5.4 Density Comparison of N^{3}ASICs with CMOS at different technology nodes 34

5.5 Transistor width distribution in 16nm CMOS-WISP-0 35

6.1 A two-level two-bit adder (Top view) 38

6.2 Partitioned N^{3}ASICs two-bit adder 39

6.3 Partitioned (7,3) counter 40

6.4 Sample functional unit 42

6.5 Functional unit with 4-phase clocking scheme 42

6.6 Functional unit with 6-phase clocking scheme 43

6.7 6-phase clocking scheme 44

6.8 (7,3) counter with 4-phase clocking showing the identity tile 44

7.1 Patterned nanowires (larger than logic nanowires) could be used as Moire patterns for alignment 47

7.2 Depiction of mask registration and alignment markers during contact creation step 48

7.3 Mask registration during functionalization step 49

7.4 Mask overlay limited Yield vs. Overlay for 3D integrated fabric 50

CHAPTER 1

INTRODUCTION AND MOTIVATION

As dimensional scaling of CMOS is approaching fundamental limits, several new materials, devices and information processing paradigms are being explored to sustain the historical trend of integrated circuit scaling and reduction of cost per function. For example, spin waves [28], QCAs [6], carbon nanotubes [14], semiconductor nanowires [13] [11] etc are under investigation as potential replacements for CMOS. However, reliable manufacturing of integrated nanosystems incorporating these novel nanodevices continues to be challenging. Specifically, assembly of nanostructures, achieving reconfigurable devices, interfacing and overlay considerations are key issues for nanoscale computing fabrics. While nanofabrics such as NASICs [16] [21] [37] [17] [38] [19], CMOL [12] and FPNI [29] have been proposed minimizing certain manufacturing constraints, some or all of the aforementioned concerns still exist.

Unconventional/self-assembly based manufacturing techniques like Nano Imprint Lithography (NIL) [18] and Superlattice Nanowire Pattern transfer (SNAP) [34] [33], are able to produce ultra-high density nanostructures. For e.g., it has been shown that 7nm width with 13nm pitch nanowires can be patterned with SNAP [10]. However these and other unconventional techniques have very poor overlay with respect to previously formed patterns. Overlay imprecision for NIL is as high as $3\sigma = \pm105$nm [25]. Further, interfacing and integration with external CMOS (e.g. for control, input/output functions) becomes challenging when unconventional techniques are employed.

On the other hand photolithography has an excellent overlay and alignment precision. According to International Technology Roadmap for Semiconductors (ITRS) [1] 16nm CMOS is projected to have an overlay imprecision of $3\sigma = \pm 3.3$nm. Also, CMOS manufacturing flow has very low defect rates compared to the self-assembly based approaches. However, conventional manufacturing flow has reduced density benefits when compared to the unconventional approaches.

Our goal in this thesis is to develop an approach by which we can combine unconventional and conventional manufacturing approaches while retaining the benefits of both. Unconventional nanomanufacturing is used in conjunction with conventional CMOS lithography and design rules to build a new class of 3-D integrated nanofabrics without any additional manufacturing constraints. A new nanofabric, called N^3ASICs (Nanoscale 3-D Application Specific Integrated Circuits [24]) is presented. This fabric can achieve the high densities obtained from unconventional manufacturing along with the reliability and overlay precision of conventional photolithography.

One possible variant of N^3ASICs is discussed in this thesis. The key idea is the use of standard pin-based 3D integration following design rules. Other versions might be envisioned by relaxing some of the design rules and/or reducing the pins/vias to achieve greater density benefits, or using programmable devices. The main contributions of this thesis are

- We present N^3ASICs, a new hybrid nano/CMOS computational fabric with no special manufacturing constraints.

- We show a layer-by-layer assembly sequence for N^3ASICs depicting how the complete fabric (including devices, interconnect and interfacing) may be realized on a single Silicon-on-Insulator (SOI) wafer.

- We show how fine-grained integration between nanoscale and CMOS features can be achieved using standard area distributed pins/vias and design rules.

- Novel dual-channel crossed nanowire field effect transistors (2C-xnwFETs) are proposed. Extensive characterization of these devices is done using Synopsys Sentaurus.

- We validate the fabric using an integrated device-circuit methodology. Behavioral models are developed and verified using detailed HSPICE circuit level simulation.

- We evaluate key system-level metrics such as density, performance and power for N^3ASICs and compare it against an equivalent 16nm CMOS design.

The rest of the thesis is organized as follows: Chapter 2 presents the physical fabric vision. Chapter 3 describes the N^3ASICs fabric in detail. Chapter 4 describes N^3ASICs devices, behavioral models and circuits. System level evaluations such as area, power and performance comparison are presented in Chapter 5. Chapter 6 describes how some of the limitations of the two-level logic approach can be overcome and presents an approach to build complex logic functions. Systematic yield implications due to mask-overlay misalignment are discussed in Chapter 7. Chapter 8 concludes the thesis.

CHAPTER 2

PHYSICAL FABRIC VISION

In this chapter, we discuss how unconventional/self-assembly and conventional manufacturing techniques can be combined to build a 3-D integrated fabric, with careful consideration of manufacturing and overlay requirements. Different integration approaches are discussed and challenges are outlined. Based on this understanding, a physical fabric vision for a hybrid nano-CMOS fabric is presented.

2.1 Approaches to Build a Nano-CMOS Hybrid Fabric

One approach to build a fully integrated 3-D fabric is to use only optical lithography for all the process steps. The extremely good overlay precision of CMOS is the key advantage of this approach. Therefore, yield obtained will be comparable to CMOS process yield. However, the approach is expected to have low density when compared to techniques that use self-assembly/unconventional nanofabrication techniques since it is limited by optical lithography.

A second approach would be to use unconventional approaches on top of a conventional manufacturing flow to obtain a 3D integrated fabric of high density. Such an approach has been examined in CMOL [12] and FPNI [29] nanofabrics, where unconventional techniques such as nanoimprint are necessary after the fabrication of CMOS layers. Overlay alignment precision needed for imprint lithography is $3\sigma=\pm105$nm [25], which implies significant challenges in alignment against previously defined lithographic features. Such a large overlay misalignment can contribute to significant yield

loss (or conversely trading-off much of the density benefit using well separated features for acceptable yield) and is not ideal.

In our current work we propose a nano-CMOS integration approach which considers the order of manufacturing process steps along with fabric design choices which aids in mitigating mask overlay while still achieving an ultra dense fabric. Given that unconventional techniques have very high overlay imprecision, a simple and intuitive way of overcoming this limitation is to make use of NIL/SNAP as the first step in the manufacturing process. This overcomes the overlay limitation of nano-manufacturing, since first step of the manufacturing sequence will not have any overlay requirement. All subsequent steps use conventional lithography and have excellent overlay alignment.

2.2 Physical Fabric Vision

Based on the latter approach, we propose a new physical fabric that consists of nanowire arrays at the bottom (built using unconventional manufacturing) with a conventional CMOS metal stack for interconnect (built using photolithography) on top. All active devices and logic implementation is achieved on the ultra-dense nanowire arrays which can be direct-patterned on an ultra-thin Silicon-On-Insulator (SOI) wafer. The patterning can be achieved using techniques like NIL or SNAP.

In this approach, patterning of high-density nanostructures is carried out *prior* to all lithography steps. Furthermore if the defined nanostructure pattern is regular (e.g. parallel arrays), the first lithographic mask has overlay tolerance, i.e. it may be offset over the array without yield loss. Subsequent steps make use of conventional photolithography. The *a priori* assembly/direct-patterning of sub-lithographic features on the densest NW layer before any conventional lithographic step (e.g., for contacts/vias) means 3D overlay alignment requirements exist only between subsequent lithographic masks, projected to be $3\sigma = \pm 3.3$nm for 16nm CMOS [1]. This

approach achieves 3-D integration without any special manufacturing requirements while ensuring finer nanoscale resolution (and consequently higher density) than can be achieved with lithography at the bottom.

To enable full and fine-grained integration with CMOS without new manufacturing requirements, lithographic design rules need to be followed. Standard lithography design rules are used for lithographic functionalization steps including defining positions of transistors, power and control rails, vias, interconnect etc. Lithographically defined vias or area-distributed interfaces connect the nanowire arrays through a CMOS metal stack. Metal interconnects are used for routing the signals in 3D. These are described in the subsequent chapters.

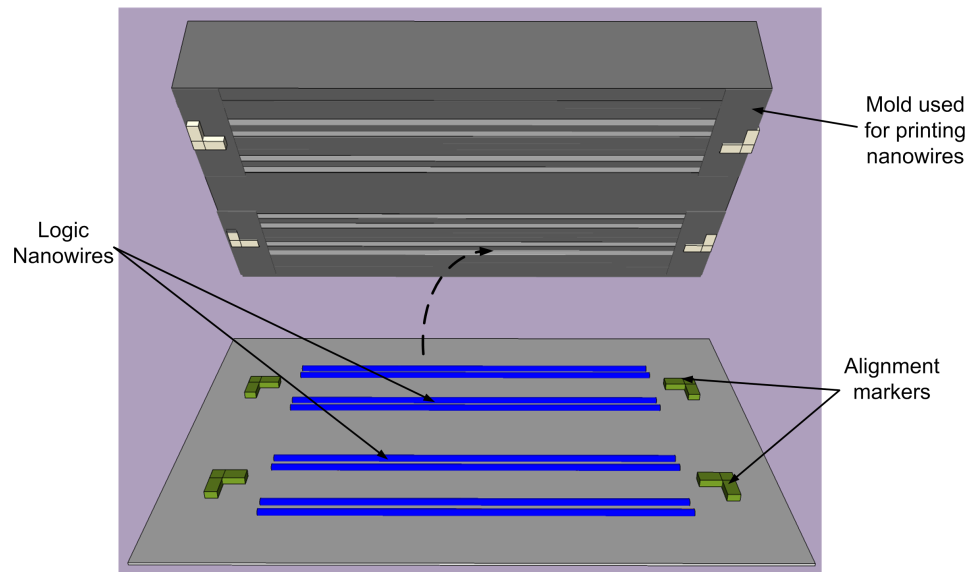

Figure 2.1. Nanowires and alignment markers in the same mold for NIL technique

In order to aid registration of photo lithographic steps, additional alignment markers can be created at the same time as the logic nanowires. If NIL is used, alignment markers for subsequent lithography steps and logic nanowires can be part of the same mold and hence transferred to the substrate in a self-aligned fashion as shown in Fig. 2.1. In the case of SNAP, where an arbitrary alignment marker may be diffi-

cult to achieve, patterned nanowires of different dimensions can be used as Moire patterns/fringes [39].

2.3 Chapter Summary

Different approaches to build a nano-CMOS hybrid fabric were presented. A nano-CMOS integration approach with careful consideration to the order of manufacturing process steps was developed. This manufacturing approach does not introduce any new manufacturing constraints. A single unconventional step is carried out and all the subsequent steps make use of conventional lithography. Further the use of conventional lithography is possible because all the layers adhere to the CMOS design rules.

The next chapter discusses N^3ASICs, a fabric incorporating these principles of 3-D integration, and shows how CMOS design rules can be applied to this Nano-CMOS hybrid fabric. Detailed assembly sequence is presented.

CHAPTER 3

N^{3}ASICS FABRIC

3.1 Introduction

In this chapter we present the 3-D integrated N^{3}ASICs fabric built using the physical fabric vision presented in the previous chapter. The fabric can be built on a single ultra-thin SOI wafer, with a direct-patterned nanowire logic plane surrounded by support CMOS circuitry (e.g. for external control). Fine-grained lithographically defined vias or area-distributed interfaces connect the nanowire arrays through a CMOS metal stack. Detailed N^{3}ASICs description and evaluations are presented in the following chapters.

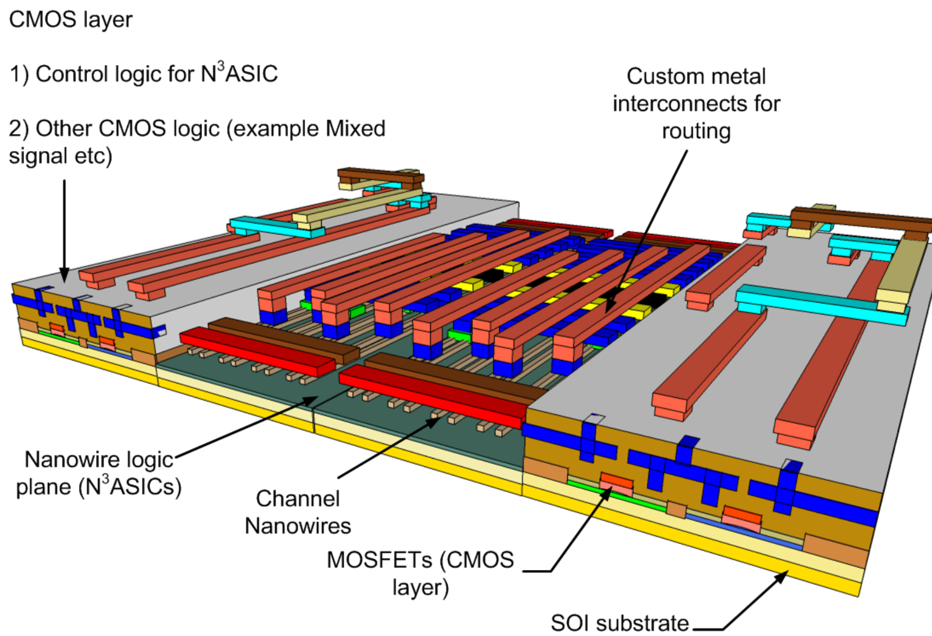

Figure 3.1. Nano-CMOS integrated N^{3}ASICs fabric

Fig. 3.1 shows the envisioned N^{3}ASICs fabric built on a standard Silicon-on-Insulator (SOI) wafer. It consists of uniform parallel semiconductor nanowire arrays on which logic/memory is implemented. Active devices in N^{3}ASICs are single type, doped dual channel crossed nanowire transistors (2C-xnwFETs). Area-distributed interfaces or vias are used to connect outputs of nanowire stages to a standard CMOS metal stack. Metal interconnections between vias achieve arbitrary routing. The nanowire logic plane is surrounded by CMOS circuitry. The peripheral CMOS circuitry can be used for control logic, dynamic clocking, mixed signal etc.

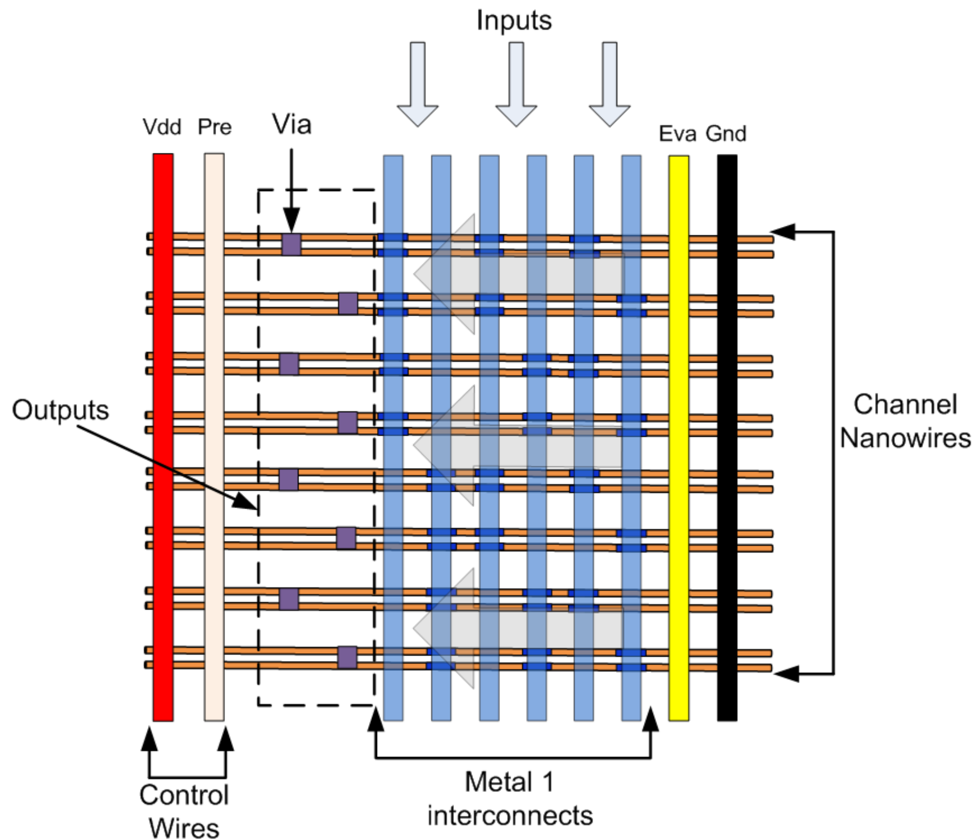

Figure 3.2. N^{3}ASICs input-output organization

Since vias and metal interconnects are used to contact the nanowires, fine-grained integration is possible. Fine-grained integration refers to the fact that every nanowire gate is able to communicate with a CMOS gate. The communication between the

Nano-CMOS layers is not limited to the periphery. Each input/output of a nanowire gate can be connected to the input/output of a CMOS gate. Fig. 3.2 shows the input-output organization in a N^3ASICs tile. All the channel nanowires are horizontal. The inputs are fed from the top onto metal 1 layer. VDD and GND contacts define the boundary of the single stage of a N^3ASICs tile. The outputs are available on the vias (shown in Fig. 3.2). These outputs can be routed to any other tile using metal interconnects.

3.2 CMOS Design Rules Applied to the Fabric

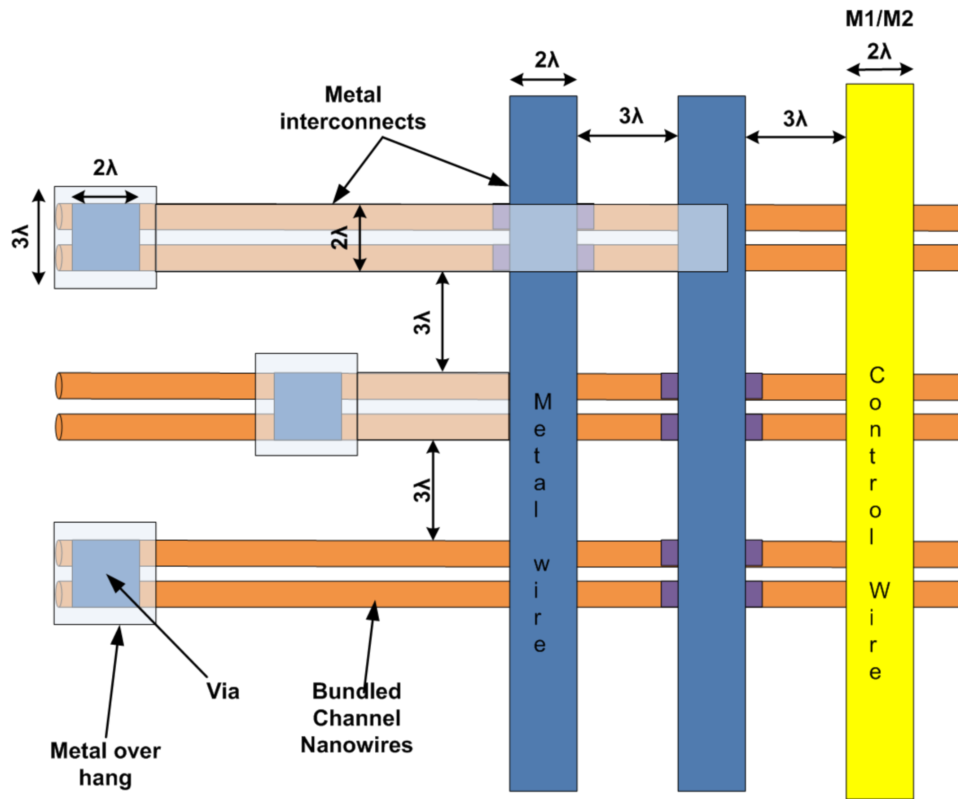

Figure 3.3. CMOS Design rules applied to N^3ASICs

To enable full and fine-grained integration with CMOS without any new manufacturing requirements, lithographic design rules need to be followed. Fig. 3.3 shows

representative λ design rules applied to the N^3ASICs fabric. All design rule requirements like Metal-Metal spacing, Metal-via spacing and Via-overhang are followed. C. Bencher et. al. [7] project that the metal 1(M1) pitch for the 16nm technology node is 40nm. This is equal to 5λ where λ=8nm for 16nm technology node.

Since metal vias are used to contact nanowires, the nanowire spacing should adhere to CMOS design rules. Given that nanowires can have much smaller dimensions than vias, more sub-lithographically patterned nanowires may be bundled within the same via dimension without any density impact. Having more than one nanowire per via allows for better contact, performance and inherent defect resilience, as will be shown in the subsequent chapters.

Fig. 3.3 shows how bundled pair of nanowires are contacted using a via. Metal 1 interconnects is used to connect the inputs of the transistors. Metal 2 interconnects are used to connect the output on the nanowires to the subsequent stages.

3.3 Assembly Sequence

We have seen that the order of manufacturing process helps in mitigating the manufacturing constraints when unconventional and conventional processes are used in conjunction. Here we present a simplified assembly sequence followed in building the N^3ASICs fabric.

The assembly sequence is as follows

- Creation of uniform semiconductor nanowire array
- Creation of lithographic contacts for VDD, GND, precharge and evaluate
- Metal gate deposition to define transistor positions (for any arbitrary functionality)
- Metal1 vias and interconnects to connect the inputs

- Metal2 interconnects to connect the signals across the logic planes

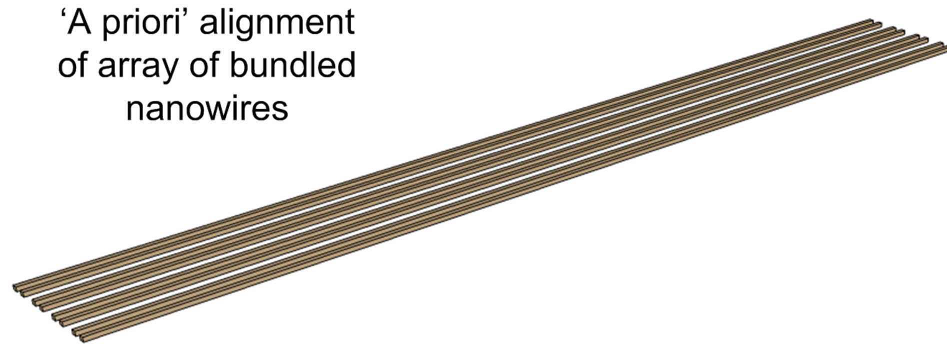

Figure 3.4. Patterned Nanowires

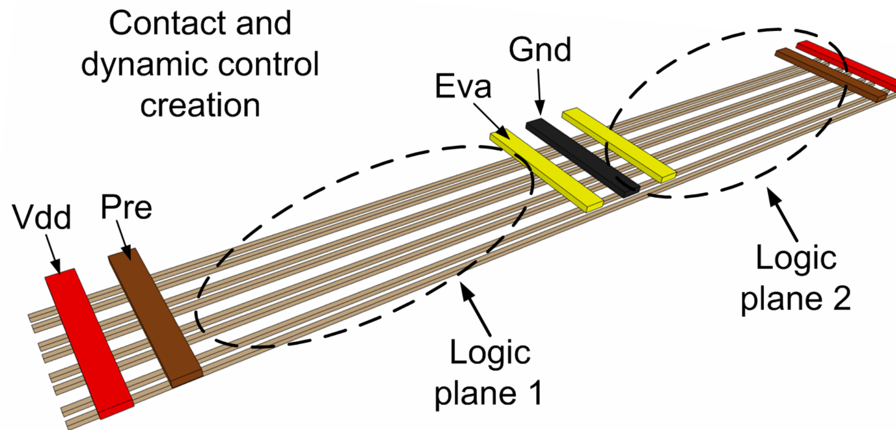

Figure 3.5. Creation of Lithographic contacts and dynamic control rails

At the bottom of the fabric is a uniform semiconductor nanowire array. This can be direct patterned on ultra-thin Silicon-On-Insulator. Nanowires can be bundled in pairs in order to achieve better contact with the vias. Fig. 3.4 shows the uniform dense nanowire array created *a priori* to any lithographic step.

Fig. 3.5 shows the contact creation for VDD and GND, precharge and evaluate. This diagram depicts the scenario of two stages cascaded next to each other. This can be treated as two logic planes as shown in the figure. We can use interconnects to route signals across the logic planes. Logic plane 1 is on the left and logic plane 2 is on the right

Fig. 3.6 shows the metal gate deposition step. Metal gates (shown in green) are deposited at certain positions to define 2C-xnwFETs using conventional lithography and masks. Initially the nanowires are doped p-type. A self-aligning ion implantation is then used to create n+/p/n+ source/channel/drain structures. This creates enhancement mode 2C-xnwFETs similar to conventional MOSFETs in CMOS. All device channels are oriented along the same direction and lie on the substrate itself.

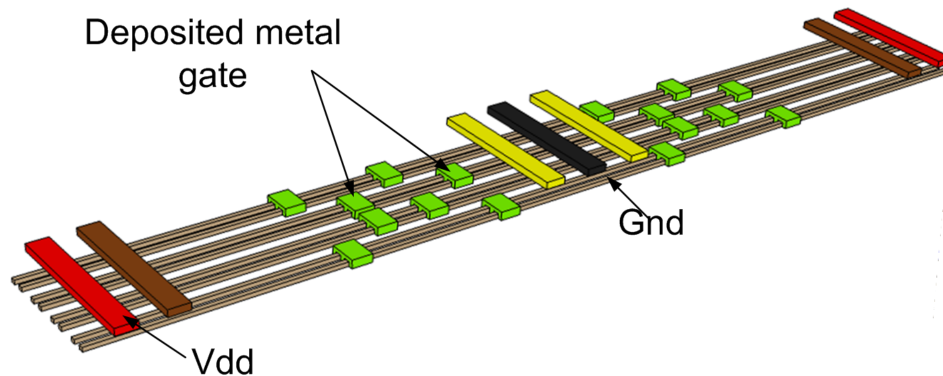

Figure 3.6. Metal gate deposition step

Fig. 3.7 shows the Metal 1 vias and interconnects. Metal lines and vias are laid down for interconnection. Inputs are received through an M1 array (light blue lines) and vias are dropped on to the nanowires to tap the outputs (blue dots).

As shown in Fig. 3.8, outputs from the left logic plane are cascaded to the inputs of the right plane using M2 (orange lines). The output of the second logic plane can

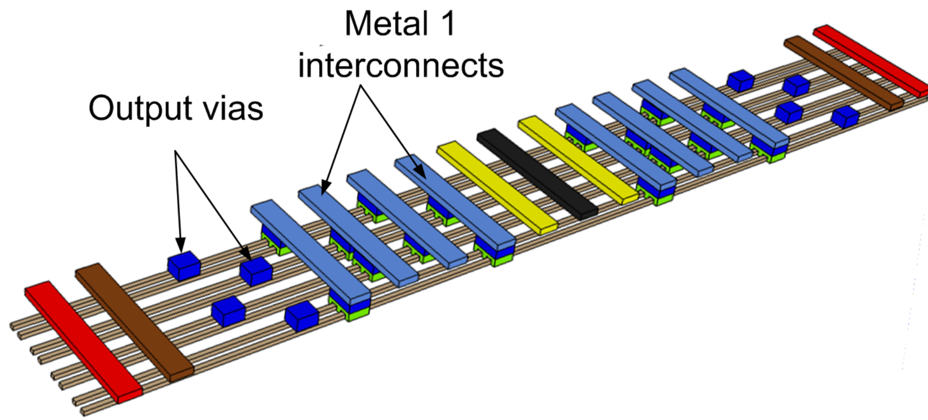

Figure 3.7. Metal 1 vias and interconnects

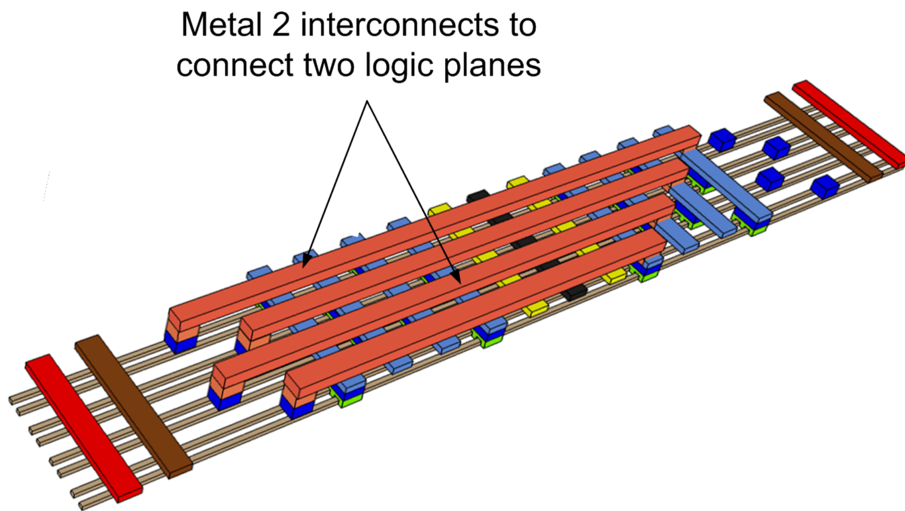

Figure 3.8. Metal 2 interconnects to route across logic planes

be routed to other tiles using higher metal layers in the metal stack. This allows us to achieve arbitrary routing between two different tiles. All local routing within a single stage is achieved on the nanowires themselves. This helps in reducing the routing overhead of the design.

3.4 Chapter Summary

In this chapter core concepts of the N^3ASICs fabric were introduced. It was shown how the CMOS design rules can be applied to the N^3ASICs fabric. A layer-by-layer assembly sequence was shown demonstrating how the fabric may be realized on a single Silicon-on-Insulator (SOI) wafer. This approach can be scaled to a large scale design with multiple cascaded logic planes.

In subsequent chapters novel dual-channel Crossed Nanowire Field Effect Transistors (2C-xnwFETs), the active devices in N^3ASICs are presented, associated circuit styles and interconnection approach are described and validated for functionality, a nanoprocessor design is implemented on N^3ASICs, and key system-level metrics, including area, power and performance are evaluated.

CHAPTER 4

N^{3}ASICS DEVICES AND CIRCUITS

4.1 Introduction

N^{3}ASICs evaluations were carried out at device, circuit and architecture level. An integrated device-fabric exploration methodology originally proposed for NASIC fabric was adopted [20]. The methodology is summarized in Fig. 4.1

Physical fabric choices impact the structure and properties of N^{3}ASICs devices. For e.g. if SNAP is used to pattern the bottom most ultra-dense nanowire layer, nanowires with square cross section will be obtained. Further, use of CMOS design rules facilitates bundling of nanowires because of the larger via dimension compared to nanowires. Hence, dual-channel devices can be used in N^{3}ASICs. For this device structure the electrical properties are obtained from Synopsys Sentaurus^{TM} [5]. Using this data, behavioral model compatible with HSPICE [3] is created. This behavioral model is used to carry out circuit and system level evaluations.

The device and the circuit level evaluations will be presented in this chapter. System level evaluation and comparison with 16nm CMOS will be presented in the next chapter.

4.2 Device Structure

The use of standard design rules and lithography for manufacturing determines device structure and dimensions. Given that channel nanowires could have much smaller dimensions than metal vias, they are bundled into pairs to make better contact, and provide for dual channel FETs. The 2C-xnwFET with an omega-like metal

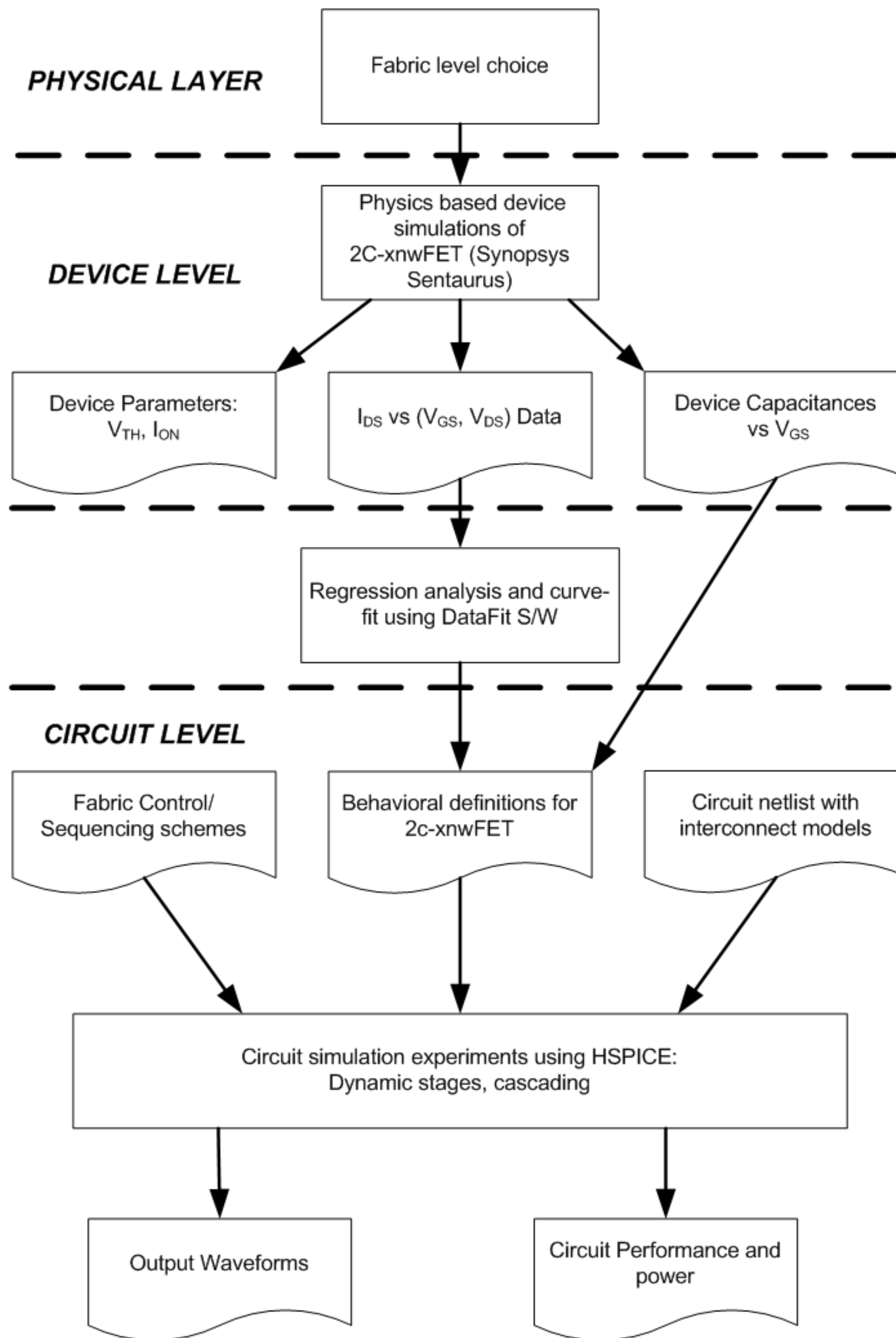

Figure 4.1. Integrated Device-fabric exploration methodology

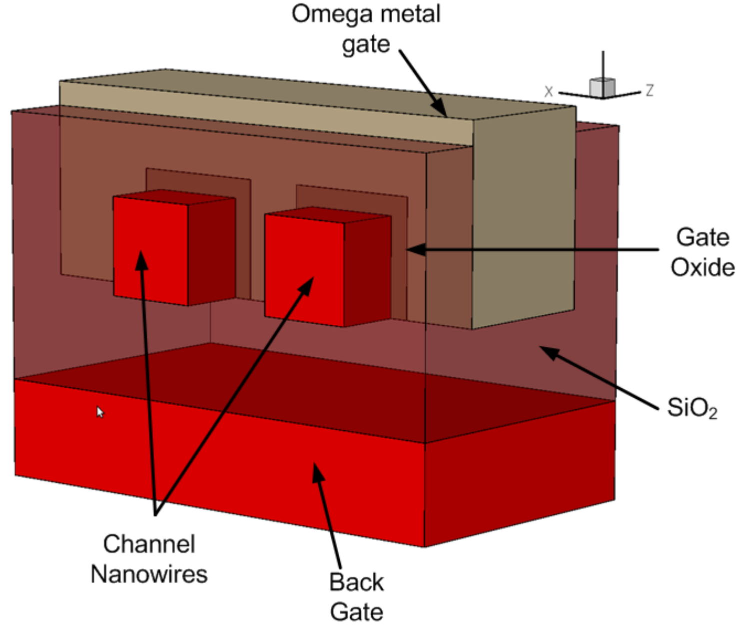

Figure 4.2. 3D structure of N^3ASICs device (2C-xnwFET)

gate is shown in Fig. 4.2. The gate width and the channel length of the device are defined by the technology node as they are lithographically defined. So, for the purpose of study, devices with 16nm gate lengths were simulated. A high-k dielectric (HfO_2 [9]) was used as gate oxide material. A gate self-aligned process with etch back can be used for defining the oxide structure.

As HfO_2 (high-k gate dielectric) is used, metal gates [27] are preferred over the regular poly silicon gates. Polysilicon gates are not suitable with HfO_2 as they cause V_{TH} instabilities and mobility degradation [9]. Moreover fully silicided metal gates have very low resistivity and do not have the problem of gate depletion either with SiO_2 or HfO_2. Further they allow work function engineering for V_{TH} tuning. Gate first [8] or gate last [15] processes can be employed in order to build the gate.

As opposed to the conventional top-gated device structures, the Omega-gated structure (somewhat similar to multi gate FETs [23]), provides better electrostatic control of the channel. A better electrostatic control over the channel gives a higher on

to off current ratio. The use of dual channels implies higher on-current, with potential benefits for system-level performance. Furthermore, the dual-channel structure implies inherent defect resilience against broken nanowires and some types of stuck-off defects, without any density impact. Even a single correctly functioning nanowire can still produce the correct output (but with a larger delay). In general, Stuck-off defects are very difficult to mask and dual channel provides a way of alleviating it. On the other hand, stuck-on defects can be masked fairly easily with structural redundancy.

4.3 Device Simulations

Device simulations were done using Synopsys Sentaurus. These device-level simulations provide 3 sets of data: i) Current data (I_{DS}) for different values of drain-source (V_{DS}) and gate-source (V_{GS}) voltages, ii) Device capacitances at different values of V_{GS}, and iii) device parameters that determine noise margins and performance of the devices such as the on-currents (I_{ON}), threshold voltage (V_{TH}). We can adjust these device parameters by changing the metal gate workfunction or substrate bias (e.g. a higher threshold voltage may be obtained by modifying the metal work function or using a more negative back gate bias).

Dual-Channel Crossed Nanowire FETs (2C-xnwFETs, Fig. 4.2) were extensively characterized using accurate physics-based 3D simulation of the electrostatics and operations using Synopsys Sentaurus^{TM}. The 2C-xnwFETs employ metal Omega gate structures for tighter electrostatic control. Gate material work function is 4.6 eV. 16nm channel devices were simulated given that it is the minimum feature size for lithographically defined gates. The notation N^3ASICs-16 represents N^3ASICs constructed with 16nm CMOS design rules, which implies λ the scale length, is equal to 8nm. The channels are doped p-type of the order of 10^{18} cm^{-3} and the source/drain regions were doped n-type of the order of 10^{20} cm^{-3}. A substrate bias of -3V was assumed to deplete the channel and adjust device parameters such as threshold voltage

Table 4.1. Devices Simulation Parameters

Parameter	Value
Gate Material	Metal
Gate Workfunction(eV)	4.6
Channel Doping (cm^{-3})	10^{18}
Gate Oxide Material	HfO_2
Gate oxide thickness (nm)	3
Bottom oxide material	SiO_2
Bottom oxide thickness (nm)	10
Back Gate bias (V)	-3
Source/Drain doping (cm^{-3})	10^{20}

and on/off current ratios for correct cascading. A high-k HfO_2 material is used for gate oxide.

The gate oxide thickness was 3nm. Drift diffusion transport models [30] were used to simulate the 3D devices. Simulations were calibrated to account for interface scattering, surface roughness and interface trapped charges as explained in [20]. Table 4.1 summarizes the parameters used for Device simulations.

Drain current vs. drain voltage (I_{DS}-V_{DS}), drain current vs. gate voltage (I_{DS}-V_{GS}), and different parasitic capacitances vs. gate voltage (C vs V_{GS}) were simulated. On-current (I_{ON}) and on/off (I_{ON}/I_{OFF}) current ratio were extracted. Fig. 4.3 shows the I_{DS}-V_{DS} curve for different V_{GS} values. Fig. 4.4 shows the I_{DS}-V_{GS} curves for different V_{DS} values. These simulations verify inversion mode behavior for 2C-xnwFETs with a positive threshold voltage.

Table 4.2 shows key device simulation results for N^3ASICs-16 2C-xnwFET. With a high on current, $V_{TH} > 0.2$, and $I_{ON}/I_{OFF} > 10^4$ the devices meet circuit requirements for correct functionality and noise.

Various capacitances at different values of V_{GS} were extracted from Synopsys Sentaurus. The figure shows the Gate capacitance with respect to V_{GS}. A plot of the gate capacitance C_G vs V_{GS} is as shown in Fig. 4.5.

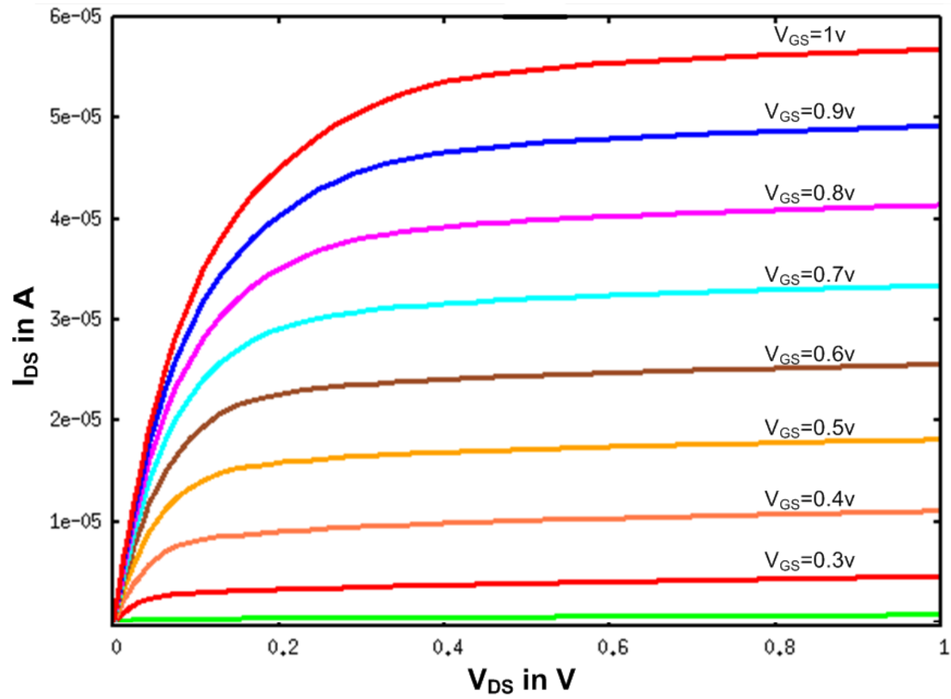

Figure 4.3. I_{DS} vs V_{DS} with varying V_{GS} for 2C-xnwFET

Table 4.2. Devices Simulation output

Parameter	N^3ASICs-16 2C-xnwFET
V_{TH}	0.27
I_{ON}	39.6μA
I_{ON}/I_{OFF}	26218

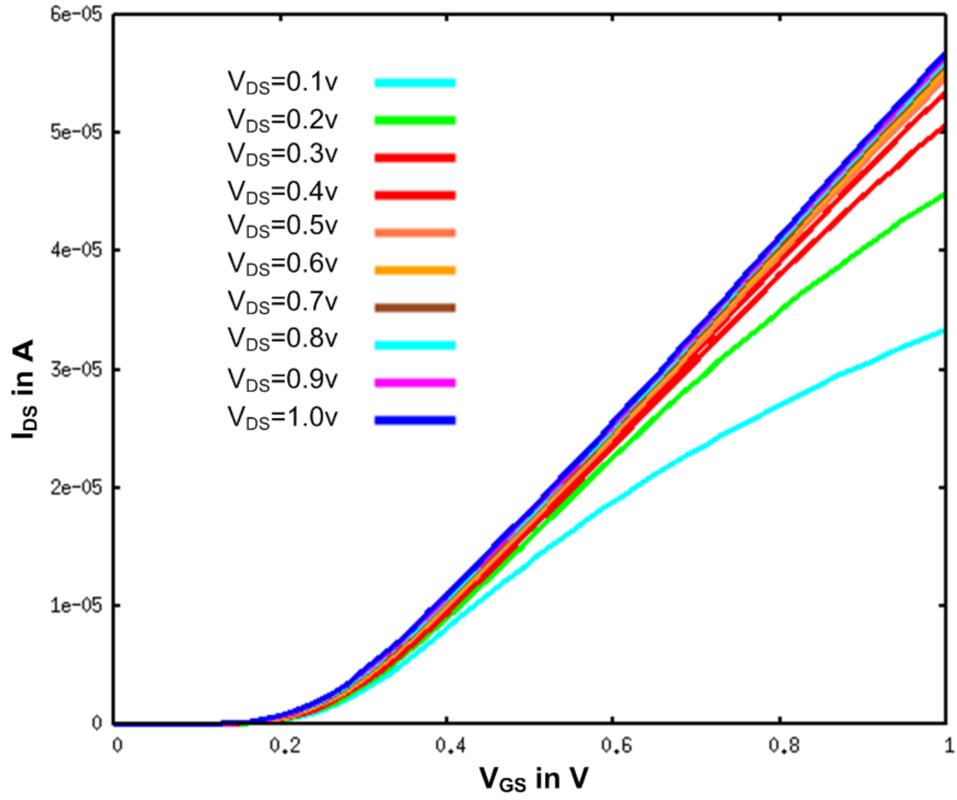

Figure 4.4. I_{DS} vs V_{GS} with varying V_{DS} for 2C-xnwFET

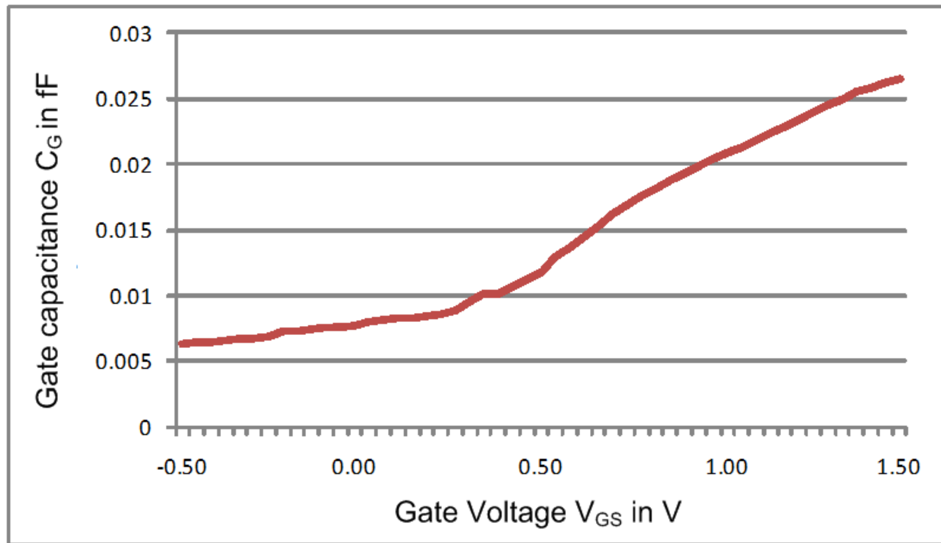

Figure 4.5. Gate capacitance vs V_{GS} for 2C-xnwFET

We see that the gate capacitance increases with increases in gate source voltage. The maximum gate voltage is 1V. Hence the maximum gate capacitance seen at any input will be around 20 aF.

4.4 Behavioral Model Creation for Circuit Simulation in HSPICE

The current data is fitted as a function of V_{GS} and V_{DS} using regression analysis and curve fitting [22]. An expression representing the current as a mathematical function of V_{GS} and V_{DS} is obtained from the curve-fit. The expression for the current, in conjunction with a piecewise linear approximation for the device capacitances forms a behavioral model of the xnwFET, which may be incorporated into HSPICE to carry out circuit level evaluations.

A regression based [22] approach is very generic and can be used to fit arbitrary device characteristics. Coefficients extracted from regression data fits are representative of the device behavior over sweeps of drain-source and gate-source voltages. This is in contrast to conventional in-built models in SPICE for MOSFETs and other devices, which use analytical equations derived from theory and physical parameters such as channel length and width. The regression coefficients in our approach may not directly correspond to conventional physical parameters. Therefore different regression fits will need to be extracted for devices with varying geometries, doping etc.

4.5 Circuit Style and Evaluations

N^3ASICs uses a dynamic circuit style similar to the circuit style employed by NASICs [22]. These dynamic circuit styles are amenable to implementation on regular nanowire arrays without the need for complementary devices, arbitrary sizing or placement, simplifying manufacturing requirements of N^3ASICs. It uses single type of

FETs to realize logic without the need for complementary devices or arbitrary doping profiles which significantly reduces customization and manufacturing requirements.

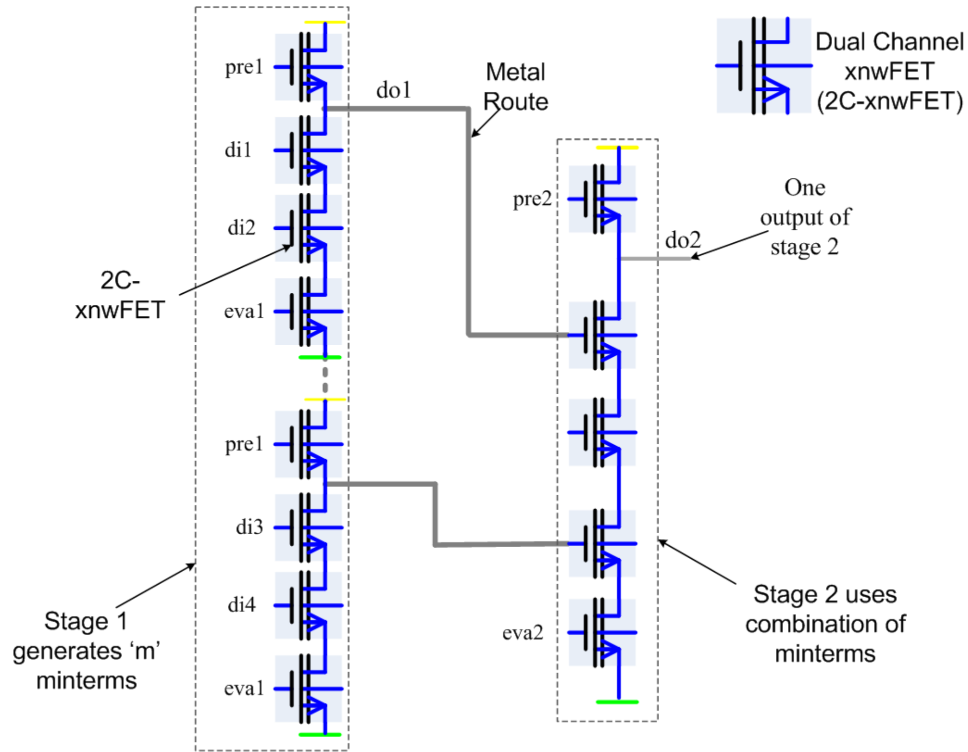

Figure 4.6. Schematic diagram of a sample circuit to illustrate how 2 stages of N^3ASICs are connected

Fig. 4.6 shows a circuit-level abstraction of cascaded NAND-NAND stages realized on the N^3ASICs fabric using n-type 2C-xnwFETs. All the outputs are precharged to logic 1 and if all inputs are logic 1, the output discharges to logic 0. All the control signals (precharge and evaluate) are active high. The outputs of the first stage act as inputs to the second stage. Logic customization is limited to defining the positions of the 2C-xnwFETs on the logic planes.

One dynamic sequencing scheme for cascading is shown in Fig. 4.7 [20]. In this scheme, successive stages are clocked using different precharge and evaluate signals, with hold phases inserted for correct cascading. During a hold phase, the output node of a given stage is implicitly latched, and used for evaluation of the next stage, similar

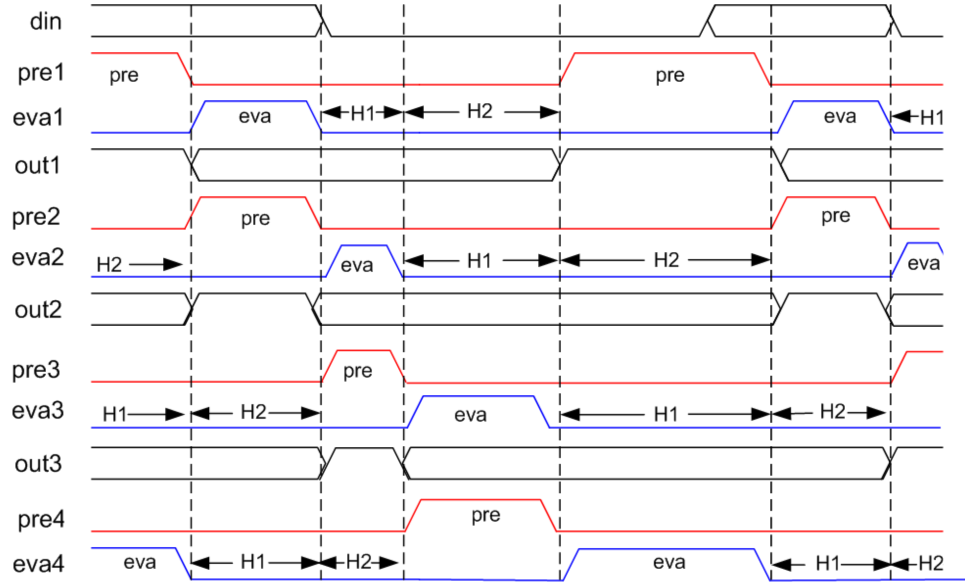

Figure 4.7. Four phase clocking scheme

to [20] [22] [35]. Implicit latching implies that area expensive latches or flip-flops requiring complementary devices/local feedback paths are not needed.

Fig. 4.8 shows the top view of a 1-bit full adder circuit built using two N^{3}ASICs logic planes. Stage 1 generates the minterms based on the inputs (marked stage 1 outputs). Minterms are fed to stage 2 using horizontal metal interconnects. Stage 2, using a combination of minterms generates different outputs. The outputs available on the right side of this stage can be routed to subsequent tiles using additional metal interconnects. Fig. 4.9 shows the cross sectional view of a cross point in the N^{3}ASICs tile.

Simulations were carried out using the behavioral models in HSPICE to evaluate the performance and power of N^{3}ASICs design. Since vias and metal interconnects are used to route signals, CMOS interconnect models are necessary to evaluate the performance of N^{3}ASICs. The interconnects were modeled using the Predictive Technology Model (PTM) [2] [40] models. The dimensions and parameters for scaled CMOS interconnect were chosen as projected by ITRS [1] and [7]. With the help of

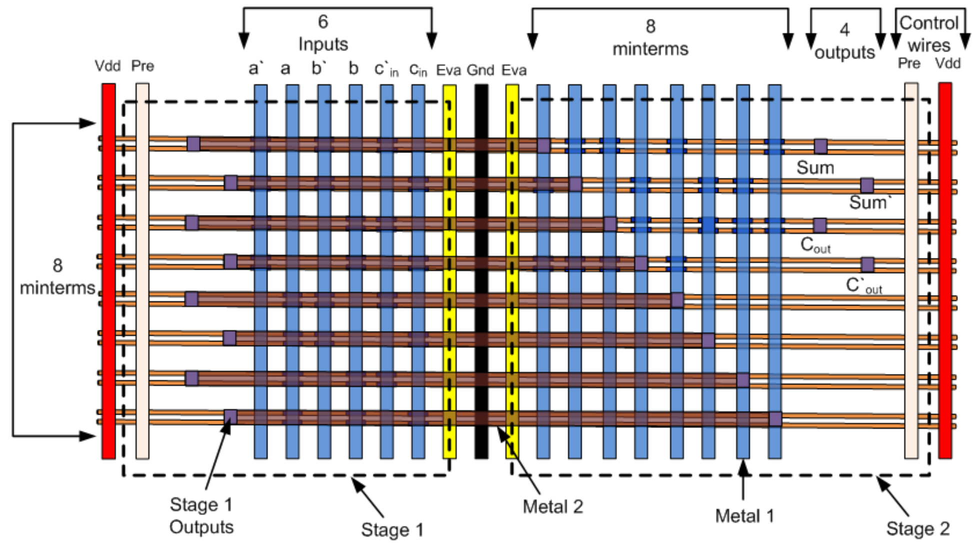

Figure 4.8. N^{3}ASICs 1 bit full adder top view

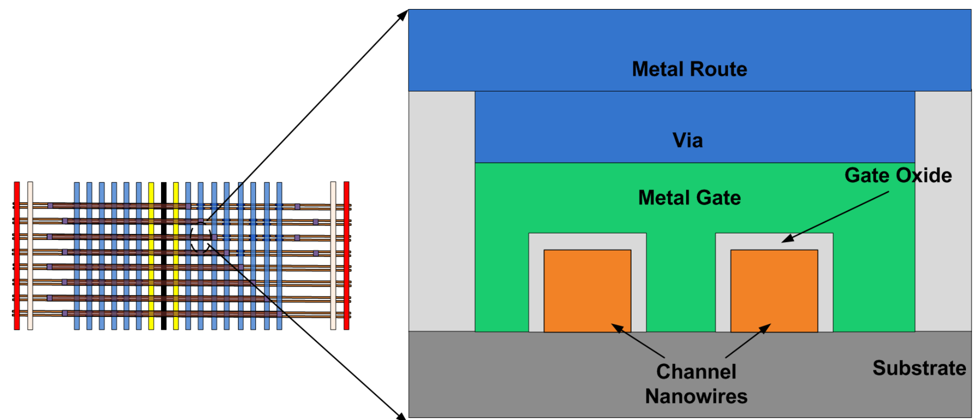

Figure 4.9. Cross sectional view of a cross point in N^{3}ASICs

behavioral models, HSPICE simulations were carried out to verify functionality and measure the performance and power of N^3ASICs.

The full-adder in Fig. 4.8 was simulated in HSPICE to verify expected circuit level behavior. Fig. 4.10 shows the output waveforms of the one bit full adder simulated in HSPICE with the behavioral model. These simulations verify functionality of the circuits and adequate noise margins. It can be noted that the data on the output node is latched during the hold phases thereby exhibiting the implicit latching behavior.

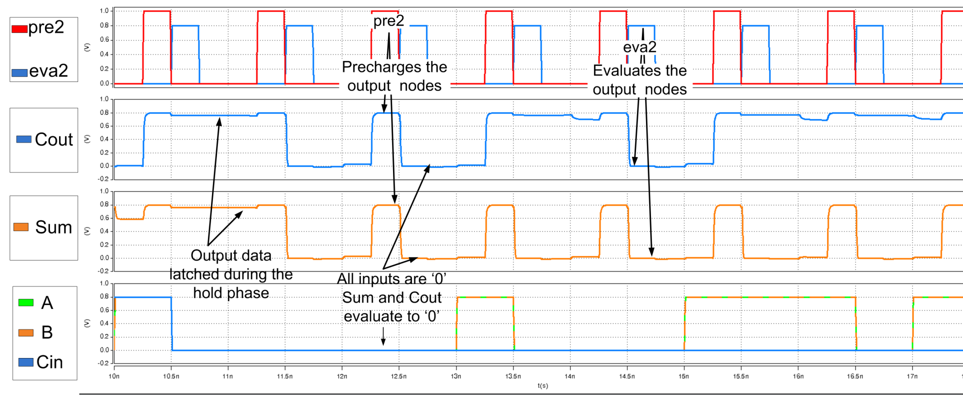

Figure 4.10. Simulation waveforms of N^3ASICs One bit full adder

4.6 Chapter Summary

In this chapter device-fabric exploration methodology was introduced. Extensive device simulations of 2C-xnwFETs were shown. It was seen that the simulated devices met the circuit requirements with positive V_{TH} and 4 orders of magnitude I_{ON}/I_{OFF} ratio. Using the device level data, HSPICE compatible behavioral models were created. Circuit simulations were carried out to validate the N^3ASICs circuits. One possible sequencing scheme was shown here, while other variants can also be used. We show how careful device design and sequencing schemes help us achieve implicit latching on the nanowires which means area expensive flip-flops and latches are not necessary to latch the data.

In the following chapter, we will present the system level evaluations of N^{3}ASICs. The results obtained will be compared against an equivalent 16nm CMOS design.

CHAPTER 5

SYSTEM LEVEL EVALUATION

In the previous chapter we looked at the Device I-V and C-V characteristics, reflecting accurate 3-D physics. Circuit level simulations were carried out to verify functionality. In this chapter, system-level metrics such as density, power and performance are evaluated for a N^3ASICs processor design WISP-0 and compared against a 16nm CMOS baseline.

5.1 WISP-0

This section provides details about WISP-0 [16] [36], used to evaluate N^3ASICs. Wire Streaming Processor version-0 (WISP-0) is a stream processor that implements a 5-stage microprocessor pipeline architecture including fetch, decode, register file, execute and write back stages. WISP-0 consists of five nanotiles: Program Counter (PC), ROM, Decoder (DEC), Register File (RF) and Arithmetic Logic Unit (ALU). Fig. 5.1 shows its layout. It uses dynamic circuits and pipelining on the wires to eliminate the need for explicit flip-fops and therefore improve the density considerably. WISP-0 is used as a design prototype for evaluating key metrics such as area, performance and power. 16nm CMOS equivalent of WISP-0 was developed to compare N^3ASICs-16.

5.2 CMOS Baseline WISP-0

A 16nm static CMOS baseline was created using the following methodology. A functional description of WISP-0 was written in Verilog. Using Synopsys Design

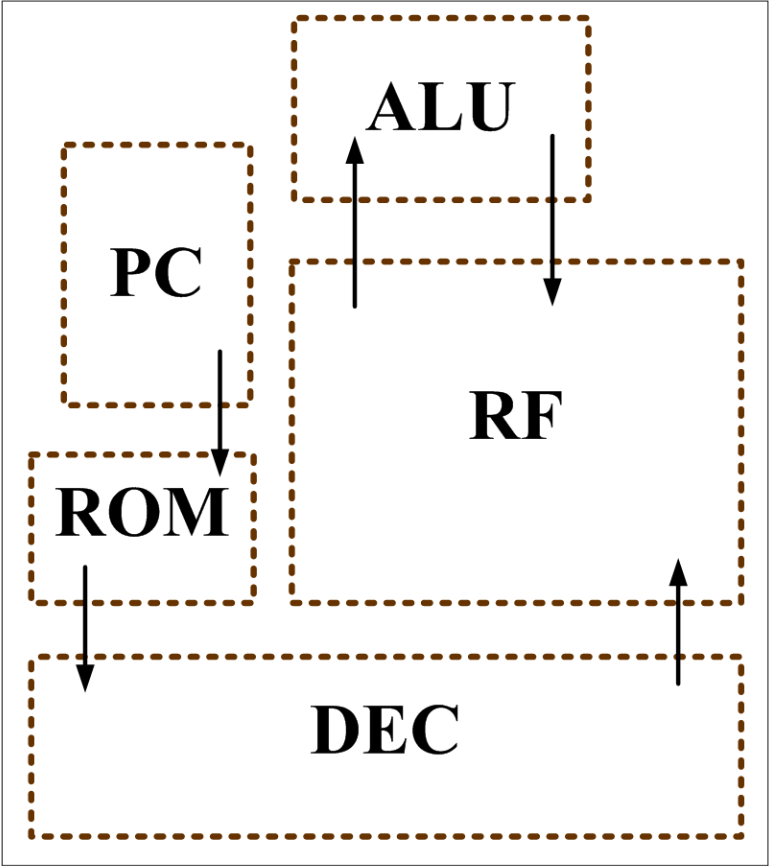

Figure 5.1. WISP-0 Nanoprocessor layout

Compiler [4] and a 45nm IBM standard cell library, a gate level Verilog netlist was created. This was converted to a SPICE netlist using the nettran utility. A standard cell library for SPICE was obtained and device dimensions were scaled to the 16nm technology node. The SPICE netlist, library and PTM 16nm MOSFET models were used to run circuit level simulations in Synopsys HSPICE to characterize the power and performance of the CMOS design. This methodology is summarized in the flow diagram in Fig. 5.2. It is seen that the best operating frequency for a 16nm CMOS design at the nominal voltage of 0.7V is 6.25GHz. Power consumption of WISP-0 was obtained from HSPICE.

In order to obtain the area estimate of 16nm WISP-0, placement and routing was carried out on the 45nm synthesized netlist. The area numbers so obtained were quadratically scaled to obtain the 16nm area numbers.

5.3 N^{3}ASICs WISP-0

A HSPICE circuit definition of the entire WISP-0 was created with proper interconnects to calculate the power and performance of N^{3}ASICs-16 WISP-0. The behavioral models created for 2C-xnwFETs were used. It is important to model the metal interconnects while estimating the power and performance, since metal interconnects is used to route the signals in N^{3}ASICs. PTM interconnects models were used to obtain the RC value of interconnects. The parameters chosen for the interconnects were in accordance with ITRS and [7].

The area of the N^{3}ASICs WISP-0 was calculated based on the design rules and the number of metal tracks. The area of each tile depends on the number of inputs, outputs and the number of minterms used to realize the logic. This is a two stage NAND-NAND logic. Minterms are generated in the first stage and a combination of minterms is used to produce the outputs.

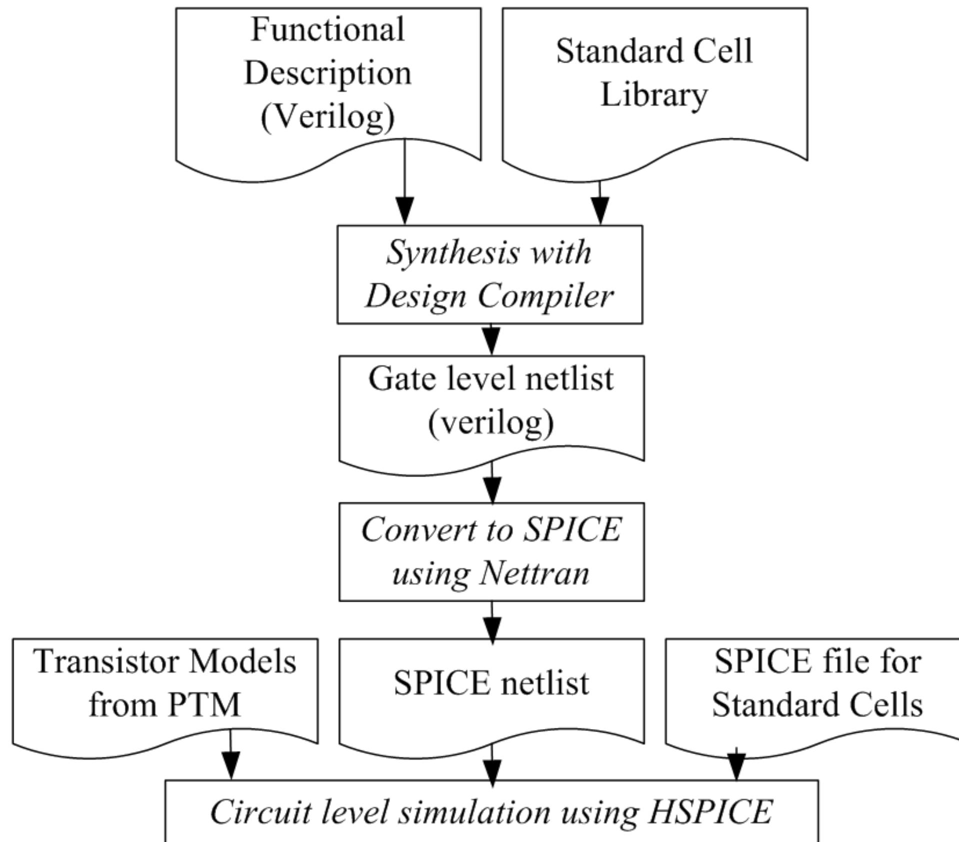

Figure 5.2. Methodology for performance characterization of 16nm static CMOS baseline

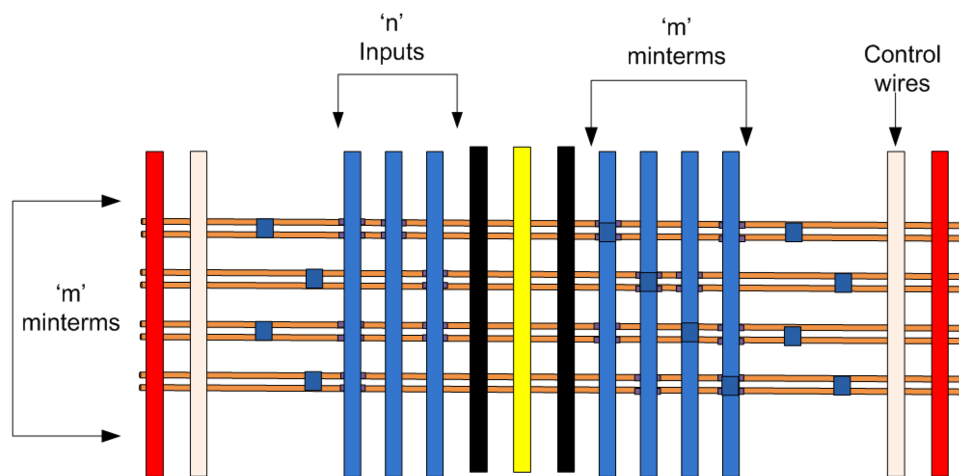

Figure 5.3. A N^3ASICs tile. Area calculation example

The area of a tile (shown in Fig. 5.3) with n inputs, o ouputs and m minterms will be

$$(n * 5\lambda + 7 * 5\lambda + m * 5\lambda + 24\lambda)X(m * 5\lambda) \tag{5.1}$$

where, 5λ is the Metal 1 pitch. The components in equation 5.1 are

- Components in the length dimension
  - n * 5λ - n inputs pitch of M1 layer
  - 7 * 5λ Metal rails for contacts and dynamic clocking
  - m * 5λ -m minterms generated in the first stage which act as inputs to the second stage
  - 24λ- for the vias on either side
- Components in the width dimension
  - m * 5λ m minterms and the pitch of Metal 1

5.4 N^{3}ASICs-16 and 16nm CMOS Comparison

Fig. 5.4 shows the density advantage of N^{3}ASICs at various technology nodes. The proposed N^{3}ASICs-16 is 3X denser compared to 16nm CMOS. The density advantage of N^{3}ASICs is due to the dense nanowire array at the bottom (implying the use of devices with smaller dimensions when compared to conventional CMOS FETs), use of single type FET to realize logic, implicit latching on the nanowires (which ensures that there is no need for area expensive latches and flip-flops) and finally reduced transistor count compared to CMOS. Since CMOS design rules are used for pitch and spacing, the scaling trend is almost constant across different technology nodes considered.

As the nanowire layer confirms to CMOS design rules, the spacing between the nanowires is greater compared to a 2-D grid based NASIC fabric. While the NASIC

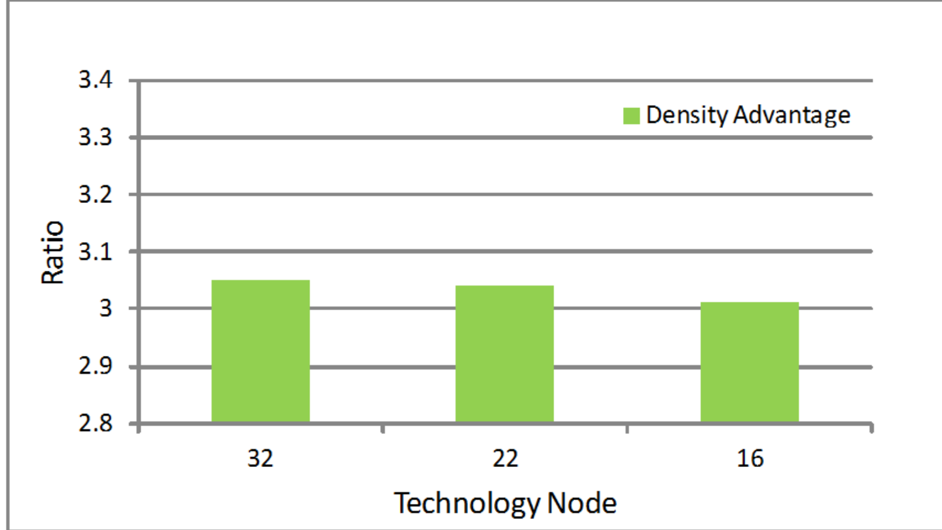

Figure 5.4. Density Comparison of N^3ASICs with CMOS at different technology nodes

Table 5.1. Key system level metrics for WISP-0

	Area(μm^2)	Performance(GHz)	Power(μW)
CMOS Baseline(16nm)	66.24	6.25	77.90
N^3ASICs-16	22	6.32	14.36
Relative Improvement	3.01	1.01	5.42

fabric is 33X denser [16] than functionally equivalent CMOS WISP-0 design, the use of design rules, while alleviating manufacturing requirements, reduces the density advantage of N^3ASICs to 3X. The evaluation results are summarized in the table.

Power and performance comparisons are shown in Table 5.1. We notice that the performance of N^3ASICs-16 is comparable to that of 16nm CMOS equivalent WISP-0. These simulations do not consider key optimizations for 2C-xnwFETs making comparisons pessimistic. For example, while the PTM models employ strained silicon, no straining was assumed for 2C-xnwFETs. It is expected that a better mobility and hence better performance could be obtained when straining techniques are employed in N^3ASICs.

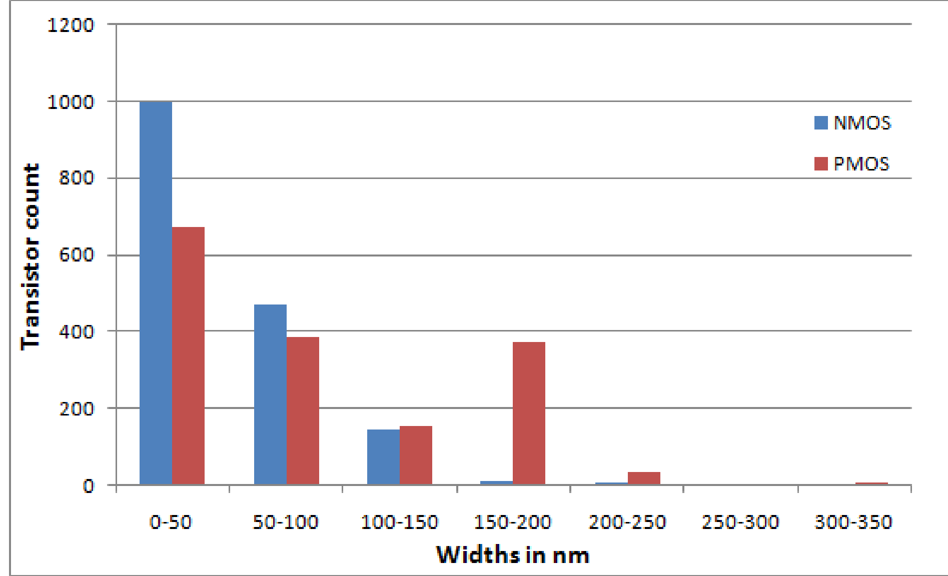

Figure 5.5. Transistor width distribution in 16nm CMOS-WISP-0

A significant reduction in average power of 5.4X was observed in case of N^{3}ASICs-16. To clearly explain this, experiments were carried out with different circuits and varying number of inputs. With the voltage and the frequency of operation being the same, the capacitances were investigated. Since there is no arbitrary sizing in the case of N^{3}ASICs and all 2C-xnwFETs are identical, the maximum input gate capacitance is always 20.42aF (Fig. 4.5). In case of the CMOS WISP-0 design, the transistors are sized, contributing to increased gate capacitance. The input gate capacitance in the case of minimum sized inverter in CMOS is 75.14aF which is more than 3.5X that of the N^{3}ASICs. The largest NMOS device used has a gate capacitance of 135.4aF and the largest PMOS device has a gate capacitance of 372.38aF. A plot of the distribution of the transistor widths in the case of WISP 0-CMOS is shown in Fig. 5.5. Since a dynamic logic style with only single type FET is used, N^{3}ASICs-16 uses a fewer number of transistors to realize the logic. Implicit latching [35] [36] of signals on the nanowires further reduces the number of transistors required. The transistor counts were 1306 and 3252 in case of N^{3}ASICs and CMOS respectively. With the

use of transistors of various widths, the gate capacitance further increases leading to increased dynamic power consumption for CMOS WISP-0.

5.5 Chapter Summary

Detailed system level evaluations were carried out using WISP-0 nanoprocessor as the test case. 16nm CMOS equivalent of WISP-0 was developed in order to compare the area, power and performance. N^3ASICs design is 3X denser than 16nm CMOS equivalent design. It was seen that N^3ASICs was able to achieve comparable performance at 5X lower power consumption. This might be a pessimistic comparison because the PTM models used to compare the results make use of straining where as the N^3ASICs devices don't. In the further chapters we will look at some logic partitioning examples and its impact on area, power and performance.Also systematic yield implications of mask overlay misalignment on the fabric will be presented.

CHAPTER 6

LOGIC PARTITIONING STUDY

6.1 Introduction

All the logic tiles in N^{3}ASICs are implemented as NAND-NAND stages. While two-level logic can implement any arbitrary function, in general it does not scale well with increasing number of inputs. As the complexity of the implemented function increases, there is an exponential increase in the number of product terms required to realize the logic. This in-turn leads to increased transistor count and might degrade the power and performance of the system. Further it might not be the most area efficient way to realize a given functionality. In order to overcome these limitations we investigate how complex logic can be realized in N^{3}ASICs while retaining the dynamic logic style, single type of FETs and taking advantage of implicit latching

With the help of smaller two-level logic tiles in conjunction with intelligent clocking schemes we can realize complex logic with less overhead compared to a brute force two-level logic approach. The key idea is to divide the logic into smaller tiles and leverage metal routing stacks to connect the tiles to realize the complex logic function. While individual tiles still implement two-level logic, we expect the overall area/performance impact to be less when compared to a full blown two-level logic implementation. This approach is proposed with careful consideration to the fabric vision that was developed. This does not make use of complementary devices and there is no arbitrary sizing or placement of the devices. In the following section we will present the motivation for such an approach and subsequently present the clocking schemes for the same.

6.2 Case study

In this section we evaluate two different circuits to examine how two level logic scales with increased complexity of the function that is being implemented.

6.2.1 Two bit adder

Two approaches that are compared here are

1. A two-level logic implementation of 2 bit adder (without partitioning)

2. Two one bit full adder tiles connected to form a 2bit ripple carry adder (with partitioning)

6.2.1.1 Two-level implementation of a 2bit full adder

In this approach the outputs are expressed as sum-of-products of the inputs and are directly realized using two-level NAND-NAND stages (Fig. 6.1).

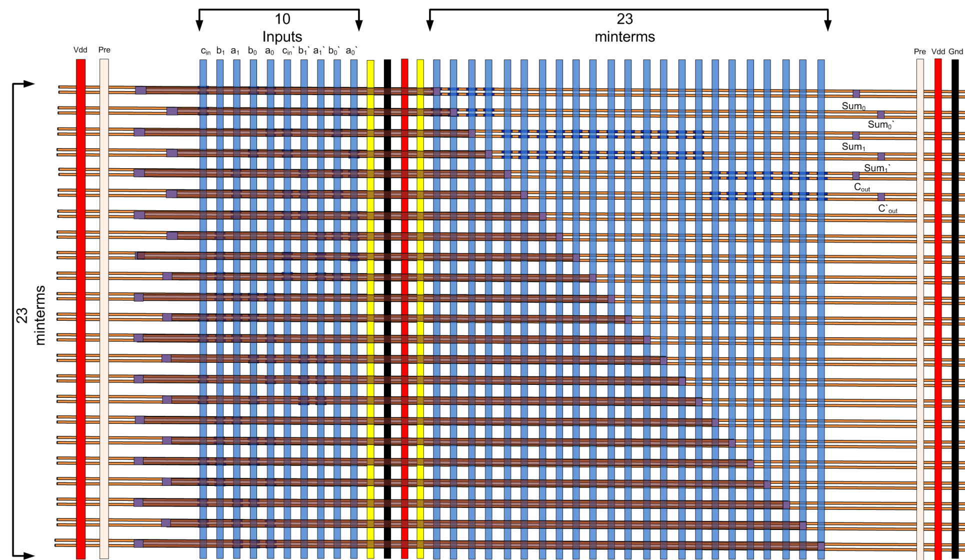

Figure 6.1. A two-level two-bit adder (Top view)

Table 6.1. Comparison of different metrics of the two approaches for a 2bit adder

Metric	Without partitioning	With partitioning
Number of transistors	182	128
Number of product terms	23	8 in each tile
Max Delay (ps)	126.47	97.16
Average power (μW)	3.40	2.14
Area(μ^2)	1.607	0.644

6.2.1.2 With partitioning

In this approach the two bit adder is realized in a ripple carry fashion. The ripple carry adder is as shown in Fig. 6.2, comprising of two one-bit full adders.

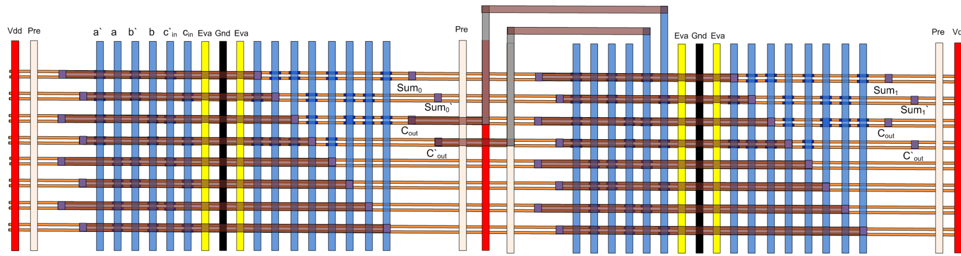

Figure 6.2. Partitioned N^{3}ASICs two-bit adder

6.2.1.3 Comparison of the two approaches

The Table .6.1 compares the two different approaches of the adders implemented in N^{3}ASICs. From the table it is clear that as we increase the design complexity, the overhead increases without partitioning. While the number of transistors required in implementing a 1bit Full adder is 64, the transistor count increases to 182 for a 2bit adder. The delay without partitioning increases due to increased number of transistors on the evaluate stack in the second stage. There is a maximum of 12 transistors in the evaluate path without partitioning and requires 2.5X more area, 1.5X more power. This increases further as we go to higher bit-widths. For example,

Table 6.2. Comparison of different metrics for the two approaches for a (7,3) counter

Metric	Without partitioning	With partitioning
Number of transistors	1316	256
Number of product terms	127	8 in each tile
Max Delay (ps)	768.41	145.74
Average power (μW)	4.36	1.23
Area(μ^2)	30.27	1.29

a two-level 4-bit adder without partitioning would require 988 more transistors and 26X more area compared to a partitioned 4-bit adder.

6.2.2 (7,3)-Counter

(n,m) parallel counters count the number of logic 1s out of n input bits and yield m = $\log_2$(n+1) output bits and are commonly used in fast multipliers. In this section, we investigate how partitioning impacts area, power and performance of a (7, 3) parallel counter and compare it to a design without partitioning. With partitioning the (7,3) counter is realized using four 1-bit full adders as shown in Fig. 6.3. From the

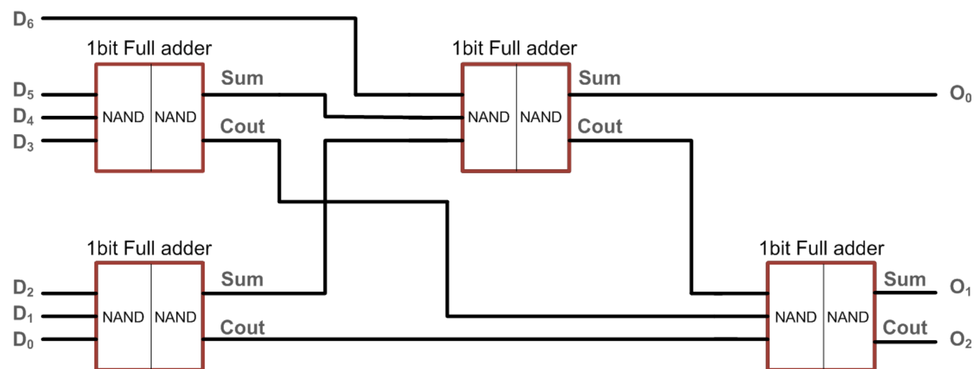

Figure 6.3. Partitioned (7,3) counter

Table. 6.2 partitioned design is 5X faster than the one without partitioning. This is due to the fact that the maximum number of transistors on the evaluate stack for the unpartitioned design is 64, which significantly impacts the evaluation delay. Also the

number of transistors required in case of an unpartitioned design is almost 5 times that of the partitioned design and it is 23X denser.

A similar study was carried out for the NASIC fabric. It was shown that partitioning of the ALU block into smaller tiles helped in achieving better performance [31]. Further, fewer transistors are required in case of the partitioned approach compared to the original design. Hence, partitioning the design aids in realizing complex logic functions.

Since, these dynamic circuits exhibit implicit latching behavior, additional latching overhead is not incurred when partitioning the design. The partitioning algorithms used for partitioning PLAs [26] can be adopted to partition the design into smaller tiles.

In the next section we will discuss some of the timing schemes that can be adopted with the partitioned design. These timing schemes can be modified to tune the circuit to obtain area, power or performance benefits.

6.3 Study of Clocking schemes for partitioning approach

With a partitioned design, a variety of sequencing schemes might be employed. We can tailor the sequencing schemes to obtain better performance at the cost of area and power or we can tailor the sequencing schemes to obtain low power and lesser area at the cost of performance. We demonstrate this with the help of two generic sequencing schemes evaluated for various circuits. One of the sequencing schemes is the 4-phase scheme that was presented earlier (Fig. 4.7). The above representative functional unit (Fig. 6.4) when implemented with 4-phase clocking scheme would require additional identity tiles in order to balance the path as shown in Fig. 6.5. Another approach is by having more number of clock phases. This would mean that we would need less identity tiles to balance the paths. For sample functional unit, we

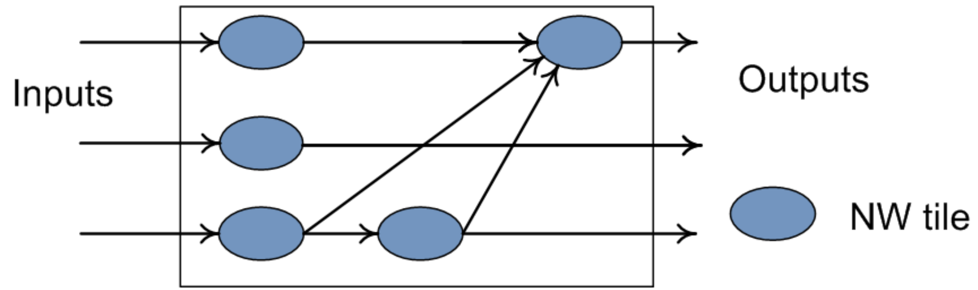

Figure 6.4. Sample functional unit

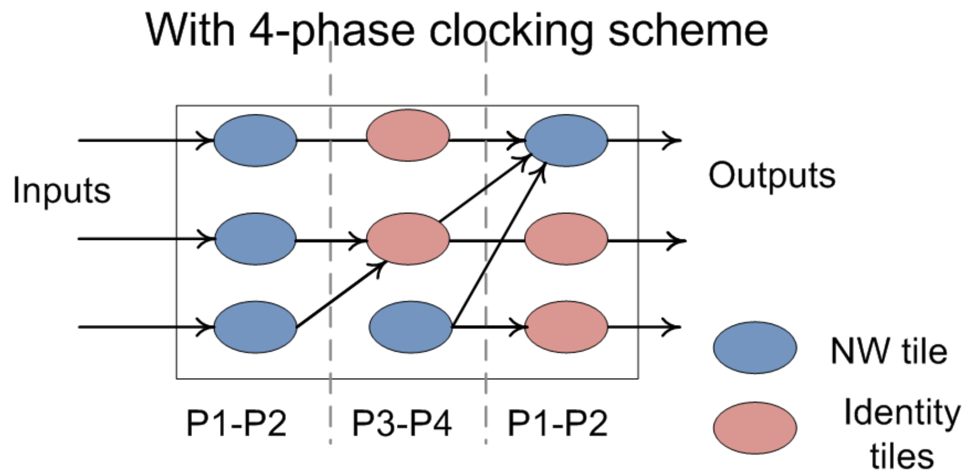

Figure 6.5. Functional unit with 4-phase clocking scheme

Table 6.3. Comparison of 4-phase and 6-phase clocking schemes for (7,3) counter

	4-phase clocking	6-phase clocking
Number of transistors	2926	256
Max Delay (ps)	97.16	145.74
Average power (μW)	5.28	4.48
Area(μ^2)	1.53	1.29

can make use of a 6-phase clocking scheme as shown in Fig. 6.6. A representative 6-phase clocking scheme is as shown in Fig. 6.7. It has four hold phases.

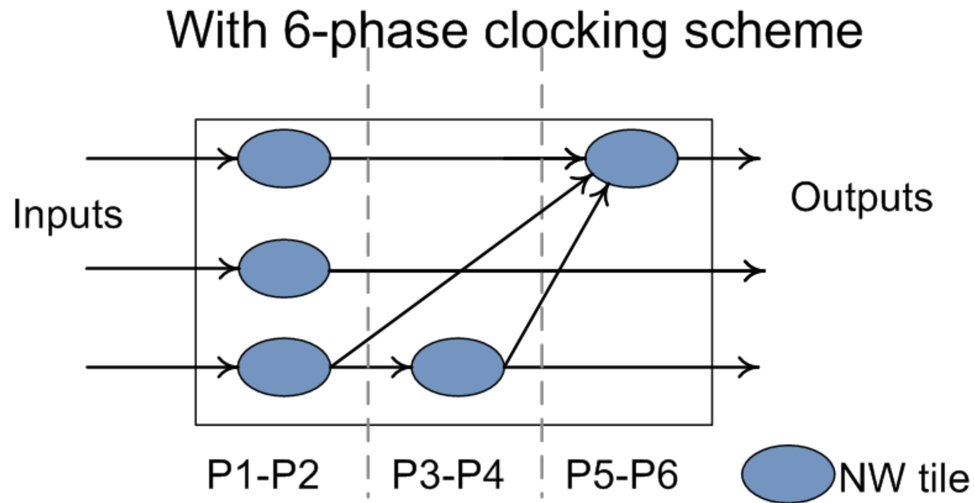

Figure 6.6. Functional unit with 6-phase clocking scheme

The Table. 6.3 shows the area, power and performance for (7,3) counter design with 4-phase and 6-phase clocking scheme. Fig. 6.8 shows the (7,3) counter with identity tiles.

The Table. 6.4 below provides the results for a 3bit adder with different clocking schemes. The 4-phase clocking scheme achieves better performance but has area penalty but the 6-phase clocking scheme has lower throughput at lesser area and power. The 4-phase clocking scheme has better throughput as every stage evaluates once in 4 cycles, when compared to 6 cycles in a 6-phase clocking scheme. The 4-

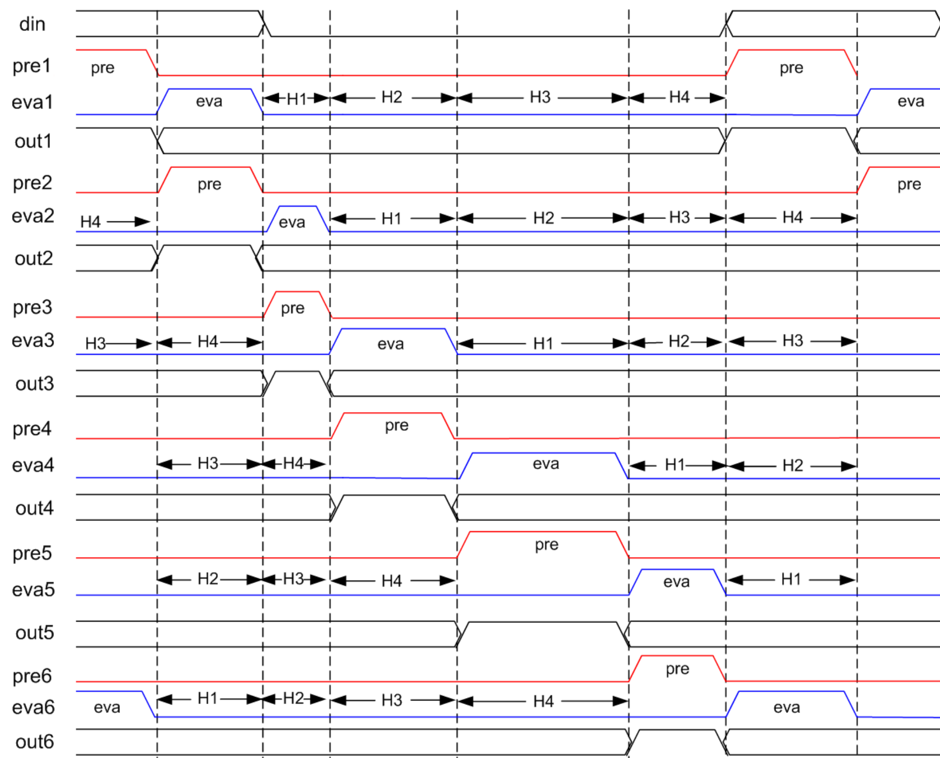

Figure 6.7. 6-phase clocking scheme

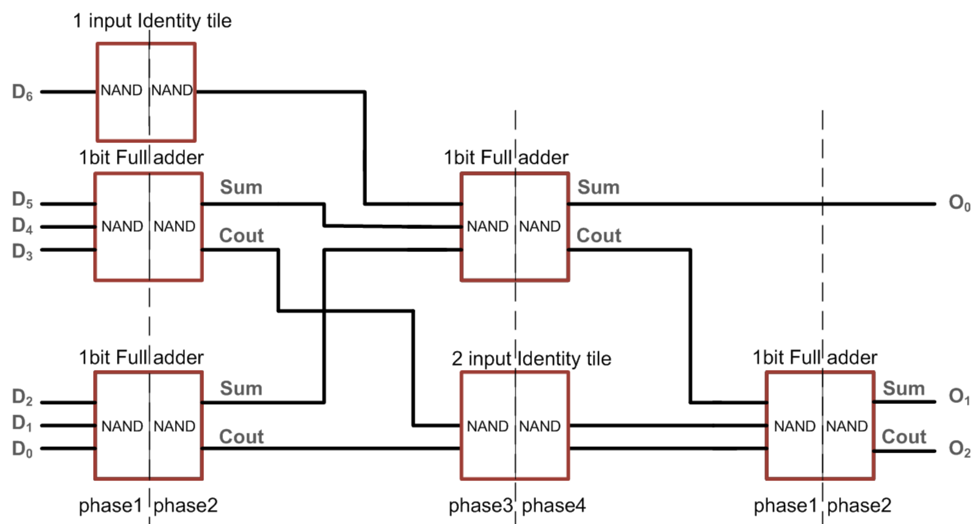

Figure 6.8. (7,3) counter with 4-phase clocking showing the identity tile

Table 6.4. Comparison of 4-phase and 6-phase clocking schemes for 3bit addder

	4-phase clocking	6-phase clocking
Number of transistors	240	192
Max Delay (ps)	91.36	141.54
Average power (μW)	4.43	3.23
Area(μ^2)	1.21	0.97

phase clocking scheme consumes more power as additional identity tiles are required to balance the paths (shown in Fig. 6.8). These identity tiles have an area penalty. In a 6-phase clocking scheme, fewer identity tiles are required to balance the path and hence it is more area and power efficient.

6.4 Chapter Summary

In this chapter, we showed partitioning and clocking schemes that can be adopted in order to overcome the limitations of the two-level logic schemes. The partitioning has been proposed with careful consideration to the fabric vision presented earlier. With increased complexity of the function being implemented, partitioning yields better results when compared to the regular two-level logic approach. Partitioning algorithms developed for PLAs can be adopted in order to divide a large tile into smaller blocks. Since the use of CMOS interconnects enables arbitrary routing, the smaller tiles can be easily routed. Further, we can modify the clocking schemes to suit the requirements (Area, power and performance). With the help of partitioning and intelligent clocking schemes, we can realize complex logic functions without significant penalty.

CHAPTER 7

IMPACT OF MASK OVERLAY

7.1 Introduction

As shown in earlier sections, the N^3ASICs fabric vision was developed with careful consideration to the order of manufacturing process. A single unconventional step is carried out *a priori*, without any overlay or registration requirement. All subsequent steps make use of conventional photolithography, which has excellent overlay precision ($3\sigma = \pm 3.3$nm). Therefore, while many lithographic masks will be employed for manufacturing N^3ASICs, the overlay-limited yield is expected to be high. This section investigates the impact of mask overlay imprecision on N^3ASICs yield. Specifically, we address the following questions: (i) How much overlay precision is necessary between process steps? (ii) What is the impact on yield if different overlays are used?

To study the impact of mask overlay a methodology was previously developed for a 2D-grid based NASIC fabric [32]. Overlay misalignment between successive masks were modeled as Gaussian random variables and Monte Carlo simulations were carried out in a custom simulator to determine the number of functioning chips. The same methodology was adopted for the N^3ASIC fabric.

7.2 Alignment and Mask overlay

Nanowire patterning may be carried out using NIL [18] or SNAP [10]. This step does not have any overlay requirement since it is carried out *a priori* to any lithographic step. In addition, self-aligned alignment markers can be patterned on the substrate at the same time as the logic nanowires. These alignment markers can

be used by subsequent lithographic steps for registering nanowire positions. If NIL is used, alignment markers and logic nanowires can be part of the same imprint mold. This can be transferred to the substrate in a self-aligned fashion. In the case of SNAP, where an arbitrary alignment marker may be difficult to achieve, patterned nanowires of different dimensions can be used as Moire patterns/fringes [39] as shown in Fig. 7.1

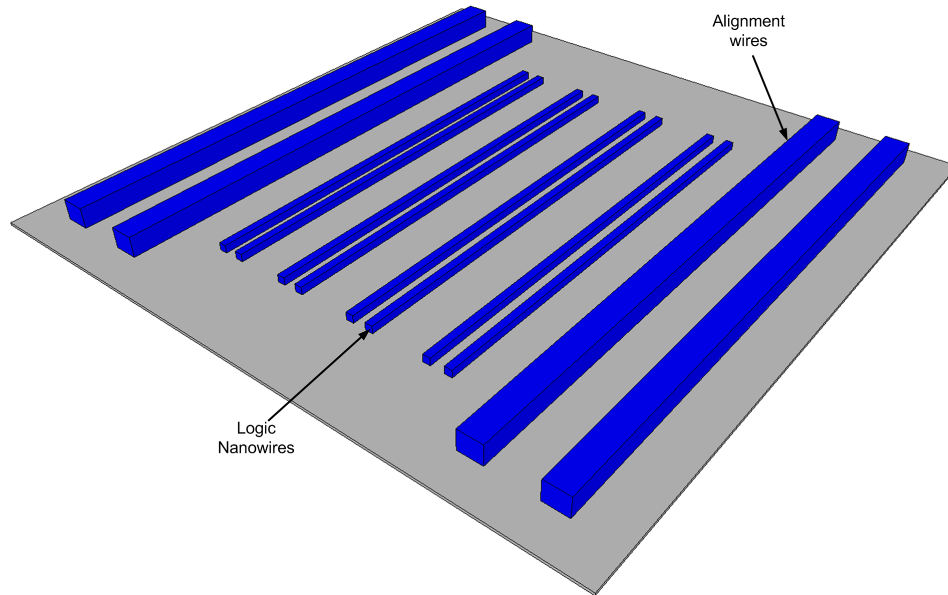

Figure 7.1. Patterned nanowires (larger than logic nanowires) could be used as Moire patterns for alignment

Since the underlying pattern of nanowires is uniform, this allows the first lithographic mask to be horizontally offset with some tolerance and still achieve correct functionality. Fig. 7.2 depicts the mask registration process during contact creation step. Fig. 7.2(a) shows the nanowires and the alignment markers created using the initial patterning technique (e.g. NIL). Fig. 7.2(b) shows the desired alignment scenario for the first lithographic step. Alignment marker (AM# 1) 1 is used as the alignment target and the litho-mask is perfectly aligned in this case. New alignment markers (AM# 2) created during this step, may be used as the alignment target for

the subsequent mask. Fig. 7.2(c) shows an excessive misalignment case which results in nanowires being not contacted by the power rails resulting in a defective chip.

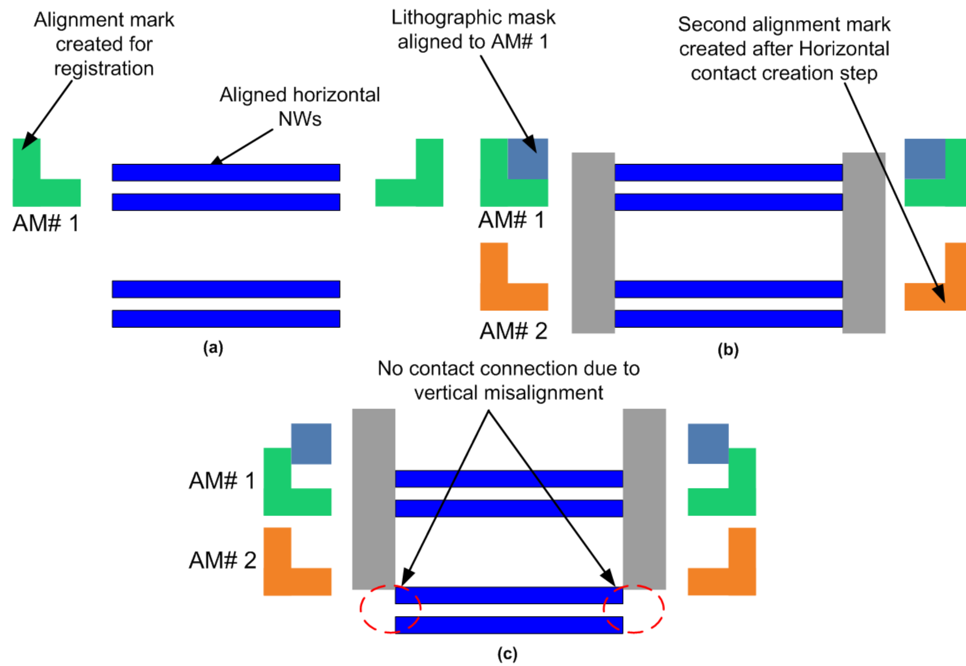

Figure 7.2. Depiction of mask registration and alignment markers during contact creation step

Fig. 7.3 depicts the impact of mask misalignment during functionalization to create metal gates and 2C-xnwFETs [24]. An incorrectly shorted device can be formed due to large vertical misalignment, impacting the yield. Also, this step has little tolerance to horizontal misalignment as contacts have already been defined. Fig. 7.3 shows correctly functionalized devices despite some overlay misalignment demonstrating the misalignment tolerance in this step. Fig. 7.3(c) shows shorted devices due to excessive overlay misalignment. During this step additional alignment markers (not shown in Fig. 7.3) will be created which will be the alignment targets for the subsequent step.

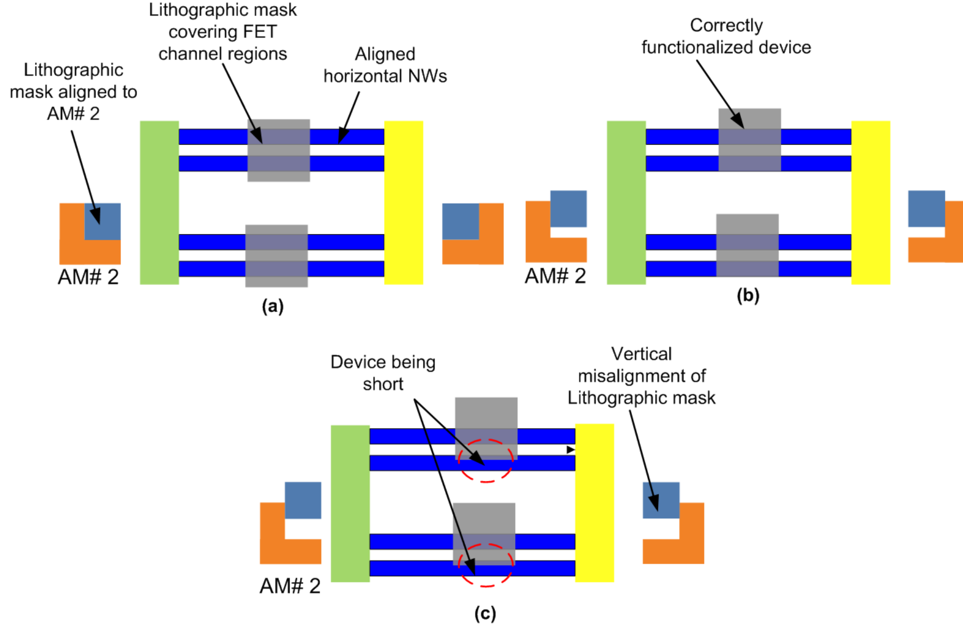

Figure 7.3. Mask registration during functionalization step

7.3 Mask Overlay simulation

The manufacturing of 3D integrated fabric employs lithographic masks. The contact creation and metal gate deposition steps involve alignment to the smallest features, and hence they are most critical to mask overlay and contribute significantly to the yield loss. Yield loss due to mask overlay during metal stack creation is minimal (identical to conventional CMOS). Hence metal stacks higher than M2 layer have not been considered in these simulations. The WISP-0 [36] nanoscale processor design was mapped onto the N^3ASIC fabric. Several 3σ overlay misalignment values projected by ITRS 2009 [1] were used to carry out the simulations.

The results in Fig. 7.4 show that close to 99% mask overlay limited yield may be obtained for $3\sigma = \pm 9$nm overlay (manufacturing solutions known as per ITRS 2009) when constructing a uniform nanowire bundle with λ=8nm (16nm technology node) in the 3D integrated fabric. Within a bundle the width of nanowires is 5nm each, with 6nm spacing to accommodate 16nm vias. Fig. 7.4 shows that even with a pessimistic

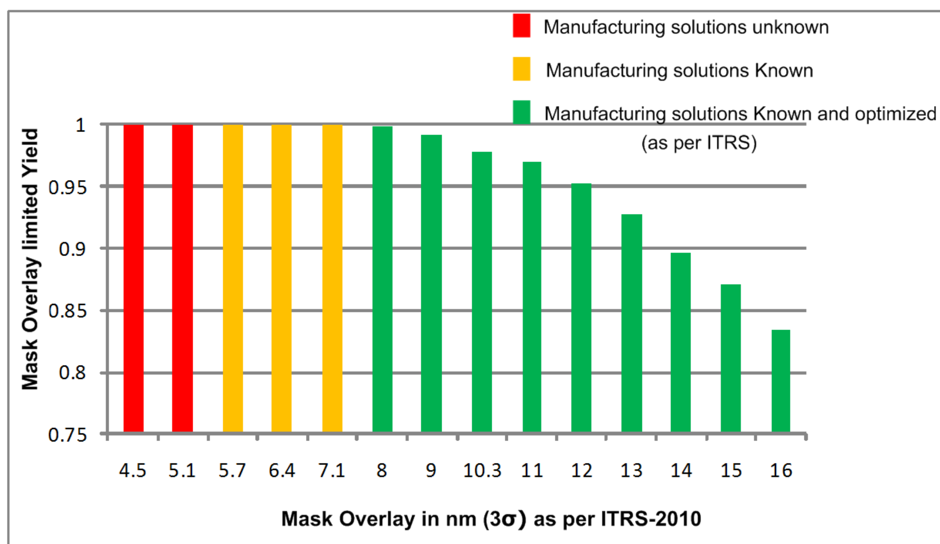

Figure 7.4. Mask overlay limited Yield vs. Overlay for 3D integrated fabric

mask overlay projection of 3σ=±16nm a mask overlay limited yield of 83% can be observed. These overlay requirements are far less stringent than the requirement for 16nm CMOS (3σ=±3.3nm for 16nm CMOS, per ITRS 2009).

It is evident from the results that the use of regular structure (like the nanowire arrays in N^{3}ASICs) does not impose stringent constraints on overlay precision requirement. Further, fewer masks are required to manufacture this fabric compared to a CMOS design which is beneficial from both yield and cost perspective.

The simulation methodology employed enables addressing key overlay and registration requirements. It is possible to estimate the overlay-limited yield for a range of overlay projections. It is also possible to address sensitivity of the overlay-limited yield to key fabric parameters such as the width and pitch of nanowires.

7.4 Chapter Summary

We have shown that by analyzing the available design choices and careful consideration of the order of manufacturing processes, the impact of mask overlay can

be alleviated. The N^3ASIC 3-D nanofabric, built using these principles, is realizable with available manufacturing techniques at very minimal yield loss. Assuming an overlay precision of 9nm or better results in a mask overlay limited yield of 100%. In contrast, irregular structures would have more stringent mask overlay requirements. For example, the proposed approach also has considerably greater tolerance (3X) to overlay imprecision than 16nm CMOS that requires a 3.3nm precision at 16nm node as per ITRS 2009.

CHAPTER 8
CONCLUSION

A 3-D integrated nano-CMOS hybrid fabric N^3ASICs was presented. A physical fabric vision was developed to enable the self-assembly/unconventional manufacturing approach and conventional photolithography, to be employed in conjunction while retaining the benefits of both the approaches. To facilitate the use of photolithography CMOS design rules were followed at all levels. No special manufacturing constraints were introduced. A detailed layer-by-layer assembly sequence of the fabric was presented. Fabric evaluations were carried out at device, circuit and system levels. A nanoprocessor implemented using the proposed N^3ASIC fabric was shown to be 3X denser than equivalent CMOS design and 5X power efficient for a comparable performance. Systematic yield implications due to mask overlay misalignment were analyzed. Results show that a yield of 83% was obtained even for a pessimistic overlay misalignment of $3\sigma = \pm 16$nm. An approach to scale the design in order to realize complex logic functions was presented.

BIBLIOGRAPHY

[1] 2009 ITRS http://www.itrs.net/Links/2009ITRS/Home2009.html.

[2] Predictive technology model (PTM). http://ptm.asu.edu/.

[3] HSPICE simulation and analysis guide, 2009.

[4] Synopsys- design compiler user guide, 2009.

[5] Synopsys- sentaurus user guide, 2009.

[6] Augustine, C., Behin-Aein, B., Fong, Xuanyao, and Roy, K. A design methodology and device/circuit/architecture compatible simulation framework for low-power magnetic quantum cellular automata systems. In *Design Automation Conference, 2009. ASP-DAC 2009. Asia and South Pacific* (2009), pp. 847–852.

[7] Bencher, Christopher, Dai, Huixiong, and Chen, Yongmei. Gridded design rule scaling: taking the CPU toward the 16nm node. In *Proceedings of SPIE* (San Jose, CA, USA, 2009), pp. 72740G–72740G–10.

[8] Gottlob, Heinrich D. B, Mollenhauer, Thomas, Wahlbrink, Thorsten, Schmidt, Mathias, Echtermeyer, Tim, Efavi, Johnson K, Lemme, Max C, and Kurz, Heinrich. Scalable gate first process for silicon on insulator metal oxide semiconductor field effect transistors with epitaxial high-k dielectrics. *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures 24*, 2 (Mar. 2006), 710–714.

[9] Guha, Supratik, and Narayanan, Vijay. High-/Metal gate science and technology. http://www.annualreviews.org/doi/abs/10.1146/annurev-matsci-082908-145320, July 2009.

[10] Heath, James R. Superlattice nanowire pattern transfer (SNAP). *Accounts of Chemical Research 41*, 12 (Dec. 2008), 1609–1617.

[11] Law, Matt, Goldberger, Joshua, and Yang, Peidong. SEMICONDUCTOR NANOWIRES AND NANOTUBES. *Annual Review of Materials Research 34*, 1 (Aug. 2004), 83–122.

[12] Likharev, Konstantin K. CMOL: second life for silicon? *Microelectronics Journal 39* (Feb. 2008), 177183. ACM ID: 1342758.

[13] Lu, Wei, and Lieber, Charles M. Semiconductor nanowires. *Journal of Physics D: Applied Physics 39*, 21 (Nov. 2006), R387–R406.

[14] Martel, R., Derycke, V., Appenzeller, J., Wind, S., and Avouris, Ph. Carbon nanotube field-effect transistors and logic circuits. In *Proceedings of the 39th conference on Design automation - DAC '02* (New Orleans, Louisiana, USA, 2002), p. 94.

[15] Mistry, K., Allen, C., Auth, C., Beattie, B., Bergstrom, D., Bost, M., Brazier, M., Buehler, M., Cappellani, A., Chau, R., Choi, C. -H, Ding, G., Fischer, K., Ghani, T., Grover, R., Han, W., Hanken, D., Hattendorf, M., He, J., Hicks, J., Huessner, R., Ingerly, D., Jain, P., James, R., Jong, L., Joshi, S., Kenyon, C., Kuhn, K., Lee, K., Liu, H., Maiz, J., Mclntyre, B., Moon, P., Neirynck, J., Pae, S., Parker, C., Parsons, D., Prasad, C., Pipes, L., Prince, M., Ranade, P., Reynolds, T., Sandford, J., Shifren, L., Sebastian, J., Seiple, J., Simon, D., Sivakumar, S., Smith, P., Thomas, C., Troeger, T., Vandervoorn, P., Williams, S., and Zawadzki, K. A 45nm logic technology with High-k+Metal gate transistors, strained silicon, 9 cu interconnect layers, 193nm dry patterning, and 100% pb-free packaging. In *Electron Devices Meeting, 2007. IEDM 2007. IEEE International* (Dec. 2007), IEEE, pp. 247–250.

[16] Moritz, Csaba Andras, Narayanan, Pritish, and Chui, Chi On. Nanoscale Application-Specific integrated circuits. In *Nanoelectronic Circuit Design*, Niraj K Jha and Deming Chen, Eds. Springer New York, 2011, pp. 215–275.

[17] Moritz, Csaba Andras, Wang, Teng, Narayanan, Pritish, Leuchtenburg, Michael, Guo, Yao, Dezan, Catherine, and Bennaser, Mahmoud. Fault-Tolerant nanoscale processors on semiconductor nanowire grids. *IEEE Transactions on Circuits and Systems I: Regular Papers 54*, 11 (Nov. 2007), 2422–2437.

[18] Mrtensson, Thomas, Carlberg, Patrick, Borgstrm, Magnus, Montelius, Lars, Seifert, Werner, and Samuelson, Lars. Nanowire arrays defined by nanoimprint lithography. *Nano Letters 4*, 4 (Apr. 2004), 699–702.

[19] Narayanan, P., Park, Kyoung Won, Chui, Chi On, and Moritz, C. A. Manufacturing pathway and associated challenges for nanoscale computational systems. In *Nanotechnology, 2009. IEEE-NANO 2009. 9th IEEE Conference on* (July 2009), pp. 119 –122.

[20] Narayanan, Pritish, Kina, Jorge, Panchapakeshan, Pavan, Chui, Chi On, and Moritz, Csaba Andras. Integrated Device-Fabric explorations and noise mitigation in nanoscale fabrics. *Submitted to TNANO under review*.

[21] Narayanan, Pritish, Leuchtenburg, Michael, Wang, Teng, and Moritz, Csaba Andras. CMOS control enabled Single-Type FET NASIC. In *2008 IEEE Computer Society Annual Symposium on VLSI* (Montpellier, France, 2008), pp. 191–196.

[22] Narayanan, Pritish, Moritz, Csaba Andras, Park, Kyoung Won, and Chui, Chi On. Validating cascading of crossbar circuits with an integrated device-circuit exploration. In *2009 IEEE/ACM International Symposium on Nanoscale Architectures* (San Francisco, CA, USA, July 2009), pp. 37–42.

[23] Pacha, C., von Arnim, K., Schulz, T., Xiong, Weize, Gostkowski, M., Knoblinger, G., Marshall, A., Nirschl, T., Berthold, J., Russ, C., Gossner, H., Duvvury, C., Patruno, P., Cleavelin, R., and Schruefer, K. Circuit design issues in multi-gate FET CMOS technologies. In *2006 IEEE International Solid State Circuits Conference - Digest of Technical Papers* (San Francisco, CA, 2006), pp. 1656–1665.

[24] Panchapakeshan, Pavan, Narayanan, Pritish, and Moritz, Csaba Andras. N3ASICs: designing nanofabrics with fine-grained CMOS integration. In *2011 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)* (June 2011), IEEE, pp. 196–202.

[25] Picciotto, Carl, Gao, Jun, Yu, Zhaoning, and Wu, Wei. Alignment for imprint lithography using nDSE and shallow molds. *Nanotechnology 20*, 25 (June 2009), 255304.

[26] Roy, S., and Narayanan, H. A new approach to the problem of PLA partitioning using the theory of the principal lattice of partitions of a submodular function. In *ASIC Conference and Exhibit, 1991. Proceedings., Fourth Annual IEEE International* (Sept. 1991), IEEE, pp. P2–4/1–4.

[27] Samavedam, S. B, Tseng, H. H, Tobin, P. J, Mogab, J., Dakshina-Murthy, S., La, L. B, Smith, J., Schaeffer, J., Zavala, M., Martin, R., Nguyen, B. -Y, Hebert, L., Adetutu, O., Dhandapani, V., Luo, T. -Y, Garcia, R., Abramowitz, P., Moosa, M., Gilmer, D. C, Hobbs, C., Taylor, W. J, Grant, J. M, Hegde, R., Bagchi, S., Luckowski, E., Arunachalam, V., and Azrak, M. Metal gate MOSFETs with HfO2 gate dielectric. In *2002 Symposium on VLSI Technology, 2002. Digest of Technical Papers* (2002), IEEE, pp. 24– 25.

[28] Shabadi, Prasad, Khitun, Alexander, Narayanan, Pritish, Bao, Mingqiang, Koren, Israel, Wang, Kang L., and Moritz, C. Andras. Towards logic functions as the device. In *2010 IEEE/ACM International Symposium on Nanoscale Architectures* (Anaheim, CA, USA, June 2010), pp. 11–16.

[29] Snider, Gregory S, and Williams, R Stanley. Nano/CMOS architectures using a field-programmable nanowire interconnect. *Nanotechnology 18*, 3 (Jan. 2007), 035204.

[30] Streetman, and Banerjee. *Solid state electronic devices*, 6th ed. ed. Prentice-Hall, Englewood Cliffs, NJ, 2010.

[31] Vijayakumar, P., Narayanan, P., Koren, I., Krishna, C. M, and Moritz, C. A. Incorporating heterogeneous redundancy in a nanoprocessor for improved yield

and performance. In *2010 IEEE 25th International Symposium on Defect and Fault Tolerance in VLSI Systems (DFT)* (Oct. 2010), IEEE, pp. 273–279.

[32] Vijayakumar, P., Narayanan, P., Koren, I., Krishna, C. M, and Moritz, C. A. Impact of nanomanufacturing flow on systematic yield losses in nanoscale fabrics. In *2011 IEEE/ACM International Symposium on Nanoscale Architectures* (2011).

[33] Wang, Dunwei, Bunimovich, Yuri, Boukai, Akram, and Heath, James R. Two-dimensional single-crystal nanowire arrays. http://www.nanoarchive.org/1853/, Dec. 2007.

[34] Wang, Dunwei, Sheriff, Bonnie A., McAlpine, Michael, and Heath, James R. Development of ultra-high density silicon nanowire arrays for electronics applications. *Nano Research 1*, 1 (July 2008), 9–21.

[35] Wang, Teng. Fault tolerant nanoscale microprocessor design on semiconductor nanowire grids. *Open Access Dissertations 86* (2009).

[36] Wang, Teng, Ben-naser, Mahmoud, Guo, Yao, and Moritz, Csaba Andras. Wire-streaming processors on 2-D nanowire fabrics. *NANOTECH 2005, NANO SCIENCE AND TECHNOLOGY INSTITUTE* (2005).

[37] Wang, Teng, Narayanan, P., and Andras Moritz, C. Heterogeneous Two-Level logic and its density and fault tolerance implications in nanoscale fabrics. *IEEE Transactions on Nanotechnology 8*, 1 (Jan. 2009), 22–30.

[38] Wang, Teng, Narayanan, Pritish, and Moritz, Csaba Andras. Combining 2-level logic families in grid-based nanoscale fabrics. In *2007 IEEE International Symposium on Nanoscale Architectures* (San Jose, CA, USA, Oct. 2007), pp. 101–108.

[39] Zaidi, Saleem H. Moire interferometric alignment and overlay techniques. In *Proceedings of SPIE* (San Jose, CA, USA, 1994), pp. 371–382.

[40] Zhao, Wei, and Cao, Yu. New generation of predictive technology model for sub-45nm design exploration. In *Proceedings of the 7th International Symposium on Quality Electronic Design* (2006), IEEE Computer Society, pp. 585–590.