

Circuit Design for FPGAs in Sub-Threshold Ultra-Low Power Systems

A Thesis

Presented to
the faculty of the School of Engineering and Applied Science
University of Virginia

in partial fulfillment
of the requirements for the degree

Master of Science

by

Yu Huang

August

2015

APPROVAL SHEET

The thesis
is submitted in partial fulfillment of the requirements
for the degree of
Master of Science



AUTHOR

The thesis has been read and approved by the examining committee:

Benton H. Calhoun

Advisor

Joanne Bechta Dugan

John Stankovic

Accepted for the School of Engineering and Applied Science:



Dean, School of Engineering and Applied Science

August
2015

Acknowledgement

I would like to express my gratitude to my advisor, Professor Benton H. Calhoun for his useful comments, remarks, and engagement through the learning process of my Master's thesis. Without his support and encouragement throughout my academic work at the University of Virginia, this work would not have been completed. I would also like to thank Professor Joanne Bechta Dugan and Professor John Stankovic for giving me useful suggestions whenever I needed them. Furthermore, I want to thank Aatmesh Shrivastava, He Qi, and Oluseyi Ayorinde, who willingly shared their precious time and given me their assistance throughout our collaboration. And also, I want to thank everyone in the Robust Low Power VLSI group as well as my friends here in UVa who have helped me and spent so many happy times in work and life with me: Yousef Shaksheer, Yanqing Zhang, Kyle Craig, Peter Beshay, Ke Wang, Jiaqi Gong, James Boely, Alicia Klinefelter, Patricia Gonzalez, Arijit Banerjee, Divya Akella, Abhishek Roy, Chris Lukas, Farah Yahya, Hash Patel, Ningxi Liu, Manula Pathirana and Dilip Vasudevan. Last but not least, I owe more thanks to my parents, my boyfriend Kevin, and his family. Without their unconditional love and support, it would not have been possible for me to finish my degree.

Abstract

Field Programmable Gate Array (FPGA) is the most promising type of programmable devices in the era of ubiquitous computing. Limited by the design cost, energy consumption, portability constraints, and flexibility demands, FPGAs compensate the gap between Application Specific Integrated Circuits (ASICs) and General Purpose Processors (GPPs). The vision for ubiquitous computing also requires us to deploy a large number of very small form-factor, long-lasting electronic systems with highly constrained energy consumption. Thus, a sub-threshold FPGA will provide energy efficient digital circuits for a variety of ultra low power ubiquitous systems at low unit cost and enable a shift in how computing and communication platforms are designed.

However, studies show that 60% - 70% of power is dissipated in the FPGA interconnect fabrics. Additionally, interconnect dominates delay and area in modern FPGAs. Driven by the goal of energy efficiency, we propose an optimization technique in sub-threshold FPGA design which focuses on the FPGA interconnect. According to a typical FPGA interconnect structure, this optimization work explores the switch boxes, connection boxes, drivers, sense amplifiers, and the signal degradation along the interconnect path to study the need for inserting repeaters to remain the functionality in sub-threshold. With the concern of energy and delay, we used energy delay product (EDP) as our metric. We fabricated a chip and both simulation and measurement results are presented in a 130nm CMOS technology.

In the modern IC area, voltage scaling is an effective and common method used in energy reduction. The special structure of FPGA interconnect, which is driven by a driver at the beginning of each path (e.g., output of a basic logic element), makes further energy reduction possible by applying a voltage scaling technique. We propose a programmable header structure to implement the voltage scaling and studied on the characteristics of typical FPGA applications by mapping MCNC benchmarks. We found that voltage scaling reduces energy consumption by an average 68.6%. This provides a very promising direction in FPGA interconnect architecture design.

Different voltage domains are very common in modern IC design. In such systems, especially ultra low power SoCs, a level converter is an essential component to shift signals between low and high voltage domains. In an energy harvesting system, which operates depending on the energy stored in an energy harvesting capacitor, the shifting capability of level converters implicates the capacity of the energy in the

capacitor being used by the system. In a system heavily constrained by energy consumption, an ultra low swing level converter is integral to lower down the system threshold voltage. We propose a 145mV (from measurement) single end level converter which can both be used both in a FPGA circuit and a low voltage IC. This work introduces the design concept of inserting a sub-threshold charge pump to further extend the shifting ability. We also fabricated a chip using 130nm CMOS technology and present both the simulation and measurement results.

Contents

1	Introduction	7
1.1	Contributions of this thesis	8
1.2	Outline of the thesis	8
2	Optimization of Energy Efficient Low-Swing Interconnect for Subthreshold FPGAs	9
2.1	Introduction	9
2.2	Circuit model of the global interconnect	11
2.2.1	Low-Swing Interconnect	11
2.2.2	Interconnect Path Distribution Exploration	12
2.2.3	Custom Interconnect Model	14
2.3	Interconnect Circuit Optimization	14
2.3.1	Optimal Voltage of the Dual- V_{DD} Scheme	14
2.3.2	Signal Degradation	15
2.3.3	Repeater Number Optimization	17
2.3.4	Connection Box (CB) Topology Optimization	17
2.3.5	Switch and Driver Size Optimization	18
2.4	Comparison of Designs	21
2.5	Test Chip and Measurement Results	22
2.6	Conclusion	26
3	Voltage Scaling on FPGA Interconnects	27
3.1	Introduction	27
3.2	Background	28
3.2.1	Conventional Island Style FPGA Interconnect	28
3.2.2	Subthreshold FPGA Interconnect	30
3.3	Motivation	32
3.4	Voltage scaling technique for subthreshold interconnect	36
3.4.1	Performance and energy exploration	36
3.4.2	Header-based voltage programmability	37

3.5	Simulations	38
3.6	Conclusion	40
4	A single ended level converter circuit design for ultra low power low voltage ICs	43
4.1	Introduction	43
4.2	Sub-threshold charge pump	45
4.3	Implementation of the level converter	45
4.4	Measurement Results	48
4.5	Conclusion	50
5	Conclusion and future work	51
5.1	Summary	51
5.2	Contributions	52
5.3	Future work	54
	References	55
	Publications	57

1 Introduction

The increased importance of power is more notable in recent years for energy-constrained systems. This type of application requires the operation in the sub-threshold region to reduce energy consumption. At the same time, massive amounts of information, increased control, and awareness of the ambient environment has led technology to ubiquitous computing, where sensors and other integrated circuits play an important role. In a typical ubiquitous computing sensor system, a large number of sensors work simultaneously in different environments, most of which are portable and wearable devices. However, this type of application presents challenges such as reducing energy consumption and maintaining flexibility. To address these constraints, the reconfigurability of Field Programmable Gate Arrays (FPGAs) helps compensate the gap between Application Specific Integrated Circuits (ASICs) and General Purpose Processors (GPPs).

Industrial companies like Microsemi and Lattice Semiconductor have their own low power FPGA products (IGLOO nano FPGA Fabric, iCE40 Ultra Family). But those devices still consume tens of milliwatts in active mode, which is high for the UbiComp requirements. Specifically, for ultra low power systems in UbiComputing, low-power sub-threshold FPGA design focuses both on energy savings and flexibility. Customized FPGAs are necessary to fit the requirements. On the other hand, in a FPGA chip, the interconnect dominates most of the energy and delay consumption, so it is important to study on how to optimize the interconnect design of FPGAs. Unfortunately, it is impossible to test the interconnect structure or any other parameters through commercial FPGAs. Commercial FPGA companies, like Xilinx and Altera, have their own packaged FPGA products which allow users to load their own verilog/VHDL code to implement the functions, but the circuit-level design is out of the user's reach. Thus, customized FPGAs are necessary to conduct the research on FPGA interconnect. Interconnect optimization is the first and important step of designing a customized FPGA.

This thesis focuses on the optimization of the interconnect with a specific interest on sub-threshold customized FPGAs. Further, we study the voltage scaling potentials on FPGA interconnects to further save energy. We also propose a subthreshold ultra low swing level converter which can be used in both a voltage scaling design and other ULP SoCs.

1.1 Contributions of this thesis

In this thesis, we optimize sub-threshold FPGA interconnect design, study on the potential energy saving in FPGA interconnects by scaling voltages, and proposed new ideas of designing an ultra-low swing single ended level converter. We discuss results of this exploration and suggest the optimal design parameters for a sub-threshold FPGA. We further investigate the voltage scaling techniques to further reduce the energy consumption on FPGA interconnects. Finally, we introduce a design of level converters based on sub-threshold charge pumps. For all the work, we fabricated test chips with a 130 nm CMOS technology.

1.2 Outline of the thesis

In chapter 2, we introduce the optimization work on low-power FPGA interconnects. This chapter includes the optimization of switch boxes, drivers, connection boxes and a study of the signal degradation.

In chapter 3, we propose a dual-VDD voltage scaling technique to further reduce the energy consumption of FPGA interconnects. This chapter applies this idea onto the MCNC benchmarks and conducted transistor-level simulations.

Chapter 4 proposes an ultra low swing level converter design which can be applied in a low voltage ICs to implement the communications between blocks and further take use of the energy in an energy harvesting system.

Chapter 5 concludes the work discussed and summarizes the contribution of the work.

2 Optimization of Energy Efficient Low-Swing Interconnect for Subthreshold FPGAs

¹ FPGA interconnect traditionally dominates energy and delay, and designs such as low-swing interconnect have been proven to reduce the interconnect burden for low energy FPGAs. We present an optimized low-swing dual-VDD interconnect for FPGAs operating in the sub-threshold region. We optimize the topology of switch boxes and connection boxes, transistor sizes, and the value of supply voltages to reduce energy and to improve energy efficiency. We also address signal degradation along lengthy interconnect paths and examine strategies for inserting low-switching-threshold repeaters. A 130nm test chip implementing low-swing dual-VDD interconnect meshes with different circuit parameters is measured. The results show that optimization of the low-swing interconnect provides up to 60.2% lower energy-delay-product (EDP) than a straightforward, unoptimized low-swing design. Furthermore, the simulation results show that the optimized low-swing interconnect is 97.7% faster and 42.7% lower energy than a traditional unidirectional interconnect.

2.1 Introduction

Existing hardware solutions for ubiquitous computing include ultra-low-power (ULP) ASICs and ULP microprocessors working in sub-threshold region. However, the development of ULP ASICs for these applications is costly and time-consuming due to high design complexity. On the other hand, ULP microprocessors consume too much power. Sub-threshold FPGAs, which are flexible and consume a reasonable amount of power, have become a highly desirable solution. However, an FPGA design implementation consumes 7X - 14X more power than a functionally equivalent ASIC design [16], so power reduction of FPGAs is critical for applying them to ULP applications. The global interconnect is the major power consumer in FPGAs. Studies have shown that 60%-70% of power is dissipated in the interconnection fabric [20, 24, 27]. In addition, interconnect also dominates delay and area in modern FPGAs. Researchers reduce power of the FPGA interconnect in different ways. In [2], a new FPGA routing switch design that is programmable to operate in three different modes was introduced. In low-power mode, leakage power was reduced by up to 52% and active power was reduced by up to 31% comparing to in high-speed mode. In [9] and [21],

¹This chapter is mainly from publication [2].

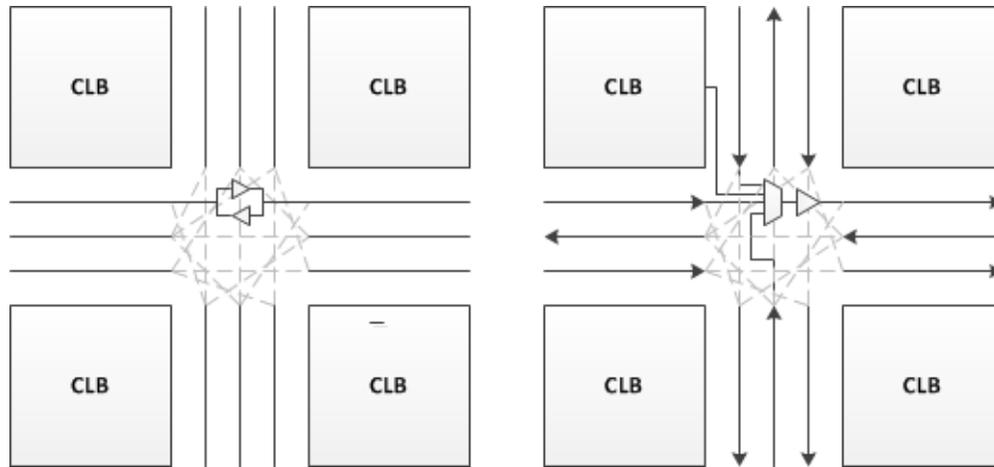


Figure 1: (a) Bi-directional switch box (b) uni-directional switch box

researchers applied a dual-VDD scheme in the routing blocks and saved up to 61% of power. Researchers in [25] and [7] exploited dual-VT scheme, which allowed mixed usage of low and high threshold transistors in routing switches in order to reduce leakage current. These works reduced routing power effectively, but ubiquitous computing applications have strict requirements on both speed and power that make energy and energy-delay-product (EDP) reduction of FPGA routing fabrics a driving challenge.

The routing fabric in FPGAs is defined as the electrical connectivity hardware between complex logic blocks (CLBs). It is comprised of connection boxes (CBs) that connect CLBs to the routing channel, switch boxes (SBs) that form the connectivity of routing paths, and wire segments. The traditional bi-directional and uni-directional SBs are shown in Figure 1 (a) and (b) respectively. Each bi-directional routing switch is comprised of 2 tri-state buffers, while each uni-directional switch is comprised of an N-input multiplexer followed by a buffer, where N represents the number of tracks that can connect to the track that this switch drives [12, 18, 19].

The traditional routing fabric is not energy efficient. The large number of buffers and multiplexers results in a highly capacitive routing channel and uses full swing signaling, which both contribute to the active energy. In [26], researchers reduced both delay and energy by implementing a new low-swing interconnect fabric operating in sub-threshold, where the supply voltage VDD is less than the threshold voltage VT of a single transistor. They used a pass-gate (PG) based design to replace the multiplexers and buffers in the routing switches. Both the capacitance and signal swing are then reduced. Drivers and sense amps (SAs) are located at the outputs and inputs of CLBs to form the two ends of each routing path. In addition, a low

switching threshold (VM) SA was introduced in their work to reduce delay and variation. Dual-VDD was also applied by using a higher VDD in the config bits to drive the PG gate terminals, reducing delay while only incurring a slight leakage penalty in the high VT configuration bits. The low-swing design made a big step towards energy reduction, however, the circuit level implementation can be greatly optimized for further reduction.

In this work, we study the influence of the main supply voltage (VDD) and the boosted voltage (VDDC) on EDP and energy. In addition, we compare the topology and size of CBs, routing switches, and drivers in terms of EDP and energy. We also examine the influence of inserting low-VM repeaters into routing paths. A test chip was fabricated to compare different circuits for the low-swing design. The measured data shows the best circuit options are 61.7% faster and 60.2% lower in EDP than a first-pass, unoptimized design at 0.4V for a 40-switch path.

In Section II, we introduce our low-swing global interconnect model based on path distribution. The circuit optimization details including design space exploration and low-VM repeater insertion are discussed in Section III, followed by the simulation results comparisons of traditional uni-directional interconnect and our optimized low-swing design. Finally, the measurement results are shown in Section V.

2.2 Circuit model of the global interconnect

2.2.1 Low-Swing Interconnect

Traditional FPGA interconnect uses multiplexers and buffers to implement routing switches to achieve high speed, but it suffers from high energy cost. Reducing supply voltage for conventional interconnect circuits to the sub-threshold region helps to solve the energy problem. However, since driver and buffer current decreases exponentially in sub-threshold, delay is increased exponentially as well. Upsizing drivers and buffers does not help, since speed depends linearly on device size but exponentially on VDD in sub-threshold. The low-swing interconnect design in [13] [26] replaces the multiplexers and buffers structure with PGs. Its basic structure is shown Figure 2. This new topology eliminates the energy consumed by buffers. Also, the signal swing along the interconnect paths is reduced due to the transfer characteristics of the sub-threshold PGs, and this lower swing further decreases energy consumption. Since active energy equals $C \times V_{DD} \times \delta V$, where C denotes the total lumped capacitance along the path and δV is the signal swing, reducing signal

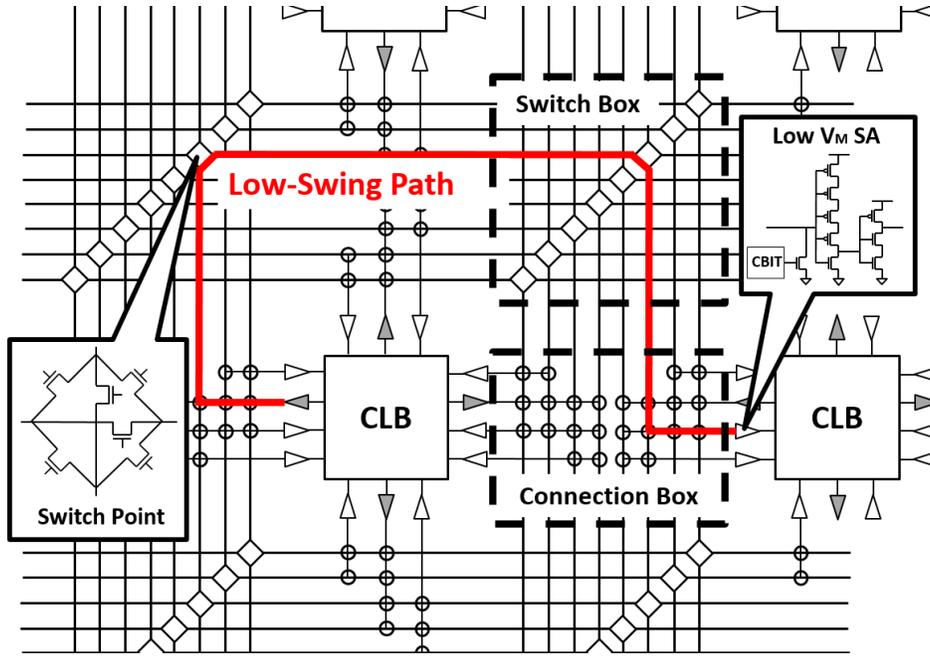


Figure 2: Basic structure of low-swing interconnect

swing reduces energy effectively. Furthermore, the low- V_M SA that receives the reduced swing signals at the input to the CLBs reduces delay by detecting the signal earlier in its transition than traditional receivers or SAs. A separate voltage rail V_{DDC} is also used to control the gate voltage of switches. Increasing V_{DDC} can reduce delay with small energy penalty.

2.2.2 Interconnect Path Distribution Exploration

We define the length of a global interconnect path as the number of switch boxes on the path from the start CLB to the destination CLB. The length of paths varies from 1 to over 100 and is not equally distributed. To understand the length of the majority of paths that this work is aiming at optimizing, we run the VPR [3] tool set on the MCNC benchmarks [32] to investigate the path distribution of the global interconnect. An Altera Stratix IV architecture (Stratix IV Device Handbook, available at www.altera.com), with fracturable LUTs, multipliers, and block RAMs, is selected as the target fabric to map the benchmarks. This architecture should be able to represent modern FPGAs.

The path distribution bar plot is shown in Figure 3. In the plot, paths are divided into 6 categories based on path length. The blue and green bars represent the path count distribution and the energy distribution. The red bar represents the average percentage of switches from the path that fall on branches rather than

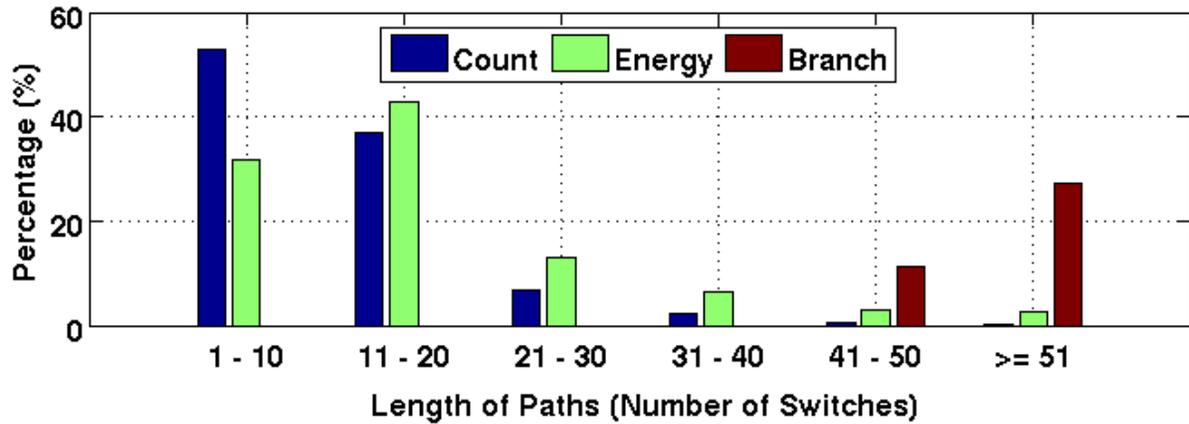


Figure 3: Path and branch distribution

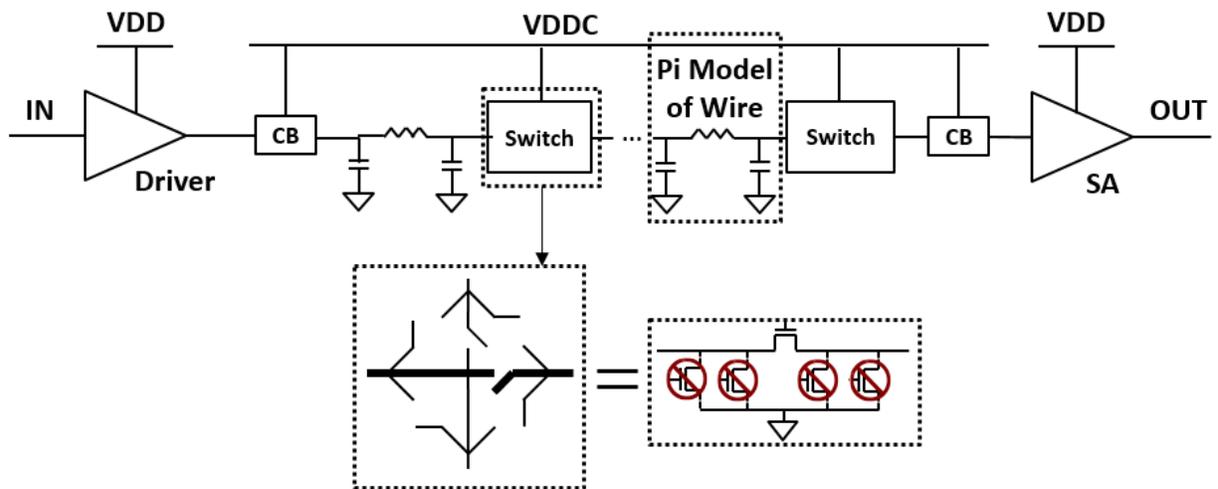


Figure 4: Diagram of the global interconnect path model

the main path. As indicated by the plot, paths shorter than length 40 take about 98% of the total path count and consume about 94% of the total global interconnect energy. Although branches are very common in the FPGA interconnect network, there are few branches on paths shorter than 40. Such analysis indicates that in order to increase energy efficiency of FPGA interconnect, circuit level optimization should mainly focus on paths shorter than 40 without branches. Some results of longer paths are also given and explained to cover a wider range of path length.

2.2.3 Custom Interconnect Model

Figure 4 shows the diagram of the global interconnect model used in this work. As mentioned in the above sections, a global interconnect path is defined as the circuit starting from the driver at an output of a CLB, passing CBs and switches, then ending at a SA of the destination CLB. We use the SA from [26] to receive low-swing signals coming out of the PG interconnect. Each wire segment is modeled as a Pi structure to represent the highly capacitive long wires. Each routing switch is modeled as one turned-on switch and four turned-off switches connected to ground, representing the signal path and the leakage paths respectively. Each CB is modeled as a multiplexer. A separate V_{DDC} voltage is applied to routing switches and CBs by high V_T configuration bits to provide flexibility in delay and energy. Low- V_M repeaters, having the same structure as a SA, can be inserted between two switches when regeneration is needed due to signal degradation. To optimize the circuit, parameters including the value of V_{DD} , V_{DDC} , the topology and size of CBs and switches, and the number of low- V_M repeaters will be varied and the corresponding influence on energy efficiency will be evaluated and discussed in the following sections.

2.3 Interconnect Circuit Optimization

2.3.1 Optimal Voltage of the Dual- V_{DD} Scheme

Supply voltage V_{DD} is a dominant knob for EDP. There are three components contributing to EDP: delay, active energy, and leakage energy. V_{DD} affects all of the important parameters for energy efficient FPGAs. Path delay decreases exponentially in the sub-threshold region at lower V_{DD} , while it only decreases quadratically in the above-threshold region. Energy is lower in the sub-threshold region and is dominated by leakage energy, while active energy, which decreases quadratically with V_{DD} , dominates total energy for super threshold operation [5]. In this work, V_{DD} is swept from 0.3V to 0.6V for paths with length of 10, 20, and 40. V_{DDC} is swept from 0 to 0.8V above V_{DD} . For 130nm CMOS, the minimum EDP is obtained at $V_{DD} = 0.5V$. Increasing V_{DD} from 0.5V to higher cannot further decrease EDP, but increases energy. On the other hand, reducing V_{DD} to 0.4V is very beneficial when energy is more important than energy efficiency, because much smaller energy can be achieved with small EDP overhead. However, reducing V_{DD} to 0.3V results in rapidly increased EDP but relatively smaller energy reduction.

Besides V_{DD} , energy and delay also depend on V_{DDC} . The active energy of the paths equals to $C \times V_{DD} \times \delta V$, where C is the equivalent lumped capacitance, V_{DD} is the supply voltage of the driver and the SA, and V is the voltage swing. For smaller V_{DDC} , the equivalent resistance of switches is large due to sub-threshold operation. Larger resistance leads to increased voltage drop and decreased voltage swing δV . Consequently, active energy and speed are both low. Applying a higher V_{DDC} , on the other hand, results in higher active energy but substantially reduced delay. In this work, V_{DDC} is swept with $V_{DD} = 0.4V$. The delay decreases sharply as V_{DDC} increases in the range of $V_{DD} \leq V_{DDC} \leq V_{DD} + 0.2V$. Keeping increasing V_{DDC} to above $V_{DD} + 0.2V$ can no longer reduce delay as significantly as before. On the other hand, energy increases slowly as V_{DDC} increases when $V_{DD} \leq V_{DDC} \leq V_{DD} + 0.2V$, while it experiences a much faster increase followed by a smaller one when $V_{DDC} \geq V_{DD} + 0.2V$. Similar to delay, the EDP decreases sharply at low V_{DDC} and slowly at high V_{DDC} . The “sharp-to-slow transition point” varies with path length. It can reach 0.3V above V_{DD} for paths longer than 40 and 0.1V for paths shorter than 10. The normalized data of sweeping V_{DD} and V_{DDC} (Figure 13 (a) & (b)) collected from measurement are discussed below.

2.3.2 Signal Degradation

In the sub-threshold region, the equivalent resistance between the drain and source of a transistor results in an IR drop for the signal passing through the channel. Since PGs are used to implement the routing switches of the low-swing interconnect, the signal swing will keep degrading along the path. As a result, the signal can become too small to be captured by the SAs. Although the switching threshold of a low- V_M SA in [26] can be as low as 0.09V at $V_{DD} = 0.4V$, repeaters are still needed to regenerate the signal when the signal swing degrades to be smaller than 0.09V.

Figure 5 shows the signal swing change after passing through different numbers of switches at $V_{DD} = 0.4V$. In the figure, the x-axis represents the number of routing switches signals have passed through, while the y-axis represents the value of the signal swing at the end of the path. The areas in different colors represent the $\mu \pm 2\sigma$ range (from Monte Carlo simulations in SPICE) of the swing at different V_{DDC} values. The areas in red, grey, and green represent V_{DDC} of 0.6V, 0.5V, and 0.4V, respectively. The black horizontal line represents the mean value of the V_M of the SA. The x-value where the V_M of the SA and the signal swing intersect represents the maximum number of switches signals can pass through without requiring any repeaters. The design of a low- V_M repeater in this work is the same as a low- V_M SA. If variation is ignored,

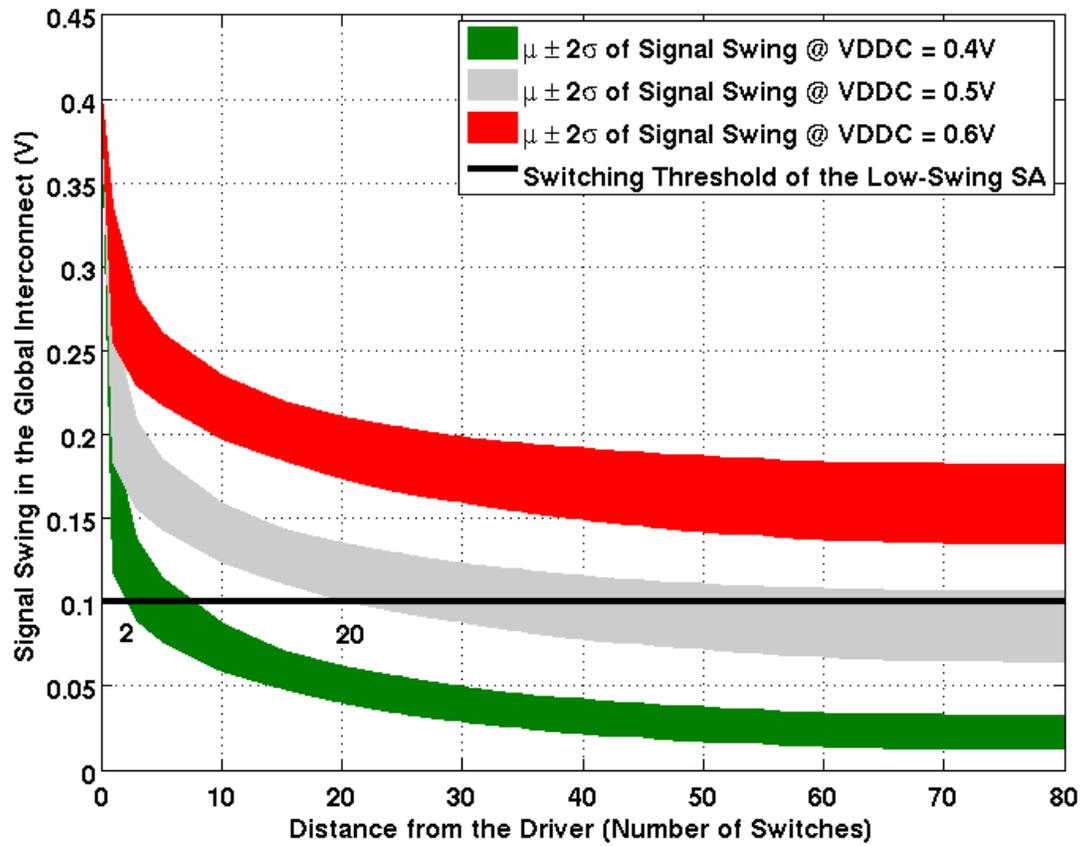


Figure 5: Range of signal swing for varying path length from Monte Carlo (MC) simulations with PG interconnect compared to the V_M of SA @ $V_{DD} = 0.4V$

a repeater is needed after the signal passes through 5, 40, or over 80 switches when V_{DDC} equals to 0.4V, 0.5V, and 0.6V, respectively. If considering variation, the switch numbers just mentioned become 2, 20, and over 80. When $V_{DDC} > 0.6V$, no repeaters are needed to maintain functionality of a path shorter than 80. Researchers in [26] also showed that the low- V_M SAs and repeaters can reduce variation effectively.

2.3.3 Repeater Number Optimization

Inserting repeaters implicates not only functionality, but delay and energy as well. Inserting repeaters increases the lumped capacitance load in the routing channel, resulting in increased active energy. However, the influence on delay after inserting repeaters is unclear. In this work, the number of low- V_M repeaters is varied. The results show that increasing the number of repeaters increases both delay and energy for paths shorter than 80. In these cases, the optimal number of repeaters in terms of energy and delay is zero. The detailed data (Figure 12) collected from measurement will be shown later in this chapter.

2.3.4 Connection Box (CB) Topology Optimization

The CBs in FPGAs targeting high performance are implemented by multiplexers with buffers to make connections between the routing fabric and the CLBs. For low energy FPGAs, buffers are removed. According to our simulation results, CBs contributes 13.4% of total delay and 2.6% of total energy to a low-swing path with length of 40. To reduce delay and energy of CBs, architecture optimization is needed.

Figure 6 shows three candidate topologies of the CBs for sub-threshold FPGAs. The 1-stage design has the smallest delay because it adds only one transistor delay to the interconnect path. However, the capacitance load of this design is the sum of all drain/source capacitance of N transistors, where N represents the number of inputs of the multiplexer. In addition, the signal swing is also large. As a result, the 1-stage design suffers from high energy. In contrast, the full multiplexer benefits from both low active and leakage energy, but suffers from slow speed. Both of the two designs cannot guarantee the maximum energy efficiency in sub-threshold. The 2-stage multiplexer is a good alternative to balance energy and delay. The ED curves, histograms from MC simulations, and area of the three topologies are compared in Figure 7 (a), (b), and (c), respectively. As shown in the figure, the delay of the 2-stage multiplexer is 16% smaller than the full multiplexer, while the energy of the 2-stage multiplexer is 5% lower than the 1-stage design. In addition, the

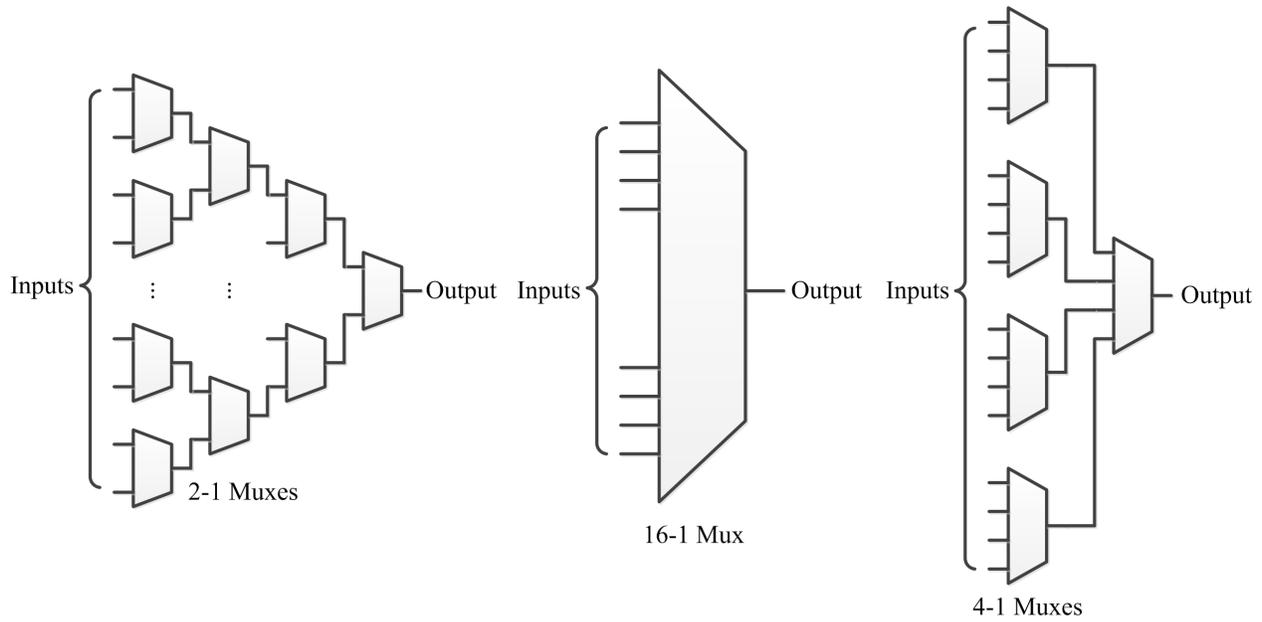


Figure 6: Schematic of different CB topologies: (a) full multiplexer (b) 1-stage multiplexer (c) 2-stage multiplexer

2-stage design has the smallest variation among the 3 candidates. The overhead of using a 2-stage design is area (2.6X larger than a full multiplexer when $N = 40$). Considering energy efficiency and variation, the 2-stage design is optimal.

2.3.5 Switch and Driver Size Optimization

Since no buffers in the routing switches, drivers are the only consumer of the active energy in low-swing interconnect. To achieve low energy, large drivers are not acceptable. However, simply reducing energy by decreasing driver size as much as possible is also not a good choice when delay is already large in the sub-threshold region. Under these circumstances, finding a driver size to balance energy and delay becomes a problem. The transistor sizes of the routing switches also need to be optimized for the same reason. Routing switches with a larger size introduce larger capacitance load into the interconnect fabric but result in larger signal swing and smaller delay.

Figure 8 (a) shows the simulated ED curve of a path of length 40 sweeping the driver size from 5X to 20X. Increasing the size of drivers from 5X to 20X reduces delay by 55% with a 39% energy overhead. This result implies that a larger driver may result in a smaller EDP. Figure 8 (c) shows the histograms of the same

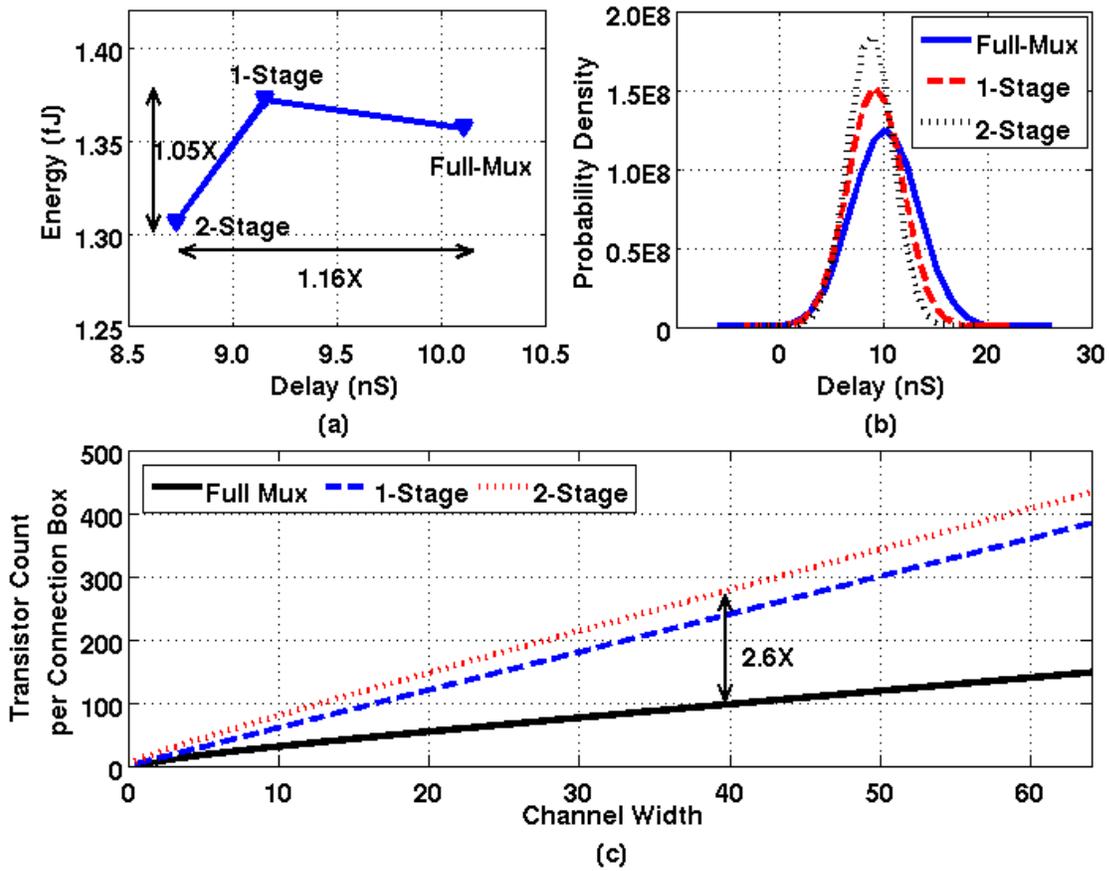


Figure 7: . Comparison of different CB topologies (a) ED curve @ $V_{DD} = 0.4V$ (b) variation @ $V_{DD} = 0.4V$ (c) area

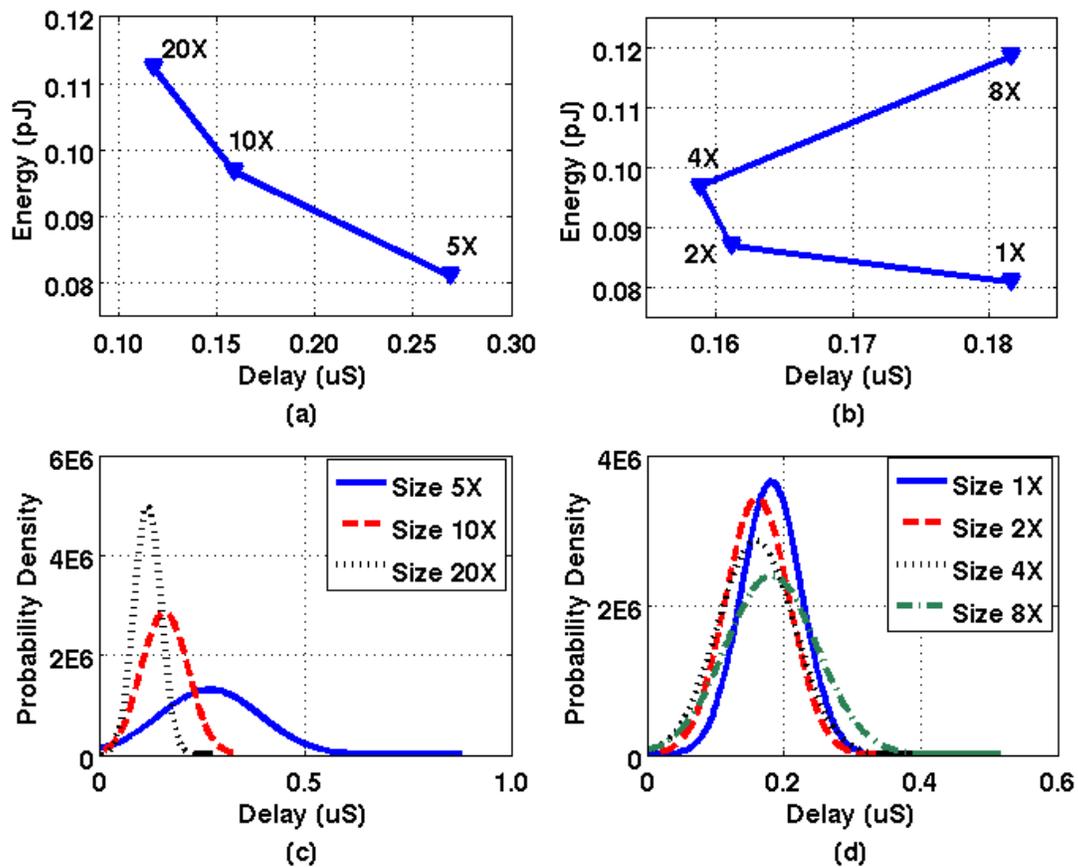


Figure 8: (a) The ED curve for a length 40 path with varying driver size @ $V_{DD} = 0.4V$ (b) with varying switch size @ $V_{DD} = 0.4V$ (c) histograms of length 40 path delay with varying driver sizes @ $V_{DD} = 0.4V$ (d) and with varying switch sizes @ $V_{DD} = 0.4V$

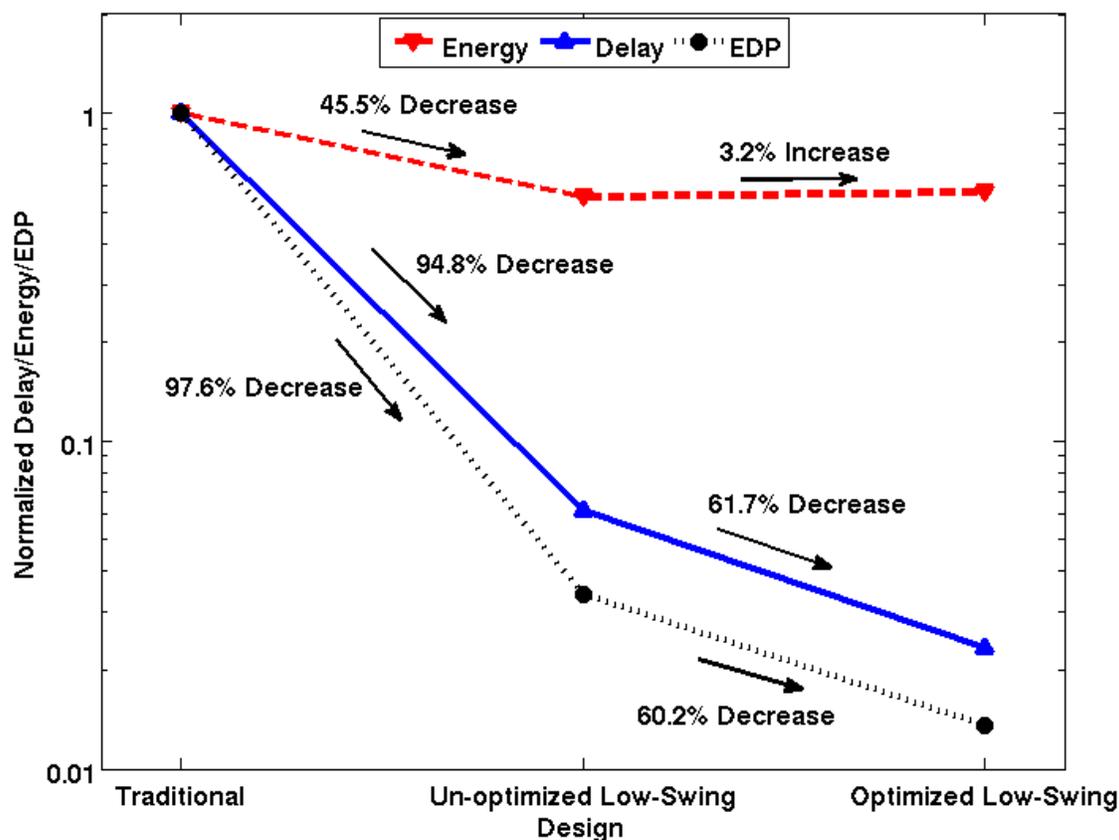


Figure 9: Comparison of the normalized delay, energy, and EDP @ VDD=0.4V

path with different driver sizes from MC simulations. Larger driver size leads to smaller variation because of larger current in the path. Furthermore, increasing the driver size above 10X results in diminishing variation reduction. Figure 8 (b) and (d) show the ED curve and histograms of a path with a length of 40 for varying sized routing switches from 1X to 8X. Across the design space, up to 13% delay reduction and 33% energy reduction can be achieved by using the optimal switch size. The histograms of using different PG sizes are similar. In the next section, we will show the measured data from a test chip. The energy overhead, delay reduction, and the optimized size of drivers and switches on real silicon will then be shown.

2.4 Comparison of Designs

The simulation results of the traditional uni-directional interconnect, un-optimized low-swing design, and optimized design are compared in Figure 9. The optimized design has 61.7% smaller delay, 60.2% lower

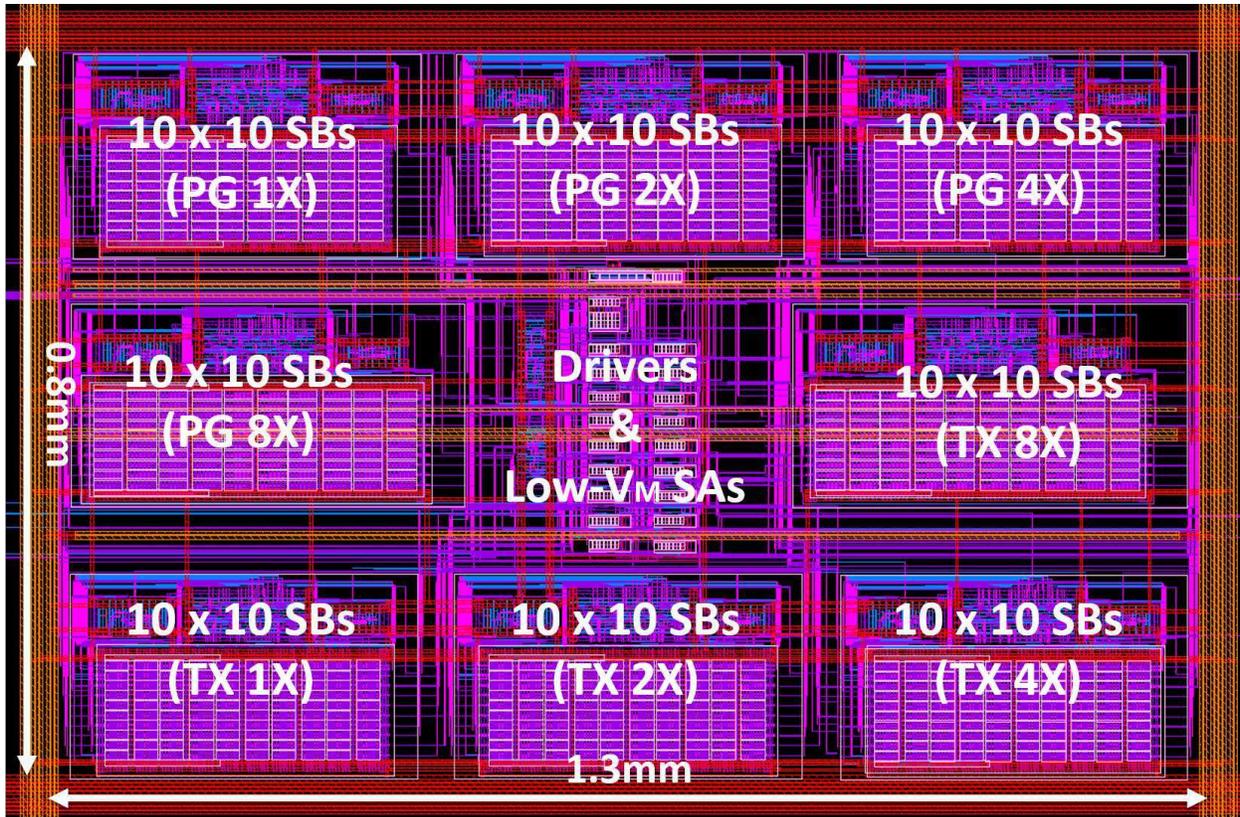


Figure 10: Block diagram of the test chip.

EDP, and 3.2% higher energy than the unoptimized design. The EDP is sharply reduced with very small energy overhead. Comparing to the traditional uni-directional design, the optimized low-swing design has 97.7% smaller delay and 42.7% lower energy.

2.5 Test Chip and Measurement Results

We implemented eight 10-by-10 dual- V_{DD} low-swing FPGA interconnect meshes with different topologies (PG and Transmission-gate (TX)) and sizes (1X, 2X, 4X, and 8X) of routing switches in 130nm bulk CMOS technology.

Wire segments are intentionally inserted between switches to imitate the RC of long wires in real FPGA fabrics. The meshes are driven by a driver block on the die. The driver block comprises drivers with different sizes followed by switches that can be configured to be turned on or off. The annotated layout of the test chip is shown in Figure 10.

	VDDC = 0.4V	VDDC = 0.5V	VDDC = 0.6V	VDDC = 0.7V	VDDC = 0.8V	VDDC = 0.9V	VDDC = 1.0V	VDDC = 1.1V	VDDC = 1.2V
Path Length = 10 Switches	Green								
Path Length = 20 Switches	Green								
Path Length = 40 Switches	Green								
Path Length = 50 Switches	Green								
Path Length = 60 Switches	Red	Green							
Path Length = 70 Switches	Red	Green							
Path Length = 80 Switches	Red	Green							
Path Length = 90 Switches	Red	Green							
Path Length = 100 Switches	Red	Green							

Figure 11: Measured shmoo plot of signal degradation @ $V_{DD} = 0.4V$, driver size 5X, and switch size 1X

The Shmoo plot in Figure 11 shows the measured functionality of paths including signal degradation at $V_{DD} = 0.4V$. In the figure, green indicates the signal can be captured by the SA after passing through the corresponding number of switches at the corresponding V_{DDC} , and red indicates the signal swing is too small to be captured. As shown, the SA successfully captures the signals after passing through at least 100 switches when $V_{DDC} \geq 0.5V$, but can only capture signals in paths shorter than 60 when $V_{DDC} = 0.4V$.

Figure 12 shows the measured ED curves of paths with different length and varying numbers of inserted repeaters. The number beside each point represents the number of repeaters inserted. The result indicates that inserting repeaters increases both delay and energy of all paths in the silicon.

As shown in Figure 13 (a), the measured EDP of a path with length of 40 decreases by 75% and the energy increases by 20% when increasing V_{DD} from 0.3V to 0.4V. Further increasing V_{DD} from 0.4V to 0.5V will decrease the EDP by 15% and increase the energy by 30%. If energy efficiency is considered, the optimal V_{DD} value is 0.5V. However, 0.4V is more desirable if we want to achieve lower energy with a small EDP overhead. Figure 13 (b) shows the EDP and energy of the same path as V_{DDC} changes. Increasing V_{DDC} from V_{DD} to $V_{DD} + 0.2V$ results in 40% EDP reduction with very small energy overhead. Increasing VDDC further cannot reduce EDP, but can increase the energy by 15%. In Figure 13 (c), the minimum EDP of the same path is obtained at a PG size of 4X and is 15% lower than the EDP at a PG size of 1X. In addition, the EDP of transmission gates is always larger than PGs. We also noticed in simulation that the optimal switch size is sensitive to the RC value of wires. If ignoring wire RC, the optimal switch size is 1X. On the other

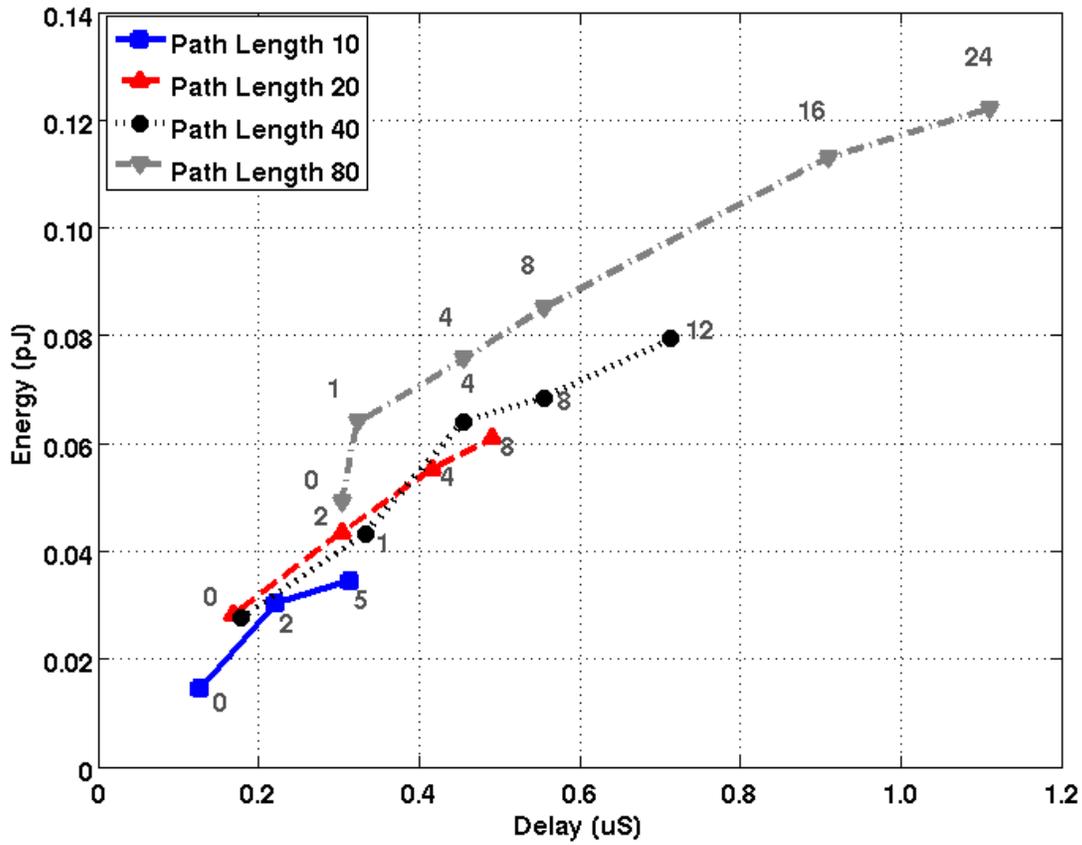


Figure 12: Measured ED curves for paths of varying length with different numbers of inserted repeaters @ $V_{DD} = 0.4V$

hand, 2X switches are needed when wires are shorter than 45m, while 4X switches are needed for longer wires. Figure 13 (d) shows that increasing the driver size from 5X to 10X reduces the EDP by 42% with a 2% energy overhead. Further increasing the driver size to 20X can decrease the EDP by 10% with a 10% energy overhead. Path with length of 10 has the similar conclusions.

The measurement results confirm the optimal choices of the topologies and sizes of the circuit components (driver size is 10X, switch topology is PG, switch size is 4X), the optimal value of supply voltages ($V_{DD} = 0.4/0.5V$, $V_{DDC}-V_{DD} = 0.2V$), the number of switches signals can pass through without repeaters (over 100), and the optimal number of inserted repeaters (no repeaters).

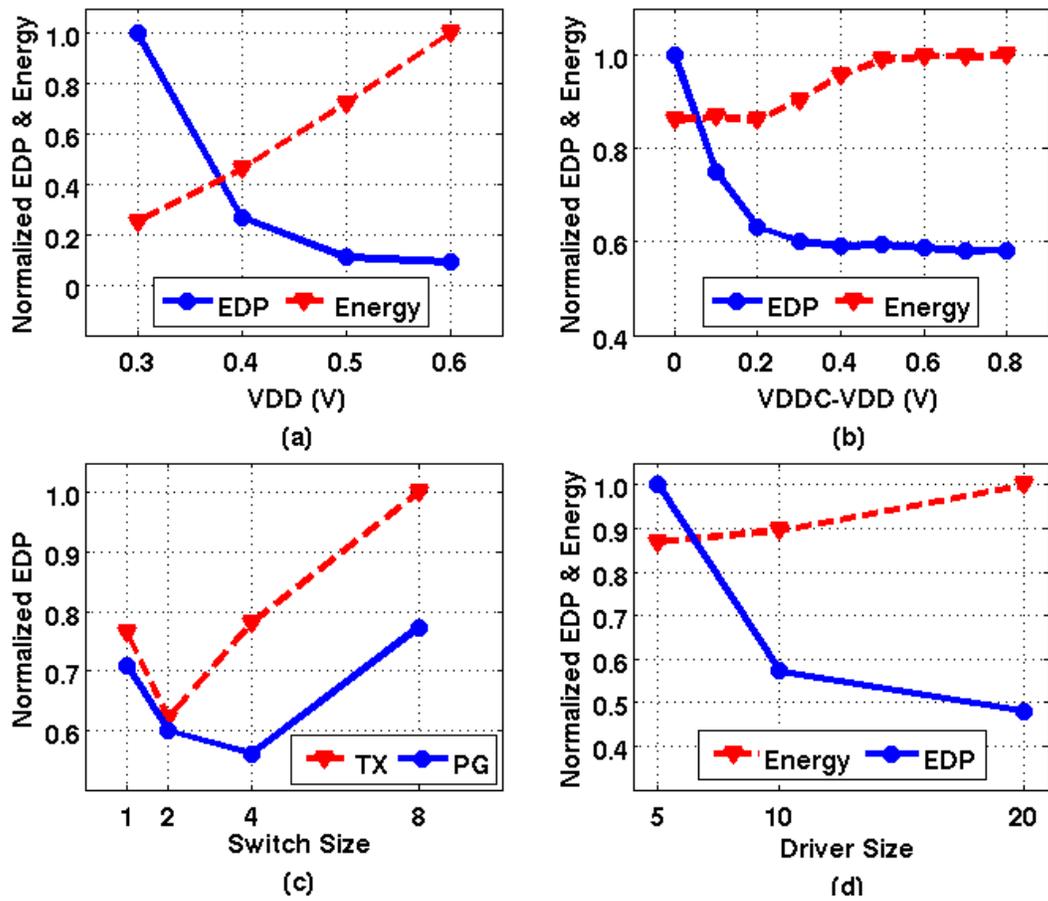


Figure 13: Measured path with length 40 for (a) V_{DD} optimization (b) V_{DDC} optimization @ $V_{DD} = 0.4V$ (c) switch size optimization @ $V_{DD} = 0.4V$ (d) driver size optimization @ $V_{DD} = 0.4V$

2.6 Conclusion

In this work, we presented an optimized low-swing dual- V_{DD} interconnect for FPGAs operating in the sub-threshold region. Considering both the energy and energy efficiency, we find the optimal topology (PG) and size (4X) of the routing switches, the best topology (2-stage design) of CBs, and the best driver size (10X). We also find the optimal voltage values ($V_{DD} = 0.4/0.5\text{V}$ and $V_{DDC}-V_{DD} = 0.2\text{V}$) for a 130nm process. In addition, signals can be captured by the low-VM SAs after passing through as many as 100 switches in series without repeaters in measured results. Inserting repeaters increases both the delay and energy of interconnect paths. A test chip in 130nm CMOS is fabricated. The measured data shows that the optimized design is 60.2% lower in EDP than a straightforward, un-optimized design at 0.4V for a 40-switch path. In simulation, the optimized low-swing design has 97.7% smaller delay and 42.7% lower energy than the traditional uni-directional design.

3 Voltage Scaling on FPGA Interconnects

As we introduced in the beginning of the thesis, power consumption in FPGAs is dominated by interconnect. Based on the work in superthreshold FPGAs, in this chapter we analyze the specialties in subthreshold FPGA interconnects and propose a voltage scaling technique for interconnects that optimizes the energy efficiency. We design a header-based voltage scaling technique and apply the voltage programmability to the single driver of each net in the interconnect. High V_{DD} is maintained for the critical path of the circuit while low V_{DD} is applied to short paths to reduce energy consumption. This design has a much lower area penalty in comparison with previous work and no performance degradation. A quantitative study is introduced on MCNC benchmarks. We make transistor-level simulations to show the energy of interconnect power is lowered by an average of 68.6% by applying the voltage scaling technique to the representatives of MCNC benchmarks [32]. Also, we show that the benchmarks can be applied with this programmable technique with an average of 98% of all the nets. Thus, this proposed design idea shows promise.

3.1 Introduction

For all the low power applications, FPGA is a competitive and attractive design option due to its high flexibility and low NRE (non-recurring engineering) cost. The increasing importance of power in FPGA has led to a lot of related work. Tuan and Lai [30] analyzes the leakage power of a superthreshold commercial FPGA architecture using 90nm technology and introduces some techniques to reduce the power of FPGAs. [1] works on the technique to reduce the active leakage power of multiplexers in FPGAs. [22] introduces a pre-defined dual- V_{DD} /dual- V_t FPGA to reduce both dynamic and leakage power. However, these works concern the techniques to reduce the logic block power in FPGAs. In [13], the authors propose a fine-grained power gating technique to the LUTs and apply it to an image processing application. [29] proposes a new DVS algorithm to the logic blocks to make them self-adaptive in operations. In [28], the authors summarize the current work on low power FPGA including device level technology, a dual voltage technique, and clock gating, which are mostly on the architecture level or logic block level. However, the logic power contains only the power of LUTs, flip-flops and MUXes which occupies less than 35% [21] of the total energy, while the interconnect of a FPGA consumes 68% of the total energy. In [21], they mention this and shift the main content of work to the interconnect of FPGAs and propose a programmable V_{dd} structure to the routing

switches of FPGA interconnects to reduce the power.

However, most of the work are based on the system- or architecture-level analysis of FPGAs. Due to the characteristics of FPGAs, it is difficult to analyze an FPGA's performance and energy efficiency at the transistor-level (SPICE simulation), which is mostly used in almost all VLSI areas or system design flows. Besides, as the ultra low power demands are increasing in recent years, subthreshold operations in FPGA are a good solution, but most of the work is not in this domain. [17] introduces a subthreshold FPGA using graphene interconnects and measures data from an FPGA test chip fabricated in a 0.18- μm SOI process which can function at supply voltages as low as 0.26V. In [4], it introduces the challenges in subthreshold CMOS and specifically in FPGAs. In this chapter, we apply a programmable V_{DD} structure to the interconnects. We do not focus on designing SRAM bit-cells, path drivers, or exploring architecture of interconnects. We will use the dual voltage scheme (the gate voltage of the routing switches is pulled up) for the routing switches as the base case of subthreshold FPGAs.

The rest of this chapter is organized as follows: Section 2 discusses background knowledge, including the conventional FPGA interconnect and the subthreshold FPGA interconnect. Section 3 introduces the opportunity and motivation we have in applying voltage scaling technique to subthreshold FPGA interconnects. Section 4 discusses our design flow. Section 5 gives the simulation results.

3.2 Background

3.2.1 Conventional Island Style FPGA Interconnect

FPGA interconnects consume almost 80% of the area and 70% of the power. Similarly, as introduced in [21], Figure 14a shows the conventional FPGA interconnect architecture, which is the most widely used island style FPGA architecture. Configurable logic blocks (CLB) are consisted with basic logic elements (BLE), which are basically Look-Up-Tables (LUT). However, we do not discuss them here. CLBs are surrounded by routing channels which consists of wire segments. Wire segments connects all CLBs, routing switches and connection switches. The inputs and outputs of CLB are connected to the routing channels via connection boxes, as showed in Figure 14b. In the intersection of horizontal channels and vertical channels, a switch box (SB) is used to route the channels, as showed in Figure 14c. Figure 14c shows the most widely used routing algorithm in island style FPGA interconnect. All the channels with the same number can be connected

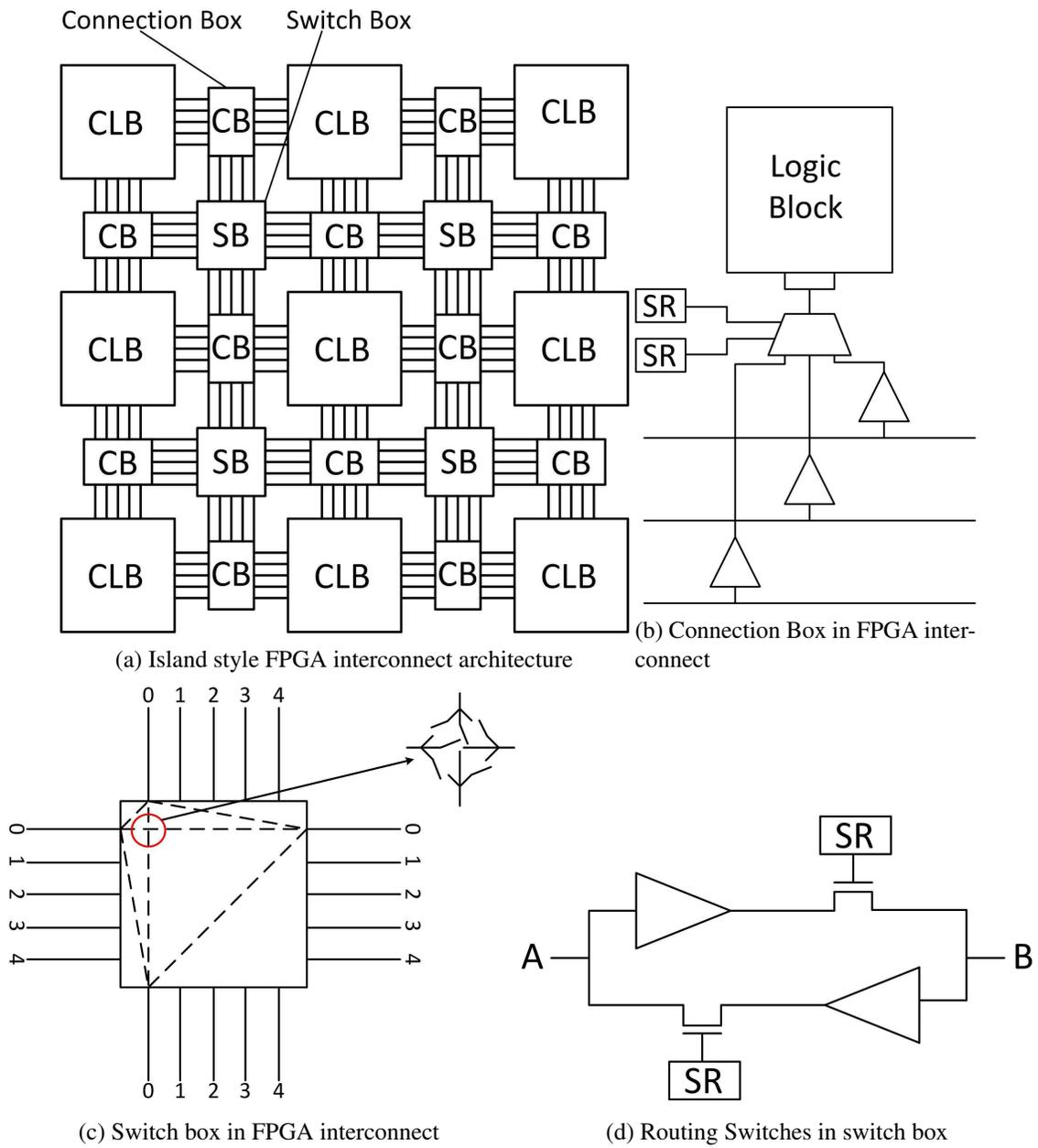


Figure 14: Conventional FPGA interconnect architecture

with each other by programming through the SRAM bitcells. Thus, in each switch point, which refers to the intersection of the channels with the same name, there are six routing switches in total to implement the routing ability. In a conventional FPGA interconnect, the routing switch in SBs use a bi-directional structure. Tri-state buffers are used to implement the independent programmable connection.

In this thesis, we use VPR [3] to place and route the MCNC benchmark set. For the architecture parameters, we use a standard FPGA architecture: a cluster of 10 in BLE (6 inputs per LUT). For the channel width, in order to let the placing and routing affect the energy analysis the least, we let VPR to route the benchmarks with a smallest channel width number for each benchmark. Since the transistor-level simulation (SPICE) is time consuming and all the MCNC benchmarks have a similar net distribution, so we pick up 7 of the benchmarks to show the simulation results.

3.2.2 Subthreshold FPGA Interconnect

The design of subthreshold FPGA requires a low power design goal and the guarantee of robustness in subthreshold domain. As showed in Figure 15a different from the design of the superthreshold FPGA, in which tri-state buffers are employed in each of the switching point, so that the transition of signals can have a swing compensation while going through the path in the circuit, in subthreshold FPGA design, the energy consumed by buffers are saved by replacing them with pass gate transistors. The gate of pass gate transistors are configured by a SRAM bitcell. The signals are driven by the driver in the CLBs while the lost in signal swings are compensated by the end of the path, a level translation circuit (LTC) as shown in Figure 15c. This is a revised buffer and both of the stages are sized to intentionally strengthen or weaken the PUN or PDN. The stack transistors can also reduce the leakage power. The basic subthreshold FPGA interconnect path is showed in Figure 15d. For connection box, in this figure it gives an example of transmission gate, it could also be a multiplexer-based connection box.

The CLB design are also different in subthreshold and superthreshold FPGAs. The choice of CLB topologies and architecture affects the performance and power consumption of a FPGA. In our work, we do not discuss CLB design.

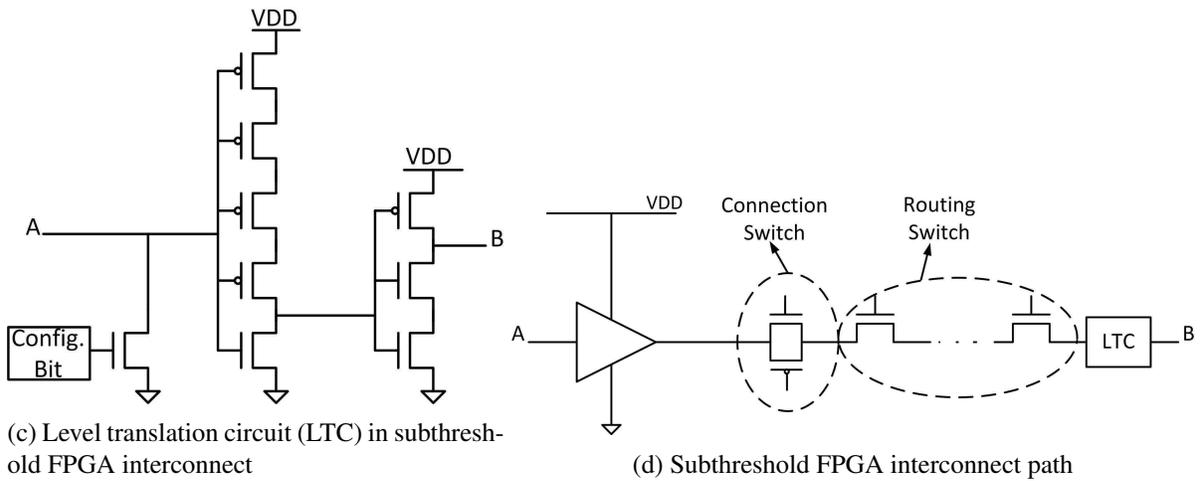
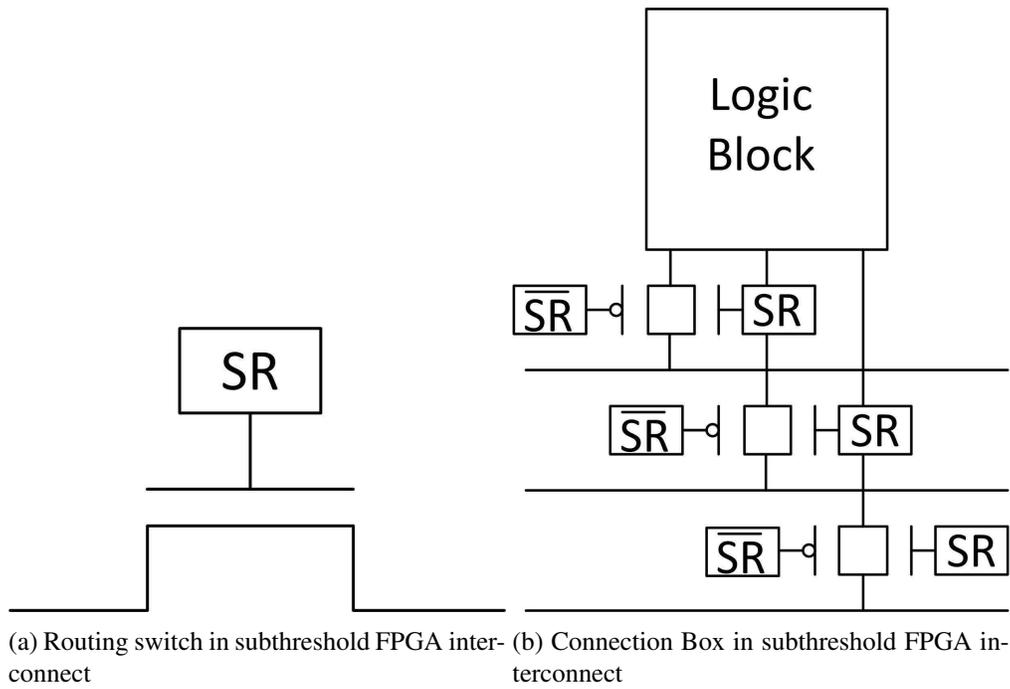


Figure 15: Subthreshold FPGA interconnect

Table 1: Extracted path information of MCNC Benchmarks

Benchmark	Total Switch #	Length of Longest Path	Average Switch#	Average Path Length
alu4	8,078	41	11.61	7.06
apex2	11,459	24	11.84	5.98
apex4	8,039	24	11.52	6.53
bigkey	6,191	19	6.05	4.48
clma	68,031	53	14.13	8.87
des	8,327	27	8.36	5.85
diffeq	6,734	34	7.15	5.14
dsip	5,944	19	8.61	5.92
elliptic	21,405	44	11.23	7.37
ex5p	7,313	25	10.95	6.68
ex1010	32,109	50	12.49	6.80
frisc	26,985	54	15.45	9.12
misex	7,624	21	10.66	5.87
pdc	41,282	39	18.02	9.14
s298	7,075	25	9.81	6.06
s38417	29,246	62	8.20	5.77
s38584.1	31,219	68	8.58	6.22
seq	10,867	25	12.38	6.53
spla	27,362	42	15.14	7.73
tseng	3,667	22	6.25	4.36
Average	N/A	N/A	11	7
Largest	N/A	68	N/A	N/A

3.3 Motivation

The difference in the interconnect design of subthreshold and superthreshold FPGA design and the specialty of an FPGA circuit provide an opportunity of a voltage scaling implementation space for increasing the energy efficiency in subthreshold FPGA interconnect. In this section, we will explore the prospects on scaling the energy of an subthreshold FPGA circuit without the penalty of performance degradation.

We use the MCNC benchmark set to analyze the distribution of nets in a subthreshold FPGA. By running VPR, we get the placing and routing information of each net in the benchmarks as shown in Table 1. For each of the 20 benchmarks, we analyze the length and breadth of each net of them. We take ALU4 as a representative from the 20 benchmarks and analyze its nets distribution. Figure 17a shows the distribution of the longest net lengths in all the paths of ALU4 after mapping by VPR. And Figure 17b shows the distribution of the total switch count of the whole ALU4 benchmark. For both of the longest net lengths and the total switch count, the distributions show a strong long tail shape, which means, most of the nets

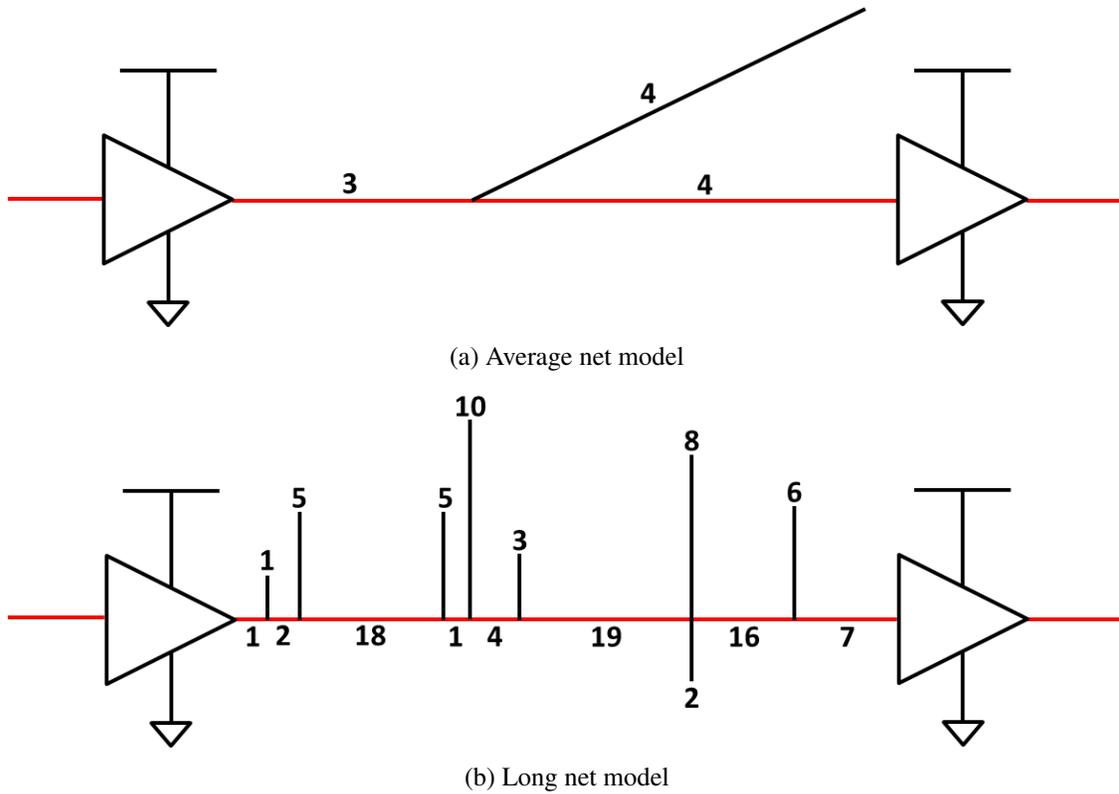
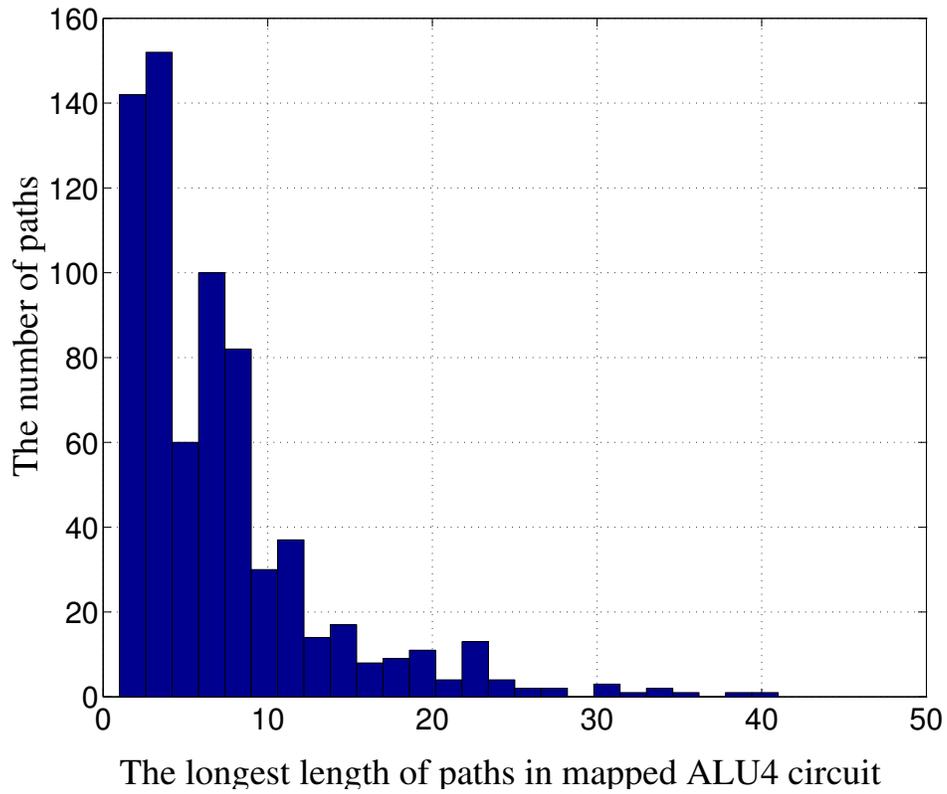
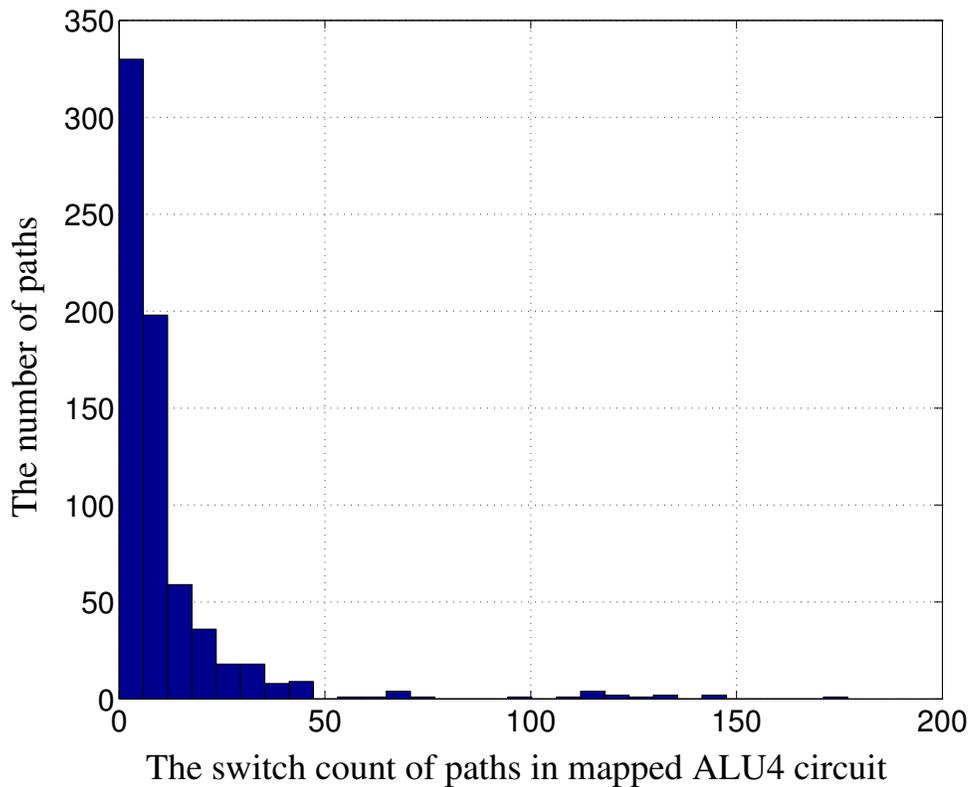


Figure 16: Interconnect circuit models

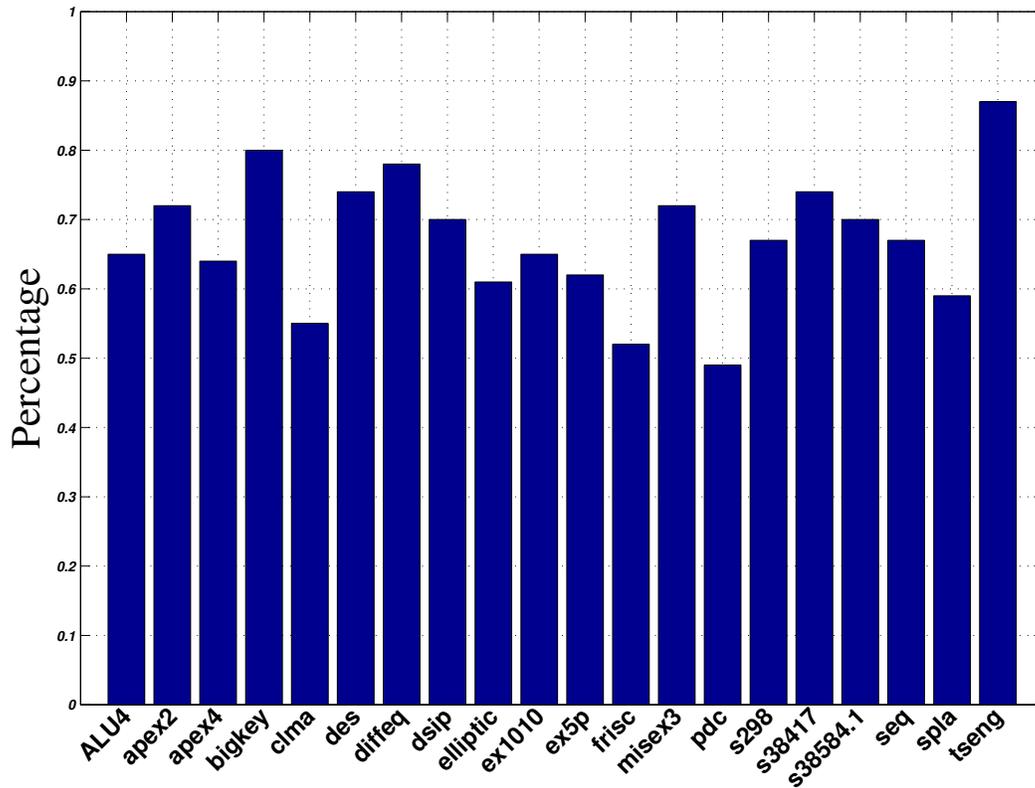
in ALU4 are actually very short while only a small amount of nets are long nets including the critical path of the whole circuit. We cannot put the distribution figures for all the benchmarks here, but all of them do show the same characteristics. Instead, based on the statistics of all 20 benchmarks, we extracted two models here: long net model (LM) in Figure 16a and average net model (AM) in Figure 16b, which refer to the longest net and the average net in all 20 benchmarks respectively. In order to make sure that the models are reasonable for the paths study of an subthreshold FPGA, we compared the nets of all 20 benchmarks with the AM since LM is the biggest net on all 20 benchmarks. As shown in Figure 18a and Figure 18b, we count the number of nets in each benchmark which are shorter than AM both on the main path length and the total switch count. The results show that for each benchmark, more than 50% of the paths are shorter than AM in both views.



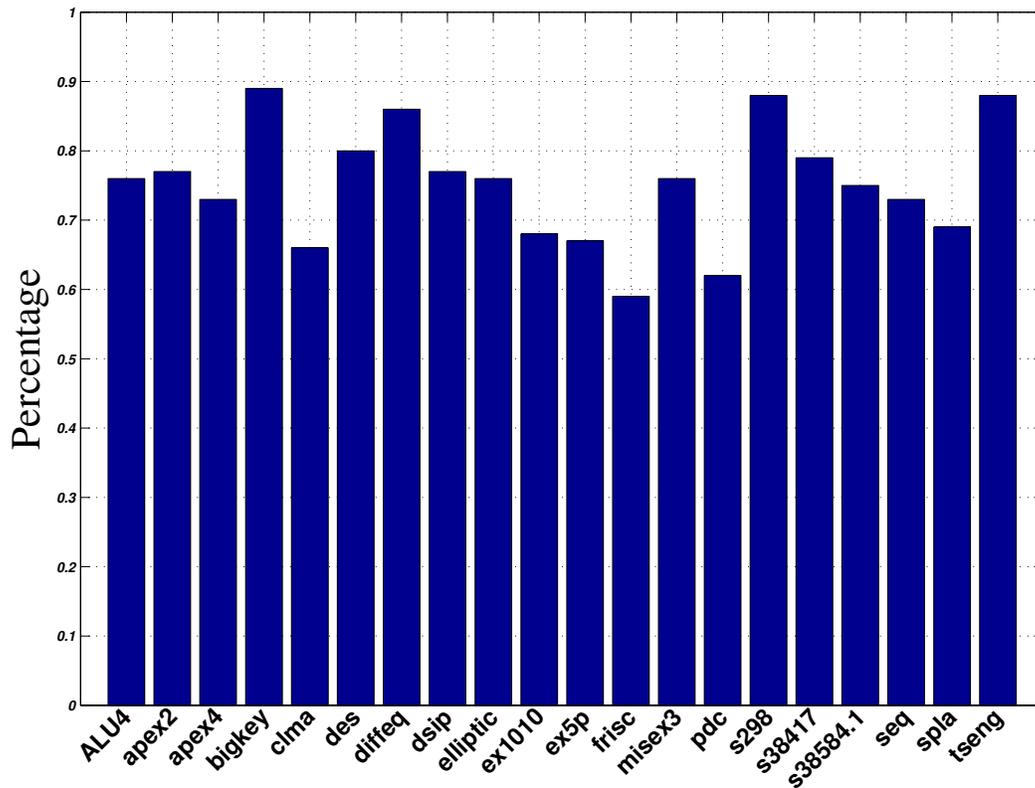
(a) Longest net distribution in ALU4



(b) Switch count distribution in ALU4



(a) Percentage of the longest nets shorter than AM circuit in 20 MCNC benchmarks



(b) Percentage of the switch count of nets shorter than AM in 20 MCNC benchmarks

Figure 18: Percentage of the paths in MCNC benchmarks in comparison with customized circuit models

3.4 Voltage scaling technique for subthreshold interconnect

In this section, we are going to introduce the voltage scaling technique of subthreshold FPGA interconnect.

3.4.1 Performance and energy exploration

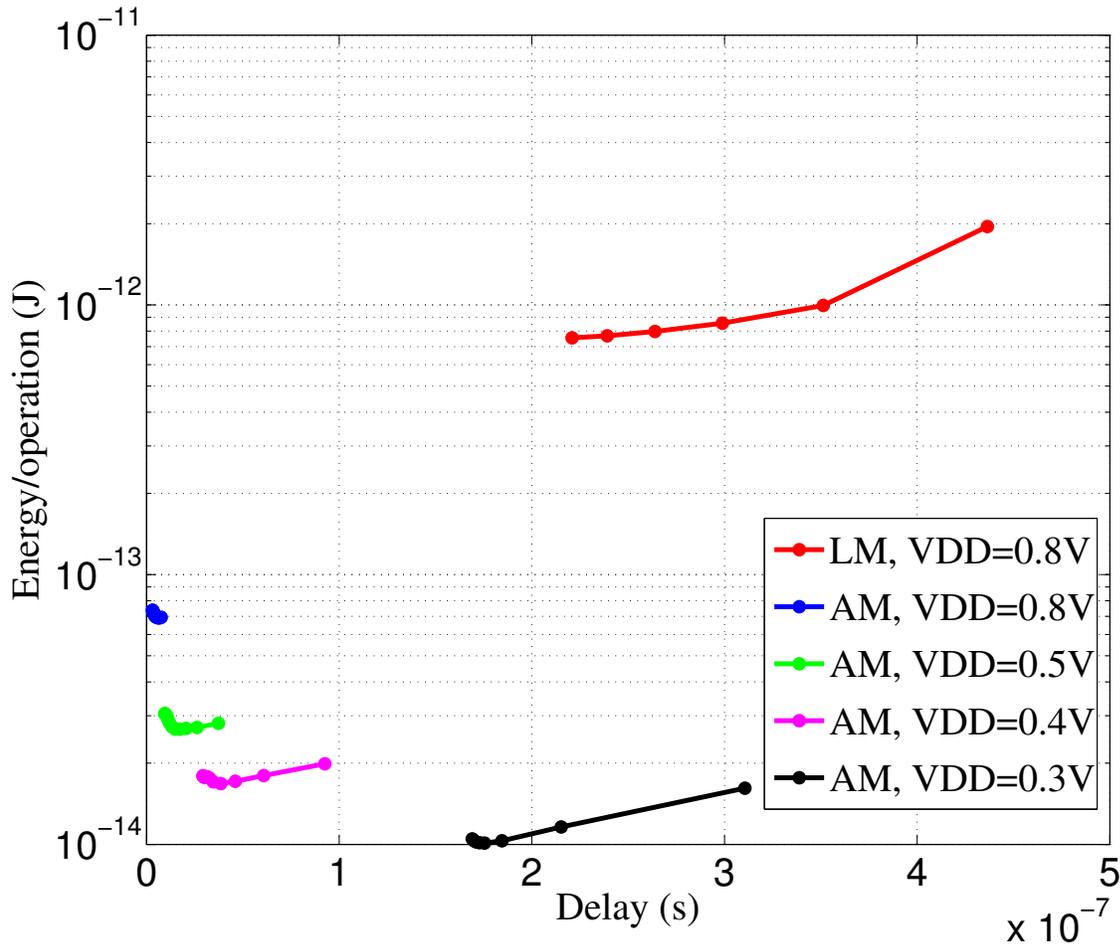


Figure 19: Energy-delay curve of LM and AM circuits with different VDDs

In this section, we are going to explore the interconnect circuits of a subthreshold FPGA. This exploration is based on the AM circuit we discussed in section 3. As we mentioned before, the subthreshold FPGA we consider uses a dual- V_{DD} scheme—that is, the switch points in the whole FPGA are pulled up by a higher voltage supply V_{DDC} both to compensate the voltage loss and get the best energy-delay performance. According to the previous work, V_{DDC} is set to be 0.15 higher than V_{DD} as a baseline setting. Under this

setting methodology, the FPGA circuits achieve the best operating point under the view of both energy efficiency and performance. In this simulation, we run AM circuit through a set of different V_{DD} s and plot the energy-delay curves in Figure 19. We also plot the energy-delay curve of LM at $V_{DD}=0.8V$. As we can see from the ED curves, a V_{DD} of 0.8V consumes almost 8X more energy than a V_{DD} of 0.3V. The LM circuit has a much higher delay than the AM circuit. In other words, lowering V_{DD} achieves a promising gain of energy efficiency with a relatively lower delay than the critical path.

3.4.2 Header-based voltage programmability

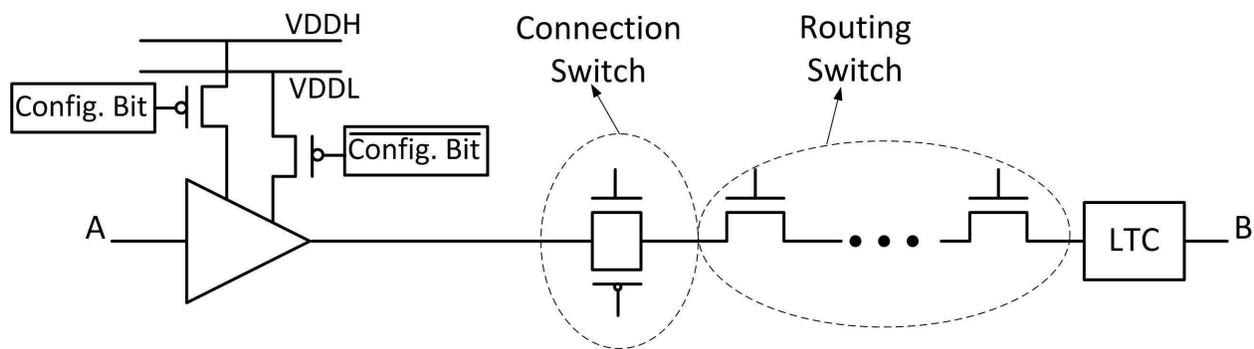
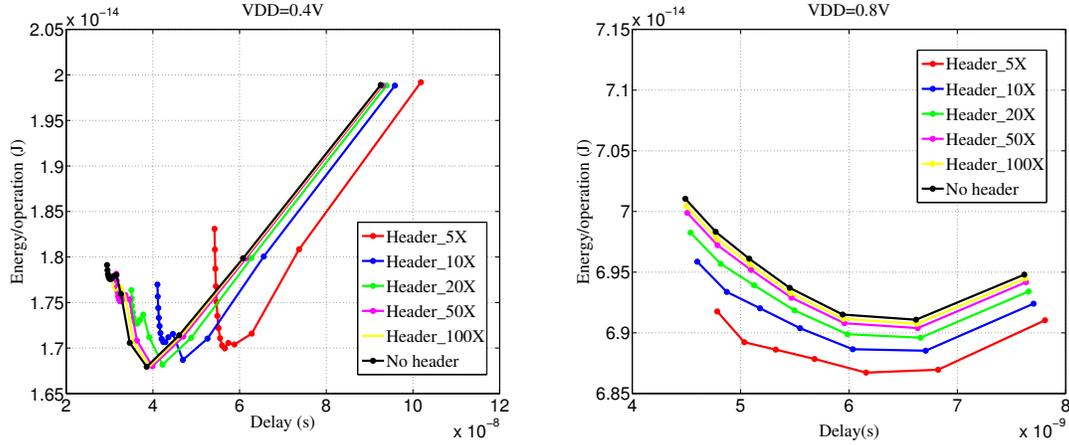


Figure 20: Header-based voltage scaling technique in subthreshold interconnect

We propose to use a PMOS header structure to implement the voltage scaling technique. As shown in Figure 20, the PMOS transistor configured by a configuration bit, which is a SRAM bitcell. The driver is connected with two different voltage rails through the PMOS transistors. By configuring the bitcell connected with the gate of the PMOS, different supply can be applied to the driver in order to tune the paths energy and performance. We have 2 configuration options here: a higher voltage V_{DDH} and a lower voltage V_{DDL} . Actually, this can be achieved by only one SRAM bitcell by using the not logic output of the bitcell.

We sweep the sizes of headers to explore the effect of headers to the circuit. As shown in Figure 21, we simulate the AM circuit at different V_{DD} s and show the results of $V_{DD}=0.4V$ and $0.8V$. Larger headers have the most similar performance and energy in comparison of the circuit without headers (black curve). But using headers can bring benefits of performance with higher V_{DD} s while benefits of energy with lower V_{DD} s. In our work, to balance the area, performance and energy, we choose size 20X as the header size.



(a) Energy-delay curves when sweeping header sizes at VDD=0.4V (b) Energy-delay curves when sweeping header sizes at VDD=0.8V

Figure 21: Header size exploration

3.5 Simulations

In this section, we discuss the transistor-level simulations we have done based on the voltage scaling technique.

As shown in Figure 18, all the MCNC benchmarks have similar nets distributions. Specifically, we run SPICE simulations on 7 out of the 20 benchmarks: ALU4, dsip, seq, s298, spla, tseng, and apex2.

Specifically, we first list the detailed simulation results for ALU4. In the simulation results shown for ALU4, we set the applicable factor to be 60%, which means 60% of the nets are applied with V_{DDL} , while the rest long nets remain controlled by V_{DDH} . Figure 22a first shows the delay of all nets in ALU4 and Figure 22b shows the delay after applying the header-based voltage scaling technique. The right part of the delay distribution remain the same while the delays of short nets shift to the right without passing the critical delay of the whole circuit. Accordingly, Figure 23a and Figure 23b give the energy change of the circuit without and with the header-based voltage scaling technique respectively. After applying with the scaling technique, the energy is reduced by 17.3% without any penalty of the performance (applicable factor is 60%).

We increase the applicable factors of every of the 7 benchmarks until it cannot be raised further. From this, we get the maximum applicable factors for each of the 7 benchmarks. In other words, we apply V_{DDL} to more nets according to the net's size until the critical path is exercised (some net with $V_{DDL}=0.4V$ consumes

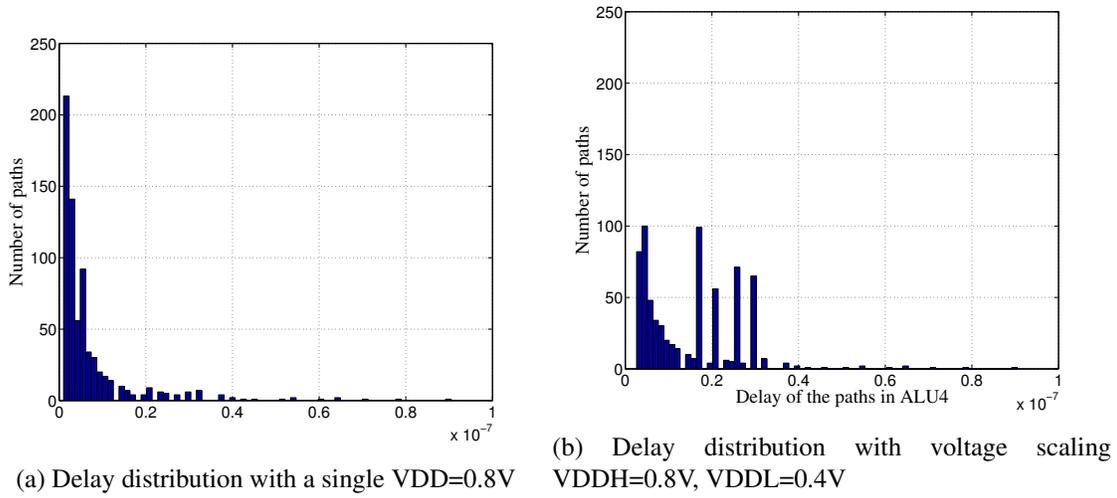


Figure 22: Delay of ALU4 with and without voltage scaling technique

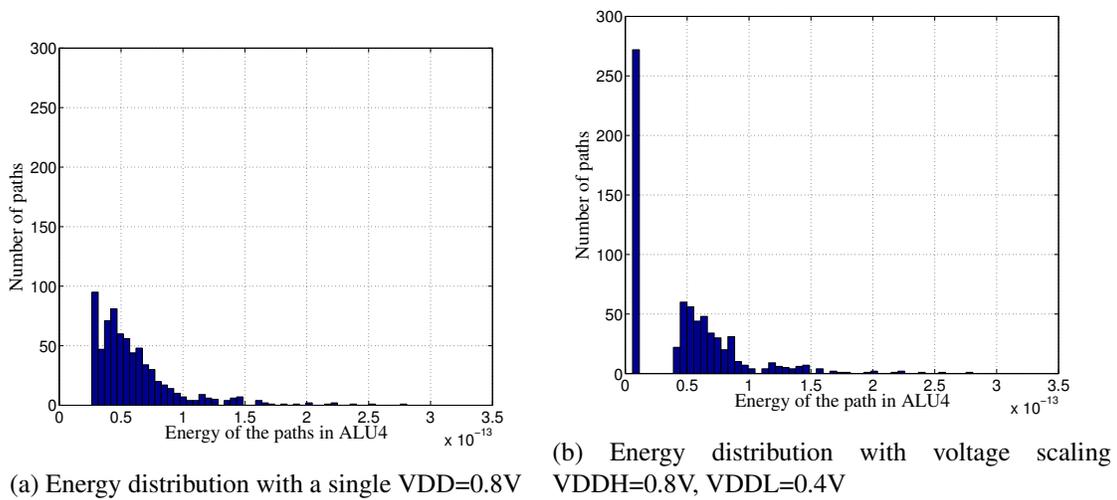


Figure 23: Energy of ALU4 with and without voltage scaling technique

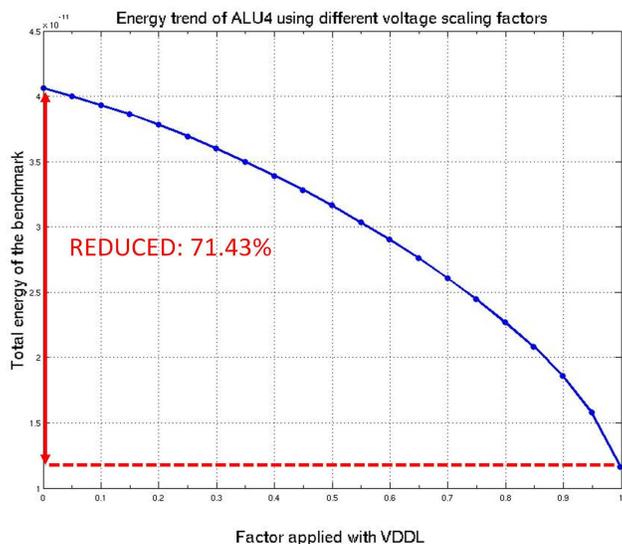


Figure 24: The effect of the applicable factors on energy saving for ALU4

longer delay than the critical path). In Figure 24, we show the total energy consumed per operation of ALU4 as the applicable factor increases from 0 to the maximum applicable factor (more than 99% for ALU4). With the maximum applicable factor, energy is reduced by 71.43%. Similarly, we conducted the same simulation to all the 7 benchmarks, and Figure 25 shows the maximum applicable factors for all 7 benchmarks. The average maximum applicable factor for all 7 benchmarks is as high as 98.00%. This is a strong potential to reduce energy consumption by using this proposed programmable voltage scaling technique. Figure 26 shows the energy saving with its own maximum applicable factor for each of the 7 benchmarks. The average energy savings is 68.60%.

3.6 Conclusion

In this chapter, we discussed a programmable voltage scaling technique to reduce energy consumption subthreshold FPGA interconnects by using a programmable header structure and showed the simulation results of the energy saving by using this idea. Our proposed header-based voltage scaling technique saves more area than the dual-VDD programmability design in [21], and our work applies to different application domain. Verified by simulation under the scenario of a 0.8V/0.4V voltage combination, the average portion of nets (applicable factor) of 7 MCNC benchmarks is as high as 98% and by applying the technique, we achieve an average of 68.6% energy reduction in the 7 MCNC benchmarks using the maximum applicable

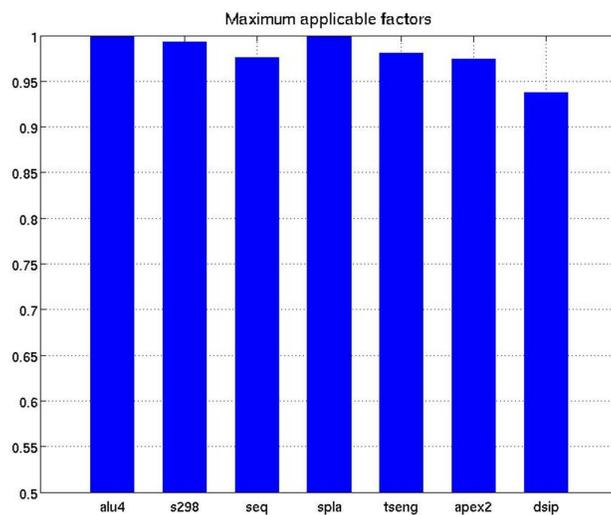


Figure 25: The maximum applicable factors for all the 7 MCNC benchmarks

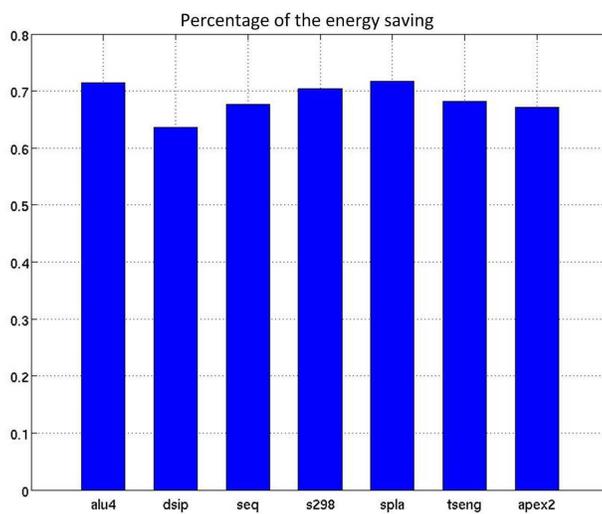


Figure 26: The energy saving with maximum factors for all the 7 MCNC benchmarks

factors. This idea gives a promising design prospect in optimizing energy efficiency in subthreshold FPGA interconnect. Future work must include fine-grained study on the voltage tuning algorithm, which is able to apply the proper voltage to every path precisely to achieve an ultra-optimized power consumption reduction or gives a dynamic voltage scaling implementation on subthreshold FPGA interconnect.

4 A single ended level converter circuit design for ultra low power low voltage ICs

² In this chapter, we discuss the design of an ultra low swing level converter, which can be employed in a sub-threshold FPGA circuit to implement voltage scaling, and also can be applied in a ultra low power system that requires a low voltage swing (e.g., an energy harvesting system).

We introduce the motivation of this charge pump based ultra low swing level converter design including the potential application of it and the state of the art. Second, we discuss the charge pump design and how it works. Third, we discuss the level converter design based on the sub-threshold charge pump and the simulation results. Finally, we show the measurement results and the comparison with prior work.

4.1 Introduction

Energy autonomy is a critical feature required to enable the large scale deployment of ultra low power (ULP) systems in the internet of things (IoT), with energy harvesting being accepted as a more viable means to provide power. However, many challenges face energy harvesting circuits, which require operation at very low power and voltage levels [14]. Figure 27 shows the block diagram of a generic energy harvesting system. The lifetime of the system depends on the energy stored on the energy harvesting capacitor C to provide power for the system. At runtime, as the energy stored on C is being consumed, the voltage on the capacitor, V_{cap} , decreases. The voltage at which the system stops operating (system threshold voltage) must be brought down to increase system lifetime. From the energy utilization perspective, the system threshold voltage should be brought down as low as possible to make full use of the stored energy. In order to more fully take advantage of the energy stored on the energy harvesting capacitor, SoCs under ultra-low voltage have been proposed in [15], which operate below 160mV. Typical ULP SoCs frequently use timers to keep the circuit functional even when the voltage is very low [11]. However, the outputs of these ULP sub-threshold circuits also operate at a very low voltage level, which causes communication problems with the core voltage levels off-chip or with other peripheral circuits. Level converters are necessary in such a system to interface between the low voltage domain and the nominal voltage domain. In this chapter, we

²This chapter is mainly from publication [1].

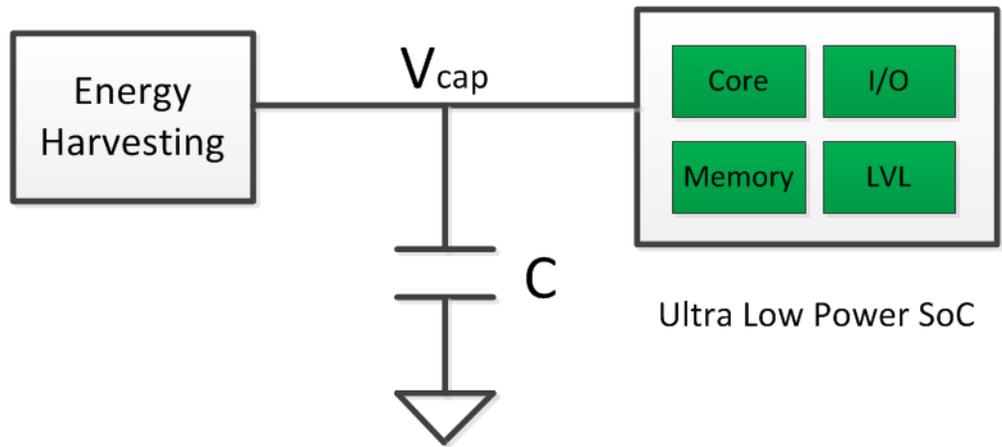
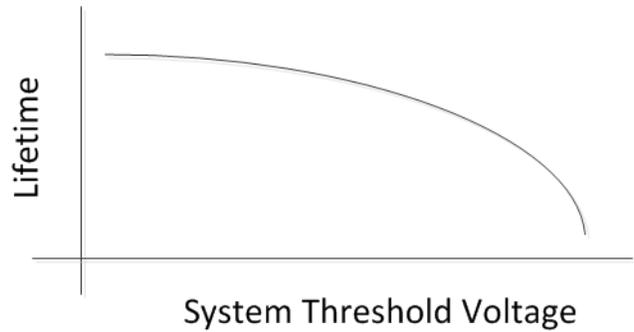


Figure 27: Generic energy harvesting based SoC.

present a low swing level converter that can convert from 100mV (simulation) and 145mV (measurement) level input signals to 1.2V using a single ended charge-pump based topology.

A traditional level converter can convert from nearly 400mV to 1.2V via a cross coupled stage. 400mV is still higher than required in an energy harvesting ULP SoC. Lower input signals can kill the positive feedback and prevent conversion with the traditional design. Several low voltage level converter circuits have been proposed in the literature. A low swing level converter can convert from 210mV to 1.2V with a bootstrapping technique [8]. A dynamic logic level converter can convert 300mV to 2.5V [6]. However, dynamic logic uses more power and area in ULP applications. A two-stage ULP level converter can convert from 188mV to 1.2V achieving ULP operation [31].

In this work, we design a level converter that can potentially convert 100mV to 1.2V using a charge-pump. The charge-pump stage increases the swing before level conversion, which helps in initiating the positive feedback. Also, a 130nm CMOS chip has been fabricated and the measurement results show a robust conversion from 145mV to 1.2V.

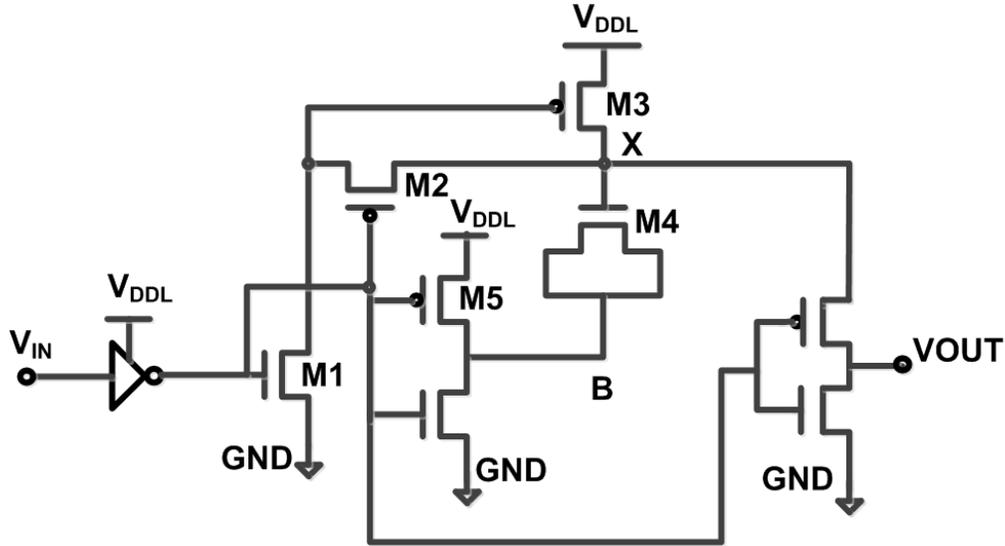


Figure 28: Schematic of the 2X charge pump used in the level converter.

4.2 Sub-threshold charge pump

Figure 28 shows the schematic of a 2x charge pump used in the proposed work. When V_{IN} is low, M1 turns on which turns on M3. X is pulled up to V_{DDL} while B is pulled down to GND by the inverter connected to it. Next, V_{IN} goes high and turns on M2 and M5, which leads to the upconversion of B from 0 to V_{DDL} . Since X was charged to V_{DDL} previously, the upconversion of B causes X to go from V_{DDL} to $2xV_{DDL}$ at the output of the charge pump. In deep sub-threshold operation with a V_{DD} between 100mV and 300mV, node X falls ideally at 200mV and 600mV, respectively. But in sub-threshold, the low slew rate prevents a full doubling of voltage when V_{DD} is very low ($<200mV$) because of the higher discharge caused by leakage. In this charge pump design, we do not require an additional body bias control circuit.

4.3 Implementation of the level converter

Figure 29 shows the architecture of the proposed topology, which combines two charge pumps and a level converter design. The first stage provides the differential inputs doubled by the 2x charge pumps. The second stage is a cross-coupled differential inverter (e.g., the traditional level converter shown in Figure 30) that restores the final output to full swing (0 to V_{DDH}). The output of the charge pump stage overpowers the equilibrium of the second stage and drives the PMOS to pull up the internal node (A or B) and trigger the positive feedback within the conversion stage.

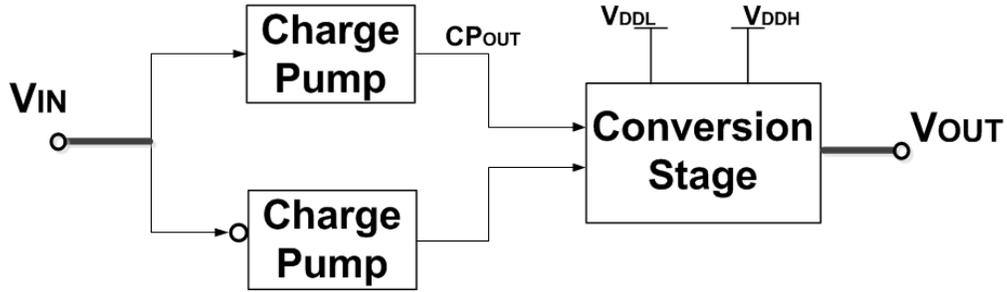


Figure 29: Architecture of the proposed level converter.

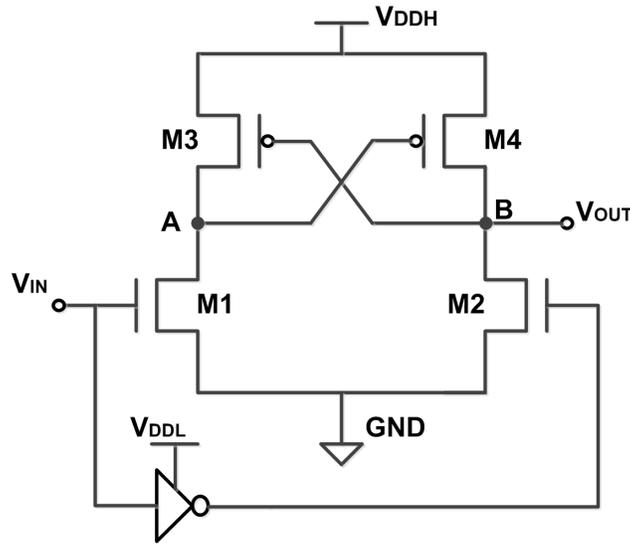


Figure 30: Schematic of the traditional level converter.

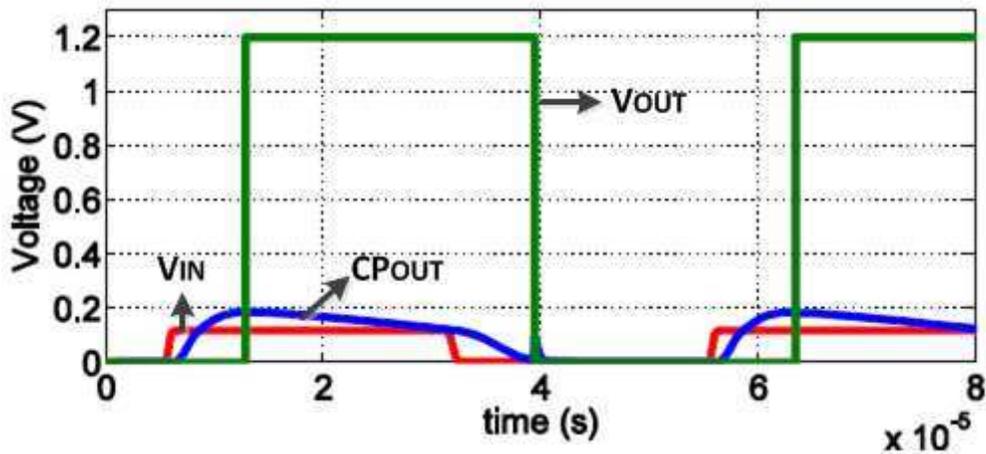


Figure 31: Functional waveform of CPBULS @ $V_{DDL}=120\text{mV}$

We propose two designs that use charge pump outputs to drive a traditional level converter and a different ultra-low swing (ULS) level converter structure from [31], respectively. We call the former proposed

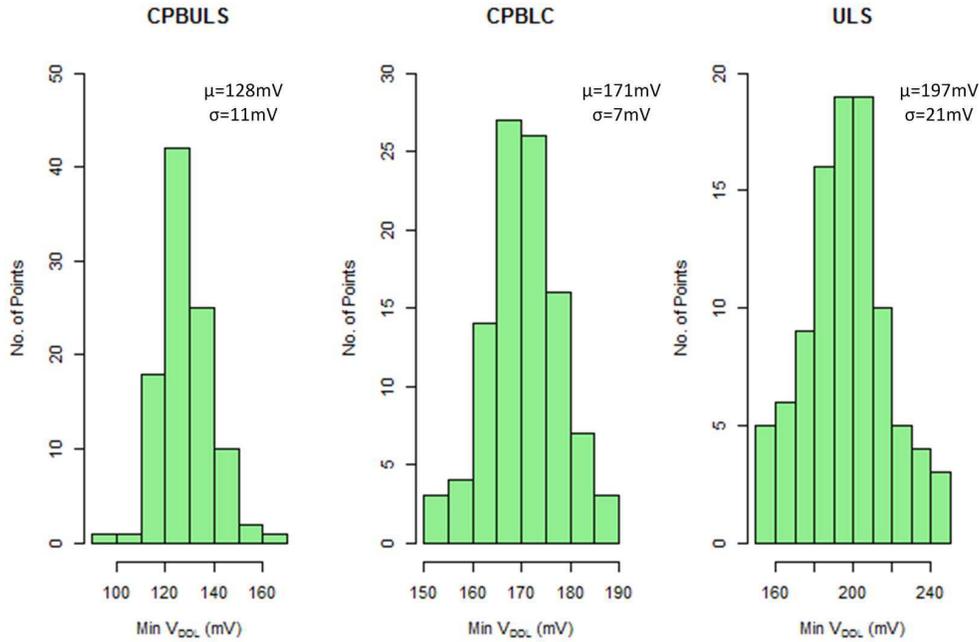


Figure 32: Monte Carlo simulation results of minimum converting input voltage of CPBULS, CPBLC and ULS level converters, $t=27^{\circ}\text{C}$.

level converter the Charge Pump Boosted Level Converter (CPBLC) , and we call the latter proposed level converter the Charge Pump Boosted Ultra Low Swing Level Converter (CPBULS). Figure 31 shows the simulation of the CPBULS at 120mV. The signals labeled in Figure 31 correspond to the signals in Figure 29. As V_{IN} goes high or goes low, one of the charge pump outputs, e.g., CPOUT, increases and initiates the positive feedback resulting in voltage conversion. Figure 32 shows the minimum input swing results of 100 Monte Carlo simulations for CPBULS, CPBLC, and ULS level converters. The charge pump technique decreases the minimum operating voltage of [31] (ULS), further lowered down to an average of 128mV, while the best case (among the 100 iterations) is 99.6mV in CPBULS, and an average of 171mV in CPBLC.

Figure 33 shows the simulation results of the minimum input voltage of CPBULS and CPBLC level converters under different temperatures. At -20°C , CPBULS and CPBLC can work at 145.4mV and 192.8mV respectively, while at 100°C , they can work at 116.4mV and 144.3mV respectively. Simulation shows that our charge-pump based level converter has lower temperature dependence for minimum operating voltage.

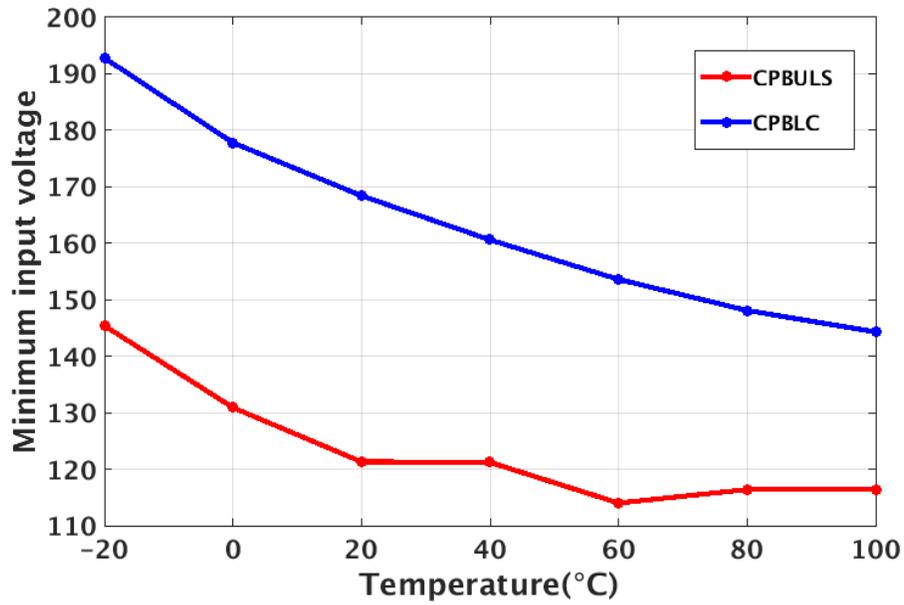


Figure 33: Simulation results of the minimum input voltage vs. temperature of CPBULS and CPBLC level converters.

4.4 Measurement Results

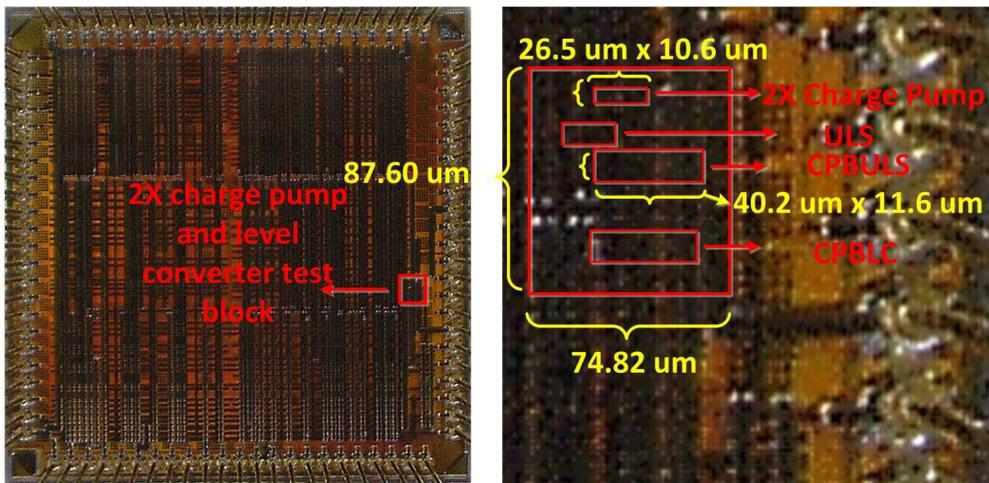


Figure 34: Die photo of the 130nm CMOS technology chip.

The design was fabricated in a 130nm CMOS process. Figure 34 shows the die photo of the test chip, the 2x charge pump consumes about $260 \mu m^2$ while the CPBULS level converter consumes about $466 \mu m^2$. Figure 35 is the testing measurements of the 2x charge pump, which starts working from a 170mV input in the worst case. The blue lines are the measurement results while the red line is from simulation. After VIN is higher than 200mV, the boosting factor is stable at 2x. Figure 36 shows the measurement results of

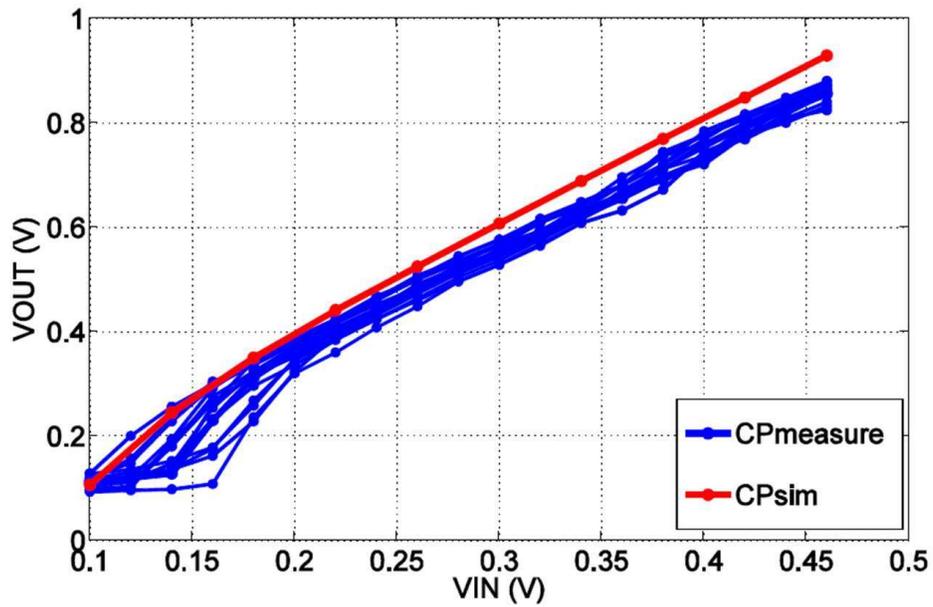


Figure 35: Simulation and measurement results of the input vs. output voltage of the charge pump stage of the level converter.

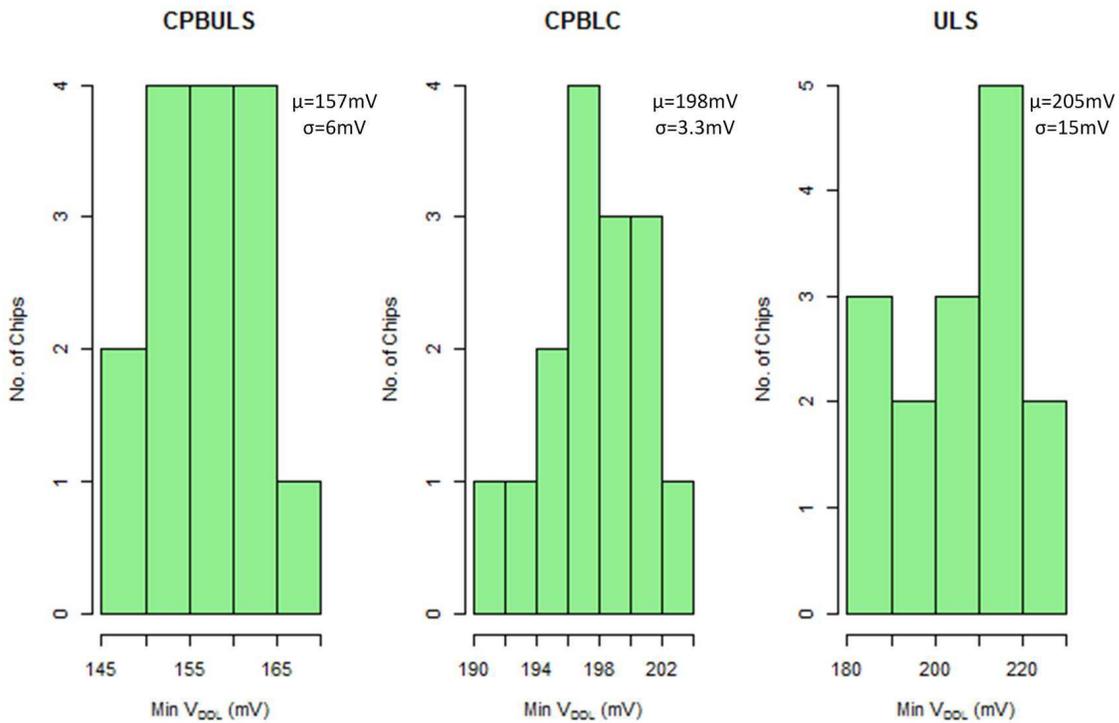


Figure 36: Measurement results of minimum converting input voltage of CPBULS, CPBLC and ULS level converters.

the minimum operational input swing for CPBULS, CPBLC, and ULS level converters across 15 chips. The limitation of this design is slower transition times that lead to higher energy per conversion due to the extra leakage.

4.5 Conclusion

Table 2: Comparison between prior work and the proposed work

	[31]	[23]	[10]	[6]	This Work
Minimum V_{DDL}	188mV	200mV	400mV	300mV	145mV
Energy/bit	-	10fJ	327fJ	850fJ	1.2pJ
Chip/Simulation	Chip	Simulation	Simulation	Chip	Chip
Maximum Frequency	17.3MHz	10MHz	1MHz	8MHz	8kHz
Area(μm^2)	-	-	120.9	112000	466
Technology	130nm	90nm	180nm	130nm	130nm

Table 2 compares with prior work. This includes the work both of chips and simulations. As we showed in the previous sections, this charge pump based level converter CPBULS upconverts reliably from 145mV to 1.2V from chip testing results while the simulation shows the minimum converting voltage can be as low as 99.6mV, which is a wider conversion range. The best energy per conversion is reported as 10fJ in [23] from simulation results. This work achieves 1.2pJ energy per conversion which is 1.5x of that in [6] from chip measurement, but a 2x conversion ability.

This proposed work can work under a lower operating voltage which can further improve the energy utilization of an ultra low power system such as an energy harvesting system. In such a system, energy stored in the energy harvesting capacitor can be further used with less limitation from the voltage.

5 Conclusion and future work

Ubiquitous computing is emerging to play an important role in the past decade including the ultra low power miniature devices such as body sensor network systems. Developing such systems are very costly and the limitation of resources can lead to high design complexity. As we discussed in the thesis, sub-threshold FPGA compensates both of energy efficiency and design flexibility. Since FPGA interconnect dominates the energy and delay, this thesis proposed an optimization work of energy efficient interconnect in sub-threshold FPGA. We also discussed a programmable voltage scaling technique on sub-threshold FPGA interconnects in the thesis to further reduce the energy consumption. Lastly, we proposed an ultra low-swing level converter which can not only be applied to such an sub-threshold FPGA but also any ultra-low power system that requires low-swing operation and better efficiency of energy.

5.1 Summary

In this thesis, we first optimized the topology of switch boxes, connection boxes, transistor sizes and the value of supply voltages to reduce the energy and improve energy efficiency.

We studied the dual- V_{DD} scheme: main supply voltage (V_{DD}) and boosted voltage (V_{DDC}). With a 130 nm CMOS technology chip and simulations, the optimal voltage values are a V_{DD} of 0.4V - 0.5V while V_{DDC} is 0.2V higher than V_{DD} .

We explored the routing switches topology and connection boxes topology. The optimal topology of switch boxes is a pass gate (PG) with size 4X. The optimal connection box topology is to use a two-stage design. The optimal driver size is 10X.

In the exploration of signal degradation, under $V_{DD}=0.4V$, signals can be captured by the low-VM sense amplifiers after passing through as many as 100 switches in series without inserting a repeater. And generally, inserting repeaters will increase both delay and energy of an interconnect path.

The optimized sub-threshold FPGA design is 60.2% lower in EDP than an unoptimized design at 0.4V for a 40-switch interconnect path. Specifically, the optimized design has a 97.7% smaller delay and 42.7% lower energy than the traditional uni-directional design.

Following the sub-threshold FPGA interconnect optimization, we proposed a programmable voltage scaling

technique by using a configurable header of the driver of the interconnect to further reduce energy consumption without any performance penalty. According to the characteristics of FPGA interconnects, the average of 98% of the nets can be applied with a lower voltage thus achieve an average of 68.6% energy reduction.

In the end, we proposed a single ended level converter circuit. This level converter can be applied to a low power system especially sub-threshold ICs, such as an energy harvesting system. This level converter uses sub-threshold charge pumps to boost the ultra low swing signals. It consists of two stages. After the boosting stage, the conversion stage will convert the low-voltage signals to nominal voltage of the system (1.2V). This level converter is required in such a low voltage IC to implement the communication between low voltage and nominal voltage domain. The simulation results show the CPBULS level converter can convert from 99.6 mV while the measurement results show it can convert from 145mV. In the measurement results, the CPBULS level converter achieves an energy per conversion of 1.2fJ and a frequency of 8kHz.

5.2 Contributions

In this thesis, we presented a comprehensive exploration on sub-threshold FPGA interconnects and new ideas of designing an ultra-low swing single ended level converter. We demonstrated the conclusion of the exploration and suggested the optimal design parameters for a sub-threshold FPGA. Next, we introduced a design of level converters based on sub-threshold charge pumps. For all the work we fabricated test chips with a 130 nm CMOS technology.

The following lists the contribution of this thesis:

- Explores the circuit level parameters: topology and size of different parts of FPGA interconnect focusing on sub-threshold domain. And gives the optimal design parameters:
 - Switch box: pass gate (PG), size 4X of the technology size (130nm CMOS)
 - Connection box: a two-stage multiplexer design.
 - Driver at the beginning of an interconnect path: 10X of the technology size.
- Presents the best voltage settings based on delay and energy in the dual-VDD scheme of an sub-threshold FPGA:

– $V_{DD} = 0.4V$

– $V_{DDC} = V_{DD} + 0.2V$

- Studies the signal degradation along sub-threshold FPGA interconnect path and the need of repeaters to boost the signals: signals do not need a repeater to be boosted through at least 100 switches in series by using the low-VM sense amplifiers and inserting repeaters will increase both delay and energy of an interconnect path.
- In general, the proposed optimized sub-threshold FPGA interconnect design is 60.2% lower in EDP than an un-optimized design at 0.4V for a 40-switch interconnect path and the optimized design has a 97.7% smaller delay and 42.7% lower energy than the traditional uni-directional design. This can further improve the performance and energy efficiency of a FPGA in sub-threshold domain.
- Proposes the idea of programmable voltage scaling on FPGA interconnect by using a header structure. This depends on the unique FPGA interconnect structure that we can control the driver of each net to adjust the delay and energy.
- Explores the applicable portion of nets of MCNC benchmarks by transistor-level simulations and showed 98% of the nets of MCNC benchmarks can be applied with a lower voltage on average without any performance penalty, but an average of 68.6% energy reduction per operation.
- Introduces the idea of inserting a sub-threshold charge pump in the design of level converters. This charge pumping stage can extend the level converter's converting ability. If we take the ULS level converter as an example, the increased converting ability is around 45mV lower.
- The design of the 145mV CPBULS level converter can potentially convert from 99.6mV from simulation results. So by improving the layout it can potentially reach a wider converting range. This is so far the reported lowest converting voltage level converter.
- The CPBULS level converter can further lower down a system threshold voltage. In an energy harvesting system, energy can be further taken use of by allying this level converter to implement the communication between low voltage blocks and nominal voltage blocks.

5.3 Future work

In future work, we will need to be able to build a customized sub-threshold FPGA using the optimized interconnect structure and the programmable voltage scaling header structure. Most importantly, we will build a tool flow which can map the verilog/VHDL functions onto the customized FPGA similar to existing commercial FPGA tool flows. Furthermore, this tool flow can not only be able to map the functions from code but also be capable of configuring the voltage scaling header to set between V_{DDH}/V_{DDL} of each net according to the mapping results. And also, fine-grained programmable dynamic voltage scaling scheme can be studied for a more accurate configuring and energy consumption.

References

- [1] ANDERSON, J. H., AND NAJM, F. N. Active leakage power optimization for fpgas. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on* 25, 3 (2006), 423–437.
- [2] ANDERSON, J. H., AND NAJM, F. N. Low-power programmable fpga routing circuitry. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on* 17, 8 (2009), 1048–1060.
- [3] BETZ, V., AND ROSE, J. Vpr: A new packing, placement and routing tool for fpga research. In *Field-Programmable Logic and Applications* (1997), Springer, pp. 213–222.
- [4] CALHOUN, B. H., RYAN, J. F., KHANNA, S., PUTIC, M., AND LACH, J. Flexible circuits and architectures for ultralow power. *Proceedings of the IEEE* 98, 2 (2010), 267–282.
- [5] CALHOUN, B. H., WANG, A., AND CHANDRAKASAN, A. Modeling and sizing for minimum energy operation in subthreshold circuits. *Solid-State Circuits, IEEE Journal of* 40, 9 (2005), 1778–1786.
- [6] CHANG, I. J., KIM, J.-J., KIM, K., AND ROY, K. Robust level converter for sub-threshold/super-threshold operation:100 mv to 2.5 v. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on* 19, 8 (Aug 2011), 1429–1437.
- [7] CICCARELLI, L., LODI, A., AND CANEGALLO, R. Low leakage circuit design for fpgas. In *Custom Integrated Circuits Conference, 2004. Proceedings of the IEEE 2004* (2004), IEEE, pp. 715–718.
- [8] GARCIA, J., MONTIEL-NELSON, J., AND NOOSHABADI, S. High performance bootstrapped cmos low to high-swing level-converter for on-chip interconnects. In *Circuit Theory and Design, 2007. ECCTD 2007. 18th European Conference on* (Aug 2007), pp. 795–798.
- [9] GAYASEN, A., LEE, K., VIJAYKRISHNAN, N., KANDEMIR, M., IRWIN, M. J., AND TUAN, T. A dual-v dd low power fpga architecture. In *Field Programmable Logic and Application*. Springer, 2004, pp. 145–157.
- [10] HOSSEINI, S., SABERI, M., AND LOTFI, R. A low-power subthreshold to above-threshold voltage level shifter. *Circuits and Systems II: Express Briefs, IEEE Transactions on* 61, 10 (Oct 2014), 753–757.
- [11] HSIAO, K.-J. 17.7 a 1.89nw/0.15v self-charged xo for real-time clock generation. In *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014 IEEE International* (Feb 2014), pp. 298–299.
- [12] HUDA, S., ANDERSON, J., AND TAMURA, H. Charge recycling for power reduction in fpga interconnect. In *Field Programmable Logic and Applications (FPL), 2013 23rd International Conference on* (2013), IEEE, pp. 1–8.
- [13] ISHIHARA, S., HARIYAMA, M., AND KAMEYAMA, M. A low-power fpga based on autonomous fine-grain power gating. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on* 19, 8 (2011), 1394–1406.
- [14] KLINEFELTER, A., ROBERTS, N. E., SHAKHSHEER, Y., GONZALEZ, P., SHRIVASTAVA, A., ROY, A., CRAIG, K., FAISAL, M., BOLEY, J., OH, S., ET AL. 21.3 a 6.45 μ w self-powered iot soc with integrated energy-harvesting power management and ulp asymmetric radios. In *Solid-State Circuits Conference-(ISSCC), 2015 IEEE International* (2015), IEEE, pp. 1–3.
- [15] KULKARNI, J., KIM, K., AND ROY, K. A 160 mv, fully differential, robust schmitt trigger based sub-threshold sram. In *Low Power Electronics and Design (ISLPED), 2007 ACM/IEEE International Symposium on* (Aug 2007), pp. 171–176.
- [16] KUON, I., AND ROSE, J. Measuring the gap between fpgas and asics. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on* 26, 2 (Feb 2007), 203–215.
- [17] LEE, K.-J., PARK, H., KONG, J., AND CHANDRAKASAN, A. P. Demonstration of a subthreshold fpga using monolithically integrated graphene interconnects. *Electron Devices, IEEE Transactions on* 60, 1 (2013), 383–390.
- [18] LEMIEUX, G., LEE, E., TOM, M., AND YU, A. Directional and single-driver wires in fpga interconnect. In *Field-Programmable Technology, 2004. Proceedings. 2004 IEEE International Conference on* (2004), IEEE, pp. 41–48.

- [19] LEWIS, D., BETZ, V., JEFFERSON, D., LEE, A., LANE, C., LEVENTIS, P., MARQUARDT, S., MCCLINTOCK, C., PEDERSEN, B., POWELL, G., ET AL. The stratix π routing and logic architecture. In *Proceedings of the 2003 ACM/SIGDA eleventh international symposium on Field programmable gate arrays* (2003), ACM, pp. 12–20.
- [20] LI, F., CHEN, D., HE, L., AND CONG, J. Architecture evaluation for power-efficient fpgas. In *Proceedings of the 2003 ACM/SIGDA eleventh international symposium on Field programmable gate arrays* (2003), ACM, pp. 175–184.
- [21] LI, F., LIN, Y., AND HE, L. Vdd programmability to reduce fpga interconnect power. In *Proceedings of the 2004 IEEE/ACM International conference on Computer-aided design* (2004), IEEE Computer Society, pp. 760–765.
- [22] LI, F., LIN, Y., HE, L., AND CONG, J. Low-power fpga using pre-defined dual-vdd/dual-vt fabrics. In *Proceedings of the 2004 ACM/SIGDA 12th international symposium on Field programmable gate arrays* (2004), ACM, pp. 42–50.
- [23] LUO, S.-C., HUANG, C.-R., AND CHIOU, L.-Y. Minimum convertible voltage analysis for ratioless and robust subthreshold level conversion. In *Circuits and Systems (ISCAS), 2012 IEEE International Symposium on* (May 2012), pp. 2553–2556.
- [24] POON, K. K., YAN, A., AND WILTON, S. J. A flexible power model for fpgas. In *Field-Programmable Logic and Applications: Reconfigurable Computing Is Going Mainstream*. Springer, 2002, pp. 312–321.
- [25] RAHMAN, A., AND POLAVARAPUV, V. Evaluation of low-leakage design techniques for field programmable gate arrays. In *Proceedings of the 2004 ACM/SIGDA 12th international symposium on Field programmable gate arrays* (2004), ACM, pp. 23–30.
- [26] RYAN, J. F., AND CALHOUN, B. H. A sub-threshold fpga with low-swing dual-vdd interconnect in 90nm cmos. In *CICC* (2010), pp. 1–4.
- [27] SHANG, L., KAVIANI, A., AND BATHALA, K. Dynamic power consumption in the virtex-ii fpga family. In *ACM/SIGDA International Symposium on Field Programmable Gate Arrays*, pp. 157–164.
- [28] SINGH, P., AND VISHVAKARMA, S. K. Device/circuit/architectural techniques for ultra-low power fpga design. *Microelectronics and Solid State Electronics* 2, A (2013), 1–15.
- [29] SREENIVAAS, V., PRASAD, D., KAMALANATHAN, M., KUMAR, V., GAYATHRI, S., AND NANDINI, M. A novel dynamic voltage scaling technique for low-power fpga systems. In *Signal Processing and Communications (SPCOM), 2010 International Conference on* (2010), IEEE, pp. 1–5.
- [30] TUAN, T., AND LAI, B. Leakage power analysis of a 90nm fpga. In *Custom Integrated Circuits Conference, 2003. Proceedings of the IEEE 2003* (2003), IEEE, pp. 57–60.
- [31] WOOTERS, S., CALHOUN, B., AND BLALOCK, T. An energy-efficient subthreshold level converter in 130-nm cmos. *Circuits and Systems II: Express Briefs, IEEE Transactions on* 57, 4 (April 2010), 290–294.
- [32] YANG, S. *Logic synthesis and optimization benchmarks user guide: version 3.0*. Microelectronics Center of North Carolina (MCNC), 1991.

Publications

- [1] **Yu Huang**, Aatmesh Shrivastava, Benton H. Calhoun. “**A 145mV to 1.2V single ended level converter circuit for ultra-low power low voltage ICs.**” In *S3S Conference*. *Accepted*
- [2] He Qi, Oluseyi Ayorinde, **Yu Huang**, Benton H. Calhoun. “**Optimizing energy efficient low-swing interconnect for sub-threshold FPGAs.**” In *Field-programmable Logic and Applications (FPL)*. *Accepted*
- [3] Oluseyi Ayorinde, He Qi, **Yu Huang**, Benton H. Calhoun. “**Using island-style bi-directional intra-CLB routing in low-power FPGAs.**” In *Field-programmable Logic and Applications (FPL)*. *Accepted*