Optimizing Energy Efficient Low-Swing Interconnect for Sub-threshold FPGAs

He Qi
Department of Electrical and Computer Engineering
University of Virginia
Charlottesville, VA
hq5tj@virginia.edu

Oluseyi Ayorinde
Department of Electrical and Computer Engineering
University of Virginia
Charlottesville, VA
oaa4bj@virginia.edu

Yu Huang
Department of Electrical and Computer Engineering
University of Virginia
Charlottesville, VA
yh3cf@virginia.edu

Benton Calhoun
Department of Electrical and Computer Engineering
University of Virginia
Charlottesville, VA
bcalhoun@virginia.edu

***Abstract*—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. This paper presents an optimized low-swing interconnect for FPGAs operating in the sub-threshold region. 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 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, un-optimized low-swing design at V_{DD} = 0.4V. Furthermore, the simulation results show that the optimized low-swing interconnect is 97.7% faster and 42.7% lower energy than a traditional uni-directional interconnect at V_{DD} = 0.4V.**

Keywords— Sub-threshold Interconnect; Energy Efficiency; Circuit Optimization; Voltage Optimization

I. INTRODUCTION

Low energy miniature systems for ubiquitous computing such as wireless sensor network have been developing rapidly in the past decade. Existing hardware solutions for ubiquitous computing include ultra-low-power (ULP) ASICs and ULP microprocessors. 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 are too power-consuming. Sub-threshold FPGAs, flexible and not high in power cost, become a highly desired solution. However, an FPGA design implementation consumes 7X ~ 14X more power than a functionally equivalent ASIC design [1], 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 the majority of power is dissipated in the interconnection fabric [2]. Researchers reduce power of the FPGA interconnect in different ways, such as introducing multiple power mode in [3], dual-V_{DD} structure in [4], and multi-V_T scheme in [5]. 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 routing fabrics shown in [6] are 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 [7], researchers reduced both delay and energy by implementing a new low-swing interconnect fabric operating in sub-threshold, where the supply voltage V_{DD} is less than the threshold voltage V_T of a single transistor. 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 supply voltage 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 switching threshold (V_M) repeaters into routing paths. A test chip was fabricated to compare different circuits for the low-swing design. The measured data showed the best circuit options are 61.7% faster and 60.2% lower in EDP than a first-pass, un-optimized design at V_{DD} = 0.4V for a 40-switch path.

II. CIRCUIT MODEL OF GLOBAL INTERCONNECT

A. 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 V_{DD} in sub-threshold. The low-swing interconnect design in [7] replaces the multiplexers and buffers structure with pass-gates (PGs). Its basic structure is shown Fig. 1. 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 to $C*V_{DD}*\Delta V$, where C denotes the total lumped capacitance

along the path and ΔV is the signal swing, reducing signal swing reduces energy effectively. Furthermore, the low-V_M sense amp (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 boosted voltage V_{DDC} is also used to control the gate voltage of switches. Increasing V_{DDC} can reduce delay with small energy penalty.

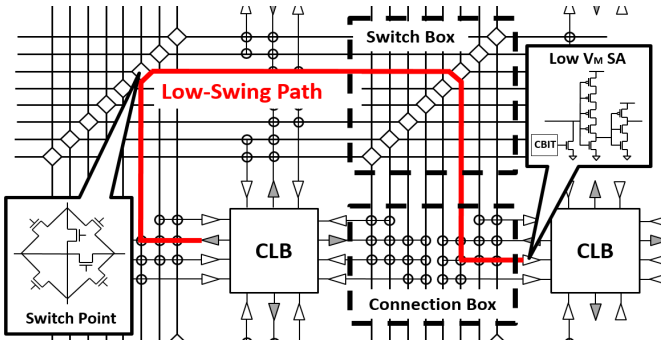

Fig. 1. Basic structure of low-swing interconnect

Fig. 2 shows the diagram of the global interconnect model used in this work. We use the SA from [7] 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 config 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.

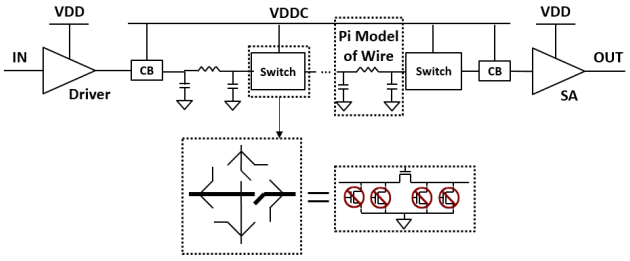

Fig. 2. Diagram of the global interconnect path model

B. *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 [8] tool set on the MCNC benchmarks [9] to investigate the path distribution of the global interconnect. An Altera Stratix IV architecture [10], 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 length distribution 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.

III. Interconnect Circuit Optimization

A. *Optimal V_{DD} and V_{DDC}*

Supply voltage V_{DD} is a dominant knob for EDP. Path delay decreases exponentially in the sub-threshold region at lower V_{DD}. Energy is lower in the sub-threshold region and is dominated by leakage energy, while active energy dominates total energy for above-threshold operation [11]. 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}. Besides V_{DD}, energy and delay also depend on V_{DDC}. 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 detailed measured data will be shown later in this paper.

B. *Signal Degradation & Repeater Insertion*

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. 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 if without repeaters to regenerate the signal.

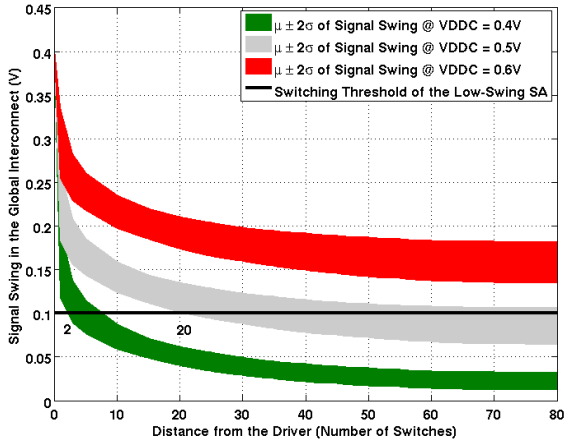

Fig. 3. Range of signal swing for varying path length from Monte Carlo (MC) simulations with PG interconnect compared to the V_M of SA @ V_{DD} = 0.4V

Fig. 3 shows the signal swing change after passing through different numbers of switches at V_{DD} = 0.4V. 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 μ ± 2σ 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 the repeater in this work is the same as a low-V_M SA. If considering variation, a repeater is needed after every 2 (20) switches at V_{DDC} = 0.4V (0.5V). When V_{DDC} ≧ 0.6V, no repeaters are needed to maintain functionality of a path shorter than 80.

C. Connection Box (CB) Topology Optimization

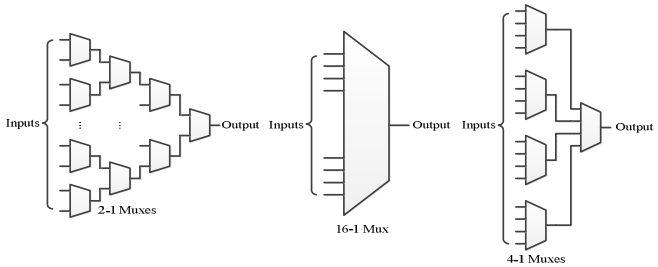

Fig. 4. Schematic of different CB topologies: (a) full multiplexer (b) 1-stage multiplexer (c) 2-stage multiplexer

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.

Fig. 4 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 Fig. 5 (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 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 with considering the area of SRAM bitcells). Considering energy efficiency and variation, the 2-stage design is optimal.

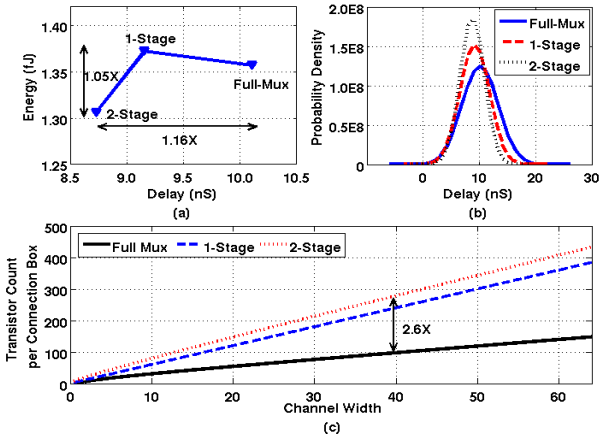

Fig. 5. Comparison of different CB topologies (a) ED curve @ V_{DD} = 0.4V (b) variation @ V_{DD} = 0.4V (c) area

D. Switch & 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. In this work, we sweep driver size from 1X to 20X and switch size from 1X to 8X. The detailed measured data will be shown later in this paper.

IV. Comparison of Designs

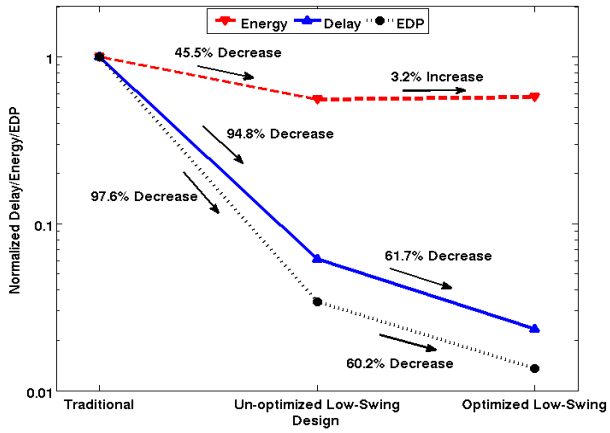

Fig. 6. Comparison of the normalized delay, energy, and EDP @ VDD=0.4V

The simulation results of the traditional uni-directional interconnect, un-optimized low-swing design (minimum size for all transistors, V_{DDC} = V_{DD} + 0.2V), and optimized design (10X for drivers, 4X for routing switches, V_{DDC} = V_{DD} + 0.2V) are compared in Fig. 6. The optimized design has 61.7% smaller delay, 60.2% lower EDP, and 3.2% higher energy than

the un-optimized 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 at V_{DD} = 0.4V.

V. Test Chip & Measurement Results

We have implemented eight 10-by-10 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.

At V_{DD} = 0.4V, the measured results shows that the swing of signals degrades along the path. The SA can successfully capture 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. Furthermore, the measured results shows that inserting repeaters increases both delay and energy of all paths in the silicon.

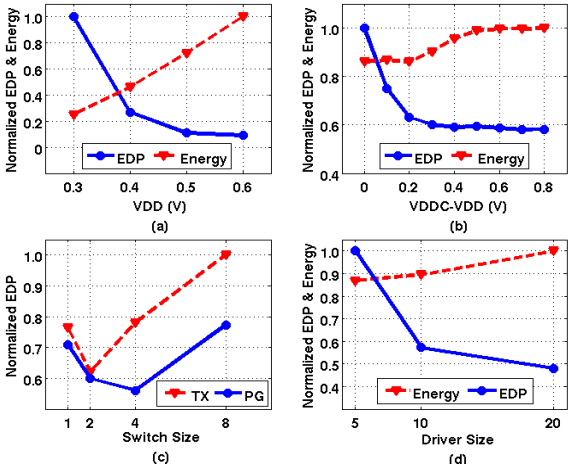

Fig. 7. Measured path with length 40 for (a) V_{DD} optimization (b) V_{DDC} optimization @ V_{DD} = 0.4V (c) switch size optimization @ V_{DD} = 0.4V (d) driver size optimization @ V_{DD} = 0.4V

As shown in Fig. 7 (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%. Fig. 7 (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 V_{DDC} further cannot reduce EDP, but can increase the energy by 15%. In Fig. 7 (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. Fig. 7 (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.

VI. Conclusion

In this work, we presented an optimized low-swing interconnect for FPGAs operating in the sub-threshold region. A test chip in 130nm CMOS is fabricated. 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 (VDD = 0.4/0.5V and VDDC-VDD= 0.2V). 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 at V_{DD} = 0.4V.

Acknowledgment

This work was partially sponsored by Defense Advanced Research Projects Agency (DARPA) under agreement [HR0011-13-2-0004]. The content, views and conclusions presented in this document do not necessarily reflect the position or the policy of DARPA or the U.S. Government, no official endorsement should be inferred.

References

[1] I. Kuon and J. Rose, “Measuring the gap between FPGAs and ASICs,” IEEE Trans. On CAD, vol. 26, no. 2, pp. 203–215, Feb. 2007.

[2] Betz, Vaughn, Jonathan Rose, and Alexander Marquardt. Architecture and CAD for deep-submicron FPGAs. Vol. 497. Springer Science & Business Media, 2012.

[3] Anderson, Jason Helge, and Farid N. Najm. "Low-power programmable FPGA routing circuitry." Very Large Scale Integration (VLSI) Systems, IEEE Transactions on 17, no. 8 (2009): 1048-1060.

[4] A. Gayasen, K. Lee, N. Vijaykrishnan, M. Kandemir, M. Irwin, and T. Tuan, “A dual-Vdd low power FPGA architecture,” in Proc. Int. Conf. Field-Programmable Logic Appl., Antwerp, Belgium, 2004, pp. 145–157.

[5] Ciccarelli, Luca, Andrea Lodi, and Roberto Canegallo. "Low leakage circuit design for FPGAs." In Custom Integrated Circuits Conference, 2004. Proceedings of the IEEE 2004, pp. 715-718. IEEE, 2004.

[6] Lemieux, Guy, Edmund Lee, Marvin Tom, and Anthony Yu. "Directional and single-driver wires in FPGA interconnect." In Field-Programmable Technology, 2004. Proceedings. 2004 IEEE International Conference on, pp. 41-48. IEEE, 2004.

[7] Ryan, Joseph F., and Benton H. Calhoun. "A sub-threshold FPGA with low-swing dual-VDD interconnect in 90nm CMOS." In *CICC*, pp. 1-4. 2010.

[8] Betz, Vaughn, and Jonathan Rose. "VPR: A new packing, placement and routing tool for FPGA research." In *Field-Programmable Logic and Applications*, pp. 213-222. Springer Berlin Heidelberg, 1997.

[9] S. Yang, "Logic Synthesis and Optimization Benchmarks, Version 3.0," Tech. Report, Microelectronics Center of North Carolina, 1991.

[10] Stratix IV Device Handbook, available at www.altera.com.

[11] Calhoun, Benton H., Alice Wang, and Anantha Chandrakasan. "Modeling and sizing for minimum energy operation in subthreshold circuits." *Solid-State Circuits, IEEE Journal of* 40, no. 9 (2005): 1778-1786.