An Energy and Power Consumption Analysis of FPGA Routing Architectures

Peter Jamieson,
Elec. and Comp. Eng.,
Miami University

Wayne Luk,
Dep. of Computing,
Imperial College

Steve J.E. Wilton,
Elec. and Comp. Eng.,
University of British Columbia

George A. Constantinides,
Dep. of EEE,
Imperial College

Abstract—In this work, we evaluate bi-directional and uni-directional FPGA routing architectures in terms of energy and power consumption using an updated power estimation framework compatible with VPR 5.0. The goal of this research is to help FPGA vendors find the best FPGA architectures. Initially, we make some general observations on how two types of routing architectures affect speed, area consumption, and power consumption. We observe how routing buffer sizing affects both the critical path delay and power and energy consumption of FPGAs with certain routing architectures. Our results show that uni-directional routing architecture, in all but one case, is the most energy efficient choice both in the traditional FPGA domain and the mobile domain where clock frequencies are fixed.

I. INTRODUCTION

Power and energy consumption are major concerns for FPGA users and vendors. High power and energy leads to decreased battery life and increased costs for packaging and cooling. This is especially important if FPGAs are used in the hand-held mobile devices.

Although it is challenging to use reconfigurable devices in the mobile domain due to the high power consumption, both industry and academia have suggested this possibility. For example, FPGA vendors such as Actel [1] and Silicon-Blue [13] have designed FPGAs to target mass produced hand-held computing devices. Also, researchers have demonstrated improved energy efficiency of reconfigurable architectures in some mobile applications [14]. Nokia, in a joint effort with academia, has created a benchmark for evaluating recon-figurable architectures in the mobile domain - GroundHog 2009 [6]. These activities suggest that low-power FPGAs may become a viable technological solution within this market.

Regardless of the target domain, FPGAs have not been anal-ysed in terms of what impact different routing architectures has on power and energy consumption. This is due to the lack of infrastructure to perform such experiments.

In this work, our goal is to analyse modern FPGA routing architectures to help FPGA vendors build the best possible FPGAs. To do this, we compare two FPGA routing architec-tures focusing on the following question: What is the impact of routing architecture on the power and energy consumption of an FPGA when targeting both the mobile domain and traditional domain?

The contributions of this paper are as follows:

1) We show that the uni-directional routing architecture is the best choice for power and energy consumption compared to bi-directional routing for both the mobile and traditional domains in most cases. For slower clock frequencies (in the KHz to 10MHz range), we show that bi-directional routing may be better.

2) To perform this study, we have updated the power estimation framework created for VPR 4.3 by Poon *et. al.* [11], for VPR 5.0 [10] to estimate power and energy consumption of both uni-directional and bi-directional routing.

The remainder of this paper is organised as follows: Sec-tion II provides a basic description of FPGA architectures, and we review research on FPGA power estimation and optimisa-tion. Section III looks at some of the details of the routing architectures and their impact on speed, area consumption, and power consumption. Section IV describes the changes made to the power estimation framework to work in VPR 5.0. Section V explains our experimental setup and shows results for analysing bi-directional and uni-directional FPGA architectures in terms of power and energy consumption, and finally, Section VI concludes the paper.

II. FPGA BACKGROUND

FPGAs are programmable chips that can implement a variety of digital designs. An FPGA architecture has many parameters that define its connectivity and makeup, including the number of Basic Logic Elements (BLEs) per cluster (N), the input size of a LUT (K) in a BLE, the number of routing tracks per channel (W), the input connectivity to the BLEs in a soft logic cluster (F_{Cin}), the output connectivity from the BLEs to the routing tracks (F_{Cout}), logical wire length in terms of the number of clusters spanned (L), and the switch block flexibility connecting routing tracks with each other (F_s).

Programmable routing architectures that connects pro-grammable blocks together are called bi-directional or uni-directional (or a mixture of both). In bi-directional routing, the output of the cluster tile is connected via a buffer, and alternatively, in uni-directional routing the output is connected to a multiplexer with other wires (from both cluster outputs and other wires in the channel), and one buffer then drives the output of this multiplexer. We will further discuss the differences of the two routing architectures in Section III.

A. Low-Power FPGA Research

Power consumption is a combination of static and dynamic power. Static power is caused by leakage currents inside transistors, and dynamic power is caused by switching activity

by the charging and discharging of load capacitance, C, as well as short-circuit currents when transistors switch.

$$P_{dynamic} = \alpha \cdot C \cdot V^2 \cdot f \qquad (1)$$

$$P_{total} = P_{static}(T) + P_{dynamic}(f) \qquad (2)$$

Dynamic power is given by equation (1). It has linear dependency on the clock frequency f and a quadratic dependency on the supply voltage V. In an FPGA, the load capacitance depends on the number of logic and routing elements used in a design. The factor α is the activity or toggle rate of an element and is dependent on the design and its input stimuli. Total power is defined in equation (2) and is a combination of static power (which is dependent on temperature T) and dynamic power.

There have been a number of efforts made in estimating FPGA power consumption. Shang *et. al.* [12] show the dynamic power consumption of a Xilinx Virtex-II FPGA using an internal simulated benchmark and estimations of the FPGAs capacitance and connectivity. This internal benchmark includes input stimuli, which they use to calculate the switching activity of their design. Anderson *et. al.* [3] have created a methodology to predict pre-layout switching activity of a design mapped to an FPGA. These estimations can be used in CAD algorithms to minimize activity on inter-cluster programmable routing. Poon *et. al.* [11] create a power estimation framework within VPR and use their power models to evaluate a variety FPGA architectures.

Reducing FPGA power consumption is an important goal when designing commercial FPGAs, including Altera's StratixIII FPGA [2] and Xilinx's Virtex 5 [15] architectures. Only a few vendors are specifically targeting their devices to the low-power mobile domain. Actel's Igloo FPGAs [1] and SiliconBlue's iCE FPGAs [13] are low cost FPGAs (approximately 1 to 2 USD) that are the leading edge in low-power FPGAs targeting a handheld market.

III. OBSERVATIONS FOR UNI-DIRECTIONAL AND BI-DIRECTIONAL ROUTING

The goal of this work is to investigate the power consumption of FPGAs for both uni-directional and bi-directional routing. In this section, we review the details of both routing architectures and make some observations to better understand the differences between the routing architectures and how these differences impact the speed, area, and power of a design mapped to an FPGA.

The basic structure of the two routing architectures is best understood diagrammatically. Figure 1(a) shows the output of a cluster and a switch box using bi-directional routing, and Figure 1(b) shows the same connectivity, but using uni-directional routing.

Based on the structure of the routing architectures, we have listed some basic observations in Table I. Column one shows the number of the observation so we can reference it, column two describes the observation, column three shows the metrics that the observation relates to, and column four shows the routing architecture type that benefits based on the observation.
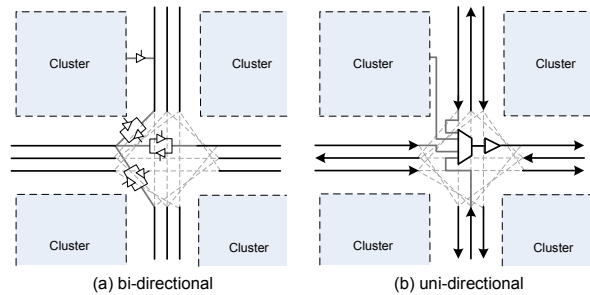

Fig. 1. Bi-directional (a) and uni-directional (b) basic structure

In terms of area, uni-directional routing consumes less area than bi-directional routing (observation 1 in Table I). This area reduction is due to buffer sharing facilitated by the multiplexer/driver routing switch. The overall effect of this observation is shown by Lemieux *et. al.* [9] and Lewis *et. al.* [5]. In both works, they show that the channel width (W) increases for uni-directional routing architectures because of reduced flexibility (due to directionality), but because of buffer sharing the FPGA area consumption is reduced.

In Table I, observation 2 says that the output from the cluster in uni-directional routing architectures connects directly to the switch block, and this removes delay and capacitance due to the lack of this buffer, which is present in bi-directional routing. Similarly, uni-directional routing architectures have routing nets that connect to fewer switching points compared to bi-directional architectures resulting in less capacitive load on a routed path (Table I, observation 3).

Observation 4, in Table I indicates that a routing path in bi-directional routing, which goes through at least one switch box, has fewer programmable switches compared to uni-directional routing. The reason for this is that bi-directional routing is built of tri-state buffers only where as uni-directional routing uses multiplexers with buffers. These multiplexers add additional transistors to a routed path. This fact is important when we consider power consumption, and more specifically, dynamic power consumption. The reason for this is dynamic power consumption is due to switching on *used* routing paths. So, even though bi-directional architectures have more programmable switches, the dynamic power consumption is a function of *used* programmable switches.

This is a concept to remember for FPGA power consumption. A statement such as, power correlates to area used (normally held by ASIC designers), should be modified when dealing with FPGAs to state, static power correlates to FPGA area, and dynamic power correlates to *used* FPGA area.

IV. POWER ESTIMATION IN VPR 5.0

To analyse the two routing architectures for power consumption we need a power estimation framework. Our work merges Poon's estimation framework (built for VPR 4.3) with the recently released VPR 5.0 [10].

Poon's power estimation framework first estimates switching activity on the of a clustered design on the connecting routing nets (wires making up a connection between one output and a number of inputs). Next, these activation estimates,

TABLE I
OBSERVATIONS FOR THE ROUTING ARCHITECTURES AND HOW THEY AFFECT SPEED, AREA, AND POWER CONSUMPTION

	Observation	Metrics Affected	Benefits
1	Uni-directional routing consumes less area than bi-directional routing due to buffer sharing.	Area	Uni
2	Uni-directional routing has one fewer buffer at the outputs of a cluster resulting in lower capacitance for this part of a routing path and less area.	Area, Speed, Power	Uni
3	Uni-directional routing has programmable routing nets that connect to fewer switching points resulting in less capacitive load on the net.	Speed, Power	Uni
4	Bi-directional routing paths, going through at least one switch box, consist of less programmable switches resulting in lower capacitance on these programmably routed paths.	Speed, Power	Bi

the FPGA architecture, and the placed and routed design are used to estimate the dynamic and static power consumption of that design by estimating power consumed for the transistors and the switching activity on paths in the architecture. We have updated VPR 5.0 to include this framework.

V. ENERGY AND POWER CONSUMPTION OF ROUTING ARCHITECTURES

In this section, we show the effect of transistor sizing on power and energy consumption focusing on which architecture is best in a given domain.

A. Experimental Setup

The architectures have routing parameters F_{cin}, F_{cout}, and F_s selected based on the ranges given in previous work by Lemieux *et. al.* [8]. For the rest of the parameters in the architecture we use the files created by Kuon [7]. In particular, we used the architecture file for an $N = 10$, $K = 4$, and transistors sized based on AreaDelay for 180nm technology. This method is reasonable given that there is no open-source tool to size the architecture for the full range of parameters.

For each experiment, we will evaluate the two architectures in terms of critical path delay, area consumption, and power consumption. Our benchmark set includes the 20 largest MCNC benchmarks [16], which are mapped to FPGAs in VPR 5.0 using a typical CAD flow. The activity estimation of nets in the FPGA is done using ACE 1.0, a tool developed by Poon *et. al.* [11]. This activation estimation tool uses transition density models, assuming all primary inputs have a transition density of 0.5.

For each benchmark and for each given architecture, we map the benchmark to the architecture and find the minimum array size and channel width (W). The channel width is then increased by 20% to alleviate routing pressure. The critical path delay, area consumption, and power consumption results are geometrically averaged for all 20 benchmarks.

B. Impact of Transistor Sizing

Our experiment focuses on speed and power consumption of the routing architectures as the transistors in the routing are increased in size. For these experiments, we keep the LUT input size (K) to 5 and the cluster size (N) to 10 and only change the size of the routing buffers. In the case of the uni-directional routing we keep the multiplexer transistors sized to minimum width. Our model for the capacitance and resistance of the routing switches is a simple model that linearly scales in terms of delay and capacitance depending on size.

In these experiments, we will seek the best architecture for both the traditional and mobile domain. One of the distinguishing features of mobile devices is that there are defined clocks available within the device. Normally, two clocks are present. One clock is the main clock and is commonly set to 100MHz or 33MHz, and a second clock, commonly called the heartbeat, is set to a much lower frequency (in the kHz range) and is used to operate circuits that wake up and control the device in low-power states. In these experiments, we will consider clock frequencies of 100MHz, 33MHz, and 150kHz as well as maximum operating frequency.
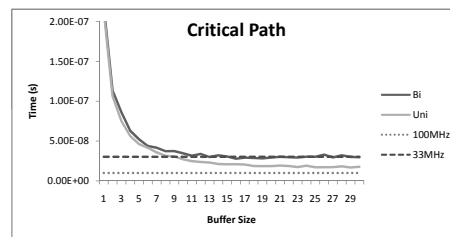

Fig. 2. Critical path delay as buffer size increases

Figure 2 shows the critical path delay of the the two routing architectures as routing buffer size increases from 1x to 30x. Also, in this figure we plot dashed 33MHz and 100MHz lines. If the critical path delay is below these lines then the architecture meets the timing requirement for the respective frequency. For buffer sizes of 1x and 2x, the bi-directional routing architecture is about equal in terms of speed. However, as buffer size increases and the capacitance on the routing nets increases (observation 2 in Table I) uni-directional routing has shorter critical path delays and is the faster choice.

Figure 3 shows energy consumption to maximum operating frequency plot for each architecture. The architectures with the best frequency or energy consumption have points outlined in black, and these points sit on the lower (better energy consumption) right (better frequency) for the two routing architecture types. As described above, we can see that bi-directional architectures are slightly better for very low operating frequencies, but very quickly uni-directional routing dominates in terms of both energy consumption and speed.

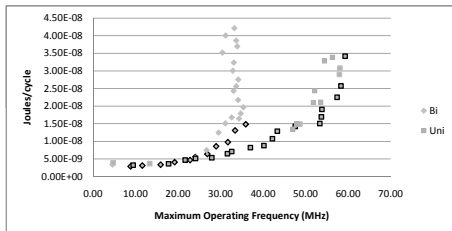Bi-directional routing is in some cases the better energy

Fig. 3. Scatter plot of the architectures in terms of Energy vs. Operating Frequency

consumer. For example, for a 150kHz clock (and up to a 10MHz block), both architectures meet the clocking requirements when the buffers are sized to 1x. In this case, bi-directional routing consumes 15% less energy at an increased cost of 63% in area compared to uni-directional routing. Bi-directional architectures low-power consumption at these speed requirements may be useful in not only the mobile domain's heartbeat clock, but in applications such as sensor networks ([4]) where low-power devices that can operate in the kHz to 10 MHz range are used.

TABLE II
POINTS WHERE A NUMBER OF BENCHMARKS MEET A 33MHZ CLOCK REQUIREMENT

Routing Architecture	Buffer Size	Geometric avg. Critical Path (seconds)	Benchmarks meet 33MHz	Energy (Joules/cycle)
Bi	10x	2.90E-08	16	6.09E-09
Uni	8x	2.81E-08	14	5.13E-09
Bi	25x	2.45E-08	18	2.03E-08
Uni	10x	2.39E-08	18	5.88E-09

If, however, a 33MHz clock is the clock requirement Table II shows details for two threshold points for each routing architecture satisfying this speed requirement; the first comparison point occurs when the respective architecture's routing buffer size results in the geometrically averaged critical path delay meeting the 33MHz threshold, and the second comparison point is for when the respective architecture's routing buffer size results in 18 of 20 benchmarks meeting the 33MHz requirement. Columns 2 through 5 of Table II show the routing buffer size, geometrically averaged critical path delay, the number of benchmarks meeting the threshold, and the energy consumption per clock cycle. In both cases, the energy consumption for uni-directional routing is better than bi-directional. In addition to the energy benefit, the uni-directional architectures at these points consume less area.

In the traditional FPGA domain, an application is mapped to an FPGA such that it can be clocked as fast as possible or at a very high frequency (compared to the clock frequencies of the mobile domain), and it is clear that in this domain uni-directional architecture dominates in terms of speed, area, and energy consumption. However, bi-directional routing architectures may have a use when using heartbeat clocks in the mobile domain. This would be when reconfigurable circuitry is created to handle low-power states.

VI. CONCLUSION

In this work, we have investigated the power and energy consumption of bi-directional and uni-directional routing architectures. We made four observations on how the two routing architectures differ in terms of speed, area, and power consumption.

We investigated the impact of routing buffer size on the two types of routing architectures. Our main goal was to find the best energy consuming architecture depending on the required clock frequency. Our results show that uni-directional routing architectures are superior in all instances except when in the kHz to 10MHz operating frequency range where bi-directional architectures consume less energy at an area cost. This result suggests that with some clever design of the uni-directional routing switch, there may be room for improvement in power and energy consumption of such architectures in all the domains.

VII. ACKNOWLEDGMENT

REFERENCES

[1] Actel. *Igloo Handbook*, Jan 2008.
[2] Altera. *Stratix III Device Handbook*, 2006.
[3] J. Anderson and F. Najm. Power estimation techniques for FPGAs. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 12(10):1015–1027, October 2004.
[4] M. Atia, J. Bowles, D. W. Clarke, M. P. Henry, I. Page, J. Randall, and J. Yang. A self-validating temperature sensor implemented in fpgas. In *FPL*, pages 321–330, 1995.
[5] D. Lewis *et. al.* The Stratix II Logic and Routing Architecture. In *ACM/SIGDA International Symposium on FPGAs*, pages 14–20, Feb 2005.
[6] P. Jamieson, T. Becker, W. Luk, P. Cheung, and T. Rissa. Benchmarking Reconfigurable Architectures in the Mobile Domain. In *Proceedings of the IEEE Symposium on Field-Programmable Custom Computing Machines*, 2009.
[7] I. Kuon and J. Rose. Automated transistor sizing for FPGA architecture exploration. In *DAC '08*, pages 792–795, 2008.
[8] G. Lemieux and D. Lewis. Using Sparse Crossbars within LUT clusters. In *ACM/SIGDA International Symposium on FPGAs*, pages 59–68, Feb 2001.
[9] G. Lemieux and D. Lewis. Directional and Single-Driver Wires in FPGA Interconnect. In *IEEE International Conference on Field-Programmable Technology*, pages 41–48, Dec 2004.
[10] J. Luu, I. Kuon, P. Jamieson, T. Campbell, A. Ye, W. M. Fang, and J. Rose. VPR 5.0: FPGA CAD and Architecture xploration Tools with ngle-Driver Routing, Heterogeneity and Process Scaling. In *ACM/SIGDA International Symposium on FPGAs*, Feb 2009.
[11] K. K. W. Poon, S. J. E. Wilton, and A. Yan. A detailed power model for field-programmable gate arrays. *ACM Trans. Des. Autom. Electron. Syst.*, 10(2):279–302, 2005.
[12] L. Shang, A. S. Kaviani, and K. Bathala. Dynamic power consumption in Virtex-II FPGA family. In *FPGA '02: Proceedings of the 2002 ACM/SIGDA tenth international symposium on Field-programmable gate arrays*, pages 157–164, 2002.
[13] SiliconBlue. *iCE DiCE: iCE65L04 Ultra Low-Power FPGA Known Good Die*, Sep 2008.
[14] T. Todman, G. Constantinides, S. Wilton, O. Mencer, W. Luk, and P. Cheung. Reconfigurable computing: architectures and design methods. *IEE Proceedings - Computers and Digital Techniques*, 152(2):193–207, 2005.
[15] Xilinx. *Virtex-5 Family Overview*, June 2006.
[16] S. Yang. Logic Synthesis and Optimization Benchmarks, Version 3.0. Tech. Report. Microelectronics Centre of North Carolina. P.O. Box 12889, Research Triangle Park, NC 27709 USA, 1991.