Design and implementation of clock network for nanometer FPGA

Lei Li(a), Jinmei Lai(b)

1ASIC and System State Key Lab, Dept. of Microelectronics, Fudan University
825 Zhangheng Rd., 201203, Shanghai, China

a) 11110720044@fudan.edu.cn
b) jmlai@fudan.edu.cn

Abstract: This paper is committed to design and implement FPGA clock network with overall considerations of speed and power in circuit design perspective. Mixed structure MUX, programmable delay adjustment unit and power-down bit strategies are presented to optimize its latency, skew and power. This clock network is implemented with 65nm process and applied to own-designed FPGA. Test results indicate 21.7% reduction in latency and 54.5% reduction in skew, compared to counterpart FPGA device, while maintaining lower power consumption.

Keywords: Clock network, FPGA, skew, latency, DAU

Classification: Integrated circuit

References

[1] Friedman and E.G.: Proceedings of the IEEE 89 [5] (2001) 665. DOI:10.1109/5.929649
[2] M. Pedram: ACM TODAES 1 [1] (1996) 3. DOI:10.1145/225871.225877
[3] L. Shang, A. Kaviani and K. Bathala: Proceedings of FPGA '02 (2002) 157. DOI:10.1145/503048.503072
[4] V. Degalahal and T. Tuan: Proceedings of ASP-DAC '05 1 (2005) 657.
[5] W. Shen, Y. Cai, X. Hong, J. Hu and B. Lu: Integration VLSI 41 [3] (2008) 426. DOI:10.1016/j.vlsi.2007.10.004
[6] G. Tellez and M. Sarrafzadeh: IEEE Trans. CAD of IC and System 16 [4] (1997) 333. DOI:10.1109/43.602470
[7] X. Zeng, D. Zhou and W. Li: Proceedings of ISPD ’99 (1999) 36. DOI:10.1145/299996.300015
[8] J. Lamoureux and S. Wilton: Proceedings of FPGA ’06 (2006) 101. DOI:10.1145/1117201.1117216
[9] K. Zhu and D. Wong: Proceedings of DAC ’94 (1994) 232. DOI:10.1145/196244.196362
[10] K. Poon, S. Wilton and A. Yan: ACM TODAES 10 [2] (2005) 279. DOI:10.1145/1059876.1059881
[11] Altera. Stratix V Device Handbook, (2013)
http://www.altera.com/literature/hb/stratix-v/stratix5_handbook.pdf.
[12] Xilinx. Virtex-5 FPGA User Guide, (2007) http://www.xilinx.com/support/documentation/user_guides/ug190.pdf.

[13] D. Garrett, M. Stan and A. Dean: Proceedings of ISLPED '99 (1999) 176. DOI:10.1145/313817.313913

[14] S. Tawfik and V. Kursun: ICECS '07 (2007) 845. DOI:10.1109/ICECS.2007.4511123

[15] V. Chickermane, P. Gallagher, J. Sage, P. Yuan and K. Chakravadhanula: ITC '08 (2008) 1. DOI:10.1109/TEST.2008.4700572

[16] Q. Wang, S. Gupta and J. Anderson: Proceedings of FPGA ’09 (2009) 13. DOI:10.1145/1508128.1508132

[17] Xilinx. Virtex-5 FPGA Datasheet, (2007) http://www.xilinx.com/support/documentation/data_sheets/ds202.pdf.

1 Introduction

Clock network is a dedicated network for distributing multiple clock signals to every logic module in a system, playing a crucial part to the chip’s performance [1, 2, 3, 4]. Hence the design of clock network becomes the utmost issue. Be significantly different from ASIC where the clock tree is custom built by users, clock network in FPGA is usually fixed after chip fabrication and cannot be changed for different user circuits. This raises much more challenges to FPGA clock network design, for the tradeoff between speed, area and power should be taken into consideration. With regard to the general clock network design, recent researches can be classified into several categories. Symmetric structures such as H-tree [1] and X-tree [5] can minimize clock skew even achieve zero-skew clock routing. But the absolutely symmetrical structure is extremely sensitive to the process variation. The other kinds of buffered tree such as irregular trees focused on inserting minimum number of buffers in clock trees with latency and skew constraints [6, 7]. However, too many buffers inserted will bring considerable latency, resulting in reduced speed as well as significant power consumption. Considering these factors, it would be inappropriate for those clock networks to be applied to a scalable FPGA.

There are also some studies on the design of FPGA clock network. The principle of FPGA clock network design is discussed briefly in [8], but no design and implementation details were given. They only describe the parameterized FPGA clock network model and have not taken skew into account, just assuming it can be improved by PLL later. Other open literatures on FPGA clock network have mostly targeted the simplified clock structures [9, 10] which are impractical in actual FPGA. Considering commercial FPGAs, clock networks in Altera FPGAs [11] are based on highly symmetric trees. These structures are sensitive to process variation as mentioned above, but they can achieve ideal zero-skew indeed. Xilinx FPGA clock networks [12] adopt fishbone topology as a basis. Although fishbone network has clock skew in nature, it is less sensitive to symmetry and process variation, which is better for scalable FPGA.

In addition to speed, power optimization is also a hotspot in the domain. Previous researches on low power had established various approaches such as clock gating
[13], buffer insertion and sizing [14], multi-voltage designs [15]. Among these approaches, clock gating used widely in high performance VLSI designs has been implanted to FPGA clock network. However, Xilinx FPGAs make use of clock gating in a limited way [12, 16], where the clock enable is only built on global clock MUXs or buffers [12]. Clock power control in Altera FPGA only available for global clock or regional MUXs [11]. In other words, they are impossible to independently gate each one of clock signals in FPGA.

Based on fishbone topology, this paper presents design and implementation of FPGA clock network with overall considerations of speed and power in circuit design perspective. Series of strategies are presented to optimize latency, skew, and power. Bypassing strategy gives shortcut to evade larger MUXs, and mixed MUXs are employed to reduce latency balancing delay and area. Programmable delay adjustment unit (DAU) is proposed to minimize clock skew by proper configurations. Power-down bit is introduced to improve conventional clock gating technique in FPGA domain, for it can handle each clock individually without affecting other clocks’ normal operation.

The reminder of this paper is organized as follows. Section 2 describes the design details of clock network for scalable FPGA, giving several schemes to minimize latency, skew and power. Section 3 gives physical implementation and test results of our design. Finally, Section 4 concludes this paper.

2 Design of FDP5 clock network

Fig. 1. (a) FDP5 Clock network. (b) Detailed view of each clock region.

Fig. 1(a) shows the simplified diagram of clock network based on fishbone topology. And Fig. 1(b) gives a detailed view of each region in this clock network. This architecture is apt to scale up, which is essential to FPGA’s scalability. Therefore, it is adopted by own-designed Fudan Programmable-5 (FDP5) FPGA. FDP5 is a scalable FPGA based on columns, including a lot of homogeneous columns, such as CLBs, IOs, BRAMs, DSPs, spanning from side to side, and one heterogeneous column with PLLs and Specific IOs (SIO) located centrally. The logic array of FDP5 is 64×30. Clock network in Fig. 1 can be easily integrated to this tiled scalable FDP5 FPGA and expanded easily with reasonable area. Global
lock multiplexers (GCMUX) in network first choose global clock signals (GCLK) from various clock sources such as PLLs, general interconnects and external clocks supplied with SIOs. Those selected GCLKs can also be turned back to the inputs of GCMUX and acted as looping back signals, which provide users with convenience to dynamically switch among more clocks and substantially improve flexibility of clock network. GCLKs are vertically distributed in the trunk. There are 32 GCMUXs as well as 32 corresponding GCLKs in FDP5. The whole FDP5 can be divided into 4 regions, each contains several rows. Considering the left and right sides, there are actually a total of 8 clock regions in FDP5. When GCLKs transmit to each region, they can be transformed to horizontal clock signals (HCLK) in primary branches through HCLK module. HCLK module contains a group of programmable MUXs which couple the GCLK trunk to each primary HCLK branch, providing several HCLK signals to the left and right respectively. Each clock region contains 8 HCLKs and they are tapped into every column and converted to leaf clock signals (LCLK), sending up and down to connect the leaf nodes clock in the top and bottom half of each column. Aside from dedicated clock LCLKs, general interconnects are also added to leaf node clock connection, ensuring the diversity of leaf clocks.

2.1 Optimization of clock latency

Clock latency in FPGA refers to the delay from the clock input port to clock pins of flip-flops inside FPGA. It contains the delay of MUX, drivers as well as the interconnect lines. As the latency caused by lines will be optimized in physical implementation, this section only focuses on MUX design to reduce latency.

(1) Mixed structure MUX

Too many candidate clock signals make the MUX in clock path very large, resulting in large path latency and area. Different MUX implementations have different delay and area overhead. Fig. 2(a) shows two common structures MUX. Taking 4:1 MUX for example, one 4:1 decoded MUX requires 4 SRAM cells for proper operation. One 4:1 encoded MUX needs only 2 SRAM cells but with longer delay path of two pass transistors. 16:1 MUX can be built up with two levels 4:1 MUXs. Simplex decoded and encoded 16:1 MUXs ask for 8 and 4 SRAM cells respectively. On the other hand, 16:1 mixed MUX shown in Fig. 2(b) only requires 6 SRAM cells in total. It has decoded MUXs as the first stage and encoded MUX as the second stage. Different size MUXs have been implemented in decoded, encoded and mixed structures separately. Fig. 2(b) illustrates mixed structure MUXs, such as 24:1, 36:1 MUX. Due to the availability of Hspice models and 65nm CMOS technology library, post-layout simulations are done to evaluate the potentials of various MUX structures, comparing delay and area under different PVT conditions. (P means process corner, including FF, TT, SS, FS, SF. V represents supply voltage, from 0.8V gradually increased to 1.3V. T means temperature, considering the lowest, room and highest temperature are -45℃, 27℃, 125℃ respectively. All the simulations in this paper have been done with Hspice software under different PVT conditions mentioned above, if not otherwise stated.) Fig. 2(c) shows the comprehensive comparison results under typical case
(TT, 1V, 27°C) for simplicity. The area is actual layout area in 65nm CMOS process and the delay is obtained by post-layout simulations. The unit of area-delay product in Fig. 2(c) is µm²·ms. It can be concluded that mixed structure MUX has smallest area-delay product, providing the best delay-area tradeoff compared with other two structures. Thus, mixed structure combining decoded and encoded MUX is employed in our design to reduce latency in a rational way.

(2) Bypassing strategy

Multiple alternative clocks also lead to larger size and more level MUXs in clock paths. Clock signals follow these paths have to pass through larger MUXs and travel longer distances. However, in a complicated application with more than one clocks, clock signals to the critical parts require smaller latency than others. To reduce latency of those particular clocks, bypassing strategy is introduced. Extra 2:1 MUX is added in clock path to bypass larger MUXs, shown in Fig. 2(d). When 2:1 bypassing MUX is set to select B, preferential clock PRE_CLK is permitted to
be connected to GCLK trunk via bypassing 2:1 MUX directly, passing through substantially less MUX circuitry. Otherwise it has to pass through larger MUX resulting in larger latency. Making use of bypassing MUX, preferential clocks can travel through the shortcut path to evade larger MUXs. It can be field configured by users as needed. In a word, bypassing strategy gives preferential treatment to latency reduction while maintaining flexibility of clock network.

2.2 Optimization of clock skew
In FPGA clock network signal reaches to clock pins of different flip-flops through different paths. Clock skew is defined as the latency difference from clock source to clock pins of flip-flops, and the maximum skew can be expressed as below:

$$\Delta = \text{Max}(d_i) - \text{Min}(d_i).$$

Where $d_i$ represents the latency of clock path. Obviously, the skew results from unbalanced paths. Since clock skew may limit system performance and even cause function failure, it is important to reduce clock skew to achieve high performance. In FPGA clock network, one approach is to distribute clock pins appropriately using placement algorithms to minimize skew [9, 16]. However, this paper will solve the skew problem in circuit design perspective.

There is certain skew in FDP5 clock network caused by different clock paths, such as skew between different HCLK branches and internal skew within each HCLK branch and each column. To settle this problem, programmable delay adjustment unit (DAU) shown in Fig. 3(a) is properly inserted into the clock network to minimize those skews. It has field programmable feature and can be reconfigured by users to adjust delay according to different demands.

In each DAU, there are four tuning transistors divided into two separate groups, P1/N1 and P2/N2, controlled by two SRAMs. By turning on or shutting off tuning transistors, the delay can be finely tuned. The delay is minimum when the four tuning transistors are turned on. Shutting off P2/N2 will increase the delay. When P1/N1 are further shut off, the delay becomes larger and the maximum delay occurs when P1/N1/P2/N2 are all shut down. The key parameter of DAU is the adjustment range. According to the post-layout simulation of FDP5 clock network, the maximum skew never exceeds 80ps. Thus the adjustment range is defined as 120ps, with 50% margin added to ensure design robustness. Moreover, to
increase the resolution of adjustment, we carefully decide the size of tuning transistors and use two-level DAUs. Fig. 3(b) illustrates HCLK MUX followed by two-level DAUs. With different configurations of Na<3:0> according to binary code, adjustment range and resolution can achieve 120ps and 7.5ps respectively.

2.3 Optimization of power consumption
Components in clock network burn static power which has little to do with frequency and is typically constant. However, the dynamic power is proportional to clock frequency, mainly caused by MUXs and drivers which drive the LCLKs into each logic module. The conventional clock gating [16] is only used for global clock buffers and impossible to gate each clock signal individually. Therefore, power-down bit named E_SRAMb controlled by SRAM cell is introduced to ameliorate this situation and further reduce power. Compared to clock gating, power-down bits are not only added to global clock buffers but also to each MUX and driver shown in Fig.4. In this manner, each clock signal can be handled independently through configuration of power-down bit directly without affecting other clocks’ operation. Moreover, once MUXs and drivers are shut down, the downstream leaf clocks do not toggle, in turn corresponding components do not dissipate dynamic power. Actually, the power switches as well as control logics and SRAMs will not consume any dynamic power after configuration. But they are indeed bringing about extra static power. Therefore, the power switch and control logic adopt high-Vt and small size transistors to minimize the static power, making sure that the total power can be reduced. Moreover, clock network power results in the next section contain the power dissipated by power switches and control logics.

3 Implementation and test results
3.1 Physical implementation
This clock network is implemented with 65nm process and applied to FDP5 FPGA, shown in Fig.5. FDP5 is designed by full-custom method. The layout area of clock

![Fig. 4. Power-down bit for power reduction.](image)

![Fig. 5. (a) Layout of FDP5. (b) The FDP5 chip.](image)
network is 1.36 mm\(^2\), costing about 4% of the total area. In nanometer technology, interconnect delay becomes comparable with transistor gate delay, even exceeds it. Therefore, clock lines in FDP5 employ the top level metal (TM) to achieve high speed performance, for TM is thicker and with larger space to adjacent lines.

3.2 Test results
To verify the clock network and evaluate its performance, a test platform in Fig. 6(a) is built up, including test software system, an FDP5 test board and other measure equipment such as signal generator, high precision oscilloscope and logic analyzer, current probe, programmable power supply, PC, display screen, etc. The FDP5 test board is placed at the bottom-right corner as can be seen. The test procedure of FDP5 is shown in Fig. 6(b). First, decide the related configuration and generate the bit file by test software. Then download the bit file to FDP5. Finally, load test vectors to the programmable I/Os and observe the output response on the test board, oscilloscope and display screen.

FDP5 clock network has been systematically tested on real chip. According to the size of FDP5, there are a total of 8 clock regions, 32 GCMUXs and 32 GCLKs. And each clock region contains 8 HCLKs and 1536 Leaf clock pins. All clocks resources have been tested through systematical tests. Test patterns covered paths from clock sources, passing through different GCMUXs to reach the corresponding GCLKs, and then entering different clock regions, selecting corresponding HCLKs, finally arriving every leaf clock pin of DFF in FPGA. In this manner, there are a total of 3,145,728 patterns in need to cover all clock resources in FDP5. We adopt traversal test methods to traverse each pattern and make sure that all clocks and their relevant paths as well as components have been tested. All the DFFs give the correct results. Results proved that FDP5 clock network can operate correctly and efficiently. Based on these systematical tests, comprehensive circuits such as color rotation system has successfully implemented on FDP5 shown in Fig. 6(a).

Among these test patterns, the maximum latency of clock network (T\(_{\text{latency, max}}\)) occurs from clock source to the farthest leaf clock pins and minimum latency (T\(_{\text{latency, min}}\)) occurs from source to the nearest clock pins. Firstly, we probe related
signals to observe and measure the delay. Then subtract the delay caused by I/O pads and equipment outside FPGA. Finally obtain the latency from clock source to sink inside FPGA. Therefore, the maximum skew \((T_{\text{skew}})\) between two leaf clock pins which are farthest apart can be derived from the difference between \(T_{\text{latency, max}}\) and \(T_{\text{latency, min}}\).

### Table I. Test results

| Parameter          | Description                        | FDP5       | Equivalent FDP5 (FDP5 x 1.36) | XC5VLX30[17] | Improved ratio |
|--------------------|------------------------------------|------------|-------------------------------|--------------|----------------|
| \(V_{IC}(\text{V})\) | Internal supply voltage            | 1.0        | 1.0                           | 1.0          | ---            |
| \(T_{\text{latency, max}}(\text{ns})\) | Maximum latency of clock network  | 1.96       | 2.67                          | 3.41         | 21.7%          |
| \(T_{\text{latency, min}}(\text{ns})\) | Minimum latency of clock network  | 1.89       | 2.57                          | --           | --             |
| \(T_{\text{skew}}(\text{ps})\) | Maximum skew of clock network      | 70         | 100                           | 220          | 54.5%          |

The test data are shown in Table I and they are measured under 1.0V standard internal supply voltage of 65nm process. FDP5 clock network is able to support 600MHz clock in test which can be comparable with XC5VLX30 FPGA device [17]. Maximum and minimum clock latency in FDP5 clock network is 1.96ns and 1.89ns respectively. Therefore, the maximum skew is 70ps. Results of FDP5 in Table I are directly measured and obtained by test.

As XC5VLX30 FPGA device [17] is similar to FDP5 in terms of 65nm process and system resources, the performance of XC5VLX30 is also given in Table I. Actually, The array size of FDP5 is 64×30 and each logic block contains 16 LUTs (4-inputs LUT), while the array size of XC5VLX30 is 80×30 and each logic block contains 8 LUTs (6-inputs LUT). According to the array size and logic capacity, FDP5 would like to be multiplied by a factor of 1.36 in order to be approximately equivalent to XC5VLX30 device and ensure a rational comparison. Test results of FDP5 have also been multiplied by 1.36 and listed as Equivalent FDP5 in Table I. Results of XC5VLX30 in Table I are obtained from Xilinx Virtex-5 Datasheet [17]. Their values listed are specified for industrial grade and derived from measuring internal test patterns [17]. Its internal supply voltage is also 1.0V in average and the latency value listed has already subtracted the delay of I/O pads and equipment. It is observed that our design equivalently achieved up to 54.5\% improvement for skew and 21.7\% for latency, compared to XC5VLX30 device. In this case, we can say that FDP5 is quite comparable with XC5VLX30 device in performance.

For power test, as the separate clock power cannot be directly obtained by test, we

![Fig. 7. Comparison of total device power (FDP5 vs. XC5VLX30).](image)
directly measured total power dissipated on FDP5 chip by current probe on test platform. A lot of ISCAS’89 sequential benchmark circuits (including s382, s1423, s1488, s15850, s38417, etc.) have been tested under 600MHz clock on FDP5. And their power results are plotted with blue line in Fig. 7. For counterpart comparison, we adopt Xpower tools to analyze the power of XC5VLX30 device. Xpower is a standard power analysis tools provided by Xilinx. Although there are tiny bit of differences from actual test value, it would still be a credible analysis tool for reference. The same benchmark circuits are also implemented on XC5VLX30 and the total device power are plotted with orange line in Fig.7. It is obvious that total power consumption in FDP5 is a bit lower than XC5VLX30. However, test results of total device power may not be a good reflection for the separate clock power (P_{CLK}) dissipated only at clock network. Due to separate clock power cannot be directly measured, we added the post-layout simulation results of FDP5 clock network for the relative comparison.

| Description | Equivalent FDP5 Results of Post-layout simulation | XC5VLX30[17] Results of Xpower Tools |
|-------------|---------------------------------------------------|--------------------------------------|
|             | P_{CLK} (uW/MHz)                                  |                                      |
|             | s382     | s1423    | s1488    | s15850   | s38417   |
| Equivalent FDP5 Results of Post-layout simulation | 421  | 438  | 416  | 456  | 475  |
| XC5VLX30[17] Results of Xpower Tools | 455  | 456  | 430  | 495  | 516  |

For brevity, P_{CLK} of equivalent FDP5 listed in Table II are the results under typical simulation case (TT, 1.0V, 27°C). And they have already contained the power dissipated on control logic and power switches. Xpower tools can also give the separate clock power of XC5VLX30 device and the values are also listed in Table II. Results indicate that the FDP5 clock network is comparable with XC5VLX30 device and FDP5 a reasonable design.

4 Conclusion
In this paper, clock network for scalable FPGA is designed and implemented, achieving optimization of speed and power. Bypassing strategy gives a shortcut to evade larger MUXs, and mixed structure MUXs are employed to reduce latency balancing area and power consumption. Programmable DAU can minimize clock skew by proper configurations, providing a nature low skew network. Power-down bit ameliorates conventional clock gating technique in FPGA domain, for it can not only control all clocks together but also handle each clock individually, without affecting normal operation of other clocks. Finally, test results validate our proposed clock network is well-designed and could support 600MHz clock with 54.5% and 21.7% improvement on skew and latency, compared to commercial FPGA while retaining lower power dissipation.

Acknowledgments
This project is supported by the National High Technology Research and Development (863) Thematic Program of China, No. 2012AA012001.