

A CODEC, Tiles to NoC Router Interface, for Next Generation FPGAs with Embedded NoCs

Alaa Salaheldin¹, Hassan Mostafa^{1,2} and Ahmed M. Soliman¹

¹Electronics and Communications Engineering Department, Cairo University, Giza 12613, Egypt,

²Center for Nanoelectronics and Devices, AUC and Zewail City of Science and Technology, New Cairo 11835, Egypt.

{alaa.salaheldin89@gmail.com, hmoustafa@aucegypt.edu, hmostafa@uwaterloo.ca, asoliman@ieee.org}

Abstract—Nowadays, SoC uses Network on Chip (NoC) to connect its increasing number of building blocks. FPGAs, like SoCs, can use NoC to connect its increasing number of tiles, memories, DSP slices and embedded processors. But one drawback of using NoC is that increasing its router ports will affect the area, power and frequency of the system significantly. For FPGAs to benefit from the NoC approach we have to find a way how to interface with large number of blocks without increasing NoC router ports. In this paper we use a concentrator module or a *Codec* to connect between routers and multiple *Tiles* (FPGA basic building block). Using this codec will reduce the effect of increasing tile count on the area, power and frequency of the FPGA routing network. A 64-tiles network with codec would consume less than 15% area, less than 50% power of a NoC-only network and operates with 2.5X frequency.

I. INTRODUCTION

FPGAs are facing a big challenge, which is the area, power and delay overhead of its routing network, a medium FPGA design includes multiple IPs and modules that can be placed in different locations across the FPGA; short wires are used to connect nodes within the same block or module and long wires are used to connect modules far apart. Increasing the FPGA area and embedded blocks requires more long wires which will introduce large delay and consume more power.

Network on chip (NoC) is widely used in different SoCs, it overcomes the conventional interconnect problems of high power consumption and latency. NoC has the benefits of independent implementation, simplified customization, scalability and support for different network topologies. PNoC [1], SOTA [2] and Split-Merge [3] are NoC router examples.

NoC routers are implemented either as hard or soft, a hard router is more area and power efficient than a soft router but the latter is more flexible and configurable. Design guidelines for soft and hard routers are introduced in [4], [5], tradeoffs between soft and hard routers are discussed in [6], [7].

Using the NoC approach instead of depending totally on long interconnect wires solves the conventional interconnects problems; because NoC uses high speed optimized lanes to transfer packets between the routers, these routers interface with the application blocks through configurable number of input/output ports solving most of the problems introduced by long and medium size routing wires.

By applying the NoC approach in FPGAs, we would have a major problem because of the large number of tiles that needs

to be connected to the network, as shown in [8], NoC routers' area, delay and power consumption are increasing significantly with the input/output port count, in order to overcome this problem and to make the NoC approach useful in designing the next generation of FPGAs, a tile to NoC router (or a codec) can be used to enable multiple tiles to share single router port, given that the required rate for all multiplexed tiles will not exceed the router port's maximum rate.

The rest of the paper is organized as follows, in section II, a brief background of CONNECT router is introduced. Section III discusses Modelling and MATLAB simulation of the codec. Implementation of the codec circuit and different network configurations are shown in Section IV. Section V includes a comparison between CONNECT-only networks (network A) and CONNECT with codec networks (network B). Section VI is an acknowledgement and finally a Conclusion in Section VII.

II. BACKGROUND

CONNECT is a soft router designed for FPGAs [9], it adds new features like virtual links and peak flow control. It maximizes routing resources utilization by using wider buses between routers. CONNECT has configurable number of input/output ports, packet length and buffer depth. It is implemented using Bluespec System Verilog (BSV) and provides a flexible design, Fig. 1 shows its architecture, it uses a single stage pipeline leading to lower cost and latency and it supports four variations of separable input-output allocators.

Few FPGA blocks(memories, embedded processors) make use of CONNECT features like peak flow control, buffering and virtual channels, but most of the other blocks and tiles will not use these features as they only need simple net connection to other nodes, therefore it is better to group tiles as much as we can and connect them to one router port, this will reduce number of required ports per router.

III. MODELLING AND SIMULATION

A. MATLAB modelling

A Simulink system level model is built using SimEvents toolbox to measure the throughput difference between two networks, Network A which includes routers only, and Network B which includes routers and codecs. SimEvents provides a discrete-event simulation engine and component library for

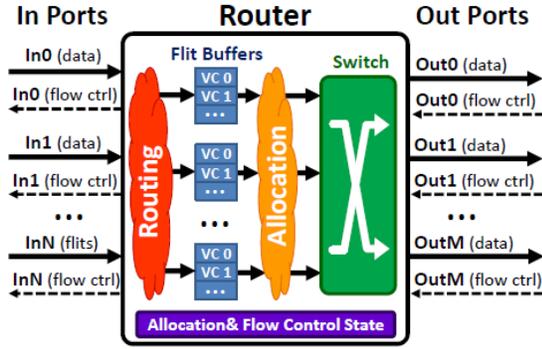


Figure 1: CONNECT router

analyzing event-driven system models and optimizing performance characteristics such as latency, throughput, and packet loss.

Both networks is a 2x2 mesh network that interfaces with sixteen tiles, each router is connected to four tiles; either by direct port connections or through a codec. Network A is without codec, so each router interfaces with four tiles and two neighbor routers (6-port router shown in Fig.2).

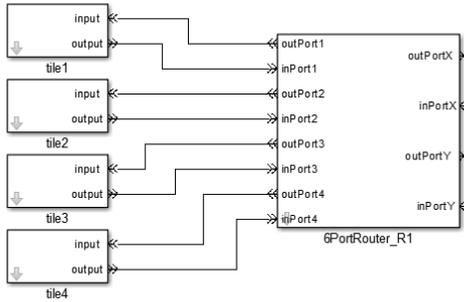


Figure 2: 6-port router in network A

In network B, each router uses a codec to interface with four tiles so the used routers are 3-port routers (shown in Fig. 3). The packet length is increased by two bits in this case to handle the switching required from codec to tiles.

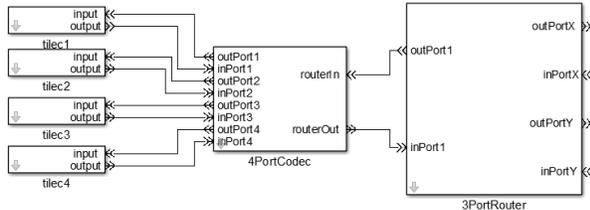


Figure 3: 3-port router and codec in network B

A router consists of routing core and input/output queues, the routing core (shown in Fig. 4) consists of routing logic and output switch, its routing logic is implemented as a delay server to model the packet processing latency and a routing table to determine which output port the packet will go through. We assumed that the packet processing time in a 6-port router is longer than in a 3-port router.

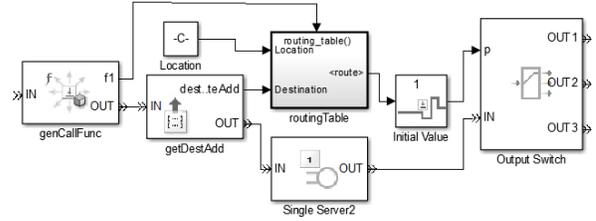


Figure 4: Router core of a 3-port router used in network B

Fig. 5 shows a codec, in the sending path it acts as multiplexer or path combiner and in the receiving path it is similar to a normal router but with shorter server processing time.

B. Simulation results

The two network models were run for the same time given these parameters; port buffer depth is set to four packets, router server time is set to three for 3-port routers, six for 6-port routers and one for codecs. A uniform distribution packet generator is used at each tile output to simulate the existence of FPGA tiles.

After running the simulation, the received packets at each tile is counted, Table I shows the comparison between received packets for some tiles in network A compared to network B and it shows significant increase. The total number of received packets to all tiles is 281 packets in network A and 677 packets in network B.

We found that the throughput is enhanced by using codecs in network B.

Table I: Simulink models comparison

Tile number	1	2	5	6	9	10	13	14	All tiles
Network A	18	19	12	17	21	17	16	18	281
Network B	54	44	43	41	47	39	45	39	677

IV. IMPLEMENTATION AND NETWORK SYNTHESIS

A. Codec implementation

The transmitting path of a codec (from tile to router) acts as a path combiner, it rotates across all attached tiles and checks for available payloads to send.

In the receiving path (Lower part of Fig. 5), it is similar to a router, a packet is examined once received from a router and according to the destination address a tile is chosen to send the packet to. Codec design is made modular to facilitate generating any network configuration with different tile widths and count, it inherits some parameter from CONNECT design in order to be compatible with it.

B. Network configurations

The comparison is done between two network models both has four routers in a 2x2 mesh topology, first model is network A that uses CONNECT routers only, we built three configurations of this model to be interfaced with 16-tiles, 32-tiles and 54-tiles. A 64-tiles network could not be built

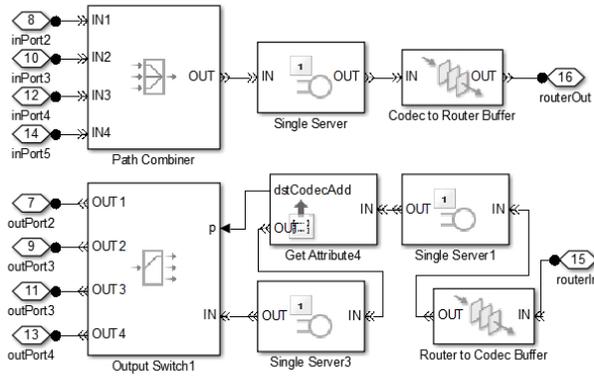


Figure 5: Codec Simulink model

as CONNECT generation tool [10] it limited to 16-port per router, two ports out of the 16 are used to connect with neighbor routers and 14 are left to interface with the tiles, this gives us total of 54 ports for the four routers inside the network.

Second model is network B that uses codecs to interface with 16-tiles, 32-tiles and 64-tiles, for each group we build three configurations for different number of codecs per router, Table II illustrates the codec network configurations, CpR is the codecs per router and TpC is the tiles per codec. For example to build the 16-tiles configuration of network B we can use either one codec per router (1CpR) and each codec connects to four tiles(4TpR), or we can use two codec per router (2CpR) and each codec connects to two tile (2TpC).

Table II: Network B CpR and TpC configurations

Configuration\Tiles	16 Tiles	32 Tiles	64 Tiles
1CpR	4TpC	8TpC	16TpC
2CpR	2TpC	4TpC	8TpC
4CpR	-	2TpC	4TpC

V. COMPARISON RESULTS

Altera Arria II GX EP2AGX260 FPGA is used as a target to compare synthesis results, it has 205200 combinational ALUTs, 102600 memory ALUTs and 692 IO pins, Quartus II 12.0 is used with ModelSim Altera Starter Edition for synthesis and RTL simulation.

The logic utilization values shown in the comparison figures are sum of both combinational and memory resources consumed on the FPGA, the frequency values are the average between the values at 100 and -40 degrees Celsius. Quartus PowerPlay Analyzer tool is used to estimate the consumed power.

A. Frequency

As shown in Fig. 6, the maximum operating frequency of network A is decreasing significantly while increasing the number of tiles, on the other hand network B starts at higher frequency and it decreases slightly as the number of tiles increases.

The reason for the significant difference between 1CpR network and 2CpR network shown in Fig. 7 is not the change in the codec circuit, but because of the difference between 3-port CONNECT router used in 1CpR network and 4-port router used in 2CpR network, the 4-port router would take more area and operate on lower frequency than a 3-port router.

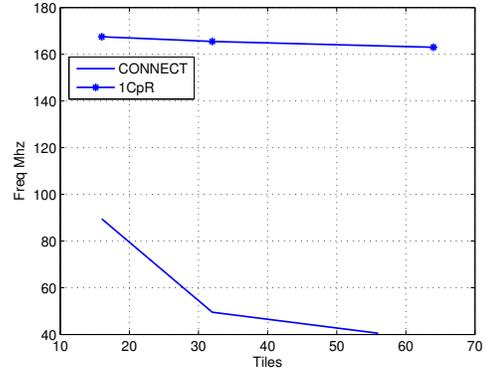


Figure 6: Maximum operating frequency comparison between network A and network B with 1CpR

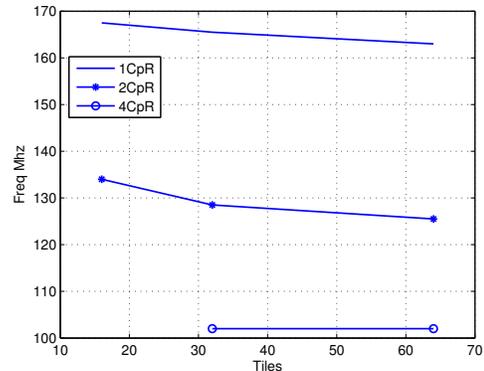


Figure 7: Maximum operating frequency comparison between different network B configurations

B. Logic utilization

As shown in Fig. 8, a 56-tiles network A uses 30% of the FPGA resources, in this network each of the four routers has sixteen ports in order to be able to interface with 14 tiles, however a 64-tiles network B with 1CpR uses at maximum 2% of the resources as each of the four routers has only three ports, two ports to interface with adjacent routers and one to interface with the codec which connects to sixteen tiles.

The reason for this large logic utilization difference is that a 16-port CONNECT router consumes larger area than a 16-port codec.

In Fig. 9, a comparison between 1CpR, 2CpR and 4CpR network B configurations is shown, the 4CpR configuration consume the largest area because each router has six ports, two to connect with adjacent routers and four to connect with four codecs. 2CpR network configuration has a 4-port router and 1CpR has a 3-port router.

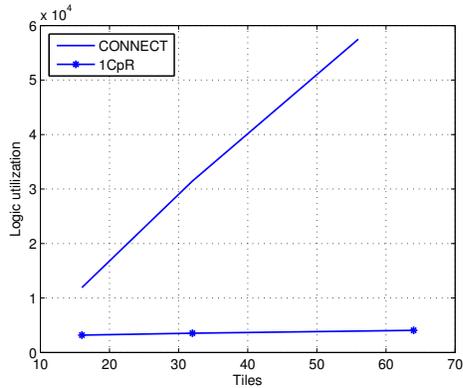


Figure 8: Logic utilization comparison between network A and network B with 1CpR

Required codec ports count will increase as we increase number of connected tiles, this explains the slight logic utilization increase shown in the comparison.

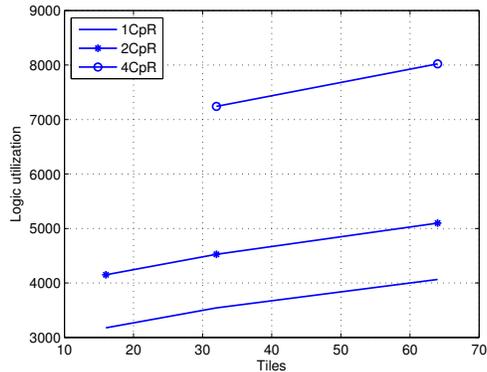


Figure 9: Logic utilization comparison between different network B configurations

C. Power consumption

As shown in Fig. 10, network A consumes more power than all network B configurations because of the large area consumed by CONNECT 6-port routers. Also it is shown that increasing the input/output port count affects power consumption of network A more significantly than network B.

VI. ACKNOWLEDGEMENT

This research was funded by NTRA, ITIDA, Cairo University, Zewail City of Science and Technology, AUC, the STDF, Intel, Mentor Graphics, MCIT.

VII. CONCLUSIONS

In this paper, we assumed that hard NoC routers will be used on next generation FPGAs to replace its conventional long wires. We focused on one problem of NoC routers which is the effect of increasing input/output ports on its area, power and frequency.

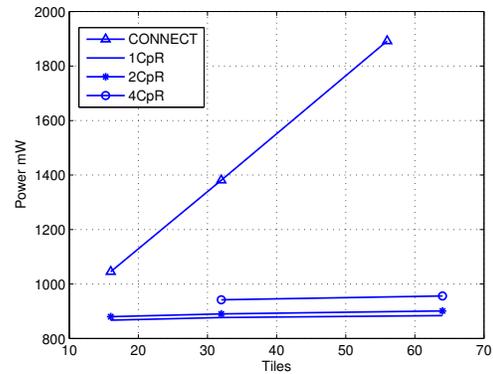


Figure 10: Power consumption comparison between network A and different network B configurations

Instead of increasing the router port count in order to interface with the increased number of tiles, we used a codec module that will concatenate as much as tiles together and interface them with one port of the router. Codec is a time division multiplexing block that is simpler than a NoC router, its area and power doesn't scale significantly with the increased number of tiles and also its frequency will not be decreased greatly like a NoC router.

When comparing two 2x2 networks, one 64 tiles with codec and another 56 tiles with router only, we found that the codec network takes less than 15% area, consume less than 50% power of a router only network and operates with 2.5x frequency.

REFERENCES

- [1] C. Hilton and B. Nelson. *PNoC: A Flexible Circuit-Switched NoC for FPGA-based Systems*. IEEE Proceedings Computers and Digital Techniques, 2006, pp. 181-188.
- [2] Daniel U. Becker. *Efficient Microarchitecture for Network-on-Chip Router*. Ph.D. dissertation, Stanford University, 2012.
- [3] Y. Huan and A. DeHon. *FPGA Optimized Packet-Switched NoC using Split and Merge Primitives*. In IEEE International Conference on Field-Programmable Technology, 2012.
- [4] N. Gamal, H. Fahmy, Y. Ismail, and H. Mostafa, "Design Guidelines for Embedded NoCs on FPGAs", IEEE International Conference on Quality Electronic Design (ISQED'2016), Santa Clara, California, USA, IEEE, pp. 69-74, 2016.
- [5] N. Gamal, H. A. H. Fahmy, Y. Ismail, T. Ismail, M. Mohie-Eldin, and H. Mostafa, "Design Guidelines for Soft Implementations to Embedded NoCs of FPGAs", International Design and Test Symposium (IDT 2016), Hammamet, Tunisia, IEEE, pp. 37-42, 2016.
- [6] M. S. Abdelfattah and V. Betz. *Design Tradeoffs for Hard and Soft FPGA-based Networks-on-Chip*. In IEEE International Conference on Field-Programmable Technology, 2012, pp. 95-103.
- [7] K. A. Helal, S. Attia, T. Ismail, and H. Mostafa, "Comparative Review of NOCs in the Context of ASICs and FPGAs", IEEE International Symposium on Circuits and Systems (ISCAS 2015), Lisbon, Portugal, IEEE, pp. 1866-1869, 2015.
- [8] A. Salaheldin, K. Abdalah, N. Gamal and H. Mostafa. *Review of NoC-based FPGAs architectures*. ICEAC, 2015, pp.1-4.
- [9] M. K. Papamichael and J. C. Hoe. *CONNECT: Re-Examining Conventional Wisdom for Designing NoCs in the Context of FPGAs*. Proceedings of the ACM/SIGDA international symposium on Field Programmable Gate Arrays, 2012, pp. 37-46.
- [10] M. K. Papamichael and J. C. Hoe. *CONNECT: CONFIGurable NETWORK Creation Tool*. <<http://users.ece.cmu.edu/mpapamic/connect/>>, 2012.