Exploration of FPGA-Based Packet Switches for Rack-Scale Computers on a Board

Jong Hun Han, Neelakandan Manihatty-Bojan, Andrew W. Moore
Computer Laboratory, University of Cambridge, UK
Email:{jong.han, neelakandan.manihatty-bojan, andrew.moore}@cl.cam.ac.uk

***Abstract*—This work explores the design space (bandwidth and port configuration) for an FPGA-based top-of-rack switch and, use our implementation, to provide an insight on which of these options is the best. We also propose an architecture for a rack-scale computer built on a printed circuit board (PCB) exploiting the FPGA-based switch.**

***Keywords*-FPGA, Rack-Scale Computer, Top of Rack, Switch.**

I. Rack-Scale Computer on a Board

One of the objectives of the datacenter is to scale economically and enhance user experience by effective network resource utilization. A computer rack built with tens of servers connected through a top-of-rack (ToR) switch can be treated as a primitive cell of the datacenter. Research on rack-scale architecture for datacenters has gained significant interest because it enables incremental upgradation of infrastructure to meet the performance requirements [1]. In this work, we propose a rack-scale computer-on-a-board (RCoB) with System-on-Chip (SoC) servers and an FPGA-based switch. We argue that the switch interconnecting the servers can be built using FPGAs instead of a merchant silicon [2].

The RCoB consists of a number of SoC servers connected to each other through an FPGA-based ToR (F-ToR) switch. Figure 1 illustrates a high level block diagram of the RCoB architecture. As shown in the figure, the F-ToR switch and the SoC servers are connected to each other directly on the PCB through high speed traces for 10Gbps to 50Gbps Ethernet. A general purpose CPU can be connected to the F-ToR switch over the PCI express for controlling and configuring the switch. In the RCoB, all servers and F-ToR switch can be synchronized using an interrupt signal and communication between the server and switch can happen through custom IO pins for transacting data and messages.

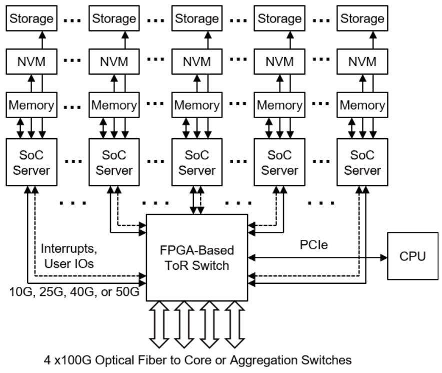

Figure 1: A block diagram of the RCoB.

Table I: F-ToR switch port number and per-port bandwidth configuration.

10Gbps(48)× 100Gbps(4)	25Gbps(16)× 100Gbps(4)	40Gbps(12)× 100Gbps(4)	50Gbps(8)× 100Gbps(4)
10Gbps(48)× 50Gbps(8)	25Gbps(16)× 50Gbps(8)	10Gbps(48)× 40Gbps(12)	25Gbps(16)× 40Gbps(12)

II. Implementation Results

Table I shows a list of the port numbers and per-port-bandwidth configuration used in this work for the F-ToR switch implementation.

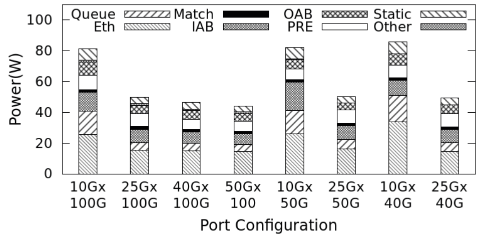

Figure 2: Power consumption results for the F-ToR switches implemented with the configuration in Table I.

Figure 2 shows the total power consumption results broken down into the modules of the F-ToR switches implemented on a Xilinx Ultrascale Virtex-7 xcvu190-flga2577-3-e. The static power (Static) of each configuration was less than 10% of the total power consumption. In each configuration, the Ethernet cores (Eth) consumed the most power. It is also observed that the F-ToR switches configured with the 10Gbps Ethernet core consumes more power than other configuration, which implies that the number of the cores is more dominant factor in the power consumption than the bandwidth. While the 10Gbps×40Gbps F-ToR switch consumes the highest power compared to other configuration, the 10Gbps×50Gbps consumed the least power among the different configurations.

Acknowledgment

This work is supported by the EU Horizon 2020 SSICLOPS (grant no. 644866) and EU Horizon 2020 ENDEAVOUR (grant no. 644960) projects.

References

[1] P. Costa, H. Ballani, K. Razavi, and I. Kash. ”R2C2: A Network Stack for Rack-scale Computers”. *ACM SIGCOMM Comput. Commun. Rev.*, 2015.

[2] N. Farrington, E. Rubow, and A. Vahdat. ”Data Center Switch Architecture in the Age of Merchant Silicon”. In *IEEE Sym. HOTI*, 2009.

978-1-5386-4037-1/17 $31.00 © 2017 IEEE
DOI 10.1109/FCCM.2017.35