POLITECNICO DI TORINO
Repository ISTITUZIONALE

3D-Hyper-FleX-LION: A Flat and Reconfigurable Hyper-X Network for Datacenters

Original
3D-Hyper-FleX-LION: A Flat and Reconfigurable Hyper-X Network for Datacenters / Liu, G., Proietti, R., Fariborz, M., Fotouhi, P., Xiao, X., Yoo, S.J.B.. - ELETTRONICO. - (2020). (OSA Advanced Photonics Congress (AP) Washington, DC United States 13–16 July 2020) [10.1364/PSC.2020.PsW1F.3].

Availability:
This version is available at: 11583/2973644 since: 2022-12-05T18:01:11Z

Publisher:
Optical Society of America

Published
DOI:10.1364/PSC.2020.PsW1F.3

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
Optica Publishing Group (formely OSA) postprint/Author's Accepted Manuscript

"© 2020 Optica Publishing Group. One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modifications of the content of this paper are prohibited."

(Article begins on next page)

23 June 2026

3D-Hyper-FleX-LION: A Flat and Reconfigurable Hyper-X Network for Datacenters

Gengchen Liu, Roberto Proietti, Marjan Fariborz, Pouya Fotouhi, Xian Xiao, and S. J. Ben Yoo
Electrical and Computer Engineering, University of California, Davis, CA, USA
Author e-mail address: genliu@ucdavis.edu, sbyoo@ucdavis.edu

Abstract: We propose a flat datacenter network using silicon photonic switches. Simulations show up to 2× improvement in throughput-per-watt over a non-oversubscribed Fat-Tree while providing > 2× reduction in the number of switching ASICs and transceivers. © 2020 The Author(s)

1. Introduction

Modern hyperscale datacenters such as Facebook' F16 or Microsoft's Quantum 10 are built with classic or modified versions of the Fat-Tree architecture, whose scalability, latency, and power consumption are dictated by the available port-count of merchant silicon switch ASICs [1]. While the Fat-Tree network can provide full bisection bandwidth and scalability, it has several drawbacks. First, by cascading many hierarchies of switches (top-of-rack, aggregation, and core switches), the multi-stage nature of Fat-Tree networks introduces large power consumption and end-to-end packet transmission latency. Second, while the communication patterns between servers are typically unevenly distributed and cause heavily congested or underutilized links, current networks are incapable of changing their network topology and link bandwidth to adapt to the significant variations of traffic patterns [2]. Overall, it is challenging for architectures like Fat-Tree to balance energy efficiency and network bandwidth utilization. A flat and reconfigurable photonic interconnect architecture has the potential to overcome the above issues associated with Fat-Tree architecture. By leveraging recent advances in silicon photonic switch fabrics [3] and low-diameter directly connected architectures like the HyperX topology [4], this paper presents 3D-Hyper-FleX-LION, an optically reconfigurable Hyper-X network. When compared to a Fat-Tree with the same oversubscription, 3D-Hyper-FleX-LION provides up to 2× improvement in throughput per watt while delivering a 2.5× reduction in the number of switching ASICs, a 4× reduction in the number of transceivers, and a 10× reduction in the number of fibers.

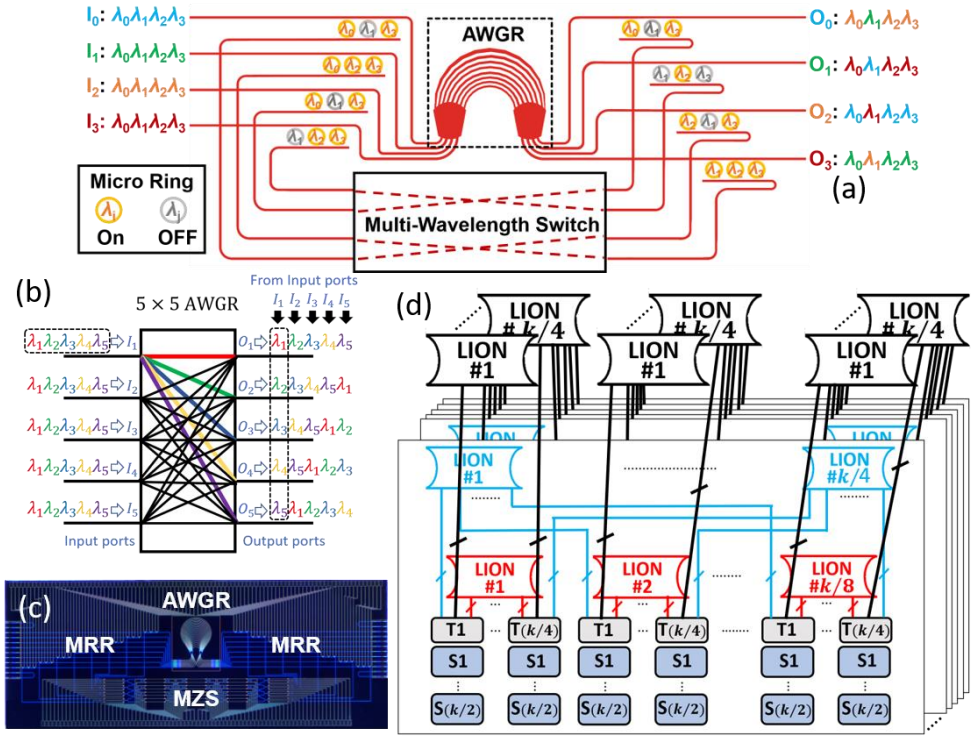

Figure 1. (a) $N\times N$ Flex-LION fabric with N=4 AWGR and b=3 MRR, and $N\times N$ MZI crossbar switch; (b) All-to-all connectivity of AWGR; (c) Microscope image of a fabricated 8×8 Flex-LIONS; (d) . The proposed 3D-Hyper-FleX-LION architecture.

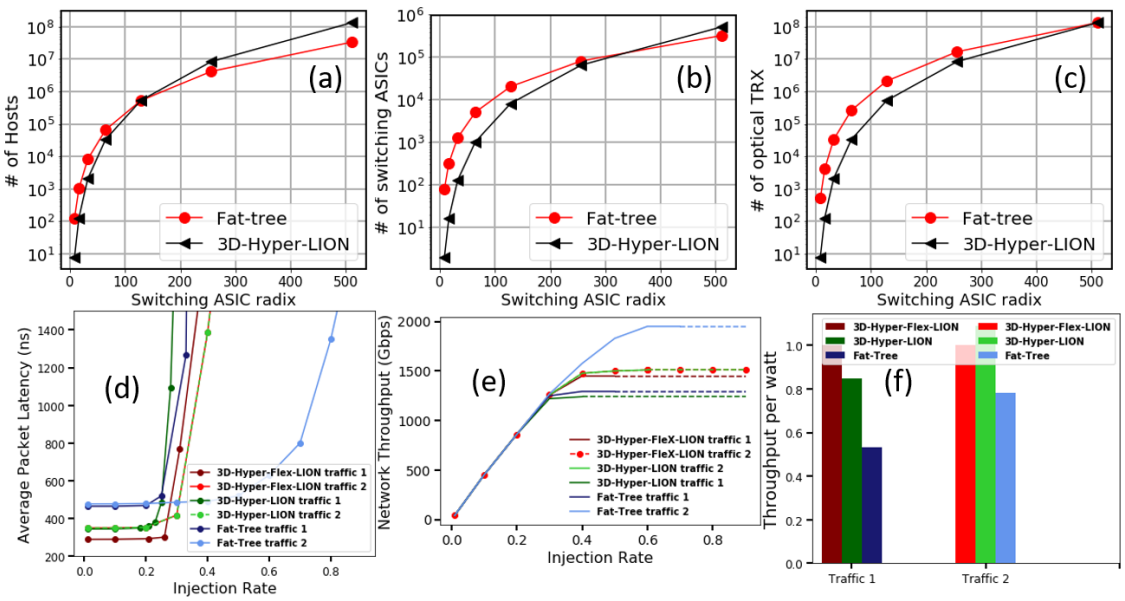

Figure 2. (a) Maximum number of supported host vs. switching ASIC radix; (b) Number of required ASICs vs. switching ASIC radix; (c) Number of required optical transceivers vs. switching ASIC radix; (d) Latency vs. injection rate; (e) Throughput vs. injection rate; (f) Throughput per watt for two different traffic patterns. Hyper-LION represents the case where MRR reconfiguration is disabled.

2. 3D-Hyper-Flex-LIONS Architecture and Network Simulation Results

Figure 1(a) shows the block diagram of Flex-LIONS (Flexible Low-Latency Interconnect Optical Network Switch), which consists of an $N\times N$ AWGR and an $N\times N$ colorless optical switch placed in between b-port MRR-based wavelength selective switches. For uniform random traffic, the MRRs can be set off-resonance so that each input port of the AWGR receives N wavelengths. In that case, all-to-all interconnection is achieved according to the wavelength-routing property of AWGR shown in Figure 1(b). Alternatively, the MRRs can be tuned in resonance to route certain wavelengths to the colorless optical switch rather than the AWGR. Figure 1(c) shows a Flex-LIONs device with N=8 and b=3 with ~5 µs switching time (as demonstrated in [3]). Although a single Flex-LIONS device requires N wavelengths from each node to interconnect N nodes, it is also possible to leverage the AWGR in Thin-CLOS architecture to realize the $N\times N$ connectivity using only the W wavelengths [5]. Figure 1(d) describes the proposed 3D-Hyper-FleX-LION architecture based on Thin-CLOS Flex-LIONSs and identical k-port TORs organized in pods, clusters, and full system. Each TOR uses half of its ports ($^{k}/_{2}$) for hosts connections. The other half of the TOR ports are partitioned into three groups used for intra-pod, inter-pod, and inter-cluster communications, where each port is used to drive a wavelength channel of a W-wavelength optical transceiver. The 3D partitioning offers the proposed architecture great scalability as it can support $^{k^4}/_{512}$ number of hosts, which is larger than what Fat-Tree can support $^{k^3}/_{4}$ when the switch ASIC radix is equal or larger than 128. Figure 2(a-c) present the theoretical analysis between a Fat-Tree and the proposed architecture with switch ASIC radix ranging from 8 to 512 ports. When k is equal to 128, the proposed architecture achieves 4× reduction in the number of optical transceivers and 2.5× reduction in the number of switching ASICs with the same number of supported hosts. We evaluated the performance of the proposed network using gem5 [6] with two types of traffic patterns. In traffic 1, 50% of hosts create a background traffic with uniform random distribution, while the other 50% of hosts create hotspot links between every two TORs. In traffic 2, all of the hosts inject uniform random traffic. Figure 2(d-f) shows the performance results based on different injection rates. The proposed network improves the packet latency and power consumption when there are hotspot links in the network. In conclusion, the results demonstrate that the proposed network outperforms Fat-Tree in terms of scalability and power consumption while providing better latency in the case of hotspot traffic and higher throughput per watt.

3. References

[1] Facebook. Reinventing Facebook's data center network. https://engineering.fb.com/data-center-engineering/f16-minipack/
[2] M. Al-Fares et al, "A scalable, commodity, data center network architecture" in Proceedings of the ACM SIGCOMM conference, 2008.
[3] X. Xiao et al, "Silicon Photonic Flex-LIONS for Bandwidth-Reconfigurable Optical Interconnects", IEEE JSTQE 26(2), 1-10, 2019.
[4] J. Ahn, et al, "HyperX: Topology, Routing, and Packaging of Efficient Large-Scale Networks", in ACM SC, 2009.
[5] R. Proietti et al, "Experimental demonstration of a 64-port wavelength routing thin-CLOS system for data center switching architecture" JOCN 10(7), 49-57, 2018.
[6] Binkert et al, "The gem5 simulator," ACM SIGARCH Comput. Archit. News, vol. 39, no. 2, p. 1, Aug. 2011.
This work was supported in part by DoD contract # H98230-16-C-0820 and NSF grant # 1611560.