Bringing minimal routing back to HPC through silicon photonics: The Flexfly architecture

Jeremiah Wilke*, Sébastien Rumley† and Keren Bergman†
* Sandia National Laboratories, Livermore, CA xxxxx
† Department of Electrical Engineering, Columbia University, New York, NY 10027
jjwilke@sandia.gov

Performance of optical transmission cannot be matched for long-distance data movement. However, in supercomputers as in large data centers, high optical bandwidth densities aren't fully leveraged, yet [1]. To date, implementing a high performance interconnect for a system with 10K compute nodes entirely with optical cables and switches remains cost-prohibitive. Current supercomputing systems maintained by the Department of Energy implement topologies primarily with electrical packet switches with optical cables only for long-reach links. While torus topologies previously dominated, high-radix packet switches have made hierarchical topologies like dragonfly and fat-tree more popular. The dragonfly topology [2] provides low-diameter connectivity with all-to-all global links at the inter-group level, and aims at minimizing the number of long distance link, synonym of expensive optical links. Because optical links are minimized, minimal routing is generally insufficient for dragonfly, and adaptive routing is required to utilize bandwidth of non-minimal paths [2]. In other terms, neighboring inter-group links are leveraged to compensate the lack of bandwidth available on the direct link. However, this results in additional network hops, lower power efficiency, and more complicated router implementations.

An obvious way to deal with these issues would consists in multiplying inter-group links, potentially to the point where a dragonfly topology becomes a 2D-Flattened-Butterfly, or in boosting bandwidth of existing inter-group dragonfly links. This may be realized in the future, when more cost-effective optical transmission links will be available.

A more subtle way to avoid non-minimal routing and its associated downsides consists in realizing bandwidth steering with optical switching. In this talk, we show how optical switches, and more specifically, low-radix ones, can alleviate some of these design challenges faced by HPC networks. We briefly survey existing optical technologies, their cost tradeoffs, and how they might be incorporated into HPC networks. We then present the Flexfly architecture, a dragonfly design incorporating low-radix silicon photonics switches [3]. Rather than adaptively route around congested global links, Flexfly can reconfigure bandwidth to match traffic patterns. We present initial results from a Flexfly study of several scientific applications, showing nearly order of magnitude improvements over basic minimal routing and over 50% improvement relative to adaptive routing. The SST/Macro simulator [4] used in the study is introduced. We finally discuss outstanding research and implementation questions surrounding Flexfly, in particular how this approach might enable simplifications at the router architecture level, and its future viability as an exascale network architecture.

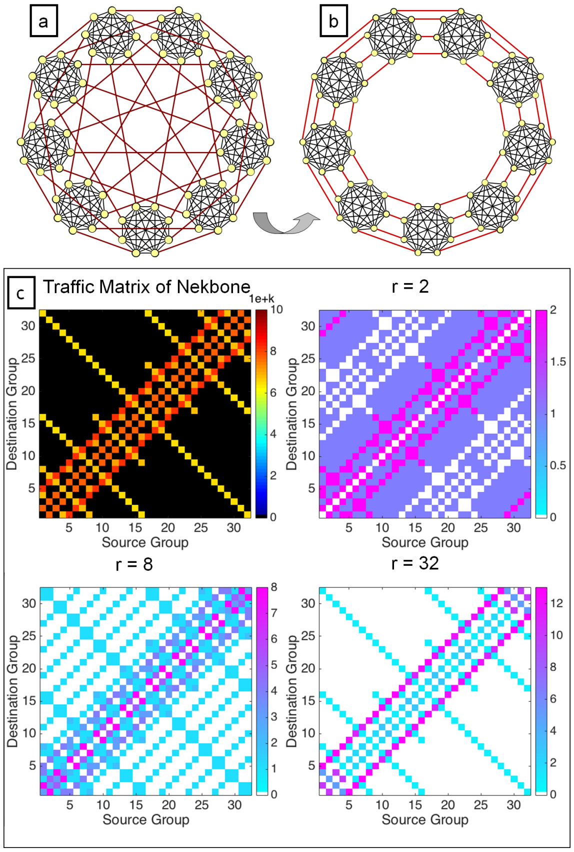

Figure 1: (a) Regular Dragonfly topology with all-to-all inter-group links (b) Re-organized topology after bandwidth steering using optical switching (c) Top-left matrix shows traffic distribution across pairs of dragonfly groups during execution of Nekbone workload; other matrices shows how bandwidth can be steered (as function of the optical switch radix) to mimic workload distribution.

1. REFERENCES

[1] S. Rumley *et al.*, "End-to-end modeling and optimization of power consumption in hpc interconnects," in *International Conference on Parallel Processing Workshops (ICPPW)*, Aug 2016.

[2] J. Kim *et al.*, "Technology-driven, highly-scalable dragonfly topology," *SIGARCH Comput. Archit. News*, vol. 36, June 2008.

[3] K. Wen *et al.*, "Flexfly: Enabling a reconfigurable dragonfly through silicon photonics," in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, Nov 2016.

[4] "http://sst.sandia.gov/about_sstmacro.html."

AISTECS 2017, January 25 2017, Stockholm, Sweden

ACM ISBN 978-1-4503-4084-7/16/01...$15.00

DOI: http://dx.doi.org/10.1145/2857058.2857061