# Harnessing Path Diversity for Laser Control in Data Center Optical Networks

Y. Demir[#], N. Terzenidis[‡], H. Han[†], D. Syrivelis[*], G. T. Kanellos[‡],
N. Hardavellas[†], N. Pleros[‡], S. Kandula[¥], and F. Bustamante[†]
[#]Intel, Hillsboro, OR, USA, yigit@u.northwestern.edu
[‡]Aristotle University of Thessaloniki, Thessaloniki, Greece, {nterzeni, gtkanellos, npleros}@csd.auth.gr
[†]Northwestern University, Evanston, IL, USA, {nikos, fabianb}@northwestern.edu
[*]Center for Research and Technology - Hellas (CERTH), Thessaloniki, Greece, dimitris.syrivelis@gmail.com
[¥]Microsoft Research, Redmond, WA, USA, srikanth@microsoft.com

***Abstract*—Optical interconnects are already the dominant technology in large-scale datacenter networks. Unfortunately, the high optical loss of many optical components, coupled with the low efficiency of laser sources, result in high aggregate power requirements for the thousands of optical transceivers that such networks employ. As optical interconnects stay always on, even during periods of system inactivity, most of this power is wasted. Ideally we would like to turn off the transceivers when a network link is idle (i.e., "power gate" the lasers), and turn them back on right before the next transmission. The danger with this approach is that it may expose the laser turn-on delay and lead to higher network latency. However, data center networks typically employ network topologies with path diversity and facilitate multiple paths for each source-destination pair. Based on this observation, we propose an optical network architecture where redundant paths are turned off when the extra bandwidth they provide is not needed, and they turn back on when traffic increases beyond a high watermark to decongest the network. Maintaining full connectivity removes the laser turn-on latency from the critical path and results in minimal performance degradation, while at the same time power-gating the lasers saves 60% of the laser power on average on a variety of data center traffic scenarios.**

***Keywords*—Data Center Networks; Optical Networks; Energy Proportionality; Energy Efficiency; Laser Gating**

## I. Introduction

Photonics have emerged as a promising solution to meet the growing demand for high-bandwidth, low-latency, and energy-efficient communication in manycore processors [14], chip-to-chip interconnects [3,6,10], and large-scale data centers [1,8,12]. Lasers are a major contributor to the total power consumption of these networks. As we argue, most of this energy is wasted.

Lasers consume a significant amount of power because their output power needs to be high enough to compensate for the optical loss of couplers, waveguides and other optical components. For example, crossing two 3 *dB* couplers on either side of a single link increases the laser power by 4×. Process variations force designers to increase the laser power even higher to maintain a safety margin. To make matters worse, WDM-compatible lasers are only 5–30% efficient. Thus, the required wall-plug laser power can easily grow by more than 10× when all the losses and inefficiencies are factored in.

However, most of the laser energy is wasted. Compute-intensive execution phases underutilize the interconnect, and servers in the cloud often stay idle or exhibit load imbalances (for example, servers in Google-scale data centers are typically 20–30% utilized [2]). While the full laser power is required to support periods of high network activity, the laser is wasted during idle times because photonic interconnects are always on. A naive form of "*laser gating*" (i.e., turning off the transceiver of an idle link to save energy [9]) risks exposing multiple laser turn-on delays to the application. A packet typically crosses multiple links to its destination. Every time it wakes up a link it needs to wait for the laser to turn on.

We observe that we can hide this delay by allowing packets to take alternative paths to their destination while a link turns on. This is easily achieved in modern data center networks as they employ topologies that provide "*path diversity*", i.e., multiple paths to each source-destination pair. For example, Facebook [12] and Microsoft [8] data centers use clos topologies, while flattened butterfly has been proposed as a cost-efficient alternative by Google [1]. Instead of turning off links arbitrarily and severing end-to-end paths, which exposes the laser turn-on delay, we propose to turn off only redundant links when utilization is low, and turn them on again when bandwidth demand grows [5]. Moreover, we propose to control the server-to-top-of-rack (ToR) switch links by intercepting socket write calls at the OS level and giving an early warning to the network interface card (NIC) lasers to turn on. We propose a network architecture, *Staged Laser Control for Data Centers (SLaC-DC)*, that embodies these ideas. We evaluate SLaC-DC on a variety of data traffic scenarios and show that it saves on average 60% of the optical transceivers power (68% max) at the cost of 6% higher packet delay.

This work is partially supported by NSF CAREER award CCF-1453853.

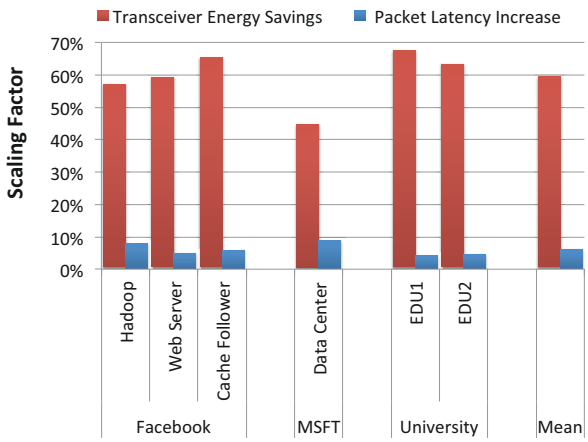


**FIGURE 1. Impact of SLaC-DC on Tx energy and packet latency.**

## II. Motivation

Power and energy efficiency have been at the forefront of research in circuits and computer architecture for at least 15 years [11]. Innovations in these fields coupled with advanced materials, semiconductor processes and packaging, yield lower-power devices for logic, memory and storage at each technology node [13]. The end result is the rapidly increasing energy efficiency of each server node. At the same time, the high power demands of modern data centers have pushed data center operators to drastically reduce inefficiencies and operate with power overheads for cooling, electrical and mechanical systems as low as 7% today [7]. As servers deliver increasingly higher performance per Watt, and data center overheads are aggressively eliminated, their contribution to the overall data center power consumption drops, exposing other components that have not yet received similar attention [9]. Data center networks are one of these components, and its relative power consumption rises as innovations in other sectors reduce the power draw of other data center components. While the switches in data centers today consume 5–10% of the overall data center power, we estimate that with the advent of fully energy-proportional servers, specialized computing, and the efficiencies of end-of-roadmap logic and storage devices [13], switches and NIC transceivers will consume more than 30% of data center power.

## III. Evaluation

We evaluate SLaC-DC on a simulation of a clos data center network similar to Facebook's [12] fed by traffic generators that model the traffic of Facebook [12] and Microsoft [8] data centers, as well as on traces of packets passing through a router in a university datacenter [4]. We find that SLaC-DC can save between 45–68% of the optical transceiver latency (60% average) at a cost of 6% higher average packet latency (Figure 1).

## IV. Conclusion

As servers and other data center components become increasingly more power efficient and approximate full energy proportionality, the power inefficiencies of optical interconnects will propel them to major power consumers. To address this issue we propose SLaC-DC, a data center network architecture that harnesses path diversity to turn off idle paths and save 60% of the overall optical transceiver energy. We believe these results justify further exploration of laser gating at data centers and the development of proof-of-concept prototypes.

## References


[1] D. Abts, M. R. Marty, P. M. Wells, P. Klausler, and H. Liu. Energy proportional datacenter networks. In *Proc. of the 37th Annual International Symposium on Computer Architecture*, 2010.

[2] L. A. Barroso and U. Holzle. The case for energy-proportional computing. *IEEE Computer*, 40(12):33–37, 2007.

[3] S. Beamer, K. Asanovic, C. Batten, A. Joshi, and V. Stojanovic. Designing multi-socket systems using silicon photonics. In *Proc. of the International Conference on Supercomputing*, 2009.

[4] T. Benson, A. Akella, and D. A. Maltz. Network traffic characteristics of data centers in the wild. In *Proc. of the 10th ACM SIGCOMM Conference on Internet Measurement*, 2010.

[5] Y. Demir and N. Hardavellas. SLaC: Stage laser control for a flattened butterfly network. In *Proc. of the 22nd IEEE International Symposium on High Performance Computer Architecture,* 2016.

[6] Y. Demir, Y. Pan, S. Song, N. Hardavellas, J. Kim, and G. Memik. Galaxy: A high-performance energy-efficient multi-chip architecture using photonic interconnects. In *Proc. of the 28th ACM International Conference on Supercomputing*, 2014.

[7] Facebook. Prineville. https://www.facebook.com/PrinevilleDataCenter/app/399244020173259/. Retrieved July 2016.

[8] A. Greenberg, J. R. Hamilton, N. Jain, S. Kandula, C. Kim, P. Lahiri, D. A. Maltz, P. Patel, and S. Sengupta. VL2: A scalable and flexible data center network. In *Proc. of the ACM SIGCOMM Conference on Data Communication*, 2009.

[9] B. Heller, S. Seetharaman, P. Mahadevan, Y. Yiakoumis, P. Sharma, S. Banerjee, and N. McKeown. Elastictree: Saving energy in data center networks. In *Proc. of the 7th USENIX Conference on Networked Systems Design and Implementation*, 2010.

[10] A. Krishnamoorthy, R. Ho, X. Zheng, H. Schwetman, J. Lexau, P. Koka, G. Li, I. Shubin, and J. Cunningham. Computer systems based on silicon photonic interconnects. *Proceedings of the IEEE*, 97(7):1337–1361, 2009.

[11] T. Mudge. Power: A first-class architectural design constraint. *Computer*, 34(4):52–58, 2001.

[12] A. Roy, H. Zeng, J. Bagga, G. Porter, and A. C. Snoeren. Inside the social network s (datacenter) network. In *Proc. of the ACM SIGCOMM Conference on Data Communication*, 2015.

[13] Semiconductor Industry Association. The international technology roadmap for semiconductors (itrs). http://www.itrs.net/, 2015.

[14] D. Vantrease, R. Schreiber, M. Monchiero, M. McLaren, N. P. Jouppi, M. Fiorentino, A. Davis, N. Binkert, R. G. Beausoleil, and J. H. Ahn. Corona: System implications of emerging nanophotonic technology. In *Proc. of the 35th Annual International Symposium on Computer Architecture*, 2008.