Meeting Exascale Interconnect Requirements with Integrated Photonics

Sébastien Rumley* and Keren Bergman
Columbia University, New York, USA
* Corresponding author: rumley@ee.columbia.edu

Massive computation capabilities of Supercomputers are exploited in an increasing number of research areas. In some of these fields, simulations models involving 10^{15} floating-point operations per second (Petaflops) already exist and run on Petascale class Supercomputers [1]. Modelers, programmers, and computer architects alike thus already look at the next grand challenge: build an Exaflops (10^{18}) capable Supercomputer, and develop simulation codes able to effectively leverage this computing power.

Neither the clock speed nor complexity of processors will progress much in the next years (as they have not progressed in the last years). The Exascale mark will thus mainly be attained by mean of parallelism increase. An Exascale machine will roughly require 100 millions of processors (10^8) each capable of roughly 10 Gigaflops (10^{10}). This is ~30 times more than the three millions of core active in today's most powerful SuperComputer, Tianhne-2. Some of this parallelism will be "absorbed" by multi-core architectures: Intel's Knights Corner included in Tianhne-2 involves 50 cores on the same chip, yielding around 1 Teraflop. Upcoming chips will likely deliver 10-30 Teraflops. However, from 30 to 100 thousands of these inflated chips will still be necessary to obtain an Exaflops. Massive interconnects regrouping up to 100K end-points, capable of ultra-short latencies, and delivering ultra-high bandwidth will thus be required [2].

Scaling current interconnects is, however, a challenge in itself. In the last ten years, the bandwidths of Supercomputer interconnect sluggishly evolved from around 10 Gbps to 100 Gbps. Signaling rates have been with difficulty multiplied by a factor of 5, from 5 to 25 Ghz [3]. As for processor clock frequencies, CMOS limits are being reached, and parallel lanes are increasingly used compensate for this fact. Processor parallelism cost (and to some extent, power dissipation) can be mitigated by further transistor downsizing. Cabling parallelism, in contrast, induces extra money and energy costs. As a result, parallel programmers tend to resign themselves to work in a bandwidth scarce environment [4].

Optics is present in Supercomputers for a decade now since ASCI Purple, first Supercomputer with optical cables, was introduced in 2005. So far, however, optical systems only act as mere subcontractors. They are used for links whose distance cannot be overcome with metallic cables. This might change with the advent of highly integrated photonics, as well as hybrid or even monolithic integration of integrated photonics with conventional electronics. Optical WDM transceivers capable of Terabit/s order bandwidths could be directly integrated within the multi-teraflop capable node, limiting the connection towards the interconnect to a single fiber. Furthermore, with end-points directly emitting and receiving on the optical domain, the use of optical switching at the interconnect core can be envisioned. Realizing such an integration could reverse the trend with respect to bandwidth available for each end-point. This would ease the job of parallel programmers, and potentially make Exascale simulation accessible to more areas.

In this talk, we will present in more details this opportunity for photonics. We will review the major obstacles that need to be overcome for this integration to be realized. A strong emphasis will be placed on power consumption, which must be kept within a well-defined envelope. The key concept of energy proportionality will be introduced.

References

[1] L. Wolf, G.W. Pieper, "Biology and Medical Research at the Exascale," SciDAC Rev. 16, Spec. ed (2010).
[2] S. Rumley et al., "Silicon Photonics for Exascale Systems", Journal of Lightwave Technology. **33**, 4 (2015).
[3] J. F. Bulzacchelli et al., "A 28-Gb/s 4-Tap FFE/15-Tap DFE Serial Link Transceiver in 32-nm SOI CMOS Technology", IEEE Journal of Solid State Circuits **47**, 12, (2012)
[4] R. F. Barrett, et al. On the Role of Co-design in High performance Computing, IOS Press, 2013.