Overview of Photonics Systems from the Physical Layer to Photonic-Based Architectures

Keren Bergman, Columbia University

A wealth of computing platforms, from handheld mobile devices, to embedded systems, and data centers are leveraged to produce major societal benefits as we embark on the age of ubiquitous computing. Continuous connectivity is an essential property driving applications among a myriad of computing devices that interact with the physical world and the high-performance parallel systems, servers, and data-centers that form the backbone of this ecosystem. The continued progress in our capabilities to unlock new or previously intractable applications is dependent more than ever on the scalability of this critical ubiquitous computing infrastructure.

While over the past decades computing progress has been largely fueled by fundamental advances in semiconductor technology, the recent emergence of multicore architectures and vastly increased parallelism signifies a profound transformation in how future computing scales. With the extraordinary growth in parallelism at all system scales, *data movement* rather than computation dominates system performance. The fundamental limitations imposed by increasing energy consumption associated with moving vast data among the growing parallel compute resources has led to the so called 'bandwidth taper' prevalent in current system architectures. The interwoven bandwidth-energy challenges essentially reduce by orders of magnitude the available communication bandwidth as data propagates from the chip, across the die and the system racks. At the same time, the communication energy costs grow exponentially with system 'distance' and even under severely reduced bandwidths can typically consume a substantial fraction of the overall power budget. The energetic consequences of these challenges are daunting, as energy consumption is completely dominated by costs of data movement.

Among the technologies that are emerging in the age of end-of-scaling electrical links, silicon photonics is perhaps the most promising to enable a transition toward a new generation of computing systems. Unlike prior generations of photonic technologies, the remarkable capabilities of silicon integrated photonics, enabled by their small device footprints, ultra-low capacitances, and the tight proximity of electronic drivers, offer the possibility of generating and receiving optical signals with fundamentally superior energy efficiencies. The insertion of photonic interconnection networks further change the energy scaling rules: once a photonic path is established, the data are transmitted end-to-end without the need for power consuming repeaters, regenerators, or buffers.

Besides improvements that optical communication offers in terms of bandwidth density, energy-efficiency, and transmission latency, the critical advantage enabled by silicon photonics is the close proximity of electronic driver circuitry. This allows for low energy transmission and reception of optical data that is distance independent: once the data is encoded in the optical domain, the propagation distance, whether it's 1cm, 10cm, or even 100m, is immaterial to the link performance. The bandwidth-distance decoupling property of optical communication uniquely enables scalable global connectivity across the entire system.