Reducing power in information processing – ultra-low energy optoelectronics?

David A. B. Miller
Ginzton Laboratory, Stanford University, U.S.A.

Power has become the dominant constraint in information processing, and will continue to be for the foreseeable future and beyond. Already we cannot continuously run all the transistors on a chip because of power limits. Even if the topological and other limitations of electrical wiring did not already limit interconnections, we could not anyway connect all the transistors and processors to one another or to memory or the outside world at their optimal bandwidths without generating too much heat. This power dissipation not only constrains performance within the box; it also leads to economic and environmental limits. The energy cost of running a server for its lifetime is larger than its purchase price. The power consumption of servers already is several percent of all electricity and of the human carbon footprint, and that number will grow. Just the interconnect power dissipation in server farms is larger than the total generating capacity of current solar power sources (see [1] for discussion of these arguments and for detailed references).

Electrical wires operate at energies of ~1 pJ/bit or higher for off-chip interconnects and ~ a few 100 fJ/bit to ~1 pJ/bit for global on-chip interconnects. It is not clear if these energies can be substantially reduced in electrical systems. Such energies are certainly too high to allow the interconnect bandwidth to keep up with the transistor performance. As a result, even if electrical lines could physically handle the necessary bandwidth, the ratio of interconnect bandwidth to computational power – the number of bytes of memory access per floating point operation (FLOP) cannot be sustained in future machines. A target of 1 byte/FLOP is often considered desirable in computer architectures, though current large machines already cannot sustain this. With electrical lines, there is no reasonable prospect of preserving this ratio and, as a result, future computer architectures are severely constrained.

Currently researched optical interconnect device technologies, especially for efficient modulators and ultra-low-capacitance photodetectors, may be able to operate with 10 fJ/bit transmitter energies, a number that, on energy grounds alone, could allow substantial improvement in bandwidth for given power. With interconnect device energies in the 1–10 fJ/bit range, it might be possible to achieve and retain 1 byte/FLOP in future processors. These are aggressive numbers for optoelectronics, but not unphysical. Nanophotonic resonators and nanometallic antennas for confining light to very small subwavelength scales might allow devices to operate into the aJ/bit range; such interconnect energies could transform information processing systems. Physics and prospects for such devices will be discussed (again, see [1] for a comparative summary of approaches).

1. D. A. B. Miller, "Device requirements for optical interconnects to silicon chips", to appear in *Proc. IEEE* (2009), preprint available at *http://www-ee.stanford.edu/~dabm/370.pdf*