Inside the Box:  
A New Hope for Optics?  

Vladimir Stojanović  

Integrated Systems Group  
Massachusetts Institute of Technology
High-speed links needed everywhere

Backbone Router Rack

PC or Console

Integrated Systems Group
What makes it challenging

- Requires sophisticated equalization circuits

![Channel response graph]

source: Rambus
Chip-to-chip I/O scaling problem

- Bandwidth need grows faster than energy/bit drops
- Creates exponentially increasing I/O power consumption
  - In power constrained systems (like processors and anything inside the box) – this limits the available bandwidth
Parallel off-chip links

- Often share clock generation and synch
- Limited equalization (few taps)
- Most power burned to drive the 50 Ω line
  - Current-mode – 200 mV swing (4 mW off 1 V supply,
    - Data rate independent
    - With receiver and pre-driver, at 10 Gb/s energy budget 500 fJ/bit
  - Voltage-mode – (2 pJ/bit state-of-the-art, dynamic power)
    - Can possibly scale to 500 fJ/bit, but not much further
Convergence of platforms

Only way to meet future system feature set, design cost, power, and performance requirements is by programming a processor array:
- Multiple parallel general-purpose processors (GPPs)
- Multiple application-specific processors (ASPs)

**Intel Network Processor**
- 1 GPP Core
- 16 ASPs (128 threads)

**IBM Cell**
- 1 GPP (2 threads)
- 8 ASPs

**Sun Niagara**
- 8 GPP cores (32 threads)

**Picochip DSP**
- 1 GPP core
- 248 ASPs

**Cisco CSR-1**
- 188 Tensilica GPPs

**Intel 4004 (1971):**
- 4-bit processor,
- 2312 transistors,
- ~100 KIPS,
- 10 micron PMOS,
- 11 mm² chip

“The Processor is the new Transistor”
[Rowen]
On-chip network opportunities

- Multiple cores on chip need to communicate
  - On and off-chip
  - Need short latency and large bandwidth
  - Current throughputs up to 1-2Tb/s
    - Need to be extremely energy-efficient since CPU power limited
    - Need to be area efficient as well
Example – 90nm CMOS, 10mm wire

- Conventional repeaters ~ 2-5pJ/bit
- Equalized point-to-point links ~ 0.2-0.5pJ/bit (10x better)
  - Latency < 1 clock cycle for 20 mm x 20 mm die and <10 GHz clocks
  - Sets the on-chip photonic link budget to <100 fJ/bit
Si-photonics may be more efficient

- Modulation speeds approaching 10Gb/s
- Energy-efficiency 1-2pJ/bit
- Potential for high-density WDM

- Off-chip
  - Great if coupled into optical backplanes
- On-chip
  - Need to improve energy-efficiency by 10-100x
- Big challenges:
  - Impact of thermal control
  - Process variation
  - Coupling to external waveguides
Target system - year 2010

- 32 x 32 core chip
  - Each core has a GPU+vector unit+local storage
    - Optional L2 cache slice
- 45 nm CMOS technology
- 30 Tb/s available data throughput
  - 60 waveguides, 50 wavelengths per waveguide, 10 Gb/s per wavelength
  - 3000 addressable DRAM banks (total >200 GB)
- 0.1 mm waveguide pitch for I/O
  - A single pad
Link 1: Fixed L2 slice-to-DRAM channel

- Tile-to-off-chip-DRAM link with dedicated photonic network
- The core-to-core network is electrical
  - Message/packet routing network
Link 2: Multiple-access L1 slice-to-DRAM network

- Tile-to-off-chip-DRAM with multiple-access photonic network
  - Network has to resolve multiple access problem
    - Many cores to same DRAM bank (wavelength channel)
- Remove L2 cache (hit rate only 50%)
  - Add more cores
- On-chip and off-chip networks are aggregated into one
Photonic DRAM interface

- 1 single-mode fiber per DIMM
  - 50 wavelengths per DIMM (50 DRAM banks)
- Hope to spread the traffic uniformly to get maximum from dedicated links
Density comparison

- **On-chip**
  - Assume 10µm pitch per modulated waveguide
    - 2µm for waveguide, 8um for modulator/add-drop filter
    - Maximum 10Gb/s channel data rate (avoid SerDes)
    - Photonic link data rate density 1Gb/s/um x WDM factor
      - Photonic links have higher density by the WDM factor (number of wavelengths per waveguide)
  - **Example – aggregate throughput 30 Tb/s**
    - From 1000 cores to I/O or shared L2 cache
    - Requires 30 mm of electrical wiring (1 Gb/s/um density)
      - Almost two full metal layers
    - Requires 0.6 mm of photonic bus (with 50 wavelengths per waveguide) – Link 1
      - 9 mm for Link 2,3
  - **Off-chip**
    - Fiber V-groove pitch 0.1mm – same as wirebond pad
      - Best density improvement WDM factor
      - Less with C4 balls – but still > 10x with 50 wavelengths per waveguide
Summary

- Inside the box battle
- All about density and energy-efficiency
- Discrete photonics does not stand a chance
- Si-photonics is the biggest hope
  - Need to see if it can be scaled
Perspective

- Path to a 30 Tb/s, 200 GB+ kiloprocessor on-a-chip interconnect system
  - Density and throughput advantage over electrical
- Circuit-switched vs. packet switched trade-offs
  - Network topology tied to device performance
- Device designs show promise to scale
  - 100-500 fJ/bit energy budgets at 10 Gb/s/channel
  - Device design driven by process information
- Critical to adopt a mainstream process for high-volume applications
  - Processors and DRAM