Challenges in Building a Flat-Bandwidth Memory Hierarchy for a Large-Scale Computer with Proximity Communication

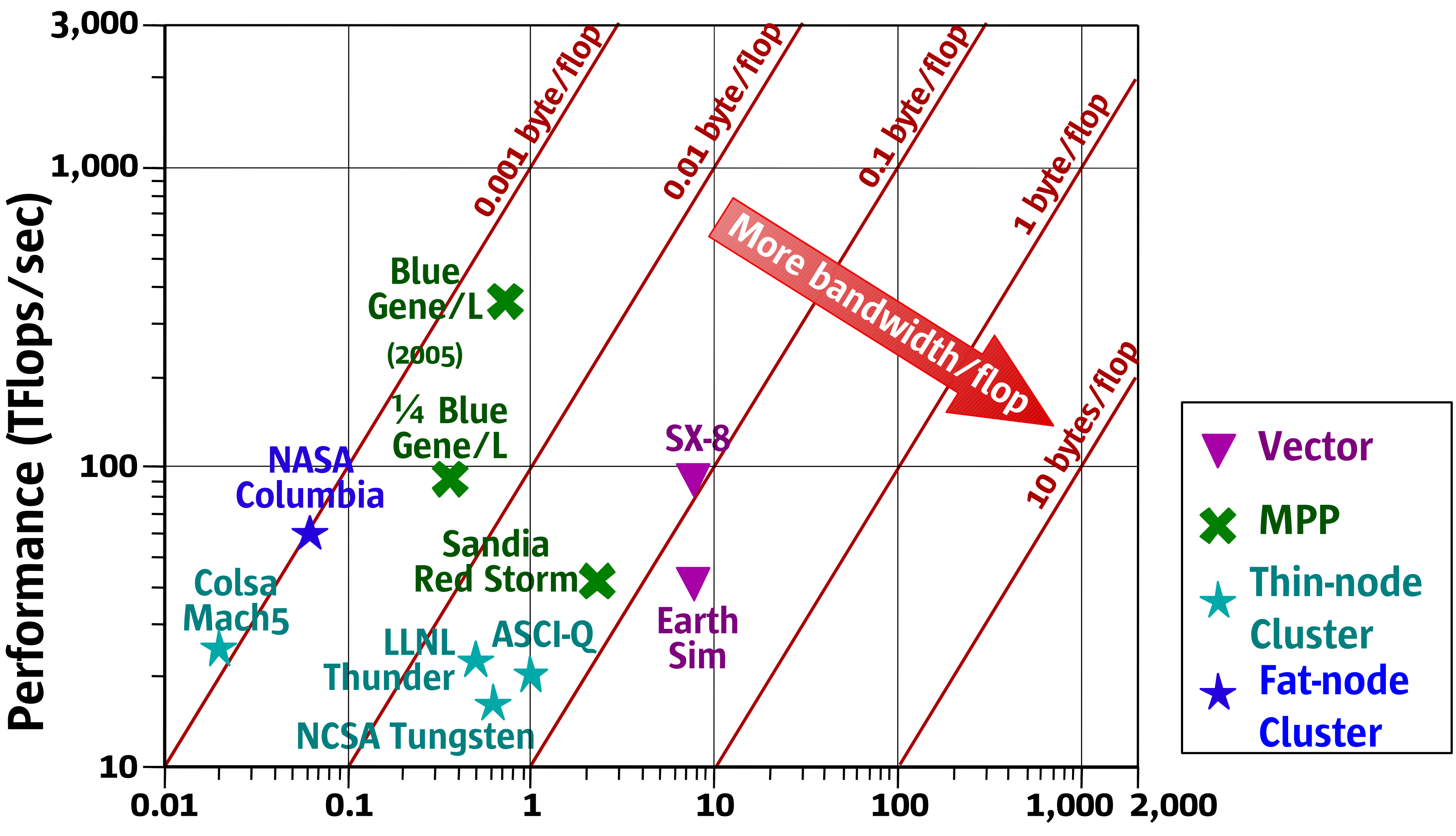
Robert Drost, Craig Forrest, Bruce Guenin, Ron Ho, Ashok V. Krishnamoorthy, Danny Cohen, John E. Cunningham, Bernard Tourancheau, Arthur Zingher, Alex Chow, Gary Lauterbach, and Ivan Sutherland

Sun Microsystems Laboratories

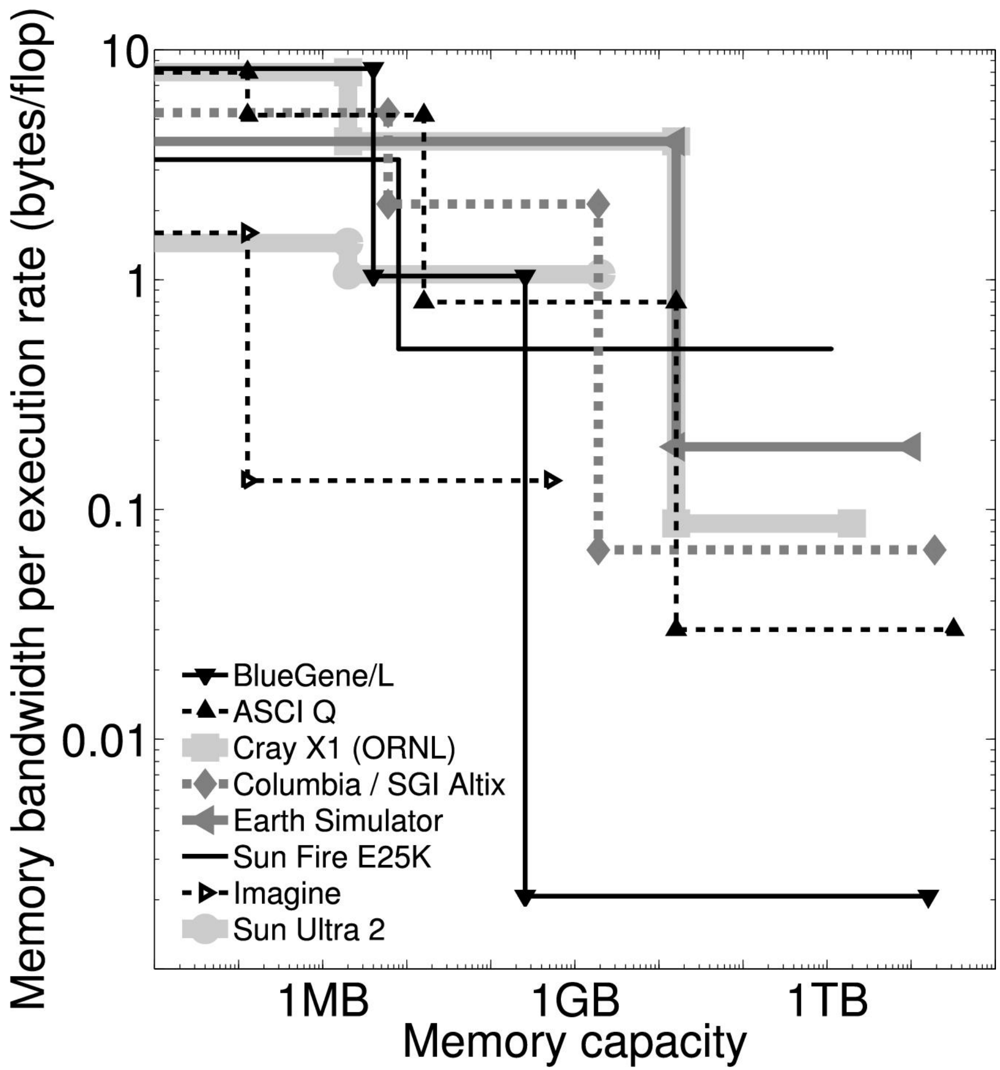
Outline

- **Overview**
- Component technologies to increase bandwidth
  - > Proximity Communication
  - > Low-power on-chip networks
  - > Dense optical communication
- Sea-of-Anything transport layer
- Summary

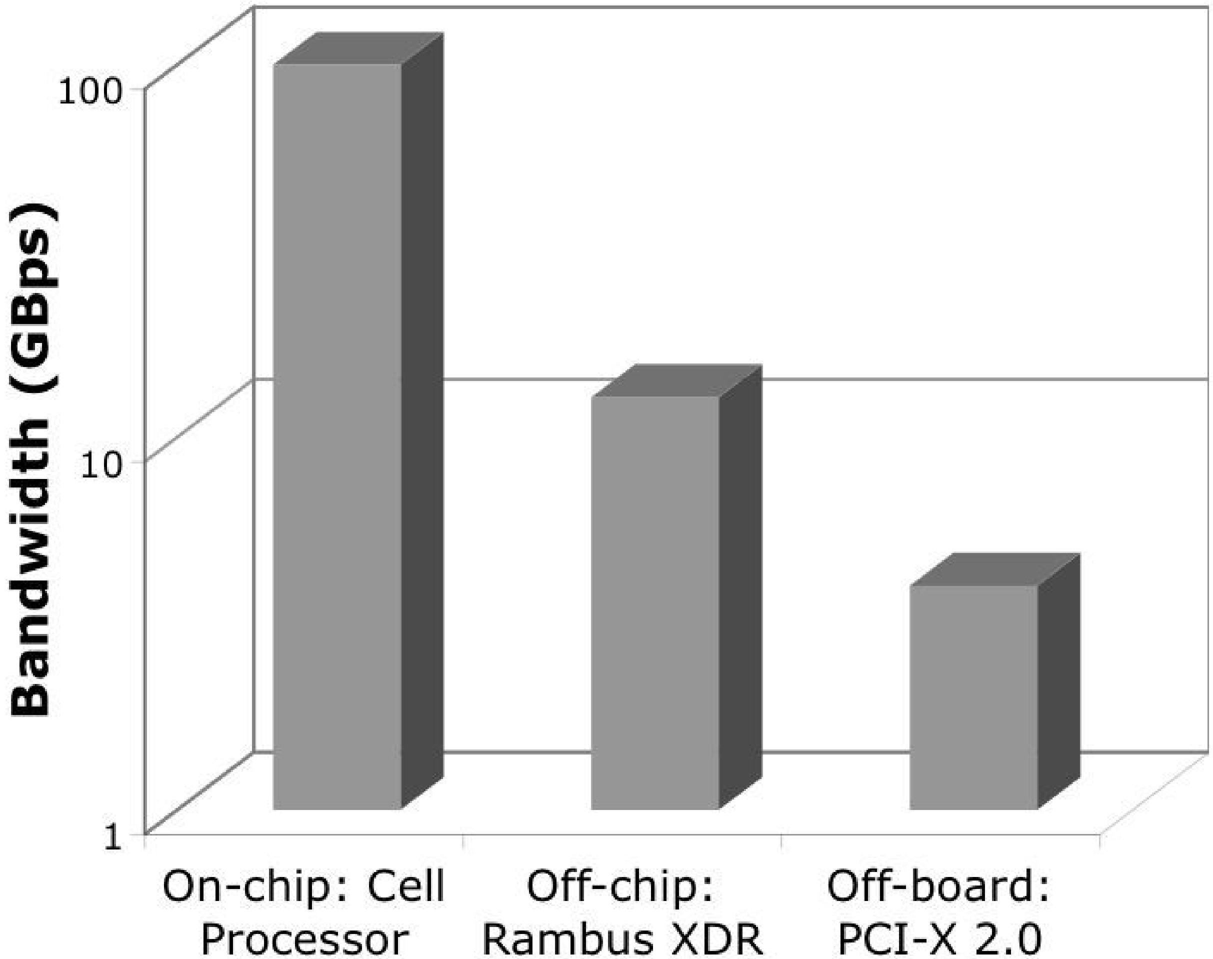
Motivation: A balanced architecture



Bandwidth versus Memory Capacity



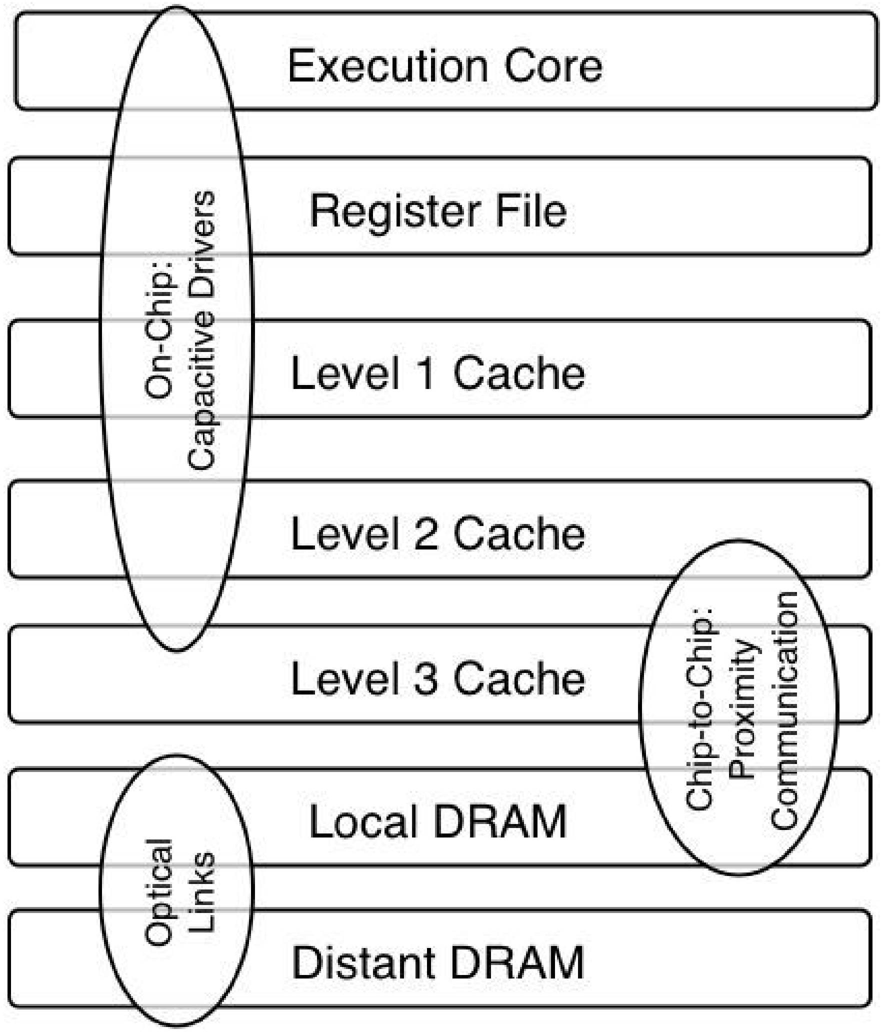
Example Physical Bandwidths



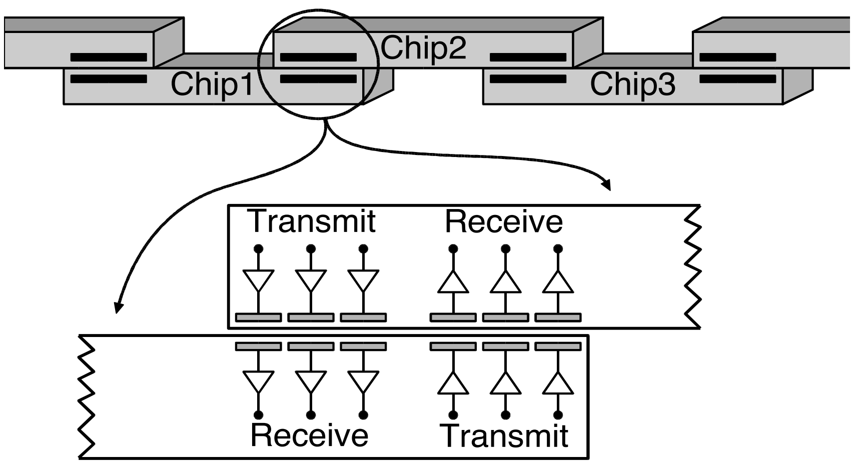
Outline

- Overview
- **Component technologies to increase bandwidth**
  - **Proximity Communication**
  - **Low-power on-chip networks**
  - **Dense optical communication**
- Sea-of-Anything transport layer
- Summary

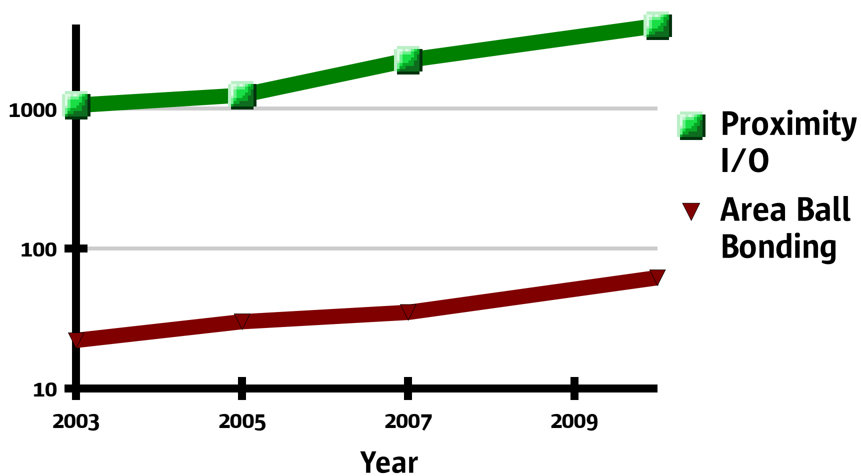
Physical to Functional Mapping

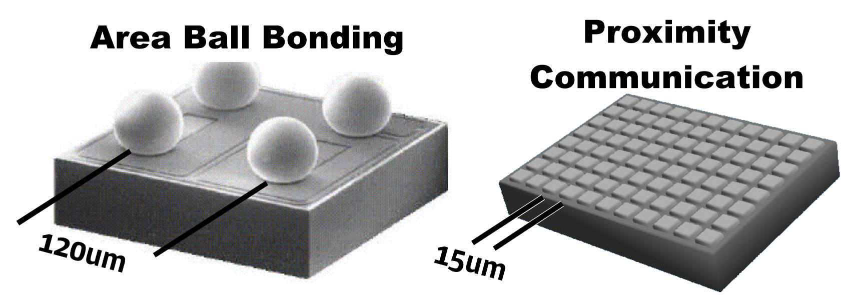


Proximity Communication



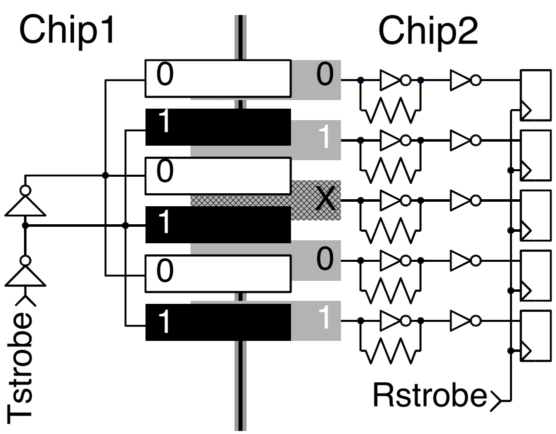
- Avoids Off-Chip Wires
- Increases Bandwidth/Area
- Makes Chips Replaceable
- Enhances Testing Capability
- Enables Smaller Chips
- Obviates ESD Protection
- Shrinks Transceiver Circuits



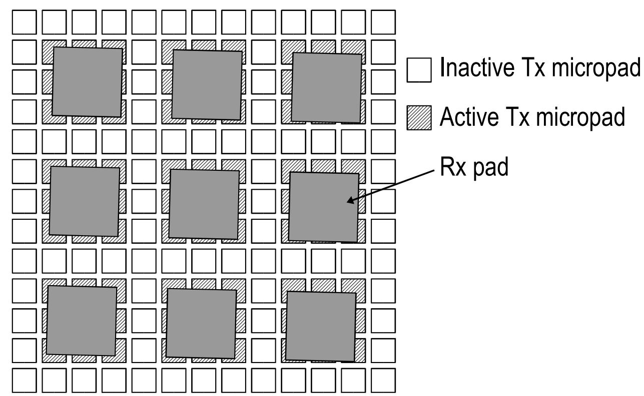


Alignment is the major challenge

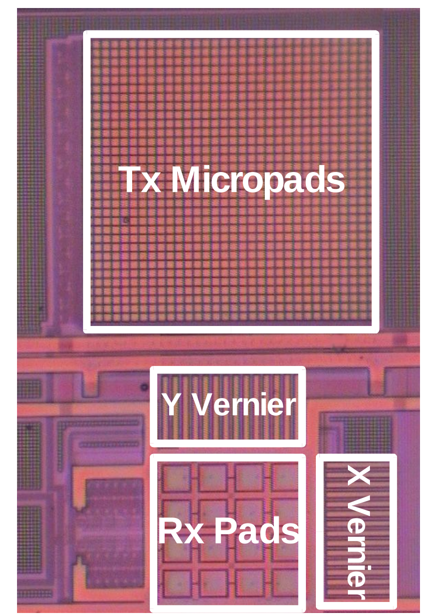
- Must align chips in X, Y, Θ, Z, Ψ, Φ
- X, Y, Θ misalignments are corrected electronically



Measure...



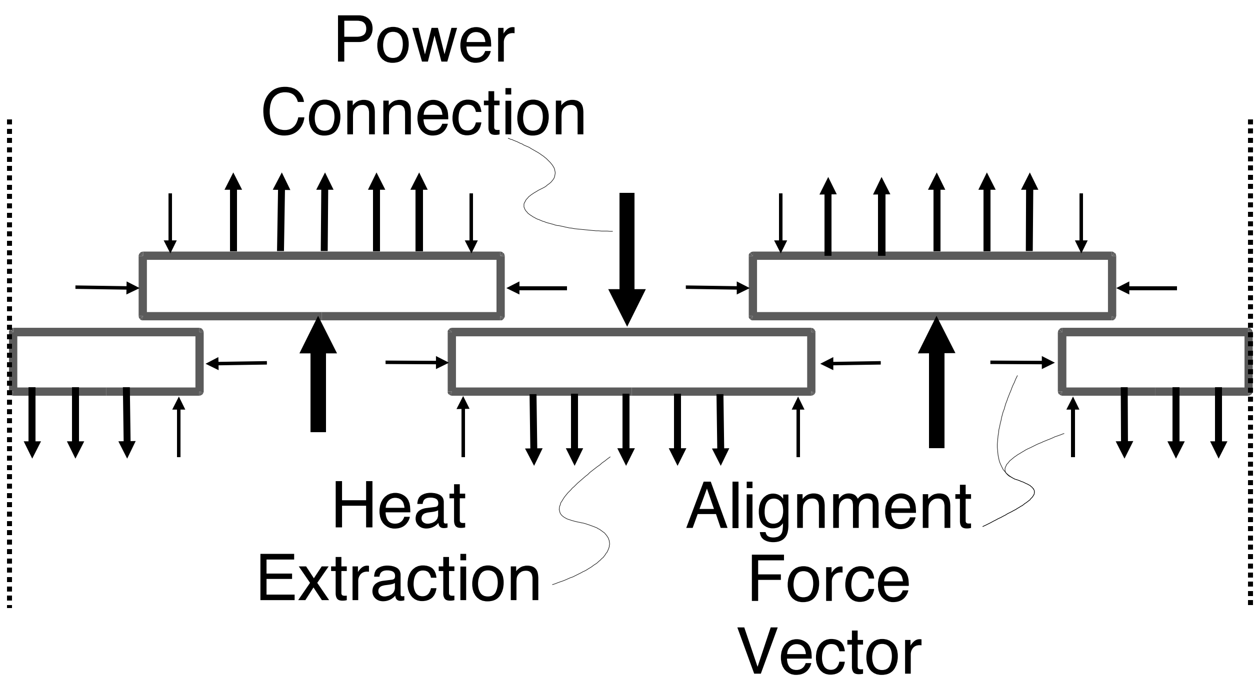
...and correct...



...on-chip

Proximity Packaging Challenges

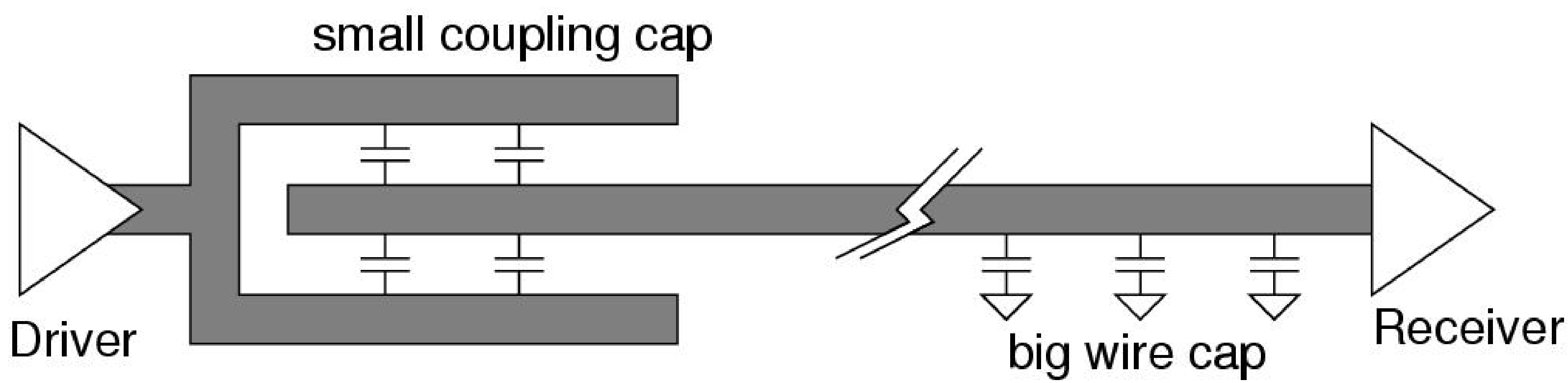
- Performance is a function of Z, Ψ, Φ misalignments



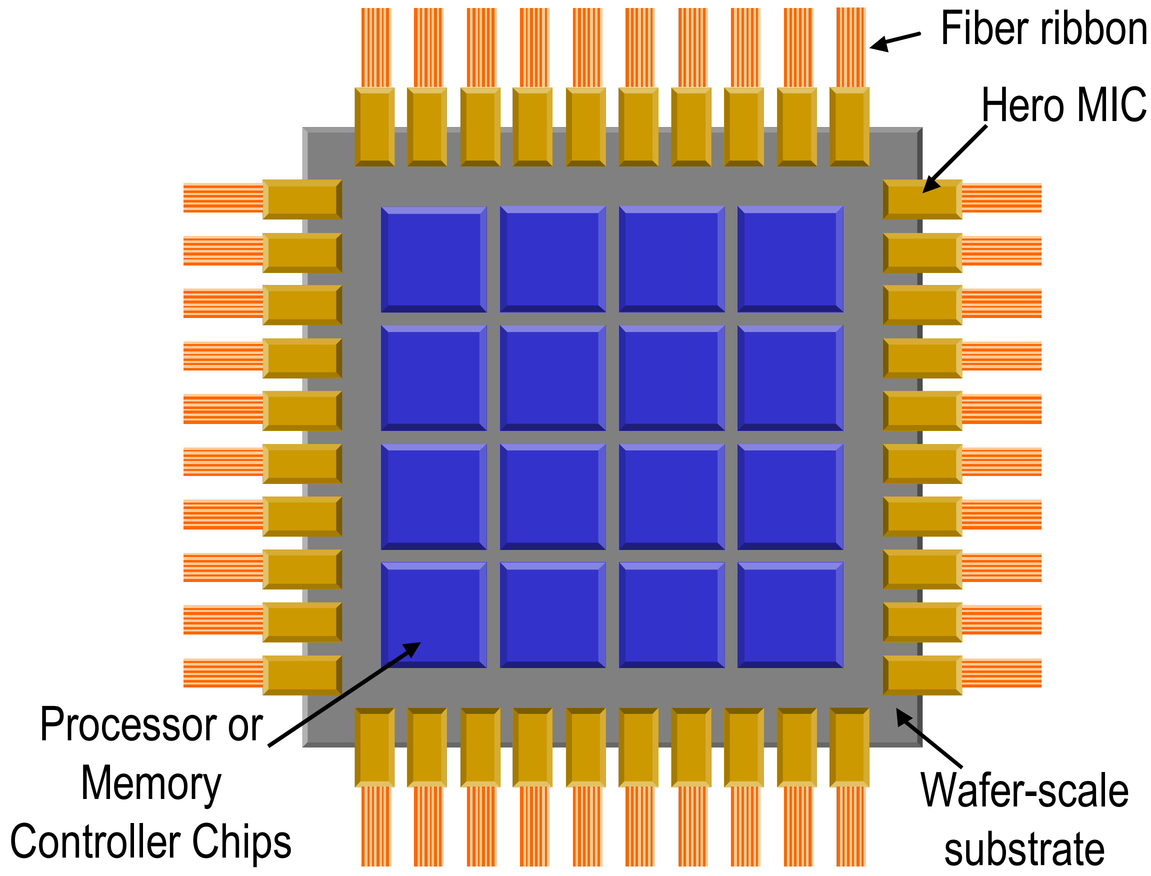
- With reasonable misalignment control tens of Tbps bandwidth per chip can be realized

Low-power on-chip signaling

- Gains in bandwidth lead to high power as well unless efficiency of signaling is improved...
- Capacitively coupled on-chip drivers reduce power, avoid the use of a low-voltage supply, pre-emphasize to correct R-C signal distortion



Off-module optical signaling



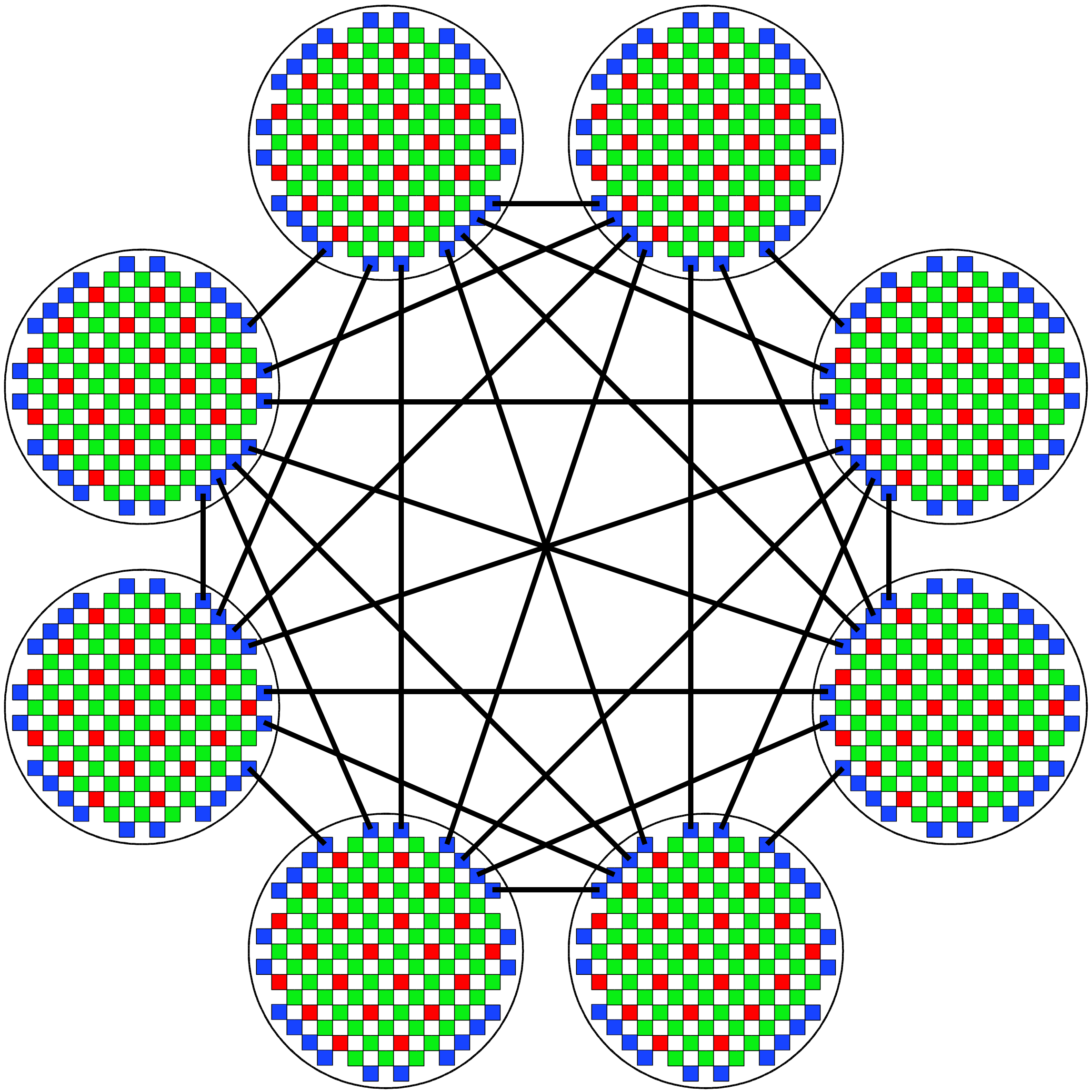
Outline

- Overview
- Component technologies to increase bandwidth
  - > Proximity Communication
  - > Low-power on-chip networks
  - > Dense optical communication
- **Sea-of-Anything transport layer**
- **Summary**

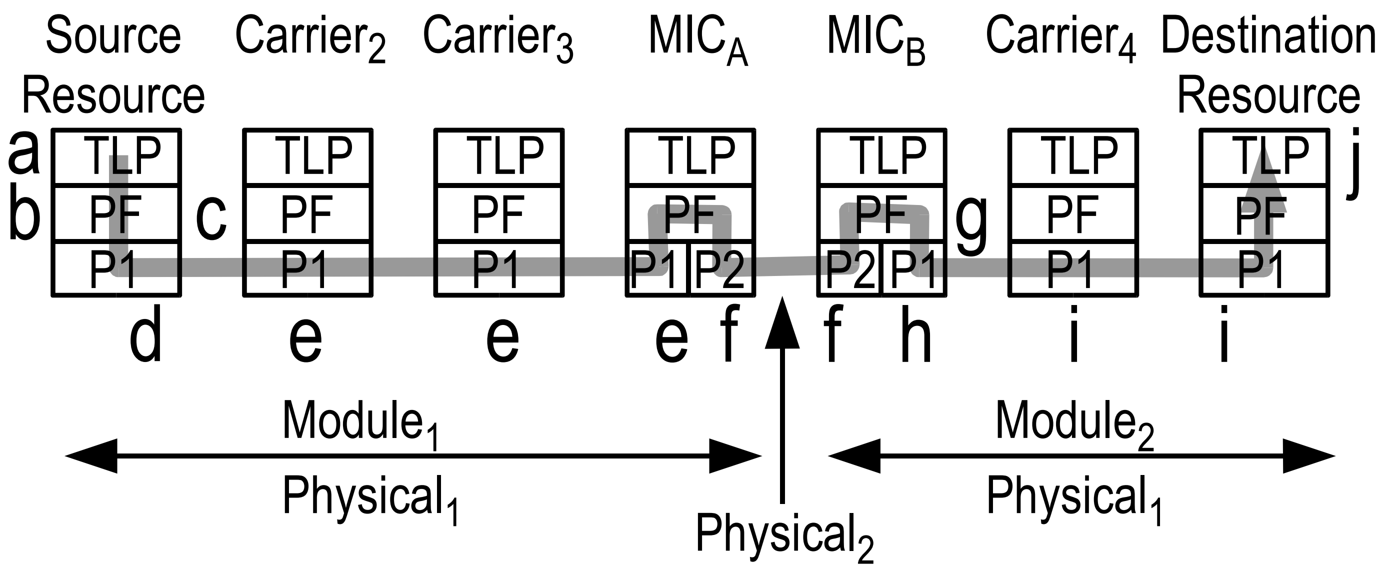
Sea-of-Anything (SoX) Interconnect

- Distributed switching transport interconnect
- Enables any resources to communicate using
  - > Proximity Communication Manhattan meshes
  - > Point-to-point optical links
- Low latency
  - > Cut-through switching
  - > Worm-hole routing
  - > Consistent packet format
- Deadlock avoidance

Multi-module Interconnect



SoX sample bus cycle transaction



PF = Path Formation
TLP = Transaction Level Protocol

P1 = Physical_1
P2 = Physical_2

Summary

- Examined memory bottleneck in large-scale computers
- Explored three high-bandwidth technologies
  - Proximity Communication
  - Low-power on-chip networks
  - Dense optical communication
- Discussed provably deadlock-free transport layer, SoX
- Together these can maintain high bytes/flop across the full memory footprint of a large-scale computer

Challenges in Building a Flat-Bandwidth Memory Hierarchy for a Large-Scale Computer with Proximity Communication

Robert Drost, Ph.D.
robert.drost@sun.com

Speaking Points

Outline: Talk about balanced system design, where the balance is processing versus bandwidth. VLSI scaling affords us tremendous processing power improvements through parallelized processing cores. However communicating data to keep these cores doing useful work is much more challenging.

Motivation: A balanced architecture

This chart presents balanced architecture. Y axis is processing performance, X is bisection bandwidth. A series of diagonal lines illustrate balances of 10 bytes/flop all the way down to .001 bytes/flop. You can see that systems like the Earth Sim in Japan provided notable performance increase by moving to the right on this graph. In contrast IBM with Blue Gene/L went vertical, not attemptingto provide bisection bandwidth comensurate with performance improvements