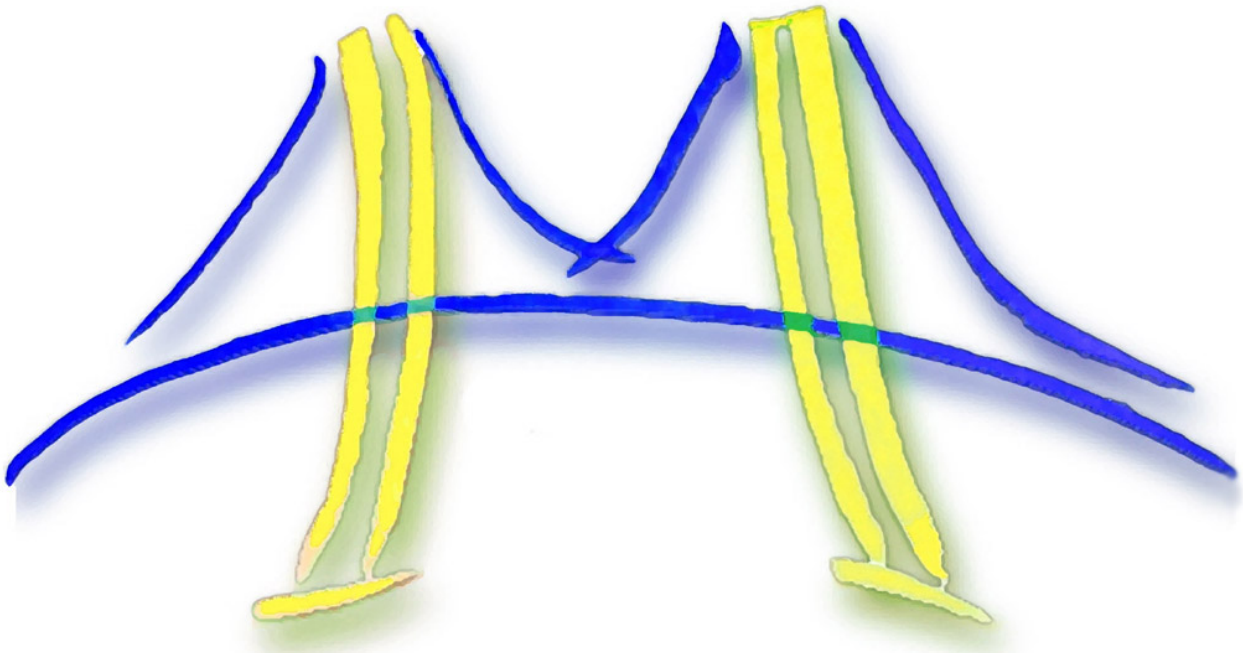
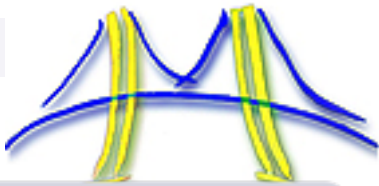


Performance and Energy Comparison of Electrical and Hybrid Photonic Networks for CMPs

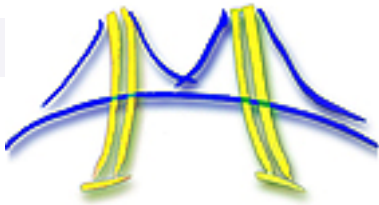
Ankit Jain, Shoaib Kamil, Marghoob Mohiyuddin, John Shalf, John Kubiatowicz

UC Berkeley ParLab/LBNL



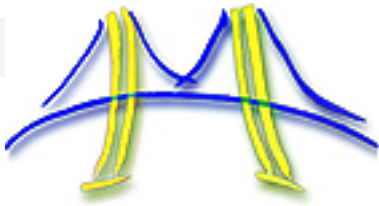
Introduction	• Motivation • Contributions • Baseline Architecture
Electrical Network	• Architecture • Simulator • Model
Hybrid Network	• Architecture • Simulator • Model
Studied Communications	• Synthetic Communication • Real Applications
Results	• Synthetic Results • Application Results • Process to Processor Mapping
Conclusions and Future Work	

Motivation



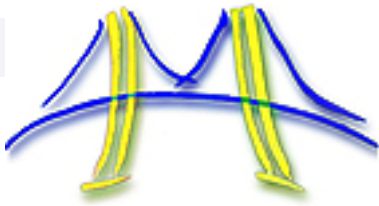
- Manycore: NoCs key to translating raw performance → sustained performance
- Electrical NoC performance/energy constrained by process technology
  - Also, every joule saved counts
- Photonic NoC promising
  - Enabled by recent advances in photonics & chip fabrication
  - Potentially high performance at low energy cost
  - But cannot do packet switching
- Use hybrid network
  - Small packets ➔ electrical NoC
  - Large packets ➔ optical NoC

Contributions

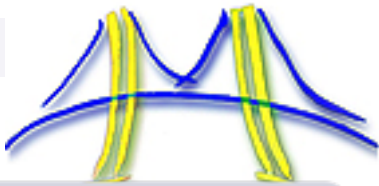


- Use both synthetic traces and real application traces to compare electrical vs. hybrid photonic networks
- Construct cycle-accurate simulators and compare with simple analytic models
- Programmability: How important is process-to-processor mapping?

Baseline Architecture



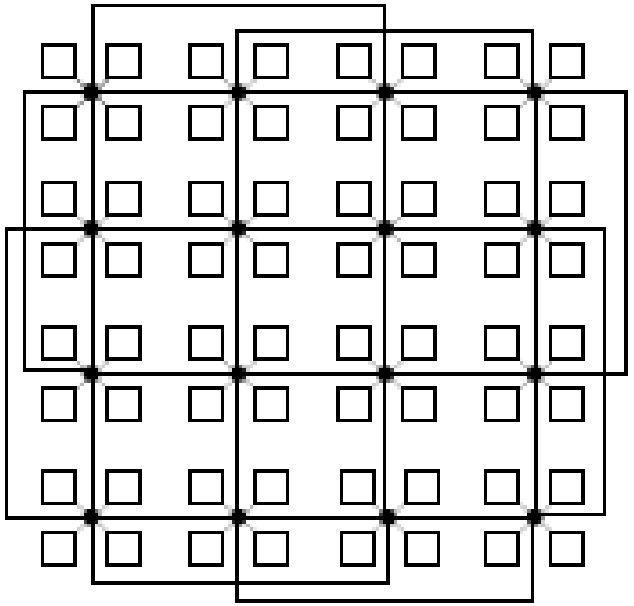
- 64 small, homogenous cores on a CMP
- Cores ~1.5mm x 1.5mm
- 22nm process, 5GHz
- 3D Integrated CMOS
  - layer for processors, layers for memory
- We examine two interconnect architectures to compare performance & energy efficiency



Introduction	• Motivation • Contributions • Baseline Architecture
Electrical Network	• Architecture • Simulator • Model
Hybrid Network	• Architecture • Simulator • Model
Studied Communications	• Synthetic Communication • Real Applications
Results	• Synthetic Results • Application Results • Process to Processor Mapping
Conclusions and Future Work	

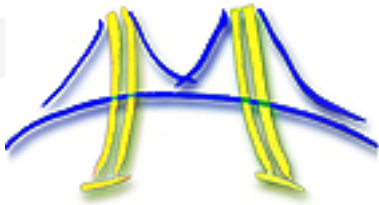
Electrical NoC

- Bill Dally's CMesh topology
- Wormhole routed
- Virtual channels
- Single electrical layer with multiple memory layers



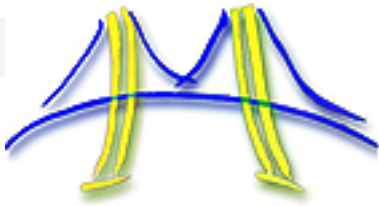
(b) CMesh

Electrical Simulator



- Processor
  - Ignore computation
  - Communication divided into “phases” (SPMD-style)
    - Send and receive all messages in a phase as fast as possible
- Router
  - XY dimension order routing
  - Express links on periphery
  - Virtual channels & wormhole routing
  - Credit based flow control
  - 8 input ports → 8x8 switch

Analytic Model for Electrical NoC

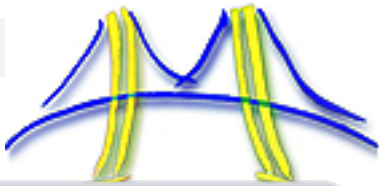


- Time
  - Bandwidth-only model
  - Assume virtual channels + wormhole routing hide latency

$$T_{msg} = \frac{size_{msg}}{bandwidth}$$

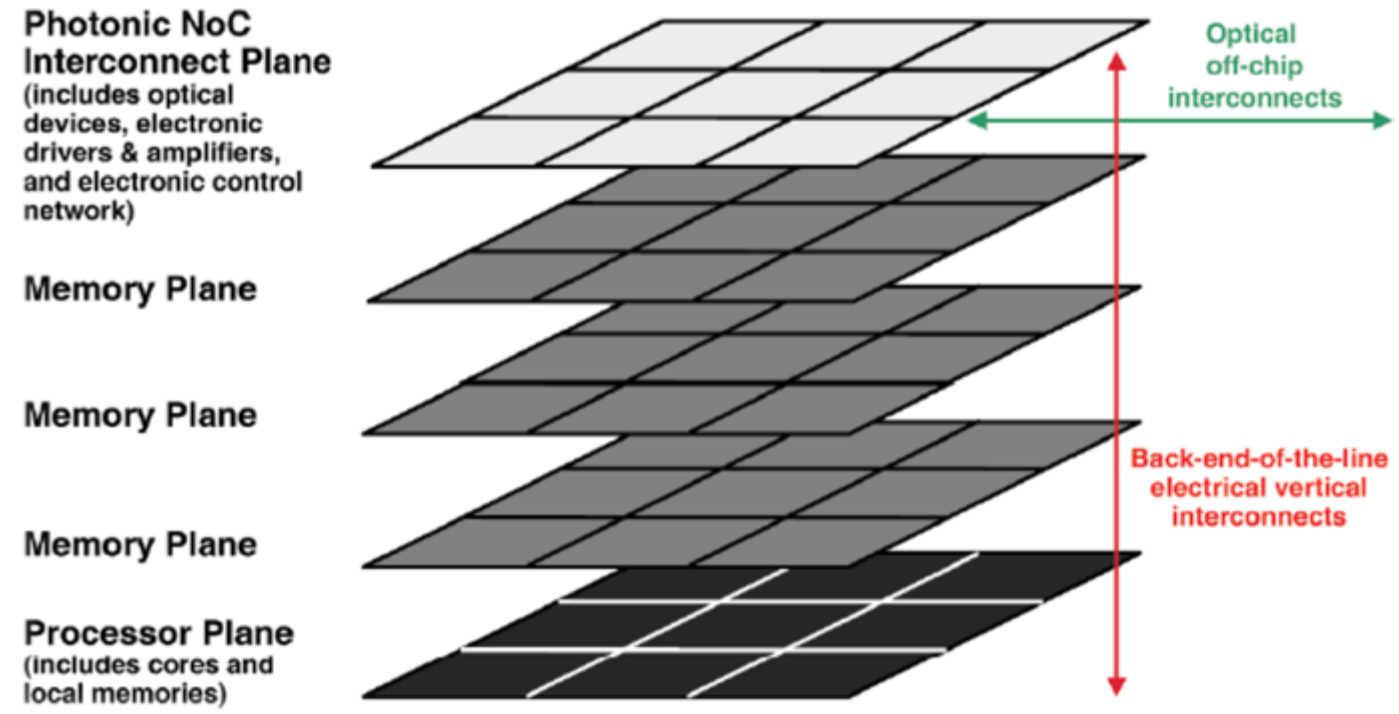
- Energy
  - Each hop incurs a set amount of energy
  - Link crossing + Router traversal
  - Parameters from Dally et al, scaled via ITRS

$$E_{total} = \sum (|L_{msg}| \times E_{hop})$$



Introduction	• Motivation • Contributions • Baseline Architecture
Electrical Network	• Architecture • Simulator • Model
Hybrid Network	• Architecture • Simulator • Model
Studied Communications	• Synthetic Communication • Real Applications
Results	• Synthetic Results • Application Results • Process to Processor Mapping
Conclusions and Future Work	

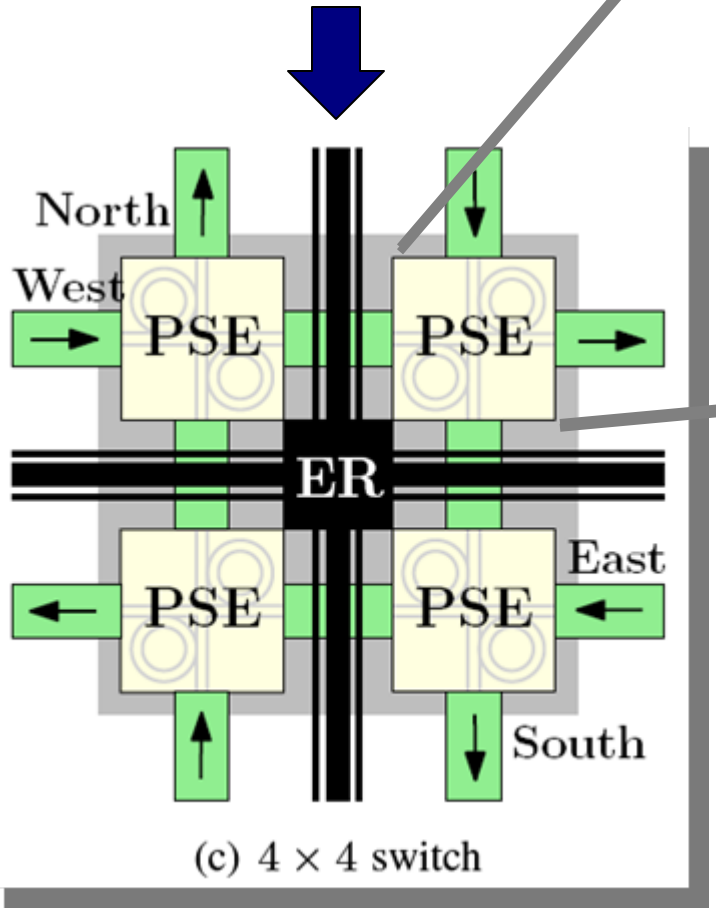
Hybrid NoC



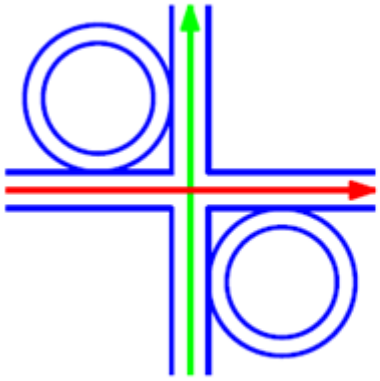
- Mesh Topology
- “Electrical Control Network” (ECN) on Processor Plane
- Multiple optical networks on Photonic Plane
- Small setup messages on ECN and bulk data transfer on optical network

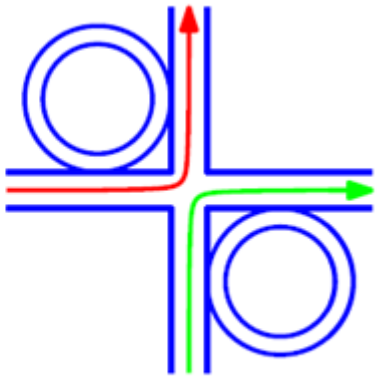
Blocking Photonic Switch

Capable of routing a single path from any source to any destination



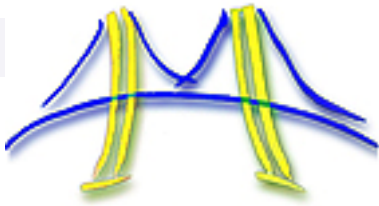
(c) 4 × 4 switch


(a) Off state


(b) On state

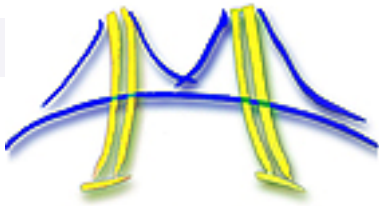
- On ➔ message turns
- No inactive power consumption
- Small switching cost
- Small active power while switched on

Deadlock in Hybrid NoC



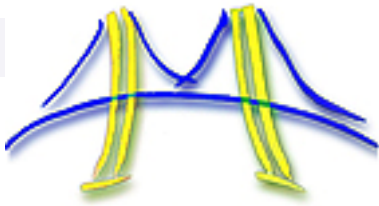
- Blocking 4x4 switch
  - Only one path can be routed at a time through a switch
- Deadlock is a known issue in circuit switching. Avoid deadlock with:
  - Exponential backoff
  - Dimension order routing
  - Multiple optical networks
    - Results in more possible paths
    - Since photonic elements are quite small, this is doable

Hybrid Simulator



- 1:1 processor to electrical router mapping
  - Each electrical router buffers up to 8 path setup messages from its corresponding processor
  - Electrical router does not use virtual channels or wormhole routing (unnecessary and consume energy)
- Path setup packets are minimally sized: take one cycle to traverse between 2 routers
- Energy includes Electro-Opto-Electrical conversions at the endpoints
  - Most expensive operation energy-wise
  - Did not include off-chip laser energy cost

Analytic Model for Hybrid NoC

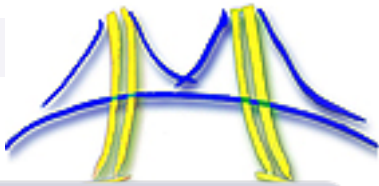


- Time
  - Must account for latency of electrical network, bandwidth limits, and contention

$$T_{msg} = |L_{msg}| \times 2 \times latency_{electrical} + \frac{size_{msg}}{bandwidth_{optical}}$$

  - For contention, serialize “most-used” link
    - Only one message can be sent along link at a time
    - Overall time is time to send all messages on busiest link
- Energy
  - Each message incurs energy cost on electrical network, plus the costs on the photonic network

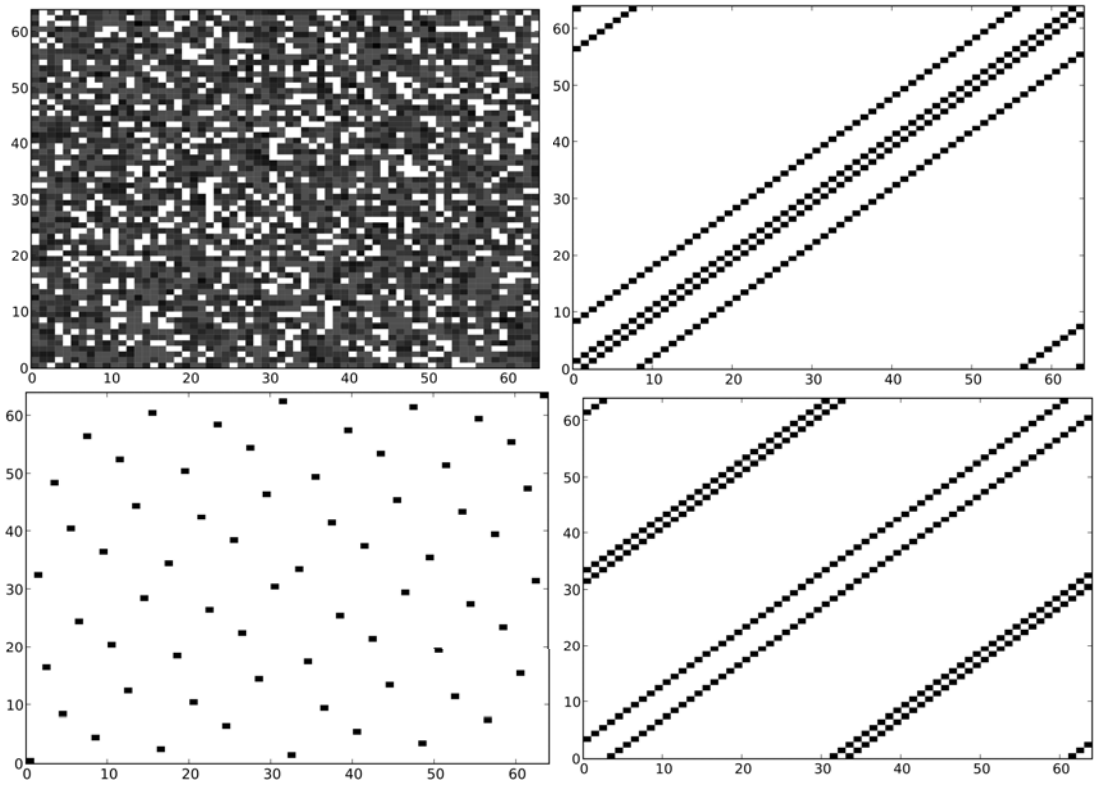
$$E_{total} = \sum (|L_{msg}| \times E_{electrical} + E_{PSEswitching} + T_{msg} \times E_{PSEactive} + E_{EOE} \times size_{msg})$$



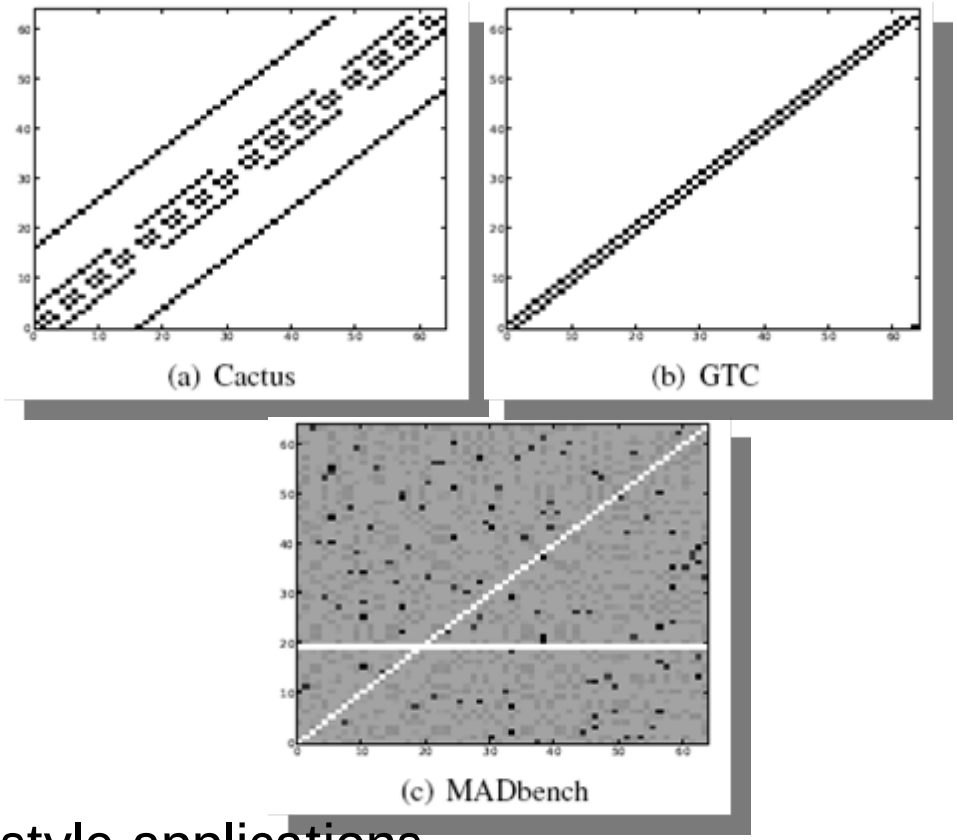
Introduction	• Motivation • Contributions • Baseline Architecture
Electrical Network	• Architecture • Simulator • Model
Hybrid Network	• Architecture • Simulator • Model
Studied Communications	• Synthetic Communication • Real Applications
Results	• Synthetic Results • Application Results • Process to Processor Mapping
Conclusions and Future Work	

Synthetic Traces

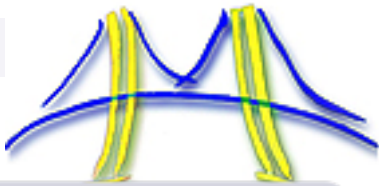
- Random messages
- Nearest-Neighbor
- Bitreverse
- Tornado
- Look at both small & large messages



Real Applications

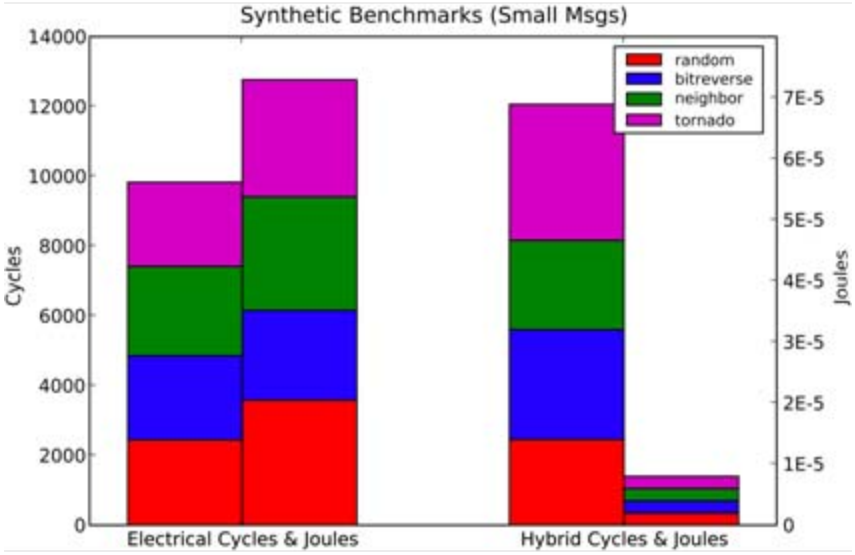


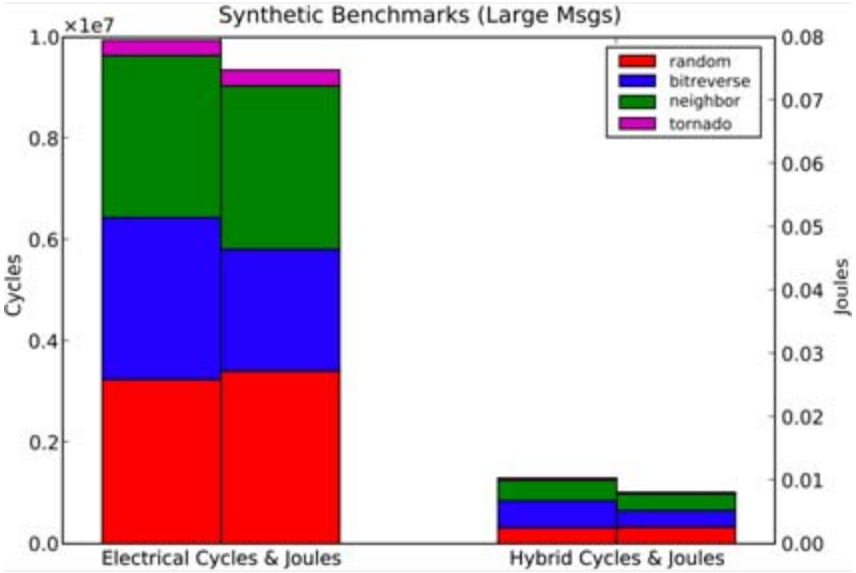
- SPMD style applications
- From DOE/NERSC workloads
- Broken into multiple phases of communication
  - implicit barrier is assumed at the end of a communication phase



Introduction	• Motivation • Contributions • Baseline Architecture
Electrical Network	• Architecture • Simulator • Model
Hybrid Network	• Architecture • Simulator • Model
Studied Communications	• Synthetic Communication • Real Applications
Results	• Synthetic Results • Application Results • Process to Processor Mapping
Conclusions and Future Work	

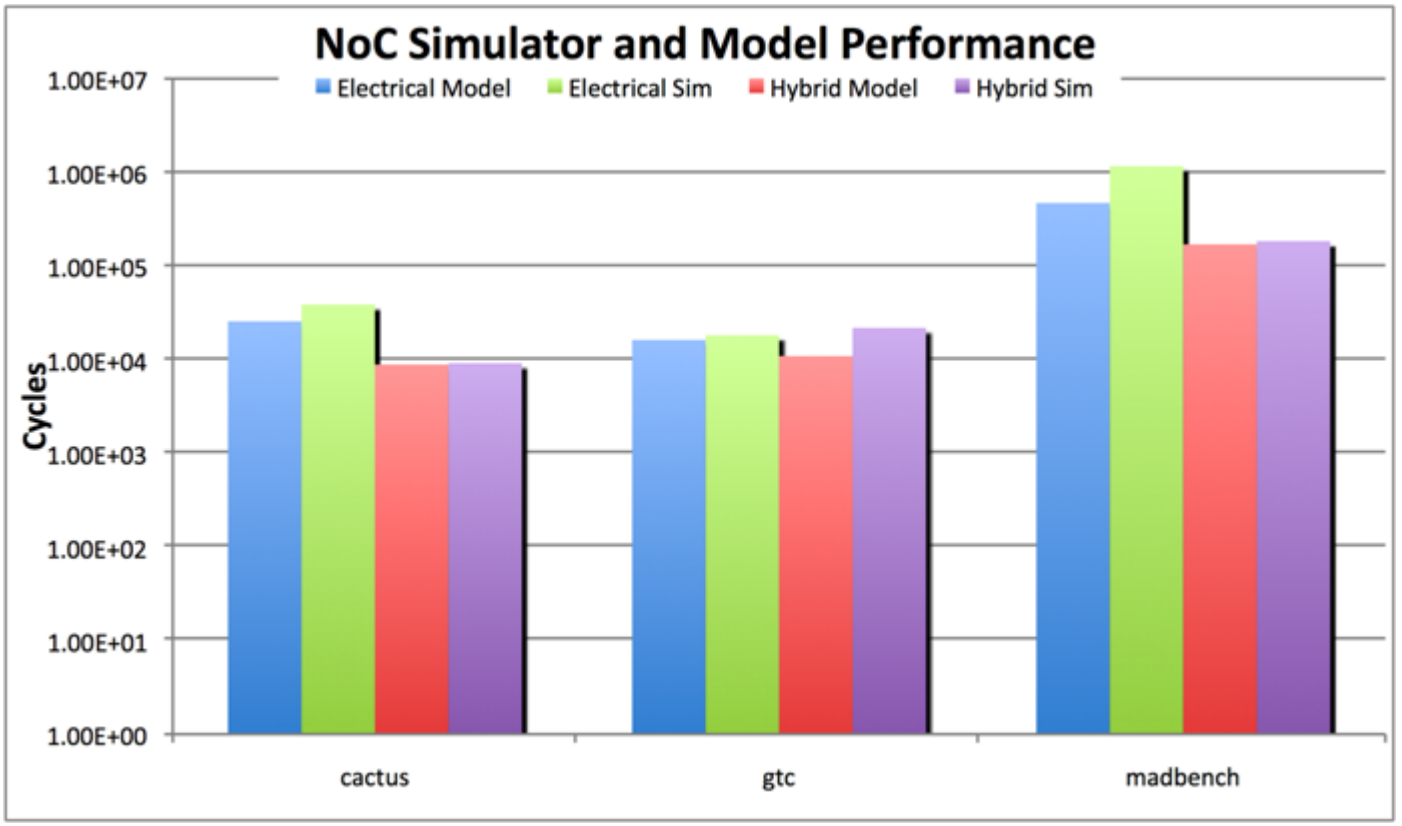
Synthetic Trace Results



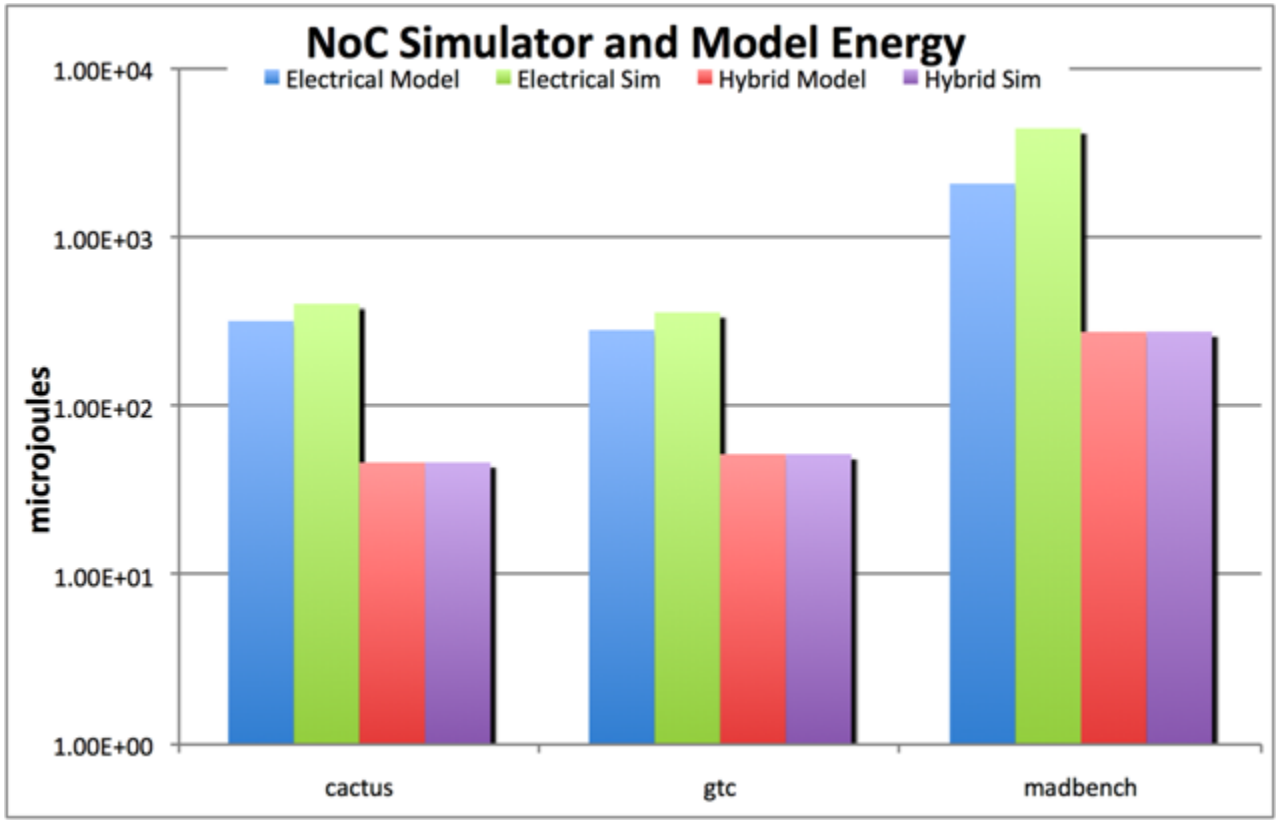


- For small messages, setup latency for the hybrid network makes it slower than electrical
- Hybrid network outperforms electrical-only on large messages, and uses far less energy in both cases

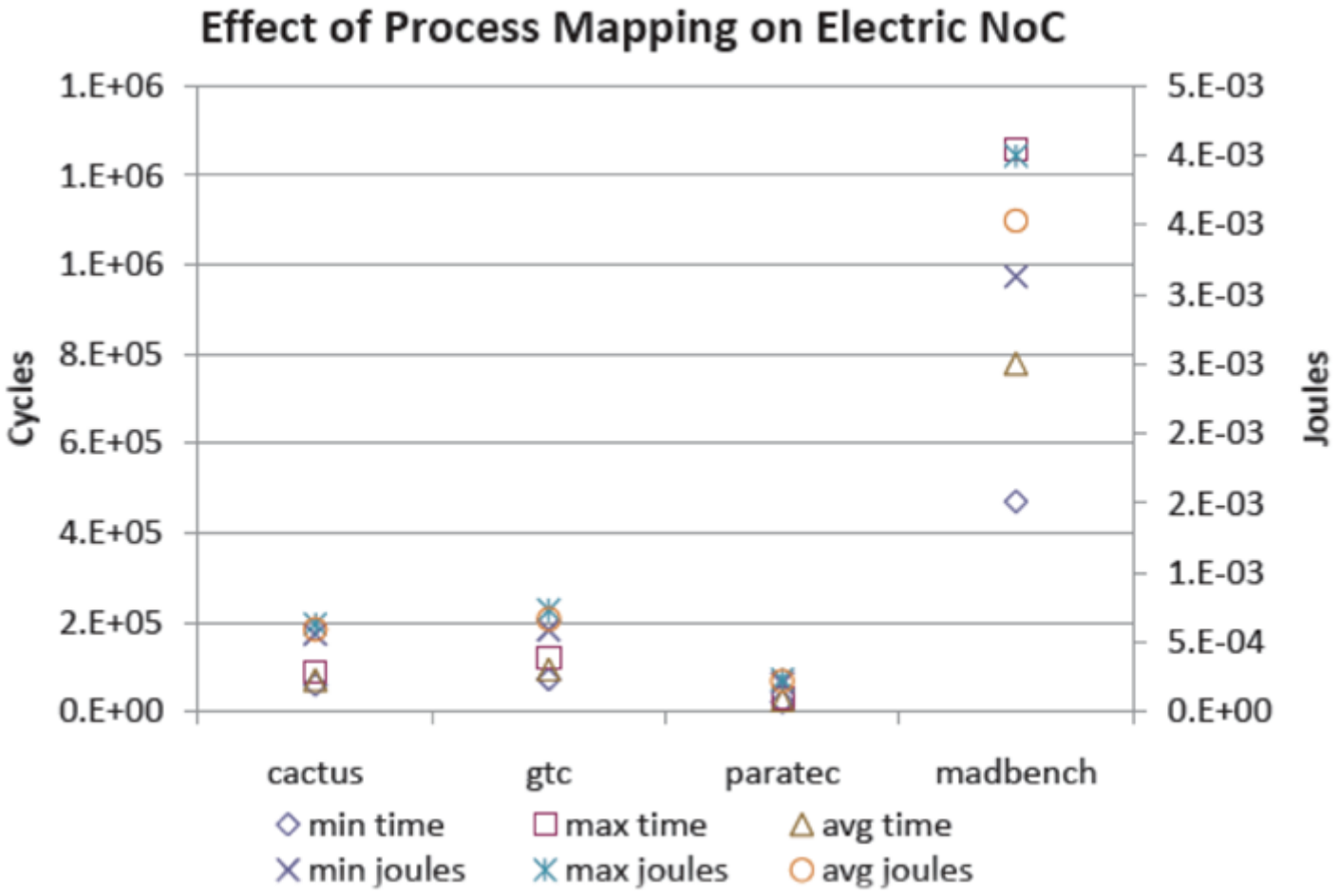
Application Performance



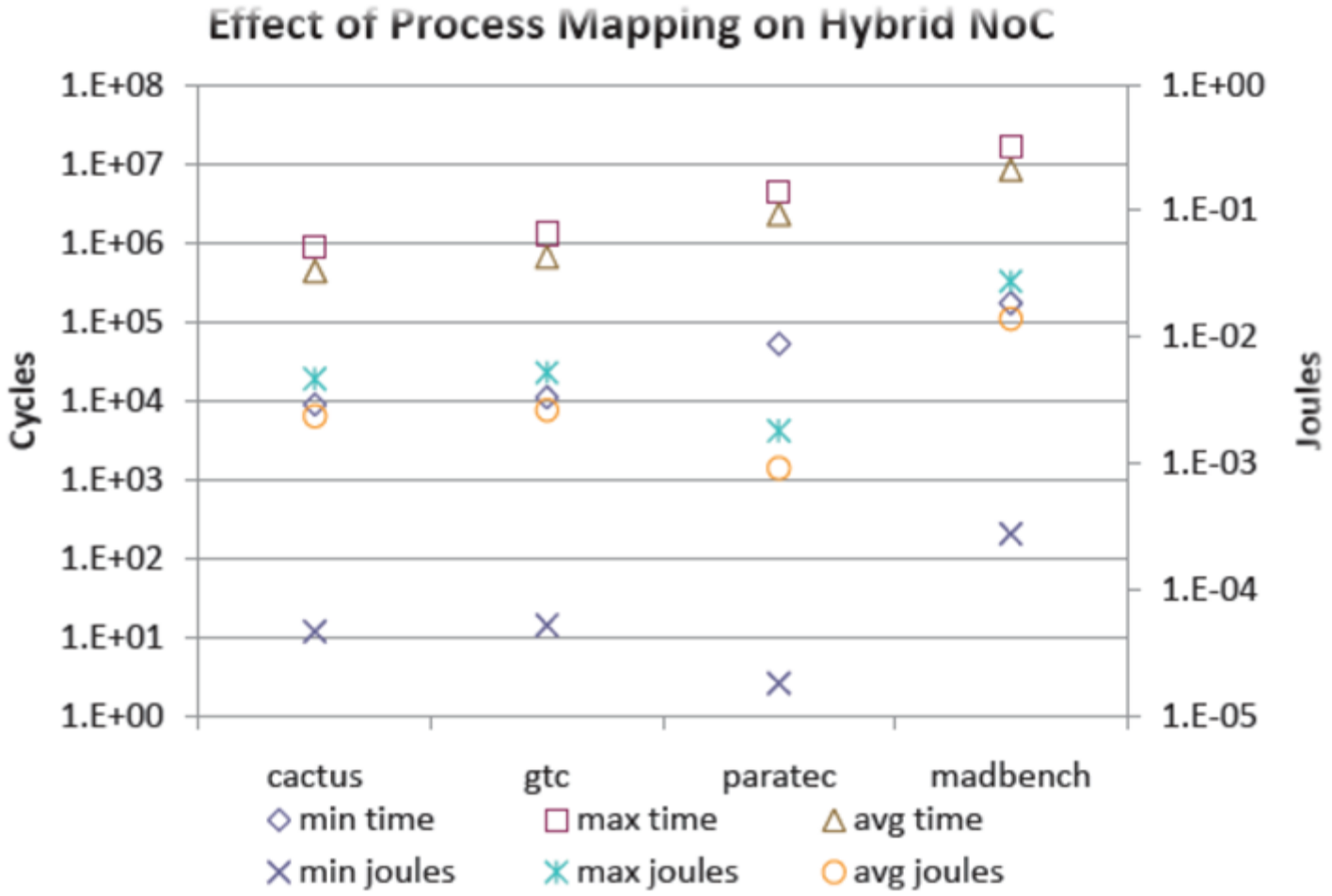
Application Energy

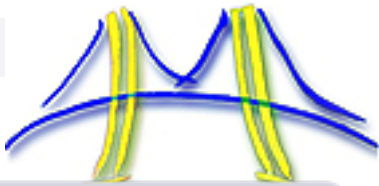


Process-Processor Mapping (1/2)



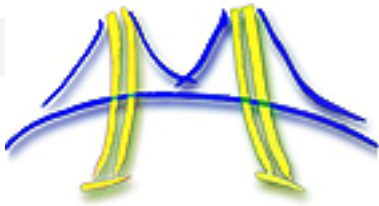
Process-Processor Mapping (2/2)





Introduction	• Motivation • Contributions • Baseline Architecture
Electrical Network	• Architecture • Simulator • Model
Hybrid Network	• Architecture • Simulator • Model
Studied Communications	• Synthetic Communication • Real Applications
Results	• Synthetic Results • Application Results • Process to Processor Mapping
Conclusions and Future Work	

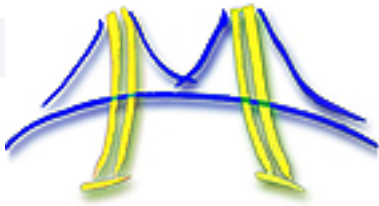
Conclusions



- Simple analytic models accurately predict both performance and energy consumption
- Hybrid NoC: Majority of energy due to Optical-to-Electrical and Electrical-to-Optical conv. (>94%).
- Hybrid NoC performs better for larger messages; energy consumption is much lower
- Process-to-processor mapping can significantly impact performance as well as energy consumption.
  - Finding the optimal mapping is not always of utmost importance— making sure not to use a ‘bad’ mapping is.

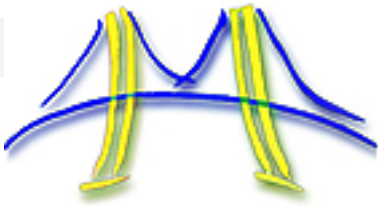
- Overall, hybrid photonic on-chip networks are promising

Future Work



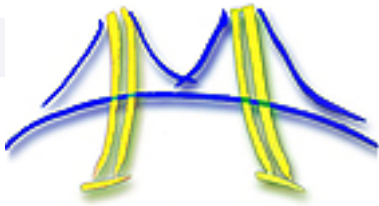
- Non-blocking optical mesh interconnection network
- Account for data transfer onto chip
- More accurate full system simulators (for both performance and energy)
  - simulate FP operations & memory traffic
  - as photonic technologies are explored by materials/hardware designers, use input to revise/refine simulators
- Explore applications with less synchronous communication models
  - Not SPMD
  - Overlap of computation and communication

Acknowledgements

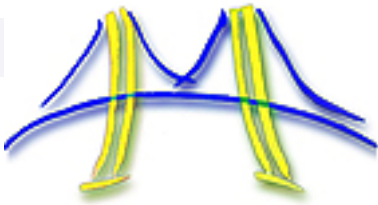


- Katherine Yelick (UC Berkeley ParLab & NERSC/LBNL)
- Assam Schacham, Luca Carloni and Dr. Keren Bergman (Columbia University)
  - Our exploration is based on their earlier work (see references)
- BeBOP Research Group (UC Berkeley Computer Science Dept)

References



- [1] Assaf Shacham, Keren Bergman, and Luca Carloni. On the Design of a Photonic Network-on-Chip. In Proceedings of the First International Symposium on Networks-on-Chip, 2007.
- [2] James Balfour, and William Dally. Design Tradeoffs for Tiled CMP On-Chip Networks. In Proceedings of the International Conference on Supercomputing, 2006.
- [3] Shoaib Kamil, Ali Pinar, Daniel Gunter, Michael Lijewski, Leonid Oliker, and John Shalf. Reconfigurable Hybrid Interconnection for Static and Dynamic Applications. In Proceedings of the ACM International Conference on Computing Frontiers, 2007.
- [4] Bergman et. al.. Topology Exploration for Photonic NoCs for Chip Multiprocessors. Unpublished to date.
- [5] Cactus Homepage. http://www.cactuscode.org, 2004.
- [6] Z. Lin, S. Ethier, T.S. Hahm, and W.M. Tang. Size Scaling of Turbulent Transport in Magnetically Confined Plasmas. Phys. Rev. Lett., 88, 2002.
- [7] Julian Borrill, Jonathan Carter, Leonid Oliker, David Skinner, and R. Biswas. Integrated performance monitoring of a cosmology application on leading hec platforms. In Proceedings of the International Conference on Parallel Processing (ICPP), 2005.
- [8] A. Canning, L.W. Wang, A. Williamson, and A. Zunger. Parallel Empirical Pseudopotential Electronic Structure Calculations for Million Atom Systems. J. Comput. Phys., 160:29, 2000.
- [9] Xiaoye S. Li and James W. Demmel. SuperLU-dist: A Scalable Distributed-Memory Sparse Direct Solver for Unsymmetric Linear Systems. ACM Trans. Mathematical Software, 29(2):110140, June 2003.
- [10] J. Qiang, M. Furman, and R. Ryne. A Parallel Particle-in-Cell Model for Beam-Beam Interactions in High Energy Ring Colliders. J. Comp. Phys., 198, 2004.
- [11] IPM Homepage. http://www.nersc.gov/projects/ipm, 2005



Backup Slides

Analytic Model

- Three Models
  - Bandwidth Model
    - For electrical network: assume virtual channels hide latency
  - Bandwidth + Latency Model
  - Bandwidth + Latency + Contention Model

ELECTRICAL HYBRID

$$T_{msg} = \frac{size_{msg}}{bandwidth}$$

$$T_{msg} = |L_{msg}| \times 2 \times latency_{electrical} + \frac{size_{msg}}{bandwidth_{optical}}$$

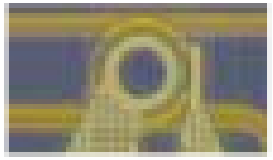
$$E_{total} = \sum (|L_{msg}| \times E_{hop})$$

$$E_{total} = \sum (|L_{msg}| \times E_{electrical} + E_{PSEswitching} + T_{msg} \times E_{PSEactive} + E_{EOE} \times size_{msg})$$

Silicon Photonics Today

Needed for on-chip

Ring modulator
Radius 30μm

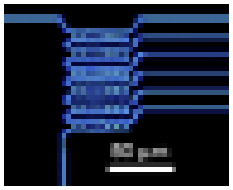


Area 25X
Power 10X

Radius 6μm

WDM filters
1.4mmx0.6mm=0.85mm²
Add. power for tuning

Area 150X
No active tuning
Thermally stable



50x100μm=0.005mm²
No active feedback, accuracy is defined by fabrication

Deflection Switch
Macroscopic

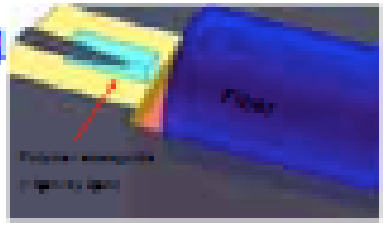
Nanoscale
No active tuning
Thermally stable

30° temp range

On-chip optical amplifier
Macroscopic
Not integrated

Mass manufacturable III-V on Si
(~50/chip)
Wafer scale

Optical Coupling
Hand-aligned
Low numbers



Mass manufacturable alignment (~50 fibers)
Package in presence of C4s and heatsink

Electrical Simulator (2/2)

- Channels
  - Buffering at both ends
  - Maximum wire length = side of processor core

ELECTRICAL SIMULATOR PARAMETERS

Model Parameter	Sim Parameter	Value
$latency_{electrical}$	Router Latency	2 cycles
	Router-Router Link Latency	2 cycles
	Virtual channels	1,2,4,6
	Buffer size in flits	1,2,3,4
	Frequency	5 GHz
$bandwidth_{electrical}$	Electrical Bandwidth	640 Gb/sec
$E_{electrical}$	Joules Per Electrical Hop	0.82e-12

Hybrid Simulator (2/2)

HYBRID SIMULATOR PARAMETERS

Model Parameter	Sim Parameter	Value
$latency_{electrical}$	Router Latency	2 cycles
	Router-Router Link Latency	1 cycle
	Path Multiplicity	1,2,4
	Time To Timeout	2,10,20
	Frequency	5 GHz
$bandwidth_{optical}$	Optical Bandwidth	960 Gb/sec
$E_{PSESwitching}$	Joules Per PSE Switching	1.0e-12
$E_{PSEActive}$	Joules Per PSE on Per Second	1.0e-6
$E_{electrical}$	Joules Per Electrical Hop	0.82e-12
E_{EOE}	Joules Per Bit EOE[1]	0.4e-12

Parameter Exploration: Electrical NoC

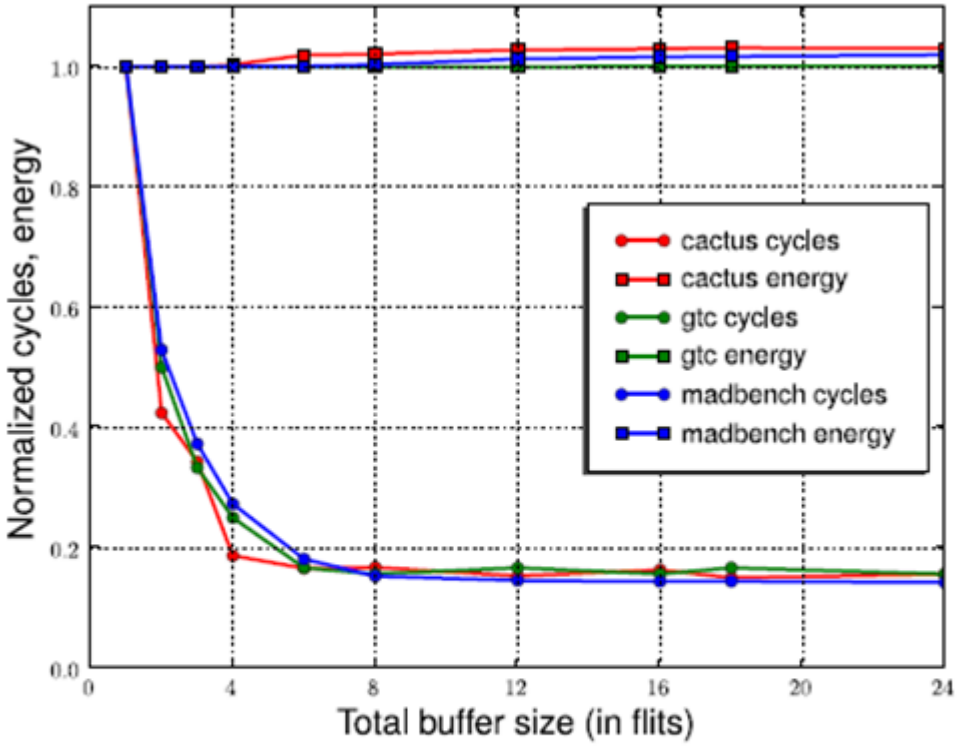
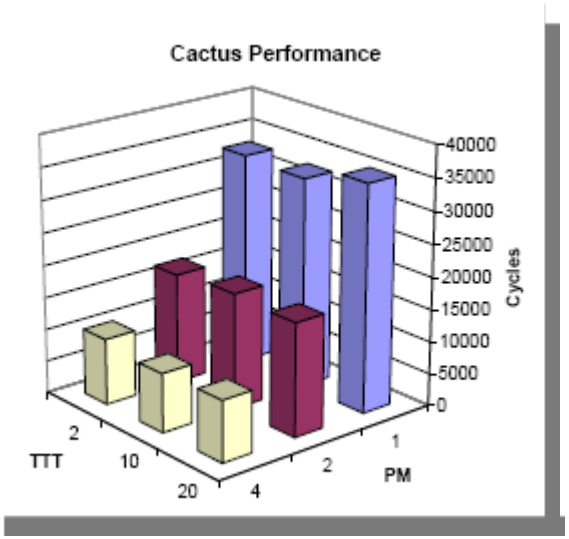
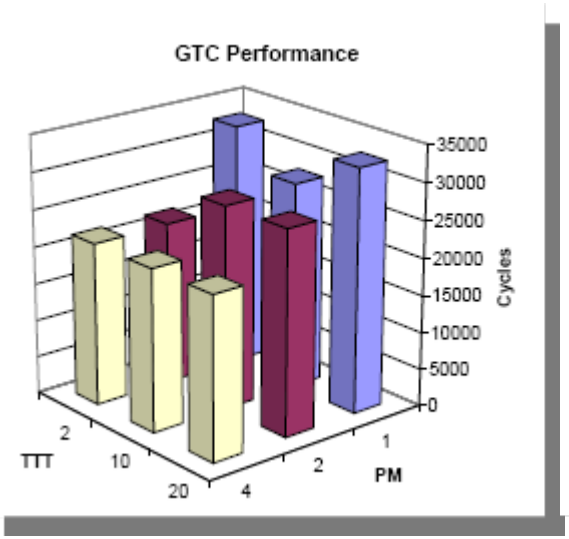


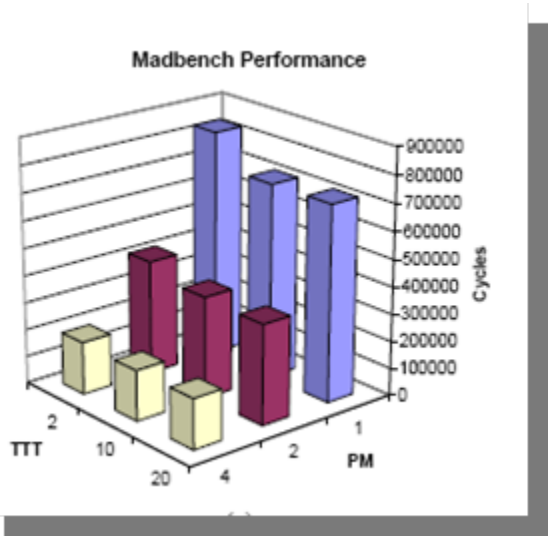
Fig. 5. Parameter exploration results for electrical NoC. Results are for communication phases only. The energy and cycle numbers are normalized w.r.t. total buffer size of 1 case (lower is better).

Total buffer size = #vcs X buffer size ➔ router

Parameter Exploration: Hybrid NoC







- Sensitive to path multiplicity
  - more available paths = less contention
- Timeouts prevent over- and under-waiting

NoC as Part of a System

- Use Merrimac FP unit numbers
- Scale to 22nm using ITRS roadmap
- Trace methodology records FP Operations
- Compare energy used in FP unit vs energy used in interconnect