INTERCONNECTION NETWORKS

Mahdi Nazm Bojnordi

Assistant Professor

School of Computing

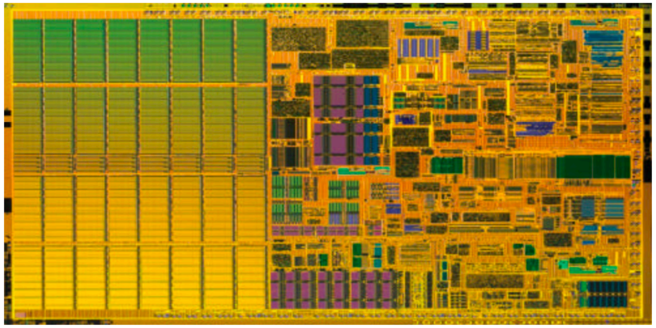
University of Utah

Overview

- Upcoming deadline
  - Feb.3rd: project group formation
    - Nathan Page and Bhavani Priya Sampath Kumar
    - Tanmay Tirpankar and Hunter Jensen
    - Ryan West, Anthony Chyr, and Jacob Larkin
- This lecture
  - Cache interconnects
  - Basics of the interconnection networks
  - Network topologies
  - Flow control

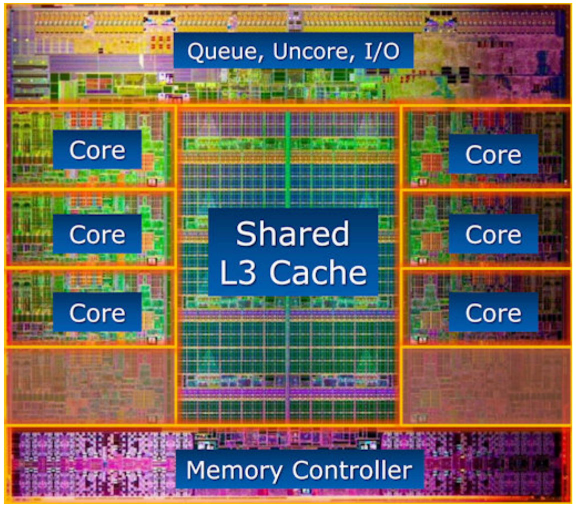
Where Interconnects Are Used?

- About 60% of the dynamic power in modern microprocessors is dissipated in on-chip interconnects



- Analysis subject: Processor, 0.13 [μm]
- 77 million transistors, die size of 88 [mm^2]
- Data sources (AF, Capacitance, Length)
- Excluded: L2 cache, global clock, analog units

[Magen'04]



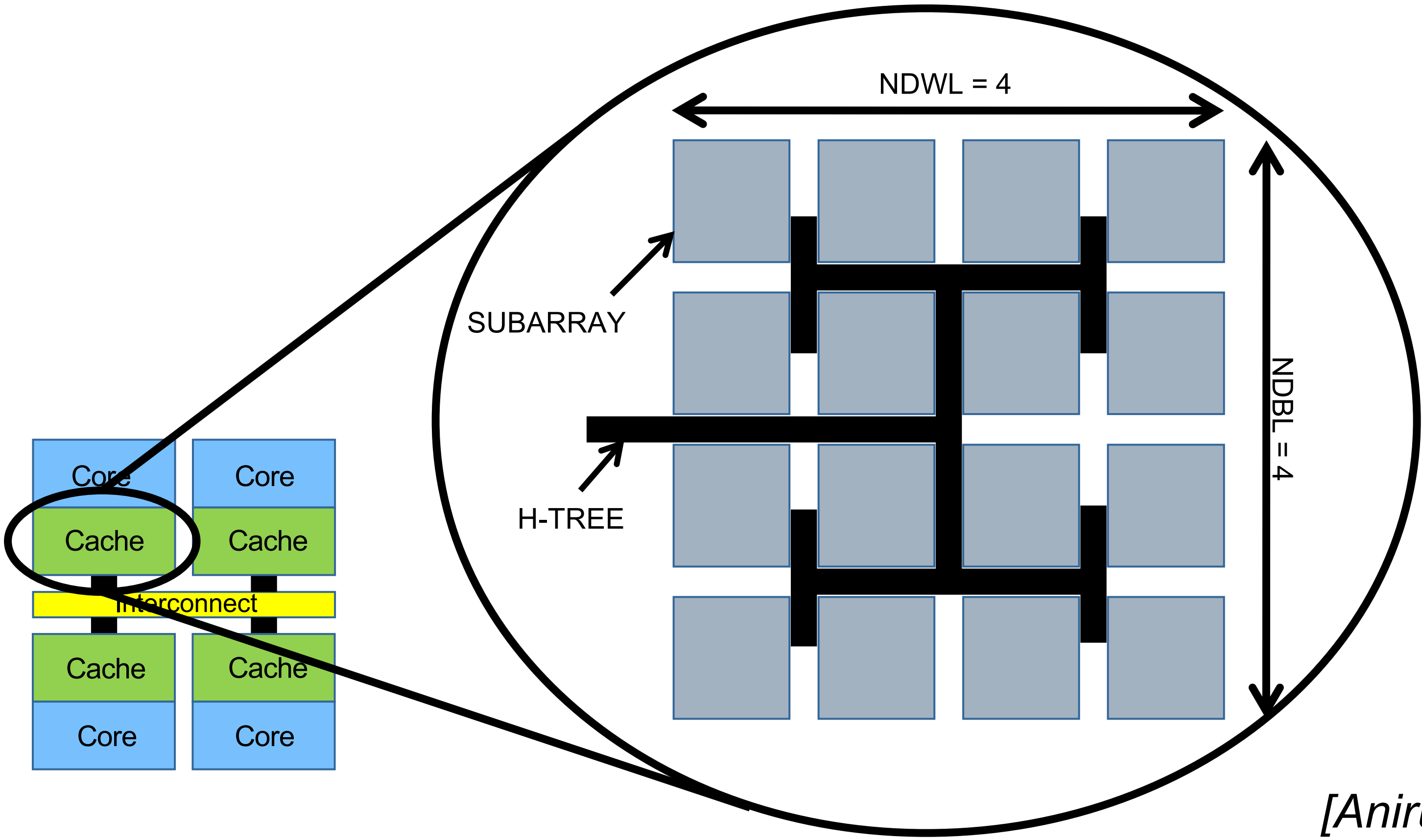
- Six processor cores
- 8MB Last level cache

[Intel Core i7]

Cache Interconnect Optimizations

Large Cache Organization

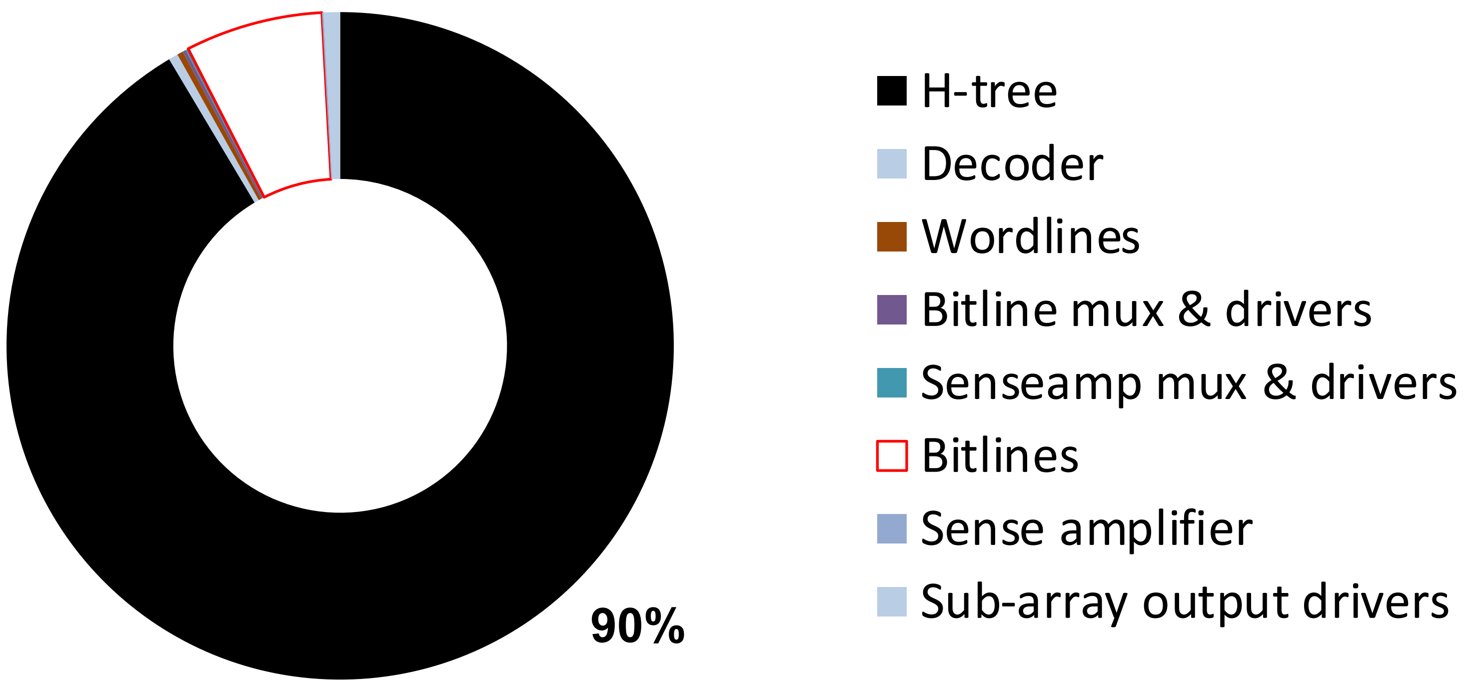
- Using fewer subarrays gives increased area efficiency, but larger delay due to longer wordlines/bitlines



[Aniruddha'09]

Large Cache Energy Consumption

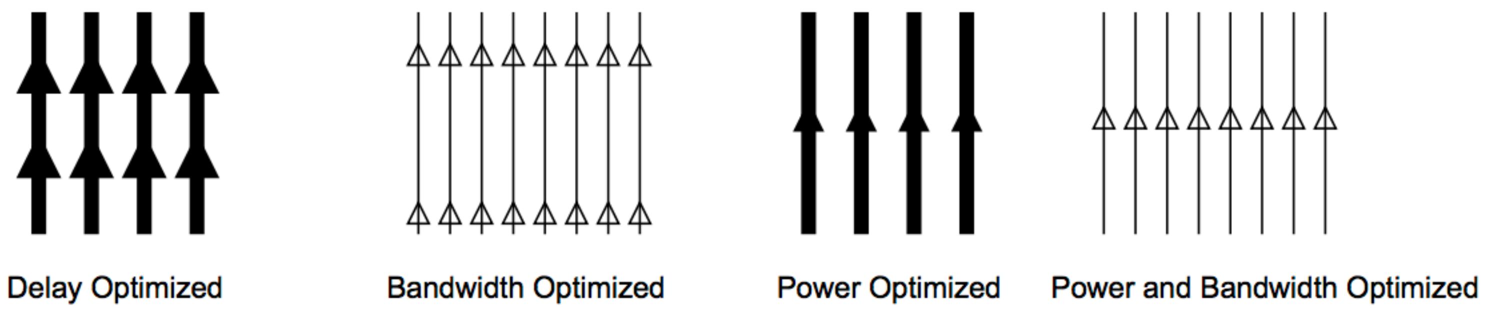
- H-tree is clearly the dominant component of energy consumption



[Aniruddha'09]

Heterogeneous Interconnects

- A global wire management at the microarchitecture level
- A heterogeneous interconnect that is comprised of wires with varying latency, bandwidth, and energy characteristics



[Balasubramonian'05]

Heterogeneous Interconnects

- Better energy-efficiency for a dynamically scheduled partitioned architecture
  - ED^2 is reduced by 11%

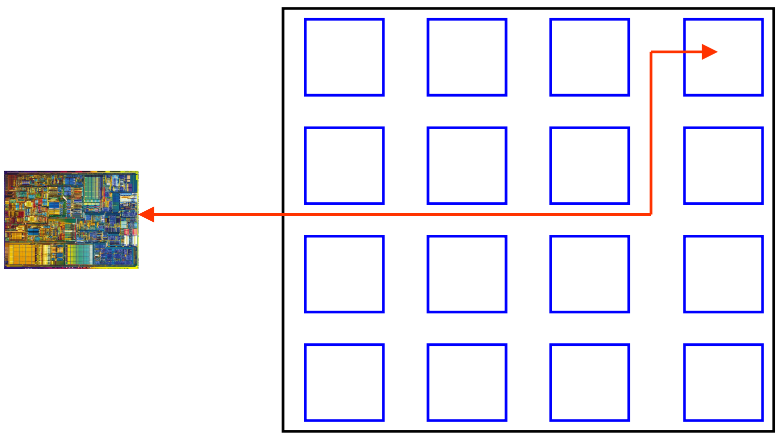
- A low-latency low-bandwidth network can be effectively used to hide wire latencies and improve performance

- A high-bandwidth low-energy network and an instruction assignment heuristic are effective at reducing contention cycles and total processor energy.

[Balasubramonian'05]

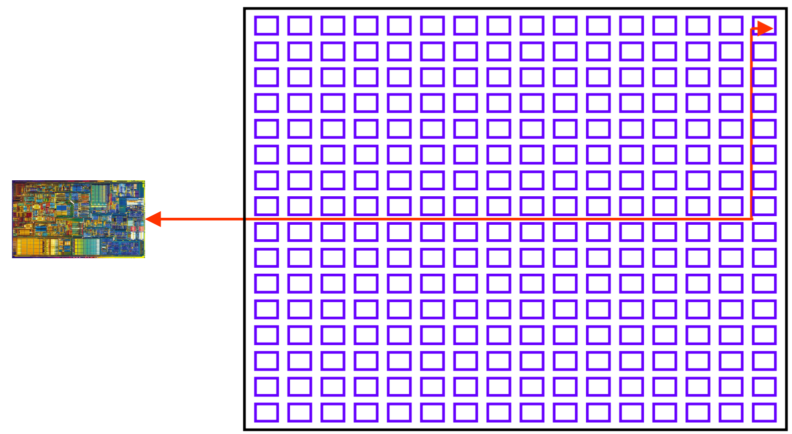
Non-Uniform Cache Architecture

- NUCA optimizes energy and time based on the proximity of the cache blocks to the cache controller.



2MB @ 130nm

Bank Access time = 3 cycles
Interconnect delay = 8 cycles



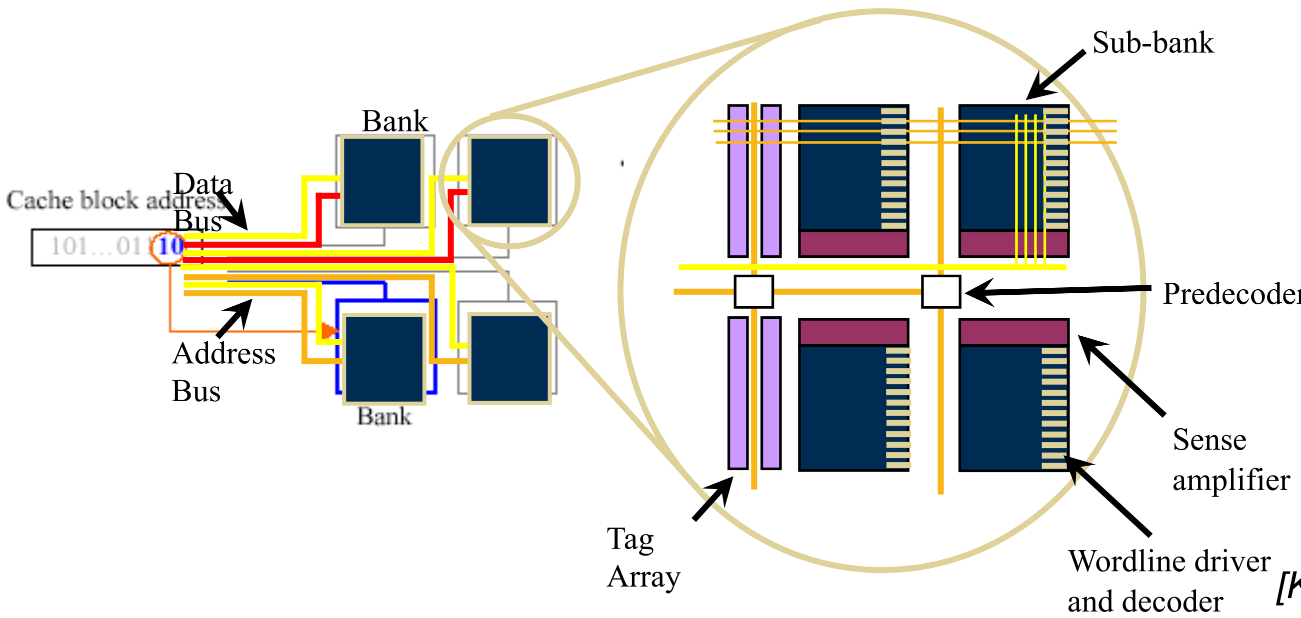
16MB @ 50nm

Bank Access time = 3 cycles
Interconnect delay = 44 cycles

[Kim'04]

Non-Uniform Cache Architecture

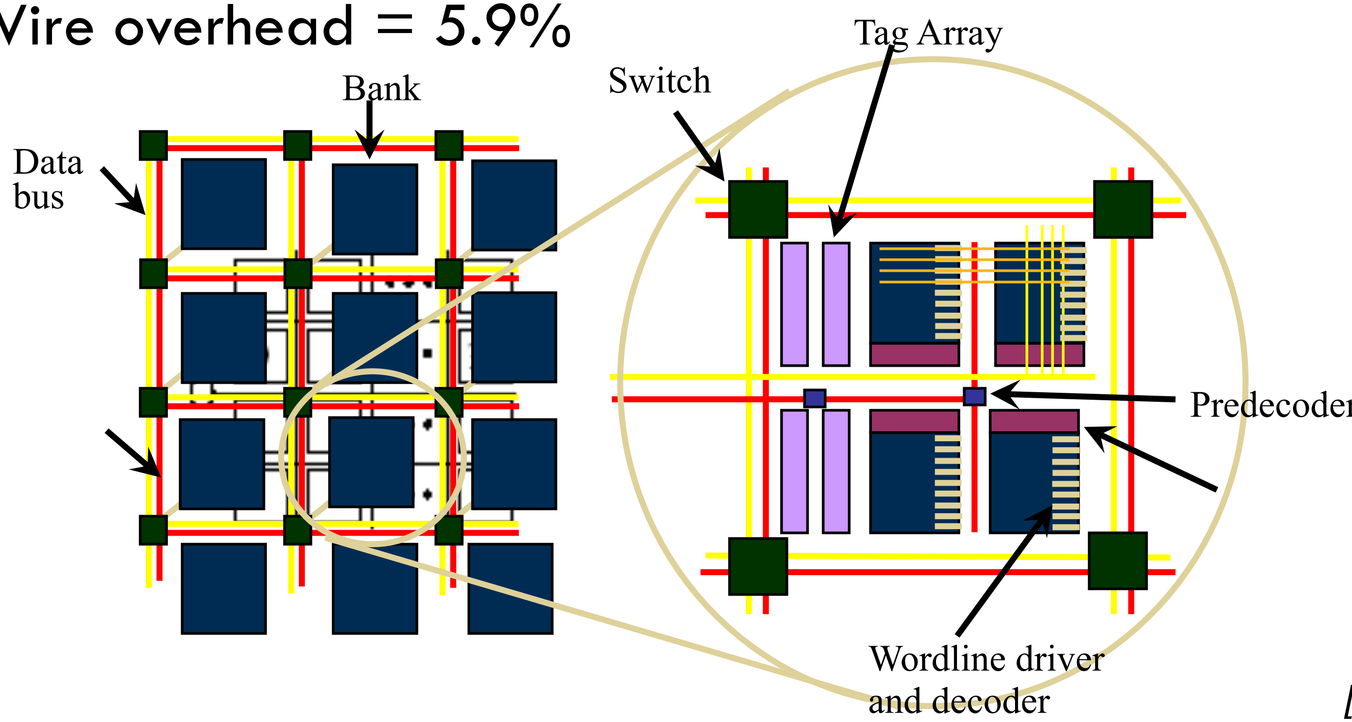
- S-NUCA-1
  - Use private per-bank channel
  - Each bank has its distinct access latency
  - Statically decide data location for its given address
  - Average access latency =34.2 cycles
  - Wire overhead = 20.9% → an issue



[Kim'04]

Non-Uniform Cache Architecture

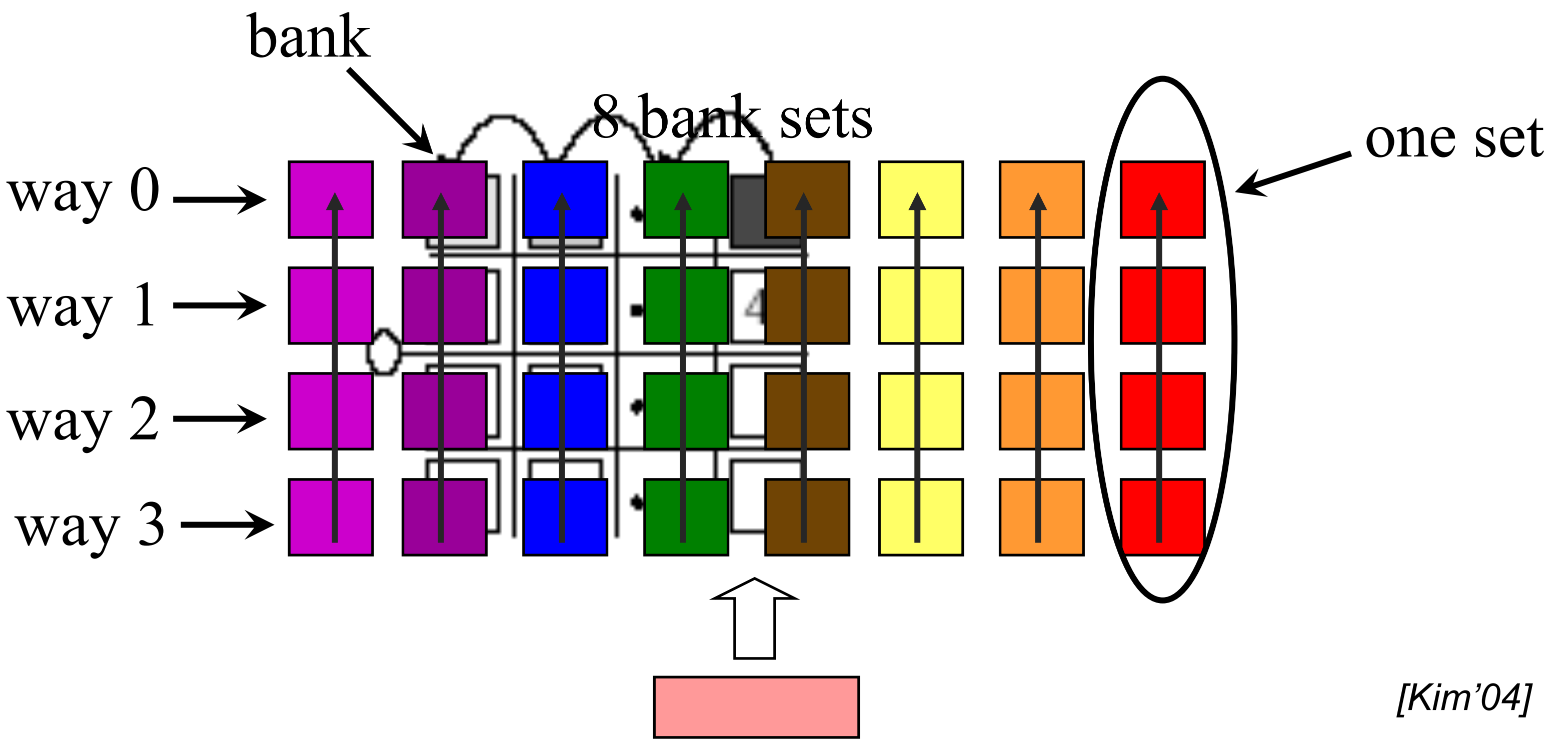
- S-NUCA-2
  - Use a 2D switched network to alleviate wire area overhead
  - Average access latency =24.2 cycles
  - Wire overhead = 5.9%



[Kim'04]

Non-Uniform Cache Architecture

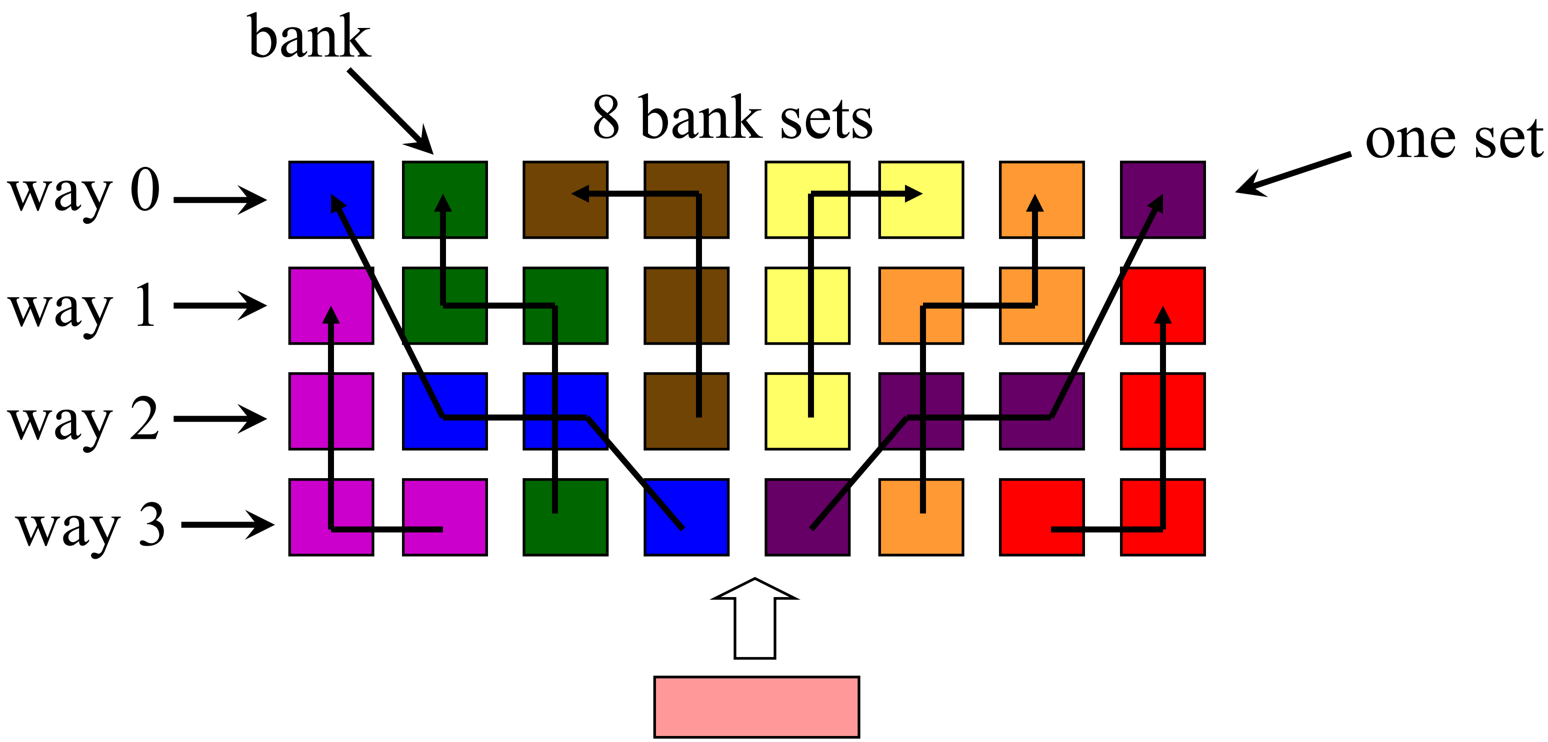
- Dynamic NUCA
  - Data can dynamically migrate
  - Move frequently used cache lines closer to CPU



[Kim'04]

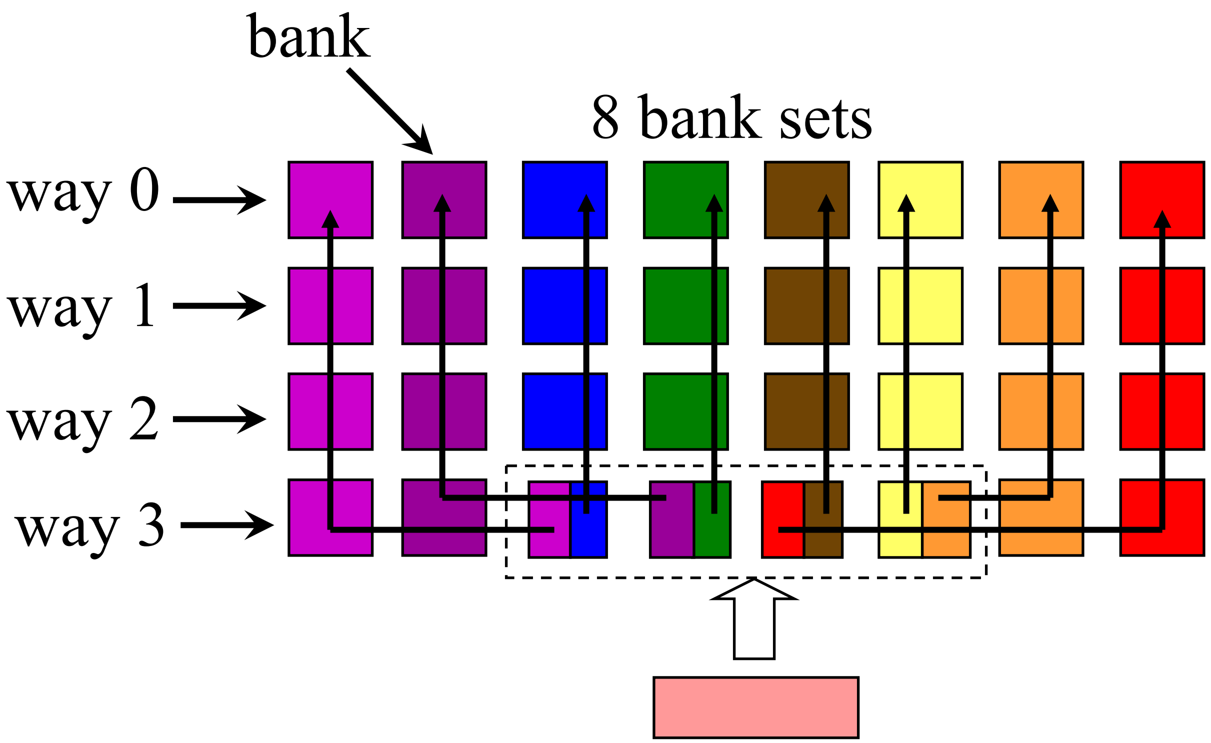
Non-Uniform Cache Architecture

- Fair mapping
  - Average access time across all bank sets are equal



Non-Uniform Cache Architecture

- Shared mapping
  - Sharing the closet banks for farther banks

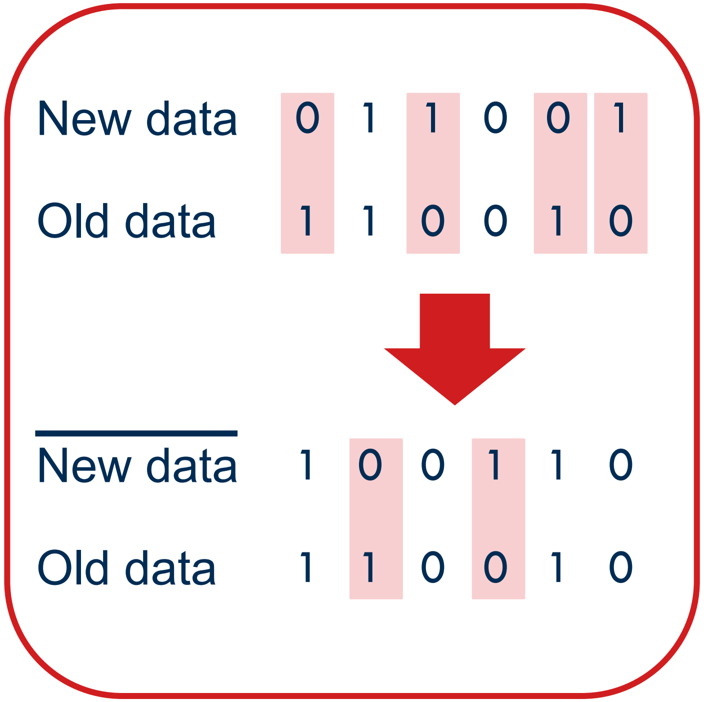


Encoding Based Optimizations

Cache Interconnect Optimizations

- Bus invert coding transfers either the data or its complement to minimize the number of bit flips on the bus.

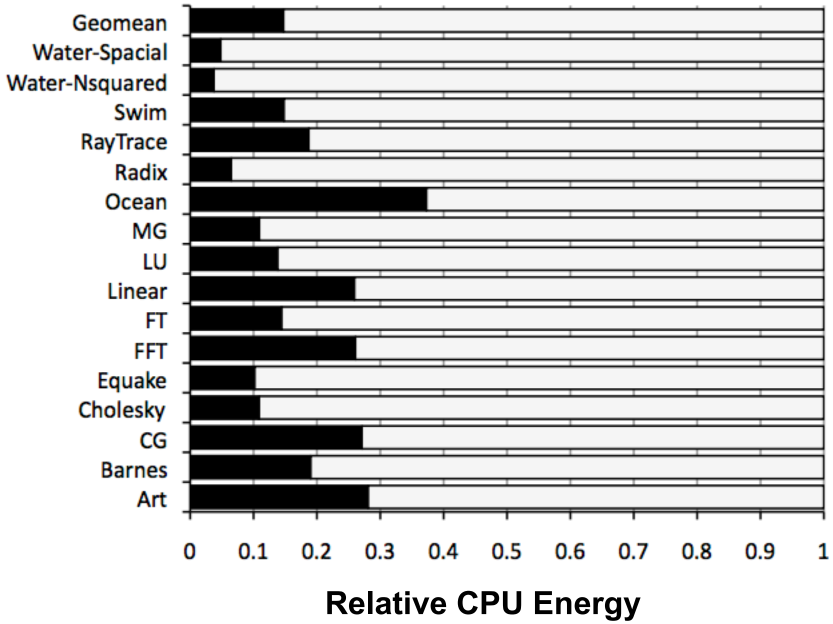
$$P_{\text{switching}} = \alpha C V_{DD}^{2} f$$



[Stan'95]

Time-Based Data Transfer

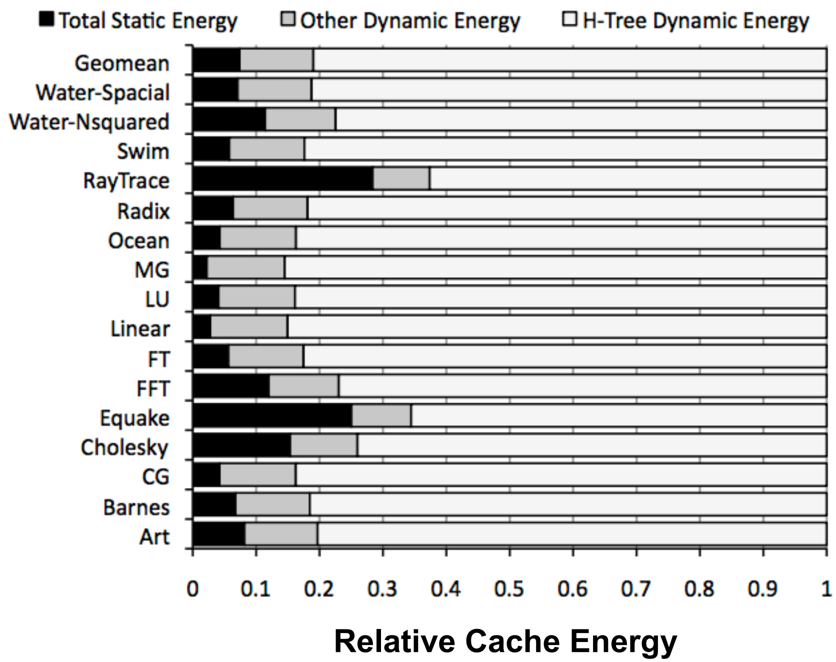
- The percentage of processor energy expended on an 8MB cache when running a set of parallel applications on a Sun Niagara-like multicore processor



[Bojnordi'13]

Time-Based Data Transfer

- Communication over the long, capacitive H-tree interconnect is the dominant source of energy consumption (80% on average) in the L2 cache

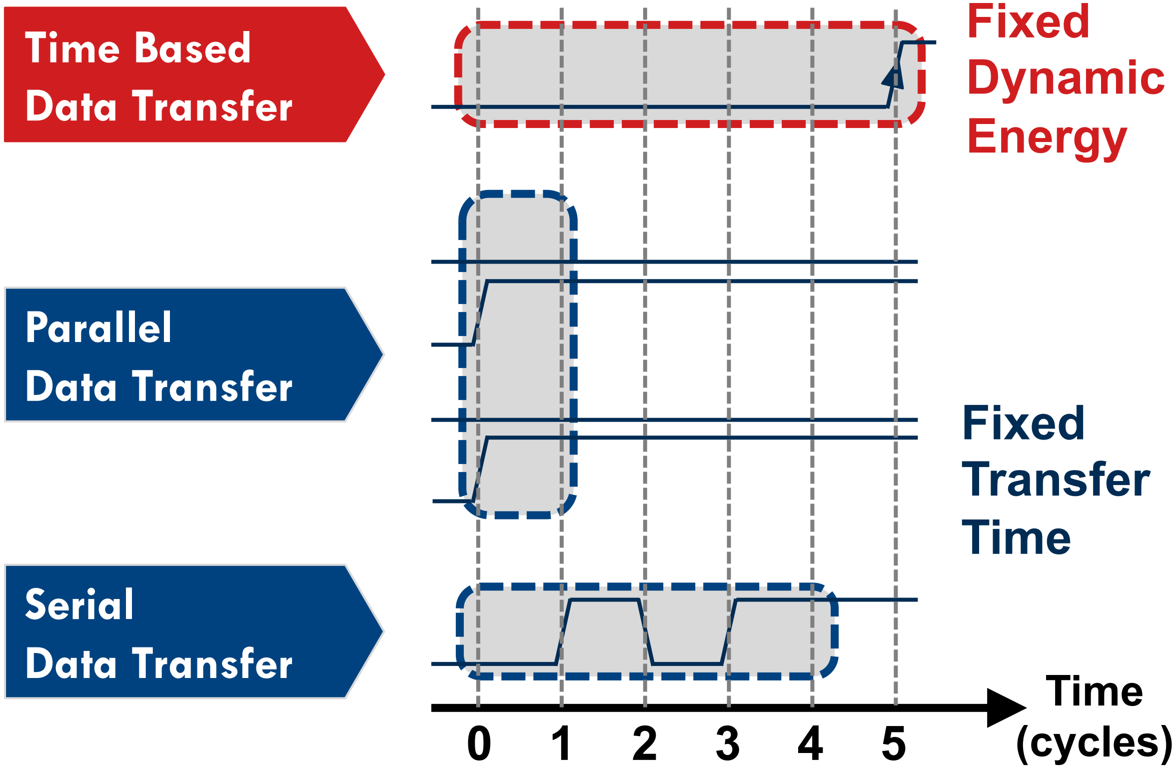


[Bojnordi'13]

Time-Based Data Transfer

Key idea: represent information by the number of clock cycles between two consecutive pulses to reduce interconnect activity factor.

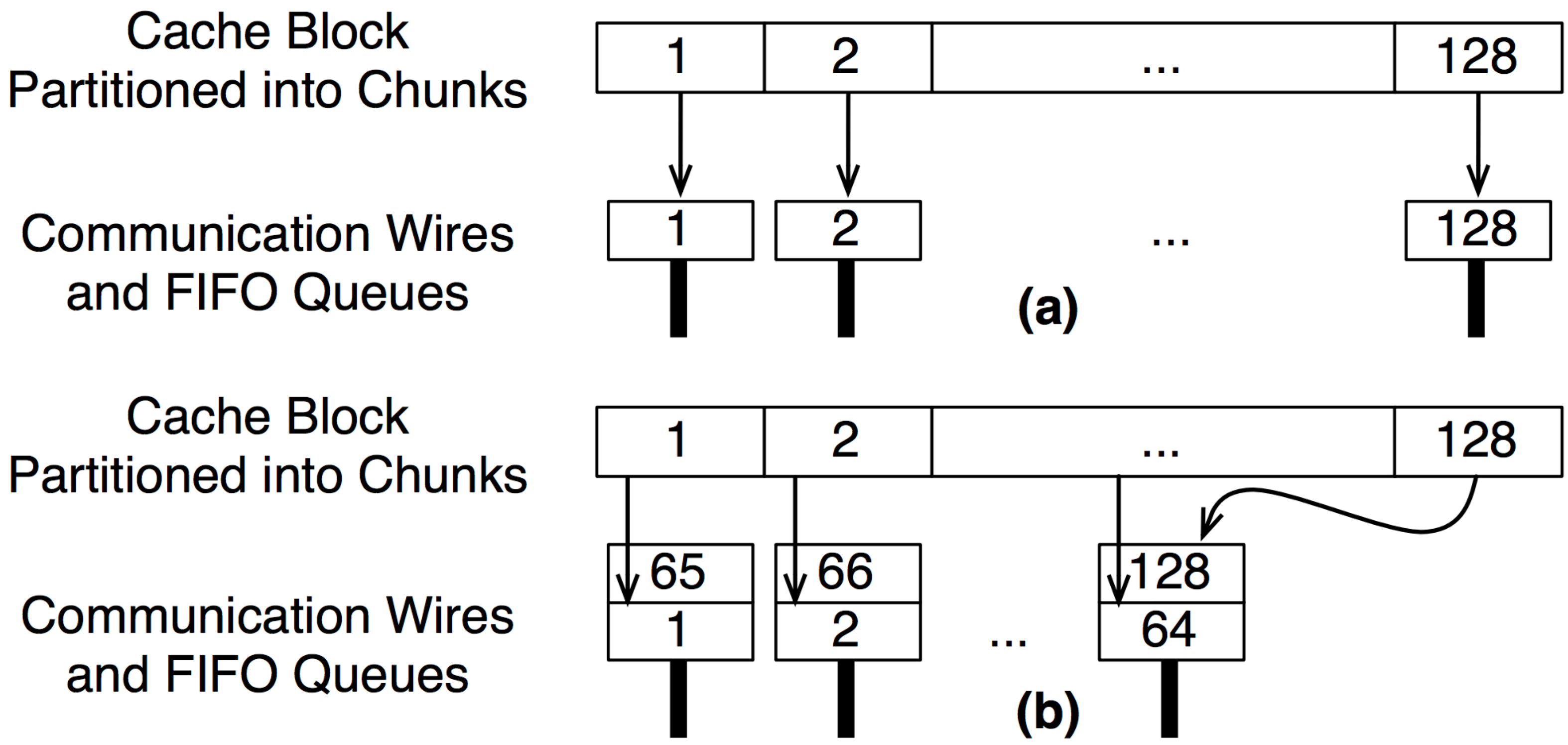
Example: transmitting the value 5



[Bojnordi'13]

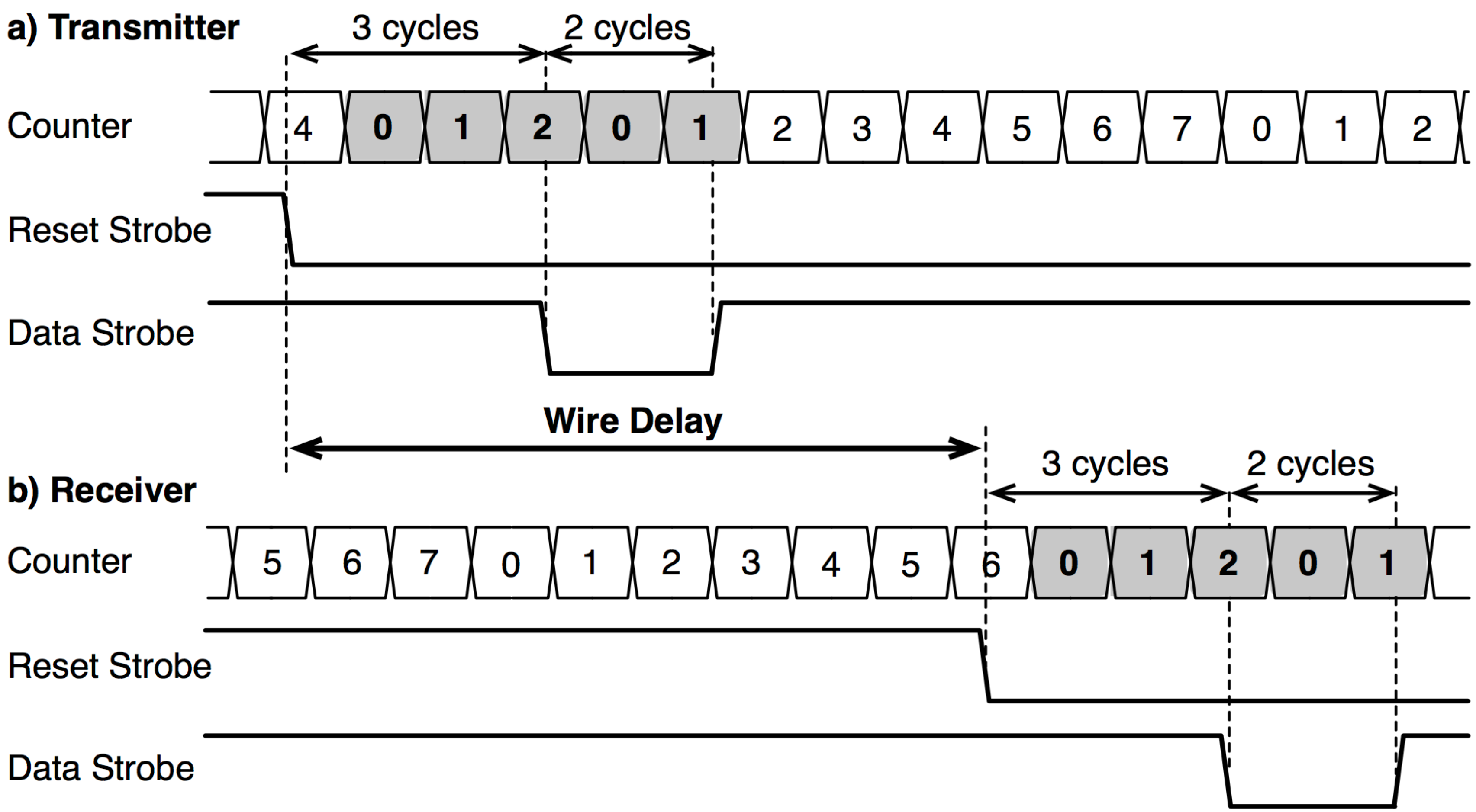
Time-Based Data Transfer

- Cache blocks are partitioned into small, contiguous chunks.



[Bojnordi'13]

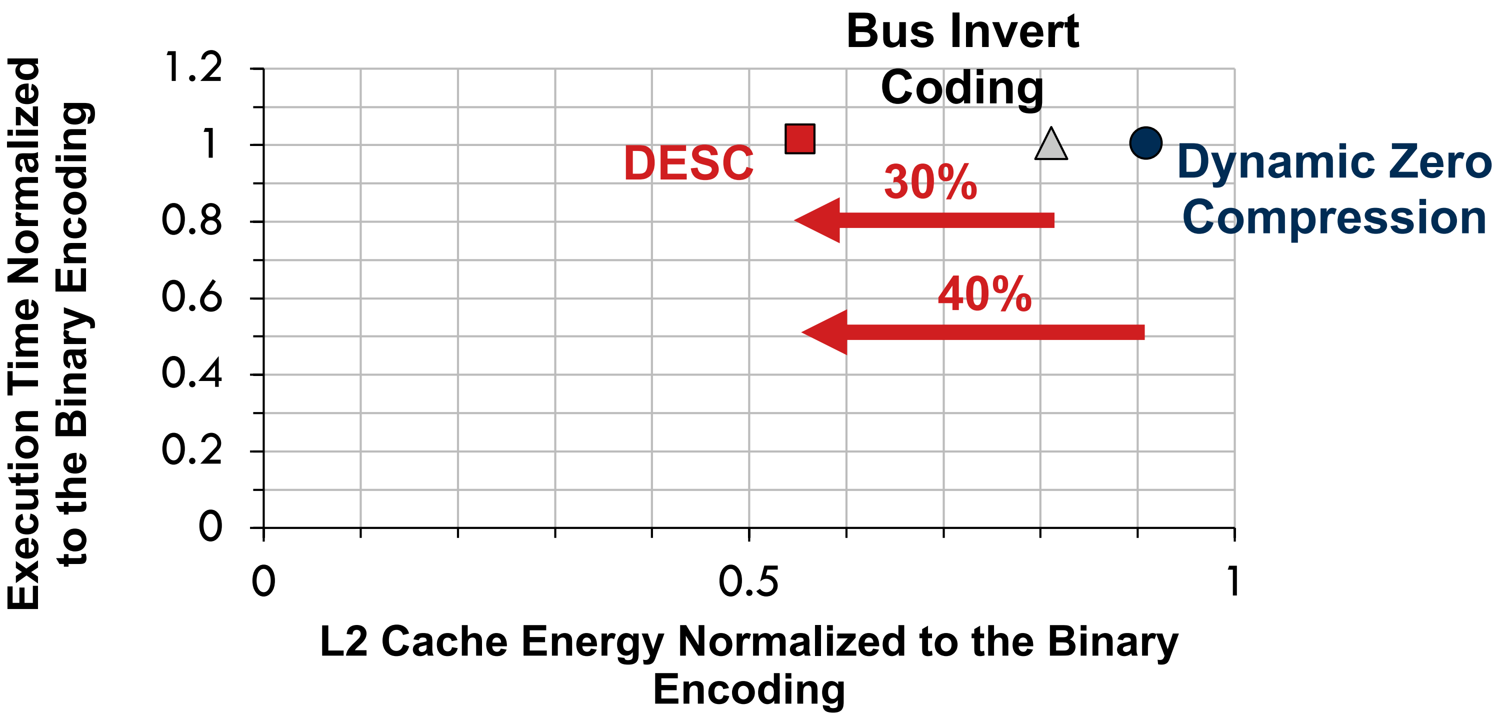
Time-Based Data Transfer



[Bojnordi'13]

Time-Based Data Transfer

- L2 cache energy is reduced by 1.8x at the cost of less than 2% increase in the execution time.

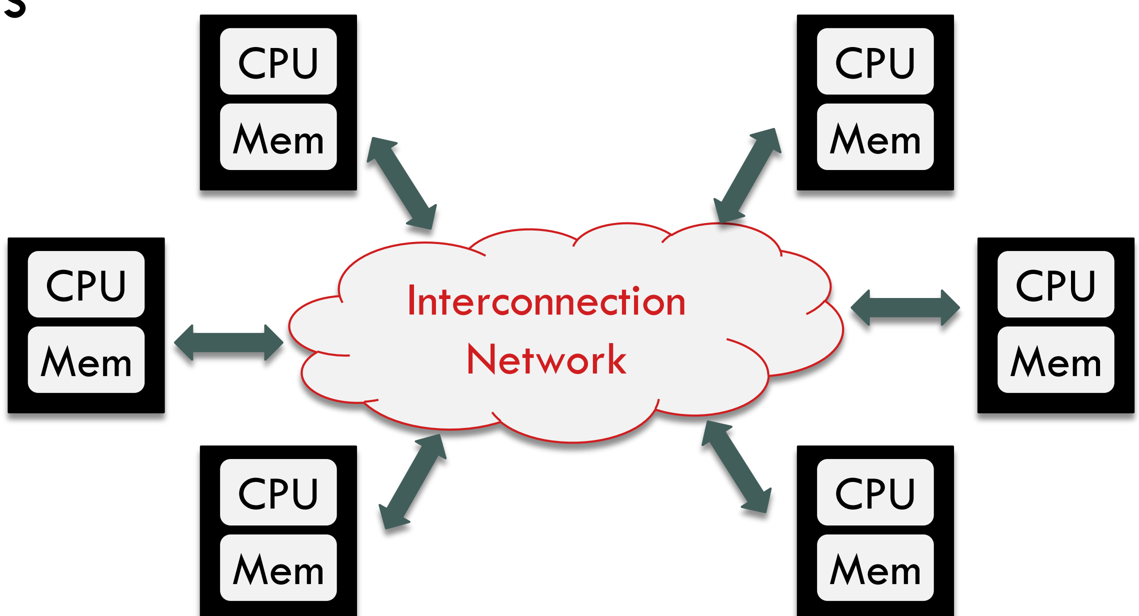


[Bojnordi'13]

Interconnection Networks

Interconnection Networks

- Goal: transfer maximum amount of information with the minimum time and power
- Connects processors, memories, caches, and I/O devices



Types of Interconnection Networks

- Four domains based on number and proximity of devices
  - On-chip networks (OCN or NOC)
    - Microarchitectural elements: cores, caches, reg. files, etc.
  - System/storage area networks (SAN)
    - Computer subsystems: storage, processor, IO device, etc.
  - Local area networks (LAN)
    - Autonomous computer systems: desktop computers etc.
  - Wide area networks (WAN)
    - Interconnected computers distributed across the globe

Basics of Interconnection Networks

- Network topology
  - How to wire switches and nodes in the network
- Routing algorithm
  - How to transfer a message from source to destination
- Flow control
  - How to control the flow messages within the network

Network Topology

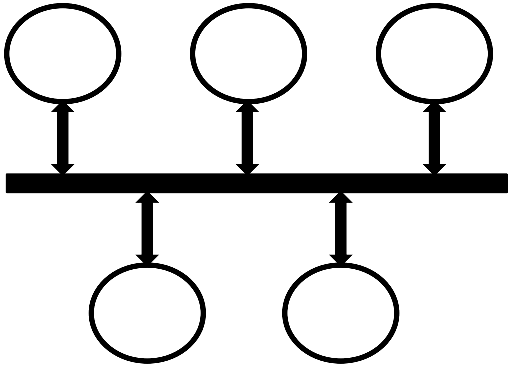
Network Topologies

- Regular vs. irregular graphs
  - Examples of regular networks are mesh and ring
- Distances in the network
  - Routing distance: number of links/hops along a route
  - Network diameter: maximum number of hops per route
  - Average distance: average number of links/hops across all valid routes

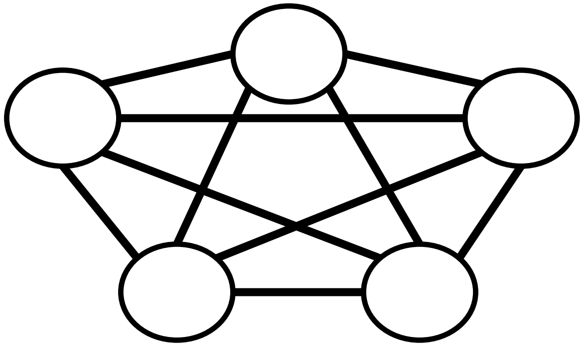
Example Topologies

- Bus
  - Simple structure; efficient for small number of nodes
  - Not scalable; highly contended
  - Used in many processors

Bus

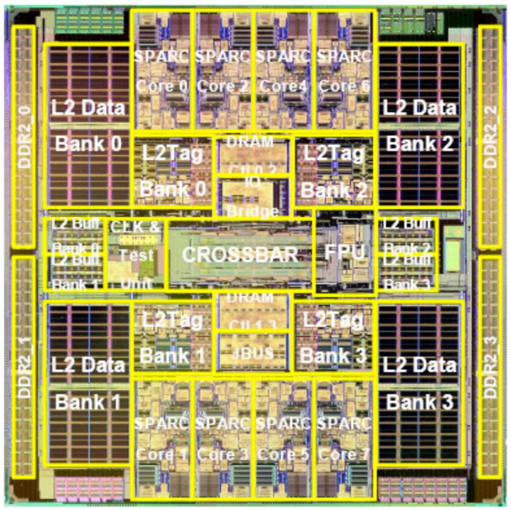


Point to Point

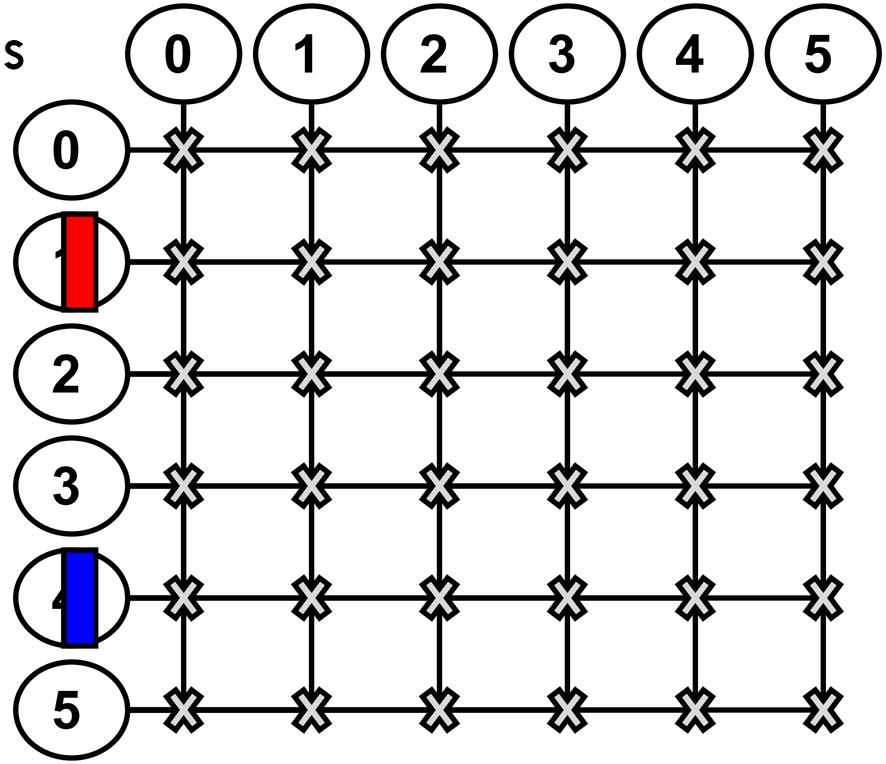


Example Topologies

- Crossbar
  - Complex arbitration
  - High throughput and fast
  - Requires a lot of resources
  - Used in Sun Niagara I/II

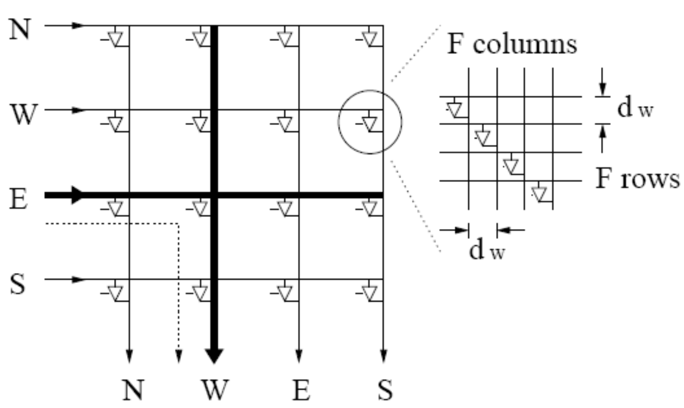


[UltraSPARC T1]

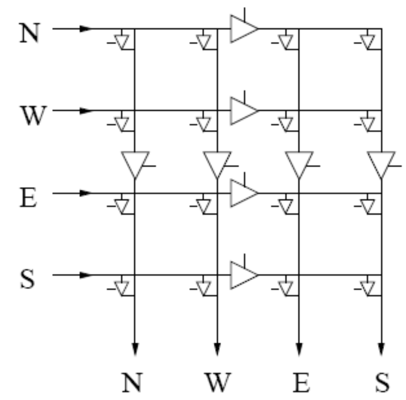


Example Topologies

- Segmented crossbar
  - Reduce switching capacitance (~15-30%)
  - Need a few additional signals to control tri-states



(a) A 4×4 matrix crossbar.



(b) A 4×4 segmented cross-bar with 2 segments per line.

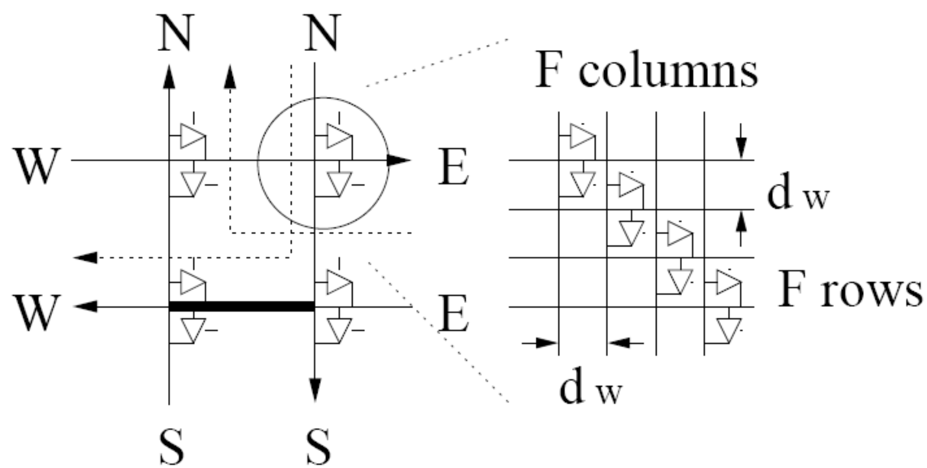
[Wang'03]

Example Topologies

- Goal: optimize for the common case
  - Straight-through traffic does not go thru tristate buffers

- Some combinations of turns are not allowed
  - Why?

Read the paper for details.

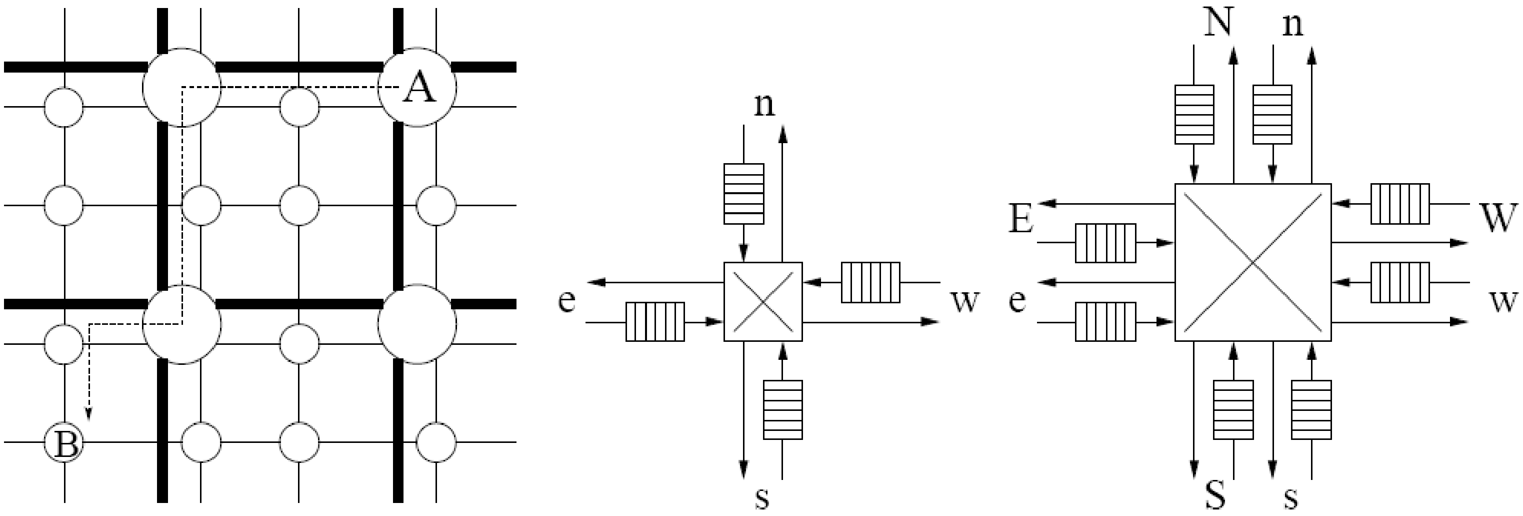


(a) A 4×4 cut-through crossbar.

[Wang'03]

Example Topologies

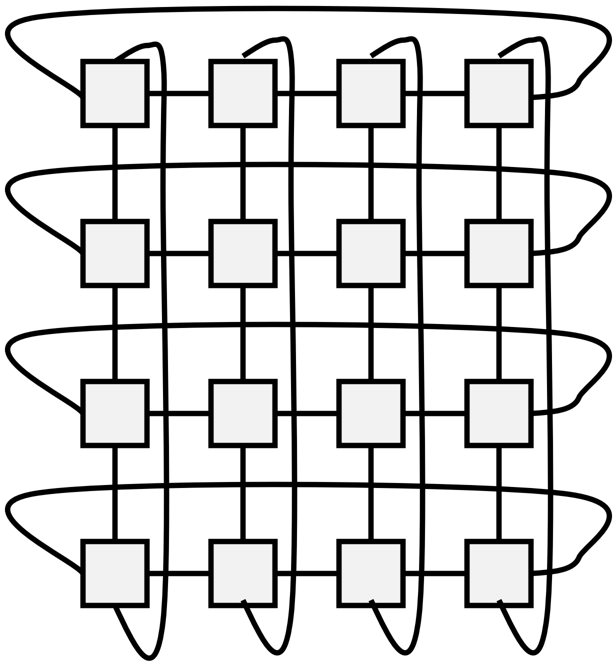
- Express channels to reduce number of hops
  - like taking the freeway



[Wang'03]

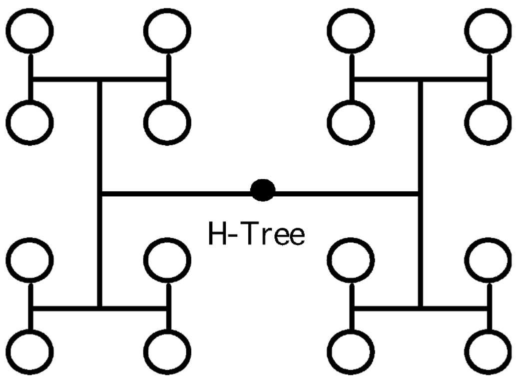
Example Topologies

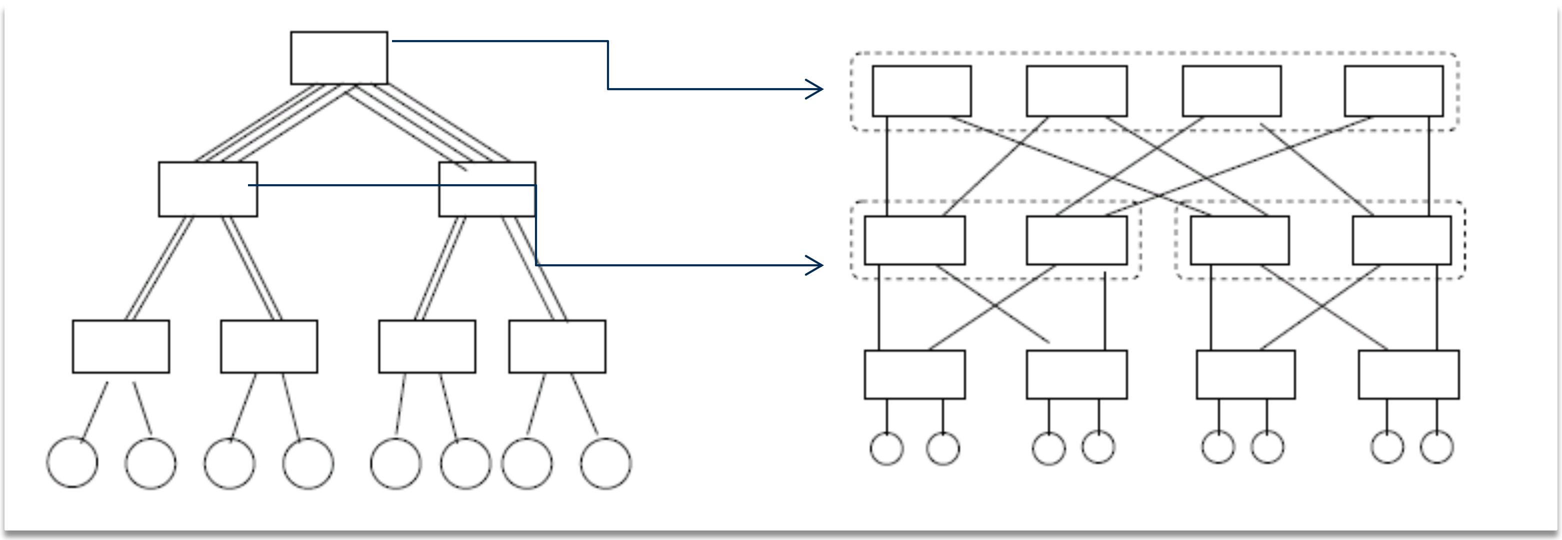
- Ring
  - Cheap; long latency
  - IBM Cell
- Mesh
  - Path diversity, efficient
  - Tilera 100-core
- Torus
  - More path diversity
  - Expensive and complex



Example Topologies

- Tree
  - Simple and low cost
  - Easy to layout
  - Efficiently handles local traffic
  - Towards root, links are heavily contended





Example Topologies

- Omega network
  - Single path from source to destination
  - Does not support all possible permutations
  - Proposed to replace costly crossbars as processor-memory interconnect

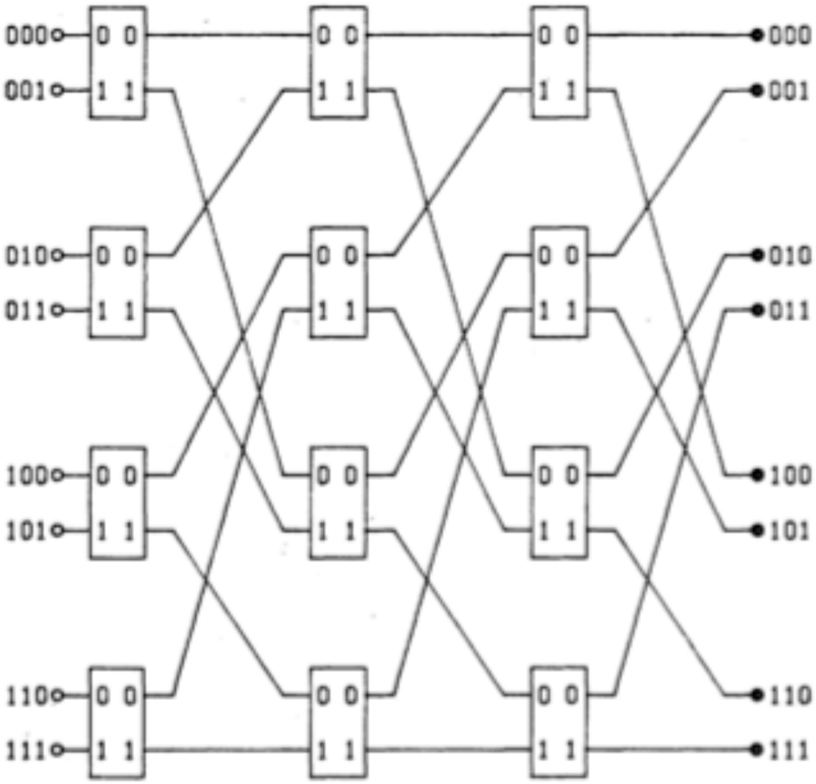


Fig. 2. Omega-network ($N = 8$).

[Gottlieb'82]

Flow Control

Sending Data in Network

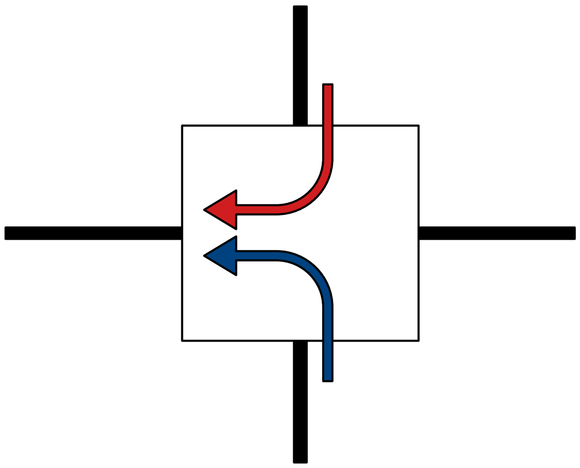
- Circuit switching
  - Establish full path; then send data
  - Everyone else using the same link has to wait
  - Setup overheads

- Packet switching
  - Route individual packets (via different paths)
  - More flexible than CS
  - May be slower than CS

Handling Contention

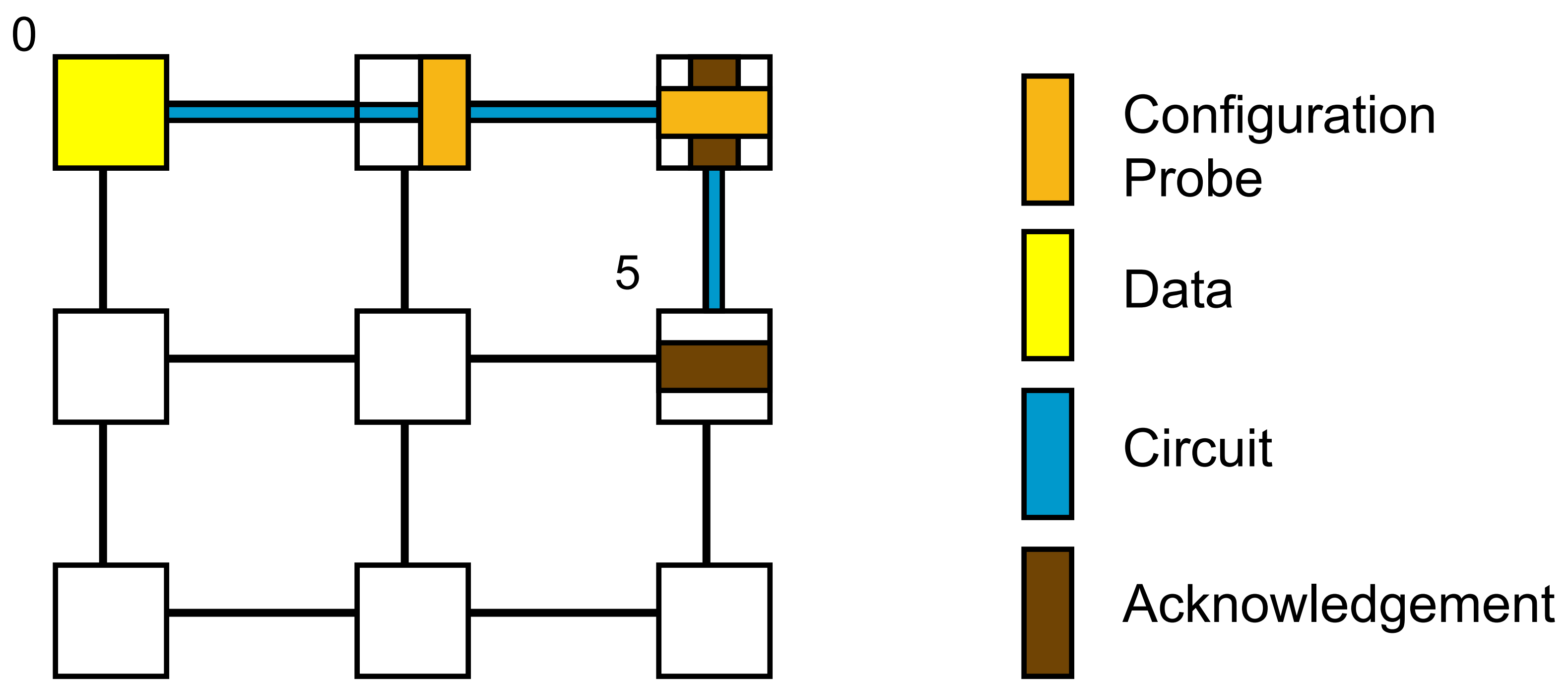
- Problem
  - Two packets want to use the same link at the same time

- Possible solutions
  - Drop one
  - Misroute one (deflection)
  - Buffer one



Circuit Switching Example

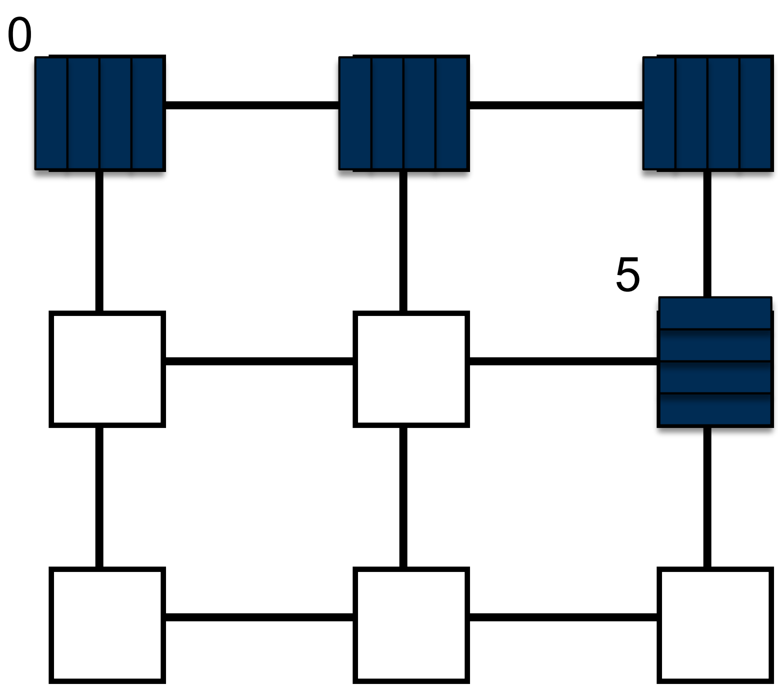
- Significant latency overhead prior to data transfer
- Other requests forced to wait for resources



[Lipasti]

Store and Forward Example

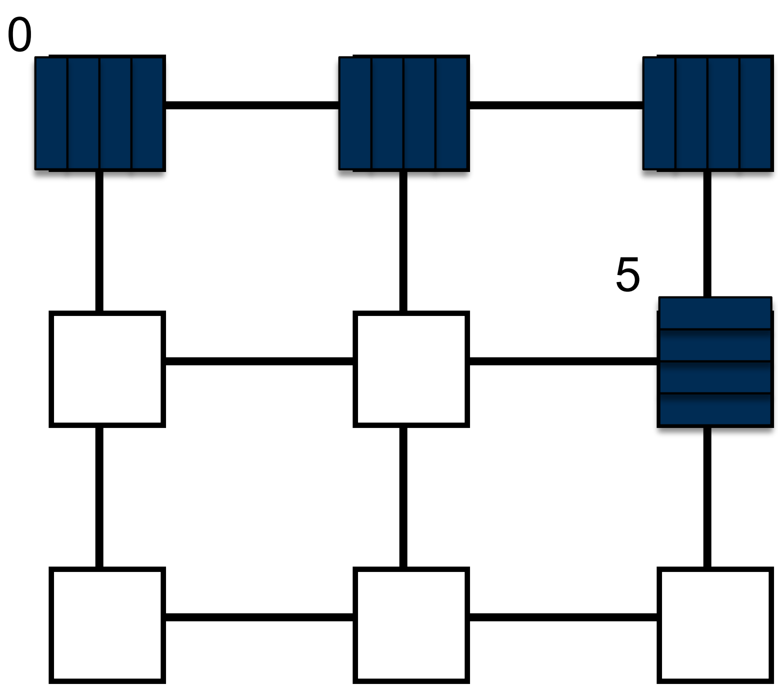
- High per-hop latency
- Larger buffering required



[Lipasti]

Virtual Cut Through Example

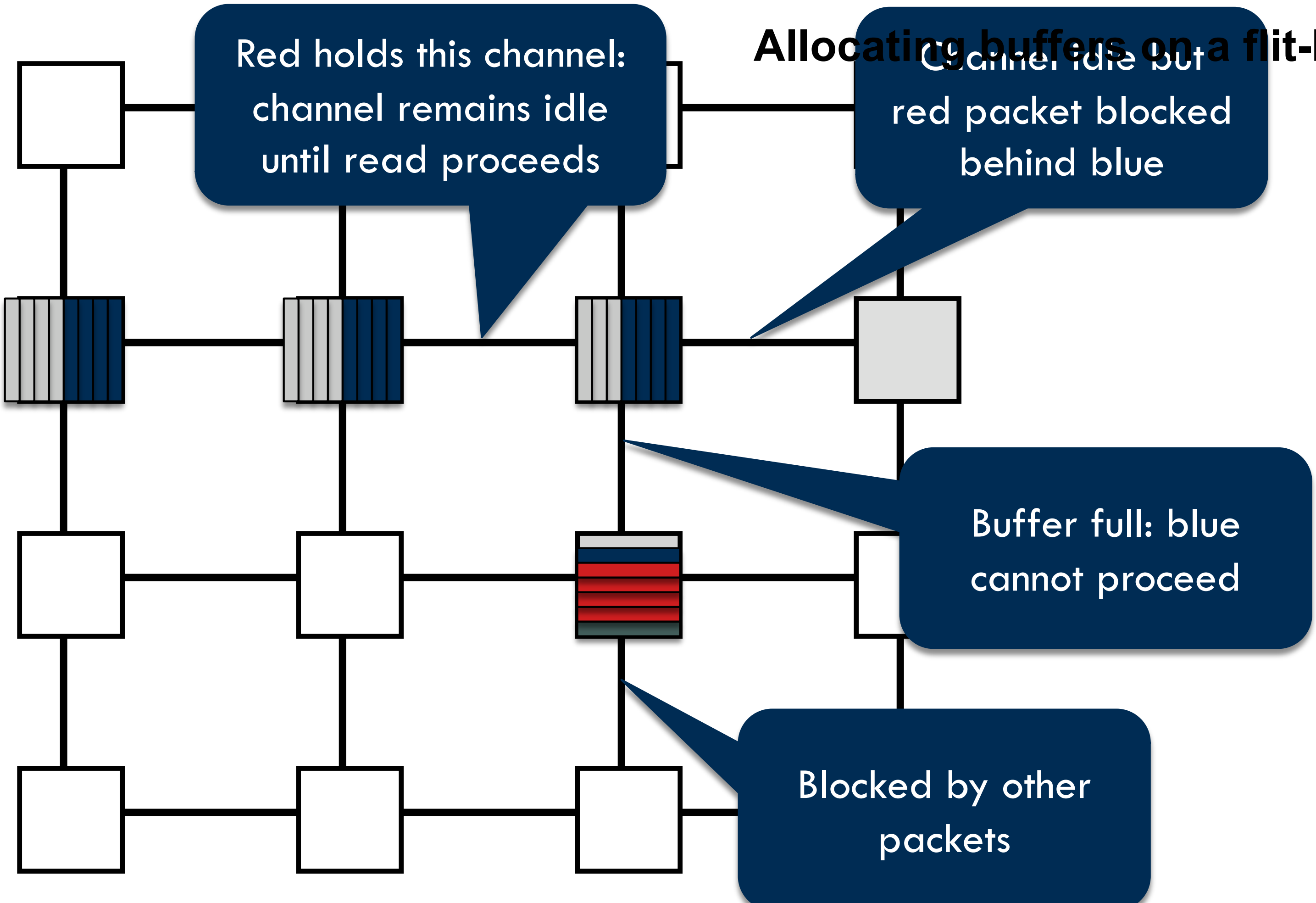
- Lower per-hop latency
- Larger buffering required



[Lipasti]

Wormhole Example

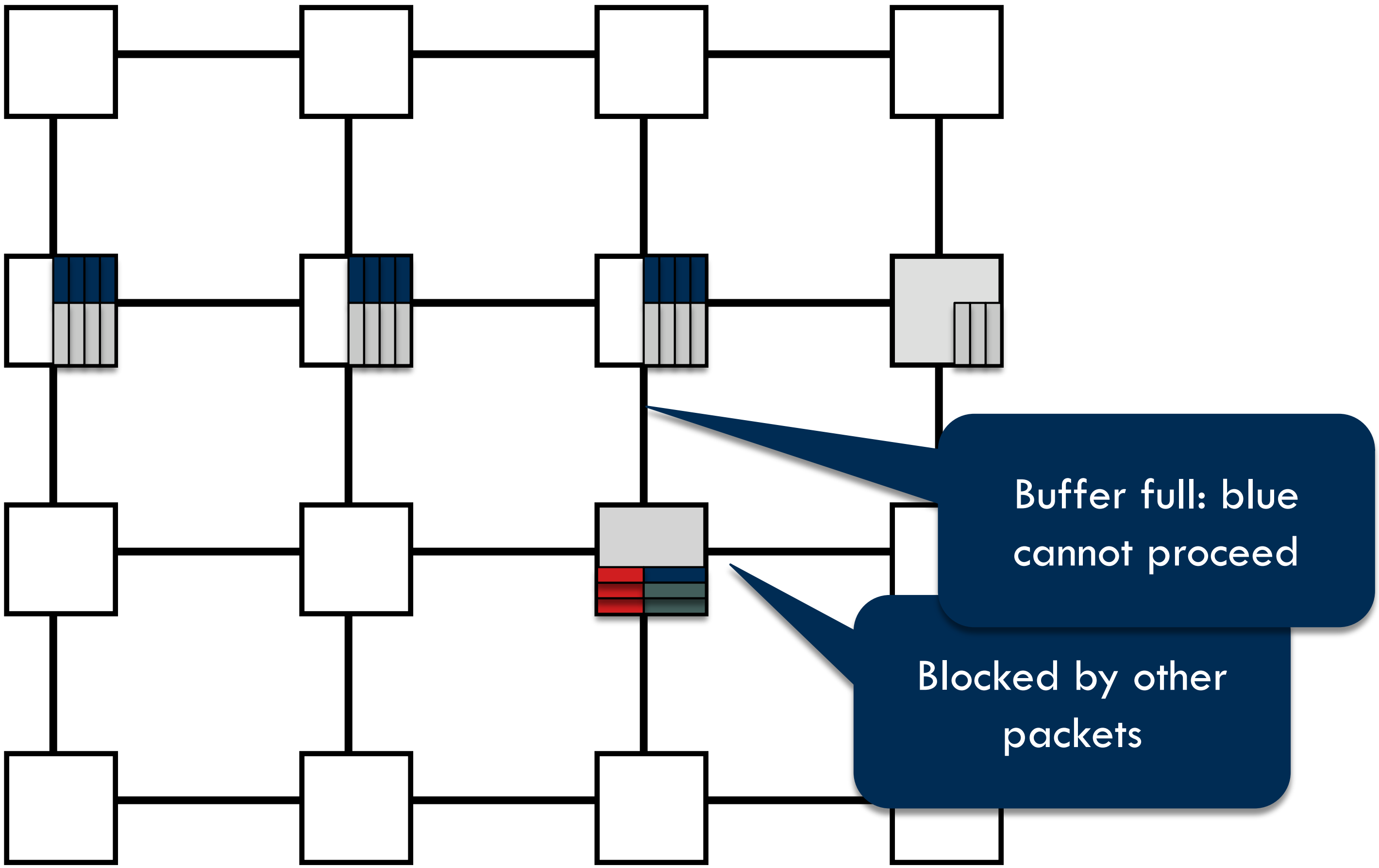
Allocating buffers on a flit-basis



[Lipasti]

Virtual Channel Example

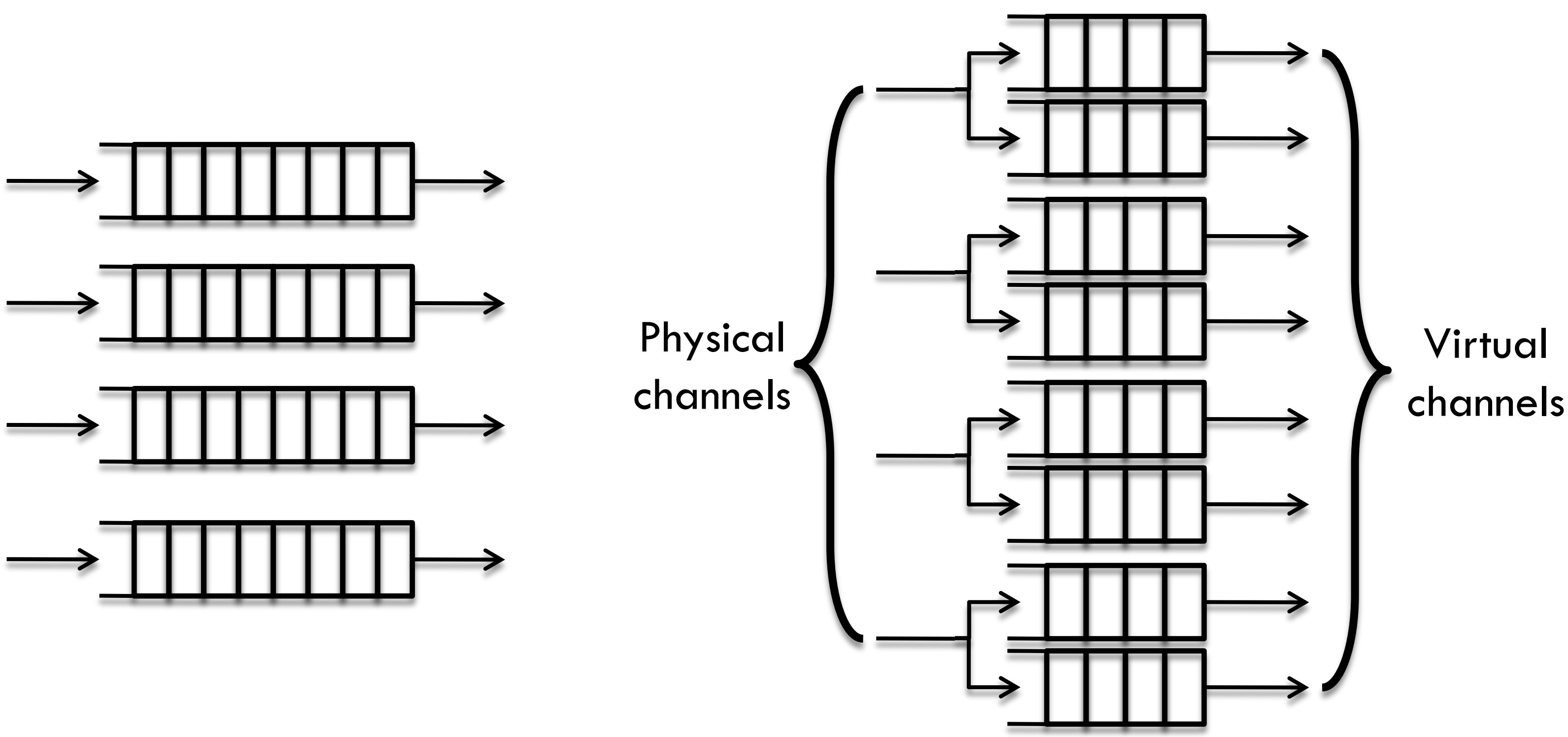
Multiple flit queues per input port



[Lipasti]

Virtual Channel Buffers

- Single buffer per input
- Multiple fixed length queues per physical channel



[Lipasti]