ACCA
ADVANCED COMMUNICATION & COMPUTER ARCHITECTURE

MINISTERIO DE EDUCACIÓN Y CIENCIA

NaNoC design platform

Heterogeneous NoC Design for Efficient Broadcast-based Coherence Protocol Support

Mario Lodde[1], Jose Flich[1], Manuel E. Acacio[2]

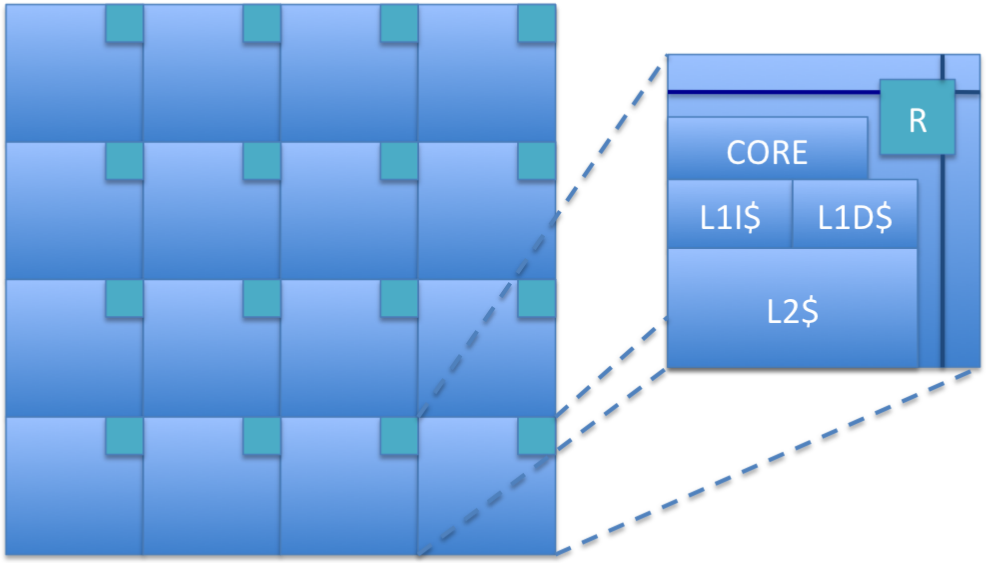
[1]Universitat Politècnica de València [2]Universidad de Murcia

Overview

- Introduction
- Broadcast-based coherence protocols
- Gather Control Network
- Implementation details
- Evaluation
- Conclusion

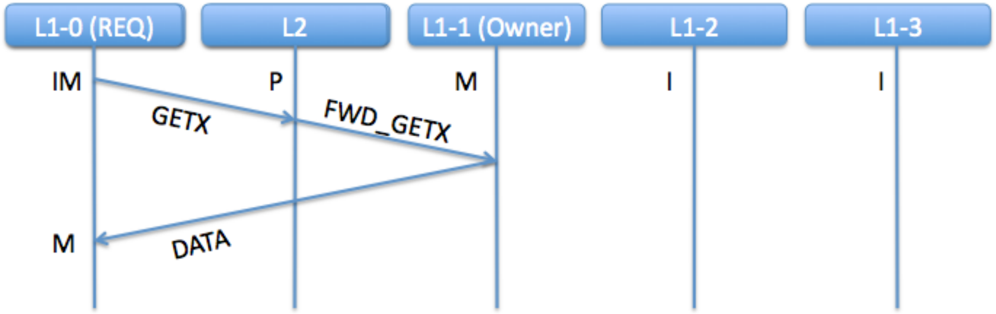
Introduction

- A tiled CMP system is made of replicated tiles, each one including a core, his cache hierarchy and a router
- Directory-based cache coherence protocols are typically employed

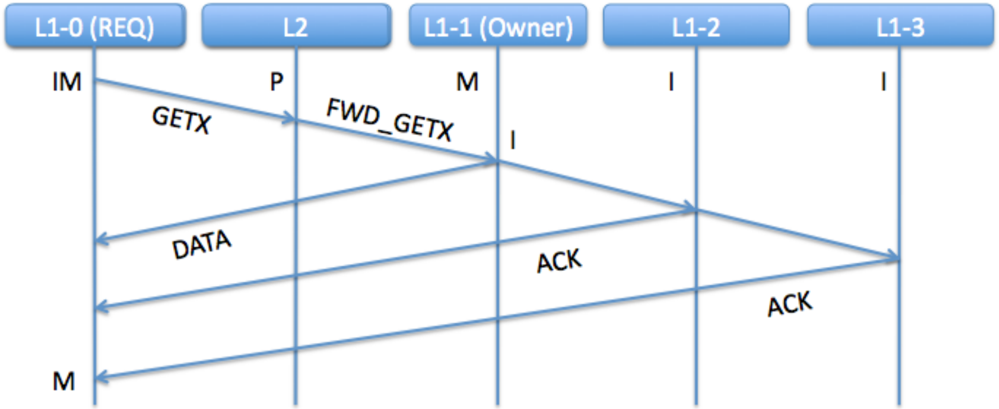


Broadcast-based protocols

- Bit vector: we know exactly which cores have a copy of the block → we only communicate with those cores

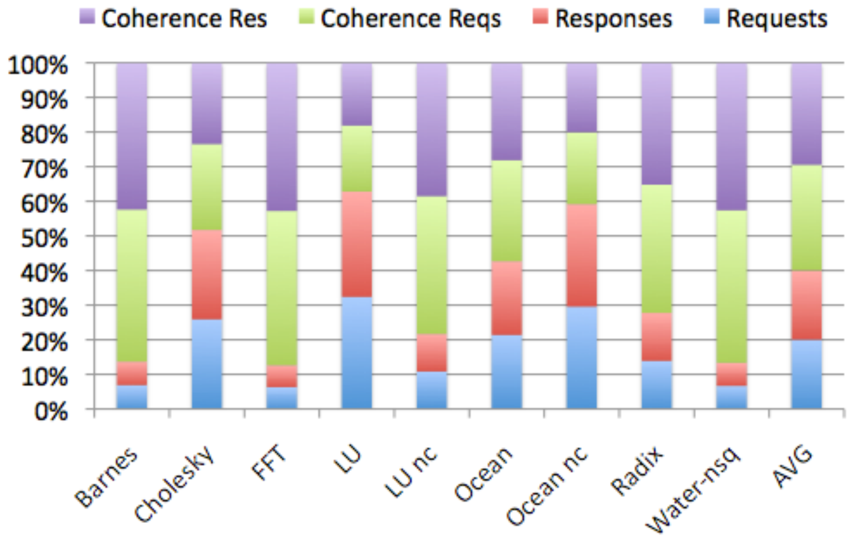


- Dir$_0$B: we don't know which cores have a copy of the block → we broadcast the request to all cores



Hammer protocol traffic breakdown

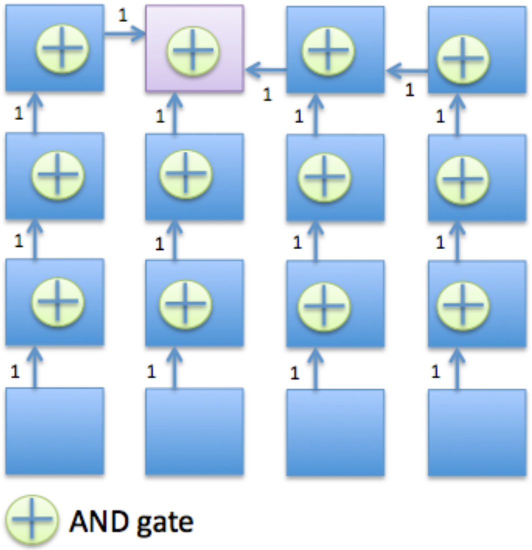
- On average, 60% of total traffic is due to coherence requests and their acknowledgements

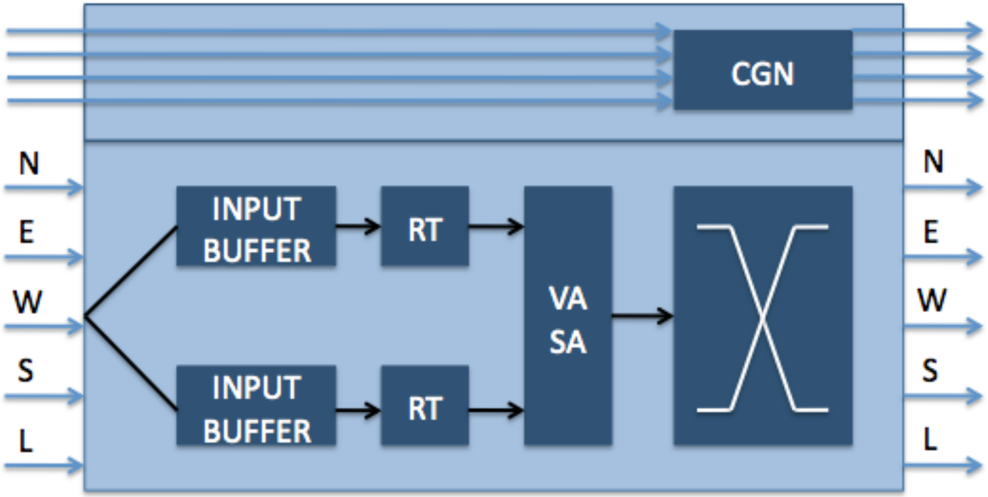


- One well-known strategy to reduce traffic generated by coherence requests is to use a NoC with multicast or broadcast support
- What shall we do with acknowledgement messages?

The Gather Control Network

- We add to the main NoC a fast and simple Gather Control Network (GCN) dedicated to collecting ACKs

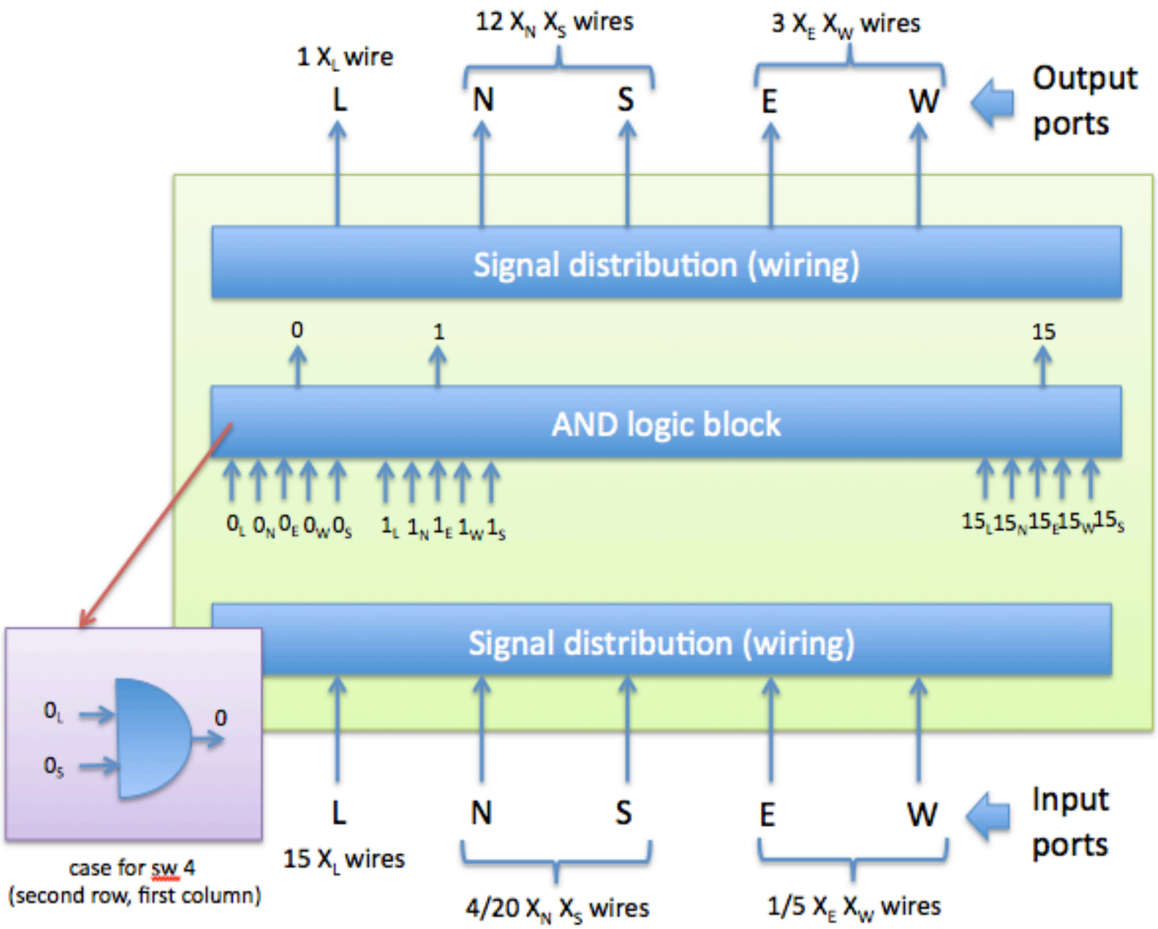




- The GCN consists of a one-bit wide subnetwork per node, each collecting the ACKs from all other nodes

GCN logic block

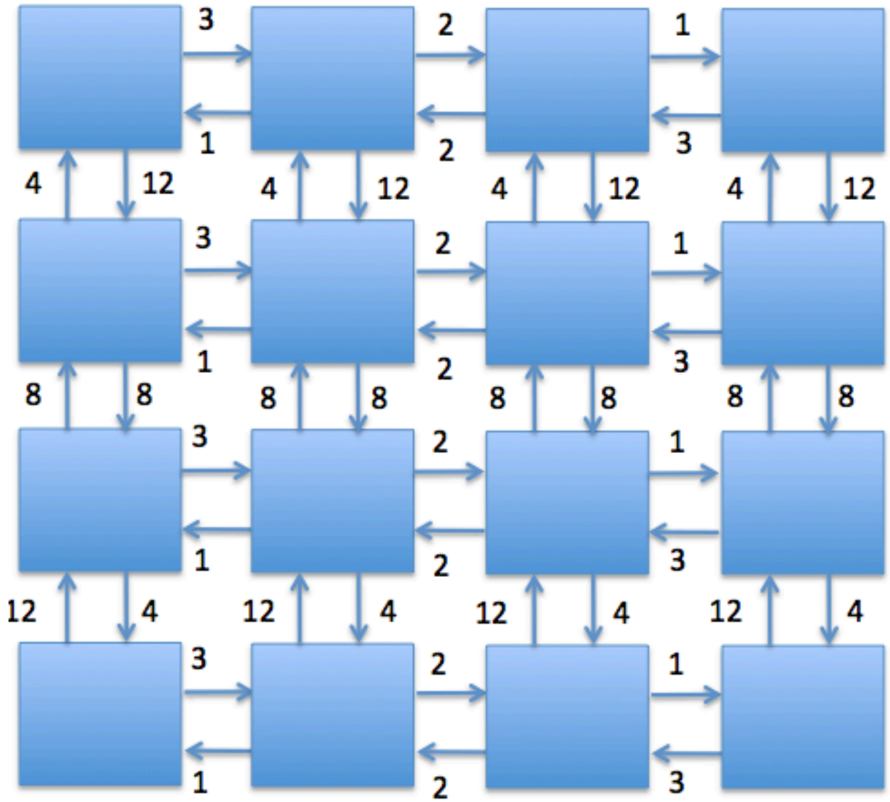
- Each GCN logic at each switch is connected to its neighbors control logic blocks with dedicated wires.
- The goal of the logic block is simply to AND the corresponding input signals and to distribute the results through the corresponding output ports, depending on the location of the switch in the mesh topology and the selected layout.



case for sw 4
(second row, first column)

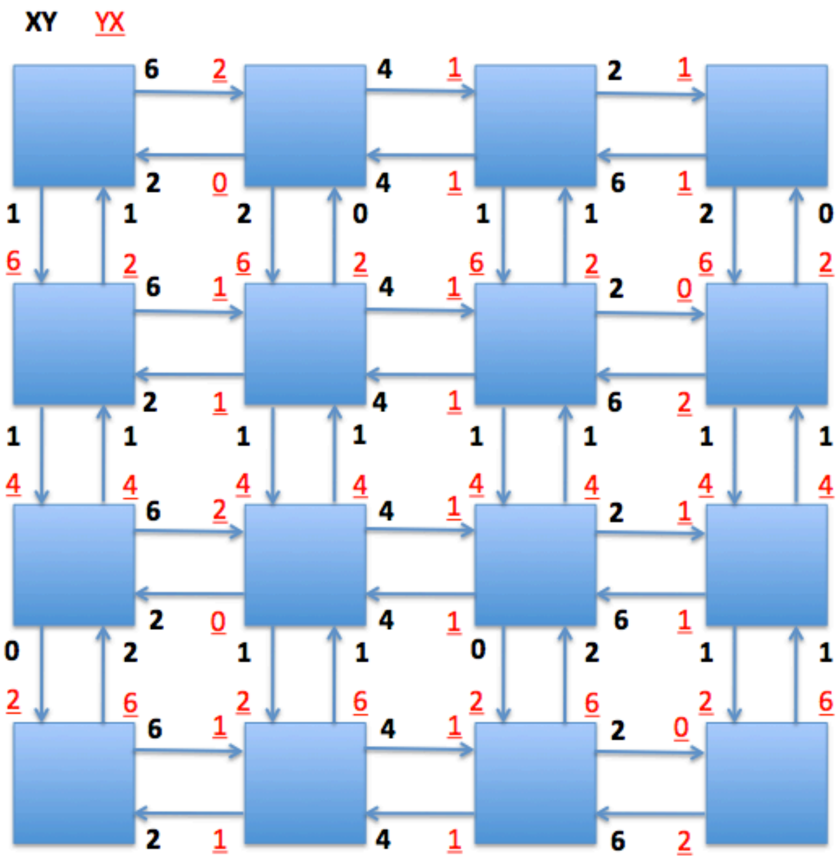
Control signals (1/2)

- YX layout
  - N signals in north and south ports
  - N x N signals in east and west ports



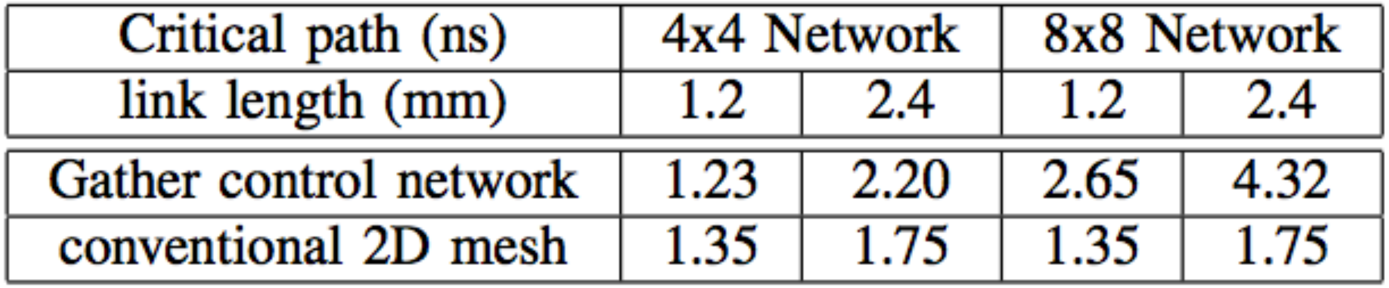
Control signals (2/2)

- Balanced layout
  - (N² + N)/2 signals per port



Implementation analysis: switch + GCN

- A 4-stage switch and the GCN control logic have been implemented using the 45 nm technology Nangate library with Synopsys DC
- The area of the GCN control logic is 2.72% the area of the switch.
- Depending on the size of the network and on link lenght, the SAGN can work at the same frequency of the NoC or may need more than one cycle (3 cycles in the worst case)



Critical path (ns)	4x4 Network		8x8 Network	
link length (mm)	1.2	2.4	1.2	2.4
Gather control network	1.23	2.20	2.65	4.32
conventional 2D mesh	1.35	1.75	1.35	1.75

Simulation tools

- We implemented the NoC and the cache hierarchy with gMemNoCsim, an in-house memory/network simulator.
- gMemNoCsim has been integrated in Graphite simulator to run applications of the SPLASH-2 suite

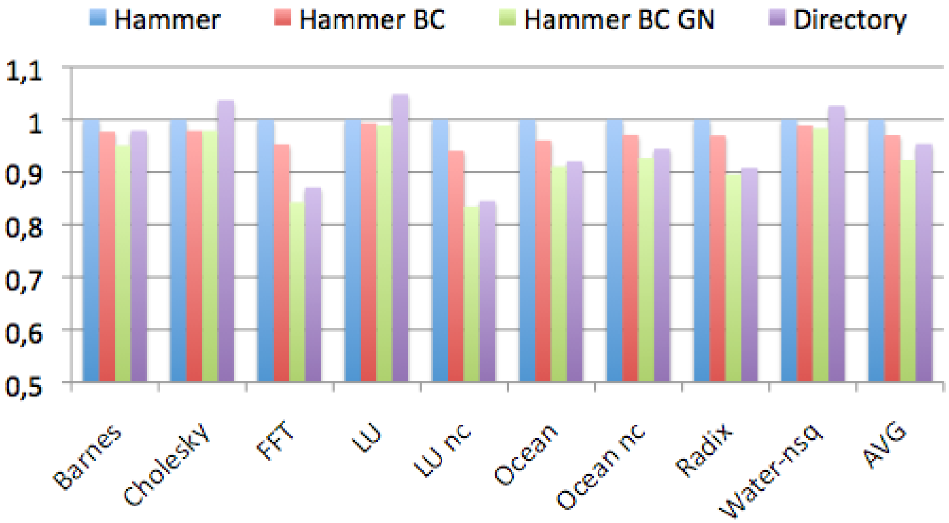
Network parameters
Wormhole switching
Stop&Go flow control
XY routing
GCN delay: 2 cycles
Flit size: 10 bytes

Cache parameters
64kB + 64kB L1 (I+D)
512kB L2
L1 tag /cache latency: 1 / 2 cycles
L2 tag/cache latency: 2 / 4 cycles

Performance Evaluation (1/3)

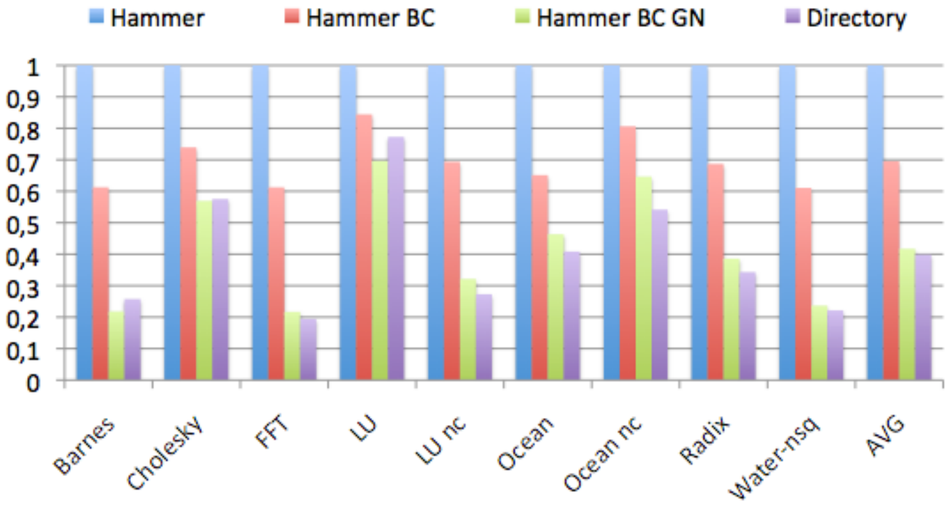
- Normalyzed execution time

  reduced by 8% on average when using NoC broadcast support and the GCN



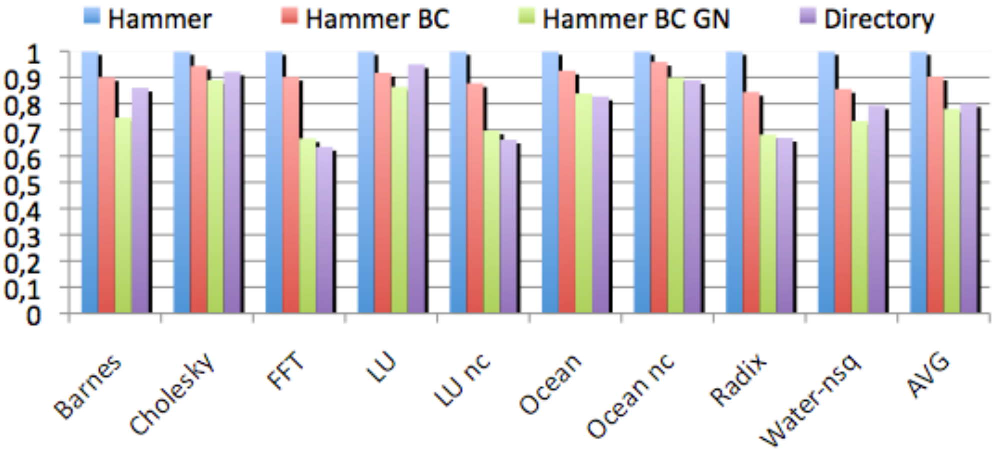
- Normalized number of injected messages

  reduced by 60% on average when using NoC broadcast support and the GCN

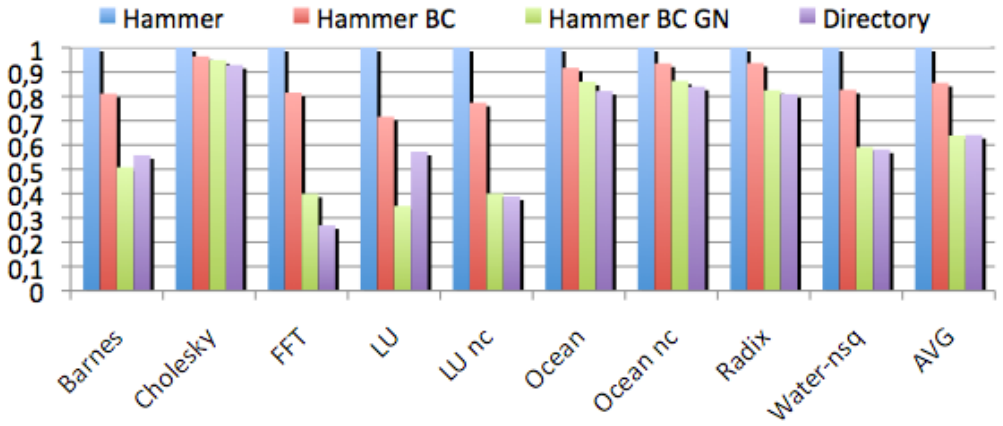


Performance Evaluation (2/3)

- Normalyzed load miss latency



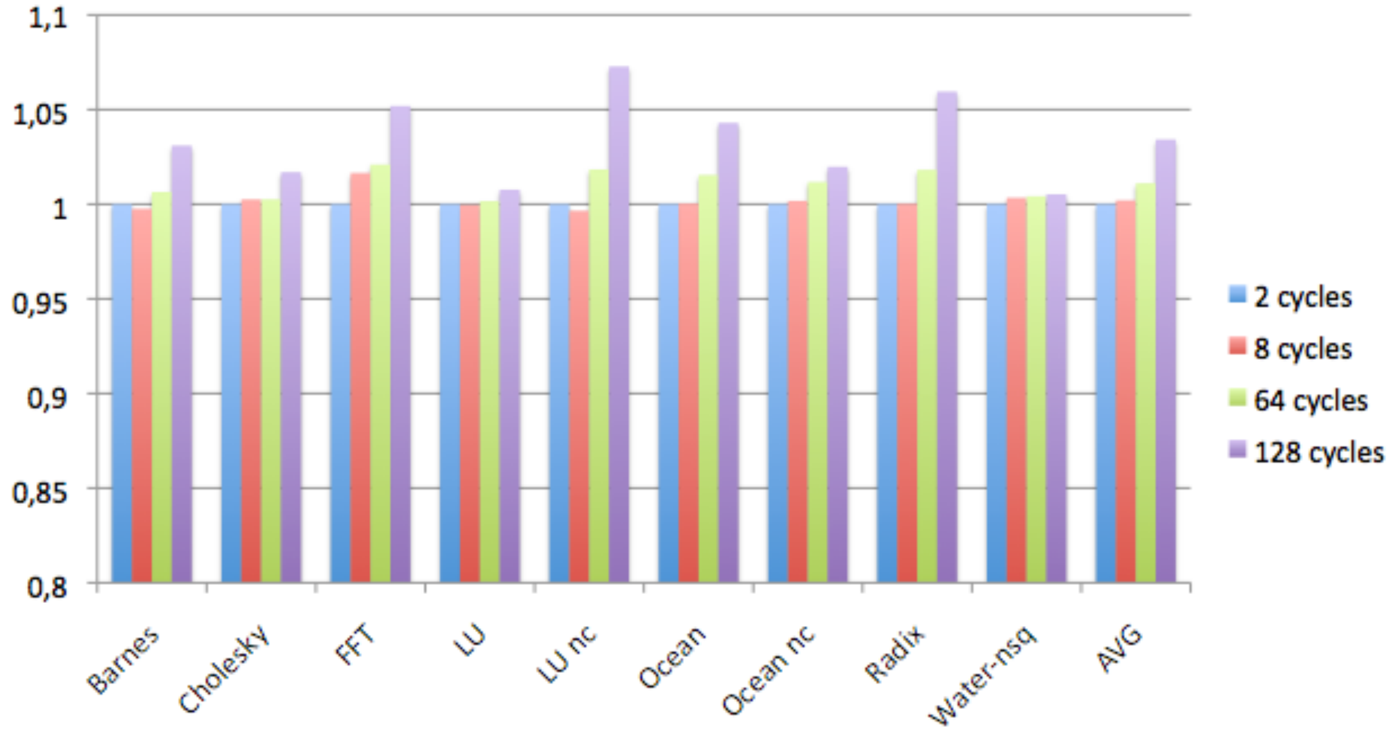
- Normalyzed store miss latency



Performance Evaluation (3/3)

- Normalized execution time with different GCN delays

Performance is not significantly affected for delays up to 64 cycles



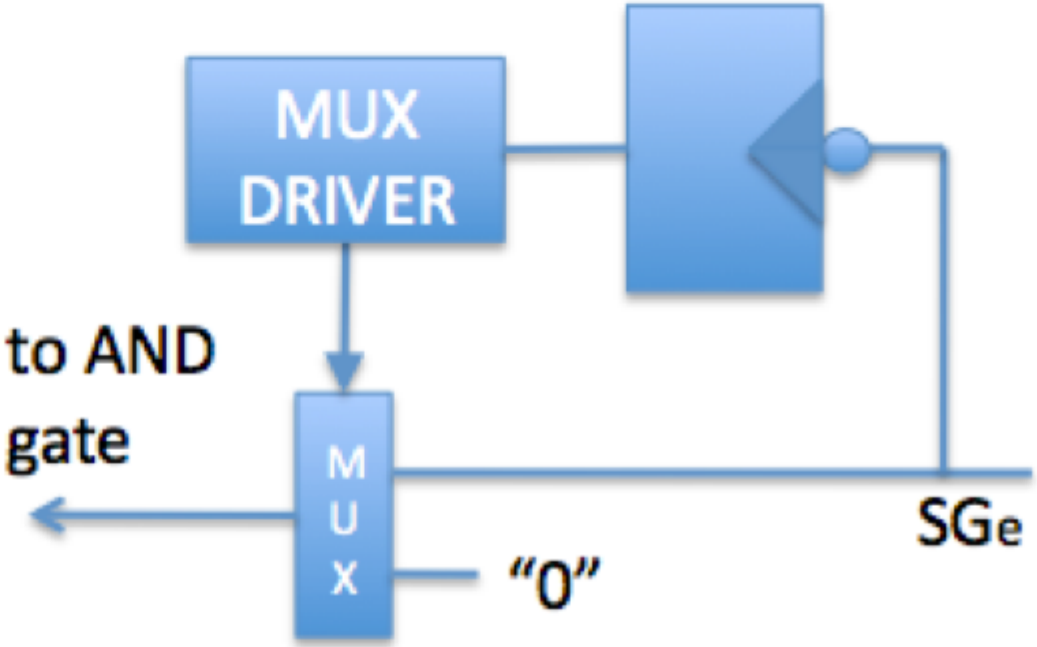
Conclusions

- Hammer protocol does not require large memory structures to keep the sharing code. However, its network traffic requirements impacts the performance significantly
- We extended a standard 2D mesh NoC and to support a dedicated gather control network to enable a fast notification of ACK messages
- Hammer protocol with NoC-level broadcast support and the GCN has better performance than directory protocol, without the area overhead due to the sharing code and generating the same amount of NoC traffic

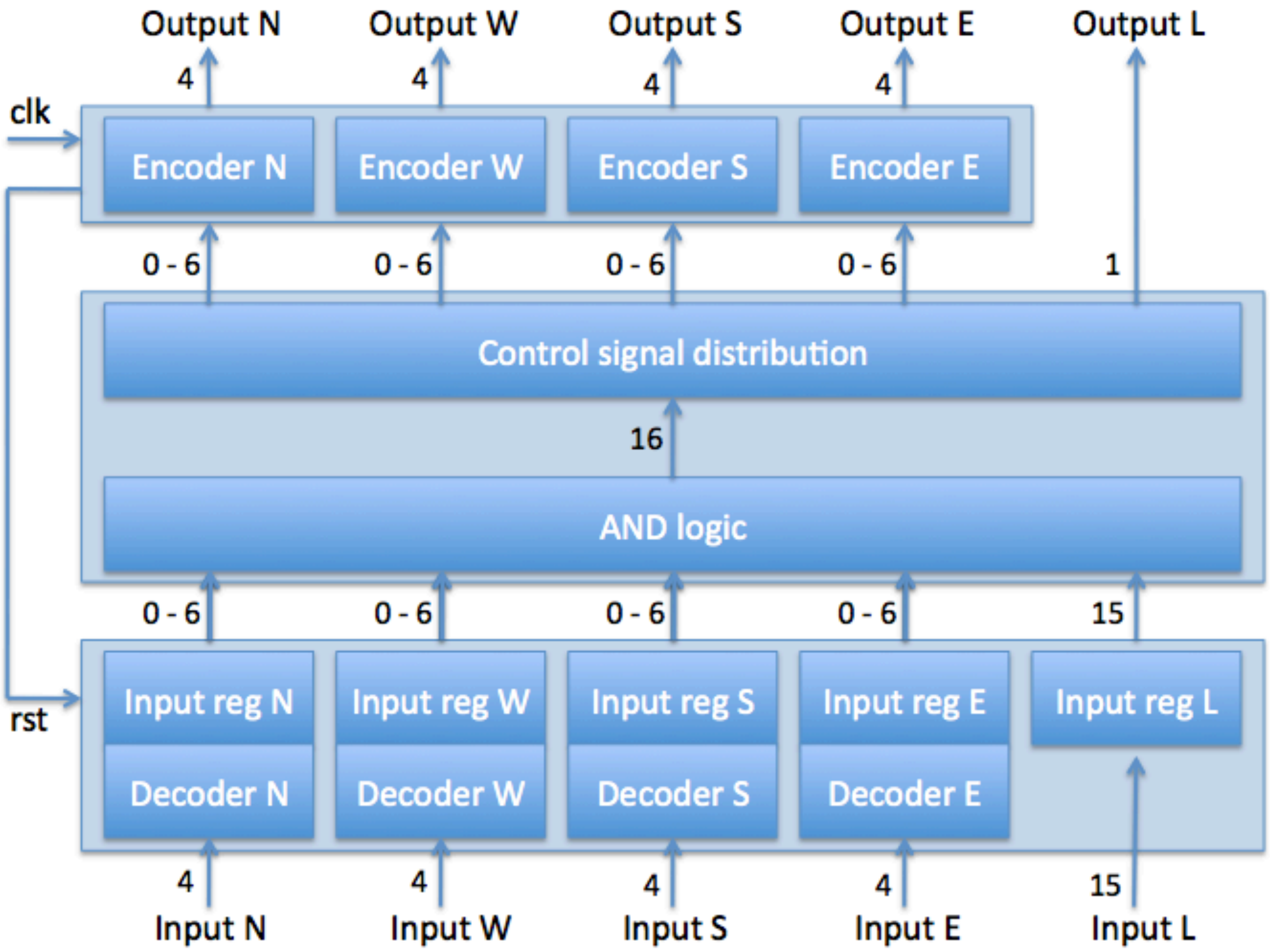
Thanks.

Any questions?

Logic at inputs of the AND gate (to reset the previous ACK gathering)



Sequential SAGN logic block


Output N
Output W
Output S
Output E
Output L
4
4
4
4
clk
Encoder N
Encoder W
Encoder S
Encoder E
0 - 6
0 - 6
0 - 6
0 - 6
1
Control signal distribution
16
AND logic
0 - 6
0 - 6
0 - 6
0 - 6
15
Input reg N
Input reg W
Input reg S
Input reg E
Input reg L
rst
Decoder N
Decoder W
Decoder S
Decoder E
4
4
4
4
15
Input N
Input W
Input S
Input E
Input L