

A Scalable and Power-Efficient Memory Hierarchy for Multicore Architectures

Enric Herrero¹, José González², Ramon Canal¹

¹Departament d'Arquitectura de Computadors
Universitat Politècnica de Catalunya

²Intel Barcelona Research Center
Microprocessor Technology Labs
Intel Corporation



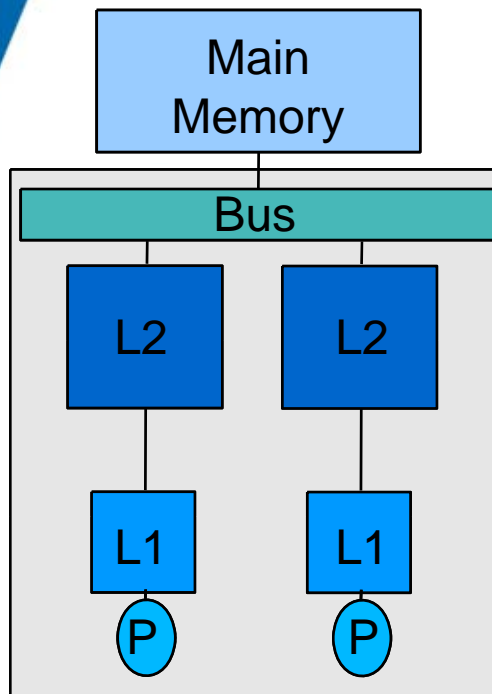
Departament d'Arquitectura
de Computadors

UNIVERSITAT POLITÈCNICA DE CATALUNYA

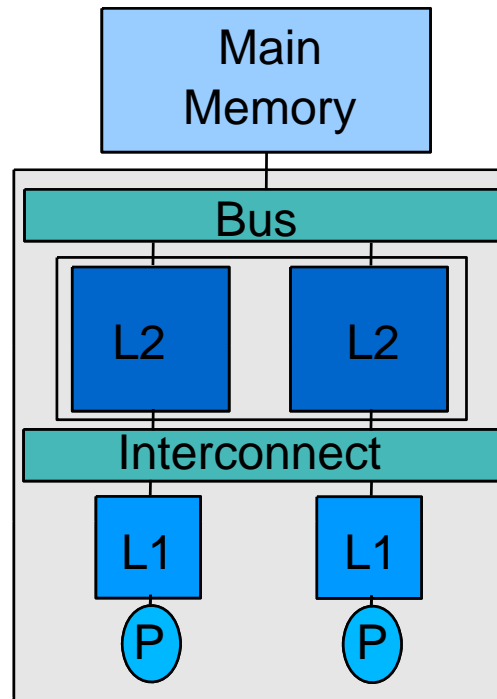
Multicore Organizations

State of the art
[Chang and Sohi '06]

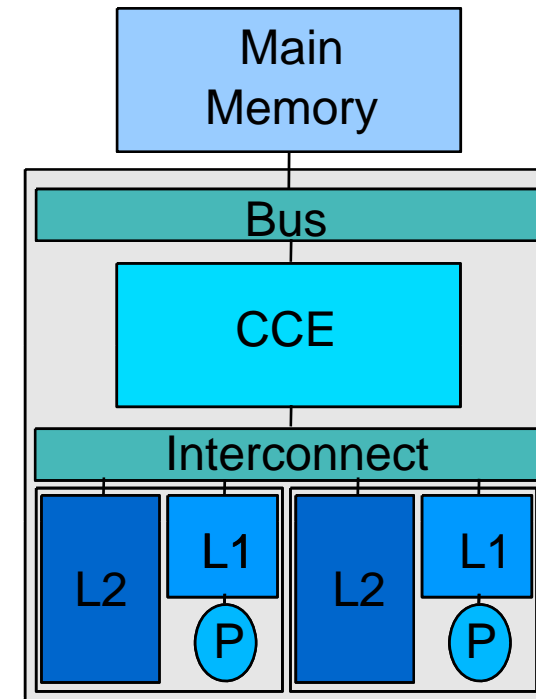
Traditional



Private Memory
Caching



Shared Memory
Physically distributed and
logically unified



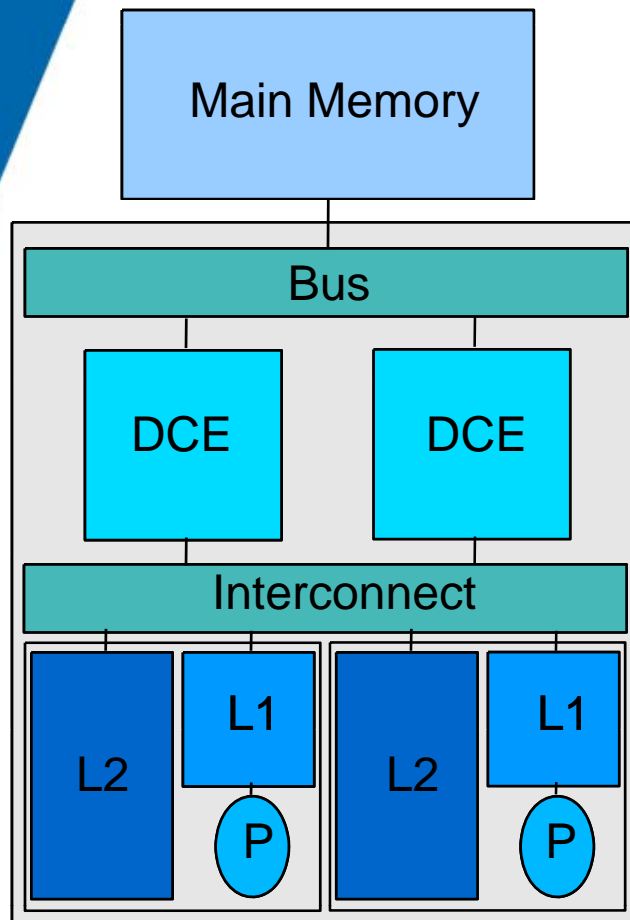
Cooperative



Our Proposal

Distributed Cooperative Caching

- Main Idea
- Private L2 to reduce access latency.
 - Distributed Directory (DCEs) to allow intra-chip sharing and reduce off-chip misses.



Distributed Coherence Engines (Distributed directory)

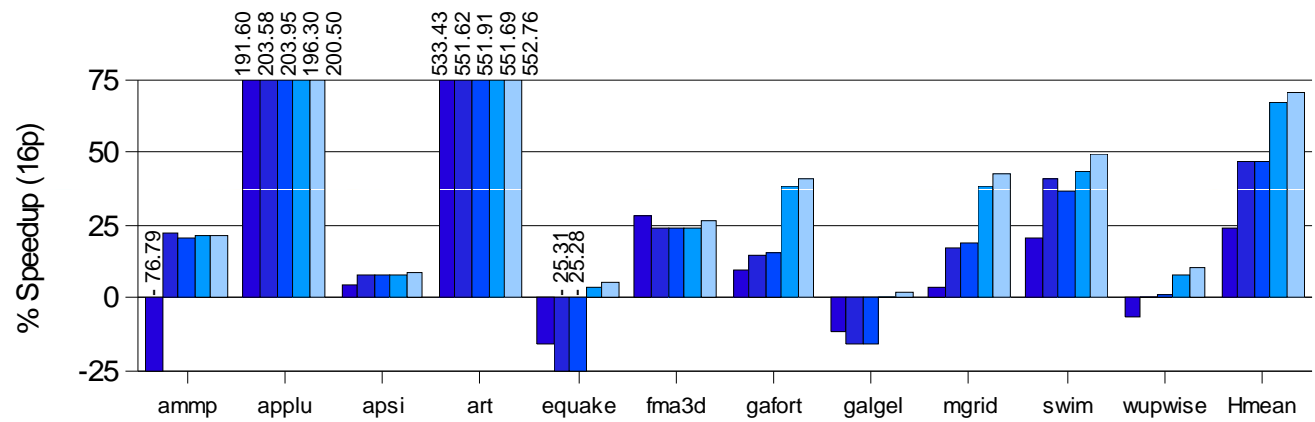
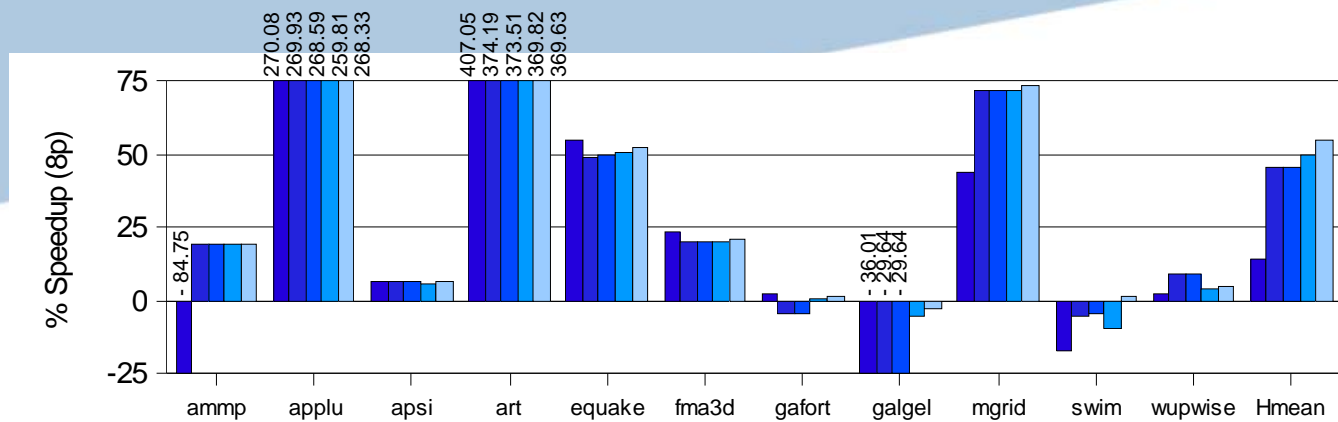
- Tags in the DCEs are not a replica of L1 and L2, every address is mapped to a unique DCE.
- Directory partitioning allows on-chip network traffic distribution and improves scalability
- Power is drastically reduced compared to Coop. Caching (less tags are accessed)
- Extra invalidation mechanism added for handling DCE replacements.

Spilling: Evicted blocks can be redirected to other L2 to make a efficient use of the cache space.

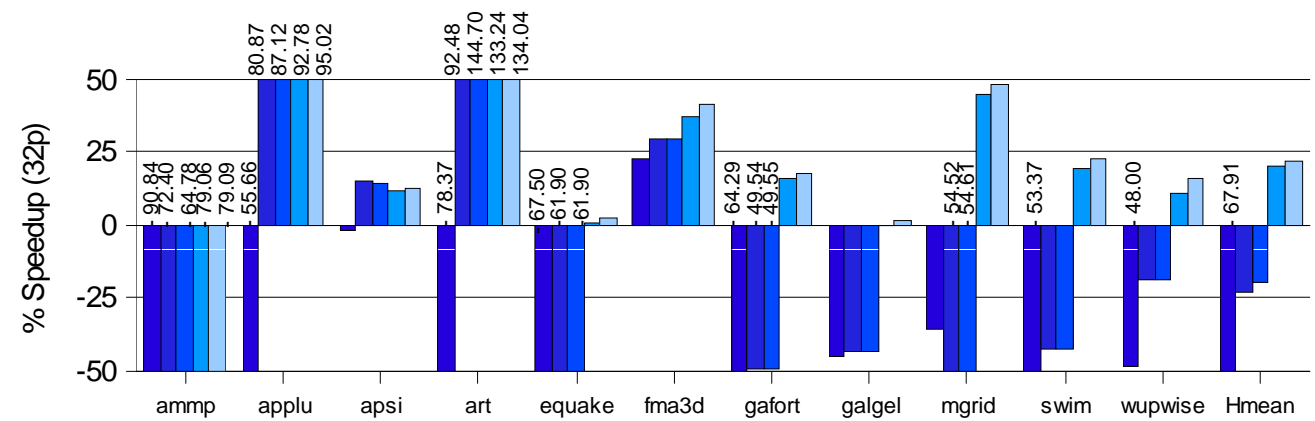
Simulation Environment

Processors	8, 16, 32	Instruction Set	Sparc V9
Processor_Type	Superscalar OoO	Instruction Window Size	16
Network Structure	Mesh with 2 VNCs	Reorder Buffer Size	48
Virtual Networks	2 (Request, Response)	Max Instr/Cycle	3
Link Bandwidth	16 bytes/cycle	Branch Predictor	YAGS
Technology	70 nm	L1 Cache	16 kb 4-way
Frequency	4 Ghz	L2 Cache	512, 256, 256 kb 4-way
Voltage	1.1 V	Memory Size	4 Gbytes
Benchmark	SPEC OMP 2001	TBE Entries	32
Simulator	Simics full-system execution-driven simulator with the GEMS toolset to provide a detailed memory hierarchy model. Extended with a power model to evaluate the energy savings of each configuration.		





Better power/performance relation (MIPS³/W).
 3.66x over Shared memory and 4.30x over Cooperative Caching



Higher performance
 21% performance increase over a Shared memory configuration and 57% over the Cooperative Caching scheme

■ Private Memory
 ■ CC2T
 ■ CC4T
 ■ DCC
 ■ DCC2x



Publications

1. E. Herrero, J. González and R. Canal, “Distributed Cooperative Caching”, In the *International Conference on Parallel Architectures and Compilation Techniques (PACT)* (October 2008).
2. E. Herrero, J. González and R. Canal, “Architectural Level Power Simulator of the Memory Hierarchy of Chip Multiprocessors”, *UPC Technical Report UPC-DAC-RR-ARCO-2008-3* (June 2008).

For further information contact Enric Herrero at eherrero@ac.upc.edu

