

Elastic Cooperative Caching:

An Autonomous Dynamically Adaptive Memory
Hierarchy for Chip Multiprocessors

Enric Herrero¹, José González², Ramon Canal¹

¹Universitat Politècnica de Catalunya

²Intel Barcelona

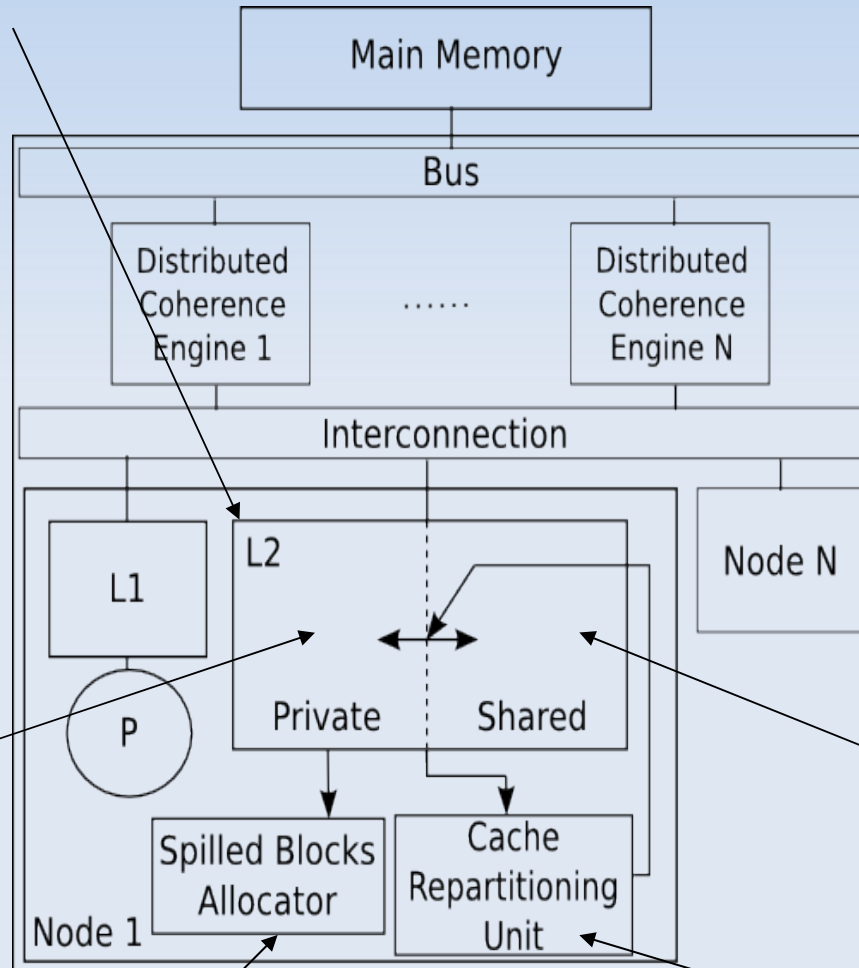
eherrero@ac.upc.edu



UNIVERSITAT POLITÈCNICA
DE CATALUNYA

ElasticCC Structure [ISCA'10]

Distributed cache among nodes.
Local allocation.



- Distributed structure for better scalability.

- Local private regions to minimize access latency and minimize inter-thread interference.

- Adaptive repartitioning to minimize off-chip misses

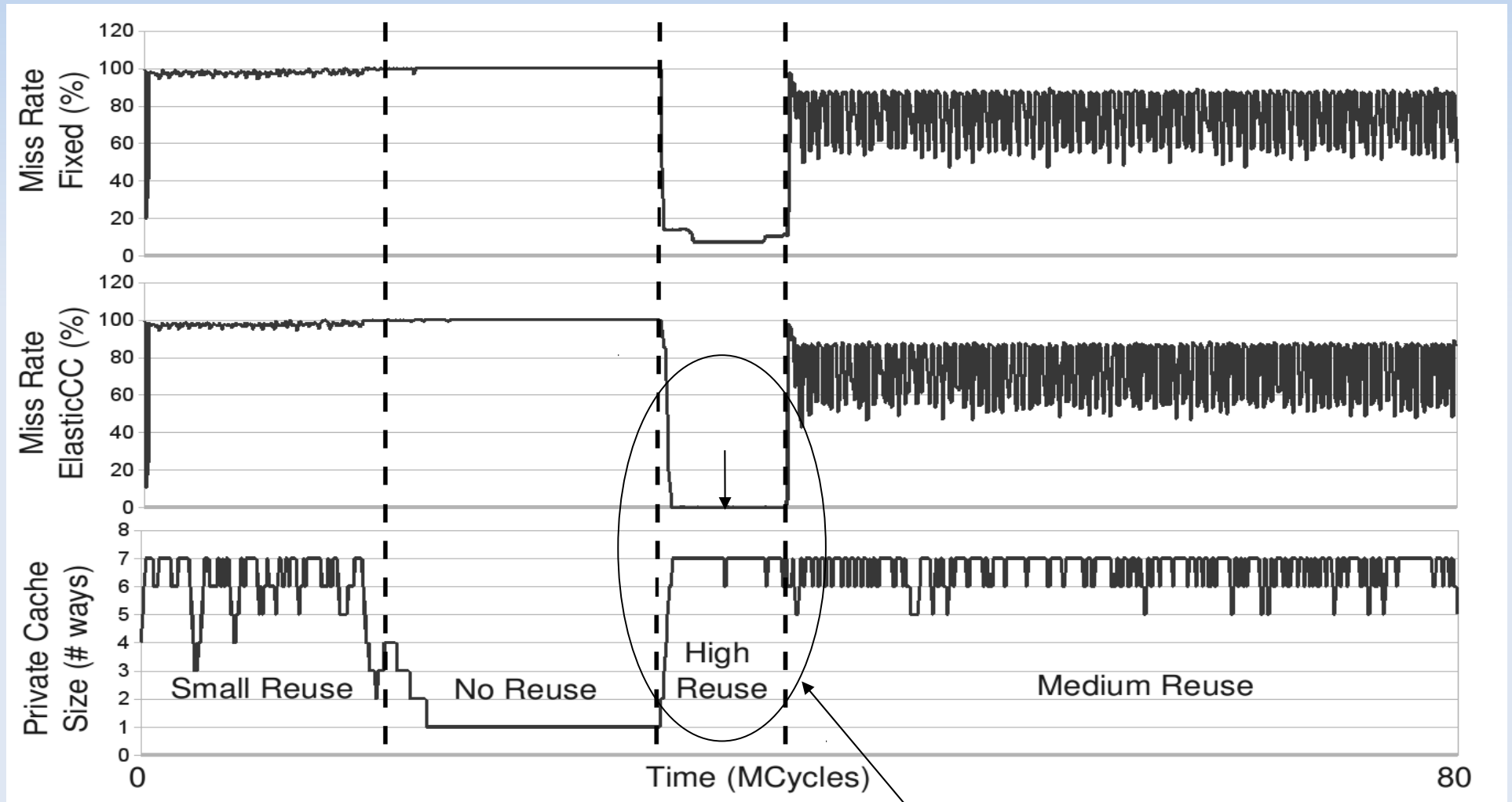
Allocates evicted blocks from private regions of all nodes

Every N cycles repartitions cache based on LRU hits in S&P partitions.

Private Regions, only local core can allocate.

Distributes evicted blocks from private partition among nodes.

Evaluation - Temporal Cache Behavior



Gafort-Equake execution, Equake Thread 1

Cache is given when it is useful

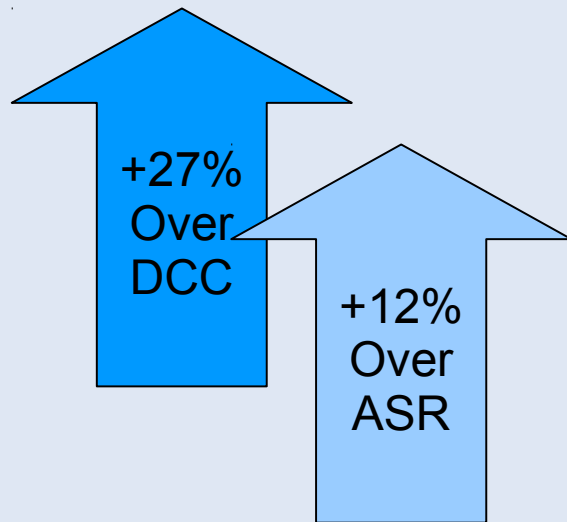
Cache Organization Comparison

Technique/ Processor	Last Level Cache type	Without software support	Distributed	Adaptive Partitioning	Adaptive Replication	Adaptive Allocation
Utility Based Partitioning [MICRO'06]	Private Partitions Centralized	Yes	No, centralized cache	Yes	No replication	Yes
Adaptive S/P NUCA [HPCA'07]	Shared/Private NUCA	Yes	No, centralized sharing engine	Yes	Limited by the fixed shared cache size	Yes
ESP-NUCA [HPCA'10]	Shared/Private/ HelpBlk NUCA	Yes	Yes	Yes	Yes	No, fixed mapping
R-NUCA [ISCA'09]	Instr/Shared/ Private NUCA	No, OS-page classification	Yes	Yes	Yes	No, fixed mapping
Adaptive Selective Replication [MICRO'06]	Private Local	Yes	Yes	No, fixed private caches	Yes	Yes
IBM Power 7	Shared Centralized	Yes	No, centralized cache	No, fixed shared cache	No replication	Yes
Intel Larrabee	Private Local	Yes	Yes	No, fixed private caches	Yes	Yes
Intel Advanced Smart Cache	Shared/Private Centralized	Yes	No, centralized cache	Yes	No replication	Yes
Elastic Cooperative Caching	Shared/Private Local	Yes	Yes	Yes	Yes	Yes

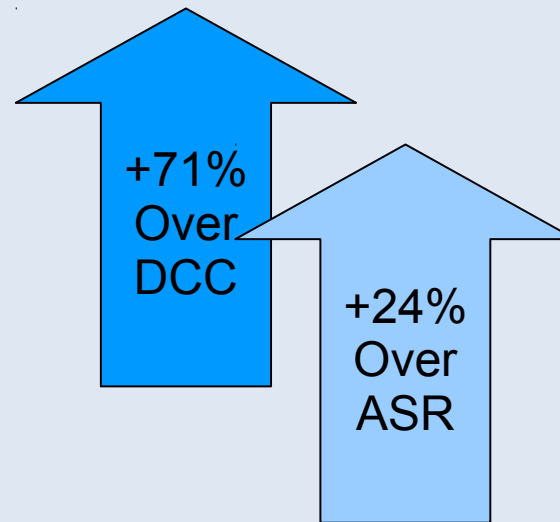
Overview

- Elastic Cooperative Caching
 - Distributed organization
 - Adaptive behavior to application requirements

Performance



Energy-Efficiency



Off-Chip Misses

