



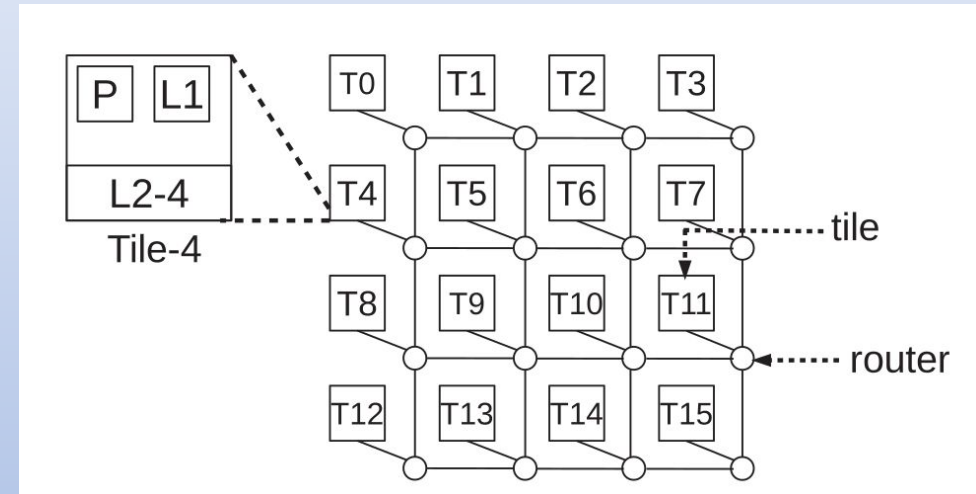
CS531: Memory Systems and Architecture

Course Instructor:

Dr. Shirshendu Das
Assistant Professor,
Department of CSE,
IIT Ropar.

shirshendu@iitrpr.ac.in

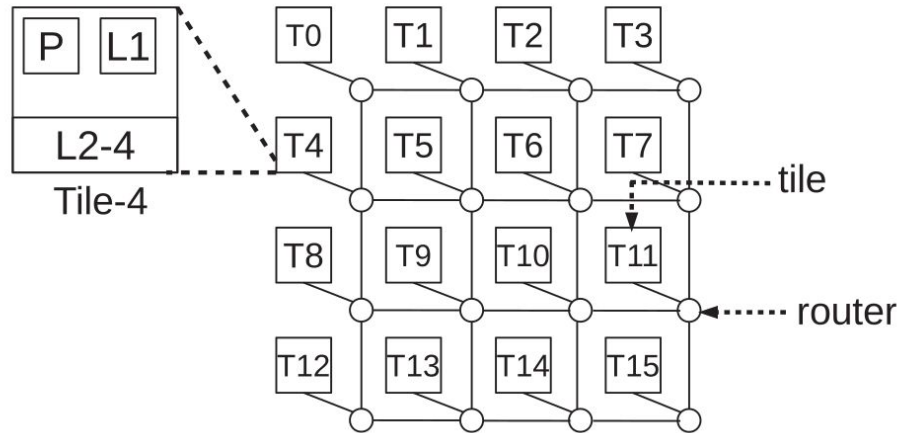
<http://cse.iitrpr.ac.in/shirshendu/shirshendu.html>



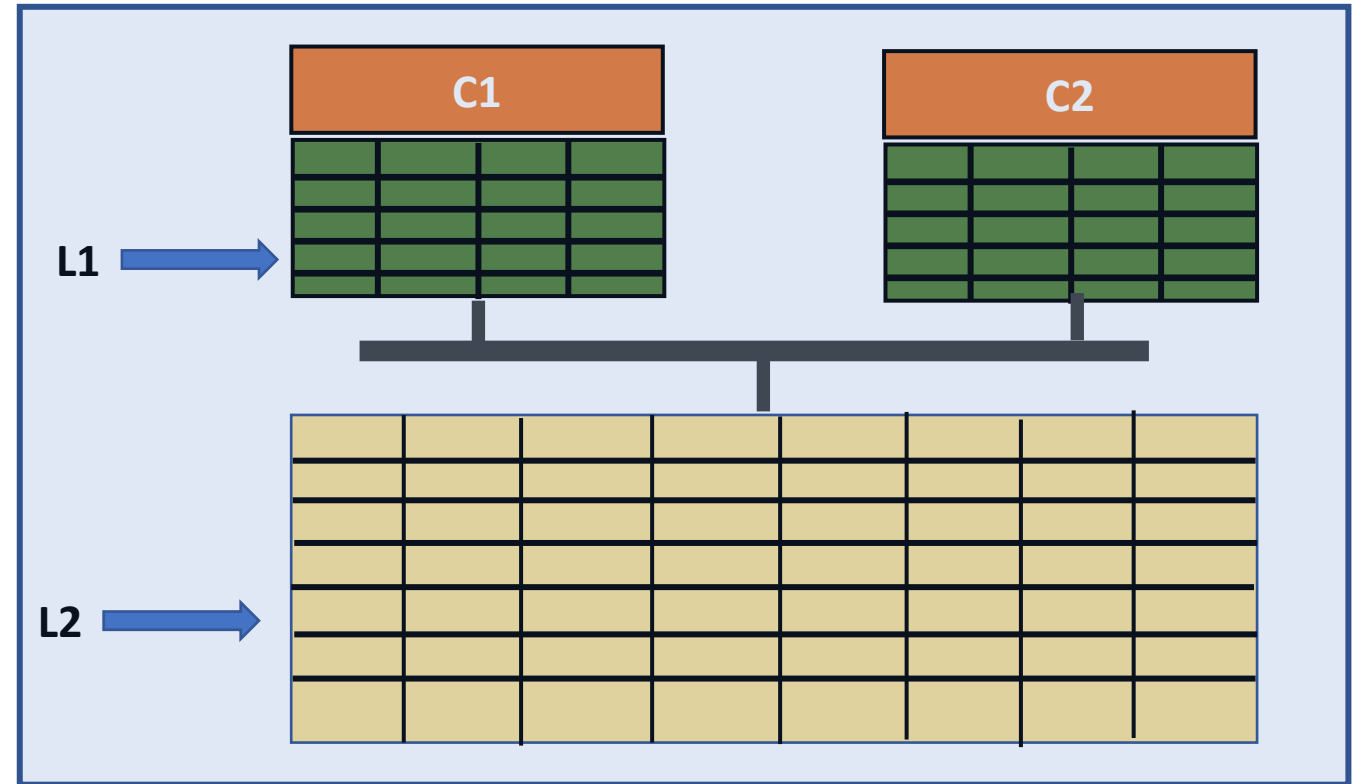
TCMP

Topic: Advancement in Replacement Policy – Part 2

Introduction 1



TCMP



CMP

◆ Important parameters:

- Performance.
- Energy consumption
- Hardware overheads.

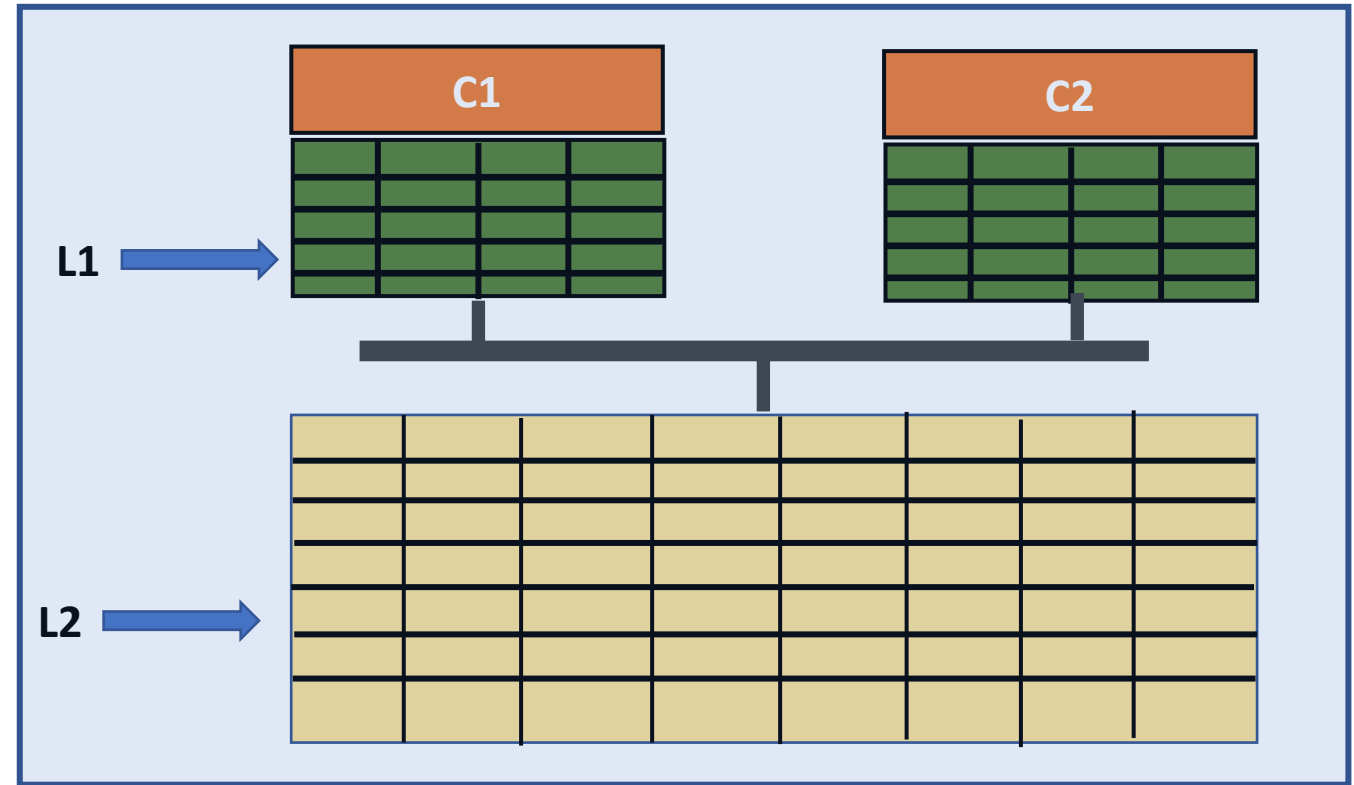
Introduction 2

❖ Important parameters:

- Performance.
- Energy consumption
- Hardware overheads.

❖ Important parameters:

- Scanning/Streaming.
- Thrashing.



CMP

Trace-1

:

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16
----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----

.....

Trace-2:

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16
----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----

Introduction 3: Coherence

❖ Execution style:

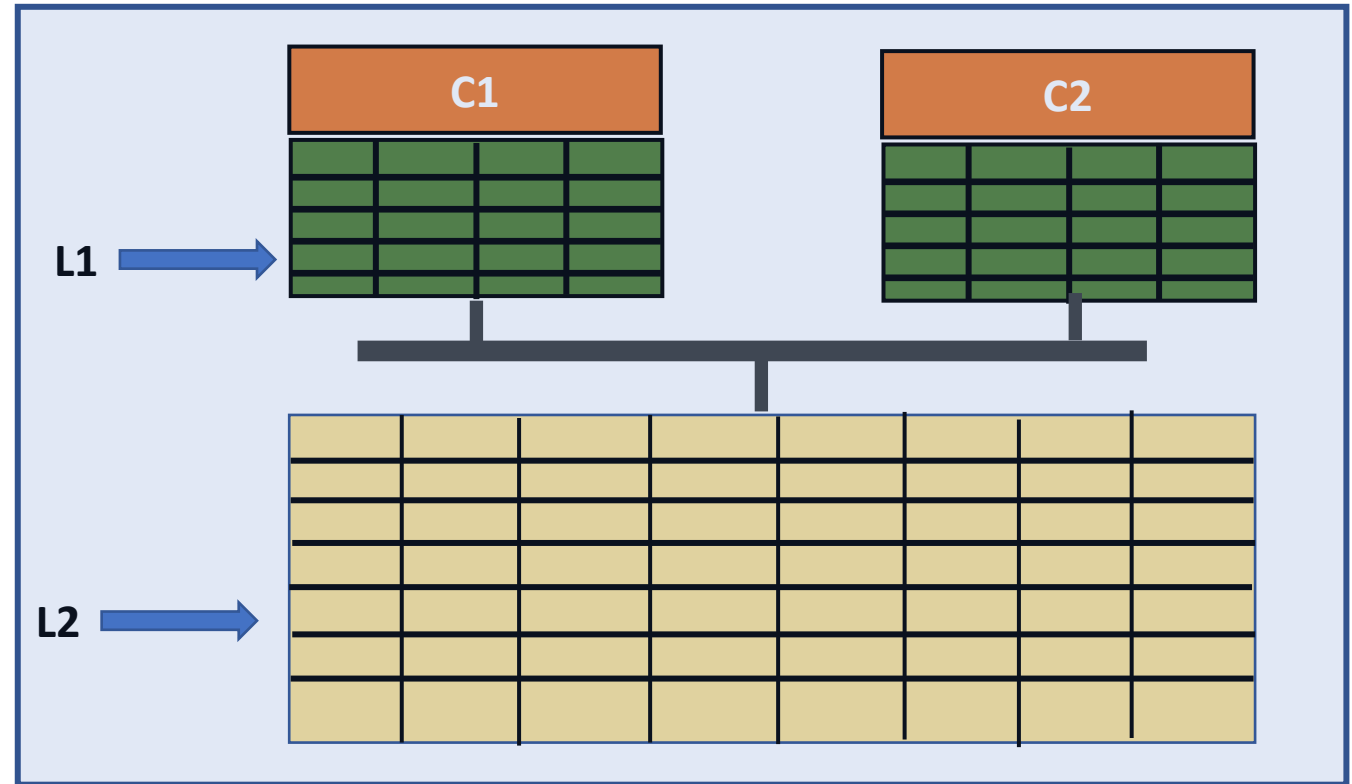
- ❑ Multiprogrammed.
- ❑ Multithreaded.

❖ Coherence Issues:

- ❖ Required to maintain the shared blocks consistent across the cores.

❖ Request Type:

- ❖ GetS
- ❖ GetX
- ❖ PutX
- ❖ Upgrade



Adaptive Insertion Policies for Managing Shared Caches

Aamer Jaleel[†] William Hasenplaugh[†] Moinuddin Qureshi[§] Julien Sebot[‡] Simon Steely Jr.[†] Joel Emer[†]

[†]Intel Corporation, VSSAD
Hudson, MA
{aamer.jaleel, william.c.hasenplaugh,
simon.c.steely.jr, joel.emer} @intel.com

[§]IBM T. J. Watson Research Center
Yorktown Heights, NY
mkquresh@us.ibm.com

[‡]Intel Israel Design Center
Haifa, Israel
julien.sebot@intel.com

*In Proceedings of the International Conference on Parallel Architectures and Compilation
Techniques (PACT 08), 2008*

Short name: TADIP

I have used some figures, tables and texts from the paper in this presentation to explain you the paper. The use is completely for academic purpose.

TADIP: Motivation 1

The LLC when shared by multiple applications or threads, cannot be handled efficiently by DIP.

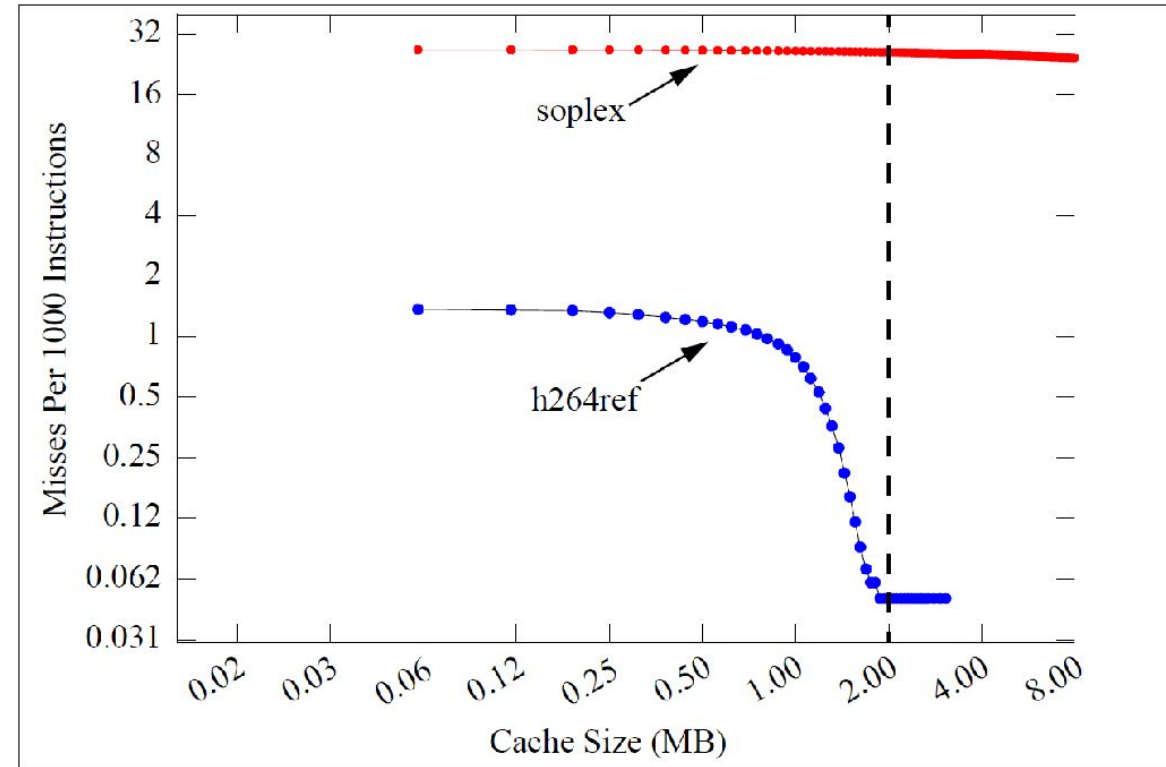
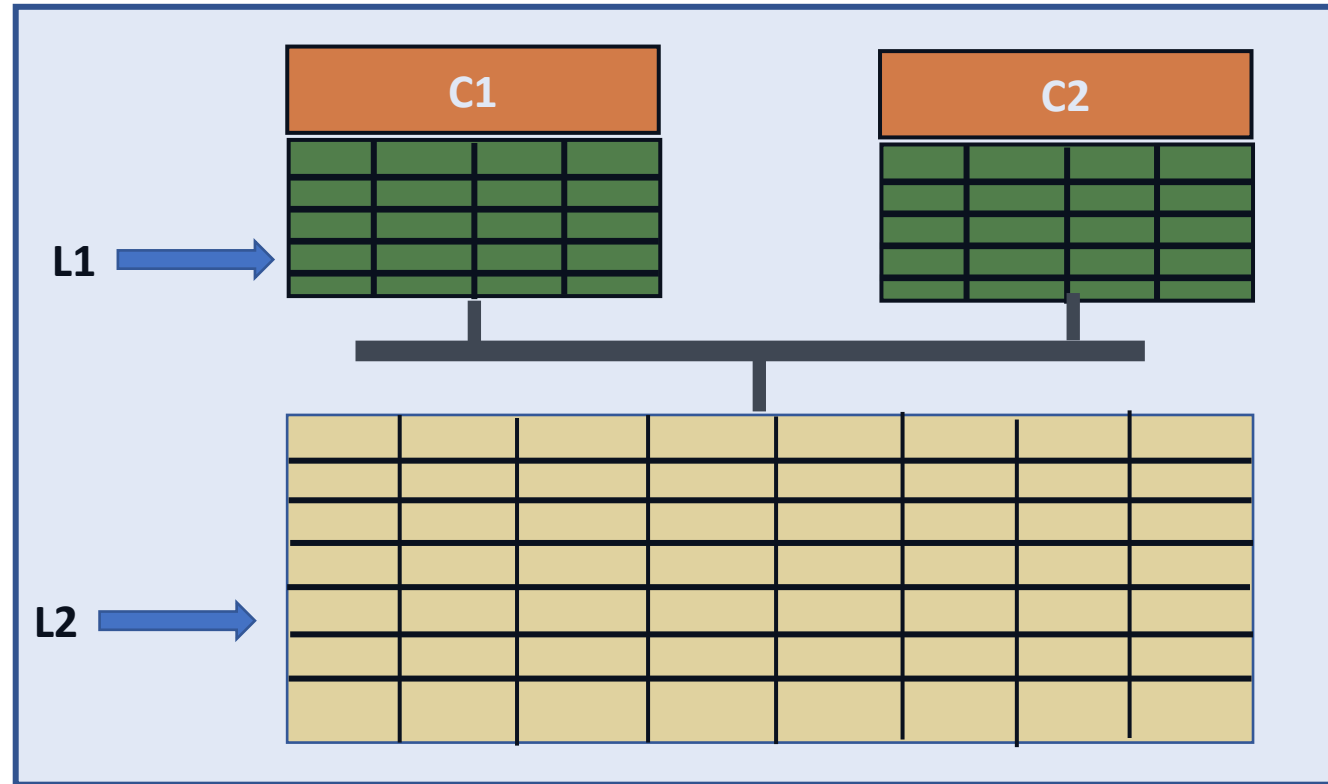


Figure 1: The Shared Cache Problem. The figure shows the cache sensitivity (under LRU) of two SPEC CPU2006 workloads. When both these workloads execute concurrently and share a 2MB cache, *soplex*, a streaming application, interferes with *h264ref*. Cache performance can be improved by reducing the interference.

TADIP: Motivation

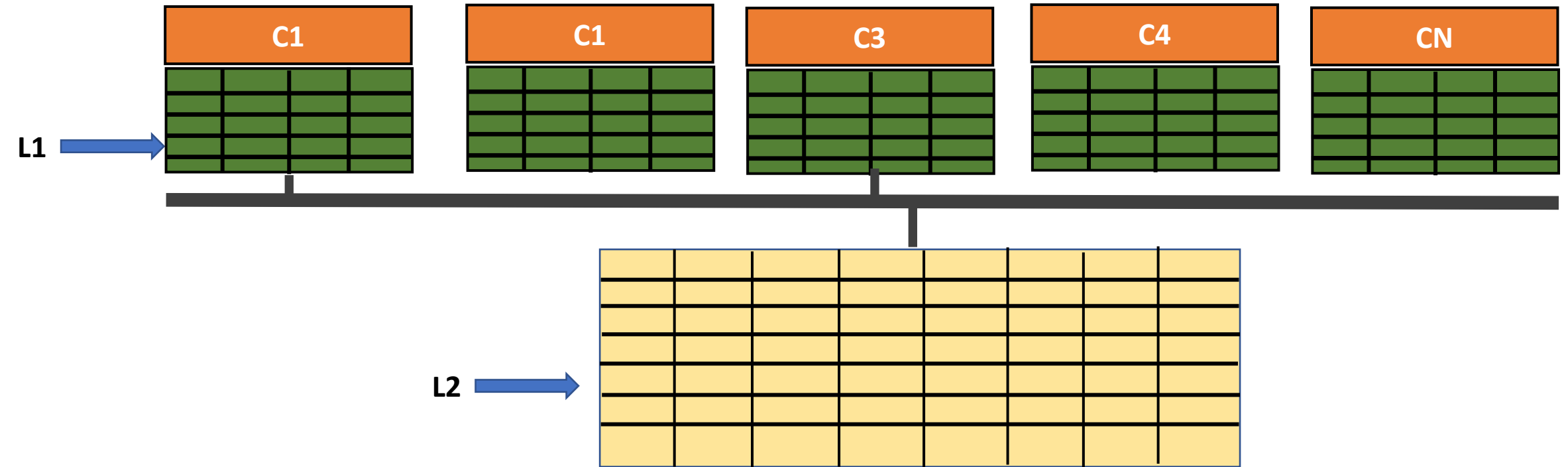


The main idea of TADIP is to apply DIP on each application (thread) individually.

0 : LRU
1 : BIP

00 01 10 11

TADIP



0 : LRU

1 : BIP

DIP: Motivation 2

◆ Difference between Cache friendly and non-cache friendly applications.

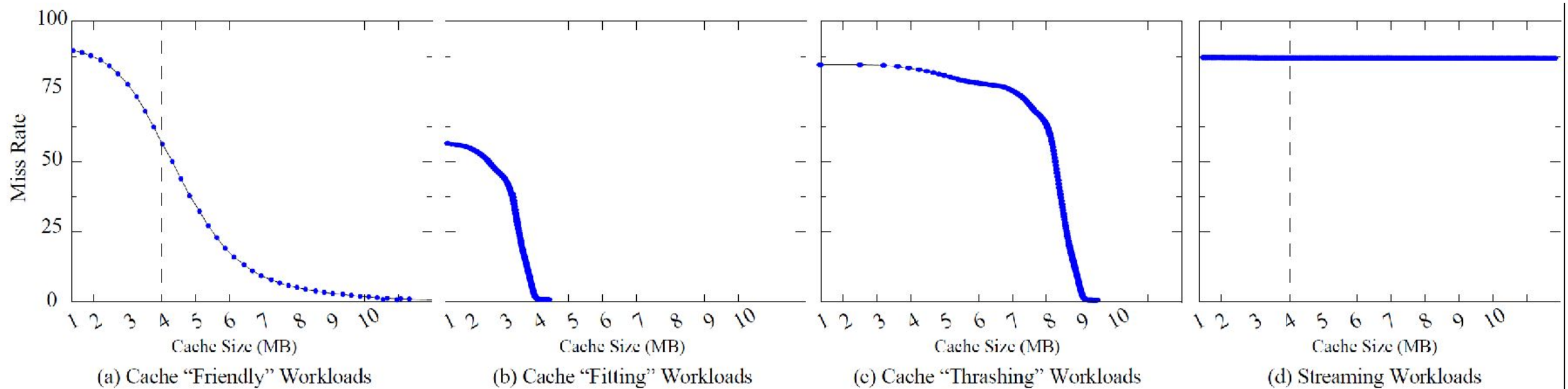
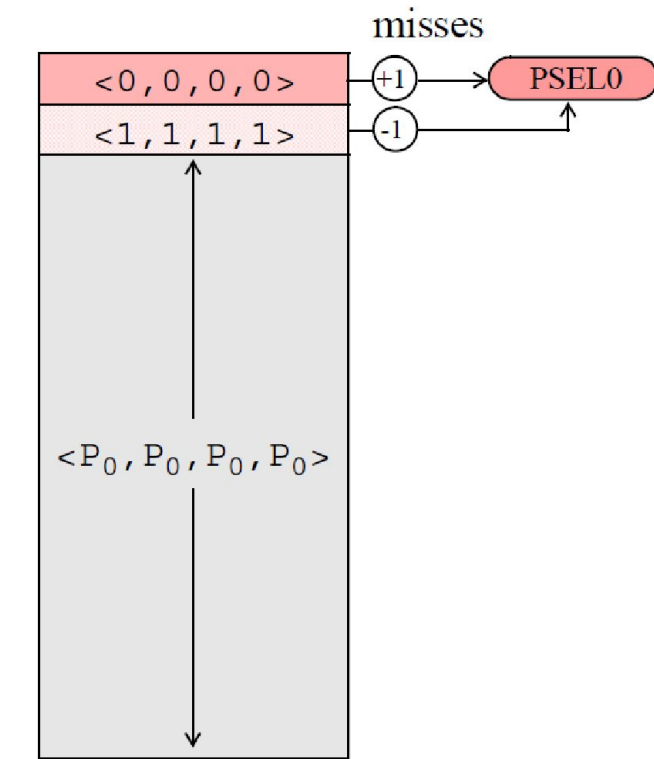
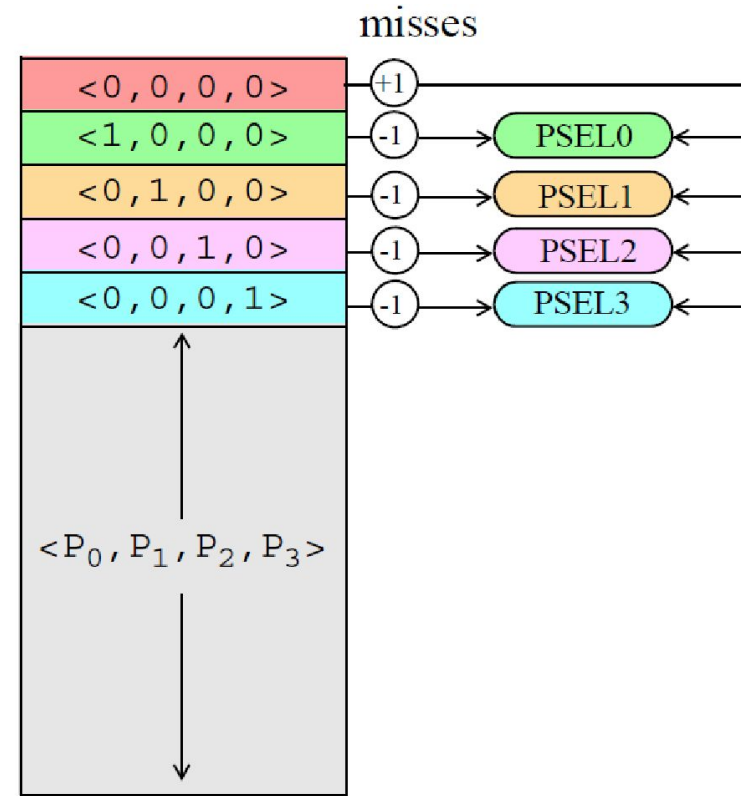


Figure 2: Workload Diversity on CMPs. Assuming a 4MB shared cache, this figure shows the diversity (in terms of cache requirements) of applications that can compete for the shared cache.

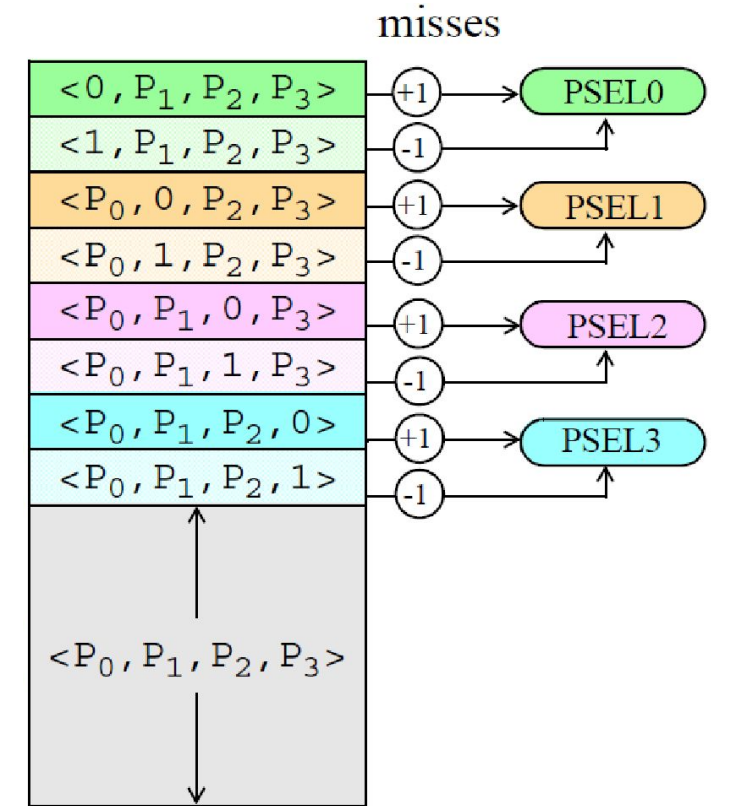
TADIP: Proposed Idea



(a) DIP (Thread Unaware)



(b) TADIP-I



(c) TADIP-F

$$P_x = \text{MSB}(\text{PSEL}_x)$$

I have explained the proposed idea from this diagrams. To understand TADIP either you watch my lecture or read the paper. Just reading this PPT will not be enough.

Thank You.