Multi core processor systems

Foreword

In a relative short time interval of about two years multicore processors became the standard. As multiple cores provide a substantially higher processing performance than their predecessors, their memory and I/O subsystems need also be largely enhanced to avoid processing bottlenecks. Consequently, while discussing multicore processors the focus of interest shifts from their microarchitecture to their macroarchitecture incorporating their cache hierarchy, the way how memory and I/O controllers are attached and how on-chip interconnects needed are implemented.

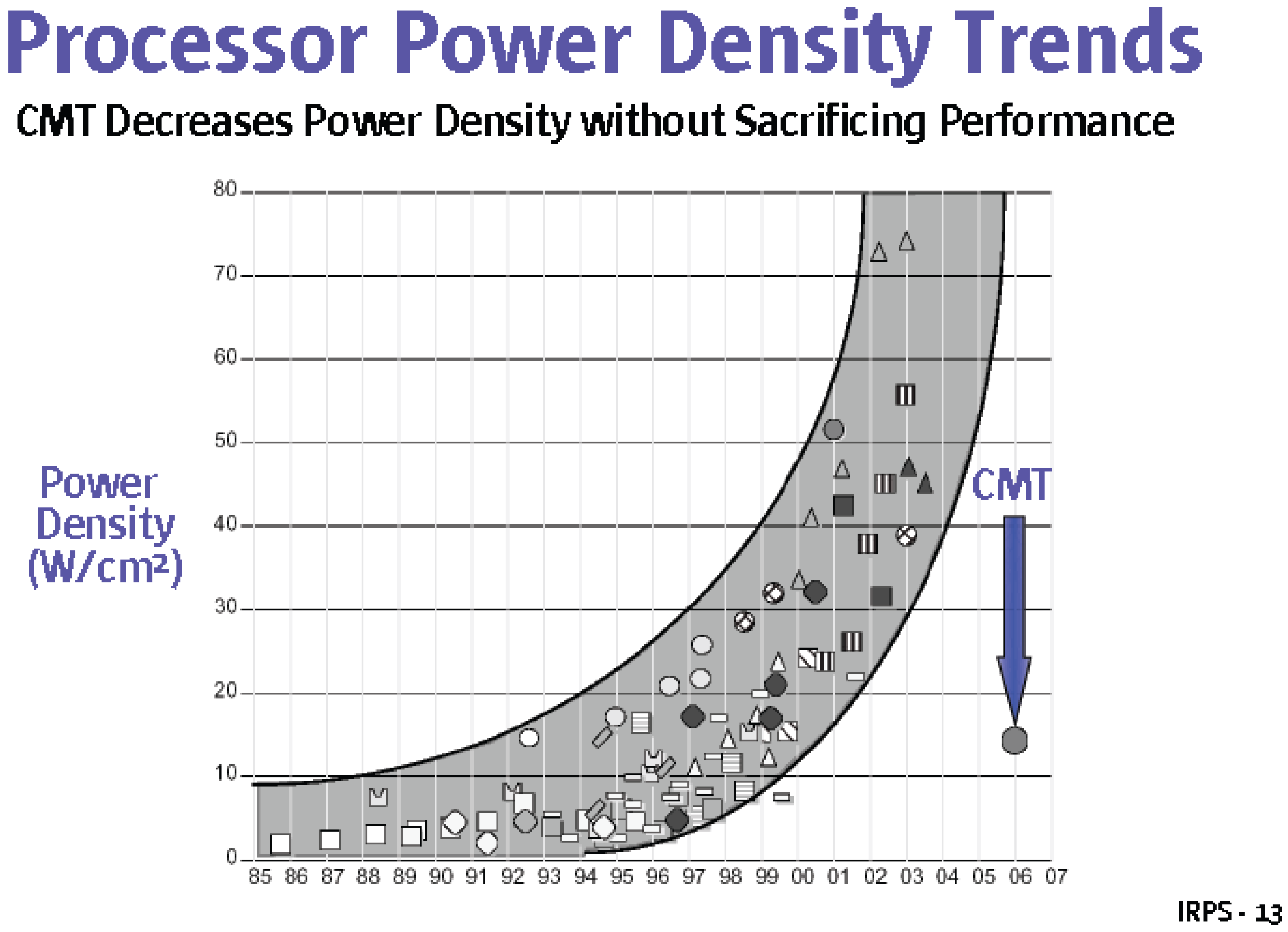
Accordingly, the first nine Sections are devoted to the design space of the macroachitecture of multicore processors. Particular Sections deal with the main dimensions that span the design space and identify how occuring concepts became implemented in major multicore lines.

Section 10 is a huge repository, providing relevant materials and facts published in connection with major multicore lines. In a concrete course appropriate parts of the repository can be selected according to the scope and attendance of the course. Each part of the repository concludes with a list of available literature to allow a more detailed study of a multicore line or processor of interest.

Overview

- 1 Introduction
- 2 Overview of MCPs
- 3 The design space of the macroarchitecture of MCPs
- 4 Layout of the cores
- 5 Layout of L2 caches
- 6 Layout of L3 caches
- 7 Layout of the memory and I/O architecture
- 8 Implementation of the on-chip interconnections
- 9 Basic alternatives of the macro-architecture of MCPs
- 10 Multicore repository

1. Introduction



Processor Power Density Trends

CMT Decreases Power Density without Sacrificing Performance

IRPS - 13

Figure 1.1: Processor power density trends

Source: D. Yen: Chip Multithreading Processors Enable Reliable High Throughput Computing
http://www.irps.org/05-43rd/IRPS_Keynote_Yen.pdf

1. Introduction (2)

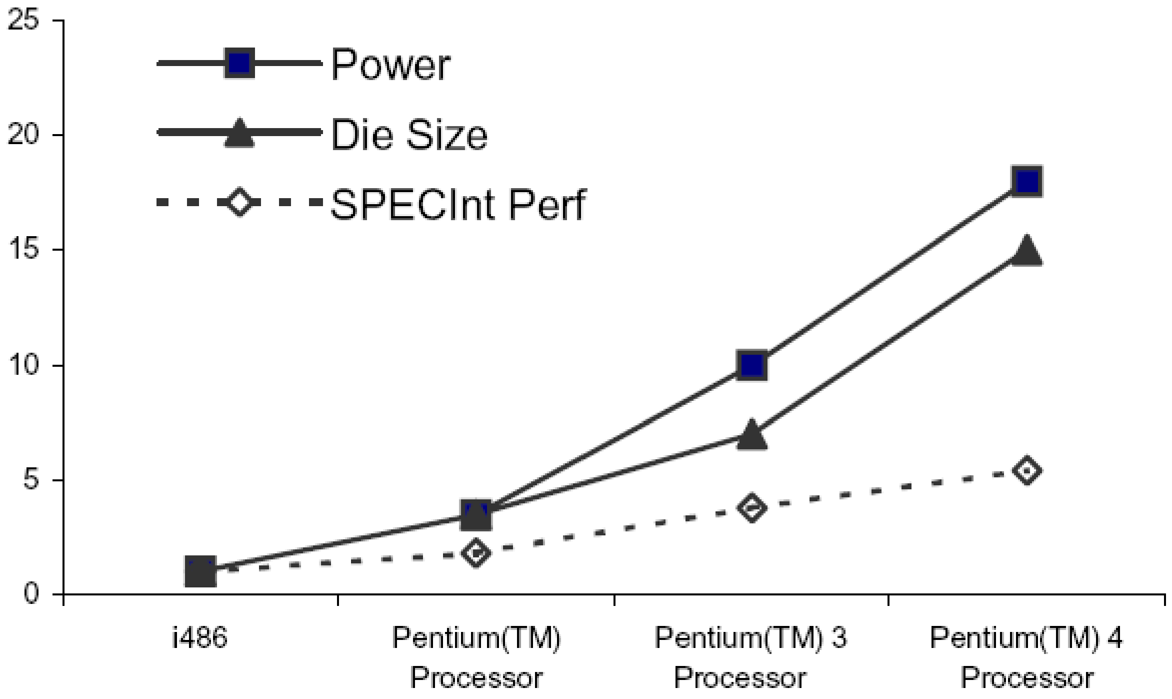


Figure 1.2: Single-stream performance vs. cost

Source: Marr T.T. et al. „Hyper-Threading Technology Architecture and Microarchitecture
Intel Technology Journal, Vol. 06, Issue 01, Febr 14, 2002, pp. 4-16

1. Introduction (3)

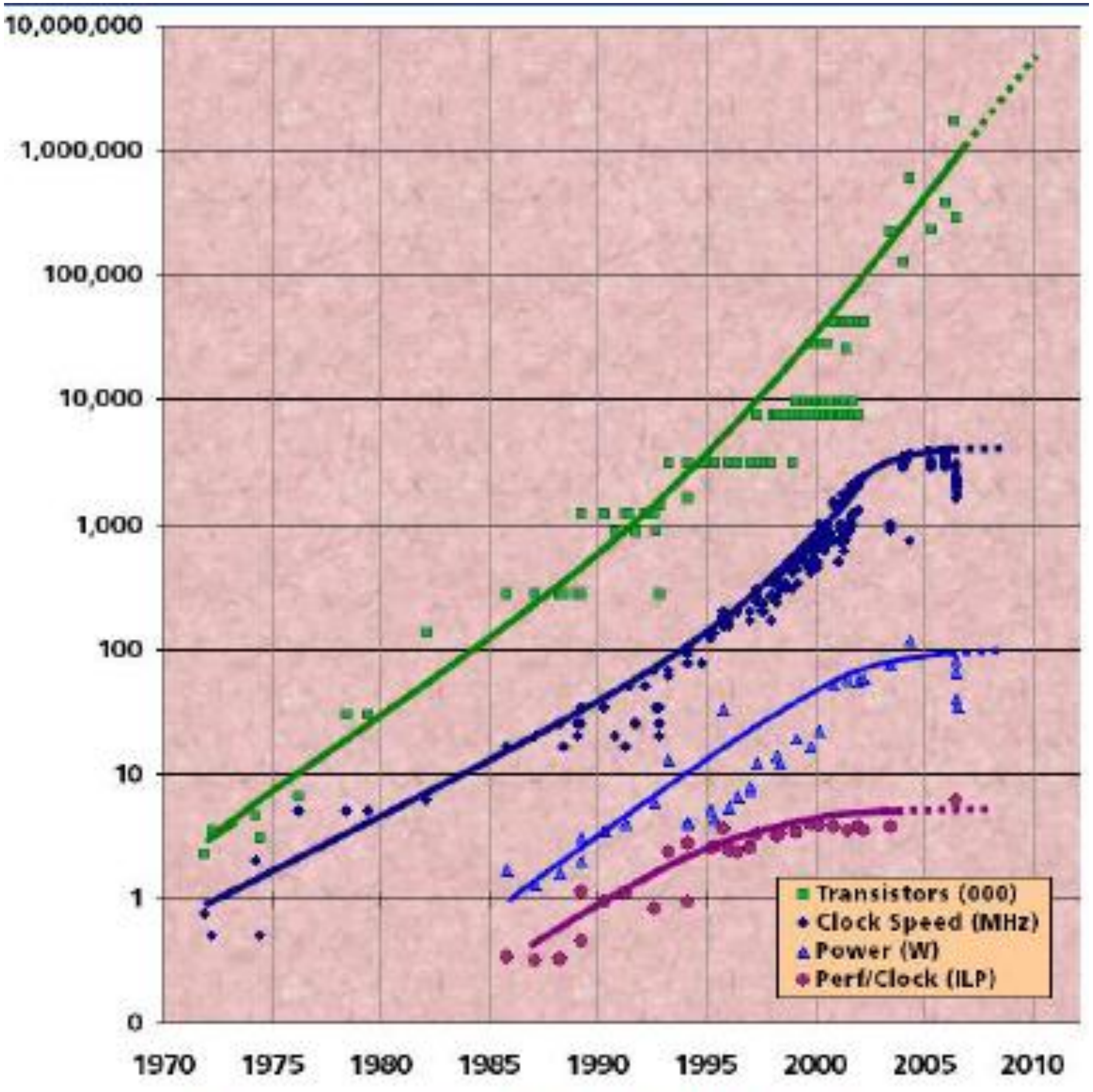


Figure courtesy of Kunle Olukotun, Lance Hammond, Herb Sutter, and Burton Smith

Figure 1.3: Historical evolution of transistor counts, clock speed, power and ILP

2. Overview of MCPs

2. Overview of MCPs (1)

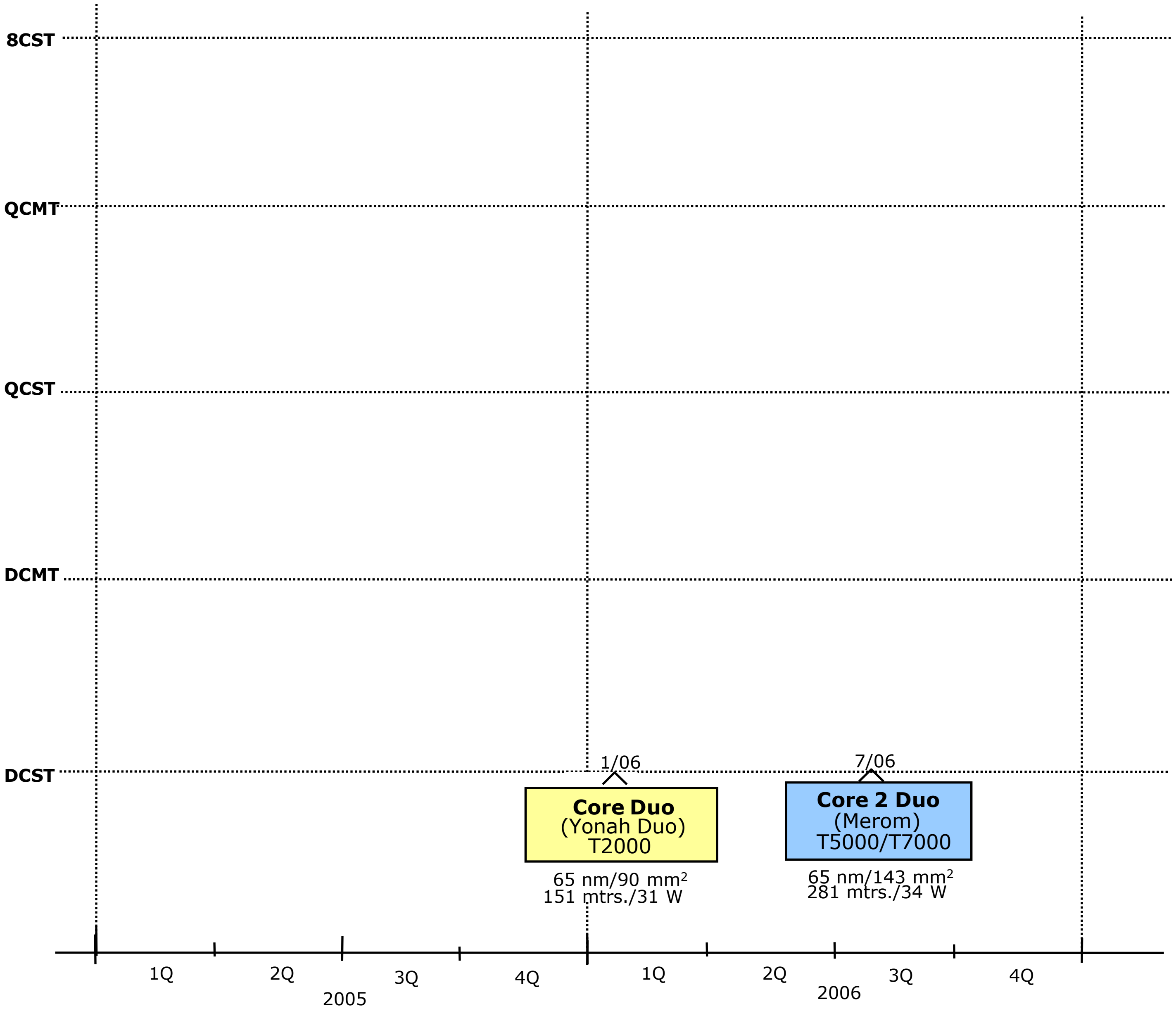


Figure 2.1: Intel’s dual-core mobile processor lines

2. Overview of MCPs (2)

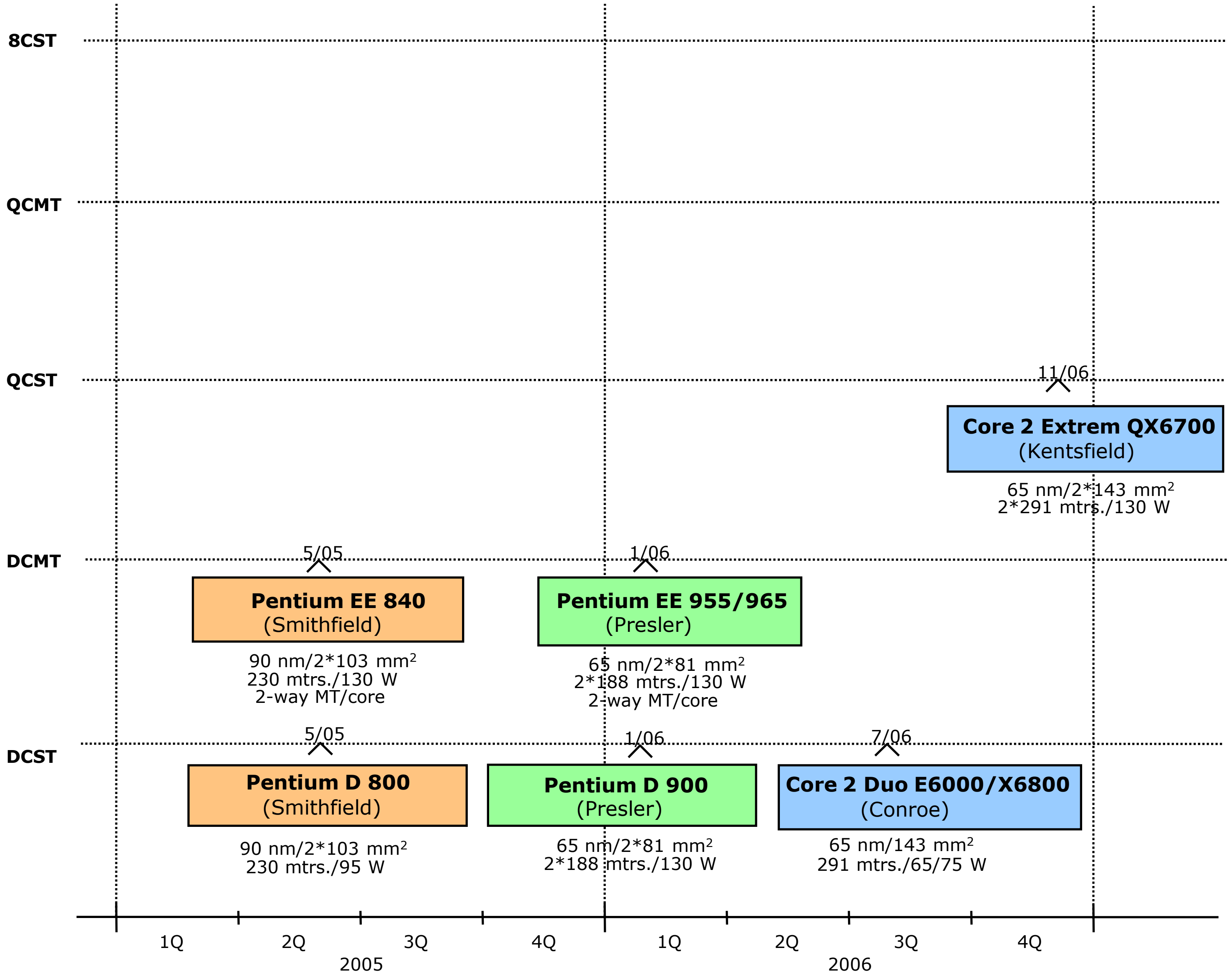


Figure 2.2: Intel’s multi-core processor lines

2. Overview of MCPs (3)

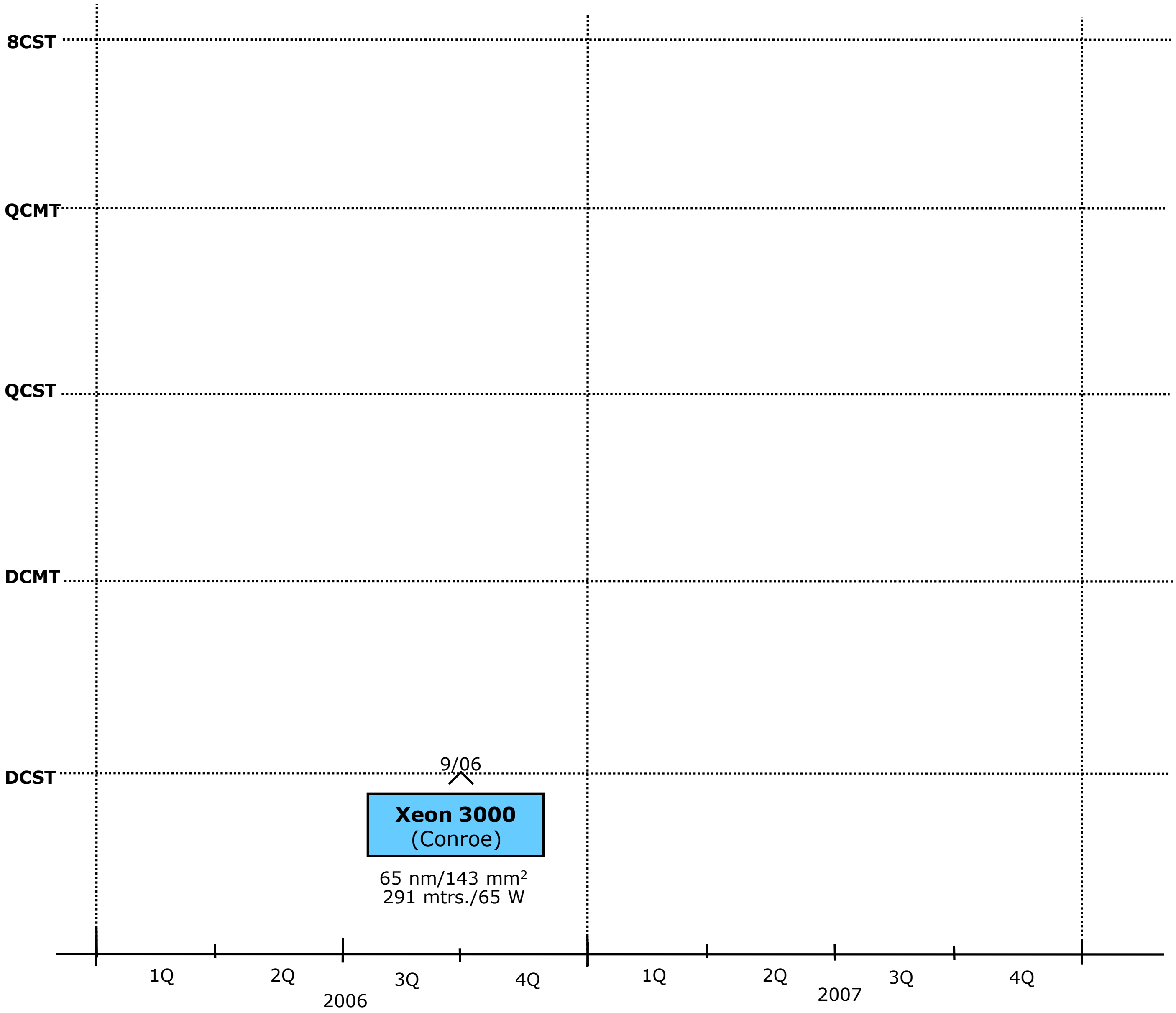


Figure 2.3: Intel’s dual-core Xeon UP-line

2. Overview of MCPs (4)

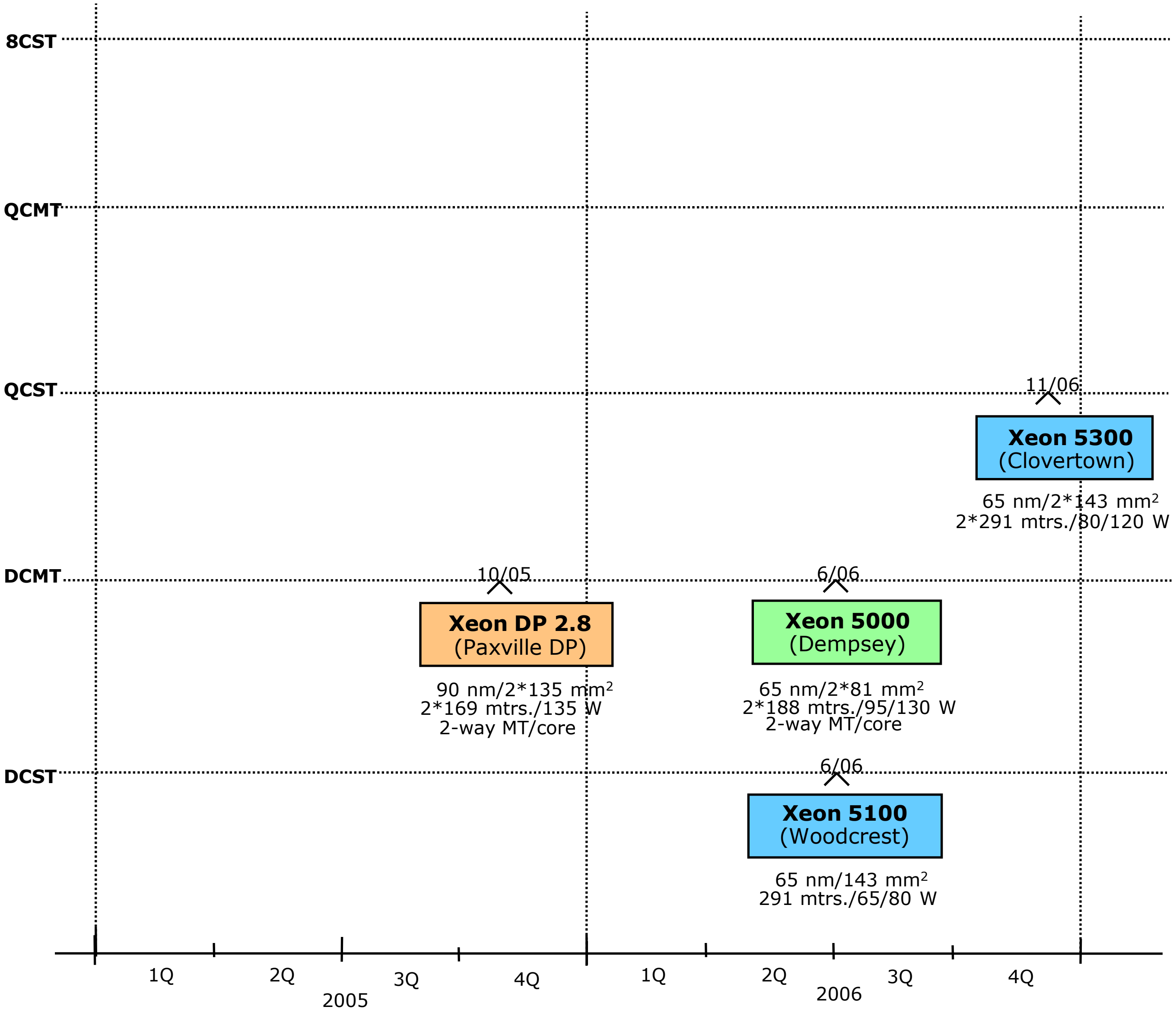


Figure 2.4.: Intel's dual-core Xeon DP-lines

2. Overview of MCPs (5)

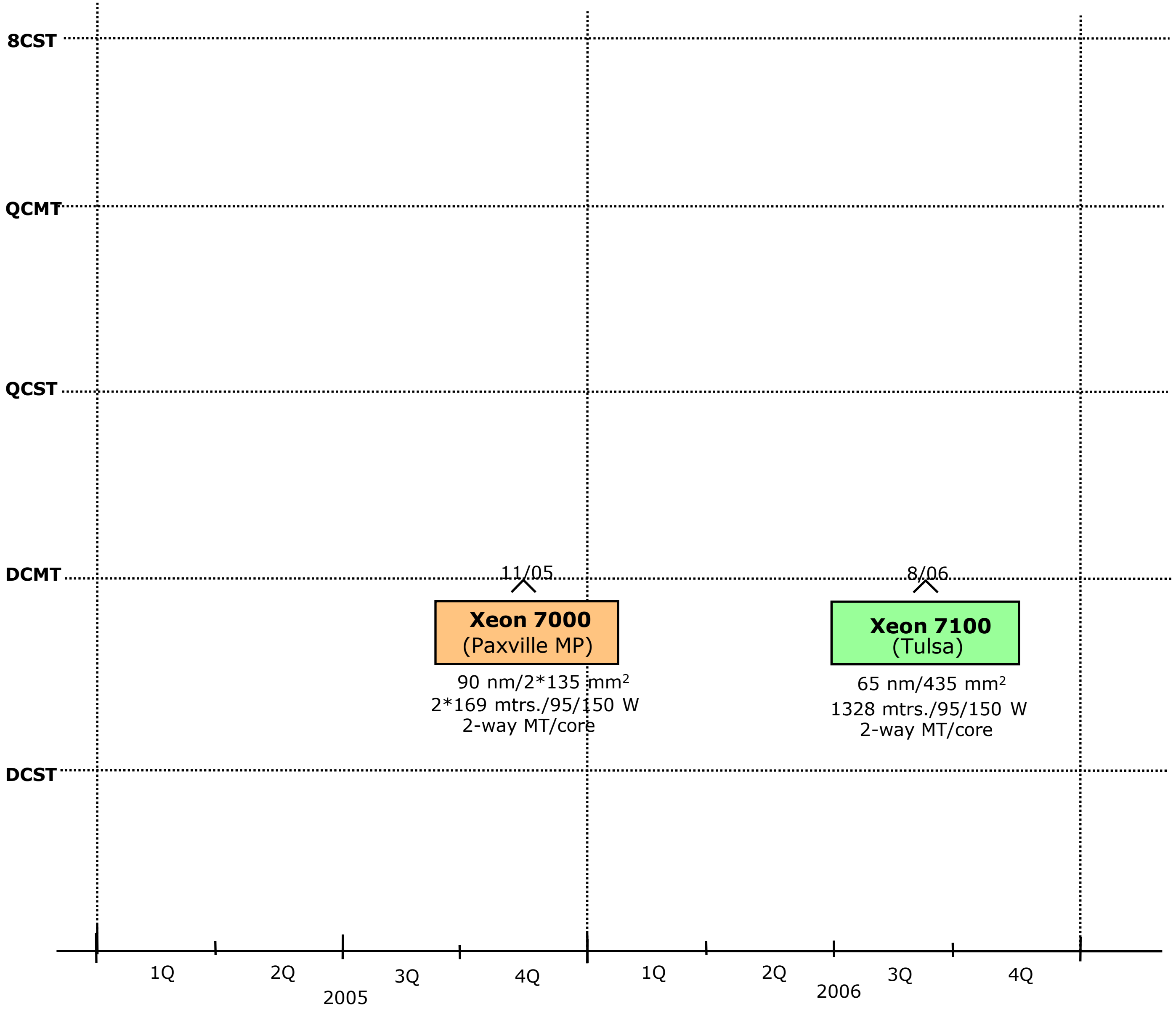


Figure 2.5.: Intel’s dual-core Xeon MP-lines

2. Overview of MCPs (6)

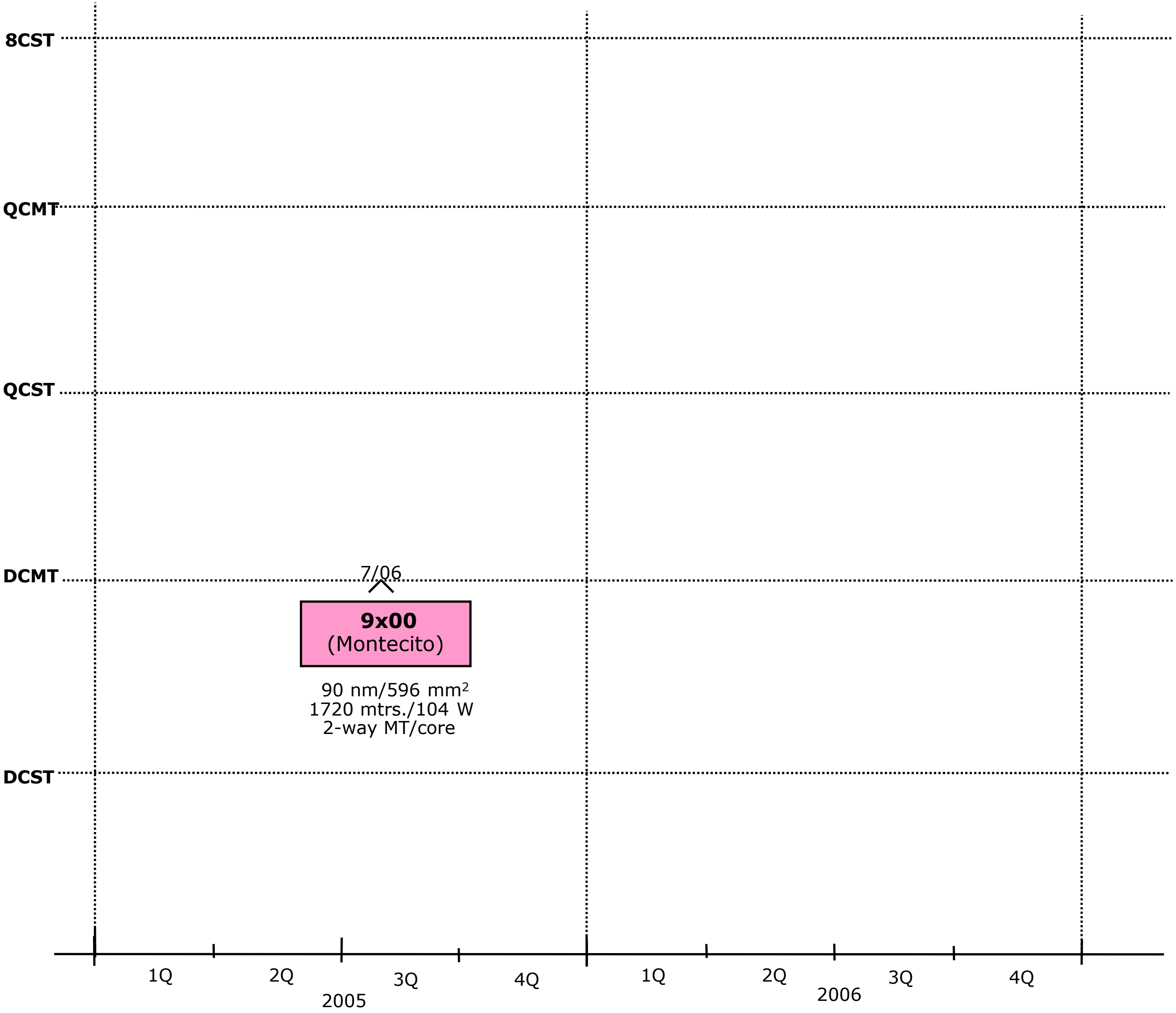


Figure 2.6.: Intel’s dual-core EPIC-based server line

2. Overview of MCPs (7)

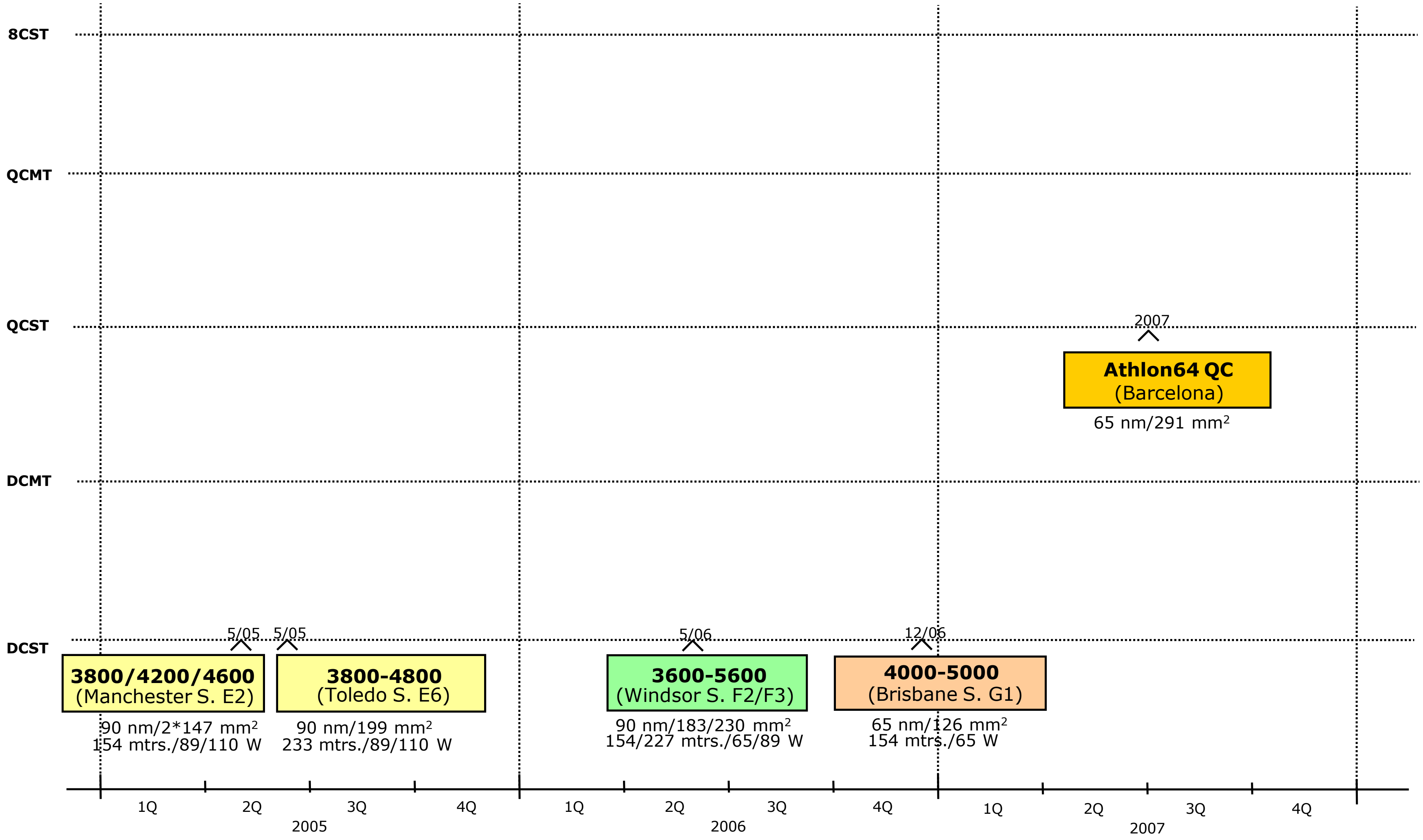


Figure 2.7.: AMD's multi-core desktop processor lines

2. Overview of MCPs (8)

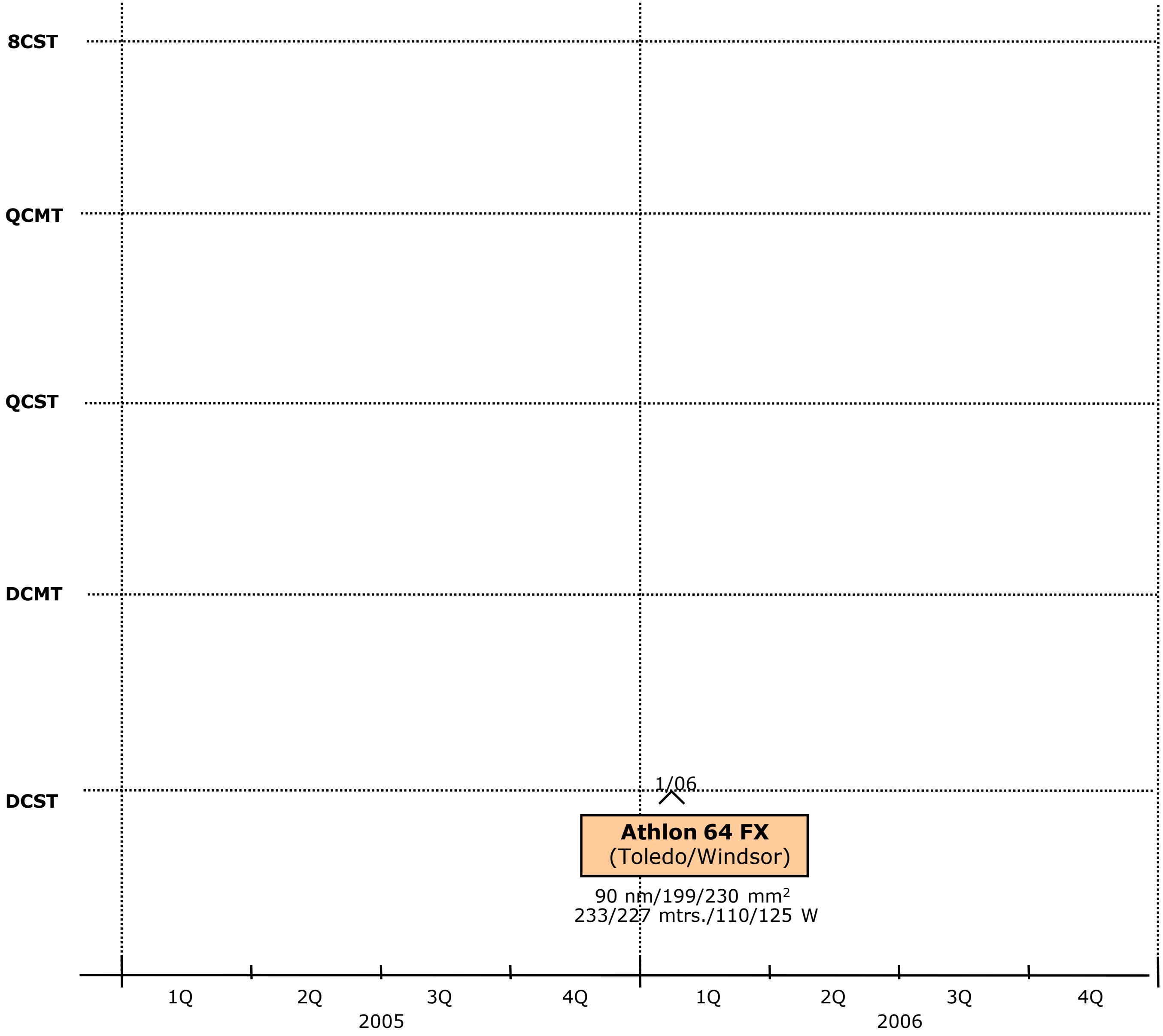


Figure 2.8.: AMD’s dual-core high-end desktop/entry level server Athlon 64 FX line

2. Overview of MCPs (9)

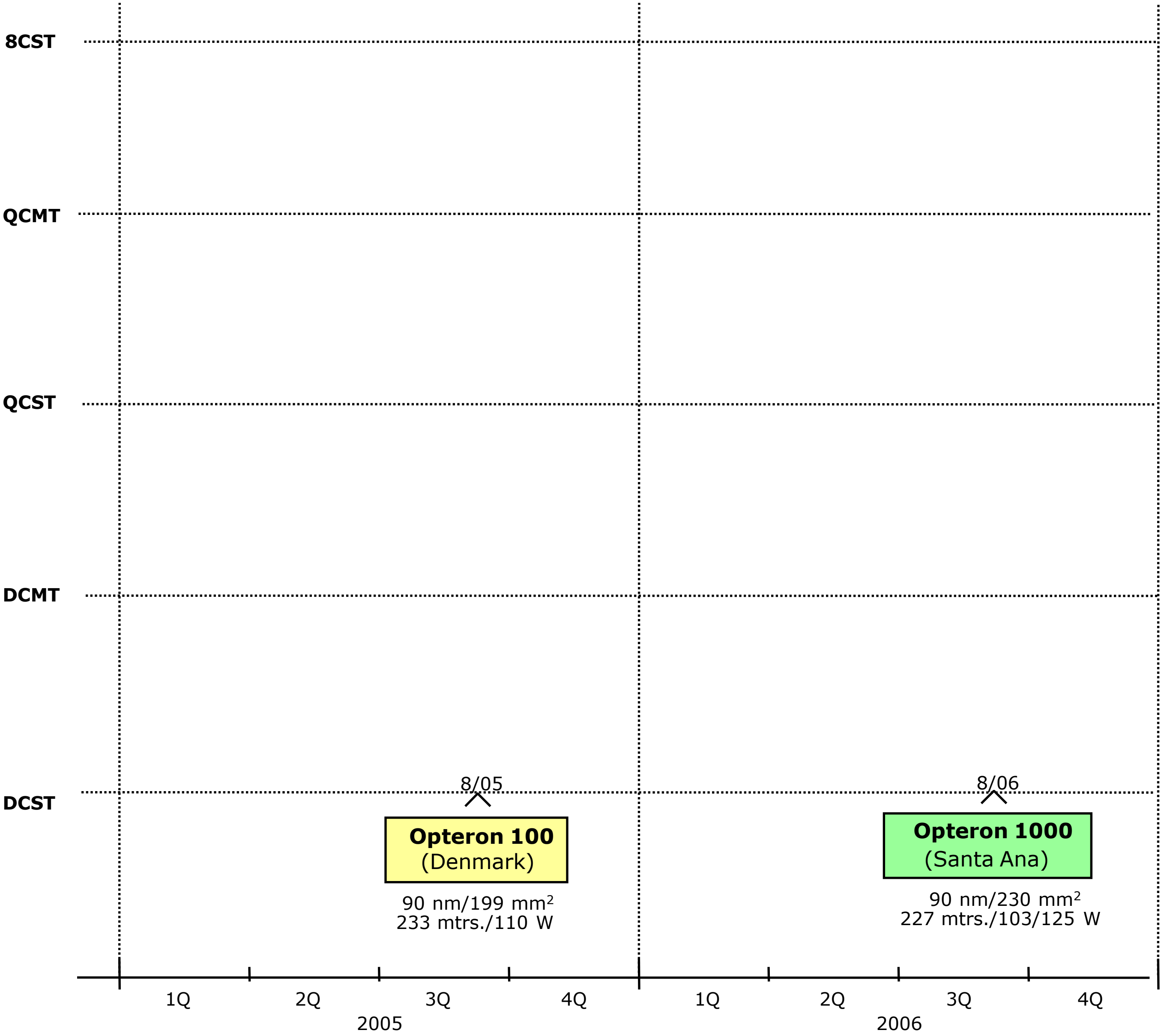


Figure 2.9.: AMD’s dual-core Opteron UP-lines

2. Overview of MCPs (10)

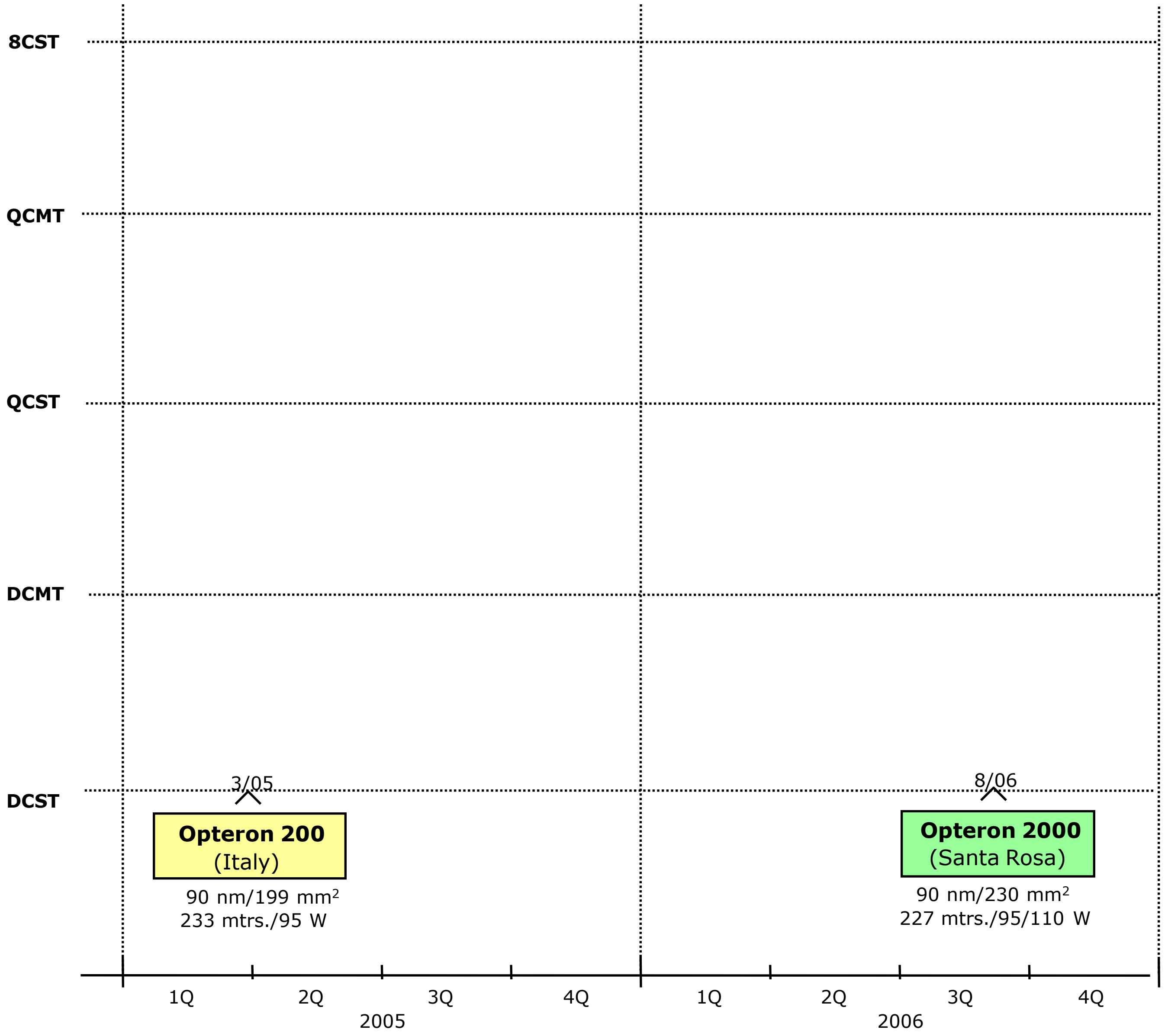


Figure 2.10.: AMD's dual-core Opteron DP-lines

2. Overview of MCPs (11)

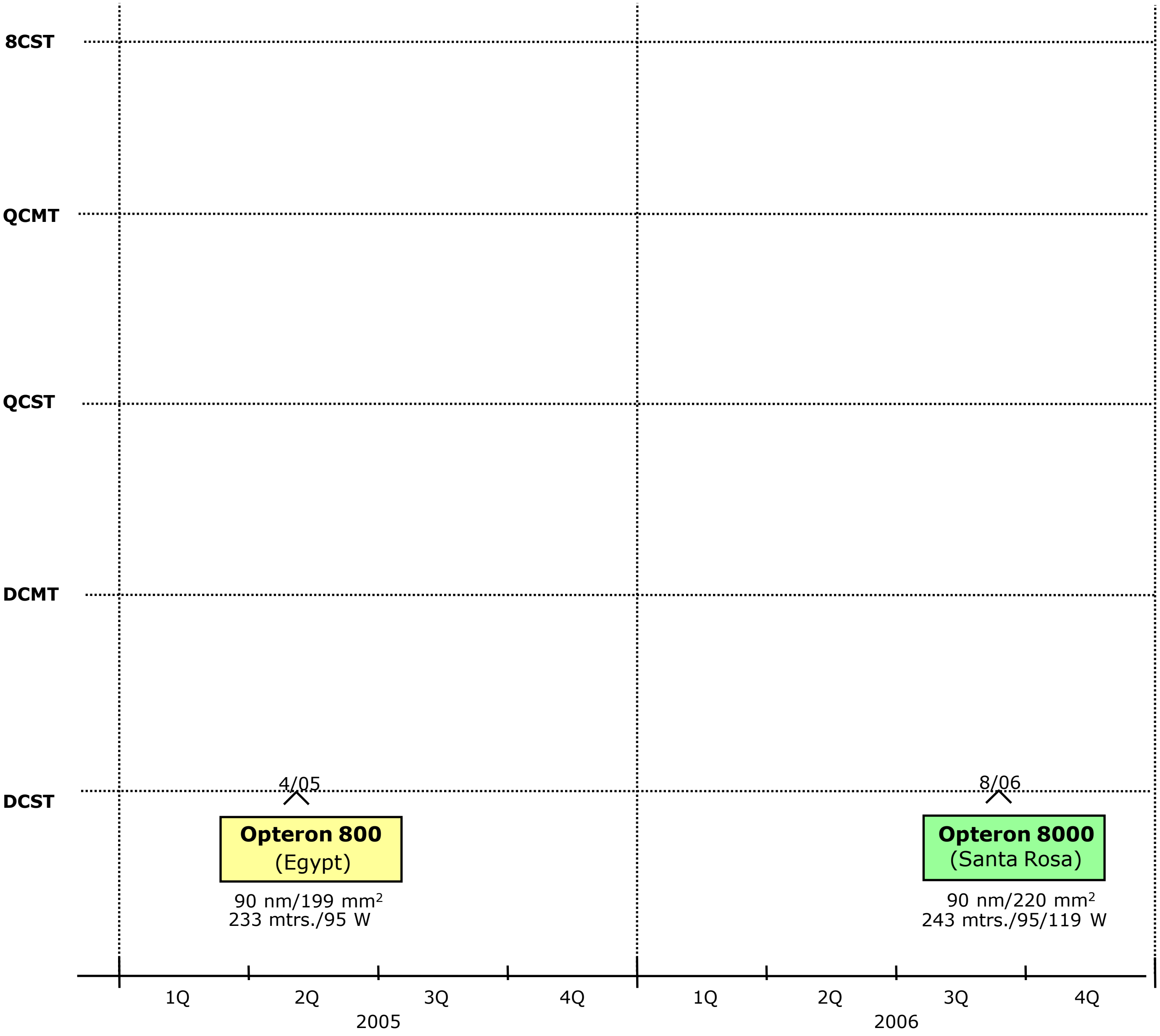


Figure 2.11.: AMD's dual-core Opteron MP-lines

2. Overview of MCPs (12)

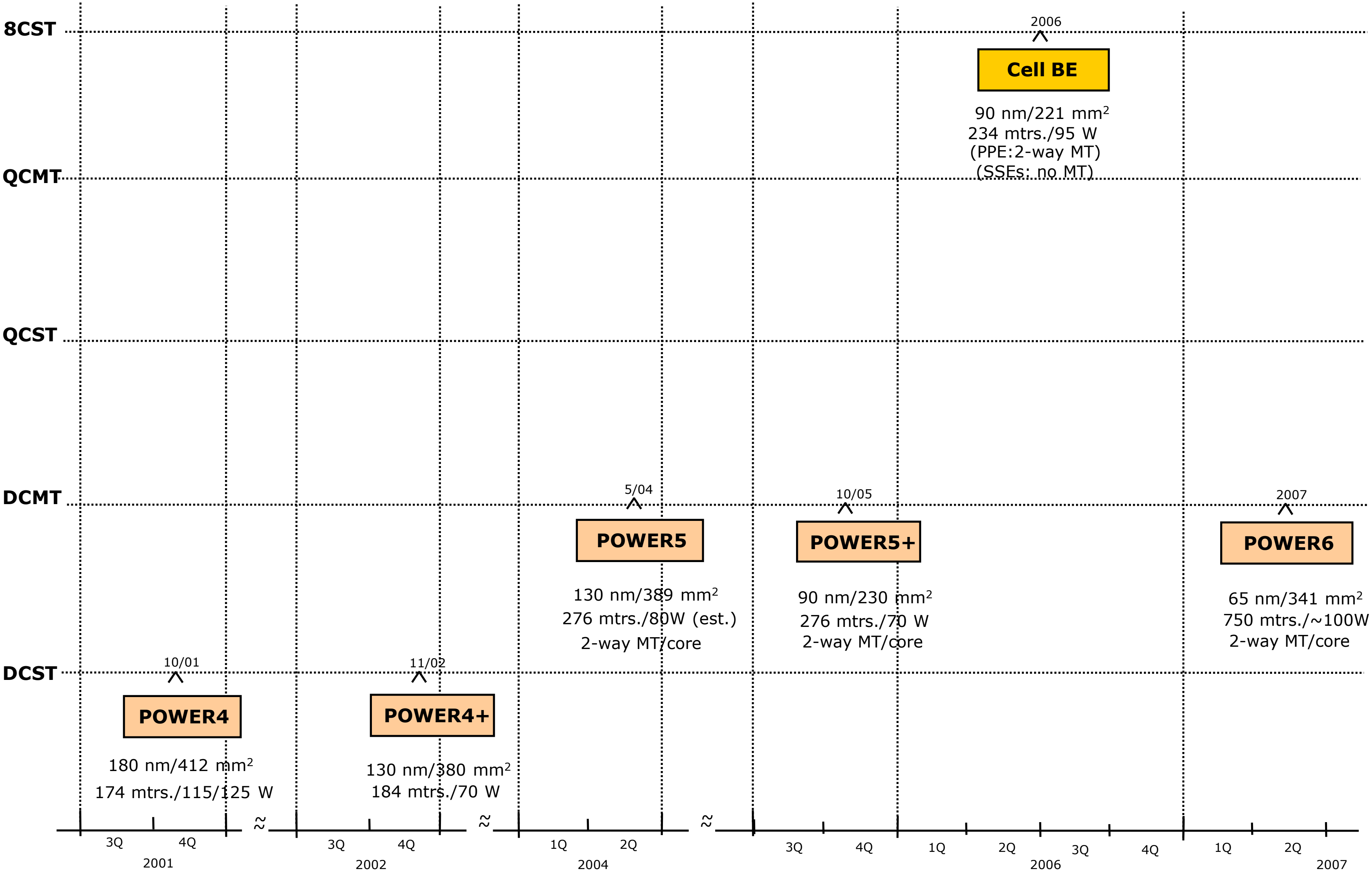


Figure 2.12.: IBM’s multi-core server lines

2. Overview of MCPs (13)

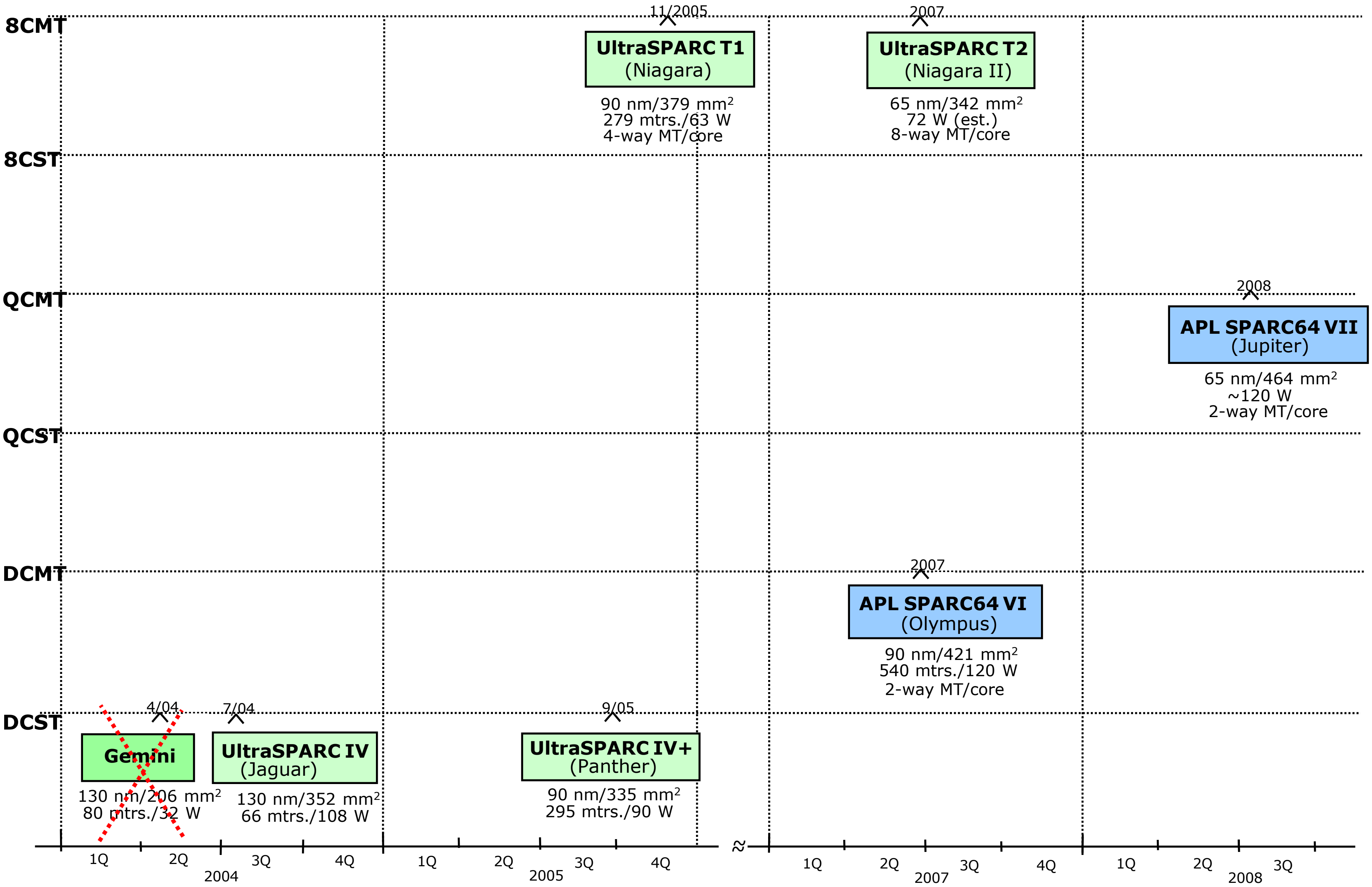


Figure 2.13.: Sun's and Fujitsu's multi-core server lines

2. Overview of MCPs (14)

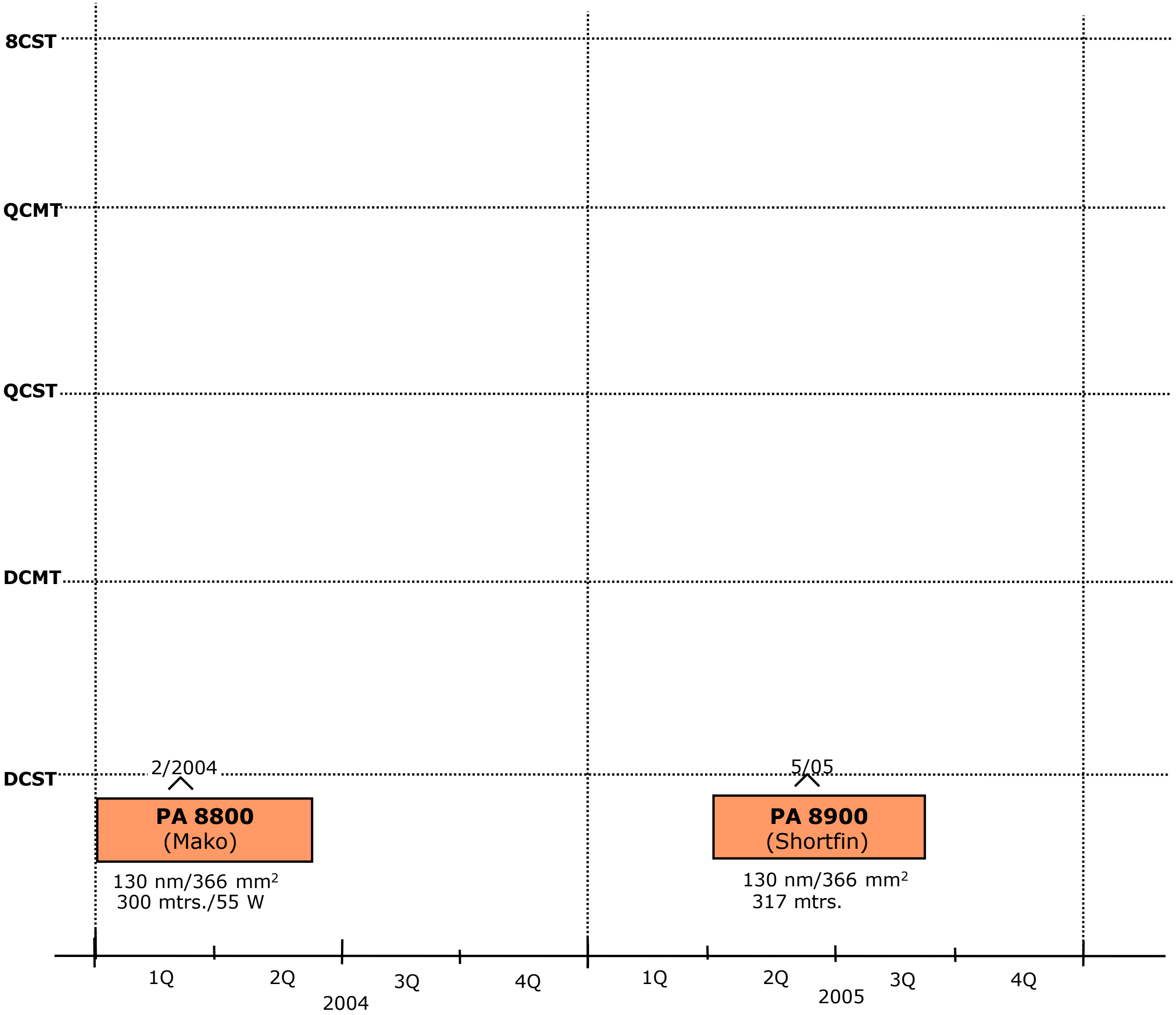


Figure 2.14.: HP's PA-8x00 server line

2. Overview of MCPs (15)

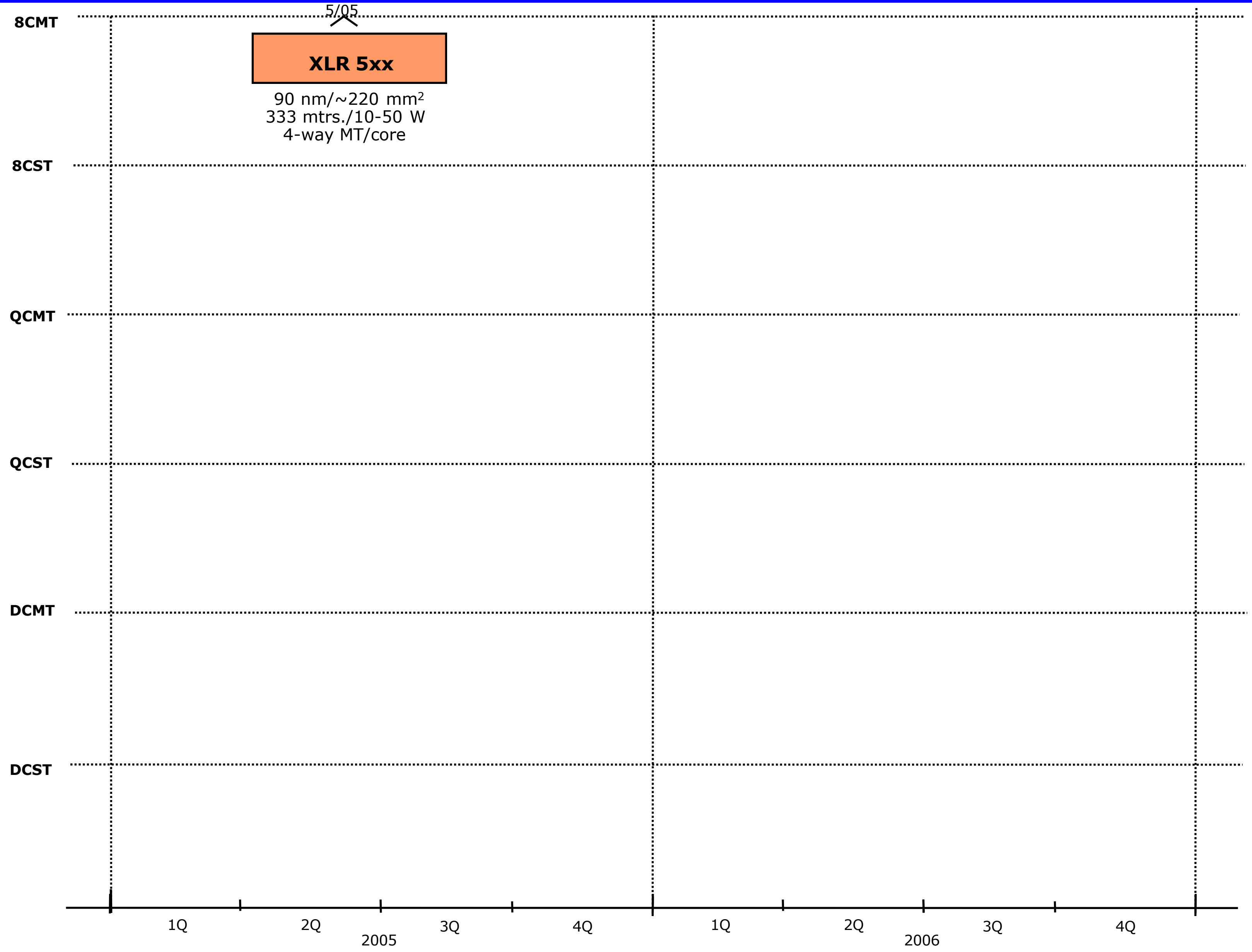


Figure 2.15.: RMI’s multi-core XLR-line (scalar RISC)

3. Design space of the macro architecture of MCPs

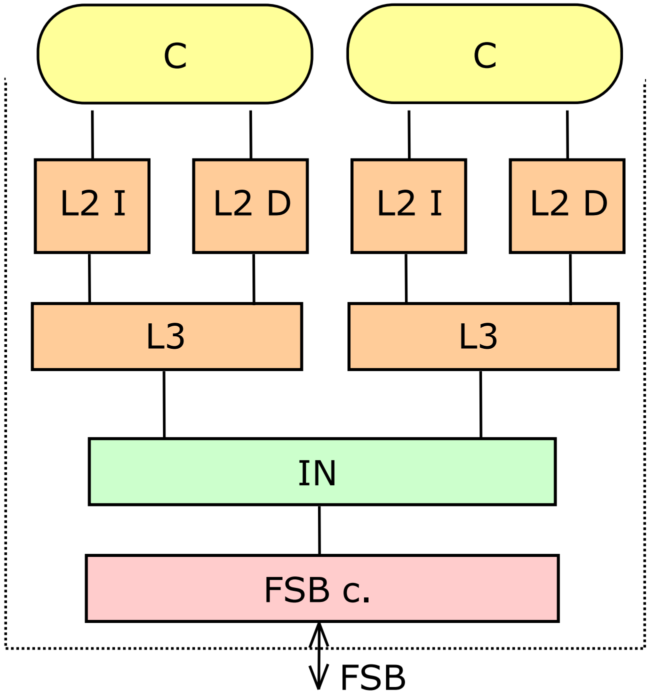
3. Design space of the macro architecture of MCPs (1)

Building blocks of multicore processors (MC)

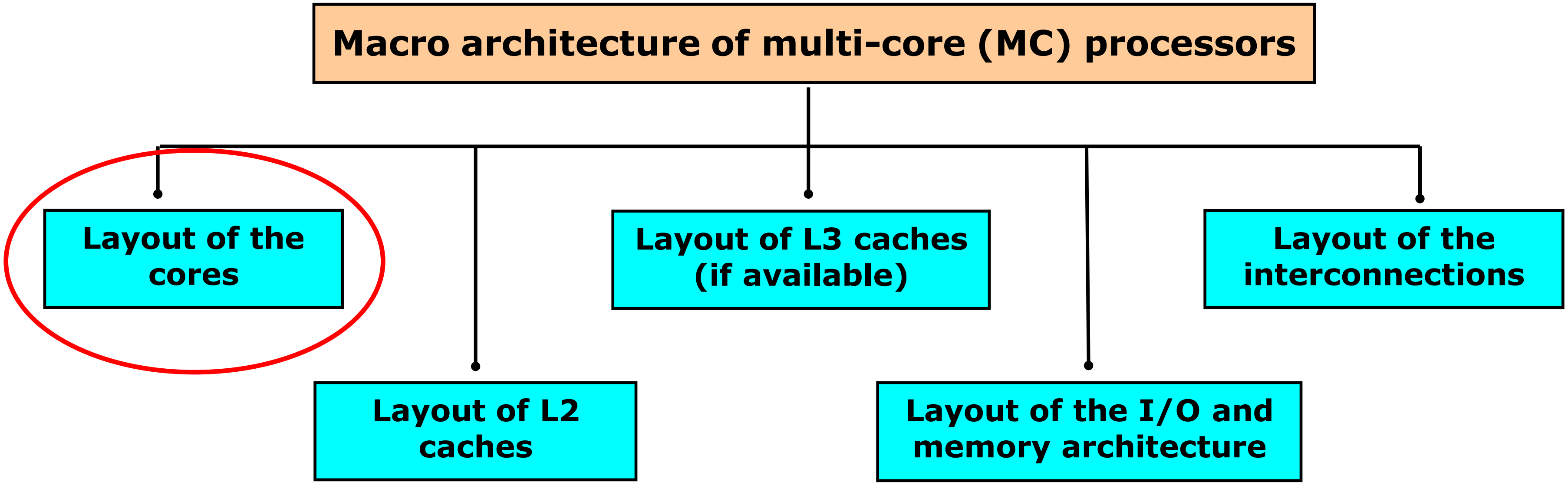
- Cores
- L2 cache(s) (L2)
- L3 cache(s), if available (L3)
- Interconnection network (IN)
- Bus controller (B. c.)

or → • FSB controller (FSB c.)

- Memory controller (M. c.)



3. Design space of the macro architecture of MCPs (2)



Macro architecture of multi-core (MC) processors

Layout of the cores

Layout of L2 caches

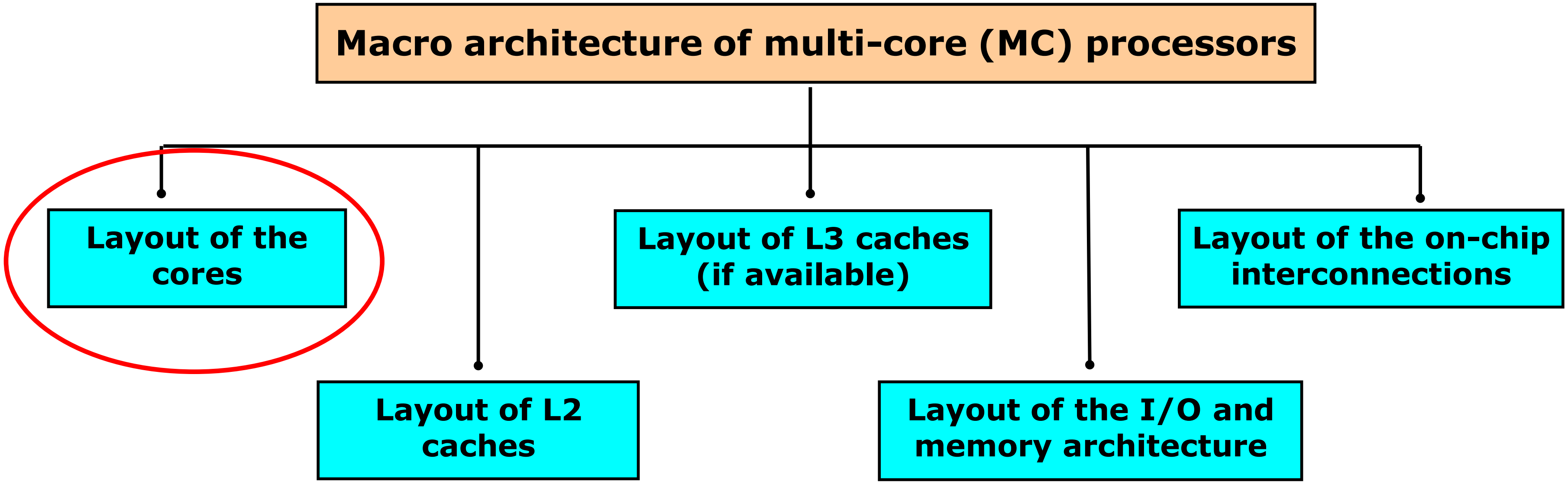
Layout of L3 caches (if available)

Layout of the I/O and memory architecture

Layout of the interconnections

4. Layout of the cores

5. Layout of the cores (1)



Macro architecture of multi-core (MC) processors

Layout of the cores

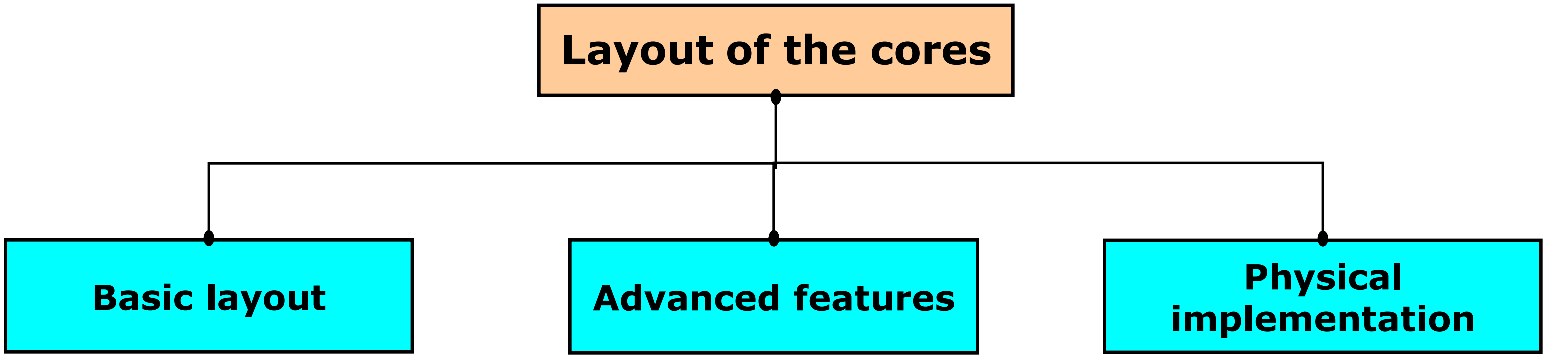
Layout of L2 caches

Layout of L3 caches (if available)

Layout of the I/O and memory architecture

Layout of the on-chip interconnections

4. Layout of the cores (2)



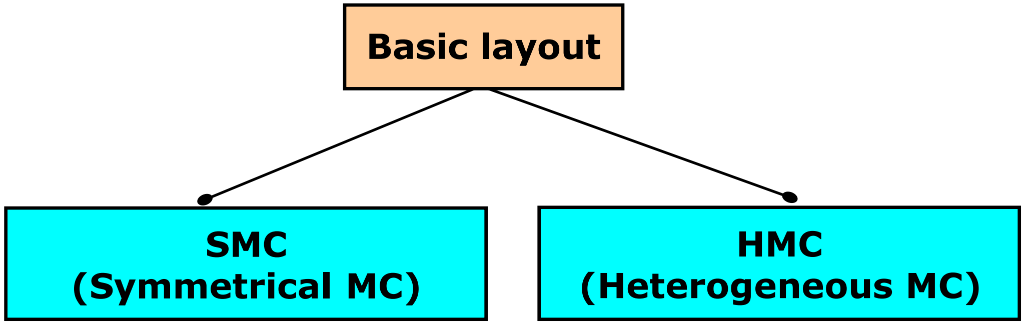
Layout of the cores

Basic layout

Advanced features

Physical implementation

4. Layout of the cores (3)



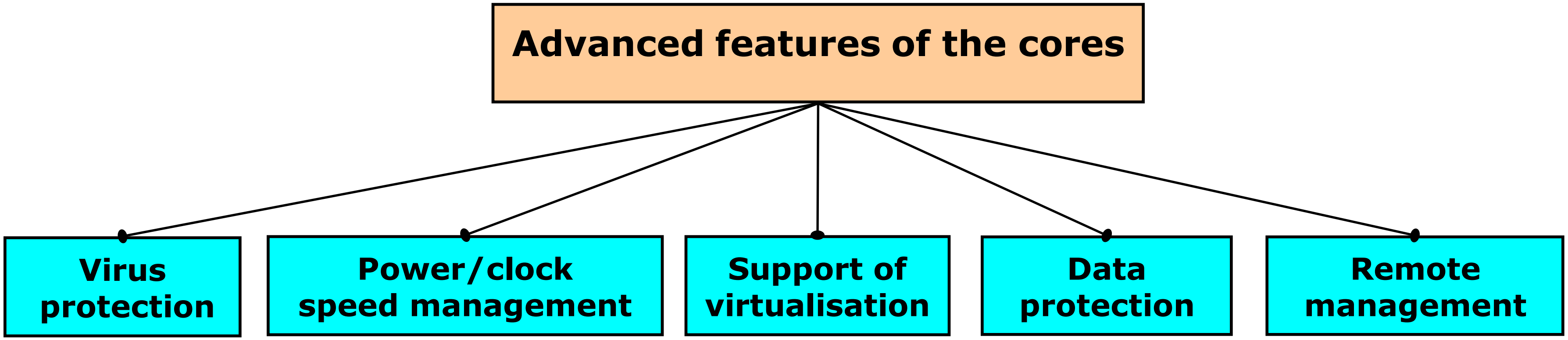
Basic layout

SMC
(Symmetrical MC)

HMC
(Heterogeneous MC)

All other MCs	*Cell BE (1 PPE+8 SPE)*

4. Layout of the cores (4)


Advanced features of the cores

Virus protection

Power/clock speed management

Support of virtualisation

Data protection

Remote management

Details on the next slide!

4. Layout of the cores (5)

	Virus protection	Power/clock speed management	Support of virtualisation	Data protection	Remote management
Intel					
Pentium M-based[1]	ED-bit	EIST	5/02	—	—
Netburst					
Prescott-based[2]	ED-bit	EIST (partly)	—	—	—
Cedar Mill-based[3]	ED-bit	EIST (partly)	P Vanderpool (partly)	—	—
Core-based[4]	ED-bit	EIST (partly)	Vanderpool (partly)	La Grande	AMT2
Itanium 2	—	DBS	Silvervale	—	—
AMD					
Athlon 64 XR	Nx-bit	Cool 'n' Quiet	Pacifica (partly)	—	—
Athlon 64 FX	Nx-bit	Cool 'n' Quiet	Pacifica	—	—
Opteron					
x00 lines	Nx-bit	PowerNow!	AMD - V	—	—
x000 lines	Nx-bit	PowerNow!	AMD - V	Presidio	—

[1] Pentium M-based line: Core Duo T2000

[2] Prescott based desktop lines:Pentium D 8xx, EE 840
Xeon lines: Paxwille DP, MP

[3] Cedar Mill based desktop lines: Pentium D 9xx, EE 955,965
Xeon lines: 5000 (Dempsey), 7100 (Tulsa)

[4] Core basd mobile line: T5xxx/T7xxx (Merom)
desktop line: Core 2 Duo E6xx, X6xxx, QX67xx
Xeon lines: 3000, 5100 (Woodcrest), 5300 (Clovertown)

Table 4.1 : Comparison of advanced features provided by Intel’s and AMD’s MC lines.
For a brief introduction of the related Intel techniques see the appropriate Intel processor descriptions

4. Layout of the cores (6)

Brief description of the advanced features pointed out while using Intel’s notations (1)

ED: Execute Disable bit (often designated as the NX bit (No Execute bit))

The ED bit with OS support is now a widely used technique to prevent certain classes of malicious buffer overflow attacks. It is used e.g. in Intel’s Itanium, Pentium M, Pentium 4 Prescott and subsequent processors as well as in AMD’s x86-64 line (designated as the NX bit (No Execute bit).

In buffer overflow attacks a malicious worm inserts its code into another program’s data space and creates a flood of code that overwhelms the processor, allowing the worm to propagate itself to the network, and to other computers.

The Execute Disable bit allows the processor to classify areas in memory by where application code can be executed and where it cannot. Actually, the ED bit is implemented as the 63. bit of the Page Table Entry. If its value is 1, code cannot be executed from that page.

When a malicious worm attempts to insert code in the buffer,
the processor disables code execution, preventing damage and worm propagation.
To became active the ED feature need to be supported by the OS.
For more information see: http://en.wikipedia.org/wiki/XD_Bit

EIST: Enhanced Intel SpeedStep Technology

First implemented in Intel’s mobile and server platforms,

It allows the system to dynamically adjust processor voltage and core frequency, to decrease average power consumption and thus average heat production.

This technology is similar to AMD’s Cool’n’Quiet and PowerNow! techniques.
For more information see http://en.wikipedia.org/wiki/SpeedStep

4. Layout of the cores (7)

Brief description of the advanced features pointed out while using Intel’s notations (2)

VT: Virtualization Technology

Virtualisation techniques allow a platform to run multiple operating systems and applications in independent partitions. The hardver support improves the performance and robustness of traditional software-based virtualization solutions.
Intel introduced hardware virtualisation support first in its Itanium line (called the Silvervale tecnique) followed by its Presler based and subsequent processors (designated as the Vanderpool technique).

AMD uses this technique in its recent lines - dubbed as Pacifica - since Mai 2006. For more information see http://en.wikipedia.org/wiki/Virtualization_Technology

La Grande Technology

It is a Trusted Execution Technology to protect users from software-based attacks aimed at stealing vital data, initiated by Intel. It includes a set of hardware enhancements intended to provide multiple separated execution environments.

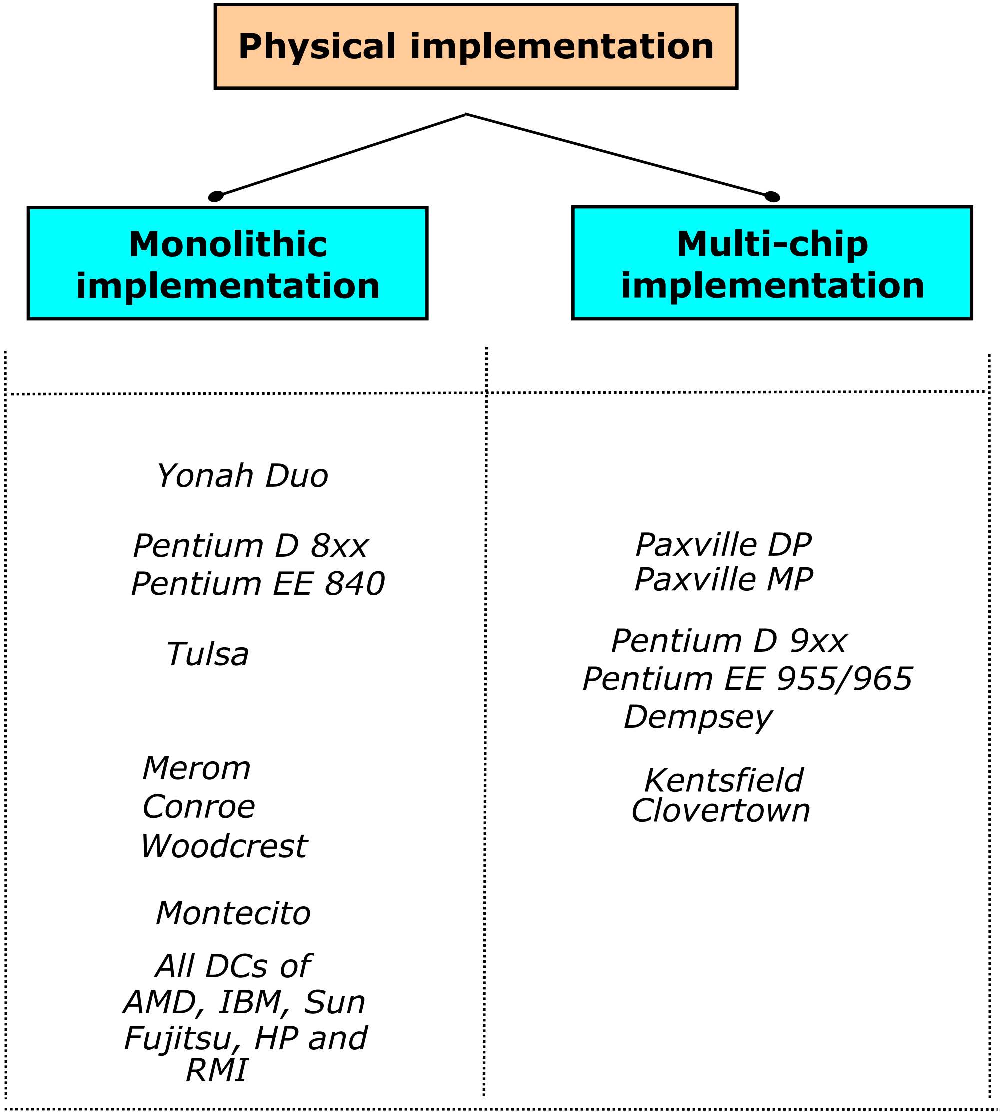
For more information see Intel® Trusted Execution Technology Preliminary Architecture Specification at http://www.intel.com/technology/security/ or http://en.wikipedia.org/wiki/LaGrande_Technology

AMT2

Intel’s Active Management Technology is a feature of its vPro technology. It allows to manage the computer system remotely, even if the computer is switched off or the hard drive has failed. By means of this technology system administrators can e.g. remotely download software updates or fix and hail system problems. It can also be utilized to protect the network by proactively blocking incoming threats.

For more information see Intel Active Management Technology at http://www3.intel.com/cd/network/communications/emea/eng/203372.htm

4. Layout of the cores (8)



Physical implementation

Monolithic implementation

Yonah Duo

Pentium D 8xx
Pentium EE 840

Tulsa

Merom
Conroe
Woodcrest

Montecito

All DCs of AMD, IBM, Sun Fujitsu, HP and RMI

Multi-chip implementation

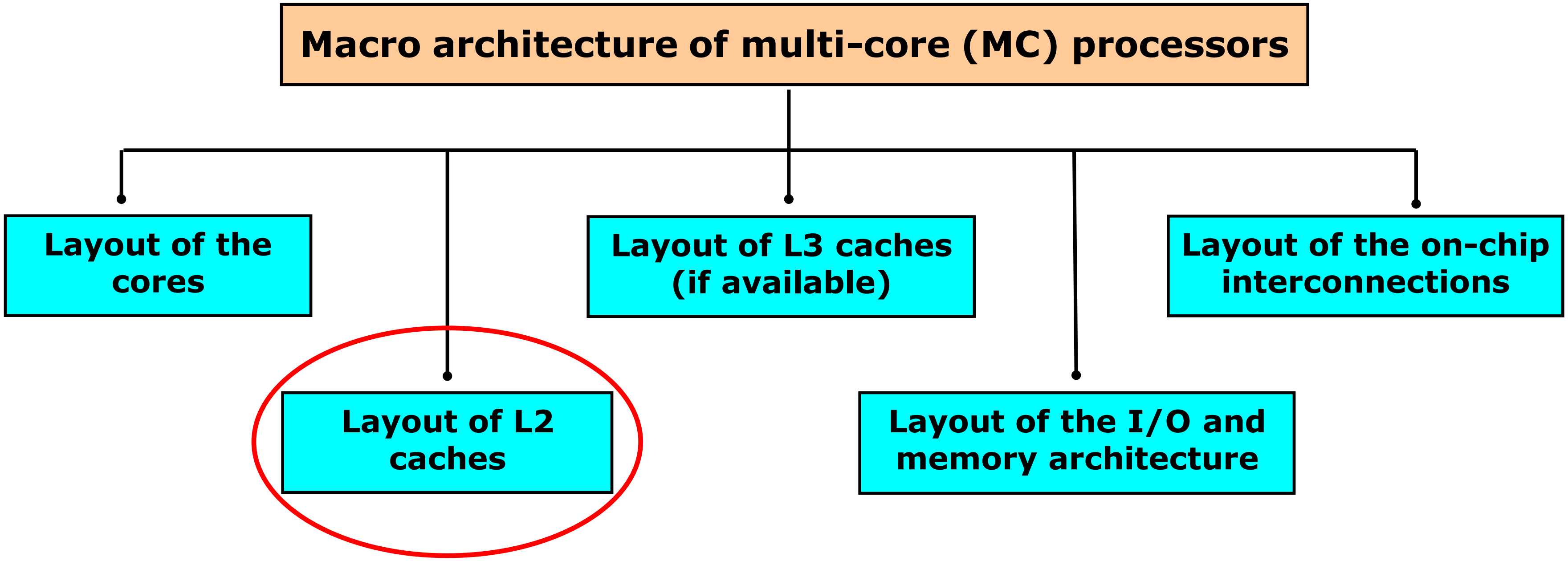
Paxville DP
Paxville MP

Pentium D 9xx
Pentium EE 955/965
Dempsey

Kentsfield
Clovertown

5. Layout of L2 caches

5. Layout of L2 caches (1)


Macro architecture of multi-core (MC) processors

Layout of the cores

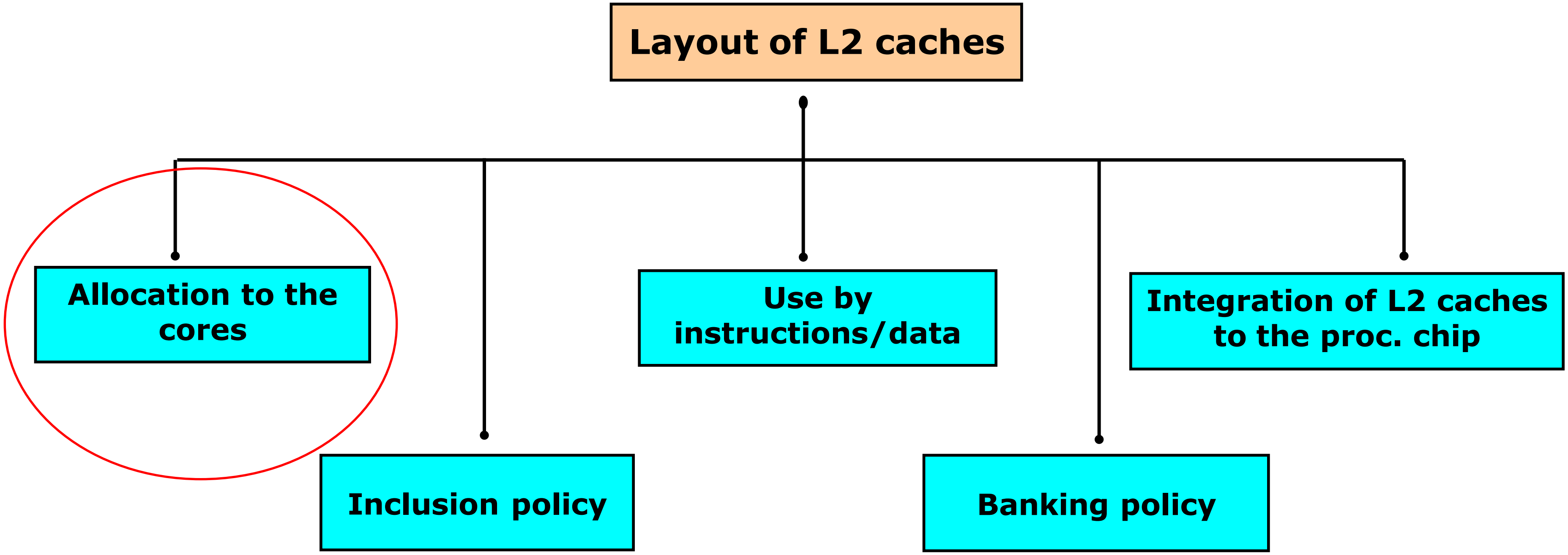
Layout of L2 caches

Layout of L3 caches (if available)

Layout of the I/O and memory architecture

Layout of the on-chip interconnections

5. Layout of L2 caches (2)



Layout of L2 caches

Allocation to the cores

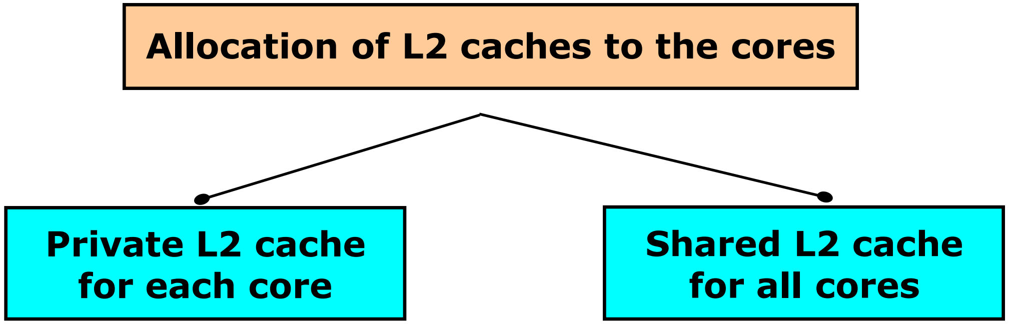
Inclusion policy

Use by instructions/data

Banking policy

Integration of L2 caches to the proc. chip

5. Layout of L2 caches (3)

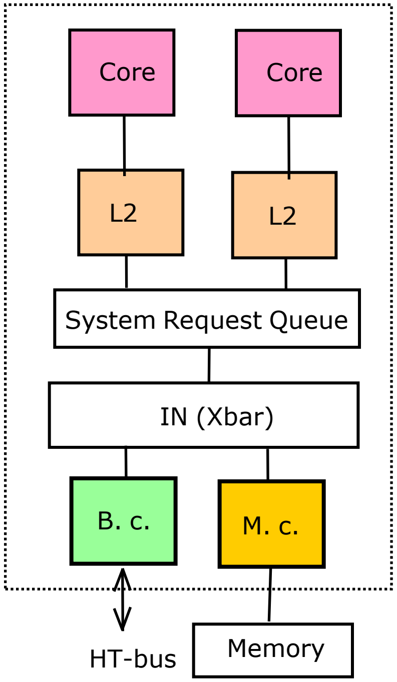


Allocation of L2 caches to the cores

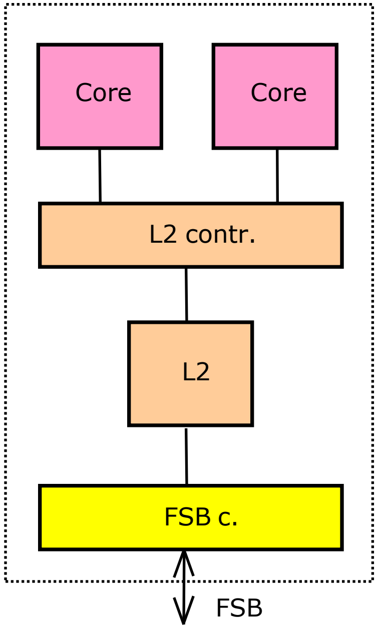
Private L2 cache for each core

Shared L2 cache for all cores

Examples:



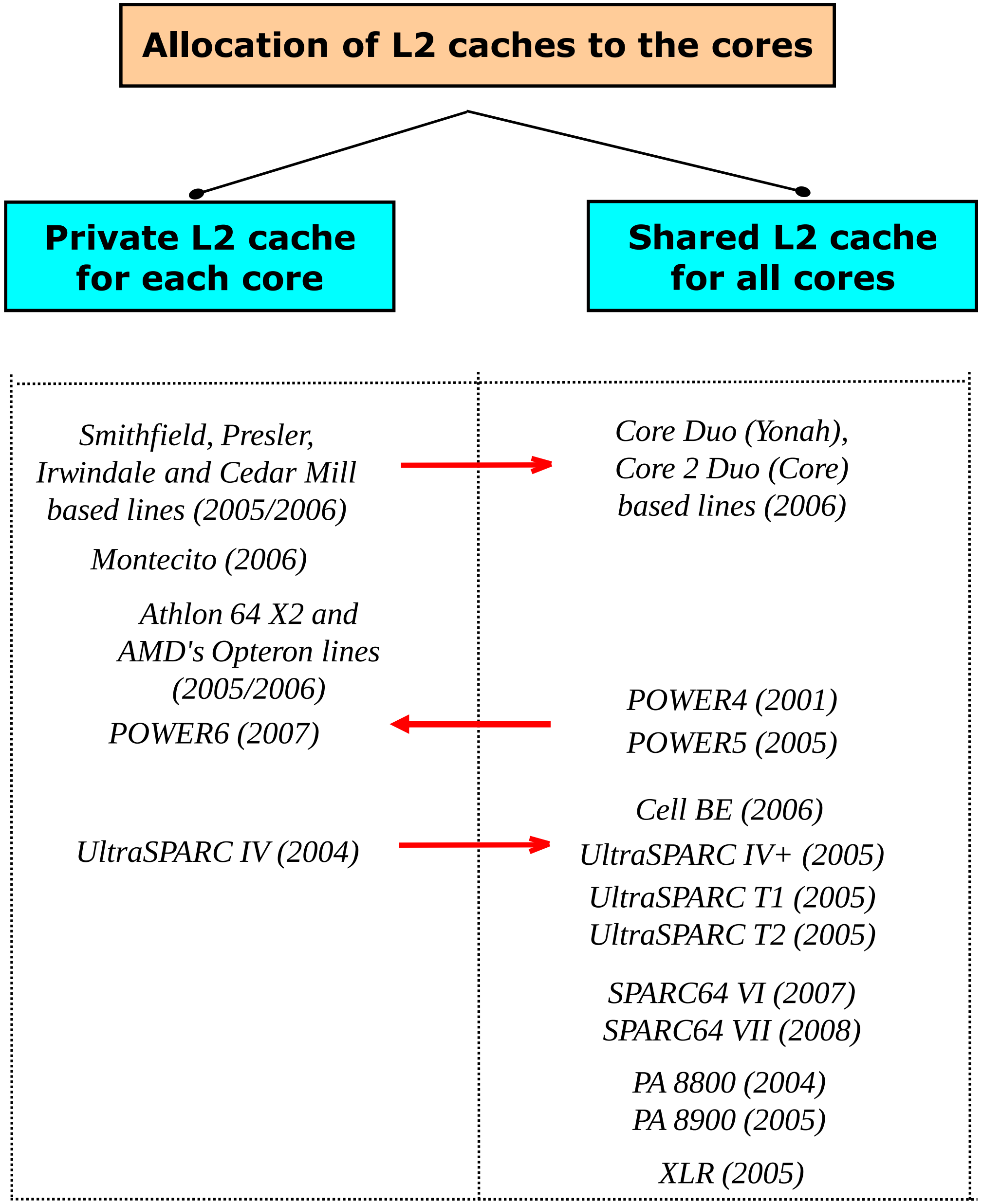
Athlon 64 X2 (2005)



Core Duo (2006)

Core 2 Duo (2006)

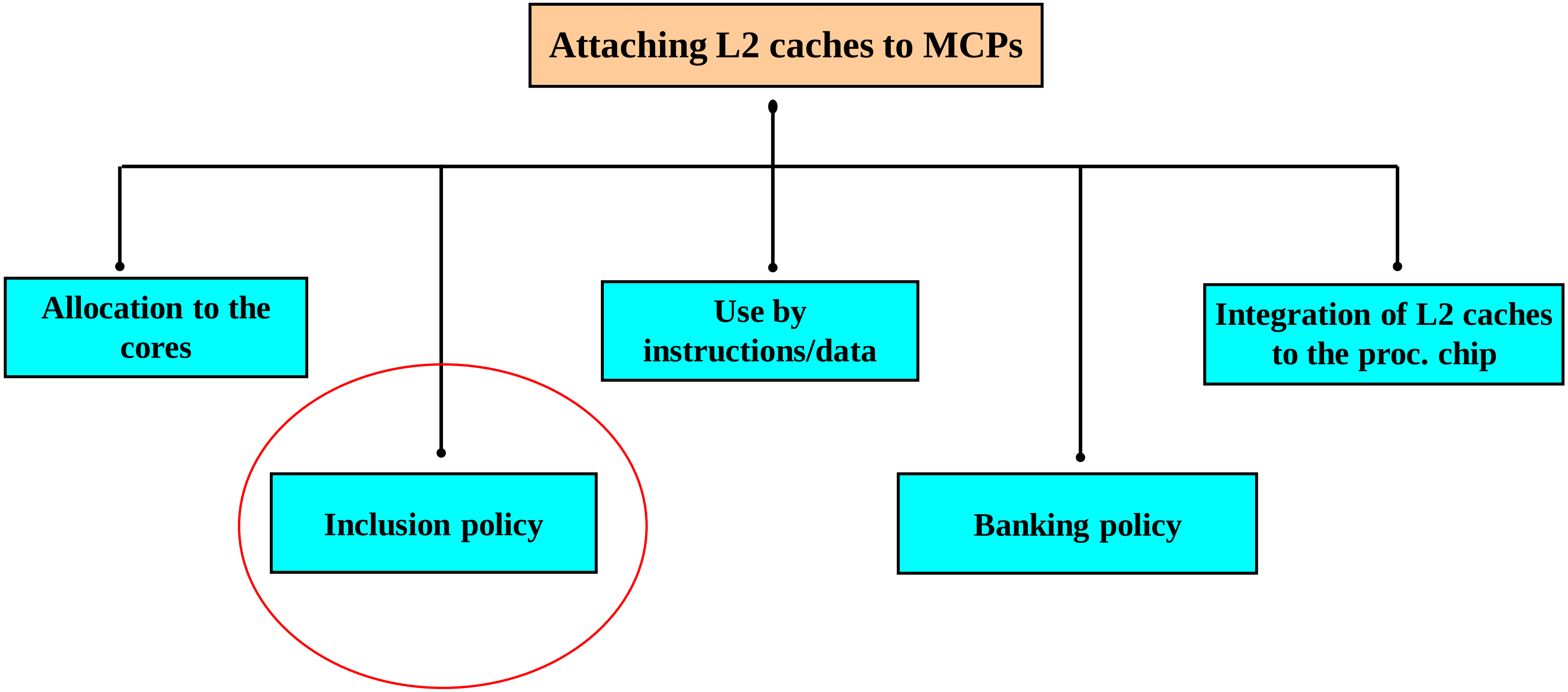
5. Layout of L2 caches (4)



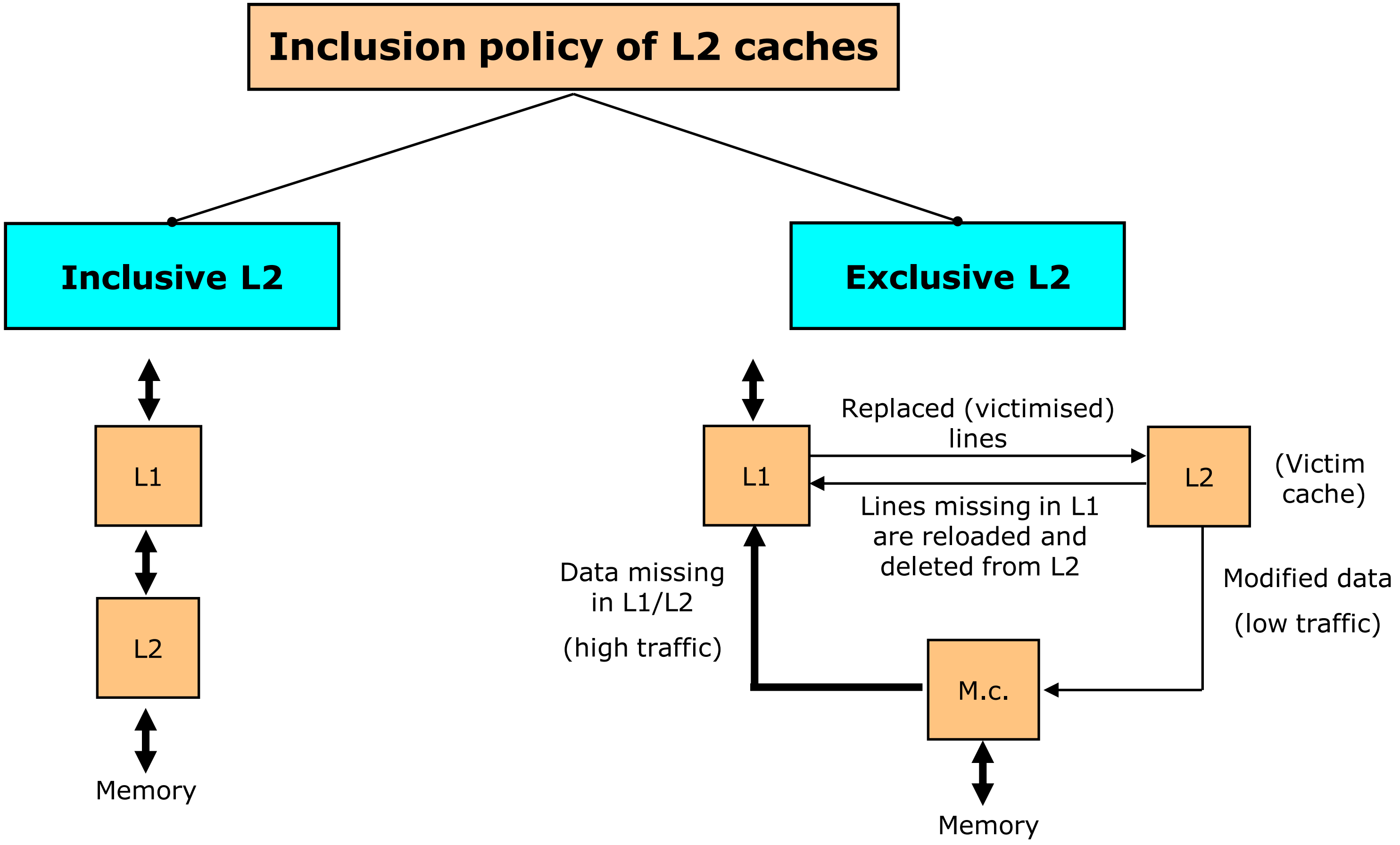
Allocation of L2 caches to the cores

Private L2 cache for each core	**Shared L2 cache for all cores**
Smithfield, Presler, Irwindale and Cedar Mill based lines (2005/2006)	*Core Duo (Yonah), Core 2 Duo (Core) based lines (2006)*
Montecito (2006)	
Athlon 64 X2 and AMD's Opteron lines (2005/2006)	
POWER6 (2007)	*POWER4 (2001)* *POWER5 (2005)*
	Cell BE (2006)
UltraSPARC IV (2004)	*UltraSPARC IV+ (2005)* *UltraSPARC T1 (2005)* *UltraSPARC T2 (2005)*
	SPARC64 VI (2007) *SPARC64 VII (2008)*
	PA 8800 (2004) *PA 8900 (2005)*
	XLR (2005)

5. Layout of L2 caches (5)



5. Layout of L2 caches (6)



- Lines replaced (victimized) in the L1 are written to L2.
- References to data missing in L1 but available in L2 initiate reloading the pertaining cache line to L1. Reloaded data is deleted from L2
- L2 operates usually as a write back cache (only modified data that is replaced in the L2 is written back to the memory),
- Unmodified data that is replaced in L2 is deleted.

5. Layout of L2 caches (7)

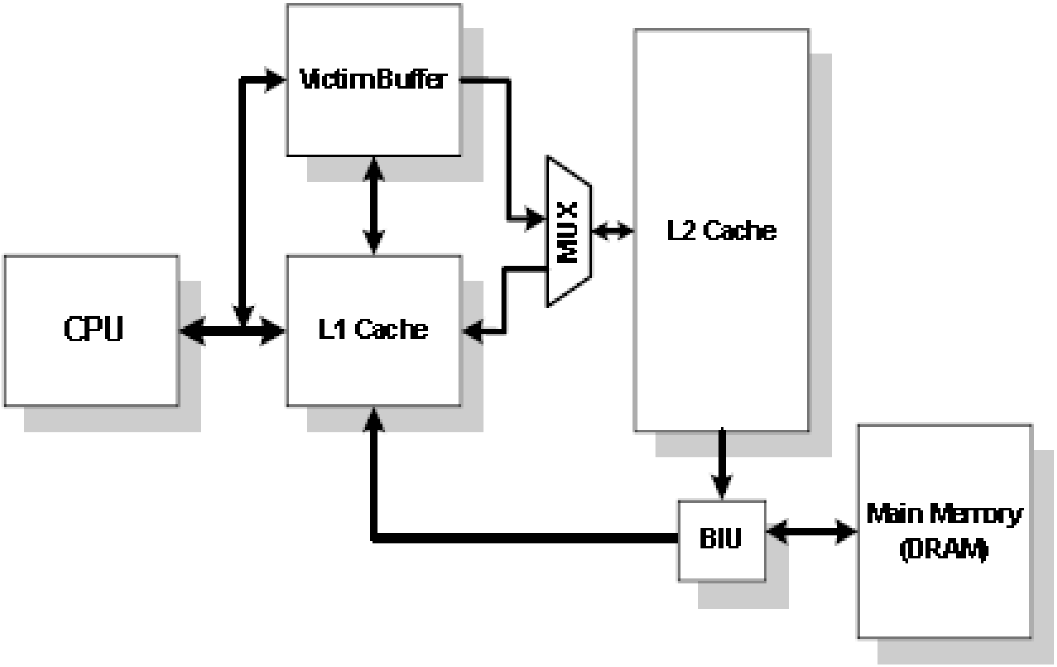
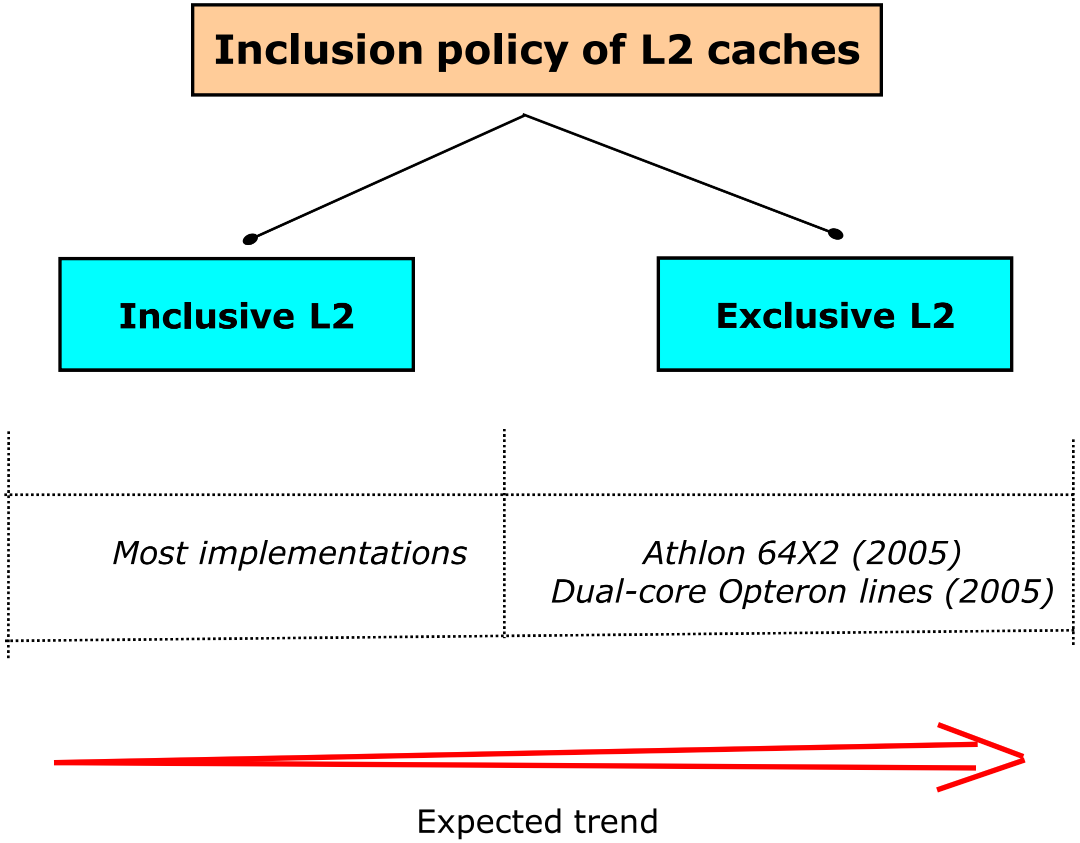


Figure 5.1: Implementation of exclusive L2 caches

Source: Zheng, Y., Davis, B.T., Jordan, M.: “ Performance evaluation of exclusive cache hierarchies”, 2004 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS), 2004, pp. 89-96.

5. Layout of L2 caches (8)

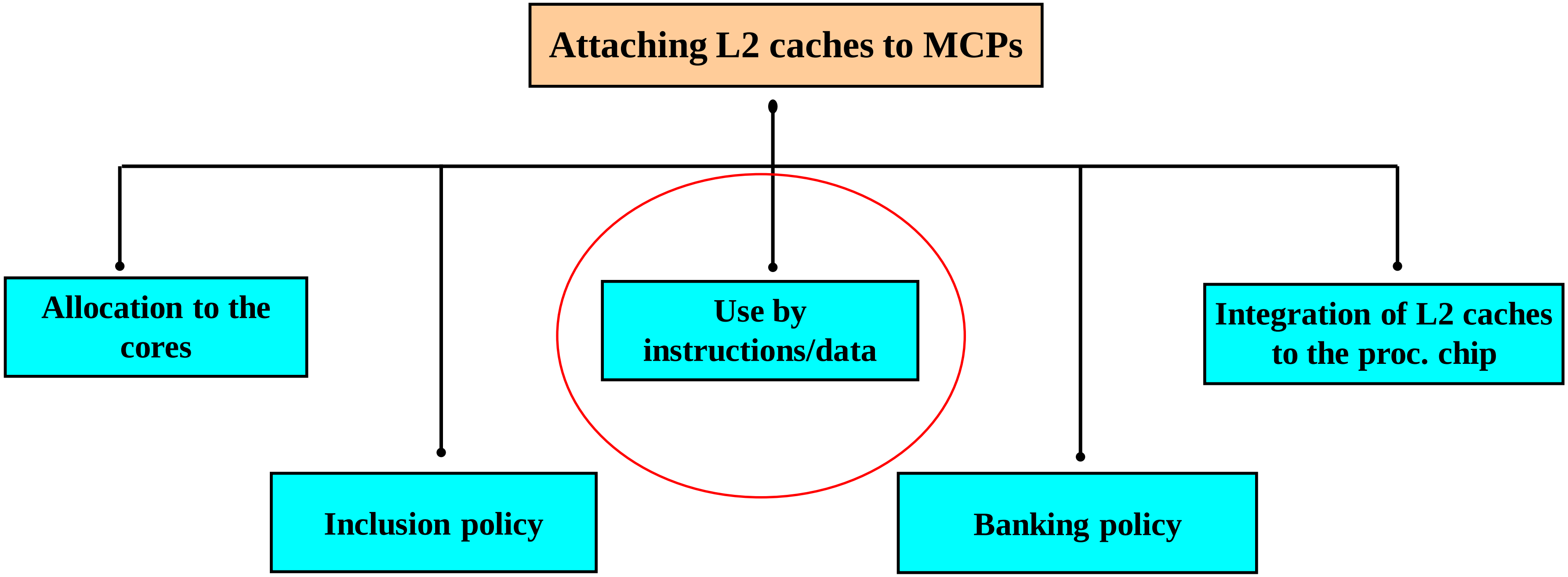


Inclusion policy of L2 caches

Inclusive L2	Exclusive L2
Most implementations	*Athlon 64X2 (2005)* *Dual-core Opteron lines (2005)*

Expected trend

5. Layout of L2 caches (9)


Attaching L2 caches to MCPs

Allocation to the cores

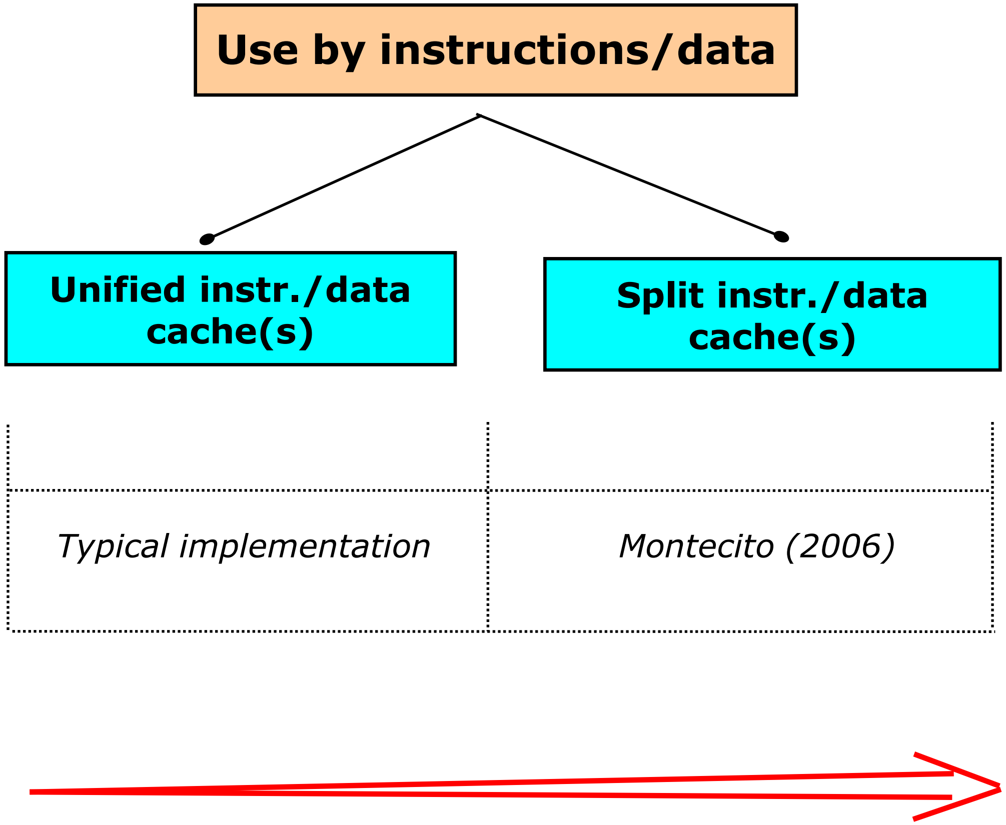
Inclusion policy

Use by instructions/data

Banking policy

Integration of L2 caches to the proc. chip

5. Layout of L2 caches (10)

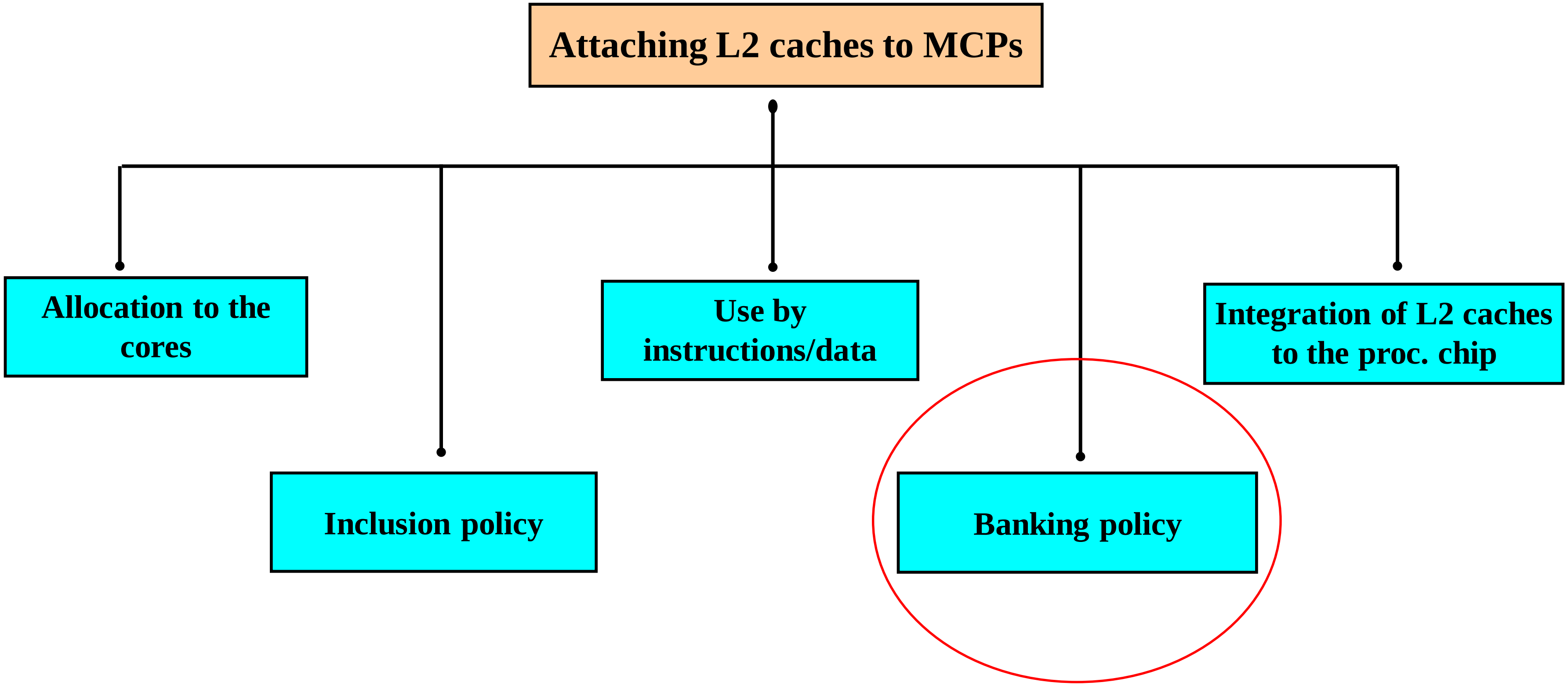


Use by instructions/data

Unified instr./data cache(s)	**Split instr./data cache(s)**
Typical implementation	*Montecito (2006)*

Expected trend

5. Layout of L2 caches (11)


Attaching L2 caches to MCPs

Allocation to the cores

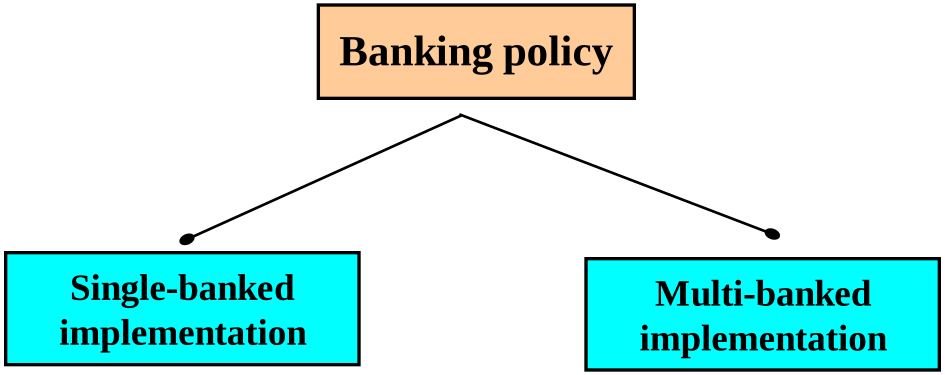
Inclusion policy

Use by instructions/data

Banking policy

Integration of L2 caches to the proc. chip

5. Layout of L2 caches (12)

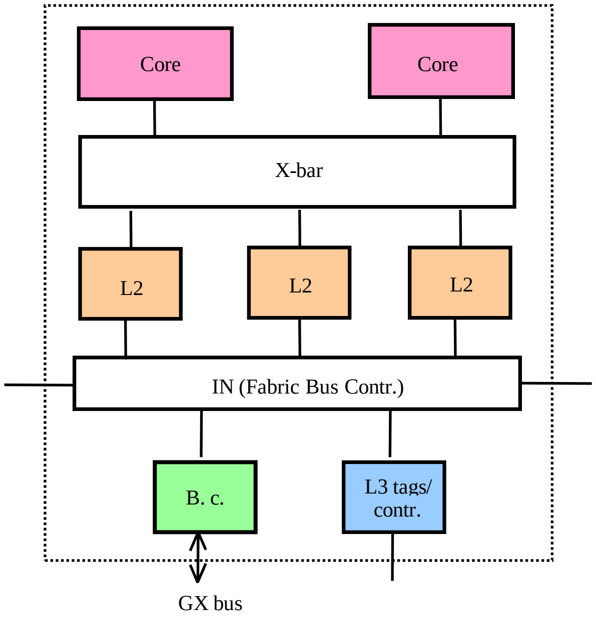


Typical implementation

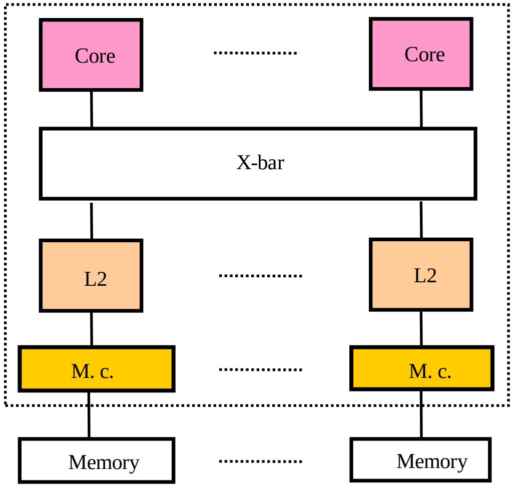
5. Layout of L2 caches (13)

Mapping of addresses to memory modules

POWER4 (2001)
POWER5 (2005)

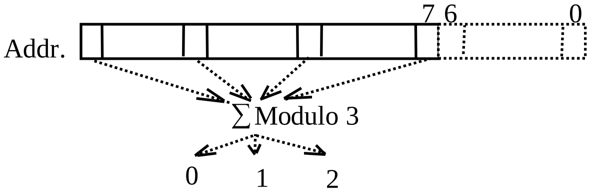


UltraSPARC T1 (2005) *(Niagara)*
(8 cores/4xL2 banks)



Mapping of addresses to the banks:

The 128-byte long L2 cache lines are hashed across the 3 modules. Hashing is performed by modulo 3 arithmetic applied on a large number of real address bits.



Mapping of addresses to the banks:

The four L2 modules are interleaved at 64-byte blocks.

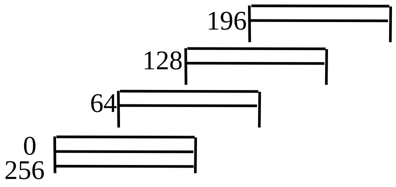
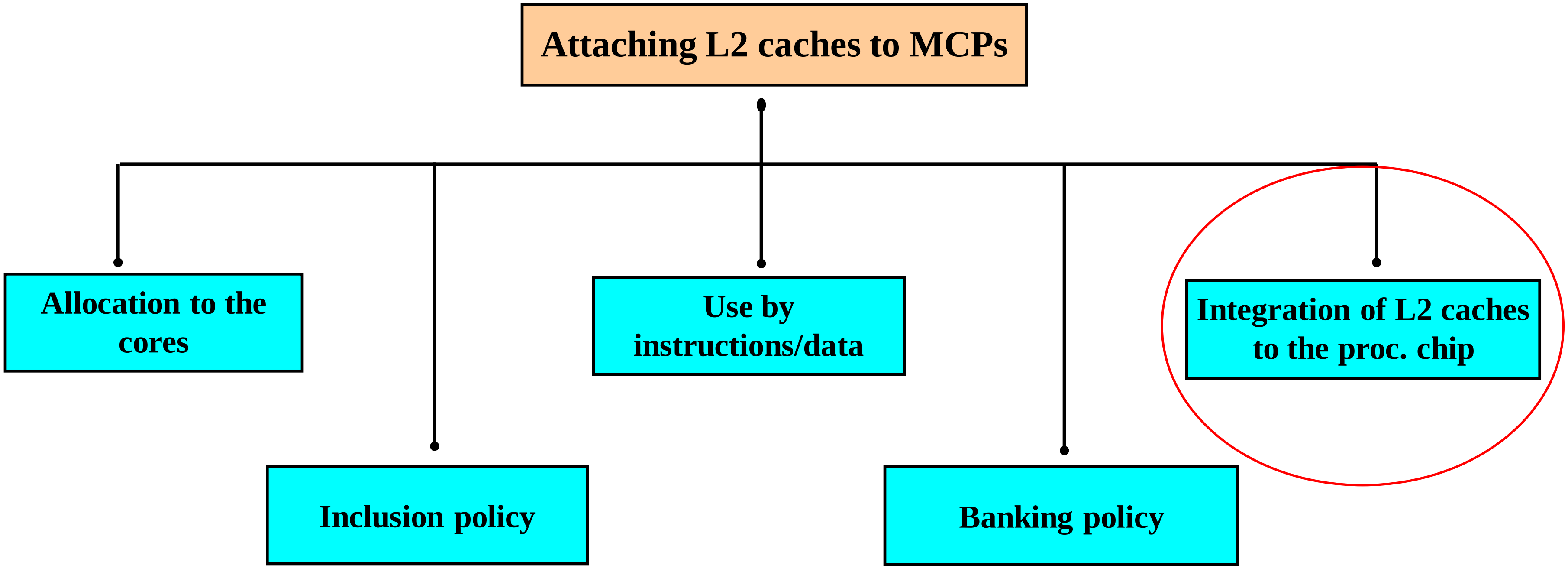


Figure 5.2: Alternatives of mapping addresses to memory modules

5. Layout of L2 caches (14)



Attaching L2 caches to MCPs

Allocation to the cores

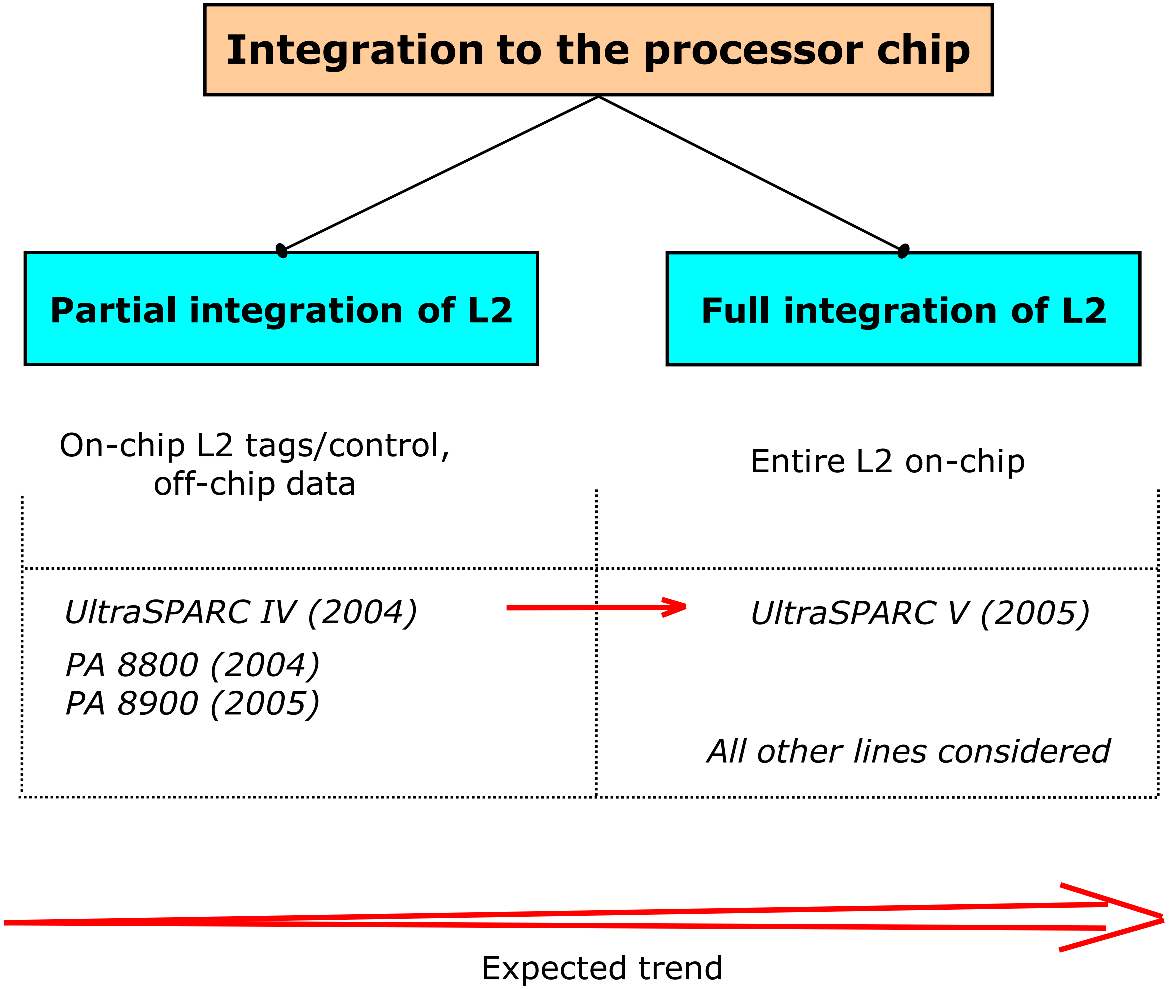
Inclusion policy

Use by instructions/data

Banking policy

Integration of L2 caches to the proc. chip

5. Layout of L2 caches (15)



Integration to the processor chip

Partial integration of L2

Full integration of L2

On-chip L2 tags/control,
off-chip data

Entire L2 on-chip

UltraSPARC IV (2004)

PA 8800 (2004)
PA 8900 (2005)

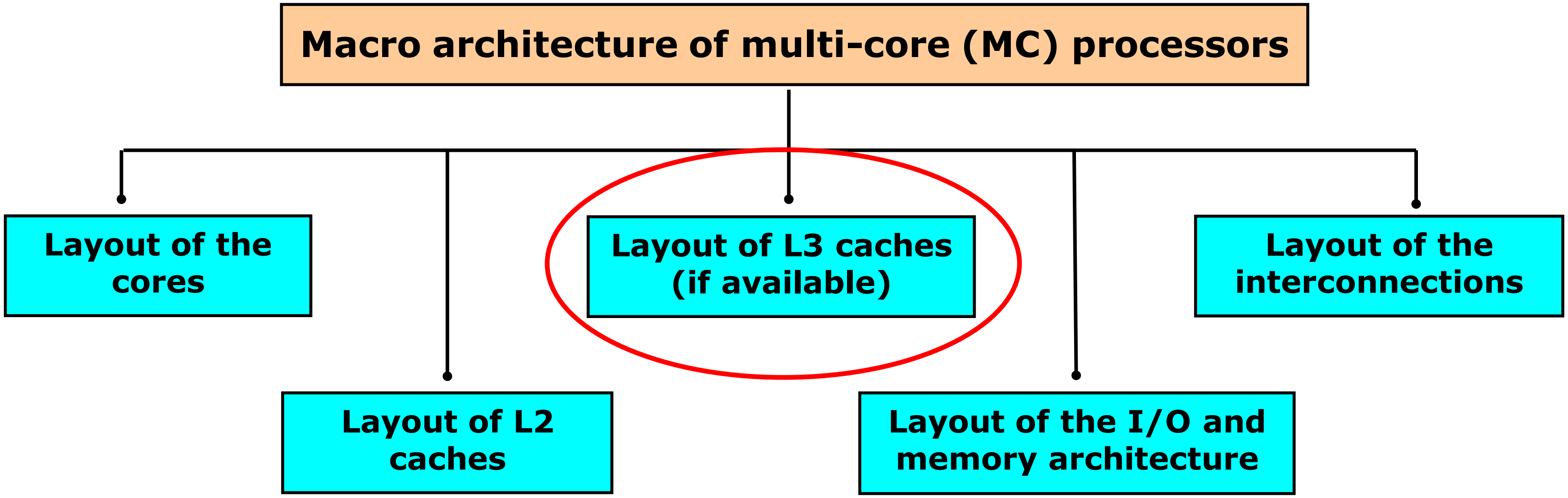
UltraSPARC V (2005)

All other lines considered

Expected trend

6. Layout of L3 caches

6. Layout of L3 caches (1)



Macro architecture of multi-core (MC) processors

Layout of the cores

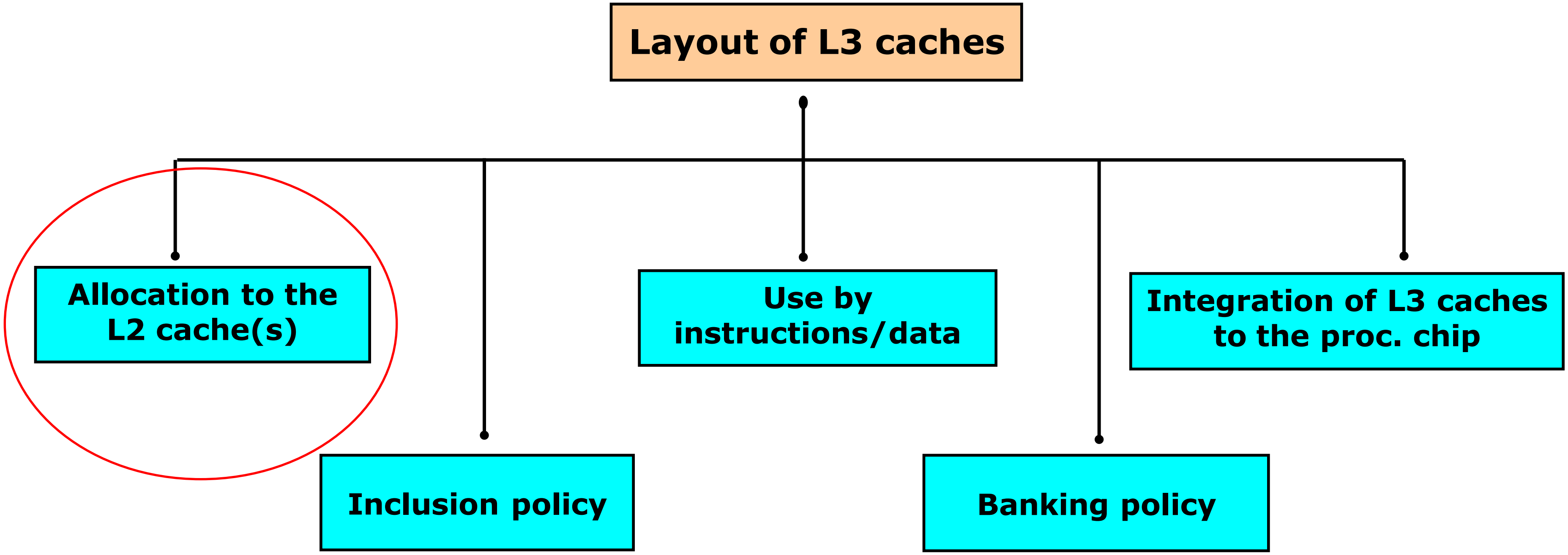
Layout of L2 caches

Layout of L3 caches (if available)

Layout of the I/O and memory architecture

Layout of the interconnections

6. Layout of L3 caches (2)



Layout of L3 caches

Allocation to the L2 cache(s)

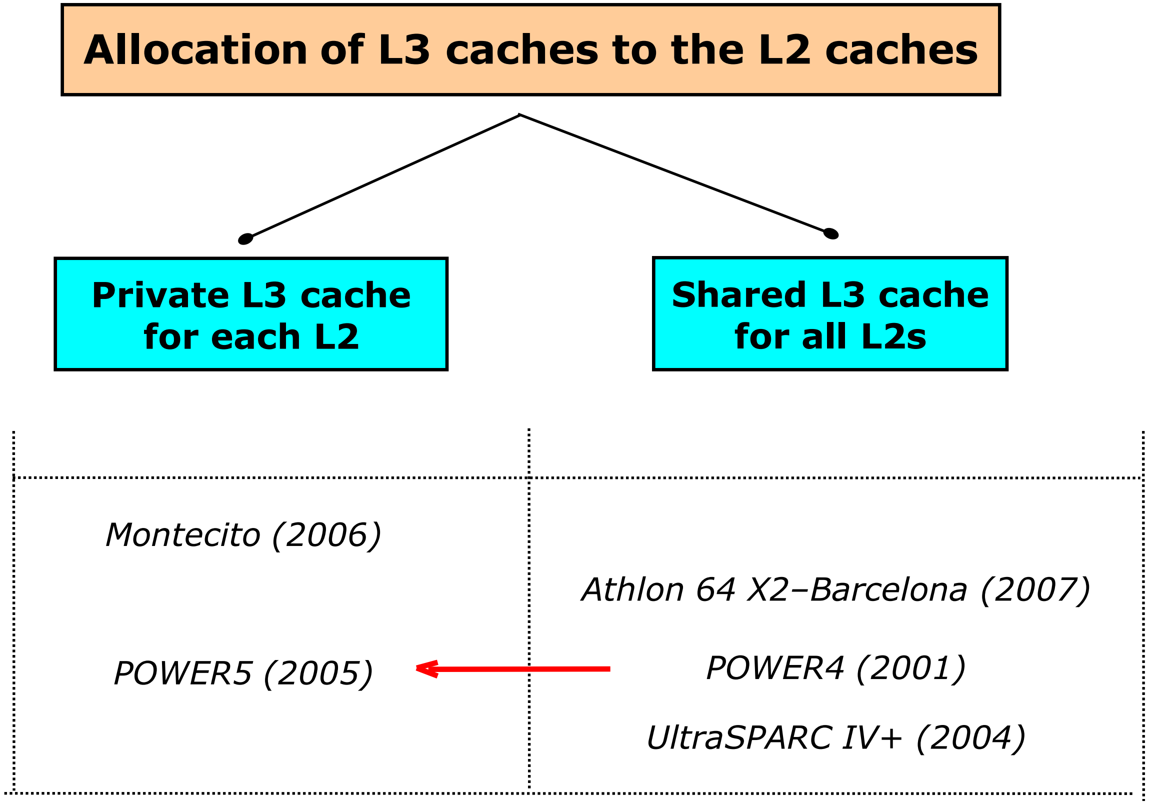
Inclusion policy

Use by instructions/data

Banking policy

Integration of L3 caches to the proc. chip

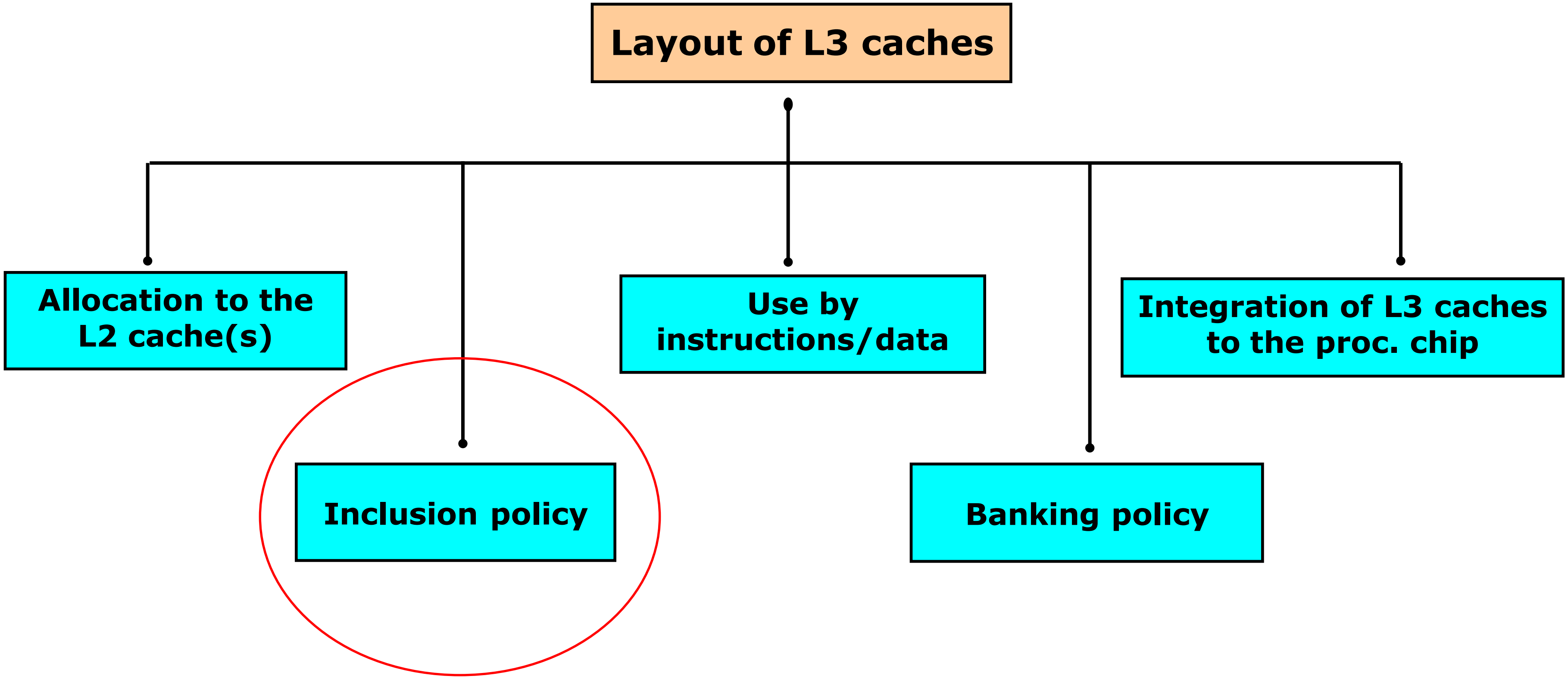
6. Layout of L3 caches (3)



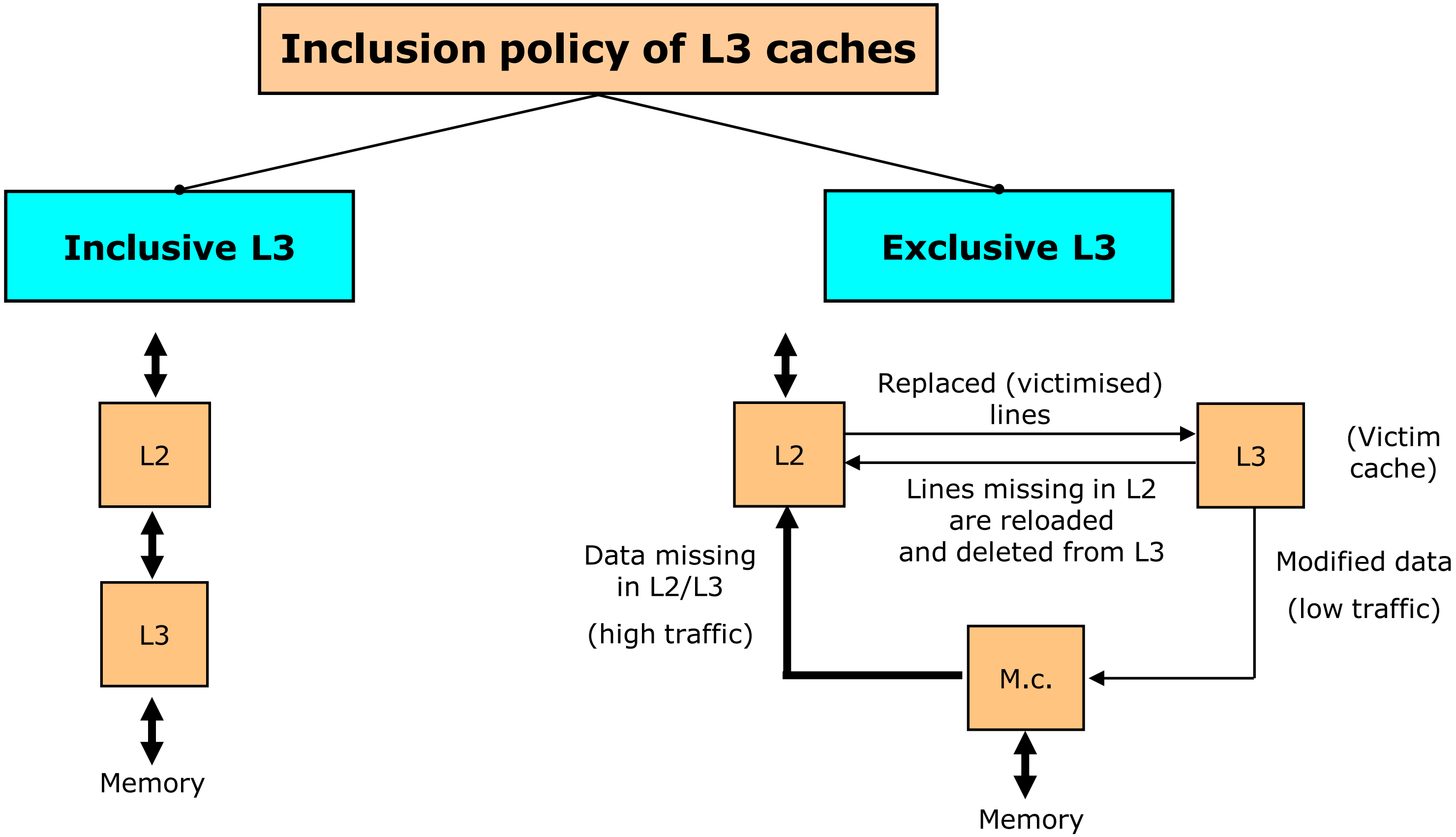
Allocation of L3 caches to the L2 caches

Private L3 cache for each L2	Shared L3 cache for all L2s
Montecito (2006)	*Athlon 64 X2–Barcelona (2007)*
POWER5 (2005)	*POWER4 (2001)*
	UltraSPARC IV+ (2004)

6. Layout of L3 caches (4)

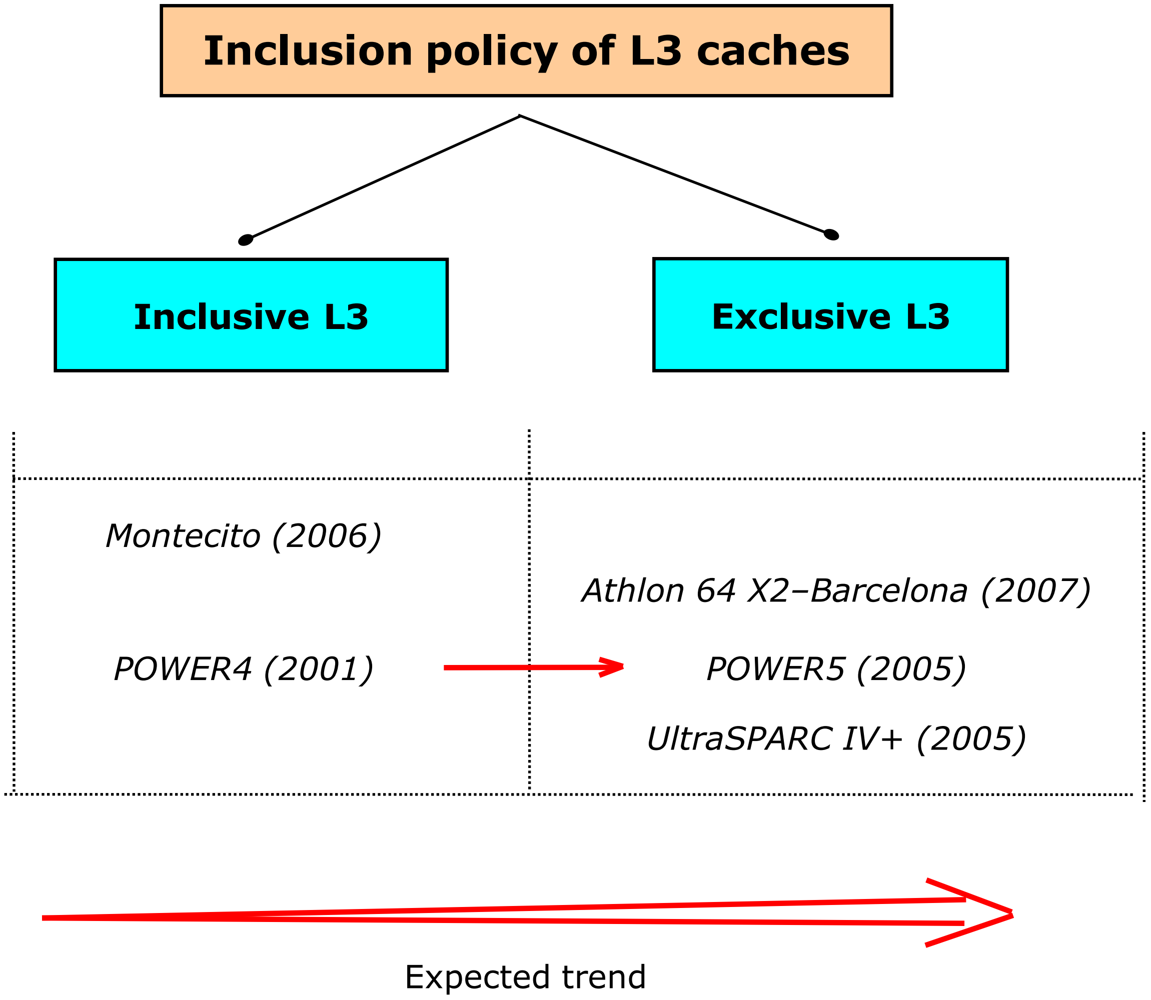


6. Layout of L3 caches (5)



- Lines replaced (victimized) in the L2 are written to L3.
- References to data missing in L2 but available in L3 initiate reloading the pertaining cache line to L2. Reloaded data is deleted from L3
- L3 operates usually as a write back cache (only modified data that is replaced in the L3 is written back to the memory),
- Unmodified data that is replaced in L3 is deleted.

6. Layout of L3 caches (6)



Inclusion policy of L3 caches

Inclusive L3

Exclusive L3

Montecito (2006)

POWER4 (2001)

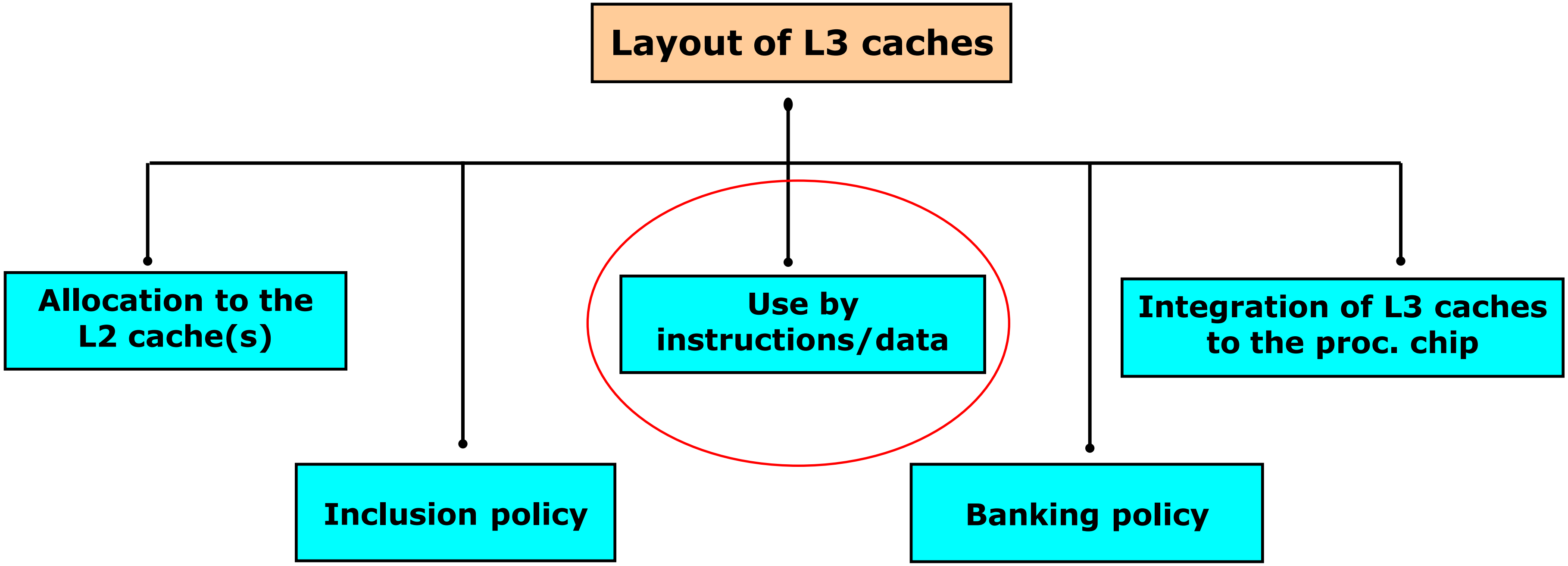
Athlon 64 X2–Barcelona (2007)

POWER5 (2005)

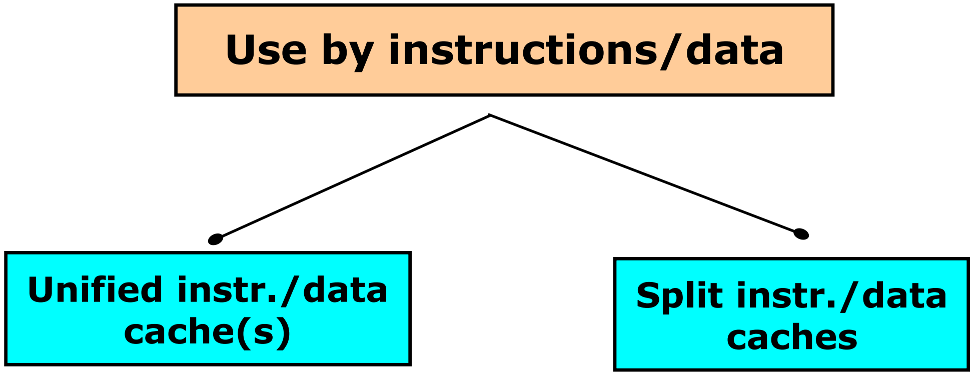
UltraSPARC IV+ (2005)

Expected trend

6. Layout of L3 caches (7)

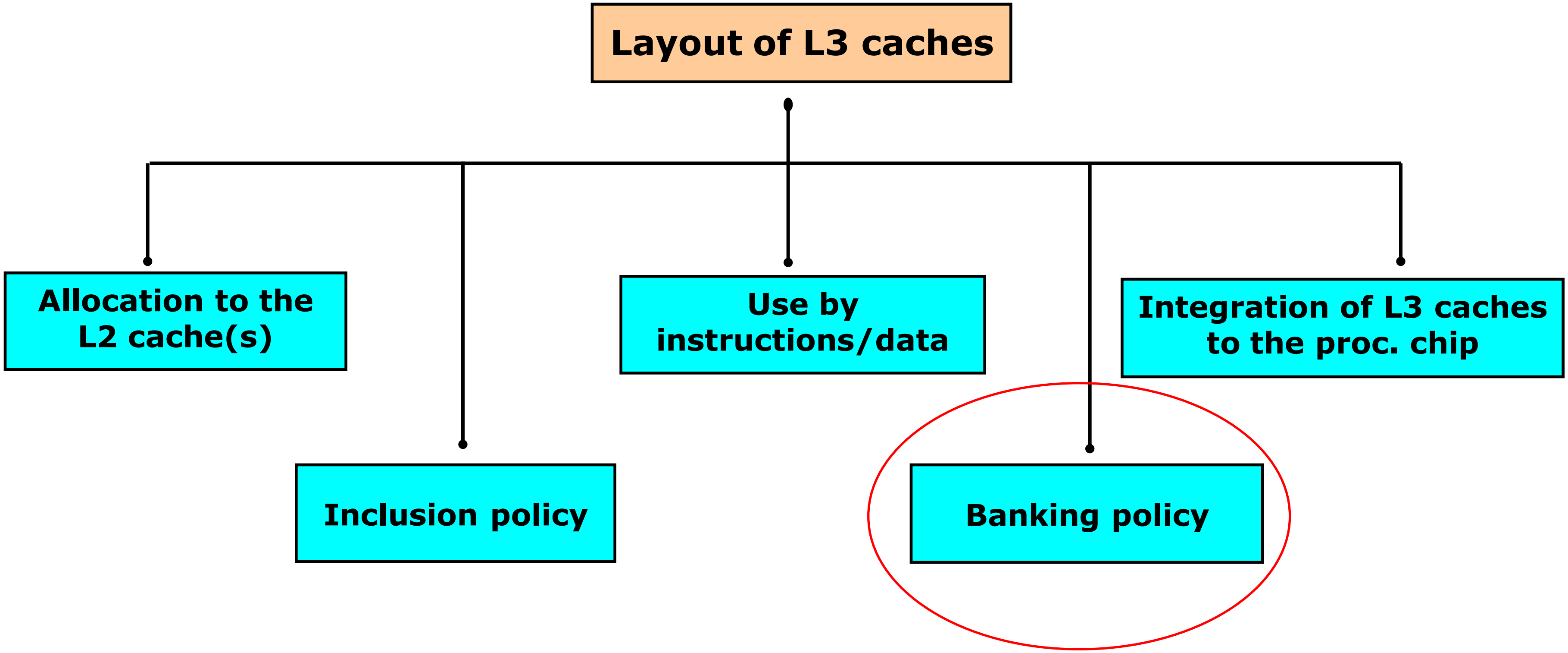


6. Layout of L3 caches (8)

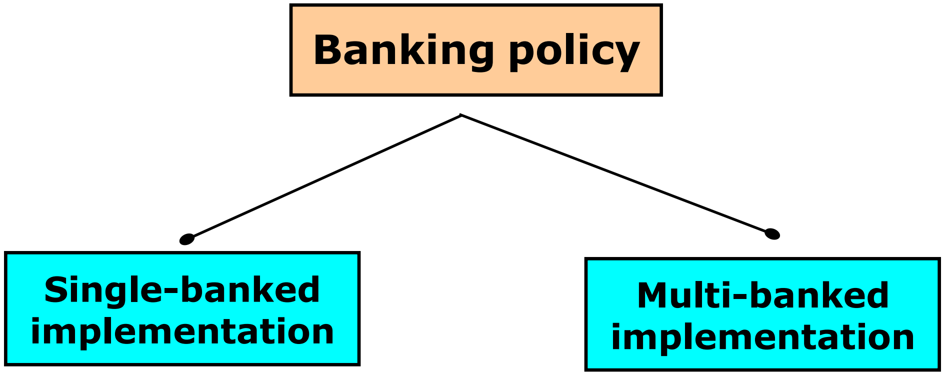


All multicore processors unveiled until now hold both instructions and data

6. Layout of L3 caches (9)



6. Layout of L3 caches (10)



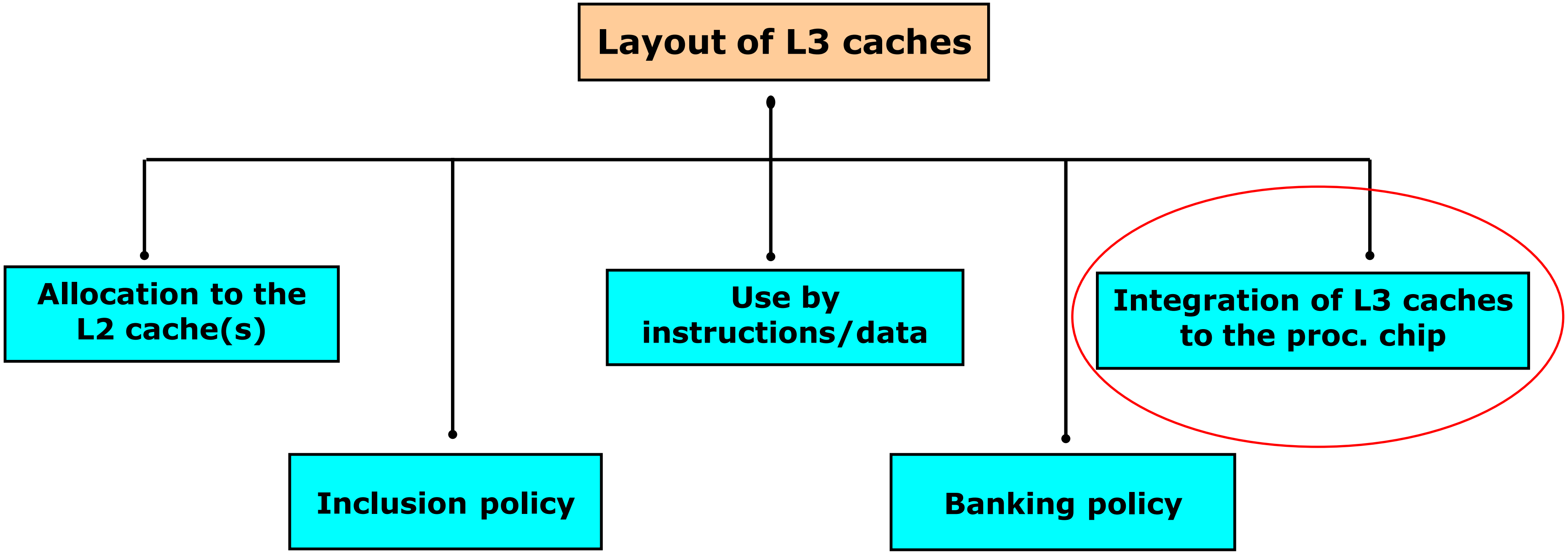
Banking policy

Single-banked implementation

Multi-banked implementation

All multicore processors unveiled until now are multi-banked

6. Layout of L3 caches (11)


Layout of L3 caches

Allocation to the L2 cache(s)

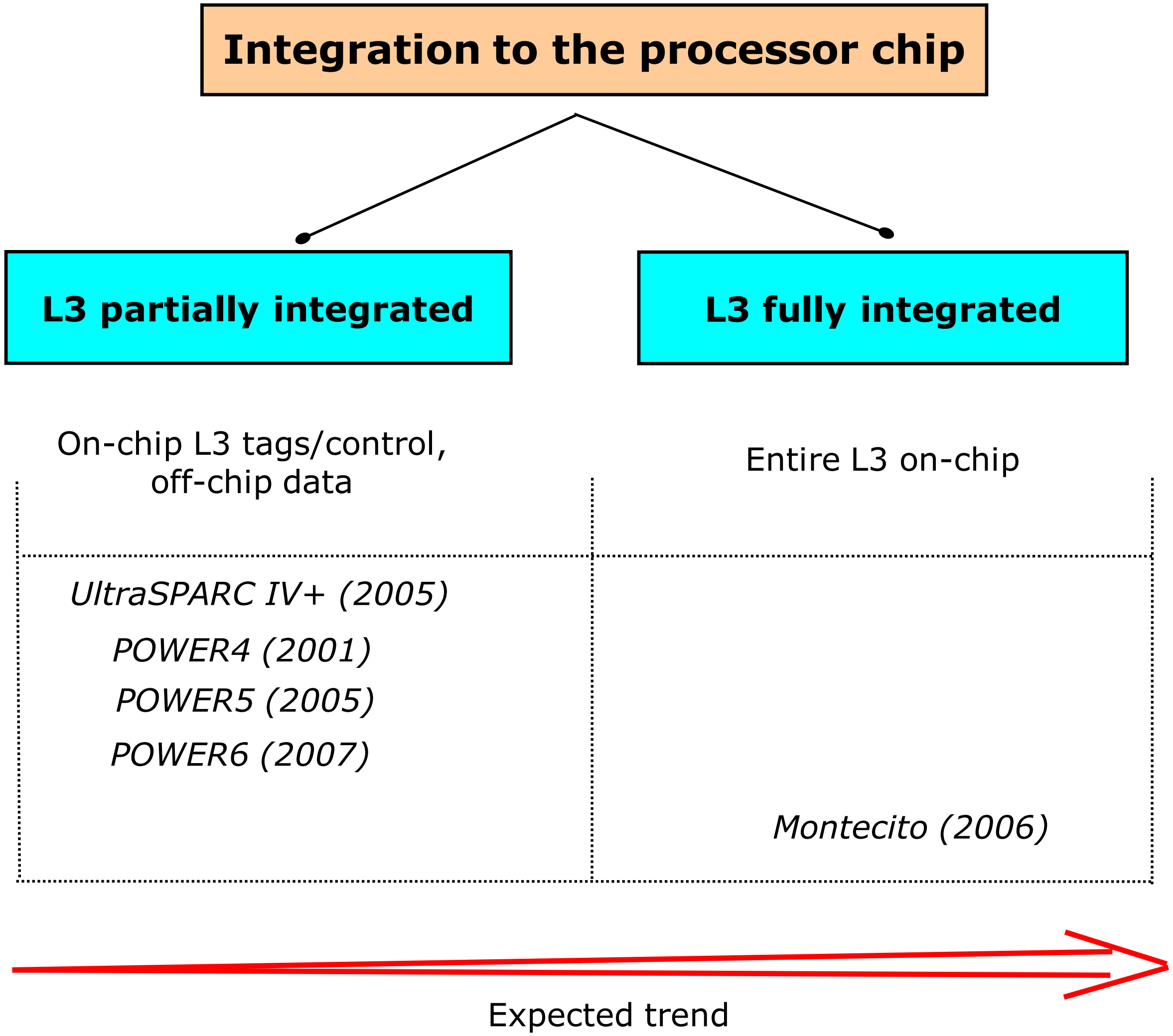
Inclusion policy

Use by instructions/data

Banking policy

Integration of L3 caches to the proc. chip

6. Layout of L3 caches (12)



Integration to the processor chip

L3 partially integrated

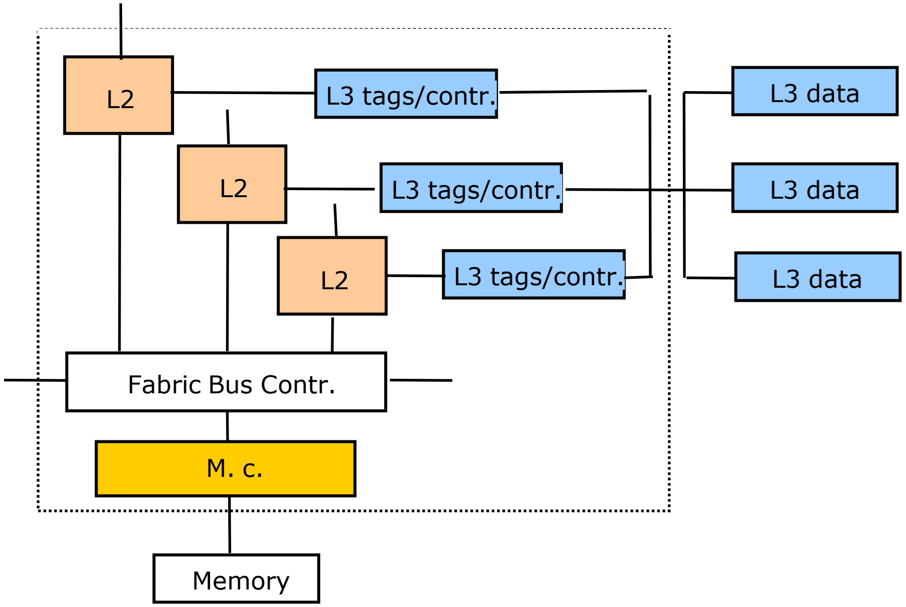
L3 fully integrated

On-chip L3 tags/control, off-chip data	Entire L3 on-chip
UltraSPARC IV+ (2005) *POWER4 (2001)* *POWER5 (2005)* *POWER6 (2007)*	*Montecito (2006)*

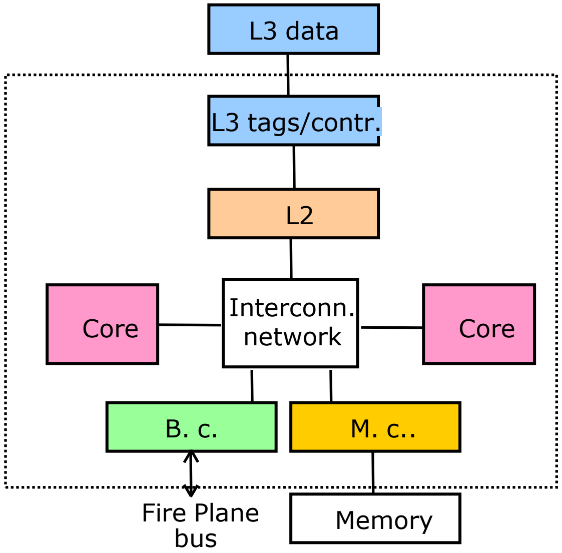
Expected trend

6. Layout of L3 caches (13)

Examples of partially integrated L3 caches:

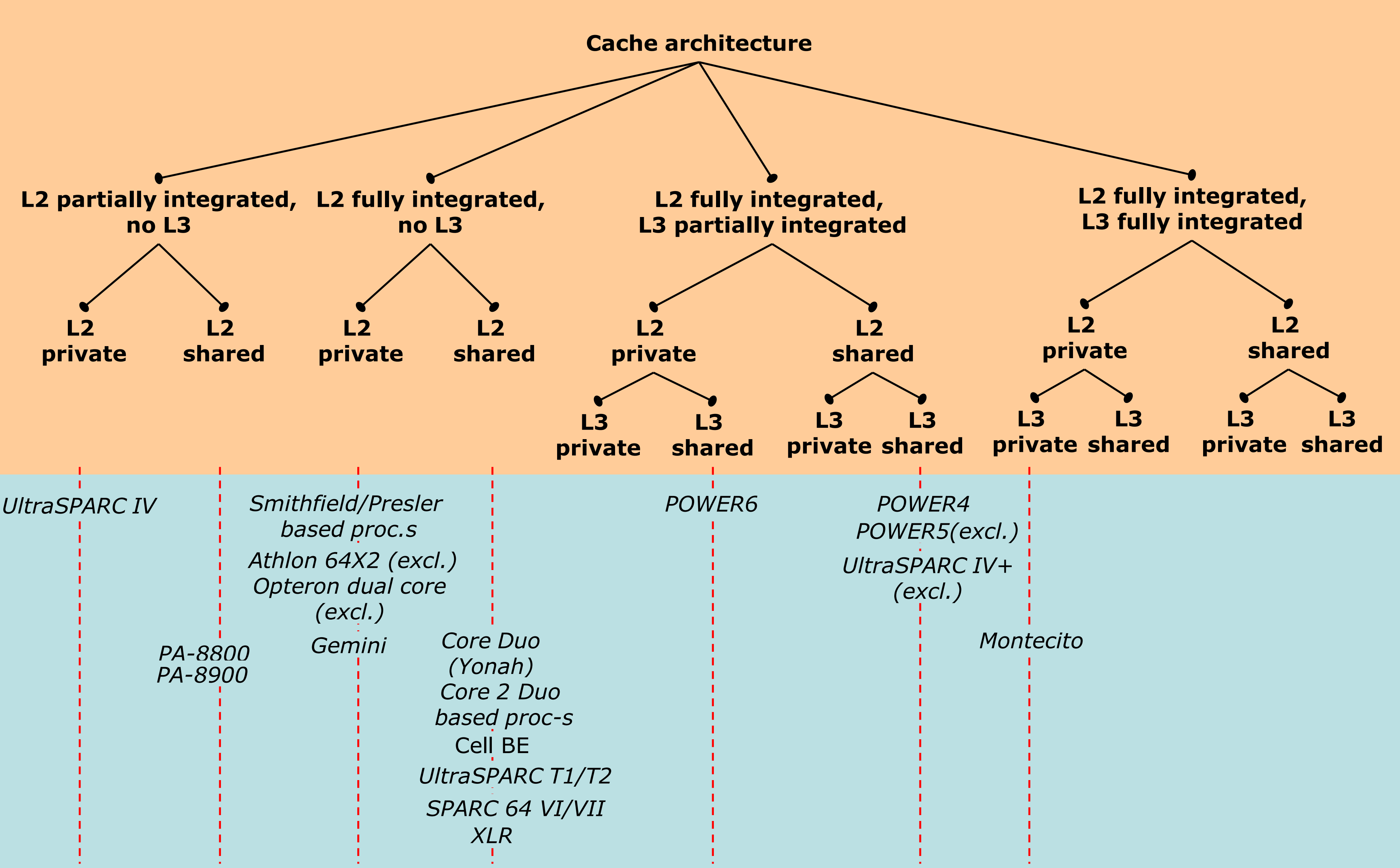


POWER5 (2005):



UltraSPARC IV+ (2005):

6. Layout of L3 caches (14)



Cache architecture

L2 partially integrated, no L3
- **L2 private**: *UltraSPARC IV*
- **L2 shared**: *PA-8800*, *PA-8900*

L2 fully integrated, no L3
- **L2 private**: *Smithfield/Presler based proc.s*, *Athlon 64X2 (excl.)*, *Opteron dual core (excl.)*, *Gemini*
- **L2 shared**: *Core Duo (Yonah)*, *Core 2 Duo based proc-s*, Cell BE, *UltraSPARC T1/T2*, *SPARC 64 VI/VII*, *XLR*

L2 fully integrated, L3 partially integrated
- **L2 private**
  - **L3 private**
  - **L3 shared**: *POWER6*
- **L2 shared**
  - **L3 private**
  - **L3 shared**: *POWER4*, *POWER5(excl.)*, *UltraSPARC IV+ (excl.)*

L2 fully integrated, L3 fully integrated
- **L2 private**
  - **L3 private**: *Montecito*
  - **L3 shared**
- **L2 shared**
  - **L3 private**
  - **L3 shared**

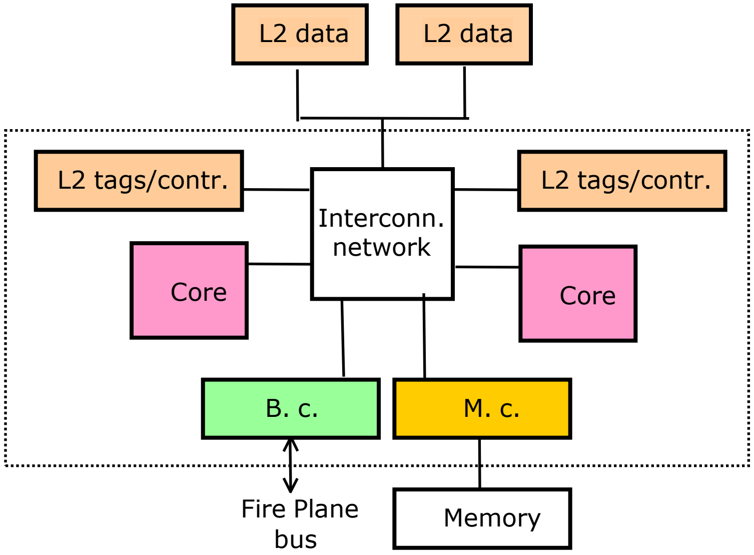
Figure 6.1: Cache architecture of MC processors

(Not differentiating between inclusive/exclusive alternatives in the figure, but denotingit in the examples)

6. Layout of L3 caches (15)

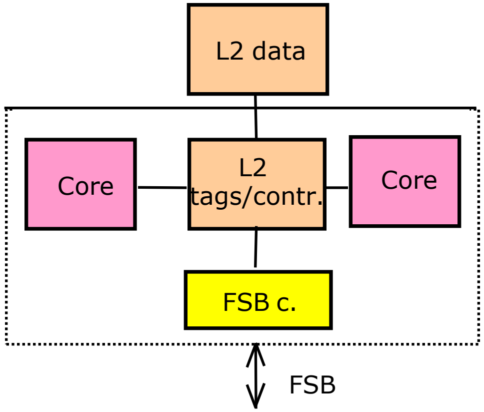
Examples for cache architectures (1)

L2 partially integrated, private
no L3



UltraSPARC IV (2004)

L2 partially integrated, shared
no L3



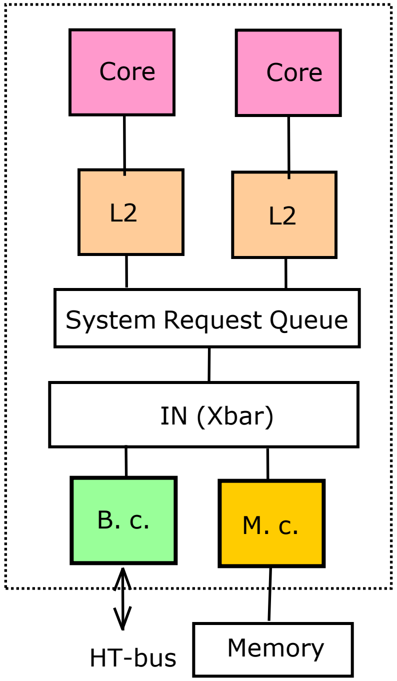
PA-8800 (2004)
PA-8900 (2005)

6. Layout of L3 caches (16)

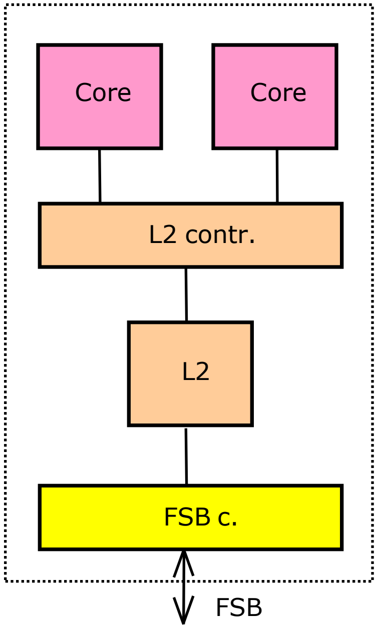
Examples for cache architectures (2)

L2 fully integrated, private no L3

L2 fully integrated, shared no L3

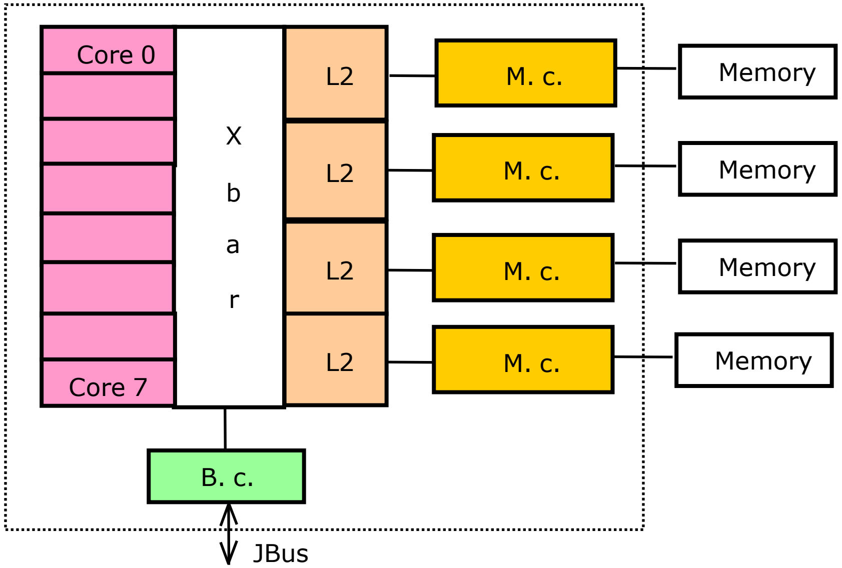


Athlon 64 X2 (2005)



Core Duo (2006)

Core 2 Duo (2006)

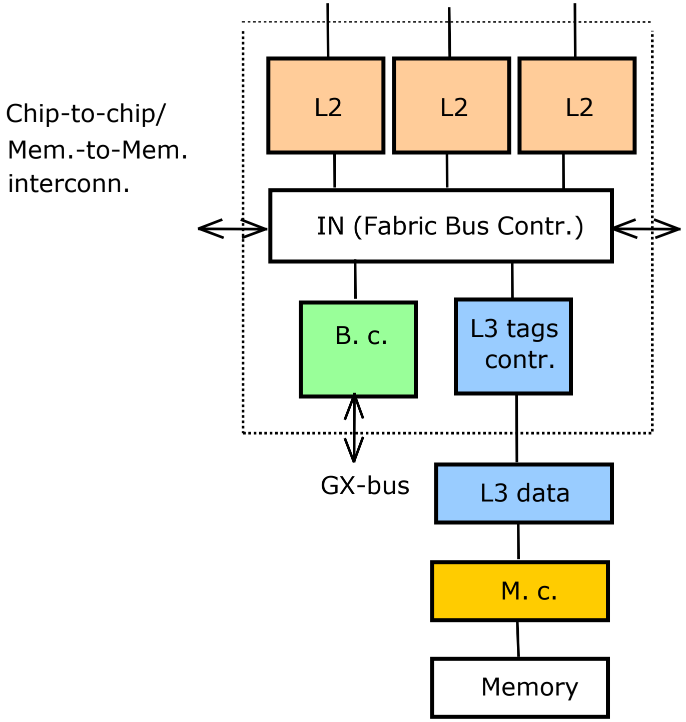


UltraSPARC T1 (2005)

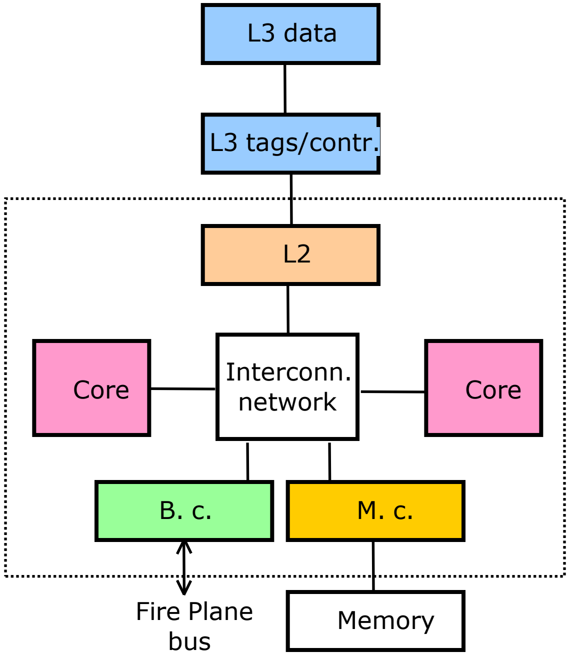
6. Layout of L3 caches (17)

Examples for cache architectures (3)

L2 fully integrated, shared
L3 partially integrated



POWER4 (2001)

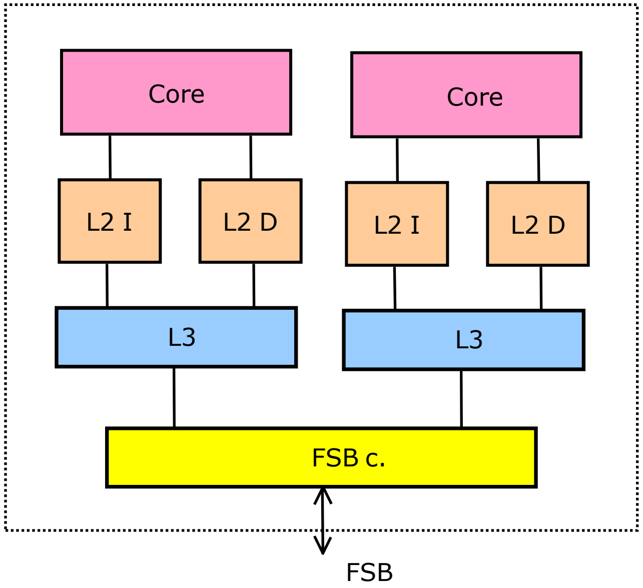


UltraSPARC IV+ (2005)

6. Layout of L3 caches (18)

Examples for cache architectures (4)

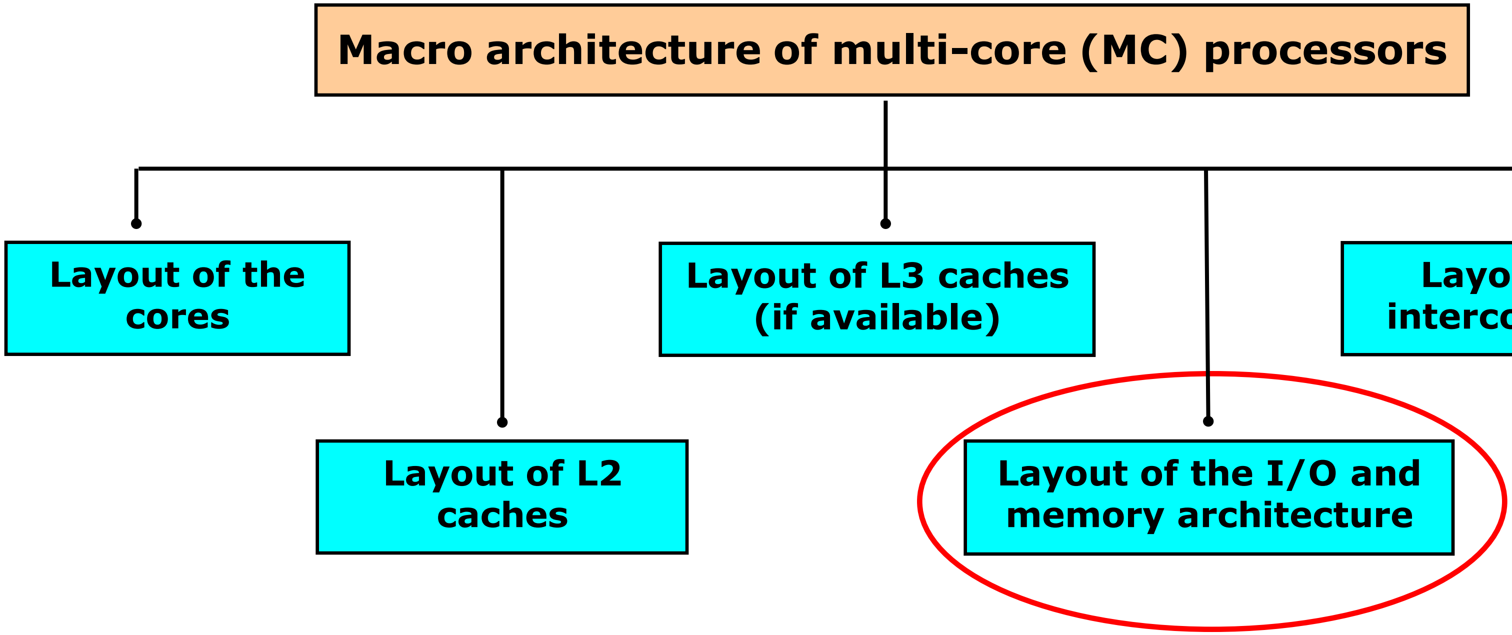
L2 fully integrated, private,
L3 fully integrated



Montecito (2006)

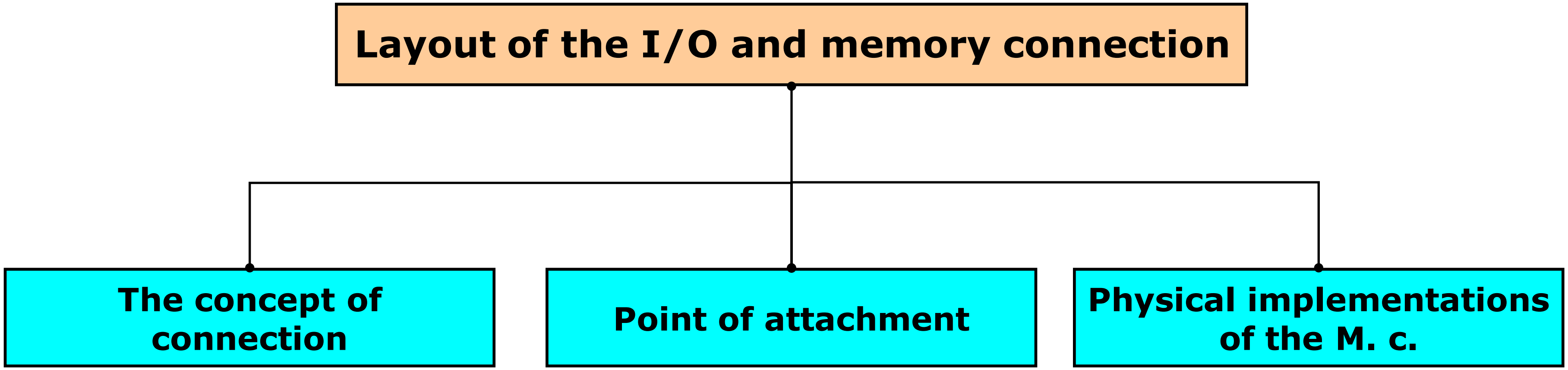
7. Layout of memory and I/O architecture

7. Layout of memory and I/O architecture (1)



Macro architecture of multi-core (MC) processors
Layout of the cores
Layout of L2 caches
Layout of L3 caches (if available)
Layout of the I/O and memory architecture
Layout of the interconnections

7. Layout of memory and I/O architecture (2)



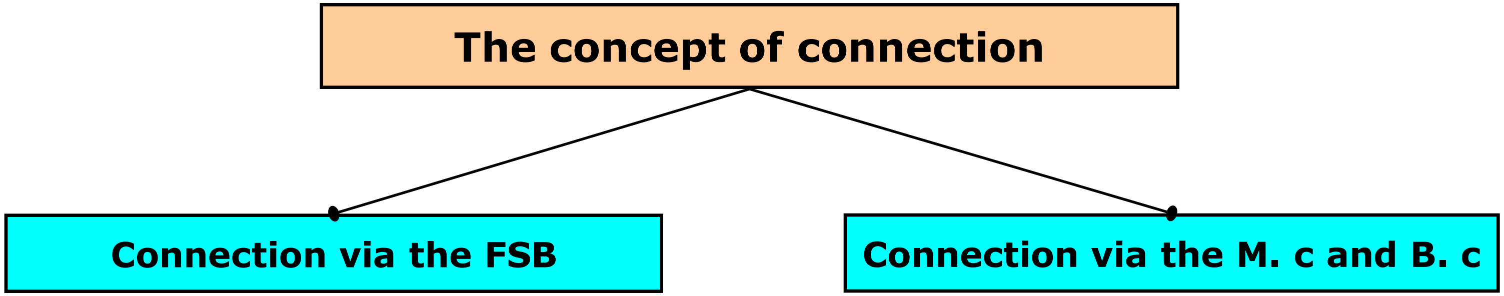
Layout of the I/O and memory connection

The concept of connection

Point of attachment

Physical implementations of the M. c.

7. Layout of memory and I/O architecture (3)



The concept of connection

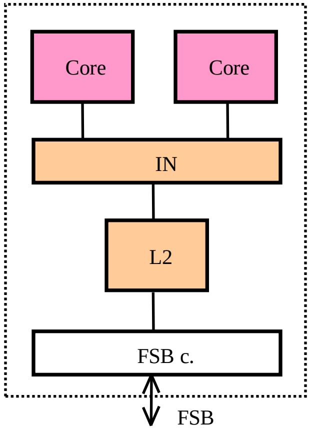
Connection via the FSB

Connection via the M. c and B. c

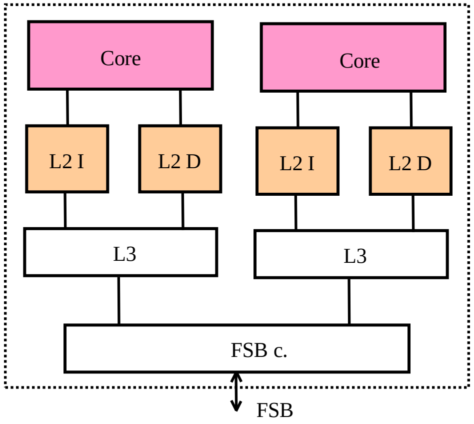
Used typically in connection with
off-chip M. c.-s

Used in connection with
on-chip M. c.-s

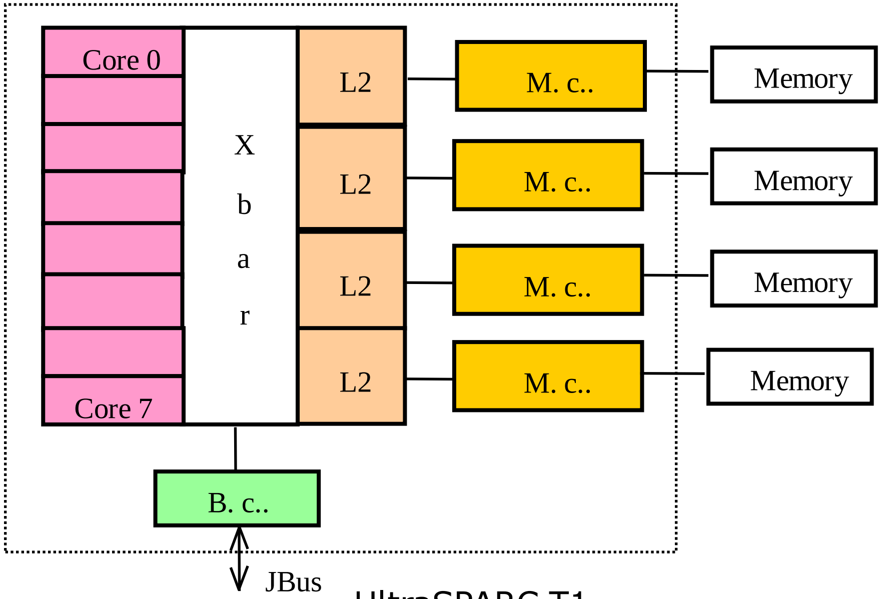
Examples:



Core 2 Duo

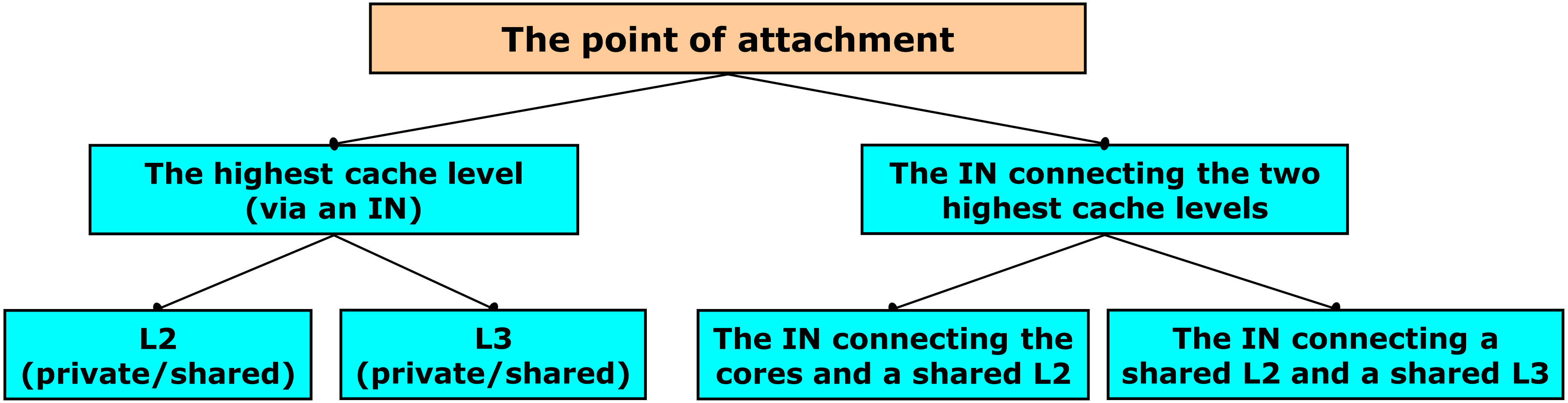


Montecito



UltraSPARC T1

7. Layout of memory and I/O architecture (4)



The point of attachment

The highest cache level (via an IN)

The IN connecting the two highest cache levels

L2 (private/shared)

L3 (private/shared)

The IN connecting the cores and a shared L2

The IN connecting a shared L2 and a shared L3

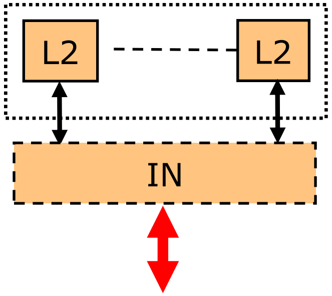
(2-level caches)

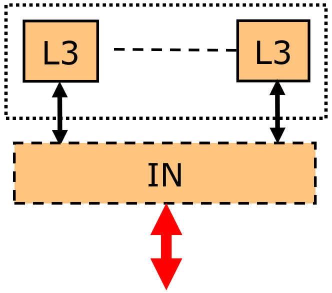
(3-level caches)

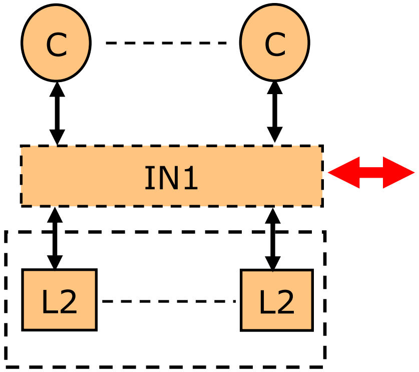
(2-level caches)

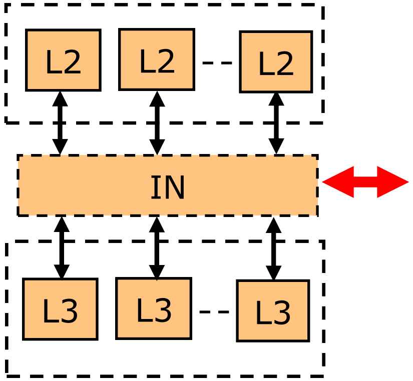
(3-level caches)

Examples:



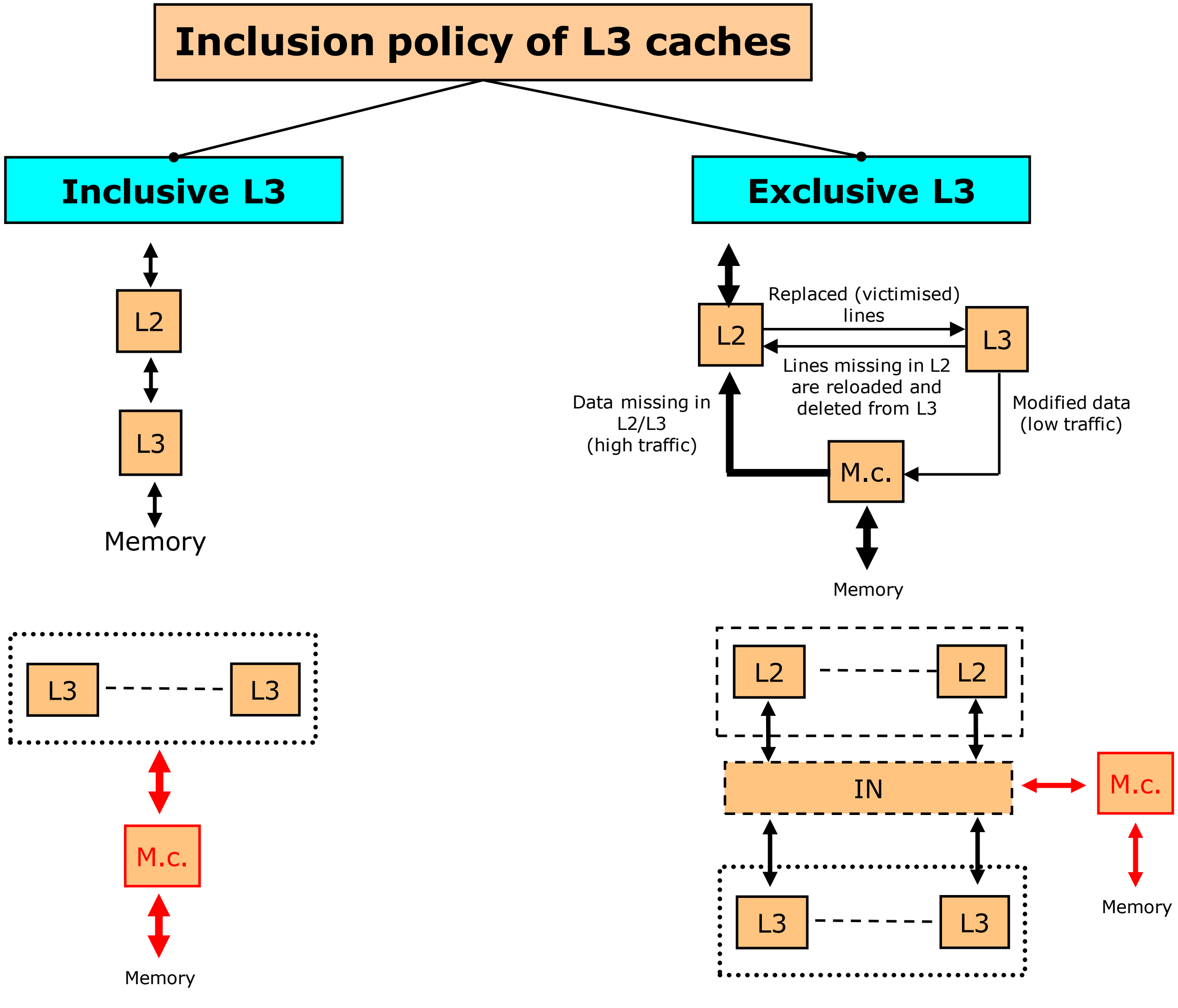






7. Layout of memory and I/O architecture (5)

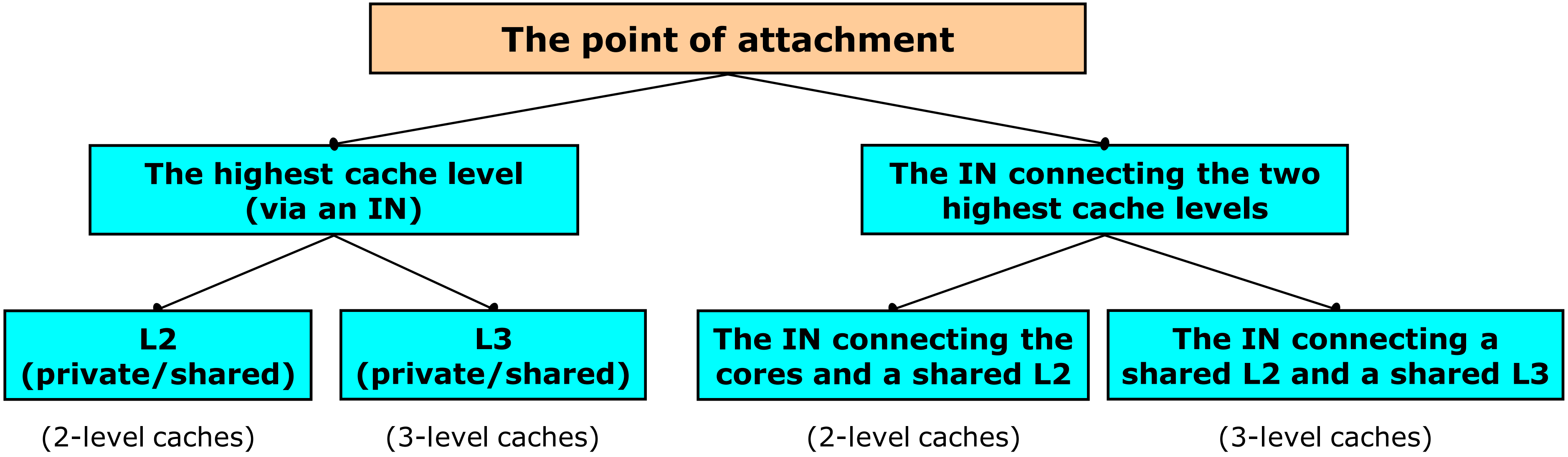
Cache inclusivity vs. memory attachment


Inclusion policy of L3 caches
Inclusive L3
Exclusive L3
L2
L3
Memory
L2
L3
Replaced (victimised) lines
Lines missing in L2 are reloaded and deleted from L3
Data missing in L2/L3 (high traffic)
Modified data (low traffic)
M.c.
Memory
L3
L3
M.c.
Memory
L2
L2
IN
M.c.
L3
L3
Memory

Montecito (2006)
POWER4 (2001)

UltraSPARC IV+ (2004)
POWER5 (2004)

7. Layout of memory and I/O architecture (6)



The point of attachment

The highest cache level (via an IN)

The IN connecting the two highest cache levels

L2 (private/shared)

(2-level caches)

L3 (private/shared)

(3-level caches)

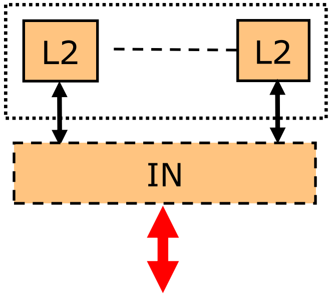
The IN connecting the cores and a shared L2

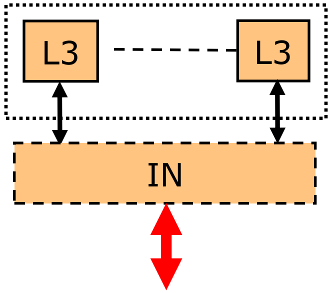
(2-level caches)

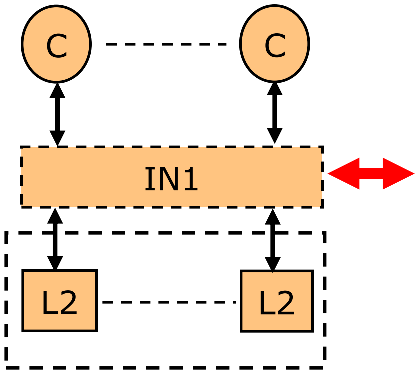
The IN connecting a shared L2 and a shared L3

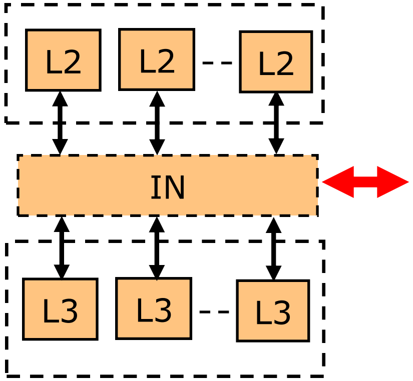
(3-level caches)

Examples:









The M. c is connected usually in this way
if the highest level cache is **inclusive**.

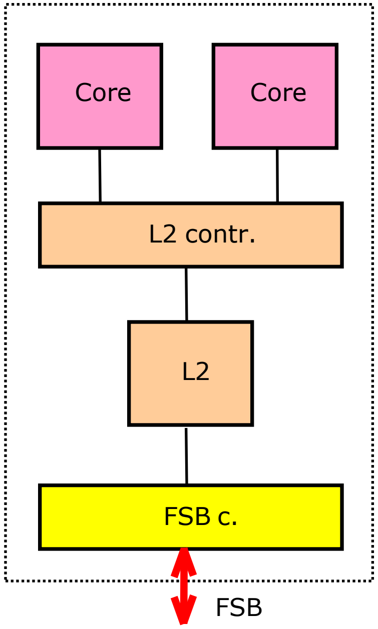
The M. c is connected usually in this way
if the highest level cache is **exclusive**.

7. Layout of memory and I/O architecture (7)

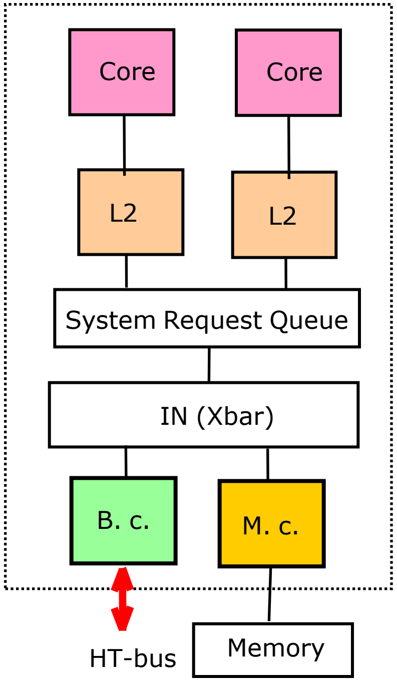
Examples for attaching memory and I/O via the highest cache level

In case of a two-level cache hierarchy

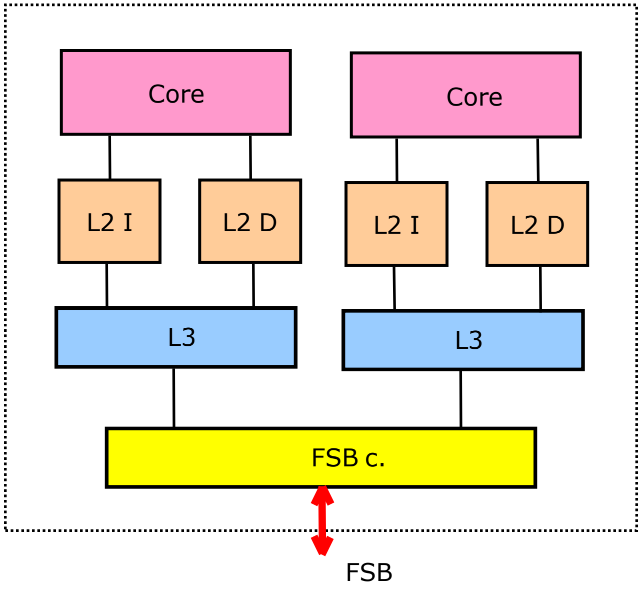
In case of a three-level cache hierarchy



Core 2 Duo (2006)



Athlon 64 X2 (2005)



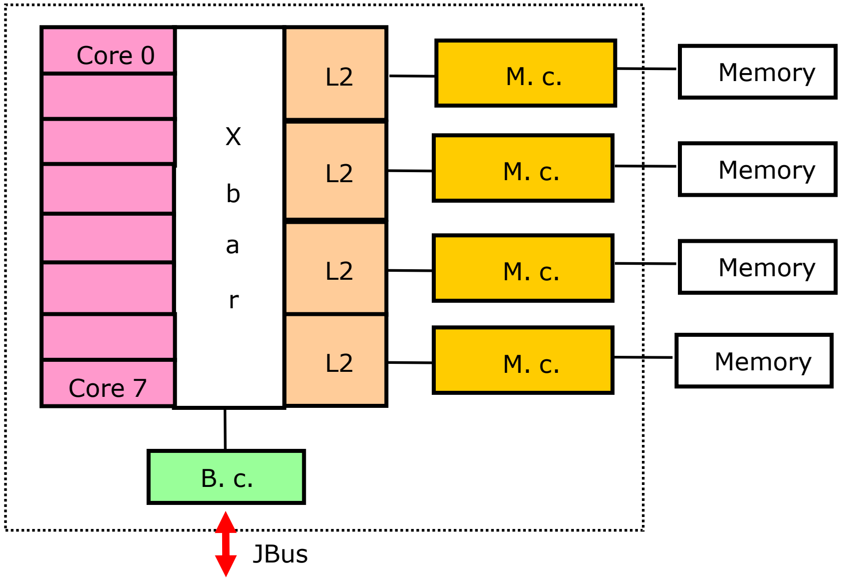
Montecito (2006)

7. Layout of memory and I/O architecture (8)

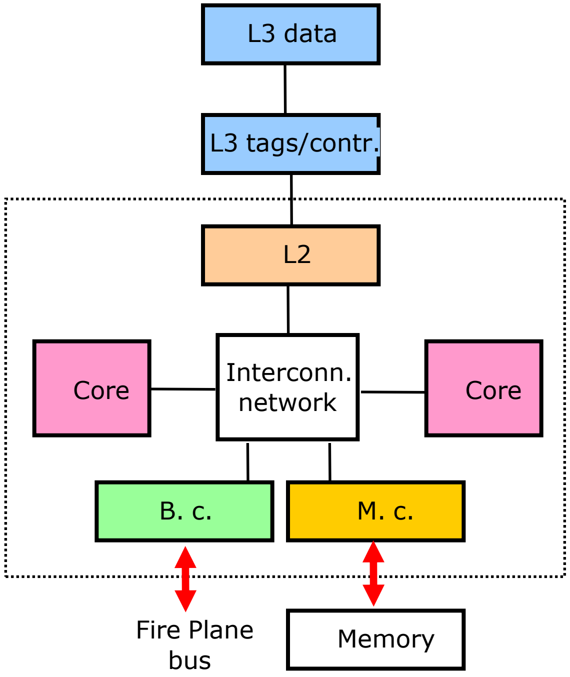
Examples for attaching memory and I/O via the interconnection network connecting the two highest levels of the cache hierarcy

In case of a two-level cache hierarchy

In case of a three-level cache hierarchy

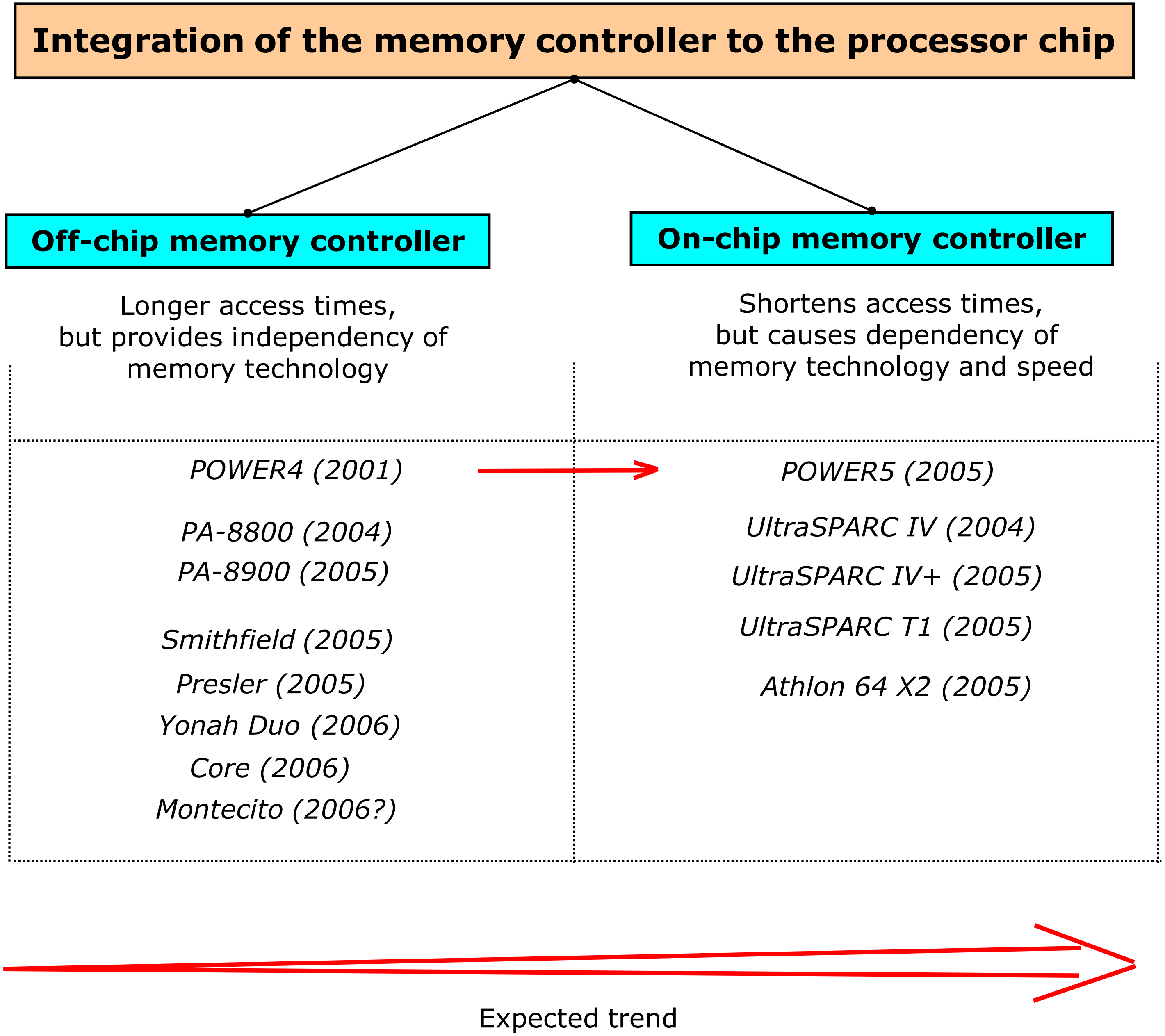


UltraSPARC T1 (2005)



UltraSPARC IV+ (2005)
(exclusive L3)

7. Layout of memory and I/O architecture (9)



Integration of the memory controller to the processor chip

Off-chip memory controller

Longer access times,
but provides independency of
memory technology

POWER4 (2001)

PA-8800 (2004)

PA-8900 (2005)

Smithfield (2005)

Presler (2005)

Yonah Duo (2006)

Core (2006)

Montecito (2006?)

On-chip memory controller

Shortens access times,
but causes dependency of
memory technology and speed

POWER5 (2005)

UltraSPARC IV (2004)

UltraSPARC IV+ (2005)

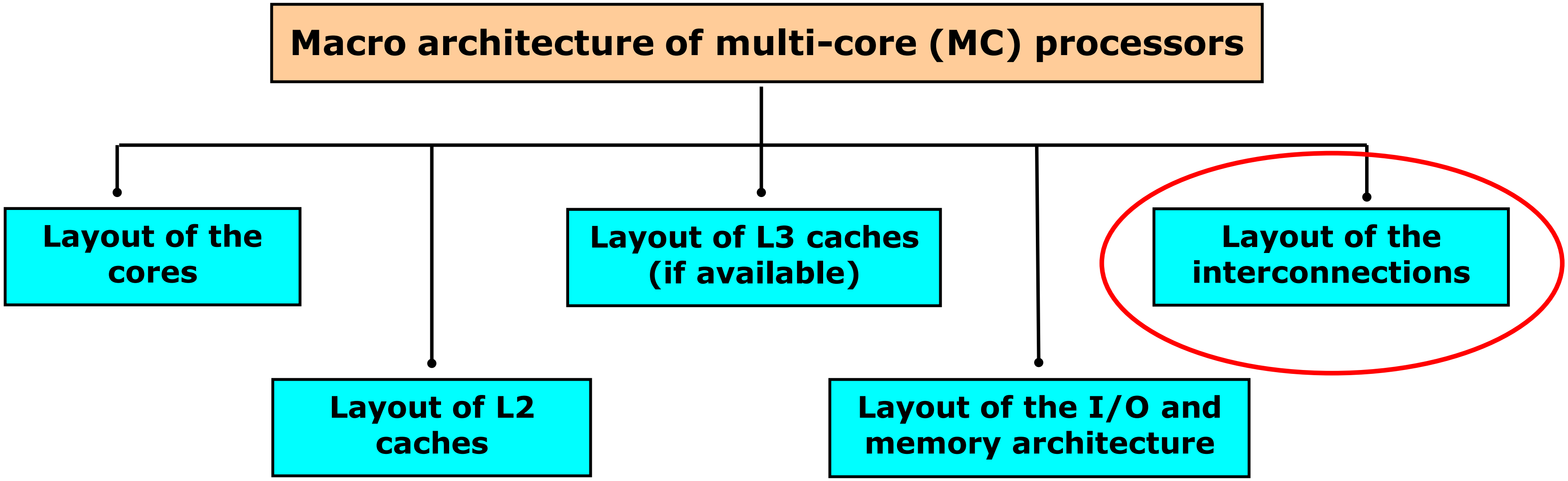
UltraSPARC T1 (2005)

Athlon 64 X2 (2005)

Expected trend

8. Layout of the on-chip interconnections

1. Layout of interconnections (1)



Macro architecture of multi-core (MC) processors

Layout of the cores

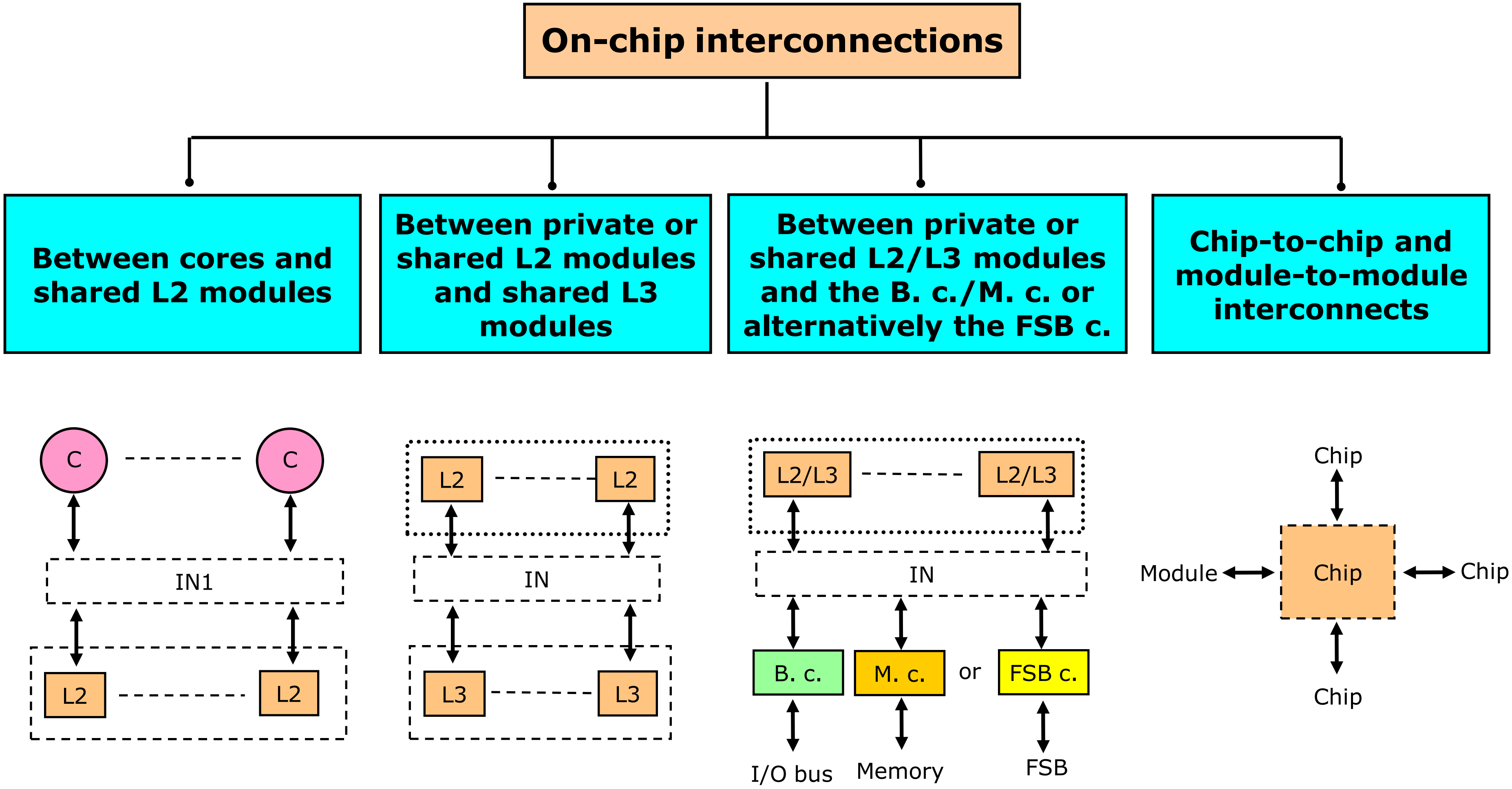
Layout of L2 caches

Layout of L3 caches (if available)

Layout of the I/O and memory architecture

Layout of the interconnections

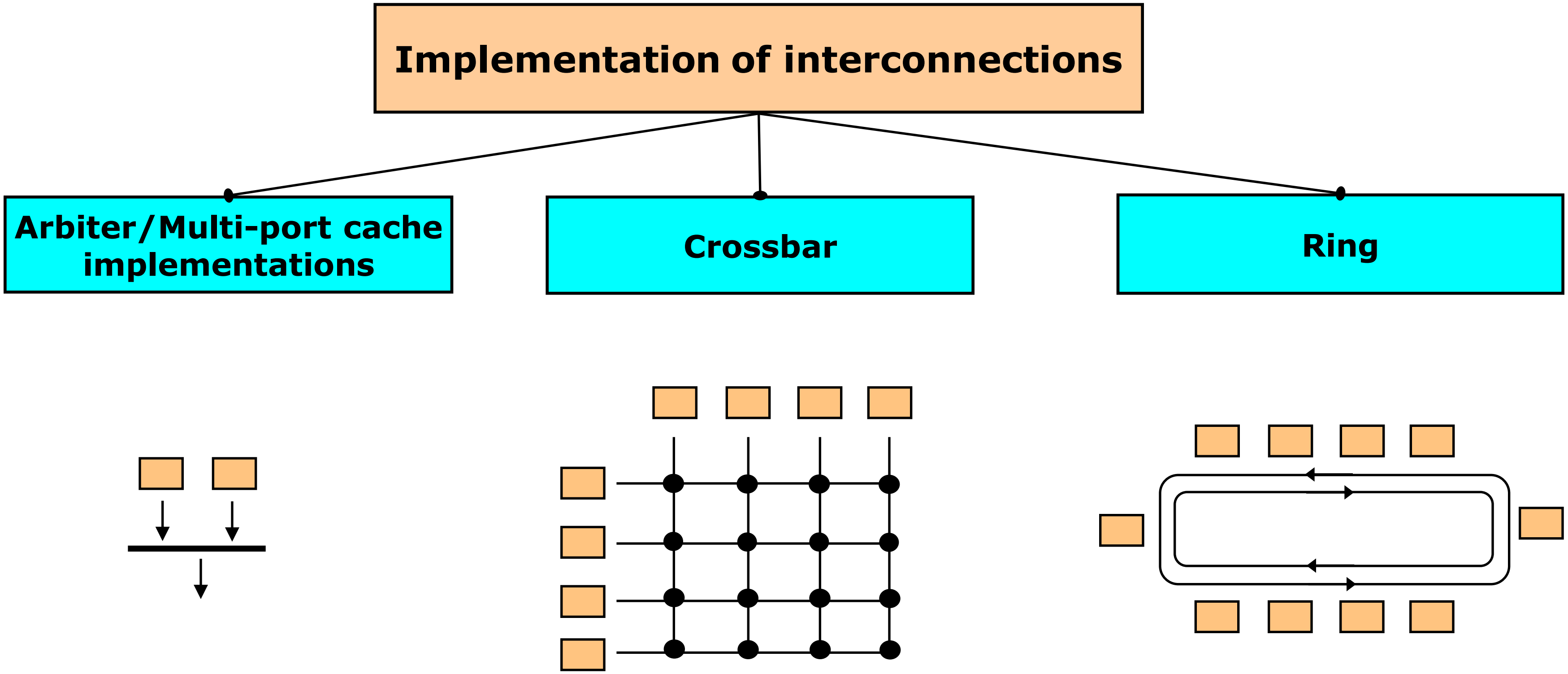
8. Layout of interconnections (2)



On-chip interconnections

- **Between cores and shared L2 modules**
- **Between private or shared L2 modules and shared L3 modules**
- **Between private or shared L2/L3 modules and the B. c./M. c. or alternatively the FSB c.**
- **Chip-to-chip and module-to-module interconnects**

8. Layout of interconnections (3)



Implementation of interconnections

Arbiter/Multi-port cache implementations

Crossbar

Ring

Quantitative aspects, such as the number of sources/destinations or bandwidth requirement affect which implementation alternative is the most beneficial.

For a small number of sources/destinations

For a larger number of sources/destinations

eg. to connect dual-cores to shared L2 caches

UltraSPARC T1 (2005)
UltraSPARC T2 (2007)

Cell BE (2006)
XRI (2005)

9. Basic alternatives of macro-architectures of MC processors

9. Basic alternatives of macro-architectures of MC processors (1)

Basic dimensions spanning the design space

- Kind of related cache architecture
- Layout of the memory and I/O connection
- Layout of the IN(s) involved

Interrelationships between particular dimensions

- Cache inclusivity affects the layout of the memory connection

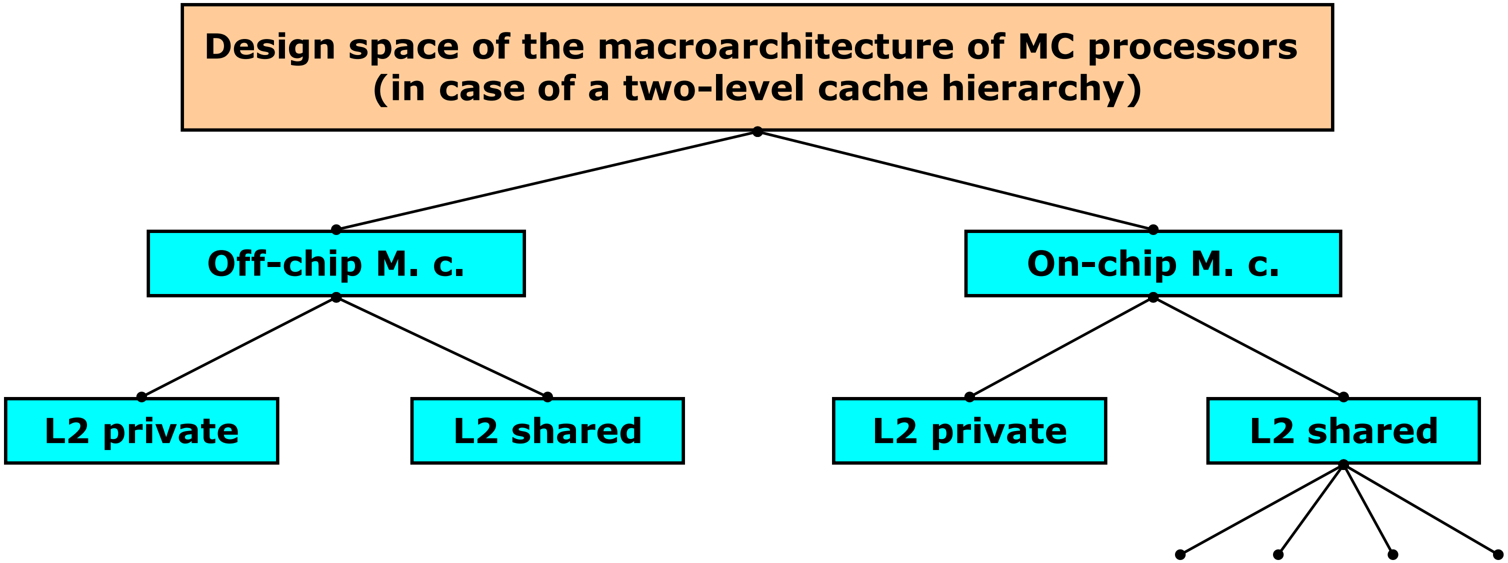
9. Basic alternatives of macro-architectures of MC processors (2)

Considering a selected part of the design space of the macroarchitecture of MC processors

Assumptions:

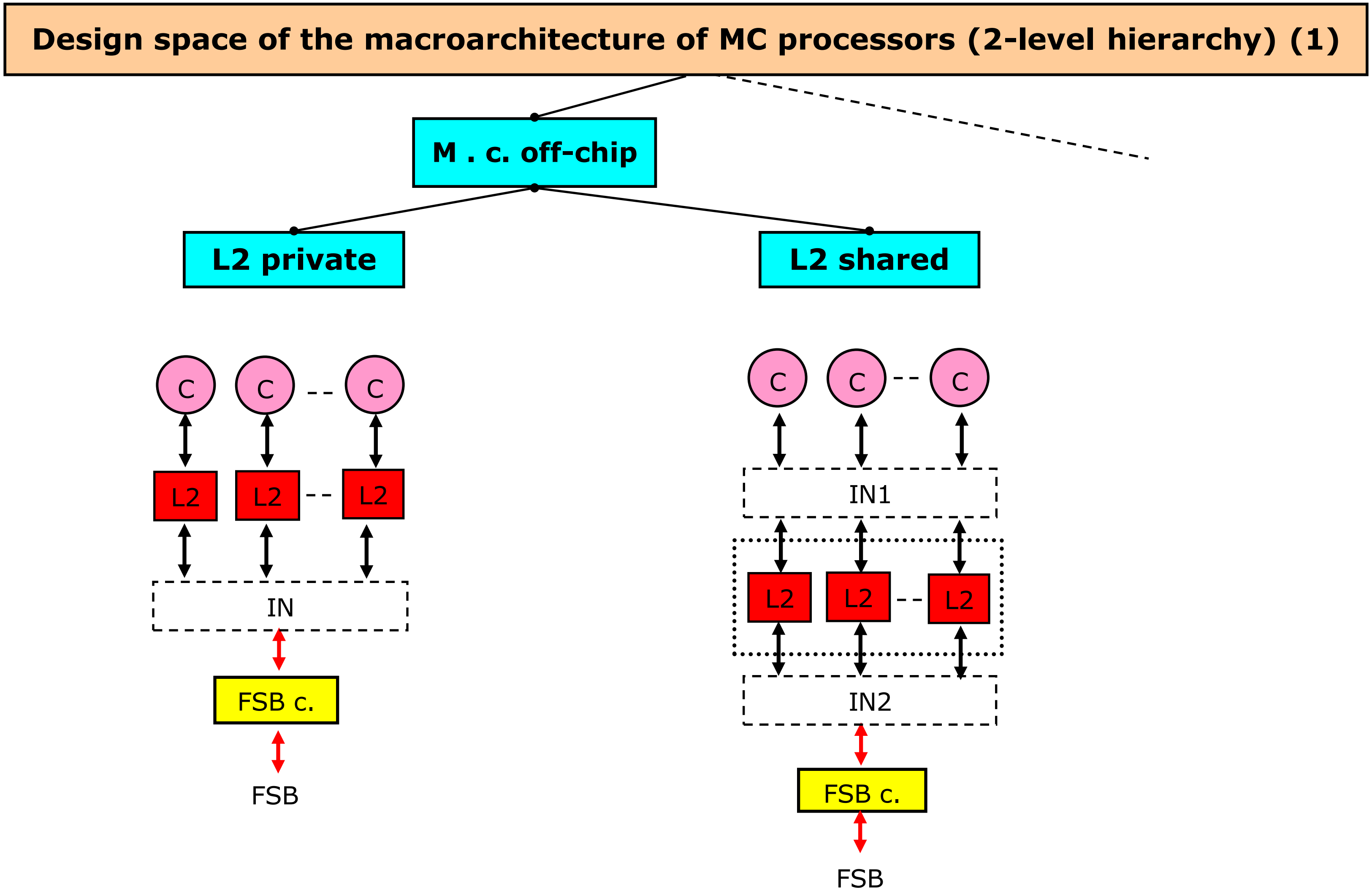
- A two-level cache hierarchy is considered (no L3).
- L2 is supposed to be partially integrated on the chip (i.e. only the tags are on the chip). This is not indicated in the design space tree but denoted in the examples.
- The kind of the implementation of the involved INs is not considered.
- In case of an off-chip M. controller the use of an FSB is assumed.

9. Basic alternatives of macro-architectures of MC processors (3)


Design space of the macroarchitecture of MC processors
(in case of a two-level cache hierarchy)
Off-chip M. c.
On-chip M. c.
L2 private
L2 shared
L2 private
L2 shared

Part of the design-space specifying the point of attachment for the memory and I/O

9. Basic alternatives of macro-architectures of MC processors (4)

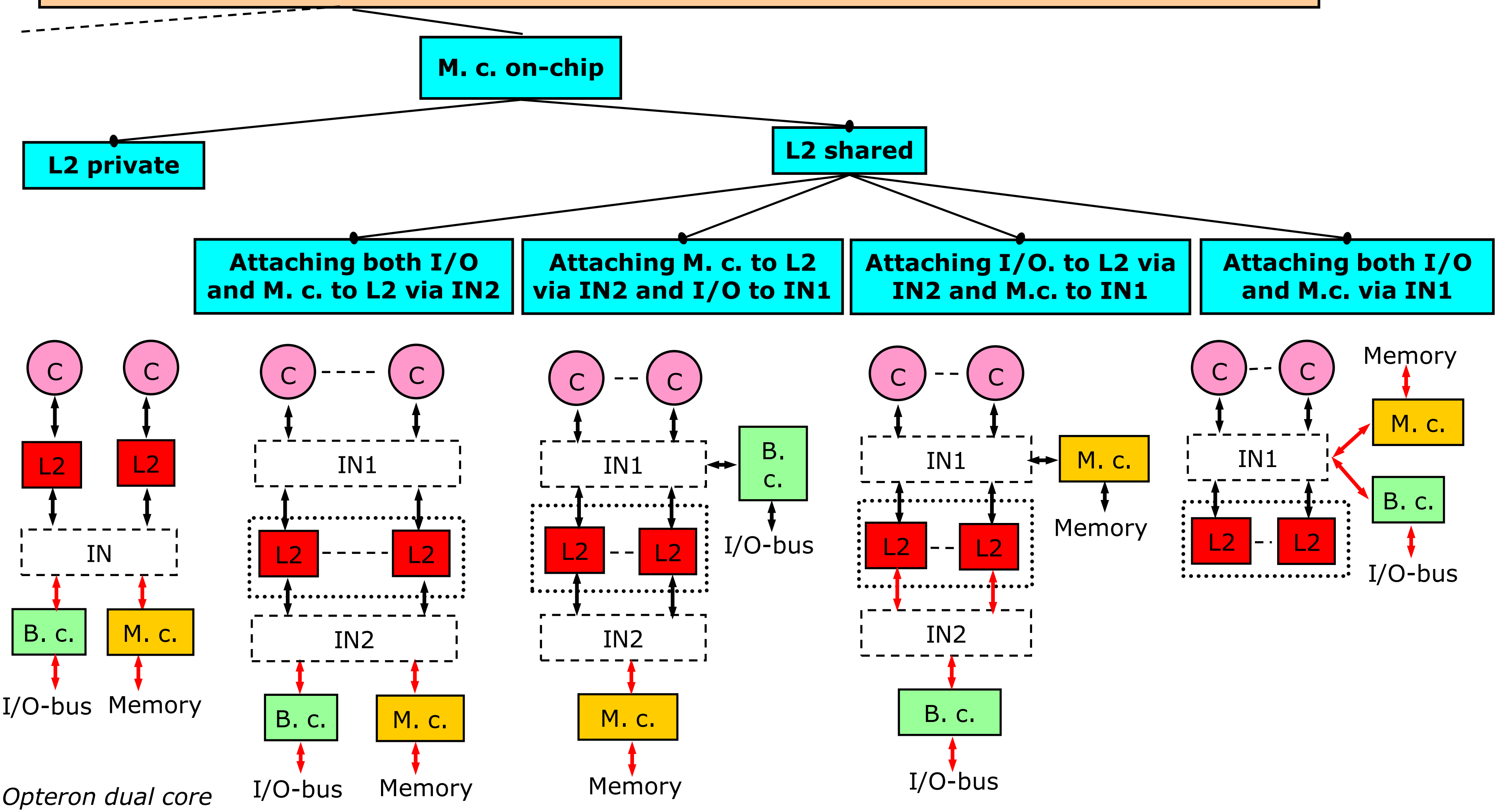

Design space of the macroarchitecture of MC processors (2-level hierarchy) (1)
M . c. off-chip
L2 private
L2 shared
C
C
C
L2
L2
L2
IN
FSB c.
FSB
C
C
C
IN1
L2
L2
L2
IN2
FSB c.
FSB

Smithfield/Presler based processors (2005/2006)
(2 cores)
PA-8800 (2004)
(2 cores, L2 data off-chip)

Core 2 Duo based processors (2006)
(2 cores)
SPARC64 VI (2007)
(2 cores)
SPARC64 VII (2008)
(4 cores)

9. Basic alternatives of macro-architectures of MC processors (5)

Design space of the macroarchitecture of MC processors (2-level hierarchy) (2)



M. c. on-chip

L2 private

L2 shared

Attaching both I/O and M. c. to L2 via IN2

Attaching M. c. to L2 via IN2 and I/O to IN1

Attaching I/O. to L2 via IN2 and M.c. to IN1

Attaching both I/O and M.c. via IN1

Opteron dual core (2 cores),(2005)

UltraSPARC IV (2 cores, L2 data off-chip), (2004)

Athlon 64 X2 (2005) (2 cores)

UltraSPARC T1 (2005) (8 cores)

Cell BE (2006) (8 cores)