

Intel's Core 2 family - TOCK lines II Nehalem to Haswell

Dezső Sima

Vers. 3.11

August 2018

Contents

- 1. Introduction
- 2. The Core 2 line
- 3. The Nehalem line
- 4. The Sandy Bridge line
- 5. The Haswell line
- 6. The Skylake line
- 7. The Kaby Lake line
- 8. The Kaby Lake Refresh line
- 9. The Coffee Lake line
- 10. The Cannon Lake line

3. The Nehalem line

- 3.1 Introduction to the 1. generation Nehalem line (Bloomfield)
- 3.2 Major innovations of the 1. gen. Nehalem line
- 3.3 Major innovations of the 2. gen. Nehalem line (Lynnfield)

3.1 Introduction to the 1. generation Nehalem line (Bloomfield)

3.1 Introduction to the 1. generation Nehalem line (Bloomfield) (1)

3.1 Introduction to the 1. generation Nehalem line (Bloomfield)

Developed at Hillsboro, Oregon, at the site where the Pentium 4 was designed.

→ Experiences with HT

→ Nehalem became a **multithreaded design**.

The design effort took about **five years and required thousands of engineers** (Ronak Singhal, lead architect of Nehalem) [37].

The **1. gen. Nehalem line targets DP servers**, yet its first implementation appeared in the desktop segment (Core i7-9xx (Bloomfield)) 4C in **11/2008**

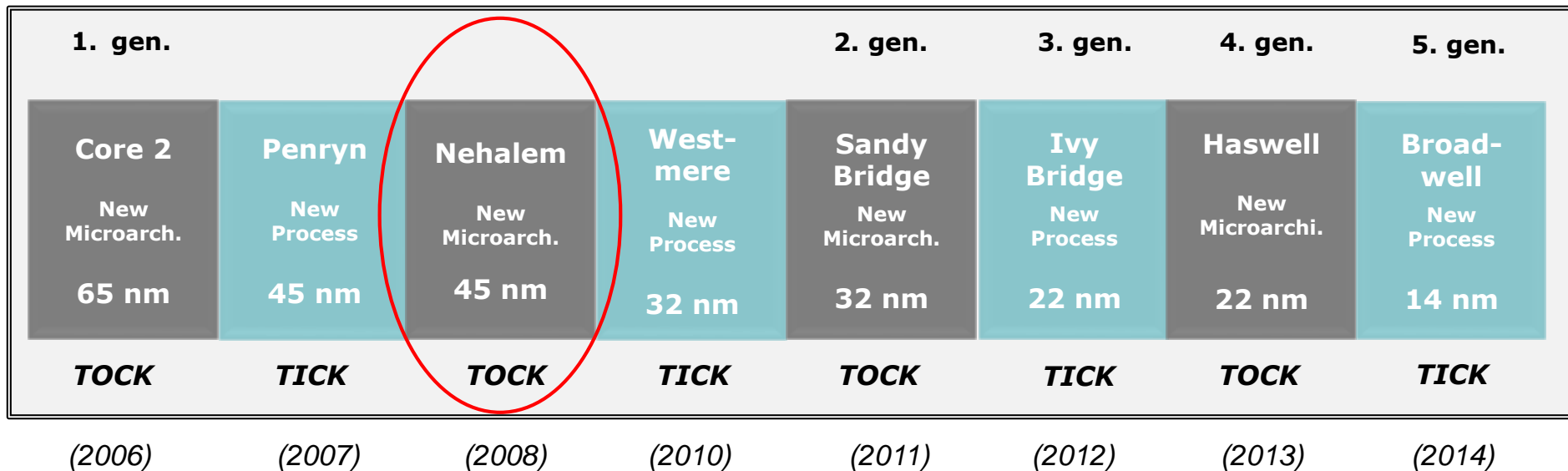
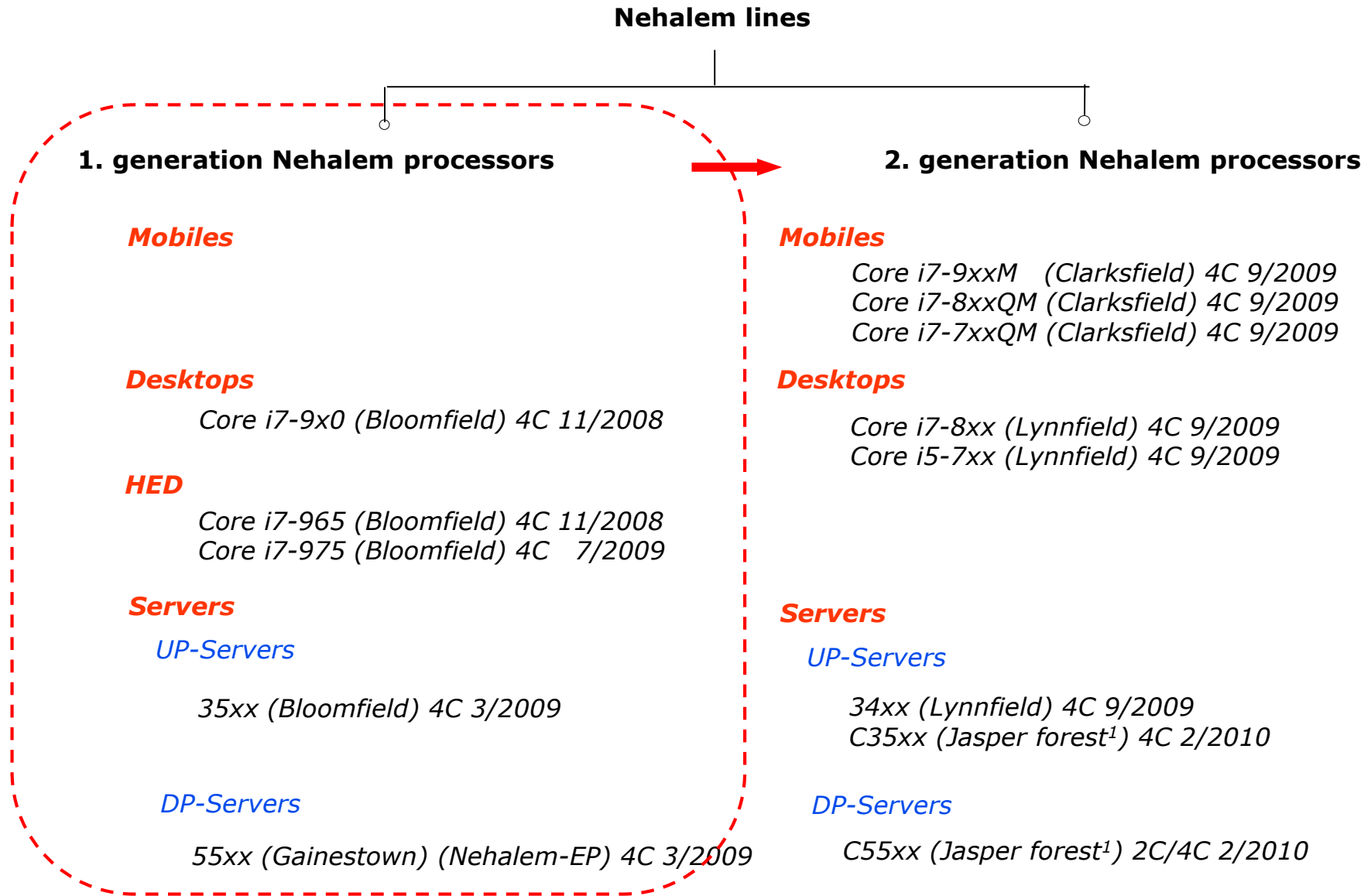


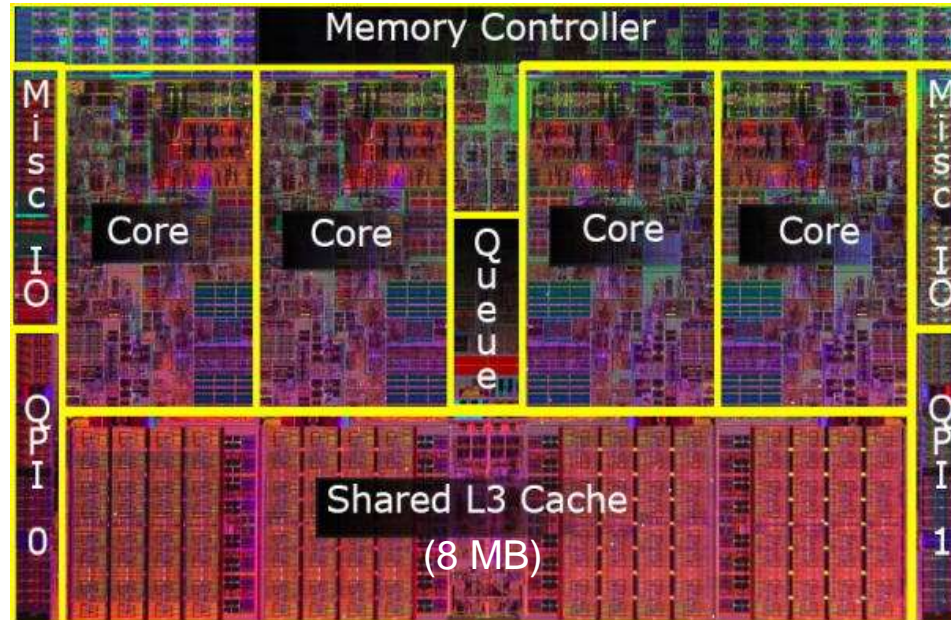
Figure : Intel's Tick-Tock development model (Based on [1])

3.1 Introduction to the 1. generation Nehalem line (Bloomfield) (2)



3.1 Introduction to the 1. generation Nehalem line (Bloomfield) (3)

Die shot of the 1. generation Nehalem desktop processor (Bloomfield) [45]



Die shot of the Bloomfield chip [45]

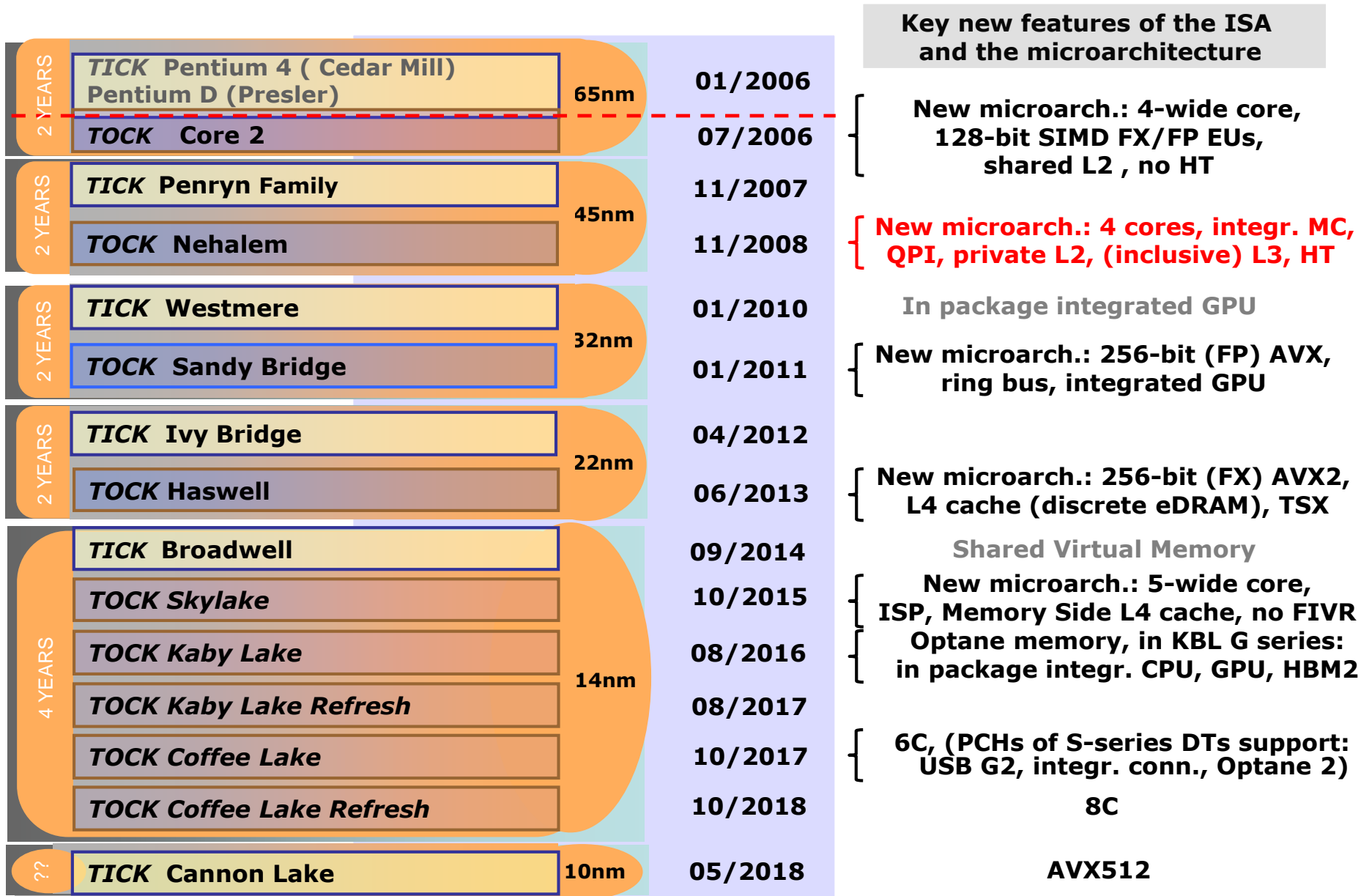
Note

- Both the desktop oriented Bloomfield chip and the DP server oriented Gainestown chip have the same layout.
- The Bloomfield die has two QPI bus controllers, in spite of the fact that they are not needed for the desktop part.

In the Bloomfield die one of the controllers is simply not activated [45], whereas both are active in the DP alternative (Gainestown).

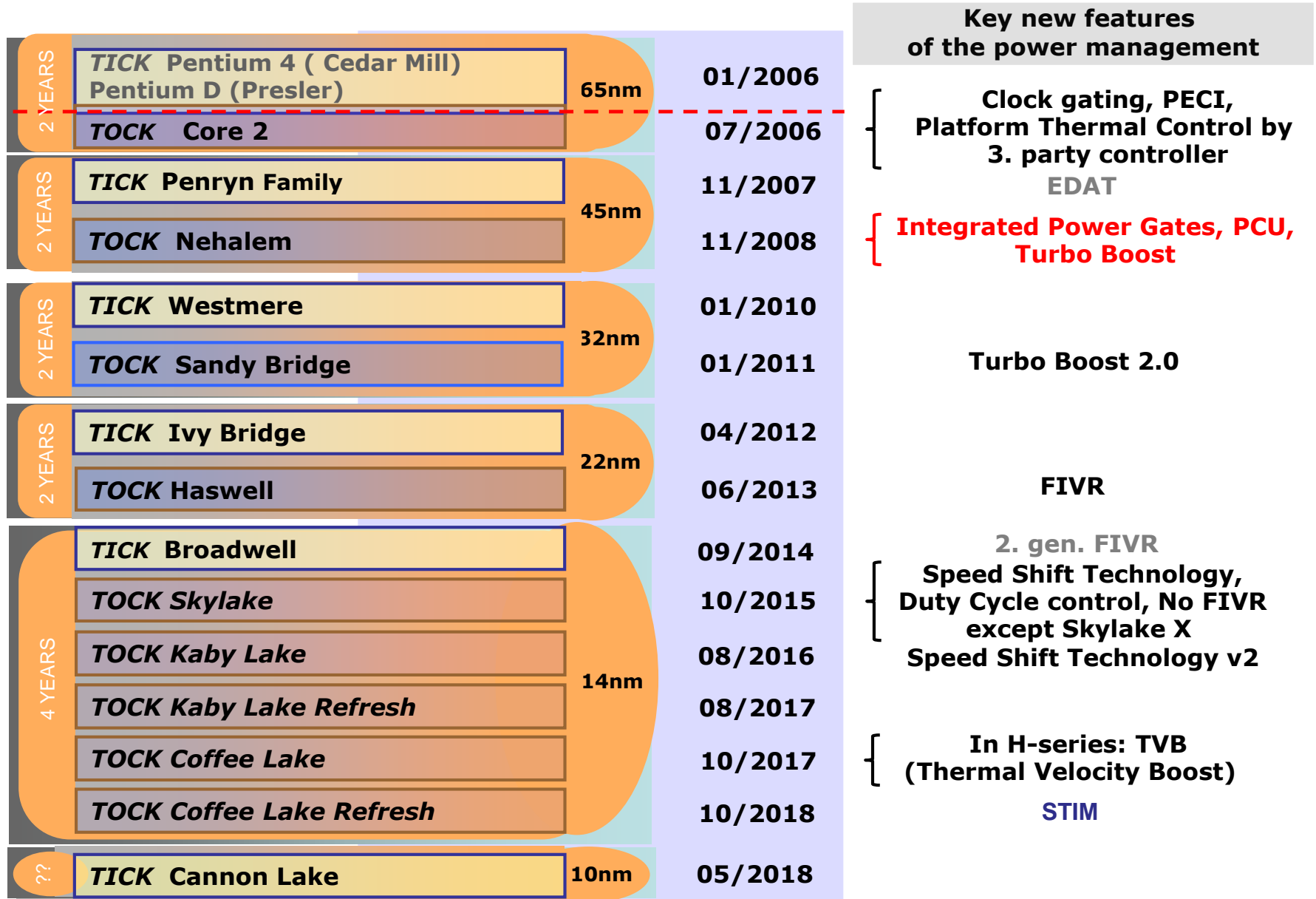
3.1 Introduction to the 1. generation Nehalem line (Bloomfield) (4)

The Nehalem line -1 (based on [3])



3.1 Introduction to the 1. generation Nehalem line (Bloomfield) (4B)

The Nehalem line -2 (based on [3])



3.2 Major innovations of the 1. generation Nehalem line

- 3.2.1 Integrated memory controller
- 3.2.2 QuickPath Interconnect bus (QPI)
- 3.2.3 New cache architecture
- 3.2.4 Simultaneous Multithreading
- 3.2.5 Enhanced power management
- 3.2.6 New socket

3.2 Major innovations of the 1. generation Nehalem line (1)

3.2 Major innovations of the 1. generation Nehalem line [54]

- **The Major incentive** for designing the microarchitecture of Nehalem: support of **4 cores**.
- 4 cores need however **twice as much bandwidth** as dual core processors, to maintain the per core memory bandwidth.
- Two memory channels used for dual core processors are more or less the limit attachable to the north bridge due to physical and electrical limitations.

Consequently, to provide enough bandwidth for 4 cores, a new memory design was necessary.

Major innovations of the 1. generation 4-core Nehalem line

- **Integrated memory controller**
(Section 3.2.1)
- **QuickPath Interconnect bus (QPI)**
(Section 3.2.2)
- **New cache architecture**
(Section 3.2.3)
- **Simultaneous Multithreading (SMT)**
(Section 3.2.4)
- **SSE 4.2 ISA extension**
(Not detailed)
- **Enhanced power management**
(Section 3.2.5)
- **Advanced virtualization**
(Not detailed)
- **New socket**
(Section 3.2.6)

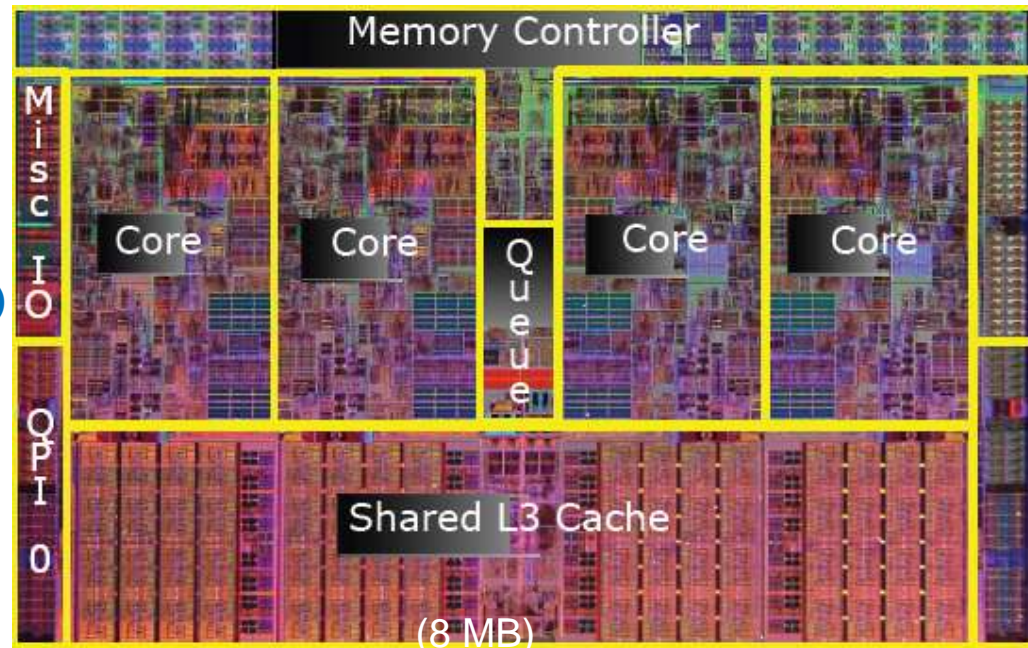


Figure 3.2.1: Die photo of the Bloomfield/Gainestown chip *

3.2.1 Integrated memory controller (1)

3.2.1 Integrated memory controller

- Traditional system architectures, as shown below for the Core 2 Duo processor, can implement **not more than two high speed memory channels connected to the MCH due to electrical and physical constraints, to be discussed in Chapter on Intel's Servers.**
- Two memory channels can however, provide enough bandwidth **only for up to dual core processors.**

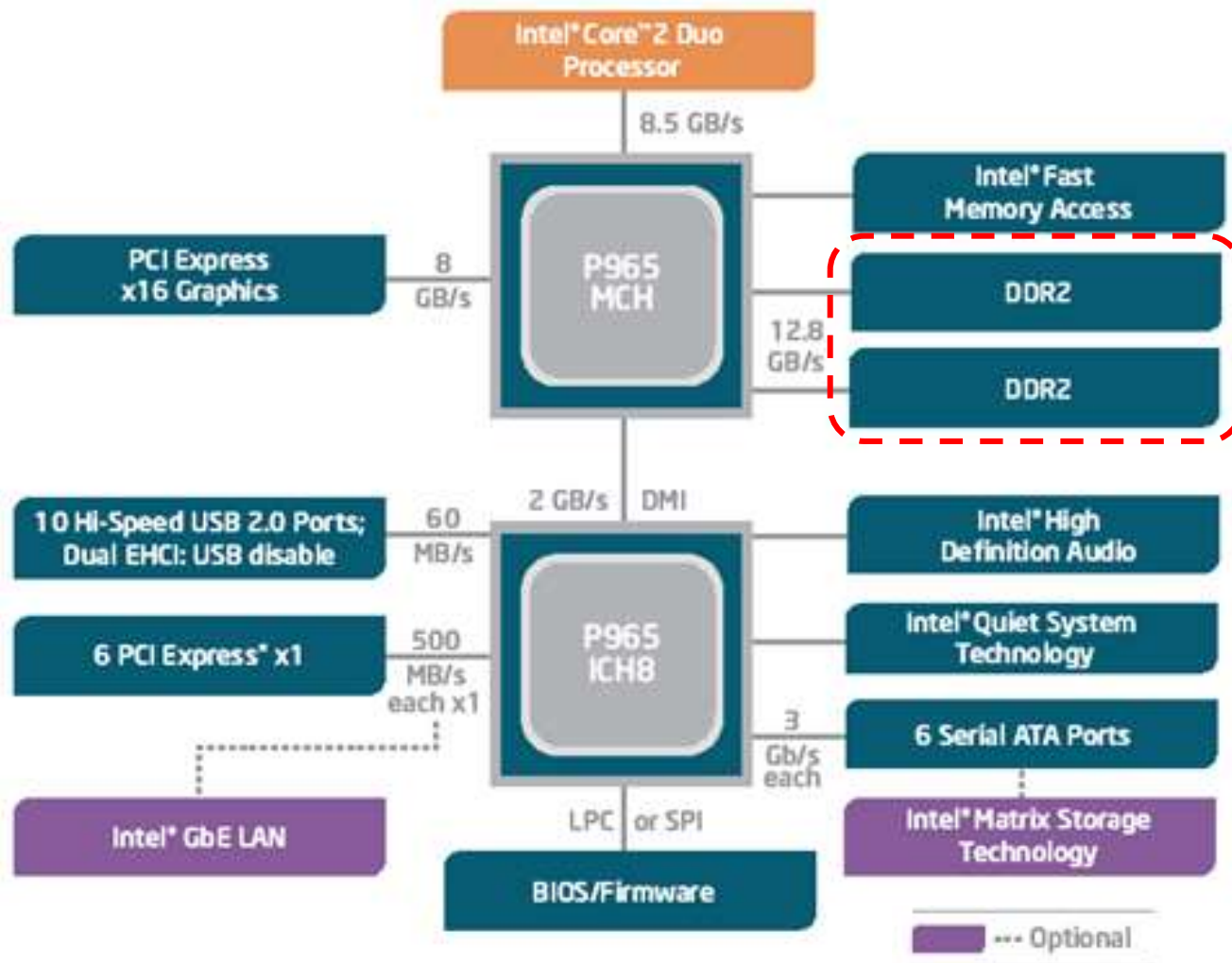


Figure: Core 2 Duo based platform [166]

3.2.1 Integrated memory controller (2)

The need for integrated memory controller in a dual processor QC Nehalem platform

n cores \rightarrow n times higher memory bandwidth need per processor
New design for attaching memory: placing memory controllers on the dies

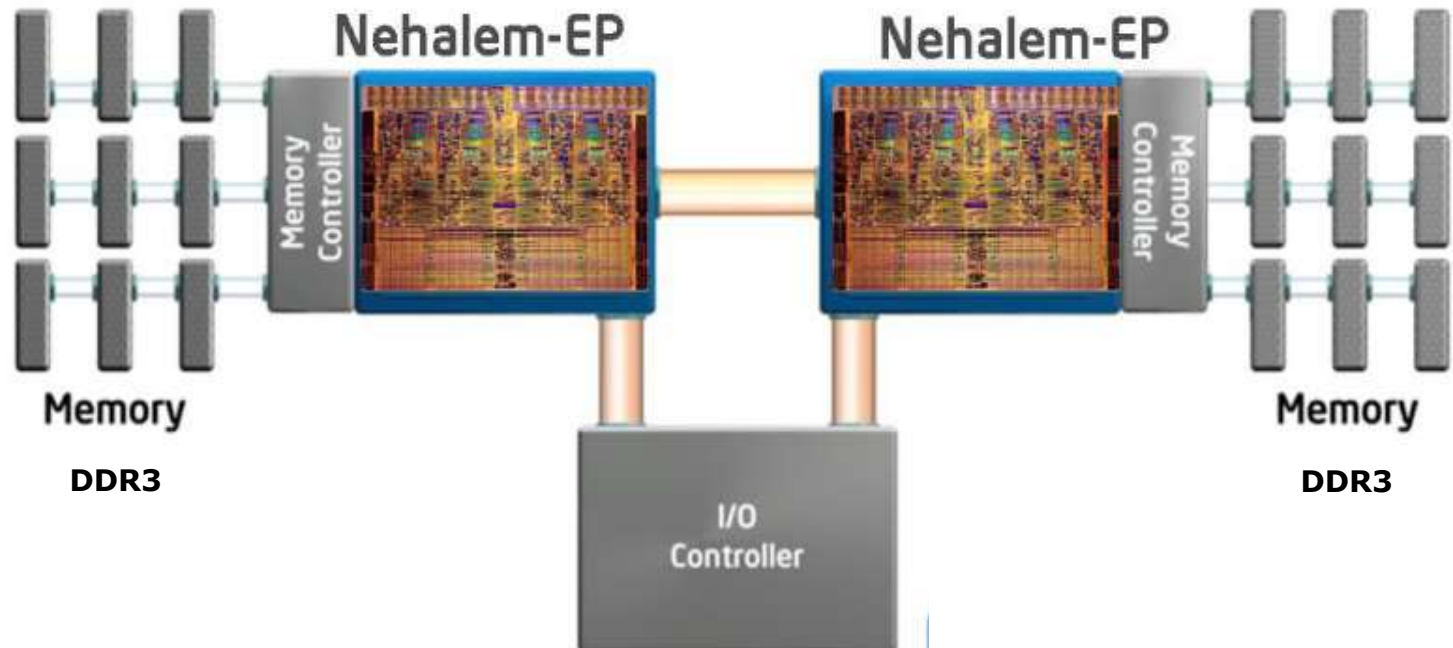
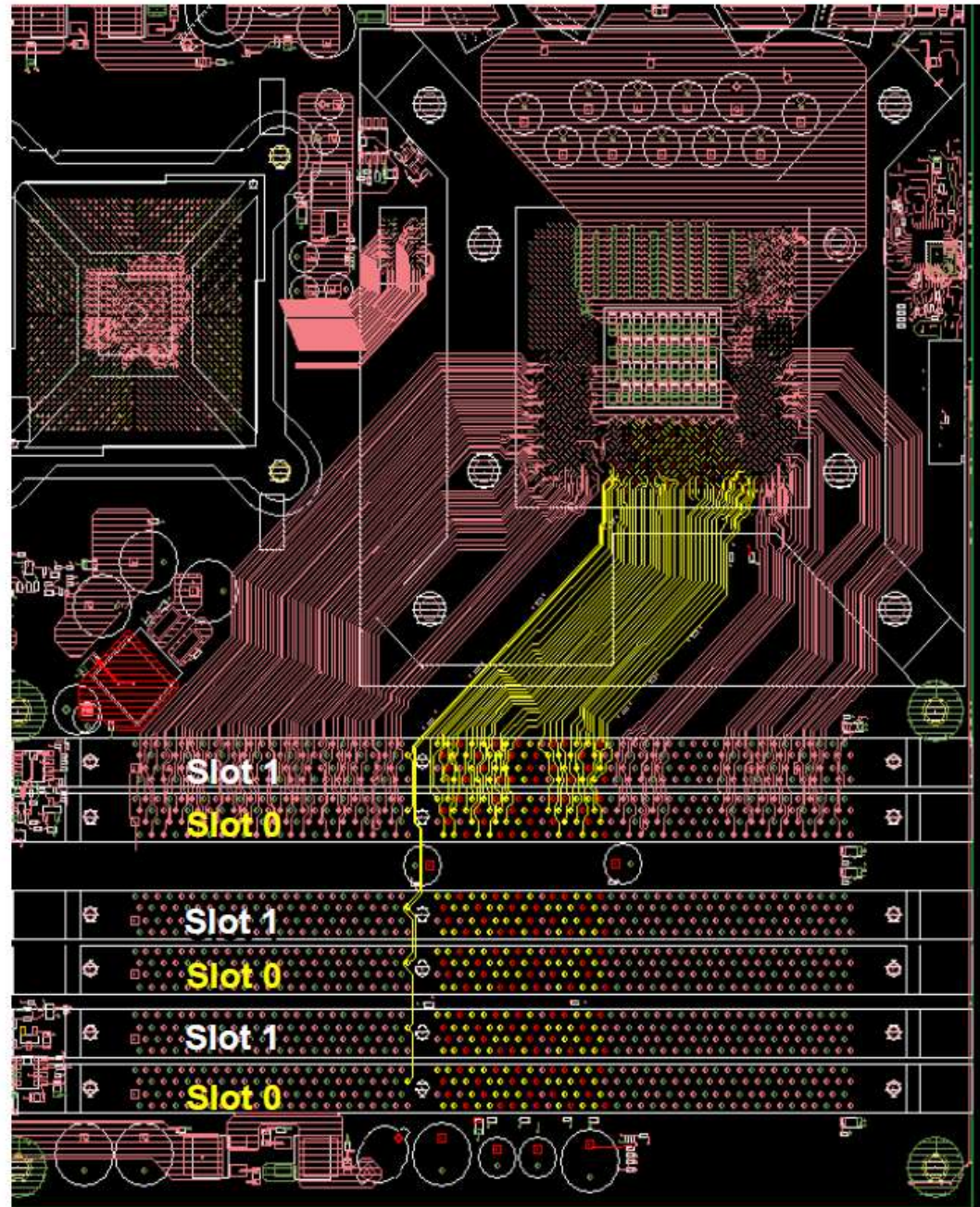


Figure 3.2.1.1: Integrated memory controller of Nehalem [33]

3.2.1 Integrated memory controller (2b)

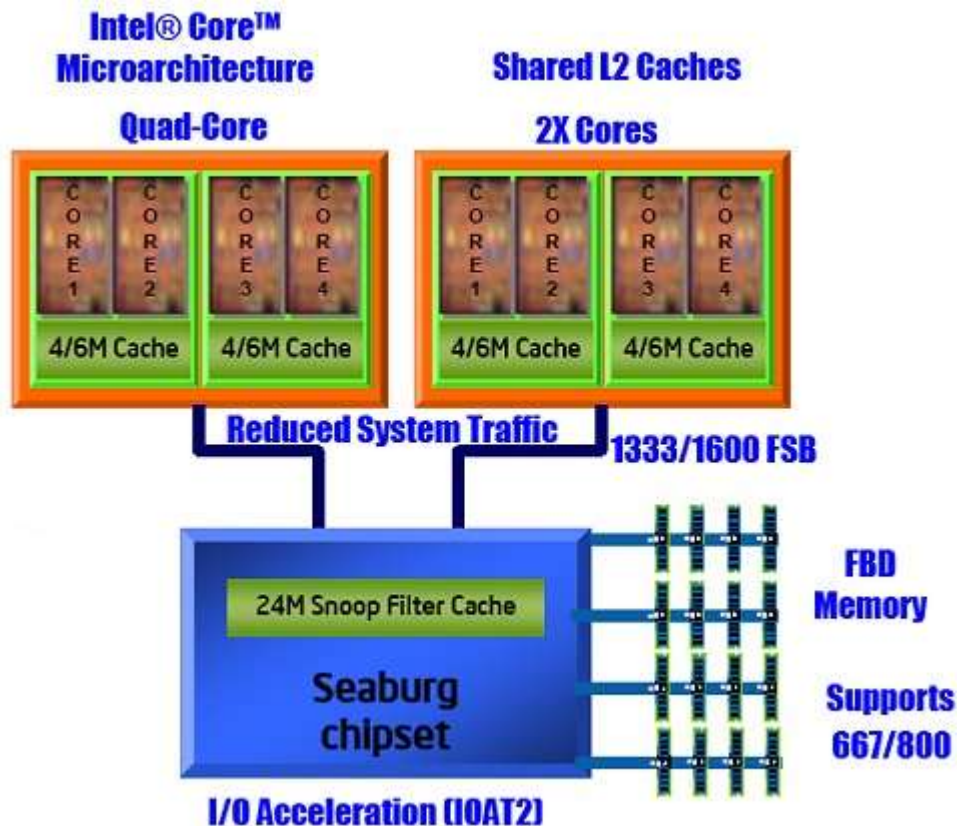
Connecting 3 DDR3 memory channels to the processor socket [242]



3.2.1 Integrated memory controller (2c)

Alternative solution: Connecting memory via (connected via low-line count serial differential interfaces)

(Harpertown (2x2 cores, 45 nm Penryn) based DP server processor [277])



Harpertown
(45 nm, 2 chips
in the same package)

FB-DIMM memory
(connected via
low-line count
serial differential
interfaces)

Benefits and drawback of integrated memory controllers

Benefits

Low memory access latency

→ important for memory intensive apps.

Drawback of integrated memory controllers

- Processor becomes memory technology dependent
- For an enhanced memory solution (e.g. for increased memory speed) a new processor modification is needed.

3.2.1 Integrated memory controller (4)

Non Uniform Memory Access (NUMA) architectures

It is a **consequence of using integrated memory controllers** in case of **multi-socket servers**

Local memory access **Remote memory access**

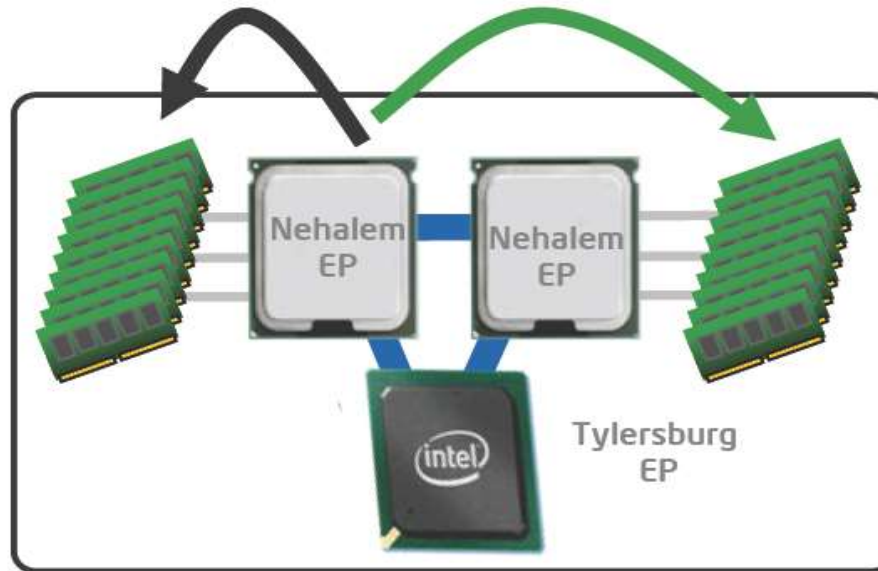


Figure 3.2.1.2: Non Uniform Memory Access (NUMA) in multi-socket servers [1]

- Advanced multi-socket platforms use NUMA
- **Remote memory access latency** $\sim 1.7 \times$ longer than **local memory access latency**
- **Demands a fast processor-to-processor interconnection to relay memory traffic (QPI)**
- Operating systems have to modify memory allocation strategies + related APIs

3.2.1 Integrated memory controller (4b)

Remark: Classification of multiprocessor server platforms according to their memory architecture

Multiprocessor server platforms classified according to their memory architecture

SMPs

(Symmetrical MultiProcessor)

Multiprocessors (Multi socket system) with **Uniform Memory Access (UMA)**

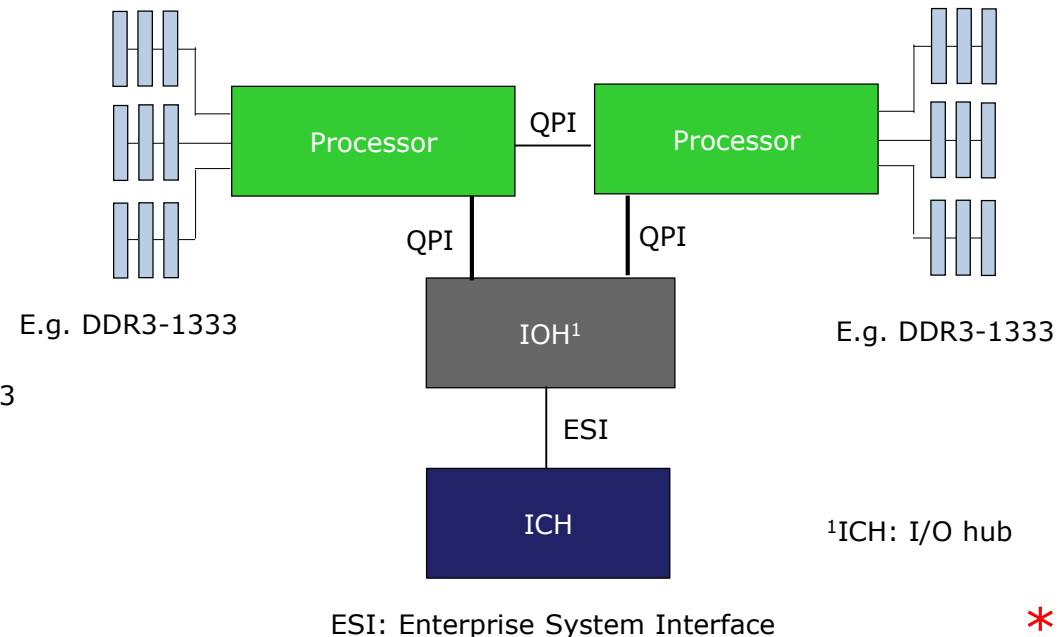
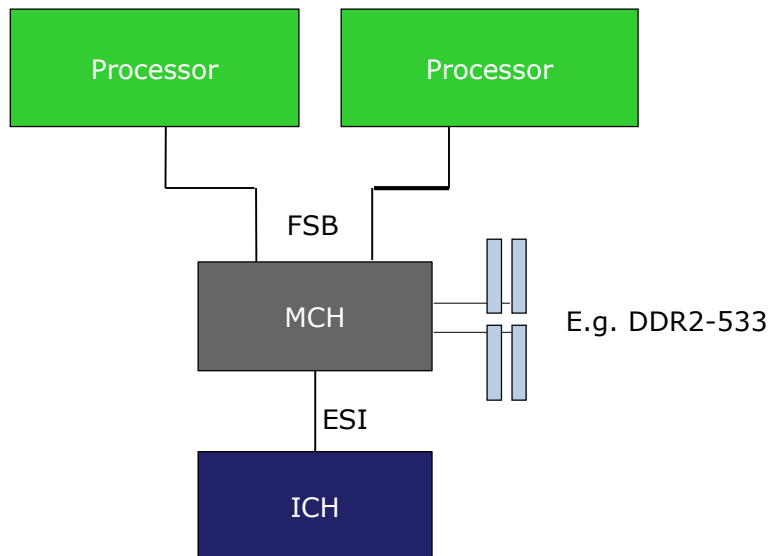
All processors access main memory by the same mechanism, (e.g. by individual FSBs and an MCH).

NUMAs

Multiprocessors (Multi socket system) with **Non-Uniform Memory Access**

Each processor is allocated a part of the main memory (with the related memory space), called the **local memory**, whereas the rest is considered as the **remote memory**.

Typical examples



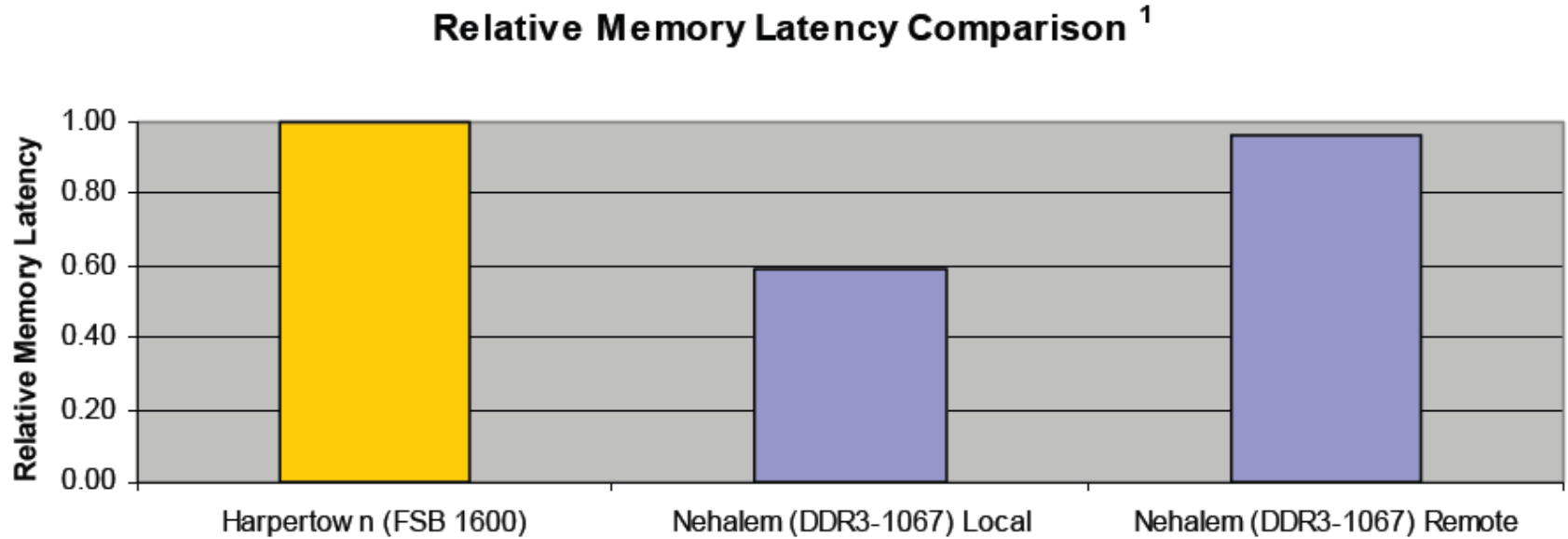
¹ICH: I/O hub

ESI: Enterprise System Interface



3.2.1 Integrated memory controller (5)

Memory latency comparison: Nehalem vs Penryn [1]



Harpertown: Quad-Core Penryn based server (Xeon 5400 series)

3.2.1 Integrated memory controller (6)

Remark

Intel's **Timna** – a forerunner to integrated memory controllers [34]

Timna (announced in 1999, due to 2H 2000, cancelled in Sept. 2000)

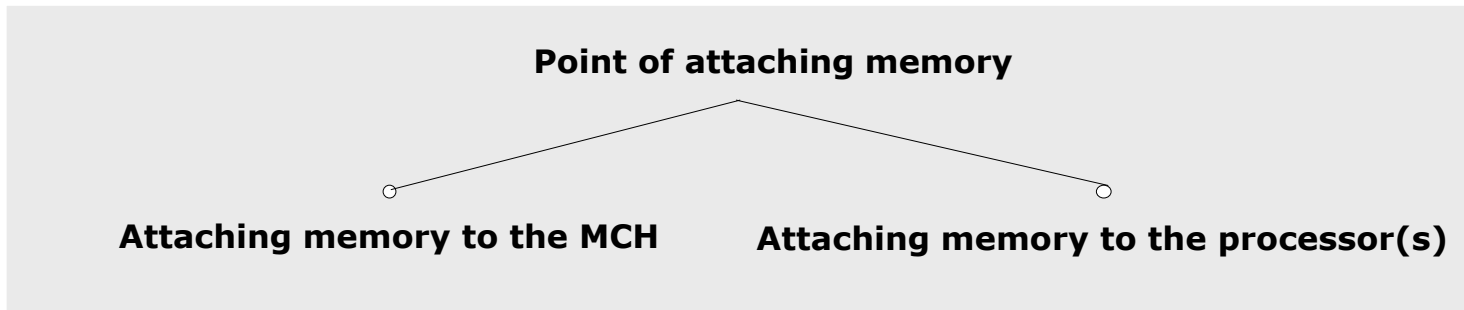
- Developed in Intel's Haifa Design and Development Center.
- Low cost microprocessor with **integrated graphics and memory controller** (for Rambus DRAMs).
- Due to design problems and lack of interest from many vendors, Intel finally cancelled Timna in Sept. 2000.



Figure 3.2.1.4: The low cost (<600 \$) Timna PC [40]

3.2.1 Integrated memory controller (7)

Point of attaching memory



Examples

<i>UltraSPARC II (1C) (~1997)</i>		<i>UltraSPARC III (2001) and all subsequent Sun lines</i>
<i>AMD's K7 lines (1C) (1999-2003)</i>		<i>Opteron server lines (2C) (2003) and all subsequent AMD lines</i>
<i>POWER4 (2C) (2001)</i>		<i>POWER5 (2C) (2005) and subsequent POWER families</i>
<i>PA-8800 (2004)</i>		
<i>PA-8900 (2005)</i>		
<i>and all previous PA lines</i>		
<i>Core 2 Duo line (2C) (2006) and all preceding Intel lines</i>		<i>Nehalem lines (4) (2008) and all subsequent Intel lines</i>
<i>Core 2 Quad line (2x2C) (2006/2007)</i>		
<i>Penryn line (2x2C) (2008)</i>		
<i>Montecito (2C) (2006)</i>		<i>Tukwila (4C) (2010)</i>

3.2.1 Integrated memory controller (8)

Main features of the dual processor QC Nehalem platform

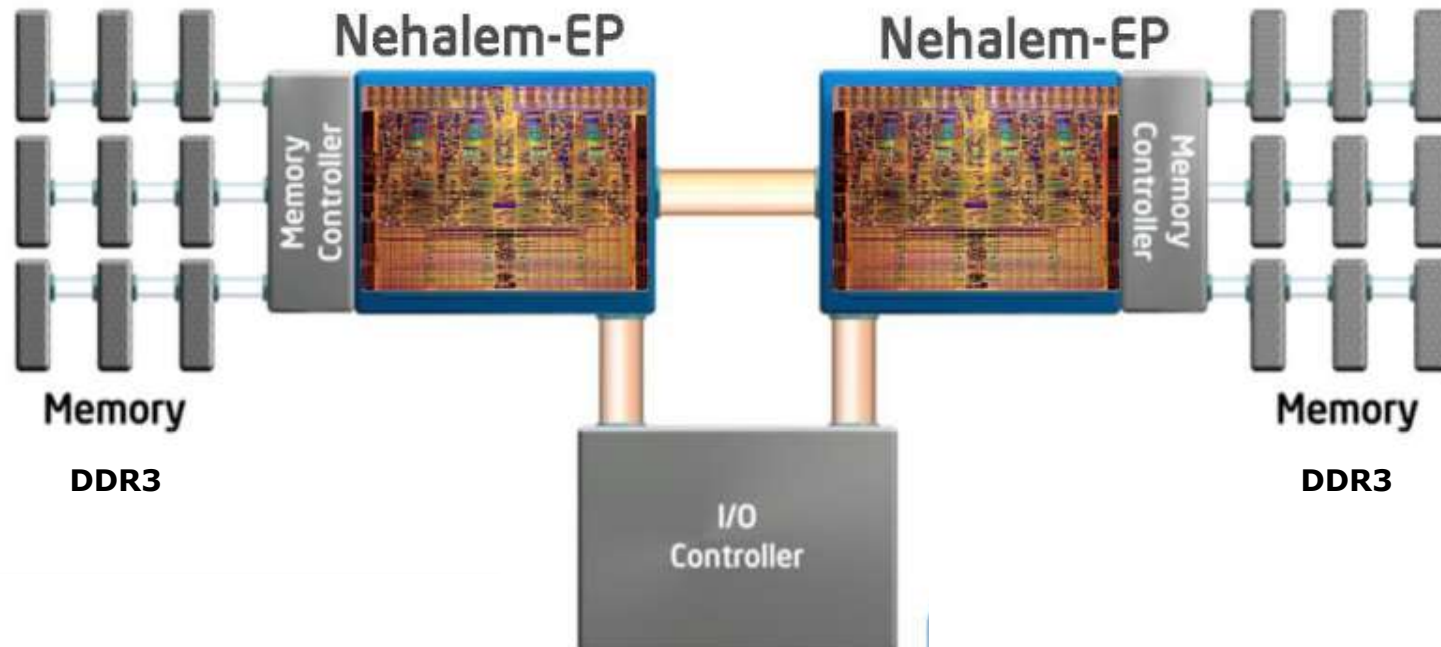


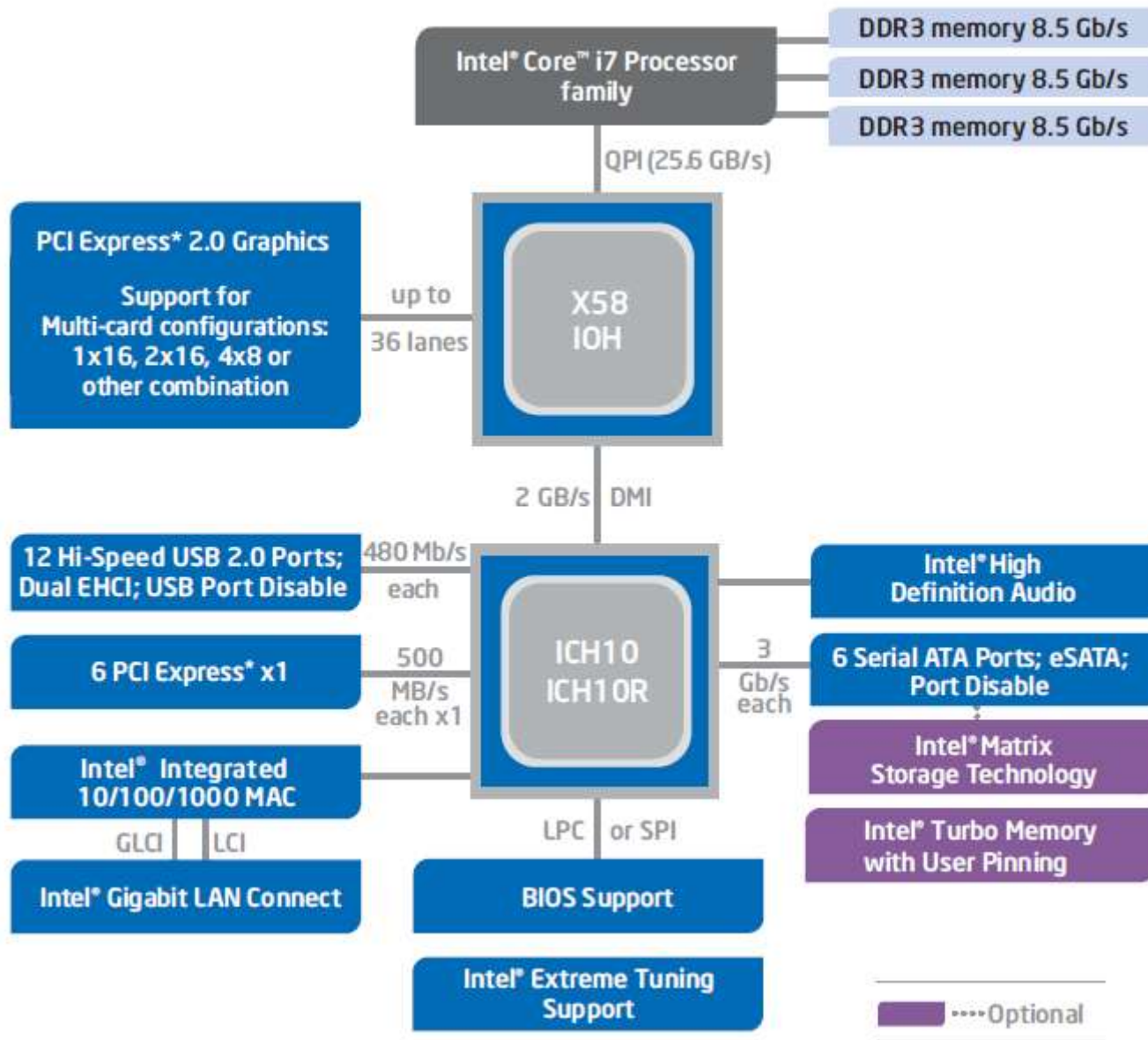
Figure 3.2.1.1: Integrated memory controller of Nehalem [33]

- 3 channels per socket
- Up to 3 DIMMs per channel (impl. dependent)
- DDR3-800, 1066, 1333
- Supports both RDIMMs and UDIMMs (impl. dependent)

Nehalem-EP (Efficient Performance):
Designation of the [server line](#)

3.2.1 Integrated memory controller (9)

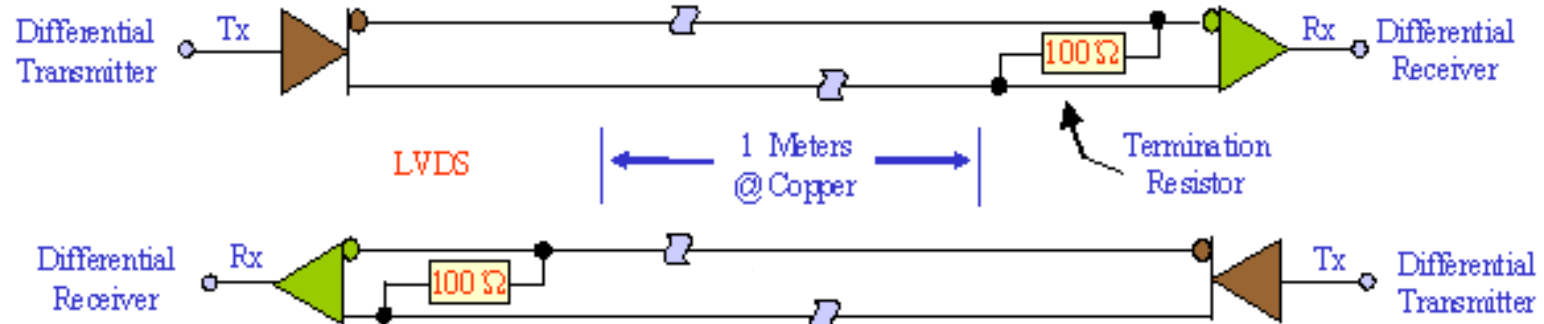
1. generation Nehalem (called Bloomfield)-based desktop platform [249]



3.2.2 QuickPath Interconnect bus (QPI) -1

- Its debut is **strongly motivated by the introduction of integrated memory controllers**, since in multiprocessors accessing data held remotely (to a given processor) needs a high-speed processor-to-processor interconnect.
- Such an interconnect will be implemented as a **serial, differential point-to-point bus**, called the **Quick Path Interconnect (QPI) bus**, similarly to AMD's HyperTransport bus, used to connect processors to processors or processors to north bridges.
- Formerly, the QPI bus was designated as the **Common System Interface bus (CSI bus)**.

Principle of differential interconnections [170]



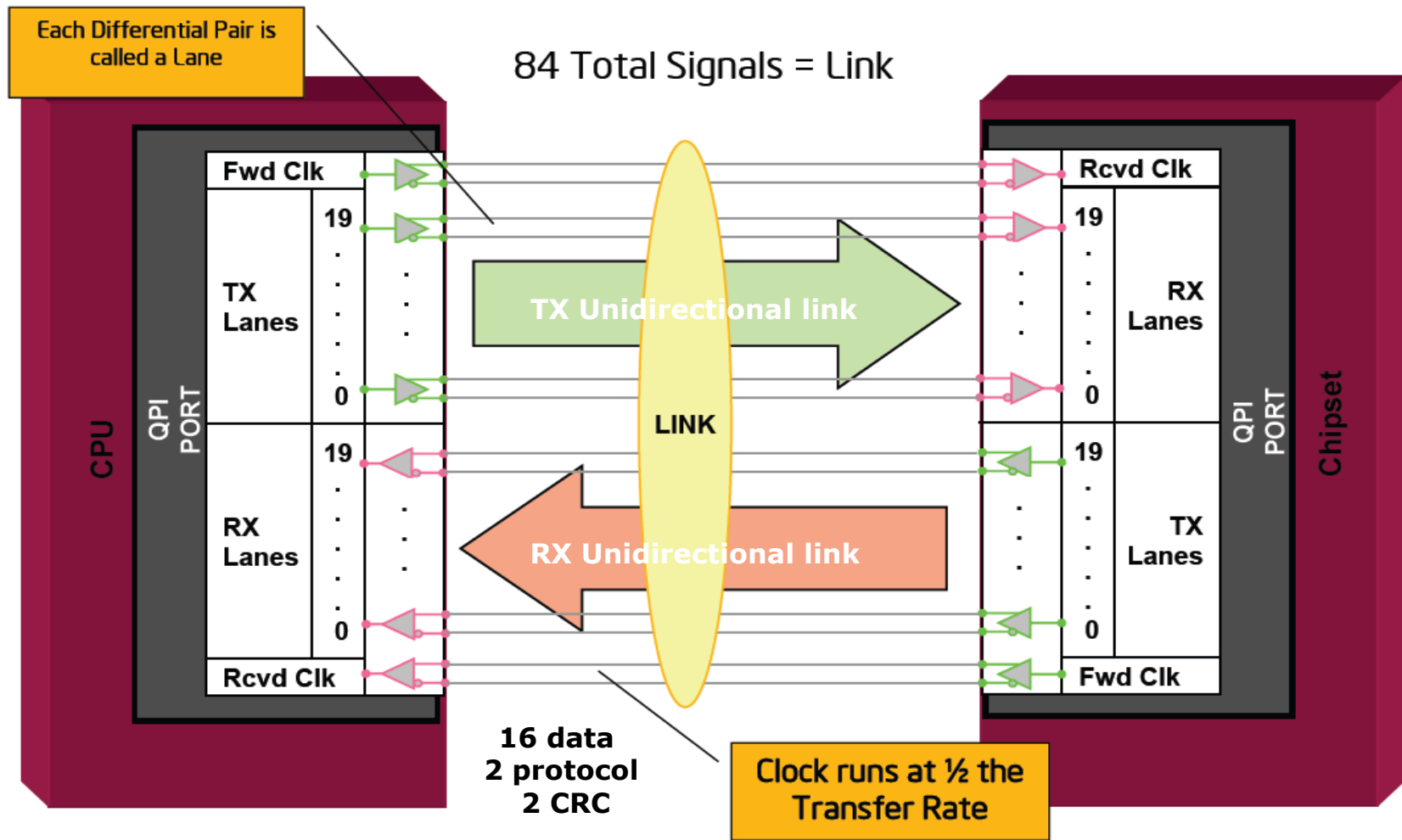
3.2.2 QuickPath Interconnect bus (QPI)- 2

- It consists of 2 **unidirectional links**, one in each directions, called the **TX** and **RX** (T for Transmit, R for Receive).

3.2.2 QuickPath Interconnect bus (QPI) (4)

Signals of the QuickPath Interconnect bus (QPI bus) [22]

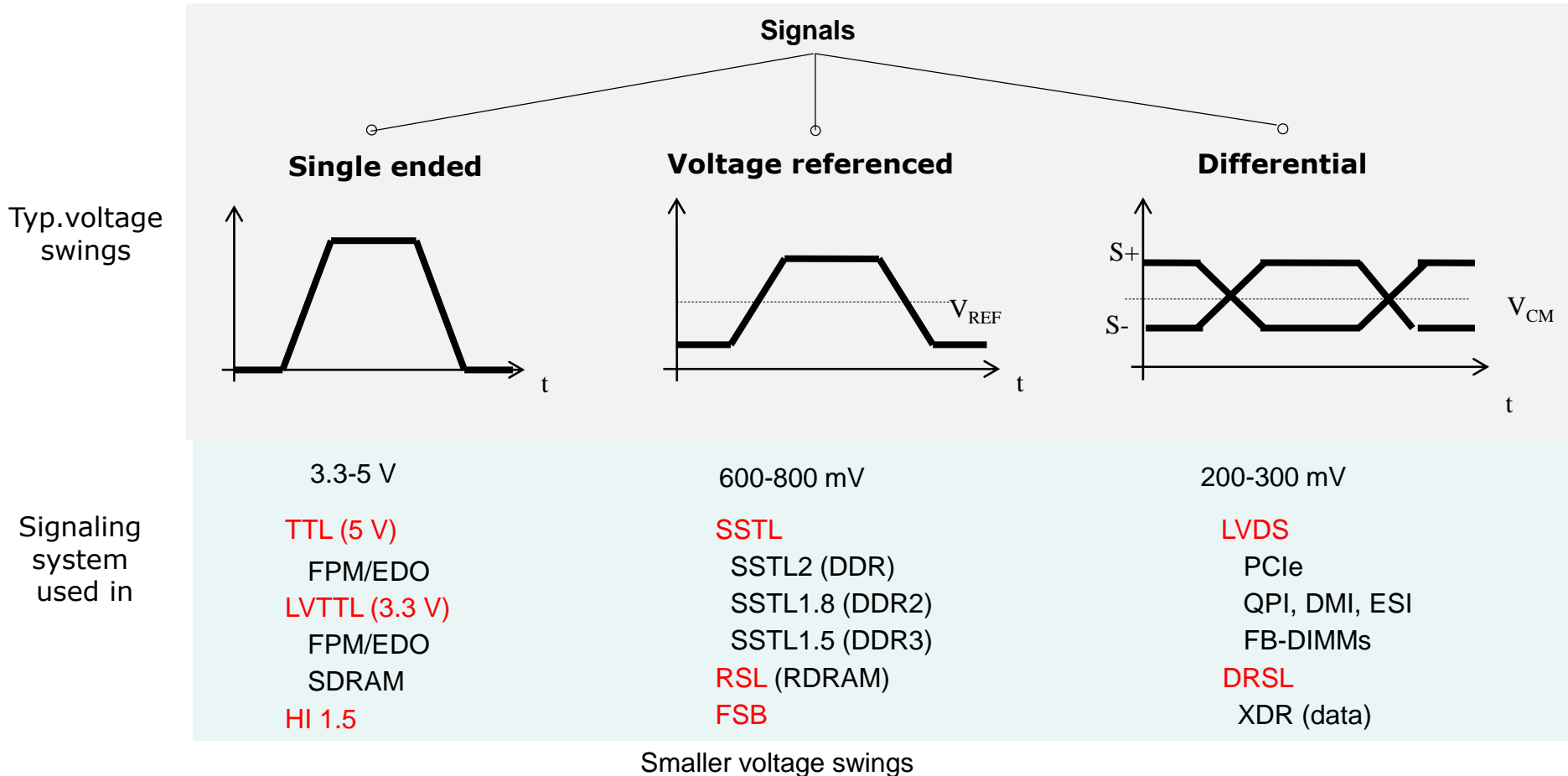
- Each unidirectional link comprises 20 data lanes and a clock lane, with each lane consisting of a pair of differential signals.



(Lane: Vonalpár)

(DDR data transfer)

Signaling systems used in buses

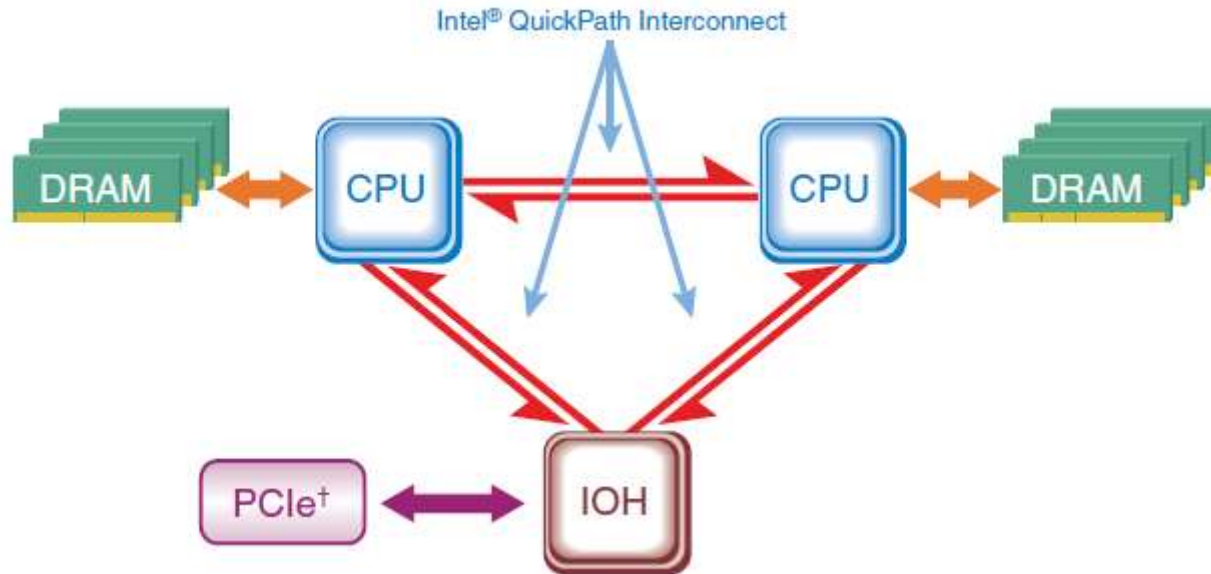


LVDS: Low Voltage Differential Signaling
 (D)RSL: (Differential) Rambus Signaling Level
 V_{CM} : Common Mode Voltage

LVTTTL: Low Voltage TTL
 SSTL: Stub Series Terminated Logic
 V_{REF} : Reference Voltage

3.2.2 QuickPath Interconnect bus (QPI) (6)

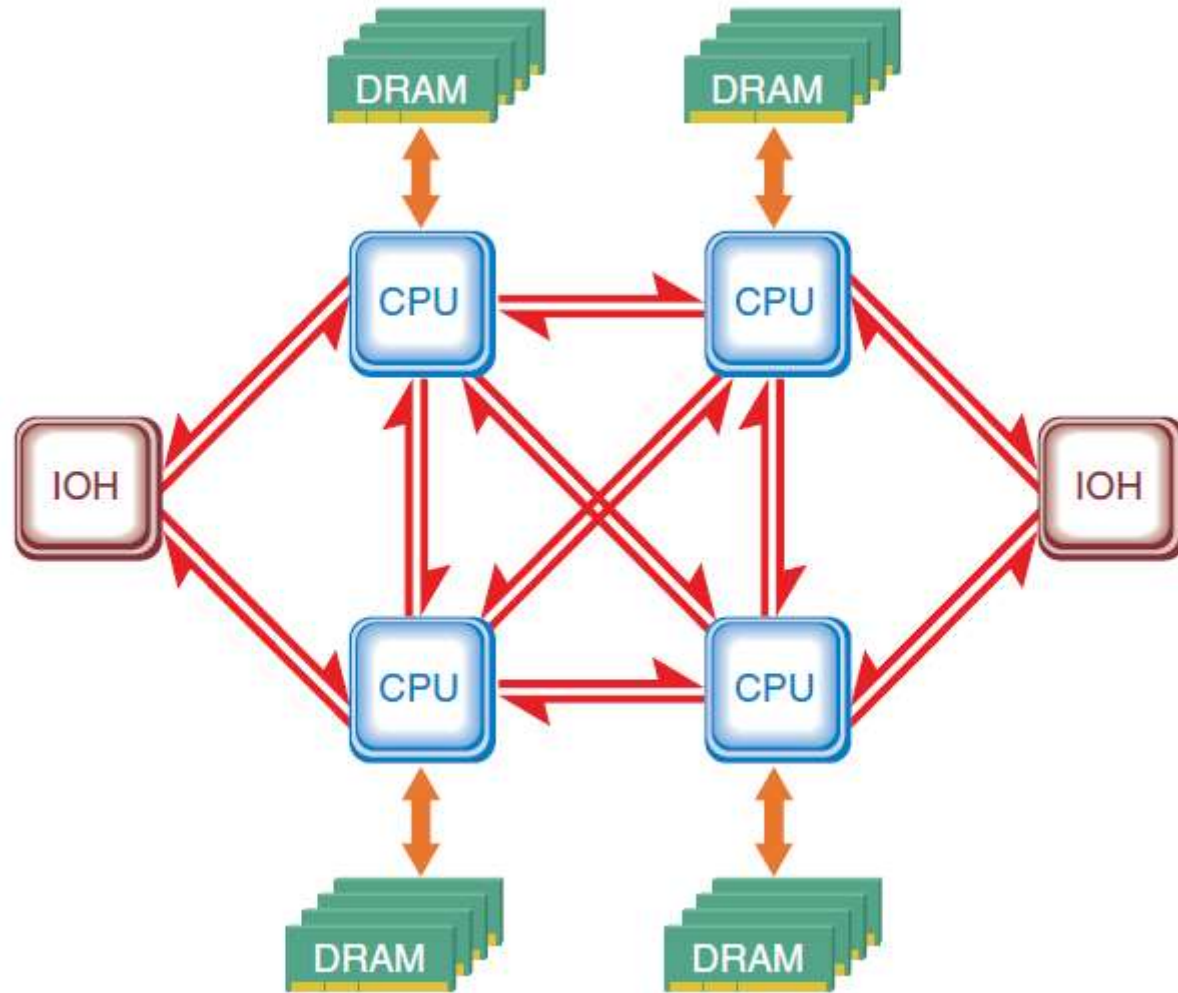
QPI based DP server architecture [169] -1



Note

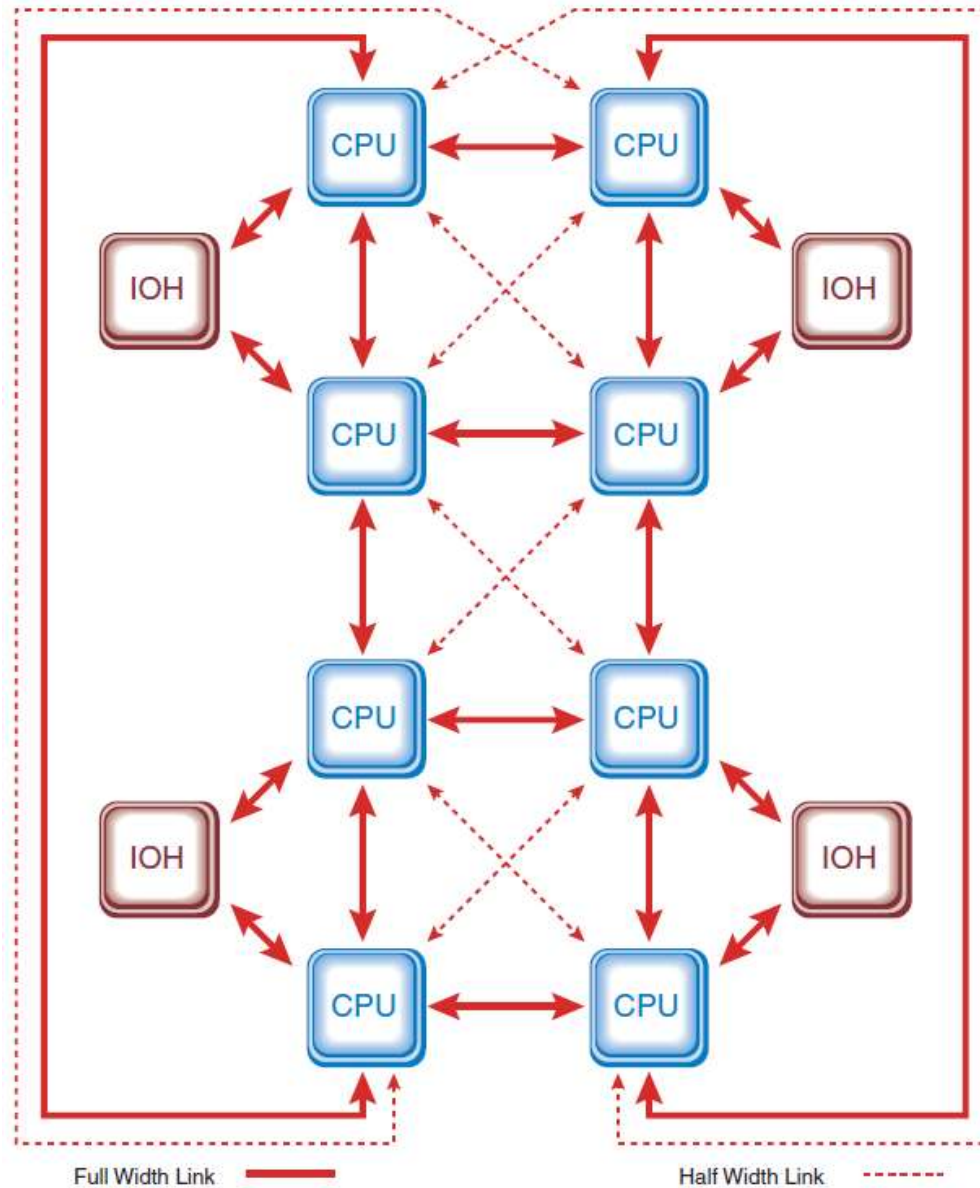
First generation Nehalem (Bloomfield) supports only DP configurations.

QPI based MP server architecture [169] -2



3.2.2 QuickPath Interconnect bus (QPI) (8)

QPI based 8-processor system architecture [169] -3



3.2.2 QuickPath Interconnect bus (QPI) (9)

Contrasting the QPI with the FSB and other serial buses

Fastest FSB

Parallel, 8 Byte data width, QDR, up to 400 MHz clock, voltage ref. signaling 12.8 GB/s → data rate

Serial links

Unidirectional point-to-point links, 2 Byte data width, DDR data rate, differential signaling

QPI	Base clock	Platforms	Data rate (up to) (in each dir.)	Year
QPI	3.2 GHz	Nehalem (server/desktop)	12.8 GB/s	2008
QPI 1.1	4.0 GHz	Sandy Bridge EN/EP Ivy Bridge-EN/EP/EX Westmere EN/EP/EX	16.0 GB/s	2010-14
QPI 1.1	4.8 GHz	Haswell EN/EP/EX Broadwell EN/EP/EX	19.2 GB/s	2014-16

UPI	Base clock	Platforms	Data rate (up to) (in each dir.)	Year
UPI	5.2 GHz	Skylake-SP	20.8 GB/s	2017

HT	Base clock	Platforms (first implemented in)	Data rate (up to) (in each dir.)	Year
HT 1.0	0.8 GHz	K8-based mobile Athlon 64/Opteron	3.2 GB/s	2003
HT 2.0	1.0 GHz	K8-based Athlon 64 desktop	4.0 GB/s	2004
HT 3.0	2.6 GHz	K10.5-based Phenom X4 desktop	8.0 GB/s	2007
HT 3.1	3.2 GHz	K10.5-based Magny Course server	12.8 GB/s	2010

3.2.3 New cache architecture

- In multiprocessors with **NUMA architectures remote memory accesses** have long access times, this strengthens the need for an enhanced cache system.
- The cache system can be enhanced by introducing a **three level cache system**, enabled by the 45 nm technology used.

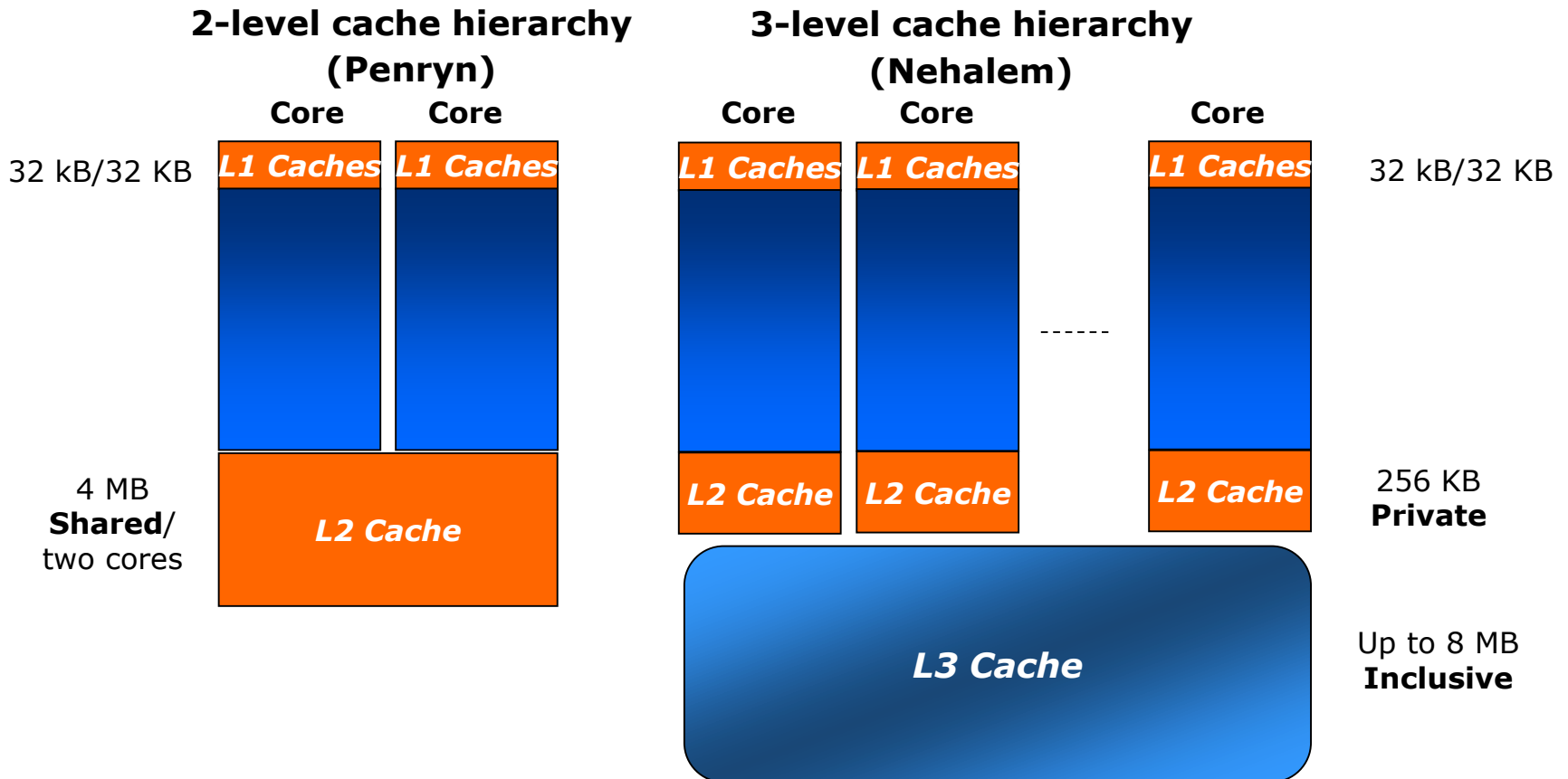


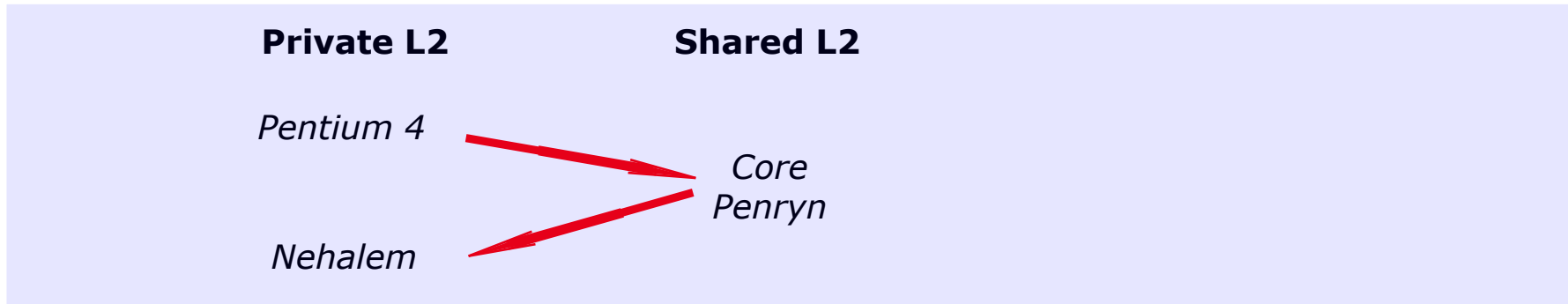
Figure 3.2.3.1: The 3-level cache architecture of Nehalem (based on [1])

Key features of the new 3-level cache architecture

- a) Using private L2 caches
- b) Changed L2 cache size
- c) Use of an inclusive L3 cache

a) Using private L2 caches

- The L2 cache is **private** again rather than shared as in the Core and Penryn processors



Assumed reason for returning to the private scheme

Private caches allow a more effective hardware prefetching than shared ones, since

- Hardware prefetchers look for memory access patterns.
- Private L2 caches have more easily detectable memory access patterns than shared L2 caches.

Remark

The POWER family had the same evolution path as above

Private L2

Shared L2

POWER4



POWER5

POWER6



b) Changed L2 cache sizes

- Without an L3 cache the optimum L2 cache size is the maximum L2 size feasible on the die.
- With an L3 cache available the optimum L2 size becomes about ¼ or ½ MB in the systems discussed.

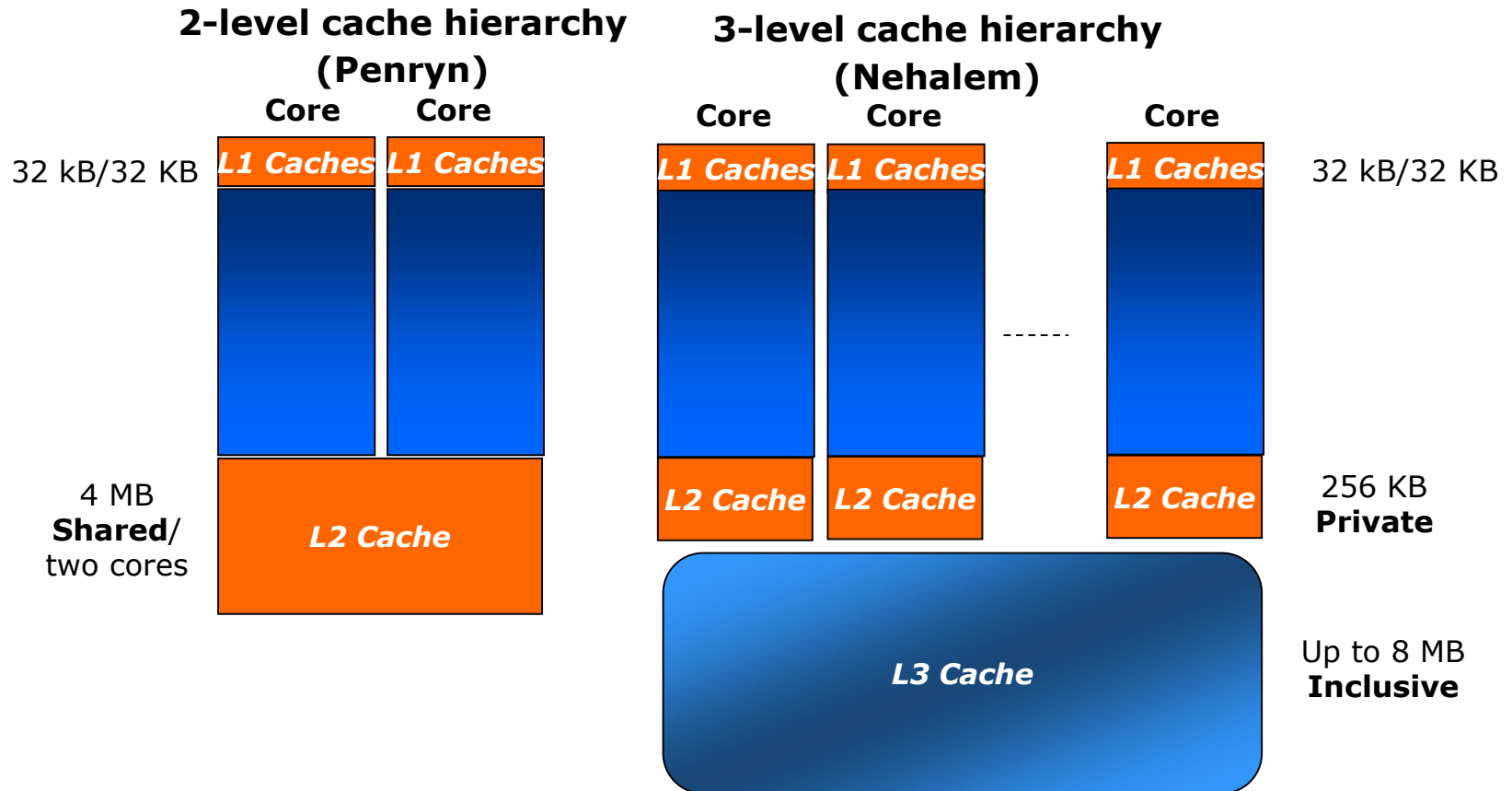


Figure 3.2.3.1: The 3-level cache architecture of Nehalem (based on [1])

Remark

The optimum cache size provides the highest system performance, since on the one side higher cache sizes lower the rate of cache misses on the other increase the cache access time. *

c) Use of an inclusive L3 cache

- The L3 cache is **inclusive** rather than exclusive

like in a number of competing designs, such as UltraSPARC IV+ (2005), POWER5 (2005), POWER6 (2007), POWER7 (2010), POWER8 (2014), AMD's K10-based processors (2007).

(An inclusive L3 cache includes the L2 cache content.)

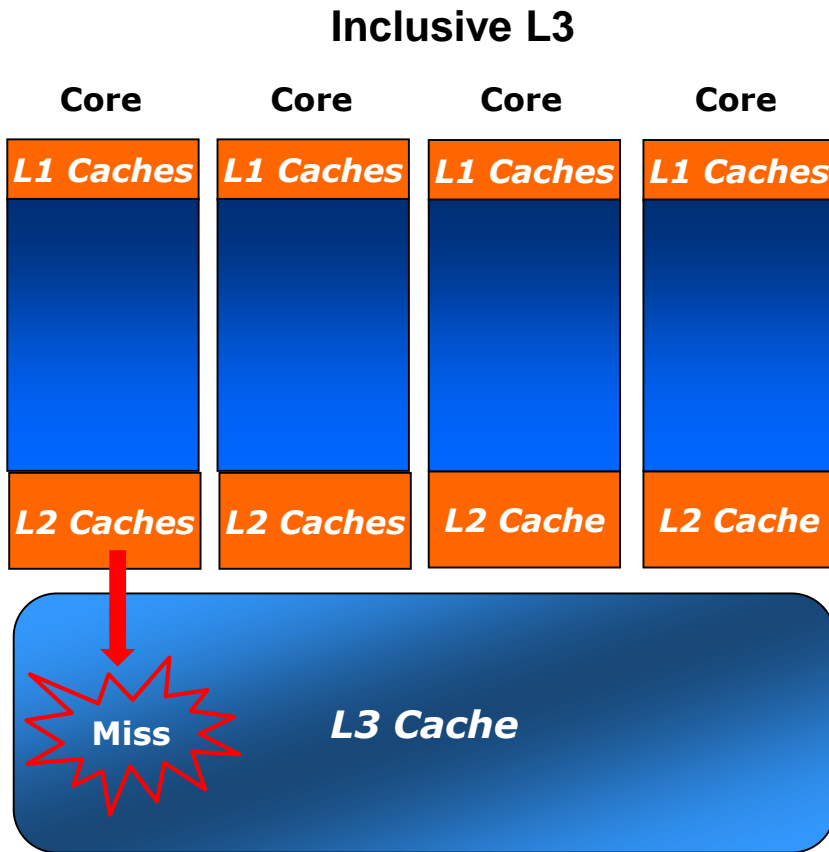
Intel's argumentation for inclusive caches [38]

Inclusive L3 caches prevent L2 snoop traffic for L3 cache misses since

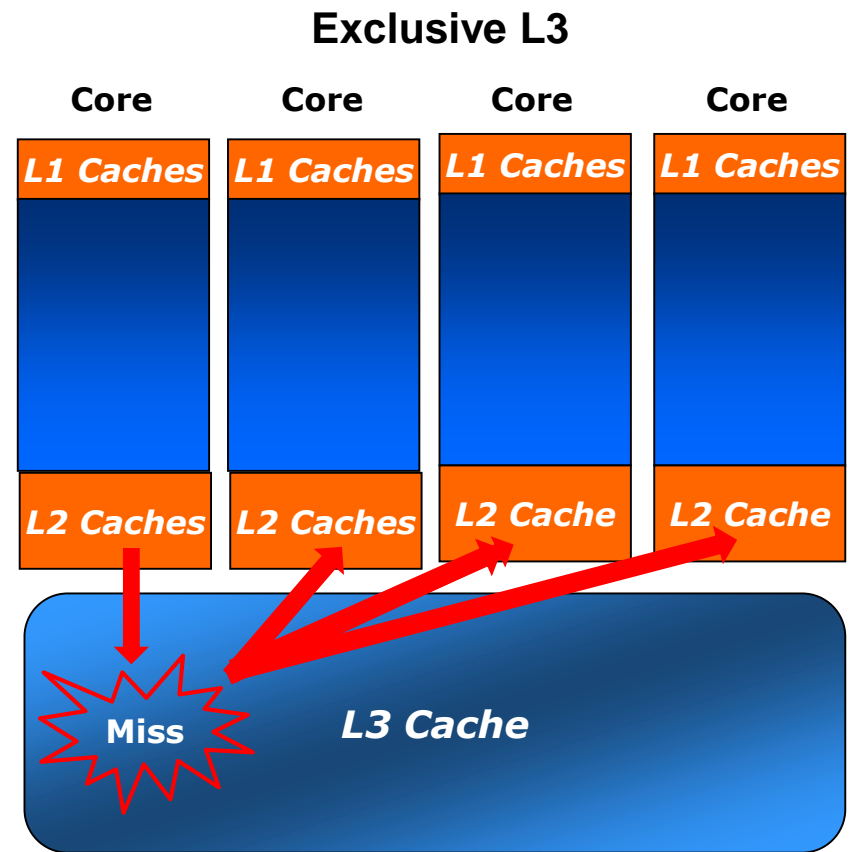
- with inclusive L3 caches an L3 cache miss means that the referenced data doesn't exist in any core's L2 caches, thus no L2 snooping is needed.
- By contrast, with exclusive L3 caches the referenced data may exist in any of the L2 caches, thus L2 snooping is required,

as indicated in the next Figure.

Benefit of inclusive L3 caches -1 (based on [209])



It is guaranteed that data is not on die



All other cores must be checked (snooped)!

Benefit of inclusive L3 caches -2 (based on [209])

Note: For higher core counts L2 snooping becomes a more demanding task and overshadows the benefits arising from the more efficient cache use of the explicit cache scheme.

3.2.3 New cache architecture (9)

Introduction of L3 caches in other processor lines

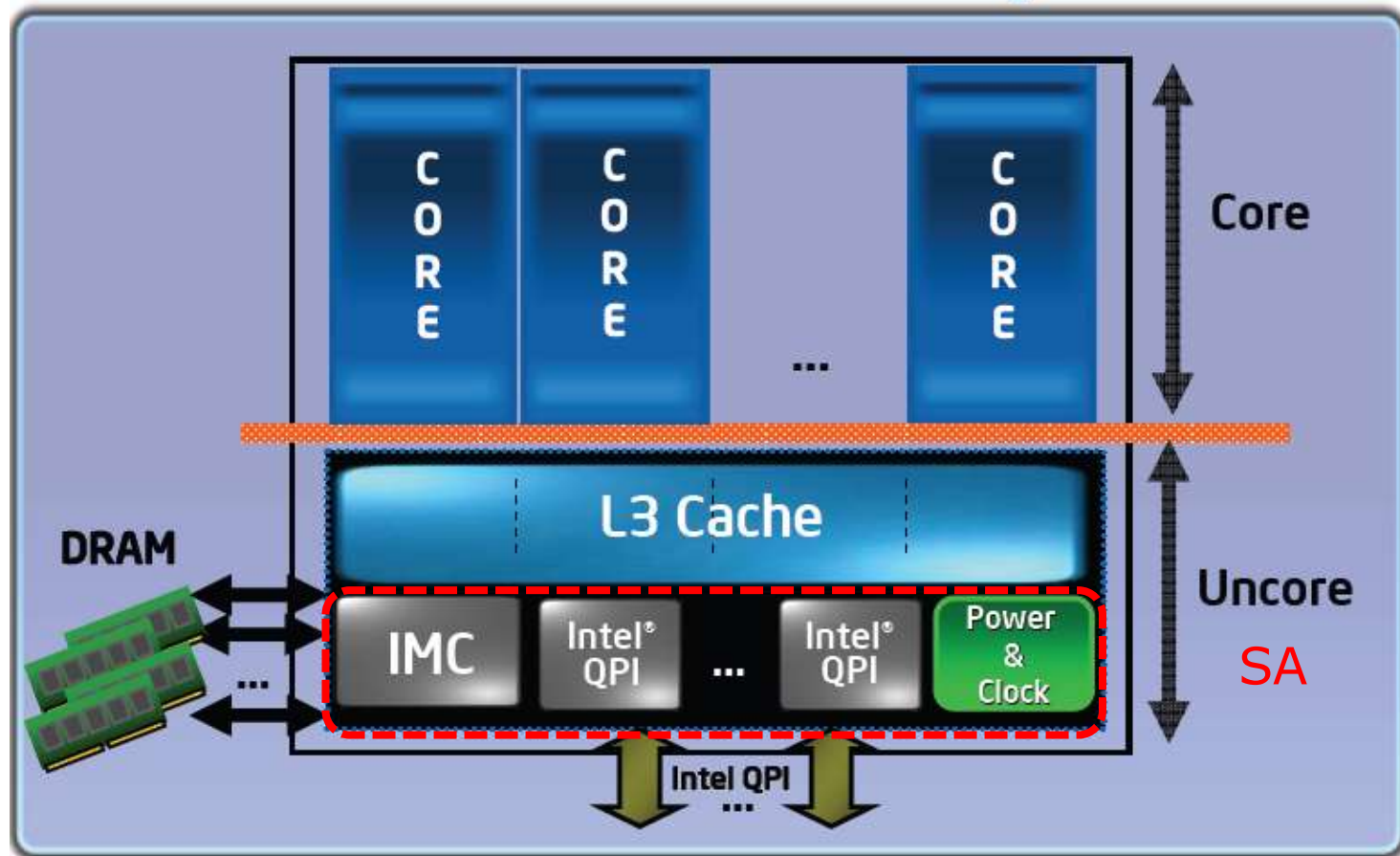
Vendor	Model	Core count	L2 MB	Year of intro.		Model	Core count	L3	Year of intro
IBM	POWER3 II	1C	16 MB off-chip	1999	→	POWER4	2C	32 MB off-chip	2001
						POWER5	2C	36 MB off-chip	2004
						POWER6	2C	32 MB off-chip	2007
						POWER7	8C	8X4MB on-chip	2010
AMD	K8 Santa Rosa	2C	2x1 MB on-chip	2006	→	K10 Barcelona	4C	2 MB on-chip	2007
Intel	Penryn	2C	6 MB on-chip	2008	→	Nehalem	4C	8 MB on-chip	2008

Remark

In the Skylake-SP server processor (2017) both

- the L2/L3 cache sizes were changed and also
- the inclusion policy from inclusive to non-inclusive (different from exclusive)

The notions of “Uncore” [1] and “System Agent”



Subsequently, Intel introduced the notion of **System Agent (SA)**, it is the L3 cache-less part of Uncore.

3.2.4 Simultaneous Multithreading (SMT)

In Nehalem Intel re-implemented SMT (since Core 2/Penryn did not support SMT)

SMT: two-way multithreading (two threads at the same time)

Each issue slot may be filled now from two threads.

Benefits

- A 4-wide core is fed more efficiently (from 2 threads).
- Hides latency of a single thread.
- More performance with low (e.g. 5%) additional die area cost.
- May provide significant performance increase on dedicated applications, as seen in the next Figure.

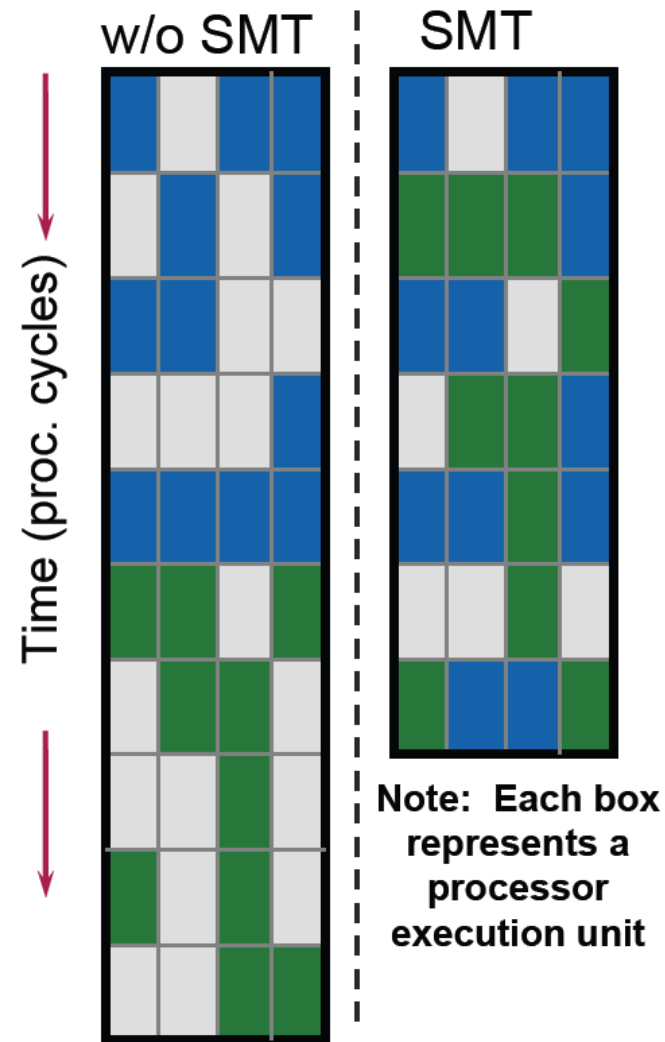
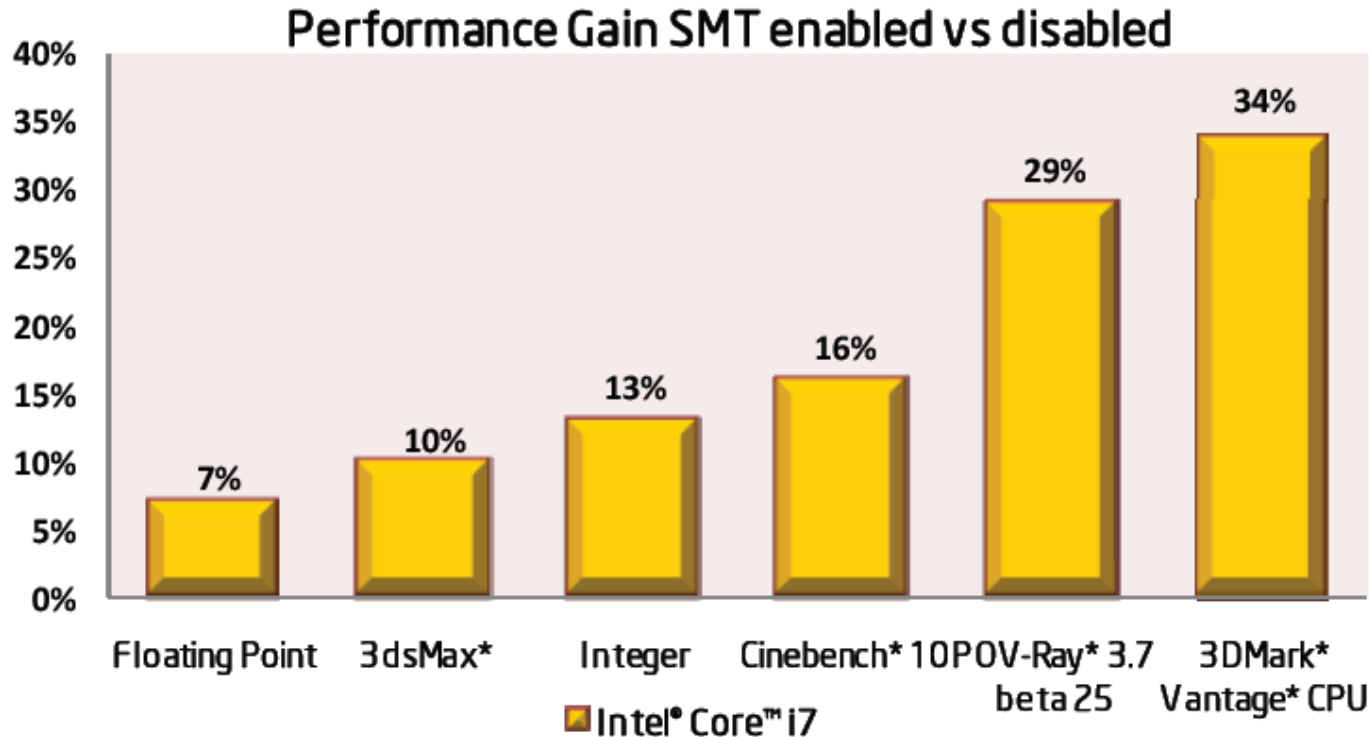


Figure 3.2.4.1: Simultaneous Multithreading (SMT) of Nehalem [1]

3.2.4 Simultaneous Multithreading (SMT) (2)

Performance gains achieved by Nehalem's SMT [1]



Floating Point is based on SPECfp_rate_base2006* estimate
Integer is based on SPECint_rate_base2006* estimate

3.2.5 Enhanced power management

3.2.5.1 Introduction

Having 4 cores instead of two clearly results in higher power consumption and this puts greater emphasis on a more sophisticated power management.

Innovations introduced to encounter this challenge:

- **Integrated power gates** (to significantly reduce power consumption)
- **Integrated Power Control Unit (PCU)** (to implement the complex task of power management)
- **Turbo Boost technology** (to convert power headroom to higher performance)

Above innovations will be discussed while we give a brief introduction into the wide spectrum of power management technics.

Power consumption: energia fogyasztás

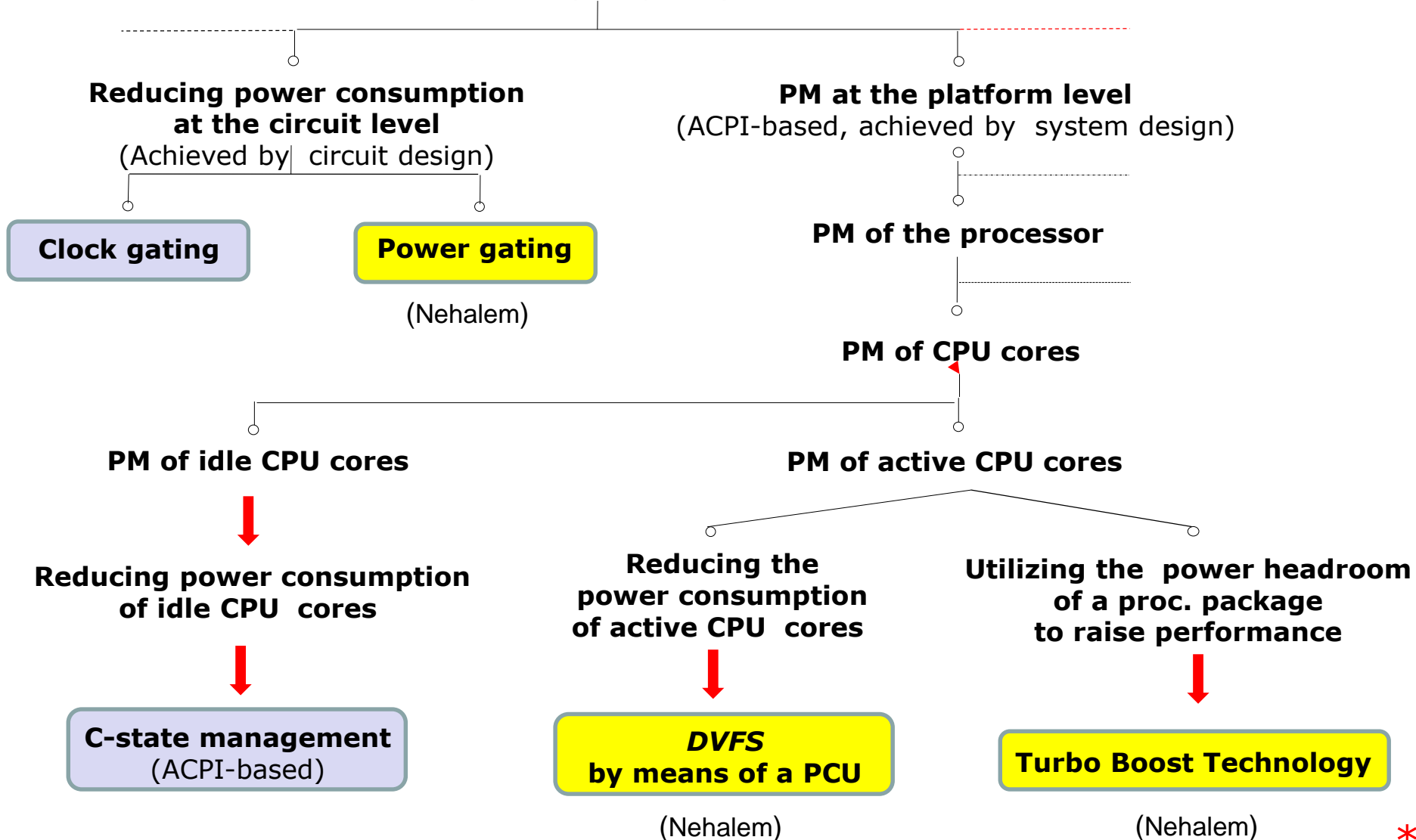
Power management: fogyasztás kezelés/disszipáció kezelés

Power gates: áramellátás kapu

Approaches and key technologies of power management in computers

Approaches for the power management of computers

(Strongly simplified)



3.2.5.2 Clock gating

- Eliminates dynamic dissipation of unused circuits by switching off their clocking.
- Clock gating was introduced in the late 1990s e.g. in DEC processor designs (in the Alpha 21264 (1996) for gating the FP unit or the StrongARM SA110 (1996)), designated at that time as conditional clocking.
- Soon fine-grained clock gating became widely used, e.g. in Intel's Pentium 4 (2000) or Pentium M (Banias) (2003).
- Recently, fine-grained clock-gating is a pervasively used technique in processors.

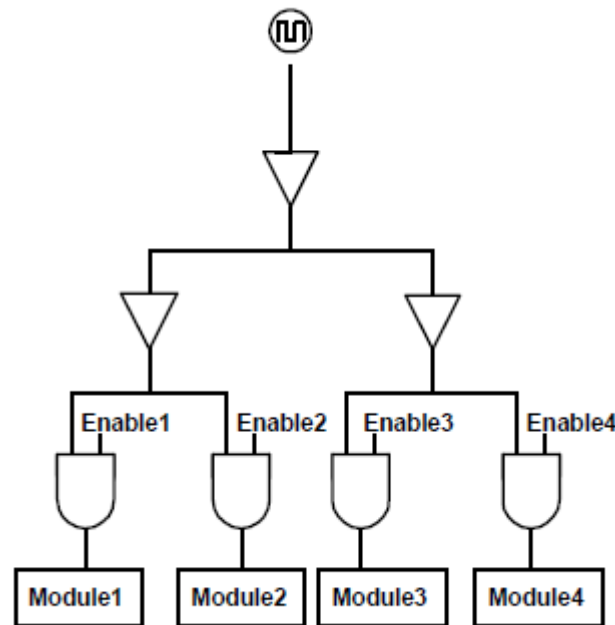


Figure: Principle of clock gating [278]

3.2.5.3 Power gating [32]

Power gating means switching off unused units from the power supply by power transistors. It eliminates both static and dynamic dissipation of unused units.

Integrated Power Gate

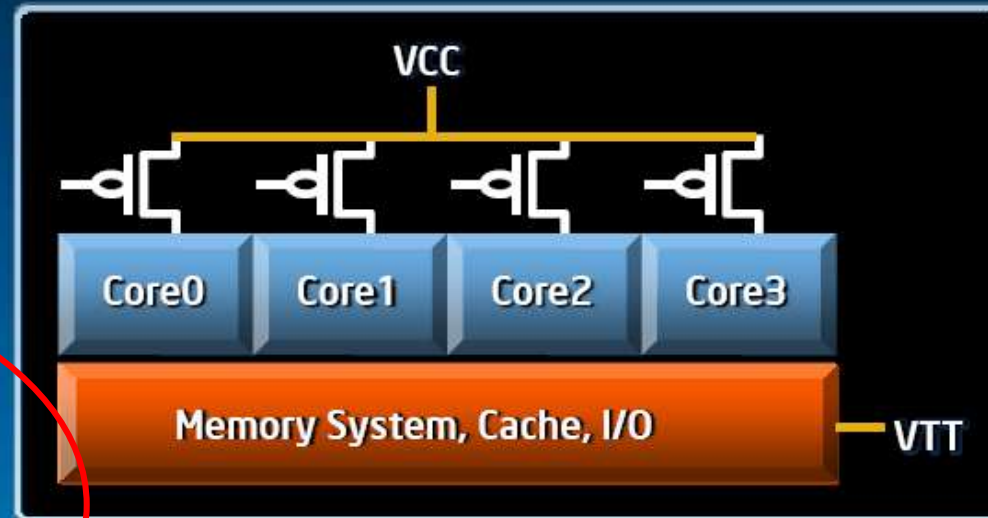
Power switches

Clock Gate

- Standard in all Intel CPUs
- Eliminates switching power
- Leakage power remains

Power Gate: New

- Eliminates switching and leakage power
- Enables idle cores to go to near zero power (C6) independently
- Transparent to platform and software
- No incremental platform cost



It is a precondition of an efficient Turbo Boost technology, since it eliminates both static and dynamic dissipation of idle cores and thus enlarges notable the power headroom.

3.2.5.3 Power gating (2)

Remark: Introducing power gating by different processor vendors

Intel introduced power gating along with their Nehalem microarchitecture in 2008, subsequently many other processor vendors followed them, as the Table below shows.

Vendor	Family	Year of intro.
Intel	Nehalem	2008
	Westmere	2010
	Sandy Bridge	2011
	Ivy Bridge	2012
	Skylake	2015
	Atom families	2010 - 2016
AMD	K12-based Llano	2011
	K14-based Bobcat	2011
	K15-based Bulldozer families	2011 - 2015
IBM	POWER7+	2012
	POWER8 (Integrated PG and DVFS)	2014

Table: Introduction of power gating

3.2.5.3 Power gating (3)

Integrated voltage regulators (FIVR) took over the task of power gates

Integrated voltage regulators
(as introduced in Intel's
Haswell and Broadwell-based lines)
allow to switch off units individually
so they supersede the use of power gating,
as the Figure on the right shows.

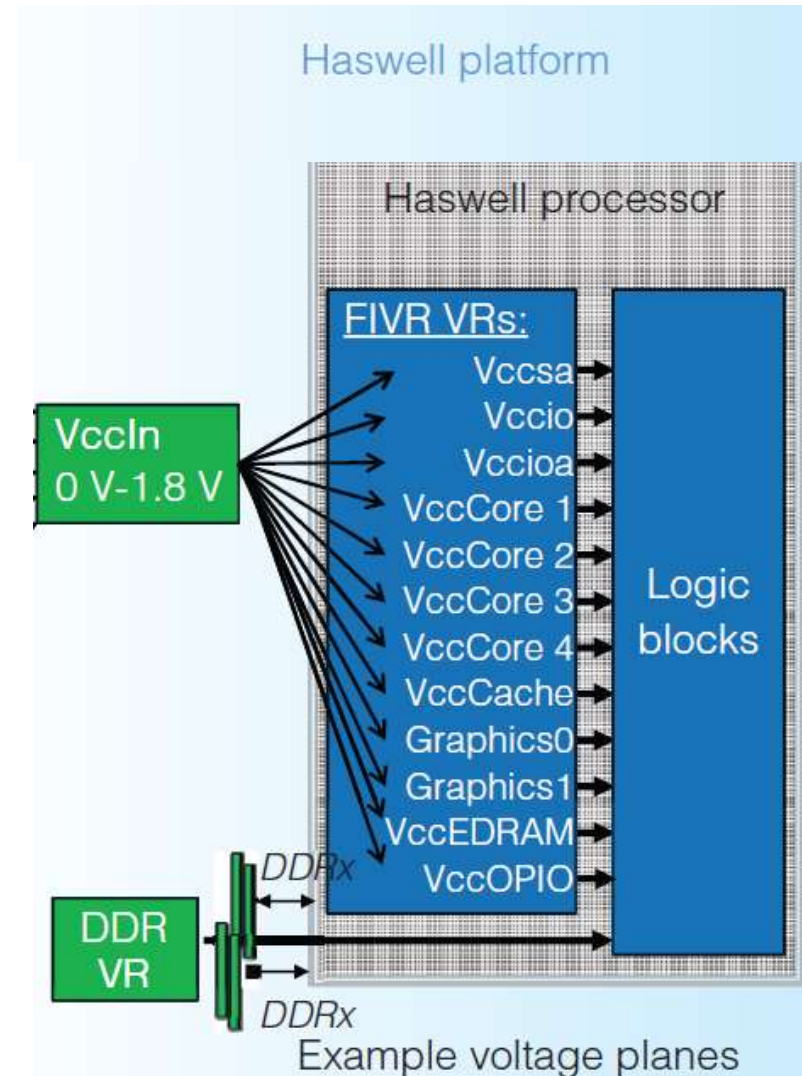


Figure: Use of integrated voltage regulators in Intel's Haswell processor (2014) [279]

3.2.5.3 Power gating (4)

Reuse of power gating after Intel has suspended the implementation of integrated voltage regulators in their Skylake line (2015)

- Integrated voltage regulators (FIVR), introduced into the Haswell and Broadwell lines unduly increased dissipation and thus reduced clock frequency.
- This is the reason why Intel omitted [integrated voltage regulators in their subsequent Skylake line](#), as indicated in the Figure on the right.

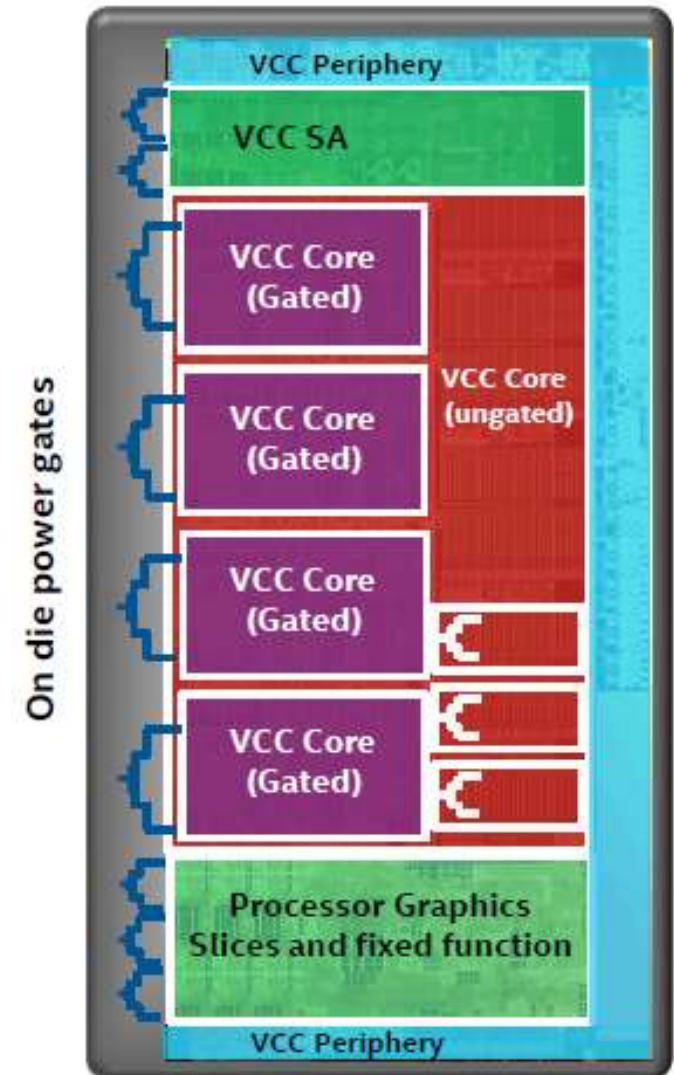


Figure: Reintroducing power gating in Intel's Skylake line [280]

3.2.5.4 The ACPI standard

- Power management can efficiently be supported by the OS, since the task scheduler “sees” the utilization of the cores or threads and this “knowledge” can be utilized for power management, as discussed later.
- OS support requires a standard interface for power management between the processor and the OS.

This need gave birth to power management standards.

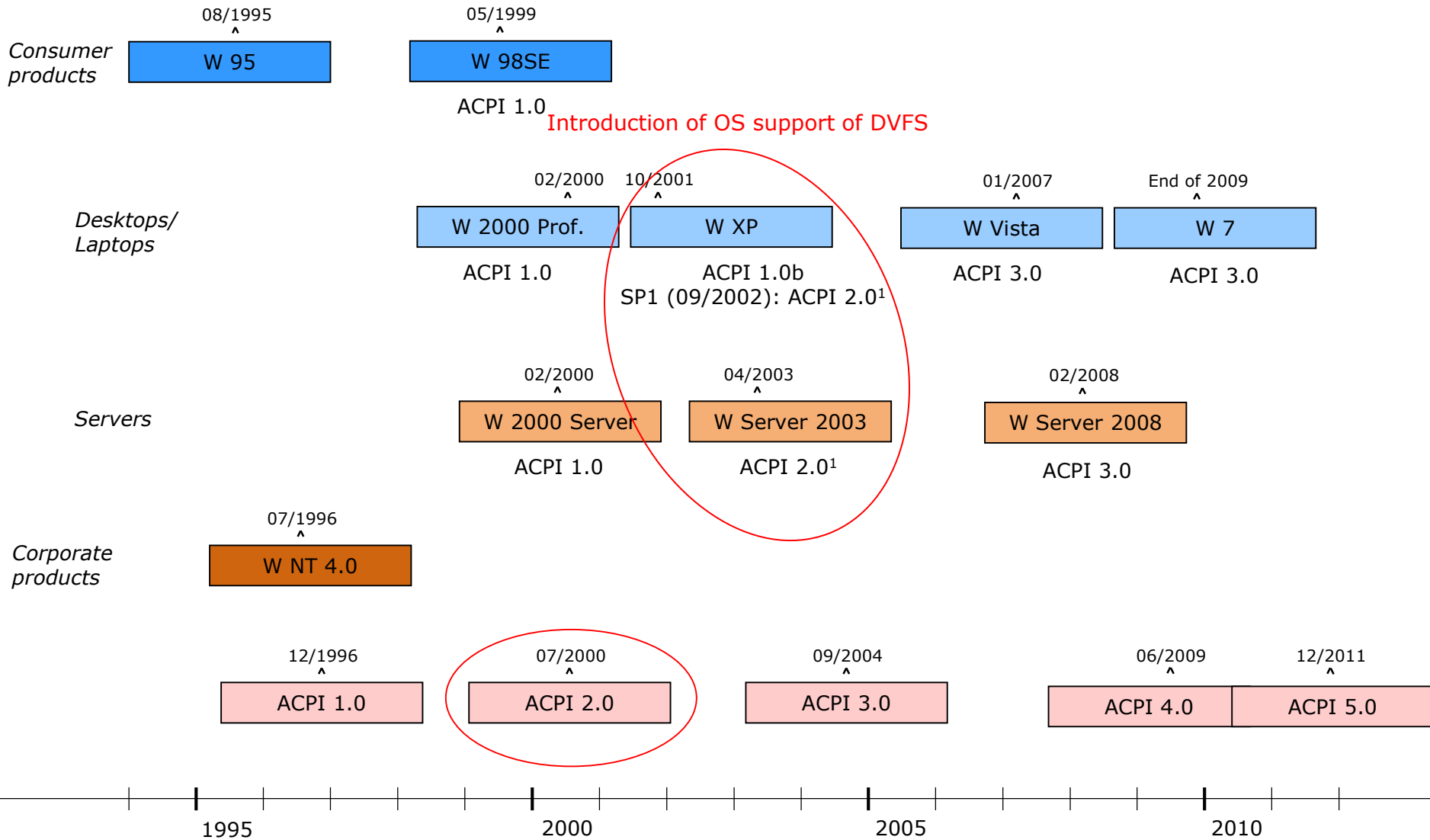
3.2.5.4 The ACPI standard (1b)

Evolution of power management standards

	Power management standards		
	(SL technology)	Advanced Power Management (APM)	Advanced Configuration and Power Interface (ACPI)
Introduced	10/1990	01/1992	12/1996
Vendor	Intel	Intel and Microsoft	Intel, Microsoft, Compaq, Phoenix and Toshiba
	A set of PM techniques	Open standard i.f. between OS and BIOS	Open standard i.f. between OS and HW
Typ. CPU scaling Done basically	SFS	DFS	DVFS
• for CPU:	by SMM	by OSPM/BIOS	by OSPM
• for devices:	by SMM	by BIOS/OSPM/OS handlers	by OSPM
First proc. supp.	386SL (embedded) 486SL (embedded) 486 family (since 06/1993)	Pentium	Pentium M and subsequent processors
First Intel's chipset supp.:	420EX 420ZX	430FX with PIIX 430HX/430VX with PIIX3	430TX with PIIX4 (ACPI 1.0)
OS support:	No OS support needed	Windows 95 (08/1995) Windows 98 (06/1998)	Windows 98 (ACPI 1.0) (06/1998) Windows XP SP1 (ACPI 2.0) (02/2002)

3.2.5.4 The ACPI standard (2)

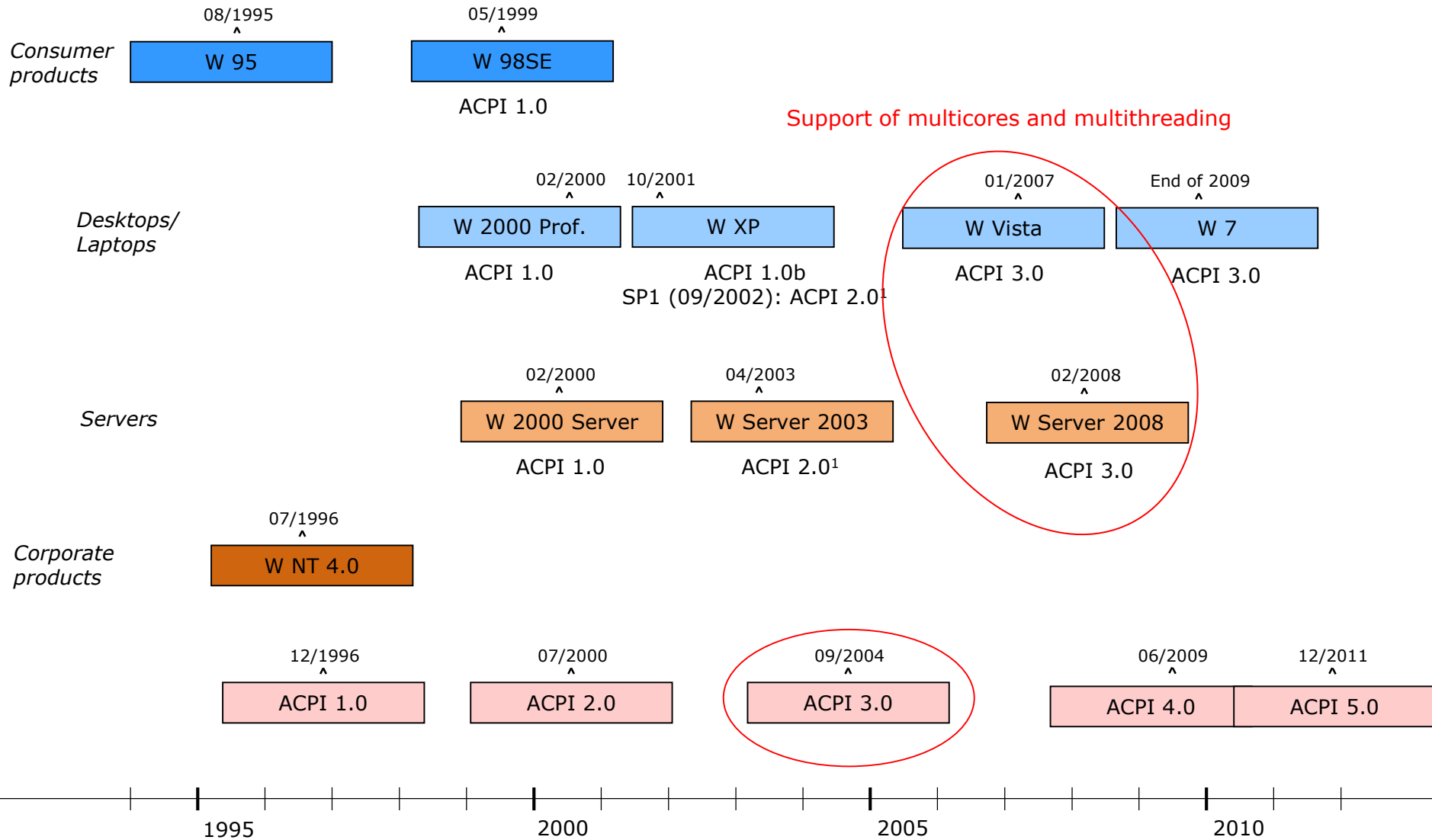
Emergence of the ACPI standard and its OS support



¹: Windows XP and Windows Server 2003 do not support all of the ACPI 2.0 specification [281]

3.2.5.4 The ACPI standard (3)

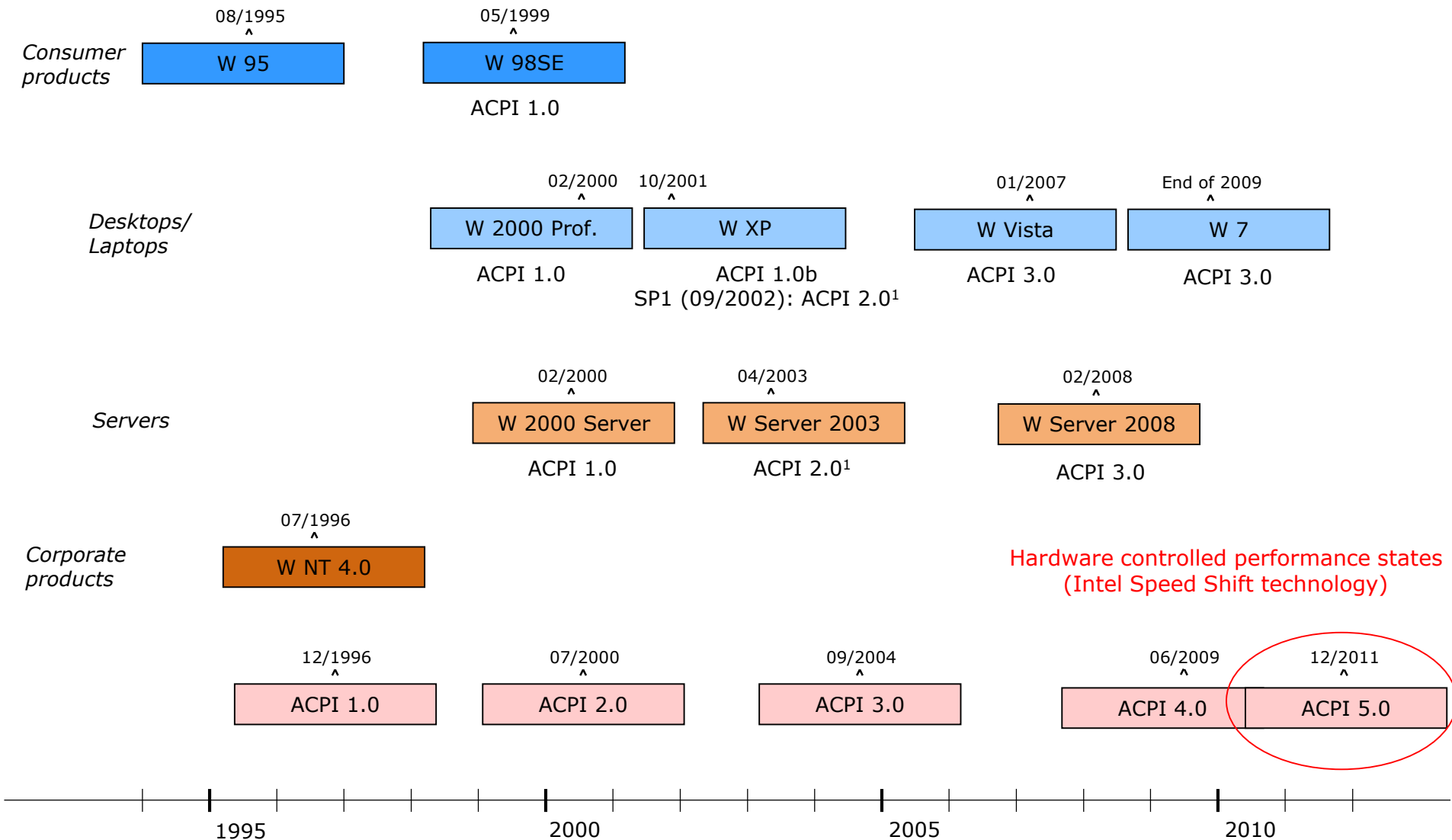
Support of multicores and multithreading in ACPI 3.0 and its OS support



¹: Windows XP and Windows Server 2003 do not support all of the ACPI 2.0 specification [281]

3.2.5.4 The ACPI standard (4)

Support of hardware controlled performance states (SpeedShift technology) in ACPI 5.0



¹: Windows XP and Windows Server 2003 do not support all of the ACPI 2.0 specification [281]

3.2.5.4 The ACPI standard (5)

Example: ACPI states in Haswell-based mobiles [282]

Pi: Performance states
(active states, since ACPI 2.0)

Ci: Idle states
(C4...Cn states since ACPI 2.0)

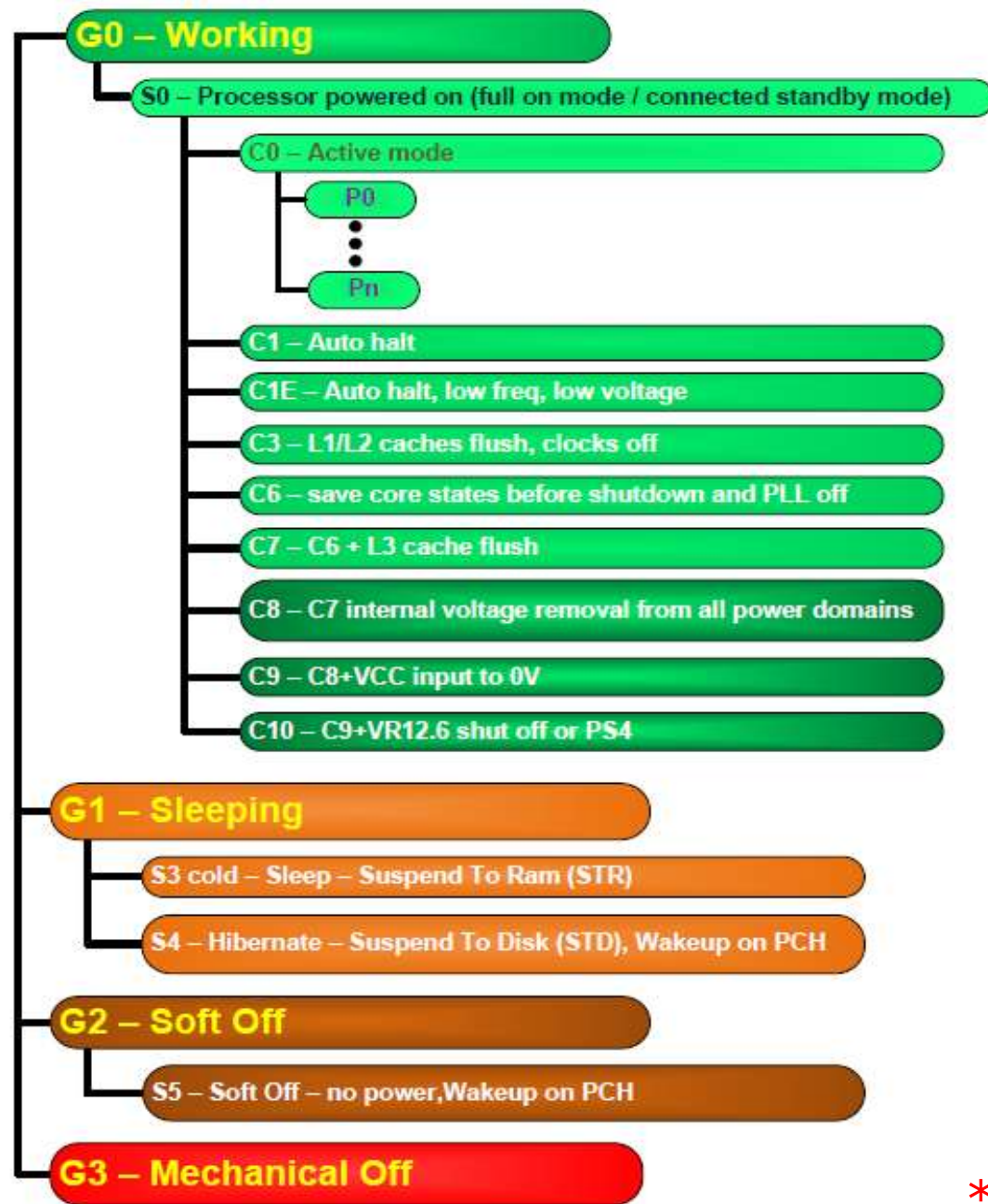
Gi: Global states

Si: System Sleep states

G1: OS-initiated, system context is saved,
no rebooting needed.

G2/Soft off: OS-initiated shut down.
Power supply remains on,
system context is not saved,
the system must be restarted.

G3/Mechanical off: Entered by
activating a mechanical switch.
The system must be restarted.



3.2.5.5 C-state management

- **Idle periods** of instruction execution **allow to reduce power consumption**, e.g. by clock gating, power gating, switching off caches etc.
- To allow **managing idle states by means of OSs in a standardized way**, ACPI introduced sog. **C-states**.

Introduction to C-states -1

- Version 1.0 of the ACPI standard introduced the C1 .. C3 idle states in 1996.
- Additional idle states C4Cn were defined in version 2.0 of this standard in 2000, as indicated in the next Figure.
- We note that the ACPI standard details the idle states C1 to C3 but does not give a detailed specification for the C4 ...Cn states, thus the C4 and higher states may be specified differently from vendor to vendor and from processor line to processor line.

3.2.5.5 C-state management (3)

Example: ACPI states in Haswell-based mobiles [282]

Pi: Performance states
(active states, since ACPI 2.0))

Ci: Idle states
(C4...Cn states since ACPI 2.0)

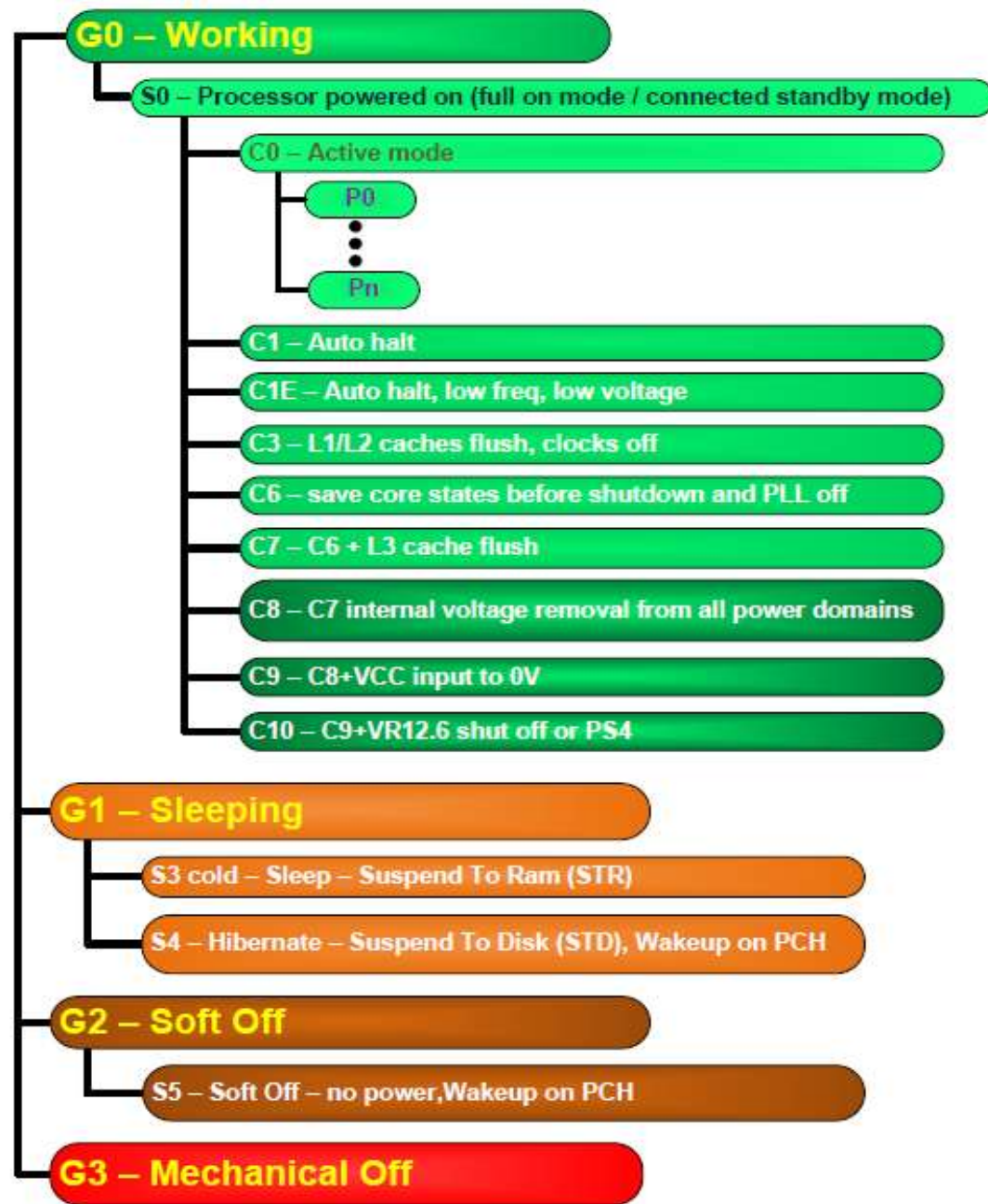
Gi: Global states

Si: System Sleep states

G1: OS-initiated, system context is saved,
no rebooting needed.

G2/Soft off: OS-initiated shut down.
Power supply remains on,
system context is not saved,
the system must be restarted.

G3/Mechanical off: Entered by
activating a mechanical switch.
The system must be restarted.



Introduction to C-states -2

- Higher numbered C-states designate increasingly deeper sleep states.
- Deeper sleep states provide higher power savings but require higher enter and exit times, as seen in the next Figure.

Power consumption vs. transfer latency of C-states

Higher numbered C states i.e. deeper idle states, result in lower power consumption but cause increasingly longer transit latencies (enter plus exit times), as indicated below for the C-states C1 - C6.

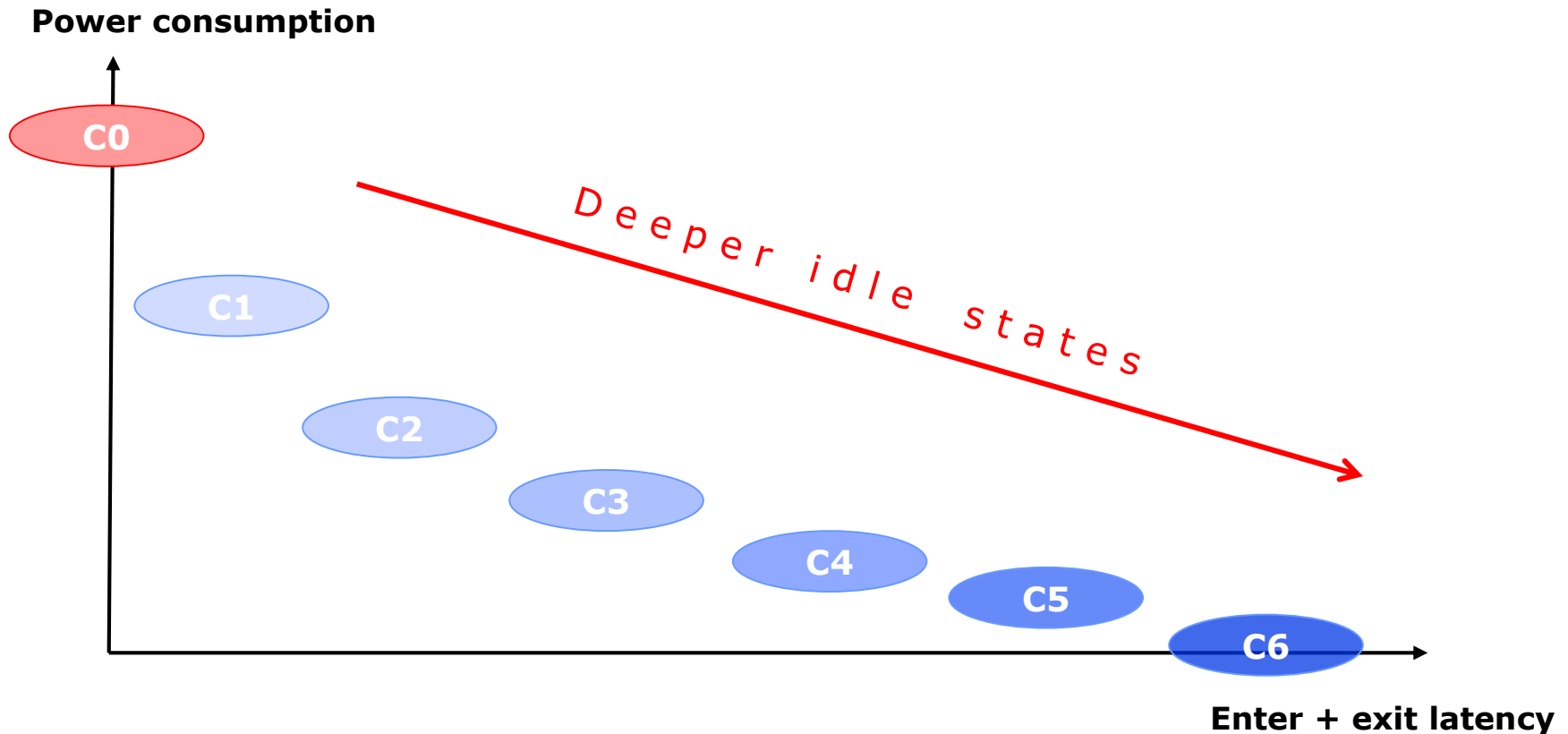
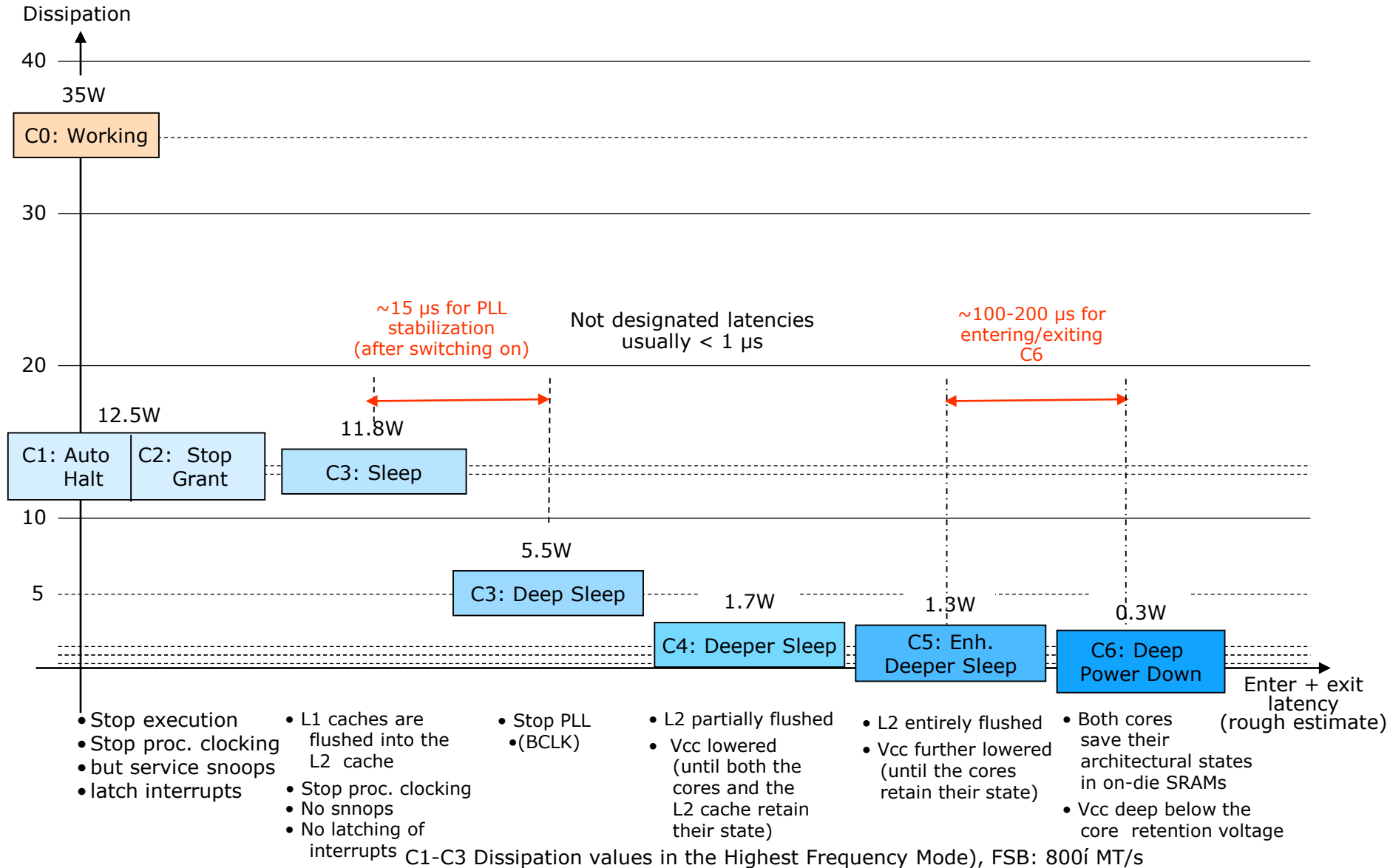


Figure: Power consumption vs. transfer latency of C-states

3.2.5.5 C-state management (6)

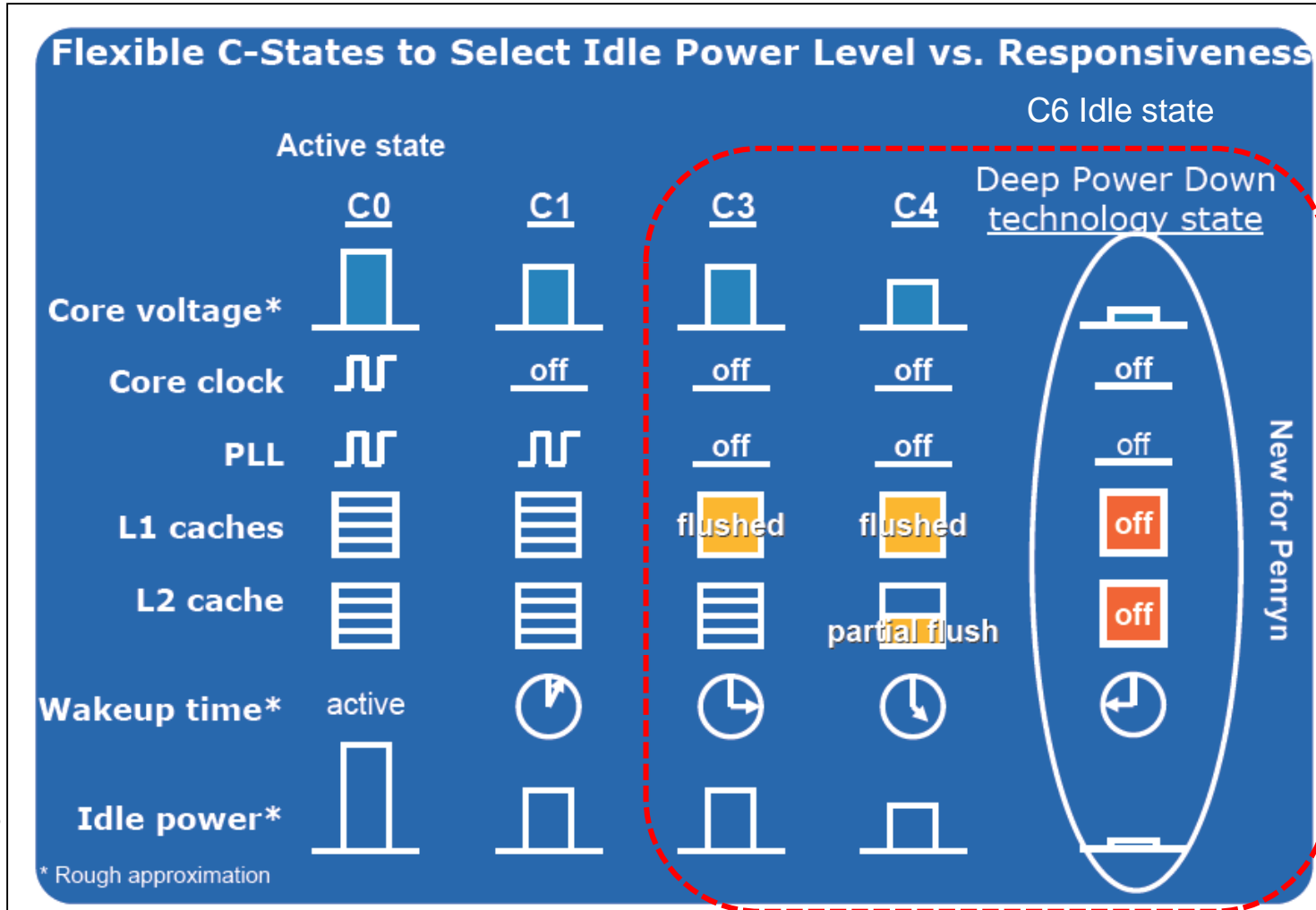
Example: Dissipation and enter + exit latencies of idle states in Intel's Penryn-based Core 2 Duo Mobile processors (e.g. T9xxx) (2008) [283]



3.2.5.5 C-state management (8)

Example: ACPI C-states in Intel's mobile Penryn-based processors [26]

- New Power Management State
- Significantly reduces processor power consumed in idle mode
- Further Extends Battery Life
- Intelligent heuristics decides when enter into.

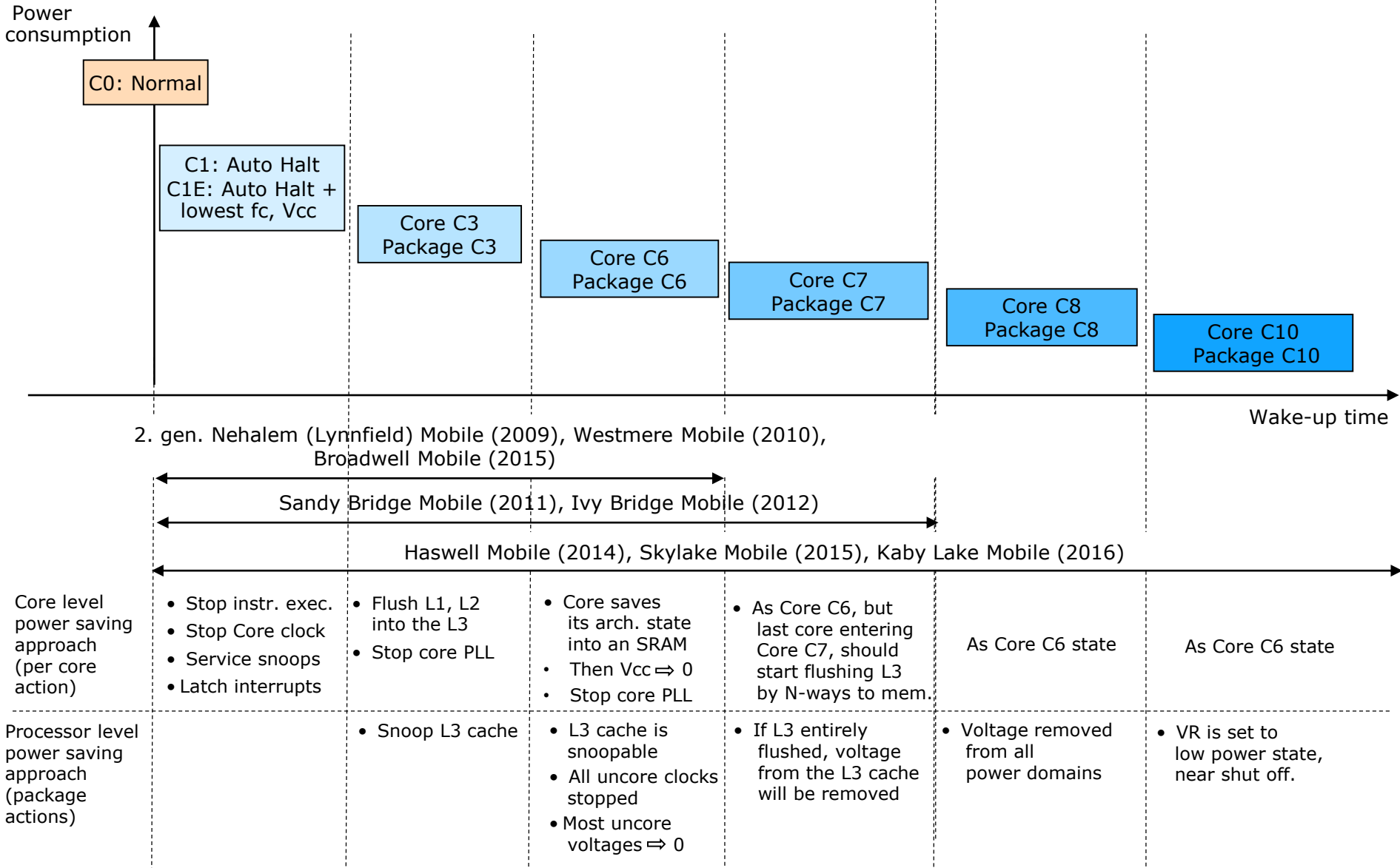


Remark

- While **mobile processors** are the most sensitive processor class concerning power consumption, these processors typically **spearhead C-state management**.
- **By contrast, desktop and server processors support often only a subset of C-states provided by mobiles**, e.g. Haswell mobile processors support C1 to C10 idle states whereas Haswell desktops and servers only the C1 to C6 idle states.
- **Subsequent processor lines** usually introduce **more idle states** with more and more sophisticated power preservation techniques.

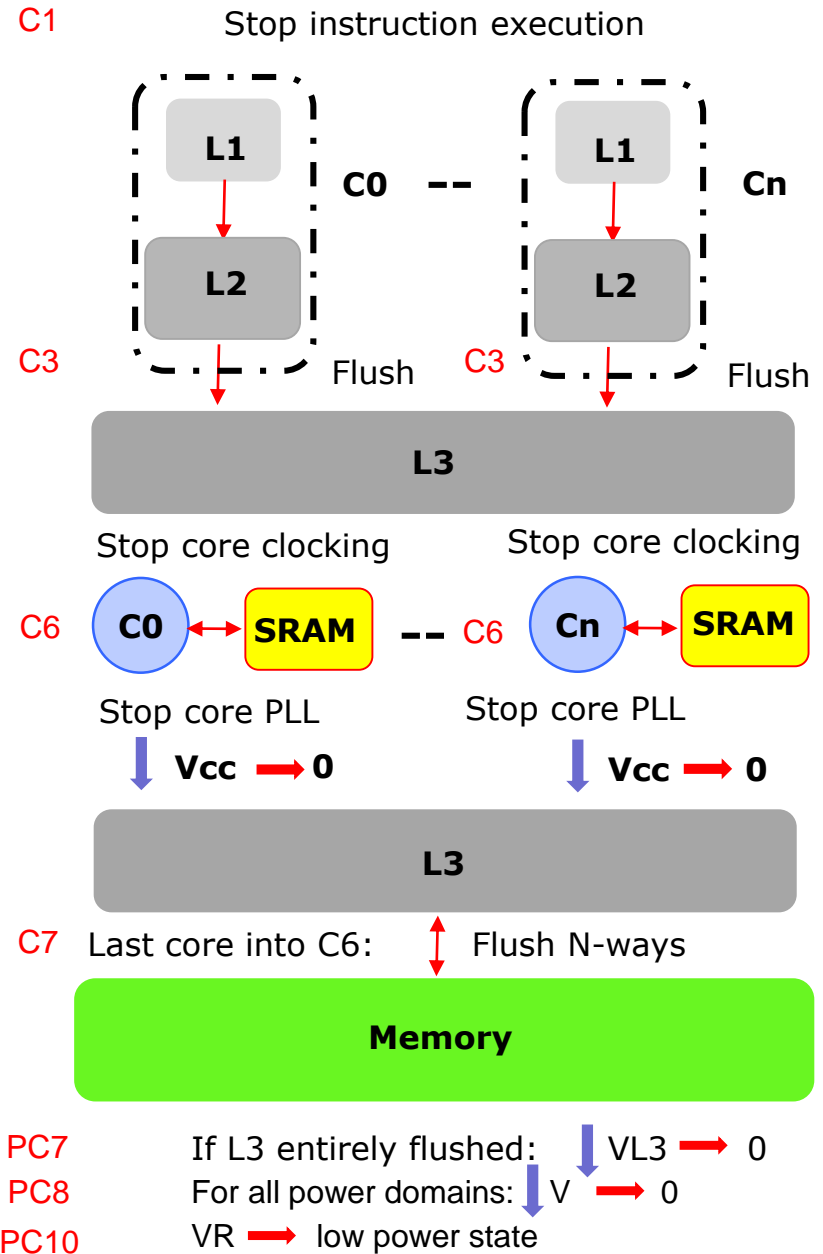
3.2.5.5 C-state management (10)

C-states and invoked power saving actions in ACPI-compliant PCU-based C-state management in multi-core mobile processors



3.2.5.5 C-state management (11)

C-states and invoked power saving actions in ACPI-compliant PCU-based C-state management in multi-cores (simplified)



Main approaches to implement idle state management

Idle state management

Intel's SL technology

The SB recognizes PM requests, like timeouts etc. and asserts the SMI# interrupt pin to notify the processor. In response, the processor enters the SMM mode, saves its internal state and the BIOS installed SMM handler performs idle state management. The last instruction of the SMM code lets restore the processor state and exit the SMM mode.

ACPI-compliant idle state management (C-state management)

OS recognizes idle periods of instruction execution and instructs the processor to manage C-states through a software interface (via MWAIT(Ci) or P_LVLn I/O READ instructions). Interrupts let exit C-states and enter the C0 operating state.

ACPI-compliant SB-based C-state management

Control logic of the SB is basically responsible for managing C-states

ACPI-compliant PCU-based C-state management

On-die PCU is basically responsible for managing C-states

Use in Intel's processors

From the 386SL (1990) on up to the embedded Pentium VRT (1998)

Section 5.2

SMM: System Management Mode

From the mobile Pentium II (1998) on up to the 1. gen. Nehalem (Bloomfield)-based lines (2008)

Section 5.3

SB: South Bridge

From the 2. gen. Nehalem (Lynnfield)-based lines on (2009)

Section 5.4

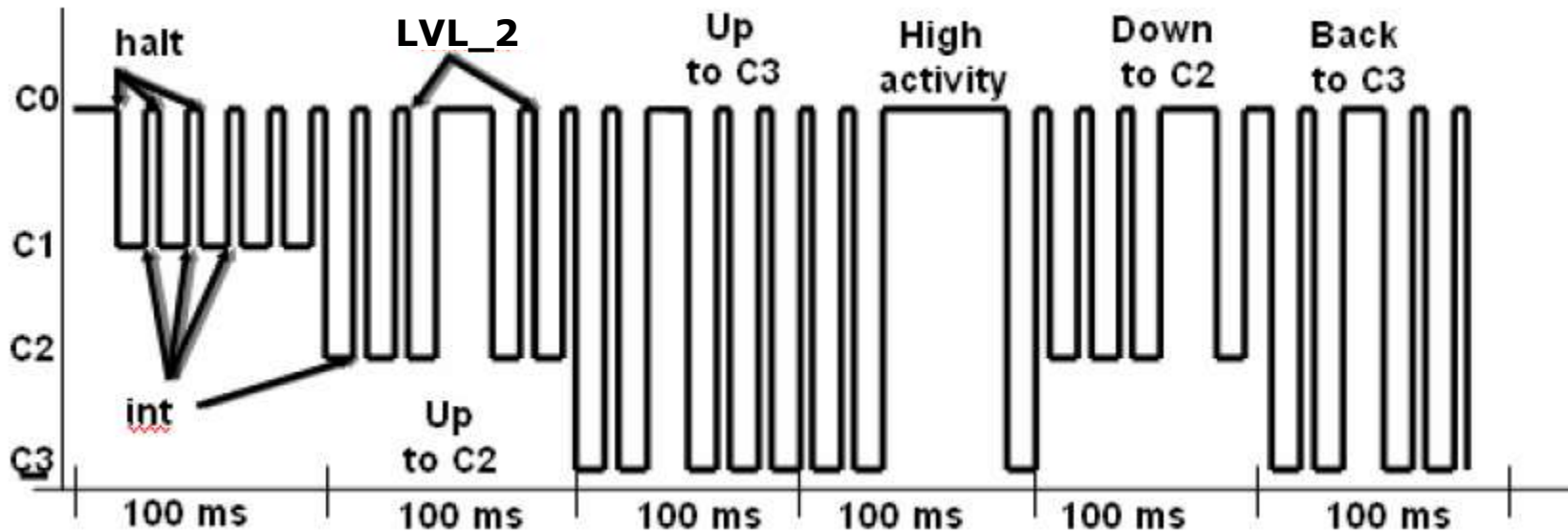
PCU: Power Control Unit *

Principle of ACPI-based idle-state management

- OS recognizes idle periods of instruction execution and instructs the processor to manage C-states through a software interface (dedicated instructions).
- The processor performs the requested C-state transition.
- Interrupts let exit C-states and enter the C0 operating state.

Managing C-state transitions by the OSPM (OS Power Manager)

- The OSPM scheduler recognizes that no work is to do for the processor, evaluates the rate of idle time in **time windows** (of e.g. 20 ms) and **initiates a transition to a target C-state** according to the actual utilization rate in the considered time window, e.g. by sending instructions to the processor.
- The following example will illustrate this.



Example: Managing C-states by the OSPM [284]

A typical OS idle loop as a basis for managing C-states [285]

```
// WorkQueue is a memory location indicating there is a thread
// ready to run. A non-zero value for WorkQueue is assumed to
// indicate the presence of work to be scheduled on the processor.
// The idle loop is entered with interrupts disabled.
WHILE (1) {
    IF (WorkQueue) THEN {
        // Schedule work at WorkQueue
    } ELSE {
// No work to do - wait in appropriate C-state handler depending
// on Idle time accumulated
        IF (IdleTime >= IdleTimeThreshold) THEN {
            // Call appropriate C1, C2, C3 state handler
            // shown below
                }
        }
    }
}
```


3.2.5.5 C-state management (16)

Example: Signal sequence generated by the SB (ICH-8) to enter/exit the C4-state in a Core 2 Duo Mobile Penryn-based platform [286]

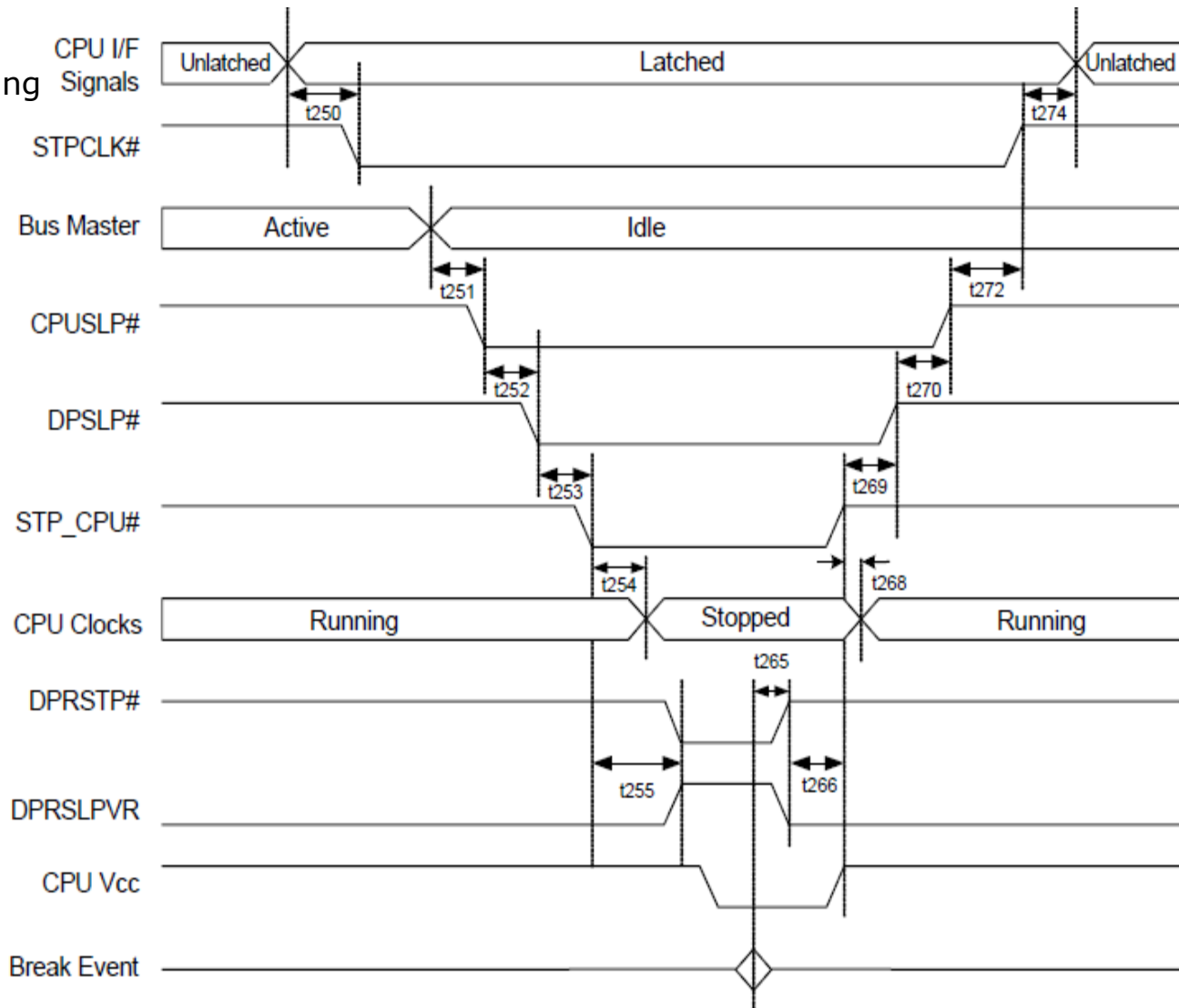
STPCLK# stops the processing of I/F signals instead it lets to latch them for later processing
Sent to the CPU,
stops CPU clocking

CPU_SLP# = SLP#, to the CPU to enter C3 Sleep, forbids snoops to bus masters
Asserts before and deasserts after STPCPU# stops PLL

STP_CPU# = DPSLP#
Sent to the CPU, stops PLL

A copy of the DPRSLPVR

Sent to the VRM (Voltage Regulator Module) to lower Vcc to Vcc4 (to a low value)



C-state management by the PCU

- In the course of the evolution of processors typically the **PCU took over the role of the SB**
 - coordinating the C-state requests
 - and performing the activities needed to implement C-state transitions.
- In Intel's **Core 2 family** this happened beginning with the **2. generation Nehalem line** (called Lynnfield) in **2009**.
- This point will not be further detailed here.

3.2.5.6 DVFS based on a PCU (Power Control Unit) (1)

3.2.5.6 DVFS based on a PCU (Power Control Unit)

Reducing the power consumption of active CPU cores

Static technique

Dynamic techniques

SVFS

DFS

DVFS

Hardware Controlled Performance States

AVFS

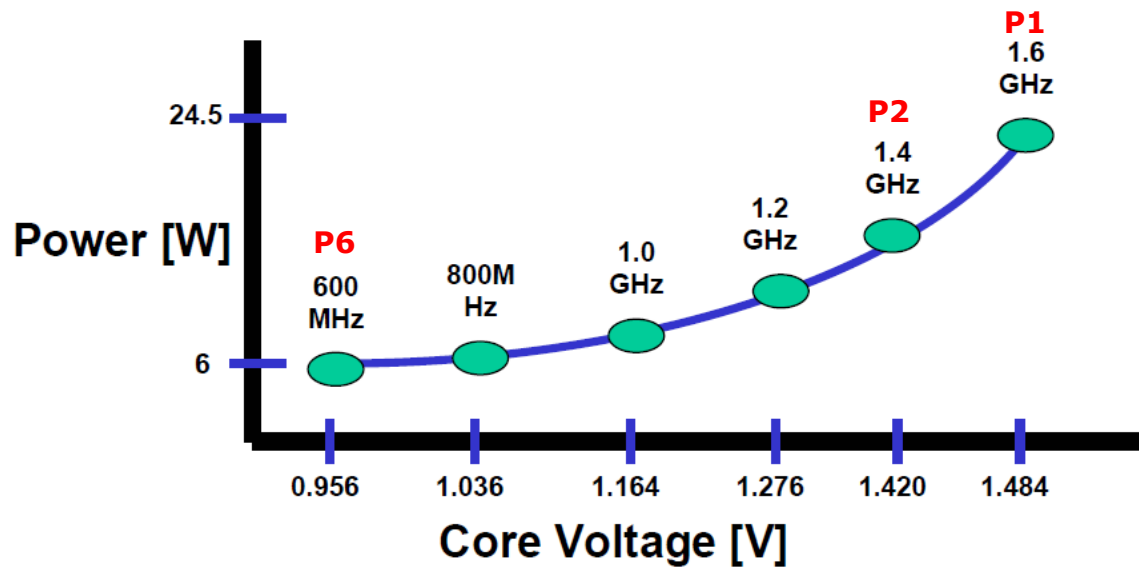
Examples

Intel	<ul style="list-style-type: none"> SpeedStep in Mobile Pentium III (2000) Mobile Pentium 4 (2002) (Northwood based) 		<ul style="list-style-type: none"> EIST in Pentium M (Banias) (2003) and subsequent lines 	<ul style="list-style-type: none"> Speed Shift in Skylake (2015) 	
VIA	LongHaul 1.0 in C3 Samuel (2000)		LongHaul 2.0 in C3 Samuel 2 step. 1 (2001)		Adaptive PowerSaver in Nano (2008)
AMD			<ul style="list-style-type: none"> PowerNow! and Cool'n'Quiet technologies in mobiles/desktops and servers (since 2000) 		<ul style="list-style-type: none"> AVFS in Excavator-based Carizzo (2015) Pure Power in Ryzen (2017)
IBM	PowerPC 750FX (2003)		<ul style="list-style-type: none"> 405LP (2002) Dynamic Power Performance Scaling in PowerPC 750GX (2004) PowerPC 970xx (2004) 		<ul style="list-style-type: none"> Energy Scale in POWER6 (2007) and subsequent processors
Samsung			<ul style="list-style-type: none"> Exynos 4 (2012) (used in Galaxy III) 		<ul style="list-style-type: none"> ASV in Exynos 7420 (2015) (used in Galaxy S6)
ARM/National					<ul style="list-style-type: none"> IEM IP (2002)

Principle of DVFS (Dynamic Voltage and Frequency Scaling) -1

Principle of operation (assuming a multithreaded single core processor):

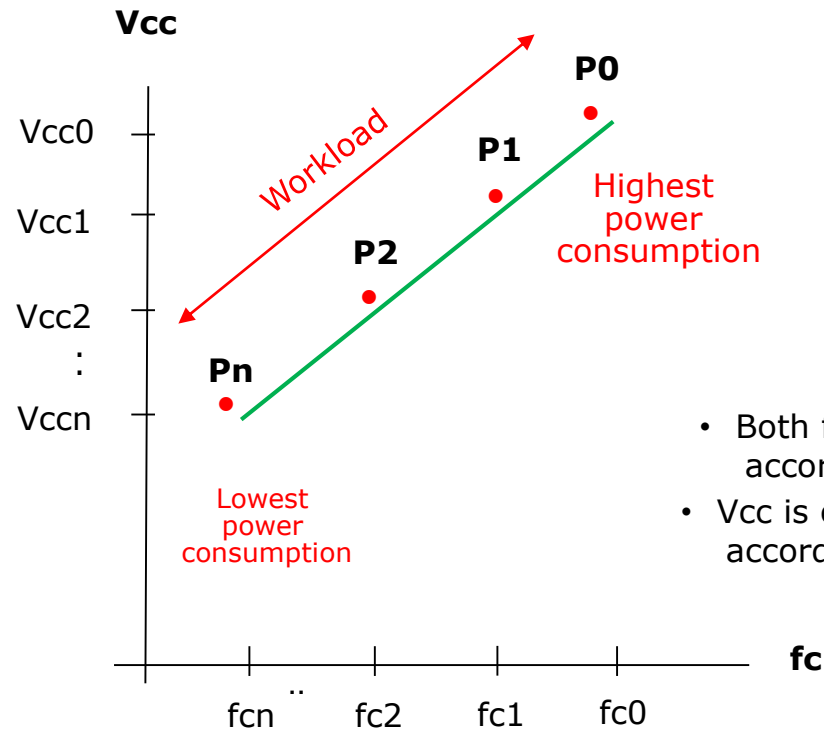
- DVFS scales the clock frequency of the cores just high enough to run the load to be executed on them to save power.
- In other words, lower than expected core utilization will be exploited to reduce core frequency.
- In multithreaded processors, core frequency will be set according to the most active thread.
- DVFS is implemented based on the ACPI P-states (Performance states), introduced in ACPI 2.0. P-states are operating points of the processor, specified by $\{f_c, V_{cc}\}$, with the highest performance P-state designated as P1 (or sometimes as P0), as shown below.



Example: Operating points of Intel's Pentium M processor (~2003) used in Intel's DVFS technology, designated as Enhanced SpeedStep technology [210]

Principle of DVFS (Dynamic Voltage and Frequency Scaling) -2

DVFS typically scales down both the clock frequency (f_c) and the core voltage (V_{cc}) as far as feasible without noticeable lengthening the run time of the workload, in order to reduce power consumption, as indicated below.



- Both f_c and V_{dd} will be scaled according to the workload intensity.
- V_{cc} is chosen with a guard band according to the actual f_c values.

Figure: Principle of DVFS

In this sense, DVFS is a **demand based scaling** of the clock frequency and voltage of the cores.

Principle of DVFS (Dynamic Voltage and Frequency Scaling) -3

- DVFS **may be directed either by the OS or otherwise** (e.g. by the PCU (Power Control Unit) by reading performance counters to calculate utilization and performing all operations needed).

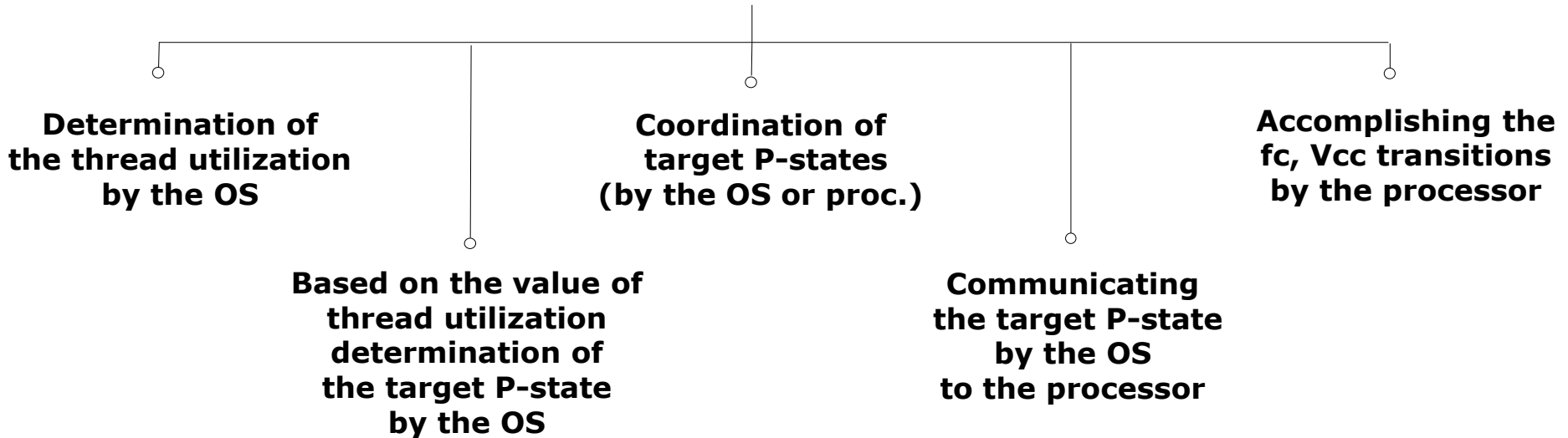
Subsequently, **we assume OS directed DVFS**.

Implementing DVFS in Intel's processors

- First implemented in the Pentium M (Banias) in 2003, designated as **EIST (Enhanced Intel SpeedStep Technology)**.
- Intel **enhanced** their DVFS implementation in the Pentium M (Yonah) in 2006 **with two 64-bit hardware counters (per thread)**, used to help OS to calculate thread/core utilization.

Main tasks of the implementation of DVFS (assuming multithreading)

Main tasks of OS directed DVFS implementation (assuming multithreading)



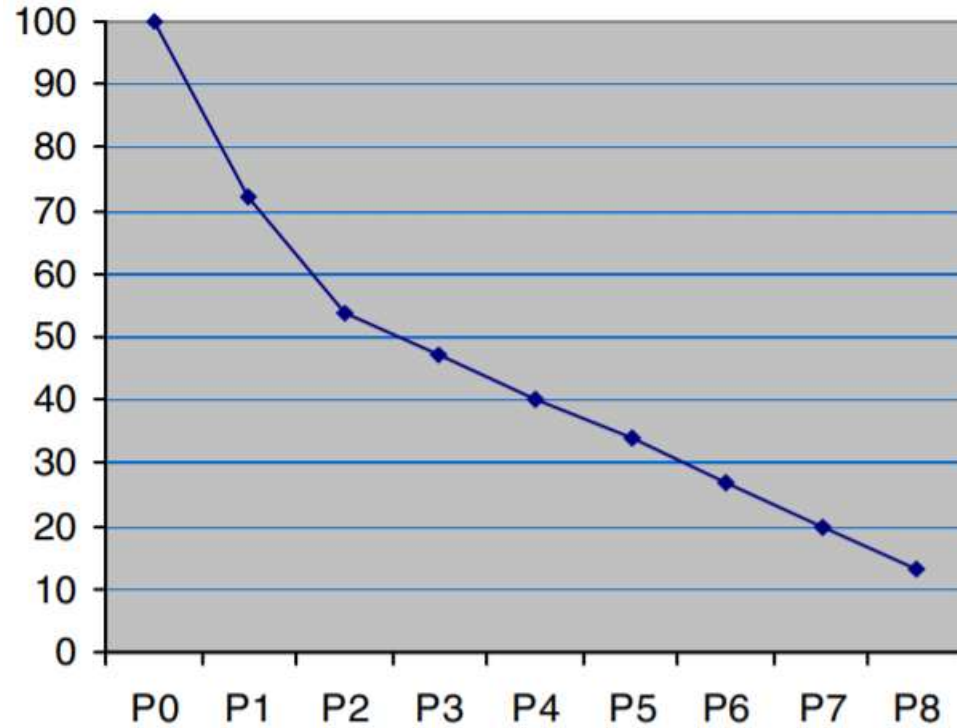
Determination of thread utilization and the target P-state by the OS (simplified) [287] -1

- The processor has **two Model Specific Registers (MSRs) per thread**, that are **actually 64-bit counters** (called also logical processor).
- **One** of the counters (IA32_MPERF MSR (0xE7h)) **increments in proportion to the base frequency**,
the **other** one (IA32_APERF MSR (0xE8h)) **increments in proportion to actual performance**.
These counters are updated only when the targeted processor is **in the C0 state**.
- Based on the readings of these counters the **OS determines the utility rate (%) of the thread as the ratio of the readings (actual/base)**.
- The OS has a **list of available P states** for the cores (specifying fc and needed Vcc).
From the available P-states the OS **selects the lowest possible** to service the actual load (i.e. utility rate).

3.2.5.6 DVFS based on a PCU (Power Control Unit) (8)

Example allocation of P-states to the rate of core utilization while running a thread [289]

Core utilization %



Coordination of P-states

- OS handles P-states of threads.
- If multiple threads are running on the same core, the target P-state (coordinated P-state) needs to be the P-state of the most demanding thread.
- In addition, if multiple cores are supplied by a common voltage or common clock, a coordination of the requested P-states for the cores is needed.

Then the target P-state of the cores becomes the P-state of the most demanding core.

Nevertheless we do not discuss further on this point.

Communicating the target P-state by the OS to the processor (simplified) [287]

- If the new P-state differs from the actual one, the OS writes the selected P-state to a given MSR (bits 0-15 of IA32_PERF_CTL (199h)) available for the thread, (actually, bits [15:8] specify the target multiplier ratio e.g. 34 for 34 x 133 MHz) and bits [7:0] the target core voltage, in a suitable coding).

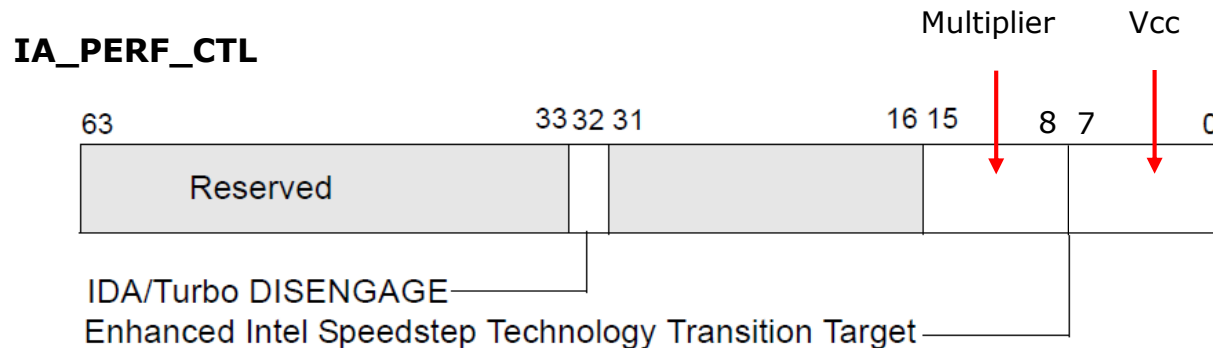


Figure: Use of the IA32_PERF_CTL (199h) MSR to set a new P-state [287]

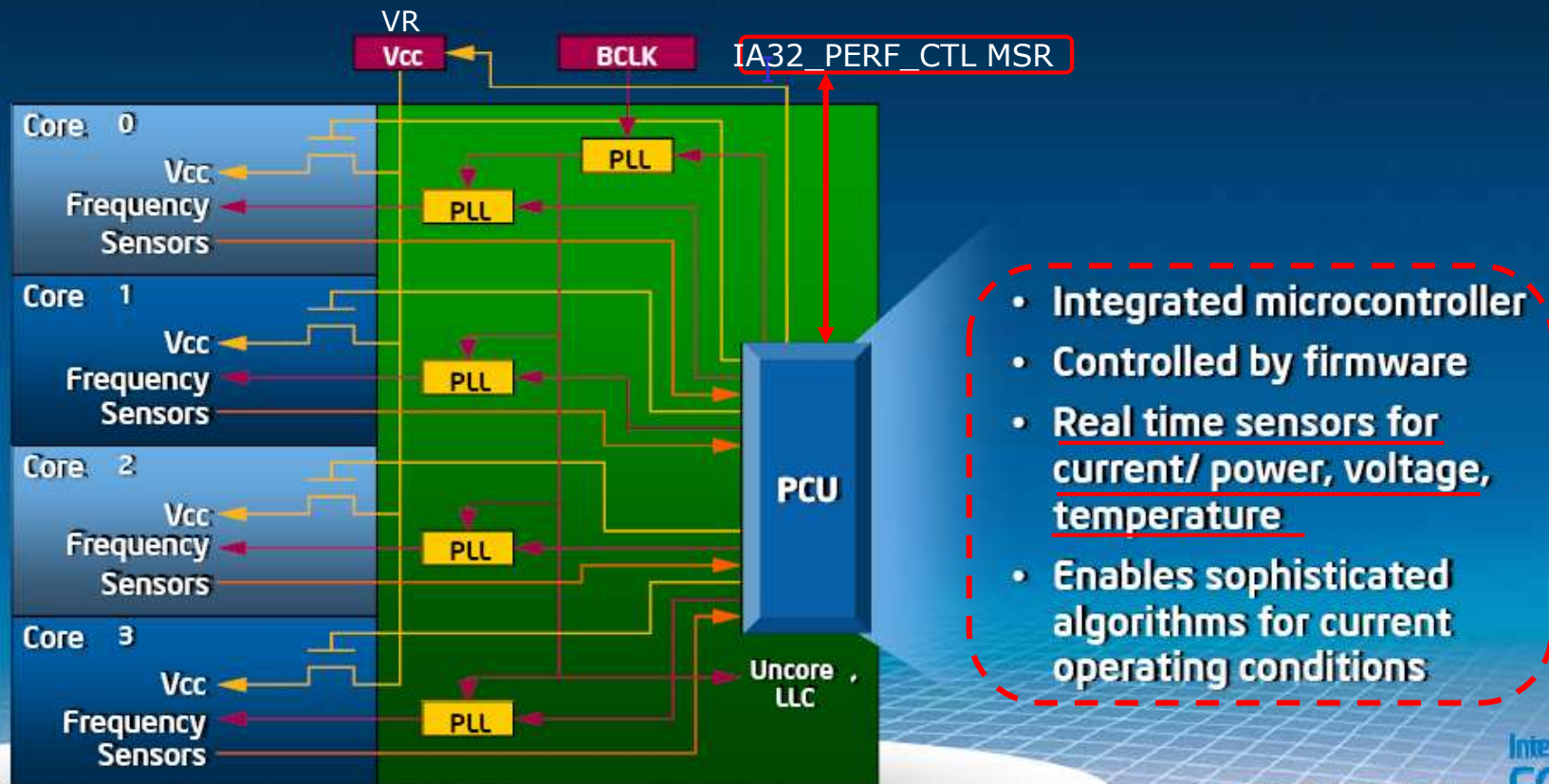
The PCU (Power Control Unit) takes notice of the state transition request and performs it by setting the PLL associated to the core running the considered thread and setting also the Voltage Regulator and initiating the transition.

Remark: The 64-bit counters used for determine thread utilization were introduced in 2006 in the Pentium M Core Duo (Yonah) processor., based on an Intel patent [288].

Implementing DVFS by means of the introduced Integrated Power Control unit (PCU) [32]

- With four cores the power consumption of the chip needs to be managed as an entity.
- This task will be overtaken by a **dedicated microcontroller** implemented on the die.
- It is also used for implementing the **Turbo Boost Mode**.

Power Management: Power Control Unit



Remark

There are two improvements of DVFS:

- **Hardware controlled performance states**
passing over the control of DVFS from the OS to the PCU, to get a faster and finer frequency and voltage scaling (introduced in Skylake (2015))
- **AVFS (Adaptive Voltage and Frequency Scaling)**
to get a more efficient scaling, to be discussed along with AMD's Zen processors.

3.2.5.6 DVFS based on a PCU (Power Control Unit) (12)

Remark: DVFS is one of the technologies used to reduce power consumption of active CPU cores

Reducing the power consumption of active CPU cores

Static technique

Dynamic techniques

SVFS

DFS

DVFS

Hardware Controlled Performance States

AVFS

Examples

Intel	<ul style="list-style-type: none"> SpeedStep in Mobile Pentium III (2000) Mobile Pentium 4 (2002) (Northwood based) 		<ul style="list-style-type: none"> EIST in Pentium M (Banias) (2003) and subsequent lines 	<ul style="list-style-type: none"> Speed Shift in Skylake (2015) 	
VIA	LongHaul 1.0 in C3 Samuel (2000)		LongHaul 2.0 in C3 Samuel 2 step. 1 (2001)		Adaptive PowerSaver in Nano (2008)
AMD			<ul style="list-style-type: none"> PowerNow! and Cool'n'Quiet technologies in mobiles/desktops and servers (since 2000) 		<ul style="list-style-type: none"> AVFS in Excavator-based Carizzo (2015) Pure Power in Ryzen (2017)
IBM	PowerPC 750FX (2003)		<ul style="list-style-type: none"> 405LP (2002) Dynamic Power Performance Scaling in PowerPC 750GX (2004) PowerPC 970xx (2004) 		<ul style="list-style-type: none"> Energy Scale in POWER6 (2007) and subsequent processors
Samsung			<ul style="list-style-type: none"> Exynos 4 (2012) (used in Galaxy III) 		<ul style="list-style-type: none"> ASV in Exynos 7420 (2015) (used in Galaxy S6)
ARM/National					<ul style="list-style-type: none"> IEM IP (2002)

3.2.5.7 Nehalem's Turbo Boost technology -1

- The **Turbo Boost technology** is strongly connected to the notion of the **TPD (Thermal Design Power)** value.

Remark

- The **TDP (Thermal Design Power)** is the **design value for the power consumption of the processor (package)**, given in W.
The TDP value of a processor model reflects the **maximum power consumed by realistic, power intensive applications**.
It serves as a **reference value for designing the cooling system** of the platform.
- The **cooling system** (called also thermal solution) of a platform has to be designed such that it **should guarantee** that the chip, more precisely the **junction temperature (T_j) does not exceed a given limit (T_{jmax} , e.g. 90 °C) while the processor dissipates TDP (given usually in Watts).**

3.2.5.7 Nehalem's Turbo Boost technology -1

- If actually, the processor dissipates less than its TDP value a power headroom arises.
- Turbo Boost technology converts power headroom to higher performance by raising, by raising the clock frequency of the active cores.

3.2.5.7 Nehalem's Turbo Boost technology (3)

Intel's forerunner of implementing Turbo Boost technology in Nehalem: EDAT in Penryn-based mobiles

- In its **dual core Penryn-based mobile** processors (Core 2 Duo Mobile) Intel introduced already a less intricate technology than the Turbo Boost technology for **utilizing available power headroom for raising single-thread performance**, termed as the (**EDAT**) **Enhanced Dynamic Acceleration Technology**, but only for **mobile processors**.
- EDAT's operation is based also on the **ACPI standard (Advanced Configuration and Power Interface)**.
- **Principle of operation:** If one of the dual cores is idle and given conditions are met, EDAT will **increase clock frequency of the active core by 1 bin** (typically 266 MHz for an FSB of 533 MHz or 333 MHz for an FSB of 666 MHz).
- The operation is controlled by **dedicated EDAT logic**.

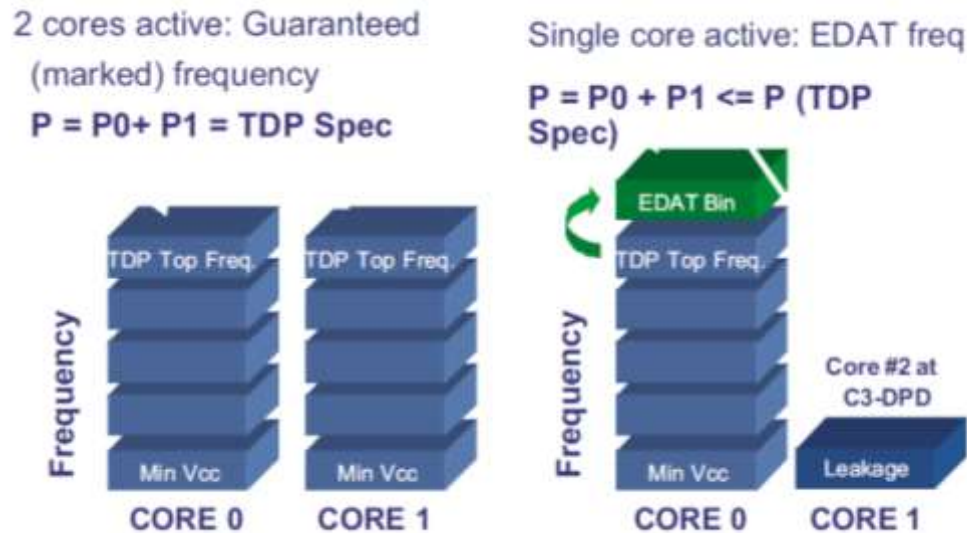
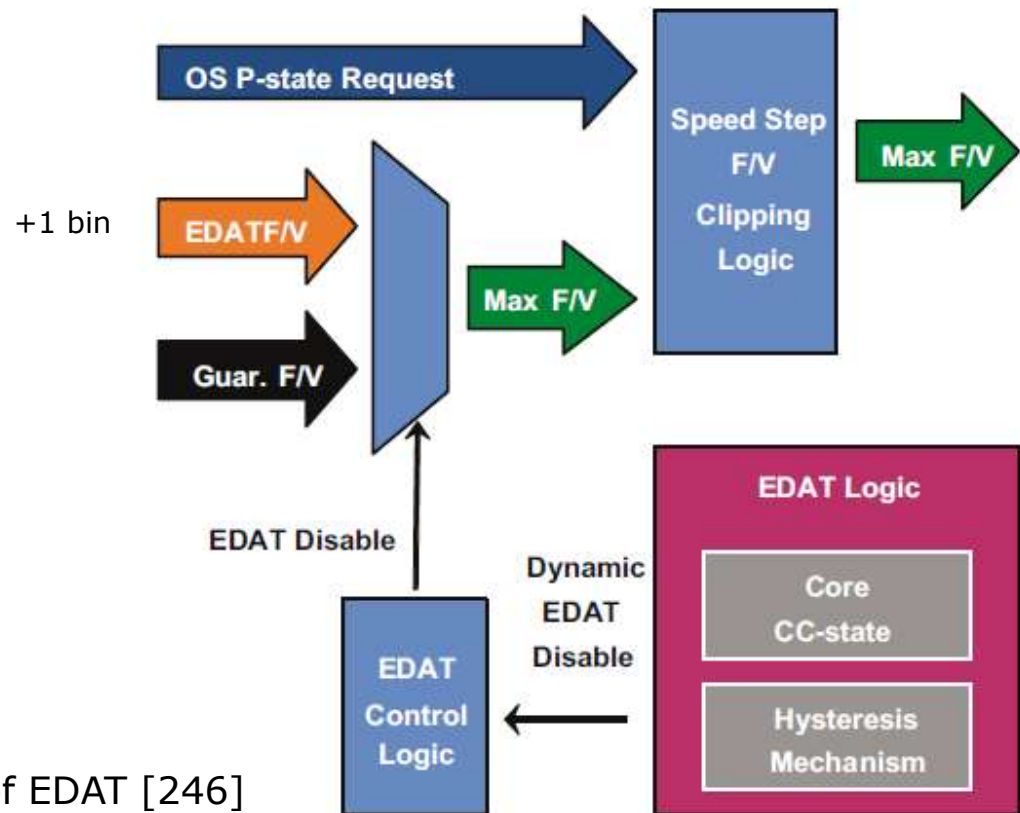


Figure: Principle of the operation of EDAT [246]

Implementation of EDAT

- EDAT logic considers a **core "active"** if it is in ACPI C0 or C1 states, whereas cores in the C3 to C6 ACPI states are considered as "idle".
- EDAT becomes activated if
 - one of the two cores becomes idle
 - the OS requests the highest P state for the active core and
 - power consumption remains below the TDP (Thermal Design Power).



CC: Core C-state
F/V: Frequency/Voltage

Figure: Principle of implementation of EDAT [246]

Extending the operation of EDAT to Penryn-based quad-core mobile processors

- Penryn based quad-core processors are in fact **MCMs (Multi Chip Modules)** with two chips properly interconnected mounted in the same package.
- In this case **each of both chips can activate EDAT independently from each other**, if one of their cores becomes idle, and the total power consumption remains below TDP.
- This technology is also designated as **Dual EDAT**.

Remark

Intel designates EDAT also as

- IDA (Intel Dynamic Acceleration Technology) in dual-core Penryn-based mobile processors or
- Dual Dynamic Acceleration Technology in quad-core (2x2-core) Penryn-based mobile processors.

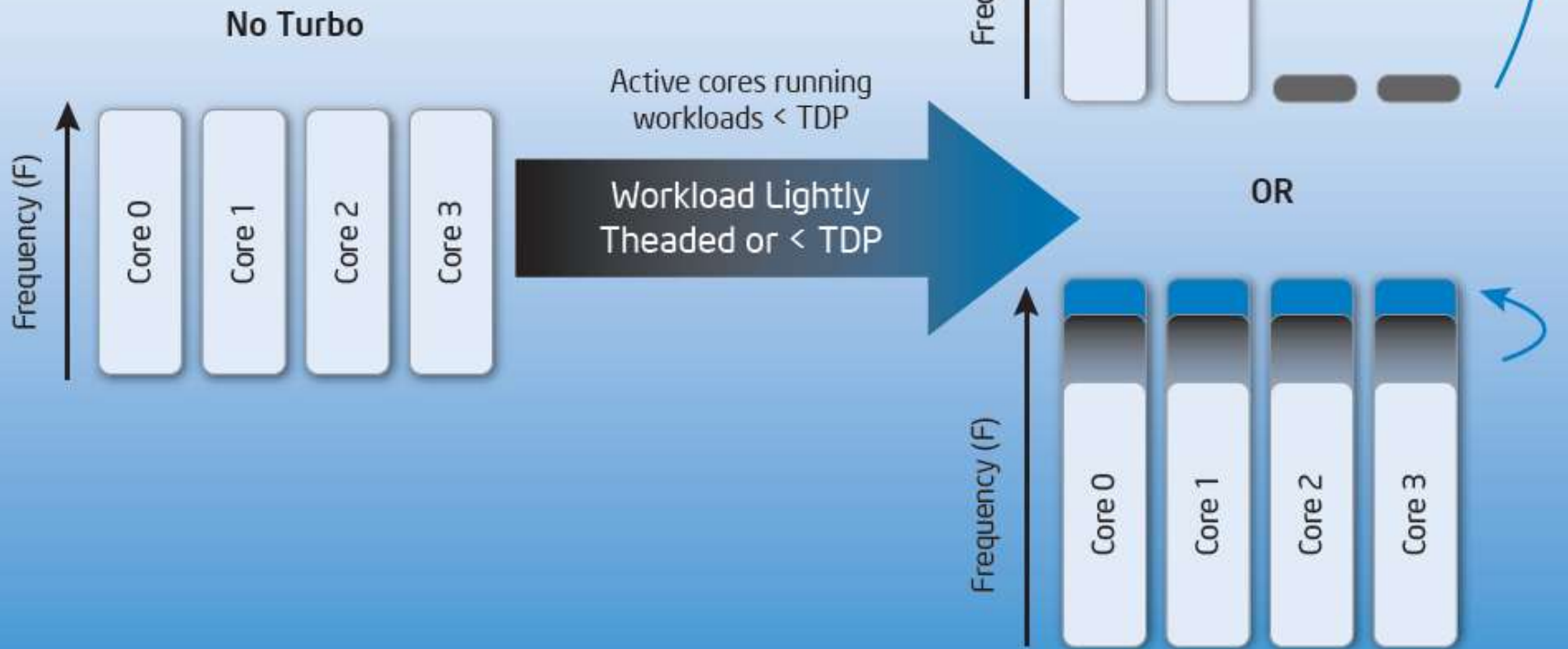
Nehalem's Turbo Boost technology as an enhancement of Penryn's EDAT

- Nehalem has already a **PCU (Power Control Unit)** that is responsible
 - for **controlling the core frequencies and core voltages** and also
 - for **checking the power dissipation of the whole package and if needed take appropriate actions.**
- Nehalem's **Turbo Boost** implementation **enhances EDAT's operation threefold:**
 - a) the PCU may increase the clock frequency of the active cores independently from the **number of active cores, even if all cores are active presuming a light workload, or more precisely that certain conditions to be discussed later, are met.**
 - c) The Turbo Boost technology is **no more restricted to mobile platforms,**
as discussed subsequently.

3.2.5.7 Nehalem's Turbo Boost technology (8)

Principle of the operation of Nehalem's Turbo Boost technology (2) [52]

Intel® Turbo Boost Technology



The Turbo mode uses the power headroom (unused power up to the TDP limit) of the proc. package.*

Precondition for activating the Turbo Boost technology

The PCU activates the Turbo Boost technology if

- the actual workload needs the highest performance state (P0),
- the actual power consumption is less than the TDP, (i.e. there is a power headroom)
- the actual current is less than a given limit and
- also the die temperature is below a given limit.

b) Principle of operation of Turbo Boost [51] -1

If the above conditions for activating the Turbo Boost technology are fulfilled

- the PCU automatically steps up core frequency in a closed loop by one bin (133.33 MHz for the Nehalem family) as long as it reaches the max. ratio of the frequency multiplier held in the MSR 1ADh for 1, 2, 3 or active cores, as seen in the next Figure.

The internal register MSR 1ADH is interpreted as follows:

MSR 1ADh (in the Nehalem family)			
Bits 31-24	Bits 23-16	Bits 15-08	Bits 07-00
Max. ratio with 4 active cores	Max. ratio with 3 active cores	Max. ratio with 2 active cores	Max. ratio with 1 active core
E.g. in the Nehalem DT i7-975 (Base clock: 3.33 GHz)			
Multiplier:	+1 bin	+1 bin	+1 bin
Max. turbo fc	3.46 GHz	3.46 GHz	3.6 GHz

- The actual turbo boost frequency results as the product of the given max. ratio times the bus clock frequency (133.33 MHz in this case).
- In each step the PCU sets the PLLs (Phase Locked Loop) of the cores and the VR (Voltage Regulator) to the appropriate values.

b) Principle of operation of Turbo Boost [51] -2

Maximum turbo frequencies are factory configured and kept in form of multiplier values in the internal registers (MSR 1ADH) of the processor, they can be read by the PCU or OS.

3.2.5.7 Nehalem's Turbo Boost technology (12)

Remarks

In subsequent processors the turbo mode achieved a higher clock boost, as seen below for a Sandy Bridge-E HED processor [243].

i7-3970X (Sandy Bridge-E), base clock: 3.50 GHz

No of active cores	1C	2C	3C	4C	5C	6C
Bins (1 bin: 100 MHz)	5	5	3	3	1	1
Turbo clock frequency	4.0 GHz	4.0 GHz	3.80 GHz	3.80 GHz	3.60 GHz	3.60 GHz

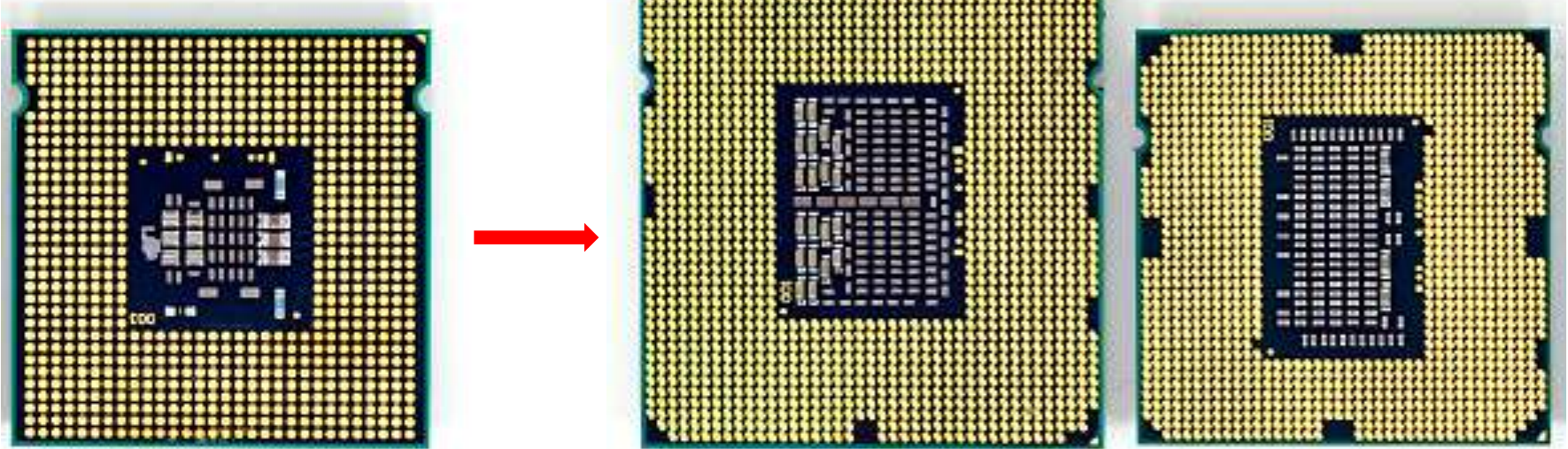
Determination of the number of active cores

- The PCU monitors the activity of all 4 cores.
- The PCU considers a core **active** if it is in the C0 (active) or C1 (Halt) state and **inactive** if it is in the C3 or the C6 state, it is the same differentiation as was done in EDAT.

Checking current, power consumption and temperature vs. specified limits [53], [50]

- To check power and temperature limits the PCU samples the current power consumption and die temperature in 5 ms intervals [53].
- Power consumption is determined by monitoring the processor current at its input pins as well as the associated voltage (V_{cc}) and calculating the power consumption as a moving average.
- The junction temperature of the cores are monitored by DTSs (Digital Thermal Sensors) with an error of $\pm 5\%$ [50].
- When any factory configured limit is surpassed (the power consumption of the processor or the junction temperature of any core) the PCU automatically steps down core frequency in increments of e.g. 133 MHz.

3.2.6 New sockets [167]



**LGA-775
(Core 2)**

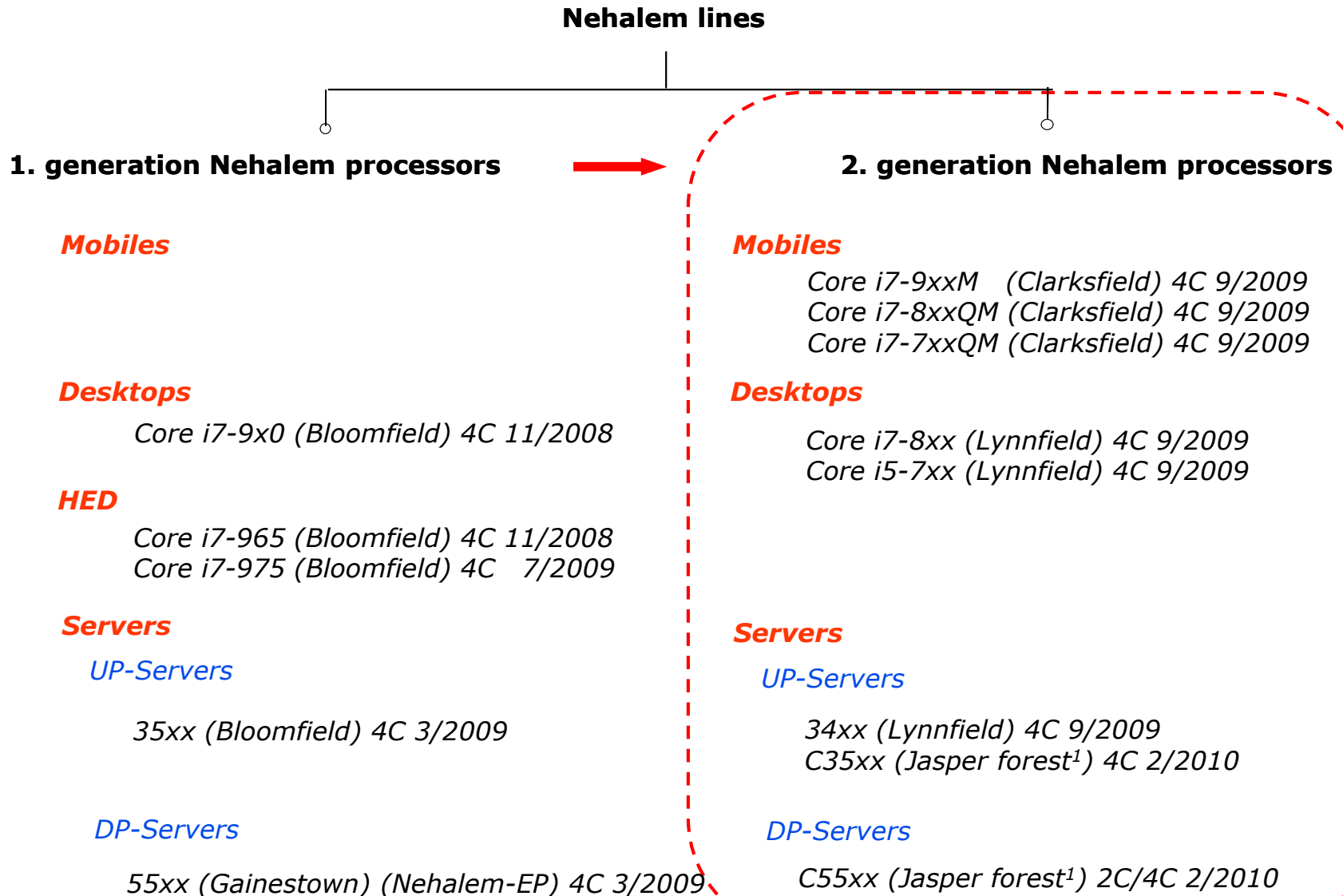
**LGA-1366
(Bloomfield)**

**LGA-1156
(Lynnfield)**

A new socket became necessary since [attaching three DDR3 memory channels needs 3x240 additional lines.](#)

3.3 Major innovations of the 2. generation Nehalem line (Lynnfield)

3.3 Major innovations of the 2. generation Nehalem line (Lynnfield) (1)



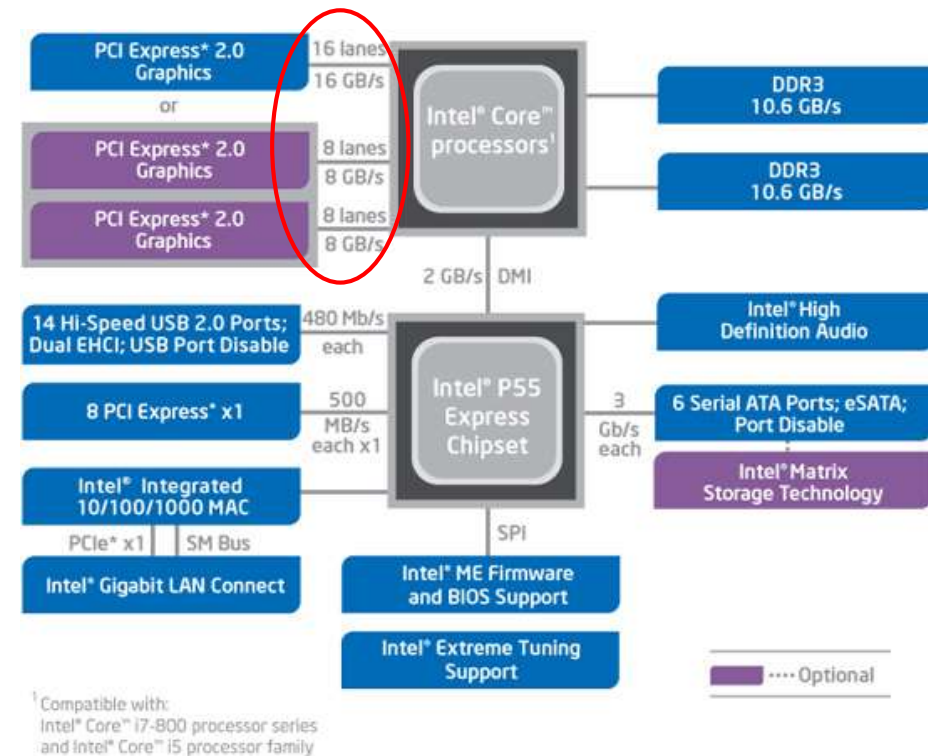
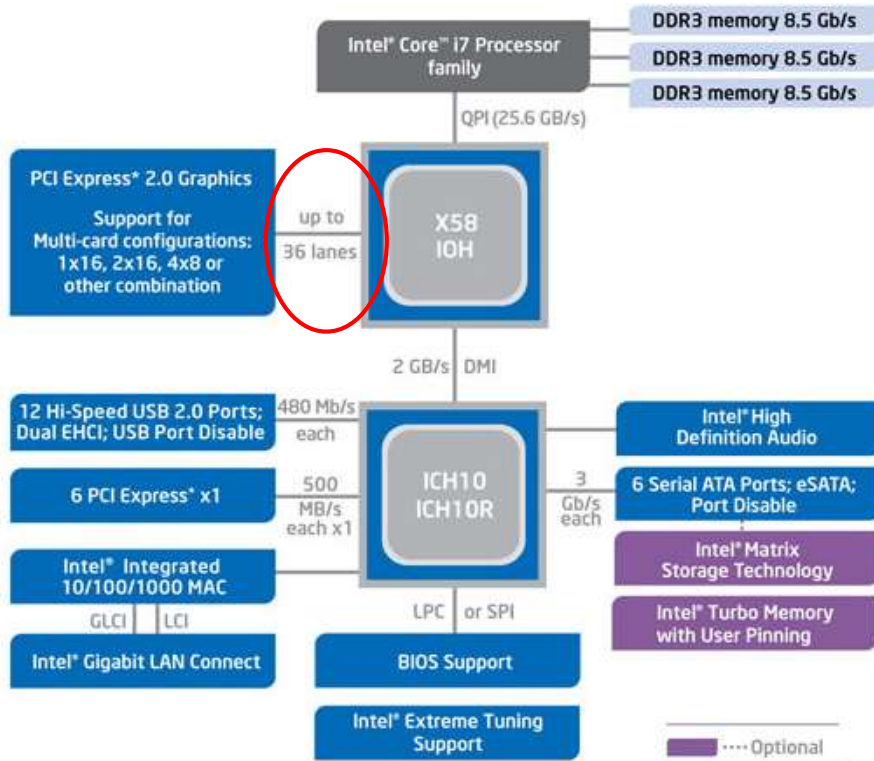
3.3 Major innovations of the 2. generation Nehalem line (Lynnfield) (2)

Major innovations of the 2. generation Nehalem line (Lynnfield) (1) [46]

The Lynnfield chip is a **major redesign** of the Bloomfield chip targeting desktops and laptops, resulting in a **cheaper** and **more efficient two-chip system solution**.

Major innovations

- a) It provides **only 16 PCIe 2.0 lanes** rather than 36 lanes for attaching graphics cards. **PCIe lanes are attached immediately to the processor** rather than to the north bridge, as in the previous generation.



The Bloomfield based platform (X58 + ICH10 / LGA-1366)

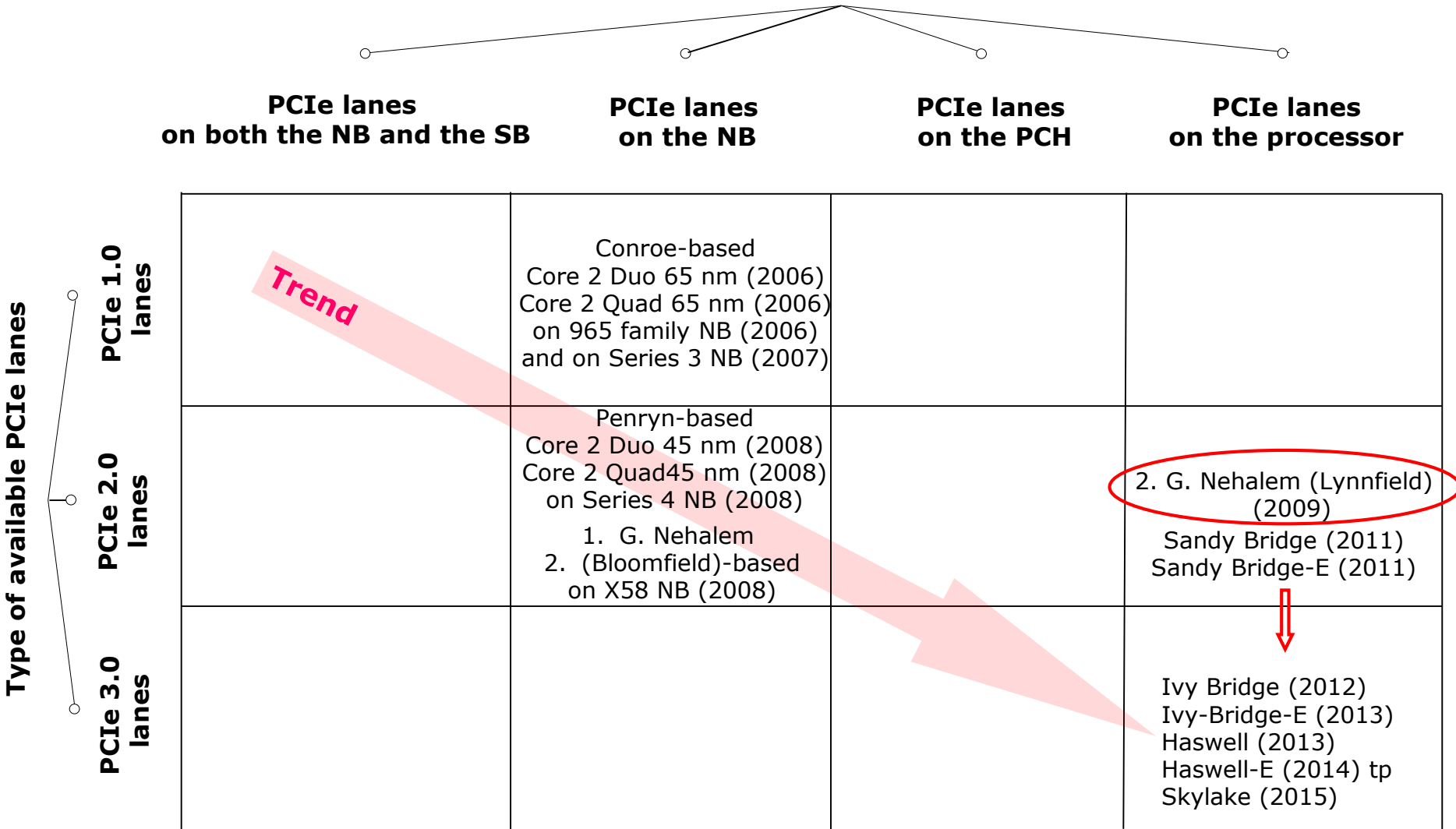
The Lynnfield based platform (P55 / LGA-1156)



3.3 Major innovations of the 2. generation Nehalem line (Lynnfield) (3)

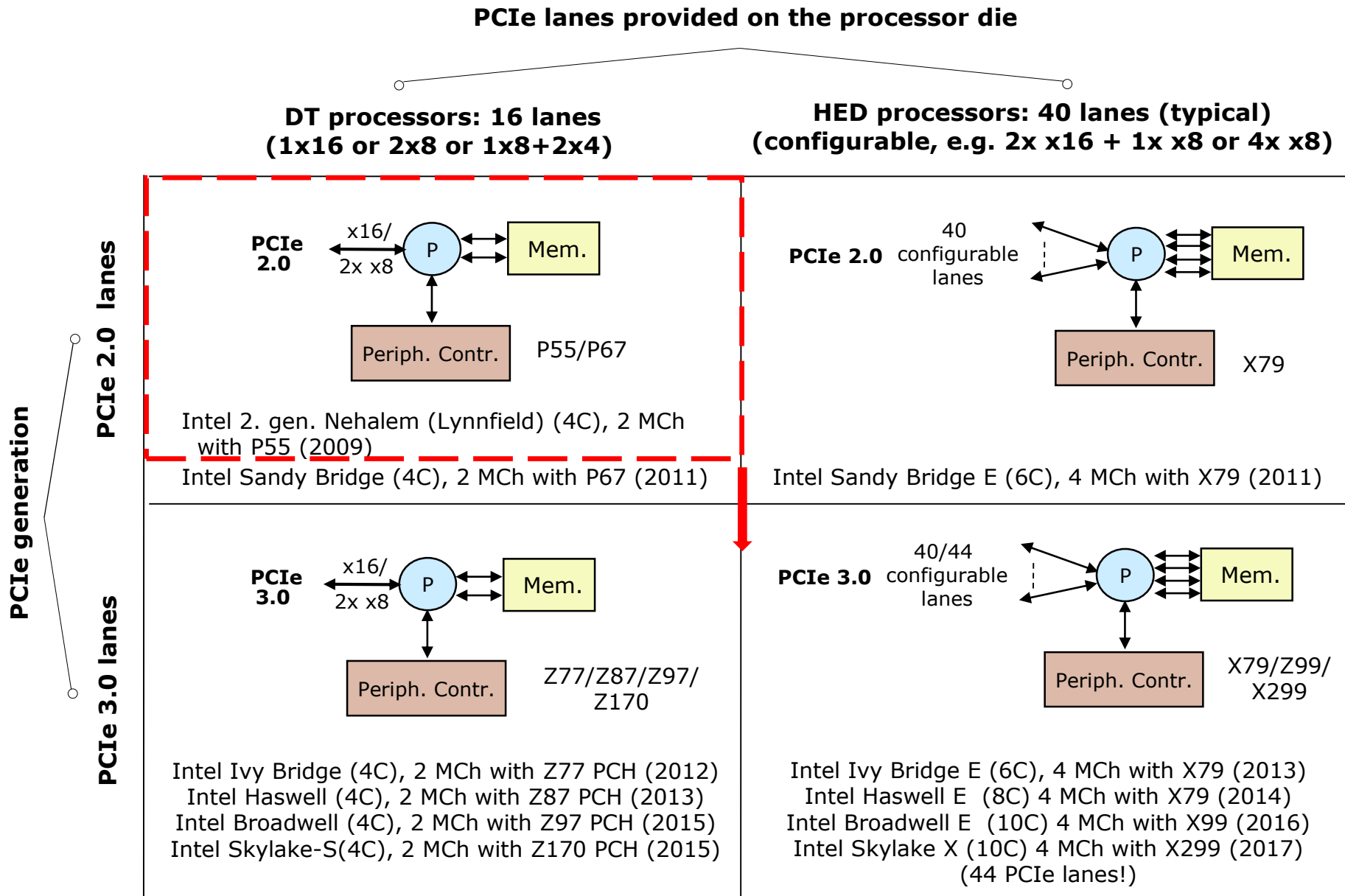
Evolution of the topology and type of available PCIe lanes for graphics cards

Topology of PCIe lanes provided for graphics cards



3.3 Major innovations of the 2. generation Nehalem line (Lynnfield) (4)

Number of on-die memory channels and PCIe lanes provided on Intel's DT and HED lines

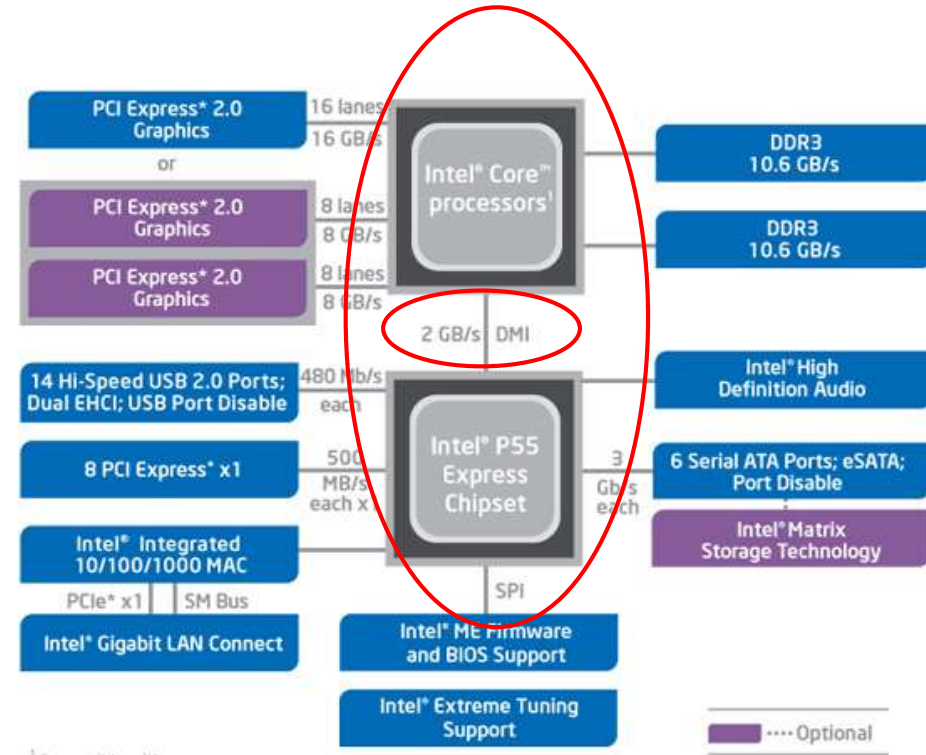
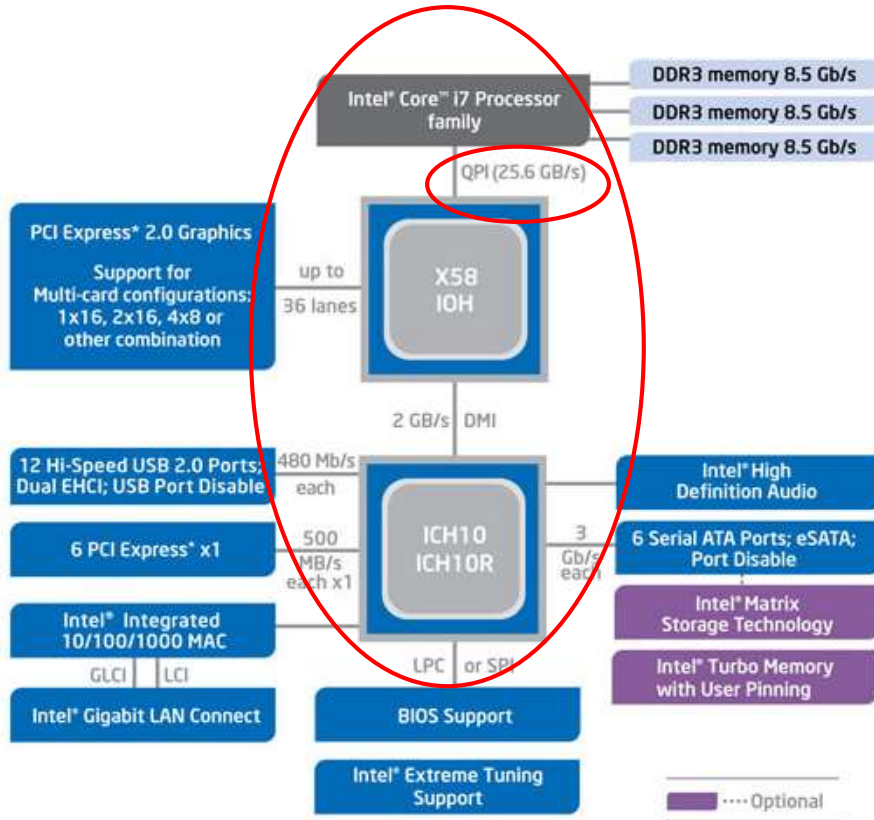


*

3.3 Major innovations of the 2. generation Nehalem line (Lynnfield) (5)

Major innovations of the 2. generation Nehalem line (Lynnfield) (2) [46]

- b) While connecting PCIe lanes directly to the processor, the previous north bridge has less functions and thus it can be integrated with the south bridge, to a PCH (Peripheral Control Hub), yielding a two chip solution.
- c) While connecting PCI lanes directly to the processor less bandwidth is needed between the processor and the PCH, thus, the high bandwidth QPI bus can be replaced by a DMI interface (i.e. by 4 PCIe lanes).



¹ Compatible with: Intel Core i7-B00 processor series and Intel Core i5 processor family

The Bloomfield based platform (X58 + ICH10 / LGA-1366)

The Lynnfield based platform (P55 / LGA-1156)

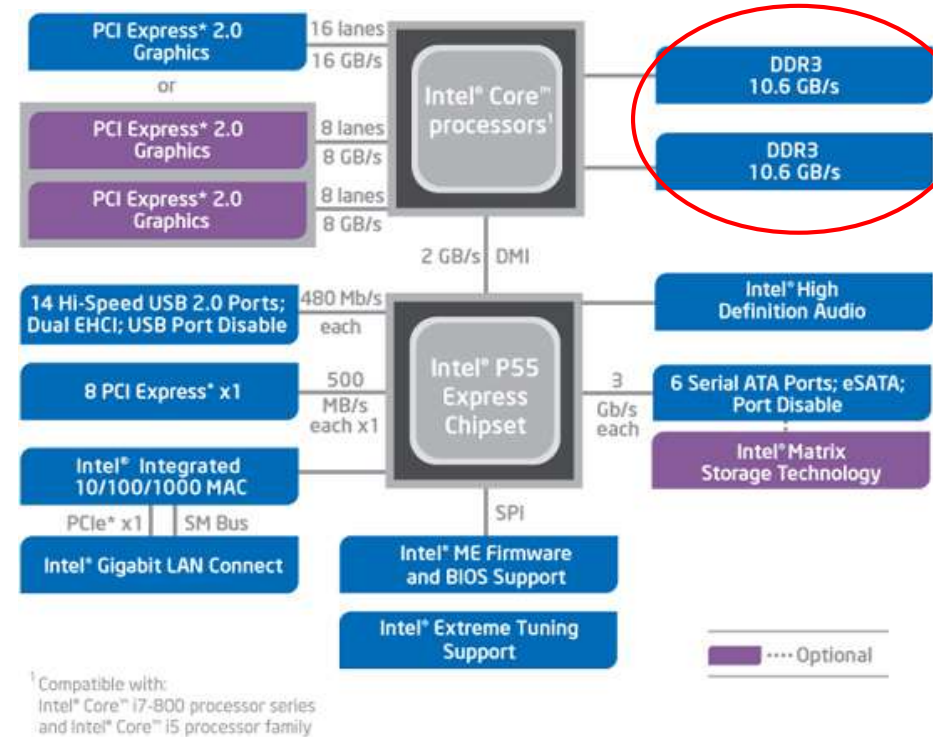
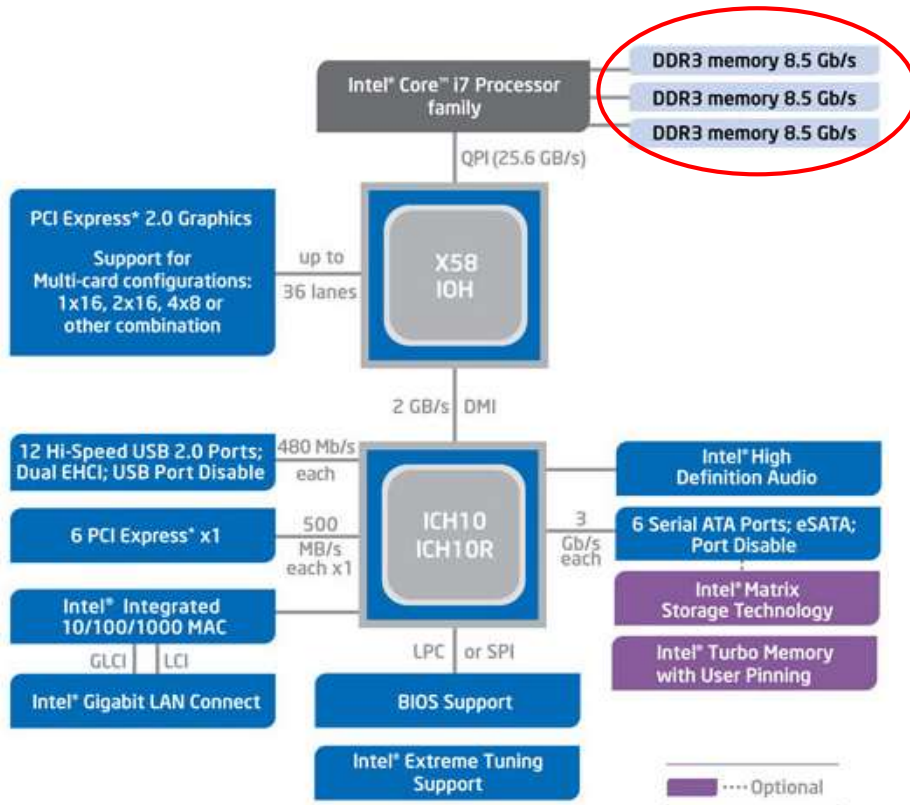


3.3 Major innovations of the 2. generation Nehalem line (Lynnfield) (6)

Major innovations of the 2. generation Nehalem line (Lynnfield) (3) [46] (cont.)

- d) It supports only **two DDR3 memory channels** instead of three as in the previous solution.
- e) Its socket needs **less connections (LGA-1156)** than the Bloomfield chip (LGA-1366).

All in all the Lynnfield chip is a **cheaper and more effective successor** of the Bloomfield chip, aiming primarily **for mobiles and desktops**

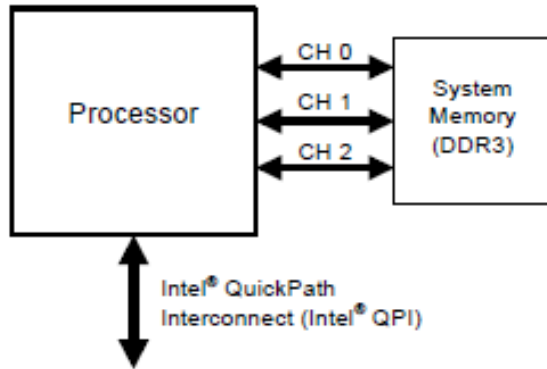


The Bloomfield based platform (X58 + ICH10 / LGA-1366)

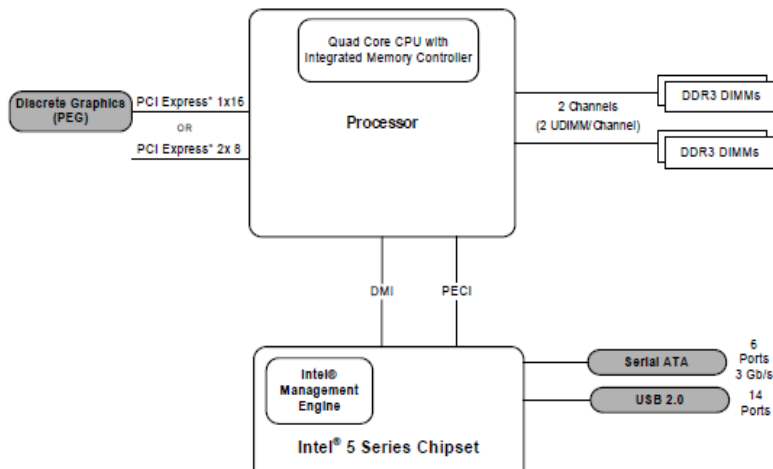
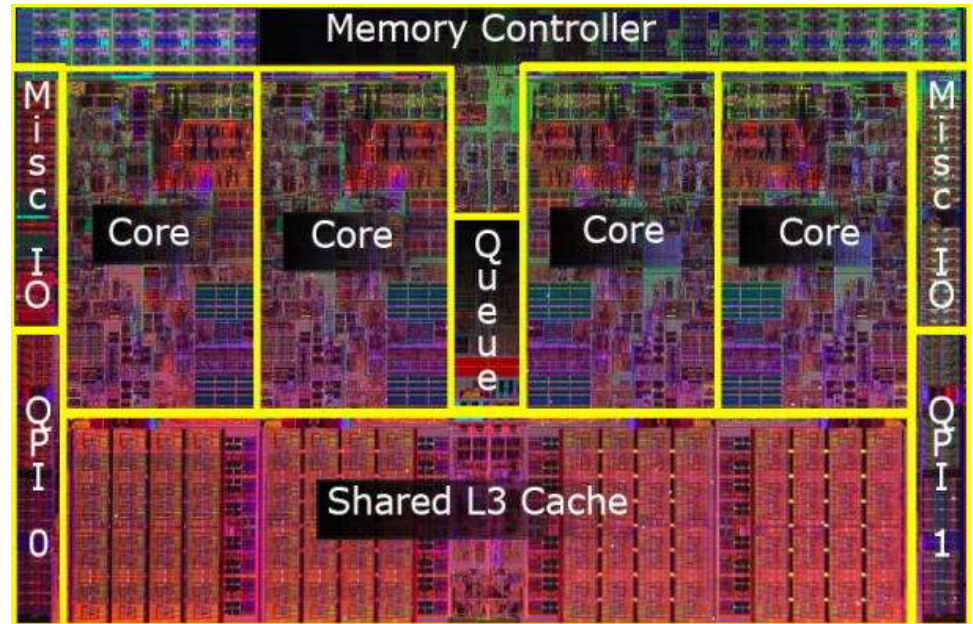
The Lynnfield based platform (P55 / LGA-1156)

3.3 Major innovations of the 2. generation Nehalem line (Lynnfield) (7)

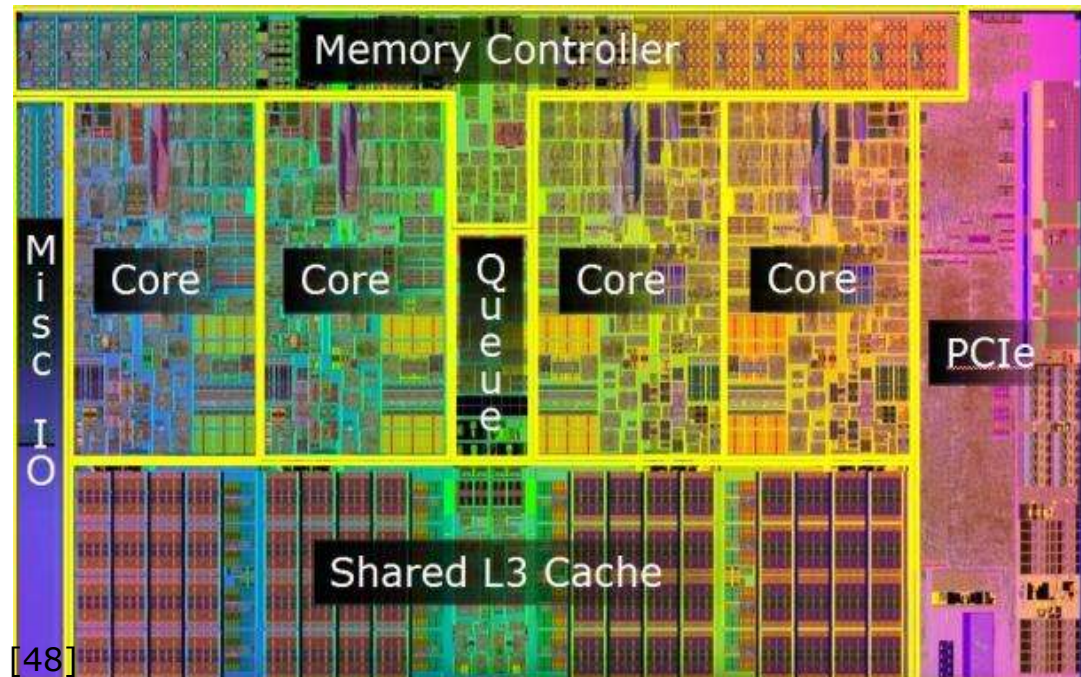
Die photos of the 1. and 2. gen. Nehalem desktop chips



First generation: **Bloomfield** chip (11/2008)
 (45 nm, 263 mm², 731 mtrs, LGA-1366)
 [45], [46], [47]



Second generation: **Lynnfield** chip (9/2009)
 45 nm, 296 mm², 774 mtrs, LGA-1156) [45] [46] [48]



4. The Sandy Bridge line

- 4.1 Introduction
- 4.2 Major innovations of the Sandy Bridge line vs. the 1. generation Nehalem line
- 4.3 Example for a Sandy Bridge based desktop platform with the H67 chipset

4.1 Introduction to the Sandy Bridge line

4.1 Introduction to the Sandy Bridge line (1)

4.1 Introduction to the Sandy Bridge line

- **Sandy Bridge** is Intel's next **new microarchitecture** using **32 nm** line width.
- Designed by Intel's Haifa design center, originally called Gesher.
- First **delivered in 1/2011**.
- It is termed also as Intel's **second generation Core processors**.

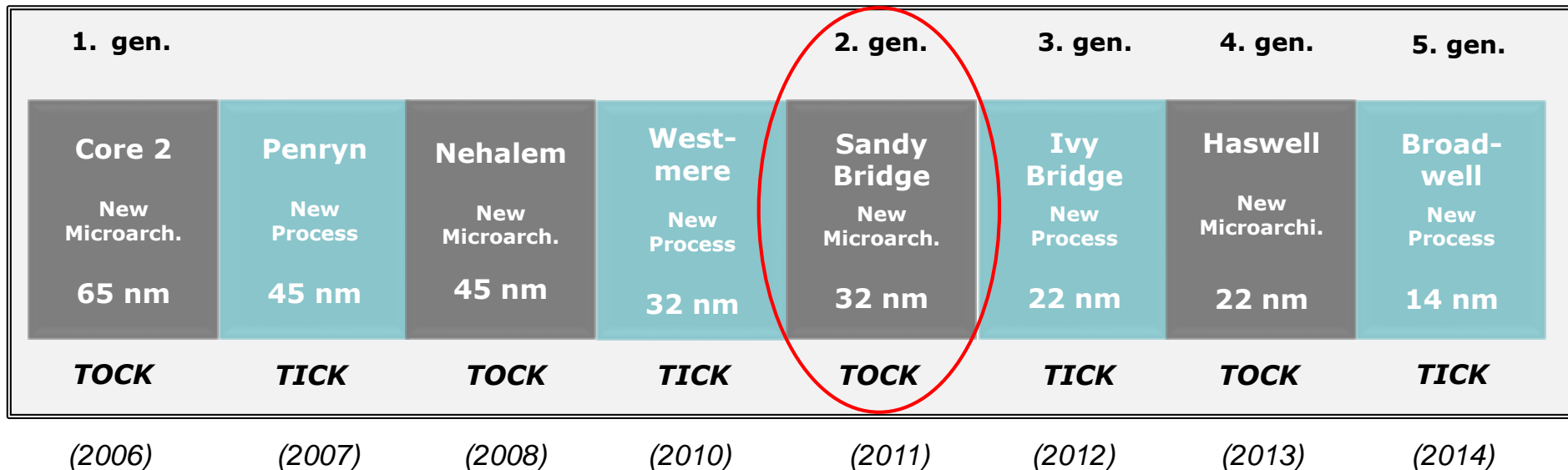
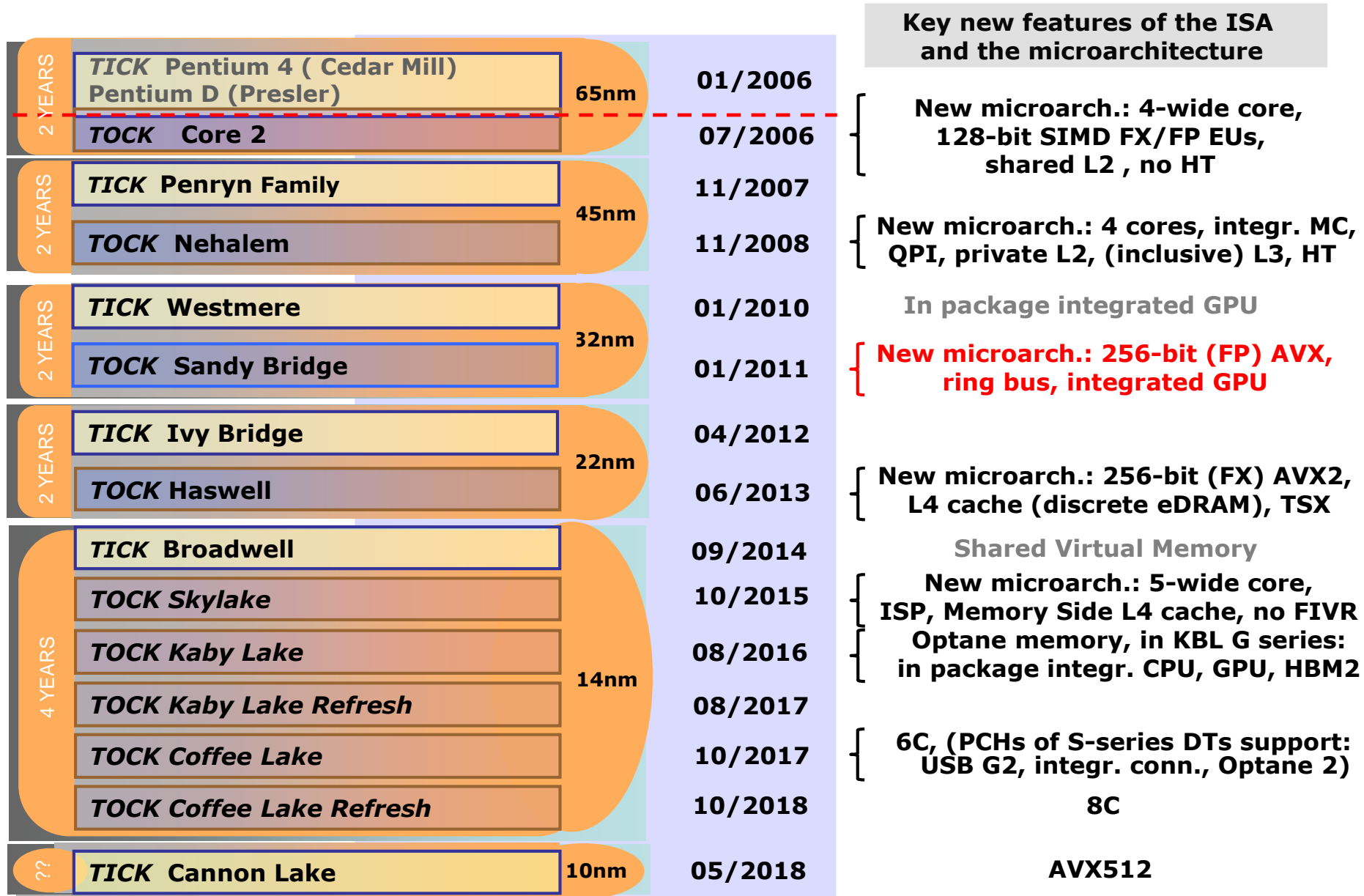


Figure : Intel's Tick-Tock development model (Based on [1])

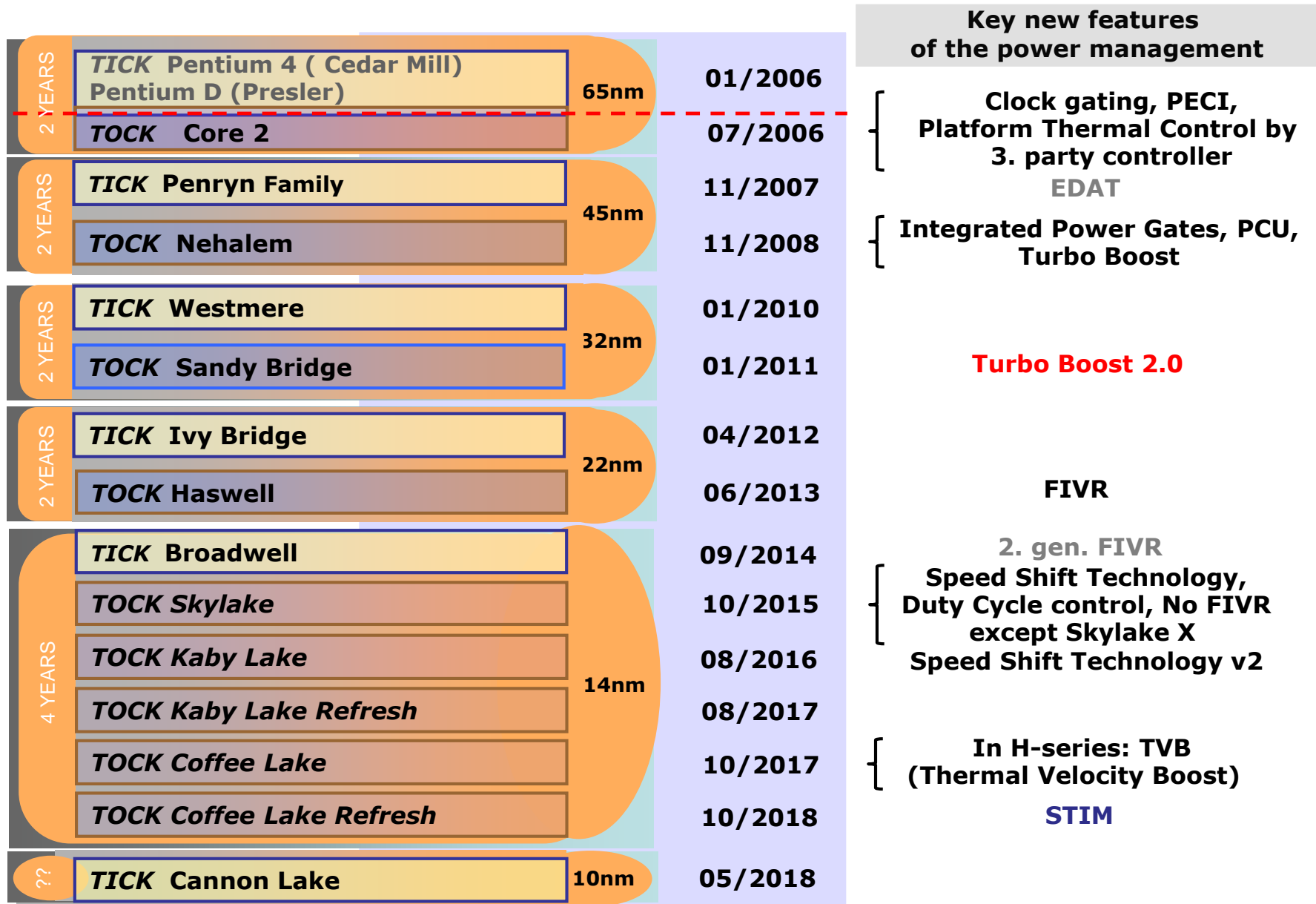
4.1. Introduction to the Sandy Bridge line (2)

The Sandy Bridge line -1 (based on [3])

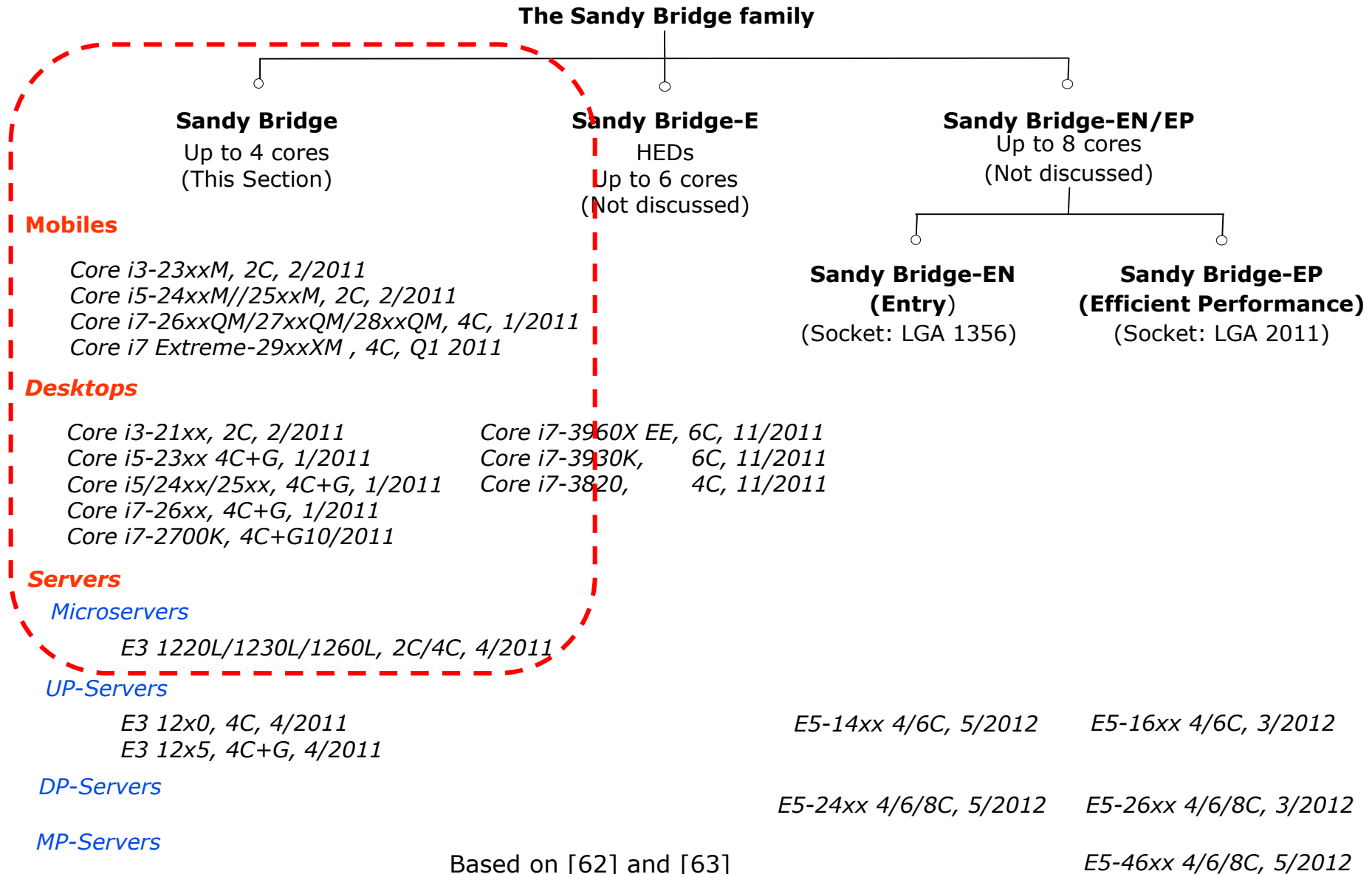


4.1. Introduction to the Sandy Bridge line (3)

The Sandy Bridge line -2 (based on [3])

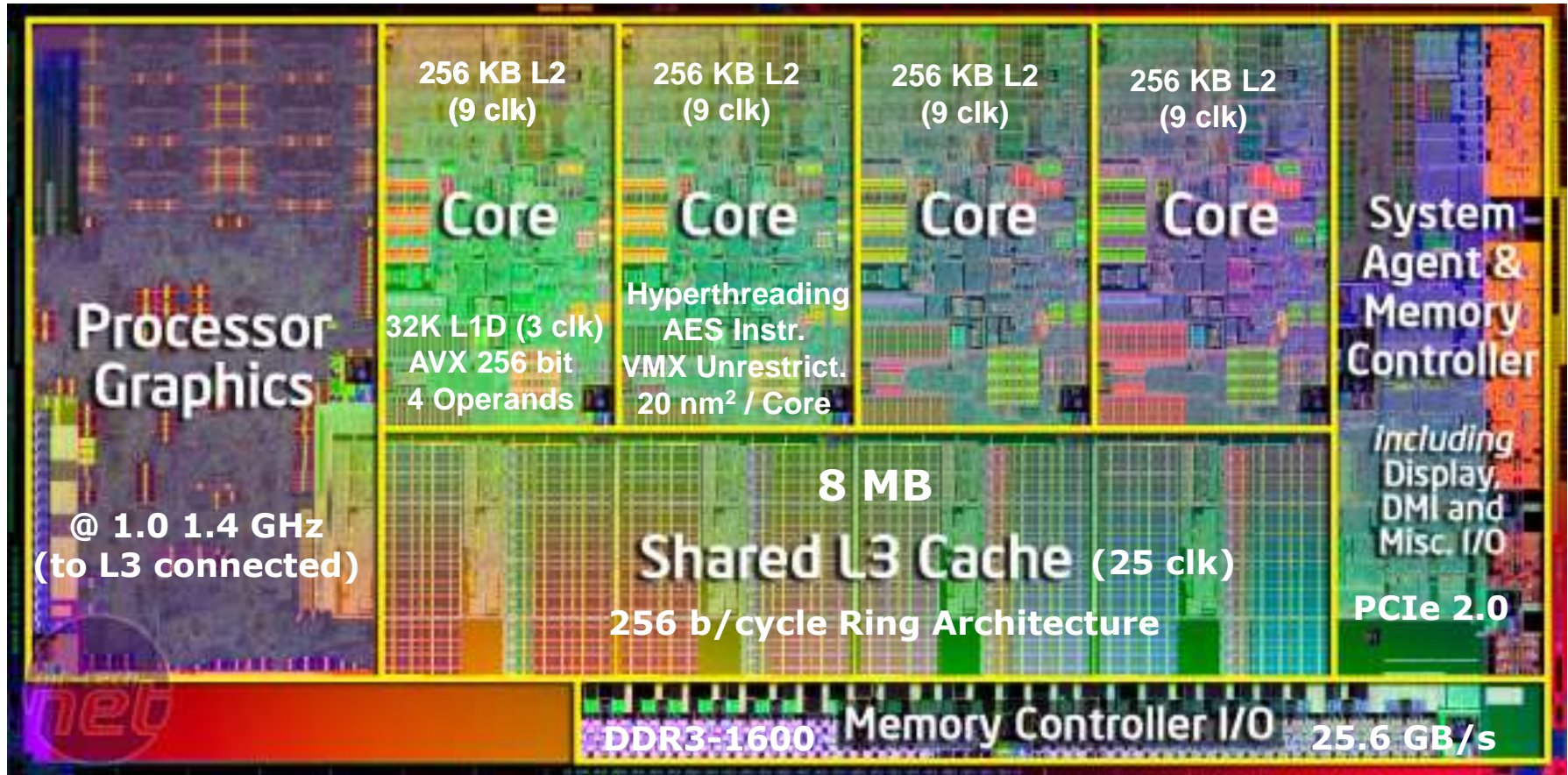


Overview of the Sandy Bridge family



4.1 Introduction to the Sandy Bridge line (5)

Main functional units of Sandy Bridge [96]



32 nm process / ~225 nm² die size. 995 mtrs, 85W TDP

Remark

Intel designates the integrated GPU as Processor Graphics (PG)

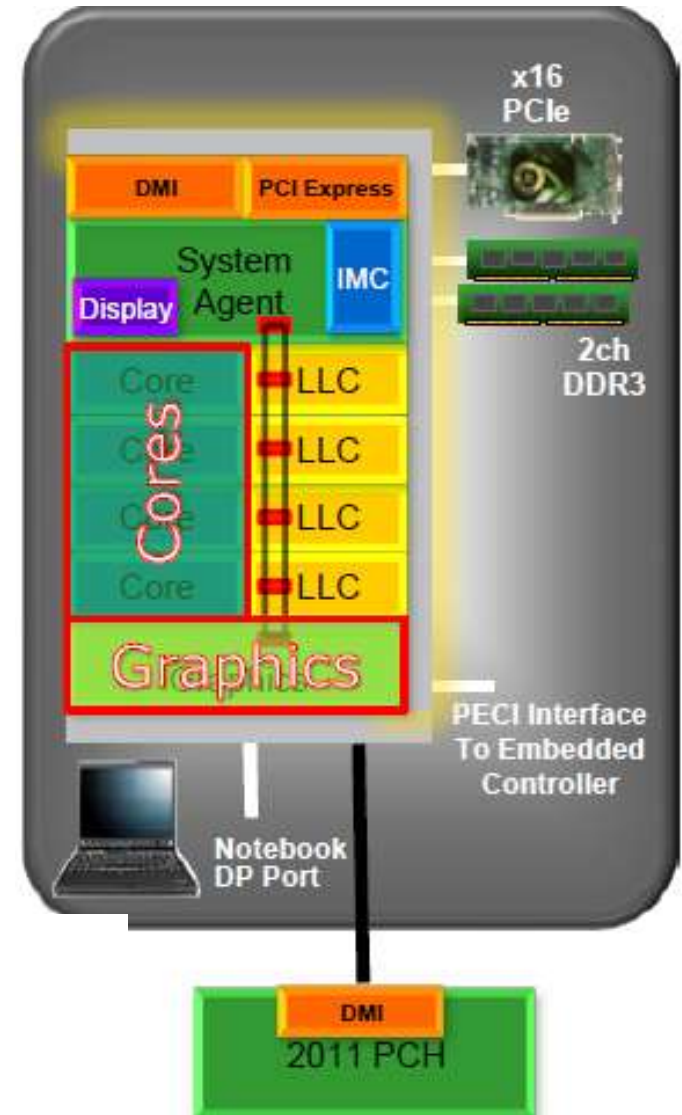
4.2 Major innovations of the Sandy Bridge line vs. the 1. generation Nehalem line

- 4.2.1 Overview
- 4.2.2 Extension of the ISA by the 256-bit AVX instruction set
- 4.2.3 New microarchitecture of the cores
- 4.2.4 On die ring interconnect bus
- 4.2.5 On die graphics unit
- 4.2.6 Turbo Boost 2.0 technology

4.2 Major innovations of the Sandy Bridge line vs. the 2. generation Nehalem line [61]

4.2.1 Overview

- Extension of the ISA by the 256-bit AVX instruction set (Section 4.2.2)
- New microarchitecture for the cores (Section 4.2.3)
- On die ring interconnect bus (Section 4.2.4)
- On-die graphics unit (Section 4.2.5)
- Turbo Boost technology 2.0 (Section 4.2.6)



4.2.2 Extension of the ISA by the 256-bit AVX instruction set

AVX: Advanced Vector Extensions

In the course of ISA extensions Intel expanded the previous **128-bit SSE SIMD instruction set** (introduced with the Pentium III in 1999) by the **256-bit AVX SIMD instruction set in the Sandy Bridge**, as follows:

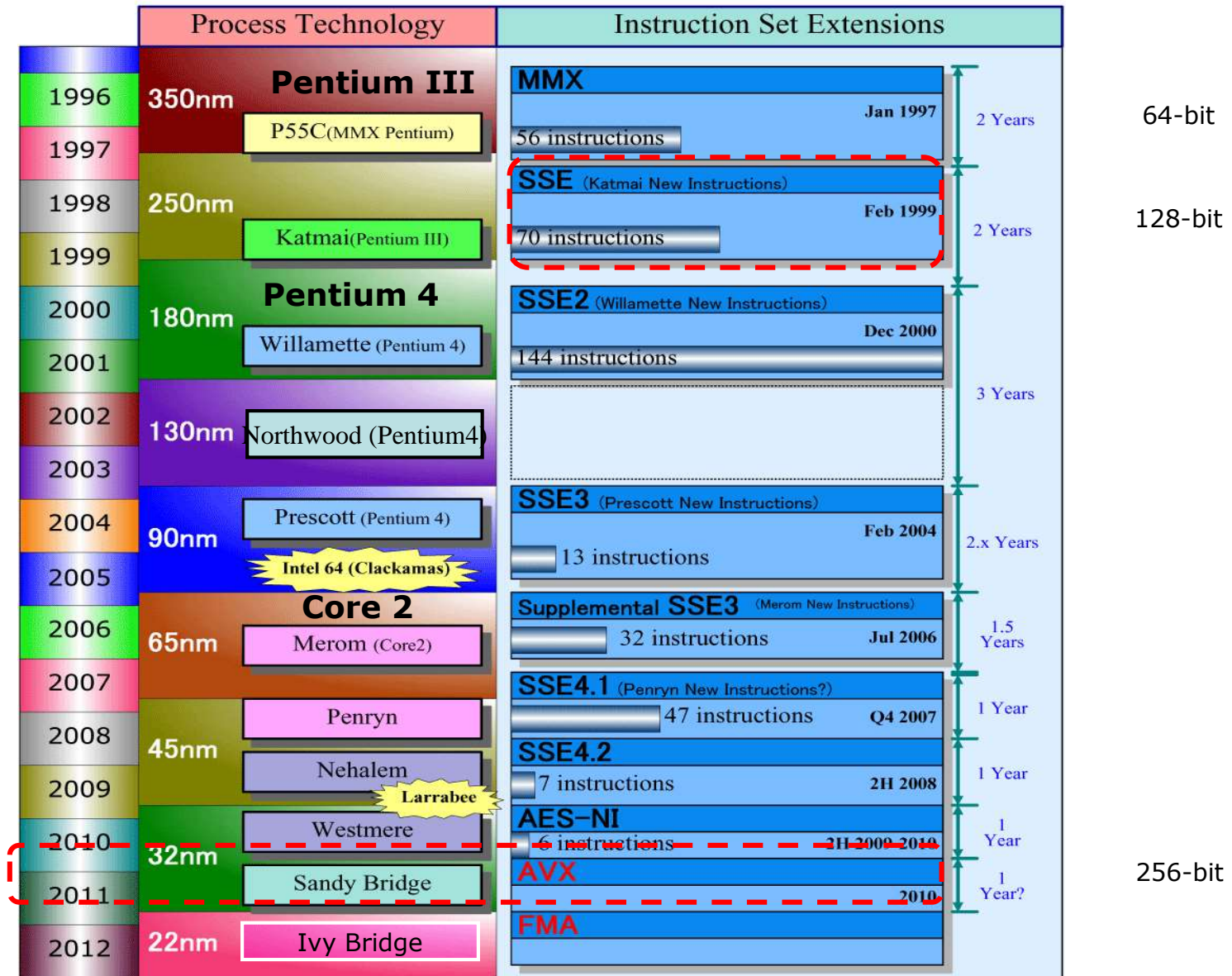
Remark

The 256-bit AVX instruction set is then expanded to the **AVX512** instruction set

- in the 14 nm Skylake-SP server processor (2017) and
- in the 10 nm Cannon Lake mobile (notebook) processor (2018)

4.2.2 Extension of the ISA by the AVX instruction set (2)

Width of Intel's subsequent SIMD extensions (Based on [18])



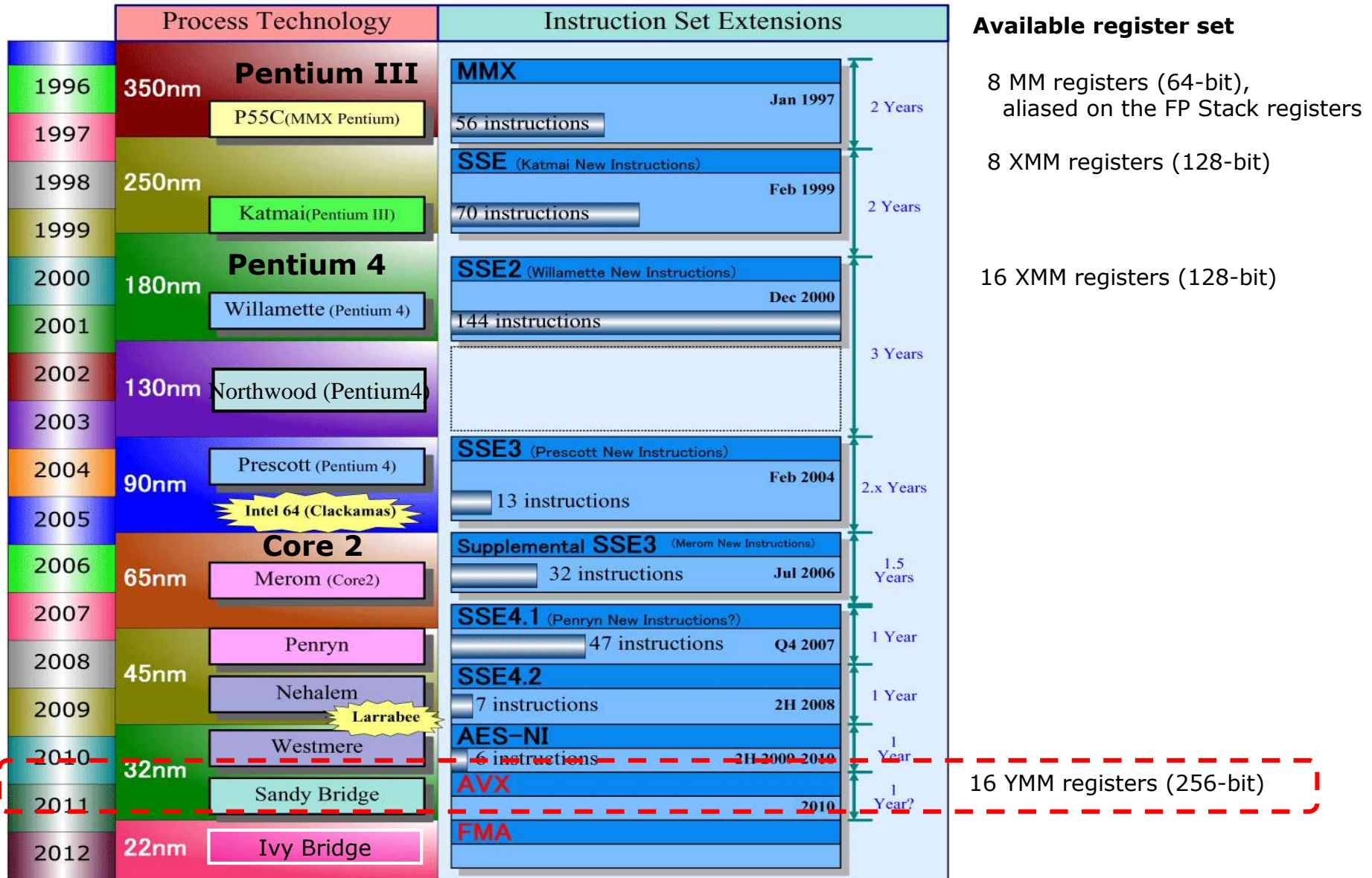
The 256-bit AVX extension

It includes

- a) **Extension**
of the 128-bit wide XMM [0, 15] SIMD **register set** to the 256-bit YMM [0, 15] register set.
- b) **Extension**
of the 128-bit SSE **instruction set** to the 256-bit instruction set.

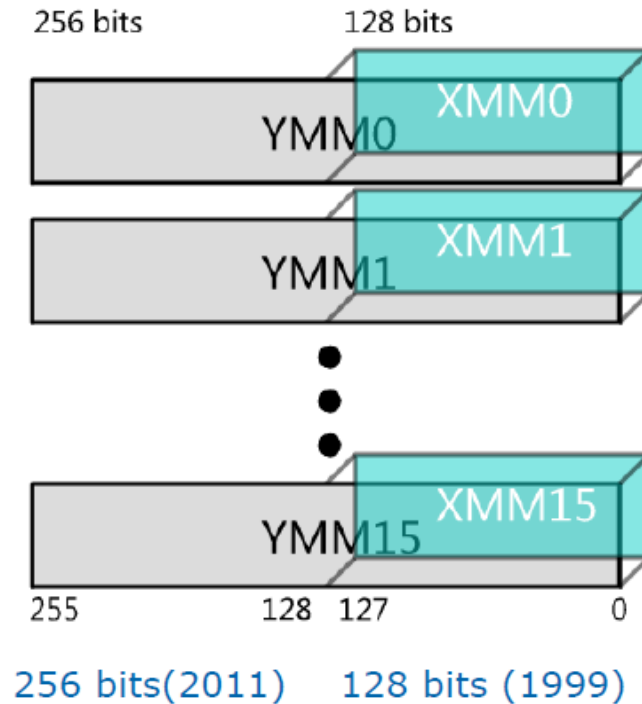
4.2.2 Extension of the ISA by the AVX instruction set (4)

Available SIMD register sets in Intel's subsequent SIMD extensions (Based on [18])



4.2.2 Extension of the ISA by the AVX instruction set (5)

- a) Extension of the 128-bit wide XMM [0, 15] SIMD register set to the 256-bit YMM [0, 15] register set [97], [168]



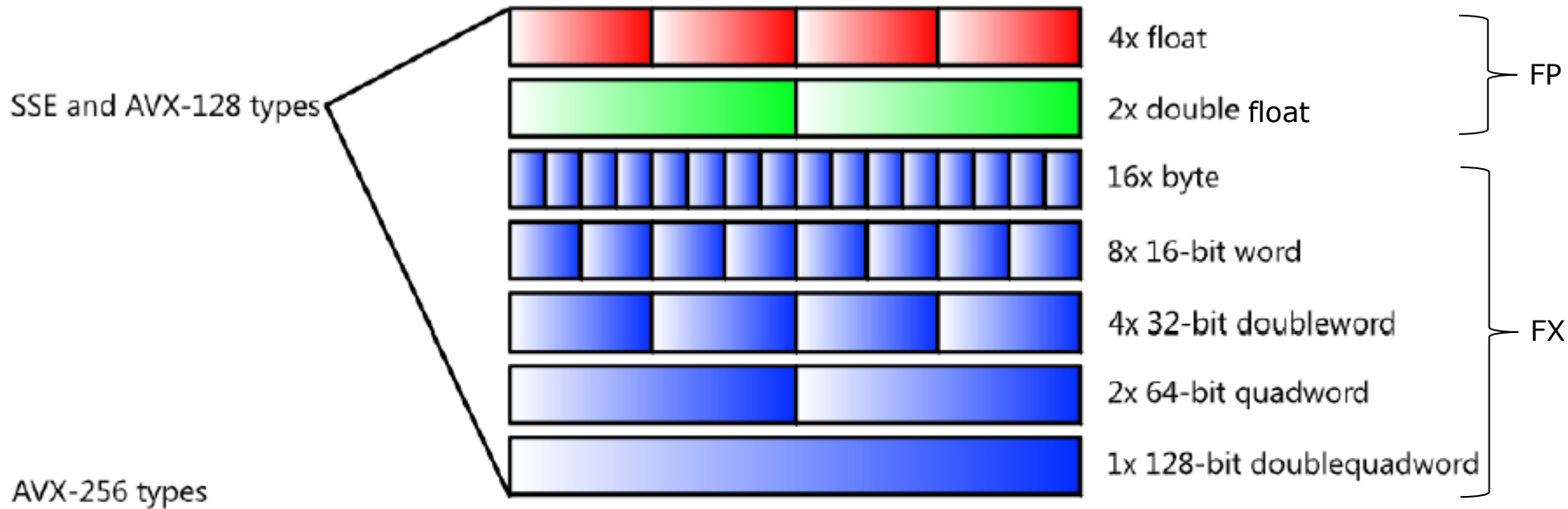
Intel AVX instructions operate on either:

- The whole 256-bits (FP only)
- The lower 128-bits (like existing Intel® SSE instructions)
 - The upper 128-bits of the register are zeroed out

4.2.2 Extension of the ISA by the AVX instruction set (6)

b) Extension of the 128-bit SSE instruction set

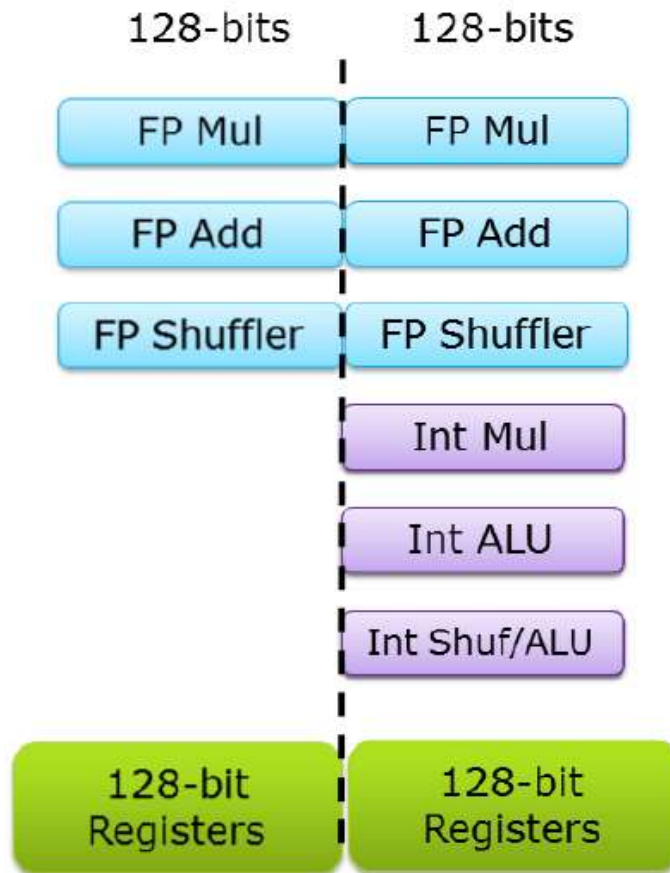
Supported data types [168]



4.2.2 Extension of the ISA by the AVX instruction set (7)

Note

AVX doubled **only FP vector width**, as indicated in the Figure below [97].



AVX doubled FP vector width
and register file width

→ Doubling peak FLPOPS

IDF2012

Implementation of AVX -1

- To implement 256-bit FP operations Intel did not widen related data paths and FP execution units to 256 bit, instead designers made use of two 128-bit data paths and two 128-bit FP execution units in the same time, as indicated in the next Figure [98].
- Sandy Bridge do not support FMA operations but it can execute up to 8 DP FP or 16 SP FP operations (additionally 4 DP SP operations or 8 SP FP operations can be executed over the Port 5).

4.2.2 Extension of the ISA by the AVX instruction set (9a)

Implementation of AVX -2

Intel redesigned large parts of the microarchitecture of the cores, as indicated by yellow boxes in the Figure below.

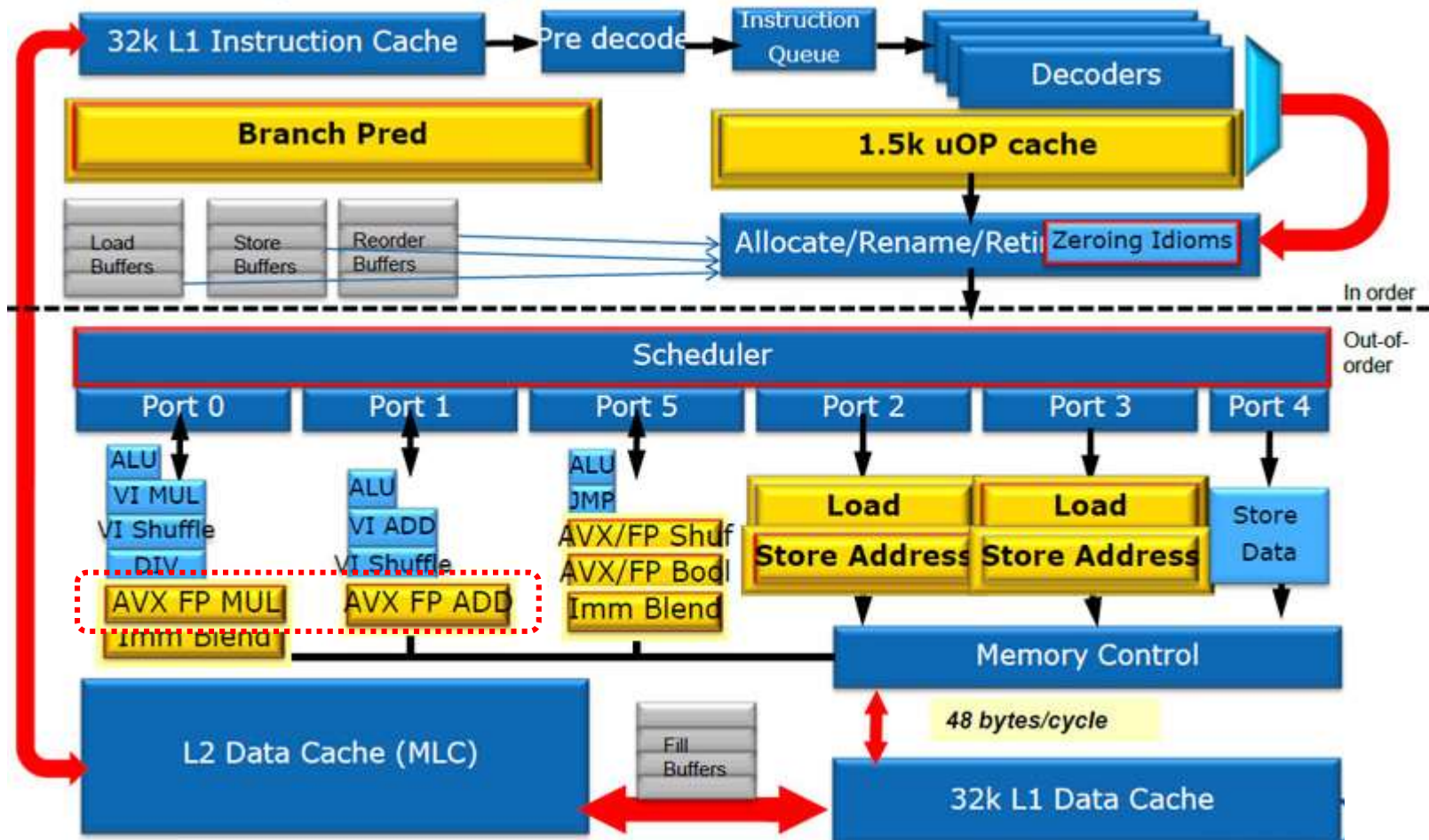
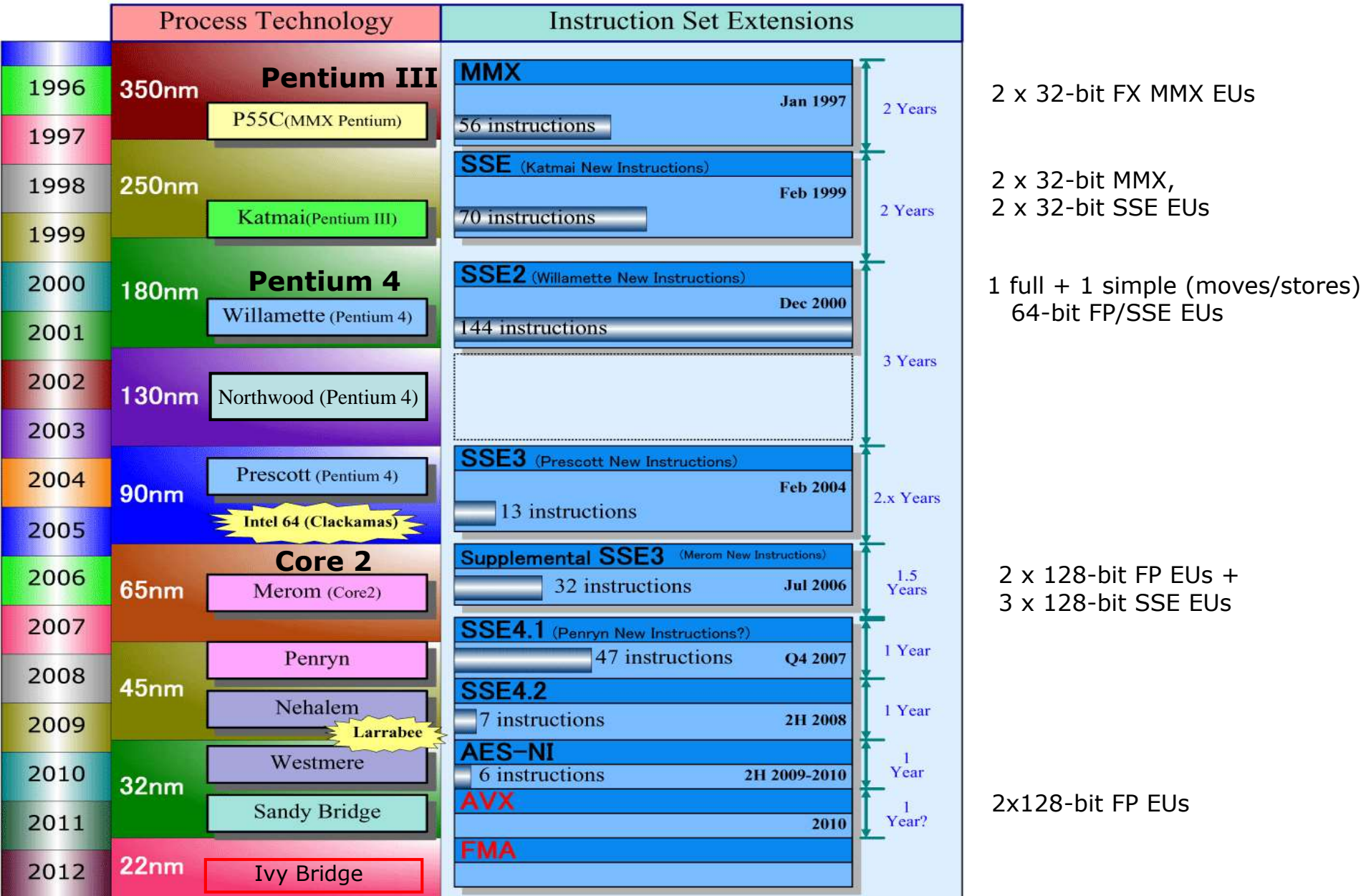


Figure 4.2.2.1: Microarchitecture of the cores of Sandy Bridge [64]

4.2.2 Extension of the ISA by the AVX instruction set (9b)

SIMD execution resources in Intel's basic processors (based on [18])



4.2.3 New microarchitectures of the cores (1)

4.2.3 New microarchitecture of the cores -1

Intel redesigned large parts of the microarchitecture of the cores, as partly indicated by yellow boxes in the Figure below.

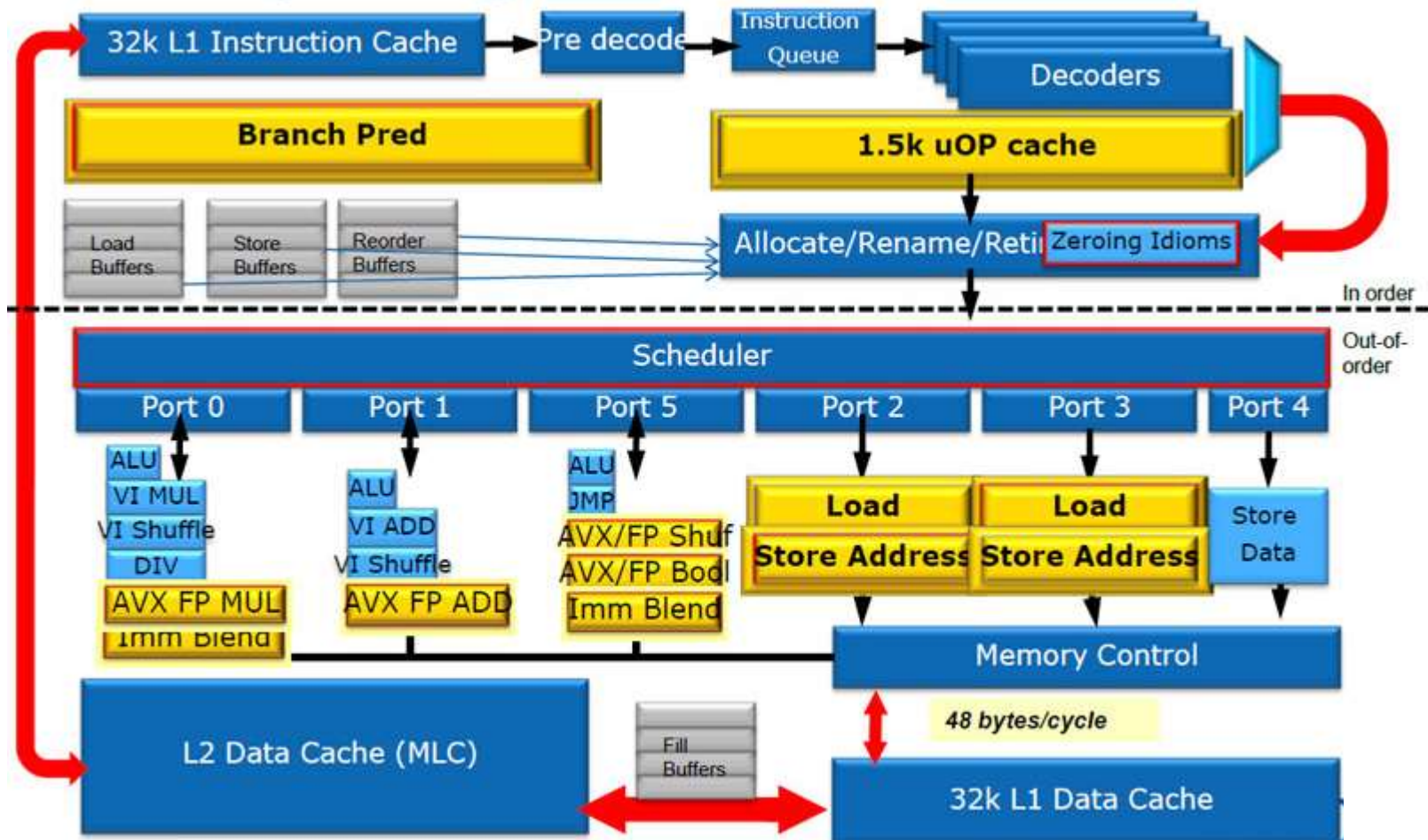


Figure: Microarchitecture of the cores of Sandy Bridge [64]

4.2.3 New microarchitecture of the cores - 2

There are **three major enhancements** of the microarchitecture, as follows:

- a) Using merged architectural and rename registers (aka physical registers) for renaming rather than the ROB, and
- b) changing the operand fetch policy from the dispatch bound to the issue bound scheme and
- c) Introducing a micro-op cache.

4.2.3 New microarchitectures of the cores (3)

a) **Using merged architectural and rename registers (aka physical registers) for renaming rather than the ROB -1**

Prior to the Core 2 line Intel [made use of the ROB](#) for renaming, as shown in the next slide.

4.2.3 New microarchitectures of the cores (4)

Using the ROB for renaming

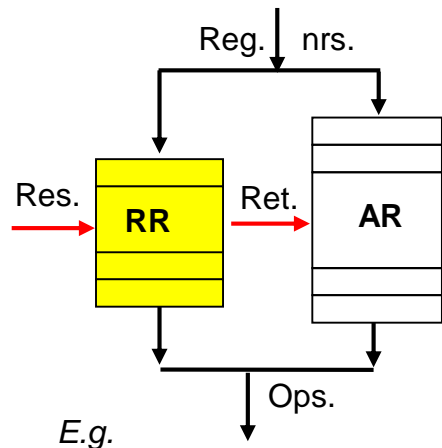
Types of rename buffers

Rename reg. file

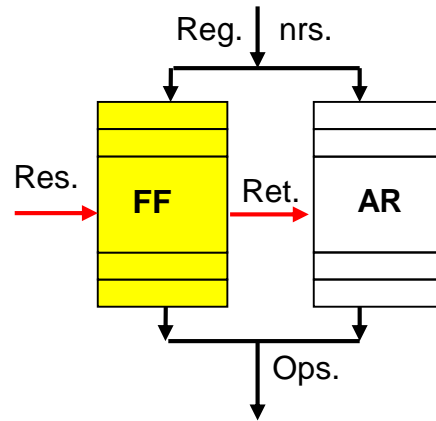
Future file

Merged arch. and rename register file

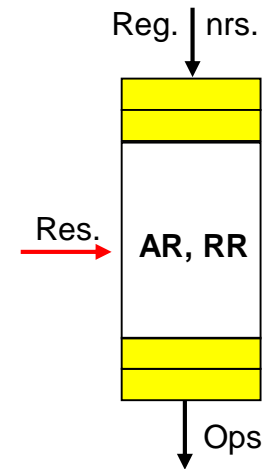
Holding renamed values in the ROB



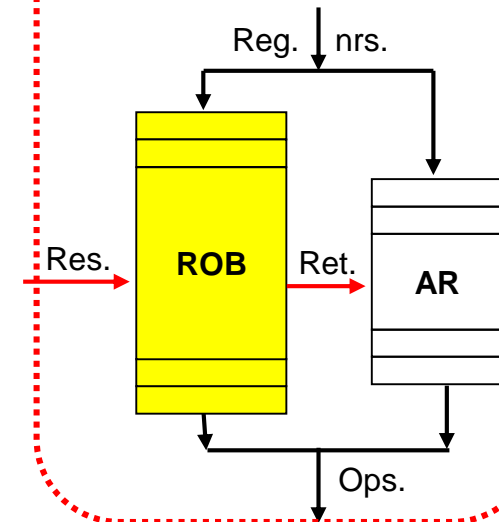
PowerPC 603 (1993)
 PowerPC 604 (1995)
 PowerPC 620 (1996)
 POWER3 (1998)
 PA 8000 (1996)
 PA 8200 (1997)
 PA 8500 (1999)
 Silbermont (2013)
 Airmont (2014)



UltraSPARC III (1999)
 K7 (FX) (1999)
 K8 (FX) (2003)



POWER1 (1990)
 POWER2 (1993)
 R10000 (1996)
 Alpha 21264 (1998)
 Pentium 4 (FP) (2000)
 K7 (FP) (1999)
 K8 (FP) (2003)
 Bulldozer (2011)
 Bobcat (2011)
 Sandy Bridge (2011) etc.
 Goldmont (2016)
 Zen (2017)



K5 (1995)
 K6 (1997)
 Pentium Pro (1995)
 Pentium II (1997)
 Pentium III (1999)
 Pentium 4 (FX) (2000)
 Pentium M (2003)
 Core 2 (2006)
 Haswell (2008)

4.2.3 New microarchitectures of the cores (5)

Using merged architectural and rename registers (aka physical registers) for renaming rather than the ROB -2

With 256-bit operands of the AVX extension ROB-based renaming became less efficient due to many reasons, like wider data-paths required, the need for forwarding result from the ROB to the architectural register file that also calls for wide data paths and for sometimes necessary multiple operand copies.

To deal with these issues the microarchitecture was changed to implement register renaming by using merged (and split) architectural and rename register files, as indicated in the next slide.

4.2.3 New microarchitectures of the cores (6)

Register renaming by means of a merged architectural and rename register file

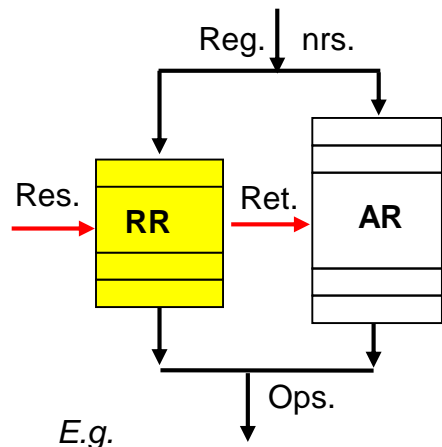
Types of rename buffers

Rename reg. file

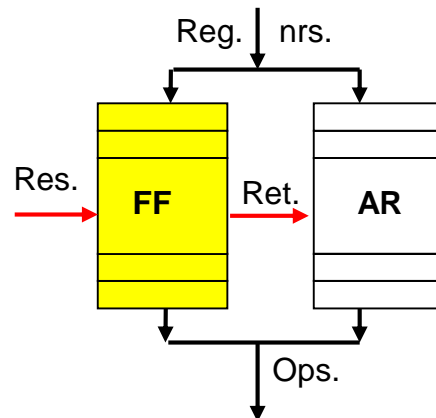
Future file

Merged arch. and rename register file

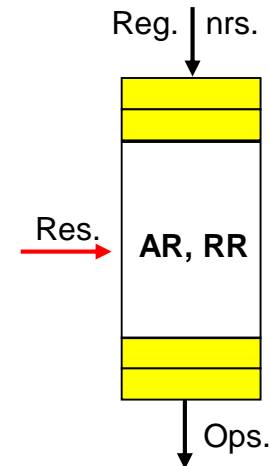
Holding renamed values in the ROB



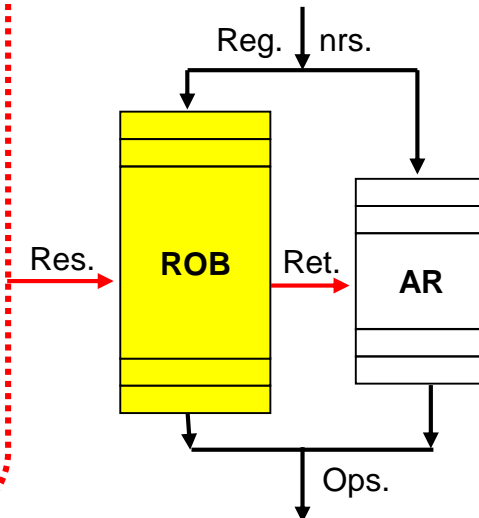
PowerPC 603 (1993)
 PowerPC 604 (1995)
 PowerPC 620 (1996)
 POWER3 (1998)
 PA 8000 (1996)
 PA 8200 (1997)
 PA 8500 (1999)
 Silbermont (2013)
 Airmont (2014)



UltraSPARC III (1999)
 K7 (FX) (1999)
 K8 (FX) (2003)



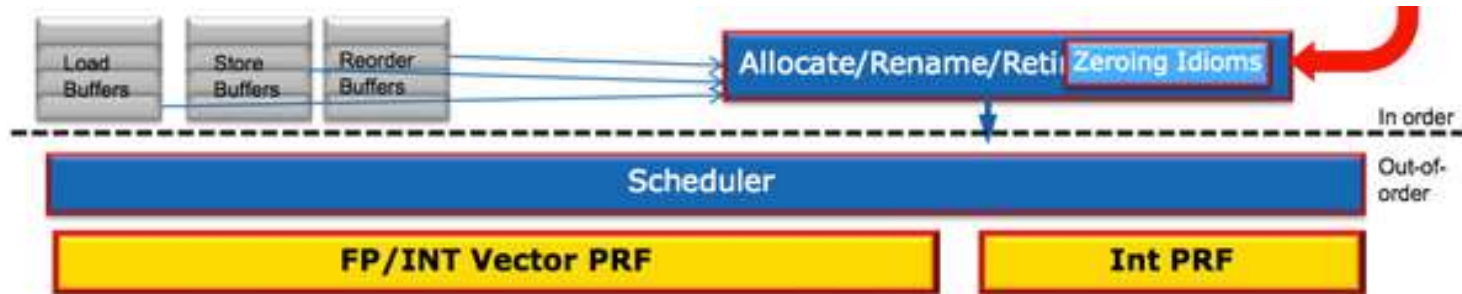
POWER1 (1990)
 POWER2 (1993)
 R10000 (1996)
 Alpha 21264 (1998)
 Pentium 4 (FP) (2000)
 K7 (FP) (1999)
 K8 (FP) (2003)
 Bulldozer (2011)
 Bobcat (2011)
Sandy Bridge (2011) etc.
 Goldmont (2016)
 Zen (2017)



K5 (1995)
 K6 (1997)
 Pentium Pro (1995)
 Pentium II (1997)
 Pentium III (1999)
 Pentium 4 (FX) (2000)
 Pentium M (2003)
 Core 2 (2006)
 Haswell (2008)

4.2.3 New microarchitectures of the cores (7)

Benefits of using merged architectural and rename register files, termed Physical Register Files for renaming in the Sandy Bridge microarchitecture [298]



- Method: Physical Reg File (PRF) instead of centralized Retirement Register File
 - Single copy of every data
 - No movement after calculation
- Allow significant increase in buffer sizes
 - Dataflow window ~33% larger

PRF is a "Cool" feature better than linear performance/power

Key enabler for Intel® Advanced Vector Extensions (Intel® AVX)

	Nehalem	Sandy Bridge
Load Buffers	48	64
Store Buffers	32	36
RS - Scheduler Entries	36	54
PRF Integer	N/A	160
PRF float-point	N/A	144
ROB Entries	128	168

b) Changing the operand fetch policy from the dispatch bound to the issue bound scheme -1

Prior to the Sandy Bridge line Intel employed the [dispatch bound operand fetch policy](#) in their Core 2 family, as the next slide indicates.

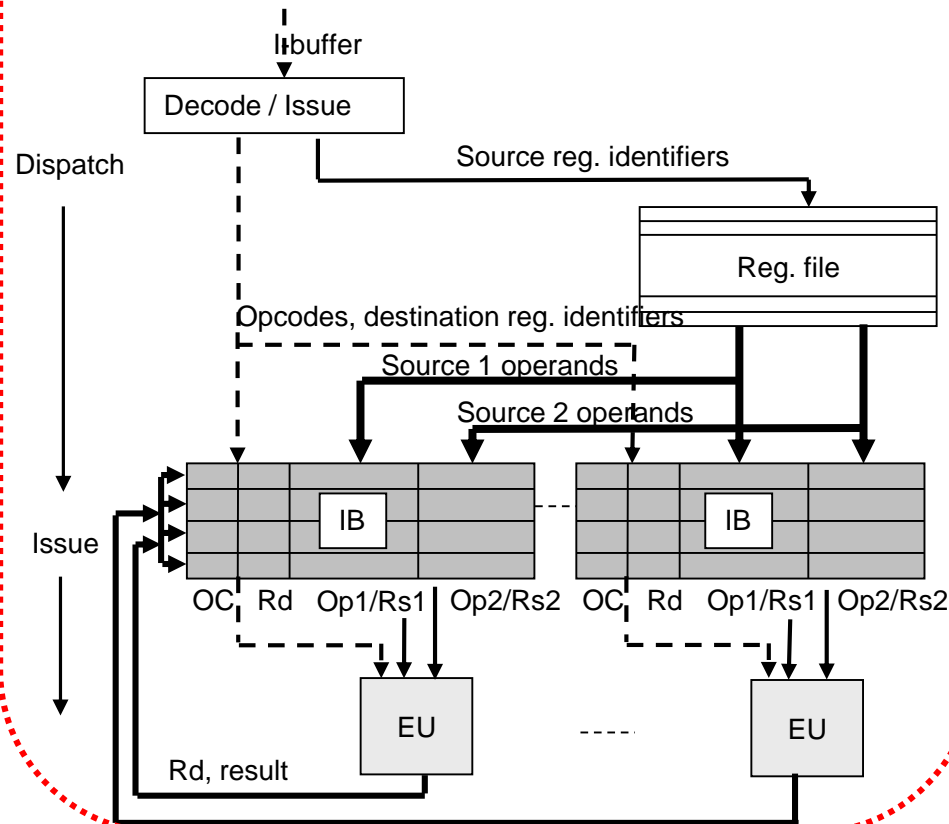
4.2.3 New microarchitectures of the cores (9)

Dispatch bound operand fetch policy

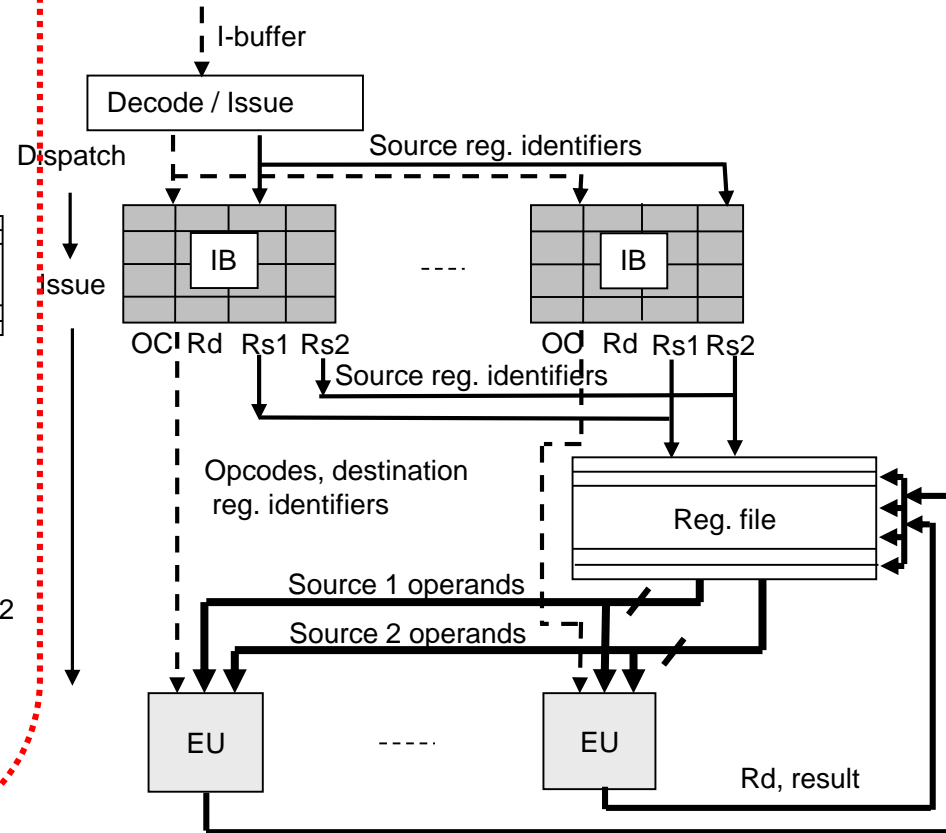
For simplicity, here we assume that no renaming is used and instruction issue is sequential (i.e. all requested operands are available)

Operand fetch policies

Dispatch bound operand fetch policy



Issue bound operand fetch policy



b) Changing the operand fetch policy from the dispatch bound to the issue bound scheme -2

Beginning with the Sandy Bridge line, however, Intel switched to issue bound operand fetch policy (see the next slide).

4.2.3 New microarchitectures of the cores (11)

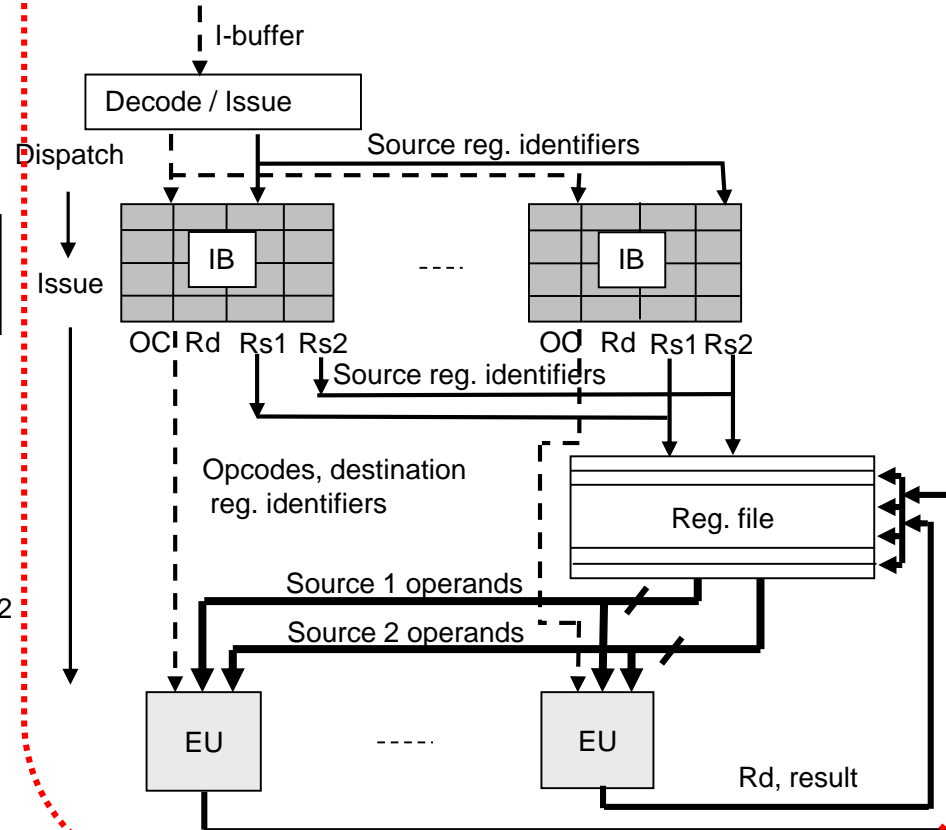
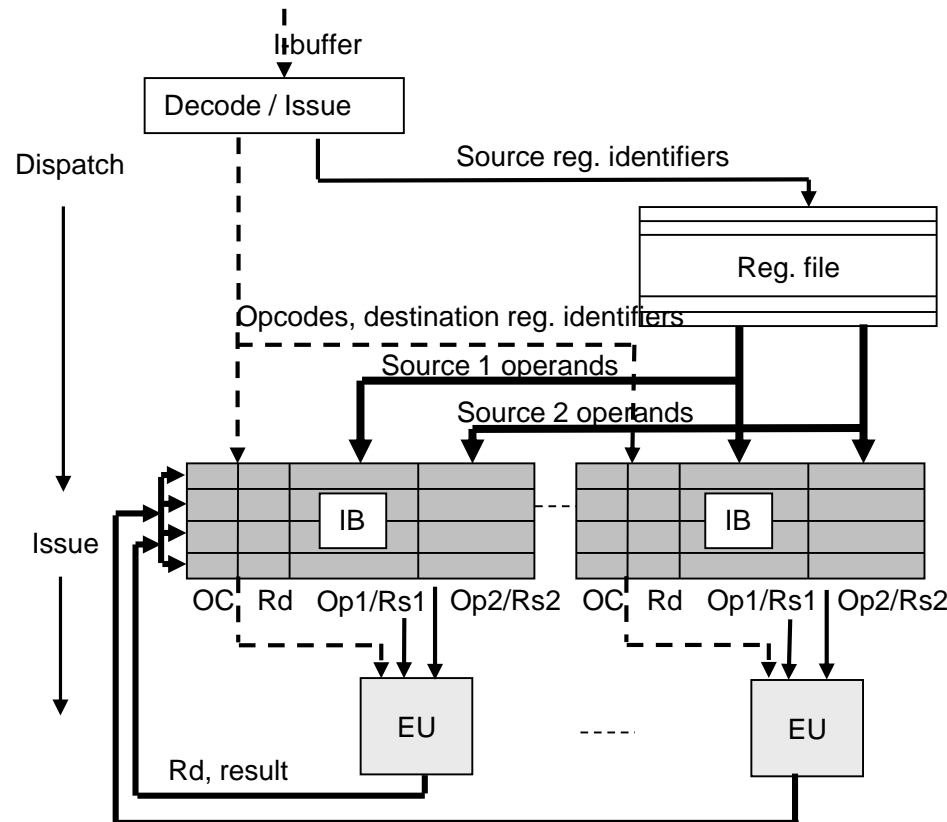
Issue bound operand fetch policy

For simplicity, here we assume that no renaming is used and instruction issue is sequential (i.e. all requested operands are available)

Operand fetch policies

Dispatch bound operand fetch policy

Issue bound operand fetch policy

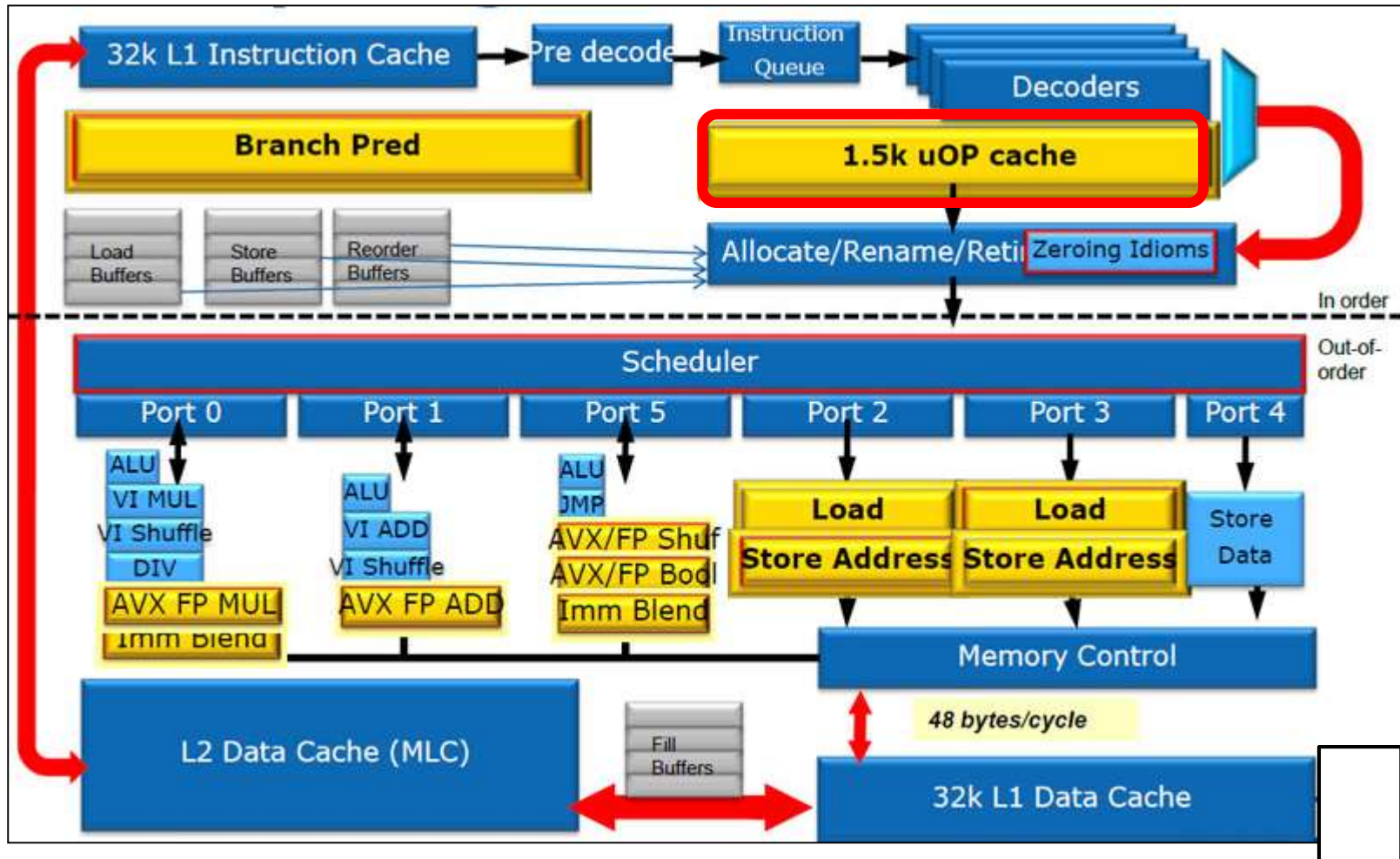


Benefits of using the issue bound operand fetch policy instead of the dispatch bound scheme

As a DSc thesis, submitted to the Hungarian Academy of Sciences in 2003 [299] points out, the most favorable datapath alternative of superscalars makes use of group reservation stations, merged architectural and rename register files and issue bound operand fetch policy (when the reservation stations hold register identifiers rather than the operands, or for yet missing operands their identifiers).

4.2.3 New microarchitectures of the cores (13)

c) Introducing a micro-op cache [213] -1



The micro-op cache [213] -2

- It can hold **1.5 K micro-operations** (micro-ops).
- Assuming an average x86 instruction length of 3.5 byte the micro-op cache is equivalent to an instruction cache of about 5.2 kB.
- The micro-op cache replaces Nehalem's **loop-buffer** that also stores micro-ops, nevertheless only up to 28 items.
- The micro-op cache **holds already decoded instructions**.
- Thus **instructions whose micro-ops are already available in the micro-op cache do not need to be fetched, predecoded, decoded and converted to micro-ops a new**.
- Here we assume that the micro-op cache has its **own branch unit** to follow instruction traces.
- According to Intel, the **hit rate** of the micro-op cache is **about 80 %**.
- This **raises performance and reduces power consumption**.

4.2.3 New microarchitectures of the cores (15)

Remark

The micro-op cache is similar to Intel's Trace Cache introduced in their Pentium 4 family in 2000.

12 K microoperations

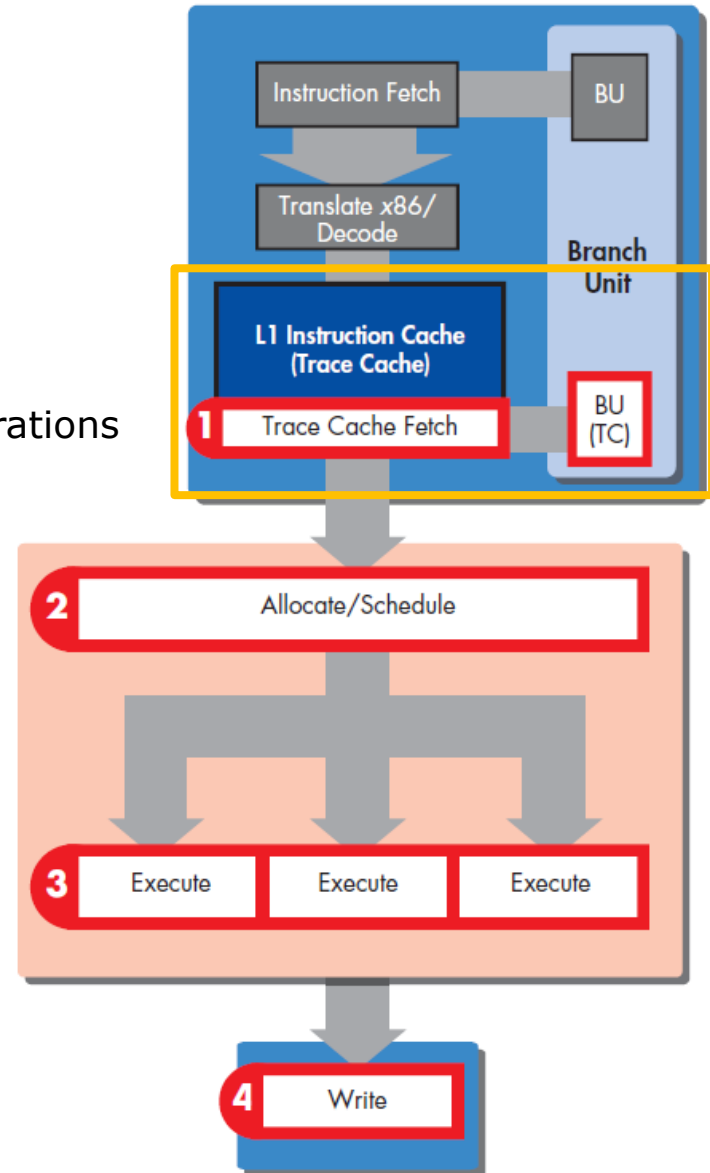


Figure: Trace Cache of the Pentium 4 [214]

4.2.3 New microarchitectures of the cores (16)

Here we do not want to go into details of the microarchitecture, but refer to two very detailed descriptions [64], [98].

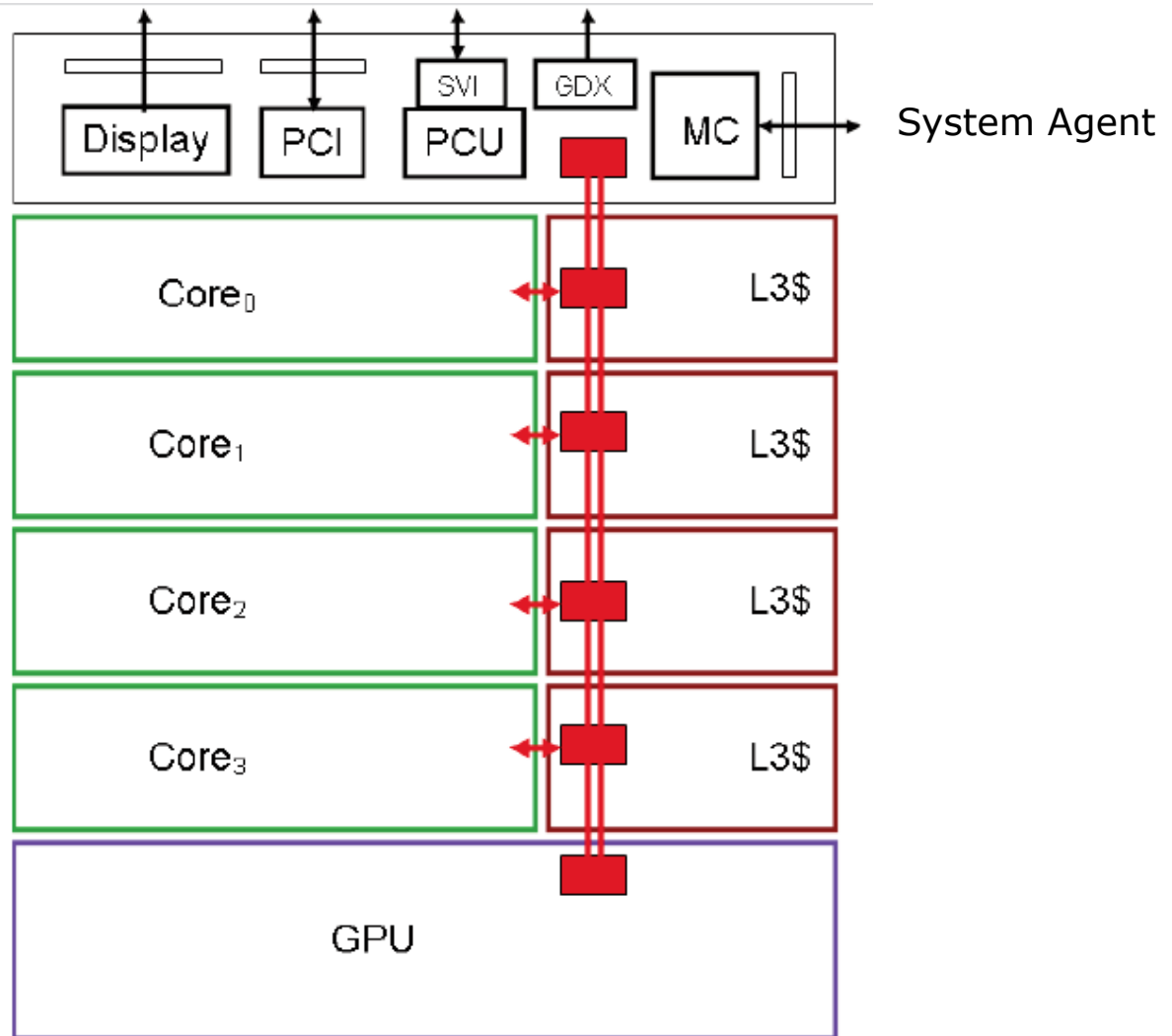
4.2.4 On-die ring interconnect bus (1)

4.2.4 On die ring interconnect bus [66]

The ring has **six bus stops** for interconnecting

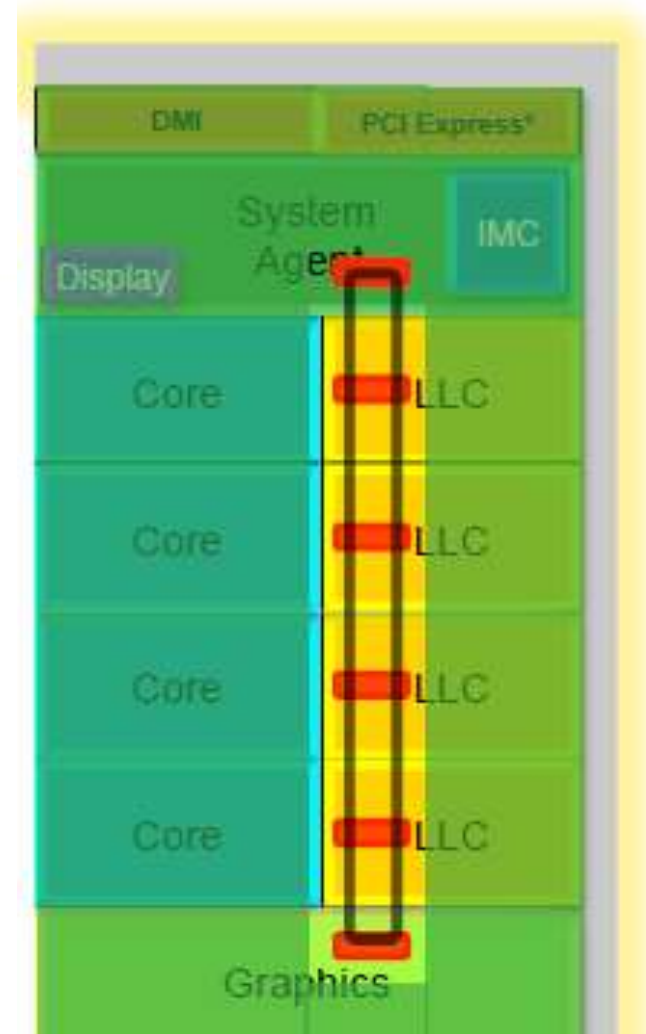
- four cores
- four L3 slices
- the GPU and
- the System Agent

The four cores and the L3 slices share the same interfaces.



Main feature of the on-die interconnect bus [64]

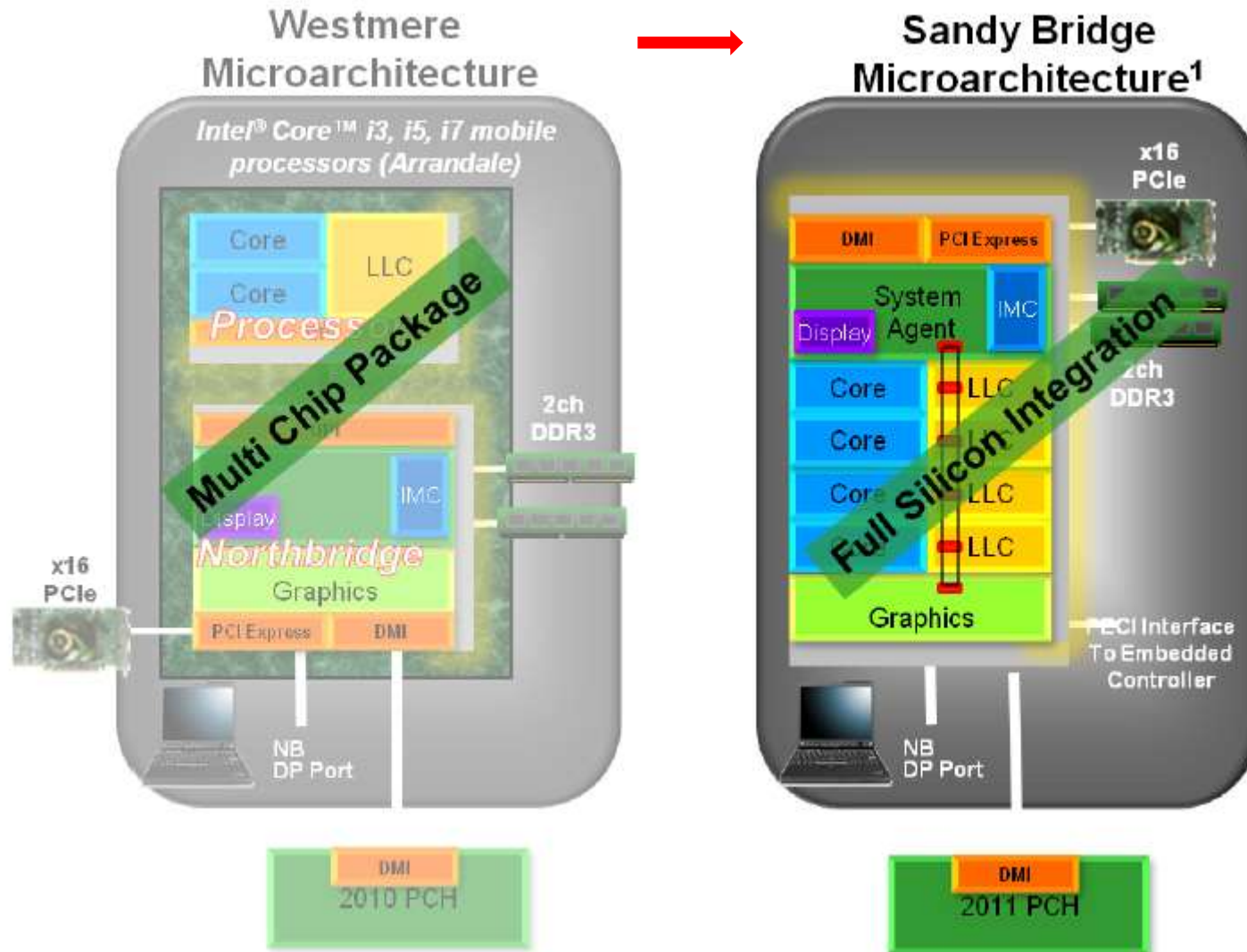
- Composed of **4 rings**
 - 32 Byte *Data* ring, *Request* ring, *Acknowledge* ring and *Snoop* ring
- It operates at core frequency (One stop/clock)
- The four rings need a considerable amount of wiring and routing.
- As the **routing runs in the upper metal layers over the LLC**, thus the rings have no real impact on the die area.



4.2.5 On die graphics unit (1)

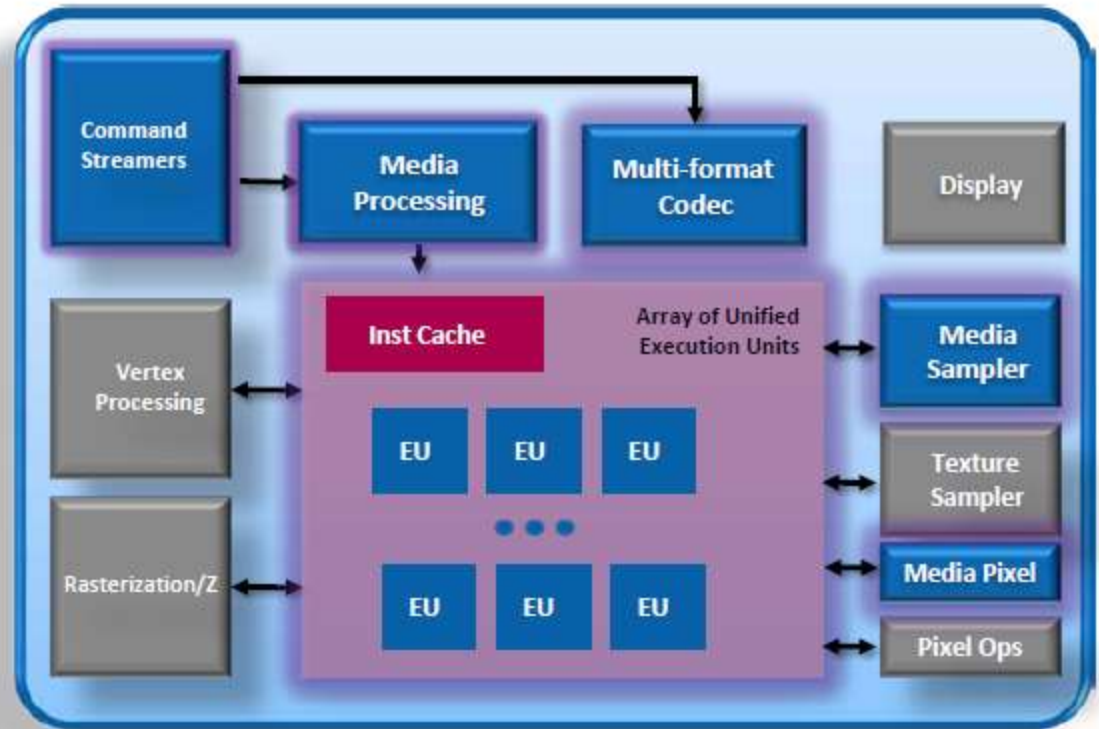
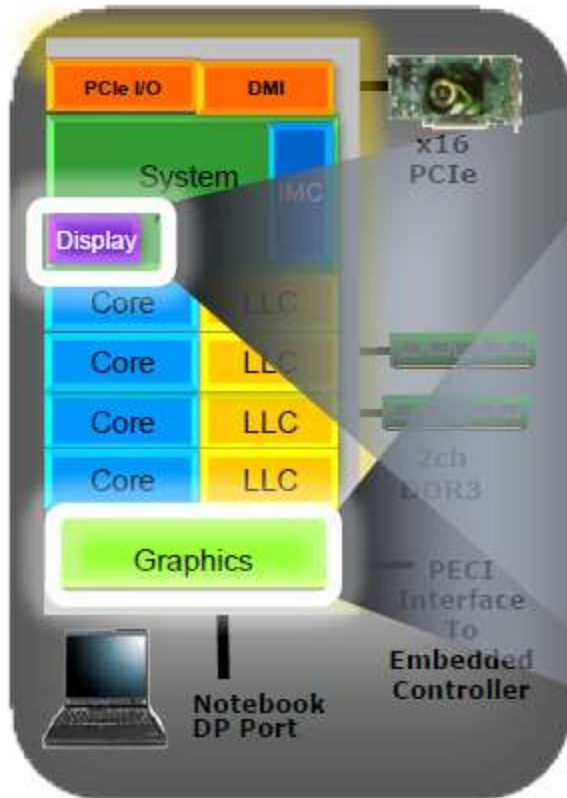
4.2.5 On die graphics unit [99]

Evolution of graphics implementation from Westmere to Sandy Bridge [99]



4.2.5 On die graphics unit (2)

Support of both media and graphics processing by the graphics unit [99]

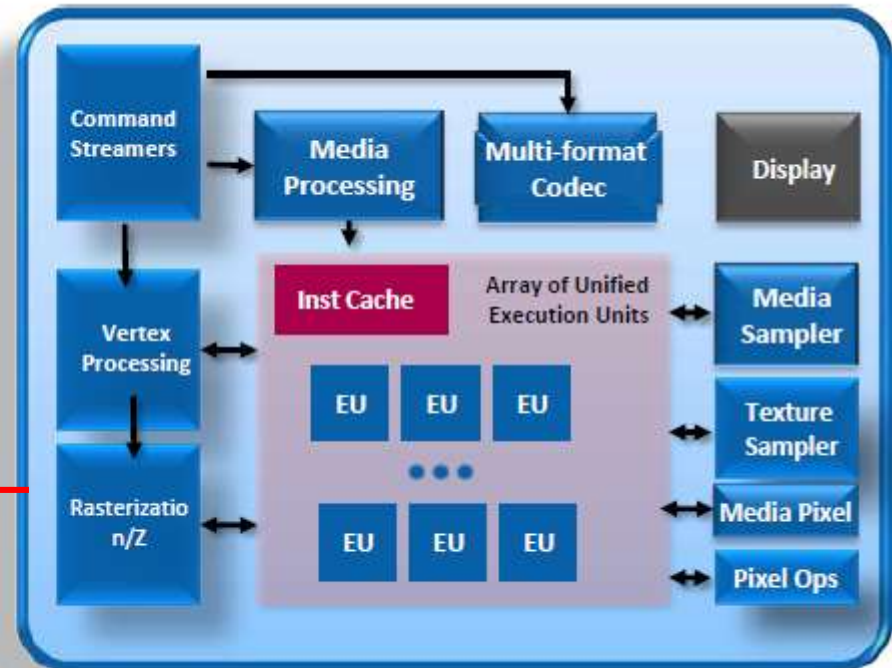


Media processing is synergistic with 3D processing

4.2.5 On die graphics unit (3)

Main features of the on die graphics unit [99]

- Now designed into the same die as CPU
 - Leading Edge 32nm Process
- Shared last level cache
 - Configurable cache partitioning
 - Higher bandwidth for Graphics
 - Lower latency
 - Reduced DRAM accesses
- Utilize CPU power management
 - Improved Graphics power efficiency
 - Best overall (CPU+Graphics) power decisions



GT1: 6 or GT2: 12 EUs

4.2.5 On die graphics unit (4)

Specification data of the HD 2000 and HD 3000 graphics [100]

Graphics	Market	Chipset/CPU	Code Name	Device ID	Core Render Clock (MHz)	Execution Units	Shader Model (Unified Shader)	API Support			Memory Bandwidth (GB/s)	DVMT (MB)	Hardware Acceleration		
								DirectX	OpenGL	OpenCL			MPEG-2	VC-1	AVC
HD Graphics 2000	Desktop	Non-K edition Core i3, Core i5, Core i7	Sandy Bridge	0102	650–1250 (Turbo)	6 (GT1)	4.1	10.1	3.0	-	21.3	1720	Full	Full	Full
				0106											
				0112											
HD Graphics 3000	Desktop	Core i5-2x00K Core i7-2x00K	Sandy Bridge	0116	650–1350 (Turbo)	12	4.1	10.1	3.0	-	21.3	1720	Full	Full	Full
				0122											
	Mobile	Core i3, Core i5, Core i7		0126 010A											
			Ivy Bridge	0080			5.0	11							

4.2.5 On die graphics unit (5)

Execution units (EU) of the graphics unit in Sandy Bridge [197]

- Each EU is basically a **4-wide SP FP SIMD** unit intended to operate on 4-component data (RGBA), capable of executing **2-operation MAD** instructions and also FX instructions.
- EUs are **5-way multithreaded** for **GT2** graphics and **4-way multithreaded** for **GT1** graphics.
- **Each thread** has a register set of **120 x 256 bit registers**.
- There is also a **fixed function Math Box** for executing **transcendental, e.g. trigonometric instructions and also FP divide**.
- The EUs do not support DP FP operations.

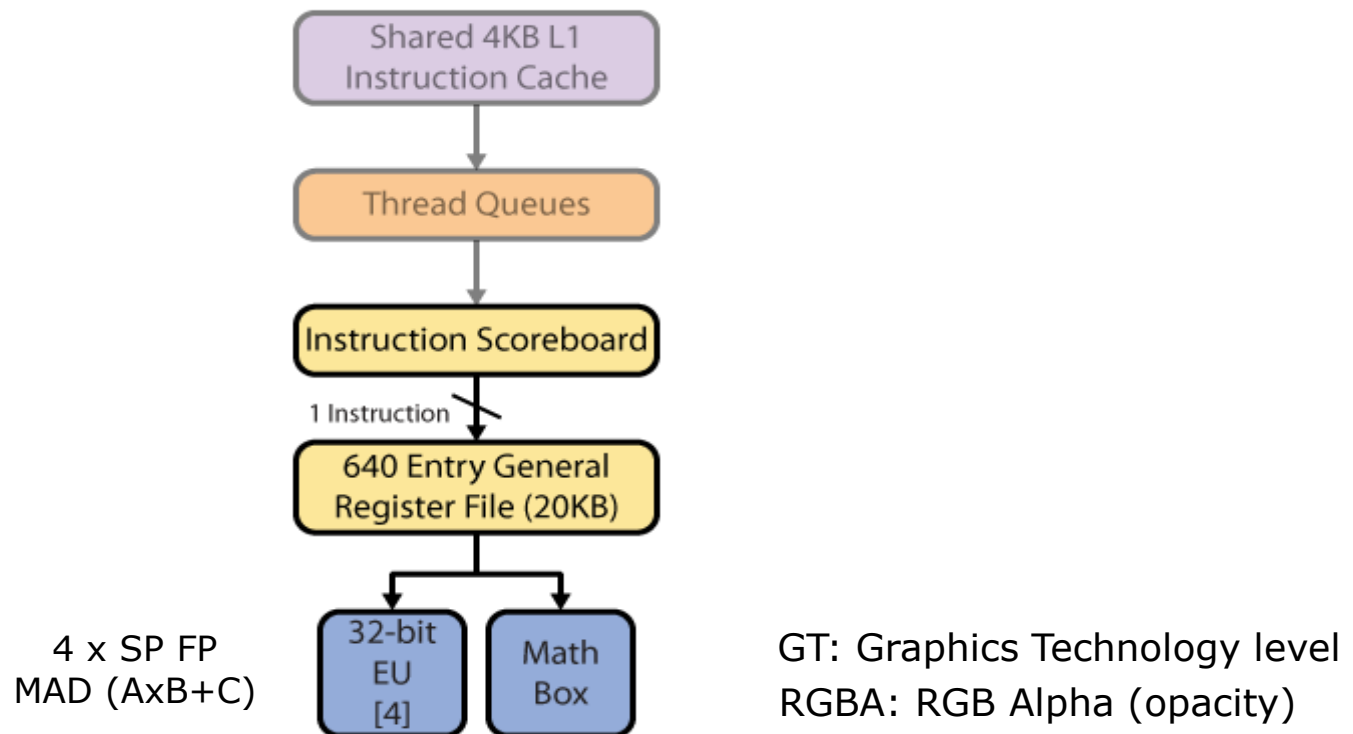
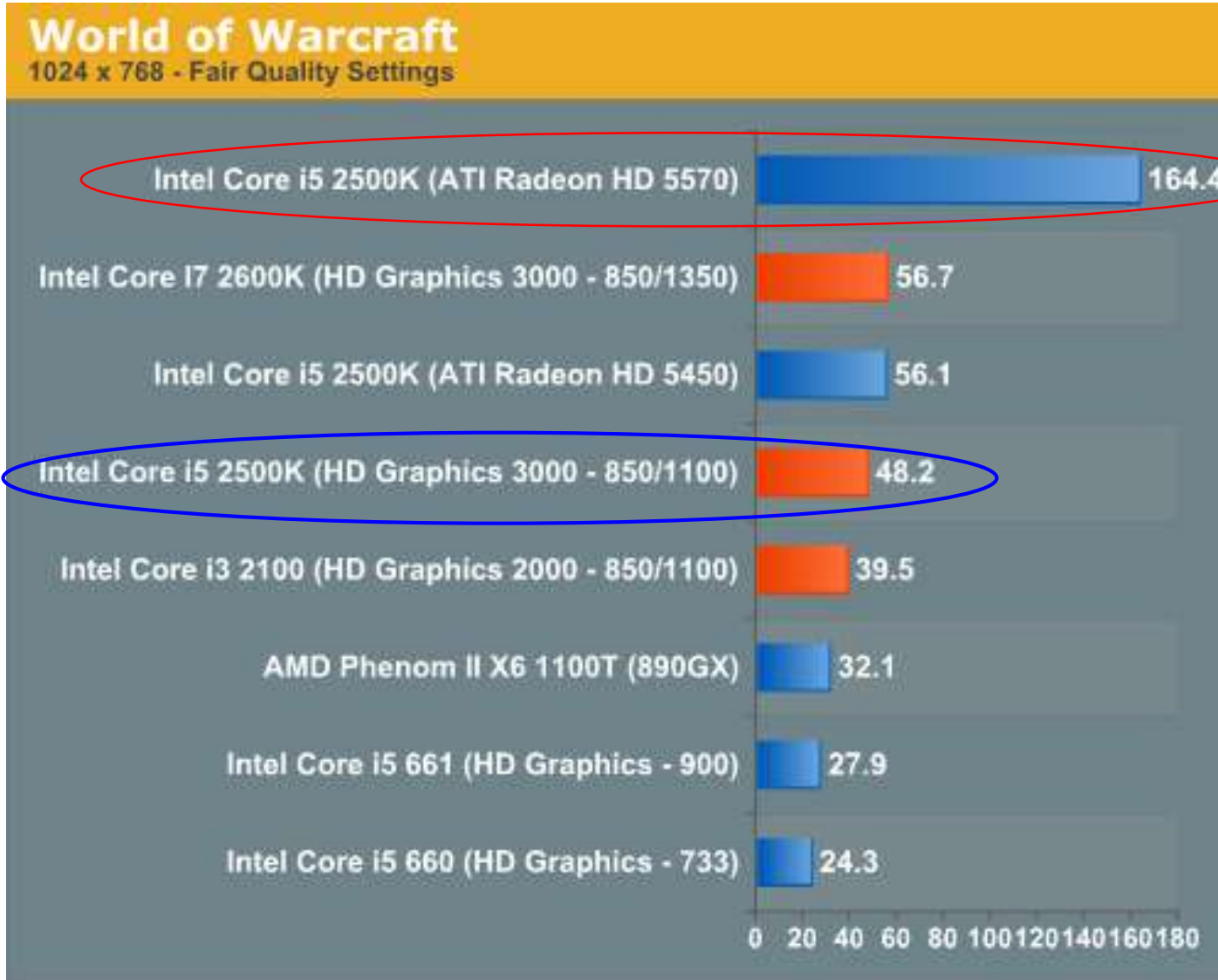


Figure: Block diagram of an EU of the graphics unit of Sandy Bridge

4.2.5 On die graphics unit (6)

Performance comparison of the Sandy Bridge's graphics: gaming [101]



HD 5570
400 ALUs

i5/i7 2xxx/3xxx:
Sandy Bridge

i5 6xx
Arrandale

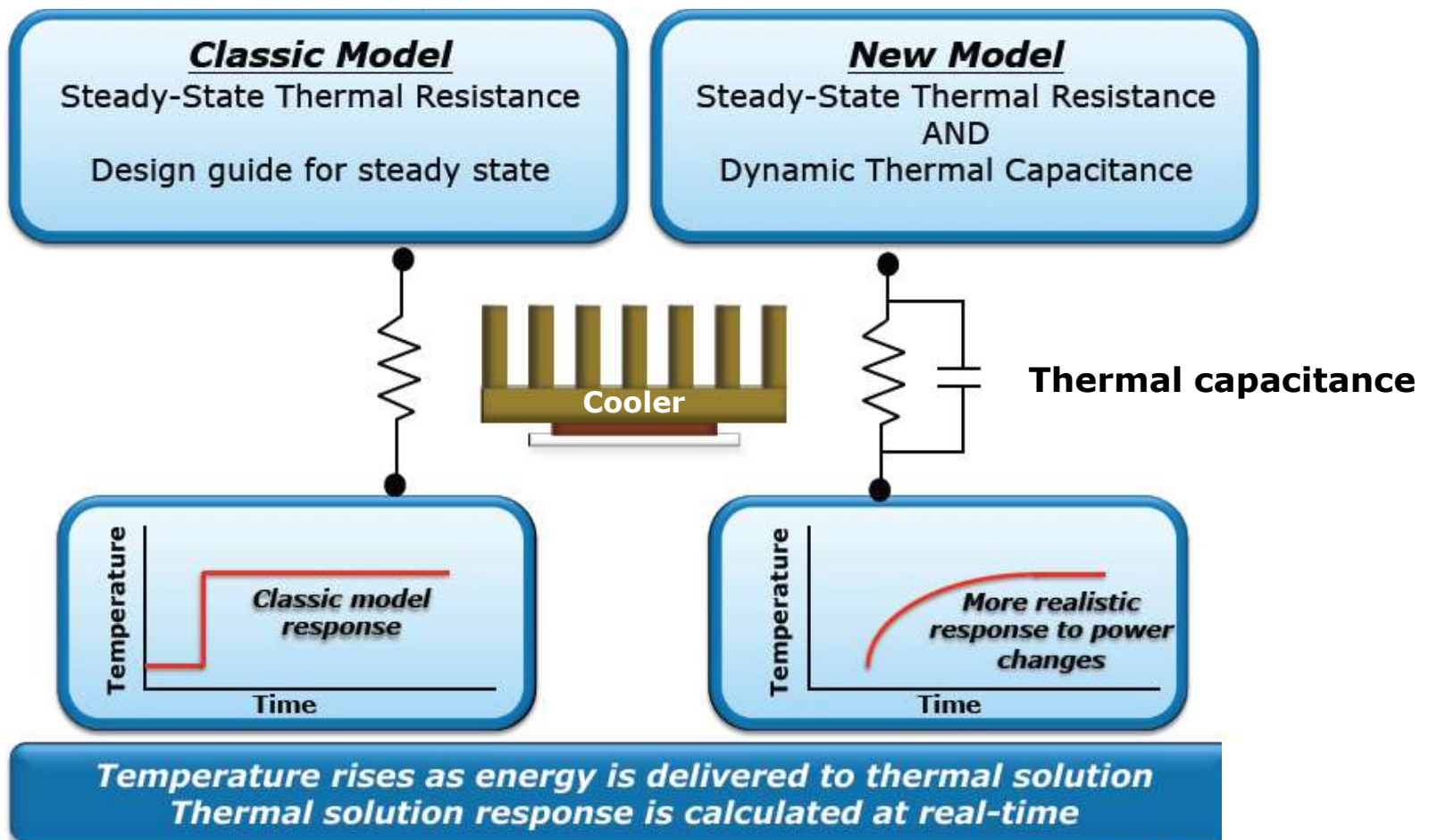
frames per sec

4.2.6 Turbo Boost technology 2.0 (1)

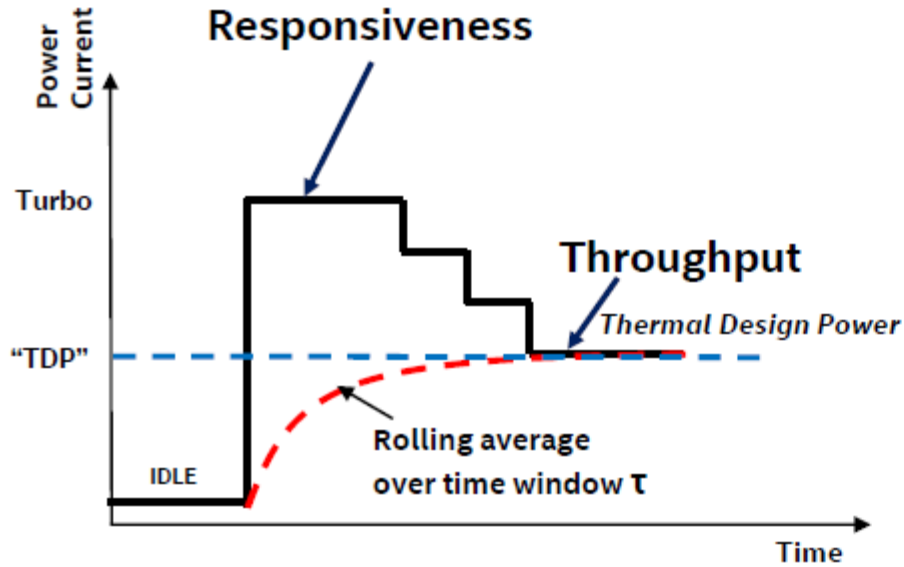
4.2.6 Turbo Boost 2.0 technology [64]

Designated also as the **2.0 generation Turbo Boost technology**.

The concept utilizes the **real temperature response** of processors to power changes in order to increase the extent of overclocking [64].



Aim of Intel's Turbo Boost Technology 2.0 [198]



- Maximize user experience within system constraints¹
- User experience:
 - Throughput performance
 - Responsiveness

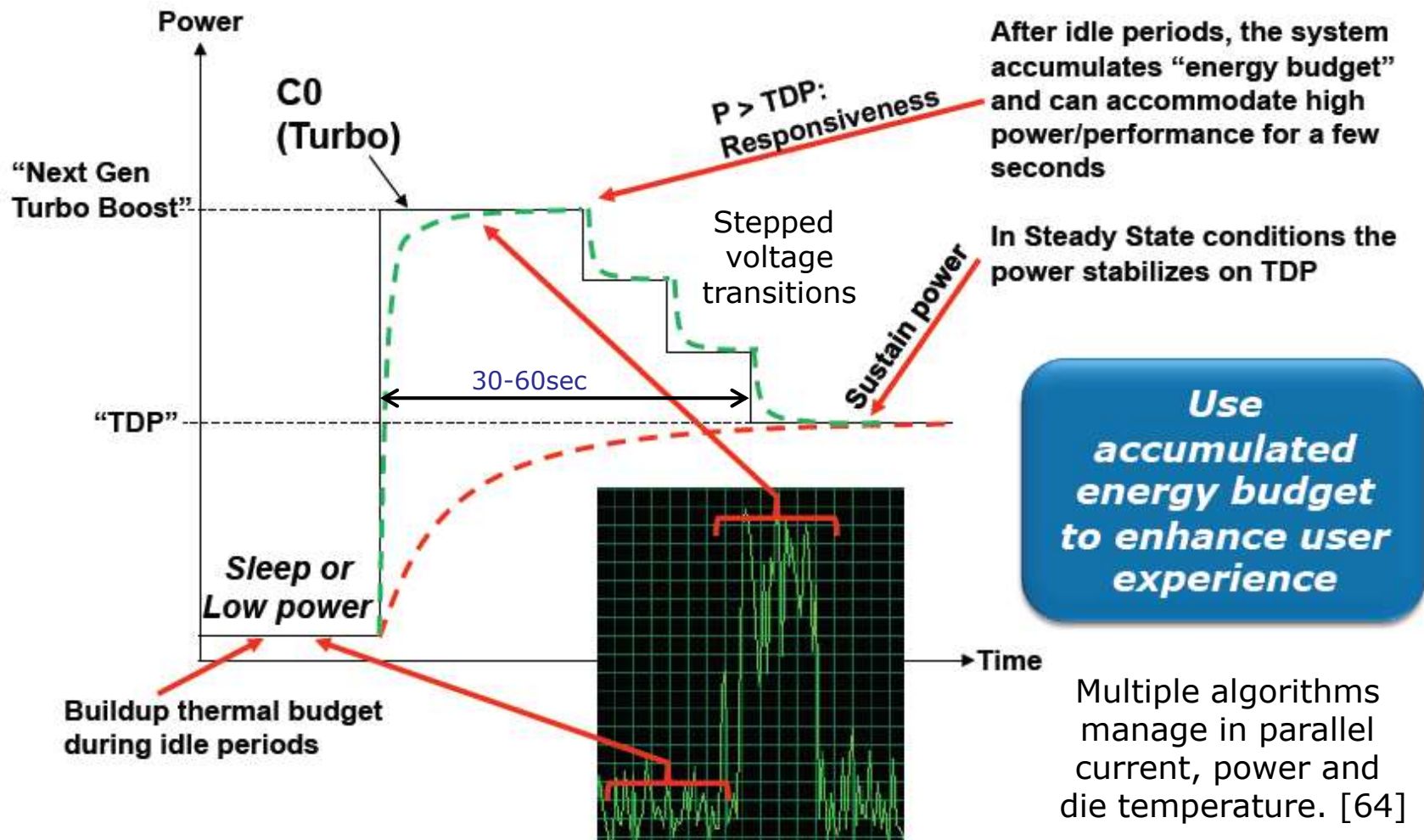
User experience: Felhasználói élmény

Responsiveness: Reakcióképesség

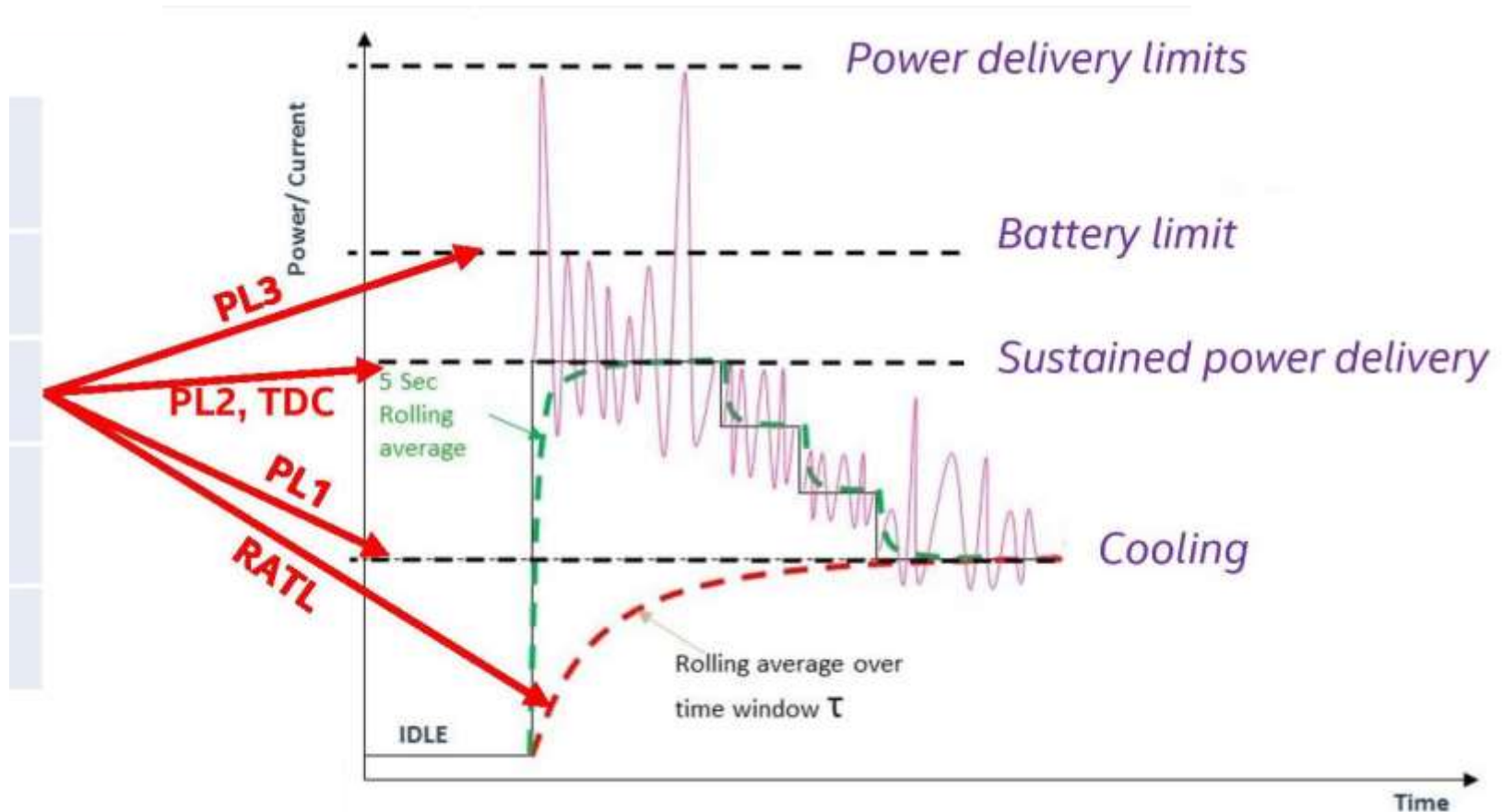
Throughput: Áteresztőképesség/Teljesítmény

Principle of the implementation of Turbo Boost 2.0 [64]

- Based on the real temperature response
the thermal energy budget accumulated during idle periods
can be utilized to push the core beyond the TDP for short periods of time (e.g. for 20 sec).



Designation of different power values related to the Turbo Boost technology [295]



PL1: The **cooling limit**, it is effectively the **TDP** value.

Here the power (and frequency) is limited by the cooling available.

PL2: The **maximum sustainable power** that the processor can take until hitting thermal limits.

This is essentially the power required **to hit the peak turbo on all cores** (E.g. 210 W for running all 8 cores of the Core i9-9900X at 4.7 GHz vs. the 95 W TDP).

Introduction of the Turbo Boost 3.0 (aka Turbo Boost Max 3.0) technology in the Broadwell-E line (2016) [248]

- Turbo Boost Max 3.0 offers an additional 100 to 200 MHz clock boost for single threaded applications.
- Max. core speeds are measured during testing the chip.
- The core with the highest possible clock speed is called the "favored core".
- It will be activated for Turbo Boosting in case when only a single core is needed.
- It needs BIOS and OS support.

Remarks to the Turbo Boost Max 3.0 technology

- In practice, motherboard manufacturers often didn't support it or they do disable it in the BIOS by default.
- If users intend to make use of it they have to install the drivers and the BIOS as well.

4.2.6 Enhanced Turbo Boost technology (4)

Intelligent power sharing between the cores and the processor graphics (PG) [64]

□ Power specification is defined for the entire package

- Monolithic die – power budget shared by CPU and PG
- Sum of component power at or below specifications

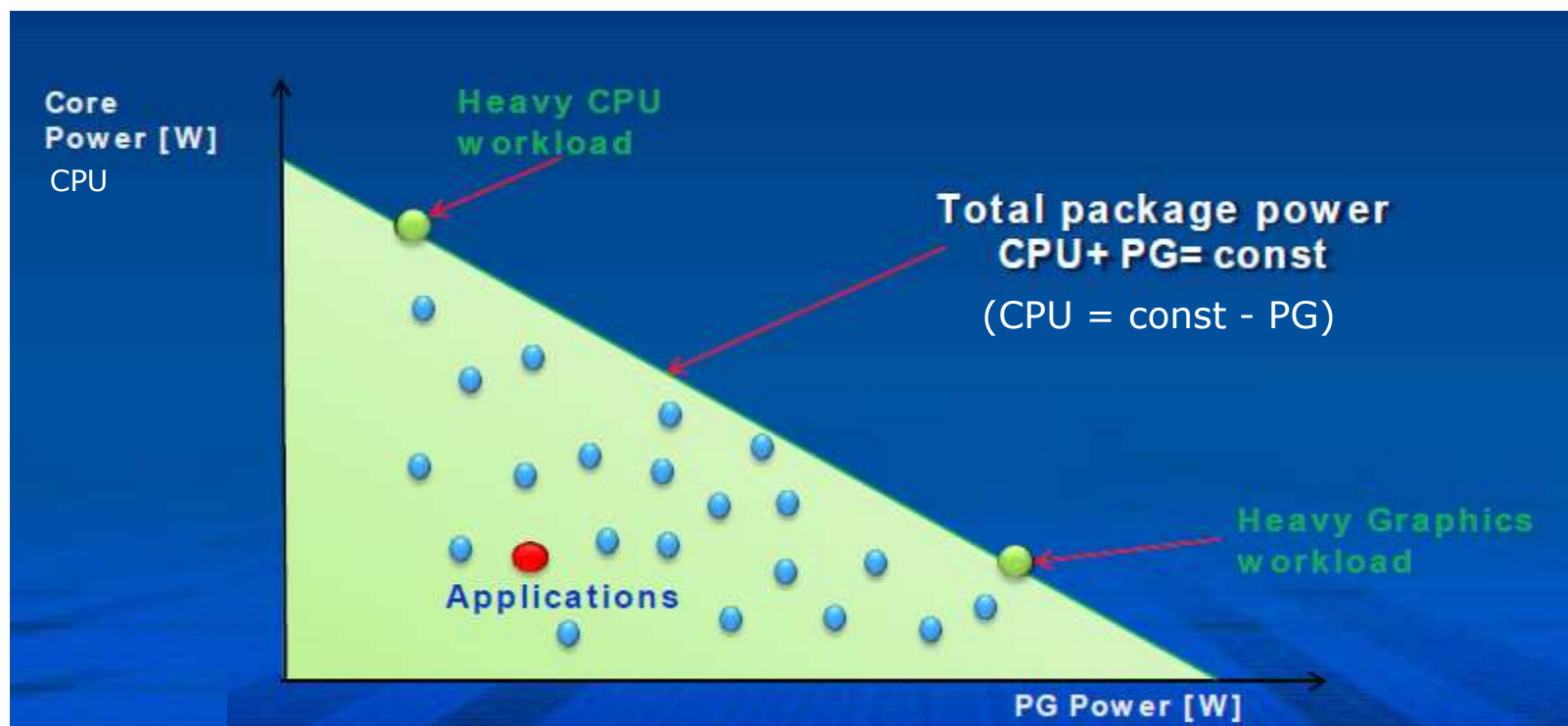

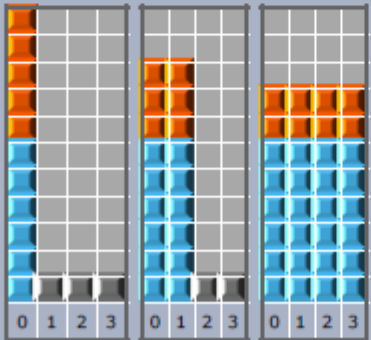
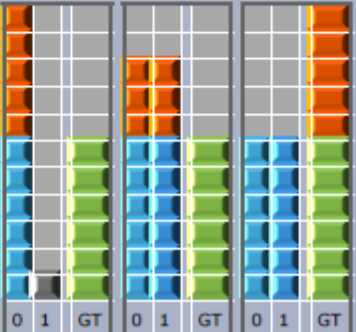



Figure: Power sharing between CPU and GPU [207]

Next Generation Intel® Turbo Boost Technology [61]

Client	Penryn (EDAT) Penryn/M	Nehalem Nehalem/M (Clarksfield) Nehalem/D (Lynnfield)	Westmere Westmere/M (Arrandale) Westmere/D (Clarkdale)	Sandy Bridge
Key New Capabilities	<ul style="list-style-type: none"> • 1 turbo bin when other core is asleep 	<ul style="list-style-type: none"> • Turbo controlled within power limit • Multi-core turbo • More turbo if cores are asleep 	<ul style="list-style-type: none"> • Graphics Dynamic Frequency • Driver controlled power sharing between IA and Graphics (Mobile) 	<ul style="list-style-type: none"> • HW controlled power sharing between IA cores and Graphics • Dynamic Turbo provides high <u>responsiveness</u> • More Turbo headroom from Improved power monitoring and control
Turbo Behavior	<p>Illustrative only. Does not represent actual number of turbo bins.</p> 	<p><u>Quad Core Die</u></p> <p>Single Core Turbo Dual Core Turbo Quad Core Turbo</p> 	<p><u>Dual Core Die</u></p> <p>Single Core Turbo Dual Core Turbo Graphics Turbo</p> 	<p>Dual Core Die Quad Core Die</p> 

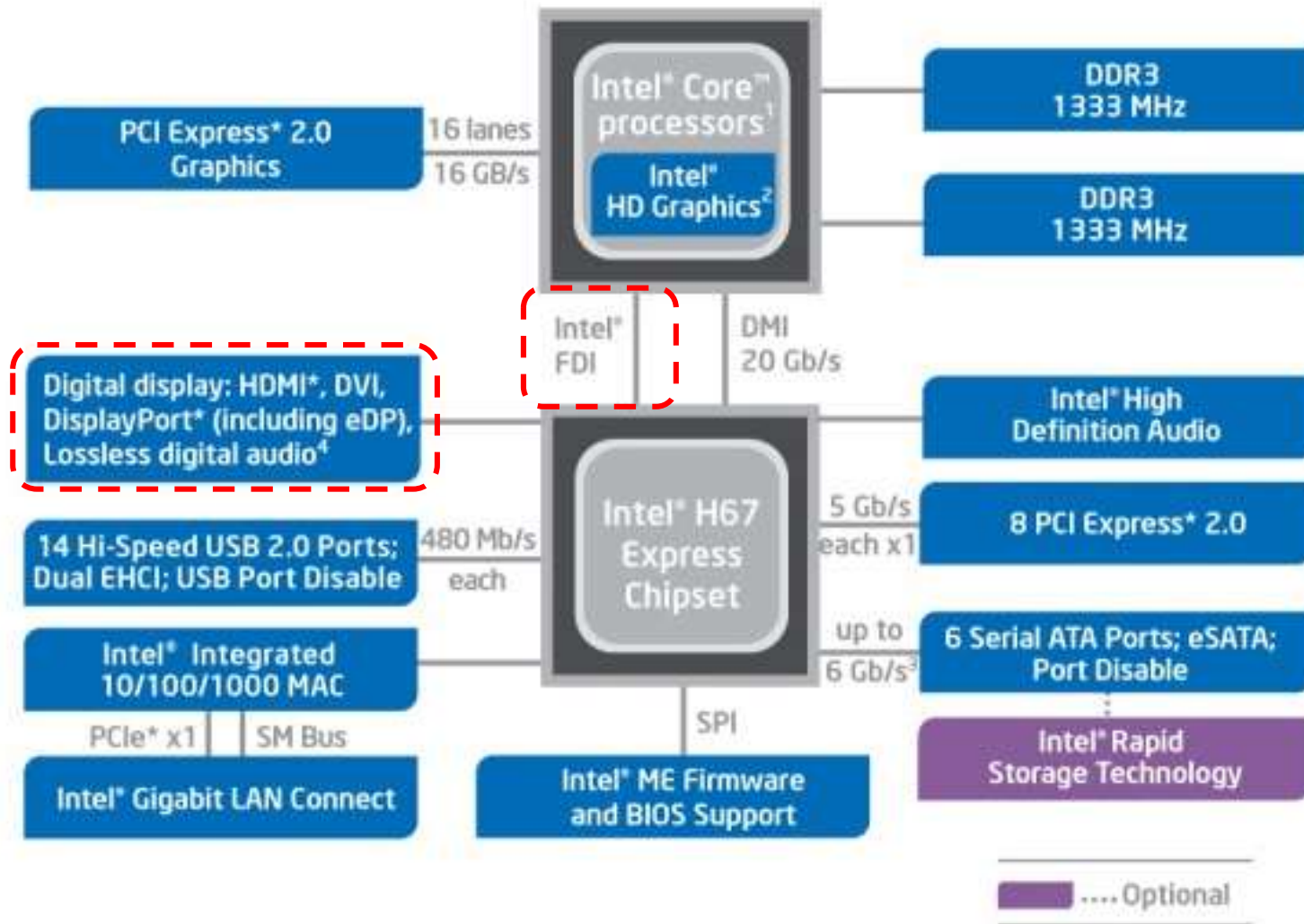
Remark

- **Active cores** run at the same clock frequency and share the same power plane.
- **Idle cores** may be shut down by power gates.

4.3 Example for a Sandy Bridge based desktop platform with the H67 chipset

4.3 Example for a Sandy Bridge based desktop platform with the H67 chipset (1)

4.3 Example for a Sandy Bridge based desktop platform with the H67 chipset [102]



FDI: Flexible Display Interface

5. The Haswell line

- 5.1 Introduction
- 5.2 Major enhancements of the Haswell line vs. the Sandy Bridge line
- 5.3 Major innovations of the Haswell line
- 5.4 Haswell based mobile and desktop processors
- 5.5 Haswell based server processors

Only Section 5.1 is discussed!

5.1 Introduction to the Haswell line

5.1 Introduction to the Haswell line (1)

5.1 Introduction to the Haswell line

Haswell processors are termed also as the **4. gen. Intel Core processors**, as indicated below.

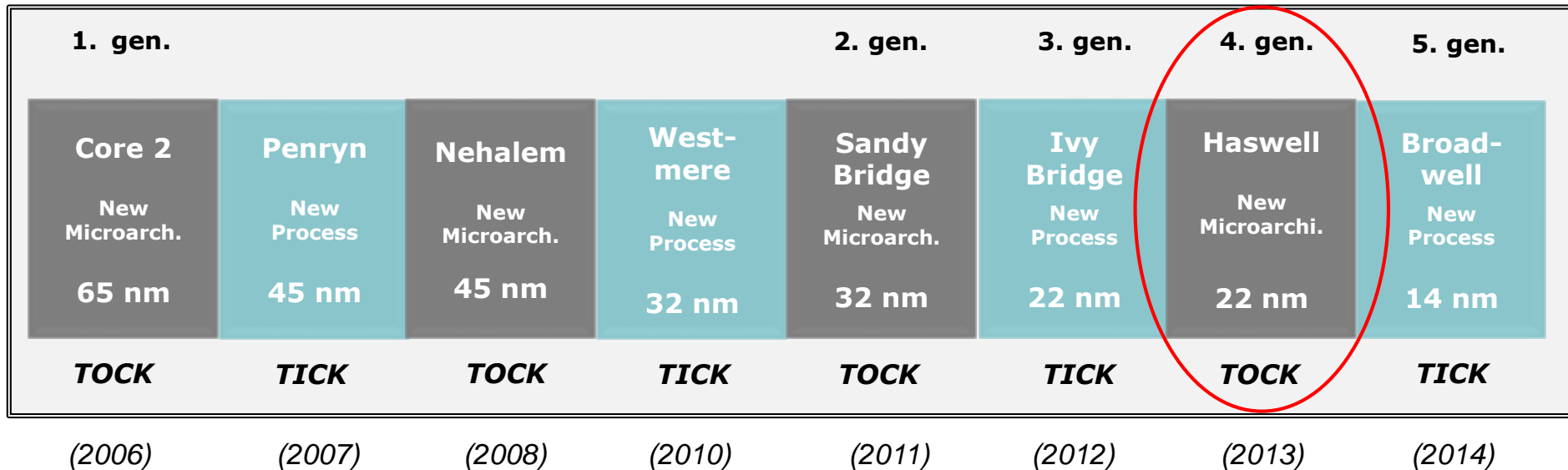


Figure : Intel's Tick-Tock development model (Based on [1])

Launched: [6/2013](#) at Computex.

The **Haswell line**

- Launched in [06/2013](#)
- [22 nm](#) IC technology

The **Haswell refresh processors** [176]

- A [second wave](#) of Haswell processors, called the **Haswell refresh** processors launched in [5/2014](#).

They do not provide any significant changes vs. the first released processors.

Actually, the manufacturing process could be made more efficient and this resulted in slight improvements in clock speeds.

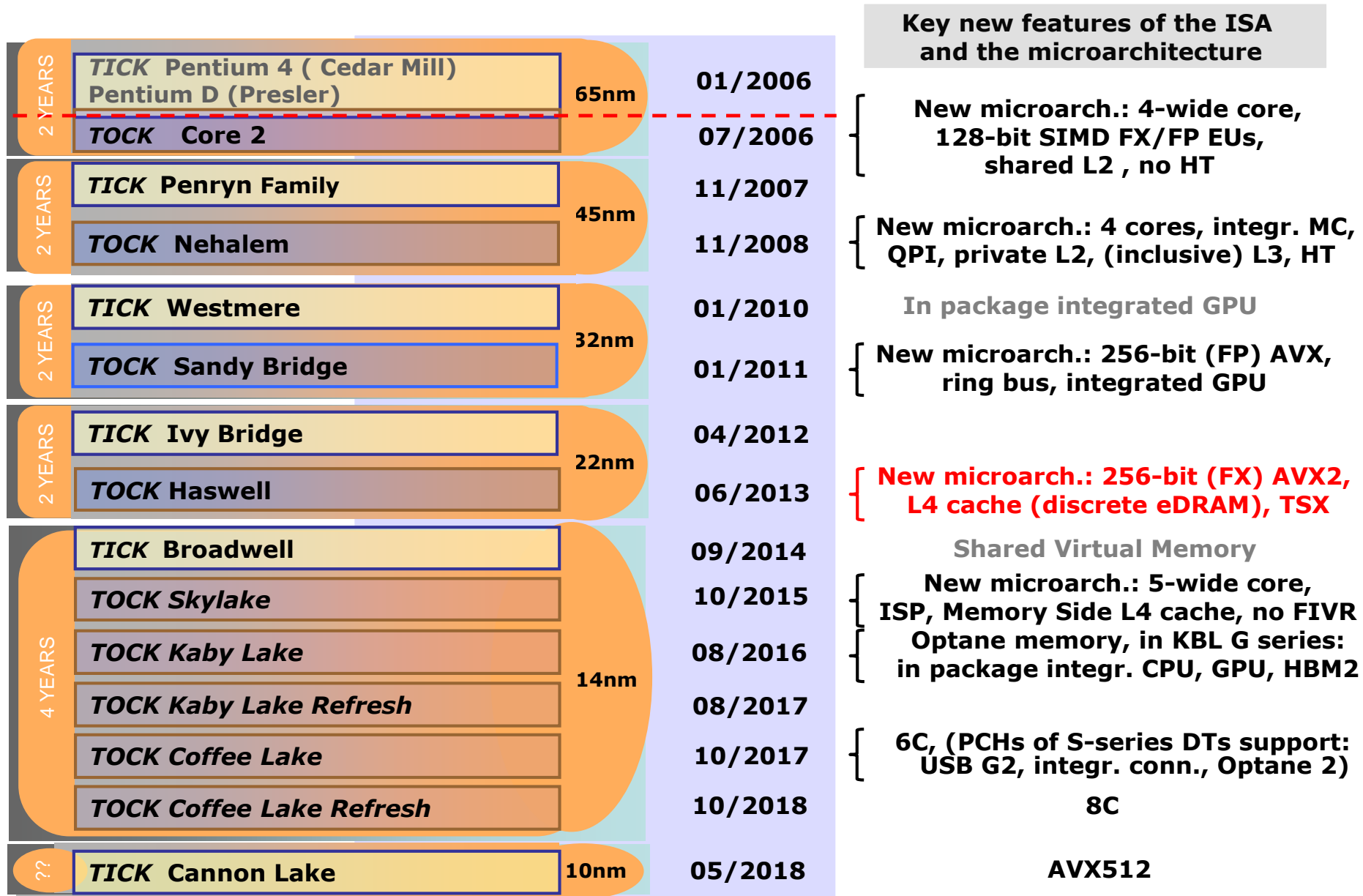
- A third [wave](#) of Haswell processors, called the **Devil's Canon line** launched in [6/2014](#). They provide higher clock speeds vs. the previous processors.

DP/MP servers (Haswell-EP, Haswell-EX)

They were launched later, in [09/2014](#) and [05/2015](#), as indicated subsequently.

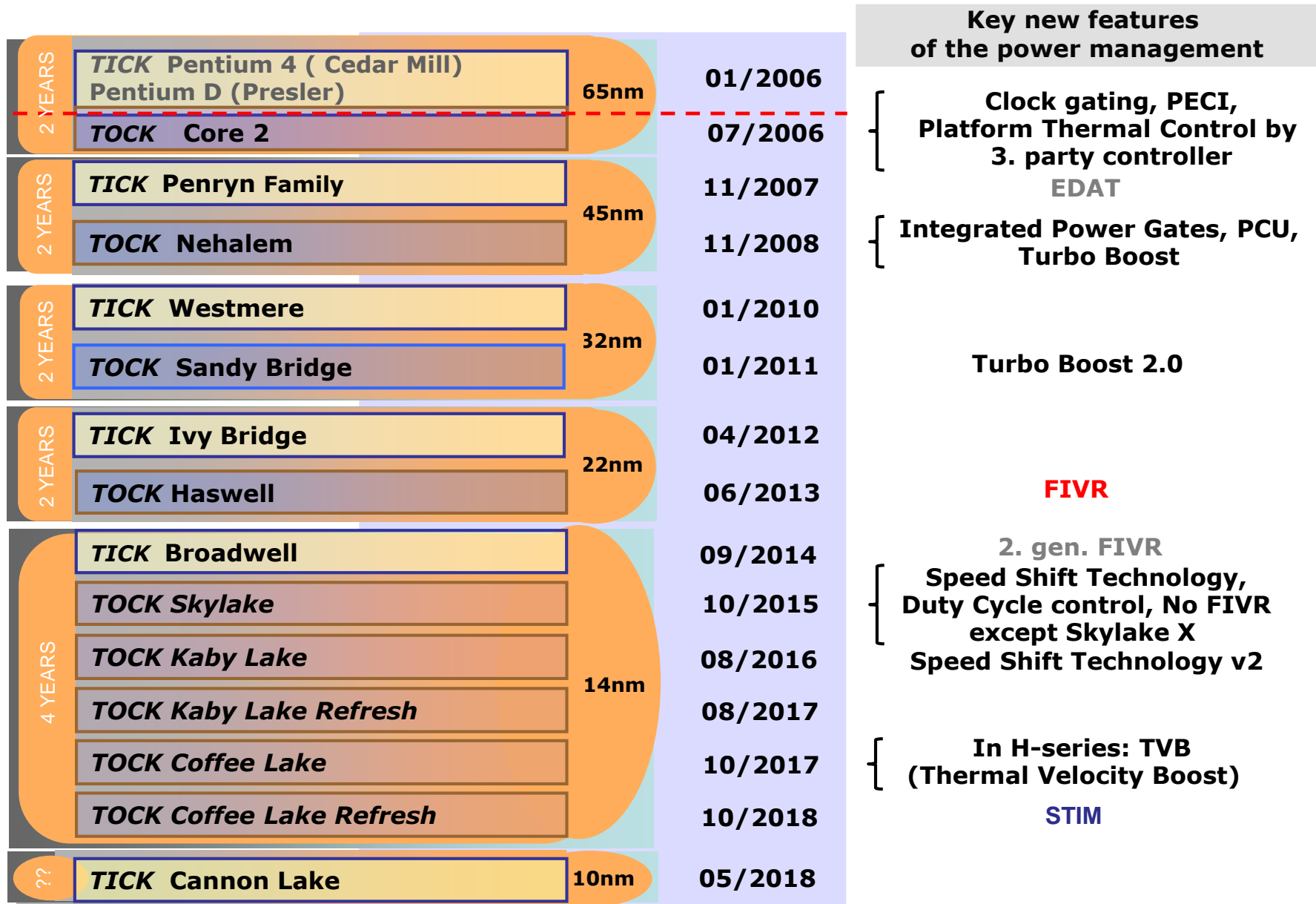
5.1 Introduction to the Haswell line (3)

The Haswell line -1 (based on [3])



5.1 Introduction to the Haswell line (4)

The Haswell line -2 (based on [3])

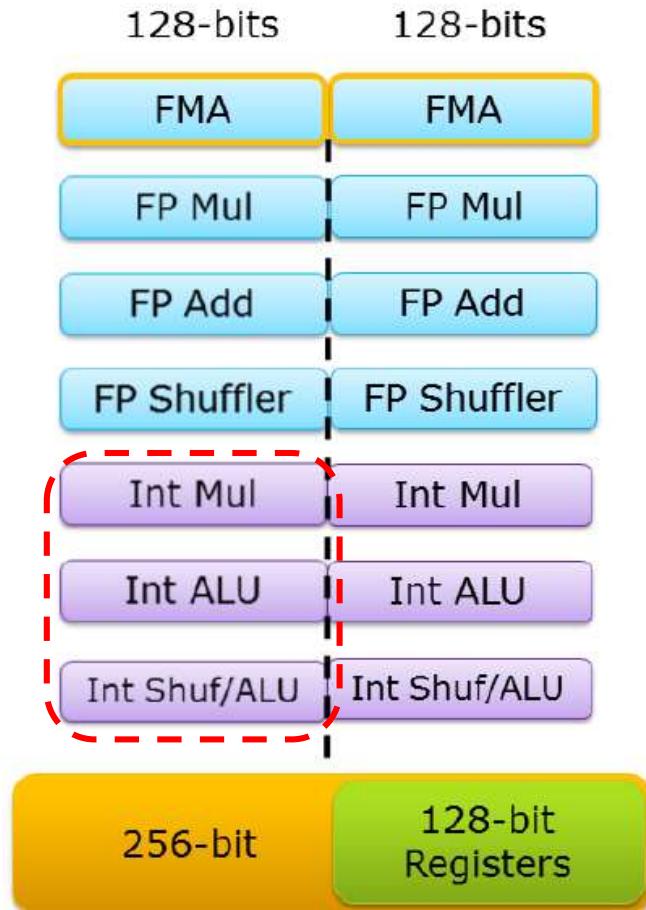


Key new features of the ISP and the microarchitecture

- a) 256-bit (FX) AVX2 ISA extension
- b) On-package eDRAM L4 cache
- c) FIVR (Fully Integrated Voltage Regulator)
- d) TSX (Transactional Synchronization Extensions)

5.1 Introduction to the Haswell line (6)

a) 256-bit (FX) AVX2 ISA extension [97]



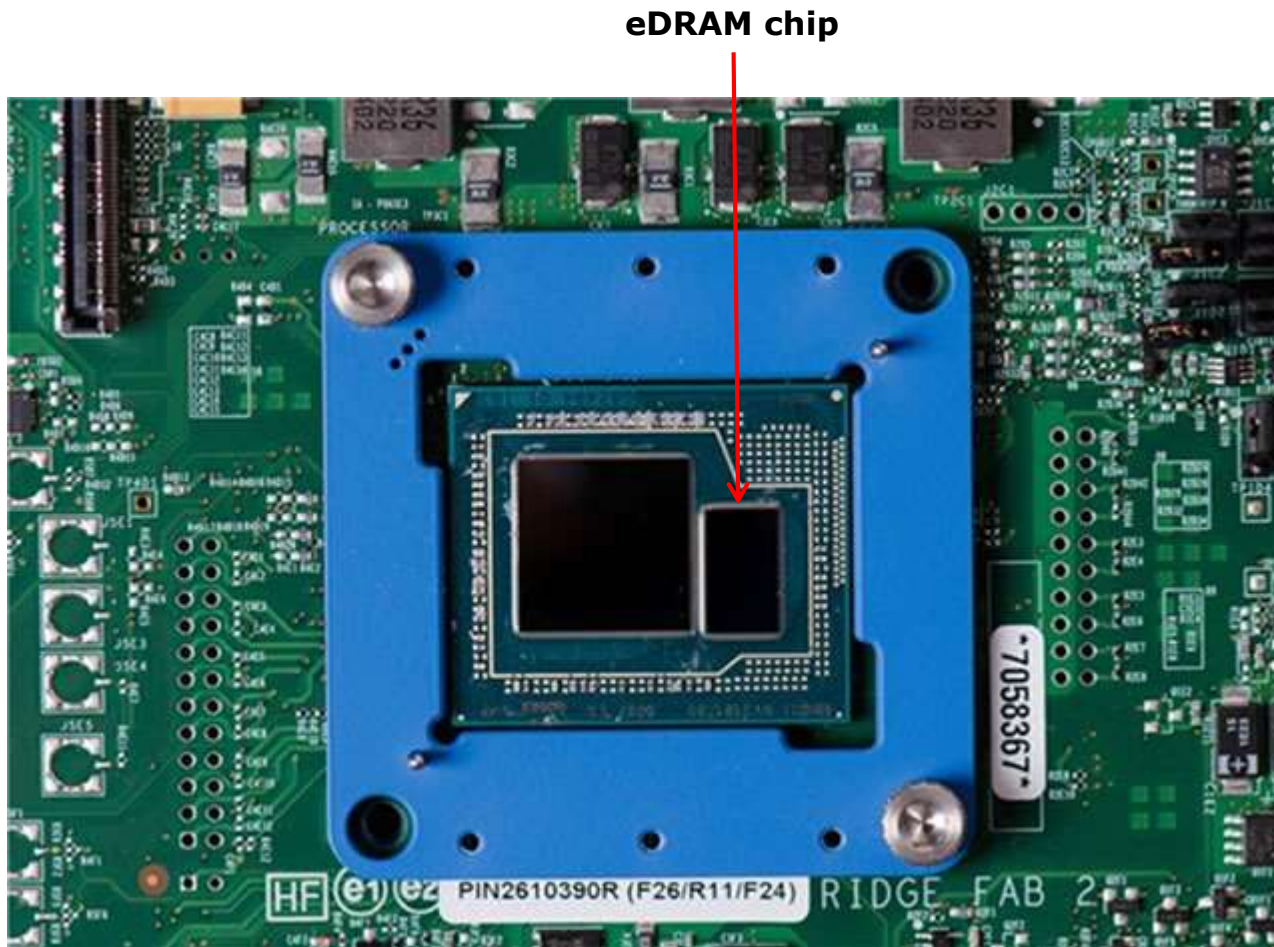
- AVX Doubled Floating Point (FP) vector width & Register File
 - Doubling peak flops
- AVX2 Doubles Integer vector width
- 2x 256-bit FMA units (Haswell)
 - Doubling peak flops again

IDF2012

FMA: Fused Multiply-Add

5.1 Introduction to the Haswell line (7)

b) On-package eDRAM L4 cache [124]



c) FIVR (Fully Integrated Voltage Regulator)

FIVR integrates legacy power delivery onto the package and the die, as shown below for Intel's Haswell processor [178]

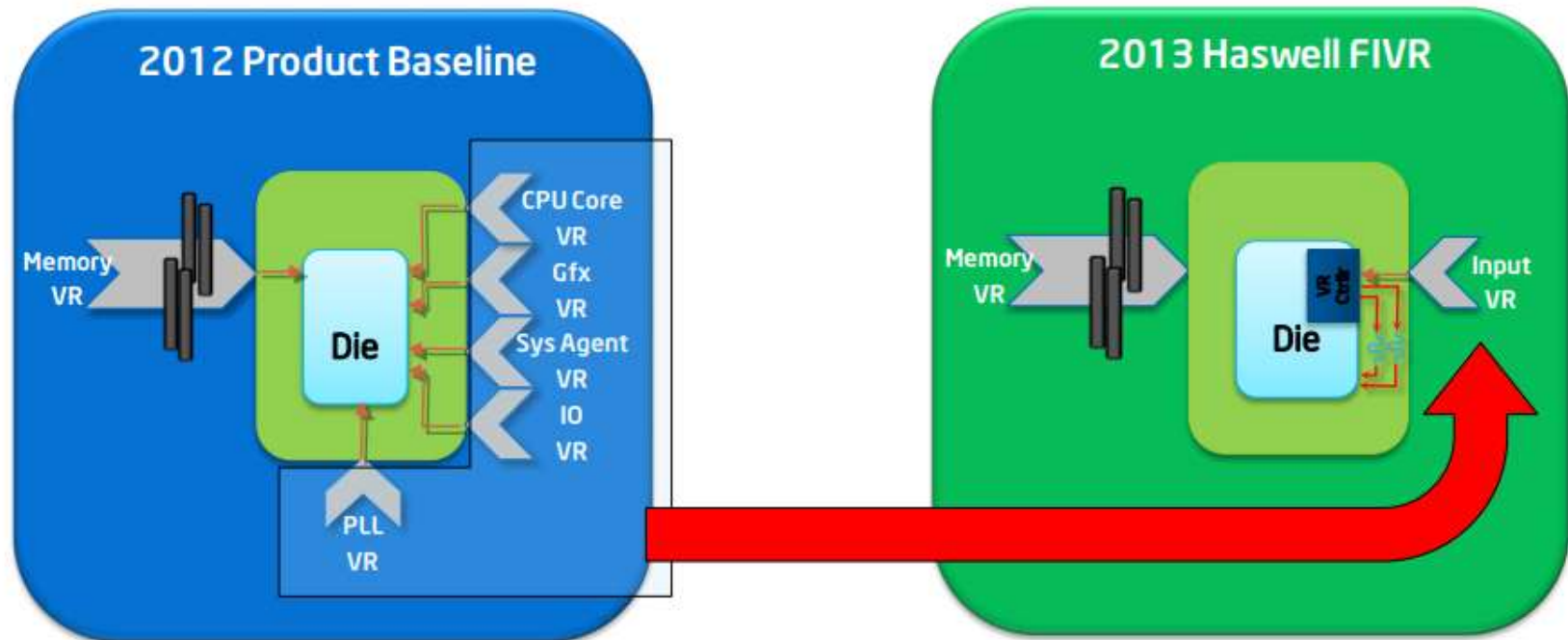


Figure: Integrating legacy power delivery onto the package and the die with FIVR [178]

This consolidates **five platform VRs into one** and thus **greatly simplifies mainboard design**.

d) TSX (Transactional Synchronization Extensions)

- Haswell also introduced a further new feature, the **Transactional Synchronization Extensions (TSX)** that was debuted on selected Haswell models (SKUs).
- **TSX supports Transactional Memory in hardware** (to be discussed later in the Chapter high end MP servers).
- Nevertheless, **in August 2014 Intel announced a bug** in the TSX implementation on all current steppings of all Haswell models and **disabled the TSX feature** on affected CPUs **via a microcode update**.
- Subsequently, TSX became **enabled** first **on a Broadwell model** (Core M-5Y70) **in 11/2014** then on the **Haswell-EX in 5/2015**.

Addressing race conditions of thread execution while accessing memory

Basically there are **two mechanisms to address race conditions** in multithreaded programs, as indicated below:

Basic mechanisms to address races in multithreaded programs



Locks

Pessimistic approach,
it intends to prevent possible conflicts
by enforcing serialization of transactions
through locks.

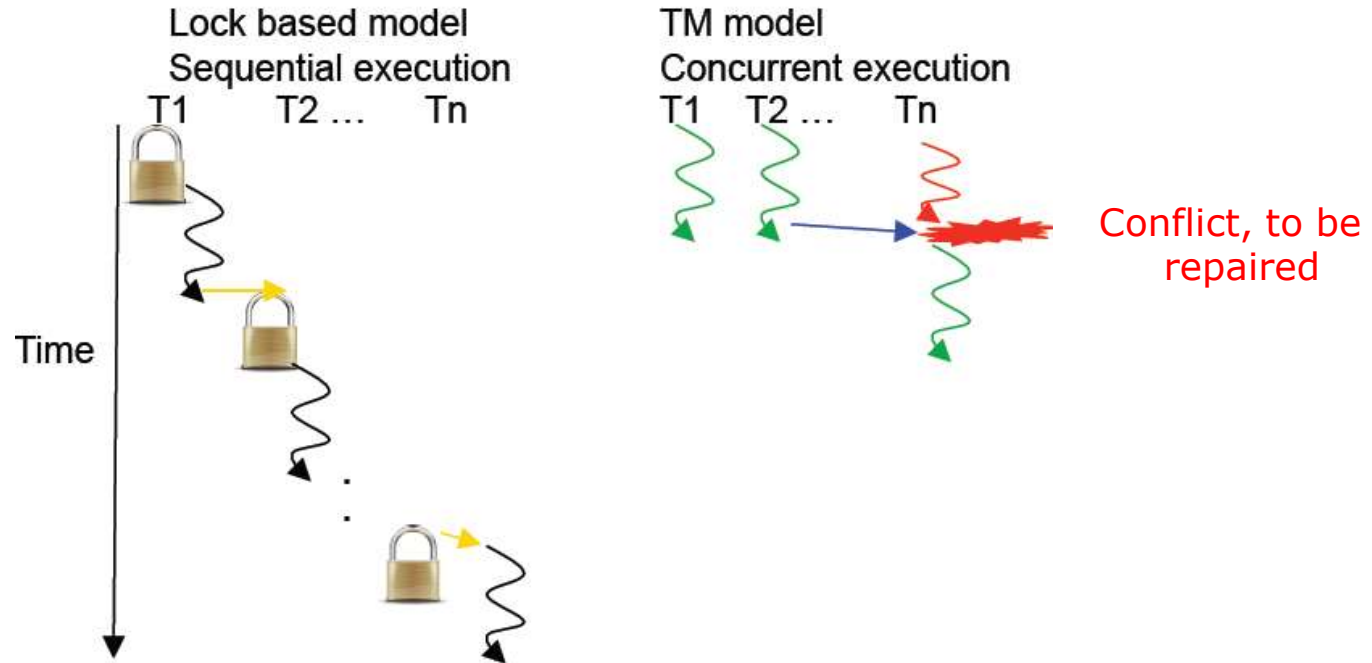
Transactional memory (TM)

Optimistic approach,
it allows access conflicts to occur
but provides a checking and repair mechanism
for managing these conflicts, i.e.
it allows all threads to access shared data simultaneously
but after completing a transaction,
it will be checked whether a conflict arose,
if yes, the transaction will be rolled back and
then replayed if feasible else
executed while using locks.

The next Figure illustrates these synchronization mechanisms.

5.1 Introduction to the Haswell line (11)

Illustration of lock based and transaction memory (TM) based thread synchronization [126]

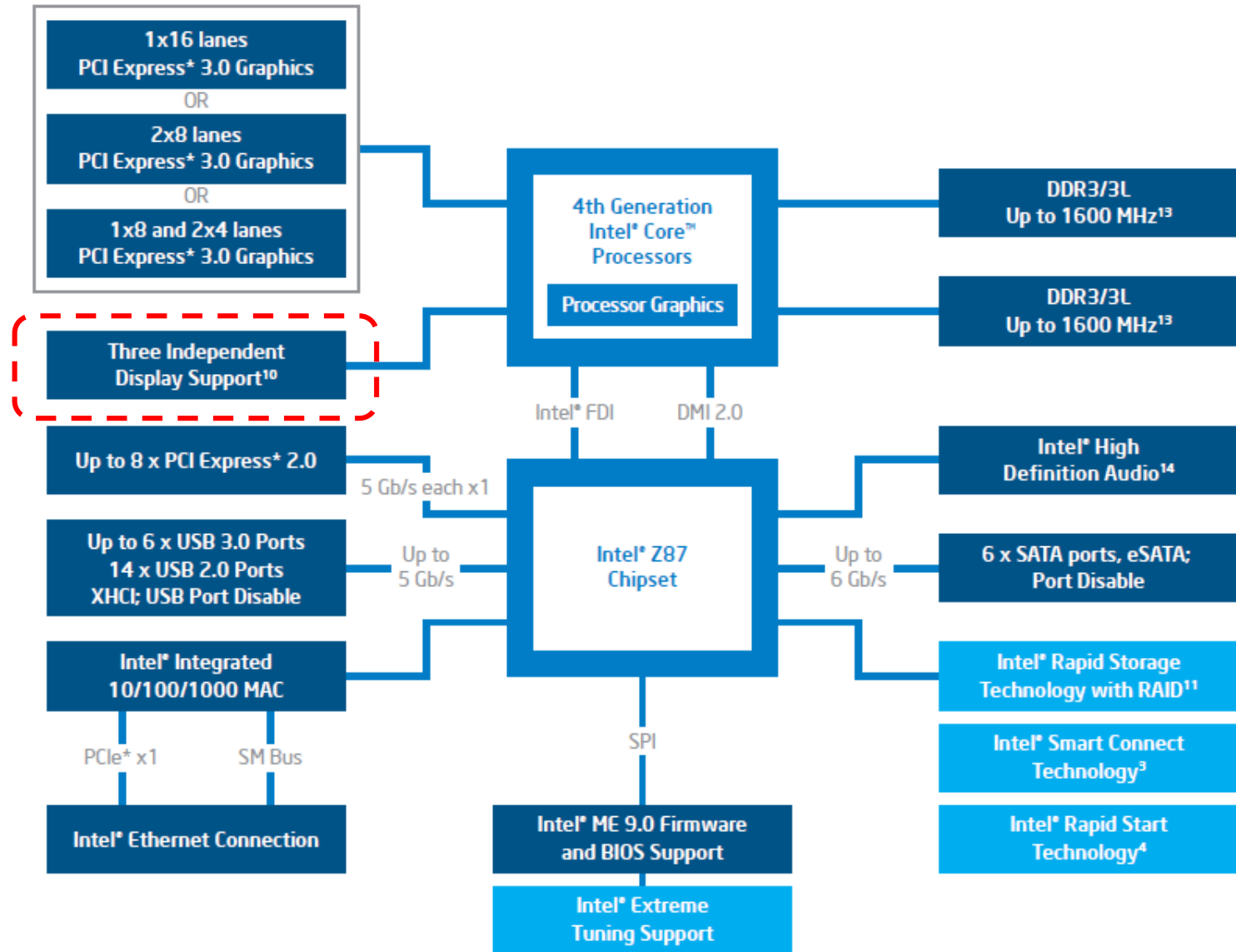


Additional platform related innovations

- a) Connecting the displays directly to the processor
- b) On-package integrated CPU and PCH for mobile processors

5.1 Introduction to the Haswell line (13)

a) Connecting the displays directly to the processor [145]



b) On-package integrated CPU and PCH for mobile processors [204]

New BGA Strategy for Ultra-Thin Devices

Integration Drives Lower Power, Smaller Designs

New! 1-Chip BGA Solution



- CPU and PCH integrated into single BGA package
- 15W & 28W TDPs, 6W and below SDP
- S0ix support
- Supports LPDDR3 and DDR3L memory

Traditional 2-Chip platform



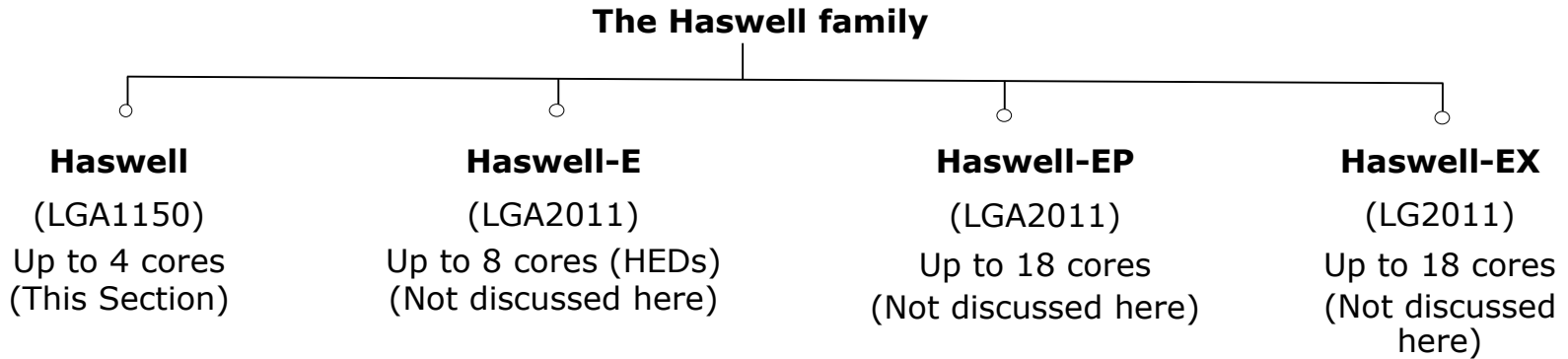
- 2 chip scalable solution: CPU and chipset
- BGA and rPGA packages
- 57W, 47W, and 37W TDPs
- Supports DDR3L Memory
- GT3e graphics

BGA: Ball Grid Array

SDP: Scenario Design Point

PGA: Pin Grid array

Overview of the Haswell family



Mobiles (SoCs)

Core i7-49xx/48xx/472x/471x/470x, 4C+G, HT, 6/2013 and 5/2014

Core i7-46xx/45xx, 2C+G, HT, 5/2013 and 6/2014

Core i5-43xx/42xx U/Y, 2C+G, HT, 6/2013 and 5/2014

Core i3-41xx/40xx, 2C+G, HT, 6/2013 and 5/2014

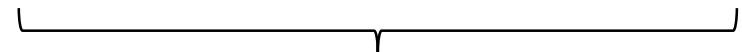
Desktops (2-chip designs)

Core i7-479x/478x/477x,476x, 4C+G, HT, 6/2013 and 5/2014

Core i5-46xx/45xx/44xx, 4C+G, HT, 6/2013 and 5/2014

Core i3-43xx/41xx, 2C+G, HT, 6/2013, 5/2014 and 3/2015

i7-5960X/5930K/5820K, 6/8 C, 8/2014

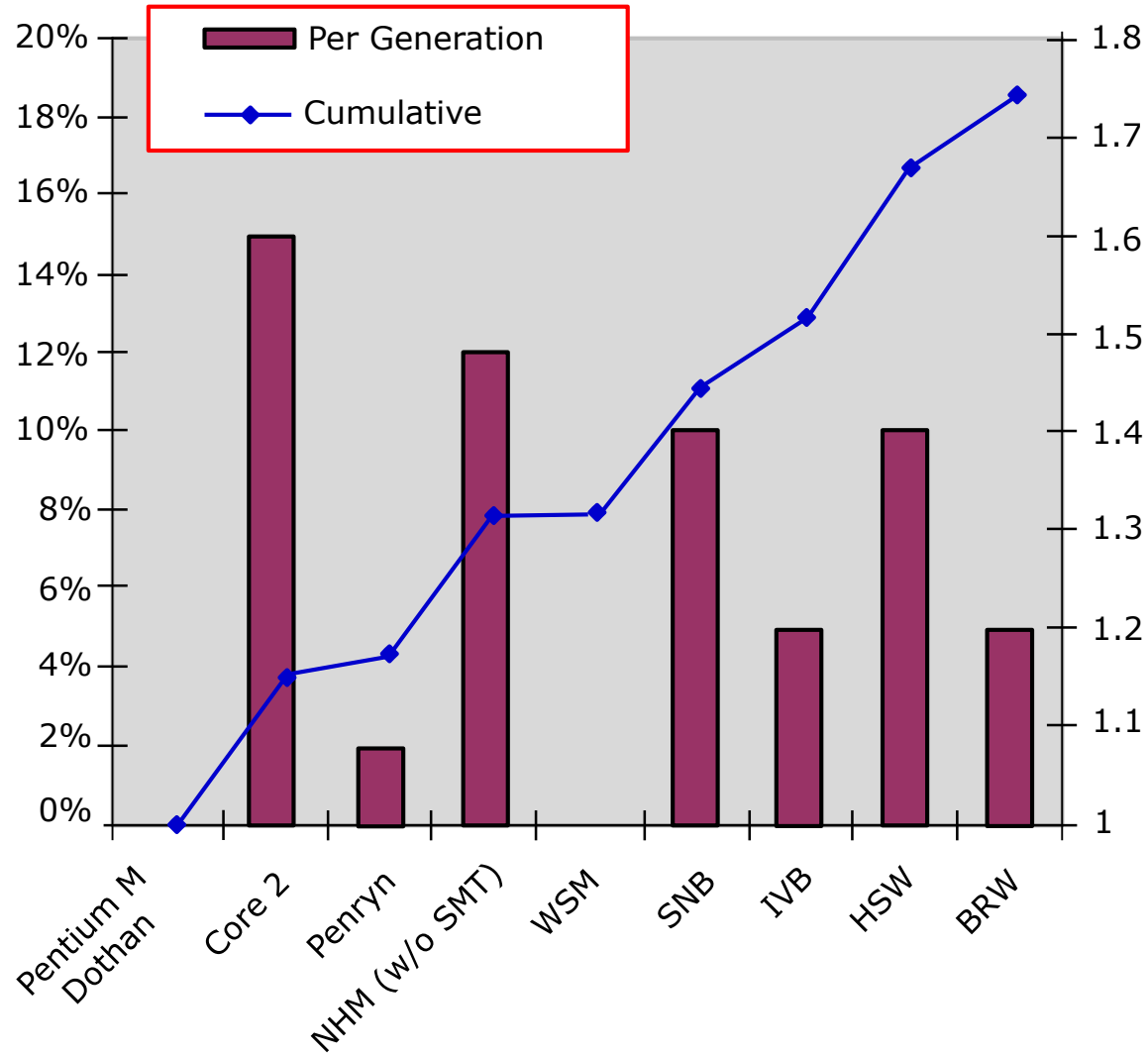


Servers

(Not discussed here)

5.1 Introduction to the Haswell line (16)

Single thread IPC in Intel's basic architectures [195]



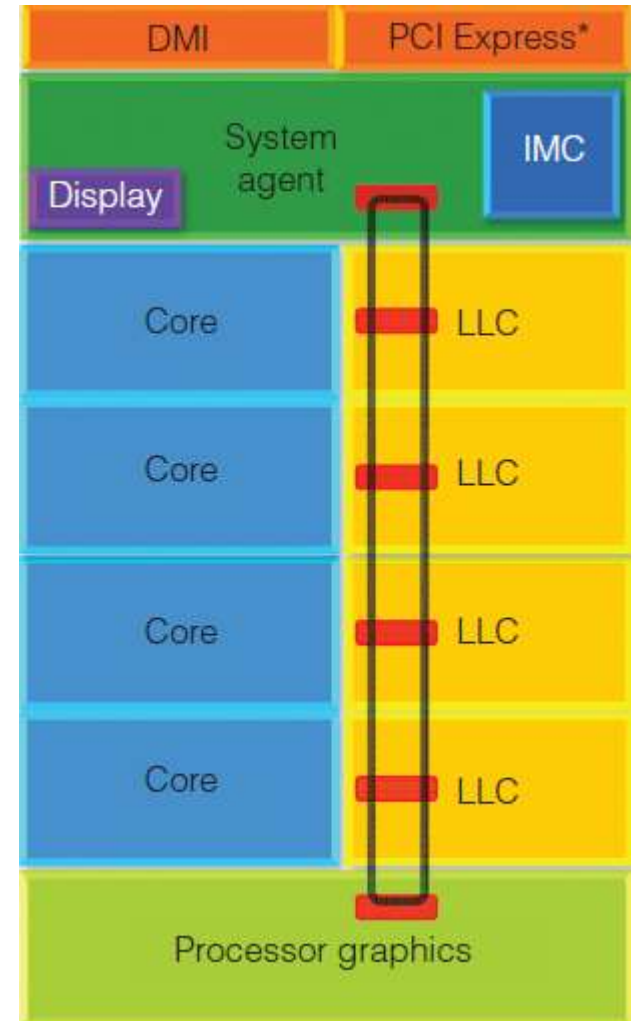
5.2 Major enhancements of the Haswell line vs. the Sandy Bridge line

- 5.2.1 Overview
- 5.2.2 ISA extension (of the cores) by the AVX2 instruction set
- 5.2.3 Enhanced microarchitecture for the cores
- 5.2.4 Enhanced graphics

5.2 Major enhancements of the Haswell line vs. the Sandy Bridge line

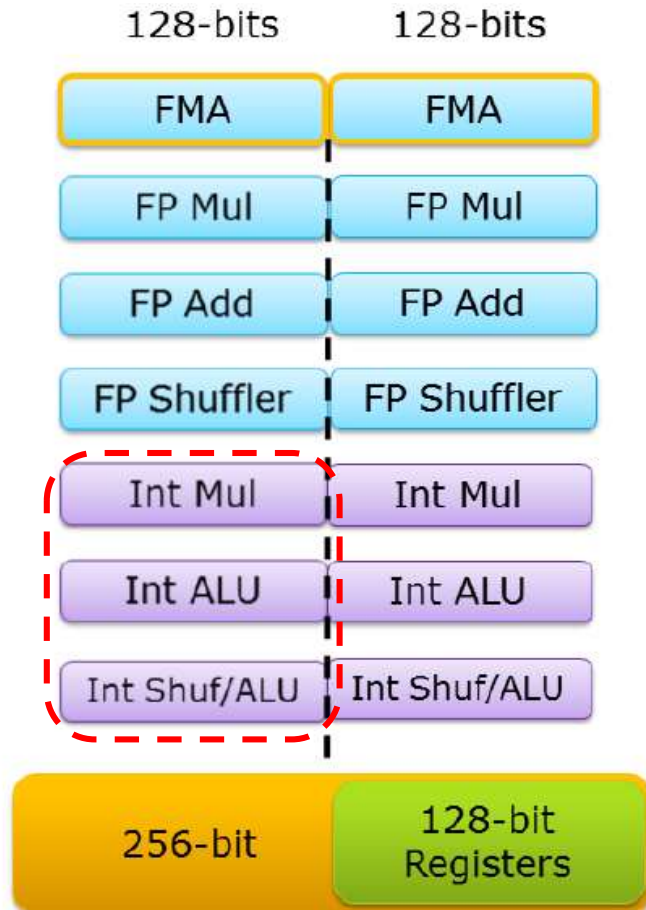
5.2.1 Overview

- ISA extension by the AVX2 instruction set (Section 5.2.2)
- Enhanced microarchitecture for the cores (Section 5.2.3)
- Enhanced graphics (Section 5.2.4)



5.2.2 ISA extension by the AVX2 instruction set (1)

5.2.2 ISA extension by the AVX2 instruction set -1 [97]



- AVX Doubled Floating Point (FP) vector width & Register File
 - Doubling peak flops
- AVX2 Doubles Integer vector width
- 2x 256-bit FMA units (Haswell)
 - Doubling peak flops again

IDF2012

FMA: Fused Multiply-Add

5.2.2 ISA extension by the AVX2 instruction set (2)

ISA extension by the AVX2 instruction set -2 [80]

- Intel® Advanced Vector Extensions 2 (Intel® AVX2)

- Includes

- 256-bit Integer vectors
 - FMA: Fused Multiply-Add
 - Full-width element permutes
 - Gather

- Benefits

- High performance computing
 - Audio & Video
 - Games

- New Integer Instructions

- Indexing and hashing
 - Cryptography
 - Endian conversion – MOVBE

	Instruction Set	SP FLOPs per cycle	DP FLOPs per cycle
Nehalem	SSE (128-bits)	8	4
Sandy Bridge	AVX (256-bits)	16	8
Haswell	AVX2 & FMA	32	16

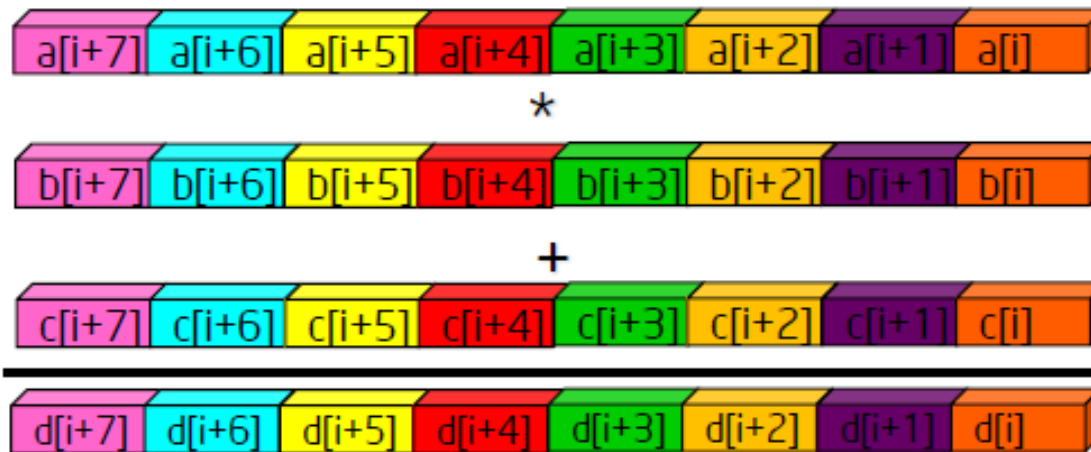
Group	Instructions
Bit Field Pack/Extract	BZHI, SHLX, SHRX, SARX, BEXTR
Variable Bit Length Stream Decode	LZCNT, TZCNT, BLSR, BLSMSK, BLSI, ANDN
Bit Gather/Scatter	PDEP, PEXT
Arbitrary Precision Arithmetic & Hashing	MULX, RORX

- Full Instruction Specification Available at: <http://software.intel.com/en-us/avx/>

5.2.2 ISA extension by the AVX2 instruction set (3)

Example for calculating $D = A \times B + C$ with 8 32-bit data vectors by using the AVX2 ISA extension [212]

```
for (i=0;i<=MAX;i++)  
    d[i]=((a[i]*b[i]) + c[i]);
```



AVX2 Vector
- One Instruction
- 16 Mathematical Operations¹

1. Number of operations per instruction varies based on the which SIMD instruction is used and the width of the operands
8 of the operations are multiplications and 8 are additions (the addition of the multiplication result to a third operand)
1. 8 Multiplication operations + 8 Addition multiplications

5.2.2 ISA extension by the AVX2 instruction set (4)

Remark

Reduced core frequency while running AVX instructions

- When the processor detects **AVX instructions** it signals it to the PCU (Power Control Unit).
- Then the **PCU delivers higher core voltage** that however increases the dissipation.
- At the same time while executing AVX instructions the **PCU reduces the clock frequency of the processor to remain within the TDP limit** and avoid overheating.
- The higher voltage will remain for 1 ms after the last AVX instruction completes and , subsequently the core voltage will return to its nominal value defined by the TDP.
- Related to this, Intel added **two AVX frequencies for their Haswell- EP (E5-1600 and E5-2600 line of processors**, as follows and demonstrated in an example.
 - **AVX base**: it is the **minimum frequency for workloíds using AVX instructions**.
 - **AVX max all core Turbo**: it is the **maximum frequency for workloads using all cores for executing AVX instructions**.

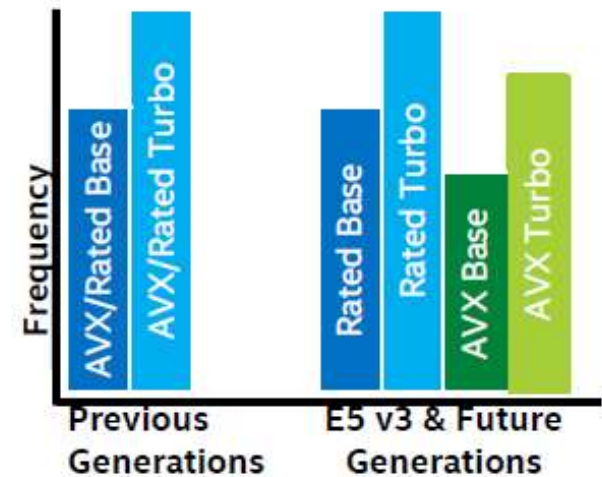
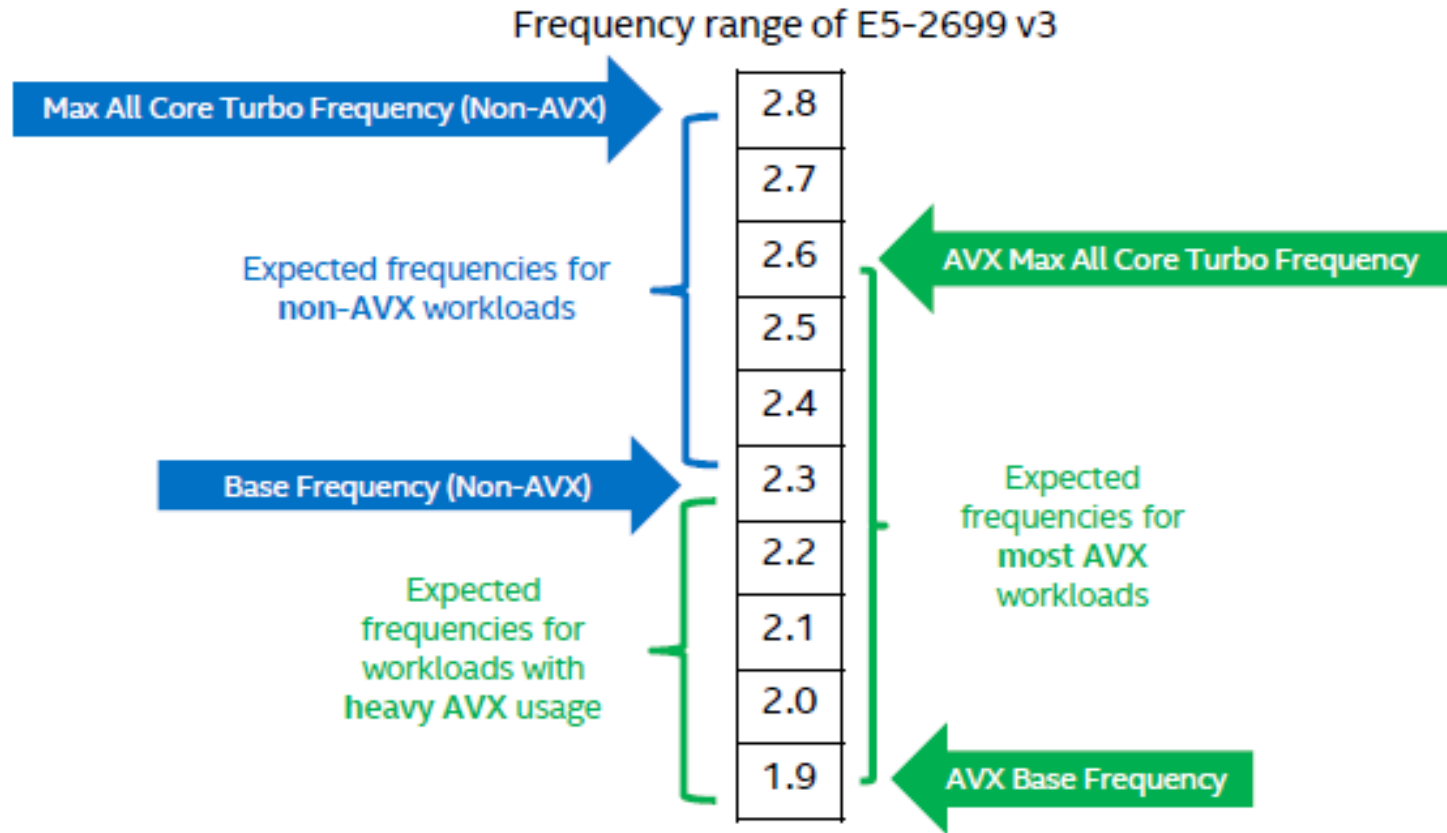


Figure: Core frequency limits in Haswell-EP and previous lines [212]

(128-bit execution)

5.2.2 ISA extension by the AVX2 instruction set (5)

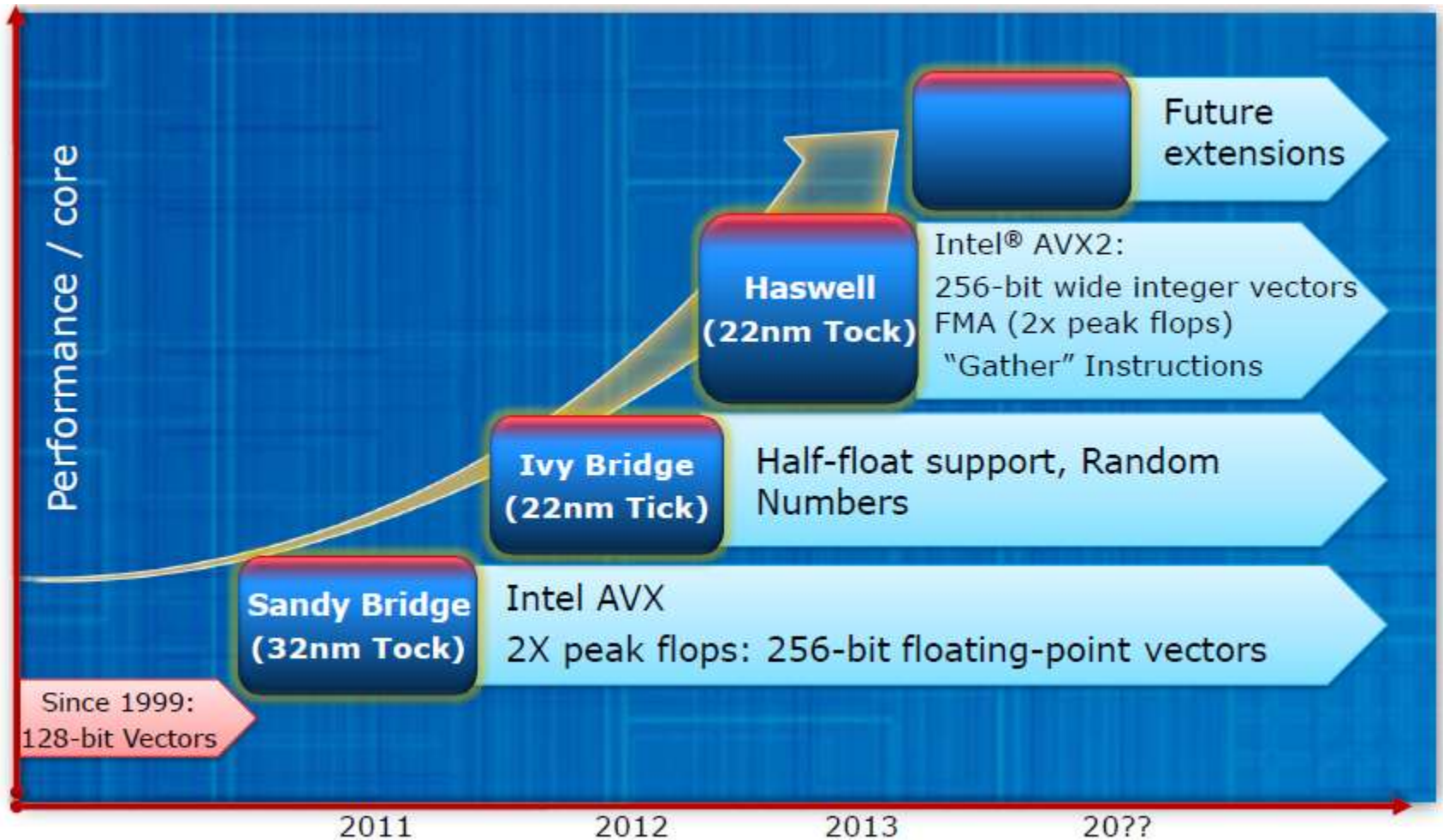
Example: Core frequency limits for Intel's E5-2699 v3 processor [212]



AVX Frequency Range Example – E5-2699 v3

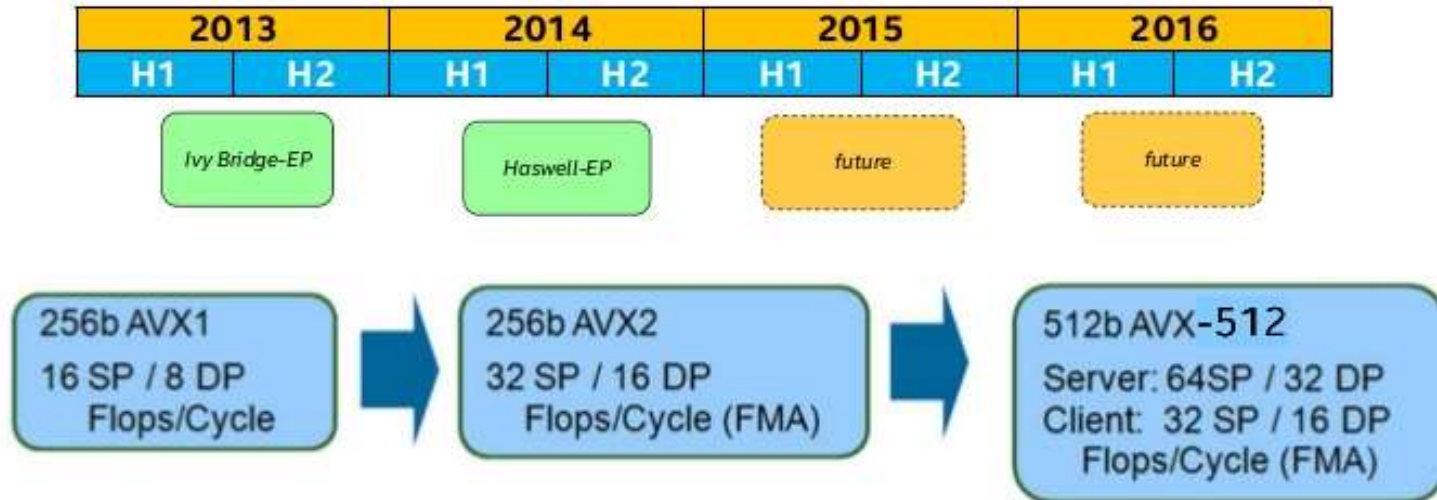
5.2.2 ISA extension by the AVX2 instruction set (6)

Evolution of the AVX ISA extensions [97]



5.2.2 ISA extension by the AVX2 instruction set (7)

Expected future evolution of AVX [165]

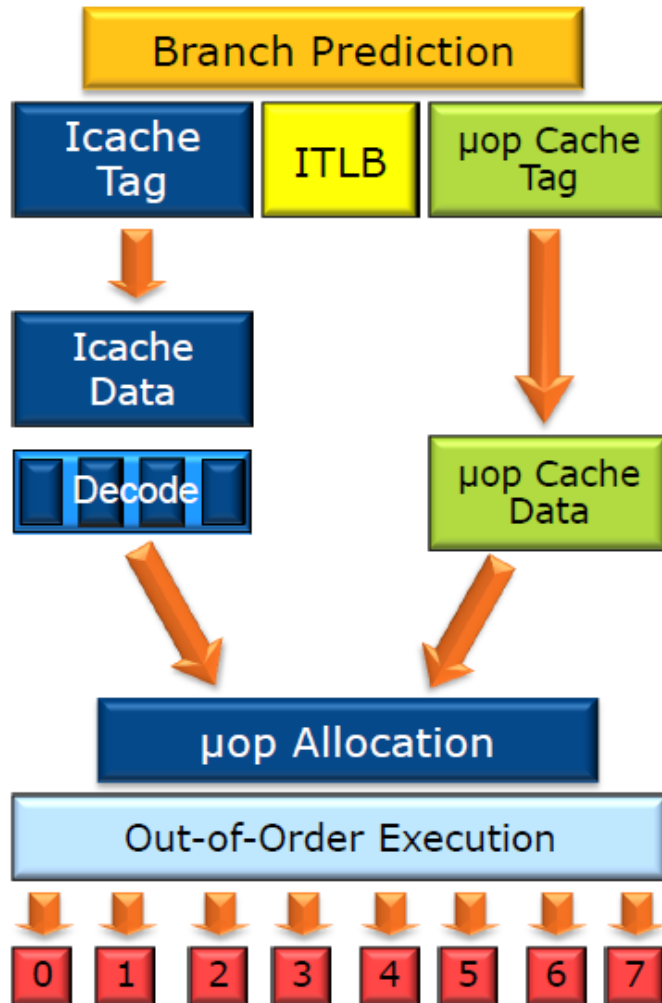


AVX	AVX2
256-bit basic FP 16 registers NDS (and AVX128) Improved blend MASKMOV Implicit unaligned	Float16 (IVB 2012) 256-bit FP FMA 256-bit integer PERMD Gather

AVX-512

512-bit FP/Integer
32 registers
8 mask registers
Embedded rounding
Embedded broadcast
Scalar/SSE/AVX "promotions"
Native media additions
Transcendental support
Gather/Scatter

5.2.3 Enhanced microarchitecture for the cores [80]



Next generation branch prediction

- Improves performance *and* saves wasted work

Improved front-end

- Initiate TLB and cache misses speculatively
- Handle cache misses in parallel to hide latency
- Leverages improved branch prediction

Deeper buffers

- Extract more instruction parallelism
- More resources when running a single thread

More execution units, shorter latencies

- Power down when not in use

More load/store bandwidth

- Better prefetching, better cache line split latency & throughput, double L2 bandwidth
- New modes save power without losing performance








No pipeline growth

- Same branch misprediction latency
- Same L1/L2 cache latency

5.2.3 Enhanced microarchitecture for the cores (2)

Buffer sizes of subsequent generations of the Core processors [80]

Extract more parallelism in every generation

	Nehalem	Sandy Bridge	Haswell	
Out-of-order Window	128	168	192	
In-flight Loads	48	64	72	
In-flight Stores	32	36	42	
Scheduler Entries	36	54	60	
Integer Register File	N/A	160	168	
FP Register File	N/A	144	168	
Allocation Queue	28/thread	28/thread	56	

5.2.3 Enhanced microarchitecture for the cores (3)

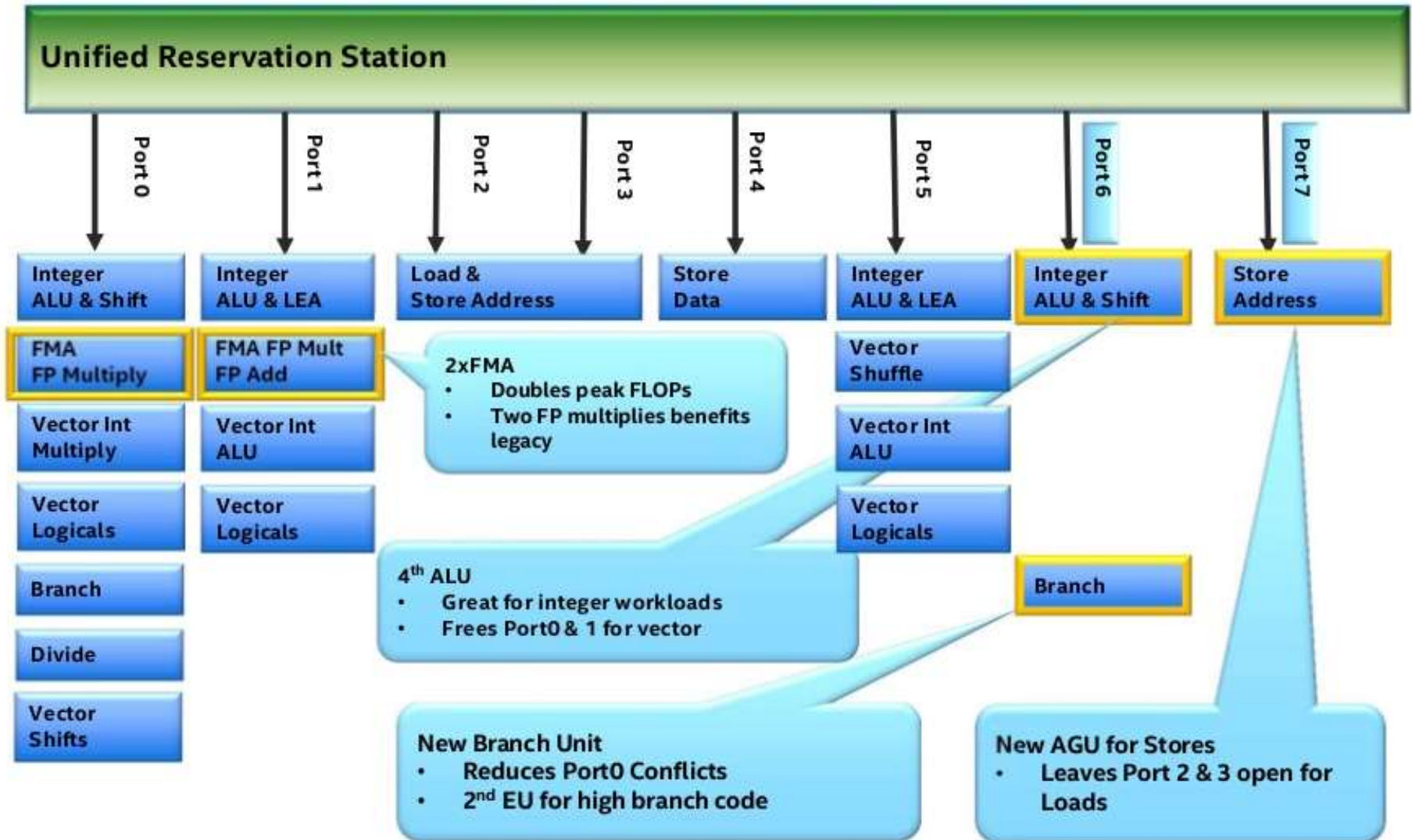
Cache sizes, latencies and bandwidth values of subsequent Core generations [122]

Metric	Nehalem	Sandy Bridge	Haswell
L1 Instruction Cache	32K, 4-way	32K, 8-way	32K, 8-way
L1 Data Cache	32K, 8-way	32K, 8-way	32K, 8-way
Fastest Load-to-use	4 cycles	4 cycles	4 cycles
Load bandwidth	16 Bytes/cycle	32 Bytes/cycle (banked)	64 Bytes/cycle
Store bandwidth	16 Bytes/cycle	16 Bytes/cycle	32 Bytes/cycle
L2 Unified Cache	256K, 8-way	256K, 8-way	256K, 8-way
Fastest load-to-use	10 cycles	11 cycles	11 cycles
Bandwidth to L1	32 Bytes/cycle	32 Bytes/cycle	64 Bytes/cycle
L1 Instruction TLB	4K: 128, 4-way 2M/4M: 7/thread	4K: 128, 4-way 2M/4M: 8/thread	4K: 128, 4-way 2M/4M: 8/thread
L1 Data TLB	4K: 64, 4-way 2M/4M: 32, 4-way 1G: fractured	4K: 64, 4-way 2M/4M: 32, 4-way 1G: 4, 4-way	4K: 64, 4-way 2M/4M: 32, 4-way 1G: 4, 4-way
L2 Unified TLB	4K: 512, 4-way	4K: 512, 4-way	4K+2M shared: 1024, 8-way

All caches use 64-byte lines

5.2.3 Enhanced microarchitecture for the cores (4)

Issue rate and execution unit enhancements of Haswell [165]



FMA: Fused Multiply-Add ($a \times b + c$), 256-bit execution (and lanes)

5.2.4 Enhanced graphics

- To compete with AMD's advanced graphics solutions Intel put a great emphasis on enhancing Haswell's integrated graphics.
- Main features of the new graphics units: are termed as Iris Pro and Iris graphics.
 - a) Sliced graphics architecture to allow scaling of EUs
 - b) Inclusion of eDRAM in high-end units.

a) Introduction of sliced graphics architecture

- The new graphics architecture of Haswell is **sliced**, to allow **scaling of EUs** by using **one or two slices/unit**.
- Each **slice** has **two sub-slices**, with up to **10 EUs/sub-slice**, as indicated in the Figure below.

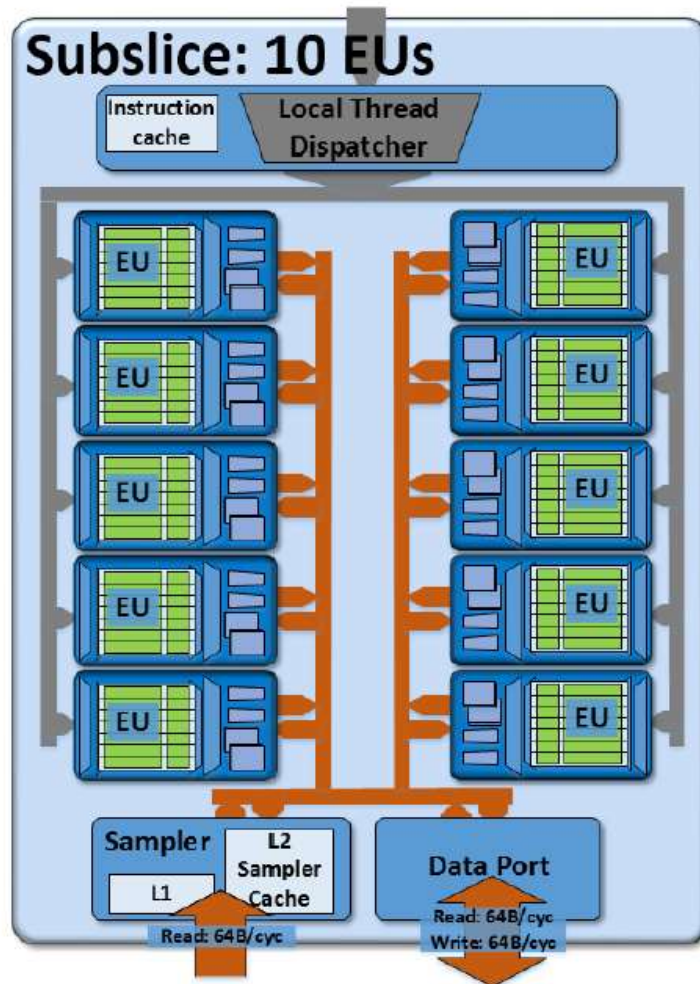
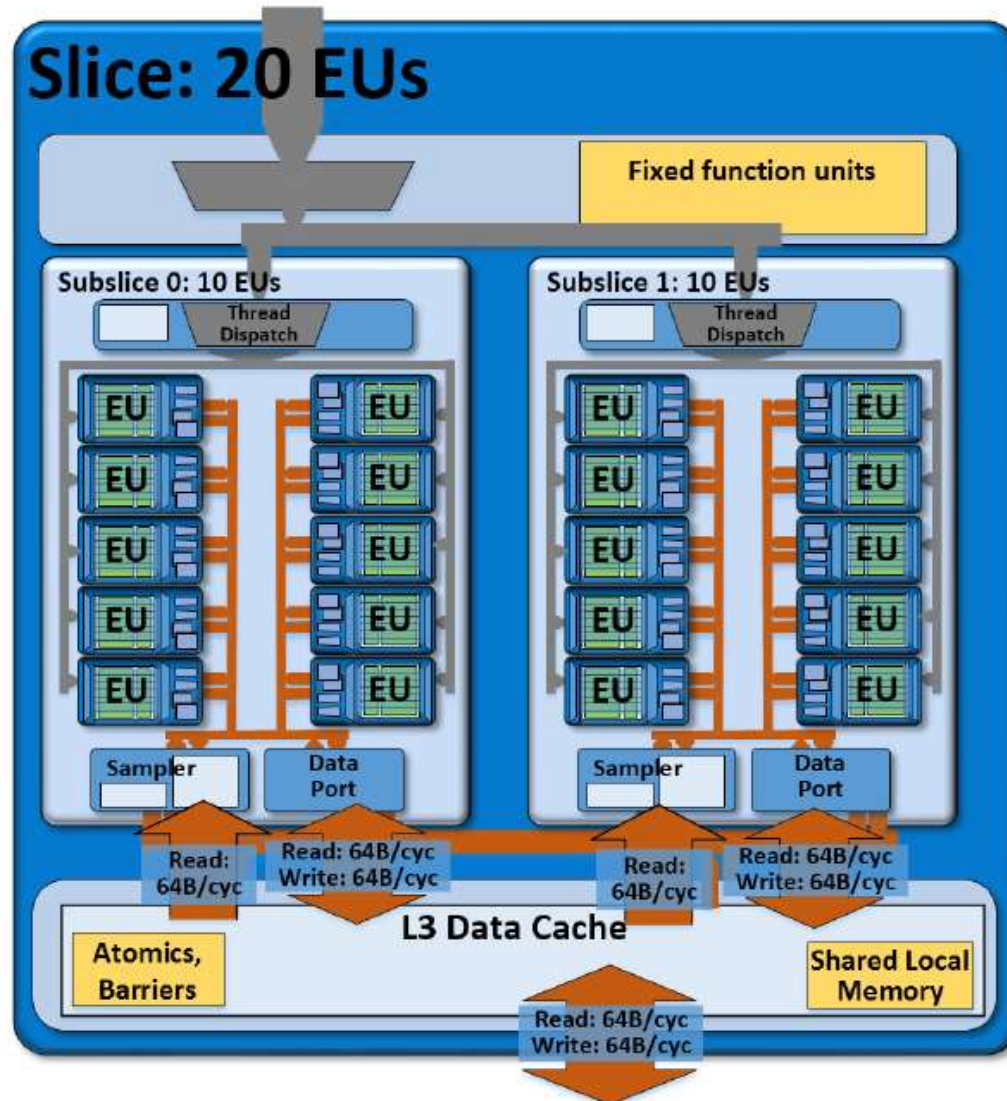


Figure: A sub-slice with 10 EUs of the graphics unit of Haswell [199]

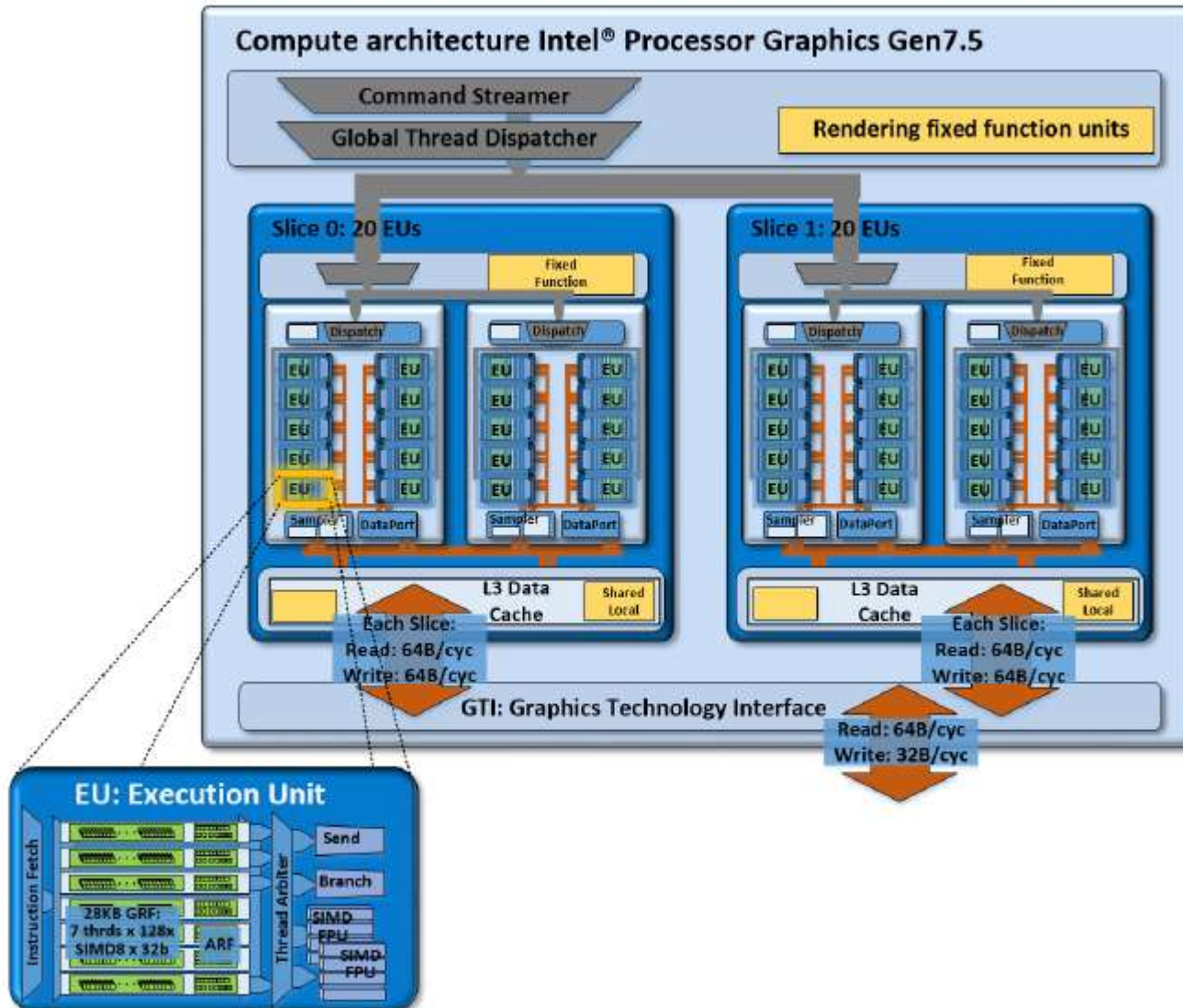
5.2.4 Enhanced graphics (3)

A slice of the graphics unit of Haswell including two sub-slices with 20 EUs [199]

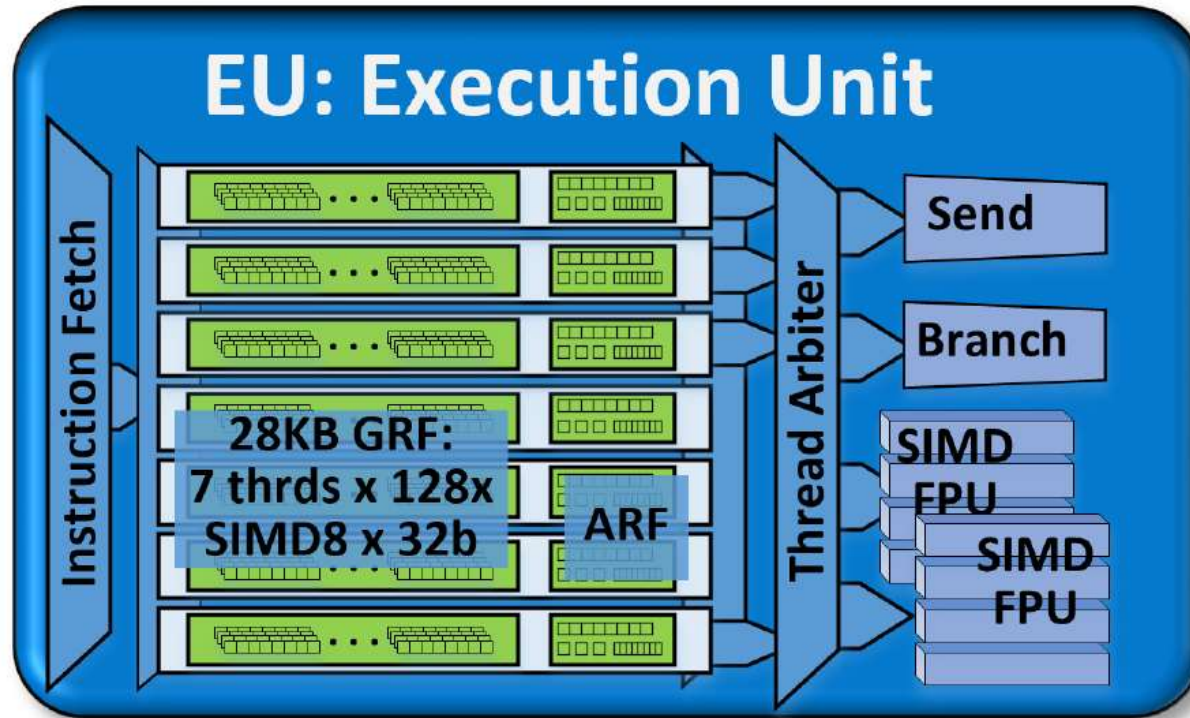


5.2.4 Enhanced graphics (4)

The architecture of a GT3 graphics unit of Haswell including two slices with 40 EUs [199]

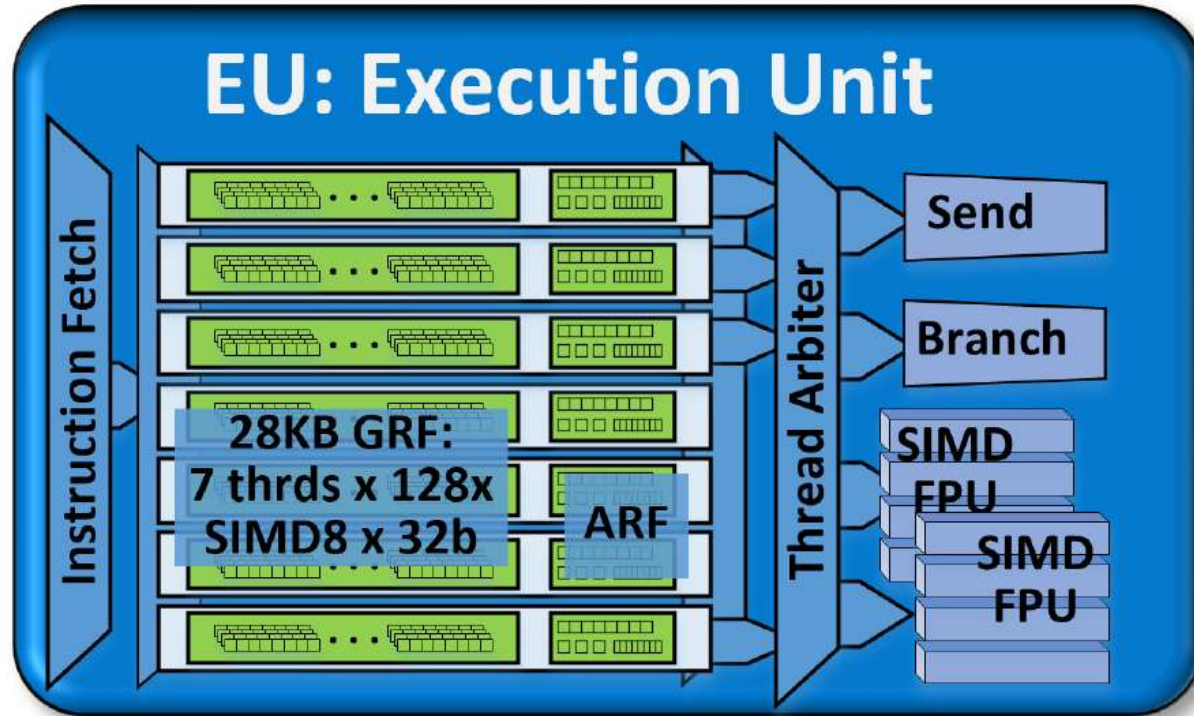


Block diagram of an EU of Haswell -1 [199]



- Each EU has **four functional units**:
 - **Two SIMD FPU** units
 - **1 Send unit** (Load/Store) and
 - **1 Branch unit**.
- An EU issues **up to 4 instructions per cycle** to the functional units.

Block diagram of an EU of Haswell -2 [199]



- Each **SIMD FPUs** can execute 4 SP FP or 1/2/4/8/16/32 bit wide FX operations.
- They can execute **MAD** instructions (Multiply-Add) per cycle.
- Thus an **EU** can execute 2 FPU x SIMD4 x 2 (MAD) = 16 SP FP operations/cycle.
- The EU is 7-way multithreaded.
- Each thread has 128 32 B registers.
- One of the FPUs also supports FX operations.
- One of the FPUs also support transcendental math functions.

Interpretation of the notions Graphics Technology (GT) for the Haswell line and subsequent lines

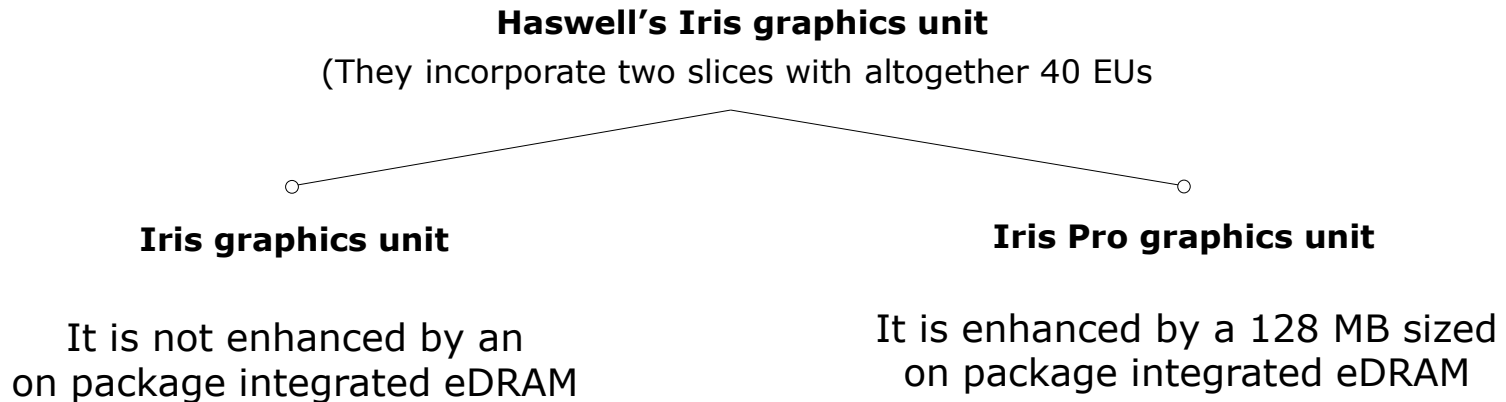
- **GT1** designates graphics with a **single slice and reduces execution resources** (less sub-slices or less EUs per sub-slice)
- **GT2** designates graphics with a **single slice** (e.g. 20 EUs in the Haswell line)
- **GT3** designates graphics with **dual slices** (e.g. 40 EUs in the Haswell line)
- **GT4** designates graphics with **triple slices**.

Introducing the notions Iris/Iris Pro graphics

Intel designates their

- high-end graphics as **Iris** graphics and
- high-end graphics enhanced with embedded DRAM (eDRAM) as **Iris Pro** graphics.

The inclusion of eDRAM will be indicated also in the GT naming by supplementing the GT level by the letter "e", so GT3e designates GT3 level with eDRAM.



b) Inclusion of eDRAM in high-end graphics units

It will be discussed in Section 5.3.1

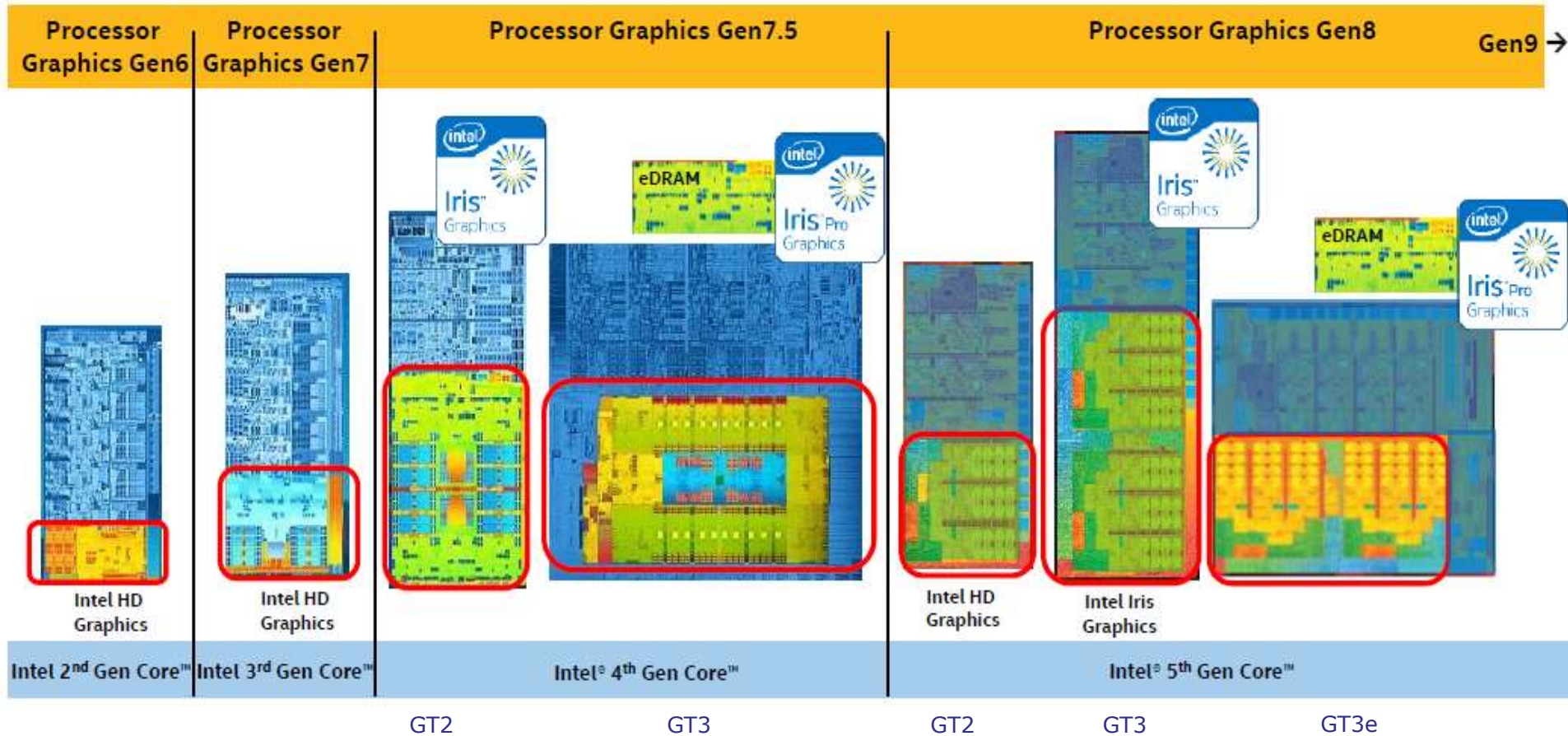
5.2.4 Enhanced graphics (10)

Evolution of main features of Intel's graphics families

Intel Core generation	Graphics generation	Models	Graphics Technology level	No. of graphics slices	No. of EUs	eDRAM	OpenGL version	DirectX version	OpenCL version
Westmere	5 th (Ironlake)	HD			12	--	2.1	10.1	n.a.
Sandy Bridge	6 th	HD 2000	GT1	1 (2x3 EU)	6		3.1/3.3	10.1	n.a.
		HD 3000	GT2	1 (4x3 EU)	12				
Ivy Bridge	7 th	HD 2500	GT1	1 (6 EU)	6	--	4.0	11.0	1.2
		HD 4000	GT2	1 (2x8 EU)	16				
Haswell	7.5 th	HD 4200- HD 4700	GT2	1 (2x10 EU)	20	--	4.3	11.1	1.2
		HD 5000 Iris 5100	GT3	2	40				
		Iris Pro 5200				128 MB			
Broadwell	8 th	HD 5300- HD 5600	GT2	1(3x8 EU)	23/24	--	4.3	11.2	2.0
		HD 6000 Iris 6100	GT3	2	47/48				
		Iris Pro 6200	GT3e	2	48	128 MB			
Skylake	9 th	HD 510	GT1	1 (3x4 EU)	12	--	4.4	12	2.0
		HD 515	GT1.5	1 (3x6 EU)	18				
		HD 520	GT2	1 (3x8 EU)	24				
		HD 535	GT3	2	48				
		HD 540	GT3e	2	48	64 MB			
		HD 580	GT4e	3	72	64/128 MB			

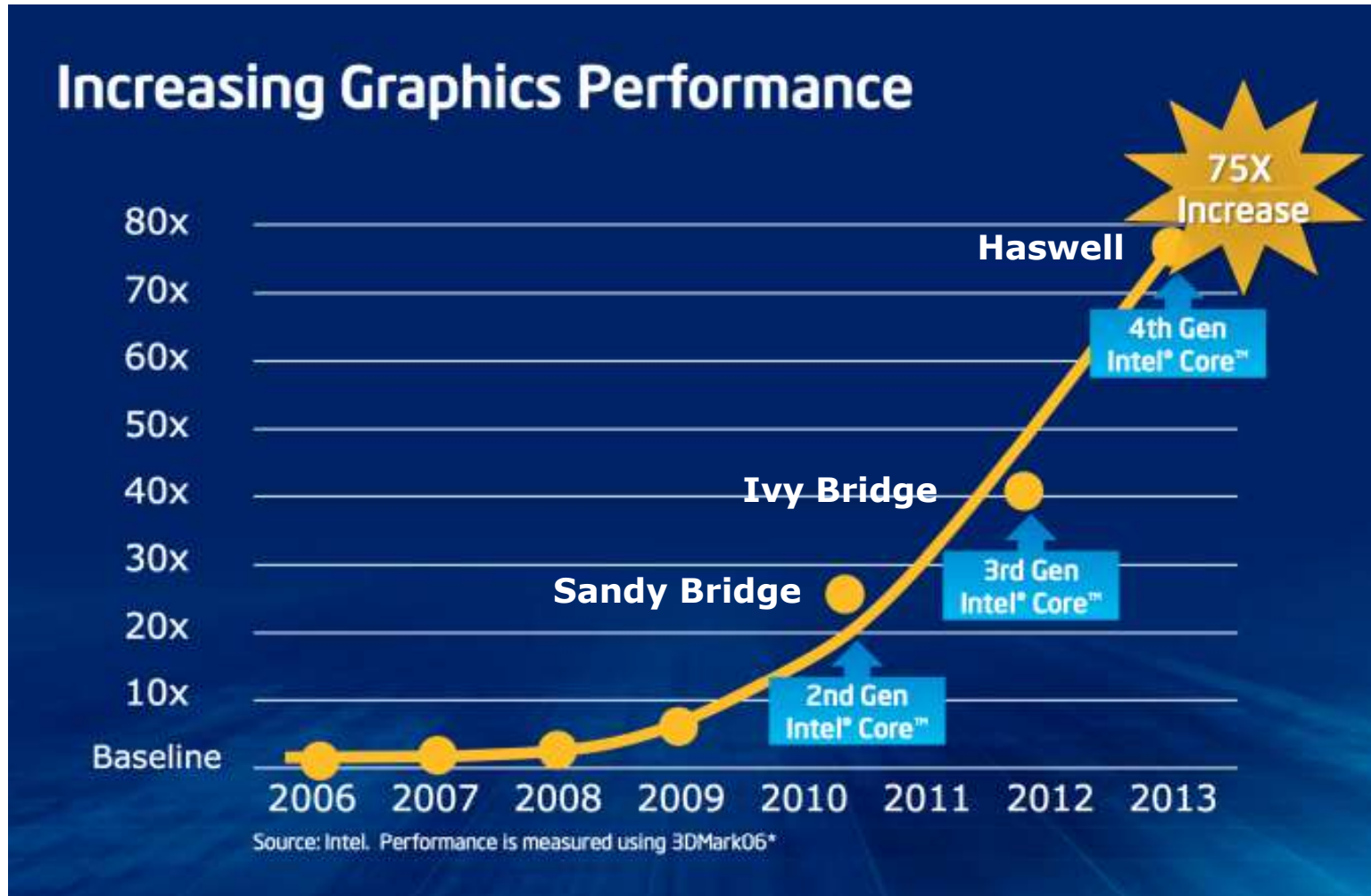
5.2.4 Enhanced graphics (11)

Evolving integrated graphics in Intel's processor generations [250]



Note that evolving processor graphics is Intel's primary interest to compete with NVIDIA and AMD.

Graphics performance increase of subsequent Core generations [117]



5.3 Major innovations of the Haswell line

- 5.3.1 In-package eDRAM cache
- 5.3.2 FIVR (Fully Integrated Voltage Regulator)
- 5.3.3 TSX (Transactional Synchronization Extensions)

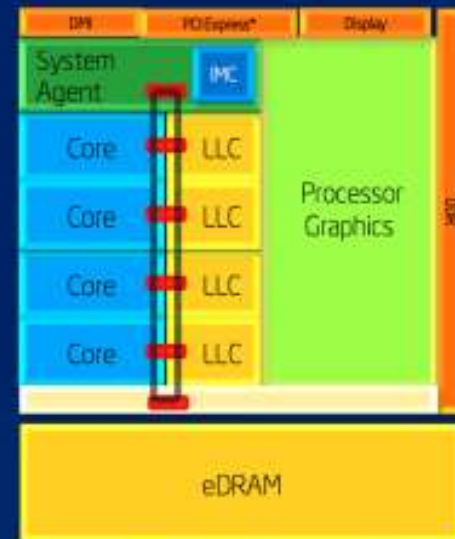
5.3 Major innovations of the Haswell line

- On-package e-DRAM cache (Section 5.3.1)
- FIVR (Fully Integrated Voltage Regulator) (Section 5.3.2)
- TSX (Transactional Synchronization Extensions) (Section 5.3)

5.3.1 On-package eDRAM cache [117]

On Package eDRAM

- Haswell introduces configurations with large graphics & an on-package eDRAM cache
- Cache attributes
 - High throughput and low latency
 - Flat power vs. sustained bandwidth curve
 - Fully shared between Graphics, Media, and Cores
- Low latency on package interface to CPU



On-package eDRAM enables discrete-class graphics performance

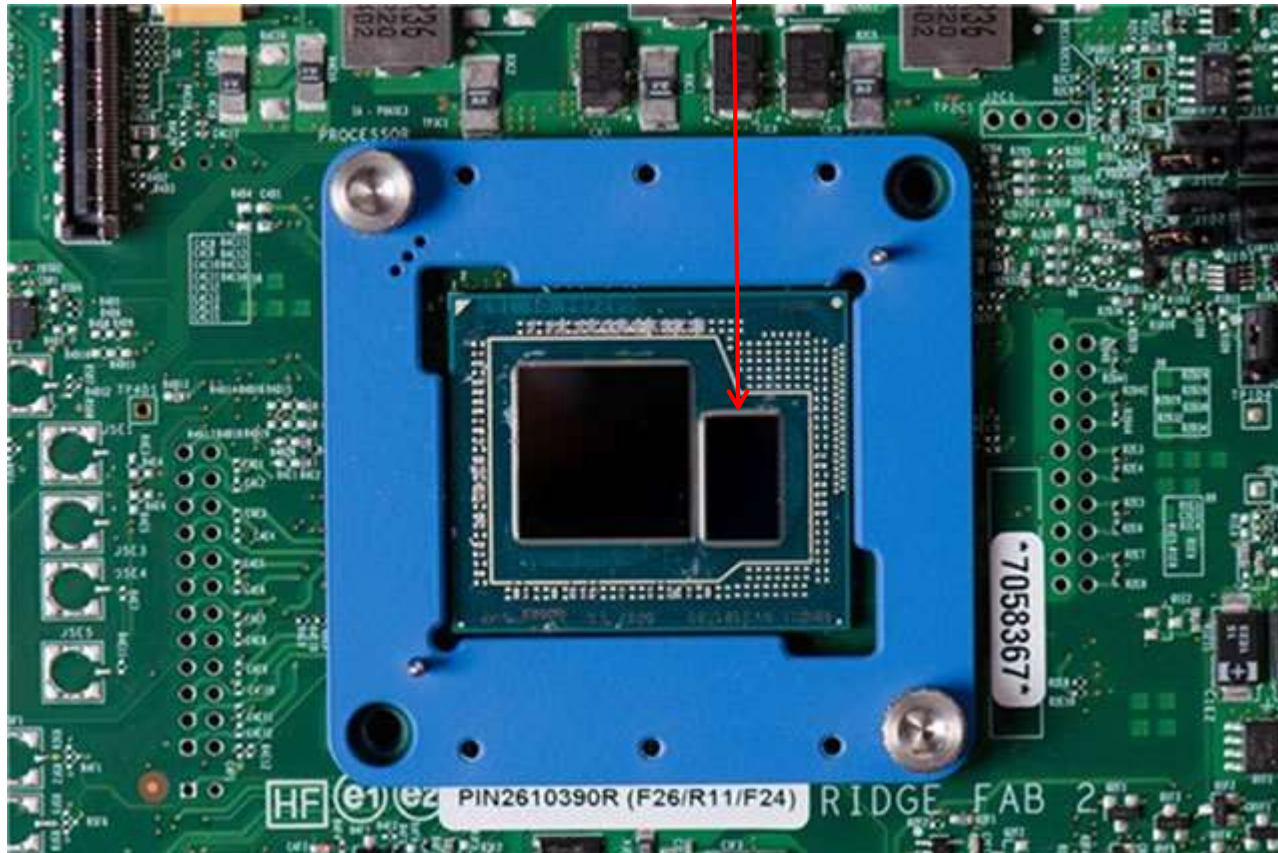
Principle of operation [117]

- The on package eDRAM, designated also as **Crystallwell**, it operates as a **true 4th level cache** of the memory hierarchy.
- It acts as a **victim buffer to the L3 cache**, in the sense that **anything evicted from the L3 cache immediately goes into the L4 cache**.
- **Both CPU and GPU requests are cached**.
- The **cache partitioning** between CPU and GPU **is dynamic**.
- If the GPU is not in use the whole L4 cache may be devoted the CPU, in this case the CPU has a **128 MB L4 cache**.
- **Access latency** after an L3 miss is **30 – 32 ns**.
- The L4 cache is capable of delivering **50 GB/s in each direction**.
- The Crystallwell die consumes between 0.5 and 1.0 W if idle and between 3.5 and 4.5 W under full load.
- The PCU (Power Control Unit) of the processor takes over the power management of the eDRAM, beyond the power management of the CPU cores, GPU, L3 cache etc.

5.3.1 On-package eDRAM cache (3)

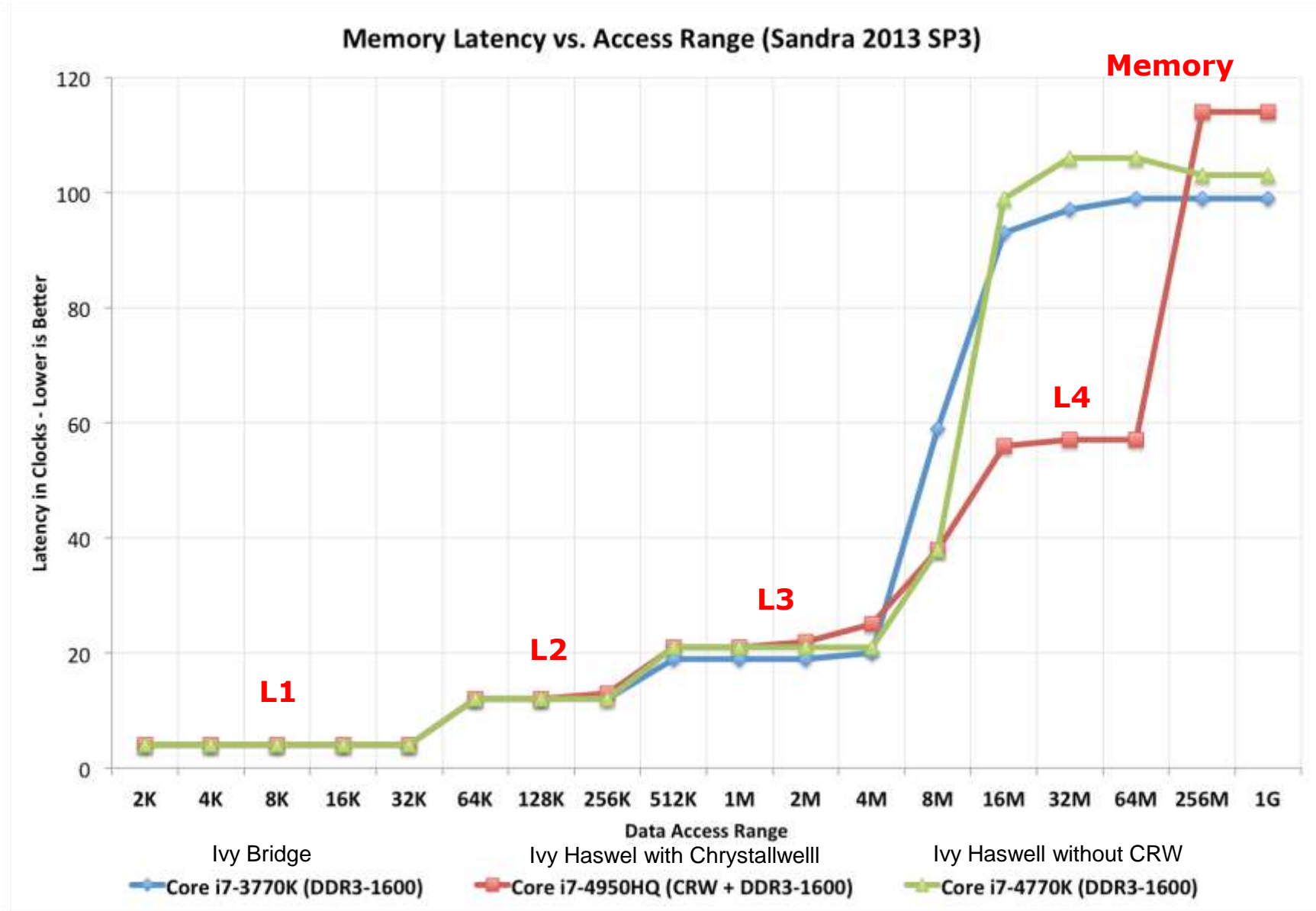
Implemented in-package eDRAM [124]

eDRAM chip



5.3.1 On-package eDRAM cache (4)

Memory latency vs. access range in a memory system with eDRAM cache (L4) [117]



5.3.2 FIVR (Fully Integrated Voltage Regulator) -1

Before introducing FIVR into the Haswell family, motherboards for Intel processors had to provide 6 different voltage regulators (VRs) to supply different voltages to the CPU cores, graphics (Gfx), System Agent (SA), IO, PLL and Memory, as indicated in the next Figure.

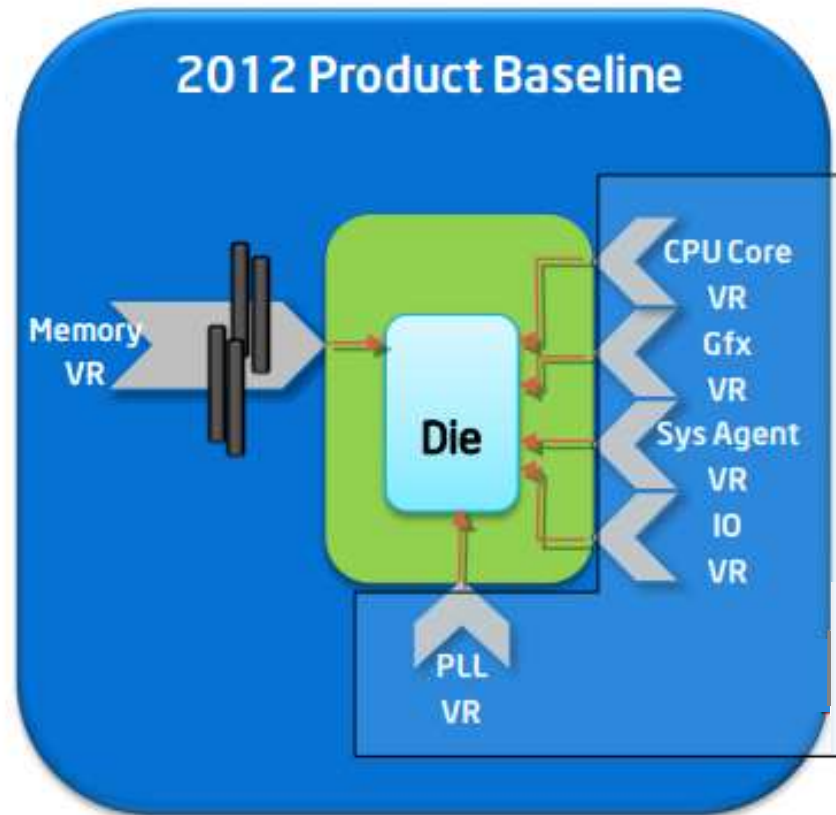


Figure: Voltage Regulators (VRs) needed in motherboards of Intel processors before introducing FIVRs [178]

5.3.2 FIVR (Fully Integrated Voltage Regulator) (2)

FIVR (Fully Integrated Voltage Regulator) -2

FIVR integrates legacy power delivery onto the package and the die, as shown below for intel's Haswell processor [178]

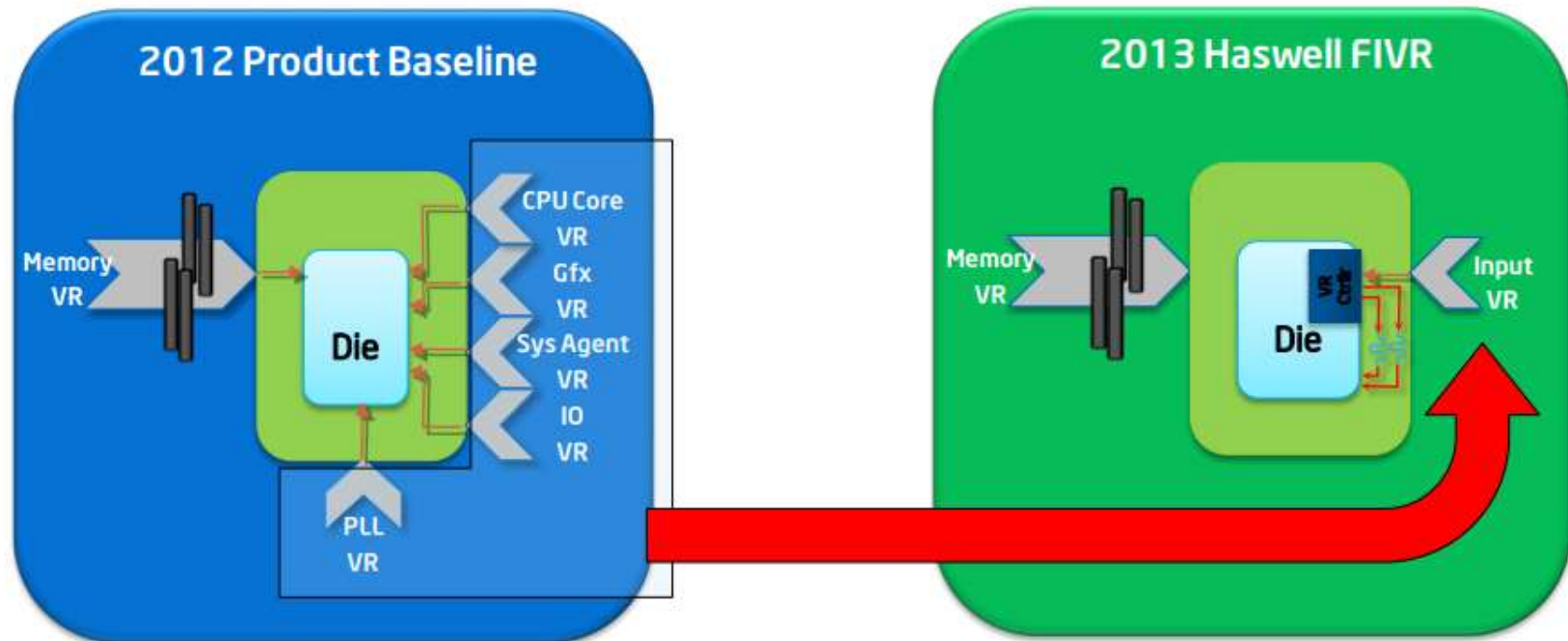
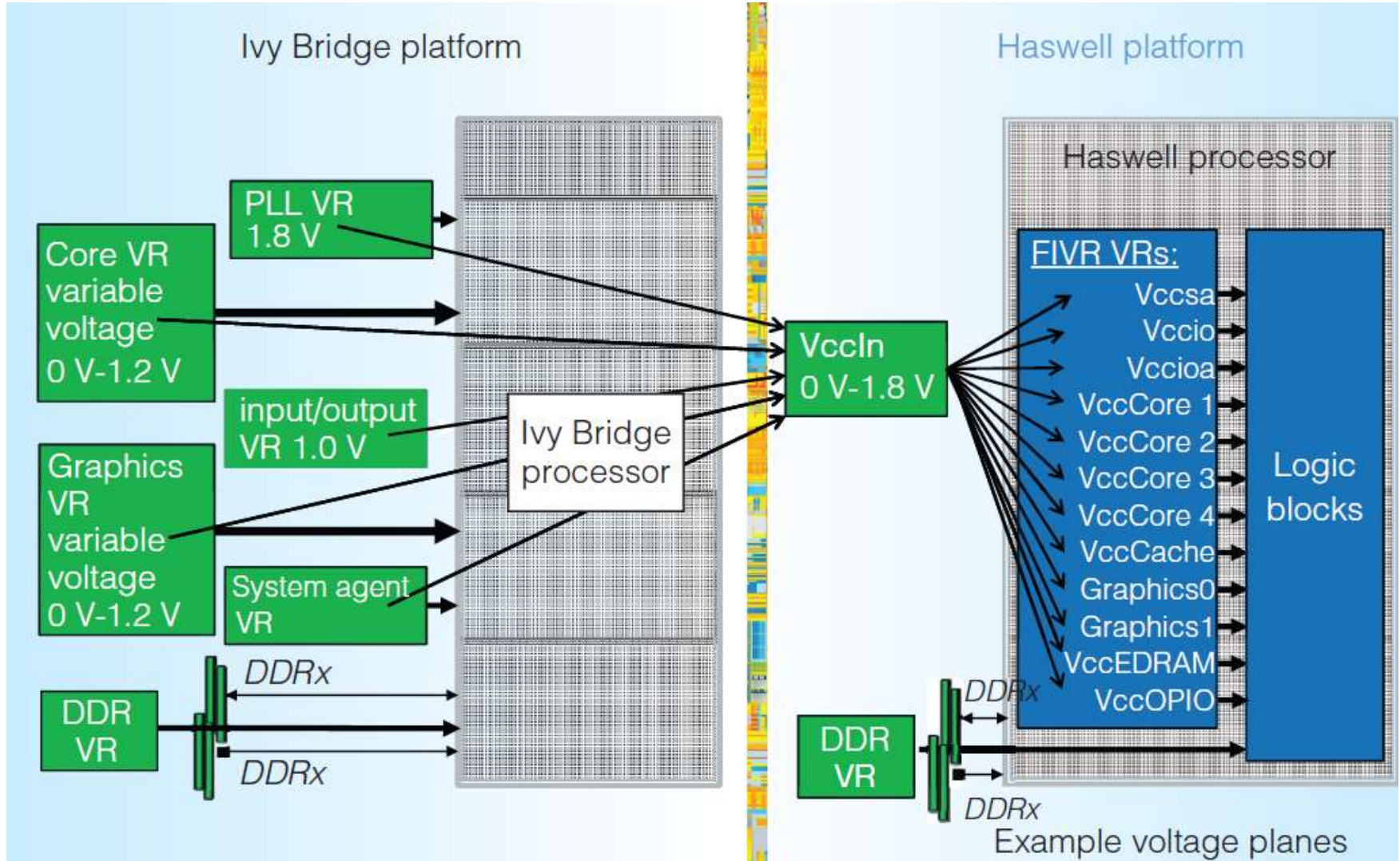


Figure: Integrating legacy power delivery onto the package and the die with FIVR [178]

This consolidates **five platform VRs into one** and thus **greatly simplifies mainboard design**.

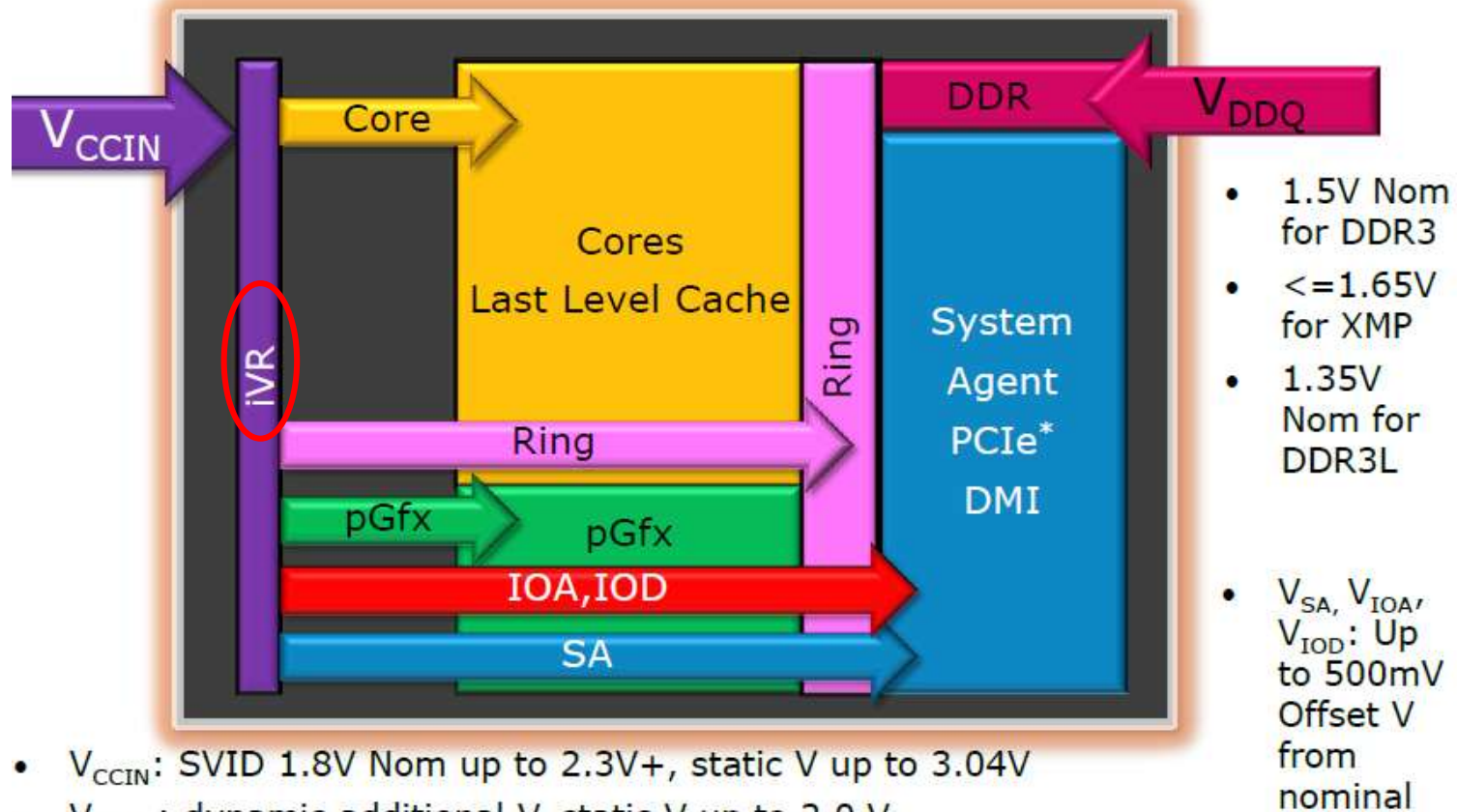
5.3.2 FIVR (Fully Integrated Voltage Regulator) (3)

Contrasting power delivery in Ivy Bridge and Haswell platforms [149]



5.3.2 FIVR (Fully Integrated Voltage Regulator) (4)

Implementation of the voltage planes in desktop and mobile Haswell processors [173]

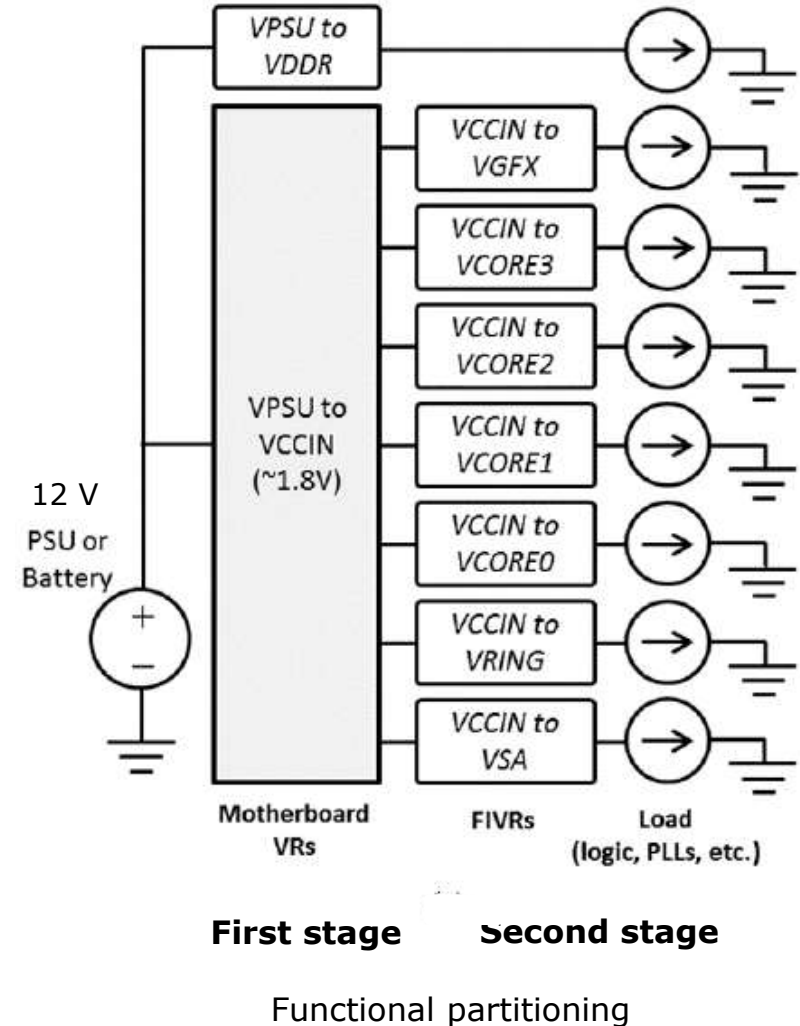


5.3.2 FIVR (Fully Integrated Voltage Regulator) (5)

Functional partitioning of Haswell's FIVR implementation [150]

FIVR is built up of two stages.

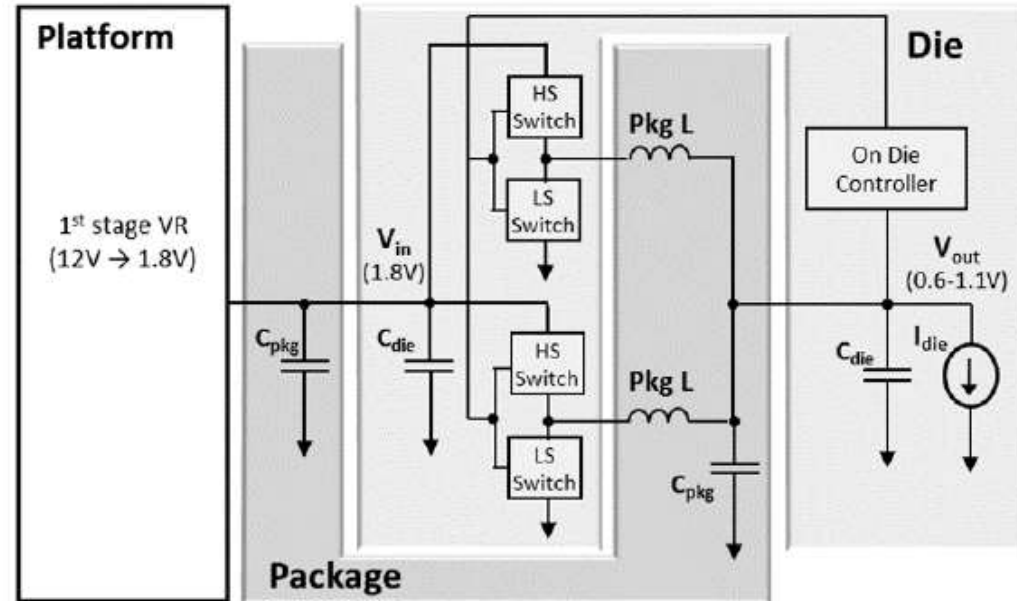
- The **first stage** of the voltage regulator converts from the PSU or battery voltage (12V) to approximately 1.8V, which is distributed across the microprocessor die.
- The **second conversion stage** is comprised of between 8 and 31 (depending on the product) FIVRs, which are 140MHz synchronous multiphase buck converters with up to 16 phases.



5.3.2 FIVR (Fully Integrated Voltage Regulator) (6)

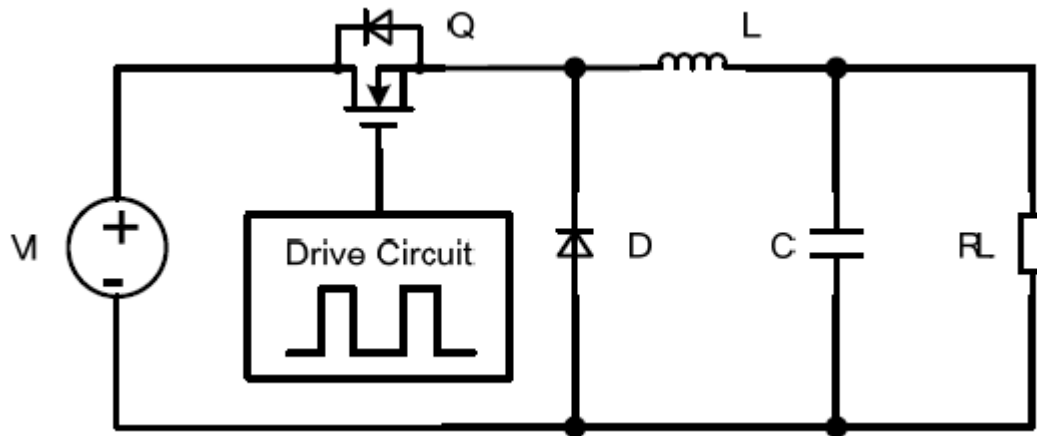
Partitioning of the implementation [150]

- The first stage is on the **motherboard**.
- The inductors and the mid-frequency decoupling capacitors are placed on the **package**.
- The power FETs, control circuitry and high frequency decoupling are **on the die**.
- Each FIVR is independently **programmable** to achieve optimal operation given the requirements of the domain it is powering.
- The settings are optimized by the **Power Control Unit (PCU)**, which specifies the input voltage, output voltage, number of operating phases, and a variety of other settings to minimize the total power consumption of the die.



5.3.2 FIVR (Fully Integrated Voltage Regulator) (7)

Principle of operation of the Buck converter [151]



Q : MOSFET
Drive circuit:
E.g. PWM modulated
(Pulse Width Modulated)

Figure: Block diagram of the Buck converter [151]

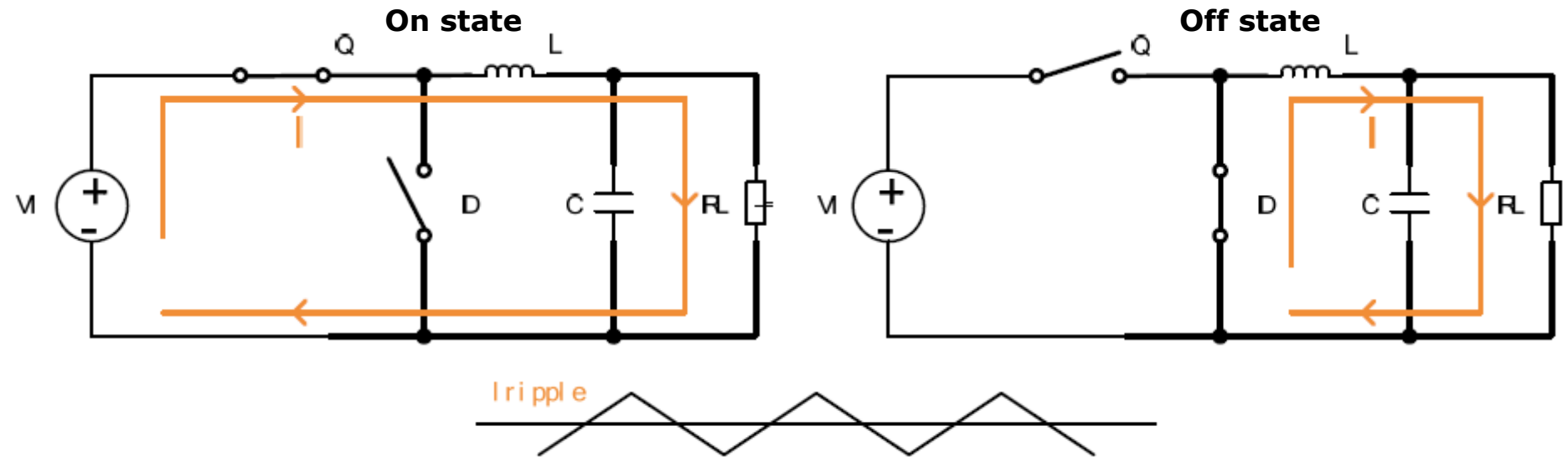


Figure: Operation of the Buck converter [151]

5.3.2 FIVR (Fully Integrated Voltage Regulator) (8)

Enhancing the buck converter to synchronous n -phase buck design to reduce ripple [152]

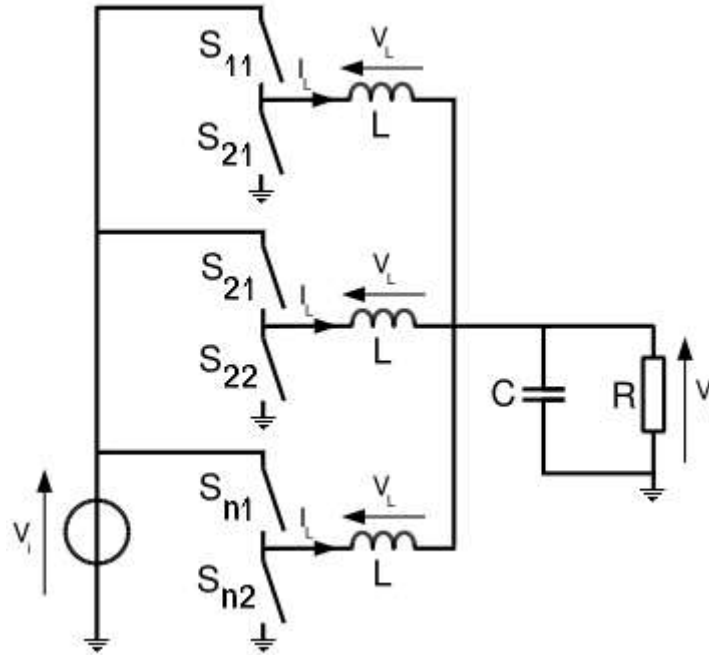
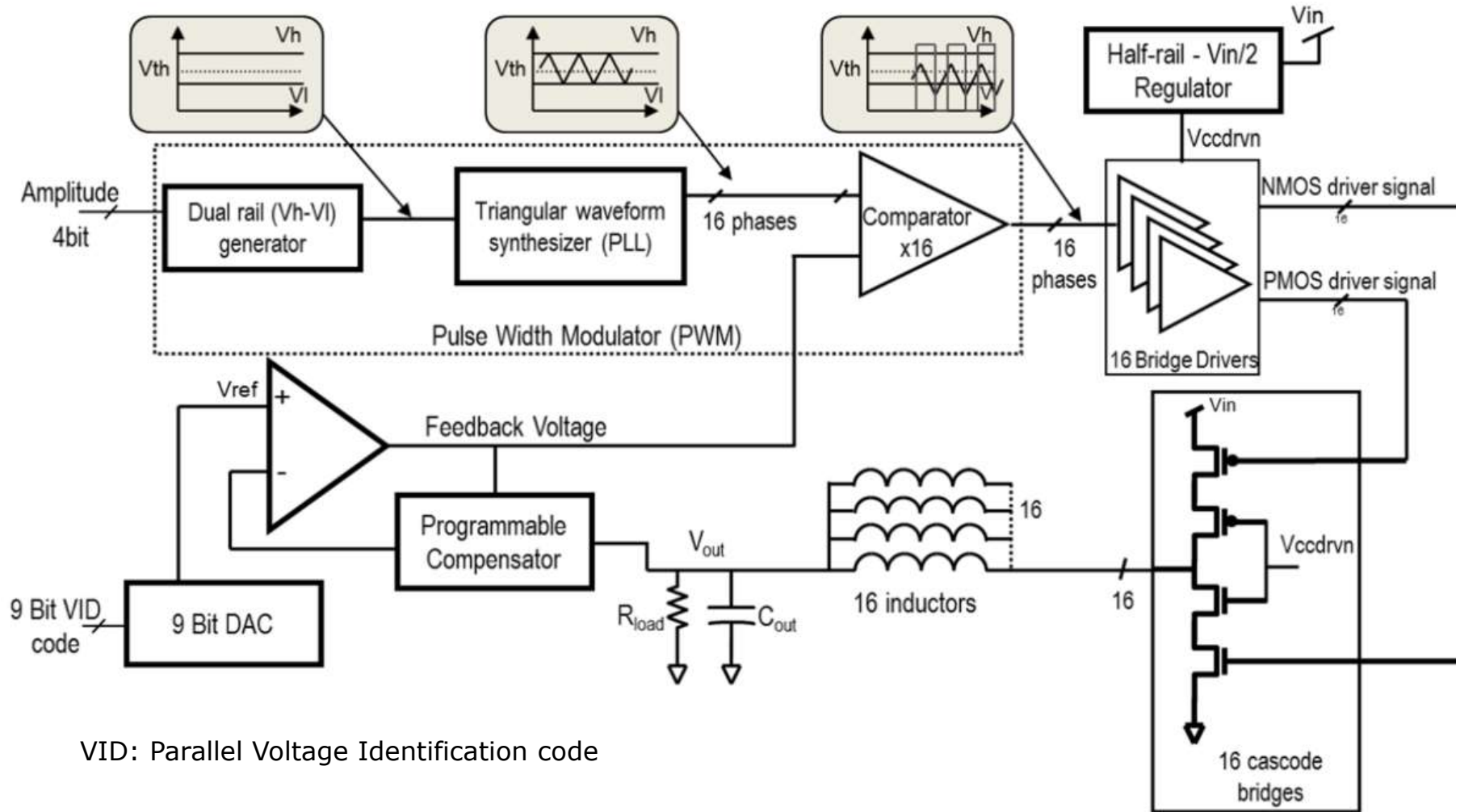


Fig. 9: Principle of a synchronous n -phase buck converter [152]

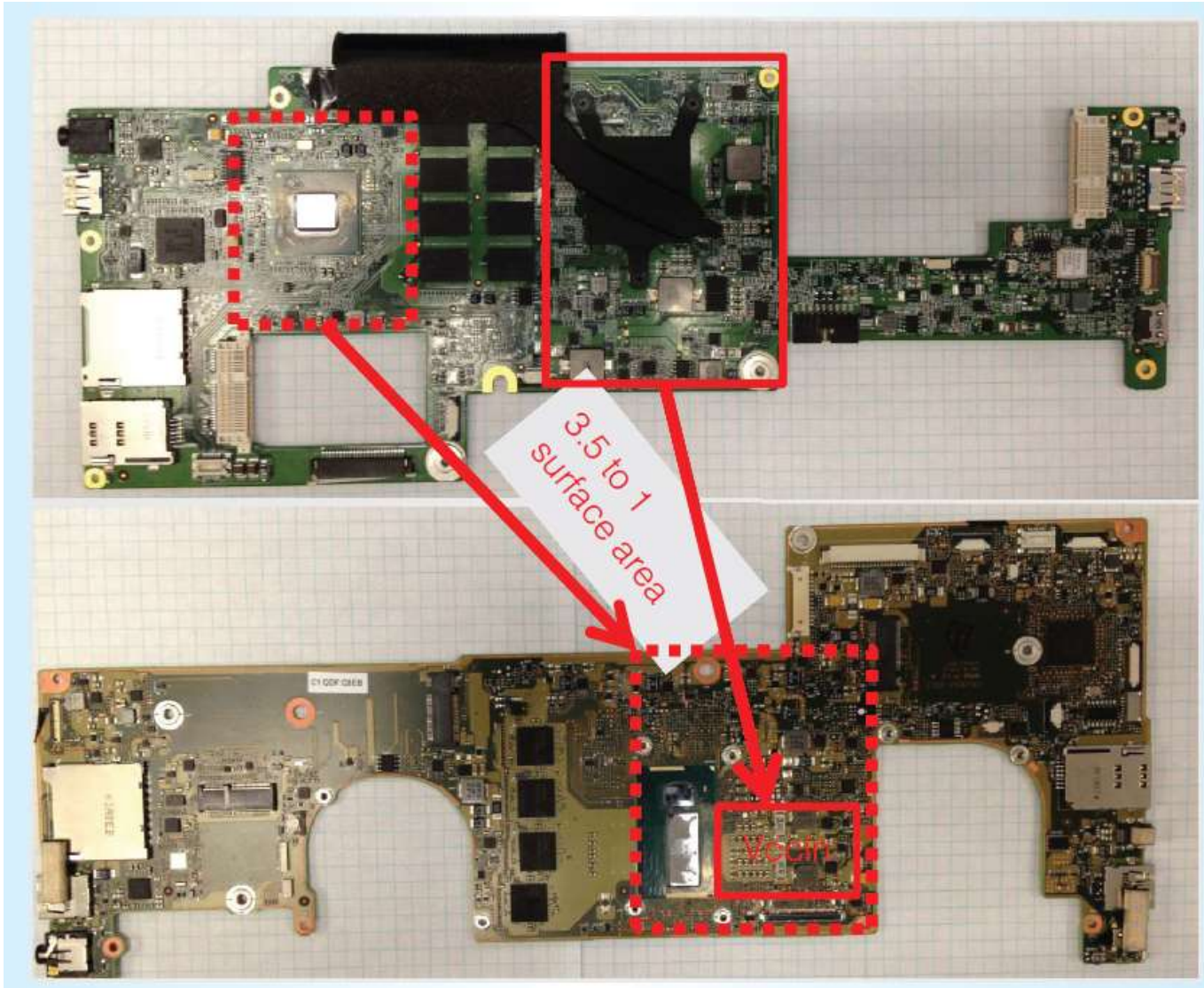
5.3.2 FIVR (Fully Integrated Voltage Regulator) (9)

Simplified block diagram of a single voltage plane in the FIVR domain of Haswell [150]



5.3.2 FIVR (Fully Integrated Voltage Regulator) (10)

Board space saving with Haswell's FIVR vs. Ivy Bridge's voltage regulator [149]



FIVR and per-core P-state control

- FIVR obviously provides an appropriate technique to deliver per-core core voltages and thus its use can greatly simplify the implementation of per-core P-state control.
- Despite this fact only the server and workstation oriented Haswell-EP lines (including the Xeon E5-1600 v3, the Xeon E5 2600 v3 and Xeon E5-4600 v3 processor lines) make use of this feature, as seen in the next slide for the Xeon E5-2600 v3, whereas mobile, desktop or Haswell-E oriented lines do not.

By contrast, all cores on Ivy Bridge and previous generations, run at the same frequency and are supplied by the same voltage.

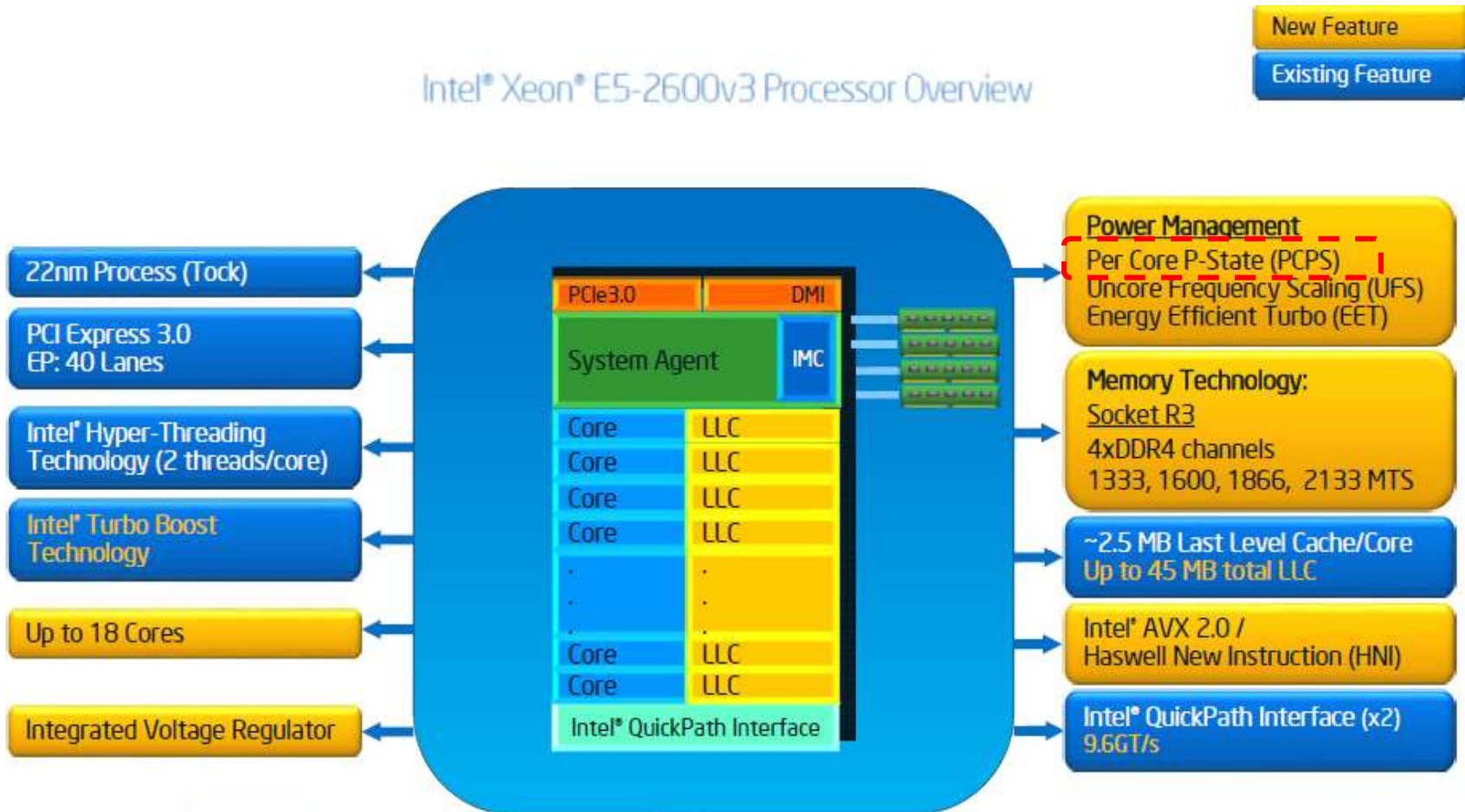
Available per-core PLLs are needed to be able to switch off PLLs of individual cores while PLLs of other cores are active.

We note that per-core P-state needs additional BIOS and OS support for scheduling the cores.

- Moreover, according to available documentation the high-end Haswell-EX server line do not implement FIVR at all presumably due to heat problems.

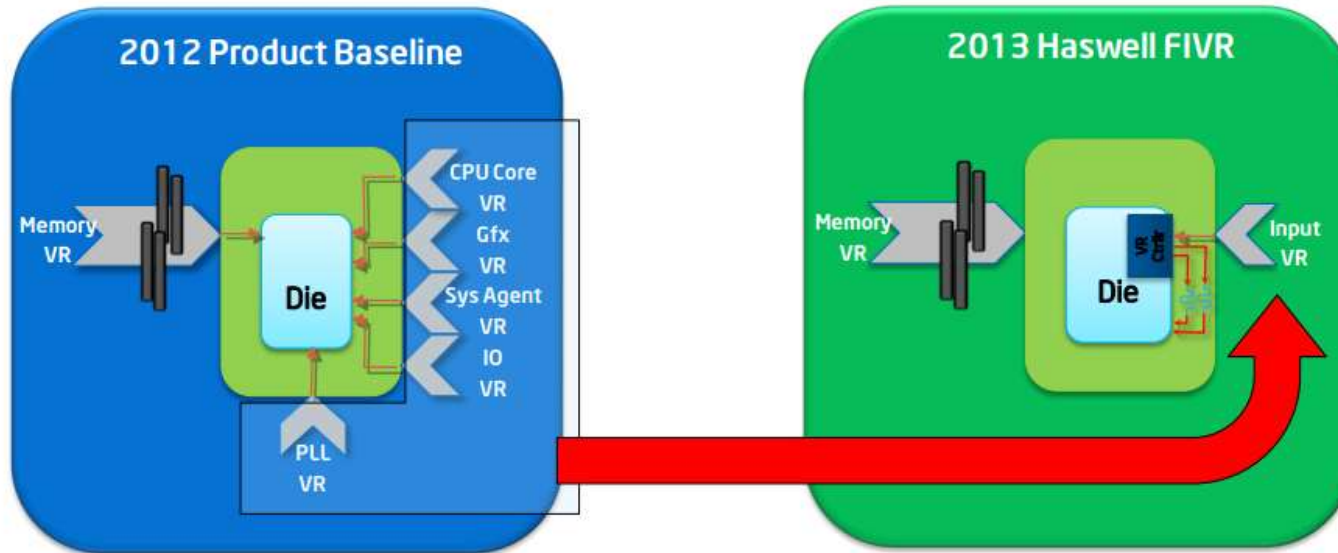
5.3.2 FIVR (Fully Integrated Voltage Regulator) (12)

Per Core P-State (PCPS) management in Intel's E5-2600 v3 (Haswell-EP) [172]



Assessment of FIVR - its benefits [178]

Fully Integrated Voltage Regulator (FIVR)



- FIVR integrates legacy power delivery onto processor pkg/die
- Greatly simplifies platform power design; consolidates 5 platform VRs down to just one
- Better arch flexibility and finer-grain on-die processor delivery control

FIVR Simplifies Platform Power Delivery Design

Assessment of FIVR - its drawbacks [178]

- FIVR clearly increases the power dissipation of the processor package and thus it has a limiting effect on the TDP and consequently on the performance.

5.3.2 FIVR (Fully Integrated Voltage Regulator) (15)

Remark

In their subsequent Broadwell family (9/2014) Intel introduced the 2nd generation FIVR with 3DL design [155]

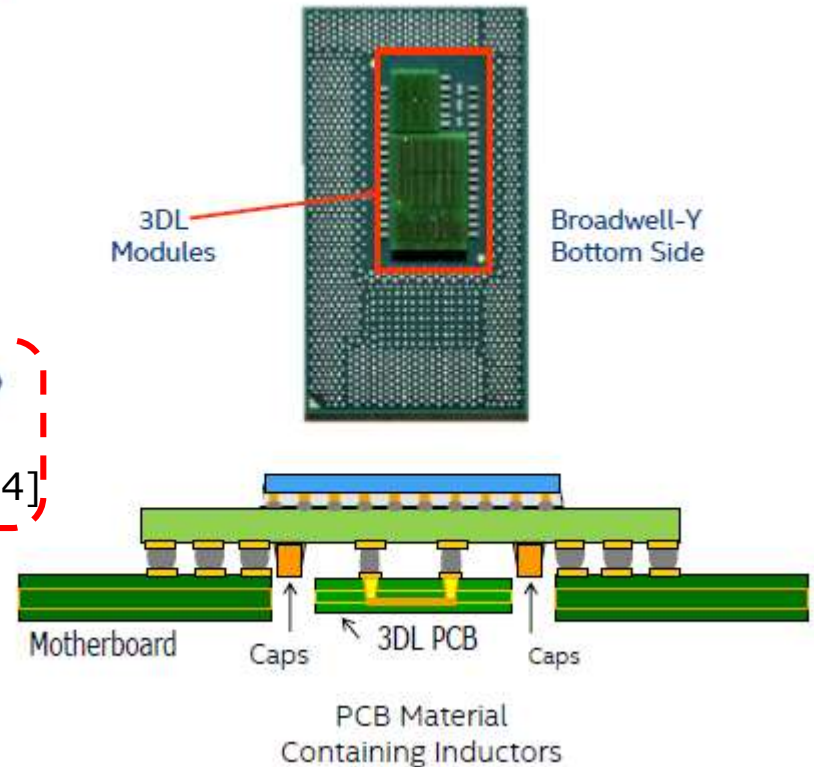
2nd Gen of FIVR enables better efficiency at lower voltages:

- Non-linear Droop Control
- Dual FIVR LVR Mode

3DL Modules:

- Inductors removed from package substrate to modules under the die. Better efficiency and package Z-height reduction of about 30 % [224]

3DL: 3D Layering
PCB: Printed Circuit Board
Caps: Capacitors
LVR: Linear Voltage Generator



The future of FIVR

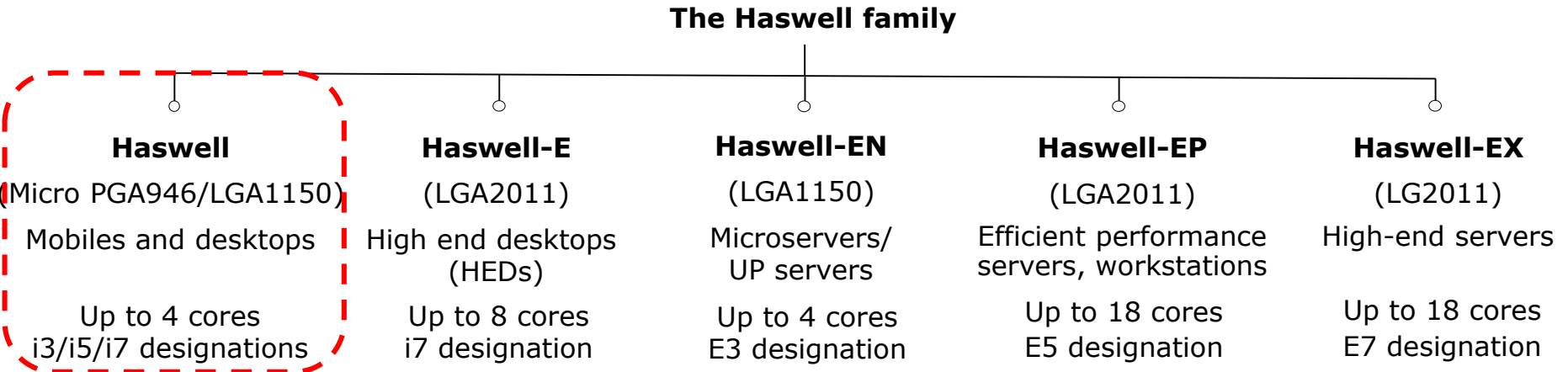
- Due to its drawbacks, after using FIVR in Broadwell, Intel decided to remove it in their upcoming Skylake and Kaby Lake processors.
Accordingly, related motherboards need again to provide appropriate power delivery.
- Nevertheless, sources state that in the subsequent 10 nm Cannon Lake processors, due to in 2017, Intel will anew make use of the FIVR technology probably as future FIVR designs will achieve higher efficiency and thus generate less heat [211].

5.3.3 TSX (Transactional Synchronization Extensions)

- Haswell also introduced a further new feature, called the **Transactional Synchronization Extensions (TSX)** that was debuted on selected Haswell models (SKUs)..
- **TSX supports Transactional Memory in hardware** (to be discussed later in the Chapter high end MP servers).
- Nevertheless, **in August 2014 Intel announced a bug** in the TSX implementation on all current steppings of all Haswell models and **disabled the TSX feature** on affected CPUs **via a microcode update**.
- Subsequently, TSX became **enabled first on a Broadwell model** (Core M-5Y70) **in 11/2014** then on the **Haswell-EX in 5/2015**.

5.4 Haswell based mobile and desktop processors

5.4 Haswell based mobile and desktop processors



Mobiles (SoCs)

Core i7-49xx/48xx/472x/471x/470x, 4C+G, HT, 6/2013 and 5/2014

Core i7-46xx/45xx, 2C+G, HT, 5/2013 and 6/2014

Core i5-43xx/42xx U/Y, 2C+G, HT, 6/2013 and 5/2014

Core i3-41xx/40xx, 2C+G, HT, 6/2013 and 5/2014

Desktops¹ (2 chips, z97 PCH)

Core i7-479x/478x/477x,476x, 4C+G, HT, 6/2013 and 5/2014

Core i5-46xx/45xx/44xx, 4C+G, HT, 6/2013 and 5/2014

Core i3-43xx/41xx, 2C+G, HT, 6/2013, 5/2014 and 3/2015

i7-5960X/5930K/5820K, 6/8 C, 8/2014

¹Additionally, in 6/2016 Intel introduced the Devil's Canon line, consisting of the i3-4360K, i5-4690K and i7-4790K models.

Interpretation of the model suffixes of the mobile and desktop lines [205]

Interpretation of the **mobile suffixes**

- M – mobile processor (Socket G3 (PGA 946))
- Q – quad-core
- U – ultra-low power (BGA 1168 packaging)
- X – "extreme"
- Y – extreme low-power (BGA 1168 packaging)
- E / H – BGA 1364 packaging

Interpretation of the **desktop suffixes**

- K – unlocked (adjustable CPU multiplier up to 63x)
- S – performance-optimized lifestyle (low power with 65 W TDP)
- T – power-optimized lifestyle (ultra low power with 35–45 W TDP)
- R – BGA packaging / High-performance GPU (currently Iris Pro 5200 (GT3e))
- X – extreme performance (adjustable CPU ratio with no ratio limit)

BGA packaging (Ball Grid Array): for soldering

LGA packaging (Land Grid Array): removable

PGA packaging (Pin Grid Array): removable

5.4 Haswell based mobile and desktop processors (3)

Main features of first introduced Haswell-based mobile Core i7 M-Series processors [125]

Intel 4th Gen Core i7 M-Series Mobile Processors					
Model	Core i7-4930MX	Core i7-4900MQ	Core i7-4800MQ	Core i7-4702MQ	Core i7-4700MQ
Cores/Threads	4/8	4/8	4/8	4/8	4/8
CPU Base Freq	3.0	2.8	2.7	2.2	2.4
Max SC Turbo	3.9	3.8	3.7	3.2	3.4
Max DC Turbo	3.8	3.7	3.6	3.1	3.3
Max QC Turbo	3.7	3.6	3.5	2.9	3.2
TDP	57W	47W	47W	37W	47W
HD Graphics	4600	4600	4600	4600	4600
GPU Clock	400-1350	400-1300	400-1300	400-1150	400-1150
L3 Cache	8MB	8MB	6MB	6MB	6MB
DDR3/DDR3L	1600	1600	1600	1600	1600
vPro/TXT/VT-d	Yes	Yes	Yes	No	No
Intel SBA	No	No	No	Yes	Yes
Price	\$1096	\$568	\$378		

SBA: Small Business Advantage (Enhanced security and productivity)

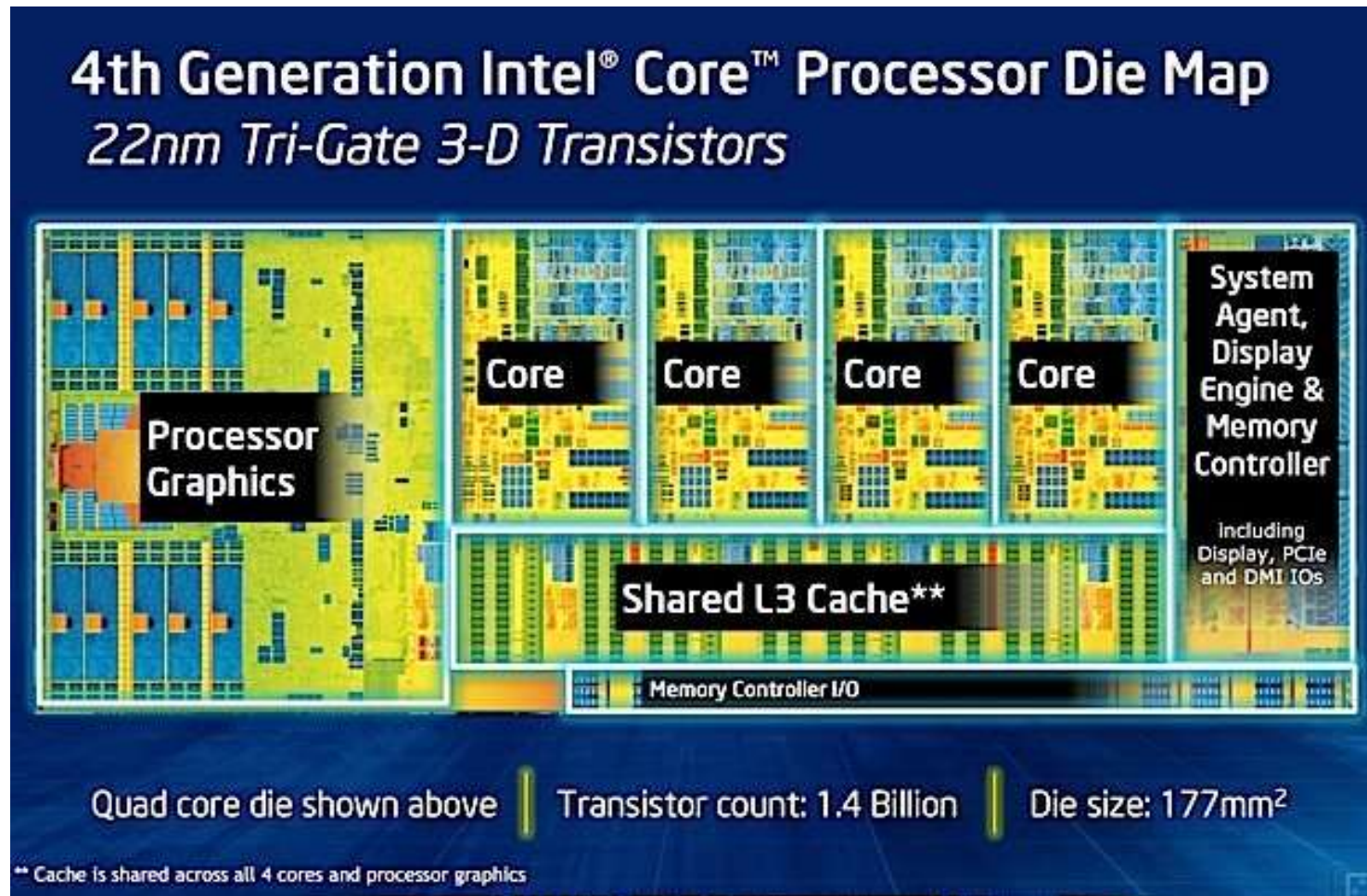
5.4 Haswell based mobile and desktop processors (4)

Main features of first issued Haswell-based Core i7 desktop processors [126]

Intel 4 th Gen Core i7 Desktop Processors						
Model	Core i7-4770K	Core i7-4770	Core i7-4770S	Core i7-4770T	Core i7-4770R	Core i7-4765T
Cores/Threads	4/8	4/8	4/8	4/8	4/8	4/8
CPU Base Freq	3.5	3.4	3.1	2.5	3.2	2.0
Max Turbo	3.9 (Unlocked)	3.9	3.9	3.7	3.9	3.0
Test TDP	84W	84W	65W	45W	65W	35W
HD Graphics	4600	4600	4600	4600	Iris Pro 5200	4600
GPU Max Clock	1250	1200	1200	1200	1300	1200
L3 Cache	8MB	8MB	8MB	8MB	6MB	8MB
DDR3 Support	1333/1600	1333/1600	1333/1600	1333/1600	1333/1600	1333/1600
vPro/TXT/VT-d/SIPP	No	Yes	Yes	Yes	No	Yes
Package	LGA-1150	LGA-1150	LGA-1150	LGA-1150	BGA	LGA-1150
Price	\$339	\$303	\$303	\$303		\$303

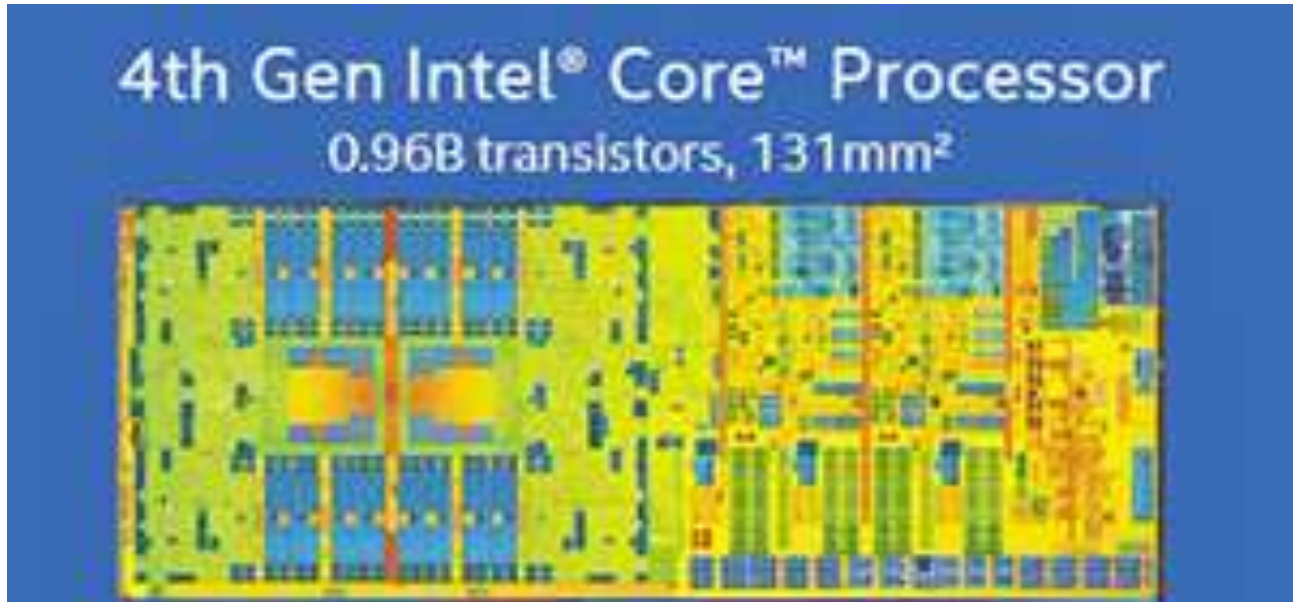
5.4 Haswell based mobile and desktop processors (5)

Die plot of a quad core Haswell processor [121]



5.4 Haswell based mobile and desktop processors (6)

Die plot of a dual core Haswell processor [177]



In package integrated CPU and PCH for mobile processors [204]

New BGA Strategy for Ultra-Thin Devices

Integration Drives Lower Power, Smaller Designs

New! 1-Chip BGA Solution



- CPU and PCH integrated into single BGA package
- 15W & 28W TDPs, 6W and below SDP
- S0ix support
- Supports LPDDR3 and DDR3L memory

BGA: Ball Grid Array

SDP: Scenario Design Point

Traditional 2-Chip platform

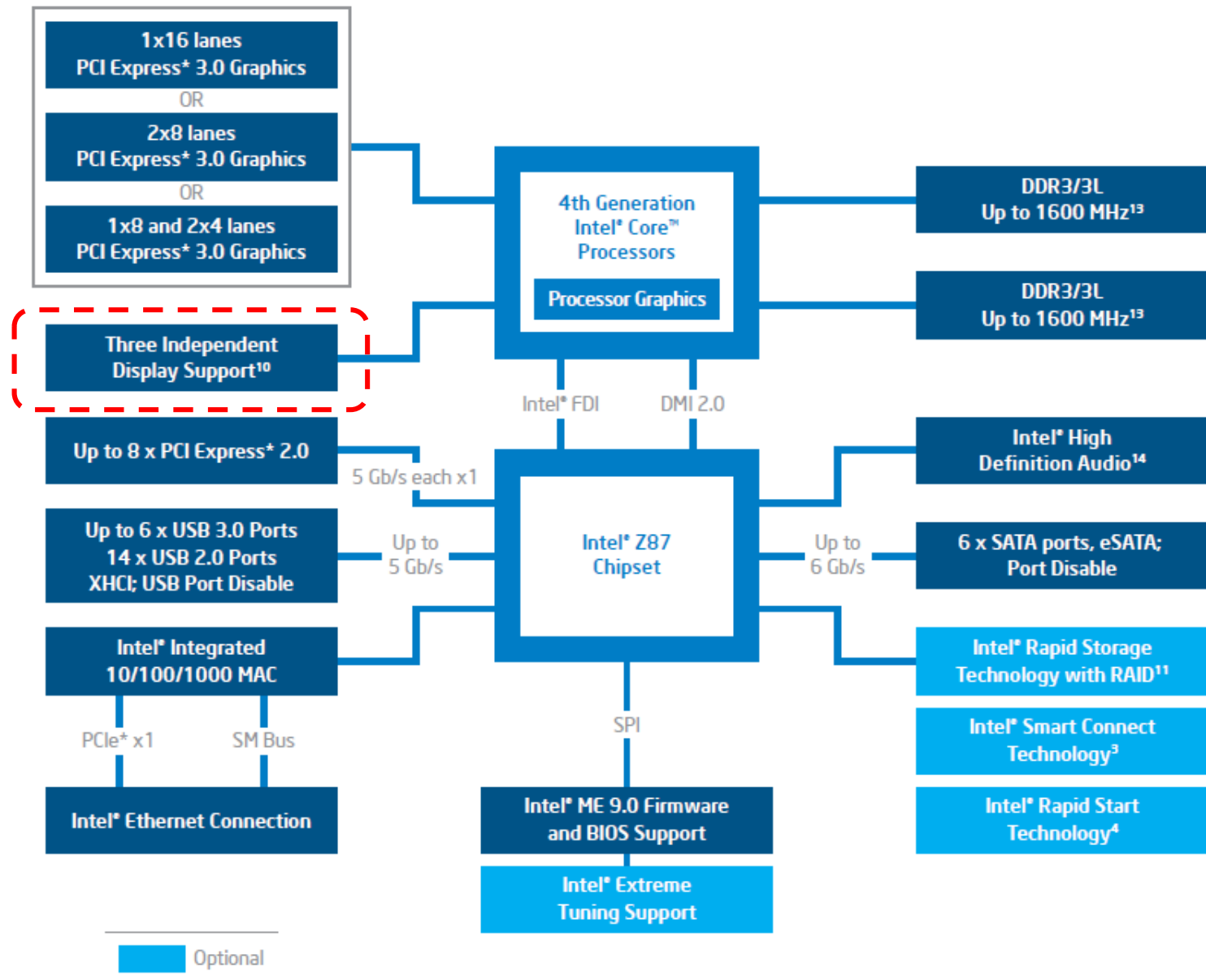


- 2 chip scalable solution: CPU and chipset
- BGA and rPGA packages
- 57W, 47W, and 37W TDPs
- Supports DDR3L Memory
- GT3e graphics

PGA: Pin Grid array

5.4 Haswell based mobile and desktop processors (8)

Example of a Haswell-based desktop platform [145]



Innovative feature of certain 8. gen. DT PCHs of Haswell processors – use of Flexible I/O (Flex I/O) lanes, called also High-Speed I/O (HSIO) lanes

- Beginning with certain 8th gen. DT PCHs (supporting the Haswell line of processors (Q87/Z87/H87) Intel introduced a new I/O bus architecture that is based on serial, differential, high-speed Flex I/O lanes called also as High-Speed I/O lanes (HSIO lanes).
- The advantage of Flex I/O lanes is that they can be configured as PCIe*, USB 3.0 or SATA lanes in accordance with the I/O needs of a platform.
- As an example the next Figure shows the mapping of the Flex I/O lanes of 8th generation PCHs to PCIe*, USB 3.0 or SATA lanes.

5.4 Haswell based mobile and desktop processors (10)

Example: Mapping of the Flex I/O lanes (ports) to PCIe*, USB 3.0 or SATA lanes in the 8 Series PCH family -1 [251]

SKU	High Speed I/O Ports																	
	Port 1	Port 2	Port 3	Port 4	Port 5	Port 6	Port 7	Port 8	Port 9	Port 10	Port 11	Port 12	Port 13	Port 14	Port 15	Port 16	Port 17	Port 18
Q87	USB 3.0 Port 1	USB 3.0 Port 2	USB 3.0 Port 5	USB 3.0 Port 6	USB 3.0 Port 3	USB 3.0 Port 4	PCIe* Port 3	PCIe* Port 4	PCIe* Port 5	PCIe* Port 6	PCIe* Port 7	PCIe* Port 8	SATA 6Gb/s Port 4	SATA 6Gb/s Port 5	SATA 6Gb/s Port 0	SATA 6Gb/s Port 1	SATA 6Gb/s Port 2	SATA 6Gb/s Port 3
					PCIe* Port 1	PCIe* Port 2							PCIe* Port 1	PCIe* Port 2				
Q85	USB 3.0 Port 1	USB 3.0 Port 2	USB 3.0 Port 5	USB 3.0 Port 6	PCIe* Port 1	PCIe* Port 2	PCIe* Port 3	PCIe* Port 4	PCIe* Port 5	PCIe* Port 6	PCIe* Port 7	PCIe* Port 8	SATA 3Gb/s Port 4	SATA 3Gb/s Port 5	SATA 6Gb/s Port 0	SATA 6Gb/s Port 1	SATA 6Gb/s Port 2	SATA 6Gb/s Port 3
B85	USB 3.0 Port 1	USB 3.0 Port 2	USB 3.0 Port 5	USB 3.0 Port 6	PCIe* Port 1	PCIe* Port 2	PCIe* Port 3	PCIe* Port 4	PCIe* Port 5	PCIe* Port 6	PCIe* Port 7	PCIe* Port 8	SATA 3Gb/s Port 4	SATA 3Gb/s Port 5	SATA 6Gb/s Port 0	SATA 6Gb/s Port 1	SATA 6Gb/s Port 2	SATA 6Gb/s Port 3
Z87	USB 3.0 Port 1	USB 3.0 Port 2	USB 3.0 Port 5	USB 3.0 Port 6	USB 3.0 Port 3	USB 3.0 Port 4	PCIe* Port 3	PCIe* Port 4	PCIe* Port 5	PCIe* Port 6	PCIe* Port 7	PCIe* Port 8	SATA 6Gb/s Port 4	SATA 6Gb/s Port 5	SATA 6Gb/s Port 0	SATA 6Gb/s Port 1	SATA 6Gb/s Port 2	SATA 6Gb/s Port 3
					PCIe* Port 1	PCIe* Port 2							PCIe* Port 1	PCIe* Port 2				
H87	USB 3.0 Port 1	USB 3.0 Port 2	USB 3.0 Port 5	USB 3.0 Port 6	USB 3.0 Port 3	USB 3.0 Port 4	PCIe* Port 3	PCIe* Port 4	PCIe* Port 5	PCIe* Port 6	PCIe* Port 7	PCIe* Port 8	SATA 6Gb/s Port 4	SATA 6Gb/s Port 5	SATA 6Gb/s Port 0	SATA 6Gb/s Port 1	SATA 6Gb/s Port 2	SATA 6Gb/s Port 3
					PCIe* Port 1	PCIe* Port 2							PCIe* Port 1	PCIe* Port 2				
H81	USB 3.0 Port 1	USB 3.0 Port 2	NA	NA	PCIe* Port 1	PCIe* Port 2	PCIe* Port 3	PCIe* Port 4	PCIe* Port 5	PCIe* Port 6	NA	NA	SATA 3Gb/s Port 4	SATA 3Gb/s Port 5	SATA 6Gb/s Port 0	SATA 6Gb/s Port 1	NA	NA

Notes:

- Ports listed with NA are not available and are disabled.

Example: Mapping of the Flex I/O lanes (ports) to PCIe*, USB 3.0 or SATA lanes in the 8 Series PCH family -2 [251]

- As the above Figure shows 8th gen. PCHs provide 18 differential lanes that are split between PCIe*, USB 3.0 or SATA lanes.
- Three of the available PCHs (Q87/B87/H87) can be configured as follows the others not.
- As seen, 4 differential pairs are multiplexed:
 - 2 multiplexed differential pairs can be configured to be used as PCIe port 1, 2 or USB3 port 3, 4, and
 - the other 2 differential pairs can be set to serve as PCIe port 1, 1y or SATA port 4, 5.

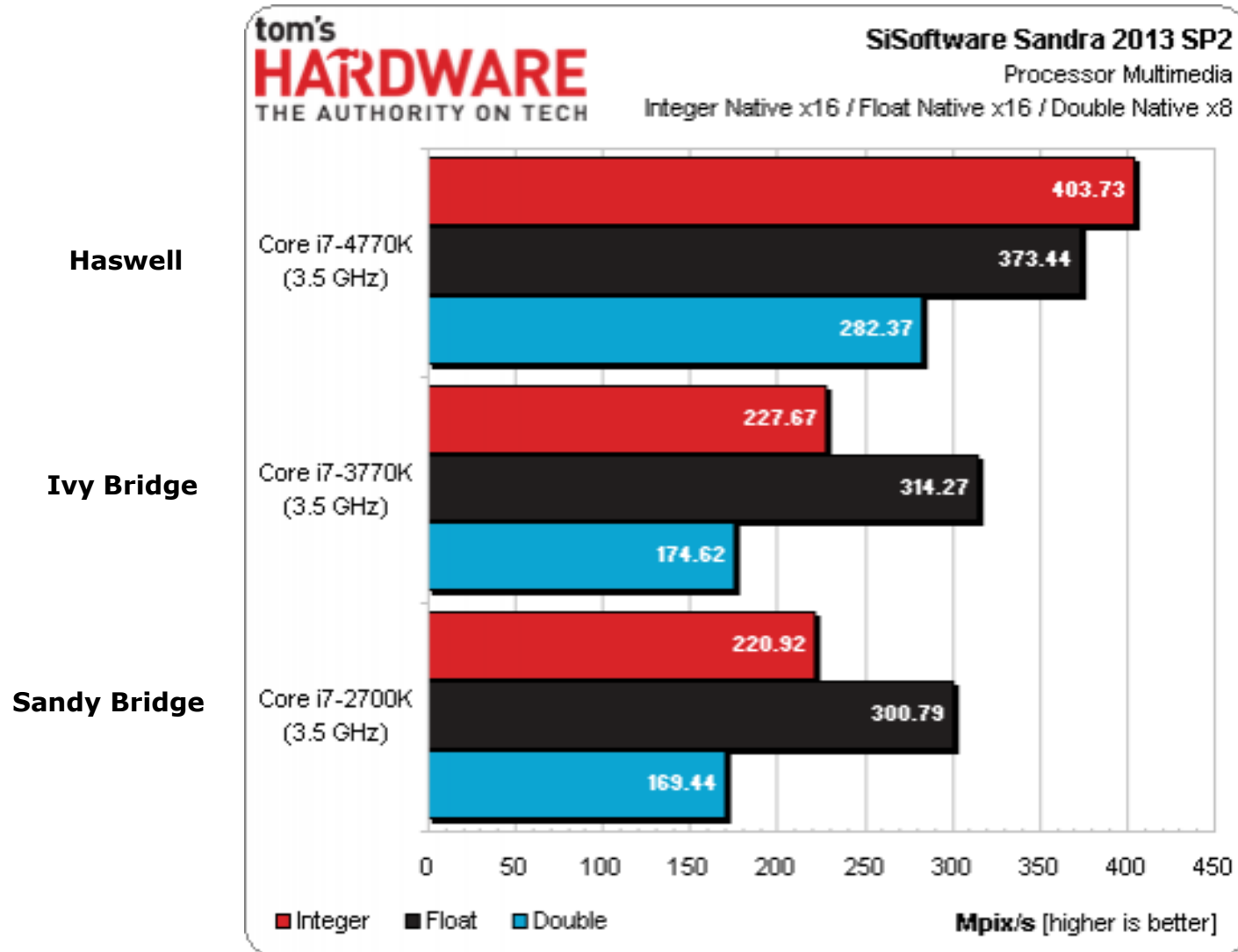
The Flexible I/O is configured through soft straps, i.e. by setting configuration parameters suitably in the system software.

Remark

Intel's Flex I/O is presumably implemented by means of embedded FPGAs as a paper of a former Intel design engineer indicates it [252].

5.4 Haswell based mobile and desktop processors (13)

Integer and FP performance of subsequent generations of desktop processors [127]



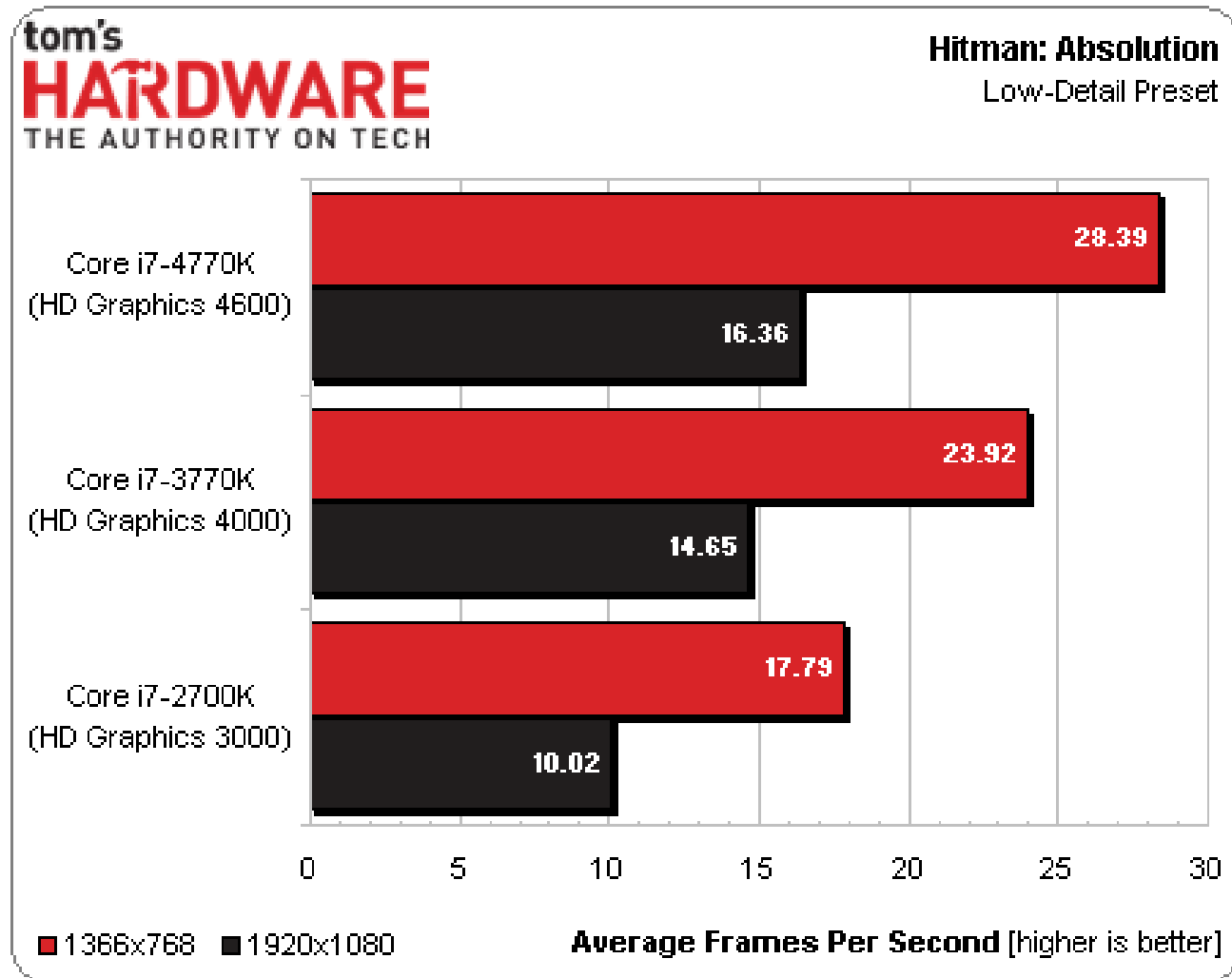
5.4 Haswell based mobile and desktop processors (14)

Note

As the AVX2 extension of the Haswell processors introduces [256-bit FX operations](#), this line has a [considerable higher integer performance](#) than previous Core generations.

5.4 Haswell based mobile and desktop processors (15)

Graphics performance of subsequent generations of desktop processors [127]



5.4 Haswell based mobile and desktop processors (17)

Main features of the Devil's Canon DT models [226]

Intel Core i5 and i7						
	i5-3570K	i5-4670K	i5-4690K	i7-3770K	i7-4770K	i7-4790K
Core Name	Ivy Bridge	Haswell	Devil's Canyon	Ivy Bridge	Haswell	Devil's Canyon
Release	April 2012	June 2013	June 2014	April 2012	June 2013	June 2014
Socket	1155	1150	1150	1155	1150	1150
CPU Frequency	3400		3500	3500		4000
CPU Turbo	3800		3900	3900		4400
L3 Cache	6 MB			8 MB		
TDP	77 W	84 W	88 W	77 W	84 W	88 W
Threads	4			8		
AVX 2.0	No	Yes		No	Yes	
FMA3	No	Yes		No	Yes	
TSX-NI	No		Yes	No		Yes
VT-d	No		Yes	No		Yes
IGP	HD 4000	HD 4600		HD 4000	HD 4600	
IGP Frequency	650	350		650	350	
IGP Turbo	1150	1200		1150	1250	
Memory	DDR3-1600					
Price	\$230	\$235	\$240	\$330	\$335	\$340

Enhancements of the Devil's Canon models -1 [226]

- [New Thermal Interface Material](#), made with a Next Generation Polymer (NGPTIM)
- [Additional capacitors](#) on the underside for smoother power delivery,
as indicated in the next Figure.

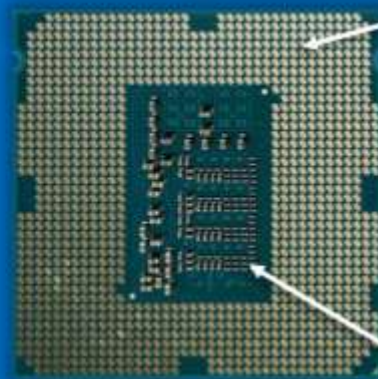
Enhancements of the Devil's Canon models -2 [226]



Enhanced Packaging and Thermals for New Unlocked 4th Gen Intel® Core™ Processors



Original



New

1150 pins in LGA package

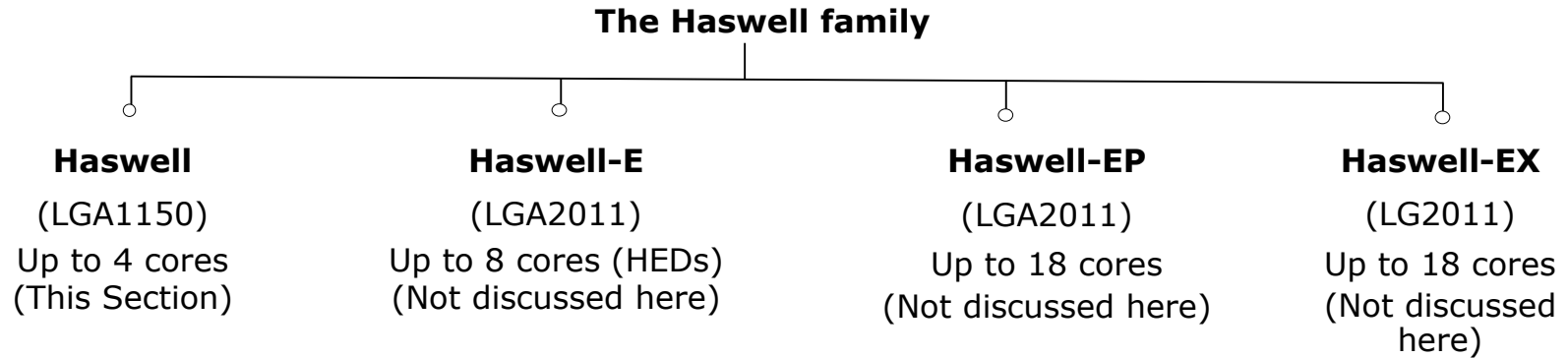
NEW! Next-Generation Polymer Thermal Interface Material (NGPTIM) that works with existing cooling solutions improvements enabling cooler & higher performing CPUs

NEW! Additional capacitors to smooth power delivery to die

Enhancements enable additional performance and headroom

5.5 Haswell based server processors

Overview of the Haswell server family



Servers

Microservers

E3-1275L/1265L v3, 4C+G, HT, 6/2013 and 5/2014
E3-1240L/1230L/1220L v3, 2C/4C, HT, 6/2013 and 5/2014

UP Servers

E3-12x5/12x6 v3, 4C+G, HT, 6/2013 and 5/2014
E3-12x0/12x1 v3, 4C, HT, 6/2013 and 5/2014

Workstations

E5-16xx v3, 4/6/8, 9/2014

DP-Servers

E5-26xx v3, 4/6/8/10/12/14/16/18C, 9/2014

MP-Servers

E5-46xx v3, 6/10/12/14/16/18C, 6/2015

E7-48xx v3, 8/10/12/14, 5/2015
E7-88xx v3, 4/10/16/18C, 5/2015

Note

As far as **Haswell-based processors** are considered

- **mobiles, desktops, microservers and UP servers** include **two or four cores**,
- **high-end desktops (E-series)** **6 to 8 cores** whereas
- **workstations, DP and MP servers** typically more, **up to 18 cores**,
as indicated in the above overviews.