Lecture 26: 80x86 Evolution 2

Seyed-Hosein Attarzadeh-Niaki

Slides from Dezső Sima

Review

- 80186, 80286, and 80386 CPUs
- Real and Protected modes
- Segmented memory
- Memory paging
- Flat memory

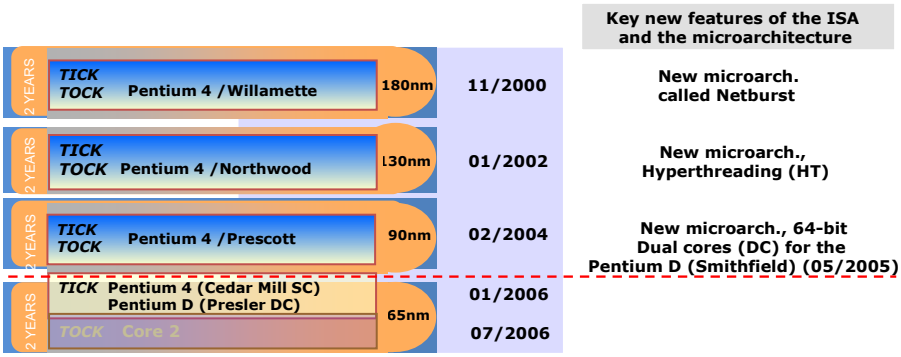
Outline

- Overview of the evolution of Intel's Pentium 4 and Core 2 families
- Evolution of desktop and laptop processors
- Evolution of HEDs (High-End desktops)
- Evolution of high-end 4S/8S servers

1.1 Overview of the evolution of Intel's Pentium 4 and Core 2 families

The Core 2 family was preceded by the Pentium 4 family

The Pentium 4 family introduced important innovations, as listed below.



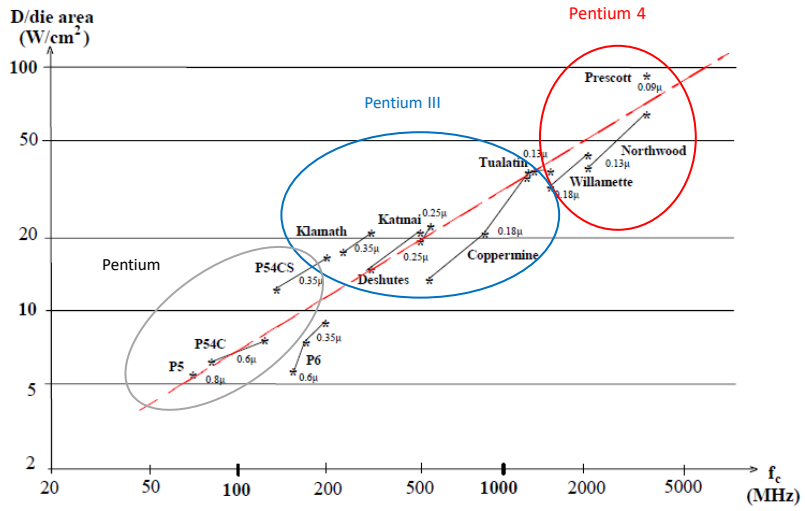
Note

This is a single phase development model, in each generation both technology and microarchitecture is changed.

Design target of the Pentium 4 family

At Pentium 4's launch (Nov. 2000) Intel's vice president (Otellini) claimed the lifespan of the Netburst microarchitecture to be 7 years and expected its clock frequency to break the 10 GHz mark in 2006.

Relative dissipation of Intel's x86 family of processors



Intel's cancellation of 4 GHz Pentium 4 devices and subsequently the Pentium 4 line

- In Oct. 2004 Intel's CEO (Chief Executing Officer) admitted that the Pentium 4 family would not achieve 4 GHz.

Figure: In Oct. 2004 Crag Barrett, Intel's then-CEO on his knees to apologize for not achieving the 4GHz mark in front of an audience of 7000 informatics professional at the IT Expo in Orlando

1.1 Overview of the evolution of Intel's Pentium 4 and Core 2 families

Changing Intel's design paradigm to cope with raising dissipation about 2003

In 2003 Intel shifted the focus of their processor development from the pure performance goal to the aspect of performance per watt, as stated in a slide from 4/2006, see below.

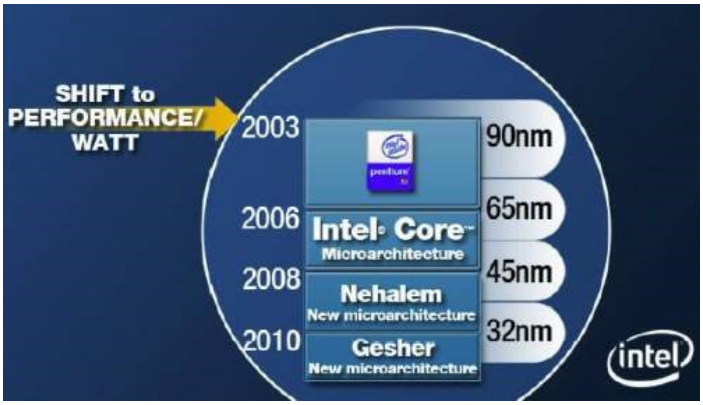


Figure 1.3: Intel's plan to develop their manufacturing technology and processor lines revealed at a shareholder's meeting back in 4/2006

1.1 Overview of the evolution of Intel's Pentium 4 and Core 2 families

Remark: A further change of the design paradigm for designing the mobile processors

With the advent of mobile devices (about 2006) a new design paradigm arose for those devices.

Mobile devices require long operating hours i.e. low power consumption, this is contrast to the design paradigm of traditional processors, as indicated below.

Traditional processors		Tablets and smartphones
High performance/power (e.g. GFLOPS/Watt)	⟷	Low power (Watt) (Number of operating hours)

Introducing a two-phase development model (Tick-Tock model) for the Core 2 family

The two-phase model reduces the complexity of the development, as

- the Tick phase focuses on the reduction of the feature size whereas
- the Tock phase focuses on enhancing the microarchitecture.

Basic designations of Intel's processor lines

M/DT	1. gen. 4/5/6xxx	6/7/8/9xxx	i5/i7-xxx	i3/i5/i7-xxx	2. gen. 2xxx	3. gen. 3xxx	4. gen. 4xxx	5. gen. 5xxx
	Core 2 New Microarch. 65 nm	Penryn New Process 45 nm	Nehalem New Microarch. 45 nm	West-mere New Process 32 nm	Sandy Bridge New Microarch. 32 nm	Ivy Bridge New Process 22 nm	Haswell New Microarchi. 22 nm	Broad-well New Process 14 nm
	TOCK	*TICK*	*TOCK*	*TICK*	*TOCK*	*TICK*	*TOCK*	*TICK*
HEDS	X/QX 6xxx 3/5/7xxx *(2006)*	QX 96/97xx 3/5/7xxx *(2007)*	i7-965/975 3/5xxx *(2008)*	980/990 E7-xxx *(2010)*	38/39xx E3/E5/E7-xxxx *(2011)*	48/49xx v2 *(2012)*	58/59xx v3 *(2013)*	68/69xx v4 *(2014)*

M/DT	6. gen. (6xxx)	7. gen. (7xxx)	8. gen.[1] (8xxx)	9. gen.
	Skylake New Microarch. 14 nm	Kaby Lake New Microarch. 14 nm	Kaby Lake Refresh New Microarch. 14 nm	Ice Lake? New Process 10 nm
	TOCK	*TOCK*	*TOCK*	*TICK*
HEDS	i7-78xx/**i9**-79xx v5/**SP** *(2015)*	v6 *(2016)*	*(2017)*	*(2018)*

[1]Astonishingly, the 8th generation encompasses

- beyond Kaby Lake Refresh also the
- Kaby Lake G with AMD Vega graphics
- Coffee Lake and
- 10 nm Cannon Lake designs.

Intel's two-phase development model (Tick-Tock model) used for the Core 2 lines

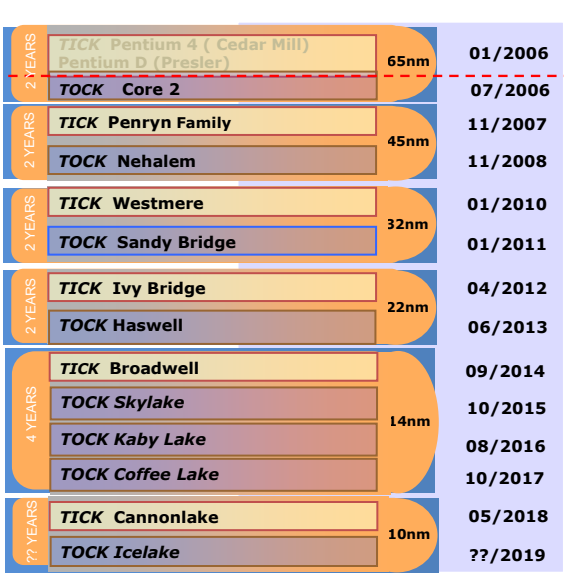
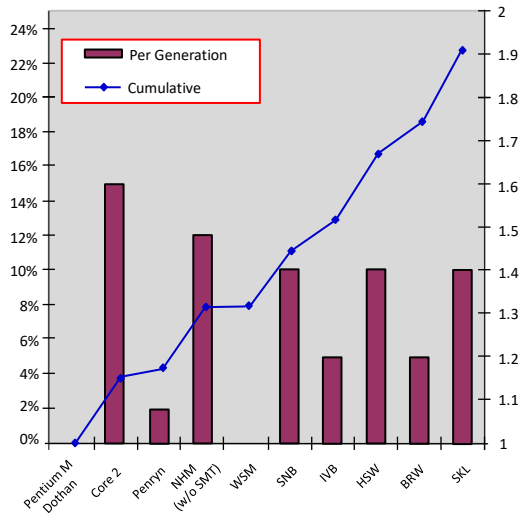


Figure is based on [3])

Raising of the single thread IPC in Intel's basic architectures



Note that Intel raised IPC in the Core family only less then 2-times in about 10 years.

Tock models contribute by about 10 %, Tick models by about 5 % for raising IPC.

Recent processor categories of Intel

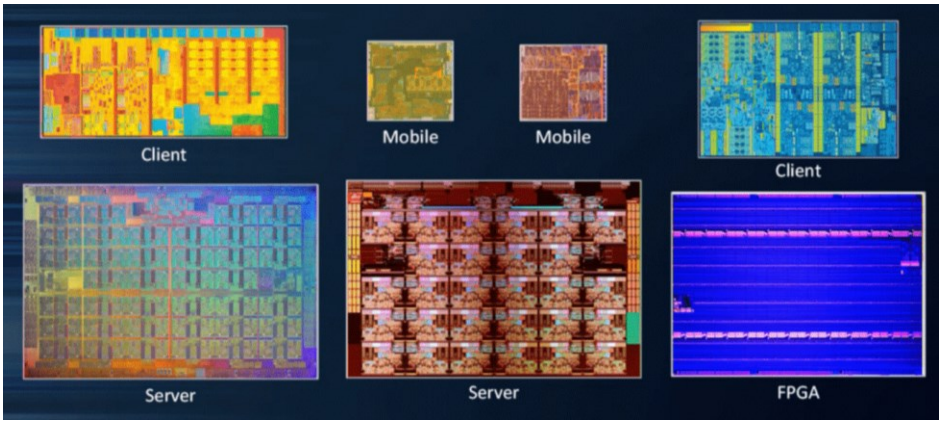
Servers	HEDs (High-End Desktops)	Desktops/Laptops (Client computing)	Tablets	Smartphones
Xeon E7/E5/E3 Platinum/Gold etc.	*Core i7/i9 (Extreme Edition or X models)*	*Core i7/i5/i3 (Basic architectures)*	*Atom lines (+ LP Core models)*	*Atom lines*
			Cancelled in 2016	Smarphones

Typical config.			
Up to 28C	2C - 18C	2C - 6C + G	(Up to 4C - 10C + G) (by other vendors)

14 nm dies of Intel's different processor categories (not on scale)

Broadwell (2C, 82 mm²)

Skylake (4C, 122 mm²)

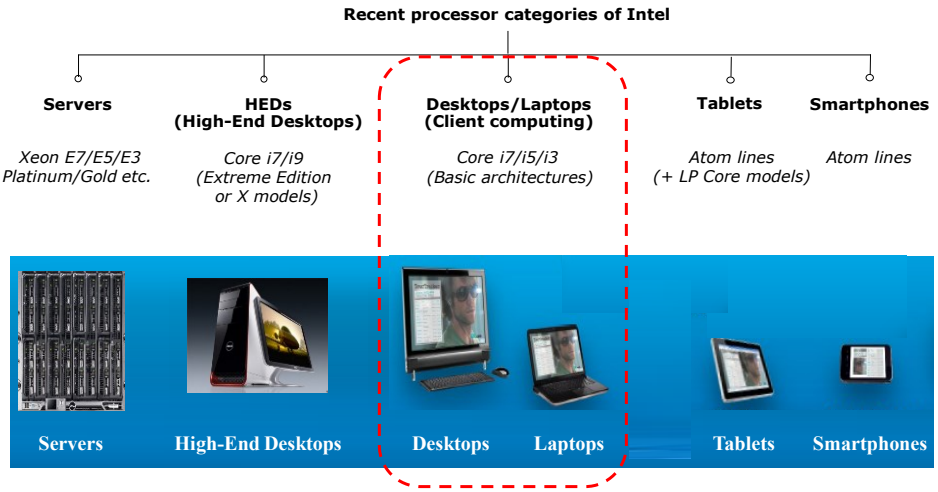


Knights Landing (72C, ~680 mm²)

Skylake-SP (28C, ~680 mm²)

1.2 Evolution of desktop and laptop processors

1.2 Evolution of desktop and laptop processors

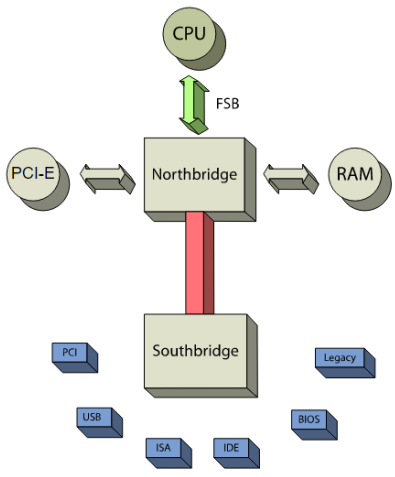


1.2 Evolution of desktop and laptop processors (3)

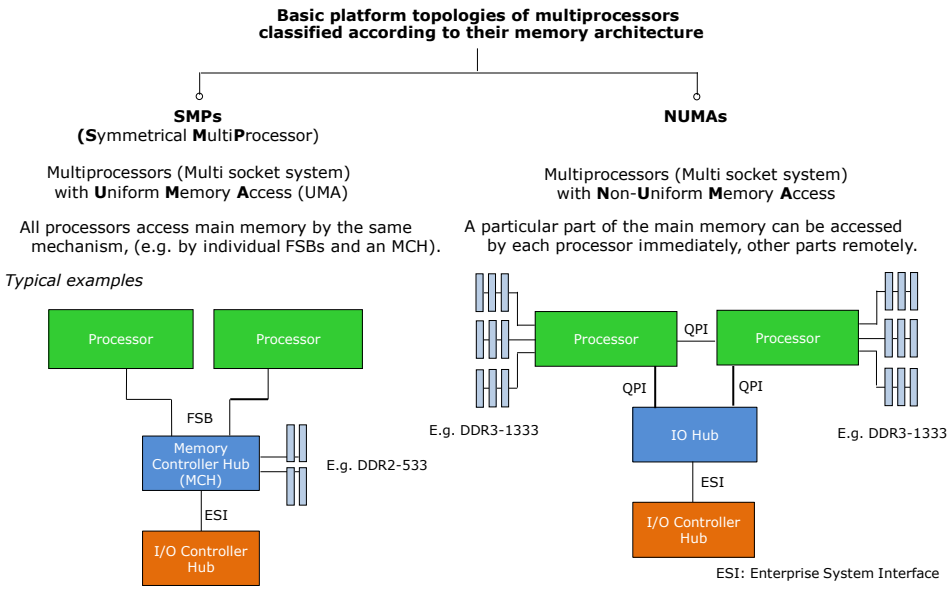
Max. number of cores, Number of memory channels, Platform topology, Speeding up the memory rate,

Designation of the platform	Platform topology	Processor	Technology	Max. no. of cores (n_c)	Year of introduction	Processor socket	MCH/PCH	Highest mem./ speed	No. of mem. channels
Anchor Creek	Off-die MC	Pentium D	90 nm	2x1	5/2005	LGA 775	945-955	DD2-667	2
Bridge Creek		Core 2	65 nm	2C	6/2006	LGA 775	945-975	DDR2-800	2
Salt Creek		Core 2 Quad	65 nm	2x2C	6/2007	LGA 775	3 Series (Bearlake)	DDR3-1067	2
Boulder Creek		Penryn	45 nm	2x2C	6/2008	LGA 775	4 Series (Eaglelake)	DDR3-1067	2
Kings Creek	On-die MC	2. G. Nehalem (Lynnfield)	45 nm	4C	9/2009	LGA 1156	5 Series (Ibex Peak)	DDR3-1333	2
		Westmere (Clarkdale)	32 nm	2C+G	1/2010	LGA 1156	5 Series (Ibex Peak)		
Sugar Bay		Sandy Bridge	32 nm	2C/4C+G	1/2011	LGA 1155	6 series (Cougar Point)	DDR3-1333	2
Maho Bay		Ivy Bridge	22 nm	4C+G	4/2012	LGA 1155	7 series (Panther Point)	DDR3-1600	2
Shark Bay		Haswell	22 nm	4C+G	6/2013	LGA 1155	8 Series (Lynx Point)	DDR3-1600	2
		Haswell refresh	22 nm	4C+G	5/2014	LGA 1150	9 Series (Wild Cat Point)	DDR3-1666	
		Broadwell	14 nm	4C+G	6/2015	LGA 1150	9 Series (Wild Cat Point)	DDR3-1866	
		Skylake	14 nm	4C+G	10/2015	LGA 1151	100 Series (Sunrise Point)	DDR4-2133	2
		Kaby Lake	14 nm	4C+G	8/2016	LGA 1151	200 Series	DDR4-2400	2
		Coffee Lake	14 nm	6C+G	10/2017	LGA 1151	300 Series	DDR4-2666	2

Intel Chipset Layout



Changing the basic platform topology of MP platforms from SMP to NUMA -2



Key features of the evolution of Intel's Core-based mainstream DT platforms

- Their core count rose soon (already in 2007) to 4 and remain at this figure.
- Two memory channels are used to connect memory.
- Memory is connected up to the Penryn via the MCH thereafter immediately via the processor.
- In a 10 year period memory speed rose about three to four times.

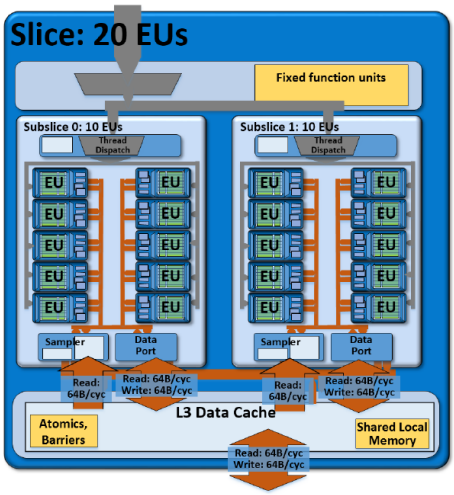
1.2 Evolution of desktop and laptop processors

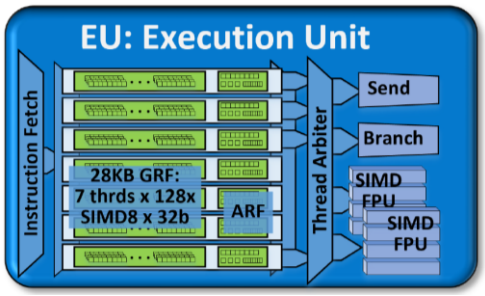
Evolution of main features of Intel's graphics families (based on [174]) **-1**

Intel Core generation	Graphics generation	Models	Graphics Technology level	No. of graphics slices	No. of EUs	eDRAM	OpenGL version	DirectX version	OpenCL version
Westmere	5th (Ironlake)	HD			12	--	2.1	10.1	n.a.
Sandy Bridge	6th	HD 2000	GT1	1 (2x3 EU)	6		3.1/3.3	10.1	n.a.
		HD 3000	GT2	1 (4x3 EU)	12				
Ivy Bridge	7th	HD 2500	GT1	1 (6 EU)	6	--	4.0	11.0	1.2
		HD 4000	GT2	1 (2x8 EU)	16				
Haswell	7.5th	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	8th	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	9th	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			

Example: A slice of the graphics unit of Haswell (6.5th gen. graphics)
1 slice (GT2): includs 20 EUs [199]

SIMD: Single Instruction Multiple Data





- 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.

1.2 Evolution of desktop and laptop processors

Evolution of main features of Intel's graphics families

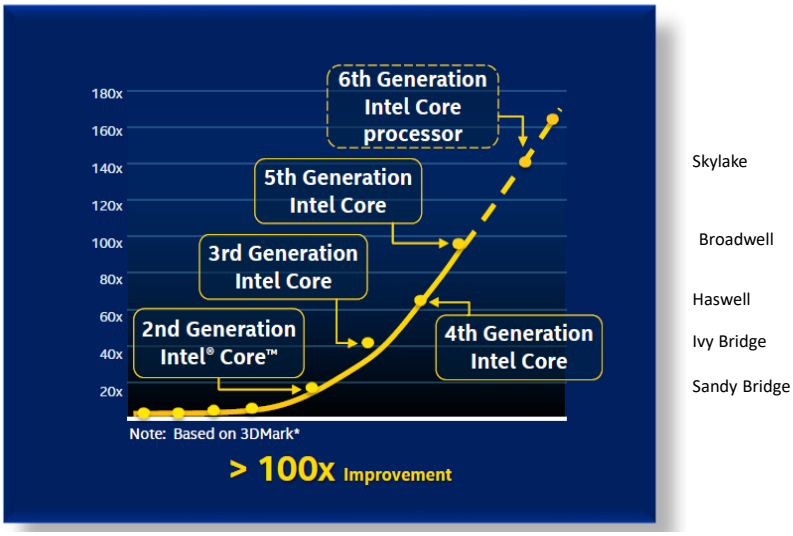
Intel Core generation	Graphics generation	Models	Graphics Technology level	No. of graphics slices	No. of EUs	eDRAM	OpenGL version	DirectX version	OpenCL version
Westmere	5th (Ironlake)	HD			12	--	2.1	10.1	n.a.
Sandy Bridge	6th	HD 2000	GT1	1 (2x3 EU)	6		3.1/3.3	10.1	n.a.
		HD 3000	GT2	1 (4x3 EU)	12				
Ivy Bridge	7th	HD 2500	GT1	1 (6 EU)	6	--	4.0	11.0	1.2
		HD 4000	GT2	1 (2x8 EU)	16				
Haswell	7.5th	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	8th	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	9th	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			

Key features of the evolution of Intel's integrated graphics families

- Intel's graphics implementations are subdivided into
  - graphics generations (5^{th} to 9^{th}) and
  - graphics technology levels (GT1-GT4).

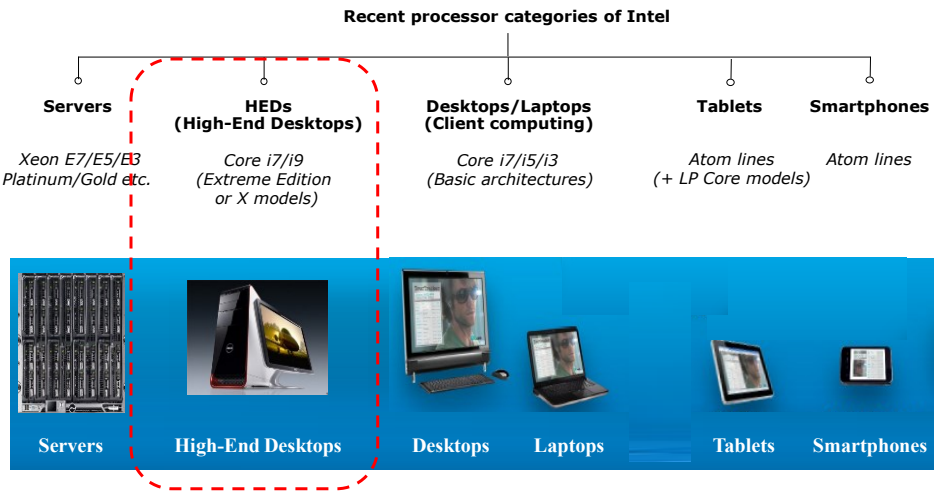
  Graphics generations are bound to the basic architectures.
  The graphics technology levels indicate the number of graphics slices (replicable sets of graphics EUs) within each graphics generation.
- The number of graphics EUs is increasing more or less according to Moore's rule (from 12 in 2010 to 72 in 2015).
- With the Haswell basic architecture graphics became eDRAM support (first 64 MB then 128 MB)
- There is continuous support of newer versions of graphics APIs and of OpenCL.

Graphics performance increase of subsequent Core generations [196]



1.3 Evolution of HEDs (High-End desktops)

1.3 Evolution of HEDs (High-End desktops)



Hardcore gamer scenario [221]

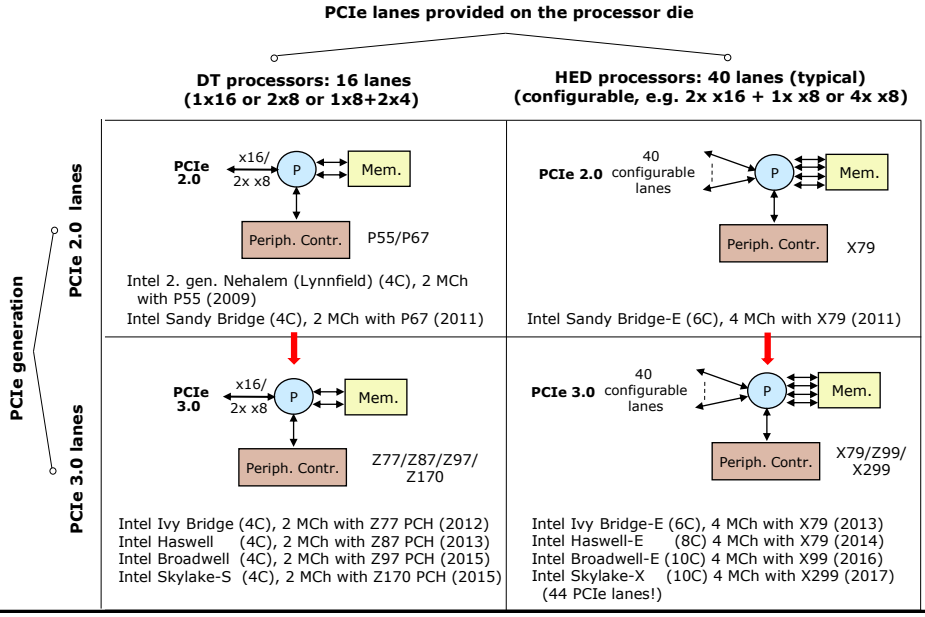
They aim at high performance desktops for hardcore gamers and graphics enthusiasts.

Number of PCIe lanes, no. of cores

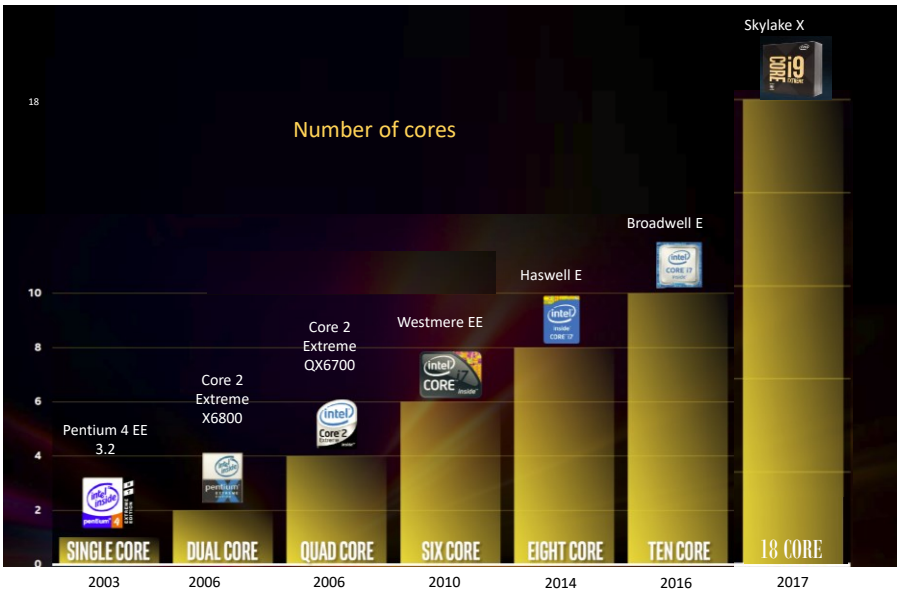
Processor	Techn.	Date of intro.	No. of cores up to	Memory attachment	Mem. speed up to	PCIe lanes	No. of mem. channels	MCH	TDP	Processor socket
Core 2 Extreme X6800	65 nm	7/2006	2C	Via MCH	DDR3-1066	32 PCIe 2.0 on the X38	2	X38	65 W	LGA 775
Core 2 Extreme QX 6xxx	65 nm	11/2006	2x2C						130 W	LGA 775
Core 2 Extreme QX 9650 (Penryn)	45 nm	11/2007	2x2C						130 W	LGA 775
Core 2 Extreme QX 9770 (Penryn)	45 nm	3/2008	2x2C		DDR3-1600	32 PCIe 2.0 on the X48	2	X48	136 W	LGA 775
1. G. Nehalem EE	45 nm	11/2008	4C	On-die MC	DDR3-1067	36 PCIe 2.0 on the X58	3	X58 (Tylersburg)	130 W	LGA 1366
Westmere EE	32 nm	3/2010	6C						130 W	LGA 1366
Sandy Bridge E	32 nm	11/2011	6C		DDR3-1600	40 PCIe 2.0 on-die	4	X79 (Patsburg)	130 W/ 150 W	LGA 2011
Ivy Bridge E	22 nm	9/2013	6C		DDR3-1866	40 PCIe 3.0 on-die	4		130 W	LGA 2011
Haswell E	14 nm	8/2014	8C		DDR4-2133		4	X99 (Wellsburg)	140 W	LGA 2011-3
Broadwell E	14 nm	5/2016	10C		DDR4-2400		4	X99 (Wellsburg)	140 W	LGA_2011-3
Skylake X	14 nm	6/2017	18C		DDR4-2666	44 PCI-3.0 on-die	4	X299 (Basin Falls)	140 W	LGA-2066
(Kaby Lake X)	14 nm	6/2017	4C		DDR4-2666	16 PCIe 3.0 on-die	4	X299 (Basin Falls)	112 W	LGA-2066

The above HED models and lines are unlocked.

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



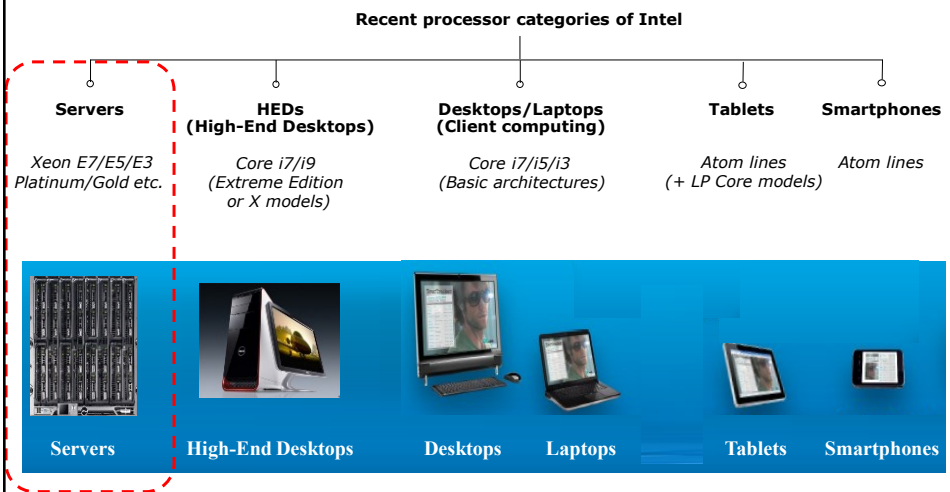
Evolution of the core counts in Intel's HED lines



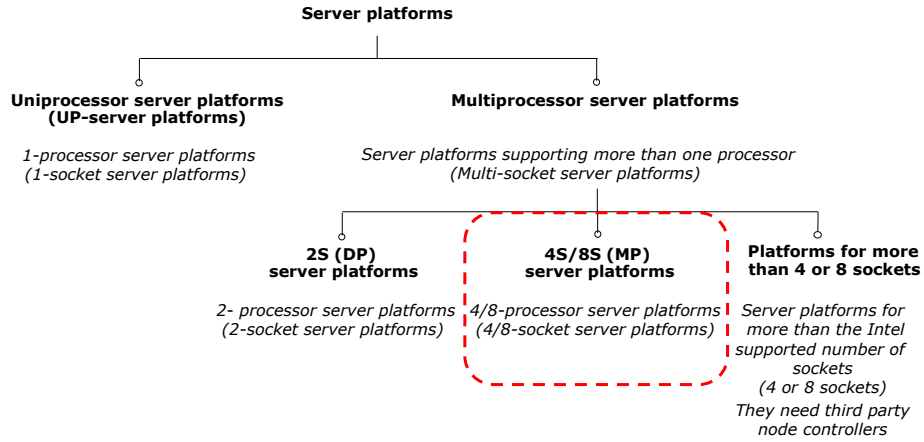
Key features of Intel's HED lines to support multiple (up to 4) discrete graphics cards:

- They typically provide vs. mainstream desktops
  - more PCIe lanes (either on the PCH or on the die)
  - more cores and
  - more memory channels (to appropriately service more processing resources)
- They are equipped with a large number of PCIe lanes (typically 32 to 44) to connect multiple discrete graphics cards vs. 16 typical in desktops,
- they do not provide integrated graphics, as it is assumed that the installation is intended to provide high quality graphics by attaching discrete graphics cards,
- they have more cores than desktops than graphics offers more parallelism than typical desktop workloads,
- HEDs have a higher TDP of 130 to 150 W vs. 65 to 95 W typical for DT processors.

1.4 Evolution of high-end 4S/8S servers



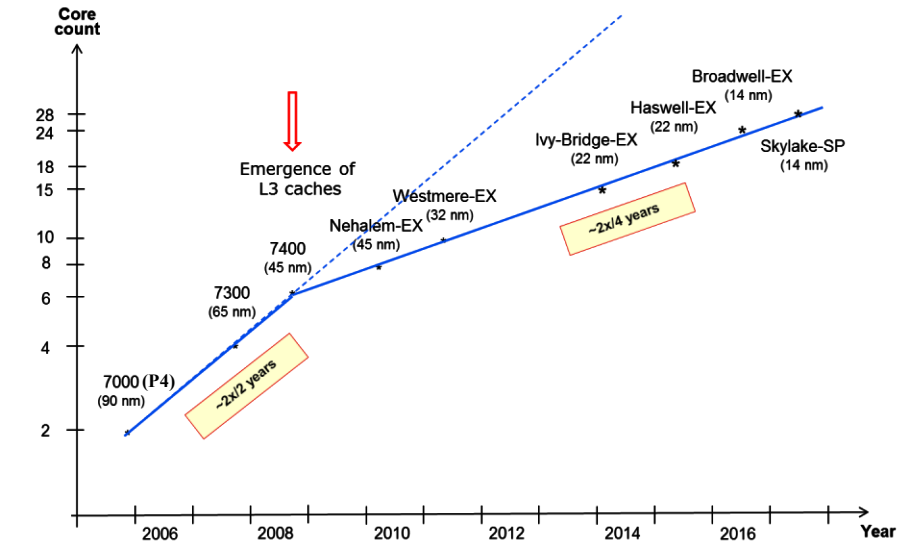
Server platforms classified according to the number of processors supported



Intel's high-end 4S/8S server platforms built up of multicore processors -- Overview

	Core	Techn.	Intro.	High-end 4S/8S server processor lines	Core count	Chipset	Proc. socket
Truland MP	Pentium 4 MP Prescott	90 nm	3/2005	90 nm Pentium 4 MP (Potomac)	1C	E8500 + ICH5	LGA 604
	Pentium 4 Presc.	90 nm	11/2005	7000 (Paxville MP)	2x1C	E8501 + ICH5	
	Pentium 4 Presc.	65 nm	8/2006	7100 (Tulsa)	2x1C		
Caneland	Core2	65 nm	9/2007	7200 (Tigerton DC) 7300 (Tigerton QC)	2C 2x2C	E7300 (Clarksboro) + 631x/632x ESB	LGA 604
	Penryn	45 nm	9/2008	7400 (Dunnington)	6C		
Boxboro-EX	Nehalem	45 nm	3/2010	7500 (Beckton/ Nehalem-EX)	8C	7500 + ICH10	LGA 1567
	Westmere	32 nm	4/2011	E7-8800 (Westmere-EX)	10C		
	Sandy Bidge	32 nm					
Brickland	Ivy Bridge	22 nm	2/2014	E7-8800 v2 (Ivy Bridge-EX)	15C	C602J (Patsburg J)	LGA 2011-1
	Haswell	22 nm	5/2015	E7-8800 v3 (Haswell-EX)	18C		
	Broadwell	14 nm	6/2016	E7-8800 v4 (Broadwell-EX)	24C		
Purley	Skylake	14 nm	7/2017	Platinum 8100 (Skylake-SP)	28C	C620 (Lewisburg)	LGA 3647

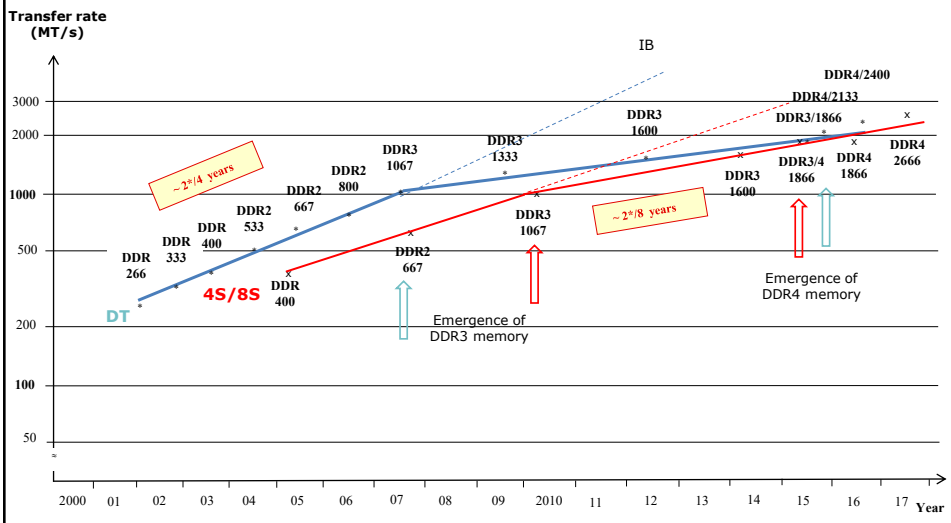
a) The rate of rising core counts in Intel's high-end 4S/8S server processors



*

2. Evolution of 4S/8S high-end servers

b) The rate of rising memory transfer rates in Intel's DT and high-end server processors



The rate of rising mem. transfer rates in Intel's high-end 4S/8S server processors

- As the above Figure shows, in the course of the evolution of high-end 4S/8S servers memory transfer rate has initially been doubled roughly every four years.
- But after DDR3 memory emerged the growth rate of memory transfer rate slowed down to doubling rougly every eight years due to the higher complexity of this technology.
- Note that in the considered time interval memory transfer rate has always been risen at a slower rate than core count, approximately at the half rate than the core count.
  - Thus the per core memory bandwidth will decrease since the memory bandwidth amounts to the product of the transfer rate and the width of the memory interface (assuming that no other changes become effective).
  - In other words, in the course of processor evolution a memory bandwidth gap arises that needs to be removed by implementing appropriately more memory channels.
  - Thus removing the memory bandwidth gap more memory channels are needed, this is the driving force of the evolution of server memory subsystems.
  - Since it is not at all a straightforward task to raise the number of memory channels, one of the focal points of server evolution is how to resolve the memory bandwidth gap by enhancing the memory subsystem.

*

Overview of Intel's x86 ISA extensions