Intel's Core family – The TOCK lines I Introduction and the Core 2 line

Dezső Sima

Vers. 3.0

August 2018

Contents

- 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 (outlook)

1. Introduction

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

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

1.1 Overview of the evolution of Intel's Pentium 4 and Core 2 families (1)

1.1 Overview of the evolution of Intel's Pentium 4 and Core 2 families (Based on [3])

The Core 2 family was preceded by the Pentium 4 family

The Pentim 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 [219].

Intel's Pentium 4 family



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 [220]. f



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 [220]

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 [74]

1.1 Overview of the evolution of Intel's Pentium 4 and Core 2 families (7)

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

High performance/power (e.g. GFLOPS/Watt)



Tablets and smartphones

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.

1.1.2 Overview of the evolution of Intel's Pentium 4 and Core 2 families - The generations

1. gen.				2	. gen.	3. gen.	4. gen.	5. gen.
Core 2 New Microarch. 65 nm	Penryn New Process 45 nm	Nehalem ^{New} Microarch. 45 nm	West- mere New Process 32 nm	S B Mi 3	Sandy Fridge New croarch. S2 nm	Ivy Bridge ^{New} Process 22 nm	Haswell New Microarchi. 22 nm	Broad- well New Process 14 nm
тоск	ТІСК	тоск	ТІСК	7	ГОСК	ТІСК	тоск	ТІСК
Haifa	Oregon	Oregon	Oregon	I	Haifa	Haifa	Oregon	Oregon
6. gen.	7. gen.	8. gen. ¹	9. gen.		¹ Aston four	ishingly, the 8 processor line	th generation s, as follows:	encompasses
Skylake New Microarch.	Kaby Lake New Microarch.	Kaby Lake R G-series Coffee Lake	Coffee Lake R New Mocroarcl	h.	• Ka • Ka • Co • 10	aby Lake Refre aby Lake G wit offee Lake (all) nm Cannon I	esh h AMD Vega g 14 nm) and th _ake designs [raphics ne 218].
14 nm	14 nm	Cannon Lake 14/10 nm	14 nm		Rem The	ark Haifa (Israel) I	Development (Center
тоск	тоск	тоск	тоск		was Pen	s also responsi ntium M (Bania	ible for the 808 is, Dothan, Jor	80 and nah
Haifa	Haifa	Haifa			(Co R: Re	ore Solo/Core l efresh	Juo)) processo	ors

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

1. gen.				2. gen.	3. gen.	4. gen.	5. gen.
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
тоск	ТІСК	тоск	ТІСК	тоск	ТІСК	тоск	ТІСК
(2006)	(2007)	(2008)	(2010)	(2011)	(2012)	(2013)	(2014)
6. gen.	7. gen.	8. gen. ¹	9. gen	. ¹ Astor four	nishingly, the & r processor line	8th generation es, as follows:	encompasses
Skylake New Microarch.	Kaby Lake New Microarch.	Kaby Lake R G-series Coffee Lake Cannon Lake	Coffee Lake F New Mocroarc	• K • K • C • 1	Caby Lake Refr Caby Lake G wi Coffee Lake and 0 nm Cannon	esh th AMD Vega o d Lake designs [graphics [218].
14 nm	14 nm	14/10 nm	14 nm				
тоск	тоск	тоск	тоск				
(2015)	(2016)	(2017/18)	(2018)				

R: Refresh

Basic designations of Intel's processor lines



i7-78xx/i9-79xx

S v5/**SP** v6

Basic microarchitectures and their related shrinks

Considered from the Pentium 4 Prescott core (the third core of Pentium 4) on.

Basic architectures	Basic architectures and their shrinks
Pentium 4 (Prescott)	2005 90 nm Pentium 4 2006 65 nm Pentium 4
Core 2	2006 65 nm Core 2 2007 45 nm Penryn
Nehalem	200845 nmNehalem201032 nmWestmere
Sandy Bridge	201132 nmSandy Bridge201222 nmIvy Bridge
Haswell	2013 22 nm Haswell 2014 14 nm Broadwell
Skylake	201514 nmSkylake201614 nmKaby Lake201714 nmCoffee Lake201810 nmCannon Lake
Icelake	2019? 10 nm Icelake

Evolution of Intel's IC technology



On Intel's Q2 2015 earnings conference call, on July 16 2015, Krzanich, Intel's CEO (Chief Executive Officer) told: in the second half of 2017, we expect to launch our first 10-nanometer product, code named Cannonlake.
The last two technology transitions have signaled that our cadence today is closer to 2.5 years than two" [180].

Intel's two-phase development model (Tick-Tock model) used for the Core 2 lines -21 (based on [3])



Intel's two-phase development model (Tick-Tock model) used for the Core 2 lines -2 (based on [3])



Raising of the single thread IPC in Intel's basic architectures (Based on [195])



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.

General implications of the evolution of transistor technology [245]



Presentation at AMD's Financial Analyst Day May 16 2017

1.1 Overview of the evolution of Intel's Pentium 4 and Core 2 families (18)



1.1 Overview of the evolution of Intel's Pentium 4 and Core 2 families (19)

14 nm dies of Intel's different processor categories (not on scale) [247]

Broadwell (2C, 82 mm²)

Skylake (4C, 122 mm²)



Knights Landing (72C, ~680 mm²) Skylake-SP (28C, ~680 mm²)

1.1 Overview of the evolution of Intel's Pentium 4 and Core 2 families (19)

14 nm dies of Intel's different processor categories (approximately on scale) [247]

Broadwell (2C, 82 mm²)



Skylake (4C, 122 mm²)



Skylake-SP (28C, ~680 mm²)



1.2 Evolution of desktop and laptop processors

1.2 Evolution of desktop and laptop processors



Intel's Core-based mainstream DT platforms

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		Pentium D	90 nm	2x1	5/2005	LGA 775	945-955	DD2-667	2			
Bridge Creek	Off-die	Core 2	65 nm	2C	6/2006	LGA 775	945-975	DDR2-800	2			
Salt Creek	MC	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			
		2. G. Nehalem (Lynnfield)	45 nm	4C	9/2009	LGA 1156	5 Series (Ibex Peak)	1222	C			
Kings Creek		Westmere (Clarkdale)	32 nm	2C+G	1/2010	LGA 1156	5 Series (Ibex Peak)		2			
Sugar Bay		Sandy Bridge	32 nm	2C/4C+G	1/2011	LGA 1155	6 series (Cougar Point)	DDR3-1333	2			
Maho Bay	On-die	Ivy Bridge	22 nm	4C +G	4/2012	LGA 1155	7 series (Panther Point)	DDR3-1600	2			
Shark Bay	МС	Haswell	22 nm	4C+G	6/2013	LGA 1155	8 Series (Lynx Point)	DDR3-1600				
		Haswell refresh	22 nm	4C+G	5/2014	LGA 1150	9 Series (Wild Cat Point)	DDR3-1666	2			
	-	-	ſ	ľ	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			

Max. number of cores

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		Pentium D	90 nm	2x1	5/2005	LGA 775	945-955	DD2-667	2			
Bridge Creek	Off-die	Core 2	65 nm	2C	6/2006	LGA 775	945-975	DDR2-800	2			
Salt Creek	MC	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			
Kingg Crack		2. G. Nehalem (Lynnfield)	45 nm	4C	9/2009	LGA 1156	5 Series (Ibex Peak)	1222	ſ			
Kings Creek		Westmere (Clarkdale)	32 nm	2C+G	1/2010	LGA 1156	5 Series (Ibex Peak)		Z			
Sugar Bay		Sandy Bridge	32 nm	2C/4C+G	1/2011	LGA 1155	6 series (Cougar Point)	DDR3-1333	2			
Maho Bay	On-die	Ivy Bridge	22 nm	4C +G	4/2012	LGA 1155	7 series (Panther Point)	DDR3-1600	2			
Shark Bay	MC	Haswell	22 nm	4C+G	6/2013	LGA 1155	8 Series (Lynx Point)	DDR3-1600				
		Haswell refresh	22 nm	4C+G	5/2014	LGA 1150	9 Series (Wild Cat Point)	DDR3-1666	2			
	F	-	-	ľ	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			

Number of memory channels

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		Pentium D	90 nm	2x1	5/2005	LGA 775	945-955	DD2-667	2			
Bridge Creek	Off-die	Core 2	65 nm	2C	6/2006	LGA 775	945-975	DDR2-800	2			
Salt Creek	MC	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			
		2. G. Nehalem (Lynnfield)	45 nm	4C	9/2009	LGA 1156	5 Series (Ibex Peak)	1222	ſ			
Kings Creek		Westmere (Clarkdale)	32 nm	2C+G	1/2010	LGA 1156	5 Series (Ibex Peak)	DDK3-1333	2			
Sugar Bay		Sandy Bridge	32 nm	2C/4C+G	1/2011	LGA 1155	6 series (Cougar Point)	DDR3-1333	2			
Maho Bay	On-die	On-die	On-die	Ivy Bridge	22 nm	4C +G	4/2012	LGA 1155	7 series (Panther Point)	DDR3-1600	2	
Shark Bay	МС	Haswell	22 nm	4C+G	6/2013	LGA 1155	8 Series (Lynx Point)	DDR3-1600				
	-	-	-	-	Haswell refresh	22 nm	4C+G	5/2014	LGA 1150	9 Series (Wild Cat Point)	DDR3-1666	2
									Broadwell	14 nm	4C+G	6/2015
		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 r1	200 Series (Union Point)	DDR4-2400	2			
		Coffee Lake	14 nm	6C +G	10/2017	LGA 1151 r2	300 Series	DDR4-2666	2			

Platform topology

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		Pentium D	90 nm	2x1	5/2005	LGA 775	945-955	DD2-667	2	
Bridge Creek	Off-die	Core 2	65 nm	2C	6/2006	LGA 775	945-975	DDR2-800	2	
Salt Creek	MC	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	
		2. G. Nehalem (Lynnfield)	45 nm	4C	9/2009	LGA 1156	5 Series (Ibex Peak)	1222	C	
Kings Creek		Westmere (Clarkdale)	32 nm	2C+G	1/2010	LGA 1156	5 Series (Ibex Peak)	DDR3-1333	Z	
Sugar Bay		Sandy Bridge	32 nm	2C/4C+G	1/2011	LGA 1155	6 series (Cougar Point)	DDR3-1333	2	
Maho Bay	On-die	Ivy Bridge	22 nm	4C +G	4/2012	LGA 1155	7 series (Panther Point)	DDR3-1600	2	
Shark Bay	МС	Haswell	22 nm	4C+G	6/2013	LGA 1155	8 Series (Lynx Point)	DDR3-1600		
		Haswell refresh	22 nm	4C+G	5/2014	LGA 1150	9 Series (Wild Cat Point)	DDR3-1666	2	
			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	

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



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		Pentium D	90 nm	2x1	5/2005	LGA 775	945-955	DD2-667	2												
Bridge Creek	Off-die MC	Core 2	65 nm	2C	6/2006	LGA 775	945-975	DDR2-800	2												
Salt Creek	(SMP)	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												
		2. G. Nehalem (Lynnfield)	45 nm	4C	9/2009	LGA 1156	5 Series (Ibex Peak)		2												
Kings Creek		Westmere (Clarkdale)	32 nm	2C+G	1/2010	LGA 1156	5 Series (Ibex Peak)	DDK3-1333	2												
Sugar Bay		Sandy Bridge	32 nm	2C/4C+G	1/2011	LGA 1155	6 series (Cougar Point)	DDR3-1333	2												
	On-die	Ivy Bridge	22 nm	4C +G	4/2012	LGA 1155	7 series (Panther Point)	DDR3-1600	2												
	(NUMA)	Haswell	22 nm	4C+G	6/2013	LGA 1155	8 Series (Lynx Point)	DDR3-1600													
		Haswell refresh	22 nm	4C+G	5/2014	LGA 1150	9 Series (Wild Cat Point)	DDR3-1666	2												
		F			Ī	Ī								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												

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.

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	5 th (Ironlake)	HD			12		2.1	10.1	n.a.
Sandy	6th	HD 2000	GT1	1 (2x3 EU)	6		3 1/3 3	10 1	na
Bridge		HD 3000	GT2	1 (4x3 EU)	12		5.1/5.5	10.1	
Tvv Bridge	7th	HD 2500	GT1	1 (6 EU)	6		4 0	11 0	12
ity bridge	7 611	HD 4000	GT2	1 (2x8 EU)	16		4.0	11.0	1.2
		HD 4200- HD 4700	GT2	1 (2x10 EU)	20				
Haswell	7.5th	HD 5000 Iris 5100	GT3	2	40		4.3	11.1	1.2
		Iris Pro 5200				128 MB			
		HD 5300- HD 5600	GT2	1 (3x8 EU)	23/24				
Broadwell	8th	HD 6000 Iris 6100	GT3	2	47/48	4.3		11.2	2.0
		Iris Pro 6200	GT3e	2	48	128 MB			
		HD 510	GT1	1 (3x4 EU)	12				
		HD 515	GT1.5	1 (3x6 EU)	18				
Skylake	9th	HD 520	GT2	1 (3x8 EU)	24		4.4	12	2.0
Skylake	5011	HD 535	GT3	2	48				210
		HD 540	GT3e	2	48	64 MB			
		HD 580	GT4e	3	72	64/128 MB			

Evolution of main features of Intel's graphics families (based on [174]) -2

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	6th	HD 2000	GT1	1 (2x3 EU)	6		3 1/3 3	10 1	na
Bridge	0011	HD 3000	GT2	1 (4x3 EU)	12		5.1/5.5	10.1	
Tvv Bridge	7th	HD 2500	GT1	1 (6 EU)	6		4 0	11 0	12
ivy bridge	7 (11	HD 4000	GT2	1 (2x8 EU)	16		4.0	11.0	1.2
		HD 4200- HD 4700	GT2	1 (2x10 EU)	20				
Haswell	7.5th	HD 5000 Iris 5100	GT3	2	40		4.3	11.1	1.2
		Iris Pro 5200				128 MB			
		HD 5300- HD 5600	GT2	1 (3x8 EU)	23/24				
Broadwell	8th	HD 6000 Iris 6100	GT3	2	47/48		4.3	11.2	2.0
		Iris Pro 6200	GT3e	2	48	128 MB			
		HD 510	GT1	1 (3x4 EU)	12				
		HD 515	GT1.5	1 (3x6 EU)	18				
Skylake	9th	HD 520	GT2	1 (3x8 EU)	24		44	12	2.0
Skylake	5011	HD 535	GT3	2	48				210
		HD 540	GT3e	2	48	64 MB			
		HD 580	GT4e	3	72	64/128 MB			

Evolution of main features of Intel's graphics families (based on [174]) -3

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	6th	HD 2000	GT1	1 (2x3 EU)	6		3 1/3 3	10.1	na
Bridge		HD 3000	GT2	1 (4x3 EU)	12		5.1/5.5	10.1	n.a.
Tvv Bridge	7th	HD 2500	GT1	1 (6 EU)	6		4.0	11.0	12
ivy bridge	7 (11	HD 4000	GT2	1 (2x8 EU)	16		4.0	11.0	1.2
		HD 4200- HD 4700	GT2	1 (2x10 EU)	20				
Haswell	7.5th	HD 5000 Iris 5100	GT3	2	40		4.3	11.1	1.2
		Iris Pro 5200			-	128 MB			
		HD 5300- HD 5600	GT2	1 (3x8 EU)	23/24				
Broadwell	8th	HD 6000 Iris 6100	GT3	2	47/48	4.3		11.2	2.0
		Iris Pro 6200	GT3e	2	48	128 MB			
		HD 510	GT1	1 (3x4 EU)	12				
		HD 515	GT1.5	1 (3x6 EU)	18				
Skylake	9th	HD 520	GT2	1 (3x8 EU)	24		4.4	12	2.0
Skylake	501	HD 535	GT3	2	48			12	210
		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]



Block diagram of an EU of Haswell -1 [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.

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



- Each SIMD FPUs can execute 4 SP FP They can execute MAD instructions (Multiply-Add)/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 (1/2/4/8/16/32 bit wide FX operations).
- One of the FPUs also support transcendental math functions.

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

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	6th	HD 2000	GT1	1 (2x3 EU)	6		3 1/3 3	10 1	na	
Bridge		HD 3000	GT2	1 (4x3 EU)	12		5.1/5.5	10.1		
Tvv Bridge	7th	HD 2500	GT1	1 (6 EU)	6		4 0	11.0	1 2	
ity bridge	, ch	HD 4000	GT2	1 (2x8 EU)	16		110	11.0	1.2	
		HD 4200- HD 4700	GT2	1 (2x10 EU)	20					
Haswell	vell 7.5th	HD 5000 Iris 5100	GT3	2	40		4.3	11.1	1.2	
		Iris Pro 5200				128 MB				
		HD 5300- HD 5600	GT2	1 (3x8 EU)	23/24					
Broadwell	8th	HD 6000 Iris 6100	GT3	2	47/48		4.3	11.2	2.0	
		Iris Pro 6200	GT3e	2	48	128 MB				
		HD 510	GT1	1 (3x4 EU)	12					
		HD 515	GT1.5	1 (3x6 EU)	18					
Skylake	9th	HD 520	GT2	1 (3x8 EU)	24		4.4	12	2.0	
Skylake	5011	HD 535	GT3	2	48				2.0	
		HD 540	GT3e	2	48	64 MB				
		HD 580	GT4e	3	72	64/128 MB			*	

Evolution of main features of Intel's graphics families (based on [174]) -4

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	6th	HD 2000	GT1	1 (2x3 EU)	6		3 1/3 3	10 1	na
Bridge	0011	HD 3000	GT2	1 (4x3 EU)	12		5.1/5.5	10.1	
Tvv Bridge	7th	HD 2500	GT1	1 (6 EU)	6		4.0	11.0	12
ivy bridge	7 (11	HD 4000	GT2	1 (2x8 EU)	16		4.0	11.0	1.2
		HD 4200- HD 4700	GT2	1 (2x10 EU)	20				
Haswell	7.5th	HD 5000 Iris 5100	GT3	2	40		4.3	11.1	1.2
		Iris Pro 5200				128 MB			
		HD 5300- HD 5600	GT2	1 (3x8 EU)	23/24				
Broadwell	8th	HD 6000 Iris 6100	GT3	2	47/48		4.3	11.2	2.0
		Iris Pro 6200	GT3e	2	48	128 MB			
		HD 510	GT1	1 (3x4 EU)	12				
		HD 515	GT1.5	1 (3x6 EU)	18				
Skylake	9th	HD 520	GT2	1 (3x8 EU)	24		44	12	2.0
Skylake	501	HD 535	GT3	2	48			12	210
		HD 540	GT3e	2	48	64 MB			
		HD 580	GT4e	3	72	64/128 MB			

Evolution of main features of Intel's graphics families (based on [174]) -5

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	6th	HD 2000	GT1	1 (2x3 EU)	6		3 1/3 3	10 1	na
Bridge	0011	HD 3000	GT2	1 (4x3 EU)	12		5.1/5.5	10.1	
Tvv Bridge	7th	HD 2500	GT1	1 (6 EU)	6		4.0	11.0	12
ity bridge	7 (11	HD 4000	GT2	1 (2x8 EU)	16		1.0	11.0	1.2
		HD 4200- HD 4700	GT2	1 (2x10 EU)	20				
Haswell	7.5th	HD 5000 Iris 5100	GT3	2	40		4.3	11.1	1.2
		Iris Pro 5200			-	128 MB			
		HD 5300- HD 5600	GT2	1 (3x8 EU)	23/24				
Broadwell	8th	HD 6000 Iris 6100	GT3	2	47/48		4.3	11.2	2.0
		Iris Pro 6200	GT3e	2	48	128 MB			
		HD 510	GT1	1 (3x4 EU)	12				
		HD 515	GT1.5	1 (3x6 EU)	18				
Skylake	9th	HD 520	GT2	1 (3x8 EU)	24		44	12	2.0
Skylake	501	HD 535	GT3	2	48			12	2.0
		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 (5th to 9th) 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)



HEDs (High-End Desktops)

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

Hardcore gamer scenario [221]



1. Introduction -2

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)

as indicated on the next Figure.

Number of PCIe lanes

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						65 W	LGA 775
Core 2 Extreme QX 6xxx	65 nm	11/2006	2x2C		DDR3-1066	32 PCIe 2.0 on the X38	2	X38	130 W	LGA 775
Core 2 Extreme QX 9650 (Penryn)	45 nm	11/2007	2x2C	Via MCH					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		DDP3-1067	36 PCIe 2.0	3	X58	130 W	LGA 1366
Westmere EE	32 nm	3/2010	6C		DDRS-1007	on the X58	J	(Tylersburg)	130 W	LGA 1366
Sandy Bridge E	32 nm	11/2011	6C	-	DDR3-1600	40 PCIe 2.0 on-die	4	X79	130 W/ 150 W	LGA 2011
Ivy Bridge E	22 nm	9/2013	6C	On-dio	DDR3-1866		4	(Patsburg)	130 W	LGA 2011
Haswell E	14 nm	8/2014	8C	MC	DDR4-2133	40 PCIe 3.0 on-die	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

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



PCIe generation

Core counts

Processor	Techn.	Date of intro.	No. of cores up to	Memory attachment	Mem. speed up to	PCIe lanes	No. of mem. channels	МСН	TDP	Processor socket
Core 2 Extreme X6800	65 nm	7/2006	2C						65 W	LGA 775
Core 2 Extreme QX 6xxx	65 nm	11/2006	2x2C		DDR3-1066	32 PCIe 2.0 on the X38	2	X38	130 W	LGA 775
Core 2 Extreme QX 9650 (Penryn)	45 nm	11/2007	2x2C	Via MCH					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		DDP3-1067	36 PCIe 2.0	2	X58	130 W	LGA 1366
Westmere EE	32 nm	3/2010	6C		DDK3-1007	on the X58	5	(Tylersburg)	130 W	LGA 1366
Sandy Bridge E	32 nm	11/2011	6C		DDR3-1600	40 PCIe 2.0 on-die	4	X79	130 W/ 150 W	LGA 2011
Ivy Bridge E	22 nm	9/2013	6C	On-die	DDR3-1866		4	(Patsburg)	130 W	LGA 2011
Haswell E	14 nm	8/2014	8C	MC	DDR4-2133	40 PCIe 3.0 on-die	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.

Remark

The Kaby Lake-X models in fact do not fit into the traditional HED line, they are actually the highest performance models of the KabyLake line and are designated as mobile models.

Evolution of the core counts in Intel's HED lines (Based on [222])



Number of memory channels

Processor	Techn.	Date of intro.	No. of cores up to	Memory attachment	Mem. speed up to	PCIe lanes	No. of mem. channels	МСН	TDP	Processor socket
Core 2 Extreme X6800	65 nm	7/2006	2C						65 W	LGA 775
Core 2 Extreme QX 6xxx	65 nm	11/2006	2x2C		DDR3-1066	32 PCIe 2.0 on the X38	2	X38	130 W	LGA 775
Core 2 Extreme QX 9650 (Penryn)	45 nm	11/2007	2x2C	Via MCH					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		DDP3-1067	36 PCIe 2.0	3	X58	130 W	LGA 1366
Westmere EE	32 nm	3/2010	6C		DDK3-1007	on the X58	5	(Tylersburg)	130 W	LGA 1366
Sandy Bridge E	32 nm	11/2011	6C		DDR3-1600	40 PCIe 2.0 on-die	4	X79	130 W/ 150 W	LGA 2011
Ivy Bridge E	22 nm	9/2013	6C	On-die	DDR3-1866		4	(Patsburg)	130 W	LGA 2011
Haswell E	14 nm	8/2014	8C	MC	DDR4-2133	40 PCIe 3.0 on-die	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

Unlocked clock multiplier

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						65 W	LGA 775
Core 2 Extreme QX 6xxx	65 nm	11/2006	2x2C		DDR3-1066	32 PCIe 2.0 on the X38	2	X38	130 W	LGA 775
Core 2 Extreme QX 9650 (Penryn)	45 nm	11/2007	2x2C	Via MCH					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		DDP3-1067	36 PCIe 2.0	2	X58	130 W	LGA 1366
Westmere EE	32 nm	3/2010	6C		DDK3-1007	on the X58	5	(Tylersburg)	130 W	LGA 1366
Sandy Bridge E	32 nm	11/2011	6C		DDR3-1600	40 PCIe 2.0 on-die	4	X79	130 W/ 150 W	LGA 2011
Ivy Bridge E	22 nm	9/2013	6C	On-die	DDR3-1866		4	(Patsburg)	130 W	LGA 2011
Haswell E	14 nm	8/2014	8C	MC	DDR4-2133	40 PCIe 3.0 on-die	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.

TDP (Thermal Design Power)

Processor	Techn.	Date of intro.	No. of cores up to	Memory attachment	Mem. speed up to	PCIe lanes	No. of mem. channels	МСН	TDP	Processor socket
Core 2 Extreme X6800	65 nm	7/2006	2C						65 W	LGA 775
Core 2 Extreme QX 6xxx	65 nm	11/2006	2x2C		DDR3-1066	32 PCIe 2.0 on the X38	2	X38	130 W	LGA 775
Core 2 Extreme QX 9650 (Penryn)	45 nm	11/2007	2x2C	Via MCH					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			36 PCIe 2.0	2	X58	130 W	LGA 1366
Westmere EE	32 nm	3/2010	6C		DDK3-1007	on the X58	J	(Tylersburg)	130 W	LGA 1366
Sandy Bridge E	32 nm	11/2011	6C	-	DDR3-1600	40 PCIe 2.0 on-die	4	X79	130 W/ 150 W	LGA 2011
Ivy Bridge E	22 nm	9/2013	6C	On-die	DDR3-1866		4	(Patsburg)	130 W	LGA 2011
Haswell E	14 nm	8/2014	8C	MC	DDR4-2133	40 PCIe 3.0 on-die	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.

Introduction -3

Key features of Intel's HED lines (cont.):

- 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,
- they are unlocked and
- 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

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
	Pentium 4 MP Prescott	90 nm	3/2005	90 nm Pentium 4 MP (Potomac)	1C	E8500 + ICH5	
Truland MP	Pentium 4 Presc.	90 nm	11/2005	7000 (Paxville MP)	2x1C	E8501 +	LGA 604
	Pentium 4 Presc.	65 nm	8/2006	7100 (Tulsa)	2x1C	ICH5	
	Core2	65 nm	9/2007	7200 (Tigerton DC) 7300 (Tigerton QC)	2C 2x2C	E7300 (Clarksboro)	
Caneland	Penryn	45 nm	9/2008	7400 (Dunnington)	6C	+ 631x/632x ESB	LGA 604
Boxboro EV	Nehalem	45 nm	3/2010	7500 (Beckton/ Nehalem-EX)	8C	7500 +	
BOXDOLO-EX	Westmere	32 nm	4/2011	E7-8800 (Westmere-EX)	10C	ICH10	LGA 1507
	Sandy Bidge	32 nm					
	Ivy Bridge	22 nm	2/2014	E7-8800 v2 (Ivy Bridge-EX)	15C		
Brickland	Haswell	22 nm	5/2015	E7-8800 v3 (Haswell-EX)	18C	C602J (Patsburg J)	LGA 2011-1
	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 -1

	Core	Techn.	Intro.	High-end 4S/8S server processor lines	Core count	Chipset	Proc. socket
	Pentium 4 MP Prescott	90 nm	3/2005	90 nm Pentium 4 MP (Potomac)	1C	E8500 + ICH5	
Truland MP	Pentium 4 Presc.	90 nm	11/2005	7000 (Paxville MP)	2x1C	E8501 +	LGA 604
	Pentium 4 Presc.	65 nm	8/2006	7100 (Tulsa)	2x1C	ICH5	
	Core2	65 nm	9/2007	7200 (Tigerton DC) 7300 (Tigerton QC)	2C 2x2C	E7300 (Clarksboro)	
Caneland	Penryn	45 nm	9/2008	7400 (Dunnington)	6C	+ 631x/632x ESB	LGA 604
Boyboro EV	Nehalem	45 nm	3/2010	7500 (Beckton/ Nehalem-EX)	8C	7500 +	
BOXDOFO-EX	Westmere	32 nm	4/2011	E7-8800 (Westmere-EX)	6C 631 8C 75 10C 10C	ICH10	LGA 1567
	Sandy Bidge	32 nm					
	Ivy Bridge	22 nm	2/2014	E7-8800 v2 (Ivy Bridge-EX)	15C		
Brickland	Haswell	22 nm	5/2015	E7-8800 v3 (Haswell-EX)	18C	C602J (Patsburg J)	LGA 2011-1
	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



Remarks to the rate of rising core counts in Intel's high-end 4S/8S server processors -1 (Not discussed)

• As the Figure above demonstrates, the core count of Intel's early L3-less 4S server processors has doubled roughly biannually.

This is obvious, since the evolving IC technology allowed to double transistor counts roughly in every new technology node.

Consequently, in L3-less processors twice as many transistors per technology node permit the doubling of core counts also roughly biannually.

 When L3 caches first appeared in the Xeon 7400 within Intel's Core 2 based server processors, the transistor count rose to 1.9 billion from 582 million in the L3-less Xeon 7300, as indicated in Table below.

Model	Technology	Date of introduction	Core count	L3 cache	Transistors (million)	Die area (mm²)
7300	65 nm	9/2007	2x2		582	286
7400	45 nm	8/2008	6	16 MB	1900	503
Nehalem-EX	45 nm	3/2010	8	8x3 MB	2300	513
Westmere-EX	32 nm	4/2011	10	10x3 MB	2600	512
Ivy Bridge-EX	22 nm	2/2014	15	15x2.5 MB	4310	541
Haswell-EX	22 nm	5/2015	18	18x2.5 MB	5700	662
Broadwell-EX	14 nm	6/2016	24	24x2.5 MB	7200	456

Table: Rising transistor counts in Intel's high-end 4S/8S server processors

Die plot of the Xeon 7400 MP implementing 16 MB L3 cache [215]



Remarks to the rate of rising core counts in Intel's high-end 4S/8S server processors -2

- However, to place twice as many cores onto the die with the same L3 size per core in every technology node, would need to double this already high transistor count in every technology node.
- Nevertheless, mainly due to power restrictions Intel could raise transistor counts on server dies only at the modest rate of doubling roughly every four years, as above Table reveals, leading to doubling core counts approximately every four years, as depicted in the above Figure.
 - Recently, the nascent lengthening of the cadence of IC technology transitions also contributes to reducing the rate at which the core count can be increased.

Model	Technology	Date of introduction	Core count	L3 cache	Transistors (million)	Die area (mm²)
7300	65 nm	9/2007	2x2		582	286
7400	45 nm	8/2008	6	16 MB	1900	503
Nehalem-EX	45 nm	3/2010	8	8x3 MB	2300	513
Westmere-EX	32 nm	4/2011	10	10x3 MB	2600	512
Ivy Bridge-EX	22 nm	2/2014	15	15x2.5 MB	4310	541
Haswell-EX	22 nm	5/2015	18	18x2.5 MB	5700	662
Broadwell-EX	14 nm	6/2016	24	24x2.5 MB	7200	456

Table: Rising transistor counts in Intel's high-end 4S/8S server processors

Remarks to the rate of rising core counts in Intel's high-end 4S/8S server processors -3



Remarks to the rate of rising core counts in Intel's high-end 4S/8S server processors -3

- There are however, two anomalies to be mentioned.
- First, the Xeon 7100 has less cores than expected by the trend for Intel's early server processors, simply due to the fact that it was implemented with a 16 MB L3 cache.
- Second, the Xeon 7400 has already an L3 cache, nevertheless it fits to the trend line of L3 cache-less processors.

The reason is a roughly doubled die area with reduced feature size vs. the preceding Xeon 7300 (see Table above) which allows to implement a 16 MB L3 cache in addition to the 6 cores without restricting the core count.

• Subsequent server processors were then implemented nearly on the same die area as the Xeon 7400, (as indicated in Table 1), so these processors fit already to the linear trend with reduced slope.

Model	Technology	Date of introduction	Core count	L3 cache	Transistors (million)	Die area (mm²)
7300	65 nm	9/2007	2x2		582	286
7400	45 nm	8/2008	6	16 MB	1900	503
Nehalem-EX	45 nm	3/2010	8	8x3 MB	2300	513
Westmere-EX	32 nm	4/2011	10	10x3 MB	2600	512
Ivy Bridge-EX	22 nm	2/2014	15	15x2.5 MB	4310	541
Haswell-EX	22 nm	5/2015	18	18x2.5 MB	5700	662
Broadwell-EX	14 nm	6/2016	24	24x2.5 MB	7200	456

Table: Rising transistor counts in Intel's high-end 4S/8S server processors

b) The rate of rising memory transfer rates in Intel's DT and high-end server processors -1



The rate of rising mem. transfer rates in Intel's high-end 4S/8S server processors -2

- 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.

The requirement for scaling the memory bandwidth with the core count

As each core may operate independently from each other the memory bandwidth provided needs to be linearly scaled with the core count, i.e. a processor with n-times more cores as a reference platform requires n-times higher bandwidth as the reference platform.
Implication of the fact that memory transfer rates are rising slower than core counts in servers

- Over generations both the core count and the memory transfer rate are raising.
- But the memory transfer rate of the reference platform rises slower 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.

• In the chapter on high-end 4S/8S servers we will discuss possible avenues of coping with this point.

1.5 Evolution of tablet and smartphone processors

1.5 Evolution of tablet and smartphone processors



Changing the design paradigm for implementing 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.



Key requirements at the introduction of mobile, i.e. low power microarchitectures



- a) Low power CPUs need narrow microarchitectures
- Example: Width of ARM's 32-bit CPUs used in most tablets and smartphones [10]



By contrast:

Intel's and AMD's recent processor lines

• aims at high performance/power (in terms of GFLOPS/Watt)

accordingly

• have 4-wide microarchitectures, as the next example shows:





Example: Width of Intel's Core 2 to Haswell processors underlying servers to laptops [10]

Consequences for Intel and AMD-1

Intel's and AMD's traditional microarchitectures are not suited for mobile devices.

Both Intel and AMD conceived, designed and introduced new, narrow (e.g. 2-wide) low-power microarchitectures.

Emergence of Intel's Atom line aiming for tablets and smartphones (Based on [2])

I	A Pro	ocessors with	Built o Optimiz	n a Comr ed Micro	non Arch -Archite	itecture ctures	e (x86)	
Optimized Power Perfo Microarchitectu	ormance ire	11/2008	1/2010	1/2011	4/2012	6/2013	9/2014	8/20	15
(intel) (intel)		Nehalem 45 nm	Nehalem 45 nmWestmere Sandy Brid 32 nm		ge Ivy Bridge Haswell 22 nm		Broadwe	Broadwell Skylake 14 nm	
inside"	ide"	4-wide out-of-order	4-wide out-of-order	4-wide out-of-order	4-wide out-of-order	4-wide out-of-order	4-wide out-of-oi	e 5-v rd out-c	wide of-order
Low Power Microarchitecture	4/2008			1/2	012	9/2013	3/	/2015	4/2016
Atom Incide			Bonnell 45 nm		Saltwell 32 nm	Silvern 22 r	nont A nm	irmont 14	Goldmont nm
			2-wide n-order		2-wide in-order	2-wid out-of-o	e ź rder out	2-wide t-of-order	3-wide out-of-orde
	2008	2009	2010	2011	2012 2	013 2	014	2015	2016

Intel's Atom-based smartphone platforms (based on [276])



Emergence of AMD's Cat line aiming for tablets and smartphones (Based on [2])



b) Low power CPUs need to be operated at low base clock frequencies



Mobile CPUs

Relative low base clock frequency (typically 1-2 GHz)

($Dd = const x fc x V^2$, in addition higher fc requires about linearly higher higher V)

Example: The max. base frequency of Skylake models resulting in different TDPs and configurations (Based on data from [202])

TDP (W)	No. of cores	Graphics	No. of graphics EUs	eDRAM	Base frequency (GHz)
4.5	2	HD 515	18		1.2
15	2	HD 540	48	64 MB	2.2
15	2	HD 520	24		2.6
28	2	HD 550	48	64 MB	3.3
35	4	HD 530	24		2.8
45	4	HD 530	24		2.9
65	4	HD 530	24		3.4
91	4				4.2

Note that low TDP can be achieved first of all by reducing the core frequency and limiting the computer resources (cores, GPU EUs) provided.

Worldwide market share of smartphone and tablet application processors in 2015 (based on revenue) [217]

Smartphone application processors worldwide market share 2015 (revenue)				
Qualcomm (USA)	42 %			
Apple (USA)	21 %			
MediaTek (Taiwan)	19 %			
Samsung (S. Korea)				
Spreadtrum (China)				

Tablet application processors worldwide market share 2015 (revenue)				
Apple (USA)	31 %			
Qualcomm (USA)	16 %			
Intel (USA)	14 %			
MediaTek (Taiwan)				
Samsung (S. Korea)				

[Source: Related press releases of Strategy Analytics]

Intel's withdrawal from the smartphone and mobil market [216]

- In 4/2016 Intel announced their withdrawal from the mobil market.
- Intel's statement says:

"I can confirm that the changes included canceling the Broxton platform as well as SoFIA 3GX, SoFIA LTE and SoFIA LTE2 commercial platforms to enable us to move resources to products that deliver higher returns and advance our strategy.

These changes are effective immediately."

• At the same time Intel laid off about 12000 employees (~ 11 % of their workforce).

AMD's cancellation of their Cat line in 2015

- The last core of AMD's Cat line was the Puma+ core (launched in 6/2015 in the Carrizo-L APU).
- In AMD's 2016 Mobility roadmap there is no sign of an APU powered by the Puma+ core or a derivative of it belonging to the Cat line.
- Instead AMD has been placed emphasis on the Zen architecture that has been launched in 2017.

NVidia's leaving the smartphone and tablet market in June 2016 [81]

- In 6/2016 (at Computex) NVIDIA's CEO J. Huang declared the firm's leaving the smartphone and tablet market by saying:
- : "We are no longer interested in that market". He adds, "Anybody can build smartphones, and we're happy to enjoy these devices, but we'll let someone else build them".
- Instead NVIDIA became interested in designing in-car computers and car infotainment systems.

2. The Core 2 line

- 2.1 Introduction Introduction to the Core 2 line
- 2.2 Major innovations of the Core 2 line
- 2.3 Performance leadership change between Intel and AMD achieved by the Core 2
- 2.4 Die plot of the Core 2
- 2.5 Overview of Core 2 based processor lines

(Not discussed)

2.1 Introduction to the Core 2 line

2.1 Introduction to the Core 2 line -1

The Pentium 4 line was cancelled due to unmanageable high dissipation figures of its third core (the 90 nm Prescott core)

(caused primarily by the design philosophy of the line that preferred clock frequency over core efficiency for increasing performance).

For the development of the next processor line dissipation became the key design issue.

The next line became based primarily on the Pentium M – Intel's first mobile line since for its designs dissipation reduction was a key issue.

(The Pentium M line was a 32-bit line with 3 subsequent cores, designed at Intel's Haifa Research and Development Center) thus the Core 2 was developed also there.

The Haifa team had no experience with multithreading.

The Core microarchitecture does not support multithreading.

2.1 Introduction to the Core 2 line -2

- The Core 2 is Intel's 1. generation new microarchitecture, it is using 65 nm line width.
- First delivered in 6/2006.

1. gen.				2. gen.	3. gen.	4. gen.	5. gen.	
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	
тоск	ТІСК	тоск	ТІСК	тоск	ТІСК	тоск	ΤΙϹΚ	
(2006)	(2007)	(2008)	(2010)	(2011)	(2012)	(2013)	(2014)	

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

The Core 2 line -1 (based on [3])



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



2.2 Major innovations of the Core 2 line

- 2.2.1 4-wide core
- 2.2.2 128-bit SIMD and FP units
- 2.2.3 Smart L2 cache
- 2.2.4 Innovations in power management

2.2 Major innovations of the Core 2 line 2.2.1 4-wide core

It means 4-wide front end and retire unit.

This is the key benefit of the Core 2 family.

By contrast

both Intel's previous Pentium 4 family and AMD's K8 have only 3-wide cores, as the next Figures show.

Figure 2.2.1.1: Block diagram of Intel's Core 2 microarchitecture [3]



Figure 2.2.1.2: Block diagram of Intel's Pentium 4 microarchitecture [5]



Figure 2.2.1.3: Block diagram of AMD's Athlon microarchitecture [4]



2.2.2 128-bit wide SIMD and FP units

a) Overview of the x86 ISA extensions in Intel's processor lines

They aim at providing enhanced media support.

2.2.2 128-bit wide SIMD and FP units (2)

Overview of Intel's x86 ISA extensions (based on [18])



MultiMedia eXtensions

Streaming SIMD Extensions

Advanced Encryption Standard Advanced Vector Extension Fused Multiply-Add instr.

2.2.2 128-bit wide SIMD and FP units (3)

Register spaces provided by Intel's x86 ISA extensions (based on [18])



Illustration of the available SIMD register spaces of the MMX to SSE3/SSSE3 ISA extensions in 32- and 64-bit ISAs [162]

32-bit x86 CP	Us	64-bit x86 CPUs
MMX 64 bits	SSE, SSE2, SSE3 128-bits wide	SSE2, SSE3 128-bits wide
mm0	xmm0	xmm0
mm1	xmm1	xmm1
mm2	xmm2	xmm2
mm3	xmm3	xmm3
mm4	xmm4	xmm4
mm5	xmm5	xmm5
mm6	xmm6	xmm6
mm7	xmm7	xmm7
80-bit	8 new integer	xmm8
floating	registers in	xmm9
point registers	SSE chips	xmm10
. oglotoro		xmm11
		xmm12
		xmm13
		xmm14
		xmm15
		8 additional integer registers added in AMD64 and EM64T chips

2.2.2 128-bit wide SIMD and FP units (5)

Intel's x86 ISA extensions - the operations introduced (based on [18])



Overview of the Supplemental Streaming SIMD Extension (SSSE3) of the Core 2 ISA [161]

SSSE3 provide 32 instructions (represented by 14 mnemonics) to accelerate computations on packed integers.

These include:

- Twelve instructions that perform horizontal addition or subtraction operations (operations on adjacent numbers.
- Six instructions that evaluate absolute values.
- Two instructions that perform multiply and add operations and speed up the evaluation of dot products.
- Two instructions that accelerate packed-integer multiply operations and produce integer values with scaling.
- Two instructions that perform a byte-wise, in-place shuffle according to the second shuffle control operand.
- Six instructions that negate packed integers in the destination operand if the signs of the corresponding element in the source operand is less than zero.
- Two instructions that align data from the composite of two operands.

2.2.2 128-bit wide SIMD and FP units (7)

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



b) Contrasting the number of 128-bit wide execution units in Intel's Pentium 4, Core 2 and AMD's K8 -1

The Core 2 has 128-bit wide and more FP and SSE execution units than Pentium 4 or AMD's K8, as detailed below.

Core 2

- two 128-bit FP units
- three 128-bit SSE units

Pentium 4

- single 64-bit FP/SSE unit
- single 64-bit FP/SSE move unit

AMD K8

- two 64-bit FP/SSE units
- single 64-bit FP/SSE move unit

The next Figure demonstrates it.
b) Contrasting the number of 128-bit wide execution units in Intel's Pentium 4, Core 2 and AMD's K8 -2 [159], [160]

32 Entry Reservation Station Integer/FP Queue, Schedulers and Registers Port 2 Port 1 Port 0 Port 0 Port 1 64 bit 64 bit ALU 64 bit 128 bit 128 bit Fast 64 bit 128 bit Fast FP/SSE Shift ALU ALU SSE SSE SSE ALU FP Move ALU Execute Branch Rotate 128 bit Full 128 bit FMUL ALU FADD FDIV ADD ADD Shift Logic SSE ADD Store FP/SSE Move SSE MUL Branch FP/SSE Store All ALU SSE DIV Shift FXCH MMX **K8** Execution Units 8 Entry 8 Entry 8 Entry 12 Entry 12 Entry 12 Entry Integer, Memory Integer, Memory Integer, Memory FP RS FP RS FP RS Scheduler Scheduler Scheduler 120 Entry FP Reg File 64 bit 64 bit 64 bit ALU ALU ALU

IMUL

Core Execution Units

64 bit

FADD

SSE

64 bit

FMUL

SSE

64 bit

FMISC FP/SSE P4 Execution Units

Core 2's achieved performance boost for gaming apps vs. AMD's Athlon 64 FX60 [13]



2.2.3 Shared L2 cache

In the Core 2 Intel makes use a shared L2 cache instead of private L2 caches and designates it as smart L2 cache, as shown below.



Private

Pentium 4-based DCs

Core 2-based DCs

Figure 2.2.3.1: Core's shared L2 cache vs. Pentium 4's private L2 caches

Benefits of shared caches vs. private caches

- Dynamic cache allocation to the individual cores
- Data sharing (no replicated data)
- + 2x bandwidth to L1 caches.

Drawbacks of shared caches

Shared caches combine memory access patterns

Reduce the efficiency of hardware prefetching vs private caches.

Choice between shared and private caches

Design decision depending on

whether benefits or drawbacks dominate as far as performance is concerned.

Trend	
Core 2 prefers a shared L2 cache	 Nehalem prefers private L2 caches
POWER5 prefers shared L2 cache	 POWER6 prefers private L2 caches

This trend is due to the fact that efficient data prefetching became more important than dynamic cache allocation or data sharing.

Intensive use of hardware prefetchers -1 [9]

Remarks

- Intel's first on-die L2 cache debuted only about two years earlier (8/1998), in the second core of the Pentium II based Celeron line (called the Mendocino core, built on 250 nm technology, with an L2 size of 128 KB)^{1,2}.
- Intel's first hardware prefetcher appeared subsequently in the Pentium 4 family, associated with the L2 cache (one instruction and one data prefetcher (11/2000).

Principle of operation of the L2 hardware prefetcher

- it monitors data access patterns and prefetches data automatically into the L2 cache,
- it attempts to stay 256 bytes ahead of the current data access location,
- the prefetcher remembers the history of cache misses to detect concurrent, independent data streams that it tries to prefetch ahead of its use.
- ¹ The first Celeron model debuted in 4/1998, it was the low cost alternative to Pentium II, and was designated as Cavington.

It was manufactured on 250 nm, run at 266 MHz, did not include an L2 cache and had 7.5 mtrs.

The second core of the Celeron line was launched in 8/1998 already with an L2 cache, run at 266 and 300 MHz and incorporated 19,2 mtrs.

²Note: The debut of L2 was corrected in 29/10/2014.

Intensive use of hardware prefetchers -2 [11]

8 prefetchers in the two-core processor

- 2 data and 1 L1 instruction prefetchers per core, able to handle multiple simultaneous patterns.
- 2 prefetchers in the L2 cache tracking multiple access patterns per core.

Hardware prefetchers in the Core 2 microarchitecture



Figure 2.2.3.2: Hardware prefetchers within the Core 2 microarchitecture [11]

2.2.4 Innovations in power management

- Shutting off not needed functional units of the processor
- Platform Thermal Control

Shutting off not needed functional units of the processor by clock gating [11]

Instruction Fetch and PreDecode Even during periods of high performance 2M/4M Instruction, Queue execution, many parts shared L2 of the chip core can be uCode Devode Cache shut off. ROM Example could be a up to **Rename/Alloc** SW memory initialization 10.4 Gb/s executing from front **FSB** end with IQ operating **Retirement Unit** as loop cache. (ReOrder Buffer) Green units not needed for the given operation, Schedulers they can be shut off. ALU ALU ALU Branch Intel Developer FP FP Store Load FP L1 D-Cache and D-TLB

65

Platform Thermal Control [11], [20]

 Processor provides its temperature reading over a multi drop single wire bus allowing efficient platform thermal control



Possible solution for the Platform Thermal Control Manager [88]



Figure 2.2.4.1: The aSC7621 hardware monitor with fan control and PECI from Andigilog

2.3 Performance leadership change between Intel and AMD achieved by the Core 2

Performance leadership change between Intel and AMD achieved by the Core 2

- In 2003 AMD introduced their K8-based processors implementing
 - the 64-bit x86 ISA and
 - the direct connect architecture concept, that includes
 - integrated memory controllers and
 - high speed point-to-point serial buses (the HyperTransport bus) used to connect processors to processors and processors to south bridges.
- AMD's K8-based processors became the performance leader, first of all on the DP and MP server market, where the 64-bit direct connect architecture has clear benefits vs Intel's 32-bit Pentium 4 based processors using shared FSBs to connect processors to north bridges.

Example 1: DP web-server performance comparison (2003)



Figure 2.3.1: DP web server performance comparison: AMD Opteron 248 vs. Intel Xeon 2.8 [6]

2.3 Performance leadership change between Intel and AMD (3)

Example 2: Summary assessment of extensive benchmark tests contrasting dual Opterons vs dual Xeons (2003) [7]

"In the extensive benchmark tests under Linux Enterprise Server 8 (32 bit as well as 64 bit), the **AMD Opteron made a good impression**.

Especially in the server disciplines, the benchmarks (MySQL, Whetstone, ARC 2D, NPB, etc.) show quite clearly that the **Dual Opteron puts the Dual Xeon in its place**".

2.3 Performance leadership change between Intel and AMD (4)

- This situation has completely changed in 2006 when Intel introduced their Core 2 microarchitecture,
- The Core 2 has
 - > a 4-wide front-end and retire unit compared to the 3-wide K8,
 - > two 128-bit FP units and three 128-bit SSE units compared to
 - > two 64-bit FP/SSE units and a single 64-bit FP/SSE move unit available in the K8.
- This and further enhancements of the Core microarchitecture, detailed subsequently, resulted in record breaking performance figures.

→ Intel regained performance leadership vs AMD.

Example: DP web-server performance comparison (2006)

Webserver Performance										
	MSI K2-102A2M Opteron 275	MSI K2-102A2M Opteron 280		Opteron 280 vs. Opteron 275	Extrapolated Xeon 51 Opteron 3 GHz 3 GHz		60			
Jsp - Peak	144	154		7%	182	230				
AMP - Peak	984	1042		6%	1178	1828				

Jsp: Java Server Page performance AMP: Apache/MySQL/PHP

> Figure 2.2.1.8: DP web server performance comparison: AMD Opteron 275/280 vs. Intel Xeon 5160 [8]

Remark

Both web-server benchmark results were published from the same source (AnandTech)

2.4 Die plot of the Core 2

2.4 Die plot of the Core 2 (1)

2.4 Die plot of the Core 2 [163]





2.5 Overview of Core 2 based processor lines

2.5 Overview of Core 2 based processor lines

Mobiles

T5xxx/T7xxx, 2C ,Merom 8/2006

Desktops

E6xxx/E4xxx, 2C, Core 2 Duo, (Conroe) 7/2006 X6800 Core 2 Extreme, 2C, (Conroe) 7/2006 E6xxx/E4xxx, Core 2 Duo, 2C, (Allendale) 1/2007 Q6xxx Core 2 Quad, 2x2C (Kentsfield) (2xConroe) 1/2007 QX6xxx Core 2 Extreme Quad, 2x2C, (Kentsfield XE) 11/2006

Servers

UP-Servers

30xx, 2C, Conroe, 9/2006 30xx, 2C, Allendale 1/2007 32xx, 2x2C, Kentsfield (2xConroe) 1/2007

DP-Servers

51xx, 2C Woodcrest, 6/2006 53xx, 2x2, Clowertown, (2xWoodcrest) 11/2006

MP-Servers

72xx, 2C, Tigerton DC, (2xMP-enhanced SC Woodcrest) 9/2007 73xx, 2x2C, Tigerton QC, (2xMP-enhanced DC Woodcrest) 9/2007

Example Core 2 system architecture [223]







Számítógép architektúrák III. BSc 2009. december 22.

