**INTEL® 64 AND IA-32 ARCHITECTURES**

**IA-32 Architecture**refers to systems based on 32-bit processors generally compatible with the Intel Pentium® II processor, (for example, Intel® Pentium® 4 processor or Intel® Xeon® processor), or processors from other manufacturers supporting the same instruction set, running a 32-bit operating system.

**Intel® 64 Architecture**refers to systems based on IA-32 architecture processors which have 64-bit architectural extensions, for example, Intel® CoreTM2 processor family), running a 64-bit operating system such as Microsoft Windows XP\* Professional x64 Edition or Microsoft Windows Vista\* x64. If the system is running a 32-bit version of the Windows operating system, then IA-32 architecture applies instead. Systems based on AMD\* processors running a 64-bit version of Windows are also supported by Intel compilers for Intel® 64 architecture applications.

**IA-64 Architecture**refers to systems based on the Intel® Itanium® processor running a 64-bit operating system.

**BRIEF HISTORY OF INTEL® 64 AND IA-32 ARCHITECTURE**

The following sections provide a summary of the major technical evolutions from IA-32 to Intel 64 architecture: starting from the Intel 8086 processor to the latest Intel® Core® 2 Duo, Core 2 Quad and Intel Xeon processor 5300 and 7300 series. Object code created for processors released as early as 1978 still executes on the latest processors in the Intel 64 and IA-32 architecture families.

**4-bit Processors(1969)**

Intel 4004 was the first commercially available single-chip microprocessor in history. It was a 4-bit CPU designed for usage in calculators, or, as we say now, designed for "embedded applications". Clocked at 740 KHz, the 4004 executed up to 92,000 single word instructions per second, could access 4 KB of program memory and 640 bytes of RAM.

**16-bit Processors and Segmentation (1978)**

The 8088 is similar to the 8086 except it has an 8-bit external data bus. The 8086/8088 introduced segmentation to the IA-32 architecture. With segmentation, a 16-bit segment register contains a pointer to a memory segment of up to 64 KBytes.

**The Intel® 286 Processor (1982)**

The Intel 286 processor introduced protected mode operation into the IA-32 architecture. Protected mode uses the segment register content as selectors or pointers into descriptor tables

**The Intel386™ Processor (1985)**

The Intel386 processor was the first 32-bit processor in the IA-32 architecture family. It introduced 32-bit registers for use both to hold operands and for addressing. The lower half of each 32-bit Intel386 register retains the properties of the 16-bit registers of earlier generations, permitting backward compatibility. The processor also provides a virtual-8086 mode that allows for even greater efficiency when executing programs created for 8086/8088 processors.

**The Intel486™ Processor (1989)**

The Intel486™ processor added more parallel execution capability by expanding the Intel386 processor’s instruction decode and execution units into five pipelined stages

**The Intel® Pentium® Processor (1993)**

The introduction of the Intel Pentium processor added a second execution pipeline to achieve superscalar performance (two pipelines, known as u and v, together can execute two instructions per clock.

**2.1.6 The P6 Family of Processors (1995-1999)**

The P6 family of processors was based on a superscalar microarchitecture. Members of this family include the following:

• The Intel Pentium Pro processor is three-way superscalar. Using parallel processing techniques, the processor is able on average to decode, dispatch, and complete execution of (retire) three instructions per clock cycle.

• The Intel Pentium II processor added Intel MMX technology to the P6 family processors along with new packaging and several hardware enhancements.

• The Pentium II Xeon processor combined the premium characteristics of previous generations of Intel processors. This includes: 4-way, 8-way (and up) scalability and a 2 MByte 2nd-Level cache running on a fullfrequency backside bus.

• The Intel Celeron processor family focused on the value PC market segment. Its introduction offers an integrated 128 KBytes of Level 2 cache and a plastic pin grid array (P.P.G.A.) form factor to lower system design cost.

• The Intel Pentium III processor introduced the Streaming SIMD Extensions (SSE) to the IA-32 architecture.

• The Pentium III Xeon processor extended the performance levels of the IA-32 processors with the enhancement of a full-speed, on-die, and Advanced Transfer Cache.

**The Intel® Pentium® 4 Processor Family (2000-2006)**

The Intel Pentium 4 processor family is based on Intel NetBurst microarchitecture and introduced Streaming SIMD Extensions 2 (SSE2), Streaming SIMD Extensions 3 (SSE3). Intel 64 architecture was introduced in the Intel Pentium 4 Processor Extreme Edition supporting Hyper-Threading Technology and in the Intel Pentium 4 Processor 6xx and 5xx sequences. Intel® Virtualization Technology (Intel® VT) was introduced in the Intel Pentium 4 processor 672 and 662.

**The Intel® Xeon® Processor (2001- 2007)**

Intel Xeon processors (with exception for dual-core Intel Xeon processor LV, Intel Xeon processor 5100 series) are based on the Intel NetBurst microarchitecture. The Intel Xeon processor MP introduced support for Intel® Hyper-Threading Technology. The 64-bit Intel Xeon processor 3.60 GHz (with an 800 MHz System Bus) was used to introduce Intel 64 architecture. The Intel Xeon processor 3000 series are also based on Intel Core microarchitecture. The Intel Xeon processor 5300 series introduces four processor cores in a physical package, they are also based on Intel Core microarchitecture.

**The Intel® Pentium® M Processor (2003-2006)**

This family is designed for extending battery life and seamless integration with platform innovations that enable new usage models. Its enhanced microarchitecture includes: • Support for Intel Architecture with Dynamic Execution

**The Intel® Pentium® Processor Extreme Edition (2005)**

The Intel Pentium processor Extreme Edition introduced dual-core technology. This technology provides advanced hardware multi-threading support. The processor is based on Intel NetBurst microarchitecture and supports SSE, SSE2, SSE3, Hyper-Threading Technology, and Intel 64 architecture.

**The Intel® Core™ Duo and Intel® Core™ Solo Processors (2006-2007)**

Its enhanced microarchitecture includes: • Intel® Smart Cache which allows for efficient data sharing between two processor cores • Improved decoding and SIMD execution

**The Intel® Xeon® Processor 5100, 5300 Series and Intel® Core™2 Processor Family (2006)**

They are based on the high-performance, power-efficient Intel® Core microarchitecture built on 65 nm process technology.

**The Intel® Xeon® Processor 5200, 5400, 7400 Series and Intel® Core™2 Processor Family (2007)**

They support Intel 64 architecture and are based on the Enhanced Intel® Core microarchitecture using 45 nm process technology.

**The Intel® Atom™ Processor Family (2008)**

The first generation of Intel® AtomTM processors are built on 45 nm process technology are based on a new microarchitecture, Intel® AtomTM microarchitecture, which is optimized for ultra low power devices with support for Intel® 64 Architecture (excluding Intel Atom processor Z5xx Series)

**The Intel® Atom™ Processor Family Based on Silvermont Microarchitecture (2013)**

Intel Atom Processor C2xxx, E3xxx, S1xxx series are based on the Silvermont microarchitecture. Processors based on the Silvermont microarchitecture supports instruction set extensions up to and including SSE4.2, AESNI, and PCLMULQDQ.

**The Intel® Core™i7 Processor Family (2008)**

The Intel Core i7 processor 900 series support Intel 64 architecture based on Intel® microarchitecture code name Nehalem using 45 nm process technology

**The Intel® Xeon® Processor 7500 Series (2010)**

The Intel Xeon processor 7500 and 6500 series are based on Intel microarchitecture code name Nehalem using 45 nm process technology with up to eight cores per physical processor package

**2010 Intel® Core™ Processor Family (2010)**

2010 Intel Core processor family spans Intel Core i7, i5 and i3 processors. They are based on Intel® microarchitecture code name Westmere using 32 nm process technology.

**The Intel® Xeon® Processor 5600 Series (2010)**

The Intel Xeon processor 5600 series are based on Intel microarchitecture code name Westmere using 32 nm process technology. They support up to six cores per physical processor package and up to 12 MB enhanced Intel® Smart Cache.

**The Second Generation Intel® Core™ Processor Family (2011)**

The Second Generation Intel Core processor family spans Intel Core i7, i5 and i3 processors based on the Sandy Bridge microarchitecture. They are built from 32 nm process technology, have innovative features and provide support for multiple sockets.

**The Third Generation Intel® Core™ Processor Family (2012)**

The Third Generation Intel Core processor family spans Intel Core i7, i5 and i3 processors based on the Ivy Bridge microarchitecture and provide support for multiple sockets.

**The 4th generation Intel® Core™ processors and the Intel® Xeon® processor E3-1200 v3 product family** are based on Intel® microarchitecture code name Haswell. The Intel® Xeon® processor E5 26xx v3 family is based on the Haswell-E microarchitecture, supports Intel 64 architecture and multiple physical processor packages in a platform.

**The Intel® Core™ M processor family and 5th generation Intel® Core™ processors** are based on the Intel® microarchitecture code name Broadwell and support Intel 64 architecture.

**The 6th generation Intel® Core™ processors** are based on the Intel® microarchitecture code name Skylake and support Intel 64 architecture.

Process-Architecture-Optimization (PAO) - Intel

**Process-Architecture-Optimization** is a development model introduced by [Intel](https://en.wikichip.org/wiki/Intel) for their mainstream microprocessors in [2016](https://en.wikichip.org/wiki/2016) following the phase-out of their [Tick-Tock](https://en.wikichip.org/wiki/intel/tick-tock) model. The change is a result of the increase in cost and complexity of advancing lithography processes in the past decade. Under the new model the amount of time utilized for any given process technology is lengthened as [Moore's Law](https://en.wikichip.org/wiki/Moore%27s_Law) increases in complexity with smaller [nodes](https://en.wikichip.org/wiki/technology_node).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Before 2016 | | 2016 onwards | | |
| TICK-TOCK MODEL | | PROCESS ARCHITECTURE OPTIMIZATION MODEL | | |
| TICK | TOCK | PROCESS | ARCHITECTURE | OPTIMIZATION |
| PROCESS | ARCHITECTURE |

Under the Process-Architecture-Optimization Model:

* **Process** - With each process, Intel advances their manufacturing [process technology](https://en.wikichip.org/wiki/process_technology) in line with [Moore's Law](https://en.wikichip.org/wiki/Moore%27s_Law). Each new process introduces higher transistor density and a generally a plethora of other advantages such as higher performance and lower power consumption. During a "process", Intel retrofits their [previous](https://en.wikichip.org/wiki/intel/microarchitectures" \o "intel/microarchitectures)[microarchitecture](https://en.wikichip.org/wiki/microarchitecture) to the new process which inherently yields better performance and energy saving. During a "process", usually just a few features and improvements are introduced.
* **Architecture** - With each architecture, Intel uses the their latest manufacturing [process technology](https://en.wikichip.org/wiki/process_technology) from their "process" to manufacture a newly designed [microarchitecture](https://en.wikichip.org/wiki/microarchitecture). The new microarchitecture is designed with the new process in mind and typically introduces Intel's newest big features and functionalities. New [instructions](https://en.wikichip.org/w/index.php?title=instruction_set&action=edit&redlink=1) are often added during this cycle stage.
* **Optimization** - With each optimization, Intel improves upon their [previous](https://en.wikichip.org/wiki/intel/microarchitectures) microarchitecture by introducing incremental improvements and enhancements without introducing any large charges. Additionally the process itself enjoys various refinements as it matures. (For example with [Kaby Lake](https://en.wikichip.org/wiki/intel/microarchitectures/kaby_lake" \o "intel/microarchitectures/kaby lake), an optimized process called "14 nm+" is used. The enhanced process had a number of transistor-level modifications done to it (e.g. taller fins) allowing for higher frequency at identical voltage levels.)

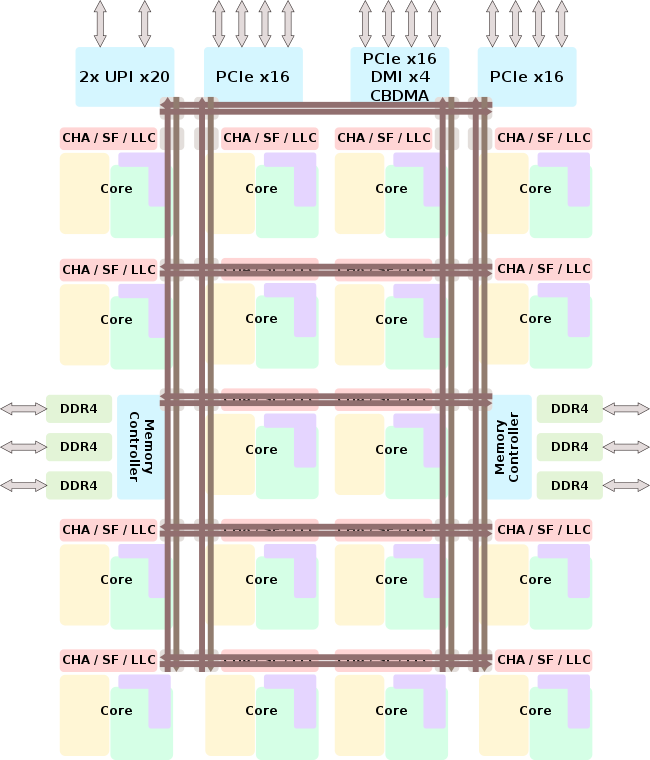
Recent ARCHITECTURES

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Codename | Haswell | Broadwell | Skylake | Kabylake | Coffeelake |
| L0 MicroOp Cache |  |  | 1,536 µOPs | 1,536 µOPs | 1,536 µOPs |
|  | Associativity |  | 8-way set associative | 8-way set associative | 8-way set associative |
|  | Sets |  | 32 | 32 | 32 |
|  | Block size |  | 6-µOP | 6-µOP | 6-µOP |
|  |  |  | statically divided between threads, per core, inclusive with L1I | statically divided between threads, per core, inclusive with L1I | statically divided between threads, per core, inclusive with L1I |
|  |  |  |  |  |  |
| L1 Instruction Cache | Cache Size | 32KiB/core | 32KiB/core | 32KiB/core | 32KiB/core |
|  | Associativity | 8-way set associative | 8-way set associative | 8-way set associative | 8-way set associative |
|  | Sets |  | 64 | 64 | 64 |
|  | Block size | 64B | 64B | 64B | 64B |
|  |  | Per core | Shared 2Thread/core | Shared 2Thread/core | Shared 2Thread/core |
|  |  |  |  |  |  |
| L1 Data Cache | Cache Size | 32KiB/core | 32KiB/core | 32KiB/core | 32KiB/core |
|  | Associativity | 8-way set associative | 8-way | 8-way | 8-way |
|  | Sets |  | 64 | 64 | 64 |
|  | Block size | 64B | 64B | 64B | 64B |
|  |  | Per core | Shared 2Thread/core | Shared 2Thread/core | Shared 2Thread/core |
|  | Fastest load-to-use |  | 4 cycles | 4 cycles | 4 cycles |
|  | Complex addresses |  | 5 cycles | 5 cycles | 5 cycles |
|  | Load Bandwidth |  | 64B/cycle | 64B/cycle | 64B/cycle |
|  | Store Bandwidth |  | 32B/cycle | 32B/cycle | 32B/cycle |
|  | Update policy |  | Write back | Write back | Write back |
|  |  |  |  |  |  |
| L2 Unified | Cache Size | 256 KiB/core | 256KiB/core | 256KiB/core | 256KiB/core |
|  | Associativity | 8-way set associative | 4-way set associative | 4-way set associative | 4-way set associative |
|  | Sets |  | 1024 | 1024 | 1024 |
|  | Block size | 64B | 64B | 64B | 64B |
|  |  | Per core | Non-Inclusive | Non-Inclusive | Non-Inclusive |
|  | Fastest load-to-use |  | 12 cycles | 12 cycles | 12 cycles |
|  | Load Bandwidth |  | 64B/cycle | 64B/cycle | 64B/cycle |
|  | Update policy | Write back | Write back | Write back | Write back |
|  |  |  |  |  |  |
| L3 Unified | Cache Size | 1.5 - 3 MiB per core | 2MiB/core | 2MiB/core | 2MiB/core |
|  |  |  | shared across all cores | shared across all cores | shared across all cores |
|  | Associativity | 16-20 -way set associative | 16-way set associative | 16-way set associative | 16-way set associative |
|  | Block size | 64B | 64B | 64B | 64B |
|  |  |  | Inclusive | Inclusive | Inclusive |
|  | Fastest load-to-use |  | 42 cycles | 42 cycles | 42 cycles |
|  | Load Bandwidth |  | 32B/cycle | 32B/cycle | 32B/cycle |
|  | Store Bandwidth |  | 32B/cycle | 32B/cycle | 32B/cycle |
|  | Update policy | Write-back policy | Write back | Write back | Write back |
|  |  |  |  |  |  |
| Side Cache | Cache Size | 128 MiB | 64MiB/core | 64MiB/core | 64MiB/core |
|  | eDRAM | shared with GPU | 128MiB | 128MiB | 128MiB |
|  |  |  | per package | per package | per package |
|  |  | only Iris Pro GPU | only Iris Pro GPU | only Iris Pro GPU | only Iris Pro GPU |
|  | Load Bandwidth |  | 32B/cycle | 32B/cycle | 32B/cycle |
|  | Store Bandwidth |  | 32B/cycle | 32B/cycle | 32B/cycle |
|  | Update policy |  | Write back | Write back | Write back |
|  |  |  |  |  |  |
| System DRAM | No of Channel |  | 2 | 2 | 2 |
|  | size |  | 8 B/cycle/channel | 8 B/cycle/channel | 8 B/cycle/channel |
|  |  |  | 42 cycles + 51 ns latency | 42 cycles + 51 ns latency | 42 cycles + 51 ns latency |
| TLB |  |  |  |  |  |
| L1 ITLB | 4 KiB page | 128 entries | 128 entries | 128 entries | 128 entries |
|  |  | 4-way set associative | 8-way set associative | 8-way set associative | 8-way set associative |
|  |  | dynamic partition | dynamic partitioning | dynamic partitioning | dynamic partitioning |
|  | 2 MiB / 4 MiB page | 8 entries per thread | 8 entries per thread | 8 entries per thread | 8 entries per thread |
|  |  | fully associative | fully associative | fully associative | fully associative |
|  |  | Duplicated for each thread | Duplicated for each thread | Duplicated for each thread | Duplicated for each thread |
|  |  |  |  |  |  |
| L1 DTLB | 4 KiB page | 64 entries | 64 entries | 64 entries | 64 entries |
|  |  | 4-way set associative | 4-way set associative | 4-way set associative | 4-way set associative |
|  |  | fixed partition | fixed partition | fixed partition | fixed partition |
|  | 2 MiB / 4 MiB page | 32 entries | 32 entries | 32 entries | 32 entries |
|  |  | 4-way set associative | 4-way set associative | 4-way set associative | 4-way set associative |
|  |  | fixed partition | fixed partition | fixed partition | fixed partition |
|  | 1GiB page | 4 entries | 4 entries | 4 entries | 4 entries |
|  |  | fully associative | fully associative | fully associative | fully associative |
|  |  | fixed partition | fixed partition | fixed partition | fixed partition |
|  |  |  |  |  |  |
| L2 TLB | 4 KiB + 2 MiB page | 1536 entries | 1536 entries | 1536 entries | 1536 entries |
|  |  | 6-way set associative | 12-way set associative | 12-way set associative | 12-way set associative |
|  |  | Shared | fixed partition | fixed partition | fixed partition |
|  | 1 GiB page | 16 entries | 16 entries | 16 entries | 16 entries |
|  |  | 4-way set associative | 4-way set associative | 4-way set associative | 4-way set associative |
|  |  | Shared | fixed partition | fixed partition | fixed partition |

**SKYLAKE architecture**

**2.1 THE SKYLAKE MICROARCHITECTURE**

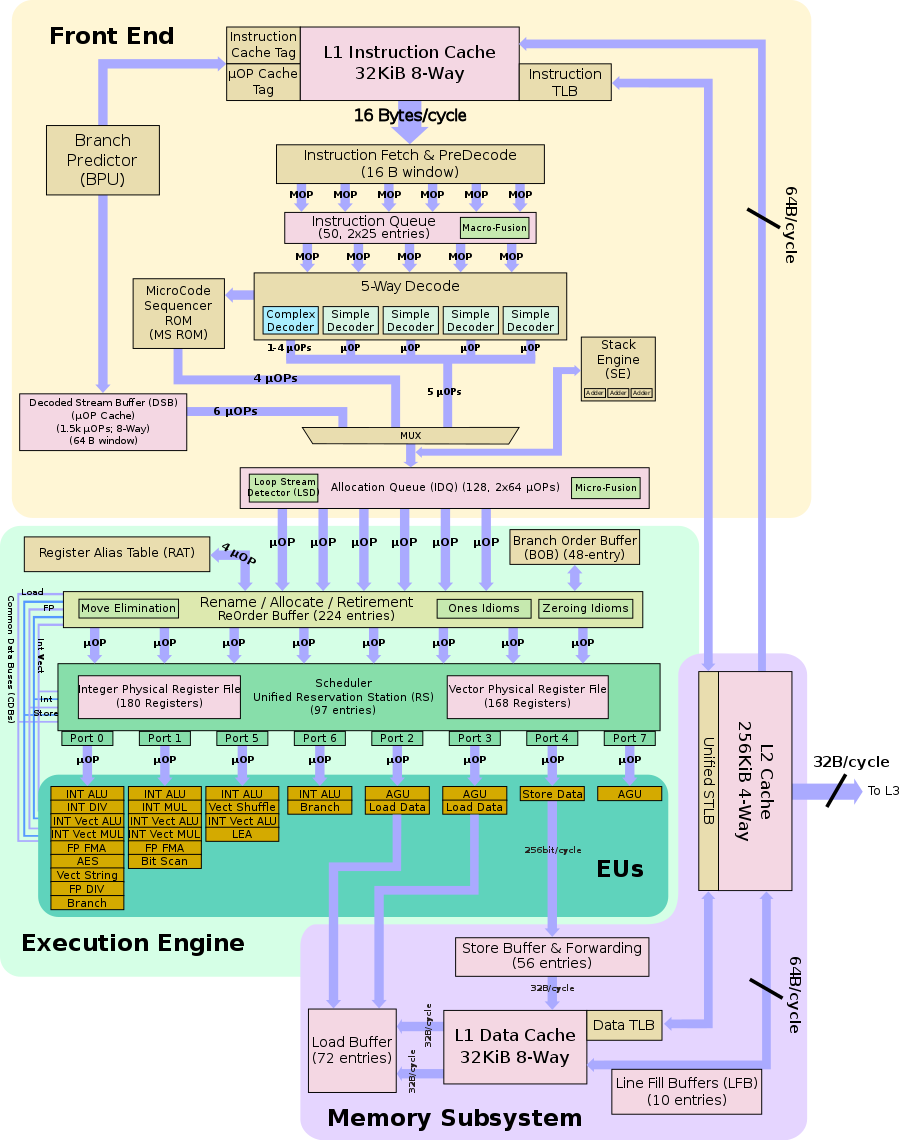
The Skylake microarchitecture builds on the successes of the Haswell and Broadwell microarchitectures.



* **CHA** - Caching and Home Agent
* **SF** - Snooping Filter

#### Entire SoC Overview (quad)

#### Entire SoC Overview (dual)



**Individual Core**

## Architecture

Overall Skylake builds upon Intel's previous microarchitecture, [Broadwell](https://en.wikichip.org/wiki/intel/microarchitectures/broadwell" \o "intel/microarchitectures/broadwell), but includes a wider and more beefed up front end, more optimized execution engine, and numerous other enhancements. Intel designed Skylake to encompass a wide range of devices and applications with a large emphasis on mobile with models ranging from as low as 4.5 W to as high as 100 W.

### Key changes from [Broadwell](https://en.wikichip.org/wiki/intel/microarchitectures/broadwell" \o "intel/microarchitectures/broadwell)

* 8x performance/watt over [Nehalem](https://en.wikichip.org/wiki/intel/microarchitectures/nehalem) (Up from 3.5x in [Haswell](https://en.wikichip.org/wiki/intel/microarchitectures/haswell" \o "intel/microarchitectures/haswell))
* Mainstream chipset
  + [Lynx Point](https://en.wikichip.org/w/index.php?title=intel/chipsets/lynx_point&action=edit&redlink=1) → [Sunrise Point](https://en.wikichip.org/w/index.php?title=intel/chipsets/sunrise_point&action=edit&redlink=1)
  + Bus/Interface to Chipset
    - [DMI 3.0](https://en.wikichip.org/w/index.php?title=intel/direct_media_interface&action=edit&redlink=1) (from 2.0)
      * Skylake S and Skylake H cores, connected by 4-lane DMI 3.0
      * [Skylake Y](https://en.wikichip.org/wiki/intel/cores/skylake_y) and Skylake U cores have chipset in the same package (simplified [OPIO](https://en.wikichip.org/w/index.php?title=intel/on_package_i/o&action=edit&redlink=1))
      * Increase in transfer rate from 5.0 GT/s to 8.0 GT/s (~3.93GB/s up from 2GB/s) per lane
      * Limits motherboard trace design to 7 inches max from (down from 8) from the CPU to chipset
  + PCIe & DMI upgraded to Gen3
  + More I/O (configurable as PCIe/SATA/USB3)
  + Lower-power I/O (eMMC, UFS, SDXC)
  + CSI-2 for the integrated IPU (mobile SKUs)
  + Intel Sensor Solution Hub integrationLarger Line Fill Buffer?
* [System Agent](https://en.wikichip.org/w/index.php?title=System_Agent&action=edit&redlink=1)
  + New Image Processing Unit (IPU)
    - Incorporates an [image signal processor](https://en.wikichip.org/w/index.php?title=image_signal_processor&action=edit&redlink=1) (ISP)
    - Mobile client models only
  + Can now have its own variable voltage and frequency
* Core
  + Front End
    - Larger legacy pipeline delivery (5 µOPs, up from 4)
      * Another simple decoder has been added.
    - Allocation Queue (IDQ)
      * Larger delivery (6 µOPs, up from 4)
      * 2.28x larger buffer (64/thread, up from 56)
      * Partitioned for each active threads (from unified)
    - Improved [branch prediction unit](https://en.wikichip.org/w/index.php?title=branch_prediction_unit&action=edit&redlink=1)
      * reduced penalty for wrong direct jump target
      * No specifics were disclosed
    - µOP Cache
      * instruction window is now 64 Bytes (from 32)
      * 1.5x bandwidth (6 µOPs/cycle, up from 4)
  + Execution Engine
    - Larger [re-order buffer](https://en.wikichip.org/w/index.php?title=re-order_buffer&action=edit&redlink=1) (224 entries, up from 192)
    - Larger scheduler (97 entries, up from 64)
      * Larger Integer Register File (180 entries, up from 168)
      * Larger Retire (4 µOPs/cycle/thread, up from 4 µOPs/cycle/core)?
  + Memory Subsystem
    - Larger store buffer (56 entries, up from 42)
    - [L2$](https://en.wikichip.org/w/index.php?title=L2$&action=edit&redlink=1) was changed from 8-way to 4-way set associative
    - Page split load penalty reduced 20-fold
    - Larger Write-back buffer
* Memory
  + Support for faster DDR-2400 memory
  + [L3$](https://en.wikichip.org/w/index.php?title=L3$&action=edit&redlink=1) re-gained 512 KiB/core (See [§eDRAM architectural changes](https://en.wikichip.org/wiki/intel/microarchitectures/skylake_(client)#eDRAM_architectural_changes) for the reason)
  + A new coherent [cache](https://en.wikichip.org/w/index.php?title=cache&action=edit&redlink=1) fabric implementation
    - doubles the throughput of the last level cache (LLC, L3$ in this case) miss handling
    - 50% improvement in bandwidth/watt
    - new [eDRAM](https://en.wikichip.org/w/index.php?title=eDRAM&action=edit&redlink=1" \o "eDRAM (page does not exist)) cache architecture for higher bandwidth
* TLBs
  + ITLB
    - 4 KiB page translations was changed from 4-way to 8-way associative
  + STLB
    - 4 KiB + 2 MiB page translations was changed from 6-way to 12-way associative
* Electrical
  + The fully integrated voltage regulator (FIVR) is moved back to the motherboard
    - Originally intended to be a cost-cutting measure by moving the FIVR on-die as well as making it more efficient, the move resulted in unintentionally making the FIVR the limiting factor when it came to overclocking.
  + DMI/PEG are now on a discrete clock domain with BCLK sitting on its own domain with full-range granularity (1 MHz intervals)
* Testability
  + New support for [Direct Connect Interface](https://en.wikichip.org/w/index.php?title=intel/direct_connect_interface&action=edit&redlink=1) (DCI), a new debugging transport protocol designed to allow debugging of closed cases (e.g. laptops, embedded) by accessing things such as [JTAG](https://en.wikichip.org/w/index.php?title=JTAG&action=edit&redlink=1" \o "JTAG (page does not exist))through any [USB 3](https://en.wikichip.org/w/index.php?title=USB_3&action=edit&redlink=1) port.
* [Gen 9 GPUs](https://en.wikichip.org/wiki/intel/microarchitectures/gen9)
  + Improved underlying implementation of the memory QoS for higher resolution displays and the integrated [image signal processor](https://en.wikichip.org/w/index.php?title=image_signal_processor&action=edit&redlink=1) (ISP)
    - Allow for higher concurrent bandwidth
  + Skylake retires VGA support, multi-monitor support for up to 3 displays via HDMI 1.4, DP 1.2, and eDP 1.3 interfaces.
  + Direct X 12, OpenCL 2.0, OpenGL 4.4
  + Up to 24 EUs GT2 (same as [Haswell](https://en.wikichip.org/wiki/intel/microarchitectures/haswell" \o "intel/microarchitectures/haswell)); 48 EUs for GT3, and up to 72 EUs on [Iris Pro Graphics](https://en.wikichip.org/wiki/intel/iris_pro_graphics)
    - 1,152 GFLOPS @ 1 GHz

#### CPU changes[[edit](https://en.wikichip.org/w/index.php?title=intel/microarchitectures/skylake_(client)&action=edit&section=10" \o "Edit section: CPU changes)]

* Most ALU operations have 4 op/cycle 1 for 8 and 32-bit registers. 64-bit ops are still limited to 3 op/cycle. (16-bit throughput varies per op, can be 4, 3.5 or 2 op/cycle).
* MOVSX and MOVZX have 4 op/cycle throughput for 16->32 and 32->64 forms, in addition to Haswell's 8->32, 8->64 and 16->64 bit forms.
* ADC and SBB have throughput of 1 op/cycle, same as Haswell.
* Vector moves have throughput of 4 op/cycle (move elimination).
* Not only zeroing vector vpXORxx and vpSUBxx ops, but also vPCMPxxx on the same register, have throughput of 4 op/cycle.
* Vector ALU ops are often "standardized" to latency of 4. for example, vADDPS and vMULPS used to have L of 3 and 5, now both are 4.
* Fused multiply-add ops have latency of 4 and throughput of 0.5 op/cycle.
* Throughput of vADDps, vSUBps, vCMPps, vMAXps, their scalar and double analogs is increased to 2 op/cycle.
* Throughput of vPSLxx and vPSRxx with immediate (i.e. fixed vector shifts) is increased to 2 op/cycle.
* Throughput of vANDps, vANDNps, vORps, vXORps, their scalar and double analogs, vPADDx, vPSUBx is increased to 3 op/cycle.
* vDIVPD, vSQRTPD have approximately twice as good throughput: from 8 to 4 and from 28 to 12 cycles/op.
* Throughput of some MMX ALU ops (such as PAND mm1, mm2) is decreased to 2 or 1 op/cycle (users are expected to use wider SSE/AVX registers instead).

#### New instructions[[edit](https://en.wikichip.org/w/index.php?title=intel/microarchitectures/skylake_(client)&action=edit&section=11" \o "Edit section: New instructions)]

*See also:*[*Server Skylake's New instructions*](https://en.wikichip.org/wiki/intel/microarchitectures/skylake_(server)#New_instructions)

Skylake introduced a number of [new instructions](https://en.wikichip.org/wiki/x86/extensions):

* [SGX1](https://en.wikichip.org/w/index.php?title=x86/sgx1&action=edit&redlink=1) - Software Guard Extensions, Version 1
* [MPX](https://en.wikichip.org/w/index.php?title=x86/mpx&action=edit&redlink=1) -Memory Protection Extensions
* [XSAVEC](https://en.wikichip.org/w/index.php?title=x86/xsavec&action=edit&redlink=1) - Save processor extended states with compaction to memory
* [XSAVES](https://en.wikichip.org/w/index.php?title=x86/xsaves&action=edit&redlink=1) - Save processor supervisor-mode extended states to memory.
* [CLFLUSHOPT](https://en.wikichip.org/w/index.php?title=x86/clflushopt&action=edit&redlink=1) - Flush & Invalidates memory operand and its associated cache line (All L1/L2/L3 etc..)

## Architecture[[edit](https://en.wikichip.org/w/index.php?title=intel/microarchitectures/kaby_lake&action=edit&section=8" \o "Edit section: Architecture)]

*See also: [Skylake § Key changes from Broadwell](https://en.wikichip.org/wiki/intel/microarchitectures/skylake" \l "Key_changes_from_Broadwell" \o "intel/microarchitectures/skylake)*

While there is no change in pure IPC over Skylake and the actual microarchitecture is largely the same, Intel introduced a number of enhancements in Kaby Lake. Note that because of the improvements done to the process and the uplift in binning, it is the mostly the ultra-low power (i.e. mobile) processors that will see the most substantial gain. Likewise, the high-end models will see very little gain. The enhanced manufacturing process allowed Kaby Lake chips to be highly [overclockable](https://en.wikichip.org/w/index.php?title=overclockable&action=edit&redlink=1" \o "overclockable (page does not exist)) with models such as the [Core i7-7700K](https://en.wikichip.org/wiki/Core_i7-7700K) capable of comfortably reaching 5 GHz for many people with a reasonable cooling setup.

### Key changes from [Skylake](https://en.wikichip.org/wiki/intel/microarchitectures/skylake" \o "intel/microarchitectures/skylake)[[edit](https://en.wikichip.org/w/index.php?title=intel/microarchitectures/kaby_lake&action=edit&section=9)]

* Enhanced "14nm+" process results in ~15% higher frequency (100 to 300 MHz increase across the board, same price)
* Same IPC as Skylake (i.e. performance/[MHz](https://en.wikichip.org/w/index.php?title=MHz&action=edit&redlink=1) is unchanged)
* 10x performance/[Watt](https://en.wikichip.org/w/index.php?title=Watt&action=edit&redlink=1) over [Nehalem](https://en.wikichip.org/wiki/intel/microarchitectures/nehalem) (Up from 8x)
* [SkyLake's Speed Shift](https://en.wikichip.org/wiki/intel/microarchitectures/skylake#.22Speed_Shift.22_.28new_power_management.29) implementation is significantly improved, cutting responsiveness by as much as 66% (down to just ~10-15ms to peak frequency).
* Mainstream chipset (See [§ Sockets](https://en.wikichip.org/wiki/intel/microarchitectures/kaby_lake#Sockets.2FPlatform))
  + [Sunrise Point](https://en.wikichip.org/w/index.php?title=intel/chipsets/sunrise_point&action=edit&redlink=1) (100 series) → [Union Point](https://en.wikichip.org/w/index.php?title=intel/chipsets/union_point&action=edit&redlink=1) (200 Series)
    - Sunrise Point is still compatible (may need firmware update)
  + Added support for [Optane](https://en.wikichip.org/w/index.php?title=intel/optane&action=edit&redlink=1" \o "intel/optane (page does not exist)) Technology
* Memory
  + Faster memory for mainstream desktops (i.e., [Kaby Lake S](https://en.wikichip.org/wiki/intel/cores/kaby_lake_s" \o "intel/cores/kaby lake s)) DDR4-2400 (from DDR4-2133)
  + Faster memory for high-perf mobile (i.e., [Kaby Lake H](https://en.wikichip.org/wiki/intel/cores/kaby_lake_h" \o "intel/cores/kaby lake h)) DDR4-2400 (from DDR4-2133)
* Interfaces
  + [Embedded DisplayPort](https://en.wikichip.org/w/index.php?title=Embedded_DisplayPort&action=edit&redlink=1) ([eDP](https://en.wikichip.org/w/index.php?title=eDP&action=edit&redlink=1" \o "eDP (page does not exist))) now supports eDP Standard 1.4 (From 1.3 in Skylake)
* [Gen 9.5](https://en.wikichip.org/wiki/intel/microarchitectures/gen_9.5) GPUs
  + Iris Plus now support HDMI (1.4a) 4096x2304 @ 30 Hz (from 24 Hz)
  + New native hardware support for 4K HEVC/VP9 (See [§ Graphics](https://en.wikichip.org/wiki/intel/microarchitectures/kaby_lake#Graphics))
  + [HD Graphics 510](https://en.wikichip.org/wiki/intel/hd_graphics_510) **→** [HD Graphics 610](https://en.wikichip.org/wiki/intel/hd_graphics_610) (12 Execution Units, no change)
  + [HD Graphics 515](https://en.wikichip.org/wiki/intel/hd_graphics_515) **→** [HD Graphics 615](https://en.wikichip.org/wiki/intel/hd_graphics_615) (24 Execution Units, no change)
  + [HD Graphics 520](https://en.wikichip.org/wiki/intel/hd_graphics_520) **→** [HD Graphics 620](https://en.wikichip.org/wiki/intel/hd_graphics_620) (24 Execution Units, no change)
  + [HD Graphics 530](https://en.wikichip.org/wiki/intel/hd_graphics_530) **→** [HD Graphics 630](https://en.wikichip.org/wiki/intel/hd_graphics_630) (24 Execution Units, no change)
  + [HD Graphics P530](https://en.wikichip.org/wiki/intel/hd_graphics_p530) **→** [HD Graphics P630](https://en.wikichip.org/wiki/intel/hd_graphics_p630) (24 Execution Units, no change)
  + [Iris Graphics 540](https://en.wikichip.org/wiki/intel/iris_graphics_540) **→** [Iris Plus Graphics 640](https://en.wikichip.org/wiki/intel/iris_plus_graphics_640) (48 Execution Units, no change)
  + [Iris Graphics 550](https://en.wikichip.org/wiki/intel/iris_graphics_550) **→** [Iris Plus Graphics 650](https://en.wikichip.org/wiki/intel/iris_plus_graphics_650) (48 Execution Units, no change)
* Families
  + [Core i3](https://en.wikichip.org/wiki/intel/core_i3) processors dropped support for ECC memory (except for Embedded models)
  + [Pentium](https://en.wikichip.org/wiki/intel/pentium_(2009)) desktop processors now have [Hyper-Threading](https://en.wikichip.org/w/index.php?title=intel/hyper-threading&action=edit&redlink=1) support (note that Mobile Pentium already had that feature.)
  + [Pentium](https://en.wikichip.org/wiki/intel/pentium_(2009)) desktop & mobile processors now have Memory Protection ([MPX](https://en.wikichip.org/w/index.php?title=intel/mpx&action=edit&redlink=1)) and [OS Guard](https://en.wikichip.org/w/index.php?title=intel/os_guard&action=edit&redlink=1) support

## Architecture[[edit](https://en.wikichip.org/w/index.php?title=intel/microarchitectures/coffee_lake&action=edit&section=8" \o "Edit section: Architecture)]

[](https://en.wikichip.org/wiki/File:intel_8th_gen_core_logs.png)

Coffee Lake is 8th Generation Core

While there is no change in pure IPC over Skylake and the actual microarchitecture is largely the same, Intel introduced a number of major architectural changes in Coffee Lake. In addition to improved performance brought by the uplift in [binning](https://en.wikichip.org/w/index.php?title=binning&action=edit&redlink=1) as a result of the enhanced process, Coffee Lake also increased the number of cores by 50%, enabling much higher multi-threaded performance. The enhanced manufacturing process should allow Coffee Lake chips to be highly [overclockable](https://en.wikichip.org/w/index.php?title=overclockable&action=edit&redlink=1" \o "overclockable (page does not exist)).

### Key changes from [Kaby Lake](https://en.wikichip.org/wiki/intel/microarchitectures/kaby_lake" \o "intel/microarchitectures/kaby lake)[[edit](https://en.wikichip.org/w/index.php?title=intel/microarchitectures/coffee_lake&action=edit&section=9)]

* Enhanced "14nm++" process results in higher turbo frequencies
* IPC improvement from larger cache for various workloads, but actual core is unchanged
* System Architecture
  + 50% more [cores](https://en.wikichip.org/wiki/physical_core) (6, from 4)
  + 50% larger [last level cache](https://en.wikichip.org/wiki/last_level_cache) (up to 12 MiB, from 8 MiB)
* Core
  + LSD has been re-enabled (Previously [disabled](https://en.wikichip.org/wiki/intel/microarchitectures/skylake_(server)#Front-end))
* Chipset
  + [200 Series chipset](https://en.wikichip.org/w/index.php?title=intel/chipsets/union_point&action=edit&redlink=1) → 300 Series chipset (Cannonlake PCH)
* Memory
  + Faster memory for mainstream desktops (i.e., [Coffee Lake S](https://en.wikichip.org/wiki/intel/cores/coffee_lake_s)) DDR4-2666 (from DDR4-2400)
* Graphics
  + [Gen 9.5](https://en.wikichip.org/wiki/intel/microarchitectures/gen_9.5) GPUs (No Change)
  + HD Graphics 6x0 **→** UHD Graphics 6x0 (Branding change only)
    - [HD Graphics 630](https://en.wikichip.org/wiki/intel/hd_graphics_630) **→** [UHD Graphics 630](https://en.wikichip.org/wiki/intel/uhd_graphics_630) (No change)
* Families
  + [Core i3](https://en.wikichip.org/wiki/intel/core_i3) gained 100% more [cores](https://en.wikichip.org/wiki/physical_core) (4, from 2) but dropped [hyper-threading](https://en.wikichip.org/w/index.php?title=intel/hyper-threading&action=edit&redlink=1) support
  + [Core i5](https://en.wikichip.org/wiki/intel/core_i5) gained 50% more cores (6, from 4)
  + [Core i7](https://en.wikichip.org/wiki/intel/core_i7) gained 50% more cores (6, from 4)

CORE i9 & XEON SCALABLE PROCESSORS

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Intel® Core™ i9 | Intel® Xeon® Platinum | Intel® Xeon® Gold | Intel® Xeon® Silver | Intel® Xeon® Bronze |
| Code Name | Skylake | Skylake | Skylake | Skylake | Skylake |
| **Essentials** | |  | | | |
| Vertical Segment | Desktop | Server | Server | Server | Server |
| Status | Launched | Launched | Launched | Launched | Launched |
| Launch Date | Q2/3'17 | Q3'17 | Q3'17 | Q3'17 | Q3'17 |
| Lithography | 14 nm | 14 nm | 14 nm | 14 nm | 14 nm |
| **Performance** | |  | | | |
| # of Cores | 10 to 18 | 4 to 28 | 4 to 22 | 4 to 12 | 6 to 8 |
| # of Threads | 20 to 36 | 8 to 56 | 8 to 44 | 8 to 24 | 6 to 8 |
| Processor Base Frequency | 2.60 to 3.30 GHz | 2.00 to 3.60 GHz | 2.00 to 3.60 GHz | 1.80 - 2.60 GHz | 1.70 GHz |
| Max Turbo Frequency | 4.20 to 4.30 GHz | 2.80 to 3.80 GHz | 3.20 to 4.20 GHz | 3.00 GHz |  |
| Cache | 13.75 to 24.75 MB L3 | 16.5 to 38.5 MB L3 | 13.75 to 30.25 MB L3 | 8.25 to 16.5 MB L3 | 8.25 to 11 MB L3 |
| Bus Speed | 8 GT/s DMI3 |  |  |  |  |
| # of QPI Links | 0 | 3 or 2 | 3 or 2 | 2 | 2 |
| Intel® Turbo Boost Max Technology 3.0 Frequency ‡ | 4.40 to 4.50 GHz |  |  |  |  |
| TDP | 140 to 165 W | 105 to 205 W | 85 to 200 W | 70 to 85 W | 85 W |
| **Supplemental Information** | |  | | | |
| Embedded Options Available | No | No yes | No yes | Yes No | Yes No |
| Conflict Free | Yes | Yes | Yes | Yes | Yes |
| **Memory Specifications** | |  | | | |
| Max Memory Size (dependent on memory type) | 128 GB | 768GB/1.5 TB | 768 GB/1.5TB | 768 GB | 768 GB |
| Memory Types | DDR4-2666 | DDR4-2666 | DDR4-2666 | DDR4-2400 | DDR4-2133 |
| Maximum Memory Speed |  | 2666 MHz | 2666 MHz | 2400 MHz | 2133 MHz |
| Max # of Memory Channels | 4 | 6 | 6 | 6 | 6 |
| ECC Memory Supported ‡ | No | Yes | Yes | Yes | Yes |
| **Expansion Options** | |  | | | |
| Scalability | 1S Only | S8S, 2S | S4S, 2S | 2S | 2S |
| PCI Express Revision | 3.0 | 3.0 | 3.0 | 3.0 | 3.0 |
| Max # of PCI Express Lanes | 44 | 48 | 48 | 48 | 48 |
| **Package Specifications** | |  | | | |
| Sockets Supported | FCLGA2066 | FCLGA3647 | FCLGA3647 | FCLGA3647 | FCLGA3647 |
| Max CPU Configuration | 1 |  |  |  |  |
| Thermal Solution Specification | PCG 2017X |  |  |  |  |
| T<sub>JUNCTION</sub> | 95°C | 84°C | 82°C | 91°C | 77°C |
| Low Halogen Options Available | See MDDS | See MDDS | See MDDS | See MDDS | See MDDS |
| **Advanced Technologies** | |  | | | |
| Intel® Optane™ Memory Supported ‡ | Yes | No | No | No | No |
| Intel® Speed Shift Technology |  | Yes | Yes | Yes | Yes |
| Intel® Turbo Boost Max Technology 3.0 ‡ | Yes | No | No | No | No |
| Intel® Turbo Boost Technology | 2.0 | 2.0 | 2.0 | 2.0 | No |
|  |  | Yes | Yes | Yes | Yes |
| Intel® Hyper-Threading Technology ‡ | Yes | Yes | Yes | Yes | No |
| Intel® Virtualization Technology (VT-x) ‡ | Yes | Yes | Yes | Yes | Yes |
| Intel® Virtualization Technology for Directed I/O (VT-d) ‡ | Yes | Yes | Yes | Yes | Yes |
| Intel® VT-x with Extended Page Tables (EPT) ‡ |  | Yes | Yes | Yes | Yes |
| Intel® TSX-NI |  | Yes | Yes | Yes | Yes |
| Intel® 64 ‡ | Yes | Yes | Yes | Yes | Yes |
| Instruction Set | 64-bit | 64-bit | 64-bit | 64-bit | 64-bit |
| Instruction Set Extensions | SSE4.1/4.2, AVX2, AVX-512 | Intel® SSE4.2, Intel® AVX, Intel® AVX2, Intel® AVX-512 | Intel® SSE4.2, Intel® AVX, Intel® AVX2, Intel® AVX-512 | Intel® SSE4.2, Intel® AVX, Intel® AVX2, Intel® AVX-512 | Intel® SSE4.2, Intel® AVX, Intel® AVX2, Intel® AVX-512 |
|  |  | 2 | 2 or 1 | 1 | 1 |
| Enhanced Intel SpeedStep® Technology | Yes | Yes | Yes | Yes | Yes |
| Intel® Volume Management Device (VMD) |  | Yes | Yes | Yes | Yes |
| Integrated Intel® Omni-Path Architecture (Intel® OPA) |  | Yes | Yes |  |  |
| **Security & Reliability** | |  | | | |
| Intel® AES New Instructions | Yes | Yes | Yes | Yes | Yes |
| Intel® Trusted Execution Technology ‡ |  | Yes | Yes | Yes | Yes |
| Execute Disable Bit ‡ | Yes | Yes | Yes | Yes | Yes |
| Intel® Run Sure Technology |  | Yes | Yes |  |  |
| Mode-based Execute Control (MBE) |  | Yes | Yes | Yes | Yes |

Future trends

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Codename** | **Introduction** | **Process** | **Predecessor** | **Succesor** | **Cycle** |
| **Haswell** | 2013 | 22nm | **Ivy Bridge** | **Broadwell** | **Tock** |
| **Broadwell** | 2014 | 14nm | Haswell | Skylake | **Process** |
| Skylake | 2015 | 14nm | Broadwell | Kabylake | **Architecture** |
| Kabylake | 2017 | 14nm | Skylake | Coffeelake, Cannon lake | Optimization |
| Coffeelake | 2017 | 14nm | Kabylake | Icelake | Optimization |
| Cannonlake | 2017 | 10nm | Kabylake | Icelake | Process |
| Icelake | 2018 | 10nm | Cannonlake | Tigerlake | Architecture |
| Tigerlake | 2019 | 10nm | Icelake | Sapphire Rapids | Optimization |
| Sapphire Rapids | 2020 | 10nm | Tigerlake | Optimization | |
|  | 2020 | 7nm | Saphire Rapids | | Process |