**INTEL® 64 AND IA-32 ARCHITECTURES**

The exponential growth of computing power and ownership has made the computer one of the most important forces shaping business and society. Intel 64 and IA-32 architectures have been at the forefront of the computer revolution and is today the preferred computer architecture, as measured by computers in use and the total computing power available in the world.

**2.1 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.

**2.1.1 16-bit Processors and Segmentation (1978)**

The IA-32 architecture family was preceded by 16-bit processors, the 8086 and 8088. The 8086 has 16-bit registers and a 16-bit external data bus, with 20-bit addressing giving a 1-MByte address space. 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. Using four segment registers at a time, 8086/8088 processors are able to address up to 256 KBytes without switching between segments. The 20-bit addresses that can be formed using a segment register and an additional 16-bit pointer provide a total address range of 1 MByte.

**2.1.2 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. Descriptors provide 24-bit base addresses with a physical memory size of up to 16 MBytes, support for virtual memory management on a segment swapping basis, and a number of protection mechanisms. These mechanisms include: • Segment limit checking • Read-only and execute-only segment options • Four privilege levels

**2.1.3 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. In addition, the Intel386 processor has support for: • A 32-bit address bus that supports up to 4-GBytes of physical memory • A segmented-memory model and a flat memory model • Paging, with a fixed 4-KByte page size providing a method for virtual memory management • Support for parallel stages

**2.1.4 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. Each stage operates in parallel with the others on up to five instructions in different stages of execution. In addition, the processor added: • An 8-KByte on-chip first-level cache that increased the percent of instructions that could execute at the scalar rate of one per clock • An integrated x87 FPU • Power saving and system management capabilities

**2.1.5 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). The on-chip first-level cache doubled, with 8 KBytes devoted to code and another 8 KBytes devoted to data. The data cache uses the MESI protocol to support more efficient write-back cache in addition to the write-through cache previously used by the Intel486 processor. Branch prediction with an on-chip branch table was added to increase performance in looping constructs. In addition, the processor added: • Extensions to make the virtual-8086 mode more efficient and allow for 4-MByte as well as 4-KByte pages • Internal data paths of 128 and 256 bits add speed to internal data transfers • Burstable external data bus was increased to 64 bits • An APIC to support systems with multiple processors • A dual processor mode to support glueless two processor systems A subsequent stepping of the Pentium family introduced Intel MMX technology (the Pentium Processor with MMX technology). Intel MMX technology uses the single-instruction, multiple-data (SIMD) execution model to perform parallel computations on packed integer data contained in 64-bit registers. See Section 2.2.7, “SIMD Instructions.”

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

The P6 family of processors was based on a superscalar microarchitecture that set new performance standards; see also Section 2.2.1, “P6 Family Microarchitecture.” One of the goals in the design of the P6 family microarchitecture was to exceed the performance of the Pentium processor significantly while using the same 0.6-micrometer, fourlayer, metal BICMOS manufacturing process. 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 Pentium Pro introduced the dynamic execution (micro-data flow analysis, out-of-order execution, superior branch prediction, and speculative execution) in a superscalar implementation. The processor was further enhanced by its caches. It has the same two on-chip 8-KByte 1st-Level caches as the Pentium processor and an additional 256-KByte Level 2 cache in the same package as the processor. • The Intel Pentium II processor added Intel MMX technology to the P6 family processors along with new packaging and several hardware enhancements. The processor core is packaged in the single edge contact cartridge (SECC). The Level l data and instruction caches were enlarged to 16 KBytes each, and Level 2 cache sizes of 256 KBytes, 512 KBytes, and 1 MByte are supported. A half-frequency backside bus connects the Level 2 cache to the processor. Multiple low-power states such as AutoHALT, Stop-Grant, Sleep, and Deep Sleep are supported to conserve power when idling. • 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. Vol. 1 2-3 INTEL® 64 AND IA-32 ARCHITECTURES • 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. SSE extensions expand the SIMD execution model introduced with the Intel MMX technology by providing a new set of 128-bit registers and the ability to perform SIMD operations on packed single-precision floatingpoint values. See Section 2.2.7, “SIMD Instructions.” • 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.

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

The Intel Pentium 4 processor family is based on Intel NetBurst microarchitecture; see Section 2.2.2, “Intel NetBurst® Microarchitecture.” The Intel Pentium 4 processor introduced Streaming SIMD Extensions 2 (SSE2); see Section 2.2.7, “SIMD Instructions.” The Intel Pentium 4 processor 3.40 GHz, supporting Hyper-Threading Technology introduced Streaming SIMD Extensions 3 (SSE3); see Section 2.2.7, “SIMD Instructions.” 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.

**2.1.8 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; see Section 2.2.2, “Intel NetBurst® Microarchitecture.” As a family, this group of IA-32 processors (more recently Intel 64 processors) is designed for use in multi-processor server systems and high-performance workstations. The Intel Xeon processor MP introduced support for Intel® Hyper-Threading Technology; see Section 2.2.8, “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 Dual-Core Intel Xeon processor includes dual core technology. The Intel Xeon processor 70xx series includes Intel Virtualization Technology. The Intel Xeon processor 5100 series introduces power-efficient, high performance Intel Core microarchitecture. This processor is based on Intel 64 architecture; it includes Intel Virtualization Technology and dual-core technology. 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.

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

The Intel Pentium M processor family is a high performance, low power mobile processor family with microarchitectural enhancements over previous generations of IA-32 Intel mobile processors. This family is designed for extending battery life and seamless integration with platform innovations that enable new usage models (such as extended mobility, ultra thin form-factors, and integrated wireless networking). Its enhanced microarchitecture includes: • Support for Intel Architecture with Dynamic Execution • A high performance, low-power core manufactured using Intel’s advanced process technology with copper interconnect • On-die, primary 32-KByte instruction cache and 32-KByte write-back data cache • On-die, second-level cache (up to 2 MByte) with Advanced Transfer Cache Architecture 2-4 Vol. 1 INTEL® 64 AND IA-32 ARCHITECTURES • Advanced Branch Prediction and Data Prefetch Logic • Support for MMX technology, Streaming SIMD instructions, and the SSE2 instruction set • A 400 or 533 MHz, Source-Synchronous Processor System Bus • Advanced power management using Enhanced Intel SpeedStep® technology

**2.1.10 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. See also: • Section 2.2.2, “Intel NetBurst® Microarchitecture” • Section 2.2.3, “Intel® Core™ Microarchitecture” • Section 2.2.7, “SIMD Instructions” • Section 2.2.8, “Intel® Hyper-Threading Technology” • Section 2.2.9, “Multi-Core Technology” • Section 2.2.10, “Intel® 64 Architecture”

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

The Intel Core Duo processor offers power-efficient, dual-core performance with a low-power design that extends battery life. This family and the single-core Intel Core Solo processor offer microarchitectural enhancements over Pentium M processor family. Its enhanced microarchitecture includes: • Intel® Smart Cache which allows for efficient data sharing between two processor cores • Improved decoding and SIMD execution • Intel® Dynamic Power Coordination and Enhanced Intel® Deeper Sleep to reduce power consumption • Intel® Advanced Thermal Manager which features digital thermal sensor interfaces • Support for power-optimized 667 MHz bus The dual-core Intel Xeon processor LV is based on the same microarchitecture as Intel Core Duo processor, and supports IA-32 architecture.

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

The Intel Xeon processor 3000, 3200, 5100, 5300, and 7300 series, Intel Pentium Dual-Core, Intel Core 2 Extreme, Intel Core 2 Quad processors, and Intel Core 2 Duo processor family support Intel 64 architecture; they are based on the high-performance, power-efficient Intel® Core microarchitecture built on 65 nm process technology. The Intel Core microarchitecture includes the following innovative features: • Intel® Wide Dynamic Execution to increase performance and execution throughput • Intel® Intelligent Power Capability to reduce power consumption • Intel® Advanced Smart Cache which allows for efficient data sharing between two processor cores • Intel® Smart Memory Access to increase data bandwidth and hide latency of memory accesses • Intel® Advanced Digital Media Boost which improves application performance using multiple generations of Streaming SIMD extensions The Intel Xeon processor 5300 series, Intel Core 2 Extreme processor QX6800 series, and Intel Core 2 Quad processors support Intel quad-core technology.

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

The Intel Xeon processor 5200, 5400, and 7400 series, Intel Core 2 Quad processor Q9000 Series, Intel Core 2 Duo processor E8000 series support Intel 64 architecture; they are based on the Enhanced Intel® Core microarchitecture using 45 nm process technology. The Enhanced Intel Core microarchitecture provides the following improved features: • A radix-16 divider, faster OS primitives further increases the performance of Intel® Wide Dynamic Execution. • Improves Intel® Advanced Smart Cache with Up to 50% larger level-two cache and up to 50% increase in wayset associativity. • A 128-bit shuffler engine significantly improves the performance of Intel® Advanced Digital Media Boost and SSE4. Intel Xeon processor 5400 series and Intel Core 2 Quad processor Q9000 Series support Intel quad-core technology. Intel Xeon processor 7400 series offers up to six processor cores and an L3 cache up to 16 MBytes.

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

The first generation of Intel® AtomTM processors are built on 45 nm process technology. They are based on a new microarchitecture, Intel® AtomTM microarchitecture, which is optimized for ultra low power devices. The Intel® AtomTM microarchitecture features two in-order execution pipelines that minimize power consumption, increase battery life, and enable ultra-small form factors. The initial Intel Atom Processor family and subsequent generations including Intel Atom processor D2000, N2000, E2000, Z2000, C1000 series provide the following features: • Enhanced Intel® SpeedStep® Technology • Intel® Hyper-Threading Technology • Deep Power Down Technology with Dynamic Cache Sizing • Support for instruction set extensions up to and including Supplemental Streaming SIMD Extensions 3 (SSSE3). • Support for Intel® Virtualization Technology • Support for Intel® 64 Architecture (excluding Intel Atom processor Z5xx Series)

**2.1.15 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.

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

The Intel Core i7 processor 900 series support Intel 64 architecture; they are based on Intel® microarchitecture code name Nehalem using 45 nm process technology. The Intel Core i7 processor and Intel Xeon processor 5500 series include the following innovative features: • Intel® Turbo Boost Technology converts thermal headroom into higher performance. • Intel® HyperThreading Technology in conjunction with Quadcore to provide four cores and eight threads. • Dedicated power control unit to reduce active and idle power consumption. • Integrated memory controller on the processor supporting three channel of DDR3 memory. • 8 MB inclusive Intel® Smart Cache. • Intel® QuickPath interconnect (QPI) providing point-to-point link to chipset. • Support for SSE4.2 and SSE4.1 instruction sets. • Second generation Intel Virtualization Technology.

**2.1.17 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. They support the same features described in Section 2.1.16, plus the following innovative features: • Up to eight cores per physical processor package. • Up to 24 MB inclusive Intel® Smart Cache. • Provides Intel® Scalable Memory Interconnect (Intel® SMI) channels with Intel® 7500 Scalable Memory Buffer to connect to system memory. • Advanced RAS supporting software recoverable machine check architecture.

**2.1.18 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 innovative features can include: • Deliver smart performance using Intel Hyper-Threading Technology plus Intel Turbo Boost Technology. • Enhanced Intel Smart Cache and integrated memory controller. • Intelligent power gating. • Repartitioned platform with on-die integration of 45 nm integrated graphics. • Range of instruction set support up to AESNI, PCLMULQDQ, SSE4.2 and SSE4.1.

**2.1.19 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 the same features described in Section 2.1.16, plus the following innovative features: • Up to six cores per physical processor package. • Up to 12 MB enhanced Intel® Smart Cache. • Support for AESNI, PCLMULQDQ, SSE4.2 and SSE4.1 instruction sets. • Flexible Intel Virtualization Technologies across processor and I/O.

**2.1.20 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 and have innovative features including: • Intel Turbo Boost Technology for Intel Core i5 and i7 processors • Intel Hyper-Threading Technology. • Enhanced Intel Smart Cache and integrated memory controller. • Processor graphics and built-in visual features like Intel® Quick Sync Video, Intel® InsiderTM etc. • Range of instruction set support up to AVX, AESNI, PCLMULQDQ, SSE4.2 and SSE4.1. Intel Xeon processor E3-1200 product family is also based on the Sandy Bridge microarchitecture. Intel Xeon processor E5-2400/1400 product families are based on the Sandy Bridge-EP microarchitecture. Intel Xeon processor E5-4600/2600/1600 product families are based on the Sandy Bridge-EP microarchitecture and provide support for multiple sockets.

**2.1.21 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. The Intel Xeon processor E7-8800/4800/2800 v2 product families and Intel Xeon processor E3- 1200 v2 product family are also based on the Ivy Bridge microarchitecture. The Intel Xeon processor E5-2400/1400 v2 product families are based on the Ivy Bridge-EP microarchitecture. The Intel Xeon processor E5-4600/2600/1600 v2 product families are based on the Ivy Bridge-EP microarchitecture and provide support for multiple sockets.

**2.1.22 The Fourth Generation Intel® Core™ Processor Family (2013)**

The Fourth Generation Intel Core processor family spans Intel Core i7, i5 and i3 processors based on the Haswell microarchitecture. Intel Xeon processor E3-1200 v3 product family is also based on the Haswell microarchitecture.

**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.

In this document, references to the Pentium 4 processor refer to processors based on the Intel NetBurst® microarchitecture. This includes the Intel Pentium 4 processor and many Intel Xeon processors based on Intel NetBurst microarchitecture. Where appropriate, differences are noted (for example, some Intel Xeon processors have third level cache). The Dual-core Intel® Xeon® processor LV is based on the same architecture as Intel® Core™ Duo and Intel® Core™ Solo processors. Intel® Atom™ processor is based on Intel® Atom™ microarchitecture.

**SKYLAKE architecture**

**2.1 THE SKYLAKE MICROARCHITECTURE**

The Skylake microarchitecture builds on the successes of the Haswell and Broadwell microarchitectures. The basic pipeline functionality of the Skylake microarchitecture is depicted in Figure 2-1.

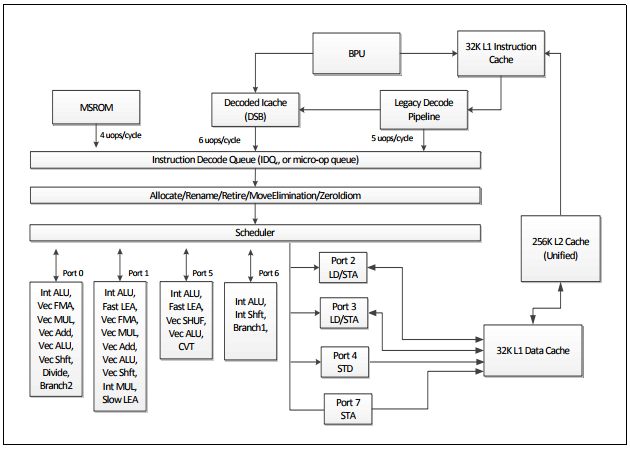


Figure 2-1. CPU Core Pipeline Functionality of the Skylake Microarchitecture

The Skylake microarchitecture offers the following enhancements: • Larger internal buffers to enable deeper OOO execution and higher cache bandwidth. • Improved front end throughput. • Improved branch predictor. • Improved divider throughput and latency. • Lower power consumption. • Improved SMT performance with Hyper-Threading Technology. • Balanced floating-point ADD, MUL, FMA throughput and latency. The microarchitecture supports flexible integration of multiple processor cores with a shared uncore subsystem consisting of a number of components including a ring interconnect to multiple slices of L3 (an off-die L4 is optional), processor graphics, integrated memory controller, interconnect fabrics, etc. A four-core configuration can be supported similar to the arrangement shown in Figure 2-3.

**2.1.1 The Front End**

The front end in the Skylake microarchitecture provides the following improvements over previous generation microarchitectures: • Legacy Decode Pipeline delivery of 5 uops per cycle to the IDQ compared to 4 uops in previous generations. • The DSB delivers 6 uops per cycle to the IDQ compared to 4 uops in previous generations. • The IDQ can hold 64 uops per logical processor vs. 28 uops per logical processor in previous generations when two sibling logical processors in the same core are active (2x64 vs. 2x28 per core). If only one logical processor is active in the core, the IDQ can hold 64 uops (64 vs. 56 uops in ST operation). • The LSD in the IDQ can detect loops up to 64 uops per logical processor irrespective ST or SMT operation. • Improved Branch Predictor. 2.1.2 The Out-of-Order Execution Engine The Out of Order and execution engine changes in Skylake microarchitecture include: • Larger buffers enable deeper OOO execution compared to previous generations. • Improved throughput and latency for divide/sqrt and approximate reciprocals. • Identical latency and throughput for all operations running on FMA units. • Longer pause latency enables better power efficiency and better SMT performance resource utilization. Table 2-1 summarizes the OOO engine’s capability to dispatch different types of operations to various ports.

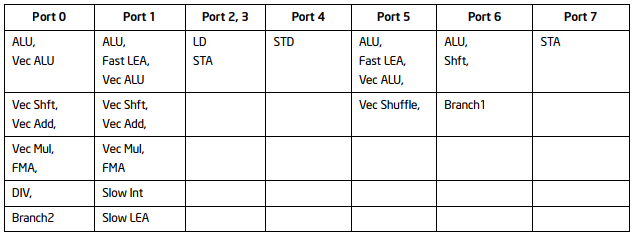


Table 2-1. Dispatch Port and Execution Stacks of the Skylake Microarchitecture

Table 2-2 lists execution units and common representative instructions that rely on these units. Throughput improvements across the SSE, AVX and general-purpose instruction sets are related to the number of units for the respective operations, and the varieties of instructions that execute using a particular unit.

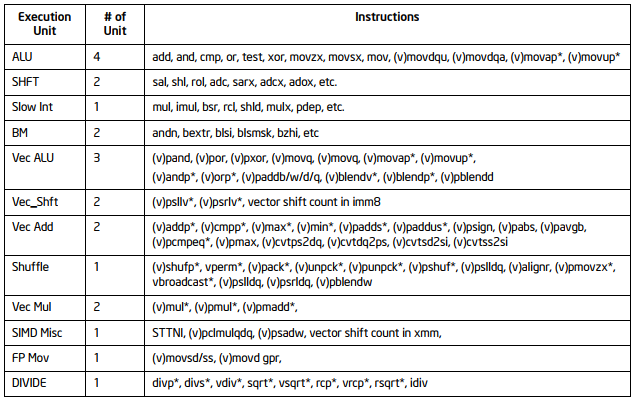


Table 2-2. Skylake Microarchitecture Execution Units and Representative Instructions1

NOTES: 1. Execution unit mapping to MMX instructions are not covered in this table. See Section 11.16.5 on MMX instruction throughput remedy

A significant portion of the SSE, AVX and general-purpose instructions also have latency improvements. Appendix C lists the specific details. Software-visible latency exposure of an instruction sometimes may include additional contributions that depend on the relationship between micro-ops flows of the producer instruction and the micro-op flows of the ensuing consumer instruction. For example, a two-uop instruction like VPMULLD may experience two cumulative bypass delays of 1 cycle each from each of the two micro-ops of VPMULLD. Table 2-3 describes the bypass delay in cycles between a producer uop and the consumer uop. The leftmost column lists a variety of situations characteristic of the producer micro-op. The top row lists a variety of situations characteristic of the consumer micro-op.

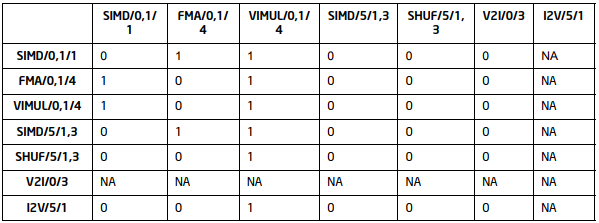


Table 2-3. Bypass Delay Between Producer and Consumer Micro-ops

The attributes that are relevant to the producer/consumer micro-ops for bypass are a triplet of abbreviation/one or more port number/latency cycle of the uop. For example: • “SIMD/0,1/1” applies to 1-cycle vector SIMD uop dispatched to either port 0 or port 1. • “VIMUL/0,1/4” applies to 4-cycle vector integer multiply uop dispatched to either port 0 or port 1. • “SIMD/5/1,3” applies to either 1-cycle or 3-cycle non-shuffle uop dispatched to port 5.

**2.1.3 Cache and Memory Subsystem**

The cache hierarchy of the Skylake microarchitecture has the following enhancements: • Higher Cache bandwidth compared to previous generations. • Simultaneous handling of more loads and stores enabled by enlarged buffers. • Processor can do two page walks in parallel compared to one in Haswell microarchitecture and earlier generations. • Page split load penalty down from 100 cycles in previous generation to 5 cycles. • L3 write bandwidth increased from 4 cycles per line in previous generation to 2 per line. • Support for the CLFLUSHOPT instruction to flush cache lines and manage memory ordering of flushed data using SFENCE. • Reduced performance penalty for a software prefetch that specifies a NULL pointer. • L2 associativity changed from 8 ways to 4 ways.

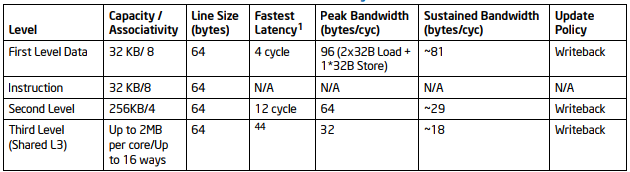


Table 2-4. Cache Parameters of the Skylake Microarchitecture

The TLB hierarchy consists of dedicated level one TLB for instruction cache, TLB for L1D, plus unified TLB for L2. The partition column of Table 2-5 indicates the resource sharing policy when Hyper-Threading Technology is active.

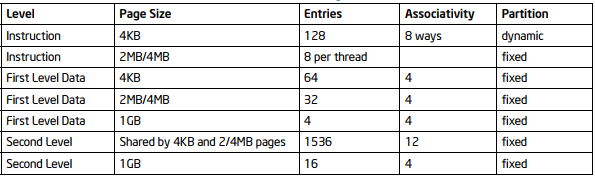


Table 2-5. TLB Parameters of the Skylake Microarchitecture