The exponential growth of both the power and breadth of usage of the computer has made it the most important force that is reshaping human technology, business, and society in the second half of the twentieth century. Further, the computer promises to continue to dominate technology growth well into the twenty-first century. Microprocessor is the brain of a computer. Intel has been the dominant player in the designing and manufacturing of microprocessors, although there are Advanced Micro Devices and Cyrix. In more than twenty years, Intel has introduced 6 generations of microprocessors. Next, we will give a brief introduction to each processor.

1. From 8086 to 80486 [1-4]

The development of microprocessors in Intel really started from 4004, which was designed in 1969. The first actual processor is the 8086 introduced in 1978, quickly followed by a more cost-effective version, the 8088. The 8086 processor has 16-bit registers and a 16-bit external data bus, with 20-bit addressing to give a 1-MB address space. The 8086 ran at a speed of 8 MHz. The 8088 processor is identical to the 8086 except for a smaller external bus of 8 bits. Segmentation of up to 64 KB in size is introduced.

The 286 processor was introduced in 1982 and ran at a speed of 12.5 MHz with 134,000 transistors on chip. It also supports the virtual memory management on a segment swapping basis, and various protection mechanism including segment limit checking, read-only and execute-only segment options, and up to 4 privilege levels to protect operating system code.

The 386 processor was introduced in 1985 and ran at a speed up to 20 MHz. It introduces 32-bit registers into the architecture for use both as operands for calculations and for addressing. The 32-bit addressing is supported with an external 32-bit address bus. Each segment can be as large as 4 GB. The original instructions are enhanced with new 32-bit operand and addressing form, and completely new instructions are added. Paging is also introduced into 386 with the fixed 4 KB page size providing a method for virtual memory management that is significantly superior compared to using segments for the purpose. The paging is much more efficient for operating systems and completely transparent to applications without significant sacrifice of execution speed. Parallel processing has been a good way to enhance performance and 386 processor is the first processor to include a number of parallel stages: six. These stages are the bus interface unit (accessing memory and I/O), the code pre-fetch unit (receiving object code from the bus unit and putting it into a 16-byte queue), the instruction decode unit (decoding object code from the pre-fetch unit into micro-codes), the execute unit (executing micro-codes), the segment unit (translating logical addresses to linear addresses and doing protection-check) and the paging unit (translating linear addresses to physical addresses, doing page based protection check, and containing a cache with information for up to 32 most recently accessed pages).

The 486 processor was introduced in 1989 and the top speed is 100 MHz. It adds more parallel execution capability by expanding the 386 processor’s instruction decode and execution
units into five pipelined stages. Each stage operates in parallel and can do its work on one
instruction in one clock cycle. On-chip level 1 cache was introduced and it is unified both for
instructions and data. The processor with a frequency at 100 MHz has a 16 KB L1 cache and uses
a write-back mechanism. The processors with lower frequencies have 8 KB write-through L1 cache.
The 486 processor provides supports for L2 cache and multiprocessing, thus reduces bus utilization
and allow multiprocessors to share a single memory bus. The 486 processor also for the first time
put the floating-point unit onto the same chip as the CPU (Earlier FPUs were on separate chips.),
thus give more efficiency to the processor. The introduction of speed-multiplying technology and
the use of the clock-multiplier allow the processor to operate at frequencies higher than the external
memory bus speed. Now, processor core speed is usually several times that of the memory bus
speed.

1.1. Pentium [1,5,6]

In March 1993, Intel introduced its fifth generation microprocessor and also changed its
naming from x86 series to Pentium. It runs at a speed up to 200 MHz and the system bus works at
66 MHz. The Pentium processor adds a second execution pipeline to achieve super-scalar
performance. It has two general-purpose integer pipelines and a pipelined floating-point unit. The
main integer pipeline (U) has five stages: pre-fetch, decode 1, decode 2, execution, and write-back.
The second general pipeline (V) is similar. The processor can issue up to two instructions per clock
cycle. The two instructions are checked and issued if possible. The first one goes to pipeline U and
the second one goes to pipeline V. If only one instruction is issued, it goes to pipeline U and the V
pipeline has no instruction. It is capable of executing up to two integer instructions in parallel,
doubling the performance relative to an equivalent frequency 486 processor. When a stall happens,
successive instructions are not allowed to pass the stalled instruction in either pipeline. The
instruction pre-fetcher has four 32-byte buffers. When a branch instruction is fetched, the branch
target buffer predicts whether the branch is going to be taken or not. If a branch is predicted not
taken, pre-fetch requests continue going. If the branch is predicted to be taken, the other pre-fetch
buffer is enabled and starts to pre-fetch as if the branch were taken. If a branch is found to be
mal-predicted, the instruction pipelines are flushed and pre-fetching starts over. The processor
uses a dynamic branch prediction scheme with a 256-entry branch target buffer. If a branch is mal-
predicted, there is a three-cycle penalty if the branch was executed in the U pipeline or a four-cycle
penalty if it was executed in the V pipe.

The Pentium processor has 2 write buffers, one for each of the integer pipelines to enhance
the performance of consecutive writes to memory. It supports strong write ordering, that is, writes
happen in the order they occur. The pipelined floating-point unit (FPU) has been completely
redesigned over the 486 processor FPU, although in 1994 Intel announced the floating-point unit
flaw in Pentium processors at a speed range of 60 to 100 MHz. The FPU is made of appending a
three stage floating-point pipeline to the integer pipeline. Thus, it has pre-fetch, decode 1, decode 2,
X1 stage, X2 stage and WF stage. Most floating-point instructions have execution latencies of more
than one clock cycle. The FPU is capable of executing two floating-point instructions per clock cycle
and achieves three to five times the performance over 486 processor FPU for common operations.
The on-chip level 1 cache system consists of two separate 8-KB 2-way set associative caches, one
for code and one for data. The caches use a write-back (in some cases, write-through as used in
486) mechanism and an LRU replacement policy. The data cache can be accessed simultaneously
from both pipes. The minimum delay for a cache miss is 4 clock cycles. The registers are still 32-bit wide. The internal data paths of 128- and 256-bit are added to speed up internal data transfer. It also uses a 64-bit wide external data bus. Pages are either 4 KB or 4 MB big. The use of advanced programmable interrupt controller (APIC) supports the systems with multiprocessors. Many instructions, which were micro-coded in earlier x86 processors, are now hardwired for increased performance. The processor now has hardware support for virtual interrupts provided through virtual interrupt flag (VIF) and virtual interrupt pending (VIP) bits in the EFLAGS register.

1.2. Pentium Pro [1,6-8]

Pentium Pro processor introduced in November 1995 was the sixth generation of processors from Intel and a big advance from the classic Pentium processor, although Intel now has discontinued the Pentium Pro and replaced it with the Pentium II processor. Pentium Pro was the first in Intel family to use the dynamic execution micro-architecture. It has a de-coupled, 12-stage, super-pipelined implementation, thus trading less work per stage for more stages. Each pipeline stage in Pentium Pro was 33% less in time than that in a Pentium processor. This helps to achieve a higher clock rate. The dynamic execution removes the constraint of linear instruction sequencing between the traditional “fetch” and “execute” phases and opens up a wide instruction window using an instruction pool. This approach allows the "execute" phase of the processor to have much more visibility into the program's instruction stream so that better scheduling may take place. Optimized scheduling requires the fundamental "execute" phase to be replaced by de-coupled "dispatch/execute" and the "retire" phases. This allows instructions to be started in any order but always be completed in the original program order. The processor was implemented as three independent engines coupled with an instruction pool as shown in figure 1.

![Figure 1: Three Engines Communicating Using An Instruction Pool](image)

The instruction fetch/decode unit fetches a stream of instructions from the L1 instruction cache and decodes them into a series of micro-operations. The micro-operation stream (still in the order of the original instruction stream) is sent to the instruction pool (also called reorder buffer) The instruction fetch unit fetches one 32-byte line per clock cycle from the instruction cache and
transmits 16 aligned bytes to the decode unit. The fetch unit computes the instruction pointer based on inputs from the branch target buffer, the exception/interrupt status and branch-misprediction indications from the integer execution units. The most important part is the branch prediction performed by the branch target buffer. Using an extension of YehUs algorithm, the 512-entry branch target buffer looks many instructions ahead of the retirement program counter. Within this instruction window there may be numerous branches, procedure calls and returns that must be correctly predicted if the dispatch/execute unit is to do useful work. A 512-entry branch target buffer (BTB) which stores the history of previously-seen branches and their targets is used to predict the direction and target of branches based on an instruction’s address. When a branch is pre-fetched, the BTB feeds the target address directly into the instruction fetch unit. Once the branch is executed, the BTB is updated with the target address. The branch target buffer prediction algorithm includes pattern matching and can track up to the last four branch directions per branch address. Thus, a loop with four or less iterations should have about 100% correct prediction. Using the branch target buffer allows dynamic prediction of previously seen branches. The instruction decoder contains three parallel decoders: two simple-instruction decoders and one complex instruction decoder. An instruction can be converted into one or more micro-operations. These micro-operations are then executed by the execution units. The instruction decoder can generate up to 6 micro-operations per clock cycle (one each for the simple instruction decoders and four for the complex instruction decoder). The instruction pool (formally known as the reorder buffer) is an array of content-addressable memory, arranged into 40 micro-operation registers. It contains micro-operations that are waiting to be executed, as well as those that have already been executed but not yet committed to machine states (memory). The dispatch/execute unit can execute instructions from the reorder buffer in any order. The dispatch/execute unit is an out-of-order unit that schedules and executes the micro-operations stored in the reorder buffer according to data dependencies and resource availability and temporarily stores the results of these speculative executions. The scheduling and dispatching of micro-operations from the instruction pool are handled by the reservation station. It continuously scans the reorder buffer for micro-operations that are ready to be executed (i.e., all the source operands are available) and dispatches them to the available units. The results of a micro-operation execution are returned to the reorder buffer and stored along with the micro-operation until it is retired. This scheduling and dispatching process supports classic out-of-order execution, where micro-operations are dispatched to the execution units strictly according to data-flow constraints and execution resource availability, without regard to the original ordering of the instructions. When two or more micro-operations of the same type (for instance, integer operations) are available at the same time, they are executed in a pseudo FIFO order in the reorder buffer. Execution of micro-operations is handled by two integer units, two floating-point units, and one memory-interface unit, allowing up to 5 micro-operations to be scheduled per clock cycle. The two integer units can handle two integer micro-operations in parallel. One of the two units is designed to handle branch micro-operations. The memory interface unit handles load and store micro-operations. It can execute both a load and a store in parallel in one clock cycle.

The processor temporarily stores each write to memory in a write buffer. The write buffer improves processor performance by allowing the processor to continue executing instructions without having to wait until a write to memory and/or to a cache is complete. The retirement unit continuously searches the instruction pool for completed micro-operations that no longer have data dependencies with other micro-operations or unresolved branch-predictions. When completed
micro-operations are found, the retirement engine commits the results of these micro-operations to memory and/or registers (the processor’s eight general-purpose registers and eight floating-point unit data registers) in the order they were originally issued taking into accounts of interrupts, exceptions, breakpoints, and branch mal-predictions and retires or removes the micro-operations from the instruction pool. The retirement unit can retire 3 micro-operations per clock cycle. These three independent engines complete the jobs (deep branch prediction, dynamic data flow analysis, and speculative execution) behind the concept of dynamic execution. Using parallel processing technique, the Pentium Pro processor was able on average to decode, dispatch and complete execution (retire) of three instructions per clock cycles. The Pentium Pro processor ran at a speed up to 200 MHz and used a 66 MHz system bus. It had a separate level 1 cache with 8 KB 4-way set associative for instructions and 8 KB dual-ported, 2-way set associative for data supporting one load and one store operation per clock cycle. The level 2 cache was 256 KB big and it was on chip, coupled to the core processor through a full clock-speed 64-bit cache bus. Cache coherence was maintained using the MESI (modified, exclusive, shared, invalid) cache protocol.

1.3. Pentium with MMX Technology [6,9-13]

To combat for the game market and make PC the platform of choice for games, 3-D, motion video and virtual reality, Intel developed the MMX technology to boost multimedia performance. The Pentium processor with MMX technology (running at a speed up to 233 MHz) delivers 10-20% more performance and is over 60% faster compared to the original Pentium processor at the same frequency. The performance enhancements are 28.2% and 26.0% respectively when comparisons are made at the 200 MHz and 166 MHz levels. The MMX technology is a multimedia extension to x86 code. It uses SIMD (single instruction, multiple data) technique which allows the parallel processing of many data pieces with one instruction, and thus enables increased performance on a wide variety of multimedia and communications applications. It consists of 57 new instructions, 8 64-bit MMX registers and 4 new data types. Using 64-bit MMX registers and MMX instructions, it can operate simultaneously on 8 bytes, four 16-bit words or two 32-bit pairs of words. It is designed to improve processor performance in traditional digital signal processor applications including the graphics, audio and voice processing capabilities. Pentium processor with MMX technology adds additional stages to the pipeline. The MMX pipeline is integrated with the integer pipeline, very similar to the integration of the floating-point pipeline. An additional stage (fetch stage) is added to the integer pipeline, any prefixes are decoded in the fetch stage. Up to two instructions can be executed every clock cycle. An instruction first-in, first-out (FIFO) buffer is situated between the newly introduced fetch stage and the decode 1 stage to alleviate stalls that may occur during instruction fetch and parsing. The MMX instructions that access memory or integer registers can only be executed in the U pipeline and cannot be paired with any non-MMX instructions. After updating an MMX register, one additional clock cycle must pass before that MMX register can be moved to either memory or to an integer register.

The level 1 cache in the processor has also been doubled. Now, it is 16 KB for instructions and 16 KB for data. Data cache is 4-way set associative compared to 2-way set associative in the Pentium classic. Larger L 1 cache improves performance by reducing the memory access and providing faster access to recently used instructions and data. The instructions and data caches can be accessed simultaneously while the dual-port data cache supports two data fetching simultaneously from two pipelines. The data cache also supports a write-back (or alternatively,
write-through, on a line by line basis) policy for memory updates. The penalty for a cache miss is eight internal clock cycles. The branch target buffer, used by the dynamic branch prediction to boost performance by predicting the most likely set of instructions to be executed, has been improved with the MMX technology to increase its accuracy. A pool of 4 write buffers, which was 2 in the Pentium classic, is shared between the dual pipelines to improve memory write performance.

1.4. Pentium II [1,14-18]

The Pentium II processor (code named "Klamath", introduced in May 1997) runs at speeds ranging from 233 MHz to 450 MHz. It is manufactured with the 0.25 micron manufacturing process and has over 7.5 million transistors on die resulting in more power in less space. At 450 MHz, the Pentium II processor gives 32% more integer performance (measured by SPECint95), 35% more floating-point performance (as measured by SPECfp95), and 30% more media performance (measured by Norton Media Benchmark) than the 333 MHz Pentium II processor.

The Pentium II processor is made from the addition of the MMX technology to the Pentium Pro processor architecture. Thus, it has Pentium Pro's power (although sacrificed a little bit from the removal of the L2 cache from chip for cost savings) and the enhanced multimedia performance from the MMX technology. The Pentium II processor uses a larger, separated level 1 cache, two 16-KB 4-way set associative for instructions and data to reduce the disadvantage of slower external level 2 cache. The caches employ a write-back mechanism and a pseudo-LRU replace algorithm. The level 2 cache (512 KB, unified) is removed from the chip for cost savings. There is a dedicated 64-bit cache bus with error correction code (ECC) functionality on it for applications where data intensity and reliability are essential. The transfer rates between the Pentium II processor core and the L2 cache are one-half of the processor core clock frequencies and scale with the processor core frequencies. Some versions of the Pentium II processor introduced later for embedded systems include a 256-KB on-chip cache to enhance performance. The multi-transaction system bus runs at 66 MHz and 100 MHz (on some versions). It supports for up to 2 processors, enables two-way symmetric multiprocessor. The L2 cache does not connect to the Pentium II processor system bus, the processor has a dedicated cache bus. The dual independent bus architecture separates the cache bus from the system bus and gives high bandwidth, performance and scalability. As a member of the P6 family of processors, the Pentium II processor also uses the dynamic execution micro-architecture, which incorporates a unique combination of multiple branch prediction, data flow analysis for optimized, reordered scheduling of instructions, and speculative execution for carrying out instructions based on optimized schedule and enabling the super-scalar execution units remain busy. The MMX media enhancement technology includes single instruction, multiple data (SIMD) technique, 57 new instructions, 8 64-bit MMX technology registers, and 4 new data types. It is designed to achieve a new level of performance for a wide variety of multimedia and communications applications.

Its pipelined floating-point unit (FPU) supports 32-, 64-, as well as 80-bit formats. The parity-protection address/request and response system bus signals with a retry mechanism for high data integrity and reliability.
1.5. Pentium III [1,15,19,20]

In February 1999, Intel introduced the Pentium III processor at speeds ranging from 450 MHz to 1.13 GHz. The system bus runs at 133 MHz (available on 2 versions) and 100 MHz. It supports up to 2 processors, enables symmetric multiprocessing (SMP). The newly introduced advanced system buffering consists of 4 write-back buffers, 6 fill buffers, and 8 bus-queue entries. It is an optimization in system buffer sizes and bus queue entries that result in an increase in the utilization of the available bandwidth on the 133 MHz and 100 MHz system bus. Level 1 cache is still the same, separate 16 KB for instructions and 16 KB for data. Level 2 cache is moved back on-die again. At 256 KB, L2 cache is 8-way set associative. Data move in and out of cache every 2 clock cycles. A 256-bit (32-byte) data bus to the L2 cache gives a bandwidth of 16 GB/s at 1 GHz. Some versions of the Pentium III processor include an off-die L2 cache. This L2 cache consists of a 512-KB unified cache with a dedicated 64-bit cache bus.

The internet streaming SIMD extensions (SSE), designed to enhance the performance of floating-point applications, consist of 70 new instructions and include single instruction, multiple data (SIMD) for floating-point, additional SIMD-integer and cachability control instructions. The SSE support SIMD operations on packed single-precision floating-point data types and the additional SIMD integer instructions support operations on packed quad-word data types (byte, word, or double-word). There are also eight 128-bit XMM registers. This gives the programmer the ability to develop algorithms that can mix packed single-precision floating-point and integer using both SSE and MMX instructions respectively. When SSE are used in desktop and internet applications, the benefits include higher resolution and quality images can be viewed and manipulated, high quality audio, MPEG2 video and simultaneous MPEG2 encoding and decoding, reduced CPU utilization for speech recognition as well as higher accuracy and shorter response time. Other features that are used in the Pentium III processor and remain unchanged compared to the Pentium II processor include: P6 dynamic execution micro-architecture having multiple branch prediction, data flow analysis and speculative execution functionality; Dual independent bus architecture placing the L2 cache on a dedicated, high-speed bus, freeing the system bus from cache traffic and giving improved system bandwidth, system performance and scalability; MMX media enhancement technology for enhanced multimedia and communications performance; A pipelined floating-point unit supporting the 32-, 64-, as well as, 80-bit format; Parity-protected address/request and response system bus signals with a retry mechanism for high data integrity and reliability; The testing and performance monitoring features are inherited without changes.

1.6. Pentium II Xeon and Pentium III Xeon [21,22]

Intel also has Pentium II Xeon and Pentium III Xeon processors for mid-range and higher servers and workstations. Pentium II Xeon comes with a speed of 400 MHz and 450 MHz. The system bus is 100 MHz and it supports up to 8 processors. The major difference is in the L2 cache. In Pentium II Xeon, the L2 cache is as big as 2 MB, and provides a performance boost to server applications and other applications that work on large data sets. The L2 cache bus operates at the same speed as the processor core. Pentium III Xeon is Intel’s most advanced and most powerful processor for the entry to mid-range servers and workstations. It is available at speeds up to 1 GHz. It has two processors, so is called dual processing with a system bus running at 133 MHz. Intel’s next generation processor for servers and workstations will be Itantium.
1.7. Celeron [23]

Intel’s Celeron processor is promoted for the value-priced PC market. Its speed ranges from 600 MHz to 800 MHz. The 800 MHz processor uses a 100 MHz system bus, and the rest use the 66 MHz bus. L1 cache stays the same, separated 16 KB for instructions and 16 KB for data. L2 cache is 128 KB on-chip. Other features including the P6 dynamic execution, MMX media enhancement, streaming SIMD extensions, pipelined floating-point unit, etc., remain unchanged.

1.8. Pentium 4 [24-27]

In November 2000, after a 4-month delay, Intel finally introduced its Pentium 4 processor to the computer community. A lot of changes have been made and it is a big advance from Pentium III to Pentium 4. Running at 1.30, 1.40 and 1.50 GHz, it is compatible with previous generation processors. The Pentium 4 processor is designed to deliver performance across applications and usage where end users can truly appreciate and experience the performance in the internet age. These applications include internet audio, stream video, image processing, video content creation, speech, 3D, CAD, games, multi-media and multi-tasking user environments. The so-called NetBurst micro-architecture used in Pentium 4 is designed to achieve high performance for both integer and floating-point computations at very high clock rates. It has many advanced features and improvements over the P6 micro-architecture used in Pentium Pro, Pentium II, and Pentium III.

Hyper Pipelined Technology: It doubles the pipeline depth compared to the P6 micro-architecture. One of the key pipelines, the branch prediction/recovery pipeline, is implemented in 20 stages compared to 10 stages in the P6 micro-architecture. More stages allow the processor to clock to higher frequencies and less work is performed per stage. Performance penalty of mal-predicted branch instructions may also increase.

400 MHz System Bus: A 400 MHz system bus is used, which delivers 3.2 GB of data per second in and out of the processor. This is accomplished through a physical signaling scheme of quad pumping the data transfers over a 100-MHz clocked system bus (i.e. The bus runs at 100 MHz with 4 bits of data being passed each clock cycle.) and a buffering scheme allowing for sustained 400 MHz data transfer. This new system bus delivers 3 times the bandwidth of the Pentium III processor bus and it supports for single processor configurations compared to up to 2 processors supported by Pentium III system bus.

Random Access Memory: Besides the originally used pricey dual channel Rambus DRAM (RDRAM), which was previously the only memory option for the Pentium 4, Intel also decided to use double data rate SDRAM and slower SDRAM in Pentium 4. They may appear in the market by the second half of 2001. The dual RDRAM memory channel delivers 3.2 GB/s of memory bandwidth to the processor. High memory bandwidth gives balanced platform support and provides the memory bandwidth necessary to extract full performance from the Pentium 4 processor.

Level 1 Execution Trace Cache: The processor’s micro-architecture can support up to three levels of on-chip cache. Only two levels of on-chip caches are implemented in the Pentium 4 processor. All caches use a pseudo-LRU replacement algorithm. The level 1 data cache has been reduced from 16 KB in Pentium III to 8 KB in Pentium 4 to alleviate manufacturing costs. It is 4-way set associative and uses a write-through mechanism. The Pentium 4 processor includes an execution trace cache (instruction cache) that stores up to 12 KB decoded micro-ops in the order of program execution. This increases performance by removing the IA-32 decoder from the main
execution loop and in turn removes the decoder pipeline latency, makes more efficient use of the cache storage space since it only stores instructions that are needed. The result is a mean to deliver a high volume of instructions to the processor’s execution units and a reduction in the overall time required to recover from branches that have been mal-predicted. Previously, when a branch was mal-predicted, the processor had to start from fetching and re-decoding a new instruction. With Pentium 4, the processor goes straight to the execution trace cache, retrieves the decoded micro-op it needs and sends it through the execution pipeline, thus, speeding up the process considerably.

1.8.1. 256 KB, Level 2 Advanced Transfer Cache:

The advanced transfer cache consists of a 256-bit interface that transfers data on each core clock cycle. As a result, the Pentium 4 processor at 1.50 GHz can deliver data at 48 GB/s whereas the transfer rate is 16 GB/s on Pentium III processor at 1 GHz. The L2 cache is on-die, 8-way set associative, and it uses a write-back mechanism.

1.8.2. Rapid Execution Engine:

Two arithmetic logic units (ALU) on the Pentium 4 processor are running at twice the core processor frequency. This allows basic integer instructions such as add, subtract, logical and, logical or, etc. to execute in half a clock cycle. For example, the rapid execution engine on a 1.50 GHz Pentium 4 processor runs at 3 GHz. This decreases the latency and increases the throughput of basic integer operations. The execution trace cache keeps up with the high-speed execution engine.

1.8.3. Advanced Dynamic Execution:

Regular dynamic execution technology used in previous generations of Pentium processors refers to enhanced branch prediction, data flow analysis, and speculative execution. Pentium 4 processor uses a larger 4 KB branch target buffer (versus 512-entry table in Pentium III), which stores more detail on the history of past branches, and implements an enhanced branch prediction algorithm which reduces the number of branch mal-predictions by about 33% over the P6 generation processor’s branch prediction capability. Via data flow analysis, the Pentium 4 core looks at the current set of decoded instructions and determines if they are available for processing now, or if they are dependent on more instructions. It then determines the best sequence for processing. With speculative execution, the processor executes the instructions in the most efficient way possible while looking ahead for more work to perform. The Pentium 4 is capable of viewing up to 126 instructions in flight, compared to 40 in the Pentium III and 72 in the Athlon from Advanced Micro Devices, and handling up to 48 loads and 24 stores in pipeline.

1.8.4. Enhanced Floating-Point and Multi-Media Unit:

The Pentium 4 processor expands the floating-point register to a full 128-bit and adds an additional register for data movement, separates the 128-bit floating-point move and data store port (in addition to the integer store port). This improves performance on floating-point and multi-media applications to give lifelike video and 3D graphics.
1.8.5. Internet Streaming SIMD Extensions (SSE2):

The SSE2 extends the SIMD capabilities that MMX technology and SSE technology (64-bit instructions introduced originally with SSE) delivered to the process of packed double-precision floating-point data elements and 128-bit packed integers. It introduces 144 new instructions. These instructions include SIMD 128-bit double-precision floating-point instructions, SIMD 128-bit integer arithmetic instructions, conversion instructions between floating-point and integer data, and cachability instructions. These new instructions reduce the overall number of instructions required to execute a particular program task and as a result can contribute to an overall performance increase. They accelerate a broad range of applications, including video, speech and image, photo processing, encryption, financial, engineering and scientific applications.

1.8.6. Other features and their benefits are:

The AGP (accelerated graphics port) 4X interface allows graphics controllers to access main memory at over 1 GB/s, twice that of the previous AGP platform. With RDRAM memory and the Pentium 4 processor, AGP 4X delivers high level 3D graphics performance; Two USB (universal serial bus) controllers double the bandwidth available for USB peripherals to 24 Mbps over 4 ports from 12 Mbps previously; The latest AC97 audio delivers 6 channel of audio for enhanced sound quality and full surround sound capability for live broadcast and other digital dashboard programming; LAN connection interface provides flexible network solutions such as home phone line, 10/100 Mbps ethernet, and 10/100 Mbps ethernet with LAN manageability. This simplifies network connectivity and increase ease of deployment; Dual ultra ATA/100 controllers support the fastest IDE interface for transfers to storage devices; Communication network riser card allows flexibility for multiple configurations on a single card to extend USB, LAN, and audio; Low power sleep mode saves energy; The Pentium 4 processor is manufactured based on the 0.18 micron manufacturing process. It has 48 million transistors, nearly 50% more than what Pentium III has. Smaller size helps in increasing processor core frequencies and reducing power consumption. It may shrink to 0.13 micron in the year 2001.

Unlike its previous generations, the Pentium 4 processor faces a strong competitor, the Athlon from Advanced Micro Devices, since the day it was introduced. Although, Intel has stated that the Pentium 4 was not designed to speed up standard office applications, comparisons conducted by PCWorld.com among 800 MHz Pentium III, 1.2 GHz Athlon and 1.50 GHz Pentium 4 revealed that Athlon beats Pentium 4 in almost every test such as audio file conversion test, Adobe Photoshop test, the floating-point-intensive AutoCAD, and Unreal Tournament tests. Pentium 4 pulls even with Pentium III in media encoding. The performance of Pentium 4 may improve if Intel can convince developers to optimize applications for the Pentium 4.
2. "Intel 486 Processors" intarch/intel486/index.htm
3. "Intel 386 Processors" intarch/intel386/index.htm
4. "Intel 186 Processors" intarch/intel186/index.htm
5. "Pentium Processors" intarch/pentium/pentium.htm
7. "Intel PentiumPro Processor Architecture Overview"
**Table of Contents**

1. From 8086 to 80486 [1-4]  
   1.1. Pentium [1,5,6]  
   1.2. Pentium Pro [1,6-8]  
   1.3. Pentium with MMX Technology [6,9-13]  
   1.4. Pentium II [1,14-18]  
   1.5. Pentium III [1,15,19,20]  
   1.6. Pentium II Xeon and Pentium III Xeon [21,22]  
   1.7. Celeron [23]  
1.8. Pentium 4 [24-27]  
   1.8.1. 256 KB, Level 2 Advanced Transfer Cache:  
   1.8.2. Rapid Execution Engine:  
   1.8.3. Advanced Dynamic Execution:  
   1.8.4. Enhanced Floating-Point and Multi-Media Unit:  
   1.8.5. Internet Streaming SIMD Extensions (SSE2):  
   1.8.6. Other features and their benefits are:
List of Figures

Figure 1: Three Engines Communicating Using An Instruction Pool