MODULE 5 : PENTIUM PROCESSOR

- 5.1 Pentium Architecture
- 5.2 Superscalar Architecture
- 5.3 Integer and Floating Pipeline stages
- 5.4 Branch Prediction Logic

MODULE 5 : PENTIUM PROCESSOR

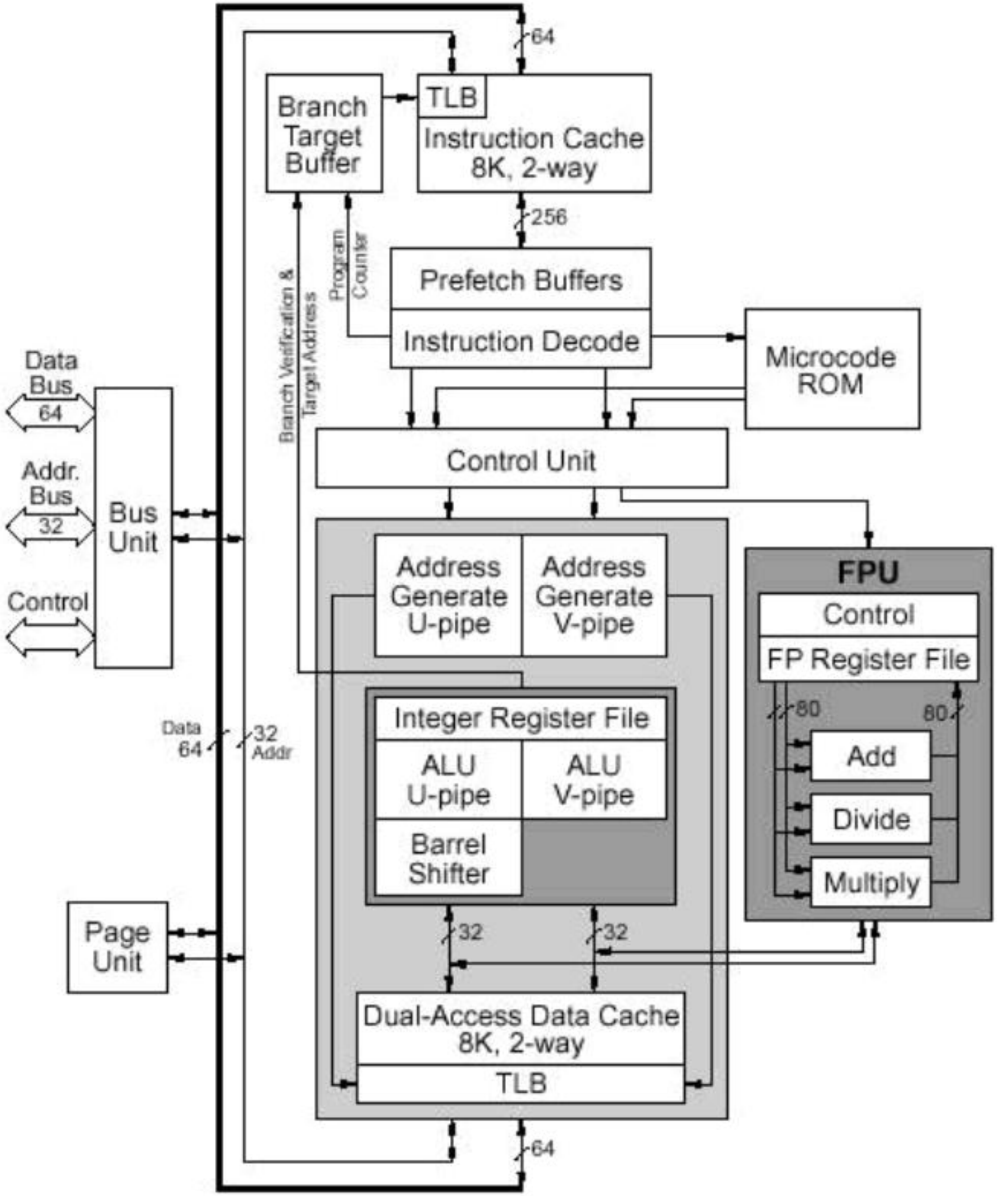
❑ Salient features of 80386

- It is a 32 bit Microprocessor.
- It has a 64 bit data bus.
- It has 8 memory banks.
- It has a 32 bit address bus.
- It can access 4 GB of physical memory.
- It has 5 Pipeline stages for integer operations.
- It has an internal Floating point unit.
- It has an 8 stage Floating point Pipeline.
- It is 2 way superscalar. This means it has two pipes called the u-pipe and the v-pipe.
- It operates on 66 MHz - 99 MHz frequency.

MODULE 5 : PENTIUM PROCESSOR

❑ Architecture :

- Pentium has a 2 way superscalar architecture giving extremely superior performance.
- It has two pipes called the u-pipe and the v-pipe. Each performs a 5-stage integer pipeline. Bus Unit1. The Bus unit is responsible for transferring data in and out of the up.2. It is connected to the external memory and I/O devices, using the system bus.



MODULE 5 : PENTIUM PROCESSOR

❑ Architecture :

- Bus Unit:
  - The Bus unit is responsible for transferring data in and out of the up.
  - It is connected to the external memory and I/O devices, using the system bus.
- L1 code cache :
  - Pentium has an on chip 8 KB LI Code cache.
  - It is 2-way It contains the most recently used instructions.
- Prefetch Unit:
  - It prefetches instructions from the L1 Code cache.
  - It has two queues each of 32 bytes.
  - One queue acts as the active queue, where as the other is used during branch prediction
- Decode Unit:
  - It decodes two instructions simultaneously for U and v pipes.
  - Simple instructions are decoded by the hardwired control unit.
  - Complex instructions are decoded by the micro programmed control unit.

MODULE 5 : PENTIUM PROCESSOR

- Integer Execution Unit
  - It can handle two integer instructions simultaneously.
  - This first one goes to u-pipe and the second to v-pipe.
  - There are address generation units for each pipe.
  - If the instruction uses memory operand the address generation unit generates physical address of the operand and fetches it form the 8 KB L1 Data Cache.
  - There are two separate ALUS for U and V Pipes.
  - The U-pipe ALU is equipped with a barrel shifter and hence can handle complex arithmetic like MUL and DIV. Both ALUs are 32-bits each.
  - The integer unit uses 32-bit integer registers like EAX. EBX etc.
- Floating point unit
  - It performs Floating Point operations.
  - It uses 80-bit F.P. Registers.
  - It has its own F.P. Control unit and independent circuits for F.P. arithmetic operations.
- Branch Prediction Logic
  - Pentium does branch prediction to minimize the pipeline penalty during branch operations
  - It uses a Branch Target Buffer with 256 entries.
  - It gives history of previous branches and helps in predicting next branch instruction.

MODULE 5 : PENTIUM PROCESSOR

❑ Pentium : Integer Pipeline stages

➢ Pentium performs integer instructions in a five-stage pipeline

- PF Prefetch
- DI Instruction Decode
- D2 Address Generate
- EX Execute - ALU and Cache Access
- WB Write-Back

MODULE 5 : PENTIUM PROCESSOR

❑ Pentium : Integer Pipeline stages

❑ Stage 1 :PF Prefetch

- ➢ Here instructions are fetched from the L1 Cache and stores them into the Prefetch queue.
- ➢ The Prefetch queue is of 32 bytes as it needs atleast two full instructions to be present inside for feeding the two pipelines, and maximum size of an instruction is 15 bytes.
- ➢ There are two Prefetch queues but only one of them is active at a time.
- ➢ It supplies the instructions to the two pipes.
- ➢ The other one is used when branch prediction logic predicts a branch to be "taken". Since the bus from L1 cache to the prefetcher is of 256 hits (32 Bytes), the entire queue can be fetched in 1 Cycle. (T State)

❑ Stage 2 :DI Instruction Decode

- ➢ The decode stage decodes the instruction opcode.
- ➢ It also checks for instruction pairing and performs branch prediction.
- ➢ Certain rules are provided for instruction pairing. Not all instructions are pairable.
- ➢ If the two instructions can be paired, the first one is given to the u pipe and the second one to the v pipe. I not, then the first one is given to the u pipe and the second one is held back and then paired with the forthcoming instruction

MODULE 5 : PENTIUM PROCESSOR

- Pentium : Integer Pipeline stages
- Stage3 :D2 Address Generate
  - It performs address generation where it generates the physical address of the required memory operand using segment translation and page translation.
  - Even protection checks are performed at this stage.
  - The address calculation is fast due to segment descriptor caches and TLB.
  - In most cases the address translation is performed in 1 cycle itself.
- Stage 4: EX Execute
  - Execution StageThe Execution stage mainly uses the ALU.
  - The U pipeline's ALU has a barrel shifter, while the V pipeline's does not.
  - Instructions involving shifting like MUL, DIV etc can only be done by U pipeline.
  - Operands are either provided by registers or by data cache (assuming a hit).
  - Both, u and v pipes can access the data cache simultaneously.
  - During execution, if the u pipe instruction stalls, the v pipe one has to also stall.
  - But if the v pipe instruction stalls, the u pipe one can continue.
- WB Write-Back
  - As the name suggests the result is written back into the appropriate registers
  - The flags are updated accordingly.

MODULE 5 : PENTIUM PROCESSOR

- Pentium : Floating Point Instruction Pipeline stages

- Stage Description
  - Prefetch: Identical to the integer prefetch stage.
  - Instruction De-code 1: Identical to the integer D1 stage
  - Instruction De-code 2: Identical to the integer D2 stage.
  - Execution Stage (Ex): Register read, memory read, or memory write performed as required by theinstruction (to access an operand).
  - FP Execution 1 stage: Information from register or memory is written into a FP register. Data is converted to floating-point format before being loaded into the floating-point unit.
  - FP Execution 2 stage: Floating-point operation performed within floating-point unit.
  - Write FP Result: Floating-point results are rounded and the result is written to the target floating-point register.
  - Error Reporting: If an error is detected, an error reporting stage is entered where the error is reported and the FPU status word is up-dated.

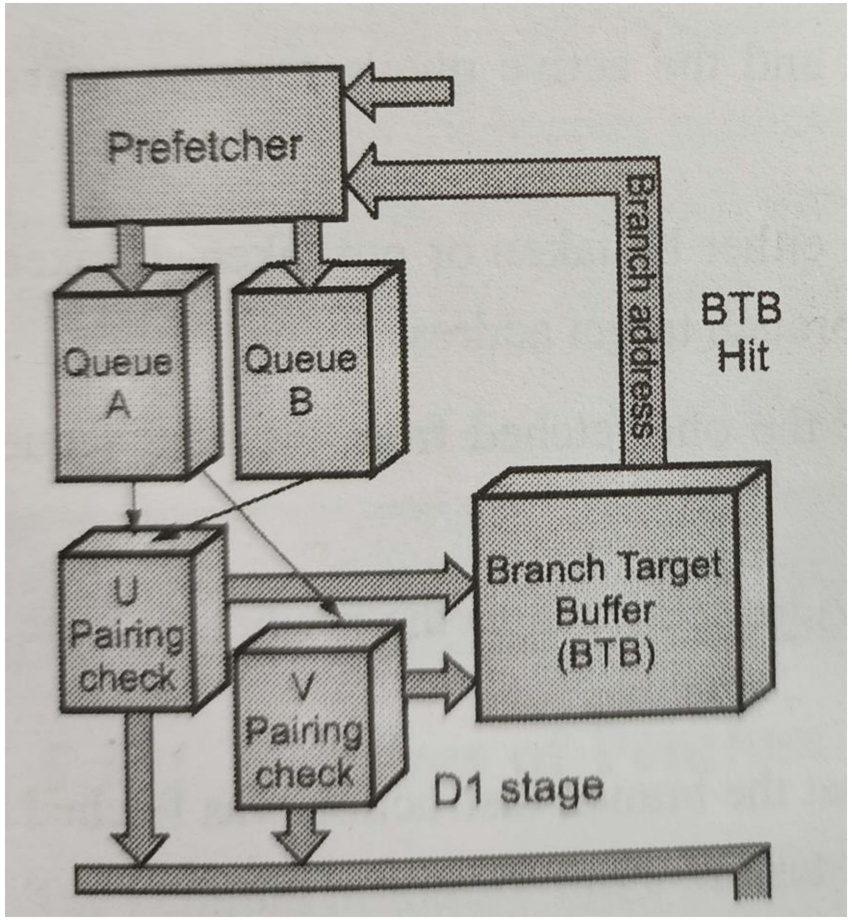
Most floating point instructions are issued singly to the U pipeline and cannot be paired with integer instructions.
It consists of eight pipeline stages.
The first four stages are shared with integer pipeline and the last four reside within the floating point unit

MODULE 5 : PENTIUM PROCESSOR

❑ Pentium : Branch Prediction Logic

❑ DI Instruction Decode

- ➢ The Pentium processor includes branch prediction logic, allowing it to minimize pipeline flushing
- ➢ When a branch operation is correctly predicted, no performance penalty is incurred.
- ➢ However, when branch prediction is not correct, a three cycle penalty is incurred if the branch is executed in the U pipeline and upto four (3+1 extra may be needed) cycle penalty if the branch is in the V pipeline.
- ➢ The prediction mechanism is implemented using a four-way, set-associative Cache with 256 entries.
- ➢ This is referred to as the Branch Target Buffer, or BTB.
- ➢ The directory entry for each line contains the following information.
  - ▪ A valid bit that indicates whether or not the entry is in use.
  - ▪ Two History bits that-track how often the branch has been taken each time that it entered the pipeline before
  - ▪ The Memory Address of the branch instruction for identification.
- ➢ The Branch Target Buffer, or BTB, is a look-aside cache that sits off to the side of the DI stages of the two pipelines and monitors for branch instructions.
- ➢ During D1 stage, when an instruction is decoded and identified as a branch instruction, the address of the instruction is searched in the BTB for a previous history.



MODULE 5 : PENTIUM PROCESSOR

❑ Pentium : Branch Prediction Logic

❑ DI Instruction Decode

- If no history exists, then prediction is made that the branch will not be taken.
- If there is a history (BTB hit), then prediction is made as follows:
  - If the History bits are 00 or 01 (Strongly Not taken or weakly not taken), then the prediction is that the branch will not be taken.
  - If the History bits are 10 or 11 (Strongly taken or weakly taken), then the prediction is that the branch will be taken.
- If the branch is predicted to be taken, then the active queue is no longer used. Instead, the prefetcher starts fetching instructions from the branch address and stores them into the second queue which now becomes the active queue. This queue now starts feeding instructions into the two pipes.. I
- f branch is predicted to be not taken, then nothing changes, and the active queue remains active and instructions are fetched from the sequentially next locations.
- When the instruction reaches the execution stage, the branch will either be taken or not taken. If taken, the next instruction to be executed should be the one fetched from the branch target address,
- If the branch is not taken the next instruction executed should be the one fetched from the next sequential .memory address after the branch instruction.
- When the branch is taken for the first time, the execution unit provides feedback to the prediction logic. The branch target address is sent back and recorded in the BTB.
- A directory entry is made containing the source memory address that the branch instruction was fetched from and history bits are set to indicate that the branch has been strongly taken.

MODULE 6 : PENTIUM 4

- 6.1 Pentium 4 : Intel NetBurst Architecture
- 6.2 Pipelining in NetBurst Architecture
- 6.3 Hyper threading technology and its use in pentium

MODULE 6 : PENTIUM 4

❑ Pentium 4 Architecture :

❑ Features :

- ➢ Specifically designed for high-end performance on bandwidth-hungry Internet and advanced applications
- ➢ Ideal solution for sophisticated end users working with complex Internet, imaging, video, vpeech, JD,and multimedia applications.
- ➢ Based on Intel NetBurst microarchitecture
- ➢ 400-MHz system bus.
- ➢ Hyper-pipelined technology Q Advanced dynamic execution
- ➢ Rapid execution engine
- ➢ Advanced transfer cache
- ➢ Execution trace cache
- ➢ Streaming SIMD (Single Instruction, Multiple Data) Extensions 2 (SSE2)

- 400-MHz System Bus: The Intel Pentium 4 processor features a 400-MHz system bus (the 100-MHz clock is quad-pumped that provides 3X bandwidth over the 133-MHz system bus in the Intel Pentium III processor (1.06 GB/s).
- Hyper-pipelined Technology : With the Intel Pentium 4 processor, Intel has doubled the pipeline depth to 20 stages, enabling a higher clock frequency.
- Hyper Threading (HT) Technology: Hyper-Threading (HT) Technology was developed to improve the performance of 1A-32 processors when executing multi-threaded operating system and application code or single-threaded applications under multi-tasking environments.
- Advanced Dynamic Execution: The Advanced Dynamic Execution engine is a very deep, out-of-order speculative execution engine that keeps the execution units executing instructions. It does so by providing a very large window of instructions from which the execution units can choose. The large out-of-order instruction window allows the processor to significantly reduce stalls that can occur while instructions are waiting for dependencies to resolve.

MODULE 6 : PENTIUM 4

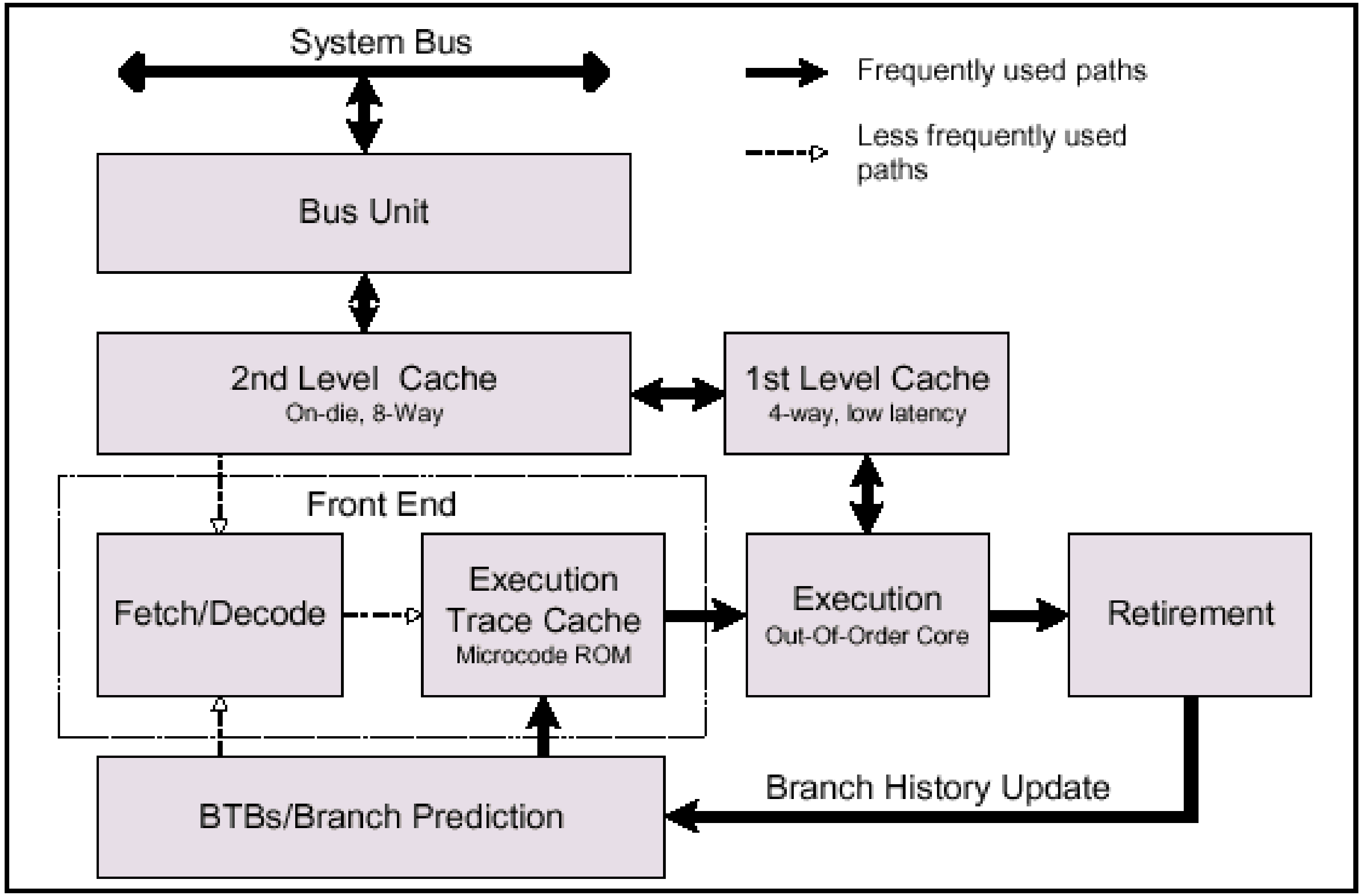
❑ Pentium 4 Architecture :

❑ Features :

- Rapid Execution Engine: The two Arithmetic Logic Units (ALUS) in the Intel Pentium 4 processor run at twice the core frequency of the processor. This makes it possible to execute basic integer instructions (such as add. subtract, logical AND, and logical OR) in half a clock cycle, with higher execution throughput and reduced latency of execution.
- Revolutionary Cache Subsystem: In order to increase performance and scalability, the Intel Pentium 4 processor features an innovative new cache subsystem designed to optimize data transfer to the core. An execution trace cache stores12K decoded instructions in the order of program flow instead of predecoded instructions that can not take code branches into consideration.
- Streaming SIMD Extensions 2 (SSE2): To make the SIMD instruction set even more powerful, the Intel Pentium 4 processor provides 144 new performance improving instructions, including 128 bit SIMD double precision floating point, 128-bitSIMD integer, and improved cache and memory management instructions.

MODULE 6 : PENTIUM 4

❑ Pentium 4 : Inetel NetBurst Architecture



MODULE 6 : PENTIUM 4

❑ Pentium 4 : Inetel NetBurst ArchitectureA

- ➢ This microarchitecture pipeline is made up of three sections:The front end pipeline.The out-of-order execution core The retirement unit.
- ➢ The concept behind the Intel NetBurst microarchitecture (Pentium 4 processor, Intel Xeon processor), was to improve the throughput, improve the efficiency of the out-of-order execution engine, and to create a processor that can reach much higher frequencies with higher performance relative to the P5 and P6 microarchitectures, while maintaining backward compatibility.
- ➢ The Intel NetBurst microarchitecture addressed some of the common problems found in high-speed,pipelined microprocessors.
- ➢ Limiting factors for processor performance were delays from pre-fetch and decoding of the instructionsto micro-operations, the efficiency of the branch prediction algorithm, and cache misses.
- ➢ The execution trace qache addresses these problems by storing decoded IA-32 instructions.
- ➢ Instructions are fetched and decoded by a translation engine, which builds the decoded instruction into sequences of uops called traces, which are then stored in the trace cache.
- ➢ The execution trace cache stores these uopsiri the path of predicted program execution flow, where theresults of branches in the code are integrated into the same cache line.
- ➢ This increases the instruction flow from the cache and makes better use of the overall cache storagespace since the cache no longer stores instructions that are branched over and never executed.

MODULE 6 : PENTIUM 4

❑ Pentium 4 : Inetel NetBurst ArchitectureA

- ➢ The trace cache delivers up to three uops per clock to the core. Branch targets are predicted based on their linear address using branch prediction logic and fetched assoon as possible.
- ➢ Branch targets are fetched from the execution trace cache if they are cached there; otherwise, they are fetched from the memory hierarchy.
- ➢ The translation engine's branch prediction information is used to form traces along the most likelypaths.
- ➢ The core's ability to execute instructions out of order remains a key factor in enabling parallelism.
- ➢ The processor employs several buffers to smooth the flow of uops.
- ➢ This implies that when one portion of the entire processor pipeline experiences a delay, that delay may be covered by other operations executing in parallel (for example, in the core) or by the execution of u.ops which were previously queued up in a buffer (for example, in the front end).
- ➢ The NetBurst microarchitecture adds further improvements to the execution units over that of the P6 microarchifecture. For example, the arithmetic logic units operate twice as fast as previous microarchitectures.
- ➢ As with the previous implementations, the retirement section receives the results of the executed uops from the execution core and processes the results so that the proper architectural state is updated according to the original program order.
- ➢ For semantically correct execution, the results of IA-32 C instructions must be committed in original program order before they are retired.

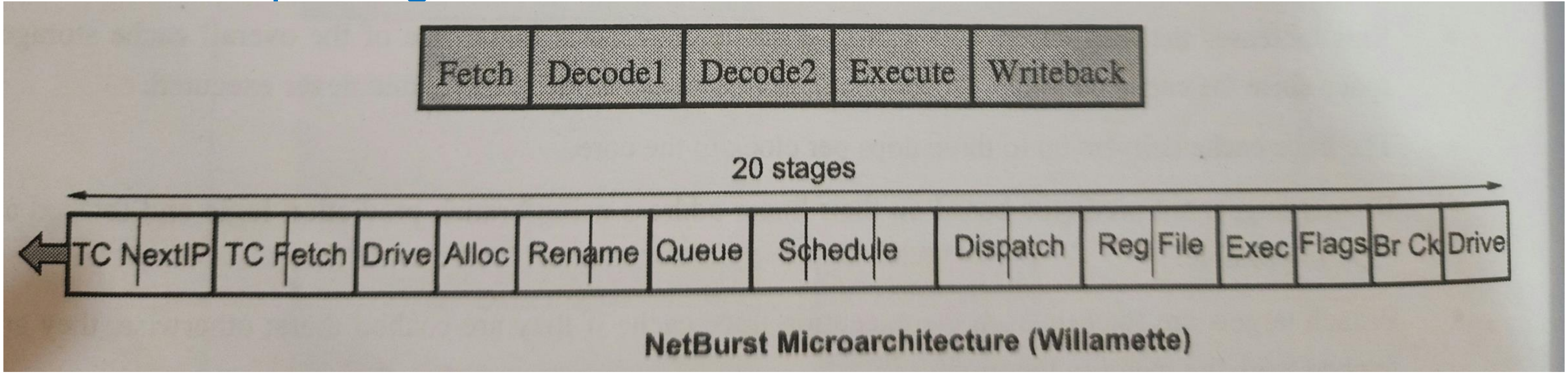
MODULE 6 : PENTIUM 4

Pentium 4 : Intel NetBurst Architecture

- Exceptions may be raised as instructions are retired.
- Thus, exceptions cannot occur speculatively, they occur in the correct order, and the machine can be correctly restarted after an exception.
- When a uop completes and writes its' result to the destination, it is retired. Up to three uops may be retired per cycle.
- Again, the ROB is the unit in the processor which buffers completed uops, updates the architectural state in order, and manages the ordering of exceptions, the retirement section also, keeps track of branches and sends updated branch target information to the branch target buffer (BTB) to update branch history.

MODULE 6 : PENTIUM 4

❑ Pentium 4 : Pipelining in NetBurst Architecture



NetBurst Microarchitecture (Willamette)

- Above figure shows that the NetBurst pipeline is 4 times as deep as that of the P5's. Increasing pipeline depth increases logic complexity and branch penalties, but it also allows clock speeds to increase.
- Trace Cache next Instruction Pointer: The trace cache fetch logic gets a pointer to the next instruction in the trace cache. Trace cache is intel's name for putting the LI cache inside of the first functional unit for speed.
- Trace Cache Fetch: use pointer to fetch an instruction from the cache.
- Drive: The two Drive stages shown in figure represent time required to move signals across the chip. No other work is done during these stages. NetBurst is the first pipeline with dedicated stages for wiredelays. This is apparently necessary for multiple-gigahertz speeds.

MODULE 6 : PENTIUM 4

❑ Pentium 4 : Pipelining in NetBurst Architecture

- Allocate and Rename: The CPU actually contains more registers than are related in the specification inorder to speed things up and be able to execute operations in a superscalar fashion (which is to say, more than one operation at once). At this time, the CPU will associate different registers with the names more of the registers.
- Queue: Operations are now placed into either the memory queue or the arithmetic (everything else)queue for scheduling.
- Schedule: In a superscalar processor, operations are often executed out of order so that they do not step on each other and so that they are completed as rapidly as possible. The P4 hasfour queues: Memory. fast ALU, Slow ALU/General FPU, and Simple FP. All instructions get dumped into one of them for later execution by the appropriate (and linked) functional instructions are waiting on them, and the number, of cycles required by an ALU (or FPU, or LSU) to complete the instruction.
- Dispatch: Instructions are moved from the queues to the functional units.
- Register Files: The instructions are now loaded into the functional units for actual execution.
- Execute: The functional units process the -instructions in the files along with data in the registers. This is the seventeenth stage. A lot has happened before we got here.
- Flags: There is a status register (sometimes called a flag register) in all CPUs of the x86 family which is used for conditional jumps. The flags are set in this stage.

MODULE 6 : PENTIUM 4

❑ Pentium 4 : Pipelining in NetBurst Architecture

- Branch Check: Now in the nineteenth out of twenty stages We finally check to see if the branch predictor - predicted incorrectly and we have to discard some operation we have just spent eighteen stages (and a cycle) on.

In other words, the Pentium 4 is split up into 20 very short pipeline stages. Some of them are so short that they aren't long enough to fit an entire function, and so that function actually takes two stages. Chopping execution up into so many short stages means that very high clock rates can be achieved, but at the same time huge penalty if branch prediction goes wrong.

MODULE 6 : PENTIUM 4

❑ Pentium 4 : Hyper threading technology and its use in Pentium

- Hyper-Threading (HT) Technology was developed to improve the performance of IA-32 processors when executing multi-threaded operating system and application code or single-threaded applications under multi-tasking environments.
- The technology enables a single physical processor to execute two or more separate code streams (threads)concurrently using shared execution resources. HT Technology is one form of hardware multi-threadingcapability in IA-32 processor families.
- It differs from multi-processor capability using separate physically distinct packages with each physical processor package mated with a physical socket. HT Technology provides hardware multi-threading capability with a single physical package by using shared execution resources in a processor core. Architecturally, an IA-32 processor that supports HT.
- Technology consists of two or more logical processors, each of which has its own LA-32 architectural state.Each logical processor consists of a full set of IA-32 data registers, segment registers, control registers, debug registers, and most of the MSRs.
- Each also has its own advanced programmable interrupt controller (APIC). Because of HT technology. OS gets illusion as if there are two (logical) processors executing code in parallel to give higher performance.

ALL THE BEST