INTEL 80386 MICROPROCESSOR

- ➔ 80386 is a 32 bit microprocessor
- ➔ It is improved version of 80286
- ➔ Software compatibility with 8086,80186 and 80286

Improvements in 80386 over 80286

- ➔ Registers and ALU are 32 bits wide
- ➔ The instruction set is extended to support 32 bit addresses and data
- ➔ The maximum size of physical memory is extended from 16 MB to 4 Gb
- ➔ Faster execution speed as 80386 runs at higher clock frequency

On-chip memory management support paging

Features of 80386:

- ➔ 80386 is available in two versions : 80386DX & 80386SX.(*Internal architecture of both versions are same,differ only in external address and data bus*)
- ➔ The 80386 is available with maximum clock speed rating of 12.5,16,20,25 or 33 Mhz

80386 SX	**80386 DX**
It has 24 bit address bus and 16 bit data bus	32 bit address bus and 32 bit data bus
80386SX is called reduced bus version of 80386	80386DX is called full version of 80386
It can address up to 16 MB	It can address up to 4 Gb
It is available in 100 pin flat Package	It is available in 132 pin PGA (Pin Grid Array) Package

Operating Modes of 80386

80386 has 3 operating modes :-

i. Real address mode
ii. Protected Virtual Address mode
iii. Virtual 8086 (V86) mode

After a hardware reset the processor will start working in th real address mode

Real Address Mode

The 80386 processor will enters in real address mode when it is resetted.

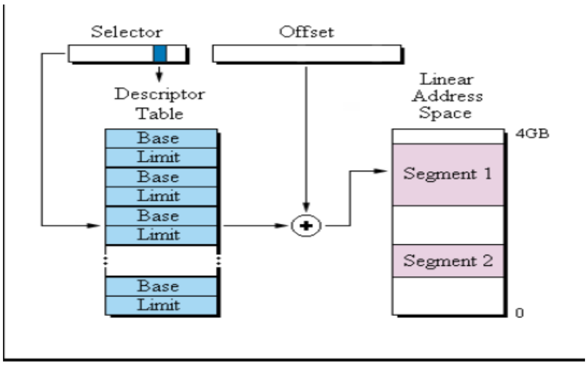
- ➔ In this mode 80386 will work similar to 8086 processor but can access 32 bit registers
- ➔ Paging is disabled in this mode
- ➔ 80386 can only access 1 Mb of physical memory like 8086 in this mode
- ➔ **20 bit physical memory is computed by shifting segment address to left 4 times and adding it to offset address**
- ➔ Only some instructions in 80386 can be executed in this mode
- ➔ Primary purpose of this mode is to set up 80386 for protected virtual address mode

Protected Virtual Address Mode

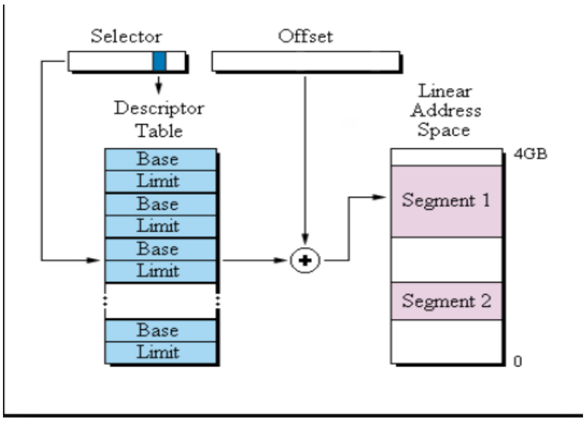
The 80386 can switch from real address mode to PVAM by setting PE bit (Protection Enable bit)

- In PVAM the processor can run all 8086 and 80286 programs
- In this mode the processor has 4 Gb of physical memory address space and 64 Tb of virtual memory address space(The virtual memory does not exist physically it still appears to be available within the system)
- The concept of VM is implemented using Physical memory that the CPU can directly access and secondary memory that is used as a storage for data and program, which are stored in secondary memory initially
- 80386 operates in two memory management modes in PVAM
- Non-Paged Mode
- Paged Mode
- paging is a memory management scheme by which a computer stores and retrieves data from secondary sttorage for use in main memory.In this scheme,the memory is divided in to fixed size pages and stores in secondary memory,when a program needs a page the OS copies certain number of pages from secondary memory to main memory.Paging is an important part of virtual memory implementations in modern operating systems.
- In PVAM virtual address can be 48 bit or 32 bit
- In non-paged mode the pysical address is computed using the selector,descriptor and offset like 80286
- In paged mode a linear address is computed using selector,descriptor amd offset,later this linear address is converted to physical address by the paging unit

Address Computation in PVAM when paging is disabled



Address Computation in PVAM when paging is enabled



Virtual Mode

- The processor can switch from PVAM to Virtual Mode by setting VM bit

- Virtual Mode permits the execution of 8086 applications with all protection features of 80386
- Virtual Mode is similiar to Real Mode
- So also in this mode the processor computes 20 bit address by shifting the segment register left by 4 times and adding it to offset
- In Virtual Mode paging can be enabled

ARCHITECTURE OF 80386

- The six functional units are: a)Bus Interface Unit

b)Code Prefetch Unit c)Instruction decode unit

d)Segmentation Unit e)Paging Unit f)Execution Unit

a)Bus Interface Unit: It interfaces with the m/m and I/O.

b)Code Prefetch Unit: It is used to prefetch code from m/m & store it in 16 bit temporary code queue.

The queue acts as a buffer between prefetch unit & the instruction decode unit.

c)Instruction decode unit :It translates the instructions from the code prefetch unit into microcodes.

d)Segmentation unit:It produces a translated linear address which paging unit translates into physical addresss.

e)Paging unit:It checks for paging violations before it sends a bus request and the address to BIU & external bus.

- **Execution Uni**t:It operates on the decoded instruction, performing the steps needed to execute it.
- It contains **control unit,data unit and protection test unit**
- **Control Unit** contains microcodes and parallel hardware for fast multiply,divided and effective address calculation.
- **Data Unit** includes ALU,8general purpose registers and 64 barrel shifter for performing multiple bit shift in one clock.

- **Protection test unit** checks for segmentation violation under the control of microcode

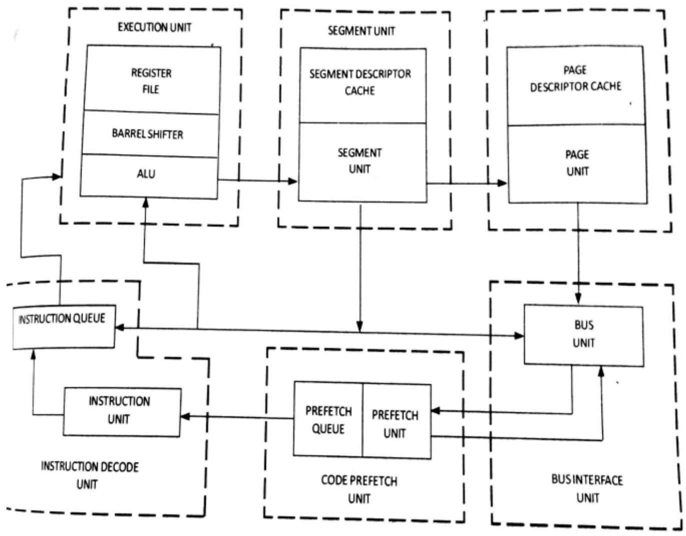


Fig. 4.1 Internal Block Diagram of 80386

Pentium Microprocessor

Pentium processor features:

- ➔ Superscalar architecture
- ➔ 64 bit data bus
- ➔ 32 bit address bus which can address up to 4 Gb of physical memory
- ➔ Dynamic branch predictions
- ➔ It consists of 3.1 million transistors
- ➔ It is available with maximum clock speed of 60-233 Mhz
- ➔ Dual processing support
- ➔ Bus cycle pipielining
- ➔ Advanced power management features
- ➔ The Pentium processor has two separate 8-kilobyte (KB) caches on chip, one for instructions and one for data

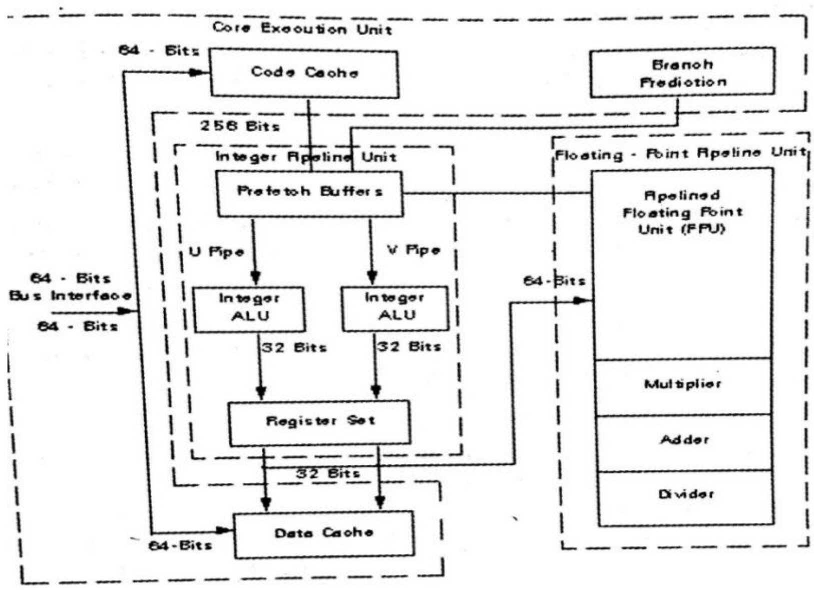


Fig. 4.2 Pentium Architecture

ARCHITECTURE OF PENTIUM

- It consists of 3 blocks: **core execution unit, integer pipeline unit and floating-point pipeline unit.**
- **External data bus is of 64-bits.**
- **CORE EXECUTION UNIT:This unit contains code cache,data cache and branch prediction**
- **Pentium uses a separate 8kB data cache and 8KB code cache.**
- **Efficient branch prediction requires that the destination of a branch can be accessed simultaneously with data reference of the previous instruction executing in the pipeline.**
- **A branch target buffer in Pentium CPU holds branch target addresses for previously executed branches.**
- **2.Intger pipeline unit:It has 5stages:prefetch,instruction decode,address generation,execution and write-back**
- **Prefetch:**CPU fetches the instructions from instruction cache,which stores the instructions to be executed**.**
- **Instruction Decode**:In this stage,the CPU decodes the instructon and generate a control word.
- **Address** Generation:The Control word from previous stage is again decoded for final execution and the CPU generates addresses for data m/m references.
- **EXECUTION**:The CPU either accesses the data cache for data operands or computes the result in ALU.
- **Write** back:CPU updates the registers contents or the status in the flag register depending on the execution result.
- **Floating point pipeline unit:The** functions of this unit are computationally expensive divide function with hard-wired implementation and speeding up function.
- It consists of 8 stages-prefetch,instruction decode,address generation,operand fetch,first execute,second execute,write float and error reporting stages.
- **4.64-bit Bus Interface**:It is used to interface external i/o devices with Pentium processor using 64-bit data bus.

Pipelining in Pentium Processor

The execution unit has two parallel pipelines named U-Pipe and V-Pipe with individual ALU for each pipe.

- Pipeline has 5 stages
- They are Prefetch (PF),Decode Stage-1 (D1),Decode Stage-2 (D2),Execute (E) and Writeback (WB)
- U is the default pipeline and slightly more capable than V-pipeline
- The U-pipeline can handle any instructions while the V-Pipeline can handle simple instructions

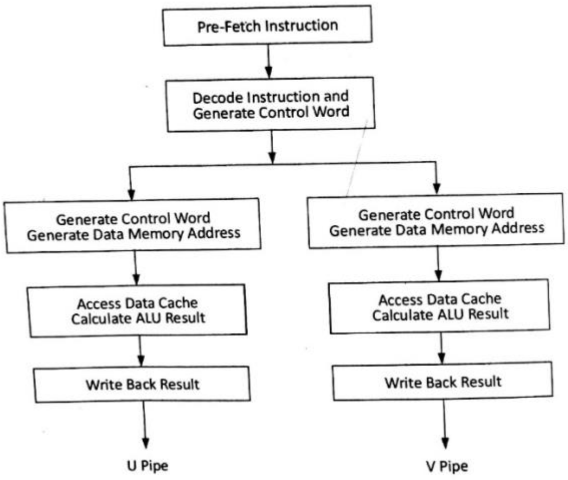


Fig. 4.3 Pentium Pipeline Stages

- **PREFETCH**:Instructions are fetched from instruction cache and aligned in prefetch buffer for decoding
- Decode 1:Instructions are decoded into Pentium's internal instruction format.Branch prediction also takes place at this stage.
- Decode 2:addresss computation takes place at this stage.
- Execute:The Integer hardware executes the instruction.
- Write-back:The results of computation are written back into register

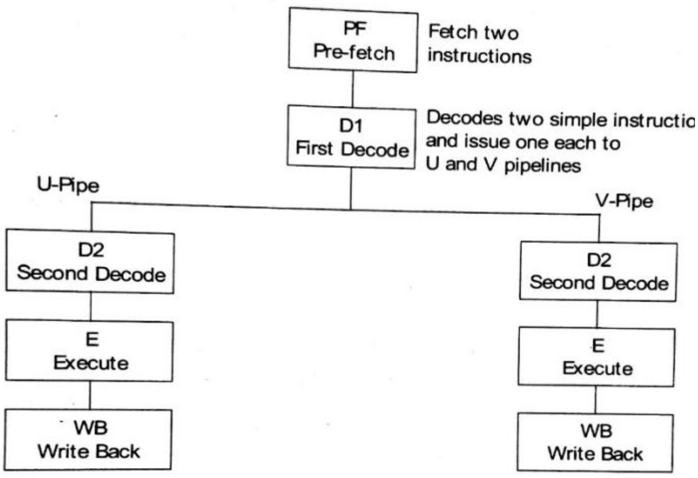


Fig. 4.4 Superscalar Execution in Pentium

SUPERSCALAR ARCHITECTURE

- The Pentium supports a two-way superscalar architecture with two integer pipelines, simultaneously executing two consecutive instructions.
- The stages PF and D1 are common to both U and V pipelines
- The D1 uses various techniques to decide if the given two consecutive instructions can be executed simultaneously considering the type of instructions and dependencies between them.
- While executing instructions the processor checks next two instructions.

- If the execution of one instruction does not depend on the other then the first instruction is issued to U-Pipe and the second instruction to V-Pipe. Thus two instruction is executed simultaneously.
- If it's not possible to execute two instructions simultaneously then two instructions are issued to U-pipe one by one

Pipe Line Hazards

Hazards are situations in pipelining which prevents the next instruction in the instruction stream from executing during the designated clock pulses

- Hazards prevents the ideal speed up gained by pipelining and are classified in to three classes :-

a) Structural Hazards

b) Data Hazards

c) Control Hazards

- Structural Hazards. They arise from resource conflicts when the hardware cannot support all possible combinations of instructions in simultaneous overlapped execution.
- Data hazards occur when instructions that exhibit data dependence modify data in different stages of a pipeline. Ignoring potential data hazards can result in race conditions (also termed race hazards). There are three situations in which a data hazard can occur:
- read after write (RAW), a *true dependency*
- write after read (WAR), an *anti-dependency*
- write after write (WAW), an *output dependency*
- **Control Hazards**.They arise from the pipelining of branches and other instructions that change the PC.

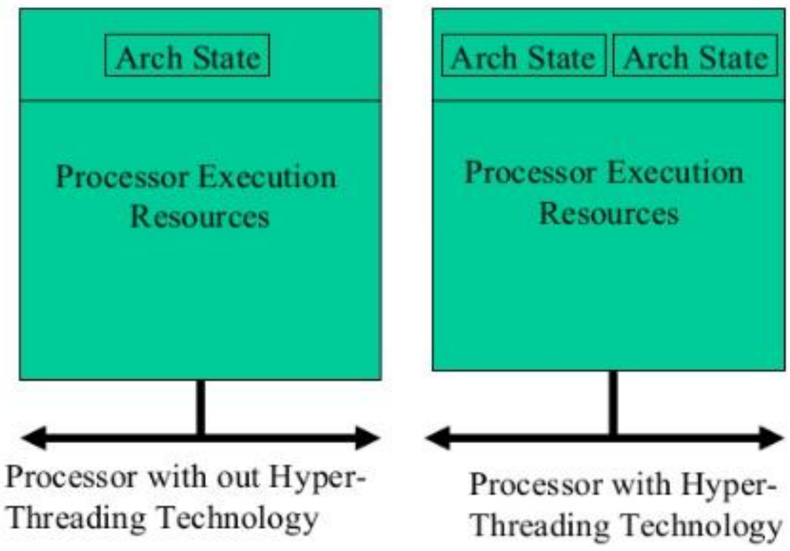
Hyperthreading Technology(HTT)

- Hyperthreading provides two logical processor in a single processor,ie a single processor acts like multiple processor
- HTT allows processor to work more efficiently.
- HTT enables different part of processor to work on different tasks concurrently.
- It divides workload in to processes and threads.
- HTT make use of resources that would sit idle and done more work in same amount of time.
- It is first introduced in INTEL Xeon processor

Hyper-Threading Technology Architecture

Each logical processor maintains one copy of the architecture state

Hyper-Threading Technology Architecture



Following resources are duplicated to support Hyper-Threading Technology

- Register Alias Tables
- Next-Instruction Pointer
- Instruction Streaming Buffers and Trace Cache Fill Buffers
- Instruction Translation Look-aside Buffer

MMX -(Multi Media Extension) Processor

The acronym "MMX" stands for MultiMedia extensions

- ➔ MMX is a Pentium microprocessor from Intel that is designed to run faster when playing multimedia applications.
- ➔ A PC with an MMX microprocessor runs a multimedia application up to 60% faster than one with a microprocessor having the same clock speed but without MMX.
- ➔ In addition, an MMX microprocessor runs other applications about 10% faster, because of increased cache.
- ➔ All of these enhancements are made while preserving compatibility with software and operating systems developed for the Intel Architecture.

The MMX technology consists of following improvements over the non-MMX Pentium microprocessor:

I. 57 new instructions have been added that are designed to handle video, audio, and graphical data more efficiently.
II. 4 new data types

III.Eight 64 bit wide mmx registers-MM0 to MM7

A new process, Single Instruction Multiple Data (SIMD), makes it possible for one instruction to perform the same operation on multiple data items.

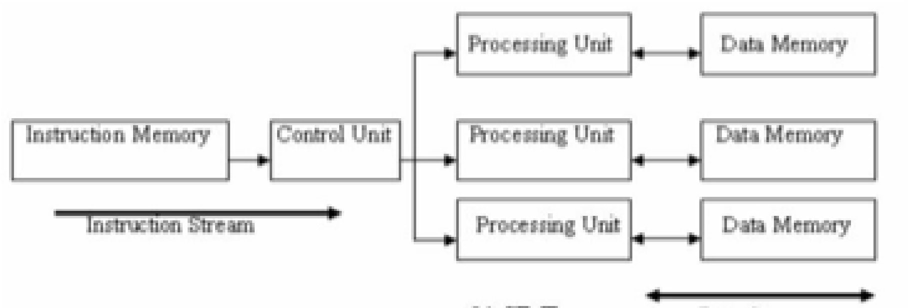
I. The memory cache on the microprocessor has increased to 32 KB

The basis of MMX technology is SIMD

Single Instruction Multiple Data

- In SIMD one instruction can process multiple data items.
- SIMD is useful when large amount of regularly organised data is processed.
- SIMD is applying same instructions to multiple data.
- SIMD instructions can greatly increase performance when exactly same operation is performed in multiple data objects.

Single Instruction Multiple Data



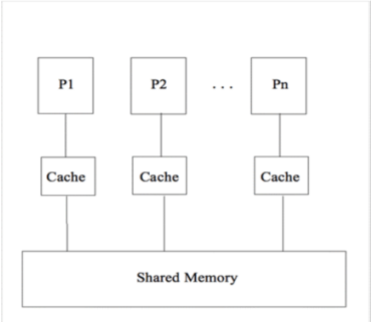
(b) SIMD

SSE

SSE is the acronym for Streaming SIMD Extensions
- SSE is a set of instruction dedicated to applications like signal processing,3D Graphics etc
- SSE contains 70 new instructions most of which work on single precision floating point data.
- SSE added 8 new 128 bit registers XMM0 to XMM7.
- There is also a new 32 bit control/status register named MXCSR.

Multicoreprocessor System
- Performance requirement of applications such as weather prediction,radar tracking,signal processing etc exceeds the capability of single processor architecture.
- A computer system which has two or more processors working simultaneously and sharing the same hard disk, memory and other memory devices

- Fig shows a multiprocessor system(P1,P2 etc up to Pn represents processors).It is clear that in Multiprocessor system memory is shared by each processor and they have different cache.



- **A Multi-core processor** is an integrated circuit in which two or more processors have been attached for enhanced performance, reduced power consumption and more efficient simultaneous processing of multiple tasks.
- The concept of multicore technology is mainly centered on the possibility of parallel computing.
- Architecture of multicore processor enables communication between all available cores to ensure that processing tasks are divided and assigned accurately.

Advantages

Reduced Cost: Multiple processors share the same resources (like power supply and mother board).

Increased Reliability: The failure of one processor does not affect the other processors though it will slow down the machine provided there is no master and slave processor.

Increased Throughput:An increase in the number of processes completes the work in less time.

Major issues in multicore processing

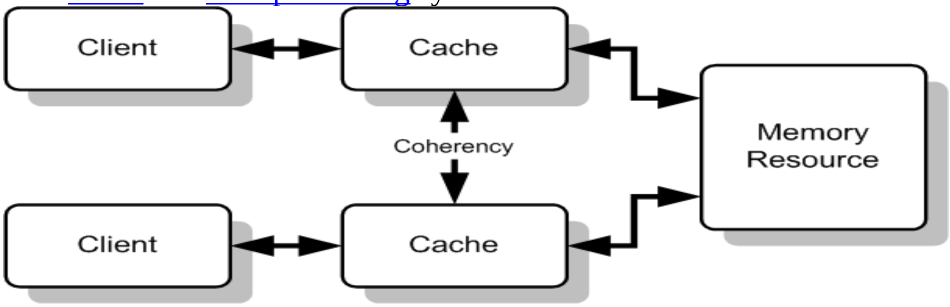
Interconnect issue

- Since there are so many components on chip in a multicore processor like cores, caches, network controllers etc., the interaction between them can affect the performance if the interconnection issues are not resolved properly
- In the initial processors, bus was used for communication between the components.
- In order to reduce the latency, crossbar and mesh topologies are used for interconnection of components.
- Also, as the parallelism increases at the thread level, communication also increases off-chip for memory access, I/O etc.
- For issues like this, packet based interconnection is actively used. This packet based interconnection has been used by Intel .

cache coherence

- cache coherence is the uniformity of shared resource data that ends up stored in multiple local caches.
- When clients in a system maintain caches of a common memory resource, problems may arise with incoherent data, which is particularly the case

with CPUs in a multiprocessing system.



- consider both the clients have a cached copy of a particular memory block from a previous read.
- Suppose the client on the bottom updates/changes that memory block, the client on the top could be left with an invalid cache of memory without any notification of the change.
- Cache coherence is intended to manage such conflicts by maintaining a coherent view of the data values in multiple **caches.**

BUS-SNOOPING PROTOCOL, a protocol for maintaining cache coherency in symmetric multiprocessing environments.

In a **directory based coherence approach**, the information about which caches have a copy of a block is maintained in a structure called *Directory*.

A directory approach can result in a substantial traffic saving compared to broadcast/snoopy approach in such applications