PART 2

SUBMITTED BY: DR. SANGEETA ARORA
(DEPT. OF COMPUTER SCIENCE AND IT)

ADVANCED COMPUTER ARCHITECTURE

PARALLEL PROCESSING MECHANISM

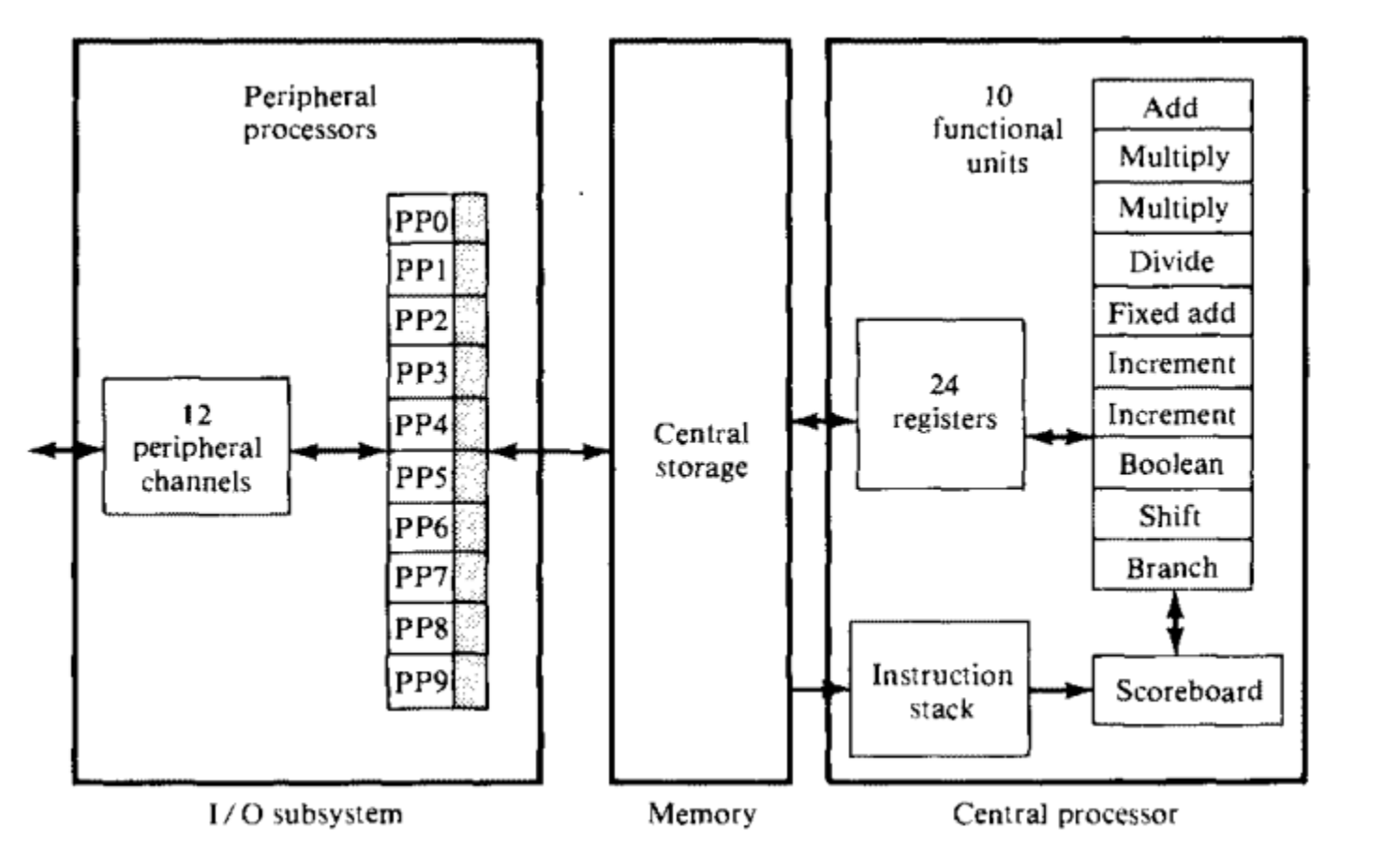
Parallel processing mechanism have been developed in uniprocessor computer:

- Multiplicity of functional unit
- Parallelism and pipelining within the CPU
- Overlapped CPU and I/O operations
- Use of hierarchical memory system
- Balancing of subsystem bandwidth
- Multiprogramming and time sharing

MULTIPLICITY OF FUNCTIONAL UNIT

- In the early year computer had only one arithmetic and logic unit in its CPU.
- The ALU could perform one function at a time
- Many of the function of the ALU can be distributed to multiple and specialized functional unit which can operate in parallel.

EXAMPLE OF CDC-6000



PARALLELISM AND PIPELINING WITH IN THE CPU

- Pipelining is a method of increasing system performance and throughput.
- Parallel adders using such techniques as carry – look ahead and carry save are now built into almost all ALUs.
- The use of multiple functional units is a form of parallelism with the CPU.
- Various phases of instruction execution are:

VARIOUS PHASES OF INSTRUCTION EXECUTION ARE-

- INSTRUCTION PHASE
  - STEP 1: fetch instruction
  - STEP2: decode instruction
- EXECUTION PHASE
  - STEP3: execute instruction
  - STEP4: store result

OVERLAPPED CPU AND I/O OPERATIONS

- I/O operations can be performed simultaneously with the cpu computations by using separate I/O controller ,channels or input output processor.
- By freeing up CPU cycles while devices are serving requests, CPU-bound processes can be executed concurrently with I/O-bound processes.
- For example, if process A is CPU-bound, and process B is I/O-bound, the system as a whole can reach high utilization by overlapping CPU and I/O processing effectively.

EXAMPLE

Process A

Loop:

90 msec of CPU

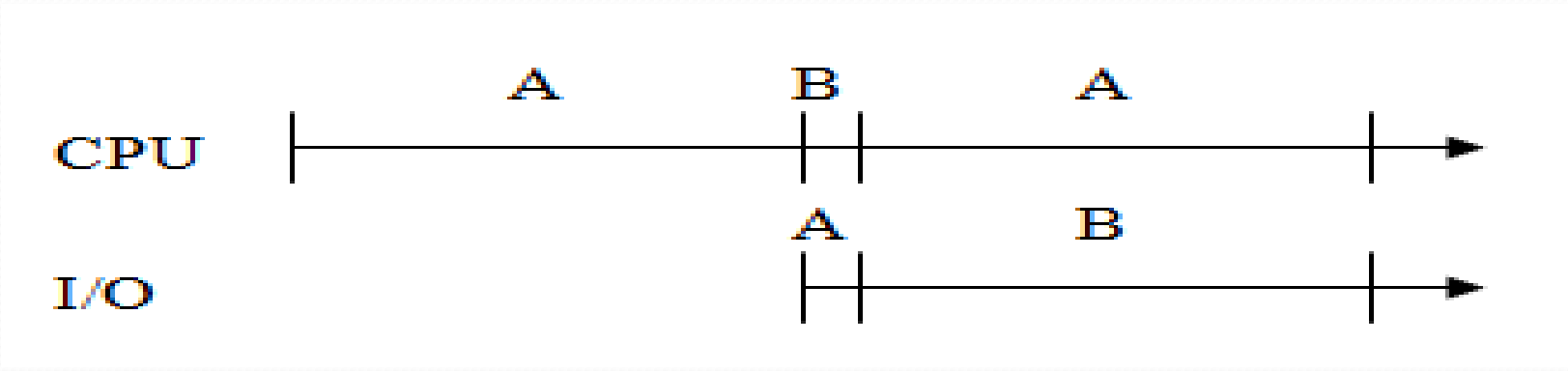
10 msec of I/O

Process B

Loop:

10 msec of CPU

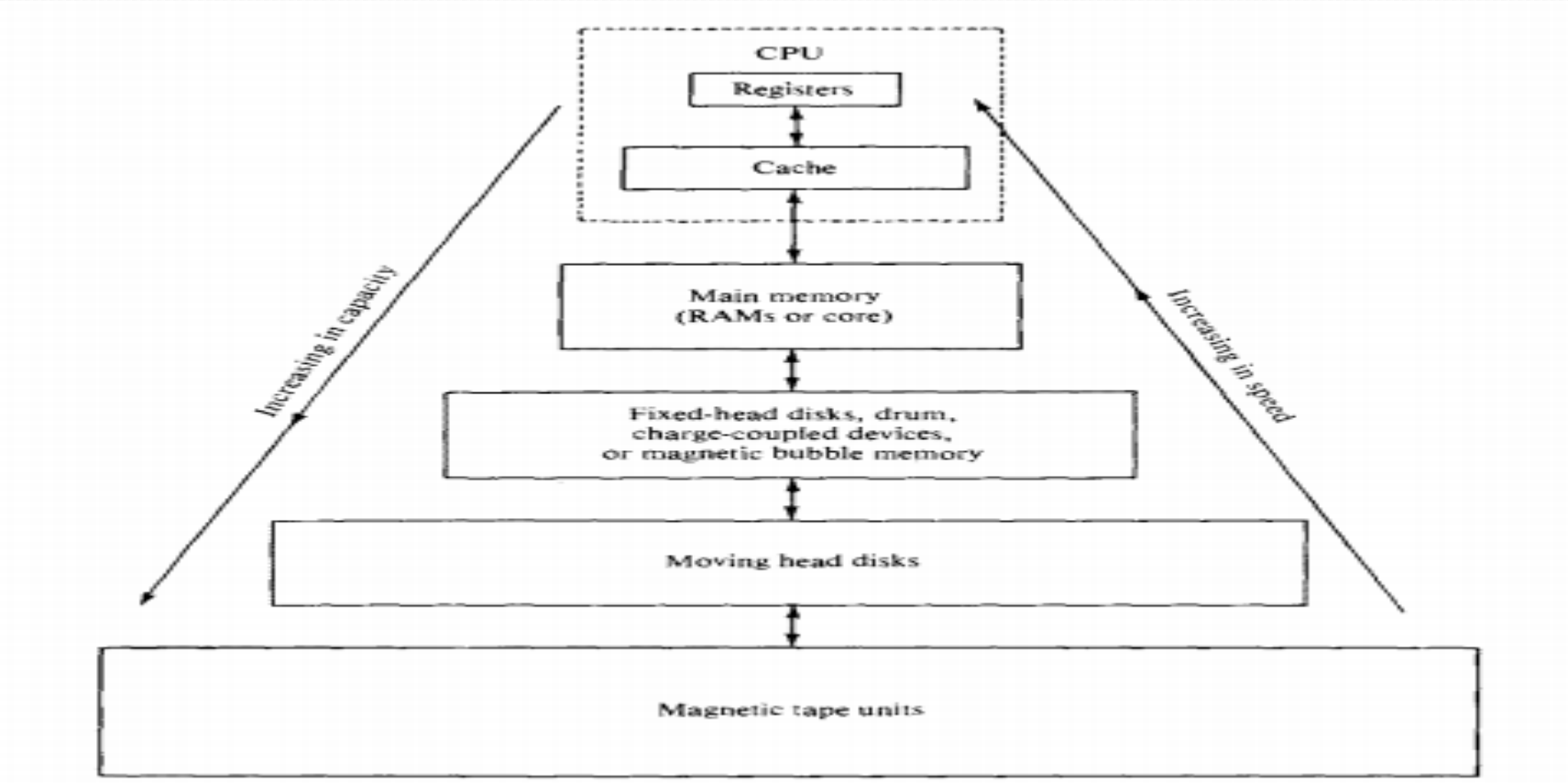
90 msec of I/O



USE OF HIERARCHICAL MEMORY SYSTEM

- Hierarchical memory can be used to close up the speed gap.
- Memory hierarchy a concept that is necessary for the CPU to be able to manipulate data.
- This is because it is only able to get instructions from cache memory.
- Cache memory is located on the processor chip, and is the fastest kind of memory.
- As a result of this, it is also the smallest, meaning that we cant hold all of our processes in it at once.

USE OF HIERARCHICAL MEMORY SYSTEM



BALANCING OF SUBSYSTEM BANDWIDTH

- The CPU is the fastest unit in a computer, within a processor cycle.
- A memory parameter is related to computing speed is bandwidth.
- Memory bandwidth affects cpu performance.
- Processing speed is limited by the rate at which it can fetch instruction and data from main memory.

Bandwidth

- Bandwidth: Number of bits that can be transmitted over a certain time -- typically per unit time.
- The memory bandwidth is measured by the number of memory words that can be accessed per unit time.
- Some people also refer it to the spectrum -- example 10 GHz.
- This typically translates to a maximum data rate.

Transmission Time

- A function of bandwidth
- If bandwidth is B, transmission time is 1/B.
- If bandwidth is 10 Mbps, the transmission time is $1/(10 \times 10^6) = 1$ Μρσ.

Propagation Delay

- Once a bit is put on a link, the time it takes to go across the link.
- Depends on the speed with which the electromagnetic signal (light) travels in the medium -- 2×10^8 m/s in fiber.
- Propagation delay = distance/speed of signal.

Queuing Delay

- At each intermediate node or router, a packet is queued.
- Thus, it has to wait prior to transmission.
- How long does it have to wait ? Dependent on the load on the network -- how many packets are traversing that router ?

Latency

- How long does a packet take to go from one host to another.
- Also called “Delay”.

Latency = Propagation Delay +
Queuing Delay +
Transmission Delay

Throughput

- Defines how efficiently channel is being used.

  Throughput = Transfer size/ Transfer time.

What is the transfer time ?

  RTT + (Transfer size/Bandwidth)

(ignoring queuing delays).

Example

- 1 MB file over a 1 Gbps network with RTT 100 milliseconds.
- Transfer time = 100 ms + (1 MB/1Gbps) = 100 ms + 8 ms = 108 ms.
- Effective throughput = 1 MB/108 ms = 74.1 Mbps.

BANDWIDTH BALANCING BETWEEN CPU AND MEMORY

- The speed gap between the CPU and the main memory can be closed up by using fast cache memory between them.
- A block of memory words is moved from the main memory into the cache so that immediate instruction/data can be available most of the time from the cache

BALANCING BETWEEN MEMORY AND I/O DEVICES

- I/O channels with different speeds can be used between the slow I/O devices and the main memory.
- To achieve totally bandwidth system, in which the entire memory bandwidth matches the bandwidth sum of the processor and I/O devices.

BALANCING BETWEEN MEMORY AND I/O DEVICES

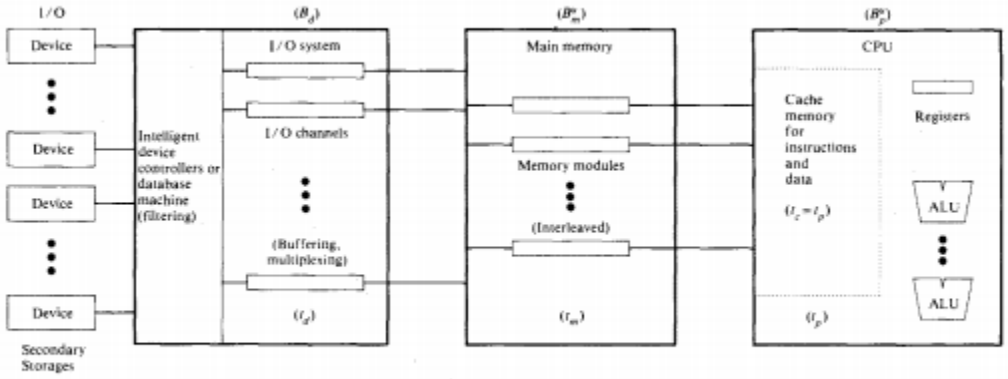


Figure 1.8 Bandwidth balancing mechanisms between CPU, memory, and I/O subsystem in a uniprocessor computer.

MULTIPROGRAMMING AND TIME SHARING

- Multiprogramming
- Processor has more than one program to execute
- The sequence in which the programs are executed depends on their relative priority and whether they are waiting for I/O.
- After an interrupt handler completes, control may not return.
- The program that was executing at the time of the interrupt.

TIME SHARING SYSTEM

- Time-Sharing Systems
- For many applications the interaction with users is
- essential.
- Processor time is shared among multiple users
- Using multiprogramming to handle multiple interactive
- jobs
- Multiple users simultaneously access the system through
- terminals, with the OS interleaving the execution of each
- user program in a short burst or quantum of computation

REFERENCE

- Advanced Computer Architecture: Parallelism, Scalability, Programmability by kai hwang.

THANK YOU