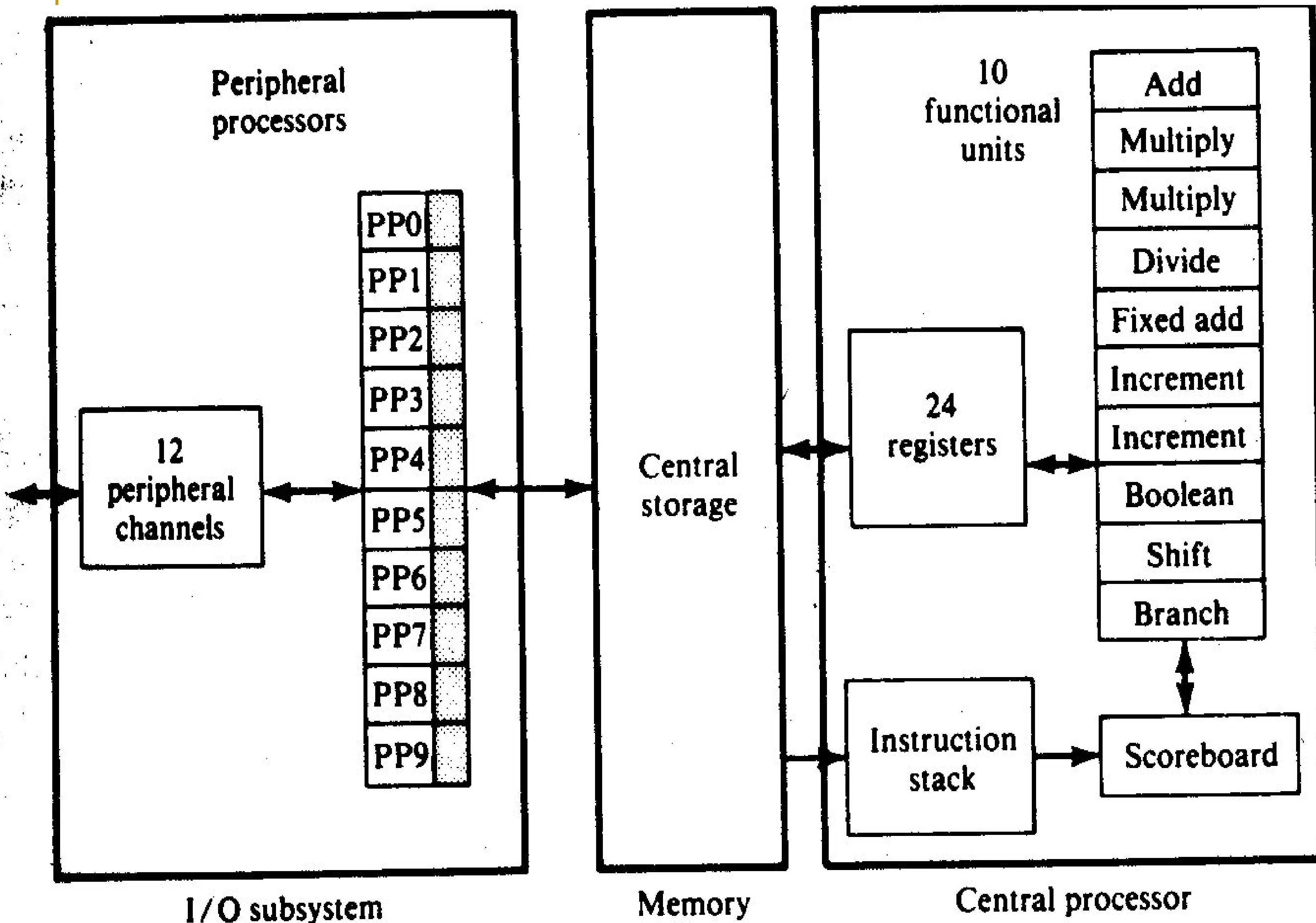
Parallel Processing Mechanisms

1. Multiplicity of functional units
2. Parallelism and pipelining within the CPU
3. Overlapped CPU and I/O operations
4. Use of a hierarchical memory system
5. Balancing of subsystem bandwidths
6. Multiprogramming and time sharing

1. Multiplicity of functional units

- Early computers
  - one ALU that perform one operation at a time.
  - Slow process
- Multiple and specialized functional units.
  - operate in parallel.
- IBM 360/91 →
  - two parallel execution units (fixed and floating point arithmetic)
- CDC-6600 →
  - 10 functional units

System Architecture of CDC-6600



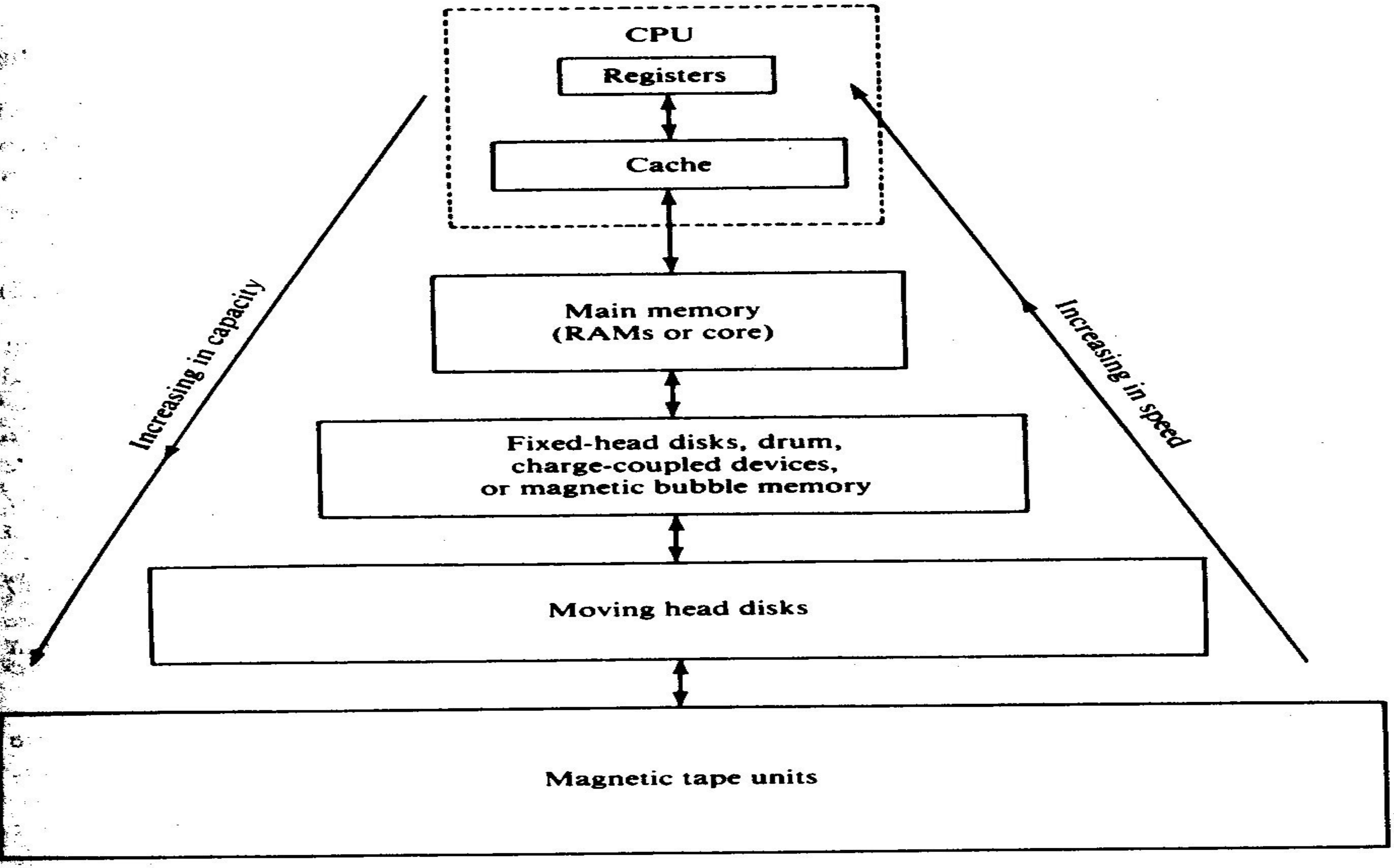
2. Parallelism and pipelining within CPU

- Parallel Adders
  - bit serial adders.
  - carry-lookahead and carry save adders.
- Multiplier recoding and convergence division.
- Phases of instruction execution are pipelined
  - Instruction fetch, decode, operand fetch, arithmetic logic execution, store result.
- Instruction Prefetch and data buffering.

3. Overlapped CPU and I/O operations

- I/O operations performed simultaneously with CPU computations
  - separate I/O controllers, channels or I/O processors.
- DMA channels – cycle stealing.

4. Use of a hierarchical memory system



5. Balancing of subsystem bandwidths

- . $t_d > t_m > t_p$
- Bandwidth of a system
  - no: of operations performed per unit time.
- Memory bandwidth (B_m)
  - no: of words that can be accessed per unit time.

$$B_m = \frac{W}{t_m}$$

- Processor bandwidth (B_p)
  - max: CPU computation rate.
  - Ex: Cray-1 $\rightarrow$ 160 MFLOPS

- I/O bandwidth (B_d)
  - Average data transfer rate.
  - Ex: Modern drives has Bd = 1 megabyte/sec
- Utilization bandwidth of memory (B_m^u)

$$B_m^u = \frac{B_m}{\sqrt{M}} \quad \text{and} \quad B_m^u \leq B_m$$

- Utilization bandwidth of CPU

$$B_p^u = \frac{R_w}{T_p}$$

- Utilization bandwidth of I/O (B_d^u)
  - lower than the actual bandwidth.
- Relationship b/w BWs

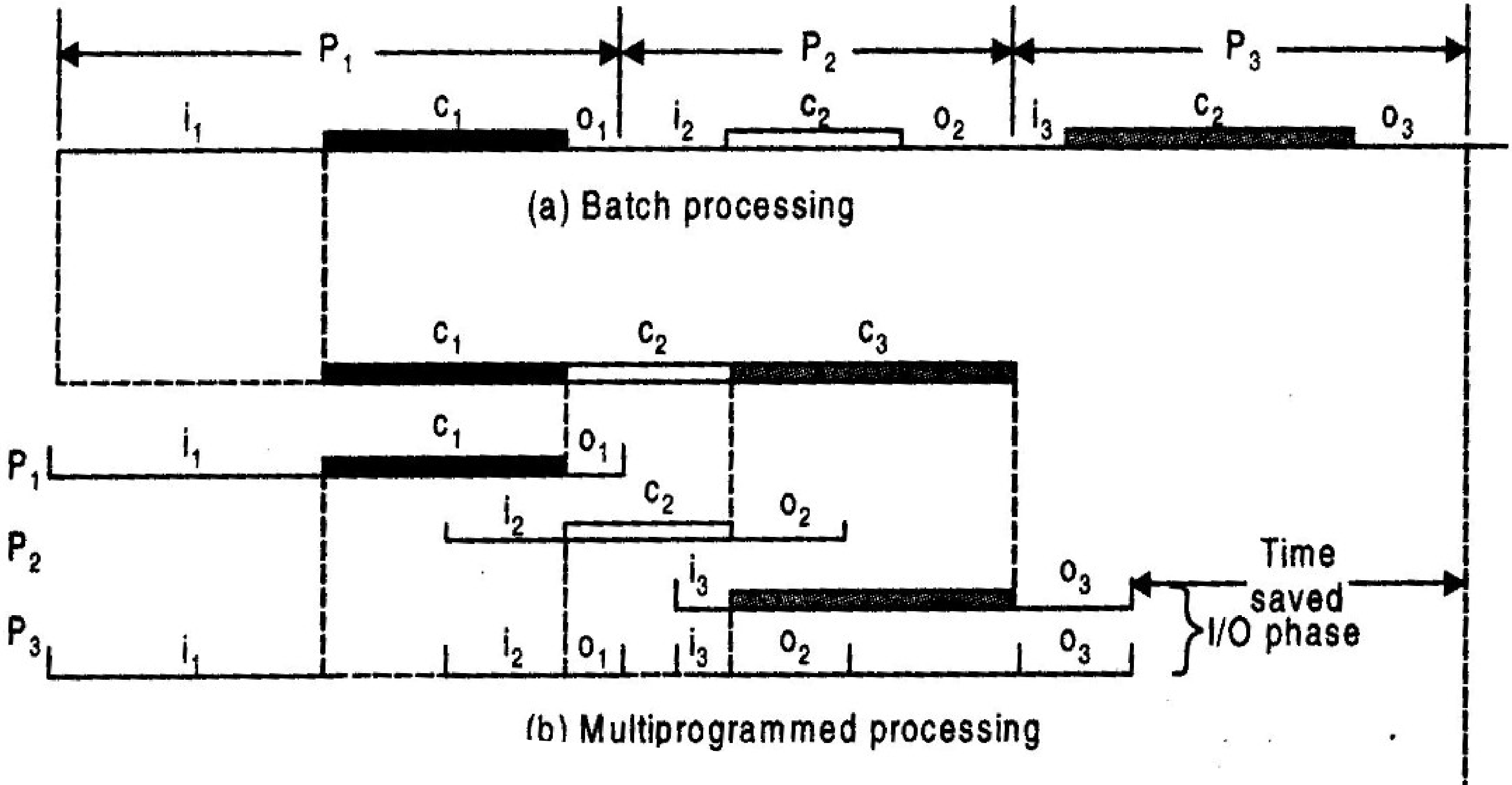
$$B_m \geq B_m^u \geq B_p \geq B_p^u \geq B_d$$

- Bandwidth balancing between CPU and memory.
  - Cache memory
  - $t_c = t_p$
- Bandwidth balancing between memory and I/O
  - Buffer
  - Multiplexing
- Totally balanced system

$$B_p^u + B_d = B_m^u$$ where $$B_p^u = B_p$$ and $$B_m^u = B_m$$

6. Multiprogramming and time sharing

- Batch processing
  - Sequential execution
- Multiprogramming
  - Interleaving of CPU and I/O operations among several programs
- Time sharing
  - Assign fixed or variable time slices to multiple programs


P1
P2
P3
i1
c1
o1
i2
c2
o2
i3
c2
o3
(a) Batch processing
c1
c2
c3
P1
i1
c1
o1
i2
c2
o2
P2
i3
o3
Time saved
I/O phase
P3
i1
i2
o1
i3
o2
o3
(b) Multiprogrammed processing

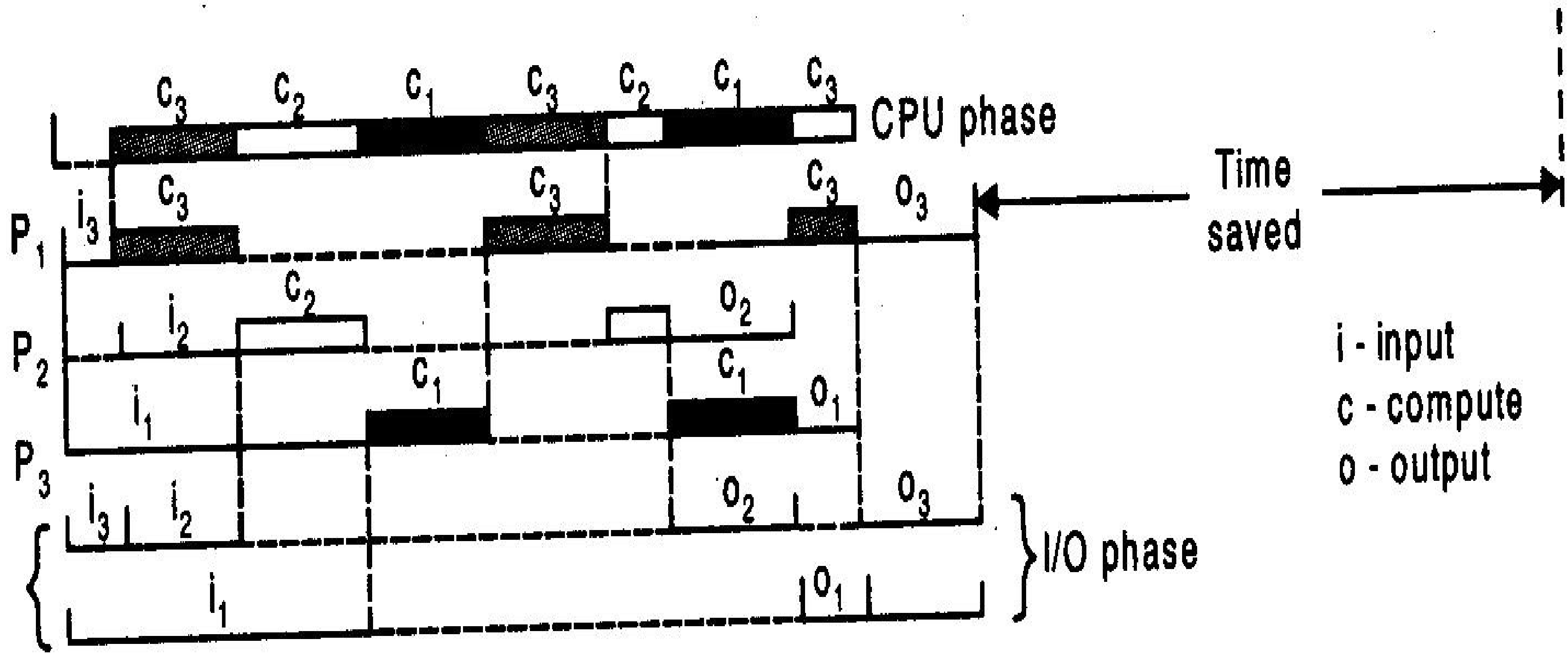


Fig. 1.8. Time Shared Processing