Parallel Computer Structures

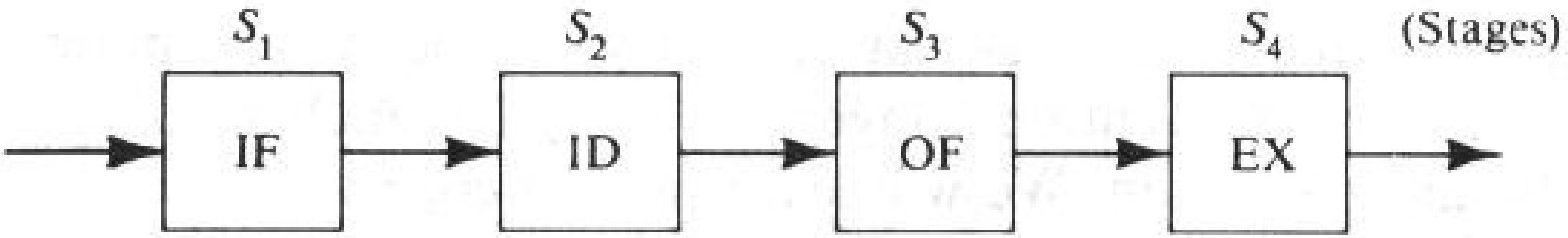
國立清華大學
National Tsing Hua University

Parallel Computer Structures

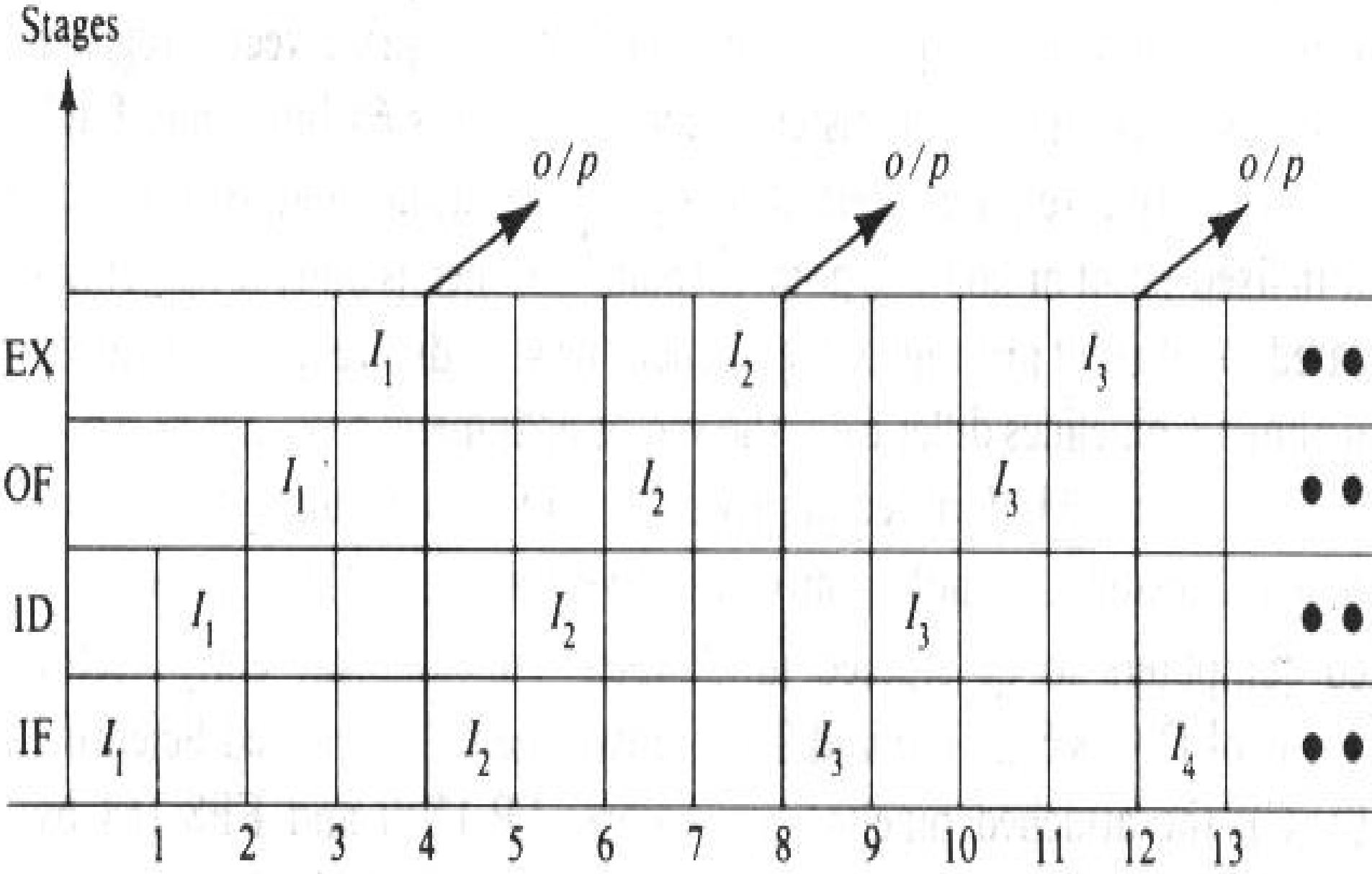
- Pipelined Computers :
  - overlapped computations
  - *temporal parallelism*.
- Array Processors :
  - multiple synchronized arithmetic logic units
  - *spatial parallelism*.
- Multiprocessor Systems :
  - *asynchronous parallelism*
  - set of interactive processors with shared resources.

Pipeline Computers

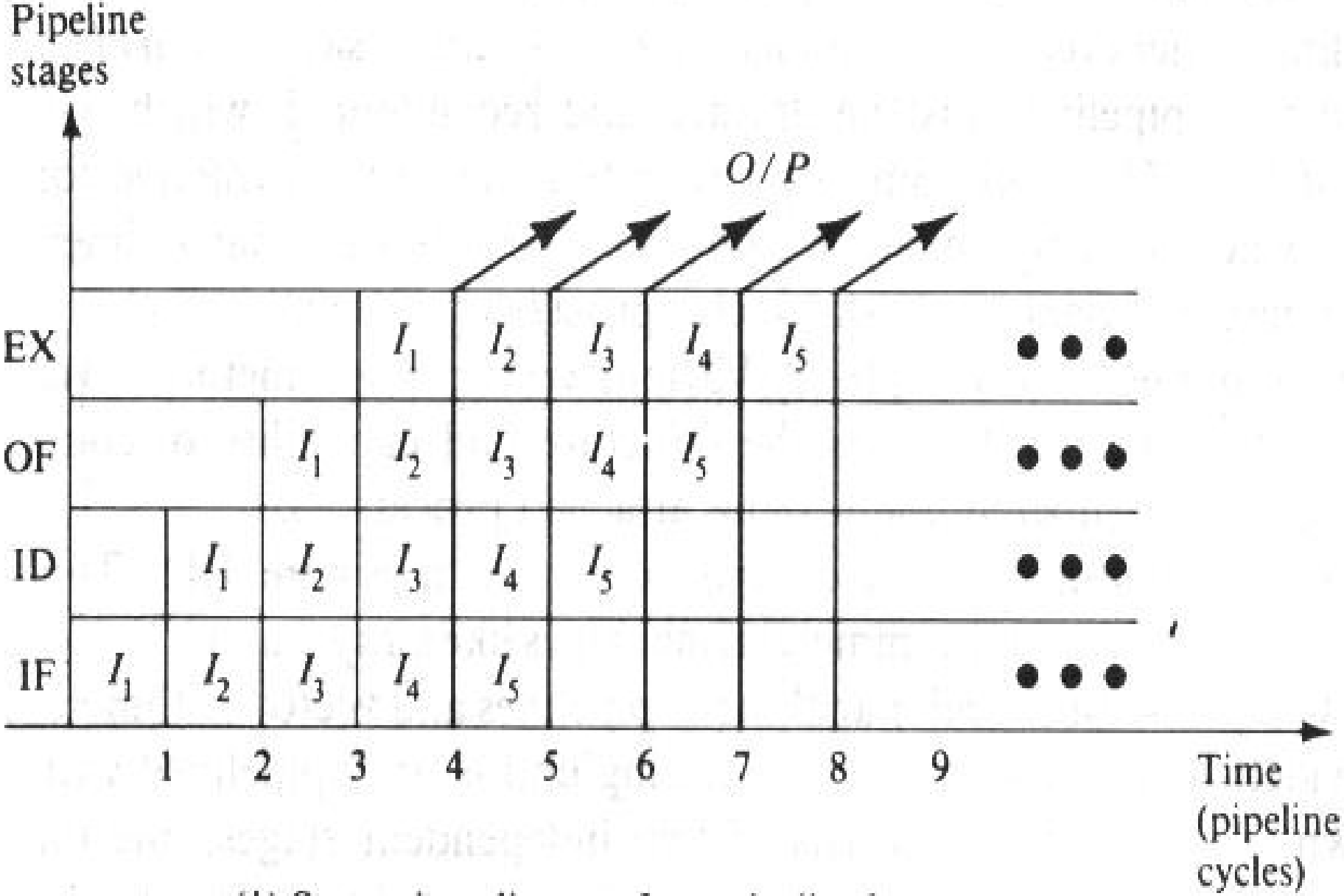
- Normally, four major steps to execute an instruction:
  - Instruction Fetch (IF)
  - Instruction Decoding (ID)
  - Operand Fetch (OF)
  - Execution (EX)



Space-time diagram for a nonpipelined processor



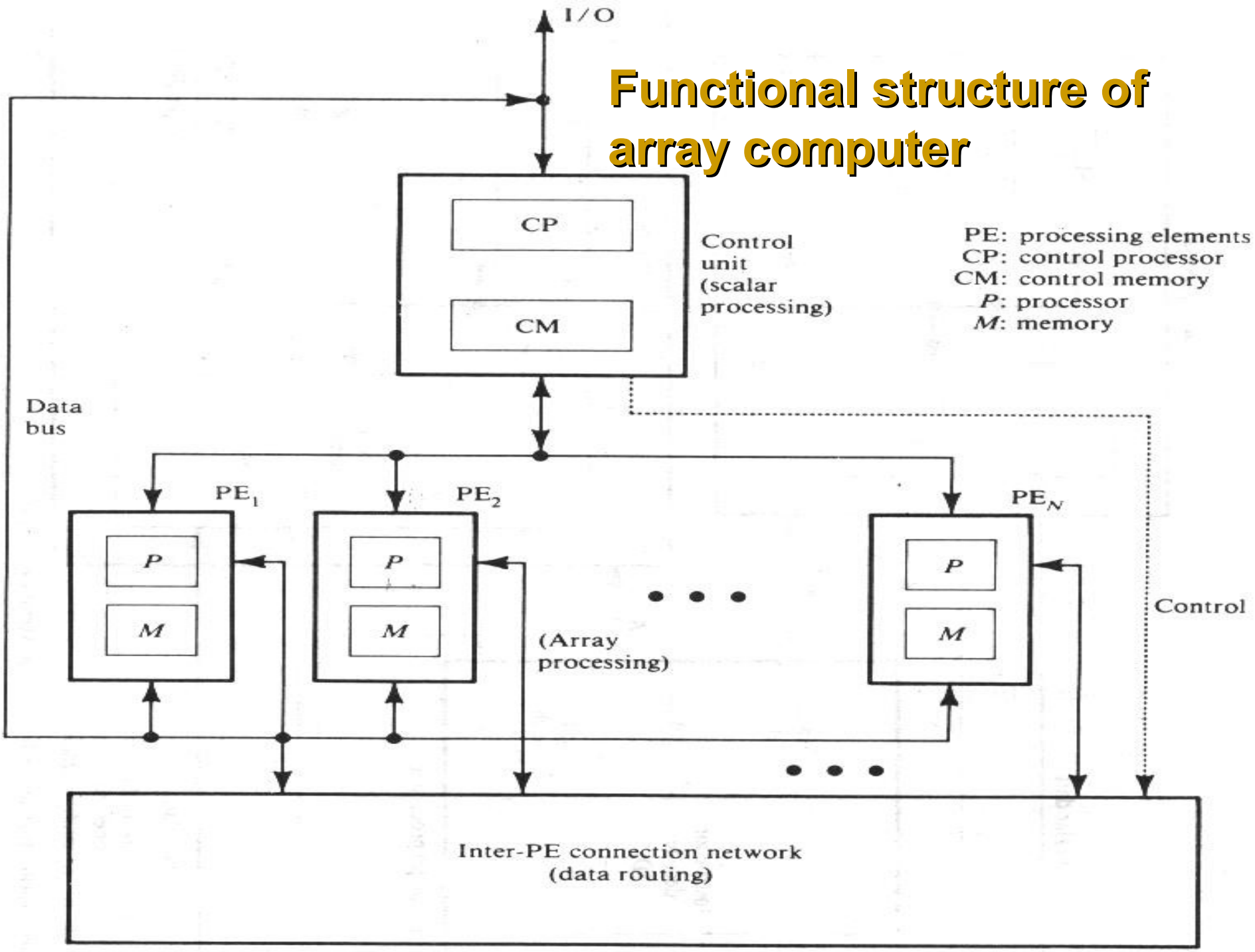
Space-time diagram for a pipelined processor



Array Computers

- A synchronous parallel computer
- Multiple arithmetic logic units
  - *processing elements* (PE)
  - operate in parallel.
- PEs are synchronized
  - perform the same function at the same time.
- Appropriate data routing mechanism must be established among the PEs.

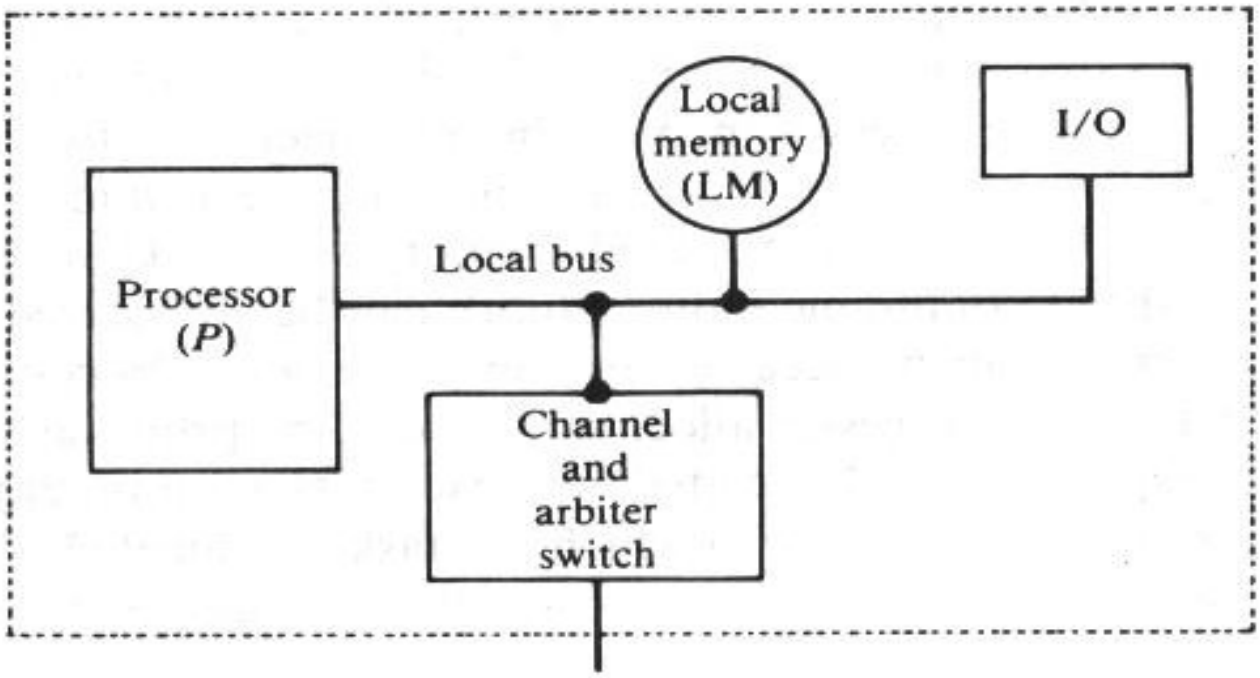
Functional structure of array computer



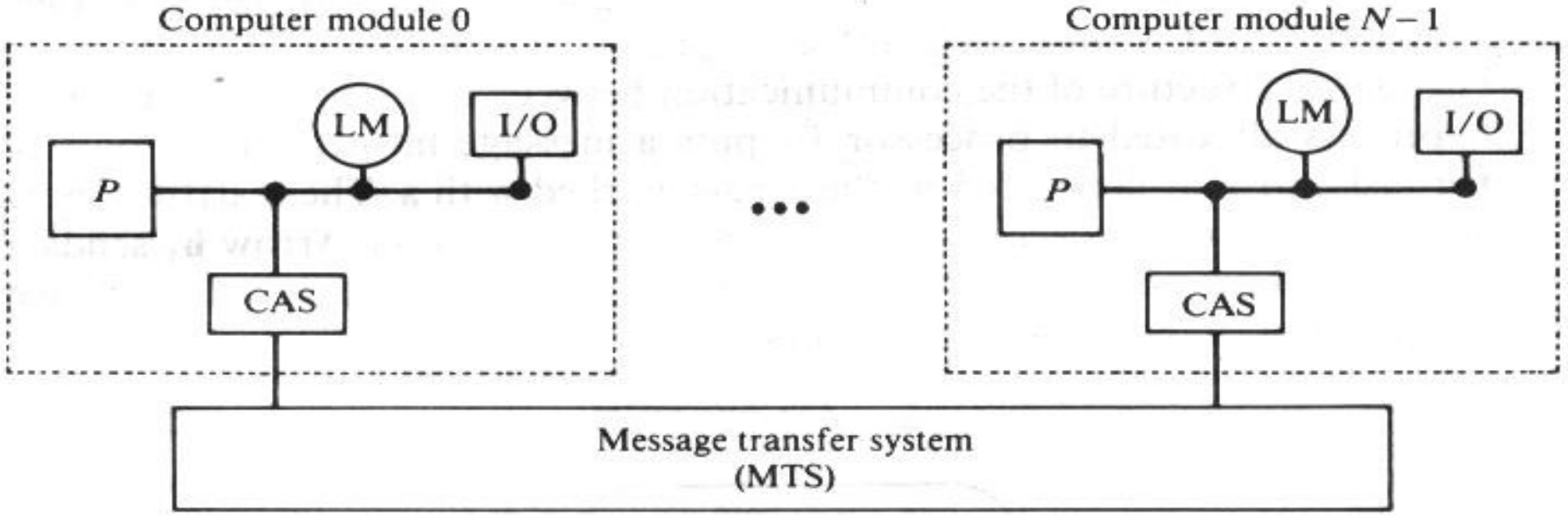
Multiprocessor Systems

- A single computer that includes multiple processors (computer modules).
- All processors share memory modules, I/O channels and peripheral devices.
- Controlled by one operating system
  - provides interaction between processors and their programs.
- Local memory and private devices.

- Processor communication
  - sending messages
  - sharing a common memory or through an interrupt network.
- Interconnections
    - Time-shared common bus
    - Crossbar switch network
    - Multiport memories

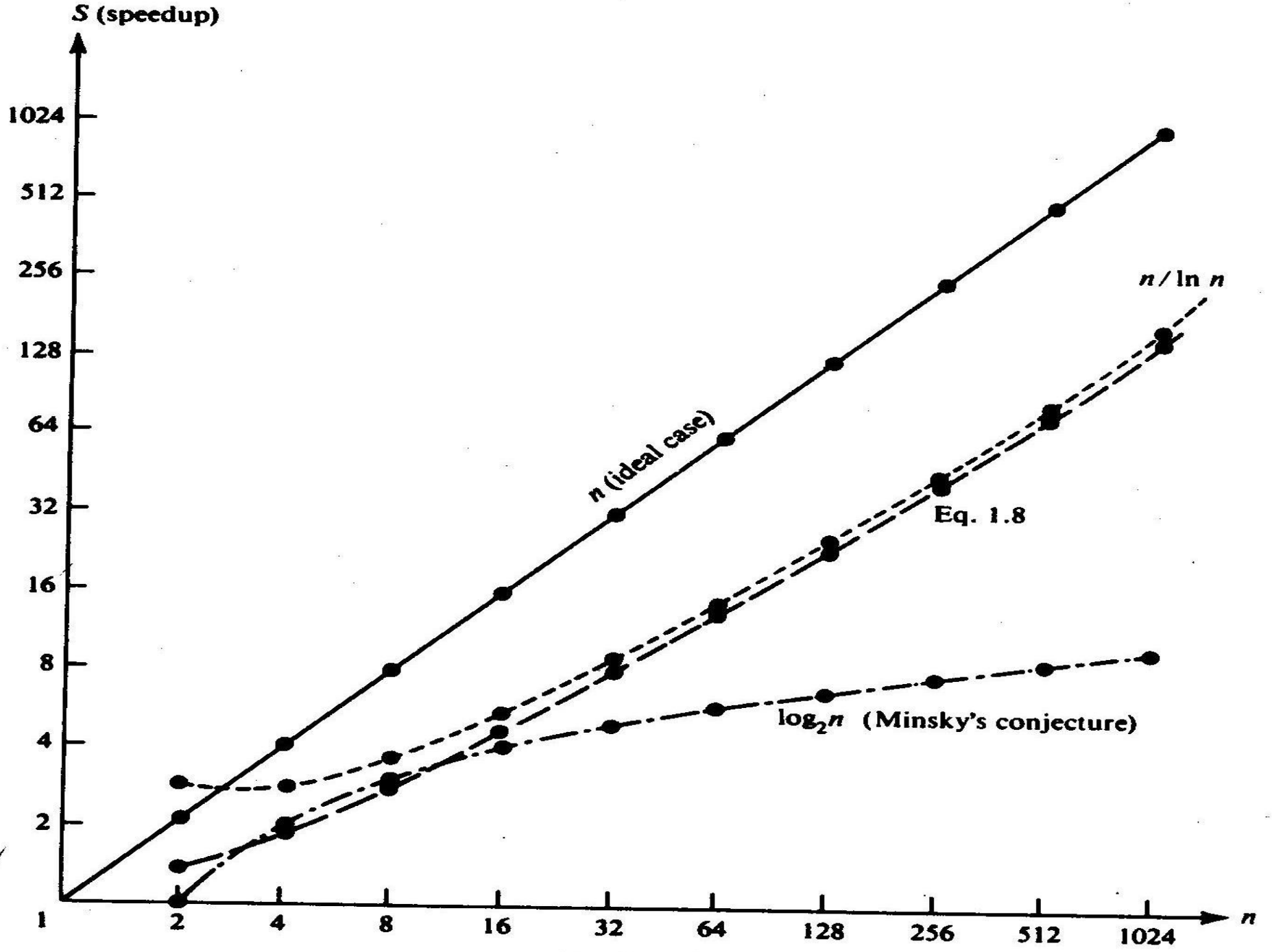


(*a*) A computer module



Performance of Parallel Computers

- Speedup of a parallel computer with n identical processors
  - n times faster than single processor.
- Speedup is much less(in practice)
  - some processors are idle.
- Actual speedup
  - lower-bound *$\log_2 n$ (Minsky's conjecture)*
  - upper bound *$n/\ln n$*
- *So Commercial processor system consists of only 2 or 4 processors.*


S (speedup)
1024
512
256
128
64
32
16
8
4
2
1
2
4
8
16
32
64
128
256
512
1024
n
n (ideal case)
n / ln n
Eq. 1.8
log2n (Minsky's conjecture)

- $T_1=1$
- f_i – probability of assigning the same problem to i processors.
- Average load = $d_i = 1/i$ per processor
- $f_i = 1/n$.
- Average time required to solve the problem on an n- processor system is

$$T_n = \sum_{i=1}^{n} f_i . d_i = \frac{\sum_{i=1}^{n} \frac{1}{i}}{n}$$

- Average speedup $S = \frac{T_1}{T_n} = \frac{n}{\sum_{i=1}^{n} \frac{1}{i}} \leq \frac{n}{\ln n}$

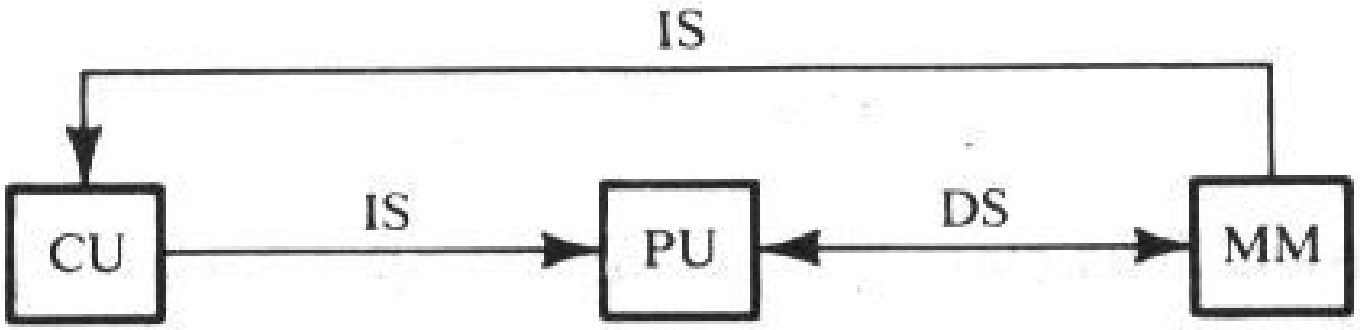
Architectural Classification Schemes

1. Flynn's Classification

- Multiplicity of data and instruction streams.
- *Instruction stream*
  - a sequence of instructions as executed by the machine.
- *Data stream*
  - a sequence of data including input, partial, or temporary results, called for by the instruction stream.
- Flynn's four machine organizations :
  - SISD, SIMD, MISD, MIMD.

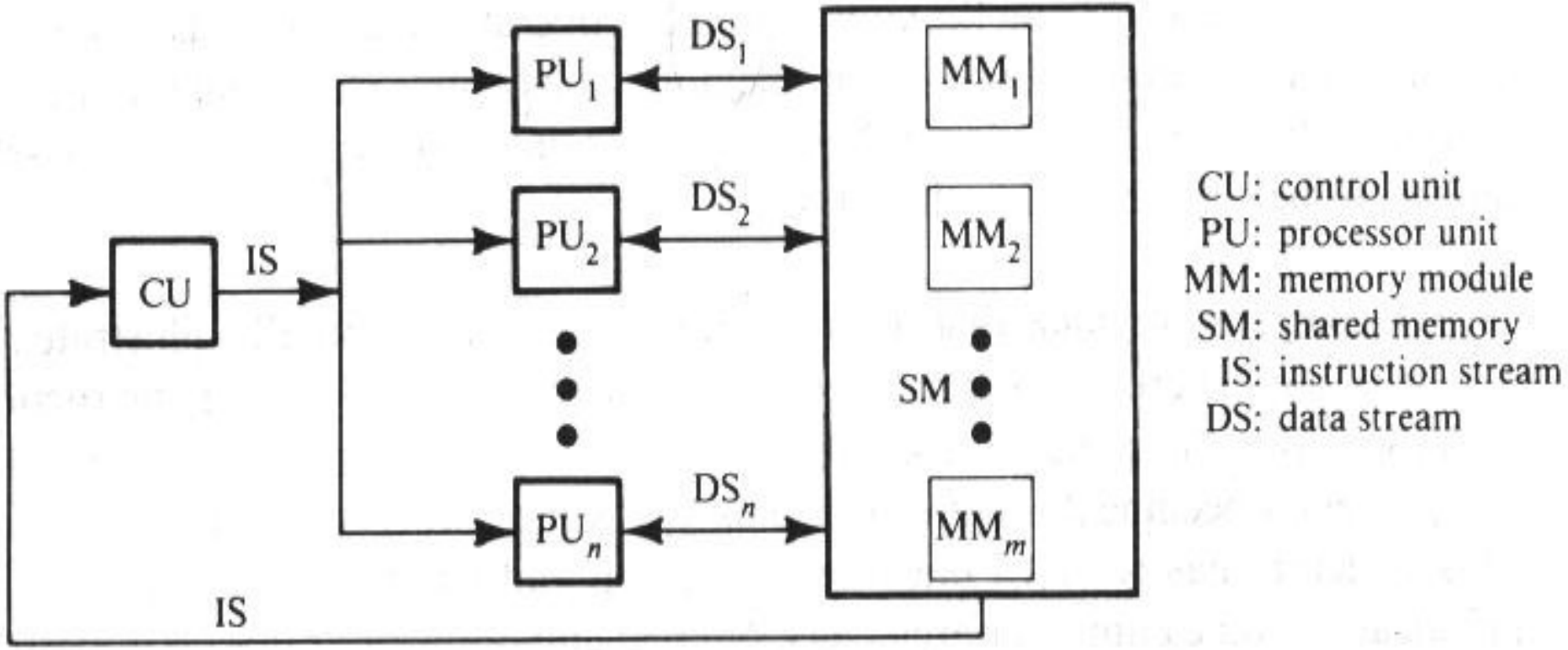
SISD

- Single Instruction stream-Single Data stream
- Instructions are executed sequentially but may be overlapped in their execution stages (pipelining).



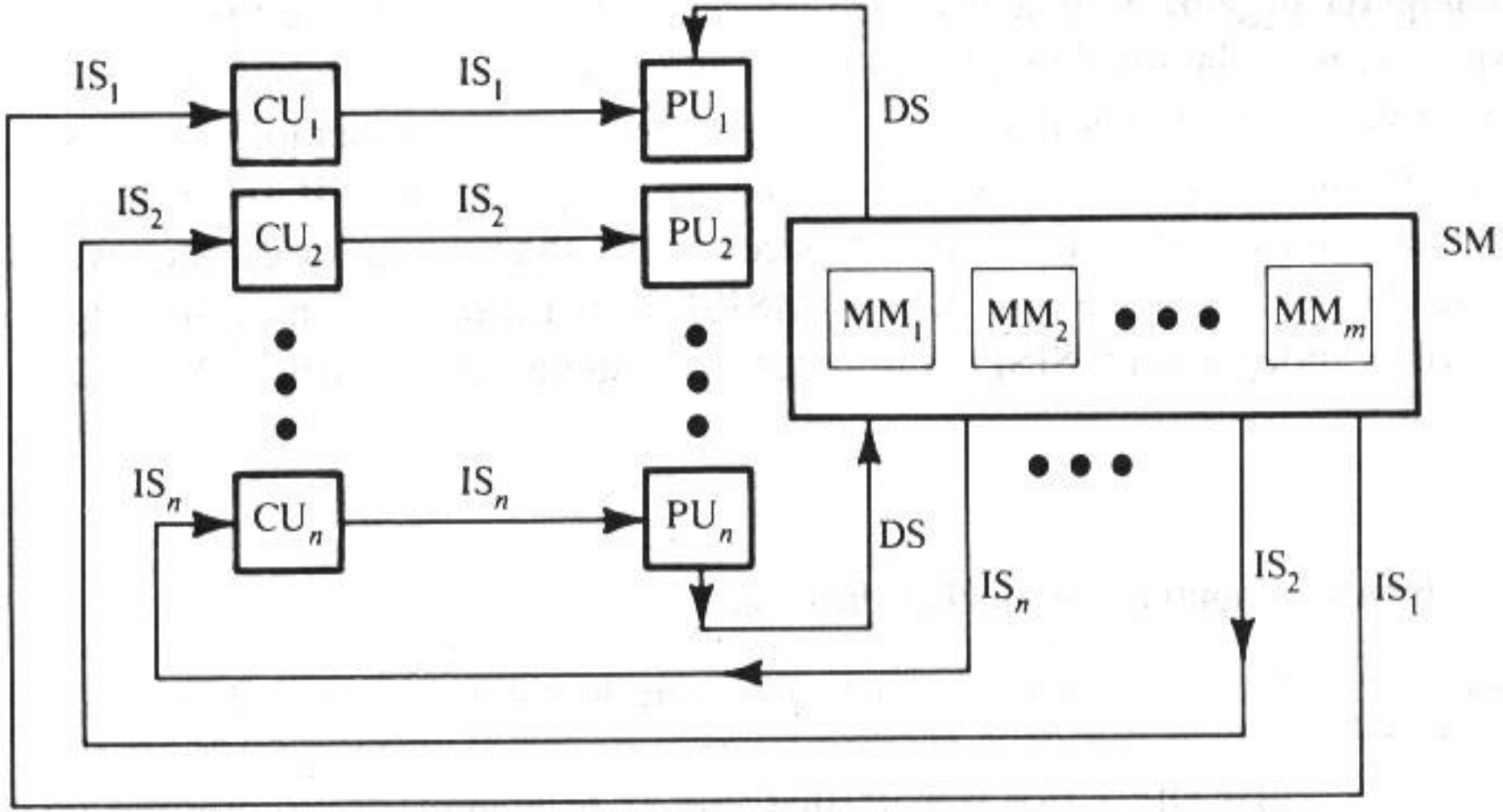
SIMD

- Single Instruction stream-Multiple Data stream
- There are multiple PEs supervised by the same control unit.



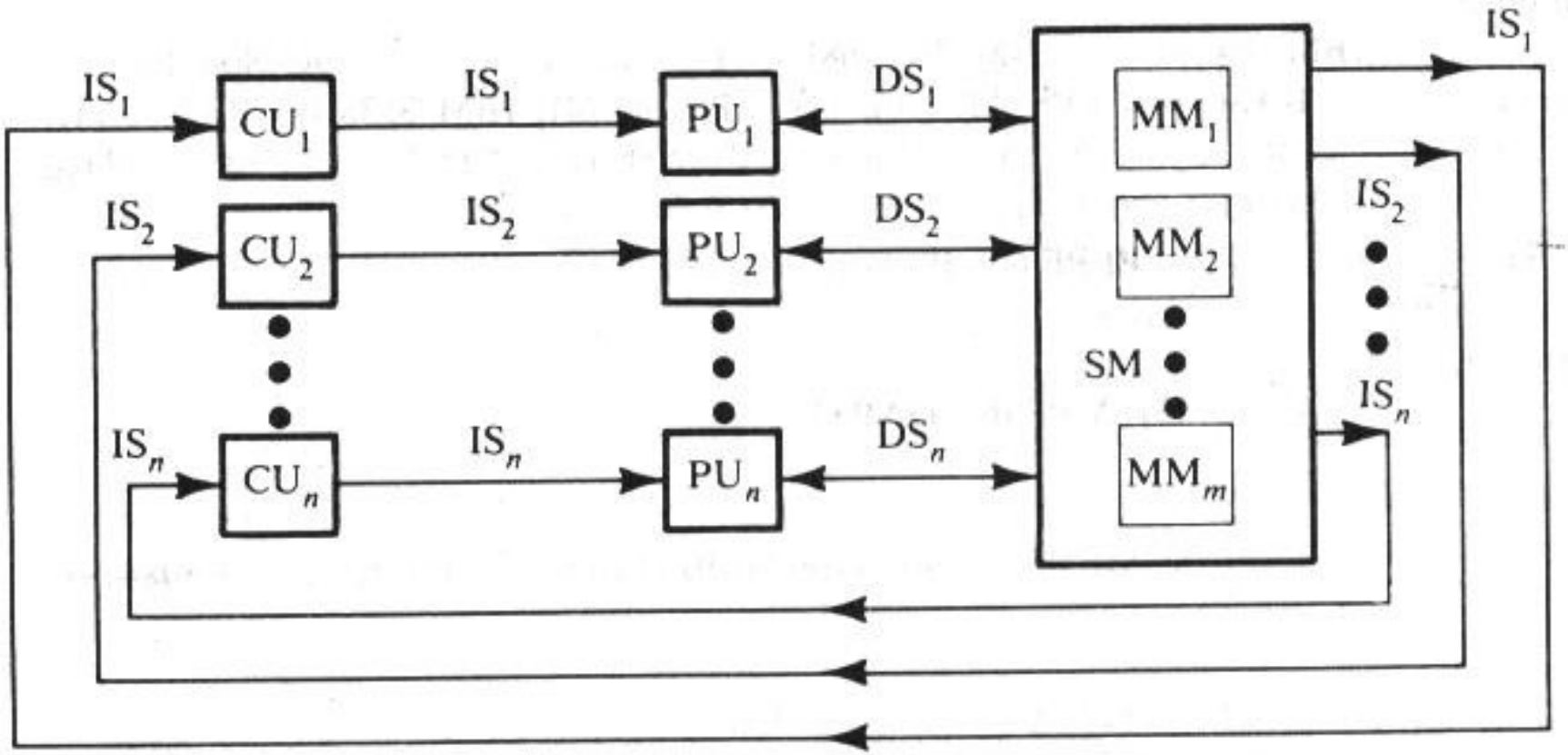
MISD

- Multiple Instruction stream-Single Data stream



MIMD

- Multiple Instruction stream-Multiple Data stream
- Multiprocessor systems and Multi- computer systems.



2. Feng's classification

- Use the degree of parallelism
- Maximum degree of parallelism
  - max: no: of bits that can be processed within a unit time (P).
- Average parallelism degree

$$P_a = \frac{\sum_{i=1}^{T} P_i}{T}$$

- Utilization Rate

$$\mu = \frac{P_a}{P} = \frac{\sum_{i=1}^{T} P_i}{T.P}$$

- Max: parallelism degree, *P(C)* of a computer system C is *P(C)= n.m*
- Four type of processing methods

1. Word-serial and bit-serial (WSBS)
   - n=m=1
   - Bit serial processing as one bit is processed at a time.
2. Word-parallel and bit-serial (WPBS)
   - n=1, m>1
   - Bis(Bit-slice) processing as an m-bit-slice is processed at a time.

3. **Word-serial and bit-parallel (WSBP)**
   - n>1, m=1
   - Word-slice processing as it processed one word of n-bits at a time.
4. **Word-parallel and bit-parallel (WPBP)**
   - n>1, m>1
   - Parallel processing as an array of *(n x m)* bits is processed at a time.

3. Handler's Classification

- Degree of parallelism and pipelining built in to the hardware structures of a computer system.
- Parallel pipeline processing is considered at three subsystem levels:
  - PCU, ALU, BLC
- $T(X) = < K \times K' , D \times D' , W \times W' >$
- *Ex:*

  TI - ASC <1, 4, 64 x 8>