PARALLEL PROCESSING

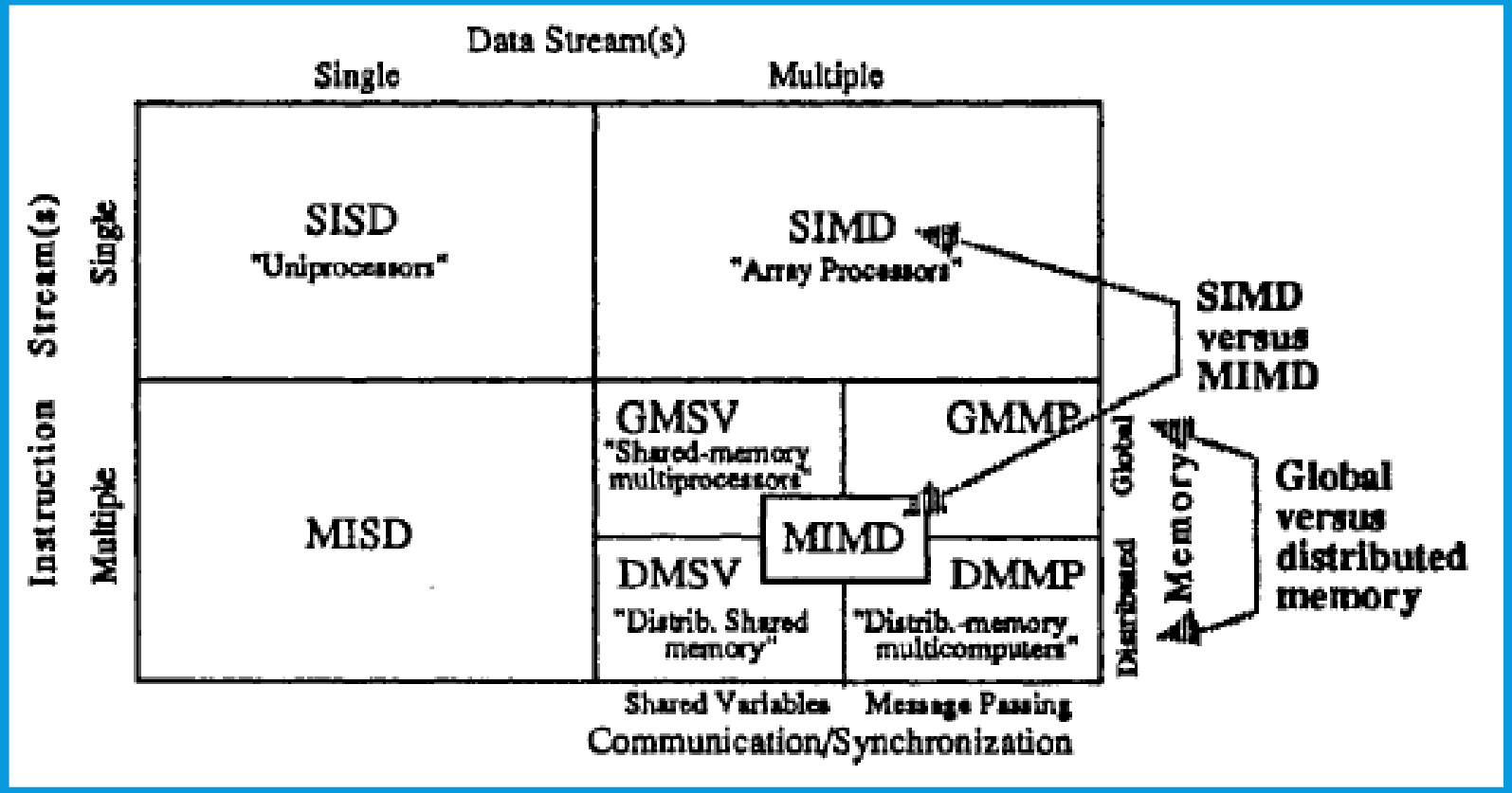
Mohammed Alabdulkareem

kareem@ksu.edu.sa

Office 2247

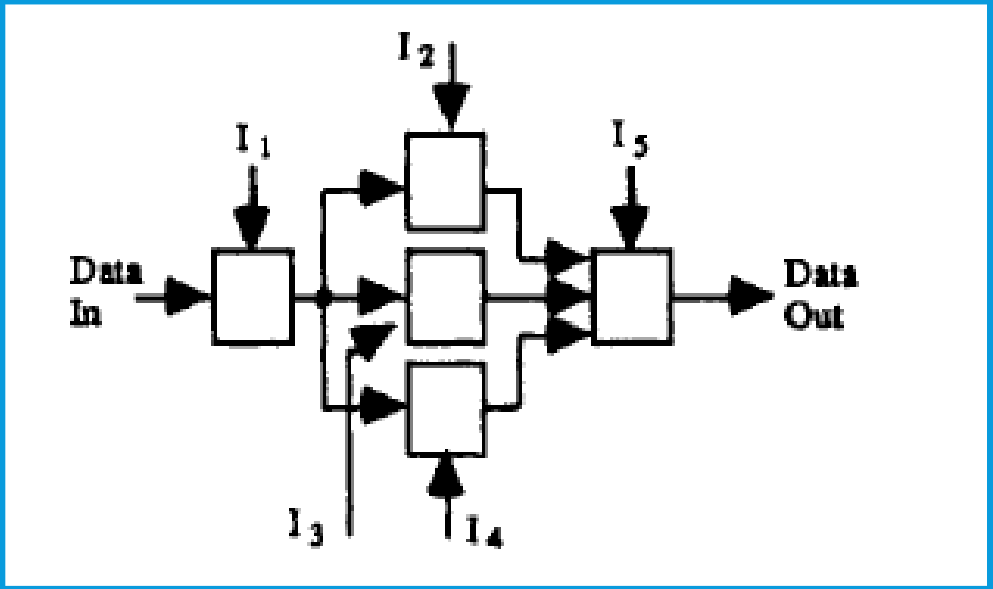
7

MODELS OF PARALLEL PROCESSING



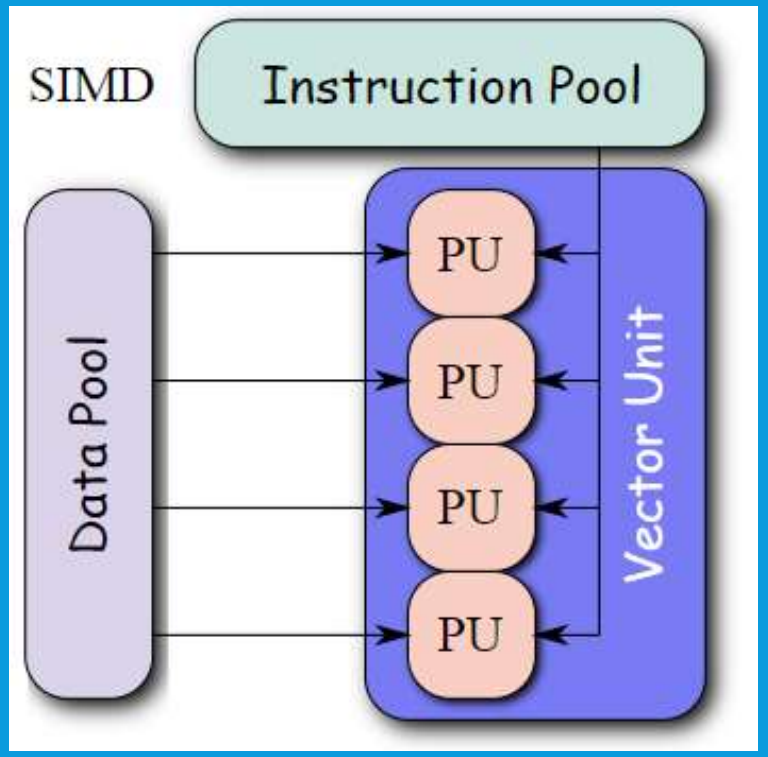
MODELS OF PARALLEL PROCESSING

- SISD class is the standard uniprocessor systems with all processing enhancements.
- MISD class has not found widespread application. It is impossible to design a general-purpose architecture of this type.



MODELS OF PARALLEL PROCESSING

- **SIMD ARCHITECTURES**



MODELS OF PARALLEL PROCESSING

- SIMD ARCHITECTURES

  - The early parallel machines where SIMD machines.
  - A central unit fetches and interprets the instructions and then broadcasts appropriate control signals to a number of processors.
  - The applications at the early stages where suitable such as air traffic control and linear-algebra computations.

MODELS OF PARALLEL PROCESSING

- **SIMD Design Choices** (*Synchronous or Asynchronous*)

  - Each processor can execute or ignore the instruction being broadcast based on its local state or data dependent conditions.
  - For example, an "if-then-else" statement is executed by first enabling the processors for which the condition is satisfied and then flipping the "enable" bit before getting into the "else" part. On average 50% of the processors will be idle.
  - This will lead to some inefficiency.

MODELS OF PARALLEL PROCESSING

- **SIMD Design Choices (*Synchronous or Asynchronous*)**

  - SPMD (Single Program Multiple Data) is a solution where each processor execute its own copy of the program.
  - Now it needs occasional synchronization and higher complexity of each processor, which must now have a program memory and instruction fetch/decode logic.

MODELS OF PARALLEL PROCESSING

SIMD Design Choices (*Custom-chip or commodity-chip)*

- ***Commodity-chip***
  - SIMD machine can be designed based on commodity (off-the-shelf) components **or** with custom chips.
  - Commodity components tend to be inexpensive because of mass production.
  - However, such general-purpose components will likely contain elements that may not be needed for a particular design.
  - These extra components may complicate the design, manufacture, and testing of the SIMD machine and may introduce speed penalties as well.

MODELS OF PARALLEL PROCESSING

SIMD Design Choices (*Custom-chip or commodity-chip)*

- ***Custom-chip***
  - Custom components (including ASICs = application-specific ICs, multichip modules, or WSI = wafer-scale integrated circuits) generally offer better performance
  - It leads to much higher cost in view of their development costs being borne by a relatively small number of parallel machine users.

MODELS OF PARALLEL PROCESSING

- **SIMD Story**
  - **The first commercially successful machine was CRAY-1.**
  - **Read more about SIMD vector machines:**

https://www.hpcwire.com/2016/09/26/vectors-old-became-new-supercomputing/

  - **Do we have Vector Machines theses days? Explain?**

MODELS OF PARALLEL PROCESSING

- **SIMD Design Choices (***Custom-chip vs Commodity-chip)*

	Cost	Performance
Commodity-chip	low	low
Custom-chip	High	High

MODELS OF PARALLEL PROCESSING

MIMD

MIMD paradigm has become more popular recently for the following reasons:

- High flexibility of the MIMD architectures.
- Taking advantage of commodity microprocessors.
- MIMD machines are most effective for medium-to coarse-grain parallel applications, where the computation is divided into relatively large sub-computations or tasks whose executions are assigned to the various processors.

MODELS OF PARALLEL PROCESSING

MIMD advantages

- Flexibility in exploiting various forms of parallelism.
- Easer of partitioning into smaller independent parallel processors in a multiuser environment.
- More scalable.

MIMD disadvantages

- Inter-processor communication overhead.
- more difficult programming.

MODELS OF PARALLEL PROCESSING

MIMD Design Choices (MPP)

- *Massively Parallel Processor*
- Build a parallel machine out of a large number of small processors.
- Usually 1000 or more processors.
- We gain higher speed-up for the parallelizable part of the program.
- The alternative is to use small number of powerful processors, this works better on the inherently sequential part of the program.

MODELS OF PARALLEL PROCESSING

MIMD Design Choices (*Tightly versus loosely coupled)*

- Instead of using specialized multiprocessor/ multicomputer (tightly-coupled) we may use a cluster of computers or workstations connected via commodity network (loosely-coupled).
- *Network of workstations* (NOW) or cluster computing is being more popular due to the relatively low cost of implementation.
- Depending on the problem type we may choose the suitable architecture (Tightly or loosely coupled).
- An intermediate approach is to link tightly coupled clusters of processors via commodity networks.

MODELS OF PARALLEL PROCESSING

MIMD Design Choices (MP or SM)

- ***Explicit message passing versus virtual shared memory.***
  - MP: Users must explicitly specify all messages that must be sent between processors.
  - VSM: Users program in an abstract higher-level model, with the required messages automatically generated by the system software.

- **Which is better?**
  - Using explicit message passing is more efficient but need more effort.
  - Using virtual shred memory is less efficient but easier for the programmer.

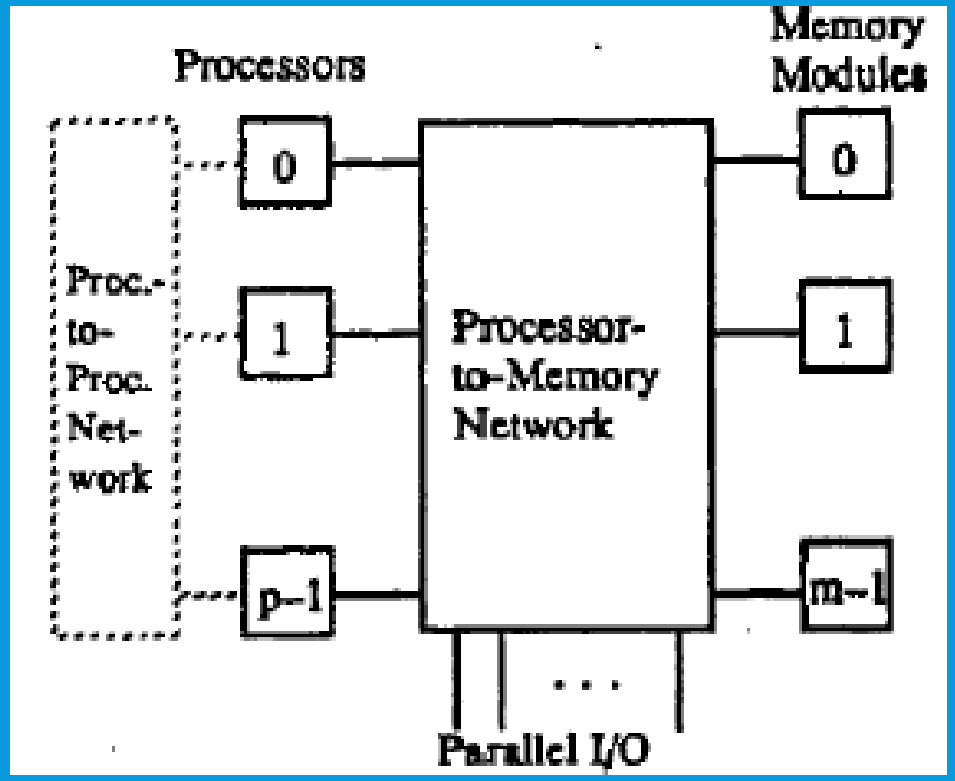
MODELS OF PARALLEL PROCESSING

GLOBAL VERSUS DISTRIBUTED MEMORY

- Global memory may be visualized as being in a central location where all processors can access it.
- Processors can access memory through a special processor-to-memory network.
- The interconnection network must have very low latency (challenging for large number of processors).
- Very high network bandwidth is a must.

MODELS OF PARALLEL PROCESSING

- GLOBAL MEMORY



MODELS OF PARALLEL PROCESSING

- **GLOBAL MEMORY**
- A global-memory multiprocessor is characterized by the type and number of processors ***p***, the capacity and number of memory modules ***m***, and the network architecture.
- ***p*** and ***m*** are independent parameters. However, achieving high performance typically requires that they be comparable in magnitude.

MODELS OF PARALLEL PROCESSING

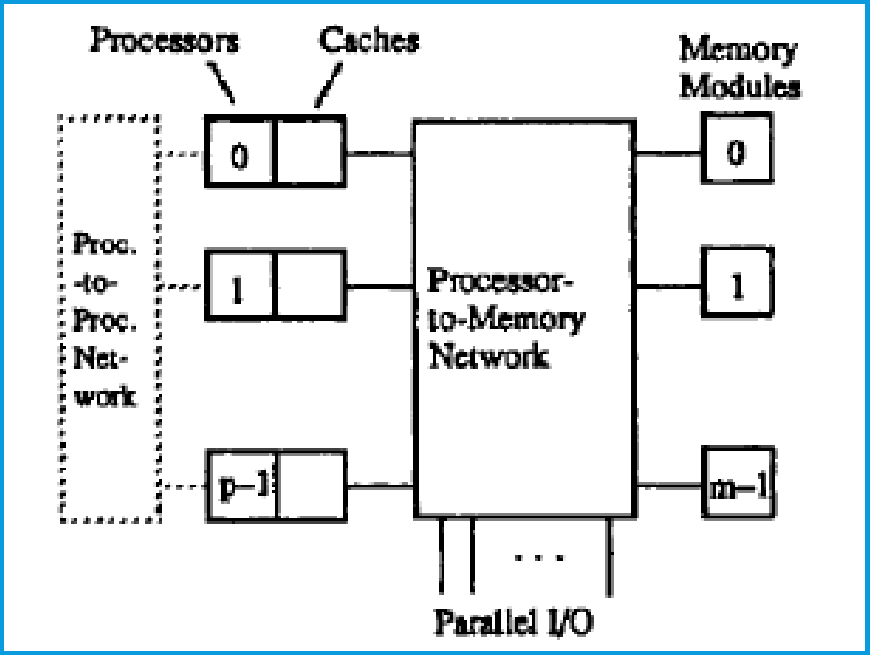
GLOBAL MEMORY

- Interconnection network between processors and memory modules or between processors can be:
  - Crossbar switch
  - Single or multiple buses
- Multistage interconnection network (lower cost than crossbar switch and higher band-width than bus)

MODELS OF PARALLEL PROCESSING

GLOBAL MEMORY

- Use cache to enhance performance but may cause hazard.



MODELS OF PARALLEL PROCESSING

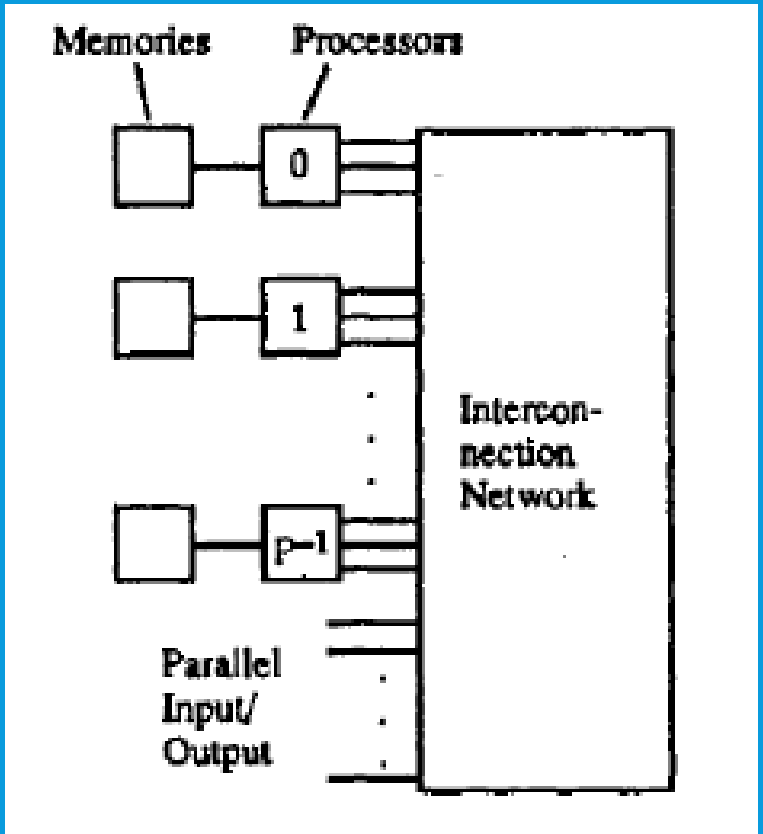
DISTRIBUTED MEMORY

- A collection of ***p*** processors, each with its own private memory, communicates through an interconnection network.
- The latency of the interconnection network may be less critical, as each processor is likely to access its own local memory most of the time.
- Depending on the type of parallel applications the interconnection latency may or may not be critical.

MODELS OF PARALLEL PROCESSING

جامعة الملك سعود
King Saud University

DISTRIBUTED MEMORY



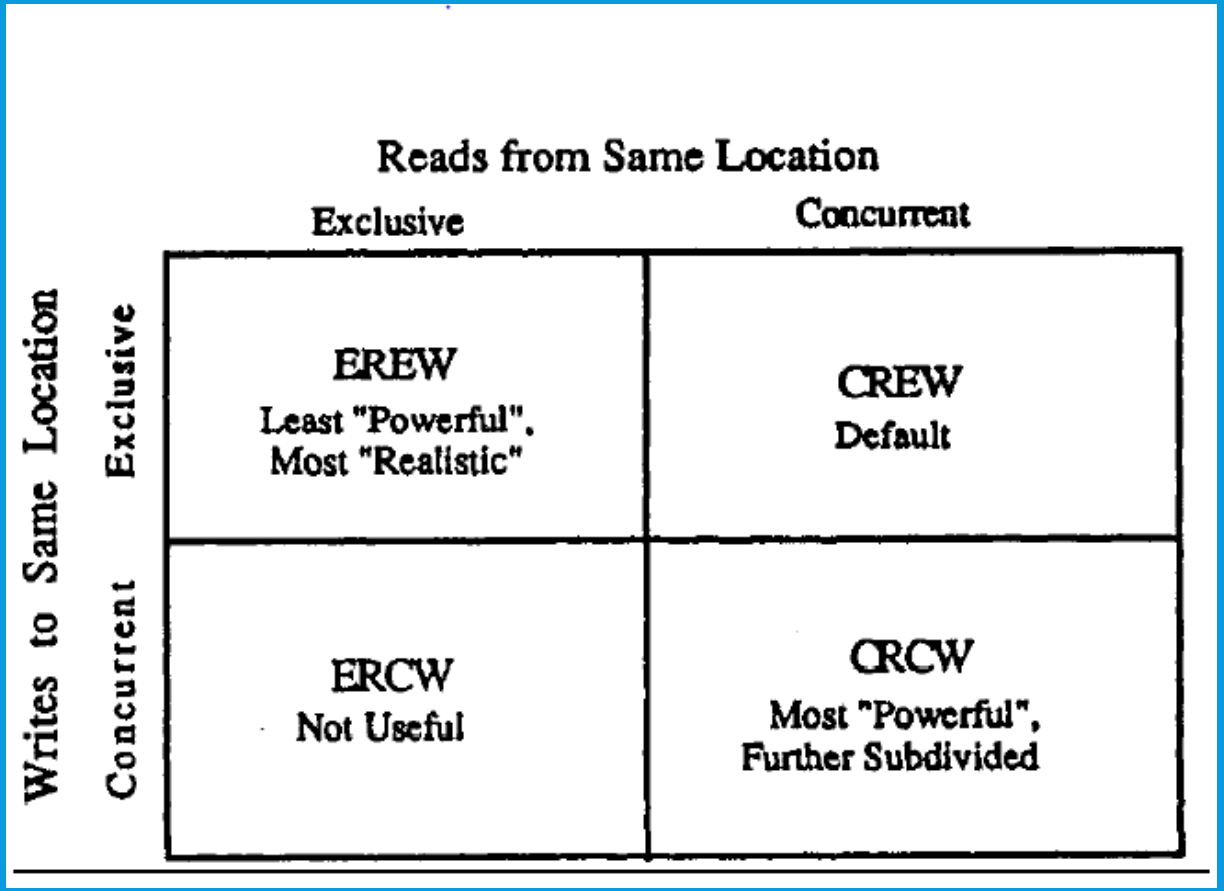
MODELS OF PARALLEL PROCESSING

- The PRAM model prescribes the concurrent operation of *p* processors (in SIMD or MIMD mode) on data that are accessible to all processors in an *m*-word shared memory.
- Processor *i* can do the following in the three phases of one cycle:
    1. Fetch an operand from the source address *si* in the shared memory
    2. Perform some computations on the data held in local registers
    3. Store a value into the destination address *di* in the shared memory

MODELS OF PARALLEL PROCESSING

- It is possible that several processors may read data from the same memory location or write their values into a common location.
- That can be shown in the PRAM sub models:
  - EREW
  - CREW
  - ERCW
  - CRCW

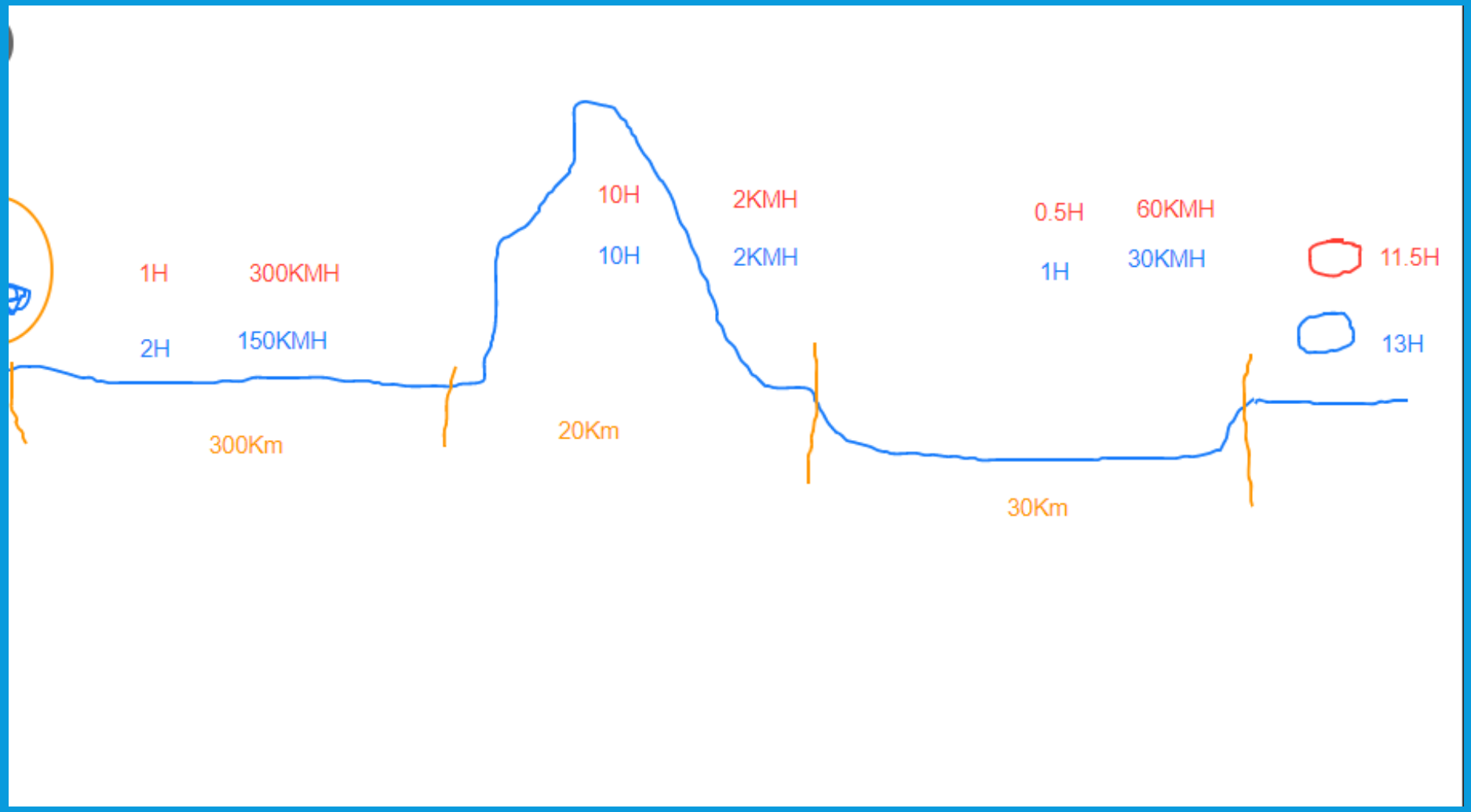
MODELS OF PARALLEL PROCESSING



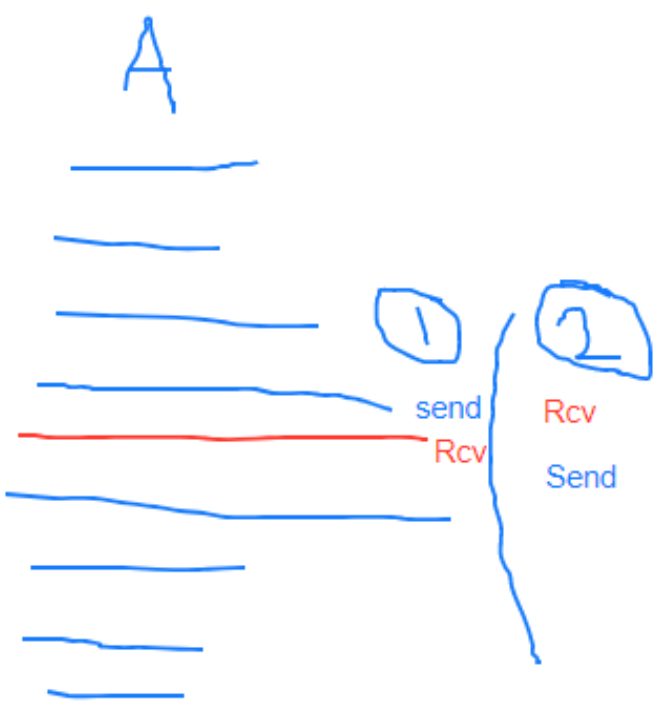
MODELS OF PARALLEL PROCESSING

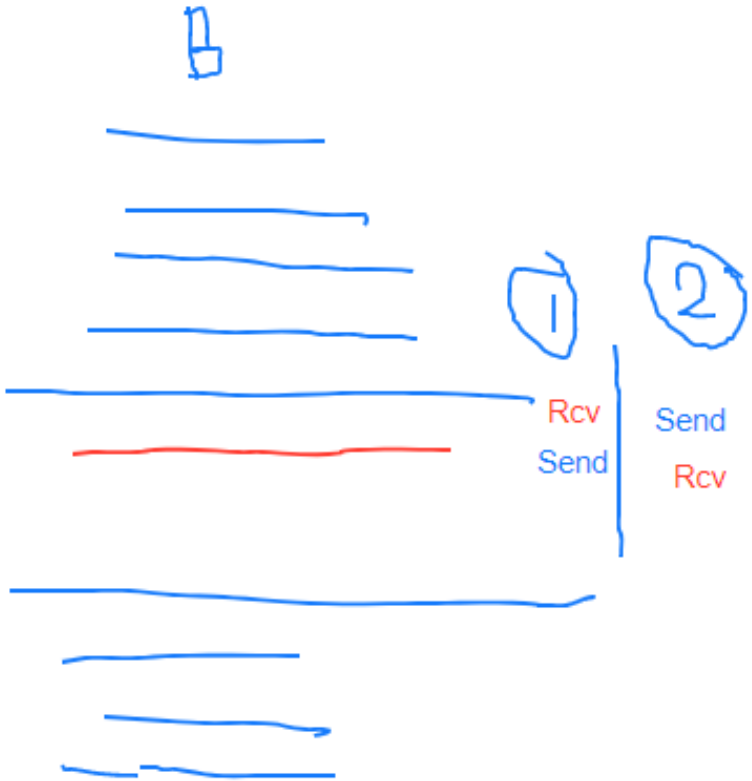
- CRCW:
  - **Common**: Multiple writes allowed only if all store the same value (CRCW-C).
  - **Random**: The value written is randomly chosen from among those offered (CRCWR).
  - **Priority**: The processor with the lowest index succeeds in writing its value (CRCW-P).
  - **Max/Min**: The largest/smallest of the multiple values is written (CRCW-M).
  - **Reduction**: The arithmetic sum (CRCW-S), logical AND (CRCW-A), logical XOR (CRCW-X), or some other combination of the multiple values is written.

MODELS OF PARALLEL PROCESSING



MODELS OF PARALLEL PROCESSING





MODELS OF PARALLEL PROCESSING