

# Paralelismo

Computadores de alto-desempenho são utilizados em diversas áreas:

- análise estrutural;
- previsão de tempo;
- exploração de petróleo;
- pesquisa em fusão de energia;
- diagnóstico médico;
- simulações aerodinâmicas;
- automação industrial;
- sensoriamento remoto;
- engenharia genética;
- e outras.

# Paralelismo

A obtenção de alto desempenho não depende somente em utilizar dispositivos de *hardware* mais rápidos, mas também em melhorias na arquitetura dos computadores e técnicas de processamento.

Arquiteturas avançadas de computadores estão baseadas no conceito de processamento paralelo:

- *pipeline*;
- arranjo de processadores;
- multiprocessadores.

Processamento paralelo é uma forma eficiente de processamento de informação que explora eventos concorrentes no processo de computação.

# Paralelismo

Mecanismos de processamento paralelo foram desenvolvidos para computadores com um único processador:

- múltiplas unidades funcionais;
- paralelismo e *pipeline* dentro da CPU;
- superposição das operações da CPU e de entrada/saída;
- hierarquia de memória;
- multiprogramação e compartilhamento de tempo.

Muitas das funções da ALU podem ser distribuídas em unidades funcionais especializadas que podem operar em paralelo.

## Paralelismo

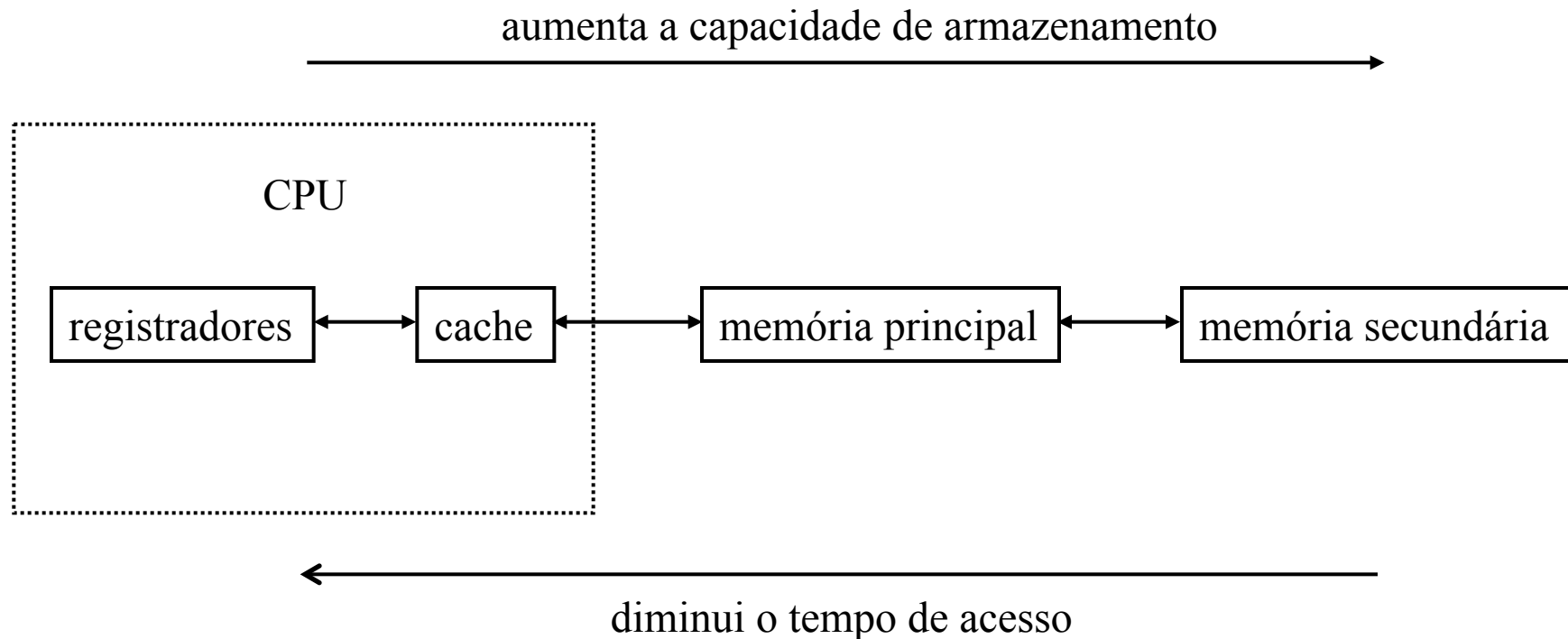
O CDC-6600 tem 10 unidades funcionais dentro da CPU, operando independentemente e simultaneamente. Um quadro de marcação registra a disponibilidade das unidades funcionais e dos registros sendo requeridos.

As operações de entrada/saída podem ser realizadas simultaneamente com as operações da CPU através do uso de controladores de entrada/saída, canais de entrada/saída ou processadores de entrada/saída.

Acesso direto à memória (DMA) pode ser utilizado para permitir transferência direta entre dispositivos de entrada/saída e a memória, numa forma de roubo de ciclo, transparente à CPU.

# Paralelismo

Hierarquia de memória oferece a possibilidade de reduzir a diferença entre a velocidade de processamento da CPU e o tempo de acesso à memória.



# Paralelismo

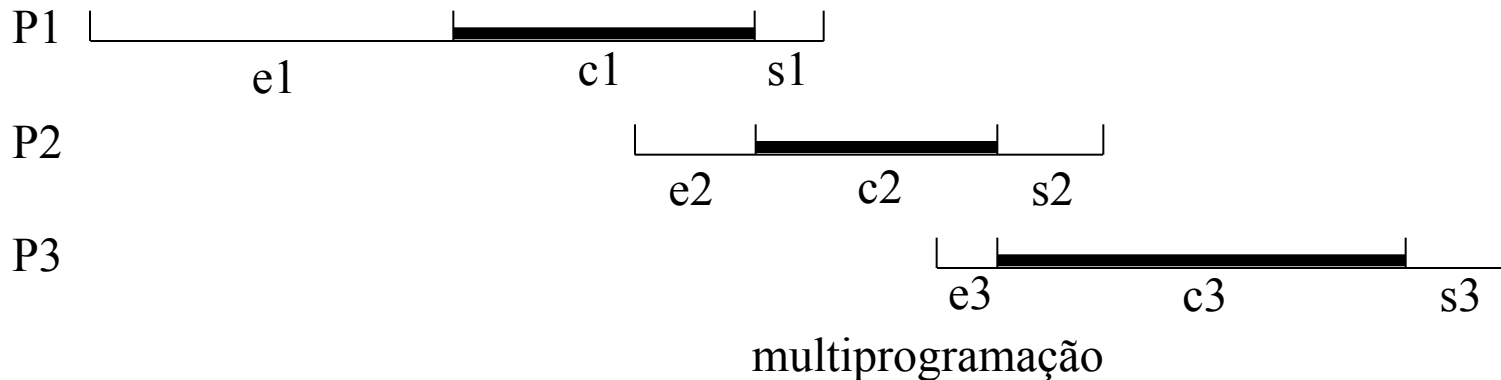
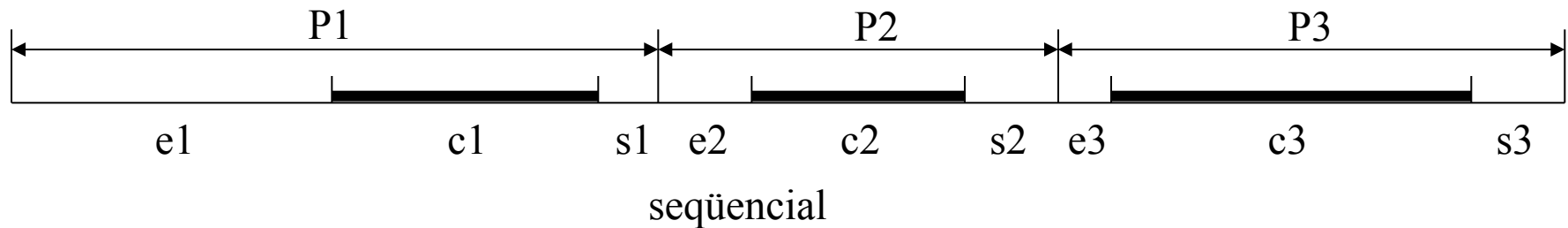
Multiprogramação e compartilhamento de tempo são soluções de *software* para obter concorrência num sistema com um processador.

Dentro de um mesmo intervalo de tempo podem haver múltiplos processos ativos no computador, competindo por recursos de memória, entrada/saída e CPU.

Multiprogramação consiste do intercalamento de programas, permitindo melhor utilização dos recursos através da superposição das operações de entrada/saída e da CPU.

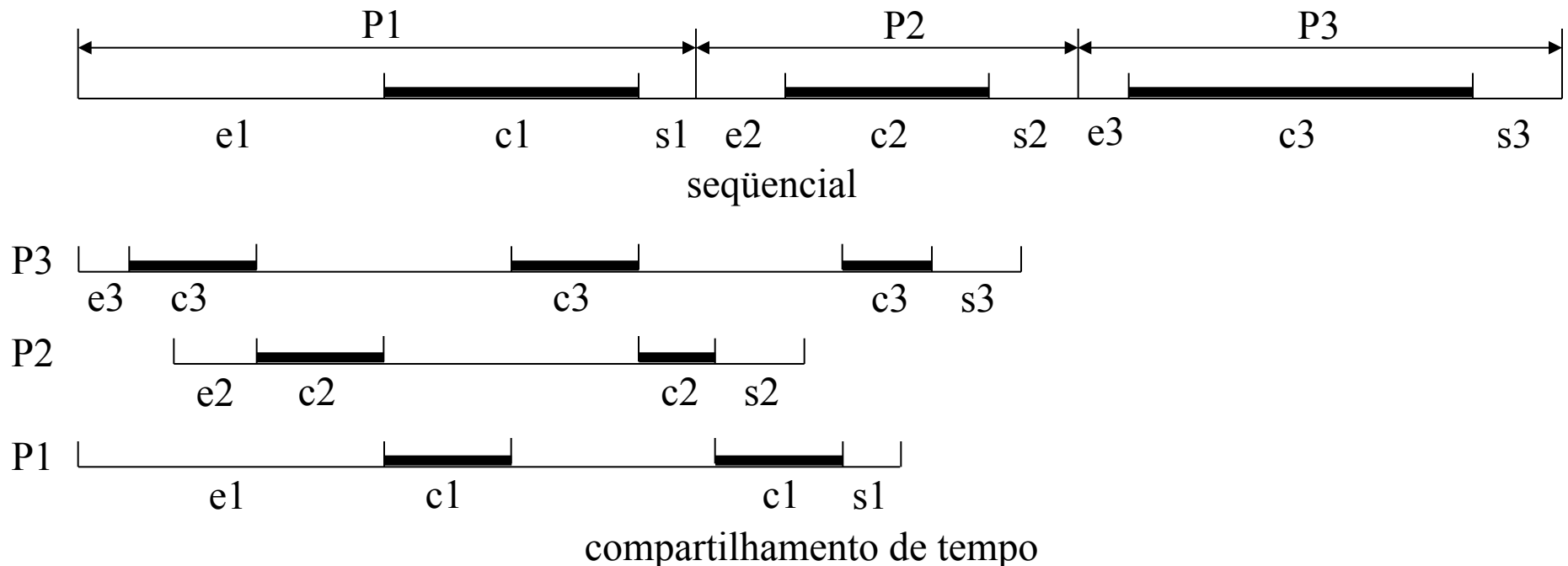
# Paralelismo

Considere os processos P1, P2 e P3, realizando operações de entrada (e), saída (s) e computação (c):



# Paralelismo

Compartilhamento de tempo consiste em atribuir fatias iguais ou variáveis do tempo de CPU para vários programas, evitando que um determinado programa venha a ocupar a CPU por um tempo muito longo.





# Paralelismo

Computadores paralelos são sistemas que suportam processamento paralelo. Podem ser divididos em três configurações arquiteturais: *pipeline*, arranjo de processadores e multiprocessadores.

*Pipeline* realiza computações superpostas a fim de explorar paralelismo temporal.

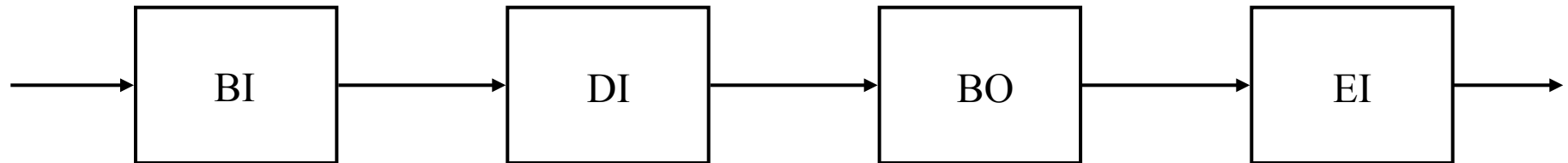
Arranjo de processadores utiliza múltiplas unidades lógicas e aritméticas síncronas para obter paralelismo espacial.

Multiprocessadores implementam paralelismo assíncrono através de um conjunto de processadores interativos com compartilhamento de recursos.

# Paralelismo

A execução de uma instrução em um computador pode ser resumida em quatro passos principais:

- busca de instrução (BI);
- decodificação de instrução (DI);
- busca de operando (BO);
- execução da instrução (EI).



# Paralelismo

O diagrama espaço-tempo abaixo mostra a seqüência de eventos para um processador *pipeline*.

tempo	1	2	3	4	5	6	7	8	9	10
estágio										
BI	I1	I2	I3	I4	I5					
DI		I1	I2	I3	I4	I5				
BO			I1	I2	I3	I4	I5			
EI				I1	I2	I3	I4	I5		

## Paralelismo

Um ciclo de instrução consiste de múltiplos ciclos do *pipeline*, que pode ser igual ao atraso do estágio mais demorado.

O fluxo de dados de um estágio a outro e as operações de todos os estágios são controlados e sincronizados por um *clock* comum do *pipeline*.

Registradores de interface são utilizados entre estágios adjacentes para armazenar resultados intermediários.

Uma vez que o *pipeline* esteja cheio, um resultado é produzido a cada ciclo. No exemplo, o ciclo de instrução foi reduzido em  $\frac{1}{4}$  do ciclo sem *pipeline*.

## Paralelismo

Teoricamente, um processador *pipeline* com  $k$  estágios poderia ser  $k$  vezes mais rápido que um sem *pipeline*.

Esse *speedup* ideal pode não ser atingido devido a conflitos de memória, dependência de dados, desvios e interrupções.

Alguns dos principais problemas no projeto de um *pipeline* são:

- seqüenciamento do programa;
- prevenção contra colisão;
- controle de congestionamento;
- tratamento de desvios;
- reconfiguração.

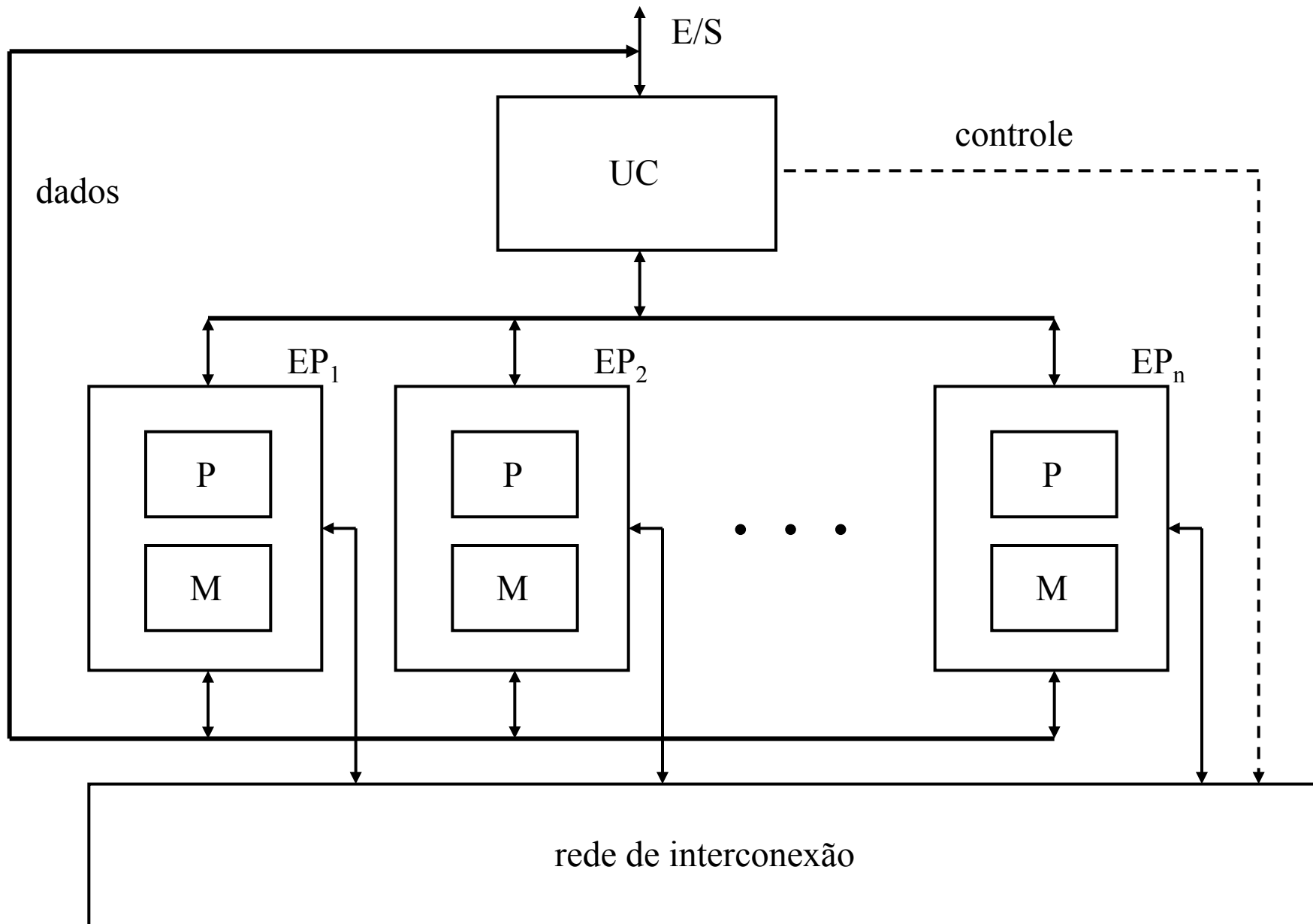
## Paralelismo

Um arranjo de processadores é um computador paralelo síncrono com múltiplas unidades lógicas e aritméticas, chamadas elementos de processamento (EP), que podem operar em paralelo.

Os EPs são sincronizados para realizar a mesma operação ao mesmo tempo, sendo constituídos de uma ALU com registradores e uma memória local, interconectados por uma rede de roteamento de dados.

A unidade de controle (UC) realiza busca e decodificação de instrução.

Em seguida, a UC controla as interconexões entre EPs e envia as instruções vetoriais a esses elementos, para serem executadas sobre os operandos obtidos das memórias locais.



## **Paralelismo**

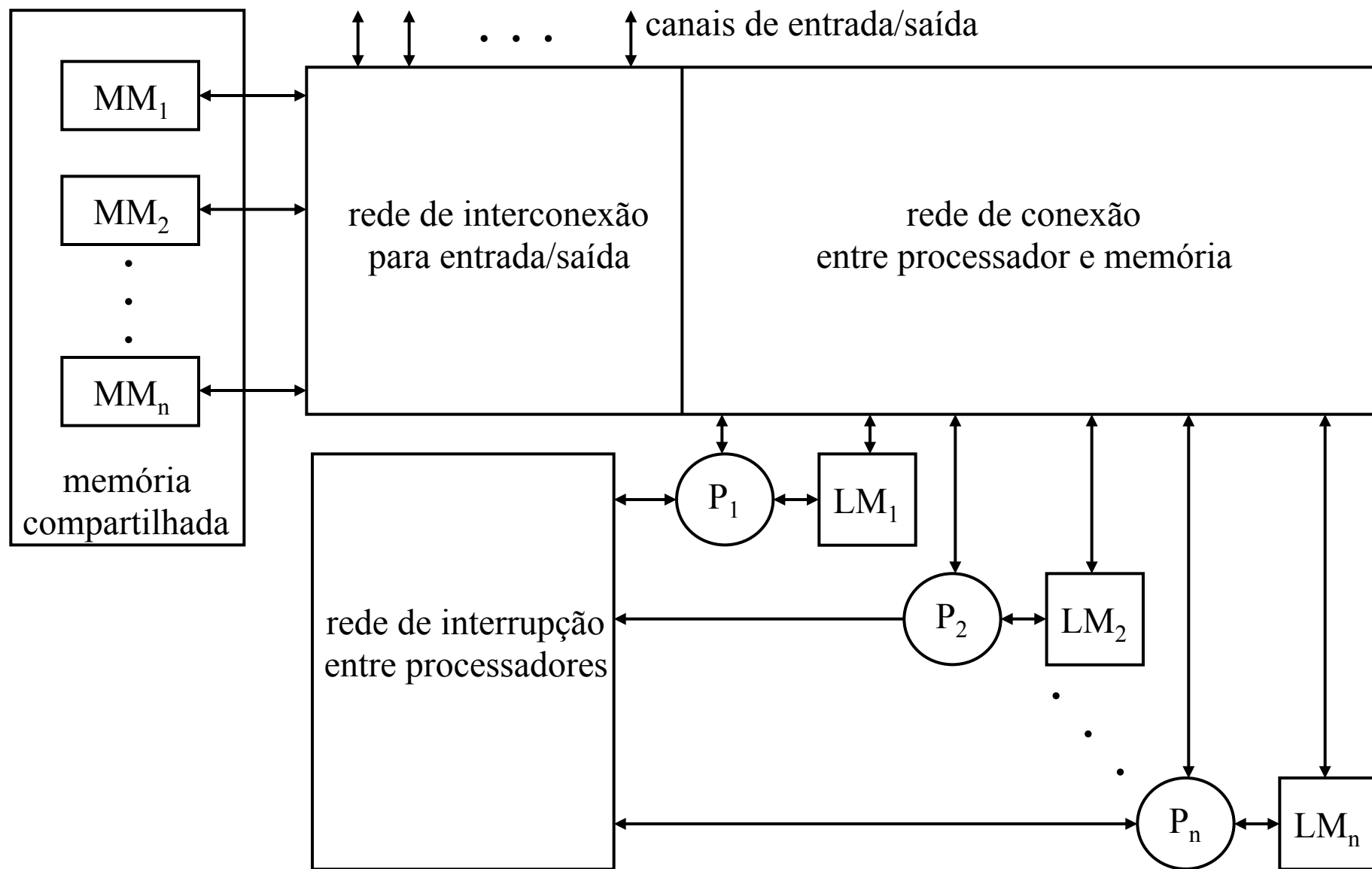
Sistemas multiprocessadores contêm dois ou mais processadores com capacidades aproximadamente idênticas, compartilhando módulos de memória, canais de entrada/saída e dispositivos periféricos.

O sistema multiprocessador deve ser controlado por um único sistema operacional, controlando as interações entre processadores e programas.

Cada processador possui a sua memória local e dispositivos próprios, sendo que a comunicação entre processadores se faz através das memórias compartilhadas ou de uma rede de interrupção.

Há três tipos de interconexão: barramento comum compartilhado no tempo, rede de chaves cruzadas e memórias de múltiplas portas.





# Paralelismo

A classificação de Flynn é baseada na multiplicidade do fluxo de instruções e de dados num computador, em que o principal processo computacional é a execução de uma seqüência de instruções sobre um conjunto de dados.

Flynn definiu quatro organizações de máquina:

- Fluxo único de instrução-fluxo único de dados (SISD);
- Fluxo único de instrução-fluxo múltiplo de dados (SIMD);
- Fluxo múltiplo de instrução-fluxo único de dados (MISD);
- Fluxo múltiplo de instrução-fluxo múltiplo de dados (MIMD).

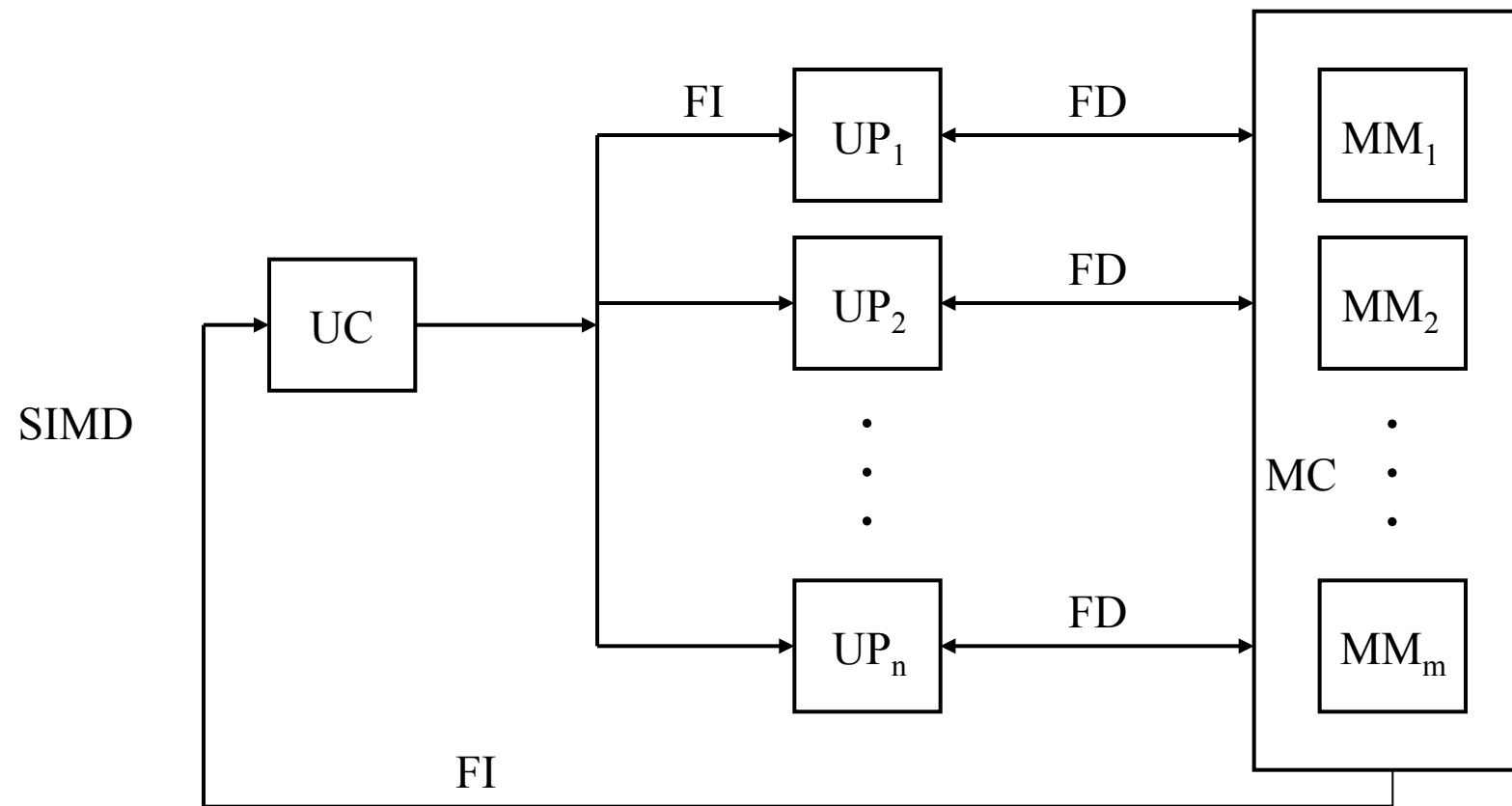
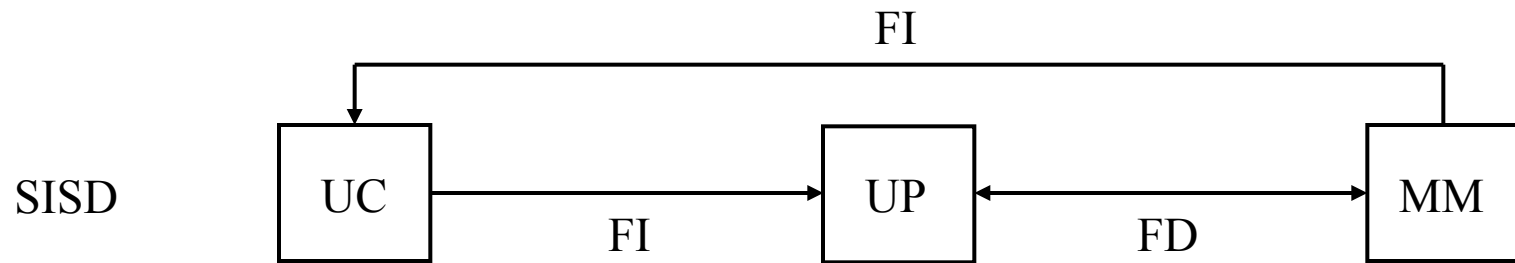
## Paralelismo

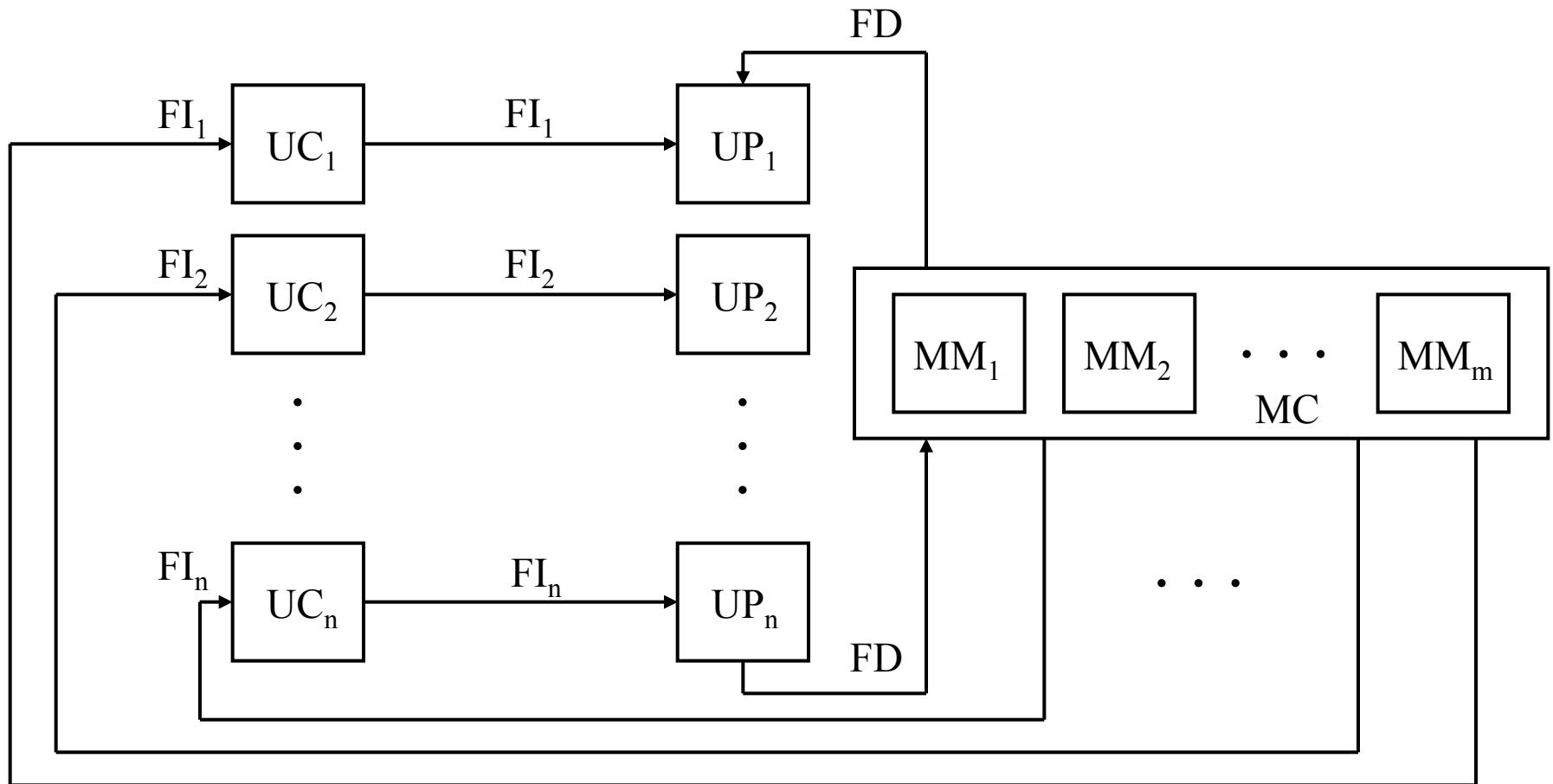
Tanto as instruções quanto os dados são buscados dos módulos de memória (MM).

As instruções são decodificadas pela unidade de controle (UC) e, em seguida, são enviadas às unidades processadoras (UP) para execução.

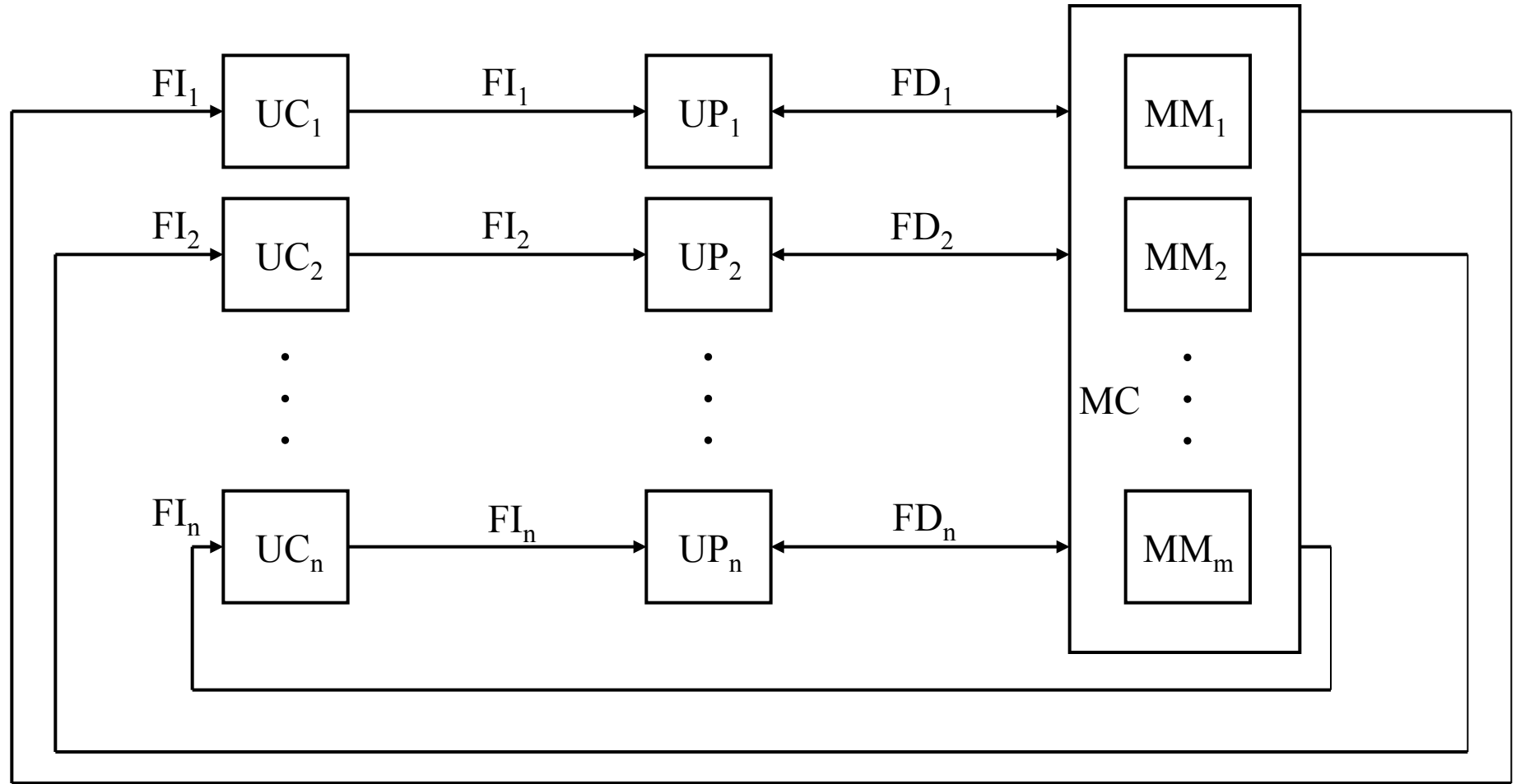
Cada fluxo de instrução (FI) é gerado por uma unidade de controle independente.

Fluxos múltiplos de dados (FD) são originados dos módulos de memória compartilhados.





MISD



MIMD

## Paralelismo

SISD é encontrada na maioria dos computadores seqüenciais disponíveis. As instruções são executadas seqüencialmente, mas podem ser superpostas (*pipeline*). Pode apresentar mais de uma unidade funcional, mas todas sob a supervisão de uma única unidade de controle.

SIMD corresponde ao arranjo de processadores, onde há múltiplos elementos processadores supervisionados pela mesma unidade de controle. Todas as unidades processadoras recebem a mesma instrução distribuída pela unidade de controle, mas operam sobre diferentes conjuntos de dados. A memória compartilhada pode conter múltiplos módulos.

## Paralelismo

MISD apresenta  $n$  unidades processadoras, cada uma recebendo instruções distintas, operando sobre o mesmo conjunto de dados. Não há implementação deste tipo de arquitetura.

MIMD corresponde a sistemas multiprocessadores. Diz-se que um sistema é fortemente acoplado se o grau de interações entre os processadores é muito alto. Caso contrário, é dito fracamente acoplado, sendo este o mais comum.