

Arquitetura de Computadores Paralelos

Tipos e Paralelismo

Dispositivos HPC

Métodos e SpeedUp

Justificativa

- Computadores paralelos fazem parte de quase todos os ambientes científicos
- Muitas empresas os empregam para acelerar seus processos
- A tecnologia HPC permeia cada vez mais a arquitetura de computadores pessoais

Tipos de paralelismo

- Paralelismo intrínseco (componentes)
- Paralelismo de hardware por função
- Multicore (== Multitask)
- Sistemas Multiprocessados
(memória compartilhada)
- Sistemas computacionais, clusters (múltiplos computadores)

Aplicações

Aplicações que exigem alto desempenho computacional: alta capacidade e desempenho sustentado (velocidade)

Exemplos de aplicações típicas:

- **modelagem física:** estrutura, fluidos, deformações. Indústria automobilística, aeroespacial, petróleo;
- **modelagem bioquímica:** estrutura molecular. Indústria química e farmacêutica;
- **análise de sinais:** astronomia, neurociências;
- **previsão de tempo;**
- **pesquisa científica em geral;**
- **órgãos governamentais:** indústria armamentista, modelagem nuclear.

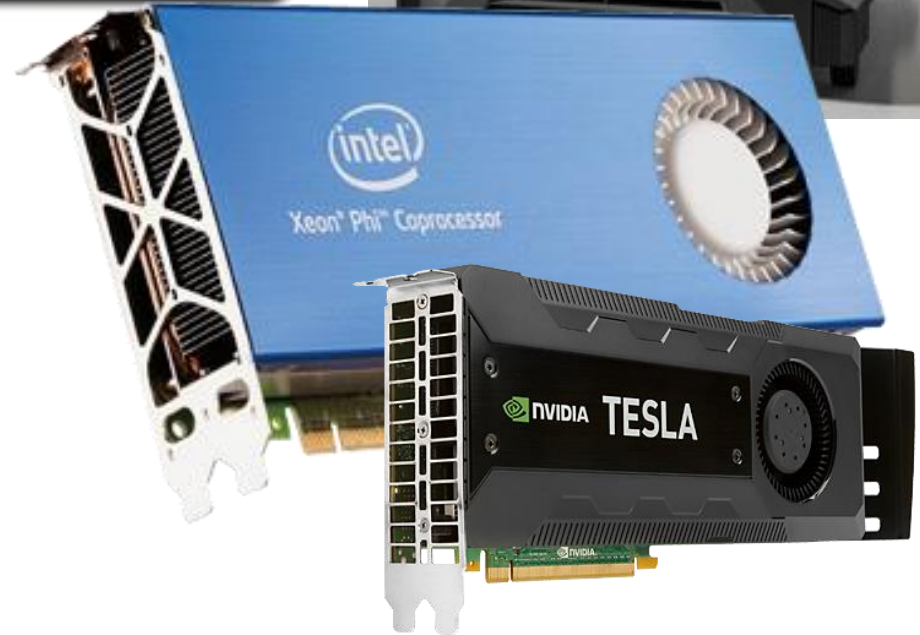
Processamento X Programação

Características do **problema** definem a **arquitetura**

- **processamento vetorial**: arquitetura principal dos supercomputadores
- **multiprocessamento** (MPP): computadores com processadores massivamente paralelos
- **clusters**: arquiteturas compostas por computadores comerciais (servers e workstations)

Principais Fabricantes

- SGI (CRAY)
- SUN
- NEC
- IBM
- INTEL
- NVIDIA
- AMD
- MICROSOFT



TOP500 (June 2015)

www.top500.org

Rank	Site	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	National Super Computer Center in Guangzhou China	Tianhe-2 (MilkyWay-2) - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P NUDT	3,120,000	33,862.7	54,902.4	17,808
2	DOE/SC/Oak Ridge National Laboratory United States	Titan - Cray XK7 , Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x Cray Inc.	560,640	17,590.0	27,112.5	8,209
3	DOE/NNSA/LLNL United States	Sequoia - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM	1,572,864	17,173.2	20,132.7	7,890
4	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect Fujitsu	705,024	10,510.0	11,280.4	12,660
5	DOE/SC/Argonne National Laboratory United States	Mira - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM	786,432	8,586.6	10,066.3	3,945

#1 China



[Home](#) / [National Supercomputin...](#) / [Sunway TaihuLight - Sunway MPP, Sunwa...](#)

Sunway TaihuLight - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway

Site:	National Supercomputing Center in Wuxi
Manufacturer:	NRCP
Cores:	10,649,600
Linpack Performance (Rmax)	93,014.6 TFlop/s
Theoretical Peak (Rpeak)	125,436 TFlop/s
Nmax	12,288,000
Power:	15,371.00 kW (Submitted)
Memory:	1,310,720 GB
Processor:	Sunway SW26010 260C 1.45GHz
Interconnect:	Sunway
Operating System:	Sunway RaiseOS 2.0.5

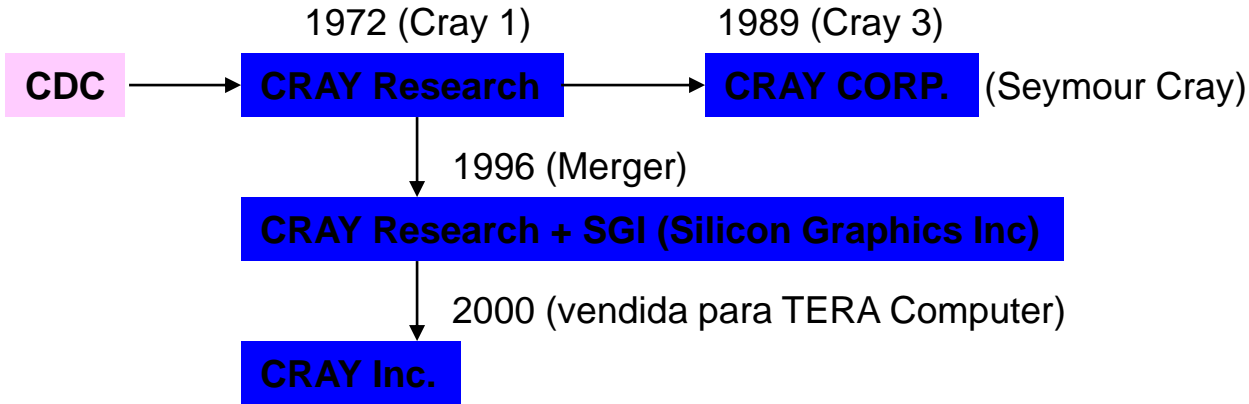
Tianhe-2 (MilkyWay-2)

TH-IVB-FEP Cluster, Intel Xeon, TH Express-2, Intel Xeon Phi 31S1P

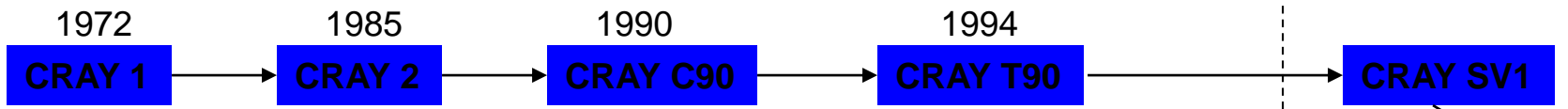
Site:	<u>National Super Computer Center in Guangzhou</u>
Manufacturer:	NUDT
Cores:	3,120,000
Linpack Performance (Rmax)	33,862.7 TFlop/s
Theoretical Peak (Rpeak)	54,902.4 TFlop/s
Nmax	9,960,000
Power:	17,808.00 kW
Memory:	1,024,000 GB
Processor:	Intel Xeon E5-2692v2 12C 2.2GHz
Interconnect:	TH Express-2
Operating System:	Kylin Linux
Compiler:	icc
Math Library:	Intel MKL-11.0.0
MPI:	MPICH2 with a customized GLEX channel

HISTÓRICO - CRAY

SEYMOUR CRAY
(1925 - 1996)



VECTOR PROCESSORS



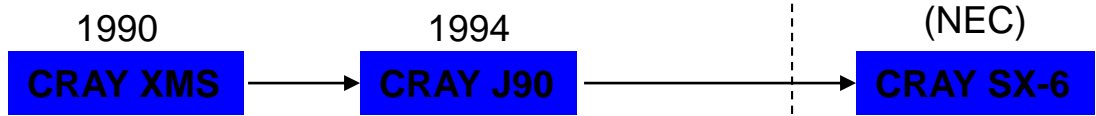
ATUAL

MULTIPROCESSORS



MPP

MINISUPER



CRAY T3E



- 40 a 2176 Elementos RISC
- 256 ou 512 MB RAM cada
- conexão Toroidal duplo
- 42 a 166 GB/s

CRAY MTA



- projeto original da TERA
- paralelismo compilador
- multi-thread
- 16 a 256 processadores
- memória compartilhada até 1 TB
- facilidade programação

CRAY SV1



- arquitetura clássica vetorial, até 24 elementos
- CPU de 8 GFlop
- Memória 128 GB

CRAY HPC-Cluster



- conjunto de Elementos comerciais de baixo custo: Dell PowerEdge
- interconexão rede de alta velocidade
- software da Cray (Linux)
- balanceamento

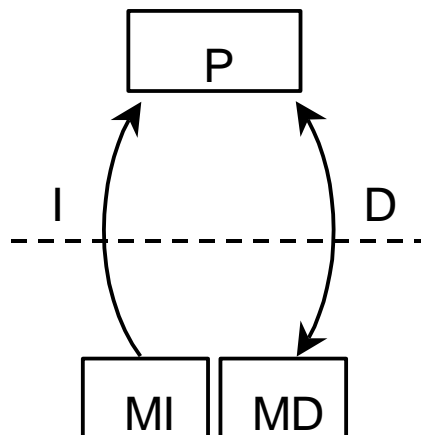
CLASSIFICAÇÃO DOS COMPUTADORES PARALELOS

Uma classificação genérica de um sistema de computador que perdura até hoje devida à sua simplicidade, é aquela clássica, apresentada por Flynn [FLYNN, 1972], na qual uma máquina é caracterizada através da maneira como trata as suas instruções e os seus dados. As categorias são:

- *SISD* - “Single Instruction, Single Data”;
- *SIMD* - “Single Instruction, Multiple Data”;
- *MISD* - “Multiple Instruction, Single Data”;
- *MIMD* - “Multiple Instruction, Multiple Data”.

SISD

Computador von Neumann



Legenda:

P: Processador

MI: Memória de Instruções

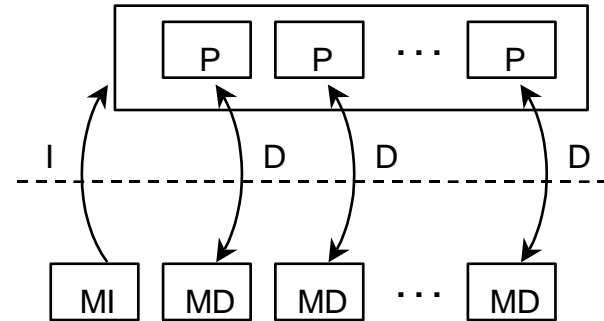
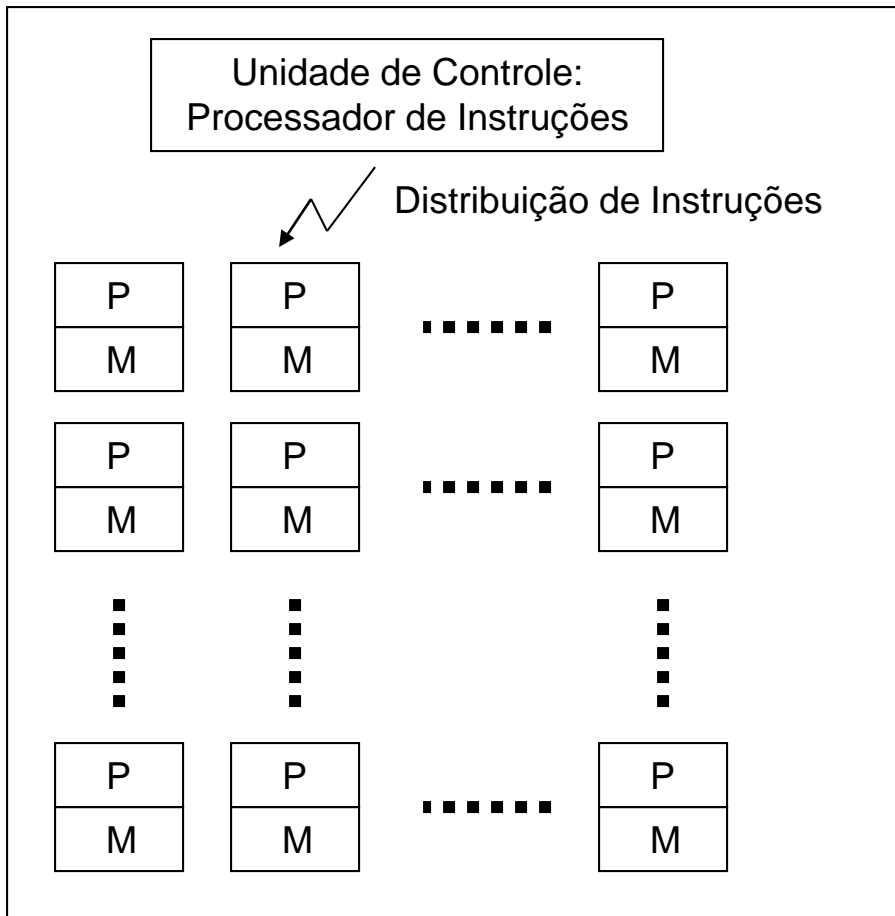
MD: Memória de Dados

I: Fluxo de Instruções

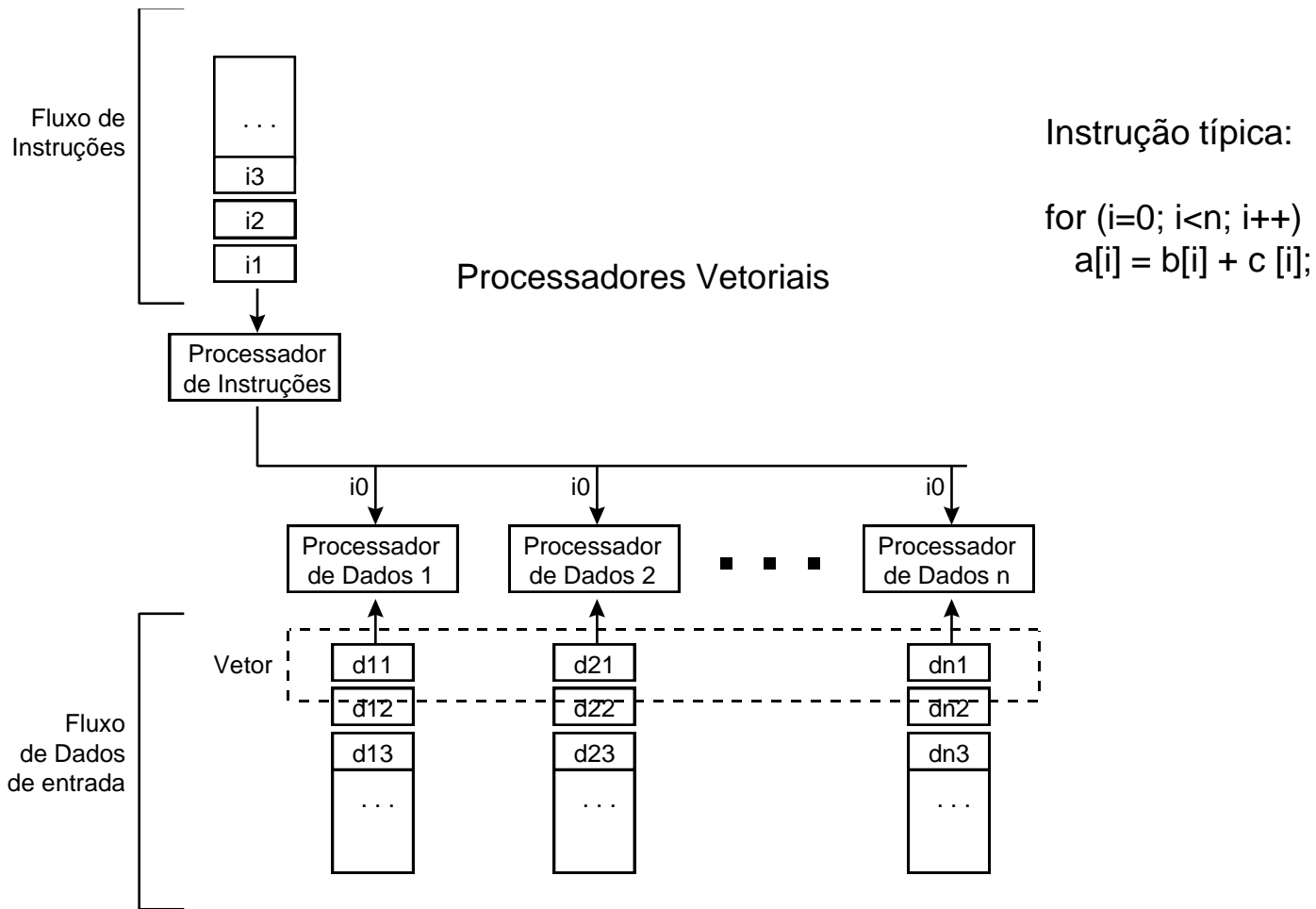
D: Fluxo de Dados

SIMD

Array Processors



MÁQUINAS VECTORIAIS



Obs.: por simplificação, não foram representados os demais operandos e os dados de saída

Exemplo de Computador Vetorial SIMD

Supercomputador Cray-1

- computador mais poderoso da época (meio a fins de 1970)
- clock 80 MHz e 8 MB de memória principal

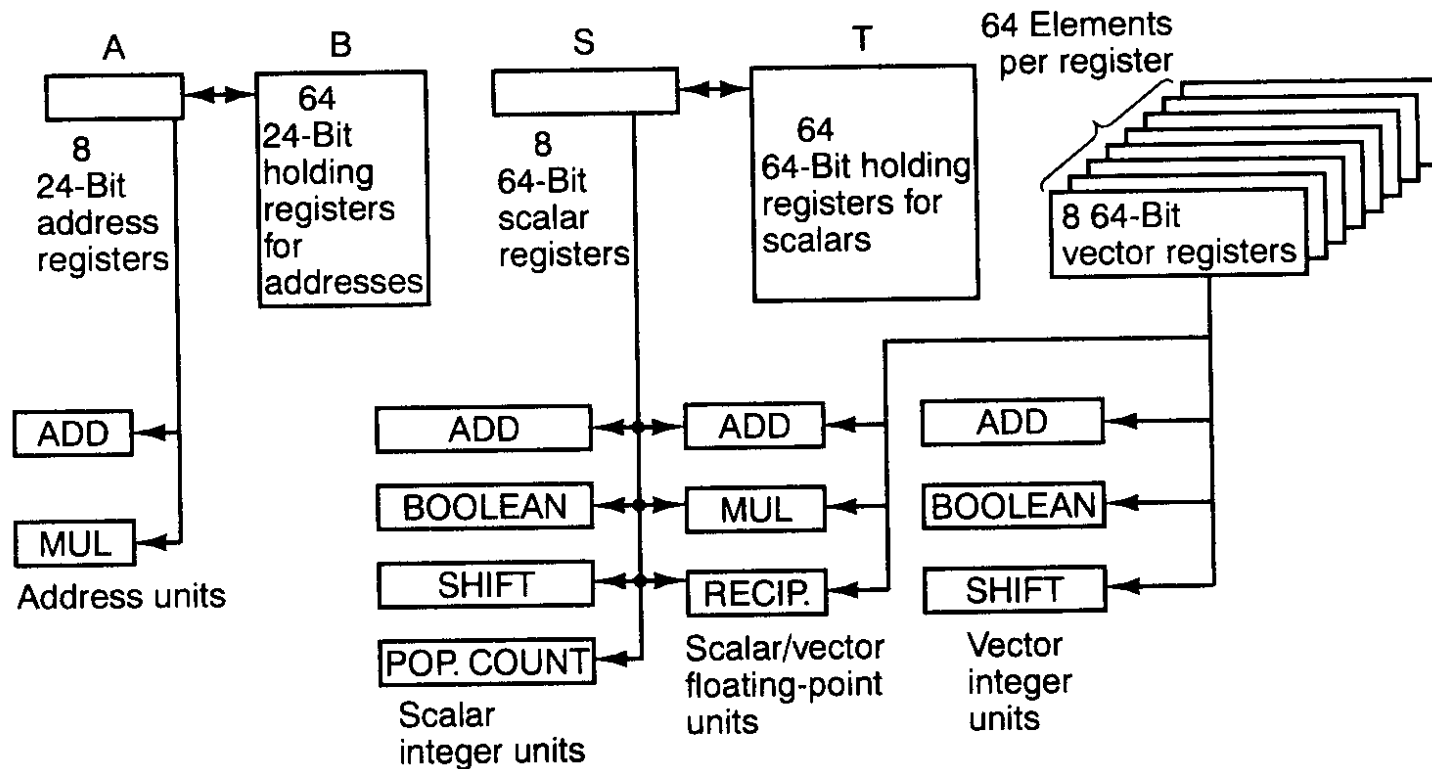
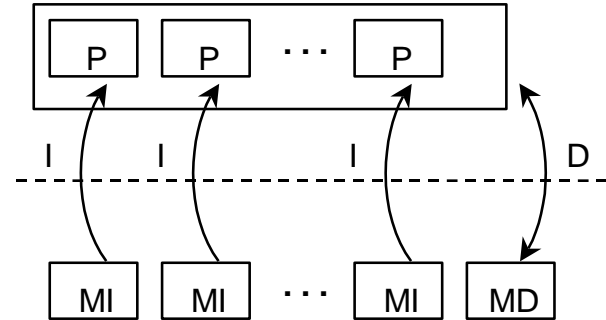
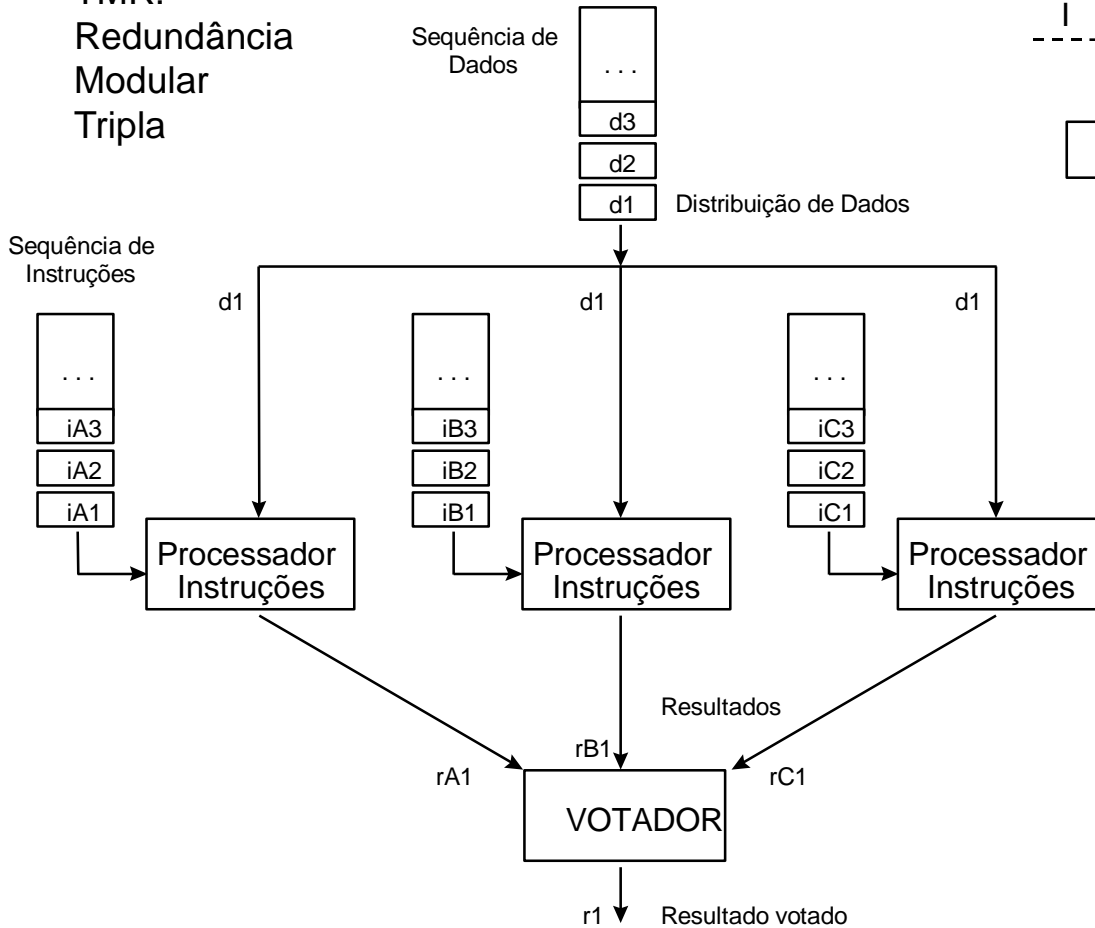


Figure 8-19. Registers and functional units of the Cray-1

MISD

TMR:
Redundância
Modular
Tripla

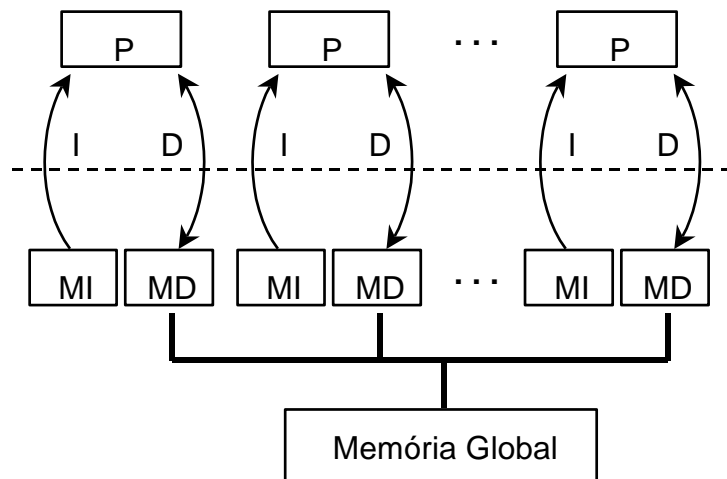


Arquitetura de processadores
Tolerantes a Falha

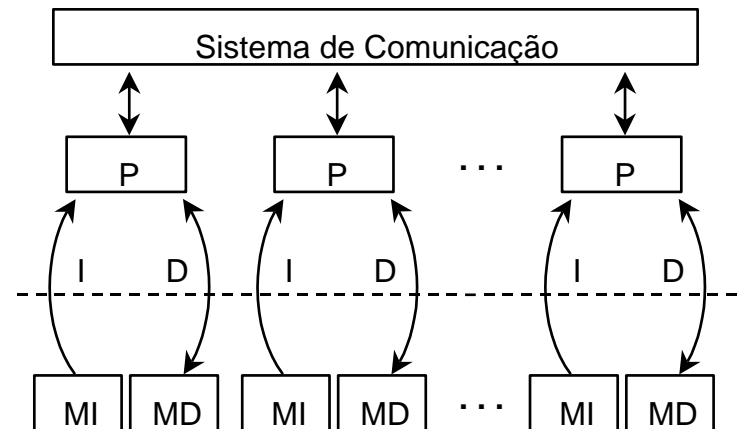
MIMD

Multiprocessamento: Sistemas Massivamente Paralelos (MMP)
Multicomputadores: Sistemas Distribuídos

MIMD com memória compartilhada



MIMD com troca de mensagens



MIMD MEMÓRIA CACHE

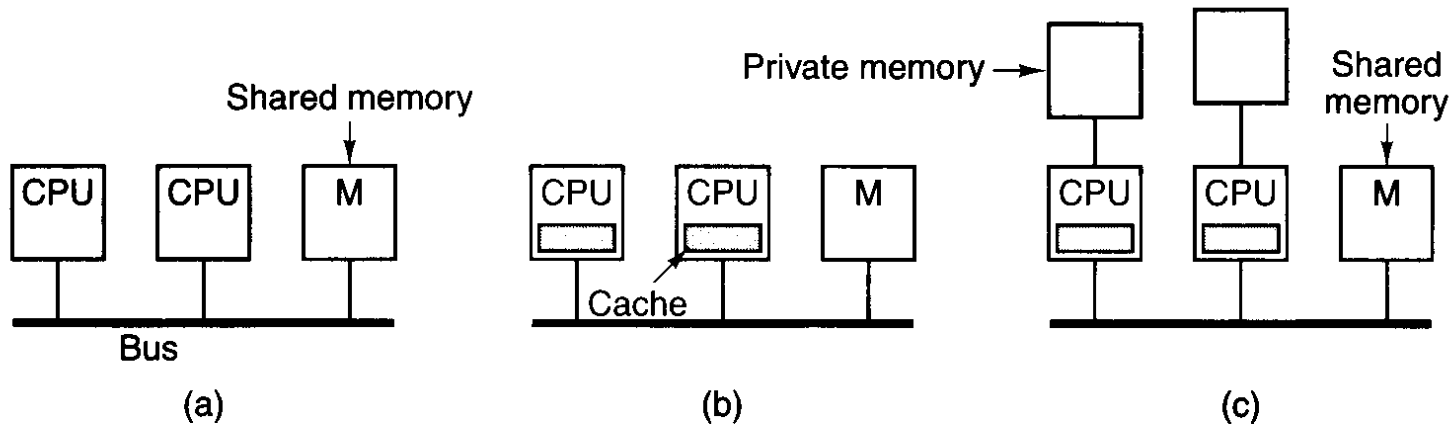
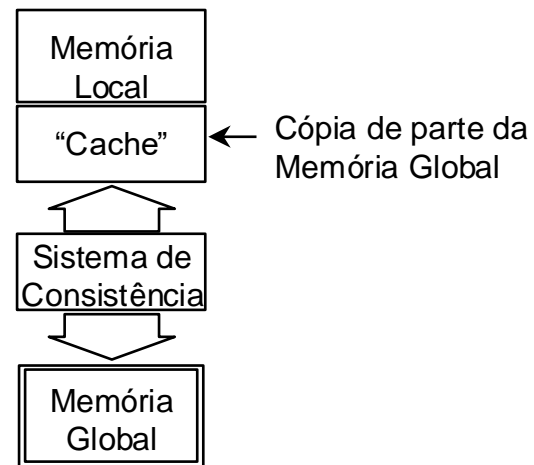


Figure 8-22. Three bus-based multiprocessors. (a) Without caching. (b) With caching. (c) With caching and private memories.

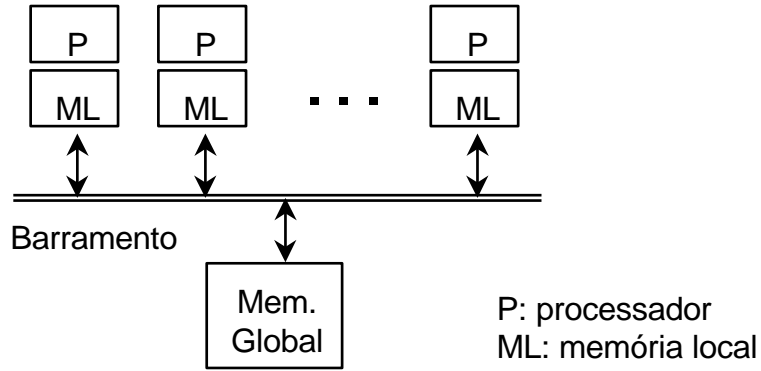
Na prática:

- barramento é o gargalo do sistema
- dificuldade em manter a consistência do Cache
- limite no número de UCPs (dezenas)

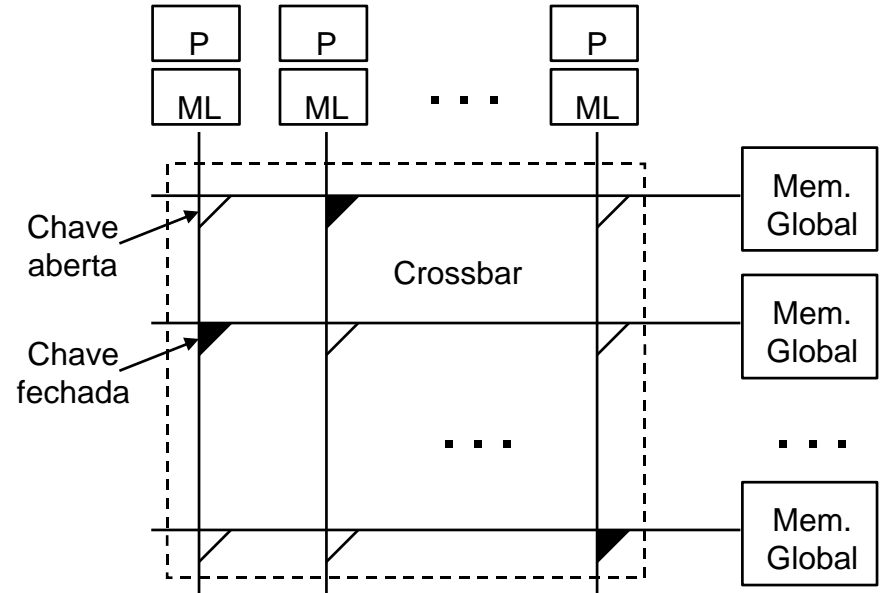


MIMD MEMÓRIA COMPARTILHADA

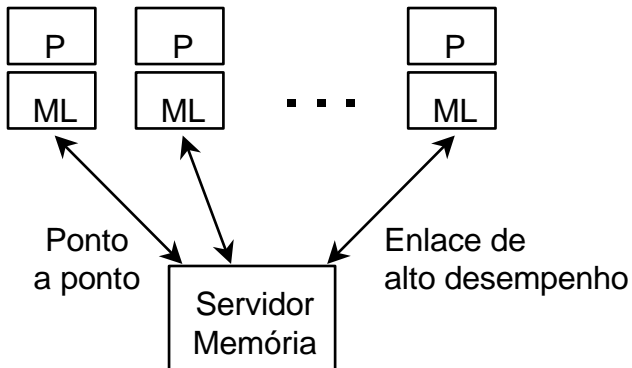
Através de Barramento



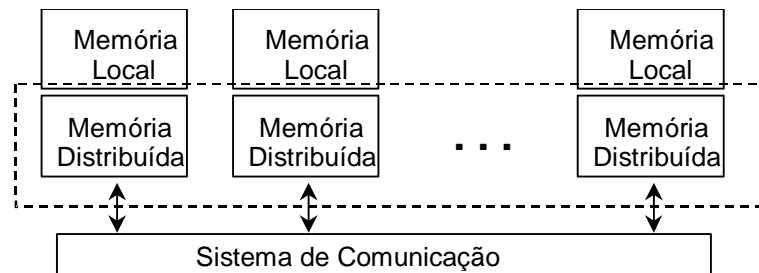
Através de Crossbar



Através de Servidor de Memória

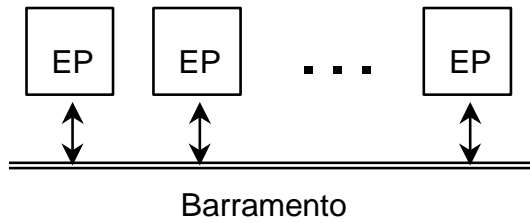


Memória Distribuída com controle de integridade

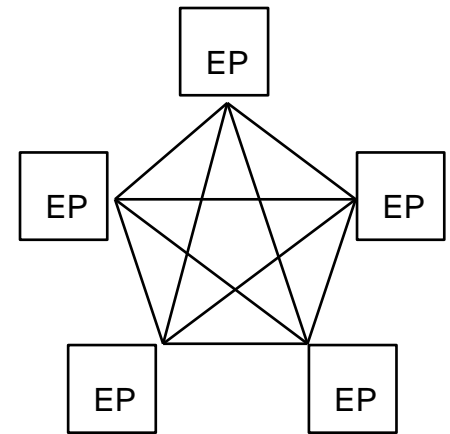


MIMD TROCA DE MENSAGENS

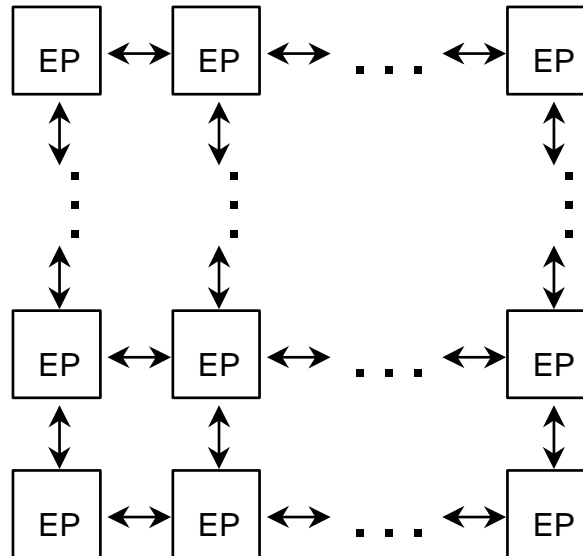
Interconexão por barramento serial ou paralelo



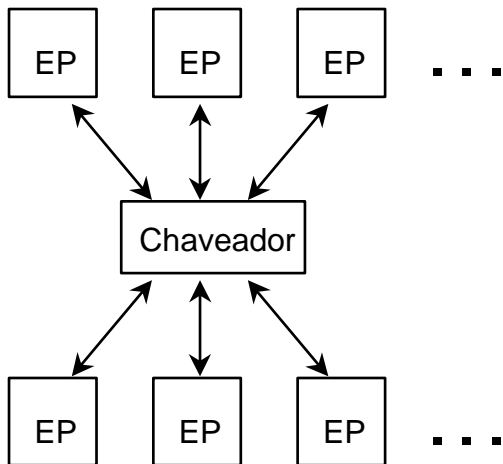
Conexão total ponto a ponto



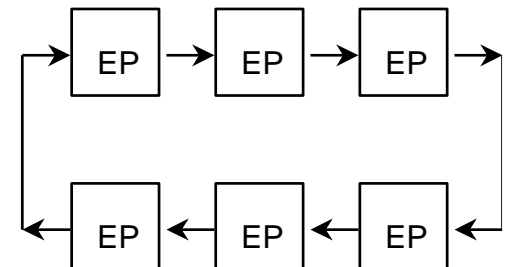
Conexão em Malha (Mesh)



Conexão por chaveador



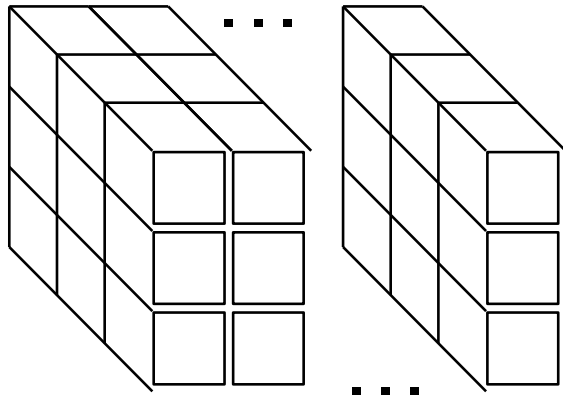
Conexão em anel



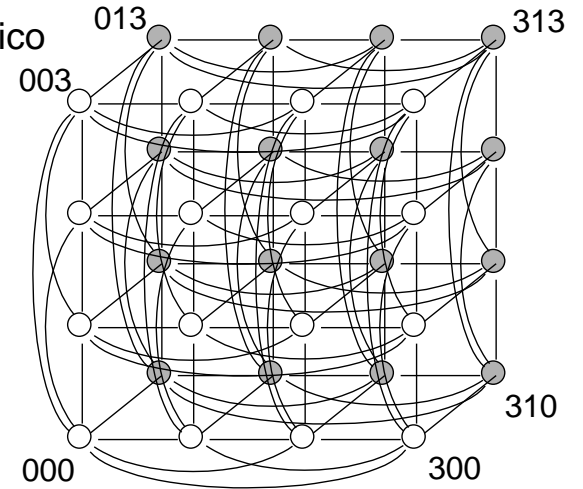
Problema básico: como interligar milhares de Elementos Processadores - Questão de topologia de rede

MIMD TROCA DE MENSAGENS: HIPERCUBOS

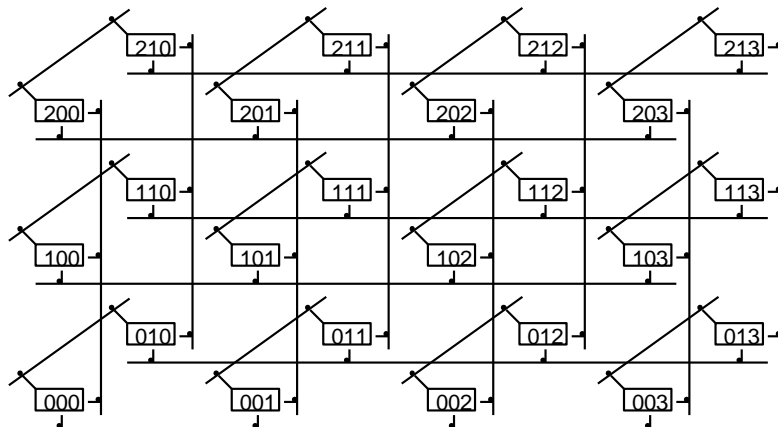
Hipercubo Clássico

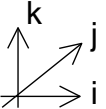


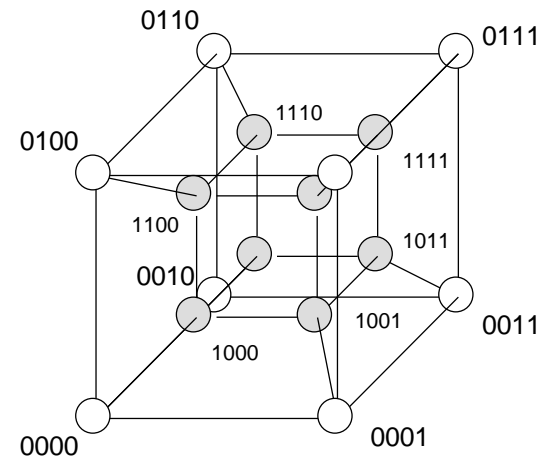
Hipercubo Genérico



Hipercubo com Múltiplo Acesso



Coordenadas: ijk  **Booleano**



TOPOLOGIAS DE INTERCONEXÃO

Conexão de:
Memórias
Processadores
Computadores

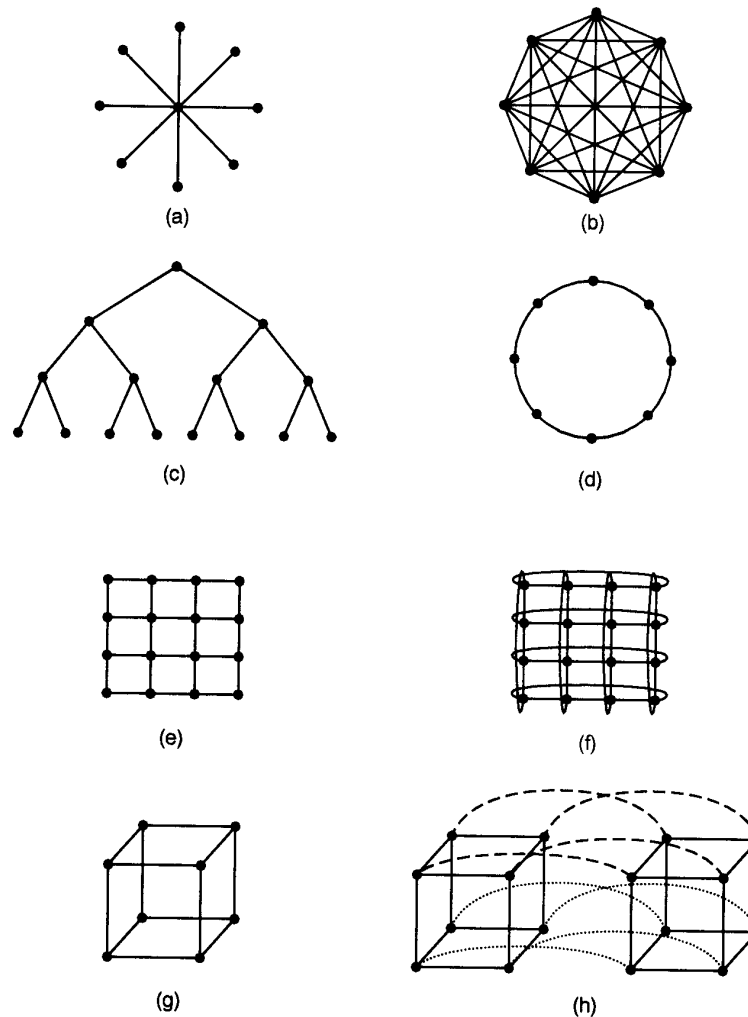


Figure 8-4. Various topologies. The heavy dots represent switches. The CPUs and memories are not shown. (a) A star. (b) A complete interconnect. (c) A tree. (d) A ring. (e) A grid. (f) A double torus. (g) A cube. (h) A 4D hypercube.

QUESTÕES BÁSICAS

NÍVEIS DE PARALELISMO:

divisão a partir de uma atividade única

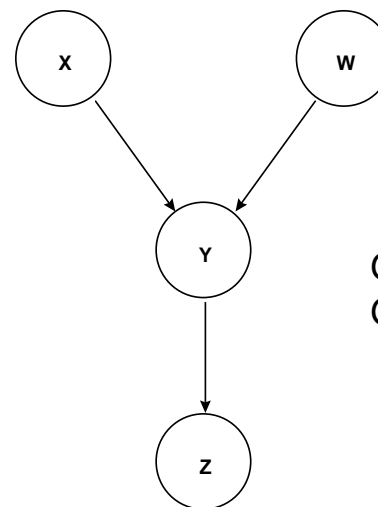
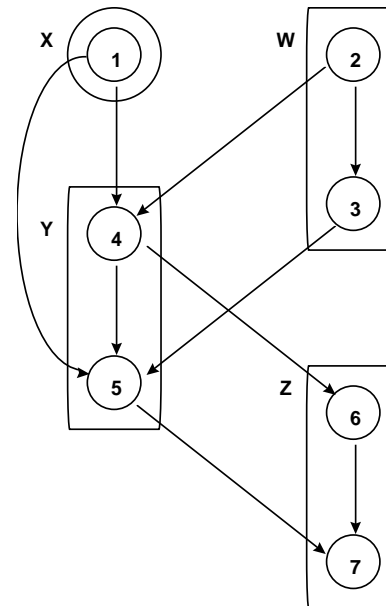
- execução de trabalho ("job"),
- execução de tarefa ("task"),
- execução de função ("thread"),
- execução de instrução,
- transferência de registros e dispositivos lógicos (hardware).

ORGANIZAÇÃO para execução paralela

- particionamento / granularidade;
- escalação e distribuição (fork);
- sincronização;
- comunicação (join).

Grau de ACOPLAMENTO:

- fraco: sistemas distribuídos
- forte: nível de registros



PARADIGMAS DE PROCESSAMENTO

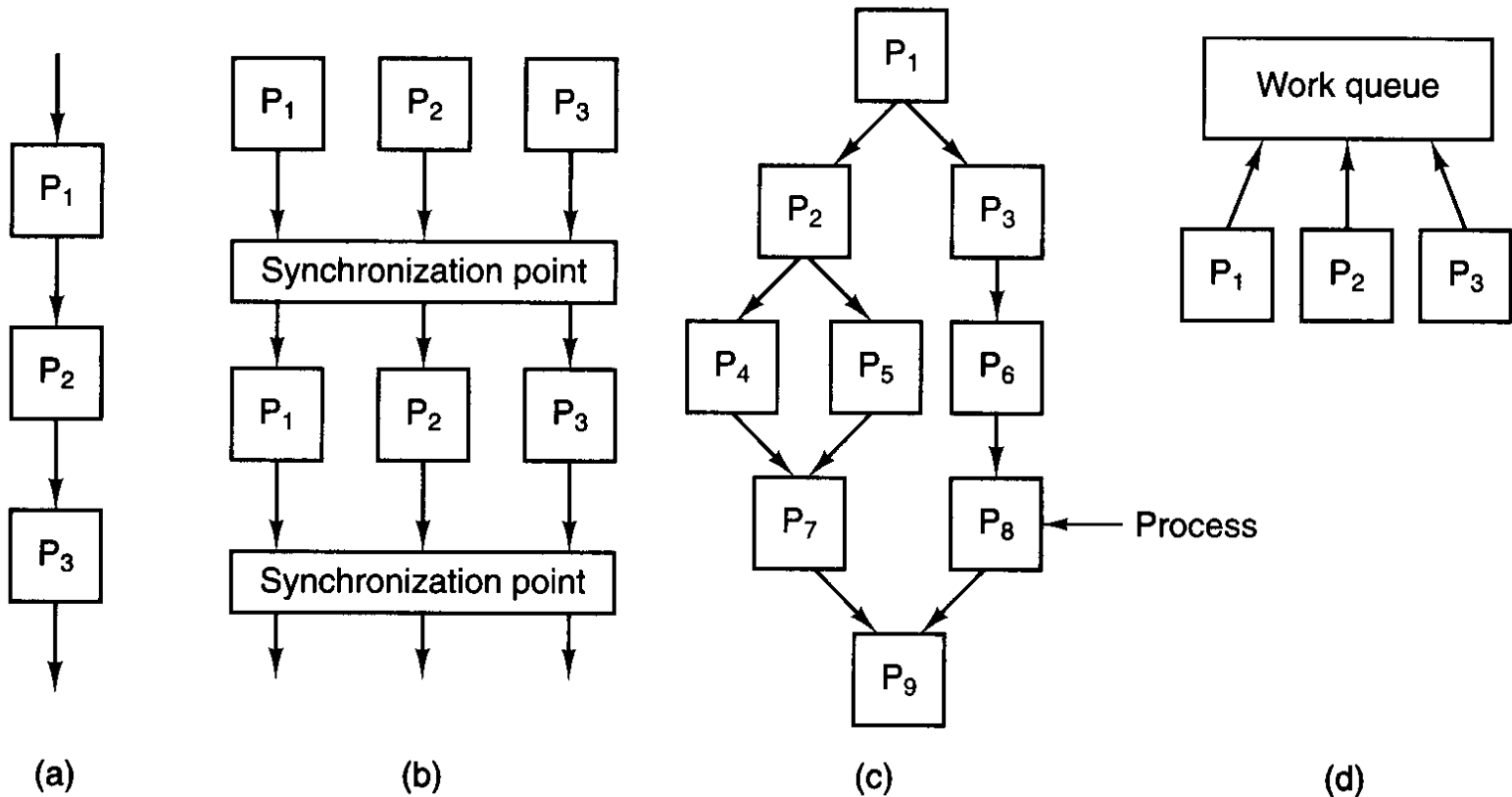
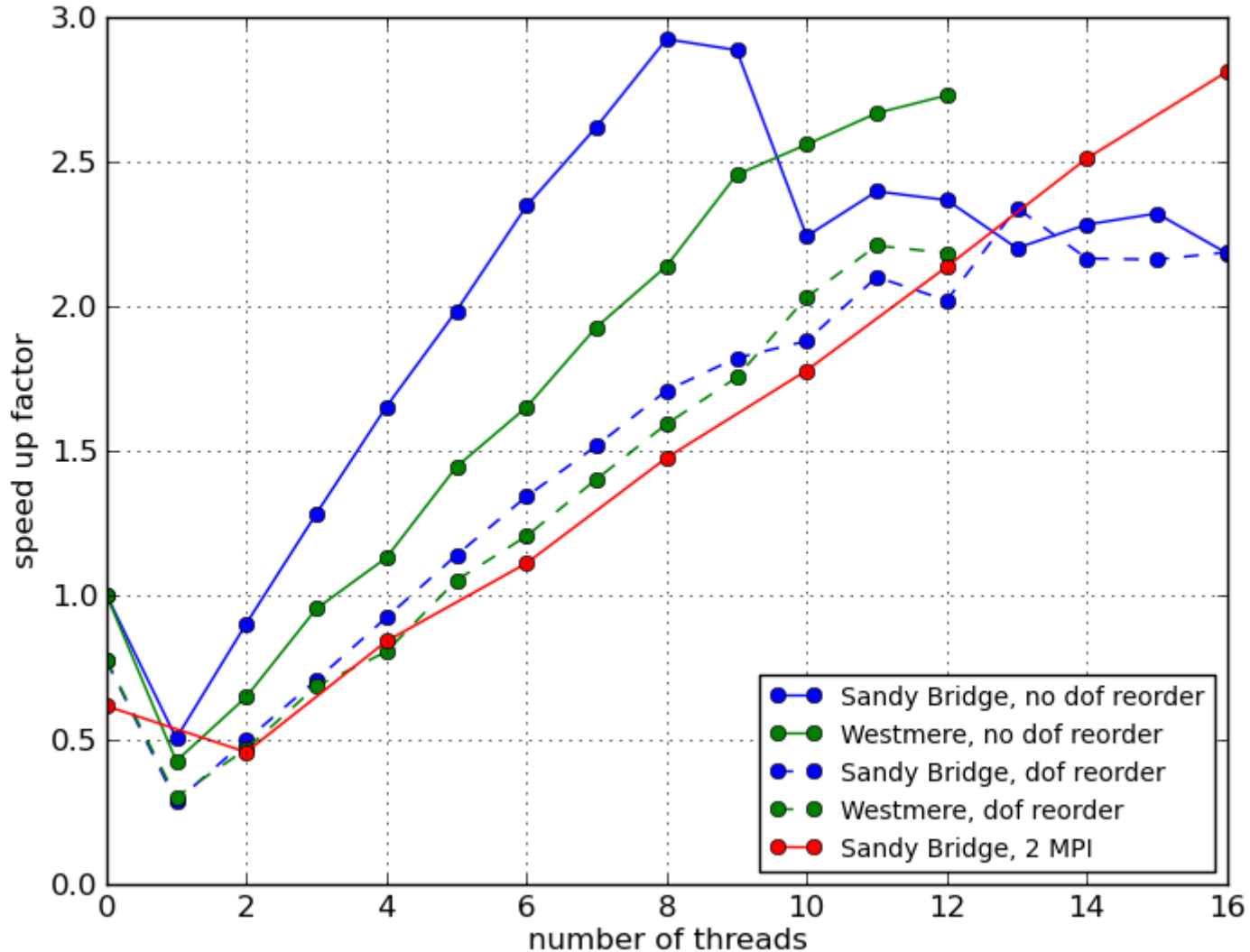


Figure 8-11. Computational paradigms. (a) Pipeline. (b) Phased computation. (c) Divide and conquer. (d) Replicated worker.

SPEED-UP

Quantas vezes mais rápido o programa é executado?



GRANULARIDADE

$$G = \frac{\text{Quantidade de computação}}{\text{Quantidade de comunicação}}$$

Paralelismo granular

Fine grained = Blocos pequenos concentrados em múltiplas tarefas

Coarse grained = Blocos grandes concentrados em poucas tarefas

Quanto mais fina a granularidade, maior o potencial de paralelismo e consequentemente, speed-ups, porém, maior o overhead (comunicação)

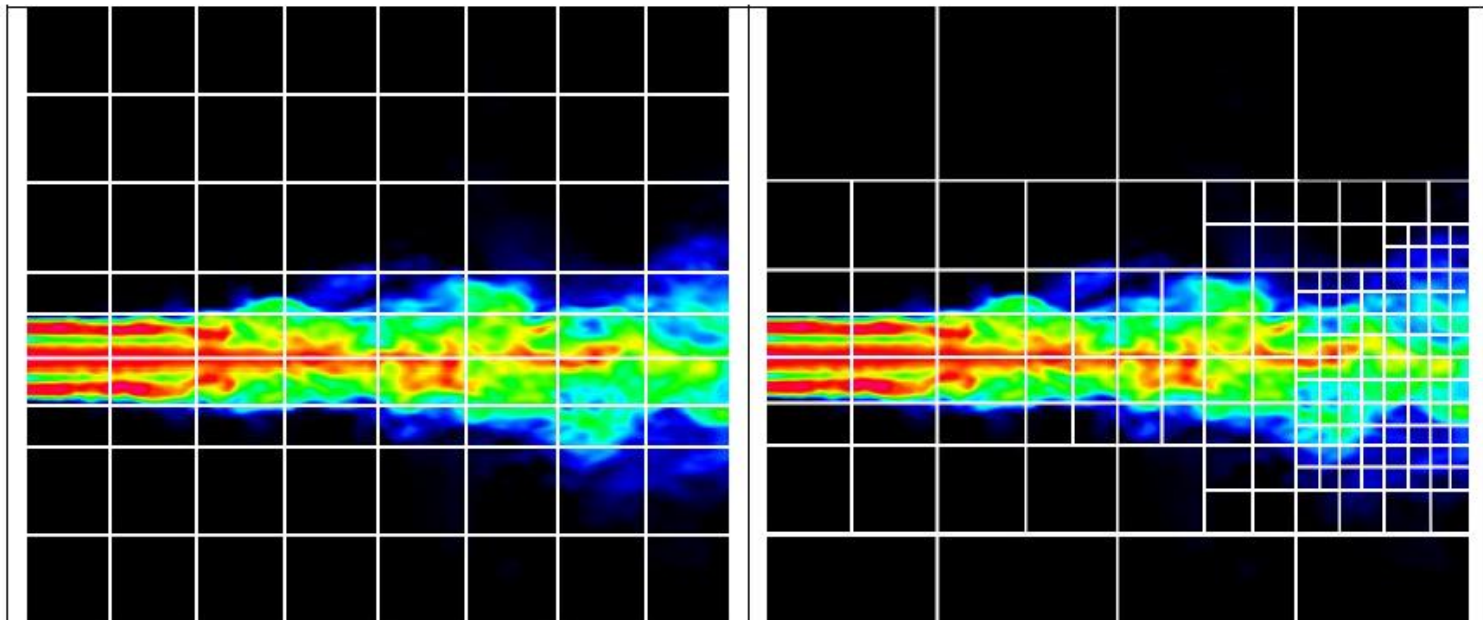
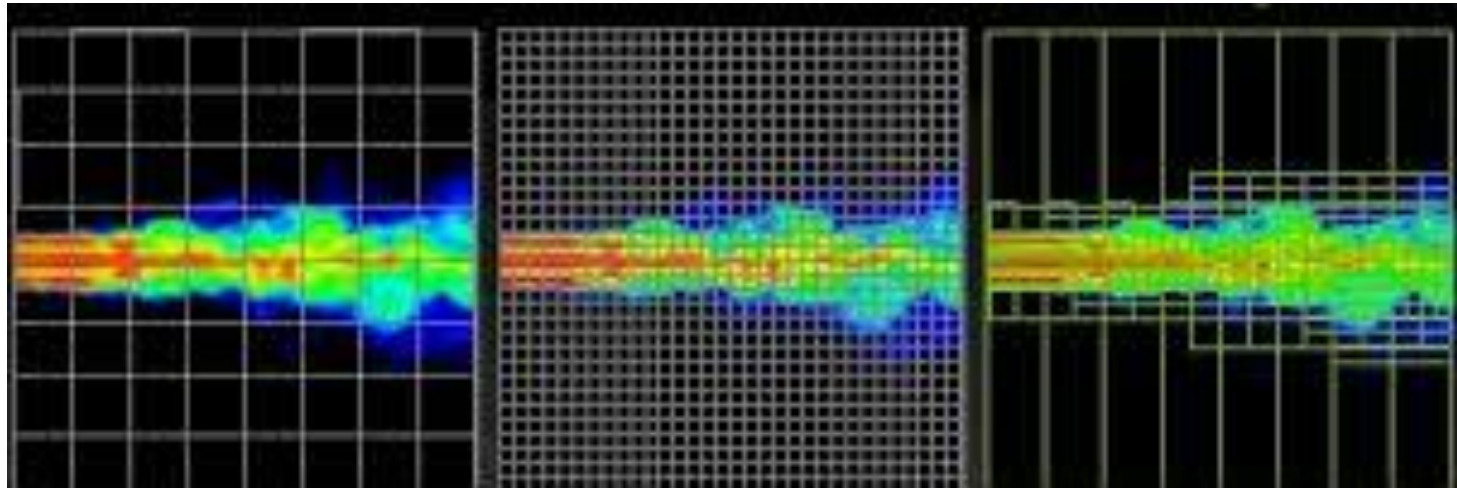
A estratégia de particionamento depende do problema, da abordagem e da arquitetura computacional empregada

PARTICIONAMENTO

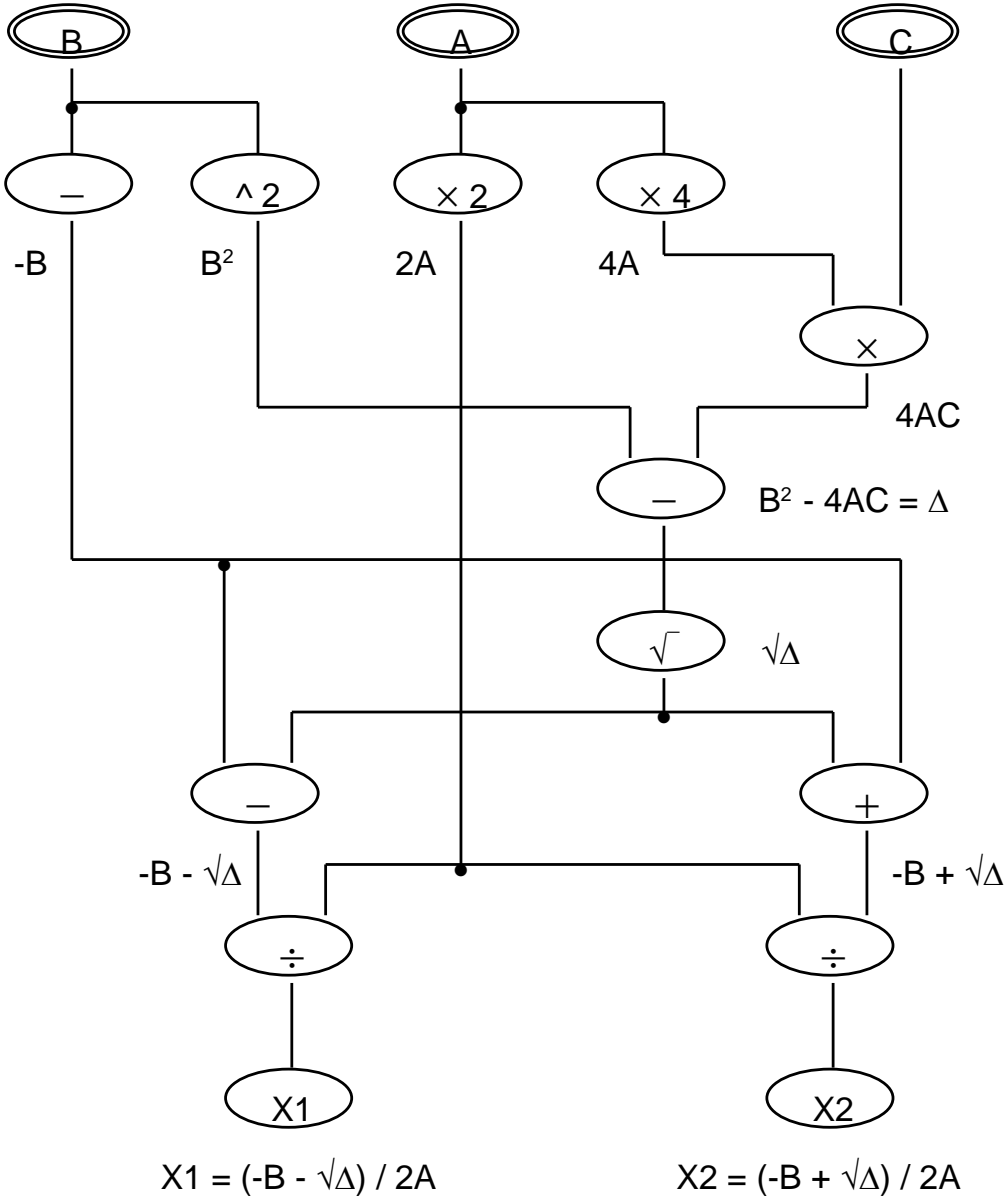
Coarse Grain

Fine Grain

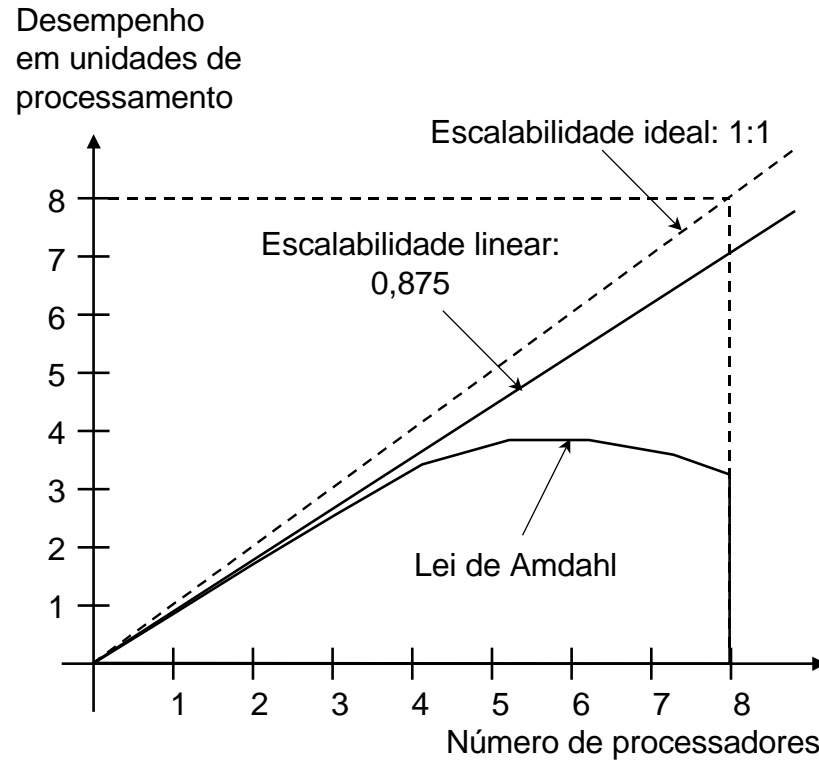
Adaptive scaling



PARADIGMA DATA-FLOW (FLOWCHART)



“LEI” DE AMDAHL



“Lei de Amdahl” (1967): interligar computadores em paralelo para realizar uma tarefa implica em coordenar esta solução cooperativa, gerando um "overhead" de natureza seqüencial. Assim, este "overhead" suplantaria a capacidade agregada, estabelecendo um limite superior ao incremento da capacidade de processamento.