

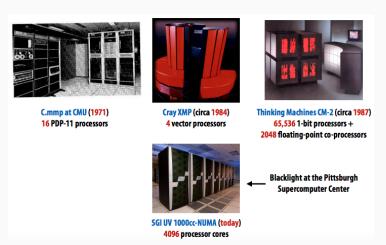
Arquiteturas paralelas – Parte 1

Processamento Paralelo

Prof. Oberlan Romão Departamento de Computação e Eletrônica – DCEL Centro Universitário Norte do Espírito Santo – CEUNES Universidade Federal do Espírito Santo

Uma breve história da computação paralela

 Foco Inicial (a partir dos anos 70): "Supercomputadores" voltados para Computação Científica



Uma breve história da computação paralela

- Foco Inicial (a partir dos anos 70): "Supercomputadores" voltados para Computação Científica
- Outra aplicação de tendência (a partir do início dos anos 90): bancos de dados



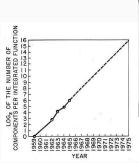
Uma breve história de desempenho do processador

- Aumento do tamanho da palavra do processador
 - 4 bit \rightarrow 8 bit \rightarrow 16 bit \rightarrow 32 bit \rightarrow 64 bit
- Pipeline mais eficiente
- Exploração do paralelismo de nível de instrução (ILP)
- Taxas de clock mais rápidas
 - 10 MHz \rightarrow 200 MHz \rightarrow 3 GHz
- Durante os anos 80 e 90: grandes ganhos de desempenho exponencial
 - e depois...

Uma breve história da computação paralela

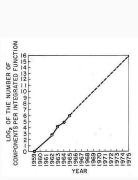
- Foco Inicial (a partir dos anos 70): "Supercomputadores" voltados para Computação Científica
- Outra aplicação de tendência (a partir do início dos anos 90): bancos de dados
- Ponto de inflexão em 2004: Intel atinge a barreira de Densidade de Energia



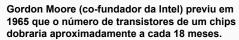


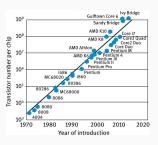


Gordon Moore (co-fundador da Intel) previu em 1965 que o número de transistores de um chips dobraria aproximadamente a cada 18 meses.

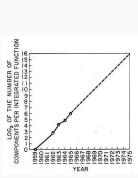






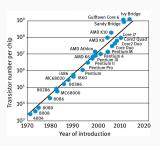


2x transistores/chip a cada 1,5 ano



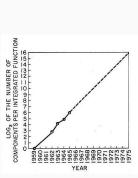


Gordon Moore (co-fundador da Intel) previu em 1965 que o número de transistores de um chips dobraria aproximadamente a cada 18 meses.



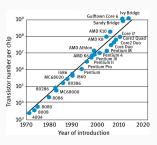
2x transistores/chip a cada 1,5 ano

Chamada de Lei de Moore





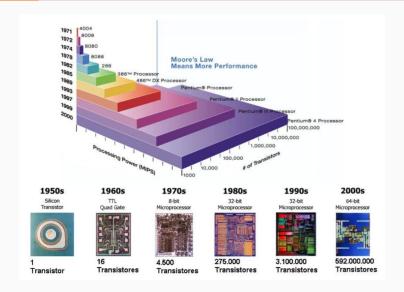
Gordon Moore (co-fundador da Intel) previu em 1965 que o número de transistores de um chips dobraria aproximadamente a cada 18 meses.

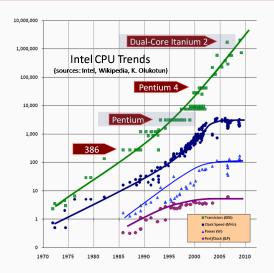


2x transistores/chip a cada 1,5 ano

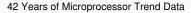
Chamada de Lei de Moore

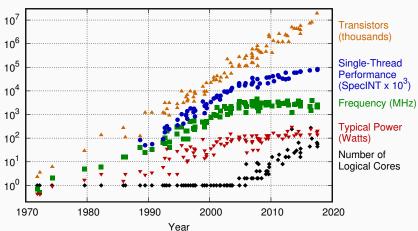
Os processadores tornaram-se menores, mais densos e mais poderosos.





- Densidade do chip continuou a dobrar a cada 1,5 ano
- O clock não





Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten New plot and data collected for 2010-2017 by K. Rupp

Quais as dificuldades em aumentar o clock de processadores de um único núcleo?

• Problemas de temperatura

- Problemas de temperatura
 - Transitores menores ⇒ processadores mais rápidos

- Problemas de temperatura
 - Transitores menores ⇒ processadores mais rápidos
 - Processadores mais rápidos ⇒ aumento do consumo de energia

- Problemas de temperatura
 - Transitores menores ⇒ processadores mais rápidos
 - Processadores mais rápidos ⇒ aumento do consumo de energia
 - Aumento do consumo de energia ⇒ aumento do calor

- Problemas de temperatura
 - Transitores menores ⇒ processadores mais rápidos
 - ullet Processadores mais rápidos \Rightarrow aumento do consumo de energia
 - Aumento do consumo de energia ⇒ aumento do calor
 - Aumento do calor \Rightarrow processadores não confiáveis

- Problemas de temperatura
 - Transitores menores ⇒ processadores mais rápidos
 - Processadores mais rápidos ⇒ aumento do consumo de energia
 - Aumento do consumo de energia ⇒ aumento do calor
 - Aumento do calor ⇒ processadores não confiáveis
- Problemas de velocidade da luz

- Problemas de temperatura
 - Transitores menores ⇒ processadores mais rápidos
 - Processadores mais rápidos ⇒ aumento do consumo de energia
 - Aumento do consumo de energia ⇒ aumento do calor
 - Aumento do calor ⇒ processadores não confiáveis
- Problemas de velocidade da luz
- Limitações de litografia (tamanho dos transistores)

- Problemas de temperatura
 - Transitores menores ⇒ processadores mais rápidos
 - Processadores mais rápidos ⇒ aumento do consumo de energia
 - Aumento do consumo de energia ⇒ aumento do calor
 - Aumento do calor ⇒ processadores não confiáveis
- Problemas de velocidade da luz
- Limitações de litografia (tamanho dos transistores)
- Produção e verificação difíceis

- Problemas de temperatura
 - Transitores menores ⇒ processadores mais rápidos
 - Processadores mais rápidos ⇒ aumento do consumo de energia
 - Aumento do consumo de energia ⇒ aumento do calor
 - Aumento do calor ⇒ processadores não confiáveis
- Problemas de velocidade da luz
- Limitações de litografia (tamanho dos transistores)
- Produção e verificação difíceis

Quais as dificuldades em aumentar o clock de processadores de um único núcleo?

- Problemas de temperatura
 - Transitores menores ⇒ processadores mais rápidos
 - Processadores mais rápidos ⇒ aumento do consumo de energia
 - Aumento do consumo de energia ⇒ aumento do calor
 - Aumento do calor ⇒ processadores não confiáveis
- Problemas de velocidade da luz
- Limitações de litografia (tamanho dos transistores)
- Produção e verificação difíceis

Como continuar aumentando o poder de processamento dos processadores?

Quais as dificuldades em aumentar o clock de processadores de um único núcleo?

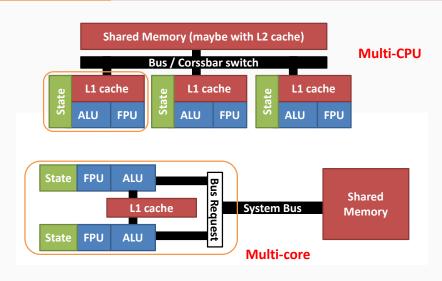
- Problemas de temperatura
 - Transitores menores ⇒ processadores mais rápidos
 - Processadores mais rápidos ⇒ aumento do consumo de energia
 - Aumento do consumo de energia ⇒ aumento do calor
 - Aumento do calor ⇒ processadores não confiáveis
- Problemas de velocidade da luz
- Limitações de litografia (tamanho dos transistores)
- Produção e verificação difíceis

Como continuar aumentando o poder de processamento dos processadores?

• Processadores multicore



Multiprocessadores modernos



- FPU: Floating-point unit
- ALU: Arithmetic logic unit

Arquitetura de von Neumann

- Composta por quatro componentes principais:
 - Memória
 - Unidade de controle
 - Unidade aritmética e lógica
 - Entrada/Saída



- Memória de acesso aleatório é usada para armazenar as instruções e os dados do programa
 - Instruções: dados codificados que dizem ao computador o que fazer
 - Dados: informações a serem usadas pelo programa
- A unidade de controle obtém instruções/dados da memória, decodifica as instruções e, em seguida, coordena sequencialmente as operações para realizar a tarefa programada
- Unidade aritmética e lógica executa operações aritméticas básicas
- Entrada/Saída é a interface para o operador humano

Arquitetura de von Neumann

So what? Who cares?

Arquitetura de von Neumann

So what? Who cares?

- Computadores paralelos ainda seguem este design básico, apenas multiplicado em unidades
- A arquitetura básica e fundamental continua a ser a mesma

O que é um computador paralelo?

Uma definição comum

• Um computador paralelo é uma coleção de elementos de processamento que cooperam para resolver problemas rapidamente

Por que precisamos saber sobre hardware?

- Porque as características da máquina realmente importam
- Porque você se preocupa com a eficiência e o desempenho (você está escrevendo programas paralelos afinal de contas!)

Classificação de sistemas paralelos

- Existem diversas classificações para as arquiteturas paralelas
- Devido a constante evolução, nenhuma classificação consegue abranger todas as arquiteturas existentes

Classificação de sistemas paralelos

Classificação de Flynn (1972)

- "Some computer organizations and their effectiveness", IEEE Transactions on Computers, vol. C-21, pp. 948-960, 1972
- Baseia-se na unicidade e multiplicidade do fluxo de dados e instruções
- Mais conhecida

Classificação de Duncan (1990)

- "A survey of parallel computer architectures", IEEE Computer, pp. 5-16, Fevereiro, 1990
- Classificação mais recente e abrangente
- Menos conhecida

Classificação de sistemas paralelos

Taxonomia de Flynn

- Proposta por Michael J. Flynn em 1972
- Distingue arquiteturas de computadores de acordo a quantidade de instruções e dados processados em um determinado momento
- Uma das metodologias mais conhecidas e utilizadas para classificar a arquitetura de um computador ou conjunto de computadores
- Classifica a arquitetura dos computadores segundo duas dimensões independentes: instruções e dados, em que cada dimensão pode tomar apenas um de dois valores distintos: simples ou múltiplo.

Taxonomia de Flynn

Fluxo de dados

	Único	Múltiplo
Único	SISD Single Instruction Single Data Instrução Simples de Dados Simples	SIMD Single Instruction Multiple Data Instrução Simples de Múltiplos Dados
Múltiplo	MISD Multiple Instruction Single Data Instrução Múltipla de Dados Simples	MIMD <i>Multiple Instruction Multiple Data</i> Instrução Múltipla de Dados Múltiplos

Taxonomia de Flynn

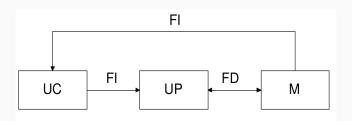
SISD: único fluxo de instrução, único fluxo de dados

- Classe que representa os computadores convencionais (seriais), a arquitetura básica de von Neumann
 - Fluxo Simples de Instruções: Apenas um fluxo de instruções e processado de cada vez
 - Fluxo Simples de Dados: Um único fluxo de dados existe entre processador e memória
 - Execução determinística
- As instruções são executadas serialmente, porém os estágios (busca da instrução, decodificação, busca do operando e execução) podem ser sobrepostos (pipeline)
- Pode-se saber o que está ocorrendo exatamente em cada instante de tempo e reproduzir o processo passo a passo mais tarde
 Single Instruction Multiple In

SIMD

Taxonomia de Flynn

SISD: único fluxo de instrução, único fluxo de dados



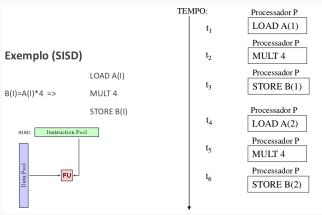
FI - Fluxo de instruções M - Memória

FD - Fluxo de dados UC - Unidade de Controle

UP - Unidade de Processamento

	Single Instruction	Multiple Instruction
Single Data	SISD	MISD
Multiple Data	SIMD	MIMD

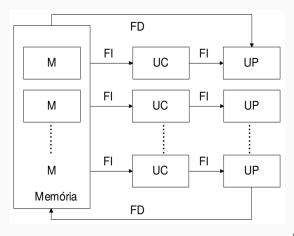
SISD: único fluxo de instrução, único fluxo de dados



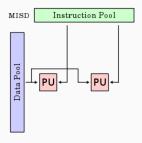
	Single Instruction	Multiple Instruction
Single Data	SISD	MISD
Multiple Data	SIMD	MIMD

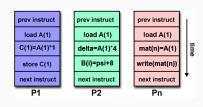
- Vários processadores, onde cada um recebe instruções distintas, mas operam sobre o mesmo conjunto de dados
- Poucos exemplos
 - Múltiplos filtros de frequência operando sobre um único fluxo de sinal
 - Múltiplos algoritmos de criptografia para decodificar uma mensagem





	Single Instruction	Multiple Instruction
Single Data	SISD	MISD
Multiple Data	SIMD	MIMD

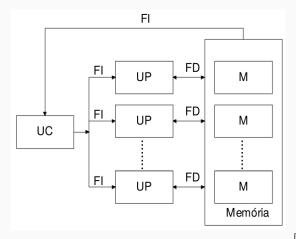




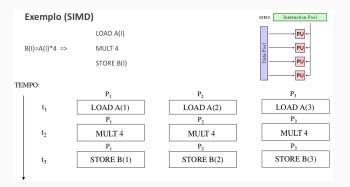
	Single Instruction	Multiple Instruction
Single Data	SISD	MISD
Multiple Data	SIMD	MIMD

- Classe que representa os processadores matriciais, paralelos e associativos
 - Única instrução: Todas as unidades de processamento executam a mesma instrução
 - Vários dados: Cada unidade de processamento pode operar em um elemento de dados diferentes
- Uma única unidade de controle que envia um fluxo de instruções para vários processadores
- Os processadores recebem a mesma instrução ao mesmo tempo e atuam sobre diferentes fluxos de dados
- GPUs são normalmente implementações SIMD

	Single Instruction	Multiple Instruction
Single Data	SISD	MISD
Multiple Data	SIMD	MIMD



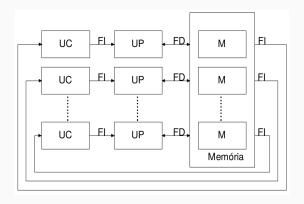
	Single Instruction	Multiple Instruction
Single Data	SISD	MISD
Multiple Data	SIMD	MIMD



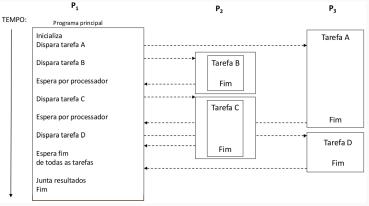
	Single Instruction	Multiple Instruction
Single Data	SISD	MISD
Multiple Data	SIMD	MIMD

- Cada processador pode executar um fluxo de instruções diferentes
- Cada processador pode trabalhar com diferentes fluxos de dados
- Podem ser síncronos ou assíncronos
- Atualmente o tipo mais comum de computador paralelo

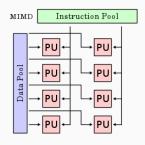
	Single Instruction	Multiple Instruction
Single Data	SISD	MISD
Multiple Data	SIMD	MIMD

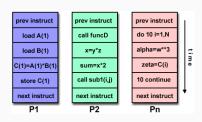


	Single Instruction	Multiple Instruction
Single Data	SISD	MISD
Multiple Data	SIMD	MIMD

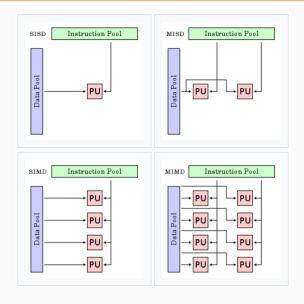


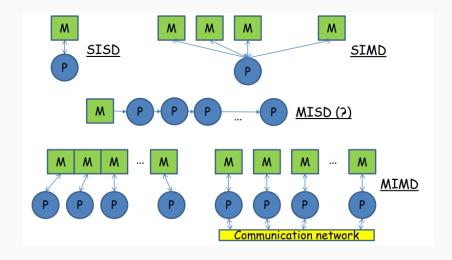
	Single Instruction	Multiple Instruction
Single Data	SISD	MISD
Multiple Data	SIMD	MIMD



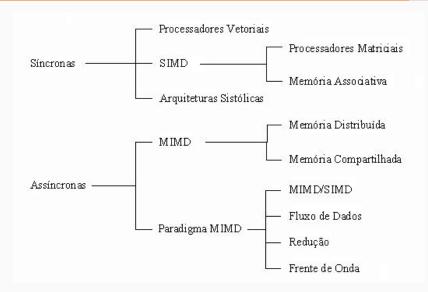


	Single Instruction	Multiple Instruction
Single Data	SISD	MISD
Multiple Data	SIMD	MIMD





Taxonomia de Duncan



Taxonomia de Duncan

Arquiteturas Síncronas

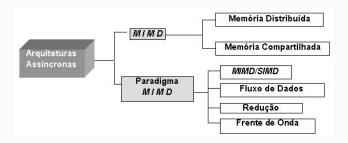
- As operações são coordenadas através de um relógio global, dessa forma os processadores permanecem síncronos
- Pouca flexibilidade para desenvolvimento de algoritmos paralelos



Taxonomia de Duncan

Arquiteturas Assíncronas

- Caracterizam-se pelo controle descentralizado de hardware, de maneira que os processadores sejam independentes entre si
- Formada basicamente pelas arquiteturas MIMD, sejam convencionais ou não



Para pensar...

- 1. Por que é tão difícil/complicado construir processadores com transistores menores?
- 2. "Se um único processador consegue resolver um problema em n segundos, n processadores conseguem resolver o mesmo problema em 1 segundo?"

Referências

- https://en.wikipedia.org/wiki/Flynn%27s_taxonomy
- De-Mystifying Software Performance Optimization https://software.intel.com/en-us/articles/ de-mystifying-software-performance-optimization
- Introduction to Parallel Computing
 https://computing.llnl.gov/tutorials/parallel_comp/
- Material do Prof. Fernando Santos Osório
 http://conteudo.icmc.usp.br/pessoas/fosorio/fosorio-usp.html