

# Commercial Applications in High Performance Computing

## PEE-5793 Computação de Alto Desempenho para Aplicações Comerciais

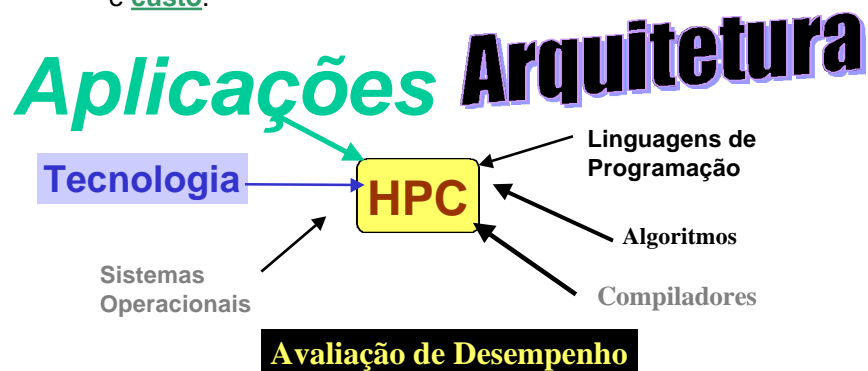
Prof. Dr. Edward David Moreno  
Prof. Dr. Sergio Takeo Kofuji

1

## Computação de Alto Desempenho e Arquiteturas

Sob o ponto de Vista de Projeto (Arquitetura de Comp.):

Projetar e caracterizar os diferentes níveis de sistemas computacionais visando obter bom desempenho e facilidade de programação, sempre considerando os limitantes de tecnologia e custo.



2

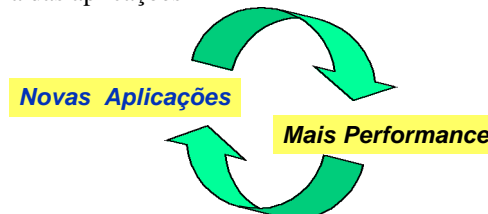
## Tendências em “HPC” (Paralela)

- **Demanda das Aplicações:** Há um sempre desejo de mais e mais computação
  - *Computação científica:* Todas as áreas (CFD, Biologia, Química, Física, ...)
  - *Computação de Propósito-Geral:* Vídeo, Gráfica, CAD, Databases, TP...
- **Tendências Tecnológicas**
  - Maior número de transistores on-chip
  - Elementos de Projeto cada vez mais rápidos: CPUs, DRAMs, Disks, etc.
- **Tendências Arquiteturais**
  - ILP crescente.
  - Aproveitar paralelismo de granularidade maior (não fina).
- **Econômico**
- **Tendências Recentes:**
  - Os microprocessadores têm suporte para multiprocessador
  - Servidores e Workstations estão cada vez mais próximos de MPPs.
  - Os microprocessadores futuros serão multiprocessadores.

3

## Tendências das Aplicações

- As aplicações requerem melhoramentos de performance e de hardware, o que habilita e estimula a criação de novas aplicações, e.....
  - Esse ciclo tem possibilitado um aumento exponencial na performance de CPU
  - Melhorar arquiteturas de alto desempenho é um pouco mais complexo: para a maioria das aplicações



- A necessidade de performance em certos valores, pode ocasionar um aumento no custo.

4

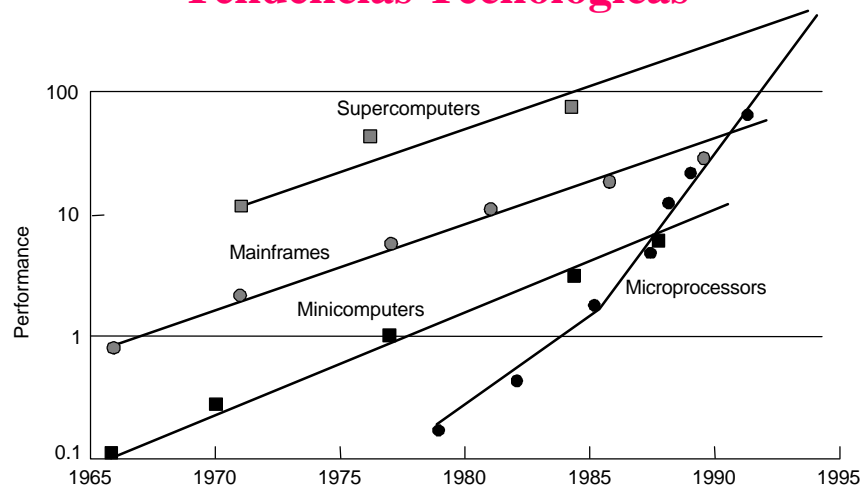
## Tendências das Aplicações (2)

- Nos últimos anos as transições e análise de alto desempenho têm sido focalizados para aplicações científicas e de engenharia.
- Grande avanço em computação comercial
  - Banco de dados e transações, área financeira.
  - Geralmente pequena escala, mas grandes sistemas são também usados.
- Desktop também usam programas multithreaded, os quais são similares aos programas paralelos.
- Há uma grande demanda por melhorar a performance em workloads sequenciais.
  - Isso motiva a utilização de multiprocessadores de pequena escala.

**Há uma demanda real e crescente de aplicações de alto desempenho: Aplicações Comerciais.**

5

## Tendências Tecnológicas



Há uma grande chance de ter multiprocessadores on-chip, pelo menos já há vários processadores on-board.  
Ex. Quad Pentium Pro, os quais são rápidos.

6

## Tendências Arquiteturais


- Arquitetura transforma os avanços tecnológicos em desempenho e capacidade.
- Soluciona a interrelação entre paralelismo e localidade.
  - Uproc. Atuais: 1/3 execução real, 1/3 cache, 1/3 comunicação externa.
  - Essas correlações podem mudar com a escala e os avanços tecnológicos.
- Há um maior conhecimento das tendências arquiteturais de Uproc.
  - Ajuda a estabelecer critérios em relação a questões de projeto de máquinas de alto desempenho.
  - Mostra a importância do paralelismo ainda para máquinas “seqüenciais”.
- Um grande impacto da tecnologia VLSI, tanto que o paralelismo explorado está grandemente relacionado com os projetos VLSI. ILP e VLSI !!
  - pipelining e simples ISA, + avanços de compilador
  - caches on-chip e unidades funcionais => execução superscalar
  - Maior complexidade: Execução out of order, especulativa e predição.
    - Confiar com transferência de controle e problemas de latência.

### PARALELISMO AO NÍVEL DE THREADS

7

## Tendências Arquiteturais: MPs baseados em Barramentos

- Uproc. num chip torna fácil e natural o modelo de memória compartilhada
  - domina o mercado de servidores e das companhias, após os desktops
- Uproc. Rápidos saturam o bus, então houve avanços na tecnologia de bus, HOJE
  - Uma ampla faixa de sistemas baseados em bus, desktops até grandes servidores.

 **Econômico**

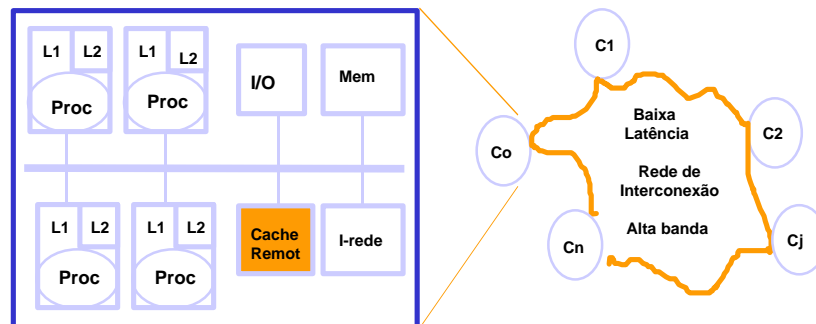
- Uproc. Padronizados são mais rápidos e muito mais baratos.
- Custo de projeto de MPPs gira em torno de dezenas de milhões de dólares (5-100).

**Há uma possível padronização para utilizar multiprocessadores baseados em nós de SMPs com vários processadores conectados através de um barramento.**

8

## Arquiteturas Escaláveis

- Grandemente atrativo: Econômico, Tecnologia, Arquitetura, Aplicações.
- Paralelismo explorado em diferentes níveis: ILP, Servidores multiprocessadores, Multiprocessadores de grande escala.
- O mesmo acontece para o sistema de memória: BW ↑ Latência. ↓
- Arquiteturas distintas ⇒ Custo, Desempenho e Escalabilidade também distintos



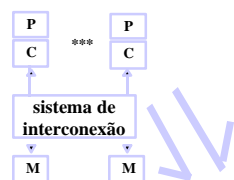
9

## ARQUITETURAS Típicas

Clusters ??  
Redes de Workstations

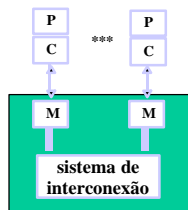
Hoje, as três vertentes mais interessantes na construção de multiprocessadores de grande-escala são:

### CC-NUMA - SMPs



DASH  
NUMAChine  
MIT Alewife  
Sting  
NUMA-Q  
SGI Origin

### DSMs



FLASH  
Typhoon  
Shrimp  
NCP2

### COMA



### Comparação:

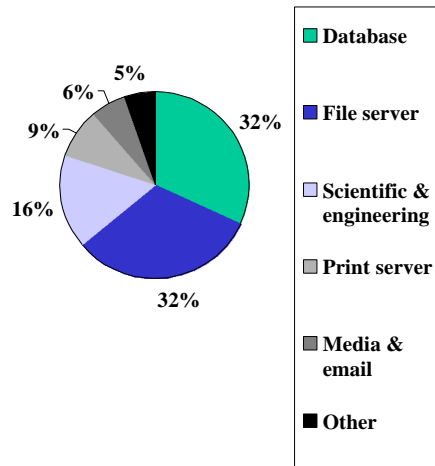
- Sistema de Memória mais flexível
- Em aplicações dominadas por falhas de coerência
- CC-NUMA e DSM obtêm melhor desempenho

10

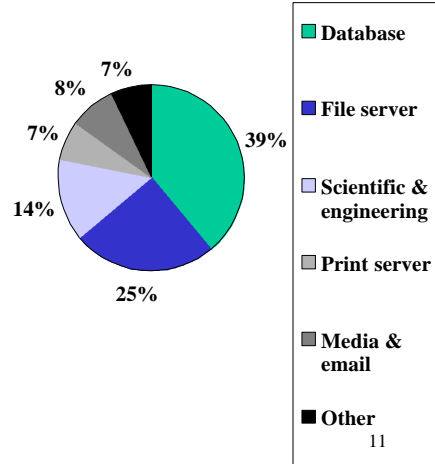
## Motivação: Tendências no Volume do Mercado

Fonte: Stenstrom, et al., *IEEE Computer*, December 1997

**Ano 1995: Servidores  
Volume**

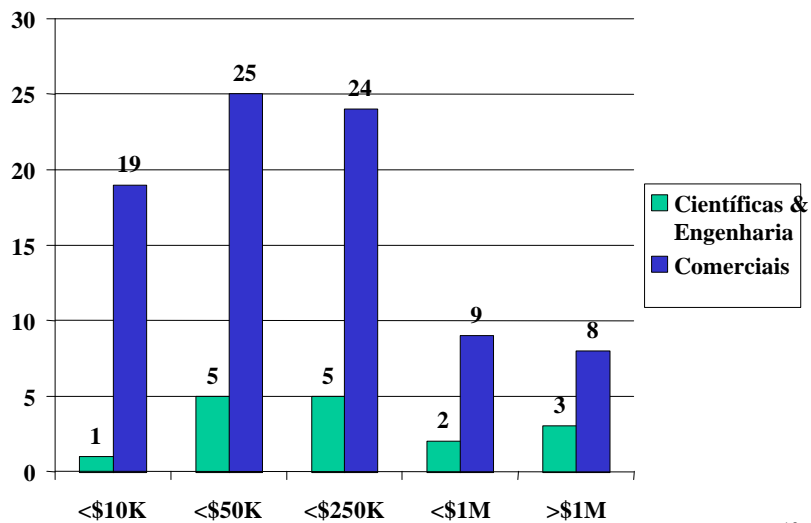


**Ano 2000 Servidores  
Volume**

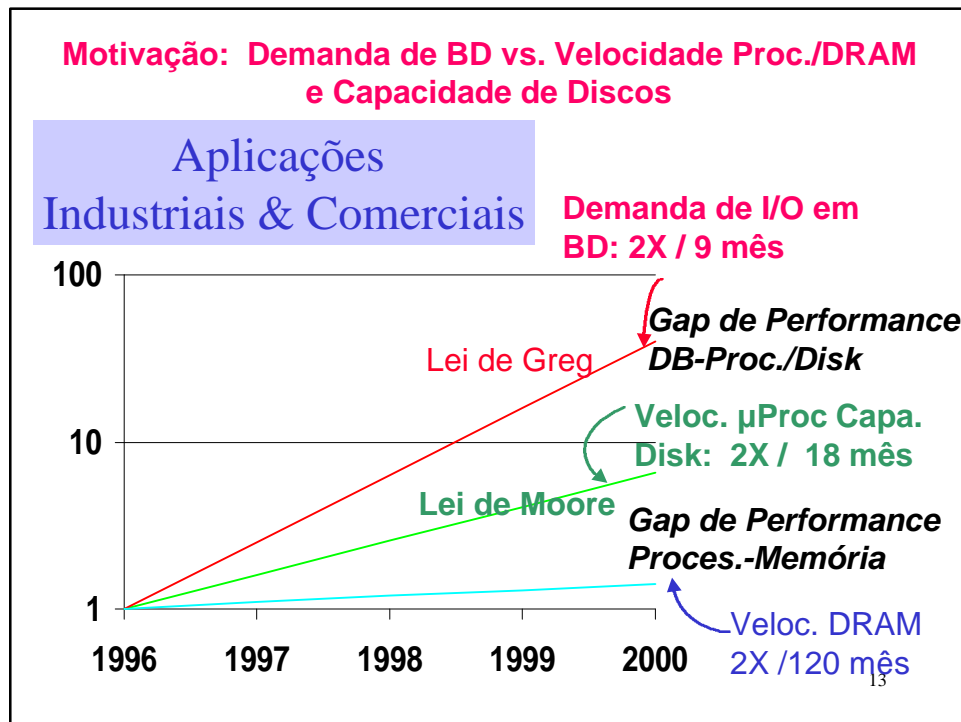


## Ano 1995: Volume de Mercado vs. Custo de Máquina

Fonte: Stenstrom, et al., *IEEE Computer*, December 1997



12



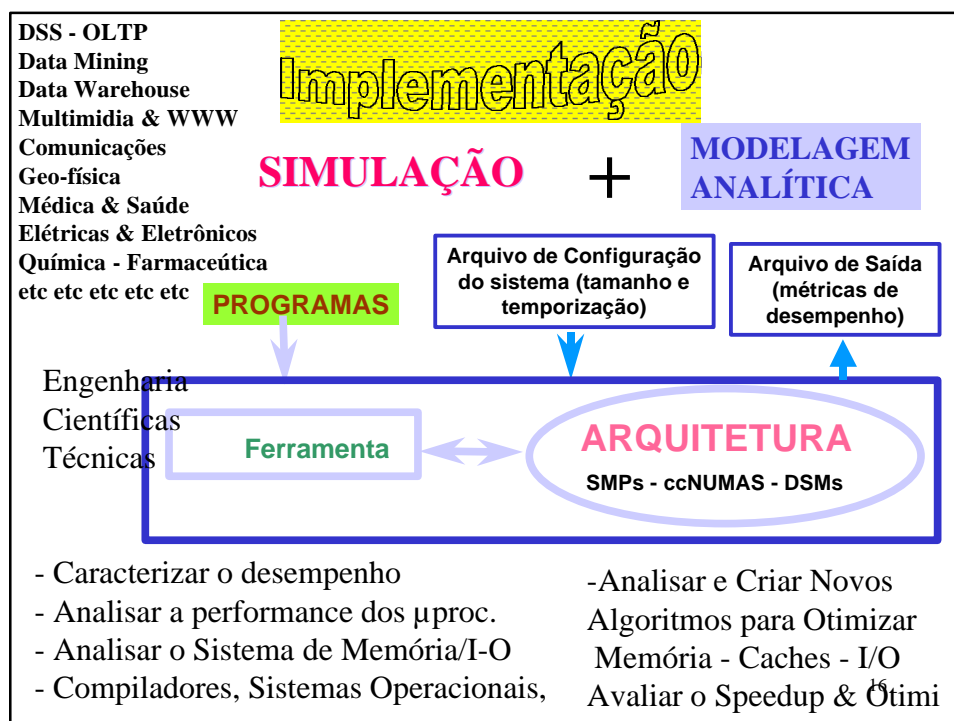
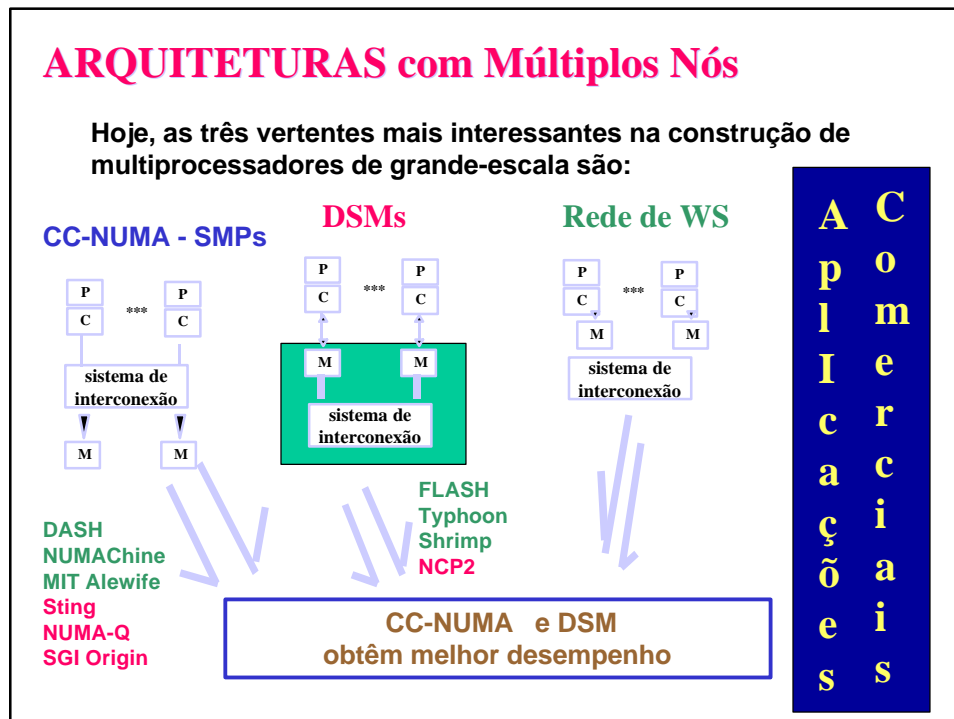
## Limitações Processador: TPC-C

Pentium Pro	SPEC-int95	TPC-C
– Caches Multi-nível: Miss rate (cache L2=1MB)	0.5%	5%
– Superescalar (2-3 instr. requisitadas/clock): % clks	40%	10%
– Speedup de execução “Out-of-Order”	2.0X	1.4X
– Clocks por Instrução (CPI)	0.8	3.4
• % Máxima Performance	40%	10%

Fonte:

- Kim Keeton, Dave Patterson, Y. Q. He, R. C. Raphael, and Walter Baker. "Performance Characterization of a Quad Pentium Pro SMP Using OLTP Workloads," *Proc. 25th Int'l. Symp. on Computer Architecture*, June 1998. ([www.cs.berkeley.edu/~kkeeton/Papers/](http://www.cs.berkeley.edu/~kkeeton/Papers/)).
- Bhandarkar, D.; Ding, J. "Performance characterization of the Pentium Pro processor." *Proc. 3rd Int'l. Symp. on High-Performance Computer Architecture*, Feb 1997. p. 288-97.

14





## Atividades Iniciais

*OLTPs DM DWH*

*Comerciais*

**Objetivo:** Caracterização, Análise, Otimizações e Projeto de Sistemas de Memória em Máquinas de Alto Desempenho.

**Sistemas:** (i) CC-NUMAs (ii) SMPs, (iii) DSMs (iv) Rede WS

### Aspectos de Performance e Projeto:

- Protocolos de Coerência de Cache (invalidação vs. atualização, HW vs SW, híbridos).
- Técnicas de tolerância a latência (prefetching, multithreading, modelos de consistência não sequencial, etc.)
- Otimizações baseadas em Compilador.
- Otimizações baseadas em Algoritmos.
- **Interrelação com o Sistema Operacional !!**

Outros: Identificação de gargalos, visualização de desempenho, Novas técnicas de Simulação e Validação, Modelagem Analítica.

## Resultados Esperados

### Acadêmico:

- Projetos Pesquisa
- Dissertações de Mestrado
- Teses de Doutorado
- Livro

### Industrial:

- Projetos conjuntos com grupos não-acadêmicos

**Resultado:** **Conhecimento** + **Financeiro**

- O paralelismo pode ser difícil de obter.
- Pouco ou moderado paralelismo é importante na escalabilidade
- Difícil estabelecer comparações entre diferentes plataformas.

18

# **CRONOGRAMA**

**Ver Tabela de Programa na Pagina WEB**

**<http://www.lsi.usp.br/~edmoreno>**

19

## **Avaliação**

Realizar e expor um projeto.

- Título
- Objetivos
- Introdução
- Motivação
- Metodologia
- Trabalhos Correlatos
- Resultados e Análise
- Conclusões
- Referências

- Na forma de paper

O mais completo que for possível

- Um estudo de avaliação de desempenho

- Uma análise crítica e comparativa de trabalhos já existentes e modernos

**Entregar integrantes e possível projeto: Outubro/04**

**Entregar um rascunho: Outubro/25**

**Motivação, Introdução, Objetivos e Referências**

**Apresentações de Projetos: Dezembro 06&13/99**

20

