

Sistemas Embarcados Multiprocessados

Exame de Qualificação:
Mateus Beck Rutzig

mbrutzig@inf.ufrgs.br

Orientador: Luigi Carro

carro@inf.ufrgs.br



Laboratório de Sistemas Embarcados
Instituto de Informática
Universidade Federal do Rio Grande do Sul

Introdução

▶ Dispositivos embarcados dos anos 90



▶ Processamento de Vídeo



- ▶ Video Conferência
- ▶ Processamento de Vídeo
- ▶ Reconhecimento de Imagens



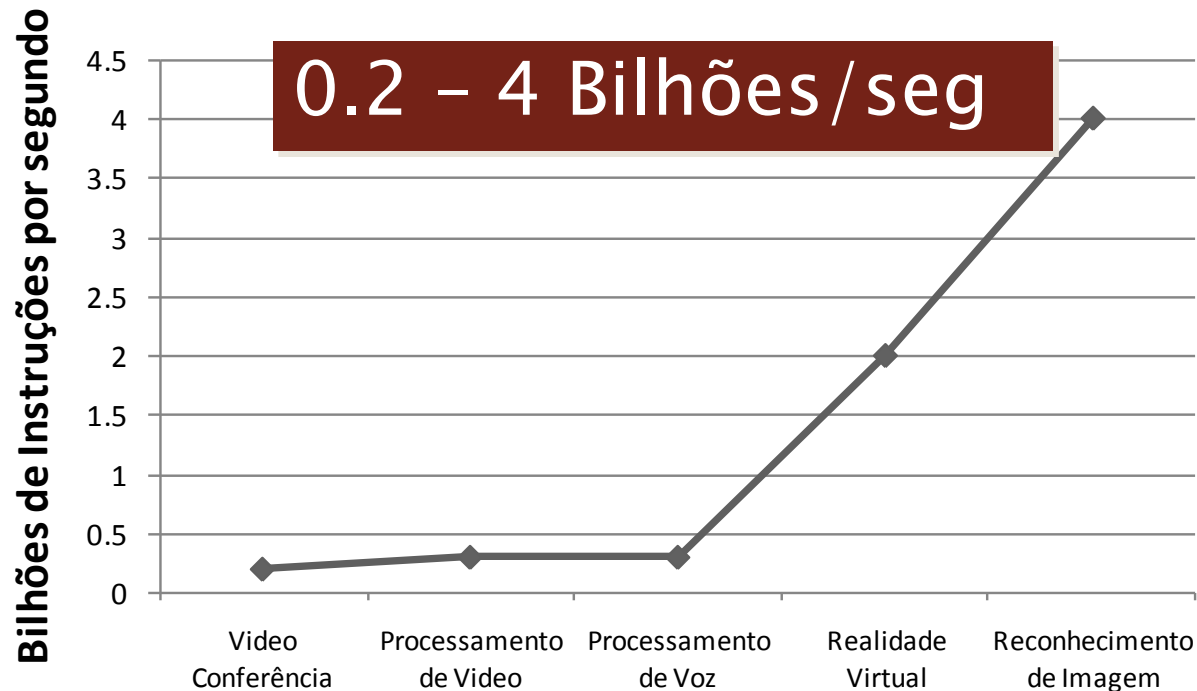
▶ Processamento de voz



▶ Realidade Virtual

Introdução

- Requisitos de desempenho para CADA dispositivo



Processadores com diferentes capacidades de processamento

Introdução

▶ Dispositivos atuais



▶ Processamento de Vídeo



▶ Video Conferência
▶ Processamento de Vídeo
▶ Reconhecimento de Imagens



▶ Processamento de voz



▶ Realidade Virtual

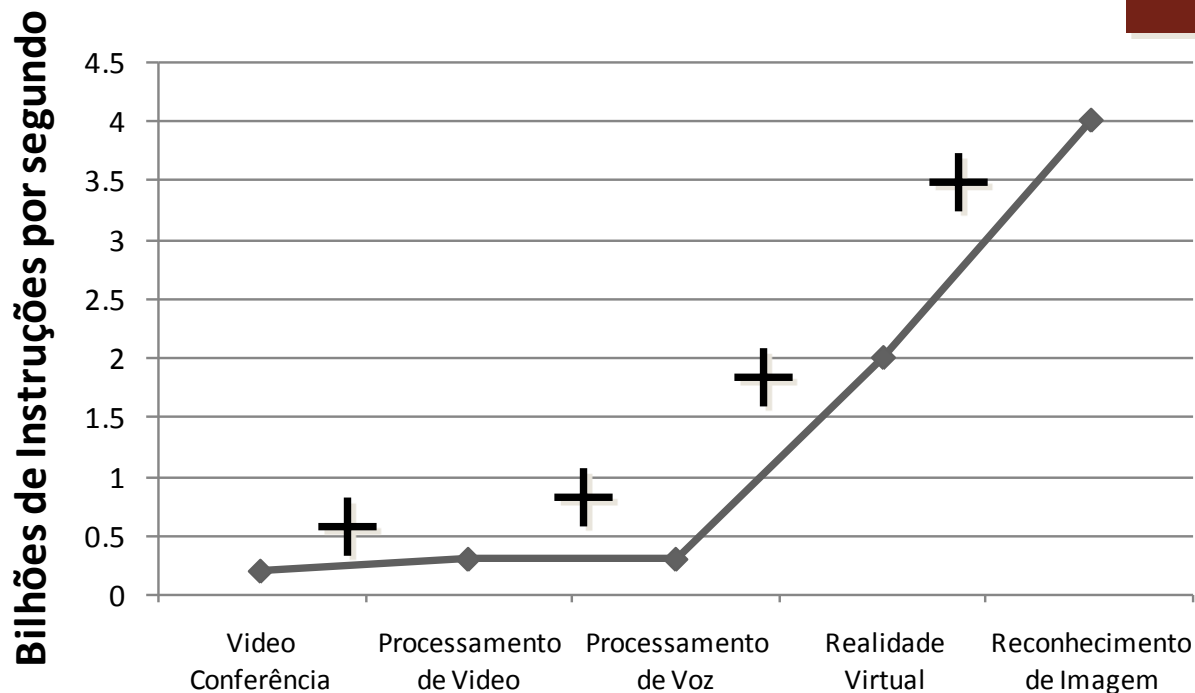
Aplicações são as mesmas dos anos 90...

...mas agora convergiram para um único dispositivos

Introdução

- Requisitos de desempenho para os dispositivos atuais

= ~ 6.5 bi/sec



Introdução

- ▶ A busca pelo desempenho caminha em paralelo com rígidas restrições de projeto que manipulam orçamentos de...

▶ Potência

▶ Área

▶ Energia



Redução do Time-to-Market

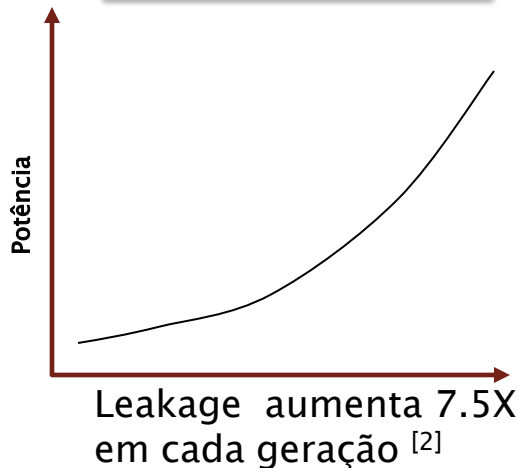
Introdução

- ▶ Possíveis soluções para alcançar os requisitos de desempenho, energia, potência e “time-to-market”
 - Um processador com infinitos recursos para explorar ILP
 - Multicore homogêneo
 - Multicore heterogêneo

Introdução

- ▶ Possíveis soluções para alcançar os requisitos de desempenho, energia, potência e “time-to-market”
 - Um processador com infinitos recursos para explorar ILP

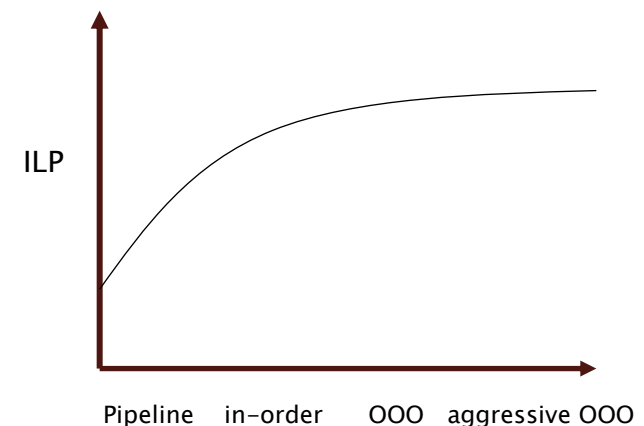
Parede da Potência



Parede da Frequência

Não há muito espaço para diminuir o tempo de estágio do pipeline [3]

Single Thread Performance



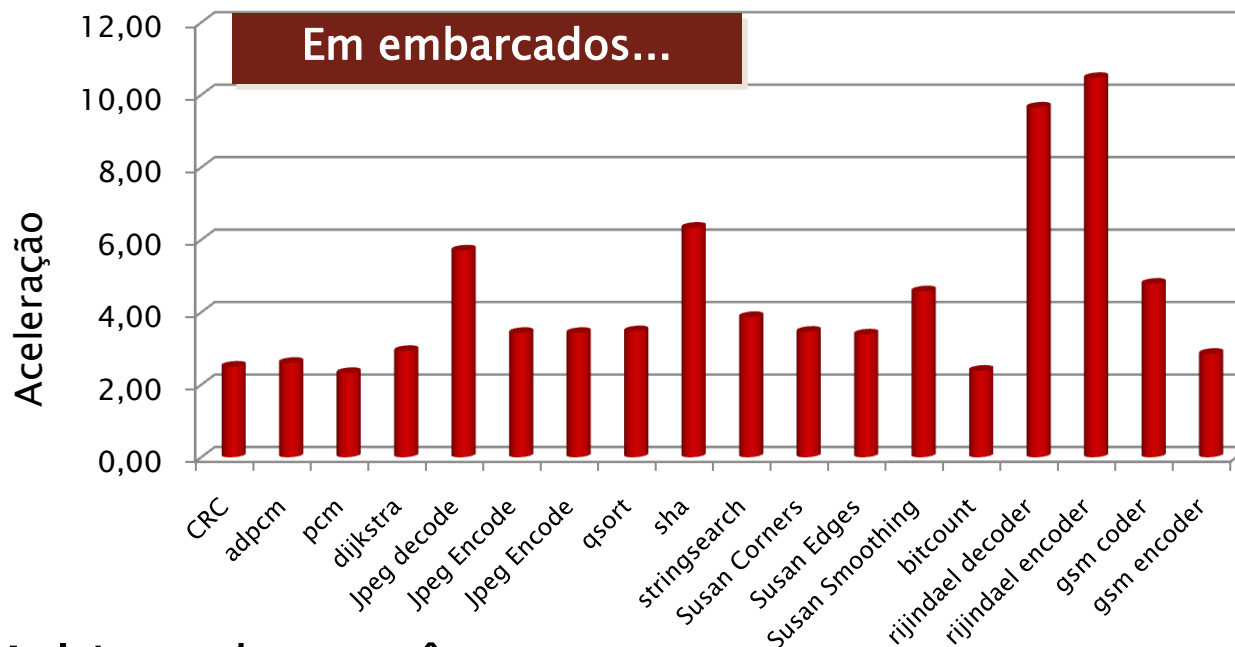
- Multicore homogêneo
- Multicore heterogêneo

[2] S. Borkar “Design Challenges of Technology Scaling” Micro 1999

[3] Vikas Agarwal, M. S. Hrishikesh, Stephen W. Keckler, Doug Burger. Clock rate versus IPC: the end of the road for conventional microarchitectures. In ISCA 2000

Introdução

- ▶ Possíveis soluções para alcançar os requisitos de desempenho, energia, potência e “time-to-market”
 - Um processador com infinitos recursos para explorar ILP



- Multicore homogêneo
- Multicore heterogêneo

...o desempenho é satisfatório para aplicações heterogêneas

Introdução

- ▶ Possíveis soluções para alcançar os requisitos de desempenho, energia, potência e “time-to-market”
 - Um processador com infinitos recursos para explorar ILP



- Multicore homogêneo
- Multicore heterogêneo

...o compromisso entre desempenho e energia não é satisfatório

Introdução

- ▶ Possíveis soluções para alcançar os requisitos de desempenho, energia, potência e “time-to-market”
 - Um processador com infinitos recursos para explorar ILP ✖
 - Multi-core homogêneo ?
 - Multi-core heterogêneo ?

Introdução

- ▶ Possíveis soluções para alcançar os requisitos de desempenho, energia, potência e “time-to-market”
 - Um processador com infinitos recursos para explorar ILP ✖
 - Multicore homogêneo ?
 - Multicore heterogêneo ?
 - **Vantagens**
 - Tempo de projeto reduzido
 - Aumento do “Yield”
 - Escalabilidade
 - **Desvantagens**
 - Mapeamento das aplicações nos processadores
 - Re-escrita do código
 - As aplicações nem sempre oferecem paralelismo neste grão

Sumário

1. Paralelismo de Grão Grosso
2. Arquiteturas Multiprocessadas

Sumário

1. **Paralelismo de Grão Grosso**
2. Arquiteturas Multiprocessadas

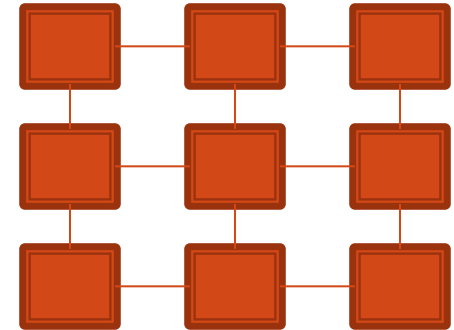
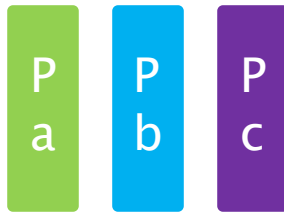
Paralelismo de Grão Grosso

- ▶ Níveis de exploração
 - Processos
 - Threads

Paralelismo de Grão Grosso

► Níveis de exploração

- Processos



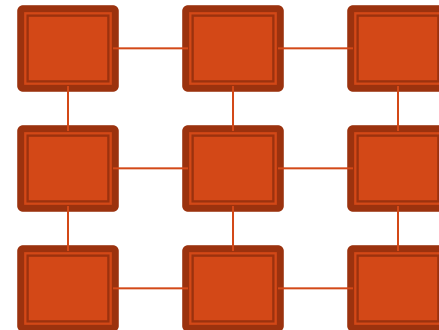
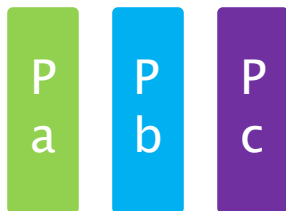
Paralelismo extraído automaticamente pelo Sistema Operacional

- Threads

Paralelismo de Grão Grosso

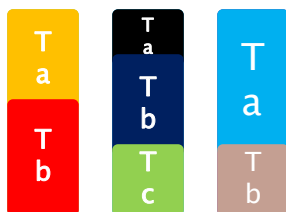
► Níveis de exploração

◦ Processos



Paralelismo extraído automaticamente pelo Sistema Operacional

◦ Threads



Projetista do software é responsável pela extração do paralelismo

Sistema Operacional mapeia as threads nos processadores

Paralelismo de Grão Grosso

- ▶ Níveis de exploração
 - Processos
 - ▶ API (Application Programming Interface)
 - ▶ MPI (Message Passing Interface)
 - ▶ Comunicação baseada em troca de mensagens
 - ▶ Mais utilizada para comunicação em clusters
 - Threads
 - ▶ OpenMP (Open Multi-Processing)
 - ▶ Comunicação baseada em memória compartilhada
 - ▶ Mais utilizada para comunicação intra-chip

Projetista do software é responsável pela extração do paralelismo

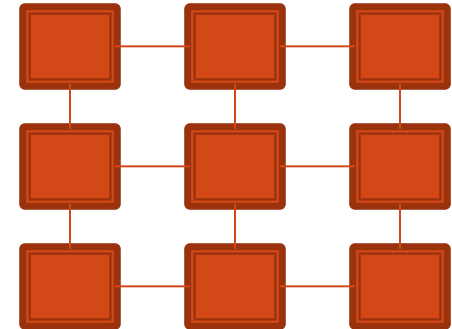
Sistema Operacional mapeia as threads nos processadores

Paralelismo de Grão Grosso

► Exemplos OpenMP

```
#define N 4
int main(int argc, char **argv)
{
    int i, a[N];
    #pragma omp parallel for
    for (i = 0; i < N; i++)
        a[i] = 2 * i;
    return 0;
}
```

T
a
T
b
T
c
T
d



Nenhuma dependência de dados entre as iterações

Paralelismo de Grão Grosso

► Exemplos OpenMP

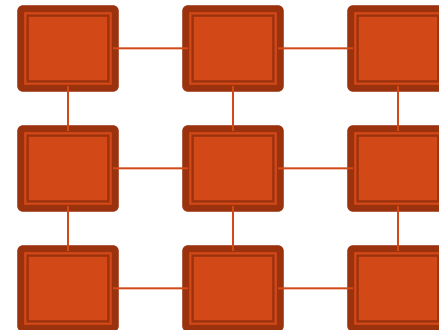
```
#define N 4
int main(int argc, char **argv)
{
    int i, a[N];
    #pragma omp parallel for
    for (i = 0; i < N; i++)
        a[i] = 2 * i;
    return 0;
}
```

Nem todo o software é capaz de explorar uma abordagem multiprocessada

```
#define N 4
int main(int argc, char **argv)
{
    int i, a[N];
    #pragma omp parallel for
    for (i = 0; i < N; i++)
        a[i] = 2 * a[i-1];
    return 0;
}
```

T
a

Existe dependência de dados entre as iterações



Sumário

1. Paralelismo de Grão Grosso
2. **Arquiteturas Multiprocessadas**

Sumário

1. Paralelismo de Grão Grosso
2. **Arquiteturas Multiprocessadas**
 - **Propósito Geral**
 - Embarcados

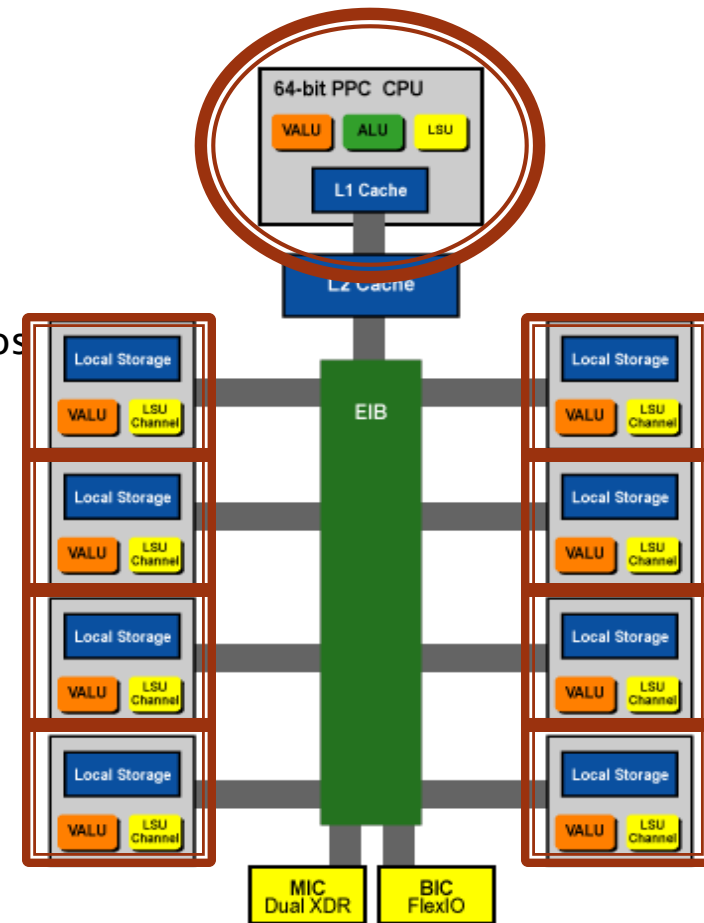
Arquiteturas Multiprocessadas

► Cell Processor (Playstation 3 Console)

- Processador central de alto desempenho
 - Power Processing Element (PPE)
 - IBM PowerPC
 - Multithreaded (2 Threads)
 - Cache L1 (32kb I e 32 Kb D)
 - Cache L2– 512 Kb
 - Frequência: 3.2 GHz
 - Atua como controlador dos processadores periféricos
- Oito processadores periféricos
 - Synergistic Processing Elements (SPE)
 - Baseados em Processamento Vetorial (SIMD)
 - 256 Kb Local Storage
 - Controlados por software

Dificuldades para codificar software

Desenvolvedores de SW são responsáveis pelo gerenciamento dos dados das Local Storages

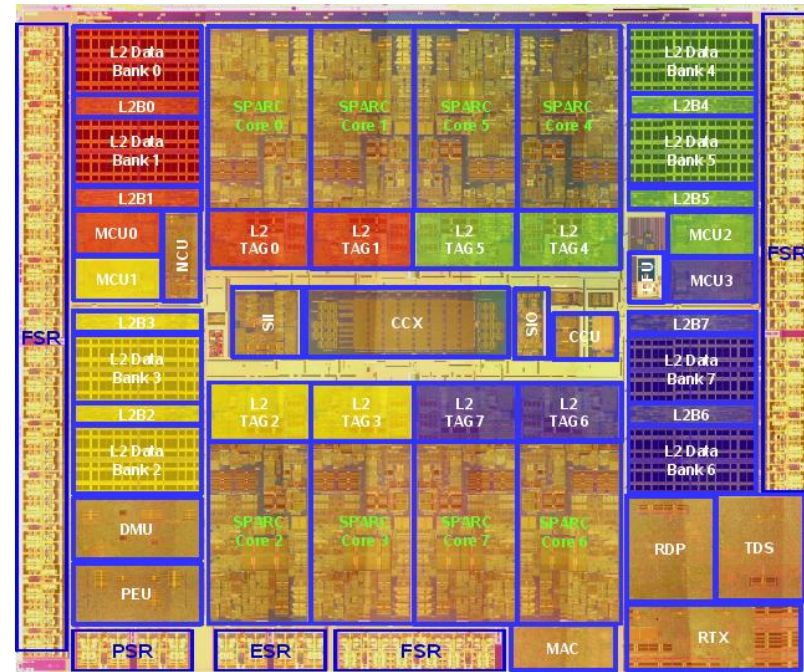


The CELL Architecture

Arquiteturas Multiprocessadas

► UltraSparc T2 (Niagara 2)

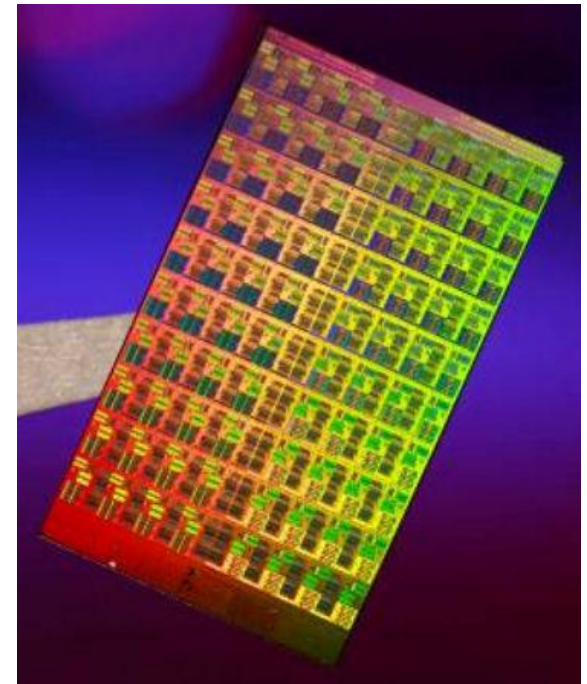
- 8 processadores baseados no instruction set Sparc-V9
 - Cada processador é capaz de executar 8 threads concorrentemente (total de 64 threads)
 - “Server on a Chip”
 - PCI express
 - Duas portas 10 Gigabit Ethernet
 - 4 controladores dual-channel FBDIMM
- Pipeline de 8 estágios (1.6Ghz)
- Cache L2 4MB
 - 8 Bancos
 - Associativa por conjunto 16
- Aplicações
 - WebServers
 - Database



Arquiteturas Multiprocessadas

► Intel 80 Cores

- 1 trilhão de operações de ponto flutuante por segundo
- Consumo de potência: 62 Watts
- Baseado em uma arquitetura VLIW
 - Exploração do ILP via compilador
 - Término da Era x86
 - Quebra da compatibilidade de SW
 - Core2Duo
 - 291 milhões de transistores
 - Intel 80 Cores
 - 1000 milhões de transistores
- Problema de interconexão!
 - Network mesh-based



Simplicidade dos Cores:
Remoção do HW superescalar

Como codificar software com este tipo de interconexão?

Arquiteturas Multiprocessadas

► Intel Larrabee

- Arquitetura CPU–GPU híbrida
- Replicações da arquitetura P54C
 - Pentium 75–120 MHz (1995)
 - Processador Superescalar In–Order
- Poderosa execução de processamento vetorial
 - Melhorias na tecnologia SSE
- Extensões Larrabee no instruction set
 - Voltadas para o processamento gráfico
- Grande largura de banda com a memória

Arquiteturas Multiprocessadas

► NVIDIA Tesla D870

- Contém 2 placas Tesla C870
 - 3 GB DDR3
 - 256 streaming processors
 - Consumo máximo 170 Watts
- GPGPU (*General-Purpose Graphics Processing Unit ou Unidade de Processamento Gráfico de Uso Geral*)
 - 1 trilhão de operações de ponto flutuante por segundo
- Suporte a tecnologia Compute Unified Device Architecture (CUDA)
 - Extensões no Compilador C/C++



Arquiteturas Multiprocessadas

- **Propósito Geral**

- Utilizam arquitetura homogênea
 - Facilidade no gerenciamento de distribuição das threads
 - Tempo curto de validação
- Utilizam poucas técnicas de redução de consumo de potência
 - DFVS (Dynamic Frequency and Voltage Scaling)
- Tendência (já acontecendo!)
 - Replicação de cores simples para explorar paralelismo de grão grosso

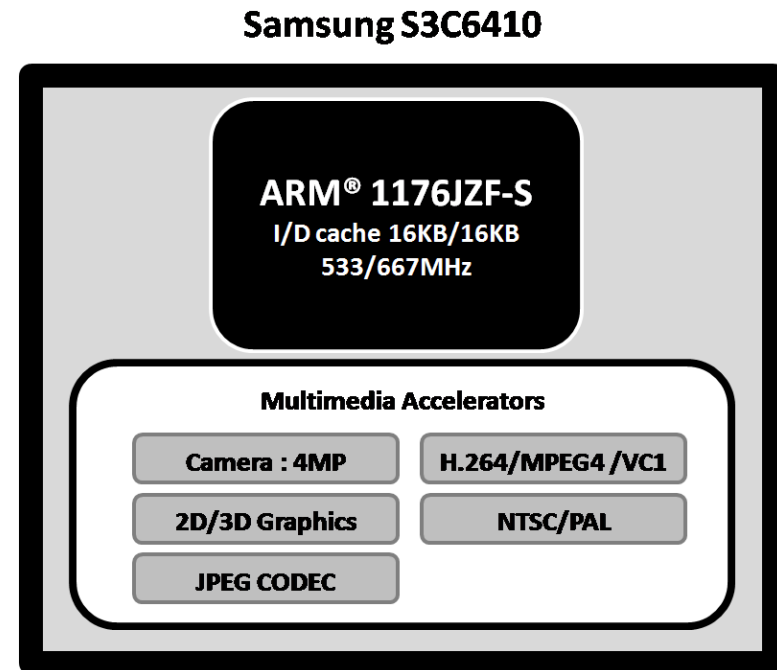
Sumário

1. Paralelismo de Grão Grosso
2. **Arquiteturas Multiprocessadas**
 - Propósito Geral
 - **Embarcados**
3. Tendências para o Futuro Embarcado

Arquiteturas Multiprocessadas

▶ Samsung S3C6410 (iPhone 3G)

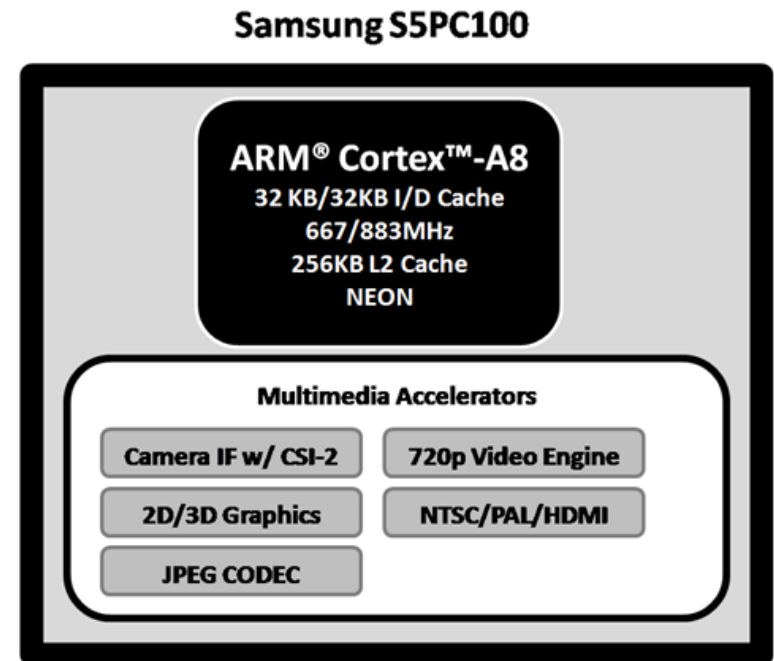
- ARM 1176JZF-S processador de propósito geral (667 MHz)
 - Pipeline de 9 estágios (ARMv6)
 - Processamento vetorial (SIMD)
 - Tecnologia Jazelle – Execução Java Nativa
 - Thumb Instruction Set– compactação de instruções
 - Extensões DSP
- Aceleradores multimedia
 - Formatos: MPEG4, VCI, H.264
- 128 MB RAM
- Gerenciamento de energia
 - ARM Intelligent Energy Management
 - ▶ Desligamento de partes do hardware não utilizadas
 - ▶ Potência Estática



Arquiteturas Multiprocessadas

► Samsung S5PC100 (iPhone 3G S)

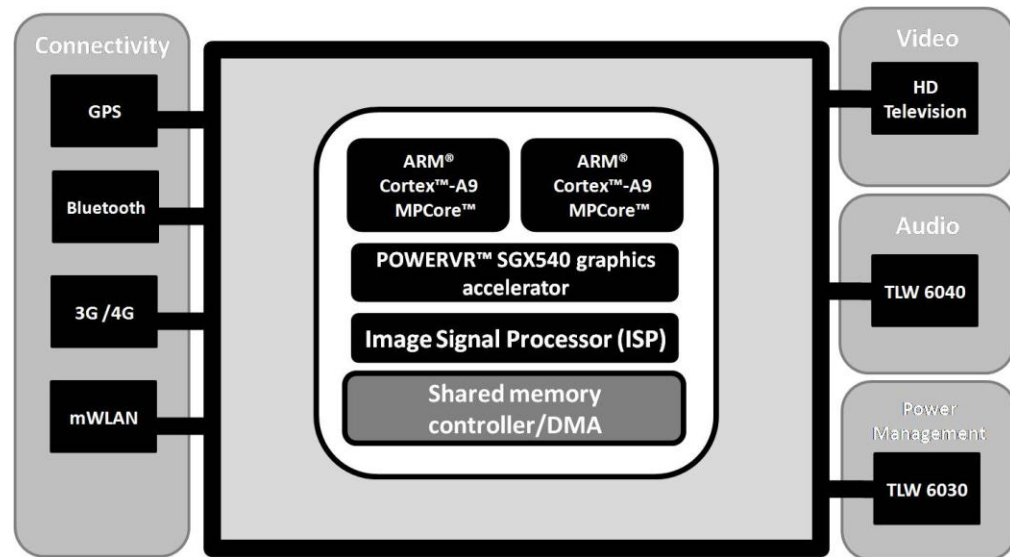
- ARM Cortex A8 Processador de Propósito Geral (600MHz)
 - Pipeline 13 estágios superescalar (ARMv7) com preditor de saltos dinâmico
 - Dobro de cache L1, além da inserção de 256Kb de cache L2
 - Tecnologia NEON ARM
 - Extensões para acelerar H.264 e MP3
 - Desempenho 2 vezes melhor que a anterior
 - Thumb2 Instruction Set
- Alta Definição manipulando vídeos
 - 720p
- 256 MB RAM
- Gerenciamento de energia
 - Potência Dinâmica:
 - 3 níveis de *clock gating*
 - Grosso – nível de arquitetura
 - Médio– nível de blocos (NEON)
 - Fino– nível de componente lógico
 - Potência estática (multi-Vt)



Arquiteturas Multiprocessadas

▶ Omap Platform 4440

- Dual-core ARM Cortex A9 MPCore Processador de Propósito Geral
 - Cada ARM Cortex A9 pode ser composto de até 4 cores
- POWERVR™ SGX540 multi-threaded graphics accelerator
 - Suporte as APIs OpenGL 2.0 Directx 10.1
 - Reprodução de vídeo em alta definição (1080p)
- Gerenciamento de energia
 - ▶ Dynamic Power Switching (DPS)
 - ▶ Troca de modo de alimentação baseado na atividade do sistema
 - ▶ Static Leakage Management (SLM)
 - ▶ Coloca o sistema em modo standby (mais agressivo)



Propósito Geral x Embarcados

	Processador	Arquitetura	Organização	Cores	Multithreaded Cores?	Interconexão
Embarcados	OMAP4440	Heterogêneo	Heterogêneo/ Homogêneo	2 ARM Cortex A9 1 PowerVR graphics accelerator 1 Image Signal Processor	Não	Barramento
	Samsung S3C6410	Heterogêneo	Heterogêneo	1 ARM1176JZF-S + Multimedia Accelerators	Não	Barramento
	Samsung S5PC100	Heterogêneo	Heterogêneo	1 ARM Cortex A8 + Multimedia Accelerators	Não	Barramento
Propósito Geral	Cell	Heterogêneo	Heterogêneo	1 PowerPC 8 SPE	Não	Barramento
	Niagara 2	Homogêneo	Homogêneo	8 SPARC V9 ISA	Sim (8 threads)	Crossbar
	Intel 80-Cores	Homogêneo	Homogêneo	80 VLIW	Não	NoC Mesh
	Intel Larrabee	Homogêneo	Homogêneo	n P54C x86 cores SIMD execution	Não	Barramento
	NVIDIA Tesla (GeForce8800)	Homogêneo	Homogêneo	128 Stream Processors	Sim (até 768 threads)	"Network"

.....

Propósito Geral x Embarcados

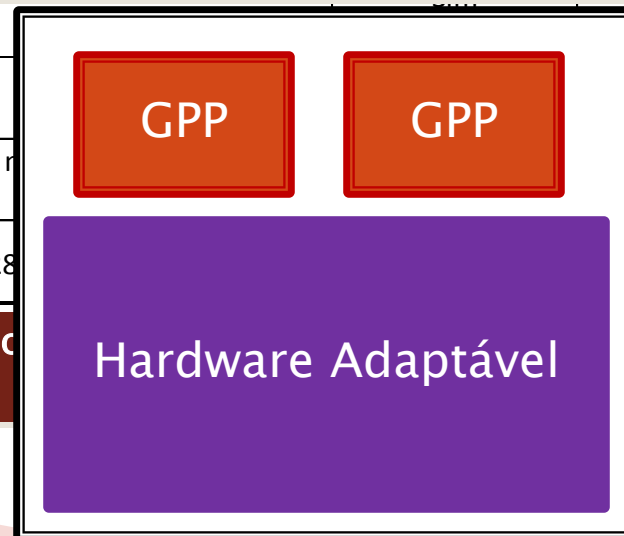
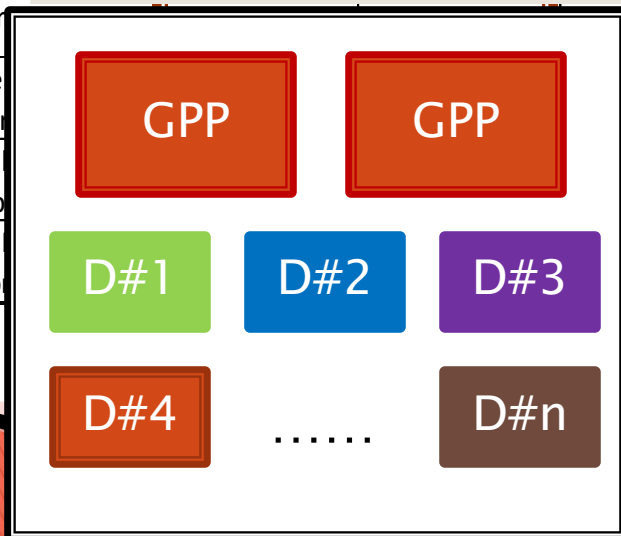
	Processador	Arquitetura	Organização	Cores	Multithreaded Cores?	Interconexão
Embarcados	OMAP4440	Heterogêneo	Heterogêneo/ Homogêneo	2 ARM Cortex A9 1 PowerVR graphics accelerator 1 Image Signal Processor	Não	Barramento
	Samsung S3C6410	Heterogêneo	Heterogêneo	1 ARM1176JZF-S + Multimedia Accelerators	Não	Barramento
	Samsung S5PC100	Heterogêneo	Heterogêneo	1 ARM Cortex A8 + Multimedia Accelerators	Não	Barramento
Propósito Geral	Cell	Heterogêneo	Heterogêneo	Fronteira entre os 2 domínios		Barramento
	Niagara 2	Homogêneo	Homogêneo	8 SPARC V9 ISA	Sim (8 threads)	Crossbar
	Intel 80-Cores	Homogêneo	Homogêneo	80 VLIW	Não	NoC Mesh
	Intel Larrabee	Homogêneo	Homogêneo	n P54C x86 cores SIMD execution	Não	Barramento
	NVIDIA Tesla (GeForce8800)	Homogêneo	Homogêneo	128 Stream Processors	Sim (até 768 threads)	"Network"

Replicação de cores com o mesmo
Instruction Set

.....

Propósito Geral x Embarcados

	Processador	Arquitetura	Organização	Cores	Multithreaded Cores?	Interconexão
Embarcados	OMAP4440	Heterogêneo	Heterogêneo/ Homogêneo	2 ARM Cortex A9 1 PowerVR graphics accelerator 1 Image Signal Processor	Não	Barramento
	Samsung S3C6410	Heterogêneo	Heterogêneo	1 ARM1176JZF-S + Multimedia Accelerators	Não	Barramento
	Samsung S5PC100	Heterogêneo	Heterogêneo	1 ARM Cortex A8 + Multimedia Accelerators	Não	Barramento
Propósito Geral	Cell	O que acontecerá quando novos domínios de aplicações migrarem para dispositivos embarcados?				Barramento
	Niagara					Crossbar
	Intel 80-Core					NoC Mesh
	Intel Larrabee					Barramento
	NVIDIA Tesla (GeForce)					"Network"



Conclusão

	Processador	Arquitetura	Organização	Cores	Multithreaded Cores?	Interconexão
Embarcados	OMAP4440	Heterogêneo	Heterogêneo/ Homogêneo	2 ARM Cortex A9 1 PowerVR graphics accelerator 1 Image Signal Processor	Não	Barramento
	Samsung S3C6410	Heterogêneo	Heterogêneo	1 ARM1176JZF-S + Multimedia Accelerators	Não	Barramento
	Samsung S5PC100	Heterogêneo	Heterogêneo	1 ARM Cortex A8 + Multimedia Accelerators	Não	Barramento
Propósito Geral	Cell			O que acontecerá quando novos domínios de aplicações migrarem para dispositivos embarcados?		
	Niagara 2	Homogêneo	Homogêneo	8 SPARC V9 ISA		Crossbar
	80-					oC Mesh
	Lar					rramento
	NV					Network"
	Tesla (Gerenciador)				(até 700 threads)	

Futuro próximo:
 -Hardware dedicado para executar os domínios atuais

Futuro não tão distante:
 -Hardware adaptável para qualquer domínio de aplicação

Exame de Qualificação:

Sistemas Embarcados Multiprocessados

Mateus Beck Rutzig

mbrutzig@inf.ufrgs.br

Orientador: Luigi Carro

carro@inf.ufrgs.br



Laboratório de Sistemas Embarcados
Instituto de Informática
Universidade Federal do Rio Grande do Sul