A interface hardware / software

Arquitetura de Computadores III

Arquiteturas Multicore de propósito geral (história)

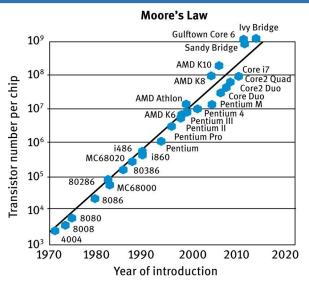
2024

Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

221

221

Lei de Moore



https://www.ncbi.nlm.nih.gov/books/NBK321721/figure/oin_tutorial.F3/

2024

Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

Dennard Scaling

- Há redução em área, tensão e atraso do circuito, com impacto em:
 - Aumento da frequência.
 - Redução em todas as distâncias levando a uma redução na capacitância.
 - Consequentemente, há uma redução no consumo de potência.
- Portanto, mesmo com o aumento em 2x na densidade de transistores (Lei de Moore) o consumo de potência permanece o mesmo.
- No entanto, Dennard ignorou as correntes de fuga e limites de tensão e, portanto:
 - Com os transistores ficando menores, a potência aumenta porque não escala em função do tamanho.
 - Isso ficou conhecido como Power Wall que limita na prática processadores com frequência de até 4GHz, desde 2006.

https://en.wikipedia.org/wiki/Dennard_scaling

2024

Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

22

223

Qual a "origem" do processador multicore?

Propósito geral

- Em uma fábrica de processadores tão tão distante...
 - O diretor para o arquiteto: Precisamos de mais desempenho!
 - O arquiteto para o diretor: Não é possível aumentar o paralelismo de instruções!
 - O diretor para o engenheiro: Precisamos de mais desempenho!
 - O engenheiro para o diretor: Não é possível aumentar a frequencia, o chip vai "queimar"!

Dennard scaling?

Quem poderá nos socorrer!?

Se este miniconto coincidir com fatos, é pura sorte!

2024

Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

Qual a "origem" do processador multicore?

Propósito geral

- Em uma fábrica de processadores tão tão distante...
 - Alguém diz: E a Lei de Moore!?
 - A Lei de Moore está relacionada à capacidade de integração.
 - Não está relacionada ao aumento de frequencia.
 - Está relacionada a quantidade de transistores em um mesmo espaço.
 - Se diminuirmos o tamanho dos transistores?
 - 180 nm, 130 nm, 90 nm, 65 nm, 45 nm, 32 nm, 22 nm....
 - Vamos aumentar a quantidade de processadores dentro do chip de processador!

 Dennard scaling?
 - Vamos chamá-los de núcleos!
 - Portanto, não adianta apenas aumentar paralelismo de instruções, nem frequência de operação!
 - Precisamos aumentar a quantidade de núcleos!

Se este miniconto coincidir com fatos, é pura sorte!

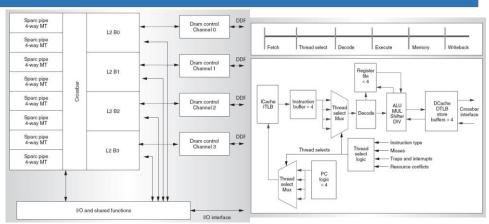
2024

Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

225

225

UltraSparc-T1 (Niagara 1)



- Suporte a 4 threads IMT por núcleo (32 threads ativas, 8 threads simultâneas).
- Crossbar switch de 134,4 GB/s.
- · 4 canais DDR 23GB/s.

KONGETIRA, P. et al., Niagara: a 32-way multithreaded Sparc processor, IEEE MICRO, v. 25, Issue 2, p. 21-29, March-April 2005.

• Potência < 80W.

2024

Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

226

Power 4 e Power 5

Processor Processor Processor Processor cache Fabric Fabric Power 4 L3 cache L3 cache Memory Memory Memory Memory (a) Processor Processor Processor Processor L3 cache L3 cache cache Power 5 Fabric Fabric controller controller Memory controller Memory controller Memory Memory (b)

- Os processadores Power 4 e Power5 possuem códigos binários e estrutura compatíveis.
- Diferença no acesso a cache L3.
- A grande diferença: o Power 5 suporte a duas threads simultâneas (SMT) por núcleo.
 - O Power5 possui dois núcleos físicos, mas quatro núcleos lógicos de processamento.

KALLA, R., SINHAROY, B., TENDLER, J. M., IBM Power5 Chip: A Dual-Core Multithreaded Processor, IEEE MICRO, v. 24, Issue 2, p. 40-47, 2004

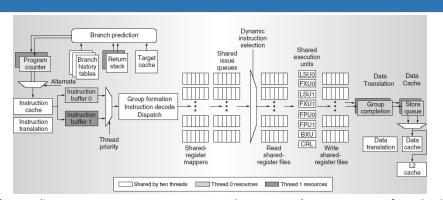
2024

Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

227

227

Power 5



- A informação mais interessante no projeto do Power5 é justamente o fato do desempenho não melhorar com o suporte SMT por núcleo. Os principais motivos são:
 - O número limitado de unidades de execução que são compartilhados entre as duas threads.
 - O alto consumo da largura de banda de memória pelas duas threads.
- Esta é a principal razão para que o Power5 também suporte apenas uma thread por núcleo.

KALLA, R., SINHAROY, B., TENDLER, J. M., IBM Power5 Chip: A Dual-Core Multithreaded Processor, IEEE MICRO, v. 24, Issue 2, p. 40-47, 2004

2024

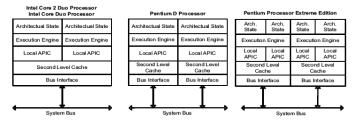
Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

228

Intel Dual Core

- Os processadores Pentium D e Core Duo, são compostos por dois núcleos, mas sem suporte a hyperthreading. Fisicamente e logicamente são dois núcleos internos.
- Diferenças básicas:
 - Suporte a múltiplas threads
 - Cache nível L2

INTEL, IA-32 Intel Architecture Software Developer's Manual, Volume 1: Basic Architecture, March 2006



2024

Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

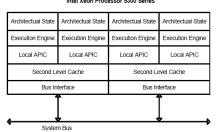
22

229

Intel Quad Core

- O projeto de um processador com 4 núcleos pela Intel é basicamente a união de dois processadores dual core.
- A cache L2 compartilhada ganhou força.
- SMT não é utilizado.

Intel Core 2 Extreme Quad-core Processor Intel Core 2 Quad Processor Intel Xeon Processor 3200 Series



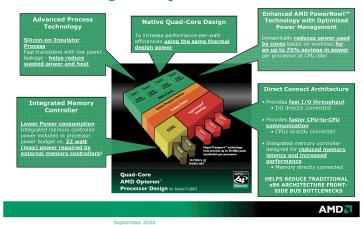
INTEL, IA-32 Intel Architecture Software Developer's Manual, Volume 1: Basic Architecture, March 2006

2024 Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

230

AMD Quad Core

Technologies Enabling Performance-Per-Watt Advantages for Quad-Core



http://ixbtlabs.com/articles2/editorial/amd-guiseppe-amato-conf-part1.html

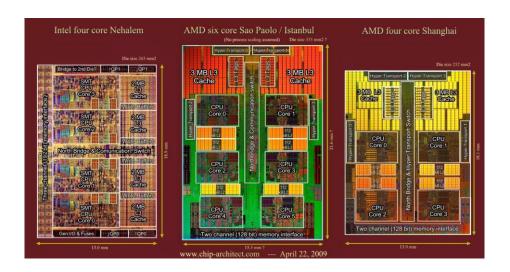
2024

Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

231

231

Nehalem, Istanbul, Shanghai

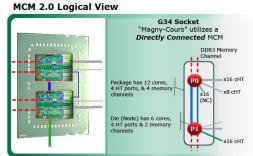


2024

Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

AMD Magny Cours (12 núcleos)

Multi-Chip Module (MCM)



Chiplets são pequenos circuitos integrados modulares que podem ser combinados para criar um sistema em um chip (SoC) mais complexo ou um design de múltiplos núcleos. Diferentemente dos chips monolíticos tradicionais, que integram todas as funcionalidades em um único chip de silício, os chiplets dividem essas funcionalidades em núcleos menores e especializados.

https://www.techpowerup.com/102445/amd-demos-48-core-magny-cours-system-details-architecture

2024

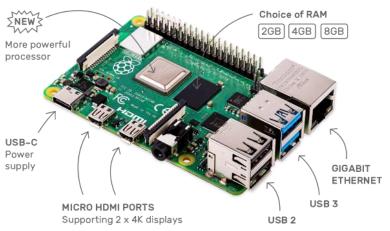
Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

233

233

Arquiteturas de Processadores Multicore

Raspberry Pi 4 Model B



https://www.raspberrypi.org/products/raspberry-pi-4-model-b/

- Broadcom BCM2711, Quad core Cortex-A72 (ARM v8) 64bit SoC @ 1.5GHz
- 2GB, 4GB or 8GB LPDDR4-3200 SDRAM (depending on model)
- 2.4 GHz and 5.0 GHz IEEE 802.11ac wireless, Bluetooth 5.0, BLE
- Raspberry Pi standard 40 pin GPIO header (fully backwards compatible with previous boards)
- Micro-SD card slot for loading operating system and data storage
- Starting at \$35

2024

Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

234

Arquiteturas de Processadores Multicore

Parallella Board



- **18-core** credit card sized computer
- #1 in energy efficiency @ 5W
- 16-core Epiphany RISC SOC
- Zynq SOC (FPGA + ARM A9)
- Gigabit Ethernet
- 1GB SDRAM
- Micro-SD storage
- Up to 48 GPIO pins
- HDMI, USB (optional)
- · Open source design files
- Runs Linux
- Starting at \$99

https://www.parallella.org/

2024

Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

235

235

Processor Intel i9



Lançamento: 2017

10 cores, 20 threads Frequência base: 3,3GHz 13.75 MB L3 Cache 140 W

Dependendo do modelo a quantidade de transistors pode variar de 3 a 22 bilhões.

2024

Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

236

Revolução Multicore - CPU

Xeon 2007: Quad-core CPUs

Xeon 2008: 6-core CPUs

• Xeon 2010: 8-core CPUs

Xeon 2013: 12-core CPUs

• Xeon 2015: 22-core CPUs

• Xeon 2017: 32-core CPUs

Xeon 2019: 48-core CPUs

Intel Xeon Cascade Lake é um processador com 48

Voltado a servidores de alto desempenho, processador Intel Xeon Cascade Lake AP tem até 48 núcleos e será lancado em 2019

Um mês depois de revelar os primeiros processadores Core de nona geração, a Intel anunciou novos chips para servidores: a linha Xeon Cascade Lake Advanced Performance (ou Xeon Cascade Lake AP), como foi batizada, pode contar com até 48 núcleos em uma única unidade.



https://tecnoblog.net/noticias/2018/11/06/intel-xeon-48-nucleos/

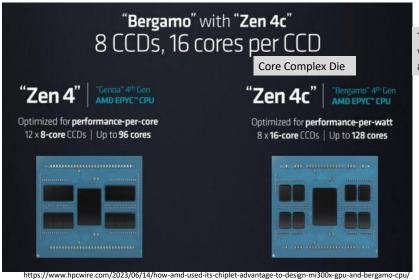
2024

Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

237

237

How AMD Used Its Chiplet Advantage to Design Bergamo CPU



The power-optimized version of the Zen 4 architecture.

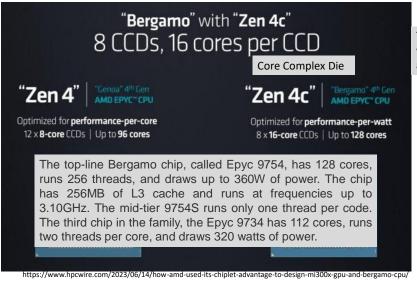
Bergamo has 82 billion transistors, and supports 128 cores per socket.

2024

Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

238

How AMD Used Its Chiplet Advantage to Design Bergamo CPU



The power-optimized version of the Zen 4 architecture.

Bergamo has 82 billion transistors, and supports 128 cores per socket.

2024

Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

23

239

Neural Processing Unit (NPU)

Processador Intel® Core™ Ultra 9 285K



Intel® Gaussian & Neural Accelerator (GNA) é um bloco de acelerador de ultrabaixa potência, projetado para executar cargas de trabalho de IA centradas em áudio e fala. O Intel® GNA é projetado para executar redes neurais baseadas em áudio com ultrabaixa potência, além de aliviar simultaneamente a CPU dessa carga de trabalho.

Número de núcleos: Nº de Performance-cores: Nº de Efficient-cores: Total de threads:

Frequência turbo max: 5.7 GHz

Frequência base do Performance-core: **3.7 GHz** Frequência base do Efficient-core: **3.2 GHz**

Cache 36 MB Intel® Smart Cache

Cache L2 total: 40 MB

Potência básica do processador: 125 W

O processador Intel® Core™ Ultra 9 285K possui 4 núcleos Xe dedicados para gráficos integrados. Esses núcleos Xe formam um total de 64 unidades de execução (EUs), ou seja, 512 shaders unificados.

2024

Arquitetura de Computadores III - Engenharia e Ciência da Computação - PUC Minas

240