

Memória Cache

L1, L2, L3 e L4

Grupo 1

Caio Cominato Petrelli	17.00100-5
Gabriel Euzébio Dicier	17.01096-9
Guilherme Keese Tabacow	17.00666-0
Lucas Negresio Libório	17.01165-5
Marcello Beer	17.00865-4

Introdução

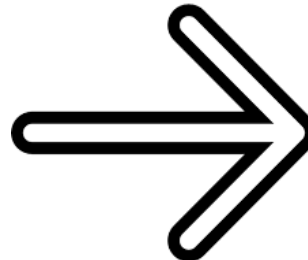
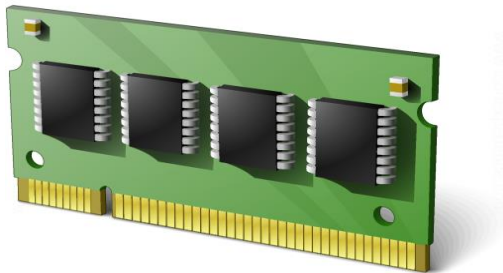


- **Objetivo**

- Potencializar o **desempenho** do chip no processamento.

- **Como?**

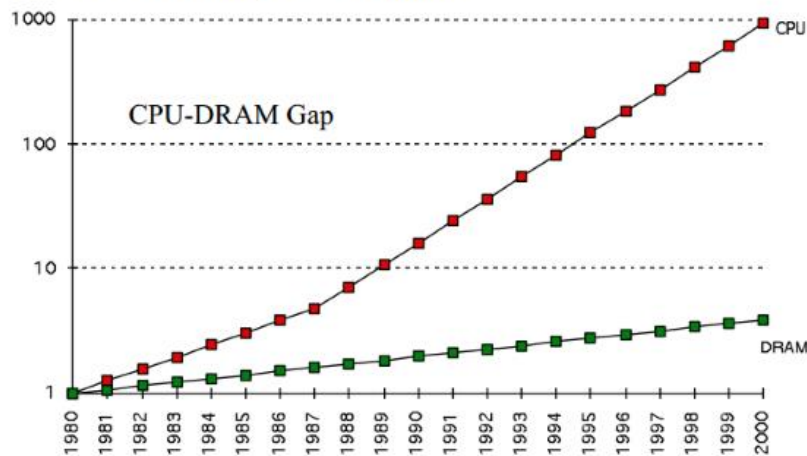
- Trabalhando em **conjunto** com processador.
- Posicionando os blocos de memória com grande probabilidade de reuso pela CPU mais próximos.



Necessidade de um Intermédio

- **Clock da Memória < Clock do Processador**
 - CPU processa dados em taxa **maior** que a RAM pode enviar

▪ Processor vs Memory Performance

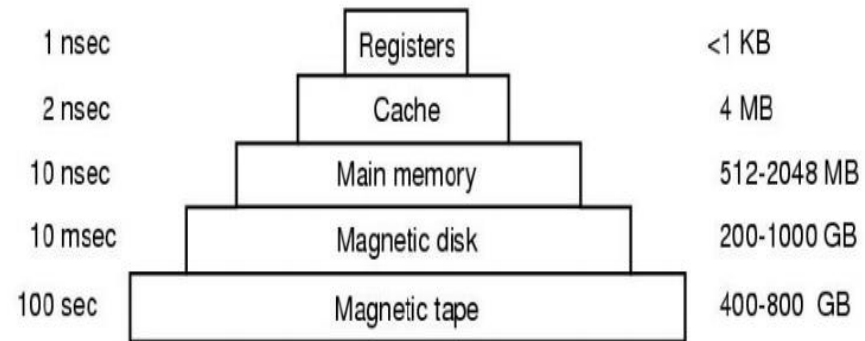


1980: no cache in microprocessor;

1995 2-level cache

Typical access time

Typical capacity



Comparação entre velocidades de mídias

Necessidade de um Intermédio

- **Memória Cache**
 - Mais rápida que a memória RAM
 - Fornece informações mais cruciais ao processador
 - Evita que o processador requisiite várias vezes dados da RAM.
- **Memoria cache é mais cara**
 - Portanto é restrita apenas às chamadas recentes e importantes.



Funcionamento do Cache

- **Caches de CPU:**

- Pequenos pacotes de memória
 - Armazenamento requerido pela CPU
 - Priorização feita por algoritmos



- **Objetivo**

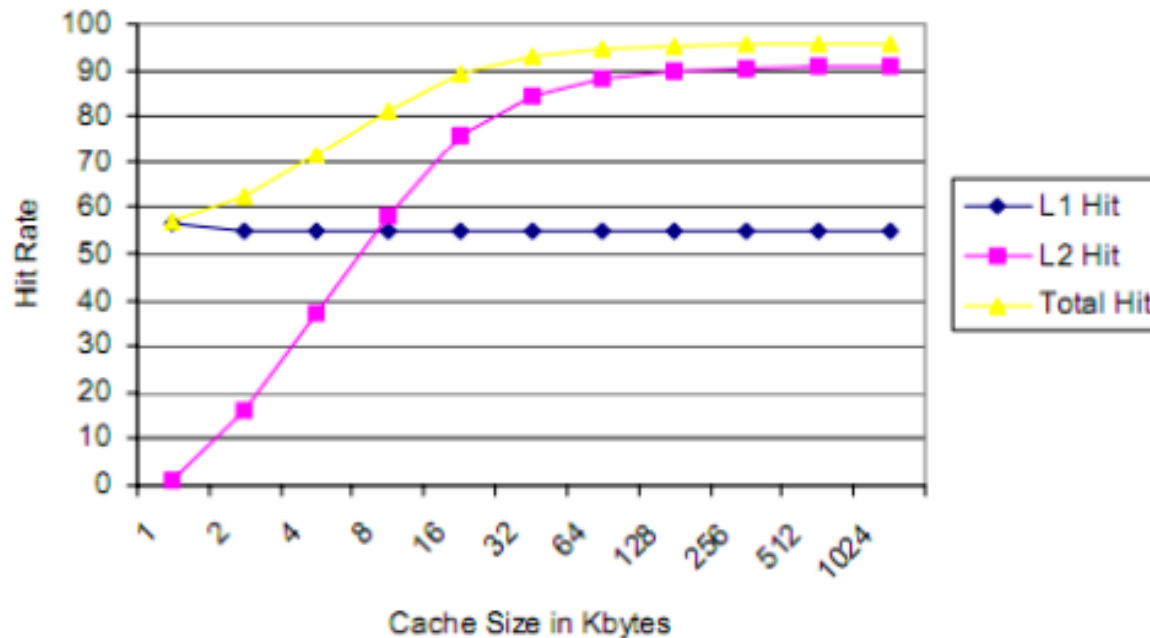
- Assegurar que a CPU tenha acesso aos dados no cache (**cache hit**).
- **cache miss** é quando a memória não é encontrada
- Neste momento entra em ação o nível de cache **L2**, em caso negativo **L3**.

- **Fabricantes desenham o cache de forma diferente:**

- Padrão na requisição da RAM

Taxas de Hit L1 x L2

Hit Rates for Constant L1, Increasing L2



- Cache L1 possui uma taxa de hit constante.
- Já o cache L2, seu tamanho é diretamente proporcional a taxa de hit
- Um L2 maior e mais lento pode prover os mesmos benefícios do L1

Memória Cache - Onde Está?

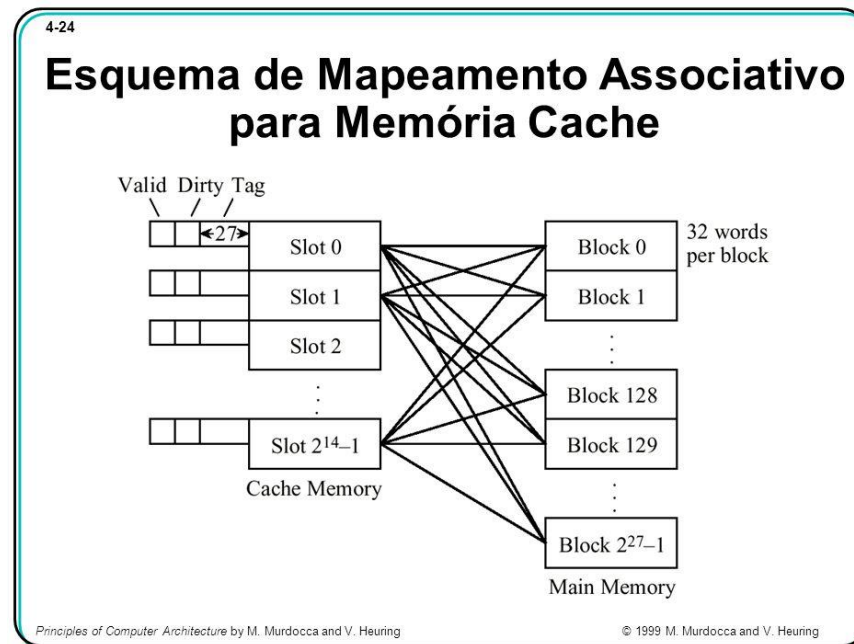
- Qualquer dispositivo que requisite um recurso a outro apresenta cache.
 - Exemplo:
 - Processadores
 - Microcontroladores e Microprocessadores
 - GPUs
 - FPGAs
 - Discos Rígidos
 - Placas-mãe
 - Clusters de Banco de dados

Cache em Níveis

- A evolução da velocidade nos processadores permitiu a divisão do cache em níveis.
- Atualmente a demanda por alta velocidade de memória é muito grande.
 - Por isso, são construídas em níveis que se diferem em tamanho versus desempenho.

Associatividade da memória cache

- Quando o cache for associativo, **qualquer** bloco de RAM pode ser armazenado em **qualquer** bloco do cache.



Cache L1

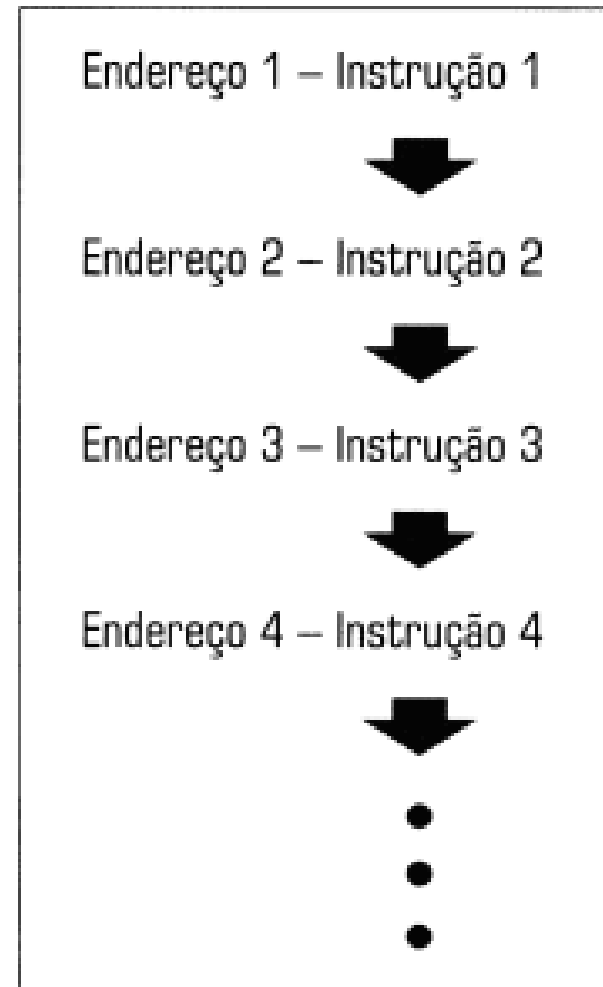
- **O cache L1 é parte da memória estática.**
 - Memória estática são memórias não-voláteis.
- **A primeira geração de caches L1**
 - Criados após os processadores Intel 80486
 - Cache unificado.
- **A partir da 5ª geração de processadores Intel,**
 - Cache L1 foi dividido em 2 (cache de dados e cache de instruções)
 - Aumentando o desempenho

Cache L1

- O circuito que carrega as instruções da RAM no processador é a unidade de busca.
- Com o surgimento do cache L1, as unidades de busca procuravam diretamente no L1.
- Circuito tentava adivinhar quais seriam as próximas instruções que a CPU deseja.

Cache L1

- Adivinhação das instruções
 - A maioria dos programas são executados sequencialmente.
- Unidade de busca pede instruções do endereço X:
 - Controlador carrega bloco de endereços sequenciais para dentro do cache L1 de instruções.



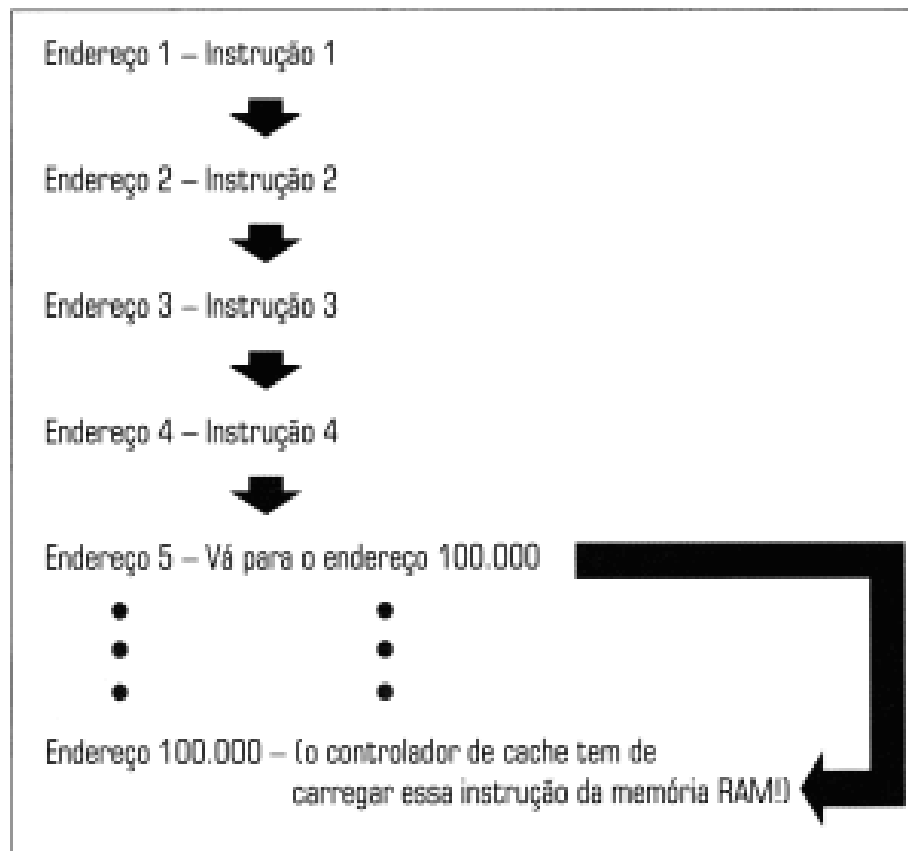
Cache L1

- Então, de modo simplista, supondo um processador com cache L1 de instruções de 16 KB.
- Se o processador pede a unidade de busca o primeiro endereço, o controlador de cache carrega logo os 16 KB seguintes.
- Cache L1 é bem pequeno comparado aos outros níveis.

Cache L1

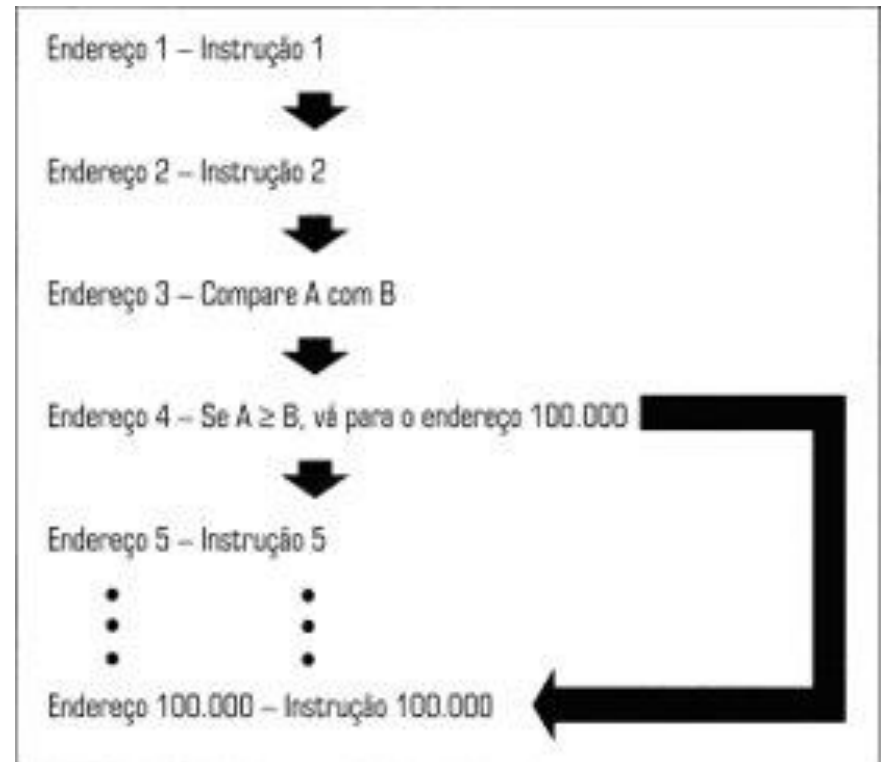
- **Este método funcionaria bem se todos os programas fossem executados de forma sequencial.**
 - Na prática podem ocorrer desvios
- **Endereço já carregado** no cache L1:
 - **Sem problemas** de desempenho.
- Controlador de cache apresenta uma **previsão de desvio**
- **Desvios incondicionais:**
 - Circuito de previsão procura instruções do tipo **JUMP**
 - Controlador carrega informações do endereço antecipadamente no L1.

Exemplo de Desvio de Instrução



Cache L1 – Desvios Condicionais

- **Desvios condicionais:**
 - Não tem como saber para qual endereço seguir
- **Circuito de previsão:**
 - Carrega as **duas ramificações** possíveis
 - Descarta a não usada



Cache L1 – Circuitos de Previsão de Desvio

- O processador tem 3 pequenas memórias para o circuito de previsão de desvio.
 - **BHT** (Tabela com Histórico de Desvios).
 - **BTB** (Buffer de destino do desvio)
 - **RAS** (Pilha de endereços de retorno).
 - Armazena o endereço de origem da instrução **CALL** (chamadas de subrotina).
- **Quanto maior**, mais eficiente, aumentando o desempenho do processador.

Cache L1 - Cache de Instruções

- O cache L1 de instruções podem ou não ter **dados acoplados**. Instrução significa tudo o que é **carregado** para dentro do **processador**.
 - Exemplo - Instrução MOV AX, FFFFh
- Existem instruções que podem **buscar ou armazenar dados de endereço** de memória RAM.
 - Exemplo - Instrução do tipo MOV [2000h], AX
- O processador executa **duas tarefas (leitura e gravação)**

Cache L1 - Cache de Dados

- Cache L1 de dados possui um **papel diferente**.
- Pedido é enviado ao cache L1 de dados pela unidade de armazenamento do processador.
 - O processador é "liberado"
 - Controlador atualiza a memória RAM com o dado.

Cache L1 - Operações de Leitura e Escrita

- Sem o cache L1: operações de escrita e leitura na memória teriam que ser diretas na RAM .
- O processador acessa a memória RAM através de seu clock externo.
- O processador envia os dados ao cache L1 em seu clock interno e efetua novas tarefas enquanto o controlador do cache atualiza a RAM.

Cache L1 - Operações de Leitura e Escrita

- Caso o endereço já tenha sido lido há pouco tempo, seu conteúdo ainda permanece do cache de dados e o controlador de cache não necessita buscar na RAM.
- Se o dado não estiver, o controlador irá procurá-lo em L2 e assim vai.

Cache L1 – Processadores Intel 7ª Geração

- Possuem arquitetura de cache L1 diferente.
 - Não possuem cache L1 de instruções
 - Possuem cache de execução de 150 KB que armazena microinstruções RISC decodificadas.
- Caches L1 de dados
 - Tamanho de apenas 8 KB
 - Conectados aos caches L2 com um caminho de 256 bits (4x maior que todos os outros processadores existentes)
 - Desempenho 4x maior.
 - Pentium III de 1 GHz: 8000 MB/s,
 - Pentium 4 de 1 GHz: 32000 MB/s

Conclusão Cache L1

- O cache L1 de dados e de instruções tem funções completamente diferentes, mas normalmente soma-se o valor dos dois para dizer a capacidade de memória de cache L1 de um processador.
- O cache dividido aumenta o desempenho do processador.
- Apenas um dispositivo pode acessá-lo por vez.
 - Cache dividido, apresentam controle mais simples e circuitos independentes que podem ser acessados simultaneamente.

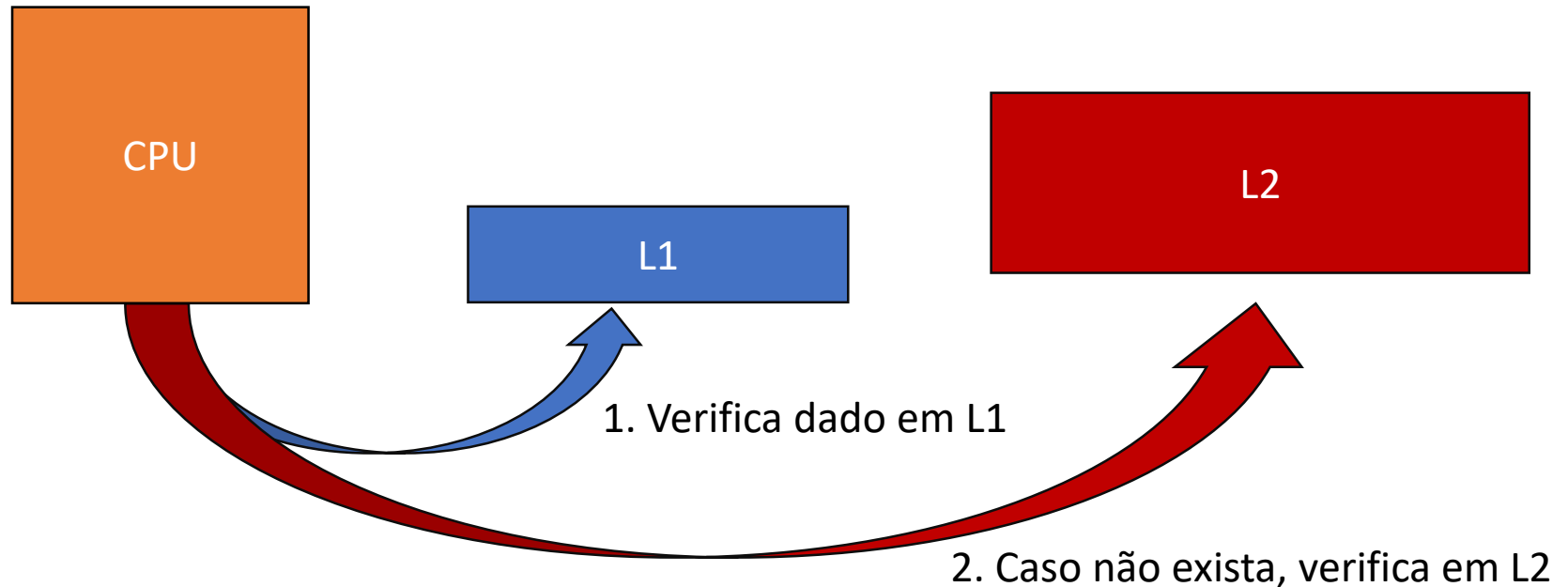
Cache L2

- Cache L1

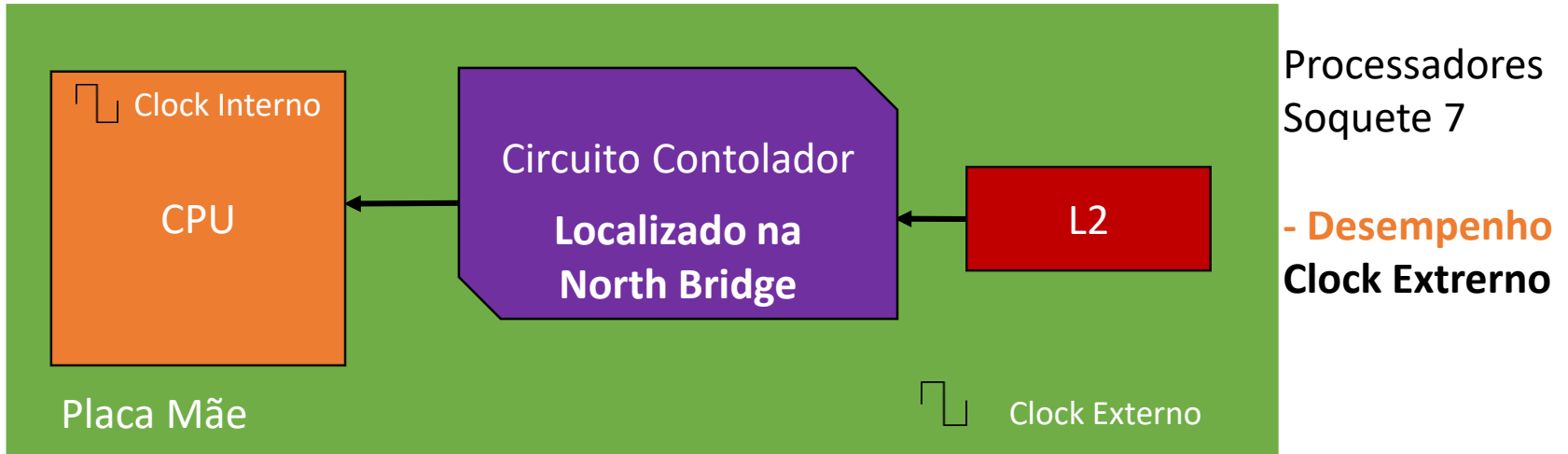
- Essencial para CPU
- Pouca capacidade
- Agiliza o carregamento de instruções pro processador.

- Cache L2

- Maior capacidade
- Carrega blocos maiores de dados da memória RAM
- Pipeline parecido com L1
 - (Carrega dados baseado em frequência)

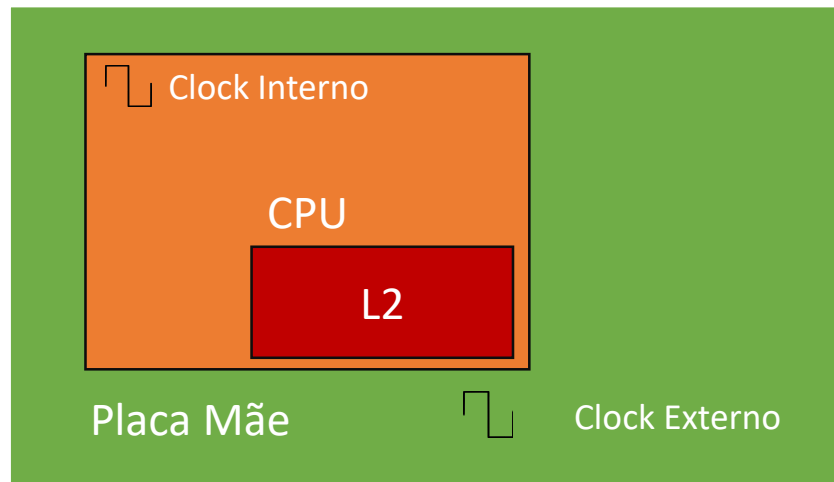


Cache L2 - Localização Física

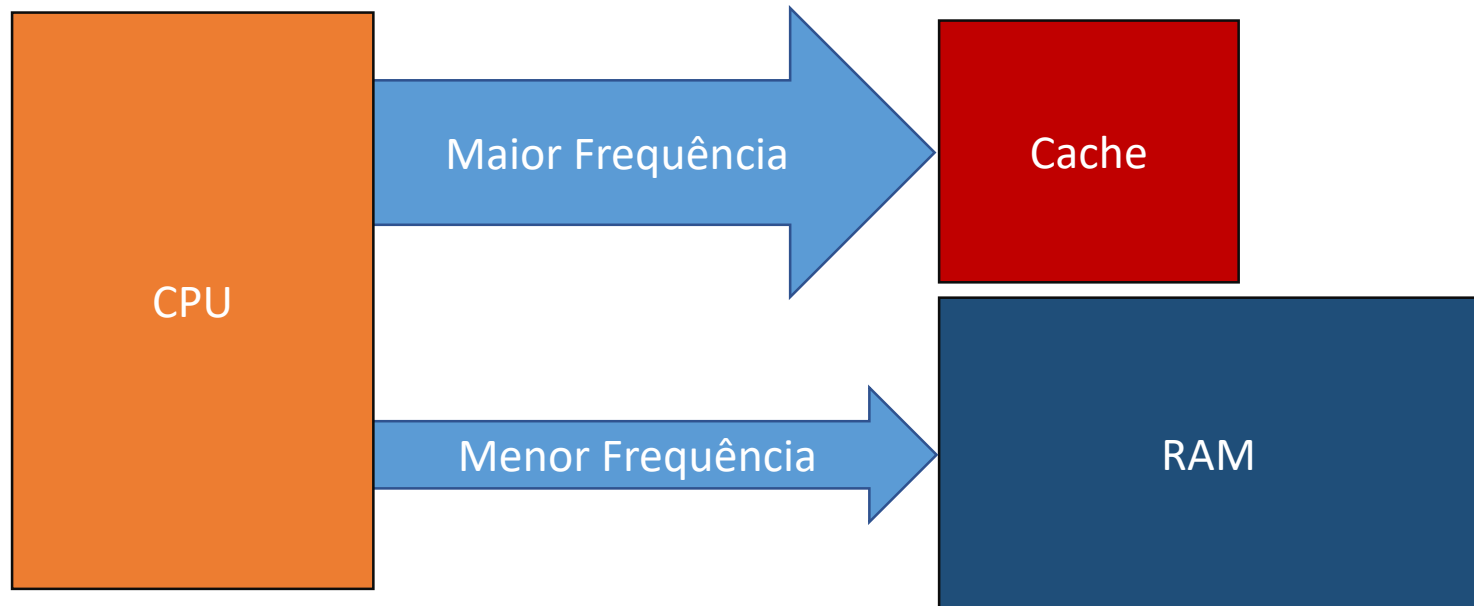


Processadores Intel
6ª Geração

+ Desempenho
Clock Interno



Cache L2



- Wait States
- Latência da Memória
 - Velocidade Menor de Barramento

Conclusão Cache L2

- Desempenho do cache L2 depende de sua quantidade, frequência em que é acessado e arquitetura.
- Antigamente, L2 era a memória compartilhada.

Cache L3

- **Só é utilizada quando a informação não é encontrada na memória L2.**
- **Memória maior em comparação com níveis mais baixos.**
- **A memória L3 é compartilhada entre todos os núcleos.**
 - Surgimento da memória L3 fez com que as memórias L1 e L2 não precisassem de tanta memória, tornando-as mais eficientes em troca de menos memória.
 - A memória L3 é compartilhada entre todos os núcleos.
- **A memória cache L3 está quase sempre localizada na placa mãe.**

Cache L4

- *"Os processadores Intel IA-64 permitem ainda a utilização de um cache L4 na placa-mãe. A utilização desse cache depende do fabricante da placa-mãe."*

Portanto são exclusivas na arquitetura IA-64.

- **Comportamento entre ela e o Cache L3 é semelhante entre os caches L1 e L2.**
- **Localizada na Placa-mãe.**
- **Descontinuação da arquitetura IA-64 em 2021.**

Conclusões

- **Memórias cache são fundamentais**
- **Clock nem sempre é o único fator determinante para um computador.**
 - Clock RAM em torno de **2.1 MHz**
 - Clock Procesasdor em torno de **3.0 GHz**

Referências Bibliográficas

- <https://www.extremetech.com/extreme/188776-how-l1-and-l2-cpu-caches-work-and-why-theyre-an-essential-part-of-modern-chips>
- <https://bootblockbios.wordpress.com/hardware/memoria-ram/cache-de-memoria/>
- <https://pt.wikipedia.org/wiki/Cache>
- <https://www.techtudo.com.br/noticias/noticia/2016/10/o-que-e-memoria-cache-entenda-sua-importancia-para-o-pc.html>

Memória Cache

L1, L2, L3 e L4

Grupo 1

Caio Cominato Petrelli	17.00100-5
Gabriel Euzébio Dicier	17.01096-9
Guilherme Keese Tabacow	17.00666-0
Lucas Negresio Libório	17.01165-5
Marcello Beer	17.00865-4