

Lab. Manut. e Mont. de Computadores

Aula 08 - Memórias e Hierarquia de Memória

Prof. Dr. Wendell Fioravante da Silva Diniz
3º Ano - Informática Integrado
Centro Federal de Educação Tecnológica de Minas Gerais
Unidade Varginha

Recapitulando...

- O barramento PCI é um dos padrões mais difundidos para interconexão de placas de expansão
- Criado pela Intel nos anos 90, teve suas patentes publicadas em domínio público e seu desenvolvimento gerido pelo consórcio PCI Interest Group
- Arquitetura em mezzanino, ligando-se ao processador por meio de uma ponte
- É um barramento multiplexado, síncrono com arbitração centralizada
- Opera a 66 MHz, com taxa de transferência máxima teórica de 528 MB/s

A Memória

- A memória são uma das partes vitais de um sistema de computação
- Sua função é armazenar os dados que serão manipulados pelo pelo sistema, permitindo que estes possam ser recuperados quando necessário
- Conceitualmente, a memória é um depósito de dados
- Na prática, a memória de um sistema computacional não é um único componente, mas um subsistema constituído de vários componentes interligados segundo uma hierarquia
- Embora simples em conceito, o sistema de memória apresenta a maior variação de tecnologia, organização, desempenho e custo em um computador



Características de um sistema de memória

- Um sistema de memória pode ser classificado segundo suas características. As mais importantes são:
- **Localização**
- **Capacidade**
- **Unidade de transferência**
- **Método de acesso**
- **Desempenho**
- **Tipo físico**
- **Características físicas**
- **Organização**

Localização

- Quanto à localização, as memórias são classificadas como:
- **Internas:** fazem parte desta classificação os registradores, os caches e a memória principal. Recebem este nome porque estão envolvidas no processo de funcionamento do sistema computacional como um todo
- **Externas:** são os discos magnéticos, discos ópticos, unidades de fita, cartões de memória. São usados sobretudo para armazenamento, não participando do funcionamento do sistema como um todo

Capacidade

- A capacidade refere-se ao número de bits que uma memória pode armazenar
- Pode-se medir a capacidade de um sistema de memória em número de palavras ou número de bytes
- Os tamanhos mais comuns de palavras são 8, 16, 32 e 64 bits
- Ex um módulo de 1024 palavras de 32 bits terá:
 - $1024 \times 32 = 32768 \text{ bits} = 4096 \text{ bytes} = 4 \text{ MiB}$

Unidade de transferência

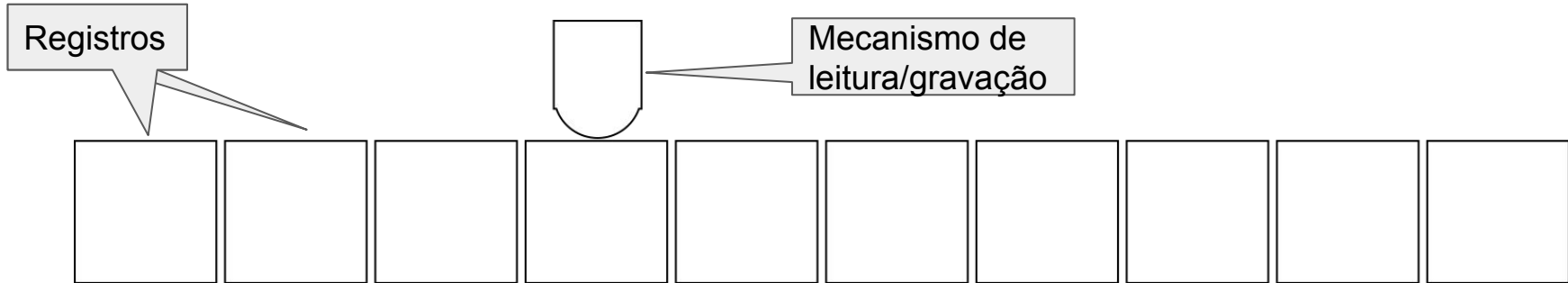
- Refere-se, no caso da memória interna, ao número de linhas elétricas disponíveis para transferência dos dados
- Pode ser igual ou maior que o tamanho da palavra
- **Palavra:** é a unidade “natural” dos dados. O tamanho da palavra no sistema geralmente refere-se ao tamanho em bits usado para armazenar um número inteiro e ao tamanho da instrução. Porém essa regra não é seguida fielmente. Ex: CRAY 90 representa inteiros com 46 bits e tem instruções de 64 bits. X86 tem instruções de tamanho variável e inteiros de 32 bits

Unidade de transferência

- **Unidade endereçável:** é o número mínimo de bits que pode ser endereçado. Alguns sistemas endereçam os dados a nível de palavras, outros a nível de bytes. Em ambos, a relação entre o número de bits (A) de endereço e quantidade de unidades endereçáveis (N) é dada por: $N = 2^A$
- **Unidade de transferência:** finalmente, refere-se ao número de bits que podem ser acessados em uma única operação. Não precisa ser necessariamente igual ao tamanho da palavra, sendo geralmente maior
- Na memória externa, os dados são transferidos em unidades muito maiores, chamadas de blocos

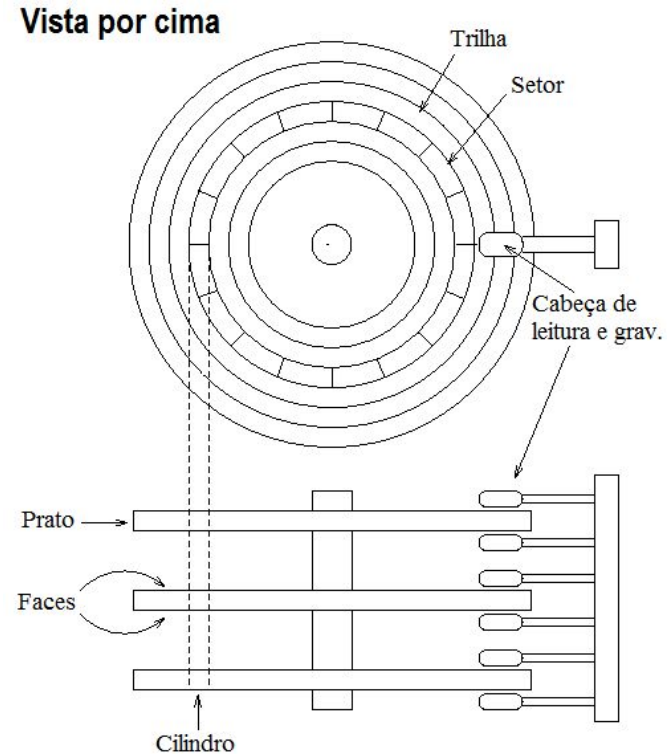
Métodos de acesso

- **Acesso sequencial:** a memória é acessada por um mecanismo de leitura/gravação compartilhado. Os dados estão organizados em unidades dispostas de forma linear em unidades chamadas registros. O tempo para acessar um registro é extremamente variável, dependendo de sua posição absoluta e da posição atualmente em uso. Ex: unidades de fita



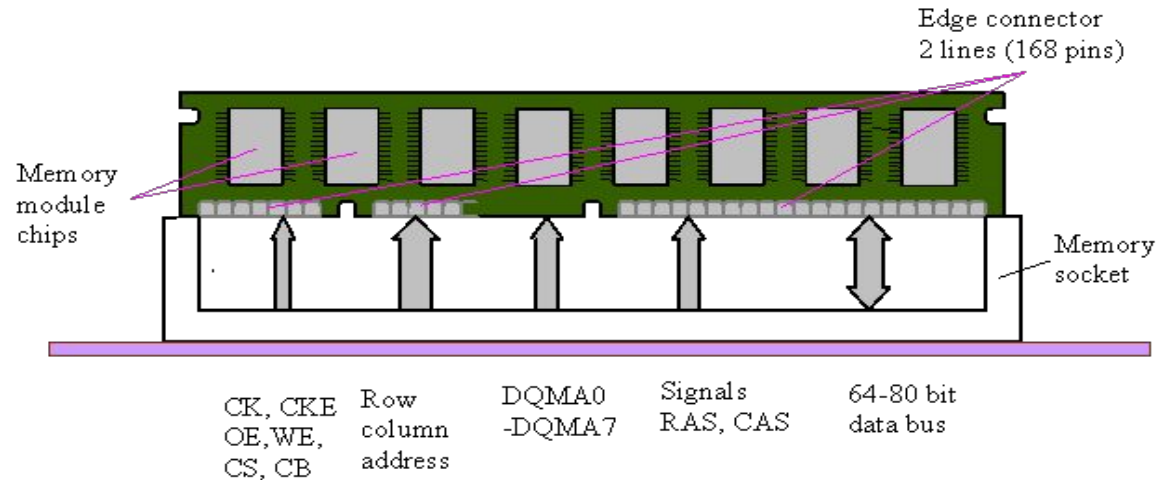
Métodos de acesso

- **Acesso direto:** como no acesso sequencial, o acesso direto envolve um mecanismo de leitura/escrita compartilhado. Entretanto, o acesso é feito em dois níveis, uma vizinhança geral acessada diretamente, mais uma busca sequencial dentro do setor. Ex. unidades de disco



Métodos de acesso

- **Acesso aleatório:** os dados são organizados em endereços únicos que podem ser acessados em qualquer ordem, independentemente de sua posição física no módulo. O tempo de acesso é constante para todos os endereços. Ex. memória principal e alguns sistemas de cache



Métodos de acesso

- **Acesso associativo:** neste método, uma chave de busca é comparada paralelamente com o conteúdo das palavras de memória. Assim, uma palavra é recuperada com base em parte de seu conteúdo. Como no acesso aleatório, cada posição tem um mecanismo de acesso único e tempo de acesso constante. Permite acessos muito rápidos, por isso é usado em alguns sistemas de cache.
- Alta complexidade e alto custo impedem seu uso em larga escala

Desempenho

- **Tempo de acesso (Latência):** para memórias de acesso aleatório, refere-se ao tempo necessário para completar uma operação de leitura ou escrita. Ou seja, o tempo decorrido entre o momento em que se apresenta um endereço à memória e o momento em que este aparece na saída, ou o tempo que os dados devem estar presentes na entrada até que a gravação seja garantida. Para a memória de acesso não-aleatório, refere-se ao tempo necessário para posicionar o mecanismo de leitura/gravação
- **Tempo de ciclo de memória:** refere-se ao tempo total incluindo o tempo de acesso mais qualquer tempo adicional necessário para que a memória esteja pronta para o próximo acesso

Desempenho

- **Taxa de transferência:** é a taxa em que os dados são transferidos para a memória. Para as memórias de acesso aleatório, é dada por 1/tempo de acesso. Para as outras memórias, existe a relação:

$$T_N = T_A + \frac{n}{R}$$

Onde: T_N = tempo médio para ler ou escrever n bits

T_A = tempo de acesso médio

n = número de bits

R = taxa de transferência em bps

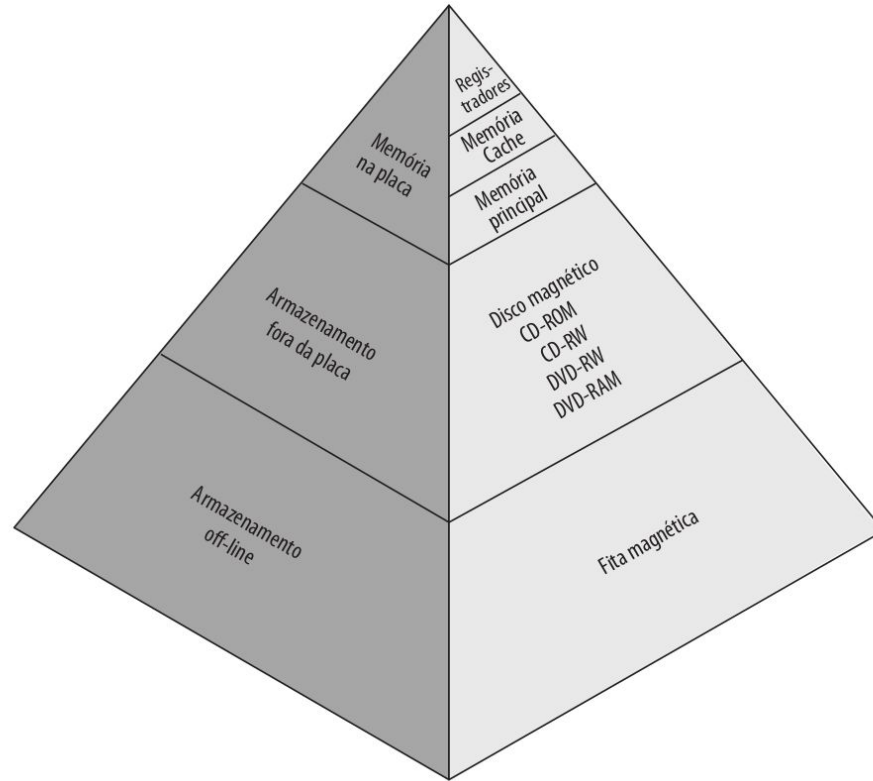
Tipo físico

- **Semicondutores:** são os módulos formados por pastilhas (chips) de silício ou outra substância semicondutora. Um circuito eletrônico pode ser ligado ou desligado para representar um 0 ou 1
- **Magnéticas:** são formados por uma camada metalizada e magnetizada. A mudança no alinhamento do campo magnético da camada representa um 0 ou 1
- **Ópticos:** usa uma substância capaz de refletir ou não um feixe de luz laser, representando um 0 ou 1
- **Optomagnéticos:** a leitura e gravação acontecem por meios ópticos, mas sob a presença de um campo magnético que controla o tipo de operação em conjunto com a potência do laser

Hierarquia de memória

- As restrições de um projeto de sistema de memória pode ser resumido em três questões:
 - Quanto?
 - Com que velocidade?
 - Com que custo?
- Existe uma relação entre essas três variáveis, a saber:
 - Menor tempo de acesso, maior custo por bit
 - Maior capacidade, menor custo por bit
 - Maior capacidade, tempo de acesso mais lento
- Para otimizar essa equação, as diversas memórias foram divididas de acordo com uma hierarquia, permitindo aproveitar as características de cada tecnologia

Hierarquia de memória



Hierarquia de memória

- Conforme se desce na memória, acontece o seguinte
 - Aumento no tempo de acesso
 - Aumento na capacidade
 - Diminuição do custo por bit
 - Diminuição da frequência de acesso pelo processador
- Desta forma, coloca-se as memórias mais rápidas mais perto do processador, em pequena quantidade e as memórias mais lentas em maior quantidade, porém mais longe do processador
- **Volatilidade:** é a capacidade da memória de reter dados sem a presença de alimentação elétrica. As memórias de maior hierarquia são voláteis, ou seja perdem seus dados quando o computador é desligado
- As memórias não voláteis são conhecidas como memória de massa ou de armazenamento

Registradores

- São internos ao processador
 - Guardam os dados e instruções que estão sendo processadas ou serão imediatamente processadas
- Por serem construídas com a mesma tecnologia do processador, tem tempo de acesso muito baixo
 - Tempo de acesso em ciclo único (1-2 ns)
- Podem armazenar um único dado, instrução ou endereço. Ex. Core i7 possui 16 registradores de 64 bits
- Voláteis
- Tecnologia semicondutora

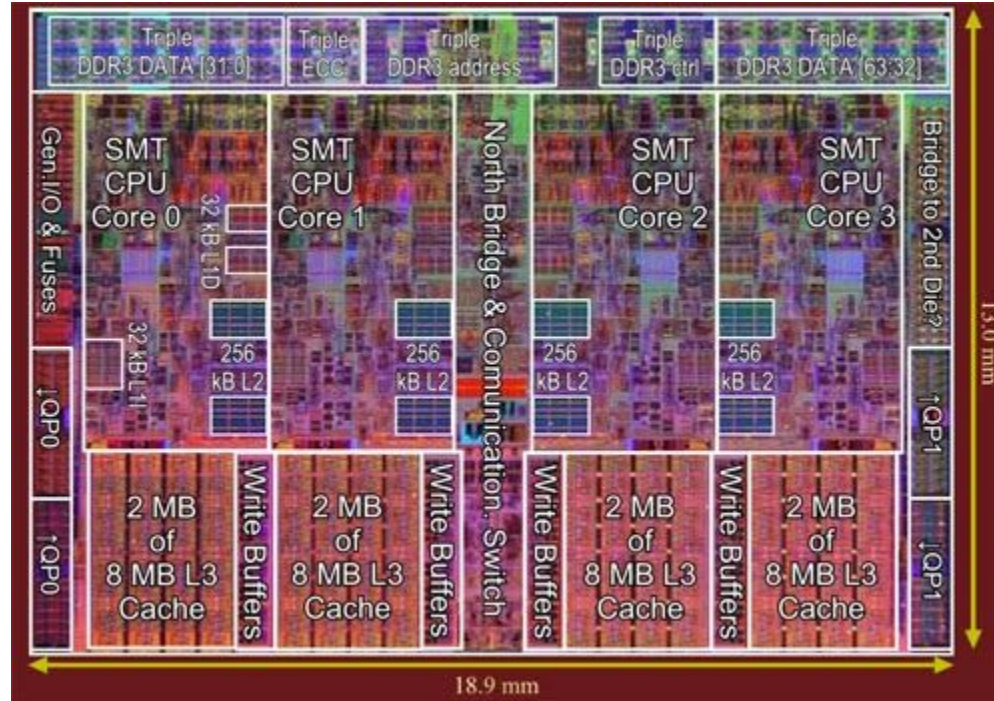
Caches

- Criada para melhorar os tempos de acesso à memória principal pelo processador
- Separadas em níveis, com quantidades e velocidades diferentes
 - L1
 - L2
 - L3
- Antigamente, o nível L1 referia-se ao cache interno ao processador. O L2 poderia ser interno ou externo e o L3 sempre externo
- Os processadores modernos integram os 3 níveis de cache

Caches

- Tempo de acesso: um ciclo, 5-20 ns
- Capacidade: varia conforme o modelo da CPU. Ex: Core i7 7500U
 - L1: 128 KiB (4x32 KiB)
 - L2: 512 KiB (2x256 KiB)
 - L3: 4 MiB (2x2 MiB)
- Voláteis e de alto custo de produção
- Tecnologia semicondutora (SRAM)

Caches

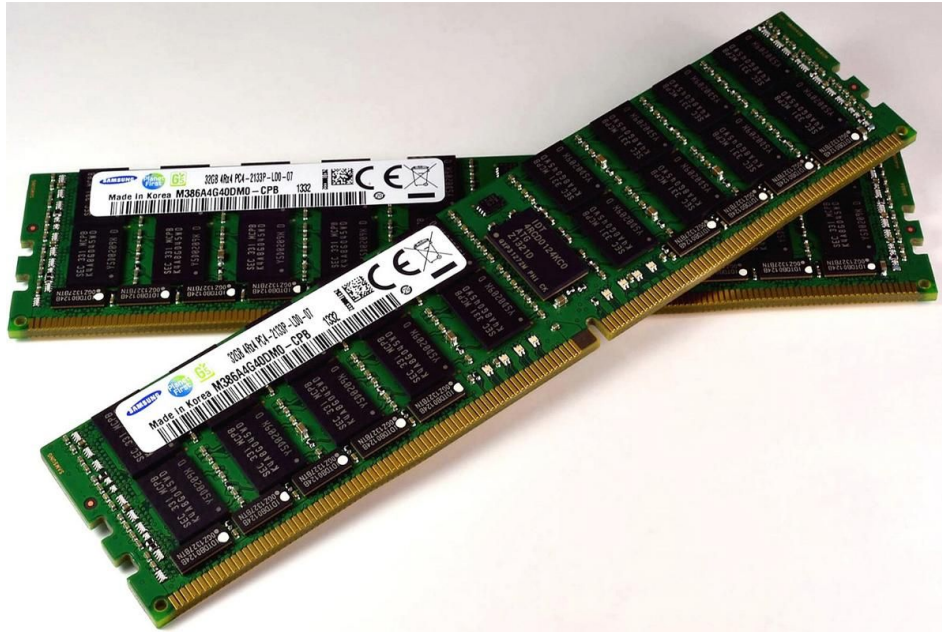


Memória Principal

- É a região conhecida como memória RAM
- É a memória básica do sistema computacional. É nela que programas e dados são armazenados e acessados pelo processador
- Tempo de acesso: 50-80 ns
- Capacidade: maior que os caches, limitada pela arquitetura do processador e capacidade do controlador de memória
 - Arquitetura de 32 bits: 2^{32} endereços = 4.294.967.296 bytes = 4 GiB
 - Arquitetura de 64 bits: 2^{32} endereços = $1,844674407 \times 10^{19}$ bytes = 17.179.869.184 GiB = 16 EiB
- Volátil
 - Existe uma pequena parte da memória principal que não é volátil. Trata-se da BIOS, programa que é executado toda a vez que o computador é ligado

Memória principal

- Tecnologia de semicondutores
- Custo mais baixo que os caches



Memória secundária

- São as memórias de armazenamento, onde os programas e dados são armazenados quando o computador é desligado
- Os discos magnéticos (HDD) são a tecnologia corrente com melhor custo-benefício
- Estão sendo substituídos pelos SSDs (Drives de estado sólidos), que são mais caros, mais rápidos e menores em capacidade
- Os discos ópticos eram usados para transportar grandes volumes de dados
- Foram substituídos pelos cartões de memória e memory keys
- Tempos de acesso
 - HDD: 8-30 ms
 - Discos ópticos: 120-300 ms
 - SSD: 0.1 a 0.3 ms

Leitura Recomendada

-