

Arquitetura e Organização de Computadores

Guilherme Henrique de Souza Nakahata

Universidade Estadual do Paraná - Unespar

21 de Maio de 2024

Conceitos básicos e evolução do computador

- **Arquitetura de computador** refere-se aos atributos de um sistema visíveis a um programador.
- **Organização de computador** refere-se às unidades operacionais e suas interconexões que percebam as especificações de arquitetura.
- Historicamente, e ainda hoje, a distinção entre arquitetura e organização tem sido importante.

- Em termos gerais, há somente quatro funções básicas que podem ser apresentadas pelo computador:
 - **Processamento de dados.**
 - **Armazenamento de dados.**
 - **Movimentação de dados.**
 - **Controle.**

- Há quatro componentes estruturais principais:
 - Unidade central de processamento (CPU — do inglês, Central Processing Unit).
 - Memória principal.
 - E/S.
 - Sistema de interconexão.

Componentes principais para computador de uso geral

- **Memória principal.**
- **Unidade Lógica e Aritmética (ULA).**
- **Unidade de Controle.**
- **Dispositivos de E/S.**

Componentes principais para computador de uso geral

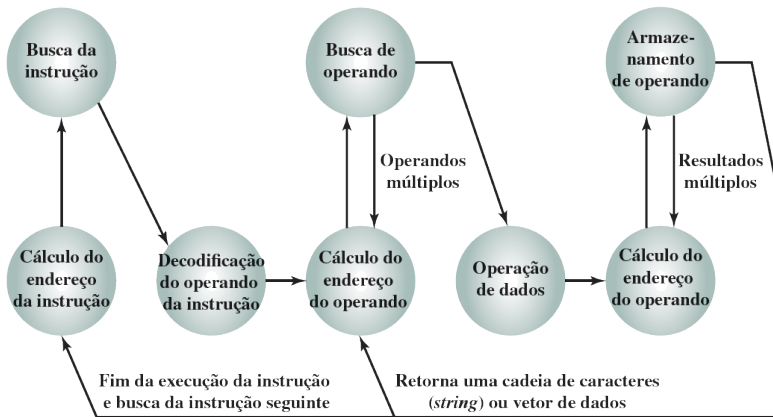
- **Memória principal:** Armazena dados e instruções;
- **Unidade Lógica e Aritmética (ULA)** Realiza operações com dados binários;
- **Unidade de Controle:** Interpreta e executa instruções armazenadas na memória;
- **Dispositivos de E/S:** Permite a interação entre o computador e o usuário.

- Alterar e entrar programas no ENIAC;
- Programa ser armazenado na memória junto com os dados;
- Obter instruções lendo da memória;
- Programas poderiam ser alterado ou criado com valores na memória;
- Conceito de **programa armazenado**;

Arquitetura de Von Neumann

- Proposto por Von Neumann em 1945;
- Baseada em um modelo de computador de único processador;
- Capaz de executar uma série de instruções armazenadas na memória;
- Consistindo em cinco componentes principais:
 - Unidade central de processamento (CPU);
 - Memória;
 - Dispositivos de entrada e saída;
 - Controlador de entrada e saída;
 - Barramentos;

Ciclo de instrução



Componentes do computador

- No início de cada ciclo de instrução, o processador busca uma instrução da memória.
- Em um processador típico, um registrador chamado **contador de programa (PC)** mantém o endereço da instrução a ser buscada em seguida.
- O processador sempre incrementa o PC após cada busca de instrução, de modo que buscará a próxima instrução em sequência (ou seja, a instrução localizada no próximo endereço de memória mais alto).

Componentes do computador

- No início de cada ciclo de instrução, o processador busca uma instrução da memória.
- Em um processador típico, um registrador chamado **contador de programa (PC)** mantém o endereço da instrução a ser buscada em seguida.
- O processador sempre incrementa o PC após cada busca de instrução, de modo que buscará a próxima instrução em sequência (ou seja, a instrução localizada no próximo endereço de memória mais alto).

Busca e execução de instruções

- O processador interpreta a instrução e realiza a ação solicitada.
- Em geral, essas ações estão em uma destas **quatro** categorias:
- Processador-memória: os dados podem ser transferidos do processador para a memória ou da memória para o processador.
- Processador-E/S: os dados podem ser transferidos de ou para um dispositivo periférico, transferindo entre o processador e um módulo de E/S.
- Processamento de dados: o processador pode realizar alguma operação aritmética ou lógica sobre os dados.
- Controle: uma instrução pode especificar que a sequência de execução seja alterada.

- As interrupções são fornecidas como um modo de melhorar a eficiência do processamento.
- Com as interrupções, o processador pode estar engajado na execução de outras instruções enquanto uma operação de E/S está em andamento.
- Depois que essas poucas instruções tiverem sido executadas, o controle retorna ao programa do usuário.
- Enquanto isso, o dispositivo externo está ocupado aceitando e imprimindo dados vindos da memória do computador.

- Essa operação de E/S é realizada simultaneamente com a execução de instruções no programa do usuário.
- Quando o dispositivo externo está pronto para ser atendido, o módulo de E/S para o dispositivo externo envia um sinal de requisição de interrupção ao processador.
- O processador responde suspendendo a operação do programa atual, desviando para um programa para atender a esse dispositivo de E/S em particular, conhecido como tratador de interrupção, e retomando a execução original depois que o dispositivo for atendido.

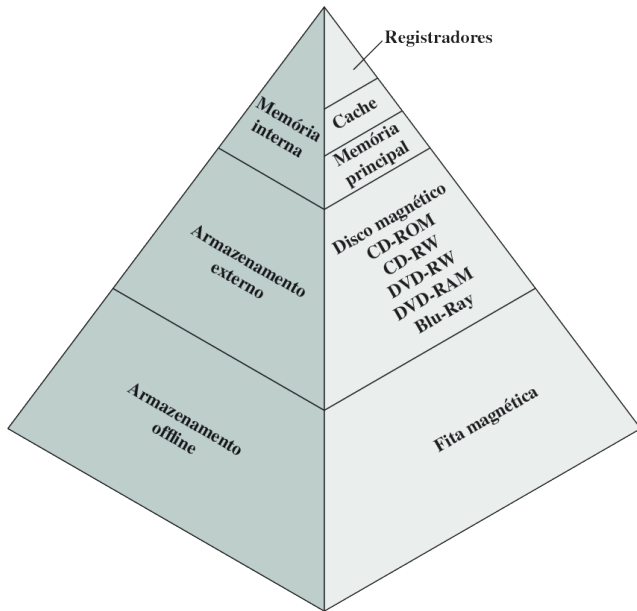
- O termo localização indica se a memória é interna ou externa ao computador.
- Uma característica importante da memória é a sua capacidade.
- Um conceito relacionado é a unidade de transferência.
- Para a memória interna, a unidade de transferência é igual ao número de linhas elétricas que chegam e que saem do módulo de memória.

- Os **métodos de acesso** das unidades de dados **inclui**:
 - **Acesso sequencial**
 - **Acesso direto**
 - **Acesso aleatório**
 - **Associativo**

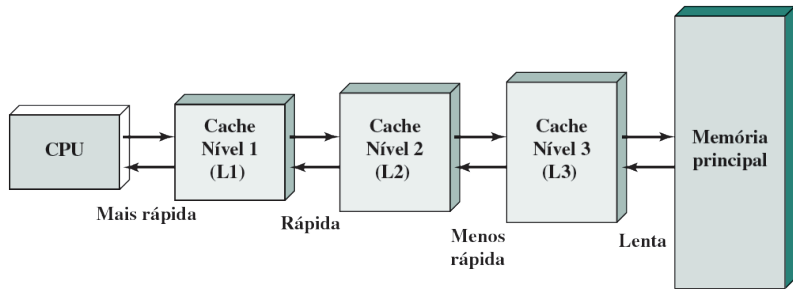
- Os **métodos de acesso** das unidades de dados **inclui**:
 - **Acesso sequencial:**
 - Memória é organizada em unidades de dados chamadas registros;
 - Sendo acessado de forma linear.
 - **Acesso direto:**
 - envolve um mecanismo compartilhado de leitura-escrita;
 - Seus blocos ou registros contém endereço único, baseado no local físico;
 - Acesso é feito por meio de um acesso direto a uma vizinhança genérica do registo, em seguida, por uma pesquisa sequencial.

- Os **métodos de acesso** das unidades de dados **inclui**:
 - **Acesso aleatório:**
 - Cada local endereçável na memória tem um mecanismo de endereçamento exclusivo, fisicamente interligado;
 - Qualquer posição pode ser selecionada de modo aleatório, sendo endereçada e acessada diretamente;
 - **Associativo:**
 - Permite fazer uma comparação de um certo número de bits com uma combinação específica;
 - Uma palavra é buscada na memória com base em uma parte de seu conteúdo, e não de acordo com seu endereço.

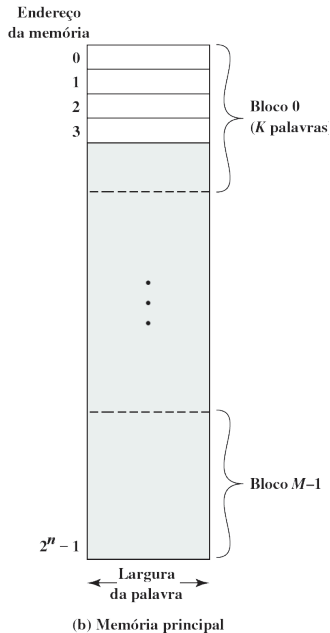
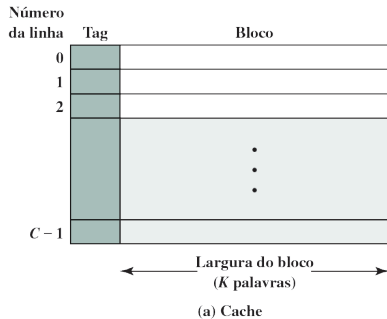
Memória cache



Memória cache



Memória cache



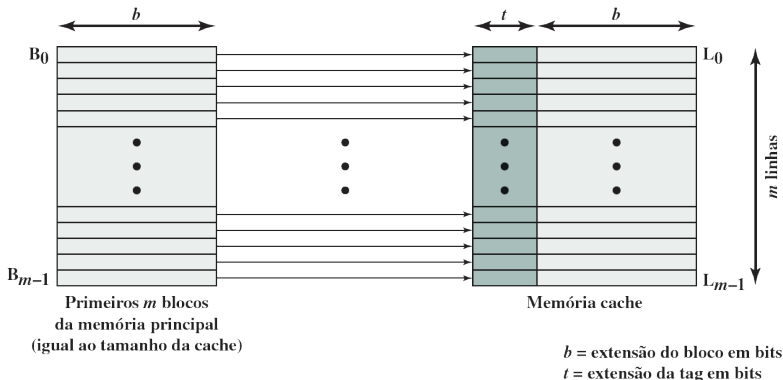
Memória cache

Endereços da cache	Política de escrita
Lógico	<i>Write through</i>
Físico	<i>Write back</i>
Tamanho da memória cache Função de mapeamento	Tamanho da linha Número de caches
Direto	Um ou dois níveis
Associativo	Unificada ou separada
Associativo em conjunto	
Algoritmo de substituição	
Usado menos recentemente (LRU — do inglês, <i>Least Recently Used</i>)	
Primeiro a entrar, primeiro a sair (FIFO — do inglês, <i>First In, First Out</i>)	
Usado menos frequentemente (LFU — do inglês, <i>Least Frequently Used</i>)	
Aleatória	

- É necessário haver um algoritmo para mapear os blocos da memória principal às linhas de cache.
- É preciso haver um meio para determinar qual bloco da memória principal atualmente ocupa uma linha da cache.
- Para função de mapeamento três técnicas podem ser utilizadas:
 - **Direta;**
 - **Associativa;**
 - **Associativa por Conjunto.**

Mapeamento direto

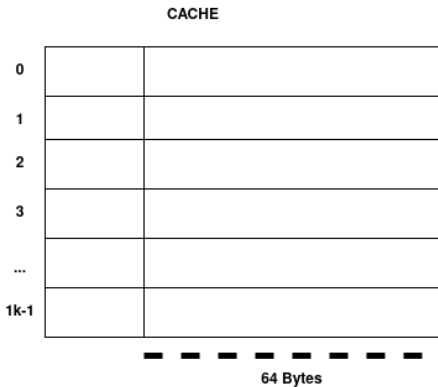
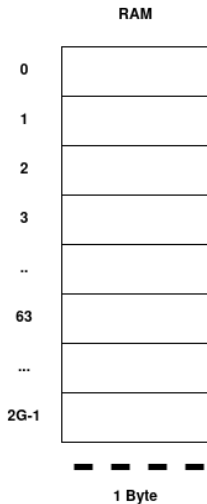
- A técnica mais simples, conhecida como mapeamento direto, mapeia cada bloco da memória principal a apenas uma linha de cache possível.



- Vantagens:
 - Simplicidade e Velocidade.
 - Hardware barato.
 - Procura simples (posição fixa).
- Desvantagens:
 - Pode ter mau aproveitamento das posições da cache.

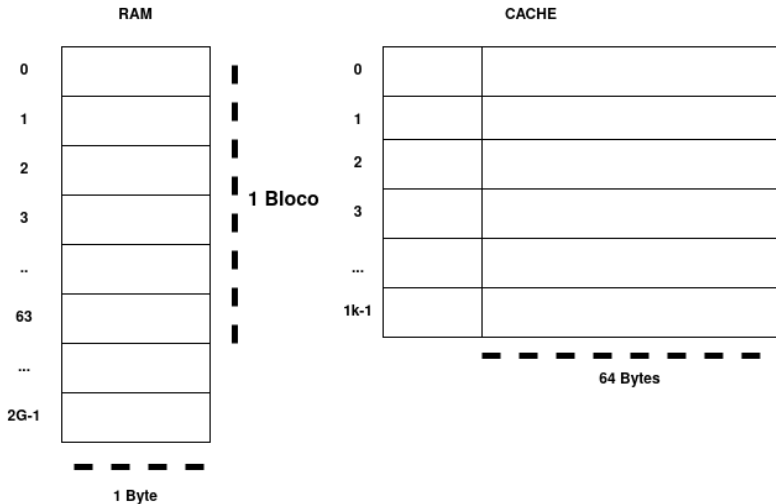
Mapeamento direto

Memória RAM de 2 GB e cada célula 1Byte
Cache de 64 kb sendo 1k linhas



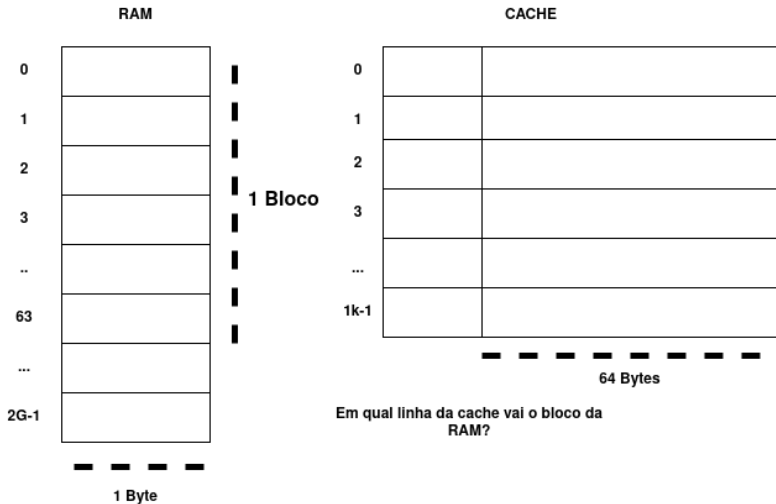
Mapeamento direto

Memória RAM de 2 GB e cada célula 1Byte
Cache de 64 kb sendo 1k linhas



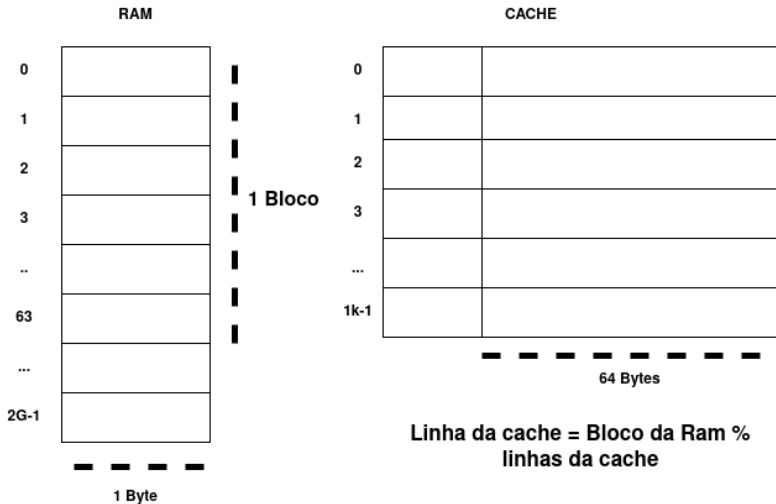
Mapeamento direto

Memória RAM de 2 GB e cada célula 1Byte
Cache de 64 kb sendo 1k linhas



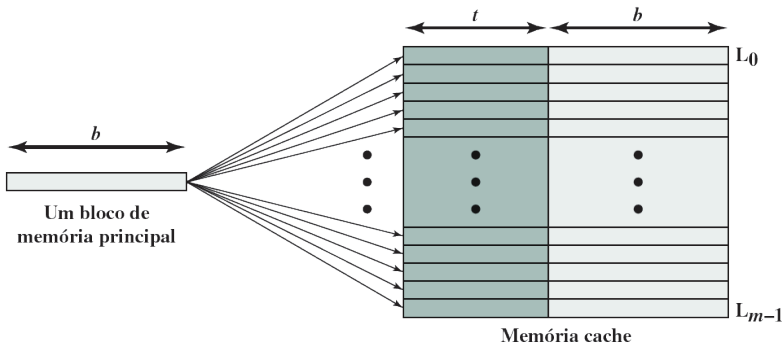
Mapeamento direto

Memória RAM de 2 GB e cada célula 1Byte
Cache de 64 kb sendo 1k linhas



Mapeamento associativo

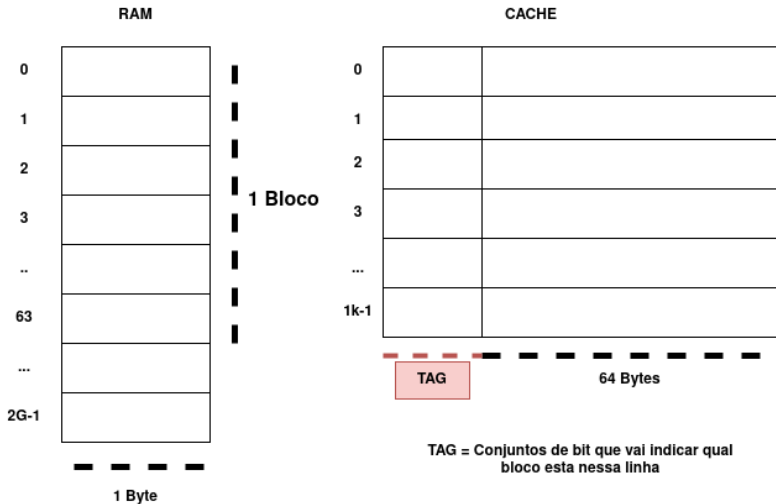
- Permite que cada bloco da memória principal seja carregado em qualquer linha da cache.



- Vantagens:
 - Melhor distribuição da informação na cache.
- Desvantagens:
 - Memória associativa tem alto custo e tamanho limitado.
 - Necessita política de substituição.

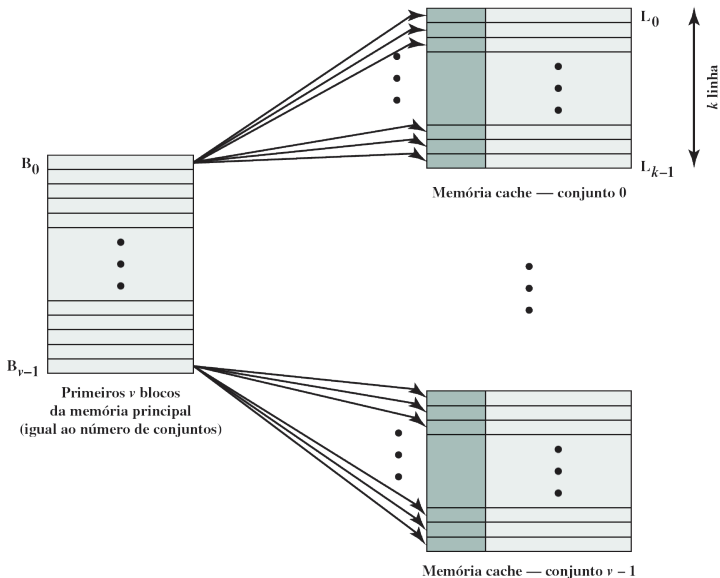
Mapeamento associativo

Memória RAM de 2 GB e cada célula 1Byte
Cache de 64 kb sendo 1k linhas



Mapeamento associativo por Conjunto

- Mapeamento associativo e direto.



Mapeamento associativo por Conjunto

- Vantagens:

- Aumenta tamanho da cache mantendo tamanho da memória associativa.

- Desvantagens:

- Memória associativa tem alto custo e tamanho limitado.
- Somente faz substituição dentro do conjunto.
- Necessita política de substituição.

- Uma vez que a cache esteja cheia, e um novo bloco seja trazido para a cache, um dos blocos existentes precisa ser substituído.
- Para as técnicas associativa e associativa em conjunto, um algoritmo de substituição é necessário.
- Usado menos recentemente (**LRU**).
- Primeiro a entrar, primeiro a sair (**FIFO**).
- Usado menos frequentemente (**LFU**).

- Quando um bloco que está residente na cache estiver para ser substituído, existem dois casos a serem considerados.
 - Se o bloco antigo na cache não tiver sido alterado, ele pode ser substituído por novo bloco sem primeiro atualizar o bloco antigo.
 - Se pelo menos uma operação de escrita tiver sido realizada em uma palavra nessa linha da cache, então a memória principal precisa ser atualizada escrevendo a linha de cache no bloco de memória antes de trazer o novo bloco.
- Diversas políticas de escrita são possíveis, com escolhas econômicas e de desempenho.

- A técnica mais simples é **write through**.
- As operações de escrita são feitas na memória principal e também na cache, garantindo que a memória principal sempre seja válida.
- Na técnica conhecida como **write back**, as atualizações são feitas apenas na cache e somente na modificação do bloco que ele é salvo na memória principal.

- A mais comum é conhecida como **memória de acesso aleatório** (RAM — do inglês, Random Access Memory).
- Uma característica distinta da memória que é designada como RAM é a possibilidade tanto de ler dados como escrever novos dados na memória de um modo fácil e rápido.
- Outra característica distinta da RAM é que ela é volátil.
- A tecnologia RAM é dividida em duas tecnologias:
 - **Dinâmica;**
 - **Estática.**

- **Principais diferenças:**

- **Velocidade:** Mais rápida SRAM, utiliza flip-flops para armazenar dados, não precisando de atualização constante.
- **Tamanho:** A SRAM requer mais espaço físico para armazenamento, utiliza mais componentes.
- **Custo:** A SRAM é mais cara devido a sua complexidade de fabricação e a quantidade de espaço físico que ocupa.
- **DRAM:** Mais adequada para aplicações que exigem alta densidade de armazenamento e são sensíveis ao custo.
- **SRAM:** Mais adequada para aplicações que exigem alta velocidade e baixa latência. Por conseguir acessar dados mais rapidamente que a DRAM porém, com custo maior.

- Uma **memória somente de leitura** (Read-Only Memory - ROM) contém um padrão permanente de dados, que não pode ser mudado.
- Uma ROM é **não volátil**, ou seja, nenhuma fonte de energia é necessária para manter os valores dos bits na memória.
- A **ROM programável (PROM)** é não volátil e pode ser gravada apenas uma vez.
- A **memória somente de leitura programável e apagável (EPROM)** memória apenas de leitura que pode ser programada uma vez e não pode ser apagada eletronicamente (Expor a luz ultravioleta).

- Uma forma mais atraente de memória principalmente de leitura é a **memória somente de leitura programável e apagável eletricamente (EEPROM)**.
- Essa é uma memória principalmente de leitura que pode ser gravada a qualquer momento sem apagar o conteúdo anterior; somente o byte ou os bytes endereçados são atualizados.
- Outra forma de memória de semicondutor é a **memória flash**.
- É intermediária entre a EPROM e a EEPROM tanto no custo quanto na funcionalidade.
- Não volátil e pode ser programada e apagada eletronicamente, armazena grande número de dados em um único chip.

- Uma nova versão da SDRAM, referida como Double-Data-Rate DRAM (DDR-DRAM), proporciona características que aumentam a taxa de dados.
- A DDR alcança taxas mais altas de dados de três maneiras:
 - A transferência de dados é sincronizada tanto na borda de subida como na de descida do clock;
 - A DDR usa frequência de clock mais alta no barramento para aumentar a taxa de transferência;
 - Um esquema de buffering é usado (memória temporária).

- Outra forma de memória semicondutora é a **memória flash**.
- Ela é usada nas aplicações tanto como memória interna como externa.
- Uma característica importante da memória flash é que ela é uma memória permanente, o que significa que ela retém dados quando não há energia aplicada à memória.

- STALLINGS, W. **Arquitetura e Organização de Computadores**. 10 ed. São Paulo: Pearson, 2017;
- TANENBAUM, A. S. **Organização Estruturada de Computadores**. 5 ed. Pearson 2007;
- HENNESY, J. PATTERSON, D. **Organização e Projeto de Computadores**. 3 ed. Editora Campus, 2005.

Obrigado! Dúvidas?

Guilherme Henrique de Souza Nakahata

guilhermenakahata@gmail.com

<https://github.com/GuilhermeNakahata/UNESPAR-2024>