



# Estruturas de dados II

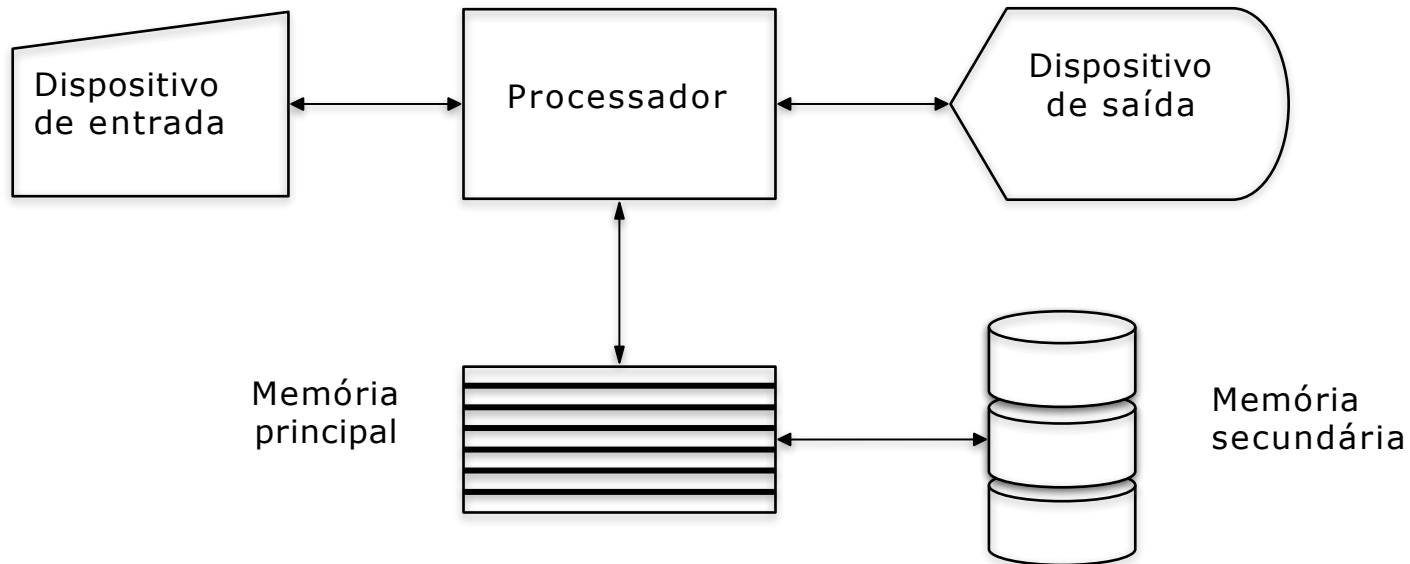
Hierarquia dos Dispositivos de Memória

André Tavares da Silva

[andre.silva@udesc.br](mailto:andre.silva@udesc.br)

# Memória

- Um dos principais componentes de um sistema de computação
- A função da memória é
  - Armazenar informações que são ou serão manipuladas pelo sistema
  - Recuperar informações previamente armazenadas, quando necessário



# Memória

- Aspectos chave da memória
  - Velocidade de acesso
  - Capacidade de armazenamento
- O ideal seria
  - Memória de tamanho ilimitado
  - Memória com um tempo de acesso muito rápido
- Solução
  - Criar uma ilusão para o processador (ou programa) de forma que a memória seja vista como ilimitada e muito rápida

# Princípio da localidade

- Um programa acessa uma parte relativamente pequena do seu espaço de endereçamento em um instante qualquer
- Existem dois tipos de localidade
  - Localidade temporal
    - Se um endereço é referenciado, ele tende a ser referenciado novamente dentro de um espaço de tempo curto
  - Localidade espacial
    - Se um endereço é referenciado, endereços adjacentes dele tendem a ser referenciados em breve

# Memória

- Na prática, a memória não se refere a um único componente
  - Subsistema constituído de vários componentes de diferentes tipos
  - Estes componentes estão interligados seguindo uma hierarquia
- Hierarquia de memória: diferentes tipos de memória interligados
  - Registradores: dispositivos de armazenamento dentro do processador
  - Cache: dispositivos de armazenamento próximo ao processador
  - Memória principal (RAM – Random Access Memory)
  - Memória secundária: dispositivos de armazenamento (SSDs, HDs, DVDs, etc.)

# Memória

- Razões para existência de diferentes tipos de memória
  - Tempo de acesso
    - Velocidade do processador é maior que o tempo de acesso à memória
- Capacidade de armazenamento
  - Necessidade de se armazenar cada vez mais informações
  - Necessidade de armazenamento permanente do dados
- Custo
  - Considerando a relação entre
    - Tempo de acesso, capacidade de armazenamento e preço

# Memória

- Existem duas operações que podem ser realizadas em uma memória
  - Armazenar: guardar um dado na memória
    - Operação de escrita ou gravação (*write*)
  - Recuperar: buscar o dado armazenado na memória
    - Operação de leitura (*read*)
- A unidade de armazenamento na memória principal é chamada célula
  - Célula é um agrupamento de bits tratado em conjunto pelo sistema
  - Usada para efeitos de armazenamento e transferência

# Memória

- Para acessar um dado armazenado na memória, deve ser especificado o endereço da célula correspondente
- Cada célula é identificada por um endereço
  - $N$  endereços: memória organizada sequencialmente entre 0 e  $N - 1$
- Os endereços não são fisicamente gravados na memória
  - Somente os dados são armazenados em cada célula
  - Os endereços são enviados pelo processador para a memória pelo barramento de endereços (BE)



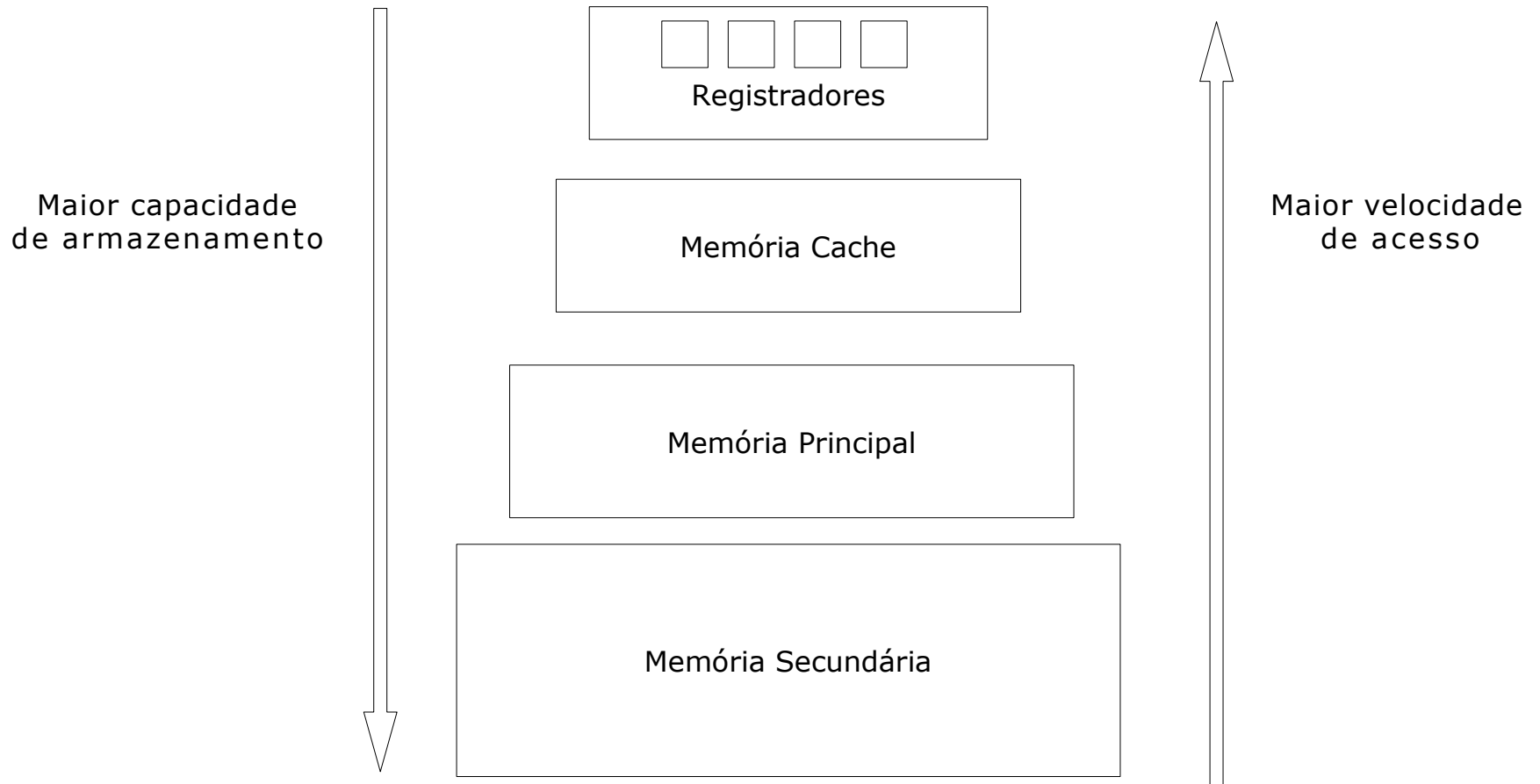
# Memória

- Processadores com endereços de 32 bits
  - Células de 1 byte cada
  - Endereçam até  $2^{32}$  células de 1 byte
- Processadores com endereços de 64 bits
  - Células de 1 byte cada
  - Endereçam até  $2^{64}$  células de um byte

# Operações sobre a memória

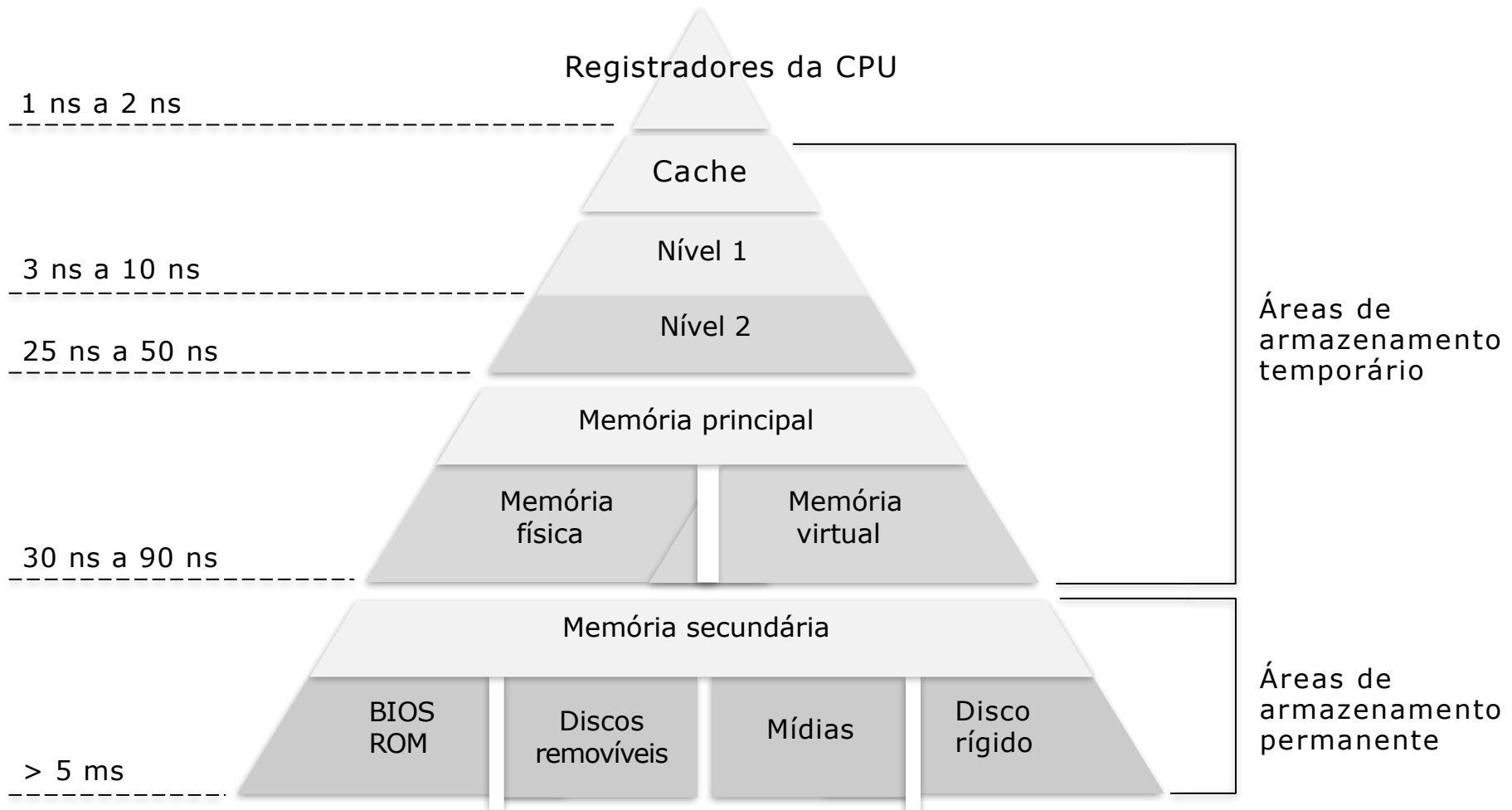
- Operação de escrita é destrutiva
  - Armazena o novo conteúdo sobre o conteúdo anterior
- Operação de leitura não é destrutiva
  - Copia o valor do local de origem, sem modificá-lo

# Hierarquia de memória



# Hierarquia de memória

Tempo de acesso



# Hierarquia de memória

- Cada tipo do componente possui suas tecnologias e características
  - Solid-State Drive (SSD)
  - Double Data Rate (DDR)  
DDR SDRAM (Synchronous Dynamic Random Access Memory)
  - Static Random-Access Memory (SRAM)
  - Dynamic Random-Access Memory (DRAM)
- Para definir a função e as diferenças de cada componente, em geral as seguintes características são analisadas:
  - Tempo de acesso
  - Capacidade
  - Volatilidade
  - Tecnologia de fabricação e custo

# Tempo de acesso

- Tempo decorrido para a memória colocar a informação no barramento de dados depois da posição ter sido endereçada
  - Tempo de acesso para leitura
- O tempo aumenta ao seguir em direção à base da hierarquia
  - Depende da tecnologia de construção e da velocidade de seus circuitos
  - Varia bastante entre os diferentes tipos de memória
  - Poucos ns para memória tipo RAM (dispositivos eletrônicos)
  - Dezenas de *ms* na memória secundária (dispositivos eletromecânicos)

# Capacidade

- Quantidade total de informação a ser armazenada em uma memória
  - Refere-se ao tamanho da memória
- Aumenta ao seguir em direção à base da hierarquia de memória
  - A unidade de medida mais comum é o *byte*

# Volatilidade

- Capacidade de reter ou não a informação quando a energia elétrica é Desligada
- Memória volátil: não retém a informação
  - Exemplo, registradores, memórias cache e principal (tipo RAM)
- Memória não volátil: retém a informação
  - Exemplo, memória tipo ROM (*read only memory*) e memória secundária



# Tecnologias de fabricação e custo

- Memórias de semicondutores (memórias eletrônicas)
  - Registradores, memória cache, principal, ROM e secundária (SSD)
  - São os tipos de memória mais caras
- Memórias de meio magnético
  - Usadas em discos rígidos (*hard disks*)
  - Não-volátil e mais baratas (dispositivos eletromecânicos)
- Memória de meio ótico
  - CDs, DVDs
  - Usa-se um feixe de luz para marcar o valor de cada *bit*
  - Não-volátil e mais baratas

# Registradores

- São internos ao processador
  - Armazenam instruções e dados que estão sendo manipulados pelas operações executadas pelo processador
- Tempo de acesso
  - Possuem o menor tempo de acesso do sistema
    - Por serem construídos com a mesma tecnologia do processador
  - Tempo de acesso equivale a um ciclo de memória (1 a 2 *ns*)
- Capacidade
  - São fabricados para armazenar um dado, instrução ou um endereço

# Memória cache

- Criada para acelerar a velocidade de transferência das informações
  - Entre memória principal e processador
- Podem ser internas ou externas ao processador
  - Cache L1 (primeiro nível) na pastilha
  - Cache L2 (segundo nível) interna
    - Embarcada no mesmo circuito integrado (*on chip*)
  - Cache L3 (terceiro nível) ou L2 externa
    - *Chip* separado instalado na placa mãe
- Tempo de acesso e capacidade
  - Um ciclo de memória (5 a 20 *ns*)
  - Valores típicos: L1 entre 32 e 48KB, L2 até 512KB e L3 até 4MB

# Avaliação da eficiência da memória cache

- Métricas mais comuns
  - *Hit*: dado encontrado no nível procurado
  - *Miss*: dado não encontrado no nível procurado
  - *Hit-rate*: percentual de hits no nível
  - *Miss-rate*: percentual de *misses* no nível (complementar ao *hit-rate*)
  - *Hit-time*: tempo de acesso ao nível
    - Incluindo o tempo para verificar se é *hit* ou *miss*
  - *Miss-penalty*: tempo médio gasto para acesso ao dado não encontrado
    - Tempo para transferir o dado de níveis mais inferiores ao nível desejado

# Memória principal

- Dispositivo onde o programa que vai ser executado é armazenado para que o processador leia as instruções
  - Memória básica de um sistema de computação
- Tempo de acesso
  - Um ciclo de memória (50 a 80 *ns*)
- Custo muito mais baixo em comparação a memória cache

# Memória secundária

- Tempo de acesso
  - HDs: 8 a 30 *ms*
  - CDs e outras mídias com a mesma natureza: 120 a 300 *ms*
- Alta capacidade de armazenamento, tipicamente superior a 1 TB
  - Não-voláteis
  - Memória de semicondutores, meio magnético ou ótico
- Custo bem mais baixo que o da memória principal
  - Alternativas modernas baseadas em sistemas de arquivos distribuídos
    - Amazon S3, Google Drive, Dropbox, dentre outras



# Estruturas de dados II

Hierarquia dos Dispositivos de Memória

André Tavares da Silva

[andre.silva@udesc.br](mailto:andre.silva@udesc.br)