

Organização de Computadores

Apostila 4

Parte 1

2011

CAPÍTULO

4**SUBSISTEMA DE MEMÓRIA****Assuntos**

UD III – Conversão de Bases e Aritmética Computacional

- 4.1 Introdução
- 4.2 Representação das informações
- 4.3 Hierarquia de Memória
- 4.4 Memória Principal
- 4.5 Erros
- 4.6 Memória Cache

Objetivos

Conhecer:

- A representação dos dados na memória
- As operações realizadas na memória
- A hierarquia de memórias
- Os erros oriundos da memória
- A memória cache
- As tecnologias de fabricação de memórias
- Saber como funciona a memória principal

4.1 Introdução

Memórias são dispositivos de **armazenamento** de dados/informações para **manipulação** pelo computador, isto é, são verdadeiros **depósitos** de dados que ficam a disposição do sistema computacional para realizar tarefas de processamento para geração de informação útil aos seus usuários.

Como sistema, existem vários componentes ou tipos de memória que se complementam para essa finalidade, cada uma para atender a fins específicos. O termo subsistema se resume pelo fato de estarmos enxergando o sistema computacional como o sistema maior e as memórias como “seus componentes integrados”. A ideia aqui é a mesma correlata da teoria de conjuntos, onde temos conjuntos e subconjuntos.

Percebam que o subsistema de memórias guardam as mesmas características de sistema, ou seja, conjunto de memórias diferentes que trabalham coordenadas para o mesmo fim – armazenar dados para sua posterior manipulação pelo Sistema Computacional.

Por quê variados tipos de memórias ?

Inicialmente observemos que existe uma diferença acentuada de manipulação de dados refletida pela alta velocidade de execução dos Processadores em contraste ao elevado tempo de acesso às memórias, na transferência de bits entre esses dois componentes do Sistema Computacional.

Isso ocorre particularmente pela diferença de tecnologia aplicada a esses dois componentes e principalmente devido às funcionalidades bem específicas de cada um deles. Deve ser considerado, também, o aumento constante do tamanho dos programas e volume de dados armazenados cada vez maiores.

A necessidade da existência de vários tipos de memória ocorre em virtude de vários fatores concorrentes. Em primeiro lugar, o aumento, sempre crescente, da velocidade das UCP, muito maior que o *tempo de acesso* da memória, ocasiona atrasos (às vezes, intoleráveis) na transferência de bits entre a memória e UCP, e vice-versa.

Outro fator relaciona-se com a capacidade de armazenamento de informações que os sistemas de computação precisam ter, cada vez maior, em face do aumento do tamanho dos programas, bem como do aumento do volume dos dados que devem ser armazenados e manipulados nos sistemas atuais.

Se existisse apenas um tipo de memória, sua velocidade (tempo de acesso) deveria ser compatível com a da UCP, de modo que esta não ficasse esperando muito tempo por um dado que estivesse sendo transferido da memória.

[Mário Monteiro]

Se existisse somente um tipo de memória, o tempo de acesso desta deveria ser compatível com a velocidade das CPU, o que não é uma realidade. Muito pelo contrário, a evolução tecnológica dos processadores, como uso de semicondutores, é muito superior a evolução das memórias. Muitas vezes são empresas fabricantes diferentes, com interesses particulares bem centrados no mercado.

Velocidades: CPU x Memórias

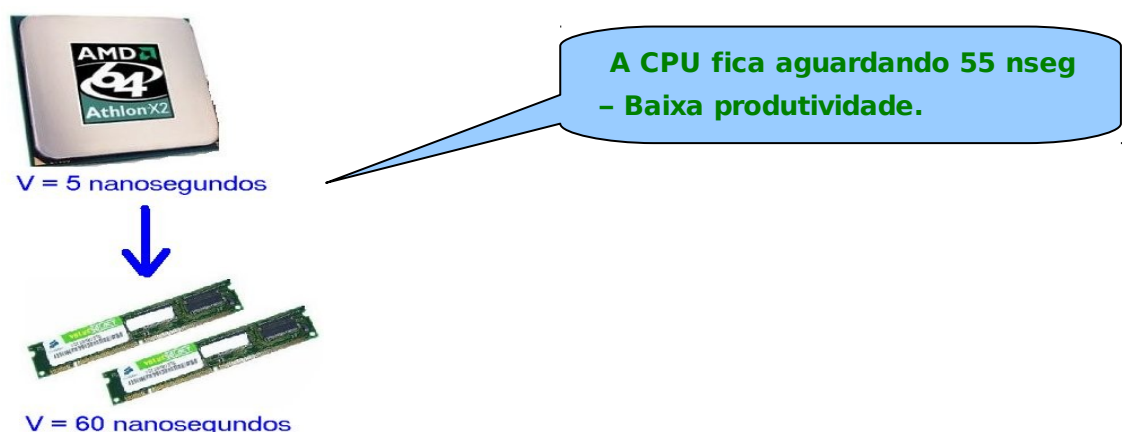


Figura 4.1 – Velocidade entre CPU e Memória

Na figura 4.1 observamos um exemplo de velocidade de Processador comparado como o tempo de acesso a memória. Percebe-se que são bem

diferentes. O processador fica 55 nanosegundos ocioso aguardando uma próxima operação da memória, refletindo desperdício de recurso computacional. Este é um grande problema para ser resolvido.

Como melhorar a velocidade das memórias?

Como podemos ver, há necessidade de se aumentar a velocidade (tempo de acesso) das memórias – Isso ocorre há muito tempo. Entretanto essa evolução gera custos muito elevado (capacidades maiores) para se tentar aproximar a velocidade da memória a do processador.

Como o avanço tecnológico (figura 4.2) a velocidade das CPU dobram a cada 1 ½ ano (atualmente esta relação aumentou consideravelmente) enquanto a velocidade das memórias aumentam 10% ao ano com o mesmo custo de fabricação, muito embora a capacidade das Memórias têm aumento de 400% a cada 3 anos (segundo Monteiro).

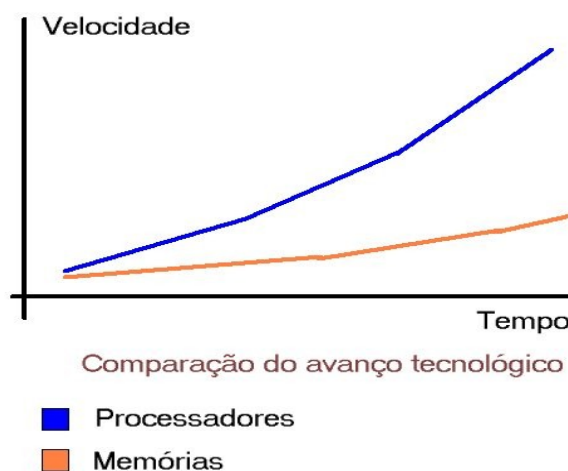


Figura 4.2 – Relação de avanço tecnológico entre CPU e Memórias

O fato é que Memórias de elevada velocidade possuem custos altíssimos, que são repassados ao mercado, além de não possuírem grandes capacidades de armazenamento. Encontra-se então um dilema a ser resolvido pelos projetistas de memórias de computadores.

Como produzir memórias mais rápidas e de maior capacidade de armazenamento a custos menos elevados?

A solução encontrada pelos projetistas e arquitetos de hardware foi a adoção de variados tipos de memórias mesclando custos, capacidades de armazenamentos e velocidades diferenciadas, formando-se uma verdadeira **hierarquia de memórias** como parte de um subsistema de memória a ser utilizado pelos sistemas computacionais.

Memórias funcionam como depósitos similares a um arquivo de fichários (figura 4.3) com várias gavetas, onde cada uma possui um endereço para se obter seu conteúdo – acesso de leitura (**read**) ou para guardar algum conteúdo – acesso de escrita (**write**).



Figura 4.3 – Memórias como Arquivos

Como outros exemplos similares podemos citar as caixas de correio de um prédio, estantes de Bibliotecas ou mesmo um Depósitos de suprimento. Observem que em todos estes exemplos a ideia central é **armazenar** e **disponibilizar** algo.

Em se tratando de memórias eletrônicas, ressalta-se de importância o fato de que sempre que forem energizadas (quando o sistema computacional for ligado), cada célula da memória **sempre possui** um valor expresso em binário – mesmo que este valor não seja útil ou não seja de controle por parte de qualquer processo que esteja ativo no computador. **Este conceito é muito útil na programação.**

4.2 Representação das Informações

Sabemos que o elemento básico de armazenamento físico no hardware é o bit e que pode ser representado como um sinal elétrico, ótico ou magnético para indicar se o bit tem valor 1 ou 0.

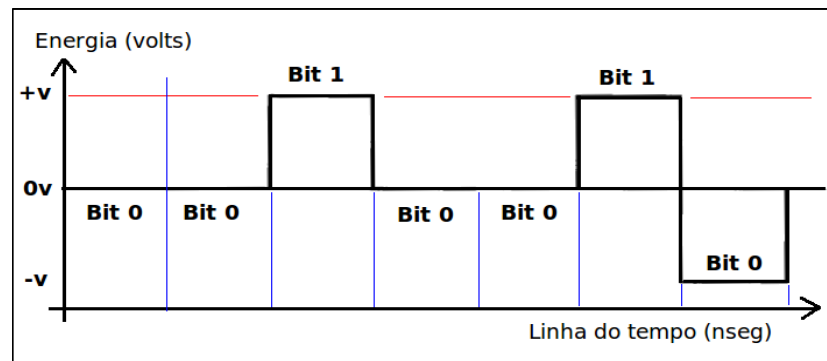
Como exemplo:

- ✓ **Um sinal elétrico** (+v = 1; -v = 0)
- ✓ **Campo magnético**(+ = 1; - = 0)
- ✓ **Presença ou ausência de ponto de luz (queima).**

A memória de um sistema de computação tem como elemento básico de armazenamento físico o *bit*. Ou seja, fisicamente ela é construída de modo a representar individualmente bit por bit (seja com seu valor 0 – zero, seja com seu valor 1 – um). O modo pelo qual cada bit é identificado na memória é variado: pode ser um sinal elétrico, como mostrado na figura 4.4, ou por um campo magnético ou ainda por presença/ausência de um ponto de luz.

Como um bit pode apenas indicar 2 (dois) valores distintos, sua utilidade individual é bastante restrita. Na prática, precisamos passar para o computador as informações que conhecemos na vida cotidiana, normalmente em forma de caracteres ou símbolos gráficos, que visualmente conseguimos distinguir. Assim, claramente distinguimos o caractere “a” do caractere “b”, como também o símbolo matemático “+” do símbolo “(”, porque todos eles têm formato visual diferente e o ser humano, através do sentido da visão, consegue distingui-los (é possível, também, para os humanos separar esses símbolos através do sentido do tato, como se faz em braille, para pessoas cegas).

[Mário Monteiro]

Figura 4.4 – Representação dos Bits 1 e 0¹

Logicamente, 1 bit não representa uma informação útil ao ser humano mas um grupo de 8 bits pode representá-la na memória, por isso adotou-se um padrão de 8 bits (Byte) para representar-se todos os caracteres existentes em nossa linguagem escrita. A IBM chamou este padrão como sendo a Tabela ASCII (*American Standard Code for Information Interchange*).

Fazendo um resgate da **arquitetura de Neumann** onde afirma que os dados são armazenados em memórias, e que por sua vez são divididos em unidades menores que possuem endereço único, a representação de armazenamento em memória segue, assim, esse padrão.

Cada unidade de memória é chamada de **Célula** de Armazenamento que consiste em um grupo de bits como se fosse um único elemento (**inseparável**). Então a manipulação de dados a serem transferidos de/para o processador, ou mesmo entre outros tipos de memórias, é realizada por meio de conteúdos completos de células, não fazendo sentido transferência de metade ou parte de uma célula para transferir dados para outro dispositivo do hardware.

Uma *célula* é, então, um grupo de bits tratado em conjunto pelo sistema, isto é, esse grupo é movido em bloco como se fosse um único elemento, sendo assim identificado para efeito de armazenamento e transferência, como uma unidade. Mais adiante, conforme veremos, o termo *célula* costuma ser utilizado apenas para identificar a unidade de

1 Baseado na figura 5.2 do livro do Mário Monteiro.

armazenamento da memória principal; nos demais tipos de memória, a unidade de armazenamento (grupo de bits que se move junto) possui outras denominações (bloco, setor, cluster, etc.).

[Mário Monteiro]

Pensando-se em memórias eletrônicas, como as Memórias Principais (veremos mais adiante), elas são compostas de milhares de células, já as demais memórias como discos, são subdivididas em blocos, setores, *clusters*, etc.

Como nossa área de interesse inicial trata-se das memórias eletrônicas, podemos então afirmar que elas são constituídas de várias células e cada uma delas possuindo um identificador único, chamado de **ENDEREÇO** de memória.

Estes endereços são definidos pelos fabricantes e normalmente o campo de endereçamento varia de **0** até **N-1**, onde N é o número de grupos (células) que possui a memória.

Em memórias mais antigas, as **Células** possuíam tamanhos variados de grupos fixos de bits como 8, 16, 24, etc. Atualmente as empresas fabricantes resolveram padronizar o tamanho das células de memória como sendo de 8 bits, muito embora esse padrão não seja formalizado em normas internacionais, mas é o que vem se adotando de maneira informal.

Assim, os **Endereços** são expressos em valores que variam de 0 até N-1, sendo seu último endereço dependente do tamanho físico da memória e normalmente são expressos em números Hexadecimais.

Exemplo:	Endereços (hexadecimal)	Células (binário)
	2A	10111001
	2B	01111110
	2C	11111100

4.3 Operações realizadas com Memória

Como nos depósitos ou repositórios, podemos:

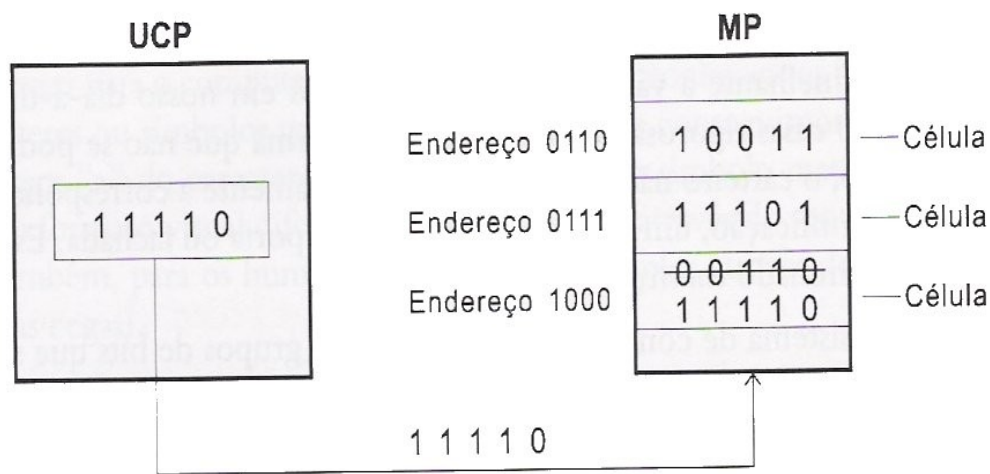
- ✓ “guardar” - **armazenar** coisas/objetos, ou
- ✓ “resgatar” - **recuperar** coisas/objetos

De maneira similar podemos realizar com as memórias já que possuem os mesmos princípios de fundamento de utilização dos depósitos. Assim, existem duas operações que podemos realizar com as memórias por meio dos seus ENDEREÇOS:

- ✓ **Operação Write (Escrita)** – que significa escrever, gravar ou armazenar (*write* ou *record*) dados em células – são **operações destrutivas**, isto é, após o dado novo ser gravado em uma determinada célula não se consegue resgatar o dado anterior que estava armazenado nela (figura 4.5); e
- ✓ **Operação Read (Leitura)** – que significa ler, recuperar (*read* ou *retrieve*) dados que estão armazenados em células – são **operações não destrutivas**, isto é, o dado é copiado da respectiva célula, mantendo seu valor original armazenado (figura 4.6).

Lembrando-se de que as memórias eletrônicas, quando energizadas, todas suas células componentes possuem valores armazenados. Em outras palavras, não ocorre a possibilidade da existência de uma memória estar energizada e qualquer uma de suas células ter ausência de valor armazenado.

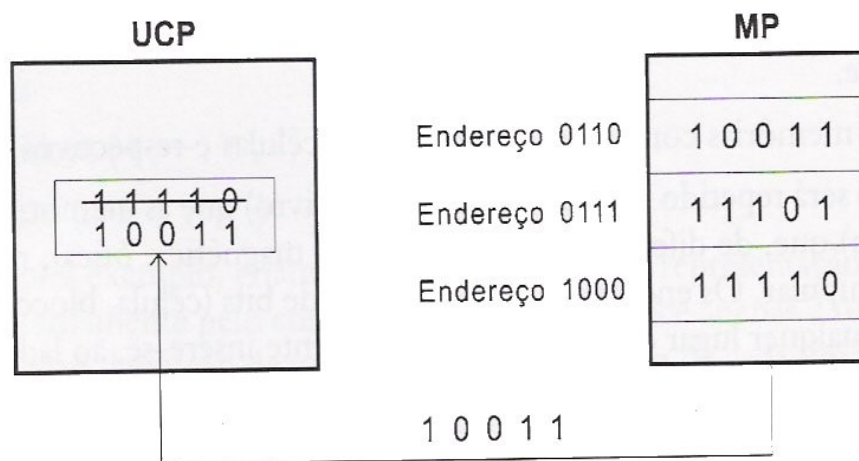
Exemplo de operação de Escrita (Write)



(a) Operação de escrita — O valor 11110 é transferido (uma cópia) da UCP—para a MP e armazenado na célula de endereço 1000, apagando o conteúdo anterior (00110).

Figura 4.5 – Esquema de uma operação de escrita em memória

Exemplo de operação de Leitura (Read)



(b) Operação de leitura — O valor 10011, armazenado no endereço da MP 0110 é transferido (cópia) para a UCP, apagando o valor anterior (11110) e armazenando no mesmo local.

Figura 4.6 – Esquema de uma operação de leitura em memória

Leia atentamente os passos para verificar como se procede cada uma das duas operações com memória.

4.4 Hierarquia de Memórias

As memórias de um sistema computacional formam um verdadeiro subsistema para manipulação de dados a serem utilizados pelo SC, e como tal existem variados tipos de memórias que são organizadas de forma hierárquica e que se interligam de maneira bem estruturada.

O objetivo maior do subsistema de memória é minimizar o tempo de espera do processador, como vimos anteriormente, e assim tornar o sistema mais efetivo e eficiente, com performance cada vez mais robusta e que atenda a evolução tecnológica crescente dos processadores e dos demais dispositivos do Sistema Computacional.

Essa hierarquia pode ser representada conforme indicada na figura 4.7, a seguir.

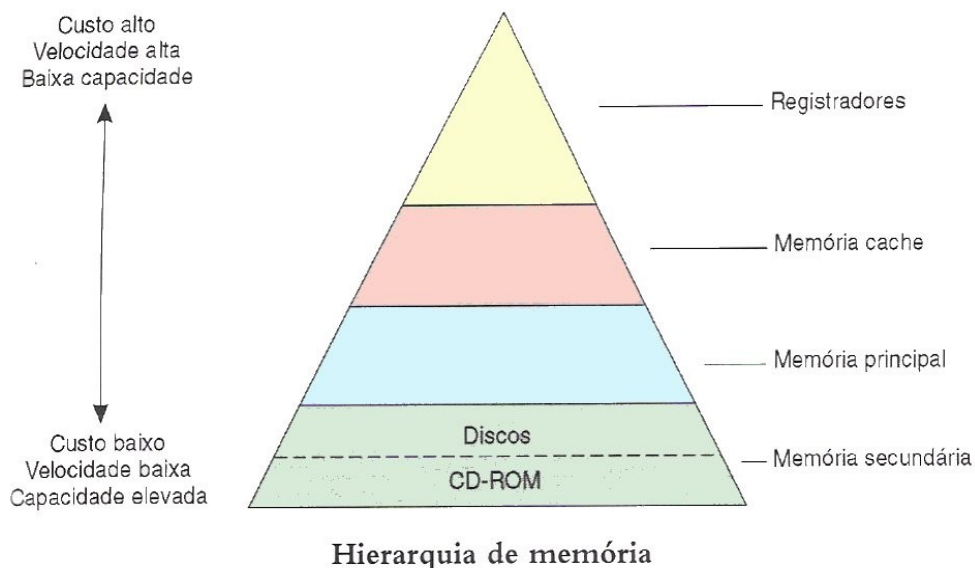


Figura 4.7 – Hierarquia de memória

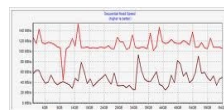
Analisando a figura 4.7 identificamos os quatro tipos de memórias existentes em um Sistema Computacional: os Registradores, Cache, Principal e Secundária.

No lado esquerdo da figura é mostrada a relação de alguns parâmetros

que diferenciam cada um dos tipos de memória da hierarquia do subsistema, e que iremos tratar de cada um deles a seguir.

4.4.1 Parâmetros de cada tipo de Memória

i) Tempo de Acesso



Indica quanto tempo a memória gasta para colocar uma informação na barra de dados após uma determinada posição ter sido endereçada. Isto é, o período de tempo decorrido desde o instante em que foi iniciada a operação de acesso (quando a origem – em geral é a UCP – passa o endereço de acesso para o sistema de memória) até que a informação requerida (instrução ou dado) tenha sido efetivamente transferida. É um dos parâmetros que pode medir o desempenho da memória. Pode ser chamado de *tempo de acesso* para leitura ou simplesmente tempo de leitura.

[Mário Monteiro]

Numa operação de leitura, por exemplo, por meio do barramento de endereço, dar-se o início da operação de acesso (informa à memória o endereço da célula), por meio do barramento de dados ocorre o recebimento dos dados da célula endereçada pelo componente requisitante, normalmente a CPU, a qual recebe uma cópia dos dados existentes na memória.

O tempo de acesso tem influência direta de qual tecnologia de construção foi utilizada para montagem dos seus circuitos internos e da velocidade desses circuitos eletrônicos. Logicamente existe também uma relação direta com o custo de fabricação desses componentes.

Exemplos:

- ✓ Memória Principal (MP): variam de poucos nanossegundos
- ✓ Memória Secundária (MS): variam de dezenas a centenas de milissegundos

Exceto as Memórias Secundárias o tempo de acesso independe do posicionamento físico das células, ou seja, a posição do dado na memória não influi diretamente no tempo de acesso desse dado, o que não ocorre com as memórias secundárias pois envolvem dispositivos eletromecânicos, que apresentam tempos de acessos diferenciados, e que dependendo da posição em que este dado se encontrar na memória secundária, fruto da ação mecânicas que compõem essas memórias, de uma maneira geral, poderá resultar num tempo de acesso ainda maior.

Em resumo:

- ✓ Registradores, Cache e MP: dispositivos eletrônicos
- ✓ Memórias Secundária: são dispositivos eletromecânicos (a maioria)

ii) Ciclo de Memória (*Memory System's Cycle Time*)



Existente somente para memórias eletrônicas (RAM, ROM) onde caracteriza-se pelo tempo entre duas operações sucessivas de escrita ou leitura (R/W).

Face a tecnologia de fabricação (Núcleo Magnético – Dinâmicas), algumas memórias precisam realizar, rotineiramente, processos de sincronização (*Refresh*) para manterem os níveis energéticos de seus componentes eletrônicos em padrões aceitáveis para não resultarem em erros de operação. Isso ocorre particularmente com as memórias do tipo DRAM ou *Dinamic* RAM.

Já as memórias do tipo SRAM ou *Static* RAM, utilizadas pelas memórias

Cache, não usam *refresh* e com isso o Tempo de Acesso passa a ser igual ao Ciclo de Memória.

O *Refresh* (figura 4.8) é como se o dispositivo (memória) estivesse aguardando o pronto para executar uma operação de leitura ou escrita resultando num tempo maior para seu pleno funcionamento. Onde presume-se que durante esse sincronismo resulta em tornar, temporariamente, as memórias que usam essa tecnologia, em estado de espera (*wait*).

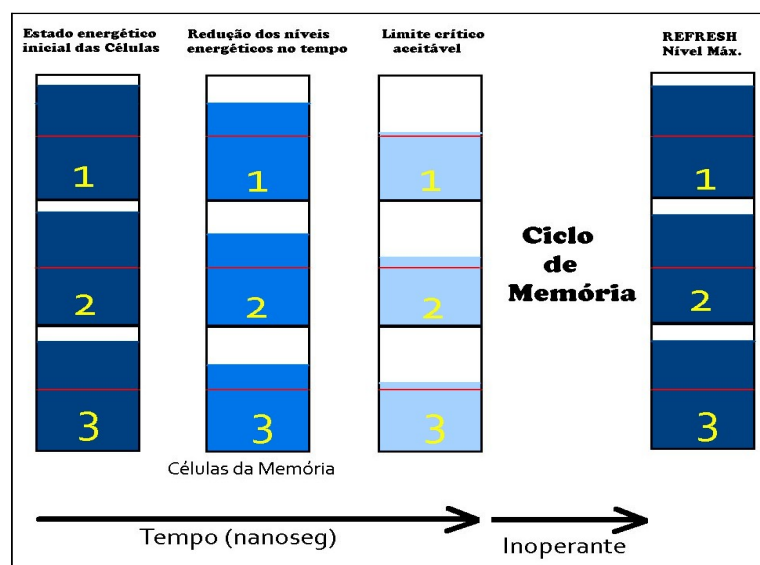


Figura 4.8 – Processo de sincronização (*refresh*) de memórias DRAM

iii) Capacidade



Caracteriza-se pela quantidade de dados que podem ser armazenados na memória. Suas principais unidades de armazenamento são o Byte (MP e algumas MS), Células (MP e Cache), Setor (discos) e Bit (Registradores).

Exemplos:

- ✓ O Registrador tem 32 bits
- ✓ A memória ROM tem 32 KBytes

- ✓ A memória RAM pode endereçar até 128 M células
- ✓ O disco C tem capacidade de 80 Gbytes
- ✓ O DVD tem capacidade de 4 GBytes

iv) Volatilidade



Volatilidade é a propriedade de perda da informação quanto a memória deixa de ser energizada. Portanto só faz sentido para memórias eletrônicas, que funcionam com energia. Assim, Memórias podem ser voláteis ou não voláteis.

- **Memória Volátil** => **perde** os dados após desligado o S. C.
- **Memória não-volátil** => **retém** os dados após desligado S. C.

Como todo computador precisa realizar, na sua inicialização, a execução de softwares básicos² armazenados pelo fabricante para torná-los operacionais, é razoável se pensar que possuem memórias não voláteis para armazenar esses software, sem o qual tornaria impossível a inicialização dos computadores.

Uma vez que um processador nada pode fazer sem instruções que indiquem a próxima operação a ser realizada, é óbvio que todo sistema de computação deve possuir alguma quantidade de memória não-volátil. Isto é, ele deve possuir, pelo menos, algumas instruções armazenadas em memória não-volátil para serem executadas inicialmente, sempre que se ligar o computador.

[Mário Monteiro]

2 Já apresentado no capítulo 1 da primeira apostila.

Exemplos:

- ✓ **Memórias voláteis:** Registradores, RAM (Random Access Memory)
- ✓ **Memórias não-voláteis:** Discos (magnéticos e óticos); Fitas (magnéticas); ROM (Read Only Memory); EPROM (Erasable Programmable ROM) (reutilizável)

v) Tecnologia de Fabricação



Este é um dos pontos que mais interferem na qualidade das memórias, onde são aplicados estudos de materiais, condutibilidade elétrica, resistência e dissipação de calor, e arquitetura de circuitos.

Existem variadas tecnologias empregadas na fabricação de componentes computacionais.

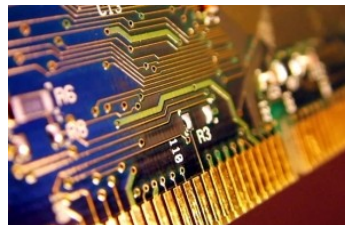


Figura 4.9 – Tecnologia de Computadores

Ao longo do tempo, diversas tecnologias vêm sendo desenvolvidas para a fabricação de memórias. Atualmente, algumas dessas tecnologias já são obsoletas, como as memórias de núcleo de ferrite (magnéticos), *veja a figura 4.10³*, e outras ainda têm uma aplicação comercial ampla, como as memórias de bolha.

[Mário Monteiro]

As primeiras memórias eletrônicas, usado até 1972, utilizavam a

3 Texto, em itálico, incluído pelo autor.

tecnologia com base em núcleo de ferrite magnético (figura a seguir) hoje considerada ultrapassada.

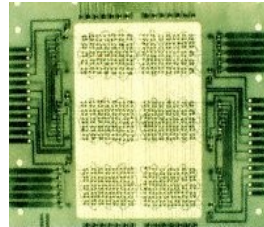


Figura 4.10 – Tecnologia com uso de núcleo de ferrite

A seguir, vamos apresentar aquelas que são mais conhecidas no mercado e em uso pelos sistemas computacionais, até hoje.

a) Semicondutores

Atualmente com o avanço dos estudos e aplicação dos semicondutores é a tecnologia predominante na fabricação de memória e outros componentes internos dos computadores.

Por serem mais rápidas e conseqüentemente relativamente mais caras, existem variadas tecnologias (figura 4.11) que utilizam esse conceito.

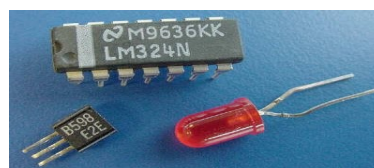


Figura 4.11 – Componentes fabricados com semicondutores

b) Meio Magnético

Memórias que utilizam essa tecnologia são encontrados nos discos rígidos (HD), disquetes, fitas DAT, etc. São memórias não voláteis e que muitas delas possuem mecanismos eletromecânicos resultando em tempos de acessos diversos (diferentes dispositivos) que invariavelmente são bem superiores às memórias eletrônicas.

Em contrapartida possuem grande capacidade de armazenamento e

baixo custo. A figura 4.12 apresenta alguns desses dispositivos e características físicas.



Figura 4.12 – Memórias magnéticas

c) Meio ótico

Encontrado nos dispositivos como CD-ROM, CD-RW, DVD que possuem características que permitem apenas leitura ou ambas operações (R/W).

Possuem tempo de vida mais limitado e seu processo de funcionamento é baseado em feixes de luz para “marcar” 0's e 1's. As figuras 4.13 e 4.14 representam um esquema de seu funcionamento.



Figura 4.13 – Esquema de funcionamento de um DVD/CD (a mídia)

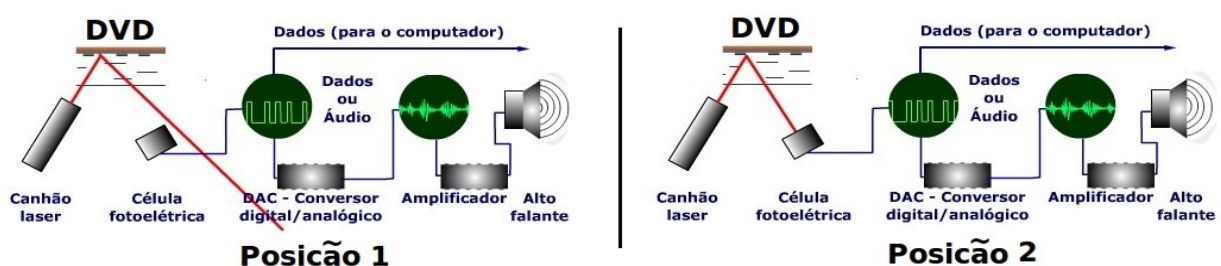


Figura 4.14 – Esquema de funcionamento de um DVD/CD (mecanismo)

vi) Temporariedade



Indica o tempo de permanência dos dados em cada tipo de memória, ou seja, no funcionamento normal do Sistema Computacional cada memória foi projetada para manter armazenado dados por tempos diferentes.

Certas memórias registram os dados por elevada grandeza de tempo ao contrário de outras que não passam de algumas frações de tempo (na ordem de nanosegundos).

Assim, as memórias, quanto a este parâmetro, caracterizam-se de dois tipos bem distintos:

- ✓ **Permanente**: maior duração de tempo de armazenamento (mais capacidade e menor custo) – menos velozes.
- ✓ **Transitória**: menor duração de tempo de armazenamento (menos capacidade e maior custo) – mais velozes.

vii) Custo



Este parâmetro é totalmente dependente da tecnologia de fabricação que envolve a memória e resultando em variados tempos de acesso, ciclo de memória e quantidade de armazenamento, refletido em seu custo final de fabricação e comercialização.

Utiliza-se como medida o preço de bytes armazenados, pois as memórias possuem diferentes tipos.

Exemplos:

- ✓ 1 HD de 80 GB custa R\$130,00 = 0,0000000015 R\$/byte
- ✓ 1 MP de 1 GB custa R\$ 240,00 = 0,00000022 R\$/byte

4.4.1 Níveis Hierárquicos – Detalhes

Outra forma de apresentar com mais detalhes a hierarquia de memórias pode ser vista na figura 4.15, onde de forma alguma invalida a escala em pirâmide vista anteriormente, porém podemos identificar, em cada um dos quatro níveis, alguns de seus dispositivos integrantes.



Figura 4.15 – Hierarquia de Memórias (detalhe)

Serão apresentadas, a seguir, as principais características de cada um dos quatro níveis hierárquicos de memória.

1º Nível: Registradores

É considerada a memória mais rápida encontrada em um SC e possui muito pouca capacidade de armazenamento e, basicamente, são utilizadas no interior dos Processadores.

Algumas delas que iremos abordar em mais detalhes quando estivermos estudando os Processadores são: Registradores de Instrução (RI) e *Programm Counter* (PC), dentre outras.

Assim os registradores são memórias internas das CPU que dão suporte a

armazenamento de dados transitórios manipulados no processamento de informações.

Dentre os parâmetros de memórias já estudados veremos como se enquadra esse tipo de memória.

Tempo de Acesso

Possui tempo de acesso igual ao ciclo de memória, isto é, não implementam as funcionalidades de *refresh* como ocorre com as memórias dinâmicas, e por isso, são extremamente rápidos, de 1 a 5 nanosegundos.

Capacidade

Como são componentes internos aos processadores, necessitam armazenar poucas informações, normalmente armazenam um único **dado**, **instrução** ou **endereço** e que normalmente possuem o tamanho da palavra.

Na verdade, alguns registradores especiais possuem capacidade de armazenamento maiores que isso, mas são exceções.

Tecnologia

São memórias baseadas na tecnologia de semicondutores, a mesma utilizada na construção dos Processadores.

Apesar disso, podem variar um pouco na tecnologia de semicondutores, com diferentes tempo de acesso, custos e capacidades. É o que ocorre no mercado quando se questiona qual processador mais rápido, se o AMD ou Pentium. Fatalmente usam tecnologias diferenciadas no uso de semicondutores.

Em linha geral a mais comum são os bipolar e MOS (Metal Oxide Semicondutor).

Temporalidade

Armazenam dados por muito pouco tempo. Deve-se ressaltar que a escala de grandeza que estamos nos referindo aqui é da ordem de nanosegundos (ns) que equivale a 10^{-9} do segundo, portanto extremamente pequeno.

Assim, quando nos referirmos a outras memórias serem mais lentas, estaremos comparando com essa ordem de grandeza. Exceção de faz às memórias secundárias, como veremos mais a frente. Os registradores normalmente retém dados por alguns nanossegundos (6 a 10 ns).

Como a CPU está constantemente requisitando dados da Memória Principal (MP), os registradores estão constantemente trocando de valores, em quantidades de 8 a 64 bits, por operações realizadas.

Volatilidade

São consideradas memórias voláteis, isto é, ao se desligar o sistema computacional elas perdem os valores armazenados.

Custo

Como usam tecnologias avançadas, de última geração, resultam em custos mais elevados (topo da pirâmide).

2º Nível: Memória Cache

Deve-se ressaltar que em sistemas mais antigos não possuíam esse tipo de memória. Os Registradores eram ligados diretamente à Memória Principal.

As tecnologias em Processadores tiveram e ainda hoje têm maior evolução que a das Memórias resultando em diferenças consideráveis de velocidade desses dois componentes.

Sua necessidade de surgimento ocorreu em virtude dessa grande diferença de velocidade existente entre o processador e a memória principal, que eram bastante afetados pelo ciclo de memória (lentos) das MP, necessitando haver uma memória intermediária, com maior velocidade, isto é, menor tempo de acesso, para minorar o tempo de espera do processador ao requisitar dados da memória, para fins de processamento.

Assim surgiram as memórias Cache, que estão no segundo nível da hierarquia da pirâmide, com objetivo de melhorar a performance do sistema

computacional deixando o processador menos ocioso, na execução de suas funcionalidades.

Possuem uma arquitetura de dois ou três níveis:

- 1º Nível é chamado de L1 (Level 1) – interna ao Processador, algumas são acopladas externamente a eles (Cache Principal) – a mais veloz entre os tipos de caches.
- 2º Nível é chamado de L2, Cache Externa ou Cache Secundária – instaladas (*Chip*) diretamente na placa-mãe (*motherboard*) do sistema (figura 4.16).

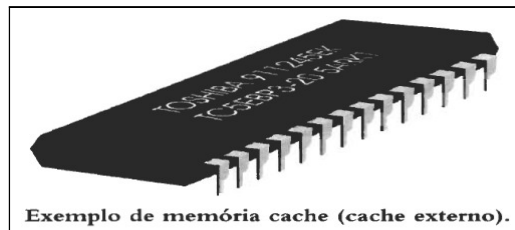


Figura 4.16 – Memória Cache L2

A função das memórias Cache

Acelerar a velocidade de transferência dos dados entre CPU e MP, aumentando seu desempenho. A aplicação deste conceito foi estendido para outros componentes internos do computador como nas Placas de vídeo, Discos, etc.

As Caches também são chamadas de RAM Cache - Random Access Memory (RAM) Cache.

Quanto aos parâmetros de memória as Caches caracterizam-se conforme descrito a seguir.

Tempo de Acesso – Ciclo de Memória

Usam semicondutores e são estáticas (SRAM). Possuem tempo de acesso entre 5 a 7 nanosegundos, portanto maiores que os dos registradores.

Sendo as memórias de semicondutores, fabricadas com tecnologia e recursos para prover menos ciclos de memória do que as memórias RAM comuns (memória principal do tipo dinâmica), elas possuem velocidade de transferência tal que lhes garante tempos de acesso menores que 5 a 7 ns (nanosegundos), sendo por esta razão situadas, na pirâmide, logo abaixo dos registradores.

[Mário Monteiro]

Capacidade

Oscilam entre: 64 KB a 2 MB para as L2 e 16 KB ou mais para as L1.

Tendo em vista que a UCP acessa primeiramente a memória cache, para buscar a informação requerida (a próxima instrução ou dados requeridos pela instrução em execução), é importante que a referida memória tenha capacidade adequada para armazenar uma apreciável quantidade de informações, uma vez que, se essas informações não foram encontradas na cache, então o sistema deverá sofrer um atraso para que as mesmas sejam transferidas da memória principal para a cache.

[Mário Monteiro]

Volatilidade

São memórias voláteis (requerem energia).

Tecnologia

Usam tecnologia avançada e em geral são estáticas – Static Random Access Memory (**SRAM**).

Temporariedade

Permanecem pouco tempo – todos os programas a utilizam e têm pouca capacidade.

O tempo de permanência de uma instrução ou dado nas memórias cache é relativamente pequeno, menor que a duração da execução do programa ao qual a referida instrução ou dado pertence. Isto porque, devido a seu tamanho não ser grande e ser utilizada por todos os programas em execução, há necessidade de alteração periódica da informação armazenada para permitir a entrada de novas informações. Embora a transitoriedade das informações na cache seja uma realidade, o período efetivo de permanência de um dado ou instrução é dependente do tipo de política de substituição de informação na cache.

[Mário Monteiro]

Custo

Alto custo. Pouco menor que dos registradores (R\$/Byte). As Cache L1 são mais caras que as L2.

3º Nível: Memória Principal

É a memória básica de um Sistema Computacional e segue a arquitetura do sistema de von Neumann (instruções e dados armazenados). A CPU busca as instruções e dados armazenados nela.

A parte 2 desta apostila tratará exclusivamente sobre a memória principal, no entanto alguns pontos básicos serão adiantados aqui.

Uma das principais características definidas no projeto de arquitetura de von Neumann, o qual se constitui na primeira geração dos computadores, consistia no fato de ser uma máquina “de programa armazenado”. O fato de as instruções, uma após a outra, poderem ser imediatamente acessadas pela UCP é que garante o automatismo do sistema e aumenta a velocidade de execução dos programas (por se tratar de uma máquina executando ações sucessivas, sem intervalos e em cansar, como não acontece com os seres humanos).

[Mário Monteiro]

Tempo de Acesso – Ciclo de Memória

A memória principal é construída com elementos cuja velocidade operacional se situa abaixo das memórias cache, embora sejam muito mais rápidas que a memória secundária. Nas gerações iniciais de computadores o tipo mais comum de memória principal era uma matriz de pequenos núcleos magnéticos, os quais armazenavam o valor 1 ou o valor 0 de bit, conforme a adição do campo magnético armazenado. Essas memórias possuíam baixa velocidade, a qual foi substancialmente elevada com o surgimento das memórias de semicondutores. Atualmente, as memórias desse tipo possuem tempo de acesso entre 7 e 15 ns.

[Mário Monteiro]

Capacidade

Maiores que as Cache, 128 MB a 4 GB (uso associado aumenta ainda mais sua capacidade).

Volatilidade

São voláteis, mas possuem uma parte (ROM) não volátil, para serem reconhecidas ao ligar o computador.

Tecnologia

As primeiras eram de núcleo de ferrite e atualmente usam semicondutores e são DINÂMICAS ou Dinamic Random Access Memory (**DRAM**).

Temporariedade

É variável e normalmente armazenam as instruções. Deve-se levar em consideração a quantidade de programas e seu tamanho (em execução).

São mais duradouras que as Cache e Registradores, isto é, os dados e instruções armazenados duram mais tempo na memória.

Para que um programa seja executado é necessário que ele esteja armazenado na memória principal juntamente com seus dados. Atualmente esta afirmação é parcialmente verdadeira, visto que não é mais necessário que o programa completo (todas as instruções) esteja na MP. Neste caso, é obrigatório apenas o armazenamento, na MP, da instrução que será acessada pela UCP (na prática, não se usa somente a *instrução* que será executada, mas sim esta e um grupo de outras). Não importa, contudo, se é o programa todo, ou parte dele, que deve estar armazenado na MP para ser utilizado pela UCP. Fica claro que, em qualquer circunstância, as instruções e os dados permanecem temporariamente na MP, enquanto durar a execução do programa (ou até menos tempo). Esta temporariedade é bastante variável, dependendo de diversas circunstâncias, como, por exemplo, o tamanho do programa e sua duração, a quantidade de programas que estão sendo processados juntos, e outras mais. No entanto, a transitoriedade com que as informações permanecem armazenadas na MP é, em geral, mais duradoura que na memória cache ou nos registradores, embora mais lenta que na memória secundária.

[Mário Monteiro]

Custo

São mais baratas que as Cache, tornando-se mais acessível para se ter um computador com maior capacidade de Memória Principal.

4º Nível: Memória Secundária

Também chamada de memória de massa ou auxiliar e que garantem um armazenamento mais **PERMANENTE**.

Possuem diferentes tipos: CDRom, DVD, PenDrive, Fita DAT, HD, etc. A grande característica é sua **NÃO-VOLATILIDADE**, além de serem as mais

baratas e têm maior tempo de acesso.

Usam tecnologias variadas e diferentes capacidades de armazenamento, incluindo-se as memórias flash, como os pen-drivers, tão difundidos atualmente.

A indústria de eletrônicos, em geral, acompanha as tendências de mercado na atualidade onde o princípio da convergência digital⁴ vem sendo o filão de mercado, com i-Pod, i-Fones, smart-fones, telefones IP e todos os i's que ainda estão por vir, inclusive com consideráveis capacidade de armazenamento de dados.

REFERÊNCIAS

MONTEIRO, M. A. Introdução a Organização de Computadores, 2004.

Sites da Internet.

4 Convergência digital é uma integração de mídias que se convergem para interagir em um único ambiente. Telefone móvel, televisão, rádio, jornal são programados para interagir e transmitidos em um único canal, gerando um comunicação multicanal. TV, rádio, celular, internet vão passar a interagir de uma forma não-linear em um único ambiente, sem que o usuário tenha que migrar de uma mídia para outra. Essas mídias e canais vão passar a transmitir seu conteúdo de forma integrada. (Wikipédia).