

UNIDADE 3 – FUNCIONAMENTO E COMPONENTES BÁSICOS

MÓDULO 1 – ARMAZENAMENTO DE INFORMAÇÕES

01

1- CIRCUITOS DE ARMAZENAMENTO

Circuito de armazenamento é a parte dos circuitos digitais que se responsabiliza pelo armazenamento de informações.

Quando falamos em circuitos de armazenamento precisamos entender que, nos sistemas digitais, em especial os computadores, as informações são representadas por meio de conjuntos de binários chamados de “PALAVRAS”. Atualmente, nos computadores, geralmente a PALAVRA tem o tamanho de 32, 64 ou 128 bits. Entretanto, tempos atrás, as estações de trabalho (computadores) usavam apenas 8 e 16 bits.

É natural que um sistema digital seja projetado para trabalhar utilizando determinado tamanho de palavra, sendo assim, deve ter os recursos de hardware que suportem simultaneamente armazenar e processar conjunto de “n” bits (lembre-se: “n” corresponde ao tamanho da palavra).

No desenvolver do estudo deste módulo, veremos os circuitos digitais responsáveis pelo armazenamento de informação, sendo eles os **registradores** e as **memórias** ROM e RAM. Dentre estes circuitos, há aqueles construídos de tal forma que conseguem manipular a informação armazenada. Dentre as operações possíveis, estão os deslocamentos (à direita e à esquerda), o incremento e o decremento.

Incremento e decremento

São operações aritméticas. Sendo que o incremento – aumenta “um” no operando (contador) e decremento – diminui “um” no operando. Um exemplo de operação de incremento seria: $k = k + 1$. Seguindo o mesmo raciocínio, o decremento é: $K = K - 1$. No desenvolvimento de programas para computadores é muito usado incremento e decremento nas estruturas de repetição na variável de controle. Um exemplo relacionado a manipulação de informações nos circuitos de armazenamento, podemos imaginar um circuito contador (Contadores digitais são circuitos implementados a partir de flip-flops) que possui uma conexão específica para cada flip-flop vizinho, de tal forma que os flip-flop de maior ordem é o responsável pela ordem de “incremento” de seu vizinho de menor ordem.

02

2 – REGISTRADORES

O registrador é um circuito digital que tem em sua formação “n” flip-flops, que de forma simultânea, consegue armazenar (de forma independente) “n” bits.

O registrador é considerado um elemento básico de armazenamento em um processador, podendo

variar de três até chegar a algumas dezenas. A localização de registradores dentro do processador tem por intuito **acelerar** o processamento, pois a manipulação dos dados acontece próximo dos recursos de processamento, reduzindo a necessidade de acessar a memória principal.

Os registradores são circuitos digitais capazes de armazenar e deslocar informações binárias, e são tipicamente usados como um dispositivo de armazenamento temporário. São empregados na execução de programas de computadores, disponibilizando um local para armazenar dados. Na maioria dos computadores modernos, quando da execução das instruções de um programa, os dados são movidos da memória principal para os registradores. Então, as instruções que utilizam estes dados são executadas pelo processador e, finalmente, os dados são movidos de volta para a memória principal.

O termo utilizado para armazenar dados em um registrador é **carregar**, embora as palavras escrever e armazenar também sejam usadas. A ação oposta a carregar um registrador é conhecida como **ler os conteúdos** de um registrador. A leitura consiste simplesmente em se conectar às saídas do registrador. Assim percebe-se que a leitura não está sincronizada com o relógio, além disso, não remove os bits do registrador nem os modifica de nenhum modo.

03

A figura abaixo mostra um registrador de quatro bits, com carga paralela, utilizando flip-flops. A letra D representa os flip-flops.

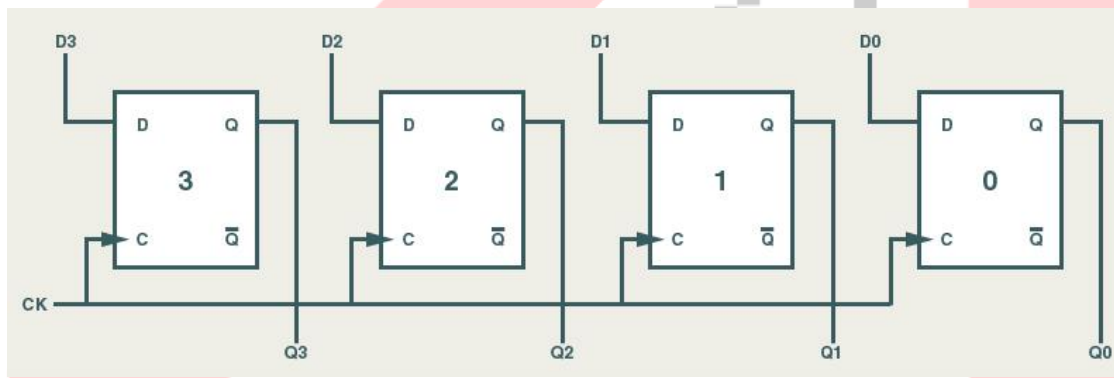


Figura 01

Observe a imagem acima e veja que cada flip-flop é o responsável por armazenar um bit, observando a notação posicional, e cada bit possui um caminho independente do restante, tanto na entrada como na saída. É por este motivo que o registrador é chamado de carga paralela. Continuando a observação da figura, vemos que o flip-flop, que tem por índice 3, armazena o bit mais significativo de uma palavra de quatro bits. O índice 0 (zero) armazena o bit menos significativo.

Podemos dizer que um registrador tem um funcionamento similar a uma barreira. Vemos na Figura 01 que os dados que estão nas entradas D0, D1, D1 e D3 serão somente copiados caso o sinal CK (relógio) passe por uma borda ascendente. As cópias dos valores feitas na passagem de uma borda ascendente ficarão armazenadas nos flip-flops até que haja uma ocorrência da próxima borda ascendente. Desta forma o registrador fica imune a uma possível mudança indesejada dos sinais. Estes sinais são

representados por D0, D1, D2 e D3. É na saída Q onde está o valor armazenado em um flip-flop qualquer. Diante desta característica que um dado que estiver armazenado no registrador poderá ser consultado por outro hardware em qualquer momento, mas, para isso acontecer, é necessário haver um caminho, um conjunto de fios, algo físico que estabeleça comunicação entre a saída do registrador e a entrada do outro dispositivo. O outro dispositivo, por exemplo, pode ser um somador/subtrator.

Carga paralela

É a carga simultânea de todos os bits, por meio da sincronização feita pelo clock (Nos circuitos digitais síncronos, o clock é um sinal usado para coordenar as ações de dois ou mais circuitos eletrônicos).

Notação posicional

É um modo de representação numérica na qual o valor de cada algarismo depende da sua posição relativa na composição do número.

04

No registrador exemplificado na Figura 01 há uma deficiência importante: toda vez que o sinal de relógio CK passar por uma borda ascendente, os valores das entradas D0, D1, D2 e D3 serão copiados, mesmo não sendo o objetivo desejado. Contudo, cabe à Unidade de Controle coordenar e ordenar os sinais dos circuitos de um computador. Na verdade, **o sinal de CK (relógio) tem por função determinar o momento no qual uma ordem deverá ser obedecida**. Sendo assim, um registrador de uma máquina (computador) deve possuir os recursos necessários para a realização da carga do dado quando o relógio passar pela borda ativa apenas se o sinal de “load” (carga) estiver ligado. Na figura abaixo é possível observar um registrador de quatro bits com sinal de carga e carga paralela.

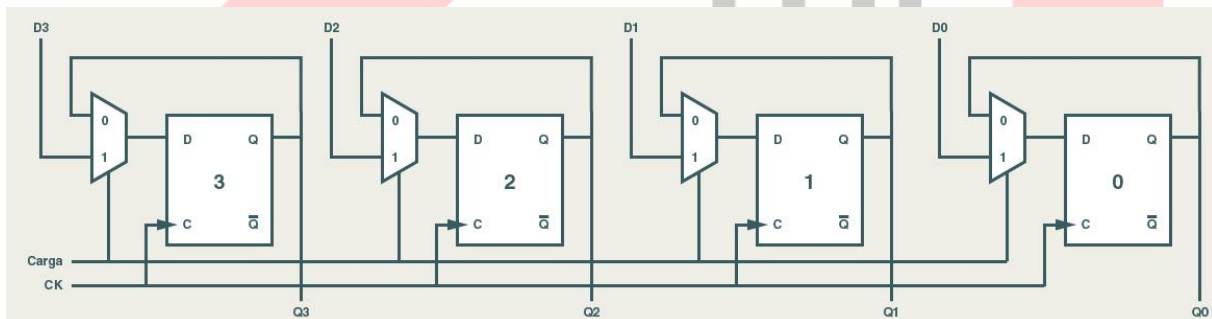


Figura 02

05

Os desenhos mais complexos de circuitos utilizam preferencialmente o símbolo utilizado na figura abaixo (Figura 03). Isso é feito para facilitar a compreensão e evitando a complexidade de um esquema completo de um circuito.

A figura abaixo simboliza um registrador de 4 bits. Ele possui uma carga paralela e sinal de carga (o sinal de carga é possível visualizar observando a figura 02 e também a figura 03). O sinal de carga é a representação da passagem de corrente (carga), que no caso da memória é um dado (parte de uma informação).

Cabe relembrar a importância do conceito de clock (relógio): é ele o responsável de zelar para que os microprocessadores trabalhem em um rígido padrão de tempo. Cabe ao clock gerar pulsos (sinais) em intervalos regulares. A cada sinal (pulso) uma ou mais instruções internas são realizadas. O nome deste registrador é “Reg”.

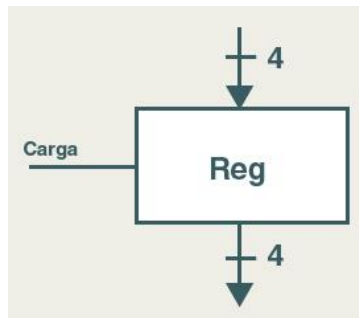


Figura 03

06

• Registrador com Carga Paralela

Na figura abaixo vemos a forma de onda da saída Q para o circuito à esquerda da figura. Este exemplo ilustra como é o funcionamento do circuito associado a um bit de determinado registrador de carga paralela que também está com sinal de carga.

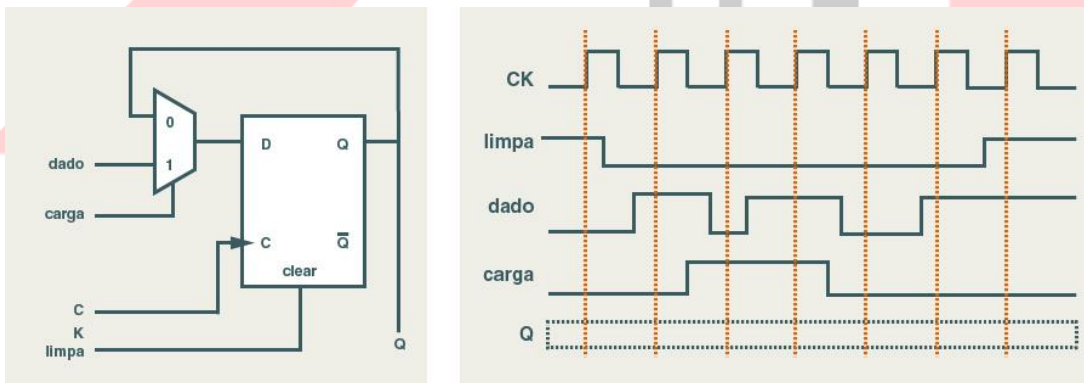
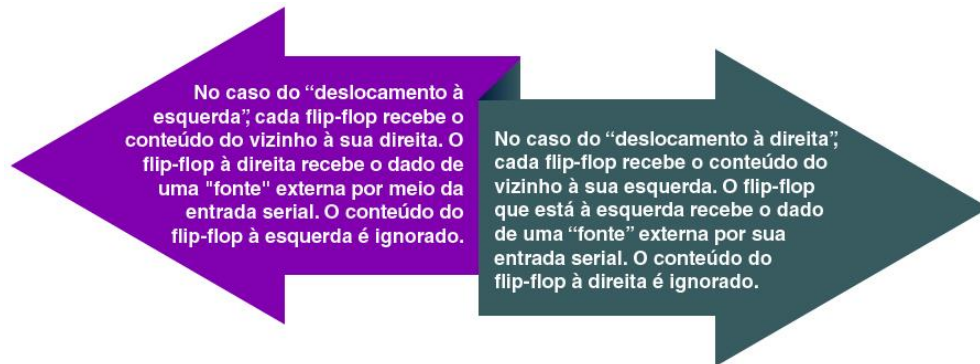


Figura 04

07

• Registradores de Deslocamento (*shift registers*)

O deslocamento de bits é uma operação importante na aritmética binária. Essa operação é basicamente o deslocamento do conteúdo de um flip-flop para o seu adjacente. O sentido da operação pode ser da esquerda para a direita, que é chamado **deslocamento à direita**. Caso o sentido da operação seja da direita para a esquerda é chamado de **deslocamento à esquerda**.



A importância das operações de deslocamento dos bits está no fato destas viabilizarem, entre outras, a realização de um grande número de operações lógicas e aritmética sem um sistema digital.

08

Na Figura 05 vemos um registrador de deslocamento à direita de 4 bits. Notamos que existe uma ligação entre a saída de cada flip-flop e a entrada do vizinho à direita. Este registrador não possui sinal de carga. Contudo, geralmente, este sinal existe.

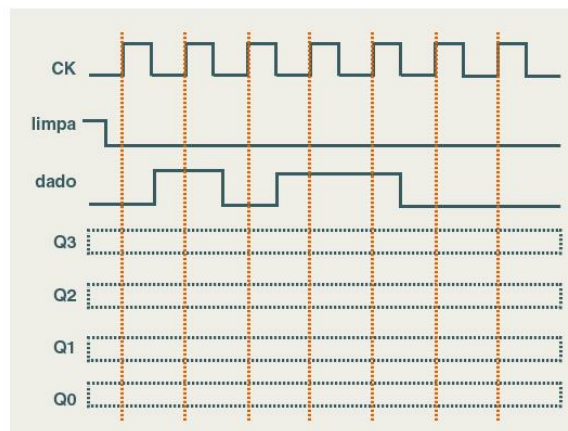
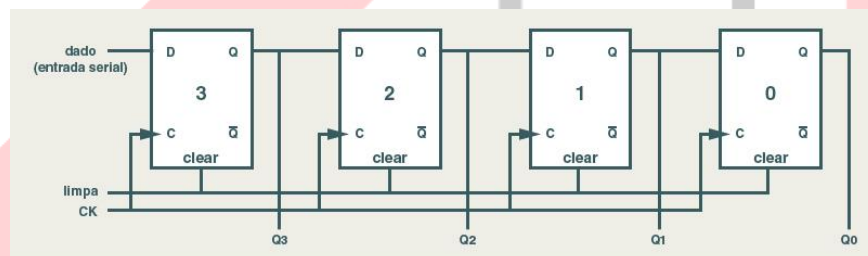


Figura 05

Na Figura 06 temos um registrador de “deslocamento à esquerda” que apresenta uma ligação entre a saída de cada flip-flop e a entrada do flip-flop logo a sua esquerda. Observe que a entrada serial está conectada ao flip-flop à direita que armazena o bit menos significativo.

Você pode ver na Figura 06 um exemplo de um registrador de deslocamento à esquerda de 4 bits com reset assíncrono.

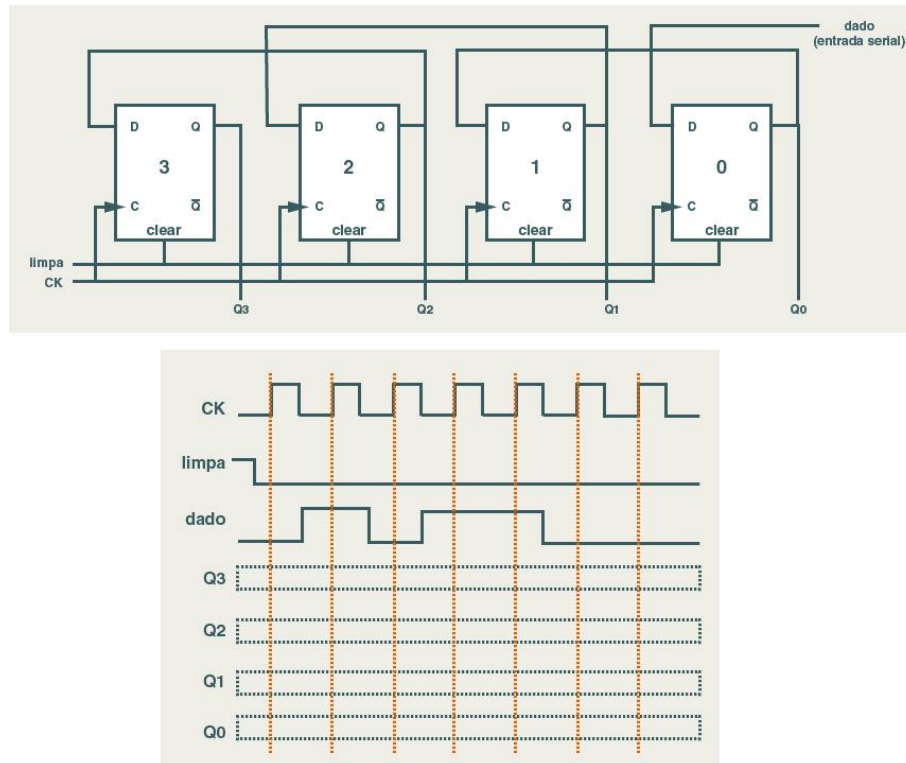


Figura 06

Reset assíncrono

Quando um flip-flop inicia suas atividades o valor de “Q” não é conhecido, podendo ter o valor 1 (um) ou valor 0 (zero). Valor 1 ativo (enable); valor 0 desativado (disable). Desta forma os flip-flops precisam ter uma forma de determinar o valor inicial de sua saída Q. Esta forma é o reset. O sinal de iniciação é chamado de sinal de reset podendo iniciar com valor um (1) ou zero (0), podendo ser síncrono ou assíncrono. O reset síncrono depende do clock sendo acionado quando o sinal do clock é quem inicia o processo. O reset assíncrono não depende do relógio (clock) sendo acionado quando estiver ativo.

- Registrador de deslocamento com sinal de carga paralela

O que torna um registrador com grau de utilidade alto é que além de possibilitar a carga paralela através de sinal de carga, também possibilita deslocamentos tanto à direita como à esquerda. Tudo isso se torna

possível em virtude de na entrada de cada flip-flop haver um seletor com a capacidade de selecionar de onde vem o dado a ser armazenado no flip-flop corrente. Este corrente pode vir de uma fonte externa (carga paralela), vindo da direita, da esquerda, quando do deslocamento ou do próprio flip-flop, quando mantido o conteúdo inalterado.

Um exemplo deste tipo de registrador pode ser observado na figura abaixo. Este registrador é chamado “registrador-deslocador” de quatro bits com reset assíncrono e sinal de carga.

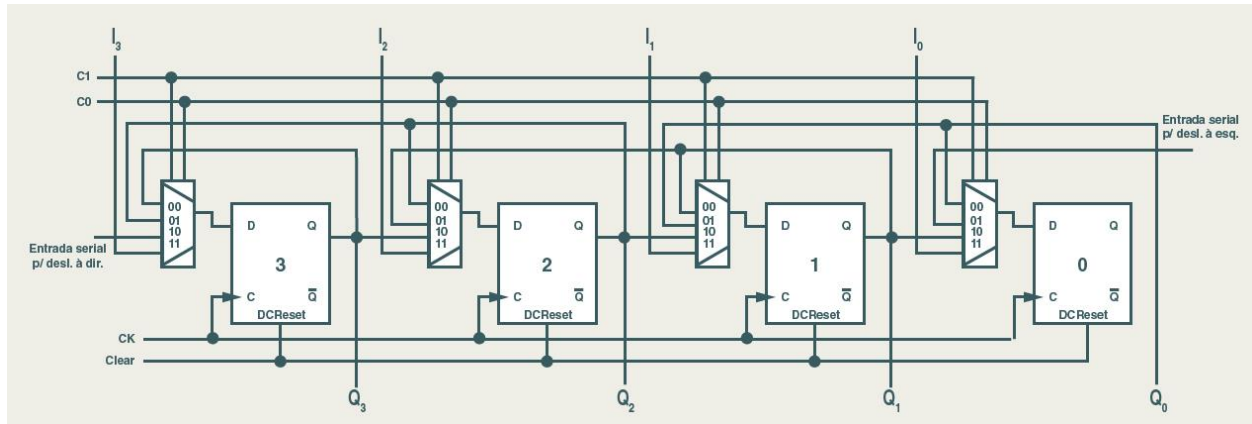


Figura 07

11

Este tipo de registrador é chamado também de **shift-register** e as seguintes operações são possíveis com ele:

- 1- Carga paralela;
- 2- Mantém seu conteúdo;
- 3- Zera o conteúdo (clear=1)
- 4- Desloca à esquerda e à direita

O seu funcionamento acontece da seguinte maneira: caso o sinal “clear” for igual a um (1), e $Q_3=Q_2=Q_1=Q_0=0$. Caso estas condições não sejam atendidas, haverá a necessidade de utilizar a seguinte tabela verdade:

CK	C1	C0	OPERAÇÃO
$\neq \uparrow$	X	X	mantém conteúdo
\uparrow	0	0	mantém conteúdo
\uparrow	0	1	desloca à esquerda (shift left)
\uparrow	1	1	desloca à direita (shift right)
\uparrow	1	0	carga paralela

Tabela Verdade 01

- Registrador Contador Assíncrono

Um registrador que faz contagem utilizando binário é chamado **incrementador** ou contador. Isso acontece por meio do sinal de relógio, então, a cada sinal, o conteúdo do registrador é incrementado de um valor (unidade). Um exemplo seria um registrador contador de 4 bits que pode contar de 0 (0000) até 15 (1111).

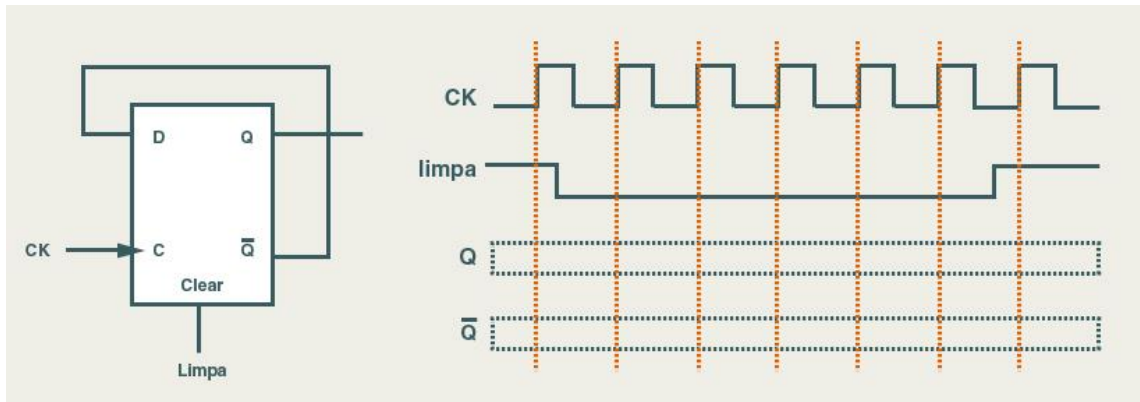


Figura 08

Na figura acima podemos ver um circuito e traçar a forma da onda de Q. Um circuito que tem como função ser contador de mais bits necessita ter conexão entre cada flip-flop vizinho, de tal maneira que cada flip-flop de ordem maior será o responsável pela ordenação do incremento de seu vizinho de ordem menor.

3- MEMÓRIAS

Vimos nos parágrafos anteriores como é feita a construção de alguns tipos de registradores. Os registradores são muito rápidos, mas em contrapartida possuem capacidade reduzida de armazenamento. Por exemplo, um registrador armazena somente uma palavra por vez.

Contudo, geralmente, nos sistemas digitais e em especial nas máquinas (computadores), uma quantidade grande de palavras necessita ser armazenada. Para atingir este objetivo de armazenamento, passa a ser prioridade que um sistema digital tenha um conjunto específico de circuitos que sejam adequados para armazenarem simultaneamente um **grande volume de palavras**. O nome dado aos circuitos que executam este armazenamento é **memória**.

Quanto ao seu funcionamento, a memória pode admitir apenas consulta no modo leitura ao seu conteúdo ou também poderá permitir o modo leitura e escrita (modificação).

Há memórias que são chamadas de **ROMs** (Read-Only Memories) em virtude de ser apenas para leitura. Todavia, as memórias chamadas **RAMs** (Random-Access Memories) permitem leitura e escrita.

Você deve ter chegado à conclusão de que a memória é a responsável por armazenar os dados com os quais o processador irá trabalhar.

Os conteúdos das ROMs podem ser gravados (escritos) durante a fabricação ou até mesmo pelo usuário, que poderia ser o fabricante do computador.

É importante salientar que:

Uma característica principal da memória **ROM** é que os dados uma vez gravados não poderão ser modificados e possuem apenas permissão de leitura.

As memórias **RAM** possuem circuitos que têm por característica a capacidade de armazenar as informações binárias que podem ser modificadas por inúmeras vezes.


14

Memória RAM

Vamos agora estudar como é a estrutura física ou os circuitos que compõem as memórias Random-Access Memory (RAM).

A memória RAM tem como organização uma matriz de 2^n linhas com m bits armazenados em cada linha, totalizando $2^n \times m$ bits. Geralmente, n está entre 16 e 32, podendo ser m enquanto for 1, 4, 8, 16 ou 32. A figura abaixo (Figura 09) mostra como é essa organização matricial. Observe que há 2^n linhas, chamadas também de posições. A cada posição há uma associação com um endereço que começa pelo endereço zero (0). Então, será preciso n bits para ocorrer a decodificação de 2^n endereços existentes. Ilustração: refazer ou formatar a imagem, tornando os dados mais nítidos.

Endereço em binário	Endereço em decimal	Conteúdo (exemplo)
0...000	0	011...0100
0...001	1	011...0100
0...010	2	101...1100
0...011	3	101...0001
0...100	4	011...0101
0...101	5	111...0110
0...110	6	101...0001
0...111	7	000...1101
	:	:
	:	:
0...110	2^n-2	000...1100
0...111	2^n-1	100...1100



 $m...bits$

Figura 09

Na Figura 09 é possível observar como é a organização de uma memória RAM e suas linhas e endereços.

15

Vamos supor que uma empresa irá fabricar apenas um circuito integrado (CHIP) em uma memória RAM com capacidade de armazenamento $2^n \times m$ bits, conforme a organização que vimos anteriormente. A figura abaixo (Figura 10) é uma representação do chip que falamos agora. Este chip precisa ter “n” entradas de endereço (A_{n-1}, \dots, A_1, A_0), de tal maneira que seja possível selecionar uma e apenas uma dentre todas as 2^n posições que há na matriz. Este chip poderá também conter “m” entradas (I_{n-1}, \dots, I_1, I_0), e “m” saídas (O_{n-1}, \dots, O_1, O_0), de tal maneira que seja possível a leitura/consulta do conteúdo de uma das 2^n linhas ou a gravação/escrita de uma informação nova em uma das 2^n linhas da matriz.

Em virtude de haver duas operações possíveis, **leitura** e **escrita**, sobre o conteúdo da matriz, é perfeitamente normal existir uma entrada de seleção da operação. Este tipo de entrada chamará RWS (Read/Write Select).

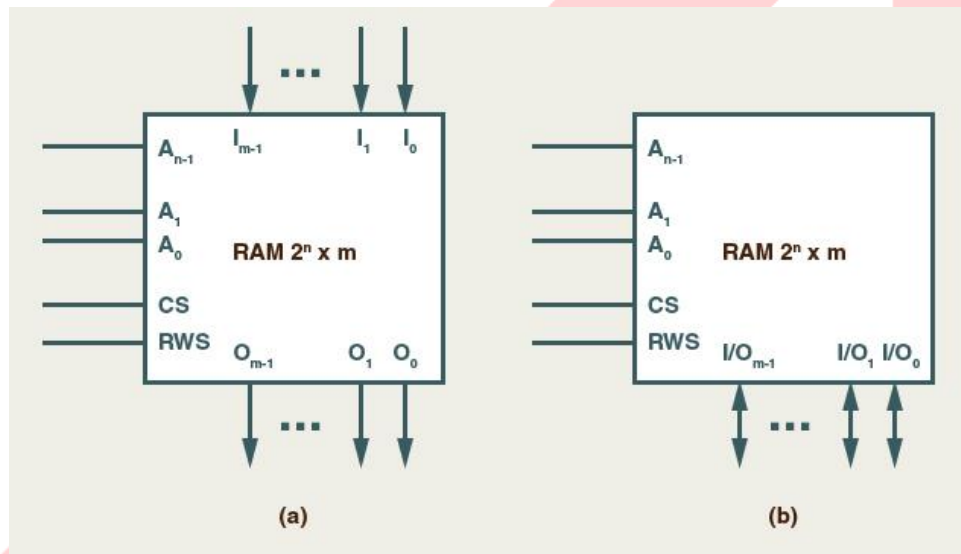


Figura 10

Em suma, há a necessidade de haver um sinal de habilitação do chip (CS-Chip Select). Para exemplificar, se o CS for igual a zero ($CS=0$) é porque o chip está desativado. O mesmo seria se o CS for igual a um ($CS=1$), é por que o chip está realizando uma operação determinada pelo valor da entrada RWS.

16

Na Figura 10 (copiada abaixo) podemos ver as representações gráficas possíveis para um chip de memória RAM.

Um sinal de habilitação do chip, CS, tem como serventia para o caso de haver necessidade de mais de um chip para implementar a memória do computador.

Observe novamente a Figura 10 (a), que é um exemplo de quando o quantitativo de bits que precisam ser armazenados em cada posição de memória “m” for pequeno, o chip RAM poderá ter entradas separadas das saídas. Contudo, o comum é que um mesmo conjunto de pinos do chip sirva tanto como entrada ou saída, situação que pode ser visualizada na figura 10 (b). Neste caso a utilidade vai depender da operação que será realizada, que é selecionada através do pino RWS. Na verdade, o detalhe é justamente este, pois o uso compartilhado irá reduzir a quantidade de pinos no chip, vindo a baratear seu custo.

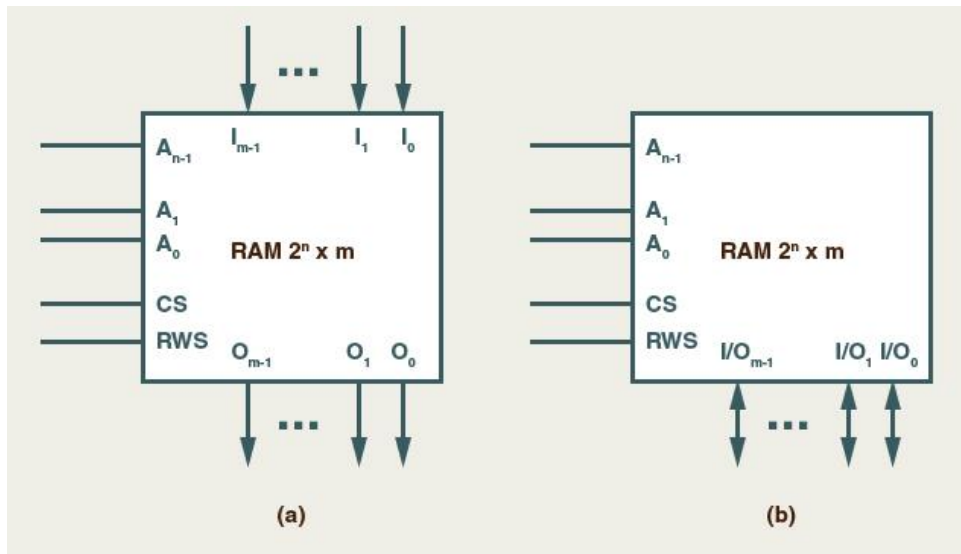


Figura 10

17

Olhando pelo lado estrutural, uma memória RAM tem sua organização feita semelhante a uma matriz de elementos básicos de memória, *buffers* de entrada/saída e decodificador de endereços. Conforme é possível observar na figura 11, uma CM (célula de memória) pode ser representada por meio de símbolo, sendo constituída por um latch D controlado cuja entrada está ligada à saída e a saída ligada à entrada de um *buffer*.

O **buffer**, também chamado driver, pode exercer duas funções:

1. **reforçar o sinal lógico**, em virtude da necessidade de quando há muitas portas conectadas a uma mesma saída ou quando o sinal vai percorrer um fio com grande extensão,
2. ou também exercer a função de uma **chave eletrônica** que isola fisicamente a saída da entrada, isto é, irá servir para disciplinar o acesso de vários sinais a um mesmo fio ou barramento.

Entendido o que é *buffer* na memória e voltando à figura 11, quando o sinal de **seleção de linha** é = 1, o bit armazenado no latch passará pelo *buffer*, permanecendo disponível na “saída” da célula. Caso o sinal de **habilitação de escrita** valha também 1, o valor presente na “entrada” será mantido no latch. O sinal de “habilitação de escrita” serve como controle para o latch.

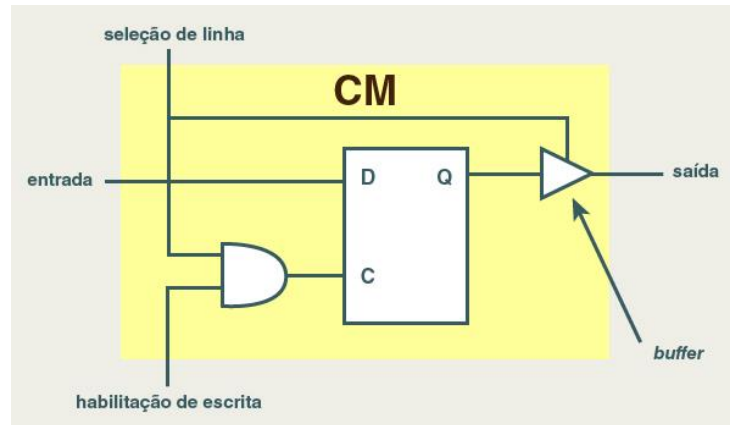


Figura 11

18

A Figura 12 é uma representação da memória RAM por meio de um diagrama de blocos. Trata-se de um exemplo de memória RAM (4X4), formada por dezesseis CMs. Para cada acesso à memória, o decodificador de endereços ativa o sinal de **seleção de linha** associado ao endereço aplicado às suas entradas, o que ativa todos os CMs da linha escolhida. Assim, caso RWS seja igual a zero (0) e CS igual a um (1), os bits que estão armazenados na linha escolhida passarão pelos *buffers* de saída das respectivas CMs e também pelos *buffers* de entrada e saída do chip.

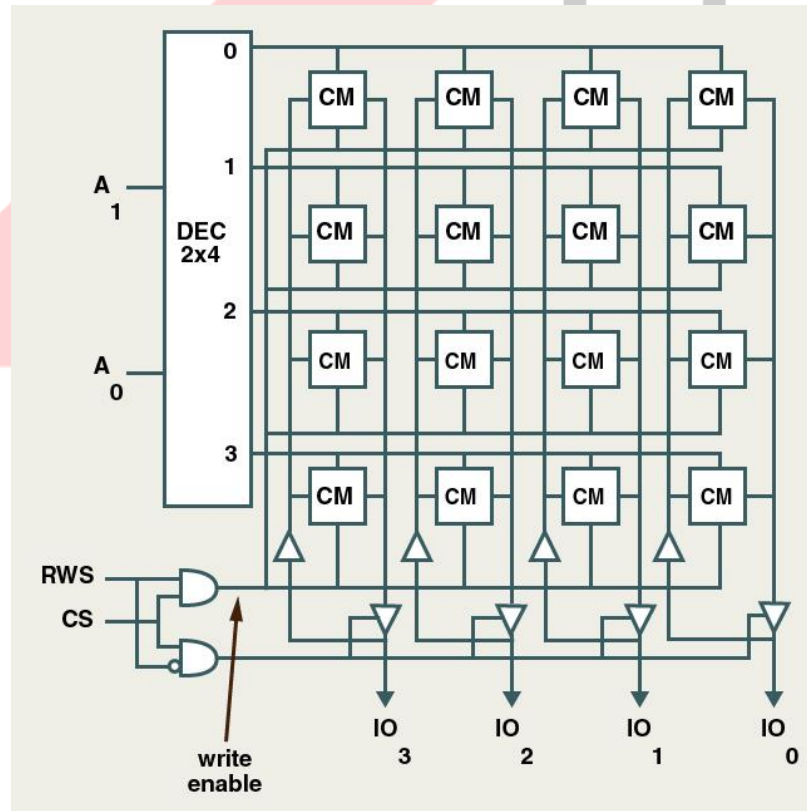


Figura 12

19

A maneira como é organizada a memória RAM impacta em restrições no tempo das operações de leitura e escrita. Por exemplo, como o caminho crítico da entrada a saída passa pelo decodificador, as entradas de endereço necessitam estar estáveis antes de qualquer outro sinal ou sinais. Isso quer dizer que, durante o ciclo de leitura mostrado na figura 13 abaixo, as entradas de endereço deverão ser fornecidas em t_0 , seguidas por CS em t_1 . Sendo assim, os dados da memória estarão disponíveis apenas em t_2 .

O atraso $t_2 - t_0$ é chamado **memory-access time** (tempo de acesso à memória), enquanto que o tempo $t_2 - t_1$ é chamado **output-enable time** (tempo de habilitação da saída). Observe que depois de os valores das entradas de endereço terem sido alterados em t_3 , os dados continuam disponíveis até t_5 . O intervalo $t_5 - t_3$ é chamado **output-hold time** (tempo de manutenção da saída). O intervalo $t_5 - t_4$ é chamado de **output-disable time** (tempo de desabilitação da saída).

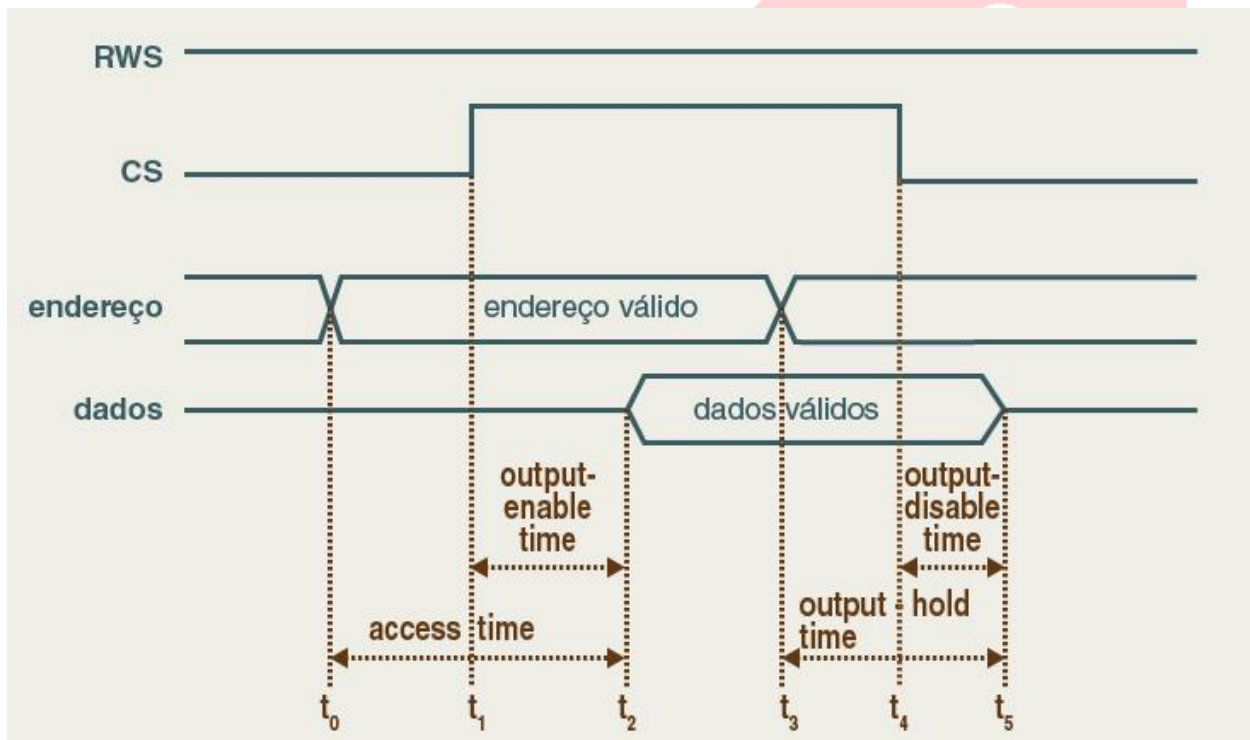


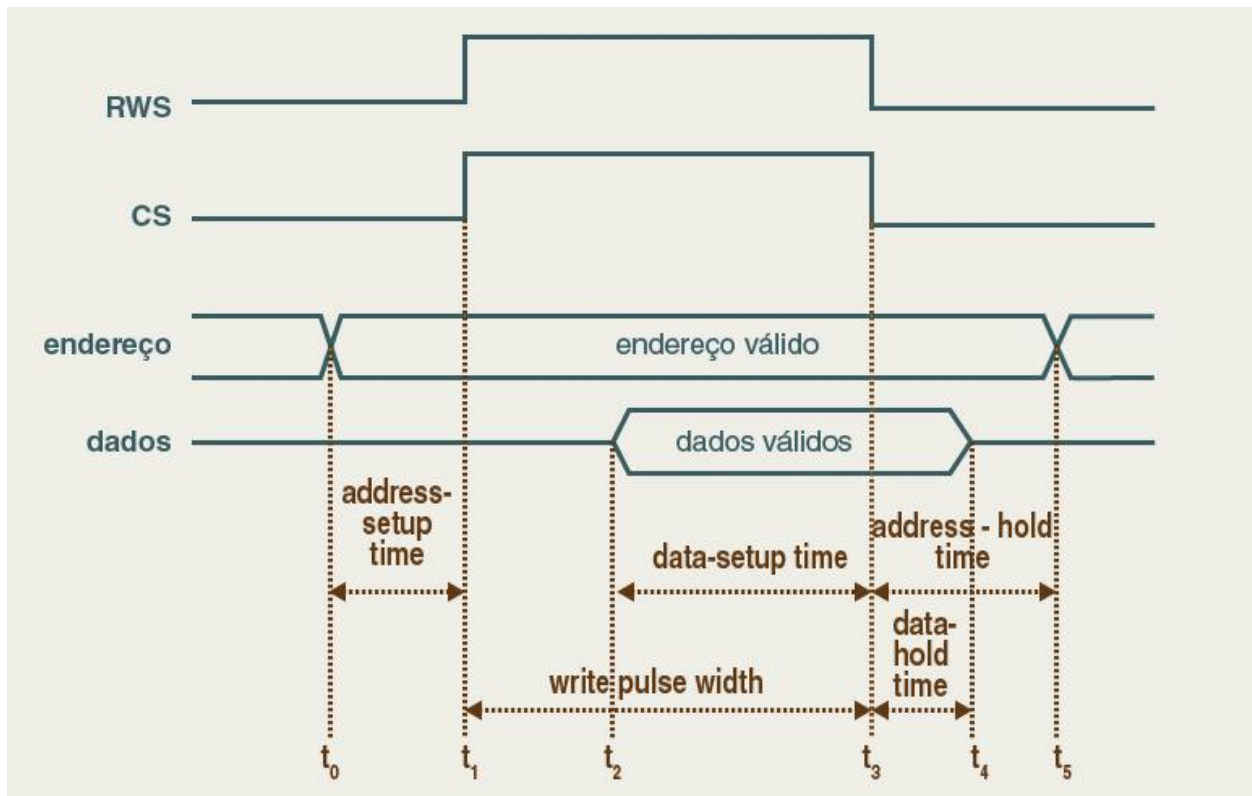
Figura 13

Em virtude de o caminho entre as entradas de endereço e as saídas ser maior do que o caminho entre CS até as saídas, o **tempo de acesso** aponta a validade dos dados sempre que o endereço e CS forem aplicados ao mesmo tempo. Em contrapartida, caso o endereço e CS não sejam mais válidos (CS igual a zero), o **tempo de desabilitação** especificará a validade dos dados.

20

A Figura 14 é possível ver as restrições temporais para a situação de um ciclo de escrita em uma memória RAM. Nesta figura observa-se por premissa que CS e RWS foram aplicados, ao mesmo tempo,

no instante t_1 . Em virtude de atraso entre o endereço e a saída ser maior do que o atraso entre CS ou RWS e a saída, o endereço deve ser aplicado um tempo antes, que neste caso, seria em t_0 .



O atraso $t_1 - t_0$ é chamado de *address setup time* (tempo de preparação do endereço). Em virtude de cada CM ser feito a partir de um latch D controlado com CS exercendo a função de controle, cada bit do dado na borda de descida de CS (t_3) será armazenado no correspondente latch. Contudo, é preciso que o dado já esteja estável um período antes e logo depois da borda de descida de CS para garantir a referida escrita. Observe a figura e veja que os tempos são descritos como *data setup time* (tempo de preparação do dado) e *data hold time* (tempo de manutenção do dado), sendo chamados respectivamente como: $t_3 - t_2$ e $t_4 - t_3$.

Cabe salientar que CS ou RWS devem ter duração igual ou superior ao intervalo de tempo $t_3 - t_1$, que é chamado de *write-pulse width* (duração do pulso de escrita). Também é importante saber que o endereço deve estar válido durante um tempo após a borda de descida de RWS ou CS. Este intervalo de tempo é denominado *address-hold time* (tempo de manutenção do endereço), que pode ser demonstrado na imagem como $t_5 - t_3$.

21

Nós vimos aqui a CM representada pela composição de um latch D e duas portas, entretanto, muitos fabricantes utilizam estruturas menos complexas e também com poucos transistores. A maneira como é feita a implementação de CMs é que determina a classificação da memória RAM em **estática** e **dinâmica**.

RAM estáticas <i>Static RAM (SRAM)</i>	RAM dinâmicas <i>Dynamic RAM (DRAM)</i>
<ul style="list-style-type: none"> • Sua CM é feita com seis transistores, sendo que quatro destes formam dois inversores conectados em laço de realimentação, exercendo a função do latch D. • Na função da porta E e do <i>buffer</i> de saída existe um transistor, para cada um, o qual serve de chave liga/desliga. • A memória SRAM é capaz de armazenar seu conteúdo por tempo indeterminado enquanto sua alimentação for mantida. • Têm o custo mais elevado e apresentam capacidade menor de armazenamento. • São mais rápidas que as DRAM, sendo indicadas quando não é preciso armazenar uma grande quantidade de dados e a performance (velocidade) e há necessidade de velocidade maior nas operações. 	<ul style="list-style-type: none"> • Cada CM é implementada com apenas um transistor, o que faz com que o conteúdo da CM se perca ao término da operação de leitura, necessitando ser reescrito. • Devido a defeitos no processo de fabricação, o conteúdo da CM é mantido por pouco tempo. • Os dois problemas acima são minimizados com a utilização de um artifício chamado de reflash, instalado no interior da memória, que de tempos em tempos reforça o conteúdo de cada linha de CMs. Durante o reflash, as operações de escrita e leitura ficam travadas, reduzindo o desempenho da memória. • Possuem grande densidade e maior capacidade de armazenamento. • Apresentam custo reduzido. • São muito utilizadas em eletrônicos.

22

RESUMO

Neste módulo estudamos os circuitos digitais responsáveis pelo armazenamento de informação, sendo eles os registradores e as memórias ROM e RAM. Vimos que o registrador é um circuito digital que tem em sua formação “n” flip-flops, que de forma simultânea, consegue armazenar (de forma independente) “n” bits. Vimos também como funcionam os registradores com carga paralela, os registradores de deslocamento, o registrador de deslocamento com sinal de carga paralela e o registrador contador assíncrono. As memórias ROM possuem organização semelhante às memórias RAM. Entretanto, a CM de uma ROM é muito mais simples, geralmente constituída por apenas um transistor ou diodo, que exerce a função para conceder acesso ao zero (=0V) lógico ou ao um (1) lógico (5V a 1,5V, conforme a tecnologia). As memórias ROM, nos computadores, tem por função o armazenamento das configurações básicas que não são alteradas. Um exemplo são as rotinas de entrada e saída que constituem a BIOS dos computadores.

UNIDADE 3 – FUNCIONAMENTO E COMPONENTES BÁSICOS

MÓDULO 2 – MEMÓRIA

01

1 - CARACTERÍSTICAS DOS SISTEMAS DE MEMÓRIAS DOS COMPUTADORES

Durante a leitura e estudo deste módulo você terá a oportunidade de entender as características dos sistemas de memória dos computadores. Dentre estas características, iremos destacar: localização, capacidade, unidade de transferência e método de acesso.

Veremos também a forma hierárquica da organização da memória no computador, esta característica é em virtude de como é a sequência de utilização e acesso à memória. Entenderemos o que é memória RAM e ROM e as respectivas diferenças entre elas. E para finalizar veremos como é simples o cálculo da capacidade da MP (Memória Principal).

Então vamos começar?

02

Para o adequado e eficaz funcionamento da manipulação das informações (instruções de um programa e dados) de e para a memória de um computador, constata-se a necessidade de se ter, em um mesmo computador, diferentes tipos de memória. Para determinadas atividades, por exemplo, é essencial que a transferência de informações seja a mais rápida possível. É o caso das atividades realizadas internamente no processador central, onde a velocidade é primordial, porém a quantidade de bits a ser manipulada é muito pequena (em geral, corresponde à quantidade de bits necessária para representar um único valor - um único dado).

O conceito de memória é aparentemente simples, pode-se até dizer que é aquele componente que apresenta maior variedade de tipos, tecnologia, organizações, desempenhos e custos, entre todos os elementos de um sistema de computador ou sistema computacional.

Devido à grande variedade de tipos de memória, não é possível implementar um sistema de computação com uma única memória, pois nenhuma das tecnologias de memória existentes satisfaz de maneira ótima todos os requisitos de armazenamento de dados em computadores.

Na realidade, há muitas memórias no computador, as quais se interligam de forma bem estruturada, constituindo um sistema em si, parte do sistema global de computação, podendo ser denominado **subsistema de memória**.

Esse subsistema é projetado de modo que seus componentes sejam organizados hierarquicamente, ou seja, um sistema computacional típico é equipado com uma **hierarquia de subsistemas de memória**, algumas internas (diretamente acessíveis pelo processador) e outras externas (acessíveis pelo processador por meio de um módulo de E/S - entrada / saída).

Os sistemas de memória de computadores podem ser facilmente compreendidos por meio de sua classificação, de acordo com suas características fundamentais. Estas características podem ser visualizadas na tabela abaixo:

Características fundamentais de sistemas de memórias de computadores	
Localização Processador Interna (principal) Externa (secundária)	Desempenho Tempo de acesso Tempo de ciclo Taxa de transferência
Capacidade Tamanho da palavra Número de palavras	Tecnologia De semicondutores Magnética Ótica Magneto-óptica
Unidade de Transferência Palavra Bloco	Características Físicas Volátil / não-volátil Apagável / não-apagável
Método de acesso Sequencial Direto Aleatório Associativo	Organização

Estudaremos com mais detalhes cada uma dessas características a seguir.

• Localização

O termo **localização** é utilizado para diferenciar se a memória é do tipo **interna** ou **externa** à máquina.

No geral a memória interna é popularmente conhecida como a **MP (memória principal)**, mas há outros tipos de memória interna. O processador necessita de uma memória local própria, formada por registradores. Consequentemente, conforme iremos estudar, a unidade de controle do processador também poderá ter sua memória interna própria. Outro exemplo de memória interna é a **memória cache**.

A **memória principal**: também chamada de primária ou real, é o local em que estão armazenados as instruções e dados. É um dispositivo de armazenamento que mantém os dados (informações) de maneira temporária dos serviços que estão em processamento no processador. É na MP que os dados ficam disponíveis para transferência para equipamentos de saída. A MP é desde o início da computação a memória básica de um sistema de computação. Em suma, é na MP onde o programa que será executado é armazenado para o processador faça a busca de instrução por instrução.

A **memória externa** são todos os dispositivos de armazenamento, também conhecidos como periféricos. Exemplo: **HD Externo**. Estes periféricos são acessíveis ao processador (UCP) por meio de controladores de E/S (entrada e saída).

A **memória secundária** também pode ser chamada de memória externa e “não-volátil”. O termo “não-volátil” significa que esta mantém as informações armazenadas mesmo após o desligamento do computador. Para facilitar o entendimento, usaremos o HD (Disco rígido) como exemplo de MS (memória secundária). É no HD que geralmente estão armazenados os programas (navegadores, jogos, leitores de PDF, leitores de texto etc.) e também os arquivos pessoais dos usuários. Contudo, em determinadas situações, o sistema operacional utilizará o HD como uma extensão da memória principal. Essa ação tem o intuito de evitar que haja travamentos do sistema. Em suma, quando falamos que o HD é um exemplo de memória secundária, estamos fazendo referência ao fato de os discos também desempenharem a função de ser uma extensão da memória principal, sendo uma segunda memória, caso não haja espaço suficiente na MP.

05

• Capacidade

Quando você resolve adquirir um notebook ou outro modelo qualquer de computador, uma das características que provavelmente é observada é a **capacidade da memória**. Você pode escutar do vendedor que o computador “vem com a capacidade de 4GB de RAM”. Mas o que é capacidade de memória?

Capacidade de memória é o número total de bits que o dispositivo é capaz de armazenar. Pode ser expresso de várias formas: 2048 bits; 2kb ou 256 Bytes.

Por exemplo: computador da marca XXX vem equipado com 4GB (gigabytes) de memória. Na memória interna, a capacidade é geralmente especificada em razão das **palavras** ou dos bytes (1 byte= 8 bits). Os tamanhos mais usados para palavras são geralmente 8, 16, 32 ou 64 bits. No caso da memória externa, a capacidade é tipicamente expressa em bytes.

• Unidade de transferência

Um conceito que está relacionado ao assunto é a **unidade de transferência de dados** (UTD). Na memória interna, a UTD é proporcional ao número de linhas de dados do módulo de memória. Geralmente o número de linhas é igual ao tamanho da palavra.

Para entender bem, veja estes três conceitos relacionados à memória interna:

Unidade endereçável

Em vários sistemas, a unidade endereçável de dados é a **palavra**. Contudo, outros sistemas permitem endereçamento de bytes. Em ambos os casos, a relação entre o tamanho em bits A de um determinado endereço e o número de unidades endereçáveis, onde N é $2^a = N$.

Palavra

É naturalmente a unidade de organização de uma memória. O tamanho de uma palavra é geralmente igual ao número de bits usados para a representação de um número inteiro e ao tamanho de uma instrução. Entretanto, há várias exceções. Um exemplo que vale a pena mencionar foi o primeiro supercomputador da história, o Cray 1, que teve seu lançamento no ano de 1976, criado pelo norte-americano Seymour Cray. No CRAY 1 uma palavra tinha 64 bits, porém a representação de números inteiros utilizava 24 bits. Outro exemplo foi a arquitetura VAX (Virtual Address eXtension), que foi uma família de minicomputadores de 32 bits, fabricado pela Digital (Digital Equipment Corporation) no ano de 1978. No VAX uma palavra tem 32 bits e há uma grande variedade de tamanhos das instruções, representados como múltiplos de byte.

Unidade de transferência

A unidade de transferência de dados da MP é o número de bits que podem ser escritos ou lidos de cada vez. Não há necessidade de ser igual a uma **palavra** ou à unidade endereçável de dados. No caso das memórias externas, as transferências são em unidades maiores que uma palavra, estas unidades são chamadas de **blocos**.

06

- **Método de acesso**

O método de acesso aos dados é uma das maneiras de diferenciar os tipos de memória. Iremos destacar o **acesso aleatório**.

No **acesso aleatório**, cada posição de memória endereçável possui um mecanismo único de endereçamento a ela fisicamente conectado. O tempo de acesso a uma determinada posição é constante e não depende da sequência de acessos anteriores.

Dessa forma, qualquer posição será selecionada aleatoriamente, sendo acessada e endereçada diretamente. A MP, bem como algumas memórias cache, são exemplos de dispositivos de memória de acesso aleatório.

- **Desempenho**

A partir da visão do usuário, as características que mais se destacam em relação à memória são seu desempenho e sua capacidade. Há parâmetros que são usados para mensurar este desempenho, são eles:

Tempo de ciclo de memória

Conceito aplicável em especial a memórias de acesso aleatório e compreende no tempo de acesso e o tempo adicional antes que um segundo acesso seja iniciado. Esse tempo a mais poderá ser necessário para que haja o desaparecimento de transientes nas linhas de sinais ou para a regeneração dos dados, caso a leitura seja para subscrever o conteúdo.

Tempo de acesso

Em uma memória de acesso aleatório, é o tempo despendido para haver uma operação de escrita ou de leitura. Também é o tempo passado desde o instante em que um endereço é apresentado à memória até o momento estão disponíveis para utilização ou armazenados.

Taxa de transferência

É a medida na qual os dados podem ser transferidos de ou para a unidade de memória. Na memória de acesso aleatório, é expressa da seguinte forma: $1/(\text{tempo de ciclo}) = 1/T_c$.

07

• Tecnologias

Várias tecnologias estão sendo usadas para fabricar memórias para computadores. Geralmente as memórias mais usadas são:

- as memórias de semicondutor ou [memórias semicondutoras](#),
- memórias de superfície magnética ([discos e fitas magnéticas](#)),
- as [memórias ópticas](#) e
- magneto-óptico.

• Características físicas

As características físicas de armazenamento são de suma importância. Em uma [memória volátil](#) há perda de dados quando a energia é desligada. Na [memória não-volátil](#), a partir do momento que os dados são gravados, permanecem disponíveis após o desligamento do computador. As memórias de semicondutor podem ser não-voláteis como voláteis. As memórias de superfície magnética (HD) são não-voláteis.

• Organização

A organização de memória é o arranjo físico dos bits para formar PALAVRAS.

A organização é um aspecto que deve ser observado no projeto de memórias de acesso aleatório. Geralmente é fornecido como sendo o número de posições de memória *versus* o número de bits por posição como, por exemplo, 2kB, ou seja, 2x1024 Bytes ou 2048x8bits.

Memórias Semicondutoras

São memórias produzidas com componentes de estado sólido e dividem-se em dois tipos básicos: RAM (Random Access Memory) e ROM (Read Only Memory).

Discos e fitas magnéticas

Pode parecer incrível, mas as fitas magnéticas e os discos ópticos são os principais representantes para armazenamento de dados. Ousamos dizer que, provavelmente, é o suporte de dados mais antigos e ainda muito utilizados em sistemas computacionais, embora tenha havido várias evoluções desde sua criação na década de 1950.

Memórias ópticas

São memórias que utilizam a polarização e refração da luz. Os CDs gravam informações em duas dimensões, mas já experimentos que são capazes de armazenar em cinco dimensões. Um exemplo prático do uso de CD e DVD são os que armazenam informações e programas, sendo utilizados na instalação de programas no computador.

Memória volátil

São memórias que perdem o conteúdo retido quando deixam de ser alimentadas (RAM).

Memória não-volátil

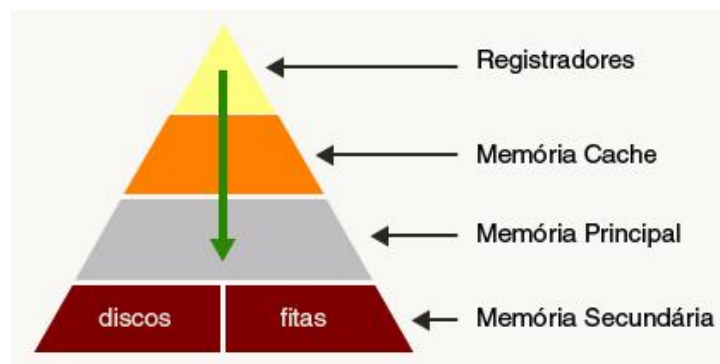
São memórias que retêm a informação mesmo quando não estão alimentadas (ROM). Dentre as memórias não-voláteis, sobressaem-se as memórias permanentes de computadores, que retêm programas de BIOS, SETUP e DIAGNÓSTICO, fabricadas pelos AWARD, PHOENIX, AMI etc.

08

2 - HIERARQUIA DE MEMÓRIA

Nos computadores a memória é organizada de forma hierárquica. Esta hierarquia começa pela proximidade física ao processador. Vamos fazer uma analogia com uma pirâmide.

Ao imaginarmos uma pirâmide, podemos visualizar que o nível mais alto (topo da pirâmide) é composto de **registradores do processador**, estão próximos fisicamente ao processador. Logo depois, do topo para baixo, há níveis de memória cache, que são chamados de caches L1, L2, L3 e L4.



Mantendo a sequência, de cima para baixo, há a chamada **memória principal**, que geralmente utiliza módulos de memória dinâmica de acesso aleatório (DynamicRandom-Access Memory – DRAM). Essas memórias são chamadas de **memórias internas** ao sistema de computação.

Esta estrutura hierárquica continua com a **memória externa**, na qual o nível subsequente é tipicamente composto por um HD fixo (disco rígido fixo), e com os níveis abaixo constituídos de meios removíveis, tais como discos ópticos, memória flash e fitas magnéticas.

09

Conforme vamos descendo pela hierarquia da memória, há também uma diminuição do custo por bit, isto é, a **capacidade de memória** vai aumentando, mas o tempo de acesso torna-se mais lento. Em virtude disso, você há de convir que o ideal é a utilização de apenas memórias mais rápidas; contudo, estas memórias são muito caras. Sendo assim, para manter um equilíbrio entre custo e o benefício, o tempo de acesso sofre diminuição em favor do custo mais baixo, e por consequência, utilizando memórias que são mais lentas.

Fica evidente que diante desta situação criaram-se artifícios para facilitar e aumentar a performance. Uma das ideias foi organizar os dados e também os programas na memória de tal forma que as palavras de memória mais requisitadas sejam localizadas e encontradas de forma mais rápida.

Geralmente, os acessos futuros à memória principal pelo processador sejam posições recentemente usadas na memória. Sendo assim, a memória cache mantém uma cópia das palavras da memória usadas recentemente. Isto agiliza o processo, pois as palavras que serão requisitadas pelo processador provavelmente estarão armazenadas na memória cache.

10

As limitações de um projeto de memória são possíveis de resumir em três itens:

- quanto ao custo,
- quanto à velocidade e
- quanto à capacidade de armazenamento.

A capacidade de armazenamento

É algo em constante evolução. Na década de 90 um computador modelo Infoway, da marca ItauTec, tinha capacidade de memória de 16MB, utilizava um processador Pentium 166Mhz e tinha o preço de aproximadamente R\$ 3.000,00 (três mil reais). Na época o salário mínimo era de aproximadamente R\$ 100,00 (cem reais), consequentemente adquirir um computador era algo para poucos. Veja que atualmente a capacidade de armazenamento aumentou de maneira estrondosa e tornou-se mais fácil adquirir uma máquina ou computador (PC).

A velocidade

Também evoluiu bastante, contudo precisamos observar que para obter um melhor desempenho, a velocidade da memória deve ser compatível com a do processador. Sendo assim, deve-se evitar que o processador fique ocioso aguardando que operações ou instruções sejam acessadas constantemente na memória durante a sua execução.

A questão custo

Deve ser observada para tornar acessível e viável comercialmente uma determinada memória, o ideal é que o preço da memória seja compatível com dos outros componentes de um sistema computacional (PC).

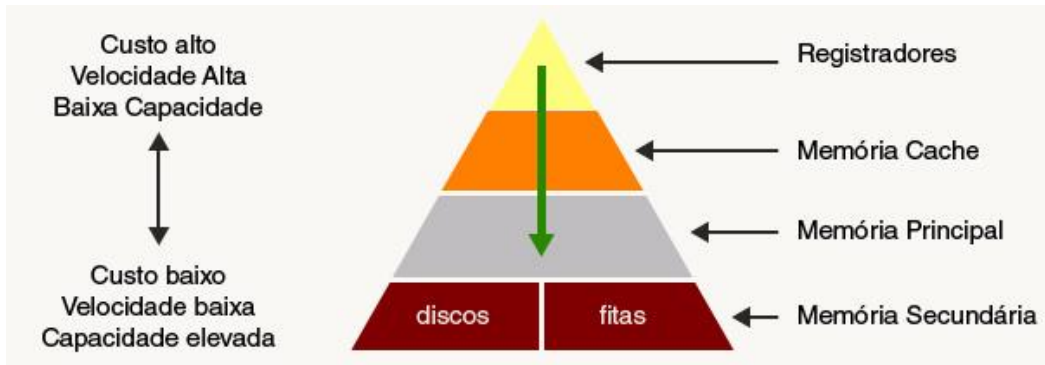
11

Uma gama de tecnologias é empregada na implementação de sistemas de memória, mas o mais importante nesta implementação é observar as seguintes relações:

- capacidade maior com um custo por bit menor;
- capacidade maior com um tempo de acesso menor.
- tempo de acesso mais rápido com o custo por bit maior;

Observamos que o projetista de um sistema de memória precisa estar atento que a tecnologia empregada forneça grande capacidade de armazenamento de dados, com desempenho melhor e com tempo de acesso menor. Uma maneira para ajudar os projetistas foi empregar uma **hierarquia de memórias**, conforme já vimos. Para ajudar no entendimento desta hierarquia, observe a figura abaixo e, tomando como referência o topo da pirâmide, à medida que descemos há as seguintes relações: o custo por bit diminui, a capacidade aumenta, o tempo de acesso aumenta e a frequência de acesso à memória pelo processador diminui.

Desta forma, as memórias menores e mais caras, porém, mais rápidas são combinadas com memórias maiores e mais baratas e também mais lentas. O grande segredo desta estrutura é a diminuição da frequência de acessos. Isto é em virtude dos registradores, memória cache e MP serem 100.000 vezes mais rápidas que as memórias (secundárias) que são os discos (HD).



12

Uma memória importante, mais cara e menor, é a **memória FLASH**, que é constituída de registradores internos do processador. Geralmente um processador possui várias centenas de registradores internos.

O personagem principal das memórias internas é a MP (**memória principal**). Cada posição da MP possui endereço único. A maioria das instruções de máquina faz referência a um ou mais endereços da MP. Geralmente a MP é combinada com uma memória cache menor e de velocidade maior. A maioria de nós (usuários) e até mesmo usuários avançados (desenvolvedores de software) não prestamos atenção à memória cache. Mesmo passando despercebida a memória cache desempenha o papel importante de ser a responsável pela organização da movimentação de dados entre os registradores do processador e a MP, com o intuito de melhorar a performance.

Observando a hierarquia, os registradores, memória cache e MP são os tipos de memória voláteis que utilizam a tecnologia de semicondutores. Este uso de 3 níveis de memória explora o fato de existir uma gama de tipos de memória de semicondutor que possuem diferenças no custo e na velocidade.

Para o armazenamento de dados de forma permanente são usados **dispositivos externos** de armazenamento de massa, dentre os mais conhecidos temos:

- os HD (Hard Disc - discos rígidos),
- discos SSD (Solid State Drive) - são parecidos com os HDs, entretanto, os SSDs não utilizam discos magnéticos ou cabeças de leitura, mas gravação por chips de memória flash (NAND) não volátil, similar aos usados em pendrives.),
- mídia removíveis (HD externo, pendrives, cartões SD),
- fitas e dispositivos ópticos (CD-ROM, DVD, Blue Ray).

13

Há também a possibilidade de utilizar parte da MP como área de armazenamento temporário de dados que depois serão gravados em disco. Essa técnica é chamada de **cache de disco**, que tem por objetivo melhorar o desempenho do sistema fazendo operações de escrita em disco de forma agrupada ou invés de fazer transferências de pequenas quantidades de dados. Outra maneira é que alguns dados a serem escritos em disco podem ser referenciados outra vez pelo programa e antes que aconteça uma nova transferência de dados da cache para esse disco.

Nos primórdios da computação, quando dos primeiros computadores, a tecnologia mais utilizada na época para fabricação da MP de acesso aleatório era utilizando um grupo de anéis de material ferromagnético, chamados de núcleos. Desde o seu aparecimento, as memórias que utilizam a tecnologia da microeletrônica são muito mais superiores às memórias de núcleo magnético. Nos dias atuais, o uso de pastilhas de semicondutores para a MP é muito difundido. Observe a figura abaixo sobre os tipos de memória de semicondutor:

Tipo de memória	Categoria	Mecanismo de apagamento	Mecanismo de escrita	Volatilidade
Memória de acesso aleatório (RAM)	Memória de leitura e de escrita	Eletricamente, em nível de bytes	Eletricamente	Volátil
Memória apenas de leitura (ROM)	Memória apenas de leitura	Não é possível	Máscaras	Não - Volátil
ROM programável (PROM)				
PROM apagável (EPROM)	Memória principalmente de leitura	Luz UV, em nível de pastilha	Eletricamente	
Memória flash		Eletricamente, em nível de blocos		
PROM eletricamente apagável (EEPROM)		Eletricamente, em nível de bytes		

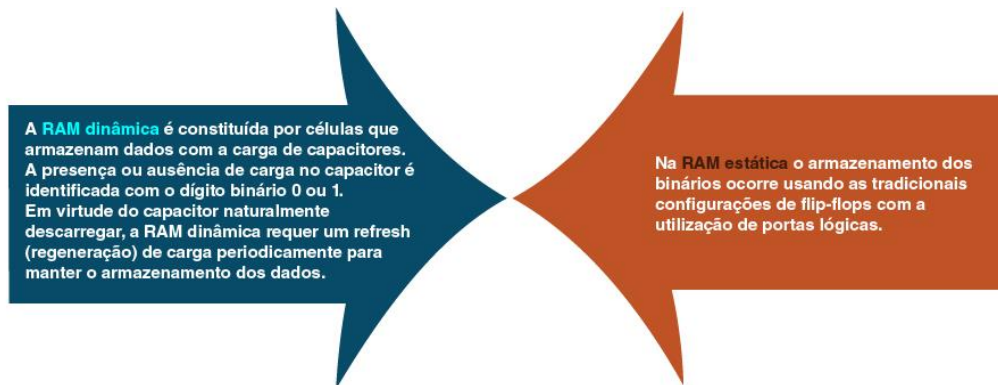
Na figura acima o tipo de memória de semicondutor mais conhecida é a **memória de acesso aleatório (RAM - Random Access Memory)**. A RAM tem uma particular característica de possibilitar a leitura e gravação (escrita) de dados sejam de maneira fácil e rápida. Esta gravação e leitura são realizadas por meio de sinais elétricos.

14

3 - MEMÓRIA RAM

Como já foi mencionado anteriormente, a memória RAM é volátil em virtude de necessitar de fornecimento de energia constante, pois caso seja interrompido este fornecimento, será perdido os dados. Consequentemente, a memória RAM é utilizada para armazenamento temporário de dados.

A tecnologia da memória RAM é dividida em **dinâmica** e **estática**.



Tanto a memória RAM dinâmica quanto a RAM estática têm por característica serem **voláteis**. O posicionamento de uma memória dinâmica é muito simples e, por consequência, tem o tamanho menor em relação à memória estática. A RAM dinâmica tem células menores que implicam em mais células por unidade de área, isto a torna mais densa e também mais barata do que a RAM estática. Contudo, ela requer um circuito de regeneração. Sendo assim, no caso de memórias de grande capacidade, o custo fixo do circuito de regeneração será compensado pelo menor custo das células da RAM dinâmica. Consequentemente, as RAM dinâmicas têm mais vantagem no quesito capacidade maior. Um detalhe importante a ser frisado é que as memórias estáticas são geralmente mais rápidas que as dinâmicas.

15

4 - MEMÓRIA ROM

Contrastando com a memória RAM, há a memória chamada de “apenas de leitura” (*Read-Only Memory – ROM*). O nome nos remete que a ROM contém mantêm os dados e que eles não devem ser alterados. Embora ela permita ler seus dados, em contrapartida, ela não permite alterá-los. A ROM tem sua importância destacada na microprogramação, nas bibliotecas de sub-rotinas, programas do sistema e tabelas de funções. Caso a capacidade de memória necessária for pequena, a vantagem da utilização de uma memória ROM é que os programas ou dados estão permanentemente armazenados na PM, não necessitando serem carregados a partir de um dispositivo de armazenamento secundário.

Uma alternativa de menor custo para ROM é a PROM (*programmable ROM*), isto geralmente ocorre quando é necessário um pequeno número de memórias ROM com dado conteúdo de memória. Como pode ser evidenciado pelo nome “programmable”, a PROM é similar à memória ROM, e a PROM também é não-volátil, sendo que os dados podem ser gravados, mas apenas uma única vez. A gravação em uma memória PROM é feita eletricamente e geralmente é feita pelo usuário ou pelo fornecedor, após a fabricação da pastilha original.

Variações principalmente da memória apenas de leitura, são mais úteis em aplicações que exigem armazenamento não-volátil e há uma certa frequência de operações de leitura em relação às operações de escrita. Existem três maneiras mais comuns de MP de leitura, que são:

- memória flash,
- EEPROM e
- EPROM.

Memória flash

A memória FLASH é outro tipo de memória de semicondutor que tem este nome em virtude da velocidade com que pode ser programada. A memória FLASH foi criada na década de 80, mostra características intermediárias entre a EEPROM e a EPROM, tanto na funcionalidade como no custo. Similarmente a EEPROM ela utiliza a tecnologia de apagamento elétrico dos dados, podendo em poucos segundos ter seu conteúdo apagado completamente. Também é possível apagar alguns blocos de memória e não a pastilha inteira.

EEPROM

A EEPROM (memória apenas de leitura programável e apagável eletricamente) os dados podem ser gravados sem que seja necessário apagar todo o seu conteúdo atual. Esta memória combina a vantagem da flexibilidade com a não volatilidade em virtude de poder ser diretamente atualizada, por meio das linhas de dados e de endereço usuais e do barramento de controle.

EPROM

EPROM (memória programável apenas de leitura) pode ser subscrita (apagada) por um processo ótico em que os dados podem ser gravados e lidos eletricamente. Todavia, antes de qualquer operação de escrita, todas as células de memória devem ser apagadas, voltando a ter o valor inicial.

16

5 - CÁLCULO DA CAPACIDADE DE UMA MEMÓRIA PRINCIPAL

Para calcular a capacidade de uma MP (Memória Principal) precisamos entender que as memórias RAM são organizadas em **células**, que podem conter parte de um dado ou o dado completo. A mensuração (medida) utilizada para sua capacidade é o bit (Kbit, Mbit etc.) ou em quantidade de células diretamente. Em virtude de cada célula conter, no máximo, um determinado dado, e cada célula é identificada por um **endereço**, o mais importante componente para a determinação da capacidade máxima de memória de um sistema é a quantidade de células (ou endereços) suportada por um sistema. Na maioria dos sistemas são usadas células de 1 Byte (8 bits), no dia a dia usamos geralmente em bytes, por exemplo, compra-se um computador com 4 megabytes de memória. Outro exemplo: dizemos que um computador tem um pente de memória de 128 MB – 128 MegaBytes. No último exemplo, estamos dizendo é o número de células.

Para realizar conversões, é necessário lembrar que:

$$2\text{Kbytes} = 2 \times 1024 \times 8 \text{ bits} = 2 \times 2^{10} \times 2^3 = 2^{14} \text{ bits.}$$

Nas células são armazenados bits, então para determinada célula de M bits, podemos ter 2^M diferentes símbolos. E uma célula com capacidade de 8 bits, podemos ter até 256 símbolos, que vão de 00000000 até 11111111.

Caso determinada memória principal possua N endereços e E é a quantidade de bits que o compõe cada um dos N endereços, consequentemente $N = 2^E$. Essa memória principal terá, por consequência, a capacidade de armazenar T igual a $N \times M = 2^E \times M$.

Então podemos resumir que para calcular o total de bits de uma memória principal é igual a $N \times M$.
Então:

$$N = 2^E, \text{ então, } MP = 2^E \times M.$$

17

RESUMO

Estudamos neste módulo que a localização, capacidade, unidade de transferência e método de acesso são características dos sistemas de memória dos computadores. Vimos que há uma hierarquia no acesso à memória conforme a sua proximidade física ao processador. Entendemos que a forma hierárquica da organização da memória no computador é em virtude de como é a sequência de utilização e acesso à memória. Por proximidade o processador acessa primeiro a memória cache, depois a Memória Principal (MP) e depois a memória secundária. Esta é a maneira básica do uso dos recursos de memória em um computador.

Várias tecnologias são empregadas na implementação de um sistema de memória. É interessante observar que o mais importante nesta implementação é conseguir uma capacidade de armazenamento maior com um custo por bit menor. Isto é, uma memória principal (MP) com maior capacidade e com um custo baixo (mais barata). Entretanto, ao compararmos com uma memória secundária (exemplo: HD) com a MP é fácil notar a diferença de capacidade de armazenamento, mas o acesso ao HD é muito mais lento do que o acesso a MP.

Vimos que há a memória RAM e ROM e as respectivas diferenças entre elas. No caso da memória RAM ela é dividida em dinâmica e estática em virtude da tecnologia utilizada. Uma RAM dinâmica é constituída por células que armazenam dados com a carga de capacitores. Contrastando com a memória RAM, há a memória chamada de “apenas de leitura” (ROM). O nome nos remete que a ROM contém e mantém os dados e que eles não devem ser alterados. Embora ela permita ler seus dados, em contrapartida, ela não permite alterá-los. A ROM tem sua importância destacada na microprogramação, nas bibliotecas de sub-rotinas, programas do sistema e tabelas de funções.

E por fim, vimos que para calcular a capacidade de uma MP precisamos entender que as memórias RAM são organizadas em células, que podem conter parte de um dado ou o dado completo. A mensuração

(medida) utilizada para sua capacidade é o bit (Kbit, Mbit etc.) ou em quantidade de células diretamente.

UNIDADE 3 – FUNCIONAMENTO E COMPONENTES BÁSICOS

MÓDULO 3 – UCP – UNIDADE CENTRAL DE PROCESSAMENTO

01

1- ESTRUTURA INTERNA E FUNCIONAMENTO DO PROCESSADOR (CPU)

Prezados estudantes, até o momento, já vimos algumas informações sobre o que é um processador. Estudamos também que há uma interação muito grande entre a UCP – unidade central de processamento ou CPU (processador), memória e dispositivos de entrada e saída. Neste módulo iremos nos aprofundar um pouco mais. Vamos observar a estrutura interna e o funcionamento da UCP. Ao final entenderemos que a UCP é composta pela unidade de controle, dos registradores, da unidade aritmética e lógica, da unidade de execução de instruções e das interconexões entre esses componentes.

Geralmente quando o assunto é processador, a comparação clássica é que ele é o cérebro do computador e, embora seja uma comparação interessante e didática, esta definição está um pouco distante do que seja realmente o funcionamento de um processador.

Agora, quando falamos do processador como componente eletrônico, dizemos que ele é um tipo especializado de *microchip* (circuito eletrônico miniaturizado). Destacamos que ele é composto por três componentes principais, que são:

- a ULA – unidade lógica aritmética,
- a UC – unidade de controle e
- os Registradores.

Veremos, oportunamente, cada um desses componentes.

02

As atividades realizadas pela CPU podem ser divididas em duas grandes categorias funcionais: função **processamento** e função **controle**.

Função processamento	Função controle
A função processamento se encarrega de realizar as atividades relacionadas com a efetiva execução de uma operação, ou seja, <u>processar</u> .	A função controle é exercida pelos componentes da CPU que se encarregam das atividades de busca, interpretação e controle da execução das instruções, bem como do controle da ação dos demais componentes do sistema de computação (memória, entrada/saída).

Dentre as várias tarefas importantes da CPU destacam-se as seguintes:

- buscar na memória a instrução a ser executada;
- interpretar que operação a instrução está explicitando;
- buscar os dados onde estiverem armazenados;
- executar efetivamente a operação com os dados e armazenar o resultado no local definido pela instrução;
- reiniciar o processo, buscando a próxima instrução.

Em suma, o processador é um poderoso **dispositivo que realiza cálculos**. As tarefas da CPU consistem, então, em:

- endereçar,
- acelerar,
- preparar dados ou resolver conforme a aplicação ou as demandas.

03

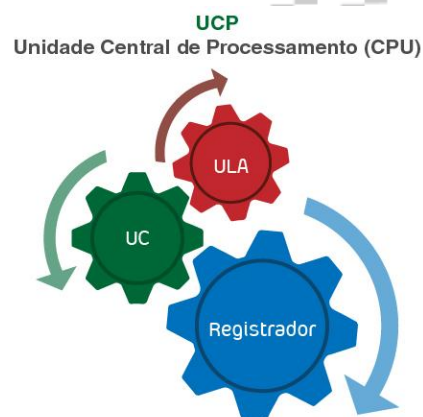
Para ajudar na compreensão sobre processador, assista a este vídeo disponível no YouTube, que trata das características e funções do processador. O vídeo é muito interessante e vale a pena investir alguns minutos assistindo.

<https://www.youtube.com/watch?v=IfOIB4-pn1k>

04

2 - COMPONENTES DO PROCESSADOR

O processador é formado por um conjunto de componentes, sendo que cada um tem um papel específico no processamento de um dado, uma informação, um programa e etc.



Enfim, é este conjunto de componentes que chamamos de processador ou UCP. Sendo assim, o papel principal do processador é **tornar possível a execução dos programas**. A seguir veremos cada um desses componentes.

• ULA - Unidade Lógica Aritmética

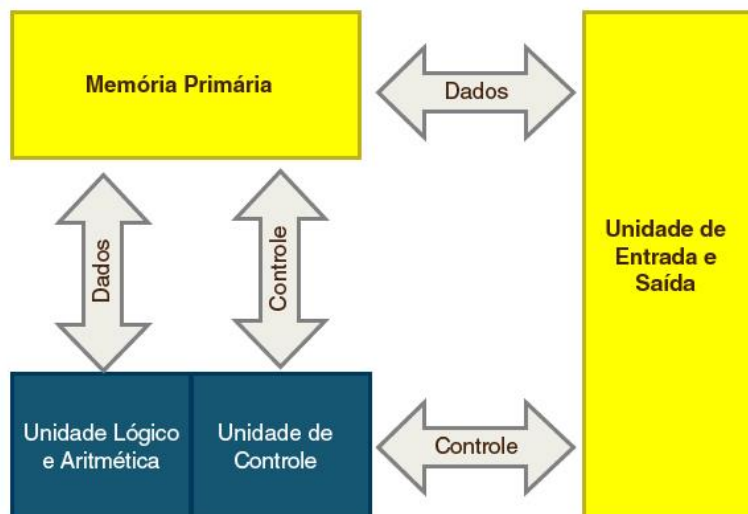
A ULA – unidade lógica aritmética, também chamada de ALU (Arithmetic Logic Unit), é o dispositivo principal da função processamento.

A ULA é a responsável principal por realizar as operações matemáticas e booleanas. Tem por incumbência, como o próprio nome já deixa claro, realizar as operações matemáticas requisitadas durante o processamento (execução) de um programa.

Quando falamos a sigla ULA parece que estamos fazendo referência à dança havaiana HULA, mas ULA (Unidade Lógica e Aritmética), como o próprio nome sugere e também já falamos, é o componente responsável em realizar as operações aritméticas e lógicas sobre os dados. Até poderíamos nos arriscar a usar o exemplo da dança havaiana HULA, para efeito de memorização, já que a ULA é a responsável pela “dança das operações aritméticas e lógicas dos dados” no processador. Todos os componentes de uma máquina (computador), como: registradores, unidade de controle, dispositivos de entrada e saída (E/S) e memória, enfim, todos eles têm como principal serventia trazer os dados para serem processados pela ULA e receber os resultados. Podemos, então, afirmar que a ULA é o núcleo ou a essência de uma máquina (computador).

Assim como todos os demais componentes eletrônicos de um computador, a ULA tem como base dispositivos lógicos digitais simples, os quais têm capacidade de armazenar dígitos binários e efetuar operações simples de lógica booleana.

Observe a figura e veja como a ULA está conectada com o restante do processador e outros componentes do computador.



Tanto os dados fornecidos à ULA, bem como o resultado de uma operação, são armazenados nos registradores. Esses registradores são áreas de armazenamento temporário dentro do processador, conectadas à ULA por meio de caminhos de sinais (barramento). A ULA pode também ativar flags (bits especiais) que indicam o resultado de uma operação.

Caso haja uma operação que exceda a capacidade de armazenamento de um registrador, isso é indicado atribuindo o valor 1 ao bit de overflow (estouro de memória). Cabe à unidade de controle o fornecimento de sinais para fazer o controle da ULA e também pela transferência de dados entre a ULA e os registradores.

Overflow

É um estouro de memória, ou seja, condição que ocorre quando os dados resultantes do processamento ou de entrada necessitam de mais bits do que os que foram fornecidos no hardware ou no software para armazenamento de dados.

07

• UC - Unidade de controle

A unidade de controle é o dispositivo mais complexo da CPU.

A UC é responsável pelo controle das atividades de todos os componentes da máquina (computador).

Um exemplo seria o controle da gravação de um dado no HD (Disco Rígido), outro exemplo seria uma determinada busca de instrução da memória principal.

Além de possuir a lógica necessária para realizar a movimentação de dados e instruções de e para a CPU, a unidade de controle, como o próprio nome sugere, controla a ação da ULA. Os sinais de controle emitidos pela UC ocorrem em vários instantes durante o período de realização de um ciclo de instrução e, de modo geral, todos possuem uma duração fixa e igual, originada em um gerador de sinais comumente conhecido como relógio (clock).

Clock

É o dispositivo gerador de pulsos cuja duração é chamada de ciclo. A quantidade de vezes em que este pulso básico se repete em um segundo define a unidade de medida do relógio, denominada frequência, a qual também usamos para definir velocidade na CPU.

08

• Registradores

Este componente do processador tem por característica ser veloz, localizado fisicamente próximo à UCP, tem por função principal guardar parte das informações que estão na memória principal. Ele funciona como uma memória interna de alta performance ou alto desempenho.

Os registradores funcionam como pequenas unidades de memória, implementadas na CPU, com as seguintes características:



Além dos registradores de dados, a CPU possui sempre outros registradores com funções específicas ou que funcionam para a área de controle, os quais não participam diretamente da função processamento.

Os registradores mais conhecidos são o **contador de programa** (PC) e o **registrador de instrução** (IR). De maneira resumida, cabe ao contador de programa sinalizar qual será a próxima instrução (programa) a ser executada e ao registrador de instrução cabe a execução de fato da instrução (programa).

09

3 - MAIS SOBRE A UCP: MEMÓRIA CACHE, MMU, HERTZ, EVOLUÇÃO

Memória cache

No módulo sobre memória vimos os vários tipos de memórias e suas características. Neste momento, iremos dar destaque para a **memória cache**. Você deve lembrar que ela é a responsável pelo armazenamento das instruções dentro do processador. Quando diante de determinada quantidade de dados que a CPU possa vir a ter que processar, será no cache que serão alocadas as informações que estão sendo constantemente requisitadas. Esta estratégia de armazenar o que está “sendo constantemente requisitado” tem por objetivo agilizar o processamento ao armazenar no processador

os dados que estarão rapidamente acessíveis. Esta técnica evita e torna desnecessário percorrer o HD ou a RAM para buscar este tipo de informação.

MMU (Memory Management Unit)

Este dispositivo é o responsável em transformar as instruções lógicas (virtuais) em endereços físicos na memória RAM (memória principal). A MMU registra onde está localizada cada informação do sistema tornando possível ao processador saber onde buscar os dados e instruções na memória. A MMU é a coordenadora do funcionamento da memória, tornando possível o processador ser rápido ao agilizar o acesso à memória RAM.

10

Comparar processadores era algo que há um tempo atrás ficava restrito a computadores, mas com a popularização de smartphones e tablets esse cenário mudou. O processador é o coração do smartphone e é o responsável por fazer tudo funcionar. É preciso prestar atenção no número dos núcleos (cores) e no clock (quantos GHz) dele para avaliar o seu desempenho. Quanto mais núcleos e mais velocidade, melhor o desempenho do smartphone.

Assim como em um computador, o processador é responsável pela velocidade de execução de tarefas (jogos, vídeos, câmeras, etc.) e aliado à memória e a outros fatores, garante a eficiência e a usabilidade do aparelho.

Assista a um vídeo que fala um pouco sobre o funcionamento do processador de um smartphone.

<http://olhardigital.uol.com.br/embed/saiba-como-funciona-o-processador-de-seu-smartphone/35102>

11

O que é HERTZ?

Durante o tempo que você investiu ao assistir o vídeo anterior deve ter notado que a palavra HERTZ foi mencionada algumas vezes. Provavelmente deve ter observado que “que o equipamento XYZ tem mais ou menos Hertz”.

Você conseguiu entender o que significa um processador ter tantos Hertz?

Então vamos entender!

Veja que interessante esta especificação de um fabricante a respeito de um processador (UCP):

3ª Geração do Processador Intel® Core™ i3-3217U (**1.8GHz**, 3Mb Cache, Placa de vídeo Intel HD Graphics Integrada).

Veja que foi destacado em negrito o 1.8GHz para destacar que, caso você não tenha compreendido o

significa Hertz (Hz), entenda que é o quanto um processador troca dados com o sistema. Utilizando o exemplo acima: Processador Intel® Core™ i3-3217U 1.8GHz . Notamos que ele oferece **1.8GHz**, isto quer dizer que ele é capaz de realizar 1,8 bilhões de ciclos por segundo.

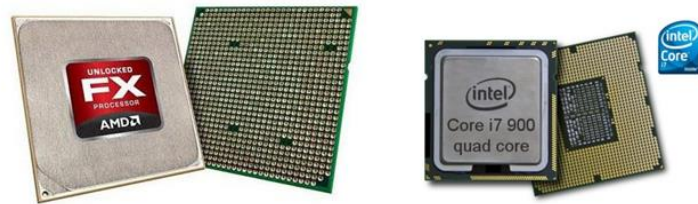
Há um componente fundamental para medir estes ciclos e orientar o fluxo de troca de informações no processador, este componente é o circuito clock, que estabelece os critérios de velocidade do processador.

12

Evolução dos processadores

Para você compreender a evolução dos processadores nos dias atuais, empresas fabricantes de processadores como Intel e AMD (**A**dvanced**M**icro **D**evelopments) estão investindo em processadores chamados APU's (Accelerated Processing Unit), que numa tradução direta significa: Unidade de Processamento Acelerada.

O diferencial em relação aos processadores normais é que as APU's trazem, no mesmo conjunto, CPU e GPU (**U**nidade de **P**rocessamento**G**ráfico ou **G**raphics **P**rocessing **U**nit), oferecendo uma velocidade muito maior de processamento gráfico. Mesmo as APU's trazendo o vídeo integrado, esta solução é diferente do modelo anterior de chips gráficos onboard (onboard, tradução “na placa”, consequentemente, quando você ouvir dizer que um computador é “onboard”, significa que a placa-mãe (motherboard / mainboard) responsável por conectar e interligar todos os componentes do computador) tem um ou mais dispositivos integrados, os quais podem incluir uma placa de rede, vídeo, som, vídeo e etc.



Processador FX da empresa AMD

Processador Core i7 da empresa Intel.

Uma das maiores vantagens é a velocidade de comunicação entre CPU e GPU, aumentando significativamente o tempo de resposta e, em consequência, otimizando todo o sistema.

13

Assista ao vídeo da AMD, disponível na página do Youtube, e veja o que é APU's - Accelerated Processing Unit.

<https://www.youtube.com/watch?v=Fd-TNIhcmbg>

4 -INSTRUÇÕES DE MÁQUINA

Bom, até agora estávamos fazendo uma introdução do que venha ser uma UCP ou CPU, seus componentes e algumas características. Todavia, precisamos nos aprofundar e falar sobre o que são instruções de máquina.

Um computador pode executar tarefas complicadas e sucessivas se for “instruído” sobre o que fazer e em que sequência isso deve ser feito. Os seres humanos, ao receberem uma instrução do tipo “trazer a pasta do funcionário Pedro”, são capazes de localizar o arquivo onde estão as pastas de todos os funcionários – geralmente por ordem alfabética – e encontrar a pasta, trazendo-a a quem solicitou. Nessa tarefa, o nosso cérebro realizou uma série de ações intermediárias para que a instrução fosse concluída com sucesso. Entretanto, se a mesma instrução fosse dada a uma máquina e ela não tivesse qualquer orientação prévia registrada, jamais conseguiria localizar e trazer a pasta solicitada. Para a máquina, é necessário que a instrução seja detalhada em pequenos passos e numa linguagem própria, visto que ela é programada para ser capaz de entender só dessa forma, ou seja, em pequenas operações. No exemplo, a máquina deveria receber um conjunto de instruções específicas para ela, as chamadas **instruções de máquina**.

O conjunto de instruções da máquina constitui o limite em que o projetista e o programador de computadores visualizam em um mesmo computador. Na visão do projetista, o conjunto de instruções de máquina fornece os requisitos funcionais para a CPU. Caso uma pessoa resolva utilizar a linguagem de máquina (Assembly) para programar, o primeiro passo será entender e conhecer sobre a estrutura de memória, sobre o conjunto de registradores da CPU, funcionamento da ULA e os tipos de dados disponíveis diretamente na máquina.

A operação de uma CPU é determinada pelas instruções que ela executa, chamadas de instruções do computador ou **instruções de máquina**. A coletânea de diversas e diferentes instruções capazes de serem executadas pela CPU é chamada de **conjunto de instruções da CPU**.

Internamente, cada instrução de uma máquina (computador) é representada como uma **sequência de bits**. Uma instrução é dividida em campos, correspondentes aos elementos da instrução. A CPU deve ser capaz de extrair os dados dos vários campos da instrução e efetuar a operação requerida.

Torna-se difícil para um desenvolvedor de software (programador) lidar com representações binárias de instruções de máquina. Consequentemente, tornou-se usual a prática de usar uma representação simbólica para instruções de máquina.

Os códigos de operação são representados por abreviações, que indicam a operação a ser efetuada. Veja na tabela abaixo um exemplo:

CÓDIGO	OPERAÇÃO
ADD	Adição.
SUB	Subtração.
MPY	Multiplicação.
DIV	Divisão.
LOAD	Carregar dados da memória.
STOR	Armazenar dados na memória.

Veja um exemplo em que os operandos são representados de maneira simbólica:

ADD W, Z

A instrução acima significa **adicionar o valor contido na posição Z com o conteúdo do registrador W**. Neste caso, Z é um endereço de uma posição de memória e W indica um registrador específico. Podemos notar que a operação é feita sobre o conteúdo da posição de memória, e não sobre seu endereço.

Operando

Em matemática, um operando é uma das entradas (argumentos) de um operador. Por exemplo, em $3 + 6 = 9$, + é o operador e 3 e 6 são os operandos. Em código de máquina é bem parecido. Um operando é um valor (um argumento) no qual a instrução opera. O operando pode ser um registrador, um endereço de memória, uma constante literal, ou um rótulo. Um exemplo simples na arquitetura PC é:

MOV DS, AX

em que o valor no operando registrador AX deve ser movido (MOV) para o registrador DS. Dependendo da instrução, pode haver zero, um, dois ou mais operandos.

16

Consequentemente, há a possibilidade de se escrever um programa utilizando-se linguagem de máquina de maneira simbólica. Cada código de operação simbólico tem um correspondente em representação binária, e cabe ao programador especificar qual o endereço de cada operando simbólico. O programador, por exemplo, pode especificar que Q = 8212 e E = 8213. Uma entrada simbólica pode ser “traduzida”, por meio de um programa, códigos de operações e referências em operandos na forma binária, tendo como resultado instruções de máquinas binárias.

Atualmente, é muito difícil encontrar programadores que gostem de utilizar (programar) utilizando linguagem de máquina. A grande parte dos programas é escrita em linguagem chamadas de **alto nível** (java, python, shell script). Mesmo assim, a linguagem de máquina simbólica ainda permanece sendo uma ferramenta muito útil para a descrição de instruções de máquina.

Um computador (estação de trabalho) deve possuir um conjunto de instruções que possibilita ao usuário criar qualquer tarefa de processamento de dados. Um programa qualquer que utiliza linguagem

de alto nível pode ser traduzido para uma linguagem de máquina, para que depois possa ser executado. Em consequência, o conjunto de instruções de máquina deve ser capaz de expressar qualquer comando de uma linguagem de alto nível.

Operando

Genericamente vários tipos de operandos que podem ser referenciados em uma instrução de máquina:

- Operandos implícitos: são aqueles que, embora não explicitados na área de operandos da instrução, serão utilizados no seu processamento.
- Operandos imediatos: são aqueles cujo valor faz parte da instrução;
- Operandos indiretos: são aqueles onde a referência contida na instrução aponta para o local onde o valor do operando está armazenado.

17

A seguir iremos relacionar os **tipos de instruções de máquina**:

- Movimentação de dados: instruções de memória;
- Processamento de dados: instruções aritméticas e lógicas;
- Armazenamento de dados: instruções de memória;
- Controle: instruções de desvio e de teste.

Instruções aritméticas proporcionam capacidade computacional para processar dados numéricos.

As **instruções booleanas** operam sobre bits de uma PALAVRA; disponibilizam, então, uma capacidade para processar todo tipo de dado que o usuário desejar. Essas operações são primariamente efetuadas em dados guardados em registradores da CPU, consequentemente, é isto que motiva a existência de instruções de memória para mover dados entre os registradores e a memória e vice-versa.

Para a transferência de programas e dados para a memória e para a transferência dos resultados retornando para o usuário é que há as **instruções de E/S** (Entrada e Saída). **Instruções de teste**, como o próprio nome diz, são para fazer o teste do valor de uma palavra de dados ou o estado de uma execução. Para o desvio da execução de um programa para uma nova instrução é que há as **instruções de desvio**, geralmente em detrimento do resultado de um teste.

Os campos de endereço, no formato de instrução típico, são relativamente pequenos. Com o intuito de possibilitar a referência a uma grande quantidade de posições de MP, técnicas variadas de endereçamento vêm sendo empregadas. A maioria dessas técnicas envolve decisões que contrapõem a quantidade de posições de memória endereçáveis e ou a flexibilidade de endereçamento ao número de referências à memória em cada instrução e/ou à complexidade do cálculo de endereços.

18

O formato de uma instrução especifica como será a disposição dos bits da instrução, conforme ele é formado.

Um formato de instrução deve ter um código de operação e zero ou mais operandos, explicitamente ou implícito.

Cada operando explícito é feito referência utilizando-se os modos de endereçamento (endereçamento direto, registrador, indireto, deslocamento, pilha). O formato deve ser um indicador de maneira explícita ou implícita do modo de endereçamento para cada operando. São utilizados mais de um formato de instrução na maioria dos conjuntos de instruções.

Para começo de um projeto sobre o tamanho de instrução, a questão primordial deste projeto é o tamanho do formato de instrução. Essa escolha interfere, e sofre interferência, pelo tamanho e pela maneira que a memória é organizada, pela complexidade e velocidade da UCP e pela estrutura de barramento (responsável em conectar à placa-mãe todos os componentes de um computador, como processadores, memórias, placas de vídeo).

O dilema que se vislumbra é o conflito entre o desejo de oferecer uma variedade poderosa de instruções e a necessidade de economia de espaço. Programadores almejam mais operandos, mais códigos de operação, mais modos e maior espaço de endereçamento. Quantidade maior de operandos e códigos de operação facilitam o trabalho de programar, possibilitando escrever programas menores para executar determinada tarefa. Da mesma forma, mais modos de endereçamento dão ao programador flexibilidade maior na implementação de certas funções, tais como manipulação de tabelas e desvios com múltiplos endereços de destino.

Saiba+

Em virtude do aumento da capacidade de armazenamento da MP e o uso crescente de memória virtual (somatório da MP e parte do HD), os programadores desejam endereçar uma maior área de memória. Códigos de operação, modos de endereçamento, operandos, endereços, requer bits na instrução, como consequência, instruções de maior tamanho. Todavia, instruções maiores podem significar desperdício, uma instrução de 32 bits ocupa duas vezes mais espaço que uma de 16 bits.

19

O tamanho da instrução deve ser proporcional ao tamanho da unidade de transferência de dados de e para a memória ou um deve ser múltiplo do outro. Caso contrário, não será obtido um número inteiro de instruções durante um ciclo de busca. Um aspecto importante é a taxa de transferência da memória. Essa taxa não está acompanhando o aumento da velocidade dos processadores. Então, a memória passará a ser o gargalo caso o processador seja mais rápido na execução das instruções. Uma solução adotada é usar instruções menores ou o uso de memória cache.

Fica explícito que fixar o tamanho de instrução conflita com a capacidade de endereçamento e o número de códigos de operação. Número maior de códigos de operação impacta em mais bits no campo de código.

O projetista se vê, então, diante de um conjunto de fatores que deverão ser considerados e equilibrados. Ainda não está bem definido até onde é crítico a escolha de uma ou outra alternativa.

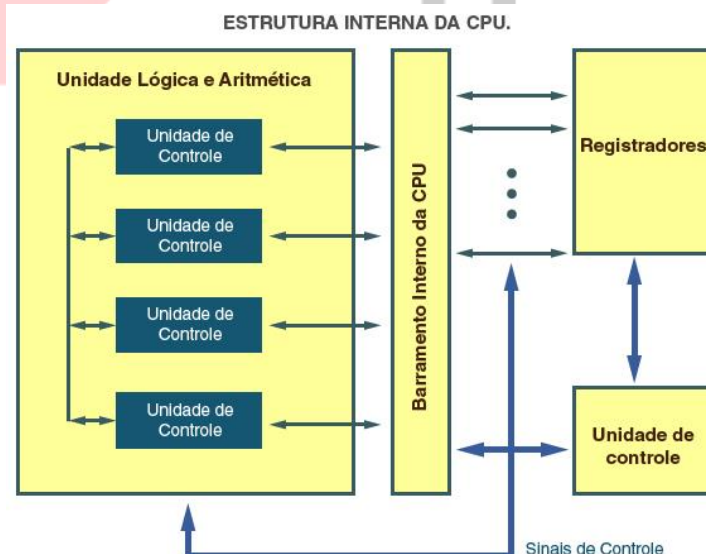
Para entendermos como é a organização da UCP (CPU), é necessário considerar as seguintes **ações** que devem ser executadas:

- **Busca de dados:** a execução de determinada instrução poderá solicitar a leitura de dados de um módulo de E/S (Entrada e Saída) ou leitura de dados da memória.
- **Interpretação de instrução:** a instrução propriamente dita;
- **Busca de instrução:** a CPU faz a leitura de determinada instrução da memória;
- **Escrita de dados:** os resultados da execução poderão solicitar escrita dos dados em um módulo de E/S ou na memória.
- **Processamento de dados:** execução de determinada instrução pode vir a solicitar que seja efetuada uma operação lógica ou aritmética sobre os dados.

20

Para a execução dessas ações, a UCP necessita guardar temporariamente alguns dados. Ela manterá a posição de memória da última instrução, para ter ciência da próxima instrução, e necessita também de armazenar temporariamente dados e instruções, enquanto determinada instrução é executada. Em suma, a UCP precisa de uma pequena memória interna. Você deve estar lembrado que os componentes mais importantes da UCP (CPU) são a ULA (unidade lógica e aritmética) e a UC (unidade de controle). Cabe à ULA efetuar o processamento de dados e a UC o controle da operação da ULA como também pela transferência de dados e instruções para fora e para dentro da UCP.

A figura abaixo mostra o **barramento interno** da UCP, que nada mais é do que os caminhos de transferência de dados e de sinais de controle. Cabe a ele ser o responsável pela transferência dos dados entre os vários registradores e a ULA, uma vez que a ULA opera apenas com dados localizados na memória interna da UCP. A figura também mostra os elementos básicos de uma ULA.



21

Como estudamos anteriormente um sistema computacional emprega uma hierarquia de memória. Nos níveis mais altos a memória é mais rápida, menor e com o custo maior por bit.

Na UCP, há um conjunto de registradores que funciona como um nível da hierarquia de memória acima da MP e da memória cache. Os **registradores da UCP possuem duas funções**:

Registradores de controle e de estado	Registradores visíveis para o usuário
Usados pela UC para controlar a operação da UCP e por programas com privilégios do sistema operacional com o objetivo de controlar a execução de programas.	Tornam possível ao programador que utiliza linguagem de máquina diminuir as referências à memória, pela otimização do uso de registradores.

22

RESUMO

Durante o estudo deste módulo você compreendeu que um processador é um poderoso dispositivo que realiza cálculos. Conforme estão sendo processadas as instruções armazenadas na memória interna do processador, ele vai respondendo a esse volume e faz o processamento da informação. Estudamos que o processador é formado por um conjunto de componentes (ULA – Unidade Lógica Aritmética, UC – Unidade de controle, e Registradores), sendo que cada um tem um papel específico no processamento de um dado, uma informação, um programa e etc. A ULA é responsável pela execução das instruções dos programas. A Unidade de Controle (UC) é responsável por controlar as ações a serem realizadas. O registrador agiliza e armazena parte das informações que estão na memória principal, no intuito de evitar acesso contínuo a esta memória. Para finalizar, vimos que o conjunto de instruções da máquina constitui o limite em que o projetista e o programador de computadores visualizam em um mesmo computador. Na visão do projetista, o conjunto de instruções de máquina fornece os requisitos funcionais para a CPU. Na visão do usuário, que programar em linguagem de máquina (Assembly), há a necessidade do conhecimento básico sobre a estrutura de memória, sobre o conjunto de registradores da CPU, funcionamento da ULA e os tipos de dados disponíveis diretamente na máquina. A operação de uma CPU é determinada pelas instruções que ela executa, chamadas de instruções de máquina. A coletânea de diversas e diferentes instruções capazes de serem executadas pela CPU é chamada de conjunto de instruções da CPU.

UNIDADE 3 – FUNCIONAMENTO E COMPONENTES BÁSICOS

MÓDULO 4 – DISPOSITIVOS DE ENTRADA E SAÍDA E BARRAMENTOS

01

1 - MÓDULOS DE ENTRADA E SAÍDA

Neste módulo iremos fazer uma pequena introdução a respeito dos dispositivos de entrada e saída (E/S). Veremos que é um dispositivo de entrada/saída, como são os métodos de realização de operações de E/S (entrada/saída) e o importante papel do barramento na comunicação de todos estes dispositivos.

Além da memória e do processador, outro elemento importante e fundamental de um sistema computacional é o **conjunto de módulos de E/S (entrada e saída)**. Cada módulo está conectado por meio do barramento ao sistema ou com o computador central e controla um ou vários dispositivos periféricos.

Um módulo de entrada e saída (E/S) não é apenas um conjunto de conectores mecânicos que estão ligando um dispositivo ao barramento do sistema. Na verdade há um certo “raciocínio” ou uma inteligência, ou seja, há uma lógica dedicada ao desempenho da função de comunicar um periférico ao barramento.

Barramento

É um conjunto de linhas de comunicação que permitem a interligação entre dispositivos, como a CPU, a memória e outros periféricos.

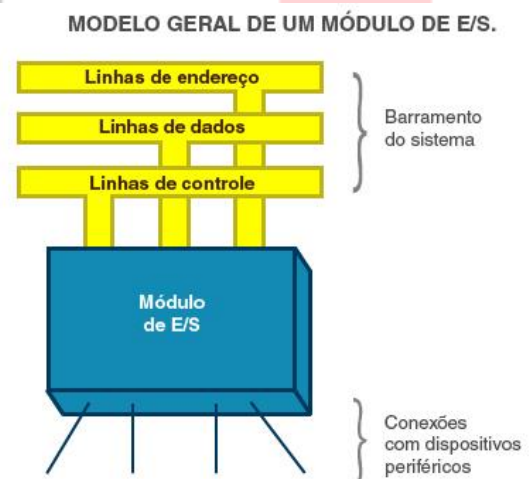
02

Você poderá ficar em dúvida e questionar o motivo dos periféricos não estarem diretamente conectados ao barramento do sistema. A resposta é simples e abaixo seguem os motivos:

- Há uma gama de periféricos, com vários e diferentes mecanismos de operação. Não seria possível incorporar ao processador a lógica necessária para poder controlar dispositivos de vários tipos.
- Em virtude da taxa de transferência de dados dos periféricos, na maioria das vezes, ser bem menor do que a taxa de transferência de dados do processador ou da memória, fica impossível utilizar barramentos do sistema de alta velocidade para comunicar-se diretamente com um periférico.
- Na maioria das vezes os periféricos utilizam tamanhos de palavras e formatos de dados diferentes dos utilizados na máquina (computador) em que estão conectados.

Em virtude das razões expostas acima, torna-se necessário um módulo de entrada e saída (E/S) que irá desempenhar as seguintes **funções**(observe a figura abaixo):

- Fazer a interface com um ou mais dispositivos periféricos, por meio de conexões adequadas de dados.
- Oferecer uma interface com a memória e com o processador, por meio do computador central ou pelo barramento.



Comutador

Dispositivo usado para inverter o sentido de uma corrente. Por exemplo, um interruptor.

03**2 - DISPOSITIVOS EXTERNOS**

Agora vamos falar dos dispositivos externos. As operações de entrada e saída são efetuadas utilizando-se uma gama de dispositivos externos, que disponibilizam um meio para efetuar a troca de dados entre o computador e o ambiente externo. Por meio de uma conexão de um módulo de entrada e saída um dispositivo externo, é conectado, por exemplo, uma impressora (conforme figura acima). Para a transferência de dados, informações de controle e informações de estado entre o módulo de entrada e saída utiliza-se esta conexão. Chama-se de periférico ou dispositivo periférico o dispositivo externo conectado a um módulo de E/S (entrada e saída).

Classificação dos dispositivos Externos

Os dispositivos externos são classificados segundo sua finalidade ou sua função. Podem ser:

- Voltados para a comunicação com dispositivos remotos;
- Voltados para a comunicação com a máquina;
- Voltados para a comunicação com o usuário.

O disco magnético (HD) é um exemplo de dispositivo voltado para a comunicação com a máquina.



Provavelmente você agora deve estar se perguntando: já estudei que HD (disco magnético) também exerce função de memória, então, HD é um dispositivo de E/S ou uma memória secundária?

Pois é, esta pergunta é totalmente pertinente e vamos explicar. O HD exerce as duas funções, de dispositivo de E/S e também de memória secundária, isto ocorre em virtude de como é o modelo de arquitetura de um sistema computacional (conjunto de hardware/software). Pela visão funcional, o HD faz parte da hierarquia de memória, e por outro lado, pela visão estrutural, o HD é controlado por um módulo de entrada e saída (E/S).

Comunicação com o usuário

São exemplos de dispositivos voltados para a comunicação com o usuário: impressoras e terminais de vídeo.

Comunicação com a máquina

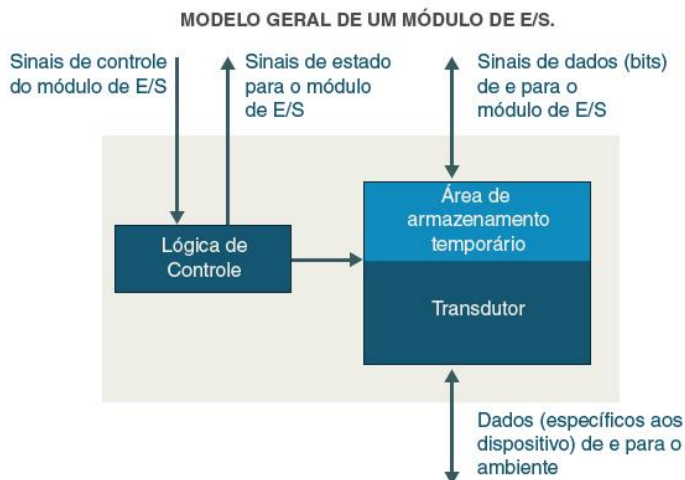
Além do HD, há outros exemplos de dispositivos voltados para a comunicação com a máquina: sensores e controladores utilizados em robótica (estudo sobre utilização de robôs).

Comunicação com dispositivos remotos

Os dispositivos que são voltados para a comunicação com dispositivos remotos tornam possível um computador trocar dados remotamente com outros dispositivos, estes dispositivos podem ser voltados para a comunicação com o usuário, assim como um terminal, poderá ser com outra máquina ou uma comunicação interna.

04

A figura abaixo é um exemplo de modelo de um **dispositivo externo**.



A **lógica de controle** associada ao dispositivo é que efetivamente faz o controle de sua operação, em resposta a um INPUT (comando recebido) no módulo de entrada e saída. Um transdutor é utilizado na conversão de dados codificados como sinais elétricos para alguma outra forma de energia, em uma operação de saída, ou dessa outra forma de energia para sinais elétricos, em uma operação de entrada. Normalmente, está associada ao transdutor uma **área de armazenamento temporário** para a transferência de dados entre o módulo de entrada e saída e um ambiente externo.

A interface com o módulo de entrada e saída é feita por:

SINAIS DE DADOS	Os dados compõem um conjunto de bits que serão recebidos ou remetidos para um módulo de entrada e saída (E/S).
SINAIS DE ESTADO	São a indicação de qual é o estado do dispositivo. Exemplo: os sinais READY (pronto) /NOT-READY (não-pronto) indicam o estado do dispositivo, informando se está pronto ou não para executar uma transferência de dados.
SINAIS DE CONTROLE	Determinam qual deverá ser a função a ser executada pelo dispositivo, assim como: <ul style="list-style-type: none"> - remeter dados para o módulo de entrada e saída (E/S) – INPUT (conexão de entrada) ou READ (leitura)-, - receber dados do módulo de entrada e saída (E/S) OUTPUT (conexão de saída) ou WRITE (escrita), - dizer qual o estado do dispositivo ou desempenhar uma função de controlar um dispositivo em particular (exemplo: movimentar o cabeçote do HD).

Transdutor

Dispositivo que recebe um sinal e o retransmite, independentemente de conversão de energia.

Interface

É o conjunto de meios planejados, sendo físicos ou lógicos com o objetivo de fazer a adaptação entre dois sistemas.

05**Interação usuário/máquina**

De forma geral a maneira mais usual de interação entre um usuário e sua máquina (computador) é feita por meio da combinação **teclado- mouse-monitor de vídeo**, embora já se percebam outras formas de interação, como o reconhecimento de voz. Inclusive, atualmente, a interação está sendo via *touch* (toque) ou tecnologia *touchscreen* (tela sensível ao toque). Os dados são oferecidos pelo usuário por meio do teclado ou pelo toque. Essa entrada é então transmitida ao computador, podendo também ser mostrada no monitor de vídeo. Além do mais, o monitor mostra os dados fornecidos como um resultado das operações realizadas pela máquina (computador).

É um caractere a unidade básica de troca de dados. A cada caractere está associado um código tipicamente de 7 ou 8 bits. A codificação mais usada é um código de 7 bits conhecido como ASCII (código americano para troca de informações). Cada caractere é representado por um código binário distinto de 7 bits; dessa maneira, podem ser representados até 128 caracteres diferentes. A tabela abaixo é um exemplo da correspondência das letras do alfabeto e o código ASCII. Veja que a letra **A** seu correspondente de 7 bits é = **1000001**.

ASCII Alphabet			
A	1000001	N	1001110
B	1000010	O	1001111
C	1000011	P	1010000
D	1000100	Q	1010001
E	1000101	R	1010010
F	1000110	S	1010011
G	1000111	T	1010100
H	1001000	U	1010101
I	1001001	V	1010110
J	1001010	W	1010111
K	1001011	X	1011000
L	1001100	Y	1011001
M	1001101	Z	1011010

Reconhecimento de voz

Você já recebeu atendimento eletrônico no qual ele pedia que você “falasse a opção desejada”? Ou conhece algum smartphone que efetua sozinho uma ligação para a pessoa cujo nome é pronunciado por você? Esses programas ouvem o som emitido pela sua voz, classificam as sílabas e é aplicado um método de busca para associar estas informações com padrões de palavras a fim de encontrar semelhanças.



Como funciona: para que o computador reconheça o som da sua voz juntamente com a fonética da palavra pronunciada e efetue a aplicação desejada, ele precisa encadear uma sequência de passos. Primeiro ele precisa digitalizar a fala que se quer reconhecer. Para isso, ele utiliza um conversor analógico-digital que capta as vibrações criadas pela sua voz e converte essas ondas em dados digitais. Em seguida, aplica-se uma medida para cada uma das ondas captadas e o som digitalizado é filtrado para separá-lo de ruídos e interferências. Então, efetua-se uma computação das características que representam o domínio espectral (frequências) contido na voz. Nessa etapa do processo, o som pode necessitar ser sincronizado, pois as pessoas não costumam utilizar o mesmo tom e nem sempre falam na mesma velocidade. Isso consiste em um ajuste com modelos de som já armazenados na memória do classificador.

Então essa digitalização é separada em frações ainda menores, ou seja, sons fonéticos não maiores do que uma sílaba. Em seguida, o programa compara os sons captados com fonemas conhecidos e presentes em seu banco de dados que correspondam ao idioma que o locutor tenha falado. Em outras palavras, é aplicado um método de busca para associar as saídas com padrões de palavras e da voz de quem as emitiu.

Finalmente, o sistema analisa o resultado e o compara com palavras e frases conhecidas e, como resultado, ele identifica o que seu usuário disse e converte para a funcionalidade desejada (texto em uma planilha, um comando, o reconhecimento do usuário, etc.).

06

Quando uma tecla é acionada pelo usuário, é gerado um sinal eletrônico, que ao ser interpretado pelo **transdutor** do teclado e faz a tradução conforme o padrão de bits correspondente ao código ASCII. Sendo assim, esse padrão de bits será transmitido para o módulo de E/S da máquina (computador). Um trabalho que você digitou poderá ser armazenado na máquina, codificado em ASCII.

Durante uma operação de saída de dados, os caracteres em código ASCII são transmitidos para um dispositivo externo através do módulo de E/S. O **transdutor** desse dispositivo realiza a interpretação dos

códigos e os envia, por meio de sinais eletrônicos correspondentes, para o dispositivo de saída, que então irá executar alguma função de controle solicitada ou irá exibir o caractere indicado.

A unidade de disco (HD) possui circuitos eletrônicos para troca de sinais de controle, de estados com o módulo de E/S, troca de dados, além de circuitos controladores do mecanismo de escrita e leitura do disco. Em um HD de cabeçote fixo, o **transdutor** faz a conversão dos padrões magnéticos da superfície do HD para bits da área de armazenagem temporária do dispositivo e vice-versa. Um HD de cabeçote móvel deve também ter a capacidade de mover o braço do disco na radialmente, sobre a superfície do HD.

07

3- FUNÇÕES DO MÓDULO DE ENTRADA/SAÍDA

As funções mais importantes de um módulo de entrada e saída (E/S) são:

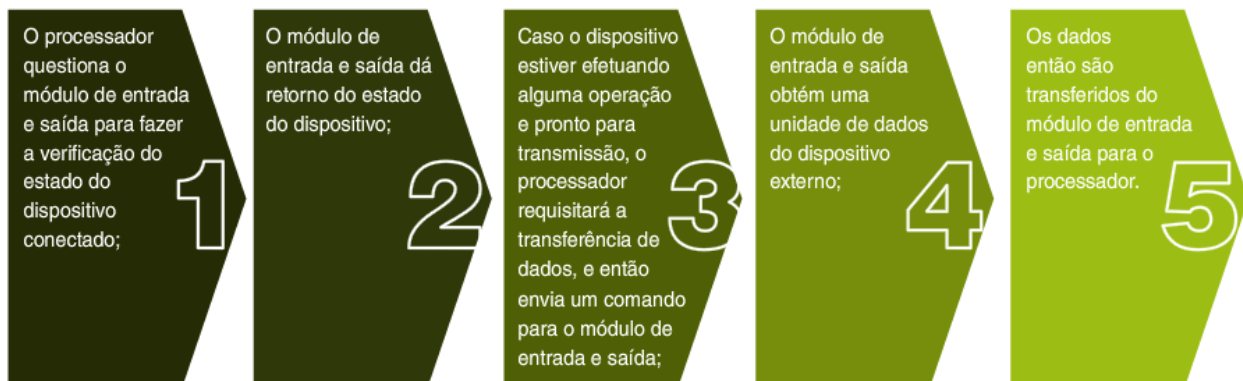
- Comunicação com dispositivos;
- Comunicação com o processador;
- Controle e temporização;
- Detecção de erros;
- Área de armazenamento temporário de dados.

O processador pode estabelecer uma comunicação em qualquer hora com um ou mais dispositivos externos, conforme as necessidades de entrada e saída do programa.

Os recursos internos do sistema, como o barramento e a MP, são compartilhados para haver a realização de várias atividades, incluindo a entrada e saída de dados. Em virtude disso, um módulo de E/S inclui funções de temporização e controle, para efetivamente monitorar o fluxo de dados entre os recursos internos e os dispositivos externos.

08

Para exemplificar, o controle de transferência de dados de um dispositivo externo para o processador envolve as seguintes etapas sequenciais:



Caso um barramento esteja sendo usado pelo sistema, cada uma das interações entre o módulo de entrada e saída e o processador irá envolver uma ou mais arbitrações (priorização) do barramento.

Para que haja comunicação com o processador há necessidade de serem observados os seguintes itens:

Decodificação de comando

O módulo de E/S recebe comandos do processador, geralmente enviados como sinais, por meio do barramento de controle. Exemplo, um módulo de entrada e saída conectada a um HD poderá aceitar os seguintes comandos: leitura e escrita de um setor, busca de determinada trilha e pesquisa de determinado registro identificador.

Dados

Por meio do barramento de dados há transferências entre o processador e o módulo de entrada e saída.

Informações de estado

Geralmente os periféricos são muito lentos, em virtude disso, é necessário conhecer o estado do módulo de entrada e saída. Exemplo, um módulo de E/S não está pronto para uma operação de leitura em virtude de estar ainda processando um comando de entrada e saída anterior. Essa situação é enviada como um sinal de estado, podendo ser o sinal BUSY (ocupado) e READY (pronto). Há a possibilidade de haver também sinais para indicar condições de erro

Reconhecimento de endereço

Cada dispositivo de entrada e saída possui um endereço, assim como uma palavra de memória. Dessa forma, o módulo de entrada e saída deve ser capaz de reconhecer distintamente um endereço para cada periférico que esteja sob seu controle.

Uma tarefa primordial de um módulo de entrada e saída é o de **armazenar temporariamente os dados**. Enquanto a taxa de transferência de dados entre a MP e o processador é muito alta, entretanto, as taxas da maioria dos dispositivos periféricos são bem menores e variam em valores. A transferência de dados da MP para o módulo de entrada e saída é feita de forma muito rápida. Esses dados são armazenados de forma temporária no módulo de entrada e saída e, conseqüentemente, remetidos em uma taxa adequada para o dispositivo periférico. No sentido oposto à transmissão os dados também são armazenados de forma temporária no módulo de entrada e saída, para não manter a memória numa taxa baixa de transferência de dados. Dessa forma, o módulo de entrada e saída deverá ter a capacidade de realização de operações tanto à velocidade de um dispositivo externo como à velocidade da memória.

Cabe também ao módulo de entrada e saída (E/S) ser o responsável pela **detecção de erros** e do envio desta informação para o processador. A detecção de erro alerta sobre mau funcionamento elétrico ou mecânico sinalizado pelo dispositivo, assim como alguma alteração no padrão de bits transmitido por um dispositivo para o módulo de entrada e saída. É utilizado um tipo de código de detecção de erros para detectar erros de transmissão, podendo ser utilizado um bit de paridade em cada caractere de dados.

Para exemplificar, sabemos que um código em ASCII ocupa 7 dos 8 bits de um byte e o valor que será atribuído ao oitavo bit de tal maneira que o número total de “1” no byte seja par (paridade par) ou ímpar (paridade ímpar). Então, ao receber um byte, o módulo de entrada e saída verifica a paridade e determina se houve erro ou não.

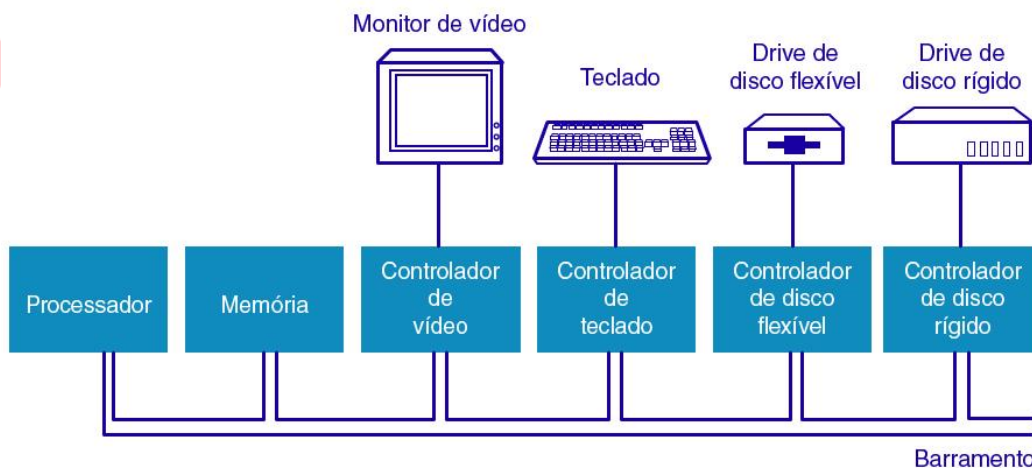
11

4 – BARRAMENTO E CONTROLADOR

Barramentos são estruturas físicas localizadas nas placas eletrônicas (exemplo: placa-mãe) que têm por finalidade ser uma interface de comunicação entre os vários tipos de periféricos que fazem parte de um computador (máquina).

Observando fisicamente uma estrutura de uma máquina, vemos que é composta por uma caixa de metal, uma placa de circuito impresso chamada de **placa-mãe**. Nessa placa há muitos chips, entre eles o chip processador e alguns slots nos quais os DIMM (Módulo de Memória em Linha Dupla), poderão ser inseridos. Há também na placa mãe um barramento que percorre todo o seu comprimento, e soquetes pelos quais são feitas as ligações dos conectores das placas aos controladores dos dispositivos de entrada e saída.

Veja a figura abaixo que é um exemplo de um barramento único, onde estão conectados o processador à memória e os dispositivos de entrada e saída (E/S).



12

Cada dispositivo de entrada e saída (E/S) possui um circuito chamado de controlador que está ligado por meio de um slot livre à placa mãe, isto poderá ser para qualquer dispositivo de entrada e saída com exceção para os dispositivos on-board.

A função do **controlador** é administrar o dispositivo de entrada e saída quanto ao acesso ao barramento.

Ao executar uma determinada tarefa, ele inicia diversos comandos através dos barramentos de uma máquina aos vários dispositivos do sistema, que por meio de seus controladores irão obter esses dados e retornarão um bit após o outro. É uma das atribuições do controlador agrupar esses bits em unidades e os escrever na memória à medida que é reconhecida cada unidade.

Um controlador poderá não utilizar o processador para ler e escrever dados na memória, esta operação é chamada de DMA (**D**irect **M**emory**A**ccess) que quer dizer **acesso direto à memória**.

Não são apenas os controladores de E/S que utilizam o barramento, o processador também o utiliza para realizar busca de informações e instruções iniciadas por algum programa em execução.

No barramento está localizado um circuito eletrônico inteligente que é capaz de dar prioridade na sequência das informações enviadas e recebidas, isto ocorre devido à presença de um chip chamado de **árbitro do barramento**, como se fosse um sinal de trânsito, que estabelece de quem é a vez. Geralmente os dispositivos de entrada e saída têm a preferência, deixando o processador em segundo lugar, pois o HD e outros dispositivos que utilizam dispositivos mecânicos (movimentar cabeça de leitura), não podem sofrer interrupções durante seus trabalhos, consequentemente forçá-los a ficar em modo de espera acarretará em perdas de dados.

Slot

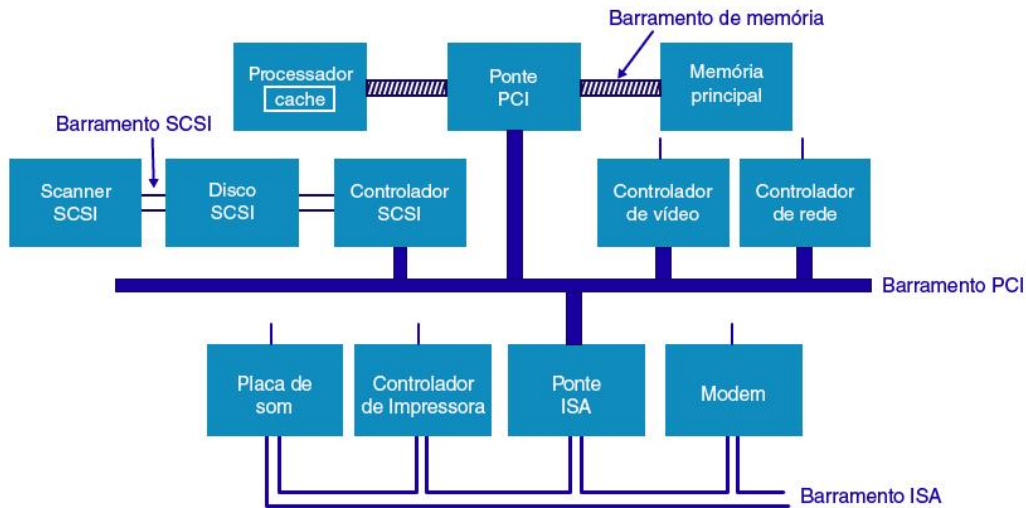
Termo em inglês que significa: ranhura, fenda, conector, encaixe. Sua função é conectar os periféricos ao barramento.

13

Quando uma operação de entrada e saída estiver sendo realizada, haverá acesso prioritário garantido ao barramento o dispositivo que estiver efetuando a tarefa. Esse processo é chamado de **Roubo de ciclo**, que influencia diretamente na velocidade da máquina (computador).

São exemplos de barramentos: Barramento ISA (**I**ndustry **S**tandard **A**rchitecture); EISA (**E**xtended **I**ndustry **S**tandard **A**rchitecture), PCI (**P**eripheral **C**omponent **I**nterconnect) e SATA (**S**erial **A**dvanced **T**echnology **A**ttachment).

Observe na figura abaixo um exemplo de barramento PCI:



Nesta figura é possível visualizar que o processador se comunica com a memória e também com os controladores por meio de uma conexão de alta velocidade e dedicada.

A empresa americana de processadores INTEL definiu o barramento PCI com suporte para interconexão de componentes e periféricos com o intuito de garantir que o mercado não se tornasse uma variedade enorme de tipos de arquitetura de barramento local, que foram lançados em curto período de tempo.

14

RESUMO

Estudamos e compreendemos durante o desenvolvimento deste módulo que, além da memória e do processador, outro elemento importante e fundamental de um sistema computacional é o conjunto de módulos de E/S (entrada e saída). As operações de entrada e saída são efetuadas utilizando-se uma gama de dispositivos externos, que disponibilizam um meio para efetuar a troca de dados entre o computador e o ambiente externo. Estudamos que há três categorias de classificação dos dispositivos externos que são: os voltados para a comunicação com dispositivos remotos; os voltados para a comunicação com a máquina e os voltados para a comunicação com o usuário.

Quando você clica em uma tecla qualquer é gerado um sinal eletrônico que ao ser interpretado pelo **transdutor** do teclado que irá fazer a tradução conforme o padrão de bits correspondente ao código ASCII.

Uma das funções mais importantes de um módulo de entrada e saída (E/S) divide-se nas seguintes categorias: Comunicação com dispositivos; Comunicação com o processador; Controle e temporização; Detecção de erros e Área de armazenamento temporário de dados.

Para finalizar vamos lembrar que ao falarmos em Barramentos precisamos entender que são estruturas físicas localizadas nas placas eletrônicas (exemplo: placa-mãe) que tem por finalidade ser uma interface de comunicação entre os vários tipos de periféricos que fazem parte de um computador (máquina).