

Sistemas de Informação
Disciplina: Arquitetura e Organização de Computadores - 1º Período
Professor: José Maurício S. Pinheiro

AULA 3: Subsistema de Memória

De acordo com o modelo de Von Neumann, a função da Unidade Central de Processamento (UCP), ou CPU, é capturar dados e instruções que compõem um programa e processá-los, não importando sua origem ou destino. Mas para que o processador possa executar os programas, seus dados e instruções devem estar armazenados na memória. Portanto, a memória dos computadores é um elemento indispensável e tão importante quanto o processador.

1. Memórias

A memória é a parte do computador onde os programas e os dados são armazenados e permite ao computador armazenar dados de forma temporária ou permanente. Sem uma memória na qual os processadores (CPU) possam ler ou escrever informações, o conceito de computador digital com programa armazenado não pode ser implementado.

Para o funcionamento adequado de um computador, são necessários diferentes tipos de memória. Em algumas tarefas é fundamental que a transferência de dados seja feita da forma mais rápida, como as tarefas realizadas pela CPU, onde a velocidade é fator preponderante, ao passo que a quantidade de bits a ser manipulada é muito pequena. Esse tipo de memória deve possuir características diferentes daquele em que a capacidade de armazenamento é mais importante que a sua velocidade de transferência de e para outros dispositivos.

A necessidade da existência de diferentes tipos de memória ocorre em virtude de vários fatores concorrentes, mas principalmente em função do aumento da velocidade das CPU's (a qual é muito maior do que o tempo de acesso da memória) e da capacidade de armazenamento. Se existisse apenas um tipo de memória, sua velocidade deveria ser compatível com a da CPU, de modo que esta não ficasse esperando muito tempo por um dado que estivesse sendo transferido. A CPU manipula um dado em 5 ns, ao passo que a memória transfere um dado em 60 ns.

Considerando os diversos tipos de memórias existentes, as quais variam em função de sua tecnologia de fabricação, capacidade de armazenamento, velocidade e custo, pode-se dizer que fica muito difícil projetar um computador utilizando-se apenas um único tipo de memória. Dessa forma, o computador possui diferentes tipos de memórias, as quais se encontram interligadas de forma bem estruturada, constituindo o que é chamado de subsistema de memória, o qual é parte do sistema computacional.

O subsistema de memória é projetado de modo que seus componentes sejam organizados hierarquicamente. A Figura 1 apresenta uma pirâmide contendo a hierarquia das memórias existentes em um computador. Observa-se que a base da pirâmide é larga, simbolizando a elevada capacidade de

armazenamento, o tempo de uso, a velocidade e o custo de sua tecnologia de fabricação. A base da pirâmide representa dispositivos de armazenamento de massa (memória secundária), de baixo custo por byte armazenado, mas ao mesmo tempo com baixa velocidade de acesso. Em direção ao topo indica que quanto mais rápidas forem as memórias, mais elevado será o seu custo em relação à tecnologia e menor a sua capacidade de armazenamento em um computador.



Figura 1 - Hierarquia de memórias

1.1. Memórias Voláteis e Não-Voláteis

Memórias voláteis são as que requerem energia para manter a informação armazenada, ao contrário das memórias permanentes ou não-voláteis, que mantêm o conteúdo mesmo após o desligamento do sistema.

A Tabela 1 apresenta as características básicas de cada tipo de memória.

Tabela 1 - Características dos diferentes tipos de memórias

	Localização	É Volátil?	Velocidade	Capacidade de armazenamento	Custo por bit
Registrador	Processador	Sim	Muito alta (opera na velocidade do processador)	Muito baixa (Bytes)	Muito alto
Cache	Processador	Sim	Alta (opera na velocidade do processador)	Baixa (KB)	Alto
Principal	Placa-mãe	RAM – sim ROM – não	Depende do tipo de memória instalada	Média (MB)	Médio (tem caído muito)
Secundária	HD, CDs, etc.	Não	Baixa (lenta)	Alta (GB)	Baixo (tem caído muito)

2. Registradores

A função da memória é armazenar dados destinados a serem utilizados pelo processador. O processador busca dados e instruções e os aloca temporariamente em seu interior para que possa realizar as operações solicitadas utilizando seus demais componentes. Os registradores são os locais onde esse conteúdo fica armazenado.

O conceito de registrador surgiu da necessidade de o processador armazenar temporariamente dados intermediários durante um processamento quando o resultado de operação precisa ser armazenado até que o resultado de uma busca da memória esteja disponível para com ele realizar uma nova operação (Figura 2).



Figura 2 - Funcionamento do registrador

Por exemplo, quando a instrução de soma de uma variável qualquer (A e B) deve ser executada, antes de a soma ocorrer, as variáveis são carregadas da memória RAM para dois registradores e só depois a soma é realizada. Isso acontece por que somente os registradores possuem comunicação direta com a CPU. O resultado dessa soma será armazenado em outro registrador antes de ser carregado na memória RAM. Esses registradores são chamados de GPR (*General Purpose Register*) ou Registradores de Propósito Geral.

Além dos registradores que não possuem uma função especial ou pré-determinada, todo microprocessador possui um conjunto de registradores com funções pré-definidas pelo fabricante, são os SFR (*Special Function Register*) ou Registradores de Funções Especiais. Os bits são conectados diretamente a circuito internos tais como: Temporizadores, conversores A/D e osciladores. Isso significa que eles estão diretamente no comando da operação do hardware sem precisar da intervenção da CPU, ou seja para modificar alguma característica do hardware que esteja associado ao SFR basta modificar os bits desse registrador. No caso de registradores de 8 bits cada SFR funcionara como 8 chaves (liga/desliga) que comandam alguns pequenos circuitos dentro do micro controlador, conforme exemplifica a Figura 3.

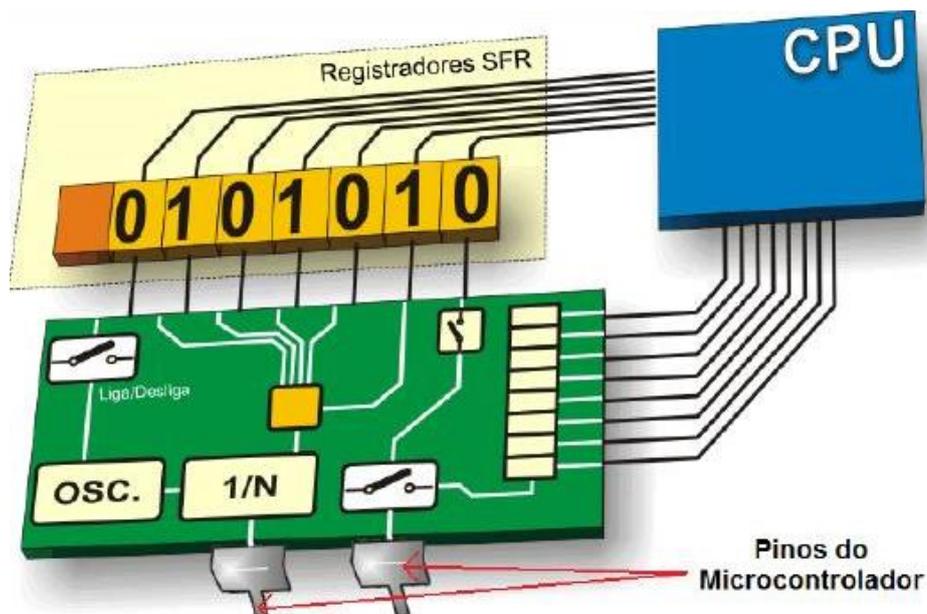


Figura 3 - Registradores SFR

Os registradores são dispositivos de armazenamento temporário localizados no interior do processador. Por causa da tecnologia utilizada, os registradores são um tipo de memória extremamente rápida e bastante cara. Por esse motivo, sua disponibilidade é muito limitada. Cada registrador possui capacidade para manter apenas um dado (uma palavra).

Os registradores ficam no topo da pirâmide, o que representa que sua velocidade de transferência de dados dentro do processador é bastante elevada; em consequência disso, sua capacidade e armazenamento é baixa e seu custo é alto.

3. Memória Cache

Como o processador precisa buscar dados e instruções na memória externa, denominada memória principal, para processá-los e, considerando que a tecnologia desenvolvida para os processadores fez com que esses dispositivos sejam bem mais rápidos que a memória principal, surgiu a necessidade de diminuir esse atraso gerado pela transferência de dados entre a memória e o processador. Na busca de soluções para a limitação imposta pela comunicação entre processador e memória, foi desenvolvida uma técnica que consiste na inclusão de um dispositivo de memória entre a memória principal e o processador. Esse dispositivo é denominado memória cache. Sua função principal é acelerar a velocidade de transferência das informações entre processador e memória principal e, com isso, aumentar o desempenho dos sistemas de computação.

A memória cache é volátil, assim como os registradores, pois dependem de energia para manter o seu conteúdo armazenado. A Figura 4 apresenta um diagrama de blocos de um processador Intel Core i7 e a distribuição da memória cache.

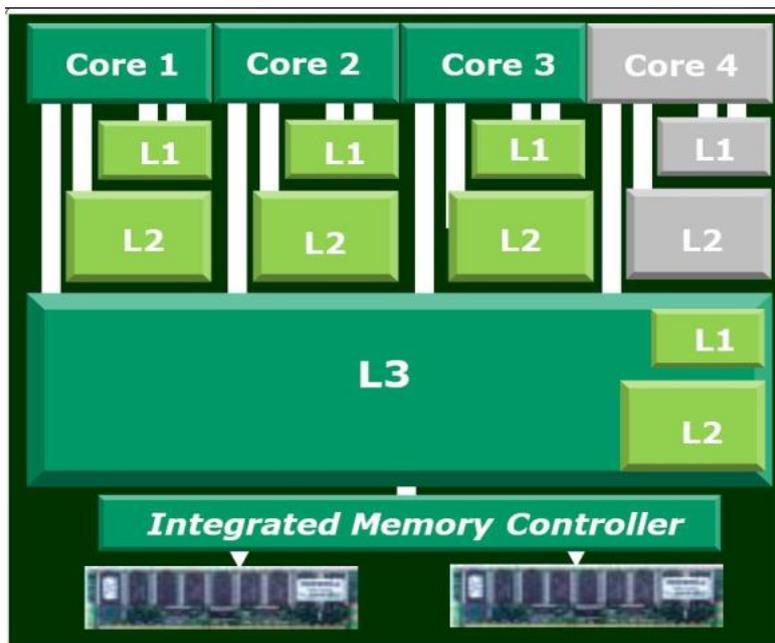


Figura 4 - Memória cache no Core i7

A memória cache é um tipo de memória construída com tecnologias semelhantes às do processador, isso eleva os custos de produção. Para amenizar o fator custo e dispor das vantagens de um sistema computacional com desempenho mais rápido, foram incorporadas ao computador pequenas porções de memória cache, localizadas internamente ao processador e entre ele e a memória principal, as quais funcionam como um espelho de parte da memória principal.

Quando o processador solicita um determinado dado e o encontra na cache, não há necessidade de requisitá-lo à memória principal, reduzindo significativamente o tempo de processamento. Ou seja, quanto mais memória

cache um processador possuir, melhor será o seu desempenho. A tecnologia de fabricação da memória cache é SRAM (*Static Random Access Memory*), a qual é bastante diferente das memórias DRAM (*Dynamic Random Access Memory*) – tecnologia da memória principal. A diferença é que nas memórias SRAM não há necessidade de *refresh* ou realimentação constante para que os dados armazenados não sejam perdidos. Assim, o *refresh* passa a não ser necessário, o que faz com que esse tipo de memória seja mais rápida e consuma menos energia.

A memória cache faz com que os dados usados com maior frequência pela CPU estejam instantaneamente disponíveis. Os processadores trabalham, basicamente, com dois tipos de cache: cache L1 (Level 1 ou Nível 1) e cache L2 (Level 2 ou Nível 2). A memória primária ou cache nível 1 (L1) está localizada dentro da CPU. O cache nível 1 é muito pequeno, normalmente variando entre 16 KB e 512 KB. A cache L2 é um pouco maior que a L1 e foi implantada quando a cache L1 se mostrou insuficiente. Nas gerações anteriores, a cache L1 ficava localizada no interior do processador e a cache L2 era externa a ele. Nas gerações de computadores atuais, ambos os tipos ficam localizados dentro do chip do processador, sendo que, em muitos casos, a cache L1 é dividida em duas partes: “L1 para dados” e “L1 para instruções”. Dependendo da arquitetura do processador, é possível computadores que tenham um terceiro nível de cache (L3). O cache L3, ou terceiro nível de memória, localizado na placa mãe, é um tipo de cache usado em sistemas complexos, com microprocessadores de vários núcleos. O cache L3 é compartilhado por todos os núcleos do processador.

Um tipo particular de RAM, a memória estática de acesso aleatório (SRAM), é usada principalmente para implementar a memória cache. A SRAM usa transistores múltiplos, normalmente de quatro a seis, para cada célula (bit) de memória. Ela possui uma matriz externa de portas, conhecida como multivibrador bi-estável (flip-flop) que chaveia entre dois estados. Isso significa que ela não tem que ficar sendo refrescada continuamente como a DRAM. Cada célula mantém seus dados, enquanto estiver energizada. Sem precisar de refresh contínuo, a SRAM pode operar com extrema rapidez. Porém, a complexidade de cada célula faz com que seja caro o seu uso como padrão de memória RAM. A Figura 5 apresenta a disposição dessas memórias.



Figura 5 - Memórias cache

4. Memória Principal

É um tipo de memória à qual o processador pode fazer acesso direto. Além de alocar os dados e instruções de programas a serem manipulados pelo processador, esse tipo de memória dá acesso às memórias secundárias, de forma a disponibilizar dados ao processador.

A memória principal é denominada memória RAM (*Random Access Memory*), corresponde a um tipo de memória volátil, ou seja, seu conteúdo fica armazenado enquanto o computador estiver ligado (energizado); ao desligar a corrente elétrica, o conteúdo da memória RAM é apagado. Esse é o motivo pelo qual muitas pessoas perdem arquivos que estão utilizando quando ocorrem fatos como, por exemplo, alguém esbarrar no cabo ligado à tomada de energia elétrica ou mesmo cessar o fornecimento de energia. A Figura 6 apresenta os diferentes modelos de memória RAM.

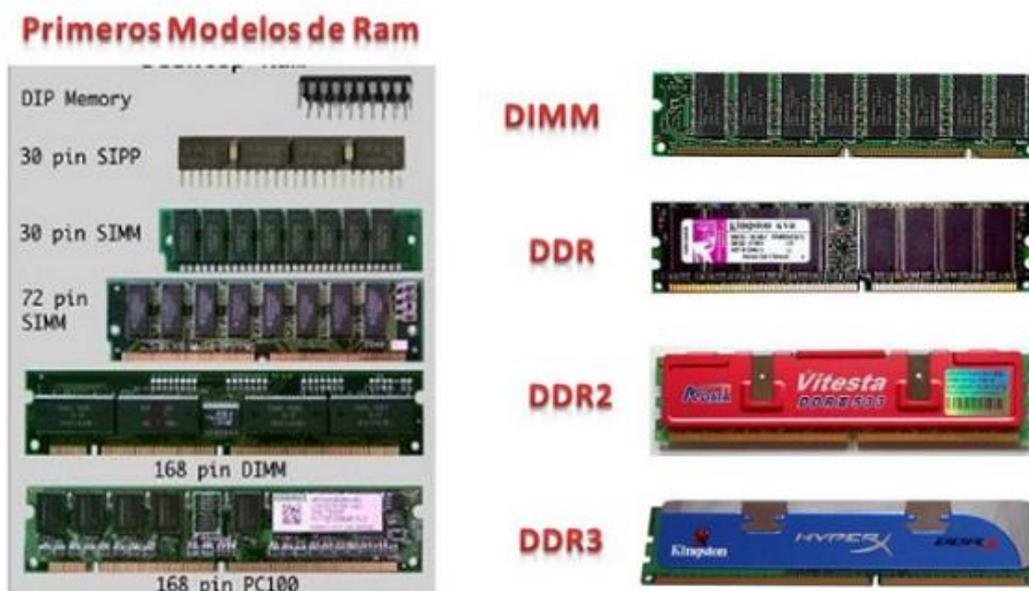


Figura 6 - Tipos de memória RAM

A memória RAM é denominada genericamente de DRAM (*Dynamic RAM*), ou RAM dinâmica, pelo fato de possuir uma característica chamada refresh de memória, que tem a finalidade de manter os dados armazenados enquanto o computador estiver ligado. Essa denominação está ligada à tecnologia utilizada na fabricação desse tipo de memória, a qual se baseia na utilização de dispositivos semicondutores, mais especificamente capacitores associados a transistores para representar bits de dados armazenados.

Pelo fato de precisarem ser “refrescadas” ou realimentadas constantemente, as memórias DRAM consomem muitos ciclos do processador para a realimentação, além de consumirem mais energia que outros tipos de memória. Por isso, são mais lentas e possuem custo muito menor e capacidade de armazenamento de dados consideravelmente maior que as memórias estáticas.

Atualmente, podemos contar com muitas opções de padrões de memória RAM, devido à busca constante por uma memória de maior capacidade, maior

velocidade de acesso, menor consumo de energia e de tempo de realimentação. Apresentamos alguns padrões de memória RAM disponíveis mais utilizados atualmente:

- DDR (Double Data Rate): duplicam o desempenho da memória, possibilitando a transferência de dois lotes de dados – entre processador e memória – por ciclo de clock.
- DDR-2: possibilitam a transferência de quatro lotes de dados por ciclo de clock e apresentam menor consumo de energia que a DDR original.
- DDR-3: transferem oito lotes de dados por ciclo de clock e consomem ainda menos energia que sua versão anterior.

A memória RAM é comercializada no formato de pentes ou módulos de memória, contendo uma determinada quantidade desses recursos. Os pentes de memória podem variar de acordo com as características apresentadas pela memória, especialmente ligadas ao desempenho ou velocidade de transferência. Sendo assim, existem diferentes modelos de módulos de memória disponíveis no mercado (Figura 7). Dentre os que são utilizados atualmente, podem-se citar:

- Módulos SIMM (Single In Line Memory Module): apresentam um pequeno orifício nas linhas de contato. Foram utilizados em memórias FPM e EDO RAM. Não se encontram disponíveis no mercado atualmente;
- Módulos DIMM (Double In Line Memory Module): apresentam contatos em ambos os lados do módulo. São utilizados atualmente em memórias DDR, DDR2 e DDR3.

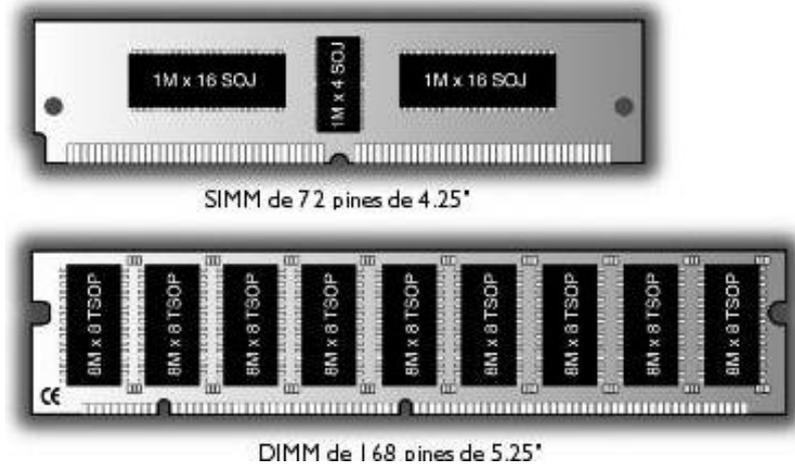


Figura 7 - Exemplos de módulos SIMM e DIMM

4.1. Organização da Memória Principal

Como a memória principal é o local onde os dados e as instruções de um programa ficam armazenados para serem utilizados pelo processador durante a sua execução, é preciso ficar claro que esse conteúdo precisa estar

organizado em uma estrutura padrão que permita a identificação do local onde estão armazenados cada um dos seus itens (Por exemplo, uma instrução ou um dado).

A memória principal encontra-se organizada em um conjunto de células, sendo que cada uma delas representa o agrupamento de uma quantidade de bits. Cada célula caracteriza uma unidade de armazenamento na memória e possui um endereço único, o qual é utilizado pelo processador para acessar seu conteúdo. Portanto, a célula é a menor unidade endereçável em um computador.

Cada célula é constituída de um conjunto de circuitos eletrônicos, baseados em semicondutores, que permitem o armazenamento de valores 0 ou 1, os quais representam um dado ou uma instrução. A quantidade de bits que pode ser armazenada em uma célula é definida pelo fabricante. Uma célula contendo N bits permite o armazenamento de 2^N combinações, o que representará a quantidade de células possíveis na memória. Um tamanho comum de célula adotado pelos fabricantes é 8 bits (1 byte). Se for possível armazenar em uma memória de 2^N combinações possíveis de células (cada uma delas contendo dados armazenados), então será possível calcular a capacidade de armazenamento da memória principal, da seguinte forma:

- Se $N = 9$ bits, tem-se que $2^9 = 512$ (células de memória);
- Se cada célula pode armazenar 8 bits, tem-se que: $512 \times 8 = 4$ KB de espaço em memória.

O acesso a cada posição (célula) de memória pode ser feito de modo aleatório, proporcionando grande flexibilidade, graças à sua tecnologia de fabricação. São duas as operações que podem ser realizadas em uma memória: escrita (write) – para o armazenamento de dados na memória – e leitura (read) – para a recuperação de dados e instruções armazenados na memória. O termo palavra também é utilizado ao se tratar de memória de um computador, mas não deve ser confundido com célula, pois palavra é utilizada para definir a unidade de transferência e processamento e o número de bits que podem ser transferidos entre a memória principal e a CPU para processamento.

5. Memória ROM

A memória ROM (*Read Only Memory*) também é considerada uma memória principal, mas apresenta algumas diferenças em relação à memória RAM. A primeira delas é o fato de ser uma memória somente de leitura, ou seja, seu conteúdo é escrito uma vez e não é mais alterado, apenas consultado. Outra característica das memórias ROM é que elas são do tipo não voláteis, isto é, os dados gravados não são perdidos na ausência de energia elétrica ao dispositivo. A Figura 8 apresenta um exemplo de módulo de memória ROM.

É dito que um software que é armazenado em uma memória ROM passa a ser chamado de firmware. Em um computador existem diversos software disponíveis em memórias ROM, pois não podem ser apagados ao desligar o computador e devem ficar disponíveis sempre que for necessário.

Dessa forma, as memórias ROM são aplicadas em um computador para armazenar três programas principais:

1. **BIOS (Basic Input Output System):** ou Sistema Básico de Entrada e Saída, é responsável por ensinar o processador da máquina a operar com os dispositivos básicos de entrada e saída;
2. **POST (Power On Self Test):** Auto teste – programa de verificação e teste que se executa após a ligação do computador, realizando diversas ações sobre o hardware (Por exemplo, contagem de memória);
3. **SETUP:** Programa que altera os parâmetros armazenados na memória de configuração (CMOS).



Figura 8 - Exemplo de módulo de memória ROM

As memórias ROM podem ser classificadas em:

- **PROM (Programmable Read-Only Memory):** este é um dos primeiros tipos de memória ROM. A gravação de dados neste tipo é realizada por meio de aparelhos que trabalham através de uma reação física com elementos elétricos. Uma vez que isso ocorre, os dados gravados na memória PROM não podem ser apagados ou alterados;
- **EPROM (Erasable Programmable Read-Only Memory):** tem como principal característica a capacidade de permitir que dados sejam regravados no dispositivo. Isso é feito com o auxílio de um componente que emite luz ultravioleta. Nesse processo, os dados gravados precisam ser apagados por completo. Somente após esse procedimento uma nova gravação pode ser realizada;
- **EEPROM (Electrically-Erasable Programmable Read-Only Memory):** este tipo de memória ROM também permite a regravação de dados; no entanto, ao contrário do que acontece com as memórias EPROM, os processos para apagar e gravar dados são feitos eletricamente, fazendo com que não seja necessário mover o dispositivo de seu lugar para um aparelho especial para que a regravação ocorra;
- **EAROM (Electrically-Alterable Programmable Read-Only Memory):** as memórias EAROM podem ser vistas como um tipo de EEPROM. Sua principal característica é o fato de que os dados gravados podem ser

alterados aos poucos, razão pela qual esse tipo é geralmente utilizado em aplicações que exigem apenas reescrita parcial de informações;

- **Flash:** podem ser vistas como um tipo de EEPROM; no entanto, o processo de gravação (e regravação) é muito mais rápido. Além disso, memórias Flash são mais duráveis e podem guardar um volume elevado de dados. Trata-se do tipo de memória utilizada em pen-drive;
- **CD-ROM, DVD-ROM** e afins: essa é uma categoria de discos ópticos onde os dados são gravados apenas uma vez, seja de fábrica, como os CD's de músicas, ou com dados próprios do usuário, quando este efetua a gravação. Há também uma categoria que pode ser comparada ao tipo EEPROM, pois permite a regravação de dados: CD-RW e DVD-RW e afins.

6. Memória Secundária

A memória secundária também é denominada de memória de massa, por possuir uma capacidade de armazenamento muito superior à das outras memórias citadas. Outra característica que difere a memória secundária das outras memórias é o fato de ser permanente (não volátil), ou seja, não perde o conteúdo armazenado caso o computador seja desligado. Por estar na base da pirâmide, apresenta o menor custo por byte armazenado.

Este tipo de memória não possui acesso direto pelo processador, sempre havendo a necessidade de carregamento de dados dos dispositivos de memória secundária para a memória principal, para que então sejam enviados ao processador. A memória secundária pode ser constituída por diferentes tipos de dispositivos, alguns diretamente ligados ao sistema para acesso imediato (por exemplo, discos rígidos) e outros que podem ser conectados quando desejado (por exemplo, pen-drive, CD/DVD).

Em relação à tecnologia de fabricação, existe uma variedade muito grande de tipos, assim como a variedade de dispositivos que se enquadra nessa categoria de memória e para cada dispositivo, existem diferentes tecnologias de fabricação. A Figura 9 apresenta os detalhes de construção de um disco rígido.



Figura 9 - Modelo de disco rígido