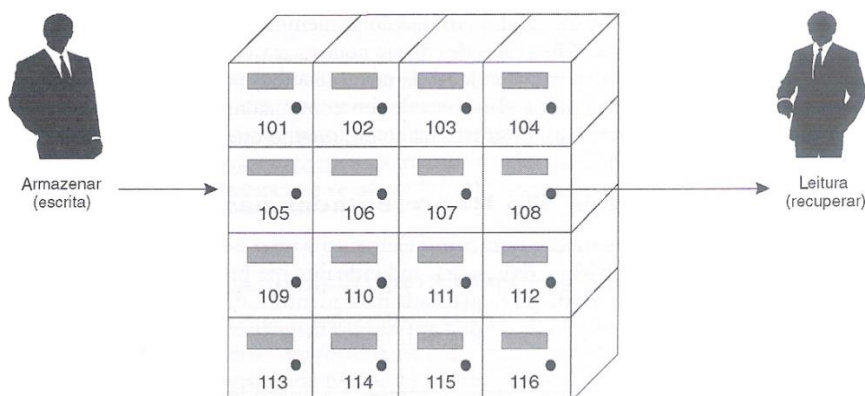


Subsistemas de memória

A memória é o componente responsável pelo armazenamento das informações (dados ou instruções) que são utilizadas no sistema computacional.

Em Monteiro (2012), podemos ver de maneira simples o esquema conceitual de memória, utilizando como exemplo um depósito, que pode ser utilizado por uma ou mais entidades, conforme a figura a seguir:



Exemplo de um típico depósito que funciona de modo semelhante a uma memória
Fonte: Monteiro (2012).

De acordo com a figura, podemos analisar alguns itens:

1) Cada caixa tem seu endereço

O endereço é essencial para que se possa acessar qualquer informação na memória. Na memória, através do endereço, pode-se acessar uma célula (ou bloco ou setor) de memória.

Segundo Monteiro (2012), “uma célula é, então, um grupo de bits tratado em conjunto pelo sistema, isto é, esse grupo é movido em bloco como se fosse um único elemento, sendo assim identificado para efeito de armazenamento e transferência, como uma unidade”.

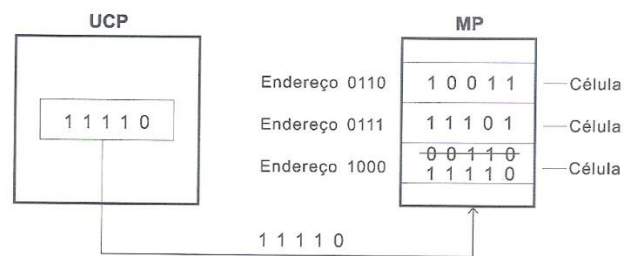
Cada uma das células, ou conjunto de bits, possui um endereço único e a memória é organizada de acordo com esses endereços, que são colocados de forma sequencial.

2) Temos duas pessoas acessando o depósito

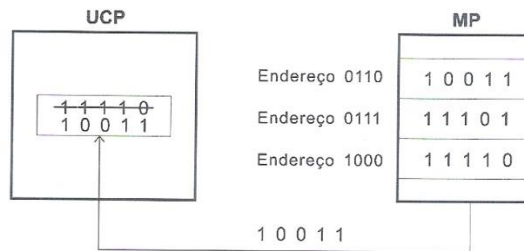
Isso representa as duas operações que podem ser realizadas com a memória: a de escrita e a de leitura.

A operação de escrita é caracterizada como o armazenamento de uma informação na memória. Essa operação também pode ser chamada de gravação (*write*). É importante lembrar que a escrita destrói o conteúdo anterior da memória, pois os bits novos serão armazenados no lugar dos originais.

A operação de leitura (*read*) é utilizada para recuperar uma informação na memória (*retrieve*). Essa informação pode ser utilizada para diversos usos: pela UCP para realizar o ciclo de instrução, como dados para operações aritméticas, para envio a dispositivos de saída, entre vários outros no sistema computacional. A operação de leitura, diferentemente da de escrita, não é destrutiva, ou seja, os bits na memória permanecem os mesmos. O que acontece é apenas a realização de uma cópia desse conteúdo.



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



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

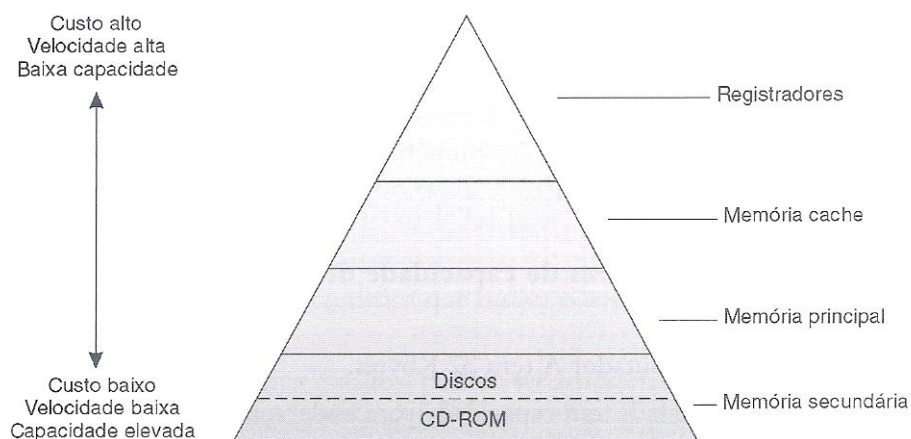
Operação de leitura e escrita na MP

Fonte: Monteiro (2012).

Não temos somente um tipo de memória em um sistema computacional. Devido a diferentes características advindas da tecnologia de construção dessas memórias, como tempo de acesso (que é diretamente ligado à velocidade), capacidade, custo e volatilidade, cada tipo de memória é utilizado para uma função.

Esses diferentes tipos de memória que existem no sistema computacional constituem o subsistema de memória. É importante lembrar que essas memórias funcionam de forma conjunta, sendo interligadas.

Podemos organizar as memórias em uma pirâmide, que representa a hierarquia de memória, conforme mostra a figura a seguir:



Hierarquia de memória

Fonte: Monteiro (2012).

Segundo Hennessy (2003), como uma memória rápida é dispendiosa, uma hierarquia de memória é organizada em vários níveis – cada um deles menor, mais rápido e caro por byte que o nível imediatamente abaixo. O objetivo é fornecer um sistema de memória com custo quase

tão baixo quanto o nível de memória mais econômico e com velocidade quase tão grande quanto o nível mais rápido.

Weber (2008, p. 55) complementa: “um sistema de computação só apresenta ganho na relação custo/desempenho se as memórias na hierarquia apresentarem velocidade, preço e tamanho significativamente diferentes”.

Parâmetros para análise das características:

1) Tempo de acesso: quanto tempo a memória gasta para colocar uma informação no barramento de dados após a posição ter sido endereçada ou o período de tempo decorrido desde o instante em que foi iniciada a operação de acesso até que a informação tenha sido transferida. É chamado também de tempo de acesso para leitura.

- depende do modo como o sistema é construído e da velocidade dos circuitos;
- o tempo de acesso das memórias eletrônicas (RAM, ROM) é igual, independente da distância física entre o local de um acesso e do próximo, diferente do caso dos dispositivos eletromecânicos (discos, fitas).
- é importante perceber que os conceitos de velocidade e tempo de acesso são inversamente proporcionais. Quanto menor o tempo de acesso, maior a velocidade e, quanto mais tempo é gasto para trabalhar com uma informação (maior tempo de acesso), menor é a velocidade da memória.

2) Capacidade: representa a quantidade de informação que pode ser armazenada.

Dependendo do tipo de memória que estamos utilizando, usamos múltiplos diferentes. Por exemplo, quando nos referimos à cache, falamos em megabytes, quando nos referimos à memória principal, falamos em megabytes e gigabytes e, quando nos referimos à memória secundária, falamos em gigabytes e terabytes. Para registradores, utilizamos bits.

Termo	Símbolo	Número aproximado de bytes
Quilobyte	K (ou KB)	um mil
Megabyte	MB	um milhão
Gigabyte	GB	um bilhão
Terabyte	TB	um trilhão
Petabyte	PB	um quadrilhão

Capacidades de armazenamento

Fonte: Capron

3) Volatilidade: diz respeito à capacidade de uma memória manter ou não suas informações na falta de alimentação elétrica.

- memória não volátil: retém a informação armazenada quando a energia é desligada → memórias magnéticas, memórias óticas, ROM, EPROM
- memória volátil: perde a informação → registradores, RAM.

4) Tecnologia de fabricação: as mais utilizadas atualmente são:

- Memórias de semicondutores: utilizadas para implementação dos registradores, memória principal e memória cache. São também chamadas de memórias eletrônicas, pois são construídas usando-se circuitos eletrônicos.

- Memórias de meio magnético: utilizadas nos discos rígidos (HDs) e nos antigos disquetes. Nessas memórias, as informações são armazenadas na forma de campos magnéticos.
- Memórias de meio ótico: utilizadas para armazenamento de informações nos CDs e DVDs. São usados feixes de luz para gravar os bits no dispositivo.
- SSD (Solid-State Drive). Segundo Techtudo, 2016:

... esse tipo de armazenamento de massa não possui nenhum disco. O dispositivo é todo formado por circuitos integrados e em seu interior não há partes móveis, o que o torna absolutamente silencioso, mais rápido e menos propenso a danos físicos do que o HD. O sistema de gravação e leitura, naturalmente, é diferente dos discos rígidos. A maior parte dos SSDs atuais armazena os dados em células de memória flash, que também são do tipo não volátil. É essa a versão usada em smartphones, tablets e computadores mais novos.

5) Custo: ligado a tecnologia de fabricação → maior ou menor tempo de acesso, ciclo de memória, quantidade de bits em espaço físico.

É importante lembrar que, quando analisamos o custo de uma memória, devemos fazer o cálculo do valor por byte e não levando em consideração o valor dos dispositivos.

Se fizermos isso, podemos chegar a conclusões errôneas, como afirmar que as memórias secundárias, como os discos rígidos, são mais caras que a memória principal (RAM).

Registradores

Em um sistema de computação, a destinação final de qualquer tipo de memória é o processador (a UCP). Isto é, o objetivo final de cada uma das memórias (ou do subsistema de memória) é armazenar informações a serem, em algum momento, utilizadas pelo processador. Ele é o responsável pela execução das instruções, pela manipulação dos dados e pela produção dos resultados das operações.

As ações operativas do processador são realizadas nas suas unidades funcionais: na unidade de aritmética e lógica – UAL, na unidade de ponto flutuante – UFP ou talvez em uma unidade de processamento vetorial. No entanto, antes que a instrução seja interpretada e as unidades da UCP sejam acionadas, o processador necessita buscar a instrução de onde ela estiver armazenada (memória cache ou principal) e armazená-la em seu próprio interior, em um dispositivo de memória denominado registrador de instrução.

Em seguida a este armazenamento da instrução, o processador deverá, na maioria das vezes, buscar dados na memória (cache, principal ou mesmo de unidades de disco ou fita) para serem manipulados na UAL. Esses dados também precisam ser armazenados em algum local da UCP até serem efetivamente utilizados. Os resultados de um processamento (de uma soma, subtração, operação lógica, etc.) também precisam, às vezes, ser guardados temporariamente na UCP, ou para serem novamente manipulados na UAL por uma outra instrução, ou para serem transferidos para uma memória externa à UCP. Esses dados são armazenados na UCP em pequenas unidades de memória, denominadas registradores.

Um registrador é, portanto, o elemento superior da pirâmide de memória, por possuir a maior velocidade de transferência dentro do sistema (menor tempo de acesso), menor capacidade de armazenamento e maior custo.

Memória cache

Na pirâmide de memória, abaixo dos registradores, encontra-se a memória cache.

Em toda execução de uma instrução, a UCP acessa a memória principal (sem cache), pelo menos uma vez, para buscar a instrução (uma cópia dela) e transferi-la para um dos registradores da UCP. E, mais ainda, muitas instruções requerem outros acessos à memória,

seja para a transferência de dados para a UCP (que serão processados pela UAL), seja para a transferência do resultado de uma operação da UCP para a memória.

Em resumo, para a realização do ciclo de uma instrução, há sempre a necessidade de ser realizado um ou mais ciclos de memória.

Considerando-se que um ciclo de memória é atualmente bem mais demorado do que o período de tempo que a UCP gasta para realizar uma operação na UAL, fica claro que a duração da execução de um ciclo de instrução é bastante afetada pela demora dos ciclos de memória.

Desde há muito, então, esta interface entre o processador e a memória principal vem sendo um ponto frágil no que se refere à performance do sistema.

Na tentativa de melhorar o desempenho dos sistemas de computação, os projetistas das UCP vêm constantemente obtendo velocidades cada vez maiores nas operações dessas unidades, o que não está acontecendo na mesma proporção com o aperfeiçoamento tecnológico das memórias utilizadas como memória principal. Assim, atualmente a diferença de velocidade entre UCP e memória principal é talvez maior do que já foi no passado.

Na busca de uma solução para este problema, foi desenvolvida uma técnica que consiste na inclusão de um dispositivo de memória entre UCP e MP, denominado memória CACHE, cuja função é acelerar a velocidade de transferência das informações entre UCP e MP e, com isso, aumentar o desempenho dos sistemas de computação.

Para tanto, esse tipo de memória é fabricado com tecnologia semelhante à da UCP e, em consequência, possui tempos de acesso compatíveis, resultando numa considerável redução da espera da UCP para receber dados e instruções da cache, ao contrário do que acontece em sistema sem cache.

Memória principal

Uma das principais características definidas no projeto de arquitetura de Von Neumann consistia no fato de ser uma máquina de programa armazenado. O fato das instruções, uma após a outra, poderem ser imediatamente acessadas pela UCP é que garante o automatismo do sistema e aumenta a velocidade de execução dos programas.

E a UCP pode acessar imediatamente uma instrução após a outra porque elas são armazenadas internamente no computador. Esta é a importância da memória.

E, desde o princípio, a memória especificada para armazenar o programa (e seus dados) a ser executado é a memória que atualmente chamamos de principal (ou memória real), para distingui-la da memória de discos e fitas (memória secundária).

A memória principal é, então, a memória básica de um sistema de computação desde seus primórdios. É o dispositivo onde o programa (e seus dados) que vai ser executado é armazenado para que a UCP vá “buscando” instrução por instrução.

Memória secundária

Na base da pirâmide que representa a hierarquia de memória em um sistema de computação encontra-se um tipo de memória com maior capacidade de armazenamento do que os outros tipos já descritos, menor custo por byte armazenado e com tempos de acesso também superiores aos outros tipos. Esta memória, denominada memória secundária, memória auxiliar ou memória de massa, tem por objetivo garantir um armazenamento mais permanente a toda a estrutura de dados e programas dos usuários, razão por que deve naturalmente possuir maior capacidade que a memória principal.

A memória secundária de um sistema de computação pode ser constituída por diferentes tipos de dispositivos, alguns diretamente ligados ao sistema para acesso imediato (discos rígidos, por exemplo), e outros que podem ser conectados quando desejado (como os disquetes, fitas de armazenamento, etc.), cuja informação armazenada se torna diretamente conectada e disponível para o específico disquete ou fita que estiver inserido no elemento de

leitura/escrita (“drive” ou acionador), enquanto os demais ficam disponíveis (“off-line”) para acesso manual pelo usuário.

Análise dos parâmetros: Tempo de acesso, Velocidade, Capacidade, Volatilidade, Tecnologia de fabricação, Custo.