



Memória

Armazenamento e Velocidade



Evolução das Memórias

O armazenamento e interpretação de dados é o princípio básico da ciência computacional. Se por um lado temos o processador se reservando a tarefa de manipulação dos dados, fica a cargo da memória a responsabilidade de armazenar os mesmo quando em uso ou arquivados.

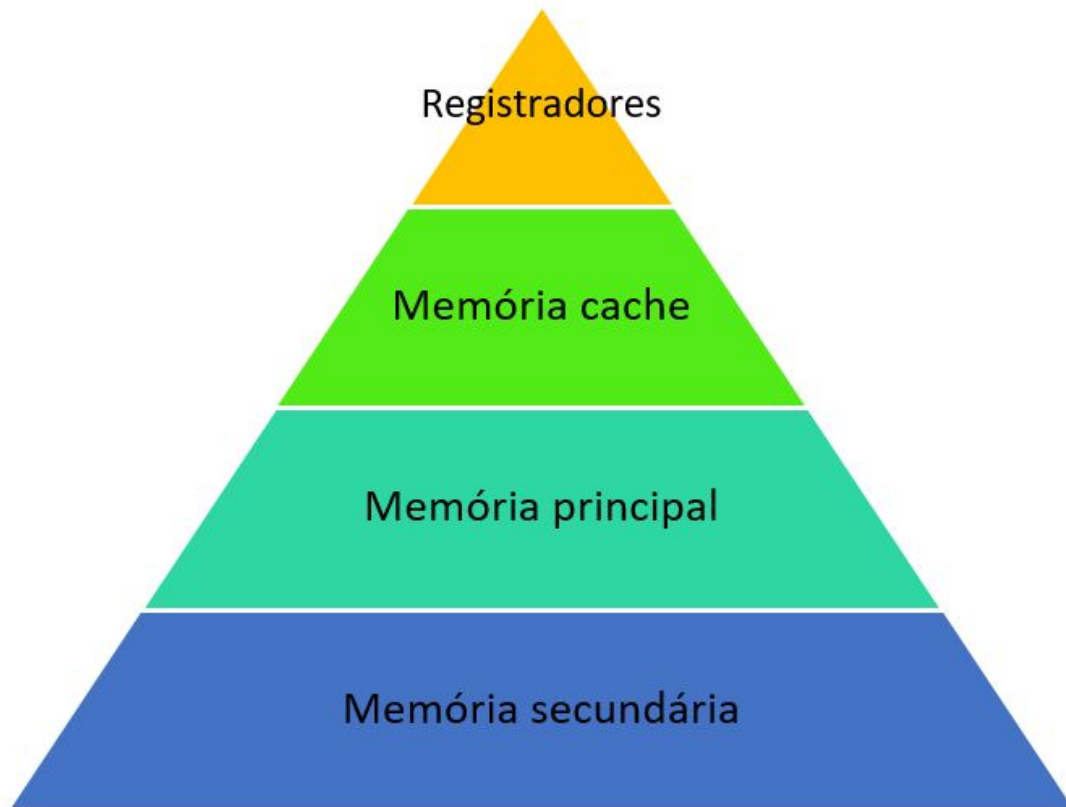
Nos últimos séculos, vários dispositivos foram desenvolvidos com essa tarefa. Dentro da era da computação moderna, temos uma jornada através de cartões perfurados, fitas magnéticas, até a chegada dos discos e memórias flash mais modernas.



Memórias na Arquitetura Atual

Todo computador é dotado de uma quantidade de memória (que pode variar de computador para computador), a qual constitui um conjunto de circuitos capazes de armazenar os dados e os programas executados pelo próprio computador. As memórias são classificadas em diferentes categorias, representadas pelo gráfico a seguir:

- Registradores
- Memória Cache
- Memória Primária
- Memória Secundária





Registradores

São as memórias de menor capacidade, encontradas dentro do núcleo do próprio processador, porém não menos importante, pois uma instrução, para ser executada, necessita estar no registrador.

- Menor tempo de acesso do sistema, sendo de um ciclo de memória (1 a 2 ns)
- Armazenam um dado, instrução ou endereço. Definem a arquitetura em 32 ou 64 bits
- Voláteis



Funcionamento do Registrador

Os registradores são memórias elaboradas com o mínimo de transistores possível, utilizando o que há de mais moderno em tecnologia de armazenamento. Elas são as memórias mais rápidas que podem ser construídas e por isso são também as mais caras. Por essa razão, elas aparecem numa quantidade muito pequena em um computador, na casa de alguns Kilo Bytes.

O tamanho e quantidade dos registradores de uma CPU é uma das principais decisões de projeto. Se forem grandes demais, ou em quantidade maior do que a necessária, podem resultar em desperdício e aumento desnecessário no preço do processador. Já se forem pequenos, ou em pouca quantidade, com certeza vão tornar o computador muito mais lento do que o desejado.

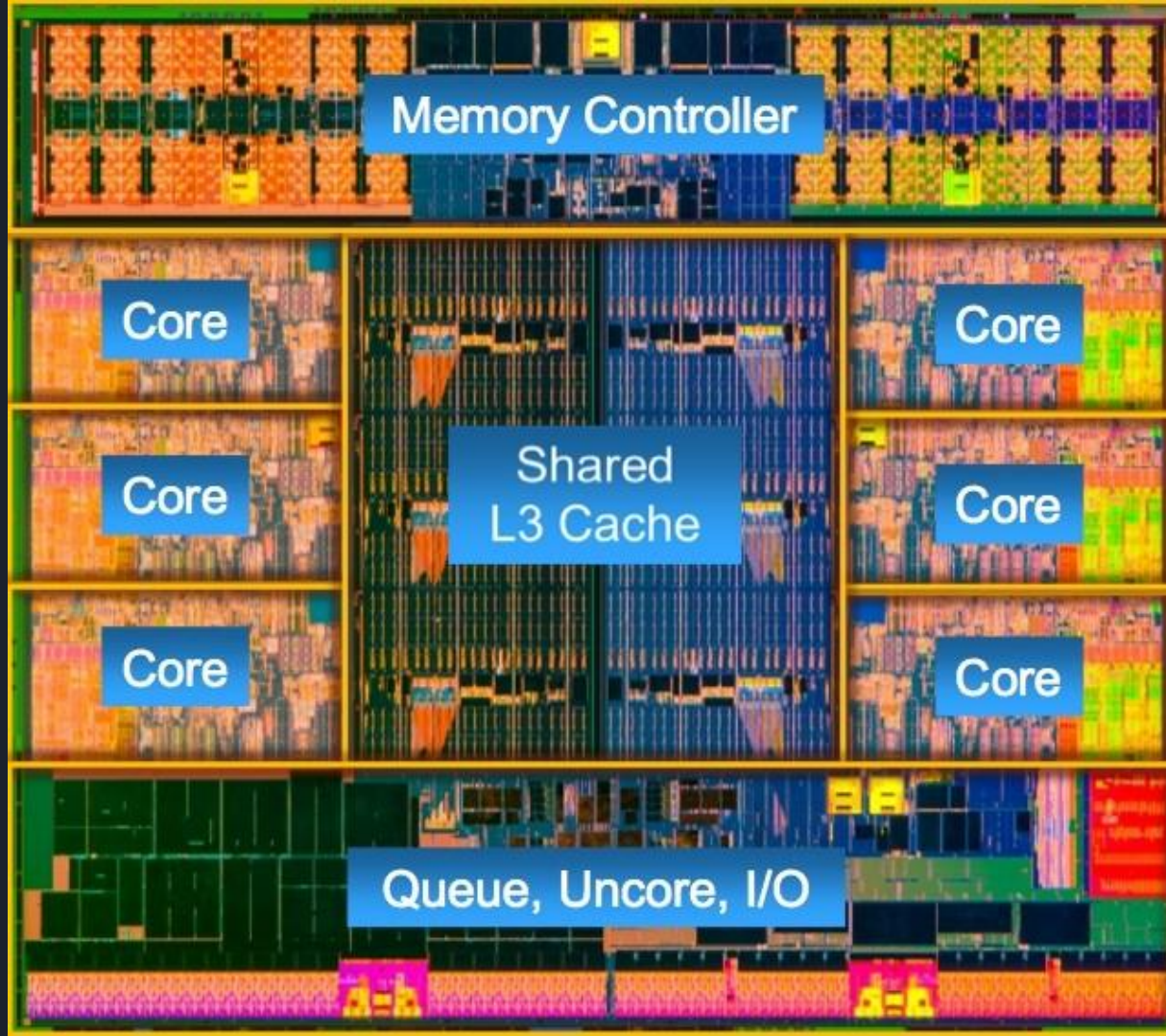
A CPU tenta ao máximo manter as variáveis mais utilizadas nos registradores. Ela faz isso guardando aquelas mais usadas nas últimas operações. Nem sempre isso funciona, mas no geral, é a melhor solução.



Memória Cache

A memória cache é uma memória pequena e rápida. Geralmente encontramos as memórias caches dentro dos processadores, pois são responsáveis por passar os dados da memória principal para o núcleo do processador. O acesso é mais rápido que o da memória principal, porém o seu tamanho é bem reduzido em relação à mesma.

- L1: Porção pequena, dividida em dois níveis (dados e instruções), com capacidade medida em KB.
- L2: Medida em MB (em média 4 MB), é uma auxiliar à L1. A memória cache L2 é programada para salvar dados e informações .
- L3: Já a memória cache L3 é uma memória de maior tamanho para salvar conteúdos “genéricos”, que supostamente podem ser requisitados por quaisquer núcleos para dar continuidade às tarefas (em média de 6~16MB).
- Tempo de acesso de 5 a 20 ns.





Funcionamento da Memória Cache

Como podemos conferir na imagem anterior, o L3 cache está a disposição de forma direta para todos os núcleos de um processador. Já os níveis L2 e L1 costumam estar junto ao núcleo, conseguindo essa alocação por terem um menor tamanho físico e lógico.

No processo de busca de instruções de um processador, o cache L1 é o primeiro a ser lido. Nele o processador pode buscar diretamente em dois diferentes níveis: o de memória de instrução ou memória de dados, agilizando o processo. Caso os dados não sejam encontrados, o processador buscará os dados no L2 e L3 cache respectivamente. Por fim, caso não encontre o dado necessário, a busca será solicitada da memória RAM.



Memória Principal

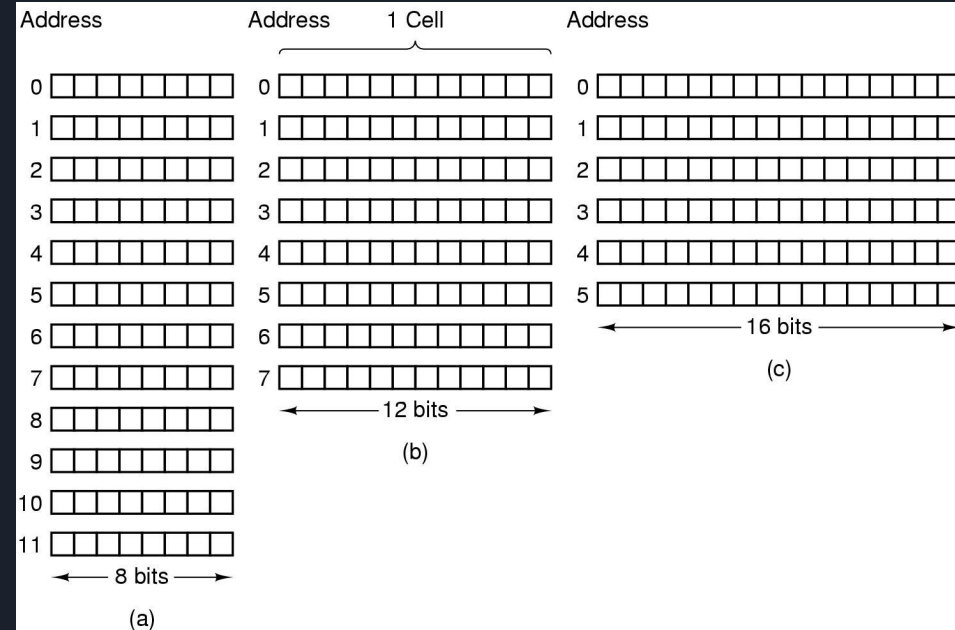
Também conhecida como memória de trabalho, onde normalmente devem estar armazenados os programas e dados a serem manipulados pelo processador. Geralmente é esta memória que se referencia na especificação de um microcomputador.

- Tempo de acesso de um ciclo de memória é de 50 a 80 ns
- Tem capacidade variante, a depender do sistema usado e arquitetura do processador.
 - Processadores 32 bits tem limite de acesso a até 4GB de RAM
 - Limitação de sistemas 64bits são definidas pelo sistema operacional em uso.
- Volátil.

Arquitetura da Memória Principal

A memória principal pode ser comparado a um armário com diversas gavetas enumeradas, as quais armazenam os dados carregados do sistema. A unidade de armazenamento na memória principal é denominada célula.

- A célula é um agrupamento de bits tratado em conjunto pelo sistema
- É usada para efeitos de armazenamento e transferência.





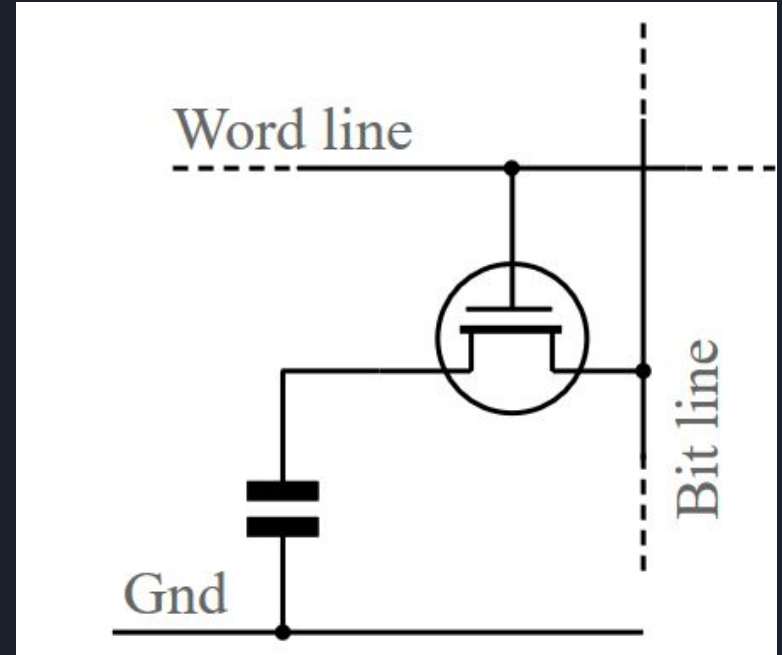
Acesso a Memória Principal

Para ter acesso a uma informação armazenada na memória, deve ser especificado o endereço da célula correspondente, sendo cada célula identificada por um único endereço.

Esses endereços não são fisicamente gravados na memória, somente as informações são armazenadas em cada célula. Eles são enviados pelo processador para a memória através de um barramento* de endereços.

Funcionamento da Memória Principal

Como vimos no slide anterior, uma memória RAM é composta por milhares de células. Cada célula contém um capacitor e um transistor. O capacitor serve para armazenamento de energia e o transistor libera ou não a passagem da mesma. Um capacitor carregado com energia representa o valor 1, um capacitor sem energia representa o valor 0.



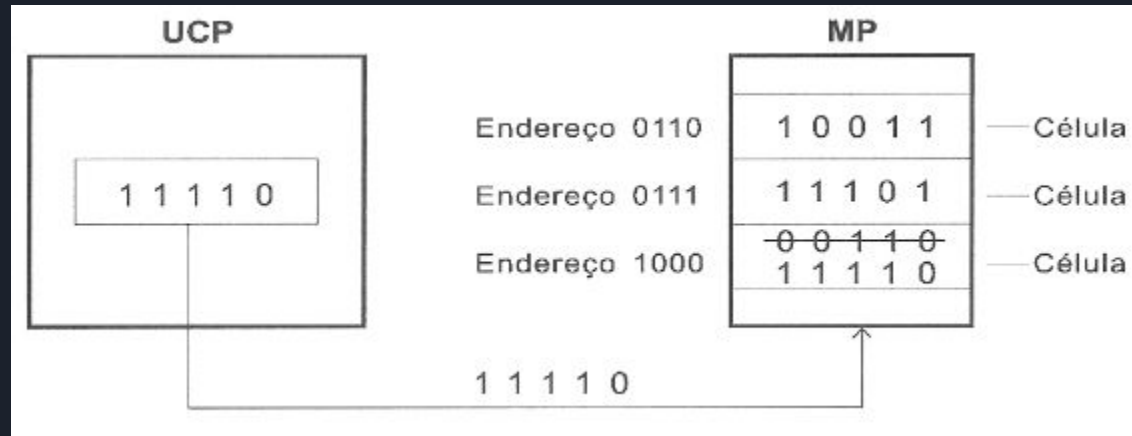
Operações de Leitura

As operações de leitura realizadas na memória principal (MP) pela unidade central de processamento (UCP) não são destrutivas. Uma cópia exata do do valor é feita no local diretamente no local de origem.



Operação de Escrita

As operações de escrita realizadas na memória principal (MP) pela unidade central de processamento (UCP) são destrutivas. É armazenado um novo conteúdo sobre o conteúdo anterior.







Memória Secundária

Este tipo de memória não é acessado diretamente pela CPU. Seu acesso é feito através de interfaces ou controladoras especiais. A memória secundária é uma memória do tipo permanente (não se apaga quando o computador está desligado), que tem uma alta capacidade de armazenamento, e um custo muito mais baixo em relação ao da memória principal.



Unidades de Memória

Bit	Um dígito binário é lógico 0 e 1
Byte	Grupo de 8 Bits
Kilobyte	1 KB = 1024 Bytes
Megabyte	1 MB = 1024 KB
Gigabyte	1 GB = 1024 MB
Terabyte	1 TB = 1024 GB
Petabyte	1 PB = 1024 TB



Capacidade de Transmissão

As capacidades de transmissão dos dispositivos de armazenamento simbolizam a quantidade de dados por segundo que os dispositivos conseguem transferir para o processador. É a mesma medida de transmissão de qualquer dispositivo visto até aqui, ou seja, KB/s, MB/s, GB/s, etc



Classificação de Memórias

Podemos definir inicialmente as memórias em duas categorias:

- ROM - Read Only Memory
 - Permite apenas as leituras dos dados e não perde a informação na ausência de energia.
- RAM - Random Access Memory
 - Permite tanto a leitura quanto a gravação dos dados, e perde informação quando não há alimentação elétrica.

Dentro dessa definição, podemos considerar as memórias RAM como voláteis e ROM como não voláteis.



Tipos de Memória ROM

- PROM - Programmable Read-Only Memory
 - Os dados gravados na PROM não podem ser apagados ou alterados.
- EPROM - Erasable Programmable Read-Only Memory
 - Permite a regravação dos dados de forma mecânica (luz ultravioleta). Os dados precisam ser apagados por completo.
- EEPROM - Electrically-Erasable Programmable Read-Only Memory
 - Permite a regravação dos dados de forma elétrica.
- EAROM - Electrically-Alterable Programmable Read-Only Memory
 - Permite a regravação dos dados como a EEPROM, podendo fazer esse processo de forma parcial.
- Flash
 - Dispositivo de memória não volátil com velocidade de acesso mais alta que pode ser eletricamente apagado e reprogramado.

