

# Memórias

UNICAP - CCT

Ciência da Computação

Arquitetura e Organização de Computadores I

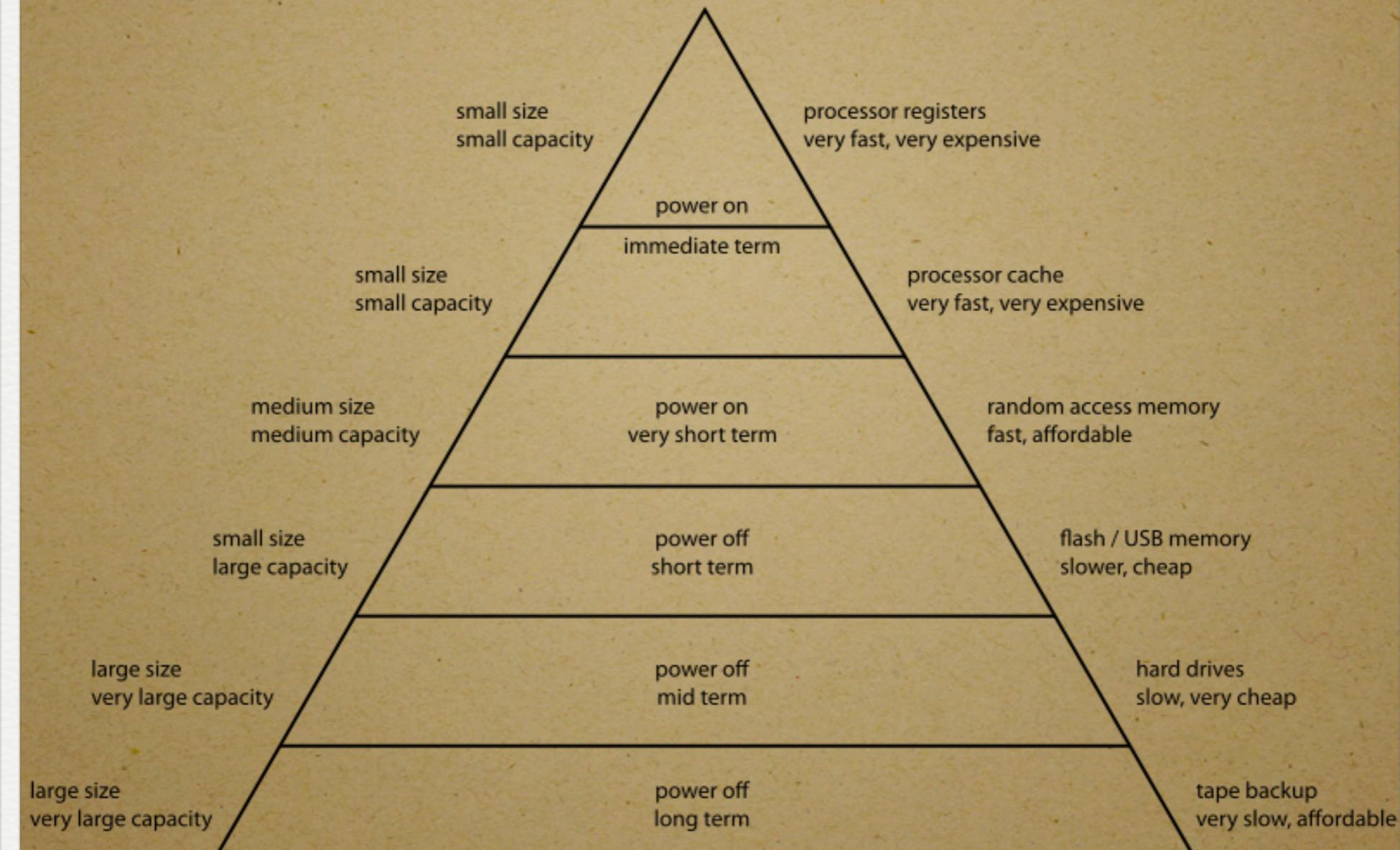
Prof. Sérgio Murilo

# Hierarquia de Memória

# Hierarquia de Memória

- O termo Hierarquia de Memória se refere a uma classificação de tipos de memória em função do desempenho. Tais memórias são categorizadas entre si através da comparação de suas características.

# Computer Memory Hierarchy



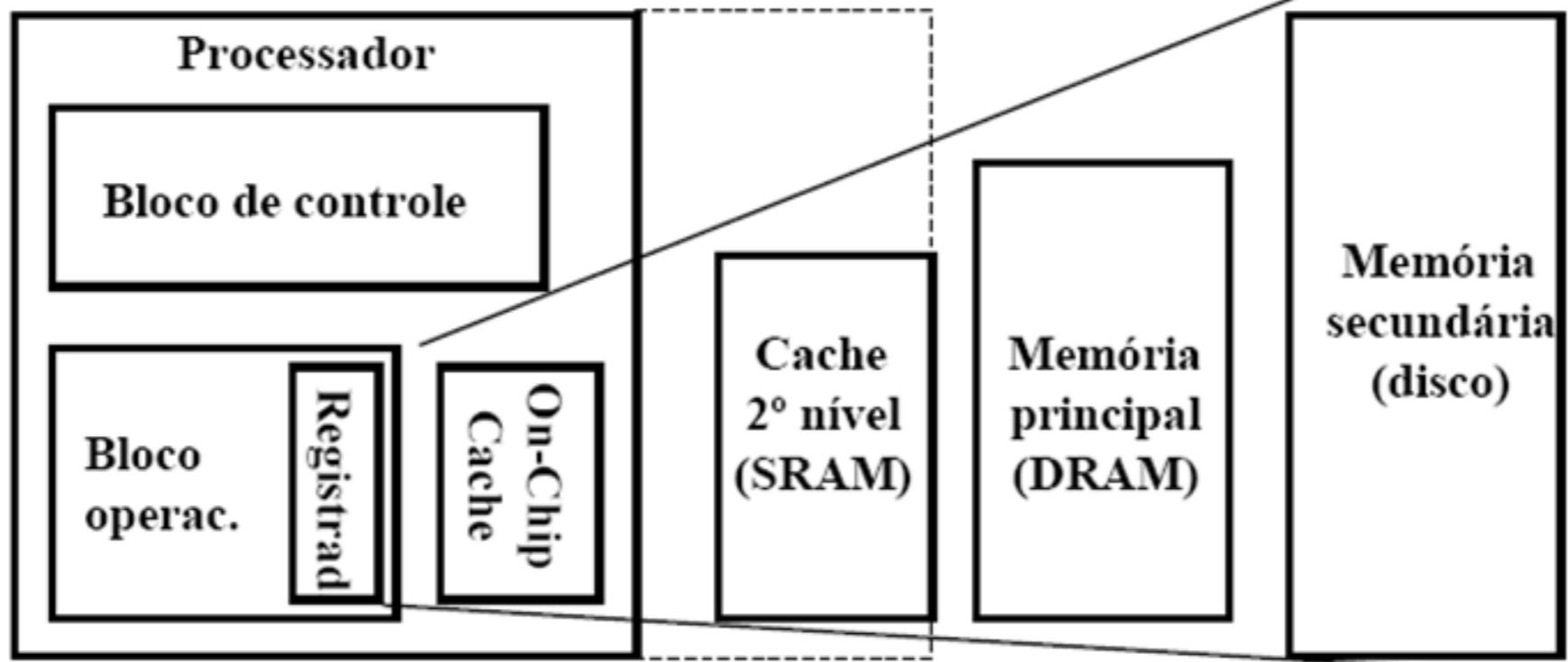
# Hierarquia de Memória

# Hierarquia de Memória

- Usando como exemplo uma comparação de tempo de acesso, poderíamos organizar a seguinte sequência:
  - Registrador → Cache → Memória RAM → Memória Secundária

# Hierarquia de Memória

- Sendo:
  - **Registrador:** Memória temporária usada pelo processador no processamento das instruções.
  - **Cache:** Armazena partes da memória principal que são usados frequentemente pelos programas.
  - **Memória RAM:** Memória principal do computador, sendo diretamente endereçável pelo processador.
  - **Memória Secundária:** Memória de armazenamento permanente.



**Velocidade (ns):** 0,1      1      2-5      10-20      10.000.000 (10 ms)

**Tamanho (bytes):** 100      16 K      512 K      256 M

Gs

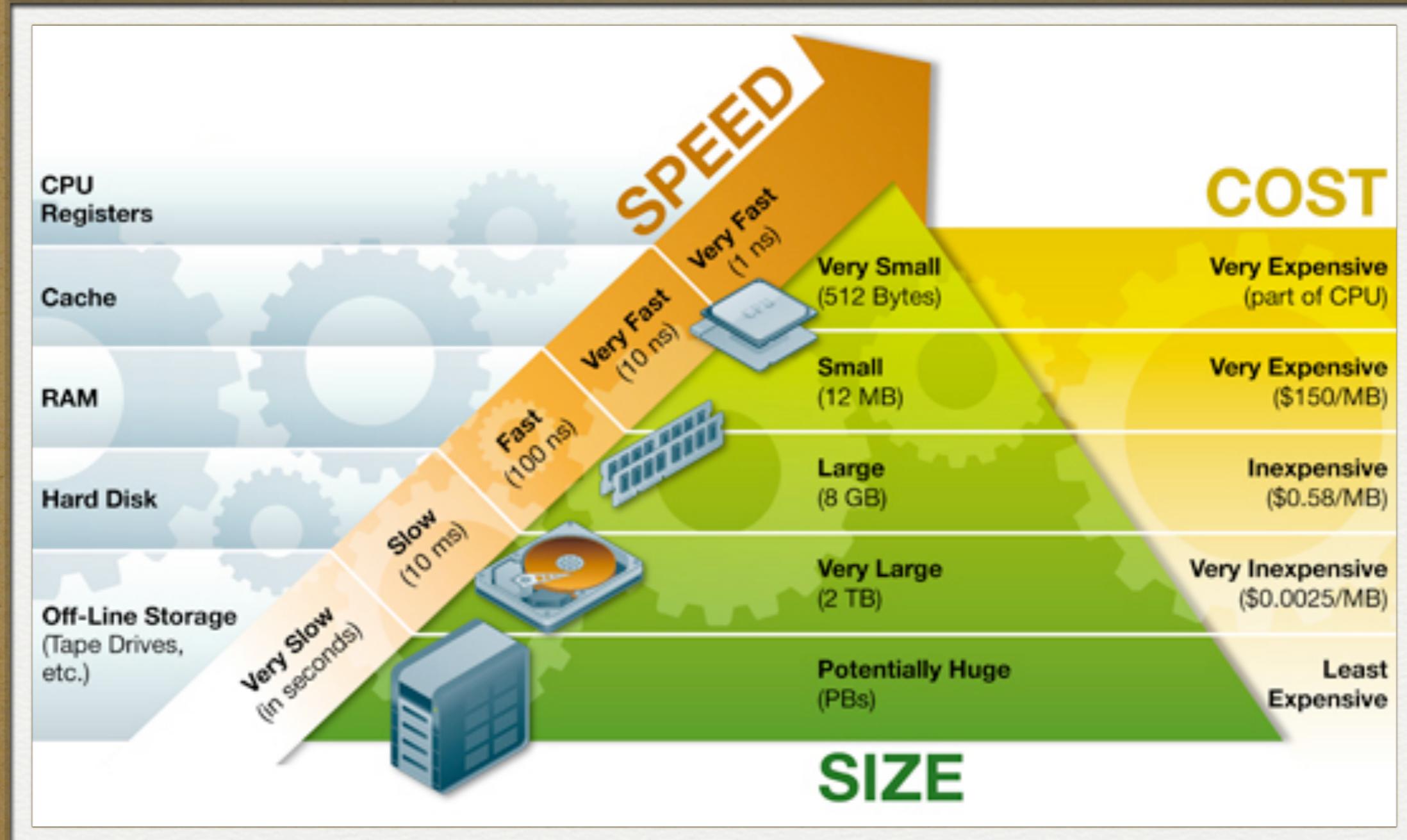
# Hierarquia de Memória

# Hierarquia de Memória

- Para obter um melhor desempenho, a velocidade da memória deve ser compatível com a do processador.
- Para que um sistema seja comercialmente viável, o custo da memória deve ser compatível com o custo dos demais componentes, o que torna as três características principais da memória conflitantes(custo, capacidade, velocidade de acesso).

# Hierarquia de Memória

- OBS: um dado não precisa passar por todos os níveis da hierarquia para chegar ao destino, ele pode ser transferido do registrador diretamente para um HD.



# Custo / Velocidade

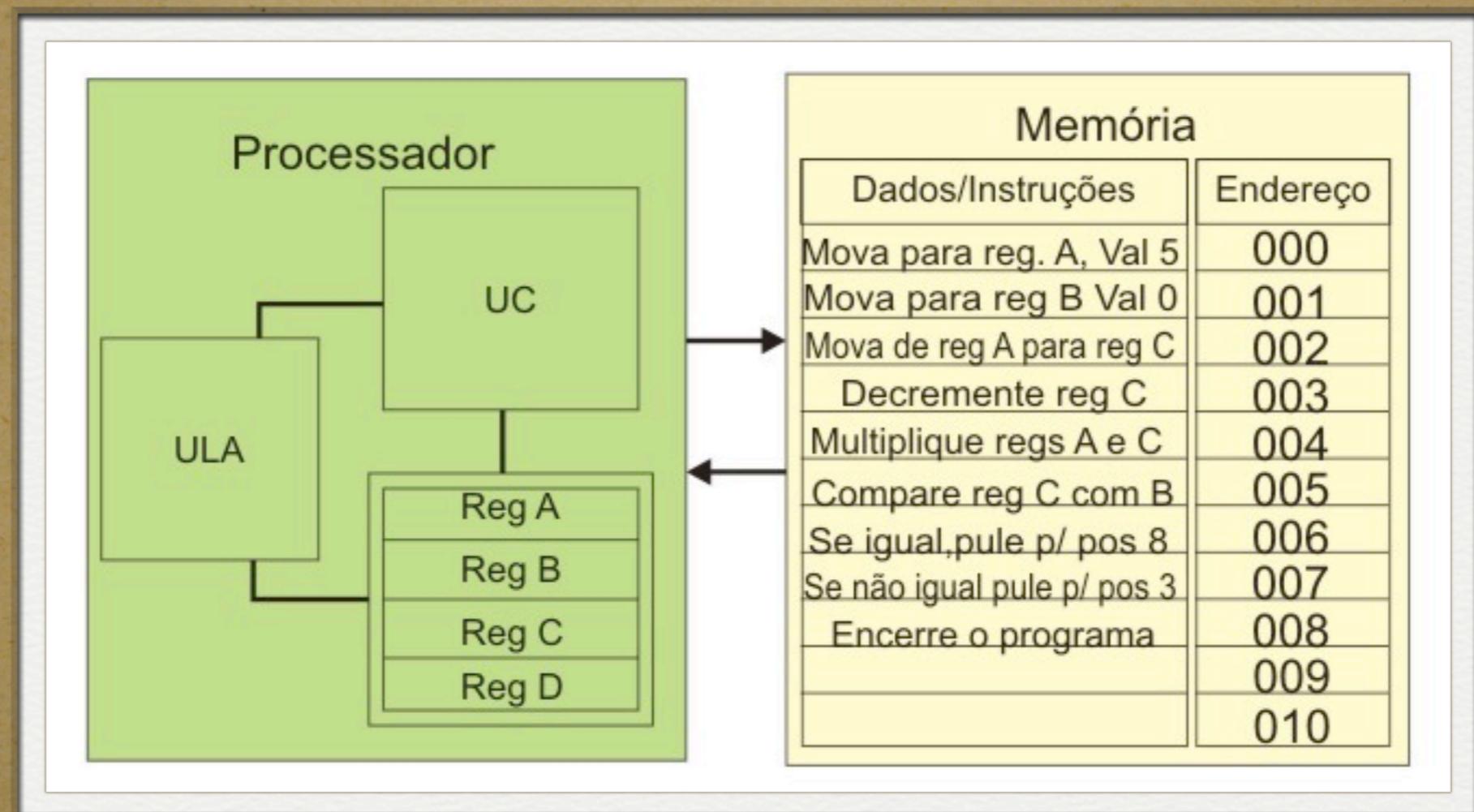
# Registradores

# Definição

- São localizações de armazenamento rápido independentes, que mantêm dados temporariamente.
- Localizados dentro da CPU (Unidade Central de Processamento).
- CPU = ULA + Unidade de Controle + Registradores

# Características

- Alta velocidade
- Específicas
- Voláteis
- Dependem da arquitetura



# Registradores

### General-Purpose Registers

31	16	15	8	7	0	16-bit	32-bit
		AH	AL		0	AX	EAX
		BH	BL			BX	EBX
		CH	CL			CX	ECX
		DH	DL			DX	EDX
			BP				EBP
			SI				ESI
			DI				EDI
			SP				ESP

# Registradores de Uso Geral (Intel)

# Registradores Básicos de Execução de Instruções (Intel IA-32)

- **Registradores de Uso Geral:** São 8, usados para guardar operandos e ponteiros.
- **Registradores de Segmento:** guardam até 6 segmentos de memória.
- **Registradores EFLAGS:** Informa a situação do programa em execução e permite controle limitado do processador.
- **Registrador EIP:** Contem um ponteiro de 32bits para a próxima instrução.

# 8 Tipos de Registradores de Uso Geral (Intel x86)

- EAX - Registrador Acumulador
- EBX - Registrador de Base
- ECX - Registrador Contador
- EDX - Registrador de Dados
- ESI - Índice de Origem
- EDI - Índice de Destino
- EBP - Ponteiro de Base
- ESP - Ponteiro de Pilha

# EAX - Registrador Acumulador

- Trabalha melhor com cálculos.
- Usa as nove operações básicas (ADD, ADC, AND, CMP, OR, SBB, SUB, TEST, and XOR) mais facilmente.

# EDX: Registrador de Dados

- Trabalha mais próximo do Acumulador.
- Serve para armazenar o excesso dos cálculos do Acumulador (funciona como uma extensão 64bits).

# ECX: Registrador Contador

- Análogo do ‘i’ usado em programação
- Serve para controlar os loops

# EDI: Índice de Destino

- Responsável por mover os resultados de cálculos para endereços de memória RAM
- Guarda um ponteiro de endereço de memória

# ESI: Índice de Origem

- Mesma função do EDI, mas serve para leitura e não para gravação.
- Pode ser usado para outros fins.

# ESP and EBP: Ponteiro de Pilha e Ponteiro de Base

- Os dois registradores principais do x86.
- ESP retorna o endereço de um instrução da pilha.
- EBP recebe o conteúdo de ESP e passa a ser usado para os cálculos.
- Importantes instruções (PUSH, POP, CALL e RET) usam o ESP.

# EBX: Registrador de Base

- Mais usado em operações 16bits.
- Nas plataformas atuais (32 e 64bits) serve como mais um registrador de dados.

Cache

# Para que serve?

- Com o passar do tempo, a diferença de velocidade entre o processador e a memória principal está crescendo cada vez mais.
- A melhor forma de consertar isso é utilizando a memória cache.

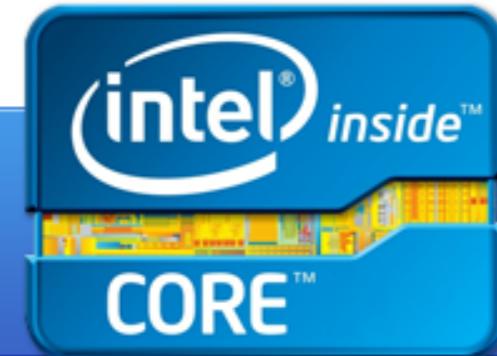
# Porque usá-la?

- A memória cache trabalha em alta velocidade e armazena poucos dados comparada à memória principal.
- O problema é econômico (cache é cara) e não tecnológico (cache é rápida)

# Características

Processadores	Fabricante	Tamanho da Cache
80486	Intel	8 KB
Athlon K7	AMD	(I - 64KB) e (D - 64KB)
Pentium	Intel	(I - 8KB) e (D - 8KB)
Pentium MMX	Intel	(I - 8KB) e (D - 24KB)
Pentium PRO	Intel	(I - 8KB) e (D - 8KB)
PowerPC 601	Motorola / IBM	32 KB
Pentium III	Intel	(I - 8KB) e (D - 24KB)
Pentium 4	Intel	(I - 8KB) e (D - 8KB)
Athlon 64	AMD	(I - 64KB) e (D - 64KB)
Power 5	IBM	64 KB
Itanium	Intel	(I - 16KB) e (D - 16KB)

# INTEL CORE QUARTA GERAÇÃO



NOME	Intel Core i7-4770TE	Intel Core i7-4770T	Intel Core i7-4770S
SOCKET	LGA1150	LGA1150	LGA1150
NÚCLEOS/THREADS	4/8	4/8	4/8
CLOCK	2,3 GHz	2,5 GHz	3,1 GHz
MÁX CLOCK	3,3 GHz	3,7 GHz	3,9 GHz
CACHE L1	128 KB (instruções) + 128 KB (dados)	128 KB (instruções) + 128 KB (dados)	128 KB (instruções) + 128 KB (dados)
CACHE L2	4 x 256 KB	4 x 256 KB	4 x 256 KB
CACHE L3	8 MB	8 MB	8 MB
TDP	45 W	45 W	65 W
GRÁFICOS	Intel® HD Graphics 4600	Intel® HD Graphics 4600	Intel® HD Graphics 4600
MEMÓRIA	32 GB (DDR3-1600)	32 GB (DDR3-1600)	32 GB (DDR3-1600)



# Princípio da Localidade

- Referências à memória feitas em qualquer intervalo curto de tempo tendem a usar apenas uma pequena fração da memória. A ideia geral é que, quando uma palavra é referenciada, ela é trazida da memória lenta (RAM) para a memória rápida (CACHE).

# Princípio da Localidade

- Princípio da Localidade Temporal : Tenderá a ser referenciada em breve novamente (loops).
- Princípio da Localidade Espacial : Itens próximos tendem a ser referenciados em breve (programa sequencial).

# Máquina executando instruções

- 1. Procurar dados (instruções) no cache.
- 2. SE dados existem no cache ENTÃO
  - 2.1. Dados são lidos do cache

# Máquina executando instruções

- 3. SENÃO
- 3.1. Procurar os dados na memória principal e copiar para o cache.
- 3.2. SE existe espaço no cache para armazenamento dos dados lidos ENTÃO
  - 3.2.1. Armazenar dados no cache.

# Máquina executando instruções

- 3.3. SENÃO
- 3.3.1. Remover dados do cache.
- 3.3.2. SE dados não foram alterados  
ENTÃO
- 3.3.2.1. Nada é feito.

# Máquina executando instruções

- 3.3.3. SENÃO
- 3.3.3.1. Copiar dados para a posição apropriada na memória principal.
- Computador mono processado
- Memória principal não precisa ser alterada imediatamente.

# Se dados do cache são modificados

- Computador monoprocessado:
  - Memória principal não precisa ser alterada imediatamente
- Computador multiprocessado (vários processadores têm acesso à mesma memória principal e mantêm seus caches privados):
  - Memória principal deve ser imediatamente atualizada quando o cache for alterado.

# Níveis de Cache

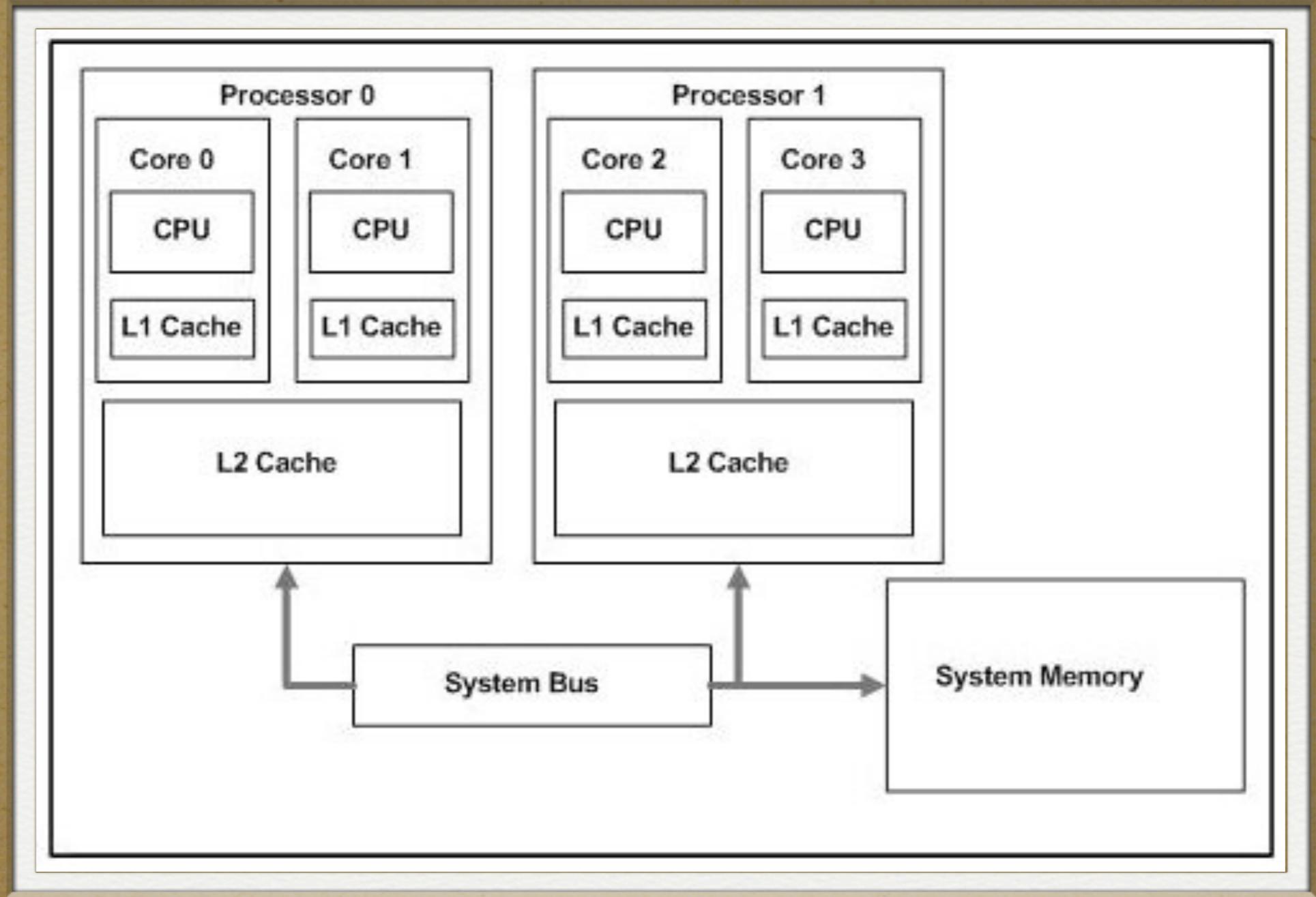
- Cache de nível 1 (L1)
  - Geralmente é dividida para instruções e dados
  - Fica localizada no mesmo chip do processador
  - Tamanho variando entre: 16 kB à 64 kB
- Cache de nível 2 (L2)
  - Fica num chip separado do processador, mas no mesmo encapsulamento
  - Tamanho variando entre: 512 kB à 1 MB

# Níveis de Cache

- Cache de nível 3 (L3)
  - Fica localizada na placa mãe
  - Contém alguns MBytes
  - Construída com tecnologia SRAM
- Os níveis funcionam de modo inclusivo, ou seja, o conteúdo do nível 1 está contido no nível 2, o conteúdo do nível 2 está contido no nível 3.

# No exemplo atual L1

- Os atuais processadores Intel contam com três níveis de memória cache. O nível L1 é dividido em duas partes: instruções e dados. Cada parte é dividida em dois, algo que propicia melhores resultados para as atuais arquiteturas da fabricante. Na tabela, foram somados os valores, por isso você vê 128 KB (instruções) + 128 KB (dados), mas o certo seria 64 KB x 2 + 64 KB x 2.



# Cache

# No exemplo atual L2

- O nível de cache L2 é composto por módulos que são utilizados de forma separada por cada núcleo. É por isso que um processador de 2 núcleos conta com  $2 \times 256$  KB de cache L2, da mesma forma que um chip de 4 núcleos utiliza  $4 \times 256$  KB.

# No exemplo atual L3

- A memória Smart Cache da Intel é o nível L3. Esse módulo é dividido entre todos os núcleos e armazena dados que podem ser úteis a todos os cores. Como você deve imaginar, quanto maior o cache L3, maior será o desempenho do processador.

# Protocolo MESI

- **Modificada:** linha da cache modificada ( $\neq$  RAM) e disponível apenas em uma cache;
- **Exclusiva:** linha da cache (= RAM) e não está presente em nenhuma outra memória cache;
- **Compartilhada:** linha da cache (= RAM) e pode estar presente em outra memória cache;
- **Inválida:** Linha da cache não contém dados válidos.

# Organização de Cache

- Cache Associativo
- Cache por mapeamento direto
- Cache associativo por conjunto

# Política de Alocação

- Estratégias : Write through e Copy back
- Write through : Quando uma palavra é escrita no cache ela é imediatamente escrita de volta à memória (maior confiabilidade → maior tráfego no barramento)
- Copy back : Não atualiza a memória quando o cache é alterado. A memória somente é atualizada quando a entrada é expurgada do cache.(mais rápida → maior complexidade)
- Se leituras >> escritas → write through
- Se escritas >> leituras → copy back

RAM

# RAM(Random Access Memory)

- Memória de acesso aleatório- capacidade de acesso a qualquer posição e em qualquer momento
- É essencial para acompanhar a velocidade doprocessador. Este tipo de memória recebe as informações doHD, e as armazena temporariamente, disponibilizando este conteúdo ao processador.
- Memória principal

# RAM(Random Access Memory)

- Quanto mais memória, melhor o desempenho - sem uma quantidade suficiente dela o sistema passa a usar memória virtual, que é lenta.
- Memória volátil
- É nesta memória que são carregados os programas em execução e os respectivos dados do utilizador. Ao ligar é necessário refazer todo o processo de carregamento, em que o sistema operacional e aplicativos usados são transferidos do HD para a memória, onde podem ser executados pelo processador.

# RAM

- A memória RAM é um chip semelhante a um micro-processador, composto por milhões de transistores e capacitores. O capacitor é uma peça capaz de armazenar elétrons. Quando ele está carregado, o sistema faz uma leitura com base no famoso código binário de “zeros e uns”. Cada leitura dessa em zero ou um significa um bit de informação. Essa leitura é feita de forma muito rápida, são muitas em poucos milésimos de segundos.
- Os chips de memória são vendidos na forma de pentes de memória. Existem pentes de várias capacidades, e normalmente as placas possuem dois ou três encaixes disponíveis.
- O primeiro tipo de memória RAM foi a núcleo magnético, desenvolvida de 1955 a 1975 e, posteriormente, utilizada na maioria dos computadores até o desenvolvimento e adoção da estática e dinâmica de circuitos integrados RAM no final dos anos 1960 e início de 1970.

# Capacidade e Velocidade

- Capacidade - medida em Bytes, KB, MB, GB, TB
- Velocidade - Hz ou Mhz. Relacionado com a quantidade de blocos de dados que podem ser transferidos durante um segundo.
- Algumas memórias RAM podem efetuar duas transferências de dados no mesmo ciclo de clock, duplicando a taxa de transferência de informação para a mesma frequência de trabalho.
- Colocação das memórias em paralelo (propriedade da arquitetura de certos sistemas) permite multiplicar a velocidade aparente da memória.

# Largura e velocidade do barramento

- A largura nos diz o número de bits que podem ser enviados ao CPU simultaneamente.
- A velocidade é o número de vezes que esse grupo de bits pode ser enviado a cada segundo.
- A memória comunica-se com o CPU, trocando dados, e completa o que se conhece como ciclo de barramento. É esse período que apresenta o desempenho da memória que, pode ser de 100MHz e 32bits, por exemplo. Isto singnifica que tal memória é capaz de enviar 32bits de dados ao processador 100 milhões de vezes por segundo.

# TIPOS

# DRAM(memória dinâmica de acesso aleatório)

- Mais barata e comum
- Um transistor e um capacitor unem-se para formar uma célula de memória, que é responsável por um bit de dados. Enquanto o capacitor conserva o bit de informação, o transistor age como um controle, que permite ao chip ler o capacitor ou mudar seu estado.
- Mais lenta, pois passa por um processo de refrescamento dos dados, o que leva tempo e deixa a memória lenta.

# SRAM (memória estática de acesso aleatório)

- Composta por Flip-flops, que contém quatro ou seis transistores e fios.
- A vantagem desse tipo é que não há necessidade de ser refrescada.
- Mais rápida
- Ocupa bem mais espaço em um chip que uma célula de memória dinâmica. O que resulta na menor quantidade de memória que se pode ter por chip, fazendo da SRAM um componente bem mais caro.

# Dual Channel

- Característica que dá a possibilidade do chipset ou o processador de comunicarem-se com duas vias de memória ao mesmo tempo. O resultado é o fornecimento do dobro de largura de dados do barramento.

# TIPOS

- SDRs são o tipo tradicional, onde o controlador de memória realiza apenas uma leitura por ciclo - usados em micros antigos: Pentium II e Pentium IIIe os primeiros Athlons e Durons soquete. Por não serem mais fabricados, eles são atualmente muito mais raros e caros que os DDR . Possuem um chanfro.
- DDR(Double Data Rate(taxa dupla de transferência)) - são mais rápidas, pois fazem duas leituras por ciclo. Possuem dois chanfros.

# TIPOS

- DDR2 - O barramento de acesso à memória trabalha ao dobro da frequência dos chips de memória propriamente ditos. Isso permite que sejam realizadas duas operações de leitura por ciclo, acessando dois endereços diferentes. Como a capacidade de realizar duas transferências por ciclo introduzida nas memórias DDR foi preservada, as memórias DDR2 são capazes de realizar um total de 4 operações de leitura por ciclo, uma marca impressionante. Existem ainda alguns ganhos secundários, como o menor consumo elétrico, útil em notebooks. Incompatíveis com as placas-mães antigas. Possuem um número maior de contatos (um total de 240, contra 184 dos pentes DDR), e o chanfro central é posicionado de forma diferente, de forma que não seja possível instalá-los nas placas antigas por engano. Muitos pentes são vendidos com um dissipador metálico, que ajuda na dissipaçāo do calor e permite que os módulos operem a frequências mais altas.

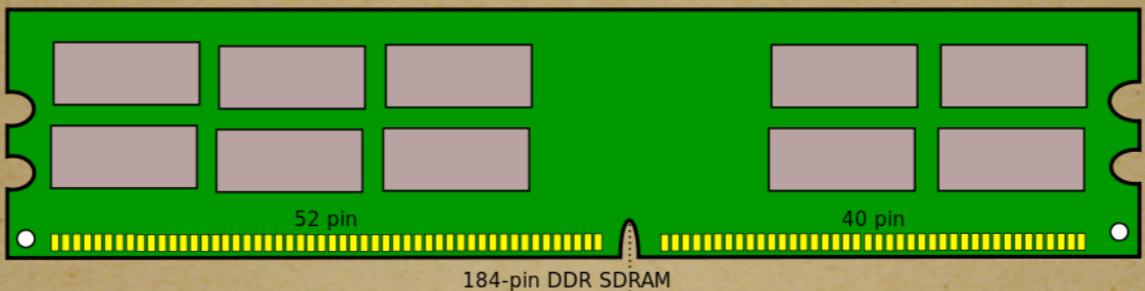
# TIPOS

- O modelos para desktops que ficaram mais conhecidos são: o SIMM, single in-line memory module; o DIMM, dual in-line memory module e o RIMM, Rambus in-line memory module
- Já o modelo SODIMM serve para notebooks. Essas são memórias que se diferenciam essencialmente nos modos como organizam os pinos nos conectores e transmitem os dados.

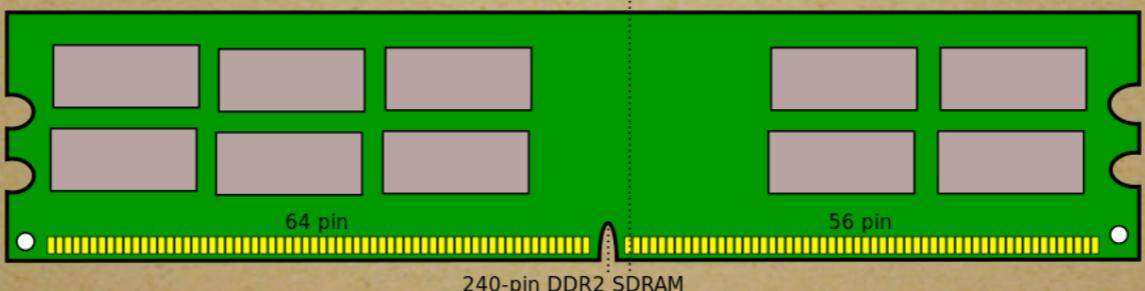
# O padrão atual DDR3

- Possível taxa de transferência duas vezes maior, de modo que permite taxas de barramento maiores, como também picos de transferência mais altos.
- Não há redução significativa de latência.
- Consome cerca de 30% menos energia (Trabalha com voltagem de 1.5V, menor que a 1.8V da DDR2 e os 2.5V da DDR).
- Alta taxa de transferência, possível graças ao buffer de 8 bits; diferente dos 4 bits da DDR2 ou dos poucos 2 bits de buffer da DDR.
- O Intel Core i7, lançado em novembro de 2008, suporta apenas DDR3, conectando-se diretamente à memória ao invés de um chip intermediário.
- Os DIMMs (Módulos de Memória em Linha Dupla) da DDR3 tem 240 pinos, o mesmo número da DDR2, e o mesmo tamanho; mas são eletricamente incompatíveis, além de possuir diferentes localizações.

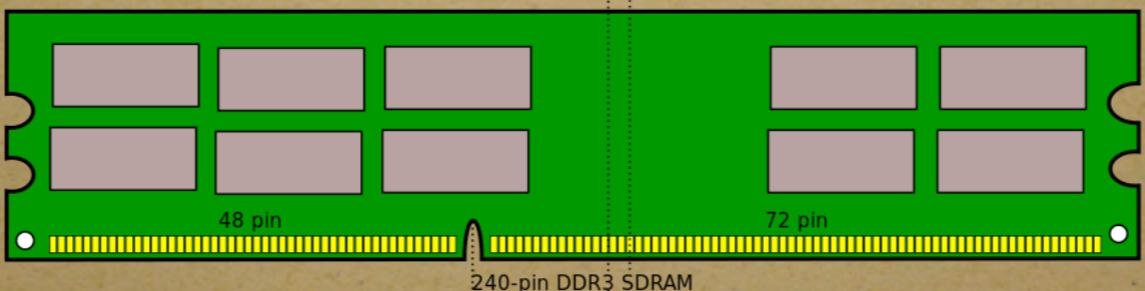
# DDR



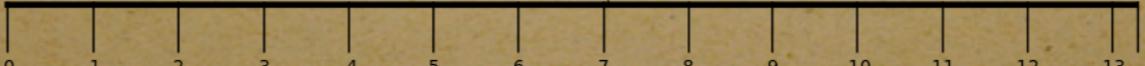
# DDR 2



# DDR 3



cm.



# DDR

ROM

# ROM

- Você certamente já se deparou com nomes como CD-ROM, DVD-ROM, entre outros.
- O ROM significa que essas mídias foram gravadas e não podem mais ser alteradas. Isto é, que o conteúdo delas apenas está disponível para leitura.

# ROM

- Memória ROM (Read Only Memory) – memória apenas para leitura, isto é, os dados armazenados nela não podem ser alterados. É uma memória de acesso aleatório, porém não é volátil, ou seja, é uma memória permanente, significa que as informações não se perdem quando o microcomputador é desligado ou quando falta energia elétrica. As informações contidas nesse tipo de memória são as gravadas pelo fabricante e os programas básicos do sistema operacional, que são executados quando a máquina é ligada.

# ROM

- Basicamente, essa é a função da memória ROM: oferecer dados apenas para leitura. Normalmente, a ROM é utilizada para armazenar firmwares, pequenos softwares que funcionam apenas no hardware para o qual foram desenvolvidos e que controlam as funções mais básicas do dispositivo.

# ROM

- Dentro da memória ROM do computador estão gravados três programas: BIOS, POST e Setup.
- O POST (Power on Self Test) é o autoteste que o micro executa sempre em que é ligado (contagem de memória etc.).
- O Setup é o programa de configuração da máquina, e é através dele que configuramos o tipo de disco rígido e outras opções relacionadas a configuração de hardware do sistema.

# ROM

- A BIOS (Basic Input Output System) é um chip que carrega as configurações mais básicas do sistema antes de inicializar o sistema operacional propriamente dito. Ela verifica se a data e hora estão certas, se a ventoinha do processador está operando, se os diversos periféricos e controladores estão recebendo tensão, bem como se as memórias RAM estão prontas para trabalhar para, enfim, "chamar" o HD que acordará o sistema operacional.

# ROM

- Essas instruções e regulagens estão escritas em um software que, em última análise, é a ROM em si. É esta analogia, do software embutido em um dispositivo, que estende o conceito de ROM para diversos outros aparelhos e sistemas.

# ROM

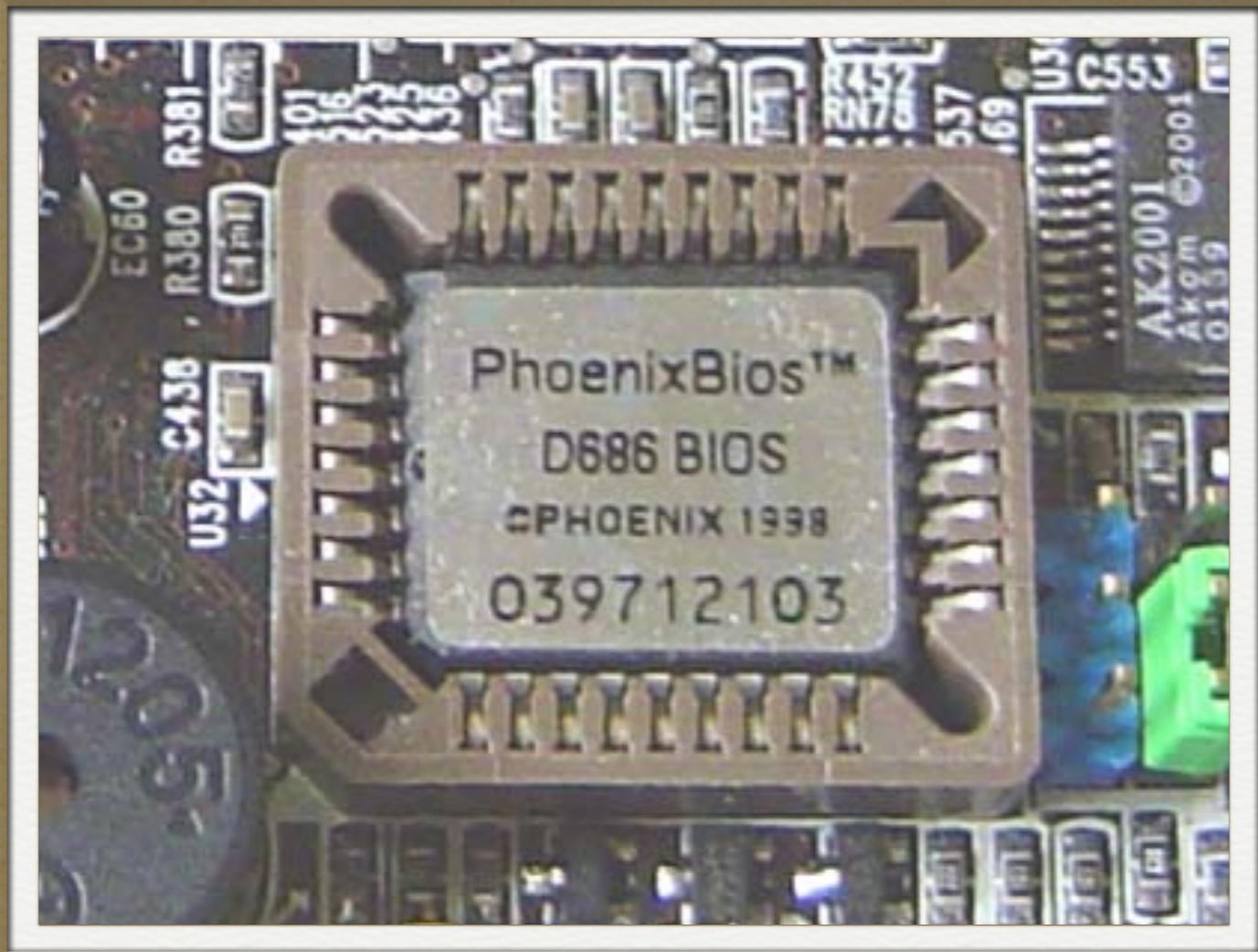
- Em relação a fabricação, uma ROM é fabricada como qualquer outra pastilha de circuito integrado, cujos dados são gravados na pastilha durante o processo de fabricação. Isso apresenta dois problemas:
  1. A etapa de gravação de dados tem um custo fixo relativamente alto, que não depende do número de cópias produzidas.
  2. Não podem ocorrer erros: se algum bit estiver errado, todo o lote de memória ROM será inutilizado.

# IMPORTANTE

- Apesar dos três programas citados acima estarem gravados na memória ROM, com o passar do tempo, muita gente passou a chamar a memória ROM do computador de BIOS, embora isto esteja muito errado. E da mesma forma que confudem ROM com BIOS, muitas pessoas costumam falar “Setup” da “BIOS”, enquanto Setup e BIOS são dois programas distintos dentro da memória ROM. Ou seja, BIOS é o programa que está gravado dentro da memória ROM (que é um chip) do computador. Setup e POST são outros dois programas que estão gravados no mesmo chip, favor não confundir.



ROM



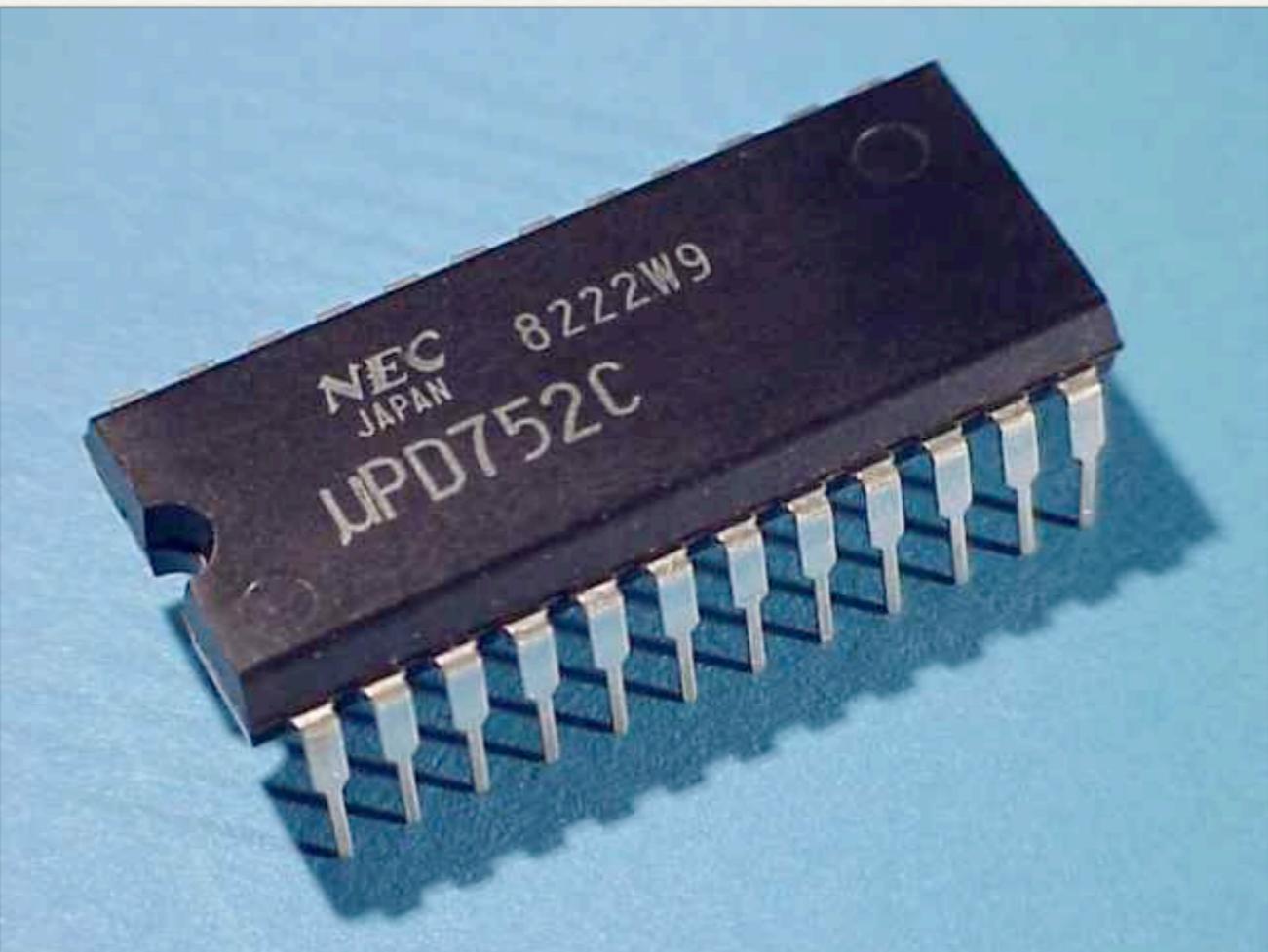
ROM

# Tipos de ROM

- Mask-ROM
- PROM
- EPROM
- EEPROM ou E2PROM
- EAROM
- Flash

# Mask-ROM

- As primeiras ROMs a serem desenvolvidas são as chamadas Mask-ROM, e são nada mais do que circuitos integrados que guardam o software ou os dados gravados durante a sua criação e que não está passível de qualquer tipo de alteração. Podemos compará-las com os CD-ROMs: o usuário acessa aquilo que comprou e não pode gravar outros dados na mídia ou chip e também a eletrodomésticos com funções digitais, como o micro-ondas (ainda não inventaram uma forma de atualizar o firmware de um deles).



# Mask-ROM

# PROM

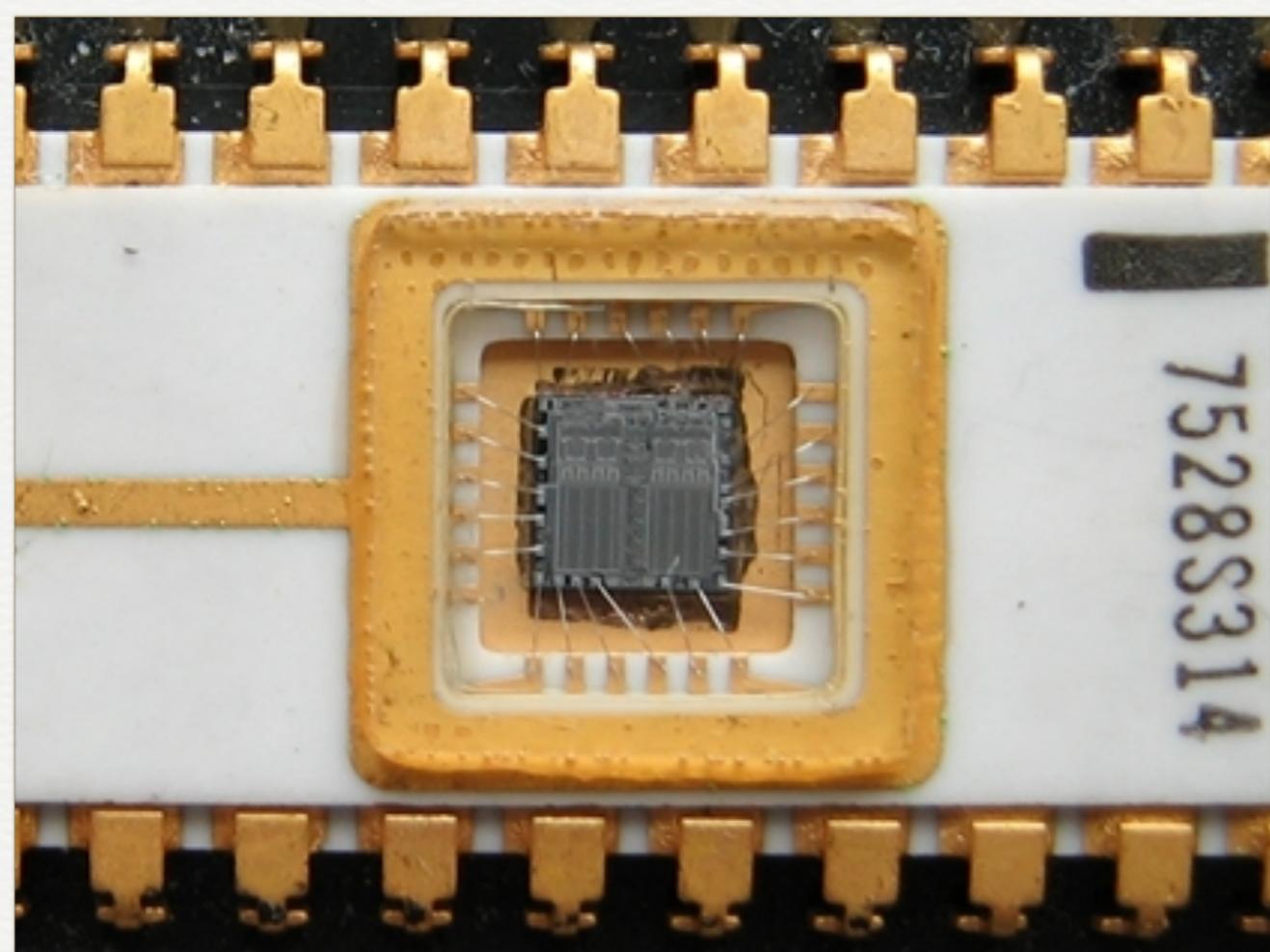
- Com o passar do tempo, foram necessárias memórias similares, mas que possibilitassem a inserção posterior de dados. A primeira dessa nova leva foi a Programmable Read-Only Memory (PROM), que permite que o conteúdo seja modificado por meio de um dispositivo conhecido como programador PROM. Porém, como o programador PROM altera fisicamente as ligações internas do chip, essa inserção pode acontecer apenas uma vez. Esse tipo de ROM pode ser encontrado em consoles de video games e em aparelhos de celulares. Além disso, podemos comparar a PROM com o CD gravável (CD-R), que também suporta apenas uma gravação.



PROM

# EPROM

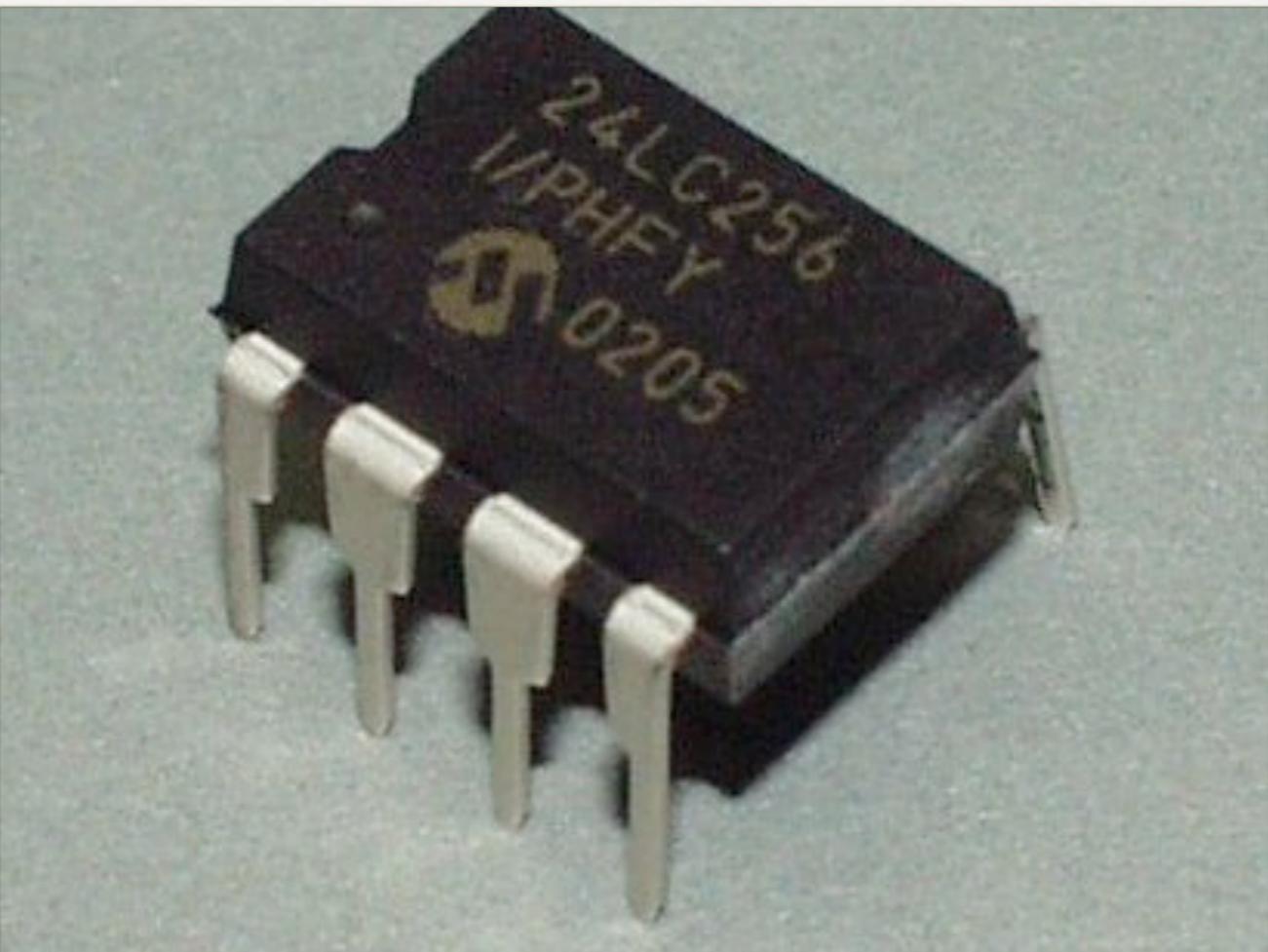
- Outro tipo muito usado é o Erasable Programmable Read-Only Memory (EPROM). É a primeira forma de ROM que pode ser zerada e reescrita. Basicamente, consiste em expor o chip à luz ultravioleta por um espaço de tempo determinado para apagá-lo, tempo esse que depende da idade da memória. A exposição zera o chip e a partir daí os dados podem ser reescritos. O processo de reescrita dos dados requer uma voltagem cada vez maior e, com isso, a número de reprogramações acaba sendo limitado. A grande inovação da EPROM é permitir a regravação de dados.



# EPROM

# EEPROM ou E2PROM

- Um tipo mais recente é a Electrically Erasable Programmable Read-Only Memory (EEPROM) que, como o próprio nome indica, permite que os dados sejam apagados e gravados com o uso de eletricidade, fazendo com que não seja necessário mover o dispositivo de seu lugar para um aparelho especial para que a regravação ocorra. É a mais utilizada pela indústria atualmente, e está presente na BIOS do seu computador, nos consoles de última geração, bem como no celular. Através dela, por exemplo, é possível atualizar o firmware de uma câmera ou de um MP3 Player de maneira muito mais prática.



EEPROM  
ou E2PROM

# EAROM

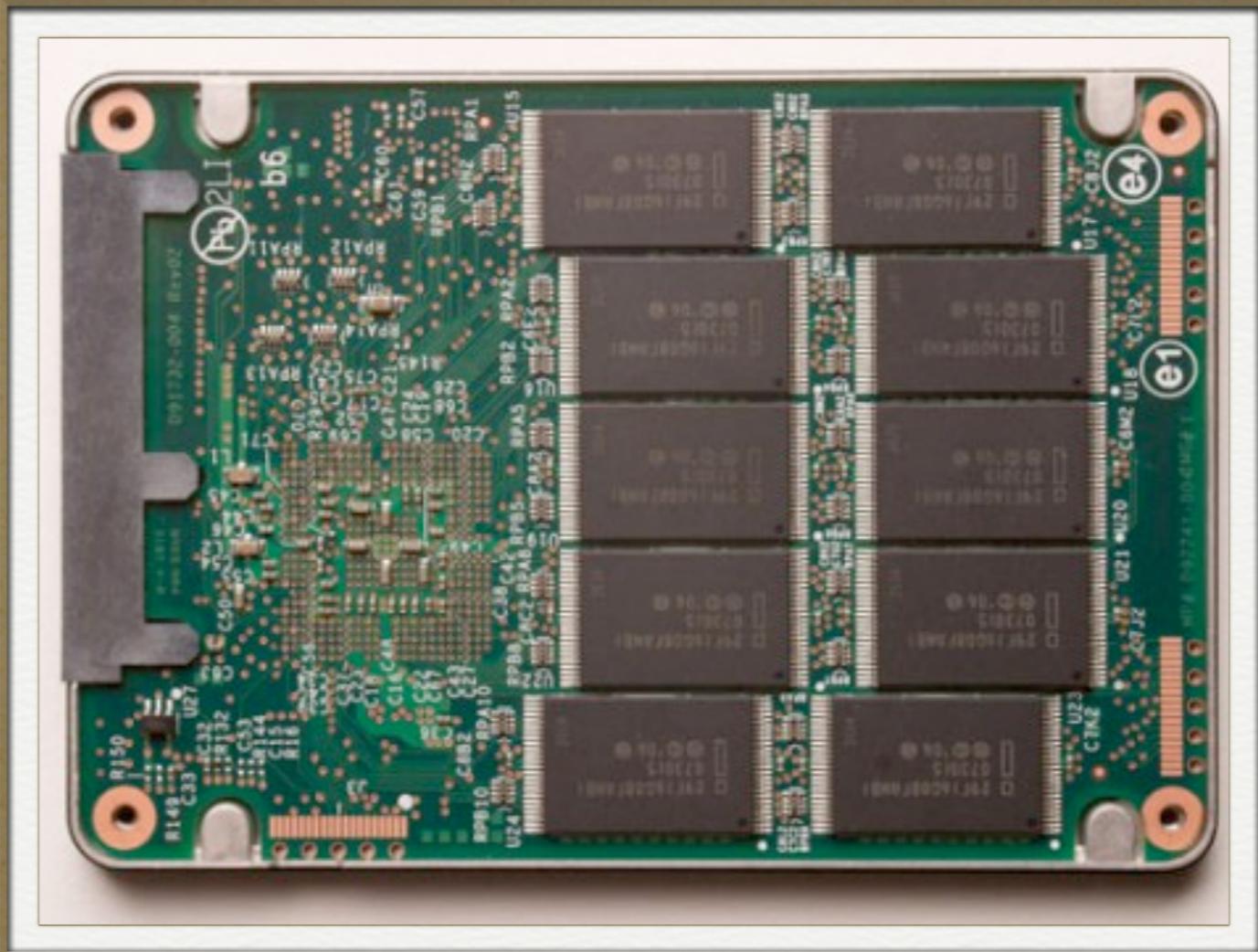
- As memórias EAROM (Electrically-Alterable Programmable Read-Only Memory ) podem ser vistas como um tipo de EEPROM. Sua principal característica é o fato de que os dados gravados podem ser alterados aos poucos (permite a alteração de um bit por vez do seu conteúdo), razão pela qual esse tipo é geralmente utilizado em aplicações que exigem apenas reescrita parcial de informações.



# EAROM

# Flash

- As memórias Flash também podem ser vistas como um tipo de EEPROM, no entanto, o processo de gravação (e regravação) é muito mais rápido, além de durar muito mais, possibilitando mais de 1 milhão de ciclos de reprogramação. A memória Flash difere das EEPROMs comuns que apagam a sua memória reescrevendo conteúdo ao mesmo tempo, o que as torna mais lentas para atualizar. A memória Flash pode apagar os dados em blocos inteiros, tornando-se a tecnologia preferida para aplicações que requerem uma atualização frequente de grandes quantidades de dados, como no caso de um cartão de memória para um dispositivo eletrônico digital.



# Flash

# GRUPO

- Calógeras Lima
- Filipe Villa Verde
- Maria Lima
- Pedro Magalhães
- Raphael Brito