



Instituto de Matemática
Departamento de Ciência da Computação

Arquitetura de Computadores

Visão geral da memória e memória interna (principal)

Prof. Marcos E Barreto

Tópicos

- Visão geral do sistema de memória
 - Organização da memória principal
 - Considerações sobre correção de erros
 - Organizações avançadas de DRAM
-
- Referência:
 - Stallings. Arquitetura e organização de computadores. Cap. 5.



Visão geral do sistema de memória

- Características dos sistemas de memória

Localização	Desempenho
Interna X Externa	Tempo de acesso
Capacidade	Tempo de ciclo
Nº palavras X nª bytes	Tempo de transferência
Unidade de transferência	Tipo físico
Bloco X palavra	Semicondutor
Método de acesso	Magnético
Sequencial	Óptico
Direto	Magneto-óptico
Aleatório	Características físicas
Associativo	Volátil X não-volátil
	Apagável X não-apagável
	Organização
	Módulo de memória

Visão geral do sistema de memória

- **Localização**

- Interna

- Diretamente acessível pelo processador
 - Ex: registradores, cache, memória principal e memória da UC.

- Externa

- Acessível via controlador de E/S
 - Ex: discos, fitas, unidades externas em geral.

Visão geral do sistema de memória

- **Capacidade**

- Normalmente expressa em bytes (1 byte = 8 bits) ou palavras.
- Tamanhos de palavra: 8, 16 e 32 bits.

Visão geral do sistema de memória

- **Unidade de transferência (UT)**

- Memória interna

- UT é igual ao número de linhas elétricas que entram e saem do módulo de memória => número de bits lidos ou escritos na memória de uma só vez.
- Pode ser igual ao **tamanho da palavra** (quantidade de bits usados para representar uma instrução ou um valor inteiro), mas geralmente é maior (64, 128 ou 256 bytes)
- UT depende da unidade endereçável (geralmente, é a palavra).

- Memória externa

- A UT é muito maior que uma palavra.
- A transferência é feita em **blocos**.

Visão geral do sistema de memória

- **Método de acesso**

- Acesso sequencial

- Memória organizada em registros, com acesso linear.
 - Mecanismo de leitura/escrita compartilhado.
 - Ex.: fitas

- Acesso direto

- Blocos e registros têm um endereço exclusivo, baseado no local físico.
 - Mecanismo de leitura/escrita compartilhado
 - O acesso direto alcança uma “vizinhança geral”, sendo seguido de um acesso linear até a posição final.
 - Ex.: discos

Visão geral do sistema de memória

- **Método de acesso**

- Acesso aleatório

- Cada posição de memória tem um mecanismo de endereçamento exclusivo, fisicamente interligado.
 - Tempo para acessar um local é constante e independente dos acessos anteriores.
 - Ex.: memória principal e alguns tipos de cache.

- Associativo

- Memória de acesso aleatório que permite fazer uma comparação de um certo número de bits desejados dentro de uma palavra para uma combinação específica.
 - Tal operação é feita em todas as palavras de memória.
 - Palavra é recuperada com base no seu conteúdo (parte dele).
 - Ex.: memória cache

Visão geral do sistema de memória

- **Desempenho**

- Tempo de acesso (latência)
 - Tempo gasto para uma operação de leitura ou escrita, desde o instante em que o endereço a ser acesso é apresentado à memória até que o dado esteja disponível para leitura ou tenha sido escrito => memória aleatória
 - Tempo gasto para movimentar o mecanismo de leitura/escrita ao local desejado => memória sequencial ou de acesso direto.

Visão geral do sistema de memória

- **Desempenho**

- Tempo de ciclo de memória
 - Aplicado principalmente à memória de acesso aleatório.
 - Consiste no tempo de acesso mais algum tempo adicional até que um segundo acesso possa ser feito.
 - Depende do barramento do sistema, e não do processador!

Visão geral do sistema de memória

- **Desempenho**

- Taxa de transferência

- Taxa em que os dados podem ser lidos ou escritos na memória.
- Para memória de acesso aleatório: $1 / \text{tempo de ciclo}$
- Para memória sequencial/direto: $TN = TA + \frac{n}{R}$
 - TN = tempo médio para ler ou escrever N bits
 - TA = tempo de acesso médio
 - n = número de bits
 - R = taxa de transferência em bits por segundo (bps)

Visão geral do sistema de memória

- **Tecnologias**

- Memória semicondutora
- Memória magnética
- Memória óptica
- Memória magneto-óptica

- **Características físicas**

- Volátil x não-volátil (caso especial: ROM)

- **Organização**

- Arranjo físico de bits para formar as palavras

Hierarquia de memória

- **Questões de projeto**

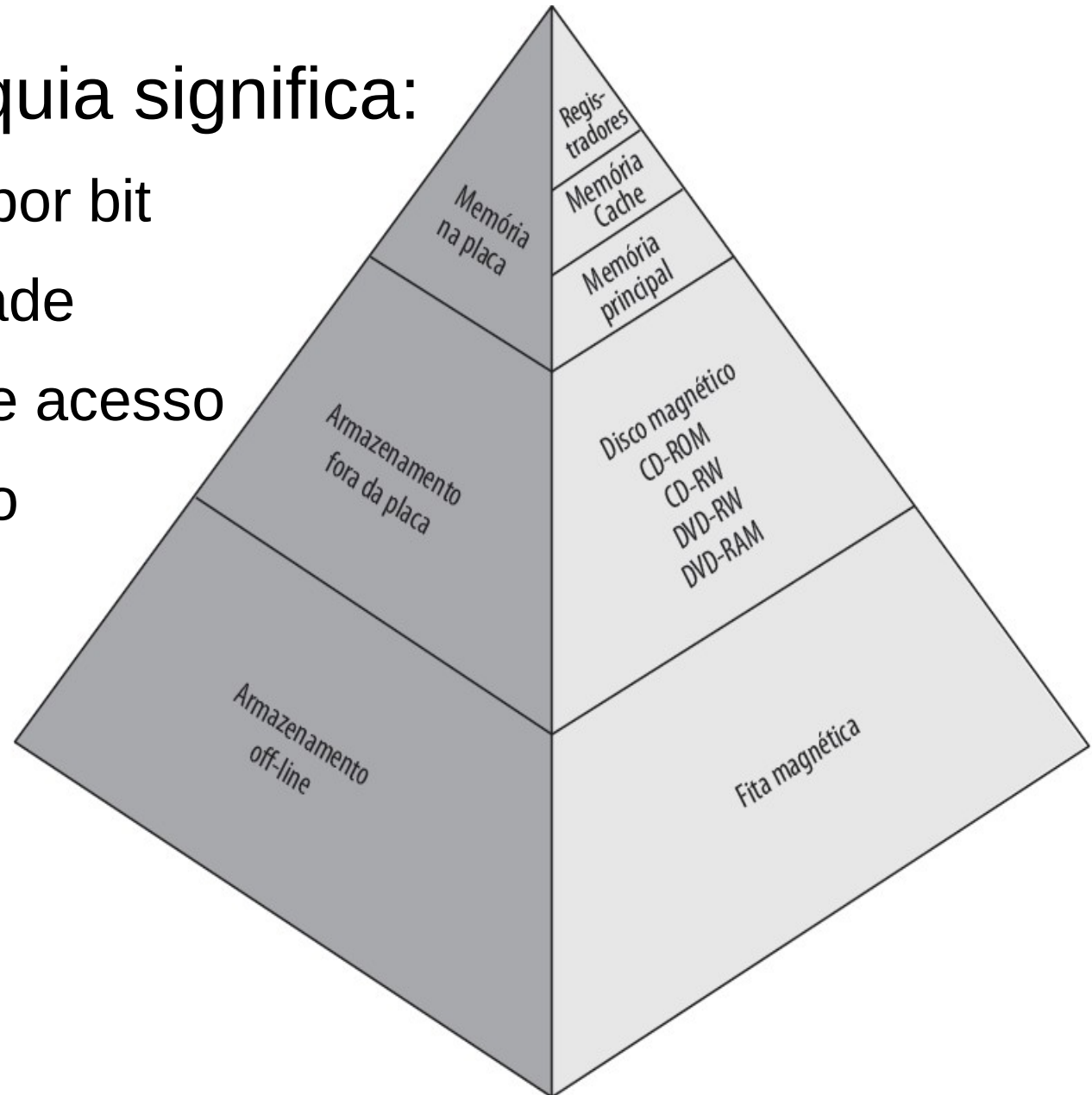
- Quanto?
- Com que velocidade?
- Com que custo?

- **Relações**

- Tempo de acesso mais rápido: maior custo por bit
- Maior capacidade: menor custo por bit
- Maior capacidade: tempo de acesso mais lento

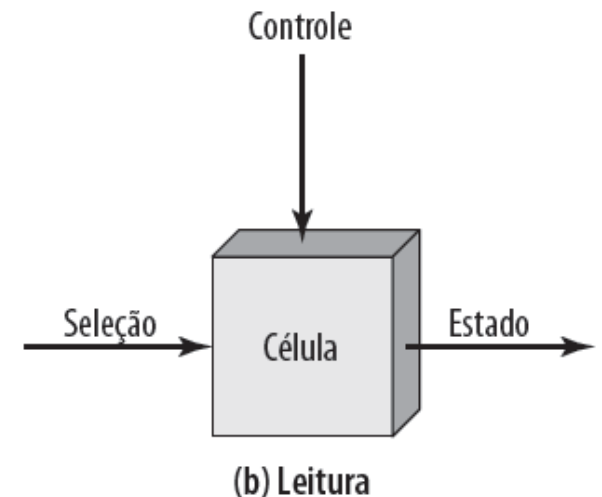
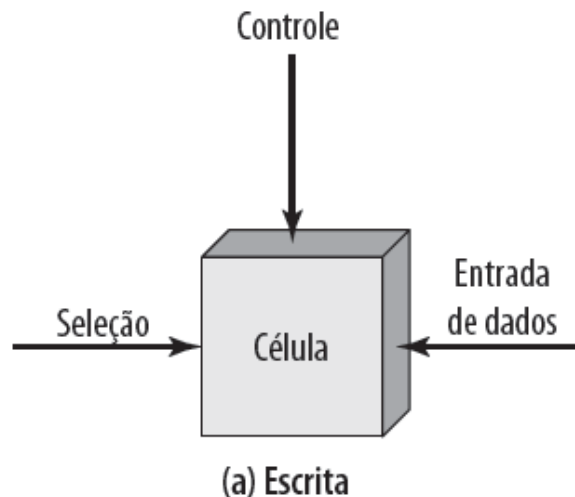
Hierarquia de memória

- Descer na hierarquia significa:
 - a) Diminuição no custo por bit
 - b) Aumento da capacidade
 - c) Aumento do tempo de acesso
 - d) Frequência de acesso



Memória principal

- Armazenamento de acesso aleatório
 - Memória semicondutora X memória de núcleo ferromagnético
- Organização básica: **célula de memória**
- Propriedades:
 - Dois estados estáveis: 0 e 1
 - Podem ser escritas (pelo menos uma vez) para definir o estado
 - Podem ser lidas para verificar o estado



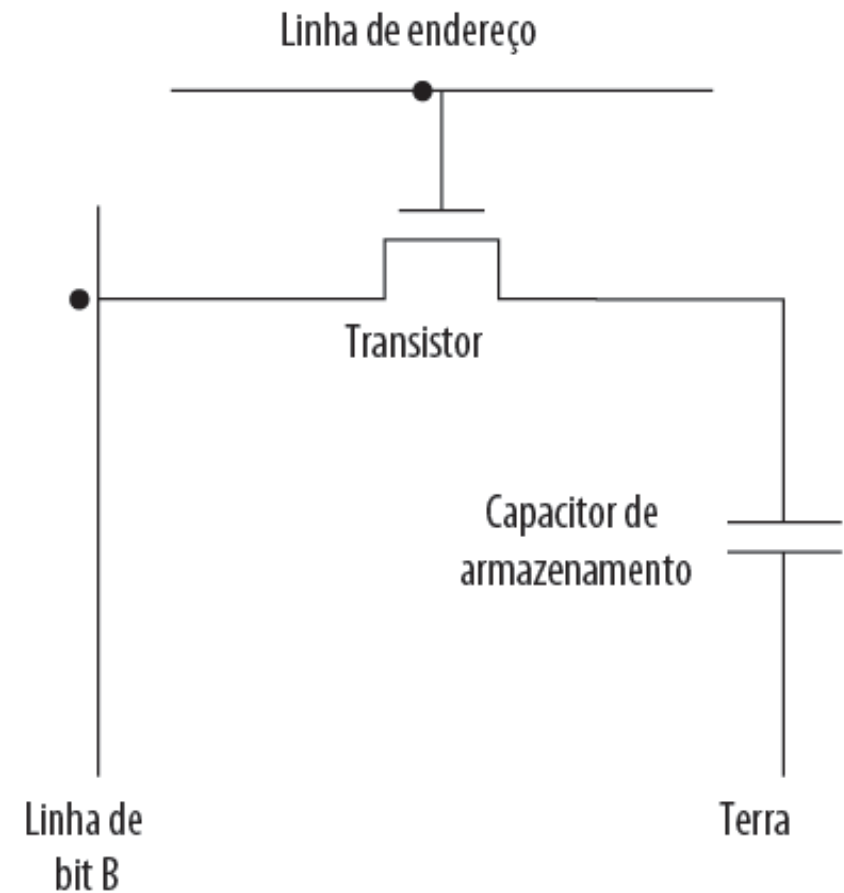
Tipos de memória

- Todos os tipos de memória são de acesso aleatório
 - Palavras individuais acessadas por meio da lógica interna de endereçamento.

Tipo de memória	Categoria	Apagamento	Mecanismo de escrita	Volatilidade
RAM	Leitura-escrita	Eletricamente, em nível de byte	Eletricamente	Volátil
ROM	Somente leitura	Não é possível	Máscaras	Não-volátil
PROM			Eletricamente	
EPROM	Luz UV, nível de chip			
EEPROM	Eletricamente, nível de byte			
Flash	Eletricamente, nível de bloco			

RAM (*random access memory*)

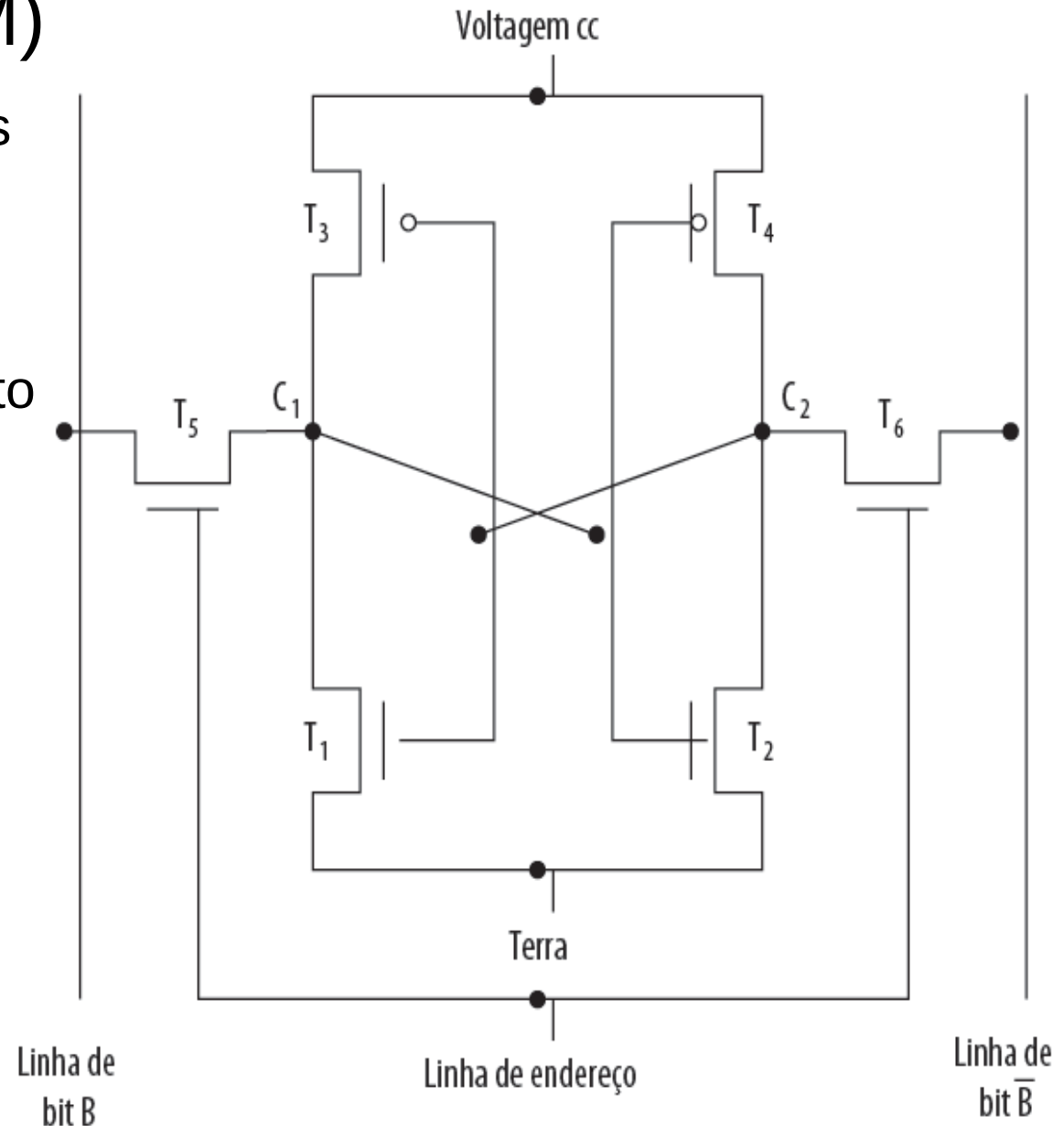
- Leitura e escrita de forma aleatória
- Memória volátil
- RAM dinâmica (DRAM)
 - Células que armazenam carga em capacitores
 - Exigem recarga periódica (*refresh*) => dinâmicas



RAM (*random access memory*)

- RAM estática (SRAM)

- Mesmos elementos lógicos de um processador
- Portas lógicas e flip-flops
- Armazena valores enquanto houver energia



DRAM x SRAM

- São memórias voláteis => potência deve ser continuamente fornecida para manter o valor
- DRAM
 - Célula é mais simples e menor => memória mais densa e barata
 - Requer suporte de um circuito de recarga (refresh)
 - Adequadas para requisições de grande capacidade => memória principal
- SRAM
 - Usadas em memória cache (no chip e fora dele).

ROM (*read-only memory*)

- Contém um padrão permanente de dados, que não pode ser alterado.
- Memória não-volátil => não requer fonte de energia
- Aplicações
 - Microprogramação, bibliotecas de funções de uso frequente, programas do sistemas e tabelas de funções
- Gravação de dados é parte do processo de fabricação da ROM
 - Custo fixo grande, independente da quantidade de memórias fabricadas
 - Não permite erros => descarte do lote inteiro

ROM (*read-only memory*)

- ROM programável (*Programmable ROM*)
 - Não-volátil
 - Pode ser escrita uma vez, de forma elétrica, através de equipamentos específicos
 - Pequeno número de ROM com determinado conteúdo
- Memória principalmente de leitura
 - Aplicações com muitas operações de leitura e poucas operações de escrita
 - Três tipos: EPROM, EEPROM e memória flash

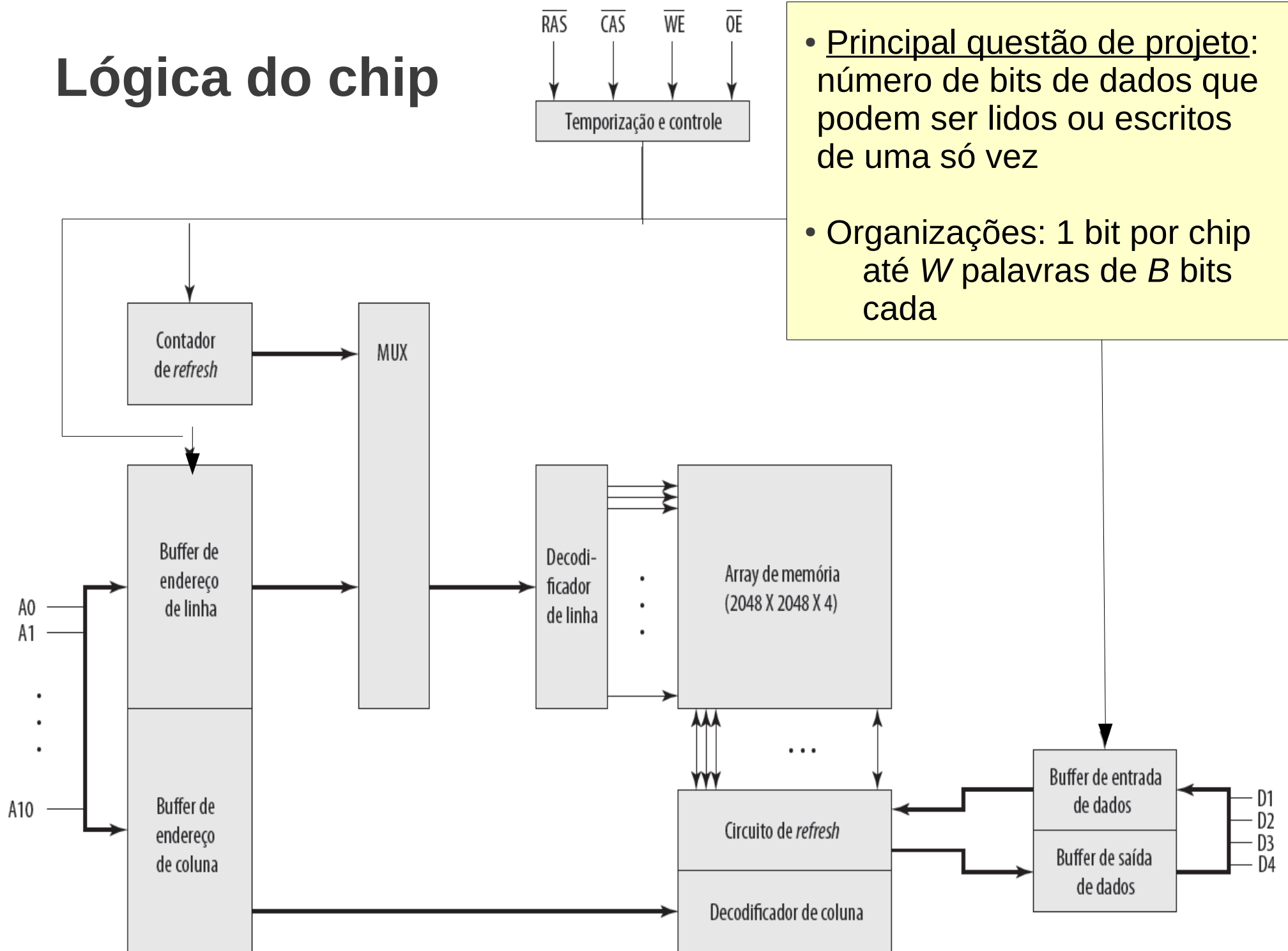
Memória principalmente de leitura

- PROM apagável (*Electrically PROM*)
 - Lida e escrita eletricamente
 - Processo de apagamento é feito através de radiação ultravioleta
- PROM apagável eletricamente (*Electrically-Erasable PROM*)
 - Pode ser escrita a qualquer momento, sem necessidade de apagamento
 - Vantagem da não-volatilidade combinada com atualização no local, através dos barramentos de dados e de endereço
 - Mais cara que a EPROM e menos densa

Memória principalmente de leitura

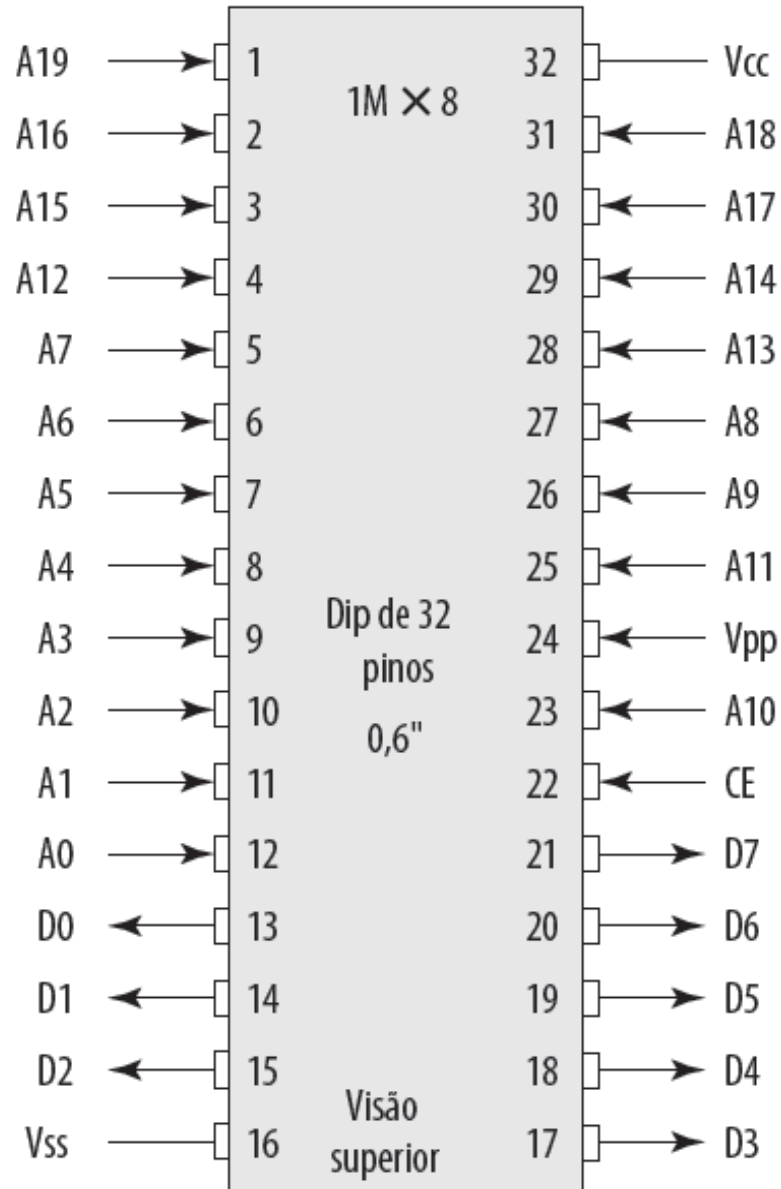
- Memória flash
 - Flash => velocidade de reprogramação
 - Memória intermediária entre EPROM e EEPROM quanto a custo e funcionalidade
 - Blocos da memória flash podem ser apagados eletricamente em poucos segundos
 - Memória com alta densidade => um transistor por bit

Lógica do chip



Empacotamento e organização do chip

Cápsula EPROM = chip 8Mbits (1M x 8)



A0 – A19: endereço da palavra

D0 – D7: dados

Vcc: fonte de alimentação

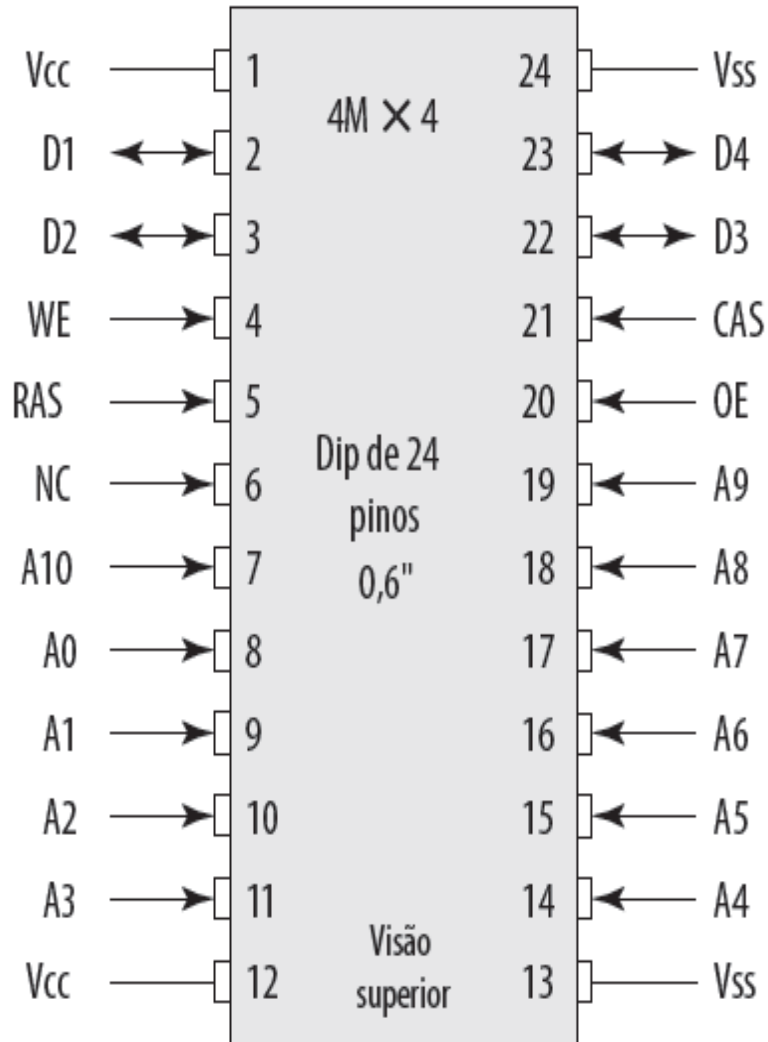
Vss: pino terra

CE: indica se o endereço é válido para este chip

Vpp: voltagem de programa

Empacotamento e organização do chip

Cápsula DRAM = chip 16Mbits (4M x 4)



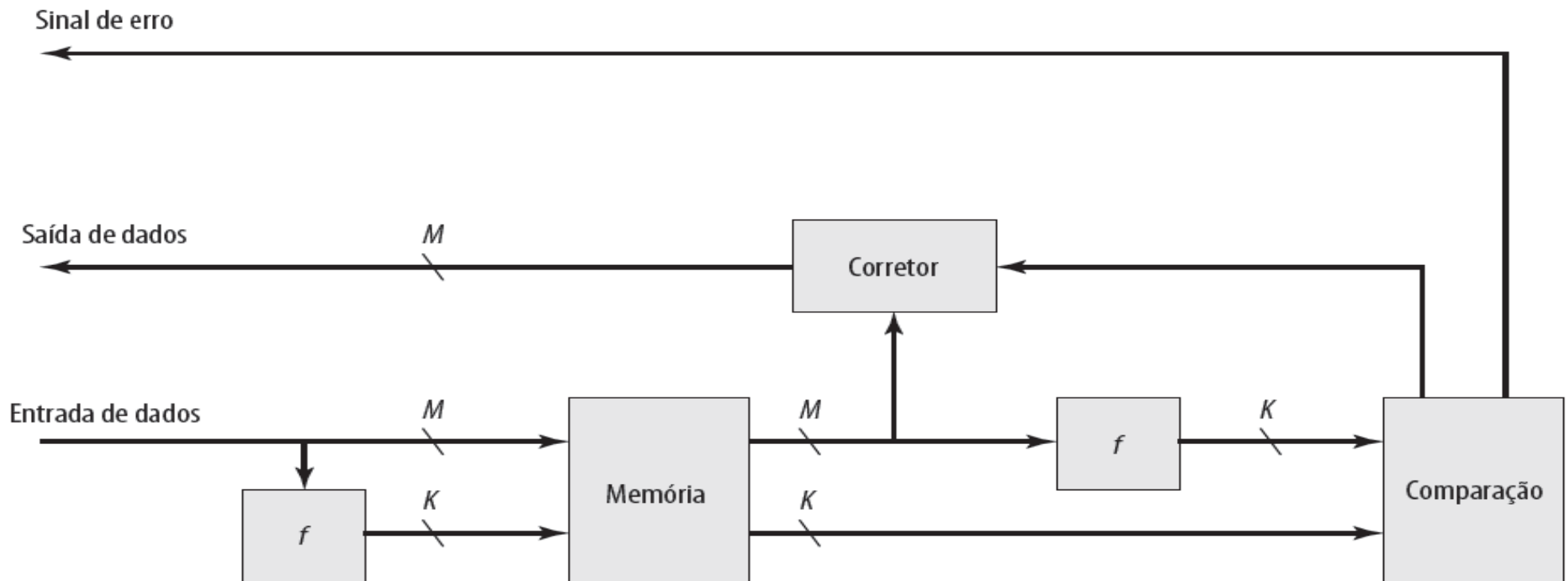
A0 – A10: endereço da palavra
D1 – D4: dados (entrada e saída)
Vcc: fonte de alimentação
Vss: pino terra
WE e OE: indicam leitura ou escrita
NC = nenhuma conexão

Memória intercalada

- Os chips que compõem a memória principal podem ser agrupados para formar um banco de memória.
- Os bancos de memória podem ser intercalados (*interleaved memory*)
- Cada banco é capaz de atender a uma requisição de leitura ou escrita => k bancos atendem k solicitações simultâneas.
- Se palavras consecutivas de memória forem armazenadas em diferentes bancos, pode-se agilizar a transferência de um bloco de memória.
 - DICA: apêndice E do livro do Stallings sobre memória intercalada.

Considerações sobre correção de erros

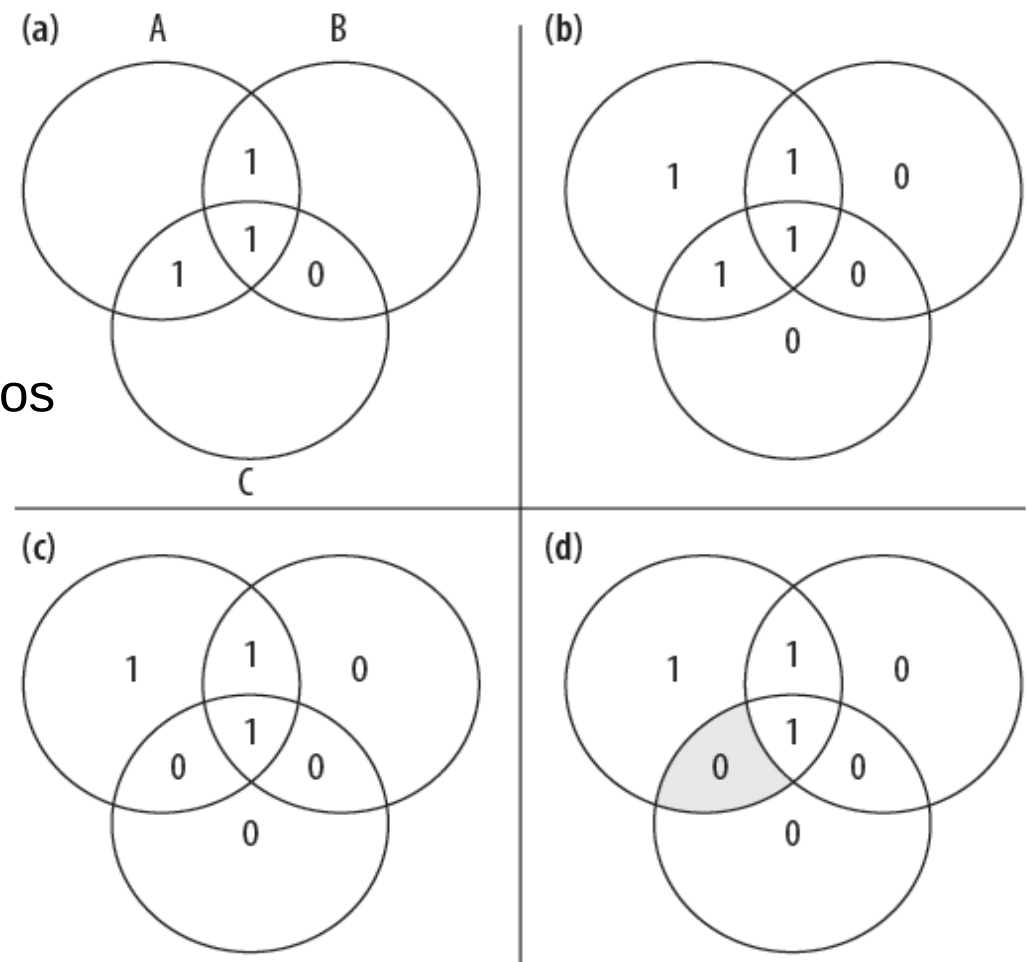
- Falha permanente X erros não-permanentes
 - Falha permanente = defeito físico
 - Erro não-permanente = evento aleatório, não-destrutivo
- Uso de técnicas de detecção e correção de erros



M = palavra de dados K = código de verificação f = função verificadora

Considerações sobre correção de erros

- Códigos de correção de erros
 - Número de erros de bits em uma palavra que podem ser corrigidos
- Exemplo:
- **código de Hamming**
 - Palavra de 4 bits atribuídos às regiões internas dos círculos
 - Bit de paridade par na parte externa de cada círculo



Organizações avançadas de DRAM

- Chip DRAM: bloco básico para montagem de memória
 - Restrições quanto à arquitetura interna e interface com o barramento de memória do processador
- Inserção de vários níveis de cache SRAM para melhorar o desempenho
- Melhorias na arquitetura básica da DRAM

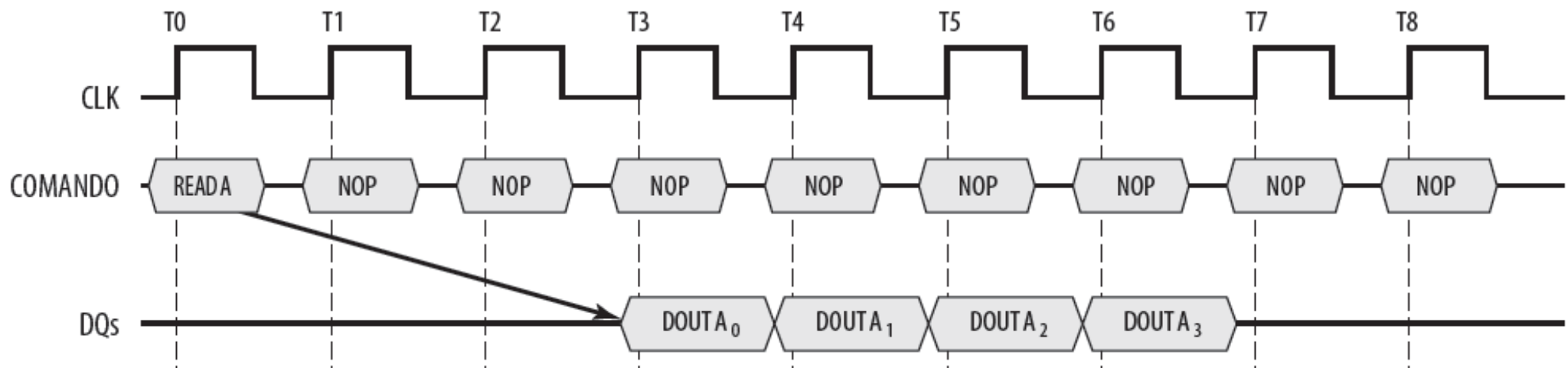
Comparação de desempenho de algumas alternativas à DRAM

Tipo de memória	Frequência de clock (MHz)	Taxa de transferência (GB/s)	Tempo de acesso (ns)	Contagem de pinos
SDRAM	166	1,3	18	168
DDR	200	3,2	12,5	184
RDRAM	600	4,8	12	162

Organizações avançadas de DRAM

- DRAM síncrona (SDRAM)
 - Comunicação com processador é sincronizada por um sinal de clock e executa na velocidade plena do barramento processador/memória
 - Emprega um modo de **rajada** (*burst*) para eliminar o tempo gasto em configuração de endereço e pré-carga de fileira de linha e coluna após o primeiro acesso.

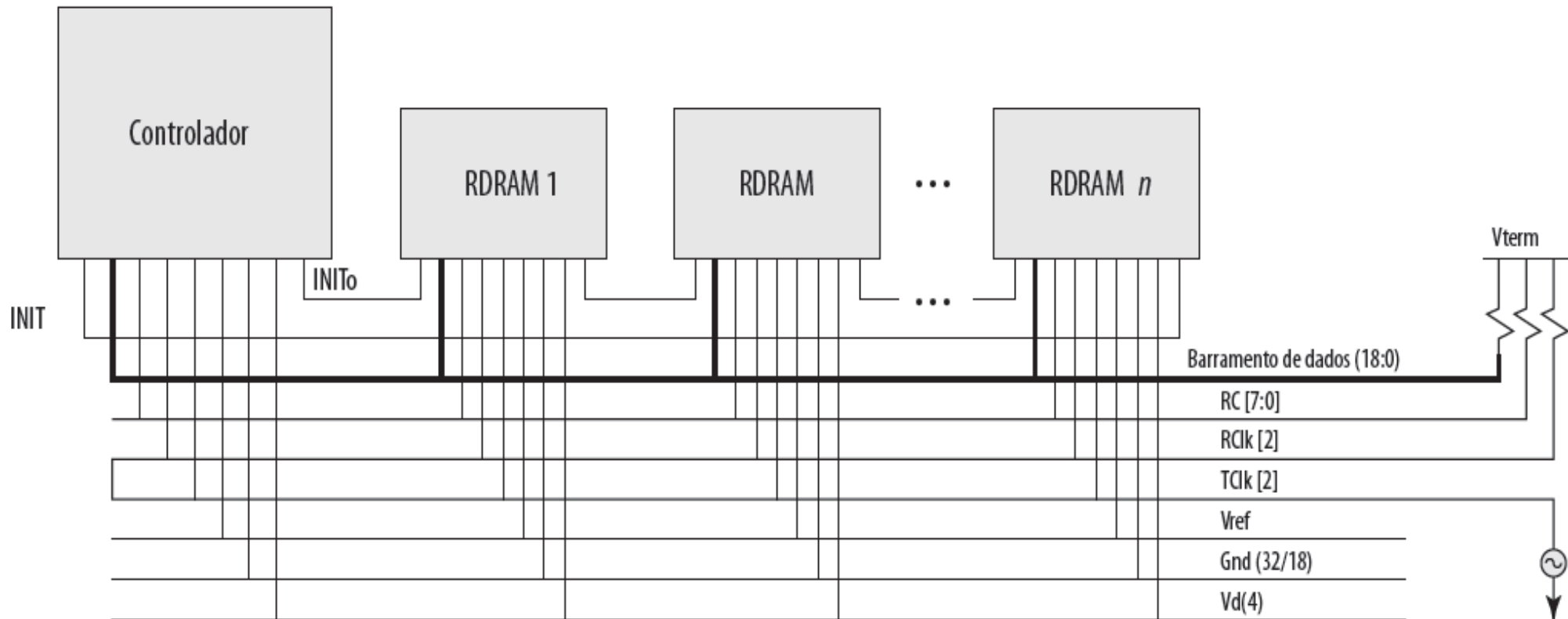
Temporização de leitura da SDRAM (tamanho de rajada = 4, latência de CAS = 2)



Organizações avançadas de DRAM

- DRAM RamBus (RDRAM)

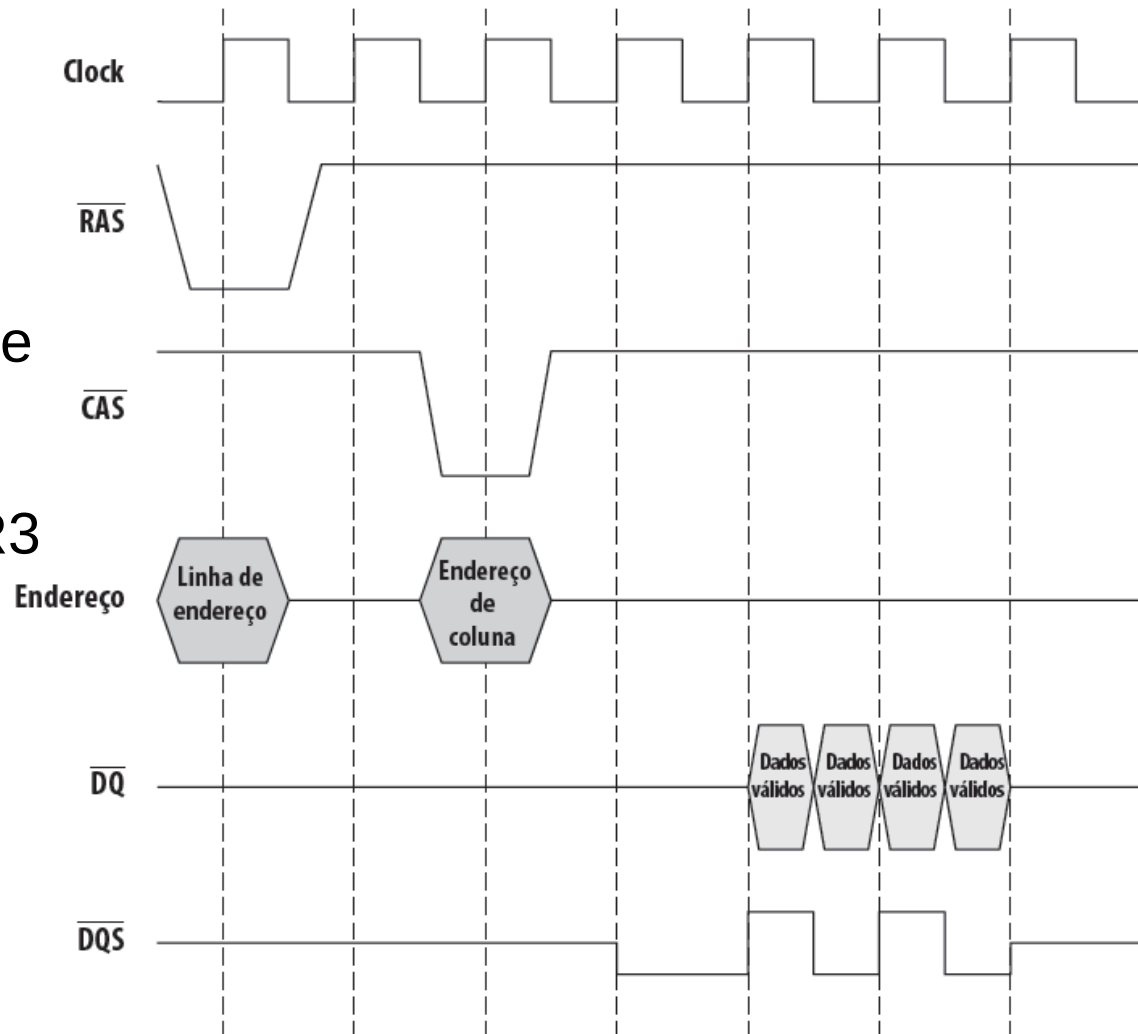
- Usado pela Intel nos chips Pentium e Itanium
- Chips são encapsulados com todos os pinos para o mesmo lado
- Barramento pode endereçar 320 chips a uma taxa de 1,6 GBps
- Barramento emprega protocolo assíncrono orientado a bloco



Organizações avançadas de DRAM

- Double Data Rate SDRAM (DDR SDRAM)

- Pode enviar dados 2 vezes por ciclo de clock (transições de subida e de descida)
- Chips usados em servidores e desktops
- Gerações: DDR, DDR2, DDR3



RAS = seleção de endereço de linha
CAS = seleção de endereço de coluna
DQ = dados (entrada e saída)
DQS = seleção DQ

Leituras recomendadas

- Apêndice E do livro do Stallings: memória intercalada
- Apêndice K do livro do Stallings sobre tecnologia DDR
- Apêndice D do livro do Stallings sobre *victim cache*
- => material disponível na Wiki da disciplina