

Arquitetura de Sistemas Embarcados

Edna Barros (ensb@cin.ufpe.br)



Centro de Informática – UFPE

Capítulo 5: Memória



Roteiro

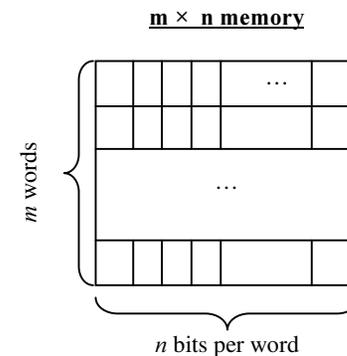
- Habilidade de escrita em memória e permanência da informação
- Tipos comuns de memórias
- Compondo memórias
- Hierarquia de Memória e Memória Cache
- RAM Avançadas

Introdução

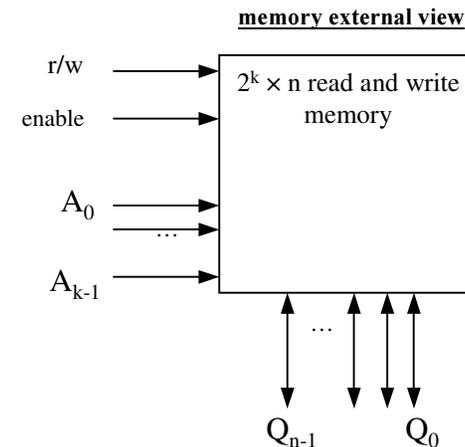
- Aspectos funcionais de sistemas embarcados
 - Processamento
 - Processadores
 - Transformação nos dados
 - Armazenamento
 - Memória
 - Retenção dos dados
 - Comunicação
 - Barramentos
 - Transferência de dados

Memória: Conceitos Básicos

- Armazena grande número de bits
 - $m \times n$: m palavras de n bits cada
 - $k = \log_2(m)$ sinais de endereço
 - ou $m = 2^k$ palavras
 - ex., Memória 4,096 x 8 :
 - 32,768 bits
 - 12 sinais de endereços
 - 8 sinais de dados (input/output)



- Acessos a Memória
 - r/w: seleciona leitura ou escrita
 - enable: sinal que permite leitura ou escrita
 - multiporta: acessos múltiplos a diferentes endereços simultaneamente



Habilidade de Escrita/Permanência da Informação

- Diferenças tradicionais entre ROM/RAM:

- ROM

- Somente leitura, bits
 - Permanencia da informação na falta de potência

- RAM

- Leitura e escrita,
 - Perda de informação na falta de potência

- Evolução tecnológica

- Tecnologia de escrita em ROM's

- e.g., EEPROM

- Tecnologia permite permanência de informação em RAMs

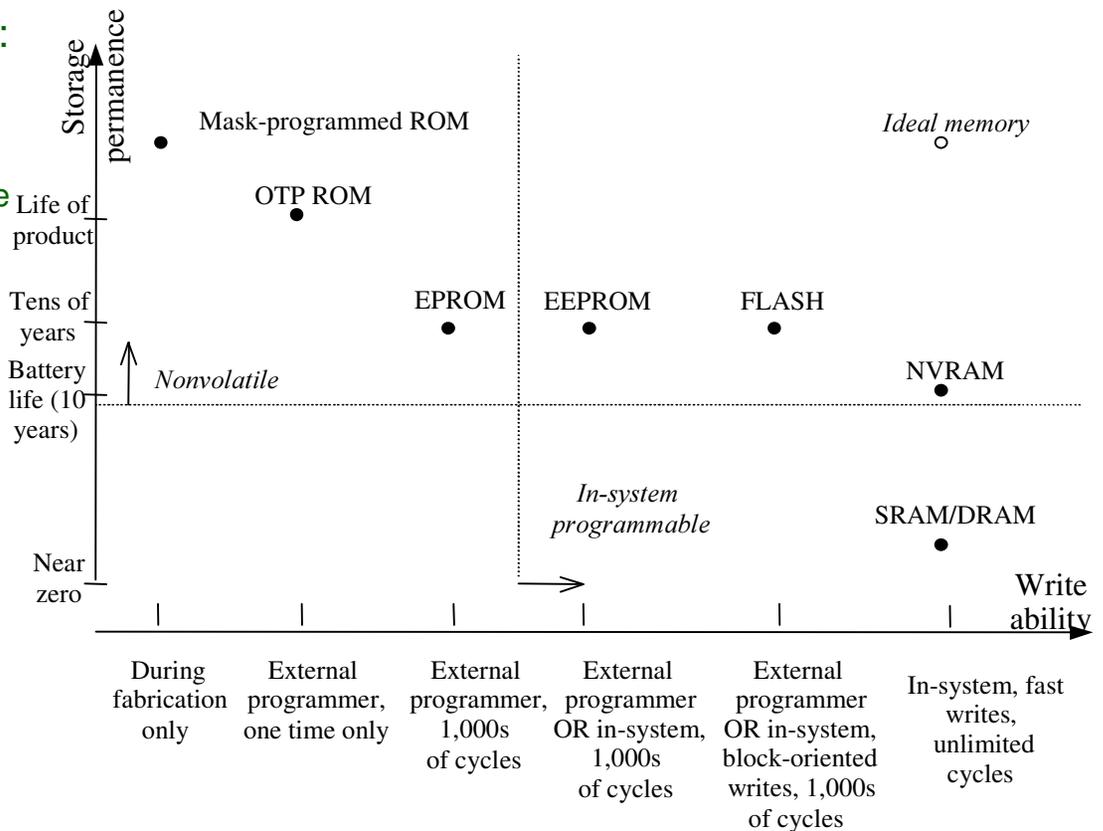
- e.g., NVRAM

- Habilidade de Escrita

- Maneira e velocidade que a memória pode ser escrita

- Permanência da Informação

- Habilidade da memória manter a informação após bits terem sido escritos



Write ability and storage permanence of memories, showing relative degrees along each axis (not to scale).

Habilidade de Escrita

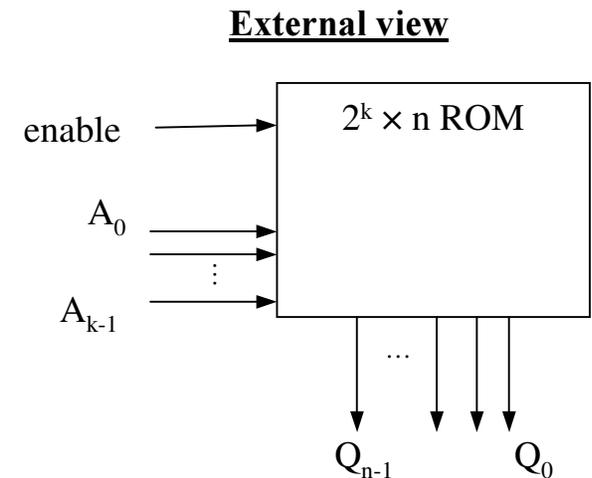
- Limites da habilidade de escrita
 - High end
 - Processador escreve na memória de forma simples e rápida
 - ex. RAM
 - Middle range
 - processador escreve na memória mas de forma lenta
 - ex. FLASH, EEPROM
 - Lower range
 - Equipamento especial “gravador” deve ser usado durante escrita
 - ex. EPROM, OTP ROM
 - Low end
 - Bits são armazenados durante gravação
 - Ex. ROM programável por máscaras
- Memória interna (In-system programmable memory)
 - Pode ser escrita pelo processador
 - Habilidade de escrita entre high end e middle end

Permanência de Armazenamento

- Limites da Permanência de Armazenamento
 - High end
 - Informação nunca é perdida
 - Ex. ROM
 - Middle range
 - Armazena informação durante dias, meses ou anos após desligamento da fonte de energia
 - Ex. NVRAM
 - Lower range
 - Armazena informação enquanto energia permanece ligada
 - Ex. SRAM
 - Low end
 - Inicia perda de informação logo após a escrita
 - Ex. DRAM
- Memória não volátil
 - Armazena informação sem fornecimento de energia

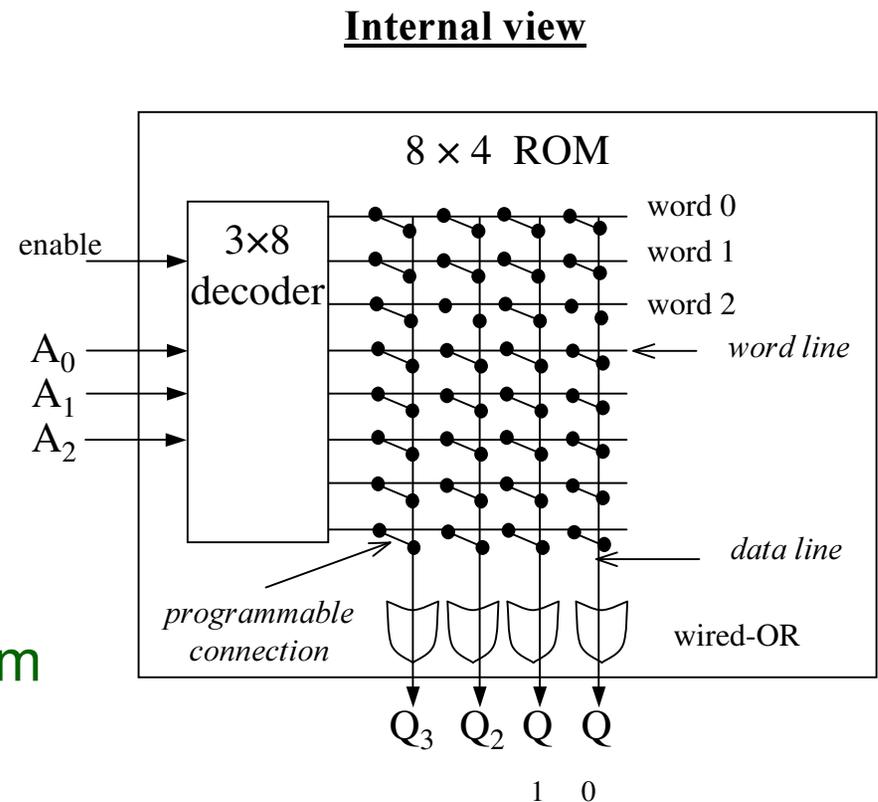
ROM: Memória “Read-Only”

- Memória não-volátil
- Pode ser lida porém não pode ser escrita pelo processador
- Programada antes de ser inserida no sistema embutido
- Usos:
 - Armazena software para processador de propósito geral
 - Instruções do programa podem ser palavras da ROM
 - Armazena dados a serem usados pelo sistema
 - Implementa Circuito combinacional



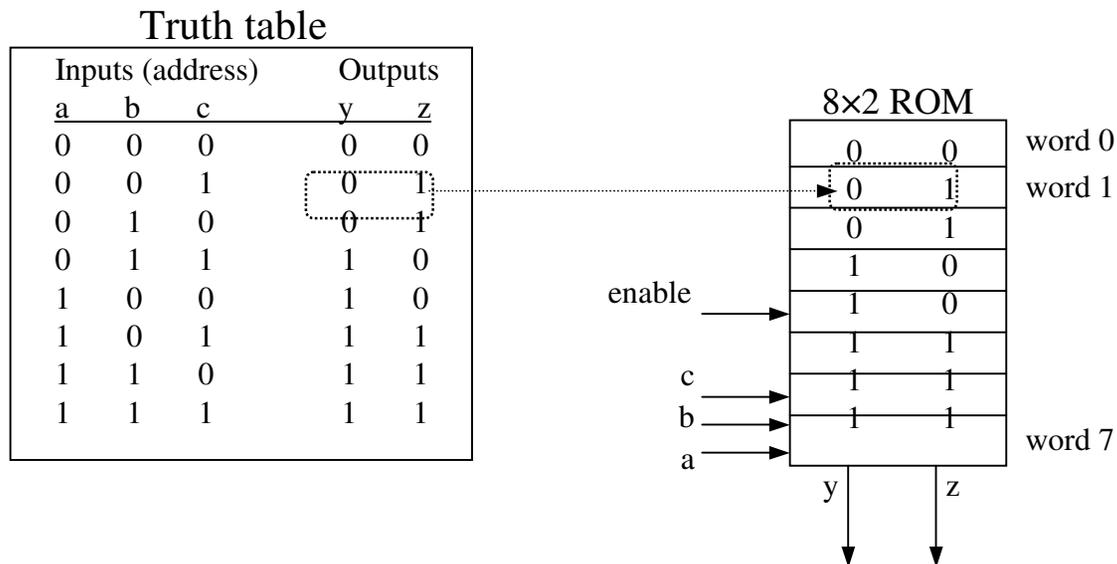
Example: 8 x 4 ROM

- Linhas horizontais = palavras
- Linhas Verticais = dados
- Decodificador seta 2a. Linha se endereço = 010
- Linhas de Dados Q₃ e Q₁ são setadas (conexão programada com palavra na 2a. Linha)
- Palavra 2 não está conectada com linhas de dados Q₂ e Q₀
- Saída é 1010



Implementando a Função Combinacional

- Qualquer circuito combinacional de n funções das mesmas k variáveis pode ser feito com uma ROM $2^k \times n$



ROM Programável por Máscara

- Conexões programadas durante fabricação
- Pouquíssima habilidade de escrita
 - Somente uma vez
- Permanência do Armazenamento
 - bits nunca mudam (exceto devido a falhas)
- Usada tipicamente no projeto final de sistemas com alto volume de produção

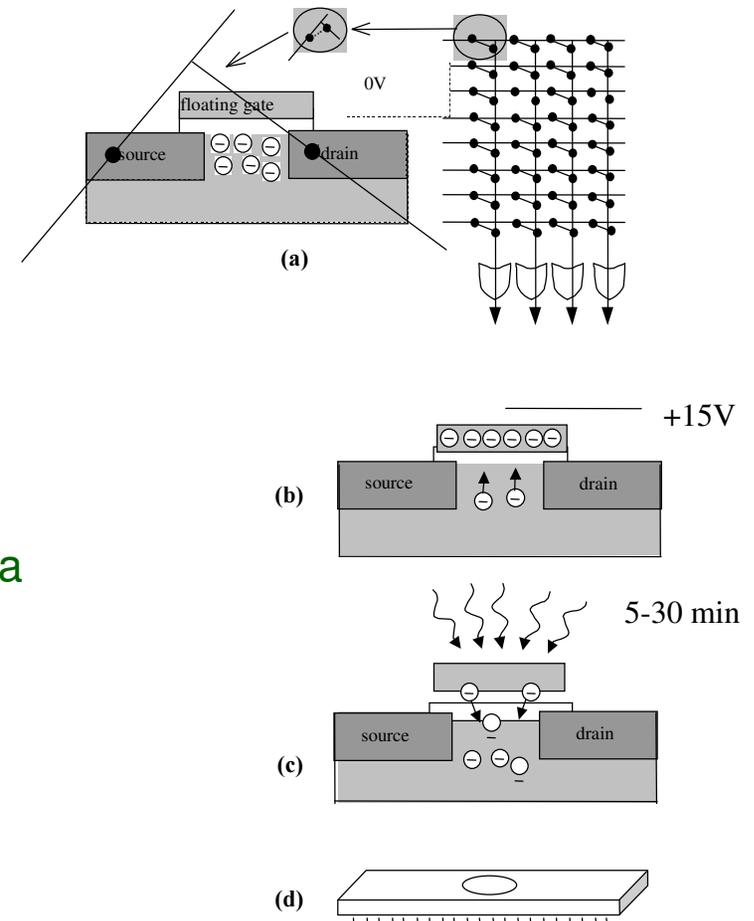
OTP ROM: One-time programmable ROM

- Conexões programadas após manufatura pelo usuário
 - Usuário fornece arquivo de conteúdo da ROM
 - Uso de gravador
 - Cada conexão programável é um fusível
 - Gravador queima fusíveis quando a conexão não existe
- Pouca habilidade de escrita
 - Necessita de gravador
- Alta permanência de armazenamento
- Comumente usada em produtos finais
 - Barata, difícil de ser modificada involuntariamente

EPRM: Erasable programmable ROM

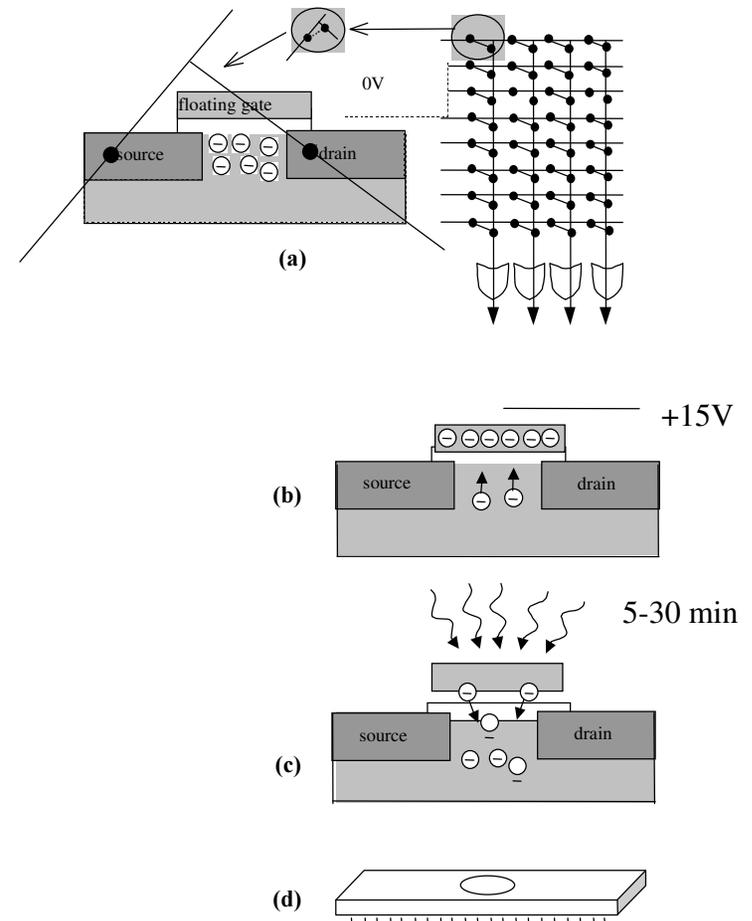
- **Componente Programável é um transistor MOS**

- Transistor tem “floating” gate
- **(a)** Carga Negativa forma canal entre source e drain armazenando valor lógico 1
- **(b)** Grande tensão positiva causa cargas negativas se moverem para fora do canal armazenando o valor lógico 0
- **(c)** (Apagamento) Aplicando raios ultra-violeta na superfície causa o retorno das cargas negativas restaurando o nível lógico 1
- **(d)** Uma EPROM possui janela de quartzo para permitir a aplicação da luz



EPRM: Erasable programmable ROM

- **Melhor habilidade de escrita**
 - Pode ser apagada e reprogramada milhares de vezes
- **Permanencia de Armazenamento Reduzida**
 - Informação permanece por cerca de 10 anos mas é sensível a radiação ou ruído elétrico
- **Usada durante desenvolvimento**



EEPROM: Electrically erasable programmable ROM

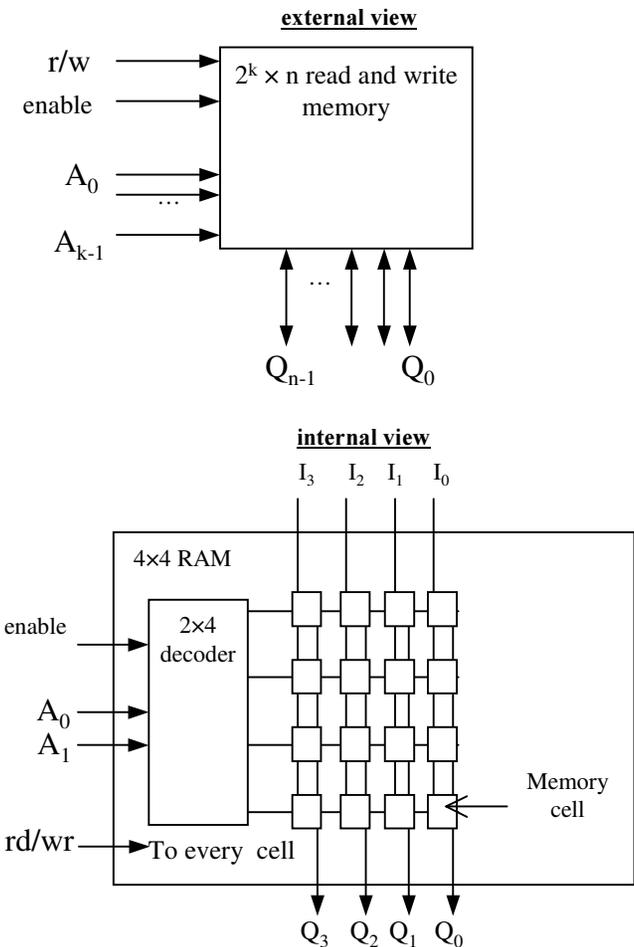
- Programada e apagada eletronicamente
 - Tipicamente usada com tensão maior que normal
 - Pode-se programar e apagar palavras individuais
- Melhor habilidade de escrita
 - Pode ser programada internamente a partir de circuito que forneça tensão maior que normal
 - Escrita bastante lenta (devido apagamento e escrita)
 - Pino “busy” pin indica ao processador que a memória ainda escreve.
 - Pode ser apagada e programada várias vezes
- Permanência similar a EPROM (cerca de 10 anos)
- Mais conveniente que EPROMs porém mais cara

Memória Flash

- Melhoramento da EEPROM
 - Mesmo principio
 - Mesma habilidade de escrita e permanência do armazenamento
- Apagamento mais rápido
 - Grandes blocos são apagados de uma vez (em vez de uma palavra)
 - Blocos possuem milhares de bytes
- Escrita de palavras simples pode ser bastante lento
 - O bloco completo deve ser lido, atualizado e o bloco completo de ser escrito antes
- Usado em sistemas embutidos para armazenar grande quantidade de informações
 - Ex. Camara digital, TV set-top boxes, telefones celulares

RAM: “Random-access” memory

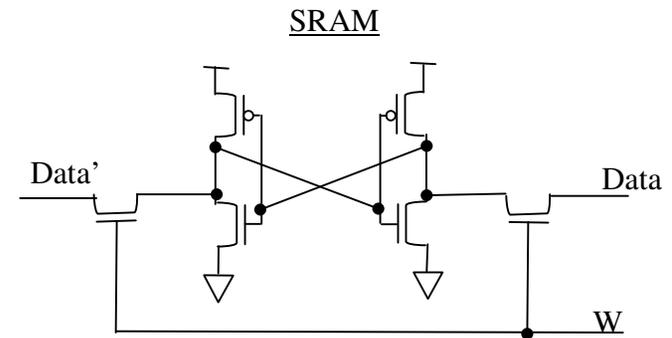
- **Memória tipicamente volátil**
- **Leitura e Escrita facilitada durante s execução do sistema**
- **Estrutura Interna mais complexa que a ROM**
 - Uma palavra consiste de várias células, cada uma armazenando 1 bit
 - Cada linha de entrada e saída conectam com as linhas de dados de entrada e saída
 - rd/wr conectado a cada célula
 - Quando uma linha é liberada pelo decodificador, cada célula tem uma lógica que armazena informação no caso de escrita ou disponibiliza informação no caso de leitura



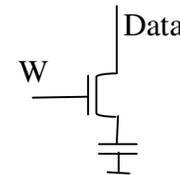
Tipos Básicos de Memória RAM

- SRAM: Static RAM
 - Célula de Memória usa flip-flop para armazenar bits
 - Requer 6 transistores
 - Armazena informação enquanto existe potência
- DRAM: Dynamic RAM
 - Célula de Memória usa transistor MOS e capacitor para armazenar um bit
 - Mais compacta que a SRAM
 - Necessidade de “Refresh”
 - Taxa de refresh: 15.625 microsec.
 - Acesso mais lento que SRAM

memory cell internals



DRAM

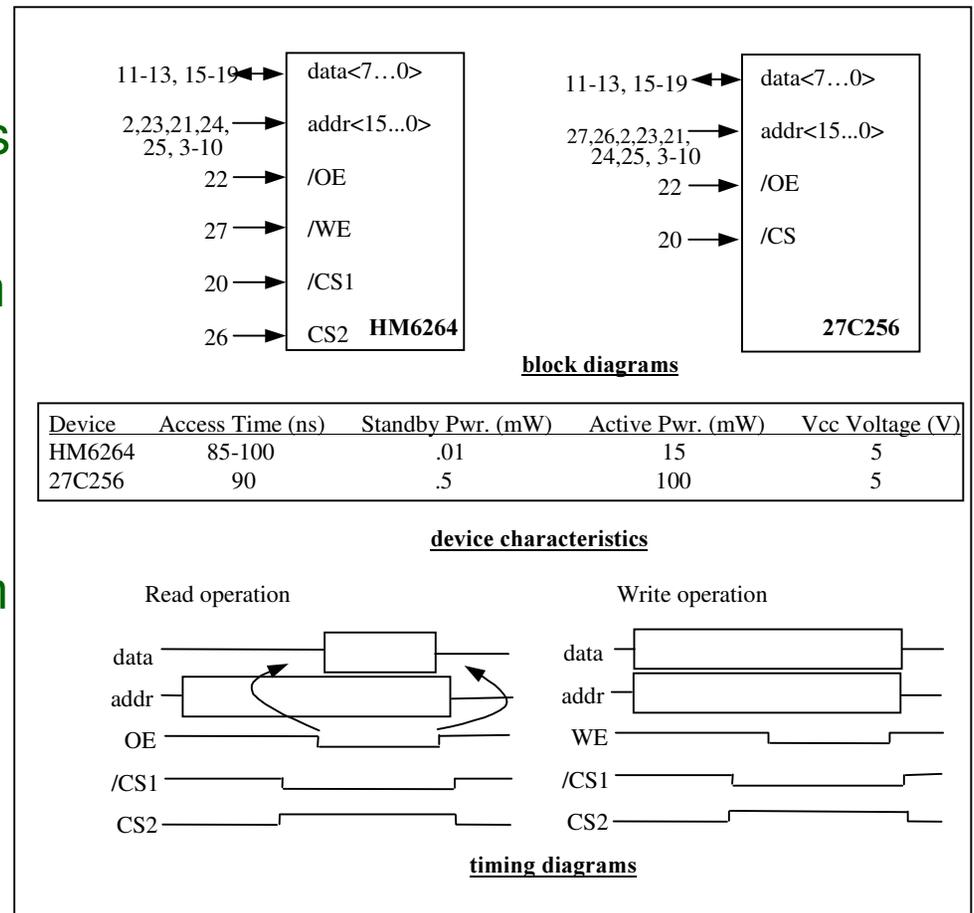


Variações de Ram

- PSRAM: Pseudo-static RAM
 - DRAM com controlador de refresh embutido
 - Alternativa popular e de baixo custo para a SRAM
- NVRAM: Nonvolatile RAM
 - Mantem informação após remoção da fonte
 - RAM com Bateria
 - SRAM com sua própria bateria conectada permanente
 - Escritas tão rápidas quanto leituras
 - Não limita número de escritas
 - SRAM com EEPROM ou flash
 - Cópia de RAM em EEPROM ou flash antes de desligamento da fonte de energia

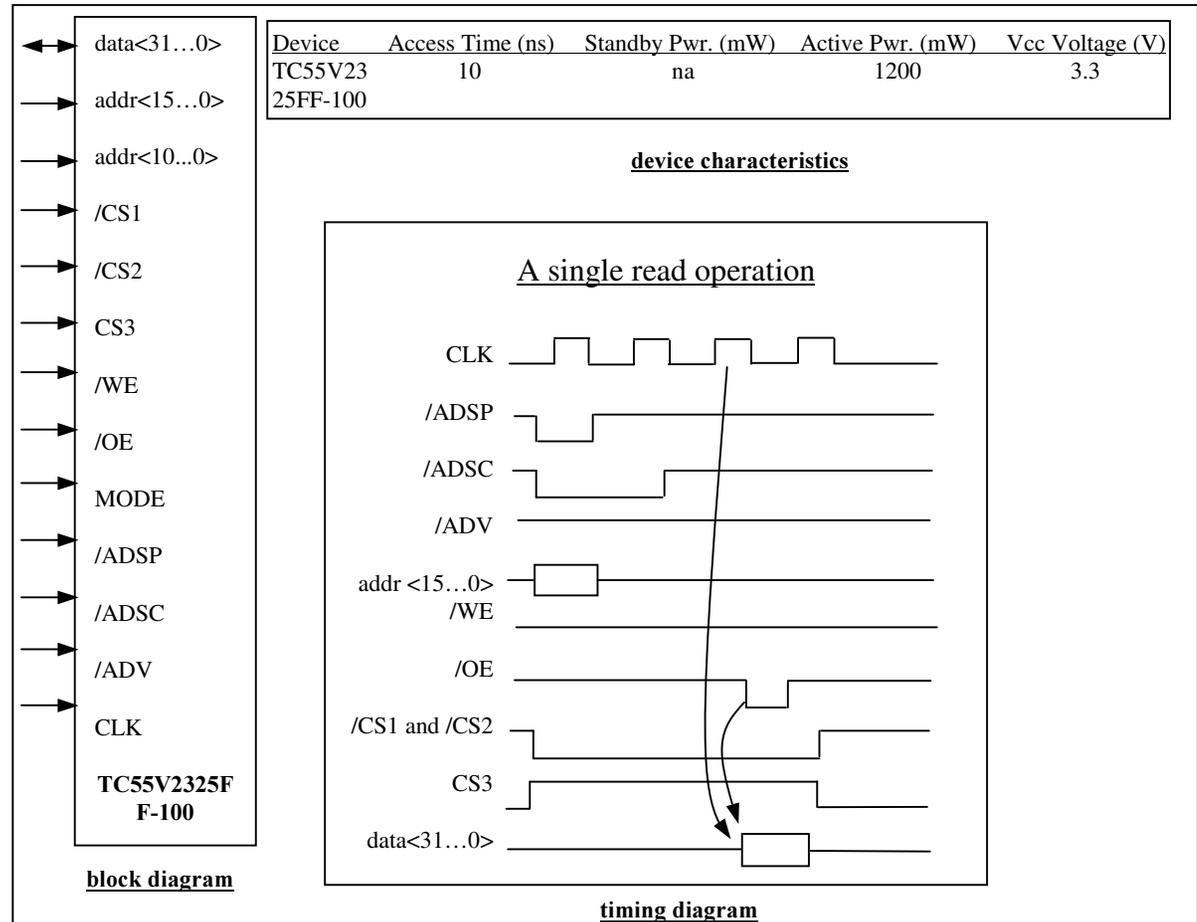
Exemple: HM6264 & 27C256 RAM/ROM

- Dispositivos de baixo custo e baixa capacidade
- Usada em microcontroladores de 8-bits
- Primeiros dois digitos indicam tipo:
 - RAM: 62
 - ROM: 27
- Digitos subsequentes indicam capacidade em kilobits



Exemplo: TC55V2325FF-100

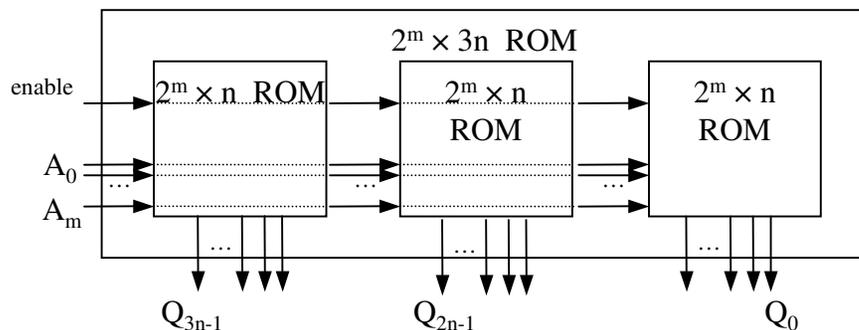
- Memória SRAM de 2-megabit, síncrona, com pipeline e burst
- Projetada para uso com processadores de 32 bits
- Capaz de leituras e escritas sequenciais rápidas



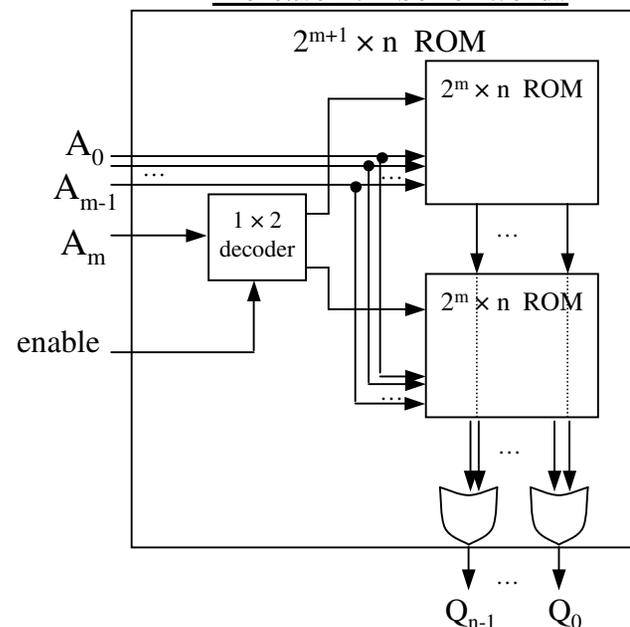
Composição de Memória

- Tamanho da memória necessário difere do tamanho das memórias disponíveis
- Quando a memória disponível é maior: ignore bits do endereço não usados
- Quando a memória disponível for menor: composição de memórias menores
 - Conexão lado a lado para aumentar palavra
 - Conexão topo com início para aumentar número de palavras

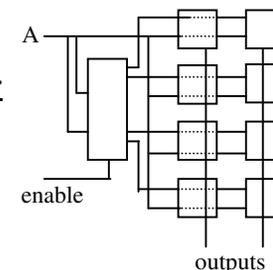
Increase width of words



Increase number of words

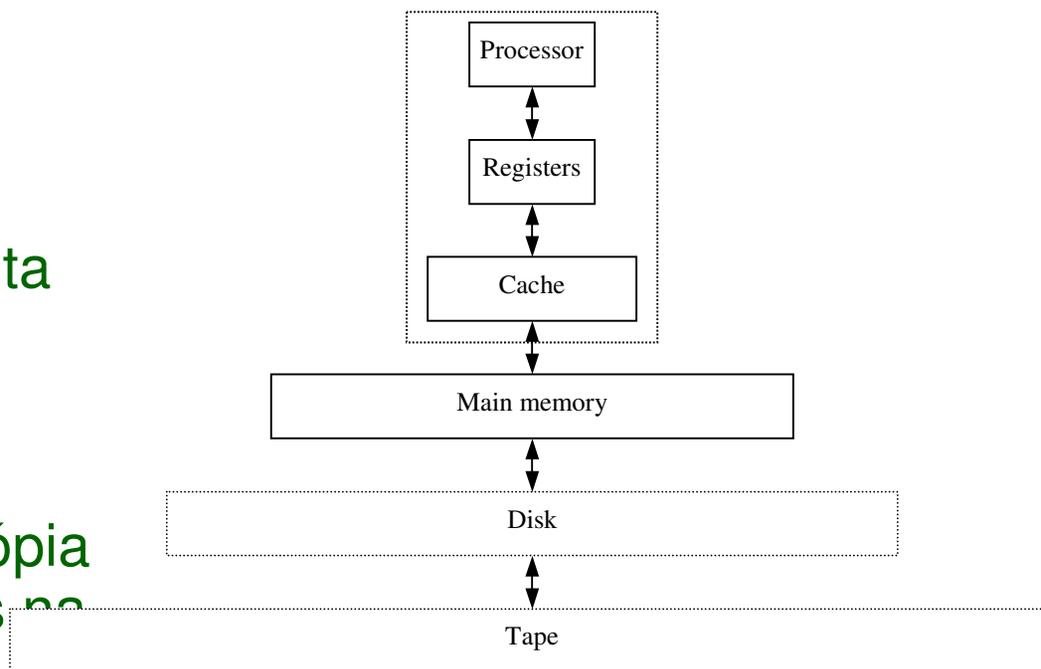


Increase number and width of words



Hierarquia de Memória

- Desejado: memória barata e rápida
- Memória principal:
 - Grande, barata e lenta
- Cache
 - Pequena, caras e rápidas armazena cópia de partes acessadas na memória principal
 - Pode haver múltiplos níveis de cache



Cache

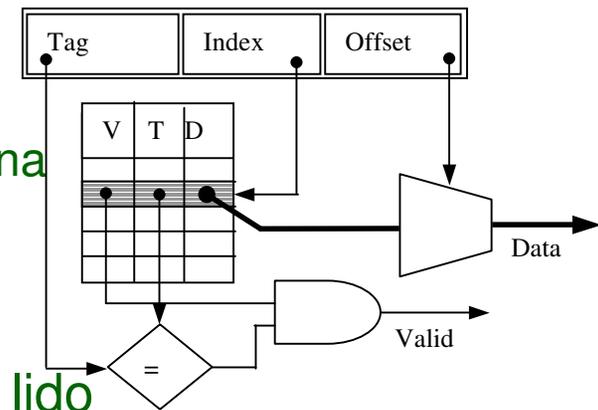
- **Usualmente feita com SRAM**
 - Rápida porém mais cara que DRAM
- **Usualmente no mesmo do processador**
 - Espaço limitado,
 - Acesso rápido (1 ciclo apenas)
- **Operação da Cache :**
 - Requisita acesso a memória principal
 - Verifique se cópia está na cache
 - cache hit
 - Cópia está na cache
 - cache miss
 - Cópia não está na cache, leitura de várias palavras de cache
- **Escolhas no Projeto da Cache**
 - Mapeamento, políticas de substituição e técnicas de escrita

Mapeamento de Cache

- Poucos endereços podem estar na cache
- Está o conteúdo de um endereço na cache?
- Mapeamento de Cache usado para atribuir endereço de memória a endereço de cache causando um hit ou um miss
- Tres técnicas básicas:
 - Mapeamento direto
 - Mapeamento Completamente Associativo
 - Associativa por conjunto
- Caches são particionadas em blocos indivisíveis ou linhas de endereços de memória adjacentes
 - Usualmente 4 ou 8 endereços por linha

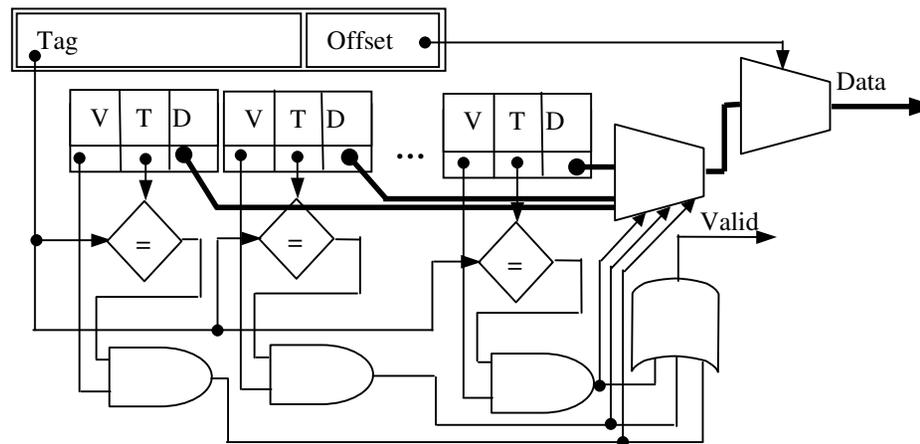
Mapeamento Direto

- Endereço é dividido em dois campos:
 - Índice
 - Endereço da cache
 - Número de bits determinado pelo tamanho da cache
 - Tag
 - É comparado com o tag armazenado na cache
- Bit Válido
 - Indica se o dado no slot é valor válido lido da memória
- Offset
 - Usado para encontrar determinada palavra na linha de cache



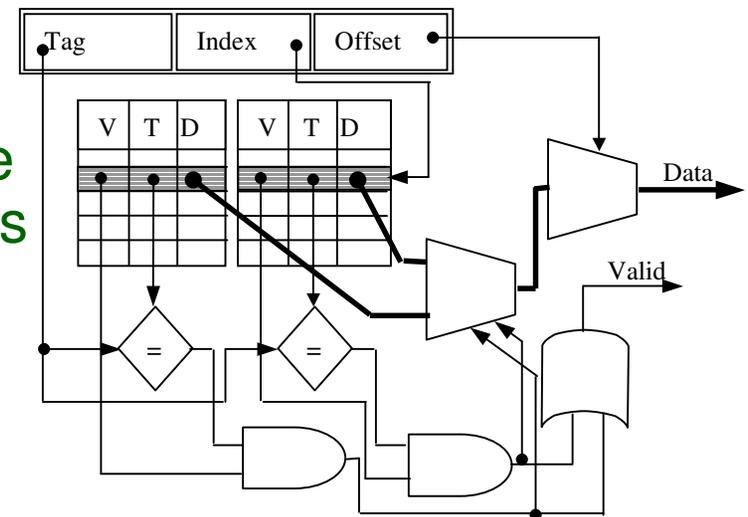
Mapeamento Completamente Associativo

- Endereço de memória completo está armazenado em cada endereço de cache
- Todos os endereços são comparados simultaneamente com o endereço desejado
- Comparação simultânea entre endereços da cache e endereço procurado
- Bit de Validade



Mapeamento Set-associativo

- Compromisso entre mapeamento direto e mapeamento completamente associativo
- Índice igual a mapeamento direto
- MAS cada endereço contem tags e conteúdo de 2 ou mais localizações de memória
- Tags de um mesmo conjunto são comparados simultaneamente
- Cache com tamanho chamada N-way set-associative
 - 2-way, 4-way, 8-ways são comuns



Política de Substituição de Cache

- Qual bloco deverá ser substituído?
 - Quando uma cache completamente associativa está cheia
 - Quando a linha de cache associativa por conjunto está cheia
- Técnicas
 - Randomica
 - LRU: least-recently used
 - Substitua bloco não acessado por mais tempo
 - FIFO: first-in-first-out
 - Empilhe bloco quando acessado
 - Substitua bloco do topo da pilha

Técnicas de Escrita em Cache

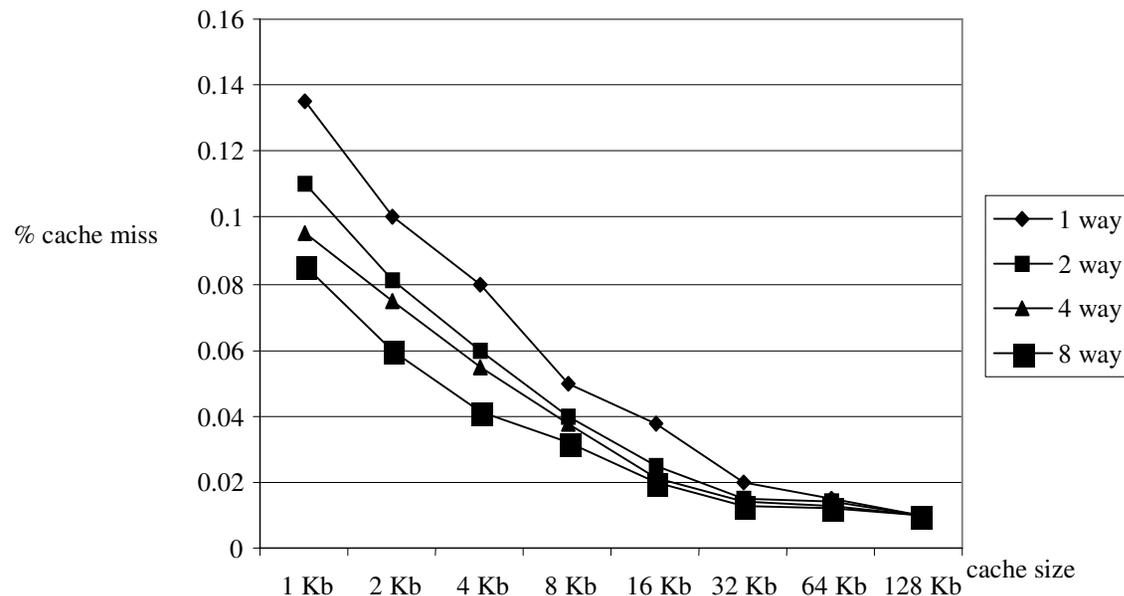
- Como manter a consistência entre endereços na cache e na memória?
- Write-through
 - Atualiza memória sempre que cache é escrita
 - Fácil de implementar
 - Processador deve esperar pela atualização da memória
 - Escritas não necessárias
- Write-back
 - Memória é atualizada quando bloco é substituído
 - Bit extra “dirty” indica escrita em cache
 - Número de acessos a memória é reduzido

Impacto da Cache no Desempenho do Sistema

- Principais parametros
 - Tamanho total da cache
 - Número total de bytes de dado da cache
 - Tags, bits de validade e outros bits de controle
 - Grau de Associatividade
 - Tamanho do Bloco
- Caches maiores permitem taxas de faltas menores mas a custo maior
 - Cache de 2 Kbyte: miss rate = 15%, custo de hit = 2 ciclos, custo de miss = 20 ciclos
 - Custo médio de acesso a memória = $(0.85 * 2) + (0.15 * 20) = 4.7$ ciclos
 - Cache de 4 Kbyte: miss rate = 6.5%, custo de hit = 3 ciclos, custo de miss: mesmo
 - Custo médio de acesso a memória = $(0.935 * 3) + (0.065 * 20) = 4.105$ ciclos
(melhora)
 - Cache de 8 Kbyte: miss rate = 5.565%, custo de hit = 4 ciclos, custo de miss: mesmo
 - Custo médio de acesso a memória = $(0.94435 * 4) + (0.05565 * 20) = 4.8904$ ciclos
(piora)

Melhorando o Desempenho da Cache

- Melhorando a taxa de hits sem aumentar tamanho da cache
 - Aumento de bloco
 - Aumento da Associatividade

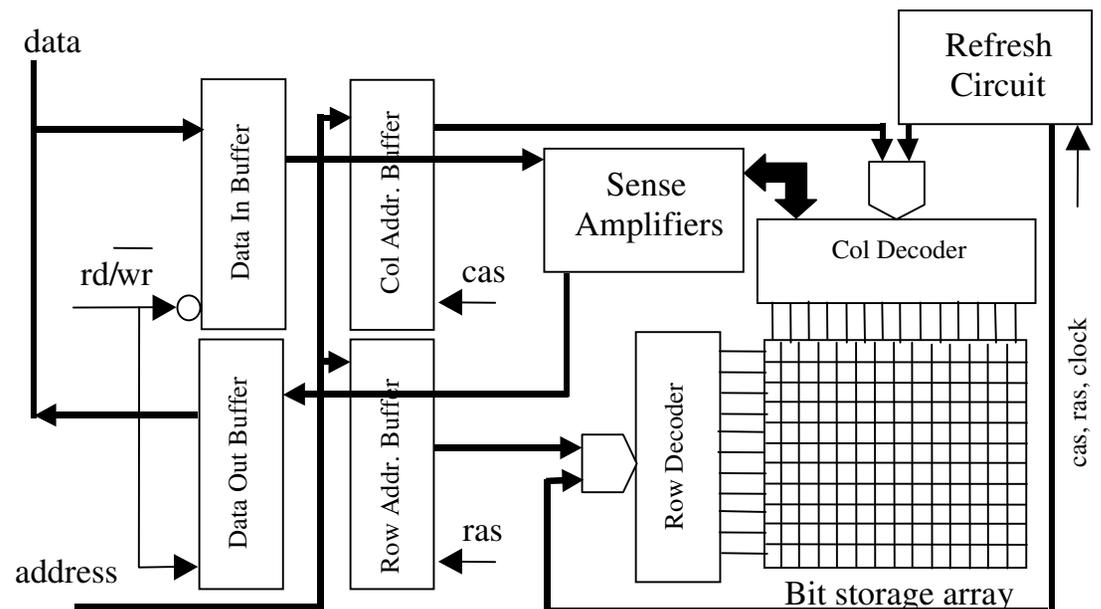


RAM Avançada

- DRAMs são usadas como memória principal em sistemas embarcados baseados em microprocessadores
 - Grande capacidade de armazenamento e baixo custo
- Variações de DRAMs
 - Necessidade de acompanhar velocidade do processador
 - FPM DRAM: fast page mode DRAM
 - EDO DRAM: extended data out DRAM
 - SDRAM/ESDRAM: synchronous and enhanced synchronous DRAM
 - RDRAM: rambus DRAM

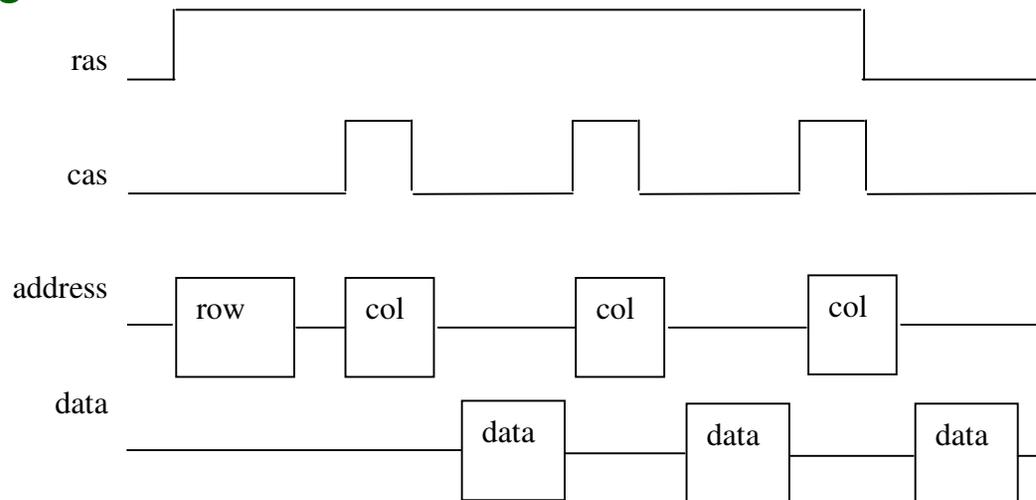
Basic DRAM

- Barramento de endereço multiplexado para acessar linha e coluna
- Endereços de linha e coluna são armazenados através da ativação dos sinais *ras* e *cas*
- Circuito de refresh pode ser interno ou externo
 - Ativar endereços consecutivos ativa processo de refresh
 - Circuito de refresh desativado durante operação de leitura ou escrita



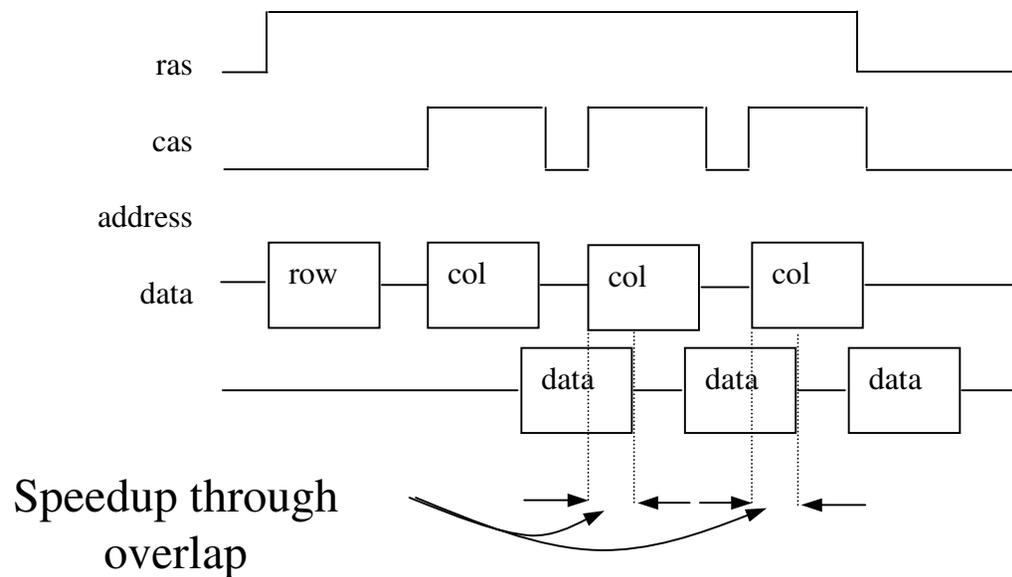
Fast Page Mode DRAM (FPM DRAM)

- Cada linha de memória é vista como página
- Página contém múltiplas palavras
- Palavras individuais endereçadas pelo endereço da coluna
- Diagrama temporal
 - Envia endereço da linha (página)
 - Leitura de 3 palavras consecutivas pelo envio de endereços das colunas



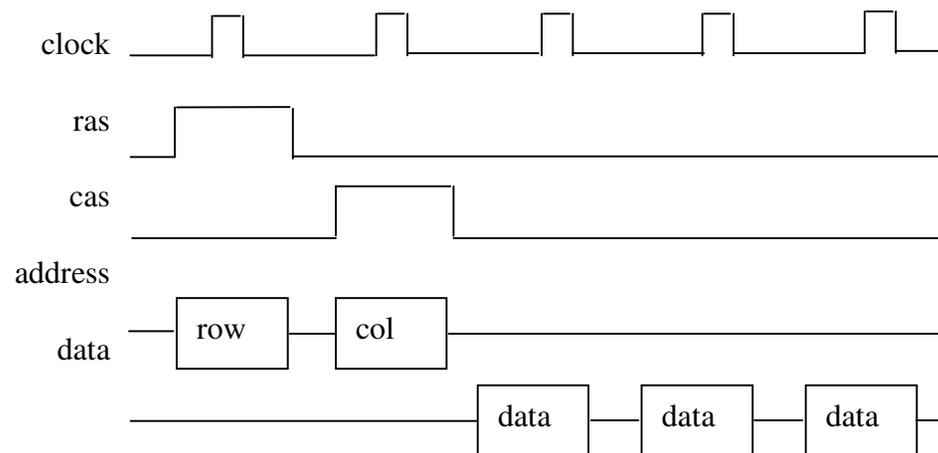
Extended data out DRAM (EDO DRAM)

- Versão melhorada da FPM DRAM
- Latch extra antes de buffer de saída
 - Permite ativação do *cas* antes de completar operação de leitura
- Reduz latência de leitura/escrita devido a ciclos adicionais



(S)ynchronous and Enhanced Synchronous (ES) DRAM

- SDRAM armazena dado a cada transição de clock
- Elimina tempo para detectar sinais *ras/cas* e *rd/wr*
- Um contador inicializado com endereço da coluna é incrementado a cada transição de clock para acessar endereços consecutivos
- ESDRAM = versão melhorada da SDRAM
 - Buffers foram adicionados para sobreposição de endereçamento de colunas
 - Clock mais rápido e menor latência de leitura/escrita



Problemas de Integração das DRAMs

- SRAM facilmente integrável no mesmo chip do processador
- DRAM integração mais difícil
 - Diferença no processo de integração entre DRAM e lógica convencional
 - Objetivos da lógica convencional:
 - Minimizar capacitância para reduzir retardo e consumo de potência
 - Objetivos dos projetistas de DRAMs:
 - Criar capacitores para reter informação armazenada

Memory Management Unit (MMU)

- Papel da MMU
 - Trata DRAM refresh, interface com barramento e arbitragem
 - Cuida do compartilhamento de memória entre vários processadores
 - Traduz endereços virtuais (do processador) para endereços físicos da DRAM
- CPUs modernas possuem MMU interno