

# Sistemas de Memória

## Conceitos Básicos

João Canas Ferreira

Outubro de 2015



Assuntos

## Tópicos

### ① Memórias

- Aspetos gerais
- Memórias Estáticas
- Memórias Dinâmicas

### ② Decodificação de endereços

- Organização geral
- Decodificação total
- Decodificação parcial

## 1 Memórias

Aspetos gerais

Memórias Estáticas

Memórias Dinâmicas

## 2 Descodificação de endereços

Organização geral

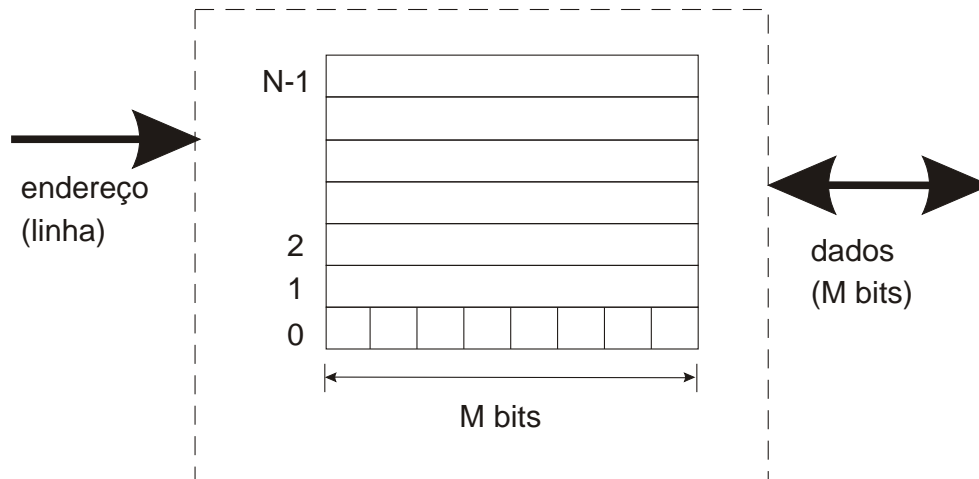
Descodificação total

Descodificação parcial

## Taxonomia

- Registos e bancos de registos permitem guardar pequenas quantidades de dados. Para maiores quantidades, usam-se **memórias de acesso direto**.
- RAM = *random access memory* (memória de acesso direto): permitem leitura e escrita em qualquer posição.
- ROM = *read-only memory*: permitem apenas leitura.
- A maior parte das memórias RAM perde os dados quando é desligada a alimentação (memória volátil). Exceções:
  - (E)EPROM: (Electrically) erasable programmable ROM
  - memórias FLASH.
- Dois tipos de memórias RAM voláteis:
  - SRAM: memória estática (cada célula de memória é um “flip-flop”);
  - DRAM: memória dinâmica (cada célula deve ser atualizada periodicamente).

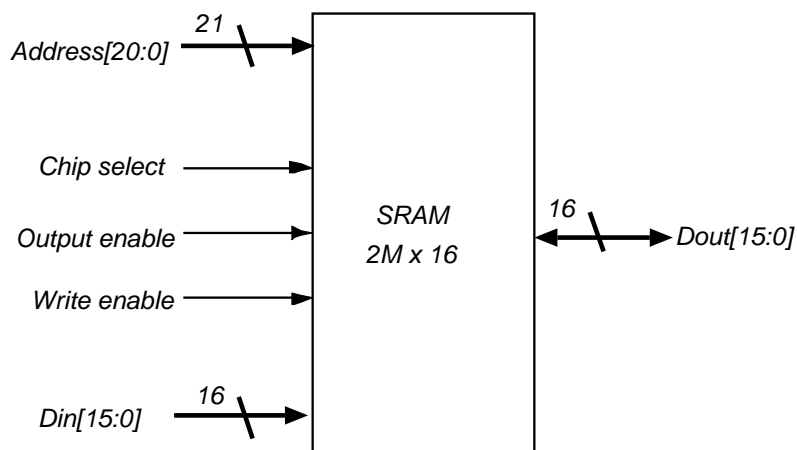
## Circuitos de memória: organização conceptual



- Para P linhas de endereço:  $N = 2^P$
- Abreviaturas:  $2^{10} = 1024 = 1 \text{ K}$ ,  $2^{20} = 1048576 = 1 \text{ M}$
- O porto de dados é bidirecional: é preciso especificar o tipo de acesso (leitura ou escrita).
- M é a *largura* da memória.

## Memórias estáticas

As memórias estáticas aproximam-se do modelo conceptual de funcionamento.



➡ Para aceder à memória:

- ativar o circuito: *chip select* (CS) ativo
- especificar o tipo de acesso:  
ativar *output enable* (leitura) **OU** *write enable* (escrita).

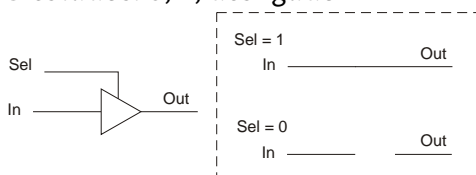
## Memórias estáticas: acessos

- Tempo de acesso para leitura: intervalo entre o instante em que *output enable* e endereço estão corretos e o aparecimento de dados na saída.
- Valores típicos para memórias estáticas:
  - rápidas: 2–4 ns
  - típicas: 8–20 ns (cerca de 32 milhões de bits)
  - de baixo consumo: 5–10 vezes mais lentas
- Durante esse tempo, um processador que execute uma instrução por ciclo e use um relógio de 2 GHz, executa:
  - 4–8 instruções
  - 16–40 instruções
- Tempo de acesso para escrita: endereços e dados devem estar estáveis antes e depois do flanco. O sinal de *write enable* é sensível ao nível (não ao flanco) e deve ter uma duração mínima para que a escrita se realize.
- O tempo de escrita é superior ao tempo de leitura.

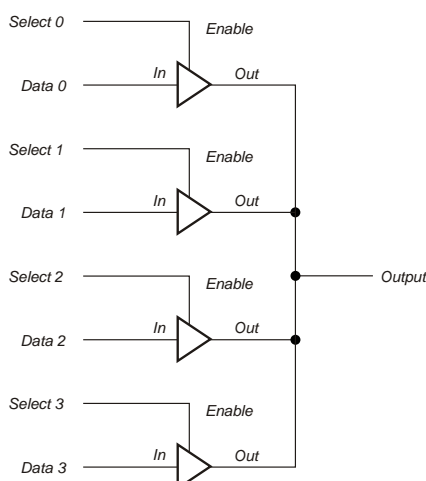
## Memórias estáticas: circuito de saída

### Buffer tristate

3 estados: 0, 1, desligado



Circuito de saída:



► Ao contrário de um banco de registos, o circuito de saída não pode ser baseado num multiplexador: uma SRAM 64K x 1 precisaria de ter um multiplexador 65536-para1.

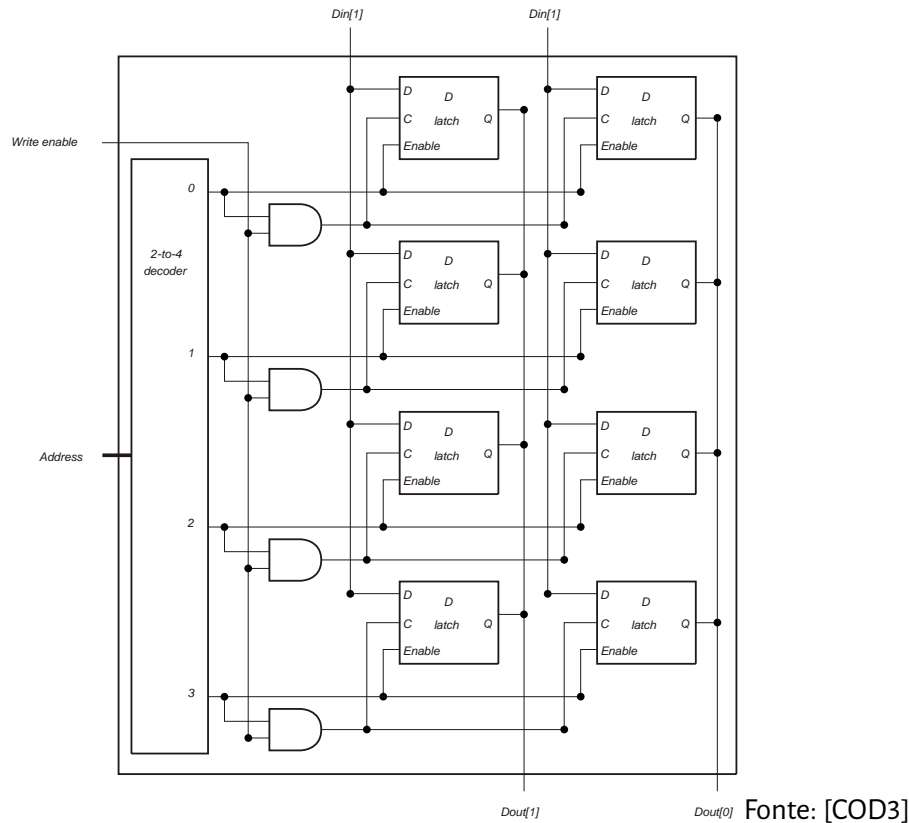
► Solução: utilizar *buffer tristate*, cuja saída pode ter 3 estados (0, 1 ou alta-impedância).

► No estado de alta-impedância, a saída do circuito está *desligada*.

► O estado da saída é determinado por uma entrada de controlo: *Sel*.

► Todas as saídas são ligadas em paralelo. Não pode existir mais que uma saída ativa (i.e., não em alta-impedância) em cada instante.

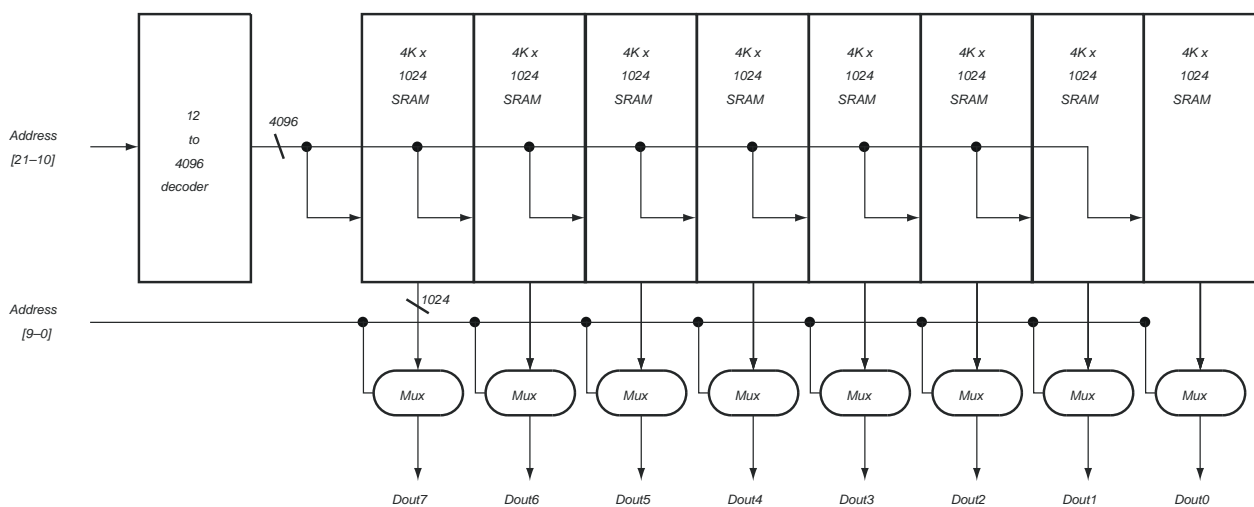
## Estrutura básica de uma memória estática



SRAM 4x2

## Memórias estáticas organizadas por bancos

Para limitar o tamanho do decodificador de endereços:



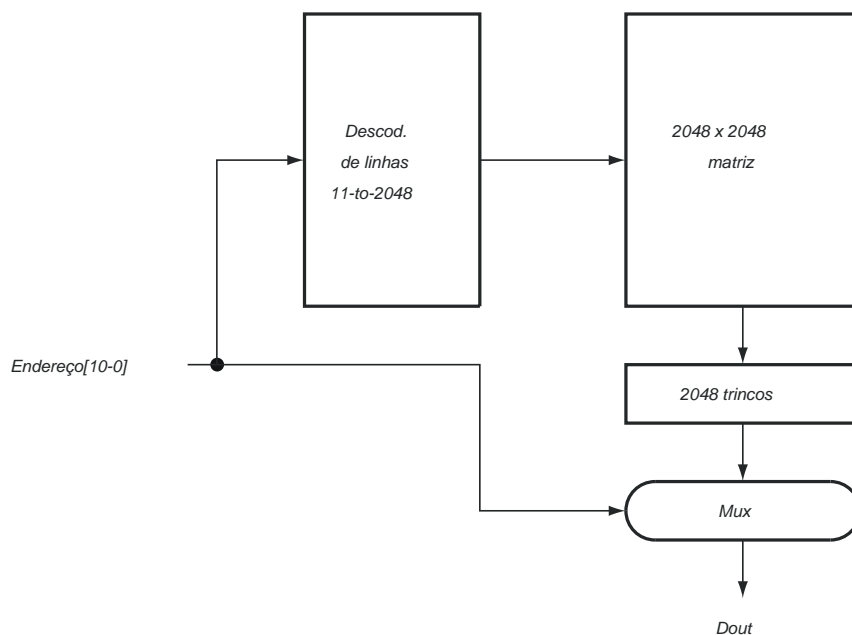
Fonte: [COD3]

- Organização típica de uma memória 4Mx8 como uma coleção de blocos de memória 4Kx1024.
- Os blocos MUX são realizados por *buffers* de três estados.

## Memória dinâmica (DRAM)

- Valor guardado como *carga num condensador*.
- O acesso é feito através de um transístor a operar como interruptor.
- Consequência: maior densidade ( $\text{bit}/\text{mm}^2$ ), logo circuitos de maior capacidade e menor custo.
- Comparação: SRAM requer 4 a 6 transístores por bit armazenado.
- Acesso a DRAM é feito em duas etapas:
  - ① seleção de coluna (usando uma parte do endereço);
  - ② seleção de linha (usando os restantes bits do endereço).
- DRAM é mais lenta que SRAM. Valor típico: 2 Gibit ( $512\text{M} \times 4$ ), tempo de acesso 55 ns.
- Condensador vai perdendo a carga e deve ser periodicamente “refrescado”, fazendo uma leitura seguida de escrita (circuito dinâmico). Refrescamento “consome” 1% a 2% dos acessos.

## Esquema geral do acesso a uma memória dinâmica

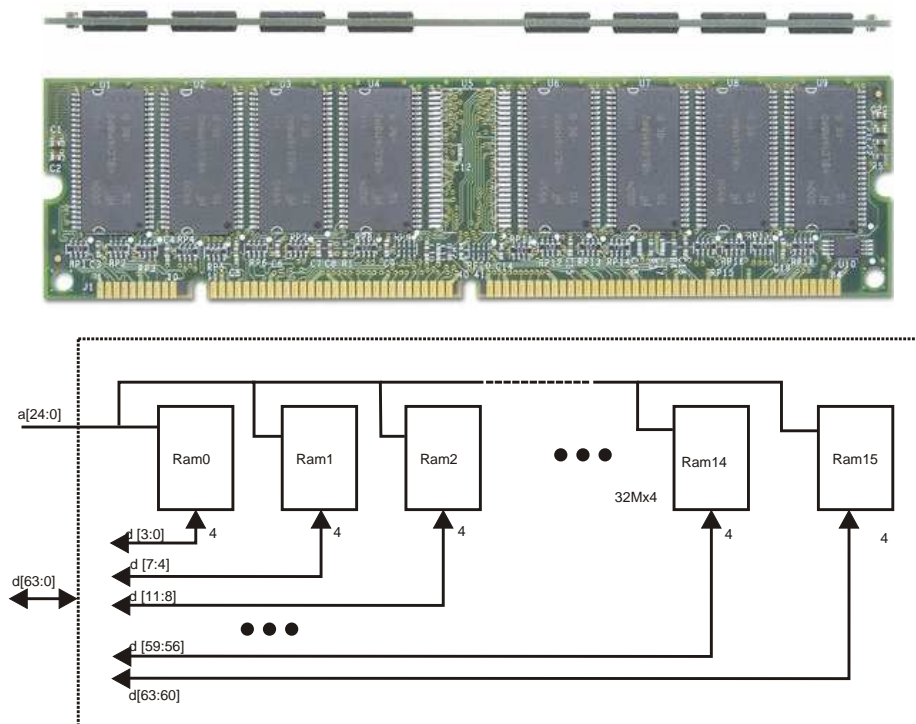


- Endereço: 11+11 bits.
- DRAM  $4\text{M} \times 1$ : 11 bits selecionam a linha, que é “copiada” para 2048 trincos.
- Multiplexador seleciona uma de 2048 entradas.

## Módulos de memória: DIMM

Chips individuais podem ser agrupados em módulos.

Ex: módulo 32Mx64 (256 MB) pode usar 16 componentes 32Mx4.



### 1 Memórias

Aspetos gerais

Memórias Estáticas

Memórias Dinâmicas

### 2 Descodificação de endereços

Organização geral

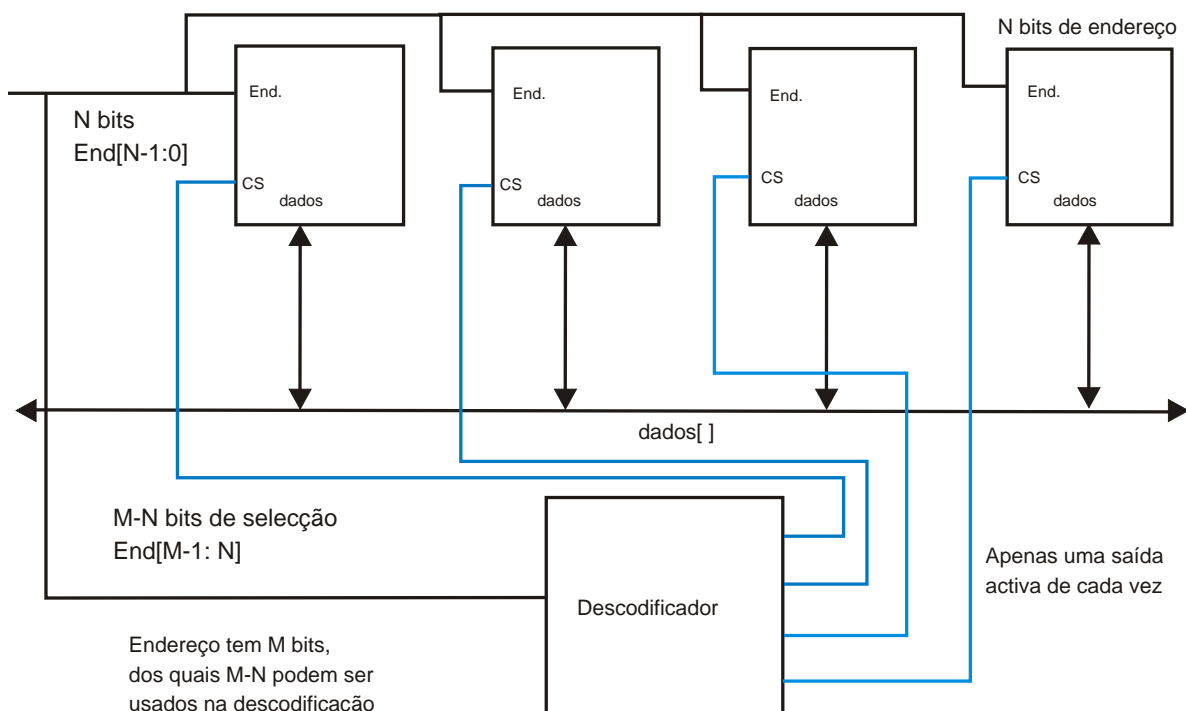
Descodificação total

Descodificação parcial

## Organização da memória de um computador

- A memória física de um computador é geralmente composta por vários módulos (circuitos integrados, DIMM, etc.) por forma a ser possível obter maiores capacidades de armazenamento.
- Para além dos módulos de memória é necessário ter um circuito de descodificação de endereços que seleciona quais os módulos ativos durante um dado acesso (com base no endereço apresentado pelo CPU).
- Organização típica: os bits menos significativos são ligados diretamente aos módulos individuais, enquanto os bits mais significativos são usados para fazer a seleção dos módulos.
- Linhas de dados podem ser “partilhadas” por mais que um módulo (usando *buffers tristate*).
- Alguns módulos usam internamente esta abordagem.

## Organização da memória: diagrama de blocos



Para memórias DRAM a descodificação de endereços é mais complicada. Apenas abordaremos o caso das memórias SRAM.



## Regras para descodificação de endereços

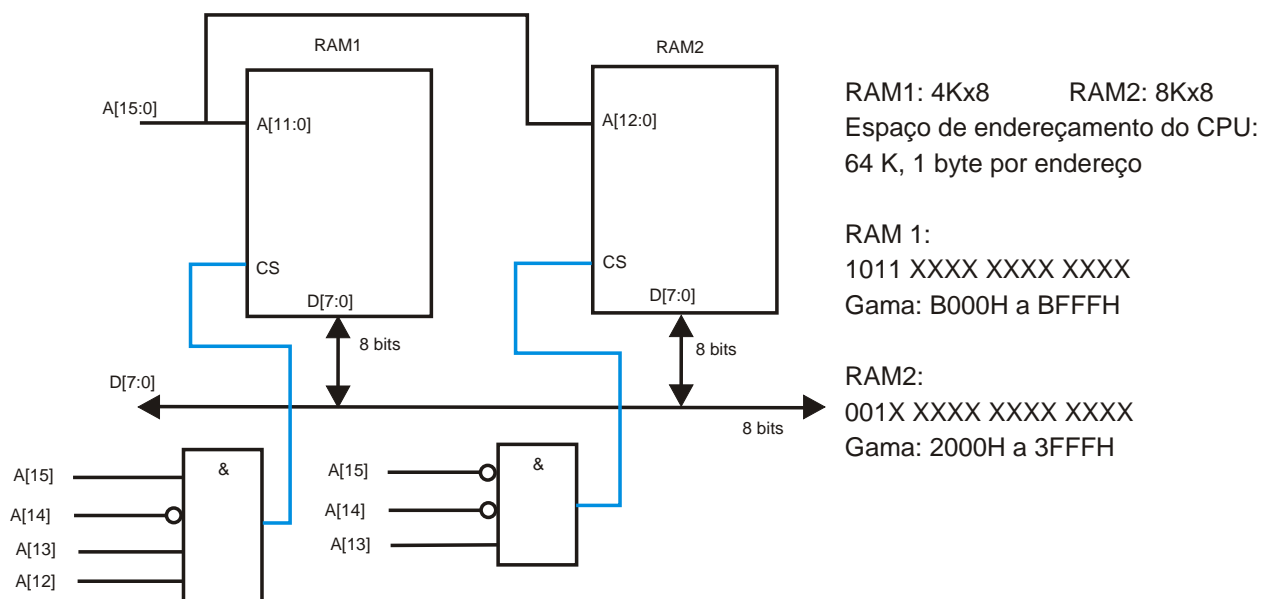
- Para que esta organização funcione bem, a descodificação de endereços deve garantir que:

Para o conjunto de todos os módulos que partilham uma mesma linha de dados: apenas um (ou nenhum) deve ser ativado durante um acesso.

Se esta condição não for respeitada, os componentes podem ser definitivamente danificados.

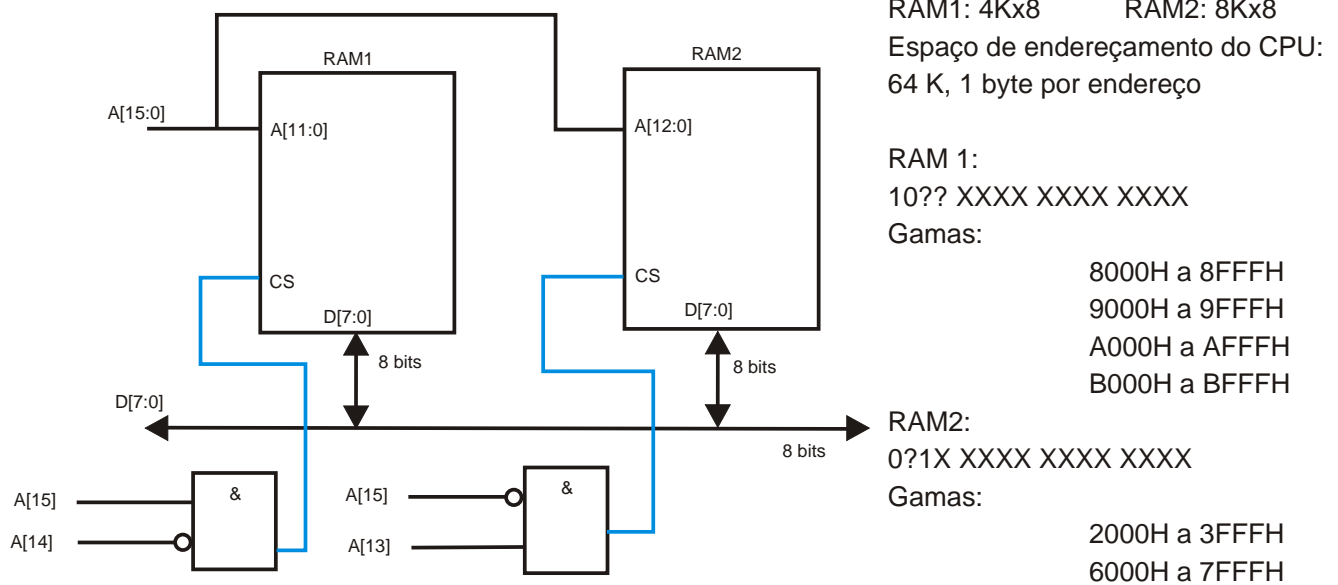
- O mapeamento de endereços para componentes pode ser classificado de acordo com o número de endereços que é mapeado na mesma posição física:
  - **total:** 1 endereço  $\rightarrow$  1 posição
  - **parcial:** N endereços  $\rightarrow$  1 posição
- Na descodificação total, todos os bits do endereço são “usados”: ligados diretamente aos componentes ou utilizados na seleção dos componentes.

## Descodificação total: exemplo



- Endereço B712H (46866)  $\rightarrow$  RAM1
- Endereço C1E0H (49632)  $\rightarrow$  nenhum circuito

## Descodificação parcial: exemplo



- O byte 10 de RAM1 pode ser acessado através de que endereços?
- 800AH, 900AH, A00AH e B00AH

## Referências

- COD4** D. A. Patterson & J. L. Hennessey, Computer Organization and Design, 4 ed.
- COD3** D. A. Patterson & J. L. Hennessey, Computer Organization and Design, 3 ed.

Os tópicos tratados nesta apresentação são descritos na seguinte secção de [COD4]:

- apêndice C, secção C.9

Também são tratados na seguinte secção de [COD3]:

- apêndice B, secção B.9