

Aulas 18 e 19

- Tecnologias de memória
- Organização genérica de um circuito de memória a partir de uma célula básica
- Memória SRAM (*Static Random Access Memory*):
 - organização de células básicas num array
 - ciclos de acesso para leitura e escrita: diagramas temporais
 - construção de módulos de memória SRAM
- Memória DRAM (*Dynamic Random Access Memory*):
 - célula básica; organização interna
 - ciclos de acesso para leitura e escrita: diagramas temporais
 - refrescamento: modo "RAS only"
 - construção de módulos de memória DRAM

José Luís Azevedo, Bernardo Cunha, Tomás O. Silva, P. Bartolomeu

Introdução – conceitos básicos

- RAM – Random Access Memory
 - Designação para memória volátil que pode ser lida e escrita
 - Acesso "random"
- ROM – Read Only Memory
 - Memória não volátil que apenas pode ser lida
 - Acesso "random"

(Acesso "random" - tempo de acesso é o mesmo para qualquer posição de memória)

Introdução – conceitos básicos

- Tecnologias:
 - Semicondutor
 - Magnética
 - Ótica
 - Magneto-ótica
- Memória volátil:
 - Informação armazenada perde-se quando o circuito é desligado da alimentação: RAM (SRAM e DRAM)
- Memória não volátil:
 - A informação armazenada mantém-se até ser deliberadamente alterada: EEPROM, Flash EEPROM, tecnologias magnéticas

Memória não volátil – evolução histórica

- **ROM** – programada durante o processo de fabrico (1965)
- **PROM** – Programmable Read Only Memory: programável uma única vez (1970)
- **EPROM** – Erasable PROM: escrita em segundos, apagamento em minutos (ambas efectuadas em dispositivos especiais) (1971)
- **EEPROM** – Electrically Erasable PROM (1976)
 - O apagamento e a escrita podem ser efetuados no próprio circuito em que a memória está integrada
 - O apagamento é feito byte a byte
 - Escrita muito mais lenta que leitura
- **Flash EEPROM** (tecnologia semelhante à EEPROM) (1985)
 - A escrita pressupõe o prévio apagamento das zonas de memória a escrever
 - O apagamento é feito por blocos (por exemplo, blocos de 4 kB) o que torna esta tecnologia mais rápida que a EEPROM
 - O apagamento e a escrita podem ser efetuados no próprio circuito em que a memória está integrada
 - Escrita muito mais lenta que leitura

Tecnologias de memória

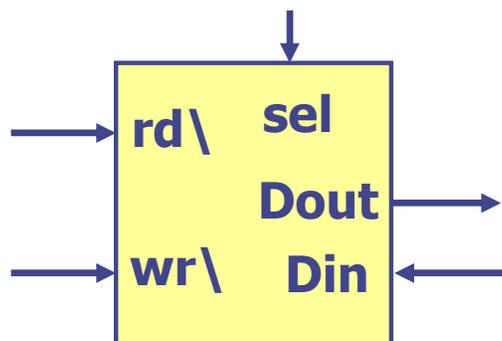
Tecnologia	Tempo Acesso	\$ / GB
SRAM	0,5 – 2,5 ns	\$500 - \$1000
DRAM	35 - 70 ns	\$10 - \$20
Flash	5 – 50 us	\$0,75 - \$1
Magnetic Disk	5 - 20 ms	\$0,005 - \$0,1

(Dados de 2012)

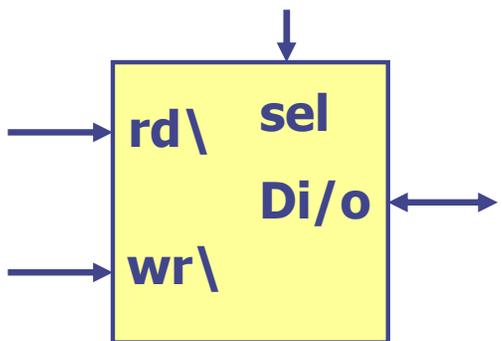
- SRAM - Static Random Access Memory
- DRAM - Dynamic Random Access Memory
- Dadas estas diferenças de custo e de tempo de acesso, é vantajoso construir o sistema de memória como uma hierarquia onde se utilizem todas estas tecnologias

Organização básica de memória

- Uma memória pode ser encarada como uma coleção de N registos de dimensão K (N x K)
- Cada registo é formado por K células, cada uma delas capaz de armazenar 1 bit
- Uma célula de memória (de 1 bit) pode ser representada por:



Din	– Data In (1 bit)
Dout	– Data Out (1 bit)
sel	– Select
rd\'	– Read\'
rw\'	– Write\'



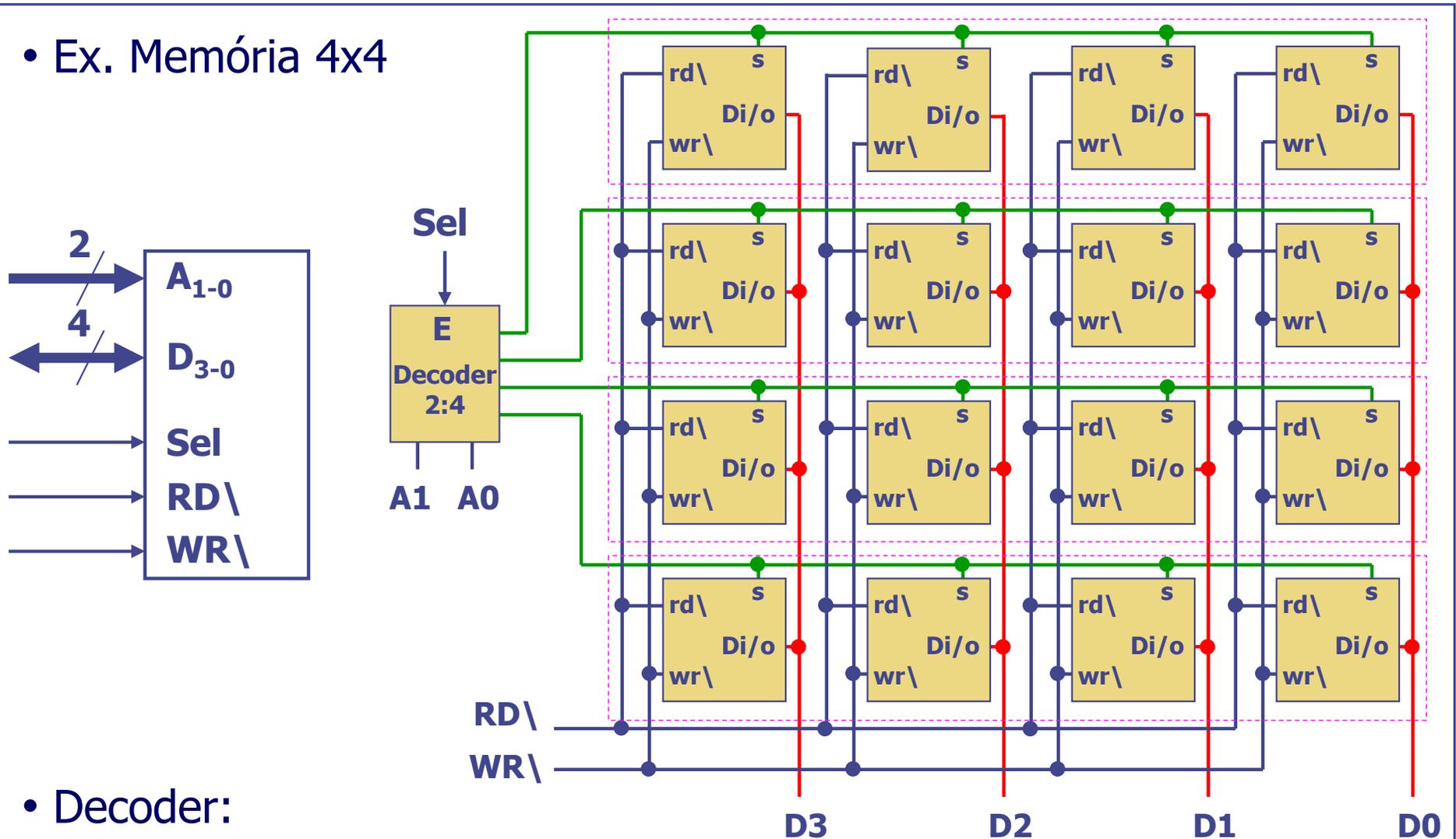
Di/o	– Data In/Out (1 bit)
sel	– Select
rd\'	– Read\'
rw\'	– Write\'

Agrupamento de células de memória

- Através do agrupamento de células-base pode formar-se uma memória de maior dimensão
- O que é necessário especificar:
 - O **número total de Words (N)** que a memória pode armazenar
 - O **Word size (K = 1, 4, 8, 16, 32, ...)**
(Número total de bits = word size * n^o words)
- Exemplos:
- 1k x 8
 - 8 bits / word
 - 1k = 2^{10} → 10 linhas de endereço → 1.024 endereços
- 1M x 4
 - 4 bits / word
 - 1M = 2^{20} → 20 linhas de endereço → 1.048.576 endereços

Organização 2D

- Ex. Memória 4x4



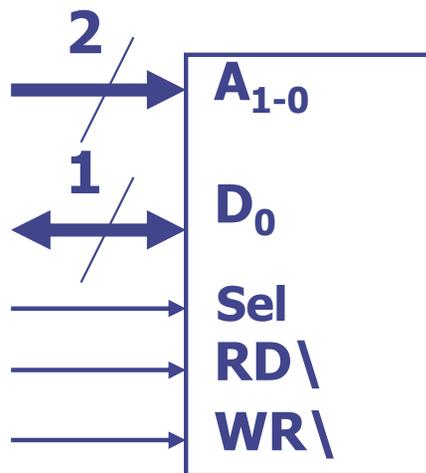
- Decoder:

- 2^N saídas

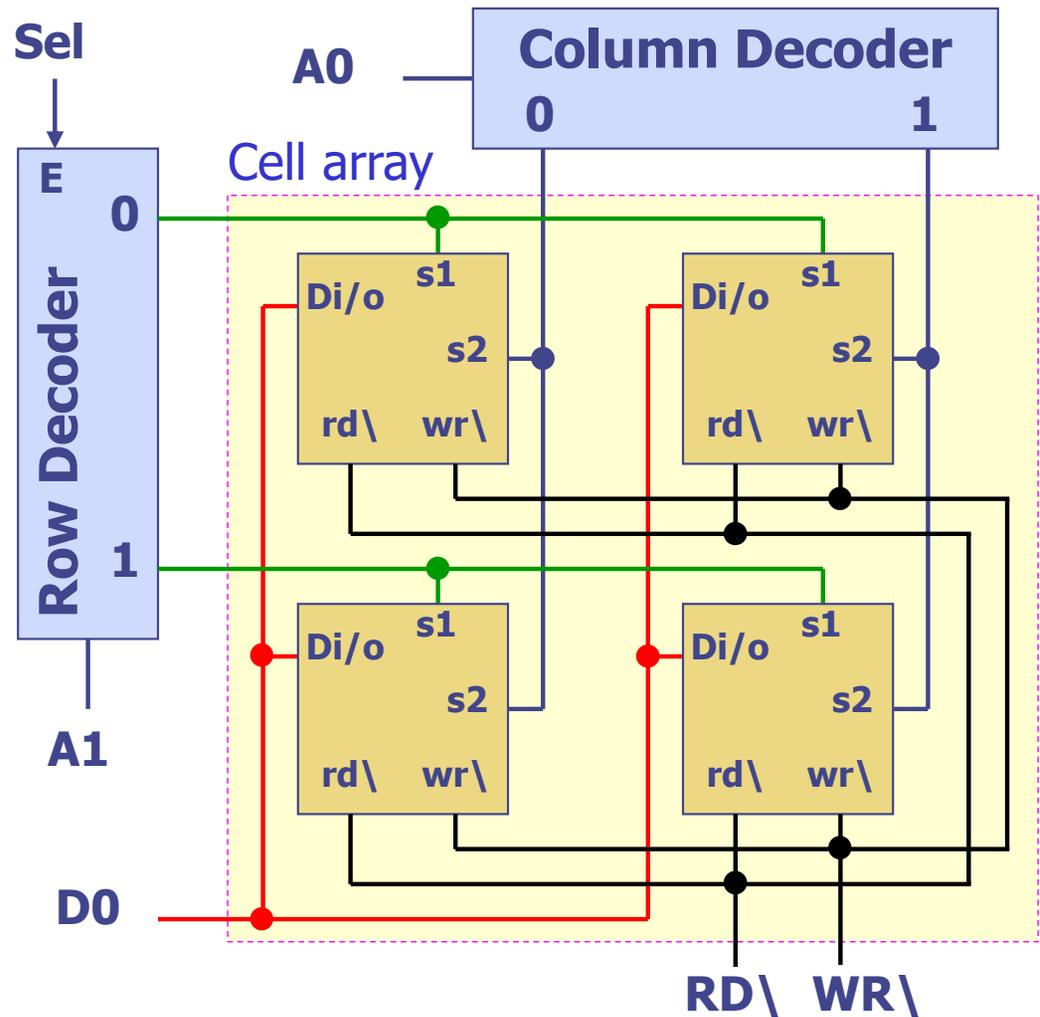
- Ex. $1M \times 4 = 2^{20} \times 4 \rightarrow 1.048.576$ saídas, N° de gates $\gg 2^{20}$

Organização em matriz (conceito)

- Ex. Memória 4x1



(Célula seleccionada: $S1.S2=1$)



Q1. E se a memória fosse de 16x1? e se fosse 8x1? e 1Mx1?

Q2. E se a memória fosse de 4x2? E se fosse 4x4?

Memória do tipo RAM (volátil)

- **SRAM – Static RAM**

- Vantagens:

- Rápida
- Informação permanece até que a alimentação seja cortada

- Inconvenientes:

- Implementações típicas: 6 transistores / célula
- Baixa densidade, elevada dissipação de potência
- Custo/bit elevado

- **DRAM – Dynamic RAM**

- Vantagens:

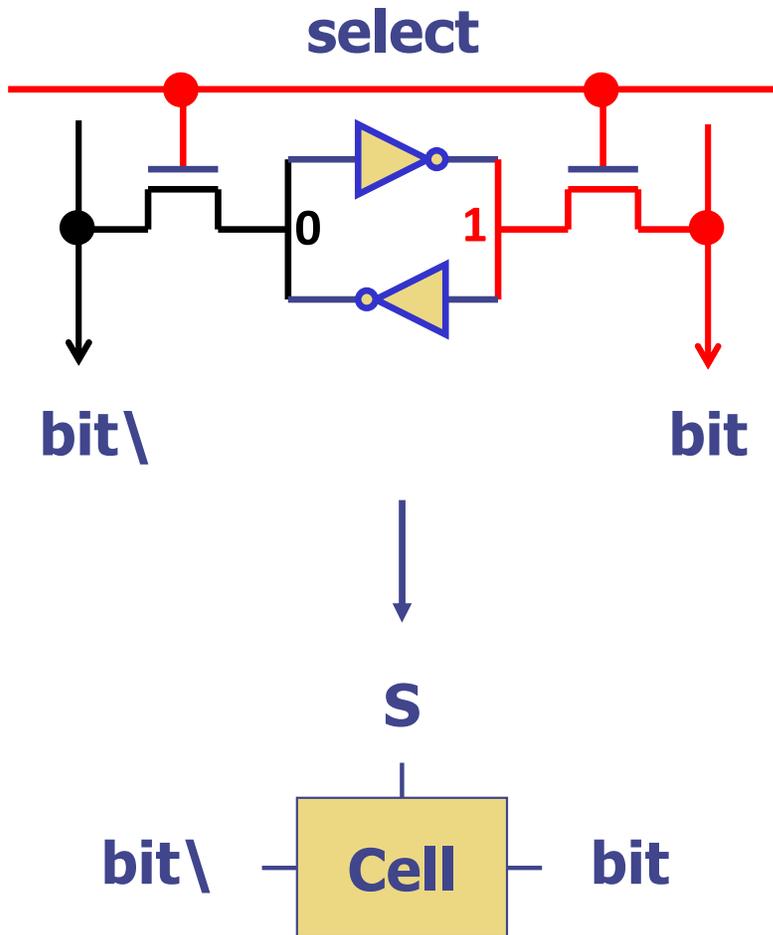
- Implementações típicas: (1 transistor + 1 condensador) / célula
- Alta densidade, baixa dissipação de potência
- Custo/bit baixo

- Inconvenientes:

- Informação permanece apenas durante alguns mili-segundos (necessita de *refresh* regular – daí a designação "dynamic")
- Mais lenta (pelo menos 1 ordem de grandeza) que a SRAM

RAM estática (SRAM)

- 6 transístores / célula



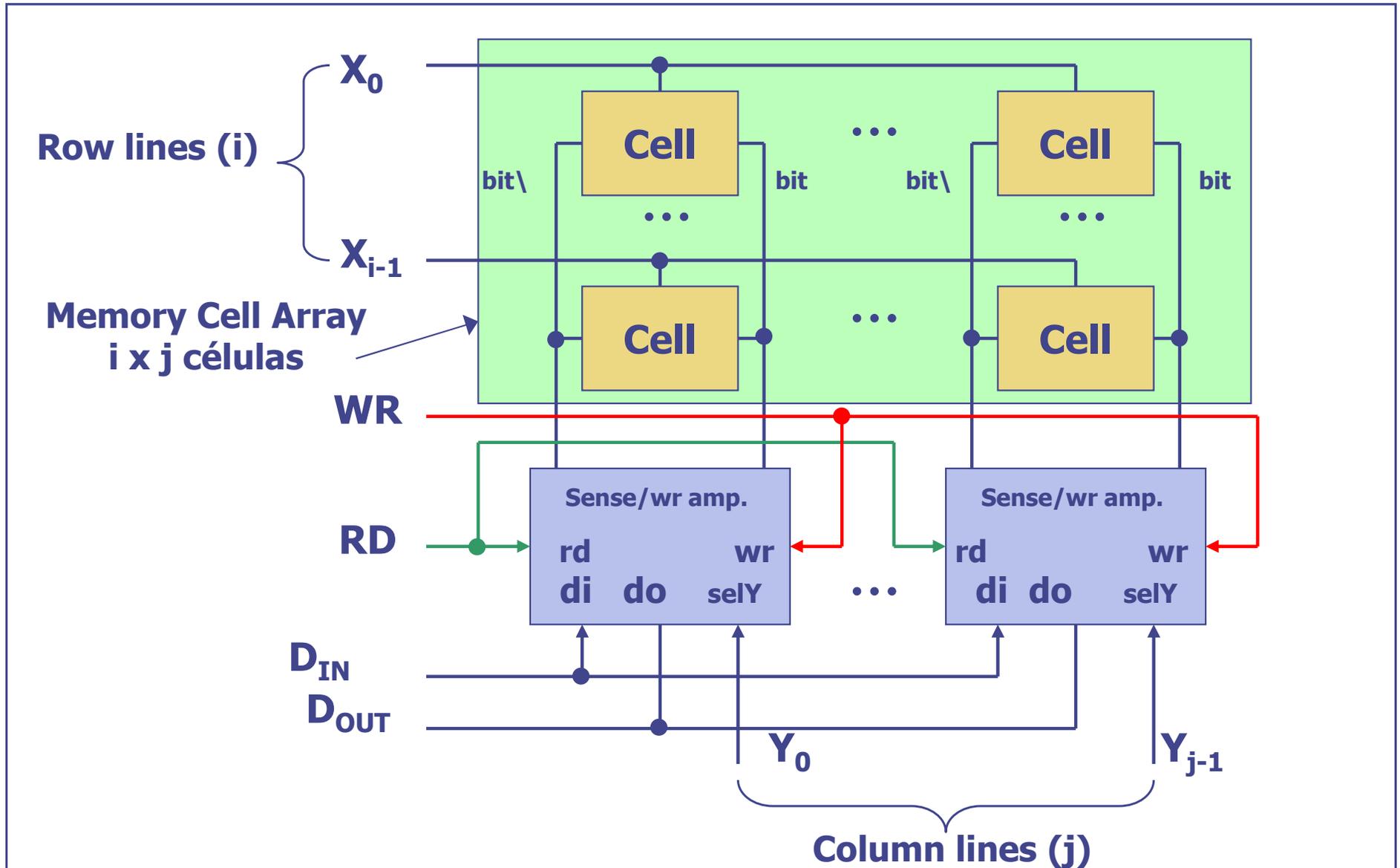
- **Write**

- Colocar a informação em "bit" (e "bit\"). Exemplo: para a escrita do valor lógico "1" – "bit"=1, "bit\"=0
- Ativar a linha "select"

- **Read**

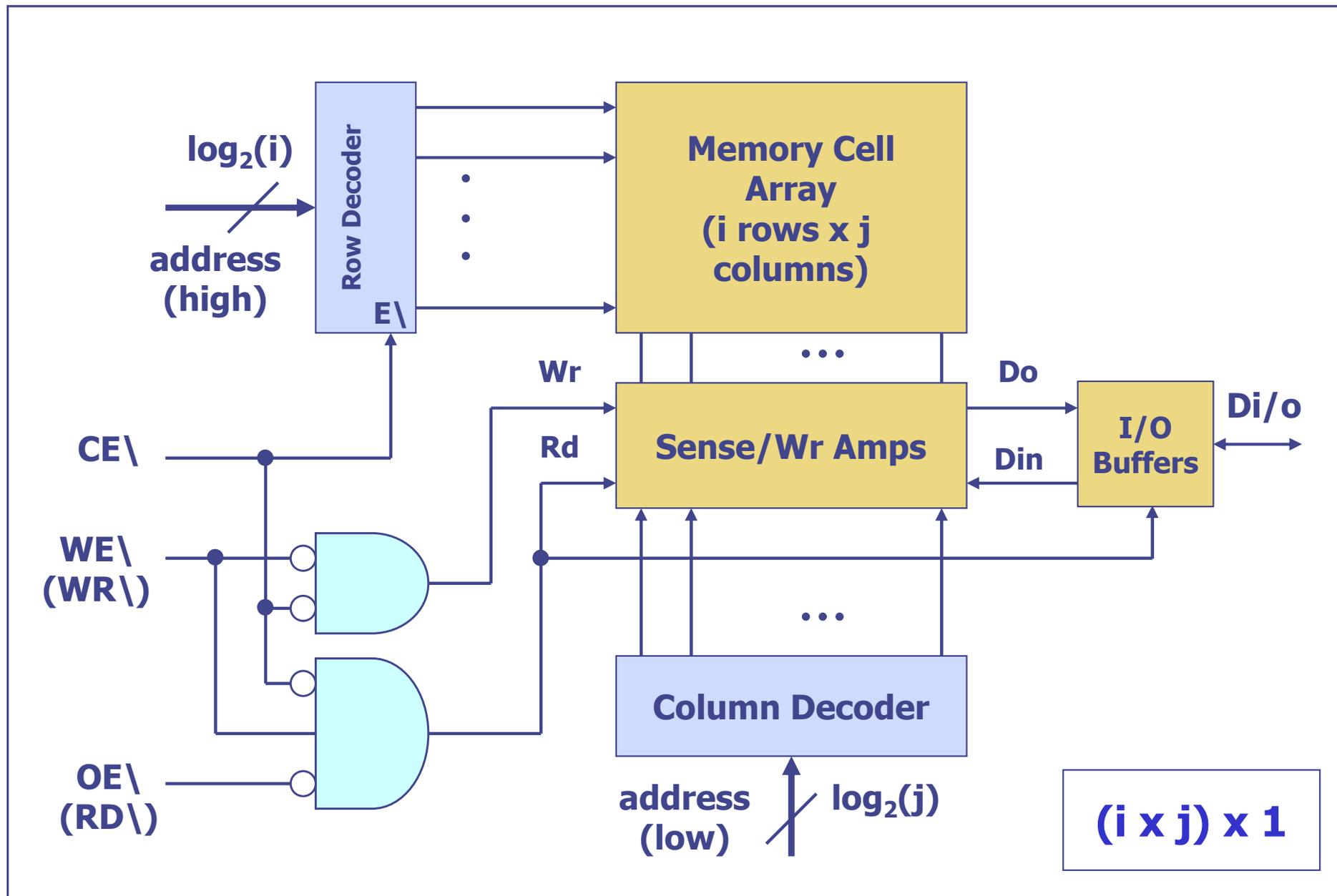
- Ativar a linha "select"
- O valor lógico armazenado na célula é detetado pela diferença de tensão entre as linhas "bit" e "bit\"

SRAM - Organização interna

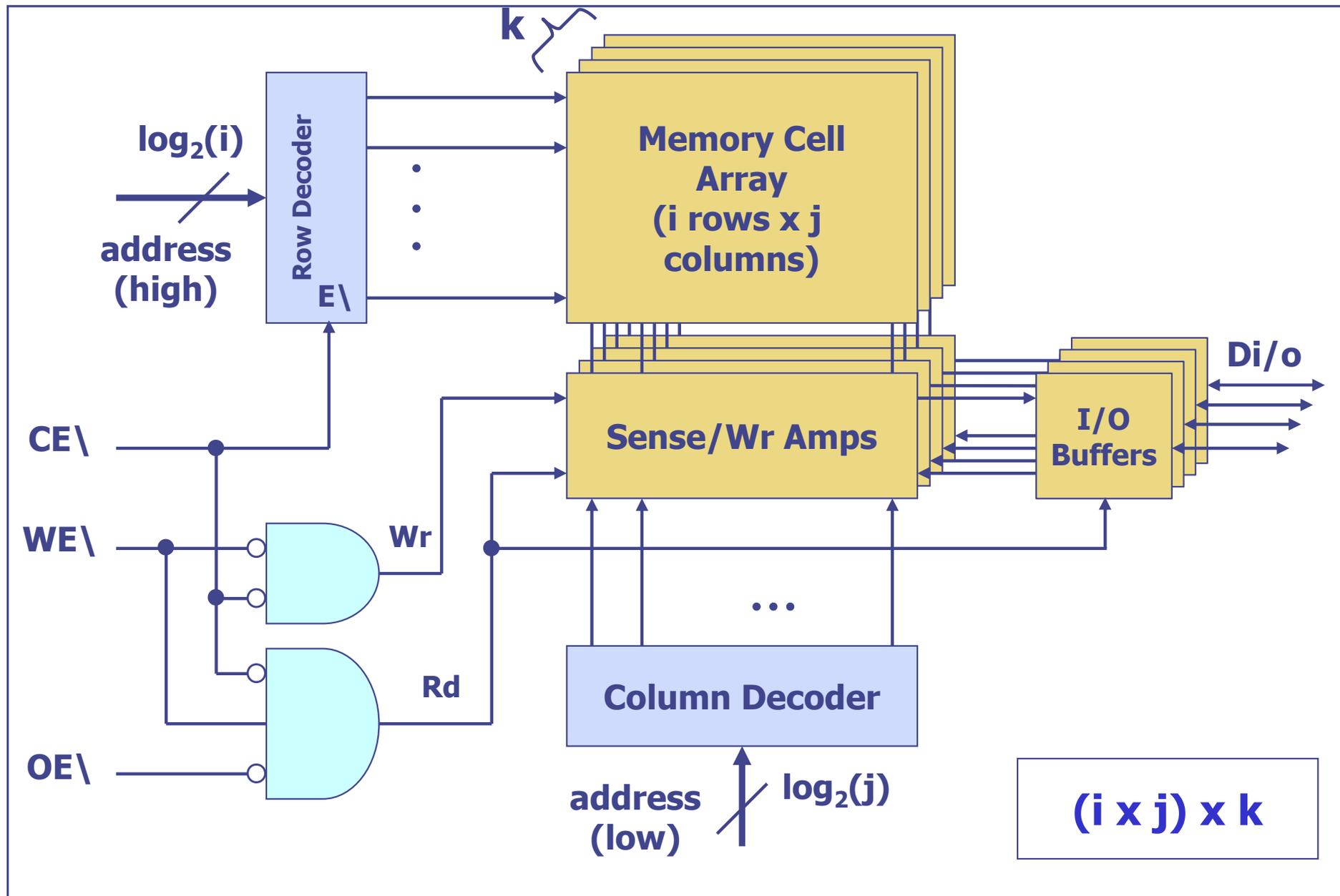


- Se $selY=0$, as linhas "bit" e "bit\" ficam em alta impedância

SRAM - Organização interna

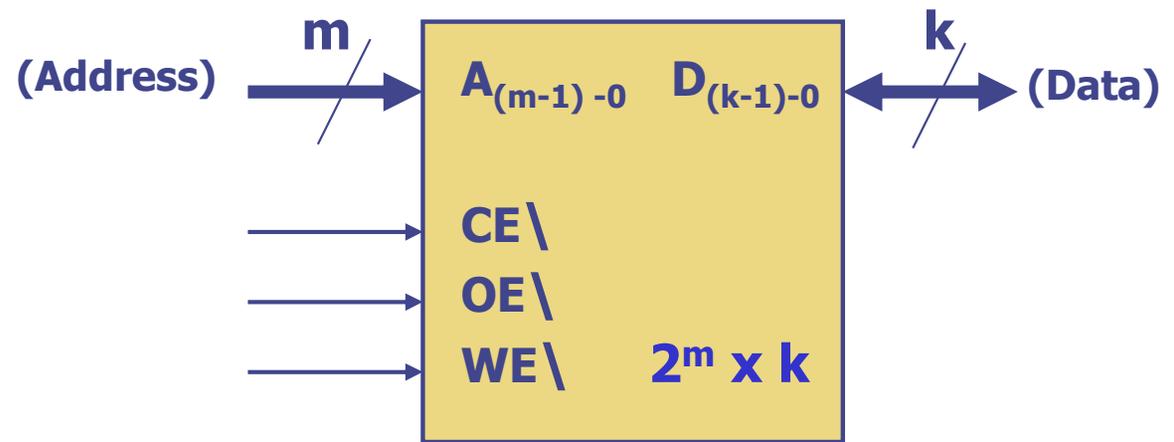


SRAM - Organização interna



SRAM - Bloco funcional

- Diagrama lógico (interface assíncrona)

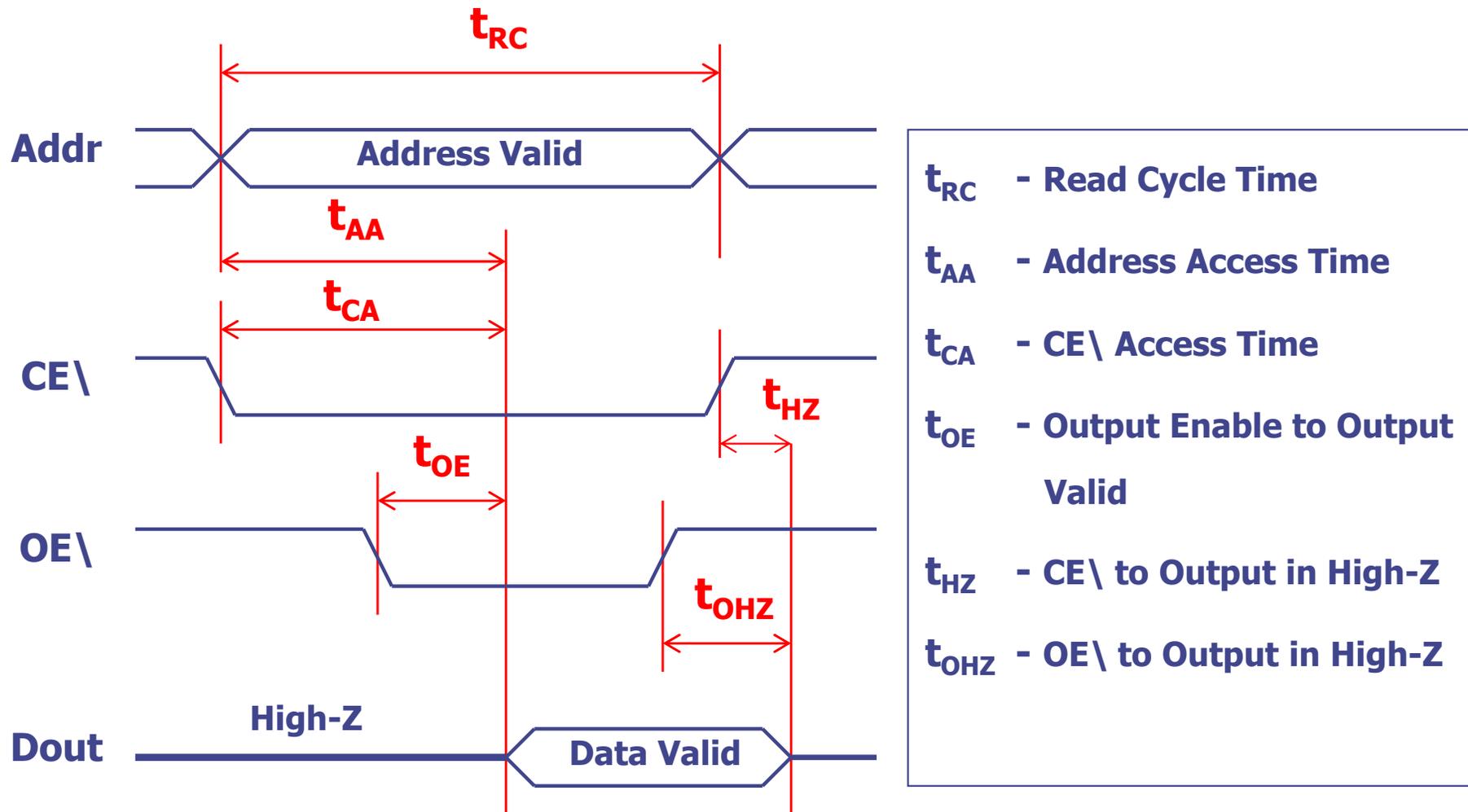


- Tabela de verdade

CE\	OE\	WE\	Operação
1	X	X	High-Z
0	1	1	High-Z
0	X	0	Escrita
0	0	1	Leitura

SRAM – Ciclo de Leitura

- Diagrama temporal típico de um ciclo de leitura de uma memória SRAM (interface assíncrona)



SRAM – Ciclo de leitura

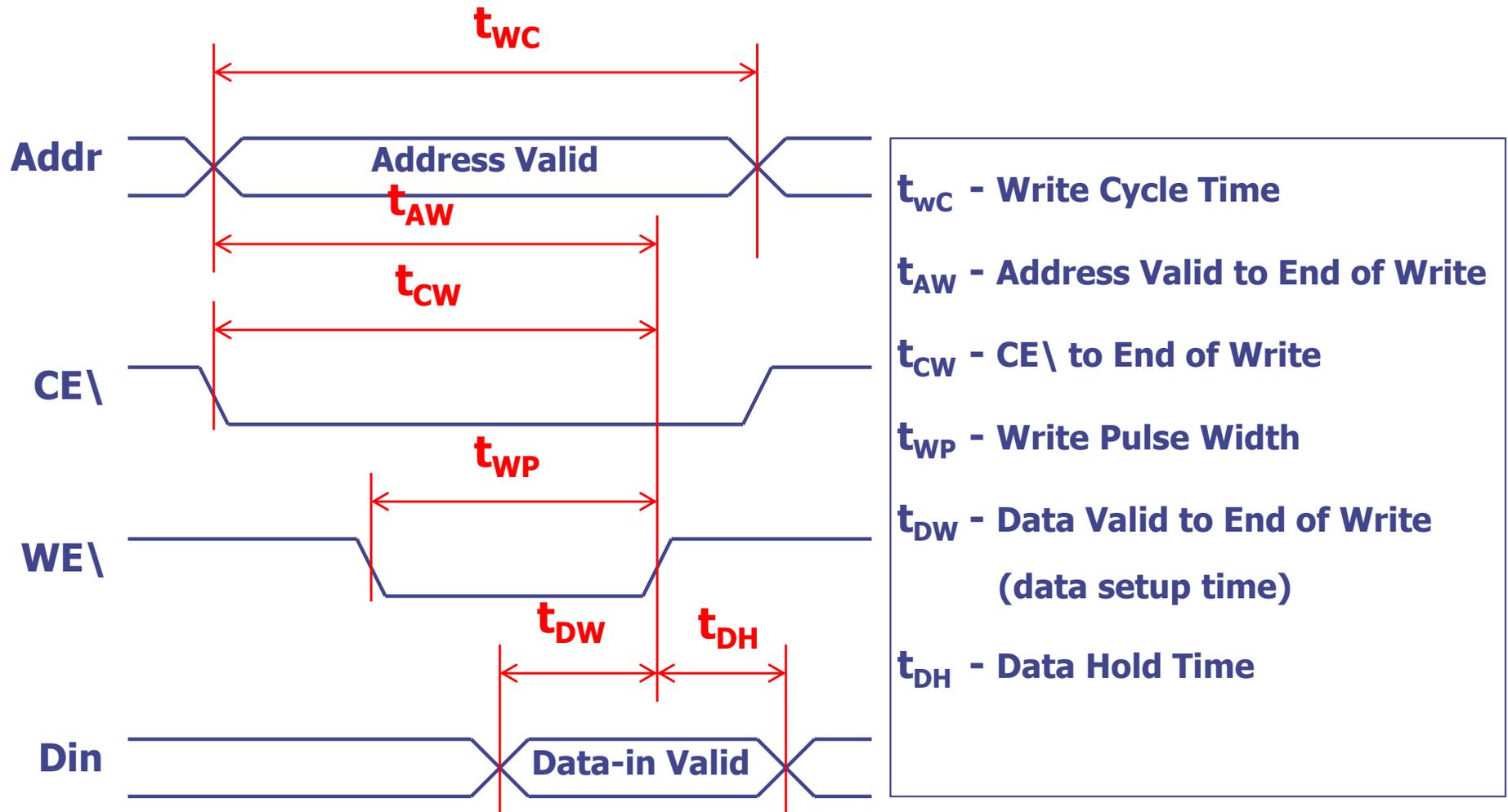
- Valores indicativos (em ns) dos parâmetros associados a um ciclo de leitura de uma memória SRAM:

Parameter	Symbol	Min.	Max.
Read Cycle Time	t_{RC}	1.5	
Address Access Time	t_{AA}		1.5
CE\ Access Time	t_{CA}		1.5
Output Enable to Output Valid	t_{OE}		0.7
CE\ to Output in High-Z	t_{HZ}		0.6
OE\ to Output in High-Z	t_{OHZ}		0.6

- **Cycle Time:** tempo de acesso mais qualquer tempo adicional necessário antes que um segundo acesso possa ter início
- **Access Time:** tempo necessário para os dados ficarem disponíveis no barramento de saída da memória
- **Taxa de transferência:** taxa a que os dados podem ser transferidos de/para uma memória ($1 / \text{cycle_time}$)

SRAM – Ciclo de Escrita

- Diagrama temporal típico de um ciclo de escrita de uma memória SRAM



SRAM – Ciclo de Escrita

- Valores indicativos (em ns) dos parâmetros associados a um ciclo de escrita de uma memória SRAM:

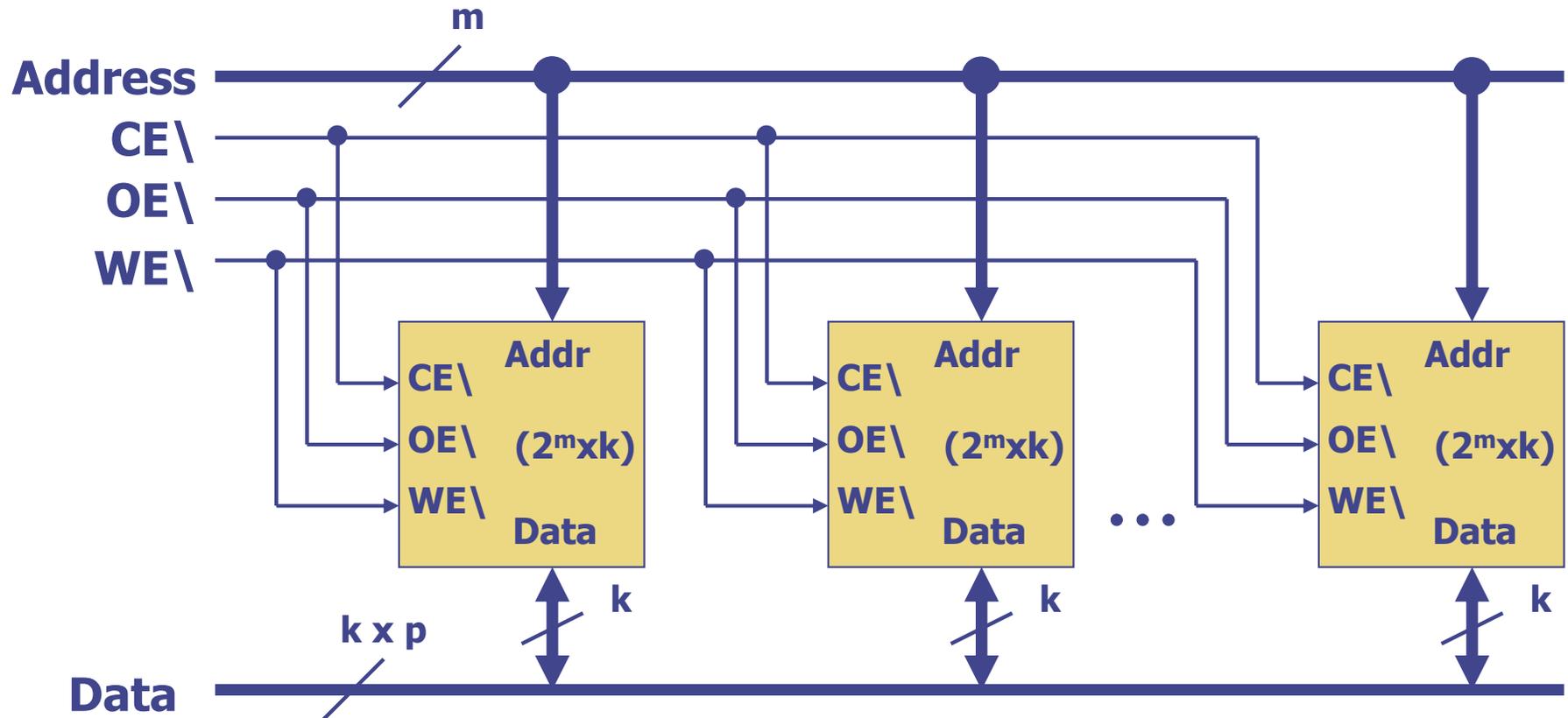
Parameter	Symbol	Min.	Max.
Write Cycle Time	t_{WC}	1.5	
Address Valid to End of Write	t_{AW}	1.0	
CE\ to End of Write	t_{CW}	1.0	
Write Pulse Width	t_{WP}	1.0	
Data Valid to End of Write	t_{DW}	0.7	
Data Hold Time	t_{DH}	0	

Aumento da capacidade de armazenamento

- É frequente ter-se necessidade de memórias com uma capacidade de armazenamento superior à capacidade individual dos circuitos disponíveis comercialmente
- Nessa situação recorre-se à construção de módulos de memória que resultam do agrupamento de circuitos de acordo com o aumento pretendido
- Assim, a construção de um módulo de memória pode envolver as duas fases seguintes, ou apenas uma delas, em função dos circuitos disponíveis e dos requisitos finais de armazenamento:
 - **Aumento da dimensão da palavra.** Exemplo: a partir de C.I.s de 32K x 1, construir uma memória de 32K x 8
 - **Aumento do número total de posições de memória.** Exemplo: a partir de C.I.s de 32K x 8, construir uma memória de 128K x 8

Módulo de memória SRAM

- Aumento da dimensão da palavra



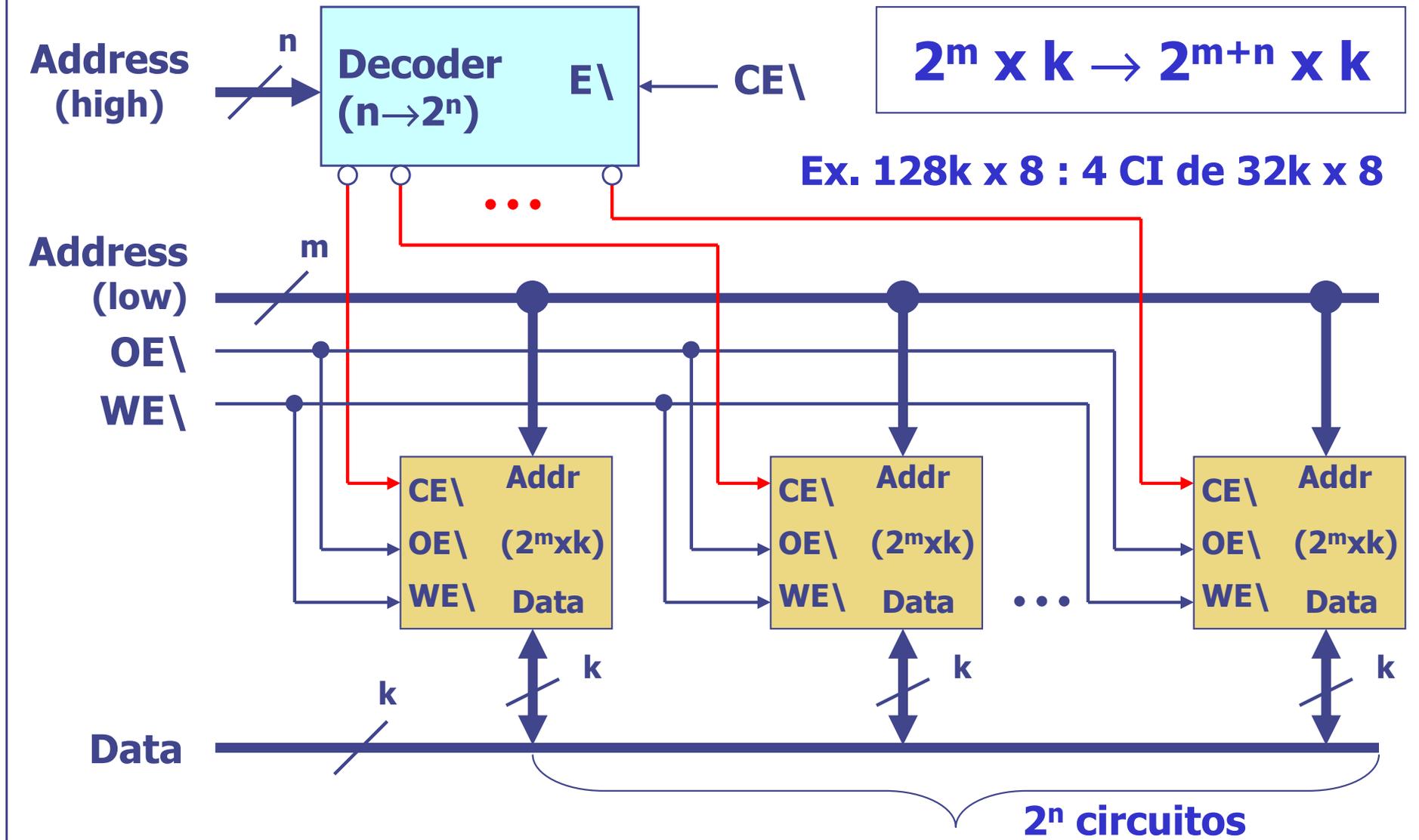
$$2^m \times k \rightarrow 2^m \times (k \times p)$$

"p" circuitos

Ex. $32k \times 8$: 8 CI de $32k \times 1$

Módulo de memória SRAM

- Aumento do número total de posições de memória



Memória do tipo RAM (volátil)

- **SRAM – Static RAM**

- Vantagens:

- Rápida
- Informação permanece até que a alimentação seja cortada

- Inconvenientes:

- Implementações típicas: 6 transistores / célula
- Baixa densidade, elevada dissipação de potência
- Custo/bit elevado

- **DRAM – Dynamic RAM**

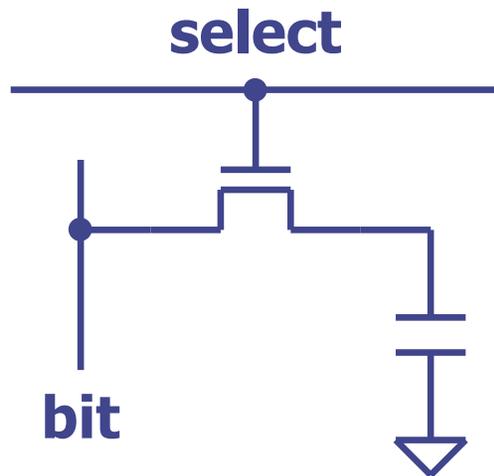
- Vantagens:

- Implementações típicas: (1 transistor + 1 condensador) / célula
- Alta densidade, baixa dissipação de potência
- Custo/bit baixo

- Inconvenientes:

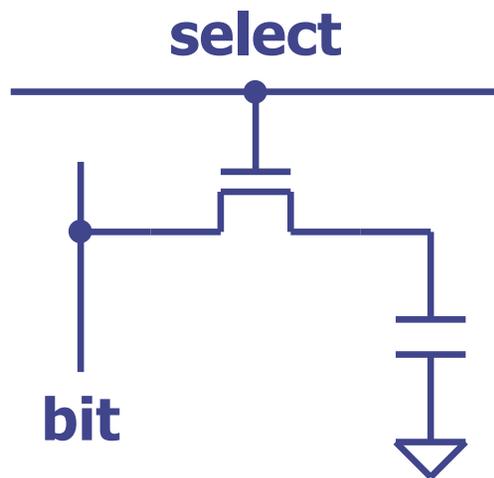
- Informação permanece apenas durante alguns mili-segundos (necessita de *refresh* regular – daí a designação "dynamic")
- Mais lenta (pelo menos 1 ordem de grandeza) que a SRAM

RAM Dinâmica (DRAM)



- Condensador com uma capacidade muito pequena (dezenas de fF ($1 \text{ fF} = 10^{-15} \text{ F}$))
- Na ausência de leitura, o condensador descarrega "lentamente"
- Informação permanece na célula apenas durante alguns milisegundos
- É obrigatório fazer o refrescamento ("refresh") periódico da carga do condensador
- A operação de leitura é destrutiva (descarrega o condensador)
- Após uma operação de leitura é necessário repor a carga no condensador

RAM Dinâmica (DRAM)



- **Write**

- Colocar dado na linha "bit"
- Ativar a linha "select"

- **Read**

- Pre-carregar a linha "bit" a $V_{DD}/2$
- Ativar a linha "select"
- Valor lógico detetado pela diferença de tensão na linha bit, relativamente a $V_{DD}/2$
- Restauro do valor da tensão no condensador (write)

- **Refresh da célula**

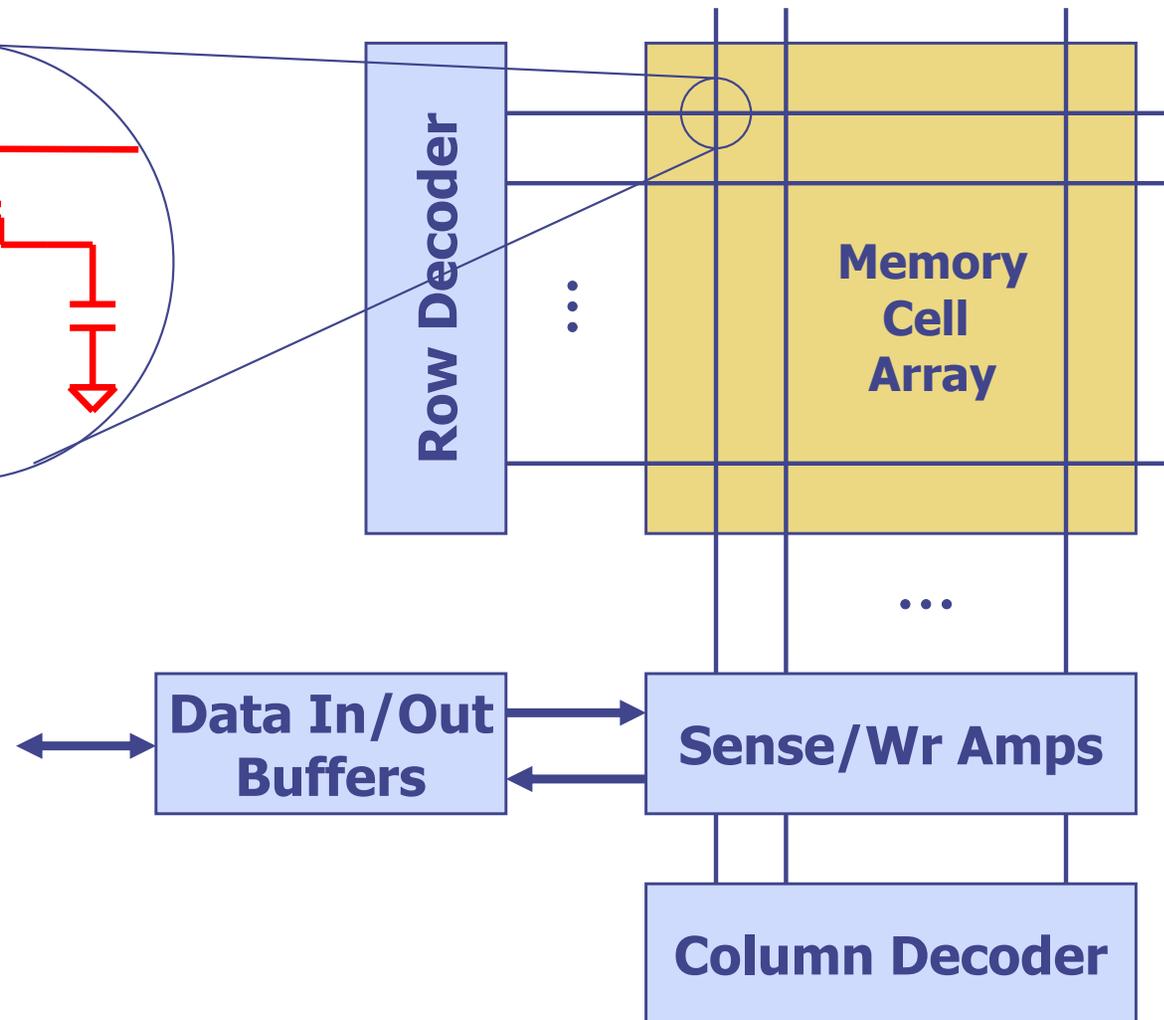
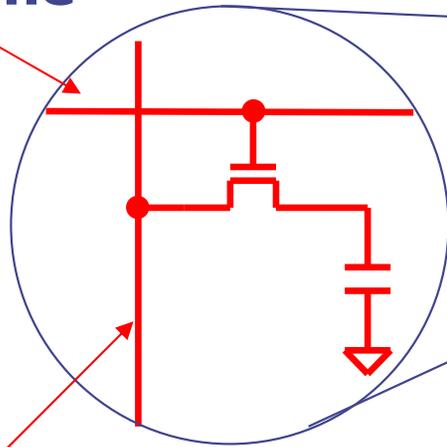
- Operação interna idêntica a uma operação de "Read"

RAM Dinâmica (DRAM)

- Organização em matriz

Select line

Bit line



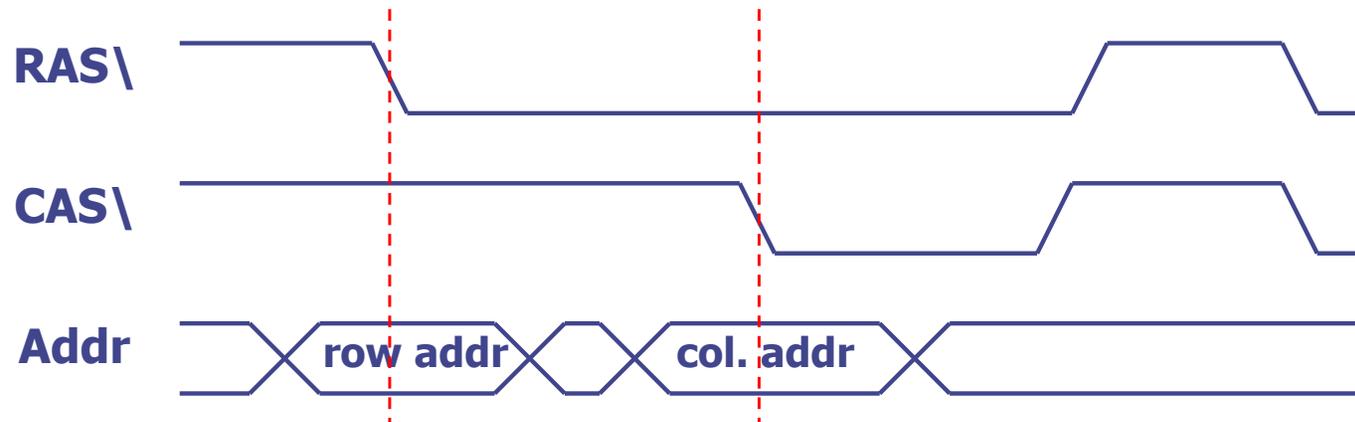
RAM Dinâmica (DRAM)

- O endereço de acesso à memória é dividido em 2 partes:

Address:

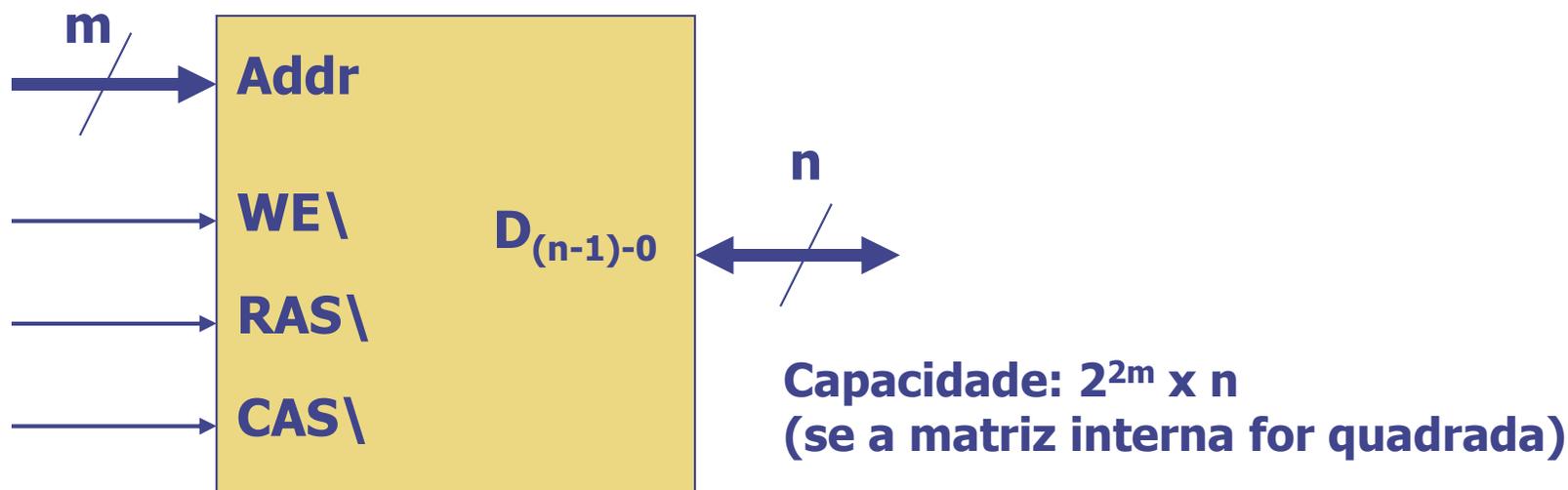
Row Address	Column Address
--------------------	-----------------------

- O barramento de endereços é multiplexado: primeiro é enviado o **endereços de linha** e depois é enviado o **endereço de coluna**
- A multiplexagem no tempo é feita com 2 strobes independentes
 - **RAS** – Row Address Strobe
 - **CAS** – Column Address Strobe



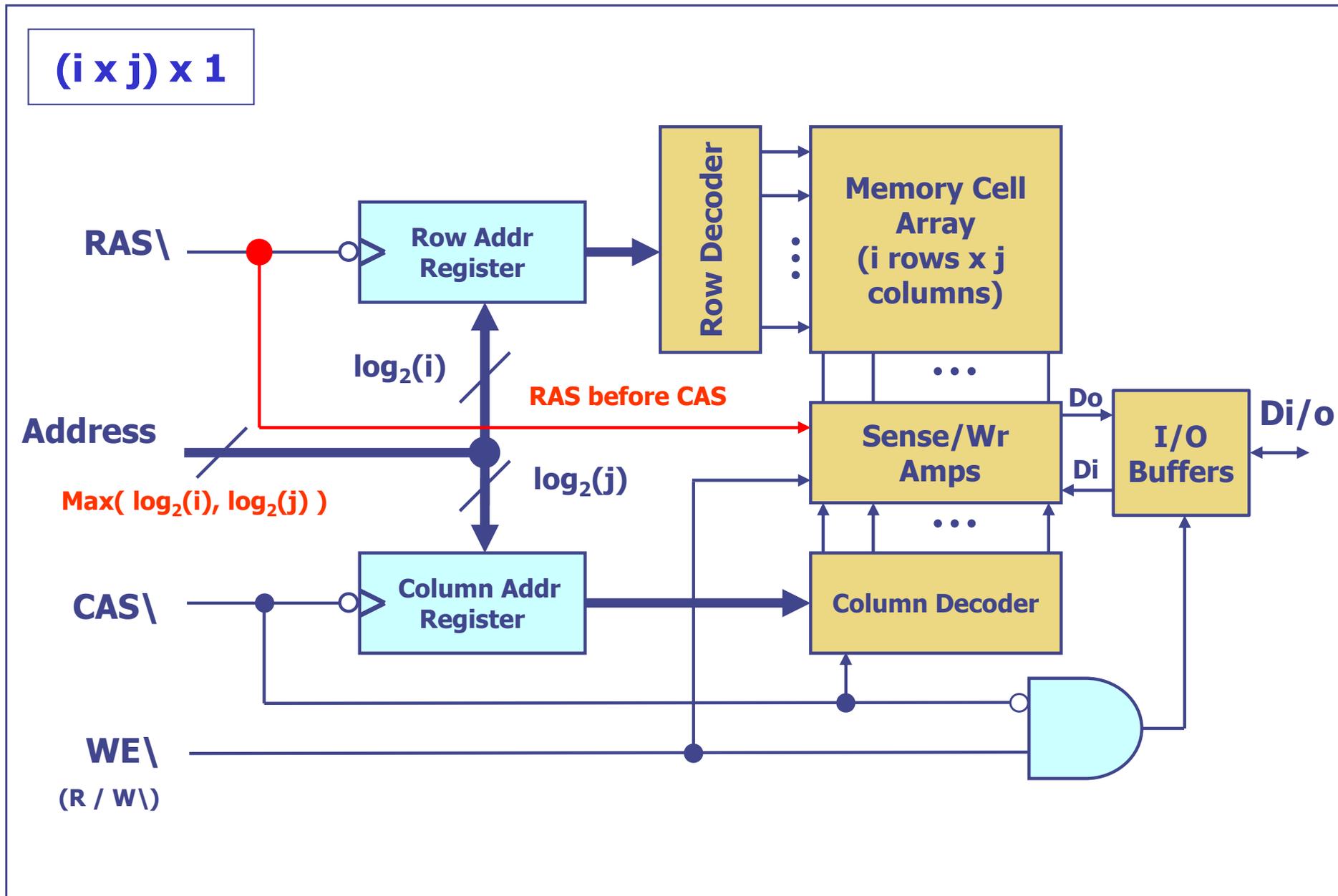
- As transições do RAS e do CAS são usadas para armazenar internamente os endereços de linha e de coluna, respectivamente
- Linha CAS funciona também como "chip-enable"

DRAM - Diagrama lógico



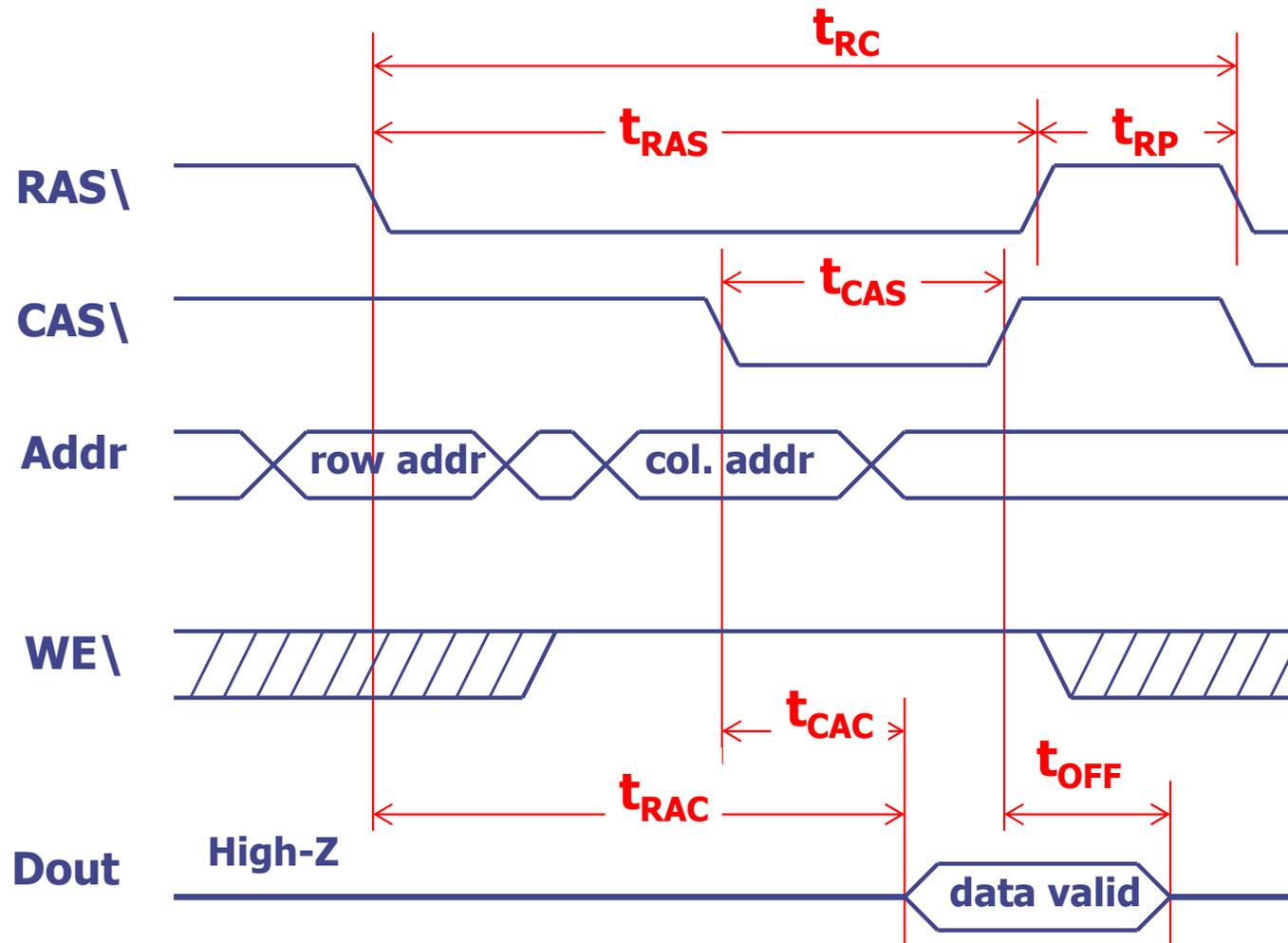
- $WE\ = 0 \rightarrow$ escrita; $WE\ = 1 \rightarrow$ leitura ($\equiv R/W\$)
- **RAS**: valida endereço da linha na transição descendente
- **CAS**: valida endereço da coluna na transição descendente

DRAM – Diagrama de blocos conceptual



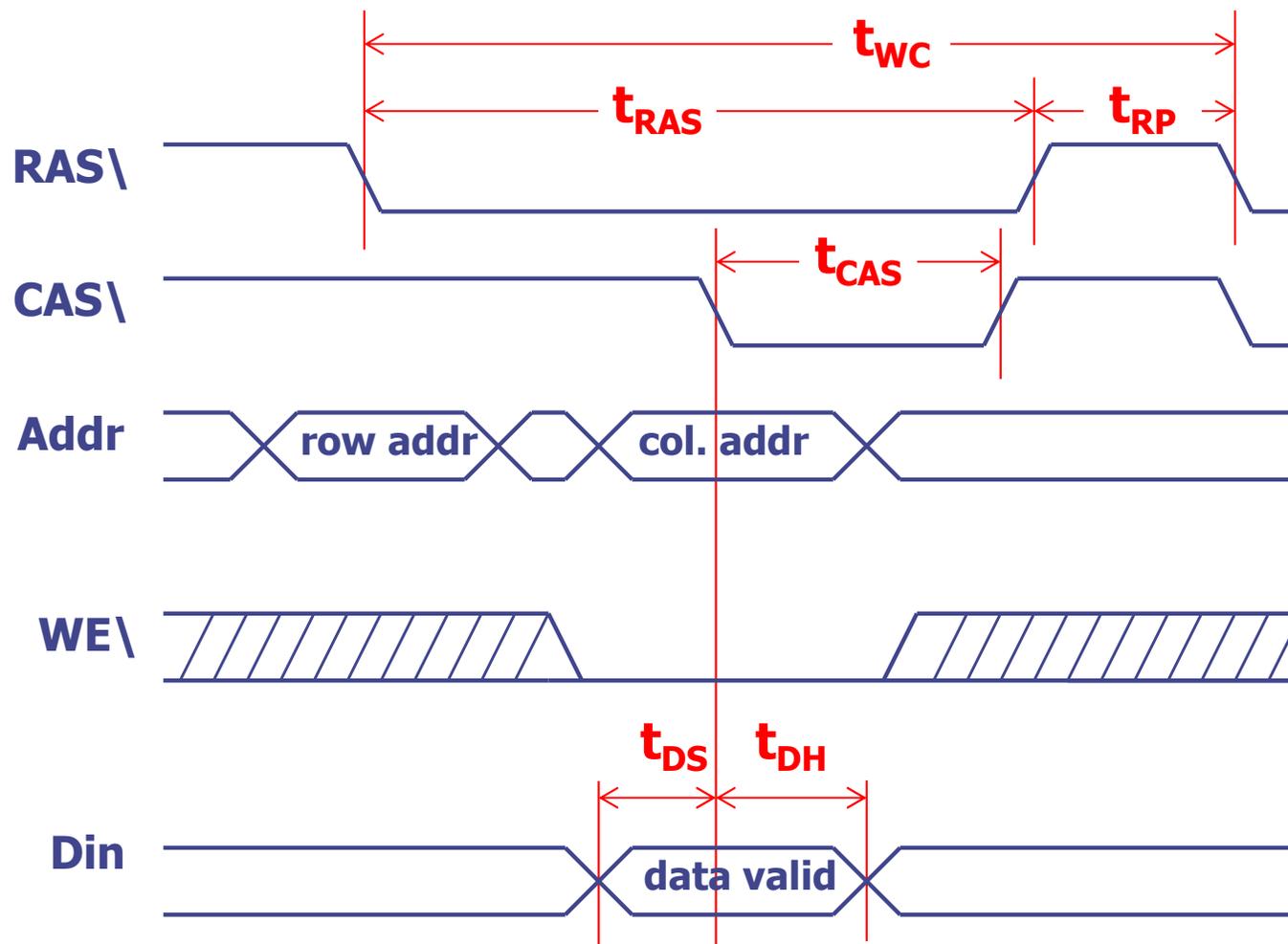
DRAM – Ciclo de Leitura

- Diagrama temporal típico de um ciclo de leitura de uma memória DRAM



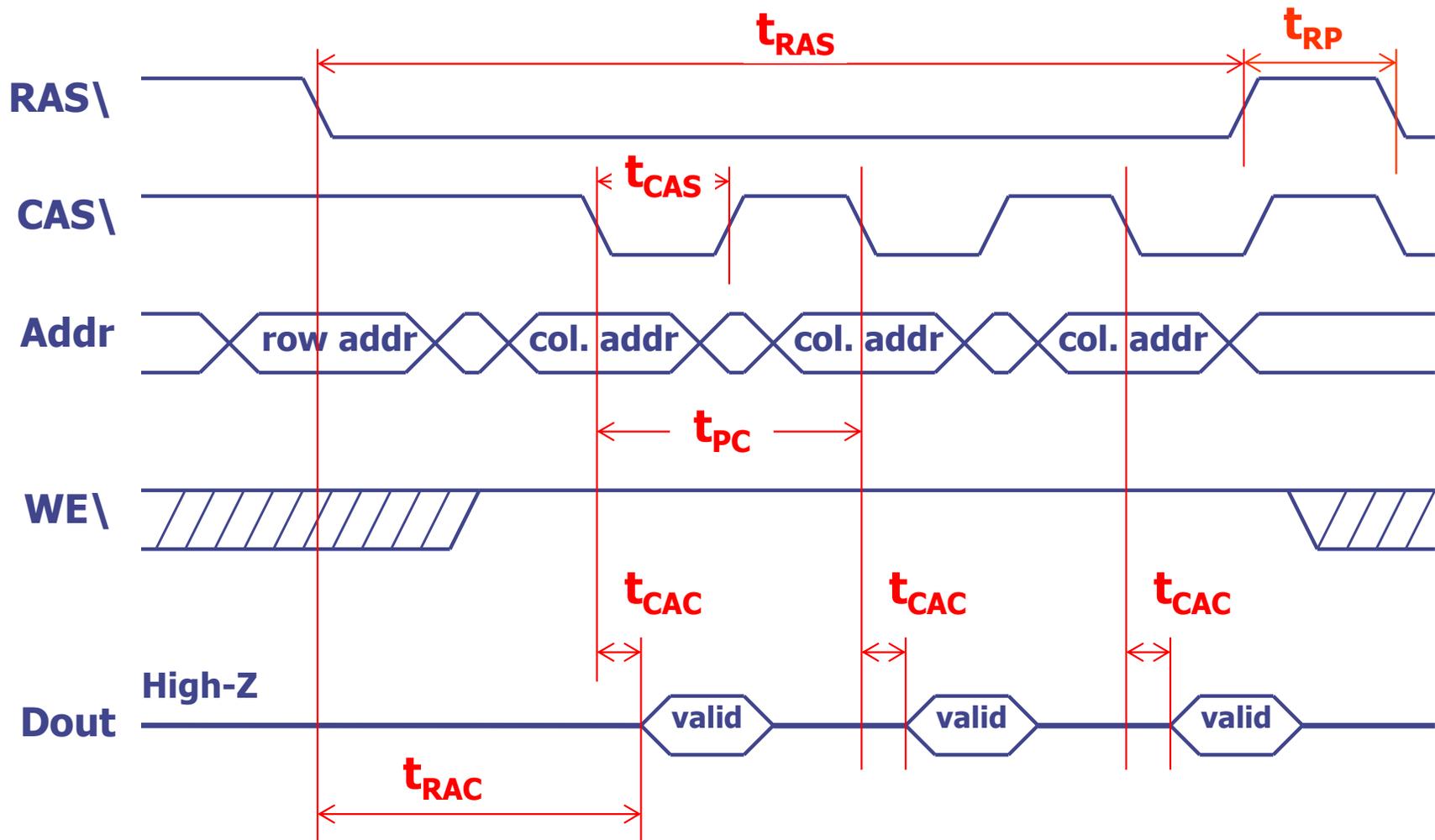
DRAM – Ciclo de Escrita

- Diagrama temporal típico de um ciclo de escrita (*early write*) de uma memória DRAM



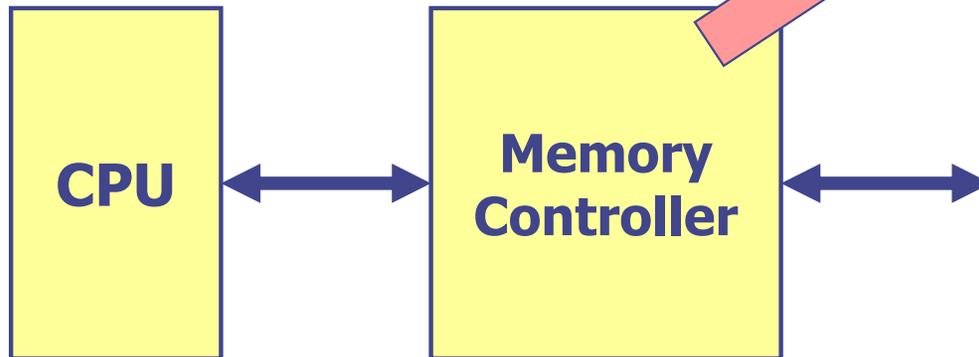
DRAM – Ciclo de Leitura em *page mode*

- Diagrama temporal típico de um ciclo de leitura de uma memória DRAM, em modo paginado (*page mode*)

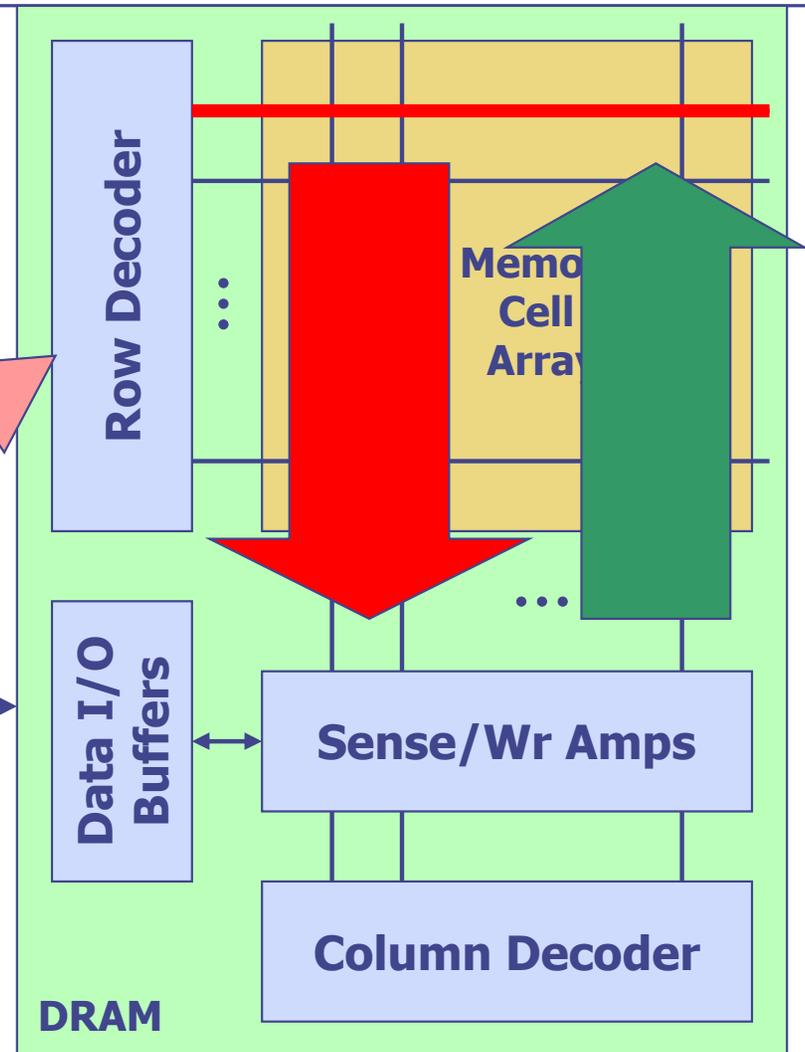


DRAM – Refrescamento

Operação interna de leitura (não há troca de informação com o exterior)

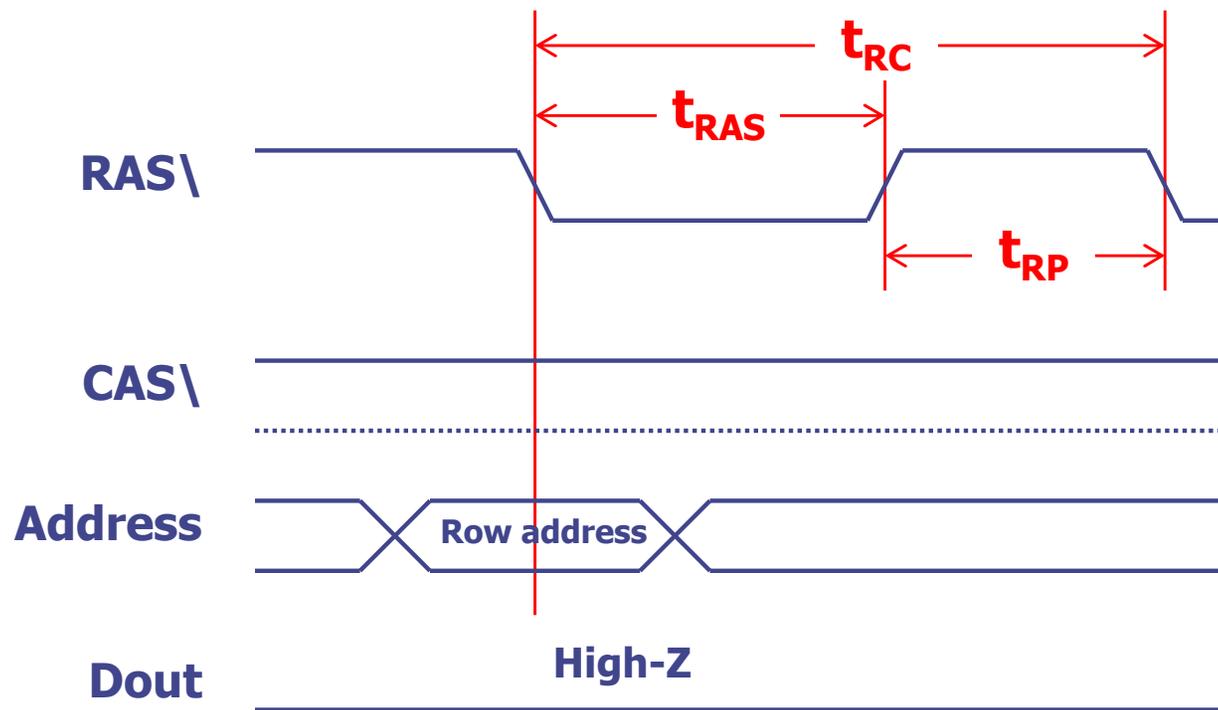


O refrescamento é efetuado, de forma sequencial, para cada uma das linhas da matriz



DRAM Refresh – RAS Only

- O *refresh* é feito simultaneamente em **todas as células da mesma linha da matriz** (especificada no address bus, no momento da ativação do sinal RAS\)
- O sinal CAS\ mantém-se inativo durante o processo



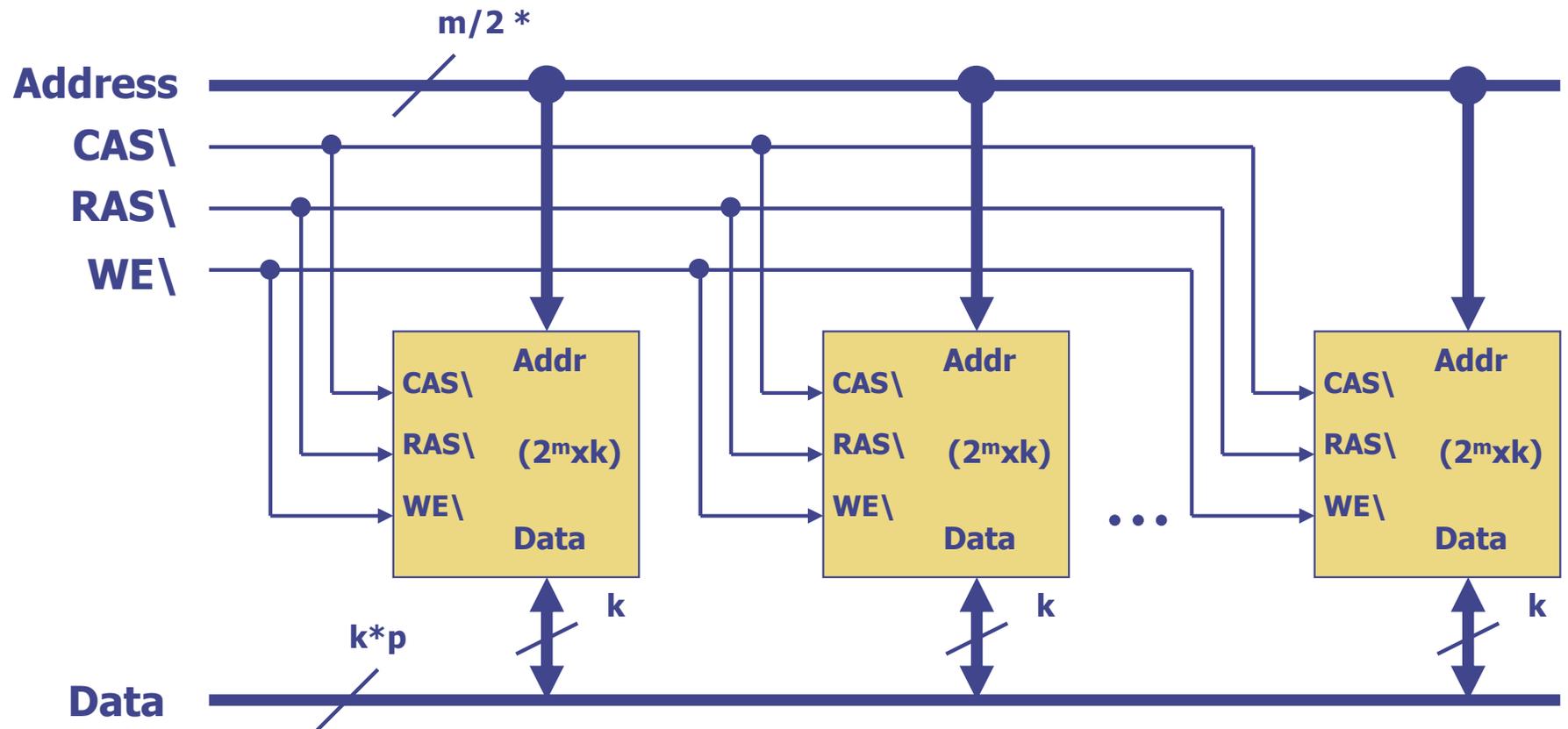
DRAM - Parâmetros principais

- Valores indicativos (em ns) dos tempos indicados nos diagramas temporais de leitura e escrita de uma memória DRAM com um tempo de acesso de 55 ns:

Parameter	Symbol	Min.	Max.
Read or Write Cycle Time	t_{RC}	100	
RAS\ precharge time	t_{RP}	45	
Page mode cycle time	t_{PC}	35	
RAS\ pulse width	t_{RAS}	55	10000
CAS\ pulse width	t_{CAS}	28	10000
Data-in setup time	t_{DS}	5	
Data-in hold time	t_{DH}	14	
Output buffer turn-off delay	t_{OFF}		15
Access time from RAS\	t_{RAC}		55
Access time from CAS\	t_{CAC}		28

Módulo de memória DRAM

- Aumento da dimensão da palavra



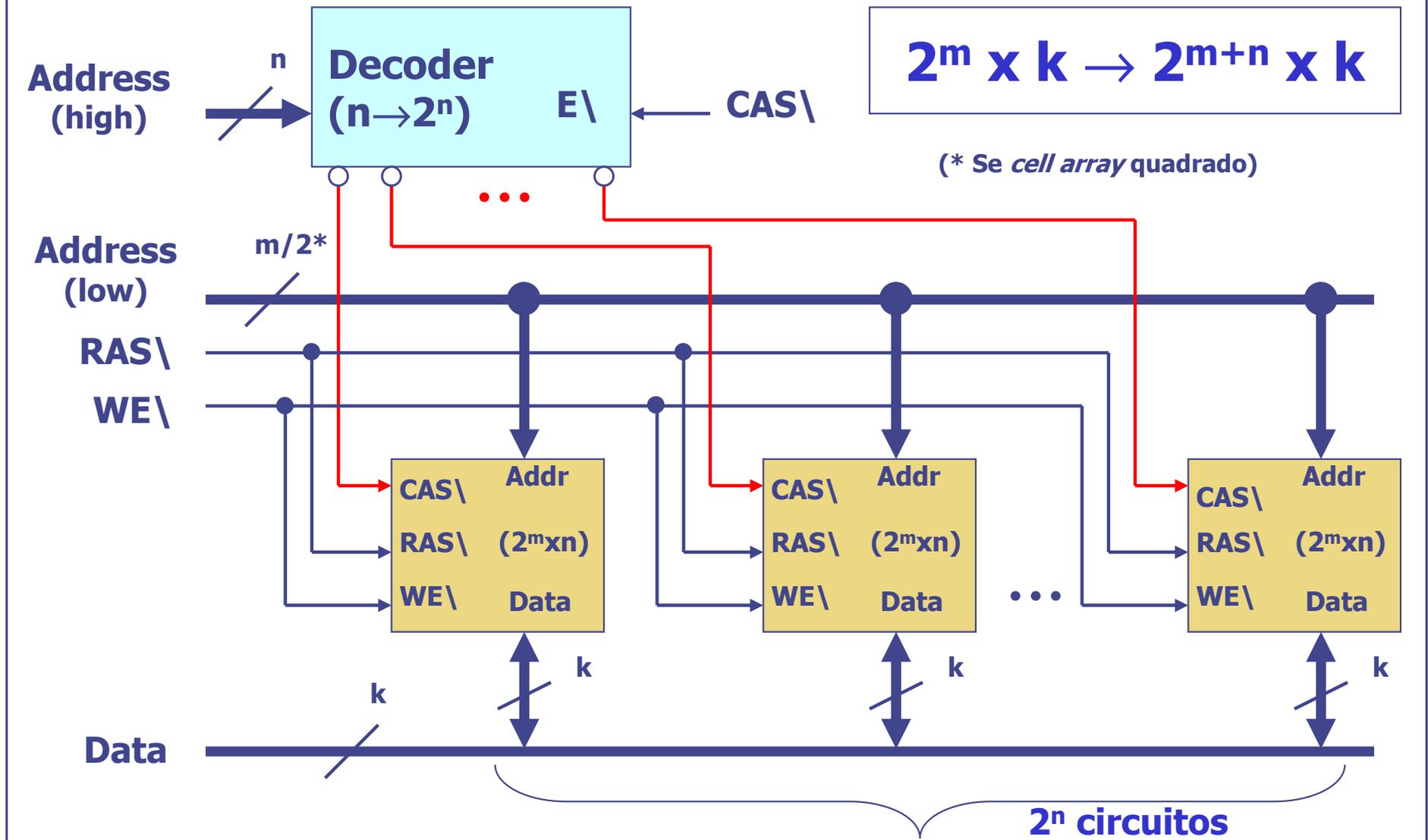
$$2^m \times k \rightarrow 2^m \times (k \cdot p)$$

p circuitos

(* Se *cell array* quadrado)

Módulo de memória DRAM

- Aumento do número total de posições de memória



Melhorias de desempenho da DRAM

- **Fast Page Mode**

- Adiciona sinais de temporização que permitem acessos repetidos ao buffer de linha (sem outro tempo de acesso à linha)

- **Synchronous DRAM (SDRAM)**

- Adiciona um sinal de relógio à interface DRAM, para facilitar a sincronização de transferências múltiplas
- Múltiplos bancos, cada um com o seu buffer de linha

- **Double Data Rate (DDR SDRAM)**

- Transferência de dados tanto no flanco ascendente como no flanco descendente do sinal de relógio (duplica a taxa de transferência de pico)
- Versão atual: DDR5 (2021->). Exemplo: DDR5-6400, 6400 Milhões de transferências por segundo, relógio de 3.2 GHz

- Estas técnicas melhoram a largura de banda, mas não a latência

Name		Release year	Chip			Bus			Voltage (V)
Gen	Standard		Clock rate (MHz)	Cycle time (ns)	Pre-fetch	Clock rate (MHz)	Transfer rate (MT/s)	Bandwidth (MB/s)	
DDR	DDR-200	1998	100	10	2n	100	200	1600	2.5
	DDR-266		133	7.5		133	266	2133 ¹ / ₃	
	DDR-400		200	5		200	400	3200	
DDR2	DDR2-400	2003	100	10	4n	200	400	3200	1.8
	DDR2-533		133 ¹ / ₃	7.5		266 ² / ₃	533 ¹ / ₃	4266 ² / ₃	
	DDR2-800		200	5		400	800	6400	
	DDR2-1066		266 ² / ₃	3.75		533 ¹ / ₃	1066 ² / ₃	8533 ¹ / ₃	
DDR3	DDR3-800	2007	100	10	8n	400	800	6400	1.5/1.35
	DDR3-1600		200	5		800	1600	12800	
	DDR3-1866		233 ¹ / ₃	4.29		933 ¹ / ₃	1866 ² / ₃	14933 ¹ / ₃	
DDR4	DDR4-1600	2014	200	5	8n	800	1600	12800	1.2/1.05
	DDR4-2400		300	3 ¹ / ₃		1200	2400	19200	
	DDR4-2666		333 ¹ / ₃	3		1333 ¹ / ₃	2666 ² / ₃	21333 ¹ / ₃	
	DDR4-3200		400	2.5		1600	3200	25600	
DDR5	DDR5-3200	2020	200	5	16n	1600	3200	25600	1.1
	DDR5-4000		250	4		2000	4000	32000	
	DDR5-5600		350	2.86		2800	5600	44800	
	DDR5-6400		400	2.5		3200	6400	51200	
	DDR5-7200		450	2.22		3600	7200	57600	