Estructura de Computadors

Tema 5 La memòria principal







Objectius

- Presentar una visió del conjunt del sistema de memòria del computador
- Revisar les característiques bàsiques i més rellevants de les memòries RAM, concretament les que determinen les seues prestacions
- Comprendre l'organització i accés a la memòria principal
- Descriure les característiques estructurals i funcionals més rellevants de les memòries RAM dinàmiques síncrones (SDRAM)
- Interpretar els paràmetres de temporització de les memòries síncrones SDRAM
- Comprendre la configuració dels mòduls de memòria
- Definir el concepte de mapa de memòria i introduir el disseny dels sistemes de descodificació
- Entendre les funcions del controlador de memòria dinàmica

Continguts

- 1. El sistema de memòria del computador
- 2. Característiques bàsiques de les memòries i mesura de prestacions
- 3. Memòria principal
- 4. Memòries RAM dinàmiques
- 5. Mòduls de memòria
- 6. Mapes de memòria
- 7. El controlador de memòria

Bibliografia

- Stallings, W.
 - Organización y arquitectura de computadores. 7a edició, Prentice Hall, 2006
 - Cap 5 (5.3)
- Hamacher, V.C., Vranesic, Z.G., Zaky, S.G.
 - Organización de computadores. 5a edició. McGraw Hill, 2002.
 - Cap 5 (5.2.4, 5.2.5, 5.2.6, 5.6.1)
- Notes en PoliformaT sobre disseny de mòduls de memòria
- Pàgines web:
 - www.micron.com (xips)
 - www.kingston.com (mòduls)
 - www.tomshardware.com

El sistema de memòria del computador

- Què és i on està la memòria?
- Concepte de jerarquia de memòria

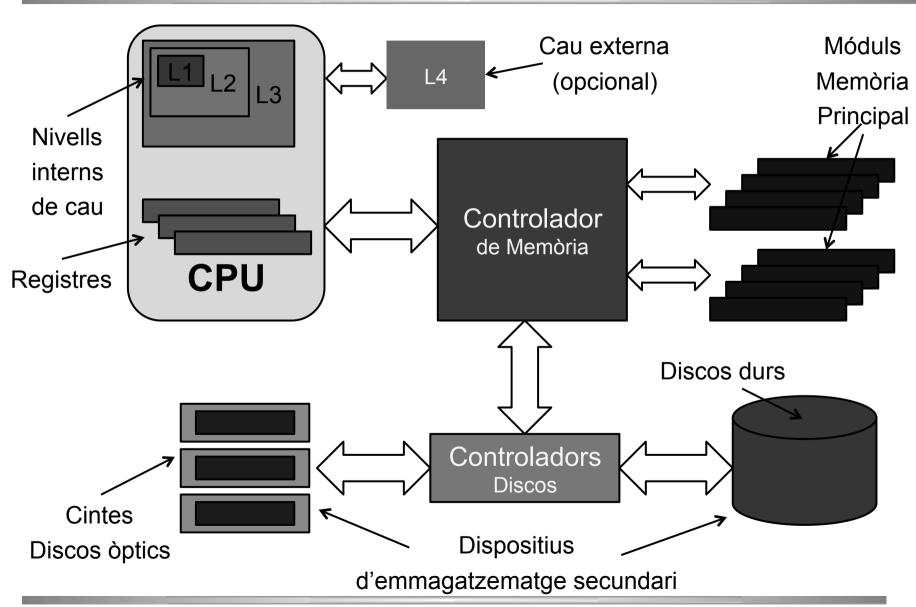
¿Què és i on està la memòria?

- Medi físic capaç d'emmagatzemar informació de forma temporal o permanent
 - Tipus d'operació: lectura i/o escriptura
- ¿On està la memòria en un computador?
 - Registres del processador
 - Memòria cache
 - Memòria principal
 - Memòria secundària
 - Discos
 - Cintes

Tecnologies diferents Fonaments físics diferents Localització en diversos llocs

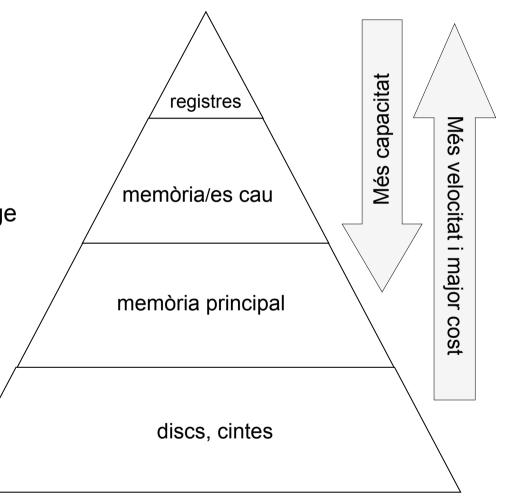
 Objectiu: dissenyar una memòria amb capacitat elevada i temps d'accés petit

Sistema de memòria d'un computador



Jerarquia de memòria: què és?

- Situació ideal: disposar de memòria ràpida i barata de gran capacitat
- Criteris tecnològics contraposats
 - Velocitat d'accés
 - Capacitat d'emmagatzematge
 - Cost per bit emmagatzemat
 - Consum d'energia
 - Fiabilitat
- Solució
 - Organització jeràrquica



2. Característiques bàsiques de les memòries

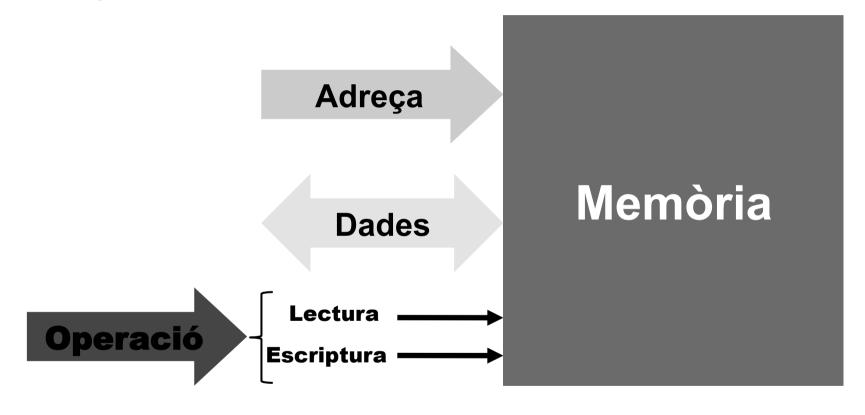
- Modes d'accés
- Capacitat d'emmagatzematge
- Temps d'accés i de cicle
- Amplada de banda

Paràmetres característics de les memòries

- Mode d'accés
 - Com accedim a la informació?
- Capacitat
 - Quanta informació hi cap?
- Velocitat de transferència o amplada de banda
 - Com de ràpid es transfereix la informació?

Modes d'accés

Operació bàsica d'accés a memòria



 El byte es la unitat mínima d'informació que pot ser accedida/adreçada per la CPU

Informació accedida en un accés

Paraula

- Unitat màxima de transferència en un accés
- La longitut en bytes sol ser una potència entera de 2 (1, 2, 4, 8, 16, ...)
- Exemple
 - Transferències entre memòria principal i processador sense memòria cache

Bloc

- Conjunt de paraules a què s'accedeix per mitjà d'una petició
- Exemples
 - Transferències entre discos i memòria principal
 - Transferències entre memòria cache i memòria principal

Modes d'accés

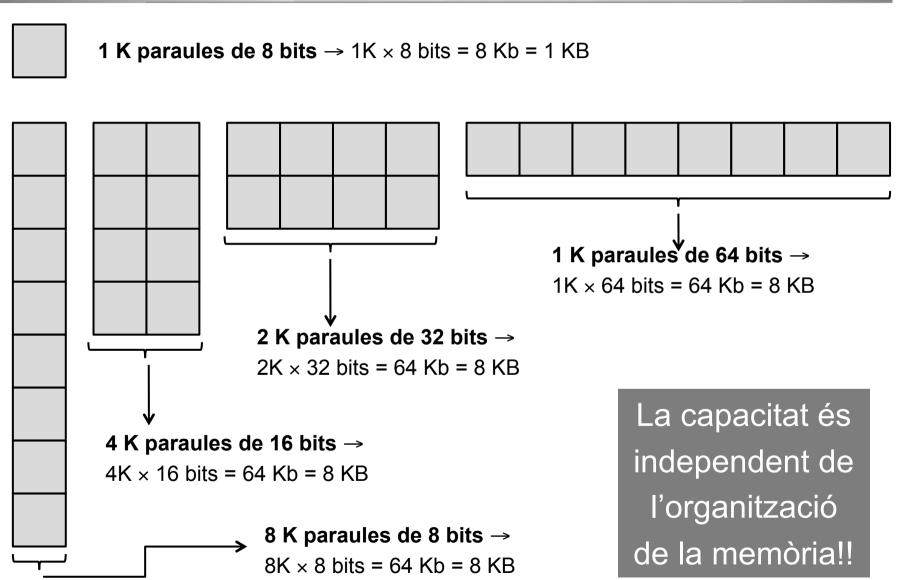
- Aleatori (random access)
 - Temps d'accés independent de la posició de la informació
 - Exemples: memòries principal i cache, memòries ROM
- Sequencial (sequential access)
 - Temps d'accés proporcional a la distància de la informació
 - Exemple: cintes magnètiques
- Directe (direct access)
 - Temps d'accés proporcional a la distància física entre la informació i el lector (capçal)
 - Hi ha dues components: l'una directa i l'altra seqüencial
 - Exemple: discos magnètics
- Accés associatiu (content-addressable memory; CAM)
 - Memòries adreçables per contingut
 - Duen un major temps d'accés per la necesitat de comparació, que és independent de la posició
 - Exemple: TLB, directori memories cache

Capacitat d'emmagatzematge

- Quantitat d'informació emmagatzemada: bits o bytes
 - Nomenclatura: B = 1 byte, b = 1 bit
- Com s'expressa la capacitat?
 - Capacitat total: en bytes o múltiples de byte
 - Nombre de paraules x Tamany de paraula (bits/bytes)
- Prefixos
 - Segons el context, són del tipus 2ⁿ o del tipus 10ⁿ
 - Exemple: la capacitat de la memòria principal sempre s'expressa en unitats del tipus 2ⁿ

Giga Tera	(K) 2 ¹⁰ ya (M) 2 ²⁰ a (G) 2 ³⁰	Valor (10 ⁿ) 10 ³ 10 ⁶ 10 ⁹ 10 ¹² 10 ¹⁵
--------------	--	--

Capacitat d'emmagatzematge i organització



Exemples de capacitat d'emmagatzematge

- Capacitat total expressada en bytes
 - $1024 \text{ bytes} = 2^{10} \text{ bytes} = 1 \text{ KB}$
- La memòria té 128K posicions de 16 bits cadascuna
 - $128K \times 16 \text{ bits} = 128K \times 2^4 \text{ bits} = 128K \times 2^1 \text{ bytes} = 256 \text{ KB}$
- La memòria té 8 MB en paraules de 32 bits
 - $8 MB = 2^{3} \times 2^{20} \times 2^{3} \text{ bits} = 2^{1} \times 2^{20} \times 2^{5} \text{ bits} = 2M \times 32 \text{ bits}$
- D'altres exemples
 - $64 \text{ Kbits} = 64 \text{K} \times 1 \text{ bits} = 2^{16} \text{ bits} = 2^{13} \times 2^{3} \text{ bits} = 2^{13} \text{ bytes}$
 - $-256 \text{ Mbits} = 28 \times 2^{20} \text{ bits} = 25 \times 2^{20} \times 2^{3} \text{ bits} = 32 \text{ MB}$

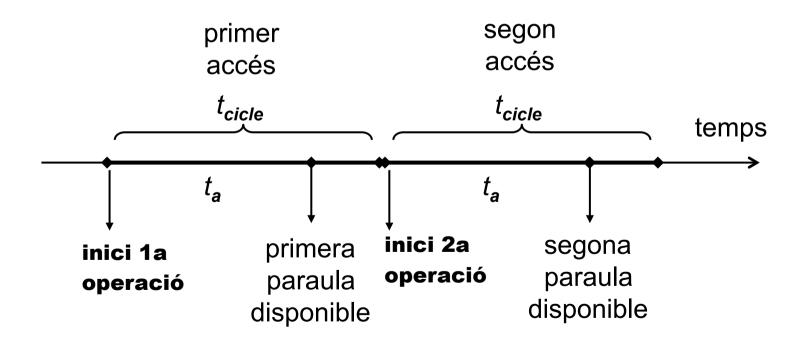
Paràmetres temporals

- Temps d'accés
 - Temps màxim entre l'inici de l'operació i l'obtenció o emmagatzematge de:
 - la paraula
 - la primera dada del bloc
 - Es mesura en unitats de temps

a les unitats de temps: m na	lom Valor nil·li (m) 10 ⁻³ nicro (μ) 10 ⁻⁶ ano (n) 10 ⁻⁹ ico (p) 10 ⁻¹²	
--------------------------------	---	--

Temporització d'un accés a paraula

- Temps de cicle
 - Temps mínim entre dues operacions consecutives de memòria
 - En general, T_{cicle} ≥ T_{accés}
- Exemple: dos accessos de lectura a paraula



Velocitat de transferència

- Amplada de banda (Bandwidth)
 - Genèricament: quantitat d'informació transferida per unitat de temps

$$B = \frac{\text{bytes/bits transferits}}{\text{temps}}$$

- Mesurada en bytes (o bits) per unitat de temps: MB/s, GB/s
 (MBps, GBps), Mb/s, Gb/s (Mbps, Gbps)
 - Els prefixos són sempre del tipus 10ⁿ
- Exemple amb accés a paraula
 - Temps de cicle: 20 ns
 - Temps d'accés: 15 ns
 - En cada accés es transmet una paraula de 32 bits

$$B = \frac{4 \text{ bytes}}{T_{cicle}} = \frac{4 \text{ bytes}}{20 \times 10^{-9} \text{ s}} = 2 \times 10^8 \frac{\text{bytes}}{\text{s}} = 200 \text{ MB/s}$$

Velocitat de transferència

- Amplada de banda (bandwidth)
 - Un altra forma alternativa d'interpretat l'amplada de banda

 $B = bytes transferits en un cicle \times f reqüència transmissió$

- Els bytes transferits en un cicle corresponen al tamany del bus de dades de la memòria
- La freqüència de transmissió correspón a l'invers del temps de cicle de la memòria

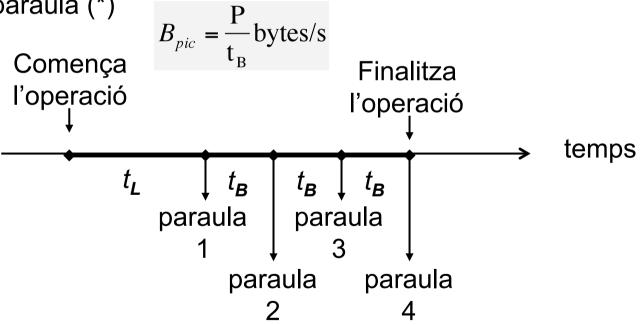
$$B = \text{Grandària_bus_dades} \times \left(\frac{1}{T_{ciclo}}\right) = 4 \text{bytes} \times \left(\frac{1}{20 \times 10^{-9} \text{s}}\right) =$$

$$= 4 \text{bytes} \times 50 \times 10^6 \text{ Hz} = 4 \text{bytes} \times 50 \text{MHz} = 200 \text{MB/s}$$

 L'amplada de banda també sol associarse al terme de productivitat del sistema de memòria

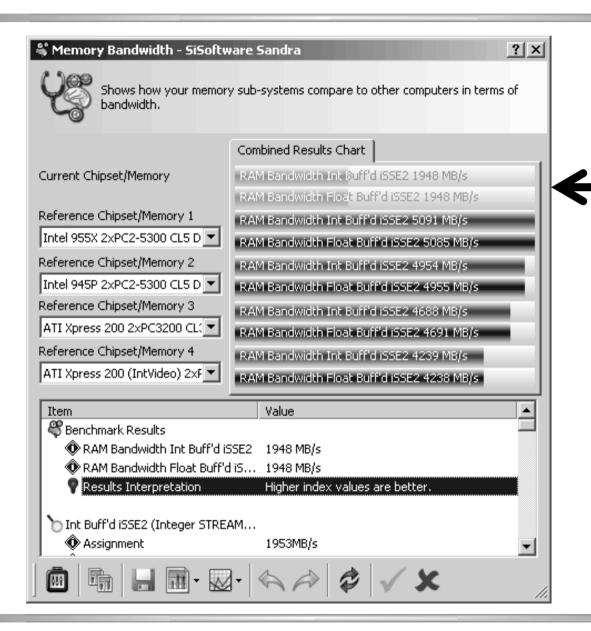
Temporització d'un accés a bloc

- Temps d'accés i amplada de banda
 - En accedir a un bloc, cal dedicar un temps d'accés o latència t_L
 fins que es transfereix la primera paraula de P bytes
 - La resta de paraules del bloc es transfereixen a intervals de t_B
 - L'amplada de banda (màxim o pic) es mesura a partir de la primera paraula (*)



(*) S'assumeix que és possible concatenar l'accés a blocs solapant el final del bloc actual amb el temps d'accés del següent bloc

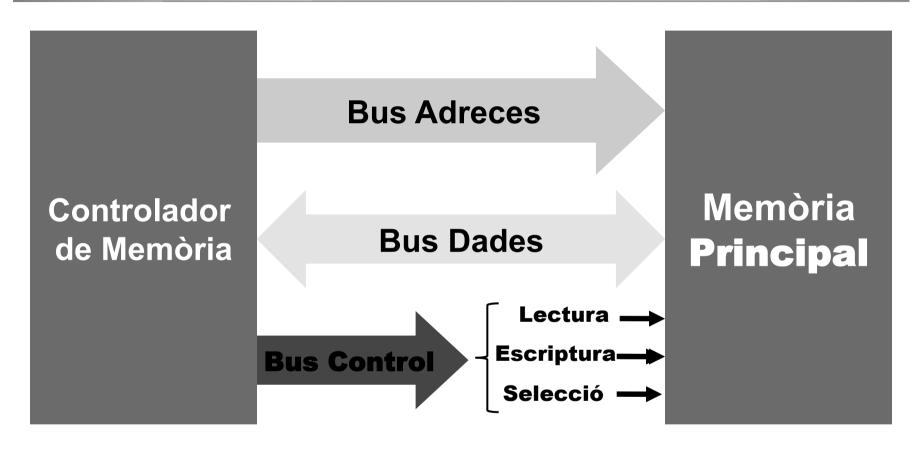
SiSoftware Sandra: amplada de banda



3. La memòria principal

- Interfície externa de memòria
- Adreçament de la memòria principal
- Formats d'emmagatzematge
- Ubicació de dades en el MIPS R2000
- Organització lògica
- Accés a memòria en el MIPS R2000
- Accés a memòria: estructura de busos

Interfície externa de memòria



 El controlador de memòria adapta els busos d'adreces/ dades/control de la CPU a les necessitats de connexió de la memòria principal

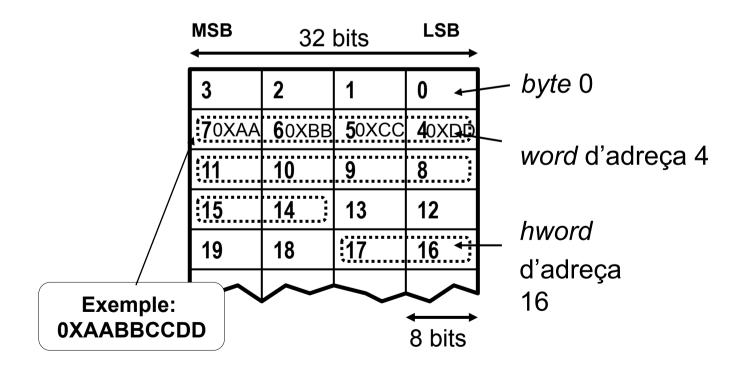
Adreçament de la memòria principal

- Adreces
 - Van referides a cada byte
 - Cada byte té una adreça única
- Paràmetres (definits per la CPU)
 - Amplada de paraula ← bus de dades
 - En bits o bytes
 - Adreçament ← bus d'adreces
 - En bits/línies d'adreça
 - En capacitat d'adreçament (unitats 2ⁿ)
 - Format emmagatzematge paraules
 - Little-endian / big-endian
- Exemple d'espai d'adreçament: MIPS R2000
 - 32 bits de dades (D₃₁ fins D₀)
 - 32 bits d'adreça (A₃₁ fins A₀)
 - Pot referenciar 2³² bytes, això és, 4 GB

Adreçament (little-endian)

MIPS R2000

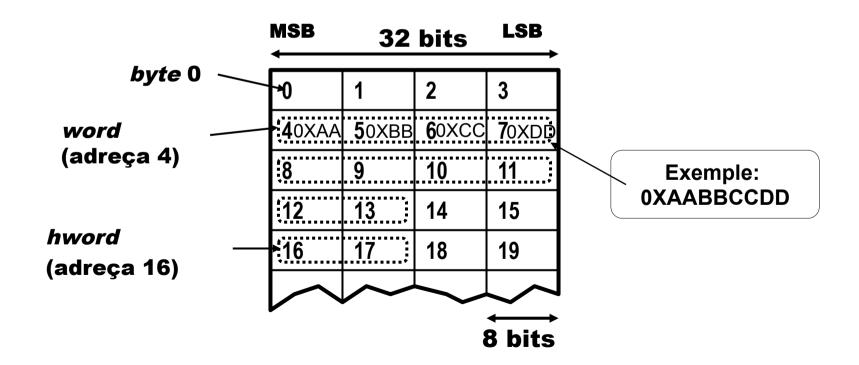
- Paraules (words) de 32 bits en adreça múltiple de 4
- Mitges paraules (hwords) de 16 bits en adreça parella
- Bytes en adreça qualsevol



Adreçament big-endian

Motorola 68000

- Paraules (words) de 32 bits en adreces multiples de 4
- Mitges paraules (hword) de 16 bits en adreces parelles
- Bytes en qualsevol adreça



Ubicació de dades en el MIPS R2000

Directives

- Octets: .byte

– Mitges paraules: .half

Paraules senceres: .word

Cadenes de caràcters:

.ascii .asciiz

Alineament automàtic

0x98	0x53	0x41	0x69
0x87	0x33	0x00	0x4B
0x72	0x65	0x50	0x54
0x00	0x00	0x00	0x65
0x12	0x34	0x56	0x78
		0x00	0x00

.data 0x1000A000

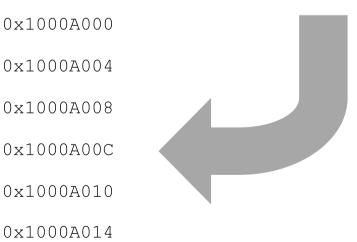
dni: .word 0x98534169

lletra: .asciiz "K"
codbar: .half 0x8733
data: .byte 0x54

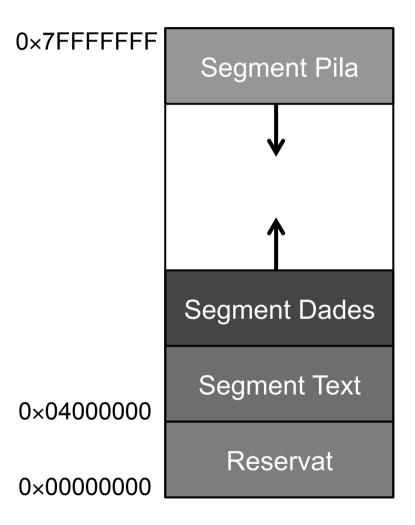
nom: .ascii "Pere"

.word 0x12345678

.space 2



Organització lògica de la memòria



- La memòria no és un simple espai d'adreçament pla
- Distintes zones o segments des del punt de vista del programari (compilador/SO)
- Els segments estableixen les àrees de memòria on ubicar els distints tipus d'informació

Tipus d'accessos a memòria en el MIPS R2000

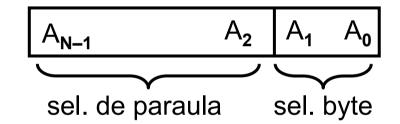
- La memòria principal es l'única visible pel programador (amb les instruccions de load i store)
- Lectura (load)
 - Octets: 1b \$3,0(\$6)
 - Mitges paraules (16 bits): 1h \$3,0 (\$6)
 - Paraules senceres (32 bits): 1w \$3,0 (\$6)
- Escriptura (store)
 - Octets: sb \$3,0(\$6)
 - Mitges paraules (16 bits): sh \$3,0 (\$6)
 - Paraules senceres (32 bits): sw \$3,0 (\$6)
- El tipus d'accés depén de:
 - Codi d'operació (codi de la instrucció màquina)
 - Valor de l'adreça (parella, senar, múltiple de 2, de 4, etc.)

Bits d'adreça i línies d'adreça en el bus

- Per tal d'accedir a una paraula, cal fer-ho al conjunt de bytes que la componen
- Els bits de l'adreça utilitzats pel processador estan estructurats:

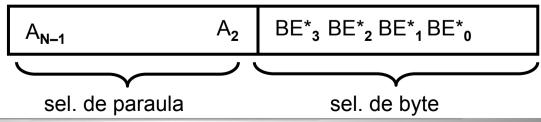
exemple amb W=32

Bus lògic d'adreces



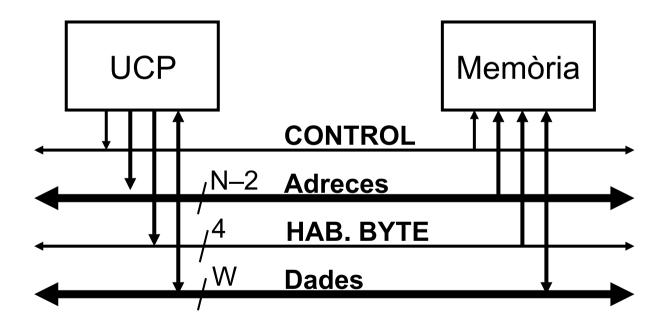
- En el bus, les línies d'adreça sempre seleccionen paraules
- La selecció dels bytes implicats en l'accés es fa per mitjà de línies d'habilitació o màscara de byte (senyals /BE_i) emeses pel processador

Bus físic d'adreces



Estructura del bus de memòria

- Exemple amb W=32 i N bits d'adreça
 - N–2 línies físiques d'adreça A_{N-1} ... A₂
 - 4 línies d'habilitació de byte BE₀* ... BE₃*
 - W línies de dades D₃₁ ... D₀
 - Línies de control



Generació dels senyals /BE_i

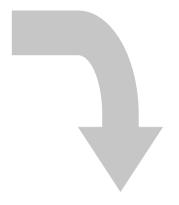
UCP de 32 bits

 El circuit de control del processador, segons el tipus d'accés i els bits menys significatius de l'adreça, activa els senyals /BE_i

Accés	A_1	A_0	BE ₃ *	BE ₂ *	BE ₁ *	BE ₀ *
a byte	0	0	1	1	1	0
	J 0	1	1	1	0	1
	1	0	1	0	1	1
	<u>l</u> 1	1	0	1	1	1
a hword	∫o	0	1	1	0	0
	1	0	0	0	1	1
a word	0	0	0	0	0	0
		ر کے				
		V		`	Y	
No estan en e			I Senyals presents			
bus d'adreces			-			

Exemple d'activacions dels senyals /BE_i

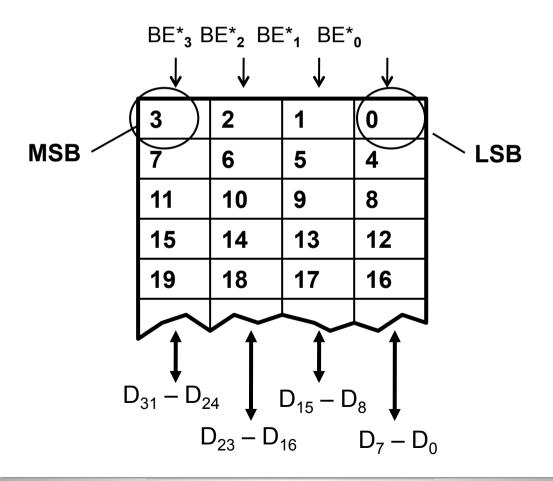
```
li $t0, 0x10000000
lw $a0, 0($t0)
lb $a1, 5($t0)
sh $a2, 6($t0)
sb $a3, 8($t0)
```



	_A ₁	A_0	BE ₃ *	BE ₂ *	BE ₁ *	BE ₀ *
lw	0	0	0	0	0	0
lb	0	1	1	1	0	1
sh	1	0	0	0	1	1
sb	0	0	1	1	1	0

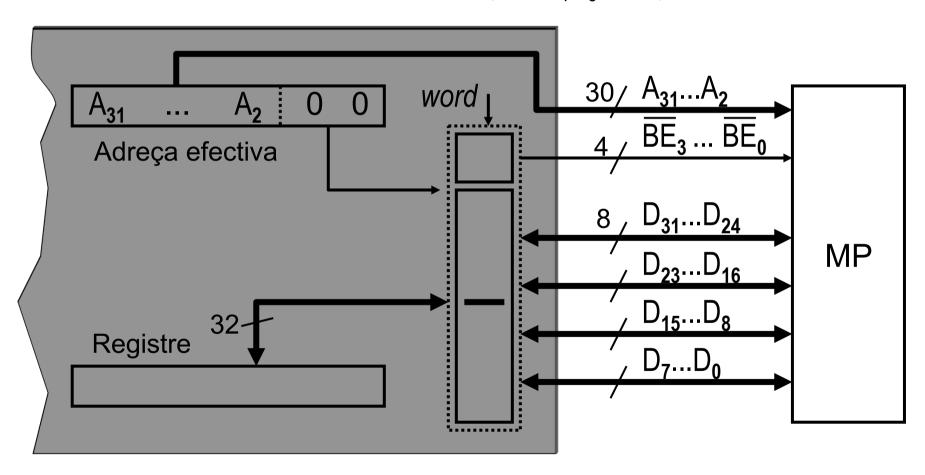
Bus d'adreces i línies de selecció de byte

 El format d'emmagatzematge determina l'associació de línies BE*_i a bytes del bus de dades



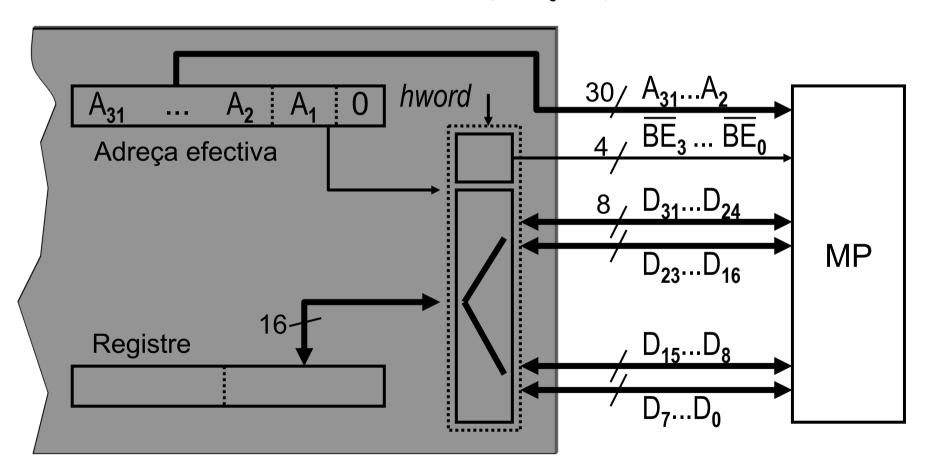
Accés a una paraula de 32 bits

- UCP de 32 bits
 - Adreça efectiva múltiple de 4 (bits $A_1A_0 = 00$)



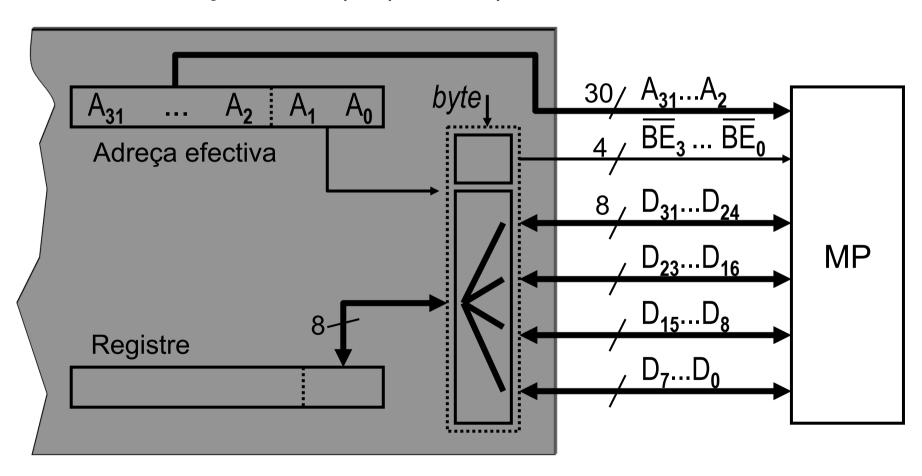
Accés a mitja paraula (16 bits)

- UCP de 32 bits
 - Adreça efectiva múltiple de 2 (bit $A_0 = 0$)



Accés a un byte (8 bits)

- UCP de 32 bits
 - L'adreça efectiva pot prendre qualsevol valor



4. Memòries RAM dinàmiques

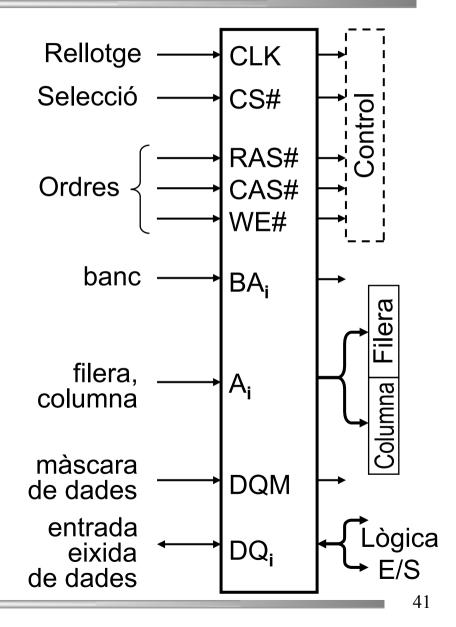
- Organització
- Adreçament
- Temporització
- Xips SDRAM i tendències tecnològiques

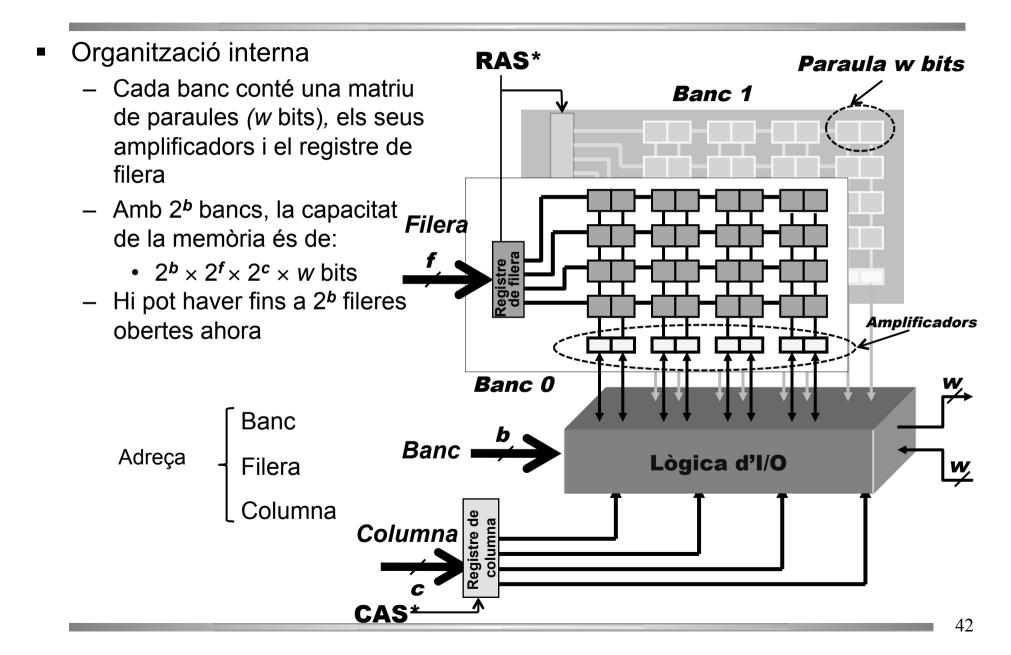
Memòries RAM dinàmiques

- Per què RAM dinàmica en la memòria principal?
 - Alta integració
 - La cel·la de memòria dinàmica (transistor+condensador) ocupa molt poc espai de xip (menor que la cel.la de la SRAM; biestable)
 - Al mateix cost de fabricació, més capacitat que SRAM
 - Menor cost per bit
 - Baix consum en relació a la SRAM
 - Amb les memòries DRAM síncrones (SDRAM) actuals es consegueix una amplada de banda major en els accésos a blocs
 - S'adapta als processadors amb memòries cache
- Inconvenients
 - Són mes lentes que les SRAM
 - Requereixen d'un refresc periòdic per evitar la pèrdua d'informació

Interfície del xip

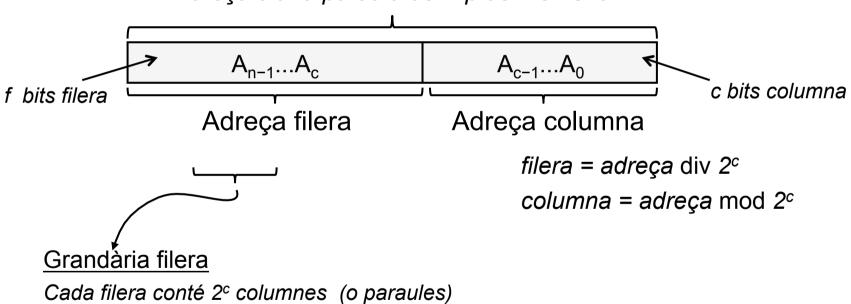
- Ordres: són combinacions de tres senyals (RAS*, CAS* y WE*)
- Banc: selecciona un dels bancs interns de la SDRAM
- Adreça: segons l'ordre, aprofiten per a especificar una filera, una columna o d'altres coses. El nombre de bits de fila sol determinar el nombre de línies
- Dades: tantes línies com l'ample de paraula
- Màscara de dades: útil per a seleccionar octets dins la paraula





- Adreces estructurades*
 - En una matru de 2^{f+c} paraules organitzades en 2^f fileres i 2^c columnes, l'adreça de n = f + c bits queda

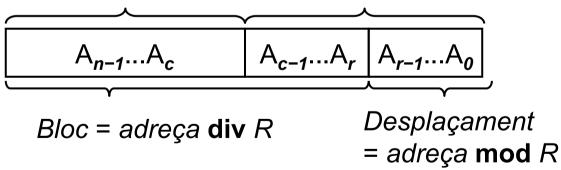
Adreça d'una paraula del xip de memòria



(*) Per simplificació, s'assumeix un únic banc o matriu SDRAM

Blocs de memòria

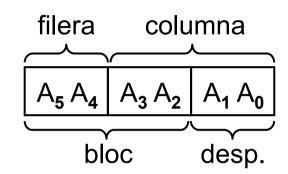
- És la unitat més frequent de comunicació entre el processador amb memòria cache i la memòria principal
- Un bloc està format per $R = 2^r$ paraules (típicament 4 o 8) consecutives amb adreça de la forma $A_{n-1}...A_rX...X$, encara que podria ser pogramat entre 1 i la grandària de la filera
- La longitud R de bloc és una constant per a un computador donat
- Dins d'un xip de DRAM, l'adreça queda estructurada de la manera següent: filera columna

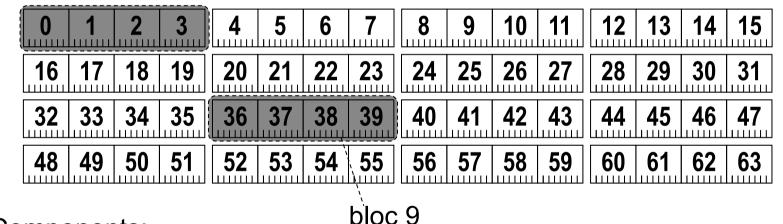


Noteu que en una filera hi ha 2^{c-r} blocs i que cap bloc es troba repartit entre dues fileres

- Exemple de matriu
 - DRAM 64x8 bits
 - 4 fileres de 16 paraules
 - Blocs de 4 paraules

bloc 0





Components:

16 x 8 amplificadors

Registre de filera de 2 bits + descodificador de 2 a 4

Registre de columna de 4 bits + descodificador de 4 a 16

Accés a blocs de memòria

- Les ordres de lectura i escriptura indiquen l'adreça d'una paraula, però el xip opera amb les R paraules del bloc que la conté
- La paraula adreçada és la primera transmesa, seguida de la resta del bloc en forma de r\u00e0fega de dades (burst)
- La seqüència de paraules accedides fins completar el bloc depén de l'adreça especificada en l'ordre

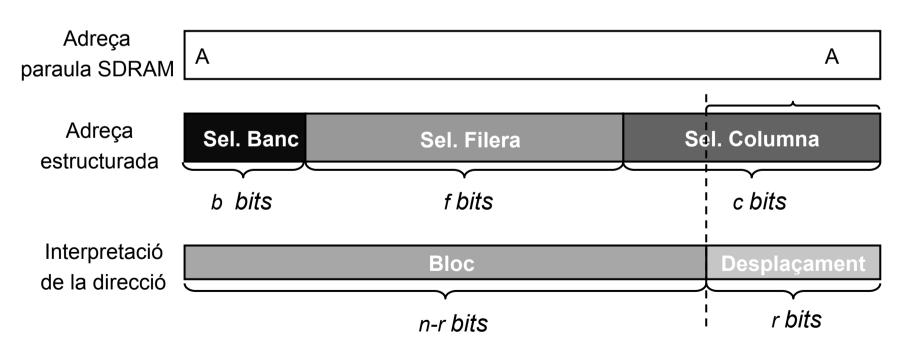
Exemple amb tamany bloc = 4



Adreça	Seqüència d'accesos		
0	0-1-2-3		
1	1-2-3-0		
2	2-3-0-1		
3	3-0-1-2		

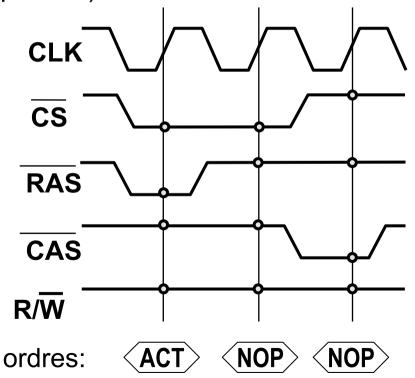
Adreçament complet SDRAM

- Grandària SDRAM: 2ⁿ × w bits
- Organització: 2^b bancs × 2^f fileres × 2^c columnes
- Format adreça: n = b + f + c
- Blocs de 2^r paraules



Emissió d'ordres

- Els flancs ascendents del senyal de rellotge marquen els instants en que el xip accepta ordres
- Si el xip està deshabilitat per CS#, l'ordre interpretada és NOP (cap operació)



Ordres més importants:

ACTIVATE

READ

WRITE

PRECHARGE

NOP (també quan CS*=1)

Activació

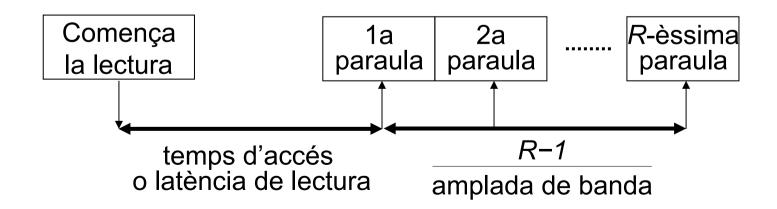
- Carrega el registre de fila, desencadenat per l'activació del senyal /RAS (Row Address Strobe). La filera seleccionada queda "oberta" i els amplificadors s'apliquen sobre la filera
- Després de l'activació, la filera completa queda retesa per amplificadors, i es pot accedir a diferents columnes repetint les operacions de lectura o escriptura sobre la filera oberta
- Operació: lectura o escriptura
 - El senyal /CAS (Column Address Strobe) carrega el registre de columna i el senyal R/W* activa la lògica d'E/E a la columna seleccionada

Precàrrega

 La precàrrega allibera (o tanca) la fila activa i els amplificadors queden en repós. Ja no es pot llegir ni escriure sobre la filera fins després de la propera activació

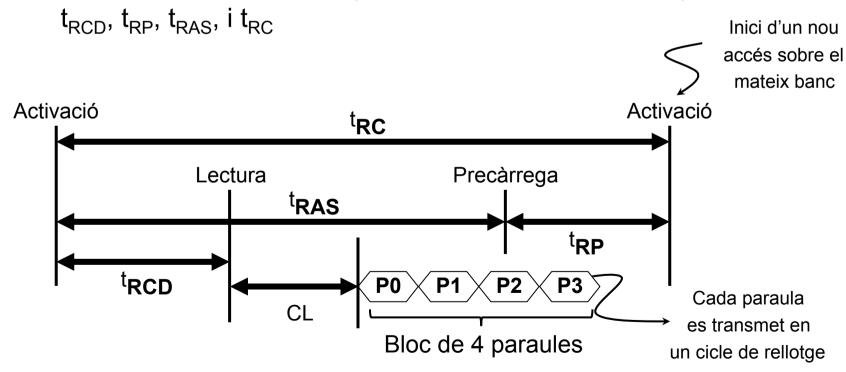
(*) Cadascuna de les ordres anteriors van lligades a un banc determinat.

- Accés a blocs de memòria: Temporització
 - Les lectures i escriptures afecten a les R paraules d'un bloc
 - Una volta la filera que conté el bloc està activa, la transferència és ràpida
 - El temps necessari per a llegir un bloc té dues components:
 - La latència o temps d'accés a la primera paraula, independent de R
 - El temps de transmissió de les paraules restants, que depén de R i de la velocitat de la lògica d'E/S del xip o amplada de banda

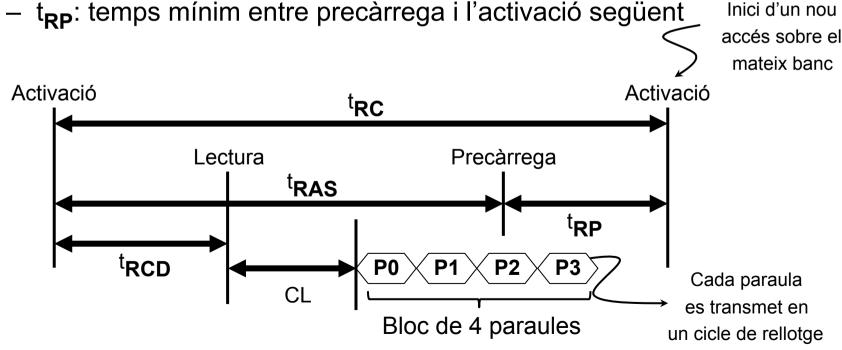


- Accés a blocs de memòria
 - La seqüència d'ordres en un accés té la forma:
 - ACTIVE READ PRECHARGE
 - ACTIVE WRITE PRECHARGE

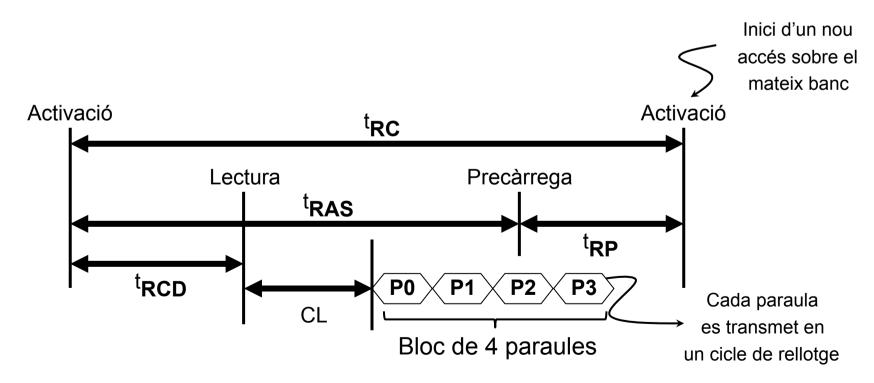
on cal inserir ordres NOP per satisfer restriccions temporals, com



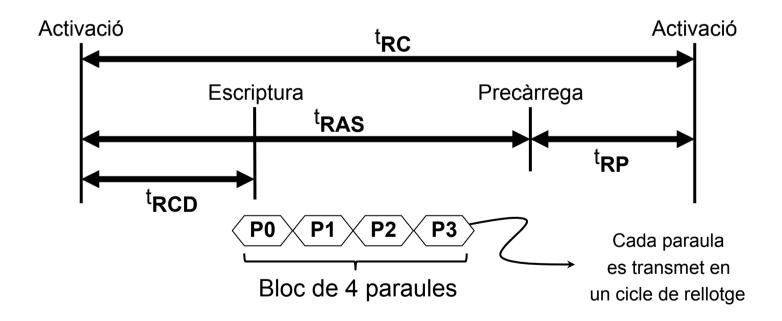
- Paràmetres temporals importants
 - t_{RCD}: temps mínim des de l'activació d'una filera fins a l'inici de l'operació. Contribueix al temps d'accés
 - t_{RAS}: temps mínim entre l'activació i la precàrrega d'una filera
 - t_{RC}: temps mínim entre dues activacions. S'entén com a temps de cicle



- Latència de CAS
 - t_{CL}: latència de CAS o temps màxim entre l'inici de la lectura i l'eixida de dades vàlides
 - Temps d'accés a les dades ≥ t_{RCD} + t_{CL}



- Operació d'escriptura
 - La transferència d'escriptura comença en el mateix cicle en el qual es dóna l'ordre WRITE



Restriccions temporals

- Un xip SDRAM té una frequència màxima (o un període de rellotge mínim) de funcionament. L'usuari pot fer servir aquesta frequència o altra inferior
- Per a una freqüència de rellotge donada el controlador haurà d'inserir cicles NOP entre les ordres ACTIVE, READ/WRITE i PRECHARGE d'una operació i també entre operacions successives per tal de satisfer les restriccions temporals de la matriu DRAM
- Exemple:

	50 MHz	100 MHz	166 MHz
	$(t_{CK} = 20 \text{ ns})$	$(t_{CK} = 10 \text{ ns})$	$(t_{CK} = 6 \text{ ns})$
t _{RCD} = 18 ns	1	2	3
t _{RAS} = 42 ns	3	5	7
t _{RC} = 60 ns	3	6	10
t _{RP} = 18 ns	1	2	3

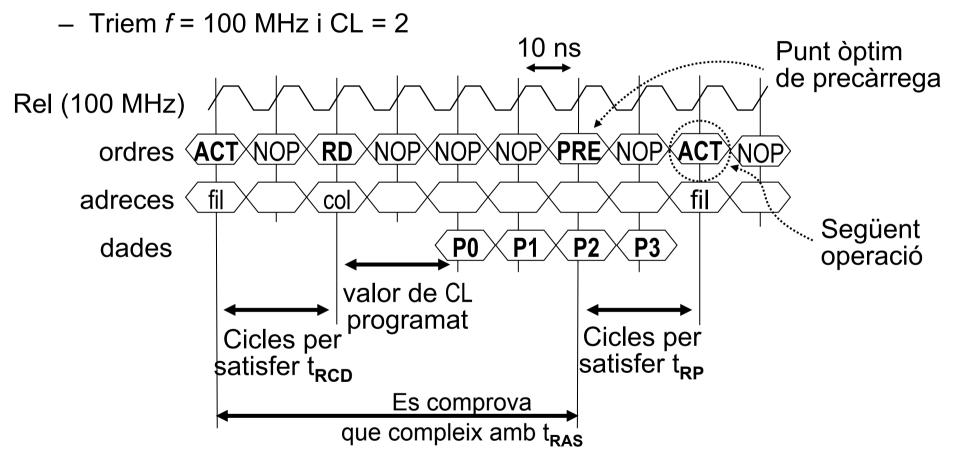
- Freqüència de rellotge i latència de CAS
 - La latència de CAS (CL) és un valor que s'escriu en el registre de mode del xip durant l'inicialització de la memòria
 - Per a una freqüència donada, el xip té un CL mínim
 - o al revés: per a fer servir un valor CL donat, el xip té una freqüència màxima (o un període mínim)

PARAMETER		SYMBOL	MIN	MAX	UNITS
Clock cycle time	CL = 3	^t CK (3)	6		ns
	CL = 2	^t CK (2)	10		ns
	CL = 1	^t CK (1)	20		ns
ACTIVE to PRECHARGE command		^t RAS	42	120k	ns
ACTIVE to ACTIVE command period		^t RC	60		ns
AUTO REFRESH period		^t RFC	60		ns
ACTIVE to READ or WRITE delay		^t RCD	18		ns
Refresh period (4,096 rows)		^t REF		64	ms
PRECHARGE command period		^t RP	18		ns
ACTIVE hank ato ACTIVE hank becommand		toon	12		nc

- Temporització de la transferència
 - El moment òptim d'enviament de l'ordre de PRECHARGE coincideix amb R⁽¹⁾ cicles després d'ordenar l'operació de READ⁽²⁾ o WRITE
 - Les ordres READ i WRITE tenen un mode automàtic que provoca la precàrrega en el moment òptim i estalvia l'enviament d'una ordre PRECHARGE explícita.

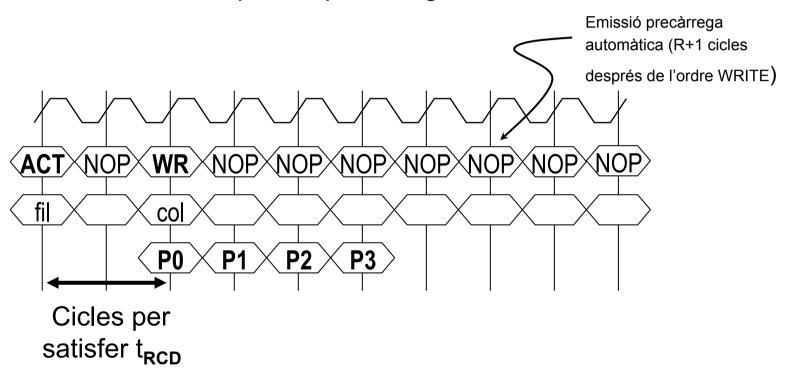
- (1) R representa la grandària del bloc (nombre de paraules que s'accedeixen)
- (2) En el cas de lectura, és equivalent a enviar l'ordre PRECHARGE CL-1 cicles abans de transferir-se la darrera paraula del bloc

Cronograma de lectura: exemple

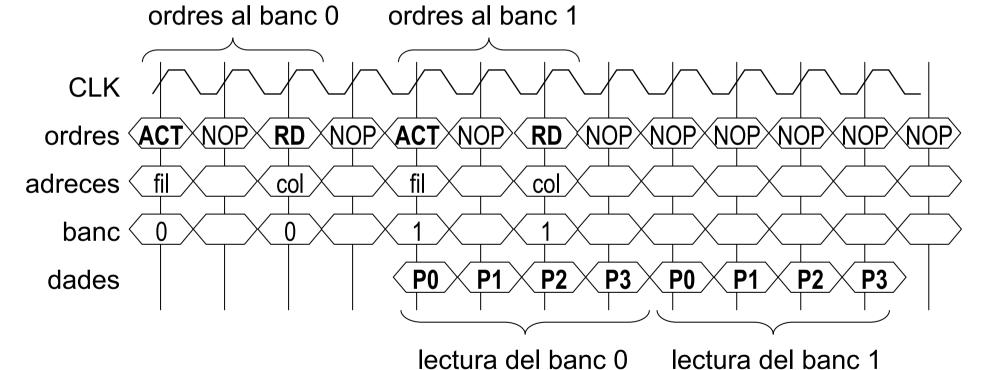


Latència de lectura resultant = $4 \times 10 = 40$ ns Amplada de banda = $100 \cdot 10^6 \times w$ bits/s

- Cronograma d'escriptura
 - A 100 MHz. Suposem precàrrega automàtica

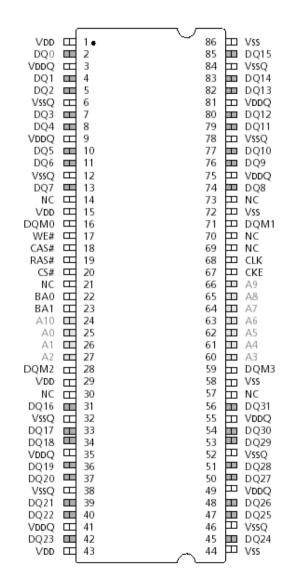


- Cronograma de lectura amb bancs
 - Es poden donar ordres a bancs independents
 - Pipelining: les dades ixen en el mateix ordre amb el qual es demanen, amb una latència constant

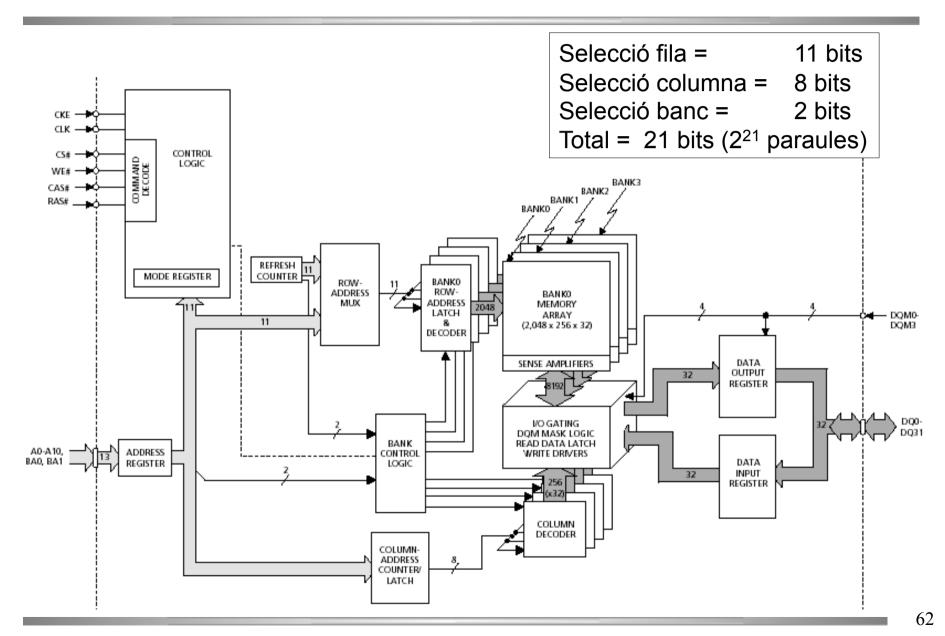


Xip comercial SDRAM 2Mx32 bits (64 Mbits)

- Model: MT49LC2M32B2 (www.micron.com)
- Encapsulat en plàstic amb 86 potetes
- Freqüències de rellotge i temps de cicle
 - 143, 166, 183 i 200 MHz
 - 7 ns, 6 ns, 5.5 ns, 5 ns, respectivament
- Latència de CAS programable
 - 143, i 166 MHz: 1, 2 ó 3 cicles
 - 183 i 200 MHz: 3 cicles
- Ràfegues programables
 - 1, 2, 4, 8 o 256 paraules de 32 bits
- Organització en bancs
 - 4 bancs de 2048×256×32 bits
 - $2Mx32 = 2^{26} = 4x2^{24} = 4 \times (2^{11} \times 2^8 \times 32)$
- Període de refresc: 64 ms (comptador: 11 bits)



Xip comercial SDRAM 2Mx32 bits (64 Mbits)



Tecnologies

- Memòries SDRAM (Synchronous DRAM): els xips tenen una interfície síncrona regulada per una ona de rellotge de freqüència característica: 100, 133, 200, etc. MHz
- Estàndards d'interés, mantinguts per JEDEC (Joint Electron Device Engineering Council)
 - SDRAM, la més antiga, de la qual deriven les tecnologies posteriors
 - DDR-SDRAM (*Double Data Rate* SDRAM), en versions successives: DDR, DDR2 i DDR3
- Les tecnologíes DDR, DDR2 i DDR3 multipliquen per 2, 4 i 8, respectivament, l'amplada de banda de la SDRAM
- Per a una tecnologia donada, l'amplada de banda depén linealment de la freqüència de rellotge

- Com explicar l'increment de l'amplada de banda?
 - JEDEC va proposar les tecnologíes DDR, DDR2 I DDR3 com una solució ràpida, simple i de baix cost per tal d'incrementar l'amplada de banda de les memòries sense incrementar la seva complexitat VLSI
 - Per aixó, es va proposar:
 - 1) que les memòries treballen amb dos rellotges, un intern per accedir a les matrius DRAM i l'altre extern per fer les transferències de dades
 - 2) utilitzar la tècnica de prebusca (PREFETCH)
 - La freqüència interna és la mateixa en SDRAM, DDR, DDR2 i DDR3 mentre qué l'externa és 2 i 4 voltes major que l'interna en DDR2 i DDR3, respectivament
 - S'accedeix simultaniament a 2, 4 o 8 paraules a la freqüència interna de rellotge i es transfereixen d'una en una al/des de l'exterior a una freqüència efectiva 2, 4 o 8 voltes major

Tendències

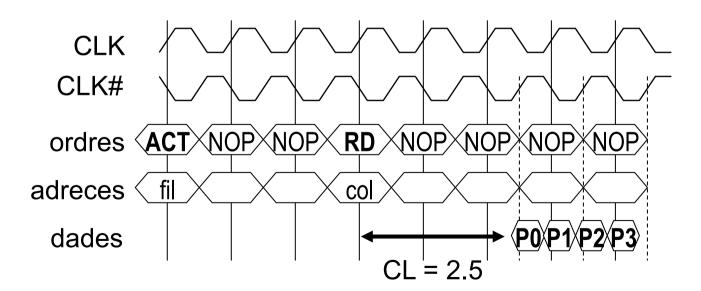
- Increment de la densitat d'integració
- Increment de l'amplada de banda apujant la freqüència del rellotge
- Baixada de la tensió d'alimentació. El calor generat (proporcional a V² i a f) limita la densitat i la freqüència dels circuits integrats

Any	Tecnologia	Densitats (bits)	Vdd (V)	Freqüències (MHz)
1997	SDRAM	64M-512 M	3.3	66–167
1999	DDR	128M-1G	2.5	133–200
2004	DDR2	256M-2G	1.8	200–400
2006	DDR3	1G-4G?	1.5	400-800?
2013	DDR4		1.2	

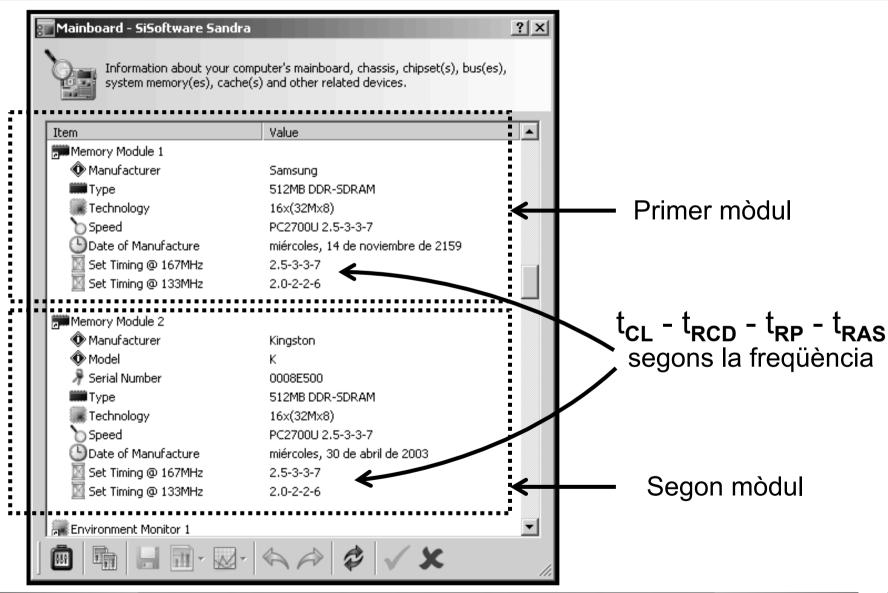
- Desenvolupament de la lògica de control per tal d'aprofitar l'amplada de banda
- Les latències de lectura DRAM baixen molt lentament al llarg del tempsaileg micron 2006

Memòries SDRAM DDR

- Interfície del xip
 - La velocitat de transmissió és el doble de la freqüència del rellotge
 - Es duplica la tasa efectiva de transferència de dades sense modificar la freqüència del rellotge mitjançant la prebusca interna de dos paraules al mateix temps
 - La latència de CAS és un múltiple de 0.5 cicles
 - Senyals afegits a la interfície SDRAM:
 - Un senyal CLK# invertit



SiSoftware Sandra: temporizació

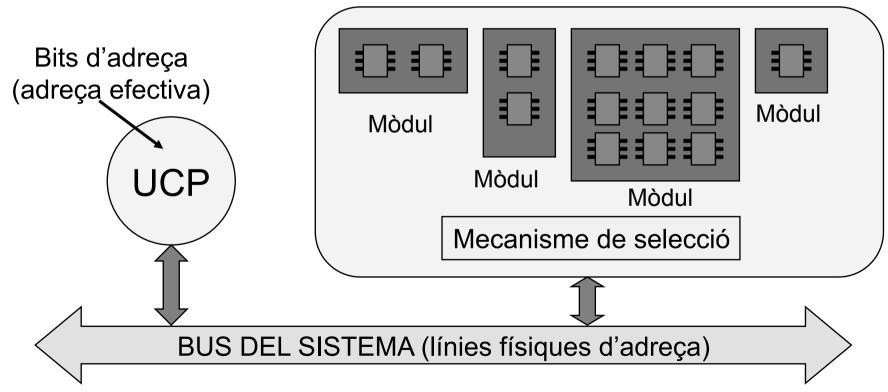


5. Mòduls de memòria

- Conceptes bàsics
- Organització interna dels mòduls
- Mòduls comercials de memòria dinàmica

Conceptes bàsics

- El sistema de memòria principal d'un computador sol estar format per un conjunt de mòduls
- El sistema de memòria i l'estructura del bus s'ajusten a una organització determinada pel processador

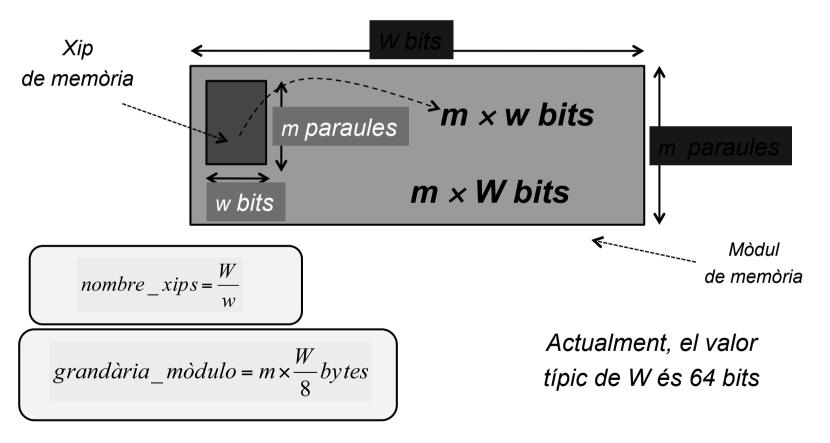


Conceptes bàsics

- Un mòdul de memòria és un conjunt de xips que satisfà uns determinats requeriments
 - Usualment, cada mòdul és d'un tipus de memòria determinat:
 ROM, SRAM, DRAM, etc.
- Cada mòdul, a través d'un mecanisme de selecció, suporta un conjunt d'adreces dins de l'espai d'adreçament global del processador
- El repartiment de les adreces per mòduls (conjunt d'adreces que pertany a cada mòdul) es denomina mapa de memòria

Organització de mòduls de memòria

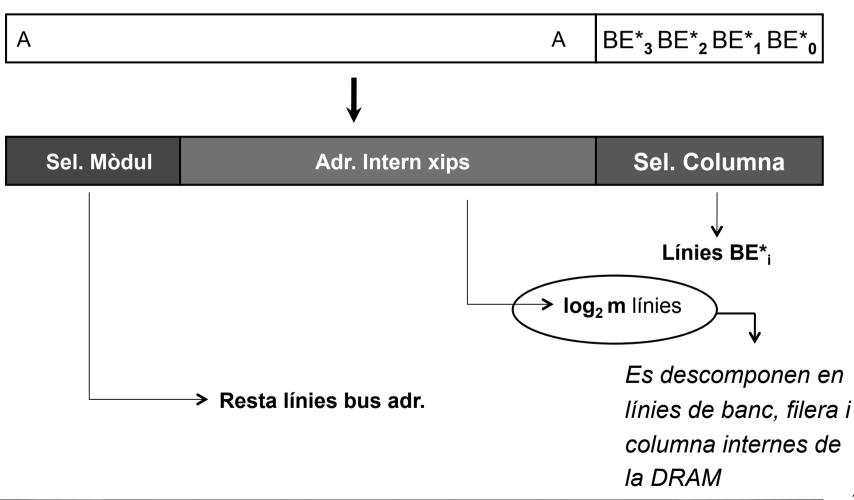
 Els xips DRAM d'un mòdul de memòria estan habitualment distribuïts al llarg d'una única filera. El nombre de xips del mòdul depén de la grandària de paraula (nombre de bits) dels xips (w) i de la grandària de paraula del mòdul (W)



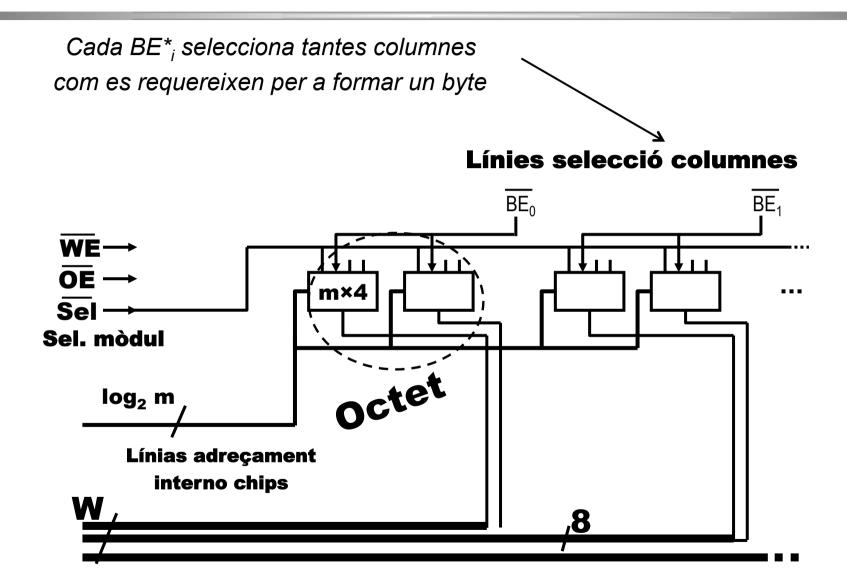
Organització de mòduls de memòria

Adreçament dels mòduls de memòria

Bus físic d'adreces (CPU 32 bits)



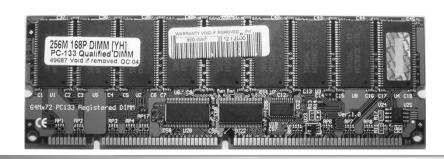
Organització de mòduls de memòria

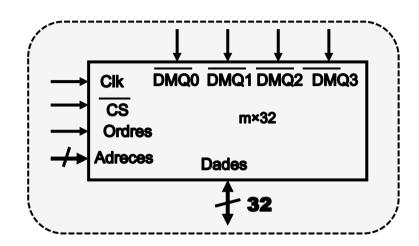


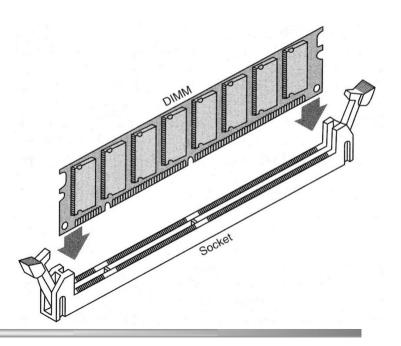
Mòduls estàndard de memòria dinàmica

Terminals

- Línies d'adreces multiplexades
- L'entrada CS* fa de línia de selecció de mòdul
- Les entrades DMQ*(CAS) són les línies d'habilitació de byte (només en escriptura)
- Els mòduls s'insereixen i extrauen per mitjà de les ranures (slots)
 - Opcionalment poden tenir un bit de paritat (informació redundant per cada byte per tal de detectar errors)

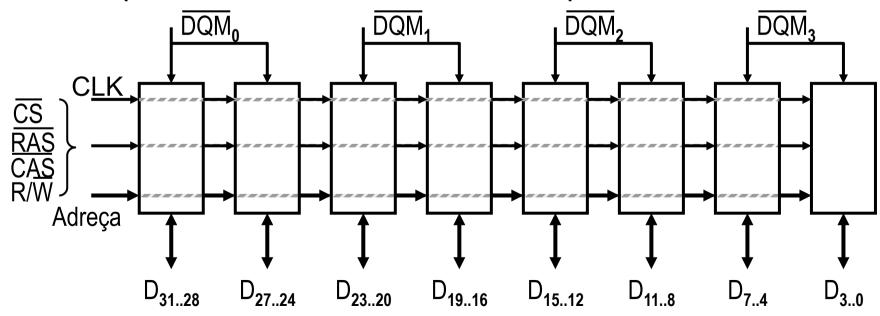






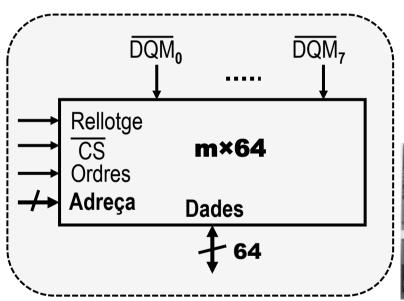
Organització del mòdul de memòria dinàmica

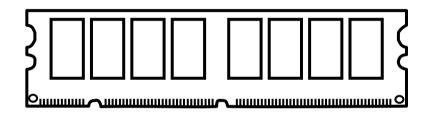
- Sol tenir una fila de xips idèntics
 - Les entrades de rellotge, selecció, ordres i adreça són comunes
 - Les entrades de màscara de dades DQM dels xips aprofiten per a seleccionar els bytes
 - L'amplada de banda del mòdul és la suma de les amplades de banda dels xips
- Exemple: mòdul de m×32 bits amb xips de m×4

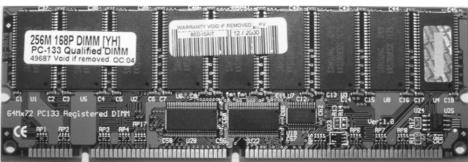


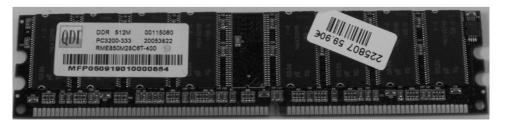
Mòduls DIMM (Dual Inline Memory Module)

 168/184/240 contactes, 64 bits de dades, 13 cm de longitud







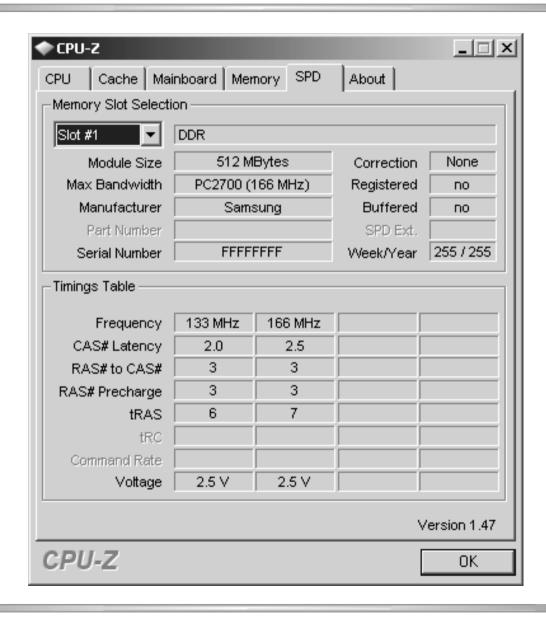


SPD (Serial Presence Detect)

- Memòria EEPROM de pocs bytes
- Emmgatzema informació sobre el mòdul de memòria
 - Temporització
 - Capacitat
 - Fabricant
 - Número de sèrie
- Permet la configuració automàtica del sistema de memòria
- Hi ha programes que poden accedir a aquesta informació
 - CPU-Z, Sisoft Sandra, entre d'altres



Informació continguda en el SPD



Comparativa de mòduls i denominacions

Empaquetament

SDRAM: DIMM 168 c

DDR SDRAM: DIMM 184 c

DDR2 SDRAM: DIMM 240 c

DDR3 SDRAM: DIMM 240 c

Amplades de banda i denominacions

SDRAM (100 MHz): 100 MHz x 8 bytes = 800 MB/s

Nomenclatura: PC100

- DDR SDRAM (100 MHz): 100 MHz x 8 bytes x 2 = 1600 MB/s

Nomenclatura: PC1600

DDR2 SDRAM (200 MHz): 200 MHz x 8 bytes x 2 = 3200 MB/s

Nomenclatura: PC2-3200

DDR3 SDRAM (400 MHz): 400 MHz x 8 bytes x 2 = 6400 MB/s

Nomenclatura: PC3-6400

Mòduls DDR estàndard

Nom	Freq.	Retard	F. rell. E/E	V. Transf.	Denom.	Taxa Transf. Màx.
DDR-200	100 MHz	10 ns	100 MHz	200 MT/s	PC1600	1.600 MB/s
DDR-266	133 MHz	7,5 ns	133 MHz	266 MT/s	PC2100	2.133 MB/s
DDR-300	150 MHz	6,6 ns	150 MHz	300 MT/s	PC2400	2.400 MB/s
DDR-333	166 MHz	6 ns	166 MHz	333 MT/s	PC2700	2.667 MB/s
DDR-366	183 MHz	5,5 ns	183 MHz	366 MT/s	PC3000	2.933 MiB/s
DDR-400	200 MHz	5 ns	200 MHz	400 MT/s	PC3200	3.200 MB/s
DDR-433	216 MHz	4,6 ns	216 MHz	433 MT/s	PC3500	3.500 MB/s
DDR-466	233 MHz	4,2 ns	233 MHz	466 MT/s	PC3700	3.700 MB/s
DDR-500	250 MHz	4 ns	250 MHz	500 MT/s	PC4000	4.000 MB/s
DDR-533	266 MHz	3,7 ns	266 MHz	533 MT/s	PC4300	4.264 MB/s

MT/s: Milions de transferències per segon

Mòduls DDR estàndard

Nom	Freq.	Retard	F. rell. E/E	V. Transf.	Denom.	Taxa Transf. Màx.
DDR2-400	100 MHz	10 ns	200 MHz	400 MT/s	PC2-3200	3.200 MB/s
DDR2-533	133 MHz	7,5 ns	266 MHz	533 MT/s	PC2-4300	4.264 MB/s
DDR2-600	150 MHz	6,7 ns	300 MHz	600 MT/s	PC2-4800	4.800 MB/s
DDR2-667	166 MHz	6 ns	333 MHz	667 MT/s	PC2-5300	5.336 MB/s
DDR2-800	200 MHz	5 ns	400 MHz	800 MT/s	PC2-6400	6.400 MB/s
DDR2-1000	250 MHz	3,75 ns	500 MHz	1.000 MT/s	PC2-8000	8.000 MB/s
DDR2-1066	266 MHz	3,75 ns	533 MHz	1.066 MT/s	PC2-8500	8.530 MB/s
DDR2-1150	286 MHz	3,5 ns	575 MHz	1.150 MT/s	PC2-9200	9.200 MB/s
DDR2-1200	300 MHz	3,3 ns	600 MHz	1.200 MT/s	PC2-9600	9.600 MB/s

MT/s: Milions de transferències per segon

Mòduls DDR estàndard

Nom	Freq.	Retard	F. rell. E/E	V. Transf.	Denom.	Taxa Transf. Màx.
DDR3-1.066	133 MHz	7,5 ns	533 MHz	1.066 MT/s	PC3-8500	8.530 MB/s
DDR3-1.200	150 MHz	6,7 ns	600 MHz	1.200 MT/s	PC3-9600	9.600 MB/s
DDR3-1.333	166 MHz	6 ns	667 MHz	1.333 MT/s	PC3-10667	10.664 MB/s
DDR3-1.375	170 MHz	5,9 ns	688 MHz	1.375 MT/s	PC3-11000	11.000 MB/s
DDR3-1.466	183 MHz	5,5 ns	733 MHz	1.466 MT/s	PC3-11700	11.700 MB/s
DDR3-1.600	200 MHz	5 ns	800 MHz	1.600 MT/s	PC3-12800	12.800 MB/s
DDR3-1.866	233 MHz	4,3 ns	933 MHz	1.866 MT/s	PC3-14900	14.930 MB/s
DDR3-2.000	250 MHz	4 ns	1000 MHz	2.000 MT/s	PC3-16000	16.000 MB/s

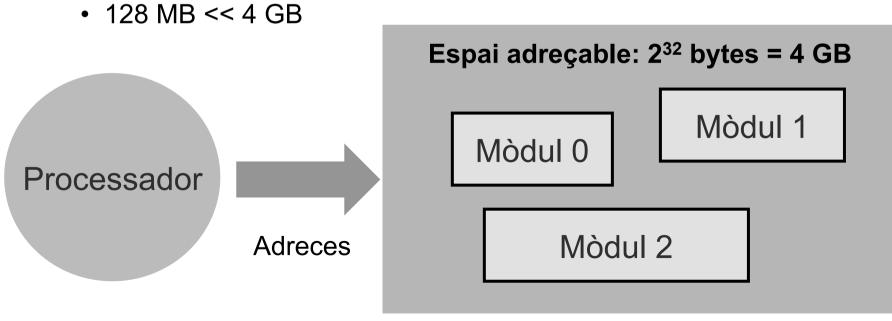
MT/s: Milions de transferències per segon

6. Mapes de memòria

- Concepte de mapa
- Funcions de selecció de mòdul

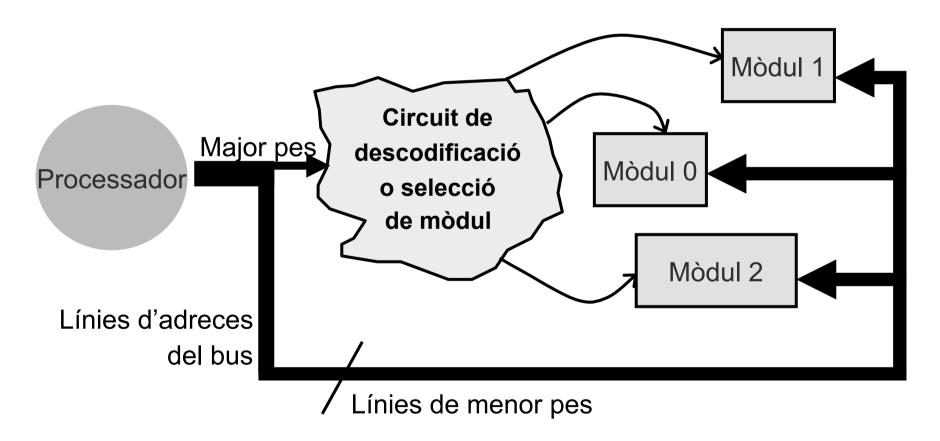
Concepte de mapa de memòria

- Distribució dels distints mòduls en l'espai físic d'adreçament del processador
 - L'espai adreçable sol ser molt més gran que l'espai ocupat pels mòduls físics de la memòria
 - Exemple: processador amb 32 bits d'adreces i memòria principal de 128 MB implementada mitjançant 2 mòduls de 32 MB i un de 64 MB



Circuit de descodificació o selecció

 Cal descodificar l'adreça emesa pel processador per tal d'identificar el mòdul referenciat



Paràmetres a considerar

- Processador
 - Espai d'adreçament
 - Depén de la longitud de les adreces efectives
- Mòdul
 - Adreça d'inici dins del mapa
 - Capacitat expressada en bytes
 - Compte: no influeix l'organització interna del mòdul
- Nivell d'activació de les funcions de selecció
 - Nivell alt o baix

Un cas particular de mapa

- Processador de 32 bits i espai d'adreçament de 256 KB
- Mòdul DRAM de 64 KB i adreça d'inici en 0x20000

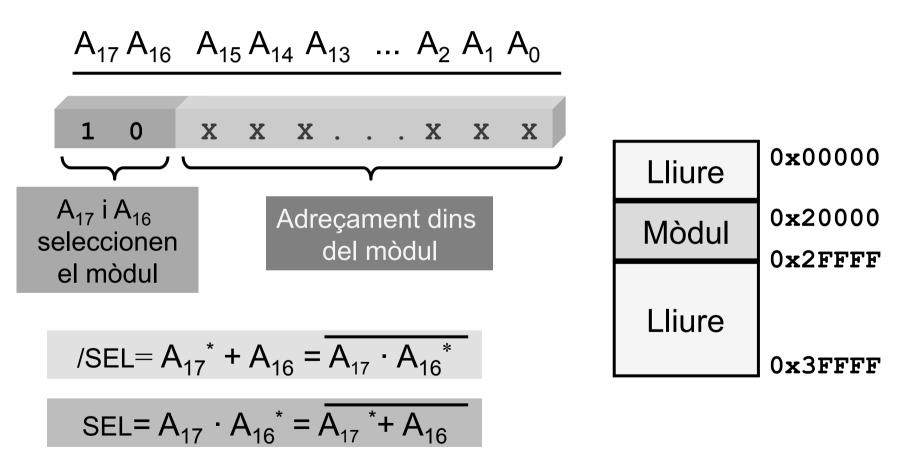
A ₁₇	A ₁₆	A ₁₅	A ₁₄	A ₁₃	•	•			A_2	A ₁	A_0	_
1	0	0	0	0	•	•	•	•	0	0	0	0x20000
1	0	0	0	0	•	•	•	•	0	0	1	0x20001
1	0	0	0	0	•	•	•	•	0	1	0	0x20002
1	0	1	1	1	•	,	•	•	1	1	0	0x2FFFE
1	0	1	1	1	•	•	•	•	1	1	1	0x2FFFF



Forma general de les adreces que pertanyen al mòdul

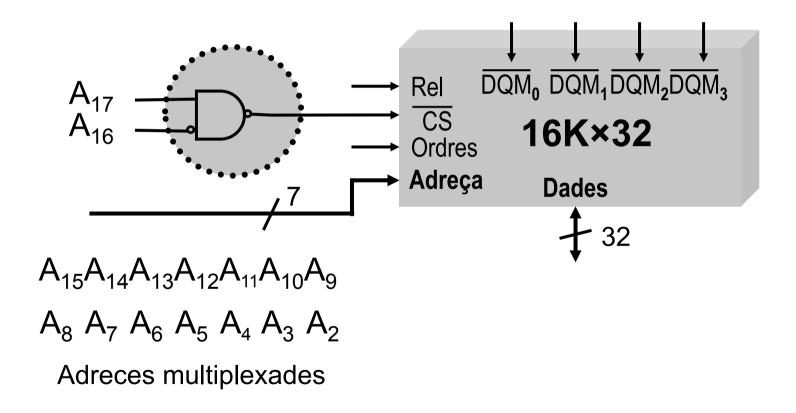
Funció de selecció del mòdul

- El mòdul conté totes les adreces de la forma 0x2????
 - Rang d'adreces : 0x20000 fins 0x2FFFF



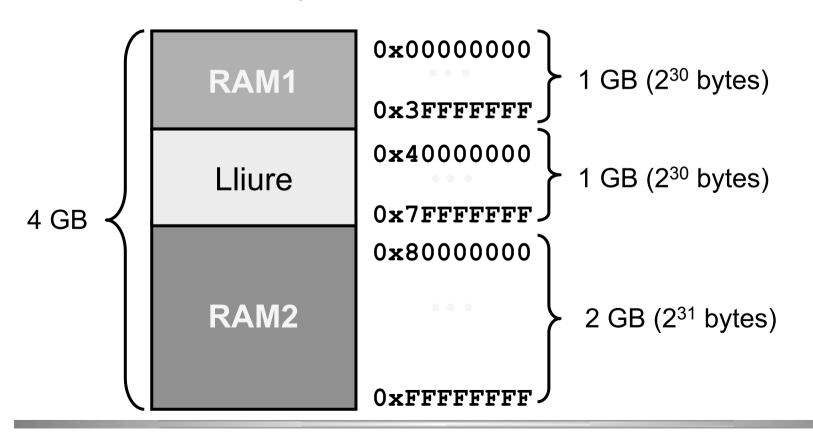
Detall de la implementació física

Cal considerar les línies físiques



Exemple de mapa amb dos mòduls

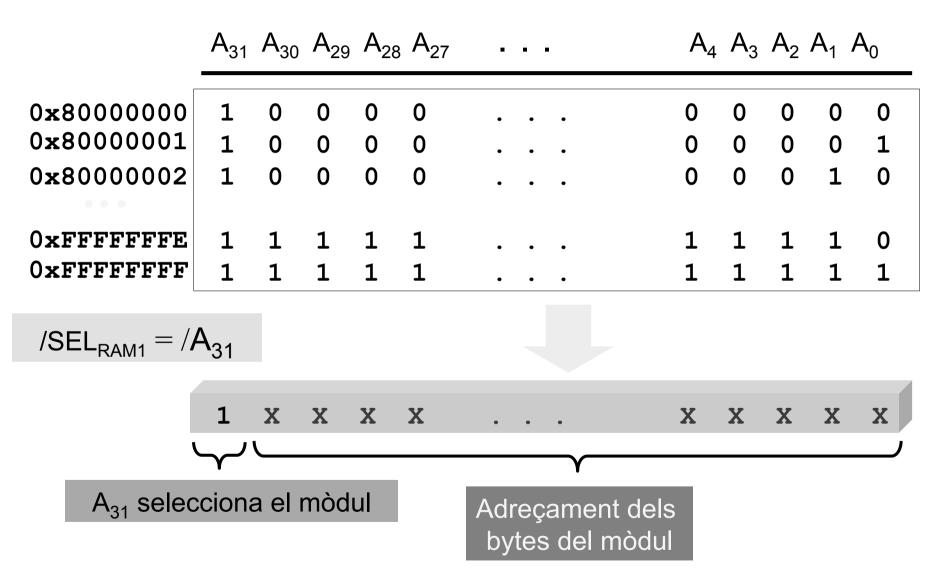
- MIPS R2000: espai d'adreçament de 4 GB (2³² bytes)
- Mòduls, capacitat i adreça de començament
 - RAM1, 1 GB, a partir de 0x0000000
 - RAM2, 2 GB, a partir de 0x80000000



Detall del mapa de memòria per a RAM1

	A ₃₁	A ₃₀	A ₂₉	A ₂₈	A ₂₇	•	= 1	•	A ₄	A ₃	A_2	A_1	4 ₀
0x00000000	0	0	0	0	0	•	•	•	0	0	0	0	0
0x0000001	0	0	0	0	0	•	•	•	0	0	0	0	1
0x0000002	0	0	0	0	0	•	•	•	0	0	0	1	0
0x3FFFFFE	0	0	1	1	1	•	•	•	1	1	1	1	0
0x3FFFFFFF	0	0	1	1	1	•	•	•	1	1	1	1	1
$/SEL_{RAM1} = A_{31} + A_{30}$													
	0	0	X	X	X	•	•	•	Х	X	X	X	X
•	$\overline{}$												
A ₃₁ i A	A	\dr	eçament (del mòdu										

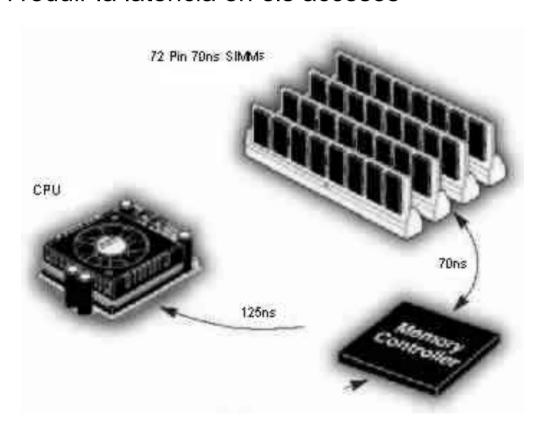
Detall del mapa de memòria per a RAM2



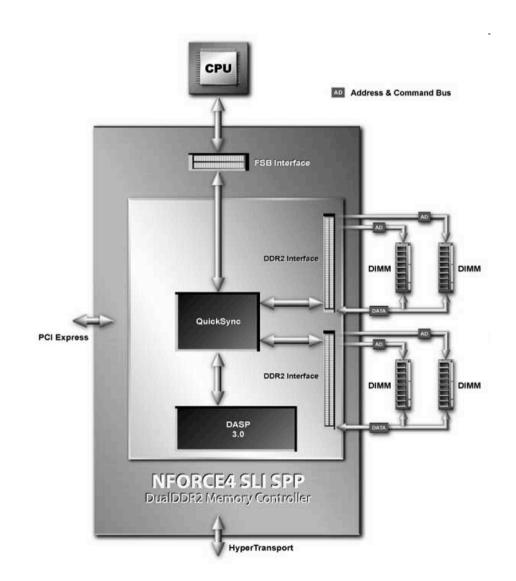
7. El controlador de memòria dinàmica

- Funcions del controlador
- Inicialització del sistema de memòria

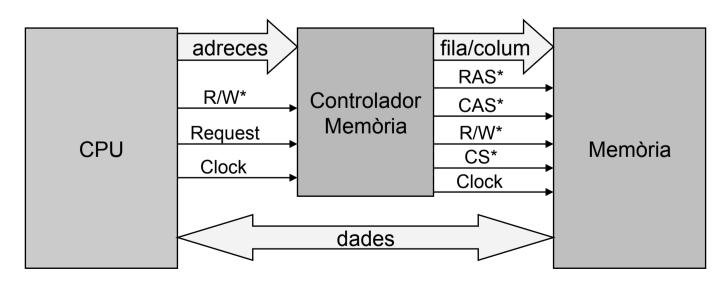
- Gestiona els mòduls de memòria dinàmica i fa d'intermediari entre aquesta i el processador
- El seu principal objectiu és maximitzar l'amplada de banda efectiva de memòria i reduir la latència en els accesos



- Disponibilitat d'un o varios canals independents
- L'amplada de banda total del sistema de memòria és la suma de les amplades de banda de cada canal
- Convé conéixer les limitacions i requeriments del controlador abans d'ampliar la memòria!!!

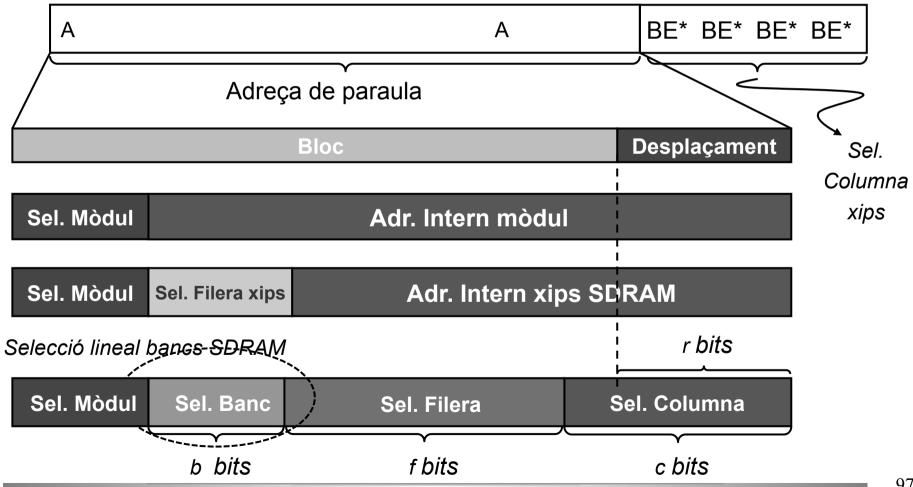


- Funcions del controlador:
 - Selecció del mòdul de DRAM segons l'adreça de paraula del bus i activació de la línia CS* corresponent
 - Traducció de l'adreça física lineal en adreces de banc, fileres i columnes (i línies BE* del bus en línies DQM*), tractant de minimitzar els conflictes de bancs en accesos adjacents i maximitzar el paralelisme proporcionat per el sistema de memòria (canal, mòdul, filera, columna)
 - Sincronització amb els xips (aplica el protocol d'accés)
 - Generació del senyal de rellotge i gestió del refresc



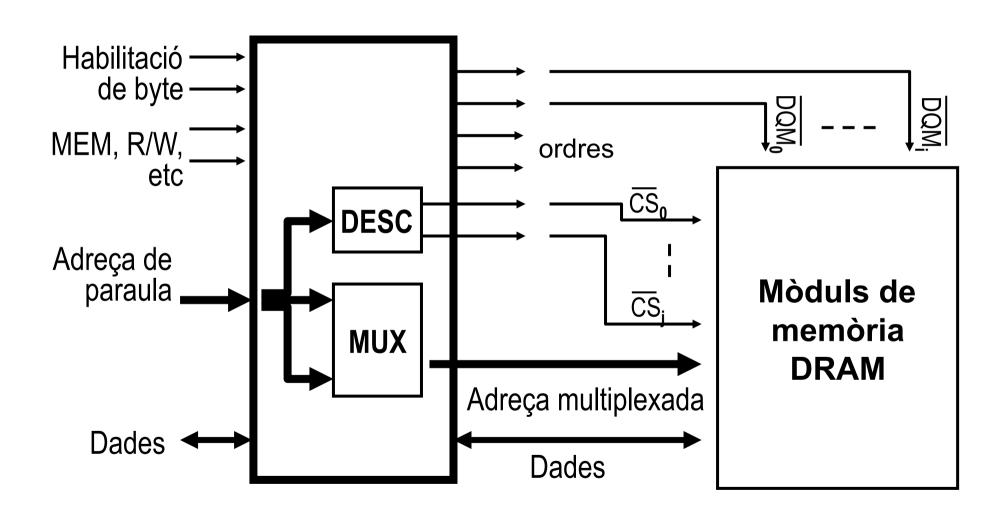
Estructura resultant de les adreces

Bus físic d'adreces (CPU 32 bits)



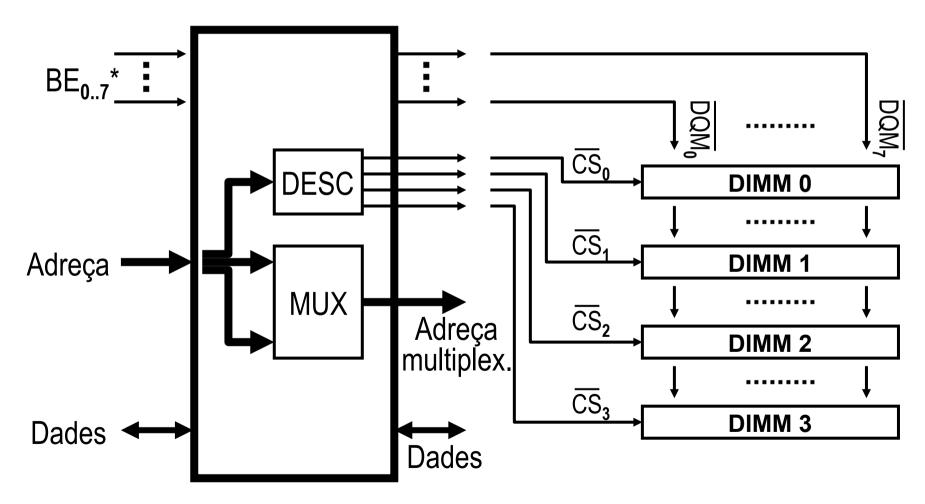
- Estructura resultant de les adreces
 - n: nombre de bits de l'adreça. Depén del processador
 - b bits d'adreça de banc. b = log₂ num_bancs
 - f bits d'adreça de fileres. f = log₂ num_fileres
 - c bits d'adreçament de columna. c = log₂ num_columnes
 - r bits de desplaçament dins del bloc. Depén del tamany de bloc determinat pel sistema de memòria

Estructura del controlador de memòria DRAM



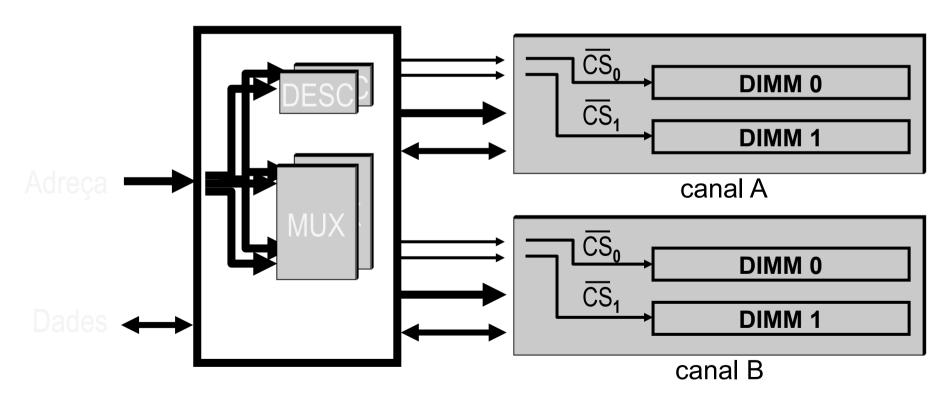
Exemple de configuració DRAM estàndard

W=64 bits, 4 mòduls DIMM



Exemple de configuració DRAM estàndard

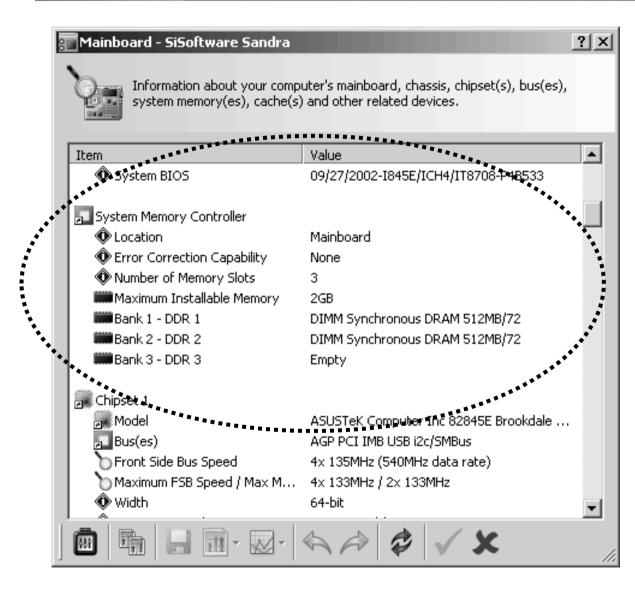
- W=64 bits, 4 mòduls DIMM en dos canals
 - Els més frequents ara (2006 en endavant)
 - Permet duplicar l'amplada de banda dels mòduls
 - S'obliga a una certa distribució de mòduls entre els canals!



Funcions del controlador de memòria SDRAM

- Inicialització del sistema de memòria
 - El controlador de memòria forma part del controlador de sistema o xipset. El xipset és un circuit de complexitat comparable al processador, i que concentra moltes funcions crítiques
 - Al engegar un computador, el controlador comprova, un a un, tots els mòduls connectats, n'obté les característiques clau a partir del SPD (capacitat, geometria de matriu, restriccions temporals, etc.) i inicialitza els xips
 - El controlador configura
 - el mapa de memòria amb els mòduls presents i en fixa les funcions de selecció
 - la freqüència de rellotge
 - la temporització adient per a accedir correctament als mòduls
 - la freqüència d'emissió d'ordres AUTOREFRESH

SiSoftware Sandra: controlador de memòria



Tres sòcols de memòria, dos d'ells ocupats per mòduls d'igual capacitat