

Memòria principal

Funció

Com s'ha indicat diferents vegades, la Memòria Principal és la que conté els Programes quan s'estan executant i les Dades necessàries per l'execució.

De fet, és la única memòria imprescindible a un ordinador. Si fem servir diferents tipus de memòries és per la senzilla raó de que no existeix un dispositiu de memòria perfecte. Les característiques que hauria de tenir una memòria ideal són:

- Ràpida com el processador o més.
- Capacitat molt gran en poc espai (alta densitat d'integració).
- Poder retenir les dades quan no hi ha energia.
- Poder Escriure i Llegir a la mateixa velocitat.
- Econòmica.

Constitució

Per tal de tenir unes bones prestacions, els requisits principals són que sigui ràpida, de gran capacitat i que es pugui llegir i escriure amb temps semblants. El primer requisit imposa que sigui una memòria basada en semiconductors.

TIPUS	Cel·la Bàsica	Característiques	Utilització
SRAM	Flip-Flop (6 transistors)	Gran Velocitat, Baixa Densitat Volàtil	Caché
DRAM	Transistor +Capacitat	Alta densitat Lenta respecte a les SRAM Volàtil Necessita Controlador	Memòria Principal (mides grans)
ROM	Fusible	Alta Densitat No Volàtil NO Reprogramable	Només per informació fixa
EEPROM	Transistor de porta flotant	S'esborra elèctricament cel·la a cel·la Baixa densitat (respecte a les Flash)	Sistemes de baixa capacitat, reprogramació y simplicitat de connexió
FLASH	Transistor de porta flotant	No Volàtil S'esborra elèctricament per blocs Alta densitat (resp. a les EEPROM)	SETUP-BIOS Sistemes reprogramables d'alta capacitat

Resum dels tipus de Memòries Semiconductores que poden formar part d'un Computador.

Memòries DRAM

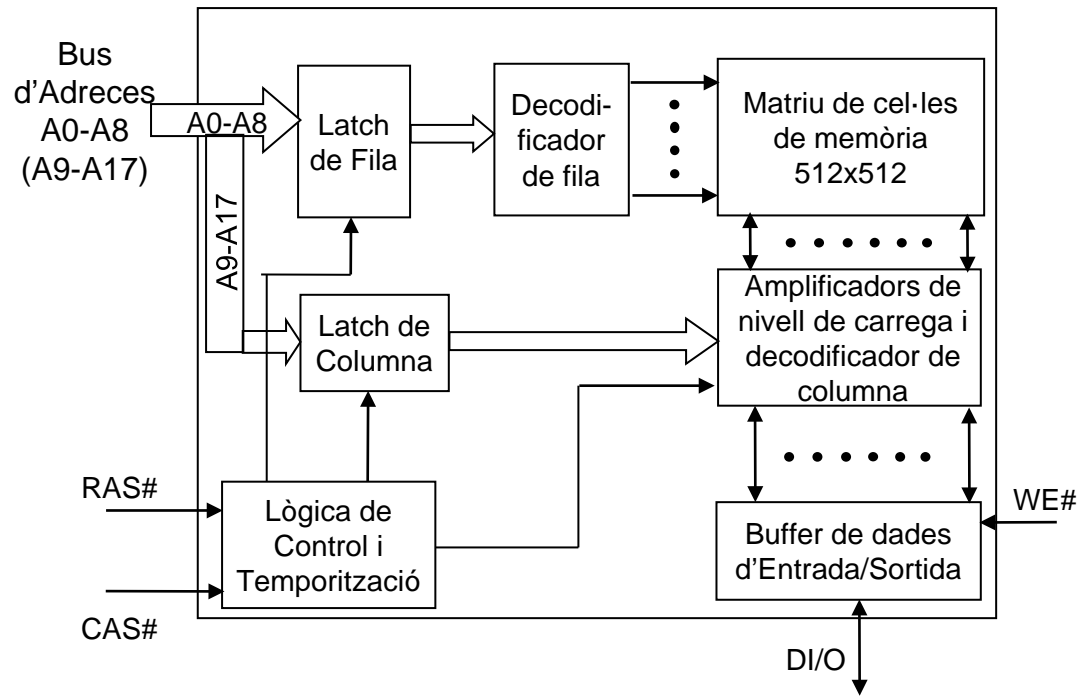
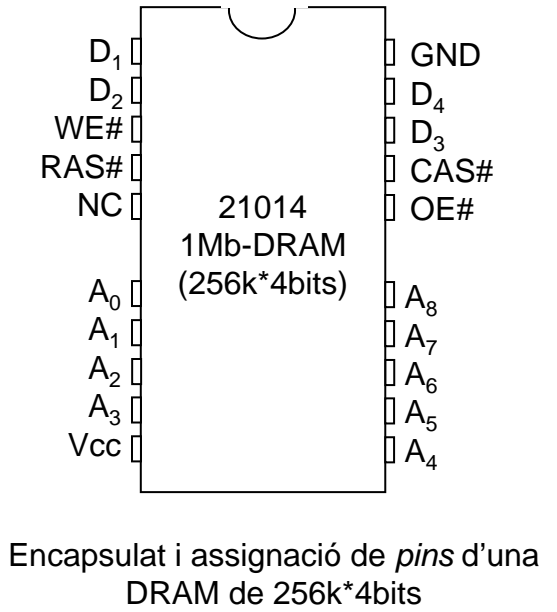


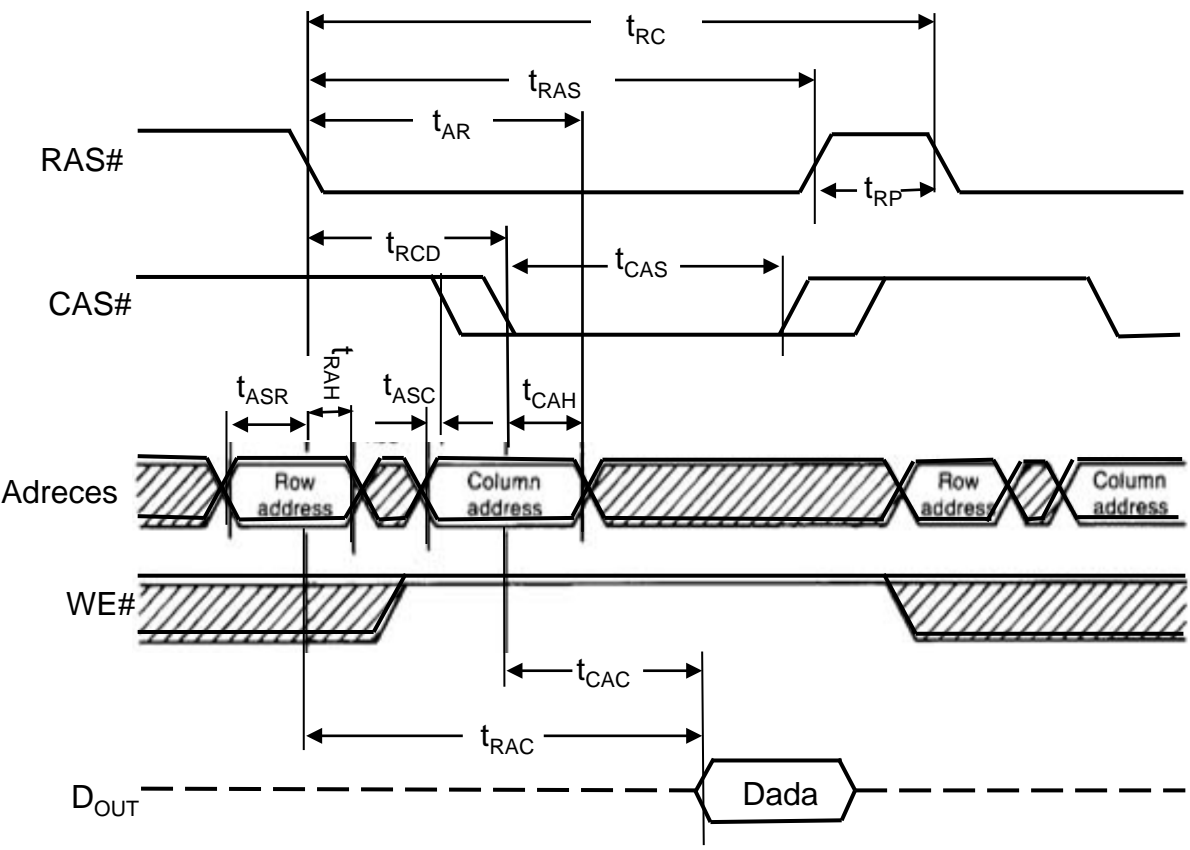
Diagrama de blocs d'una memòria DRAM de 256K x 1 bit



Encapsulat i assignació de pins d'una DRAM de 256k*4bits

Senyals de Control	
RAS# : <i>Row Address Strobe</i>	WE#
CAS# : <i>Column Address Strobe</i>	OE#
No hi ha cap senyal de <i>Chip Select</i>	

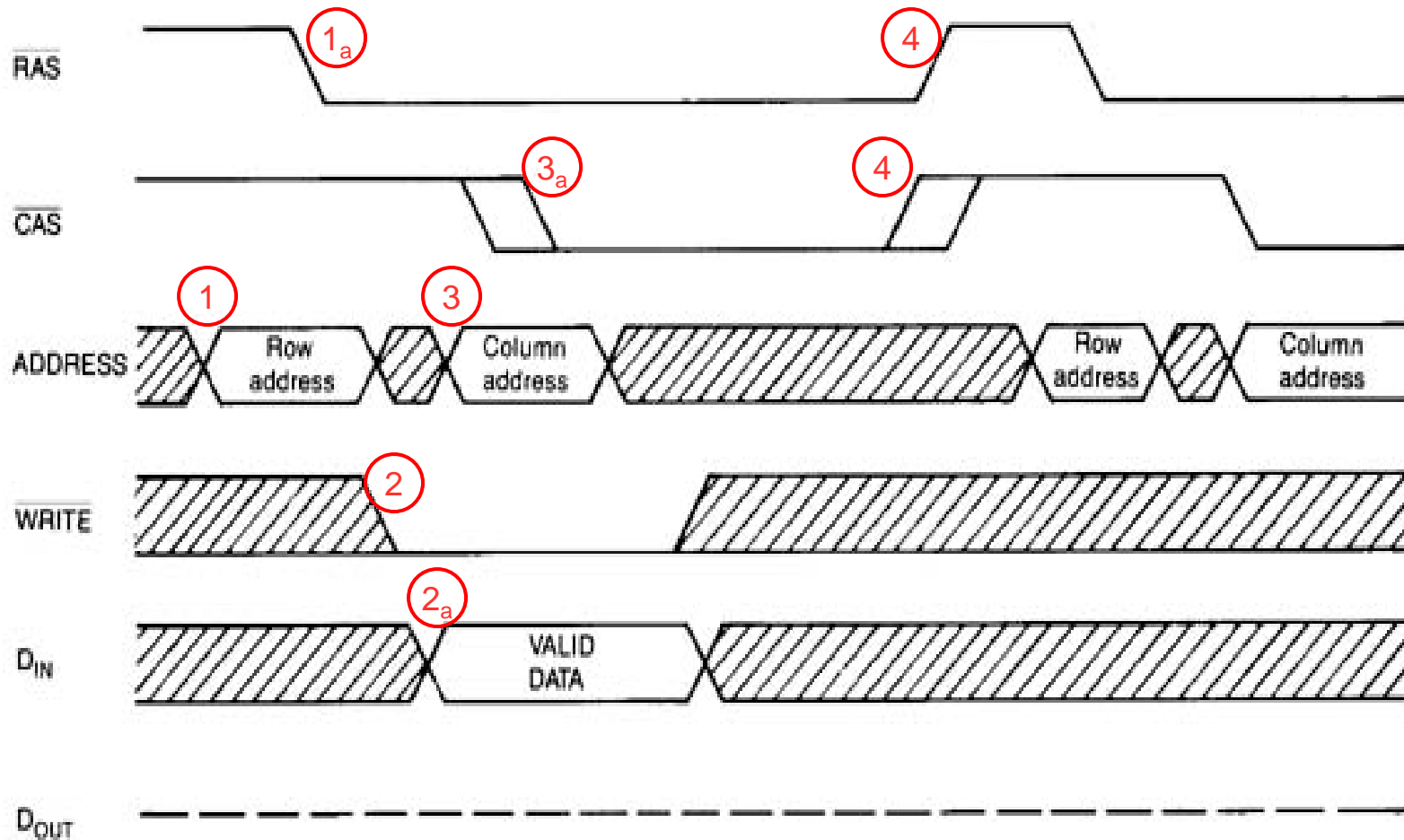
Diagrama Temporal d'una Lectura a una Memòria DRAM



Temps
t_{RC} : Temps de Cicle de Lectura
t_{RAS} : Temps d'activació del senyal RAS#
t_{CAS} : Temps d'activació del senyal CAS#
t_{AR} : Temps d'activació de les Adreces
t_{ASR} : Temps de <i>Setup</i> del senyal RAS#
t_{ASC} : Temps de <i>Setup</i> del senyal CAS#
t_{RAH} : Temps de <i>Hold</i> del senyal RAS#
t_{CAH} : Temps de <i>Hold</i> del senyal CAS#
t_{RAC} : Temps d'Accés des de activació RAS#
t_{CAC} : Temps d'Accés des de activació CAS#
t_{RCD} : Retard mínim entre RAS# i CAS#
t_{RP} : Temps de Precàrrega.

Hold : Manteniment del senyal. *Setup* : Activació prèvia del senyal.

Diagrama Temporal d'una Escriptura a una Memòria DRAM



Refresc a les Memòries DRAM

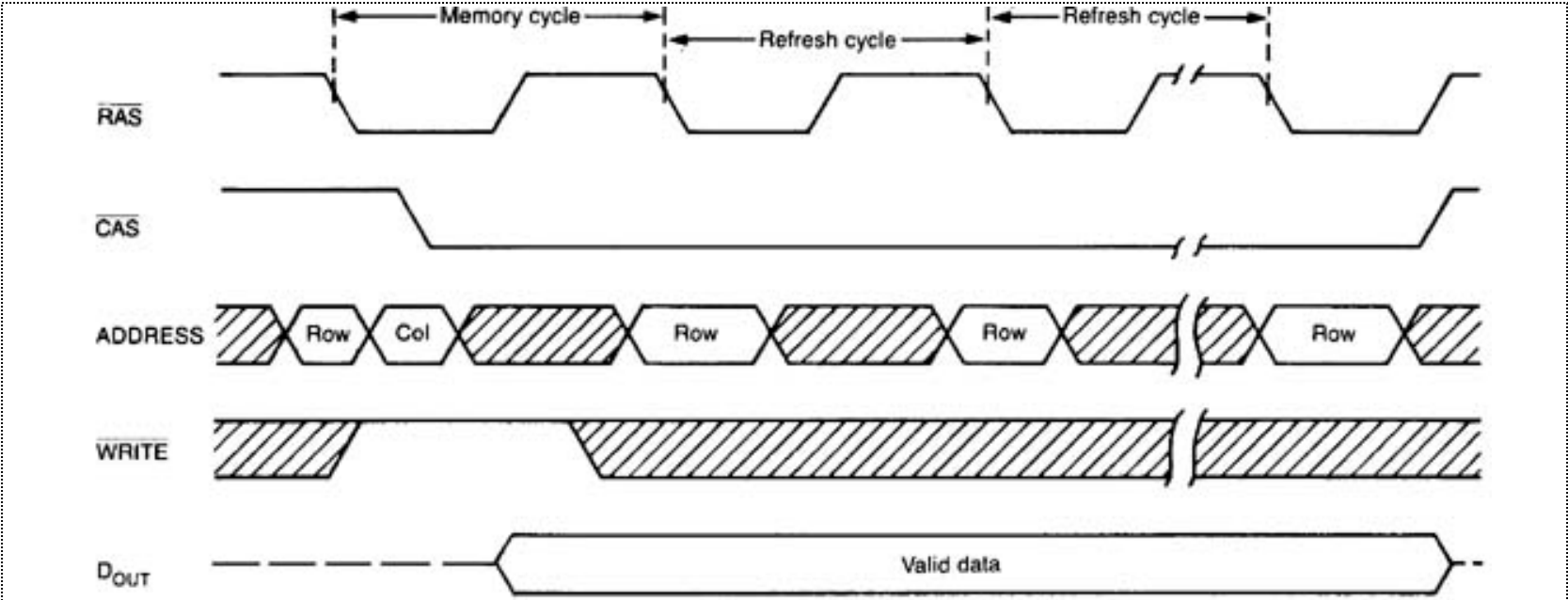
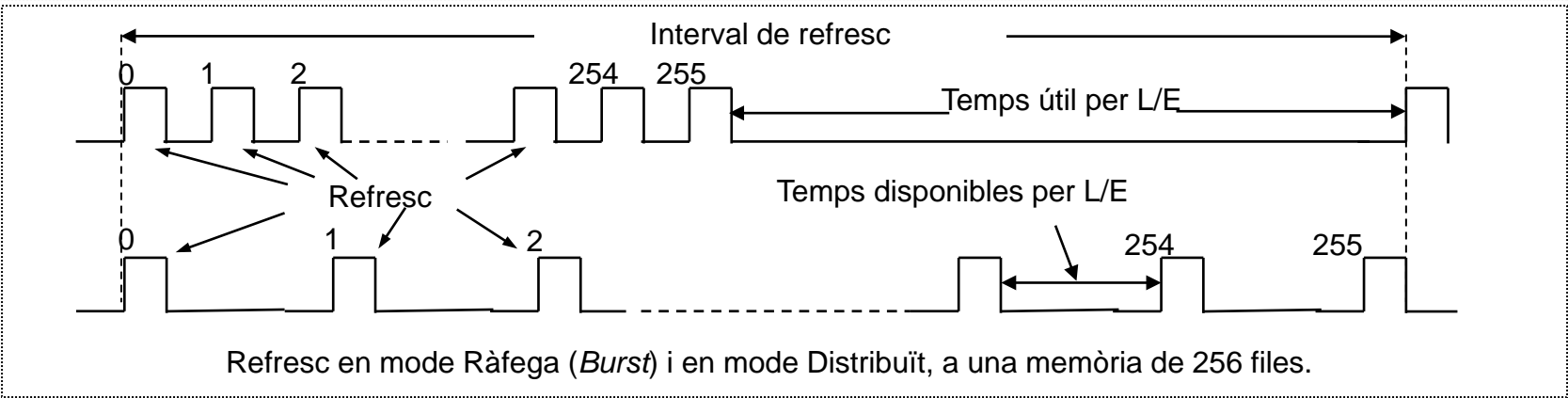
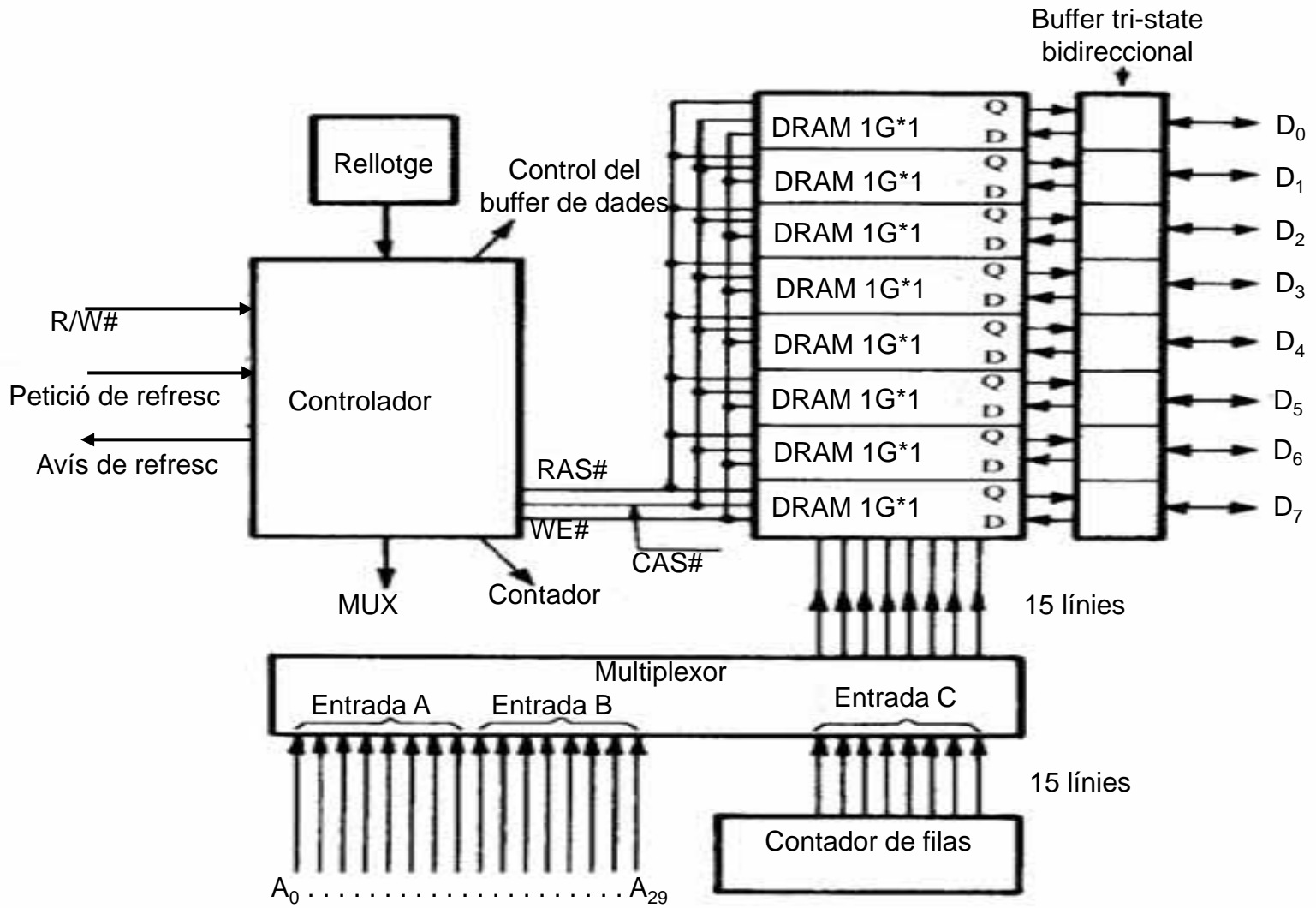


Diagrama temporal d'una operació de Refrescament Amagat (*Hidden Refresh*) d'una DRAM



Refresc en mode Ràfega (*Burst*) i en mode Distribuït, a una memòria de 256 files.

Sistema de Memòria DRAM de 1GBytes, amb xips de 1Gbits



Problemes

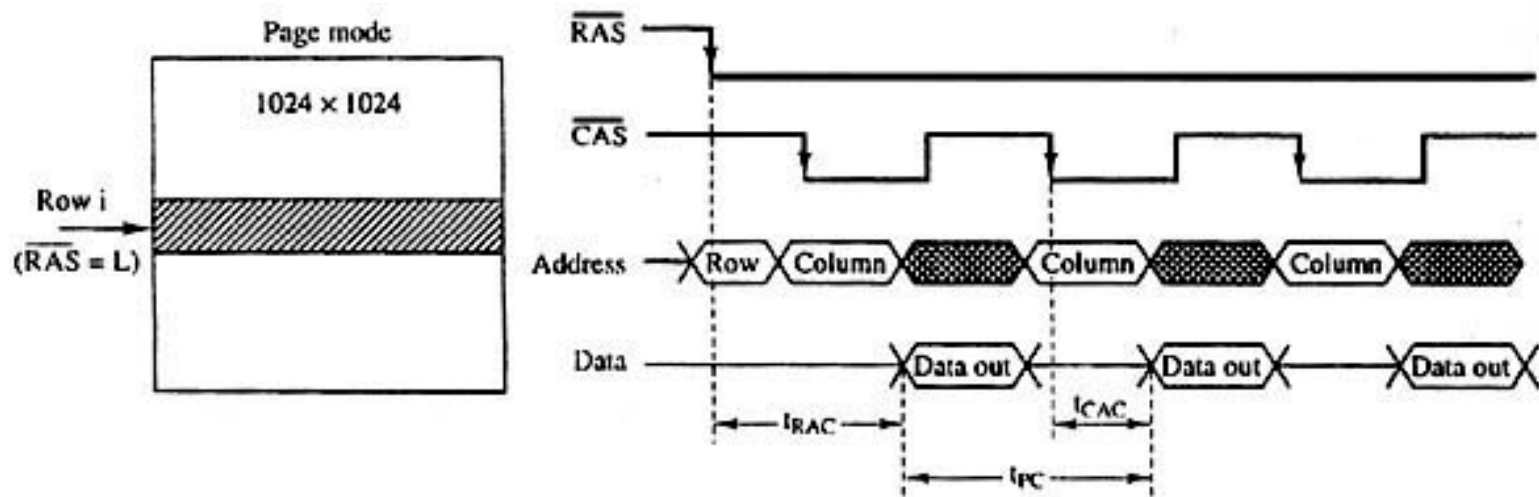
Així les memòries DRAM, tot i que són les que més s'ajusten als requisits de la Memòria Principal d'un ordinador, tenen una sèrie d'inconvenients:

1. El principal problema de les memòries DRAM continua sent el seu ample de banda (màxim número de bits que la memòria pot transferir per unitat de temps, p.e.: una memòria de temps de cicle de 100ns i 32 bits de longitud de paraula té un ample de banda màxim de 40MB/s).
2. Una altre problema és que és que operar amb DRAMs és més complicat perquè es necessita un controlador de DRAM per gestionar les tasques de:
 - Refrescament de les cel·les de memòria.
 - Gestió dels senyals RAS# i CAS#.
 - Multiplexació d'adreces.

El primer punt intentarem millorar-lo amb diferents tècniques, però el segon és insalvable.

Accés en Mode Pàgina “FPM”

Consisteix en que un cop hem fet una accés normal, com que ja tenim la fila carregada als amplificadors de càrrega, si volem accedir a una dada que està a aquesta mateixa línia, ens podem estalviar l'accés a la fila. Només hem de indicar quina és la nova columna a la que volem accedir i activar el senyal /CAS.



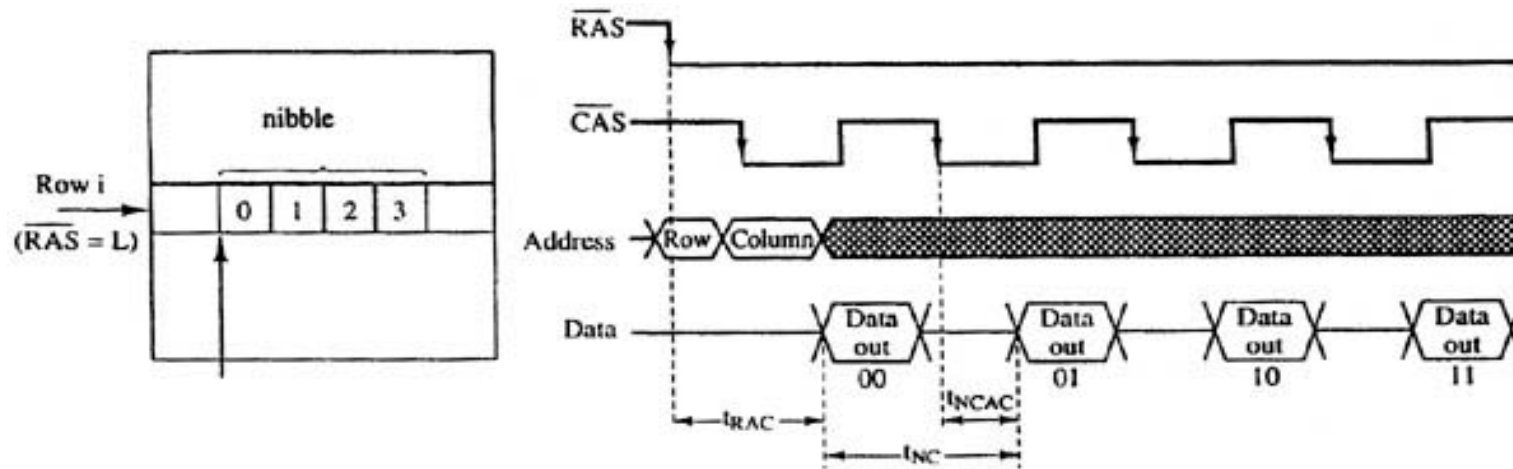
Temps d'accés en mode Pàgina = $t_{CAC} \approx t_{RAC}/3$

Temps de cicle en mode Pàgina = $t_{PC} \approx t_{RC}/3$

On t_{RAC} és el temps d'accés normal i t_{RC} el de cicle de lectura normal

Accés en Mode Nibble i Sèrie

En aquest mode d'accés, que com tots l'anteriors comença amb un accés normal, els accessos successius no necessiten que es doni cap adreça de columna. Senzillament, cada pols del senyal /CAS proporciona la dada que hi ha a la següent columna a la que hem accedit prèviament. El guany en temps per accés és més gran que als altres casos però té l'inconvenient que només pot accedir a quatre bits consecutius.



Temps d'accés en mode Nibble (Sèrie) = $t_{NCAC} \approx t_{RAC}/3$

Temps de cicle en mode Nibble (Sèrie) = $t_{NC} \approx t_{RC}/4$

On t_{RAC} és el temps d'accés normal i t_{RC} el de cicle de lectura normal

Hi ha una versió millorada, que s'anomena accés sèrie, que és la generalització del mode Nibble sense la limitació de 4 bits.

Entrellaçat de Memòria (Memory Interleaving)

- A més de solucions tecnològiques, com les dues anteriors, també hi ha millores arquitecturals per aconseguir millors amplituds de banda.
- Una d'aquestes tècniques és l'Entrellaçat de les posicions (bancs) de memòria. Amb aquesta tècnica s'intenta eliminar l'efecte degradant del Temps de Precàrrega a les memòries DRAM.

PRINCIPI DE FUNCIONAMENT: Els programes tendeixen a accedir a posicions adjacents de memòria. Nosaltres dissenyarem el sistema perquè dues posicions adjacents de memòria no estiguin al mateix banc. Això permetrà que es puguin solapar les activitats dels dos (o més) bancs i augmentar així l'amplitud de banda.

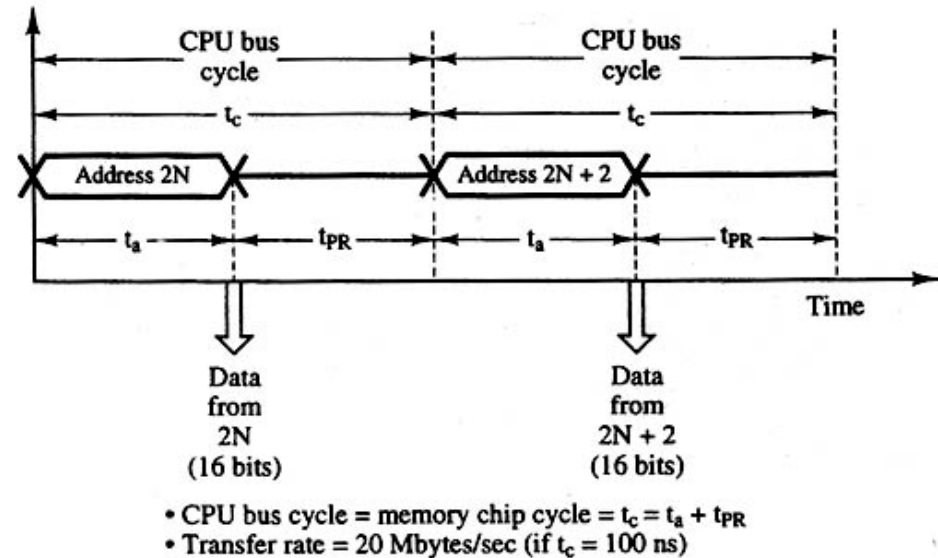
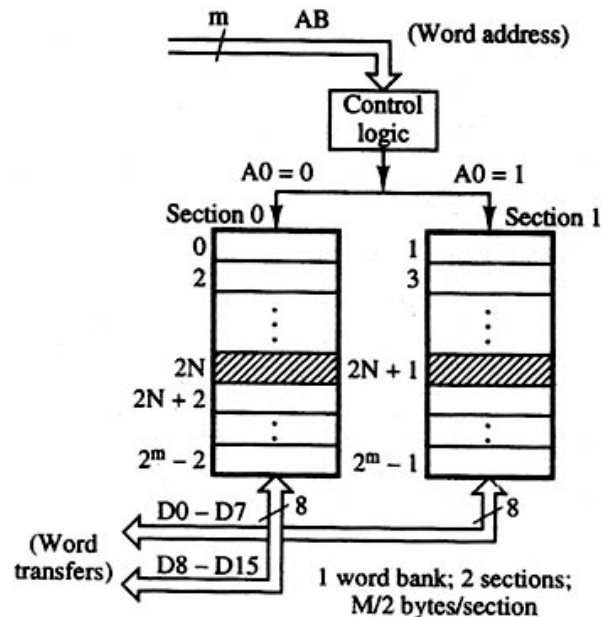
Segons la longitud de paraula del nostre sistema el disseny quedarà:

8 bits		16 bits		32 bits	
Banc 1	Banc 2	Banc 1	Banc 2	Banc 1	Banc 2
Posicions 2N	Posicions 2N+1	Posicions 4N i 4N+1	Posicions 4N+2 i 4N+3	Posicions 8N, 8N+1, 8N+2 i 8N+3	Posicions 8N+4, 8N+5, 8N+6 i 8N+7

Notació: t_a = Temps d'Accés (des de /RAS). T_{pr} = Temps de Precarrega. T_c = Temps de Cicle.

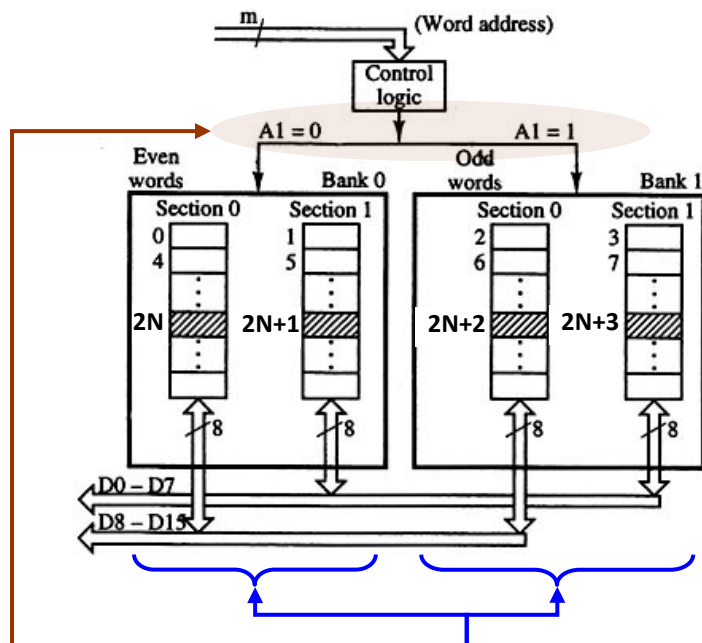
Considerarem pel l'anàlisi que $T_a = T_{pr} = T_c/2$

Accés "Normal" a una Memòria de 16 bits



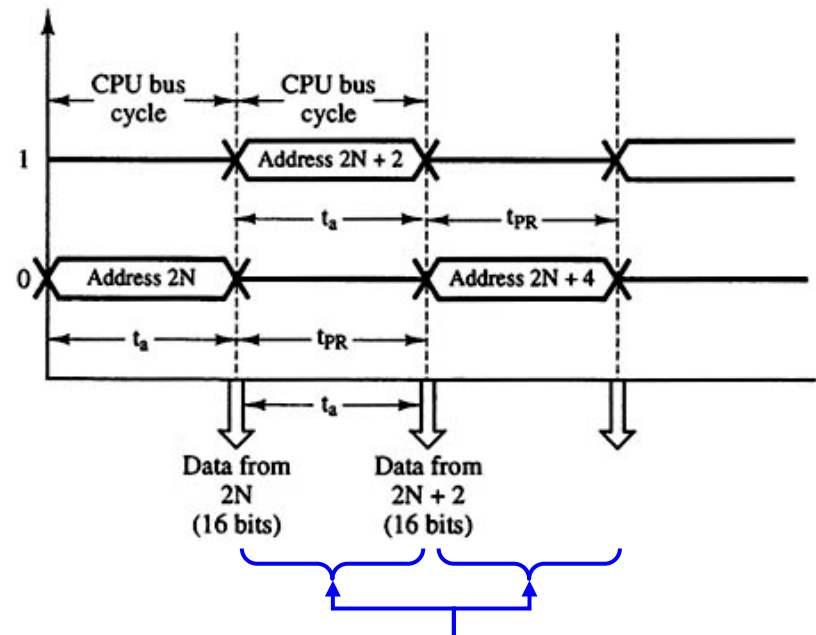
- Si el sistema és de 16 bits, la memòria la tindrem organitzada en dos seccions d'un byte.
- Tenint en compte els temps que hem comentat abans, el diagrama de temps de dos accessos consecutius seria el de la figura.

Entrellaçat de Memòria en 2 Bancs, a un Sistema de 16 bits.



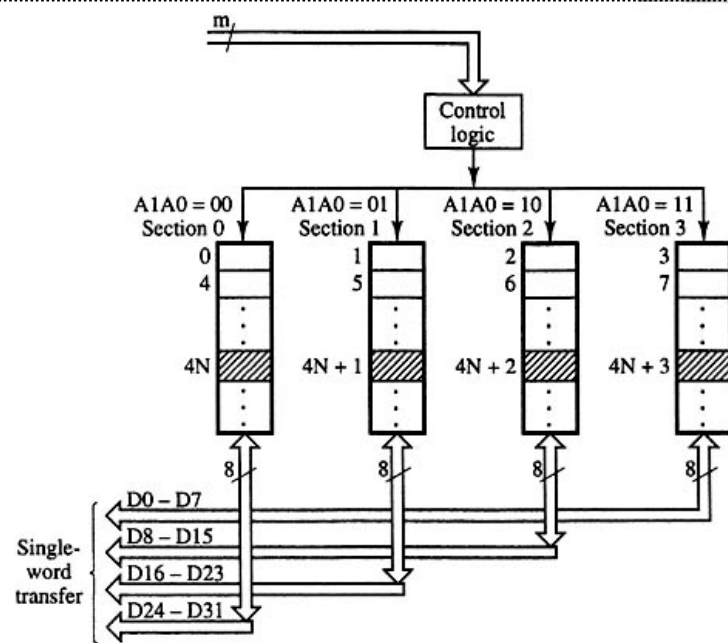
Per tal de dur a terme l'entrellaçat de memòria, ara dissenyarem dos bancs cadascú dels qual constarà de dues seccions.

Per tal de que estiguin entrellaçats, posicions consecutives han de estar a bancs diferents, el que hem de fer és fer servir el/s bit/s més petit/s possible per a fer la selecció del/s bancs. En aquest cas és el A_1 , ja que el A_0 es necessita per a seleccionar secció dintre del cada banc.

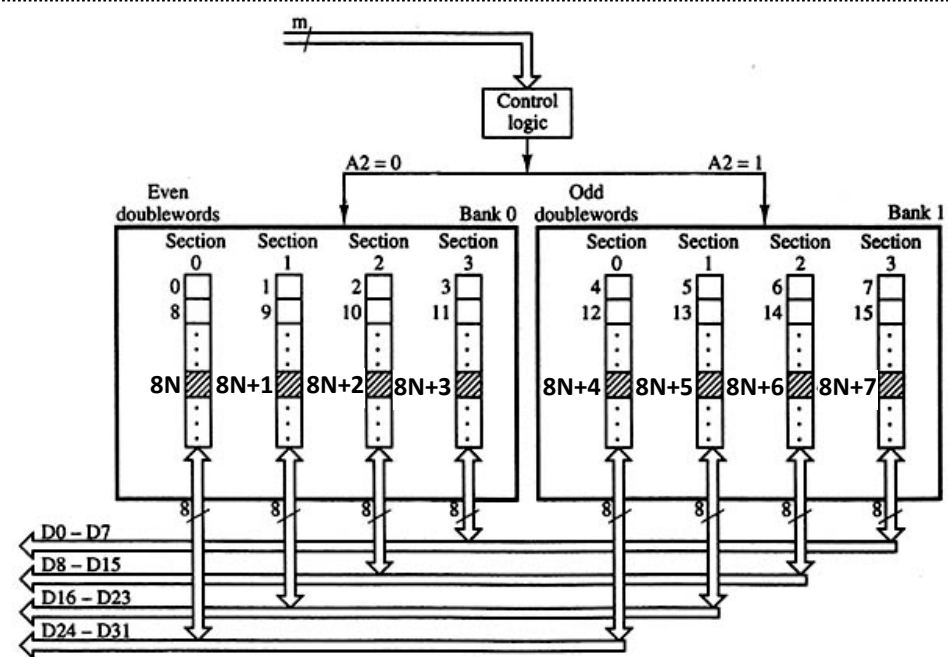


Quan accedim a posicions consecutives, o que estan a diferents bancs, llavors mentre un està fent l'accés l'altre pot fer la precàrrega....

Entrellaçat de Bancs de Memòria a sistemes de 32 bits

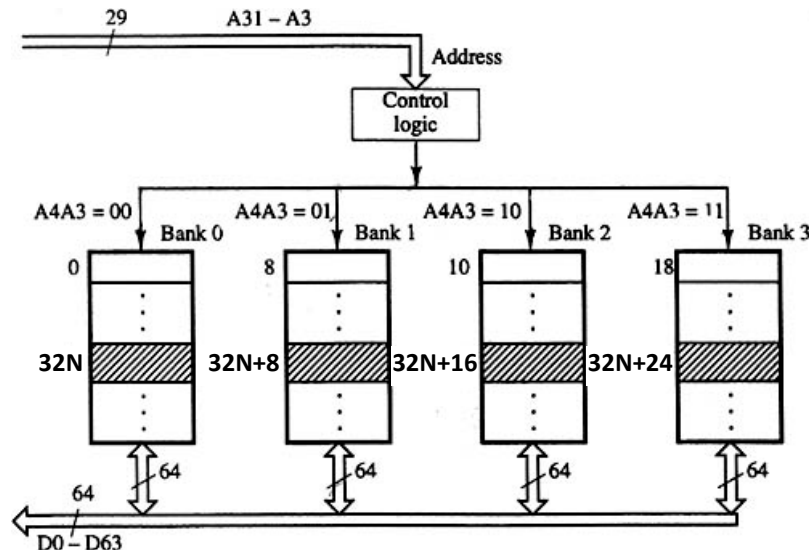


- Sistema de memòria de 32 bits sense entrellaçar. Els bits A0 i A1 els fem servir per seleccionar secció, i a partir d'aquí anirien als xips per a seleccionar posicions. Si tenim bancs, sense entrellaçar, seran els bits de més pes (després dels xips) els que farem servir per seleccionar el banc.

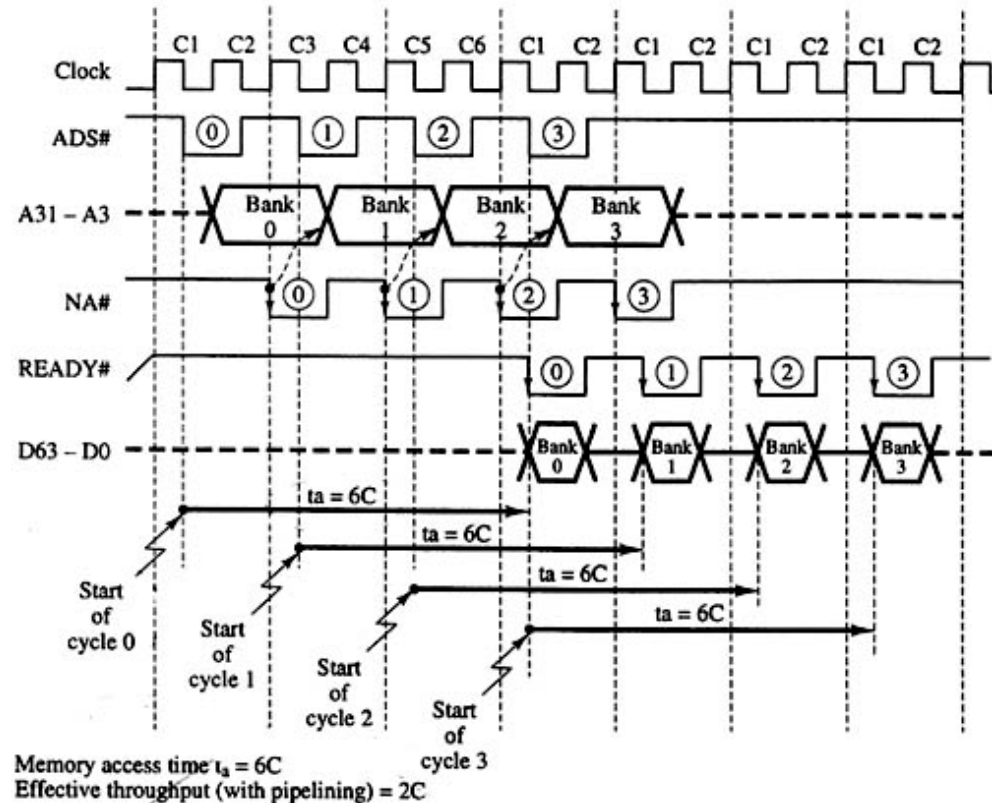


- Sistema de memòria de 32 bits amb 2 bancs entrellaçats. Igual que abans, els bits A0 i A1 els fem servir per seleccionar secció, el bit A2 servirà per seleccionar banc (això és el que fa que posicions consecutives estiguin entrellaçades). A partir d'aquí anirien als xips per a seleccionar posicions.

Exemple d'entrellaçat al i860 (4 bancs i 64bits dades)

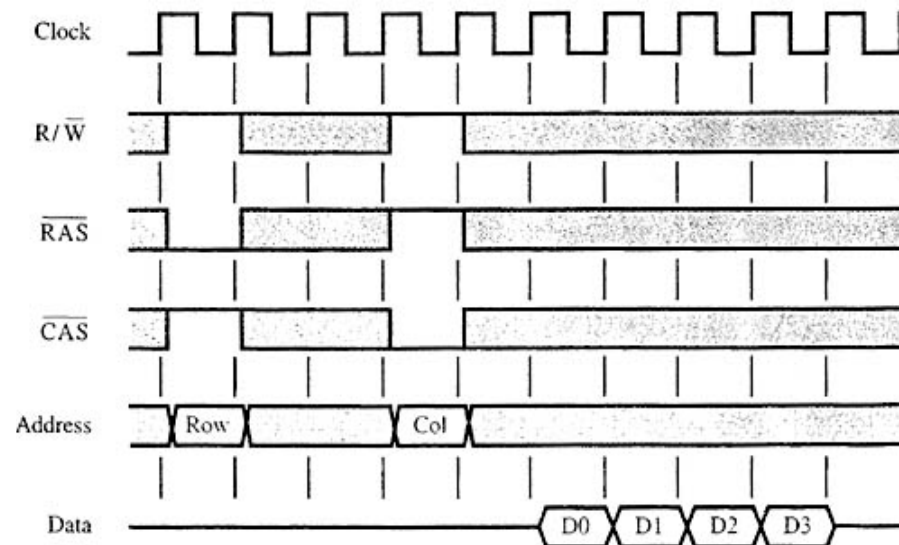
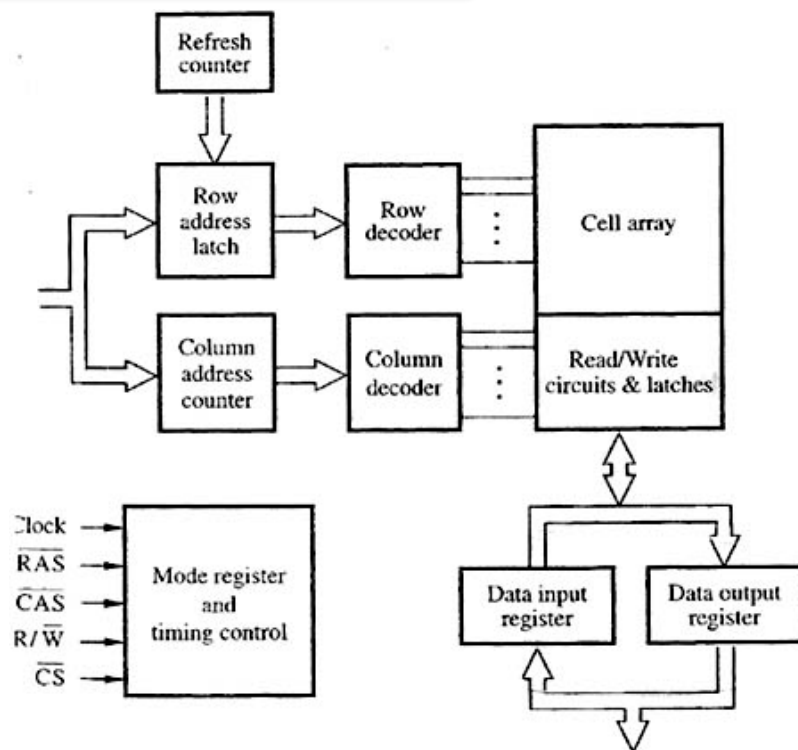


El i860 és un processador RISC d'Intel que té un bus de dades de 64bits (8 seccions, això vol dir A0, A1 i A2 per seleccionar secció) i admet accés en "Address Pipelining" de fins a 4 anticipacions. Per poder accedir a aquesta "facilitat" dissenyarem un entrellaçat de 4 bancs de manera que podem fer-los treballar simultàniament. Això vol dir que per a seleccionar bancs farem servir A3 i A4.



De fet, el entrellaçat no només serveix per "ocultar" el temps de precàrrega, si no també per a implementar accessos simultanis a diferents bancs. Encara que siguin de SRAM.

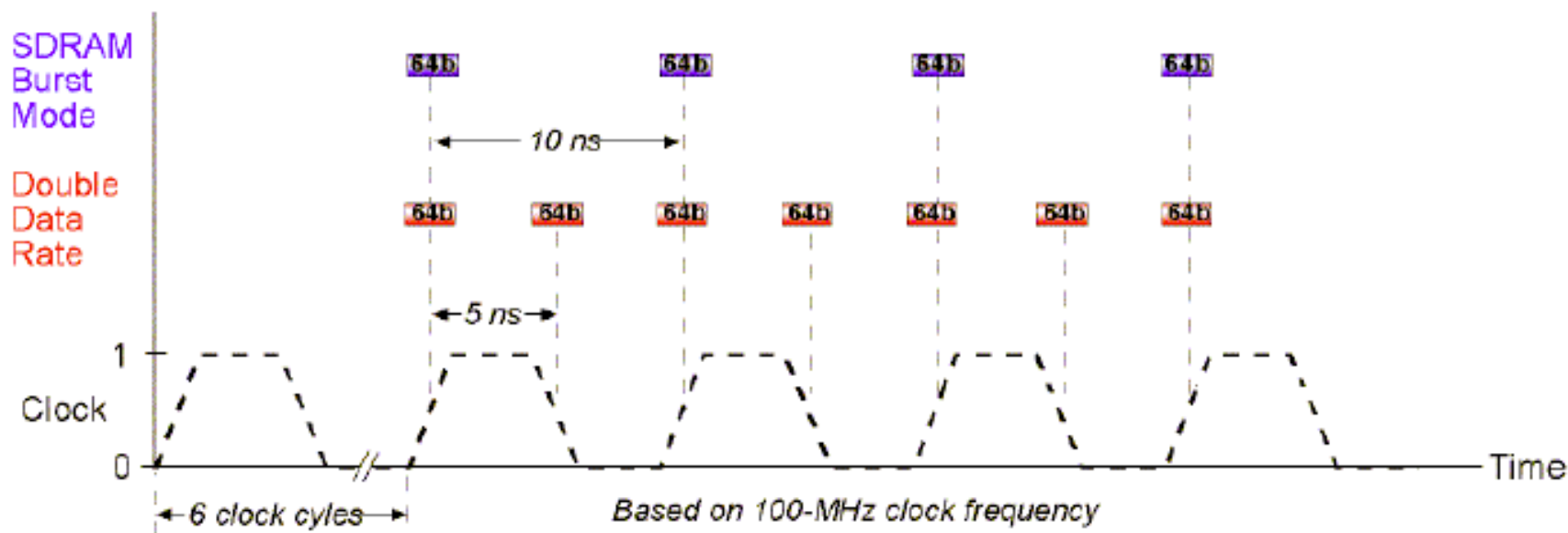
Memòries SDRAM



- Estan sincronitzades amb un senyal de rellotge amb el bus del sistema.
- Implementen bancs entrelaçats a un mateix xip.
- Les E/S de dades estan registrades.
- Les memòries SDRAM implementen la circuiteria de refresc internament.

- Poden treballar de diferents modes, els quals es seleccionen mitjançant "comandaments".
- A la figura tenim un diagrama de temps d'una transferència en ràfega (burst) 5-1-1-1.
- **CL**, t_{RAS} ..., temps de latència, cicle...
- Com es pot veure, ni tant sols fa falta que es vagi activant el senyal /CAS per a accedir a les dades consecutives. Es fa amb el "Clock" i un comptador intern que va generant les adreces següents.

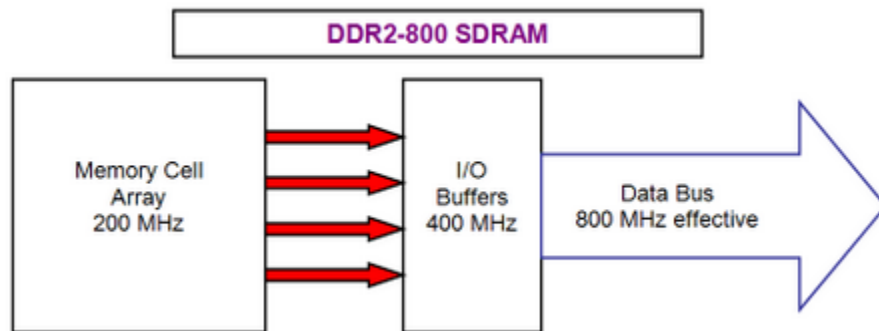
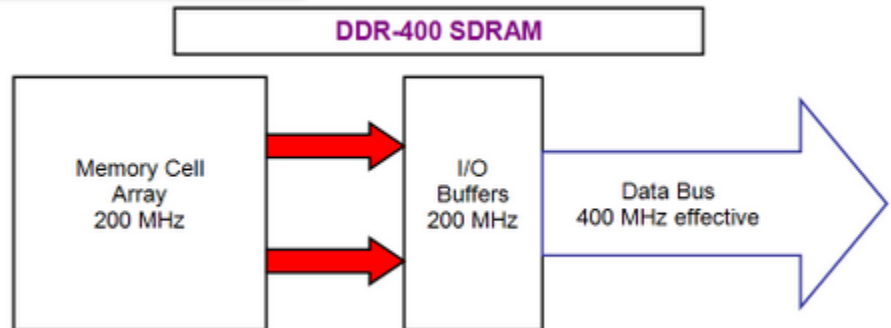
Comparació entre SDRAM i DDRAM (DDR)



- Prefetching: En cada cicle de rellotge es transfereixen 2 bits de la matriu de memòria als registres d'E/S, en comptes d'un com fan les SDRAM, existeixen dos canals de *pipeline* per fer això.
- **Dues Transicions:** Es transfereix un bit a cada flanc de rellotge, a la pujada i a la baixada.
- **Tensions menors:** Treballen a tensions menors (consum més petit i més velocitat). SDRAM a 3.3V, DDR a 2.5V, DDR2 a 1.8V i DDR3 a 1.5V.

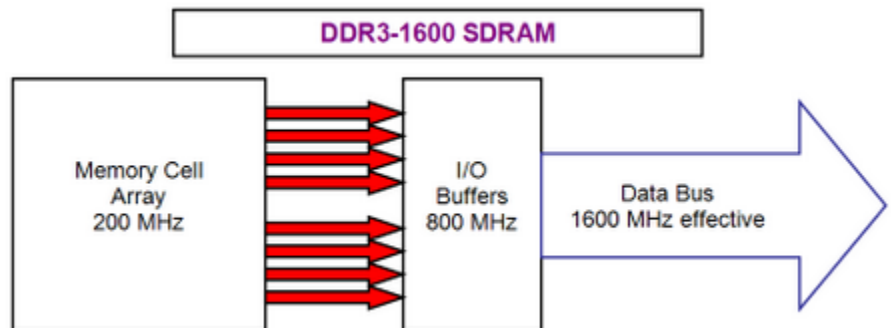
Diferències entre memòries DDR, DDR2 i DDR3

Esquema del principi de funcionament de les transferències de dades en un circuit de memòria DDR-400.

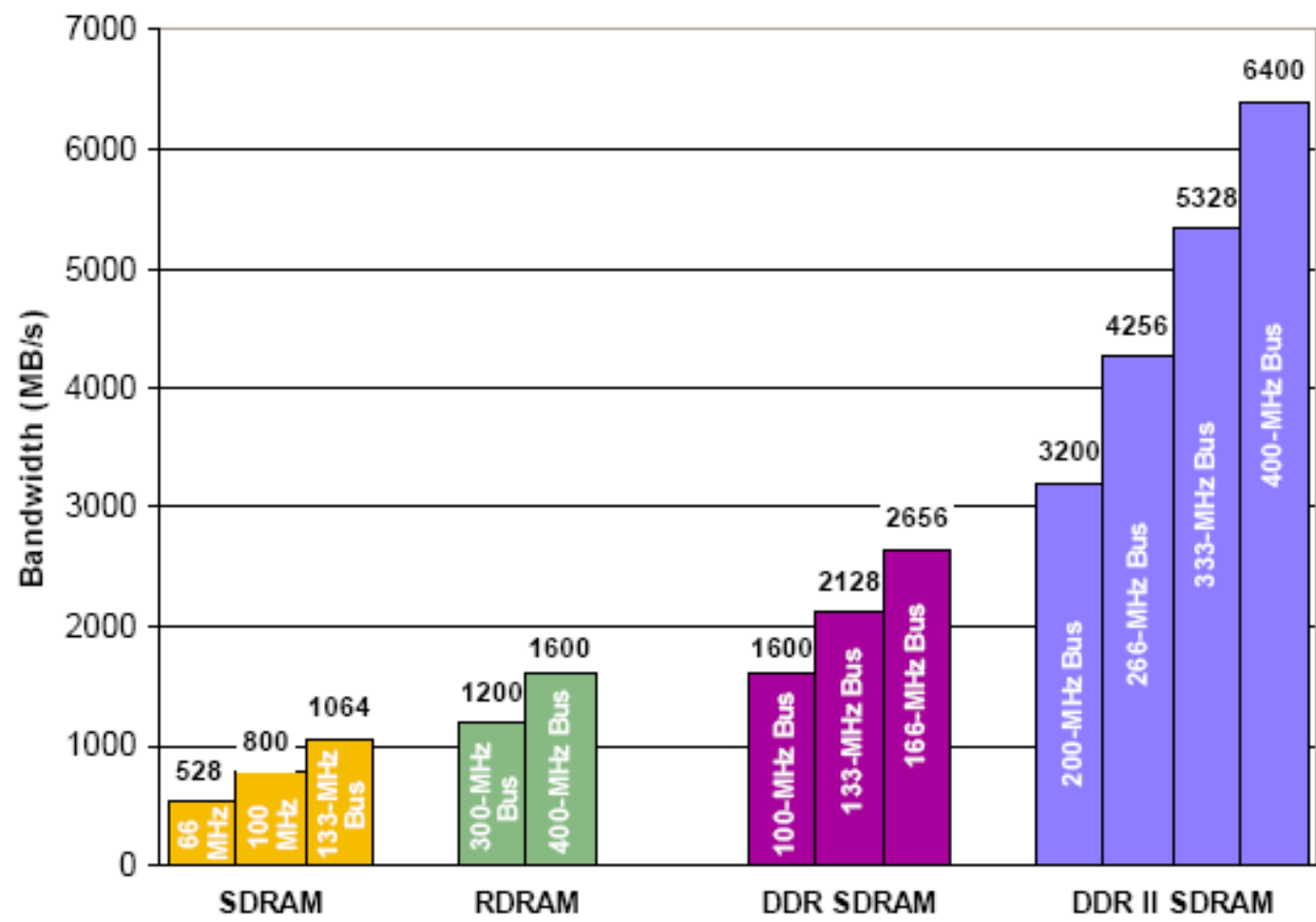


Esquema del principi de funcionament de les transferències de dades en un circuit de memòria DDR2-800.

Esquema del principi de funcionament de les transferències de dades en un circuit de memòria DDR3-1600.



Comparació entre SDRAM, DDR i RDRAM



Comparació d'Amplés de Banda (Bandwidth)

Aquí es comparen els amplex de banda “teòrics” de diferents tipus de RAM.

Tipus de RAM	Ample de Banda (Bandwidth)
SDRAM 100 MHz	100 MHz X 64 bit= 800 MB/sec
SDRAM 133 MHz	133 MHz X 64 bit= 1064 MB/sec
DDRAM 266 MHz	2 X 133 MHz X 64 bit= 2128 MB/sec
DDRAM 333 MHz	2 X 166 MHz X 64 bit= 2656 MB/sec
DDRAM 800 MHz	2 X 400 MHz X 64 bit= 6400 MB/sec
RDRAM 600 MHz	600 MHz X 16 bit= 1200 MB/sec
RDRAM 700 MHz	700 MHz X 16 bit= 1400 MB/sec
RDRAM 800 MHz	800 MHz X 16 bit= 1600 MB/sec

Temporitzacions a les memòries DDR (DDRAM)



- **DDR2-XXX:** XXX indica la màxima freqüència a que pot treballar. DDR2-400 voldria dir que pot treballar fins a 400MHz (compte que la freqüència real és la meitat: 200MHz).
- **PC2-YYYY:** YYYY indica la màxima velocitat de transferència (ample de banda) del mòdul. Ja que els mòduls transfereixen 64bits de cop, es calcula multiplicant la freqüència per 8. Una DDR2-400 seria $400 \times 8 = 3200$ MB/s (PC2-3200).



Com són les temporitzacions a les Memòria DDR

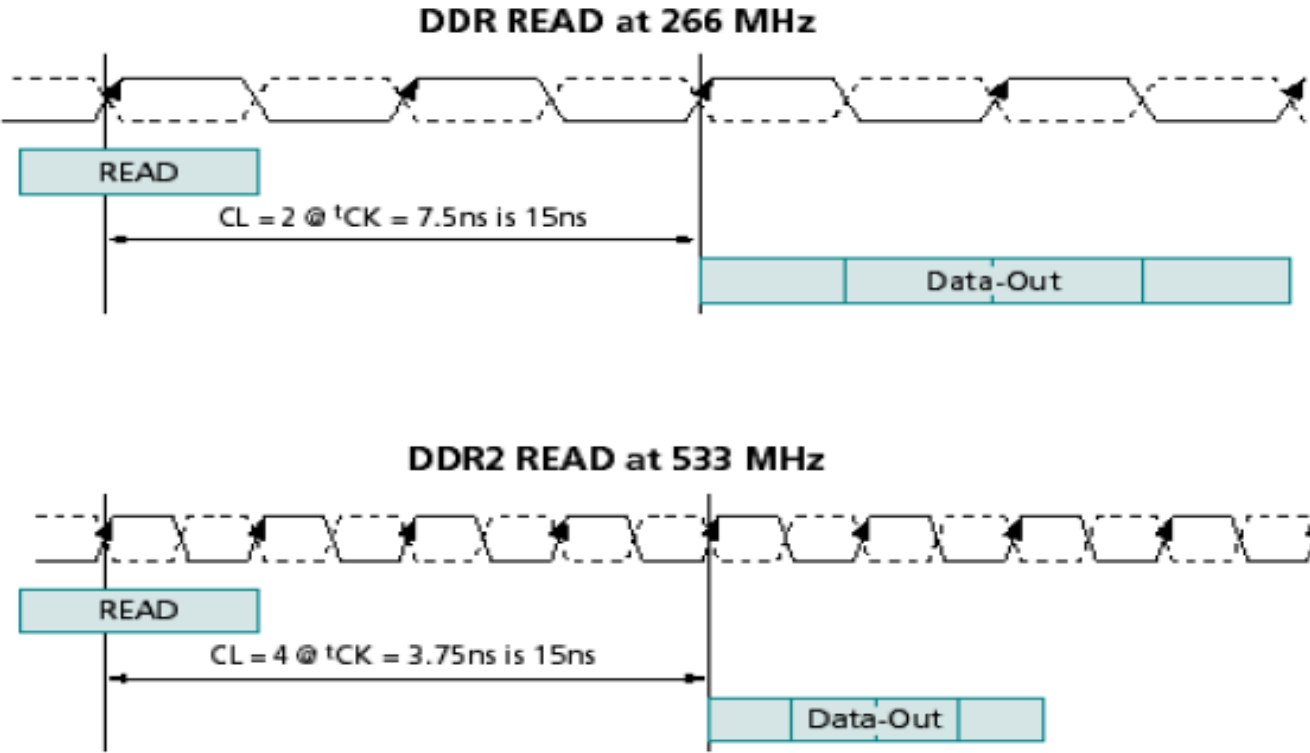


Si ens fixem bé hi ha 4 números més que ens donen més informació (a la imatge són: 2-3-3-6)

- **CL** (CAS Latency).
- **t_{RCD}** (RAS to CAS Delay): Temps que ha de transcorre entre que s'activa el senyal RAS i es pot activar el senyal CAS.
- **t_{RP}** (RAS Precharge): Temps que ha de transcorre entre que es deixa d'accedir a una fila i podem començar l'accés a un altre.
- **t_{RAS}** (Cycle Time).

Aquests temps són importants perquè les velocitats de la transparència anterior són teòriques, suposen que a un cicle de rellotge es fa una transferència, però això només es veritat a partir de la segona transferència a una ràfega. Al començament, o quan es canvia de fila..... S'han de tenir en compte aquests temps.

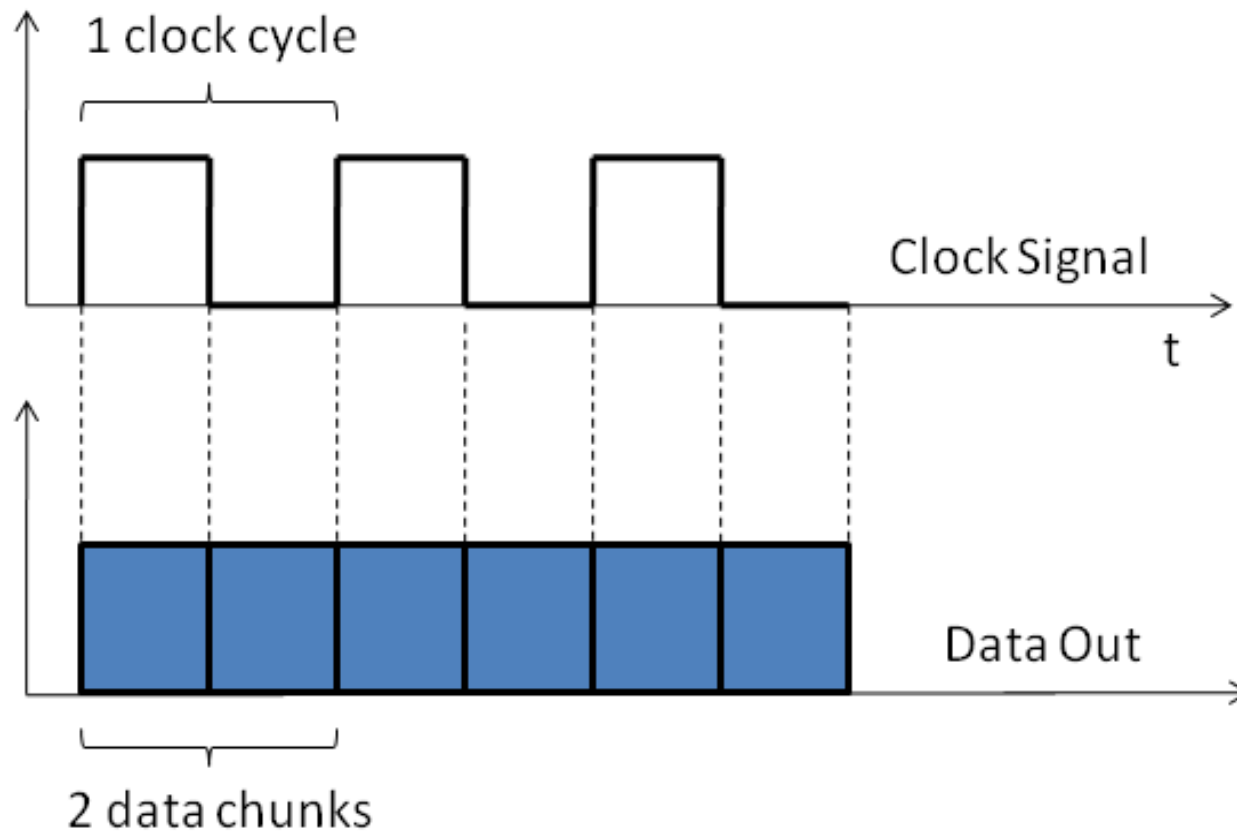
Com són les temporitzacions a les Memòria DDR



Hem d'anar en compte perquè a mida que la freqüència de les memòries augmenta, també ho fan els paràmetres de "latència" i això vol dir que el guany no és tant com ens podria semblar.

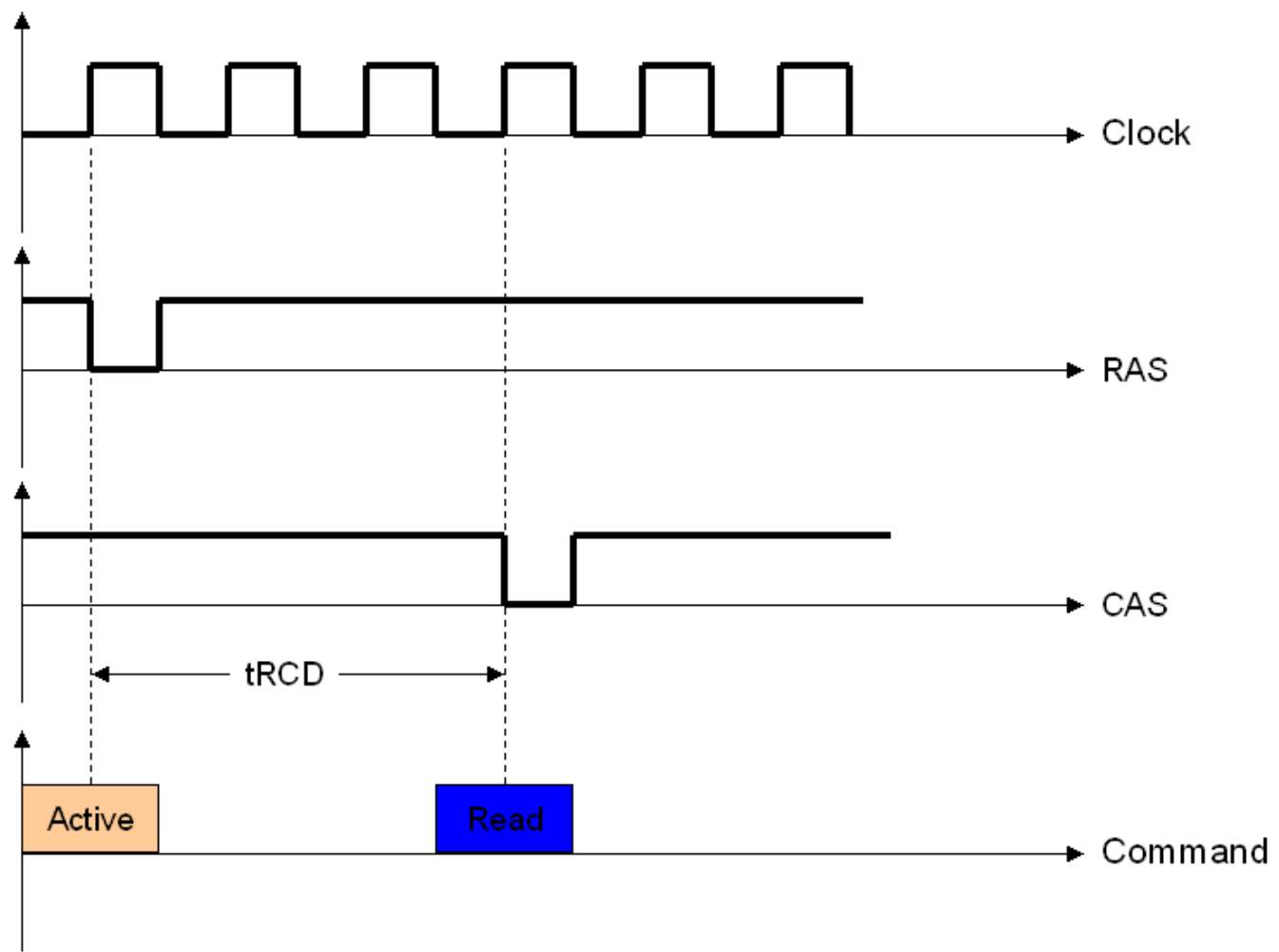
Memòria	DDR2-800	DDR2-1066	DDR3-1066	DDR3-1333	DDR3-1600
Data Rate	800MHz	1066MHz	1066MHz	1333MHz	1600MHz
CAS Latency	4	5	5	7	9
CAS Latency in ns	5 ns	4.69 ns	4.69 ns	5.25 ns	5.63 ns

DDR

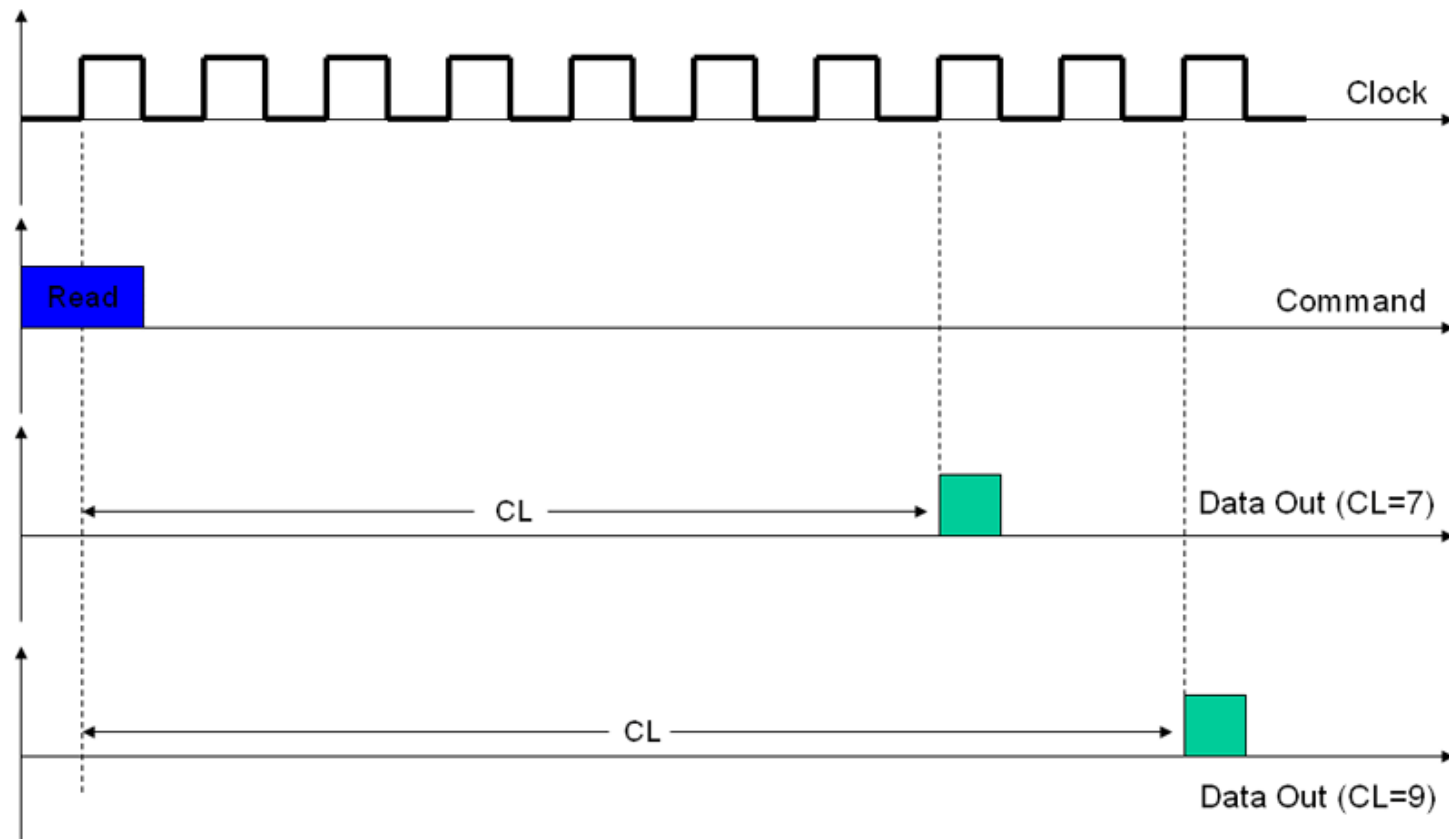


Clock signal and DDR mode

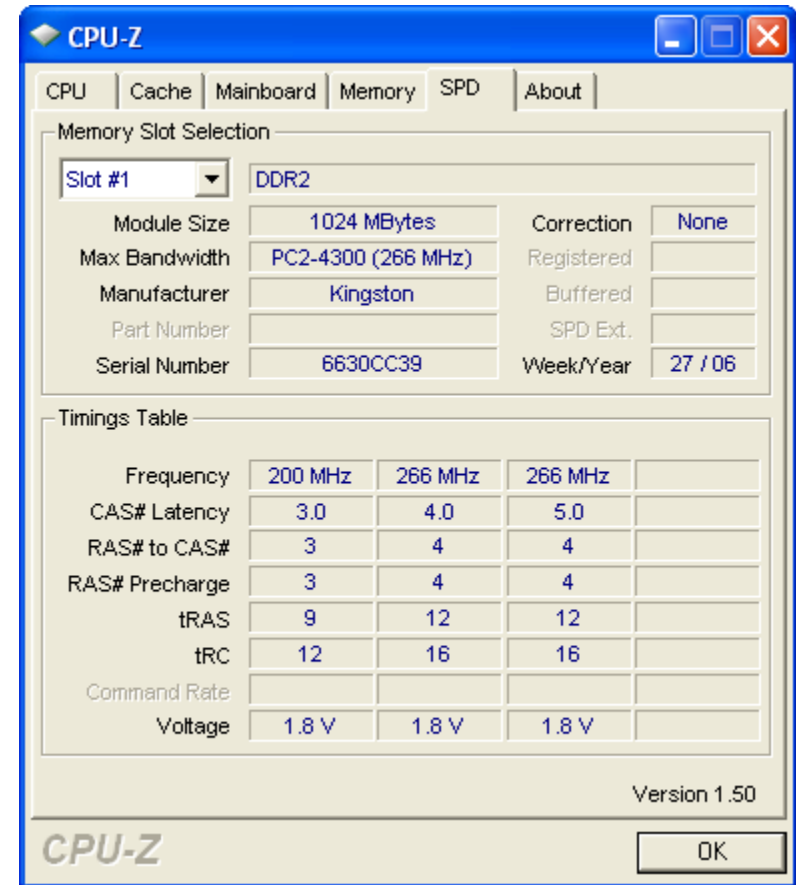
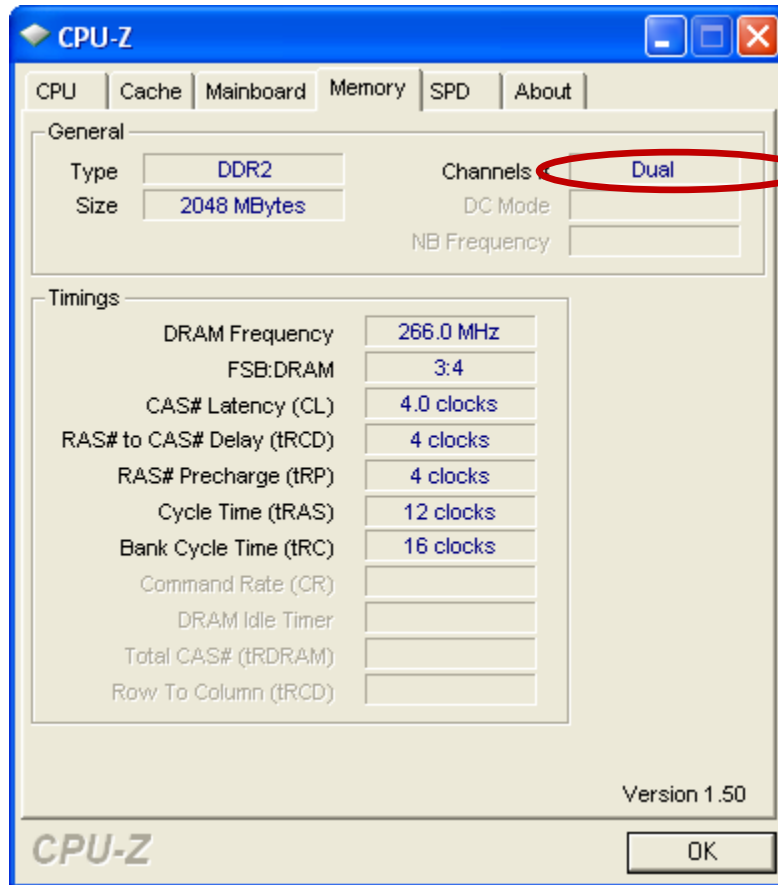
Trcd



CL



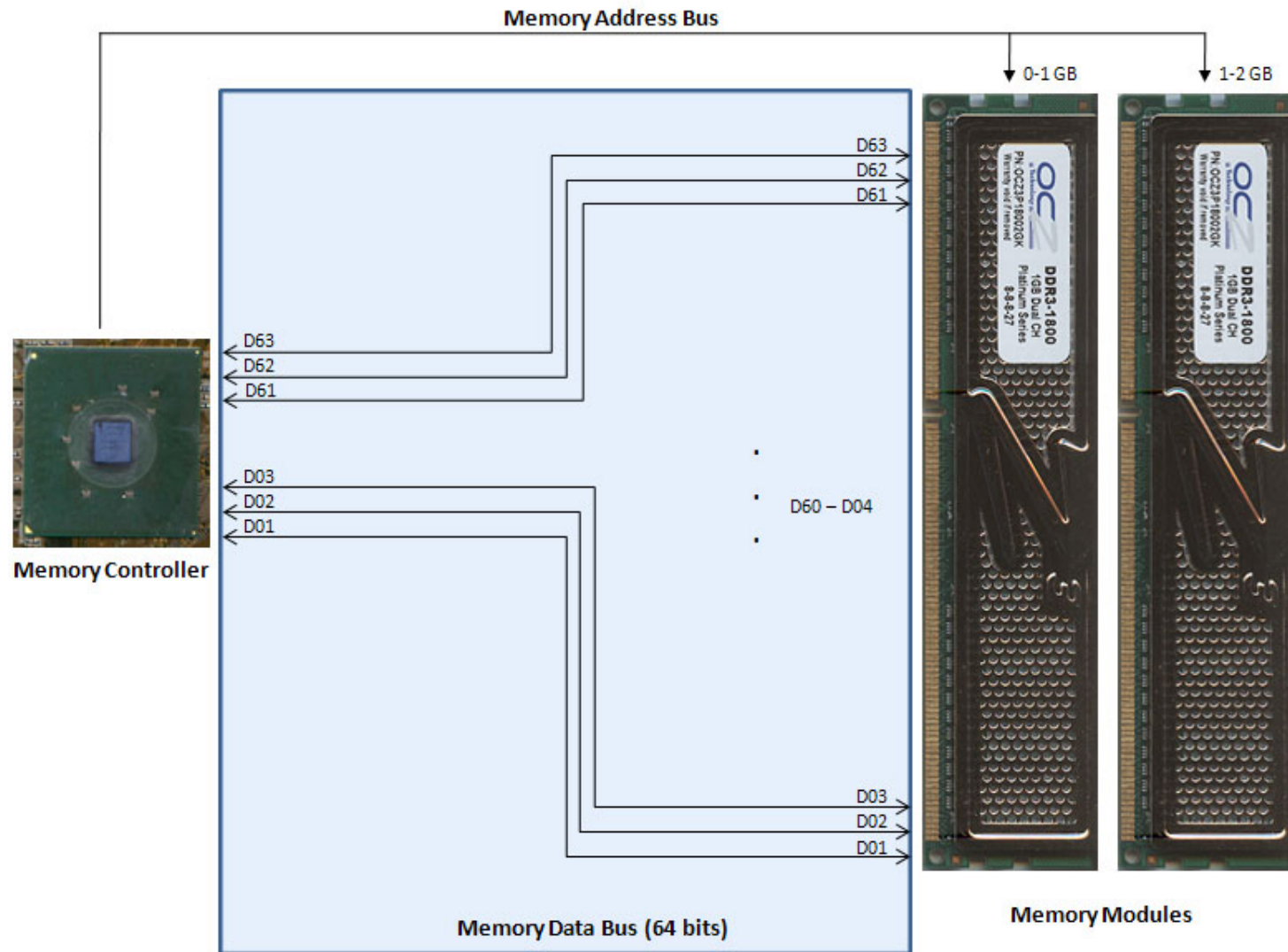
Exemple de temporitzacions a les Memòria DDR



En aquest exemple estem fent servir mòduls DIMM de memòries dinàmiques tipus:

- DDR2-533/PC4300 4-4-4-12 (Freqüència real bus 266MHz)
- Dual-Channel**

DDR Single channel architecture



DDR Dual channel architecture

