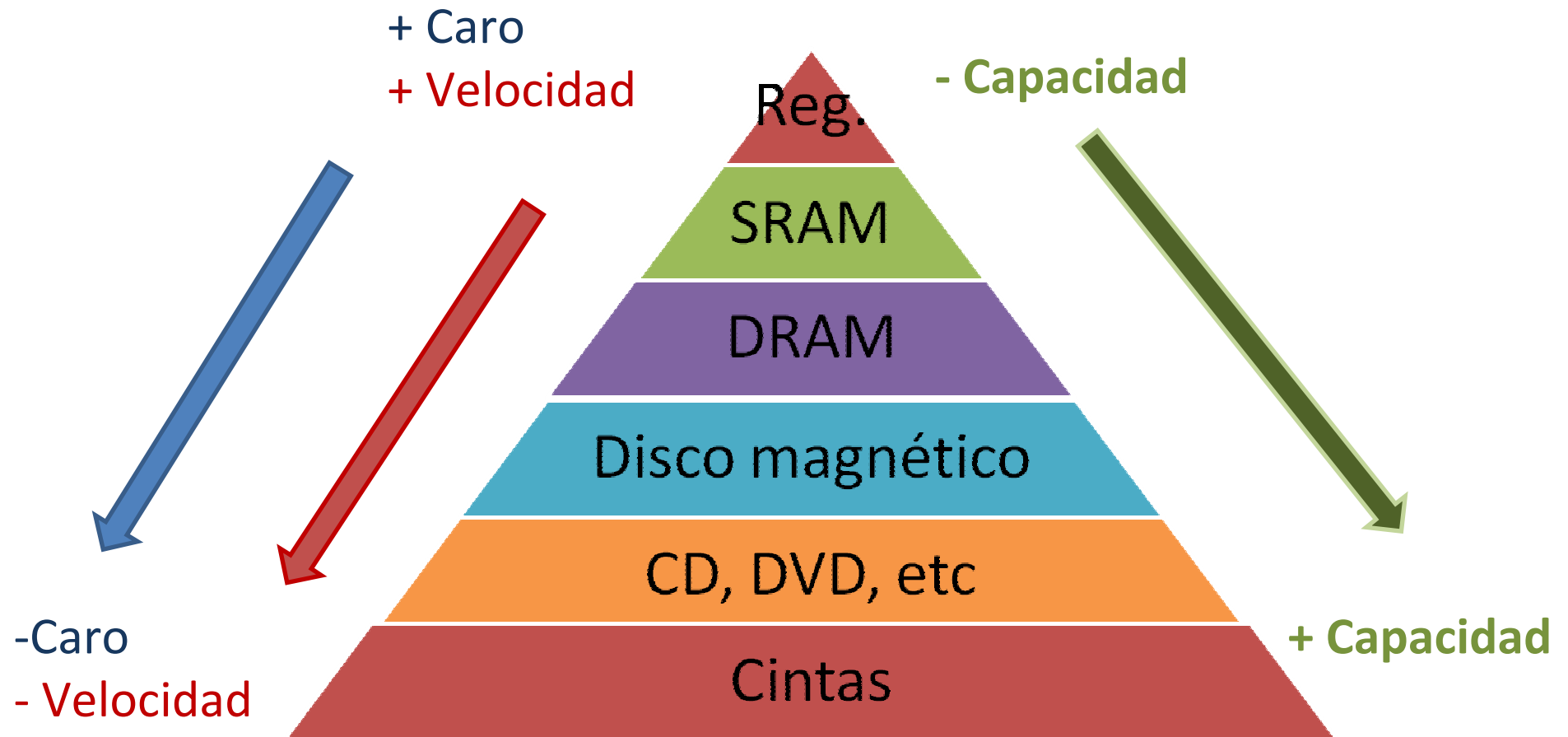


Tema 5. Sistema de memoria

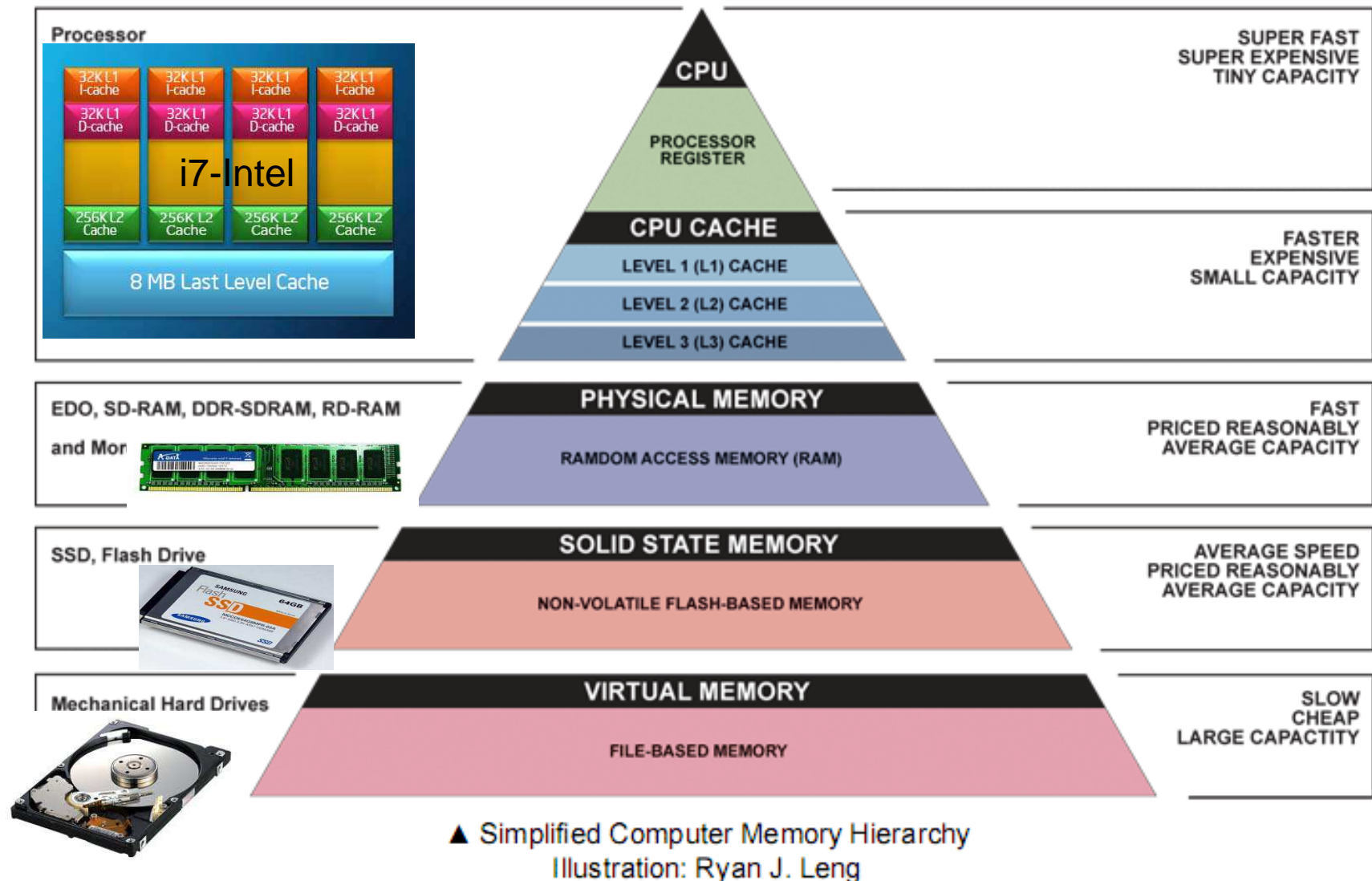
Contenidos

- 1.- Jerarquía de memoria
- 2.- Características de la memoria
- 3.- Memoria RAM dinámica
- 4.- Mapa de memoria
- 5.- Memoria caché

Jerarquía de memoria



Jerarquía de memoria



Características de la memoria

Ubicación

- CPU (registros, cache)
- Interna (RAM)
- Externa (HDD)

Capacidad

- Cantidad de bytes que puede almacenar

Unidad de transferencia

Cantidad de bits que se leen o escriben en una operación de L/E

- Palabra (RAM)
- Bloque (HDD)

Método de acceso

- Secuencial (Cintas)
- Directo (HDD)
- Aleatorio (RAM,ROM)
- Asociativo (Caché)

Tiempo de acceso

- Tiempo que tarda la memoria en suministrar/grabar una palabra desde el momento en que se presenta su dirección (RAM, cache)
- Tiempo que tarda el disco en situar el mecanismo de L/E en una posición (HDD)

Dispositivo físico

- Semiconductor (RAM,ROM,FLASH)
- Soporte magnético (HDD)
- Soporte óptico (DVD,BlueRay)
- Soporte óptico-magnético (Cintas)

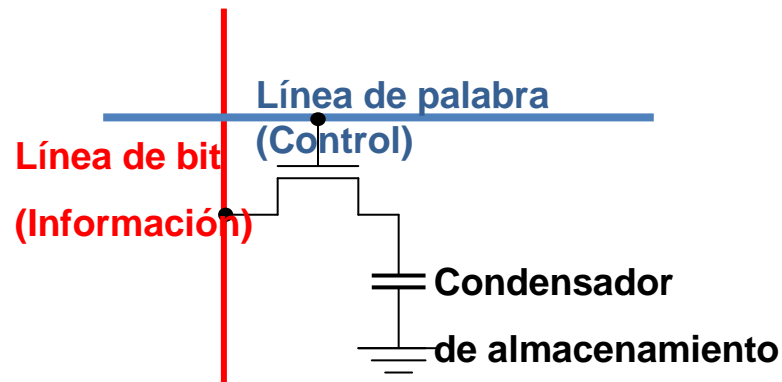
Características físicas

- No volátil (HDD,ROM) / Volátil (RAM)
- No borrrable (ROM) / Borrable (FLASH)

Memorias semiconductoras

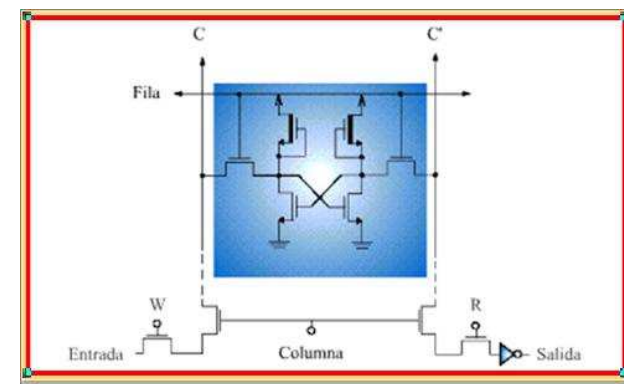
Dinámicas (DRAM)

- Necesita refresco (periodo de mseg.)
- Condensador (ocupa menos espacio que un biestable → mayor densidad)
- Mayor capacidad, más barata
- Memoria Principal



Estáticas (SRAM)

- NO Necesita refresco
- Biestable
- Menor capacidad, más cara
- Caché



DRAM: características

• Organización

- Módulos de memoria (1GB, 2GB, ...)
- Ranks (1, 2, 4)
- Chips de memoria (64Mx8, 128Mx8, ...)
- Bancos internos de un chip (4, 8)

• Prestaciones

- Latencia
- Tiempo de ciclo
- Velocidad de transferencia

• Mejoras

- ECC

- **Celda:** Unidad mínima de almacenamiento (bit, byte)

- **Palabra:** agrupación de celdas que se leen o escriben a la vez (múltiplo de un byte)

• Unidad direccionable

Mínima cantidad de bits a los que le corresponde una dirección física

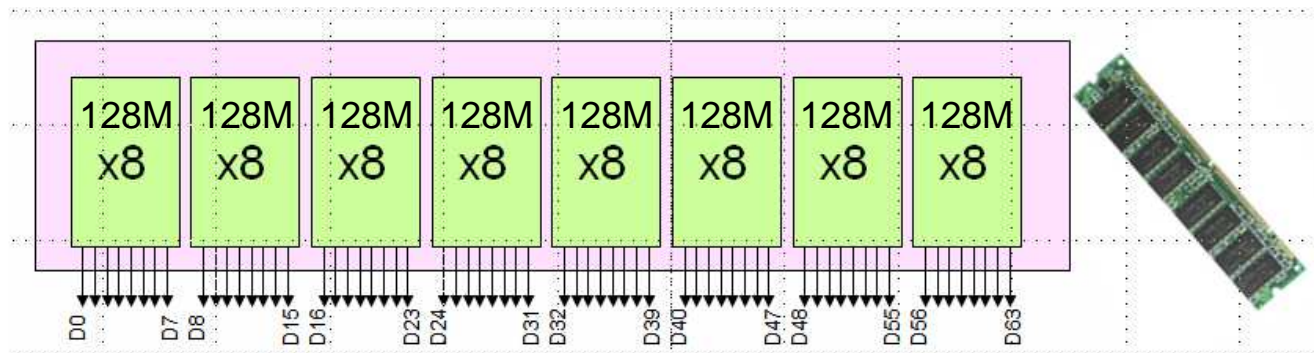
Byte → Dirección física

• Unidad de transferencia

Cantidad de bits que se leen o escriben en una operación de L/E
Palabra (64 bits)

DRAM: organización

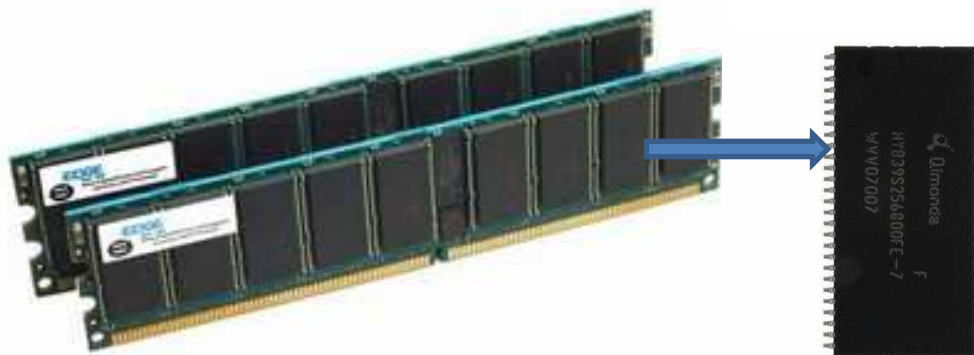
Organización del módulo



¿Cuántas palabras es capaz de almacenar este modulo?

y ¿Cuántos bytes?

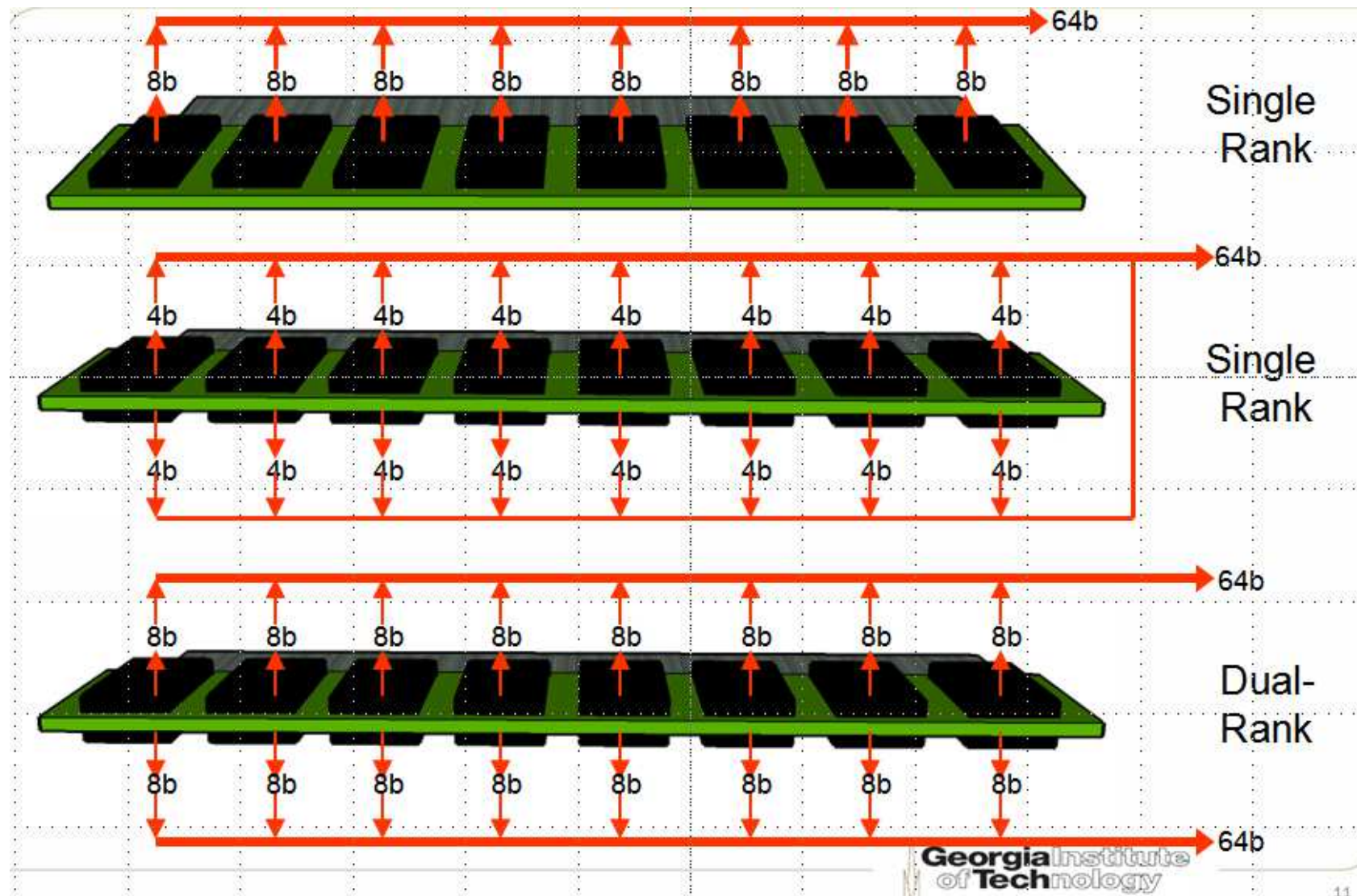
Organización del chip



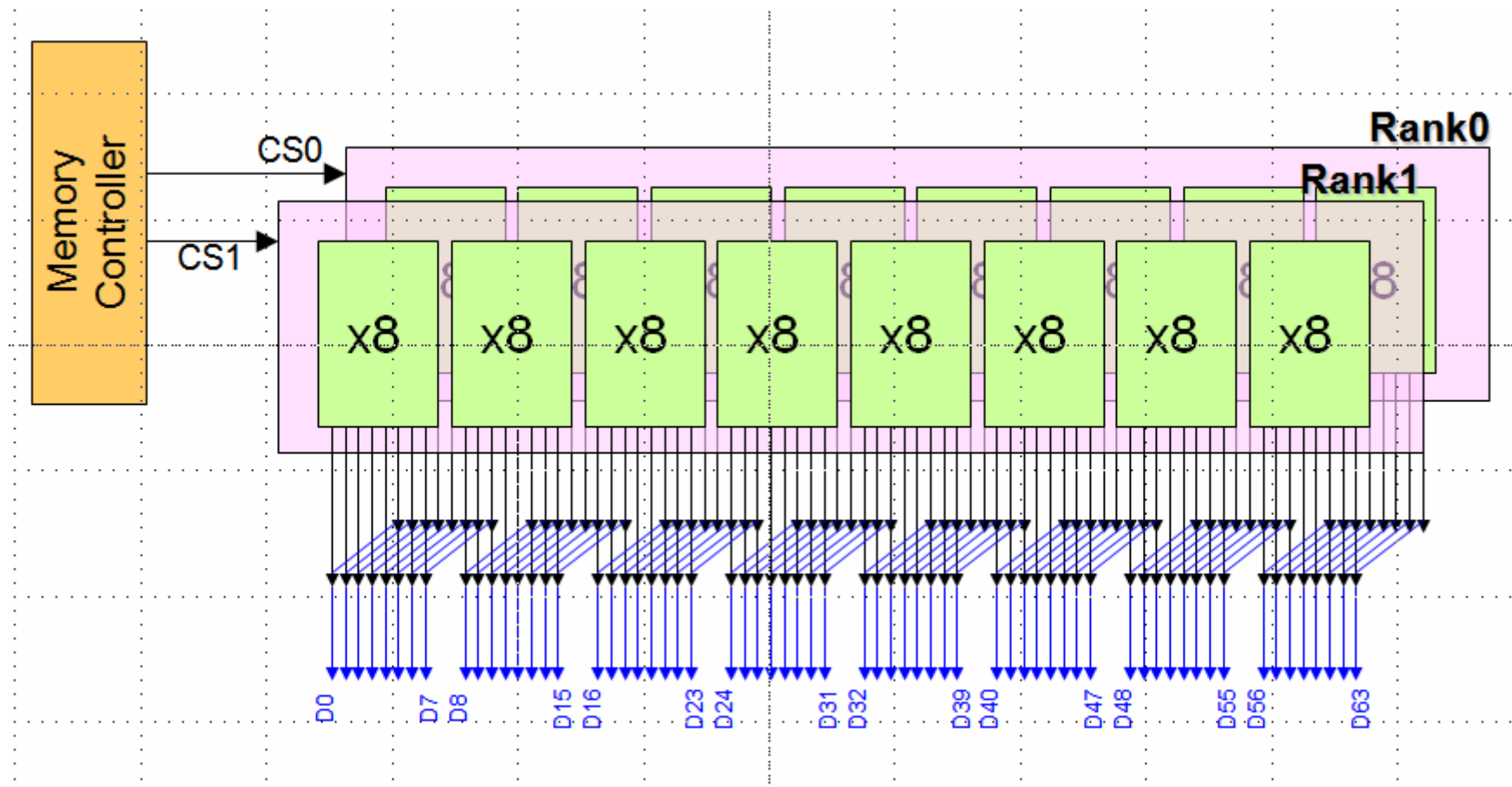
Cantidad de palabras de un chip: 128M, 256M, 512M, 1G

Ancho de palabra de un chip: x1, x4, x8, x16, x32, x64

DRAM: organización en módulos



DRAM: organización en módulos

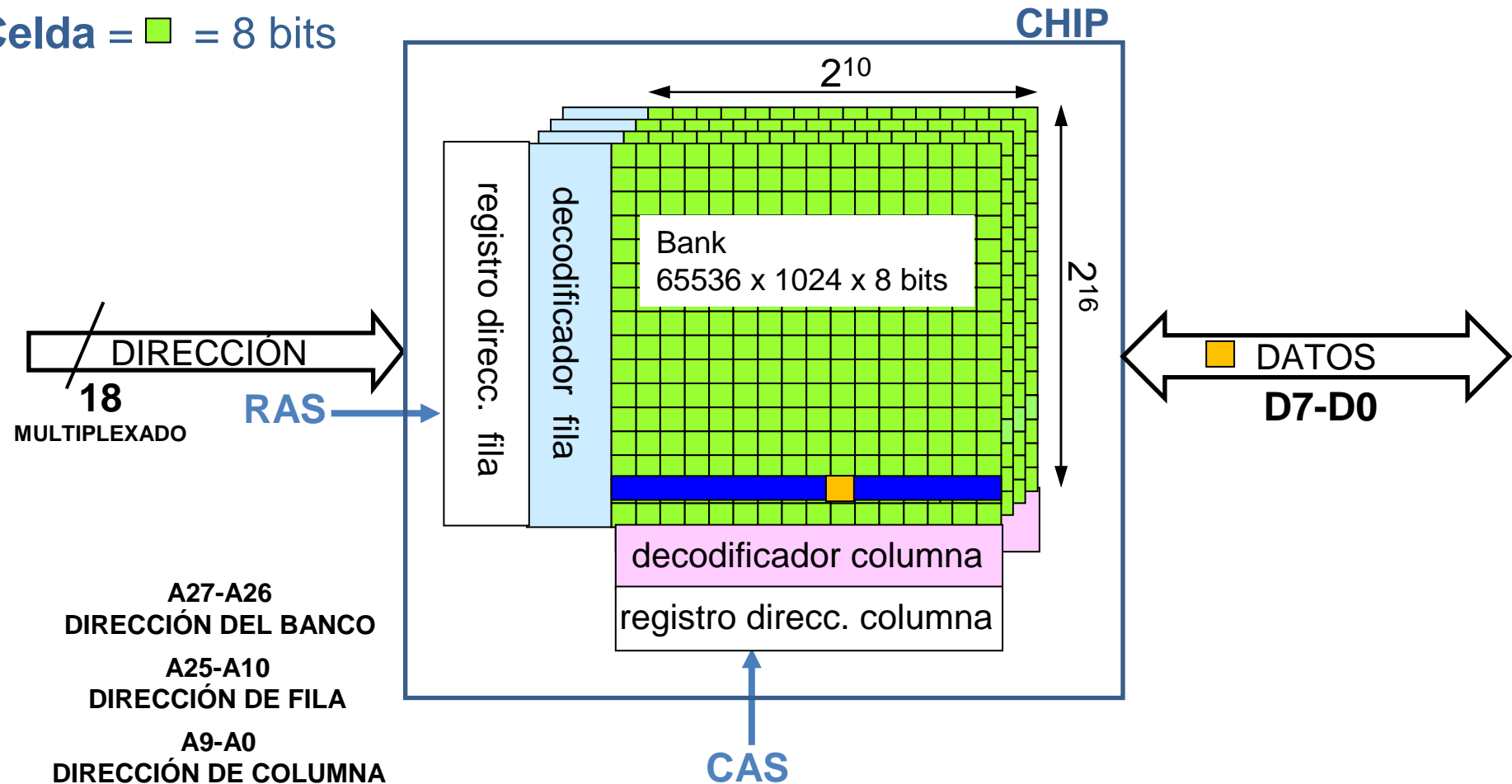


DRAM: organización dentro del chip

Ejemplo: chip 256Mx8 con 4 bancos internos

Banco = 2^{16} filas x 2^{10} columnas = 2^{26} celdas = 64Mx8

Celda = = 8 bits



DRAM: organización dentro del chip

ELPIDA

DATA SHEET

2GB Unbuffered DDR3 SDRAM DIMM

EBJ20UF8BCF0 (256M words × 64 bits, 1 Rank)

Specifications

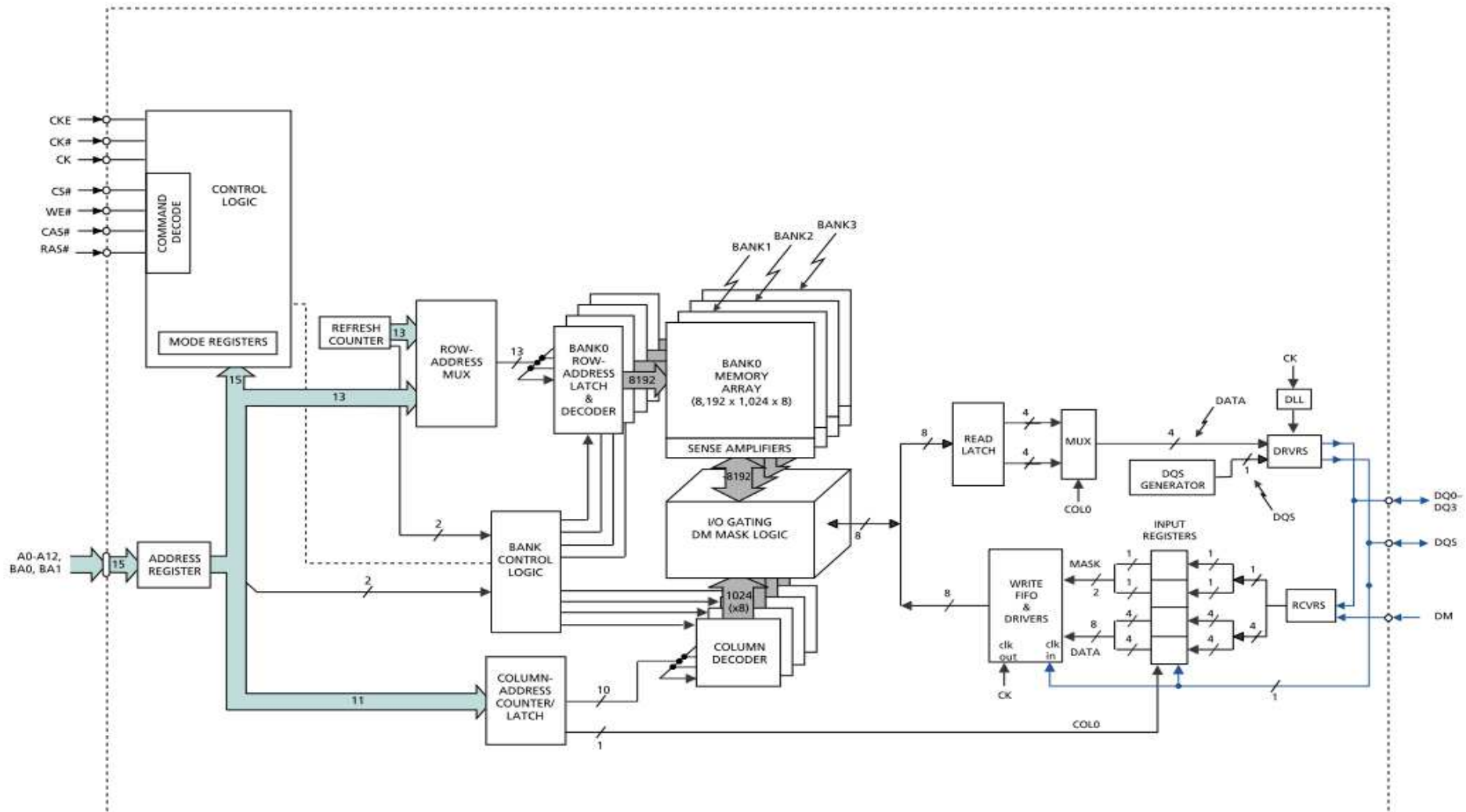
- Density: 2GB
- Organization
 - 256M words × 64 bits, 1 rank
- Mounting 8 pieces of 2G bits DDR3 SDRAM sealed in FBGA
- Package: 240-pin socket type dual in line memory module (DIMM)
 - PCB height: 30.0mm
 - Lead pitch: 1.0mm
 - Lead-free (RoHS compliant) and Halogen-free
- Power supply: VDD = 1.5V ± 0.075V
- Data rate: 1600Mbps/1333Mbps (max.)
- Eight internal banks for concurrent operation (components)

Features

- Double-data-rate architecture: two data transfers per clock cycle
- The high-speed data transfer is realized by the 8 bits prefetch pipelined architecture
- Bi-directional differential data strobe (DQS and /DQS) is transmitted/received with data for capturing data at the receiver
- DQS is edge-aligned with data for READs; center-aligned with data for WRITEs
- Differential clock inputs (CK and /CK)
- DLL aligns DQ and DQS transitions with CK transitions
- Commands entered on each positive CK edge; data and data mask referenced to both edges of DQS
- Data mask (DM) for write data

DRAM: organización dentro del chip

Figure 2: Functional Block Diagram: 64 Meg x 4



DRAM: prestaciones

Tiempo de ciclo: Tiempo entre dos accesos consecutivos a memoria.

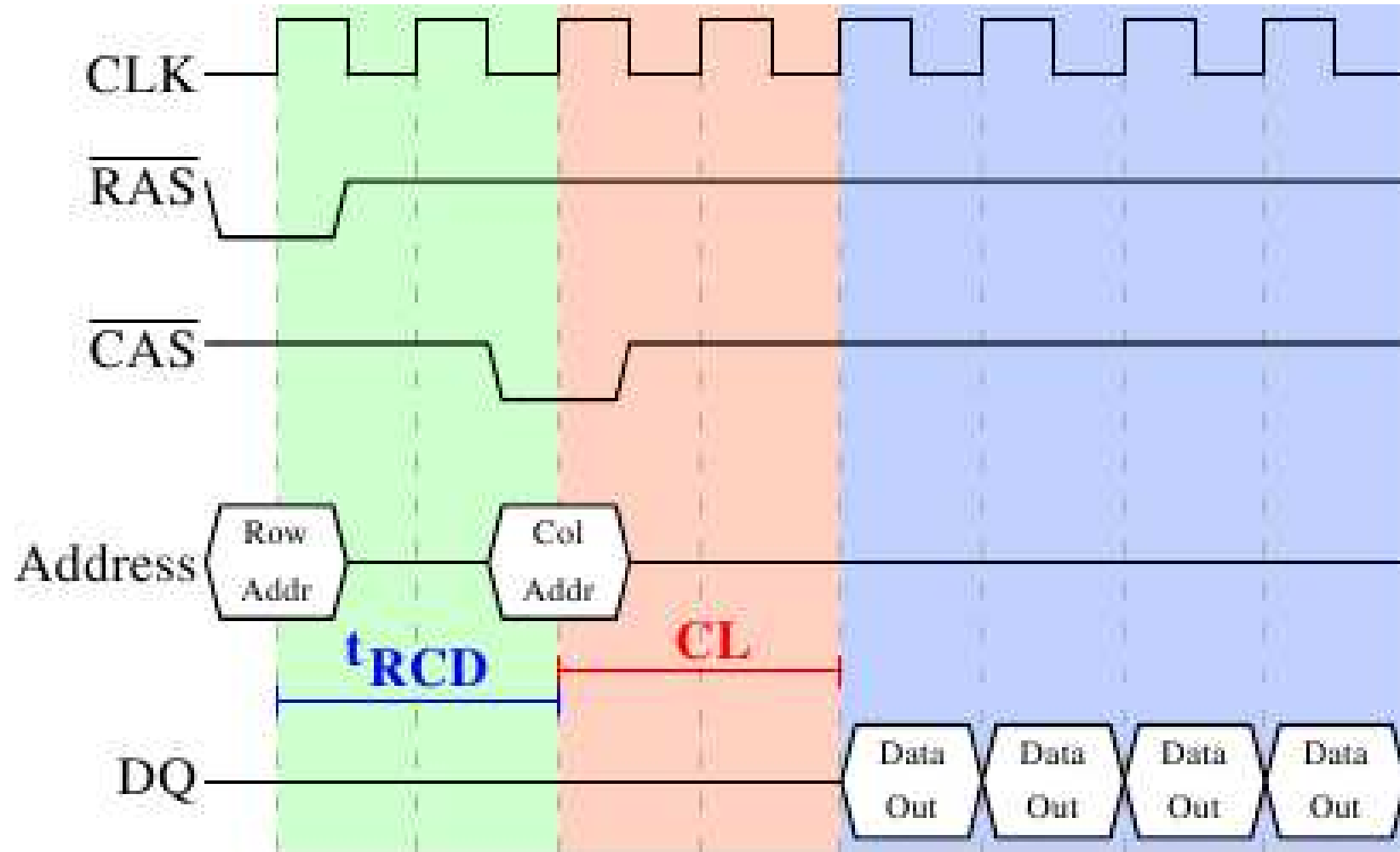
Velocidad de transferencia: es la velocidad a la que la memoria transfiere los datos, expresada en bytes por segundo.

Latencia: Tiempo transcurrido desde que el controlador de memoria solicita un dato hasta que éste está disponible en los pines de salida.

La latencia se produce por los retardos que provocan los accesos a los diferentes componentes de la memoria. Existen varias latencias, la más importante es la latencia CAS (CL) que es el número de ciclos de reloj necesarios para acceder a una columna del banco interno del chip y proporcionar el dato en los pines de salida.

DRAM: prestaciones

Timing SDRAM



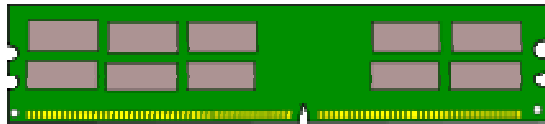
DRAM: tipo SDRAM-DDR

SDRAM (Synchronous DRAM): su funcionamiento está sincronizado con una señal de reloj.

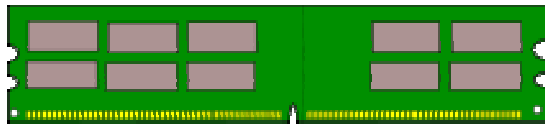
DDR (Double Data Rate): dos accesos por cada ciclo de reloj.



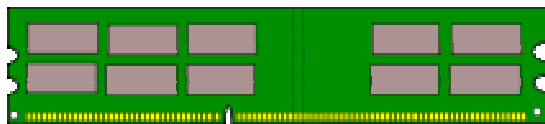
DDR



DDR 2



DDR 3



| Nombre estándar - Frecuencia efectiva | Frecuencia Array (CAS,RAS) | Frecuencia buffer E/S | Palabras transferidas por segundo | Nombre del módulo | Máxima capacidad de transferencia |
|---------------------------------------|----------------------------|-----------------------|-----------------------------------|-------------------|-----------------------------------|
| DDR2-400 | 100 MHz | 200 MHz | 400 millones | PC2-3200 | 3200 MB/s |
| DDR2-533 | 133 MHz | 266 MHz | 533 millones | PC2-4200 | 4264 MB/s |
| DDR2-600 | 150 MHz | 300 MHz | 600 millones | PC2-4800 | 4800 MB/s |
| DDR2-667 | 166 MHz | 333 MHz | 667 Millones | PC2-5300 | 5336 MB/s |
| DDR2-800 | 200 MHz | 400 MHz | 800 Millones | PC2-6400 | 6400 MB/s |
| DDR2-1000 | 250 MHz | 500 MHz | 1000 Millones | PC2-8000 | 8000 MB/s |
| DDR2-1066 | 266 MHz | 533 MHz | 1066 Millones | PC2-8500 | 8530 MB/s |
| DDR2-1150 | 286 MHz | 575 MHz | 1150 Millones | PC2-9200 | 9200 MB/s |
| DDR2-1200 | 300 MHz | 600 MHz | 1200 Millones | PC2-9600 | 9600 MB/s |

x2

Palabras/seg = frecuencia x palabras/ciclo =


= 200 millones ciclos/seg x 2 palabras/ciclo = 400 millones de palabras/seg

Velocidad de Transferencia (MB/seg) = palabras/seg x bytes/palabra =

= 400 Mpalabras/seg x 8 bytes/palabra = 3.200 MB/seg


DRAM: tipo SDRAM-DDR

x4

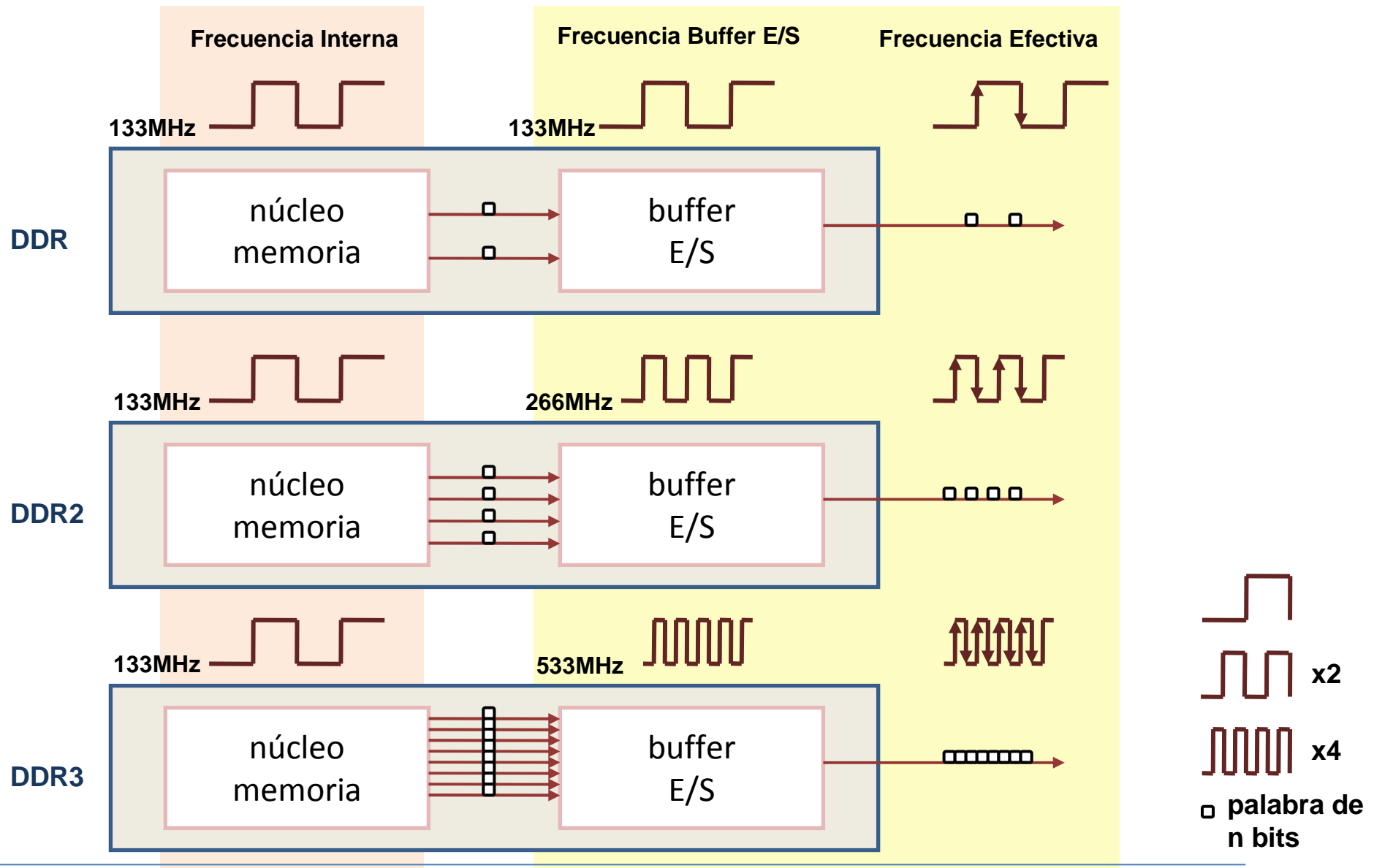


| Nombre estándar – frecuencia efectiva | Frecuencia Array (CAS,RAS) | Frecuencia buffer E/S | Palabras transferidas por segundo | Nombre del módulo | Máxima capacidad de transferencia |
|---------------------------------------|----------------------------|-----------------------|-----------------------------------|-------------------|-----------------------------------|
| DDR3-1066 | 133 MHz | 533 MHz | 1066 Millones | PC3-8500 | 8530 MB/s |
| DDR3-1200 | 150 MHz | 600 MHz | 1200 Millones | PC3-9600 | 9600 MB/s |
| DDR3-1333 | 166 MHz | 667 MHz | 1333 Millones | PC3-10667 | 10664 MB/s |
| DDR3-1375 | 170 MHz | 688 MHz | 1375 Millones | PC3-11000 | 11000 MB/s |
| DDR3-1466 | 183 MHz | 733 MHz | 1466 Millones | PC3-11700 | 11700 MB/s |
| DDR3-1600 | 200 MHz | 800 MHz | 1600 Millones | PC3-12800 | 12800 MB/s |
| DDR3-1866 | 233 MHz | 933 MHz | 1866 Millones | PC3-14900 | 14930 MB/s |
| DDR3-2000 | 250 MHz | 1000 MHz | 2000 Millones | PC3-16000 | 16000 MB/s |

x2



DRAM: tipo SDRAM-DDR



DRAM: tipo SDRAM-DDR

| Nombre estándar | Frec. Array de celdas (CAS,RAS) | Frecuencia buffer E/S | Palabras transferidas por segundo | Nombre del módulo | Máxima capacidad de transferencia (Mbytes/s) |
|-----------------|---------------------------------|-----------------------|-----------------------------------|-------------------|--|
| DDR3-1066 | 133 MHz | 533 MHz | 1066 Millones | PC3-8500 | 8530 MB/s |
| DDR3-1200 | 150 MHz | 600 MHz | 1200 Millones | PC3-9600 | 9600 MB/s |

Memoria 1 $f_{\text{array}} = 133 \text{ MHz}$, $L=6$

1) Duración ciclo array = $1/f_{\text{array}} = 7.5 \text{ ns}$

Duración hasta la salida de la 1ª palabra = $6 \times 7.5 \text{ ns} = 45\text{ns}$

¿cómo influye la latencia?

Memoria 2 $f_{\text{array}} = 133 \text{ MHz}$, $L=9$

2) Duración ciclo array = $1/f_{\text{array}} = 7.5 \text{ ns}$

Duración hasta la salida de la 1ª palabra = $9 \times 7.5 \text{ ns} = 67.5\text{ns}$

Memoria 3 $f_{\text{array}} = 150 \text{ MHz}$, $L=7$

3) Duración ciclo array = $1/f_{\text{array}} = 6.7 \text{ ns}$

Duración hasta la salida de la 1ª palabra = $7 \times 6.7 \text{ ns} = 46.9\text{ns}$

Mapa de memoria

- **Mapa de memoria:** Todo el espacio direccionable por el computador.
bus de direcciones = n bits \rightarrow mapa de memoria = 2^n direcciones

Pentium Dual Core E6000

Bus de direcciones = 36 bits \rightarrow mapa de memoria = 2^{36} direcciones \rightarrow si cada byte de la memoria lleva asociado una dirección \rightarrow se pueden direccionar 64 GB

- **Mapa disponible:** subconjunto del mapa para el que existe memoria física (módulos DRAM, ROM BIOS, etc).

- **Espacio libre:** ampliaciones.



Mapa de memoria: ejemplo 1

Intel 286 (bus direcciones = 24 bits) con 2 módulos RAM de 1MB

1º) mapa de memoria = 2^{24} direcciones → si cada byte de la memoria lleva asociado una dirección → se pueden direccionar 16 MB max.

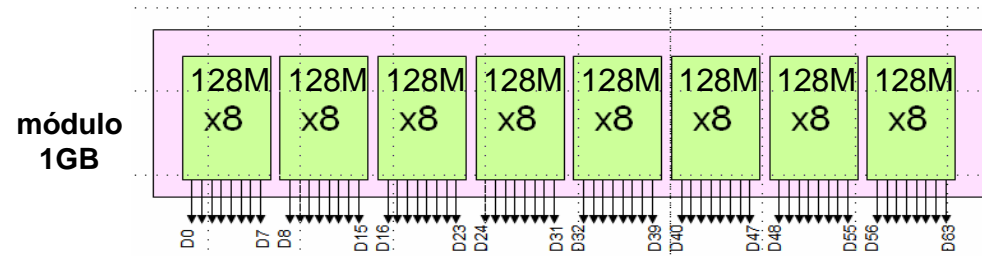
2º) para direccionar 1MB necesitamos 20 bits del bus de direcciones (utilizamos los menos significativos A19-A0).

| A23-A20 | A19-A16 | A15-A12 | A11-A8 | A7-A4 | A3-A0 | Función |
|---------|---------|---------|--------|-------|-------|-----------------------|
| 0 | 0 | 0 | 0 | 0 | 0 | Módulo1 |
| 0 | F | F | F | F | F | |
| 1 | 0 | 0 | 0 | 0 | 0 | Módulo 2 |
| 1 | F | F | F | F | F | |
| 2 | 0 | 0 | 0 | 0 | 0 | Libre (Bios, I/O,) |
| F | F | F | F | F | F | |

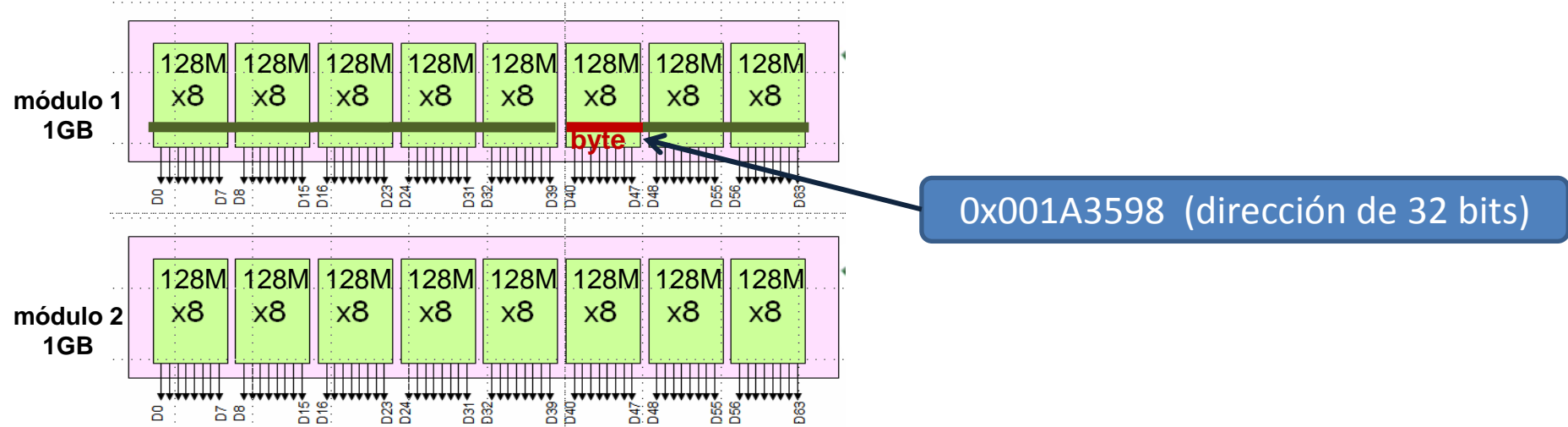
Mapa de memoria: ejemplo 2

Características del ordenador:

- Bus de direcciones = 32 bits
- Palabra = 64 bits
- Módulo RAM = 1GB
- N^o módulos RAM = 2

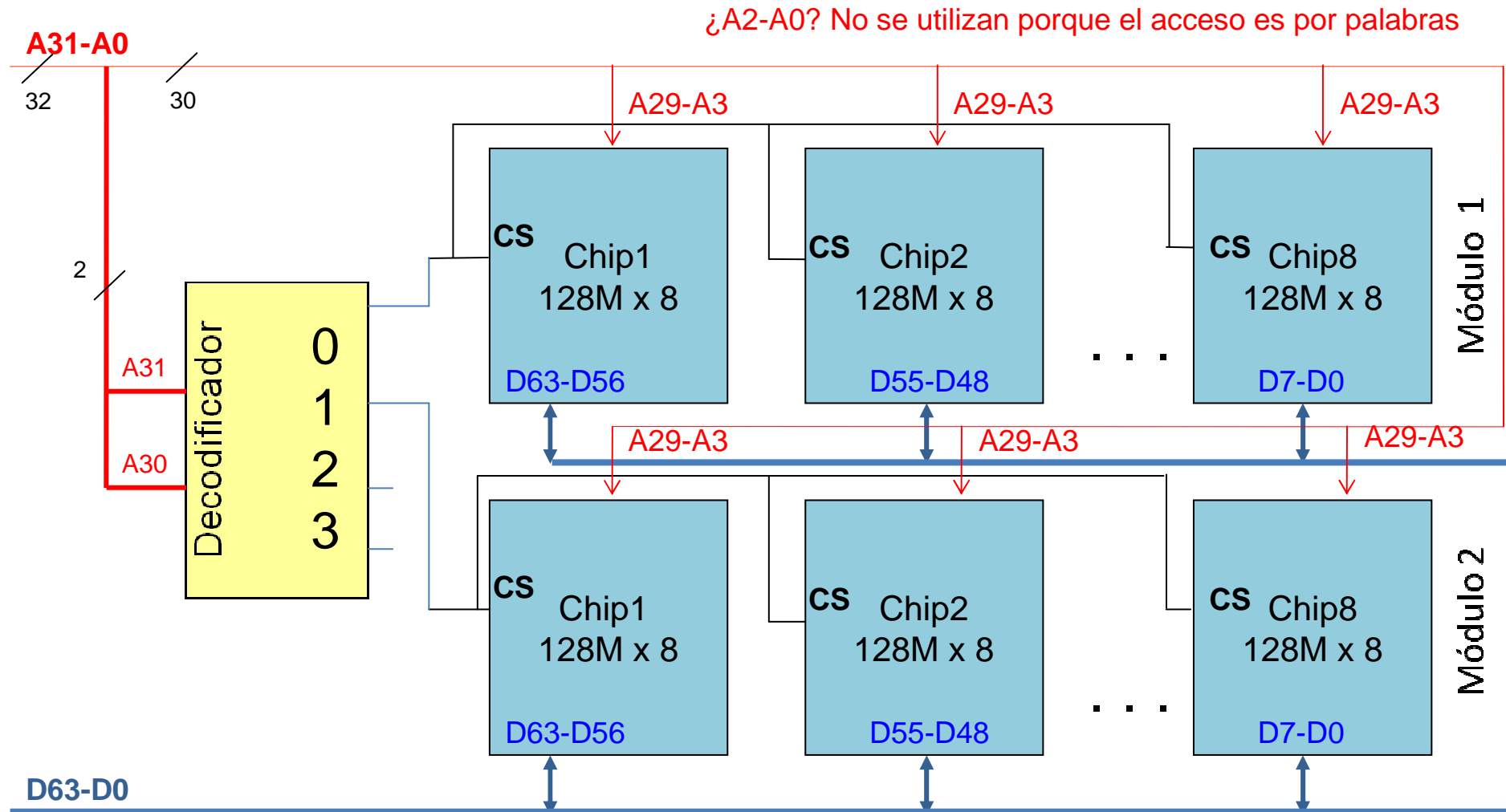


Mapa de memoria: ejemplo 2



- 1º) si una palabra tiene 8 bytes (2^3)
necesitamos 3 bits para identificar el byte en la palabra
utilizaremos las 3 líneas menos significativas del bus de direcciones **(A2-A0)**
- 2º) si un módulo tiene 128 Mpalabras (2^{27})
necesitamos 27 bits para identificar la palabra en el módulo
utilizaremos las 27 líneas siguientes del bus de direcciones **(A29-A3)**
- 3º) ¿y las líneas que sobran?
las utilizaremos para identificar el módulo **(A31-A30)**

Mapa de memoria: ejemplo 2



Mapa de memoria: ejemplo 2

| | A31 A30 ¿Qué Módulo? | A29-A3 (128 Megapalabras) ¿Qué palabra del módulo? | A2-A0 ¿Qué byte? | Mód | Dirección Hex. |
|--------|-------------------------|---|---------------------|-------|-------------------|
| 1ªDir. | 00 | 00 0000 0000 0000 0000 0000 0000 0 | 000 | 1 | 00000000 |
| Última | 00 | 11 1111 1111 1111 1111 1111 1111 1 | 111 | 1 | 3FFF FFFF |
| 1ªDir. | 01 | 00 0000 0000 0000 0000 0000 0000 0 | 000 | 2 | 40000000 |
| Última | 01 | 11 1111 1111 1111 1111 1111 1111 1 | 111 | 2 | 7FFF FFFF |
| 1ªDir. | 10 | 00 0000 0000 0000 0000 0000 0000 0 | 000 | Libre | 80000000 |
| Última | 11 | 11 1111 1111 1111 1111 1111 1111 1 | 111 | | FFFF FFFF |

bus direcciones = 32 bits →

mapa de memoria = 2^{32} direcciones →

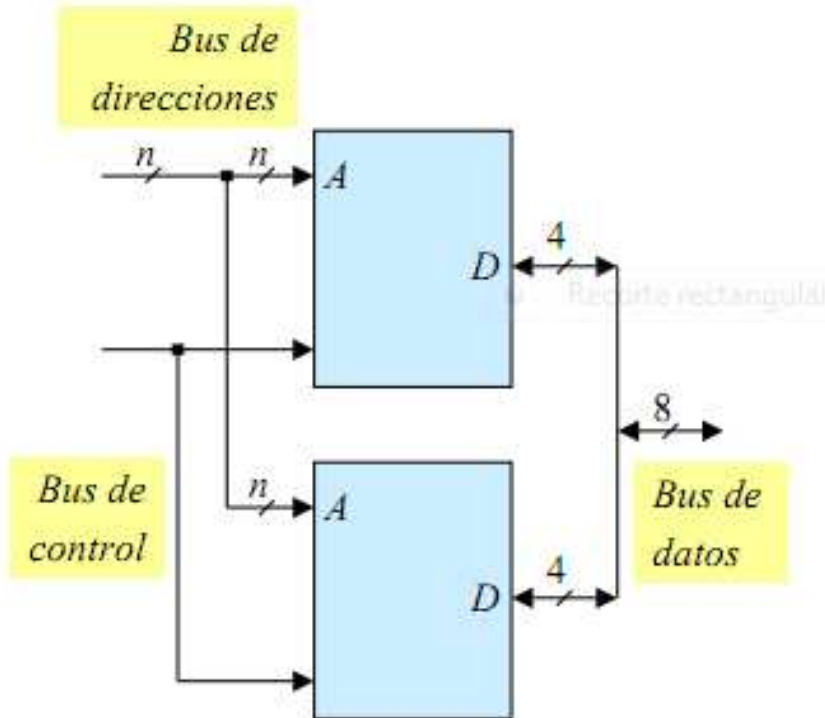
si cada byte de la memoria lleva asociado una dirección →

se pueden direccionar 4GB como máximo

Expansión de memoria

Longitud de palabra

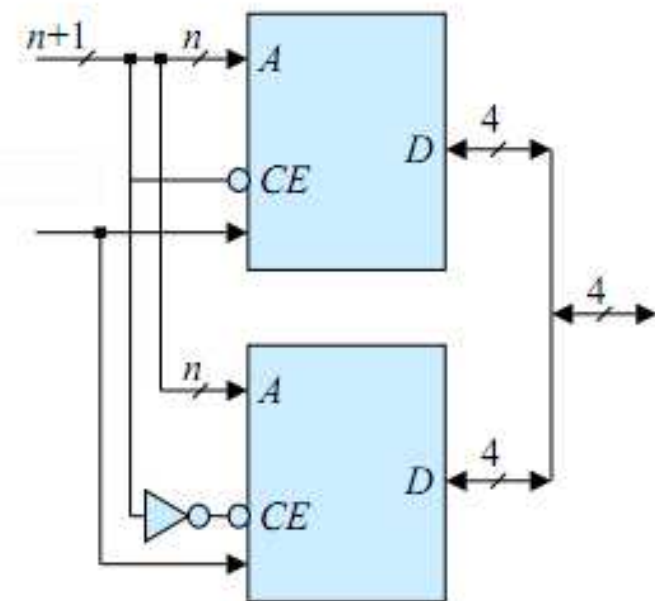
- Número bits en cada dirección



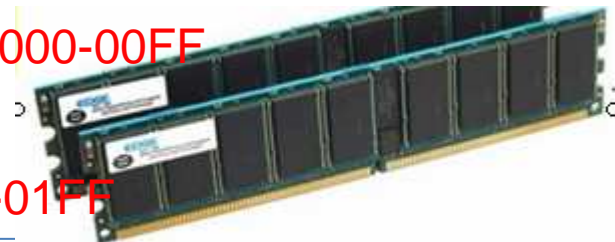
8 bits 8 bits

Capacidad de palabra

- Número de direcciones



Dir: 0000-00FF

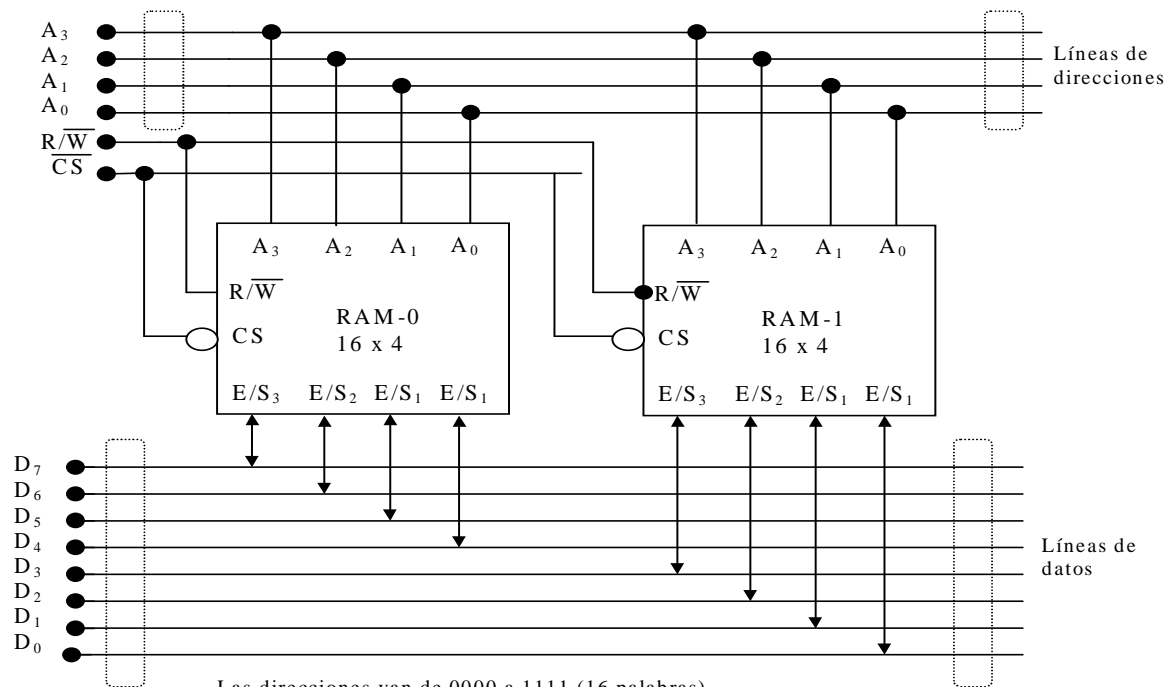


Dir: 0100-01FF

Expansión de memoria

Incrementar la longitud de palabra

Combinación de dos RAM de 16 x 4 para formar una memoria de 16 x 8



Las direcciones van de 0000 a 1111 (16 palabras)
Tamaño de palabra, 8 bits

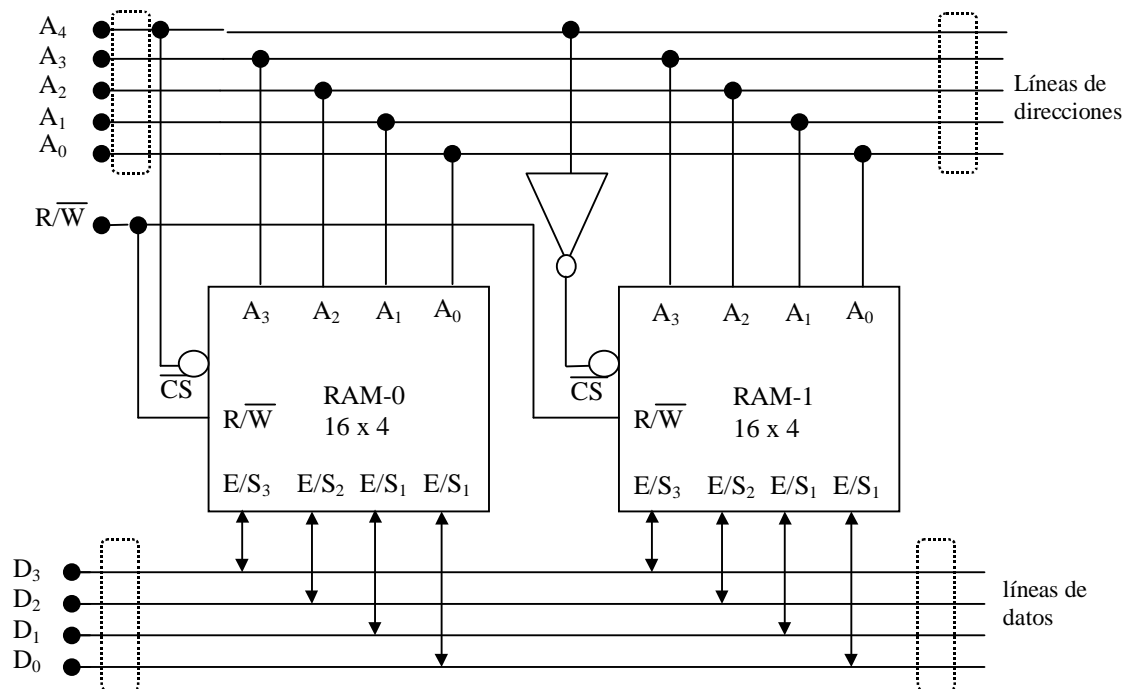
Los 4 bits de más peso de
cada palabra de almacenan
en RAM-0

Los 4 bits de menos peso
de cada palabra de
almacenan en RAM-1

Expansión de memoria

Incrementar el número de palabras de memoria

Combinación de dos RAM de 16 x 4 para formar una memoria de 32 x 4

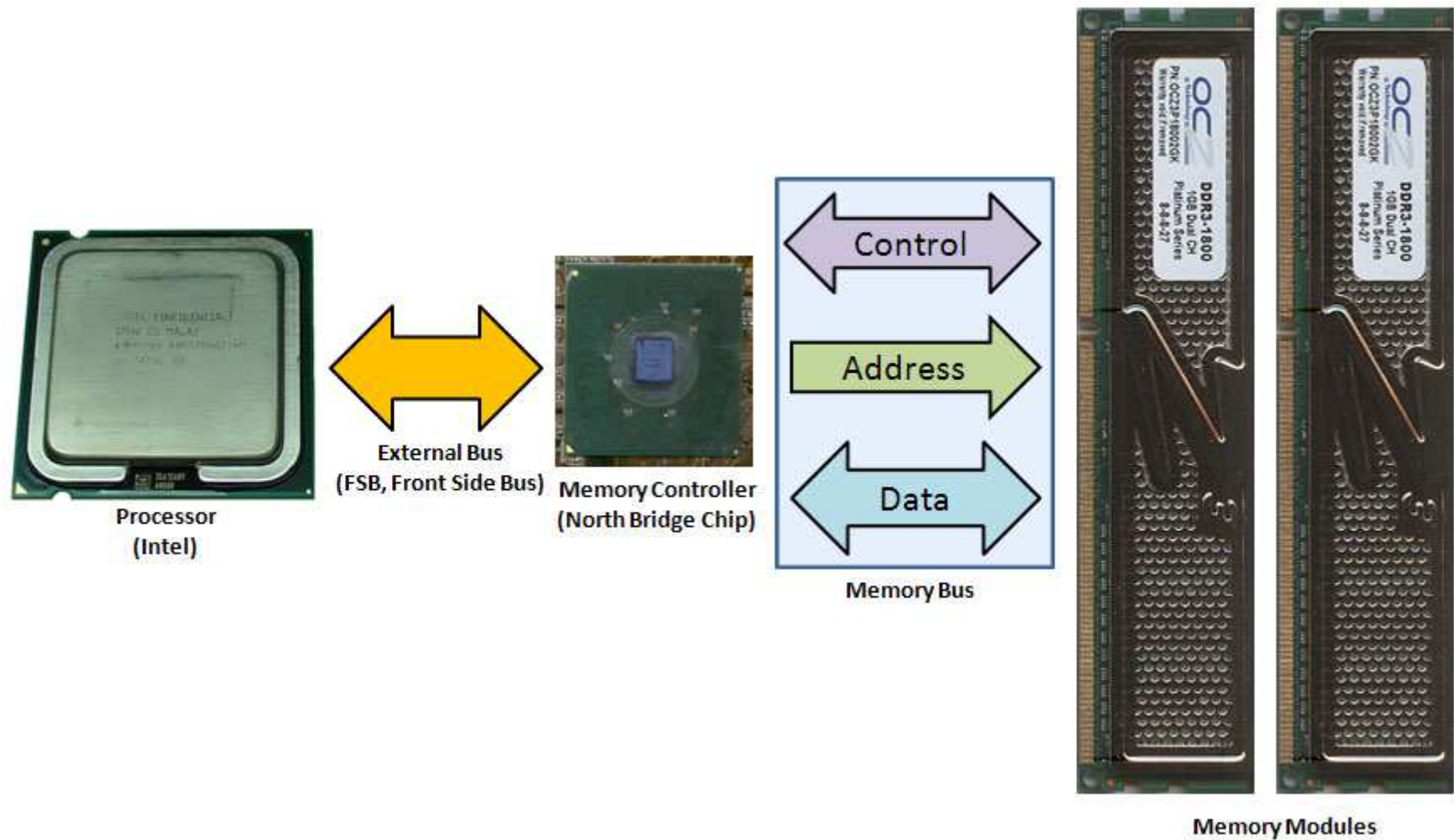


Intervalos de direcciones

00000 a 01111 – 16 palabras en RAM-0

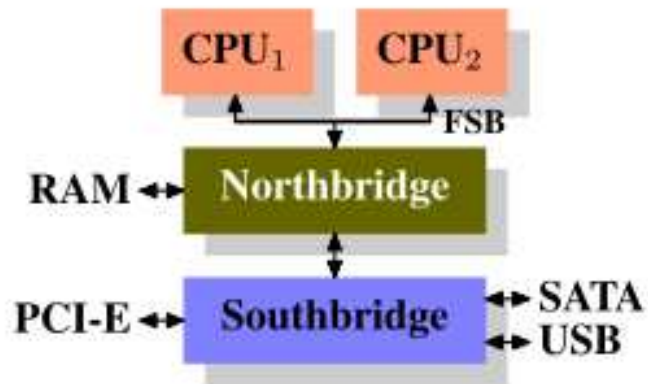
10000 a 11111 – 16 palabras en RAM-1

Comunicación Procesador-Memoria (DRAM)

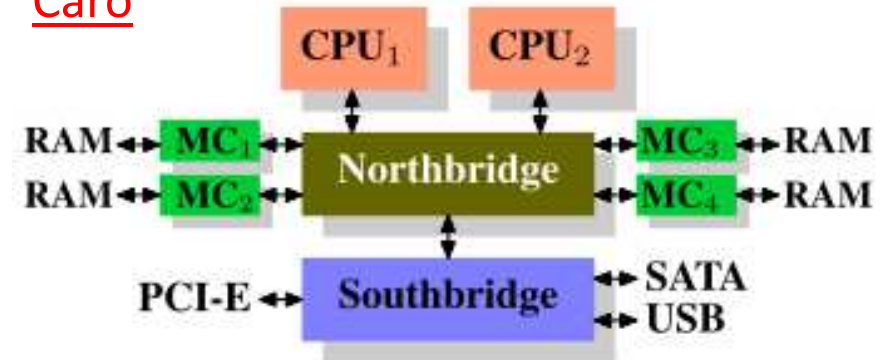


Comunicación Procesador-Memoria (DRAM)

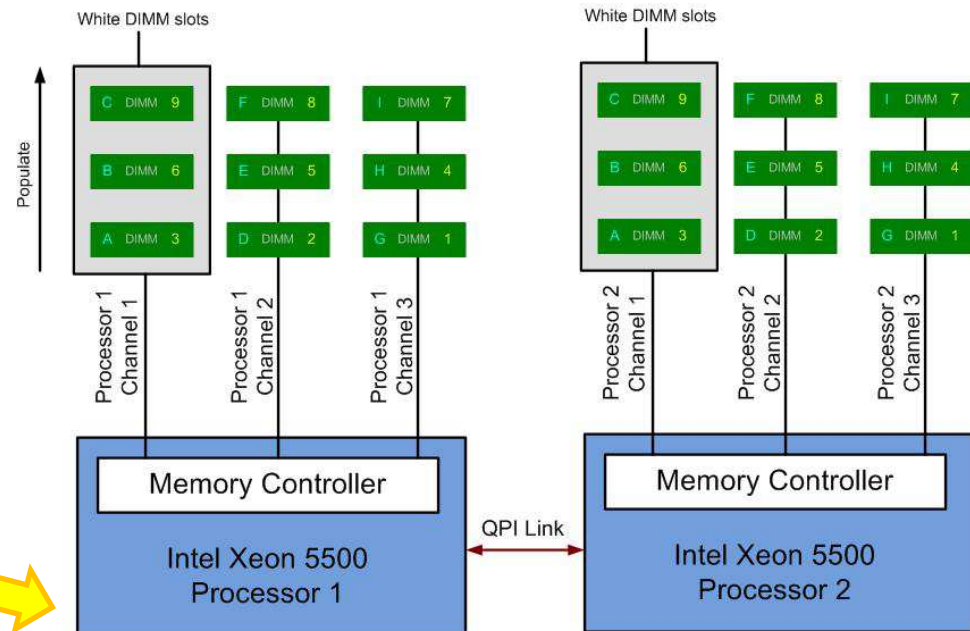
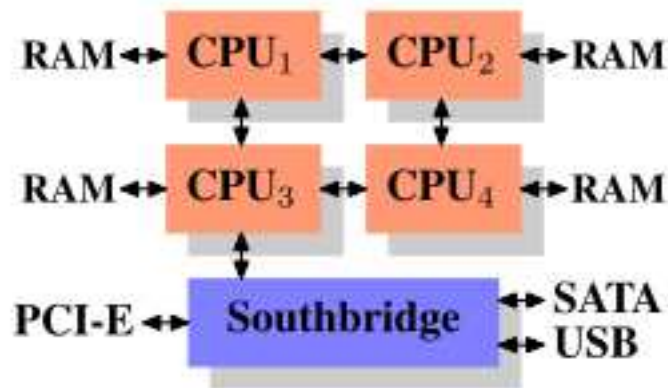
Clásico



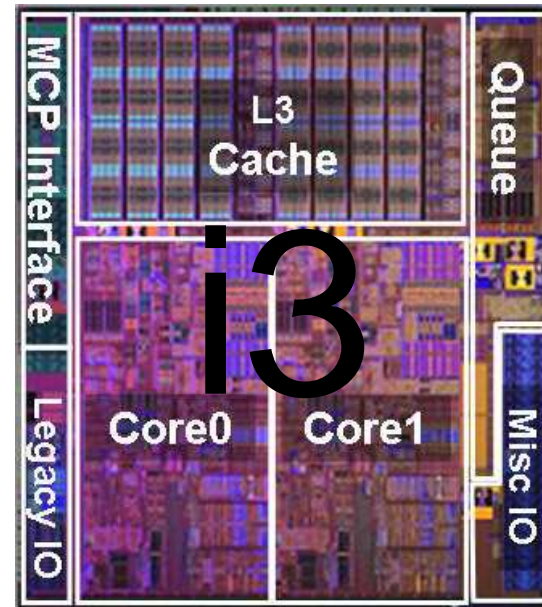
Caro



Actual (NUMA)

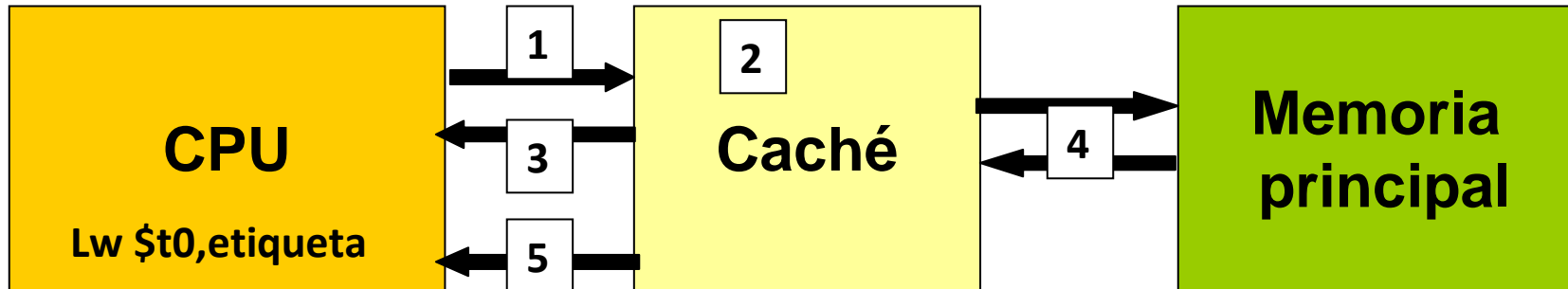


Memoria caché: niveles



- más
lenta ↓
- **L1:** Está dentro del núcleo. (64KB-256KB)
 - **L2:** Está fuera del núcleo y hay una por cada núcleo. (256KB-4MB)
 - **L3:** Solo hay una que comparten todos los núcleos. (8MB)

Memoria caché: funcionamiento



1 La CPU solicita un dato.

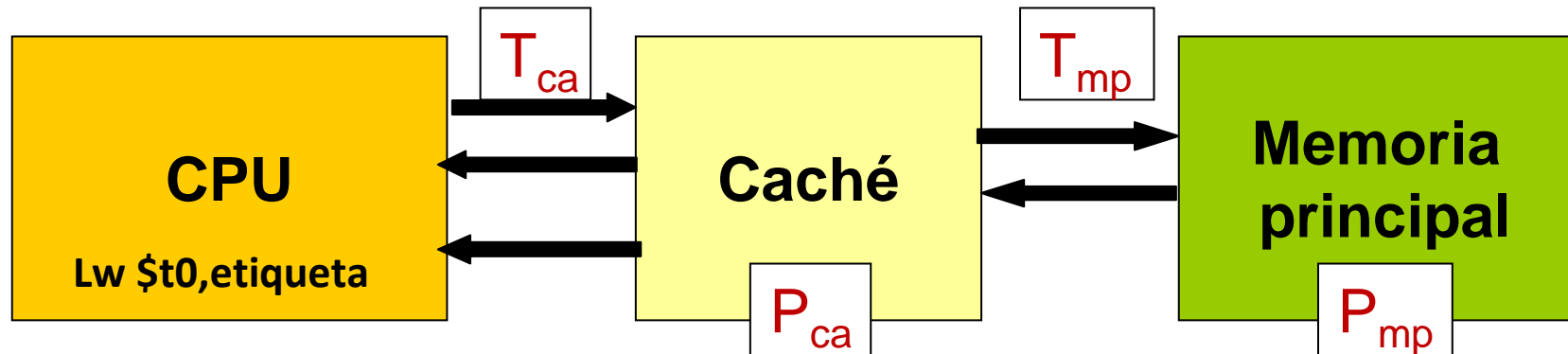
2 La caché comprueba si tiene ese dato

3 Si la caché tiene ese dato lo entrega a la CPU

Si la caché NO tiene ese dato {

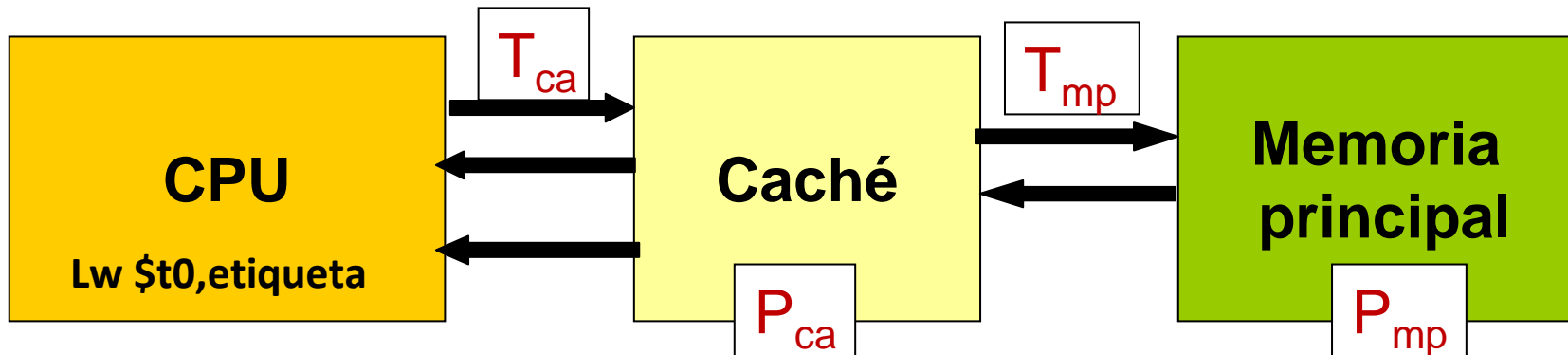
- 4** Lo solicita a memoria principal
- 5** Lo entrega a la CPU

Memoria caché: funcionamiento



$$\left. \begin{array}{l} T_m = P_{ca} * T_{ca} + P_{mp} * T_{mp} \\ P_{mp} = 1 - P_{ca} \end{array} \right\} T_{medio} = P_{ca} * T_{ca} + (1 - P_{ca}) * (T_{mp} + T_{ca})$$

Memoria caché: funcionamiento



$$T_{\text{medio}} = P_{ca} * T_{ca} + (1 - P_{ca}) * (T_{mp} + T_{ca})$$

Ejemplo:

- Tiempo de acceso a caché igual a 10ns
- Tiempo de acceso a memoria principal igual a 120ns
- Probabilidad de que el dato/inst. esté en la caché igual a 80%

$$T_{\text{medio}} = 0,80 * 10\text{ns} + (1 - 0,80) * (10\text{ ns} + 120\text{ns}) = 8\text{ns} + 26\text{ns} = 34\text{ns}$$