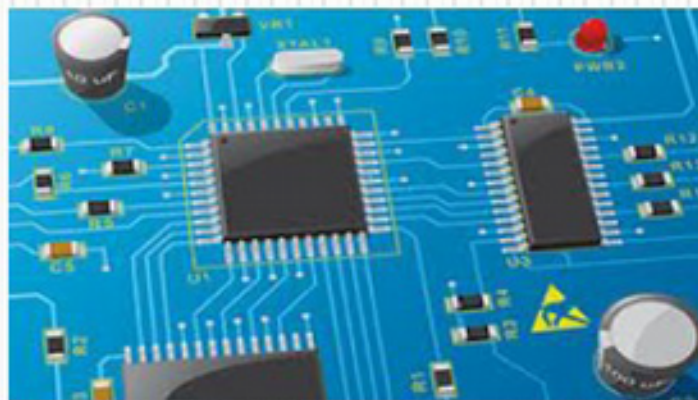




## T3. Sistema de Memoria:

### 3.2 Memoria Principal (RAM)

#### FUNDAMENTOS DE ARQUITECTURA DE COMPUTADORES



# Contenido del capítulo

- DRAM: Introducción, arquitectura y organización
- Detección y corrección de errores
- Temporización y modos de acceso
- Módulos de memoria del PC
- Tecnología DRAM
- Rendimiento de la memoria RAM
- Bibliografía

# Contenido del capítulo

- DRAM: Introducción, arquitectura y organización
- Detección y corrección de errores
- Temporización y modos de acceso
- Módulos de memoria del PC
- Tecnología DRAM
- Rendimiento de la memoria RAM
- Bibliografía

# Introducción

---

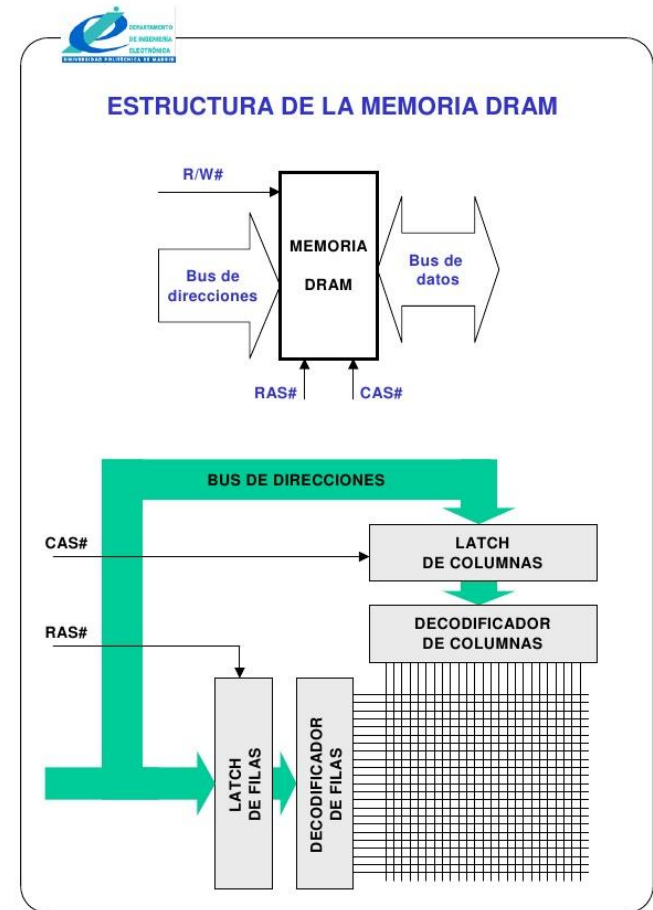
- Como sabemos, la memoria RAM es una memoria de acceso aleatorio (se puede acceder a cualquier posición en tiempo constante).
- Este tipo de memoria se usa tanto en los registros, como en la caché y en la memoria principal del ordenador.
- Distinguimos entre memoria RAM dinámica y estática, siendo la primera objeto de estudio de este capítulo.

# Características de la Tecnología DRAM

- Un chip de memoria es también un circuito integrado
- La celda básica de almacenamiento se consigue mediante la integración de los siguientes componentes:
  - Un único transistor que actúa como “interruptor” para la corriente.
  - Un condensador cargado (1) o sin corriente (0).
- Su sencillez permite una gran capacidad de integración y a un coste bajo: memoria principal de los ordenadores.
- **Necesitan ser refrescadas:**
  - El condensador implica fugas (con valor 1): cada cierto tiempo debe regenerarse la carga.
  - *Período de regeneración o refresco* es del orden de 10 a 20 ms.

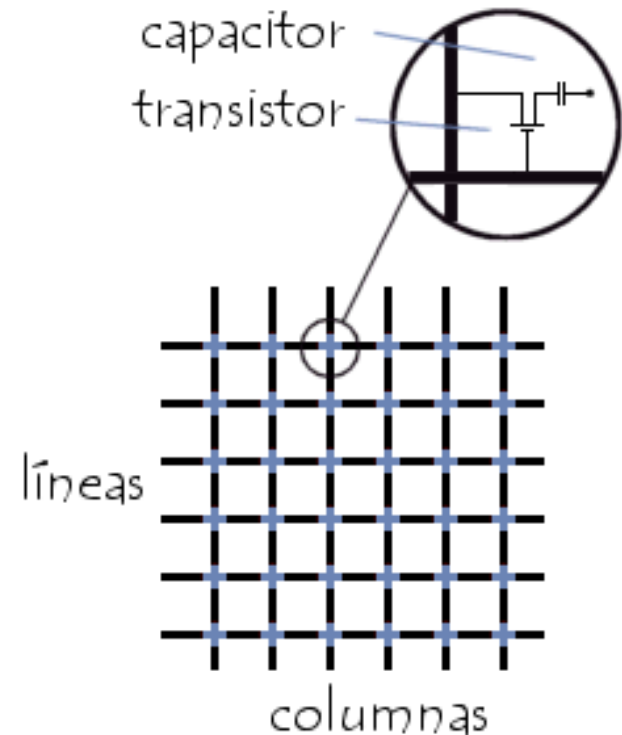
# Arquitectura y Organización DRAM

- Existen dos señales especiales:
  - RAS (Row Address Strobe): *wordlines*
  - CAS (Column Address Strobe): *bitlines*
- Pasos para el acceso:
  - Selección de fila (con la dirección y la señal RAS)
  - Selección de una columna (dirección y señal CAS)
  - R/W del dato deseado.
- Las direcciones pueden estar “*multiplexadas*”: se envía primero la fila y a continuación la col.

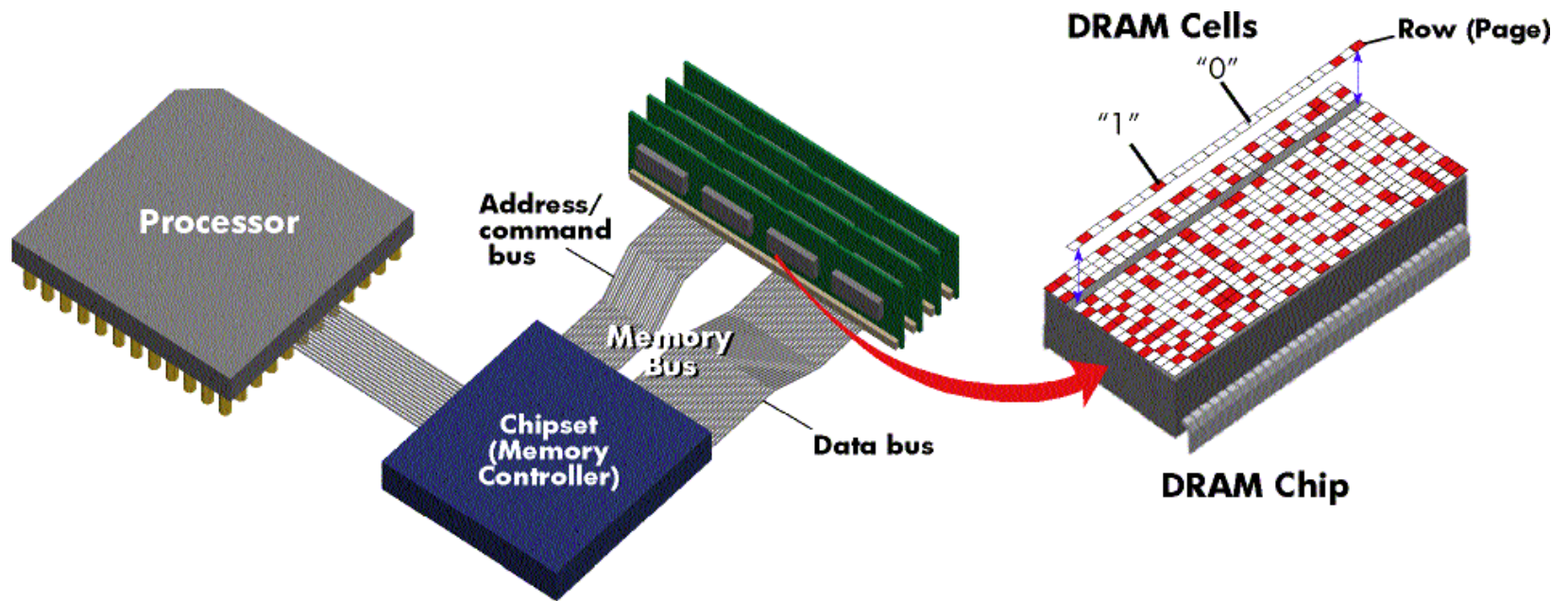


# Arquitectura y Organización DRAM (2)

- Existe un *controlador de memoria* dinámica que genera las señales con las temporizaciones adecuadas para el chip de memoria:
  - Reloj para sincronización y envío de datos.
  - Identificar cada fila y columna (RAS y CAS).
  - Llevar la cuenta de la secuencia de refresco.
  - Lectura y restauración de la señal de una celda (precarga).
  - Gestionar si una celda debe ser leída o escrita.
- Otras funciones del controlador de memoria incluyen:
  - Identificar el tipo, velocidad y
  - Detectar el tamaño de memoria
  - Chequeo de errores...



# Arquitectura y Organización DRAM (3)





# Contenido del capítulo

- DRAM: Introducción, arquitectura y organización
- Detección y corrección de errores
- Temporización y modos de acceso
- Módulos de memoria del PC
- Tecnología DRAM
- Rendimiento de la memoria RAM
- Bibliografía

# Detección y corrección de errores

- Todos los elementos de almacenamiento están sujetos a errores, especialmente la memoria de DRAM:
  - Error duro:
    - Es un defecto físico permanente, tal que la celda o celdas afectadas no pueden almacenar datos de manera segura.
    - Las celdas quedan ancladas a 0 o a 1, o conmutando erróneamente entre 0 y 1.
    - Los errores duros pueden estar causados por funcionamiento en condiciones adversas, defectos de fabricación, y desgaste.
  - Error blando:
    - Es un evento aleatorio no destructivo que altera el contenido de una o más celdas de almacenamiento, sin dañar la memoria.
    - Estos pueden deberse a problemas de la fuente de alimentación o a pequeñas emisiones radiactivas presentes en todos los materiales.

# Detección y corrección de errores (2)

- Memoria sin paridad: No realiza comprobación de errores.
- Memoria con paridad:
  - Necesita un bit adicional por cada byte de información.
  - Almacena un 1 si hay un número impar de unos y un 0 en caso contrario.
  - Evidentemente se le pueden escapar muchos errores y es imposible corregirlos.
  - En caso de detección de error normalmente aparece una interrupción que bloquea el ordenador.
  - Este tipo de comprobación no retrasa para nada el acceso a memoria.

Tipo de Módulo	Anchura de datos sin paridad	Anchura de datos con paridad
SIMM 30-contactos	8 bits	9 bits
SIMM 72-contactos	32 bits	36 bits
DIMM 168-contactos	64 bits	72 bits

# Contenido del capítulo

- DRAM: Introducción, arquitectura y organización
- Detección y corrección de errores
- Rendimiento de la memoria RAM
- Temporización y modos de acceso
- Módulos de memoria del PC
- Tecnología DRAM
- Bibliografía

# Rendimiento de la memoria RAM

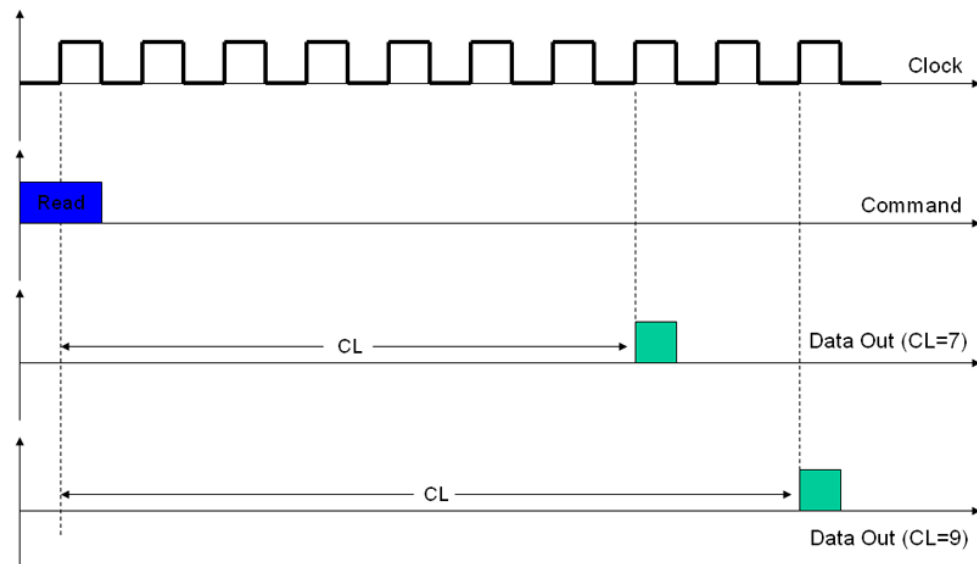
- Recordemos dos parámetros esenciales de la memoria:
  - Ancho del bus: n° de bits que pueden enviarse simultáneamente a la CPU.
  - Velocidad del bus: número de veces que un grupo de bits pueden ser enviados por segundo.
- Adicionalmente, la eficiencia de memoria se mide según dos parámetros básicos:
  - **Latencia:** tiempo de acceso.
  - **Frecuencia de reloj:** número de ciclos por segundo.

# Rendimiento de la Memoria RAM

- Actualmente, la latencia se mide en función de 4 tiempos diferentes:
  - CAS: indica el tiempo que tarda la memoria en colocarse sobre una columna o celda.
  - Delay RAS to CAS: Latencia entre la activación de la señal fila y de columna
  - RAS: indica el tiempo que tarda la memoria en colocarse sobre una fila.
  - PRECHARGE: indica el tiempo que tarda la memoria en desactivar un banco o bloque de datos.
- Los 3 primeros valores suelen ser idénticos:
  - Las latencias típicas DDR2 fueron 5-5-5-15
  - Para dispositivos DDR3 son 7-7-7-20 para DDR3-1066 y 7-7-7-24 para DDR3-1333.

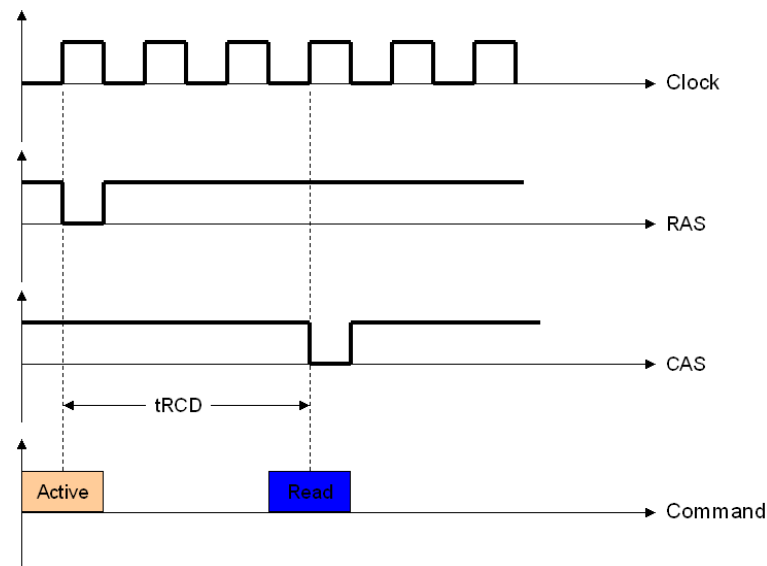
# Rendimiento de la Memoria RAM

- Latencia CAS (CL): nos indica cuántos ciclos tarda la memoria en devolver el dato requerido.
- A menor latencia CAS, más eficiente será el acceso.
- Memorias DDR2 proporcionan dos datos por orden CAS.
- En memorias Burst (ver más adelante) solo se necesita un ciclo para acceder al siguiente dato.



# Rendimiento de la Memoria RAM

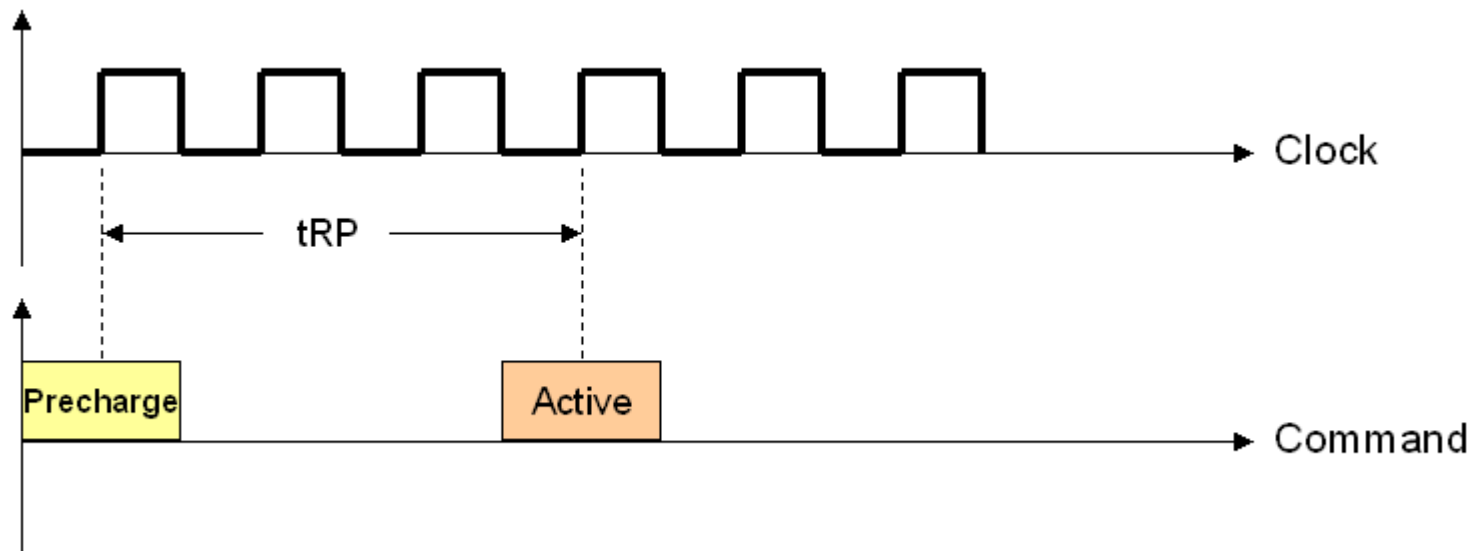
- Delay RAS to CAS (tRCD):
  - Recordemos la memoria RAM no es más que una matriz de datos
  - Para acceder a una celda, primero activamos la señal de fila y a continuación la de columna.
  - A menor tiempo entre estas dos señales, más rápido será el acceso.
  - También se indica como el número de ciclos entre la orden ACTIVE y la de Lectura/Escritura





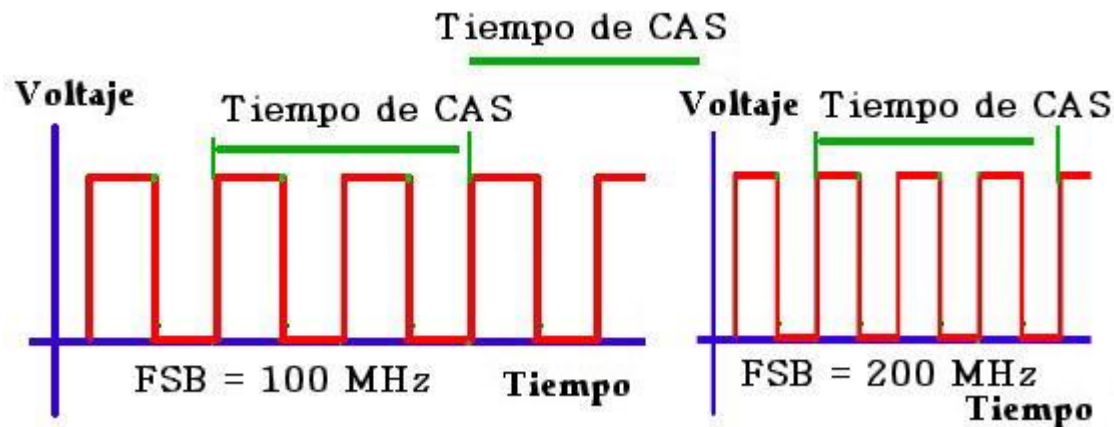
# Rendimiento de la Memoria RAM

- Precarga RAS ( $t_{RP}$ ): “Cierra” la fila de memoria usada para poder activar una nueva fila.
- Es la latencia trascurrida desde la orden de Precarga hasta que se puede reactivar la orden de R/W (Active)
- El tiempo que pasa entre la emisión de la orden “Precarga” y la obtención real de los datos es igual a  $t_{RP} + t_{RCD} + CL$



# Rendimiento memoria RAM: Overclocking

- Para mejorar el rendimiento de los módulos de MP existen dos formas (BIOS):
  - Reducir la latencia (tiempo de acceso).
  - Incrementar la frecuencia de trabajo (MHz)  $\sim$  FSB.
- Ambos son inversamente proporcionales:
  - Al incrementar el n° de ciclos por segundo  $\rightarrow$  la latencia aumenta.
  - Necesito más ciclos para agotar el tiempo de espera.



## Rend. de memoria RAM: Overclocking (2)

- Las memorias actuales son síncronas con el reloj del sistema.
- La frecuencia de reloj del FSB/HTT/QPI restringe la velocidad máxima que alcanzará nuestra memoria instalada.
- Si queremos explotar al máximo nuestra memoria RAM, debemos instalar la memoria de mayor velocidad hasta llegar al límite impuesto por el FSB del microprocesador.
- La sincronía con el FSB/HTT se realiza a través de divisores: intentar sincronizar microprocesador y memoria RAM lo máximo posible (1:1).
- En definitiva, si colocamos una memoria a mayor frecuencia que la permitida por la placa base, estamos desperdiciando rendimiento.

## Rend. de memoria RAM: Overclocking (3)

- Modificar estos valores suele implicar subir el voltaje para suministrarle más energía:
  - A veces es exigencia del fabricante.
  - Uso de disipadores de memoria RAM.
- Resulta muy dependiente del controlador de memoria:
  - Integrado en el micro (más rápido).
  - En el chipset de la placa (más lento).
- Deshabilitar la memoria virtual también hace mejorar el rendimiento si disponemos de capacidad suficiente.
- Incluso en casos óptimos, la CPU tarda más en acceder a los datos de memoria que en procesarlos: necesidad de memoria caché.

# Contenido del capítulo

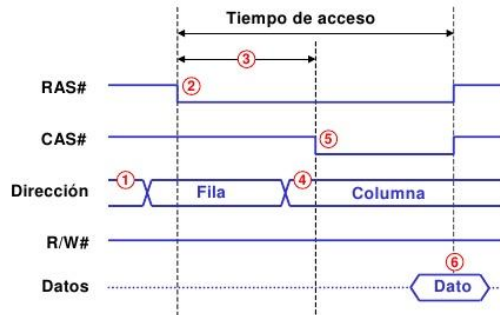
- DRAM: Introducción, arquitectura y organización
- Detección y corrección de errores
- Rendimiento de la memoria RAM
- Temporización y modos de acceso
- Módulos de memoria del PC
- Tecnología DRAM
- Bibliografía

# Temporización y modos de acceso



## SECUENCIA DE ACCESO

1. Dirección de fila
2. Activación de RAS#
3. Retardo RAS/CAS
4. Dirección de columna
5. Activación de CAS#
6. Lectura o escritura (R/W#)



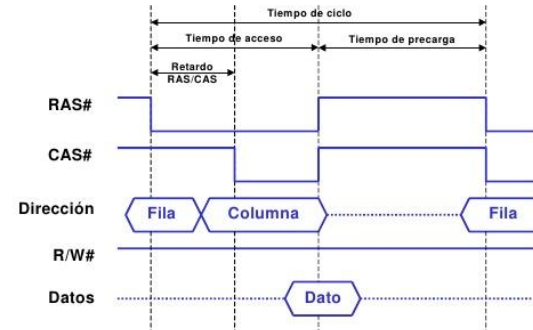
## SECUENCIA DE ACCESO

### Tiempo de pre-carga

Lectura destructiva

Necesita regeneración

El hardware interno regenera el dato

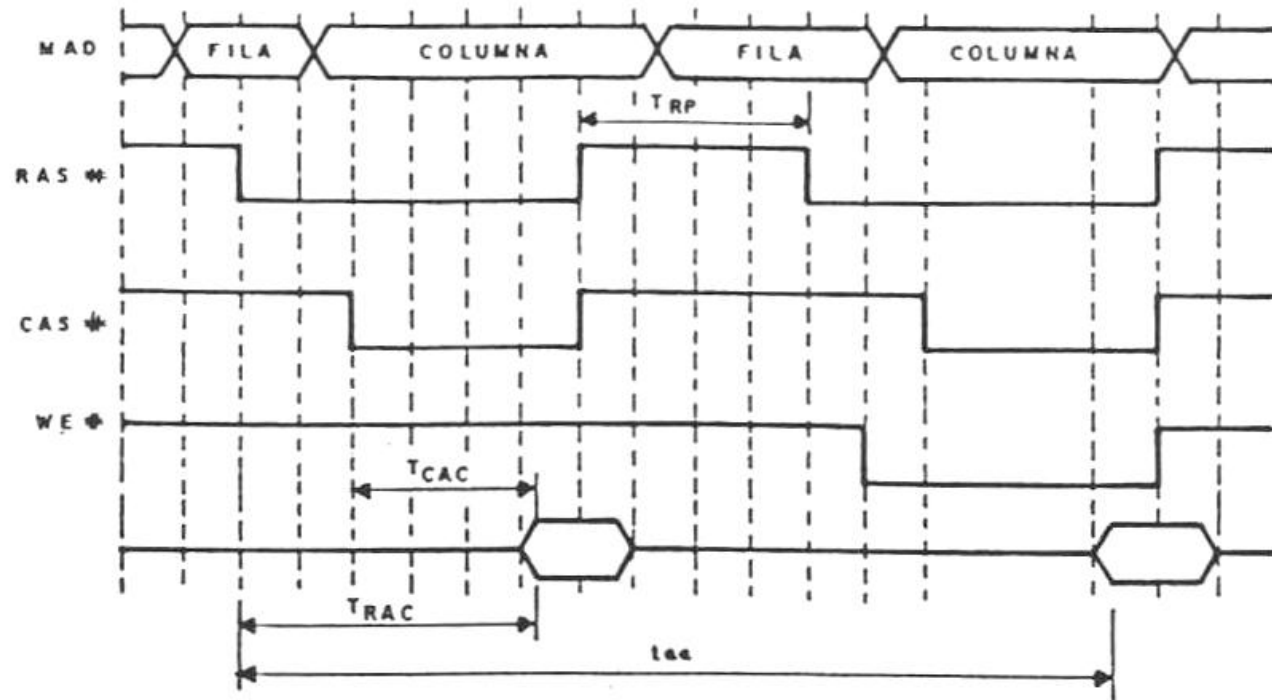


Tiempo de pre-carga ~ tpo. acceso

Tiempo real de acceso = tpo. de ciclo

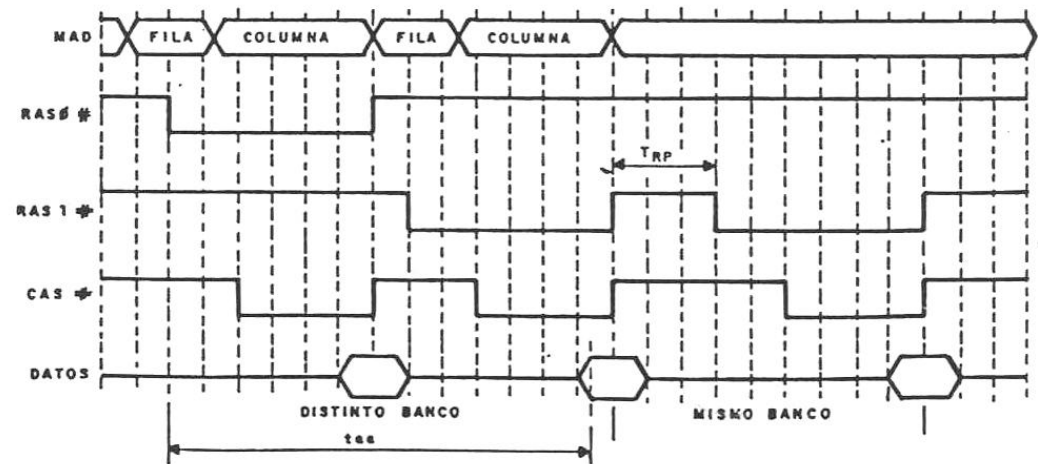
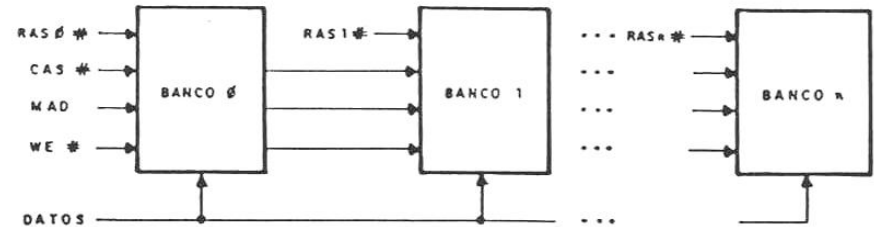
# Modos de acceso: estándar

- Tiempo de dos acceso simultáneos:
  - $T_{aa} = T_{RAS} + T_{RP} + T_{RAS}$



# Modos de acceso: entrelazado

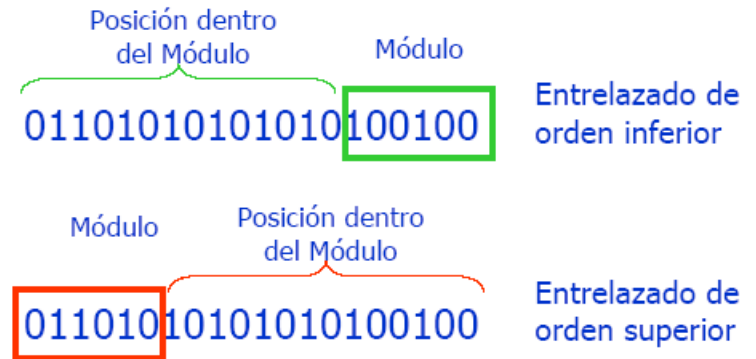
- Por bancos, módulos, chips o bloques de memoria
- Se almacenan posiciones consecutivas en distintos bloques
- Se solapan los tiempos de precarga
- Tiempo de acceso:  $t_{RAC} + (M \cdot t_{Sel})$ , donde  $t_{Sel}$  es el tiempo de conmutación entre módulos



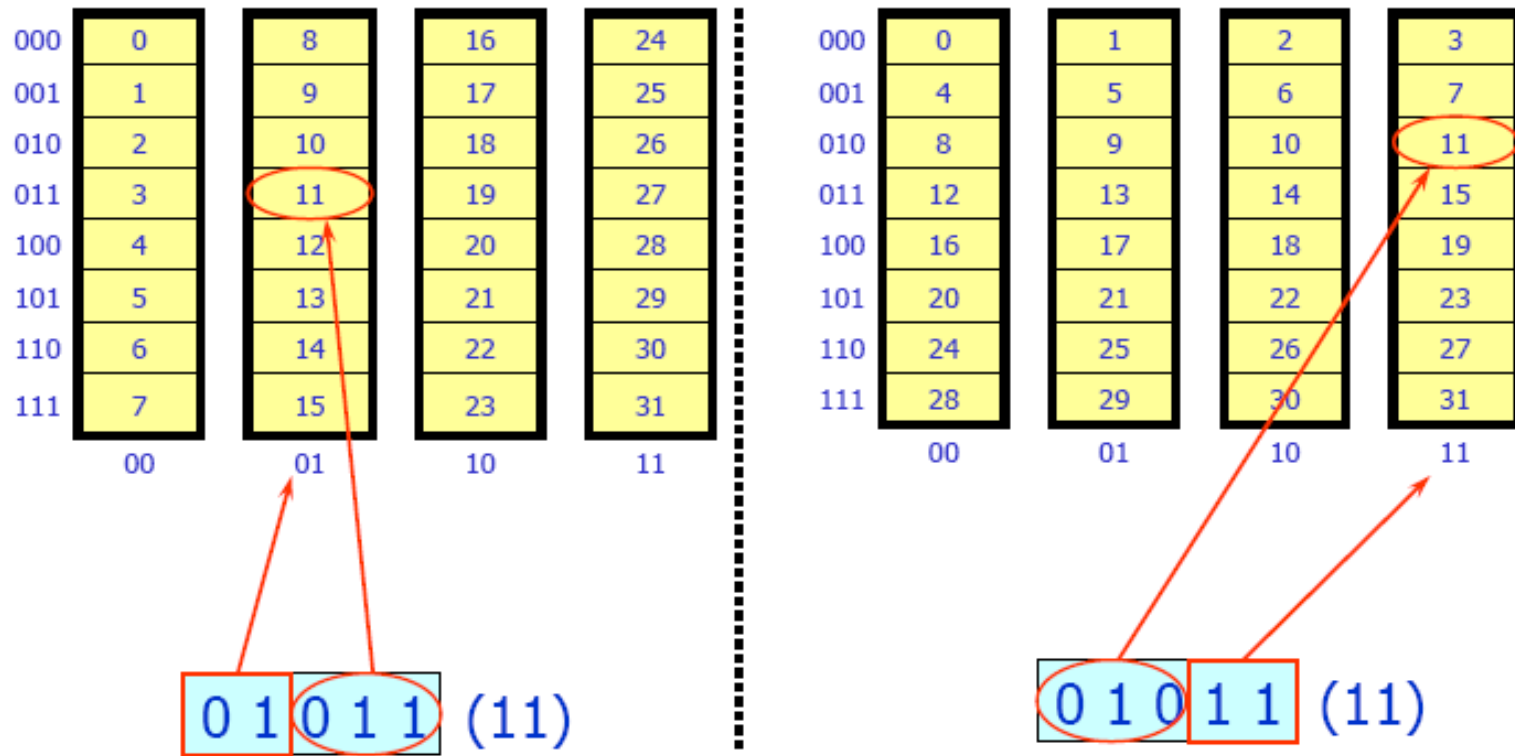


# Entrelazado

- Dirección de memoria: un conjunto de bits indican el módulo y el resto indican la posición dentro de ese módulo.
- Ej: Entrelazado de una memoria de 220 palabras en  $2^6=64$  módulos



# Entrelazado de Memoria



$2^5 = 32$  direcciones de memoria

$2^2 = 4$  módulos de  $2^3 = 8$  posiciones

Entrelazado Superior

Entrelazado Inferior

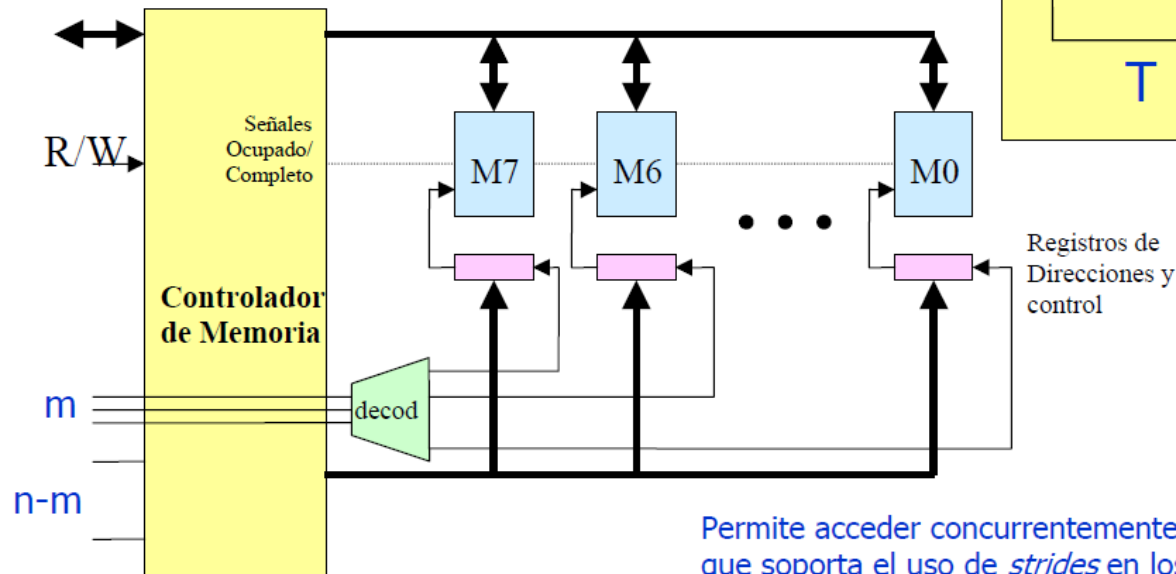
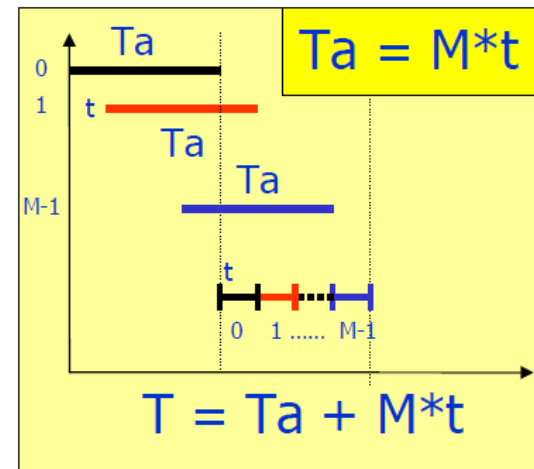
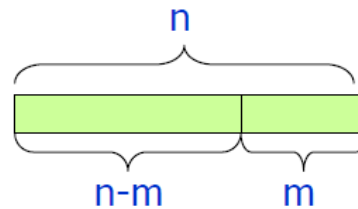
# Ej: Entrelazado de Memoria Inferior

## Con Entrelazado Inferior

$N=2^n$  direcciones

$M=2^m$  módulos

$2^{(n-m)}$  direcciones/módulo

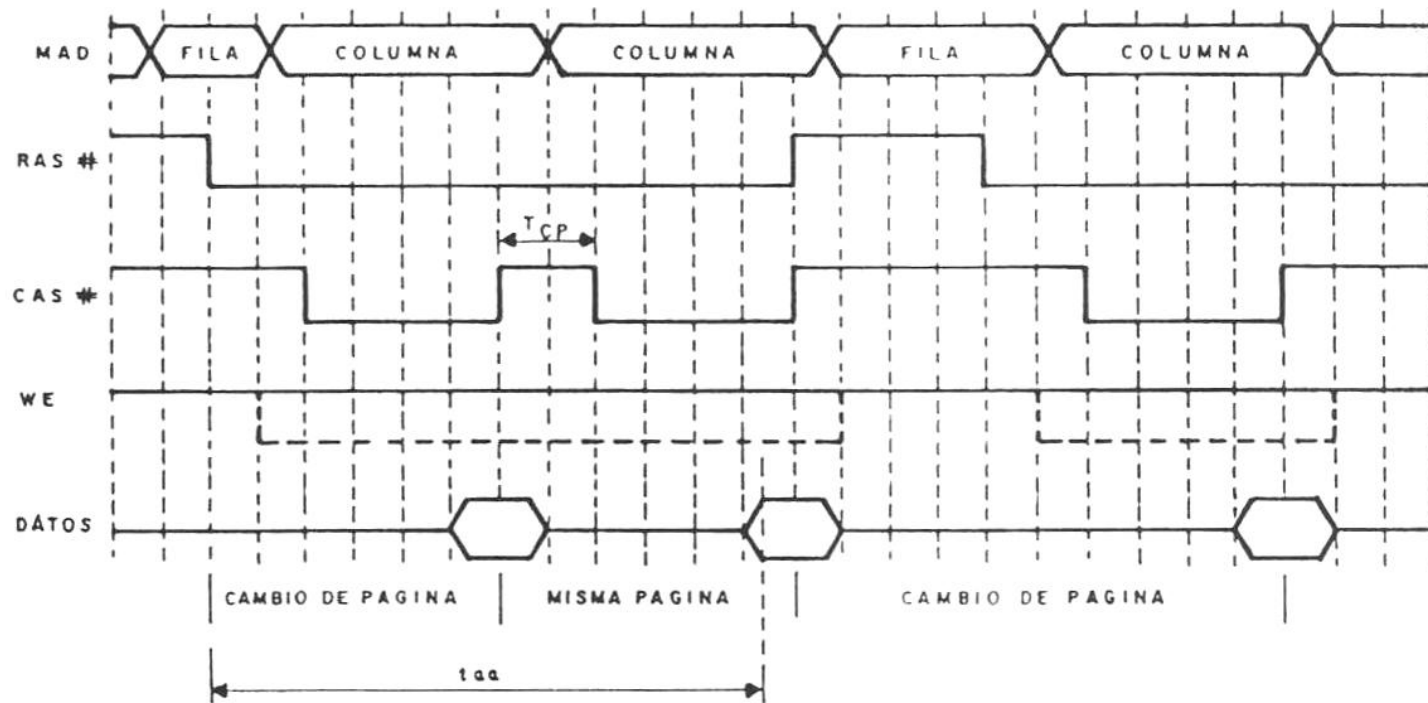


Permite acceder concurrentemente a  $M$  direcciones, con lo que soporta el uso de *strides* en los accesos a memoria

# Modos de acceso: modo página

- Acceden rápidamente a direcciones en la misma fila:
  - No intervienen los retardos de RAS
  - Limitan su acceso a la señal CAS
  - El acceso al primer dato es estándar
- Los chips se organizan por páginas
  - Cada página se corresponde con un número de fila
  - El tamaño de cada página es igual al número de columnas
- Mientras no se cambie de página no necesitamos volver a presentar la dirección de fila (RAS no cambia).

# Modos de acceso: Modo Página (2)



# Modos de acceso: modo “Burst”

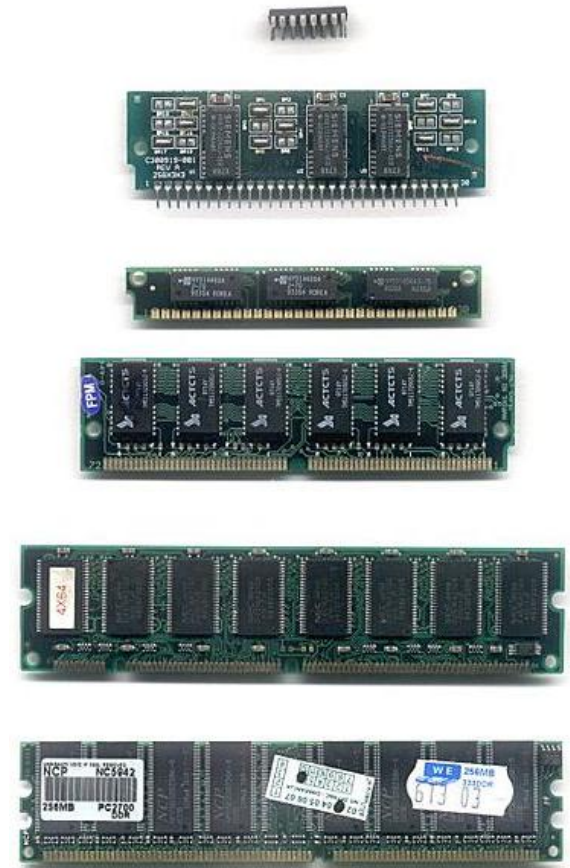
- Un acceso usual a memoria leerá una palabra completa (32 ó 64 bits).
- Al almacenar datos consecutivamente, se permite paralelizar el acceso: técnica del **entrelazado superior**.
- Es por tanto una “fusión” entre el modo página y el entrelazado.
- La señal #CAS tiene un incremento automático

# Contenido del capítulo

- DRAM: Introducción, arquitectura y organización
- Detección y corrección de errores
- Rendimiento de la memoria RAM
- Temporización y modos de acceso
- Módulos de memoria del PC
- Tecnología DRAM
- Bibliografía

# Módulos de memoria del PC

- Los chips de memoria en DIP (Dual Inline Package), a su vez empaquetados en los módulos de memoria.
- Podemos encontrar módulos de memoria de tres apariencias físicas:
  - Módulos SIMM: Formato usado en computadores antiguos. Tenían un bus de datos de 16 o 32 bits
  - Módulos DIMM: Usado en computadores de sobremesa. Se caracterizan por tener un bus de datos de 64 bits.
  - Módulos SO-DIMM: Usado en ordenadores portátiles. Formato miniaturizado de DIMM.





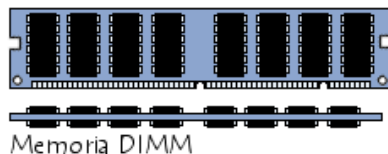
# Módulos de memoria del PC: SIMM

- Los módulos en formato SIMM (Módulo de Memoria en Línea Simple) se tratan de placas de circuito impresas, con uno de sus lados equipado con chips de memoria.
- Existen dos tipos de módulos SIMM:
  - Módulos SIMM de 30 conectores: memorias de 8 bits para PCs de primera generación (286, 386).
  - Módulos SIMM de 72 conectores: capaces de almacenar 32 bits de información en forma simultánea.
    - Estas memorias se encuentran en los PC que van desde el 386DX hasta los primeros Pentiums.
    - Pentium: el procesador funciona con un bus de información de 64 bits: necesitan estar equipados con dos módulos SIMM.



# Módulos de memoria del PC: DIMM

- Módulos en formato DIMM (Módulo de Memoria en Línea Doble), son memorias de 64 bits.
- No necesitan *emparejamiento*.
- Poseen chips de memoria en ambos lados de la placa de circuito impresa.
- Estos módulos poseen una segunda muesca que evita confusiones con SIMM (y son más grandes)
- Existen módulos más pequeños, conocidos como SO DIMM, diseñados para ordenadores portátiles.



# Contenido del capítulo

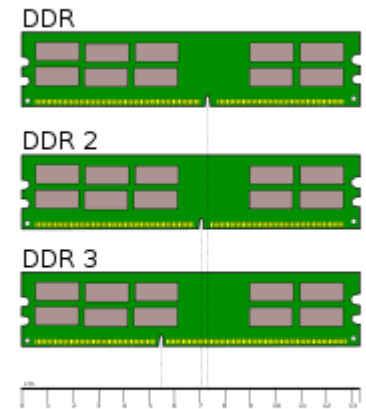
- DRAM: Introducción, arquitectura y organización
- Detección y corrección de errores
- Rendimiento de la memoria RAM
- Temporización y modos de acceso
- Módulos de memoria del PC
- Tecnología DRAM
- Bibliografía

# Tecnologías DRAM: Acceso más eficiente

- **Fast Page Mode (FPM):** Funciona en modo página y se solapan algunos tiempos de trabajo ciertos buffers internos.
- **Extended Data Out (EDO):** Físicamente igual a la FPM. Pequeñas optimizaciones en el manejo de las señales de control y a la inclusión de registros a la salida de los bloques.
- **Synchronous DRAM (SDRAM):** las señales RAS y CAS se sincronizan con la velocidad de comunicaciones del bus del sistema. Permite un diseño simplificado.
- **Double Data Rate Synchronous DRAM (DDR-SDRAM):** capaz de leer datos tanto en el flanco de subida como en el de bajada del ciclo del reloj del sistema, duplicando la velocidad de transferencia de datos.
  - DDR2 → Multiplica la velocidad de acceso x 2 + dual channel.
  - DDR3 → Multiplica la velocidad de acceso x 8 + triple channel.
- **Direct Rambus DRAM (DR-DRAM):** teóricamente más rápida que las anteriores, pero con derechos registrados (Rambus + Intel). *Obsoleto*.
- **SyncLink DRAM (SLDRAM):** extensión de SDRAM y competencia directa de Rambus. *Obsoleto*.

# Tecnologías DRAM (2): DDR SDRAM

- Físicamente incompatibles entre ellos, ya que llevan la muesca para introducirlos en la ranura DIMM en diferente posición:
  - DDR: 168 pines y voltaje hasta 2.5V
  - DDR2: 240 pines y voltaje hasta 1.8V.
  - DDR3: 240 pines y voltaje hasta 1.5V.
  - SO-DIMM: en sus tres variantes cuentan con 200 pines.
- La velocidad de reloj de E/S y la reducción de voltaje son la principal mejora en las distintas versiones:
  - RAM DDR: 400MHz - 200 Mhz E/S – 3200 MB/s
  - DDR2: 1200Mhz - 800 Mhz E/S – 9600 MB/s
  - DDR3: 2200 Mhz – 1200 Mhz E/S – 18000 MB/s
- <http://www.hardwaresecrets.com/article/167>



# Tecnologías DRAM (3): Modos DDR

- **Single Memory Channel:**
  - Los módulos de memoria intercambian información con el bus por un sólo canal.
  - Hay un único banco de slots (DDR).
- **Dual Memory Channel:**
  - Se reparten los módulos de memoria entre los dos bancos de slots diferenciados en la placa base.
  - Pueden intercambiar datos con el bus a través de dos canales simultáneos, uno para cada banco (DDR – DDR3).
- **Triple Memory Channel:**
  - Actual para Intel Core i7.
  - Los módulos de memoria se reparten de 3 en 3 (DDR3).
  - 6 slots de memoria para una sola CPU.
  - Mayor ancho de banda.



# Tecnologías DRAM (4): DDR2

- Son capaces de trabajar con 4 bits por ciclo:
  - 2 bits de ida y 2 bits de vuelta en un mismo ciclo (flanco de reloj): uso de un buffer para el envío/escritura de datos.
  - Se mejora sustancialmente el ancho de banda potencial bajo la misma frecuencia de una DDR SDRAM tradicional.
- Mayores latencias que las DDR convencionales



## Tecnologías DRAM (5): DDR3 y DDR4

- Permite transferencias de datos ocho veces más rápido.
- No hay reducción en latencia: proporcionalmente más alta.
- DDR3 permite usar integrados de 512 megabits a 8 gigabytes: módulos de hasta 16 GB.
- Nueva tecnología DDR4:
  - En fase de producción (2013 → 2014)
  - Mayor número de pines (288)
  - Se prevé un voltaje menor (1.2V) frente a DDR3 (1.5V).
  - Frecuencia de reloj desde 2133 MHz hasta 3.2GHz





# Memoria de Video: VRAM

---

- Video Random Access Memory (VRAM) es un tipo de memoria RAM que utiliza el controlador gráfico para poder manejar toda la información visual que le manda la CPU del sistema.
- La principal característica de esta clase de memoria es que es accesible de forma simultánea por dos dispositivos (CPU y Tarjeta Gráfica).
- Es posible que la CPU grabe información en ella, mientras se leen los datos que serán visualizados en el monitor en cada momento.
- Existen distintos tipos de memoria VRAM, siendo la más extendida **GDDR-SDRAM**
- Actualmente se trabaja con GDDR5 (20 GB/s con 32 bits): incluida en PS4.

# Contenido del capítulo

- DRAM: Introducción, arquitectura y organización
- Detección y corrección de errores
- Rendimiento de la memoria RAM
- Temporización y modos de acceso
- Módulos de memoria del PC
- Tecnología DRAM
- Bibliografía

# Bibliografía

---

- Patterson y Hennessy: Estructura y Diseño de Computadores. Capítulo 5.
- Murdocca y Heuring: Principios de Arquitectura de Computadoras: Capítulo 7.
- HP: Memory technology evolution: an overview of system memory technologies. Technology brief, 7th edition