

Arquitectura de Computadoras para Ingeniería

(Cód. 7526)
1° Cuatrimestre 2016

Dra. Dana K. Urribarri
DCIC - UNS

Memoria

Memoria

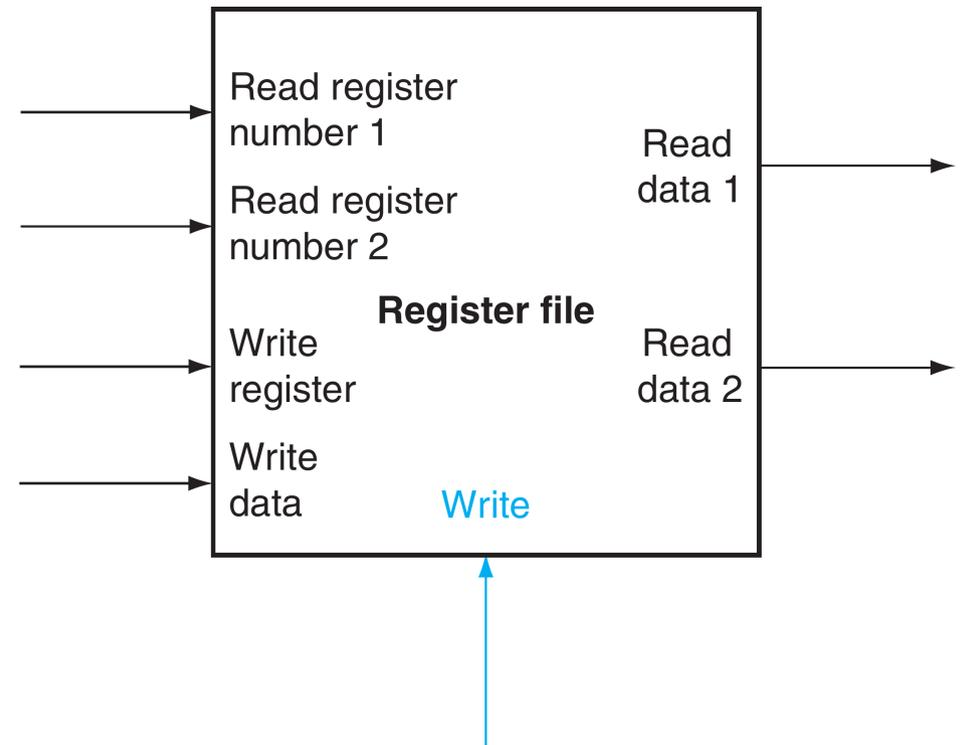
- Diferentes tecnologías de implementación de memorias.
- Difieren principalmente en
 - Costo
 - Capacidad
 - Velocidad
- El objetivo es lograr que la memoria tenga los beneficios de las tecnologías:
 - Máxima velocidad
 - Gran capacidad
 - Bajo costo

Tecnologías

- Tipos principales de memoria
 - Registros
 - SRAM (static random access memory)
 - DRAM (dynamic random access memory)
 - Flash
 - Discos magnéticos

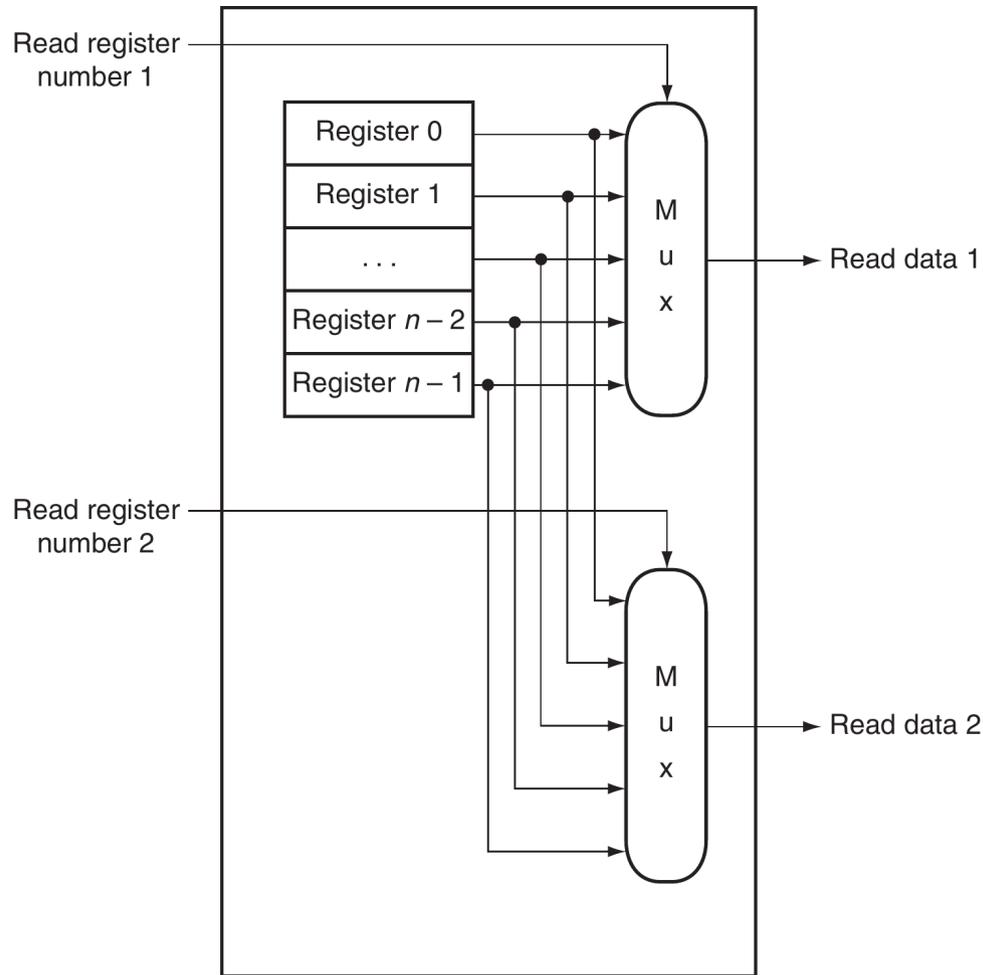
Banco de registros

- Conjunto de registros que pueden leerse y escribirse.
 - Es rápido
 - Costoso en la relación \$/bit
 - Alto consumo energético

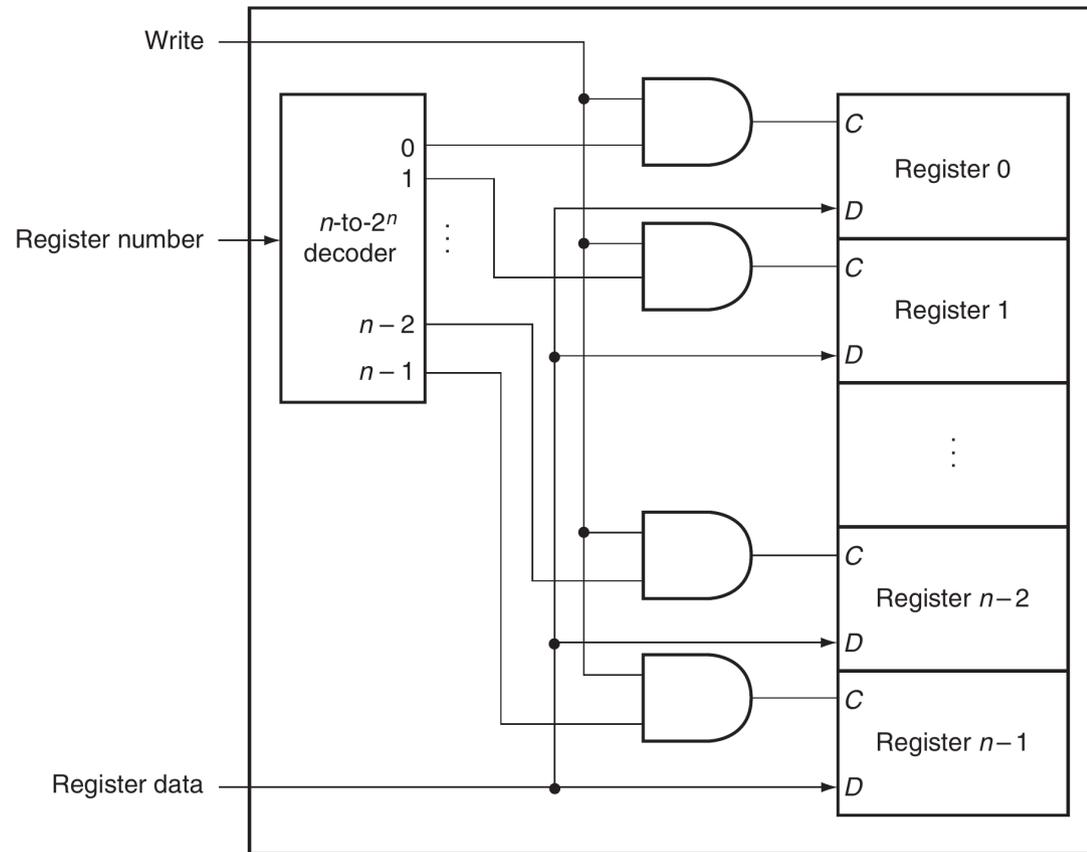


- Los registros se identifican con números.

Banco de registros



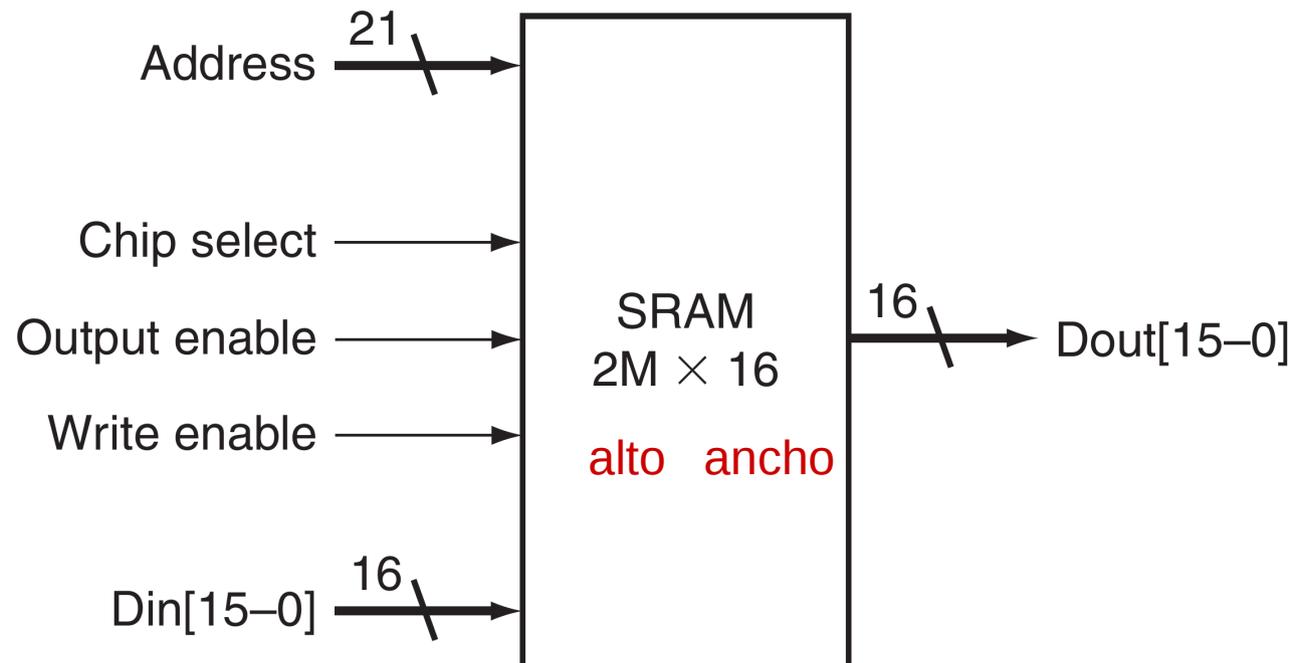
Dos puertos de lectura.



Se necesita lógica adicional para leer el mismo que se está escribiendo.

SRAM

- Circuitos integrados con un arreglo de FF.
- En general, único puerto de lectura/escritura.
- Tiempo fijo para acceder a cualquier elemento.

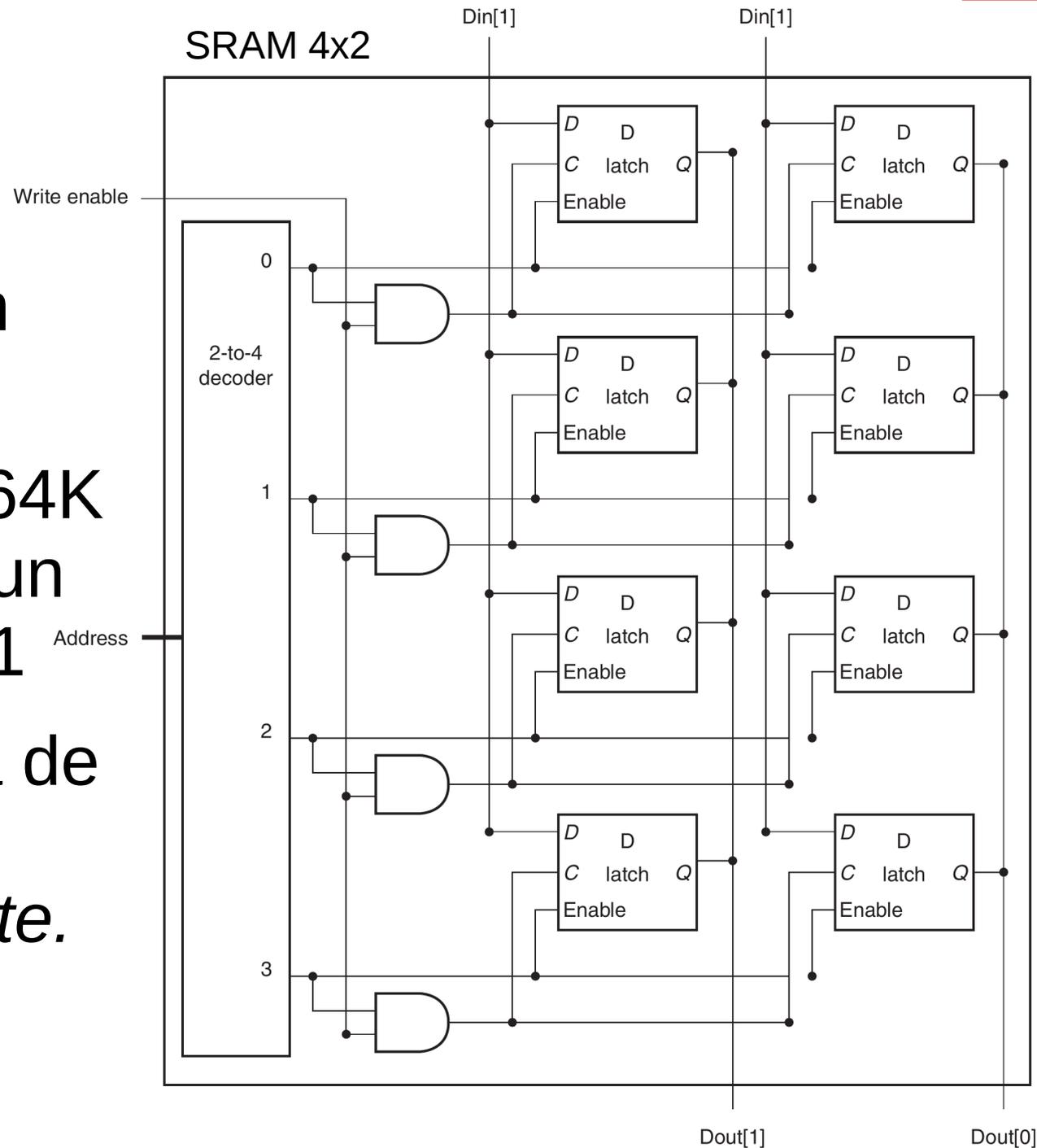


SRAM

Es impracticable implementarlo con MUX.

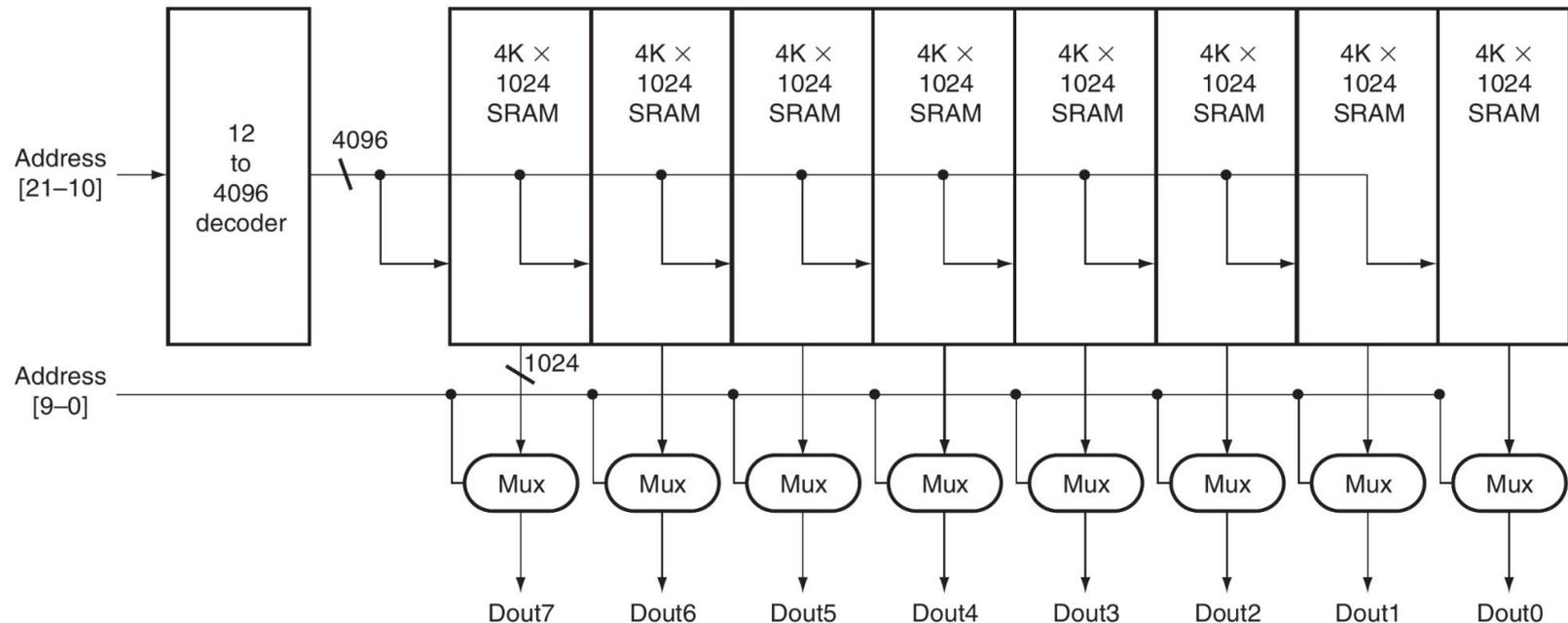
- Una SRAM de 64K x 1 necesitaría un MUX de 64K a 1

Se tiene una línea de salida común y se usan *buffers tristate*.



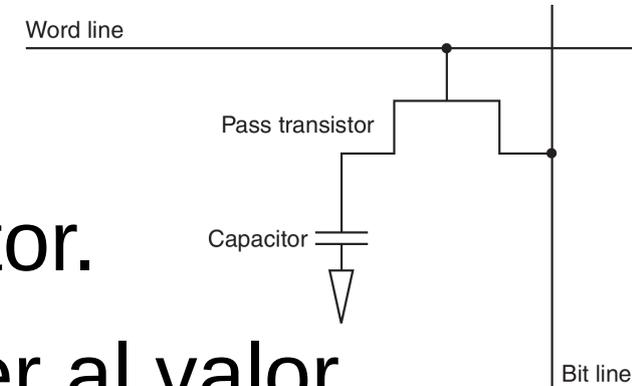
SRAM

- Una SRAM de 4M x 8 necesitaría un decoder de 22 x 4M y 4M líneas para habilitar los 8 FF de la palabra.
- Las memorias grandes se organizan en un arreglo y la decodificación se hace en 2 pasos.



DRAM

- El valor se almacena en un capacitor.
- Se utiliza un transistor para acceder al valor cargado (para lectura o escritura).
 - En comparación, las SRAM necesitan 6 transistores por bit.
 - Eso hace a las DRAM más densas y baratas por bit.
- Los capacitores se descargan con el tiempo.
Pueden mantener la carga por algunos mseg.
 - Las DRAM necesitan *refrescarse* periódicamente (leer el contenido y volver a escribirlo).
 - *Controlador de memoria*: incluido en el chip del procesador o separado.



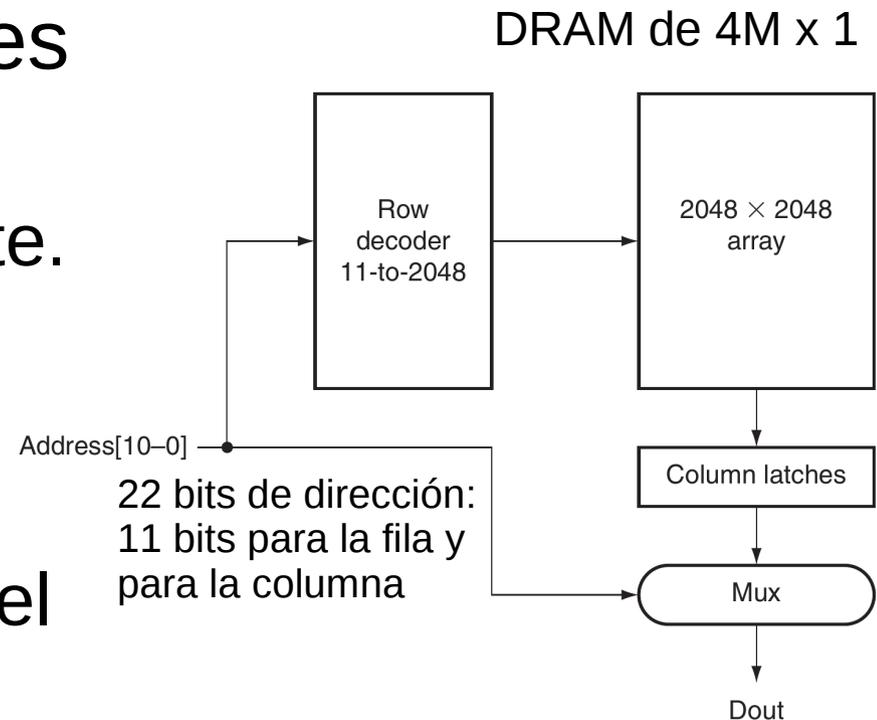
DRAM

- Decodificación de dos niveles

- 1) *Row access*: Elige una fila y activa la línea correspondiente. El contenido de todas las columnas en la línea activa se almacenan en *latches*.

- 2) *Column access*: Selecciona el dato de la columna correcta.

- Para ahorrar pines se usan los de dirección tanto para filas como para columnas. Se distingue con las señales **RAS** (*Row Access Strobe*) y **CAS** (*Column Access Strobe*).



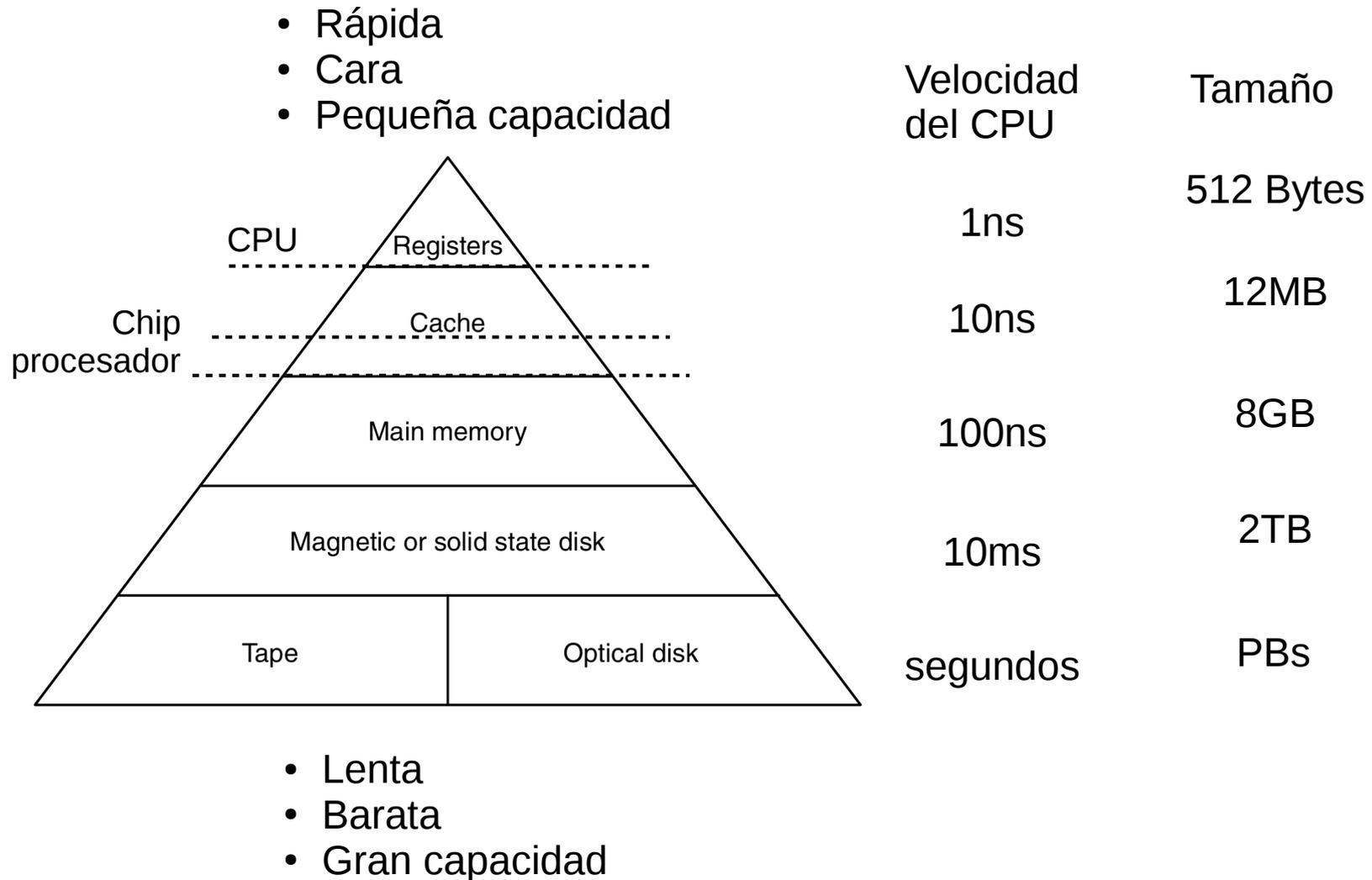
SRAM vs DRAM

- Ambas son volátiles.
- Las celdas de la DRAM son más simples y pequeñas.
 - Más densa y menos costosa
- DRAM requiere un circuito de refresco.
- Para grandes memorias, el costo (fijo) adicional del circuito de refresco se compensa con el menor costo de las celdas.
- SRAM son más rápidas que las DRAM.
- SRAM se usa para memoria cache (integrada y externa al chip)
- DRAM se usa para memoria principal.

SSRAM – SDRAM – DDRRAM

- Synchronous SRAM y Synchronous DRAM
 - Mejora la velocidad.
 - Elimina el tiempo de sincronización entre memoria y procesador.
 - Permite transferir una ráfaga de datos a una serie de direcciones consecutivas.
- Double Data Rate RAM
 - SDRAM que transfiere datos tanto en el flanco ascendente como en el descendente del reloj.

Jerarquía de memoria



Jerarquía de memoria

Level	1	2	3	4
Name	Registers	Cache	Main memory	Disk storage
Typical size	<1 KB	32 KB–8 MB	<512 GB	>1 TB
Implementation technology	Custom memory with multiple ports, CMOS	On-chip CMOS SRAM	CMOS DRAM	Magnetic disk
Access time (ns) (2006)	0.15–0.30	0.5–15	30–200	5,000,000
Bandwidth (MB/sec)	100,000–1,000,000	10,000–40,000	5000–20,000	50–500
Managed by	Compiler	Hardware	Operating system	Operating system/ operator
Backed by	Cache	Main memory	Disk	Other disks and DVD

Principio de localidad

La información no se accede de manera aleatoria

- **Localidad temporal:**
 - Los datos y códigos usados en el pasado es probable que se usen en el futuro cercano (datos en la pila, bucles)
- **Localidad espacial:**
 - En el futuro cercano es probable usar datos y código cercanos (en término de direcciones de memoria) a los datos y código actuales (recorrer arreglos, código secuencial).

Jerarquía de Memoria

- Cada nivel de la jerarquía tiene menor capacidad que el nivel inferior. → **No hay lugar para todo**
- Cuando se genera una referencia a memoria puede ocurrir que:
 - Se encuentre entre el contenido del nivel accedido
→ *hit* (éxito)
 - No se encuentre en el contenido del nivel accedido
→ *miss* (fallo o faltante)
- En el caso de un fallo en el nivel L_i , la referencia se busca en el nivel inmediatamente inferior L_{i+1} .

Bibliografía



- Capítulo 5 y Apéndice B. David A. Patterson & John L. Hennessy. *Computer Organization and Design. The Hardware/Software Interface.* Elsevier Inc. 2014, 5ta Ed.
- Capítulo 2. *Multiprocessor Architecture. From simple pipelines to chip multiprocessor.* Jean-Loup Baer. Cambridge University Press. 2010.

Suplementaria

- Apéndice B. John L. Hennessy & David A. Patterson. *Computer Architecture. A Quantitative Approach.* Elsevier Inc. 2012, 5ta Ed.