

Organización de Computadoras 2021

TURNO RECURSANTES
CLASE 8

Resumen clase 8

- Subsistema de Memoria
- Organización jerárquica de la memoria
- Clasificación de las memorias
- Organización física de la memoria
- Evolución de las RAM
- SDRAM, DDR, DDR2, DDR3

Memoria

3

- Memoria es el conjunto de elementos y dispositivos que almacenan información en un sistema de cómputo.
- Cuando se habla del Subsistema Memoria de un sistema de cómputo, hay 3 conceptos fundamentales para analizar:
 - Velocidad
 - Tamaño
 - Costo

Memoria

Velocidad de la Memoria

- La velocidad de la memoria se mide por la cantidad de bytes (o bits o words) que puede transferir por unidad de tiempo (típicamente por segundo).
- Por lo tanto depende de 2 factores principales:
 - El tiempo de acceso de la memoria
 - La cantidad de bytes que puede transferir simultáneamente.
- La tecnología permite que la velocidad de la memoria crezca ininterrumpidamente a lo largo del tiempo.
- Como referencia, se puede decir que la velocidad de la memoria aumenta un 10% cada año.

Memoria

5

Velocidad de la Memoria

- El incremento en la velocidad de la memoria es bastante inferior al aumento de velocidad del procesador.
- La velocidad del procesador depende, entre otros factores, de la tecnología y la arquitectura.
- Como referencia, aproximadamente cada 18 meses un procesador típico duplica la cantidad de instrucciones ejecutadas por segundo.

Memoria

6

Velocidad y tiempo de acceso de la Memoria

- Es decir, que el incremento en la velocidad de la memoria es bastante inferior al aumento de velocidad del procesador.
- El tiempo de acceso de la Memoria depende básicamente de 2 factores:
 - La tecnología (costo)
 - Capacidad (tamaño) de la memoria (costo)

Memoria

7

Tiempo de acceso de la Memoria

Las memorias semiconductoras tienen 2 parámetros referidos al tiempo de acceso:

- Tiempo de acceso: tiempo (máximo) que transcurre desde que se inicia la operación de lectura/escritura hasta obtener/almacenar el dato.
- Tiempo de ciclo: tiempo mínimo que tiene que haber entre dos operaciones sucesivas sobre la memoria

siendo: $t_{\text{ciclo}} > t_{\text{acceso}}$

Memoria

8

Tiempo de acceso de la Memoria

Las memorias magnéticas u ópticas tienen 3 parámetros referidos al tiempo de acceso:

$$T_{acc} = T(\text{posicionar el cabezal}) + T(\text{latencia}) + T(\text{transferencia})$$

donde:

- $T(\text{posicionar el cabezal}) = f(\text{cantidad de pistas a recorrer})$
- $T(\text{latencia}) = f(\text{distancia entre sector del disco a acceder y la posición cabeza lectora/grabadora})$
- $T(\text{transferencia}) = f(\text{velocidad de rotación})$

Memoria

9

Capacidad de la Memoria

- La capacidad de la memoria está asociada al tipo de dispositivo.
- Depende de la tecnología empleada por el dispositivo que almacena información.
- Como referencia, se estima que la capacidad de la memoria se cuadruplica cada 36 meses.

Memoria

10

Costo de la Memoria

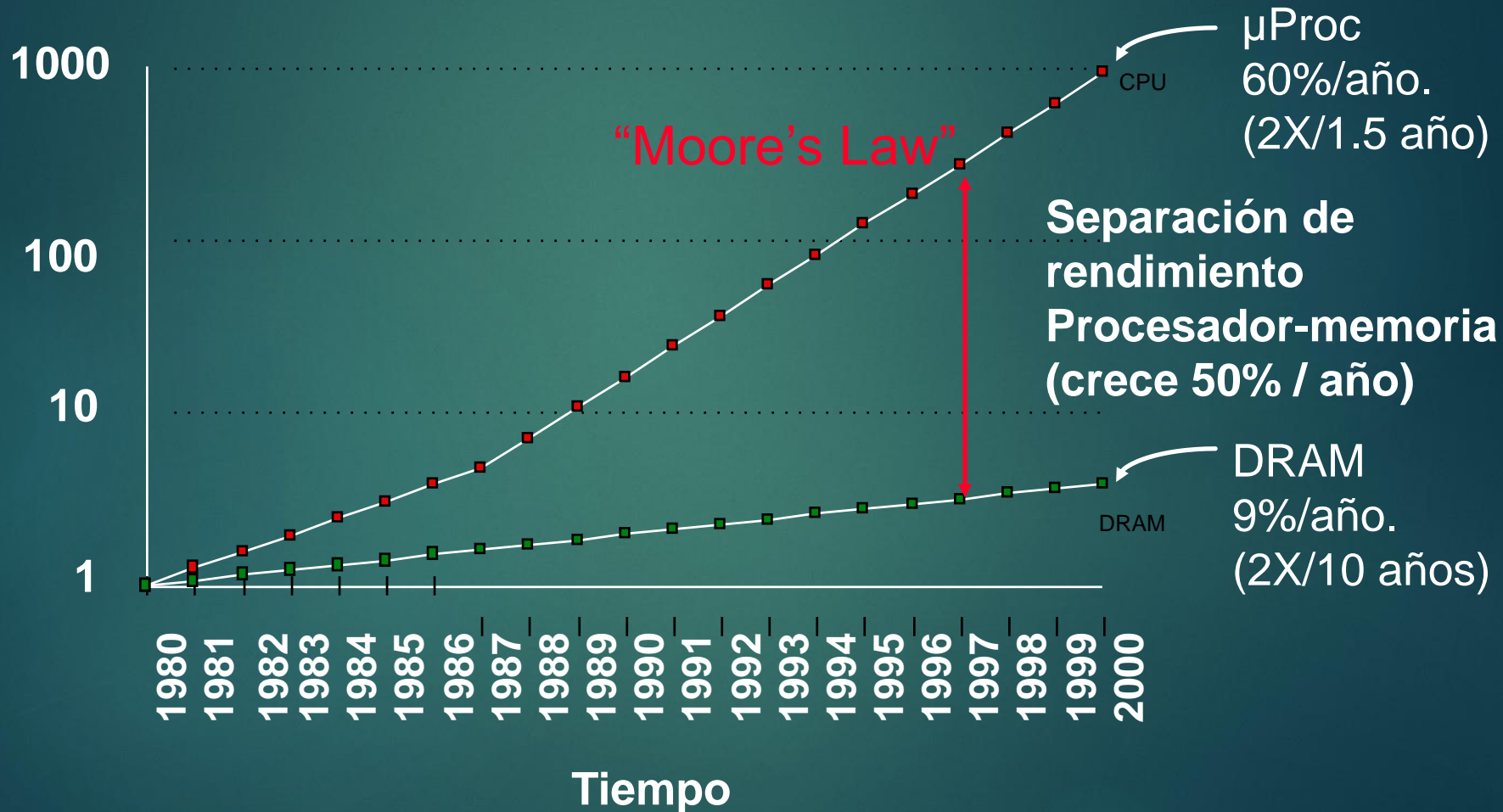
- El costo de la memoria se mide en U\$S/bit
- Depende de la tecnología: los medios de almacenamiento masivo (óptico, magnético) tienen menor costo por bit que los dispositivos de menor capacidad.
- El costo está relacionado con el tiempo de acceso: a menor tiempo de acceso mayor costo por bit.

Jerarquía de la Memoria

11

- Existe una brecha entre la velocidad del procesador y la de la memoria.
- La brecha ha ido creciendo con el tiempo, siguiendo unas curvas de crecimiento del procesador y la memoria.
- La brecha ha crecido debido a las exigencias de tamaño de la memoria (a mayor memoria, más lenta es).

Jerarquía de la Memoria



Jerarquía de la Memoria

- En el diseño de la memoria, el objetivo es balancear la diferencia de velocidades, disponer de una gran capacidad de memoria, al menor costo posible.
- Para resolver el problema de balancear velocidad, capacidad y costo, se distribuye el almacenamiento de la información en varios tipos de memorias con distintas capacidades y velocidades, pero organizadas de forma tal que se logra un comportamiento global equivalente al que tendría una memoria única, grande, rápida y barata.

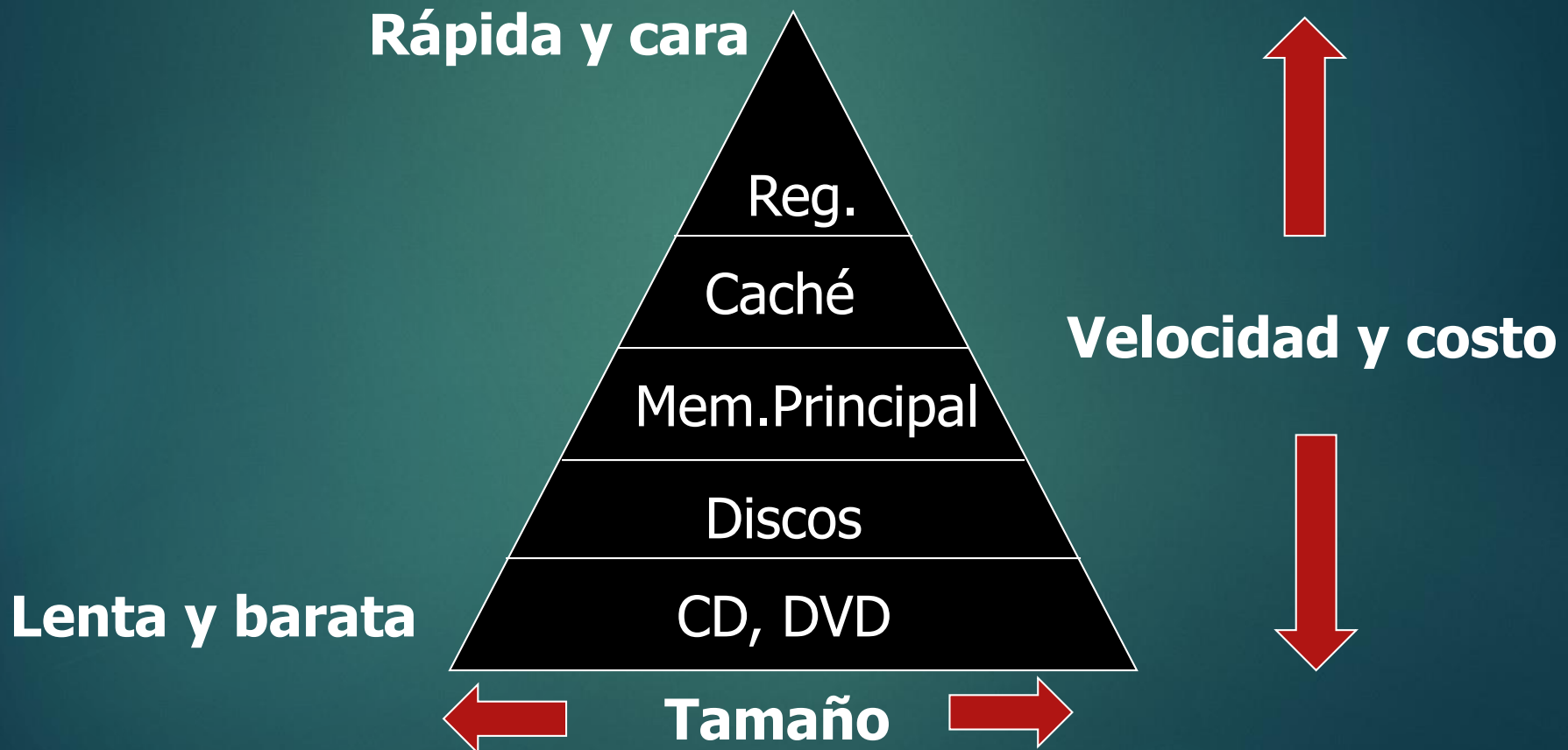
Jerarquía de la Memoria

- La forma en que se organizan estos distintos tipos de memoria es lo que se conoce como:

JERARQUIA DE MEMORIA

- La Jerarquía de Memoria se acostumbra representar como una pirámide.
- En la cima de la pirámide están los registros, en la base, las memorias de almacenamiento masivo (memorias secundarias) por ejemplo discos magnéticos y de almacenamiento “off line” (CD, DVD, cintas).

Jerarquía de la Memoria



Jerarquía de la Memoria

- A medida que ascendemos en los niveles de la jerarquía de memoria, tenemos dispositivos de memoria de mayor rendimiento y más costo por bit. Y también aumenta la frecuencia de accesos a ese tipo de memoria.
- A medida que descendemos en los niveles de la jerarquía de memoria, los dispositivos tienen mayor capacidad (de almacenamiento), son más lentos y el costo por bit disminuye.
- Se observa que existe un nivel intermedio entre los registros y la Memoria principal que se llama:

Memoria Caché

Jerarquía de la Memoria

- Los niveles jerárquicos de la memoria son administrados de manera diferente, algunos por hardware y otros por software.
- Registros: los administra el compilador, y eventualmente si tiene acceso a ellos, el programador.
- Cache: administrado por el hardware
- Memoria principal: administrado por:
 - el hardware
 - el sistema operativo (memoria virtual)
 - el programador (archivos)

Jerarquía de la Memoria

➤ En resumen, la “Memoria” del computador es un conjunto de dispositivos con:

- Diferentes tecnologías
- Diferentes fundamentos físicos
- Diferentes localizaciones

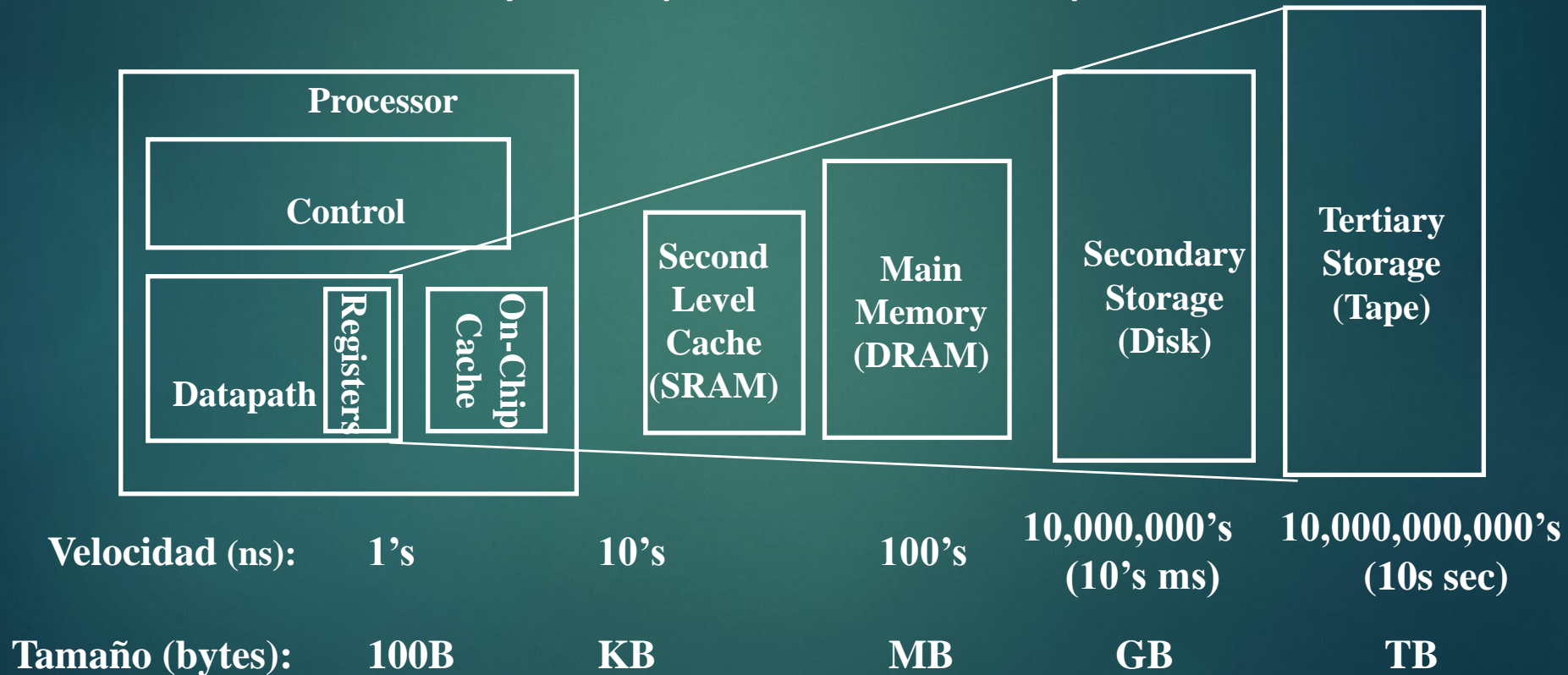
cuyo objetivo es:

- Aumentar la Capacidad de almacenamiento
- Reducir el costo
- Reducir tiempo de acceso



Jerarquía de la Memoria

- En el gráfico siguiente se muestran los principales niveles jerárquicos de la Memoria con sus capacidades de almacenamiento y tiempos de acceso típicos.



Clasificación de las memorias

- Existen diferentes formas de clasificar a las memorias, dependiendo de algunas de sus propiedades.

1) Duración de la información almacenada en la memoria

- Dependiendo de el tiempo que dura la información, las memorias pueden ser:
 - Memorias volátiles: mantienen la información mientras tengan energía. Ejemplo: RAM
 - Memorias no volátiles: no necesitan energía para retener la información. Ejemplo: discos, cintas

Clasificación de las memorias

2) Según el tipo de transferencia

- Dependiendo del tipo de transferencia empleada, las memorias pueden ser:
 - Transferencia por palabra: se transfiere una palabra por vez. Ejemplo: memoria principal
 - Transferencia por bloque: se transfiere un bloque de datos por vez. Ejemplo: discos, caché

Clasificación de las memorias

3) Según el métodos de acceso

Hay 4 mecanismos básicos de acceso a una memoria:

- Acceso aleatorio: el tiempo para acceder a una dada posición es independiente de la secuencia de accesos anteriores y es constante (RAM)

Ej: RAM, memoria principal.

- Acceso secuencial: el acceso debe hacerse en una secuencia lineal específica.

Ej: unidades de cinta.

Clasificación de las memorias

- Acceso “directo”: los bloques o registros individuales tienen una dirección única que se basa en la localización física (coordenada).

Ej: discos magnéticos (combinación de aleatoria y secuencial).

- Acceso asociativo: el acceso se hace “buscando” contenidos y no por ubicación en la memoria.

Ej: memoria caché

Las memorias asociativas se conocen normalmente como memorias “CAM”, por las siglas de Content addressable memory access.

Clasificación de las memorias

4) Según el tipo de operación que se puede ejecutar

- Memorias de lectura/escritura: son las memorias que se pueden leer y escribir. Técnicamente se conocen como memorias R/W (por Read/write memory).
 - Ejemplo: memoria principal, memoria RAM

Las memorias de lectura/escritura comúnmente se las llama memorias RAM (Random Access Memory). Estrictamente hablando, las memorias RAM son memorias donde el tiempo de acceso es independiente de la posición del dato.

- Memorias de lectura solamente: son las memorias que solo pueden ser leídas. Técnicamente se conocen como memorias ROM (por Read Only memory)
 - Ejemplo: memoria BIOS

Clasificación de las memorias

5) Según el tiempo que dura la información amacenada

- Estáticas: la información almacenada se mantiene mientras la memoria reciba alimentación eléctrica. Se conocen como memorias SRAM.
- Dinámicas: la información almacenada se debe “refrescar” periódicamente para que no se pierda. Se conocen como memorias DRAM.

Clasificación de las memorias

➤ Estáticas:

- La celda que almacena información es del tipo flip flop
- Generalmente tiene menor tiempo de acceso

➤ Dinámicas:

- La celda binaria está basada en transistores, y capacitores que almacenan carga. Debido a que los capacitores van perdiendo la carga con el tiempo, se necesita recargarlos periódicamente. Este proceso se conoce como “refresco”.
- Las celdas de memoria a capacitor tienen menor superficie que las celdas a flip flop. Como las celdas son más pequeñas, las DRAM tienen mayor densidad de celdas que las SRAM.
- Por lo tanto las memorias DRAM almacenan más información que las SRAM en la misma superficie.

Organización física de la memoria

27

- El elemento básico de una memoria semiconductora es la celda de memoria.
- Todas las celdas de una memoria semiconductora comparten 3 propiedades:
 - Dos estados estables: uno (1) y cero (0).
 - Se puede escribir en ellas (al menos una vez)
 - Se pueden leer para conocer el estado.

Organización física de la memoria

28

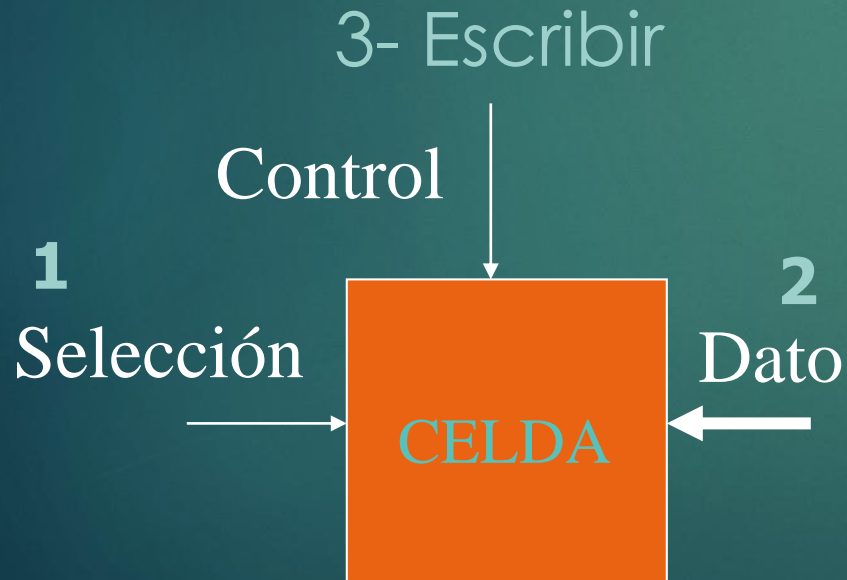
- De acuerdo a lo visto en clases anteriores, en general la celda de memoria requiere de 3 informaciones (o variables):
 - El dato almacenado en la celda de memoria.
 - La selección de la celda de memoria.
 - La indicación del tipo de acción sobre la celda: escritura o lectura

Organización física de la memoria

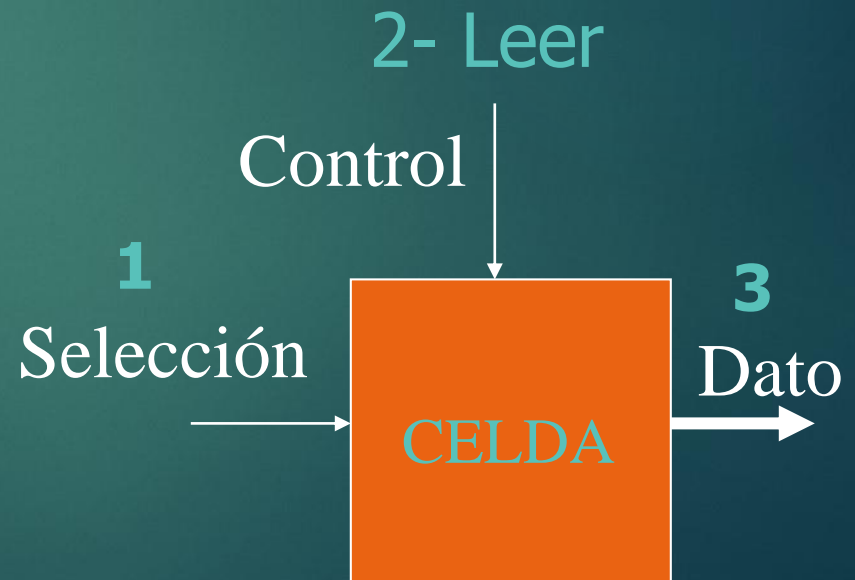
29

- El orden en el que se presentan las señales en una celda de memoria para la lectura o la escritura es el siguiente (los números indican el orden):

Escritura de Datos



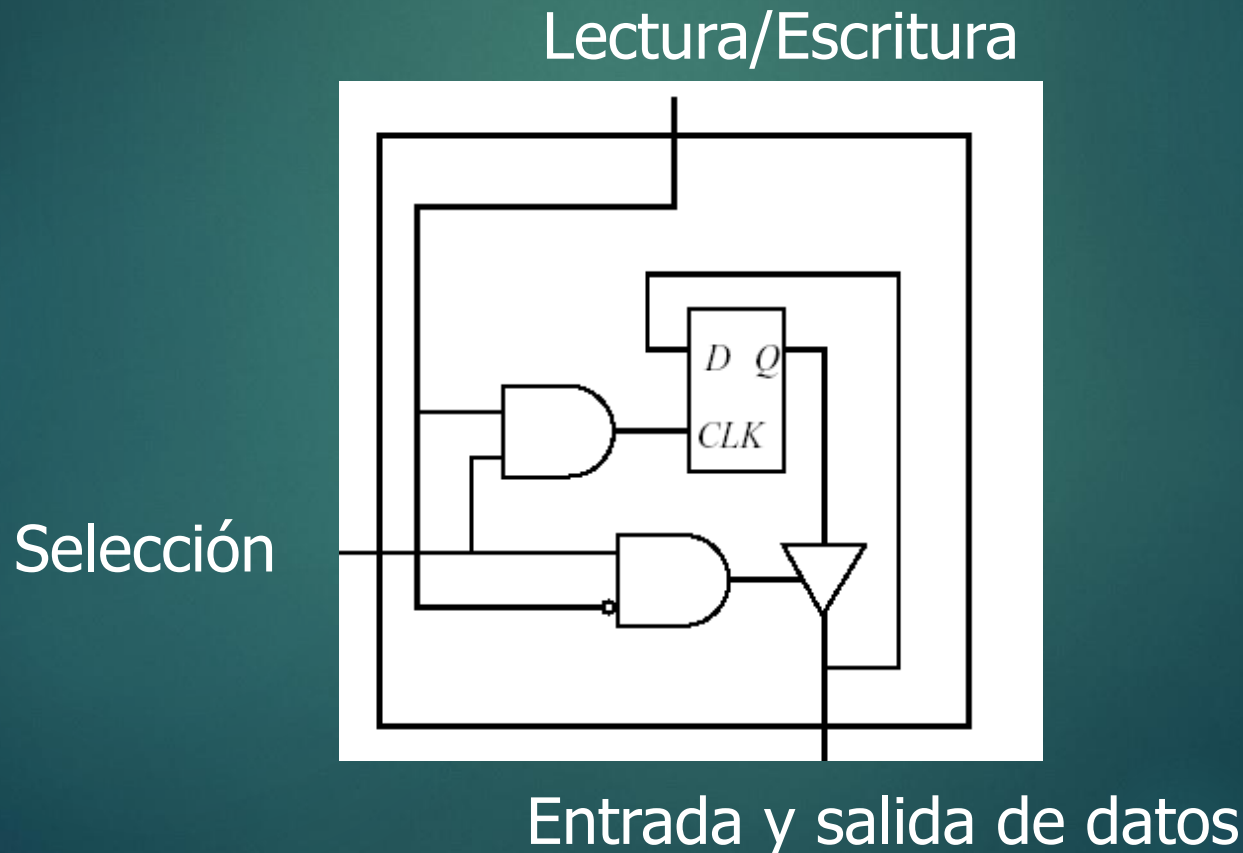
Lectura de Datos



Organización física de la memoria

30

- Físicamente, una celda de memoria basada en un FF tipo D contiene, además del FF, algunas compuertas que permiten implementar el acceso a la celda.



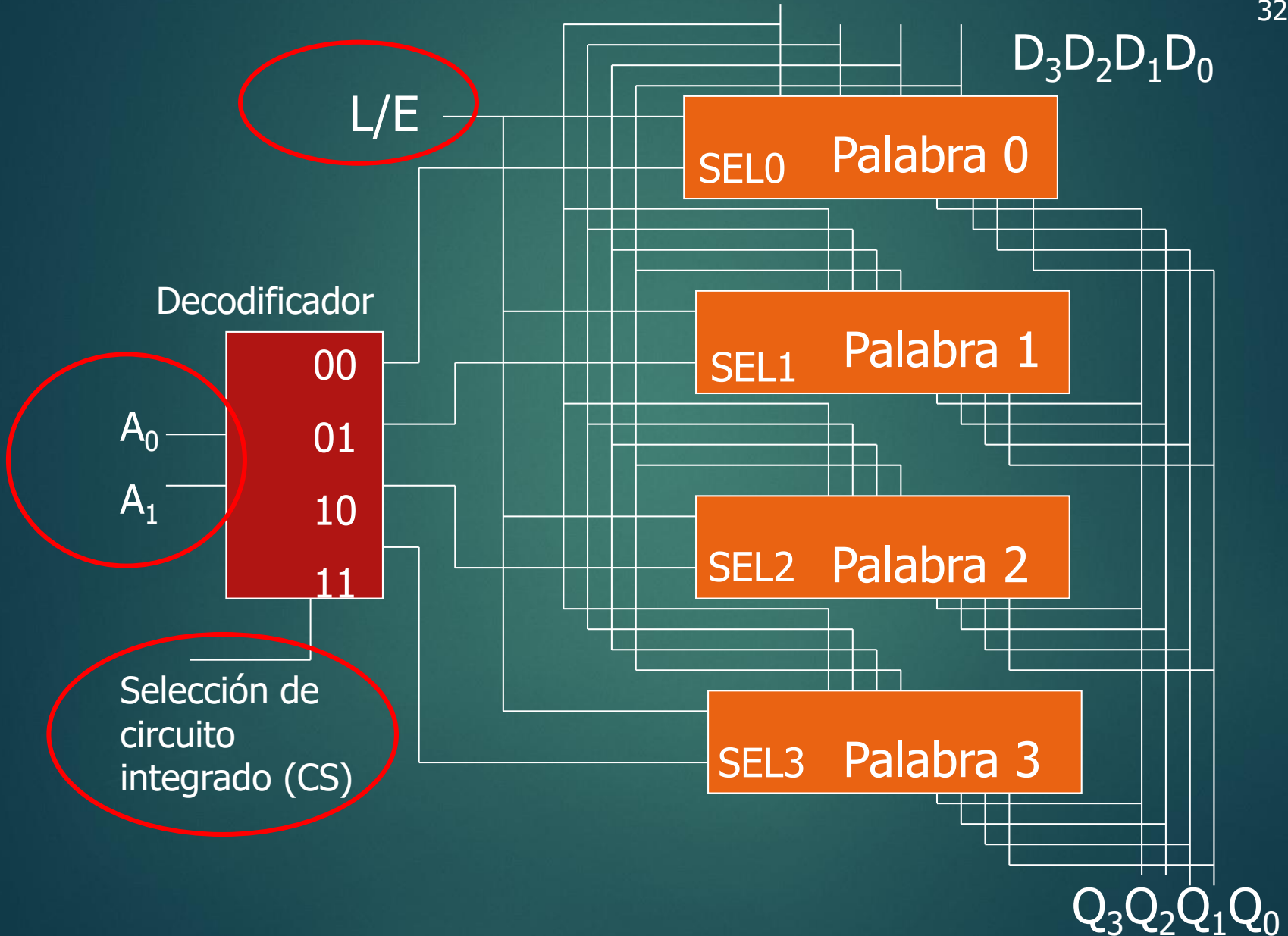
Organización física de la memoria

31

- Se puede construir una memoria de 1 bit usando un 1 FF tipo D. Se puede construir un registro de n bits mediante un arreglo de n FF tipo D (análisis en clase 3 del presente curso).
- Para construir memorias se requiere una organización un poco diferente, en la cual sea posible 'direccionar' palabras individuales.
- En la figura siguiente se muestra un arreglo de 4 palabras de memoria con las siguientes características:
 - Todas las palabras usan las mismas líneas para la entrada ($D_3D_2D_1D_0$) y salida ($Q_3Q_2Q_1Q_0$) de los datos.
 - La señal que identifica el tipo de operación (lectura/escritura) es común a todas las palabras (L/E).
 - Cada palabra tiene su señal de selección (SEL0, SEL1, SEL2, SEL3).
 - La selección de la palabra se hace por medio de 2 señales codificadas (A_0A_1) que son señales de direcciones.

Organización física de la memoria

32



Organización de chips de memoria

33

- Cuando se diseñan memorias de gran tamaño, se requiere disponer de varios circuitos integrados (“chips”) de memoria.
- Cada chip contiene un arreglo de celdas de memoria.
- De los chips de memoria importan 2 factores:
 - Cantidad de celdas de memoria, es decir la “capacidad” del chip.
 - Organización de las celdas de memoria.
- En las memorias semiconductoras se han empleado 2 enfoques organizacionales: 2D y 2½D.

Organización de chips de memoria

34

Modelo de organización de memoria 2D

- Es el modelo visto hasta ahora.
- Cada chip contiene un arreglo de bits organizado en una matriz de:

2^W filas x B columnas
- Las filas son las palabras de la memoria.
- Las columnas son el tamaño de la palabra (todas las palabras tienen el mismo tamaño de bits).
- Cada chip de memoria provee todos los bits de la palabra.

Organización de chips de memoria

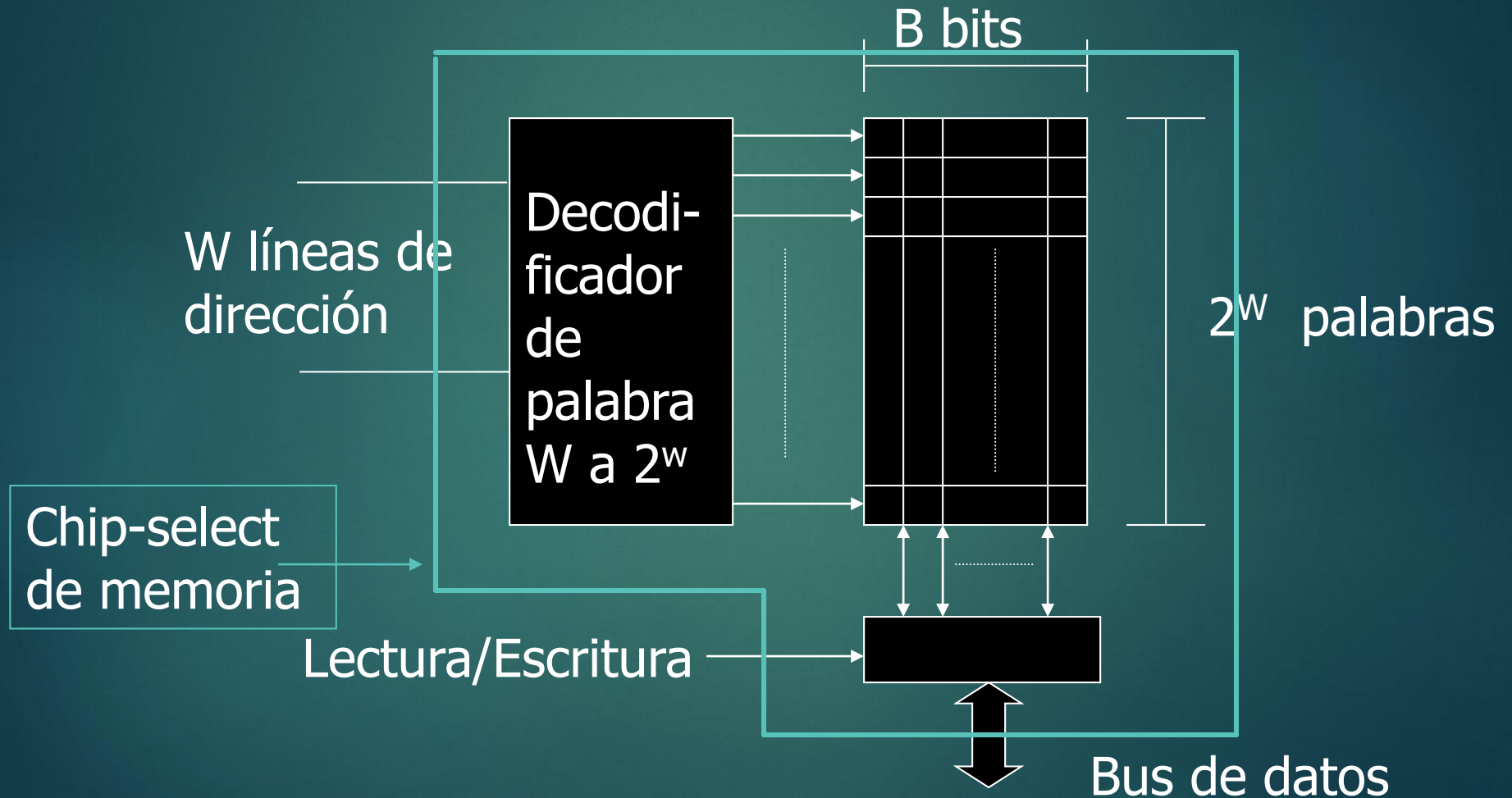
35

- Las filas se seleccionan de a 1, mediante un decodificador de W entradas y 2^W salidas.
- Las W entradas se conectan al Bus de direcciones.
- El chip se debe “seleccionar” mediante una señal adicional (puede ser alguna línea de dirección).
- Todas las celdas de todas las palabras están conectadas a las entradas y salidas de datos de la memoria.
- Las entradas/salidas de datos de la memoria se conecta al Bus de datos.

Organización de chips de memoria

36

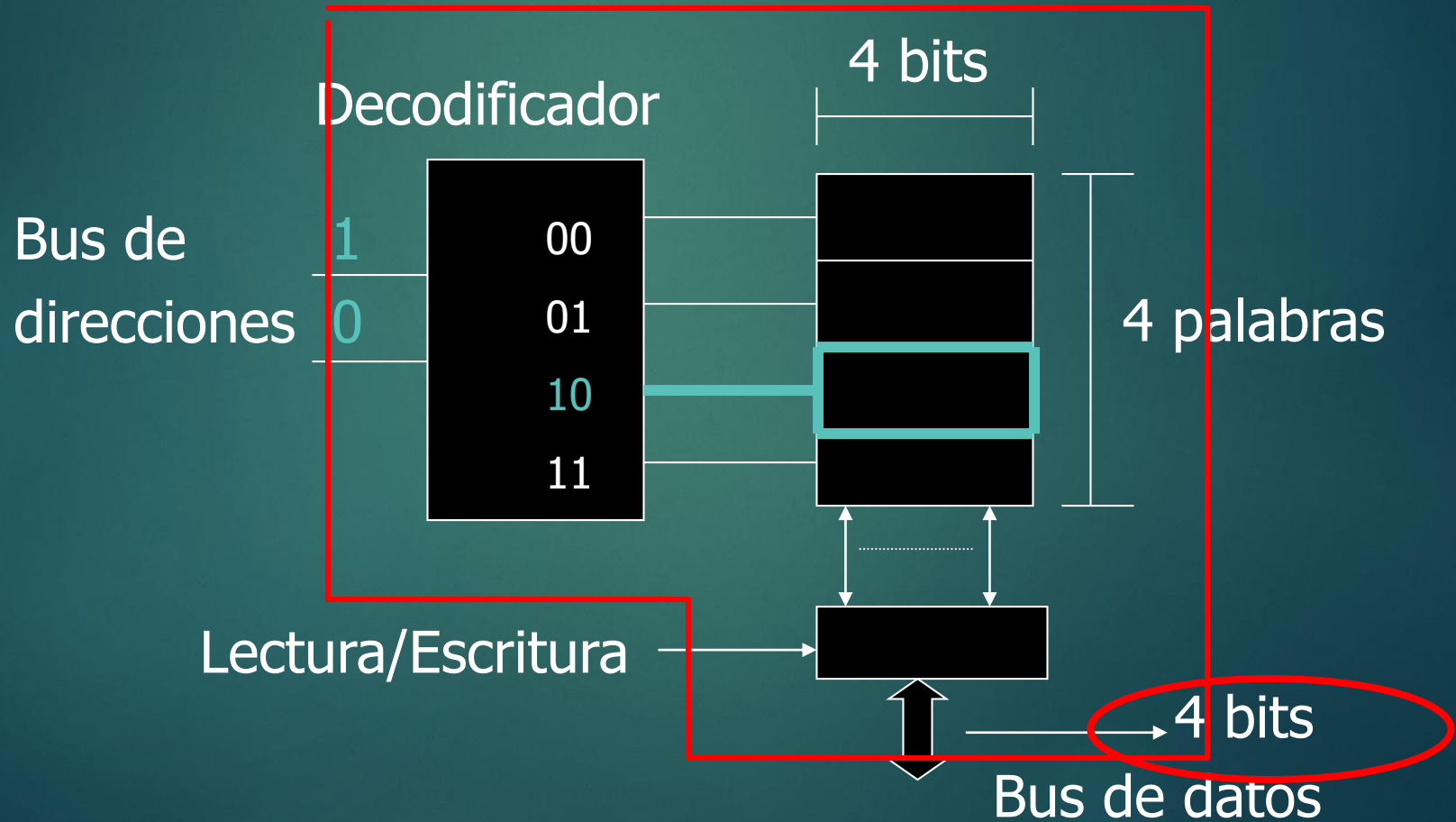
Esquema de chip de memoria para organización 2D



Organización de chips de memoria

37

Ejemplo: memoria de 16 bits organizada en 4 palabras de 4 bits por palabra



Organización de chips de memoria

38

Modelo de organización de memoria 2 ½ D

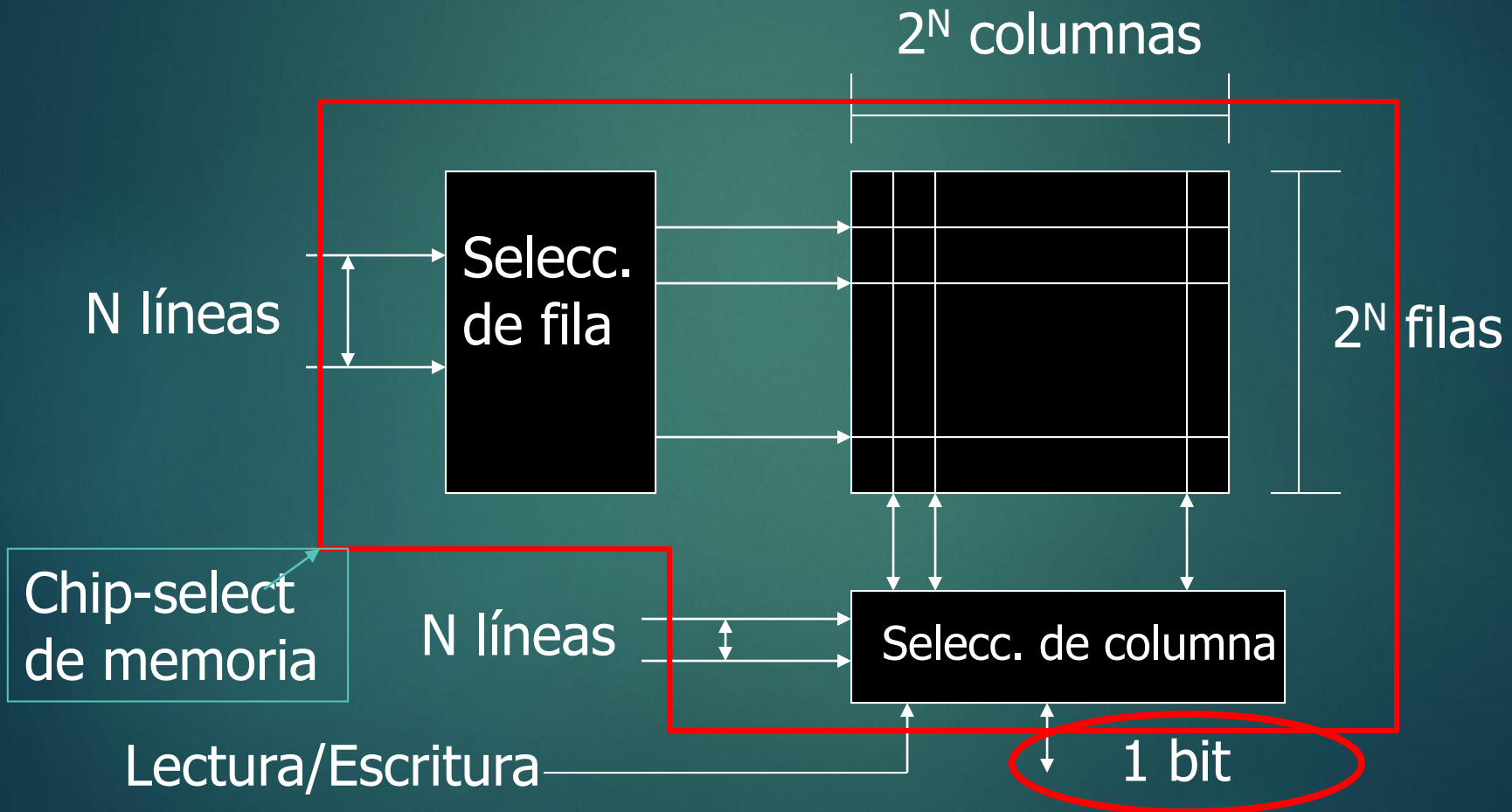
- El arreglo de memoria de cada chip es una matriz típicamente cuadrada (para reducir la cantidad de líneas de direcciones).
- Es decir, cada chip contiene un arreglo de bits organizado en una matriz de:

2^N filas x 2^N columnas
- Con los N bits de fila y los N bits de columna se selecciona 1 solo bit de la matriz, que es el único bit a ser accedido (para lectura o escritura).
- Cada chip de memoria provee solo 1 bit de la palabra.

Organización de chips de memoria

39

Esquema de chip de memoria para organización $2^{\frac{1}{2}} D$



Ej: memoria
de 64 bits

Decodificador de filas

A_2 0

A_1 1

A_0 1

A_5

A_4

A_3

000 001 010 011

000

001

010

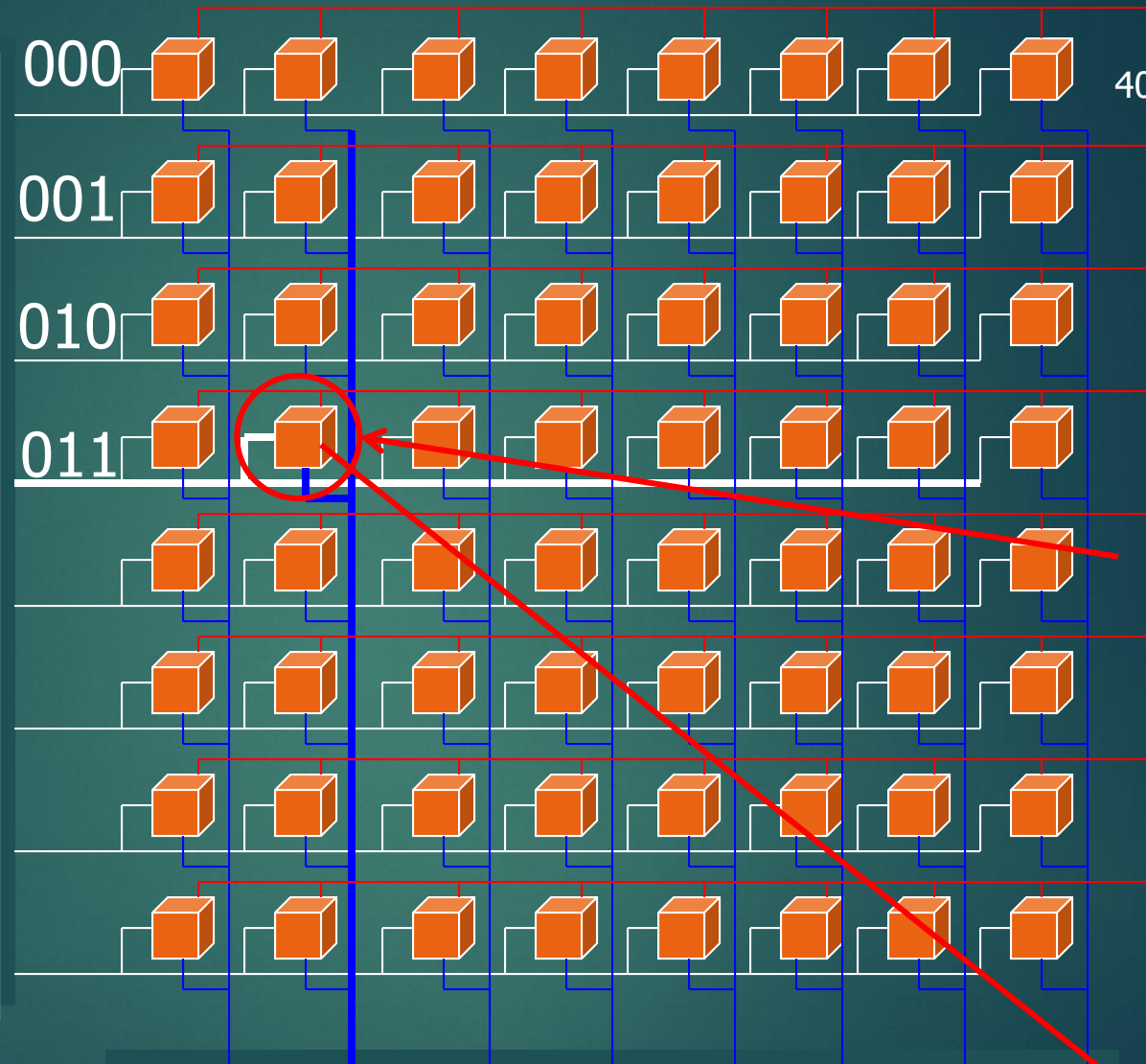
011

40

Celda

Bit

Decodificador de columnas



Organización de chips de memoria

41

Comparando los 2 sistemas de organización de la memoria, se tiene:

➤ Distribución de bits:

- En 2D todos los bits están en el mismo chip
- En 2½D los bits de una misma palabra estarán en distintos chips.

➤ Decodificación:

- En 2D cada palabra requiere una línea de selección. Para N líneas de dirección habrá 2^N líneas de selección, por lo que el tamaño (área del chip) del decodificador crece exponencialmente.
- En 2½D para N líneas de dirección, se requieren 2 decodificadores de $2^{N/2}$ líneas de selección, por lo que el tamaño (área del chip) del decodificador crece exponencialmente, pero con menor superficie que el de 2D.

Organización de chips de memoria

42

- Errores:
 - En 2D es difícil usar circuitos correctores de error porque todos los bits de una palabra están en el mismo chip.
 - En 2½D al estar los bits dispersos en distintos chips, es más fácil agregar circuitos correctores de error, y además, hay menor probabilidad de error.

Organización de chips de memoria

Arreglos combinados en 2D

43

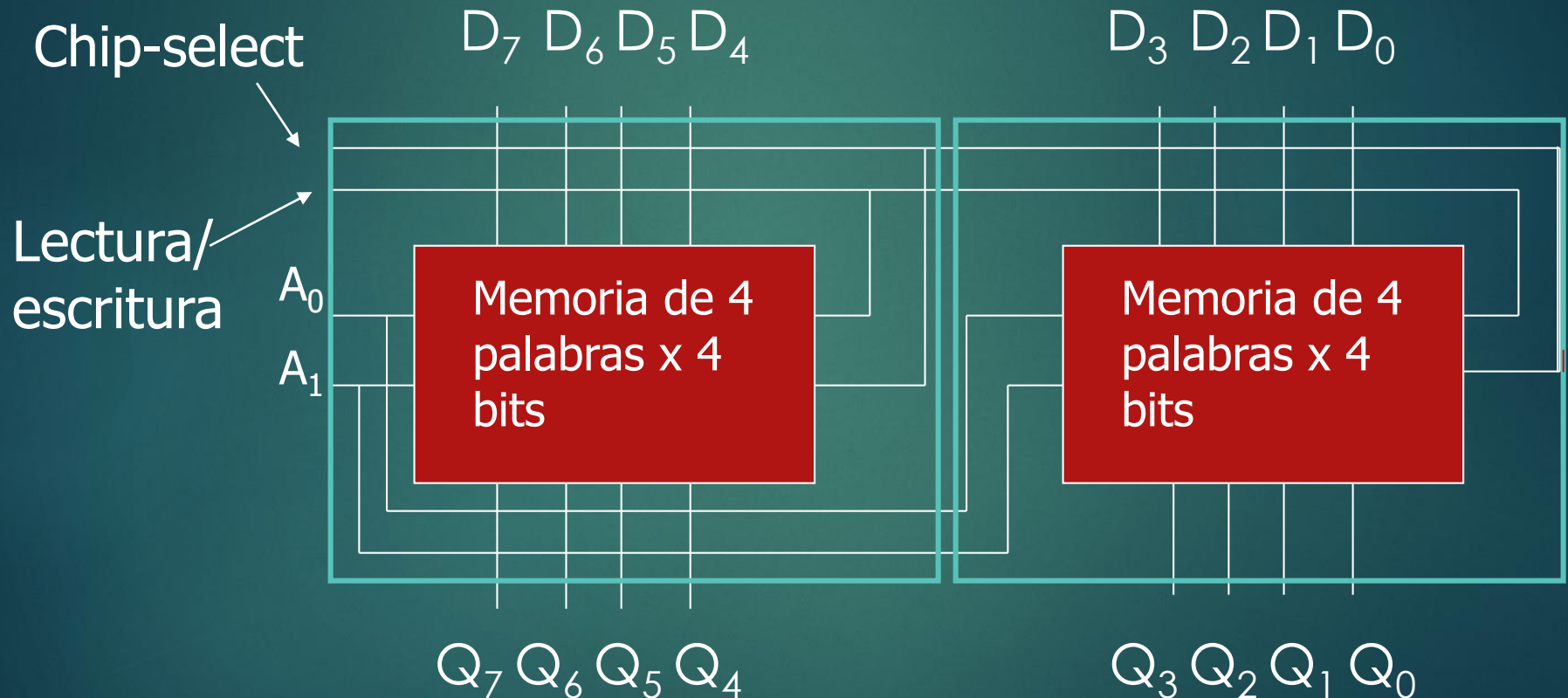
- Cuando se requiere implementar una memoria con chips de memoria de menor tamaño que el requerido, se pueden tener 2 situaciones:
 - Opción 1: El chip de memoria tiene la cantidad de palabras de memoria necesaria, pero el tamaño de la palabra es una fracción del requerido.
 - Opción 2: El chip de memoria tiene el tamaño de la palabra requerido, pero el chip tiene una fracción de la cantidad de palabras necesarias.
- En ambos casos la solución es combinar los chips de memoria.
- Hay 2 formas de combinar los chips de memoria: “paralelo” y “serie”.

Organización de chips de memoria

Arreglos combinados en 2D

44

Esquema de 2 chips de memoria 2D en “paralelo”



Organización de chips de memoria

Arreglos combinados en 2D

45

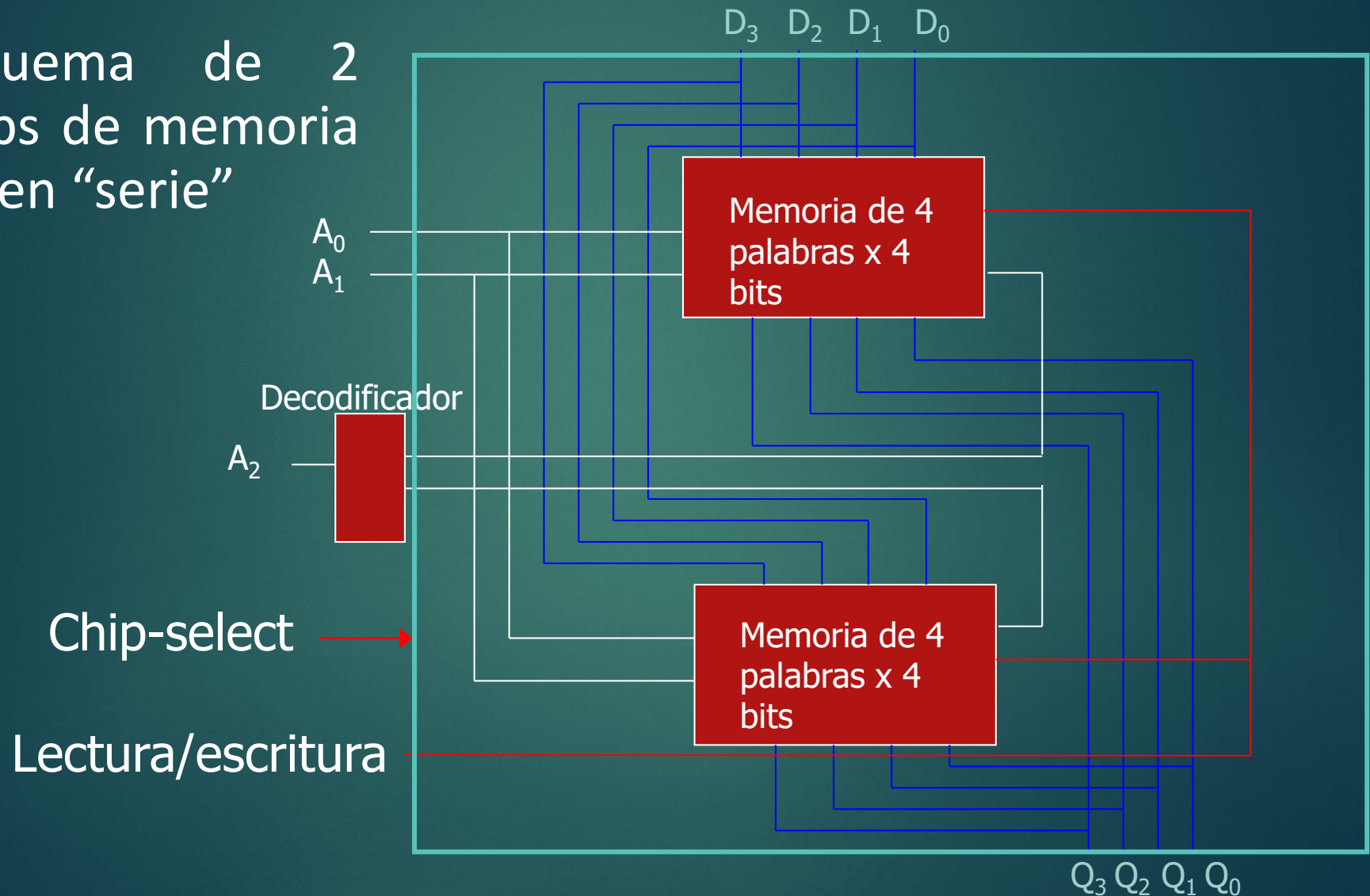
- En el arreglo “paralelo” de chips de memoria 2D, se tiene:
 - Ambos chips reciben las mismas señales de dirección, lectura/escritura y selección de chip (chip-select).
 - Cada chip aporta una fracción del dato.
 - La palabra completa del dato a ser leído o escrito, está distribuido en los chips de memoria.
 - El tamaño del dato deber ser un múltiplo entero del tamaño de la palabra del chip.

Organización de chips de memoria

Arreglos combinados en 2D

46

Esquema de 2
chips de memoria
2D en "serie"



Organización de chips de memoria

Arreglos combinados en 2D

47

- En el arreglo “serie” de chips de memoria 2D, se tiene:
 - Ambos chips reciben las mismas señales de datos, lectura/escritura y selección de chip (chip-select).
 - Cada chip aporta un rango de palabra.
 - El rango de direcciones completo es la suma de los rangos de cada chip de memoria.
 - La selección de la palabra dentro de un chip se hace con la parte baja de la dirección, que es común a todos los chips (A_0 y A_1 en el ejemplo anterior).
 - La selección del chip se hace con la fracción superior de la dirección (A_2 en el ejemplo anterior).

Evolución de las RAM

48

DRAM

- Las memorias RAM de las computadoras personales (PC) han ido evolucionando a lo largo de los años, tanto en capacidad, como en prestaciones y organización interna.
- Las primeras RAM que se incluyeron en las PC fueron las DRAM (Dynamic RAM).
- Las celdas debían refrescar periódicamente, es decir que había que accederlas en períodos regulares de tiempo para no perder la información almacenada. Este refresco se hacía mediante circuitos internos del Sistema.
- El acceso era de tipo asíncrono, es decir, sin sincronización con el reloj del Procesador.

Evolución de las RAM

49

EDRAM

- Posteriormente surgieron las memorias EDRAM (Enhanced DRAM).
- Estaban compuestas por una DRAM grande, y una RAM estática (SRAM) pequeña para aumentar la velocidad de acceso. Las memorias SRAM eran de mayor velocidad (es decir, menor tiempo de acceso) que las DRAM.
- El acceso era de tipo asincrónico, es decir, sin sincronización con el reloj del Procesador.
- La CPU no sabía cuando iba a estar disponible el dato de la memoria. La CPU debía esperar todo el tiempo hasta que la memoria le devolvía el dato.

Evolución de las RAM

50

SDRAM

- En la etapa siguiente de la evolución, surgieron las memorias SDRAM (Synchronous DRAM) (año 1996).
- Mecánicamente estaban montadas en un formato conocido como DIMM.



Evolución de las RAM

51

SDRAM

- Como su nombre lo indica, el acceso a la memoria está sincronizado con el reloj del sistema.
- De esta manera, la CPU sabe exactamente cuando los datos están disponibles (en múltiplos de períodos de reloj). Durante el tiempo que tarda la memoria en devolver el dato, la CPU puede realizar otra tarea.
- La memoria dispone, además, de un modo de operación conocido como ráfaga (“burst”) para aumentar la velocidad de transferencia de bloques de memoria.

Evolución de las RAM

52

DDRx-SRAM

- Posteriormente las memorias evolucionaron al modo DDR: Double Data Rate.
- El modo DDR permite mayor cantidad de accesos en cada ciclo de reloj.
- De esta manera aumenta la velocidad de transferencia por unidad de tiempo.
- La primer version fué DDR-SRAM, que dispone de:
 - 2 acceso a memoria por ciclo de reloj (acceso en ambos flancos del reloj).
 - Duplica aproximadamente la velocidad de una SDRAM.

Evolución de las RAM

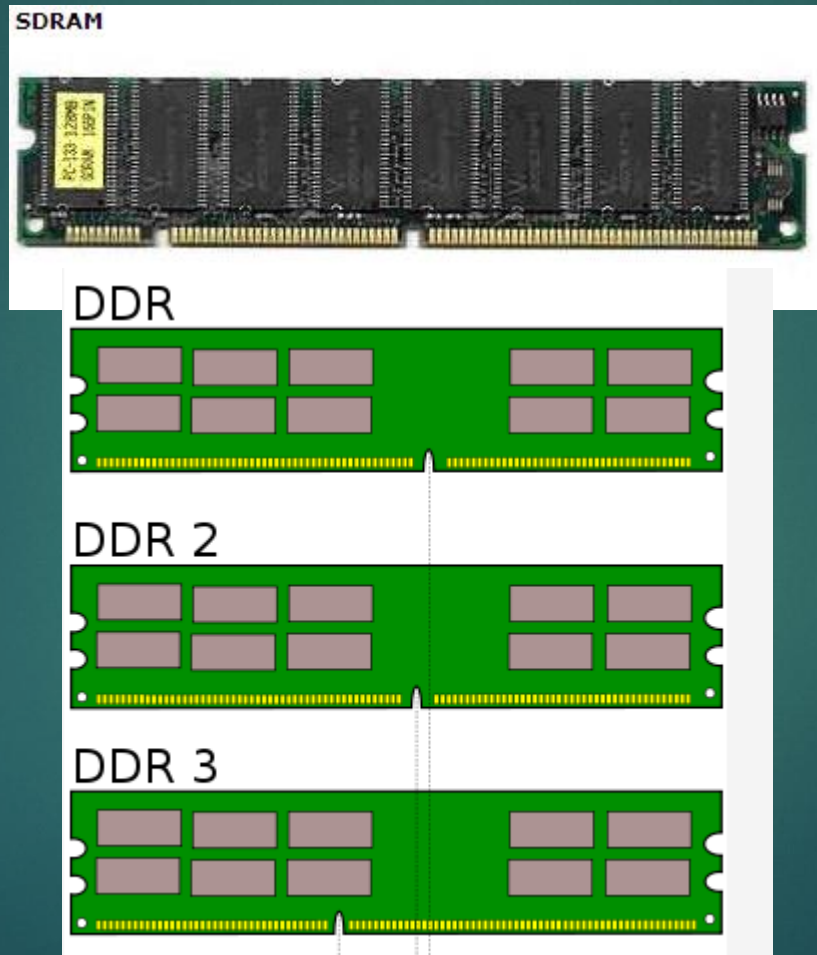
53

- DDR2-SRAM (Double data rate type 2) admite 4 accesos por ciclo de reloj.
- DDR3-SRAM (Double data rate type 3) admite 8 accesos por ciclo de reloj.

SDRAM, DDR, DDR2, DDR3

54

Formatos físicos SDRAM, DDR, DDR2, DDR3



SDRAM, DDR, DDR2, DDR3

55

Datos característicos SDRAM



Nombre estándar	Velocidad del bus	Tiempo entre señales	Velocidad del módulo	Datos transferidos por segundo	Nombre del módulo	Máxima capacidad de transferencia
SDR-66		15 ns	66,6 MHz	66 Millones	PC-66	533 MB/s
SDR-100		10 ns	100 MHz	100 Millones	PC-100	800 MB/s
SDR-133		7,5 ns	133,3 MHz	133 Millones	PC-133	1066 MB/s

SDRAM, DDR, DDR2, DDR3

56

Datos característicos DDR

DDR:



Nombre estándar	Reloj de Bus	Tiempo entre señales	Reloj de memoria	Datos transferidos por segundo	Nombre del módulo	Máxima capacidad de transferencia
DDR-200	100 MHz		100 MHz	200 Millones	PC-1600	1600 MB/s (1,6 GB/s)
DDR-266	133 MHz		133 MHz	266 Millones	PC-2100	2128 MB/s (2,1 GB/s)
DDR-333	166 MHz		166 MHz	333 Millones	PC-2700	2656 MB/s (2,6 GB/s)
DDR-400	200 MHz		200 MHz	400 Millones	PC-3200	3200 MB/s (3,2 GB/s)

SDRAM, DDR, DDR2, DDR3

57

Datos característicos DDR2

DDR2:

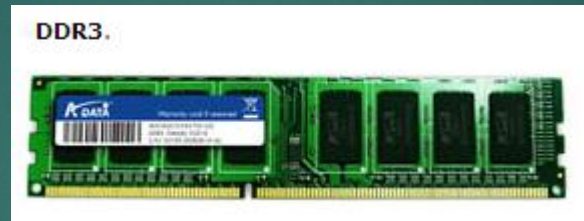


Nombre estándar	Velocidad del reloj	Tiempo entre señales	Velocidad del reloj de E/S	Datos transferidos por segundo	Nombre del módulo	Máxima capacidad de transferencia
DDR2-400	100 MHz	10 ns	200 MHz	400 millones	PC2-3200	3200 MB/s
DDR2-533	133 MHz	7,6 ns	266 MHz	533 millones	PC2-4200	4264 MB/s
DDR2-600	150 MHz	6,7 ns	300 MHz	600 millones	PC2-4800	4800 MB/s
DDR2-667	166 MHz	6 ns	333 MHz	667 Millones	PC2-5300	5336 MB/s
DDR2-800	200 MHz	5 ns	400 MHz	800 Millones	PC2-6400	6400 MB/s
DDR2-1000	250 MHz	3,75 ns	500 MHz	1000 Millones	PC2-8000	8000 MB/s
DDR2-1066	266 MHz	3,75 ns	533 MHz	1066 Millones	PC2-8500	8530 MB/s
DDR2-1150	286 MHz	3,5 ns	575 MHz	1150 Millones	PC2-9200	9200 MB/s
DDR2-1200	300 MHz	3,3 ns	800 MHz	1200 Millones	PC2-9600	9600 MB/s

SDRAM, DDR, DDR2, DDR3

58

Datos característicos DDR3



Nombre estándar	Velocidad del reloj	Tiempo entre señales	Velocidad del reloj de E/S	Datos transferidos por segundo	Nombre del módulo	Máxima capacidad de transferencia
DDR3-1066	133 MHz	7,5 ns	533 MHz	1066 Millones	PC3-8500	8530 MB/s
DDR3-1200	150 MHz	6,7 ns	600 MHz	1200 Millones	PC3-9600	9600 MB/s
DDR3-1333	166 MHz	6 ns	667 MHz	1333 Millones	PC3-10600	10 664 MB/s
DDR3-1375	170 MHz	5,9 ns	688 MHz	1375 Millones	PC3-11000	11 000 MB/s
DDR3-1466	183 MHz	5,5 ns	733 MHz	1466 Millones	PC3-11700	11 700 MB/s
DDR3-1600 ⁸	200 MHz	5 ns	800 MHz	1600 Millones	PC3-12800	12 800 MB/s
DDR3-1866	233 MHz	4,3 ns	933 MHz	1866 Millones	PC3-14900	14 930 MB/s
DDR3-2000	250 MHz	4 ns	1000 MHz	2000 Millones	PC3-16000	16 000 MB/s
DDR3-2200	350 MHz	3,3 ns	1200 MHz	2200 Millones	PC3-18000	18 000 MB/s

Referencias

- Stallings. 5ta Ed.
 - Capítulo 4: Memoria Interna
- Links de interés
 - <http://www.pctechguide.com/14Memory.htm>