

# Organización de Computadoras 2020

TURNO RECURSANTES  
CLASE 9

# Resumen clase 9

- Memoria cache
- Memoria secundaria
- Discos magnéticos
- Formatos de grabación en discos magnéticos

# CPU vs MEMORIA

3

- Históricamente la CPU ha sido más rápida que la memoria.
- El aumento en la cantidad y complejidad de los circuitos integrados que componen los chips de CPU y memoria se ha orientado principalmente a:
  - CPU: mejorar la performance, para hacerla más veloz (ej.: uso de pipeline).
  - Memoria: aumentar la capacidad de almacenamiento del chip (más memoria, más grandes decodificadores).

# CPU vs MEMORIA

- Esta diferencia de velocidades implica que después que la CPU 'emite' una solicitud de lectura a la memoria (a través de los buses de direcciones, y control) pasan varios ciclos de reloj antes de recibir la información requerida, en el bus de datos.
- En todos los ciclos de instrucción, la CPU accede a la memoria:
  - al menos una vez, para buscar la instrucción
  - varias veces, para buscar operandos.
- Dada la necesidad de acceder a la memoria, la velocidad a la cual la CPU ejecuta instrucciones está limitada parcialmente por el tiempo del ciclo de memoria.

# CPU vs MEMORIA

5

- El problema del desbalance entre la velocidad de la CPU y la de la memoria no es puramente tecnológico sino también económico.
- Se pueden construir memorias muy rápidas, casi tanto como la CPU, pero para obtener la máxima velocidad se requiere que estén dentro del chip de la CPU, y así evitar el uso del bus.
- El acceso a la memoria por el bus del sistema es inherentemente 'lento' (en comparación con la velocidad interna de la CPU).

# Memoria Caché

6

- Una forma de compensar el desbalance entre CPU y memoria es usar una Jerarquía de memoria.
- En la Jerarquía de memoria se combinan memorias muy rápidas y pequeñas con memorias grandes y más lentas, para obtener una velocidad conjunta (de ambas memorias) 'casi' (en promedio) tan rápida.
- La memoria muy rápida y pequeña que integra la Jerarquía de memoria es la Memoria Caché.

# Memoria Caché

7

El concepto de Memoria Caché es muy simple:

- La memoria Cache es una memoria muy rápida y pequeña intercalada entre la CPU y la Memoria Principal.
- Casi tan rápida como la CPU, y de tamaño miles de veces mas chica que la Principal.
- Cuando la CPU busca una información, si está en la Caché la trae desde ahí, y si no, se busca en la memoria principal.

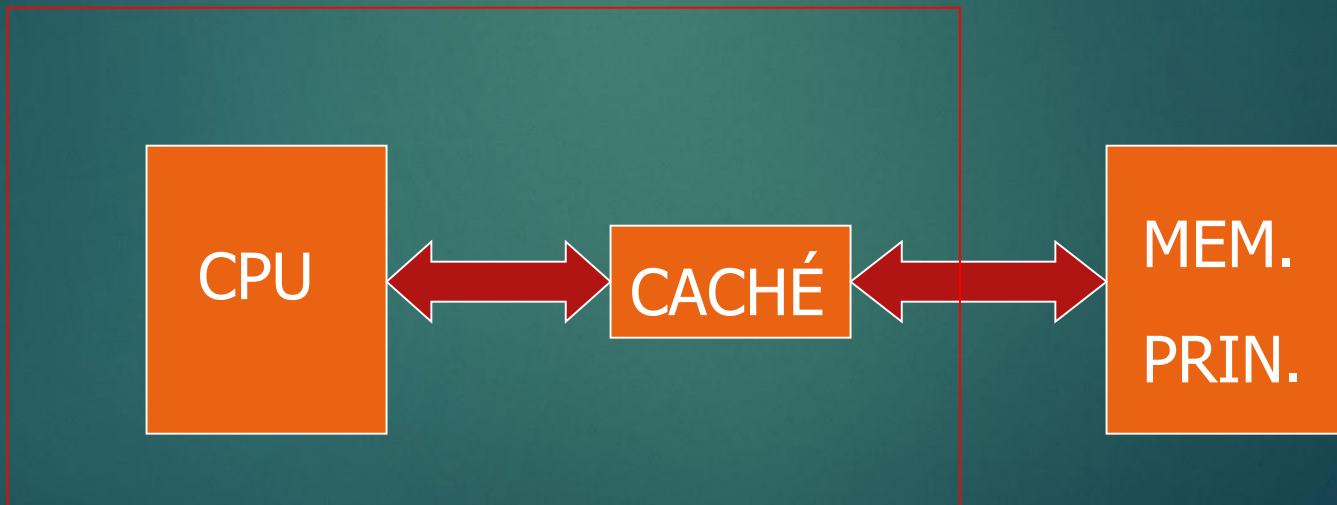


# Memoria Caché

8

## Ubicación física de la Memoria Caché

La memoria Caché físicamente está intercalada entre la CPU y la memoria Principal, y puede estar integrada en el mismo chip de la CPU o fuera de él.





# Memoria Caché

- Para poder ser muy rápida, la Memoria Caché es relativamente mucho más chica que la Memoria Principal.
- Por lo tanto, solo contiene una fracción muy pequeña de la información de la Memoria Principal.
- El uso de la Memoria Caché se sustenta en la forma determinística (no aleatoria) en que se comportan los programas.
- En otras palabras, los programas tienen un comportamiento predecible.

# Memoria Caché

10

El comportamiento determinístico se basa en 2 “principios empíricos”:

- Principio de localidad espacial (de la referencia)
  - “Cuando se accede a una palabra de memoria, es altamente probable que el próximo acceso sea en la vecindad de la palabra anterior”.
- Principio de localidad temporal (de la referencia)
  - “Cuando se accede a una posición de memoria, es altamente probable que en un lapso de tiempo corto dicha posición de memoria sea accedida nuevamente”.

# Memoria Caché

11

- El principio de Localidad espacial se sustenta en la forma normal en la que la CPU accede a las instrucciones y los datos:
  - Acceso a instrucciones (código): estadísticamente, un porcentaje elevado de instrucciones se ejecuta accediendo a instrucciones consecutivas en memoria (acceso secuencial de instrucciones).
  - Accesos a datos: Los programadores tienden a asignar variables relacionadas próximas entre sí, de tal manera que su acceso está en un entorno cercano. Por otra parte, las estructuras de datos (matrices o pilas) se almacenan en un área de memoria cercana, haciendo que su acceso cumpla el principio de localidad de la referencia espacial.

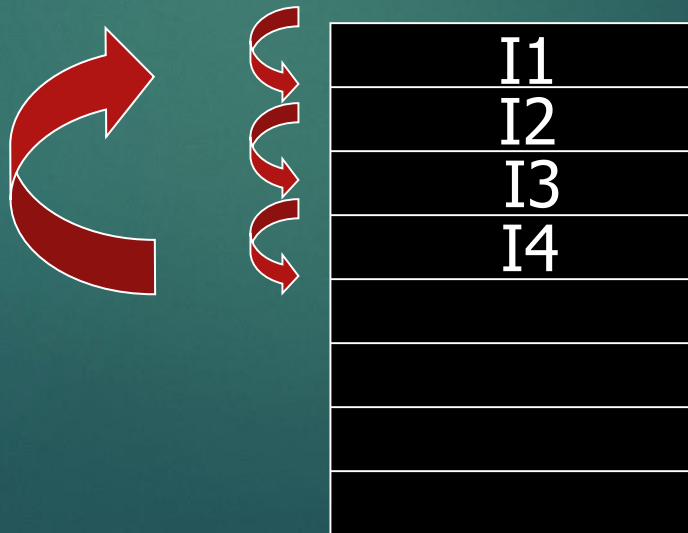
# Memoria Caché

12

- El principio de Localidad temporal, se sustenta en la forma normal en la que la CPU accede a:
  - Código:
    - Ciclos o bucles: en los ciclos o bucles la CPU busca una secuencia de instrucciones cercanas repetitivamente (la cantidad de veces que requiere el bucle).
    - Subrutinas (Procedimientos o Funciones)
  - Datos:
    - Accesos repetitivos a los mismos datos
    - Estructuras de datos (matrices, vectores, etc.)
    - Pilas

# Memoria Caché

- Un ejemplo típico que muestra el principio de la localidad de la referencia espacial y temporal es el de la ejecución de un lazo o bucle.
- Las instrucciones I1, I2, I3 e I4 se acceden secuencialmente, en un entorno cercano y en forma repetitiva, durante todo el tiempo que dura el bucle.



# Memoria Caché

14

- Otro ejemplo donde se aprecian los principios de la localidad temporal y espacial es en el caso de las siguientes sentencias:

for i=1 to i=10, do

A[i]:=0;

- Estas 2 sentencias exhiben los dos principios antes mencionados:
  - Temporal: en cada ciclo se consulta el valor de i
  - Espacial: cada asignación A[i]:=0 almacena un 0 en un elemento del arreglo (el siguiente).



# Memoria Caché

15

- Basándose en los principios de Localidad temporal y espacial, la idea de intercalar una memoria (Caché) muy rápida, entre la CPU y la Memoria Principal, que contenga las posiciones de memoria a ser usadas próximamente, puede producir resultados muy buenos, dependiendo de la cantidad de datos accedidos en la Caché.
- Por los principios de Localidad temporal y espacial cuando se hace referencia a una palabra, es altamente probable que se acceda en el corto tiempo a las vecinas. Copiando en la memoria Caché la posición de memoria a acceder y sus vecinas, los próximos accesos probablemente se hagan mucho más rápido que si se hicieran a la Memoria principal (dependiendo de la tasa de aciertos).



# Memoria Caché

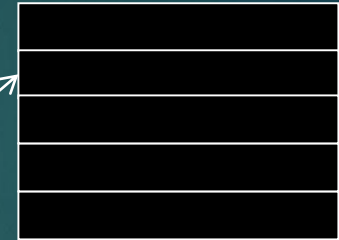
16

- Para facilitar las transferencias entre Memoria Principal y Caché, ambas memorias se organizan en bloques de palabras (de igual tamaño).

Memoria Principal



Memoria Cache



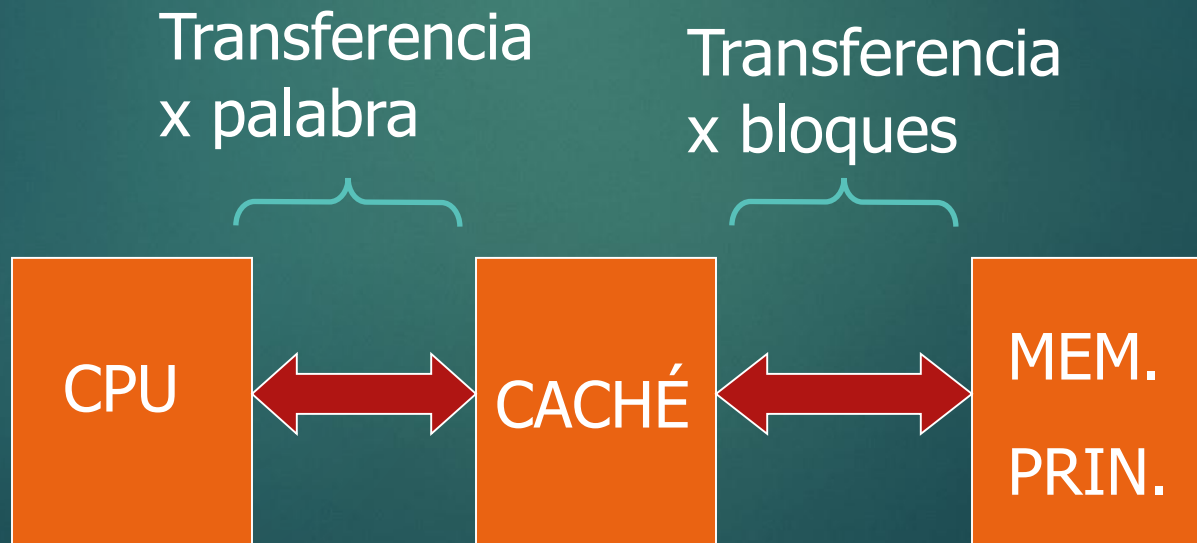
Bloque



# Memoria Caché

17

- Entre la Memoria Principal y la Caché, las transferencias son por bloque de datos, es decir, la mínima transferencia entre ambas memorias es un bloque.
- Entre la CPU y la Caché la transferencia es por palabra (porque así opera la CPU).



# Memoria Caché

18

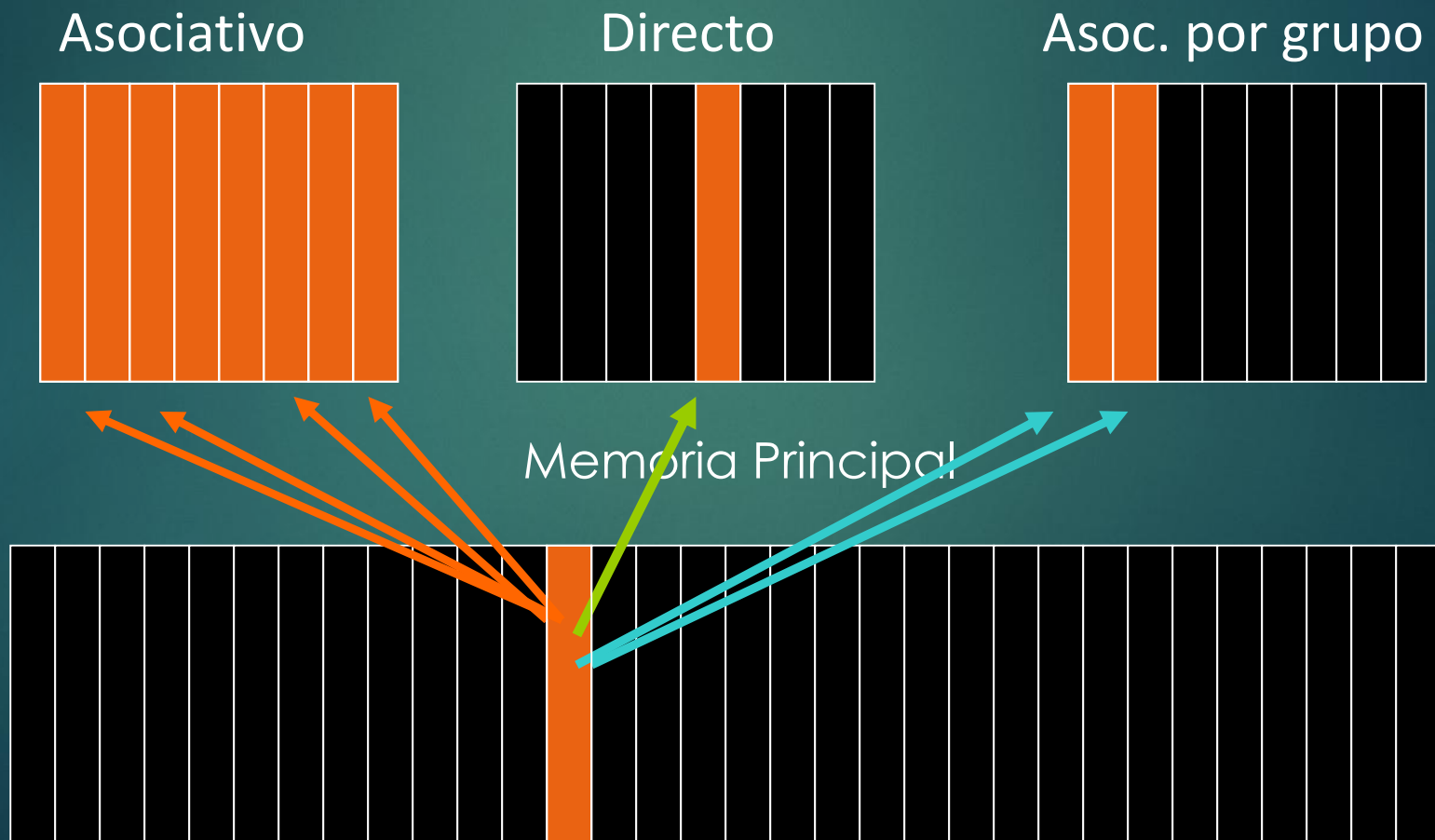
- El proceso de asignación de bloques de la Memoria Principal en la Caché se llama Función de Mapeo.
- Existen 3 técnicas básicas de mapeo:
  - Asociativa
  - Directa
  - Asociativa por grupos

# Memoria Caché

## Funciones de mapeo

19

- Las 3 funciones de mapeo están representadas en la siguiente figura.



# Memoria Caché

## Aciertos y fallos

20

- Cuando la CPU busca un dato en la Caché lo puede encontrar o no.
- Cuando la CPU encuentra el dato en la Caché ocurre un Acierto (o Hit), y la CPU obtiene el dato a alta velocidad.
- Cuando la CPU no encuentra el dato en la Caché ocurre un Fallo (o Miss), y la CPU tiene que obtenerlo de la memoria principal, a una velocidad menor.

# Memoria Caché

## Aciertos y fallos

21

- Así, la eficiencia de la Caché depende de la cantidad de veces que la CPU encuentra el dato en la Caché.
- La eficiencia de la Caché se expresa a través de la frecuencia de aciertos, es decir el número de veces que la CPU “acierta” a la Caché.

# Memoria Caché

## Niveles de caché

22

- A veces la Caché está organizada en varios niveles.
- Comúnmente identificados como L1, L2, L3, etc.
- Cuanto más alto es el nivel, mas pequeña es la memoria.
- Considerando que típicamente el porcentaje de aciertos de una memoria caché es del orden del 90% (es decir, la tasa de fallos es del 10%), organizando la memoria caché en múltiples niveles, lo que se trata es de mejorar la eficiencia del 90% de aciertos.



# Memoria secundaria

23

- Consideramos como memoria externa los dispositivos de almacenamiento que no están conectados físicamente a bus del sistema.
- Comprende dispositivos tales como:
  - Discos magnéticos
  - Discos ópticos
  - CD-ROM
  - CD-R
  - CD-RW
  - DVD

# Memoria secundaria

## Discos magnéticos

24

### Características físicas

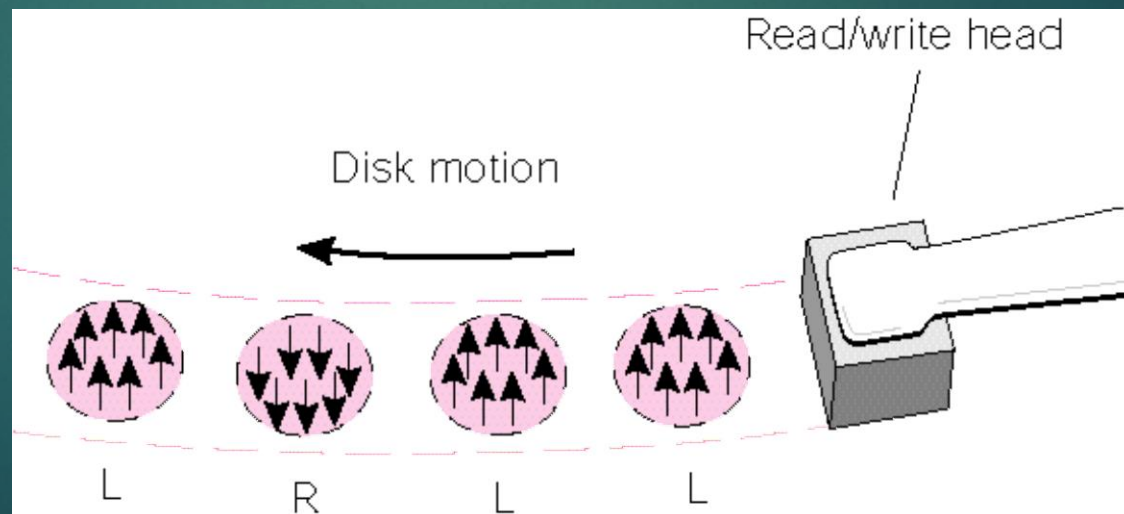
- Los discos magnéticos son dispositivos electromecánicos compuestos por 1 o más platos rígidos.
- Los platos son típicamente de aluminio o vidrio (que tiene menor coeficiente de dilatación que el aluminio)
- Los platos están recubiertos con una capa de material magnético, por ejemplo óxido de Fe.
- La superficie tiene que ser sumamente uniforme, sin defectos superficiales.

# Memoria secundaria

## Discos magnéticos

### Características físicas

- En el disco se generan pequeñas áreas que pueden ser magnetizadas en 2 direcciones por un transductor (cabeza lectora/grabadora).
- El disco está girando, de tal manera que hay un movimiento relativo entre la superficie magnetizable y el transductor, al momento de la lectura o la escritura.



# Memoria secundaria

## Discos magnéticos

26

### Características físicas

- La Lectura y escritura es a través de una cabeza transductora.
- Durante la lectura o escritura, la cabeza está estática y el plato gira.
- Los puntos magnetizados en 2 direcciones determinan 2 estados, que se los puede asociar a 1 y 0.
- El resultado, en definitiva, es el almacenamiento en círculos concéntricos de unos y ceros.

# Memoria secundaria

## Discos magnéticos

27

### Características físicas

- El movimiento relativo del disco con una posición fija de la cabeza lectora/grabadora determina un anillo o pista.
- La cabeza lectora/grabadora se mueve radialmente en pasos discretos. Para cada posición de la cabeza se tiene una pista o track.
- La cantidad de pasos discretos que puede dar la cabeza sobre la superficie del disco define la cantidad de pistas o tracks disponibles en el disco.

# Memoria secundaria

## Discos magnéticos

28

### Características físicas

- Las pistas tienen el mismo centro, es decir, son concéntricas.
- Las pistas están separadas por espacios vacíos denominados “gaps”.
- Si se reducen los gaps aumenta la capacidad del disco (porque aumenta la densidad de pistas).
- Se considera que el número de bits por pista es el igual para todas las pistas.

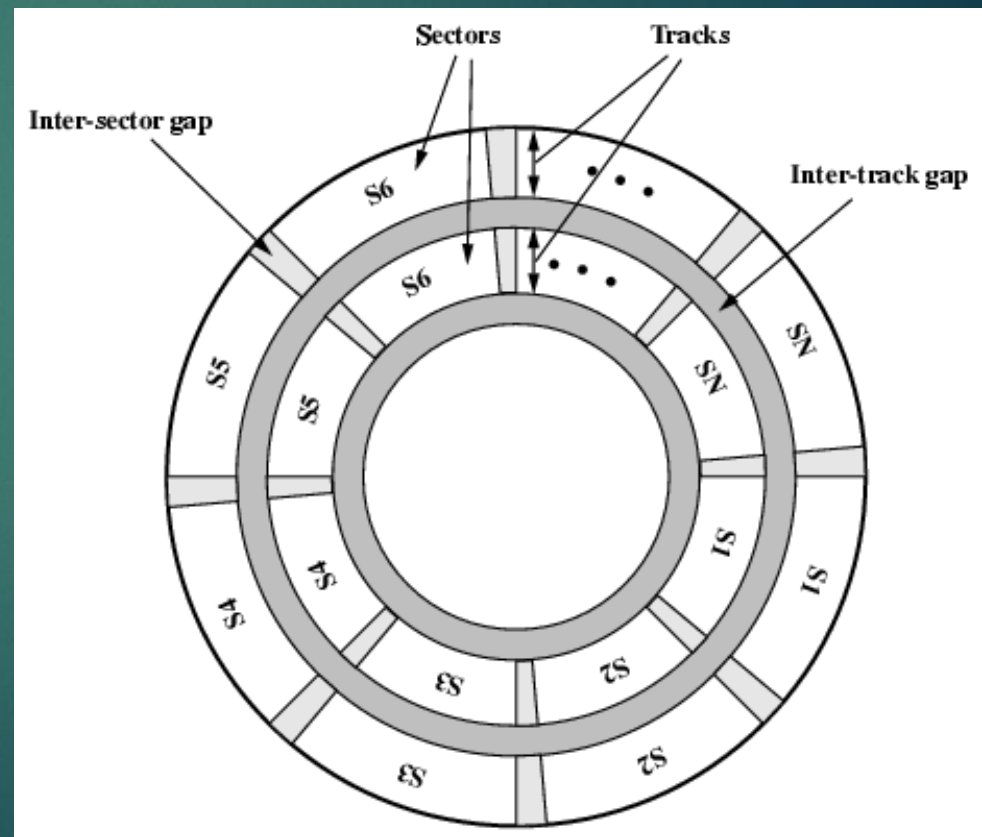


# Memoria secundaria

## Discos magnéticos

29

- Las pistas se dividen en sectores.
- Un sector es la mínima unidad de transferencia del disco.
- Cada pista tiene un número entero de sectores





# Memoria secundaria

## Discos magnéticos

### Tipos de discos magnéticos

- Cabeza fija (raro) o móvil.
- Disco removible o fijo.
- Simple ó doble lado.
- Uno ó múltiples platos.

### Tipos de cabezales

- De contacto (Floppy)
- De distancia de separación fija
- De separación aerodinámica (conocido como tipo Winchester)

# Memoria secundaria

## Discos magnéticos

### Características típicas de discos magnéticos

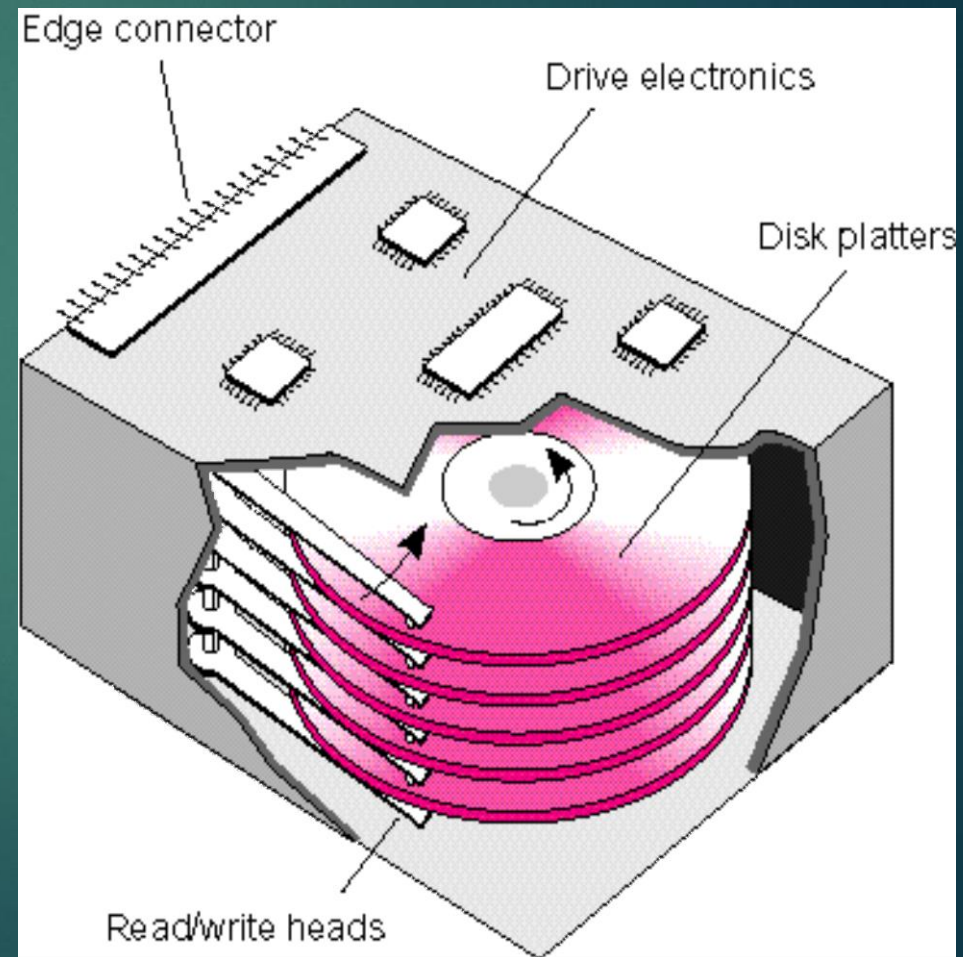
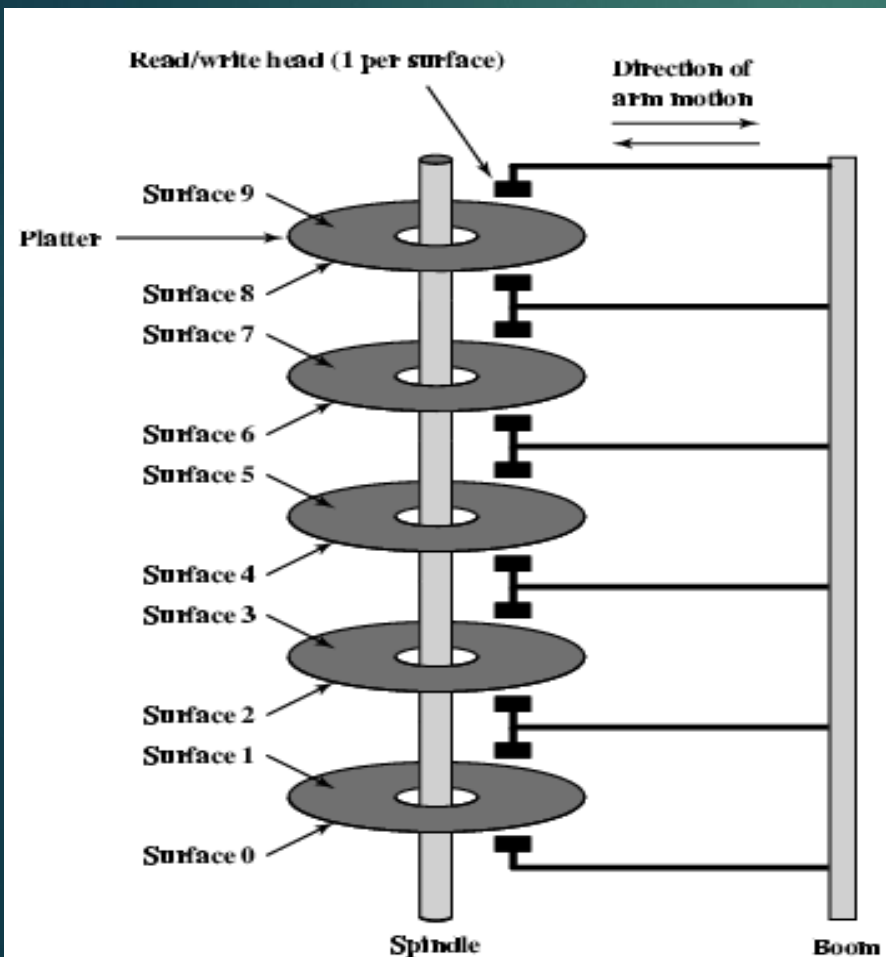
- Los discos típicamente disponen de múltiples platos.
- Una cabeza lectora/grabadora por cara o superficie.
- Todas las cabezas se mueven solidariamente.
- Las pistas alineadas de cada plato forman una estructura lógica denominada “cilindro”.
- En general, los datos son almacenados por cilindros, para reducir los movimientos de las cabezas lectora/grabadora.
- Esta operación, además, permite aumentar la velocidad de respuesta.

# Memoria secundaria

## Discos magnéticos

32

Esquema típico de disco magnético:

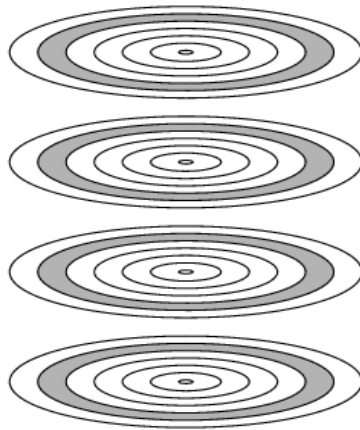


# Memoria secundaria

## Discos magnéticos

33

Cilindros compuestos por pistas alineadas en diferentes platos



**Figure 6.6 Tracks and Cylinders**

# Formato de grabación en discos

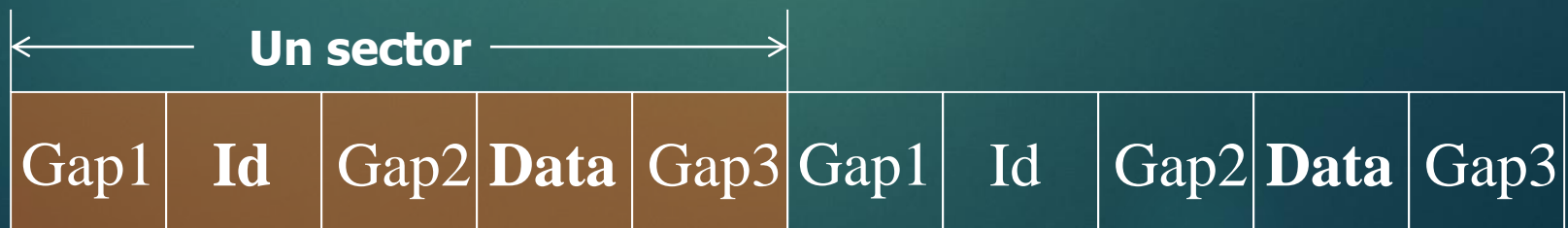
34

Formato de grabación de discos magnéticos estándar ST 506:

Cada sector tiene 5 campos:

- Gap1: separación entre sectores – inicio de sector
- Id: Identificación del sector
- Gap2: separación entre campo Id y campo de dato
- Data: campo de datos
- Gap3: separación entre sectores – fin de sector

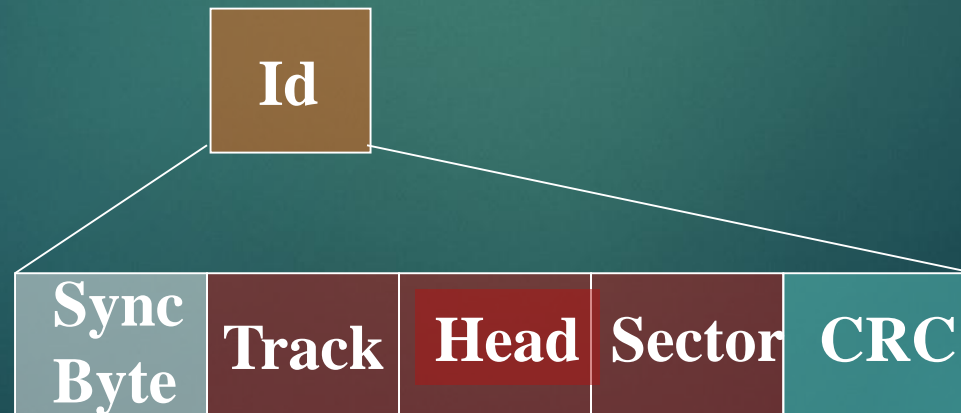
Los Gap son espacios de separación que no contienen información útil.



# Formato de grabación en discos

35

- El campo ID está compuesto por 5 campos:
  - SyncByte: patrón de bits para sincronizar la operación
  - Track: número de track o pista
  - Head: numero de cabeza lectora/grabadora
  - Sector: número de sector dentro de la pista
  - CRC: bits para detección y corrección de errores





# Formato de grabación en discos

36

- El campo Data está compuesto por 3 campos:
- Encabezado, con información para sincronizar la lectura e identificar el sector.
- Datos, de una determinada longitud en bytes, normalmente expresada como potencia de 2.
- Código para errores (CRC: Cyclic Redundancy Check), con información para detectar y/o corregir posibles errores.





# Formato de grabación en discos

37

## Capacidad de almacenamiento de un disco

- La capacidad total de un disco se determina de la siguiente manera:

$$\text{Capacidad} = \frac{\text{bytes}}{\text{sector}} \times \frac{\text{sectores}}{\text{pista}} \times \frac{\text{pistas}}{\text{superficie}} \times \# \text{ de superficies}$$

- Debido a las restricciones del formato, se desperdicia espacio en las pistas más externas.
- En los formatos de grabación posteriores, para aumentar la capacidad del disco se particiona el área en zonas. Cada zona tiene una cantidad fija de bits/pista, pero se requieren circuitos electrónicos más complejos debido a que la densidad de grabación varía con la zona.

# Formato de grabación en discos

38

## Tiempo de acceso de un disco

- El tiempo de acceso a un sector en el disco está compuesto por 2 componentes:
  - Tiempo de búsqueda (seek): es el tiempo que tarda la cabeza lectora/grabadora en llegar al cilindro o pista buscado.
  - Tiempo de latencia (por rotación): es el tiempo que hay que esperar hasta que el sector “pase” por debajo de la cabeza lectora/grabadora.

$$\text{Tiempo de Acceso} = T.\text{seek} + T.\text{latencia}$$

Una vez encontrado el sector, la transferencia se hace a una determinada velocidad. El tiempo total es:

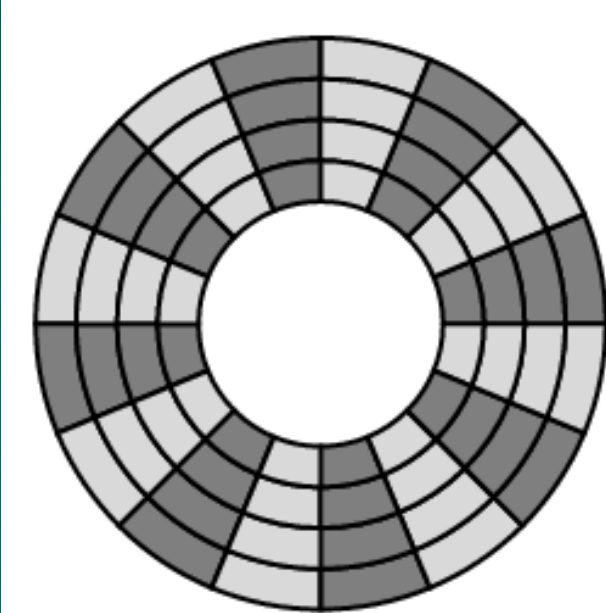
$$T_{\text{Total}} = T.\text{de Acceso} + T.\text{de Transferencia de datos}$$

# Formato de grabación en discos

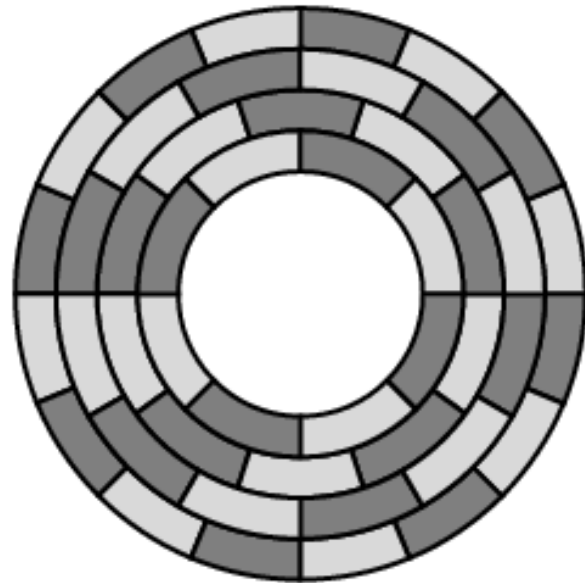
39

## Formatos de grabación fijo y variable

La velocidad de rotación angular (velocidad a la que gira el disco) puede ser fija o variable.



**Grabación en CAV**  
**(velocidad constante)**



**Grabación en VAV**  
**(velocidad variable)**

# Formato de grabación en discos

40

## Velocidad de rotación en discos – formato CAV

- La velocidad de rotación determina el tiempo de latencia y el tiempo que dura la transferencia.
- Si rota a velocidad angular constante - CAV (rpm)
  - Como la velocidad lineal (velocidad tangencial) es proporcional al radio, los bits de las pistas más cercanas al centro giran más lentamente que los bits de las pistas más alejadas.
  - Es decir, los bits y por lo tanto los sectores, ocupan distinto espacio en las diferentes pistas.

# Formato de grabación en discos

41

## Velocidad de rotación en discos – formato VAV

- Si rota a velocidad angular variable - VAV (rpm)
  - Se puede hacer que la velocidad lineal sea la misma para todos los bits.
  - Por lo tanto, los bits y los sectores ocupan el mismo tamaño en las diferentes pistas.

# Formato de grabación en discos

42

## Tipos de formatos

- El formato de grabación de un disco define:
  - Cantidad de sectores
  - Tamaño del sector
  - Funcionalidades de los distintos campos en cada pista y sector.
- Existen 2 tipos de formatos.
  - Por hardware: se usan marcas físicas para definir el tamaño de cada sector. Este era un método antiguo.
  - Por software: el tamaño del sector está determinado por el Sistema operativo.



# Referencias

- Capítulo 5: Memoria Externa
  - Stallings. 5ta Ed.
- Links de interés
  - ▶ <http://www.pctechguide.com/02Storage.htm>