

# Almacenamiento y discos

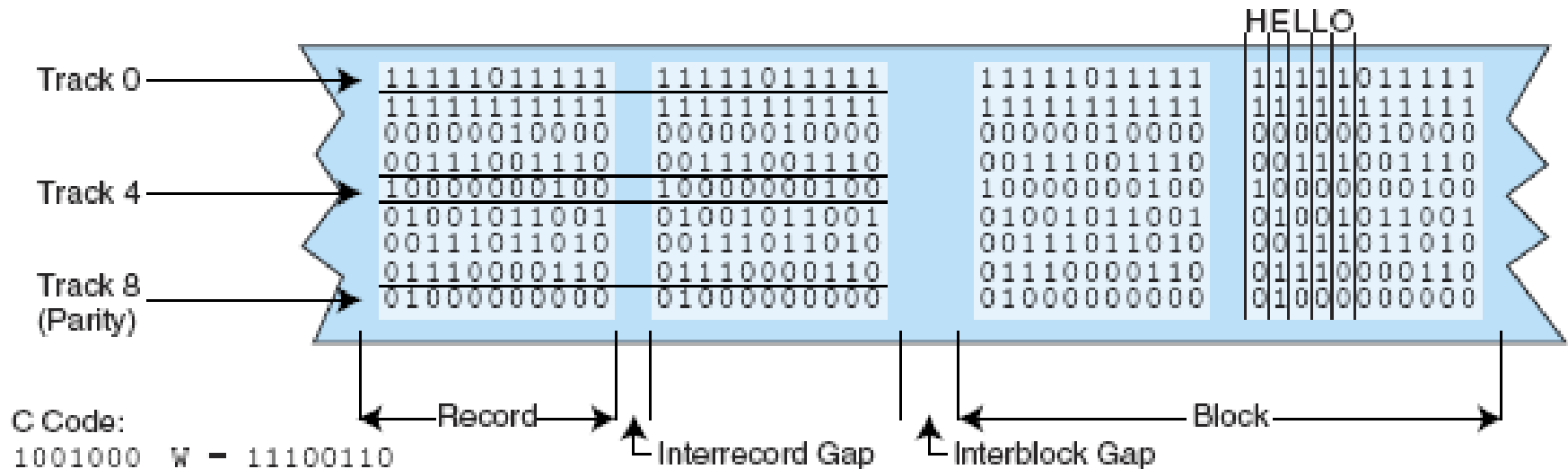
# Almacenamiento y discos

1. Motivación
2. Cintas magnéticas
3. Discos magnéticos
4. Discos ópticos
5. RAID

# Motivación

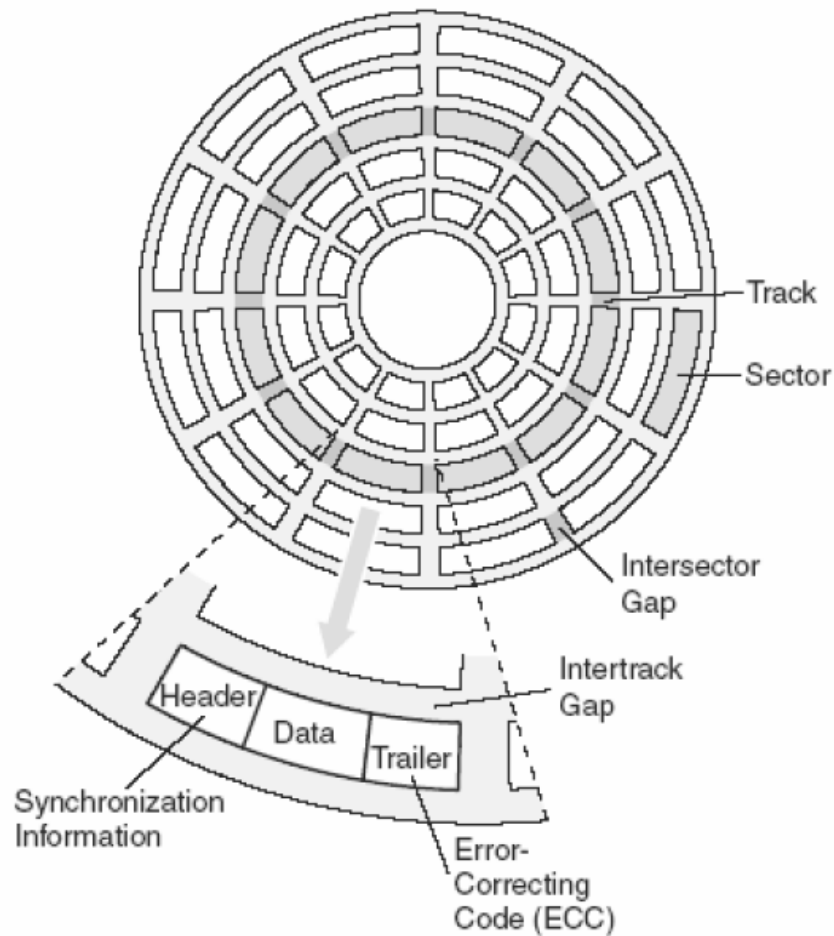
- Memoria volátil se pierde al apagarse la computadora
- Primeros medios de almacenamiento perdurables
  - Tarjetas perforadas
  - Cintas magnéticas
  - Cintas de papel

# Cintas Magnéticas



- 11 MB de capacidad
- 1/2 hora para leer/escribir la cinta completa
- Interblock gap: distancia para frenar/arrancar la cinta
- Interrecord gap: distancia para procesar el registro
- Se tiene que transferir un bloque ENTERO

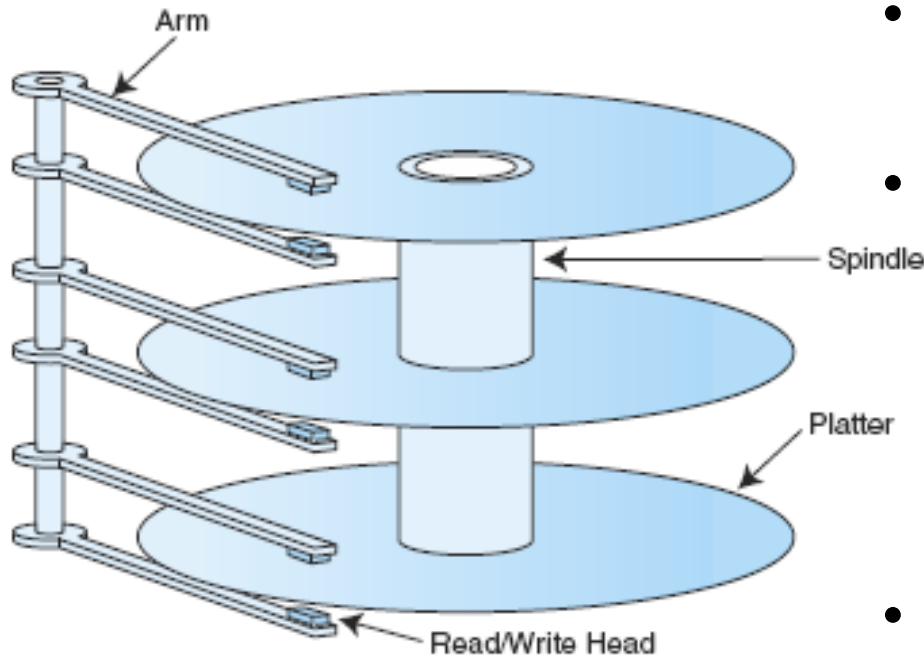
# Disco magnético



# Discos Magnéticos

- Físicamente no se “graban” bits
  - Métodos de grabación:
    - FM (Frequency Modulation),
    - RLL (Run Length Limited),
    - RLL 2.7, etc.
- La interfaz física del disco abstraee la codificación/decodificación de las cadenas de bits al medio magnético

# Discos Rígidos



- Hasta 15000 rpm, media entre 5400 rpm y 7200 rpm
- El almacenamiento magnético no está “libre” de errores
  - Codificación de los datos
  - Algoritmos de recuperación de errores
- Todo esto es realizado por la controladora del disco

# Disco rígido

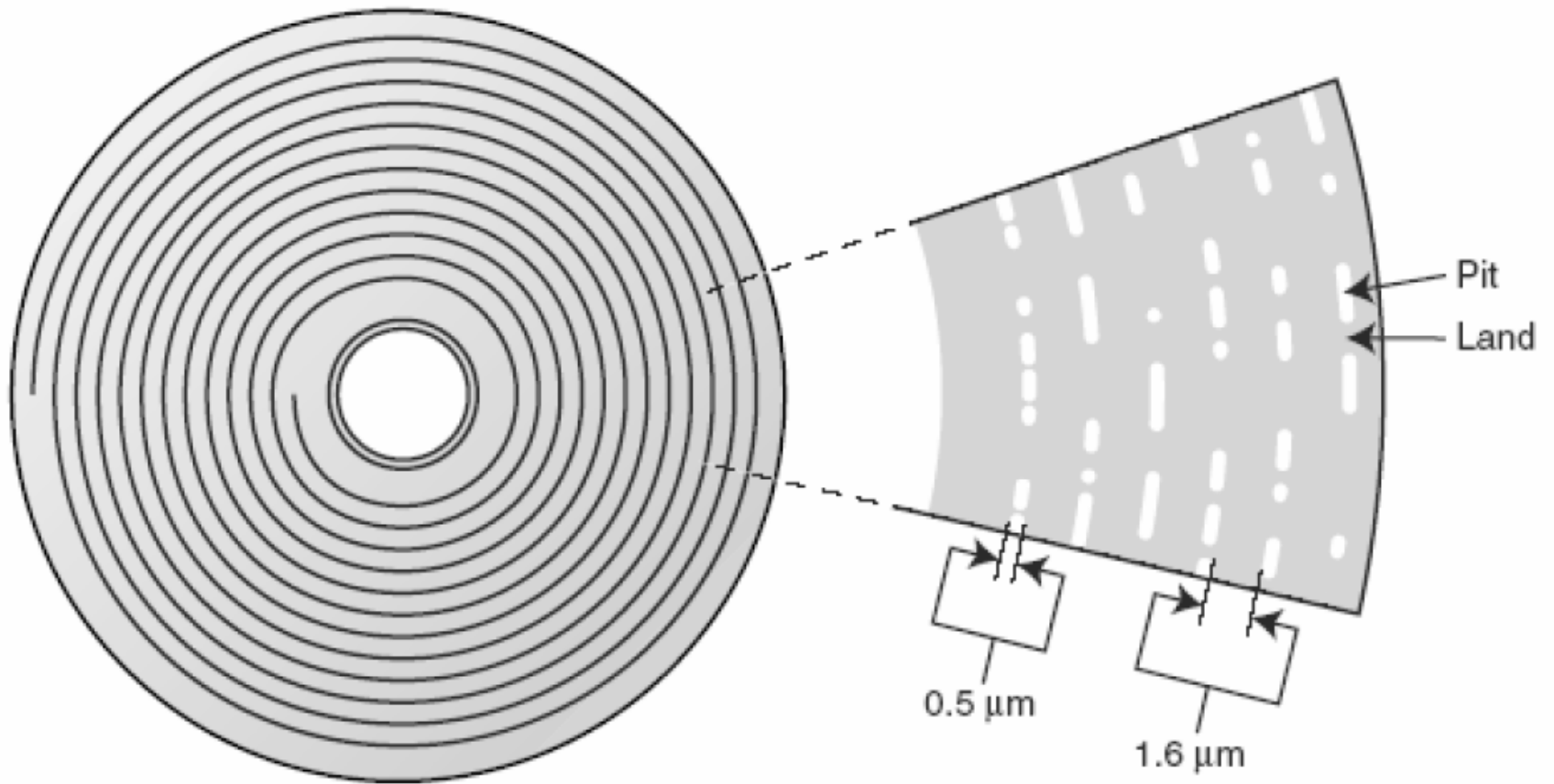
- Cantidad de cabezas (platos x 2)
- Cantidad de cilindros
  - Pistas por cara
- Cantidad de pistas
- Cantidad de sectores por pista



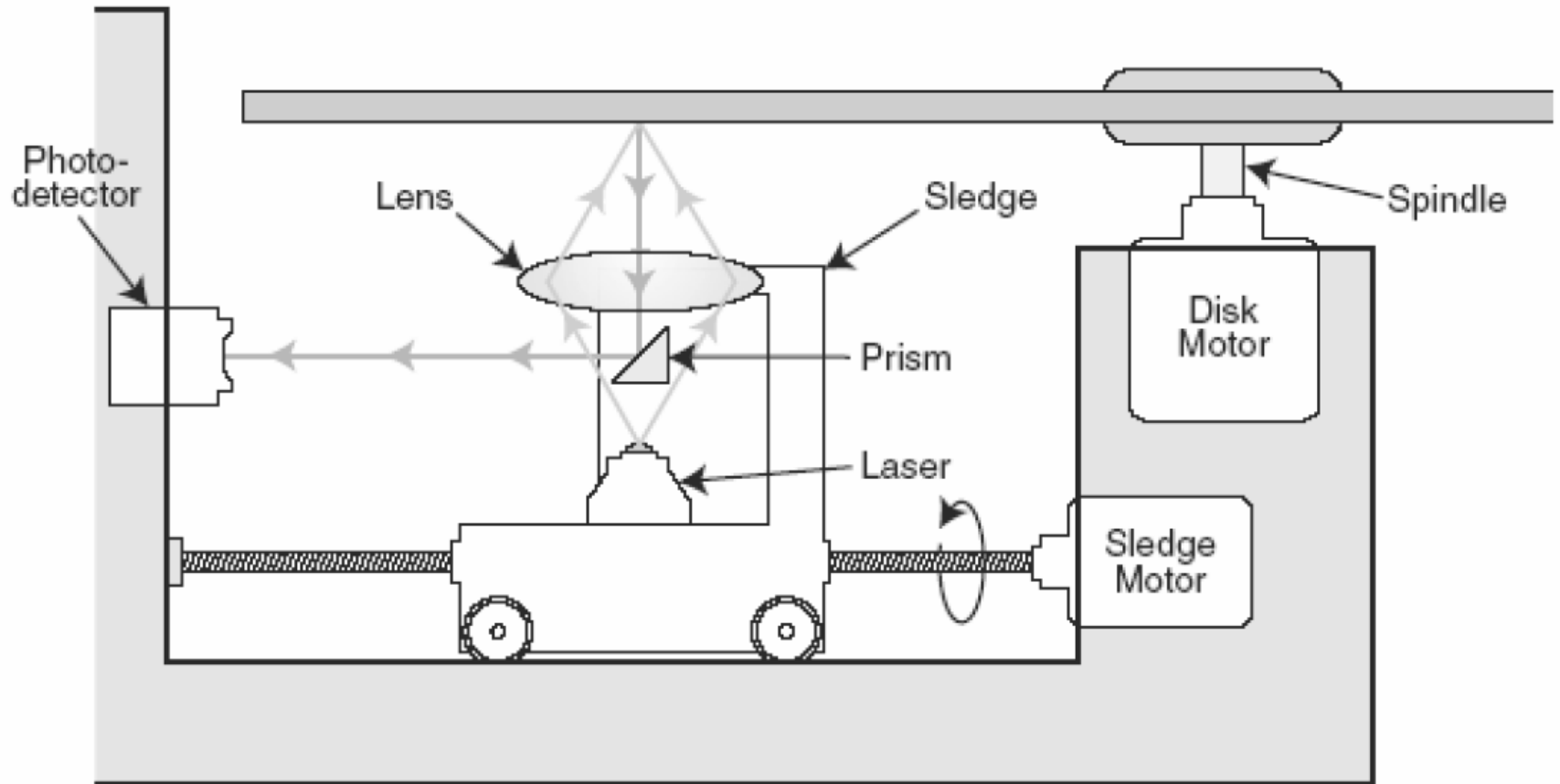
# Disco rígido

- Tiempo entre pistas adyacentes
- Tiempo promedio entre pistas
- Tiempo de latencia media
- Velocidad de rotación
- Velocidad de transferencia de la interfaz

# Discos ópticos



# Lectura disco óptico



# Discos ópticos

- El motor gira el disco más rápidamente para lograr misma velocidad de lectura
- Un sector es identificado por el número de minutos y segundos de pista que hay entre el y el comienzo del disco.
  - Se calibra bajo la asunción que el reproductor procesa 75 sectores por segundo (CD 1X)
- Para localizar un sector, la cabeza realiza una búsqueda heurística

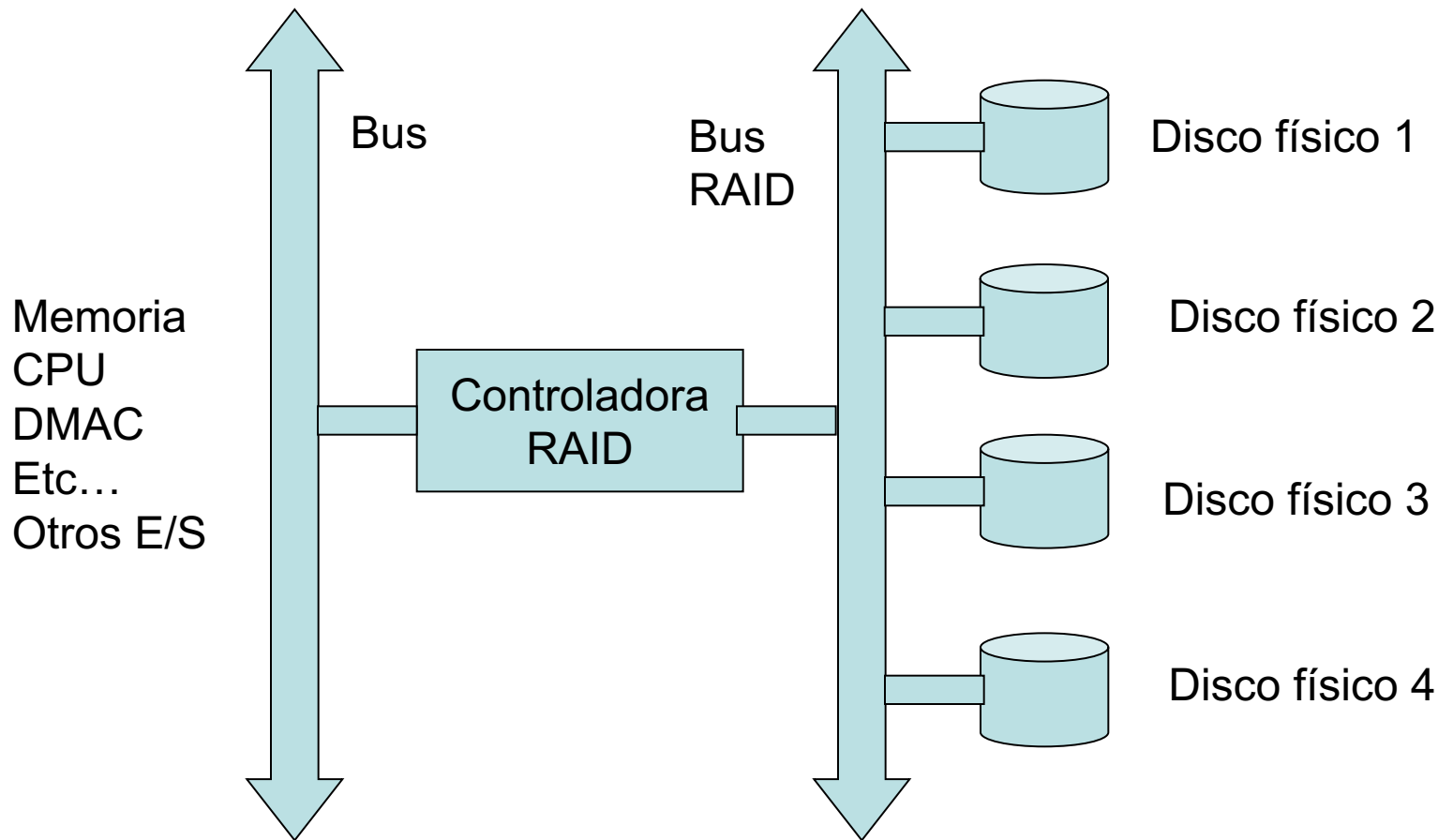
# RAID

- Redundant Array of Independent Disks
- Conjunto Redundante de Discos Baratos
- David Patterson et al., Berkeley 1998
  - Performance (velocidad de acceso y transferencia)
  - Confiabilidad (recuperación de errores, redundancia)

# RAID – Distribución de datos

- La distribución de datos puede ser gestionada por:
  - Software: el hardware es el que se encuentra disponible en el sistema (la controladora de discos proveída generalmente en el chipset)
  - Hardware: se dispone de una controladora especializada para RAID

# Controladora RAID



# RAID – Tipos (Levels)

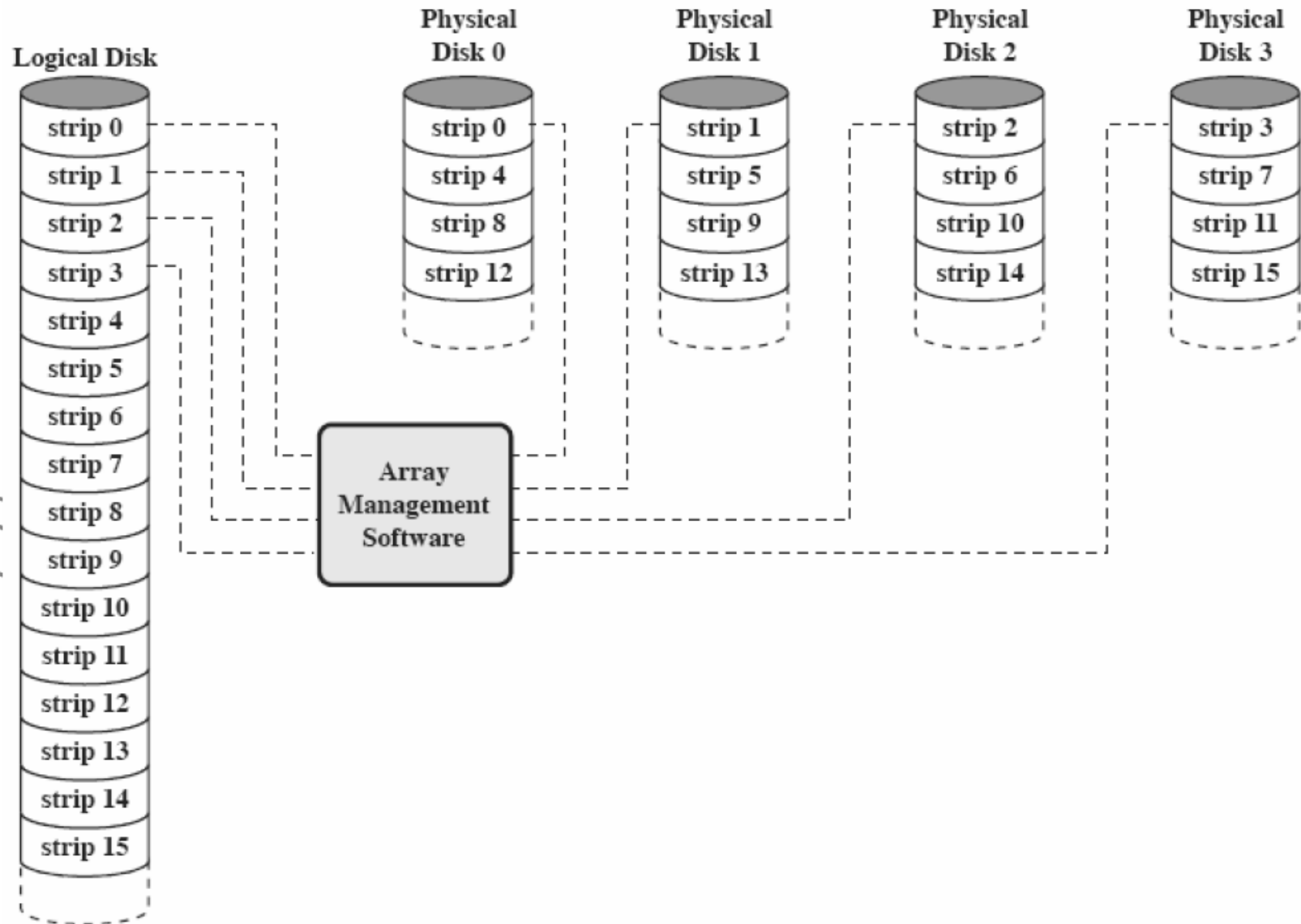
- Level 0: une varios discos como si fueran uno solo, sin tolerancia a fallos.
- Level 1: mirroring.
- Level 2: con código de corrección de errores (código de Hamming).
- Level 3, 4 y 5: con bits de paridad.
- Level 0 + 1: varios discos con mirroring



# RAID

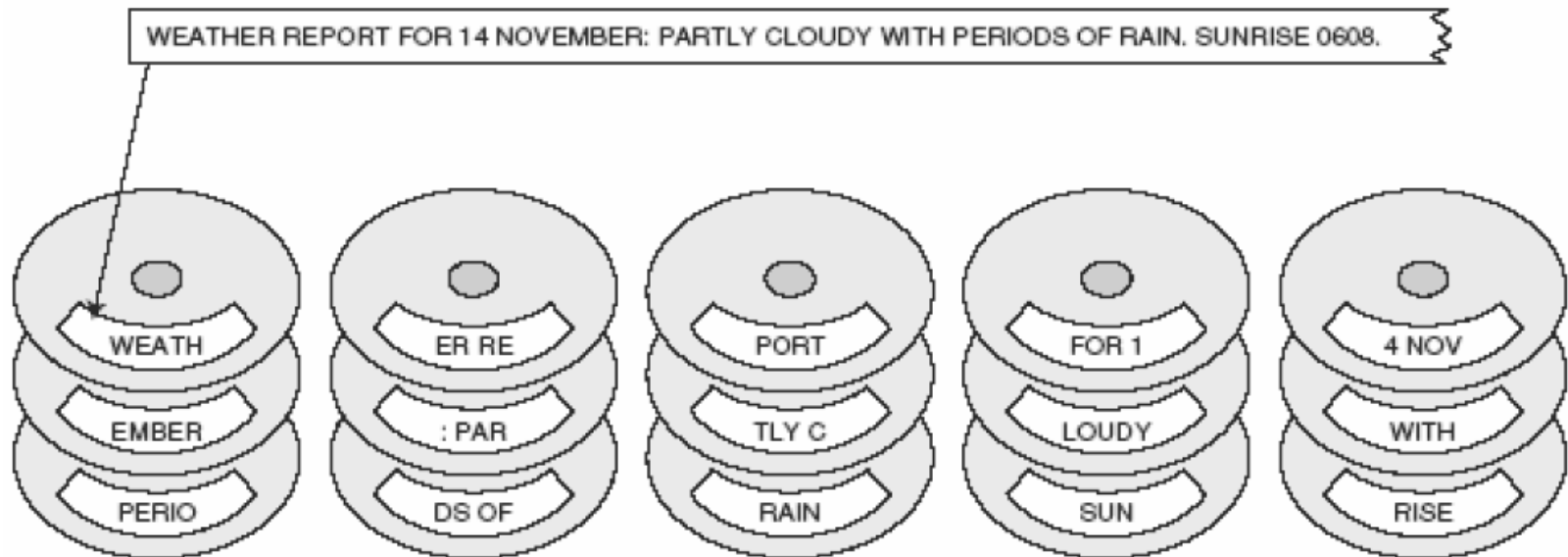
- Striping: separación de secuencias de datos, con el objeto de escribirlos en múltiples dispositivos físicos:
  - Tan pequeño como 1 bit
  - Tan grande como bloques de un tamaño específico

# RAID 0



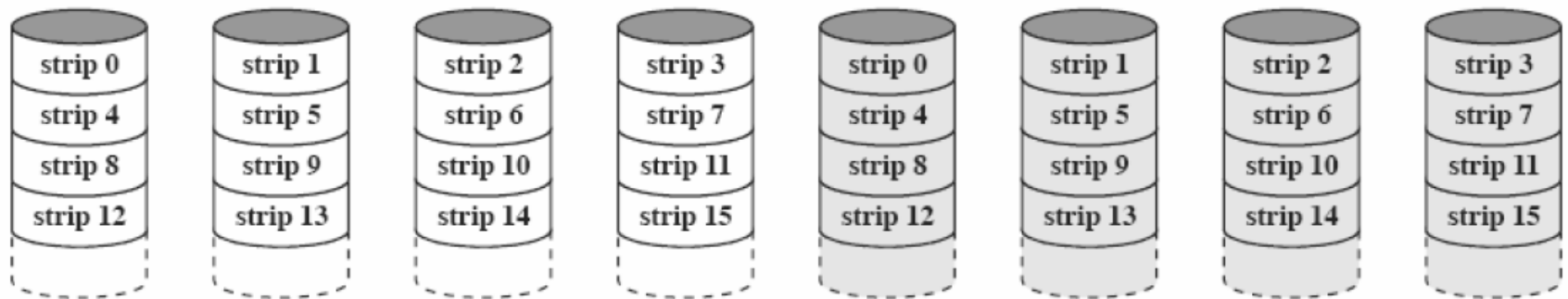
# RAID 0

- Mayor performance (si existen controladoras y cachés por cada disco físico)
- Menor confiabilidad (no hay redundancia)



# RAID 1

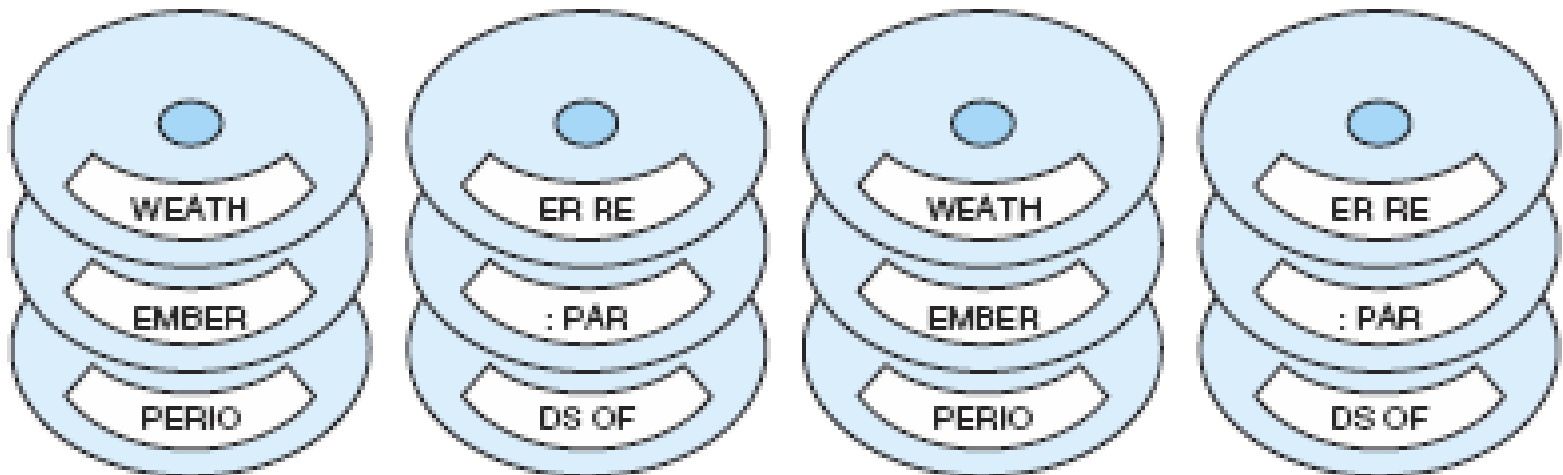
- RAID con mirroring



(b) RAID 1 (mirrored)

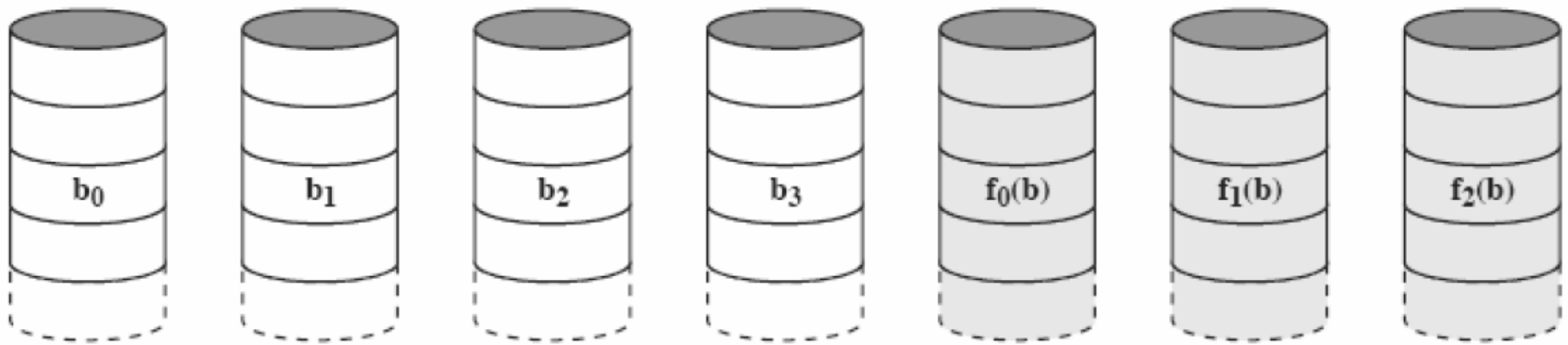
# RAID 1

- Cada dato escrito es duplicado físicamente
- Escrituras más lentas
- Lecturas más rápidas
- Necesito doble de espacio físico



# RAID 2

- Redundancia usando códigos de Hamming



(c) RAID 2 (redundancy through Hamming code)

# RAID 2

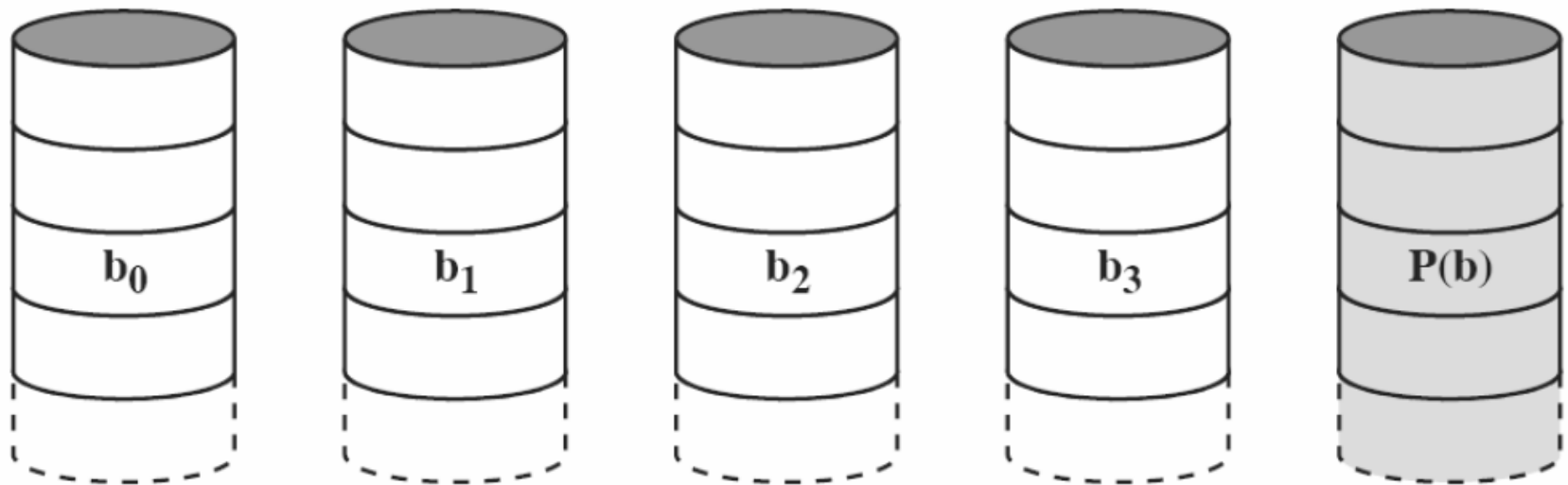
- Stripping = 1 bit (al menos 8 discos físicos para acomodar 1 byte)
- Discos adicionales son usados para guardar información para corrección de errores (código de Hamming)
- Hardware especial para computar el código de Hamming

# Código de Hamming

- Adaptación del concepto de bit de paridad
  - 0110 1100    Paridad 0
  - 1100 1110    Paridad 1
- Distancia Mínima de Hamming ( $d_{\min}$ )
  - Detecta hasta  $(d_{\min} - 1)$  bits errores
  - Corrige hasta  $(d_{\min} - 1) / 2$  errores

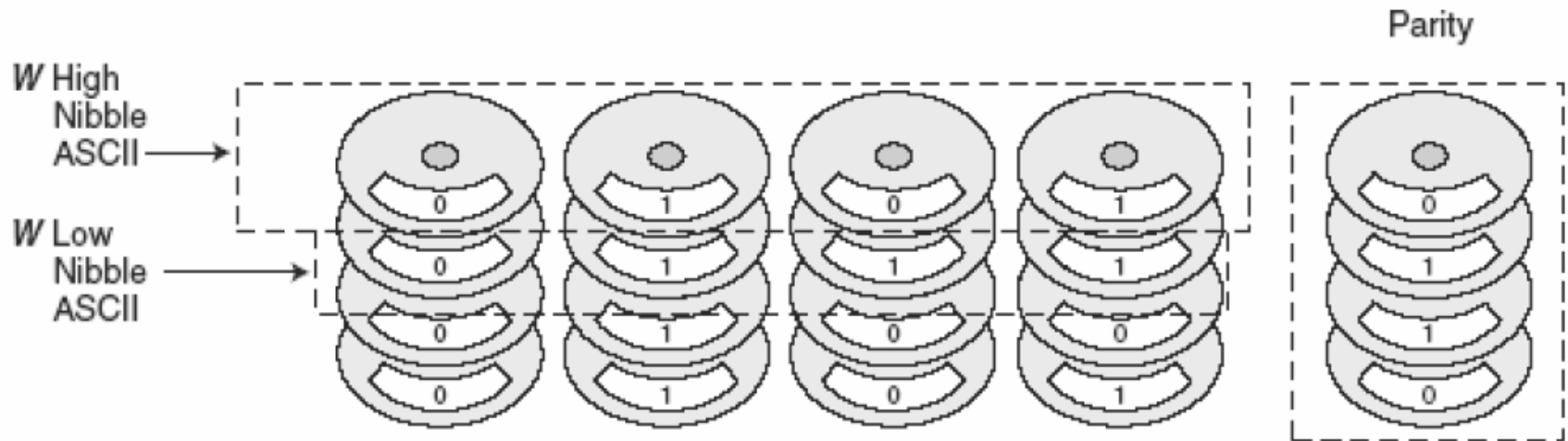


# RAID 3



**(a) RAID 3 (bit-interleaved parity)**

# RAID 3

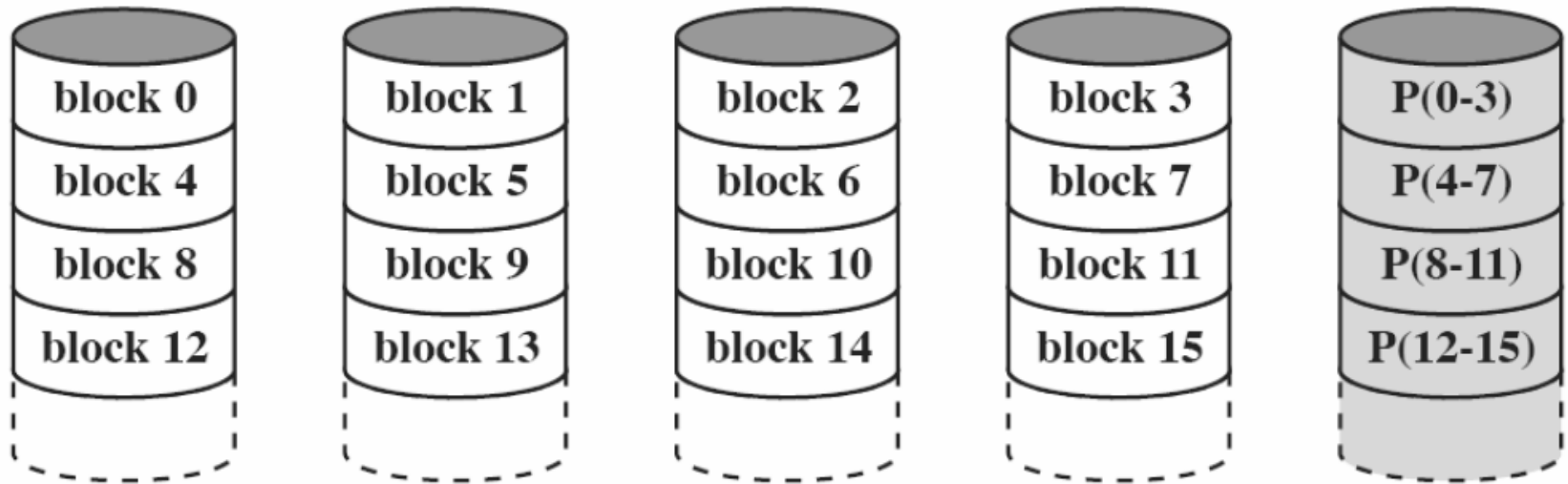


Letter	ASCII	Parity (even)	
		High Nibble	Low Nibble
W	0101 0111	0	1
E	0100 0101	1	0
A	0100 0001	1	1
T	0101 0100	0	1
H	0100 1000	1	1
E	0100 0101	1	0
R	0101 0010	0	1

# RAID 3

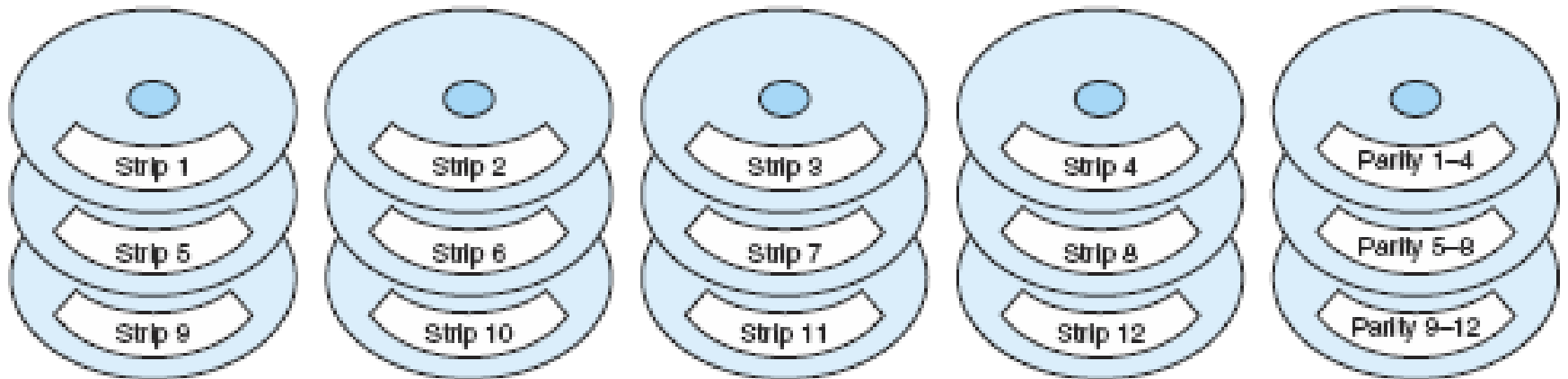
- Cómputo de paridad
  - $\text{Parity} = b_0 \text{ xor } b_1 \text{ xor } \dots \text{ xor } b_7$
- Si un disco falla (ejemplo el 6) se puede reconstruir su contenido a partir de los restantes y el disco de paridad
  - $b_6 = b_0 \text{ xor } \dots \text{ xor } b_7 \text{ xor Parity}$
- Cuello de botella: accesos al disco de paridad

# RAID 4



(b) RAID 4 (block-level parity)

# RAID 4

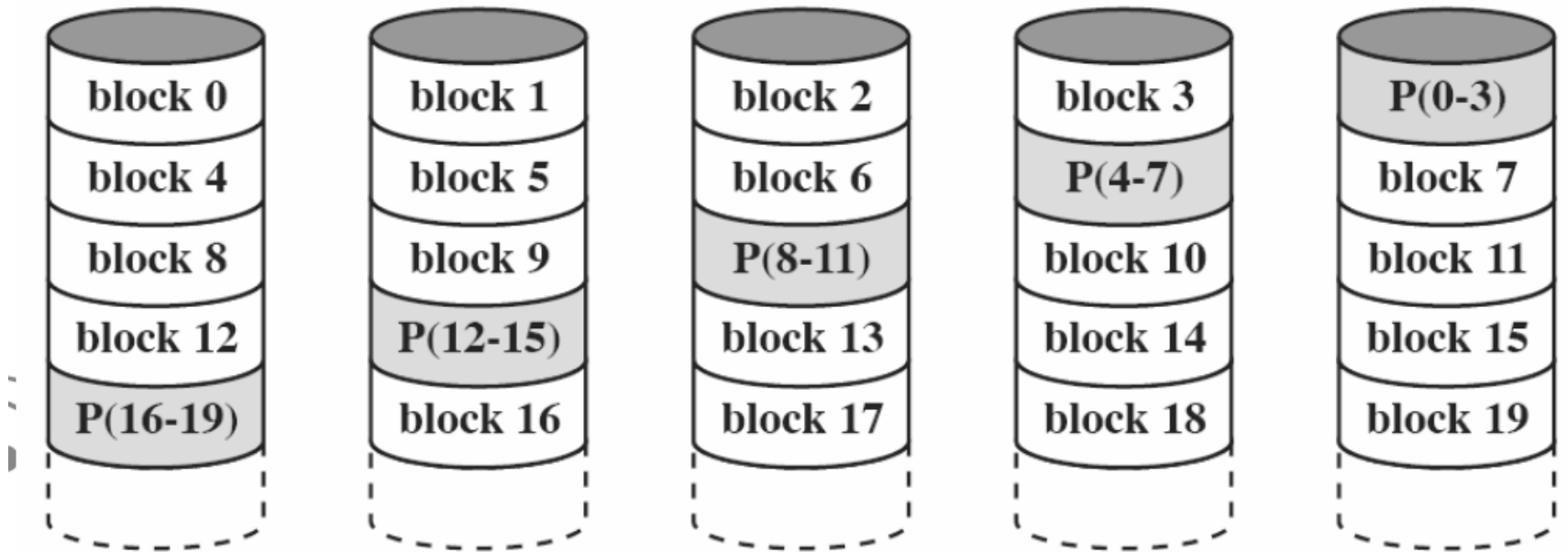


$\text{PARITY 1 - 4} = (\text{Strip 1}) \text{ XOR } (\text{Strip 2}) \text{ XOR } (\text{Strip 3}) \text{ XOR } (\text{Strip 4})$

# RAID 4

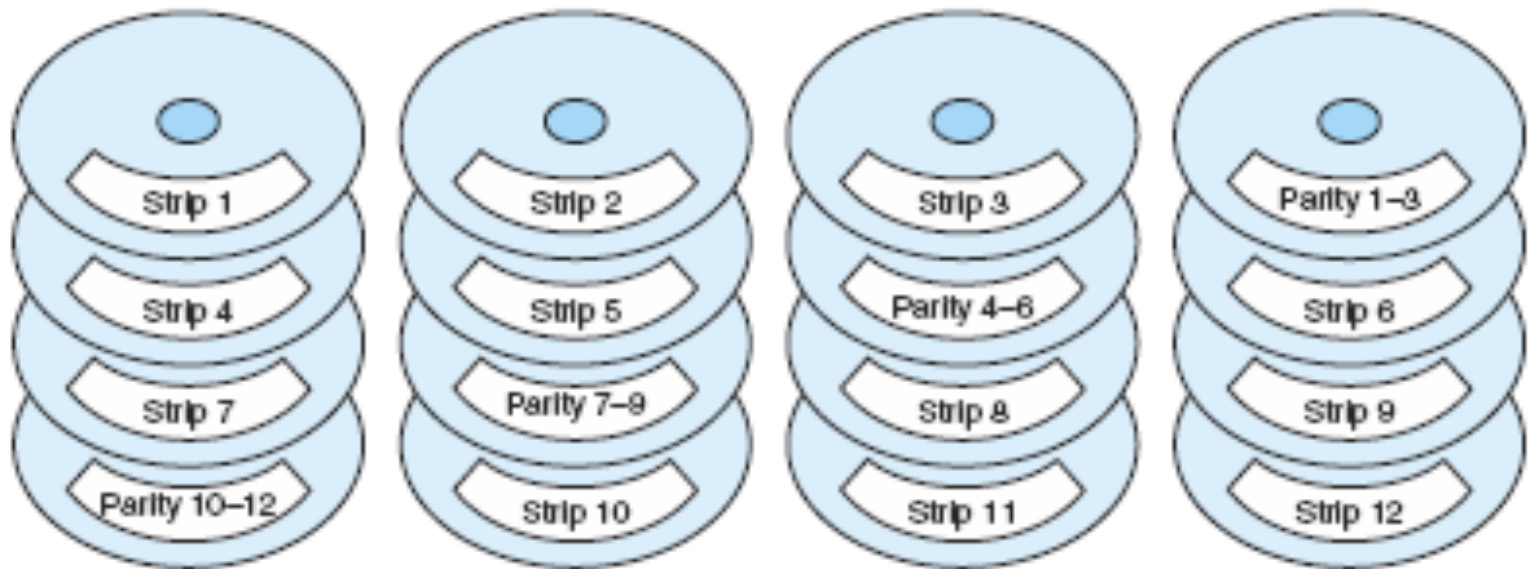
- Idem a RAID 3 pero con granularidad mayor a bit
- Cuello de botella: accesos al disco de paridad

# RAID 5



(c) RAID 5 (block-level distributed parity)

# RAID 5



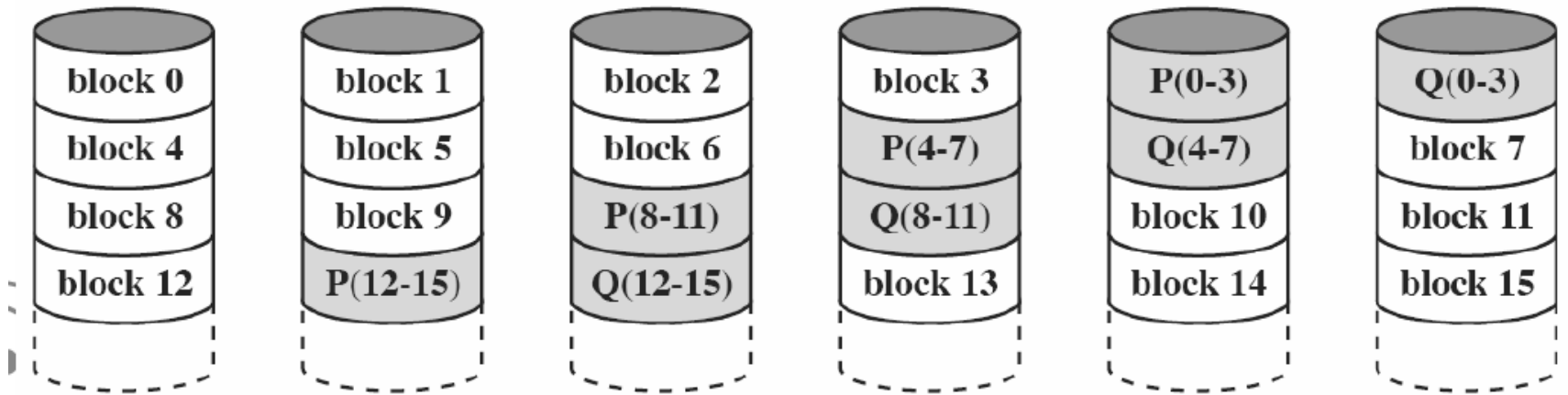
$\text{PARITY 1 - 3} = (\text{Strip 1}) \text{ XOR } (\text{Strip 2}) \text{ XOR } (\text{Strip 3})$



# RAID 5

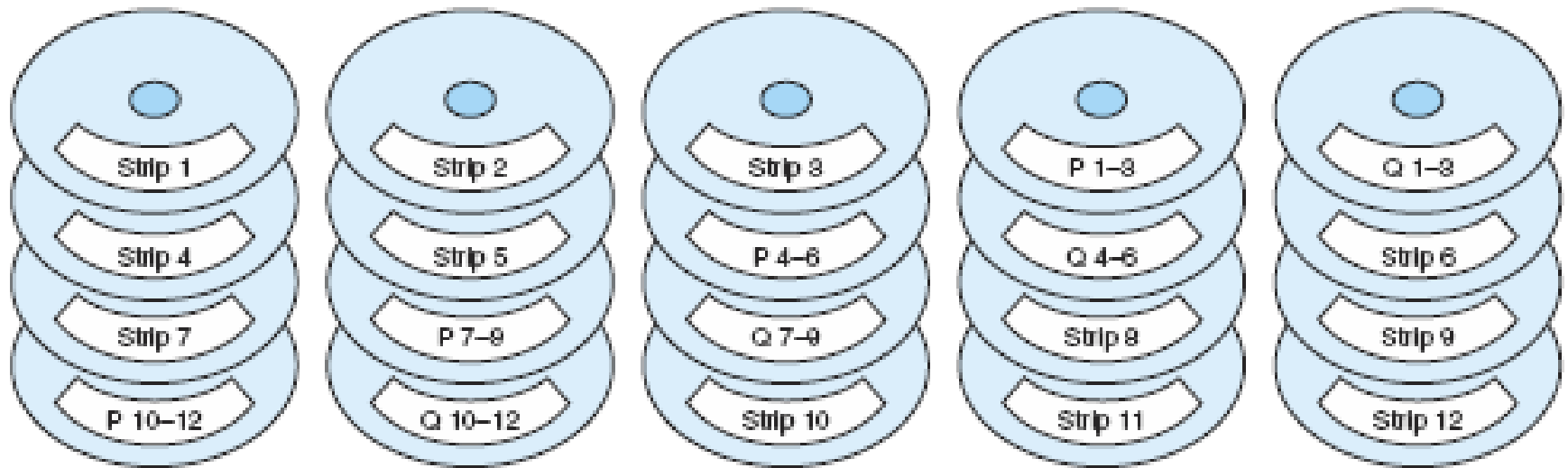
- Mejor throughput que los RAID 3 y 4
- Controladora más complicada, y por lo tanto, más cara
- Sigue soportando a lo sumo la falla de 1 disco

# RAID 6



(d) RAID 6 (dual redundancy)

# RAID 6



P = Parity  
Q = Reed-Soloman

# RAID 6

- Posee doble mecanismo de redundancia:  
Parity + Reed-Soloman
- Muy lenta la escritura (Reed-Soloman)
- Única implementación (hasta ahora)
  - RAMAC RVA 2 Turbo disk array