AAAC & TAC
Clase #6
Entrada/Salida

Facultad de Ingeniería
Universidad de la República

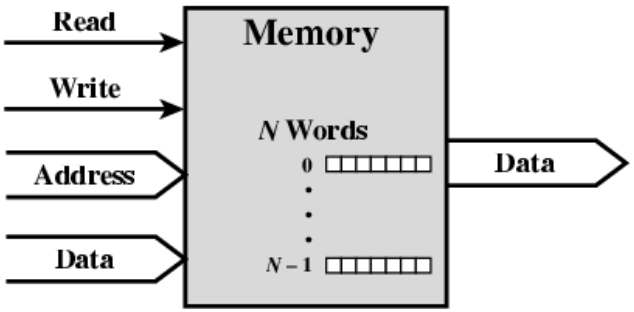
Instituto de Computación
Curso 2014

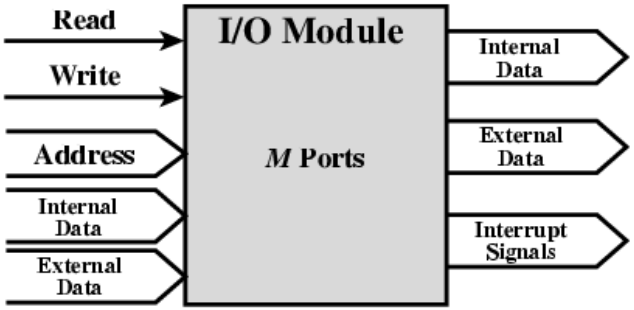
Veremos…

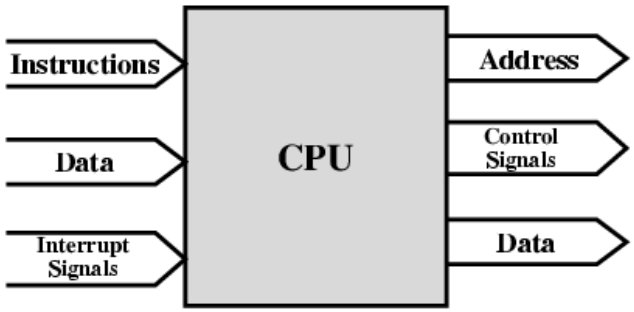
- Conexión de memoria y E/S
  - Buses del sistema
- Estructuras de Control de E/S
  - Programada (Polling)
  - Interrupciones
- E/S avanzada
  - DMA
  - Procesadores de E/S
- Periféricos
  - Interfaces de alta velocidad
  - Almacenamiento

Módulos del sistema

- Todas las unidades deben interconectarse
- El tipo de conexión varía según el módulo
  - Memoria
  - Entrada/Salida
  - CPU

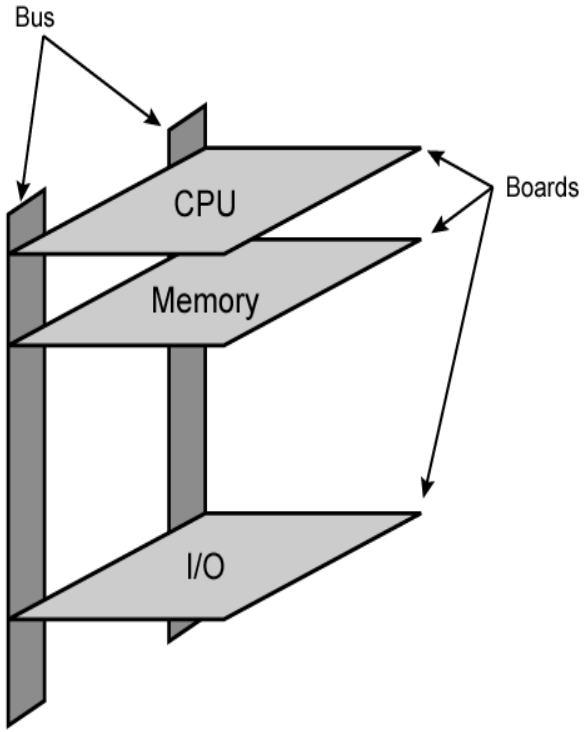






Buses

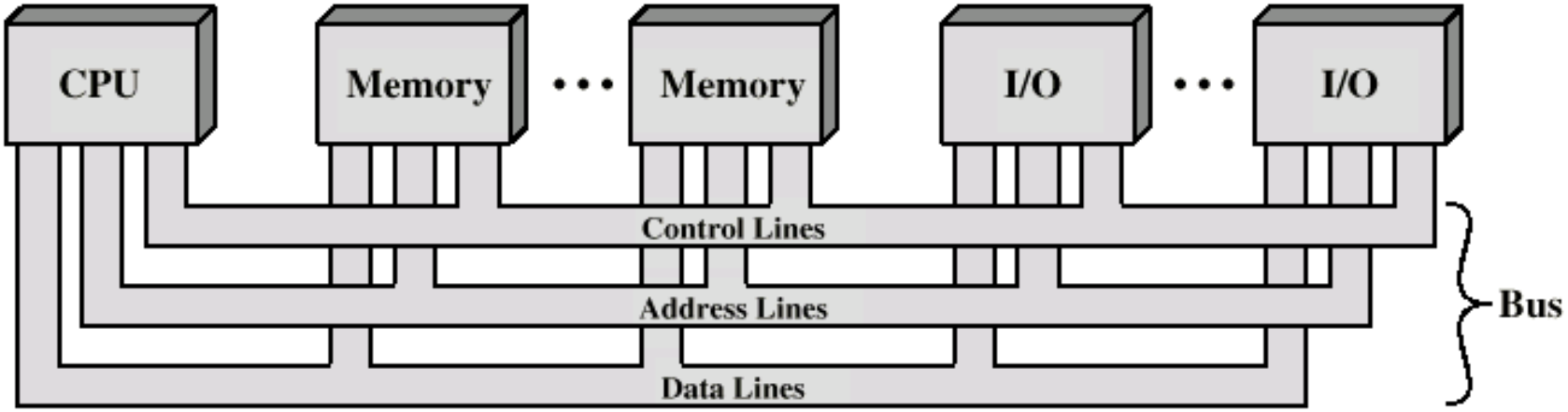
- Camino de comunicación entre dispositivos
- Normalmente broadcast
- El *bus del sistema* interconecta CPU, memoria y módulos de entrada/salida
- Líneas en paralelo, agrupadas por función
  - Ej. 32 bits de bus de datos y 32 bits de direcciones



Buses

- Bus de datos
  - Transporta datos
    - A este nivel no hay diferencia entre “datos” e “instrucciones” (Von Neuman)
  - Performance asociada al ancho del bus
    - 8, 16, 32, 64 bits
- Bus de direcciones
  - Identificar fuente o destino de los datos en memoria
  - Ancho del bus determina capacidad de memoria del sistema
    - Ej. 8080 tiene un bus de direcciones de 16 bits, resultando 64k de memoria direccionable
- Bus de Control
  - Información de control y temporización
    - Lectura/Escritura
    - Pedidos de interrupción
    - Señales de reloj

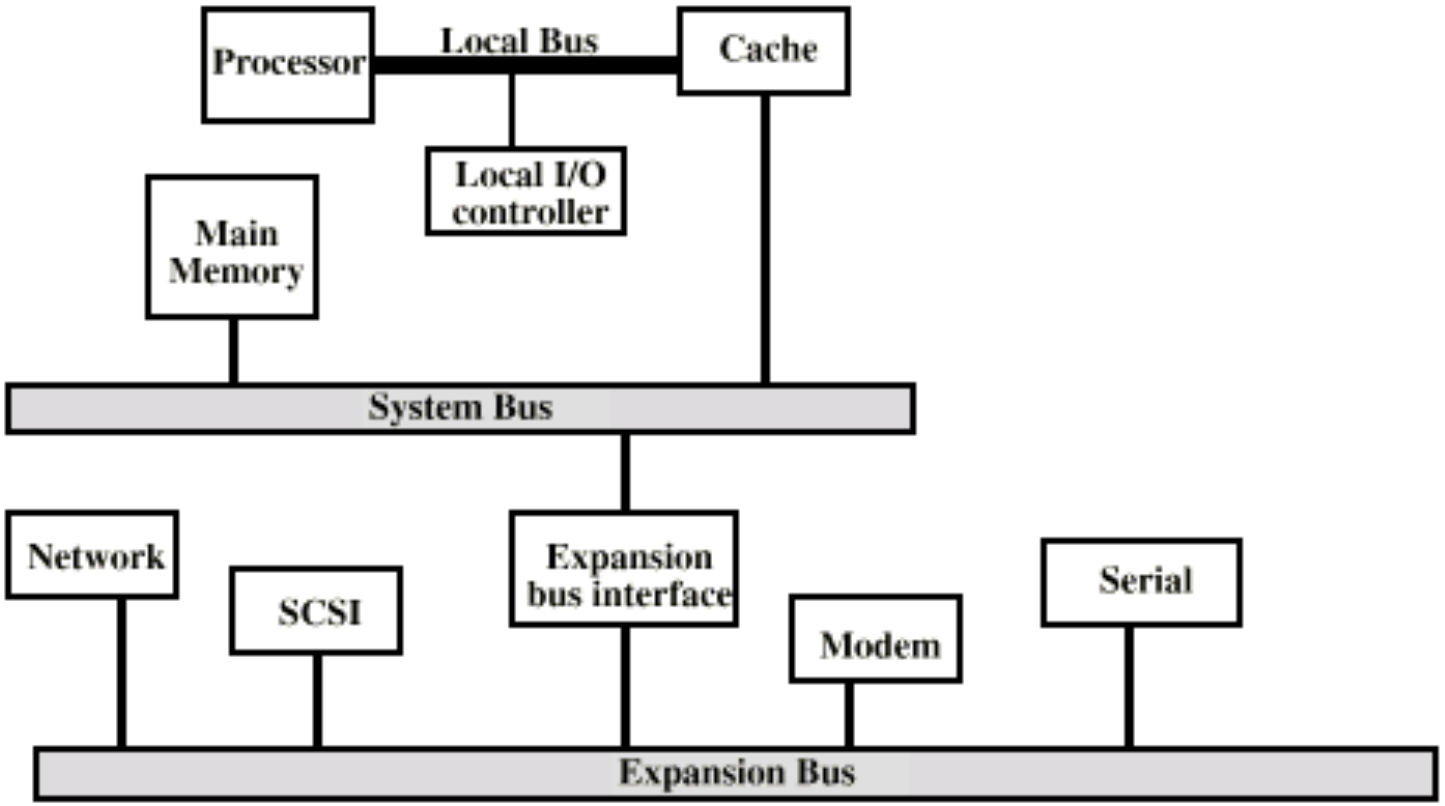
Esquema de interconexión del bus



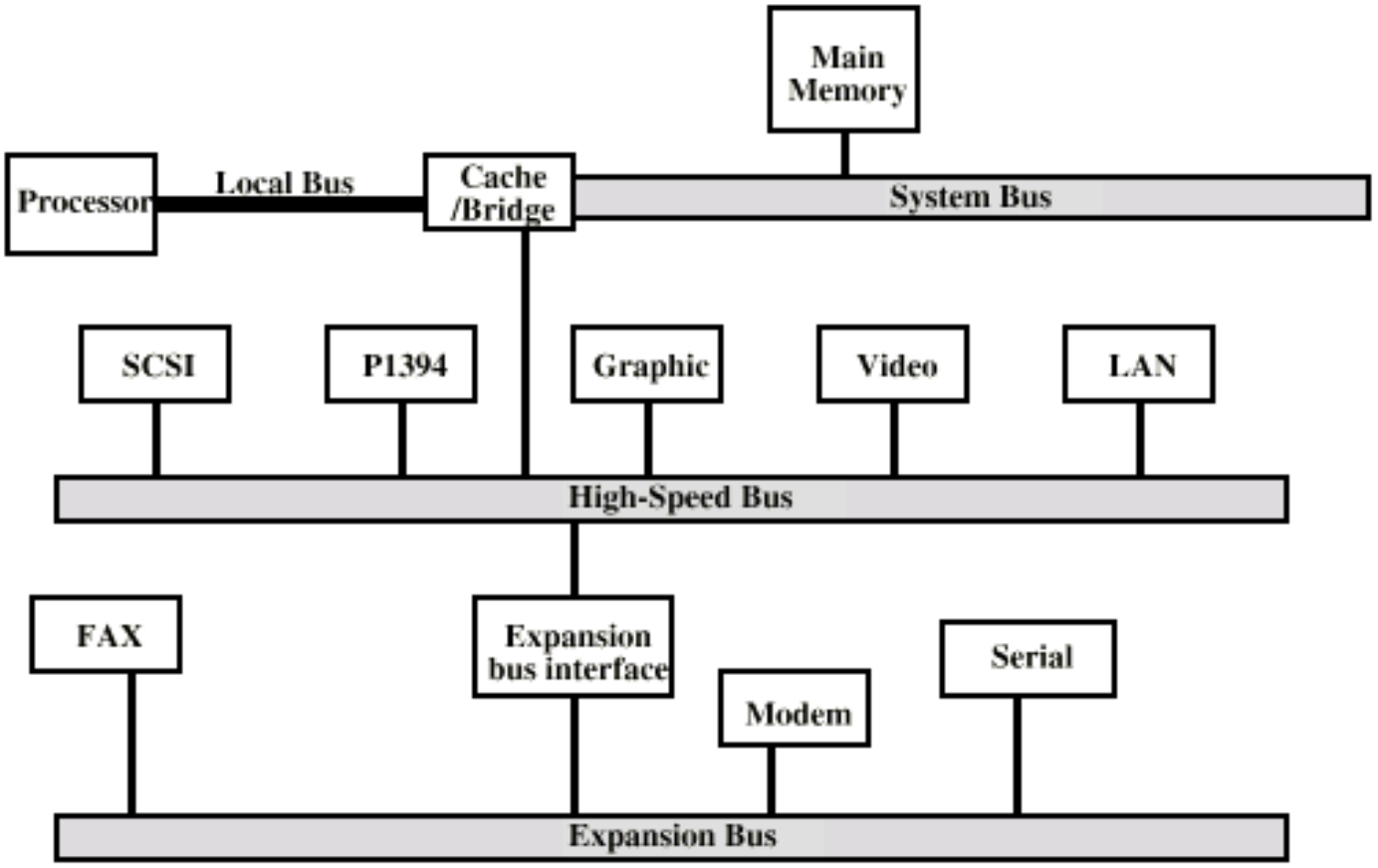
- Muchos dispositivos en el bus…
  - Retardos de propagación
    - Caminos de datos largos complican la coordinación y afectan negativamente la performance
  - Problemas eléctricos (“fan out”)

- En general se utilizan múltiples buses para contrarrestar estos problemas

Buses:
Tradicional (con cache)



Buses:
Bus de Alta Performance



Arbitraje del Bus

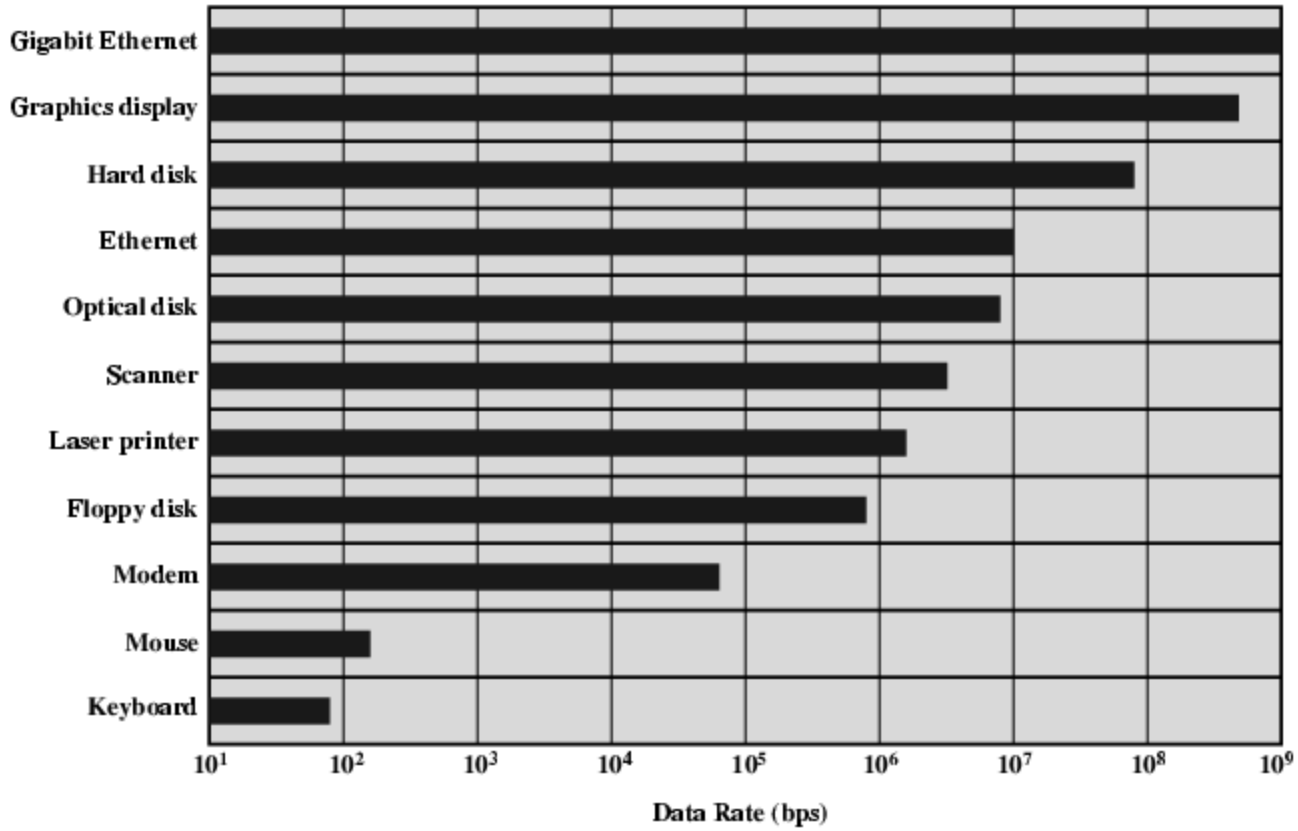
- Bus controlado por más de un módulo
  - Ej. CPU y controlador de DMA
- Solo un módulo puede controlar el bus a la vez!
- Se debe arbitrar

- Arbitraje Centralizado
  - Dispositivo hardware único controlando acceso al bus
    - Árbitro o Controlador del Bus
  - Puede ser parte de la CPU o separado
- Arbitraje Distribuido
  - Cada módulo puede reclamar el control del bus
  - Lógica de control en todos los módulos

E/S:
la interfaz con el procesador

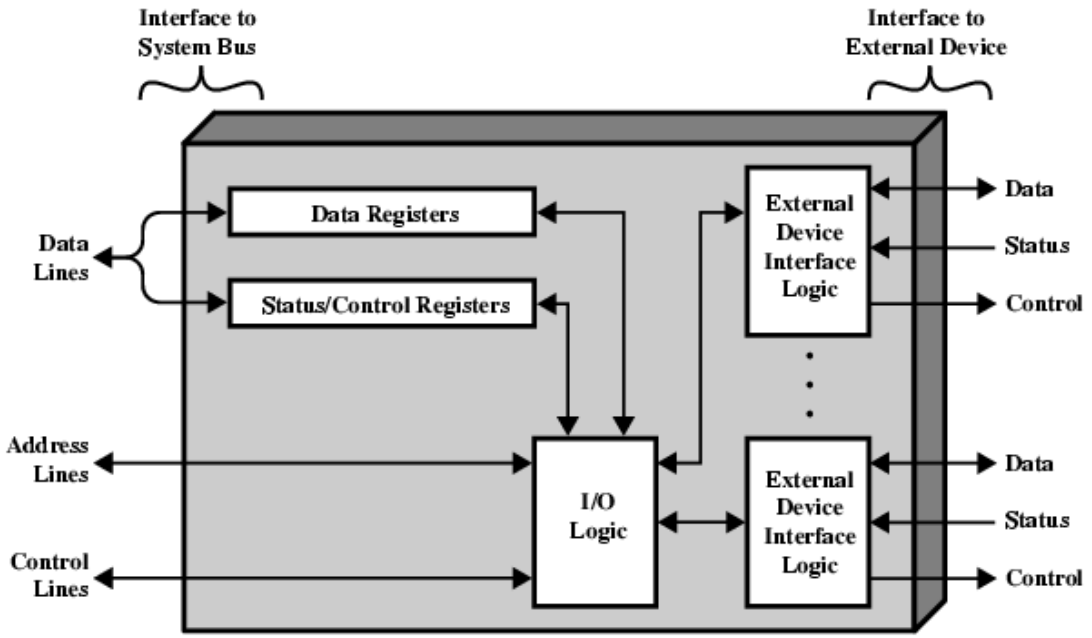
- Direccionamiento
  - E/S aislada
  - E/S mapeada en Memoria
- Control de E/S
  - Polling
  - Interrupciones
  - DMA
  - Controladores de E/S
  - Procesadores de E/S

Tasas de transferencia de E/S



Módulo de E/S

- Funciones
  - Control & Temporización
  - Comunicación con la CPU
  - Comunicación con dispositivos
  - Almacenamiento temporal (buffering) de datos
  - Detección de Errores



Módulo de E/S

- Mostrar propiedades del dispositivo a la CPU?
- Soporte de múltiples dispositivos?
- Control de funcionalidades del dispositivo, o dejar esa tarea a la CPU?

Direccionamiento de la E/S

- E/S aislada
  - Espacios de direcciones separados
  - Se necesitan líneas de selección entre E/S y memoria
  - Instrucciones específicas para E/S

- E/S mapeada en memoria
  - Dispositivos y memoria comparten espacio de direcciones
  - Accesos a E/S se ven como simples accesos a memoria
    - El HW debe resolver la diferencia
  - No hay instrucciones especiales para E/S
  - A cada dispositivo se le asigna un identificador único
  - Los comandos emitidos por la CPU contienen este identificador (dirección)

E/S Programada

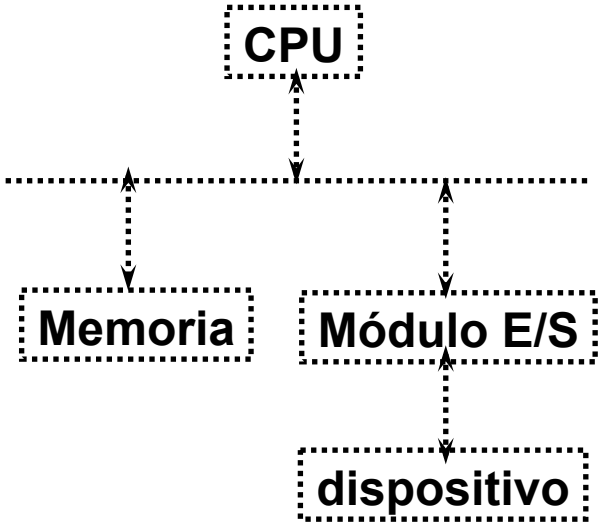
- La CPU controla directamente la E/S
  - Monitorizar estado
  - Comandos de lectura/escritura
  - Transferencia de datos
- La CPU debe esperar que el módulo de E/S complete las operaciones indicadas
  - Pérdida de tiempo de CPU!
- E/S Programada - detalle
  - CPU requiere una operación de E/S
  - Módulo de E/S ejecuta la operación
  - Módulo de E/S setea bits de estado
  - La CPU chequea los bits de estado periódicamente
  - Módulo de E/S no informa directamente a la CPU
    - O sea, no interrumpe a la CPU
  - La CPU puede esperar (“busy wait”), o rechequear más tarde

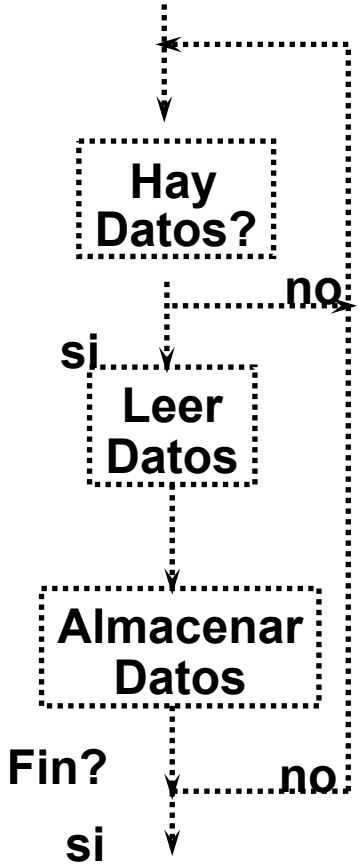
E/S Programada:
Comandos

- La CPU emite direcciones
  - Identifica módulo de E/S (y dispositivo si el módulo controla más de uno)

- La CPU emite comando
  - Control – especifica al módulo de E/S qué hacer
    - Ej. Situar cabezal de disco duro en determinada posición
  - Test - chequo de estado
    - Ej. Error?
  - Lectura/Escritura
    - El módulo de E/S transfiere datos desde/hacia el dispositivo

E/S Programada (diagrama)





Loop “busy wait”
no es una manera eficiente
de usar la CPU
a menos que el dispositivo
sea muy rápido!

E/S por Interrupciones

- Resuelve las esperas de la CPU
- No se debe hacer polling del estado
  - El módulo de E/S interrumpe cuando está pronto

- Como funciona? (Ej. lectura)
  - La CPU emite un comando de lectura
  - El módulo de E/S hace la transferencia desde el dispositivo mientras la CPU hace otra cosa
  - Una vez finalizada la transferencia, el módulo de E/S interrumpe a la CPU
  - La CPU requiere los datos
  - El módulo de E/S transfiere datos

E/S por Interrupciones:
Que hace la CPU?

- Emite comando de E/S
- Hace otra tarea
- Chequea por interrupciones al fin de cada ciclo de instrucción
  - Recordar: el procesamiento de interrupciones es uno de los desafíos complicados del control de la CPU
    - Más difícil en Superescalares
- Si hay interrupciones pendientes
  - Salvar contexto
  - Procesa interrupción

- Problemas de diseño
  - Como identificar el módulo que interrumpe?
  - Como resolver múltiples interrupciones?

E/S por Interrupciones:
Identificación del módulo que interrumpe

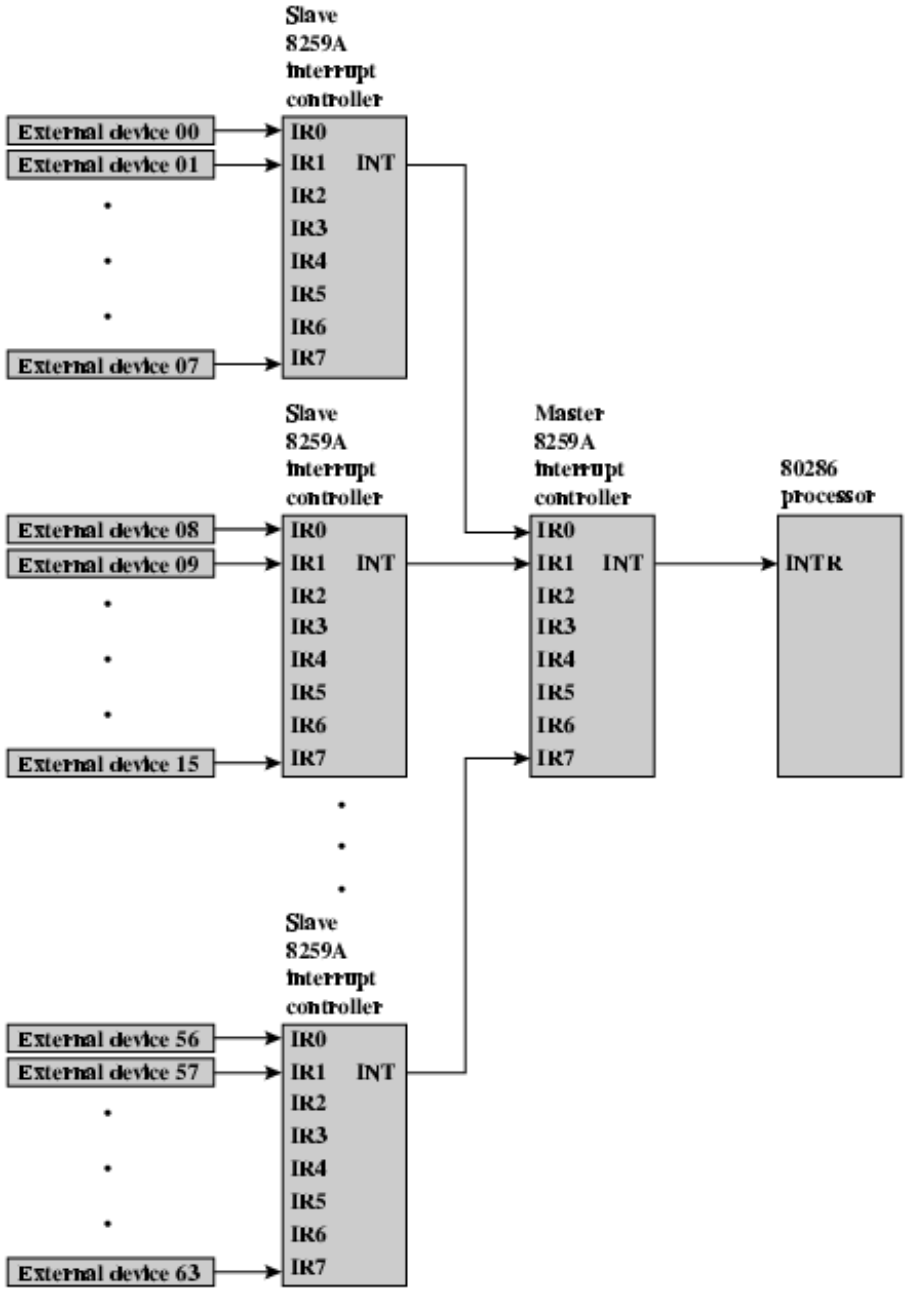
- Líneas diferentes para cada módulo
  - Número limitado de dispositivos
- Polling
  - La CPU interroga cada dispositivo (por software)
  - Lento…
- Daisy Chain o Polling de Hardware
  - Línea de interrupción compartida
  - La CPU envía "acknowledge" a una línea anidada que recorre todos los dispositivos
  - El módulo que interrumpió pone su código de identificación en el bus. A este código se le denomina *vector*
  - La CPU identifica el manejador mediante este código
- Arbitraje de bus
  - Módulo debe conseguir el bus para poder interrumpir
  - Cuando CPU reconoce la interrupción, el master pone su código de identificación en el bus.

Interrupciones múltiples

- Cada dispositivo tiene una prioridad determinada
- En una jerarquía de bus masters, solo el master actual puede interrumpir
- Ejemplo - Bus del PC
  - 80x86 tiene una sola línea de interrupción
    - Usan un controlador de interrupciones, el 8259A
    - El 8259A tiene 8 líneas de interrupción
- Secuencia de eventos
  - 8259A acepta interrupciones
  - 8259A determina prioridad
  - 8259A interrumpe al 8086 (pone un “1” en la línea INTR)
  - CPU reconoce la interrupción (pone un “1” en la línea INTA)
  - 8259A pone el identificador del dispositivo en el bus de datos
  - CPU procesa la interrupción

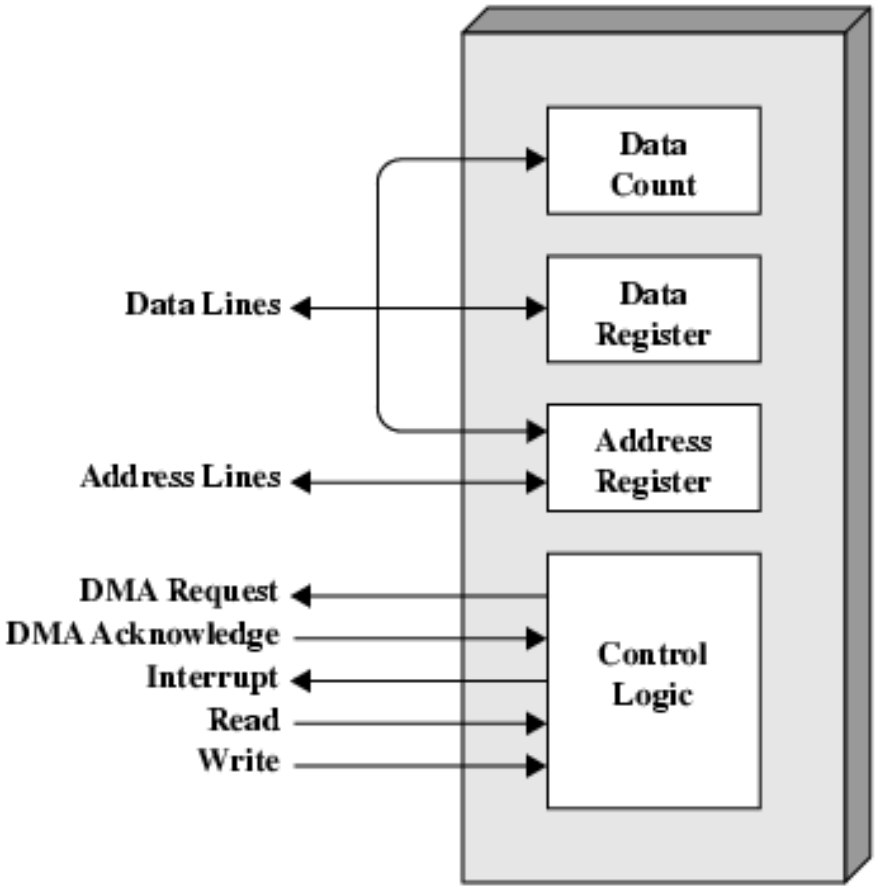
Interrupciones múltiples

- Se pueden poner en cascada como en la figura

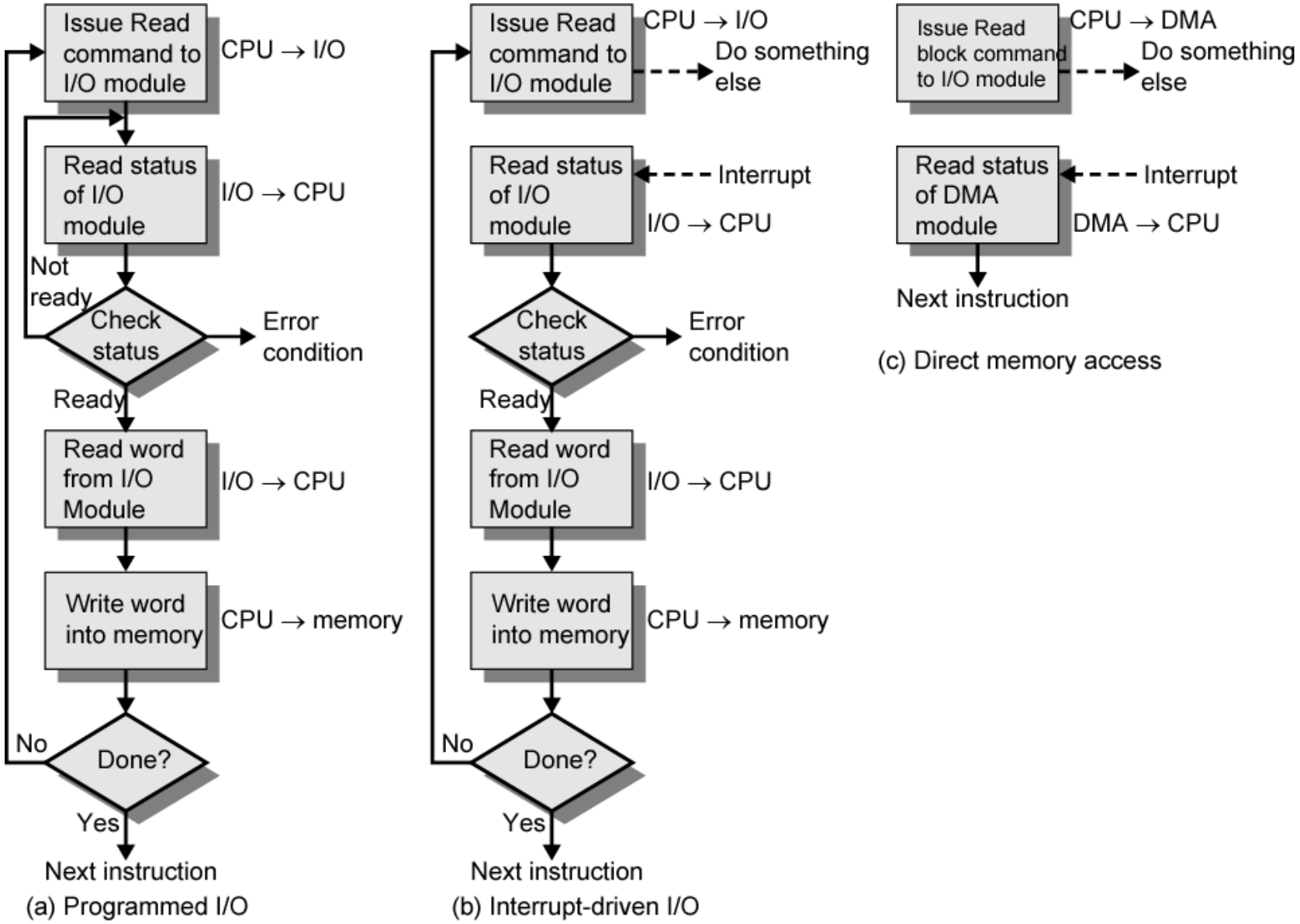


Direct Memory Access

- E/S programada y por interrupciones requieren la intervención activa de la CPU
  - Limita la velocidad de transferencia
  - CPU "queda atada" a la E/S
- Alternativa DMA
  - Módulo adicional (hardware) en el bus
  - El controlador de DMA "sustituye" a la CPU en tareas de E/S



Técnicas para leer un bloque de datos de entrada

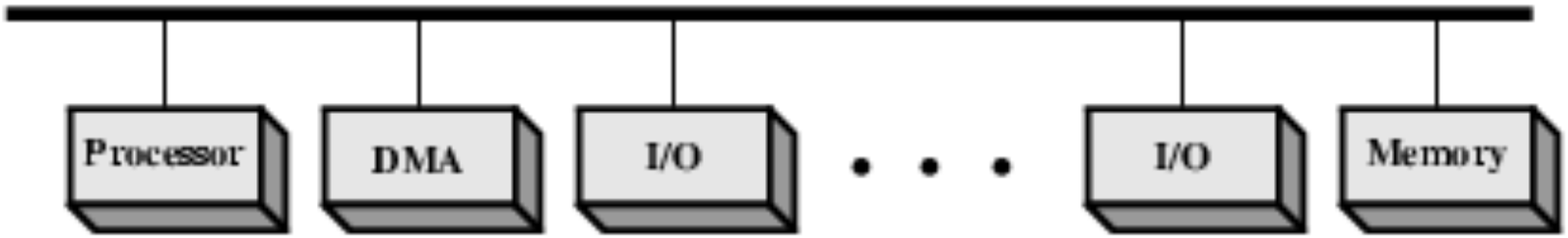


Direct Memory Access:
Cómo funciona?

- La CPU programa el controlador de DMA:
  - Lectura/Escritura
  - Dirección del dispositivo
  - Dirección de comienzo del bloque de memoria
  - Cantidad de datos a transferir
- La CPU hace otra tarea mientras el controlador de DMA se encarga de la transferencia
- El controlador de DMA interrumpe cuando finaliza

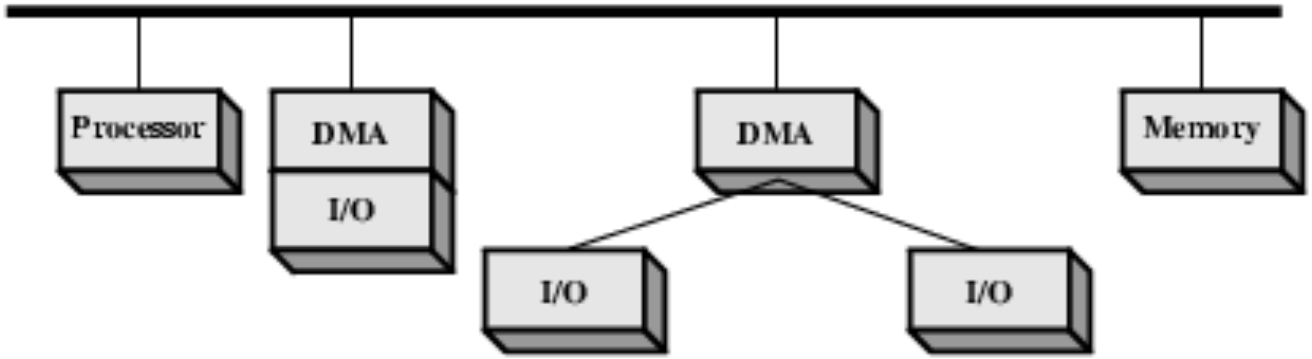
- "Robo" de ciclos del DMA
  - Por cada palabra transferida, el DMA se adueña del bus
  - No es una interrupción (no hay que cambiar de contexto)
  - La CPU queda suspendida si necesita acceder al bus
    - Ej. Antes de leer o escribir en memoria por fallo de cache
  - Puede enlentecer la CPU, pero no tanto como si la CPU tuviera que hacer la transferencia

Direct Memory Access:
Conexión al bus



- Un solo Bus, controlador de DMA desacoplado
  - Cada transferencia usa el bus dos veces
    - E/S <-> DMA, DMA <-> memoria
  - CPU puede ser suspendida hasta dos veces por cada transferencia

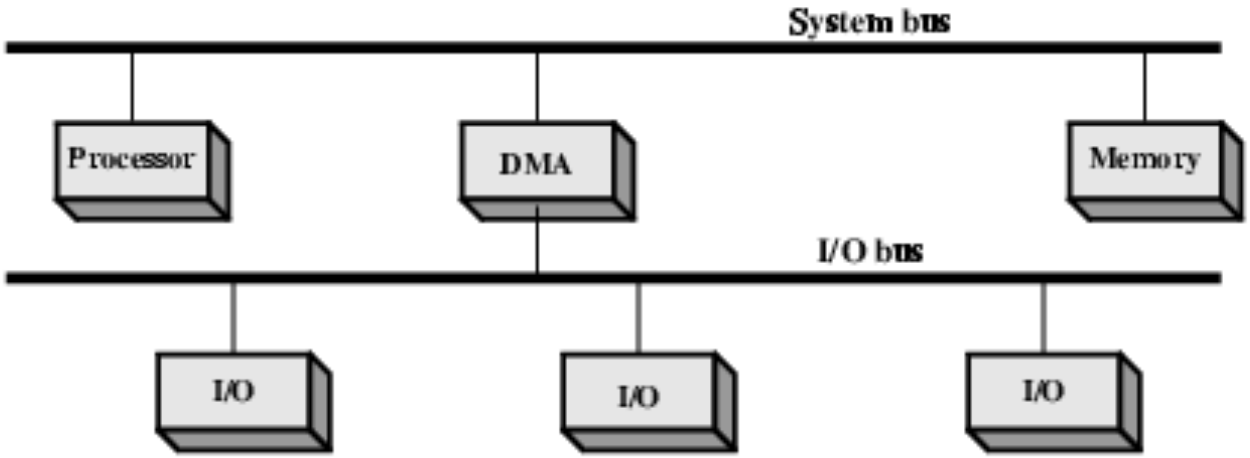
Direct Memory Access:
Conexión al bus



(b) Single-bus, Integrated DMA-I/O

- Un solo Bus, controlador de DMA integrado con E/S
  - El controlador puede soportar más de un dispositivo
  - Cada transferencia usa el bus una sola vez
    - DMA <-> memoria
- CPU suspendida una sola vez como máximo por cada transferencia

Direct Memory Access:
Conexión al bus

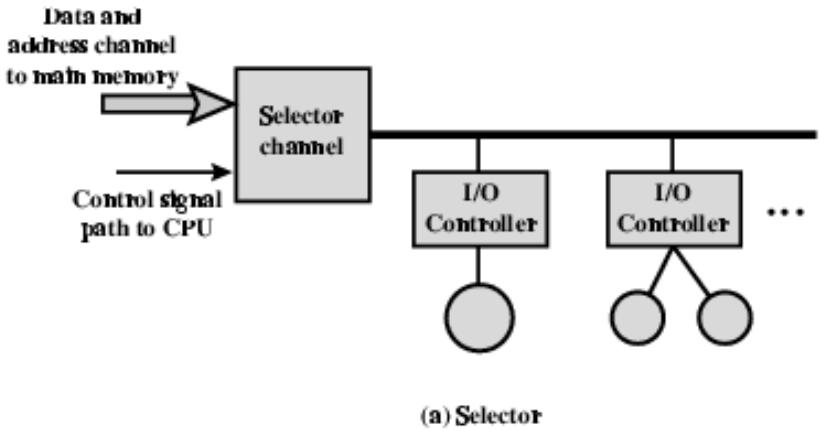


(c) I/O bus

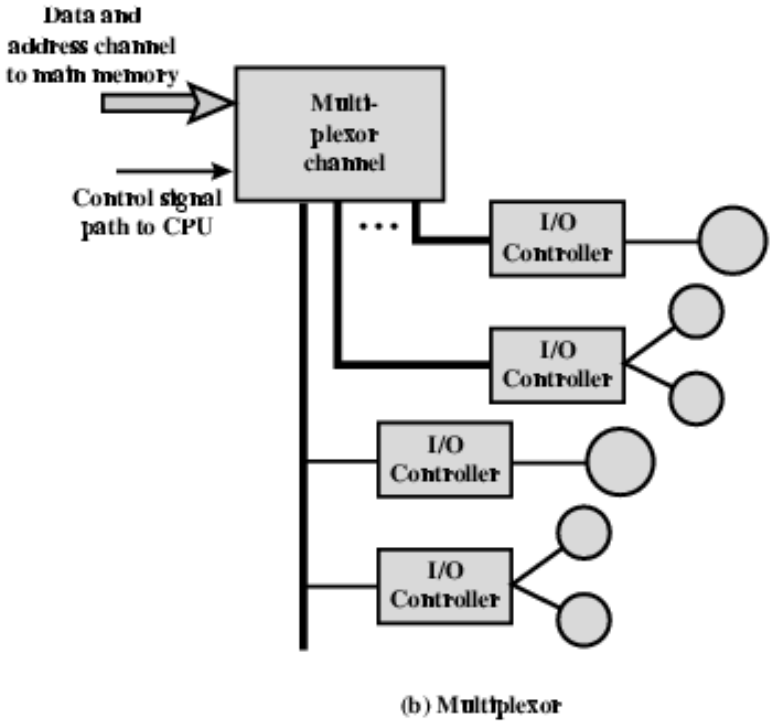
- Bus de E/S separado
  - Permite escalar el DMA
    - El bus soporta todos los dispositivos que usan el DMA
  - Cada transferencia usa el bus del sistema una sola vez
    - DMA <-> memoria
- CPU suspendida una sola vez como máximo por cada transferencia

Procesadores de Entrada/Salida

- Sofisticación de los módulos de E/S
  - Ej. placa de gráficos 3D
- Procesadores / canales / controlador de E/S
  - Pueden ejecutar su propio juego de instrucciones para realizar operaciones complejas
    - Procesador
    - Puede tener memoria local
  - A veces los términos Procesador y Canal de E/S se usan para distinguir la presencia o no de memoria local
- Ejemplo
  - Imprimir una página de 80 x 60 caracteres provocaría 4800 interrupciones. Podría delegarse el trabajo de imprimir una página completa a un procesador de E/S



(a) Selector



(b) Multiplexor

Ejemplo: GPUs

- Ejemplo 1
  - Trabajo con modelos numéricos (para simulaciones de fluidos) en 3 dimensiones, con simplificaciones en los modelos para bajar tiempo de cómputo. Es habitual no trabajar con un modelo realmente en 3 dimensiones sino realizar los cálculos por capas (en la horizontal) y cada cierta cantidad de pasos de tiempo simulado hacer una corrección verificando condiciones en la vertical.
  - Si cada paso de tiempo simulado implica 10 seg de CPU y la verificación en la vertical cuesta 50 seg, el diagrama de tiempos de la ejecución del programa será la siguiente:

  - Si se cuenta con una GPU y se implementa los cálculos de la vertical en dicho dispositivo, se puede trabajar en forma concurrente con ambos procesadores calculando cada 5 pasos de tiempo las condiciones en la vertical.

Ejemplo: GPUs

- Ejemplo 2
  - Cuando se generan imágenes por computadora se realizan diversos cálculos, uno de ellos el cálculo de la luz en cada pixel de la imagen.
  - Existen implementaciones de estos algoritmos de Radiosidad en GPU.
  - Entonces mientras la GPU calcula la luz de una imagen la CPU puede estar calculando otras propiedades de la imagen u otra etapa de la aplicación, por ejemplo la lógica de un juego de computadora.

Relaciones con otros componentes

- DMA y memorias cache
  - Potenciales problemas de coherencia de cache
  - Política de escritura write-thorugh soluciona una parte del problema (operaciones de salida)
  - Soluciones por software
    - E/S siempre a páginas marcadas como noncachable
    - Vaciado de cache forzado luego de una lectura
  - Soluciones por hardware
    - Se vigila el bus del sistema para detectar escrituras a bloques que están en cache (snooping). Se pueden duplicar tags para no enlentecer la cache
  - Técnicas análogas se pueden usar para salidas bajo una política de escritura write-back
- DMA y memoria virtual
  - Buffer en páginas no contiguas en memoria real
  - La página que se está transfiriendo podría reemplazarse
  - *DMA virtual:* incluye registros de mapeo de direcciones virtuales a físicas (tantos como páginas puedan transferirse)

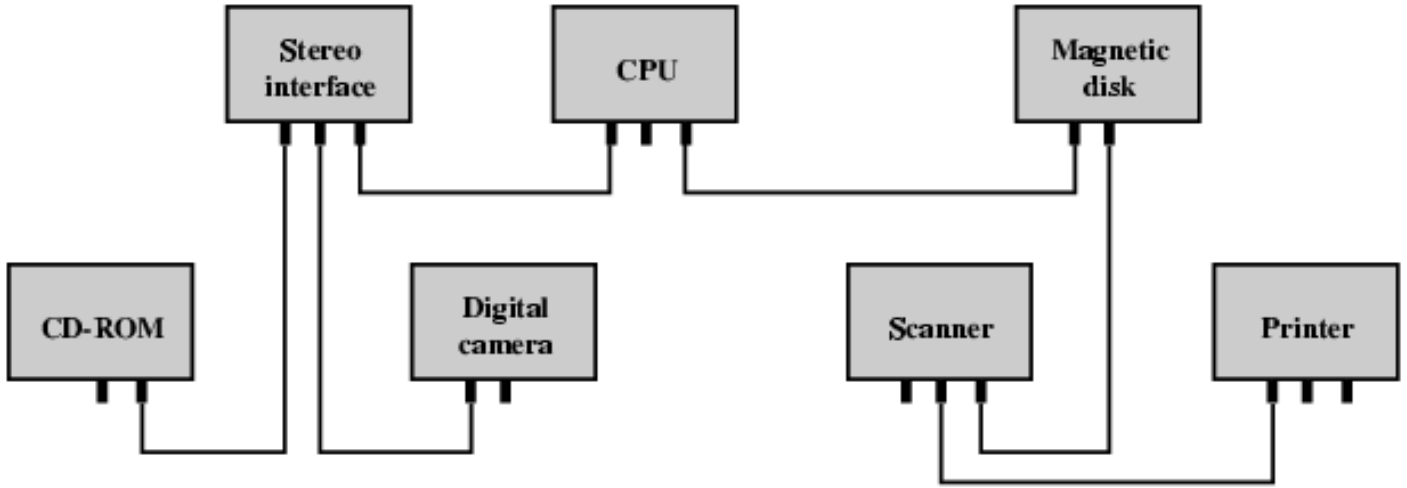
Interfaces de alta velocidad

- El objetivo es conectar dispositivos a alta velocidad
- Históricamente:
  - Procesador/memoria/buses dedicados
  - Ethernet, ATM (WAN)
  - Evolución Ethernet: 1Gbps, 10Gbps…
- Extensiones de IP/Ethernet para SANs (Storage Area Networks)
  - iSCSI
  - i Fibre Channel
- Otras tecnologías…
  - FireWire
  - Myrinet, QsNet: usadas en clusters HPC (propietarias)
  - InfiniBand

IEEE 1394 FireWire (1995)

- Bus serial de alta performance
  - Rápido
  - Bajo costo
  - Fácil de implementar
  - Usado por Apple, Sony, HP y otros
- Tasas de tranferencia desde 25 a 400Mbps
- IEEE 1394b (2002): tasas > 800 Mbps
- Similar a USB.
  - USB 2.0 (2001) llega a 480 Mbps.
  - USB 2.0 **host-based**
  - FireWire **peer-to-peer**
- Firewire S3200 y USB 3.0 en camino

IEEE 1394 FireWire:
Configuración

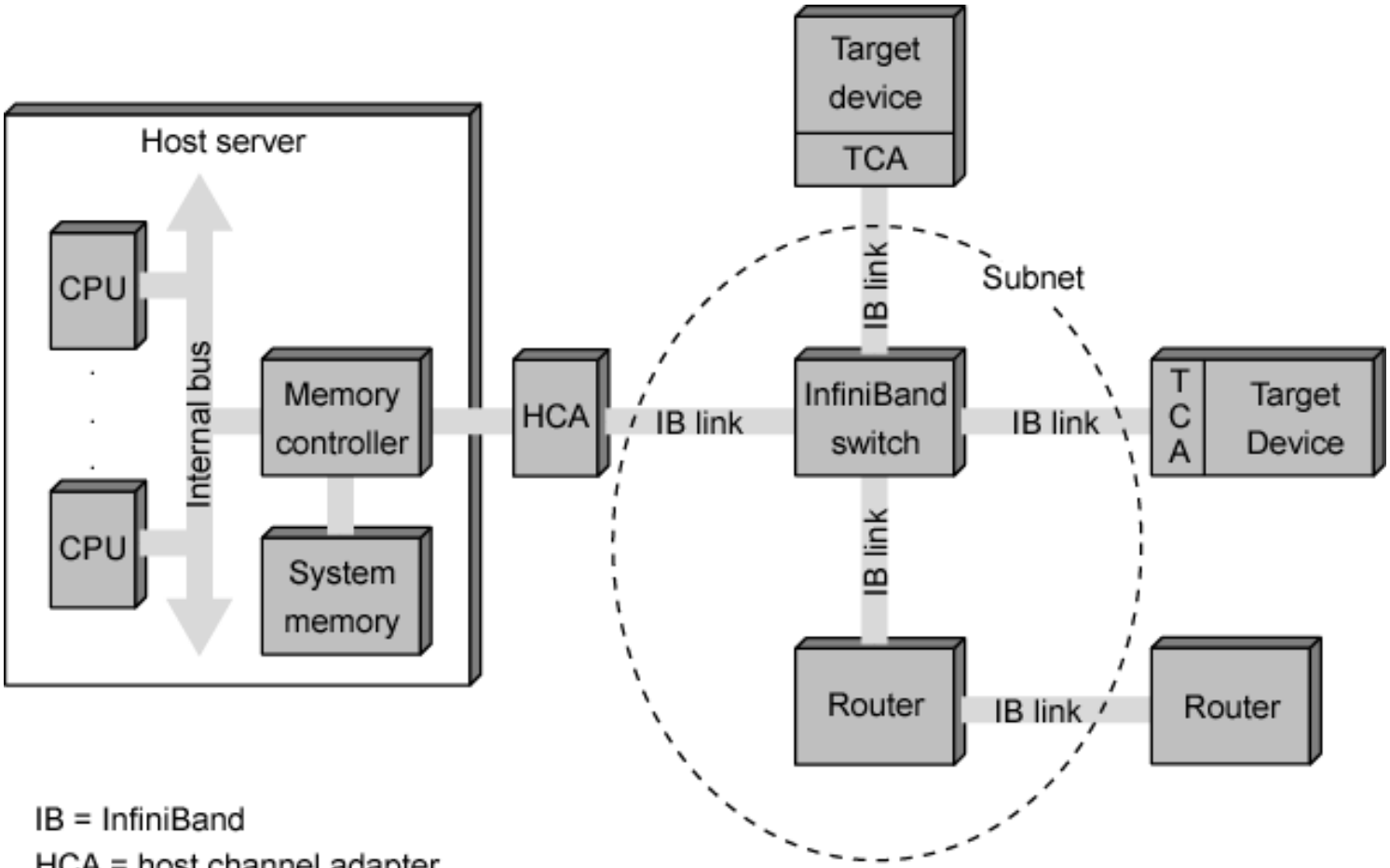


- Estructura de árbol. Daisy chain como caso particular
- Hasta 63 dispositivos en cada puerto
- Hasta 1022 buses usando puentes
- Configuración automática
- No se necesitan terminadores de bus

InfiniBand

- Especificación de E/S para servidores “high end”
  - Unión de “Future I/O” (Cisco, HP, Compaq, IBM y otros) y “Next Generation I/O” (Intel, Sun, Dell, Siemens y otros)
  - Versión 1 publicada en 2001
  - Arquitectura para flujo de datos entre dispositivos inteligentes (ej. servidor, almacenamiento, red)
  - Aplicación fundamentalmente en clusters HPC.
  - Las comunicaciones son punto a punto, a través de switches, no a través de un bus compartido.

InfiniBand Switch Fabric



InfiniBand

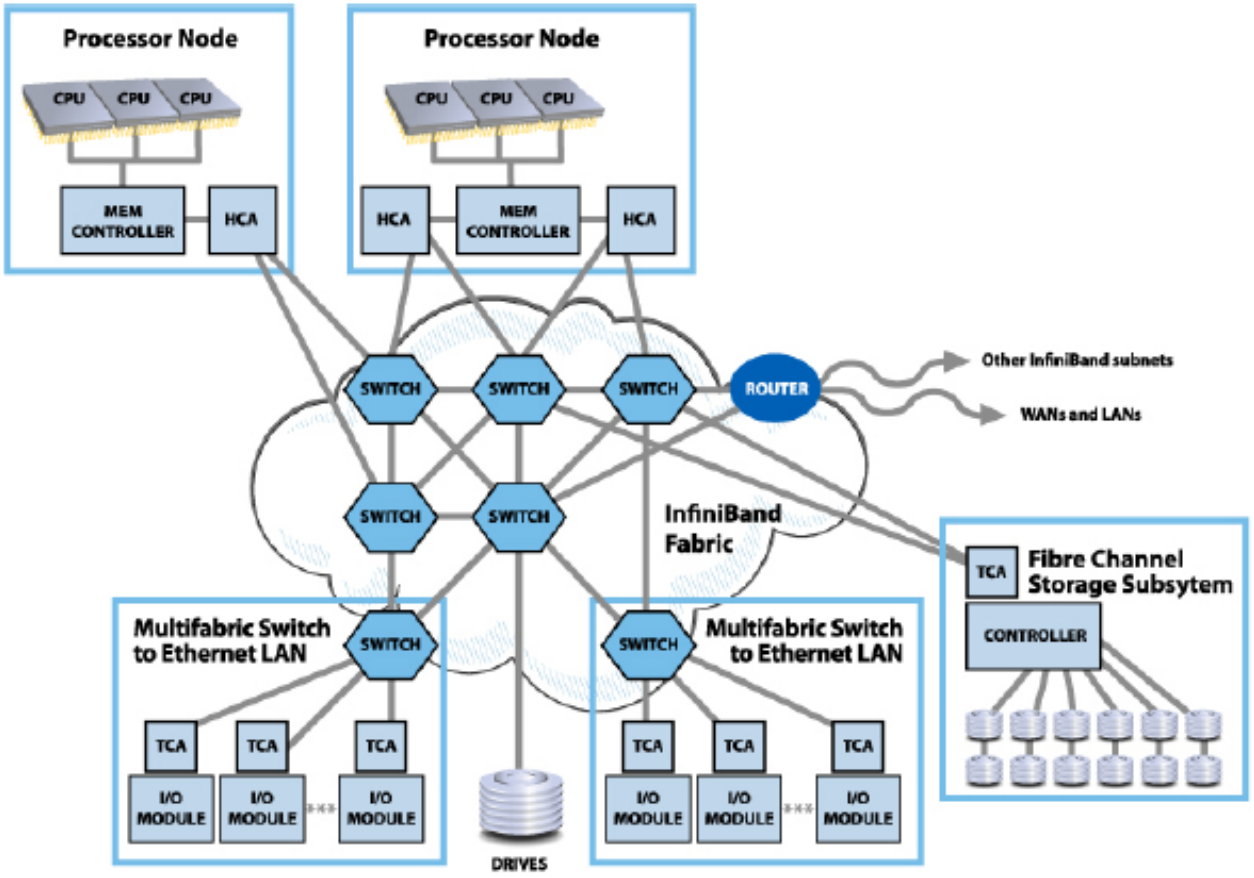
- Mayor densidad de dispositivos
- “Data center” escalable
- Nodos independientes se pueden agregar a demanda
- Distancias de la E/S al servidor
  - 17m en cobre / 300m fibra óptica
- Hasta 30 Gbps
- RDMA (Remote DMA) -> baja latencia.

InfiniBand

- InfiniBand compite / se complementa con
  - 1 y 10 Gigabit Ethernet
  - Fibre Channel

Technology	Standards Body	Signaling Speed	First Standard	Maximum Frame Size	Primary Application
Gigabit Ethernet	IEEE & IETF	1.25 Gbps	1999	1.5K	LAN: Local Area Network
Fibre Channel	ANSI	2.125 Gbps	1988	2K	SAN: Storage Area Network
InfiniBand Architecture	InfiniBand Trade Association	2.5Gbps (1x) 10Gbps (4x) 30Gbps (12x)	2001	4K	IAN: I/O Area Network

InfiniBand


Processor Node
CPU
CPU
CPU
MEM CONTROLLER
HCA
Processor Node
CPU
CPU
CPU
HCA
MEM CONTROLLER
HCA
SWITCH
SWITCH
SWITCH
ROUTER
Other InfiniBand subnets
WANs and LANs
SWITCH
SWITCH
InfiniBand Fabric
Multifabric Switch to Ethernet LAN
SWITCH
TCA
TCA
TCA
I/O MODULE
I/O MODULE
I/O MODULE
DRIVES
SWITCH
Multifabric Switch to Ethernet LAN
TCA
TCA
TCA
I/O MODULE
I/O MODULE
I/O MODULE
TCA
Fibre Channel Storage Subsytem
CONTROLLER

Tipos de Memoria Externa

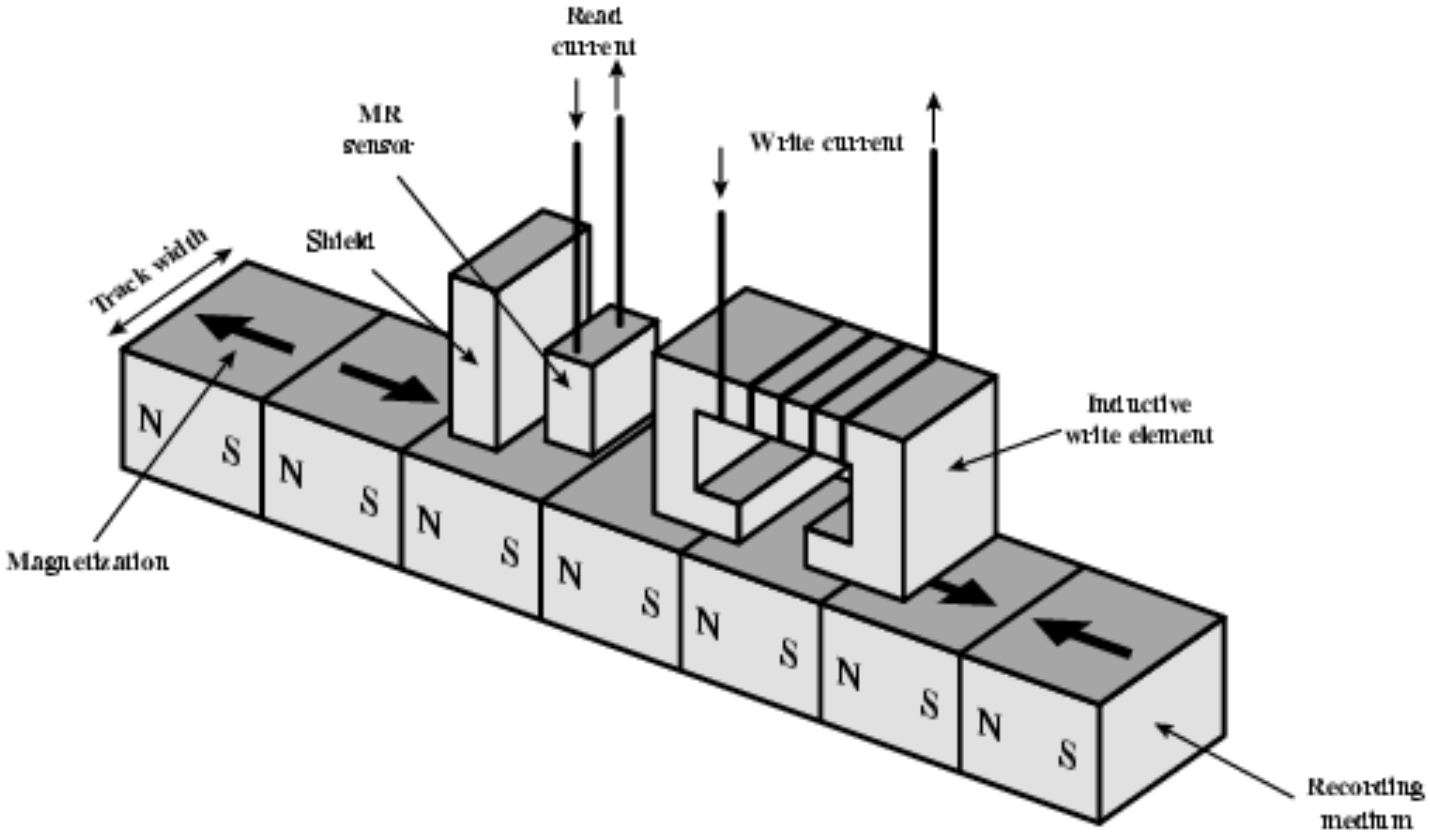
- Discos Magnéticos
  - RAID
  - Removible
- Discos Ópticos
  - CD-ROM
  - CD-Recordable (CD-R)
  - CD-R/W
  - DVD
- Cinta Magnética

Discos Magnéticos

- Sustrato cubierto por material magnetizable (p.ej. óxido de hierro)

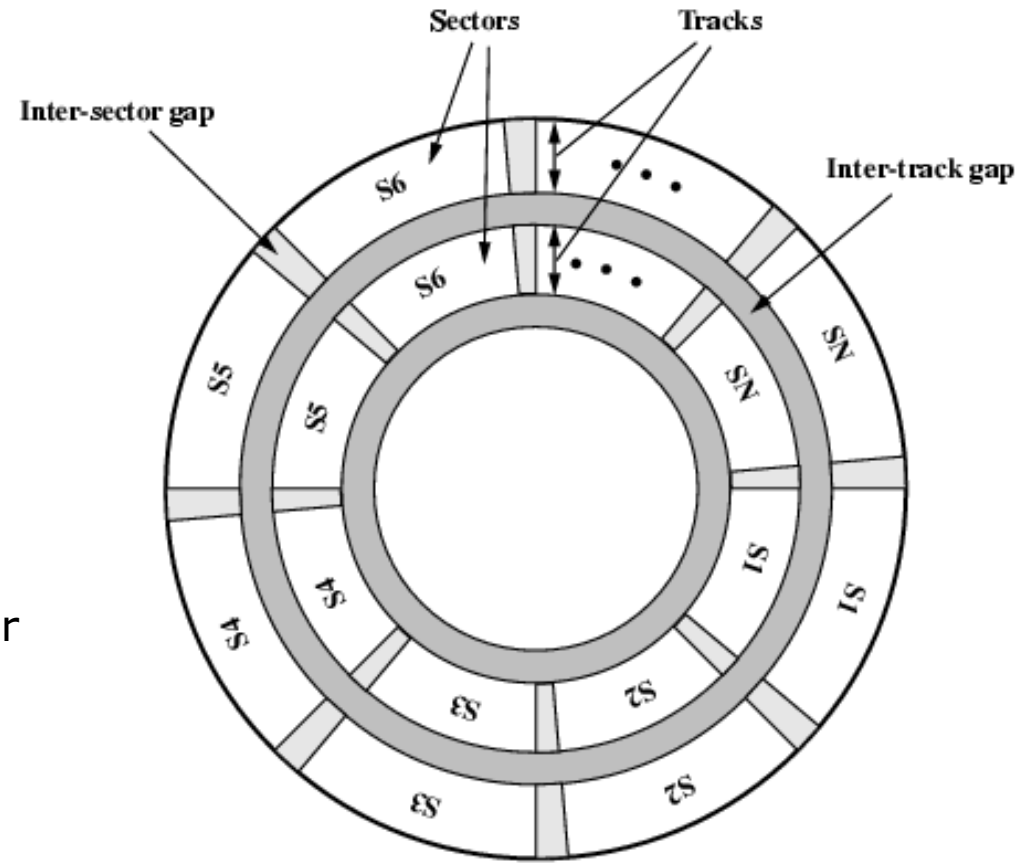
- Mecanismos de lectura y escritura
  - Grabado y recuperación mediante una **cabeza** magnética
    - Puede haber una cabeza de lectura/escritura o separadas
  - Durante la lectura/escritura la cabeza está fija y rotan los platos
  - Escritura
    - Corriente en la cabeza produce un campo magnético
    - Se aplican pulsos de corriente en la cabeza
    - Patrón magnético grabado en la superficie
  - Lectura
    - El campo magnético debido al movimiento relativo a la cabeza induce corriente
    - Actualmente se usan cabezas de lectura magneto-resistivas
    - Permiten mayor frecuencia de operación
      - Mejor densidad de almacenamiento y más velocidad de acceso

Discos Magnéticos:
Escritura inductiva, lectura MR



Discos Magnéticos:
Formato y organización de los datos

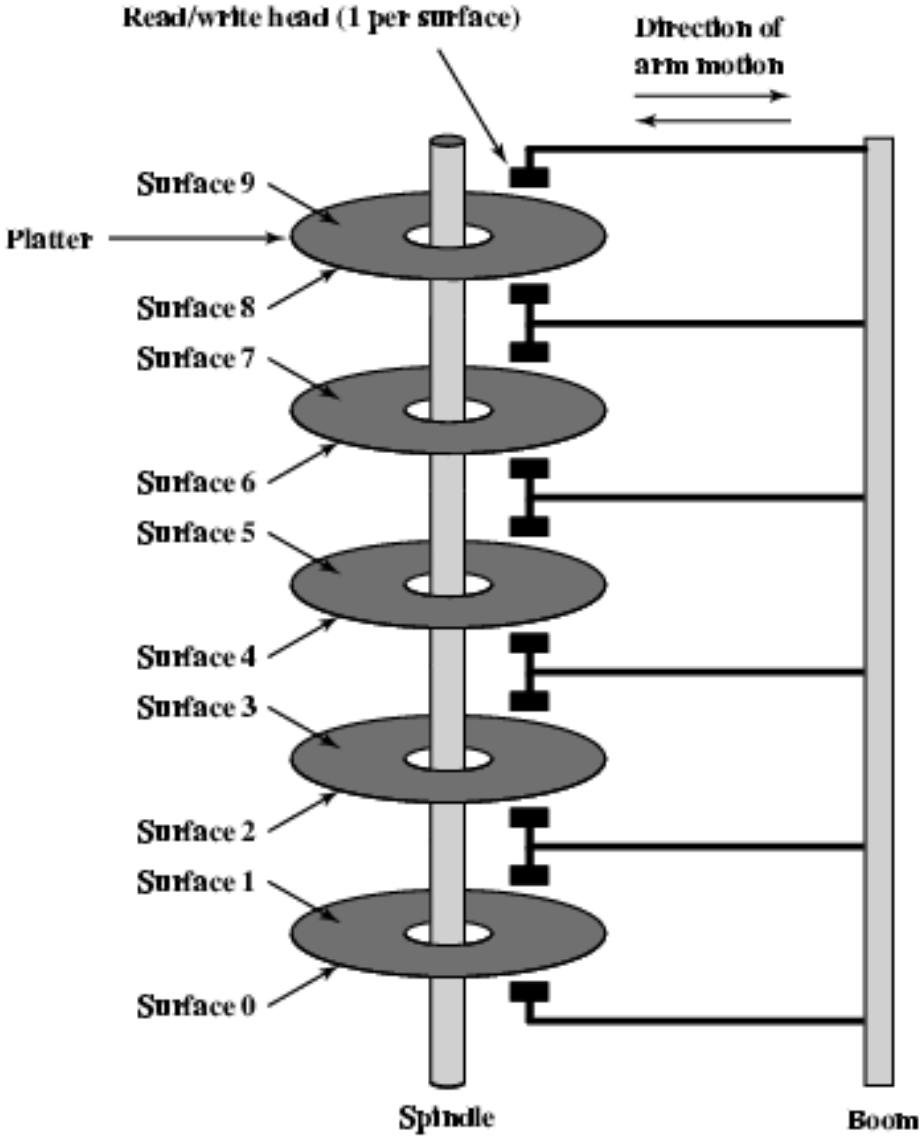
- Anillos concéntricos llamados **pistas** (**tracks)**
  - "Gaps" entre tracks
  - Gap más pequeño -> mayor capacidad
  - Igual cantidad de bits por track (densidad variable)
  - Velocidad angular constante
- Tracks divididas en **sectores**
  - "Block size" mínimo es un sector
  - Puede haber más de un sector por bloque



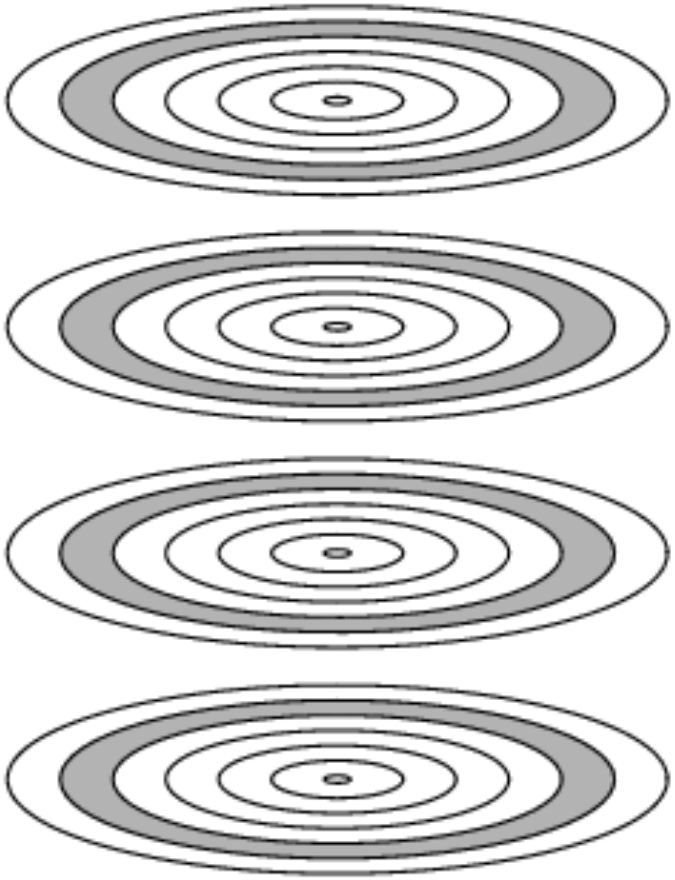
Discos Magnéticos:
Múltiples platos

- Una cabeza por lado
  - Cabezas se alinean juntas
- Traks alineadas en cada plato forman **cilindros**

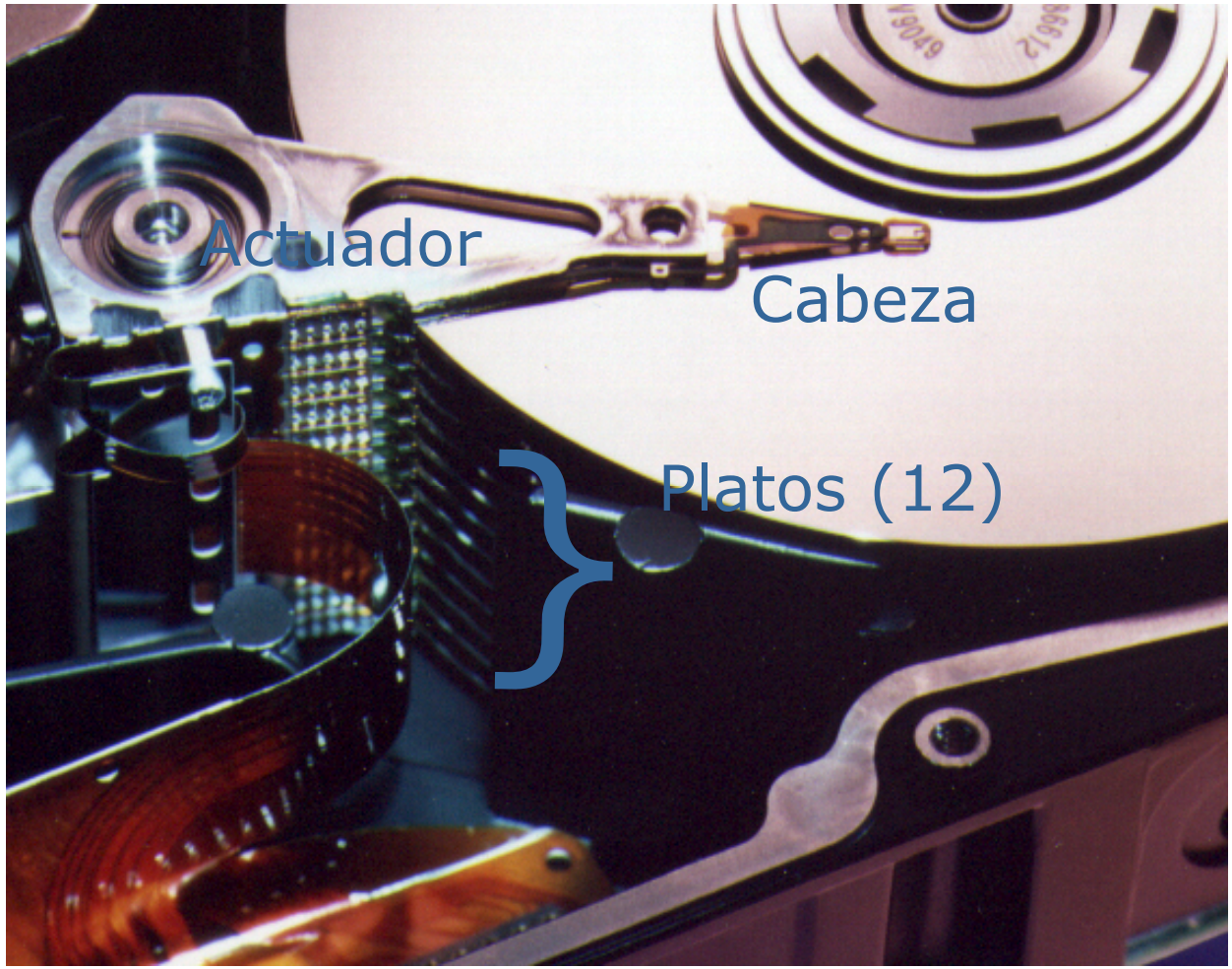
- Los datos son repartidos por cilindro
  - Reduce movimiento de las cabezas
  - Aumenta velocidad (tasa de transferencia)



Discos Magnéticos:
Cilindros



Discos Magnéticos:
De la vida real…



Discos Magnéticos:
Rendimiento (i)

- Disk Latency = Seek Time + Rotation Time + Transfer Time + Controller Overhead
  - Seek Time
    - Situar cabeza en el track correcto
  - Rotation Time
    - Esperar que los datos roten debajo de la cabeza
  - Transfer Time
    - Depende de la tasa de transferencia del disco (ancho de banda, "bit density"), tamaño del bloque requerido

- Parámetros que interesan
  - Capacidad: 100%/año (2X / año)
  - Tasa de Transferencia (BW): 40% / año (2X / 2 años)
  - Rotation + Seek time: 8% / año (1/2 en 10 años)
  - MB/$: 100%/año (2X / año)

Discos Magnéticos:
Rendimiento(ii)

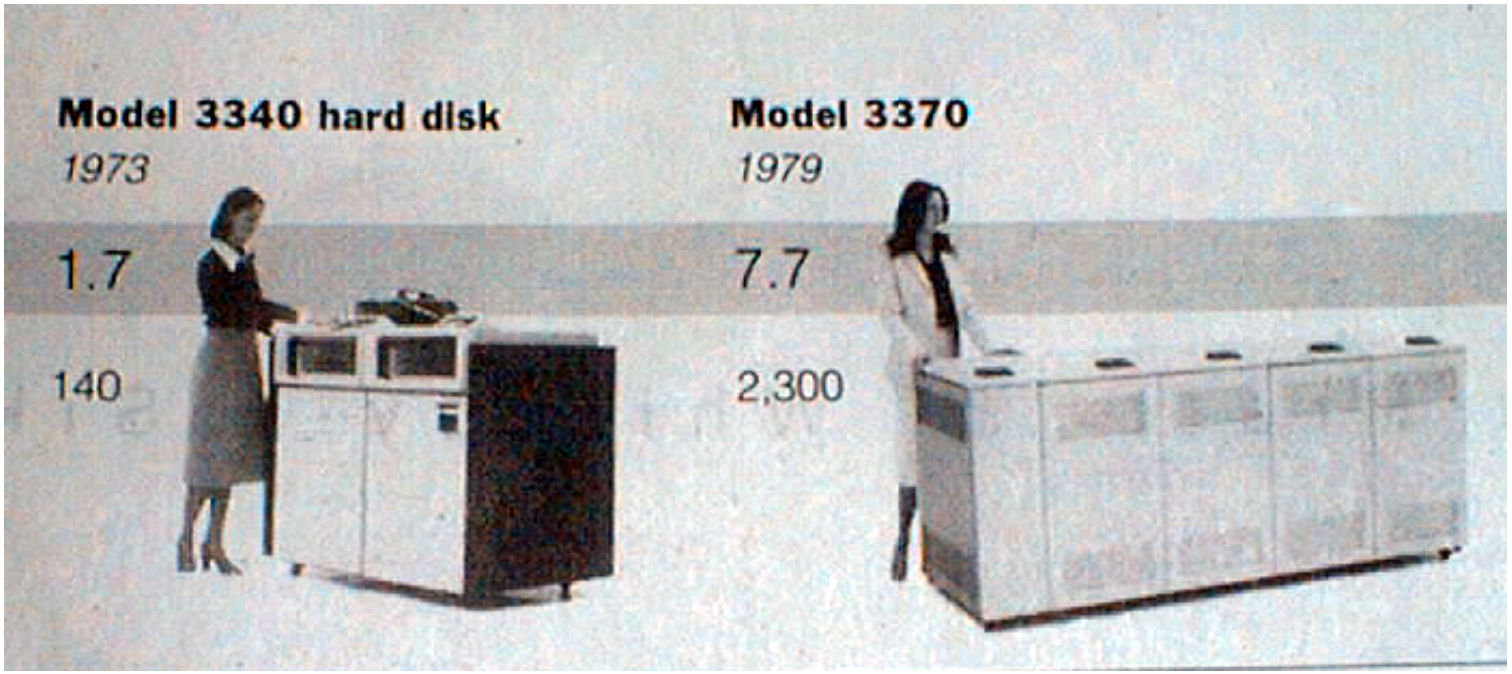
- Ejemplo: Barracuda 180
  - 181.6 GB, 3.5 inch disk
  - 12 platos, 24 superficies
  - 24,247 cilindros
  - 7,200 RPM; (4.2 ms avg. latency)
  - 7.4/8.2 ms avg. seek (r/w)
  - 63.5 to 35 MB/s (interno)
  - 0.1 ms controller time
  - 10.3 watts (idle)

- Calcular tiempo para leer 64 KB (128 sectores)
  - Disk latency = average seek time + average rotational delay + transfer time + controller overhead
    - = 7.4 ms + 0.5 * 1/(7200 RPM) + 64 KB / (63.5 MB/s) + 0.1 ms
    - ≈ 7.4 ms + 0.5 /(7200 RPM/(60000ms/M)) + 64 KB / (65 KB/ms) + 0.1 ms
    - ≈ 7.4 + 4.2 + 1.0 + 0.1 ms = **12.7 ms**

Discos Magnéticos:
Historia (i)

Data density Mbit/sq. in.

Capacity of Unit Shown Megabytes

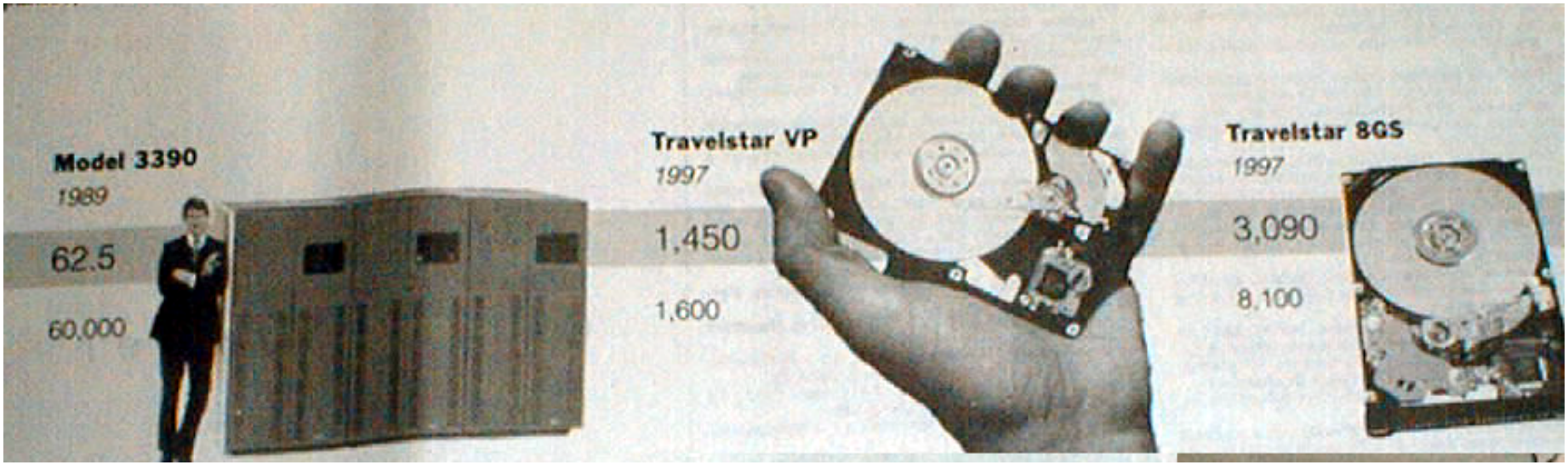


1973:
1. 7 Mbit/sq. in
140 MBytes

1979:
7. 7 Mbit/sq. in
2,300 MBytes

source: New York Times, 2/23/98, page C3,
"Makers of disk drives crowd even more data into even smaller spaces"

Discos Magnéticos:
Historia (ii)



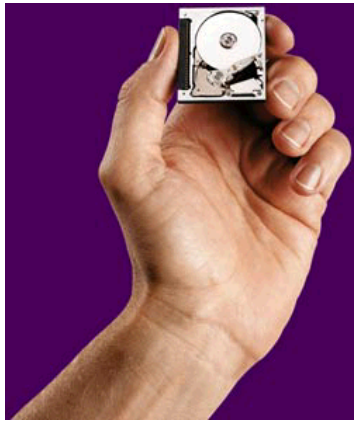
1989:
63 Mbit/sq. in
60,000 MBytes

1997:
1450 Mbit/sq. in
2300 MBytes

1997:
3090 Mbit/sq. in
8100 MBytes

source: New York Times, 2/23/98, page C3,
"Makers of disk drives crowd even more data into even smaller spaces"

Discos Magnéticos:
Historia (iii)



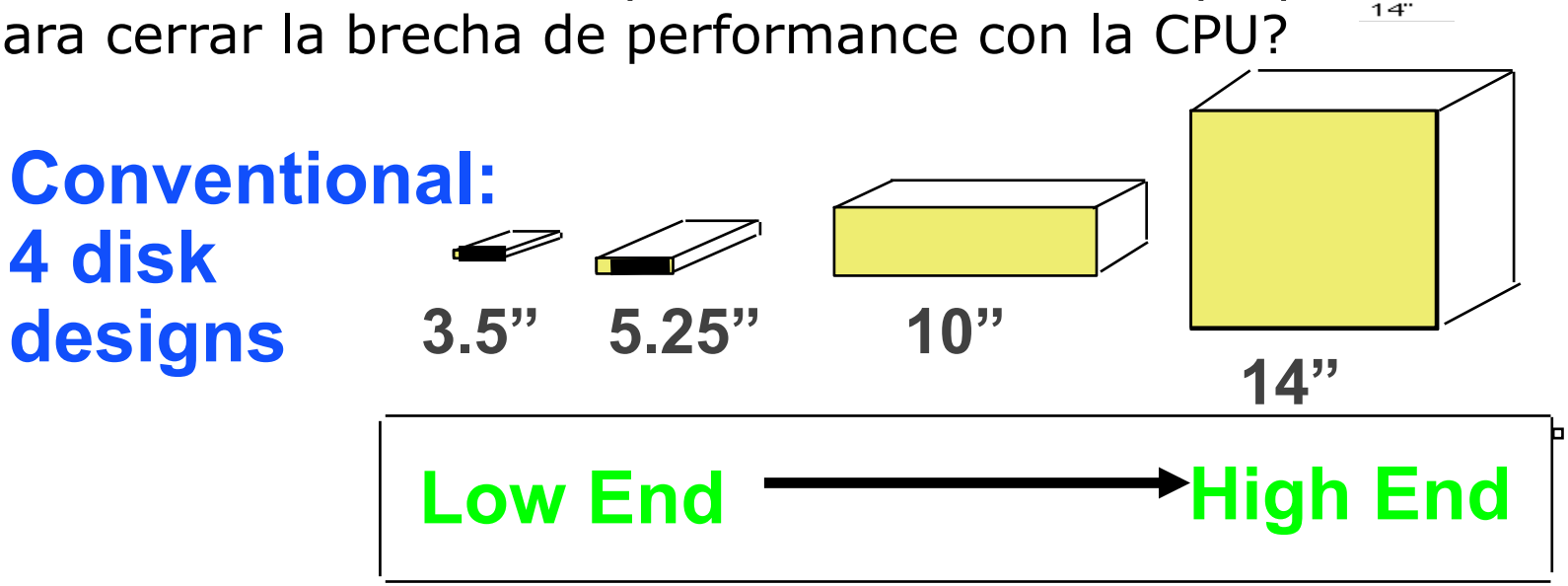
- 2000 IBM MicroDrive
  - 5mm x 35.5mm x 43.2mm
  - 1 GB, 3600 RPM, 5 MB/s, 15 ms seek

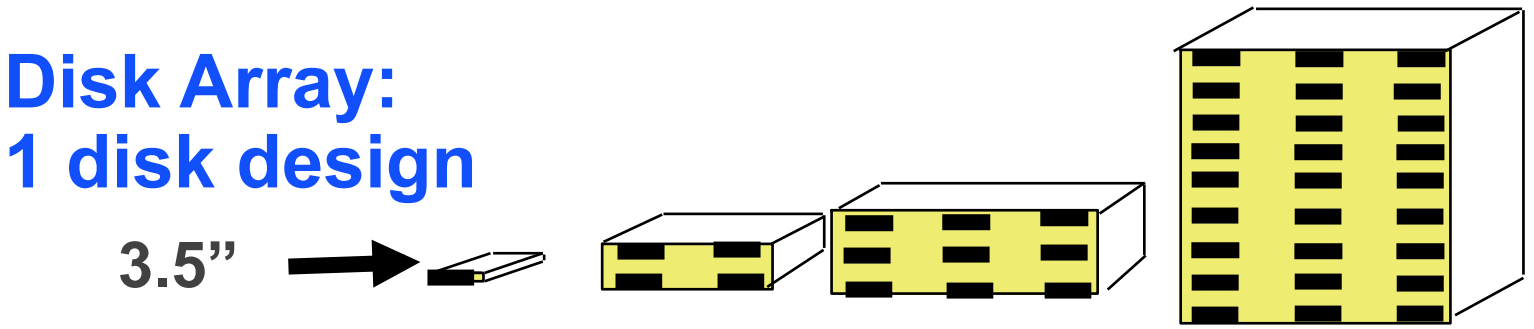
- 2010 Hitachi Travelstar Z7K320
  - 7mm x 70mm x 100mm, 95g
  - 320 (GB), 7200 RPM, 300 MB/s

- Aplicaciones:
  - Digital video camera
  - Portable digital audio
  - Portable digital video
  - Handheld navigation

Array de discos pequeños?

- Katz & Patterson, 1987: podemos usar discos pequeños para cerrar la brecha de performance con la CPU?





Array de discos pequeños?
(Discos de 1988)

	IBM 3390K	IBM 3.5" 0061	x70	
Capacity	20 GBytes	320 MBytes	22 GBytes	
Volume	97 cu. ft.	0.1 cu. ft.	11 cu. ft.	9X
Power	3 KW	11 W	1 KW	3X
Data Rate	15 MB/s	1.5 MB/s	105 MB/s	7X
I/O Rate	600 I/Os/s	55 I/Os/s	3900 IOs/s	6X
MTTF	250 KHrs	50 KHrs	??? Hrs	
Cost	$250K	$2K	$150K	

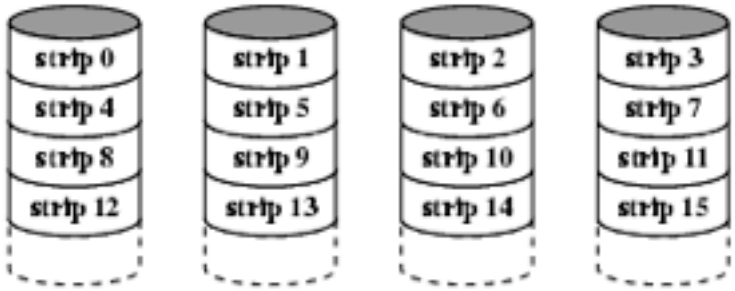
- Interesante… pero se reduce la confiabilidad
  - Confiabilidad de N discos = Confiabilidad de 1 Disco ÷ N
  - 50,000 Horas ÷ 70 discos = 700 horas -> MTTF baja de 6 años a 1 mes!
- Arrays (sin redundancia) muy poco confiables….

RAID

- Redundant Array of Independent (Inexpensive) Disks
- Conjunto de discos físicos vistos como un solo volumen lógico por el Sistema Operativo
  - Datos distribuidos en los discos físicos
  - Se usa capacidad redundante para almacenar información de paridad
  - Los archivos se “parten” en múltiples discos
- Redundancia -> alta disponibilidad
  - Disponibilidad: se mantiene el servicio al usuario aunque algunos componentes fallen
- Discos fallan…
  - Información se reconstruye a partir de la redundancia almacenada en el array
    - Penalización de capacidad debido a la redundancia
    - Penalización de Ancho de Banda para actualizar información redundante

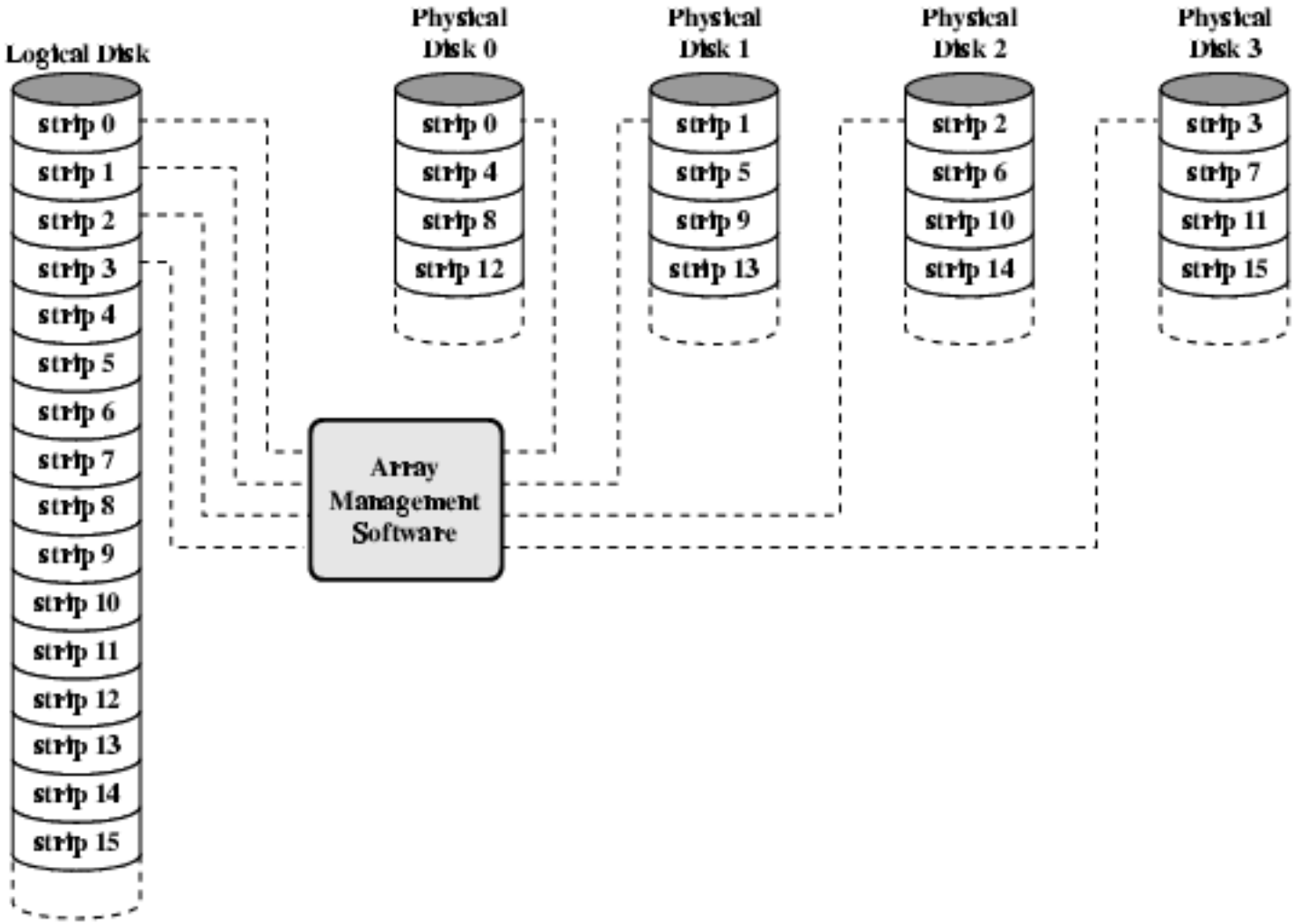
RAID 0

- No tiene redundancia
- Datos distribuidos en todos los discos
  - *Stripping* Round Robin
- Mejora velocidad
  - Es probable que requerimientos múltiples no estén en el mismo disco
  - Búsqueda en paralelo
  - Conjunto de datos repartido en muchos discos



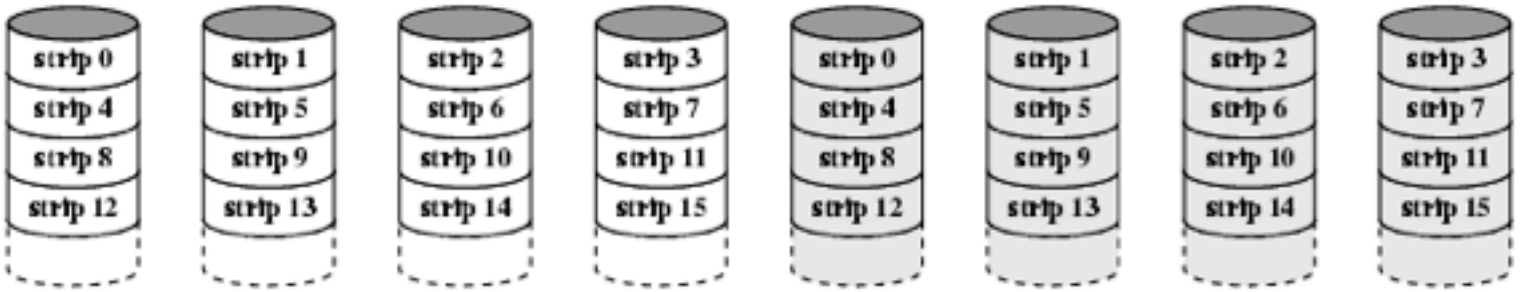
(a) RAID 0 (non-redundant)

Data Mapping en RAID 0



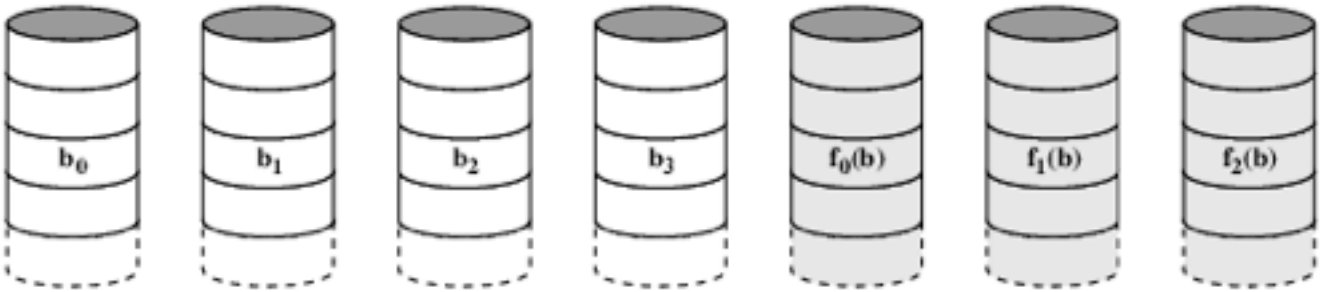
RAID 1

- RAID 1: Discos espejados (“mirrored”)
  - Los datos son distribuidos en los discos
  - 2 copias de cada “tira” de datos en discos separados
  - Se puede leer de cualquiera
  - Se escribe en ambos
    - Escritura Lógica = dos escrituras físicas
- Recuperación es simple
  - Cambiar disco fallado y re-mirror
  - No hay “down time”
- Caro!
- *Obs: la figura es RAID 0+1*



RAID 2

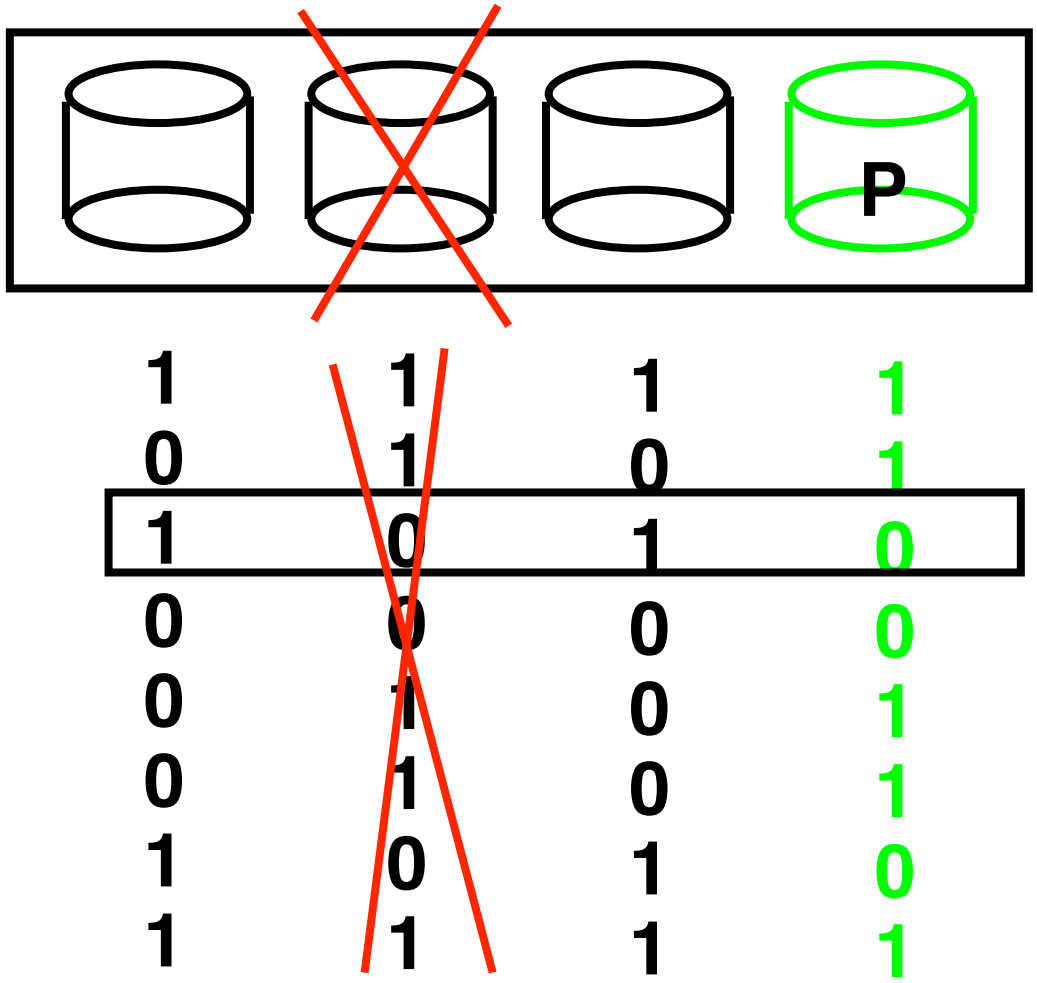
- Discos sincronizados
- *Tiras* muy pequeñas
  - Por ej. byte/word
- Corrección de errores se calcula usando bits correspondientes en los discos
- Múltiples discos de paridad almacenan código de Hamming para la corrección de errores
- Mucha redundancia
  - Caro
  - No se usa



(c) RAID 2 (redundancy through Hamming code)

RAID 3

- RAID 3: Sólo un disco redundante por array
  - Se calcula y almacena el bit de paridad en el disco P
  - Los datos del disco fallado se pueden reconstruir a partir de la información sobreviviente y el disco P
- Lógicamente se dispone de un disco de alta capacidad y alta tasa de transferencia
- Array ancho reduce costo/ capacidad, pero baja la disponibilidad

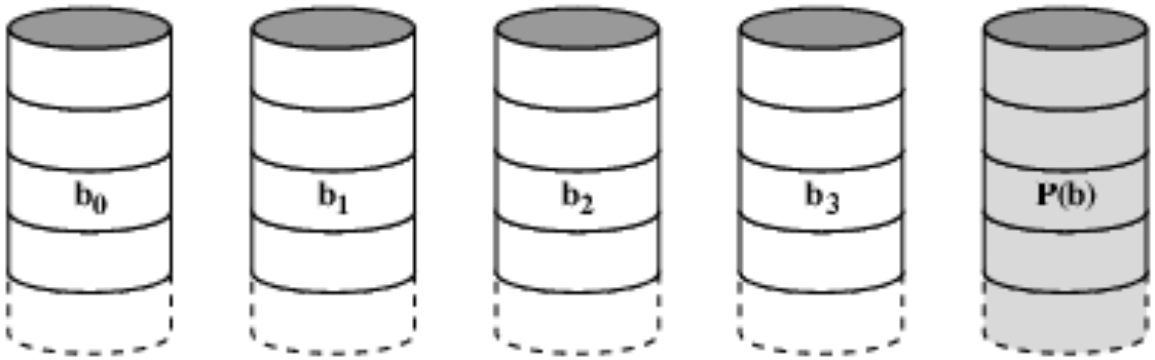


RAID 4

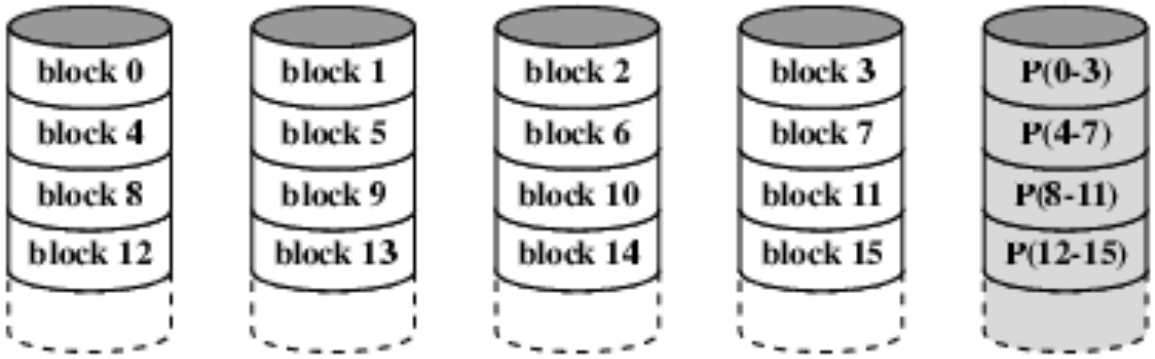
- En RAID 3 cada acceso necesita todos los discos: paridad de bit
- RAID 4 opera sobre bloques
- Cada disco puede operar en forma independiente
  - Bueno para “pequeñas lecturas”
  - Altas tasas de acceso de E/S

- Penalización de RAID 4: escrituras
  - Cuando se actualiza un bloque, se deben leer todos los bloques de una tira para actualizar el bit de paridad

RAID 3 & 4



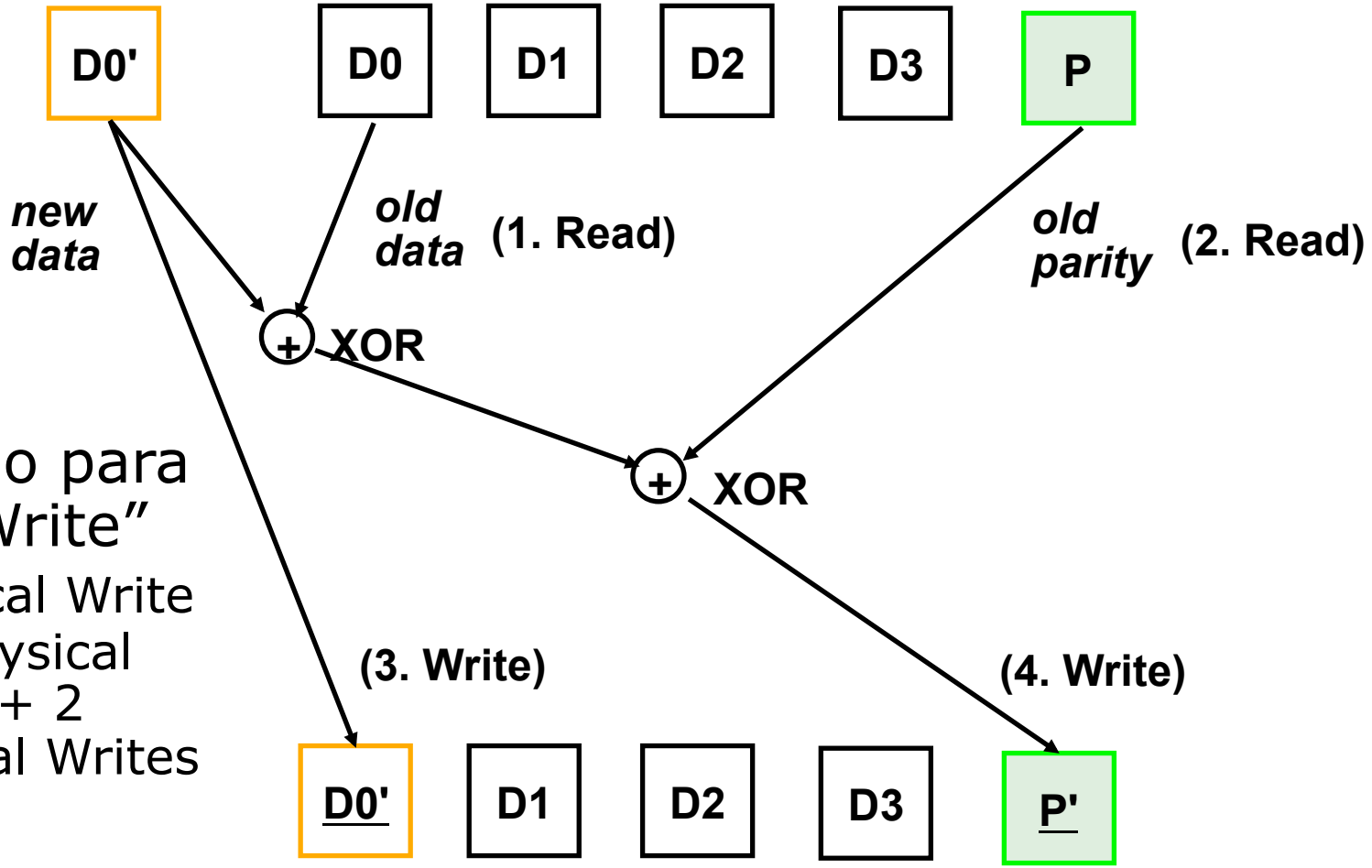
(d) RAID 3 (bit-interleaved parity)



(e) RAID 4 (block-level parity)

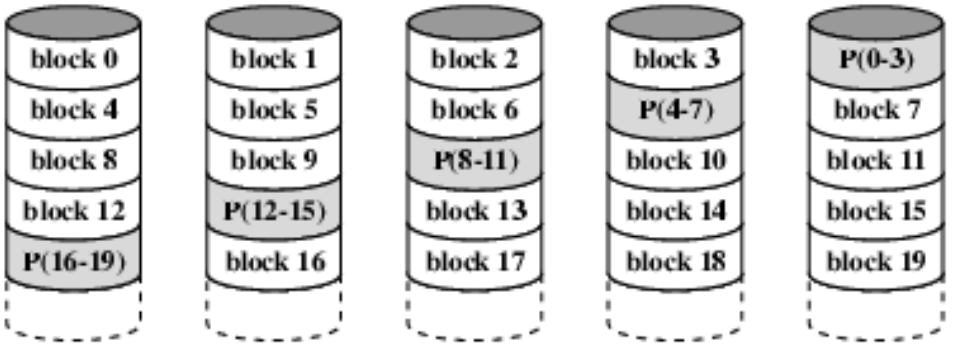
Small Writes

- Algoritmo para "Small Write"
  - 1 Logical Write = 2 Physical Reads + 2 Physical Writes

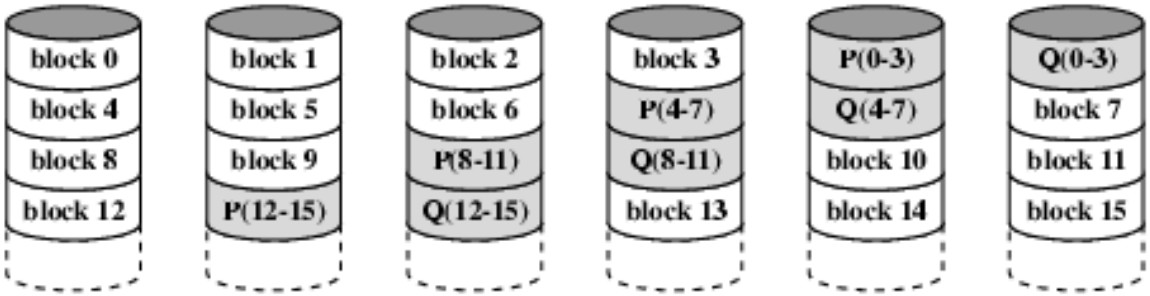


RAID 5 (y 6)

- Similar a RAID 4
- Mejora: paridad repartida en todos los discos
  - Paridad se escribe en round robin
- Elimina cuello de botella del disco de paridad
- RAID 6: doble paridad, no se usa



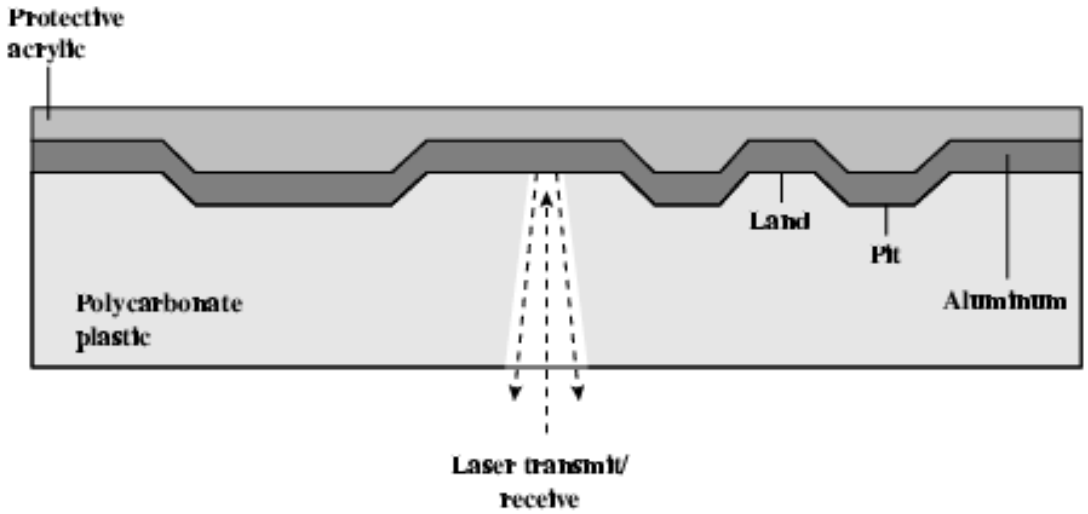
(f) RAID 5 (block-level distributed parity)



(g) RAID 6 (dual redundancy)

Almacenamiento Óptico: CD-ROM

- Originalmente para audio
- 650Mbytes, 70 minutos de audio
- Polycarbonato revestido de material espejado, usualmente aluminio
- Datos se almacenan como “huecos”
  - Se leen mediante un laser que se refleja en la superficie
  - Densidad y velocidad constante



Velocidades del CD-ROM

- El audio tiene una velocidad única
  - Velocidad lineal constante, 1.2 m/s
  - Track (espiral) tiene 5.27km de largo
  - 4391 segundos de audio = 73.2 minutos
- Otras velocidades se notan como múltiplos
  - Ej. 48x
  - Este número es el máximo

Otros Almacenamientos Ópticos

- CD-Recordable (CD-R)
  - WORM
  - Compatible con CD-ROM drives
- CD-RW
  - Borrable
  - Cada vez más barato
  - Normalmente compatible con CD-ROM drive
  - Funciona porque el material (una aleación) toma diferentes estados sólidos al calentarse a diferentes temperaturas. La luz se refleja en mayor o menor medida dependiendo del estado sólido del material.

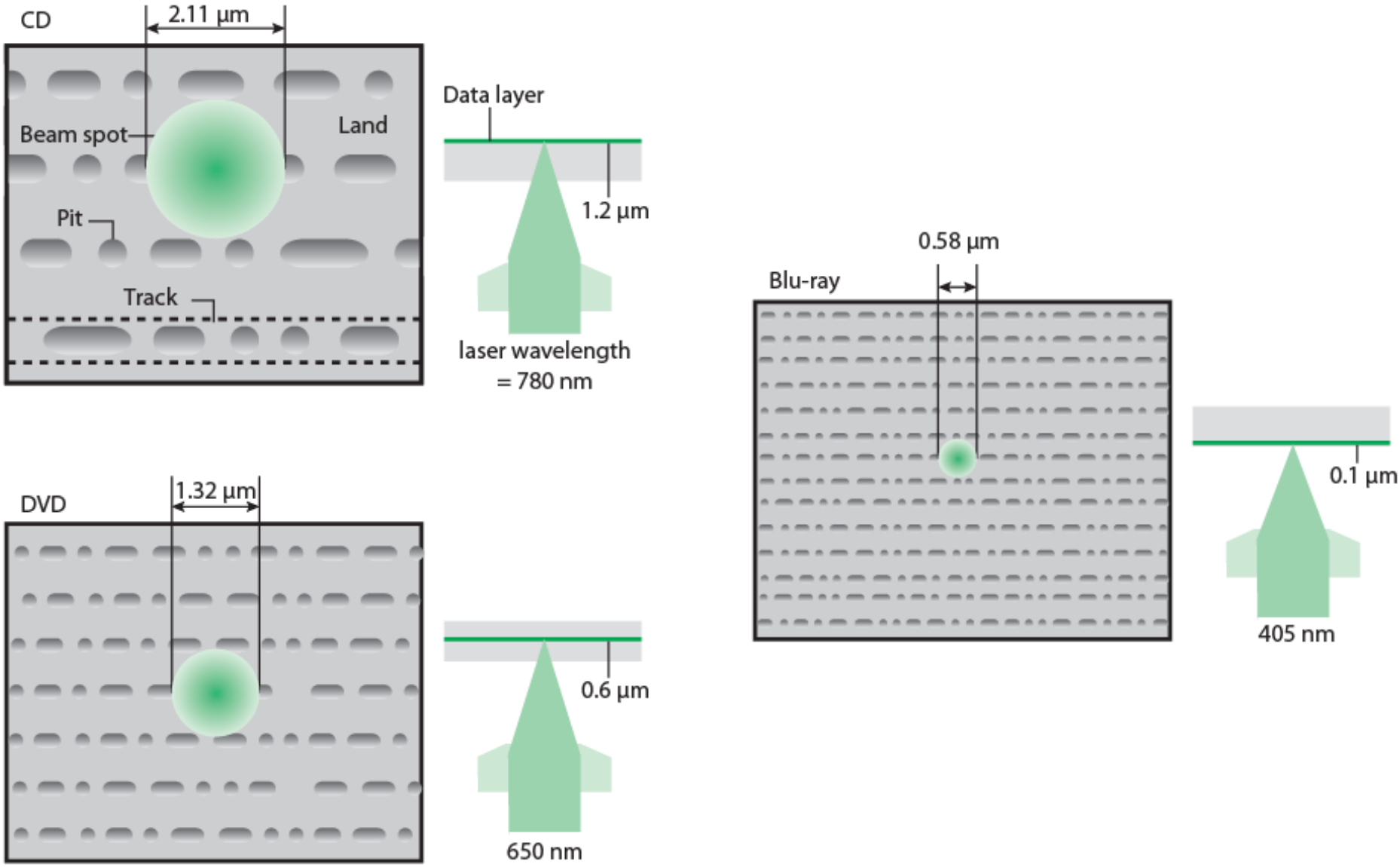
DVD

- DVD - Digital Video Disk
  - Multi-capa
  - Alta capacidad (4.7G por capa)
  - Película completa en un disco
    - Compresión MPEG

Discos Ópticos de Alta Definición (HD)

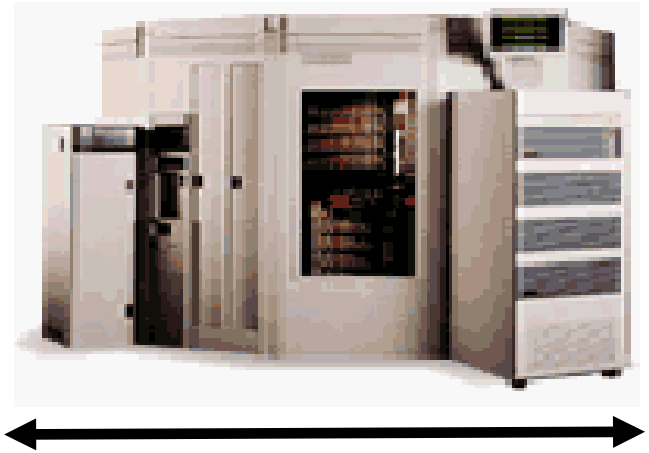
- Diseñados para videos de alta definición
- Mayor capacidad que el DVD
  - Laser de longitud de onda más corta
    - Rango azul-violeta
  - Hoyos más pequeños
- HD-DVD
  - 15GB, un solo lado, una sola capa
- Blue-ray
  - Capa de datos más cercana al laser
    - Foco más preciso, menos distorsión, hoyos más pequeños
  - 25GB en una sola capa
  - Disponible en “read only” (BD-ROM), “recordable once” (BR-R) y “re-recordable” (BR-RE)

Memoria Óptica: características



Cinta Magnética

- Acceso serial
- Lento
- Barato
- Respaldo, archivo



- Ejemplo
  - Robot de almacenamiento StorageTek Powderhorn 9310
  - 6000 x 50 GB 9830 cintas = 300 TBytes (sin comprimir)
    - Obs: todos los libros de la Biblioteca del Congreso USA son 30 TB ASCII
  - Puede cambiar 450 cintas por hora
  - 1.7 a 7.7 Mbyte/seg por lector, hasta 10 lectores