

CAPÍTULO 3

MEDIOS DE ALMACENAMIENTO DE DATOS

3.1 CLASIFICACION Y APLICACIONES

3.1.1 Clasificación

Los medios masivos de almacenamiento de datos se caracterizan por dos importantes propiedades: cantidad de información y transportabilidad. Las aplicaciones que aprovechan la posibilidad de almacenamiento de gran cantidad de información son, entre otras, el sistema operativo, las bases de datos, los archivos temporarios, las aplicaciones multimedia (imagen y sonido), etc. Los medios de almacenamiento masivo lo logran a un bajo costo por unidad de información, en un espacio reducido. La transportabilidad permite la distribución y actualización de software, las copias de seguridad, los archivos permanentes, etc.

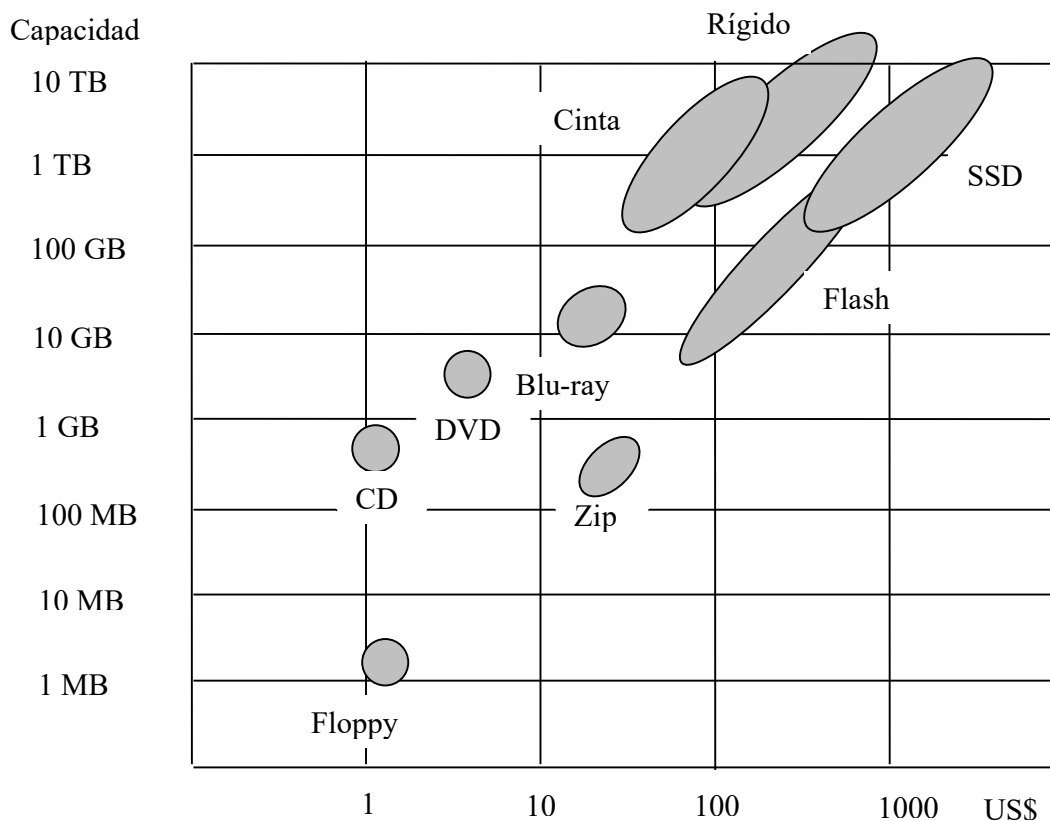


Figura 3.1.1 - Relación Costo/Capacidad

Los medios de almacenamiento masivo en uso en la actualidad son:

- discos rígidos
- cintas
- discos ópticos (CD, DVD, BLU-RAY)
- memorias flash (USB/SD)
- discos de estado sólido (SSD)

3.1.2 Aplicaciones

Cada uno de ellos tiene una aplicación específica que hace que estas distintas tecnologías puedan coexistir en el mercado.

Los discos rígidos se utilizan principalmente como *disco del sistema* cuando se emplean sistemas operativos en disco (ej. DOS, RSX, VMS, UNIX, NT, etc.). También para tener acceso rápido a bases de datos, como área de almacenamiento temporario en procesos de compilación y para alojar tareas inactivas en ambientes multitarea (checkpointing).

Los menores costos de las unidades SSD están cambiando el estándar para soporte del sistema operativo. Dada su mayor velocidad, comparada con la de los discos rígidos, están siendo empleados por la mayoría de los usuarios corporativos, nuevas laptops, gamers e incluso usuarios particulares de desktops.

Las cintas se usan para respaldar discos rígidos, como medio de almacenamiento de archivos históricos o bases de datos secuenciales muy grandes y como medio de transporte y distribución de software.

Los discos flexibles se usaban en sistemas de bajo costo como transporte y distribución de software y como respaldo de archivos de tamaño reducido. Actualmente en desuso.

Los discos Zip, una mezcla de disco flexible con tecnología de disco rígido, llegaron a tener capacidades de 250 MB en un medio removible. Actualmente en desuso.

Los discos ópticos (CD-ROM) se utilizan (cada vez menos) como medio de distribución de música y drivers. Existen versiones CD-RW.

Los DVD y Blu-Ray compiten con las cintas en su capacidad de almacenamiento y se usan (cada vez menos) como backup de datos, transporte de SW y distribución de multimedia.

3.2 INTERFAZ CON EL PROCESADOR

3.2.1 Esquema general

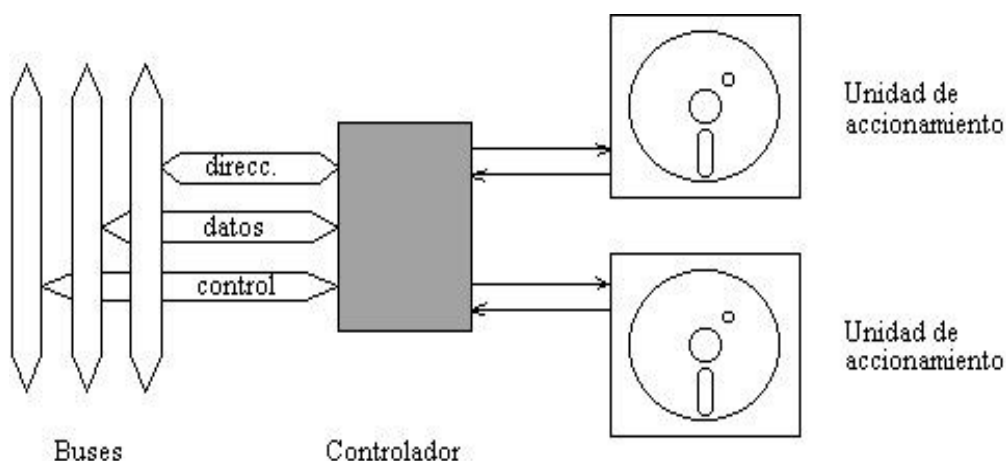


Figura 3.2.1 - Controlador de periféricos - Diagrama en bloques

En este esquema, el controlador puede manejar más de una unidad de accionamiento. Existen también las unidades con controlador incorporado, todos unidos a un bus, eliminando el cuello de botella del controlador cuando hay unidades múltiples.

El controlador contiene la lógica de DMA, memoria incorporada (caché) para compensar las diferencias de velocidad entre el medio magnético y el bus, lógica de interrupciones, conversión serie/paralelo, manejo de errores, etc. Actualmente están controlados por un microprocesador de alta velocidad o circuitos dedicados y realizan funciones de optimización.

3.2.2 Tipos de controladores/interfases

IDE: Integrated Drive Electronics.

Es un controlador cuya característica principal es que su lógica se encuentra integrada en la unidad de accionamiento. Una ventaja de este método radica en que los errores se detectan y corrigen en forma local, sin afectar la comunicación con el procesador. Su performance se sitúa entre los 4 y los 8 MB/s, cuando se encuentra conectado al bus ISA. Una limitación es que se pueden conectar sólo dos discos IDE.

La versión mejorada, Enhanced IDE, soporta hasta 4 periféricos, dos en el controlador primario y dos en el secundario. En cada controlador se debe seleccionar un dispositivo como “master” y el otro como “slave”, usualmente mediante un puente en el dispositivo.



Figura 3.2.2 – Conector de la interfaz IDE

UATA/UDMA:

En las placas madre que lo permitían (a partir de 1999) se accedía a los discos rígidos por acceso directo a memoria, a la velocidad de 33 MB/s (UDMA 33), 66 MB/s, 100 MB/s y 133 MB/s (UDMA 133) con lo que se alcanzó el límite del bus PCI (133 MHz). La conexión a partir de ATA66 requiere un cable plano especial de 80 conductores que posee los mismos conectores de 40 contactos y los otros están a masa para apantallar las señales y reducir la interferencia a esta elevada frecuencia.

SCSI: Small Computer System Interface.

En este caso existe un controlador en el bus del sistema, al que pueden conectarse hasta siete dispositivos, que pueden ser discos, cintas, CD-ROMs, aunque sólo uno de ellos se puede comunicar a la vez con el procesador. El bus de datos es de 8 bits, alcanzando velocidades de 5 MB/s. Se conecta sobre el bus ISA. La versión mejorada, SCSI-2 aumenta el ancho del bus de datos a 16 o 32 bits, pudiendo lograrse velocidades de hasta 40 MB/s.



Figura 3.2.3 – Conector de la interfaz SCSI

SAS: Serial Attached SCSI.

Es una interfaz de transferencia de datos en serie, sucesor del Small Computer System Interface (SCSI) paralelo, aunque sigue utilizando comandos SCSI para interactuar con los dispositivos SAS. Aumenta la velocidad y permite la conexión y desconexión de forma rápida. El último estándar vigente, desde 2017, es SAS-4 que opera a 22,5 Gb/s.



Figura 3.2.4 – Conector de la interfaz SAS

SATA:

Desde 2004 se utiliza la conexión Serial ATA (SATA). Se trata de una comunicación serie que opera a 1,5 Gbps, permitiendo velocidades de transferencia pico de 150 MB/s. El conector de datos es más pequeño, ya que sólo posee siete conductores, dos para transmisión, dos para recepción y tres GND y puede tener una longitud de varios metros. SATA II, una actualización de este estándar, opera a 3 Gbps (300 MB/s). En 2010 comienza a difundirse SATA III, que posee una especificación de 6 Gbps (600 MB/s). Las alimentaciones de 5 V y 12 V también poseen un conector diferente al de los antiguos IDE.

Los discos rígidos mecánicos rara vez alcanzan los 150 MB/s de velocidad de transferencia, por lo que los nuevos estándares de SATA no representan un beneficio en la performance. Sin embargo, los dispositivos de estado sólido (SSD) si aprovechan estos incrementos de velocidad de transferencia.



Figura 3.2.5 – Conector de la interfaz SATA

La especificación AHCI define a los fabricantes de hardware una estructura de memoria para intercambiar datos entre la memoria del sistema host y los dispositivos de almacenamiento conectados. AHCI proporciona a los desarrolladores de software y diseñadores de hardware un método estándar para detectar, configurar y programar adaptadores SATA/AHCI. AHCI es independiente del estándar SATA 3Gb/s, si bien expone las capacidades avanzadas de SATA (como NCQ y Hot Swap) de tal manera que los sistemas puedan utilizarlos.

NCQ es una forma de recuperar los datos según la facilidad de obtenerlos de un disco en rotación, como se ve en la figura, reduciendo los tiempos totales, aunque no se respeta el orden

original de las solicitudes. Se entenderá mejor cuando veamos el funcionamiento de una unidad de discos rígidos.

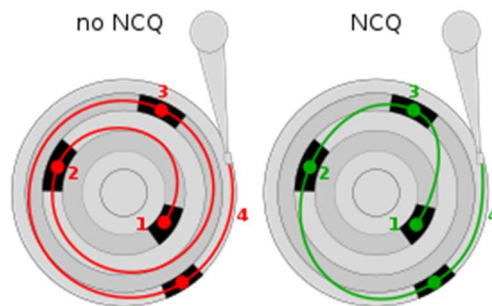


Figura 3.2.6 – Comparativa sin NCQ vs con NCQ

Hot swap está permitido por diseño en SATA. Esta interfaz soporta la conexión y desconexión en uso, es decir con la alimentación aplicada. Esto es muy útil en sistemas redundantes, como se explicará en el párrafo 3.9, al final de este capítulo.

M.2:

M.2 anteriormente conocido como NGFF (por las siglas en inglés Next Generation Form Factor o factor de forma de nueva generación) es una especificación para tarjetas de expansión interna de computadoras. Las especificaciones de M.2 son flexibles, lo que permite diferentes longitudes y anchos de módulos para las unidades de estado sólido en general y para su uso en dispositivos más pequeños como ultrabooks o tablets.

La interfaz del bus de datos que está detrás del conector M.2 es una PCI Express 3.0 (hasta cuatro slots PCI Express), Serial ATA 3.0 y USB 3.0 (un puerto lógico individual por cada uno de los dos últimos). Está a merced del fabricante del dispositivo o del puerto M.2 elegir qué interfaces se soportarán, dependiendo del nivel deseado tanto del dispositivo como del receptor. El conector M.2 puede presentar distintas muescas de codificación que denotan tanto diferentes capacidades como propósitos, evitando así el uso de módulos M.2 en dispositivos incompatibles.

Para dispositivos SSD que usen M.2 a través de PCI Express 3.0 4x, la máxima velocidad teórica es de 3938 MB/s. Si, en cambio, el puerto soporta PCI Express 4.0, la velocidad asciende a 7877 MB/s. En 2021 los SSD comerciales más rápidos pueden operar cerca de 7000 MB/s.



Figura 3.2.7 – Conector de la interfaz M.2

A continuación, una comparativa de las distintas interfases (de Wikipedia, editado).

Name	Raw data rate	Data rate	Max. cable length	Power provided	Devices per channel
eSATA	6 Gbit/s	600 MB/s	2 m	No	1 (15 with a port multiplier)
eSATAp	6 Gbit/s	600 MB/s	1 m	5 V, 12 V	
SATA Express	16 Gbit/s	1.97 GB/s	1 m	No	
SATA revision 3.0	6 Gbit/s	600 MB/s			
SATA revision 2.0	3 Gbit/s	300 MB/s			
SATA revision 1.0	1.5 Gbit/s	150 MB/s			1
PATA (IDE) 133	1.064 Gbit/s	133.3 MB/s	0.46 m	5 V (only 2.5-inch drive 44-pin conn)	2
SAS-4	22.5 Gbit/s	2.25 GB/s	10 m	Backplane conn only	1 (> 65k with expanders)
SAS-3	12 Gbit/s	1.2 GB/s			
SAS-2	6 Gbit/s	600 MB/s			
SAS-1	3 Gbit/s	300 MB/s			
IEEE 1394 (FireWire) 3200	3.144 Gbit/s	393 MB/s	100 m	15 W, 12–25 V	63 (with a hub)
IEEE 1394 (FireWire) 800	786 Mbit/s	98.25 MB/s	100 m		
IEEE 1394 (FireWire) 400	393 Mbit/s	49.13 MB/s	4.5 m		
USB 3.2 (Gen 2x2)	20 Gbit/s	2.44 GB/	1 m	Yes 100 W, 5, 12 or 20 V	127 (with a hub)
USB 3.1 (Gen 2)	10 Gbit/s	1.22 GB/s	1 m	100 W, 5, 12 or 20 V	127 (with a hub)
USB 3.0 (USB 3.2, Gen 1)	5 Gbit/s	610 MB/s	2 m	4.5 W, 5 V	
USB 2.0	480 Mbit/s	58 MB/s	5 m	2.5 W, 5 V	
USB 1.1	12 Mbit/s	1.5 MB/s	3 m	Yes	
SCSI Ultra-320	2.56 Gbit/s	320 MB/s	12 m	Only with SCA Backplane	15 excl. host bus adapter/host
PCI Express 1.0 (×4 link)	10 Gbit/s	1 GB/s	n/a	yes	n/a
PCI Express 2.0 (×4 link)	20 Gbit/s	2 GB/s	n/a	yes	n/a
PCI Express 3.0 (×4 link)	32 Gbit/s	3.938 GB/s	n/a	yes	n/a
PCI Express 4.0 (×4 link)	64 Gbit/s	7.8775 GB/s	n/a	yes	n/a

Figura 3.2.8 – Evolución de los estándares

3.3 METODOS DE GRABACION MAGNETICA

3.3.1 Introducción

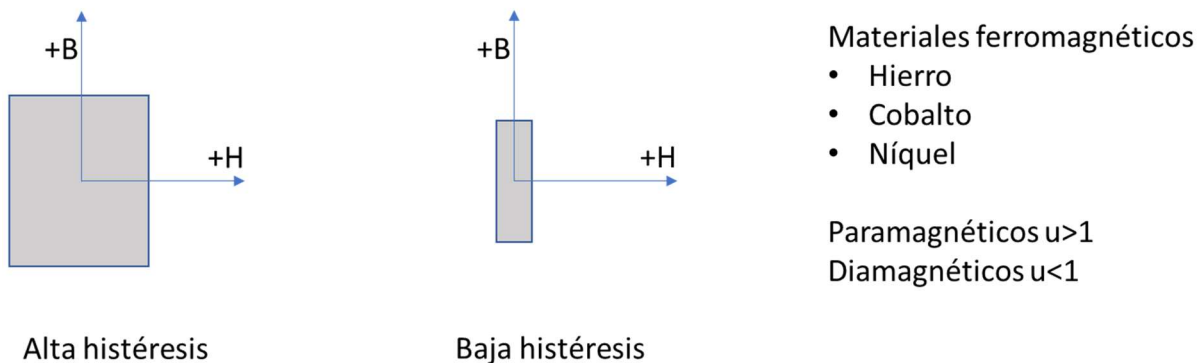


Figura 3.3.1 – Tipos de materiales

Los materiales de alta histéresis se usan en imanes permanentes en general y en nuestro caso los aprovecharemos para retener la información, como se ve en la figura 3.3.2.

En cambio, los materiales de baja histéresis se usan en relés, motores y transformadores y en nuestro caso los usamos en el cabezal de grabación como se ve en la figura 3.3.3.

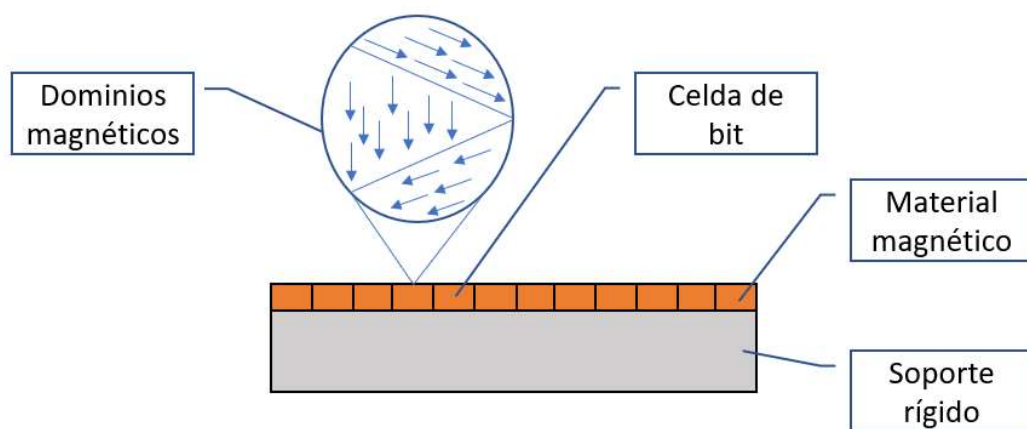


Figura 3.3.2 – Medios de almacenamiento magnéticos

El material donde se almacenan los datos es una fina capa de Níquel, un material ferromagnético con alta histéresis. Los minúsculos dominios magnéticos son grupos de dipolos que se orientan en función del campo aplicado y luego de retirado el mismo se mantienen con la misma orientación.

Las celdas de bit son las mínimas áreas que pueden orientarse sin influir en las vecinas.

El material magnético yace sobre una superficie sólida, en general de aluminio, que da el nombre a uno de los medios conocidos: el disco rígido.

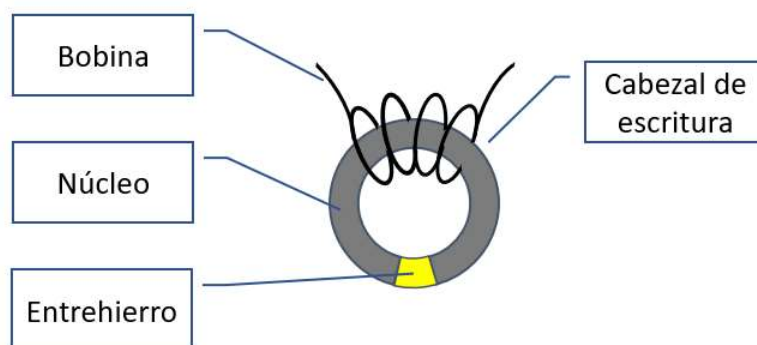


Figura 3.3.3 – Cabezal de grabación magnética

El cabezal de escritura consiste en un núcleo de material ferromagnético de baja histéresis sobre el cual esté arrollado un bobinado. Cuando se aplica corriente eléctrica al bobinado, el núcleo se magnetiza como un electroimán. Cuando se retira la corriente, el mismo se desmagnetiza.

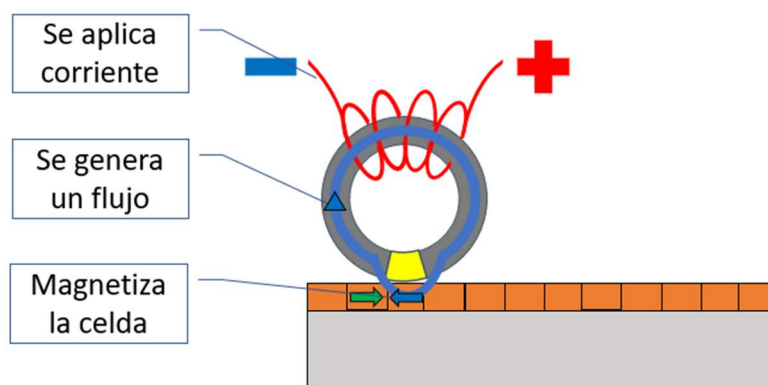


Figura 3.3.4 – Proceso de grabación

Como se ve en la figura, las líneas de flujo magnético se desvían del entrehierro para magnetizar la superficie de grabación. Este efecto se ve favorecido por la utilización de un material diamagnético (shim) en el entrehierro, el que, al llenar ese espacio, impide la acumulación de partículas extrañas.

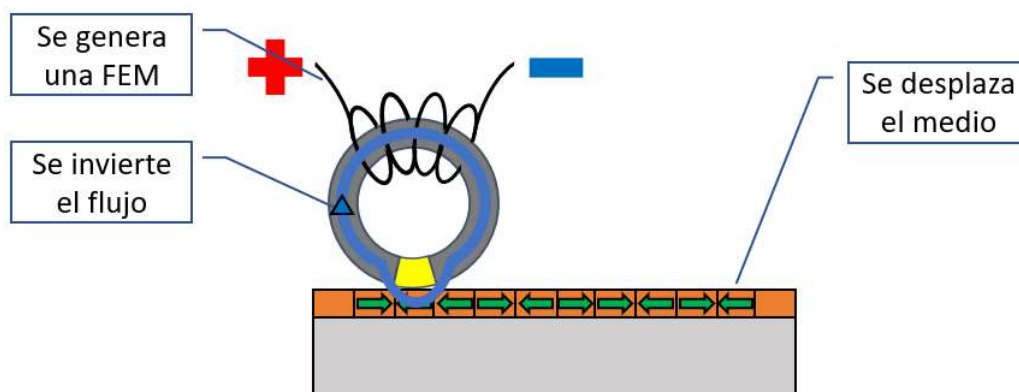


Figura 3.3.5 – Proceso de lectura

El mismo cabezal se puede utilizar para la lectura. Como se ve en la figura, cuando el medio se mueve debajo del entrehierro, las distintas direcciones del campo magnético grabado generan variaciones de flujo en el núcleo. A su vez, estas variaciones de flujo generan FEM en la bobina. Esta FEM se utiliza, luego de ser procesada y amplificada, para reconstruir la información previamente grabada en el medio.

En los discos flexibles y cintas, la cabeza roza contra la superficie del medio magnético; en los discos rígidos, la misma está separada una pequeña distancia, llamada “altura de vuelo”, explicada en detalle en 3.5.

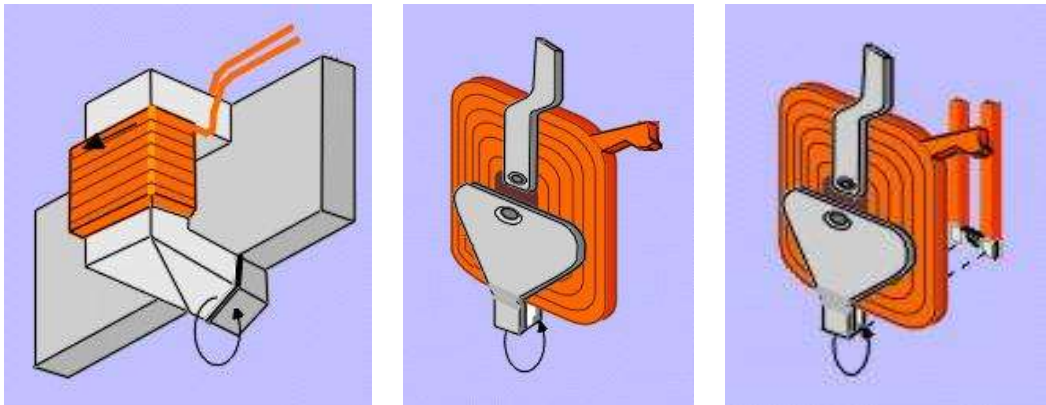


Figura 3.3.6 – Cabezales – A: Inductivo – B: Thin film – C: MR/GMR/TMR

En la actualidad, los discos rígidos poseen un cabezal de lectura separado del de escritura. El mismo fue evolucionando desde el efecto Hall, el MR (magneto-resistivo), pasando por el GMR (magneto-resistivo-gigante) y finalmente el TMR (magneto-resistivo-túnel).

La variación de tensión que se obtiene en estos sensores es mucho mayor que la fem en una bobina, para una transición de flujo dada, lo que permite detectar los menores efectos magnéticos que se dan al reducir las celdas y aumentar la densidad de grabación.

Efecto Hall

Cuando por un material circula una corriente eléctrica y estando este mismo material en el seno de un campo magnético, se comprueba que aparece una fuerza en los portadores de carga que los reagrupa dentro del material; esto es, los portadores de carga se desvían y agrupan a un lado del material conductor o semiconductor, apareciendo así una variación de resistividad.

El efecto magnetorresistivo se basa en la propiedad de un dispositivo susceptible magnético de película delgada hecho de un material de aleación ferromagnética de variar su resistencia en la presencia de un campo magnético o magnetorresistencia.

Este sensor tiene un coeficiente de temperatura pequeño y un rendimiento estable lo que lo hace apropiado para los discos rígidos. A diferencia del sensor de semiconductor, es más sensible a la dirección del campo magnético.

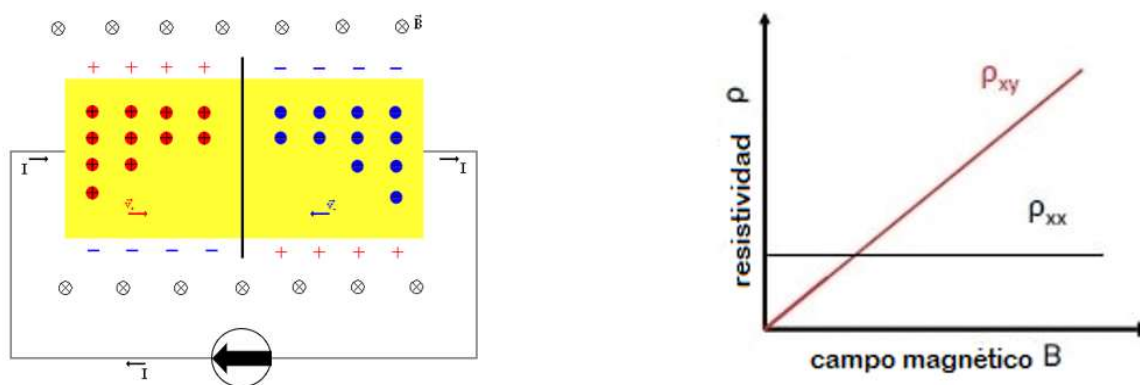


Figura 3.3.7 – Efecto Hall

Grabación perpendicular

Desde de 2008 ya se utiliza el método de magnetización perpendicular, que logra aproximadamente el doble de densidad de grabación.

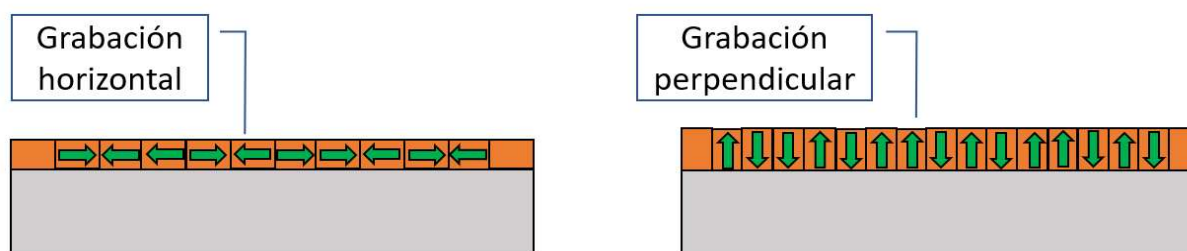


Figura 3.3.8 – Grabación horizontal vs grabación perpendicular

En la figura se puede ver la disposición del cabezal de grabación.

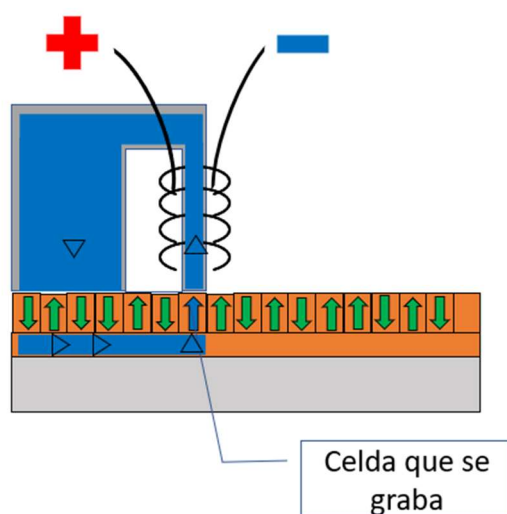


Figura 3.3.9 – Grabación perpendicular

Sólo una celda se graba en forma perpendicular. Las celdas por las que retornan las líneas de flujo no se ven afectadas porque al repartirse el flujo entre una gran superficie, el mismo es muchísimo menor al que se usó para grabarlas.

Codificación de los datos

Para codificar los datos binarios en forma de cambios o transiciones de campo magnético, se idearon distintos métodos. Comentaremos los tres principales.

3.3.2 Método FM (modulación de frecuencia)

- Transición (T) en todos los bordes de celda
- Para grabar "1": transición (T) en el centro de la celda
- Para grabar "0": sin transición (N) en el centro de la celda

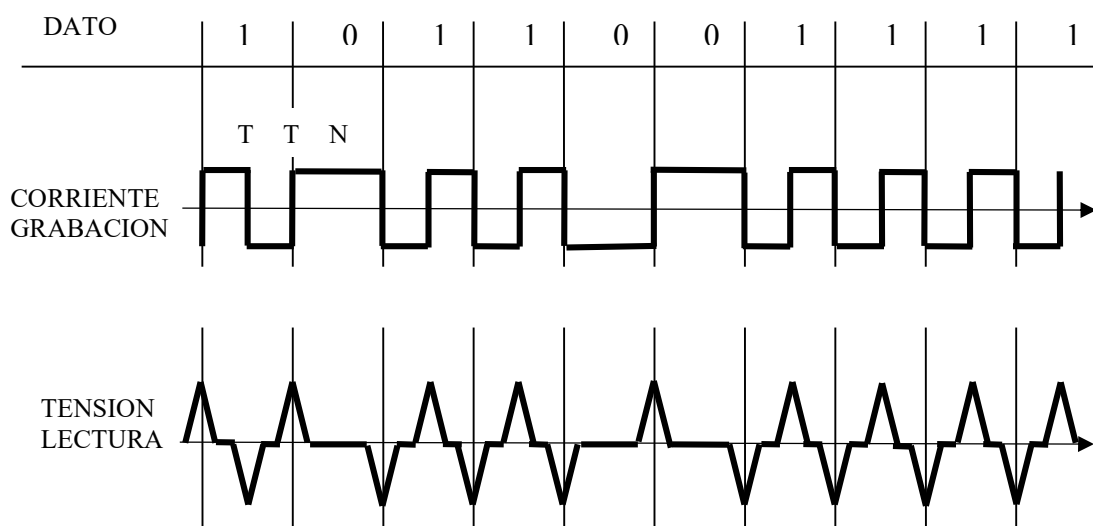


Figura 3.3.10 – Método de grabación FM

Vemos que al grabar varios "1" se obtiene una "frecuencia" doble que al grabar varios "0".

3.3.3 Método MFM (FM modificado o Miller)

- Para grabar "1": transición (T) en el centro de la celda
- Para grabar "0": transición (T) en el borde de la celda, sólo si el dato anterior era "0".

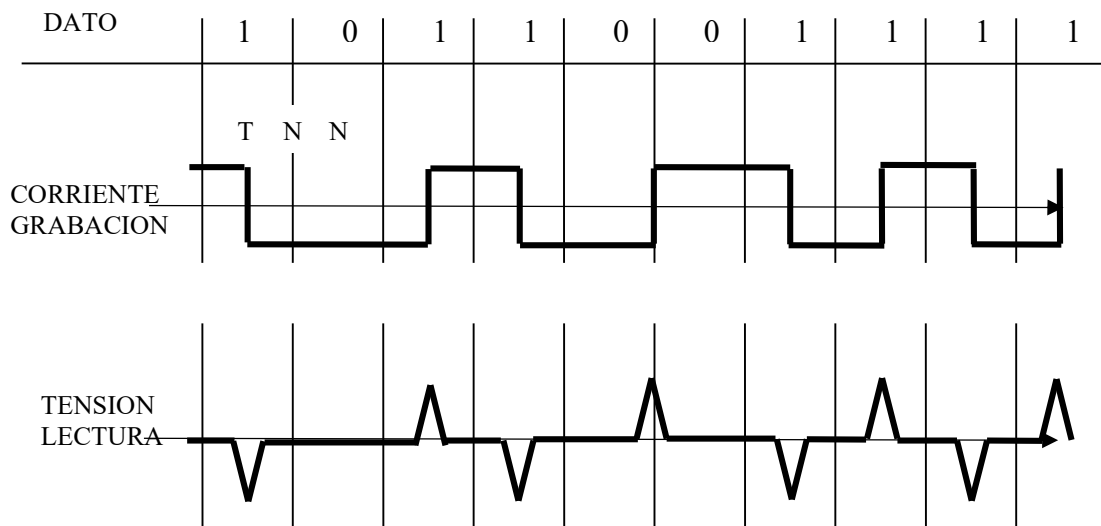


Figura 3.3.11 – Método de grabación MFM

Comparando ambos métodos, vemos que en MFM ningún dato requiere más de una transición por celda. Por el contrario, en FM se necesitan dos transiciones para grabar un "1". Teniendo en cuenta que, para un mismo medio magnético, la cantidad de información que se puede almacenar depende de la densidad de grabación (transiciones por unidad de longitud) y del método de grabación (bits por transición), con MFM se puede guardar el doble de información en la misma longitud que la que se almacenaría con FM. Ambos métodos permiten recuperar la señal de reloj, que determina el momento oportuno para leer la información de la celda.

3.3.4 Método RLL (Run Length Limited)

Aprovechando implementaciones de hardware más complejas, se pueden utilizar técnicas más elaboradas como RLL (Run Length Limited) y ARLL (Advanced RLL).

En RLL 2,7 (inventado por IBM en 1979) la codificación se basa en una tabla que traduce las distintas combinaciones de unos y ceros que se pretende grabar a una secuencia especial de transiciones de flujo. Con esto se logra mejorar la relación entre información y señal de reloj, aumentando en un 50 % la densidad de grabación respecto de MFM.

RLL n,k significa que la mínima cantidad de N seguidos entre dos T será de n y la máxima será k. La mínima determina la densidad de grabación y la máxima, la facilidad de recuperar el reloj. En la figura, que corresponde a Seagate ST11R IBM, se resaltan $n=2$ y $k=7$ para RLL 2,7.

Conjuntos de datos	Transiciones de flujo
10	NTNN
11	TNNN
000	NNNTNN
010	TNNTNN
011	NNTNNN
0010	NNTNNTNN
0011	NNNNTNNN

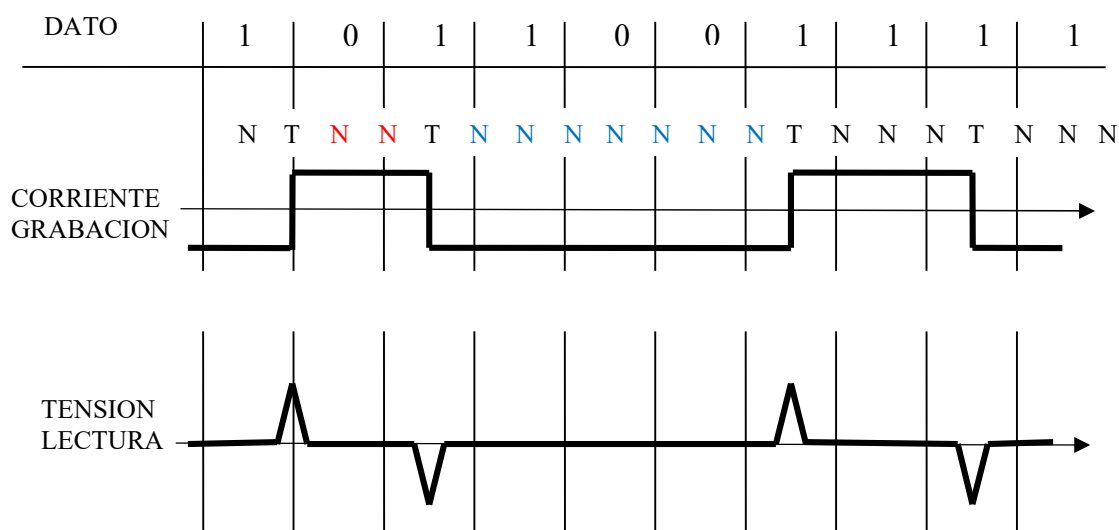


Figura 3.3.12 – Método de grabación RLL 2,7

En la figura se puede ver como se grabaría la misma información de los casos anteriores, con RLL 2,7. Véase que cualquier dato antes o después de esta secuencia respeta la codificación.

ARLL es un nombre que se da al RLL 3,9. Éste tiene una densidad de grabación que duplica al MFM ($n=3$), sin embargo, por tener menor información de reloj ($k=9$) tenía problemas de confiabilidad, lo que hizo que fuera abandonado, volviendo al RLL 2,7.

Nota 1: en la actualidad, los discos de alta capacidad codificarían en RLL 1,7, que tiene una menor densidad de grabación que el RLL 2,7 pero mayor confiabilidad. Ahora las altas densidades se logran con la mejora de los materiales y las tecnologías de fabricación.

Nota 2: estas técnicas de codificación de señales se utilizan también en comunicaciones, como emisor de infrarrojos, etc.

Referencias:

Runlength-Limited Sequences. December 1990 - Proceedings of the IEEE 78(11):1745 - 1759

Puede consultar el documento completo en https://www.researchgate.net/publication/2984369_Runlength-Limited_Sequences

3.4 ORGANIZACION Y CARACTERISTICAS

3.4.1 Organización física

Si se considera la recta perpendicular al plato que pasa por las cabezas de lectura/escritura, se observa que la misma genera una superficie cilíndrica al hacerla girar alrededor del eje de rotación. Para cada posición de las cabezas, corresponde un cilindro y la forma de numerarlos es desde el más exterior (cilindro 0) hacia adentro.

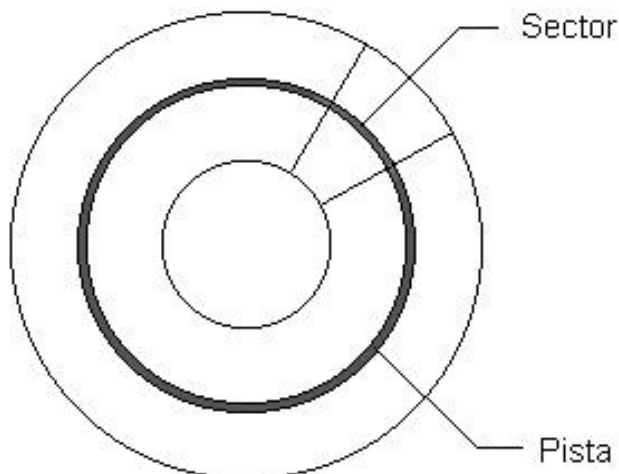


Figura 3.4.1 – Organización física del disco

La intersección de cada cilindro con cada superficie magnética, determina una circunferencia, llamada pista (track). Existen dos pistas por cada placa: una superior y otra inferior.

A su vez, cada pista se divide en arcos determinados por ángulos iguales. Son los sectores, que se numeran desde 1 hasta el último. La mínima unidad direccionable en el disco es el sector, que usualmente contiene 512 bytes de datos.

3.4.2 Formateo

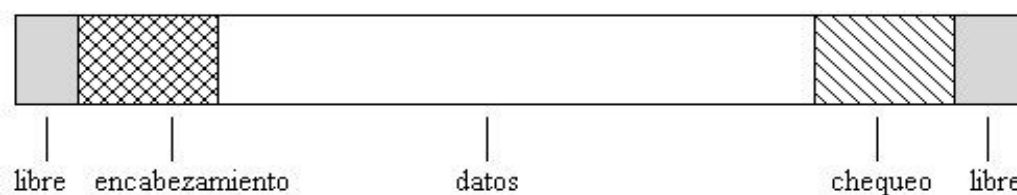
Es el proceso de inicialización del medio de almacenamiento para la unidad de accionamiento. El formateo determina la organización del medio (cantidad de pistas, sectores, bytes por sector) y por lo tanto también su capacidad.

Por ejemplo: un disco rígido de 1985 tenía 4 platos, 850 cilindros, 26 sectores y 512 bytes por sector tendrá una capacidad de:

$$4 \text{ platos} \times 2 \frac{\text{cabezas}}{\text{plato}} \times 850 \frac{\text{pista}}{\text{cabeza}} \times 26 \frac{\text{sectores}}{\text{pista}} \times 512 \frac{\text{bytes}}{\text{sector}} = 90.521.600 \text{ bytes}$$

Para una unidad de cinta con 22 pistas, 8000 sectores y 512 bytes por sector se obtendrá una capacidad de 90.112.000 bytes, es decir aproximadamente 90 Mb.

En cada sector, a su vez, se graba información de control aparte de los datos.

**Figura 3.4.2 – Formateo**

3.4.2.1 Encabezamiento (*header*)

El encabezamiento contiene la identificación de la pista la cabeza y el sector. Esto es para que luego del proceso de búsqueda se pueda confirmar el posicionamiento del sistema mecánico. En caso de no coincidir, se informa “error de búsqueda” (seek error).

También posee una secuencia de datos preestablecida para sincronizar la lectura, que consiste en una secuencia de varios ceros y luego un uno, que independientemente del sistema de grabación permite recuperar el sincronismo del reloj, para la posterior lectura de los datos.

El encabezamiento tiene su propio código de verificación.

3.4.2.2 Datos

En la zona indicada como datos es donde residen los datos del usuario (normalmente 512 bytes por sector).

3.4.2.3 Chequeo

La información de chequeo se graba con los datos y luego se verifica al recuperar los datos, para determinar posibles errores de lectura. Algunos algoritmos de chequeo permiten no sólo detectar sino también corregir errores de varios bits.

El método de chequeo usual en medios magnéticos es el CRC (cyclic redundancy check).

El mismo consiste en una función que se aplica a los datos, obteniendo un patrón de bits de longitud definida, por ejemplo 16 bits. La función es una división de polinomios (en módulo 2), donde el dividendo son los datos y el divisor un número fijo, por ejemplo, para el CRC-16 el polinomio, de 17 bits, es $x^{16}+x^{15}+x^2+1$. El CRC es el resto de la división.

Para este caso, hay un circuito que calcula el CRC a medida que se graban o leen los datos, lo que lo hace muy práctico para su uso en los medios de almacenamiento (y también muchas otras aplicaciones, como comunicaciones, etc.).

Consiste en un registro de desplazamiento con realimentaciones de compuertas XOR en los FF que corresponden a factores “1” en el polinomio divisor. En este ejemplo, el 74F401 calcula el CRC a medida que la secuencia de datos con su correspondiente reloj ingresa por la línea D (en la parte superior). Cuando el último bit es ingresado, en el registro queda el CRC.

Una propiedad interesante de este circuito es que cuando se realiza la lectura, si el CRC almacenado en el disco se envía a continuación de la secuencia de los datos, el resultado en el registro es “cero” si los datos se corresponden con el CRC. En caso de no ser “cero” se informa “error de CRC”.

Es importante hacer notar que el CRC está implementado para detectar los tipos de errores debidos a interferencia electromagnética, rayones, etc, que usualmente afectan muchos bits seguidos y de manera aleatoria.

El CRC-16 se usaba en las unidades de diskettes. En las unidades de discos rígidos se usan CRC de 32 bits, lo que permite, matemáticas mediante, también corregir secuencias erróneas.

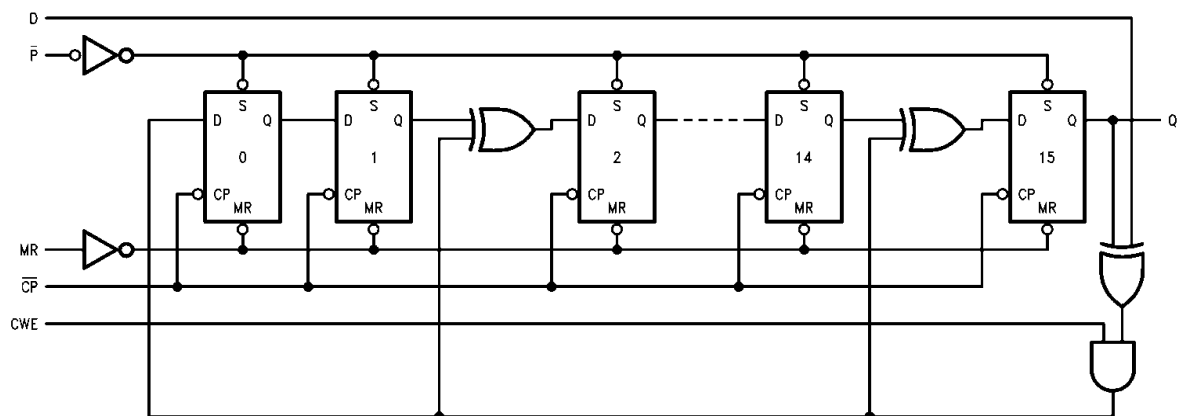


Figura 3.4.3 – Parte del 74F401 – CRC generator/checker

Nota:

Una analogía de CRC sería la paridad. Para 16 bits de datos, un bit de paridad permite detectar un error de 1 bit. Con codificación Hamming, para 16 bits de datos, 6 bits de chequeo permiten detectar errores de 2 bits y corregir errores de 1 bit (véase: memorias ECC).

3.4.2.4 Libre (gap)

Los espacios libres (gaps) se dejan para evitar superposición de grabaciones cuando se graba un solo sector sobre una pista previamente utilizada.

Debido a pequeñas diferencias de velocidad de rotación o de la frecuencia de reloj, si no estuvieran los gaps se podría sobrescribir el sector siguiente, inutilizando de esta manera el mismo.

3.4.3 Definiciones

Tiempo de búsqueda

Es el tiempo que tarda el posicionador de la cabeza en ir desde el cilindro origen hasta el cilindro destino.

Latencia

Es el tiempo que tarda la cabeza en encontrar un sector dentro de una pista determinada.

Tiempo de acceso

Es la suma del tiempo de búsqueda y la latencia. Se mide en ms (milisegundos).

Tiempo de acceso promedio

Es la mitad del peor tiempo de acceso. También se puede determinar por métodos estadísticos.

Velocidad de transferencia

Es la velocidad a la que se transfieren los datos entre la unidad de disco y la CPU. Depende de la cantidad de datos que hay en una pista y la velocidad de rotación del plato y se mide en kB/s o MB/s.

3.5 UNIDAD DE DISCOS RIGIDOS

Las placas o platos son de aluminio y están recubiertas con depósitos de metales ferromagnéticos. Las mismas giran juntas, impulsadas por un motor con control de velocidad preciso. En general se trata de un motor sincrónico, con velocidades desde 5400 rpm, para discos de portables o de videovigilancia, de 7200 rpm para aplicaciones generales y excepcionalmente de mayor velocidad para aplicaciones críticas.

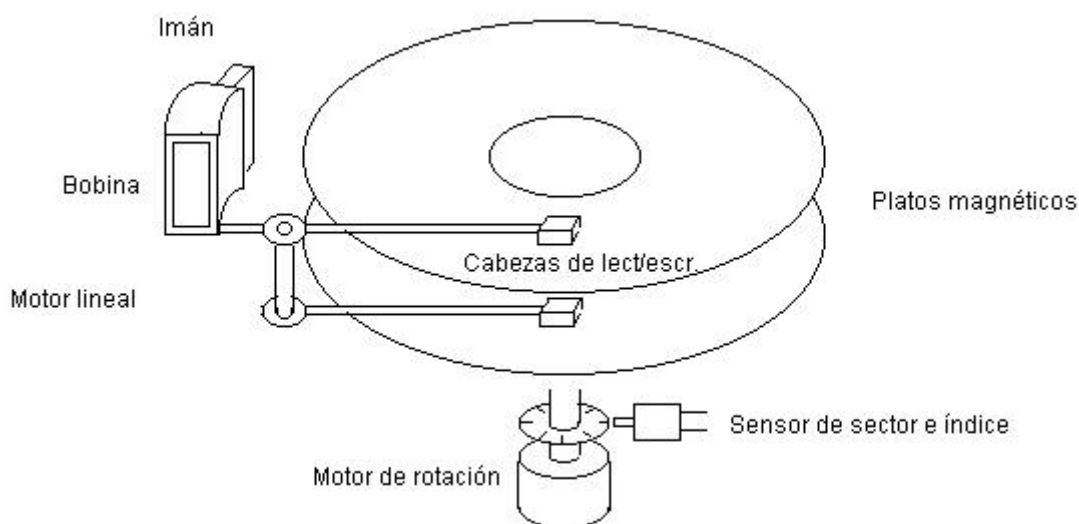


Figura 3.5.1 – Unidad de accionamiento de discos rígidos

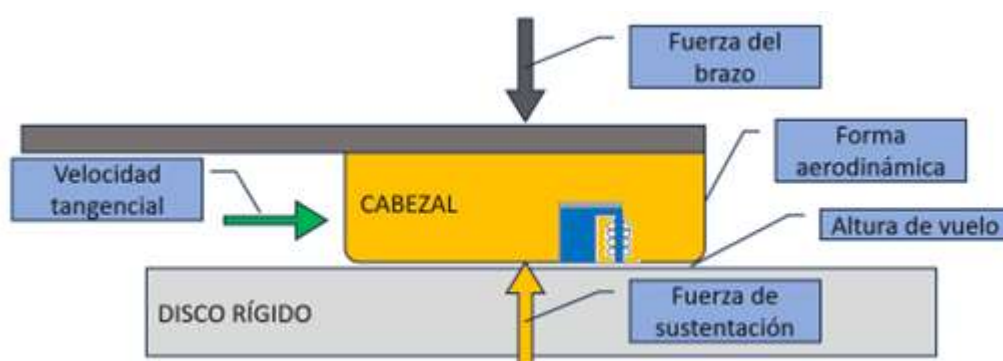


Figura 3.5.2 – Altura de vuelo del cabezal de disco rígido

Cuando el disco está detenido, las cabezas apoyan sobre el plato, usualmente en las pistas interiores (cercanas al eje de rotación).

Cuando el disco gira, las cabezas "vuelan" a pocos micrones de distancia de la superficie. Los factores que afectan a la "altura de vuelo" de la cabeza son:

- la velocidad de la cabeza respecto del plato (velocidad tangencial) producto de la velocidad angular y la distancia al eje de rotación.
- la forma aerodinámica de las cabezas.
- la fuerza del elástico que sostiene las cabezas, bastante mayor que el peso de ellas para que la altura no sea diferente en los cabezales superiores e inferiores.

Debido a estos factores, la cabeza está ligeramente más lejos del plato en las pistas exteriores que en las interiores y la unidad puede funcionar en cualquier posición.

El motor que posiciona las cabezas es de tipo lineal (voice coil) para lograr tiempos de búsqueda reducidos. Su principio de funcionamiento es similar al de un parlante de audio, donde una bobina solidaria al soporte de las cabezas está inmersa en un campo magnético creado por un imán permanente fijo al bastidor (chassis) de la unidad. Al circular corriente por la bobina aparece una fuerza perpendicular al plano de la bobina, haciendo que ésta se mueva.

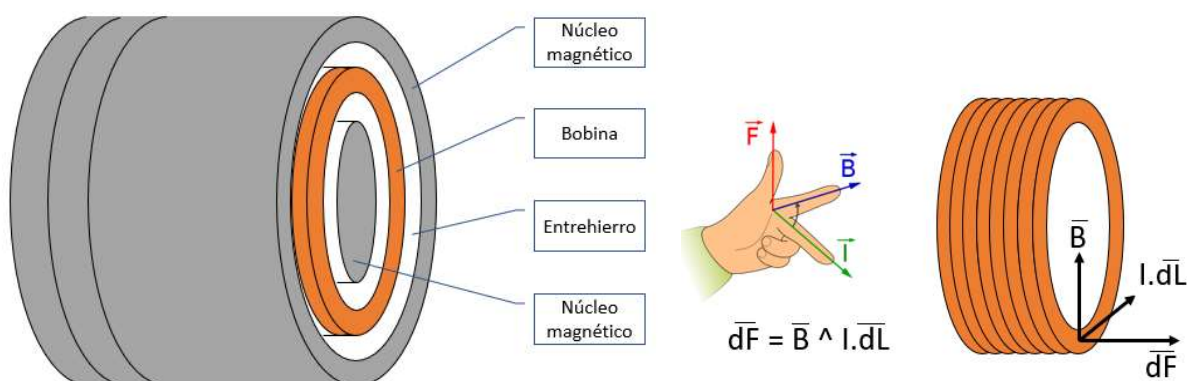


Figura 3.5.4 – Motor lineal



Figura 3.5.3 – Interior y detalle del motor de accionamiento de las cabezas

En la figura 3.5.3 se aprecia un motor con una configuración geométrica diferente, pero que tiene el mismo principio de funcionamiento que el “voice coil”. Con esto se ahorra espacio.

Un sistema realimentado (servosistema) es el encargado de regular la corriente para que la cabeza se dirija al cilindro deseado y una vez allí se mantenga.

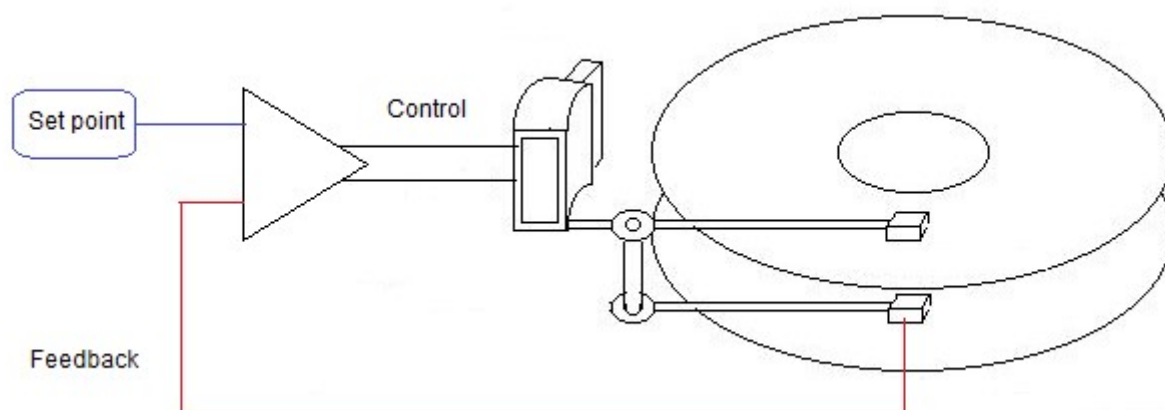


Figura 3.5.5 – Servosistema de posicionamiento

La información que sirve de referencia de posición está grabada en las pistas como se vio en formateo. Existe también otra señal, grabada a ambos lados de cada pista, que sirve para centrar la cabeza en dicha pista. Este proceso se denomina formateo de bajo nivel y lo realiza el fabricante, no pudiendo modificarse por parte del usuario.

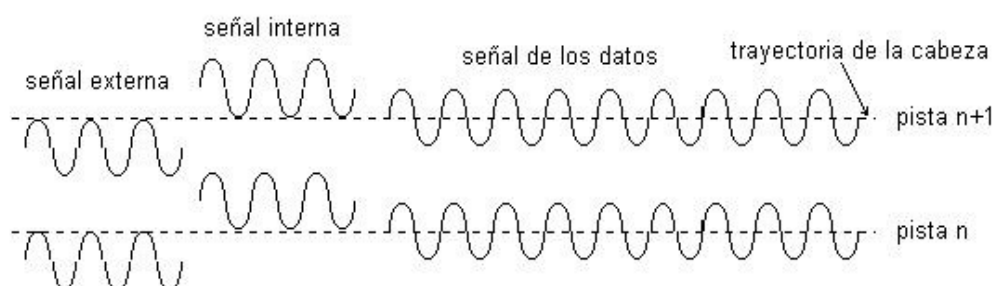


Figura 3.5.6 – Información de servo

Al comienzo de cada sector, la cabeza se centra tratando de que la señal externa y la interna tengan la misma amplitud. Este sistema tiene la ventaja que, si varían las dimensiones del disco, por ejemplo, por cambios de temperatura, las pistas serán localizadas con total precisión.

Para determinar la posición de los sectores, existía un disco ranurado que, trabajando en conjunto con un sensor inductivo permitía la generación de un impulso eléctrico por cada comienzo de sector. Una ranura doble (index), indicaba el comienzo del primer sector (sector 1). Ver la figura 3.5.1.

En discos actuales, con motor de rotación sincrónico, se utiliza un sensor Hall para indicar el sector 1 y el resto se hace por temporización precisa controlada por cristal. De esta forma se puede aprovechar mejor la superficie del disco, grabando más sectores en las pistas externas. En el controlador microprocesado de la unidad se hace la conversión de sectores físicos a lógicos y de esta forma el sistema operativo sigue viendo los sectores en la convención (CHS) o la convención lineal, en la que los mismos se numeran de forma correlativa desde el primero hasta el último.

Características técnicas de un disco rígido (año 2018)**Barracuda® ST SATA III-6Gb/s 14TB**

Specifications	
Model Number	ST14000DM001
Interface	SATA 6Gb/s
Cache	256MB
Capacity	14TB
PHYSICAL	
Height	26.1mm (1.028 in)
Width	101.85mm (4.010 in)
Length	146.99mm (5.787 in)
Weight (typical)	700g (1.543 lb)
PERFORMANCE	
Spin Speed (RPM)	7200 RPM
Sustained data transfer rate	250Mb/s
Average latency	4.16ms
Random read seek time	8.5ms
Random write seek time	9.5ms
I/O data transfer rate	600MB/s
Unrecoverable read errors	1 in 10**15
RELIABILITY	
MTBF	750,000 hours
Annual Failure Rate	0.34%
POWER	
12V start max current	2A
Average wait power	0.8W
Average idle power	5W
Average operating power	6,9W
ENVIRONMENT	
Ambient Temperature	
Operating	5°–60°C
Nonoperating	-40°–70°C
Maximum operating temperature change	20°C per hour
Maximum nonoperating temperature change	30°C per hour
Shock	
Operating Shock (max)	63 Gs for 2ms
Nonoperating Shock (max)	300 Gs for 2ms
ACOUSTICS	
Acoustics (Idle Volume)	2.8 bels
Acoustics (Seek Volume)	3.2 bels

3.6. UNIDAD DE DISCOS FLEXIBLES

En la figura se muestra un esquema de una unidad de accionamiento de discos flexibles. Las placas magnéticas son de Mylar y están recubiertas en ambos lados con óxidos magnéticos y luego con un material resistente al rozamiento. La placa está contenida en un envase con acceso para las cabezas de lectura/escritura y el aro de arrastre. El tamaño de las placas puede ser de 3,5 y 5,25 pulgadas (actualmente en desuso). Las cabezas rozan sobre la placa cuando se está realizando un acceso. Hay una cabeza de cada lado, que trabaja sobre la respectiva superficie del disco.

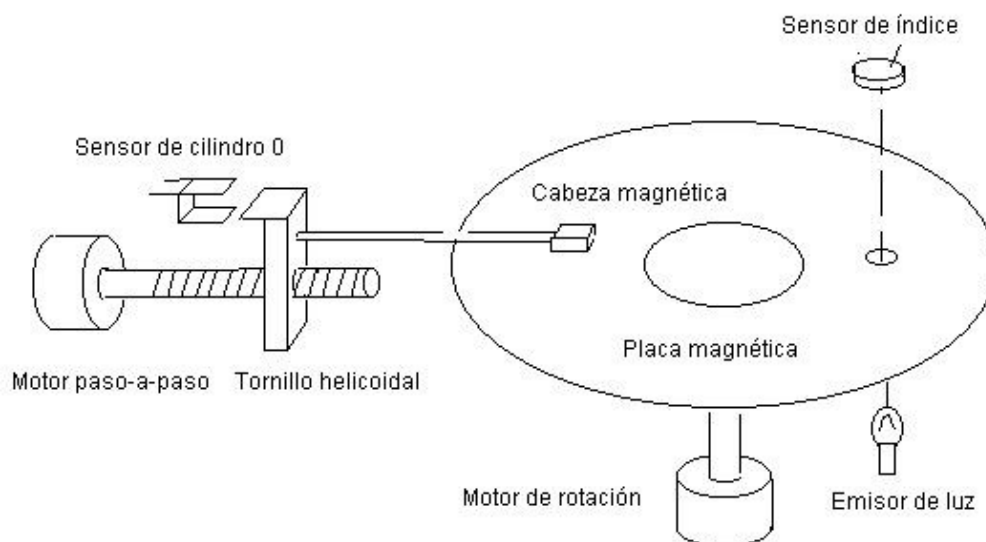


Figura 3.6.1 – Unidad de accionamiento de discos flexibles

El motor que posiciona las cabezas es del tipo paso a paso (stepper motor). La referencia para la posición la da un sensor óptico (sensor de cilindro 0) que es excitado cuando el posicionador de las cabezas retrocede hasta que la lengüeta interrumpe el haz de luz. El cilindro 0 es el más externo, a partir de este, cada paso del motor representa un nuevo cilindro.

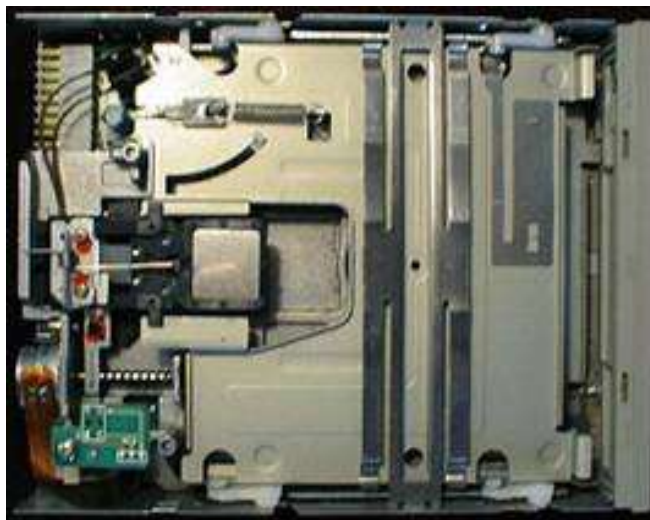


Figura 3.6.2 – Interior de la unidad de accionamiento de discos flexibles

Otro sensor óptico determina la posición del primer sector de un cilindro, gracias a una perforación que posee la placa (en los discos de 5,25 pulgadas). En los discos de 3,5 pulgadas, no es necesario el orificio, ya que el anillo metálico de arrastre posee una muesca que encastra siempre en la misma posición en la unidad de accionamiento.

Los sectores se determinan por software durante el formateo.

3.7. UNIDAD DE CINTA MAGNETICA

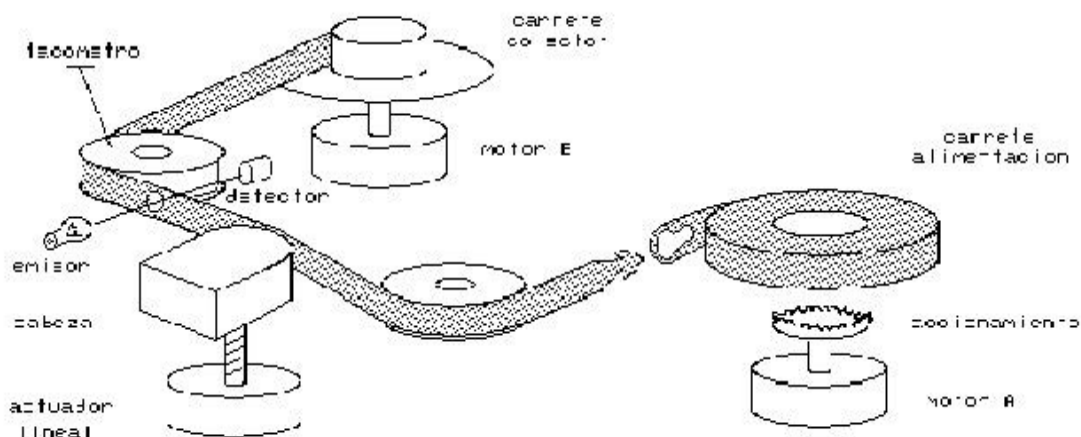


Figura 3.7.1 – unidad de accionamiento de cinta magnética

La unidad de cinta de la figura opera con cassettes de un solo carrete. Las dimensiones de la cinta son $\frac{1}{2}$ " (12,7 mm.) de ancho y 600 pies (180 m) de largo y las del cassette 10x10x2 cm. También existen cassettes de dos carretes, como los de audio.

El cabezal posee dos juegos de cabezas: una de grabación y una de lectura para cuando la cinta se mueve hacia adelante y otras dos para el movimiento hacia atrás. La disposición de cabezas de escritura y lectura alineadas y separadas unos 8 mm. permite la operación de verificación leyendo los datos escritos (read-after-write).

El cabezal puede adoptar 11 alturas diferentes gracias a un actuador paso a paso. De esta forma, dado que hay dos juegos de cabezas, se graban 22 pistas.

La combinación de los movimientos de la cinta y de la cabeza determinan pistas en forma de serpentina, que dan origen al nombre genérico de este tipo de unidades (streaming tapes).



Figura 3.7.2 – Disposición de las pistas en una "streaming-tape"

Al introducir el cassette en la unidad, se abre una puerta en la parte posterior de la caja, que permite el acceso al extremo libre de la cinta. Cuando se asegura la traba del cassette también asciende el encastre de accionamiento (motor A).

Un pescante de material plástico adosado al carrete colector, tiene su extremo especialmente diseñado para enganchar el borde libre de la cinta y tirar (mediante el motor B)

hasta que la misma quede enhebrada en ambos carretes. En este momento el cassette no puede ser retirado porque se encuentra accionada la traba de seguridad, manejada por un electroimán.

La cinta posee un orificio que activa el detector (ver figura) y le indica a la unidad que se alcanzó el comienzo de cinta. La velocidad de la cinta se controla regulando las velocidades de los motores A y B (de corriente continua) y teniendo en cuenta la información que suministra un generador de pulsos (tacómetro) solidario a una de las poleas. La tensión de la cinta se ajusta en función del consumo de los motores.

Este tipo de unidades, usualmente está controlada por microprocesador, por lo tanto posee diagnósticos internos, tanto de la parte electrónica como de la mecánica y brinda al controlador información completa de sus condiciones operativas.

La descripción precedente corresponde a una unidad de cinta que se usaba con las mini-computadoras DEC PDP-11, para realizar backup de los discos rígidos (1985).

En la figura siguiente se puede apreciar una unidad actual y su correspondiente medio. Estos medios pueden ser de lectura escritura o WORM (Write Only Read Many). Esta última se utiliza con fines de almacenamiento en aplicaciones donde los datos no puedan ser alterados una vez grabados, por ejemplo por temas legales o de seguridad informática.



Figura 3.7.3 – Cinta magnética y unidad de accionamiento de cinta

Características técnicas de una unidad de accionamiento de cinta (2011)

TS1140 tape drive

Characteristics

Recording technique	Linear Serpentine
Number of tracks	2560
Native capacity (uncompressed)	4 TB (using JC/JY media), 1.6 TB (using JB/JX media), or 500 GB (using JK media)
Native sustained data rate (uncompressed)	250 MBps
Burst data rate	800 MBps
High-speed search (max)	12.4 mps
Warranty	One year

Physical characteristics

Dimensions	95 mm H x 198 mm W x 467 mm D (3.8 in x 7.8 in x 18.4 in)
Weight	5.7 kg (12 lbs 7 oz)

3.8. UNIDAD DE DISCOS ÓPTICOS

3.8.1 Unidad de CD ROM, DVD ROM y Blu-ray

El medio, removible y de bajo costo, es un disco de policarbonato de 12 cm de diámetro en el que se graba una sola pista en forma de espiral, comenzando en el interior del mismo. La grabación se realiza colando el policarbonato en un molde o matriz que tiene la forma complementaria a la que se desea obtener.



Figura 3.8.1 – Pista en espiral

La unidad de discos ópticos utiliza un láser para la lectura de la información. Un sistema óptico dirige la luz del emisor hasta la superficie del disco y simultáneamente, a través de prismas, al detector. La luz reflejada en el disco también es dirigida al receptor, como se ve en la figura:

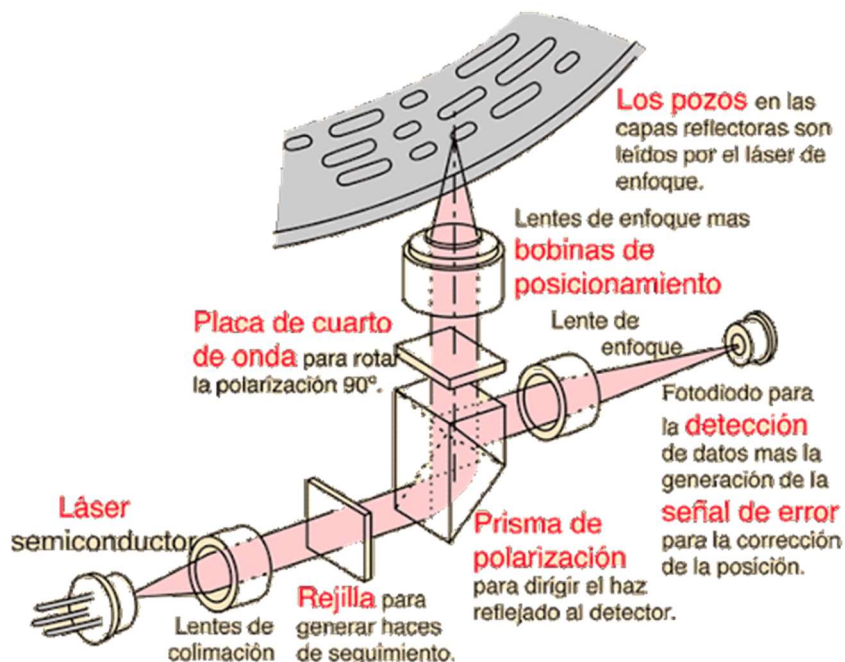


Figura 3.8.2 – Unidad de CD

Para que el sistema de lectura pueda recorrer toda la pista, se cuenta con un motor que mueve el cabezal hacia adentro o hacia afuera del eje de rotación.

El movimiento del cabezal es acompañado por unas bobinas que mueven la lente para “saltar” de pista, manteniéndose siempre enfocado a donde debe leer. La velocidad de rotación

del eje es variable, de manera de mantener una velocidad lineal constante del cabezal respecto de la pista espiral. Esto se debe a que por diseño se pretende una tasa de bits constante en toda la superficie del disco.

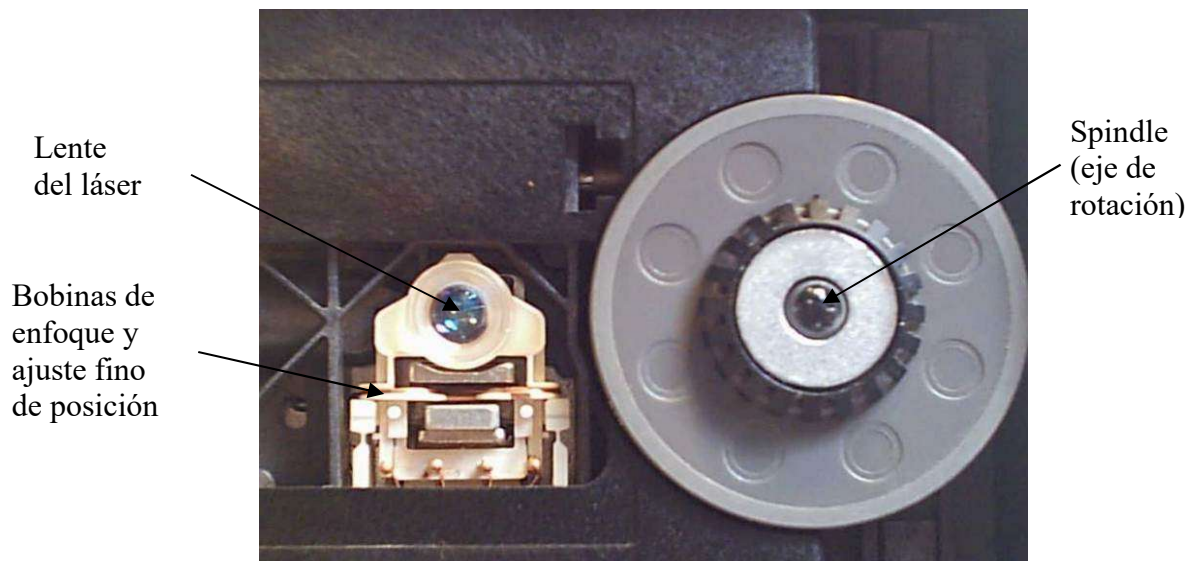


Figura 3.8.3 – Detalles de una unidad de accionamiento

Los datos se codifican mediante pozos (pits) y planos (lands) en la superficie del disco. La profundidad de un pozo (respecto del plano) corresponde a $\frac{1}{4}$ de la longitud de onda de la luz utilizada. De esta forma, como se mencionó anteriormente que tanto la luz emitida como la reflejada llegan al detector, ambas ondas se suman o restan dependiendo de si han incidido en un pozo o un plano, como muestra la figura siguiente:

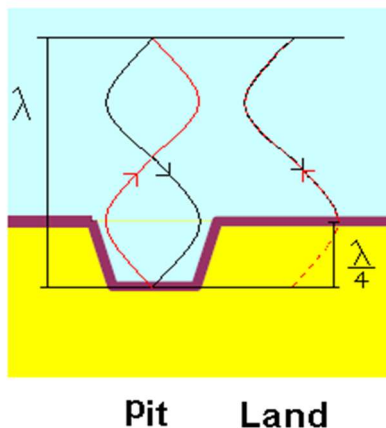


Figura 3.8.4 – lectura de un CD ROM

Así, los pozos se “ven” como “oscuros” (las ondas se restan) y los lands como “claros” (las ondas se suman). De esta manera al leerse la pista se pueden detectar las transiciones entre pozos y planos lo que asociado a la temporización permite recuperar la información digital (como hacíamos en los medios magnéticos).

En el caso de los CD el método de codificación utilizado es EFM (Eight to Fourteen Modulation), una derivación de RLL que codifica 8 bits en 14 combinaciones de T (transición) y N (no transición).

Es de destacar que en ningún caso hay pozos o planos de menos de 3 ciclos de reloj, lo que facilita la lectura y tampoco de más de 10 ciclos de reloj lo que facilita la recuperación del reloj.

Los DVD utilizan el método EFMPlus, derivado del anterior y con mayor densidad de información. Ambos fueron patentados por Philips.

El formateo de los sectores con códigos de verificación permite la corrección de errores en ciertos casos, como rayones o huellas en la superficie del disco.

Referencias:

Runlength-Limited Sequences . December 1990 - Proceedings of the IEEE 78(11):1745 - 1759

Puede consultar el documento completo en https://www.researchgate.net/publication/2984369_Runlength-Limited_Sequences

Características de los medios

En las siguientes figuras se aprecian las características comparativas entre un CD, un DVD y un Blu-ray. Nota: NA (apertura numérica de la lente).

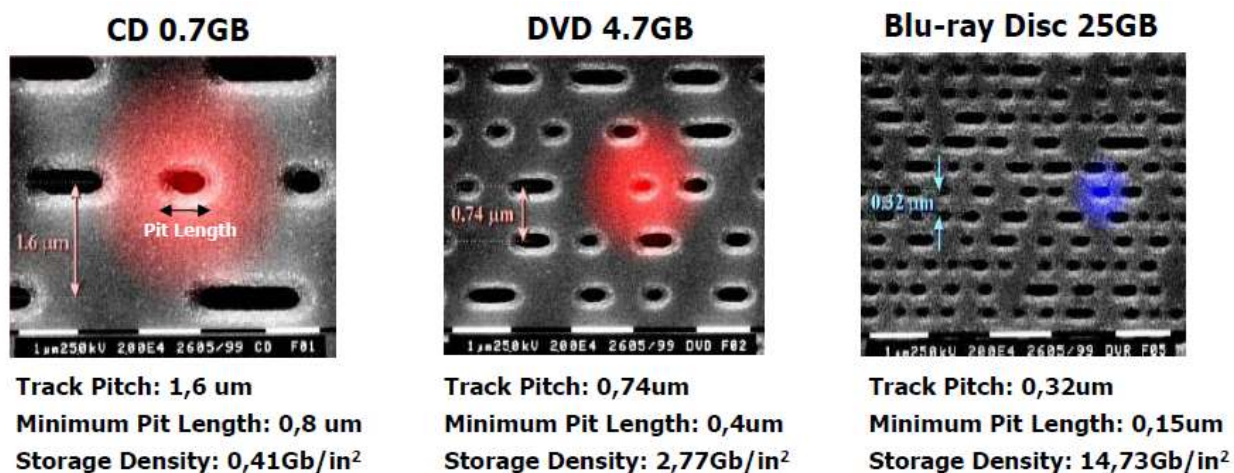


Figura 3.8.5 – Densidad de almacenamiento. Comparativa entre CD, DVD y Blu-ray



Figura 3.8.6 – Longitud de onda. Comparativa entre CD, DVD y Blu-ray

3.8.2 Unidad RW (Read/Write)

Cuando los discos RW se fabrican, el área reescribible es tal que hay surcos a lo largo de una pista espiral continua. La superficie de estos surcos es plana, con una capa reflectiva (es decir, la superficie de grabación está en ella es estado cristalino, significando que no hay datos).

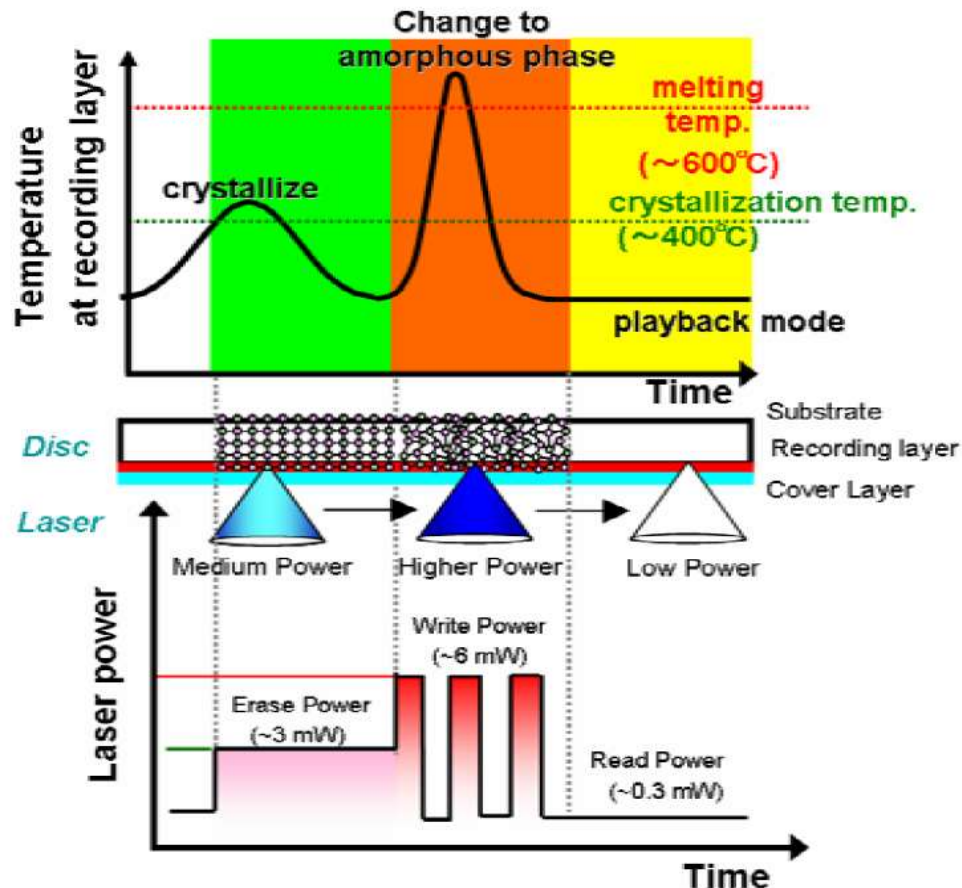


Figura 3.8.7 - Proceso de grabación de cambio de fase

La capa de grabación del disco RW se hace de un material que pueda cambiar su estado (o fase) a partir de una fase cristalina (land) a una fase amorfa (pit). La combinación de segmentos cristalinos y amorfos constituye una sección grabada.

Cuando se graba en el disco, la potencia del láser de la unidad de grabación es controlada de manera que varíe entre el punto de la cristalización, el punto de fusión, y un punto de enfriamiento, que es cercano a la temperatura ambiente. Controlando la potencia del láser de esa forma mientras el disco gira, se crean las secciones grabadas.

Referencias:

Blu-ray Disc Association - 1.C Physical Format Specifications for BD-ROM 5th Edition - March, 2007. Puede consultar el documento completo en <http://www.blu-raydisc.com/Assets/Downloadablefile/BD-ROMwhitepaper20070308-15270.pdf>.

3.9. ARREGLOS DE DISCOS (RAID)

El acrónimo RAID (Redundant Array of Independent Disks) aplica en forma genérica a una combinación de dos o más discos rígidos para mejorar las prestaciones de velocidad, confiabilidad o ambas. La idea original (y de allí el nombre) fue lograr la confiabilidad de un disco costoso con dos discos económicos, grabando los mismos datos en ambos. Luego aparecieron más combinaciones y se generó la clasificación RAID de un conjunto de discos.

Referencias: <http://es.wikipedia.org/wiki/RAID>.

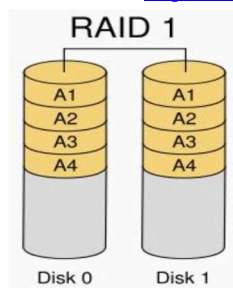


Figura 3.9.1 – RAID 1

El RAID 1 o espejado de discos aumenta la confiabilidad replicando los mismos datos en dos discos idénticos. La capacidad y la velocidad del conjunto es la misma que la de cada componente por separado. Ciertos controladores RAID permiten leer alternativamente de ambos discos, por lo que en estos casos la velocidad de lectura puede mejorar.

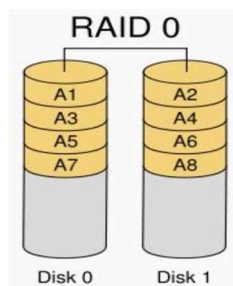


Figura 3.9.2 – RAID 0

El RAID 0, agrupa dos discos como si fuese uno solo de mayor capacidad. Esta configuración aumenta la capacidad y la velocidad de cada disco individual. Como contrapartida, reduce la confiabilidad, ya que la falla de cualquiera de ellos perjudicará al conjunto.

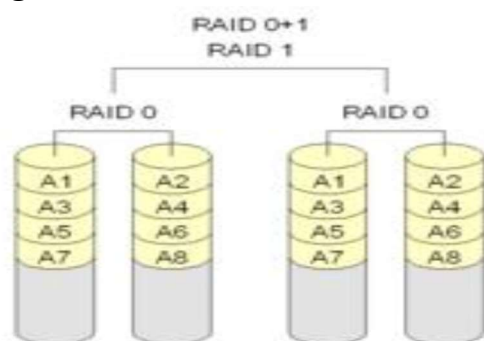


Figura 3.9.3 – RAID 0+1

RAID 0+1 es una combinación de los dos anteriores. Se trata de agrupar dos juegos RAID 0 en una configuración RAID 1. La confiabilidad del conjunto es igual a la de cada componente, se logra un incremento de velocidad y la capacidad es igual a la de cada conjunto RAID 0. Ejemplo: si se usan cuatro discos de 1 TB, se obtendrá un conjunto de 2 TB GB que la computadora ve como un disco único.

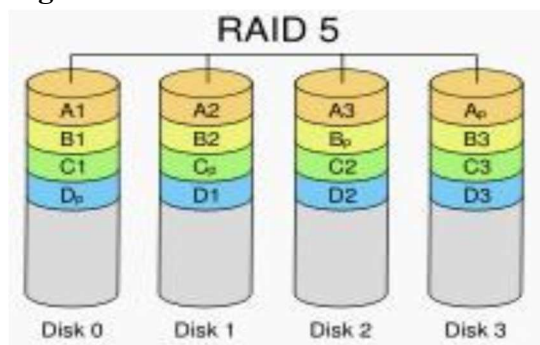


Figura 3.9.4 – RAID 5

RAID 5 requiere al menos tres discos. En el ejemplo de la figura se usan cuatro. Los archivos se dividen en “stripes” (A1, A2, A3) que se graban en discos diferentes y también se calcula y graba en otro disco la paridad correspondiente a los n strips (Ap). Si uno de los discos falla, los datos que estaban en él se pueden reconstruir a partir de la información de paridad guardada en los demás. El disco averiado se puede así reemplazar, logrando un aumento de confiabilidad con una reducida pérdida de capacidad de almacenamiento del conjunto.



Figura 3.9.5 – Unidad de almacenamiento con tecnología RAID

3.10. CONFIABILIDAD

En el apartado anterior se vio que, según la configuración utilizada, con dos discos se podía aumentar o disminuir la confiabilidad según sea la configuración utilizada.

Definimos la confiabilidad R como una función inversa de la tasa de fallas. Es decir, un equipo que falla muy seguido tendrá una baja confiabilidad.

Confiabilidad (R) = $e^{(-\lambda t)}$, donde:

- λ (lambda) es la tasa de fallas del sistema, normalmente expresada como fallas por unidad de tiempo, inversa del MTBF (Tiempo Medio entre Fallas), dato del fabricante de discos como se vio anteriormente.
- t es el periodo para el que se calcula la confiabilidad, puede ser un año.

3.10.1 – Ejemplo de cálculo – RAID 1

Supongamos para cada disco rígido una tasa de fallas (MTBF) de 1 cada 3 años, lo que equivale aproximadamente a 1 cada 1000 días. La tasa de falla es entonces 0,001.

Si los discos están configurados como RAID 1, es decir en espejo, los mismos datos están en ambos discos y la probabilidad de perder los datos a causa de una falla interna del disco estará dada por el producto de las probabilidades de falla de cada uno, dado que son sucesos independientes, entonces:

$$P(\text{RAID 1}) = P(\text{DISCO 1}) * P(\text{DISCO 2}) = 0,001 * 0,001 = 0,000001$$

El valor 0,000001 es muy pequeño, tanto que se corresponde con 1 en 2700 años.

Ver la Nota 3.1 más adelante para recordar el cálculo.

Estos cálculos están muy simplificados y merecen algunas consideraciones:

- El tiempo medio de reparación (MTTR) debe ser muy bajo en relación con el tiempo medio entre fallas (MTBF), por ello los sistemas RAID que incrementan confiabilidad poseen discos que se pueden reemplazar sin apagar el sistema. Ver figura 3.9.5.
- El aviso de que un disco está en falla debe ser inmediato, ya que de no hacerlo se estará incrementando el MTTR. Si no se cuenta con este aviso, el sistema podría estar trabajando con un disco solo y por ende su confiabilidad bajaría drásticamente, al valor del disco simple.

3.10.2 – Ejemplo de cálculo – RAID 0

Si, por el contrario, utilizamos los discos en RAID 0, entonces cualquiera de ellos que falle hará fallar al conjunto, entonces:

$$P(\text{RAID } 0) = P(\text{DISCO } 1) + P(\text{DISCO } 2) = 0,001 + 0,001 = 0,002$$

Dicho de otro modo, en el mismo período tendremos el doble de fallas, dado que tenemos el doble de discos en uso. El valor 0,002 se corresponde con 1,5 años aproximadamente.

3.10.2 – Ejemplo de cálculo – RAID 0+1

En las configuraciones más complejas, se aplican las mismas reglas:

$$P(\text{RAID } 0+1) = P(\text{RAID } 0) * P(\text{RAID } 1) = (P(\text{DISCO } 1) + P(\text{DISCO } 2)) * (P(\text{DISCO } 3) + P(\text{DISCO } 4)) = (0,001 + 0,001) * (0,001 + 0,001) = 0,002 * 0,002 = 0,000004$$

Este conjunto tiene una confiabilidad excelente, sumado a un aumento de la velocidad de transferencia al doble y una capacidad total igual a la mitad de la suma de todas las capacidades.

Si los discos son de 1 TB de capacidad y 100 MB/s de velocidad de transferencia sostenida, el conjunto RAID 0+1 tendrá una capacidad de 2 TB y una velocidad de 200 MB/s.

La tasa de fallas de 0,000004 se corresponde con 1 en 685 años.

Estas técnicas de redundancia para aumentar la confiabilidad se usan profusamente en aviación, centrales nucleares y procesos industriales críticos. En lo relacionado con procesamiento de datos, se usa esta técnica en los servidores de datacenters, etc.

En los equipos de medición y control se aplican los discos “espejados” o RAID 1 dado que en general son las únicas partes móviles del sistema y por ende las más propensas a fallas.

Nota 3.1. Probabilidad de ocurrencia de eventos independientes.

Si tenemos 2 dados, cada uno con 6 caras, la probabilidad de que al arrojarlos juntos se obtengan dos “1” es de 1/36. Esto sale de la regla del producto, que indica que la probabilidad es el producto de las probabilidades individuales (1/6). En forma gráfica, se representan todas las combinaciones posibles para los 2 dados como se ve en la figura (® creative commons).

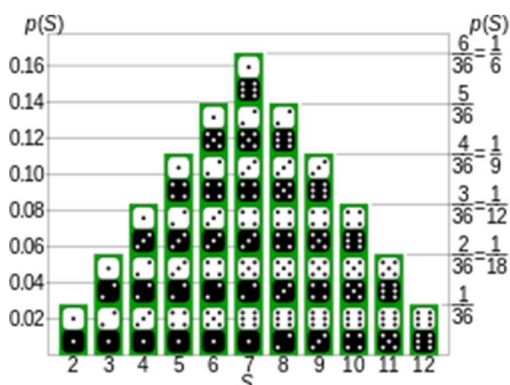


Figura 3.9.6 – Probabilidad de ocurrencia de cada número suma de dos dados