

Resumen: Final De Organización de Computadoras
por M.E.Defalco, M.E.Corróns y D.De Sousa Días

Unidad I: Computadoras Digitales

1. CONCEPTOS INTRODUCTORIOS

Informática: Viene de la contracción de las palabras INFORMación autoMÁTICA. Es la ciencia que estudia el análisis y resolución de problemas mediante computadoras.

Computadora: Es una máquina que cumple con las siguientes características

Digital: porque las señales eléctricas y la información que se procesa se representan a través de dígitos binarios (0 y 1).

Sincrónica: porque las operaciones internas están coordinadas y temporalizadas por un reloj central.

Posee cierta capacidad de cálculo numérico y lógico, gracias a la Unidad Aritmético Lógica donde se pueden realizar operaciones simples como adición, disyunción, conjunción, comparación.

Está controlada por un programa: se tienen instrucciones almacenadas en memoria que la computadora leerá, interpretará y ejecutará ordenadamente. Esta característica es la mayor diferencia entre una calculadora y una computadora y, además, convierte a las computadoras en máquinas muy versátiles (el funcionamiento de las mismas se puede modificar con sólo cambiar el programa).

Se comunica con el mundo exterior (que es analógico) utilizando dispositivos periféricos. Es decir, puede emitir y recibir información hacia y del mundo real.

Software: se refiere a los algoritmos o instrucciones que forman los programas.

Hardware: son los elementos tangibles de una computadora. Por ejemplo cables, circuitos integrados.

Firmware: es el software compuesto por un bloque de instrucciones que tienen un fin específico y que se almacena y se ejecuta desde la memoria ROM. Este software está integrado en la parte del hardware, es decir que viene incorporado con el dispositivo, por lo que el firmware es en cierto punto, hardware y software al mismo tiempo. El BIOS es un programa de tipo firmware, por ejemplo.

Arquitectura: son los atributos de un sistema que son visibles al programador. Por ejemplo, conjunto de instrucciones que tendrá una máquina, técnicas de direccionamiento, etc.

Organización: es cómo se implementan los atributos arquitectónicos. Una cuestión de organización es, por ejemplo, cómo se implementará la operación de multiplicación; será por medio de una unidad especializada o se efectuarán sumas sucesivas?

Programa: secuencia de instrucciones u órdenes ejecutables sobre una computadora que describe cómo realizar una tarea.

Algoritmo: conjunto de pasos finitos y ordenados que nos llevan a la solución de un problema, dicha solución debe ser dada en un tiempo aceptable.

2. ESTRUCTURA Y FUNCIONAMIENTO BÁSICO DE COMPUTADORES

El término **estructura** se refiere al modo en que los componentes se interrelacionan.

Con **funcionamiento** nos referimos a la operación de cada componente como parte de la estructura.

La mayoría de las computadoras actuales de propósito general presentan una estructura interna basada en la describió John Von Neumann.

Hay cuatro componentes estructurales principales:

Unidad Central de Procesamiento (UCP): (procesador o microprocesador) Es la encargada de buscar una por una las instrucciones de los programas en memoria, interpretarlas (decodificarlas) y ejecutarlas. La UCP se compone, a su vez, por varias partes: la Unidad de Control (UC), la Unidad Aritmético-Lógica (UAL) y una memoria pequeña y de alta velocidad -registros-. Dichas partes se conectan a través de un bus interno de la UCP.

Memoria Principal: se divide en Memoria de Instrucciones, donde se almacenan las instrucciones que la computadora debe ejecutar, y en Memoria de Datos, donde se guardan los datos necesarios que la computadora debe usar para resolver el problema.

Entrada/Salida: son los dispositivos que permiten la comunicación de la computadora con el mundo externo.

Bus de Comunicación: permite el paso de información de direcciones (Bus de Direcciones), datos (Bus de Datos) y control (Bus de Control), conectando los componentes de la computadora.

Funciones básicas de un procesador:

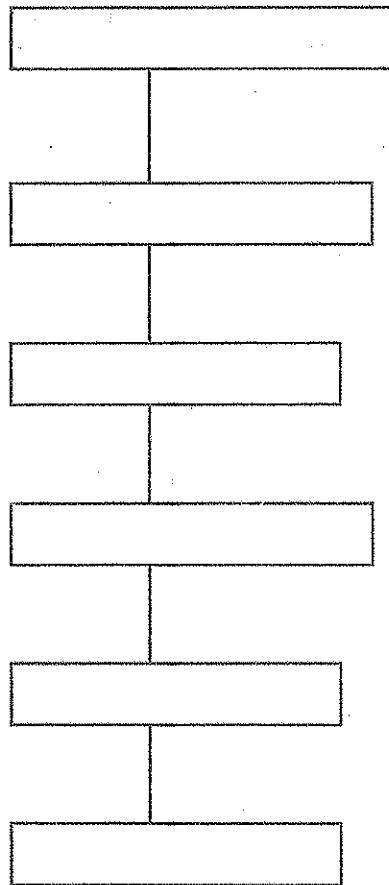
- Procesamiento de datos
- Almacenamiento de datos
- Transferencia de datos
- Control (de las tres funciones anteriores)

El funcionamiento de una computadora se puede simplificar como una secuencia infinita de los siguientes pasos:

- 1) Buscar la próxima instrucción a ejecutar en la memoria de instrucciones.
- 2) Interpretar qué hacer con la instrucción en la UC.
- 3) Ejecutar las operaciones interpretadas por la UC, usando la UAL de ser necesario. Estas operaciones pueden comprender lectura o escritura de la memoria de datos o entrada/salida por los periféricos.

Las computadoras se estructuran como una jerarquía o serie de abstracciones o capas o niveles, donde cada uno de éstos se sustenta en el que está debajo de él. Así, es posible controlar la complejidad y diseñar sistemas de cómputo de manera sistemática y organizada.

Máquina multinivel:



Nivel 0: nivel de lógica digital

- El nivel inferior sería el nivel de dispositivos ó transistores (física e ingeniería electrónica).
- En este nivel:
Puertas lógicas → son los objetos integrantes de este nivel, que combinadas forman:
 - » Circuitos integrados
 - » Circuitos combinacionales
 - » Circuitos aritméticos
 - » Relojes
 - » Memorias
 - » Microprocesadores
 - » Buses

Nivel 1: nivel de microprogramación

- Existe un programa llamado microprograma.
- La función del mismo es interpretar las instrucciones del nivel 2.
- En algunas máquinas no existe el nivel de microprogramación.

Nivel 2: nivel de lenguaje máquina

- Cada fabricante publica el "Manual de referencia del lenguaje máquina" para cada uno de los computadores.
- Las instrucciones del nivel de máquina las interpreta el microprograma.

- En las máquinas en las que no existe el nivel de microprogramación, las instrucciones del nivel de máquina son realizadas directamente por los circuitos electrónicos.

Nivel 3: nivel del sistema operativo

- La mayoría de las instrucciones de este nivel están también en el nivel 2 pero además tiene un nuevo conjunto de instrucciones, una organización diferente de la memoria, posibilidad de ejecutar 2 o más programas ...
- Las nuevas instrucciones las interpreta el sistema operativo.
- Las que son idénticas a las del nivel 2 las lleva a cabo el microprograma.

Nivel 4: nivel del lenguaje ensamblador

- El lenguaje ensamblador es una forma simbólica de los lenguajes subyacentes.
- El ensamblador es el programa que lleva a cabo la traducción de un programa del nivel 4.

Nivel 5: nivel de lenguajes de alto nivel

- Los lenguajes de alto nivel son más fáciles de utilizar que los lenguajes de niveles inferiores.
- Los traductores de programas en lenguaje de alto nivel pueden ser compiladores o intérpretes.

3. CLASIFICACIÓN DE LAS COMPUTADORAS

Se distinguen tres enfoques:

× Potencia de cálculo:

Desechables: generalmente chips individuales, como los que vienen en algunas tarjetas de felicitación.

Incorporadas: contienen un procesador, poca memoria y cierta capacidad de E/S, todo en un solo chip. Se encuentran en artículos eléctricos como lavarropas, microondas, etc.

De juego: tienen capacidades gráficas especiales pero software limitado y casi sin extensibilidad. Contienen un procesador, unos cuantos MB de memoria, una pantalla, entre otras cosas.

Personales: incluyen varios MB de memoria, un disco duro, unidad de CD-ROM, módem, etc. Cuentan con sistemas operativos complejos, muchas opciones de expansión y enorme surtido de software.

Servidores: se usan como servidores de red con configuraciones de uno o más procesadores, muchos GB de memoria y de disco y capacidad de trabajo en red de alta velocidad.

Macrocomputadoras: son extremadamente grandes y costosas. Generalmente tienen mucha capacidad de E/S.

Supercomputadoras: poseen UCP muy rápidas, muchos GB de memoria y sus discos y redes son muy rápidos. Se emplean para resolver problemas de cálculo en ciencia e ingeniería.

× Aplicación:

Comercial
Científica
De Control
De Comunicación
Base de Datos

x Estructura:

- Computadores serie
- Computadores con paralelismo interno
- Computadores con paralelismo externo

4. EVOLUCIÓN HISTÓRICA DE LAS COMPUTADORAS

Generación cero: computadoras mecánicas (1642-1945)

- 1642. Pascal construye una máquina calculadora funcional que podía sumar y restar.
- 1672. Von Leibniz construye una máquina similar a la de Pascal pero que además podía multiplicar y dividir.
- En 1823, Babbage diseña y construye su llamada máquina de diferencias, que sólo sumaba y restaba. Estaba diseñada para ejecutar un solo algoritmo y perforaba sus resultados en una placa de cobre (preursora de las tarjetas perforadas y el CD-ROM).
- Más tarde idea la máquina analítica, que sería capaz de realizar cualquier cálculo que se le programara mediante tarjetas perforadas. Ada Lovelace le ayuda a Babbage a desarrollar el sencillo lenguaje ensamblador que usaría la máquina. Por la insuficiente tecnología de la época, Babbage no logra eliminar los problemas de hardware de la máquina y no logra construirla.
- 1936. Zuse construye una serie de máquinas calculadoras automáticas empleando relevadores electromagnéticos. Sus creaciones son destruidas por bombardeos de guerra.
- Poco después Stibitz y Atanasoff realizaron trabajos similares a los de Zuse. La máquina de Atanasoff no logra entrar en operación por la falta de tecnología de hardware, pero lo que éste proponía era muy avanzado: usar aritmética binaria y que la memoria funcionara de la misma forma que hoy en día funciona la RAM.
- 1944. Aiken retoma el trabajo de Babbage y lo termina, llamando a su máquina Mark I. Las entradas y salidas se efectuaban mediante una cinta de papel perforada.

Primera generación: tubos de vacío (1945-1955)

Lo que estimuló la invención de la computadora electrónica fue la Segunda Guerra Mundial. Surgieron los tubos de vacío (válvulas), que eran tubos de vidrio del tamaño de una lámpara que albergaban circuitos eléctricos. Las máquinas que se construían eran muy grandes, costosas y de difícil operación.

- 1943. Se construye en Inglaterra la primera computadora digital del mundo: COLOSSUS, un proyecto secreto de guerra, cuyo fin era decodificar mensajes que se enviaban los alemanes.
- En 1946, Mauchley y Eckert finalizan la primer máquina electrónica de propósito general, el ENIAC (Electronic Numerical Integrator Computer), proyecto que habían comenzado en 1943 con intención de tenerlo listo antes del fin de la guerra para poder utilizarlo en aplicaciones de balística. El ENIAC era una máquina decimal, pesaba 32 toneladas, consumía 140 kw de potencia, tenía 18000 tubos de vacío y se programaba ajustando 6000 llaves de multiposición y conectando numerosas bases.
- Ese mismo año, los creadores del ENIAC comienzan a trabajar en su sucesor, el EDVAC. John Von Neumann, quien había participado en el proyecto ENIAC, propone una idea innovadora para el EDVAC, pero dicho proyecto sería abandonado por Eckert y Mauchley antes de terminarse.
- Por lo tanto, Von Neumann decide diseñar su propio EDVAC, la máquina IAS, en la cual emplearía su idea de diseño.

La máquina de Von Neumann

Von Neumann no tardó en darse cuenta que la tediosa e inflexible tarea de programar con interruptores podía ser dejada atrás.

El proceso de programación podría ser mucho más fácil si el programa se representaba en forma digital en la memoria de la computadora, junto con los datos (idea conocida como *concepto de programa-almacenado*).

Así, una computadora podría conseguir sus instrucciones al leerlas de la memoria, y se podría hacer o modificar un programa colocando los valores en una zona de memoria. La ejecución del programa sería secuencial y se aumentaría la velocidad.

Él percibió también que la aritmética decimal en serie utilizada por la ENIAC, en la que cada dígito se representaba con 10 válvulas, podía ser sustituida por una aritmética binaria. Así se simplificaría la implementación de funciones y se disminuiría la probabilidad de fallos.

Otro aspecto importante es que la memoria sería direccionable por localidad, sin importar el dato almacenado.

El computador IAS, no completado hasta 1952, es el prototipo de toda una serie de computadoras de propósito general. Su estructura consta de:

- unidad de memoria, donde se almacenan los datos e instrucciones.
- unidad aritmético lógica, capaz de hacer operaciones con datos binarios.
- unidad de control, que dirige las operaciones.
- un equipo de entrada/salida dirigido por la unidad de control que provee las instrucciones y datos (entrada) y envía los resultados (salida).

La memoria de la IAS constaba de 4096 posiciones de almacenamiento, llamadas palabras, de 40 bits cada una que podían ser 0 ó 1. Cada palabra contenía dos instrucciones de 20 bits o bien un entero con signo de 40 bits. Tanto la UC como la ALU contenían posiciones de almacenamiento, llamadas *registros*, definidos de la siguiente manera:

- Registro temporal de memoria (MBR)
- Registro de dirección de memoria (MAR)
- Registro de instrucción (IR)
- Registro temporal de instrucción (IBR)
- Contador de programa (PC)
- Acumulador (AC) y multiplicador/cociente (MQ)

IAS contaba con un total de 21 instrucciones que se pueden agrupar en los siguientes grupos: para transferencia de datos, para saltos (incondicionales y condicionales), para aritmética y para modificación de direcciones.

Con la máquina de Von Neumann se deja atrás el concepto de programar en hardware (al cambiar la tarea, hay que cambiar el hardware) y aparece el nuevo concepto de programación: programar en software.

Salvo raras excepciones, todas las computadoras de hoy en día tienen la misma estructura general y funcionamiento que la indicada por von Neumann.

1949. Computadora EDSAC. Utilizó el modelo de von Neumann, fue la primera computadora de programa almacenado (porque IAS se concluyó en 1952).

UNIVAC I, por Eckert y Mauchley. Primera computadora comercial. Primera en utilizar un compilador para traducir idioma de programa en idioma de máquinas. Su principal avance era un sistema de cintas magnéticas que podían leerse hacia delante y hacia atrás.

En 1953, IBM aparece en escena con la 701, su primer computador con programas almacenados. Se utilizó para aplicaciones científicas.

1955. Aparece la 702 de IBM, usada en aplicaciones de gestión.

1956. Aparece la 704, que tenía 4K de memoria de núcleos, instrucciones de 36 bits y hardware de punto flotante.

En 1958, IBM inicia la producción de su última máquina de bulbos, la 709.

Segunda generación: transistores (1955-1965)

El transistor fue inventado en 1948 en los laboratorios Bell por Shockley y colaboradores. Los transistores sustituyeron a los tubos de vacío ya que son más pequeños y baratos y disipan menos el calor. Dieron lugar a computadoras más pequeñas, confiables y económicas que las anteriores. En esta etapa nace la industria de las minicomputadoras.

En esta etapa también comenzaron a construirse ALU y UC más complejas, aparece el bus como un conjunto de alambres en paralelo para conectar los componentes de una computadora y lenguajes de programación de alto nivel, como el Algol 60, precursor del Pascal.

Nace el primer juego de video, de la mano de un grupo de estudiantes.

Se formaron nuevas compañías, como Digital Equipment Corporation (DEC) en 1957 que se concentró más que nada en bajar el costo del hardware. Otras compañías en cambio, como IBM, que para ese entonces ya era el principal fabricante de computadoras, y Control Data Corporation (CDC), compañía fundada en 1964, se preocuparon por construir máquinas cada vez más veloces.

Tercera Generación: circuitos integrados (1965-1980)

La invención del circuito integrado de silicio por R. Noyce en 1958 hizo posible colocar docenas de transistores en un solo chip. Este empaquetamiento permitió construir computadoras más pequeñas, más rápidas y menos costosas que sus predecesoras transistorizadas.

En 1968, el creador del circuito integrado junto con G. Moore y A. Rock formaron la Intel Corporation para fabricar chips de memoria. En 1971, de la mano de Intel, corporación que marcaría historia, nace el microprocesador al lograr fabricar un chip que contenía todos los componentes de la CPU.

Para 1964 IBM era el principal fabricante de computadoras y tenía un problema grave con sus dos máquinas de mayor éxito: eran totalmente incompatibles. IBM da entonces un paso radical: introduce una sola línea de productos, la System/360 diseñadas para computación tanto científica como comercial. Así nace el concepto de *familia de computadoras*, idea que fue muy bien recibida.

La familia de computadoras respondía a los siguientes puntos:

- a) Conjunto de instrucciones similares o idénticas
- b) Sistemas operativos similares o idénticos

Conforme se va de los modelos más bajos a los más altos de la familia:

- c) Velocidad creciente
- d) Número creciente de puertos de E/S
- e) Memoria creciente
- f) Costo creciente

Las IBM System/360 estaban basadas en circuitos integrados, la alimentación de la información era realizada por medio de tarjetas perforadas, previamente tabuladas y su almacenamiento se hacía en cintas magnéticas. IBM lanzó muchos modelos de esta serie. Su sistema operativo simplemente se llama OS (Operating System) y los lenguajes que manejaron fueron el FORTRAN, ALGOL y COBOL. Otra innovación de la 360 fue la multiprogramación (tener varios programas en memoria a la vez).

Cuarta generación: integración a grande escala (1980-1990)

Para 1980 la integración a muy grande escala había hecho posible colocar decenas de miles, luego centenares de miles y por último millones de transistores en un solo chip. Este avance dio pie a computadoras más pequeñas y rápidas. Los precios habían bajado tanto que una persona podía tener su propia computadora. Se había iniciado la era de la computadora personal.

Las computadoras personales se usaban para procesamiento de texto, hojas de cálculo y otras aplicaciones interactivas.

En 1981 IBM introduce la IBM Personal Computer. Esta utilizaba el Intel 8088 como su CPU y se convirtió de inmediato en la computadora más vendida del mundo. La versión inicial de la IBM PC venía equipada con el sistema operativo MS-DOS, creado por la entonces pequeña Microsoft Corporation.

Adelantos en:

- SO: MS-DOS, Windows, UNIX;
- Memoria: el precio por bit de memoria semiconductora cae por debajo del precio por bit de memoria de núcleo. La memoria semiconductora es mucho más veloz y no es destructiva. La capacidad se duplica cada año.

Quinta generación: desde 1990 en adelante

- Tecnología:
 - » Circuitos con más de un millón de componentes.
 - » Nuevas arquitecturas: paralelismo.
 - » Tecnología óptica.
- Modo de funcionamiento:
 - » Inteligencia artificial y sistemas expertos.

5. ANÁLISIS DE LA PERFORMANCE

El **benchmark** es una técnica utilizada para medir el rendimiento de un sistema o parte de un sistema.

Más formalmente puede entenderse que un benchmark es el resultado de la ejecución de un conjunto de programas en una máquina, con el objetivo de estimar el rendimiento de un elemento concreto o la totalidad de la misma, y poder comparar los resultados con máquinas similares.

Lo que se puede medir de un sistema a través de una aplicación o benchmark es:

- Tiempo de CPU: Tiempo de Usuario + Tiempo del Sistema
- Tiempo de Respuesta: Tiempo que el usuario espera por la primera respuesta de su programa
- Tiempo Real: Tiempo desde que se lanza un proceso hasta que este termina.
- Throughput: Número de tareas ejecutados en un determinado período de tiempo
- MIPS: Millones de instrucciones por segundo
- MFLOPS: Millones de Instrucciones de Punto Flotante por segundo
- Rendimiento Máximo Teórico: Máximo rendimiento que una computadora puede obtener.

El rendimiento global de un PC depende de cómo cada uno de sus componentes trabajan juntos para realizar una tarea. El efecto de cada componente individual puede variar dependiendo de en cuánto está interviniendo en ejecutar una aplicación particular.

El rendimiento del sistema depende del hardware al que se le este aplicando el benchmark.

En general, el rendimiento es afectado por varios factores:

- Arquitectura del Procesador
- Memoria
- Compilador
- E/S
- Sistema de Comunicación

Resumen: Final De Organización de Computadoras
por M.E.Defalco, M.E.Corróns y D.De Sousa Días

Unidad V: Memoria

1) Definiciones

- **Bit:** (dígito binario o *binary digit*) es la unidad de información más pequeña, más simple y básica de la memoria. Sólo puede tomar uno de dos valores: apagado o encendido (0 ó 1).
- **Nibble:** conjunto de 4 bits o medio octeto. Las cifras hexadecimales se representan con éstos; que también son la base del sistema de codificación BCD.
- **Byte:** conjunto de 8 bits; representado con la letra 'B'.
- **Palabra:** cadena finita de bits manejados como un conjunto por la máquina. Su tamaño o longitud hace referencia al número de bits que contiene. Los ordenadores modernos normalmente tienen un tamaño de palabra de 16, 32 ó 64 bits.
- **Palabra doble:** unidad de datos que duplica el tamaño de una palabra.

2) Sistemas de numeración

Teorema fundamental de la numeración

$$N = \sum_{i=-k}^n d_i \cdot \text{base}^i$$

d = dígito

i = posición del dígito con respecto a la coma fraccionaria

2.1. Sistema Decimal

Utiliza un conjunto de 10 símbolos (0,...9). Los números decimales se puede expresar por la fórmula del Teorema anterior, donde la base es 10. Utilizado por el hombre para contar, adoptado por contar con los diez dedos de la mano.

$$1492,36 = 1 \cdot 10^3 + 4 \cdot 10^2 + 9 \cdot 10^1 + 2 \cdot 10^0 + 3 \cdot 10^{-1} + 6 \cdot 10^{-2}$$

2.2. Sistema Binario

Sistema que utiliza internamente el hardware de las computadoras actuales; utiliza 2 símbolos (0 y 1).

$$111_2 = 1 \cdot 2^2 + 1 \cdot 2^1 + 1 \cdot 2^0 = 4 + 2 + 1 = 7_{10}$$

2.3. Sistema Hexadecimal

Sistema posicional que utiliza 16 símbolos para la representación de cantidades (0,...9, A,...F [A,...F equivalen a 10,...15 del sistema decimal])

$$\begin{aligned} 3E0, A &= 3 \times 16^2 + E \times 16^1 + 0 \times 16^0 + A \times 16^{-1} \\ &= 3 \times 256 + 14 \times 16 + 0 \times 1 + 10 \times 0,06 \\ &= 992,610 \end{aligned}$$

2.3.1. BCH (Sistema Hexadecimal codificado en Binario)

Los dígitos hexadecimales se convierten uno a uno en binario (utilizando 4 bits para cada uno).

2.3.2. BCD (Sistema Decimal codificado en Binario)

Los dígitos decimales se convierten uno a uno en binario (utilizando 4 bits); tiene dos ámbitos de aplicación:

- × E/S y periféricos: los números se codifican usando un byte por dígito:

DESEMPAQUETADO.

- **Terminales interactivos**
- **MODEM**

Terminales interactivos

La combinación de un monitor de vídeo con su correspondiente teclado se llama frecuentemente **terminal** y es normal acoplar varios terminales a un computador que se encarga de procesar las distintas tareas que cada usuario (desde su terminal) le ordena.

Podemos distinguir dos tipos de terminales:

- **Terminales no inteligentes:** Sólo son capaces de ejecutar operaciones de E/S simples.
- **Terminales inteligentes:** Capaces de ejecutar ciertos procesos tales como manipulación de texto, posibilidades gráficas o programas simples dirigidos por menús para ayudar a la entrada de datos. Esto es posible al incluir micropresesadores en los terminales.

MODEM

El MODEM es un dispositivo que permite conectar dos computadores remotos utilizando la línea telefónica de forma que puedan intercambiar información entre si. La información que maneja el computador es digital, sin embargo, por las limitaciones físicas de las líneas de transmisión, no es posible enviar información digital a través de un circuito telefónico, solo pueden transmitirse señales analógicas.

Entonces, para poder utilizar las líneas telefónicas (y en general cualquier línea de transmisión) para el envío de información entre computadoras digitales, es necesario un proceso de transformación de la información. Durante este proceso la información se adecua para ser transportada por el canal de comunicación.

Este proceso se conoce como **modulación-demodulación** y es el que se realiza en el **MODEM**.

Un módem es un dispositivo que posee conversores A/D y D/A especialmente adecuados para conectar líneas telefónicas al computador. De este modo las señales provenientes de una línea telefónica (por ejemplo por una llamada) son interpretadas y "atendidas" por el módem, permitiendo que otra computadora trasmite información directamente a la nuestra. Recíprocamente, nuestra voz, una imagen o datos guardados en nuestra computadora pueden ser manejados por el módem para ponerlos (previa llamada) sobre la línea telefónica que del otro lado tendrá una computadora receptora (o eventualmente un ser humano que descuelga un teléfono).

Existen distintos sistemas de modulación de una señal analógica para que transporte información digital, siendo los métodos más sencillos la **modulación de amplitud** y la **modulación de frecuencia**. Otros mecanismos como la modulación de fase o los métodos combinados permiten transportar más información por el mismo canal.

Baudios: es el número de veces de cambio de la señal por segundo en la línea de transmisión. Los módem envían datos como una serie de tonos a través de la línea telefónica. Los tonos se "encienden" (ON) o apagan (OFF) para indicar un 1 o un 0 digital. El baudio es el número de veces que esos tonos se ponen a ON o a OFF.

Los MODEM modernos pueden enviar 4 o mas bits por baudio.

Bits por segundo (BPS): es el número efectivo de bits/seg que se transmiten en una línea por segundo. Un MODEM de 600 baudios puede transmitir a 1200, 2400 o, incluso a 9600 BPS.

Los trazadores de gráficos (en inglés: "plotters") son dispositivos de salida que realizan dibujos sobre papel. Estos periféricos tienen gran importancia ya que con ellos se obtienen directamente del computador salidas en forma de planos, mapas, dibujos, gráficos, esquemas e imágenes en general.

El funcionamiento de un plotter se controla desde programa. El usuario puede incluir en su programa instrucciones para realizar las representaciones que deseé con sus datos.

Los registradores gráficos se fundamentan en el desplazamiento relativo de un cabezal con el elemento de escritura, con respecto al papel. Dependiendo del tipo de gráfico se moverá sólo la cabeza, o la cabeza y el papel.

Según la forma en que se realiza el dibujo, los registradores se pueden clasificar en tres tipos:

- de pluma.
- electrostáticos.
- de inyección

Microfilm

La salida de datos en microfilm (COM) es una técnica de representar los datos de salida. Las técnicas COM se usan en los bancos para llevar los registros de los balances diarios de cuentas. Esto supone un gran ahorro de papel, al evitar las salidas por impresora, al tiempo que reduce problemas de almacenamiento.

Cada "página" se representa en una pantalla y se fotografía mediante una cámara especial. La imagen de la página mide alrededor de 1,5 cm². La película se corta en microfichas del tamaño de una postal conteniendo cada una cien páginas aproximadamente. Se emplea un lector de microfichas para proyectar la imagen aumentada de una página cuando es necesario leerla.

Sintetizador de voz

Las unidades sintetizadoras de voz son dispositivos que dan los resultados de un programa emitiendo sonidos (fonemas o palabras) similares al habla humana. Estos periféricos de salida suelen incluir un microprocesador, memoria ROM con programas y datos, un conversor D/A, un amplificador de audiofrecuencia y altavoz. La mayor parte de los dispositivos sintetizadores de voz tienen memorizados digitalmente cada uno de los fonemas o palabras que son capaces de emitir. Los datos que recibe un sintetizador, procedentes del computador, corresponden a la identificación de los fonemas o palabras a emitir. Una vez que se analiza el dato, se activa una rutina encargada de generar el sonido correspondiente.

Los sonidos resultan muy metálicos. Por lo general, estos sistemas incluyen programas que enriquecen las posibilidades de los mismos, como por ejemplo, generar frases o combinaciones de palabras, incluso hay sistemas que traducen cantidades.

Conversor D/A

En general un **conversor D/A (Digital/Analógico)** permitirá poner señales físicas (una presión, una fuerza, una alarma) en el mundo real, de acuerdo a las órdenes y datos internos de una computadora.

En el ámbito industrial podemos tener numerosos ejemplos de este tipo, donde la computadora (a veces llamada "robot") automáticamente registra datos de entrada de periféricos muy variables y produce señales de salida sobre otros componentes electromecánicos para producir acciones en tiempo real (por ejemplo, armar o pintar un automóvil en una línea de montaje).

Unidades de E/S mixtas

Las impresoras electrostáticas utilizan un papel especial eléctricamente conductor (de color gris metálico). La forma de los caracteres se produce por medio de cargas eléctricas que se fijan en el papel por medio de una hilera de plumillas que abarcan el ancho del papel. Posteriormente a estar formada eléctricamente la línea, se la hace pasar, avanzando el papel, por un depósito donde se la pulveriza con un líquido que contiene suspendidas partículas de tóner (polvo de carbón). Las partículas son atraídas en los puntos que conforman el carácter. Estas impresoras de línea son muy rápidas.

i) Impresoras láser

Estas impresoras tienen en la actualidad una gran importancia por su elevada velocidad, calidad de impresión, relativo bajo precio y poder utilizar papel normal. Su fundamento es muy parecido al de las máquinas de fotocopiar. La página a imprimir se transfiere al papel por contacto, desde un tambor que contiene la imagen impregnada en tóner.

La cantidad de 'puntos por pulgada' de resolución y la cantidad de hojas por minuto son parámetros que caracterizan sus prestaciones.

La impresión se realiza mediante radiación láser, dirigida sobre el tambor cuya superficie tiene propiedades electrostáticas (se trata de un material fotoconductor, tal que si la luz incide sobre su superficie la carga eléctrica de esa superficie cambia).

Parámetros que caracterizan a una impresora

- 1) **Velocidad de escritura.** Normalmente la velocidad de impresión se expresa en las siguientes unidades:
 - Impresoras de caracteres: Caracteres por segundo (cps).
 - Impresoras de líneas: Líneas por minuto (lpm).
 - Impresoras de páginas: Páginas por minuto (ppm).
- 2) **Caracteres por línea.** Es el número máximo de caracteres que se pueden escribir en una línea.
- 3) **Ancho del papel** o longitud del carro. Se suele expresar en pulgadas.
- 4) **Densidad de líneas.** Se expresa normalmente en líneas por pulgada e indica el espaciado entre líneas.
- 5) **Tipos de letras.** Una misma impresora puede escribir con distintos tipos de letras. Las de margarita pueden cambiar de caracteres sustituyendo la margarita. Las de agujas usualmente pueden escribir en tamaño normal, comprimido y elongado. El cambio del tipo de letra se realiza por programa.
- 6) **Color.** Es la posibilidad de imprimir en colores. Usualmente los colores se forman a partir de tres básicos (como en los monitores color), pero a partir del cian, el magenta y el amarillo.
- 7) **Resolución.** Una gran parte de impresoras forman los caracteres por unión de puntos. La resolución se suele expresar como número de puntos por unidad de superficie.

Otros dispositivos de salida

Plotter o Trazador de gráficos

Son impresoras de calidad de impresión, sin embargo son relativamente lentas. Los caracteres se encuentran modelados en la parte más ancha (más externa) de los sectores (péntalos) de una rueda metálica o de plástico en forma de margarita. El juego de caracteres se puede cambiar fácilmente sin más que sustituir la margarita. Actualmente están fuera de uso.

c) Impresoras matriciales o de agujas

Estas impresoras, también denominadas de matriz de puntos, son las más utilizadas con microcomputadores y pequeños sistemas informáticos. Los caracteres se forman por medio de una matriz de agujas. Las agujas golpean la cinta entintada, trasfiriéndose al papel los puntos correspondientes a las agujas disparadas.

Los caracteres, por tanto, son punteados, siendo su calidad muy inferior a los caracteres continuos producidos por una impresora de margarita.

d) Impresoras de tambor

Dentro de estas impresoras, encontramos dos tipos: las de tambor compacto y las de tambor de ruedas.

e) Impresoras de barras y de cadenas

Los caracteres se encuentran moldeados sobre una barra de acero que se desplaza a gran velocidad, oscilando delante de la línea a escribir. El juego de caracteres está repetido varias veces (usualmente tres). El número de martillos coincide con el número de caracteres por línea.

En las impresoras de cadena, los caracteres se encuentran grabados en los eslabones de una cadena. La cadena se encuentra cerrada y girando constantemente a gran velocidad frente a la cinta entintada.

f) Impresoras térmicas

Son similares a las impresoras de agujas. Se utiliza un papel especial termosensible que se ennegrece al aplicar calor.

El calor se transfiere desde el cabezal por una matriz de pequeñas resistencias en las que al pasar una corriente eléctrica por ellas se calientan, formándose los puntos en el papel.

Estas impresoras pueden ser:

- De caracteres: Las líneas se imprimen con un cabezal móvil.
- De líneas: Contienen tantas cabezas como caracteres a imprimir por línea.

Son más rápidos.

g) Impresoras de inyección de tinta

La primera patente referente a este tipo de impresión data del año 1951, aunque hasta el año 1983, en el que Epson lanzó la SQ2000, no fueron lo suficientemente fiables y baratas para el público.

El fundamento físico es similar al de las pantallas de vídeo. Se emite un chorro de gotas de tinta ionizadas que en su recorrido es desviado por unos electrodos según la carga eléctrica de las gotas. Los caracteres se forman según una matriz de puntos. Estas impresoras son bidireccionales y hay modelos que imprimen en distintos colores.

h) Impresoras electrostáticas

ver en forma directamente inteligible para el hombre los resultados de un programa de computadora.

Las impresoras tienen dos partes diferenciadas: la parte mecánica y la parte electrónica. Aquí la parte mecánica, además de encargarse de seleccionar el carácter a partir del código de E/S correspondiente, debe dedicarse a la alimentación y arrastre del papel.

Clasificación y tipos de impresoras

Existen diversos criterios para clasificar las impresoras.

a) Calidad de impresión

Tiene en cuenta la calidad de presentación y de contraste de los caracteres impresos. Las impresoras se clasifican, atendiendo a este aspecto en:

- Impresoras normales: Como las impresoras de línea, de rueda y térmicas.
- Impresoras de semicalidad. Como algunas impresoras matriciales.
- Impresoras de calidad: Como las impresoras margarita e impresoras láser.

b) Sistema de impresión

Según la forma en que realizan la impresión. Pueden ser:

- **Por impacto de martillos o piezas móviles mecánicas:** El fundamento de las impresoras por impacto es similar al de las máquinas de escribir. Las impresoras de impacto son muy ruidosas y tradicionalmente han sido las más utilizadas. Entre ellas se encuentran las impresoras de rueda, bola, margarita, matriciales, cilindro, cadena.
- **Sin impacto:** forman los caracteres sin necesidad de golpes mecánicos y utilizan otros principios físicos para transferir las imágenes al papel. Son impresoras sin impacto las térmicas, las de inyección de tinta y las impresoras láser.

c) Forma de imprimir los caracteres

En cuanto a este aspecto, las impresoras se pueden clasificar en:

- **Impresoras de caracteres:** Realizan la impresión por medio de un cabezal que va escribiendo la línea carácter a carácter. El cabezal se desplaza a lo largo de la línea que se está imprimiendo, sólo de izquierda a derecha (impresoras unidireccionales) o bien, para conseguir mayor velocidad, de izquierda a derecha y de derecha a izquierda sucesivamente (impresoras bidireccionales).
- **Impresoras de líneas:** En estas impresoras se imprimen simultáneamente todos o varios de los caracteres correspondientes a una línea de impresión.
- **Impresoras de páginas:** Aquí se incluyen un grupo de impresoras que actúan de forma muy similar a las máquinas fotocopiadoras. Se caracterizan por contener un tambor rotativo donde se forma con tinta o polvillo especial (toner) la imagen de la página a imprimir. Esta imagen, por contacto y un proceso de fijación se transfiere al papel.

Descripción de distintos tipos de impresoras

a) Impresoras de rueda

Son impresoras de impacto y de caracteres. Estas impresoras están en desuso.

b) Impresoras de margarita

Supongamos poseer un monitor alfanumérico que puede mostrar 25 filas con 100 caracteres en cada una, ¿cuánta memoria de imagen se necesita?

Dado que podemos mostrar 25×100 caracteres si cada uno de ellos utiliza 7 bits la memoria debe poder contener

$$25 \times 100 \times 7 \text{ bits} = 17500 \text{ bits} < 2188 \text{ bytes}$$

Si esos caracteres pudieran o no titilar, estar o no subrayado y estar o no resaltado se requiere almacenar 3 bits más por cada carácter (1 BIT para cada atributo) por lo que la memoria debe contener

$$25 \times 100 \times 10 \text{ bits} = 25000 \text{ bits} = 3125 \text{ bytes.}$$

Para un monitor gráfico tipo SVGA (equivalente a que cada carácter anterior utilice una matriz de 10×30 puntos) monocromo la memoria necesaria es

$$1024 \times 768 \times 1 \text{ BIT} = 786432 \text{ bits} = 98304 \text{ bytes} = 96 \text{ KBytes}$$

y si la imagen fuera True Color la memoria debería ser

$$1024 \times 768 \times 3 \text{ bytes} = 2359296 \text{ bytes} = 2304 \text{ KBytes} = 2,25 \text{ MBytes}$$

Tipos de video

- Monitor de video, memoria de visualización y teclado armados juntos para formar un *Terminal*
 - Usualmente orientados a carácter
 - Conexión con ancho de banda pequeño (serie)
- Monitor de video con memoria de visualización que está *mapeada en memoria*.
 - Permite mostrar imágenes y movimiento
 - Conexión al bus de memoria permite cambios rápidos (ancho de banda grande)

Visualizadores (o "displays")

Los visualizadores son pequeñas unidades de salida que permiten al usuario leer una instrucción, un dato o un mensaje. Los caracteres se forman partiendo de estructuras en módulos, cada uno de los cuales sirve para visualizar un carácter. Cada módulo contiene una serie de segmentos, siendo los más habituales de 7. Un carácter concreto se visualiza activando determinados segmentos, dependiendo de la forma del carácter.

El visualizador es el elemento de salida típico de las calculadoras de bolsillo y de los relojes digitales.

Impresoras

Las impresoras son periféricos que escriben la información de salida sobre papel. Su comportamiento inicialmente era muy similar al de las máquinas de escribir, pero hoy día son mucho más sofisticadas, pareciéndose algunas en su funcionamiento a máquinas fotocopiadoras conectadas en línea con el computador. Las impresoras son, junto a las pantallas, los dispositivos más utilizados para poder

campos eléctricos en el cristal líquido. Diferentes partes de la pantalla reciben diferentes voltajes y con esto se controla la imagen que se exhibe. Regados adelante y atrás de la pantalla hay filtros polarizantes porque la tecnología lo requiere.

Clasificación de pantallas

Según la **capacidad o no de mostrar colores** se clasifican en:

- **Monitor monocromo:** Los colores usuales en un monitor monocromático son el blanco y negro, ámbar o verde.
- **Monitor color:** El color de cada punto se obtiene con mezcla de los colores rojo, verde y azul, pudiéndose programar la intensidad de cada color básico.

Según su **capacidad de representación** se pueden clasificar en:

- **Pantallas de caracteres:** Sólo admiten caracteres.
- **Pantallas gráficas:** Permiten trazados de líneas y curvas continuas.

En las pantallas de caracteres, se debe poseer una memoria de imagen (que es una memoria de tipo RAM) que almacena la información correspondiente a cada celda (compuesta por el código binario del carácter a mostrar y la información binaria de como mostrar ese carácter -los atributos-) y una memoria de sólo lectura (tipo ROM) donde se almacenan los patrones de los caracteres representados como una matriz de puntos. La memoria ROM se denomina "generador de caracteres".

Con toda esta información almacenada el proceso a seguir es el siguiente:

- Se leen de la memoria de imagen los códigos de los caracteres que corresponden a cada posición de pantalla.
- Los códigos son enviados al generador de caracteres que proporciona la matriz de puntos correspondiente.
- Los atributos se aplican a cada carácter (matriz de puntos).

En las pantallas gráficas el usuario tiene acceso al punto de imagen, pudiendo representar en ellas imágenes configuradas no sólo con las formas de caracteres incluidos en la ROM. En este caso, se requiere una memoria de imagen que pueda contener la información de cada punto de imagen (intensidad, color y otros posibles atributos), en vez de la correspondiente a cada celda. Los dibujos, a pesar de estar formados por puntos de imagen presentan una apariencia de líneas continuas. La calidad de la pantalla gráfica depende de la densidad de puntos de imagen.

Principales parámetros que caracterizan a una pantalla

- **Tamaño:** Se describen en función del tamaño de la diagonal principal, y se tiene la mala costumbre de darla en pulgadas. Las más habituales son las de 15" (381 Mm.), aunque existen las de 14" (356 Mm.) y se generalizan 17".
- **Número de celdas o caracteres:** Lo usual es una representación de 24 filas y 80 columnas de caracteres (24*80 caracteres).
- **Resolución:** Es el número de puntos de imagen en pantalla. Éste número no depende del tamaño de la pantalla. Usualmente se consideran básicamente tres tipos de resolución:

CGA 640*200 puntos

VGA 640*480 puntos

SVGA 1024*768 puntos

Actualmente hay resoluciones superiores.

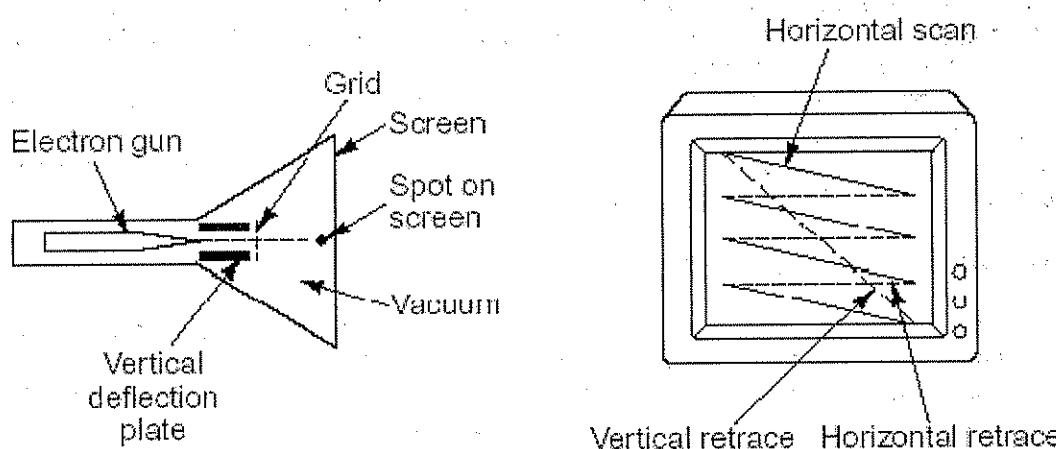
- Impresoras de tambor
- Impresoras de barras y de cadenas
- Impresoras térmicas
- Impresoras de inyección de tinta
- Impresoras electrostáticas
- Impresoras láser
- Plotter o Trazador de gráficos
- Microfilm
- Sintetizador de voz
- Conversor D/A

Monitor o Pantalla de Video

1. Tubos de Rayos Catódicos o TRC

Funcionamiento: La imagen de una pantalla de rayos catódicos (TRC) se forma al incidir un haz de electrones sobre la superficie interna de la pantalla que está recubierta de un material fosforescente, análogamente a como se forman las imágenes en un televisor.

Estas pantallas hacen desplazar el haz de electrones de izquierda a derecha y de arriba a abajo y, dependiendo de la intensidad con la que inciden los electrones en la pantalla así de brillante será cada punto de la imagen. La imagen, para ser visualizada durante un determinado tiempo debe ser repetida o refrescada periódicamente (al menos 25 veces por segundo). Estas pantallas se denominan pantallas de barrido.



Corte de un tubo de rayos catódicos (TRC) y recorrido del haz

2. Pantallas planas o LCD

Es una tecnología muy compleja, tiene muchas variaciones y está cambiando rápidamente.

Una pantalla LCD consiste en 2 placas de vidrio paralelas entre las que hay un volumen sellado que contiene un cristal líquido. Una luz detrás de la placa trasera ilumina la pantalla desde atrás. Cada placa tiene conectados electrodos transparentes. Estos unidos a cada placa sirven para crear

Un **ratón o mouse** tiene variantes tales como los trackballs de las computadoras portables, las palancas de los joysticks, las tabletas sensibles al tacto, e incluso la pantalla sensible al tacto.

Lápiz óptico

Funcionamiento: El lápiz contiene un pulsador, transmitiéndose información hacia el monitor sólo en el caso de estar presionado. Al activar el lápiz óptico frente a un punto de la pantalla se obtienen las coordenadas del lugar donde apuntaba el lápiz.

Características: Físicamente tiene la forma de una pluma o lápiz grueso, de uno de cuyos extremos sale un cable para unirlo a un monitor.

Pantalla sensible al tacto

Son pantallas que pueden detectar las coordenadas (x,y) de la zona de la propia pantalla donde se acerca algo (por ejemplo, con un dedo). Éste es un sistema muy sencillo para dar entradas o elegir opciones sin utilizar el teclado. Se utiliza para la selección de opciones dentro del menú o como ayuda en el uso de editores gráficos. Con frecuencia se ve en los denominados kioscos informativos, cada vez más difundido en grandes empresas, bancos y en puntos de información urbana. Existen pantallas con toda su superficie sensible, y otras en las que sólo una parte de ella lo es.

Scanner

Es un dispositivo que recuerda a una fotocopiadora y que se emplea para introducir imágenes en un computador. Las imágenes que se deseé capturar deben estar correctamente iluminadas para evitar brillo y tonos no deseados. Son dispositivos de entrada de datos de propósito especial que se emplean conjuntamente con paquetes software para gráficos y pantallas de alta resolución. La mayor parte de los scanners capturan imágenes en color generando una determinada cantidad de bits por cada punto. La cantidad de espacio de almacenamiento que se necesita para una imagen depende de las dimensiones máximas de la imagen y de la resolución de captura del equipo; la resolución se describe en 'cantidad de *puntos por pulgada*' ('dot per inch' en inglés o con el acrónimo 'dpi') en el sentido horizontal (eje x) y en el vertical (eje y).

Conversor A/D (Analógico/Digital)

Dispositivos que transforman datos analógicos del mundo real en datos digitales (por ejemplo un sensor de temperatura, un digitalizador de sonidos de una guitarra, un sensor de presión sanguínea, etc.). En todos estos casos alguna forma de **conversor analógico-digital** tiene la capacidad de convertir los datos del mundo real en alguna representación binaria (por ejemplo, la intensidad de un sonido) de modo que la computadora posteriormente pueda procesar los números resultantes.

Periféricos de salida de datos

- Monitor o Pantalla de Video
 - Tubos de Rayos Catódicos o TRC
 - Pantallas planas o LCD
- Visualizadores (o "displays")
- Impresoras
 - Impresoras de rueda
 - Impresoras de margarita
 - Impresoras matriciales o de agujas

Periféricos de entrada de datos

- Teclado
- Mouse
- Lápiz Óptico
- Pantalla sensible al tacto
- Scanner
- Conversor A/D (Analógico/Digital)

Teclado

Funcionamiento: Cuando utilizamos un teclado, al pulsar una tecla se cierra un conmutador que hay en el interior del teclado, esto hace que unos circuitos codificadores generen el código de E/S correspondiente al carácter seleccionado, apareciendo éste en la pantalla si no es un carácter de control.

Características: Los teclados contienen los siguientes tipos de teclas:

- **Teclado principal:** Contiene los caracteres alfabéticos, numéricos y especiales. Hay teclados que también incluyen aquí caracteres gráficos
- **Teclas de desplazamiento del cursor:** Permiten desplazar el cursor a izquierda, derecha, arriba y abajo, borrar un carácter o parte de una línea.
- **Teclado numérico:** Es habitual en los teclados de computadora que las teclas correspondientes a los caracteres numéricos (cifras decimales), signos de operaciones básicas (+, -, ...) y punto decimal estén repetidas para facilitar al usuario la introducción de datos numéricos.
- **Teclas de funciones:** Son teclas cuyas funciones son definibles por el usuario o están predefinidas en un programa.
- **Teclas de funciones locales:** Controlan funciones propias del terminal, como impresión del contenido de imagen cuando el computador está conectada a una impresora.

En algunos teclados la transmisión no se efectúa pulsación a pulsación sino que se dispone de un buffer y la transmisión se efectúa a la vez para todo un conjunto de mensajes completos cuando el usuario pulsa una tecla especial destinada a activar dicha transmisión. Esta tecla recibe distintos nombres como Return, Enter, Transmit, Intro.

Ratón o Mouse

Funcionamiento: El ratón o mouse es un pequeño periférico que está constituido por una bola que puede girar libremente, y se acciona haciéndola rodar sobre una superficie plana. Un ratón es un dispositivo pensado para apuntar y luego realizar dos operaciones:

Clic: es la operación que teniendo el ratón quieto permite elegir un objeto o una acción que previamente se ha apuntado en la pantalla.

Arrastrar: el objeto seleccionado puede ser arrastrado o utilizado para marcar una zona para luego tomar una acción específica.

En el momento de activar el mouse, se asocia su posición con la del cursor en la pantalla. Si desplazamos sobre una superficie el mouse, el cursor seguirá dichos movimientos.

Resumen: Final De Organización de Computadoras por M.E.Defalco, M.E.Corróns y D.De Sousa Días

Unidad VI: Periféricos

Periféricos

Pueden considerarse periféricos tanto aquellas unidades o dispositivos a través de los cuales el computador se comunica con el mundo exterior, como a los sistemas que almacenan o archivan la información, sirviendo de memoria auxiliar de la memoria principal.

La computadora es una máquina que no tiene sentido si no se puede comunicar con el exterior, es decir si careciera de los periféricos tales como:

- **Unidades de entrada:** para dar los programas que queremos ejecutar y los datos correspondientes.
Estas unidades transforman la información externa (instrucciones o datos tecleados) según alguno de los códigos de E/S. Así la computadora recibe dicha información en binario.
Ejemplos: Teclado, ratón, lápiz óptico, scanner, etc.
- **Unidades de salida:** nos da los resultados de los programas.
Estas unidades efectúan el proceso inverso a las de entrada, es decir, la información binaria que llega del computador se transforma de acuerdo con el código de E/S en caracteres escritos inteligibles por el usuario
Ejemplos: Monitor, impresora, plotter, etc.
- **Unidades de memoria masiva auxiliar (de E/S):** facilitan su funcionamiento y utilización.
Estas unidades transforman la información externa en señales codificadas, permitiendo su transmisión, detección, interpretación, procesamiento y almacenamiento de forma automática.
Ejemplos: Cinta magnética, disco magnético, CD-ROM, DVD.
- **Unidades de E/S mixtas:** Ejemplos: Terminal interactivo, Modem.

Otra clasificación:

- **Locales:** Se encuentran cerca del computador.
- **Remotos:** Si debido a su situación lejana la conexión hay que realizarla a través de líneas especiales de transmisión.

Las operaciones de E/S se realizan a través de una amplia gama de dispositivos. Un dispositivo externo se conecta a la computadora mediante un enlace a un módulo de E/S. El enlace se utiliza para intercambiar señales de control, estado y datos entre los módulos de E/S y el dispositivo externo (periférico).

Se pueden clasificar en:

- **Hombre-máquina:** Permite la comunicación con el usuario de la computadora. Esta comunicación se realiza a través de periféricos (monitor, etc.)
- **Máquina-máquina:** Permite la comunicación con elementos de equipo (discos magnéticos, cintas, etc.). La comunicación es digital. Si los computadores están cerca se conectan directamente en Banda Base, mediante un bus paralelo que transmite bytes o palabras.
Si existe una distancia de más de 10 o 100 mts. Se utilizan los **MODEM**.
- **Máquina-mundo físico:** Permite la comunicación con dispositivos remotos. En muchas aplicaciones, el computador se encuentra conectado directamente a un sistema físico exterior, sobre el que realiza funciones tales como adquisición de datos, control de procesos, control de instrumentación de laboratorios, etc. Para realizar esta conexión, se

emplean una serie de periféricos que se denominan de forma genérica periféricos de control.

Existen cuatro tipos de periféricos de control:

1. Cadena de lectura analógica
2. Cadena de lectura digital
3. Cadena de acción analógica
4. Cadena de acción digital

La conexión con el modulo de E/S se realiza a través de señales de control, estado y datos. Los datos se intercambian en forma de un conjunto de bits que son enviados/recibidos a/desde el modulo de E/S. Las señales de control determinan la función que debe realizar el dispositivo. Las de estado indican el estado del chip.

Maquinas auxiliares

Las maquinas auxiliares no están físicamente conectadas al computador (su funcionamiento es totalmente autónomo) y sirven para preparar o ayudar en la confección o utilización de la información que se da a, o produce, el computador.

Por ej. Hace algunos años existían maquinas autónomas para perforar tarjetas, para grabar cintas magnéticas manualmente a través de un teclado, etc.

Diferencia entre Soporte de información y periférico

Se entiende por soporte de información a todos aquellos medios físicos sobre los que va la información.

Por unidades o dispositivos periféricos se entiende a aquellos elementos encargados de transcribir la información al correspondiente soporte.

Conexión de periféricos al computador

Puertos normalizados

Permiten la conexión desde el exterior hacia la placa base.

Funcionan con las características estándar y su fabricante tiene que conocer las normas estándar para que no halla problemas de conexión con los periféricos

Las conexiones o puertos son:

- Los **puertos serie (COM)**. Transmiten la información BIT a BIT. No son adecuados para transferir grandes cantidades de información. En ellos se conectan el módem, cuando es externo.
- Los **puertos paralelo (LPT)**. Permiten una mayor velocidad en la transmisión de datos. La información se transmite en bytes o múltiplos. Es la conexión típica de impresoras o escáneres.
- Los **puertos USB (Universal Serial Bus)**. En ellos se conectan, en general, dispositivos que necesitan una alta velocidad de transferencia de datos como, por ejemplo, un escáner o una cámara digital, aunque también existen impresoras con este tipo de conexión.

Controladores (Interfaces)

Elemento entre dos fases que adapta las características diferentes entre las placas y los periféricos.

Existen también controladores por software, llamados Drivers, van a un nivel lógico y son complementarios a los anteriores.

Resumen: Final De Organización de Computadoras
por M.E.Defalco, M.E.Corróns y D.De Sousa Días

Unidad III: Lógica Digital

1) ALGEBRA DE BOOLE

La electrónica digital está fundamentada en la base matemática formada por el álgebra de Boole (George Boole, matemático inglés, 1815-1864). Este método considera que todos los elementos poseen únicamente dos estados (biestables) o dos valores, verdadero o falso (1 ó 0) que son opuestos entre sí, no permitiéndose nunca la adopción de estados intermedios. Estudiando las distintas asociaciones entre ellos se obtienen las leyes generales sobre los procesos lógicos.

Circuitos lógicos: Circuitos compuestos de elementos que solo pueden adoptar dos estados estables posibles.

Puede decirse entonces que el álgebra de Boole es el sistema matemático empleado en el diseño de circuitos lógicos, que nos permite identificar mediante símbolos el objeto de un circuito lógico de modo que su estado sea equivalente a un circuito real.

1.1) Operaciones

Se definen básicamente tres tipos de operaciones sobre las variables booleanas que son:

- La complementación
- La suma
- El producto.

La complementación

Sea una variable booleana A, si en un instante determinado posee el estado lógico 1, diremos que su estado inverso o complementado será el 0. Si por el contrario la variable A posee el estado lógico 0, su complemento será el 1. El complemento de una variable A se representa por: \bar{A} . La tabla de la verdad de los estados lógicos correspondientes a una variable y a su complementaria o inversa es la siguiente:

A	\bar{A}
0	1
1	0

La suma

La operación lógica suma entre dos o más variables booleanas se representa mediante el signo "+". Por tanto si tenemos $C = A + B$ suele leerse "C es igual a A o B". Esta operación se denomina también reunión de conjuntos. La función suma se define mediante la siguiente tabla de la verdad:

A	B	C
0	0	0
0	1	1
1	0	1
1	1	1

El resultado de esta operación es 1 lógico cuando la variable A, o la variable B valen 1. Este resultado puede generalizarse para "n" variables de entrada.

El producto

La operación producto entre dos conjuntos se representa mediante el símbolo *, Esta operación se denomina también intersección de conjuntos. Por tanto tendremos que $D = A * B$ representa un producto y se lee "D es igual a A y B." ($D=AB$). La operación producto se define mediante la siguiente tabla de la verdad:

A	B	C
0	0	0
0	1	0
1	0	0
1	1	1

El resultado del producto será 1 lógico si y solo si la variable A y la variable B son ambas 1.

1.2) Teoremas

Teorema 1. El resultado de aplicar cualquiera de las tres operaciones antes definidas, a variables booleanas, es otra variable booleana y además el resultado es único.

Teorema 2. Ley de idempotencia. Tanto la suma como el producto de una variable booleana consigo misma da como resultado la misma variable:

$$A + A = A$$

$$A * A = A$$

Teorema 3. Ley de involución. Una variable booleana negada dos veces, da como resultado la misma variable:

$$\bar{\bar{A}} = A$$

Teorema 4. Ley conmutativa. Se define respecto a la suma (y al producto) y nos dice que el orden de los sumandos (factores) no altera el resultado:

$$A + B = B + A$$

$$A \cdot B = B \cdot A$$

Teorema 5. Ley asociativa. Se define respecto a las operaciones suma y producto de la siguiente forma:

Respecto de la suma: $A + (B + C) = (A + B) + C = A + B + C$

Respecto del producto: $A(BC) = (AB)C = ABC$

Teorema 6. Ley distributiva.

Respecto de la suma: $A + BC = (A + B)(A + C)$

Respecto del producto: $A(B + C) = AB + AC$

Teorema 7. Ley de absorción.

$$A + AB = A$$

$$A(A + B) = A$$

Teorema 8. Leyes de De Morgan.

$$\sim(A + B) = \sim A \cdot \sim B$$

$$\sim(A \cdot B) = \sim A + \sim B$$

(Las leyes de De Morgan pueden ser generalizadas a "n" variables.)

A continuación se muestran algunas relaciones importantes que se deducen de las operaciones booleanas y de los teoremas anteriores:

$$0 + A = A$$

$$1 \cdot A = A$$

$$0 \cdot A = 0$$

$$1 + A = 1$$

$$\sim(\sim A) = A$$

Resumen: Final De Organización de Computadoras
por M.E.Defalco, M.E.Corróns y D.De Sousa Días

Unidad V: Memoria

1) CLASIFICACIÓN SEGÚN LAS CARACTERÍSTICAS

x Ubicación

- CPU → registros
- Interna → memoria principal
- Externa → memoria secundaria (dispositivos periféricos de almacenamiento)

x Capacidad

- Para memorias internas se expresa normalmente en bytes o palabras de longitud de 8, 16 y 32 bits (generalmente).
- Para memorias externas se expresa en bytes.

- Registros: 1KB
- Memoria principal: 128 MB
- Caché: 1 MB
- Discos: 40 GB

x Duración de la información

- Volátiles: la información se pierde o desaparece cuando se desconecta la alimentación. Ej: RAM.
- No volátiles: la información permanece grabada sin modificaciones, a menos que se realicen intencionalmente. Ej: discos, cintas.
- Permanentes: no pueden ser modificadas a menos que se destruya la unidad de almacenamiento. Ej: ROM, EPROM.

x Unidad de transferencia

- Palabras (para la memoria interna)
- Bloques, unidades más grandes que la palabra (para la memoria externa)

x Métodos de acceso

- Acceso secuencial: la memoria se organiza en unidades (registros). Para acceder a una se debe ir trasladando desde la posición actual a la deseada por todos los registros intermedios. Ej: cintas.
- Acceso directo: se accede directamente a una vecindad dad de registros o bloques, luego se hace una búsqueda secuencial contando o esperando hasta alcanzar la posición. Ej: discos.
- Acceso aleatorio (random): cada posición direccionable de memoria tiene un único mecanismo de acceso. El tiempo de acceso es constante e independiente de la secuencia de accesos previos (cualquier posición puede seleccionarse aleatoriamente y ser direccionada y accedida directamente). Ej: memoria principal y algunos sistemas de caché.
- Acceso Asociativo: es una memoria de tipo aleatorio, que permite comparar ciertas posiciones de bits dentro de una palabra buscando que coincidan con unos valores dados, y hacer esto para todas las palabras simultáneamente (esto hace que puedan ser recuperadas a través de una parte de su contenido y no de su dirección).

x Prestaciones

- Tiempo de acceso: para memorias aleatorias: tiempo en realizarse una escritura/lectura.
- para otras memorias: tiempo que se tarda en situar el mecanismo de lectura/escritura en la posición deseada.
 - tiempo de acceso de un registro: 1ns
 - tiempo de acceso de una caché: 1-20 ns
 - tiempo de acceso de la memoria principal: 60-80 ns
 - tiempo de acceso de un disco: 10ms
- Tiempo de ciclo de memoria: se aplica sólo en memorias aleatorias; es el tiempo de acceso y algún tiempo más (finalización de transiciones en las líneas de señal, ó regeneración de datos en caso de lecturas destructivas) que se requiere hasta que se pueda iniciar un segundo acceso a memoria.
- Velocidad de transferencia: velocidad a la que se pueden transferir datos a/desde la memoria. Para las memorias aleatorias la velocidad coincide con el inverso del tiempo de ciclo. Para otras memorias : el tiempo medio de escritura/lectura de N bits es igual al tiempo de acceso medio más el número de bits sobre la velocidad de transferencia en bps. ($T_n = T_a + (N/R)$).

x Dispositivos físicos

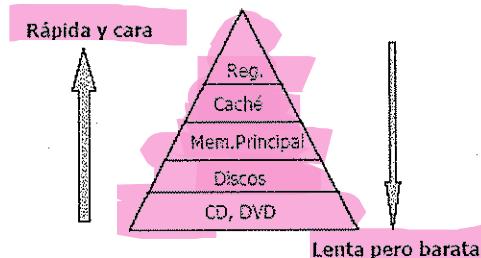
- semiconducto
- magnético
- óptico
- magnético-óptico

2) JERARQUÍA DE MEMORIA

Las tres características clave (coste, capacidad y tiempo de acceso) tienen una relación:

- a menor tiempo de acceso, mayor coste por bit.
- a mayor capacidad, menor coste por bit.
- a mayor capacidad, mayor tiempo de acceso.

Para que esto sea posible es necesario emplear una jerarquía de memoria (no sólo un componente de la misma), ya que para satisfacer las prestaciones se necesitan memorias costosas, de capacidad baja y alta velocidad (lo que se desea es utilizar tecnología de gran capacidad):



Cuando se desciende en la jerarquía:

- disminuye el coste por bit
- aumenta la capacidad
- aumenta el tiempo de acceso
- disminuye la frecuencia de accesos a la memoria por parte del procesador.

Así, se complementan memorias pequeñas, costosas y rápidas con otras grandes, económicas y lentas.

Al disminuir la frecuencia de acceso (por la agrupación de las referencias a memoria) es posible organizar los datos a través de la jerarquía, de manera que el porcentaje de accesos a cada nivel más bajo sea menor que el nivel anterior.

3) MEMORIA PRINCIPAL

3.1. Memorias semiconductoras:

- **RAM** (memoria de acceso aleatorio): es posible leer y escribir datos mediante señales eléctricas; es volátil (puede utilizarse como almacenamiento temporal). Se divide en:

· **dinámica**: hecha con celdas, que almacenan datos como cargas en transistores, que al descargarse requieren refrescos periódicos para mantener memorizados los datos; llamada **DRAM** (almacenan más información que SRAM en la misma superficie, ya que los transistores son más chicos que los flip-flops) y es utilizada como memoria principal.

· **estática**: los valores binarios se almacenan utilizando configuraciones de pueras que forman biestables (flip-flops); más rápidas, complejas y caras que las dinámicas, y no necesitan refresco; llamada **SRAM**, es más rápida que la DRAM y es utilizada como memoria caché.

- **ROM** (memoria de solo lectura): contienen un patrón de datos inalterables (que sólo se puede leer, no modificar ni escribir); no son volátiles; se construye como un chip de circuito integrado, con los datos cableados durante la fabricación (más caros y cualquier fallo provocaría el desecho de la memoria completa)

- **PROM** (memoria de solo lectura programable): no son volátiles, pueden grabarse una sola vez (incluso luego de la fabricación, con un equipo especial de grabación), son flexibles y cómodas.

- **Memorias de sobre-todo-lectura** (útil para las aplicaciones donde las lecturas son más frecuentes que las escrituras, son no volátiles)

· **EPROM** (memoria de solo lectura programable y borrable): se escribe y lee eléctricamente; antes de la escritura se deben borrar todas las celdas mediante la exposición del chip encapsulado a radiación ultravioleta; retiene su contenido indefinidamente, es más costosa que una PROM pero se puede actualizar múltiples veces.

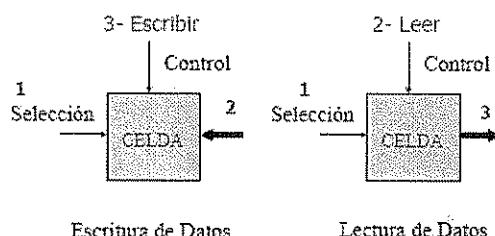
· **EEPROM** (memoria de solo lectura programable y borrable eléctricamente): se puede escribir en cualquier momento sin borrar su contenido anterior (actualizando el byte o bytes disecionados), son más costosas y menos densas que las EPROM.

· **Memoria Flash**: utilizan tecnología de borrado eléctrico de mucha más rapidez que las EPROM (todo o bloques concretos del chip, pero no a nivel de byte).

3.2. Organización de la memoria:

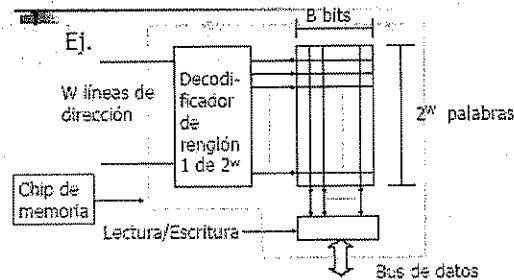
Elemento básico: **celda** de memoria, que presenta dos estados estables (que representan el 0 y 1), pueden ser escritas para fijar su contenido, pueden ser leídas para detectar su estado, posee tres terminales para transportar señales eléctricas:

1. de selección: selecciona la celda para la operación de escritura/lectura.
2. de control: indica el tipo de operación.
3. de lectura/escritura: proporciona la señal que fija el estado de la celda a 1 o 0/se utiliza como salida del estado de la celda.

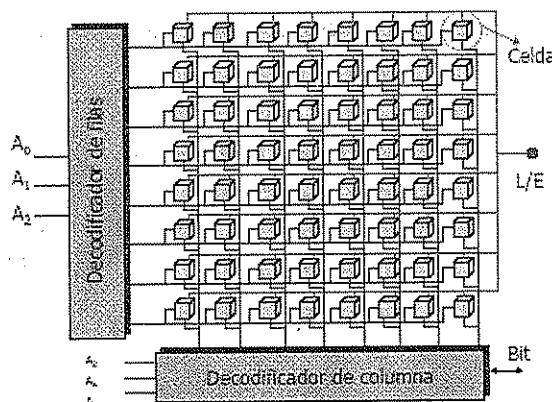


Una memoria de 1 bit la implementamos con flip-flop y armamos registros sencillos de n bits con flip-flops. Para armar memorias más grandes se requiere una organización diferente, en la cual sea posible direccionar palabras individuales: chip. Cada chip contiene un arreglo de celdas de memoria. Se han empleado dos enfoques organizacionales:

x Organización 2D: las celdas forman una matriz (todos los bits en el mismo chip) de 2^n filas y m columnas (2^n : número de palabras, m : número de bits de cada palabra). Cada fila es seleccionada por la decodificación de una configuración diferente de los n bits de dirección. La desventaja es que el decodificador (selector de palabras) crece exponencialmente con el tamaño de la memoria (ocupan mucha superficie)



x Organización 2 1/2D ó 3D: Se utilizan dos codificadores de $2^{(n/2)}$ operando en coincidencia. Las líneas de dirección se reparten entre los dos decodificadores. Para una configuración de las líneas de dirección se selecciona un único bit de la matriz. Existen varias matrices de celdas básicas, tantas como bits tenga la palabra de memoria (los bits de una misma palabra están en distintos chips), actuando en paralelo sobre ellas los circuitos de decodificación (los datos se leen/escriben por bits): disminuye la probabilidad de error.



3.3. Nuevas tecnologías RAM:

- **EDRAM (Enhanced DRAM):** Esta memoria incluye una pequeña memoria estática (SRAM) que almacena el contenido completo de la última fila leída. Con ello, ciertas peticiones pueden ser resueltas por esta rápida memoria, aumentando las prestaciones.

- **SDRAM (Synchronous DRAM [ó SDR SDRAM (Single Data Rate SDRAM)]):** 'Memoria RAM dinámica sincrónica de tasa de datos simple'; este tipo de memoria se conecta al reloj del sistema y está diseñada para ser capaz de leer o escribir a un ciclo de reloj por acceso, es decir, sin estados de espera intermedios. También incluye tecnología InterLeaving, que permite que la mitad del módulo empiece un acceso mientras la otra mitad está terminando el anterior (se presenta la dirección a la RAM, ésta encuentra los datos, SDRAM los mueve en tiempo de reloj del sistema, la CPU conoce cuándo los datos estarán listos y puede hacer otra cosa mientras tiene que esperar). SDRAM trabaja en bloques:

- **FPM-RAM (Fast Page Mode RAM):** Memoria asíncrona que incorpora un sistema de paginado debido a que considera probable que el próximo dato a acceder esté en la misma columna,

ganando tiempo en caso afirmativo; permite a la CPU acceder más rápido porque envía bloques enteros de datos.

- **BEDO-RAM** (*Burst Extended Data Output RAM*): Memoria asíncrona, variante de la anterior, es sensiblemente más rápida debido a que manda los datos en ráfagas (burst).

- **DDR SDRAM** (*Double Data Rate SDRAM*) [ó DDR]: 'Memoria de doble tasa de transferencia'; permite la transferencia de datos por dos canales distintos simultáneamente en un mismo ciclo de reloj. Los módulos DDRs soportan una capacidad máxima de 1 Gb.

- **RDRAM** (*Rambus DRAM*): introduce un módulo de memoria con niveles de rendimiento muy superiores a los módulos de memoria SDRAM. Tienen difusores de calor consistentes en una placa metálica que recubre los chips del módulo.

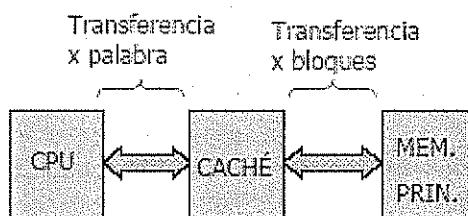
- **RAMLINK**: es una interfaz de memoria con conexiones punto a punto (forma de anillo, que permite la coordinar la actividad de muchas DRAM), cuyo tráfico se gestiona por un controlador de memoria. Los datos se intercambian en forma de paquetes.

- **CDRAM** (*Caché DRAM*): incluye una SRAM mayor que la que contiene la EDRAM. Puede utilizarse de dos modos: como una verdadera caché de X líneas de 64 bits (diferencia con la EDRAM que sólo contiene un bloque: la fila accedida recientemente) que es efectiva para accesos aleatorios a memoria usuales; y como buffer para proporcionar accesos series a un bloque de datos. (Ampliado en el punto 3.3.1)

3.3.1. Memoria Caché

Después que la CPU emite una solicitud de lectura a la memoria pasan muchos ciclos de reloj antes de reciba la palabra que necesita: llegar a la memoria por el bus del sistema es lento, por ello la solución sería combinar una cantidad pequeña de memoria rápida con una cantidad grande de memoria lenta, para obtener la velocidad de memoria casi rápida: **memoria caché**.

La caché contiene una copia de partes de la memoria principal; cuando el procesador intenta leer una palabra de memoria, se hace una comprobación para determinar si la palabra está en la caché; si es así, se entrega dicha palabra al procesador; sino, un bloque de memoria principal (con cierto número de palabras) se transfiere a la caché y, después, la palabra es entregada al procesador. Esto es: cuando se hace referencia a una palabra, ella y algunas de las vecinas se traen de la memoria grande y lenta a la caché.



Principios de utilización:

- **Principio de localidad espacial de referencia**: cuando se accede a una palabra de memoria, es muy probable que el próximo acceso sea en la vecindad de la palabra anterior. Ej: ejecución secuencial del código, tendencia de los programadores a hacer próximas entre sí variables relacionadas, acceso a estructuras tipo matriz o pila.
- **Principio de localidad temporal de referencia**: cuando se accede a una posición de memoria, es muy probable que en un lapso corto de tiempo, dicha posición de memoria sea accedida nuevamente. Ej: formación de ciclos o bucles, subrutinas, pilas.

La efectividad de la memoria caché se expresa a través de la frecuencia de aciertos de direcciones: un acierto sucede cuando los datos que necesita el procesador están almacenados en la caché (la CPU obtiene los datos a alta velocidad); cuando sucede lo contrario, se genera un fallo (la CPU tiene que obtener los datos de la memoria principal, a una velocidad menor). El porcentaje de aciertos es mucho mayor que el porcentaje de fallos (90% y 10%).

Elementos de diseño de caché:

• *Tamaño*: nos gustaría que el tamaño fuera lo suficientemente pequeño para que el coste total medio por bit se acerque al de la memoria principal, que fuera lo suficientemente grande como para que el tiempo de acceso medio total se acerque al de la caché. Cuanto más grande es, mayor es el número de puertas que direccionan la caché. Problemas: las cachés grandes tienden a ser lentas. Los tamaños óptimos sugeridos de caché se encuentran entre 1K y 512K palabras.

• *Función de correspondencia*: al haber menos líneas de caché que bloques de memoria principal, se necesita de un algoritmo que haga corresponderlos y de algún medio que determine qué bloque ocupa actualmente una línea de caché. Existen 3 técnicas de correspondencia:

- *directa*: simple y poco costosa de implementar; consiste en hacer corresponder cada bloque de memoria principal a sólo una línea posible de caché, lo que sería una desventaja ya que, si un programa referencia repetidas veces a palabras de dos bloques diferentes asignados en la misma línea, dichos bloques se estarían intercambiando continuamente en la caché, y la tasa de aciertos sería baja.
- *asociativa*: supera la desventaja de la anterior: cada bloque de memoria principal puede cargarse en cualquier línea de la caché. La lógica de control de la caché interpreta una dirección de memoria como una etiqueta y un campo (identifica únicamente un bloque de memoria) de palabra; para saber si un bloque está en la caché, la lógica de control examina simultáneamente todas las etiquetas de líneas para buscar una coincidencia. Esta correspondencia genera flexibilidad para que cualquier bloque sea reemplazado cuando se va a escribir uno nuevo en la caché (ver '*Algoritmos de sustitución*'), lo cual hace a la maximización de la tasa de aciertos. Desventaja: compleja circuitería necesaria para examinar en paralelo las etiquetas de todas las líneas de caché.
- *asociativa por conjuntos*: solución de compromiso que recoge lo positivo de las correspondencias directa y asociativa, sin sus desventajas.

• *Algoritmos de sustitución*: cuando se introduce un nuevo bloque en la caché, debe sustituirse uno de los bloques existentes (para la correspondencia asociativa, ya que en la directa no es posible). Estos algoritmos deben implementarse en hardware para conseguir una alta velocidad. Cuatro de los algoritmos probados son:

- *LRU (Last-Recent Used, Utilizado menos recientemente)*: se sustituye el bloque que se ha mantenido en la caché por más tiempo sin haber sido referenciado.
- *FIFO (First-In-First-Out, Primero en entrar-primer en salir)*: se sustituye aquel bloque del conjunto que ha estado más tiempo en la caché; puede implementarse mediante una técnica cíclica o buffer circular.
- *LFU (Least-Frequently Used, Utilizado menos frecuentemente)*: se sustituye aquel bloque del conjunto que ha experimentado menos referencias; podría implementarse asociando un contador a cada línea.
- *Técnica no basada en el grado de utilización*: toma una línea al azar entre las posibles candidatas; proporciona prestaciones sólo ligeramente inferiores a un algoritmo basado en la utilización (LRU, FIFO, LFU).

• *Política de escritura*: antes de que pueda ser reemplazado un bloque que está en una línea de caché, es necesario saber si ha sido alterado en caché pero no en memoria principal. Si no ha sido alterado puede escribirse sobre la línea de caché; sino, la memoria principal debe actualizarse. Problemas: más de un dispositivo puede tener acceso a la memoria principal (se pueden generar palabras no válidas); y, cuando varios procesadores se conectan al mismo bus y cada uno de ellos tiene su propia caché local, al modificarse una palabra de alguna de las cachés, podría invalidar una palabra de las otras. Técnicas de escritura:

- *Escritura inmediata*: la más sencilla; todas las operaciones de escritura se hacen, tanto en caché como en memoria principal, asegurando que el contenido de la memoria principal siempre es válido. Desventaja: genera tráfico sustancial a memoria.
- *Post-escritura*: técnica alternativa que minimiza las escrituras en memoria, con la cual las actualizaciones se hacen sólo en caché, ya que cuando el bloque es sustituido, es escrito en memoria principal sí, y sólo sí, las actualizaciones se han realizado. Problema: a veces hay porciones de memoria no válidas, entonces lo accesos por parte de los

secundarias más lentas, baratas y normalmente más grandes, que se usan para almacenar conjuntos de datos cuyo volumen es mucho mayor al que puede contener la memoria principal.

• *Tipos de memoria secundaria:*

- Discos magnéticos: discos flexibles (disquetes), discos rígidos (duros)
- RAID
- Discos ópticos: CD-ROM, CD-R, CD-RW, DVD
- Cintas magnéticas

4.1. Discos magnéticos

Un disco magnético es un plato circular (diámetro entre 5 y 10 pulgadas) construido con metal o plástico, cubierto por un material magnetizable. Los datos se graban en él y después se recuperan del disco a través de una bobina: 'cabeza'. Durante una operación de escritura/lectura, la cabeza permanece quieta mientras el plato rota bajo ella. El mecanismo de escritura se basa en el campo magnético producido por el flujo eléctrico que atraviesa la bobina. El mecanismo de lectura se basa en la corriente eléctrica que atraviesa la bobina, producida por un campo magnético que se mueve respecto a la bobina.

• *Organización y formato de los datos*

Los datos se organizan en un conjunto de anillos concéntricos en el plato ('pistas', entre 500 y 2000 por superficie) del mismo ancho que la cabeza. Las pistas adyacentes están separadas por bandas vacías, lo cual minimiza los errores debidos a desalineamientos de la cabeza o interferencias del campo magnético. Se suele almacenar el mismo número de bits en cada pista, esto implica que la densidad aumenta según se avanza de la pista más externa a la más interna. Los datos se transfieren en bloques, los cuales se almacenan en regiones ('sectores', entre 10 y 100 por pista) de longitud fija o variable. Para evitar imprecisiones, los sectores adyacentes se separan con intrapistas vacías (intersectores). Existen una forma para identificar las posiciones de los sectores (comienzo de la pista, principio y fin de cada sector), ésta se gestiona mediante datos de control extras que son usados sólo por el controlador del disco y no accesibles al usuario.

• *Características físicas que diferencian los distintos tipos de discos*

- Desplazamiento de cabezas:
 - cabeza fija: existe una cabeza por pista, todas se montan en un brazo rígido que se extiende a través de todas las pistas.
 - cabeza móvil: sólo una cabeza por cara del plato montada en un brazo móvil.
 - Transportabilidad de disco:
 - no extraíble: permanentemente montado en la unidad de disco (brazo, eje que hace rotar al disco y la electrónica necesaria para la E/S de datos binarios).
 - extraíble: puede ser quitado y sustituido por otro disco. Ventaja: es posible una cantidad de datos ilimitada con un número limitado de unidades de disco; además un disco puede ser utilizado en diversos computadores.
 - Superficies:
 - Doble superficie: cuando la cubierta magnetizable se aplica a ambas caras del plato.
 - Única superficie: cuando se aplica a una cara sola.
 - Platos:
 - Varios platos: apilados verticalmente, separados por una pulgada; disponen de varios brazos.
- Constituyen una unidad llamada 'paquete de disco'.
- Único
 - Mecanismos de la cabeza:
 - Contacto: la cabeza efectúa un contacto físico con el medio durante la operación de lectura/escritura (usado con los disquetes).

módulos E/S sólo podrían hacerse a través de la caché, lo que complica la circuitería y genera tráfico.

Posibles aproximaciones a la coherencia de caché (para evitar los problemas y desventajas):

- vigilancia del bus con escritura inmediata: cada controlador de caché monitorea las líneas de direcciones para detectar operaciones de escritura en memoria por parte de otros maestros del bus. Si otro maestro escribe en una posición de memoria compartida, que también reside en la memoria caché, el controlador invalida el elemento de la caché.
 - transparencia hardware: se utiliza hardware adicional para asegurar que todas las actualizaciones de memoria principal, vía caché, quedan reflejadas en todas las cachés: si un procesador modifica una palabra de su caché, esta actualización se escribe en memoria principal. Además se actualizan todas las palabras coincidentes de otras cachés.
 - memoria excluida de caché: sólo una porción de memoria principal se comparte por más de un procesador, y ésta se diseña, como no transferible a caché. Todos los accesos a la memoria compartida son fallos de caché. La memoria excluida de caché puede ser identificada utilizando lógica de selección de chip, o los bits más significativos de la dirección.
- *Tamaño de línea*: cuando se recupera y ubica en caché un bloque de datos, se recuperan la palabra y algunas palabras adyacentes. A medida que aumenta el tamaño de bloque, la tasa de aciertos primero aumenta debido al principio de localidad y más datos útiles son llevados a la caché. Sin embargo, la tasa de aciertos comenzará a decrecer cuando el tamaño de bloque aumente más y la probabilidad de utilizar la nueva información captada se haga menor que la de reutilizar la información que tiene que reemplazarse. Los bloques más grandes reducen el número de bloques que caben en la caché; a medida que un bloque se hace más grande, cada palabra adicional está más lejos de la requerida.

• *Número de cachés*: más recientemente, se ha convertido en una norma el uso de múltiples cachés. Aspectos de diseño: número de niveles de caché y el uso de caché unificada frente a caché partida. Con el aumento de densidad de integración, ha sido posible tener una caché en el mismo chip del procesador: 'caché on-chip', el cual reduce la actividad del bus externo del procesador y los tiempos de ejecución, e incrementa las prestaciones globales del sistema. Cuando la instrucción o dato requeridos se encuentran en la caché on-chip, se elimina el acceso al bus; los accesos a la caché on-chip se efectúan apreciablemente más rápidos que los ciclos de bus (el bus queda libre para realizar otras operaciones). La inclusión de una caché on-chip deja abierta la cuestión de si es deseable una caché externa ('off-chip'): los diseños más actuales incluyen ambas cachés → 'caché de dos niveles' : la caché interna: nivel 1 (L1), la caché externa: nivel 2 (L2). La razón por la cual se incluye una caché de nivel 2 es que, si no hay una caché L2 y el procesador hace una petición de acceso a una posición de memoria que no está en la caché L1, entonces el procesador debe acceder a la DRAM o la ROM a través del bus (lentitud + tiempos de acceso = bajas prestaciones). Si se utiliza una caché L2 SRAM, la información que falta puede recuperarse fácilmente; si la SRAM es rápida, los datos pueden accederse sin espera. El uso de un segundo nivel de caché mejora las prestaciones.

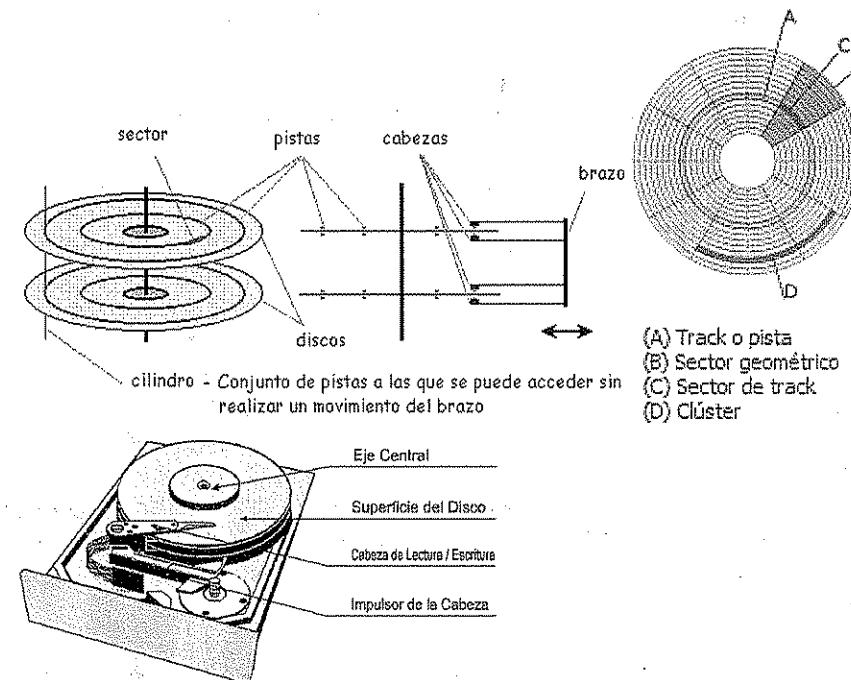
Se ha hecho normal separar la caché en dos: una dedicada a instrucciones y otra a datos.

Caché unificada: ventajas potenciales: tiene una tasa de aciertos mayor que una caché partida, ya que nivela automáticamente la carga entre captación de instrucciones y de datos, sólo se necesita diseñar e implementar una caché. A pesar de estas ventajas, la tendencia es hacia las cachés partidas: elimina la competición por la caché entre el procesador de instrucciones y la unidad de ejecución (bloqueando peticiones para terminar de realizar la que está ejecutándose; sobretodo en diseños que cuentan con segmentación de cauce de instrucciones), evitando degradaciones de prestaciones, realizar un uso eficiente del cauce segmentado de instrucciones.

4) MEMORIA SECUNDARIA

Todas las palabras de la memoria principal son accesibles directamente en un tiempo muy corto, por ello dicha memoria es cara; por ello las computadoras tienen también memorias

- Separación fija: la cabeza está a una distancia fija sobre el plato dejando una capa de aire.
- Separación aerodinámica (Winchester): la cabeza está en el contorno de una hoja de metal aerodinámica que reposa sobre la superficie del plato cuando no se mueve. Con la presión de aire generada por el giro del disco hace subir la hoja encima de la superficie.



• Parámetros para medir las prestaciones de un disco

Cuando una unidad de disco funciona, el disco rota a una velocidad constante. Para leer/escribir, la cabeza debe posicionarse en la pista deseada y al principio del sector deseado. En un sistema de cabeza móvil, el tiempo que tarde en posicionarse en la pista es el 'tiempo de búsqueda'. Una vez seleccionada la pista, se debe esperar hasta que el sector apropiado rote hasta alinearse con la cabeza; este tiempo que pasa se conoce como 'retardo rotacional', 'tiempo de latencia' ó 'latencia rotacional'. El tiempo que se tarda en llegar a la posición de lectura/escritura, 'tiempo de acceso', se conoce como la suma del 'tiempo de búsqueda' más 'tiempo de latencia'. Una vez posicionada la cabeza, se lleva a cabo la operación de lectura/escritura, lo que conlleva un tiempo de transferencia de datos. Por ende el TIEMPO TOTAL sería el 'tiempo de acceso' sumado al tiempo de transferencia de datos.

× tiempo de búsqueda = constante x pistas atravesadas + tiempo de comienzo

× tiempo de latencia = medio: 8,3 ms. en discos, entre 100 y 200 ms. en disqueteras

× tiempo de transferencia = bytes a transferir / (bytes de una pista x velocidad de rotación en revoluciones por segundo)

× tiempo de acceso medio total = tiempo e/ pistas x cantidad de pistas + tiempo de medio giro
2

• Capacidad del disco

$$\text{bytes} \times \text{sectores} \times \text{pistas} \times \text{nº de superficies}$$

sector pistas superficie

4.2. RAID (*Redundant Array of Independent Disks* – conjunto redundante de discos independientes)

Hace referencia a un conjunto de discos que operan independientemente y en paralelo. Con varios discos, las peticiones separadas de E/S se pueden gestionar en paralelo, siempre que los datos requeridos residan en discos separados. Además, se puede ejecutar en paralelo una única petición de E/S si el bloque de datos al que se va a acceder está distribuido a lo largo de varios discos.

Con el uso de varios discos, hay una amplia variedad de formas en las que se pueden organizar los datos y en las que se puede añadir redundancia para mejorar la seguridad.

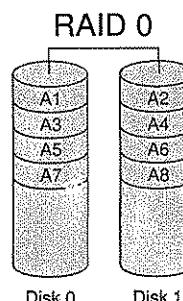
El esquema RAID consta de 6 niveles independientes que designan métodos diferentes que poseen tres características comunes:

- RAID es un conjunto de unidades físicas de disco vistas por el sistema operativo como una única unidad lógica.
- Los datos se distribuyen a través de las unidades físicas del conjunto de unidades.
- La capacidad de los discos redundantes se usa para almacenar información de paridad que garantice la recuperación de los datos en caso de fallo de disco.

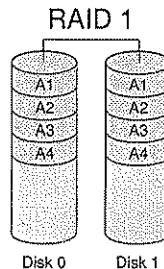
La estrategia RAID reemplaza una unidad de disco de gran capacidad por unidades de menor capacidad y distribuye los datos de forma que se puedan habilitar accesos simultáneos a los datos de varias unidades mejorando las prestaciones de E/S. Además de permitir que varias cabezas operen simultáneamente, con el incremento de los fallos, RAID utiliza la información de paridad almacenada que permite la recuperación de datos perdidos.

Niveles

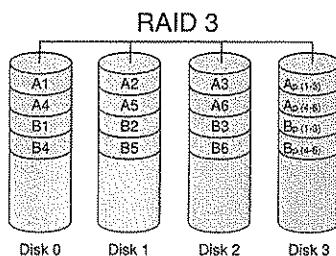
- **RAID 0:** no incluye redundancia para mejorar las prestaciones (por eso no es un verdadero miembro de la familia RAID), hay algunas aplicaciones en las que las prestaciones y la capacidad son las preocupación primaria, y un costo bajo es más importante que mejorar la seguridad; los datos del usuario y del sistema están distribuidos a lo largo de todos los discos del conjunto. Los datos son organizados en forma de tiras de datos a través de los discos disponibles, un conjunto de tiras lógicamente consecutivas, que se proyectan exactamente sobre una misma tira en cada miembro del conjunto se denomina 'franja'. Ventaja: si una única petición de E/S implica a varias tiras lógicas contiguas, éstas se pueden gestionar en paralelo, reduciendo el tiempo de transferencia de E/S.
 - × RAID 0 para alta capacidad de transferencia de datos: debe existir una capacidad de transferencia alta en todo el camino entre la memoria del anfitrión y las unidades de disco individuales. La aplicación debe hacer peticiones de E/S que se distribuyan eficientemente sobre el conjunto de discos.
 - × RAID 0 para altas frecuencias de petición de E/S: puede haber cientos de peticiones de E/S por segundo. Un conjunto de discos puede proporcionar velocidades altas de ejecución de E/S, balanceando la carga de E/S (si hay varias peticiones de E/S pendientes) a través de los distintos discos. Si la franja es relativamente grande, las peticiones de E/S que están esperando pueden ser tratadas en paralelo, reduciendo el tiempo en cola para cada petición.



- **RAID 1:** la redundancia se logra con el sencillo recurso de duplicar todos los datos, se hace una distribución de datos, cada franja lógica se proyecta en dos discos físicos separados. Puede conseguir altas velocidades de petición de E/S si la mayor parte de las peticiones son lecturas. Aspectos positivos:
 - una petición de lectura puede ser servida por cualquiera de los discos que contienen los datos pedidos.
 - una petición de escritura requiere que las dos tiras correspondientes se actualicen, y esto se puede hacer en paralelo; el resultado de la escritura viene determinada por la menos rápida de las dos escrituras.
 - la recuperación tras un fallo es sencilla.
 Desventaja: el coste requiere el doble del espacio de disco del disco lógico, por ello la RAID 1 está limitada a unidades que almacenan el software del sistema y los datos, y otros ficheros altamente críticos. Así, RAID proporciona una copia de seguridad en tiempo real de todos los datos para que, en caso de fallo de disco, todos los datos críticos están inmediatamente disponibles.

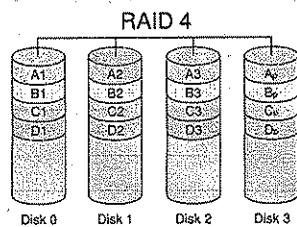


- **RAID 2:** utiliza una técnica de acceso paralelo. En un conjunto de acceso paralelo, todos los discos miembros participan en la ejecución de cada petición de E/S. El giro de cada unidad individual está sincronizado. Las tiras son muy pequeñas. Requiere menos discos que RAID 1, pero aún es bastante cara. En una sola lectura se accede a todos los discos simultáneamente. La operación de escritura debe acceder a todos los discos de datos y de paridad.
- **RAID 3:** similar a RAID 2, pero requiere sólo un disco redundante, utiliza un acceso paralelo, con datos distribuidos en pequeñas tiras.
 - × Redundancia: en el caso de un fallo en una unidad, se accede a la unidad de paridad y se reconstruyen los datos desde el resto de los dispositivos. Caso de que un disco falle, todos los datos estarán disponibles en modo reducido (para lecturas), y se requiere que ese disco se reemplace y se regenere todo su contenido en el nuevo disco.
 - × Prestaciones: velocidades de transferencia de datos muy altas: grandes transferencias; bajo rendimiento: sólo se puede ejecutar a la vez una petición de E/S.

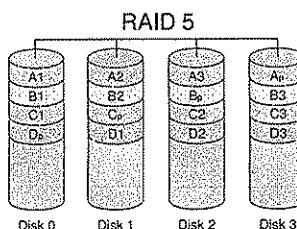


- **RAID 4:** utiliza una técnica de acceso independiente. Cada disco opera independientemente, de forma que peticiones de E/S separadas se atienden en paralelo (más adecuados para aplicaciones que requieren velocidades de petición de E/S altas, que para las que requieren velocidades altas de transferencia de datos). Las tiras son

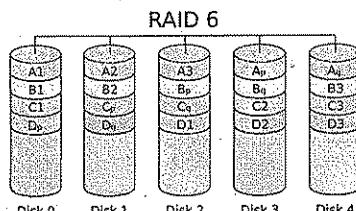
grandes. Cada vez que se realiza una escritura, el software de gestión del conjunto actualiza los datos del usuario y los bits de paridad correspondientes (generando tráfico).



- **RAID 5:** similar a RAID 4, pero distribuye las tiras de paridad a lo largo de todos los discos (evitando tráfico).



- **RAID 6:** se hacen dos cálculos de paridad distintos, que se almacenan en bloques separados en distintos discos. Posibilita la regeneración de los datos (incluso si dos de los discos que contienen los datos de los usuarios fallan). Ventaja: proporciona alta disponibilidad de datos.



4.3. Discos ópticos

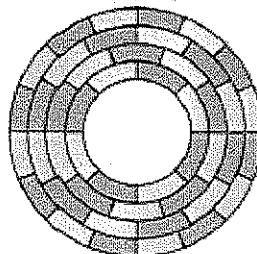
El CD (disco compacto digital de audio) es un disco no borrable, que puede almacenar más de 60 minutos de información de audio en una cara. Su éxito comercial posibilitó el desarrollo de la tecnología de memoria óptica de bajo coste. Se han introducido una gran variedad de estos discos:

- CD
- CD-ROM
- DVD
- WORM
- Disco óptico borrible
- Disco magnético-óptico

4.3.1. CD-ROM ('Compact Disc read-only memory' – memoria de disco compacto de solo lectura)

Tanto el CD de audio como el CD-ROM comparten una tecnología similar. La principal diferencia es que los lectores de CD-ROM son más robustos y tienen dispositivos de corrección de errores. El disco se forma a través de una resina, como el policarbonato, y se cubre con una superficie altamente reflectante (normalmente aluminio). La información digital se graba como una serie de hoyos microscópicos en la superficie reflectante. La superficie con los hoyos se protege con una capa final de laca transparente. La información se recupera con un láser de baja potencia,

el cual pasa a través de la capa protectora transparente mientras un motor hace girar el disco sobre el láser. La intensidad de la luz reflejada cambia si se encuentra en un hoyo. Un fotosensor detecta este cambio que se convierte en una señal digital. La información se empaqueta con densidad uniforme a lo largo del disco en segmentos del mismo tamaño y se explora a la misma velocidad, rotando el disco a una velocidad variable. Se dice que el láser lee los hoyos a una velocidad lineal constante (CLV).



Grabación en zonas

Los datos de un CD-ROM se organizan en una secuencia de bloques, que constan de:

- Sincronización: identifica el principio del bloque
- Cabecera: contiene la dirección del bloque y el byte de modo:
 - modo 0: campo de datos en blanco.
 - modo 1: uso de código de corrección de errores y 2.048 bytes de datos.
 - modo 2: 2.336 bytes de datos de usuario sin código de corrección de errores.
- Datos: datos del usuario.
- Auxiliar: - modo 2: datos del usuario adicionales.
 - modo 1: código de corrección de errores de 288 bytes.

Los CD-ROM son apropiados para la distribución de grandes cantidades de datos a un gran número de usuarios. Ventajas respecto con los discos magnéticos:

1. la capacidad de almacenamiento es mucho mayor.
2. se puede replicar en grandes cantidades de forma barata (los discos magnéticos se reproducen copiando uno a uno, usando dos unidades de disco).
3. es extraíble, puede ser usado como memoria de archivo (la mayoría de los discos magnéticos no son extraíbles, y la información que contiene tiene que ser copiada en una cinta antes de que se pueda almacenar una nueva).

Desventajas:

1. es de sólo lectura y no se puede actualizar.
2. el tiempo de acceso es mayor que el de una unidad de disco magnético, tanto como medio segundo.

Datos:

Velocidad lineal constante: 1,2 m/seg

Pista en espiral: 5,6 KM de largo (4666 seg = 77,4 minutos)

Velocidad angular constante: 200 a 530 rpm

Velocidad de reproducción: 75 sectores/ 1 segundo

Capacidad de un CD-ROM:

2 KB/sec x 75 sec/seg x 60 seg/min x 77,4 min ~ 700 MB

4.3.2.DVD ('Digital Video Disk' – disco vídeo digital)

Sustituye a las cintas VHS de vídeo analógicas con su gran capacidad de almacenamiento, y sustituye al CD-ROM en los PC y servidores: el DVD lleva al video a la edad digital. Proporciona películas con una calidad de imagen impresionante; se pueden acceder aleatoriamente como en los CDs. Se puede grabar un gran volumen de datos (7 veces más que un CD-ROM). Claves que los diferencia de los CD-ROM:

1. un DVD estándar almacena 4,7 GB por capa; los de doble capa y una cara almacenan 8,5 GB.
2. utiliza un formato de compresión (MPEG) para imágenes de pantalla completa de alta calidad.
3. un DVD de una capa puede almacenar una película de dos horas y media; uno dual una película de más de cuatro horas.

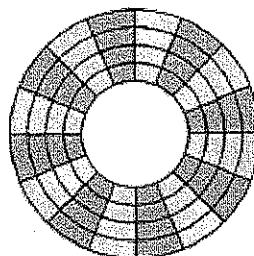
4.3.3. WORM

Es un CD de una-escritura-varias-lecturas, para las aplicaciones en las que sólo se necesitan unas pocas copias de un conjunto de datos. Para proporcionar un acceso más rápido, el WORM usa *velocidad angular constante**, sacrificando parte de su capacidad.

Técnica típica para fabricar el disco: usar un láser de alta potencia para producir una serie de ampollas en el disco; así un láser de baja potencia puede producir calor suficiente para reventar las ampollas pregrabadas.

Este disco óptico es atractivo para almacenar archivos de documentos y ficheros; proporciona una grabación permanente de grandes cantidades de datos del usuario.

* *Velocidad angular constante (CAV): incrementa el espacio lineal entre bits de información grabados en los segmentos más externos del disco; explorando la información a la misma velocidad, girando el disco a una velocidad fija.*



Grabación en CAV

4.3.4. Disco óptico borrible

Se puede escribir y reescribir repetidamente: utiliza el sistema magnético-óptico (usa la energía de un haz de láser y un campo magnético, para grabar y borrar información, invirtiendo los polos magnéticos) que al no afectar físicamente el disco se puede repetir varias veces. Usan velocidad angular constante. Ventaja: puede ser reescrito, por ende, ser usado como una verdadera memoria secundaria. Comparación con los discos magnéticos:

- Alta capacidad
- Intercambiabilidad: se pueden extraer de la unidad de disco.
- Seguridad: las tolerancias de construcción son mucho menos severas.

4.3.5. Disco magnético-óptico

La información se escribe en el disco usando un láser para calentar un pequeño punto de su superficie, y luego aplicando un campo magnético. Cuando el punto se enfria, adopta la polaridad norte-sur. Como no genera cambios físicos en el disco, esto se puede repetir varias veces.

La información se lee en el disco óptimamente: la dirección de la magnetización se puede detectar con un haz de láser polarizado (con una potencia menor a la de escritura), que refleja una luz, la cual es detectada por un sensor (similar a los lectores de discos compactos). Ventaja en comparación con los discos ópticos:

- longevidad: no se produce una degradación gradual a causa de las repetidas escrituras de los datos.
- ofrece un coste por MB inferior que en el almacenamiento magnético.

4.4. Cinta magnética

Los sistemas de cinta usan la misma técnica de lectura y grabación que los discos. El medio es una cinta de plástico flexible, cubierta por un óxido magnético, análoga a una cinta de grabación doméstica. Se estructura en un pequeño número de pistas paralelas. Las primeras usaban 9 pistas (almacén de datos de un byte, con un bit de paridad); las nuevas usan 18 o 36 pistas (una palabra o doble palabra digital); los datos se leen y escriben en bloques contiguos (registros físicos) que están separados por bandas vacías (bandas inter-registros).

Es un dispositivo de acceso secuencial (tiene que leer los registros uno a uno para llegar al requerido; tiene que rebobinar un cierto trecho y empezar a leer si el registro deseado está en alguna posición anterior) que está en movimiento sólo durante las operaciones de lectura/escritura. Los extremos contienen unas marcas metálicas pegadas denominadas BOT y EOT para la detección automática del inicio y fin de la cinta, respectivamente.

La capacidad depende de su longitud, densidad de grabación, longitud de bloque y formato de grabación.

Las cintas magnéticas fueron el primer tipo de memorias secundarias.

Tipos de unidades de cinta magnética

- *Cintas tradicionales de columnas de vacío*: las columnas de vacío tienen como objetivo mantener constante la tensión de la cinta bajo la estación de lectura/grabación.
- *Cintas tradicionales de brazos tensores*: Son más sencillas ya que no necesitan columnas de vacío pero con ellas obtiene menor velocidad.
- *Unidades de cassette de audio*: se utilizan en microcomputadoras domésticas y pequeños sistemas informáticos; el movimiento de la cinta se realiza con motores que actúan directamente sobre los carretes, no siendo tan rápidos ni precisos como las dos anteriores.
- *Unidades de cassette digitales*: disponen de cabestrantes para control de la velocidad de lectura/grabación.
- *Cartuchos*: unidades diseñadas para copias de seguridad de unidades de disco; la densidad de grabación es muy elevada y son de tamaño pequeño; tipos:
 - × De arranque/parada: intercambian bloques de información de longitud similar a las cintas tradicionales.
 - × De bobinado continuo: intercambian bloques de gran longitud.

