



UNIVERSIDAD TECNOLÓGICA DE PANAMÁ

FACULTAD DE INGENIERÍA DE SISTEMAS COMPUTACIONALES

DEPARTAMENTO DE ARQUITECTURA Y REDES DE COMPUTADORAS

FOLLETO DE ASIGNATURA ORGANIZACIÓN Y ARQUITECTURA DE COMPUTADORAS I

PROFESOR VLADIMIR VILLARREAL

2020



Esta obra está licenciada bajo la Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional. Para ver esta licencia:

<https://creativecommons.org/licenses/by-nc-sa/4.0>

Fuente del documento UTP-Ridda2:

<http://ridda2.utp.ac.pa/handle/123456789/13354>

CONTENIDO

1. Componentes generales del computador	6
1.1 Esquema general.	6
1.1.1 Unidades funcionales	6
1.1.2 Proceso de Arranque – POST (“Power On Self Test”).	8
1.2 Unidad Central de Procesamiento – CPU	8
1.2.1 Unidad de Ejecución	8
1.2.1.1 Registros generales.	9
1.2.1.2 Unidad de control	11
1.2.1.3 Unidad Aritmética Lógica.	13
1.2.1.4 Registro de Estados (banderas).	14
1.3 Elementos Internos	15
1.3.1 La tarjeta madre	15
1.3.1.1 Tipos (AT, ATX, LPX, NLX).	18
1.3.1.2 Elementos integrados	22
1.3.2 Los microprocesadores	24
1.3.2.1 Arquitectura CISC y RISC.	25
1.3.2.2 Familias.	29
1.3.3 Tarjetas de expansión.	31
1.3.3.1 Gráficas.	31
1.3.3.2 Sonido.	31
1.3.3.3 Comunicación – Red.	32
1.3.3.4 Controladoras.	33
1.3.4 Control de Energía	33
1.3.4.1 Fuente de Poder	33
1.3.4.2 Ventilador y disipador.	34
Actividades	35
Bibliografía	37
2. Interfaz del bus	39
2.1 Interconexión con los buses.	39
2.1.1 Jerarquía de buses.	40
2.1.1.1 Arquitectura tradicional.	41
2.1.1.2 Arquitectura de altas prestaciones	42
2.1.2 Estructura del bus.	44
2.1.2.1 Líneas de datos.	44
2.1.2.2 Líneas de dirección	45
2.1.2.3 Líneas de control	45
2.2 Tipos de interfaces.	46
2.2.1 Serial.	46
2.2.2 Paralela.	47
2.2.3 SCSI.	48
2.2.4 PCI	49
2.2.5 USB.	51

2.2.6 FireWire.	53
2.2.7 AGP	55
2.2.8 Otros	57
Actividades	63
Bibliografía	64
3. Memoria interna y externa	
3.1 Memoria Interna.	66
3.1.1 Características.	66
3.1.1.1 Ubicación.	66
3.1.1.2 Capacidad.	67
3.1.1.3 Unidad de transferencia.	68
3.1.1.4 Métodos de acceso.	68
3.1.1.5 Prestaciones.	69
3.1.1.6 Soporte físico.	70
3.1.2 Jerarquía de memoria.	71
3.1.3 Semiconductores de acceso aleatorio.	73
3.1.3.1 RAM (FPM, EDO, BEDO, SRAM, DRAM, SDRAM, DDR-RAM).	73
3.1.3.2 ROM (PROM, EPROM, EEPROM, FLASH).	75
3.1.3.3 Organización.	77
3.1.3.4 Lógica del chip.	78
3.1.4 Memoria Caché.	79
3.1.4.1 Principios.	81
3.1.4.2 Elementos de diseño.	83
3.1.4.3 Correspondencia.	84
3.1.4.4 Política de escritura.	86
3.1.4.5 Tamaño de línea.	87
3.1.4.6 Número de cachés.	87
3.2 Memoria Externa.	88
3.2.1 Discos magnéticos.	89
3.2.1.1 Definición.	89
3.2.1.2 Organización y formato.	90
3.2.1.3 Características.	93
3.2.1.4 Parámetros para medir prestaciones.	95
3.2.1.5 Conjunto redundante de discos independientes (RAID).	96
3.2.2 Cintas magnéticas.	98
3.2.3 Memoria Óptica.	98
3.2.3.1 CD (discos compactos).	98
3.2.3.2 Dispositivos (CD-ROM, CD-RW, DVD, WORM).	99
3.2.4 Nuevas tecnologías.	101
3.2.4.1 Unidad zip.	102
3.2.4.2 Memory stick.	102
3.2.4.3 Discos portátiles USB (flash drive).	103
3.2.4.4 Mini disc.	103
Actividades	104
Bibliografía	106

4. Dispositivos externos – periféricos.	108
4.1.1 Categorías.	109
4.1.2 Señales de comunicación.	109
4.1.2.1 Control.	110
4.1.2.2 Estado.	111
4.1.2.3 Datos.	112
4.1.3 Tipos básicos.	113
4.1.3.1 Teclado.	113
4.1.3.2 Ratón.	115
4.1.3.3 Monitor.	116
4.1.3.4 Impresora.	119
4.1.3.5 Escáner.	121
4.1.3.6 Bocinas y micrófono.	123
4.1.3.7 Otros.	123
Actividades	127
Bibliografía	129

CAPÍTULO 1: COMPONENTES GENERALES DEL COMPUTADOR

Objetivos:

- ✓ Comprender cuales son los elementos básicos de un computador según la arquitectura Von-Neumann.
- ✓ Entender cuáles son las operaciones básicas que debe realizar un computador para ejecutar un programa.
- ✓ Estudiar los componentes básicos de la CPU.
- ✓ Distinguir los componentes de una placa base.

¿De qué se trata esta sección de aprendizaje?

Este capítulo trata sobre la estructura del computador. El objetivo es presentar, tan claro como sea posible, la naturaleza y las características de los computadores de hoy en día, resaltando los atributos que impactan directo en la ejecución lógica de un programa. Se mostrará la organización de las unidades funcionales y sus interconexiones, que dan lugar a especificaciones arquitectónicas.

1. COMPONENTES GENERALES DEL COMPUTADOR.

1.1 Esquema general.

La mayoría de las computadoras actuales de propósito general presentan una estructura interna basada en la arquitectura definida por John Von Neumann. Esta estructura interna debe contener aquellos componentes que permitan realizar el procesamiento de datos útiles para el problema a resolver.

Dado que se utilizará un programa que controlará la sucesión de pasos a seguir, será necesario no solamente tener una unidad de cálculo sino también una unidad de memoria.

Podrá también, ser necesario interactuar con el mundo exterior, tanto para obtener datos como para entregar resultados, por lo que unidades que se encarguen de la entrada y la salida de valores podrán estar presentes.

Teniendo en cuenta lo anteriormente expresado, podemos esquematizarla de la siguiente manera:

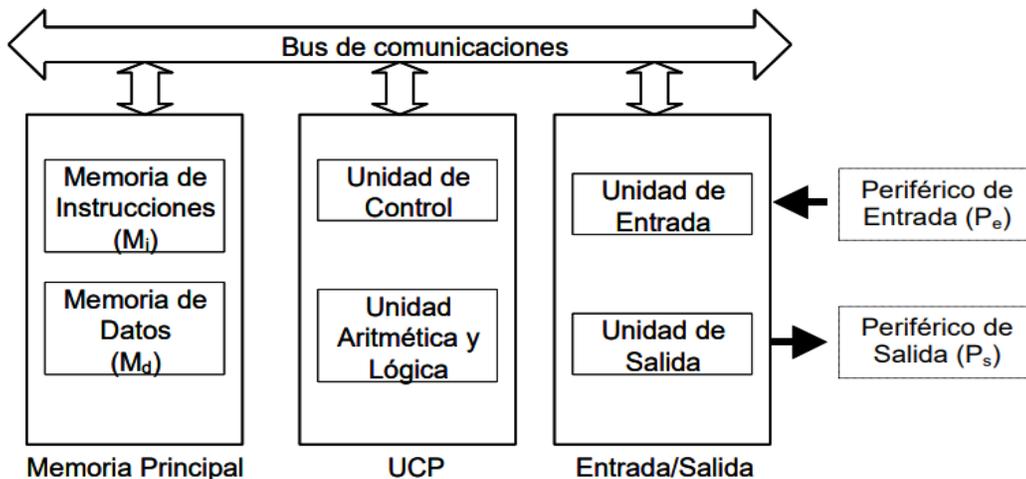


Figura 1.1: Esquema general de la arquitectura de un computador

1.1.1 Unidades funcionales

Las funcionalidades generales de un computador son:

- Procesamiento de datos
- Almacenamiento de datos
- Transferencia de datos
- Control

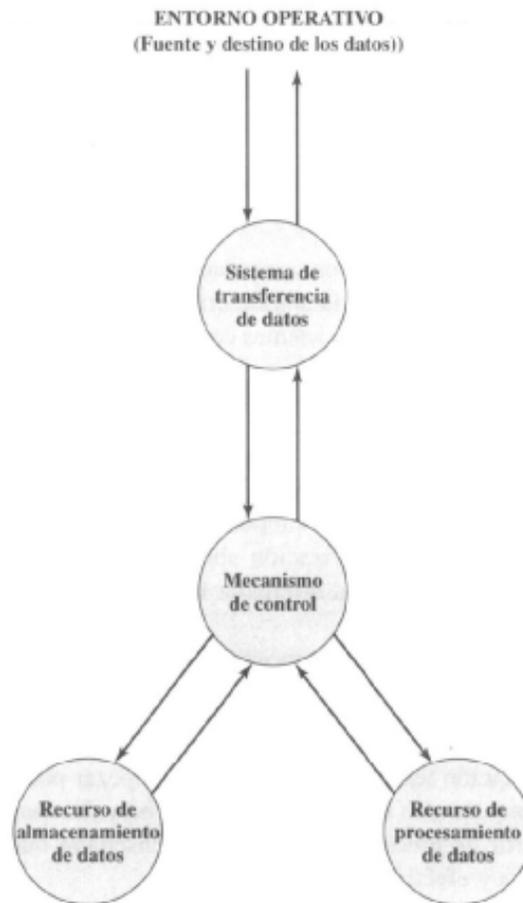


Figura 1.2: Visión funcional de un computador

El computador, tiene que ser capaz de procesar datos. Los datos pueden adoptar una gran variedad de formas, y el rango de los requisitos de procesado es amplio. También es esencial que un computador almacene datos. Incluso si el computador está procesando datos al vuelo (es decir, los datos se introducen, se procesan, y los resultados se obtienen inmediatamente), el computador tiene que guardar temporalmente al menos aquellos datos con los que está trabajando en un momento dado. Así hay al menos una función de almacenamiento de datos a corto plazo. Con igual importancia el computador lleva a cabo una función de almacenamiento de datos a largo plazo. El computador almacena ficheros de datos para que se recuperen y actualicen en el futuro.

El computador tiene que ser capaz de transferir datos entre el mismo y el mundo exterior. El entorno de operación del computador se compone de dispositivos que sirven bien como fuente o bien como destino de datos. Cuando se reciben o se

llevan datos a un dispositivo que está directamente conectado con el computador, el proceso se conoce como entrada- salida (E/S), y ese dispositivo recibe el nombre de periférico. El proceso de transferir datos a largas distancias, desde o hacia un dispositivo remoto, recibe el nombre de comunicación de datos.

Finalmente, debe haber un control de estas tres funciones. Este control es ejercido por el (los) entes que proporcionan al computador instrucciones. Dentro del computador, una unidad de control gestiona los recursos del computador y dirige las prestaciones de sus partes funcionales en respuesta a estas instrucciones.

1.1.2 Proceso de Arranque – POST (“Power On Self Test”).

Tras encender el computador, el procesador (CPU) entra en un ciclo inalterable en el que solicita instrucciones que ejecutar. Las primeras, que provienen de la memoria ROM, indican cómo empezar a usar los circuitos que operan el teclado, la pantalla, la memoria RAM y el disco duro. Esta versión preliminar del sistema operativo se conoce BIOS (Basic Input/Output System). La BIOS efectúa una comprobación o autodiagnóstico de la memoria, el procesador y del hardware detectado. Esta prueba se le conoce como POST y permite detectar un posible problema de los componentes físicos o una mala configuración de la BIOS. Cuando la fase inicial se completa el resto del sistema operativo se recupera desde la memoria secundaria, generalmente el disco duro.

El procedimiento desde que se oprime el botón de encendido hasta que el sistema operativo toma el control de la computadora, es conocido como proceso de arranque.

1.2 Unidad Central de Procesamiento – CPU

La CPU (Unidad Central de Procesamiento) es el lugar en donde se manipulan los datos, se conoce como el cerebro de la computadora, la CPU contiene un pequeño chip llamado microprocesador. Cada CPU está formada por al menos de dos partes básicas: la unidad de control y la unidad lógica-aritmética.

1.2.1 Unidad de Ejecución

Las funciones internas del microprocesador están divididas lógicamente en dos unidades de procesamiento: la BIU y la EU. Estas unidades pueden interactuar directamente, realizando las operaciones asincrónicamente.

La Unidad de Interfaz al bus (BIU) cumple las siguientes funciones:

- Busca las instrucciones en la memoria.
- Llena la cola de instrucciones que consta de 4 bytes (FIFO).
- Proporciona el control del bus.
- Proporciona a la EU los operandos a procesar.

La Unidad de ejecución (EU) recibe la instrucción buscada previamente por la BIU (Unidad de Interfaz al bus) y se encarga de ejecutarla.

La EU consta de la Unidad de Control (CU), la Unidad Lógico Aritmética (ALU) y un conjunto de 9 registros de 16 bits.

Cuando la BIU detecta que el bus externo está ocioso, envía una solicitud a la memoria para leer el siguiente byte en el flujo de instrucciones. Los bytes leídos son almacenados temporariamente en la cola de instrucciones. Cuando la EU requiere un nuevo byte de instrucción, lo toma de esta cola. La dimensión de cuatro bytes de la cola responde al compromiso de que la EU no tenga que estar esperando por un nuevo byte por un lado. Y, por otro lado, colas demasiado largas ocuparían mucho al bus llenándose con bytes que podrían no utilizarse (por ejemplo, cuando se ejecuta una instrucción de salto).

1.2.1.1 Registros generales.

Los registros son empleados para depositar transitoriamente operandos (datos) y direcciones de operandos.

Algunos registros son de propósito general (o registros multipropósito), mientras que otros tienen propósitos especiales. A continuación, se describen los registros multipropósito.

Tabla 1.1: Registros generales

Registro	Descripción
EAX (Acumulador)	Se define como un registro de 32 bits (EAX), como un registro de 16 bits (AX) o como uno de dos registros de 8 bits (AH y AL). Si se direcciona un registro de 8 o de 16 bits, sólo cambia esa porción del registro de 32 bits sin afectar a los bits restantes. El acumulador se emplea para instrucciones como multiplicación, división y algunas de las instrucciones de ajuste. En procesadores 80386 y superiores, el registro EAX puede almacenar también la dirección de desplazamiento de una posición e en el sistema de memoria.
EBX (índice base)	Éste puede direccionarse como EBX, BX, BH o BL. Algunas veces guarda la dirección de desplazamiento de una posición en el sistema de memoria, en todas las versiones del microprocesador. También permite direccionar datos de la memoria.
ECX (Conteo)	Almacena la cuenta de varias instrucciones. Además, puede guardar la dirección de desplazamiento de datos en la memoria. Las instrucciones utilizadas en un conteo son las instrucciones de cadena repetida (REP/REPE/REPNE); y las instrucciones desplazamiento (shift), rotación (rótate) y LOOP/LOOPD.
EDX (Datos)	Guarda una parte del resultado de una multiplicación, o parte del dividendo antes de una división. En el 80386 y superiores, este registro también puede direccionar datos de la memoria.
EBP (apuntador de la base)	Apunta a una posición de memoria en todas las versiones del microprocesador para las transferencias de datos de memoria. Este registro se direcciona como BP o EBP.
EDI (índice de destino)	Por lo general, direcciona datos de destino de cadenas para las instrucciones de cadenas. También funciona como un registro de propósito general de 32 bits (EDI) o de 16 bits (DI).
ESI (índice de origen)	Se utiliza como ESI o SI. Direcciona datos de cadena de origen para las instrucciones de cadenas. Se direcciona como SI al utilizarlo como registro de 16 bits; se direcciona como ESI al utilizarlo como registro de 32 bits.

1.2.1.2 Unidad de control

La unidad de control es el centro lógico de la computadora ya que los recursos de una computadora son administrados en la unidad de control, es esta unidad la que se encarga de dirigir el flujo de datos.

Las instrucciones de la CPU se encuentran incorporadas en la unidad de control, estas instrucciones o conjunto de instrucciones enumeran todas las operaciones que una CPU puede realizar. Cada instrucción es expresada en microcódigo.

Antes de que un programa sea ejecutado, cada comando debe desglosarse en instrucciones que correspondan a las que están en las instrucciones de la CPU. Al momento de ejecutar el programa, la CPU lleva a cabo las instrucciones en orden convirtiéndolas en microcódigo. A pesar de la complejidad que resulta del proceso, la computadora puede realizar este proceso a una velocidad increíble.

Cuando una CPU es desarrollada, el conjunto de instrucciones tiene los mismos comandos que su predecesora, aparte de incluirle algunos nuevos. Esto permite que el software escrito para una CPU trabaje con computadoras con procesadores más recientes, esta estrategia es llamada Compatibilidad ascendente. Esta característica permite ahorrar a los consumidores en comprar un sistema nuevo cada vez que una parte del sistema es actualizada.

Se le conoce compatibilidad decreciente o inversa, cuando un dispositivo del hardware o pieza del software puede interactuar con el mismo equipo y software que su predecesor.

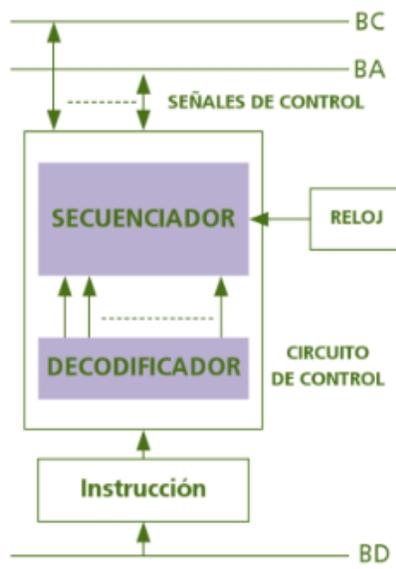


Figura 1.3: Esquema básico de la Unidad de Control

La UC se compone de las siguientes partes:

- **Circuito de control:** genera las señales de control necesarias para gobernar el ordenador. Sus partes principales son:
 - **Decodificador:** sistema combinacional que decodifica la instrucción que se va a ejecutar e informa del tipo de instrucción a un segundo módulo, el secuenciador.
 - **Secuenciado:** circuito secuencial que implementa la máquina de estados del sistema. A partir del resultado de la decodificación de la instrucción, genera de una forma ordenada en el tiempo las señales que permiten ejecutar la instrucción en curso (señales que seleccionan los registros internos adecuados o indican la dirección de memoria cuyo dato se va a tomar, comunican a la ALU el tipo de operación que ha de ejecutar, etc.).
- **Reloj:** es un circuito que genera pulsos, los cuales marcan la temporización básica del sistema. El reloj se utiliza como elemento sincronizador de todos los movimientos que se realizan en el ordenador.
- **Registros:** incluyen un contador de programa o PC (Program Counter), que almacena la dirección de memoria donde se encuentra la siguiente

instrucción a ejecutar, y un registro de instrucciones o IR (Instruction Register), donde la UC carga el código de operación de la instrucción en curso cuando la lee de memoria (cada instrucción viene representada por un número binario denominado de esta manera).

1.2.1.3 Unidad Aritmética Lógica.

Los datos almacenados en una computadora son de tipo numérico, por lo que gran parte del procesamiento involucra la comparación de números o la realización de operaciones matemáticas. Dentro de la unidad lógica aritmética de la computadora se realizan sólo dos tipos de operaciones: operaciones aritméticas y operaciones lógicas.

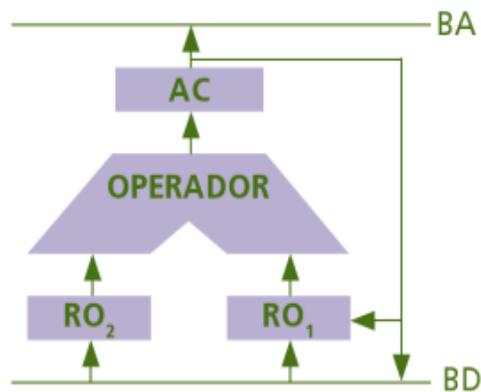


Figura 1.4: Esquema básico de la UAL

El elemento principal en UAL, es el operador. Un operador es un componente electrónico cuya misión es realizar un cálculo. Los operadores se pueden clasificar según:

- **Su ámbito:** en genéricos (pueden realizar diferentes operaciones) o específicos (sólo pueden realizar un tipo de operación).
- **El número de operandos:** en monádicos (sólo una entrada, por ejemplo, el negador), diádicos (dos entradas, por ejemplo, el operador suma) o triádicos (tres entradas, por ejemplo, el condicional).
- **Su capacidad de operar:** en serie (reciben la información en una secuencia de bits y los procesan uno a uno hasta terminar toda la lista) o en paralelo

(recepccionan la información en bloques de bits, que procesan de forma simultánea).

Toda la información que llega a la UAL se coloca en registros, que se utilizan como origen o destino de los datos que manejan los operadores de la UAL. Los registros con los que trabaja directamente el operador se llaman registros auxiliares y proporcionan la entrada de datos. El registro que recibe el resultado del operador se denomina acumulador (AC).

En el cuadro siguiente se muestra las operaciones realizadas por la unidad lógica aritmética.

Tabla 1.2: Operaciones Básicas de la UAL

Operaciones aritméticas (operaciones matemáticas)	Operaciones lógicas (comparaciones de números)
+ Suma	= , ≠ Igual a, no igual a
□ Resta	> , > Mayor que, no mayor que
× Multiplicación	< , < Menor que, no menor que
÷ División	> , > Mayor que o igual a, no mayor que ni igual a
^ Elevar a una potencia	< , < Menor que o igual a, no menor que ni igual a

La unidad lógica aritmética incluye un grupo de registros, ubicaciones de alta velocidad, las cuales se usan para conservar los datos que se procesan en ese momento

1.2.1.4 Registro de Estados (banderas).

Contienen información de estado y de control de las operaciones del microprocesador.

De los 16 bits del registro se utilizan solo 9 y cada uno de estos representa a una bandera. Existen seis banderas de estado que registran el estado del procesador, normalmente asociado a una comparación a o una instrucción aritmética:

Tabla 1.3: Registros de estado

Banderas de estado	Descripción
CF - (Carry Flag)	Bandera de acarreo. Indica acarreo en las instrucciones aritméticas.
OF - (Overflow Flag)	Bandera de desbordamiento aritmético.
ZF - (Zero Flag)	Bandera de resultado cero o comparación igual.
SF - (Sign Flag)	Bandera de resultado o comparación negativa.
PF - (Parity Flag)	Bandera de paridad, utilizada en la verificación de la transferencia de bytes entre dos componentes de la computadora.
AF - (Auxiliar Flag)	Bandera auxiliar. Indica si hay necesidad de ajuste en las operaciones aritméticas con números BCD.

Banderas de control: Registran el modo de funcionamiento de la computadora.

Tabla 1.4: Banderas de control

Banderas de control	Descripción
DF (Direction Flag)	Bandera de dirección. Controla la dirección (hacia adelante o hacia atrás) en las operaciones con cadenas de caracteres incrementando o decrementando automáticamente los registros índices (SI y DI).
IF (Interrupt Flag)	Bandera de interrupciones. Indica si están disponibles o no las interrupciones de los dispositivos externos.
TF (Trap Flag)	Bandera de atrape. Controla la operación modo paso a paso (usada por el programa DEBUG).

1.3 Elementos Internos

1.3.1 La tarjeta madre

Una computadora personal típica, en su hardware, está conformada por una placa de circuito impresa denominada placa madre o en otras bibliografías nombrada tarjeta madre (en inglés motherboard). El motherboard es una placa del tipo PCB (Printed Circuit Board) multicapa (distintas capas independientes de algún metal conductor (generalmente cobre) separadas por algún material aislante como baquelita o fibra de vidrio), que soporta toda la arquitectura que compone el ordenador en sí. Mantiene una gran cantidad de microcomponentes y diminutos chips soldados a ella.

Está unida a la carcasa o caja del ordenador mediante tornillos y soporta también todas las tarjetas necesarias para el funcionamiento del sistema. La placa base es el hardware que alberga al chip del procesador, la memoria principal y algunas

interfaces de E/S. Así mismo, tiene conectores adicionales en los que se pueden enchufar interfaces adicionales.

El bus del procesador es definido por las señales propias del chip del procesador. Los dispositivos que requieran alta velocidad de conexión con el procesador, como la memoria principal, pueden conectarse directamente a este bus. La tarjeta madre proporciona generalmente otro bus al que se pueden interconectar otros dispositivos. Los dos buses son interconectados por un circuito, que se denominará puente, que traslada desde un bus a otro las señales y protocolos. Los dispositivos conectados al bus de expansión muestran al procesador como si estuvieran conectados directamente al propio bus del procesador.

Otra de las funciones fundamentales del motherboard, es el manejo de los voltajes que alimentan a sus componentes. Además de los circuitos integrados, la placa base tiene un sistema de regulación de voltaje conformado por capacitores electrolíticos, reguladores y otros transistores que adecuan el voltaje arrojado por la fuente a cada uno de los dispositivos que integran la placa base. Este sistema permite que cada componente se mantenga estable e impide el desequilibrio del sistema por falta o exceso de corriente.

La siguiente imagen muestra un ejemplo de una tarjeta madre y sus principales elementos:

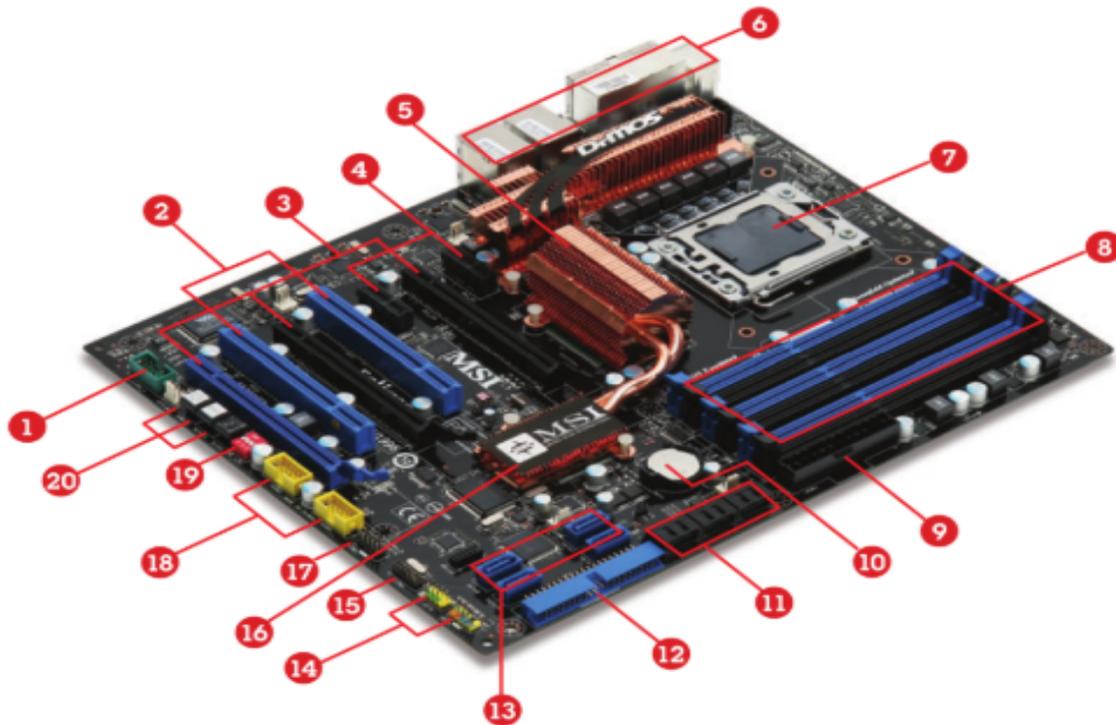


Figura 1.5: Ejemplo de una tarjeta madre

1	Conector para puerto IEEE 1394: este es el puerto por donde se conectan dispositivos para esta tecnología
2	Slot de expansión PCI convencional 32 bits: aquí se colocan placas de expansión.
3	Slot de expansión dedicado (para placas de video) PCI Express 16x: aquí sólo se colocan placas de video.
4	Slot de expansión PCI Express 1x: aquí se conectan placas de expansión
5	Chipset principal: también se le conoce como puente norte
6	Panel trasero: aquí se conectan los dispositivos y periféricos externos
7	Zócalo para el microprocesador: aquí se coloca el procesador
8	Ranuras para los módulos de memoria RAM DDR3
9	Conector principal de alimentación ATX2
10	Batería CMOS: es la que mantiene los datos del SETUP
11	Conectores SATA 2
12	Conector IDE: sirve para integrar dispositivos IDE
13	Puertos SATA 2: se utilizan para integrar dispositivos SATA
14	Pines conectores para el panel frontal: a través de estos conectores se enchufan los objetos del panel frontal
15	Conector para un chip de seguridad
16	Chipset secundario: también se le conoce como puente sur
17	Conector para agregar un puerto serial
18	Conectores para puertos USB
19	Switch de la CPU: sirve para variar el reloj del procesador
20	Botonera integrada, encendido, reset y selección del LEDs testigos

1.3.1.1 Tipos (AT, ATX, LPX, NLX).

Entre los distintos tipos de placas base cabe destacar dos grandes grupos: formatos propietarios y formatos estándar. En los primeros las especificaciones del formato no son públicas y cada fabricante realiza la placa de forma distinta intentando adaptarse al formato lo mejor posible, lo cual da como resultados problemas de compatibilidad entre placas del mismo formato, pero de diferentes fabricantes (problemas para encontrar componentes nuevos para la placa, tener que cambiar el chasis o la fuente al actualizar a una placa del mismo tipo, pero con mejores prestaciones, etc.). En el caso de los formatos estándar, las especificaciones de dicho formato son públicas, lo que favorece una mayor adaptación y expansión del formato, así como compatibilidad entre placas de distintos fabricantes.

En el primer grupo cabe destacar formatos como LPX (aunque parte de sus especificaciones son públicas) o mini-ITX, y en el segundo, y con mucha más popularidad, ATX o BTX.

A continuación, se detalla las características principales de algunos tipos de tarjetas madre:

- **AT:** En 1984, IBM lanza al mercado el IBM AT y con él, el formato AT, que adquiriría una gran repercusión en poco tiempo y se convertiría prácticamente en un estándar.

Dicho formato contaba con un procesador de 16 bits y con un bus de 16 bits. Esta placa base media 304,8 x 350.52 mm.

El gran tamaño de esta placa base dificultaba enormemente la adición de discos duros, así como presentaba el inconveniente de que el juego de cables del conector de la fuente de alimentación era confuso, produciendo que muchos usuarios quemaran la placa base al conectarla a la fuente.

- **LPX:** Al mismo tiempo que la Baby-AT se extendía rápidamente por el mercado a finales de los 80, Western Digital creaba otro formato de placa en 1987: LPX. LP viene de "Low-Profile" en referencia al hecho de que a este tipo de placas se les colocaba una tarjeta en el centro (llamada "riser card") y a dicha tarjeta se le conectaban las distintas tarjetas PCI o ISA, lo que permitía un ahorro significativo de tamaño (330x229 mm). Cabe destacar

también, que en este tipo de placas los conectores de entrada/salida se encontraban en la parte de atrás de la misma.

Las placas base LPX tuvieron un significativo éxito debido al uso de estas en sistemas de bajo coste de compañías como HP, Packard Bell, Compaq e incluso IBM. Sin embargo, presentaban una serie de desventajas que harían que finalmente fueran cayendo en desuso. En primer lugar, cabe destacar el hecho de que las especificaciones del formato LPX nunca fueron totalmente públicas, lo que ofrecía muchas dificultades para actualizar los componentes de la placa (requería comprarse una nueva riser card y dichas tarjetas eran raramente compatibles entre distintos diseños de la placa), siendo más útil cambiar de placa totalmente. De igual forma, muchas placas LPX de distintos fabricantes no eran compatibles entre sí, lo que significaba que si querías cambiar de placa muchas veces tenías que cambiar también de caja debido a las ligeras variaciones que uno u otro fabricante efectuaban en la placa.

Así mismo, el hecho de colocar una tarjeta en medio de la placa dificultaba mucho el flujo de aire, lo que suponía una gran desventaja a la hora de disipar el calor. Este defecto se acentuó con la llegada del Pentium III en 1997, que desprendía gran cantidad de calor.

- **NLX:** NLX (New Low-Profile Extended) nació en 1996 de la mano de Intel con el objetivo de reemplazar al formato LPX usado en los sistemas de tamaño reducido y bajo coste. A diferencia de LPX, las especificaciones del formato NLX son abiertas, lo que da como resultado total compatibilidad a la hora de cambiar de placa entre distintos fabricantes de esta, así como de cambiar algún componente en particular.

El formato NLX es en un principio, similar al LPX, sin embargo, NLX incorpora un gran número de mejoras para adaptarse a las mejoras tecnológicas de los últimos años, como puede ser el soportar los procesadores de gran tamaño o la inclusión de AGP o USB.

La principal característica de NLX con respecto a LPX es que la placa base se conecta directamente a la llamada “riser card”, y no al revés, como ocurría con las placas LPX. Este hecho permite que todos los cables que

normalmente se conectarían a la placa base, se conecten a la “riser card”, lo que tiene 2 resultados inmediatos: La placa no tiene ningún cable o conector interno y permite cambiar de placa con mucha facilidad al no tener que remover ningún cable o conector.

Este último punto ha sido toda una revolución en el mercado, ya que no sólo permite intercambiar una placa base con otra con una gran rapidez, sino también facilita el intercambio de componentes entre sistemas distintos.

Otras ventajas específicas del formato son:

- Soporte para un gran número de procesadores: NLX es capaz de soportar una gran gama de procesadores, así como permite adaptarse rápidamente a nuevos avances en los mismos (no hay más que cambiar de placa sin que por ello el resto de los componentes se vieran afectados, como puede ocurrir en otro tipo de placas).
- Soporte para nuevas tecnologías: NLX incluye soporte para las nuevas tecnologías tales como AGP, USB o módulos de memoria DIMM o RIMM.

Aunque NLX ofrecía varias mejoras significativas con respecto a LPX, nunca llegó a extenderse tanto como lo hizo este último. Sin embargo, gozó de bastante buena acogida entre los vendedores de equipos de bajo coste hasta la llegada en 1999 de flex-ATX o en 2001 del formato mini-ITX, que acabó por desplazar completamente a NLX

- **ATX:** El formato ATX nació en 1995 de la mano de Intel, y significó una revolución en el diseño de placas base desde el formato Baby-AT. Intel publicó las especificaciones del formato, lo que propició que dicho tipo de placa se extendiera rápidamente y se convirtiera en el formato más popular hasta nuestros días.

La placa base ATX posee un tamaño de 305 x 244 mm (superior al de antiguos formatos), lo cual nos indica que es incompatible con los formatos anteriores y que necesitamos adquirir una nueva caja y fuente de alimentación. Aunque esto pueda parecer en principio una desventaja, nada

más lejos de la realidad, pues el formato ATX recoge lo mejor de los formatos BabyAT y del LPX además de una serie propia de mejoras, a saber:

- **Facilidad para conectar la placa con la fuente:** El formato ATX incorpora un sólo conector con la fuente de alimentación (de 24 pines, con respecto a los 20 del formato AT), lo cual evita errores de conexión que pueden quemar la placa, tal y como ocurría frecuentemente con los formatos AT.
- **Recolocación de la CPU y la memoria:** En el formato ATX la CPU y los módulos de memoria se recolocan para que no puedan interferir con los buses de otras tarjetas y de esta forma actualizar la memoria o el procesador sin necesidad de quitar dichas tarjetas. Así mismo, la nueva posición de estos elementos propicia un mejor flujo de aire favoreciendo la disipación del calor.
- **Recolocación de los conectores internos de entrada/salida:** Los conectores internos de entrada/salida para los disquetes y el disco duro se colocan más cerca de dichos dispositivos, lo que propicia unos cables internos más cortos y acceder a dichos conectores no requiere quitar ninguna tarjeta o disco duro.
- **Menor costo para los fabricantes:** Las mejoras anteriormente comentadas, así como la reducción de la longitud de los cables (acortando la distancia entre los conectores y los dispositivos) o la eliminación de estos (no son necesarios cables para los puertos serie y paralelo), provoca que el formato ATX tenga un costo de fabricación mucho menor que el de formatos anteriores. Sin duda este fue un factor determinante para la gran aceptación de las ATX en el mercado.

La placa ATX es básicamente una placa Baby-AT rotada 90 °, y los slots PCI no interfieren con la memoria o la CPU ni con otros elementos.

Aunque el tamaño de una placa ATX sea similar al de una Baby-AT, normalmente los chasis de ambas son incompatibles, así como la fuente de alimentación, ya que usan diferentes conectores y proporcionan diferentes voltajes.

El formato ATX literalmente ha barrido del mercado a las placas Baby-AT y LPX, y no contentos con ello, Intel ha intentado reducir progresivamente el tamaño de las placas ATX, desde la Mini-ATX, que no es más que una ATX ligeramente más pequeña (284x208mm) hasta otros formatos como micro-ATX o flexATX que aparte de reducir su tamaño, ofrecen otra serie de características

1.3.1.2 Elementos integrados.

Los elementos integrados hacen referencia a los componentes críticos y no críticos que la placa base trae soldados a su superficie. Los que generalmente se ven son: dispositivos de video, de sonido y de red. También trae integrados los diferentes controladores para los puertos de teclado, mouse y puertos USB.

Estos componentes integrados determinan no sólo el rendimiento de los componentes internos, sino de los componentes de expansión y el de la PC en forma general.

Debemos tener presente, que los distintos tipos de tarjetas madre difieren en términos de cantidad y calidad de los elementos que lo integran.

A continuación, se detalla cada uno de los elementos integrados:

- **Buses de la tarjeta madre:** antes de comenzar con el detalle de cada uno de los elementos que encontraremos en la superficie de la placa base, es necesario saber que todos ellos se conectan entre sí por medio de los buses. Los buses son una cantidad de pistas en cuyo interior corren los datos de información. Básicamente hay tres tipos de buses: buses de datos (para que los dispositivos puedan comunicarse entre sí), bus de direcciones (cada dispositivo debe estar referenciado por una dirección), bus de control (controla las entradas y salidas de la información).
- **Zócalo del CPU:** se trata de un elemento integrado a la placa base sobre el cual se coloca la pastilla del procesador, funciona como interfaz entre el circuito integrado de la tarjeta madre y el microprocesador. Es el encargado de hacer funcionar el procesador.

- **Slot para video:** el dispositivo de video tiene como función transformar las señales eléctricas desde el motherboard hacia el monitor. Actúa como interfaz (intermediario) entre el motherboard y lo que se conoce como placa o tarjeta de video.
- **Ranura para memoria RAM:** la memoria RAM es un elemento crítico de la PC, y se instala físicamente sobre el motherboard. Para ello, la placa base necesita de una interfaz que soporta cada módulo de RAM. Estas ranuras donde se encastran los módulos de memoria RAM tienen características inherentes a su factor de forma. Es decir, cuentan con muescas de posición para diferenciar las tecnologías y se alimentan con distintos voltajes.
- **Chipset:** son un grupo de chips esencialmente diseñados para realizar determinadas tareas. Son el eje de la placa base y establecen el performance de este.

Básicamente se cuenta con dos tipos de chips. Por un lado, está el puente norte que se comunica con el bus de la CPU, el bus de memoria RAM y los buses de video. El segundo chip es el puente sur, cuya función es comunicarse con todos los demás buses de los dispositivos (bus del disco duro, el de sonido, bus USB). Estos dos puentes se comunican mediante un bus dedicado. De este modo, todos los dispositivos de la PC quedan comunicados por sus respectivos chipsets.

- **Slots de expansión:** estas ranuras tienen como función ampliar las capacidades de hardware de la PC. Por ejemplo, mediante los denominados slots de expansión. Podemos agregar placas de red, placas de sonido y puertos extras.
- **Conectores de alimentación:** los componentes electrónicos necesitan de alimentación eléctrica para funcionar. Este es el motivo por los cuales el motherboard tiene algunos conectores de alimentación. El voltaje necesario para que la tarjeta madre y los dispositivos ensamblados sobre él puedan funcionar es otorgado por otro de los elementos críticos conocidos como fuente de alimentación.

- **Conector IDE (Integrated Development Environment):** se utilizó para conectar el cable de datos de las unidades ópticas y de los discos duros.
- **Conector SATA (Serial Advanced Technology Attachment):** esta tecnología es la utilizada actualmente. Es un conector que sirve de interfaz a los discos duros y a las unidades ópticas que utilizan esta tecnología.
- **Panel Frontal:** es el conjunto de pines que se encuentra sobre una de las esquinas anteriores de la tarjeta madre. Su función es conectar la botonera de mando central de la PC que se compone del botón de encendido o power, la tecla de reset, el LED on/off y el LED de carga de disco duro. Además, se pueden encontrar otro grupo de pines que corresponden a la conexión de puertos USB y salidas /entradas de audio.
- **Panel trasero:** se encuentran en esta parte las interfaces de los dispositivos integrados, como el conector de video, las entradas y salidas de audio y la interfaz de red. También hay conectores para el teclado, el mouse y los puertos USB.

1.3.2 Los microprocesadores.

El microprocesador es un componente electrónico que realiza en una única pastilla el procesador (CPU) de una máquina programable de tratamiento de la información. Es decir, el microprocesador incluye en un solo chip, dos de las partes constituyentes de la arquitectura Von Neumann, la UC y la UP.

Es decir, el microprocesador es un componente electrónico complejo que incorpora las funciones típicas de todo un computador. Más aún, gracias al aumento en la capacidad de integración se han podido introducir más elementos de la estructura Von Neumann dentro de un único circuito. Así, algunos microprocesadores convencionales incluyen también en el mismo chip cierta cantidad de memoria (por ejemplo, la memoria caché), controladores de DMA (que forma parte del subsistema de E/S) y otros elementos.

La clave del éxito de los microprocesadores, como componente electrónico, reside en que modificando el programa almacenado en memoria pueden adaptarse a numerosas y diferentes aplicaciones, de manera que, en la actualidad, debido a su

creciente potencia de cálculo y variedad de funciones integradas, se han convertido en el estándar para la implementación de tareas, tanto de cálculo como de control. De este modo, los microprocesadores no sólo se utilizan para construir potentes computadoras, quizás su aplicación más espectacular y conocida, sino también para la realización de tareas de control en sistemas de aplicación específica (por ejemplo, el control de un horno microondas, un teléfono móvil o unos frenos ABS), sustituyendo circuitería convencional.

1.3.2.1 Arquitectura CISC y RISC.

Los microprocesadores pueden clasificarse por el número de instrucciones que incorporan en su repertorio. Así, se distingue entre microprocesadores CISC y RISC:

- **CISC:** A partir de los computadores 360 y 370 de IBM, surgidos en la década del 70, la mayoría de los procesadores (CPU) de las computadoras, incluidos los de minicomputadoras y PC personales (Pentium y el P6 de Intel, y los 680x0 de Motorola han sido CISC (Complex Instruction Set Computer). Esta denominación se debe a que pueden ejecutar desde instrucciones muy simples (como las que ordenan sumar, restar dos números que están en registros de la CPU y el resultado asignarlo a uno de esos registros), hasta instrucciones muy complejas (como los tan usados movimientos de cadenas de caracteres de gran longitud variable, en procesamiento de textos). Las instrucciones simples, luego de su decodificación, pueden ejecutarse en un pulso reloj, mientras que las complejas requieren un número de pulsos que depende de la secuencia de pasos necesarios para su ejecución. Cada paso se lleva a cabo mediante combinaciones binarias (microinstrucciones), que aparecen en las líneas de control de la UC con cada pulso de reloj, el cual activa los circuitos que intervienen en ese paso. Las sucesivas micro instrucciones que requiera la ejecución de una instrucción compleja, deben ser provistas por la ROM (firmware) de Control que las almacena, y que forma parte de la UC. Por lo tanto, un procesador CISC necesariamente debe contener una ROM con las microinstrucciones para

poder ejecutar las instrucciones complejas. Esta es una de las características CISC.

En general, cada operación que ordena una instrucción de un procesador CISC presenta variantes para ser aplicadas a diversas estructuras de datos, desde simples constantes y variables, hasta matrices y otras. Así, una instrucción que ordena suma tiene muchas variantes (códigos) en función de la estructura de datos sobre la cual opera. Es como si existieran tantas instrucciones que ordenan una misma operación como estructuras de datos típicas se han definido para operar. Este concepto fue planteado previamente bajo el nombre de "modos de direccionamiento" de una instrucción.

La existencia de muchos "modos de direccionamiento" para realizar la operación que ordena una instrucción, es otra de las características de complejidad de los procesadores CISC. Esto se manifiesta en que el repertorio ("set") de instrucciones de una máquina CISC presente un número elevado de códigos de operación. Así, una IBM 370 tiene 210 instrucciones, 300 la VAX, y 230 el 80486. Así mismo, lo anterior exige instrucciones que ocupan distinta cantidad de bytes en memoria. Resulta así, que, por tener instrucciones ejecutables en diferente cantidad de pulsos de reloj, un procesador CISC no puede aprovechar eficazmente su "pipeline" en la producción de instrucciones.

- **RISC:** Ya para los setenta, la velocidad de las memorias se acercó a la de la CPU y resultaba difícil escribir, depurar y mantener los microprogramas. Además, algunos especialistas comenzaron a analizar qué tipo de instrucciones eran las más usadas en los programas. Los resultados fueron sorprendentes:
 - El 85% de las instrucciones son de asignación, condicionales y de llamadas a procedimientos,
 - El 80% de las instrucciones de asignación son de un solo término,
 - El 41% de los procedimientos no tienen argumentos, y
 - El 80% de los procedimientos tienen 4 o menos variables locales.

Como conclusión, podemos decir que, si bien teóricamente es posible escribir programas complicados, la mayoría de los programas reales consisten en simples asignaciones, declaraciones condicionales y llamadas a procedimientos con un número reducido de parámetros.

Esta conclusión es de extrema importancia en la tendencia a agregar más y más funciones al microcódigo. Mientras que el lenguaje de máquina se hace más complicado, el microprograma se hace más grande y lento. Un número elevado de modos de direccionamiento significa que su decodificación no puede realizarse en línea (lo que implicaría repetir cientos de veces el mismo microcódigo). Peor aún, se ha sacrificado velocidad a fin de incorporar instrucciones que en la práctica rara vez se usan.

Se podría afirmar que una buena idea sería eliminar el microcódigo y que los programas se corran directamente por el hardware residiendo en una rápida memoria principal. Surgen así las computadoras con un número reducido de instrucciones, llamadas máquinas RISC (Reduced Instruction set Computer). Antes de Wilkes todas las máquinas eran RISC, y luego con la microprogramación las computadoras se hicieron más complejas y menos eficientes. Ahora, la industria está volviendo a sus raíces, construyendo máquinas sencillas y rápidas.

La denominación RISC no está del todo bien aplicada. Si bien es cierto que tienen pocas instrucciones, la característica más importante es que éstas se completan en un sólo ciclo usando pipeline intensivamente (entendiendo por ciclo la extracción de los operandos de un registro, colocarlos en el bus, ejecutarlos a través de la ALU y guardar el resultado en un registro).

Cualquier operación que no se lleve a cabo en un ciclo, no puede formar parte del conjunto de instrucciones.

Arquitectura de carga/almacenamiento: Dado que cada instrucción debe ejecutarse en un ciclo, resulta claro que aquellas que hacen referencia a memoria representan un problema.

Las instrucciones ordinarias sólo pueden tener operandos en registros (sólo está permitido el direccionamiento por registros). Las únicas instrucciones

que hacen referencia a memoria son LOAD y STORE. Para lograr que éstas se ejecuten en un ciclo se recurre a exigir que cada instrucción se inicie en cada ciclo, sin importar cuando termine. Si se logra, entonces, comenzar n instrucciones en n ciclos, se habrá alcanzado un promedio de una instrucción por ciclo.

Las instrucciones generadas por el Compilador son ejecutadas directamente por el hardware, es decir, no son interpretadas por el microcódigo. Esta es la razón de la velocidad de las RISC. La complejidad que soporta el microcódigo en las CISC, se traslada al código de usuario en las RISC. Si una instrucción compleja implica n microinstrucciones en una CISC, en una RISC implicaría un número similar de instrucciones, pero ocuparía más memoria. Sin embargo, también debe tenerse en cuenta que las instrucciones complejas representan un porcentaje menor en un programa real. Se recomienda comparar los tiempos de ciclo de instrucción de la máquina elemental con la máquina elemental microprogramada.

La razón del reducido número de instrucciones responde a la idea de simplificar el decodificador de instrucciones. En cuanto a los modos de direccionamiento conviene reducirlos al mínimo. En el formato de una instrucción RISC es posible generar distintos modos de direccionamiento.

A fin de reducir el número de cargas y almacenamientos (LOAD Y STORE), una buena parte del chip de una RISC se usa para registros (no es raro encontrar RISC con 500 registros), aprovechando que carece de firmware. La organización de estos registros es un aspecto muy importante en las RISC.

Ya que una parte importante del tráfico a memoria es consecuencia de los llamados a procedimientos (que implica transmitir parámetros, salvar registros, etc.), dio lugar a los diseñadores de las RISC a plantear una organización de registros llamada traslape de registros.

El traslape de registros consiste, en general, que en un momento dado la CPU accede a sólo un subgrupo de ellos (por lo general de 32 bits). divididos en 4 grupos de 8 cada uno. Los primeros ocho registros se encuentran en

todo momento accesibles a la CPU y representan los registros globales utilizables por cualquier procedimiento.

En cambio, desde el R8 en adelante, el grupo de 24 registros serán accesibles por la CPU dependiendo del valor de un puntero de registros. Este puntero se ajusta cuando algún procedimiento es invocado desde el actual. Esta organización permite intercambiar valores entre procedimientos sin necesidad de referir a la memoria.

A continuación, se detallan las principales diferencias entre las 2 arquitecturas de los microprocesadores:

	RISC	CISC
1	Instrucciones sencillas en un ciclo	Instrucciones complejas en varios ciclos
2	Sólo LOAD/STORE hacen referencia a memoria	Cualquier instrucción puede referencia memoria
3	Procesamiento serie de varias etapas	Poco procesamiento en serie
4	Instrucciones ejecutadas por hardware	Instrucciones interpretadas por el microprograma
5	Instrucciones de formato fijo	Instrucciones de formato variable
6	Pocas instrucciones y modos de direccionamiento	Muchas instrucciones y modos de direccionamiento
7	La complejidad está en el compilador	La complejidad está en el microprograma
8	Varios conjuntos de registros	Un sólo conjunto de registros
9	Mayor uso del Bus en Búsqueda	Menor uso del Bus en Búsqueda
10	Menor uso del Bus en Ejecución	Mayor uso del Bus en Ejecución

Figura 1.6. Comparativa entre CISC Y RISC

1.3.2.2 Familias.

En 1971 se hizo una innovación sensacional, cuando Intel desarrolló su 4004, este fue el primer chip que contenía todos los componentes del CPU en solo chip: el microprocesador había nacido. El siguiente paso importante se dio en 1974 con la introducción de Intel 8080 y hacia finales de los setenta aparece el 8086, potente microprocesador de 16 bits de propósito general.

Modelo	Características
80186	Como un 8086 mas controlado de reloj, DMA, controlador de interrupciones, etc. Actualmente se ofrece como microcontrolador
80286	Bus de dirección de 24 bits Reloj de 16 Mhz
80386	Bus de 32 biys para datos y direcciones Reloj de 33 Mhz
80486	Cache de 8 Kbytes Coprocesador matemático Reloj de 66 Mhz
Pentium	Bus de datos de 64 bits Caches separadas de datos e instrucciones de 8 Kbytes cada una Arquitectura <u>superscalar</u> Reloj hasta 1Ghz

Figura 1.7: Evolución de microprocesadores de INTEL

El siguiente paso en esta tendencia ocurrió en 1981, cuando los laboratorios Bell y Hewlett-Packard desarrollaron los microprocesadores de un solo chip de 32 bits.

La década de los 90 trajo a AMD (Advance Micro Devices – Micro Dispositivos Avanzados), una compañía norteamericana afincada en California. AMD tenía la ventaja frente a los demás desarrolladores de disponer de su propia planta de fabricación de chips, lo que concedía cierta autonomía. En 1975 AMD hizo una copia de microprocesador Intel 8080 mediante técnicas de ingeniería inversa, al cual nombró AMD 9080.

El primer chip multiprocesador comercialmente disponible se introdujo en el 2001, y era el chip POWER4 de IBM.

Intel crea el microprocesador PENTIUM, y desde allí se origina una larga familia de microprocesadores INTEL que siguen buscando la forma de mejorar sus prestaciones. Sus familias adversarias AMD e IBM continúan con las mismas

ambiciones; lo que convierte a estas tres grandes familias de microprocesadores en las más importantes del sector.

1.3.3 Tarjetas de expansión.

Las tarjetas son placas donde están incluidos circuitos impresos que hacen posible la operación con distintos periféricos.

Las tarjetas pueden realizar diferentes funciones (red, modem y/o fax), multimedia (sonido, sintonizadoras), etc. Dichas tarjetas se insertan en los slots de expansión de la placa base pudiendo ser estos AGP para video, ISA, MCA, EISA, o PCI en sus diferentes versiones.

1.3.3.1 Gráficas.

Procesa toda la información que vemos en el monitor. La mayoría de los equipos tienen una tarjeta integrada en la placa base en lugar de tener una tarjeta de video separada.

La función básica de una tarjeta gráfica es convertir la información procesada por el ordenador, o la propia tarjeta, en una señal que pueda entender el monitor, para mostrarla en forma de imagen en la pantalla. En el caso de las tarjetas aceleradoras 3D, éstas también realizan la función de procesar las imágenes tridimensionales, liberando al procesador de esta tarea.

Físicamente, las tarjetas aceleradoras consisten en una placa de circuito, cuyo circuito electrónico es casi un miniordenador, pues incluye su propio procesador gráfico (llamado GPU) y su propia memoria RAM (llamada memoria gráfica). Se conecta a la placa base a través de una ranura específica, que puede ser de tipo AGP (Accelerated Graphics Port, Puerto de Gráficos Acelerado) o PCI Express 16x (Peripheral Component Interconnect, Interconexión de Componentes Periféricos), que es más común hoy en día.

1.3.3.2 Sonido.

Esta tarjeta es la responsable de lo que los ordenadores reproduzcan o graben sonido. La mayoría de las computadoras tienen integrada la tarjeta de sonido dentro

de su placa madre, pero ésta se puede actualizar comprando una que tenga mayor calidad. La tarjeta de sonido tiene una doble función:

- Convertir la información digital contenida en archivos de sonido (de tipo WAV, MP3...) en una señal de sonido analógica que pueda ser transmitida a unos altavoces u otro aparato de sonido analógico.
- Grabar la señal de sonido procedente de una fuente analógica (micrófono, magnetófono, reproductor de CD...) en un archivo de sonido digital.

Físicamente es una placa de circuito impreso, que contiene componentes electrónicos específicos, conexiones internas y externas, así como la interfaz de conexión a la placa base, que es de tipo PCI-Express. Actualmente las placas base incorporan las funciones de una tarjeta de sonido en el chip correspondiente al Puente Sur. Pero si queremos tener un sonido de mayor calidad tendremos que añadir una tarjeta de sonido en una ranura de expansión que esté libre.

1.3.3.3 Comunicación – Red.

El punto de entrada de la red a los equipos es la tarjeta de red, también conocidas por sus siglas en inglés NIC (Network Interface Card), son elementos electrónicos que posibilitan la conexión de un equipo a una red. Podemos encontrar tarjetas de red que se instalan en los slots PCI de la placa, aunque la mayor parte de las tarjetas de red están incluidas en la propia placa base del equipo.

El diseño y las funciones implementadas en una tarjeta de red dependen de la tecnología para la que se va a utilizar. Actualmente, la tecnología dominante es ethernet, por tanto, todas las tarjetas de red que se pueden encontrar son ethernet. Las tarjetas de red llevan a cabo todo el procesamiento de las funciones 1 y 2 del modelo de referencia OSI en el equipo.

Para que una tarjeta de red funcione adecuadamente en un equipo con el sistema operativo, es necesario que exista un controlador o driver. Dicho controlador posibilita que el sistema operativo pueda intercambiar información con la tarjeta, es decir, permite la comunicación entre el sistema operativo y la propia tarjeta de red.

1.3.3.4 Controladoras.

Cada uno de los periféricos conectados al ordenador, independientemente de que sean internos o externos, necesitan utilizar alguna técnica o medio (controladores) para comunicarse entre ellos y con la propia CPU.

Podemos definir controlador como el traductor existente entre cada uno de los dispositivos periféricos y la CPU. Estos dispositivos son teclado, monitor, disco duro, DVD, etc.

Las funciones básicas de las tarjetas controladoras o controladores se pueden resumir en las siguientes:

- Convertir los datos de un formato a otro, cuando ocurra que los datos entre los distintos dispositivos y la CPU tengan distintos formatos.
- Permite la unión de los dispositivos hardware (disco duro, DVD, etc.,) con los drives del sistema operativo.
- Adecúan las velocidades de los distintos dispositivos, cuando estas son diferentes, para hacerlas compatibles

1.3.4 Control de Energía.

1.3.4.1 Fuente de Poder.

En toda computadora existe un reloj del sistema, este reloj es accionado por un cristal de cuarzo que al momento de aplicarle electricidad las moléculas en el cristal vibran muchas veces cada segundo. Estas vibraciones son usadas por la computadora para medir sus operaciones de procesamiento.

Al paso de los años las velocidades de los relojes se han incrementado constantemente. La primera PC operaba a 4.77 mega Hertz (millones de ciclos por segundo). Un ciclo de reloj es el tiempo que le lleva un transistor en apagarse y encenderse.

La fuente de alimentación eléctrica de una computadora proporciona a ésta la energía necesaria a los circuitos de la unidad central de proceso, y demás elementos hardware que la conforman.

Esta fuente de alimentación tiene como objetivo reducir las variaciones de voltaje existentes en la toma de corriente. Internamente se encuentra provisto de un

ventilador que disipa el calor desprendido por el propio bloque y en términos generales también por la computadora.

En la entrada de esta fuente se encuentra un cable que se conecta a la red eléctrica; en la salida varios cables en paralelo que se conectan con la tarjeta madre, así mismo sirven como fuente de alimentación de las unidades de disco y demás dispositivos.

Esta alimentación provee energía a la salida garantizando el funcionamiento adecuado de la computadora. La fuente estabiliza las tensiones presentadas. El fabricante especifica el rango de variación que se tolera. Las tensiones normalizadas son de $\pm 12V$ y $\pm 5V$.

1.3.4.2 Ventilador y disipador.

El procesador es alimentado por cierto voltaje que arroja la fuente de alimentación y esto genera inevitablemente temperatura. El procesador debe trabajar dentro de un rango calórico que oscila entre los 35 y 60 grados centígrados. Si este valor es superado, el sistema podría dejar de funcionar y es muy probable que el procesador se dañe.

Para evitar los excesos de temperatura, el procesador cuenta con un equipo de refrigeración conformado por un disipador y un ventilador o cooler. Estos dos dispositivos se montan sobre el procesador y se ajustan a unas pestañas de sujeción del zócalo del procesador. Entre el procesador y el disipador hay un elemento conductor de calor que permite que la temperatura del procesador busque su punto de fuga hacia el disipador. El disipador, a su vez, es refrigerado por el aire que genera el ventilador. De este modo, el procesador mantiene su temperatura dentro de los parámetros convencionales de funcionamiento.

ACTIVIDADES

Simulación de las unidades funcionales de un ordenador

Objetivos:

- Asimilar el concepto de unidad funcional
- Aplicar el concepto de unidad funcional, de forma intuitiva, a diferentes escenarios.

Desarrollo: Simular el comportamiento de una unidad funcional de un ordenador en el desempeño de la acción “Quiero saber el resultado de 23×5 ”

1. Para ello se dividirá la clase en cuatro grupos. Cada grupo será una unidad funcional

2. El profesor indicará a cada grupo que unidad funcional es. Él hará el papel de usuario.

3. Para cada uno de los enunciados, todos los grupos harán un planteamiento como este:

¿Qué hace la UC?

¿Qué hace la UAL?

¿Qué contiene la UM?

4. El profesor leerá el enunciado y acto seguido la UC planificará los pasos que tendrá que seguir para cumplir su misión adecuadamente. A continuación, se los comunicará al resto de la clase y comenzará la simulación.

5. La simulación comienza en el momento en que la UC da permiso a la Unidad de E/S para que le dé la información que ha introducido el usuario en el equipo.

6. Cada orden de la UC se llevará a cabo sin condición. Al final habrá que discutir si la secuencia ha sido la correcta.

7. Discutir con los compañeros sobre la posibilidad de complicar más el proceso mediante la descomposición de muchas de las órdenes que ha dado la UC.

8. Ahora los grupos rotan en el desempeño de la unidad funcional. Y se prueban con otros enunciados, problemas o instrucciones.

Definición de arquitectura de computadoras

Es el año 2021 y se le pide escribir un artículo, que ocupe una página, acerca de la arquitectura de computadoras para una enciclopedia infantil de ciencia y tecnología. ¿Cómo describiría la arquitectura de computadoras a los niños de primaria? Se le permite usar un diagrama, si fuera necesario.

Bibliografía

- Vázquez, J., (2012). Arquitectura de computadoras I. Primera edición. México. Red Tercer Milenio S.C. ISBN 978-607-733-091-2.
- Rocabado, S., Arias, D., (2009). Arquitectura y organización de la computadora. Primera edición. Argentina. Editorial: FUNTICs – Fundación para la Investigación y Desarrollo en Nuevas Tecnologías. I.S.B.N. N°: 978-987-25293-0-7.
- Daniel M. Argüello, Santiago C. Pérez e Higinio A. Facchini. (2018). Arquitectura de computadoras. Universidad Tecnológica Nacional – Facultad Regional Mendoza. E-book – 4° Edición. Argentina. ISBN: 978-950-42-0158-8.
- Cottino, Damián., (2009). Hardware desde cero. 1ª ed. Banfield-Lomas de Zamora: Gradi. 352 pág. ISBN 978-987-663-001-6. Recuperado en https://books.google.com.pa/books?hl=es&lr=&id=euYd0BVMrTQC&oi=fnd&pg=PA28&dq=tipos+de+motherboard&ots=tg_Uj3iTX2&sig=LjEv59UmUyN5fknvnMsfLQZ7fyl&redir_esc=y#v=onepage&q=tipos%20de%20motherboard&f=false
- Gaumé, Sylvain. (2016). Mantenimiento y reparación de un PC en red. 4ª ed. Barcelona: Ediciones ENI. ISBN: 978-2-409-00046-1. Recuperado en [https://books.google.com.pa/books?hl=es&lr=&id=iHfkAPrtThAC&oi=fnd&pg=PA17&dq=Proceso+de+Arranque+%E2%80%93+POST+\(%E2%80%9CPower+On+Self+Test%E2%80%9D\).&ots=7d9VRQNxEk&sig=xP_VdzFuqjUOhczCNVcINGhHZ4Q&redir_esc=y#v=onepage&q&f=false](https://books.google.com.pa/books?hl=es&lr=&id=iHfkAPrtThAC&oi=fnd&pg=PA17&dq=Proceso+de+Arranque+%E2%80%93+POST+(%E2%80%9CPower+On+Self+Test%E2%80%9D).&ots=7d9VRQNxEk&sig=xP_VdzFuqjUOhczCNVcINGhHZ4Q&redir_esc=y#v=onepage&q&f=false)
- Vasconcelos, Jorge. (2018). Introducción a la Computación. 1ª ed. México: Grupo Editorial Patria, S.A. ISBN ebook: 978-607-744-953-9.
- Del Brío, Bonifacio. (1999). Sistemas Electrónicos basados en microprocesadores y microcontroladores. España: Editorial Prensas Universitarias de Zaragoza. ISBN 84-7733-516-8.
- Cruz, S. I. O. Comparación de la arquitectura de microprocesadores Intel y AMD. Instituto Tecnológico de Tijuana. México.

CAPÍTULO 2: INTERFAZ DEL BUS

Objetivos:

- ✓ Comprender la estructura y el funcionamiento de un bus como punto de comunicación entre las unidades de un computador.
- ✓ Analizar la forma de sincronización que utilizan los buses para que la comunicación se realice de manera efectiva.
- ✓ Presentar la estructura jerárquica que adoptan los buses en un computador a fin de equilibrar las diferentes velocidades de los dispositivos a ellos conectados.
- ✓ Revisar el funcionamiento de algunos buses estándar.

¿De qué se trata esta sección de aprendizaje?

Un bus es un medio compartido de comunicación formado por un grupo de caminos (conductores) que entrelazan los diferentes elementos de un computador. La principal función de un bus será, pues, servir de soporte para la realización de transferencias de información entre dichos elementos.

El capítulo describe la forma como el procesador se conecta con los diferentes buses de comunicación dentro del sistema y detalla las características de los principales miembros de esta red interna en el computador.

2. LA INTERFAZ DEL BUS

2.1 Interconexión con los buses.

En un sistema de computadores, los diversos subsistemas deben tener interfaces entre sí; por ejemplo, la memoria y la CPU necesitan comunicarse, también la CPU y los dispositivos de E/S. Esto, comúnmente, se realiza con un bus.

Hennessy y Patterson (1993) especificaron que el bus sirve como enlace de comunicación compartido entre los subsistemas. Las dos principales ventajas de la organización bus son el bajo coste y la versatilidad. Al definir un sencillo esquema de interconexión, se pueden añadir fácilmente nuevos dispositivos y los periféricos pueden, incluso, compartirse entre sistemas de computadores que utilicen un bus común. El coste es bajo, ya que un simple conjunto de cables es un camino múltiple compartido.

Un bus es un camino de comunicación entre dos o más dispositivos. Una característica clave de un bus es que se trata de un medio de transmisión compartido. Al bus se conectan varios dispositivos, y cualquier señal transmitida por uno de esos dispositivos está disponible para que los otros dispositivos conectados al bus puedan acceder a ella. Si dos dispositivos transmiten durante el mismo periodo de tiempo, sus señales pueden solaparse y distorsionarse. Consiguientemente, solo un dispositivo puede transmitir con éxito en un momento dado.

Usualmente, un bus está constituido por varios caminos de comunicación o líneas. Cada línea es capaz de transmitir señales binarias representadas por 1 y por 0. En un intervalo de tiempo, se puede transmitir una secuencia de dígitos binarios a través de una única línea. Se pueden utilizar varias líneas del bus para transmitir dígitos simultáneamente (en paralelo).

Como forma de reafirmación Stallings (2006) indica que el bus que conecta los componentes principales del computador (procesador, memoria y E/S) se denomina Bus del sistema (System bus). Las estructuras de interconexión más comunes dentro de un computador están basadas en el uso de uno o más buses del sistema.

2.1.1 Jerarquía de buses.

Los buses presentan 2 principales problemas, que son solucionados mediante la implementación de la jerarquía de buses.

- El bus se convierte en un cuello de botella, pues todas las transferencias de información pasan a través de él.
- El bus debe soportar elementos de velocidades muy dispares, lo que trae consigo un diseño de bus poco óptimo.

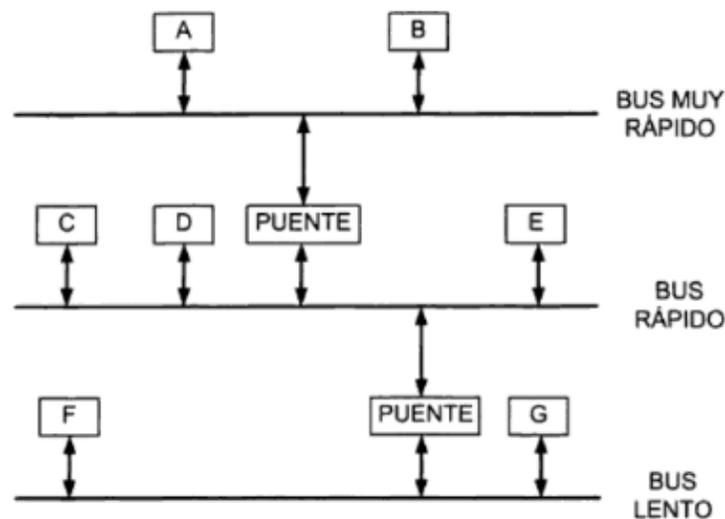


Figura 2.1: Ejemplo de jerarquía de buses

Para solucionar los problemas anteriores, la técnica que se emplea en los computadores actuales es implementar una jerarquía de buses. La figura 2.1 muestra un ejemplo de jerarquía de buses.

Como puede observarse en la figura hay tres buses de velocidades diferentes. Por ejemplo, el bus superior podría estar conectado al procesador y el inferior a dispositivos de E/S lentos. Los buses están interconectados empleando puentes (bridges).

La jerarquía de buses soluciona el problema del número de dispositivos por bus, pues los dispositivos se distribuyen en diferentes buses. El segundo problema se aminora en gran medida pues es posible llevar a cabo transferencias simultáneas en diferentes buses. Por ejemplo, es posible llevar a cabo simultáneamente las transferencias A->B, E->C Y F->G, o las transferencias A->D y F->G de la figura

2.1. Además, puesto que cada bus se especializa en un rango de velocidades de transferencia, el aprovechamiento de buses es mucho mayor.

En la práctica, la jerarquía de buses forma parte de una jerarquía de interconexión más general dentro del computador.

2.1.1.1 Arquitectura tradicional.

Una arquitectura típica de buses organizados jerárquicamente se muestra en la figura 2.2. Hay un bus local que conecta el procesador a una memoria caché y al que pueden conectarse también uno o más dispositivos locales. El controlador de memoria caché conecta la caché no solo al bus local sino también al bus del sistema, donde se conectan todos los módulos de memoria principal. El uso de caché alivia la exigencia de soportar los accesos frecuentes del procesador a memoria principal. De hecho, la memoria principal puede pasar del bus local al bus del sistema. De esta forma, las transferencias de E/S con la memoria principal a través del bus de sistema no interfieren la actividad del procesador.

Es posible conectar controladores de E/S directamente al bus de sistema. Una solución más eficiente consiste en utilizar uno o más buses de expansión. La interfaz del bus de expansión regula las transferencias de datos entre el bus de sistema y los controladores conectados al bus de expansión. Esta disposición permite conectar al sistema una amplia variedad de dispositivos de E/S y al mismo tiempo aislar el tráfico de información entre la memoria y el procesador del tráfico correspondiente a las E/S.

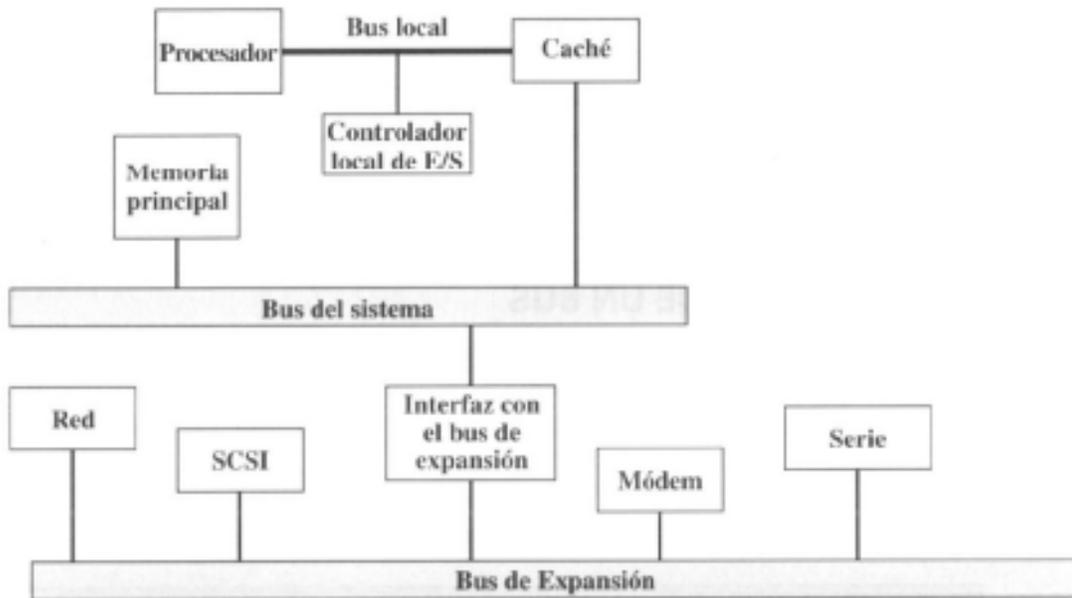


Figura 2.2: Arquitectura de bus tradicional

La figura 2.2 muestra algunos ejemplos típicos de dispositivos de E/S que pueden estar conectados al bus de expansión, como pueden ser conexiones de red, puerto serie (impresora o escáner), interfaz SCI, etc.

Esta arquitectura de buses es razonablemente eficiente, pero muestra su debilidad a medida que los dispositivos de E/S ofrecen prestaciones cada vez mayores.

2.1.1.2 Arquitectura de altas prestaciones.

Como respuesta a la debilidad de la jerarquía de buses tradicional, la industria propone un bus de alta velocidad que está estrechamente integrado con el resto del sistema, y requiere solo un adaptador (bridge) entre el bus del procesador y el bus de alta velocidad.

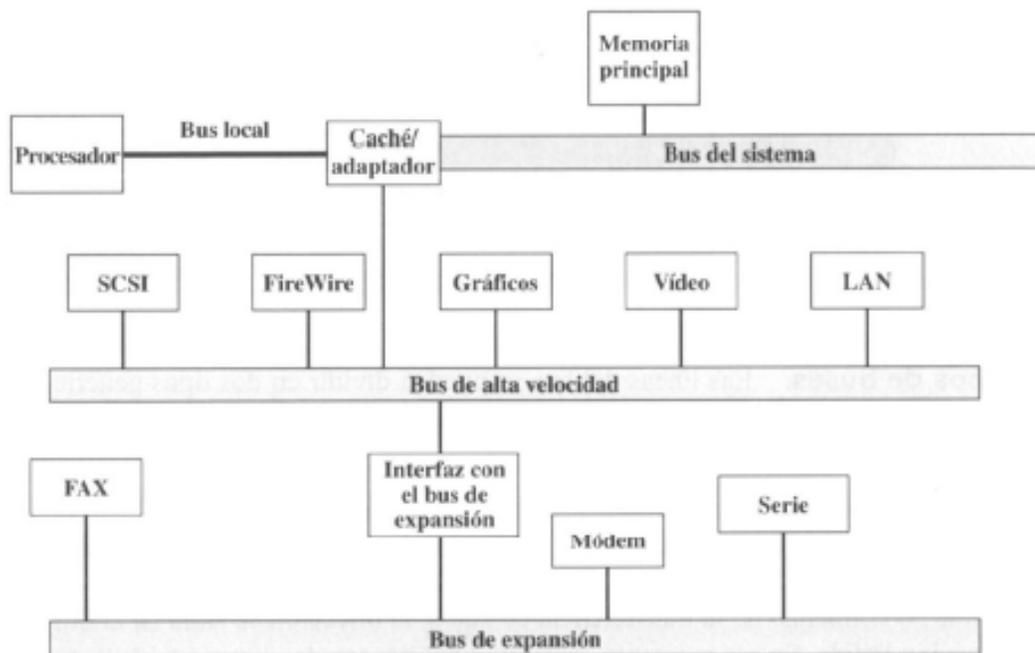


Figura 2.3.: Arquitectura de altas prestaciones

La figura 2.3 muestra un ejemplo de esta aproximación. De nuevo, hay un bus local que conecta el procesador a un controlador de caché, que a su vez está conectado al bus del sistema que soporta a la memoria principal. El controlador de caché está integrado junto con el adaptador, o dispositivo de acoplo, que permite la conexión al bus de alta velocidad. Este bus permite la conexión de LAN de alta velocidad, controladores de estaciones de trabajo específicos para aplicaciones gráficas y n de vídeo. Los dispositivos de velocidad menor pueden conectarse al bus de expansión, que utiliza una interfaz para adaptar el tráfico entre el bus de expansión y el bus de alta velocidad.

La ventaja de esta organización es que el bus de alta velocidad acerca al procesador los dispositivos que exigen prestaciones elevadas y al mismo tiempo es independiente del procesador. Así, se pueden tolerar las diferencias de velocidad entre el procesador y el bus de altas prestaciones y las variaciones en la definición de las líneas de los buses. Los cambios en la arquitectura del procesador no afectan al bus de alta velocidad y viceversa.

2.1.2 Estructura del bus.

Tanto en la arquitectura de bus tradicional como la del bus de altas prestaciones, hay un bus que se repite: el bus de sistema.

Primero hemos de considerar que el bus de sistema conecta los componentes principales del

computador, CPU, memoria, E/S, etc., por lo que es un bus de muy altas prestaciones a través

del cual la información circula por medio de señales eléctricas.

El bus de sistema está constituido, usualmente, por desde 50 hasta más de 500 líneas, dependiendo mucho del tipo de computador. Las líneas de señales que forman el bus tienen una función dedicada a un sólo uso. A pesar de que estos buses son muy diferentes entre los

computadores/microprocesadores actuales, todos tienen algo en común: que estas líneas se

pueden agrupar en tres grandes funcionalidades: líneas de datos, líneas de dirección y líneas de control.

2.3.2.1 Líneas de datos.

Las líneas del bus de sistema que se conocen como líneas de datos, son las encargadas de proporcionar un camino para transmitir datos entre los módulos que se conectan al bus.

Generalmente, un bus está compuesto por 8, 16 o 32 líneas; y los de muy alto rendimiento pueden llegar a las 64, 128 o 256 líneas distintas, cuyo número se conoce como anchura del bus de datos. Puesto que cada línea solo puede transportar un bit cada vez, el número de líneas determina cuantos bits se pueden transferir al mismo tiempo.

La anchura del bus es un factor clave a la hora de determinar las prestaciones del conjunto del sistema. Por ejemplo, si el bus de datos tiene una anchura de 8 bits, y las instrucciones son de 16 bits, entonces la CPU debe acceder al módulo de memoria dos veces por cada ciclo

de instrucción. Además, en los computadores actuales el bus de datos se optimiza al tamaño de palabra sobre el tipo de datos en el que el procesador es óptimo en su funcionamiento. Lo que nos permite tener una referencia de las prestaciones de las unidades de sistemas.

2.3.2.2 Líneas de dirección

Las líneas del bus de sistema que se agrupan con el nombre de líneas de dirección se utilizan para designar la fuente o el destino del dato situado en el bus de datos. Por ejemplo, si la CPU desea leer un dato de la memoria, sitúa la dirección de la palabra deseada en las líneas de direcciones.

Claramente, la anchura del bus de direcciones determina la máxima capacidad de memoria física que se podría poner a un sistema computacional.

Las líneas de direcciones, generalmente se utilizan también para direccionar los puertos de E/S. Usualmente, los bits de orden más alto se utilizan para discriminar entre una posición de memoria y un puerto de E/S dentro de un módulo.

2.3.2.3 Líneas de control.

Las líneas de control se utilizan para controlar el acceso y el uso de las líneas de datos y de direcciones. Puesto que las líneas de datos y de direcciones son compartidas por todos los componentes, debe existir una forma de controlar su uso.

A diferencia de las líneas de datos o de direcciones, las líneas de un bus dedicadas a control no tienen una repercusión directa en la eficiencia de la máquina. El número de líneas de control son simplemente las que tienen que ser: exactamente las mínimas necesarias para que cumplan con su cometido.

Las señales de control transmiten tanto órdenes (especifican las operaciones a realizar) como información de temporización (indican la validez de los datos y las direcciones) entre los módulos del sistema.

2.2 Tipos de interfaces.

2.2.1 Serial.

La interfaz serial transmite en forma secuencial en el tiempo todos los bits de la palabra, uno tras otro, por una sola línea de datos.

Los sistemas de comunicaciones serie han alcanzado un alto grado de estandarización desde hace tiempo. Existen normas universalmente aceptadas que fijan completamente todos los detalles de la comunicación, incluyendo aspectos mecánicos (tipo de conector y distribución de señales en las patillas de este), aspectos eléctricos (niveles y formas de las señales) y aspectos lógicos (sistemas de codificación y sincronización, y descripción de todos los circuitos de datos, control y temporizado). Estos estándares han conducido a que la mayoría de los fabricantes de procesadores y periféricos incorporen en sus equipos interfaces serie que cumplen las normas especificadas, de forma que se pueda realizar con toda facilidad la conexión indistinta de cualquier terminal o periférico con cualquier procesador. Así, se utilizan interfaces serie para conectar periféricos, como terminales de pantalla o impresoras, a computadores, aunque su distancia sea reducida y puedan, por tanto, usarse interfaces de tipo paralelo.

Por último, un tercer campo en que se utilizan sistemas de manipulación de datos en serie es el de los controladores de unidades de almacenamiento de informaciones digitales sobre soportes magnéticos (discos, cassettes y diskettes). En ellos se graban y se leen los datos en forma serie, presentándose problemas de codificación comunes con los sistemas de comunicaciones. En los dispositivos de almacenamiento, los datos son almacenados en serie, por lo que, aunque el dispositivo se conecte al sistema central mediante una interfaz paralela, el canal de lectura/escritura trabaja siempre en modo serie.

Problemas en las transmisiones serie

Cuando se transmiten informaciones a través de una línea serie es necesario utilizar un sistema de codificación que permita resolver los siguientes problemas: Sincronización de bit, Sincronización de carácter, Sincronización de mensaje.

Sincronización de bit: El receptor necesita saber exactamente dónde empieza y dónde termina cada bit en la señal recibida para efectuar el muestreo de la misma en el centro de la celda de bit.

Sincronización de carácter: La información en serie se transmite, por definición, bit a bit, pero la misma tiene sentido en palabras, por ejemplo, de 8 bits. El sistema de codificación usado debe permitir distinguir sin ambigüedades dentro de una corriente de bits cuáles son los 8 bits que forman una palabra. Normalmente se resuelve enviando los bits de cada carácter separados por alguna señal de sincronismo.

Otros sistemas, utilizados usualmente en los sistemas de comunicaciones serie, son: **Asíncrono:** Cada carácter va señalado mediante dos bits, uno al principio, bit de arranque, y otro al final, bit de parada. Estos bits permiten reconocer las fronteras de los caracteres. **Síncrono:** Cada mensaje o bloque de transmisión va precedido por unos caracteres de sincronismo. Cuando el receptor identifica una configuración de bits igual a la de los caracteres de sincronismo, da por detectado el inicio de los datos y a continuación, por conteo de bits y caracteres identifica todos los caracteres del bloque.

Sincronización de mensaje: En un sistema de comunicaciones generalmente las informaciones se transmiten en bloques de caracteres. Por sincronización de mensaje entendemos el mecanismo por el cual un conjunto de palabras es interpretado correctamente. Este problema normalmente no incumbe a los circuitos de codificación, sino al procesador que lo utiliza. El conjunto de reglas (protocolo) que permiten interpretar correctamente los mensajes suele estar controlado por una tarea software (un programa) que ejecuta el ordenador, aunque actualmente hay ciertos circuitos integrados LSI que efectúan alguna de estas tareas.

2.2.2 Paralela.

La interfaz paralela transmite simultáneamente, por líneas separadas, todos los bits de la palabra, junto con una señal de reloj que indica el momento en que está presente una palabra de información en las líneas de datos.

El método paralelo es utilizado para transmisiones a alta velocidad entre dos sistemas; no obstante, cuando la distancia entre ambos aumenta, el coste de la

línea y el de los amplificadores de transmisión y recepción puede llegar a crecer de forma tal que, desde el punto de vista económico, sea preferible utilizar un sistema serie de comunicaciones.

Las normas referidas a las interfaces paralelas han aparecido más recientemente ya que durante más tiempo han permanecido como interfaces propietarias de los distintos fabricantes. Actualmente el nivel de normalización de éstos los sitúa a la altura de las interfaces serie.

2.2.3 SCSI.

La conexión SCSI (Small Computer System Interface) es parecida a la conexión paralela en el sentido de que los datos están se transmiten en bloques de ocho. Se puede ver la configuración de los conectores en la figura 2.4. Es mejor porque se puede conectar hasta siete dispositivos (que tienen un puerto SCSI) con el mismo puerto SCSI de la unidad del sistema. Son populares para discos duros y CD-ROM externos debido a su velocidad de conexión.



Figura 2.4: Conector SCSI

Antes de la aparición de USB y FireWire, especialmente la segunda versión de cada uno, las conexiones SCSI eran comunes en equipos de altas prestaciones debido a velocidad de transferencia de datos. La razón principal por la que SCSI nunca hubiera sido aceptada universalmente como conexión de alta velocidad era la ausencia de un estándar. Cada empresa tenía su propia idea sobre como los dispositivos SCSI debían funcionar. Así que un elemento de hardware SCSI con su adaptador no funcionaba con un adaptador de otra empresa.

SCSI-1 fue estandarizado por el ANSI19 en 1986. Aunque detallaba las conexiones eléctricas y físicas de SCSI, no incluyó un conjunto de instrucciones que cada empresa podría implementar. Más adelante, se produjo un conjunto mínimo de 18 instrucciones llamado CCS (Common Command Set). A partir de entonces todos

los dispositivos SCSI soportaban el CCS. SCSI-2 se basó en el CCS para producir una versión SCSI para nuevos dispositivos e incluso dos variantes, una versión con más velocidad llamada Fast SCSI y otra versión de 16-bit llamada Wide SCSI. Se está desarrollando actualmente un nuevo estándar SCSI-3. Como parte de ese estándar existe el Ultra SCSI, que va al doble de velocidad que el Fast SCSI. Se puede ver un resumen de las características de SCSI en la tabla 2.1. También hay otra variante de SCSI-3 prevista que se llama Serial SCSI, que va a permitir velocidades de transferencia de hasta 100MB.

Tabla 2.1: Características principales de SCSI

Ancho de banda del bus	SCSI	Fast SCSI	Ultra SCSI	Cable
8-bit	5 MB/seg	10 MB/seg	20 MB/seg	50-pin
16-bit	10 MB/seg	20 MB/seg	40 MB/seg	68-pin

Aunque en muchos círculos se ve SCSI como más limitado que otros protocolos como USB y FireWire, la realidad es que SCSI proporciona el mejor soporte para múltiples dispositivos y no produce carga en la CPU porque funciona independientemente. Se podrían conectar 15 o 39 dispositivos (en uno o dos canales) sin problemas de rendimiento. Por lo tanto, aunque no ha sido muy popular en ámbitos no profesionales, es la interfaz preferida para conseguir un rendimiento y flexibilidad máximos.

2.2.4 PCI

El bus PCI (Peripheral Component Interconnect) ha sido el sistema estándar usado durante los últimos diez años, pero de acuerdo con el tiempo de vida en el diseño de este, está comenzando a ser obsoleto. Las extensiones al modelo de PCI clásico, como son los conectores de 64 bits en velocidades de 66 a 100 MHz son muy caras, por lo que no van a la par de la tecnología que surge cada año y que cada vez es menos costosa.

Desde que surgió la primera computadora personal en 1980, han existido muchos cambios en la forma como son transmitidos los datos entre una computadora y sus dispositivos periféricos, con el fin de mejorar la velocidad de transmisión y la eficiencia de los datos. El bus USB, el Serial ATA o la RDRAM son ejemplos del cambio desde una arquitectura en paralelo a un formato serie de alta velocidad. Los dispositivos anteriores fueron diseñados para permitir el máximo ancho de banda y ofrecer escalabilidad entre los nuevos modelos de computadoras que surgen año con año. La compañía Intel presentó en el año de 1991, la especificación 1.0 del bus PCI. El grupo PCI Special Interest Group, se encargó a partir de ese momento del desarrollo de las nuevas características del Bus, más tarde anunció la nueva revisión, la 2.0, en mayo de 1993. En ese tiempo el bus más popular era el Bus VESA (Local Bus (VL-bus o VLB)), esta tecnología fue creada por la asociación de estándares de Video (Video Electronics Standards Association), y consistía en un bus de 32 bits integrado a los dos conectores de los buses ISA clásicos, que operaba a una velocidad de 33 MHz, con un rendimiento muy significativo sobre el Bus ISA. Curiosamente esta última característica fue, irónicamente, la razón principal de su hundimiento, ya que en esencia fue una extensión directa del bus externo del procesador 486, y funcionaba, por tanto, a la misma velocidad que el procesador por lo que de ahí viene su nombre de local bus. Se presentó el problema de la sobrecarga del bus: no se podían conectar demasiados dispositivos, pues podía ocurrir que las pérdidas de corriente hiciesen las señales completamente ininteligibles para los dispositivos o para el procesador. VESA aconsejaba no usar más de dos dispositivos si se trabajaba a 33 MHz, aunque se podía subir hasta tres si se añadían buffers intermedios. Sin embargo, el problema más grave era que, al trabajar a la misma velocidad que el procesador, surgieron serios problemas, pues a mayor velocidad de los periféricos, más caros serán. En la práctica, se fabricaron muy pocos dispositivos capaces de trabajar a 40 o más MHz.

El bus PCI tiene varias ventajas sobre el bus anterior, para empezar, está aislado del bus de la CPU, pero permite a los periféricos acceder a la memoria del sistema. Además, también es capaz de actuar asíncronamente respecto del procesador, pudiendo trabajar a 25, 30 o 33 MHz. Esto significa que la velocidad del bus se

mantiene constante, aunque aumente la velocidad del procesador. Además, el bus PCI permite 5 o más conectores, duplicando la oferta del bus local, y, además, sin restricciones respecto a la velocidad del procesador.

Otra característica del PCI es la simplicidad de uso. El Plug and Play permite la configuración automática de los periféricos, sin que el usuario necesite asignar la IRQ, el DMA o los puertos de entrada y salida. Además, permite que varios periféricos compartan la misma interrupción, corrigiendo así los errores clásicos que tenían las microcomputadoras. Una característica adicional es el “bus mastering”, la cual permite a los dispositivos tomar control del bus y realizar transferencias entre ellos y otros dispositivos, o la memoria, sin que intervengan los procesos de la CPU, lo que reduce la latencia y la carga de trabajo del procesador.

Su introducción en los sistemas Pentium, junto con sus claros beneficios sobre sus rivales, ayudaron al PCI a ganar la guerra de los buses en 1994. Desde entonces, prácticamente todos los periféricos, desde controladores de disco duro y tarjetas de sonido hasta tarjetas de video, han sido fabricados para este bus.

2.2.5 USB.

La conexión USB (Universal Serial Bus) es el nuevo estándar para la conexión de periféricos al PC. USB es el resultado de un trabajo que se empezó en los años 90 por ingenieros de las empresas Compaq, DEC, IBM, Intel, Microsoft, NEC y Northern Telecom, para producir una especificación para una interfaz serie (en la que los datos van en secuencia, no en paralelo) que sustituyera a los puertos RS232 y Centronics. La motivación para el trabajo era que un usuario debía poder utilizar múltiples periféricos en su computador sin necesidad de instalar tarjetas especializadas, ni asignar recursos del sistema manualmente, ni reiniciar la máquina. Cada puerto USB conectado a la placa madre puede tener hasta siete puertos adicionales integrados, y cada puerto adicional puede tener hasta siete puertos más integrados, etc., hasta llegar al máximo de 127 dispositivos USB. Se lleva a cabo la división de un puerto USB utilizando un hub, que es un centro de distribución de comunicaciones que permite la conexión de más periféricos USB al bus (un canal con protocolo propio). Es común que los dispositivos USB como

monitores y teclados vengan con hubs integrados. La separación máxima entre cada hub y el periférico ha de ser 5 metros. Técnicamente USB no es un bus como PCI, porque aporta más que una simple interfaz al computador. Tampoco es un protocolo de comunicación como los puertos serie y paralelo, que requieren una interfaz externa para conectar con el computador. Se suele decir que USB proporciona una comunicación bi-direccional entre el computador y los periféricos utilizando un bus de topología de estrella en capas (tiered star topology; similar a una estructura arbórea con ramificaciones), como se puede ver en la figura 2.5. Aun así, cada periférico se comunica con el computador como si tuviera su propia línea de comunicación directa y los hubs fueran transparentes.

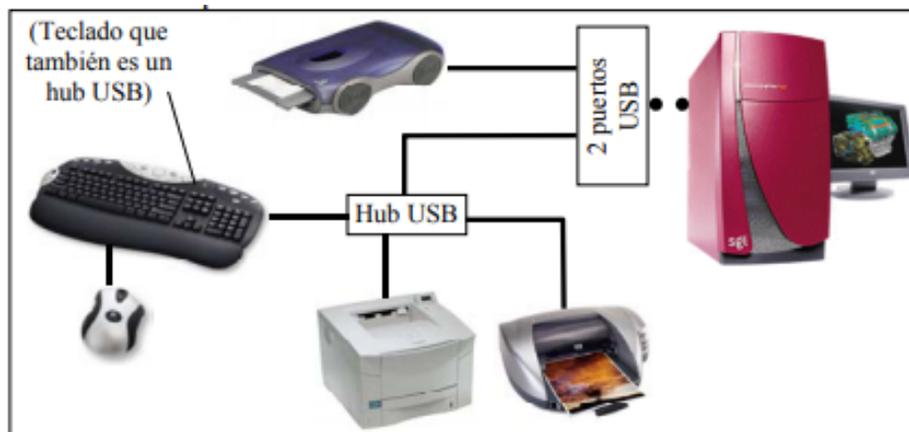


Figura 2.5: Ejemplo de la estructura de un bus USB

Los primeros dispositivos que incluyeron conexiones USB fueron los periféricos de interfaz humana, los teclados, ratones, etc. Ahora existe una gama muy amplia como los módems, los escáneres, las impresoras, las unidades de almacenamiento de distintos tipos (discos duro, unidades de cinta, las unidades de CD-ROM, etc). La velocidad de transferencia de 12 Mbits hace al USB ideal para algunos tipos de transferencia de vídeo en tiempo real. La aparición de controladores USB baratos hace que sean un tipo de conexión buena y popular en los computadores de sobremesa. Para poder asegurar la interoperabilidad de los dispositivos con USB era necesario definir estándares tanto del hardware, que permite la conexión física, como del software, que lleva a cabo el control y la transferencia de datos. Los dispositivos más sofisticados requieren software adicional para permitir su conexión

al sistema operativo. Otra característica interesante de USB es que lleva una conexión de corriente eléctrica a los dispositivos conectados que se puede usar para regularlos. Los dispositivos USB se clasifican en función de la cantidad de corriente que requieren. Los de baja energía no necesitan más de 100 mA y los de energía alta pueden necesitar más de 500 mA. Los hubs pueden ser de ambos tipos o incluso llevar su propia fuente de alimentación; dicha característica del USB, aunque útil en la mayoría de los sistemas de sobremesa donde no hay problemas de alimentación, puede producir problemas para los portátiles que están utilizando baterías. La versión 1.1 de USB tiene una velocidad de transferencia de 12 Mbits y no está diseñada para transferencias de alta velocidad que serían necesarias para, por ejemplo, hacer copias de seguridad de grandes cantidades de datos en un tiempo razonable o imprimir en color en impresoras de alta resolución, etc. Así que la nueva versión de la especificación USB, la 2, permite velocidades de transferencia de hasta 480 Mbits. Es compatible con los dispositivos USB 1.1. La limitación principal de USB es la dificultad en conectar dispositivos que utilizan otros protocolos de comunicación. Hay empresas que se dedican a producir hardware que puede servir como una interfaz en esos casos.

2.2.6 FireWire.

El estándar IEEE-P1394 se denomina también Firewire, aunque en realidad Firewire no es más que una de las implementaciones posibles que obedecen a este estándar. Apple desarrolló el estándar IEEE 1394 en 1987. Ya especificación de Nubus90, aplicada en los modelos Quadra de Macintosh y con ligeras modificaciones también en ordenadores NeXT, preveía un bus serie auxiliar de estas características.

Con el fin de estos modelos en un primer momento dio la impresión de que el estándar IEEE 1394 también caería en el olvido, pero recuperó el interés con el auge de los sistemas de bus serie para periféricos (IEEE 1394, USB, SSA, Fibre Channel). En 1994 se constituyó la 1394 Trade Organization, compuesta por compañías como Adaptec, AMD, Intel y Microsoft, así también Sony, JVC y

Yamaha. Lo que significa que Firewire se desmarca como una buena solución para aplicaciones de Vídeo y Audio. Así mismo, fabricantes de instrumentos de medida como Hewlett-Packard o National Instruments consideran a Firewire como un posible sucesor del viejo bus IEC para tecnologías de medida con soporte informático.

El protocolo IEEE-1394 abarca tres capas: Transaction, Link y Physical Layer, con sus registros Standard Control y Status, la capa Transaction constituye la conexión a un bus paralelo, normalmente PCI.

La capa Link convierte las informaciones de registro de la capa de Transaction, realiza la formación de paquetes y el control (Cycle Control) del modo síncrono. La capa Physical es responsable de la inicialización de los nodos y de realizar el proceso de arbitraje, que también lleva a cabo la conversión de datos serie para la capa Link.

Para una aplicación de sonido, y mucho más de vídeo, es importante que esté garantizado un ancho de banda para datos y con él un flujo de datos continuo, un requisito fundamental en todas las aplicaciones de tiempo real. Una red normal Ethernet o Token Ring trabaja en modo asíncrono, lo que significa que aquellos paquetes de datos que primero acceden a la red también son los que transportan primero y los siguientes sufren un retardo que el usuario puede ignorar porque es insignificante para el contenido de la información.

Por el contrario, el retraso de un frame de vídeo es perceptible en la imagen y molesto. Por tanto, un sistema de bus para aplicaciones de vídeo deberá disponer de un mecanismo que impida los retardos, y para ello Firewire posee el modo síncrono. Este término griego significa aquí que a una transmisión de datos se concede un período o intervalo constante, disponiendo entonces de un ancho de banda garantizado. Esta es una de las características esenciales de Firewire, que también está definida para el puerto Universal Serial Bus (USB) aunque en este segundo caso con un ancho de banda sustancialmente menor (12 Mbits/s frente a 200 Mbits/s).

El estándar comprende de definiciones para una conexión Backplane y también para conexión de cables de unidades IEEE-1394. La solución de cable está dirigida

al mercado del PC mientras que la versión Backplane encuentra aplicación en sistemas industriales.

De momento hay definidos tres ratios de transferencia: 100, 200 o 400 Mbits/s. El dispositivo más lento conectado al bus determina la velocidad máxima.

Firewire utiliza 64 bits de direccionamiento y trabaja según el estándar Control and Status Register Management para bus de Microcomputer (IEEE 1291). Los 16 bits superiores están reservados para direccionar los dispositivos (nodos) y los 48 bits restantes para la implementación de registro y memoria.

La nueva especificación, IEEE 1394b, soporta velocidades de 800, 1600 y 3200Mbps. Es un bus de alta velocidad muy apropiado para aplicaciones de multimedia y comunicaciones. Su velocidad lo convierte en una tecnología muy importante para la convergencia de la electrónica doméstica (las cámaras de vídeo, reproductores DVD y de CD, etc.) y los computadores. Antes eran necesarias tarjetas especiales para poder conectar todos esos elementos. Para poder realizar estos tipos de conexiones los dispositivos necesitan velocidades de transferencia garantizadas. De no ser así, al transferirse los datos, como el vídeo en tiempo real, se podría perder información. FireWire lo garantiza con la transferencia síncrona (una transferencia que ocurre cada cierto tiempo) de datos. El bus es también “plug-and-play” porque se reconfigura automáticamente cada vez que se conecta un dispositivo nuevo y los dispositivos pueden comunicarse entre ellos sin la necesidad de un sistema central. La configuración de los 63 dispositivos que se pueden conectar como máximo es flexible y se puede tener un máximo de 16 cables de hasta 4,5 metros. En general, los dispositivos FireWire tienen un puerto o varios. Un dispositivo de tres puertos se llama un ramo o hub, uno de dos puertos, un nodo “pass-through” y uno de un puerto, un nodo hoja.

2.2.7 AGP

El bus AGP (Accelerated Graphics Port, Puerto de gráficos acelerado) es una especificación de Intel desarrollada en 1996 para tratar de paliar el cuello de botella que se estaba creando para los gráficos en el bus PCI, ya que el sistema gráfico

siempre es el que más recursos necesita en los sistemas multimedia modernos. Lo que se pretende con este nuevo bus es proporcionar un gran ancho de banda entre el procesador y la memoria de video, ya que las aplicaciones de gráficos tridimensionales demandan una velocidad que el bus PCI no puede proporcionar.

El bus AGP es de 32 bits y posee 8 canales adicionales para acceso a la memoria principal. Además, puede acceder directamente a esta a través del puente norte, pudiendo emular así memoria de video en la memoria principal, con lo que la memoria de video ya no está limitada a la instalada en la tarjeta gráfica. Con eso se consigue utilizar parte de la memoria principal como memoria de texturas tridimensionales para ayudar al proceso de gráficos 3D.

La velocidad del bus es de 66 MHz, ya que en realidad es el mismo bus local del procesador, por lo que el límite del ancho de banda es 528 MB/s.

El bus AGP cuenta también con varios modos de funcionamiento que son los siguientes:

- AGP 1X: velocidad 66 MHz con una tasa de transferencia de 266 MB/s.
- AGP 2X: velocidad 133 MHz con una tasa de transferencia de 532 MB/s.
- AGP 4X: velocidad 266 MHz con una tasa de transferencia de 1 GB/s.
- AGP 8X: velocidad 533 MHz con una tasa de transferencia de 2 GB/s.

Estas tasas de transferencias se consiguen utilizando un multiplicador sobre el reloj del bus AGP que funciona a 66 MHz.

El bus AGP se utiliza exclusivamente para conectar tarjetas gráficas y sólo tiene un slot. Este slot mide 8 cm y se encuentra sobre los slots PCI y desplazado hacia la derecha con respecto a estos. Suele ser de color marrón para distinguirlo rápidamente de los otros slots de tarjetas.

2.2.8 Otros

PCI Express

El bus PCI Express, antes conocido como 3GIO, está diseñado para reemplazar al PCI y cubrir las necesidades de interconexión para la próxima década. Está diseñado para soportar diversos segmentos del mercado, y como una arquitectura de E/S que unifique los equipos de escritorio, portátiles, servidores, estaciones de trabajo y dispositivos empotrados.

El objetivo principal es conseguir un menor costo que los dispositivos PCI, tanto en bajos como altos volúmenes de producción. Para ello emplea un bus serie en vez de paralelo, pues al requerir un menor número de pistas en las placas, se reducen los costes de diseño a la vez que se aumenta el rendimiento en cuanto a espacio consumido.

El PCI Express aparece ante el sistema operativo exactamente igual que el antiguo bus PCI, por lo que no habrá que hacer modificaciones sustanciales en ellos. Lo mismo se aplica a nivel de configuración y controladores de dispositivos, que serán compatibles con los actuales para PCI. La escalabilidad en el rendimiento se consigue aumentando la frecuencia y añadiendo rutas al bus (habría varios buses serie funcionando en paralelo, cada uno independiente de los demás).

El diseño está pensado para ofrecer una alta velocidad de transferencia en cada ruta, con baja sobrecarga y latencia. Permite, además, varios canales virtuales en cada ruta o enlace físico. Finalmente, al ser una conexión punto a punto, permite que cada dispositivo tenga una conexión dedicada, evitando compartir el bus. Otras características avanzadas son:

- Soporte de múltiples estructuras de datos.
- Capacidades avanzadas de gestión de energía.
- Conexión y desconexión «en caliente» de periféricos.
- Comprobación de errores de transmisión
- Capacidad de transferencia isócrona
- Protocolo dividido en capas y basado en envío de paquetes.

Visto a alto nivel, un sistema PCI Express está formado por un sistema raíz (que estará bien en el North Bridge o en el South Bridge), uno o varios conmutadores

(switches) y los dispositivos finales. La novedad aquí es el switch, el cual permite la comunicación punto a punto entre dispositivos finales, evitando enviar tráfico hasta el bridge si éste no supone problemas de coherencia en las cachés (por tratarse, por ejemplo, de transferencias a memoria).

En la parte inferior está la capa física. El enlace más simple para un sistema PCI Express consiste en dos señales diferenciales por corriente. Se incluye una señal de reloj usando el sistema 8/10b para conseguir altas velocidades de transferencia. La frecuencia inicial es de 2'5 Gb/s en cada sentido, y se espera que los avances en la tecnología del silicio permitan aumentarlo hasta los 10 Gb/s.

Una de las características más sobresalientes es la capacidad del PCI Express de aumentar la velocidad mediante el añadido de nuevos enlaces formando múltiples rutas paralelas. La capa física soporta anchos X1, X2, X4, X8, X12, X16 y X32. La transmisión sobre múltiples rutas es transparente al resto de las capas.

La capa de enlace es la encargada de garantizar la fiabilidad y la integridad de los datos para cada paquete enviado a través de un enlace PCI Express. Junto con un número de secuencia y un CRC, un protocolo de control de flujo garantiza que los paquetes son transmitidos sólo cuando hay un buffer disponible para recibirlos en el otro extremo. Los paquetes corruptos se retransmiten automáticamente.

La capa de transacción crea los paquetes con las peticiones de la capa de software a la capa de enlace, implementándolas como transacciones. Cada paquete tiene un identificador único, soportando direccionamiento de 32 y 64 bits. Otros atributos extra incluyen «no-snoop», «relaxed ordering» y prioridad, y se usan para el enrutado y la calidad de servicio.

La capa de transacción se encarga de cuatro espacios de direccionamiento: memoria, I/O, configuración (estos tres ya existían en la especificación PCI) y el nuevo espacio Mensajes. Este reemplaza a ciertas señales en la especificación PCI 2.2 y elimina los ciclos especiales del viejo formato, lo que incluye las interrupciones, las peticiones de gestión de energía y el reinicio.

Finalmente, la capa software constituye la clave para conseguir la compatibilidad software. La inicialización y el runtime no se han cambiado respecto al PCI debido a que se quiere que los sistemas operativos puedan usar PCI Express sin necesidad

de modificarse. Los dispositivos son enumerados de forma que el sistema operativo pueda encontrarlos y asignarles recursos, mientras que el runtime reutiliza el modelo cargar/ almacenar y memoria compartida del PCI.

Las primeras implementaciones estas diseanadas para coexistir con los actuales conectores PCI. Ası, un conector 1X encaja a continuacion de un conector PCI, entre este y el borde de la placa madre, de forma tal que se puede usar cualquiera de las dos tarjetas.

BUS IEEE-488

La interfaz IEEE-488 es el resultado de la normalizacion de un bus propietario de la compana Hewlett-Packard. Esta empresa comenzo el desarrollo del bus en 1965, para la interconexion de los instrumentos de laboratorio de esta. El objeto era un bus de proposito general, destinado a simplificar el diseno y la integracion de equipos de medida entre sı y especialmente de estos con el ordenador. Dicha simplificacion se consigue al reducir al mınimo los problemas tanto electricos como mecnicos y de compatibilidad funcional entre equipos, poseyendo la suficiente flexibilidad para acomodar un amplio y creciente numero de productos. El nombre inicial fue el de HPIB (Hewlett Packard Interface Bus). Rapidamente, numerosas empresas empezaron a comercializar equipos que incorporaban este tipo de interfaz con el nombre mas habitual de GPIB (General Purpose Interface Bus). Estos trabajos iniciales fueron del interes de la comision Electrotecnica Internacional (IEC) y del Instituto de Ingenieros Electricos y Electronicos (IEEE) y respaldaron el borrador propuesto por Hewlett-Packard. El IEC-625 y el IEEE-488/1978 son los nombres oficiales de las normas publicadas por los dos organismos anteriormente mencionados. La interfaz IEEE-488 se aplica a sistemas de interconexion de instrumentos en los cuales:

- ✓ El intercambio de informacion entre los equipos interconectados sea de naturaleza digital. Este aspecto parece obvio, puesto que todos los perifericos que hemos tratado hasta ahora han sido de naturaleza digital. Sin embargo, este bus se concibio para interconectar instrumentos de medida, como generadores de seal, osciloscopios, analizadores de espectro,

fuentes de alimentación, multímetros, etc. que en principio tienen una naturaleza analógica.

- ✓ El número de equipos a interconectar no exceda de 15. La limitación a 15 aparatos se debe a que por ser un procedimiento asíncrono la sobrecarga del bus hace que el funcionamiento no sea fiable para una carga mayor.
- ✓ Las longitudes totales de transmisión sobre los cables de interconexión no excedan de 20 metros o de dos metros por equipo, cuando no se utilicen técnicas especiales de ampliación del bus.
- ✓ La velocidad de los datos en la interconexión y en cualquier línea de esta no supere la cantidad de 1 MByte/segundo. Consideraciones prácticas hacen que el límite de velocidad de transmisión de datos sea del orden de 250 KBytes/s. No obstante, la revisión IEEE-488.2 garantiza velocidades de 1MB/segundo y casi cualquier interfaz comercial supera con creces esta velocidad.

EL INTERFAZ MIDI

El interfaz MIDI fue diseñado originalmente para la interconexión de instrumentos musicales digitales entre sí y de estos con un ordenador. No es de extrañar por tanto que su gestación y sus características estén estrechamente relacionada con el mundo de los instrumentos musicales y de la música y el espectáculo en general. A comienzos de los setenta comenzaron a aparecer los primeros sintetizadores electrónicos de tipo analógico. Un sintetizador es un instrumento que genera los sonidos musicales a partir de elementos electrónicos básicos, como osciladores, generadores de envolvente o de rampa, filtros, etc. Este primer tipo de sintetizadores se podían conectar entre sí de forma analógica con una señal que proporcionaba una tensión de 1 voltio por octava. De esta forma, señales que se diferenciaban entre sí en un voltio, representaban la misma nota, pero de octavas adyacentes. Con este sencillo interfaz electrónico, se podía gobernar un sintetizador de un fabricante con un teclado de otro. Rápidamente se comenzó a introducir la digitalización y el control por ordenador de los múltiples osciladores de los sintetizadores polifónicos. A partir de este momento, el sencillo interfaz analógico

de un voltio por octava dejó de ser aplicable y los instrumentos volvieron a estar incomunicados entre sí. Algunas compañías, a la vista de las limitaciones impuestas a los usuarios, comenzaron a desarrollar estructuras de bus capaces de permitir distintas expansiones. Algunas de ellas empleaban el sistema de bus serie, con objeto de rebajar el coste, mientras otras elegían el bus paralelo debido a su mayor rapidez. En lo único que estaban todas de acuerdo era en la necesidad de desarrollar una interfaz apropiada y común a todas ellas. En diciembre de 1982, Sequential Circuits Inc. (fabricante del Prophet, primer sintetizador polifónico de difusión masiva) lanzó las primeras unidades del Prophet 600. Una de sus características más interesantes era que disponía de una conexión de interfaz serie, y que Dave Smith, presidente de Sequential denominó entonces como Universal Synthesizer Interface (USI). Durante la Feria de Invierno de Fabricantes de Música que se desarrolló aquel mismo año, técnicos de Sequential, Yamaha y algunos otros fabricantes celebraron una reunión informal en la que comenzaron a discutir las bases de una posible estandarización. Como resultado de estas conversaciones apareció un protocolo muy similar al USI de Dave Smith, y que ofrecía la mejor relación entre velocidad, simplicidad y bajo costo. El junio de 1983 se conectó un Prophet 6000 a un Yamaha DX-7. El resultado careció por sí mismo de espectacularidad, pero motivó el que, en agosto de 1983, representantes de Sequential, Roland, Yamaha, Korg y Kawai sentaran en Tokio las bases de la norma MIDI 1.0 (Musical Instrument Digital Interface). Esta interfaz es quizá la más extendida y prácticamente única dentro de su campo de aplicación pese a no haber sido respaldada por ningún organismo internacional de normalización, pero no hay instrumento musical electrónico que se precie, que no disponga de una conexión MIDI, e incluso resulta muy económico incorporar este tipo de interfaz a cualquier ordenador. De hecho, casi cualquier tarjeta de sonido convencional incluye uno.

La característica técnica más importante del MIDI reside en que, para abaratar los cables y las conexiones, se ha usado un protocolo serie, básicamente el mismo que el RS-232 con un bit de comienzo (START), 8 bits de datos y dos bits de fin (STOP). Funciona a una velocidad de 31,25 kilobaudios, lo que a primera vista sin duda parece un poco extraño. Pero 31,25 Kbd. no es una velocidad tan extraña si

consideramos el aún popular (y bastante barato hoy día) adaptador de comunicaciones asincrónicas 6850, cuyos registros de control internos actúan en modo división por 64, y son controlados externamente por un reloj de transmisión/recepción que funciona a 2 MHz., lo que arroja un resultado de 31,25 Kbd. A diferencia del RS-232 (que usa una tensión bipolar), el MIDI utiliza un bucle de corriente de 1,5 mA con un optoacoplador en la entrada del receptor.

ACTIVIDADES

1. Busque en el manual de la placa base de tu ordenador las velocidades de transferencias que soportan los diferentes buses en ella (serial, usb, etc.); además de sus principales características. Sustenta y compara tus resultados con tus compañeros.

2. Cuestionario. Responda Verdadero o Falso a las preguntas y sustente su respuesta.

- Dentro de la visión multinivel de la estructura de un computador, el nivel de “Arquitectura” se encuentra por encima del nivel “Físico” y por debajo del nivel de “Aplicación”.
- Una forma de aumentar las prestaciones de un computador consiste en introducir el paralelismo en el diseño, lo cual se puede conseguir replicando algunas partes de su estructura.
- La comparación de rendimiento utilizando como métrica las instrucciones por segundo sólo es fiable cuando las arquitecturas son similares.
- Los buses de Entrada /Salida suelen estar estandarizados y permiten conectar un número indeterminado de dispositivos.
- La conexión de todos los dispositivos a un único bus puede provocar incompatibilidades y una pérdida global en el rendimiento del sistema.
- En una jerarquía de buses, los más cortos y rápidos deben ser los buses de expansión para poder conectar de manera eficaz cualquier periférico.
- En un computador, el bus de sistema y el bus de memoria se pueden optimizar ya que se conectan a un número fijo de dispositivos de prestaciones conocida.
- El incremento de rendimiento de un bus se puede conseguir aumentando su ancho de datos, pero tiene sus limitaciones debido al mayor espacio que ocupa, al incremento de interferencias y al retraso en la propagación de las señales.
- Cuanto mayor es a sobrecarga de un bus, mayor es la cantidad de información que se puede transmitir por el bus.

Bibliografía

- Morán, C., Morán F. Mantenimiento de equipos informáticos. España.
- Richarte, J., (2012). Motherboards. 1a ed. Buenos Aires: Fox Andina; DALAGA, S.A. 192 pág. ISBN 978-987-1857-47-0. Recuperado en https://books.google.com.pa/books?hl=es&lr=&id=VUG3i30I2QC&oi=fnd&pg=PT13&dq=tipos+de+motherboard&ots=4xCCszjsLg&sig=UpT9tF0xMDAh7-IOW8alxzJARIQ&redir_esc=y#v=onepage&q=tipos%20de%20motherboard&f=false
- Moreno, J., Santos, M., (2014). Sistemas Informáticos y redes locales. Madrid: RA-MA, S.A., Editorial y publicaciones. ISBN: 978-84-9964-368-7.
- Martín, P., Oliva, J., Zarco, C. (2010). Montaje y mantenimiento de equipos. Madrid: Ediciones Paraninfo, S.A. ISBN: 978-84-9732-763-3.
- Alloza, Jesús., (2014). UF0465: Montaje de componentes y periféricos microinformáticos. IC Editorial. 344 pág. ISBN: 978-84-16173-53-2.
- García, M., López, J., (2007). Apuntes de Organización de Computadores. Asturias: Ediciones de la Universidad de Oviedo. ISBN: 978-84-8317-606-1.
- Minguet, J., Read, T. Informática Básica. Editorial Centro de Estudios Ramón Areces, S.A.
- Gallego, J., Folgado, L., (2011). Montaje y mantenimiento de equipos. Madrid: Editex, S.A.
- Álvarez Cedillo, Jesús Antonio, & García Arregui, Macario (2005). El Bus PCI Express: Generalidades. *Polibits*, (32),8-11.[fecha de Consulta 23 de Agosto de 2020]. ISSN: 1870-9044. Disponible en: <https://www.redalyc.org/articulo.oa?id=4026/402640445002>
- Stalling, W. (2006). *Organización y Arquitectura de Computadoras*. 7ª. Edición. España: Prentice Hall.
- Represa, C., Rodríguez, C. (2013). 202 Problemas de Arquitectura de Computadores. Universidad de Burgos. Recuperado en https://riubu.ubu.es/bitstream/handle/10259/3602/202_Problemas_ARCO.pdf?sequence=1

CAPÍTULO 3: LA MEMORIA

INTERNA Y EXTERNA

Objetivos:

- ✓ Conocer las características básicas de una memoria.
- ✓ Comprender los conceptos básicos sobre la organización de la jerarquía de memoria de un computador.
- ✓ Analizar cómo se gestionan los datos entre los diferentes niveles de la jerarquía de memorias, especialmente en la memoria caché.
- ✓ Obtener una visión generalizada basada en los distintos tipos de clasificación de las memorias.

¿De qué trata esta sección de aprendizaje?

Las memorias de las computadoras se utilizan para retener los bits que constituyen el alfabeto digital para representar programas y datos o cualquier otra entidad operable para una computadora. Uno de los parámetros más importantes para medir el rendimiento es la capacidad del procesador para gestionar los accesos a sus memorias.

Este capítulo abarca la descripción de memorias tanto internas como externas, las características de cada una, y cómo éstas contribuyen a que los usuarios disfruten de las principales funciones de un computador: procesar, almacenar, transferir y controlar datos.

3.1 Memoria Interna.

La memoria interna corresponde a la memoria principal (memoria RAM del computador) y adicionalmente un nivel de memoria caché o varios.

En la memoria principal se almacenan los programas que se deben ejecutar y sus datos, es la memoria visible para el programador mediante su espacio de direcciones.

La memoria principal se implementa utilizando diferentes chips conectados a la placa principal del computador y tiene una capacidad mucho más elevada que la memoria caché (del orden de Gbytes o de Tbytes en supercomputadores). Utiliza tecnología DRAM (Dynamic RAM), que es más lenta que la SRAM, pero con una capacidad de integración mucho más elevada, hecho que permite obtener más capacidad en menos espacio.

3.1.1 Características.

Las características más importantes de los diferentes tipos de memoria son la localización, la capacidad, el método de acceso, la organización de los datos en una memoria, el tiempo de acceso, velocidad, y el coste.

3.1.1.1 Ubicación.

Podemos clasificar los tipos de memoria según su ubicación dentro del computador. Básicamente, se pueden distinguir: memoria dentro del chip del procesador, memoria interna (memoria en la placa base del computador) y memoria externa. Dentro del chip del procesador habitualmente están los registros y uno o varios niveles de memoria caché. La memoria interna corresponde a la memoria principal (memoria RAM del computador) y adicionalmente un nivel de memoria caché o varios. La memoria externa corresponde a los dispositivos de almacenamiento secundario, como discos duros, unidades ópticas (CD-ROM, DVD, o Blu-ray), unidades de cinta, etc.

3.1.1.2 Capacidad.

La capacidad (o tamaño de la memoria) hace referencia a la cantidad de información que se puede almacenar. La unidad utilizada para especificar la capacidad de almacenamiento de información es el byte (1 byte = 8 bits), y a la hora de indicar la capacidad, se utilizan diferentes prefijos que representan múltiplos del byte. En el sistema internacional de medidas (SI) se utilizan prefijos que representan múltiplos y submúltiplos de una unidad; estos prefijos SI corresponden siempre a potencias de 10. Cada prefijo del sistema internacional recibe un nombre diferente y utiliza un símbolo para representarlo.

Ejemplos de prefijos SI:

$$10^3 \text{ bytes} = 1.000 \text{ bytes} = 1 \text{ Kilobyte (KB o Kbyte)}$$

$$10^6 \text{ bytes} = 10^3 \text{ KB} = 1.000 \text{ KB} = 1 \text{ Megabyte (MB o Mbyte)}$$

$$10^9 \text{ bytes} = 10^3 \text{ MB} = 1.000 \text{ MB} = 1 \text{ Gigabyte (GB o Gbyte)}$$

$$10^{12} \text{ bytes} = 10^3 \text{ GB} = 1.000 \text{ GB} = 1 \text{ Terabyte (TB o Tbyte)}$$

Ahora bien, en informática, la capacidad de almacenamiento habitualmente se indica en múltiplos que sean potencias de 2; en este caso se utilizan los prefijos definidos por la International Electrotechnical Commission (IEC). Prefijos IEC Los prefijos IEC representan múltiplos para las unidades de información bit y byte. Los nombres se han creado añadiendo el término binario a los prefijos SI. Por ejemplo, kibi sería la contracción de kilo binario.

Ejemplos de prefijos IEC:

$$2^{10} \text{ bytes} = 1.024 \text{ bytes} = 1 \text{ KiB (kibibyte)}$$

$$2^{20} \text{ bytes} = 1.024 \text{ KiB} = 1 \text{ MiB (mebibyte)}$$

$$2^{30} \text{ bytes} = 1.024 \text{ MiB} = 1 \text{ GiB (gibibyte)}$$

$$2^{40} \text{ bytes} = 1.024 \text{ GiB} = 1 \text{ TiB (tebibyte)}$$

3.1.1.3 Unidad de transferencia.

En un acceso a memoria se puede acceder a un byte o a varios, con un máximo que vendrá determinado por el número de bytes de una palabra de memoria; es decir, en un solo acceso se leen o escriben uno o varios bytes. Cuando se especifica la dirección de memoria a la que se quiere acceder, se accede a partir de esta dirección a tantos bytes como indique la operación de lectura o escritura. En memoria externa, se accede habitualmente a un bloque de datos de tamaño muy superior a una palabra. En discos es habitual transferir bloques del orden de los Kbytes.

Ejemplo:

En los procesadores x86 de 32 y 64 bits, la unidad de direccionamiento es de un byte, pero el tamaño de la palabra de memoria es de 4 bytes (32 bits). Los registros del procesador (accesibles para el programador) habitualmente tienen un tamaño igual al tamaño de la palabra de memoria; por ejemplo, en un procesador de 32 bits (como el Intel 386) el tamaño de los registros era de 32 bits (4 bytes).

Los procesadores x86-64 son procesadores con registros de 64 bits, pero en cambio el tamaño de la palabra de memoria continúa siendo de 32 bits; eso es así para mantener la compatibilidad con procesadores anteriores. No hay que olvidar que la arquitectura x86-64 es una extensión de la arquitectura de 32 bits x86-32. El tamaño de la palabra de memoria de los procesadores x86 de 32 y 64 bits es de 32 bits (4 bytes) y en un ciclo de memoria se puede acceder a 1, 2 o 4 bytes. En la arquitectura CISCA el tamaño de palabra es también de 32 bits pero accedemos siempre a una palabra de 4 bytes.

3.1.1.4 Métodos de acceso.

Cada tipo de memoria utiliza un método a la hora de acceder a las posiciones de memoria. Hay métodos de acceso diferentes característicos de cada tipo de memoria:

- **Secuencial.** Se accede desde la última posición a la que se ha accedido, leyendo en orden todas las posiciones de memoria hasta llegar a la posición deseada. El tiempo de acceso depende de la posición a la que se quiere acceder y de la posición a la que se ha accedido anteriormente.
- **Directo.** La memoria se organiza en bloques y cada bloque de memoria tiene una dirección única, se accede directamente al principio de un bloque y dentro de este se hace un acceso secuencial hasta llegar a la posición de memoria deseada. El tiempo de acceso depende de la posición a la que se quiere acceder y de la última posición a la que se ha accedido.
- **Aleatorio.** La memoria se organiza como un vector, en el que cada elemento individual de memoria tiene una dirección única. Se accede a una posición determinada proporcionando la dirección. El tiempo de acceso es independiente de la posición a la que se ha accedido y es independiente de la última posición a la que se ha accedido.
- **Asociativo:** es una memoria del tipo de acceso aleatorio que permite hacer una comparación de ciertas posiciones de bits dentro de una palabra buscando que coincidan con unos valores dados, y hacer esto para todas las palabras simultáneamente. Una palabra es por tanto recuperada basándose en una porción de su contenido en lugar de su dirección. Como en las memorias de acceso aleatorio convencionales, cada posición tiene su propio mecanismo de direccionamiento, y el tiempo de recuperación de un dato es una constante independiente de la posición o de los patrones de acceso anteriores. Las memorias caché pueden emplear acceso asociativo.

3.1.1.5 Prestaciones.

Se utilizan 3 parámetros para medir las prestaciones en una memoria:

- **Tiempo de acceso (latencia):** para memorias de acceso aleatorio es el tiempo que tarda en realizarse una operación de escritura o de lectura, es decir, el tiempo que transcurre desde el instante en el que se presenta una dirección a la memoria hasta que el dato, o ha sido memorizado, o está disponible para su uso. Para memorias de otro tipo, el tiempo de acceso es

el que se tarda en situar el mecanismo de lectura/escritura en la posición deseada.

- **Tiempo de ciclo de memoria:** este concepto se aplica principalmente a las memorias de acceso aleatorio y consiste en el tiempo de acceso y algún tiempo más que se requiere antes de que pueda iniciarse un segundo acceso a memoria. Este tipo adicional puede que sea necesario para que se finalicen las transiciones en las líneas de señal o para regenerar los datos en el caso de las lecturas destructivas. Se debe tener en cuenta que el tiempo de ciclo de memoria depende de las características del bus del sistema y no del procesador.
- **Velocidad de transferencia:** es la velocidad a la que se pueden transferir datos a, o desde, una unidad de memoria. Para memorias de acceso aleatorio coincide con el inverso del tiempo de ciclo.

3.1.1.6 Soporte físico.

Las memorias pueden tener características físicas diferentes; básicamente podemos distinguir dos clasificaciones. La primera distingue entre:

- **Memoria volátil:** memoria que necesita una corriente eléctrica para mantener su estado; estas memorias incluyen registros, memoria caché y memoria principal.
- **Memoria no volátil:** mantiene el estado sin necesidad de corriente eléctrica, incluye memorias de solo lectura, memorias programables, memoria flash, dispositivos de almacenamiento magnético y óptico.

La segunda clasificación distingue entre:

- **Memoria de semiconductores:** es una memoria que utiliza elementos semiconductores, transistores, en su construcción; incluye: registros, memoria caché, memoria principal, memorias de solo lectura, memoria flash.
- **Memoria magnética:** utiliza superficies imantadas para guardar la información; dentro de esta categoría se incluyen básicamente discos y cintas magnéticas.

- **Memoria óptica:** utiliza elementos de almacenamiento que pueden ser leídos y escritos mediante luz láser; se incluyen dispositivos de CD, DVD, Blu-ray.

3.1.2 Jerarquía de memoria.

Para equiparar la rapidez del procesador, la memoria que retiene las instrucciones y datos del programa debe ser accesible en 1 ns o menos. Se puede construir tal memoria, pero sólo en tamaños pequeños (por ejemplo, un archivo de registro). El volumen del programa y sus datos se deben mantener en memoria más lenta y grande. El reto es diseñar el sistema de memoria global de modo que parezca tener la rapidez del componente más rápido y el costo más barato. Aunque aumentar el ancho de banda de la memoria lenta puede ayudar a solventar la brecha de rapidez, este enfoque requiere métodos cada vez más sofisticados para esconder la latencia de memoria y eventualmente caer cuando la brecha crezca lo suficientemente ancha.

El hecho de que se pueden construir memorias pequeñas con tiempos de acceso de no más de unos cuantos nanosegundos lleva a la idea de una *memoria buffer* que retiene los segmentos de programa y datos actualmente más útiles (aquellos a los que se hace referencia frecuente o se sabe que serán necesarios muy pronto), en forma muy parecida a como los registros retienen algunos de los elementos de datos actualmente activos para rápido acceso por el procesador. Este *buffer* o *memoria caché* quizá es demasiado pequeña para retener todos los programas y datos requeridos, de modo que los objetos se deben llevar ahí desde la memoria principal. Tal caché no salva por completo la brecha de rapidez entre el procesador y la memoria principal, de modo que con frecuencia se usa una memoria caché de segundo nivel. Estos dos niveles de caché, junto con los registros, la memoria principal, la memoria secundaria (usualmente disco) y la memoria terciara o de archivo, se muestran en la figura 3.1:

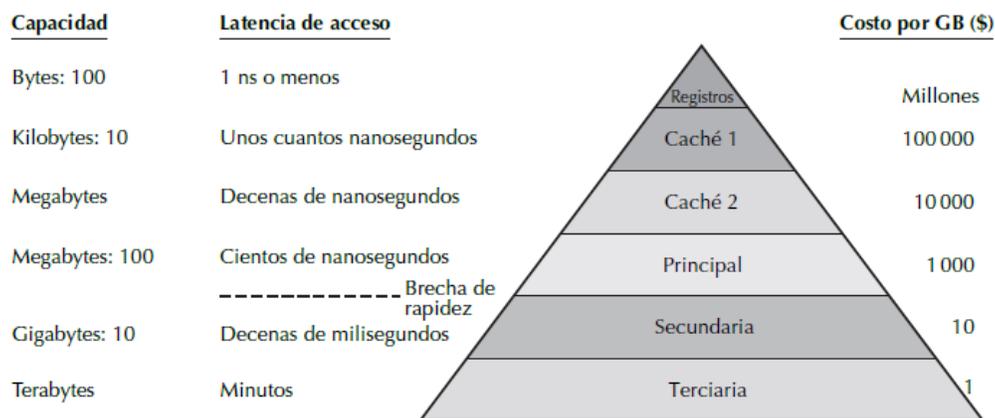


Figura 3.1: Ejemplo de jerarquía de memoria

Se dice que las memorias en interacción de diversas capacidades, rapidez y costos que se muestran en la figura 3.1 constituyen una jerarquía de memoria dentro de la cual los datos y componentes de programa más útiles de algún modo se mueven hacia niveles superiores, donde son accesibles más fácilmente. En la práctica, cada sistema de cómputo moderno tiene una memoria jerárquica, aunque no todos los niveles que se muestran en la figura 3.1 están presentes en todos los casos.

La forma piramidal de la figura 3.1 tiene la intención de presentar el tamaño más pequeño de unidades de memoria en lo alto y las mayores capacidades en el fondo. Los archivos de registro rápidos tienen capacidades que se miden en cientos o cuando muchos miles de bytes. En la actualidad, son prácticas las memorias terciarias multiterabytes, y la introducción de unidades petabyte está bajo consideración. Desde luego, el costo y la rapidez varían en la dirección opuesta, siendo las memorias más pequeñas cerca de lo alto las más rápidas y costosas. Al examinar las latencias de acceso dadas en la figura 3.1, se observa que las razones de latencias para niveles sucesivos son bastante pequeñas cerca de la cima (diez o menos). Súbitamente, la razón crece en casi 105 entre las memorias principal y secundaria, y también es grande, aunque no tan mal, entre memorias secundaria y terciaria.

Esta brecha de rapidez entre memorias semiconductoras y magnético ópticas es un importante

impedimento para alto rendimiento en aplicaciones de datos intensos.

Los niveles en la jerarquía de memoria de la figura 3.1 están numerados como sigue: 0 para registros, 1 y 2 para caché (conocidos como caché L1 y L2), 3 para memoria principal, etcétera. A veces los registros se excluyen de la jerarquía de memoria, en parte porque los mecanismos para mover datos adentro y afuera de los registros (instrucciones carga/almacenamiento explícitas) son diferentes de los usados para transferencias de datos entre otros niveles. Además, en algunas arquitecturas, ciertos elementos de datos nunca se mueven en los registros y se procesan directamente de la memoria caché.

3.1.3 Semiconductores de acceso aleatorio.

Todos los tipos de memoria interna se implementan utilizando tecnología de semiconductores y tienen el transistor como elemento básico de su construcción. El elemento básico en toda memoria es la celda. Una celda permite almacenar un bit, un valor 0 o 1 definido por una diferencia de potencial eléctrico. La manera de construir una celda de memoria varía según la tecnología utilizada. La memoria interna es una memoria de acceso aleatorio; se puede acceder a cualquier palabra de memoria especificando una dirección de memoria.

Una manera de clasificar la memoria interna según la perdurabilidad es la siguiente:

- **Memoria volátil:** SRAM (static random access memory), RAM (dynamic random access memory)
- **Memoria no volátil:** ROM (read only memory), PROM (programmable read only memory), EPROM (erasable programmable read only memory), EEPROM (electrically erasable programmable read only memory).

3.1.3.1 RAM (FPM, EDO, BEDO, SRAM, DRAM, SDRAM, DDR-RAM).

FPM: Fast Page Mode. Más rápida, tanto por su estructura, como por ser de 70 ó 60 ns. Usada hasta con los primeros Pentium. Es de tipo SIMMs de 30 ó 72 contactos (los de 72 en los Pentium y algunos 486).

EDO: Extended Data Output-RAM. Es una mejora de la FPM. Permite introducir nuevos datos mientras los anteriores están saliendo (haciendo su Output). Permite

mover un bloque completo de memoria a la caché interna del procesador para un acceso más rápido por parte de este. Este estándar se encontraba con refrescos de 70, 60 ó 50 ns. Se instala en SIMMs de 72 contactos.

BEDO: (Burst Extended Data Output). Diseñada para soportar mayores velocidades de BUS. Transfiere datos al procesador en cada ciclo de reloj, pero no de forma continuada, sino a ráfagas (bursts). Reduce los tiempos de espera del procesador para escribir o leer datos de memoria.

SRAM: La memoria estática de acceso aleatorio (SRAM) implementa cada celda de memoria utilizando un flip-flop básico para almacenar un bit de información, y mantiene la información mientras el circuito de memoria recibe alimentación eléctrica.

Para implementar cada celda de memoria son necesarios varios transistores, típicamente seis, por lo que la memoria tiene una capacidad de integración limitada y su coste es elevado en relación con otros tipos de memoria RAM, como la DRAM; sin embargo, es el tipo de memoria RAM más rápido.

DRAM: La memoria dinámica implementa cada celda de memoria utilizando la carga de un condensador. A diferencia de los flip-flops, los condensadores con el tiempo pierden la carga almacenada y necesitan un circuito de refresco para mantener la carga y mantener, por lo tanto, el valor de cada bit almacenado. Eso provoca que tenga un tiempo de acceso mayor que la SRAM.

Cada celda de memoria está formada por solo un transistor y un condensador; por lo tanto, las celdas de memoria son mucho más pequeñas que las celdas de memoria SRAM, lo que garantiza una gran escala de integración y al mismo tiempo permite hacer memorias más grandes en menos espacio.

SDRAM: Sincronic-RAM. Es un tipo síncrono de memoria donde el procesador puede obtener información en cada ciclo de reloj, sin estados de espera. Es de tipo DIMMs de 168 contactos; es la opción para ordenadores nuevos.

RDRAM: (Direct Rambus DRAM). Memoria de 64 bits que puede producir ráfagas de 2ns y puede alcanzar tasas de transferencia de 533 MHz, con picos de 1,6 GB/s. Se integra muy bien en las tarjetas gráficas AGP. Evita los cuellos de botella en la

transferencia entre la tarjeta gráfica y la memoria de sistema durante el acceso directo a memoria (DIME) para el almacenamiento de texturas gráficas.

DDR-SRAM: (Double Data Rate SDRAM o SDRAM-II). Funciona a velocidades de 83, 100 y 125MHz o el doble de estas en el proceso de transferencia de datos a memoria. Es una extensión de la memoria SDRAM, facilitando la implementación por la mayoría de los fabricantes.

3.1.3.2 ROM (PROM, EPROM, EEPROM, FLASH).

Tal como indica su nombre, se trata de memorias de solo lectura que no permiten operaciones de escritura y, por lo tanto, la información que contienen no se puede borrar ni modificar. Este tipo de memorias se pueden utilizar para almacenar los microprogramas en una unidad de control microprogramada; también se pueden utilizar en dispositivos que necesitan trabajar siempre con la misma información. La grabación de la información en este tipo de memorias forma parte del proceso de fabricación del chip de memoria. Estos procesos implican la fabricación de un gran volumen de memorias ROM con la misma información; es un proceso costoso y no es rentable para un número reducido de unidades.

Memoria programable de solo lectura o PROM (programmable read only memory): Cuando hay que fabricar un número reducido de memorias ROM con la misma información grabada, se recurre a otro tipo de memorias ROM: las memorias ROM programables (PROM).

A diferencia de las anteriores, la grabación no forma parte del proceso de fabricación de los chips de memoria, sino que se efectúa posteriormente con un proceso eléctrico utilizando un hardware especializado para la grabación de memorias de este tipo. Como el proceso de programación no forma parte del proceso de fabricación, el usuario final de este tipo de memorias puede grabar el contenido según sus necesidades. Cabe destacar que, tal como sucede con las memorias ROM, el proceso de grabación o programación solo se puede realizar una vez. Este tipo de memoria tiene unas aplicaciones parecidas a las de las memorias ROM.

Memoria reprogramable mayoritariamente de lectura. Esta puede ser de tres tipos:

- **EPROM (erasable programmable read only memory).** Se trata de memorias en las que habitualmente se hacen operaciones de lectura, pero cuyo contenido puede ser borrado y grabado de nuevo. Hay que destacar que el proceso de borrar es un proceso que borra completamente todo el contenido de la memoria; no se puede borrar solo una parte. Para borrar, se aplica luz ultravioleta sobre el chip de memoria EPROM; para permitir este proceso, el chip dispone de una pequeña ventana sobre la cual se aplica la luz ultravioleta. La grabación de la memoria se hace mediante un proceso eléctrico utilizando un hardware específico. Tanto para el proceso de borrar como para el proceso de grabar hay que sacar el chip de memoria de su localización de uso habitual, ya que la realización de estas dos tareas implica la utilización de hardware específico.
- **EEPROM (electrically erasable programmable read only memory).** Tiene un funcionamiento parecido a la EPROM, permite borrar el contenido y grabar información nueva; sin embargo, a diferencia de las memorias EPROM, todas las operaciones son realizadas eléctricamente. Para grabar datos no hay que borrarlos previamente; se permite modificar directamente solo uno o varios bytes sin modificar el resto de la información. Son memorias mayoritariamente de lectura, ya que el proceso de escritura es considerablemente más lento que el proceso de lectura.

Memoria flash: La memoria flash es un tipo de memoria parecida a las memorias EEPROM, en las que el borrado es eléctrico, con la ventaja de que el proceso de borrar y grabar es muy rápido. La velocidad de lectura es superior a la velocidad de escritura, pero las dos son del mismo orden de magnitud. Este tipo de memoria no permite borrar la información byte a byte, sino que se deben borrar bloques de datos enteros; eso lleva a que todo el contenido de la memoria se pueda borrar en pocos segundos, pero también modera el proceso de escritura, ya que para escribir un dato nuevo hay que borrar previamente todo el bloque. Tiene una capacidad de integración muy elevada y por este motivo también se utiliza para dispositivos de almacenamiento externo.

3.1.3.3 Organización.

El elemento básico de una memoria semiconductora es la celda de memoria. Aunque se utilizan diversas tecnologías electrónicas; todas las celdas de memoria semiconductoras comparten ciertas propiedades:

- Presentan dos estados estables (o semi estables), que pueden emplearse para representar el 1 y el 0 en binarios.
- Pueden escribirse en ellas (al menos una vez) para fijar su estado.
- Pueden leerse para detectar su estado.

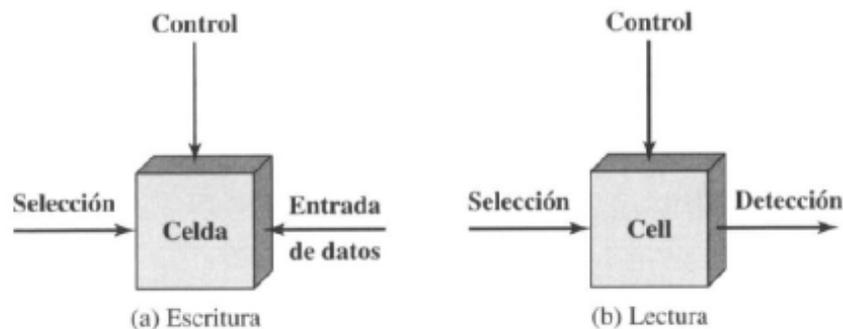


Figura 3.2: Funcionamiento de una celda de memoria.

La figura 3.2 describe el funcionamiento de una celda de memoria. Lo más común es que la celda tenga tres terminales para transportar señales eléctricas. El terminal de selección, como se indica en su nombre, selecciona la celda para que pueda realizarse una operación de escritura o lectura. El terminal de control indica si se trata de una lectura o escritura. Para la escritura, el tercer terminal proporciona la señal que fija el estado de la celda a uno o cero. En una lectura, el tercer terminal se utiliza como salida del estado de la celda.

Básicamente, los elementos que hemos de tener en cuenta son los siguientes:

- **Palabra de memoria.** Es la unidad de organización de la memoria desde el punto de vista del procesador; el tamaño de la palabra de memoria se especifica en bytes o bits. Es el número de bytes máximo que se pueden leer o escribir en un solo ciclo de acceso a la memoria.

Ejemplo Memoria de 2Kbytes con una palabra de memoria de 2 bytes. Por lo tanto, necesitaremos 10 bits para poder hacer referencia a las 1.024 (2¹⁰) posiciones de memoria que almacenarán 2 bytes (16 bits) cada una.

- **Unidad de direccionamiento.** La memoria interna se puede ver como un vector de elementos, una colección de datos contiguos, en la que cada dato es accesible indicando su posición o dirección dentro del vector.

La unidad de direccionamiento especifica cuál es el tamaño de cada elemento de este vector; habitualmente a la memoria se accede como un vector de bytes (cada byte tendrá su dirección), aunque puede haber sistemas que accedan a la memoria como un vector de palabras, en los que cada dirección corresponda a una palabra. El número de bits utilizados para especificar una dirección de memoria fija el límite máximo de elementos dirigibles, el tamaño del mapa de memoria; si tenemos n bits para las direcciones de memoria, el número máximo de elementos dirigibles será de 2ⁿ.

- **Ordenación de los bytes en memoria.** Aunque normalmente la unidad de direccionamiento de la memoria es el byte, es habitual que se puedan hacer accesos a memoria en múltiplos de byte, hasta el tamaño de la palabra (2, 4, e incluso 8 bytes). En este caso, solo se indica la dirección del primer byte de la palabra y se utilizan dos métodos a la hora de acceder a la palabra:

Big-endian: la dirección especificada corresponde al byte de más peso de la palabra. **Little-endian:** la dirección especificada corresponde al byte de menos peso de la palabra.

3.1.3.4 Lógica del chip.

Como otros circuitos integrados, las memorias semiconductoras vienen en chips encapsulados. Cada chip contiene una matriz de celdas de memoria.

En toda jerarquía de memoria existen compromisos de velocidad, capacidad y coste. Estos compromisos existen también cuando consideramos la organización de celdas de memoria y del resto de funciones lógicas de un chip de memoria. Para las memorias semiconductoras, uno de los aspectos fundamentales de diseño es el

número de bits de datos que pueden ser leídos/escritos a la vez. En un extremo está la estructura en la disposición física de las celdas de la matriz es la misma que la disposición lógica (tal como la percibe el procesador) de las palabras de memoria. La matriz está organizada en W palabras de B bits cada una. Por ejemplo, un chip de 16 Mb podría estar estructurado en 1 M palabras de 16 bits. En el otro extremo está la estructura denominada un bit-por-chip, en la que los datos se escriben/leen por bits

3.1.4 Memoria Caché.

Las memorias caché son memorias de capacidad reducida, pero más rápidas que la memoria principal, que utilizan un método de acceso asociativo. Se pueden encontrar dentro del chip del procesador o cerca de él y están diseñadas para reducir el tiempo de acceso a la memoria. En la memoria caché se almacenan los datos que se prevé que se utilizarán más habitualmente, de manera que sea posible reducir el número de accesos que debe hacer el procesador a la memoria principal (ya que el tiempo de acceso a la memoria principal siempre es superior al tiempo de acceso a la memoria caché). No es accesible por parte del programador, es gestionada por el hardware y el sistema operativo y se implementa utilizando tecnología SRAM. Los procesadores modernos utilizan diferentes niveles de memoria caché, lo que se conoce como memoria caché de primer nivel, segundo nivel, etc. Actualmente es habitual disponer de hasta tres niveles de memoria caché, referidos como L1, L2 y L3. Cada vez es más frecuente que algunos de estos niveles se implementen dentro del chip del procesador y que el nivel más próximo al procesador esté dividido en dos partes: una dedicada a las instrucciones y otra dedicada a los datos.

La memoria caché se sitúa entre la memoria principal y el procesador, puede estar formada por uno o varios niveles. En este apartado explicaremos el funcionamiento de la memoria caché considerando un único nivel, pero el funcionamiento es parecido si tiene varios. La memoria caché tiene un tiempo de acceso inferior al de la memoria principal con el objetivo de reducir el tiempo de acceso medio a los datos, pero también tiene un tamaño mucho más reducido que la memoria principal.

Si un dato está en la memoria caché, es posible proporcionarlo al procesador sin acceder a la memoria principal, si no, primero se lleva el dato de la memoria principal a la memoria caché y después se proporciona el dato al procesador. Si, en la mayoría de los accesos a memoria, el dato está en la memoria caché, el tiempo de acceso medio será próximo al tiempo de acceso a la memoria caché. Eso es factible gracias a la característica de proximidad referencial de los programas.

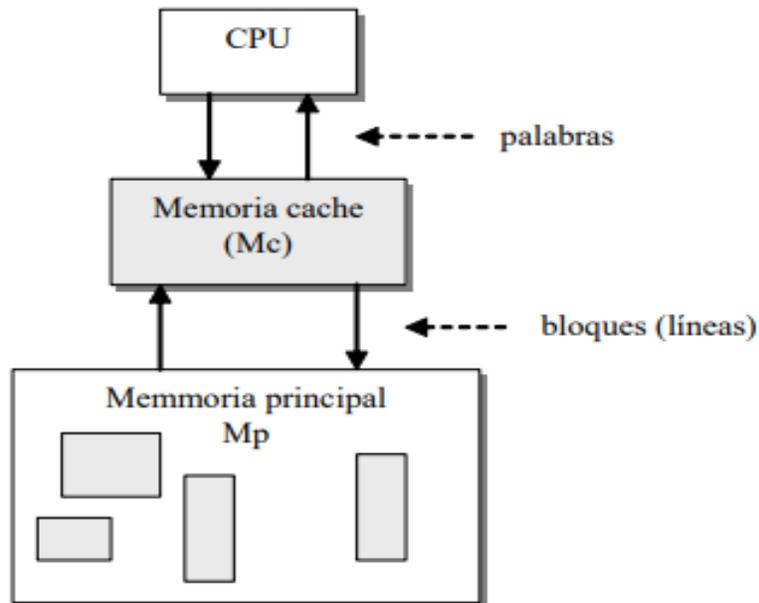


Figura 3.3: Organización de memoria caché

Para trabajar con memoria caché, la memoria principal se organiza en bloques de palabras, de manera que cuando hay que trasladar datos de la memoria principal a la memoria caché se lleva un bloque entero de palabras de memoria (figura 3.3), no se trabaja con palabras individuales. La memoria caché también se organiza en bloques que se denominan líneas. Cada línea está formada por un conjunto de palabras (el mismo número de palabras que tenga un bloque de memoria principal), más una etiqueta compuesta por unos cuantos bits. El contenido de la etiqueta permitirá saber qué bloque de la memoria principal se encuentra en cada línea de la memoria caché en un momento dado.

3.1.4.1 Principios.

El objetivo de la memoria caché es lograr que la velocidad de la memoria sea lo más rápida posible, consiguiendo al mismo tiempo tamaño grande al precio de memorias semiconductoras menos costosas.

La memoria caché es una memoria pequeña y rápida que se interpone entre la CPU y la memoria principal para que el conjunto opere a mayor velocidad. Para ello es necesario mantener en la caché aquellas zonas de la memoria principal con mayor probabilidad de ser referenciadas. Esto es posible gracias a la propiedad de localidad de referencia de los programas: *localidad de referencia: temporal y espacial*. Los programas manifiestan una propiedad que se explota en el diseño del sistema de gestión de memoria de los computadores en general y de la memoria caché en particular, la localidad de referencias: los programas tienden a reutilizar los datos e instrucciones que utilizaron recientemente. Una regla empírica que se suele cumplir en la mayoría de los programas revela que gastan el 90% de su tiempo de ejecución sobre sólo el 10% de su código. Una consecuencia de la localidad de referencia es que se puede predecir con razonable precisión las instrucciones y datos que el programa utilizará en el futuro cercano a partir del conocimiento de los accesos a memoria realizados en el pasado reciente. La localidad de referencia se manifiesta en una doble dimensión: temporal y espacial.

- **Localidad temporal:** las palabras de memoria accedidas recientemente tienen una alta probabilidad de volver a ser accedidas en el futuro cercano. La localidad temporal de los programas viene motivada principalmente por la existencia de bucles.
- **Localidad espacial:** las palabras próximas en el espacio de memoria a las recientemente referenciadas tienen una alta probabilidad de ser también referenciadas en el futuro cercano. Es decir, que las palabras próximas en memoria tienden a ser referenciadas juntas en el tiempo. La localidad espacial viene motivada fundamentalmente por la linealidad de los programas (secuenciamiento lineal de las instrucciones) y el acceso a las estructuras de datos regulares.

Para implementar el mecanismo de actualización de la caché con los datos con mayor probabilidad de ser referenciados se divide la memoria principal en bloques de un número de bytes (4,8,16 etc.) y la caché en marcos de bloque o líneas de igual tamaño. El bloque será, pues, la unidad de intercambio de información entre la memoria principal y la caché, mientras que entre la caché y la CPU sigue siendo la palabra. El directorio contiene la información de qué bloques de Mp se encuentran ubicados en Mc.

El funcionamiento de la memoria caché se puede resumir en el diagrama de flujo de la siguiente figura 3.4. En él se describe el proceso de traducción de la dirección física procedente de la CPU (en el supuesto que el procesador no disponga de memoria virtual o esté desactivado) en el dato ubicado en la posición de memoria determinada por dicha dirección:

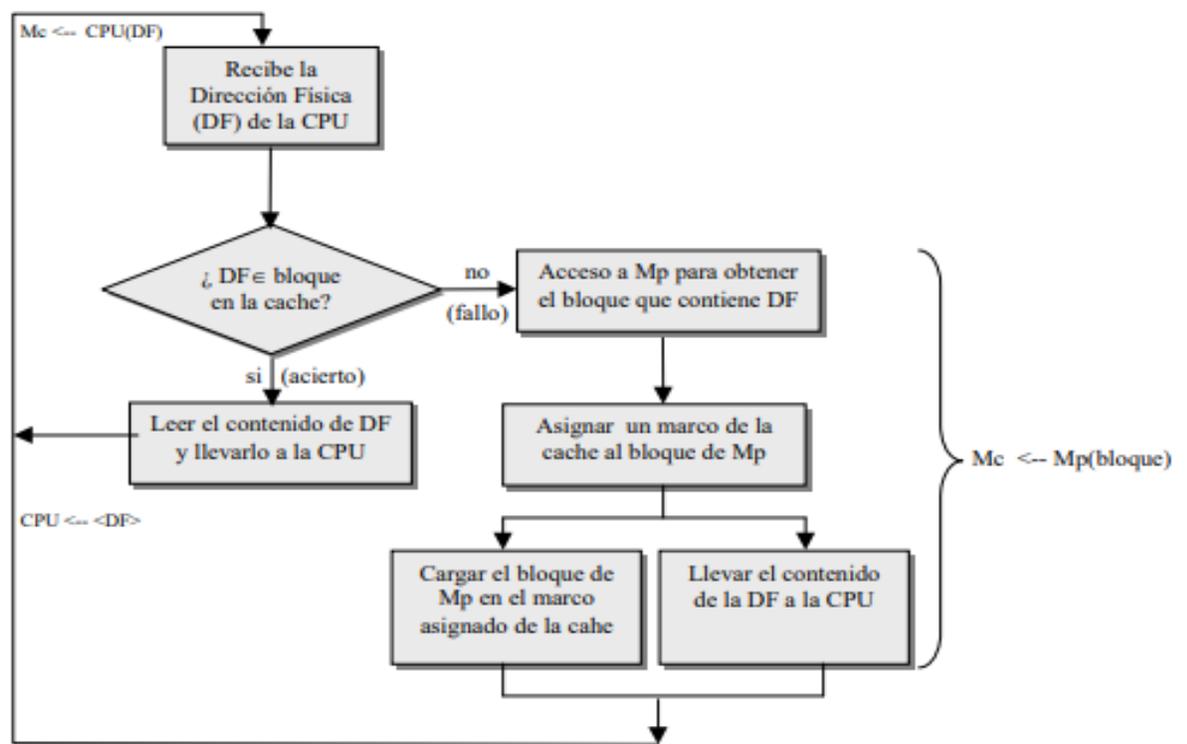


Figura 3.4 Ejemplo de funcionamiento de una memoria caché

Cada vez que el procesador quiere acceder a una palabra de memoria, primero se accede a la memoria caché; si la palabra de memoria se encuentra almacenada en la memoria caché, se proporciona al procesador y diremos que se ha producido un

acierto. En caso contrario, se lleva el bloque de datos de la memoria principal que contiene la palabra de memoria hacia la memoria caché y, cuando la palabra ya está en la memoria caché, se proporciona al procesador; en este caso diremos que se ha producido un fallo. Cuando hay un fallo, el hardware de la memoria caché debe realizar la secuencia de tareas siguiente:

1. Solicitar a la memoria principal el bloque en el que está el dato que ha producido el fallo.
2. Llevar el bloque de datos solicitado a la memoria caché. Las operaciones realizadas en esta tarea dependerán de las políticas de asignación y algoritmos de reemplazo que veremos más adelante.
3. El procesador obtiene el dato de la memoria caché como si se hubiera producido un acierto.

Un acceso con fallo en la memoria caché puede ser bastante más costoso en tiempo que un acceso con acierto, por lo que es muy importante tener un número reducido de fallos.

3.1.4.2 Elementos de diseño.

Una memoria caché se caracteriza por diversos parámetros de diseño que influyen su costo de implementación y rendimiento (tasa de impacto). La siguiente descripción se refiere a la suposición de un solo nivel de caché; esto es, no hay caché de nivel 2. Los más importantes parámetros de caché son:

- **Tamaño de caché en bytes o palabras.** Una caché más grande puede retener más datos útiles del programa, pero es más costosa y quizá más lenta.
- **Tamaño de bloque o ancho de línea de caché,** que se define como la unidad de transferencia de datos entre el caché y la memoria principal. Con una línea de caché más grande, se llevan más datos a la caché con cada fallo. Lo anterior puede mejorar la tasa de impacto, pero también tiende a ligar partes de la caché con datos de menor utilidad.
- **Política de colocación.** Determinar dónde se puede almacenar una línea de caché entrante. Políticas más flexibles implican mayor costo de hardware y

puede o no tener beneficios de rendimiento como consecuencia de sus procesos más complejos y lentos, para alojar los datos requeridos en la caché.

- **Política de sustitución.** Determinar cuál de los muchos bloques de caché (en los que se puede mapear una nueva línea de caché) se debe sobrescribir. Las políticas típicas incluyen la elección de un bloque aleatorio y del bloque menos utilizado.
- **Política de grabación.** Determinar si las actualizaciones a palabras de caché se adelantan inmediatamente a la memoria principal (política de *grabación directa*) o bloques de caché modificados se copian a la memoria principal en su totalidad y cuándo deben sustituirse en la caché (política de *escritura inversa* o *copy-back*).
- **Política de búsqueda de bloques:** determina la causa que desencadena la llevada de un bloque a la caché (normalmente un fallo en la referencia)
- **Cachés independientes para datos e instrucciones:** frente a cachés unificadas.

Estos parámetros están cercanamente relacionados; el hecho de que se cambie uno con frecuencia significa que los otros también necesitan cambios para asegurar un rendimiento óptimo de memoria.

3.1.4.3 Correspondencia.

La función de correspondencia determina las posibles líneas de la caché (marcos de bloque) en las que se puede ubicar un determinado bloque de la memoria principal que ha sido referenciado por el programa y hay que llevarlo a memoria caché.

Existen tres funciones de correspondencia para definir la posible ubicación de un bloque de memoria principal (Mp) en la memoria caché (Mc): directa, asociativa y asociativa por conjuntos.

En el primer caso (función de correspondencia directa), un bloque de Mp sólo puede ubicarse en una línea de la caché, aquella que coincide con el bloque cuando superponemos Mc sobre Mp respetando fronteras de Mc, es decir, sobre espacios

de M_p que son múltiplos del tamaño de M_c . En la correspondencia directa el bloque B_j de M_p se puede ubicar sólo en el marco de bloque o línea M_{Bi} que cumple la siguiente relación $i = j \text{ mod } m$, donde m es el número total de líneas que tiene la caché.

En la correspondencia asociativa un bloque puede ubicarse en cualquier línea de M_c .

Permite que cada bloque de memoria principal pueda cargarse en cualquier línea de la caché. En este caso, la lógica de control de la caché interpreta una dirección de memoria simplemente como una etiqueta y un campo de palabra. El campo de etiqueta identifica unívocamente un bloque de memoria principal. Para determinar si un bloque está en la caché, su lógica de control debe examinar simultáneamente todas las etiquetas de líneas para buscar coincidencias.

Finalmente, la correspondencia asociativa por conjuntos es un compromiso entre las dos anteriores. En esta se agrupan en $v=2d$ conjuntos con k líneas/conjunto o vías cada uno. Se cumple que el número total de marcos de bloque (líneas) que tiene la caché $m = v*k$. Un bloque B_j de M_p se puede ubicar sólo en el conjunto C_i de M_c que cumple la siguiente relación $i = j \text{ mod } v$.

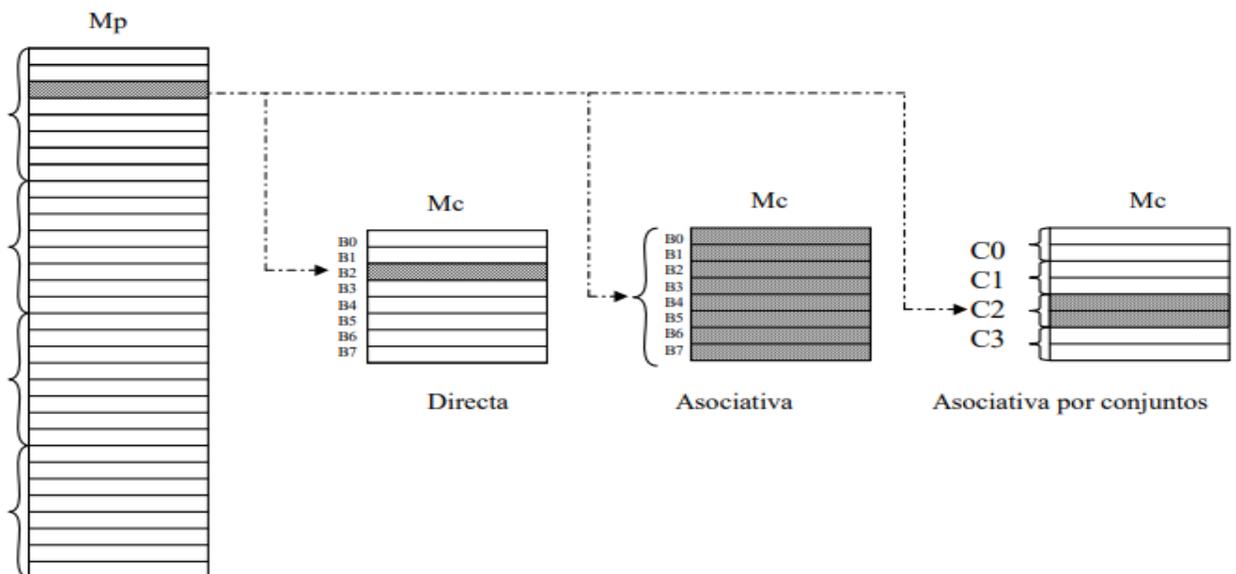


Figura 3.5: Tipos de correspondencia en memoria caché

3.1.4.4 Política de escritura.

Cuando accedemos a la memoria caché, podemos hacer lecturas o escrituras; hasta ahora hemos visto la problemática de acceder a la memoria caché para leer un dato. Cuando se debe realizar una operación de escritura, aparecen nuevos problemas porque los datos que tenemos en la memoria caché son una copia de los datos que tenemos en la memoria principal y hay que garantizar la coherencia de los datos. Analizaremos el caso con un único procesador y nivel de memoria caché entre el procesador y la memoria principal. Si tenemos más de un procesador con una memoria caché local para cada procesador, la modificación de un dato en una de estas memorias caché invalida el valor del dato en la memoria principal, pero también invalida el valor del dato si se encuentra en otra memoria caché. De manera parecida, si tenemos otros dispositivos que puedan modificar directamente un dato de la memoria principal, el valor de este dato queda invalidado en las memorias caché donde se pueda encontrar. A continuación, se describen diferentes políticas para gestionar las escrituras y mantener la coherencia entre los datos de la memoria caché y la memoria principal:

1. **Escritura inmediata (write trough):** cuando se escribe en la memoria caché, también se escribe en la memoria principal transfiriendo todo el bloque que contiene el dato modificado; de esta manera en todo momento la copia que tenemos en la caché es idéntica a la que tenemos en la memoria principal. La política de escritura inmediata es la más fácil de implementar, pero su inconveniente es que produce un gran flujo de información entre la memoria caché y la memoria principal.
2. **Escritura aplazada (write back):** las escrituras se efectúan solo sobre la memoria caché. La memoria principal se actualiza cuando se elimina una línea de la memoria caché que ha sido modificada. Eso implica añadir algunos bits a cada línea de la memoria caché para saber si la línea se ha modificado o no. Si hay que reemplazar una línea que ha sido modificada, primero es necesario copiar la línea modificada a la memoria principal y a continuación llevar el nuevo bloque, lo que aumenta significativamente el tiempo para acceder al dato.

3.1.4.5 Tamaño de línea.

Hemos visto que la memoria caché se organiza en líneas; una línea está formada básicamente por un conjunto de palabras más una etiqueta que identifica qué bloque de la memoria principal ocupa aquella línea de la memoria caché. La línea de memoria caché es la unidad de transferencia entre la memoria caché y la memoria principal. El tamaño de la línea es uno de los parámetros fundamentales del diseño de la memoria caché. Hay que decidir cuántas palabras se almacenarán en una línea de memoria caché, es decir, cuál es el tamaño de una línea.

Como ya sabemos, los datos se trasladan de la memoria principal a la memoria caché cuando hay un fallo. Si se produce un fallo, se lleva a la memoria caché el dato que lo ha provocado y el resto de los datos del bloque de memoria donde se encuentra este dato. De esta manera, se espera que los accesos siguientes sean aciertos en la memoria caché. El tamaño de la línea es de unos cuantos bytes de información (un tamaño habitual está entre los 32 bytes y 128 bytes). Aumentar el tamaño de la línea permite aprovechar la localidad espacial, pero hasta cierto punto. Cuando se produce un fallo, el tiempo necesario para trasladar una línea más grande aumenta; además, disminuye el número de líneas disponibles de la memoria caché (el tamaño de la memoria caché es fijo) y tendremos más competencia para conseguir un bloque, lo que hará que se saquen de la caché líneas que todavía no se han utilizado en su totalidad y se reducirá el efecto de la localidad espacial, y todo ello puede representar un aumento en la tasa de fallos.

3.1.4.6 Número de cachés.

En las computadoras actuales es común la utilización de varios niveles de caché. La denominada caché de nivel 1 es de tamaño reducido y se ubica funcionalmente, interceptando

los datos que entran en el microprocesador desde la memoria o los que salen ya procesados

hacia la memoria.

La caché de nivel 2 es de tamaño mayor y se ubica (si se incluyó en el sistema) entre la caché de primer nivel y la memoria principal. Su carácter optativo nos obliga

a considerar cuales son los beneficios de su incorporación. En primer lugar, podemos indicar que atiende pedidos menos frecuentes que la primaria, ya que normalmente el éxito en el acceso se logrará accediendo al primer nivel; esto significa que se utiliza como ampliación de la anterior. Dada su mayor capacidad de almacenamiento, no solo contendrá la misma información almacenada en la primaria sino unos cuantos Kbytes más.

Las ventajas de contar con varios niveles se empiezan a notar cuando las aplicaciones son muy grandes o requieren acceso frecuente a datos en la memoria principal. En la medida en que la eficiencia de un nivel secundario, esto es, en la medida en que se acceda a él con mayor frecuencia por no haber encontrado la información en la primaria, se reducirá el número de ciclos de bus utilizados en el acceso a RAM, que resulta en un mejor rendimiento global del sistema.

Por otra parte, si la computadora es un servidor en una red de computadoras, el acceso a su memoria principal es elevado, así como el acceso a sus dispositivos de E/S; esto produce un alto tráfico en el bus, que se beneficia con la implementación de estos niveles de memoria.

En el caso de sistemas multiprocesadores su utilización se justifica en gran medida, ya que en estos sistemas muchos comparten la misma memoria, por lo tanto, el acceso frecuente al bus puede llegar a condiciones de saturación.

3.2 Memoria Externa.

La memoria externa corresponde a dispositivos de almacenamiento secundario: discos magnéticos, cintas magnéticas, discos ópticos, dispositivos de memoria flash, etc., y también se pueden considerar sistemas de almacenamiento en red. Estos dispositivos son gestionados por el sistema de ficheros del sistema operativo mediante el sistema de entrada/salida. Los dispositivos que forman la memoria externa se conectan al computador con algún tipo de bus (serie o paralelo). Estos dispositivos se pueden encontrar físicamente dentro del computador conectados por buses internos del computador (IDE, SATA, SCSI, etc.) o pueden estar fuera del computador conectados por buses externos (USB, Firewire, eSATA, Infiniband, etc.).

3.2.1 Discos magnéticos.

El almacenamiento permanente de datos en una computadora opera de acuerdo con principios ópticos, magnéticos o una combinación de ambos.

Pardo (1993) señala que la idea del disco magnético surgió del tambor (cilindro en cuya superficie y en pistas paralelas se grababa la información), para aumentar, en algunos casos, su corta capacidad de almacenamiento y, en otros, por el deseo de reducir su masa.

En un principio, los discos fueron poco aceptados por ser un medio caro que presentaba un elevado índice de averías a causa de las deficiencias mecánicas que tenía el posicionamiento de las cabezas de lectura y grabación sobre la superficie de los platos.

Patterson y Hennessy (2000) mencionan que existen básicamente dos tipos de discos magnéticos, los flexibles y los duros. Ambos tipos se basan en un plato rotatorio cubierto por una superficie magnética, y usan una cabeza de lectura/escritura móvil para acceder al disco. El almacenamiento en los discos es no volátil, es decir, la información permanece aun cuando se desconecta la alimentación. Gracias a que los platos de los discos duros son de metal, estos tienen ventajas significativas sobre los discos flexibles:

- Un disco duro puede ser más grande porque es más rígido.
- El disco duro puede tener una mayor densidad porque puede controlarse con mayor precisión.
- El disco duro tiene una mayor velocidad de acceso a los datos porque puede rotar más rápido.
- Los discos duros pueden incorporar más de un plato.

3.2.1.1 Definición.

Un disco magnético es un plato circular construido con un material no magnético, llamado sustrato, cubierto por un material magnetizable. Tradicionalmente el sustrato es aluminio o una aleación de aluminio. También se usa sustratos de cristal.

En el caso del almacenamiento magnético, se almacena un flujo de bits de datos binarios (unos y ceros) magnetizando diminutas partículas de metal incrustadas en la superficie de un disco o cinta, en un patrón que representa los datos. Posteriormente, ese patrón magnético puede ser leído y convertido nuevamente en un flujo de bits exactamente igual al original.

La figura 3.6: muestra las partes en que está compuesto un disco duro.

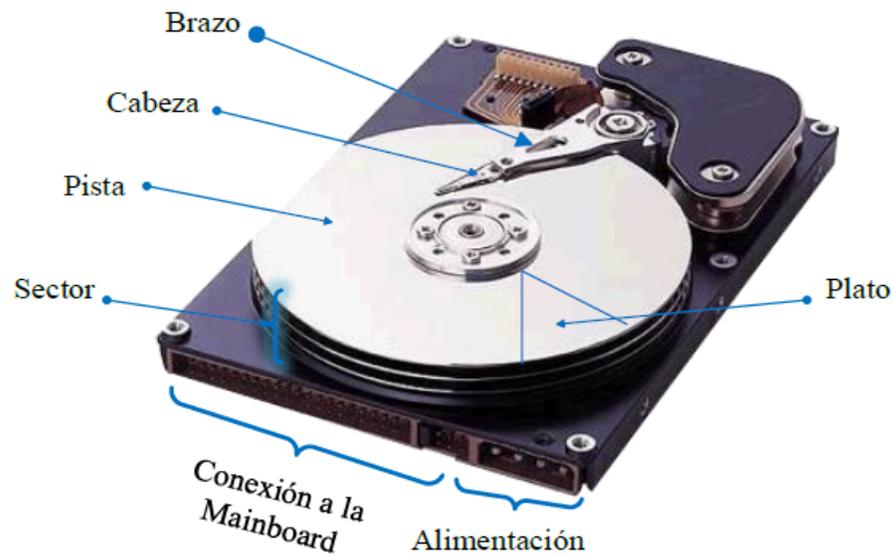


Figura 3.6: Partes de un disco duro

3.2.1.2 Organización y formato.

La cabeza es un dispositivo relativamente pequeño, capaz de leer o escribir en una zona del plato que rota bajo ella. Esto da lugar a que los datos se organicen en un conjunto de anillos concéntricos en el plato, llamados pistas. Cada pista es del mismo ancho que la cabeza. Usualmente hay cientos de pistas por superficie.

En la figura 3.7 se puede ver la disposición de los datos. Las pistas están separadas por bandas vacías. Esto previene, o por lo menos minimiza, los errores debidos a desalineamientos de la cabeza o simplemente a interferencias del campo magnético.

Los datos se transfieren al y desde el disco en sectores. Normalmente hay cientos de sectores por pista, y estos pueden tener una longitud variable o fija. En la mayoría de los sistemas de hoy se utilizan sectores de longitud fija, siendo 512 bytes el tamaño casi universal de un sector. Para evitar imposiciones de precisión ilógicas del sistema, los sectores adyacentes se separan con intrapistas vacías.

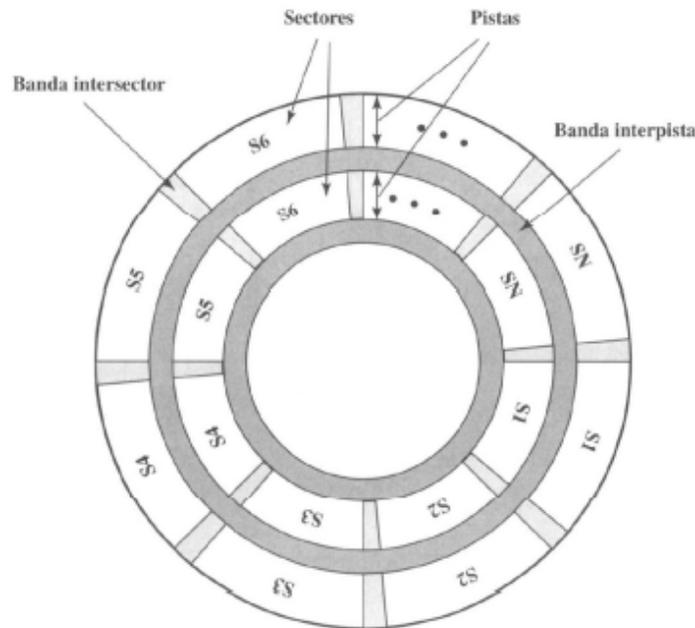


Figura 3.7: Organización de los datos en el disco

Un bit cercano al centro de un disco girando, pasa por punto fijo (como la cabeza de lectura-escritura) más despacio que un bit más externo. Por tanto, debe haber alguna forma de compensar la variación de la velocidad de forma que la cabeza pueda leer todos bits a la misma velocidad. Esto se puede hacer incrementando el espacio entre bits de la información grabada en los segmentos del disco. La información se puede escanear a la misma velocidad rotando el disco a una velocidad fija, conocida como velocidad angular constante (constant angular velocity, CAV). La figura muestra la estructura de un disco que usa CAV. El disco se divide en una serie de sectores en forma de trozo de tarta y en una serie de pistas concéntricas. La ventaja de usar CAV es que los bloques individuales de datos se

pueden direccionar directamente con la pista y sector. Para mover la cabeza desde su actual posición a una dirección específica, sólo hay que mover ligeramente la cabeza a la pista específica y esperar a que el sector se sitúe bajo la cabeza. La desventaja de CAV es que la cantidad de datos que se puede almacenar en las pistas más extensas es solo la misma que la de las pistas internas.

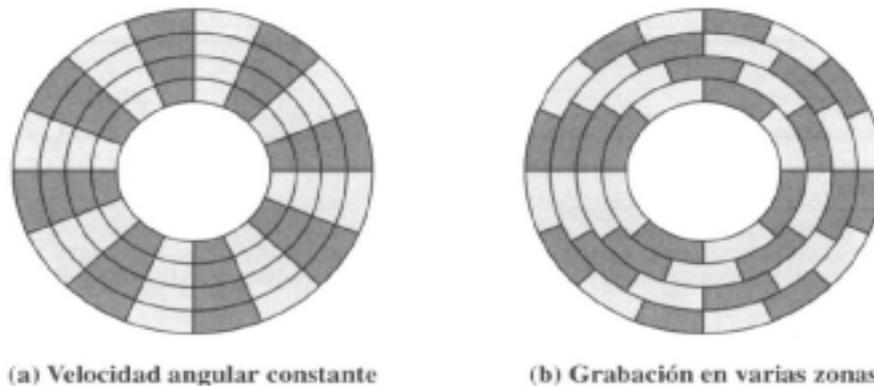


Figura 3.8: Comparación de los métodos de organización en un disco

Debido a que la densidad, en bits por pulgada, aumenta a medida que nos movemos desde la pista más externa a la más interna, la capacidad de almacenamiento de un disco con un sistema CAV sencillo viene limitada por la máxima densidad de grabación que se puede llevar a cabo en la pista más interna. Para aumentar la capacidad, los discos duros modernos utilizan una técnica conocida como grabación en varias zonas (multiple zone recording), en la que la superficie se divide en varias zonas concéntricas (usualmente 16). Dentro de una zona, el número de bits por pista es constante. Las zonas más lejanas del centro contienen más bits (más sectores) que las zonas próximas al centro. Esto permite capacidades de almacenamiento mayores a expensas de una circuitería de alguna forma más compleja. Como la cabeza del disco se mueve de una zona a otra, la longitud (a lo largo de la pista) de los bits individuales cambia, provocando un cambio en el tiempo de lectura y escritura. La figura 3.8 sugiere la naturaleza de la grabación en varias zonas; en esta figura, cada zona es una sola pista.

Algún procedimiento es necesario para situar las posiciones del sector en una pista. Claramente, debe haber algún punto de comienzo de la pista y una manera de

identificar el principio y el fin de cada sector. Estos requisitos son gestionados mediante datos de control grabados en el disco. Por tanto, el disco se graba con un formato que contiene algunos datos extra usados solo por el controlador del disco y no accesibles al usuario

3.2.1.3 Características.

Existen características que diferencian los distintos tipos de discos. Primero, las cabezas pueden ser fijas o móviles con respecto a la dirección radial del plato. En un disco de cabeza fija hay una cabeza de lectura/escritura por pista, Todas las cabezas se montan en un brazo rígido que se extiende a través de todas las pistas. En un disco de cabeza móvil, hay solo una cabeza de lectura/escritura. Como antes, la cabeza se monta en un brazo. Como la cabeza debe poder posicionarse encima de cualquier pista, el brazo debe extenderse o retraerse para este propósito.

El disco mismo, se monta en una unidad de disco, que consta del brazo, un eje que rota el disco, y la electrónica necesaria para la entrada y salida de datos binarios. Un disco no extraíble está permanentemente montado en la unidad de disco. Un disco extraíble, puede ser quitado y sustituido por otro disco. La ventaja de este último tipo es que es posible una cantidad de datos ilimitada con un número limitado de unidades de disco. Además, un disco puede ser utilizado en diversos computadores.

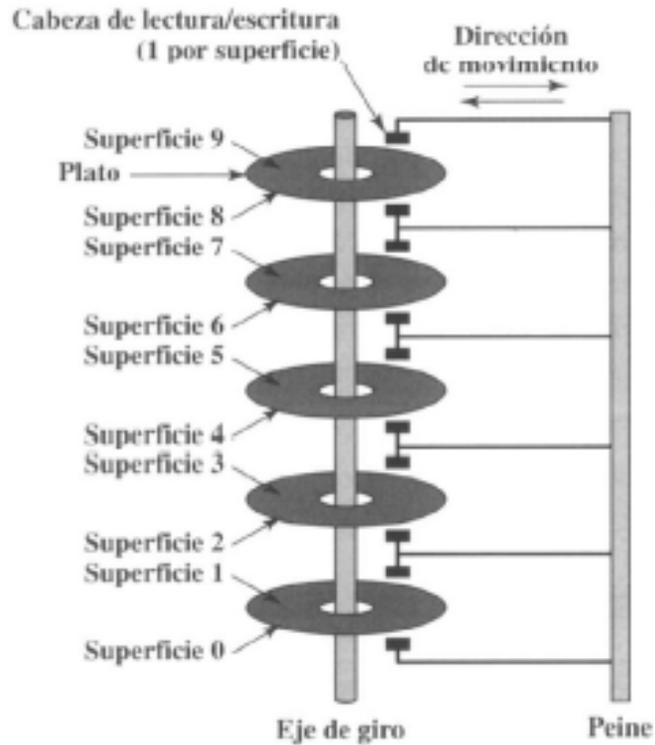


Figura 3.9: Componentes de una unidad de disco

En la mayoría de los discos, la cubierta magnetizable se aplica a ambas caras del plano, denominándose estos discos de doble superficie.

Algunas unidades de disco poseen varios platos apilados verticalmente y separados por una distancia de alrededor de una pulgada. Disponen de varios brazos. Los discos de varios platos utilizan una cabeza que se mueve, con una cabeza de lectura-escritura para cada superficie del plato. El conjunto de todas las pistas que tiene la misma posición relativa en el plato se denomina cilindro.

Las características físicas de los discos se pueden resumir de la siguiente manera:

- Movimiento de la cabeza: cabeza fija (una por pista), cabeza móvil (una por superficie).
- Portabilidad de los discos: discos no extraíbles, disco extraíble.
- Caras: una cara, dos caras.
- Platos: un plato, varios platos
- Mecanismo de la cabeza: contacto (disquete), separación fija, separación aerodinámica.

3.2.1.4 Parámetros para medir prestaciones.

Tipo de Disco

- Indica la tecnología y estructura física del mismo.

Capacidad

- Indica el contenido en octetos (bytes) que es capaz de almacenar el disco.

Tamaño

- Indica el diámetro del plato donde se encuentran las superficies magnetizables.

Tiempo medio de acceso

- Es el tiempo medio en que la cabeza lectora/grabadora tarda en acceder a un sector determinado.

Velocidad de transferencia

- Indica el número de bytes que se transfieren por unidad de tiempo entre el ordenador central y el dispositivo o viceversa.

Velocidad de rotación

- Es el número de revoluciones por minuto a que gira el plato que contiene la superficie magnetizable.

Número de superficies

- Es el número de superficies gravables.

Número de cabezas

- Es el número de cabezas lectoras/grabadoras de que consta la unidad.

Número de pistas

- Es el número de circunferencias gravables. Se suele indicar el número de pistas por superficie.

Número de sectores por pistas

- Es el número de bloques o registros físicos que hay en una pista.

Número de palabras por sector

- Es el número de palabras que pueden grabarse en un sector.

Bits por palabras

- Indica el número de bits que utilizan las palabras grabadas.

Densidad máxima

- Indica la densidad máxima de grabación en las pistas, es decir, la densidad de grabación en la pista más interior.

3.2.1.5 Conjunto redundante de discos independientes (RAID).

RAID (Redundant Array of Independent Disks o Matriz redundante de discos independientes) es una tecnología sencilla que mejora el rendimiento de las soluciones de almacenamiento externo. RAID le permite elegir la mejor forma de utilizar su dispositivo para que se ajuste a sus necesidades. En pocas palabras, la tecnología RAID divide o duplica la tarea de un disco entre varios (al menos dos) discos, tanto para mejorar el rendimiento como para duplicar los datos en un supuesto fallo de la unidad. Puede decidir cómo gestiona los datos el dispositivo configurando el modo RAID.

Niveles de RAID

RAID 0, es el modo RAID más rápido. Se necesitan al menos 2 unidades, RAID 0 distribuye los datos en cada disco. Las capacidades disponibles de cada disco se añaden juntas, de modo que se monta un solo volumen en el ordenador. Si falla una unidad física en la matriz, los datos de todos los discos se hacen inaccesibles porque se han escrito partes de los datos en todos los discos. Es ideal para los usuarios que necesitan la máxima velocidad y capacidad. Los editores de vídeo que trabajan con archivos grandes pueden utilizar RAID 0 al editar varios flujos de vídeo para un rendimiento óptimo de la reproducción.

RAID 1, es un modo RAID seguro que requiere al menos 2 unidades y que trabaja con pares de unidades. Se monta un volumen lógico en el ordenador y la capacidad disponible conjunta de ambas unidades está limitada a la del disco de menor capacidad. Si falla uno de los discos físicos, los datos están disponibles al instante en el segundo disco. Los datos no se pierden si falla uno de los discos. Proporciona la máxima seguridad de los datos en el caso de un fallo de disco único, aunque debido a que los datos se escriben dos veces, el rendimiento se reduce ligeramente durante la escritura. RAID 1 es una excelente elección cuando la seguridad es más importante que la velocidad.

RAID 3 utiliza distribución a nivel de byte con un disco de paridad dedicada, de forma que se monta un volumen en el ordenador. Una matriz RAID 3 tolera la avería de un único disco sin pérdida de datos. Si un disco físico falla, los datos del disco averiado pueden reconstruirse en un disco de recambio. Si un segundo disco falla

antes de que pueda reconstruirse en un disco de recambio, se perderán todos los datos de la matriz.

RAID 3 proporciona una excelente seguridad de los datos para entornos donde se leen archivos largos y secuenciales, como archivos de vídeo. La avería del disco no produce una interrupción del servicio, porque los datos se leen desde bloques de paridad. RAID 3 es útil para las personas que necesitan rendimiento y un acceso constante a sus datos, como editores de vídeo. No se recomienda RAID 3 para uso intensivo con archivos no secuenciales porque el rendimiento de lectura aleatoria se ve obstaculizado por la paridad de discos.

RAID 5 combina la distribución en bandas del RAID 0 con la redundancia de datos en una matriz que tenga un mínimo de tres discos. La diferencia entre RAID 3 y un RAID 5 es que una configuración RAID 3 ofrecerá mejor rendimiento a expensas de una capacidad total ligeramente menor. Los datos se distribuyen en bandas entre todos los discos y en cada banda se escribe un bloque de paridad (P) para cada bloque de datos. Si un disco físico falla, los datos del disco averiado pueden reconstruirse en un disco de recambio. Los datos no se pierden en caso de avería de un solo disco, pero si falla un segundo disco antes de que se reconstruyan los datos en una unidad de repuesto, se perderán todos los datos de la matriz.

RAID 5 combina la seguridad de datos con la utilización eficaz del espacio de disco. La avería del disco no produce una interrupción del servicio, porque los datos se leen desde bloques de paridad. RAID 5 es útil para el archivo y para las personas que necesitan rendimiento y un acceso constante a sus datos, como editores de vídeo.

En RAID 6, los datos se distribuyen en bandas entre todos los discos (cuatro como mínimo) y se escriben dos bloques de paridad (p y q en el diagrama de la derecha) en la misma banda por cada bloque de datos. Si un disco físico falla, los datos del disco averiado pueden reconstruirse en un disco de recambio. Este modo RAID puede soportar hasta dos averías de disco sin pérdida de datos. RAID 6 proporciona una reconstrucción más rápida de los datos de un disco averiado.

RAID 6 proporciona fiabilidad de datos con el añadido de una reconstrucción eficaz en caso de avería de disco. Por tanto, RAID 6 es útil para personas que necesitan auténtica seguridad con menos énfasis en el rendimiento.

3.2.2 Cintas magnéticas.

Cómo indica Beekman (1999) las unidades de cintas son dispositivos de almacenamiento comunes en la mayoría de los computadores centrales y algunos computadores personales. La cinta magnética es usada como unidad de almacenamiento. Una cinta puede almacenar enormes cantidades de información en un espacio pequeño y a un costo relativamente bajo.

La cinta magnética tiene una desventaja clara, es un medio de acceso secuencial, sea la información que la cinta contenga, el usuario debe pasar secuencialmente por la información, en el mismo orden en que fue grabada.

Las cintas usadas en los ordenadores son las cintas streamer y las data pack. Para usar estas cintas, se requiere instalar la unidad correspondiente de lectura/escritura.

3.2.3 Memoria Óptica.

Las memorias ópticas como sistemas poseen densidades de almacenamiento relativamente altas, el almacenamiento óptico también es adecuado para la utilización de sistemas extraíbles, y los sistemas ópticos son más resistentes y fiables que los magnéticos, al no existir la posibilidad de colisión de las cabezas que hay en éstos, además de ser mucho más duraderos, un sistema de almacenamiento óptico utiliza métodos ópticos para leer y/o escribir datos sobre un soporte de disco, en esencia un haz láser que explora las variaciones de dos estados de reflexión sobre una superficie especial. Existen distintas tecnologías que llevan a cabo estas operaciones: hay que distinguir entre unidades de disco de sólo lectura y de lectura-escritura.

3.2.3.1 CD (discos compactos).

El CD es un estándar creado en 1985 por Sony y Philips para almacenar audio digital, si bien su uso se ha extendido como soporte de almacenamiento de datos, al ser un método económico y similar a los discos duros de la misma época. Además

de como medio de distribución de contenidos por parte de la industria (discográfica en el caso del CD), también se empezaron a emplear rápidamente (y todavía se sigue haciendo) como medio de distribución de software o incluso como medio para realizar copias de seguridad. Hoy en día, esta última ha quedado en desuso por la gran diferencia entre la capacidad de almacenamiento de medios magnéticos y ópticos en general. La capacidad de los CD varía ligeramente. En primer lugar, encontramos 2 tamaños: discos de 8 cm (también llamados MiniCD) y de 12 cm. Este último es el tamaño más habitual, empleado también en los demás soportes ópticos. La capacidad estándar de estos discos es de 650 Mb o 74 min. de música.

3.2.3.2 Dispositivos (CD-ROM, CD-RW, DVD, WORM).

Los primeros medios ópticos empleados fueron los CD-ROM (Compact Disc-Read Only Memory). Son discos de sólo lectura, que sólo se pueden escribir una vez. Emplean la misma tecnología que los CD de audio. De hecho, las especificaciones técnicas de los distintos formatos están publicados en una serie de libros identificados por colores, entre los que tenemos:

- Libro rojo CD-DA (compact disc-digital audio): son los discos de música
- Libro amarillo CD-ROM: discos compactos de sólo lectura
- Libro naranja CD-R y CD-RW: discos compactos grabables y regrabables.
- Libro blanco VCD: disco de vídeo, antecesor del DVD (e incompatible con él)

La velocidad de acceso a los datos depende del dispositivo lector/grabador. La velocidad de reproducción de audio en tiempo real equivale a una tasa de transmisión de datos de 150 kb por segundo. Pero para transmitir datos podemos incrementar la velocidad. Las medidas se dan como múltiplos de esta velocidad. Así, un lector de 2X duplica la velocidad y es capaz de leer datos a 300 kb por segundo. Las medidas que aparecen en los lectores/grabadores suelen especificar tres datos: la velocidad de lectura, la velocidad de grabación de discos CD-R y la velocidad de grabación de discos CD-RW. Los datos en un disco óptico están grabados formando una espiral. Para representar unos y ceros lo que se hace es perforar la superficie usando un láser de una determinada frecuencia que hace que la temperatura se eleve por encima de los 600° C y el material pierda sus

propiedades cristalinas,⁵ apareciendo un pozo o «pit» que refleja la luz de forma diferente. Las áreas que quedan intactas se denominan «land». La diferencia de capacidad existente entre CD, DVD y discos Blue-Ray se debe simplemente el tamaño de estas marcas: cuanto menor sean, más datos se pueden representar en la misma superficie y el resultado es un disco de más capacidad.

En el caso de DVD, tenemos una gran variedad de formatos. Inicialmente, aparecieron los formatos generados para la distribución de contenidos por parte de la industria en DVD de vídeo o, menos habituales, de audio. Estos discos, al igual que ocurre con los CD, son discos de solo lectura y se denominan DVD-ROM. Para los discos de datos (grabables) aparecieron varias opciones excluyentes e incompatibles entre sí: se trata de los DVD-RAM, los formatos -R y los +R. Los DVD-RAM se emplearon principalmente en cámaras de vídeo y grabadores de televisión. La diferencia con otros formatos en que en ellos la información está almacenada en círculos concéntricos, como en los discos magnéticos, en lugar de en espiral. Los discos -R y +R son propuestas de grupos de empresas distintas, agrupadas en el DVD Forum y en el DVD+RW Alliance respectivamente. En los inicios, los dispositivos lectores y grabadores solo soportaban uno de estos formatos. Poco a poco, se fueron fabricando dispositivos que podían leer cualquier tipo de disco, pero solo podían grabar uno de ellos. Actualmente, los dispositivos actuales pueden trabajar indistintamente con las dos tecnologías. El formato DVD-R es equivalente a los CD-R. Nunca ha existido un CD+R o CD+RW. Y al igual que en los CD, los discos DVD-R y DVD+R son discos que se pueden grabar una sola vez y los discos DVD-RW y DVD+RW pueden borrarse y grabarse varias veces. El número de veces que un disco se puede borrar tiene un límite físico impuesto por el material que se usa, se estima que un disco grabable puede borrarse unas 1.000 veces. Los DVD admiten la grabación en varias capas. Según si los discos tienen una capa o dos y si están grabados por una de las caras o por las dos encontraremos:

- Una cara, una capa (SS SL single-side single-layer) son los discos tradicionales de 4,7 Gb (DVD-5). Solo tienen una capa de material reflectante para almacenar los datos. La otra cara se suele usar para la portada.

- Doble cara, una capa (DS SL double-side single-layer) Aunque no son usuales, pueden crearse discos con una capa por cada lado. En ese caso, hay que darle la vuelta al disco en el lector para leer la otra capa. Se identifican fácilmente porque no llevan ningún tipo de serigrafía y las dos caras son exactamente iguales.
- Una cara, doble capa (SS DL single-side double-layer) son los discos conocidos como de doble capa (DL). Tienen una capacidad de almacenamiento de 8,5 Gb y se denominan DVD-9. Las dos capas están superpuestas y se usa un láser con una frecuencia distinta que atraviesa la primera capa y lee sólo la segunda. Los DVD pregrabados comerciales son de este formato. En ocasiones, se puede percibir una pequeña pausa en un punto de la película cuando el láser cambia para leer la segunda capa.
- Doble cara, doble capa (DS DL double-side double-layer) son discos en los que se colocan dos capas en cada una de las caras. Pueden almacenar hasta 18 Gb de información. A veces también se les llama DVD-WORM, de Write Once Read Many (escribe una vez lee muchas).

3.2.4 Nuevas tecnologías.

Una tecnología reciente que cada vez se está expandiendo más rápidamente debido al aumento de la capacidad de los dispositivos que se pueden construir son las comúnmente conocidas como memorias flash y cuyo nombre técnico es el de memoria de estado sólido. La tecnología está basada en materiales similares a los empleados en las memorias, pero con la salvedad de que son capaces de mantener la información de forma definitiva sin necesidad de corriente eléctrica. Este tipo de tecnología no ha desplazado a la memoria RAM en los ordenadores porque es más lenta y porque el número de ciclos de escritura está limitado, por lo que con el tiempo iríamos perdiendo capacidad de memoria (más rápidamente que en el caso de un disco). Dentro de esta categoría de dispositivos de almacenamiento encontramos las tarjetas de memoria de diversos tipos: compact flash (CF), secure digital (SD), memory stick™ y multimedia card (MMC) entre otras, incluyendo también los cartuchos de las video consolas y la memoria interna de reproductores portátiles de

audio y vídeo. Para el almacenamiento de datos extraíbles en un ordenador personal, el dispositivo rey es la memoria USB. Recientemente han empezado a comercializarse discos duros como los convencionales (magnéticos) que emplean esta tecnología. Se conocen como Solid State Drives (SDD) y su punto de entrada han sido los ordenadores portátiles, debido a la reducción del consumo y a la menor tasa de fallos al no incorporar partes móviles que puedan estropearse como consecuencia de golpes. Sin embargo, todavía son dispositivos demasiado caros, con menos capacidad que los discos duros magnéticos y con dos problemas importantes que todavía es necesario resolver: el tiempo de acceso para escritura y el número máximo de operaciones en una misma celda. Sin embargo, el acceso para lectura es mucho más rápido, lo que se nota especialmente en el proceso de arranque (en torno a los 20 segundos con la máquina dispuesta para trabajar) o en la carga de aplicaciones (prácticamente instantánea).

3.2.4.1 Unidad zip.

Cuando se habla de ZIP muchos piensan en el programa WIN ZIP y no en las unidades los cuales las primeras unidades Zip tenían capacidad de 100MB, poco después salieron unas de 250MB y actualmente tenemos las de 750MB, estas últimas tiene un desempeño mejorado en cuanto a velocidad. Los discos para las unidades Zip, son un poco más gruesos que los disquetes pero tiene la misma forma y diseño, generalmente las podemos encontrar como unidades externas.

3.2.4.2 Memory stick.

Este formato o tipo de memoria flash pertenece a Sony que la usan en todos sus productos como por ejemplo cámaras digitales, PDA's y reproductores de música digital, cuando recién salieron al mercado tenían una capacidad de 128MB, pero ahora hay hasta de 64GB en una sola tarjeta, es un poco más cara que otro tipo de memoria flash.

3.2.4.3 Discos portátiles USB (flash drive).

Es un pequeño dispositivo que tenemos que conectar al puerto USB para poder transferir datos sin complicaciones, los primeros USB drivers tenían capacidad de 8MB, pero ahora podemos encontrar algunos de 78 hasta 6 4gigas, tienen una alta portabilidad y compatibilidad.

3.2.4.4 Mini disc.

Uno de los primeros fabricantes de estos discos, que funcionan de forma similar al de un computador, fue IBM, en el 2000. En ese entonces, la novedad era que venían en tres tipos: 340 MB, 512 MB y 1 GB de espacio, lo cual sigue siendo muy bueno actualmente. Sin embargo, otros fabricantes, como Hitachi y Toshiba, ofrecen discos de 2 y 4 GB con un tamaño de tan solo 2,5 centímetros de diámetro. Se utilizan en cámaras digitales y reproductores de MP3 portátiles, principalmente. Según los fabricantes, los minidiscos o 'microdrives' ofrecen tasas de transferencia superiores que las tarjetas Flash. Eso quiere decir que se demoran menos en pasar información, por ejemplo, a un computador.

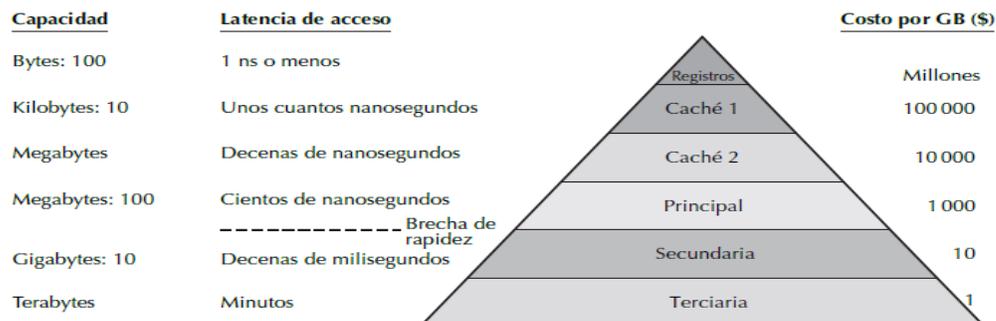
ACTIVIDADES

1. Elección de tecnología de memoria

- Las especificaciones de una cámara digital indican que su subsistema de almacenamiento de fotografías contiene una unidad de memoria SRAM muy grande como para retener un par de fotografías y un módulo de memoria tipo *flash* que puede almacenar unas 100 fotografías. Discuta las razones que son motivo de las elecciones de tecnologías y capacidades para los dos componentes de memoria.
- Repita la parte a) para un organizador electrónico, que almacene varios miles de nombres e información de contacto asociada, con la misma combinación de memorias: una unidad de memoria SRAM pequeña y un módulo de memoria *flash* muy grande.

2. Características de jerarquía de memoria

Con base en la figura, ¿cuál nivel de una memoria jerárquica es probable que tenga mayor costo? Observe que se pregunta el costo total, no el costo por byte.



3. Analogía para la memoria jerárquica de una computadora

Considere la forma jerárquica en la que una persona trata con los números telefónicos. Tiene memorizados algunos de números telefónicos importantes. Los números para otros contactos clave se mantienen en un directorio telefónico de bolsillo o un organizador electrónico.

Si en este caso se avanza más por el equivalente de la figura del problema 2, se llega al directorio telefónico de la ciudad y, finalmente, a la colección de directorios de todo el país disponibles en la biblioteca local. Dibuje y etiquete adecuadamente un diagrama piramidal para representar este sistema jerárquico, y mencione las características importantes de cada nivel.

Bibliografía

- Guijarro Rodríguez, . A. A., García Arias, P. M., & Yanza Montalván, . Ángela O. (2020). Organización y Arquitectura de computadoras. Un enfoque práctico. *Etecam*. Recuperado a partir de <http://www.etecam.com/index.php/etecam/article/view/13>.
- Rebollo, M.. (2008). Dispositivos de almacenamiento. Universidad Politécnica de Valencia. España bajo licencia Creative Commons Reconocimiento 3.0 (CC BY). Recuperado a partir de https://www.researchgate.net/publication/50838253_Dispositivos_de_almacenamiento
- Hilari, S., (2006). Dispositivos de almacenamiento. *Revista de Bibliotecología y ciencias de la información*. Vol. 10. 15 de diciembre de 2006. Pág., 75-81.
- Rodríguez, M. (2007). Recuperación de Información en discos duros. Tesis para optar al título de Ingeniero en Comunicaciones y Electrónica. Instituto Politécnico Nacional. México.
- Orenga, M., Manonellas, G. Sistema de memoria. Universidad Oberta de Catalunya. España.
- Stalling, W. (2006). Organización y Arquitectura de Computadoras. 7ª. Edición. España: Prentice Hall.
- León Socha, F. A. (2017). Arquitectura de hardware. Fundación Universitaria del Área Andina. Bogotá. 978-958-5460-46-1.

CAPÍTULO 4:

ENTRADA/SALIDA

Objetivos:

- ✓ Clasificar los diversos elementos que permiten la captura / despliegue de la información de acuerdo con la entrada / salida de los datos.
- ✓ Describir las características de periféricos de E/S y la forma en la que impactan en la sociedad moderna.

¿De qué se trata esta sección de aprendizaje?

Muchos de los datos que se procesan con una computadora se originan fuera de la máquina, los resultados del cálculo se obtienen mediante la ejecución de programas que con mucha frecuencia van hacia fuentes externas de interpretación, archivado o mayor procesamiento. En este contexto, la latencia de las operaciones de entrada y salida puede tener un impacto significativo sobre la percepción del rendimiento de una computadora; importa poco que un CPU corra un programa en una fracción de segundo si tarda muchos minutos en proporcionar los datos que deben procesarse y obtener de vuelta los resultados. Por tanto, es imperativo, para diseño y uso efectivo de computadora, que se entienda cómo funcionan los dispositivos de entrada/salida, cómo interactúan con la memoria y el CPU, y cómo se puede usar una computadora para recolectar datos desde fuentes externas o para activar y controlar dispositivos externos.

En el capítulo 4 se revisan los tipos de dispositivos de entrada/salida y sus detalles de implementación que tienen un vínculo con el costo, el rendimiento o las tasas de datos requeridas para apoyar sus interacciones con la memoria y el CPU.

4. Dispositivos externos – periféricos.

El tercer componente fundamental de una computadora es la Unidad de Entradas y Salidas (UE/S), además de la CPU y la memoria. En realidad, se trata de un conjunto de módulos de E/S especializados que se conectan habitualmente al sistema de buses (bus de direcciones, bus de datos y bus de control) y controlan la comunicación con uno o más dispositivos periféricos (o simplemente periféricos). Un módulo de E/S tiene la capacidad necesaria para permitir la comunicación entre el periférico y el sistema de buses.

Los periféricos no se conectan directamente al sistema de buses, para alcanzar a la CPU y la memoria (salvo algún caso especial), por diversos motivos. Los periféricos son muy diferentes entre sí, y debería disponerse de una lógica adicional dentro de la CPU para cada dispositivo. Además, la mayoría de los periféricos son lentos respecto a la CPU y la memoria, por lo que se desaprovecharía tiempo y prestaciones conectándolos a un sistema de buses de alta velocidad. También es inconveniente que la CPU y la memoria gestionen las transferencias de algunos periféricos de muy altos rendimientos. Finalmente, los periféricos son distintos constructivamente, y utilizan datos con formatos y tamaños distintos a los de la computadora con la que se comunican.

Los módulos de E/S son la interfaz o nexo entre el procesador y la memoria, mediante el conjunto de buses, por un lado, y los dispositivos periféricos usando enlaces de datos específicos.

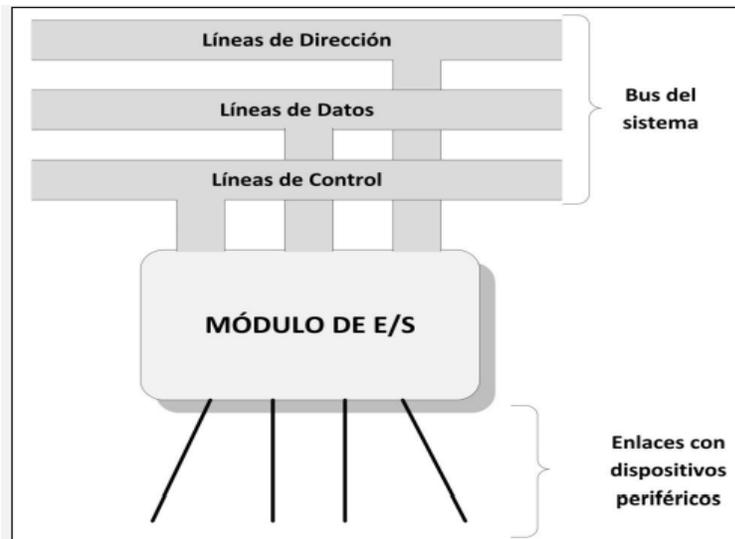


Figura 4.1: Módulo de E/S

4.1.1 Categorías.

Los periféricos se clasifican según:

- El sentido de la transferencia de los datos en: Periféricos de entrada, Periféricos de salida y Periféricos de entrada-salida.
- Con quién interactúan en: de interacción con humanos y de interacción con máquinas

4.1.2 Señales de comunicación.

Los periféricos se comunican a la computadora a través de un enlace a un módulo de E/S específico. El enlace permite el intercambio de señales de control, de estado y los datos entre el módulo de E/S y el dispositivo externo. Con las señales de control se determina la función que debe realizar el periférico. Mientras que con las señales de estado se indica el estado del dispositivo, por ejemplo, para señalar si está preparado o no para una transferencia. Finalmente, con los datos se produce el efectivo intercambio de la información digital que se envía o recibe desde el módulo de E/S.

Por lo expuesto, cada periférico deberá tener una lógica de control que controla la operación con el módulo de E/S. Las señales eléctricas de los datos pueden requerir una transducción o adecuación en el caso de una salida o una entrada dependiendo de las características operativas del periférico. Finalmente, el dispositivo puede

disponer de una memoria o buffer para almacenar temporalmente los datos que van o vienen desde el módulo de E/S hacia el periférico. Estos aspectos se muestran en la Figura 4.2.

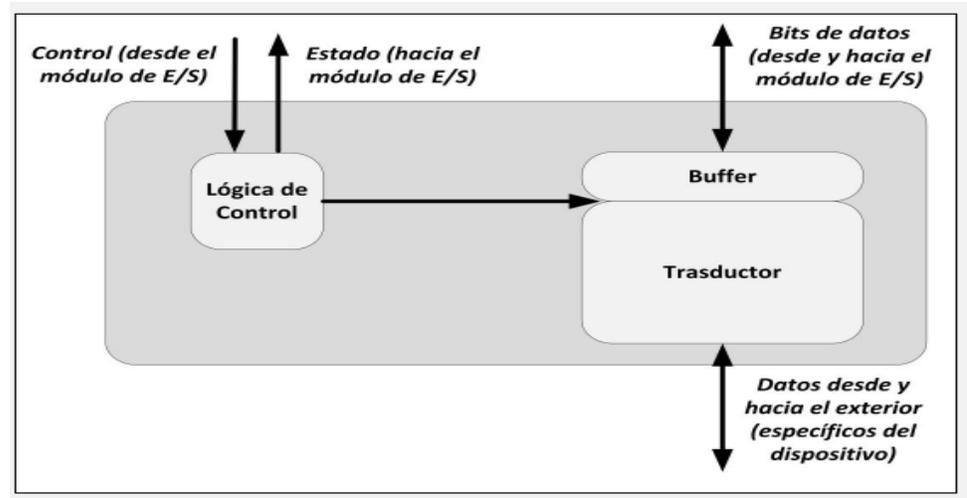


Figura 4.2: Diagrama en bloques de un dispositivo externo o periférico

4.1.2.1 Control.

Las señales de control y tiempo se utilizan para regular la transferencia elemental, indicando cómo y cuándo debe ocurrir. O sea que permiten contestar las siguientes preguntas:

- ¿Puede producirse una transferencia ahora?
- ¿Qué tipo de transferencia se va a realizar?
- ¿Cuándo comienza?
- ¿Cuándo termina?
- ¿Se realizó con éxito?

Estas preguntas nos permiten deducir que las funciones de la "lógica o inteligencia" de un bus de E/S son:

- Comunicarse con el periférico y el sistema CPU-memoria.
- Controlar la temporización durante la transferencia.
- Almacenar temporalmente bits para "paliar" la diferencia de velocidad entre emisor y receptor.
- Detectar si se produjeron errores durante la transferencia.

Cuando la actividad del bus de E/S es sincrónica se utilizan señales del *dock* que regulan la transferencia, por lo tanto, la frecuencia de *dock* es un parámetro para considerar respecto de la velocidad de la transferencia. Un bus sincrónico requiere que los dispositivos conectados a él estén "sintonizados" a esa frecuencia y que todas las actividades se produzcan en intervalos de tiempo fijos.

Cuando la actividad del bus es asincrónica, las señales del *dock* no regulan su operación.

Por lo tanto, su velocidad depende de los tiempos de los dispositivos conectados a ellos.

El siguiente es un resumen de las señales diferentes comunes a la mayoría de los buses de E/S para "controlar" la transferencia:

- Señal de *dock*, que permite sincronizar el ciclo del bus con la operación de la CPU (CLK).
- Señal que puede habilitar un tiempo de espera.
- Señal de lectura o escritura a un dispositivo de E/S o a la memoria (IOR o IOW).
- Señales de interrupción a la CPU (IRQ_n).
- Señales de reconocimiento (aceptación o requerimiento), que permiten el diálogo de los dispositivos durante la transferencia (p. ej., la técnica de *hand shaking* o "apretón de manos").
- Señal de bus cedido u ocupado.

4.1.2.2 Estado.

Las señales de estado nos dan información del estado del módulo de E/S, como saber si el módulo está disponible o está ocupado, si hay un dato preparado, si se ha acabado una operación, si el periférico está puesto en marcha o parado, qué operación está haciendo, o si se ha producido algún error y qué tipo de error. Estas señales se actualizan generalmente mediante la lógica del módulo de E/S y se almacenan en el registro de estado.

4.1.2.3 Datos.

Se considera a un **dato** como un ente abstracto que contiene información. Para lograr su representación física utilizamos las ondas electromagnéticas. A esta representación la llamamos **señal**. Las señales electromagnéticas pueden ser **continuas** cuando su intensidad varía suavemente en el tiempo, o **discretas** cuando su intensidad cambia bruscamente en el tiempo. En este último caso, si la cantidad de valores que puede adoptar una señal en el tiempo son 2 (nivel alto y nivel bajo), la señal se llama señal binaria. La Figura 4.3 indica una señal continua y una señal discreta.

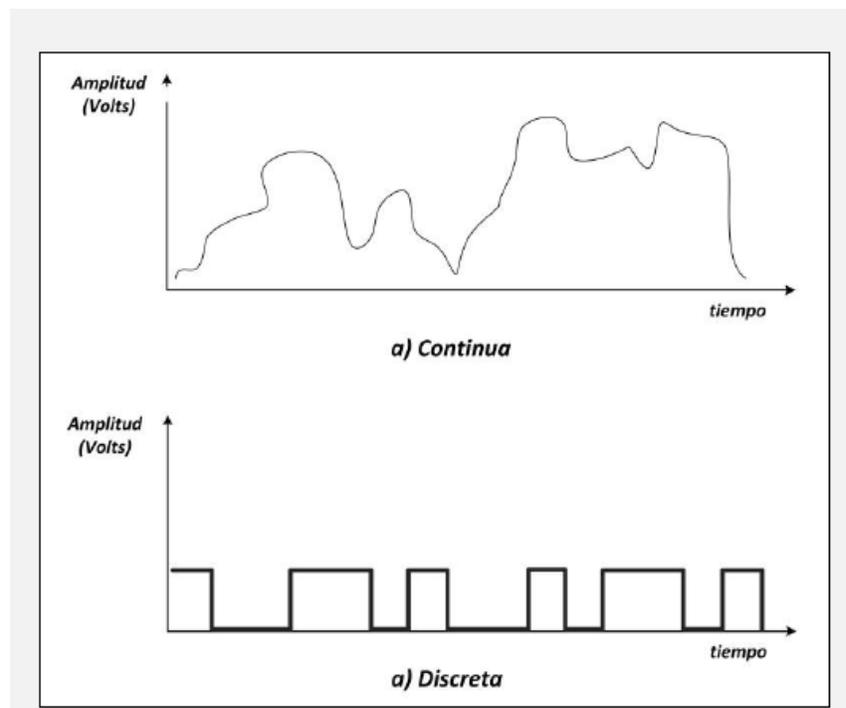


Figura 4.3: Señales continuas y discretas

Los datos, como las señales, pueden ser continuos o discretos. A los datos continuos se los llama **analógicos** y a los discretos se los llama **digitales**.

Tanto los datos analógicos como los digitales pueden transportarse desde una fuente a un destino. Este transporte mediante la propagación y procesamiento de señales recibe el nombre de **transmisión**.

Las señales de dato representan los bits del mensaje que se ha de transferir ($Dn = \text{data "sub"}n$). El “ancho de bus” o bien la cantidad de líneas afectadas a la transferencia elemental dan una medida de potencial de trabajo; esto es, a mayor cantidad de bits transferidos en paralelo mayor capacidad de transferencia.

4.1.3 Tipos básicos.

4.1.3.1 Teclado.

Un teclado consiste en un arreglo de teclas que se usa para ingresar información a la computadora u otro dispositivo digital. Los teclados más pequeños, como los que se encuentran en los teléfonos, o en un área separada en muchas computadoras de escritorio, se les conoce como teclado numérico (keypad). Casi todos los teclados alfanuméricos siguen la plantilla QWERTY, nombre derivado de las etiquetas de las primeras teclas en la hilera superior de letras en el teclado de máquina de escribir estándar. Con los años se han hecho serios intentos por introducir otras plantillas de teclas que harían más fácil alcanzar letras del alfabeto de uso más frecuente, y aumentar la rapidez de entrada de datos.

Se dice que la plantilla QWERTY se eligió para reducir deliberadamente la rapidez de tecleo con intención de evitar el atasco de los martillos mecánicos de las primeras máquinas de escribir. Desafortunadamente, la familiaridad con la plantilla QWERTY y la gran experiencia en su uso han evitado la adopción de estas plantillas más sensibles. Sin embargo, se ha progresado algo en el área de colocaciones de teclas no convencionales para dar al usuario comodidad durante el tecleo. Los teclados que siguen tales diseños se denominan teclados ergonómicos.

Los interruptores mecánicos, se usan en teclados de escritorio estándar en vista de la retroalimentación táctil y su construcción más fuerte. Puesto que los contactos se pueden ensuciar con el paso del tiempo, otros tipos de interruptores se apoyan en la fuerza inducida forma magnética en lugar de mecánicamente para cerrar un par de contactos. De esta forma, los contactos se pueden colocar en un coto sellado para aumentar la confiabilidad.

Sin importar qué tipo de contacto se use, el circuito se puede reabrir y volver a cerrar muchas

veces con cada presión de tecla. Este efecto, conocido como rebote de contacto, se puede evitar al no usar la señal de la tecla directamente sino dejarla que active un *flip-flop*, cuya salida se vuelve alta con una presión de tecla y permanece así sin importar la longitud o severidad del rebote de contacto.

También es posible lidiar con el efecto del rebote de contacto con diseño adecuado del software que maneja adquisición de datos del teclado.

Sin importar la plantilla física, las teclas de un teclado o teclado numérico con frecuencias se

ordenan en un arreglo cuadrado o rectangular 2D desde un punto de vista lógico. Cada tecla está en la intersección de un circuito de hilera y de columna. Presionar una tecla provoca cambios eléctricos en los circuitos de hilera y columna asociados con la tecla, ello permite la identificación de la tecla que se presionó. Entonces un codificador convierte las identidades de hilera y columna en código de símbolo único (usualmente ASCII), o secuencia de códigos, para que se transmitan a la computadora. La plantilla física del teclado y el ordenamiento de las teclas no son relevantes para los procesos de detección y codificación.

Con el propósito obtener la salida de cuatro bits que representa la tecla presionada en el teclado numérico hexa, las señales de hilera se postulan a su vez, acaso con el uso de un contador de anillo de cuatro bits que sigue la secuencia de conteo 0001, 0010, 0100 y 1000. Conforme una señal se postula en una hilera particular, las señales de columna se observan. La codificación de dos bits del número de hilera unido a la codificación de dos bits del número de columna proporciona la salida de cuatro bits. Un teclado de 64 teclas puede estar ordenado lógicamente en un arreglo 8×8 (aun cuando el ordenamiento físico de las teclas pueda no ser cuadrado). Entonces se produce el código de salida de seis bits en forma similar al procedimiento para el teclado hexa. Si se desea la representación ASCII del símbolo, se puede usar una tabla de consulta con 64 entradas.

4.1.3.2 Ratón.

Además de un teclado, la mayoría de las computadoras de escritorio y laptop tiene dispositivos

apuntadores que permitirán al usuario elegir opciones de menús, seleccionar texto y objetos gráficos para edición, y realizar una diversidad de otras operaciones. De hecho, algunos dispositivos sin teclado se apoyan exclusivamente en el señalamiento para funciones de control y captura de datos. El dispositivo apuntador usado más comúnmente es el *ratón*, llamado así por su forma en algunos de los primeros diseños y el alambre con forma de cola que lo conecta a la computadora. Los ratones modernos vienen en varias formas, muchos de ellos usan enlaces inalámbricos con la computadora.

En un ratón mecánico, dos contadores se asocian con los rodillos x y y. Levantar el ratón establece los contadores a cero. Cuando el ratón se arrastra, los contadores aumentan o disminuyen de acuerdo con la dirección de movimiento. La computadora usa los valores del contador para determinar la dirección y extensión del movimiento que se aplicará al cursor. Los ratones ópticos son más precisos y menos proclives a fallo debido a la reunión de polvo e hilachos en sus partes. Los ratones ópticos más simples detectan líneas de retícula en un tapete de ratón especial.

Las versiones nuevas, más avanzadas, usan pequeñas cámaras digitales que les permiten detectar movimiento sobre cualquier superficie.

Con frecuencia, un *touchpad* (almohadilla táctil) sustituye un ratón para aplicaciones en las laptops.

La ubicación de un dedo que toca el tapete se percibe (a través de circuitos de hilera y columna) y el movimiento del cursor se determina según cuán lejos, rápido y en qué dirección se movió el dedo.

Una pantalla táctil es similar, pero, en lugar de un tapete separado, la superficie de la pantalla de presentación se usa para apuntar. Otros dispositivos apuntadores incluyen la *trackball* (esfera móvil, que en esencia representa un ratón mecánico boca abajo cuya gran bola se manipula a mano) y el *joystick* (palanca de control) que encuentra aplicaciones en muchos juegos de computadora, así como en

escenarios de control industrial. Algunas computadoras laptop usan un pequeño joystick incrustado en el teclado para apuntar.

4.1.3.3 Monitor.

La presentación visual de símbolos e imágenes es el método de salida primario para la mayoría de las computadoras. Las opciones disponibles varían desde pequeñas pantallas monocromas (de los tipos usados en los teléfonos celulares y calculadoras baratos) hasta los grandes dispositivos de presentación de alta resolución, y más bien costosos, dirigidos al uso de artistas gráficos y publicistas profesionales.

Hasta hace poco el *tubo de rayos catódicos* (CRT) era el tipo principal de dispositivo de presentación visual para las computadoras de escritorio, las pantallas de panel plano se reservaban para su uso en laptop y otros dispositivos digitales portátiles. La mayoría de los voluminosos, pesados y grandes consumidores de electricidad CRT gradualmente son sustituidos por las unidades de pantalla de panel plano, cuyos costos son accesibles y cuyas pequeñas huellas y menor generación de calor son ventajas importantes para el hogar y la oficina.

Las pantallas CRT funcionan mediante un haz de electrones que barre la superficie de una pantalla sensible, y crea un pixel oscuro o claro conforme pasa cada punto.

Los primeros CRT

eran *monocromáticos*. En tales CRT, el haz de electrones golpea una capa de fósforo en la parte posterior del vidrio de la pantalla. Esta capa de fósforo emite luz porque es golpeada por una corriente de electrones que viaja a muy alta rapidez, y la intensidad del haz de electrones que pasa un punto específico determina el nivel de brillantez. Para proteger la capa de fósforo del bombardeo directo de los electrones, detrás de ella se coloca una capa de aluminio; ésta actúa como contacto eléctrico. La tecnología del CRT de color pasó por varias etapas. Los primeros intentos usaron un CRT monocromático que se veía mediante la rotación de filtros de color mientras se desplegaban a la vez los pixeles asociados con varios colores. Otros diseños incluyeron la sustitución de los filtros mecánicos con otros controlados electrónicamente y el uso de múltiples capas de fósforo, cada uno emitiendo luz de color diferente.

Los CRT modernos en uso común actualmente se basan en el método tricolor de “máscara de sombra” introducido por los televisores RCA en la década de 1950. Los detalles de diseño varían, pero es representativo el esquema que se usa en los tubos Trinitron de Sony. Estrechas tiras de fósforo que producen tres colores de luz diferentes y que se colocan en la parte posterior de la cara de vidrio del tubo. Los tres colores son rojo, verde y azul, que propicia el nombre “RGB” (por sus siglas en inglés) al esquema resultante. Tres haces de electrones separados exploran la superficie de la pantalla, cada uno llega en un ángulo ligeramente diferente. Una máscara de sombra, que representa una placa metálica con aberturas u hoyos, fuerza a cada haz a impactar sólo tiras de fósforo de un color particular. La separación de dos tiras consecutivas del mismo color, conocida como *densidad de presentación (display pitch)*, dicta la resolución de la imagen resultante.

Los tres haces de electrones se controlan con base en una representación de la imagen que se desea presentar. Todo pixel está asociado con cuatro (imágenes simples de 16 colores) a 32 bits de datos en un *buffer de marco*. Si la resolución de la pantalla es de 1 K × 1 K pixeles, y cada pixel tiene 64K colores posibles, entonces el *buffer* de marco necesita 2 MB de espacio. Este espacio se puede proporcionar en una *memoria de video* dedicada o como parte del espacio de dirección de memoria principal (*memoria compartida*). Con frecuencia, las memorias de video dedicadas son de puerto dual para permitir el acceso simultáneo mediante el CPU para modificar la imagen o por el controlador de pantalla para presentarla. A tales memorias de video a veces se les conoce como VRAM.

Físicamente, las pantallas CRT son voluminosas, requieren alto consumo de energía y generan gran cantidad de calor. Además, la imagen presentada por un CRT sufre de distorsión, así como de foco y color no uniformes. Por estas razones, las pantallas de cristal líquido en panel plano (LCD) son populares, en especial a la luz de su continua mejora en la calidad de imagen y su costo económico.

En una pantalla LCD, cada punto de intersección entre una hilera y una columna representa una pequeña persiana óptica que se controla mediante la diferencia entre los voltajes de hilera y columna. Puesto que el voltaje de hilera se proporciona a toda celda en una hilera y las celdas además reciben voltaje cruzado de otras

hileras, el contraste y la resolución tienden a ser bajos. Las *matrices activas* son similares, excepto que en cada intersección se planta un transistor de película delgada. Un transistor que se enciende permite que las celdas de cristal líquido asociadas se carguen a los voltajes en las líneas de columna. De esta forma, una imagen que se “escribe” hilera por hilera se conserva hasta el siguiente ciclo de regeneración.

Otras tecnologías de pantalla de panel plano incluyen las que se basan en diodos emisores de luz (LED, por sus siglas en inglés) y el fenómeno de plasma. Las pantallas LED consisten en arreglos de LED de uno o más colores, y cada LED se “direcciona” a través de un índice de hilera y otro de columna. Los LED electrónicos son adecuados para presentaciones que se deben ver en exteriores, requieren enorme grado de brillantez. Un desarrollo reciente en las pantallas de LED es el surgimiento de los LED orgánicos (OLED, por sus siglas en inglés), que requieren menos energía que los LED ordinarios y ofrecen una variedad de colores, incluido el blanco.

La operación de las pantallas de plasma es similar a la de una lámpara de neón. Explotan la propiedad de ciertas mezclas de gases que se descomponen en plasma cuando se sujetan a un campo eléctrico muy intenso. El plasma conduce la electricidad y convierte una parte de la energía eléctrica en luz visible. En la actualidad, las pantallas de plasma son muy caras; por tanto, tienen aplicaciones limitadas.

Muchas computadoras también ofrecen salida compatible con el formato de presentación en televisión, ello permite que cualquier TV actúe como unidad de presentación. La misma señal compatible con TV se puede proporcionar a un *proyector de video* que produce una réplica de la imagen en el CRT o unidad de pantalla de panel plano de la computadora en una pantalla del tamaño de una pared. Estas tecnologías de presentación son útiles para presentaciones basadas en computadora en todas partes, desde pequeñas salas de juntas o salones de clase hasta grandes auditorios.

4.1.3.4 Impresora.

Las primeras impresoras para computadora funcionaban en forma parecida a las antiguas máquinas de escribir mecánicas; usaban dispositivos de formación de caracteres que golpeaban sobre una banda de tela o plástico mediante un mecanismo parecido a un martillo para imprimir los caracteres sobre papel uno tras otro (impresoras de caracteres). Para aumentar la rapidez de tales impresoras, se desarrollaron mecanismos especiales que permitían la impresión de líneas completas a la vez, ello condujo a una amplia variedad de impresoras de línea. Gradualmente, las impresoras de línea evolucionaron de las ruidosas y voluminosas máquinas (con tamaño de refrigerador) a unidades más pequeñas y silenciosas.

Pronto hubo conciencia de que la formación de caracteres mediante la selección de un subconjunto de puntos en una matriz 2D de puntos, conduciría a mayor flexibilidad en los conjuntos de caracteres arbitrarios de soporte, así como en la formación de imágenes. Por tanto, las impresoras de matriz de punto gradualmente sustituyeron a muchas tecnologías de impresión del tipo impacto.

Las impresoras modernas básicamente imprimen una gran imagen en matriz de punto de la página que está compuesta de archivos de texto o gráfico, mediante *postscript* u otros formatos intermedios de archivo de impresora. Cada punto en la imagen de la página se caracteriza por intensidad y color.

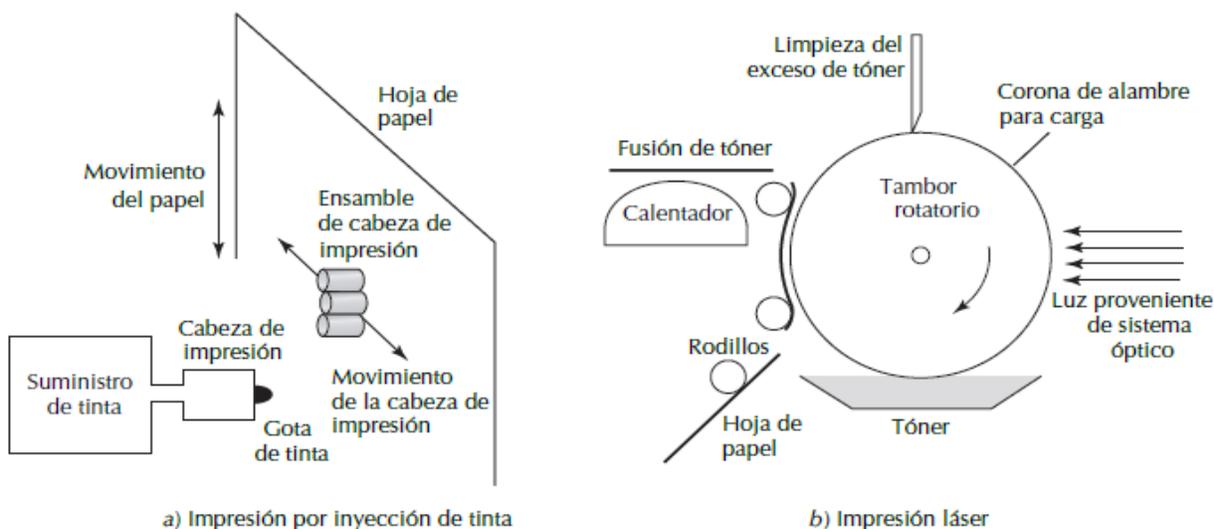


Figura 4.4: Impresoras de inyección de tinta y láser.

Las impresoras de inyección de tinta, relativamente lentas y baratas, imprimen los puntos uno o pocos a la vez. Las gotas de tinta se expulsan desde la cabeza de impresión mediante varios mecanismos como calentamiento (que conduce a la expansión de una burbuja de aire dentro de la cabeza) o generación de ondas de choque a través de un transductor de cristal piezoeléctrico al lado del depósito. Las impresoras láser más grandes y rápidas, forman una copia de la imagen a imprimir en la forma de patrones de carga eléctrica en un tambor rotatorio e imprimen toda la página a alta rapidez.

Al igual que un escáner, una impresora de matriz de punto se caracteriza por su resolución espacial, expresada en número de píxeles o puntos por pulgada (por ejemplo, 1 200 dpi), y fidelidad de color (número de colores). El rendimiento total de impresión, otra característica clave de rendimiento para las impresoras de computadora, varía de unas cuantas a muchos centenares de páginas por minuto (ppm), en ocasiones varían incluso para la misma impresora dependiendo de la resolución y requisitos de color. Muchas impresoras antiguas requerían papel especial cubierto con químicos o empacado en rollos para simplificar el mecanismo de alimentación de papel. Las impresoras modernas usan papel ordinario, se denominan impresoras de papel normal.

Las impresoras para blanco y negro depositan gotas de tinta o funden partículas de tóner en el papel de acuerdo con los requisitos del documento. Los niveles de gris se crean al dejar que se muestre parte del blanco del papel subyacente. Las impresoras de color funcionan de modo similar a los CRT de color en que crean varios colores a partir de tres colores. Sin embargo, los tres colores que se usan en las impresoras son cian (azul-verde), magenta y amarillo (*yellow*), que en conjunto forman el esquema de color CMY (por sus siglas en inglés), es diferente del RGB de los CRT. Las razones para la diferencia tienen que ver con la forma en que el ojo humano percibe el color. El esquema de color CMY representa la ausencia de los colores RGB (cian es la ausencia de rojo, etc.). De este modo, el cian absorbe rojo, el magenta absorbe verde y el amarillo absorbe azul. El esquema de color RGB es aditivo, ello significa que un color deseado se crea al sumar la cantidad adecuada de cada uno de los tres colores primarios.

El esquema CMY es sustractivo y forma un color deseado al remover los componentes adecuados de la luz blanca. La mezcla de estos tres colores en cantidades iguales absorbería los tres colores primarios y dejaría negro. Sin embargo, el blanco así producido es más bien insatisfactorio, especialmente en vista de la extrema sensibilidad del ojo humano ante cualquier color corrido en negro. Por esta razón, la mayoría de las impresoras de color usan el esquema CMYK, donde K representa al negro.

La impresión a color es mucho más complicada que la pantalla a color. Las razones incluyen la dificultad de controlar exactamente el tamaño y alineación de los puntos de los diversos colores y la posibilidad de que los colores se corran si se colocan muy juntos. Mayores problemas surgen de la resolución reducida cuando se deben soportar diferentes niveles de intensidad (escalas de grises en negro y blanco). Por ejemplo, una forma de crear la ilusión de cinco niveles de gris (0, 25, 50, 75 y 100%) es dividir el área de impresión en bloques de píxeles de 2×2 , y colocar 0-4 píxeles negros en un bloque para crear los cinco niveles. Sin embargo, esto último reduce la resolución por un factor de 2 en cada dirección (un global cuadruplicado).

4.1.3.5 Escáner.

Un escáner o digitalizador funciona como una unidad de pantalla CRT, pero en dirección contraria. Mientras que una pantalla CRT convierte señales electrónicas en niveles de color y brillo con un movimiento de barrido, un escáner percibe el color y la intensidad y los convierte en señales electrónicas conforme recorre una imagen línea por línea y punto a punto dentro de cada línea. Esta información se almacena en memoria en uno de muchos formatos estándar.

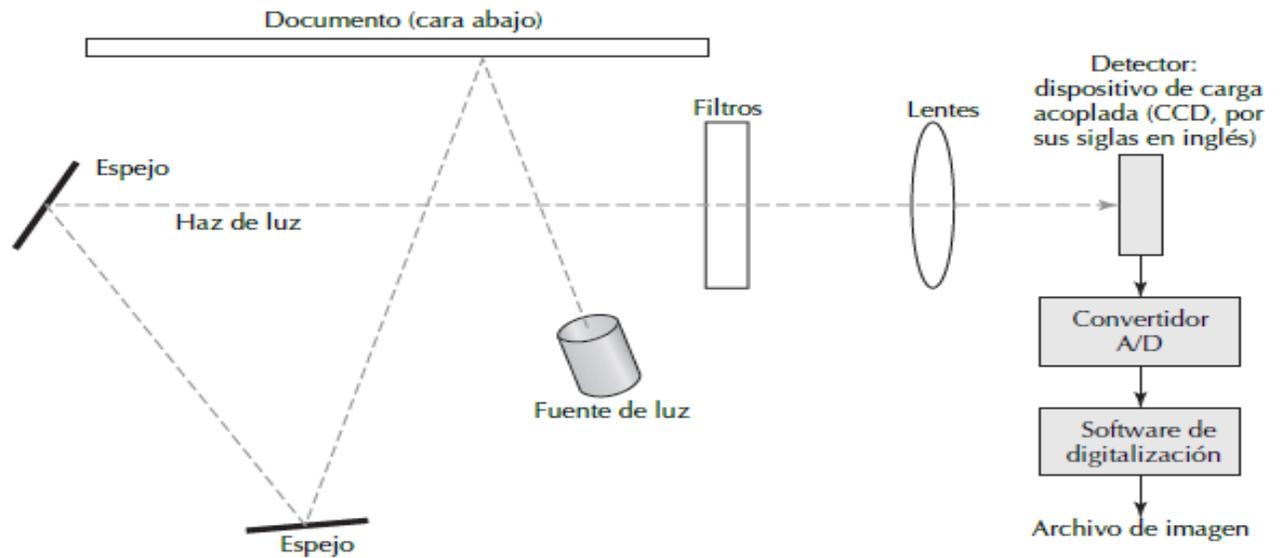


Figura 4.5: Mecanismo de digitalización para entrada de copia impresa.

Los escáneres se caracterizan por resolución espacial, expresada en número de píxeles o puntos por pulgada (por ejemplo, 1 200 dpi) y fidelidad de color (número de colores). En relación con el tipo de entrada de documento al escáner, tales dispositivos se dividen en los tipos alimentador de hoja, cama plana, de cabeza y portátiles. Los escáneres de cama plana tienen la ventaja de no provocar daño al documento original debido a doblado y permiten la digitalización de un libro y páginas de periódico. Los escáneres de cabeza y portátiles pueden digitalizar documentos grandes sin ser voluminosos o costosos. Los escáneres baratos están haciendo obsoletas las máquinas de fax.

Para el caso del texto digitalizado, la imagen se puede convertir en forma simbólica a través de software de *reconocimiento óptico de caracteres* (OCR, por sus siglas en inglés) y se almacena como archivo de texto. Este tipo de conversión de imagen a texto mediante digitalización y OCR se usa comúnmente para poner en línea libros antiguos y otros documentos. El reconocimiento de la escritura manual también se ha vuelto cada vez más importante conforme proliferan las aplicaciones de los asistentes digitales personales (PDA, por sus siglas en inglés) y las computadoras portátiles sin teclado.

4.1.3.6 Bocinas y micrófono.

La entrada de audio se captura mediante micrófonos. La mayoría de las computadoras de escritorio y laptop vienen con micrófono, o un par estéreo, y contiene una tarjeta sonora que puede capturar sonidos mediante un puerto de micrófono. Para almacenar en una computadora, el sonido se digitaliza al tomar muestras de la forma ondulatoria a intervalos regulares. La fidelidad del sonido se mejora al aumentar la tasa de muestreo y al usar más bits para representar cada muestra (por decir, 24 bits para resultados profesionales, en lugar de 16 bits). En virtud de que estas provisiones conducen a aumento en los requisitos de almacenamiento, usualmente el sonido se comprime antes de almacenarlo. Los ejemplos de formatos estándar para audio comprimido incluyen MP3 (abreviatura para MPEG-1, capa 3), Real Audio, Shockwave Audio y el formato WAV de Microsoft Windows. MP3 comprime los archivos de audio hasta casi 1 MB por minuto y conduce a audio con calidad de CD porque en el curso de la compresión remueve sólo componentes de sonido que superan el rango auditivo humano. Además de esos formatos, se pueden usar los de película MPEG y Quick Time para almacenar sonido al ignorar el componente de video.

4.1.3.7 Otros.

Además de las memorias secundaria y terciaria que constituyen los dispositivos I/O usados comúnmente para almacenamiento estable y archivado de datos, están disponibles muchos otros tipos de unidades de entrada y/o salida.

Una cámara digital fija o de video captura imágenes para ingresar a una computadora en forma muy similar a un escáner. La luz entrante proveniente del exterior de la cámara se convierte en píxeles y se almacena en memoria *flash* o algún otro tipo de unidad de memoria no volátil. Las cámaras digitales fijas se caracterizan por su resolución en términos del número de píxeles en cada imagen capturada.

Por ejemplo, una cámara de un megapixel puede tener una resolución de 1 280 × 960. Las cámaras con resoluciones de cinco o más megapíxeles son muy costosas y usualmente no se necesitan en aplicaciones sofisticadas, pues una cámara de dos

megapíxeles puede entregar impresiones de calidad fotográfica de 20 × 25 cm en impresoras de inyección de tinta. Las cámaras digitales pueden usar zoom óptico tradicional para acercar los objetos; también pueden tener zoom digital que usa algoritmos de software para acercar una parte particular de la imagen digital, pero en el proceso se reduce la calidad de la imagen. Algunas cámaras digitales fijas pueden tomar películas cortas de resolución más bien baja.

Las videocámaras digitales son capaces de capturar videos de gran calidad, mientras que las webcams se usan para capturar imágenes en movimiento con el propósito de dar seguimiento, videoconferencias y aplicaciones similares en las que la calidad y suavidad del movimiento de la imagen no son cruciales.

Las fotografías y películas se almacenan en las computadoras en una diversidad de formatos estándar, usualmente en forma comprimida para reducir los requisitos de almacenamiento. Los ejemplos más comunes incluyen JPEG (por sus siglas en inglés, *Joint Photographic Experts Group* = Grupo conjunto de expertos fotográficos) y GIF (por sus siglas en inglés, *Graphic Interchange Format* = Formato de intercambio gráfico), para imágenes, y MPEG y Quick Time para películas.

Imágenes tridimensionales se pueden capturar para su proceso por computadora mediante escáneres de volumen. Un método consiste en proyectar una línea láser sobre el objeto 3D para digitalizarlo y capturar la imagen de la línea donde interseca el objeto mediante cámaras de alta resolución. A partir de imágenes capturadas y la información acerca de la posición y orientación de la cabeza que digitaliza, se calculan las coordenadas de superficie del objeto.

Ahora es frecuente el uso de sensores para proporcionar información acerca del ambiente y otras condiciones de interés para las computadoras. Por ejemplo, existen tableros de sensores en un automóvil y muchos miles de aquéllos en una planta industrial moderna. A continuación, se presenta una lista parcial de sensores usados comúnmente y sus aplicaciones.

- Las fotoceldas constituyen los *sensores de luz* más simples para controlar luces nocturnas, semáforos, sistemas de seguridad, cámaras y juguetes. Una fotocelda incorpora una resistencia variable que cambia con la luz (desde

muchos miles de ohms en la oscuridad hasta casi un kilohm en luz brillante), ello facilita detectar la cantidad de luz mediante un convertidor analógico-digital.

- Los sensores de temperatura son de dos tipos. Un detector de contacto mide su propia temperatura y deduce un objeto con el que está en contacto, si supone equilibrio térmico. La medición se basa en la variación de las propiedades del material (como resistencia eléctrica) con la temperatura. Los detectores de no contacto miden la radiación infrarroja u óptica que reciben de un objeto.
- Los detectores de presión convierten la deformación o alargamiento en materiales a cambios en algunas propiedades eléctricas mensurables. Por ejemplo, el alambre en zigzag incrustado en un sustrato plástico en un *extensímetro* debido a deformación, con lo que su resistencia cambia ligeramente. Los detectores de presión microelectromecánica ofrecen mayor sensibilidad y precisión. Estos detectores de presión se usan en tuberías, control de motores, alas de aviones etcétera.

Las innovaciones en el diseño de sensores son de inmenso interés como resultado de la creciente demanda en el control incrustado, los sistemas de seguridad y aplicaciones militares. Las nuevas tecnologías que se siguen incluyen sensores microelectromecánicos (MEM), particularmente para presión, y biosensores, que incorporan componentes biológicos, ya sea en los mecanismos usados para detectar o en el fenómeno detectado. Los sensores MEM ofrecen mayor sensibilidad y precisión, así como tamaño pequeño. Las ventajas de los biosensores incluyen bajo costo, mayor sensibilidad y economía de energía. Se usan en control de la contaminación, análisis bacterial, diagnóstico médico y minado, entre otras aplicaciones.

La salida de imagen mediante computadoras se produce al “renderizar” varios tipos de dispositivos de presentación visual o imprimir/graficar en papel. De manera adicional, las imágenes se pueden transferir directamente a microfilme para archivar o proyectar en una pantalla durante presentaciones audiovisuales. Lo último se hace usualmente al conectar un videoprojector al puerto de salida de pantalla de una

computadora de escritorio o laptop o a un puerto especial que ofrece salida para formato de TV. Las salidas gráficas más exóticas incluyen imágenes estereográficas para aplicaciones de realidad virtual e imágenes holográficas.

Trackball: es un dispositivo en el que el usuario debe rotar directamente una bola para mover el puntero en la dirección deseada. Es como un ratón mecánico panza arriba. También tiene una versión de sobremesa. Estos son los más comunes, pero existen otros dispositivos apuntadores cuyo uso se ha extendido sólo en ciertas áreas

Palanca de juegos (joystick): es una palanca cuya inclinación provoca el movimiento del objeto que se esté manejando en la pantalla. Soy muy habituales como controlador en videojuegos.

Tableta gráfica: son muy populares en diseño por ordenador, animación o dibujo. Son como un panel táctil pero en lugar del dedo debe emplearse un lápiz especial. Suele detectar distintos niveles de presión, que se traducen, por ejemplo, en un mayor o menor grosor del trazo

ACTIVIDADES

Teclados y sus interruptores

- a) Los interruptores mecánicos de algunos teclados parecen requerir gran cantidad de movimiento vertical para cerrar el contacto eléctrico. Si examina el teclado en una moderna computadora notebook ultradelgada, observará que las teclas se mueven muy poco. Investigue el diseño de tales interruptores de movimiento pequeño.
- b) Algunos teclados se publicitan como “a prueba de derrames”. ¿Cuáles son las implicaciones de esta propiedad sobre el diseño de interruptores y otras partes de un teclado?
- c) Mencione otras dos propiedades físicas de un teclado que considere importantes para un usuario.
- d) Mencione dos propiedades negativas de un teclado que conducirían a insatisfacción o rechazo del usuario.

Codificación de teclado

- a) La mayoría de los teclados soportan dobles presiones de teclas (por ejemplo, una tecla ordinaria más una de las teclas especiales shift, alt o control) para permitir un conjunto más amplio de símbolos. Especule acerca de cómo se logra la codificación en este caso.
- b) En máquinas basadas en Windows, al menos se soporta una triple presión de tecla (Alt + control + delete). ¿Puede pensar en una buena razón para incluir esta característica, que ciertamente complica el proceso de codificación?
- c) En las previsiones de las partes a) y b), múltiples presiones de tecla conducen a la transmisión de un símbolo del teclado a la computadora. Lo opuesto también puede ser cierto para algunos teclados: múltiples símbolos enviados como resultado de una sola presión de tecla. ¿Por qué es útil esta característica? Describa una forma de efectuar el codificador requerido.

Tecnologías de impresoras

Con base en la investigación que realice acerca de impresoras de inyección de tinta y láser, compare y contraste las dos tecnologías con respecto a los siguientes atributos:

- a) Calidad de salida en términos de resolución y contraste.
- b) Calidad de salida en términos de durabilidad (falta de borrado con el tiempo).
- c) Latencia y rendimiento total de la impresión.
- d) Costo de la tinta o tóner por página impresa.
- e) Costo total de propiedad por página impresa.
- f) Facilidad de uso, que incluye dimensiones físicas, ruido y generación de calor.

Dispositivos especiales de entrada

Investigue los siguientes temas en relación con los dispositivos especiales de entrada y escriba un informe de dos páginas acerca de cada uno.

- a) Por qué los números impresos al fondo de la mayoría de los cheques bancarios (incluidos el número ID del banco y el número de cuenta) tiene formas inusuales.
- b) Cómo el código de producto universal (UPC) o código de barras, cómo se leen mediante un escáner en el contador de salida.
- c) Cómo funcionan los lectores de tarjetas de crédito, que se encuentran en muchas tiendas, o cajeros automáticos.
- d) Cómo el texto escrito a mano, que se ingresa mediante un estilete, se percibe y acepta en las PDA o PC tablet.
- e) Cómo se comunican con la computadora los teclados, ratones y controles inalámbricos (para cambiar diapositivas en una presentación).

Bibliografía

- Parhami, B., (2007). *Arquitectura de computadoras. De los microprocesadores a las supercomputadoras*. México. McGraw-Hill.
- Daniel M. Argüello, Santiago C. Pérez e Higinio A. Facchini. (2018). *Arquitectura de computadoras*. Universidad Tecnológica Nacional – Facultad Regional Mendoza. E-book – 4° Edición. Argentina. ISBN: 978-950-42-0158-8.
- Quiroga, P. (2010). *Arquitectura de computadoras*. 1a ed. Buenos Aires: Alfaomega. 372 pp. ISBN 978-987-1609-06-2.
- Carrión González, A. N. (2020). Control del sistema de E/S: Sistemas Operativos. *Ciencia Huasteca Boletín Científico De La Escuela Superior De Huejutla*, 8(15), 19-25. <https://doi.org/10.29057/esh.v8i15.5254>.
- Stalling, W. (2006). *Organización y Arquitectura de Computadoras*. 7ª. Edición. España: Prentice Hall.
- Pedruelo, Rebollo. *Dispositivos de Entrada*. Universidad Politécnica de Valencia. España.
- Orenge, M., Manonellas, G. *Sistema de entrada/salida*. España. Universidad Oberta de Catalunya.