

Capítulo 2



La Motherboard y el Microprocesador

La Motherboard

La *motherboard* (placa madre o placa base) es una tarjeta en donde se ubican los componentes clave de una computadora. Contiene el microprocesador, la memoria y otros circuitos fundamentales para el funcionamiento de la PC. La misma fue diseñada de tal forma que las placas que se conectan con el mundo exterior sean circuitos adicionales y no estén integradas con la motherboard (es decir, que no se encuentren en la misma placa).

Estas tarjetas se conectan en ranuras de expansión, que pueden ser de diferentes tipos, y serán analizadas más adelante. La idea es muy buena, pues si se quiere actualizar alguna tarjeta que se comunique con el monitor para que soporte mayores resoluciones y dibuje la pantalla más rápidamente, no deberemos cambiar la motherboard completa por una nueva, sino que reemplazando simplemente la tarjeta de video, se actualiza el sistema.

Es importante aclarar que hoy en día, la mayoría de las motherboards integran las tarjetas controladoras de discos rígidos y los puertos y buses de E/S tradicionales en la misma placa del circuito, para ahorrar ranuras de expansión y dejar más de éstas libres para otra clase de tarjetas. En algunos casos, también incorporan a la tarjeta de video, la de red, la de sonido y el módem/fax.



La motherboard tiene todos los componentes necesarios para el procesamiento de datos, como ser el microprocesador, los buses de datos, control y direcciones, la memoria del sistema, el *chipset* (juego de chips) y el BIOS (*Basic Input Output System* - Sistema Básico de Entrada y Salida), que serán descriptos uno a uno a continuación.

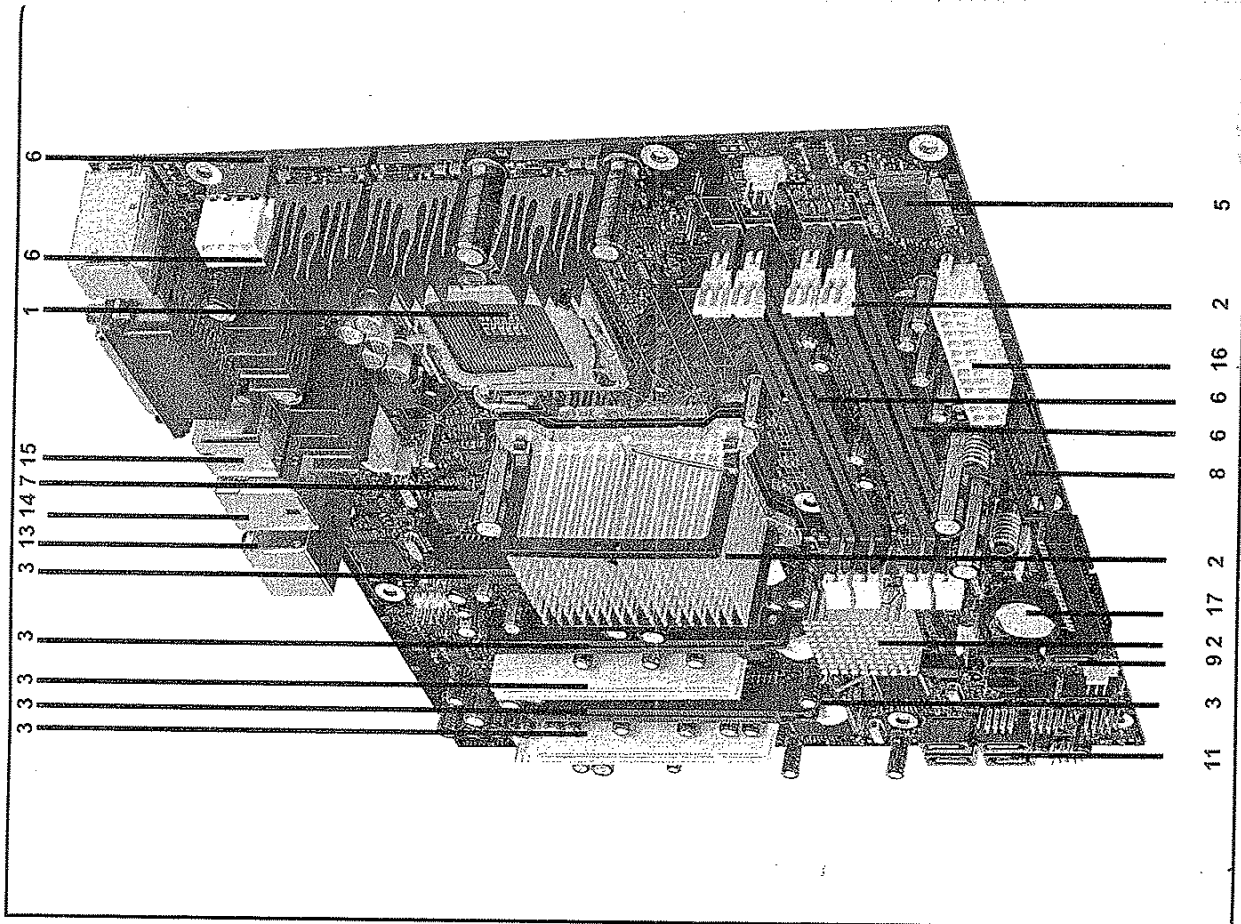


Fig. 2.1. Partes de una motherboard de última generación.

En las Figs. 2.1 y 2.2 se puede observar una motherboard, con sus diferentes componentes. Los números que se indican en ésta representan a los principales componentes de la misma, y son:

1. El zócalo para el microprocesador o procesador.
2. El *chipset* (juego de chips).
3. Las ranuras de expansión de los diferentes buses de expansión (PCI, PCI-X, PCI Express, AGP Pro, etc.).
4. Una ranura para WLAN (LAN inalámbrica) WiFi (en algunas motherboards).
5. El BIOS (*Flash*).
6. Los bancos de memoria.
7. La tarjeta de red LAN integrada (*LAN chip*).
8. Múltiples canales para discos rígidos E-IDE o dispositivos ATAPI. Las controladoras E-IDE/ATAPI están integradas en la motherboard.
9. Múltiples canales para discos rígidos Serial ATA. Las controladoras Serial ATA están integradas en la motherboard.
10. Múltiples canales con funciones RAID para discos rígidos E-IDE. Generalmente, utilizan conectores con un color diferente a los que no ofrecen funciones RAID.
11. Múltiples canales con funciones RAID para discos rígidos Serial ATA. Generalmente, utilizan conectores con un color diferente a los que no ofrecen funciones RAID.

RAID (Redundant Array of Inexpensive Disks - Arreglo redundante de discos de bajo costo) o bien (Redundant Array of Independent Disks - Arreglo redundante de discos independientes) es un estándar de diferentes configuraciones de conjuntos de discos rígidos, pensadas para conseguir mejores rendimientos, seguridad y/o escalabilidad. Muchas motherboards modernas incluyen controladoras adicionales a las convencionales que ofrecen soporte a diferentes configuraciones RAID. Para mayor información sobre las configuraciones RAID, puede consultar el libro *Servidores de Redes - Diseño, Actualización y Reparación*, de Editorial HASA.

12. Controladora de unidades de diskette (*FDD controller*).
13. Tarjeta de sonido integrada (*audio chip*).
14. Controlador IEEE 1394 (*FireWire*) con salida para dos puertos.
15. Controlador USB 2.0 con salida para cuatro u ocho puertos.

El Microprocesador

El microprocesador, al cual también lo llamaremos procesador, CPU (*Central Processing Unit* - Unidad Central de Procesamiento) o UCP, constituye el cerebro de una computadora. Éste se encarga de tomar la información que recibe de las diferentes fuentes, efectuar a dicha información los procesos necesarios y enviar el resultado al destino que se le indicó. A ese esquema se lo conoce con las siglas EPS (Entrada de información; Procesamiento; Salida) y lo podemos observar en la Fig. 2.3. Este diseño básico respeta las bases de la máquina de von Neumann, desarrollada por este destacado matemático en el año 1952.

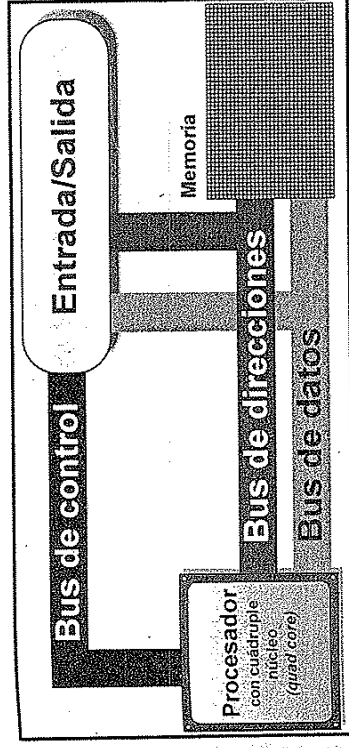


Fig. 2.3. Diagrama básico del funcionamiento de una computadora.

Características y Clasificaciones

El microprocesador puede dividirse en varios bloques o componentes, de los cuales vamos a mencionar los más importantes para analizar su funcionamiento:

- La CU o UC (Control Unit - Unidad de control).
- La ALU o UAL (Arithmetic and Logic Unit - Unidad aritmética y lógica).
- La FPU o UPF (Floating Point Unit - Unidad de Punto Flotante), también conocida como coprocesador matemático integrado.
- Los registros.
- Las memorias caché:
 - Nivel 1 ó L1 (antes conocida como interna).
 - Nivel 2 ó L2 (antes conocida como externa).
- Los buses internos y externos:
 - Bus de control (control bus).
 - Bus de direcciones (address bus).

26 - Estructura Interna de la PC - 5ta. Edición

16. Los conectores para la alimentación.
17. La batería del CMOS.
18. El conector del teclado. Ver la Fig. 2.2.
19. El conector del mouse. Ver la Fig. 2.2.
20. El o los puertos serie. Ver la Fig. 2.2.
21. El puerto paralelo. Ver la Fig. 2.2.
22. Los puertos USB 2.0. Ver la Fig. 2.2.
23. Los puertos IEEE 1394. Ver la Fig. 2.2.
24. Salida de audio S/PDIF óptica. Ver la Fig. 2.2.
25. Salida de audio S/PDIF vía coaxil. Ver la Fig. 2.2.
26. El conector RJ-45 para conexión a una LAN a través de un cable UTP. Ver la Fig. 2.2.
27. Salidas de audio para cinco o seis canales. Ver la Fig. 2.2.
28. Entrada de audio o para el micrófono. Ver la Fig. 2.2.

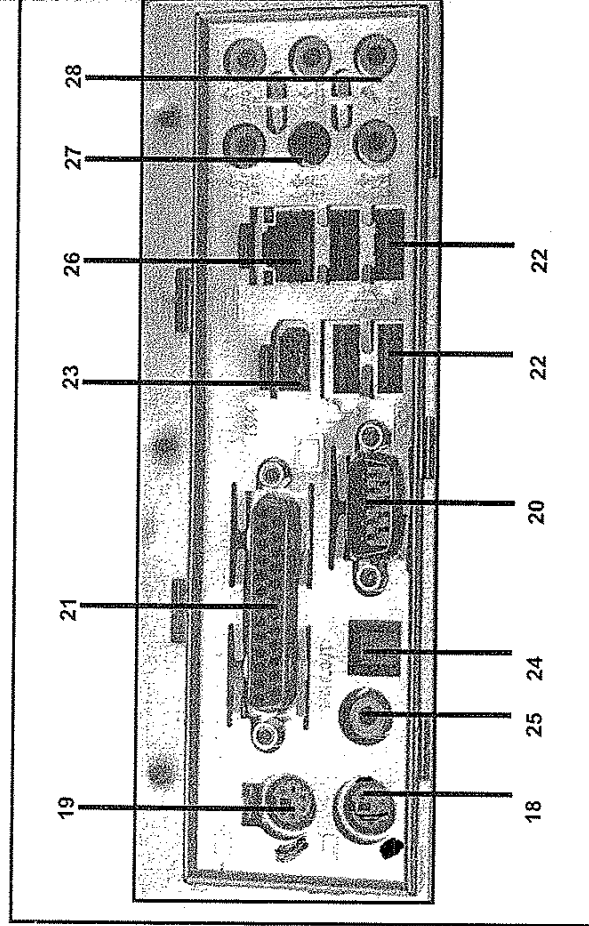


Fig. 2.2. Panel de conectores de una motherboard de última generación.

- Bus de datos (*data bus*).
- Bus hacia la caché de Nivel 1 ó L1.
- Bus hacia la caché de Nivel 2 ó L2.

Los microprocesadores que poseen múltiples núcleos (cores) de procesamiento en un único encapsulado tienen repetidos estos bloques o componentes cierta cantidad de veces, de acuerdo a su arquitectura interna y a la organización y comunicación de los diferentes núcleos de procesamiento.

A su vez, podemos catalogar y, por ende comparar, a los diferentes microprocesadores por las siguientes características que los distinguen:

- La cantidad de núcleos (*cores*) de procesamiento individuales o combinados que contienen.
- El tamaño de palabra (*word size*) o ancho máximo de las instrucciones o datos que puede interpretar en una única instrucción.
- La velocidad del reloj.
 - Interna: dentro del microprocesador.
 - Externa: del bus local, del sistema o FSB (*Front Side Bus* - Bus frontal).
- La arquitectura interna, también conocida como micro-arquitectura.
- La cantidad y organización de las unidades de ejecución.
- La cantidad y organización de canales o cañerías (*pipelines*) de instrucciones.
- Las capacidades y la eficiencia del procesamiento.
- El juego de instrucciones que soporta.
- Las instrucciones adicionales y específicas.
- La compatibilidad.
- El ancho y la velocidad de los buses internos y externos.
 - Bus de control.
 - Bus de direcciones.
 - Bus de datos.
 - Bus hacia la caché interna.
 - Bus hacia la caché externa.
 - Bus entre los diferentes núcleos de ejecución.

Y por último, tenemos las características físicas de un microprocesador, que también ocupan un rol muy importante:

- El embalaje (*package*) y sus dimensiones físicas.
- El zócalo (*socket; slot*).
- La tecnología y el proceso de fabricación (*manufacturing process*).
- La tensión de trabajo.

A continuación, vamos a analizar bien en detalle todas estas clasificaciones y características, además de comprender el funcionamiento básico de los microprocesadores.

Principios Básicos de Funcionamiento

Para comprender los principios básicos de funcionamiento, vamos a basarnos en un microprocesador con un único núcleo (*core*) de procesamiento en su encapsulado. El microprocesador posee una cierta cantidad de pequeñas memorias temporales de rápido acceso, conocidas como **registros**. En ellas almacena algunos datos necesarios y otros fundamentales para continuar procesando instrucciones sin perderse. Un ejemplo de registros del microprocesador es el registro CP (Contador de Programa), que contiene la dirección de la próxima instrucción a ejecutar. Como los registros están dentro del microprocesador, éste puede acceder al contenido de los mismos muchísimo más rápido que si lo hiciera a la memoria principal. La cantidad de registros que tiene el microprocesador depende exclusivamente de su diseño. Explicar el funcionamiento de cada uno de los registros justificaría una obra orientada exclusivamente a los microprocesadores, por lo tanto, nos limitaremos a las definiciones anteriores.

El tamaño de palabra (*word-size*), o bien el ancho máximo de las instrucciones, o los datos que puede interpretar el microprocesador en una única operación, se mide en bits y puede ser de 2^n bits: 8; 16; 32; 64 bits y así sucesivamente. Como vimos anteriormente es uno de los ítems más importantes para clasificar a los microprocesadores.



Un microprocesador de 64 bits puede procesar un dato o instrucción de 64 bits como máximo en una sola operación.

A medida que fue aumentando el tamaño de palabra, los procesadores tuvieron que agregar nuevos registros, los cuales deben respetar el tamaño de palabra. Por ejemplo, un procesador de 64 bits posee el conjunto de registros de 32 bits para mantener la compatibilidad con sus predecesores y agrega un nuevo conjunto de registros de 64 bits. Con la evolución de los procesadores, también se agregaron registros específicos con mayor tamaño que el de palabra.

La familia de microprocesadores que utilizan las PC se conoce como familia 80x86, debido a su compatibilidad hacia atrás con el antiguo procesador 8086 de Intel.

Los microprocesadores pertenecientes a esta familia tienen la capacidad de manejar una gran cantidad de instrucciones de longitud variable y se conocen como procesadores CISC (Complex Instruction Set Code - Código de conjuntos de instrucciones complejas).

Los procesadores RISC (Reduced Instruction Set Code - Código de conjuntos de instrucciones reducido) tienen una cantidad de instrucciones menor que 100 y todas son de la misma longitud (la misma cantidad de bits), por lo tanto, estos procesadores son más rápidos.

Los procesadores más avanzados de las PC compatibles logran mayores velocidades combinando las dos tecnologías, es decir, transforman las instrucciones CISC en instrucciones RISC simples de longitud fija y se ejecutan más rápidamente mediante múltiples unidades de ejecución RISC. Entonces, la micro-arquitectura de estos procesadores es una mezcla de CISC con unidades RISC.

El funcionamiento básico del microprocesador para la ejecución de instrucciones es el siguiente: la unidad de control se encarga de recibir las instrucciones provenientes de la memoria a través del bus de datos, decodifica cada instrucción y la ejecuta, enviando datos a cualquier dispositivo que la instrucción indique o realizando diferentes tipos de procesos con los datos que llegan de la memoria a través del bus de datos. Si no se le indica lo contrario mediante una instrucción de bifurcación, el procesador ejecutará las instrucciones a medida que le van llegando. El anexo del bus de datos delimita la cantidad de información que puede recibir el procesador en cada ciclo de reloj. El mismo puede ser diferente para el interior del procesador y el exterior. Si el exterior es menor al interior, se necesitan dos ciclos para alimentar al interior, por lo tanto, el rendimiento del procesador es mucho menor. Muchas versiones económicas de los procesadores recurrieron a esta técnica para abaratar costos.

La unidad aritmética y lógica (ALU) es una parte del procesador que se encarga de realizar las operaciones lógicas y aritméticas con los datos recibidos. Si la instrucción indica que se realicen algunas de estas operaciones la ALU tomará los datos de registros, los cuales se han llenado previamente con otras instrucciones para tal fin, y depositará los resultados de la operación efectuada en un registro. El microprocesador está gobernado por el reloj del sistema, un cristal de cuarzo regulado que cumple el papel de un metrónomo electrónico. A mayor velocidad del reloj, más operaciones se podrán realizar en el mismo período de tiempo.

La velocidad de este reloj se mide en Hertz (Hz), que indican la cantidad de ciclos por segundo. Hace algunos años se expresaban en MegaHertz (MHz), donde 1 MHz = 1 millón de ciclos por segundo. En los procesadores modernos se miden en GigaHertz (GHz), donde 1 GHz = 1.000 millones de ciclos por segundo.

Se puede deducir entonces que un microprocesador funcionando a 2 GHz puede ejecutar el doble de instrucciones en un segundo, que uno de 1 GHz (2.000 millones de instrucciones contra 1.000 millones). Pero, esto no es así en todos los microprocesadores, pues algunos tienen una velocidad de reloj menor pero pueden ejecutar mayor cantidad de instrucciones en un solo ciclo de reloj. La deducción anterior se limita a microprocesadores de características equivalentes pero diferente velocidad, es decir, en aquéllos en los cuales siendo todas las características que hemos descrito anteriormente iguales, solamente varíe la velocidad de reloj.

Los principales fabricantes de microprocesadores para PC son Intel y AMD (Advanced Micro Devices), sin embargo VIA (ex Cyrix) e IBM también producen procesadores compatibles.



Evolución de los Microprocesadores

A continuación, analizaremos todas las características de los principales microprocesadores del mundo de las PC: desde el 8086 hasta el Core 2 Extreme, pasando por el Pentium, las tecnologías MMX e Internet SSE 1; 2 y 3, por un lado, y 3DNow junto a Enhanced 3DNow! y 3DNow! Professional, por el otro (desarrolladas por Intel y AMD, respectivamente), el Pentium D; 4; III y II, y el Athlon 64 X2; 64 y XP, entre otros. Aunque, primero vamos a hacer una recorrida por su evolución y veremos en detalle a sus principales predecesores, quienes, a pesar de su antigüedad, siguen constituyendo las bases del funcionamiento de los más modernos. En la Tabla 2.1 podemos ver la evolución de los principales microprocesadores de Intel para PC.

A medida que fueron evolucionando los microprocesadores, se consiguieron mejores tecnologías de fabricación, de menor cantidad de micrones que permiten condensar mayor cantidad de transistores en el mismo espacio de silicio, el cual se encuentra dentro del embalaje del procesador.

Los transistores constituyen el componente principal de la electrónica de los microprocesadores y a mayor cantidad, éstos pueden ofrecer mayores funcionalidades y ser más complejos y conseguir, de esta manera, velocidades de trabajo superiores así como mayor cantidad de núcleos de ejecución en el mismo espacio físico.

El Pentium MMX de 233 MHz, que utilizaba una tecnología de 0,35 micrones, tenía 4,5 millones de transistores en un trozo de silicio escondido dentro del chip, mientras que el Core 2 Extreme, el cual tiene cuatro núcleos de ejecución (quad

core), utiliza una tecnología de 0,065 micrones, tiene 582 millones de transistores (aumentó más de 129 veces con respecto al anterior).

Pentium D de primera generación y al que se fabrica en 0,065 micrones se lo denomina de segunda generación. Esta nomenclatura es aplicable a todos los procesadores y a las futuras versiones. Muchas veces, las siguientes generaciones de un mismo procesador suelen agregar varias mejoras a las anteriores.

Tabla 2.1. Evolución de los principales microprocesadores de Intel para PC.

Fecha de lanzamiento	Versión original	Versión de bajo costo	Tecnología de fabricación (en micrones)	Cantidad de transistores (en millones)
Junio 1978	8086		3	0,02
Junio 1979	8088		3	0,02
Febrero 1982	80286		1,5	0,13
Octubre 1985	80386 (DX) / 80387 (SX)	(SX)	1,5 y 1	0,27
Abril 1989	80486 (DX, DX2 y DX4)	(SX)	1; 0,8 y 0,6	1,2 a 1,6
Marzo 1993	Pentium			
Marzo 1994	Pentium P54C		0,8; 0,6 y 0,35	3,1 a 4,5
Enero 1997	Pentium MMX			
Noviembre 1995	Pentium Pro		0,6 y 0,35	5,5
Mayo 1997	Pentium II	Celeron Celeron A*	0,35 y 0,25	7,5
Febrero 1999	Pentium III	Celeron	0,25 y 0,18	9,5 a 20
Noviembre 2000	Pentium 4	Pentium 4 Celeron	0,18 y 0,13	42 a 55
Noviembre 2002	Pentium 4 (Hyper-Threading)	Pentium 4 Celeron Celeron D	0,13; 0,09 y 0,065	55
Abril 2005	Pentium Processor Extreme Edition		0,09 y 0,065	230
Mayo 2005	Pentium D		0,09 y 0,065	376
Enero 2006	Core Duo	Core Solo	0,065	151
Julio 2006	Core 2 Duo		0,065	291
Noviembre 2006	Core 2 Extreme		0,065	582

Como podemos ver en la Tabla 2.1, hay varios procesadores que figuran con diferentes tecnologías de fabricación. Cuando un procesador que utiliza una tecnología determinada se empieza a fabricar con otra mejor, cambia de generación. Es por ello que al Pentium D de 0,09 micrones se lo conoce como

Embalajes y Zócalos para el Microprocesador

El trozo de silicio que esconde el embalaje de un microprocesador moderno suele ocupar menos que la yema de un dedo (ver la Fig. 2.4). Sin embargo, el embalaje para todas las señales que entran y salen del procesador. Además, tiene una función adicional muy importante: proveer la superficie para algún mecanismo de disipación de calor.

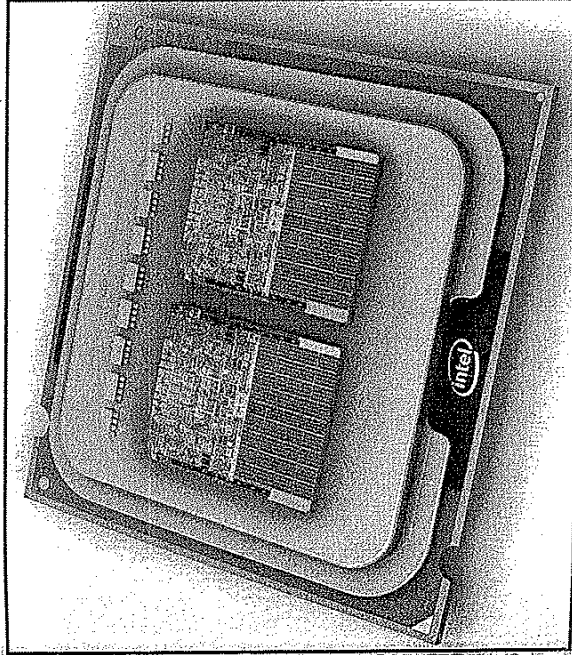


Fig. 2.3. Ilustración del trozo de silicio que conforma el Core 2 Extreme, con tecnología de 0,065 micrones y cuatro núcleos de ejecución (quad-core), en su embalaje (Cortesía de Intel).

En la Tabla 2.2 resumimos los diferentes zócalos para los diversos microprocesadores con los que nos podemos encontrar. Todos estos zócalos, cuyo nombre comienza con *Socket* o *Super* son del tipo SPGA, CPGA o CSPGA (*Ceramic Staggered Pin Grid Array* - Arreglo en grilla de pines en forma Escalonada de Cerámica) o PPGA (*Plastic Staggered Pin Grid Array* - Arreglo en Grilla de Pines en Forma Escalonada de Plástico).

A diferencia de los anteriores, en los zócalos LGA (*Land Grid Array / Leadless Grid Array* - Arreglo en grilla plano / sin pines), el microprocesador no tiene pines que se conectan en el zócalo para establecer los contactos con las pistas del circuito impreso de la motherboard. En cambio, tiene *pads* planos que se montan sobre una grilla de *pads* planos y así hace contacto, por lo tanto, es de montaje superficial. Ver la Fig. 2.5.

Por otro lado, comienzan con *Slot* o con PAC los del tipo SEC (*Single Edge Contact* - Contacto de lado único). Sin embargo, esta clase de zócalos ha quedado obsoleta y solamente se sigue utilizando en algunos procesadores para servidores.

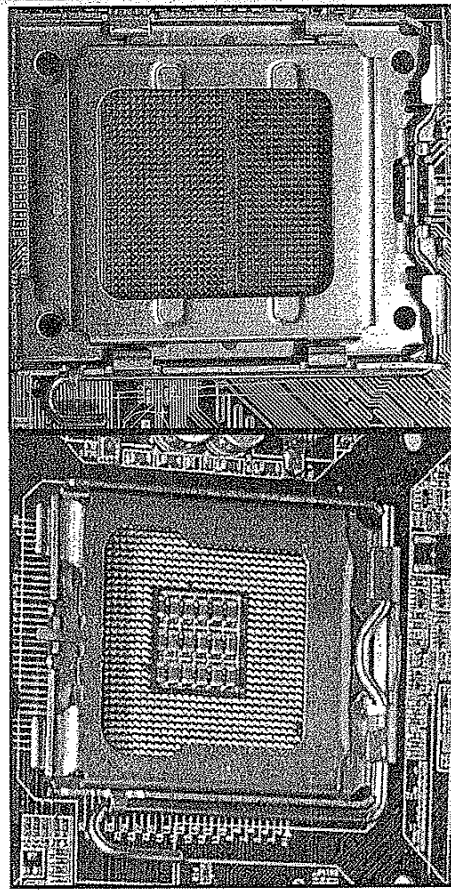


Fig. 2.5. Un zócalo Socket 775 (LGA775) (izquierda) y un zócalo Socket F (LGA1207FX) (derecha), ambos del tipo LGA.

La mayoría de las motherboards modernas que ofrecen zócalos del tipo SPGA o PPGA son del tipo ZIF (*Zero Insertion Force* - Fuerza de Colocación Cero), que permiten cambiar fácilmente un chip de procesador moviendo la traba que lo retiene o libera. Los zócalos del tipo SEC proveen el mecanismo más sencillo para cambiar un microprocesador, ya que se insertan y retiran fácilmente de la ranura correspondiente.

Tamaño de Palabra, Buses de Direcciones y de Datos

Como vimos anteriormente, el tamaño de palabra constituye uno de los elementos más importantes a la hora de clasificar a los microprocesadores. La mayoría de las PC modernas trabajan con procesadores con un tamaño de palabra de 32 ó 64 bits,

a los cuales se los denomina microprocesadores de 32 y 64 bits, respectivamente. El primer procesador de 32 bits de la familia 80x86 fue el 80386DX de Intel.

Tabla 2.2. Zócalos para los diferentes microprocesadores.

Nombre del zócalo	Cantidad de pines	Microprocesadores
Slot 1	242	Pentium II; Pentium III (1ª generación) y Celeron
Slot A	337	Athlon (K7) (1ª generación) y Thunderbird (K7) (1ª generación)
Socket 1	168/169	486SX; 486DX; 486DX2 OverDrive y 486DX4 OverDrive
Socket 2	238	486SX; 486DX; 486DX2 OverDrive; 486DX4 OverDrive; 5x86 y Pentium OverDrive
Socket 3	237	486SX; 486DX; 486DX2; 486DX4 y 5x86
Socket 370/PPGA	370	Celeron "A"
Socket 370 FC-PGA (Flat Chip)	370	Pentium III; Celeron "A" y Pentium III Xeon
Socket 4	273	Pentium (60 y 66 MHz)
Socket 423 (PGA 423)	423	Pentium 4 (1ª generación)
Socket 478 (mPGA78B)	478	Pentium 4 (2ª generación); Pentium 4 Hyper-Threading; Pentium 4; Celeron y Xeon
Socket 5	320	Pentium P54C (75; 90; 100; 120; 133; 150; 166 y 180 MHz); Pentium MMX OverDrive; 6x86 y MediaGX
Socket 6	235	486DX4 y 5x86
Socket 603 (mPGA603)	603	Xeon Pentium 4; Xeon Hyper-Threading; Xeon DP y MP; y Xeon Dual-Core
Socket 604 (mPGA604)	604	
Socket 7	321	Pentium MMX (P55C); 6x86MX (MII); hasta 188 MHz; K6; y K6-2
Socket 754	754	Athlon 64; Sempron
Socket 775 (LGA775)	775	Pentium 4 (3ª generación; 90 nm); Celeron D; Pentium 4 Extreme Edition; Pentium Processor Extreme Edition; Pentium D Core 2 Duo; Core 2 Extreme
Socket 8	387	Pentium Pro y Pentium II OverDrive (para Pentium Pro)
Socket 939	939	Athlon 64; Athlon 64 FX y Athlon 64 X2
Socket 940 (mPGA940)	940	Athlon 64 FX; Opteron 100; 200 y 800; y Opteron Dual-Core
Socket A (Socket 462)	337 462	Athlon (K7); Duron (K7); Thunderbird (K7); Athlon XP (Thoroughbred) (1ª generación); Athlon XP (Palomino) (2ª generación); Athlon XP (Bariton) (3ª generación); Athlon MP y Geode
Socket AM2 (Socket M2)	940	Athlon 64 X2; Athlon 64 FX y Sempron (2ª generación)
Socket F (LGA1207FX)	1207	Athlon 64 FX
Super Socket 7 (Super 7)	321	K6-2 y 6x86MX (MII) > 188 MHz

Los buses internos del microprocesador permiten que sus diferentes componentes envíen y reciban información. Trabajan a una velocidad determinada y tienen un ancho de datos, los cuales se miden en Hz y bits, respectivamente.

El ancho del bus de direcciones de un microprocesador determina la cantidad máxima de memoria a la cual se puede acceder (direccionar). Imaginemos a la memoria como un gigantesco tablero, con miles de casilleros, los cuales se encuentran numerados. Si queremos indicar que se cambie el valor de un casillero x, necesitamos identificarlo por su número. El ancho del bus de direcciones determina el máximo número de casilleros a los que se puede llegar a identificar. Como este valor se expresa en bits, elevando 2 al número de bits que forman el ancho del bus, obtenemos la máxima cantidad de memoria en bytes a la cual se puede acceder (direccionar), debido a que cada casillero representa 8 bits (un byte). El 8086 y el 8088 tenían un bus de direcciones de 20 bits, por lo tanto podían direccionar un máximo de $2^{20} = 1.048.576$ bytes (1024 KB, es decir, 1 MB), ver la Fig. 2.6. El 80286, el 80386, el Pentium Pro y los procesadores AMD de 64 bits, junto con los Intel con extensiones EM64T fueron aumentando el ancho de este bus a 24; 32; 36 y 40 bits, respectivamente (ver la Tabla 2.3). Este tema lo analizaremos más adelante en el **Capítulo 4: La Memoria**.

Tabla 2.3. Comparación de las capacidades de direccionamiento de memoria según el ancho del bus de direcciones.

Ancho del bus de direcciones (bits)	Máximo direccionamiento de memoria
20	1.024 KB (1 MB)
24	16.384 KB (16 MB)
32	4.096 MB (4 GB)
36	65.536 MB (64 GB)
40	1.024 GB (1 TB)
44	16.384 GB
50	1.048.576 GB (1024 TB)

Con el 80386DX, el primer procesador con un tamaño de palabra de 32 bits de la familia 80x86, apareció un nuevo concepto que sería heredado por todos sus sucesores: el de la **memoria virtual**. Cualquier dispositivo de entrada y salida (por ejemplo, un disco rígido) puede tener asignado un rango de direcciones de memoria y trabajar como ésta, de manera que si se posee memoria insuficiente, es posible recurrir al disco y utilizarlo de la misma forma que la memoria, aunque obviamente, no se tendrán las mismas velocidades conseguidas con la memoria física.

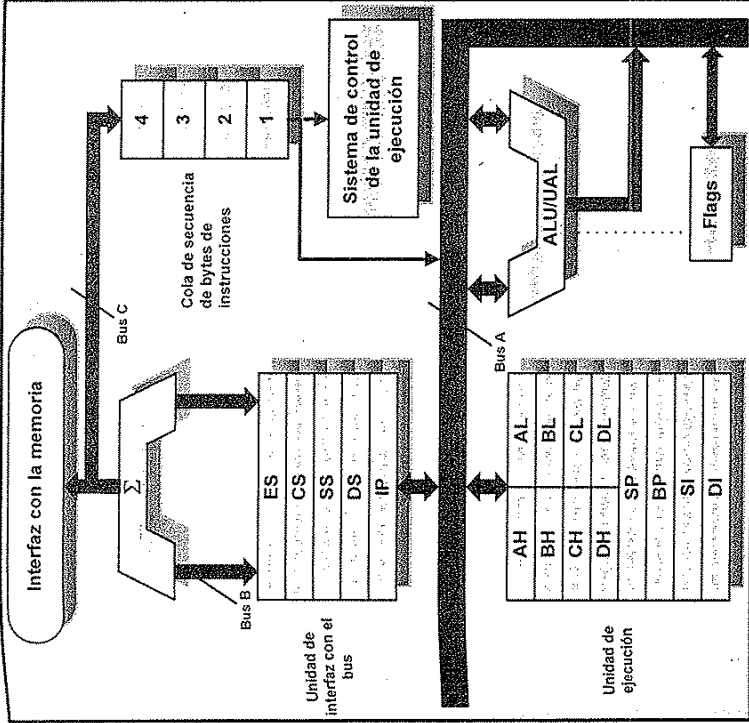


Fig. 2.6. Diagrama en bloques del prehistórico y muy simple microprocesador 8088.

Esta capacidad es propia del procesador y no del sistema operativo, como se suele pensar. Si el sistema operativo la aprovecha o no, es otra cosa. Además del ancho del bus de datos, hay que tener en cuenta la velocidad interna con la cual trabaja este bus, pues es un determinante importante del rendimiento general del procesador.

Los Procesadores de 64 bits: AMD64 y EM64T

Los microprocesadores con tamaño de palabra de 64 bits pueden procesar un dato o instrucción de 64 bits como máximo en una sola operación, lo cual implica exactamente el doble que los de 32 bits.

Hace muchos años, Intel había desarrollado los microprocesadores Itanium, que usaban un juego de instrucciones de 64 bits conocido como IA-64. Sin embargo, estos procesadores se utilizaron en servidores de alto rendimiento, pero debido a sus altísimos precios, nunca llegaron a formar parte de las PC hogareñas o de oficina.

Por su parte, AMD desarrolló un juego de instrucciones de 64 bits denominado AMD64, también conocido como x86-64, y lo incorporó en los microprocesadores Athlon 64. Como tuvieron un gran éxito, este juego de instrucciones se transformó en el estándar de 64 bits para las PC hogareñas y de oficina, y también se extendió a estaciones de trabajo de alto rango y servidores. Todos los microprocesadores modernos de AMD ofrecen soporte a este juego de instrucciones de 64 bits.

Luego, Intel incorporó extensiones de 64 bits a su línea de microprocesadores para que fueran compatibles con AMD64 y no se quedara fuera del mercado masivo de los 64 bits. A éstas las denominó Intel EM64T (*Extended Memory 64 Technology*). Tecnología de memoria extendida 64). Todos los microprocesadores modernos de Intel ofrecen soporte a este juego de instrucciones de 64 bits.

La gran ventaja de estos microprocesadores con los que están basados en IA-64 es que pueden ejecutar código de 32 bits cuando están trabajando en modo de 64 bits por lo cual, también ofrecen un excelente rendimiento con las aplicaciones de 32 bits, que actualmente constituyen la mayoría del mercado. Esto fue uno de los puntos que llevó al éxito a estos microprocesadores, además de sus costos accesibles.

Tanto los procesadores de AMD como los de Intel, que ofrecen x86-64 y EM64T, incrementaron el bus de direcciones a 40 bits, por lo cual, pueden direccionar un máximo de 1.024 GB (1 TB) de memoria (ver la Tabla 2.3). Más que suficiente para las aplicaciones más exigentes.

Sin embargo, como sucedió hace unas décadas cuando se iniciaba la guerra de los 16 bits versus los 32 bits, las aplicaciones desarrolladas para correr sobre los procesadores de 32 bits no tendrán ventaja alguna al ejecutarse con los de 64 bits. Esto se debe a que el software debe estar programado con instrucciones de 64 bits para sacar provecho de sus características especiales, lo cual tiene la gran desventaja de hacerlo totalmente incompatible con los procesadores de 32 bits. En la actualidad, todavía la gran mayoría de las aplicaciones están preparadas para el juego de instrucciones de 32 bits y no sacan provecho de estos juegos de instrucciones de 64 bits.

Sin embargo, tanto Windows como Linux tienen versiones de 64 bits para x86-64 y EM64T, por lo cual, si bien todavía el software y las aplicaciones para aprovechar estas arquitecturas están apareciendo a cuentagotas, ya están dadas las bases para comenzar a sacar el jugo de estos juegos de instrucciones.

Vamos a tomar un ejemplo sencillo para comprender la diferencia entre un tamaño de palabra de 64 bits y otro de 32 bits.

Se quiere copiar una cadena de texto de 1 MB (1.024 bytes) de memoria de una posición determinada a otra.

Utilizando el juego de instrucciones IA-32, con un tamaño de palabra de 32 bits, se pueden copiar 32 bits por instrucción desde memoria a un registro y con otra los 32 bits desde un registro a memoria. Por lo tanto, se necesitan 2 instrucciones por cada 32 bits ($32 \text{ bits} / 8 = 4 \text{ bytes}$) de memoria a copiar.

Para copiar 1 MB (1.024 bytes), el procesador necesitará ejecutar: $(1.024 \text{ bytes} \times 2) / 4 \text{ bytes por instrucción} = 512 \text{ instrucciones}$

En cambio, si se utiliza el juego de instrucciones x86-64, con un tamaño de palabra de 64 bits, se pueden copiar 64 bits por instrucción desde memoria a un registro y con otra los 64 bits desde un registro a memoria. Se siguen necesitando 2 instrucciones, pero logran movilizar el doble de la memoria, 64 bits ($64 \text{ bits} / 8 = 8 \text{ bytes}$).

En este caso, para copiar 1.024 bytes, el procesador necesitará ejecutar:

$(1.024 \text{ bytes} \times 2) / 8 \text{ bytes por instrucción} = 256 \text{ instrucciones}$

Exactamente la mitad del caso anterior.

Las Memorias Caché Incorporadas

El 80386DX fue el primer microprocesador de PC en utilizar en la motherboard una memoria caché externa, con el objetivo de aumentar el rendimiento y solucionar el problema de la lentitud de las memorias con respecto a la velocidad del procesador.

El 80486DX fue más allá e integró una memoria caché y su controladora dentro del procesador, que se conoce como memoria caché interna, de nivel 1 ó L1 y de ahí en más todos sus sucesores de Intel, AMD y Via fueron integrando y mejorando diferentes niveles de memorias caché dentro del procesador.

Con cada nuevo procesador, se fueron mejorando los buses que se utilizan para que las distintas unidades y los registros del procesador se comuniquen con las diferentes memorias caché. Se fue duplicando el ancho de los mismos, se aumentaron sus velocidades de reloj y se dividieron las memorias caché en bloques separados para que guarden diferentes clases de datos y se puedan acceder en forma simultánea por canales independientes (caché de instrucciones y de datos).

En algunos casos, las memorias caché están incorporadas en el mismo embalaje del procesador, pero funcionan a menores velocidades que la del núcleo principal. Mientras que en otros, forman parte del mismo trozo de silicio y están

perfectamente sincronizadas con el núcleo, consiguiendo de esta manera el mejor rendimiento. La desventaja de la segunda opción es su mayor costo.

Estas diferencias en las memorias caché tienen una gran incidencia en el costo de fabricación de los microprocesadores de última generación y afectan en un gran porcentaje al precio de venta de los mismos. Las versiones más económicas de los

microprocesadores suelen reducir el tamaño y la velocidad de las memorias caché que incorporan.

En el **Capítulo 4: La Memoria**, vamos a analizar detalladamente la memoria principal del sistema, junto con las diferentes memorias caché y todas sus variantes, pues su funcionamiento tiene mucho que ver con los principios de trabajo de las memorias. También en este capítulo vamos a ofrecer una comparación muy detallada de las memorias caché que utilizan los microprocesadores modernos y sus diferentes arquitecturas y organizaciones.

La Unidad de Punto Flotante

Los primeros microprocesadores no incluían un juego de instrucciones capaz de realizar operaciones con números de punto flotante (números fraccionarios o bien números con decimales). Como podemos ver en el esquema del 8088 (ver la Fig. 2.6), la única unidad de ejecución es la UAL, la cual solamente es capaz de realizar operaciones con números enteros.

Se pueden realizar operaciones aritméticas (sumas, restas, multiplicaciones y divisiones) con números de punto flotante utilizando instrucciones de la UAL para números enteros. Sin embargo, se requiere ejecutar un gran número de operaciones consecutivas para lograr, por ejemplo, multiplicar dos números de punto flotante, lo cual toma bastante tiempo del procesador. Esto se debe simplemente a que la UAL es especialista en números enteros.

Para solucionar este cuello de botella, existían los coprocesadores matemáticos, los cuales eran microprocesadores que se podían conectar en un zócalo disponible en motherboard para comunicarse con el procesador principal. Estos incluían una unidad de punto flotante, conocida como UFP o FPU (*Floating Point Unit*), la cual ofrece un conjunto de instrucciones específicamente diseñadas para realizar operaciones con números de punto flotante. De esta manera, con una única instrucción se consigue lo que antes mantenía a la UAL ocupada por un gran conjunto de instrucciones.

Un ejemplo es el 80387, que era el coprocesador matemático para el 80386DX.

Al derivarse estos cálculos pesados hacia la UFP, el procesador aumenta el rendimiento global del sistema de una manera apreciable, especialmente si las aplicaciones que se utilizan efectúan cálculos como tangentes, raíces, potencias exponenciales, etc.

Los programas de tratamiento de imágenes, autoedición y dibujo vectorial, los juegos y aplicaciones con animaciones en 2 y 3 dimensiones en tiempo real y las planillas de cálculo, entre otra gran cantidad de aplicaciones, hacen un uso intensivo de la UFP. Sus continuas operaciones matemáticas, basadas en números de punto flotante, requieren la ejecución de cálculos muy complejos.

El microprocesador 80486DX fue el primero en incluir una UFP en su arquitectura, es decir, dentro de la misma oblea de silicio. Esto hizo que el rendimiento fuera muy superior al de un coprocesador matemático externo, pues la UFP trabaja y se comunica a la misma velocidad interna del procesador principal y, además, aprovecha todas las ventajas de su microarquitectura.

A partir de esta integración de la UFP en el 80486DX, todos los procesadores que lo sucedieron tanto de Intel, como de AMD y Via, incluyeron una y luego múltiples UFP, rediseñadas y optimizadas para ser cada vez más rápidas. En la actualidad no existen procesadores que no incluyan varias UFP.

Procesadores Superescalares

Un microprocesador es superescalar cuando puede ejecutar más de una instrucción por ciclo de reloj. El Pentium de Intel fue el primero superescalar de la familia 80x86, pues podía procesar dos instrucciones en un mismo ciclo, utilizando dos canales para la ejecución de las mismas en paralelo. Todos sus sucesores tanto de Intel como de AMD, son superescalares y fueron aumentando la cantidad de instrucciones que son capaces de ejecutar en un ciclo de reloj en sus diferentes unidades de ejecución: UAL, UFP, etc.

Para tomar ventaja de los procesadores superescalares, las aplicaciones deben estar programadas para aprovechar las capacidades ofrecidas por los mismos. No todos los programas desarrollados aumentan su rendimiento drásticamente al ejecutarse en un procesador con mayores unidades de ejecución que otro, si no está preparado para sacar provecho a las ventajas adicionales.

Pipelines: Los Canales de los Procesadores

Los antiguos procesadores, como el 80486DX, recibían una instrucción por vez, la ejecutaban y guardaban los resultados en los registros correspondientes del procesador, como ya hemos descrito anteriormente. A este esquema básico se lo conoce como de un único canal, cañería o *pipeline*.

Los procesadores superescalares poseen obviamente, al menos dos canales, y para conseguir aumentar su rendimiento fueron mejorando el funcionamiento de estos canales para permitir la ejecución de las instrucciones en paralelo y en forma más eficiente.

Vamos a tomar como ejemplo el pionero en la familia 80x86: el Pentium, el cual sentó las bases para todos sus sucesores. Este procesador decodifica parcialmente la primera instrucción y determina si se puede ejecutar en paralelo con la siguiente o no. Por ejemplo: si la primera instrucción es sumar $C = A + B$ y la segunda es $D = B + C$, dichas instrucciones no pueden ser ejecutadas en paralelo, porque para

ejecutar la segunda instrucción hay que saber el valor de C. Si no sucede lo anteriormente descrito, se ejecutan las instrucciones por canales separados. Los dos canales, cañerías o *pipelines* paralelos cumplen los cinco pasos (ver la Fig. 2.7):

1. **Prebúsqueda.** Tomar la siguiente instrucción de la memoria caché o bien de la principal.
2. **Decodificación de la instrucción.**
3. **Generación de direcciones de memoria,** en dónde se ubican él o los datos (parámetros) que necesita la instrucción para ejecutarse, trayéndolos a registros.
4. **Ejecución.**
5. **Reescritura.** Guardar los resultados en los registros correspondientes.

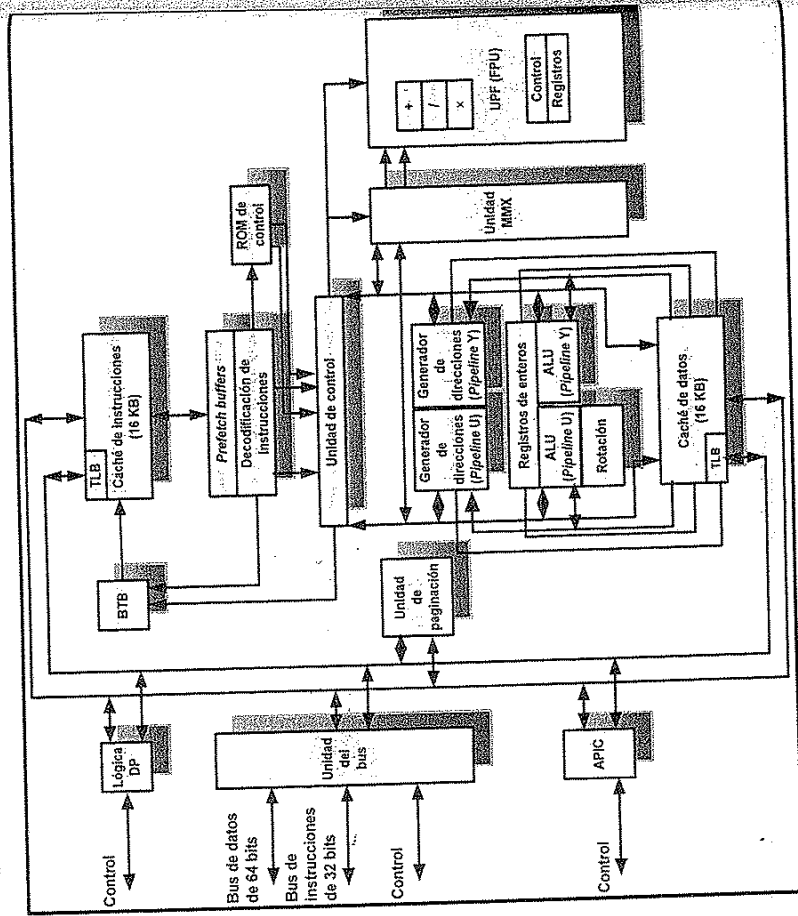


Fig. 2.7. Diagrama en bloques del Pentium MMX con sus dos canales: Pipelines U y V.

Si a estos cinco pasos se los ejecuta en forma secuencial y recién se vuelve a tomar una nueva instrucción una vez finalizado el quinto y último paso del canal, tendremos un esquema como el de la Fig. 2.8.

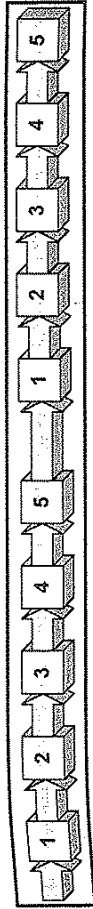


Fig. 2.8. Ejecución de cinco pasos en forma secuencial pura.

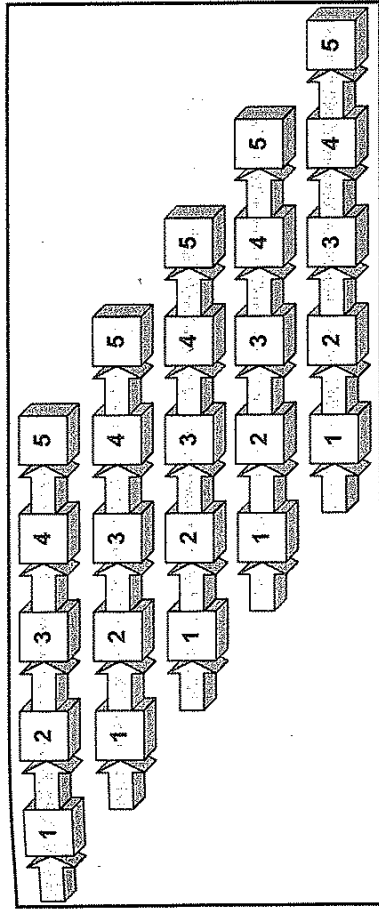


Fig. 2.9. Ejecución de cinco pasos con un esquema de pipelining (cascada).

Cuando una instrucción pasa de la prebúsqueda (paso 1) a la decodificación de instrucciones (paso 2), el canal ya está libre para ejecutar este primer paso en una nueva instrucción, pues el control de la instrucción anterior ya pasó al segundo elemento de la cadena de cinco pasos del canal. Esto sucede con todos los eslabones de esta cadena y, por lo tanto, aplicando este esquema conocido con el nombre de *pipelining*, llegamos a una ejecución en cascada como se muestra en la Fig. 2.9.

El pipelining reduce los tiempos totales necesarios para ejecutar un conjunto determinado de instrucciones y, por lo tanto, aumenta en gran proporción el rendimiento del canal con respecto a una ejecución secuencial pura.

El Pentium utiliza el esquema de *pipelining* para ambos canales, pudiendo además cada uno de ellos actuar de manera independiente. Cada uno tiene su propia unidad aritmético-lógica (ALU o UAL), circuitos propios de generación de direcciones y buses independientes con las memorias caché.

Otra de las mejoras introducidas en este procesador es el BTB (*Branch Target Buffer* - Buffer de Destino de Bifurcación), que es una memoria temporal (*buffer*) inteligente que intenta **predecir** cuándo se bifurcará un proceso, es decir, cuándo una instrucción salta a otro procedimiento ubicado en otra posición de memoria no consecutiva a las cuales se están procesando. La cantidad de entradas del BTB han aumentado en cada nueva generación de procesadores para conseguir mejor precisión y eficiencia en las predicciones.

El Pentium Pro y sus sucesores fueron un paso más allá del BTB, incorporando la tecnología de ejecución dinámica (*Dynamic Execution Technology*), la cual permite predecir múltiples bifurcaciones en las instrucciones. Así se acelera el flujo de procesamiento del microprocesador mediante un análisis del flujo de datos, para crear un plan optimizado de ejecución de instrucciones luego de analizar los datos cruzados entre las mismas. En este paso, también se puede cambiar el orden de ejecución de éstas para aprovechar al máximo los múltiples canales de procesamiento, aunque sin alterar los resultados finales y siempre que sea posible. Esto hace que el paralelismo sea mucho más probable, especialmente en aquellas aplicaciones que no fueron bien programadas para aprovechar los procesadores superescalares.

Los microprocesadores evolucionaron y, tomando como base este esquema fueron modificando los pasos que conforman la cadena de los canales, los hicieron más eficientes, agregaron unidades específicas para aumentar la velocidad de ejecución de los pasos más utilizados, entre otras mejoras y muchas que todavía se pueden implementar. También, los canales pueden dividirse en varios subcanales para ejecutar determinados sub-pasos en paralelo.

La tendencia consistió en incrementar la cantidad de pasos de la cadena, para así conseguir una mayor división del trabajo total y sacar mayor provecho del *pipelining*, consiguiendo de esta manera la mayor probabilidad para el paralelismo. Uno de los grandes avances para conseguir el mayor paralelismo posible fue la incorporación de la ejecución de instrucciones fuera de orden (*out-of-order*).

Mientras una instrucción que ingresó a uno de los canales se encuentra esperando por un dato o bien por un recurso que se encuentra ocupado, los canales libres siguen procesando las siguientes instrucciones a la cual se encuentra demorada, para ir ganando tiempo y luego seguir con el flujo de instrucciones. Esto logra aprovechar el paralelismo, que de otra forma sería desperdiciado.

La ejecución fuera de orden se fue incorporando en cada vez más unidades de los canales. A su vez, algunas de éstas pueden funcionar a diferentes velocidades de reloj con respecto a la general del procesador. A partir del Pentium 4, las UAL corren al doble de velocidad de reloj, pues es la unidad más utilizada en general por las aplicaciones.



En síntesis, se abrieron una infinidad de nuevas micro-arquitecturas centradas en poder ejecutar mayor cantidad de instrucciones en menor cantidad de tiempo, optimizando la estructura interna de estos canales.

El Bus Frontal (FSB) y los Multiplicadores del Reloj

El 80486DX2 fue el primer microprocesador que trabajaba con dos velocidades de reloj diferentes: una interna y otra externa. Fue el encargado de definir los principios de trabajo de todos los procesadores que lo sucedieron, al ser el primero en **multiplicar el reloj**.

La velocidad de reloj interna es aquélla a la cual trabajan todos los componentes del microprocesador. Por otro lado, para comunicarse con el mundo exterior, como por ejemplo, con el controlador de memoria o del sistema, utiliza otra velocidad de reloj: la externa. Esta última suele ser menor que la primera, pues el diseño de motherboards y *chipssets* con velocidades de reloj tan altas como las de los procesadores es muy costoso y complejo.



El controlador de memoria o del sistema se conoce con el nombre de NorthBridge (Puente Norte) y es parte del chipset, el cual analizaremos en detalle más adelante.

Esta técnica de tener dos velocidades de reloj permite que los procesadores puedan aumentar sus velocidades de reloj sin tener que diseñar una nueva motherboard por cada nueva velocidad de procesador. Y permitió, por ejemplo, que un 80486DX2 con una velocidad interna de 66 MHz se pudiera encajar en una motherboard diseñada para trabajar a 33 MHz ($2 \times 33 = 66$). En ese caso, el factor de multiplicación del reloj era de dos.

Un Pentium III de 933 MHz usaba un FSB de 133 MHz, mientras que un Core 2 Extreme con dos núcleos de ejecución de 2,93 GHz utiliza uno de 1066 MHz o 1,066 GHz (en realidad se trata de uno de 266 MHz con su velocidad de reloj cuadruplicada). Por lo tanto, sus multiplicadores de reloj son de $7 (7 \times 133 = 933)$ y $11 (11 \times 266 = 2933)$.

El bus que establece un canal de comunicación entre el microprocesador y el controlador del sistema (NorthBridge) se conoce con el nombre de FSB (Front Side Bus - Bus frontal), bus del sistema (system bus) o bus local. La velocidad de reloj externa es aquélla a la cual trabaja este bus y tiene una incidencia muy importante en el rendimiento general del sistema.

Podemos ver una comparación detallada de los rangos de velocidades de reloj y de velocidades del FSB que ofrecen los diferentes microprocesadores en la Tabla 2.4.

Tabla 2.4. Velocidades de reloj y del FSB de los principales microprocesadores.

Microprocesador	Cantidad de núcleos de procesamiento	Velocidad de reloj (interna y de cada núcleo)		Velocidad del FSB (externa)	
		Mínima	Máxima	Mínima	Máxima
8086/8088	1	4,77 MHz	12 MHz	4,77 MHz	12 MHz
80286	1	6 MHz	20 MHz	6 MHz	20 MHz
80386 (DX y SX)	1	16 MHz	40 MHz	16 MHz	40 MHz
80486 (DX, DX2 y DX4)	1	25 MHz	100 MHz	25 MHz	50 MHz
Pentium, P54C y MMX	1	60 MHz	233 MHz	60 MHz	66 MHz
Pentium Pro	1	150 MHz	266 MHz	60 MHz	75 MHz
Pentium II	1	233 MHz	450 MHz	66 MHz	100 MHz
Pentium III	1	450 MHz	1 GHz	100 MHz	133 MHz
Pentium 4	1	1,40 GHz	3,06 GHz	400 MHz	533 MHz
Pentium 4 (Hyper-Threading)	1	2,80 GHz	3,80 GHz	533 MHz	800 MHz
Celeron D	1	2,13 GHz	3,46 GHz	533 MHz	533 MHz
Pentium D	2	2,86 GHz	3,60 GHz	533 MHz	800 MHz
Pentium Processor Extreme Edition	2	3,20 GHz	3,73 GHz	800 MHz	1066 MHz
Core 2 Duo	2	1,86 GHz	2,66 GHz (*)	1066 MHz	1066 MHz (*)
Core 2 Extreme	2	2,66 GHz	2,66 GHz (*)	1066 MHz	1066 MHz (*)
Core 2 Extreme Q	4	2,66 GHz	2,66 GHz (*)	1066 MHz	1066 MHz (*)
K5	1	100 MHz	166 MHz	66 MHz	66 MHz
K6	1	166 MHz	300 MHz	66 MHz	66 MHz
K6-2	1	266 MHz	450 MHz	66 MHz	100 MHz
K6-III	1	400 MHz	450 MHz	100 MHz	100 MHz
6x86	1	120 MHz	200 MHz	66 MHz	75 MHz
6x86MX (M2)	1	166 MHz	350 MHz	66 MHz	100 MHz
Sempron (**)	1	1600 MHz	2200 MHz	1600 MHz	1600 MHz
Athlon 64 (**)	1	1800 MHz	2400 MHz	1600 MHz	2000 MHz
Athlon 64 FX (**)	1	2200 MHz	3000 MHz	1600 MHz	2000 MHz
Athlon 64 X2 (**)	2	2000 MHz	2800 MHz (*)	2000 MHz	2000 MHz (*)

(*) Continúa evolucionando

(**) Las velocidades del FSB mostradas en realidad son las del bus HyperTransport, pues en estos microprocesadores el controlador de memoria (NorthBridge) está incorporado y no es externo

El Pentium Pro inauguró una arquitectura de buses conocida como DIB (Dual Independent Bus - Doble Bus Independiente), que luego sería aplicada a todos sus sucesores. Existen dos buses independientes (de ahí el nombre): el FSB y el que comunica al procesador con la caché L2 incorporada. Todo esto se encuentra encerrado en el mismo embalaje, por lo cual la velocidad del segundo bus no será limitada por el diseño de la motherboard, sino que podrá crecer en forma independiente, sin limitaciones externas.

Esto fue otro paso más para independizar el crecimiento de los canales de comunicación del procesador con elementos que inciden mucho en su rendimiento. Pero, esta tendencia no termina allí y en los microprocesadores de AMD a partir del Athlon 64 también se incorpora el controlador de memoria al microprocesador, para mejorar su rendimiento. Es de esperar que en las siguientes generaciones de procesadores se siga con esta tendencia, sin embargo, por el momento, Intel solamente la aplica a la gama alta de servidores en sus procesadores Itanium 1 y 2.

Las Instrucciones Adicionales Específicas

Todos los procesadores de 64 bits compatibles con la familia 80x86 brindan soporte completo al juego de instrucciones conocido como IA-32 (Intel Architecture 32 bits - Arquitectura Intel de 32 bits) y, según el caso a x86-64 ó EM64T o bien a IA-64.

Con los cambios en su arquitectura y el aumento de sus velocidades de reloj, se intenta ejecutar más rápidamente las instrucciones de estos juegos. De esta manera, sin realizar ningún tipo de modificación en su código, una misma aplicación es más veloz si se la ejecuta en un microprocesador que posea mejores optimizaciones para procesar en menor tiempo las instrucciones IA-32; x86-64; EM64T o IA-64.

Otra técnica para conseguir mejorar el rendimiento de las aplicaciones, consiste en ofrecer nuevos conjuntos de instrucciones en el procesador que realicen tareas específicas a gran velocidad, las cuales sean muy comunes en determinadas aplicaciones. De esta manera, reescribiendo el código de estas últimas para sacar provecho de estas instrucciones, se consigue utilizar al máximo las nuevas bondades que ofrece un microprocesador y se mejora el rendimiento.

La gran desventaja de esta técnica radica en la necesidad de reescribir las aplicaciones para aprovechar las nuevas instrucciones. Las aplicaciones que no estén programadas para el nuevo juego de instrucciones, no obtendrán ningún beneficio y no aumentarán su rendimiento al ejecutarse utilizando estos procesadores.

Con el objetivo de mejorar drásticamente el rendimiento de los juegos y aplicaciones con contenido multimedia (audio, video, animaciones, realidad virtual, gráficos 3D), así como las comunicaciones, aparecen instrucciones adicionales

específicas que se agregan al juego de instrucciones básico. Este tipo de aplicaciones suele generar una gran carga de procesamiento a los microprocesadores al utilizar las instrucciones de 32 y 64 bits convencionales, por lo tanto, le quitan posibilidades a otras tareas para que sean ejecutadas al mismo tiempo.

respectivamente. Siguiendo con la línea de toda la familia 80x86, los nuevos procesadores que van surgiendo son compatibles hacia atrás con los juegos de instrucciones adicionales presentados por sus predecesores. A continuación, vamos a analizarlas en detalle. Ver la Tabla 2.5.

MMX: Extensiones Multimedia

En marzo de 1996, Intel presentó la tecnología MMX (*MultiMedia Extensions* – Extensiones multimedia). Entre otras cosas, incorpora las siguientes características a los microprocesadores de la familia 80x86 que la adoptan:

- 57 instrucciones nuevas que incorporan la técnica SIMD (*Single Instruction, Multiple Data* – Una instrucción, múltiples datos).
- Ocho registros de 64 bits nuevos para uso exclusivo de las instrucciones MMX.
- Cuatro tipos de datos nuevos, también para uso exclusivo de estas instrucciones.

La técnica SIMD permite que una instrucción vaya acompañada por datos en los cuales se encuentren empaquetados grupos de datos de menor tamaño y de esta forma, mediante una única instrucción, se pueden procesar en paralelo varias porciones de información. Esta técnica no es nueva, pues existe hace más de 30 años, pero en procesadores que no formaban parte de las PC.

Esta técnica va de la mano con los cuatro nuevos tipos de datos que presenta MMX, todos ellos de 64 bits:

- **Packed byte.** Ocho bytes empaquetados en una unidad de 64 bits ($8 \times 8 = 64$).
- **Packed word.** Cuatro unidades de 16 bits en una de 64 bits.
- **Packed doubleword.** dos unidades de 32 bits en una de 64 bits.
- **Quadword.** Una unidad de 64 bits.

El aumento en el rendimiento logrado al utilizar MMX se debe a que, gracias a la técnica SIMD y a los nuevos tipos de datos introducidos, se requieren menor cantidad de instrucciones del procesador para realizar determinadas operaciones. A su vez, estas instrucciones efectúan operaciones sobre los datos empaquetados en paralelo.

Por ejemplo, si tenemos que comparar el contenido de color de 8 pixels que está representado por un byte, con MMX solamente necesitamos empaquetar la información de los 8 en un *packed byte*, cargarlo en un registro y ejecutar una instrucción de comparación con ese dato.

Tabla 2.5. Resumen de instrucciones adicionales específicas para cada procesador.

Procesador	Instrucciones Adicionales Específicas									
	Desarrolladas por Intel					Desarrolladas por AMD				
	MMX	SSE	SSE 2	PNI/SSE 3	EM64T	3DNow!	Enhanced 3DNow!	3DNow! Professional	x86-64 (AMD64)	
Athlon 64	•									
Athlon 64 FX	•									
Athlon 64 X2	•									
Athlon XP	•									
Core 2 Duo	•									
Core 2 Extreme	•									
Core Duo	•									
K6	•									
K6-2	•									
Pentium 4	•									
Pentium 4 Hyper-Threading	•									
Pentium 4 Prescott Hyper-Threading	•									
Pentium D	•									
Pentium MMX	•									
Pentium Processor Extreme Edition	•									
Sempron	•									

Existen extensiones desarrolladas por Intel y otras por AMD, las cuales fueron apareciendo con los microprocesadores a partir del Pentium MMX y el K6,

Utilizando las instrucciones convencionales, habría que hacerlo en ocho operaciones de comparación las cuales llevarían un byte como dato cada una. En este ejemplo, el rendimiento es un 800% superior.

Las 57 instrucciones MMX cubren las siguientes áreas funcionales, siempre sobre los cuatro tipos de datos mencionados anteriormente:

- Operaciones aritméticas básicas (suma, resta, multiplicación, etc.).
- Operaciones de comparación.
- Operaciones lógicas.
- Operaciones de arrastre de bits (*shift*).
- Transferencias de datos entre registros, entre éstos y la memoria, tanto para 64 como para 32 bits.
- Conversión entre los diferentes tipos de datos, empaquetar y desempaquetar.



MMX se podría definir sencillamente como una ALU vectorial.

Cuando el software está escrito para tomar ventaja de MMX, el aumento en el rendimiento con respecto a su versión convencional llega a ser de un 400% en la mayoría de los casos, aunque depende exclusivamente del tipo de aplicación.

Internet SSE: De la Mano del Pentium III

Con la llegada del Pentium III, Intel presentó la siguiente generación de la tecnología MMX, las Internet SSE (*Streaming SIMD Extensions* - Extensiones SIMD en flujos), con el objetivo de mejorar el rendimiento de las aplicaciones que hacen uso intensivo de:

- Gráficos 3D.
- Procesamiento de imágenes de alta calidad.
- Manejo de audio de alta calidad.
- Decodificación y/o codificación MPEG2 para la reproducción o grabación de videos.
- Reconocimiento de voz.

Entre otras cosas, esta tecnología incorpora a la MMX existente las siguientes características...

- 70 instrucciones nuevas que utilizan la ya conocida técnica SIMD, pero con muchas mejoras. 50 de las 70 fueron diseñadas para efectuar operaciones de punto flotante.

- Ocho nuevos registros de 128 bits, para valores de punto flotante.
- Un tipo de datos nuevo para uso exclusivo de estas instrucciones: una unidad de 128 bits que permite empaquetar cuatro valores de punto flotante de 32 bits y procesarlos todos juntos con una única instrucción SIMD de las nuevas.
- Varias mejoras a la arquitectura MMX y a la eficiencia de las operaciones Internet SSE, para que no existan los cuellos de botella ya detectados en la versión anterior y se puedan ejecutar instrucciones de los dos tipos en forma concurrente.

Al permitir operaciones sobre valores de punto flotante, Internet SSE representa la transición de una ALU vectorial (MMX) a una UPF vectorial, la cual es más razonable y mucho más útil en las aplicaciones.

HASTA AQUÍ SE VA...

Internet SSE 2: De la Mano del Pentium 4

Con la llegada del Pentium 4, Intel presentó una nueva generación de instrucciones SIMD: las Internet SSE 2, con el objetivo de mejorar aún más el rendimiento de las aplicaciones que ya aprovechaban Internet SSE y también para aumentar la velocidad de la encriptación y desencriptación de datos en tiempo real.

Entre otras, esta tecnología agrega las siguientes características:

- 144 instrucciones nuevas que utilizan la técnica SIMD. Muchas de las instrucciones del juego de Internet SSE y de MMX fueron optimizadas para mejorar su rendimiento.
- Operaciones aritméticas básicas para enteros de 64 bits.
- Conversión entre los diferentes tipos de datos nuevos y antiguos, empaquetar y desempaquetar.
- Más cantidad de instrucciones para entremezclado de datos, muy utilizadas en las técnicas de encriptación de datos.
- Mayor soporte para operaciones de ordenamiento y de administración específica de la memoria caché.

Además de las nuevas instrucciones, Internet SSE 2 agrega varios tipos de datos nuevos para uso exclusivo de éstas, los cuales permiten procesar todos los valores empaquetados juntos con una única instrucción SIMD:

- Una unidad de 128 bits que permite empaquetar dos valores de punto flotante de 64 bits.
- Una unidad de 128 bits que empaqueta el doble de las que permitían los primeros cuatro tipos de datos que introdujo MMX.

- o **Doble packed byte.** Dieciséis bytes empaquetados en una unidad de 128 bits ($16 \times 8 = 64$).
- o **Doble packed word.** Ocho unidades de 16 bits en una de 128 bits.
- o **Doble packed doubleword.** Cuatro unidades de 32 bits en una de 128 bits.
- o **Doble quadword.** Dos unidades de 64 bits en una de 128 bits.

SSE 3: De la Mano del Pentium 4 Prescott

Con la llegada del Pentium 4, desarrollado con un proceso de fabricación de 90 nanómetros, Intel presentó una nueva generación de instrucciones SIMD: las SSE 3, también conocidas por su nombre anterior PNI (*Prescott New Instructions* - Nuevas instrucciones de Prescott), debido al nombre del núcleo del nuevo microprocesador que las presentaba.

Entre otras, esta tecnología agrega las siguientes características:

- Solamente 13 instrucciones nuevas que utilizan la técnica SIMD y mejoran las prestaciones de las existentes en versiones anteriores.
- Mejoras para la sincronización de múltiples hilos de ejecución (*threads*).
- Operaciones aritméticas horizontales con registros SSE2, para la optimización del procesamiento de gráficos 3D.
- Operaciones con números complejos, útiles para el procesamiento digital de señales.

3DNow! de AMD

El K6 de AMD incorporó la tecnología MMX. Como ya vimos anteriormente, MMX solamente permitía ejecutar operaciones SIMD sobre enteros, es por ello que AMD incluyó en sus procesadores K6-2 el agregado de la tecnología 3DNow!, la cual sumó 21 instrucciones SIMD nuevas a las originales de MMX, incluyendo un conjunto especial para realizar operaciones sobre valores de punto flotante.

3DNow! Enhanced: De la Mano del Athlon

AMD debía ofrecer una alternativa contra las Internet SSE que, a diferencia de MMX (que sí estaban presente en sus procesadores), no se incluían en el K7 (Athlon), por lo cual se mejoró la tecnología 3DNow! con 24 instrucciones nuevas, según la siguiente estratificación:

- 12 instrucciones para aumentar la velocidad de las operaciones matemáticas con enteros, para mejorar el rendimiento de la codificación de video y sonido. Compatibles con Internet SSE.

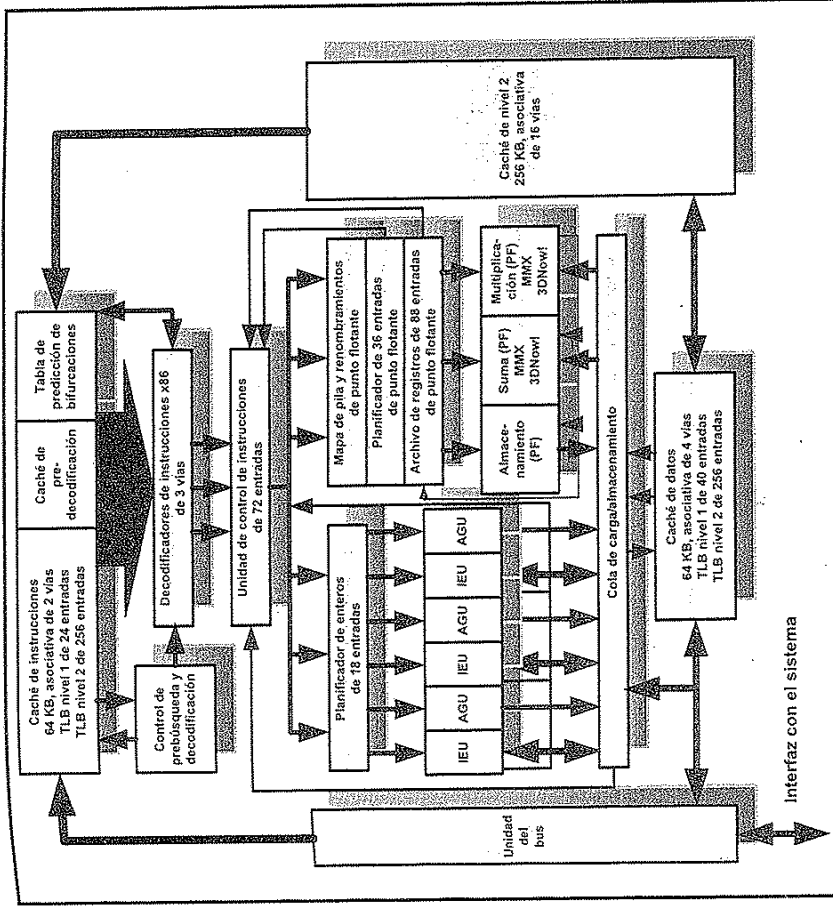


Fig. 2.10. Diagrama en bloques de la microarquitectura del Athlon XP.

- 7 instrucciones para incrementar la velocidad de movimiento de datos. Compatibles con Internet SSE.
- 5 instrucciones DSP (*Digital Signal Processing* - Procesamiento de Señales Digitales) para mejorar el rendimiento de módems por software, sonido digitalizado y la codificación y decodificación de sonido en formato MP3. A esta actualización se la conoce con el nombre de Enhanced 3DNow! (3DNow! Mejorado).

3DNow! Professional: De la Mano del Athlon XP

El Athlon XP agregó soporte a 51 instrucciones de Internet SSE y de esta manera, junto con las 19 (12 + 7) que ya se incluían en 3DNow! Enhanced, se logra la total compatibilidad con Internet SSE. A esta tecnología, se la denominó 3DNow! Professional. Sin embargo, luego, AMD fue incorporando en sus procesadores todos los juegos de instrucciones adicionales específicas desarrollados por Intel, hasta SSE 3.

Micro-Arquitecturas Básicas

Las dos micro-arquitecturas representativas de los pilares para la evolución de los procesadores de 32 bits a los 64 bits y a la integración de múltiples núcleos de ejecución fueron la *Quantispeed* del Athlon XP y la *NetBurst* del Pentium 4. Con algunos agregados y modificaciones, sentaron las bases para los siguientes microprocesadores de AMD e Intel. Vamos a analizar las novedades que introdujeron estas micro-arquitecturas en su diseño.

El Athlon XP presentó (ver la Fig. 2.10):

- Tres decodificadores de instrucciones x86.
- Tres canales (*pipelines*) independientes para operaciones con enteros.
- Tres UPF, totalmente canalizadas (*pipelined*) en tres vías, con capacidad de ejecución de instrucciones fuera de orden.
- Tres canales (*pipelines*) de cálculo de direcciones.
- Una estructura de traducción de datos y direcciones de dos niveles.

El Pentium 4, en cambio, introdujo:

- Un canal (*pipeline*) con tres secciones de primer nivel (ver la Fig. 2.11):
 - El primer nivel, con ejecuciones ordenadas.
 - El núcleo de ejecución superescalar fuera de orden.
 - La unidad de retiro.
- Un decodificador de instrucciones x86.
- Cuatro puertos para la ejecución de diferentes clases de instrucciones en paralelo, fuera de orden y a diferentes velocidades (ver la Fig. 2.12).
- Dos UAL trabajando al doble de la velocidad de reloj del procesador.
- Dos UPF, en puertos para ejecución separadas y paralelas, una específica para instrucciones de movimiento de datos y otra para el resto de las instrucciones.

- Un puerto para cargas de memoria.
- Un puerto para almacenamiento en memoria.

Los temas relacionados con los tamaños, las tecnologías y las arquitecturas de las memorias caché de los procesadores los vamos a analizar detalladamente en el **Capítulo 4: La Memoria.**

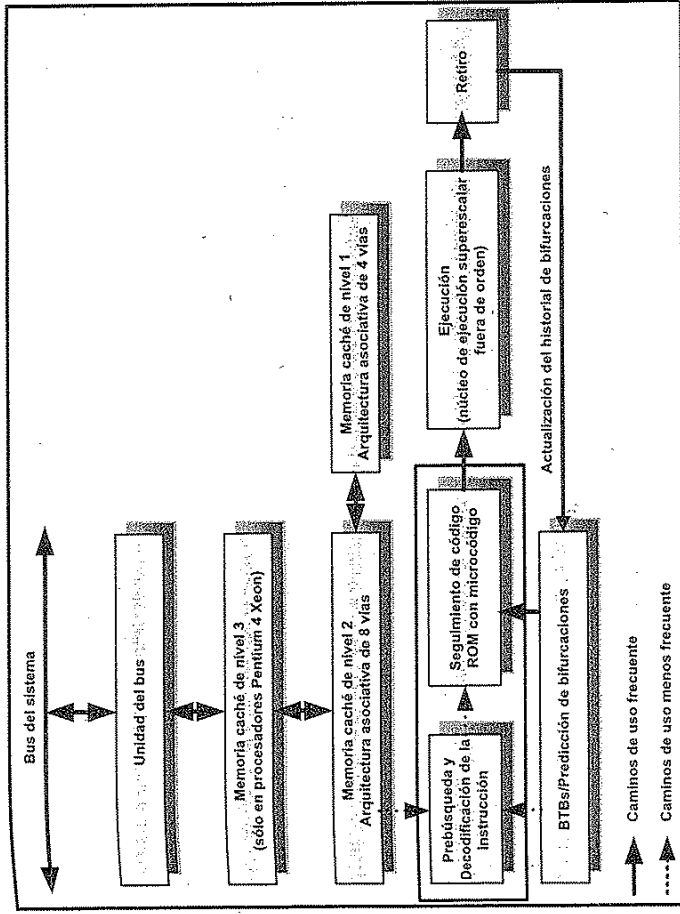


Fig. 2.11. Diagrama en bloques de la micro-arquitectura NetBurst del Pentium 4.

La tercera generación de los procesadores Pentium 4 presentaron la tecnología *Hyper-Threading* (Híper hilos de ejecución) de Intel, la cual sería la base para los siguientes procesadores con múltiples núcleos de ejecución. Esta tecnología hace que un único Pentium 4 se presente ante el sistema operativo y el resto del software como si fueran dos procesadores. Eso lo consigue simulando la existencia de dos procesadores, duplicando sus estructuras visibles ante el software, como ser sus registros.

Las ventajas ofrecidas por presentarse como dos procesadores consistían en que el software existente no estaba preparado para sacar el mayor provecho a las

micro-arquitecturas de los procesadores de última generación, por lo tanto, desperdiciaban muchos recursos provistos por los mismos. En cambio, si se presentaban como dos procesadores y se ejecutaban varias aplicaciones al mismo tiempo, con múltiples hilos de ejecución cada una, se aprovechaban los dos procesadores con los que se encontraban, al utilizar uno cada aplicación o hilo de ejecución que se estaba ejecutando y así sacaban un mayor provecho de sus posibilidades de paralelismo.

además de las mejoras en las micro-arquitecturas, se ha empleado durante muchos años y, como se pudo observar en las Tablas 2.1 y 2.4, evolucionó desde los humildes 4,77 MHz del 8086/8088 hasta varios GHz de los procesadores modernos.

Sin embargo, cada vez resulta más difícil aumentar la frecuencia del reloj, pues el proceso de fabricación se hace más complejo y aparecen grandes problemas para poder disipar el calor generado en forma eficiente y económica.

Es por ello que la otra alternativa es mejorar aún más la micro-arquitectura de los microprocesadores y dotarlos de estructuras de procesamiento duplicadas, cuadruplicadas y así sucesivamente, de manera tal que en un mismo encapsulado, se puedan tener múltiples sub-procesadores, denominados núcleos de ejecución (*execution cores*).

Sin entrar en mayores detalles técnicos, podemos decir que los microprocesadores con múltiples núcleos de ejecución, también conocidos como multi-core, ofrecen varios núcleos completos de ejecución interconectados entre sí en un único encapsulado. Es decir, físicamente tienen un aspecto muy similar a un microprocesador convencional, pero internamente serían el equivalente a algo así como dos o más microprocesadores pero contenidos en un mismo trozo de silicio.

Los primeros en aparecer con esta novedosa micro-arquitectura fueron los microprocesadores con doble núcleo de ejecución, también conocidos como *dual-core*, como el Pentium D y el Athlon 64 X2. En la Tabla 2.4 pudimos ver el resumen de las velocidades de trabajo de los diferentes procesadores y allí se indica también la cantidad de núcleos de ejecución que poseen.

Ya existen procesadores con cuatro núcleos de ejecución en un mismo encapsulado, como el Core 2 Extreme Q de Intel.

En el mismo espacio en el cual se tenía un microprocesador con un único núcleo de ejecución, ahora es posible tener dos o cuatro núcleos y se espera que en el futuro se vayan duplicando la cantidad de núcleos.

Una de las ventajas que ofrecen estos microprocesadores es que se pueden reaprovechar las motherboards diseñadas para procesadores con un único núcleo de procesamiento y aumentar el rendimiento reemplazando simplemente el micro por uno de dos o cuatro núcleos, siempre que el encapsulado sea compatible.

A partir del agregado de núcleos de ejecución, se empiezan a multiplicar las posibilidades de combinaciones de arquitecturas de comunicación, así como de recursos propios o compartidos por los diferentes núcleos. Cada posibilidad tiene

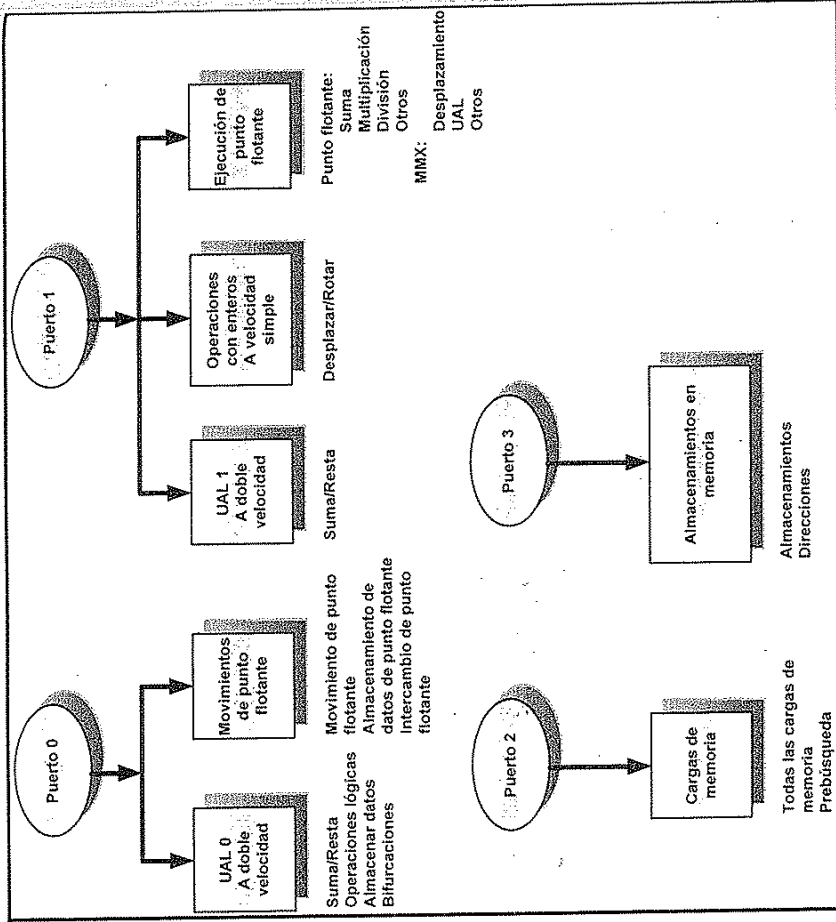


Fig. 2.12. Unidades de ejecución del Pentium 4.

Múltiples Núcleos de Ejecución (Multi-Core)

Para mejorar la capacidad de procesamiento, una de las técnicas consiste en ir aumentando la frecuencia de trabajo de los microprocesadores. Esta técnica,

sus ventajas para el rendimiento y sus desventajas en los costos de fabricación, por lo cual, se abrió un nuevo abanico en la oferta de microprocesadores.

En algunos casos, cada núcleo de ejecución incluye a las memorias caché L1 y L2, mientras que en otros, comparten las memorias caché L2 entre dos o más núcleos. Por supuesto, mientras más recursos contenga cada núcleo y menos comparta con los otros, mayor será la velocidad de procesamiento.

En estas micro-arquitecturas con múltiples núcleos de ejecución, uno de los temas más importantes es conseguir una eficiencia en los accesos a la memoria externa, los cuales resultan extremadamente costosos en tiempo, comparado con la velocidad interna de los núcleos, y hay que evitarlos a toda costa para conseguir el mejor rendimiento posible. Seguiremos profundizando en el análisis de este aspecto en el **Capítulo 4: La Memoria**.

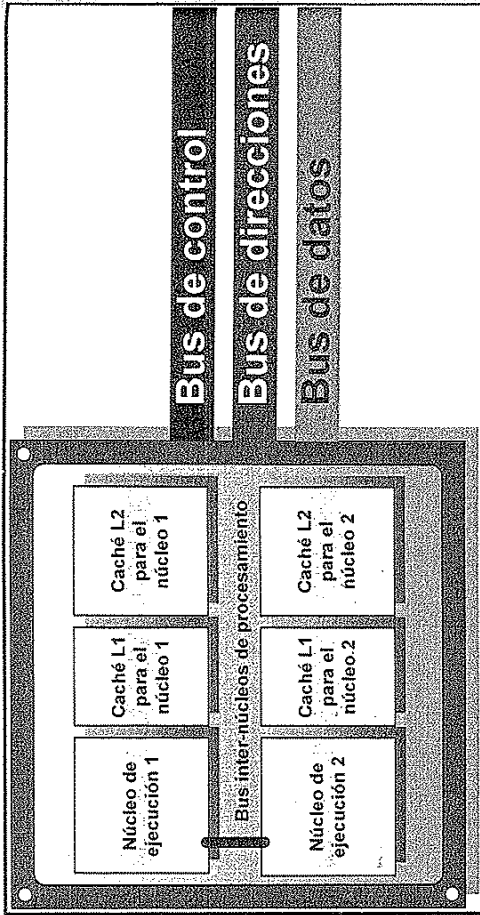


Fig. 2.13. Diagrama en bloques de una arquitectura de un procesador de doble núcleo.

En la Fig. 2.12 podemos ver la micro-arquitectura de un microprocesador con doble núcleo (*dual-core*). En este caso, las memorias caché no las comparten. La mayoría de los microprocesadores de doble núcleo utilizan ese esquema.

Uno de los puntos más importantes a tener en cuenta es el bus interno que se usa para que los diferentes núcleos de procesamiento se comuniquen entre sí. Algunos microprocesadores utilizan uno o varios buses específicos con altísimas velocidades de trabajo, mientras que otros lo hacen a través del FSB (mucho menos eficiente). Cuando hay más de dos núcleos, la arquitectura puede ser una combinación de ambos, como es el caso del Core 2 Extreme Q, el cual está

conformado por dos pares de núcleos, cada uno utiliza un bus específico para que los dos núcleos intercambien información, pero los dos pares se comunican a través del FSB. En realidad, este microprocesador combina dos procesadores de doble núcleo, que comparten la caché L2, con una comunicación a través del FSB (ver la Fig. 2.14).

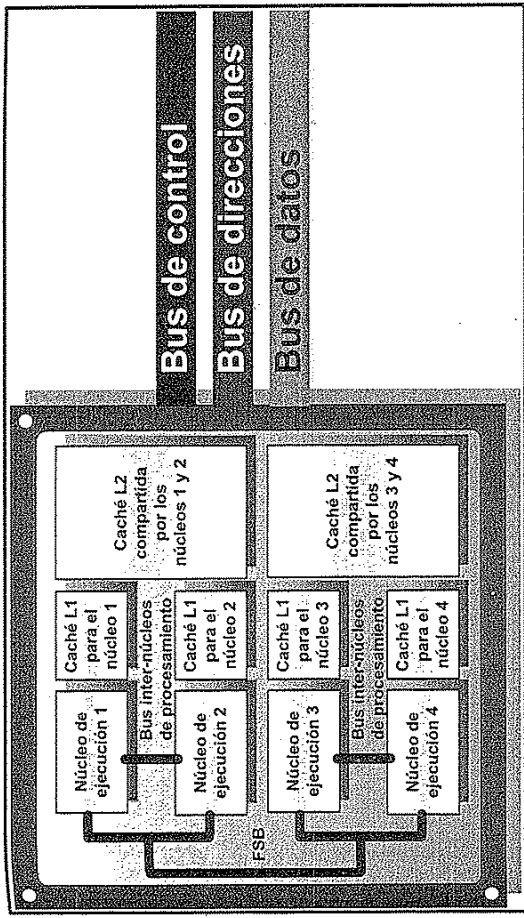


Fig. 2.14. Diagrama en bloques de una arquitectura de un procesador de cuádruple núcleo con conexión de dos pares de dobles núcleos a través del FSB.

El esquema de mayor eficiencia de un microprocesador con cuádruple núcleo es el que se muestra en la Fig. 2.15. Pero, su gran desventaja es que acarrea un costo de fabricación superior al de la Fig. 2.14.

Debemos tener en cuenta que los microprocesadores con múltiples núcleos de ejecución no son exactamente equivalentes a múltiples microprocesadores trabajando en un esquema de multiprocesamiento asimétrico.

En el multiprocesamiento simétrico (*symmetrical multiprocessing*), también conocido como SMP, cualquiera de los microprocesadores presentes que esté disponible puede ejecutar tareas. La variante más utilizada y eficiente es el multiprocesamiento simétrico de *n* vías (*n way*) o *n* caminos, donde *n* se reemplaza por la cantidad de microprocesadores utilizados. En este esquema, cada microprocesador puede ejecutar una tarea separada del resto, inclusive cuando el software no está escrito pensado para correr en un sistema con multiprocesamiento.

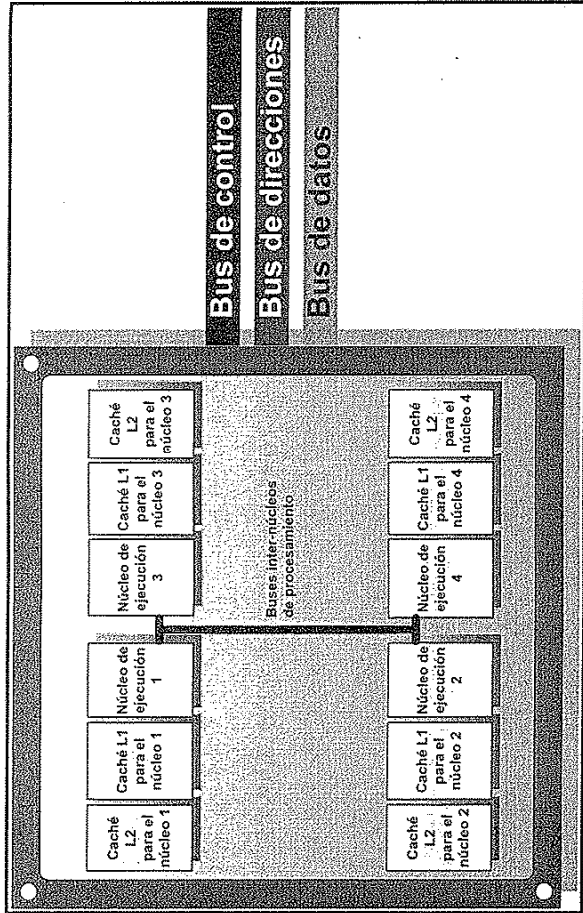


Fig. 2.15. Diagrama en bloques de una arquitectura de un procesador de cuádruple núcleo con cuatro núcleos completos independientes.

En el multiprocesamiento asimétrico, al existir varios microprocesadores físicamente, cada uno tiene su FSB disponible para acceder a la memoria externa. En cambio, en un microprocesador con múltiples núcleos, todos ellos entran y salen al mundo exterior a través de un único FSB compartido que hace de gran portón de entrada y salida desde y hacia la memoria principal. Es por ello que las tareas de coordinación para evitar conflictos por este bus requieren mayor tiempo de procesamiento que en un sistema de multiprocesamiento asimétrico de n vías.

Además, las probabilidades que el FSB se transforme en un gran cuello de botella son muy altas si las aplicaciones no están optimizadas para sacar el máximo provecho posible de las memorias caché que incorpora cada núcleo de ejecución. Por eso, en estas micro-arquitecturas se debe evitar a toda costa tener que salir frecuentemente a la memoria principal.

También, muchos sistemas de multiprocesamiento asimétrico utilizan esquemas duplicados de comunicación con la memoria principal, los cuales por el momento no aparecen en los microprocesadores con múltiples núcleos de ejecución.

Es por ello que si comparamos el rendimiento de una aplicación en particular en un sistema con determinados microprocesadores en sus versiones de

multiprocesamiento asimétrico de 2 vías y en otro con esas mismas características pero con un único microprocesador con dos núcleos de ejecución, se conseguirá un mejor resultado en el primer caso.

Sin embargo, la diferencia de costos hace inclinar fuertemente la balanza hacia los microprocesadores con múltiples núcleos de ejecución, los cuales a su vez pueden ser partes de un sistema de multiprocesamiento asimétrico de n vías, si utilizamos varios de ellos. Algo que puede resultar muy atractivo para estaciones de trabajo de alto rango y servidores.



Para mayor información sobre el funcionamiento de equipos con múltiples microprocesadores, puede consultar el libro *Servidores de Redes*, de Editorial HASA.

Sin lugar a dudas, los microprocesadores van a seguir incorporando más y más núcleos de procesamiento en los próximos años y resultan en una excelente opción para los equipos que tienen cada vez mayores aplicaciones ejecutándose al mismo tiempo. Por suerte, los sistemas operativos Windows y Linux ya están preparados para aprovechar las características de estos procesadores y las aplicaciones se irán adaptando con el tiempo para que les saquen el máximo beneficio.

El Chipset y Su Importancia

Es uno de los componentes más importantes de la motherboard a la hora de determinar el rendimiento general de la misma. Se llama *chipset* al conjunto de circuitos integrados que se encargan de ayudar al microprocesador a realizar ciertas tareas que éste delega en ellos.

Generalmente está conformado por dos circuitos integrados (ver la Fig. 2.16):

- El controlador de memoria o del sistema, conocido con el nombre de *NorthBridge* (Puente Norte). Se encarga de controlar y organizar el acceso a la memoria principal, al bus de datos y de direcciones. Determina la clase de tarjetas, las tecnologías, las velocidades y el tamaño máximo de memoria con la cual puede trabajar la motherboard. También, define la velocidad del bus de datos con la memoria. Debemos recordar que en las motherboards para los procesadores de última generación de AMD, el *NorthBridge* está incorporado en los procesadores, por lo tanto, encontramos al controlador de los buses *HyperTransport* en su lugar.
- El controlador de entrada/salida, conocido con el nombre de *SouthBridge* (Puente Sur). Se encarga de controlar los buses de E/S, los cuales analizaremos en detalle en el siguiente capítulo.

El *chipset* debe estar diseñado pensando en el procesador que se utiliza en la motherboard. Generalmente, soportan un conjunto limitado de microprocesadores.

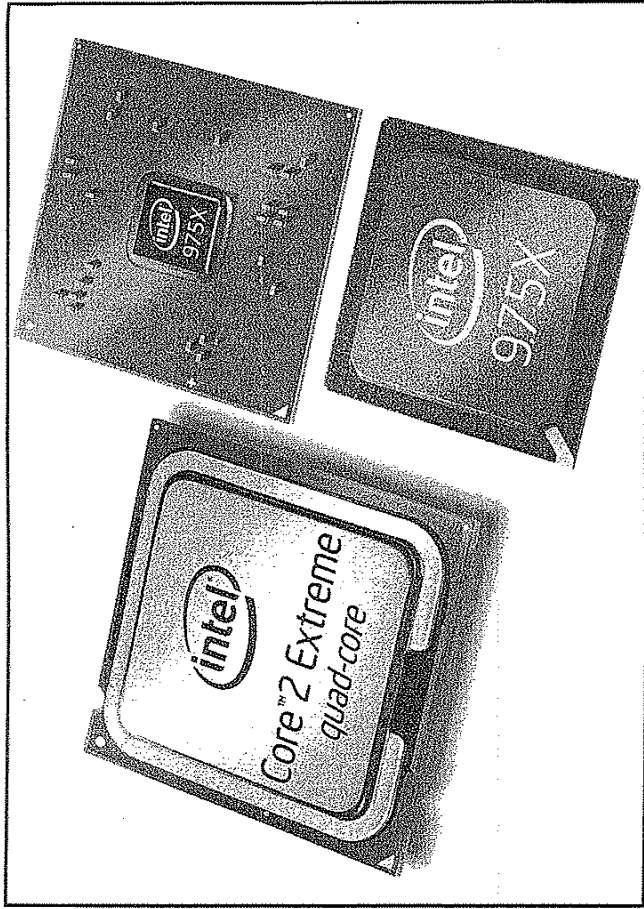


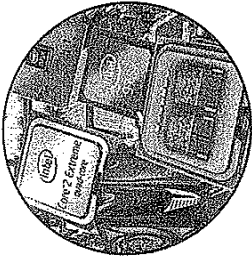
Fig. 2-15. Diagrama en bloques de una arquitectura de un procesador de *cuádruple núcleo* con cuatro núcleos completos independientes.

Al juego de chips se lo reconoce fácilmente, pues está soldado en forma fija a la motherboard y su marca suele estar con letra grande y blanca. Sin embargo, en las motherboards modernas, aparecen con importantes disipadores.

En los últimos tiempos, los juegos de chips han incorporado muchas bondades adicionales, que varios años atrás eran propias de las tarjetas de expansión que se agregaban a la motherboard. Entre otras, podemos mencionar las siguientes funciones:

- Interfaces para discos rígidos y otros dispositivos de almacenamiento: Serial ATA y Ultra ATA (IDE).
- Buses de E/S USB 2.0.
- Buses de E/S IEEE 1394.
- Tarjetas de video.
- Tarjetas de sonido.
- Interfaces para unidades de diskettes. Todavía hay algunas motherboards que las incluyen, aunque están en extinción.

Capítulo 3



Buses Internos: Las Autopistas de los Microprocesadores

Buses Internos

Los **buses internos** son los caminos que unen a los microprocesadores con los demás componentes internos de entrada y salida. Por ejemplo, se encargan de comunicar al procesador con las controladoras de discos rígidos, con la tarjeta de video conectada en una ranura de expansión y encargada de enviar información al monitor, con los buses de E/S USB e IEEE 1394, etc. Estos mismos caminos trasladan los datos entre el procesador y la memoria que reside en la motherboard y en el caso de procesadores antiguos que tenían un coprocesador matemático externo (ahora están incluidos en el chip), también los comunicaba con este último.

La comunicación con el resto de los componentes la realiza a través de vías de trazos metálicos en el circuito impreso. Estas líneas de cobre se encuentran en paralelo y la cantidad dependerá del ancho del bus con el cual trabaje el procesador externamente. Para un procesador 386SX, eran 16 líneas de cobre, para un 486 eran 32 y para un Pentium y sus sucesores son 64.

El bus interno vincula al procesador con los dispositivos de entrada y salida, pero también enlaza a estos dispositivos con la memoria. Por lo tanto, el bus puede transportar los datos que un dispositivo de entrada le entrega a la memoria, sin necesidad que la CPU pierda tiempo participando en este evento y librándola de esa tarea. También, la memoria puede entregar datos a un dispositivo de salida sin necesidad de interrumpir el funcionamiento de la CPU. Para cumplir con esa función, el dispositivo debe avisar al procesador que utilizará el bus, de manera que éste no trate de accederlo al mismo tiempo y produzca una mezcla, choque o corrupción de los datos existentes en el mismo. Para esto, al dispositivo de E/S que trabaje de esta forma, se le asigna un canal de acceso directo a memoria (DMA). La información más explícita de los canales de DMA se verá más adelante, porque ahora nuestro interés está en los buses y sus formas de trabajo y el tema del acceso

directo a memoria es demasiado extenso e importante por lo cual le dedicaremos en forma exclusiva una sección de este libro.

El bus es uno de los elementos más importantes en el rendimiento global de un sistema de computación. Imaginemos la siguiente situación, si tengo una empleada que sabe mecanografía y escribe a una velocidad mayor de 100 palabras por minuto, pero cada 3 minutos se queda sin hojas y manda al cadete a comprar 1 sola. El cadete tarda media hora en traer la hoja, porque va caminando despacio, se queda mirando vidrieras por el camino, se toma algo y después llega con la hoja. He desperdiciado la velocidad de la empleada porque el muchacho que se encarga de traer las hojas malgasta el tiempo y trae poca cantidad cada vez que viene.

El mismo concepto anteriormente explicado se aplica a los sistemas de computación. Podemos tener el procesador más rápido existente en el mercado, pero si el sistema que lo comunica con el mundo exterior es lento, no sirve de nada, porque el procesador se pasará la mayoría del tiempo esperando que el bus lleve y traiga los datos. Si el bus es rápido pero trae pocos datos al mismo tiempo, tampoco es bueno del todo, como sucedería si yo reemplazara al cadete por uno que me traiga 1 hoja por vez, pero corriendo. En cambio, si pongo uno que me traiga 500 hojas por vez, solamente lo tendré que enviar a comprar hojas un par de veces en un día.

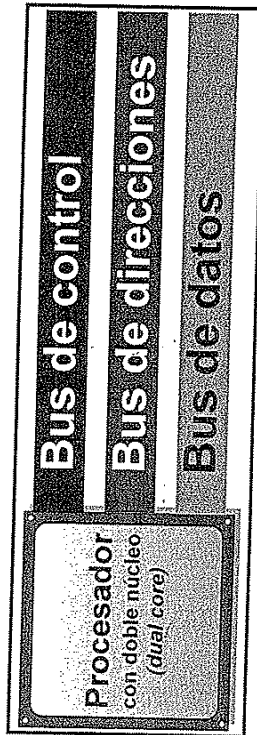


Fig. 3.1.
El procesador utiliza varios buses diferentes para transferir la información al lugar adecuado.

Diferentes Tipos de Bus y Sus Componentes

El bus de datos es el camino por donde el procesador transporta los datos desde y hacia los dispositivos de entrada y salida, pero, los datos no sirven de nada si no se establece el destino de los mismos. Para cumplir con este fin aparece el bus de direcciones, que fue nombrado tantas veces cuando se describían los diferentes procesadores. Ver la Fig. 3.1.

Otro más es el bus del sistema o bus de control, que como lo indica su nombre, se encarga de administrar la forma como serán encarados los procesos a llevar a cabo, es decir, que el procesador lo utiliza para enviar comandos (órdenes). Por ejemplo, administrar si se escribirá a la memoria o se tomará un dato de la misma para llevarlo directamente a un dispositivo de salida.

Hay varios componentes que forman el controlador del bus y se encargan de administrar de manera eficaz todos los conductores que lo conforman. Su función es comparable a la de un agente de tránsito, tratará de dirigir los datos por el camino adecuado y evitará cualquier tipo de colisión.

Como se puede deducir de lo anteriormente explicado, cada tipo de procesador deberá tener su propio sistema de transporte de datos. A mayor velocidad del procesador, más rápido deberá ser el bus para seguirle el ritmo y no tenerlo esperando mucho tiempo. Debido a que el diseño de un bus rápido trae bastantes complicaciones y requiere de componentes costosos, Intel prefirió seguir aumentando la velocidad interna del procesador antes de diseñar el nuevo bus que debutaría con la aparición del primer sistema Pentium.



Diferentes Ranuras de Expansión

Las ranuras de expansión son conectores que se encuentran sobre la motherboard y que se conectan con el sistema de bus. En éstas se enchufan las placas adicionales a la motherboard, como la tarjeta de video con aceleración 3D, la tarjeta sintonizadora de TV, la placa de red inalámbrica y la tarjeta de sonido.

Al existir diferentes tipos de bus, hay ranuras de expansión adecuadas para cada uno de ellos. Hace algunos años, siempre ofrecían compatibilidad con sus predecesores, pero actualmente, los buses se han diversificado y no es así, pues una tarjeta diseñada para un bus AGP Pro no se puede conectar en una PCI Express 16x. Esto se debe a que para conseguir mejores rendimientos, se fue dejando de lado la prioridad de mantener la compatibilidad con sus predecesores a la hora de desarrollar nuevos estándares, así como nuevas ranuras de expansión que trabajarán a mayor velocidad.



También conocidas como slots de expansión, estas ranuras son fácilmente reconocibles, pues tienen diferentes tamaños, formas y longitudes. Por lo tanto, a simple vista, podemos determinar qué tipo de ranura es. Aunque, no hay que dejarse engañar por la longitud o el tamaño, pues por ejemplo, las ranuras de un bus de 64 bits son más pequeñas que las de un antiguo bus de 16 bits. Con el pasar de los años, cada tipo de ranura de expansión se fue estandarizando junto con el bus, de manera tal que existiera una garantía total de compatibilidad.

A continuación se analizará cada tipo de bus de expansión en profundidad, comenzando por el prehistórico bus ISA de 8 bits y llegando hasta los buses PCI, PCI-X 1.0 y PCI-X 2.0 de 64 bits, así como PCI Express 1x a 32x. En la Tabla 3.1

podemos ver una comparación completa de los diferentes tipos de bus con sus características principales.

Tabla 3.1. Comparación de los diferentes buses de expansión.

Bus	Ancho de datos	Velocidad máxima	Transferencia máxima	Estado actual
ISA	16 bits	8 MHz	8 MBps	Extinguído.
EISA	32 bits	8 MHz	33 MBps	Extinguído.
MCA	32 bits	10 ó 16 MHz	40 MBps	Extinguído.
VL-Bus 1.0	32 bits	33 ó 40 MHz	133 MBps (33 MHz) 148 MBps (40 MHz)	Se aplicaron a las 486 y se extinguieron junto con estas.
VL-Bus 2.0	32 bits	50 MHz	267 MBps (50 MHz)	
PCI 1.0	32 bits	33 MHz	132 MBps	
PCI 2.0	64 bits	33 MHz	264 MBps	El más popular de la actualidad.
PCI 2.1	64 bits	33 MHz	264 MBps	
PCI 2.2	64 bits	66 MHz	528 MBps	
PCI Express x1	1 par	2,5 Gts	250 MBps	El nuevo estándar en tarjetas de video
PCI Express x2	2 pares	2,5 Gts	500 MBps	con aceleración 3D.
PCI Express x4	4 pares	2,5 Gts	1 GBps	También se perfila como el
PCI Express x8	8 pares	2,5 Gts	2 GBps	reemplazante del bus PCI
PCI Express x16	16 pares	2,5 Gts	4 GBps	
PCI Express x32	32 pares	2,5 Gts	8 GBps	
PCI-X 1.0	64 bits	66 MHz	528 MBps	Se utiliza en servidores y estaciones de trabajo de alto rango. Es mucho menos popular que PCI Express.
	64 bits	100 MHz	800 MBps	
	64 bits	133 MHz	1,06 GBps	
PCI-X 2.0	64 bits	266 MHz	2,13 GBps	
	64 bits	533 MHz	4,26 GBps	
AGP	32 bits	66 MHz	264 MBps	Un estándar en tarjetas de video
AGP Modo 2 ó 2x	32 bits	66 MHz	528 MBps	con aceleración 3D
AGP Pro (4x)	32 bits	66 MHz	1 GBps	ahora reemplazado por PCI Express
AGP Pro (8x)	32 bits	66 MHz	2 GBps	

MBps = Megabytes por segundo; GBps = Gigabytes por segundo; Gts = Gigatransferencias

Los Buses Antiguos y Extinguidos

A lo largo de la historia, los buses se han visto obligados a conseguir mejorar su rendimiento para acompañar el aumento de la velocidad de trabajo de los microprocesadores, y dejar así de ser los cuellos de botella de los sistemas de computación.

Primero, fueron aumentando la cantidad de datos que transmitían en una misma unidad de tiempo. De 16 bits pasaron a 32 bits y luego a 64 bits. Por otro lado, incrementaron su velocidad máxima de trabajo, también para conseguir transportar mayor cantidad de información en menos tiempo. Sin embargo, esto no fue suficiente para dejar de ser cuellos de botella para los subsistemas de entrada y salida de datos, sino que se replantearon completamente sus arquitecturas para adecuarse a las nuevas tecnologías de transmisión de datos serie a altísimas velocidades utilizando pares diferenciales, dando origen al más simple, pero no por ello menos revolucionario, de todos los buses: PCI Express.

El primer sistema de bus utilizado en una PC constaba de 62 líneas de comunicación ó 62 conectores en una ranura corta, que aceptaban 8 bits de datos y 20 bits de direcciones, pues el procesador para el cual estaba diseñado era el 8088, que tenía un bus de datos de ese mismo número de bits. Solamente 8 de esos 62 conductores eran de datos y otros 20 formaban el bus de direcciones, los demás servían para otros propósitos, como la reducción del ruido.

En 1984, nació el bus AT, que expandió el sistema de bus original hacia un ancho de datos de 16 bits y 24 bits de direccionamiento, de esta forma se eliminaba el límite de los 8 bits impuesto por el bus anterior. Al ser una expansión al bus anterior, es totalmente compatible con las placas de 8 bits, con una ranura de expansión idéntica, pero con el agregado de otra más con los conductos de datos y direcciones adicionales necesarios para alcanzar los nuevos valores. Funcionaba a una frecuencia de reloj de 8 MHz, con una transferencia de datos máxima de 8 megabytes por segundo, más que suficiente para esa época (ver la Tabla 3.1).

Cuando los procesadores tenían velocidades de reloj bajas, el bus podía trabajar a la misma velocidad del procesador, pero a medida que fue aumentando dicha velocidad surgieron problemas para trabajar con el sistema del bus a la misma frecuencia. Es por eso que cuando Compaq introduce su AT de 12 MHz al mercado, lo hace con un bus que trabajaba a 8 MHz. Todos copiaron la idea y así nació el estándar en sistemas de bus de 16 bits, que se denominó de la misma forma que el bus anterior de 8 bits: ISA (*Industrial Standard Architecture - Arquitectura industrial estandarizada*). Cuando se habla del bus ISA, se hace referencia al bus AT de 16 bits (ver la Fig. 3.2).

Este bus se ha extinguido hace bastante tiempo, pero fue quien marcó la arquitectura de buses de expansión de las PC por más de una década.

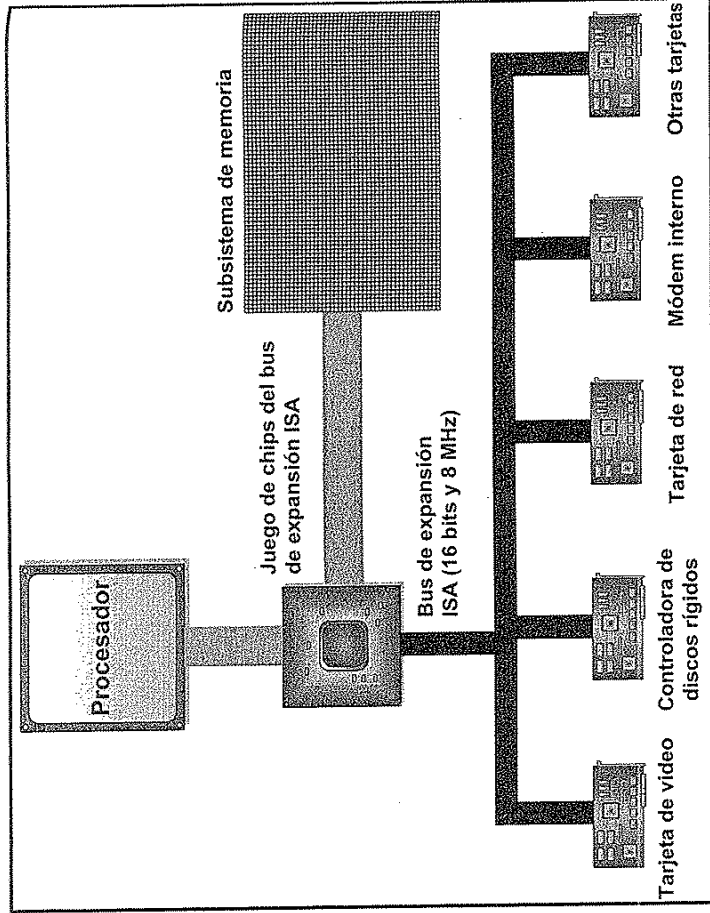


Fig. 3.2. Arquitectura del extinguido bus ISA de 16 bits.

Con la llegada de los procesadores de 32 bits comenzaron a aparecer ranuras específicamente diseñadas para trabajar con el nuevo ancho de datos, pero no eran estándar y cada fabricante la adaptaba a su manera. De esta forma, solamente se conseguían placas para esos conectores de los fabricantes de la motherboard, limitando las posibilidades de ampliación. Era necesaria una normalización, que llegó entre 1987 y 1988 con los buses MCA (*MicroChannel Architecture* - Arquitectura Microcanal) y EISA (*Enhanced Industrial Standard Architecture* - Arquitectura industrial estandarizada y ampliada). Cinco años más tarde, en 1993, se desarrollaron el VESA Local Bus 1.0 y el PCI 1.0. En la actualidad, el único que sigue vigente de estos cuatro buses es el PCI en una versión más avanzada, como veremos más adelante.



Los otros buses mencionados, si bien implementaron un ancho de datos de 32 bits, quedaron en desuso, por lo cual analizaremos en detalle los que están vigentes en las PC modernas.

Bus Mastering o Multiusuario

En las PC con un solo procesador, el bus es controlado por las acciones de éste. Pero, en PC con varios procesadores o con un único procesador con múltiples núcleos de ejecución, en las cuales cada uno de ellos se encarga de controlar tareas específicas, sería conveniente que todos los procesadores se comuniquen entre sí. Todos los buses modernos soportan esta capacidad de posibilitar el acceso de varios procesadores a un mismo bus.

El principio del *bus mastering* es que cualquier dispositivo, procesador o núcleo de procesamiento, puede pedir permiso para tomar el control del bus en un período de tiempo corto, el procesador principal cede entonces el control del bus al procesador, dispositivo o núcleo de procesamiento que le pidió hacerlo, acelerando la comunicación de éste con el sistema de E/S correspondiente, al llevarla a cabo en forma directa.

Esto se puede ejemplificar de la siguiente manera, si un señor tiene que ir a un lugar determinado y llama por su teléfono celular a su hijo para que lo guíe fijándose en un plano y se detiene en cada esquina para preguntarle por dónde debe seguir, va a tardar más tiempo que si el hijo que sabe el camino toma el auto y va directamente. Esto mismo es lo que sucede con el *bus mastering* o capacidad multiusuario del bus, si el dispositivo que toma el control del mismo tiene la capacidad suficiente para controlar el bus, no va a quitarle tiempo al procesador.

Soporte de Configuración por Software

Todos los buses modernos están ideados para soportar la configuración de las tarjetas mediante programas y no mediante *jumpers* ni interruptores en la misma placa. Este concepto permite que se enchufe la tarjeta nueva, se corra un programa y éste detecte automáticamente la configuración más adecuada para la misma, de manera tal que no hay que luchar contra los interruptores o *jumpers* ni sacar y volver a enchufar la placa reiteradas veces hasta encontrar la configuración adecuada.

El Problema del Ruido

A medida que se aumenta la cantidad de datos transferidos por un sistema de bus en una misma unidad de tiempo, éstos pueden ser afectados por el ruido (toda señal eléctrica que no sea información). Ver la Fig. 3.3.

El bus ISA utiliza la activación de interrupciones por marginalidad, esto quiere decir que cuando el nivel de la señal transferida por el bus excede cierta tensión (nivel de amplitud de la señal), esta información será interpretada como datos. Este tipo de activación de interrupciones tiene la desventaja de ser afectada por el ruido

y ante la posible generación de transitorios (picos de tensión de muy corto tiempo) éstos se interpretan como datos y generan errores considerables.

La solución implementada inicialmente en los antiguos y extinguidos buses EISA y MCA para resolver este problema era reemplazar el sistema de activación de interrupciones en forma marginal por el sistema de activación por nivel. Este sistema requiere que el nivel de tensión se mantenga alto durante la secuencia de transmisión de datos para que sean tomados como tal. De esta manera se elimina la probabilidad de tomar como inicio de una cadena de datos un pico de tensión. Ver la Fig. 3.3. Los buses PCI y PCI-X, en sus diferentes versiones siguieron utilizando este esquema.

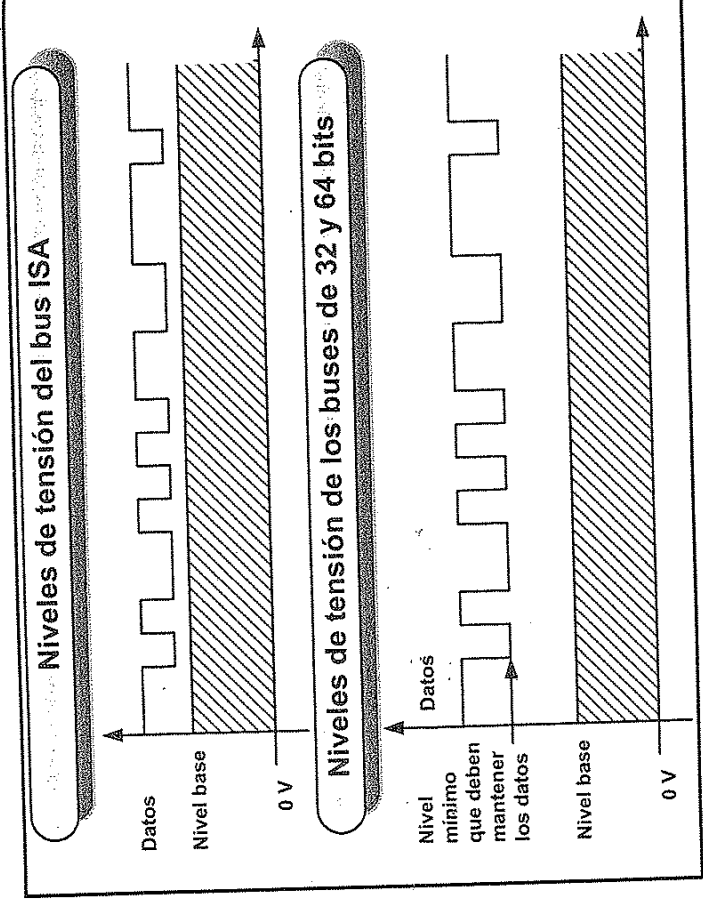


Fig. 3.3. Comparación de los niveles de tensión utilizados por diferentes buses.

Buses de 32 y 64 bits

Los buses PCI 1.0 y AGP trabajan con un ancho de datos de 32 bits en todas sus versiones.

Los buses PCI 2.x y PCI-X trabajan con un ancho de datos de 64 bits en todas sus versiones.

A continuación, analizaremos en detalle sus características.

PCI de 32 y 64 Bits (1.0; 2.x)

Este sistema de bus fue diseñado por Intel en 1993 como base de sistema de bus para el Pentium. PCI (*Peripheral Component Interconnect* - Interconexión de Componentes Periféricos) fue ideado para acomodar procesadores más rápidos sin necesidad de actualizar constantemente los dispositivos y rediseñarlos.

La arquitectura del bus PCI consiste en establecer un puente (*PCI Bridge*) entre el bus local y el procesador con un *buffer* (memorias temporarias) FIFO (*First In, First Out* - el primero que entra, el primero que sale). Intel lo denomina un bus intermedio porque está diseñado para desacoplar al procesador del bus de expansión mientras se mantiene un ancho de datos de 32 ó 64 bits y una velocidad de 33 MHz para comunicarse con los dispositivos. Ver la Fig. 3.4.

Esto es así porque es muy costoso diseñar y construir una motherboard cuyas ranuras de expansión trabajen a tan altas frecuencias como las que ha alcanzado el FSB de los procesadores actuales. Además, si esto se hiciera, los fabricantes deberían desarrollar tarjetas de expansión que pudieran trabajar a estas nuevas velocidades. Es por ello que el bus PCI cumple su función de desacoplar y no siguió con la arquitectura del VL-Bus, que solamente funcionó con un procesador 486.

Un Core 2 Duo tiene un bus local de 1066 MHz, mientras que las tarjetas conectadas a las ranuras de expansión del bus PCI trabajan a 33 MHz, a una frecuencia 32,30 veces menor que el bus local del procesador (1066 / 32,30 ≈ 33).

El bus PCI tiene las siguientes características y limitaciones:

- Permite un máximo de 10 dispositivos trabajando sobre el mismo bus PCI (incluyendo el controlador de PCI y un controlador de bus de expansión ISA, EISA o MCA opcional, que no se usa en los equipos actuales). En la mayoría de las motherboards modernas, ya no se incluyen ranuras ISA.
- Originalmente estaba limitado a 33 MHz. Sin embargo, algunas motherboards ofrecen la posibilidad de aumentar en algunos puntos la velocidad del trabajo del bus PCI, llegando a una máxima de 40 MHz. Hace un par de años apareció la versión de 66 MHz, que nunca llegó a popularizarse. Ver la Tabla 3.1.
- El ancho de los datos puede ser de 32 bits o de 64 bits (Versión 2.0, desarrollada en 1994). Esta última versión tuvo poca repercusión.

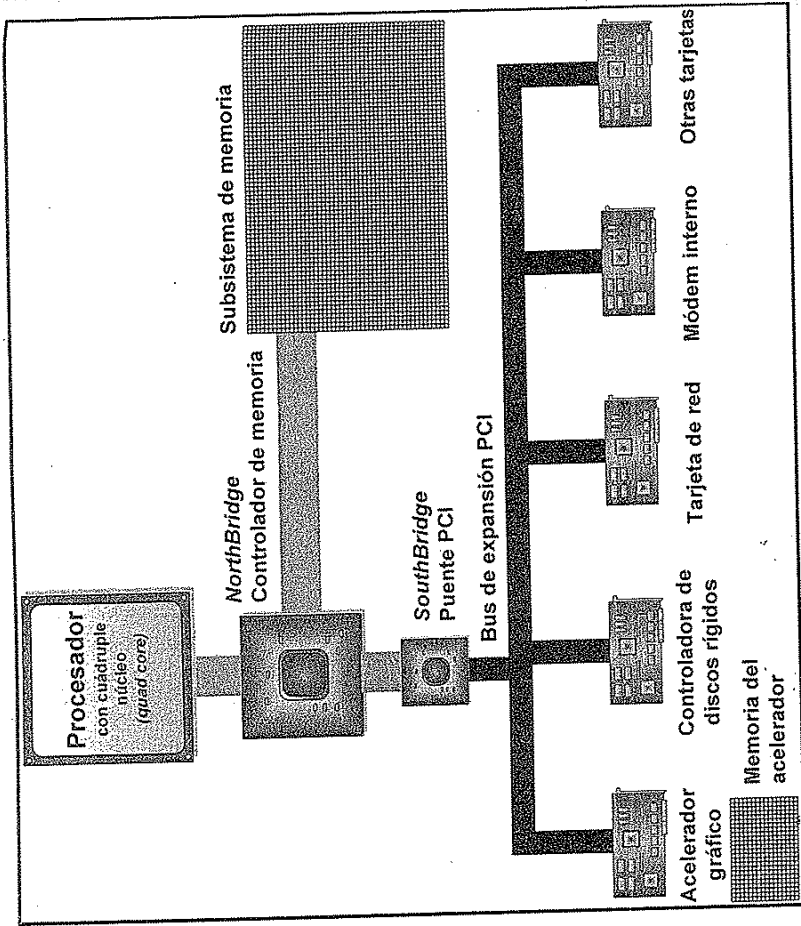


Fig. 3.4. Arquitectura del bus de expansión PCI.

- Permite al procesador efectuar sus tareas al mismo tiempo que se realiza el *bus mastering* con otros dispositivos, una ganancia de rendimiento que resultó muy atractiva para las aplicaciones multimedia y servidores de alto rango.
 - Habilita un modo rápido para lecturas y escrituras (VL-Bus solamente lo permitía en lecturas).
- El diseño de PCI requiere que los dispositivos devuelvan estados de configuración al controlador. También utiliza la técnica de multiplexado para enviar más de una información en un solo canal eléctrico, con el beneficio de la reducción de los pines necesarios para comunicarse con los chips.

El sistema PCI utiliza un diseño eléctrico de onda reflectiva, de manera tal que se necesita la mitad de amplificación para el pasaje de señales a través del bus. Esto soluciona algunos problemas de inductancias parásitas y crea menor ruido.

Este sistema fue introducido con las primeras PC basadas en el Pentium y se transformó rápidamente en el estándar para los siguientes microprocesadores. Su rendimiento es similar al extinguido VL-Bus: la máxima transferencia de datos es de 132 MBps para PCI 1.0 con 32 bits de ancho de datos y 33 MHz de velocidad de reloj y de 246 MBps para PCI 2.0/2.1 con 64 bits y 33 MHz.

Ahora, con sus implementaciones de 64 bits y 66 MHz, alcanza los 528 MBps. Actualmente PCI es un estándar y está presente en todas las motherboards modernas, sin embargo está siendo reemplazado por PCI Express y PCI-X (ver las Figs. 3.5 y 3.6).

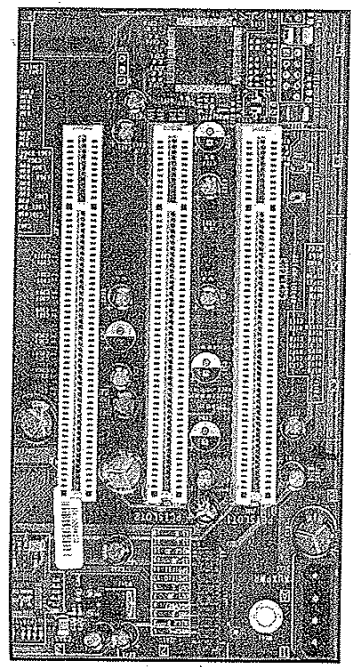


Fig. 3.5. Ranuras de expansión PCI.

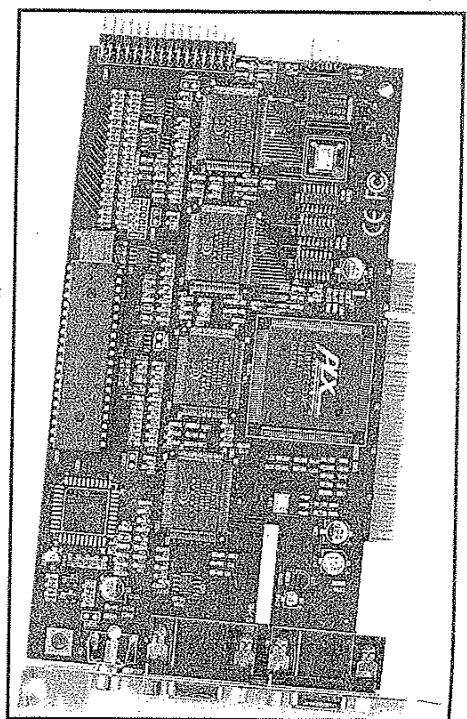


Fig. 3.6. Una tarjeta DVR para el bus de expansión PCI.

PCI-X 1.0 y 2.0

La aparición de procesadores, dispositivos y buses de entrada y salida más veloces y a su vez demandantes hizo que cada vez, los humildes 32 bits y 33 MHz del PCI más difundido no fueran suficientes para seguirle el ritmo a las nuevas tecnologías. El bus PCI convencional o bien el de 64 bits representan un enorme cuello de botella para las velocidades de bus local de los procesadores de última generación. Es así como surgió una nueva mejora al bus PCI para aumentar su rendimiento y hacerlo trabajar a tres velocidades de reloj: 66; 100 y 133 MHz, todas con un ancho de datos de 64 bits. Esta nueva versión se conoce con el nombre de PCI-X (versión 1.0). Con una velocidad de reloj de 133 MHz, alcanza una transferencia de datos máxima de 1,06 GBps (ver la Tabla 3.1).

La arquitectura de PCI-X es idéntica a la original del bus PCI

convencional. PCI-X 1.0 es totalmente compatible con los modos de trabajo y las tarjetas PCI convencionales (1.0 y 2.x), por lo cual se puede utilizar una ranura PCI-X para conectar una tarjeta PCI y ésta funcionará sin inconvenientes.

Las ranuras PCI-X incorporan una extensión de 64 bits (ver las Figs. 3.7 y 3.8) y tienen el mismo formato que las PCI de 64 bits. La diferencia radica en que está habilitada para trabajar con mayores velocidades.

Las tarjetas desarrolladas para un bus PCI de 64 bits y 66 MHz insertadas en una ranura PCI-X habilitada para trabajar en 133 MHz, funcionarán a sólo 66 MHz, debido a que no están diseñadas para los modos PCI-X.



Fig. 3.7.
Una ranura PCI-X, con la extensión de 64 bits.

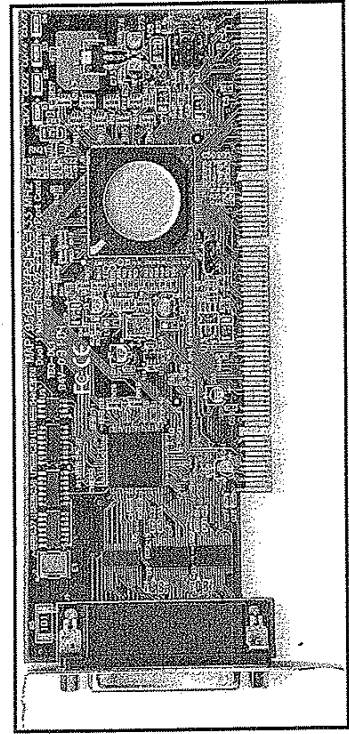


Fig. 3.8.
Una tarjeta con puertos para el bus PCI-X, con la extensión de 64 bits en el conector.

Al poco tiempo surgió una nueva versión, PCI-X 2.0, la cual agregó dos nuevas velocidades de trabajo con el mismo ancho de datos de 64 bits: 266 y 533 MHz, alcanzando velocidades máximas de transferencia de datos de 2,13 y 4,26 GBps, respectivamente (ver la Tabla 3.1). PCI-X 2.0 también es compatible hacia atrás con PCI-X 1.0 y por lo tanto con el bus PCI convencional y utiliza las mismas ranuras de expansión PCI de 64 bits (ver las Figs. 3.7 y 3.8).

PCI-X 2.0 utiliza la tecnología DDR (*Double Data Rate* – Doble tasa de transferencia de datos) para conseguir duplicar los 133 MHz de PCI-X 1.0 y llegar a los 266 MHz. Y hace uso de la tecnología QDR (*Quadruple Data Rate* – Cuádruple tasa de transferencia de datos) para cuadruplicar los 133 MHz y llegar a los 533 MHz. Esto significa que con una velocidad de reloj real de 133 MHz, envía el doble o el cuádruple de transferencias de datos por cada ciclo de reloj y así consigue elevar la velocidad final (ver la Fig. 3.9).

PCI-X 2.0 también agrega nuevas funcionalidades al bus, además de simplemente incrementar la velocidad de trabajo. Una de las características más interesantes, además de los agregados en corrección de errores, es la implementación de transacciones basadas en mensajes, mediante las cuales los dispositivos conectados al bus se pueden enviar varios mensajes entre sí con esquemas transaccionales. Algo que si bien resulta complejo de implementar puede dar origen a hardware más inteligente y con menos necesidades de utilización del procesador.

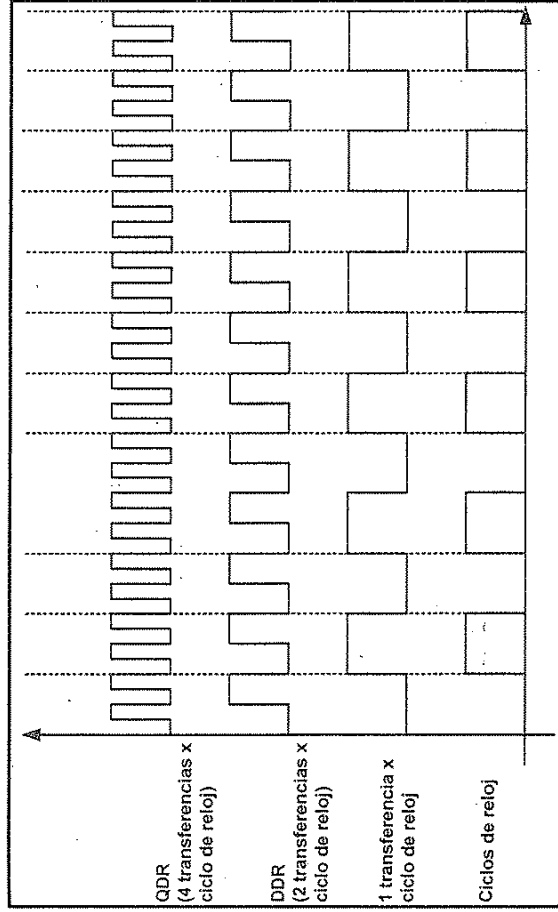


Fig. 3.9. Diferentes modos de transferencia de datos para un mismo ciclo de reloj.

AGP y AGP Pro

El bus PCI de 32 bits a 33 MHz, con un máximo de 132 MBps, desperdiciaba la potencia de los procesadores gráficos de alto rendimiento y generaba un gran cuello de botella. Para solucionar este problema, Intel desarrolló una nueva interfaz entre el bus del sistema y la tarjeta de video: el AGP (*Accelerated Graphics Port* – Puerto de Gráficos Acelerado). Este desarrollo se basó en la especificación de un bus PCI de 32 bits funcionando a 66 MHz, el cual permite una velocidad de transferencia de datos máxima de 264 MBps, conocido como Modo 1 (ver la Tabla 3.1 y las Figs. 3.10; 3.11 y 3.12).

A dicha especificación se le agregaron varias mejoras para obtener un mayor rendimiento en las tarjetas gráficas y así poder ofrecer una nueva plataforma para desarrollar nuevas aplicaciones con gráficos 3D que consigan altísimos rendimientos. Dichas mejoras se detallan a continuación:

- Ejecutar operaciones de lectura y escritura de memoria en paralelo, permitiendo aprovechar el período de latencia del acceso a memoria para seguir enviando pedidos.
- Utilizar un mecanismo de multiplexado para transmitir datos y direcciones en el bus y así incrementar la eficiencia del mismo.
- Utilizar el flanco de subida y de bajada del reloj para sincronizar la transferencia de datos, logrando duplicar la velocidad del bus para obtener una transferencia de hasta 528 MBps, conocido como Modo 2 ó 2x.
- Incorporar al bus en la segunda versión de AGP, conocida como AGP Pro, el funcionamiento con una nueva especificación eléctrica de baja tensión, que permite cuatro transferencias de datos por cada pulso del reloj, aumentando la velocidad de transferencia de datos del bus hasta 1 GBps. A este nuevo modo se lo conoce como Modo 4x.
- Aumentar nuevamente la velocidad de transferencia de datos hasta 2 GBps, en la tercera versión de AGP, siendo ésta la segunda de AGP Pro. Presentó el nuevo Modo 8x.

AGP fue diseñado teniendo en mente su funcionamiento en conjunto con la arquitectura DIB, que se presenta en los Pentium II o superiores, pues está preparada para que, mientras el procesador principal efectúa operaciones de punto flotante, el acelerador AGP puede tomar texturas de la memoria principal.

Las tarjetas AGP y AGP Pro ofrecen un rendimiento muy superior a las que se conectan a un bus PCI en las aplicaciones con gráficos 3D y en la reproducción de video cuando se tienen los *drivers* adecuados y las aplicaciones están optimizadas para este tipo de tarjetas. Durante varios años fue el estándar para tarjetas de video aceleradas, pero se lo está reemplazando por PCI Express 16x y 32x.

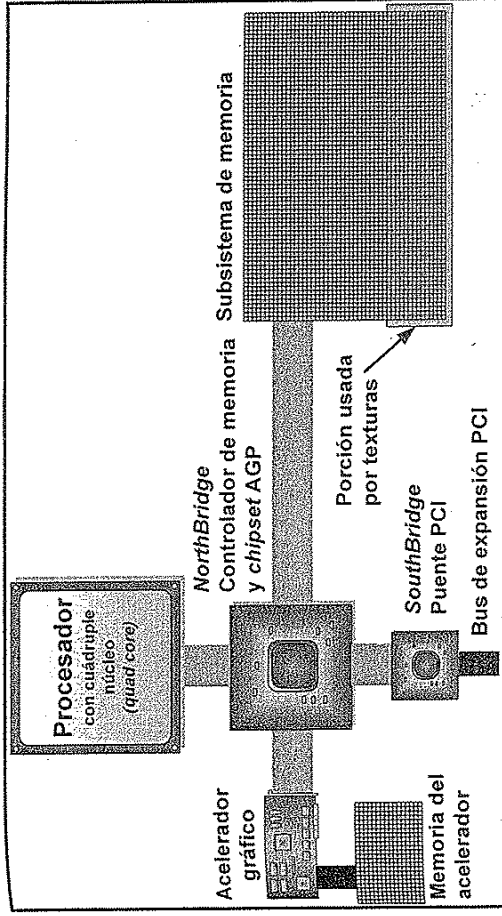


Fig. 3.10. La arquitectura de AGP y la ubicación del acelerador gráfico.



Fig. 3.11. Una ranura AGP en una motherboard.

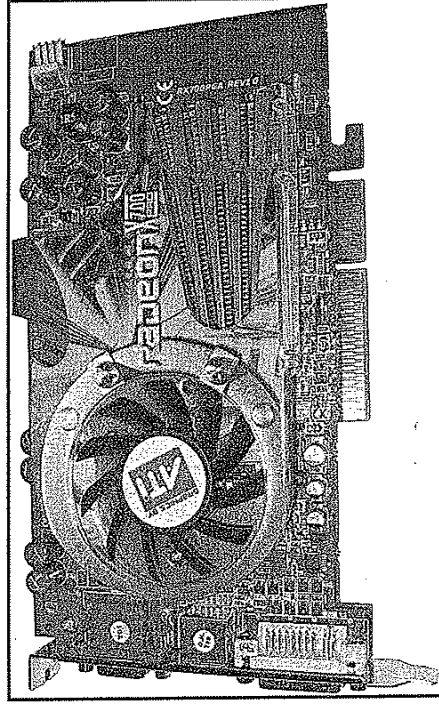


Fig. 3.12. Una tarjeta de video AGP Pro.

Las tarjetas AGP y AGP Pro (ver la Fig. 3.12) pueden utilizar su propia memoria de video como el resto de las tarjetas y al tener una conexión directa con el bus del sistema, pueden utilizar sin problemas la memoria del sistema sin necesidad de

molestar al procesador para *Z-buffering*, almacenamiento temporal de texturas o también para ampliar la memoria de vídeo de la tarjeta y poder trabajar en resoluciones más altas.

PCI Express 1x...32x

Los buses PCI, AGP y PCI-X han alcanzado sus límites de rendimiento y se hizo complicado seguir aumentando la velocidad de transferencia de datos y bajar la tensión de trabajo manteniendo sus arquitecturas de buses paralelos, para seguirle el ritmo al avance en la velocidad de trabajo de los microprocesadores y de las memorias.

Es así como surgió la necesidad de tener un sistema de bus de tercera generación con una nueva arquitectura que pueda escalar fácilmente para seguirle el ritmo al avance de los microprocesadores y las memorias y a las crecientes necesidades de mayor ancho de banda en las comunicaciones entre estos componentes y las tarjetas de E/S.

El bus PCI Express, también conocido como PCIe y 3GIO, a diferencia de todos los buses analizados hasta ahora, es un bus serie con una cantidad de pares diferenciales (sendas de transmisión y recepción de datos) escalables para multiplicar fácilmente la velocidad de transferencia de datos.

Este bus tiene las siguientes ventajas comparado con sus predecesores:

- Logra velocidades de transferencia de datos de 250 MBps hasta 8 GBps en cada dirección (ver la Tabla 3.1), según la cantidad de pares diferenciales que utilice y puede escalar fácilmente agregando nuevos pares diferenciales y seguir multiplicando las velocidades sin tener que modificar la arquitectura básica y sin que aparezcan problemas de ruido o limitaciones físicas, como sucedió con los buses cuando se intentaba incrementar la cantidad de bits transmitidos en paralelo. A diferencia de los buses paralelos, las velocidades son para cada dirección dado que es un bus *full-duplex*, por lo cual como puede transmitir y recibir datos en forma simultánea, la velocidad máxima se multiplica por 2, por lo cual, va desde 500 MBps hasta 16 GBps.
- Presenta una arquitectura de E/S unificada para PC de diferentes rangos, así como para dispositivos móviles y sistemas industriales y embebidos.
- Los costos de las tarjetas y los dispositivos, tanto como del bus, son iguales o aún menores que los de otros buses.
- La configuración de los dispositivos y los *drivers* son totalmente compatibles con la arquitectura de PCI, por lo tanto, el soporte de los sistemas operativos y del software está garantizado.

- El rendimiento que se puede conseguir del bus puede escalar con el tiempo, agregando nuevos pares diferenciales para la transmisión de datos y aumentando la frecuencia de trabajo.
- Tiene una menor sobrecarga que PCI y PCI-X, pues utiliza un conmutador que establece una conexión punto a punto entre los dispositivos que se quieren comunicar, en vez de tener que pasar por un puente (*bridge*).
- Ofrece características avanzadas como la estructuración de diferentes tipos de datos, la administración de energía, la seguridad de la calidad del servicio (QoS - *Quality of Service*), la integridad de datos y el manejo de errores, el soporte para conexión en caliente (*hot plug*) y el intercambio en caliente (*hot swap*). También, ofrece el soporte Plug & Play igual que lo hacía el bus PCI y con las mismas interfaces para administrarlo.
- Soporta transmisión isócrona, por lo cual es posible garantizar un ancho de banda determinado y asegurar el tiempo que durará la misma, para conseguir, por ejemplo, una transmisión de datos en tiempo real.

La arquitectura consiste en establecer un bus serie punto a punto entre dos dispositivos, con uno o más grupos de dos pares diferenciales que permiten la transmisión y recepción de datos cada uno de ellos. Un conmutador (*switch*) es el encargado de establecer la comunicación entre los dos dispositivos que se quieren comunicar, como el procesador y una tarjeta de vídeo, habilitando y coordinando las transferencias de datos a través de los pares (ver la Fig. 3.13). De esta manera, se pueden agregar múltiples conmutadores para que no se transformen en un cuello de botella, especialmente en los subsistemas de E/S que requieren los mayores anchos de banda.

Cada par diferencial trabaja a una frecuencia de 2,5 Gigatransferencias por segundo por dirección (envío o recepción de datos), también conocidas como Gts. Sin embargo, al igual que sucede con Serial ATA, como veremos más adelante en el

Capítulo 10: Discos Rígidos, se debe tener cuidado a la hora de evaluar la velocidad de este bus. Pues, sucede que si bien hay 2,5 Gts por cada par, como los datos se transmiten utilizando la codificación 8b/10b, la eficiencia del uso del canal es del 80%. Por lo tanto, si bien podemos decir que el ancho de banda inicial de cada par diferencial es de $2,5 \text{ Gbps} = 2.500.000.000 \text{ bits por segundo} / 8 = 312.500.000 \text{ bytes por segundo}$, su eficiencia es del 80%, por lo cual, la velocidad de transferencia de datos máxima real de cada par diferencial será de $312.500.000 \text{ bytes por segundo} \times 0,8$ (80% de eficiencia por usar la codificación 8b/10b) = $250.000.000 \approx 250 \text{ MBps}$.

La señal codificada en 8b/10b contiene un reloj embebido (*embedded clock*) para facilitar la sincronización durante las transmisiones a muy altas velocidades y permite conseguir distancias entre los componentes de hasta 50 cm.

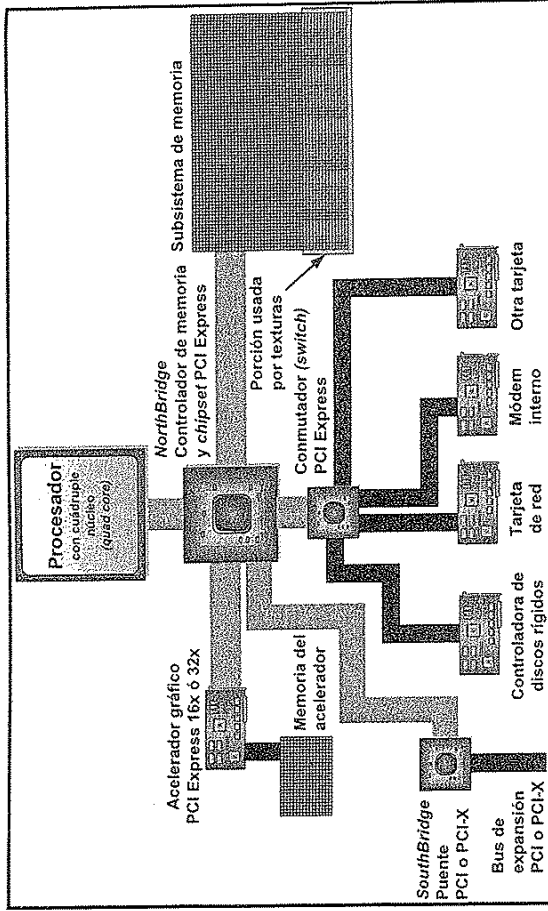


Fig. 3.13. La arquitectura de PCI Express (observar la diferencia con los buses analizados anteriormente).

Como la arquitectura del bus está preparada para que se consigan unas 10 Gts, todavía se puede esperar una gran escalabilidad para el bus PCI Express.

Tabla 3.2. Conectores para el bus PCI Express.

PCI Express	Cantidad de pines	Compuesto por	Admite tarjetas PCI Express
1x	36	Conector PCI Express 1x	1x
4x	64	Conector PCI Express 1x + extensión 4x	1x y 4x
8x	98	Conector PCI Express 1x + extensión 4x + extensión 8x	1x, 4x y 8x
16x	164	Conector PCI Express 1x + extensión 4x + extensión 8x + extensión 16x	1x, 4x, 8x y 16x
32x	294	Conector PCI Express 1x + extensión 4x + extensión 8x + extensión 16x + extensión 32x	1x, 4x, 8x, 16x y 32x

El bus utiliza un conector diferente al del bus PCI, pues trabaja con conexiones en serie diferenciales. A su vez, los conectores PCI Express varían su tamaño y cantidad de pines de acuerdo a la velocidad para la cual están preparados para trabajar y se resumen en la Tabla 3.2. Estos conectores se van conformando agregando bloques de pines, de manera tal que son compatibles hacia atrás. Por

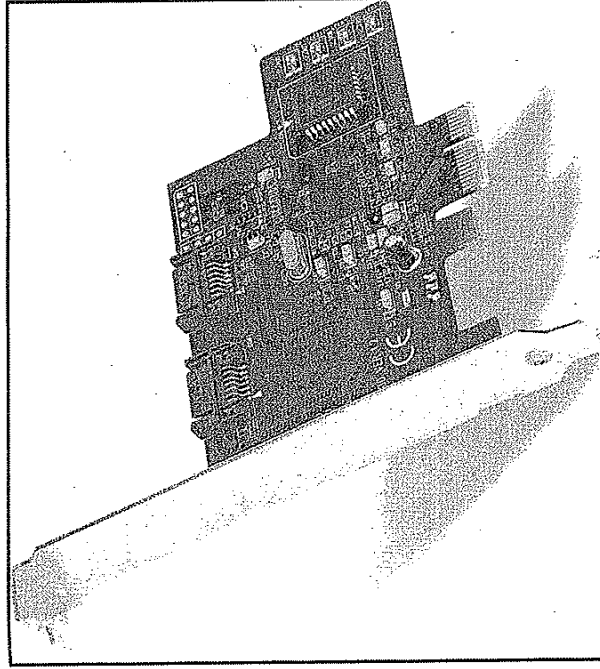


Fig. 3.14. Una tarjeta con puertos USB 2.0 para el bus PCI Express 1x.

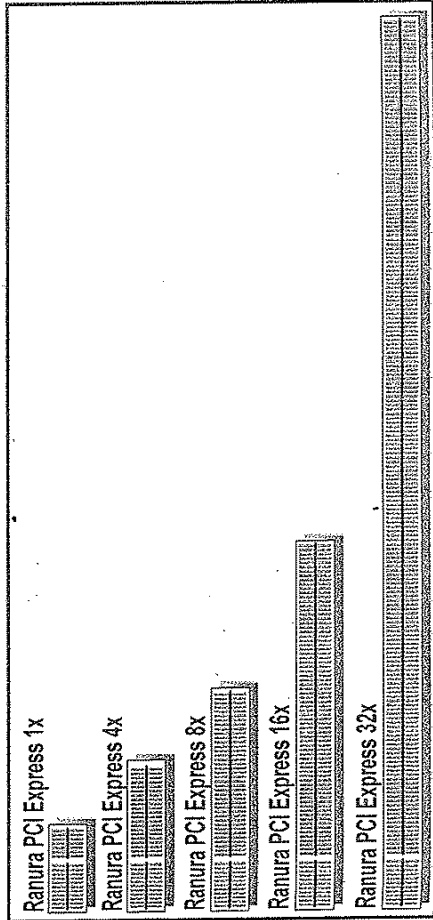


Fig. 3.15. Esquema de las ranuras PCI Express 1x; 4x; 8x; 16x y 32x.



El bus PCI Express es el sucesor de PCI y está presente en la mayoría de las motherboards modernas.

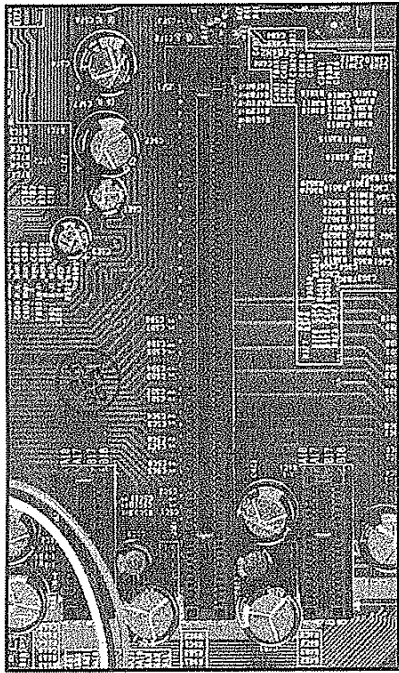


Fig. 3.15.
Ranuras PCI Express 1x y 16x en una motherboard.

Otros Buses

Además de todos los que hemos analizado en este capítulo, existen otros buses que serán explicados en el **Capítulo 6: Puertos y Buses de E/S**, debido a que están asociados a otros temas. Estos son el USB y el IEEE 1394 ó FireWire.



Para mayor información sobre la configuración, la optimización y la solución de problemas con los buses analizados anteriormente y su estructura eléctrica, puede consultar el libro **Reparando y Configurando Motherboards de PC**, de Editorial HASA.

Capítulo 4



La Memoria

La memoria es uno de los componentes fundamentales de las computadoras, pues sin ellas, éstas no podrían procesar información de ninguna manera porque no tendrían un medio de almacenamiento de información temporal.

La memoria principal es el medio de almacenamiento temporal en el que la CPU puede escribir, leer o modificar información. A esta memoria se la conoce como RAM (*Random Access Memory* - Memoria de acceso aleatorio).

Se puede considerar a la memoria como una gigantesca cajonera en donde, cada casilla representa un byte y tiene una dirección. La CPU puede leer, escribir o modificar la información en cada una de las casillas, pero para ello debe indicar la dirección de la misma y la información que desea depositar o leer, es por eso que se la llama de acceso aleatorio. Ver la Fig. 4.1.

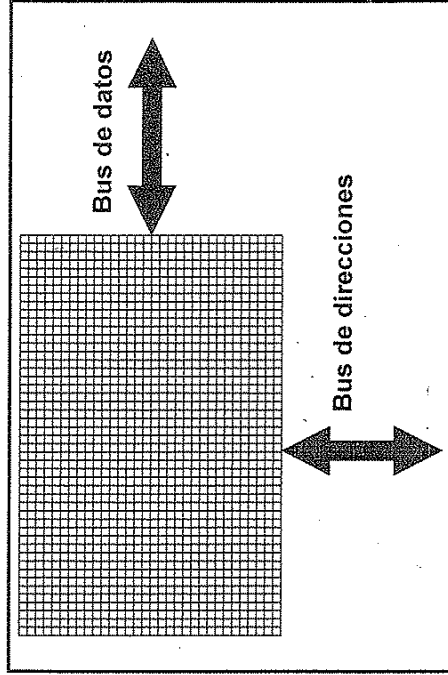


Fig. 4.1.
La memoria se puede comparar con una gigantesca cajonera.

La memoria principal tiene un tiempo de acceso bajo, algunas decenas de nanosegundos (1×10^{-9} segundos), es decir, que se tarda ese

tiempo en encontrar la casilla por su dirección y actualizar la información existente en la misma por otra. La velocidad de la memoria tendrá que ser adecuada para la CPU con la cual se trabaje.

Cuando llamamos temporal a la memoria principal, significa que al apagar la computadora, se perderá toda la información almacenada en ésta. Cuando un programa se ejecuta, se cargan partes del mismo en la memoria para alimentar al procesador con las instrucciones a seguir. Cuando se necesita que la información se mantenga constante, se utilizan los dispositivos de almacenamiento, por ejemplo: los discos rígidos, unidades USB, CD-ROM, DVD, HD DVD, etc.



¿Cuánta Memoria Necesito?

Hace algunos años, la mayoría de las PC venían con entre 256 y 512 MB de RAM, luego 1 GB pasó a ser un estándar, pero los nuevos sistemas operativos y las aplicaciones de hoy en día determinaron que se deba tener un equipo con 1,5 GB, como mínimo, para poder ejecutar a estos últimos. La cantidad de memoria que se necesita depende, de manera proporcional, del volumen de información con la cual se va a trabajar. Si se tienen muchas aplicaciones corriendo al mismo tiempo y cada una de ellas trabaja con un volumen de datos importante, se necesitará más memoria que si se utiliza solamente un procesador de textos y una planilla de cálculos con poca información.

Las aplicaciones que combinan audio, video y animación computada, como las que abundan en Internet y los juegos, requieren mucha capacidad de la CPU, así como una buena cantidad de memoria para poder trabajar cómodamente. Un equipo destinado a trabajar con video, sonido y animaciones de mediano rango, debe disponer como mínimo de 1,5 GB de memoria y un procesador con dos núcleos de ejecución de 2 GHz cada uno, como mínimo (u otro de similar rendimiento).

Para saber la cantidad de memoria que se necesita, hay que guiarse por la siguiente premisa: Cuanto más, mejor, pero no vale la pena que sobre demasiado, y se debe acordar del bolsillo, pues la memoria es uno de los componentes que más incide en el costo final de una computadora.

Los Límites en el Direccionamiento

La cantidad de memoria máxima direccionable está limitada por el tipo de procesador. Los antiguos procesadores de 32 bits (386DX o superiores) tenían un bus de direcciones de 32 bits, por lo tanto, estaban limitados a direccionar un máximo de 4 GB. En cambio, a partir del Pentium Pro, Pentium II y superiores pasaron a tener uno de 36 bits, capaz de direccionar hasta 64 GB.

Los procesadores AMD de 64 bits (AMD64) y luego la segunda generación de los Intel con extensiones EM64T, ampliaron el bus de direcciones a 40 bits, con una capacidad de direccionar hasta 1.024 GB (1 TB). Sin embargo, en modo real, es decir, durante el arranque del equipo, mantienen la compatibilidad con el manejo de memoria original del 8088 y hasta que pasan a modo protegido solamente pueden acceder a los primeros 640 KB.

Esta limitación se debe a que el antiguo 8088 tenía un bus de direcciones de 20 bits, por lo tanto podía direccionar solamente 1 MB. De ese megabyte, sólo los primeros 640 KB eran aprovechables por los programas para ejecutarse, mientras que los 384 KB restantes son utilizados por los fabricantes de computadoras para almacenar las operaciones de la ROM (*Read Only Memory* - Memoria de sólo lectura, pregrabada por el fabricante) del BIOS (*Basic Input/Output System* - Sistema básico de entrada y salida), la memoria de video, etc. Por esa razón, esta sección de la memoria no se puede utilizar por programas, aunque quedan huecos de memoria sin utilizar que sí es posible aprovecharlos. Esto se debe a que todos los procesadores siguen siendo compatibles con el 8088 original y arrancan en modo real como analizaremos más adelante.

La memoria queda formada de la siguiente manera: un área de memoria plana de 640 KB, un par de huecos de memoria libres y luego otra zona de memoria plana, que será la que va más allá de los 1024 KB (del primer megabyte). Ver la Fig. 4.2.

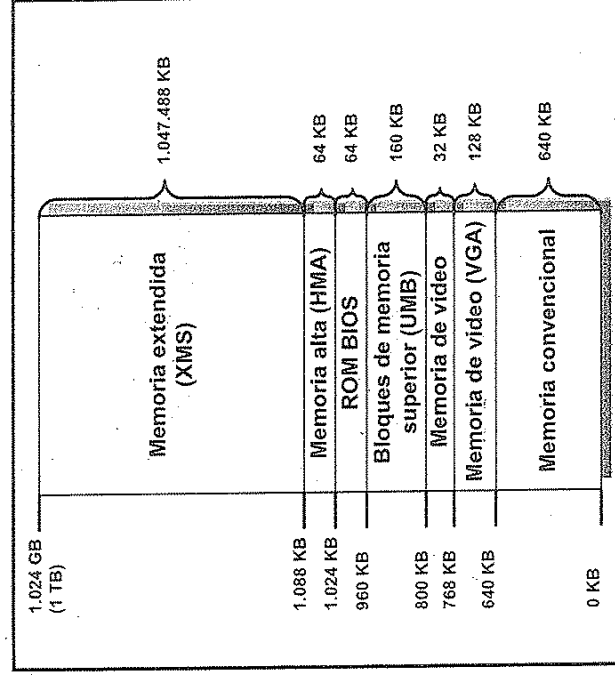


Fig. 4.2.
Mapa de memoria de una PC con un bus de direcciones de 40 bits.

La Memoria Convencional

Este tipo de memoria consiste en los primeros 640 KB. La utilizan los sistemas operativos para cargar su parte fija en la memoria, la cual les permite ejecutar el código necesario para poder acceder luego a las otras áreas de la memoria. Por más moderno que sea el sistema operativo, primero debe cargar y ejecutar código en los 640 KB iniciales. Es un área de memoria plana, accesible en forma directa, y utilizada para almacenar y ejecutar programas.

La Memoria Superior (UMB)

La memoria superior está compuesta por los huecos de memoria no utilizados entre los 640 KB y 1024 KB. Los sistemas operativos modernos como Windows Vista, XP, 2003 y Linux aprovechan estos huecos de memoria sin inconvenientes.

A los huecos libres se los conoce con el nombre UMB (*Upper Memory Blocks* - Bloques de memoria superior). En DOS existían administradores de memoria, cuya función era encargarse que los controladores cargados allí aprovecharan al máximo dichos huecos. Hoy ya no hace falta preocuparse por esos temas.

La Antigua Memoria Expandida

Como se ha visto anteriormente, para aprovechar las nuevas capacidades de direccionamiento de memoria, el procesador debe pasar al **modo protegido**.

En 1985, Lotus, Intel y Microsoft desarrollaron en conjunto un estándar para plaquetas y controladores de memoria expandida para eludir las limitaciones del DOS paginando la memoria. La versión 3.2 de LIM EMS (*Expanded Memory Specification* - Especificación de memoria expandida) permitió manejar hasta 4 MB de memoria expandida paginando en bloques de 64 KB en memoria superior.

La versión 3.2 permitía solamente almacenar datos. La versión 4.0, desarrollada en 1988, posibilitaba ejecutar programas y direccionar hasta 32 MB. Este tipo de memoria fue utilizado principalmente por las planillas de cálculo, que requerían grandes cantidades de memoria para procesar mucha información.



La forma de trabajo de este tipo de memoria consistía en tomar 64 KB de la **memoria superior** (bloque comprendido entre los primeros 640 KB y 1024 KB) para utilizarlos como 4 páginas de 16 KB que corresponden a otras direcciones más altas. Esto quiere decir que se puede tomar un bloque de 16 KB de memoria que se encuentre más allá del primer megabyte, copiar en una página existente en la memoria superior su contenido, allí modificarlo y luego reescribir el bloque con los cambios realizados en la posición original. De esta forma se logra la posibilidad de elegir un bloque de 16 KB de la memoria expandida y leer o modificar la

información en el mismo sin necesidad de acceder a esa dirección de memoria en forma directa, sino pidiendo al controlador que quiero modificar dicho sector y haciéndolo en el número de página que elijo (de 1 a 4) y los cambios que haga, automáticamente se reflejarán en la memoria expandida, que puede estar más allá de los 1024 KB.

Esto posibilitó que, sin salir del modo real compatible con el 8088, se pudiera tener acceso a más que 1 MB de memoria. La mayoría de los sistemas operativos modernos ofrecen modos de emulación de memoria expandida para aquellas aplicaciones DOS antiguas que la utilizaron. Ninguna aplicación moderna hace uso de este antiguo modo de direccionamiento, pues los procesadores actuales pueden ver la memoria en forma plana y no necesitan pasos intermedios.

La Memoria Extendida y el Modo Protegido

La **memoria extendida** es aquella que va más allá del primer megabyte y puede ser direccionada por procesadores 80286 ó superiores. Pero, para direccionar esa cantidad de memoria se debe pasar al **modo protegido**.

A partir del 80286, se amplió por primera vez el **bus de direcciones**, de manera tal que estos procesadores pudieran manejar mayor cantidad de memoria. Sin embargo, los requerimientos de compatibilidad con el 8088 hicieron que se mantenga un modo de compatibilidad y se limite el bus de direcciones a 20 bits cuando estos procesadores trabajen en el modo de **direccionamiento real**, reduciendo la cantidad de memoria máxima accesible a 1 MB.

Por lo tanto, para aprovechar la nueva capacidad de direccionamiento de los procesadores, éstos deben operar en el **modo protegido**, abandonando el modo real compatible con el 8088. En este modo, se utilizan a pleno las capacidades de direccionamiento de los mismos, utilizando la memoria extendida. Cuando el procesador se introduce en este modo de trabajo, es totalmente incompatible con sus predecesores (procesadores anteriores al 80286), pues se agrega la capacidad de protección de memoria.

La **protección de memoria** es un principio tomado de las grandes máquinas (*mainframes*), las cuales están preparadas para correr varias aplicaciones al mismo tiempo. Al hacer esto, cada una debe tener un límite de memoria para poder trabajar en ésta y tener cuidado que otra aplicación que se esté ejecutando al mismo tiempo no utilice una parte de la memoria de la otra, porque se produciría una mezcla de datos que nadie sabe en qué desastre puede terminar.

Cuando un programa trabaja en modo protegido, le pide un bloque de memoria al sistema operativo y luego accede directamente a ese bloque (ver la Fig. 4.3). Luego, otro programa pide una porción de memoria para comenzar a trabajar y el sistema operativo le da el bloque a continuación del que le asignó al anterior.

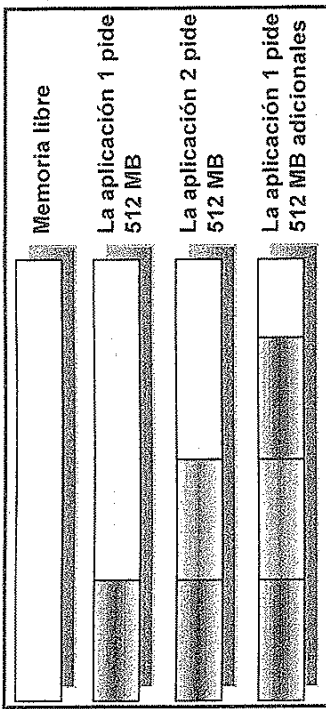


Fig. 4.3.
Respuesta del modo protegido ante las solicitudes de memoria de varias aplicaciones.

Si el primer programa vuelve a pedir un bloque de memoria más, el sistema operativo le dará uno nuevo a continuación del que le entregó al otro programa, pero si no hay suficiente memoria como para guardar un bloque, le contesta que no pudo completar la operación.

Ahora, qué sucede si uno de los programas se pasa al territorio del otro, es decir, accede directamente a la memoria fuera del bloque que reservó sin permiso alguno del sistema operativo: actúa la protección de memoria, el programa que violó los límites es detenido en su intento y, si es necesario, se termina su ejecución mientras las demás aplicaciones continúan funcionando perfectamente. En Windows aparecen los carteles de violación de acceso o de fallas en aplicaciones.

Los sistemas operativos modernos Windows Vista; XP; 2003, todas las distribuciones y versiones de Linux y Mac OS trabajan en modo protegido.

La **memoria alta** se encuentra entre los 1024 KB y los 1088 KB. En los sistemas operativos que trabajan en modo protegido se puede acceder para cualquier fin sin inconvenientes.

El Modo Largo de 64 bits

Los procesadores AMD de 64 bits (AMD64) y luego la segunda generación de los Intel con extensiones EM64T, agregaron un nuevo modo de trabajo al protegido, el **modo largo (long mode)**, el cual permite acceder al bus de direcciones extendido a 40 bits y mayores, con una capacidad de direccionar hasta 1.024 GB (1 TB) y utilizar las instrucciones y los registros de 64 bits que poseen estos procesadores.

Otros Modos de Trabajo

Existen otros modos de trabajo que surgieron para facilitar el funcionamiento de las antiguas aplicaciones DOS en modo protegido:

- Simulación de memoria expandida (LIM EMS).
- VCPI (*Virtual Controlled Program Interface* - Interfaz de programas con control virtual).
- DPMI (*DOS Protected Mode Interface* - Interfaz de Modo Protegido de DOS).

Actualmente, la mayoría de los sistemas operativos y programas funcionan en modo protegido o modo largo, pero uno nunca se termina de liberar de un programa que tenga un par de años, sea bastante útil y requiera memoria expandida para acceder a más de 640 KB.

Windows y cualquier otro sistema operativo que funcione en modo protegido y ejecute aplicaciones DOS permiten definir qué cantidad de memoria queremos destinar para que la aplicación la tenga disponible como memoria expandida.

Cualquier sistema operativo o administrador de memoria que brinde servicios VCPI cada vez que se ejecute un programa, verificará si hay memoria de algún tipo determinado disponible. Ante esta pregunta, le responde que están disponibles tanto memoria expandida como extendida. El programa toma la que quiere usar y el administrador se la suministra.

Windows en todas sus versiones y los sistemas operativos que permiten ejecutar aplicaciones DOS brindan servicios DPMI.

El Área de Memoria Reservada

El segmento de memoria entre los 640 KB y los 1024 KB es un área de memoria que en la PC original era una porción *memoria reservada* para que la utilice el hardware del sistema. Esta parte de la memoria se utiliza para:

- Memoria de video.
- Direcciones de memoria que necesitan algunas placas de expansión.
- La ROM BIOS.

La ROM BIOS

Se llama ROM (*Read Only Memory* - Memoria de sólo lectura) a un tipo de memoria de la que se pueden leer datos, pero no se pueden escribir nuevos ni modificar los existentes. Este tipo de memoria se encuentra pregrabada por el fabricante y contiene instrucciones para el manejo de algunas tarjetas o las **operaciones principales** de la PC. Este último tipo de ROM se llama BIOS (*Basic Input/Output System* - Sistema básico de entrada/salida), que es el traductor de los llamados del sistema operativo a los dispositivos de entrada/salida. El BIOS se ubica habitualmente en los últimos 64 KB de la **memoria reservada**.

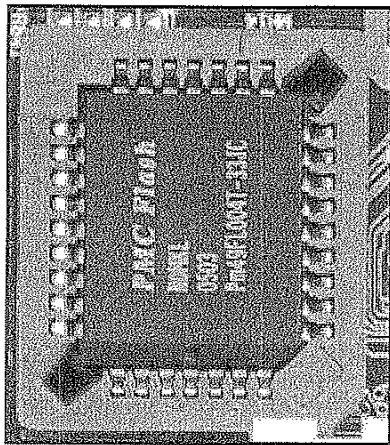


Fig. 4.4.
El chip que constituye un Flash BIOS del sistema.

El BIOS (ver la Fig. 4.4) indica el nivel de compatibilidad que tiene la PC con el hardware instalado y con los dispositivos con los cuales se va a comunicar. Un BIOS viejo puede significar tener que utilizar un controlador para poder acceder a un disco rígido de alta capacidad o un nuevo tipo de unidad de disco que no estaba disponible en los planes de los fabricantes del BIOS.

Muchas veces, cuando se llama al soporte técnico de un software determinado, especialmente de un sistema operativo, le preguntan el fabricante y fecha del BIOS, para saber si éste tiene inconvenientes con la aplicación. Siempre que haya algún problema de incompatibilidad con el acceso a un dispositivo de entrada/salida, es posible que el BIOS no esté trabajando bien con ese dispositivo, porque no lo conoce o no es compatible con el software que se está corriendo. Por ejemplo, cada versión nueva de Windows, o de cualquier otro sistema operativo, presenta inconvenientes al trabajar con ciertos BIOS (ver la Fig. 4.5).

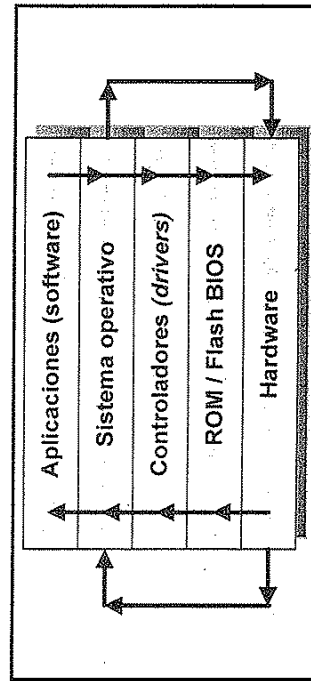


Fig. 4.5.
La ROM / el Flash BIOS como interfaz entre el hardware y el software.

El BIOS está conformado por un conjunto de pequeños programas que se encuentran grabados en la ROM (estos programas pregrabados en la memoria se conocen con el nombre de *firmware*). Los mismos se encargan de administrar a

bajo nivel el hardware de la PC. Cuando uno enciende la PC, ellos son los encargados de inicializar y verificar el hardware y los dispositivos, para luego realizar el proceso de carga y transferencia el control al sistema operativo (**booteo**).

El BIOS contiene también las rutinas que nos permiten acceder al CMOS Setup, mediante el cual podemos configurar muchos parámetros de nuestra PC y, dada su complejidad, lo analizaremos en detalle en el Capítulo 18: El CMOS Setup.

Si bien acceder a los dispositivos mediante las rutinas de la ROM BIOS asegura una total compatibilidad, los sistemas operativos modernos suelen dialogar directamente con el hardware saltando por encima del BIOS para conseguir un mayor rendimiento. La desventaja es que para esto, el sistema operativo debe tomar total conocimiento de los dispositivos con quienes quiere dialogar (ver la Fig. 4.5), y es por ello que surgieron los *drivers* (**controladores**), a los cuales analizaremos en detalle en el **Capítulo 17: Sistemas Operativos**.

Los *chips* del BIOS son fáciles de identificar, pues suelen estar montados sobre algún tipo de zócalo y son más grandes que otros chips de montaje superficial. La mayoría tienen una etiqueta o un holograma pegado encima con el nombre del fabricante y la versión.

Las ROM BIOS de las tarjetas de video y tarjetas adaptadoras SCSI contienen los procedimientos necesarios para que el sistema entienda la función de las mismas y se pueda comunicar con éstas sin inconvenientes.

Las PC Ecológicas y el Ahorro de Energía

Las **PC ecológicas** son aquellas PC de escritorio que permiten ahorrar energía cuando no se utilizan ciertos dispositivos, dejándolos en estado de espera, *stand-by* o *sleep*. Cuando toda la computadora se encuentra en el *modo sleep* puede llegar a reducir el consumo a 30 watt solamente.

Por ejemplo, una PC que no acceda al disco rígido por un tiempo determinado (por ejemplo, 10 minutos) la apaga, o mejor dicho, hace que los platos dejen de girar continuamente y cuando sea necesario leer información del mismo, vuelven a girar. En el caso de los monitores, cuando no se toca una tecla ni se mueve el mouse por cierto tiempo, se apagan automáticamente y se vuelven a encender por sí solos cuando se presiona una tecla nuevamente o se use el mouse, con el consiguiente ahorro de energía. El mismo concepto se aplica a las impresoras.

Las PC ecológicas ofrecen dicho soporte a través del BIOS, del microprocesador (éste debe tener incluida la capacidad de administración de consumo) y de la

configuración apropiada del CMOS Setup. Todos los sistemas operativos modernos proveen soporte a los modos de ahorro de energía.

Flash BIOS

En las antiguas PC, si se tenía algún problema con el BIOS, se debía proceder a su reemplazo, pero esto implicaba levantar la carcasa del gabinete, retirar el chip y colocar cuidadosamente uno nuevo. Sin embargo, a veces no se debe a un problema, sino que se necesita un BIOS más moderno para utilizar nuevos dispositivos, como ofrecer soporte a discos rígidos de más de 137 GB, arrancar a través de un dispositivo de almacenamiento conectado al bus de E/S USB 2.0, etc. La mayoría de las motherboards modernas utilizan un *Flash BIOS* (ver la Fig. 4.4). Se llama *Flash BIOS* a los chips del BIOS que se pueden reprogramar con un software especial sin necesidad de reemplazarlos. Esto significa que si necesita actualizarlo, solamente se deberá contactar con el fabricante para conseguir el software necesario para hacerlo, y así agregarle soporte a las últimas tecnologías y poder mantener su sistema al día.

Para obtener mayor información sobre los diferentes procedimientos y técnicas para actualizar un BIOS a una nueva versión, puede consultar el libro Reparando y Configurando Motherboards de PC, de Editorial HASA.



En la documentación de la motherboard debería figurar si posee un sistema de *Flash BIOS* o no, pero también se puede dar cuenta mirando el chip, que incluirá la palabra *Flash* en caso que sea de este tipo. La mayoría de las empresas que fabrican motherboards ofrecen actualizaciones para la mayoría de los BIOS en sus sitios de Internet para soportar nuevas tecnologías, si es posible, o también para solucionar incompatibilidades.

Direcciones de Memoria de Tarjetas de Expansión

Algunas tarjetas de expansión requieren cierta cantidad de memoria temporal para uso propio. Por ejemplo, una tarjeta red requiere un *buffer* (área de almacenamiento temporal) para guardar allí la información de las transmisiones y acelerar las comunicaciones.

Dos tarjetas diferentes no pueden utilizar la misma dirección de memoria como *buffer*, pues se modificarían los datos mutuamente y se perdería información. Por esa razón, la mayoría de las tarjetas ofrecen varias direcciones de memoria diferentes de las cuales se deberá elegir la que se encuentre libre. Si bien las primeras tarjetas solamente utilizaban sectores del área de memoria superior para sus propias memorias, en la actualidad mantienen esa compatibilidad, pero

aprovechan otros sectores de memoria que van más allá de esa área. Este tema se analizará más adelante cuando se trate el agregado de tarjetas.

La Memoria de Video

Como se puede deducir, es la memoria utilizada por las tarjetas de video para almacenar el contenido de la pantalla. Cuando se modifican los datos directamente en la memoria de video, éstos se verán reflejados en la pantalla, encargándose de ello la tarjeta de video. Cuando se diseñó la primera PC, se guardaron 128 KB para esta memoria, aunque en estos tiempos de multimedia, color real y aceleración 3D, esa cantidad de memoria es ínfima e insuficiente.

Las tarjetas de video modernas con aceleración 3D poseen mucha mayor cantidad de memoria que 128 KB. Aún en las tarjetas más elaboradas, los primeros 128 KB de memoria de video se siguen ubicando en el área de memoria superior, simplemente para mantener la compatibilidad con el diseño original de la PC y con el estándar VGA. Hace tiempo, se utilizaban técnicas similares a las de paginación de la memoria expandida para acceder a la memoria adicional al estándar antiguo, pero eran poco eficientes para las demandas continuas de un mejor desempeño en los gráficos.

Las tarjetas de video que poseen aceleración 3D pueden acceder a sus 256 MB; 512 MB; 1 GB o más de memoria en forma directa, sin necesidad de paginar, debido a que utilizan nuevos modos de direccionamiento. La paginación es necesaria cuando trabajan en modos compatibles con el antiguo estándar VGA. Estos temas los vamos a analizar con mayor detalle en el *Capítulo 8: Las Tarjetas de Video*.

La ROM BIOS de las Tarjetas de Video

El BIOS original de la IBM PC incluía soporte para MDA y CGA solamente. Sabiendo que en el futuro, se producirían expansiones en la mayoría de los campos manejados por el BIOS, se diseñó una forma de extenderlo, de manera tal que se podría adaptar a los nuevos productos siempre que se cumpla con un formato estandarizado de instrucciones básicas para mantener la compatibilidad. A partir de los adaptadores EGA y VGA se aprovechó esta capacidad del BIOS.

Las funciones concentradas en el BIOS de estas placas, comprenden operaciones básicas para facilitar las modificaciones a la pantalla sin necesidad de alterar registros propios de las tarjetas de video, que pueden resultar peligrosos para el monitor o la estabilidad del sistema. Algunas de las funciones son: detección de monitor color o monocromático, selección del modo de operación, manejo del cursor, selección de la página activa, manejo de ventanas de texto dentro de la

pantalla, lectura y escritura de caracteres y atributos, selección de colores, activación y lectura de pixels, cambio del juego de caracteres, etc.

Cuando arranca la PC se comunica con el BIOS de vídeo y le indica qué debe mostrar en pantalla, las instrucciones residentes en esta ROM se encargan de modificar la memoria de vídeo y los cambios se muestran en el monitor. Pero, hay una gran desventaja con este método, si bien todas las placas son compatibles con este sistema, la memoria ROM es lenta, muchísimo más que la memoria RAM, por lo tanto, para solucionar este problema, se aplica el método de *ROM Shadow* que se analizará a continuación. El DOS y Windows cuando arrancan en modo texto, utilizan la ROM BIOS de vídeo.

El Método de ROM Shadow

La memoria de sólo lectura es mucho más lenta que la memoria de escritura y lectura, por lo tanto, para solucionar el problema de la lentitud del BIOS, se instala desde el *CMOS-Setup* de las PC la opción de *Shadow RAM* o *ROM Shadow*. Si esta opción está activada, se copian los contenidos del BIOS en la memoria RAM (de ahí su nombre, pues quedan copiados como una *sombra*). De esta manera se aceleran las operaciones que lleva a cabo el BIOS, pues se encuentran en la memoria RAM que es mucho más rápida. Ver la Fig. 4.6.

Las opciones para configurar cómo se hará el ROM Shadow dependerán del BIOS que posee la PC. La mayoría permite activar o desactivar la opción para el System BIOS (BIOS del Sistema) y el Video BIOS (BIOS de la tarjeta de vídeo). Muchos BIOS permiten seleccionar cada uno de los bloques de memoria de la ROM que se quieren sombreadar en la memoria RAM. En las PC modernas no existen contraindicaciones para trasladar todos los bloques de ROM a RAM.

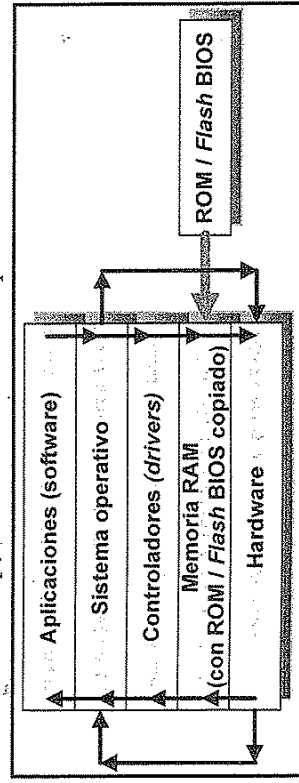


Fig. 4.6.
El método de ROM Shadow.

La Memoria Caché

Antes de comenzar con la descripción técnica de la utilidad de esta memoria ultrarrápida, pensemos en la siguiente situación. Como todo ser humano, necesito

alimentarme para subsistir, ahora, para comprar comida voy al supermercado. Si cada vez que tengo hambre debo ir hasta el supermercado, paso mucho tiempo viajando hacia éste, sin embargo, si mensualmente hago las compras generales y guardo todo en mi casa y solamente voy a buscar algo al supermercado cuando no lo tengo, aprovecharé mucho más el tiempo.

La situación anterior es la misma que vive el procesador cada vez que accede a la memoria y es el principio de funcionamiento de las **memorias caché**. Cuando los procesadores comenzaron a tener velocidades de reloj cada vez mayores y funcionaban más y más rápido, los tiempos de acceso de las memorias RAM dinámicas no alcanzaban para seguirle el ritmo.

De esa forma, el procesador debía esperar que la memoria reaccione cada vez que necesitaba un dato de ella, con lo cual se desperdiciaba una gran cantidad de tiempo que podría estar utilizando para procesar, que es su función principal y no esperar. Una alternativa sería utilizar memorias con menores tiempos de acceso: las RAM estáticas, pero esta solución es muy costosa porque las memorias estáticas cuestan más del doble que las dinámicas. Estos tipos de memoria se analizarán en profundidad más adelante.

Otra vez se recurrió a los principios utilizados en los *mainframes*, usar una memoria estática lo más rápida posible, de menor tamaño que la memoria total del sistema, como memoria caché. Se utiliza una pequeña cantidad de memoria mucho más rápida que la memoria de trabajo para almacenar datos de la memoria principal y que la CPU tome los datos de esta memoria más rápida y no deba acceder tan frecuentemente a la otra, que es mucho más lenta.

A medida que los microprocesadores fueron evolucionando, las memorias caché se fueron integrando en el procesador para ser aún más rápidas y también se agregaron diferentes niveles de memorias caché. Hace muchos años existía una caché interna muy veloz, incorporada en el procesador y una caché externa un poco más lenta, como un conjunto de chips o una tarjeta de memoria de alta velocidad adicional en la motherboard.

La gran mayoría de los microprocesadores modernos tienen estos dos niveles de memoria caché incorporados, con diferentes tamaños y velocidades, por lo cual, sus nombres pasaron a ser caché L1 (Level 1 - Nivel 1) para la más cercana los núcleos de procesamiento y caché L2 (Level 2 - Nivel 2) para la secundaria. Algunos microprocesadores también incorporan otros niveles de caché adicionales como caché L3 (Level 3 - Nivel 3), como determinados modelos de los Xeon e Itanium de Intel.



Debemos tener en cuenta que uno de los factores que más inciden en el costo de un microprocesador es el tamaño de las memorias caché que poseen. Es por ello que

los modelos más económicos de los principales fabricantes de microprocesadores consisten en ofrecer uno con menor tamaño de memorias caché y una frecuencia de reloj inferior.

Arquitecturas de Caché y Métodos de Escritura

La arquitectura de la memoria caché determina qué método se utiliza para organizar y localizar los datos en las memorias caché en los diferentes niveles; los tres tipos de diseño más comunes son los siguientes (ver la Fig. 4.7):

- **Mapeo directo.** Cada sector específico de la caché corresponde a varios de la memoria principal. Por lo tanto, cada vez que el procesador pide un dato, la controladora de caché debe ir a esa ubicación para encontrar la información, ahorrando tiempo de búsqueda. La gran desventaja de esta arquitectura de caché es que al corresponder varios bloques de memoria RAM al mismo bloque en la caché, si un programa accede a estos datos, la caché estará constantemente accediendo a la memoria principal para actualizar la información. Esta arquitectura no presenta una ganancia de rendimiento importante en la mayoría de los casos.

- **Asociativo completo.** Los bloques de la memoria principal se trasladan en cualquier bloque libre de la caché. La ventaja de esta arquitectura es que la información a la que se accedió recientemente se encuentra presente casi siempre en la caché y también los datos concurrentes. La desventaja es que la búsqueda de datos se debe hacer en forma indexada (utilizando índices) y lenta, lo cual significa que se deben recorrer los sectores de la caché en búsqueda del bloque de memoria deseado.

- **Asociativo en conjuntos de varias vías.** Combina las características de las dos arquitecturas anteriores. Está dividida en varios conjuntos: 2; 4; 8; 16; 32; etc., con múltiples sectores y cada uno de ellos conteniendo un bloque de datos. Cada bloque de memoria principal puede estar representado en cualquiera de los conjuntos, pero solamente se puede almacenar una sección de memoria por conjunto. En una caché asociativa en conjuntos de dos vías, que tiene dos conjuntos, existen solamente dos lugares para un bloque específico de memoria. De esta manera, se reduce el tiempo de búsqueda y también disminuye la posibilidad que los datos utilizados frecuentemente se sobrescriban entre sí. La desventaja de este tipo de cachés es que son más difíciles de implementar que las anteriores y más costosas.

Por otro lado, la información que se encuentra en la caché puede ser modificada por el procesador o por uno de los núcleos de ejecución contenidos en éste. La memoria caché actualiza la información modificada en la memoria principal utilizando dos métodos diferentes de escritura:

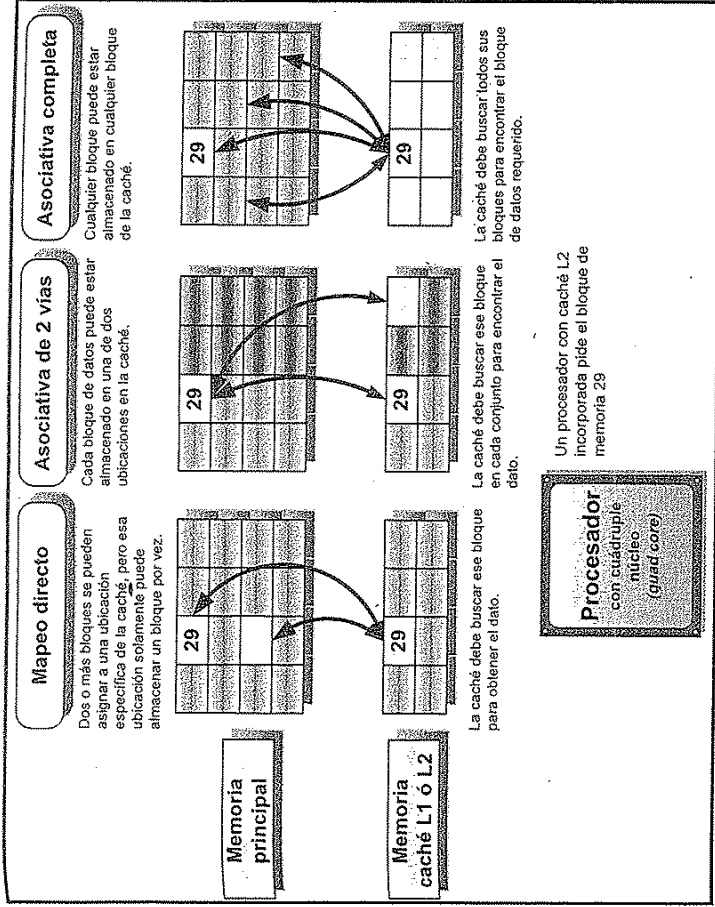


Fig. 4.7. Diagramas esquemáticos del funcionamiento de las diferentes arquitecturas de caché ante una solicitud de un bloque de memoria.

- **Write-through.** Empleado en los primeros 486, todo lo que el procesador escriba en la caché se actualizará automáticamente en la memoria principal. Es muy ineficiente.
- **Write-back.** Utilizado por la mayoría de los sistemas modernos, es el más conveniente y el más rápido. La información que el procesador escribe en la memoria caché queda almacenada en ésta y solamente se pasa a la memoria principal bajo alguna de las siguientes situaciones:
 - Cuando el bloque de memoria debe ser reemplazado por uno nuevo; cuando ese sector de la memoria se requiere para almacenar una nueva información.
 - Cuando otra parte del sistema, como un dispositivo que utilice el bus *mastering* y haya tomado el control del bus, trata de acceder a ese bloque de datos apuntando a la memoria principal.

Cuando el algoritmo del controlador de caché determina que los datos han permanecido en ésta por demasiado tiempo como para ser actualizados.

La mayoría de los microprocesadores modernos utilizan memorias caché asociativas en conjuntos de 4; 8 ó 16 vías, con el método de escritura write-back.



Las Caché con Múltiples Núcleos

Otro detalle muy importante a tener en cuenta es que los microprocesadores con múltiples núcleos de ejecución pueden tener cachés L2 unificadas o segmentadas; o bien, combinaciones de éstas. Por ejemplo, si tiene dos núcleos de ejecución y una caché de 4 MB unificada, ambos núcleos pueden utilizar esos 4 MB completos si así lo necesitaran, restando por supuesto la parte que está en uso por el otro. En cambio, si están segmentadas, cada núcleo podrá acceder solamente a 2 MB y no a los otros 2 MB que pertenecen al otro núcleo.

En algunas versiones de los procesadores de cuatro núcleos, consistentes en dos procesadores de dos núcleos sin una caché L2 unificada, como las primeras versiones del Core 2 Quad de Intel, cada doble núcleo tiene su propia caché L2 unificada, pero a su vez, podemos decir que la caché L2 está segmentada entre los dos grupos de dos núcleos. Pues, un único núcleo solamente podrá acceder a la mitad de la caché L2 total, la cual es la que comparte con el otro núcleo. La otra mitad quedará accesible para el segundo grupo de dos núcleos. En ese caso, tenemos dos cachés L2 unificadas para cada par de núcleos, pero una segmentada en dos para los cuatro núcleos. Estos esquemas son más económicos de fabricar, pero menos eficientes que las caché L2 totalmente unificadas.

Los mejores rendimientos de los microprocesadores con múltiples núcleos se consiguen con las caché L2 totalmente unificadas para que todos los núcleos puedan acceder a ésta. Especialmente, por las características del software actual que no está optimizado para su ejecución múltiples núcleos con cachés segmentadas.



El Protocolo MESI

Cuando comenzaron a aparecer los sistemas con múltiples procesadores y luego los microprocesadores con múltiples núcleos de ejecución, cada uno de ellos con sus propias cachés L2 segmentadas, compartidas o una combinación de estas dos, surgió la necesidad de establecer un control en la coherencia de los datos que contiene cada una de las memorias caché. Si bien este problema se podría haber solucionado utilizando una única caché externa para todos los procesadores o

núcleos de ejecución, se perdía la posibilidad de tener un ancho de banda mayor hacia la memoria principal. Para simplificar los ejemplos, vamos a utilizar la palabra procesador para identificar a una unidad de procesamiento, la cual puede ser tanto un microprocesador físico, así como un núcleo de ejecución dentro de un microprocesador.

El problema se puede ilustrar con el siguiente ejemplo: supongamos que el procesador A lee el valor de una dirección determinada de la memoria principal, el cual quedará en la caché conectada con este procesador. Al poco tiempo, el procesador B cambia el valor de esa posición de memoria en su memoria caché, que actualizará el cambio a la memoria principal. Como la caché del procesador A no se entera del cambio efectuado, ésta mantiene un dato inválido, que será utilizado si se corre alguna instrucción en este procesador que requiera nuevamente un dato de esa posición de memoria. Éste es uno de los problemas que se pueden tener si no se mantiene la coherencia de los datos de las cachés.

Para poder garantizar la coherencia de los datos en la memoria principal con los que se encuentran en todas las memorias caché de los múltiples núcleos o procesadores residentes en un sistema, Intel diseñó el protocolo MESI (*Modified Exclusive Shared Invalid* – Modificado exclusivo compartido inválido) para sincronizar los cambios en las distintas memorias.

Este protocolo está incorporado en todos los microprocesadores modernos, los cuales están preparados para compartir la motherboard con otros procesadores de su mismo tipo o bien poseen múltiples núcleos de ejecución. Éste garantiza que los datos ubicados en la memoria caché serán válidos o de otro modo presentarán una marca que indicará su estado de invalidez.

Al implementar el protocolo MESI, el caso presentado como ejemplo se soluciona de la siguiente manera: cuando el procesador B cambia el valor de esa posición de memoria en su caché externa y consecuentemente en la memoria principal, se marca la copia del valor que se encuentra en la caché del procesador A como inválida y si este último intentara acceder a esa posición de memoria, el protocolo garantiza que vaya a leerla a la memoria principal, obteniendo el valor correcto.

Tamaños de Cachés y Rendimiento

La caché L1 generalmente está hiper-integrada en el mismo trozo de silicio del microprocesador. Su ventaja es que el procesador la tiene más cerca y puede acceder a la misma a través de un bus interno con un ancho de datos mucho mayor que el usado para la caché L2. La caché L1 es mucho más pequeña que la L2, pero es mucho más rápida que esta última (ver la Tabla 4.1, teniendo en cuenta que varios modelos de microprocesadores siguen evolucionando). Al utilizar una caché integrada, los procesadores reducen drásticamente el tiempo necesario para acceder

a los datos en memoria. Si no están en la caché L1, podrán estar en la L2 y, en el peor de los casos, habrá que recurrir a la memoria principal.

Tabla 4.1. Algunos tamaños de caché L1 y L2 de los microprocesadores modernos (se suman las memorias caché de todos los núcleos de ejecución).

Procesador	Caché L1 total (mínima; máxima)	Caché L2 total (mínima; máxima)
Athlon 64	128 KB	512 KB; 1 MB
Athlon 64 FX	128 KB; 256 KB	1 MB; 2 MB
Athlon 64 X2	256 KB; 256 KB	1 MB; 2 MB
Athlon XP	64 KB; 128 KB	256 KB; 512 KB
Celeron D	16 KB; 32 KB	256 KB; 512 KB
Core 2 Duo	32 KB	2 MB; 4 MB
Core 2 Extreme Quad-Core	32 KB	4 MB; 8 MB
Core Duo	32 KB	2 MB
Pentium 4	8 KB para datos y caché de seguimiento de ejecución con espacio para 12.000 micro-instrucciones	256 KB; 512 KB
Pentium 4 Extreme Edition	16 KB; 32 KB	512 KB; 2 MB
Pentium 4 Hyper-Threading	16 KB	512 KB; 2 MB
Pentium D	16 KB	2 MB; 4 MB
Pentium Extreme Edition	16 KB	2 MB; 4 MB
Pentium II	64 KB	512 KB (incorporada al cartucho)
Pentium III	64 KB	256 KB; 512 KB
Sempron	128 KB	128 KB; 256 KB

La caché interna de seguimiento de ejecución presentada por el Pentium 4 fue una versión mejorada de la caché de código, la cual permite ir almacenando micro-operaciones ya decodificadas y de esta manera elimina la lógica de decodificación de instrucciones de los principales caminos de ejecución del procesador y así lograba, en teoría, mejorar aún más su rendimiento. Sin embargo, los siguientes microprocesadores de Intel abandonaron este esquema y volvieron a la caché L1 tradicional, pues el software existente se beneficia más de esta última arquitectura.



Los procesadores que también incorporan una caché L2 dentro del mismo chip, también tienen una caché L1, la cual es mucho más rápida que la L2.

El tamaño de la memoria caché L2 de todos los núcleos de ejecución suele variar entre 128 KB y 32 MB. En la Tabla 4.1 figuran los valores típicos para los procesadores modernos. En la Fig. 4.8 se muestra el rendimiento promedio de un mismo procesador, con la misma micro-arquitectura e idéntico tamaño de caché L1, pero con diferentes tamaños de caché L2 para aplicaciones modernas con un uso muy intensivo de grandes cantidades de memoria (en el pasado, este gráfico era diferente por el gran cambio que ha sufrido el software).

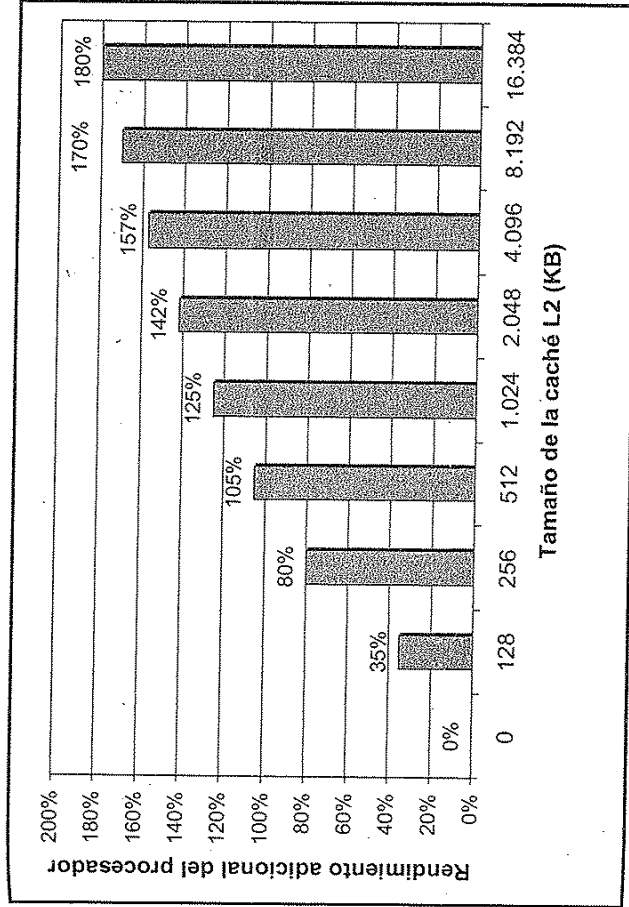


Fig. 4.8. Rendimiento del procesador con diferentes tamaños de caché L2.

El funcionamiento típico de un controlador de caché es el siguiente: cuando el microprocesador necesita leer una dirección de memoria, ésta se busca en la memoria caché muy rápidamente, si no se encuentra allí, se toma de la memoria principal y mientras el procesador continúa trabajando, los circuitos controladores de caché hacen que se almacenen en ésta las siguientes posiciones de memoria consecutivas a la que pidió el procesador.

La mayoría de los programas se van ejecutando una instrucción después de la otra y piden datos de la memoria uno después del otro, por lo tanto, cuando el microprocesador necesite leer las direcciones de memoria siguientes a la que leyó anteriormente, el contenido se encuentra en la memoria caché que es muchísimo más rápida que la memoria principal. De esta forma se logra un aumento de velocidad de procesamiento muy grande.

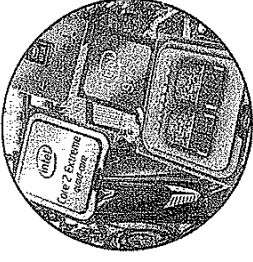


Es por ello que la caché unificada para los múltiples núcleos de ejecución resulta mucho más eficiente para la mayoría de las aplicaciones actuales. Si bien es mucho más costosa, permite que se aproveche más la capacidad de procesamiento ofrecida por los múltiples núcleos.

Cuando el procesador tiene que escribir información en la memoria principal, ésta pasa por la caché en vez de escribirse directamente en la memoria principal (mucho más lenta). El controlador de caché decidirá cuando los datos modificados serán escritos en la memoria principal. Los métodos de escritura para hacerlo y las diferentes arquitecturas las hemos analizado anteriormente.

A medida que se vayan incorporando más núcleos de ejecución en un mismo microprocesador, la arquitectura de las memorias caché se volverá aún más importante para el desempeño del software que ejecuten.

Capítulo 5



Módulos de Memoria

Cuando hablábamos de memoria, nos referíamos a la capacidad de las computadoras de almacenar información en forma temporal. Ahora bien, como todo lo que se almacena necesita un lugar en el espacio, la memoria debería ocupar un lugar físico, y lo hace.

La memoria de una computadora está compuesta por los módulos de memoria, los cuales a su vez contienen un conjunto de chips de memoria que son capaces de almacenar una cierta cantidad de datos con un determinado tiempo para acceder a la información que tienen guardada y otro tiempo para escribir o leer dicha información. También estos chips se pueden presentar con distintas capacidades.

Tipos de Chips y Módulos de Memoria

Los chips de memoria se presentan en distintos tipos y se combinan de manera diferente según el tamaño y alcance de la memoria de trabajo. Es importante aclarar que para realizar arapliaciones de memoria, si bien puede ser necesario el reemplazo de módulos, no hará falta el uso de un soldador, pues los grupos de chips se insertan en zócalos o ranuras especialmente diseñados para alojarlos.

Los chips que se utilizan para la memoria de trabajo de las PC pueden ser de diferentes tipos, y a su vez se clasifican según su capacidad, parámetros de velocidad, tecnología y costo. A continuación, presentaremos el significado de las letras que forman los nombres de estas clases de chips de memoria:

- DRAM (*Dynamic Random Access Memory* – Memoria Dinámica de Acceso Aleatorio).
- DRAM FPM (*Fast Page Mode* – Modo de Paginación Veloz).
- DRAM EDO (*Enhanced Data Output* – Transferencia de Datos Mejorada).
- SDRAM (*Synchronous DRAM* – DRAM Sincrónica).
- Direct RDRAM (*Rambus DRAM* – DRAM Rambus).

- SRAM (*Static Random Access Memory* - RAM estática).
- BSRAM (*Burst SRAM* - SRAM fugaz).
- VCM SDRAM (*Virtual Channel Memory SDRAM* - Canal Virtual de SDRAM). También conocida como VCRAM.
- DDR SDRAM (*Double Date Rate SDRAM* - SDRAM de Doble Velocidad de Transferencia de Datos).
- DDR2 SDRAM o DDR-II SDRAM (*Quadruple Date Rate SDRAM* - SDRAM de Cuádruple Velocidad de Transferencia de Datos o (*Double Date Rate 2 SDRAM* - SDRAM de Doble Velocidad de Transferencia de Datos 2).
- DDR3 SDRAM (*Double Date Rate 3 SDRAM* - SDRAM de Doble Velocidad de Transferencia de Datos 3).

A su vez, estos chips pueden presentarse individualmente o agrupados en forma de:

- SIP (*Single In-Line Packages* - Paquetes Simples de Memoria en Línea).
- SIMM (*Single In-Line Memory Module* - Módulos Simples de Memoria en Línea) en sus versiones de 30 y 72 pines representando un ancho de datos de 8 y 32 bits respectivamente.
- DIMM (*Dual In-Line Memory Module* - Módulos de Memoria Dual en Línea) en sus versiones de 168 pines, 184 pines (DDR) y 240 pines (DDR2 y DDR3); representando todas un ancho de datos de 64 bits.
- RIMM (*Rambus In-Line Memory Module* - Módulos de Memoria Rambus en Línea) en su versión de 184 pines que puede enchufarse en conectores de 168 pines, representando un ancho de datos de 16 bits. También existen RIMM con anchos de datos de 32 bits y mayores.

Los DIMM son las agrupaciones más utilizadas en la actualidad. Todas también presentan variaciones en cuanto a capacidad, tecnología, conectores, velocidad, etc., (ver la Tabla 5.1). A continuación, analizaremos profundamente los conceptos de todos los tipos de chips más utilizados en los equipos modernos y sus posibles agrupaciones, junto con todas las clasificaciones de acuerdo a los diferentes parámetros.

El Ciclo de Refresco

Los chips de RAM dinámica, que conforman la base de todas las tecnologías de memoria que vamos a estudiar, están disponibles en capacidades que van desde los 64 Kb (kilobits) hasta varios Gb (gigabits). También se presentan en capacidades expresadas en Knybbles, donde 1 nybble es igual a un grupo de 4 bits.

Tabla 5.1. Comparación de las diferentes clases de módulos de memoria.

Módulo de memoria	Bus de datos (sin tener en cuenta el control de paridad)	Bus de datos + control de paridad
SIP	8 bits	9 bits
SIMM de 30 pines	8 bits	9 bits
SIMM de 72 pines	32 bits	36 bits
DIMM	64 bits	72 bits
RIMM	16 bits	18 bits

Este tipo de chips de memoria recibe el nombre de dinámicos porque el contenido de la memoria se debe refrescar continuamente, lo cual significa que para mantener los datos almacenados, estos chips se deberán someter a un ciclo de reescritura continuo.



El origen de este problema son los componentes utilizados para armar los chips: los capacitores. Los capacitores tienen la propiedad de cargarse con una tensión y permanecer así durante un cierto tiempo, pero luego se descargan, es por eso que requieren que se lea su valor y se lo reescriba periódicamente.

En un chip de un megabit, se necesitarán 1.048.576 capacitores, que se concentrarán en el chip de la memoria. En uno de un gigabit, la cantidad de capacitores asciende a 1.073.741.824.

Como se explicó anteriormente, un capacitor se descarga después de un tiempo, siendo el mismo bastante corto. Por lo tanto, para mantener la información almacenada en el chip durante más tiempo del que permanece cargado un capacitor, será necesario leer el estado de los capacitores del chip antes que se descarguen y una vez leídos los datos, recargarlos. A este proceso se lo denomina **ciclo de refresco**.

Mientras el chip está cumpliendo con el mencionado ciclo de refresco, no se puede acceder a la información almacenada en el mismo, por lo tanto cuanto menores sean los intervalos de tiempo de los ciclos de refresco, también lo va a ser el tiempo que necesitará el chip para responder a un pedido, y entonces será más rápido. Es por ello que todas las tecnologías de memoria aparecidas tuvieron como objetivo la reducción de este tiempo.

Tiempos de Acceso

El tiempo de acceso de los chips se expresa en nanosegundos (donde 1 nanosegundo = 1×10^{-9} segundos = 0,000000001 segundos) y representa el tiempo que tarda la memoria en responder a un pedido de lectura o escritura. Aunque se

piense que solamente es este último quién determina el tiempo requerido por la memoria para leer o escribir datos, existe otro tiempo más. El **tiempo de carga** de la memoria es el tiempo que tarda una de las celdas del chip en cambiar de un estado al otro, este tiempo puede ser igual al de acceso o mayor.

Entonces, el tiempo de ciclo total de la memoria va a ser igual a la suma de los tiempos de acceso y de carga:

$$\text{tiempo de ciclo} = \text{tiempo de acceso} + \text{tiempo de carga}$$

Generalmente, los datos suministrados en el chip por los fabricantes son solamente los del tiempo de acceso.



Estados de Espera

La memoria debe ser capaz de responder a los pedidos del microprocesador cada dos tics del reloj. Un tic del reloj o cronómetro es igual a 1 sobre la frecuencia del reloj.

$$\text{tic del reloj (en segundos)} = 1 / (\text{frecuencia del reloj (en Hertz)})$$

Por lo tanto, para un sistema con un procesador que funcione a 600 MHz, un tic del reloj sería igual a $1/500.000.000 = 0,000000002$ segundos $= 2 \times 10^{-9}$ segundos $= 2$ nanosegundos. Por lo tanto, se necesitará una memoria capaz de responder en 2×2 ns $= 4$ ns. Cuando se analizaron los tiempos de acceso en las memorias DRAM, se indicó que el tiempo total de respuesta de la memoria es el tiempo de ciclo. En el ejemplo anterior, se habló de una memoria que responda en 4 ns, pero eso no quiere decir que se deba instalar una memoria de 4 ns de tiempo de acceso, pues debemos tener en cuenta el tiempo de ciclo.

Por lo tanto, podemos llegar a tener que utilizar una memoria con un tiempo de acceso de 2 ns para un requerimiento de tiempo de respuesta de 4 ns, debido al incremento en el tiempo de respuesta que representa el tiempo de carga.

A medida que fueron aumentando las velocidades del reloj de los procesadores, los tics del reloj se fueron haciendo cada vez más cortos, requiriendo un tiempo de respuesta menor por parte de la memoria. Pero, una vez más, el costo impide que se puedan utilizar memorias demasiado rápidas y, como reducir la velocidad del procesador sería una gran contradicción al avance tecnológico, se recurre al agregado de estados de espera.

Un estado de espera es básicamente un tic adicional de reloj que se agrega a cada acceso a memoria. De esta forma, en vez de utilizar una memoria con un tiempo de ciclo de 2 tics de reloj, se puede usar una con un tiempo de ciclo equivalente a 3 tics de reloj.

Agregar estados de espera significa una reducción importante en el rendimiento del sistema cuando se accede a la memoria, aunque es la única solución económica al incesante requerimiento de memorias más y más rápidas.



Organización Básica de los Chips DRAM

Los chips nombrados anteriormente tienen capacidades que varían entre 64 y miles de millones de kilobits. Como podemos ver, la memoria de trabajo de las PC está dispuesta en segmentos de kilobytes en vez de kilobits. Y, como un byte equivale a ocho bits, ocho chips de 1 gigabit cada uno forman 1 GB de memoria.

En ciertos casos, en vez de utilizar ocho chips, se utilizan nueve. Este chip adicional se utiliza para el **control de paridad**, allí se almacena constantemente el resultado de la suma de todos los datos de los ocho bits, de manera que si la suma da un número par, se representa un 0 en el bit de paridad y si es impar, un 1. Si hay un error en algunos de los chips de memoria, se podrá detectar mediante el control de paridad.

Como este chip de control de paridad no es necesario para el almacenamiento de datos (sí para la seguridad), se puede obviar, abaratando los costos. Pero, para esto se debe anular el control de paridad, reduciendo la seguridad que los datos en memoria sean los correctos. De esta forma aparecen las memorias más económicas que no incluyen el control de paridad.



También existen chips de nybbles, en vez de bits, siendo un nybble equivalente a 4 bits. Por lo tanto, estos chips almacenan el cuádruple de capacidad que un chip de bits: uno de 64 meganybbles tiene una capacidad de 256 megabits.

De esta manera, en vez de utilizar nueve chips de 512 megabits, se utilizan 2 chips de 512 meganybbles y 1 chip de 512 megabits para el control de paridad: 4 bits + 4 bits + 1 bit = 9 bits (nybble + nybble + bit paridad).

Este método constituyó una importante reducción en el espacio ocupado por la memoria, pues con tres chips se puede realizar lo que se conseguía con nueve.

Bancos de Memoria según el Procesador

Cuando estudiamos detalladamente los microprocesadores que aparecieron en el campo de las PC, desde el abuelo 8088 hasta el Core 2 Quad de Intel y el Athlon 64 X2 de AMD, hablábamos de diferencias entre los procesadores más allá de la velocidad y capacidad de procesamiento. Estas diferencias se referían a la forma

como se comunicaban con la memoria, es decir, el bus de datos y el bus de direcciones.

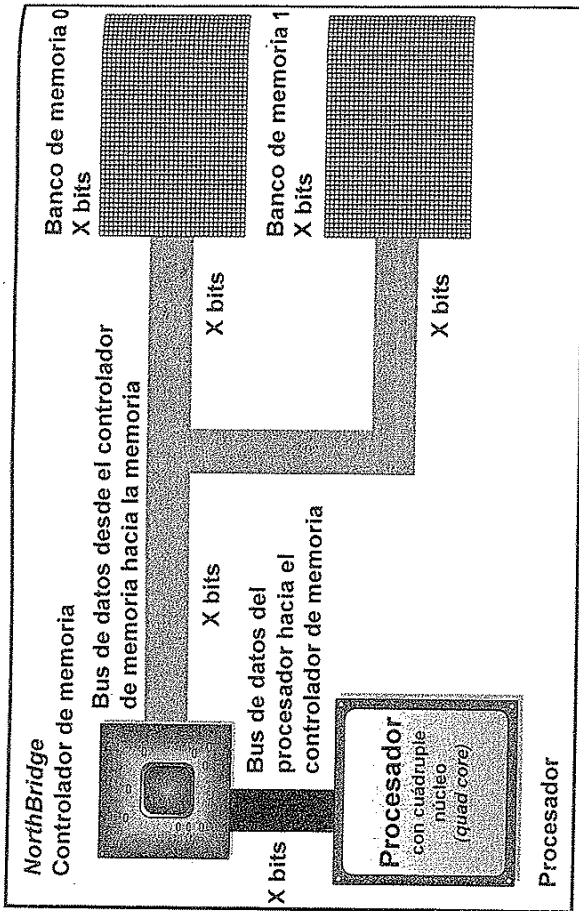


Fig. 5.1. Bancos de memoria en configuración sin paridad, con un único canal, donde X bits se debe reemplazar por el ancho del bus de datos del procesador.

El bus de direcciones define la mayor cantidad de memoria con la cual puede conversar (direccionar) el procesador, y el bus de datos define la cantidad de datos que iba a llevar y traer a memoria de una sola pasada, es decir, a mayor ancho de datos, mayor capacidad para comunicarse con más cantidad de memoria en menor tiempo.

Se conoce con el nombre de **banco de memoria** al conjunto de uno o más chips o tarjetas de memoria que agrupados conforman la misma cantidad de bits de datos que el bus de datos del procesador utilizado para recibir y transmitir información a la memoria (ver las Figs. 5.1 y 5.2).

Es más sencillo comprenderlo con ejemplos. Si tenemos un procesador como el Core 2 Extreme Q; Core 2 Quad; Pentium D; Athlon 64 X2 u otro que utilice un bus de datos de 64 bits con la memoria, cada banco de memoria puede estar compuesto por las configuraciones mostradas en la Tabla 5.2 (ver la Fig. 5.3). Las configuraciones con tarjetas SIMM ya no se utilizan, se mencionan solamente a modo de comparación.

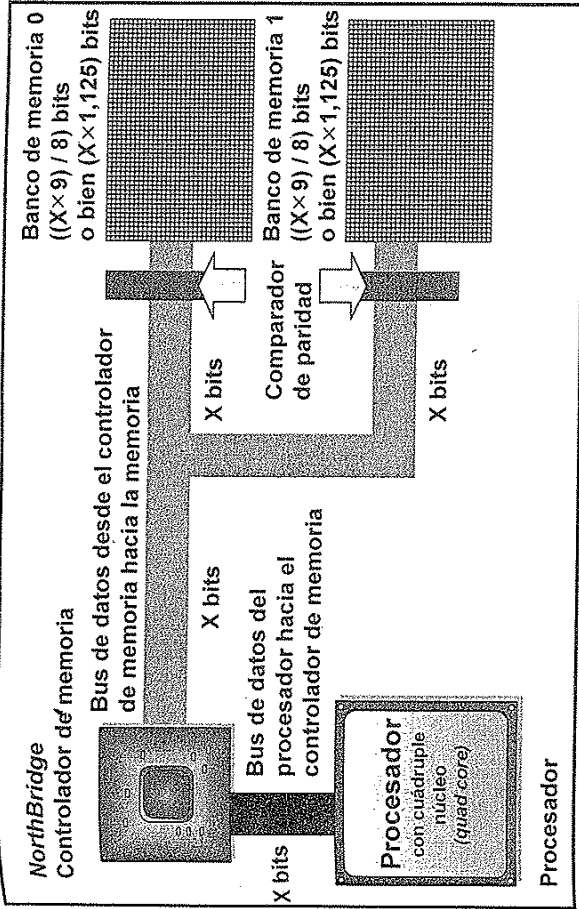


Fig. 5.2. Bancos de memoria en configuración con paridad, con un único canal, donde X bits se debe reemplazar por el ancho del bus de datos del procesador.

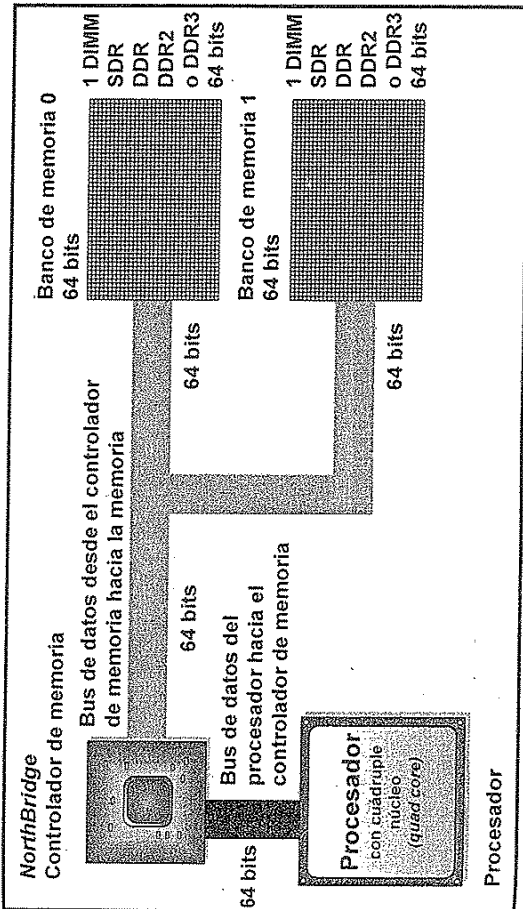


Fig. 5.3. Bancos de memoria de 64 bits con un único canal, conformados con módulos DIMM, sin control de paridad.

Tabla 5.3. Posibles configuraciones para un banco de memoria de 64 bits.

Módulo de memoria	Bits por módulo (sin paridad)	Bits por módulo (con paridad)	Cantidad de módulos para formar un banco
SIMM de 30 pines	8	9	8
SIMM de 72 pines	32	36	8
DIMM SDR, DDR, DDR2 o DDR3	64	72	1

Como conclusión, dependiendo del ancho del bus de datos con la memoria utilizada por el procesador, del tipo de módulo de memoria empleado por la motherboard y de la cantidad de canales de comunicación en paralelo que puede manejar el controlador de memoria, vamos a saber qué cantidad y cuál clase de módulos necesitamos para completar un banco de memoria.

Consideraciones para los Bancos de Memoria

Debemos tener en cuenta algunas reglas a la hora de configurar la memoria en las máquinas que utilizan cualquiera de los módulos de memoria mencionados. A continuación, describimos las reglas a respetar:

1. Cuando los bancos de memoria requieren más de un módulo para completarse, para que el sistema funcione correctamente, se deberá llenar el banco completo. En esos casos, no es posible llenar un banco por la mitad.
2. Se deben ir completando los bancos en forma ascendente. Siempre se debe comenzar a llenar primero el banco 0, luego el 1, después el 2 y así sucesivamente. No puede estar vacío el banco 0 y lleno el 1, pero sí puede estar completo el 0 y vacío el 1 y los subsiguientes.
3. Cada banco debe tener módulos de memoria de la misma capacidad y velocidad. No se pueden utilizar módulos de diferente tamaño o velocidad en un mismo banco, pero sí es válida la utilización de tarjetas de diferente tamaño en bancos separados. Por ejemplo: el banco 0 se llena con módulos de 512 MB y el banco 1 con tarjetas de 256 MB.
4. Es conveniente que los módulos que formen parte de un banco sean de la misma marca y partida. Aunque sean de la misma capacidad y velocidad, algunos fabricantes miden las velocidades con métodos diferentes y si se mezclan módulos de distintas marcas se pueden presentar problemas.
5. Si bien es posible incluir en un banco un módulo de cualquier capacidad, es muy frecuente que algunas motherboards tengan límites en las capacidades máximas para determinados bancos, por lo cual, es conveniente consultar la documentación antes de realizar una compra.

6. Algunas motherboards no permiten que el segundo banco sea menor que el primero. Por ejemplo: si se colocaron en el primer banco módulos de 1 GB, no se pueden colocar en el segundo banco de 512 MB.

7. Para estar seguro de las limitaciones de las motherboards con respecto a estas reglas, es conveniente consultar con el manual de la misma para estar seguro antes de reemplazar o agregar tarjetas de memoria, pues se pueden presentar ciertas limitaciones imprevistas, como las explicadas en las reglas 5 y 6.

Los Módulos Obsoletos

Los módulos SIMM de 30 y 72 pines fueron muy populares como tarjetas de memoria para los procesadores 386DX; 486 y Pentium. Y también los utilizaron tarjetas de video y de sonido. Sin embargo, en la actualidad han quedado en desuso, por lo cual, no vamos a describirlos en detalle.

Otros módulos que prometieron mucho, pero quedaron completamente en desuso son los RIMM (*Rambus In-Line Memory Module* - Módulos de Memoria Rambus en Línea). Si bien aparecieron como la tecnología de memoria que reemplazaría a los DIMM SDRAM, solamente se utilizaron en algunos equipos Pentium III y Pentium 4, pero, debido a sus altísimos costos, no se siguieron utilizando en PC con procesadores más modernos.

Módulos DIMM

Los DIMM (*Dual In-Line Memory Module* - Módulos de memoria dual en línea) fueron diseñados especialmente para utilizarse con los microprocesadores Pentium o superiores que duplicaban el bus de datos a 64 bits, en vez de los 32 bits que usaban sus predecesores.

Físicamente son similares a los antiguos SIMM, con la diferencia que presentan tres ranuras en la parte de los contactos y son un poco más largos que estos últimos, pues tienen 168 pines. Los DIMM (ver las Figs. 5.4 y 5.5) pueden ser de 64 ó 72 bits. Si son de 64 bits, quiere decir que no incorporan el control de paridad ($8 \times 8 = 64$). En cambio, si son de 72 poseen control de paridad ($9 \times 8 = 72$). Ver la Tabla 5.1.

Cuando aparecieron, los DIMM estaban compuestos por conjuntos de chips EDO DRAM, mientras que más tarde agruparon a las tecnologías SDRAM; VCM SDRAM (VGRAM); DDR; DDR2 y DDR3. Todas estas tecnologías de memorias las analizaremos más adelante en este mismo capítulo.

Los DIMM vienen en diferentes versiones, dependiendo de la tecnología de memoria que utilicen, como se muestra en la Tabla 5.4, en la cual, podemos ver

que a medida que mejoraron las velocidades, se fue reduciendo la tensión de alimentación con la cual trabajan los módulos.

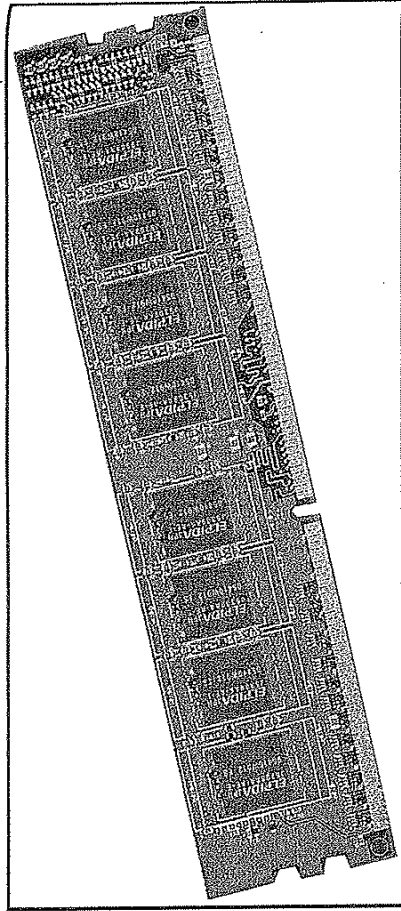


Fig. 5.4. Un DIMM DDR3 de 240 pines (observar el único corte).

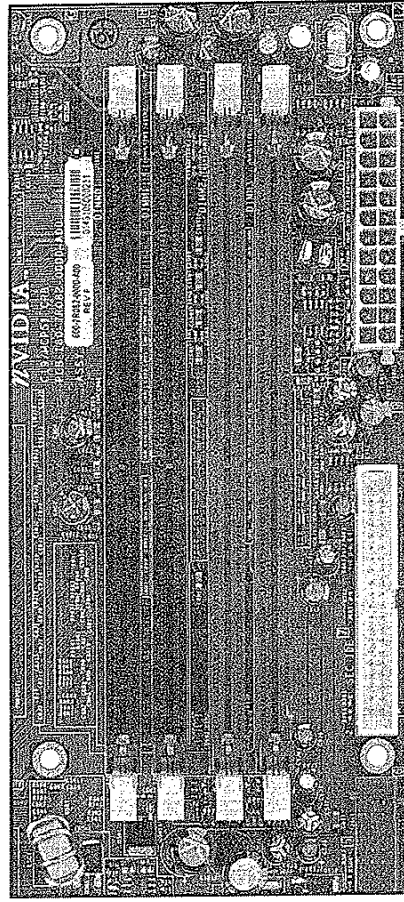


Fig. 5.5. Bancos de memoria para DIMM DDR3 de 240 pines.

Para cada una de las tecnologías de memoria que agrupan, existen distintos tipos de DIMM, diferenciados físicamente por la ubicación de una muesca que no nos permite instalar uno en una ranura preparada para otro (dado que de poder hacerlo se dañarían en forma permanente). A continuación detallamos los más utilizados en la actualidad:

- SDR y DDR SDRAM:
 - o Registered (registrado).

- o Unbuffered (sin buffer).
- o Buffered (con buffer).
- DDR2 y DDR3 SDRAM:
 - o Registered (registrado).
 - o Unbuffered (sin buffer).
 - o Fully Buffered (con buffer total)

Tabla 5.4. Diferentes versiones de los DIMM.

Tecnología de memoria	Cantidad de pines	Tensiones de trabajo
EDO DRAM	168	5 V
SDR SDRAM	168	3,3 V
VCM SDRAM (VDRAM)	168	3,3 V
DDR	184	2,5 V
DDR2	240	1,8 V
DDR3	240	1,5 V

Los DIMM también pueden ofrecer características de ECC y utilizan mayor cantidad de bits para efectuar el control y la corrección de errores que los SIMM convencionales, sin embargo, el *chipset* y el controlador de memoria deben estar preparados para ECC.

Múltiples Canales de Memoria

A partir de los procesadores Pentium 4 y, más aún, con los procesadores con múltiples núcleos de ejecución, comenzaron a aparecer motherboards que ofrecen dos o más canales de comunicación desde el controlador de memoria (*NorthBridge*) hacia la memoria. Por lo cual, cada banco de datos para estos procesadores se conforma por tanta cantidad de módulos como canales de memoria existan. Por ejemplo, el caso más común es que hayan dos canales de memoria entre el controlador de memoria y los módulos, por lo cual, se requerirán dos módulos para completar un banco.

De esta manera, se multiplica el ancho de banda entre el controlador de memoria y los bancos de memoria, consiguiendo un fuerte incremento en el rendimiento del subsistema de memoria. Pues, si bien el canal entre el procesador y el controlador de memoria se mantiene igual, el controlador de memoria logra mejorar los accesos a varios bloques a través de sus múltiples canales con la memoria y luego los

empaqueta y los envía a través del único canal con un ancho de banda reducido hacia el microprocesador (ver la Fig. 5.6).

Esta técnica mejora el rendimiento de las aplicaciones que hacen un uso intensivo de grandes cantidades de memoria, como las de retoque fotográfico y los sistemas de bases de datos. Su desventaja es que para aprovecharla se deben instalar módulos de memoria acorde a la cantidad de canales de comunicación.

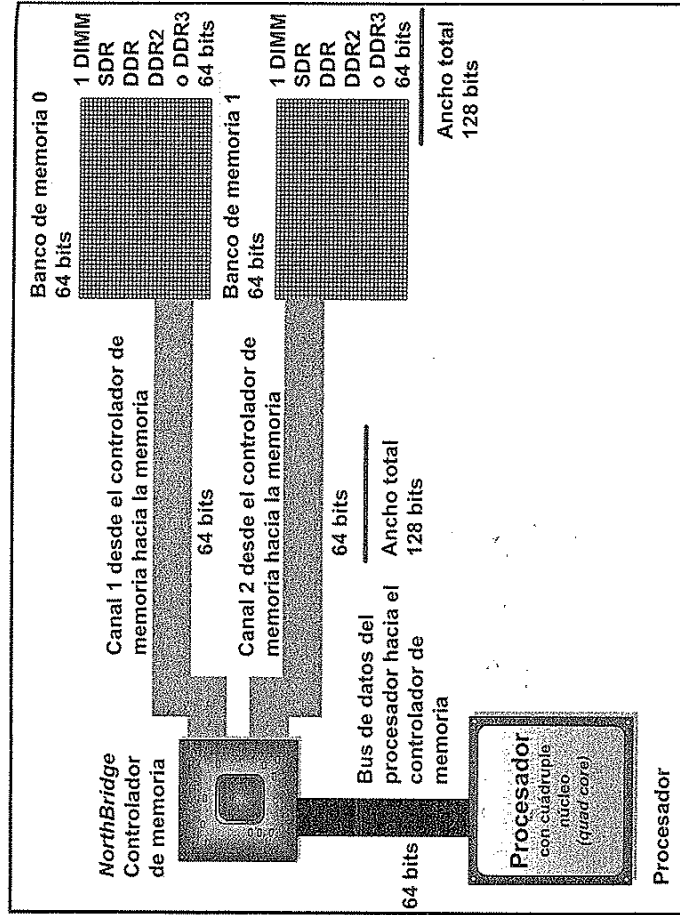


Fig. 5.6. Bancos de memoria de 64 bits con un doble canal, conformados con módulos DIMM, sin control de paridad.

Tecnologías de Memoria

Ahora que ya conocemos las diferentes agrupaciones de chips y el funcionamiento de los bancos de memoria, vamos a entrar en detalle en cada una de las numerosas tecnologías que podemos encontrar en las mencionadas agrupaciones.

Fast Page y EDO

A mediados de 1995, apareció una nueva tecnología de memoria conocida como EDO (*Enhanced Data Output* – Transferencia de datos mejorada) o memoria de alto rendimiento, debido a que sus tiempos de acceso eran similares a los de los SIMM de 72 pines convencionales, pero se diferencian en que estaban optimizadas para entregar una mayor transferencia de datos cuando se necesitan varios bloques de memoria.

Estos tipos de SIMM mejoraron el rendimiento del subsistema de memoria cuando se acceden a varios bloques de memoria, pero no cuando se lo hace en forma aleatoria: los accesos esporádicos no son más rápidos, pero sí lo son aquéllos realizados en conjunto. A este tipo de memoria también se la denominó HPM (*Hyper Page Mode* – Modo de hiper paginación).



Los SIMM de 72 pines convencionales y los de 30 pines utilizaban chips DRAM clásicos. A esta tecnología de memoria se la conoce también con el nombre de Fast Page (Paginación Veloz).

Las memorias EDO tienen solamente versiones en SIMM de 72 pines y DIMM de 168 pines.

SDRAM: Sincrónica Es Mejor

A mediados de 1996, apareció una nueva tecnología de memoria que lograba ofrecer tiempos de acceso y transferencias de datos mucho mejores que las de EDO y más aún que *Fast Page*: SDRAM (*Synchronous DRAM* – DRAM sincrónica). En algunos casos, llega a ser hasta un 400% más rápido que la DRAM convencional.

Esta nueva memoria logra trabajar en forma sincronizada con la velocidad del bus que la comunica con el microprocesador (de ahí su nombre), evitando muchos estados de espera que son característicos en las memorias EDO y *Fast Page*, pues estas últimas trabajan en forma asincrónica. Además, SDRAM incorpora modos de transferencia de datos continuos en ráfaga (*burst mode*) que aceleran los accesos a datos en posiciones contiguas y a su vez permite comenzar a acceder otro dato en paralelo (*pipeline*) sin haber completado el primero de ellos.

El tiempo de acceso promedio de la SDRAM es de 15 ns para 66 MHz, de 10 ns para 100 MHz, de 8 ns para 125 MHz y 6 ns para 133 MHz, prácticamente tan bajos como el de la mayoría de las memorias caché externas de esa época. Se presenta en varias velocidades: 66; 100; 125 y 133 MHz. Las versiones de 100 MHz y superiores se adaptan a la velocidad del bus, pudiendo descender su velocidad de trabajo a 83; 75 y 66 MHz, según sea necesario. Sin embargo, al ser DRAM sigue requiriendo el ciclo de refresco.

Los DIMM FPM fueron poco comunes, por lo cual cuando se trata de DIMM EDO muchas veces no se hace la aclaración, pero cuando son SDRAM simplemente se le agregan esas cinco letras en algún lado.

DDR; DDR2 y DDR3

En el Capítulo 3 conocimos por primera vez la aplicación de la tecnología DDR (*Double Data Rate* – Doble tasa de transferencia de datos) que utilizó PCI-X 1.0. Esta misma tecnología se aplicó sobre SDRAM, para que utilice los flancos de subida y de bajada del reloj, en vez de usar únicamente el de subida y de esta forma consigue duplicar la transferencia de datos utilizando la misma velocidad de reloj real (ver la Fig. 3.9 en el Capítulo 3). Ver la Tabla 5.5.

A estas memorias se las conoce con el nombre de DDR SDRAM y son las más utilizadas en los procesadores modernos, desafiando a la tecnología RDRAM.

Forma de Cálculo:

- SDR:** Velocidad máxima (en bytes) =
 $(\text{Velocidad del reloj (en Hz)} \times \text{ancho del bus de datos (en bits)}) / 8$
- DDR:** Velocidad máxima (en bytes) =
 $(\text{Velocidad del reloj (en Hz)} \times 2 \times \text{ancho del bus de datos (en bits)}) / 8$
- DDR2:** Velocidad máxima (en bytes) =
 $(\text{Velocidad del reloj (en Hz)} \times 4 \times \text{ancho del bus de datos (en bits)}) / 8$
- DDR3:** Velocidad máxima (en bytes) =
 $(\text{Velocidad del reloj (en Hz)} \times 8 \times \text{ancho del bus de datos (en bits)}) / 8$
- (Velocidad del reloj (en Hz) × ancho del bus de datos (en bits)) / 8 =
 (Velocidad del reloj (en Hz) × ancho del bus de datos (en bytes)) / 8 =

Como era de suponer, luego de aplicar DDR, se puede hacer uso de la tecnología QDR (*Quadruple Data Rate* – Cuádruple tasa de transferencia de datos) para cuadruplicar la transferencia de datos utilizando la misma velocidad de reloj real. Esto también lo hemos explicado en el Capítulo 3 (ver la Fig. 3.9). A las memorias SDRAM que utilizan esta tecnología se las conoce como DDR2; DDR-II o QBM (*Quad Band Memory* – Memorias de Cuádruple Banda). Luego, se volvió a duplicar la eficiencia del reloj y apareció la tecnología DDR3 o memorias DD3 de tercera generación. Ver la Tabla 5.6.

Cómo se Expresan los Diferentes DIMM SDRAM; DDR; DDR2 y DDR3

Intel desarrolló una especificación destinada a los fabricantes de memorias para normalizar los detalles de los DIMM SDRAM. Entre otras cosas, figura la

nomenclatura estándar que utilizan los DIMM para ser fácilmente identificados, que se detalla a continuación.

Tabla 5.5. Comparación completa de SDRAM; DDR; DDR2 y DDR-3.

Tecnología	Nombre del DIMM	Nombres alternativos	Velocidad del reloj sin multiplicador	Ancho del bus de datos	Velocidad máxima de transferencia de datos (por cada canal)
SDR	PC66	SDR66	66 MHz	64 bits	0,53 GBps
SDR	PC100	SDR100	100 MHz	64 bits	0,80 GBps
SDR	PC125	SDR125	125 MHz	64 bits	1,00 GBps
SDR	PC133	SDR133	133 MHz	64 bits	1,06 GBps
DDR	PC1600	DDR200 y PC200	100 MHz	64 bits	1,60 GBps
DDR	PC2100	DDR266 y PC266	133 MHz	64 bits	2,13 GBps
DDR	PC2700	DDR333 y PC333	166 MHz	64 bits	2,66 GBps
DDR	PC3200	DDR400 y PC400	200 MHz	64 bits	3,20 GBps
DDR	PC3700	DDR466 y PC466	233 MHz	64 bits	3,73 GBps
DDR	PC4200	DDR533 y PC533	266 MHz	64 bits	4,26 GBps
DDR2	Dual PC1600	PC2-1600; QBM200 y DDR-II-200	100 MHz	64 bits	1,60 GBps
DDR2	Dual PC2100	PC2-2100; QBM266 y DDR-II-266	133 MHz	64 bits	2,10 GBps
DDR2	Dual PC2700	PC2-2700; QBM333 y DDR-II-333	166 MHz	64 bits	2,70 GBps
DDR2	Dual PC3200	PC2-3200; QBM400 y DDR-II-400	200 MHz	64 bits	3,20 GBps
DDR2	Dual PC3700	PC2-3700; QBM466 y DDR-II-466	233 MHz	64 bits	3,70 GBps
DDR2	Dual PC4200	PC2-4200; QBM533 y DDR-II-533	266 MHz	64 bits	4,20 GBps
DDR2	Dual PC5300	PC2-5300; QBM677 y DDR-II-677	333 MHz	64 bits	5,30 GBps
DDR2	Dual PC6400	PC2-6400; QBM800 y DDR-II-800	400 MHz	64 bits	6,4 GBps
DDR3	PC3-6400	DDR3-800	100 MHz	64 bits	6,4 GBps
DDR3	PC3-8500	DDR3-1066	133 MHz	64 bits	8,53 GBps
DDR3	PC3-10600	DDR3-1333	166 MHz	64 bits	10,66 GBps
DDR3	PC3-12800	DDR3-1600	200 MHz	64 bits	12,80 GBps
DDR3	PC3-14900	DDR3-1866	233 MHz	64 bits	14,93 GBps

En alguna parte del DIMM, en una etiqueta o en alguno de sus componentes, aparece PCX-abc-def, donde X es la velocidad de trabajo de SDRAM en MHz,

mientras que a, b, c, d, e y f son valores numéricos que no nos interesan. Por lo tanto, podemos reconocer a un DIMM SDRAM de 100 MHz porque aparecerá como PC133-322-620 y uno de 133 MHz como PC133-322-620.

A los DIMM DDR, se los conoce con la sigla PC, seguida de la velocidad máxima de transferencia de datos. Por ejemplo, un DIMM DDR PC1600, trabaja con una velocidad de reloj real de 100 MHz y una velocidad máxima de transferencia de 1,60 GBps ($1,60 \times 1000 = 1600$ y de ahí PC1600). Ver la Tabla 5.5 para encontrar las nomenclaturas para los diferentes DIMM DDR.

A los DDR2, se los conoce de la misma manera que los DDR, pero anteponiendo el prefijo Dual o PC2 seguido de la velocidad máxima de transferencia en MBps, como por ejemplo, Dual PC 1600 ó PC2-1600 para uno que trabaja con una velocidad de reloj real de 100 MHz y una velocidad máxima de transferencia de 1,60 GBps. Ver la Tabla 5.5.

A los DD3 se les antepone PC3 seguido de la velocidad máxima de transferencia en MBps. Por ejemplo, PC3-10500 para uno que trabaja con una velocidad de reloj real de 200 MHz (aprovechado como 1600 MHz), consiguiendo una velocidad máxima de transferencia de 10,5 GBps. Ver la Tabla 5.5.

Memoria ECC (Error Correcting Code)

La memoria ECC (*Error Correcting Code* – Código de corrección de errores) es capaz de detectar y corregir los errores en el almacenamiento de datos en altas capacidades de memoria. Este tipo de memoria es ideal para servidores de aplicaciones que pueden llegar a tener 64 GB o más memoria y un error en un solo bit sería trágico. Es muy difícil conseguir este tipo de memoria (no confundirla con la memoria con control de paridad) y también muchísimo más costosa que la convencional, pero en servidores, en los cuales la información que corre en ellos vale oro, el costo adicional se justifica.



Para mayor información sobre la arquitectura de los servidores y el manejo de la memoria en esta clase de equipos, puede consultar el libro *Servidores de Redes*, de Editorial HASA.



Para mayor información sobre la instalación, configuración y solución de problemas con los diferentes tipos de memorias presentes en una PC, puede consultar los libros *Reparación y Actualización de PC* y *Reparación y Actualización de PC Visual*, ambos de Editorial HASA.

Capítulo 6



Puertos y Buses de E/S

Los puertos y buses de entrada y salida (E/S) son quienes permiten a las PC comunicarse con el mundo exterior, mientras que los dispositivos de entrada y salida les ofrecen la posibilidad que los usuarios dialoguen con las PC. Todo lo relacionado con los puertos y buses lo analizaremos en detalle en este capítulo. Nos vamos a centrar en los puertos y buses de E/S más difundidos en las PC modernas, pues existen muchos otros que son más específicos para uso industrial o para ciertas aplicaciones que van más allá de las PC hogareña y de oficina.

Puertos de Entrada y Salida

Los puertos y los buses de entrada y salida son los medios para que las PC se comuniquen con el mundo exterior. El nombre de puertos se debe a que cumplen con una función similar a los puertos de barcos. En éstos, los barcos pueden cargar y descargar productos, mientras que en los puertos de entrada y salida se posibilita la transmisión de información entre las PC y cualquier dispositivo externo.

En la Fig. 6.1 veremos el resumen de los puertos y buses de entrada y salida que pueden tener las PC hogareñas y de oficina, de acuerdo a su clasificación principal: **serie** o **paralelo**. En la Tabla 6.1 observamos una comparación completa de las diferentes velocidades máximas de transferencia de datos para todos los puertos y buses que podemos encontrar en la PC.

Los puertos se clasifican según el procedimiento que se utiliza para transmitir los datos:

- Puertos paralelos.
- Puertos serie.

Los puertos paralelos pueden transmitir varios bits al mismo tiempo, a través de varios conductores, uno para cada bit, conectados en paralelo (de ahí su nombre).

Los puertos serie son capaces de transmitir un bit detrás del otro a través de un solo conductor, es decir, en serie.

Tabla 6.1. Comparación de puertos y buses de E/S de una PC.

Nombre del puerto o bus	Velocidad de transferencia máxima
USB 1.0	12 Mbps (1,5 MBps)
USB 1.1	12 Mbps (1,5 MBps)
USB 2.0 (Hi-Speed USB)	100 Mbps a 480 Mbps (12,5 MBps a 60 MBps)
Wireless USB (USB inalámbrico)	53,3 Mbps a 480 Mbps (6,66 MBps a 60 MBps)
Bluetooth 1.2	1 Mbps (0,125 MBps)
Bluetooth 2.0 con EDR	3 Mbps (0,375 MBps)
SCSI	40 Mbps a 5120 Mbps (5 MBps a 640 MBps)
IEEE 1394-1995 (FireWire)	100 a 400 Mbps (12,5 MBps a 50 MBps)
IEEE 1394a-2000	100 a 400 Mbps (12,5 MBps a 50 MBps)
IEEE 1394b	800 Mbps a 3200 Mbps (100 MBps a 400 MBps)
IrDA	921,6 Kbps (115,2 KBps)
FastIrDA (FIR)	4 MBps (32 MBps)
Puerto Paralelo IEEE 1284 (Modo Nibble, Unidireccional)	400 Kbps a 2 Mbps (50 KBps a 250 KBps)
Puerto Paralelo IEEE 1284 (Modo Byte, Unidireccional)	1,2 Mbps a 4 Mbps (150 KBps a 500 KBps)
Puerto Paralelo IEEE 1284 (Modo Compatibilidad, Unidireccional)	1,2 Mbps a 4 Mbps (150 KBps a 500 KBps)
Puerto Paralelo IEEE 1284 Modo ECP (Enhanced Capabilities Port – Puerto de Capacidades Mejoradas, Bidireccional)	6,4 Mbps a 16 Mbps (800 KBps a 2 MBps)
Puerto Paralelo IEEE 1284 Modo EPP (Enhanced Parallel Port – Puerto Paralelo Mejorado, Bidireccional)	6,4 Mbps a 16 Mbps (800 KBps a 2 MBps)
Puerto Serie RS-232	110 bps a 921600 bps (115200 Bps)

bps = Bits por segundo; Bps = Bytes por segundo

Mbps = Megabits por segundo; MBps = Megabytes por segundo

Kbps = Kilobits por segundo; KBps = Kilobytes por segundo

Por características técnicas, a la misma velocidad de sincronización del reloj, los puertos o buses paralelos consiguen velocidades de transmisión muchísimo mayores que las logradas por los serie en el mejor de los casos, debido, simplemente, al método de transmisión que utiliza cada uno. Sin embargo, la gran ventaja del puerto o bus serie es la posibilidad de utilizar un cable mucho más largo

que el paralelo y las conexiones son mucho más seguras. Como la cantidad de conductores que requiere un cable para conectar un dispositivo a un puerto o bus serie es mucho menor que los necesarios para uno paralelo, si éste se encuentra a una distancia apreciable, resultará más económica la utilización de un cable serie. Es por ello que los buses más modernos, USB e IEEE 1394, cuyo objetivo es interconectar muchos dispositivos externos a la PC, son serie.

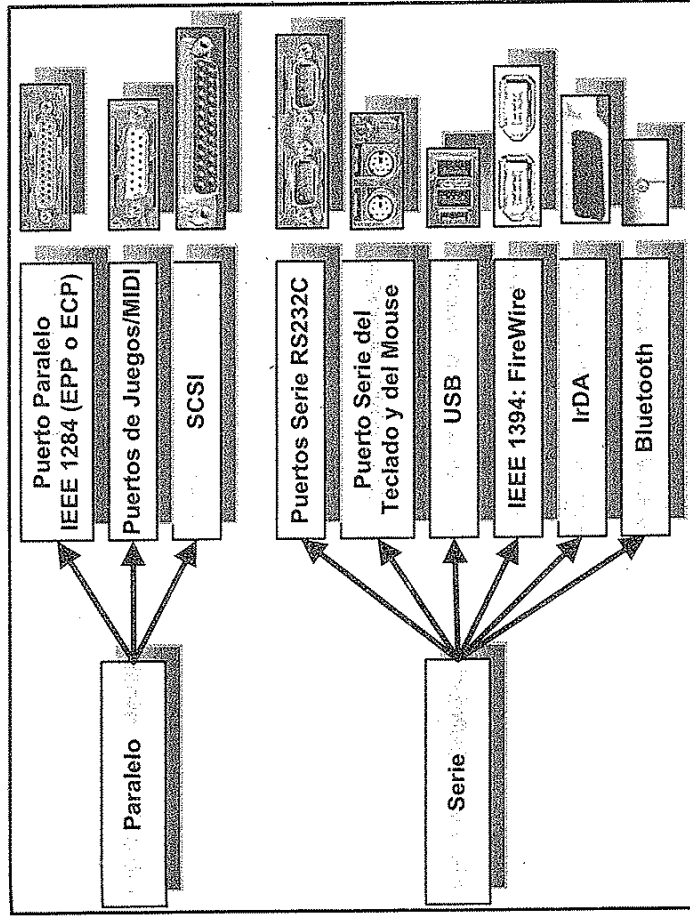











Fig. 6.1. Puertos y buses de entrada y salida más usados en una PC.

Por razones de compatibilidad con sus predecesoras, las PC ofrecen la posibilidad de incorporar un máximo de tres puertos paralelos IEEE 1284 y un máximo de cuatro puertos serie RS-232. Estos números no incluyen el puerto serie que comunica al teclado con la PC ni el puerto de juegos (game port) que permite conectar un máximo de dos joysticks a la PC. Sin embargo, estas limitaciones se pueden superar utilizando conversores de puertos serie RS-232 y de puertos paralelos IEEE 1284 a USB, a IEEE 1394 o a Bluetooth.

La mayoría de las PC vienen equipadas con dos puertos serie RS-232, un puerto paralelo IEEE 1284, otro para juegos y entre cuatro y dieciséis conexiones al bus

USB (ver la Tabla 6.2). Las motherboards modernas los incorporan, pero muchas los eliminan para ganar espacio y, en caso de no tenerlos, deberemos recurrir a tarjetas adaptadoras, como ser tarjetas que provean conexiones IEEE 1394. En la mayoría de las PC modernas, los puertos USB, IrDA, Bluetooth e IEEE 1394 aparecen en la parte delantera del gabinete, para facilitar la conexión de los dispositivos.

Tabla 6.2. Imágenes de los conectores de los puertos y buses de E/S más vistos en las PC modernas.

Nombre del puerto o bus	Figura del(los) conector(es)	Nombre del puerto o bus	Figura del(los) conector(es)
Bluetooth		Puerto Paralelo IEEE 1284	
IEEE 1394		Puerto Serie RS-232	
IrDA / Fast IrDA (FIR)		SCSI	
Mouse PS/2		USB	
Teclado PS/2			

Un adaptador anfitrión SCSI también es un puerto de E/S, aunque su difusión se limita a estaciones de trabajo y servidores de alto rango. Los adaptadores SCSI se analizan en profundidad en el **Capítulo 10: Discos Rígidos**.

Los puertos paralelos se conocen como LPTx (donde x indica el número de puerto paralelo) y los puertos serie se conocen como COMx (donde x indica el número de puerto serie). Las PC normalmente vienen con los siguientes puertos: LPT1, COM1 y COM2.

Todos los puertos poseen un rango de direcciones de entrada y salida (dirección de E/S), mediante los cuales, el procesador puede recibir datos de un puerto específico o enviarle datos, utilizando dicha dirección para acceder al mismo. A su vez, según el tipo de dispositivo que se conecte al puerto, puede que necesite su propio IRQ (línea de solicitud de interrupción), que le permite enviarle un pedido de atención al procesador para que deje de hacer lo que estaba haciendo y el dispositivo conectado al puerto pueda comandar una acción determinada sin intervención del procesador. Esto se analizará en profundidad en el **Capítulo 14: Plug & Play**.

Los diferentes puertos no pueden compartir el mismo IRQ ni las mismas direcciones de E/S, pues si lo hacen no podrán transmitir ni recibir datos correctamente del dispositivo al que están conectados. Las modificaciones de estas configuraciones para que no produzcan conflictos se analizan en profundidad en el capítulo que se nombró en el párrafo anterior.

El problema se presenta con los puertos serie, debido a que COM3 comparte el mismo IRQ que se le asigna a COM1 y COM2 hace lo mismo con COM4. Por lo tanto, si se conecta un dispositivo que requiera un IRQ en COM1 y también se enchufa otro que utilice un IRQ en COM3, ninguno de los dos funcionará correctamente y tampoco lo hará el sistema. En cambio, si el dispositivo conectado en COM3 no utiliza un IRQ, podrá trabajar sin problemas.

Si tenemos un módem que utiliza un puerto serie RS-232 y un mouse conectado a otro puerto serie RS-232, como ambos son dispositivos que requieren de un IRQ, uno de ellos deberá conectarse a un COM1/COM3 y el otro a un COM2/COM4, de lo contrario, se tendrán problemas. Hoy en día, como estos puertos se encuentran incorporados en las motherboards, la configuración se lleva a cabo mediante el CMOS Setup directamente. Conectando el mouse al puerto PS/2 destinado a tal fin o a USB, evitamos los problemas de IRQ compartidos en los puertos COMx con cualquier dispositivo que lo utilice. La gran mayoría de las motherboards modernas traen un puerto PS/2 incorporado para el mouse.

USB 1.0; 1.1 y 2.0: El Bus Serie Universal

Como su nombre lo indica, USB (*Universal Serial Bus* - Bus serie universal) es un bus serie que permite la interconexión de dispositivos de diferentes naturalezas a la PC a través de un único tipo de conector. USB es totalmente compatible con Plug & Play y permite la conexión y desconexión de dispositivos en caliente (sin necesidad de apagar nada). De esta manera se facilita la conexión y configuración de los nuevos dispositivos, pues no es necesario configurar IRQs, DIP switches, agregar tarjetas con interfaces propietarias al sistema ni enchufar una fuente de alimentación externa (si el consumo del dispositivo no es muy grande), todo lo provee el bus.

Como podemos ver en la Tabla 6.1, este bus ofreció en sus primeras versiones (1.0 y 1.1) velocidades que varían entre 1 Mbps (megabits por segundo) y 12 Mbps (1,5 MBps). Aunque, como todos los dispositivos comparten el mismo bus, solamente uno de los que estén conectados a éste podrá alcanzar la velocidad máxima si ningún otro utiliza el bus al mismo tiempo (algo poco probable).

Se debe tener en cuenta que los puertos paralelos IEEE 1284 en modos EPP y ECP ofrecen velocidades máximas de transferencia superiores a las que se pueden

conseguir con USB 1.0 y 1.1. Sin embargo, con USB 2.0, la historia cambia y éste pasa al frente con una gran ventaja. Ver la Tabla 6.1. Inicialmente, USB no fue diseñada para trabajar con dispositivos demasiado rápidos, como unidades de almacenamiento externas o cámaras que capturen video en movimiento. Los dispositivos preferidos para aprovechar USB eran teclados, impresoras, *joysticks*, mouse, cámaras digitales estáticas, scanners, videotelefonos, módems, parlantes y otros similares.

Sin embargo, el bus fue evolucionando y con la implementación de la versión 2.0 (también conocida como *High Speed USB* – USB de Alta Velocidad) alcanza velocidades de hasta 480 Mbps (60 MBps), cuarenta veces más rápido que la máxima velocidad de su predecesor. Esto le permite trabajar con nuevos dispositivos externos más exigentes en este bus: unidades *flash* portátiles (también conocidas como llaves USB o *pen drives*), DVD-RW, discos rígidos, unidades de alta capacidad y otros medios de almacenamiento masivo.

Se debe tener en cuenta que USB 2.0 es totalmente compatible hacia atrás con sus predecesores, por lo cual podemos conectar un dispositivo preparado para USB 1.0 ó 1.1 en un conector USB 2.0. Las motherboards modernas traen incorporadas entre cuatro y dieciseis conectores USB 2.0.



La interfaz que requieren los dispositivos de baja velocidad (hasta 1,5 Mbps) para conectarse a USB es sencilla de implementar y muy económica, por lo cual éstos no se encarecen. Además, como en la mayoría de los casos reciben la alimentación del mismo bus, no requieren una fuente externa. Uno de los dispositivos en donde se nota esta diferencia es en los módem USB y en las unidades *flash* USB, los cuales solamente se conectan a este bus, sin necesidad de alimentación externa.

La conexión de los dispositivos USB al bus se lleva a cabo mediante una topología estrella (*star-tiered*) como se muestra en la Fig. 6.2. La PC cumple el papel de *host* (anfitrión) y *hub* (concentrador) mientras que cada uno de los dispositivos se enchufan a los conectores provistos en los *hubs*. Algunos equipos como la PC, el monitor y el teclado actúan como *hubs* ofreciendo dos o más conectores para que se unan otros dispositivos al bus. Esto simplifica el cableado externo mediante conexiones en cascada distribuidas, aunque no se deben armar ciclos cerrados (lazos o *loops*). La longitud máxima de cada conexión es de 5 metros y se pueden conectar un máximo de 127 dispositivos.

Si nos quedamos sin conectores USB, podemos adquirir *hubs* nuevos y seguir conectando dispositivos. Estos *hubs* adicionales también sirven para aquellos dispositivos que tengan un consumo de corriente que no le puedan entregar otros conectores presentes en la motherboard, pues se deben conectar a una fuente de alimentación externa para alimentar a los dispositivos USB (ver la Fig. 6.3).

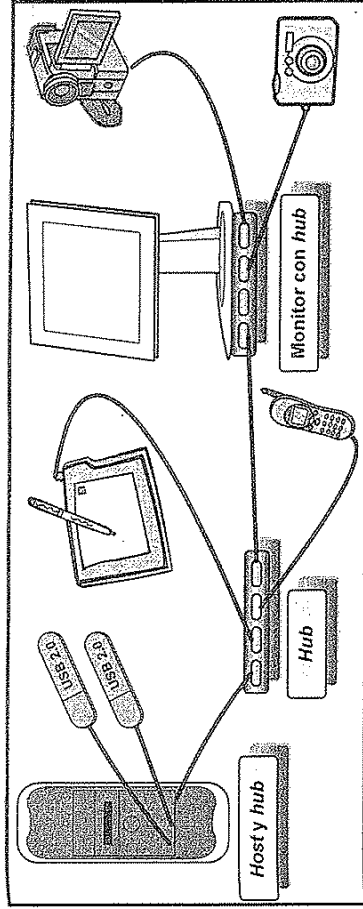


Fig. 6.2. La topología de conexión de dispositivos USB.

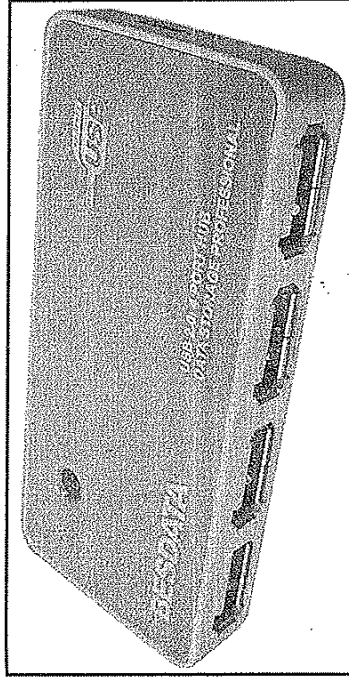


Fig. 6.3.

Un hub USB que ofrece cuatro puertos USB 2.0.

Si bien la mayoría de las motherboards ofrecen conectores USB incorporados, debemos verificar con el manual de la misma a qué versión del bus responden: 1.0; 1.1 ó 2.0, pues si necesitamos conectar dispositivos preparados para USB 2.0 y no tenemos este tipo de conectores en la motherboard, deberemos instalar tarjetas adaptadoras que suelen traer entre dos, cuatro y seis conectores.



USB se ha convertido en el bus de mayor difusión para la interconexión de dispositivos con la PC y también ha logrado gran éxito con otros equipos. Por ejemplo, Home Theatre, equipos de audio, autoestéreos, teléfonos celulares, reproductores y grabadores de DVD y cámaras, ofrecen puertos USB para conectarse a unidades de almacenamiento basadas en memorias diseñadas para este bus serie universal.

Prácticamente todos los teléfonos celulares, cámaras digitales y reproductores de MP3 se conectan a la PC mediante el bus USB.

Wireless USB: Inalámbrico

El bus más exitoso de los últimos años dio un paso más adelante al ofrecer la posibilidad de tener un *host* o varios *hubs* o bien con capacidad de comunicaciones inalámbricas por radio, en vez de utilizar cables y al estándar se lo denomina *Wireless USB*.

Wireless USB soporta un máximo de 127 dispositivos, al igual que USB con cables. Las velocidades de transmisión dependerán de las capacidades de los dispositivos USB inalámbricos, así como de las interferencias. El estándar permite trabajar a velocidades 53,3; 80; 106,7; 200; 320; 400 y 480 Mbps (Megabits por segundo), por lo tanto, según las especificaciones, permite trabajar a la máxima velocidad que lo hace el USB cableado. Sin embargo, debemos tener en cuenta que los protocolos adicionales requeridos por las comunicaciones inalámbricas hacen que se disminuya un poco la velocidad con respecto a la versión cableada. Además, tenemos que considerar que la velocidad de 106,7 Mbps se puede conseguir con una distancia máxima entre transmisor y receptor de 10 metros y, por otro lado, los 480 Mbps solamente con una de 3 metros.

La topología de Wireless USB es idéntica a la explicada para USB con cables, excepto que se pueden reemplazar estos últimos por conexiones inalámbricas, tanto entre un host o hub y un dispositivo como entre un host y un hub o entre dos hubs.



IEEE 1394: FireWire

Inicialmente fue conocido por el número de estándar de la IEEE y luego rebautizado con un nombre más comercial: *FireWire*, de aquí en adelante 1394.

Comparte muchas cosas en común con USB y se lo puede considerar como una versión que en los comienzos era de mucha mayor velocidad que USB 1.0 y ofrecía muchas de las prestaciones del bus serie que analizamos anteriormente.

Al igual que USB, 1394 es un bus serie estandarizado (nada menos que por la IEEE) que permite la interconexión de dispositivos con la PC, es Plug & Play de nacimiento, permite la conexión y desconexión de dispositivos en caliente y también provee alimentación a estos últimos, eliminando la necesidad de fuentes externas en la mayoría de los casos.

USB y 1394 se diferencian en varios aspectos. 1394 es un bus digital serie de alta velocidad al cual se pueden conectar dispositivos que requieren tasas de transferencia de datos muy rápidas, como ser: videocámaras digitales, reproductores de HD-DVD, dispositivos de almacenamiento masivo de datos como discos rígidos, discos de alta capacidad, etc. Las especificaciones 1394-1995 y

1394-2000 establecieron velocidades de trabajo para el bus que comienzan en los 100 Mbps (Megabits por segundo) y llegan a un máximo de 400 Mbps. Luego apareció 1394B, que aumentó la velocidad del bus a 800 y luego hasta 3200 Mbps (400 MBps) y es totalmente compatible hacia atrás con las otras dos, pero este último todavía no ha logrado gran difusión. Ver la Tabla 6.1.

El lanzamiento ubicaba al 1394 como el bus del futuro para los discos rígidos y otros dispositivos de almacenamiento masivo, sin embargo, sus competidores no se quedaron atrás. La versión 2.0 de USB hace que este bus compita con 1394 y las nuevas especificaciones de SCSI tampoco dan tregua a un crecimiento de 1394. Si bien 1394 se postuló como el reemplazante de SCSI como bus para la conexión de dispositivos de almacenamiento, van a seguir coexistiendo muchos sistemas de bus por varios años.

A diferencia de USB, la implementación del 1394 en los dispositivos es un poco más costosa, por lo que aumenta levemente el precio de los mismos y no permite su uso en aquellos que trabajen a bajas velocidades y sean económicos. Por el momento, USB y 1394 van a coexistir, cada uno de ellos con el mercado de dispositivos al que apuntan. Por ejemplo: no saldrán al mercado teclados para 1394, pero sí para USB.

La conexión de los dispositivos 1394 al bus se lleva a cabo mediante una topología árbol (*tree*) como se muestra en la Fig. 6.4. Cualquier dispositivo se puede conectar a cualquier otro, otra vez, sin que se armen ciclos cerrados (*loops*). Todos los dispositivos incorporan conectores 1394 que permiten armar una cadena de conexiones que se abre en cascada como las ramas de un árbol (de allí el nombre de la topología). El bus soporta un máximo de 63 dispositivos.

Una de las mayores ventajas del 1394 es que ofrece conectividad par a par entre dispositivos, de manera tal que estos pueden comunicarse entre sí sin que la PC se entere, es decir, los dispositivos pueden actuar en forma independiente y en conjunto sin que la PC los tenga que controlar. Por ejemplo, como se muestra en la Fig. 6.4, una imagen escaneada con un *scanner* de página completa se puede transmitir directamente a la impresora sin que la PC se entere.

Muchas motherboards modernas ya incorporan conectores 1394. Si tenemos una que no los ofrece, hay que adquirir tarjetas adaptadoras para conseguirlos. Irónicamente, algunos adaptadores SCSI modernos incluyen conectores 1394. Existe un gran mercado detrás del 1394, que está retando nada menos que a los viejos guerreros como IDE, SCSI e inclusive al mismísimo PCI, pero ya han respondido todos con mejoras y han aparecido otros nuevos, por lo cual la batalla por dominar las conexiones de los dispositivos de las PC será larga y dura. Y hasta ahora, el gran ganador, por muchos cuerpos de ventajita, es el USB 2.0.



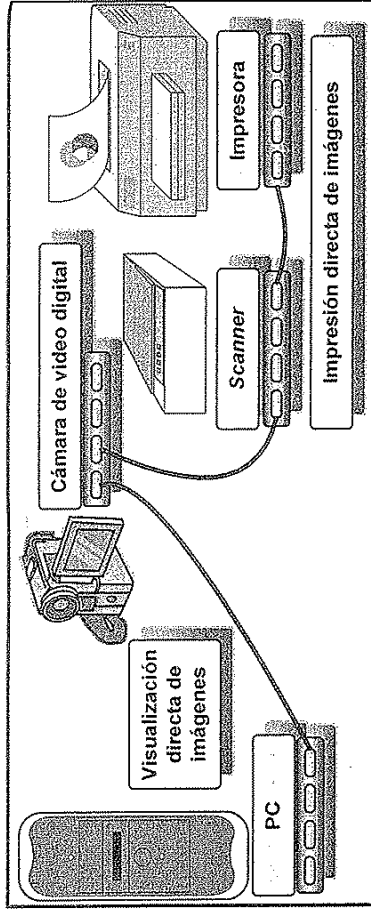


Fig. 6.4. La topología de conexión de dispositivos de IEEE 1394.

Bluetooth

La tecnología inalámbrica Bluetooth permite la comunicación de dispositivos entre sí utilizando radiofrecuencia (la banda de 2,4 a 2,485 GHz). Está diseñada para comunicaciones en distancias cortas, en las cuales, los dispositivos puedan entrar en un radio determinado de redes *ad hoc* conocidas como *piconets*. Cada dispositivo se puede comunicar en forma simultánea con hasta otros siete dispositivos utilizando una única *piconet*. A su vez, cada dispositivo puede pertenecer a varias *piconets* al mismo tiempo, las cuales se establecen en forma dinámica a medida que los dispositivos se aproximan y/o se alejan de los radios de transmisión soportados.

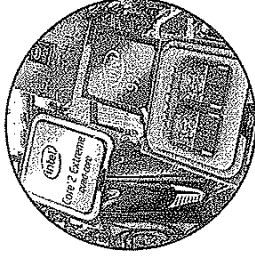
Bluetooth está bastante difundido en los teléfonos celulares, las laptops, las PDA (*Personal Digital Assistant* – Asistente digital personal) y, a raíz de ello, se está incorporando en muchas PC modernas.

Los dispositivos Bluetooth pueden estar diseñados para tres clases, que determinan el radio máximo para participar de una transmisión y recepción con otros:

- **Clase 1.** Su uso es el más común en entornos industriales. El rango máximo es de aproximadamente 100 metros.
- **Clase 2.** Su mayor difusión es en los dispositivos móviles, con un rango máximo de 10 metros.
- **Clase 3.** Solamente provee un rango máximo de 1 metro.

Las velocidades de transmisión de datos son bajas comparadas con Wireless USB, pero tienen la ventaja de estar diseñado para evitar interferencias. La versión 1.2 soporta un máximo de 1 Mbps y la versión 2.0 con EDR (*Enhanced Data Rate* – Tasa de datos mejorada), alcanza los 3 Mbps.

Capítulo 7



Dispositivos de Entrada

Los dispositivos de entrada son aquellos capaces de ingresar información a la PC solamente, por lo tanto son útiles en el ingreso de datos. Todo lo relacionado con los principales dispositivos de entrada lo vamos a analizar en detalle en este capítulo. Pero, no debemos dejar de tener en cuenta que estos dispositivos están en constante evolución y presentan novedosos agregados a sus funciones básicas, para satisfacer necesidades de los usuarios o generar nuevas.

Los dispositivos de entrada más populares en las PC son los siguientes:

- Teclado.
- *Mouse* (ratón).
- *Joystick* (palanca de mando).
- *Gamepad*.
- Volante y pedal.
- *Scanner* (Escáner).
- Tableta digitalizadora.
- Cámara digital estática.
- Videocámara digital y *Webcam*.
- Teclado musical.
- Micrófono.
- Monitor *touch-screen* que simula a un teclado y mouse.

Cualquier otro dispositivo, el cual no figure en la lista anterior, pero pueda enviar información a la PC mediante alguna interfaz estandarizada, es un dispositivo de entrada.

A continuación, analizaremos en detalle los dispositivos de entrada más conocidos en las PC.

El Teclado

El teclado es el dispositivo de entrada estándar de cualquier PC, ya que si bien hay otros dispositivos de entrada capaces de comunicarse con la PC, el ingreso de datos se hará principalmente por el teclado.

La mayoría de las aplicaciones que se van a utilizar en la PC van a requerir el uso del teclado, las planillas de cálculo, por ejemplo, requieren que se les ingresen todos los números necesarios para hacer un cálculo mediante el teclado, así como en un procesador de textos, la función del mismo es el ingreso del texto a procesar.

El único dispositivo que puede llegar a desplazar al teclado en el ingreso de datos es el reconocimiento de la voz humana por parte de la PC. Vale aclarar que desde hace tiempo hay productos en el mercado que permiten el reconocimiento de la voz, y hasta el dictado de textos hacia la PC, aunque la eficacia, velocidad, el entrenamiento que requieren y la ineficacia en ambientes en donde hay varias personas hablando, siguen dejando bastante que desear.

El teclado está formado por un grupo de interruptores o teclas ubicadas en un gabinete plástico. En este mismo gabinete, se encuentra la circuitería encargada de enviar a la motherboard la información de las teclas pulsadas a través de un solo cable en serie. Si no existiese dicha circuitería en el teclado, tendría que salir un cable por cada tecla, algo que resultaría poco práctico.

Un grupo de luces indicadoras se encargan de mostrar el estado de las funciones *Num Lock* (Bloqueo de números), *Caps Lock* (Bloqueo de mayúsculas), y *Scroll Lock* (Bloqueo de desplazamiento). Algunos teclados también utilizan una luz indicadora de funcionamiento.

La mayoría de los teclados cumplen con la misma disposición de teclas que el teclado expandido de IBM, de 101 ó 102 teclas. Aunque, hace un tiempo comenzaron a aparecer teclados ergonómicos, con formas diferentes al convencional para que las manos no se fatiguen al escribir, siendo un ejemplo de éstos el Microsoft Natural Keyboard, quien a su vez tiene tres teclas especiales para manejar Windows. En la actualidad, la mayoría de los teclados están preparados para brindar funciones especiales en Windows (ver la Fig. 7.1) y ofrecen algunas como la de iniciar el Explorador de Internet, teclas para ejecutar aplicaciones frecuentes, apagar y encender la PC, controlar el volumen del sonido, etc. Ya quedaron atrás los tiempos de los teclados de solamente 101 ó 102 teclas.

Contactos de las Teclas

Existen diferentes tecnologías en el sistema de contacto de las teclas:

- Contactos por lámina.
- Contactos por goma conductora.

- Contactos por conmutadores individuales, de doble contacto bañados en oro.

Los contactos por lámina tienen el menor costo de fabricación, por lo tanto son los más utilizados. Su funcionamiento se basa en una lámina de fibra plástica en la que se encuentran impresos los contactos correspondientes a cada tecla. Al pulsar las teclas, los contactos cierran un circuito y el procesador del teclado recibe la información y la deriva a la PC a través de pulsos en serie, según la tecla presionada. Esta vuelve a su posición inicial mediante una pieza en espiral, un plástico flexible o un capuchón de goma con forma de sopapa, que se ubican entre la tecla de plástico y las láminas.

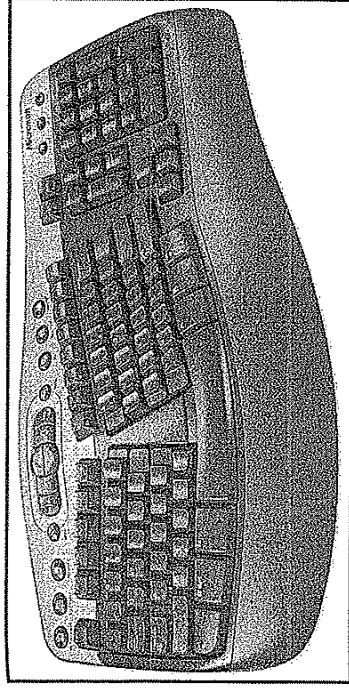


Fig. 7.1.
Un teclado
ergonómico
inalámbrico con
muchas teclas
adicionales.

En este tipo de teclados, a veces se cortan los caminos conductores de plata de las láminas y esto hace que fallen un conjunto de teclas que están conectadas al mismo camino, resultando en un comportamiento extraño del teclado.

Los contactos por goma conductora son similares en su funcionamiento a aquéllos con lámina, con la diferencia que en vez de ser una lámina de fibra es una goma flexible, grafitada en la zona donde deben producirse los contactos.

Los contactos por conmutadores individuales utilizan un conmutador o *switch* de doble contacto bañados en oro u otro material, por cada tecla. Al pulsar una tecla, se cierra el contacto y se envía la información al procesador del teclado, que realiza la misma función descrita en los dos casos anteriores. Esta tecnología de contactos es la de mayor calidad y la más cara, por lo tanto, tendrá que buscar los teclados con estos contactos y pagarlos un poco más caro, pero valen la pena.

Los teclados de buena calidad imprimen los caracteres correspondientes a cada tecla con rayo láser, por lo tanto, esta inscripción durará una gran cantidad de tiempo. En otros teclados de baja calidad, las inscripciones de los caracteres están adheridas al plástico, y al poco tiempo de uso comienzan a borrarse las teclas más utilizadas.



Conectores del Teclado

El conector más utilizado en la actualidad por los teclados de PC es el Mini-DIN 6, conocido más comúnmente por PS/2. Sin embargo, cada vez se están haciendo más populares aquéllos que se conectan al bus USB, el cual hemos explicado en detalle en el **Capítulo 6: Puertos y Buses de E/S**.

También se están popularizando los teclados inalámbricos, los cuales generalmente se comunican utilizando Bluetooth con una central que está conectada a la PC a través de un cable. Estos brindan una gran comodidad para la ubicación del teclado en cualquier lugar dentro del radio de transmisión permitido y liberan de las ataduras de los cables. Aunque, como desventaja, son un poco más caros y consumen baterías.

Configuración y Mantenimiento del Teclado

Si tiene un teclado español o inglés, la disposición de las teclas cambiará notablemente, aunque deberá indicarle al sistema operativo dicha disposición, pues para éste, el código que envía el teclado de la tecla presionada representa a la tecla original del teclado inglés, salvo que se le indique lo contrario.

Los teclados que poseen accesorios adicionales al estándar, como ser teclas especiales más allá de las de Windows, requieren que le especifiquemos al sistema operativo el modelo de teclado o dicho de otra forma, que instalemos el *driver* apropiado, pues sino no se podrán aprovechar estas funciones del mismo y se comportará como uno convencional.

Cuando las teclas comienzan a fallar, lo más probable es que estén sucios los contactos y haya que limpiarlos, pues muchas veces el teclado sufre el derrame de líquidos, pequeños restos de comidas y malos tratos por parte del usuario. También, como mencionamos anteriormente, se pueden cortar las pistas internas de las láminas.

Si el teclado falla demasiado, seguramente lo ha tratado mal y se lo deberá reparar o cambiar por uno nuevo. Si bien un teclado convencional suele ser un dispositivo económico, aquéllos que incorporan funciones adicionales no lo son tanto, por lo cual no hay que descartar la opción de repararlos.

El Mouse

El *mouse* (ratón) es el dispositivo de entrada preferido de las interfaces gráficas como Windows. En la actualidad, todas las PC se venden con este dispositivo incluido, mientras que en sus comienzos, hace muchos años, solamente era un accesorio para los diseñadores gráficos y operadores de sistemas CAD. La facilidad de uso aportada por el mouse a los programas hace que el usuario novato se pueda

comunicar sin dificultades con los mismos y a los usuarios experimentados les simplifica las tareas con un simple click de un botón.

La función básica del mouse es registrar los movimientos físicos en la PC, es decir, trasladar los desplazamientos efectuados con el mouse en la mano a un señalador en la pantalla que indica la posición del mismo, conocido como puntero.

Estructura del Mouse

El mouse está compuesto por una carcasa plástica preparada para ser desplazada tanto por una mesa como por un *pad*, especialmente diseñada para que éste se mueva libremente. La forma de dicha carcasa dependerá del modelo y el fabricante del mouse, pues algunos optan por la forma de jabón y otros buscan una más agradable a las manos, dependerá del usuario decidir cuál de ellos es más cómodo para su mano.

En los primeros mouse, en su superficie inferior sobresalía una parte de una bola de acero recubierta de una goma especial. Dicha bola se encargaba de transmitir el desplazamiento a dos ruedas: una para captar el movimiento vertical y la otra para el horizontal. En la mayoría de los casos, se utilizaba una solución optoelectrónica para transformar los movimientos en impulsos eléctricos. Al girar estas ruedas con perforaciones, un elemento fotosensible (que transforma la luz en impulsos eléctricos) iba captando cuando la luz emitida por un emisor infrarrojo atraviesa o no las perforaciones.

Sin embargo, esta estructura traía muchos inconvenientes, pues cada tanto el puntero del mouse en la pantalla no se deslizaba suavemente, sino que lo hacía en forma de saltos y el mouse parecía quedarse trabado sobre la superficie en la cual se lo estaba desplazando, pues era hora de limpiarlo. Era muy común que se depositara suciedad en los rodillos que transmiten el movimiento de la bola a la rueda perforada, por lo tanto, se debía proceder a limpiar dichos rodillos para que el mouse funcionase perfectamente.

En la actualidad, se utilizan mucho más los mouse ópticos, los cuales no tienen la antigua bola de acero. En cambio, utilizan una solución optoelectrónica directamente contra la superficie sobre la cual se están desplazando, para así transformar los movimientos en impulsos eléctricos (ver la Fig. 7.2). La evolución del mouse óptico es el mouse láser, el cual brinda una mayor precisión y se adapta a prácticamente cualquier superficie.

La gran ventaja de los mouse ópticos y láser es que no tiene piezas mecánicas, son totalmente electrónicos, por lo tanto, no se ensucian y el puntero del mouse siempre tiene un movimiento agradable en la pantalla. La desventaja es que son un poco más caros que los convencionales con bola de acero.



Cualquiera sea la estructura interna de los mouse, en la parte superior, se presentan dos o tres botones, que varían en su utilidad según el programa utilizado. Existen muy pocos programas en la actualidad que aprovechen el botón del medio, por lo tanto, si el mouse posee tres botones, en vez de dos, no presenta demasiados beneficios.

Con el auge de Internet, aparecieron agregados al mouse para darle funciones adicionales a los usuarios, haciendo foco en la navegación por las páginas Web y el trabajo con las barras de desplazamiento verticales específicamente. Surgieron las ruedas de desplazamiento entre los dos botones, así como botones adicionales ubicados sobre los laterales. Todo esto le brindó al mouse más funcionalidades que apuntar y hacer click e hizo que el usuario deba acceder menos al teclado.

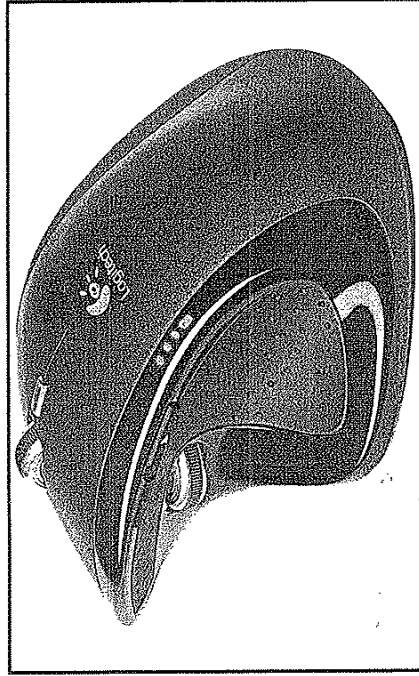


Fig. 7.2.
Un mouse láser inalámbrico con doble rueda de desplazamiento.

Conectores del Mouse

El conector más utilizado en la actualidad por los mouse de PC es el Mini-DIN 6, conocido más comúnmente por PS/2. Sin embargo, cada vez se están haciendo más populares aquéllos que se conectan al bus USB.

También se están popularizando los mouse inalámbricos, los cuales generalmente se comunican utilizando Bluetooth con una central que está conectada a la PC a través de un cable. Estos brindan una gran comodidad para desplazar el mouse sin que se quede trabado el cable siempre y cuando uno se mantenga dentro del radio de transmisión permitido. Aunque, como desventaja, son un poco más caros y consumen baterías.

Si se trata de un mouse con un conector PS/2, se debe tener la precaución de no conectarlo en el lugar en donde va el teclado que utiliza el mismo tipo de conector, pues sino no funcionará ninguno de los dos dispositivos al arrancar la PC. Vale

aclamar que tampoco sufrirán daño alguno, sólo bastará con invertir las conexiones para que todo quede funcionando como corresponde.



El conector PS/2 correspondiente al mouse se suele indicar con un color verde claro.

El Joystick

El joystick, también conocido como palanca de mando, cumple con una función similar al mouse: registrar diferentes movimientos y transmitirlos a la PC, pero el primero es de uso reservado para los juegos. Si bien el joystick no es totalmente necesario para utilizar los juegos, facilita el manejo de los mismos y es esencial para aprovechar los simuladores de vuelo, de carreras de autos, de naves espaciales, etc.

La estructura clásica de un joystick es una pequeña palanca plástica introducida en una estructura de ese mismo material. Dicha palanca se puede mover hacia cualquiera de las direcciones posibles y está conectada a dos potenciómetros, que se encargan de registrar las variaciones según la posición horizontal y vertical en la cual se encontraba ésta. Al variar la posición de los potenciómetros, también lo hace la resistencia que éstos imponen a un circuito de control, el cual se encarga de transformar los cambios de resistencia en variaciones eléctricas.

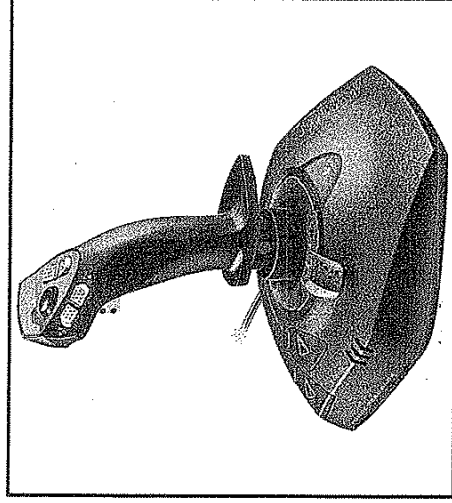


Fig. 7.3.
Un joystick moderno con muchos botones.

De ahí el nombre de joystick analógico, pues la variación captada será proporcional a la posición de la palanca, que no se limita a un par de posiciones predefinidas, como sucedía en los joysticks para las computadoras hogareñas de la prehistoria,

como la Commodore 64, que se limitaban a detectar cuando la palanca estaba a la derecha, arriba, abajo, izquierda o en diagonal.

Los *joysticks* poseen generalmente dos, cuatro o más botones, cuya función dependerá del juego que se utilice, éstos pueden estar ubicados tanto en la misma palanca como en cualquier otra parte de la estructura plástica (ver la Fig. 7.3). Unas ruedas de calibración vertical y horizontal suelen también estar presentes para ajustar la posición central de la palanca.

Debido a la existencia de una buena cantidad de simuladores de vuelo, los fabricantes de *joysticks* decidieron lanzar al mercado modelos especiales para simular la palanca de mando de los helicópteros, aviones y naves espaciales. Todos estos modelos cumplen con la misma función que el *joystick* convencional, con la diferencia que presentan diferentes estructuras para su utilización en ciertos juegos en particular.



El Gamepad

El *gamepad* es muy similar al *joystick* en sus funciones, con la diferencia que no tiene forma de palanca de mando.

Tiene el estilo de los dispositivos para controlar los juegos ofrecidos por las consolas Sega tipo PlayStation 2 y 3 ó Nintendo Wii y Gamecube 64. Se utiliza con dos manos en vez de una y suelen incluir un indicador de direcciones de un lado y un conjunto de botones del otro (ver la Fig. 7.4).

Para algunos tipos de juegos, suelen ser más cómodos que el *joystick*. Sin embargo, depende exclusivamente de los gustos.

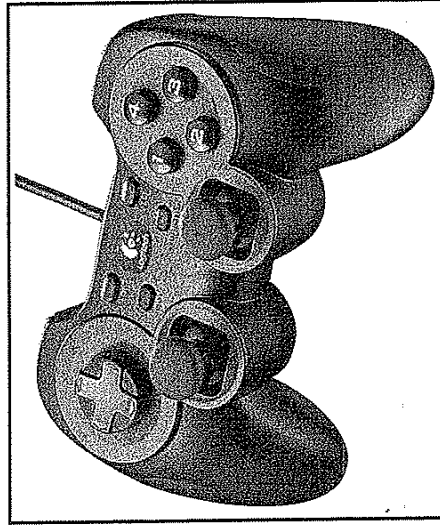


Fig. 7.4.
Un gamepad.

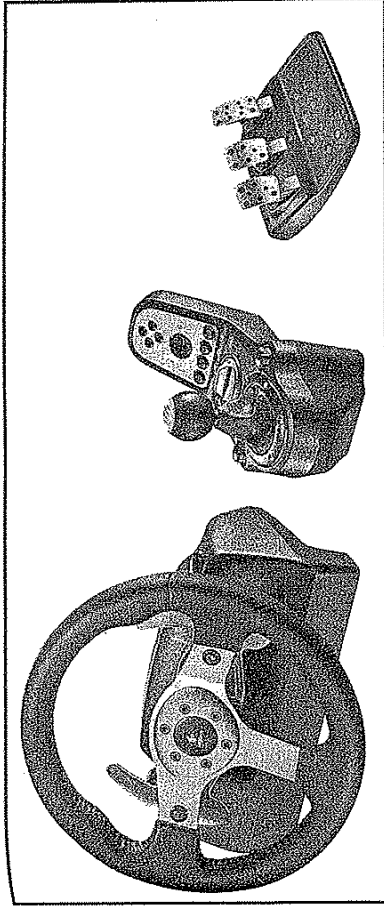


Fig. 7.5. Un dispositivo de volante, palanca de cambios y pedal.

El Volante y Pedal

Para jugar un simulador de Fórmula 1, ni un *joystick* ni un *gamepad* resultan apropiados, la mejor opción es contar con un volante y pedales. Este tipo de dispositivos existen y consiguen que la experiencia de pasar el tiempo con un juego de autos sea muy real (ver la Fig. 7.5).

Los volantes y pedales que brindan el mayor realismo son los que poseen *Force Feedback* (Retroalimentación de esfuerzo), pues generan vibraciones y resistencia del volante cuando, por ejemplo, estamos marchando sobre un empedrado, césped, montañas o queremos girar fuera de la pista.

Conexión de los Dispositivos para Juegos

Los *joysticks*, *gamepads* y volantes de hace muchos años solamente se podían conectar a un puerto de juegos (*game port*), el cual suele estar incorporado en las motherboards o bien formar parte de la mayoría de las tarjetas de sonido. Utilizaban un conector de 15 pines (DB15) que permite enchufar un máximo de dos *joysticks*, aunque para conectar más de uno al mismo conector se debía adquirir un conector Y, especial para dicha función.



La mayoría de los dispositivos de juegos modernos se conectan al bus USB, lo cual facilita la conexión de varios *joysticks*, *gamepads* y volantes con pedales a la PC.

Los dispositivos de juegos que ofrecen capacidades muy específicas se deben utilizar con juegos que saquen provecho de los mismos. Hay que tener en cuenta este tema antes de hacer una inversión en un dispositivo.

El Scanner

El scanner, también conocido como escáner, es un dispositivo de entrada que ofrece la posibilidad de ingresar imágenes digitalizadas a la computadora desde cualquier tipo de material. Mediante el scanner es posible digitalizar fotografías, gráficos, textos, etc.

El funcionamiento del scanner se basa en la optoelectrónica, es decir, en componentes electrónicos emisores y receptores de luz, los cuales emiten una secuencia de bits a la computadora indicando el estado de cada uno de los puntos que conforman la imagen. Esta secuencia de bits es interpretada por el programa que controla al scanner y puede ser visualizada en la pantalla del monitor para retocarla, imprimirla o grabarla en disco mediante cualquier programa de retoque fotográfico como el Photoshop o el software que acompaña al scanner.

Como todos los dispositivos que se conectan a la PC, los scanners tienen diferentes características y la elección dependerá de la clase de trabajos que queramos efectuar con éste.

Los scanners de página completa pueden ser de tres clases:

- Planos (*flatbed*).
- Alimentados por hojas (*sheet fed*).
- Dentro de cartuchos de impresoras de chorro de tinta.

Los scanners de página completa planos son similares a una pequeña fotocopiadora: se introduce la foto, hoja o recorte a explorar, se cierra la tapa y la luz rastreadora se encargará de explorar totalmente la superficie (ver la Fig. 7.6).

La ventaja de estos scanners es que la exploración será casi perfecta, pues no existe la posibilidad que se incline la luz rastreadora y el tamaño de los documentos que permiten explorar será de 8 1/2" x 11" (tamaño carta) como mínimo, aunque existen muchos scanners de página completa que admiten mayores tamaños de hoja. Su gran desventaja es el espacio que ocupan. Para solucionar este problema del espacio, muchas impresoras conocidas como multifunción incorporan un scanner en la parte superior.

Los scanners alimentados por hojas permiten la exploración de hojas de papel sueltas a través de un sistema de alimentación vertical de hojas, similar al de algunas impresoras. Su ventaja es el menor costo y el ahorro de espacio, comparados con uno plano. Pero, su gran desventaja es la imposibilidad de ser utilizado con fotografías, pues sólo recibe hojas de papel con un determinado grosor y que puedan ingresar y salir sin inconvenientes del sistema de alimentación. Además, carece de gran precisión, debido a las dificultades de los mecanismos de movimiento del papel.

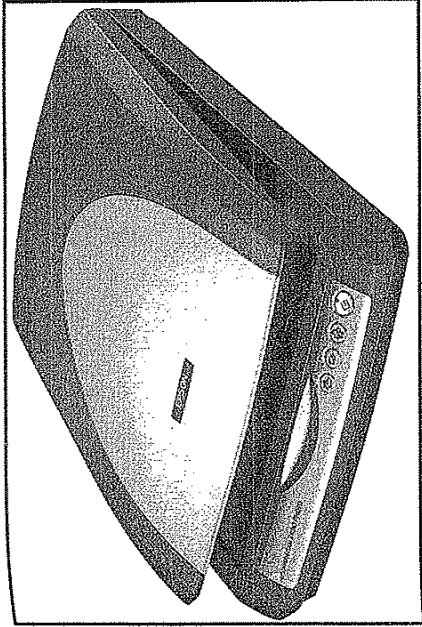


Fig. 7.6.
Un scanner de página completa plano.

Algunas impresoras de chorro de tinta ofrecen la posibilidad de reemplazar uno de los cartuchos de tinta por uno especial que contiene un sensor óptico incorporado y de esta manera, es posible transformar la impresora en un scanner. Es una buena opción económica para aquéllos que no puedan realizar la inversión o no cuenten con suficiente espacio para adquirir uno plano.

Resolución: Óptica e Interpolación

La resolución óptica de un scanner indica la cantidad de puntos que será capaz de explorar a partir de una imagen en papel u otro medio, utilizando sus sensores ópticos, conocidos como CCD, y la precisión de los movimientos de sus mecanismos de desplazamiento. La resolución se especifica en PPP (Puntos Por Pulgada), denominación más conocida como DPI (*Dots Per Inch*).

Los sensores ópticos utilizados por los scanners y las cámaras digitales se conocen con el nombre de CCD (Charge Coupled Devices - Dispositivos Acoplados por Carga) y generalmente se trata de transistores que transforman la luz percibida en cargas o impulsos eléctricos, es decir, en señales que pueden ser digitalizadas.



Como todas las resoluciones, se dividen en horizontal y vertical. En la mayoría de las ocasiones nos encontramos con resoluciones horizontales iguales o mayores que las verticales. Por ejemplo, un scanner de 1200 x 600 DPI puede reconocer 1200 puntos por pulgada en forma horizontal, pero solamente 600 en forma vertical en ese mismo espacio. Es mucho más sencillo aumentar la resolución óptica horizontal que la vertical, pues en la vertical intervienen los difícilmente precisos mecanismos de desplazamiento y por ello encontramos esta asimetría. Debemos tener en cuenta este hecho a la hora de comparar un scanner.

Un *scanner* que posea una resolución óptica de 1200 x 600 DPI, será capaz de transformar una imagen en papel de 2" x 3" en una imagen en pantalla de 2400 pixels x 1800 pixels, tamaño que no será posible visualizar en una pantalla de 1280 x 1024, por lo tanto habrá que recorrer la imagen por la pantalla utilizando las clásicas barras de desplazamiento.

Se podría deducir que a mayor resolución, mayor calidad tendrá la imagen en pantalla y en la impresora, pero también consumirá mayor cantidad de memoria para ser procesada y ocupará más espacio en disco para ser almacenada. El software para capturar imágenes desde el *scanner* permite indicar la resolución con la que se quiere trabajar.

Existe un proceso para aumentar la resolución con la que se pueden rastrear las imágenes, conocido como interpolación. El mismo consiste en rellenar los valores que se encuentran entre un punto y otro de los que conforman la resolución óptica con el promedio de ambos valores y de esta manera se generan puntos adicionales que permiten alcanzar mayores resoluciones.

Este mecanismo se suele realizar fuera del *scanner*, es decir, por software, en el controlador del mismo, por lo que requerirá mayor cantidad de memoria y tiempo del procesador. Los resultados que produce pueden alejar a la imagen explorada de la real, pues se están generando puntos que no existían en la imagen original.

La mayoría de los catálogos suelen incluir las resoluciones que pueden alcanzar los scanners por interpolación, las cuales logran ser tan altas como 9600 DPI. Sin embargo, debemos tener en cuenta solamente la resolución óptica. La interpolación se puede realizar con cualquier software de retoque fotográfico, si fuera necesaria.



Aplicaciones del Scanner

Las aplicaciones de un *scanner* son casi infinitas, pues este dispositivo permite digitalizar cualquier gráfico, logotipo, fotografía y recorte que se desee manipular desde la PC. Pero, además de su utilización para el trabajo con las imágenes exploradas, existen programas capaces de convertir un texto o documento explorado por un *scanner* en texto que se puede modificar desde cualquier procesador de textos.

Estos programas se conocen como OCR (*Optical Character Recognition* - Reconocimiento Óptico de Caracteres) y en la actualidad ofrecen una muy buena efectividad, es decir, si el documento es rastreado correctamente y no presenta manchas de impresión ni tonalidades que confundan a las técnicas de reconocimiento, estos programas serán capaces de transformarlo en texto como si se hubiera tipeado desde el teclado e inclusive manteniendo la disposición del mismo en la hoja y los formatos aplicados a los caracteres (negrita, cursiva, etc.).

Profundidades de Color

Los *scanners* más antiguos ofrecían un máximo de 256 colores, pero la gran mayoría de los modernos trabajan como mínimo con color verdadero de 24 bits: 16,7 millones de colores. Otros permiten la detección de 32, 36 y 48 bits de color. Los más recomendables son los de 32 bits de profundidad de colores, pues brindan una gama más que suficiente para cualquier tipo de aplicación.

Los *scanners* de color deben incluir una plantilla para el ajuste de los colores, para que lo explorado aparezca en pantalla con los colores en fase con los de la imagen original.

La Interfaz del Scanner

La mayoría de los *scanners* modernos suelen ofrecer conexión al bus USB y al puerto paralelo IEEE 1284. Estos últimos incluyen otro conector para poder tener también una impresora conectada al mismo tiempo.

También los hay SCSI e IEEE 1384. Sin embargo, la opción más compatible, flexible y con la mejor relación precio/rendimiento la alcanzan los USB.

El Estándar TWAIN

TWAIN es un estándar en el sistema de adquisición de imágenes y ofrece un sistema de control de *scanners* estándar para que todos los programas que quieran trabajar con cualquier tipo de *scanner* simplemente envíen órdenes al controlador compatible con TWAIN y reciban la información devuelta por el *scanner* en formato TWAIN. De esta forma, al adquirir un *scanner* compatible con TWAIN, se asegurará de que éste podrá funcionar con todos los programas que brinden soporte a este estándar. En la actualidad, todos los programas de retoque fotográfico y OCR ofrecen soporte a dispositivos TWAIN.

La Tableta Digitalizadora

La tableta digitalizadora es un dispositivo de entrada que permite digitalizar figuras y gráficos vectoriales: Este dispositivo cumple una función similar a la del mouse, con la diferencia que permite mayor rapidez y precisión a la hora de realizar los dibujos (ver la Fig. 7.7).

La tableta es un tablero de plástico con sensibilidad electromagnética para la detección de una especie de lápiz que se incluye con la misma y un dispositivo similar al mouse, generalmente equipado con una lupa con dos líneas que se cortan en el centro, para mayor precisión. El circuito electrónico de la tableta digitalizadora es bastante complejo, pues se encarga de detectar y traducir la

posición del lápiz y el señalador en una señal serie que se transmite, generalmente, mediante el bus USB o en forma inalámbrica a través de Bluetooth.

Una de las capacidades principales de este dispositivo consiste en permitir situar un dibujo encima de la tableta y registrar los movimientos del lápiz, posibilitando una especie de calcaado del dibujo. Sin embargo, el precio de estas tarjetas es muy superior al del mouse más caro, por lo tanto su utilización suele quedar restringida a los arquitectos y diseñadores profesionales.

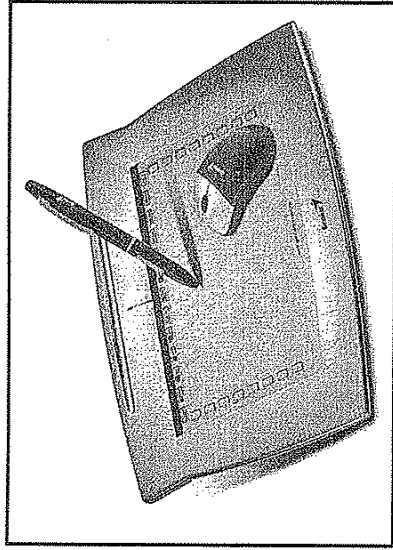


Fig. 7.7.
Una tableta
digitalizadora.

La Cámara Digital Estática

Las cámaras digitales permiten sacar fotos y transmitir las en formatos digitalizados directamente a la PC, eliminando de esta manera el rollo, el proceso de revelado y el posterior escaneo de las fotos que sería necesario con una cámara tradicional.

Estas cámaras poseen lentes que, en vez de transmitir la luminosidad recibida a un rollo, la hacen llegar a un pequeño dispositivo que posee una gran cantidad de sensores tipo CCD, como los utilizados por los *scanners*. Estos sensores se encargan de transformar la luz recibida en impulsos eléctricos y de esta manera consiguen formar una imagen en la memoria de la cámara, con una determinada resolución horizontal y vertical, y con una profundidad de colores (ver la Fig. 7.8).

La resolución máxima depende de la calidad de la lente, de la cantidad de sensores que posee y de la capacidad de procesamiento ofrecida por la cámara. Las especificaciones técnicas de las cámaras definen su capacidad de resolución en megapixels (millones de pixels), que resultan del producto de la resolución horizontal por la vertical.

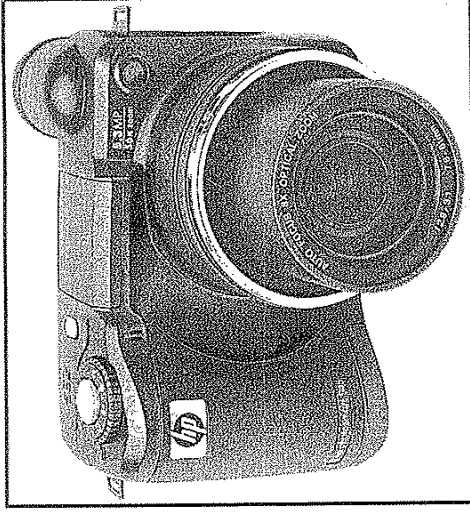


Fig. 7.8.
Una cámara
digital estática.

Resolución máxima (en megapixels) = resolución horizontal (en pixels) × resolución vertical (en pixels)

Por ejemplo, una cámara capaz de capturar 1,92 megapixels puede generar imágenes digitalizadas de 1600 × 1200. Sin embargo, si queremos realizar impresiones de calidad de las fotos o utilizarlas en publicaciones, será necesario buscar una que ofrezca resoluciones muy superiores a las requeridas para visualizarlas en pantalla.

Estas cámaras suelen incluir tarjetas de memoria *flash* para ir guardando allí las imágenes digitalizadas, aunque también existen algunas que utilizan un CD-R, CD-RW o DVD-RW. También poseen una interfaz para transmitir a la PC, utilizando una aplicación que acompaña a la cámara. Las más convenientes y difundidas en la actualidad son aquellas que utilizan una interfaz con el bus USB. Muchos teléfonos celulares modernos ofrecen las funcionalidades de una cámara digital estática y se conectan a la PC mediante un cable USB para la transmisión de las fotos.

La Videocámara Digital y Webcam

Las videocámaras digitales diseñadas exclusivamente para conectarse a la PC, también conocidas como *Webcam* (Cámara para la Web), tienen los mismos principios de funcionamiento detallados para las cámaras digitales estáticas. La diferencia es que agregan movimiento, es decir, van tomando muchas fotos por segundo y las transmiten a la PC mediante una interfaz y ésta las muestra en pantalla con una resolución y profundidad de colores (ver la Fig. 7.9).

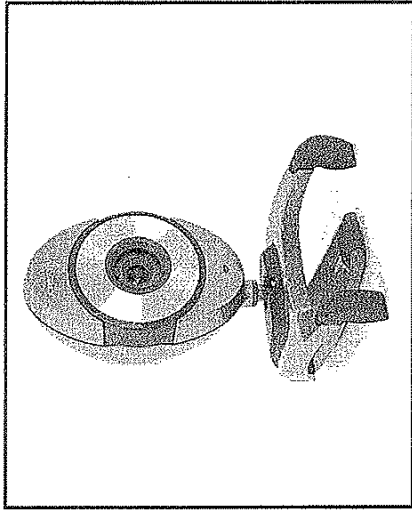


Fig. 7.8.
Una Webcam.

La cantidad de fotos por segundo que es capaz de tomar y transmitir se denomina cuadros por segundo o fps (Frames Per Second). Para tener una sensación de movimiento real, sin saltos, se deberían tener como mínimo 15 cuadros por segundo.

Estas cámaras suelen trabajar con resoluciones horizontales y verticales mucho menores que las de las pantallas, pues el movimiento de 15 cuadros por segundo en resoluciones como 1024 x 768 pixels con 16, 7 millones de colores, requieren una capacidad de procesamiento muy alta y una velocidad de transferencia de datos de casi 34 MBps.

En cambio, en resoluciones como 176 x 144, con las mismas características, se necesitan transferir solamente 1,08 MBps. Es por ello que las resoluciones más comunes están muy por debajo de 1024 x 768 y, más aún, de las utilizadas en las cámaras digitales estáticas.

La velocidad de transferencia de datos necesaria se determina por la siguiente fórmula:

velocidad de transferencia (en bytes por segundo) = (resolución horizontal (en pixels) x resolución vertical (en pixels) x profundidad de color (en bits por pixel) x cuadros por segundo) / 8

Estas cámaras son útiles para realizar videoconferencias y tener conversaciones por la Web en Internet mirándose las caras. Sin embargo, para digitalizar video sus resoluciones no suelen ser lo suficientemente altas.

Las más antiguas se conectaban al puerto paralelo IEEE 1284, sin embargo, en la actualidad la mayoría posee una interfaz para el bus USB. Las cámaras digitales que se conectan a USB 2.0 y a IEEE 1384 pueden ofrecer mejores resoluciones y mayor cantidad de cuadros por segundo.

Capítulo 8



Las Tarjetas de Video

El monitor es el medio que utiliza la computadora para comunicarse con nosotros, así como usamos el teclado, el *mouse* (ratón) y otros dispositivos de entrada (como el reconocimiento de la voz) para comunicarnos con ella. El monitor nos presenta el estado de los programas que se están ejecutando y las respuestas de la PC ante las órdenes que le damos.

El sistema de salida de información a la pantalla de la PC está compuesto por dos componentes:

- La tarjeta de video.
- El monitor.

Tarjetas de Video

Las tarjetas de video poseen una memoria en la cual se almacena toda la información que se debe presentar en pantalla, siendo éstas las encargadas de traducir los contenidos de esa memoria en información que pueda entender el monitor al cual se encuentran conectadas. De esta forma, si se cambia el contenido de la memoria que se aloja en esta tarjeta, se modifica el contenido de la pantalla. Ver la Fig. 8.1.

Se conectan en una de las ranuras de expansión de la motherboard y el monitor se enchufa a la salida de las mismas a través de un conector, el cual puede ser de diferentes tipos. De esta forma, el microprocesador se comunica con la tarjeta de video a través del bus de expansión.

La tarjeta de video está compuesta básicamente por la memoria de video, un controlador de video y un generador de caracteres. El controlador de video se fija cada cierto tiempo en la información que se encuentra en la memoria de video y se la transfiere al monitor como señal de video, constituyendo un proceso de traducción. La cantidad de veces por segundo que se traduce el contenido de la memoria de video al monitor se llama frecuencia de refresco de pantalla.

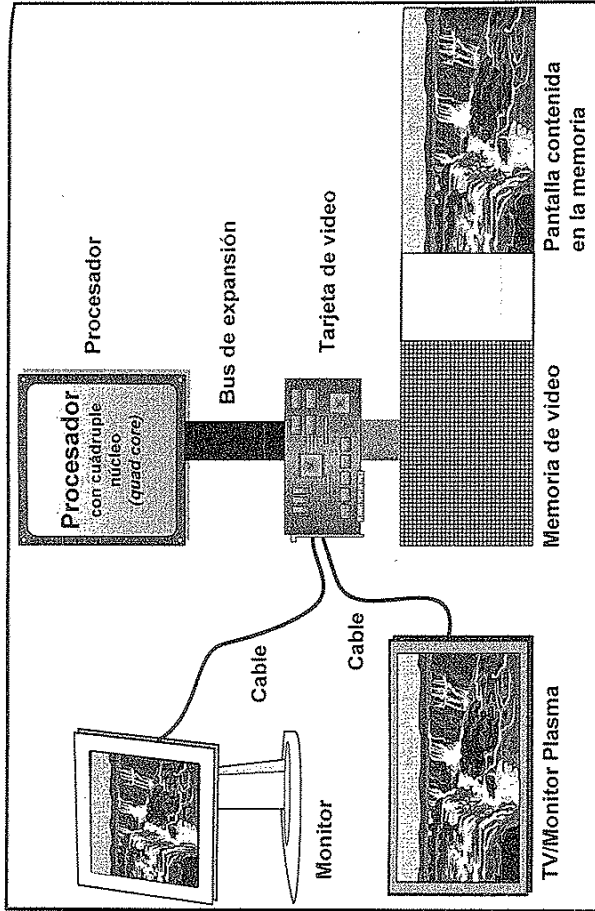


Fig. 8.1. El subsistema de salida de información a la pantalla de la PC.

Resoluciones y Pixels

Podemos definir en forma muy breve y simple a la pantalla de un monitor de PC como una gigantesca matriz (o cuadrícula). En el mismo, cada uno de los pequeños cuadros que conforman la totalidad de la pantalla pueden estar encendidos mostrando un punto de un determinado color o bien apagados, es decir, sin generar ningún tipo de luz en la pantalla.

La memoria que se encuentra en la tarjeta de video se encarga de almacenar el estado de cada uno de los puntos que conforman esta inmensa matriz. Su estado es el color que debe mostrarse o bien apagado, en caso de tratarse de una pantalla a colores. Si hablamos de una antigua pantalla blanco y negro, solamente puede ser de encendido o apagado.



La tarjeta de video es la encargada de recibir los cambios en esa memoria y traducirlos a las señales que puede interpretar el monitor para mostrarlo finalmente ante los ojos del usuario de la PC.

Se conoce como pixel al punto de menor tamaño posible en la pantalla del monitor, siendo la unidad mínima de presentación de información. Cada pixel es un elemento (punto o cuadro) dentro de la inmensa matriz que conforma la pantalla.

Cuando se analicen los monitores, se explicará detalladamente el significado de esta palabra y su relación con la resolución de un monitor.

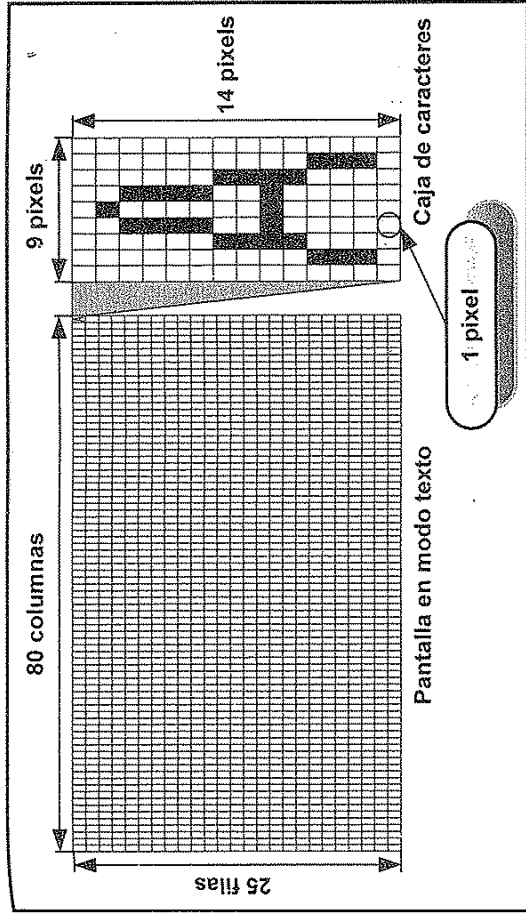


Fig. 8.2. Diagramas de la pantalla en modo texto, la caja de caracteres y los pixels.

Como toda matriz, podemos medir su tamaño y esto lo hacemos contando la cantidad de pixels (puntos o cuadros) que la conforman en forma horizontal y vertical, siempre expresado en ese orden. La resolución se mide en pixels.

Si tenemos una pantalla cuyo contenido está representado en una matriz compuesta por 800 pixels en forma horizontal y 600 en forma vertical, decimos que tiene una resolución horizontal de 800 pixels y una vertical de 600 pixels. Simplificando, esa pantalla tiene entonces una resolución de 800 x 600 pixels, con lo cual la matriz contiene un total de 480.000 pixels.

Modos de Trabajo: Texto y Gráficos

Todas las tarjetas de video que se utilizan en las PC modernas poseen dos modos de trabajo:

- **Modo texto.** Dispone de un espacio compuesto por un conjunto de columnas (horizontal) y filas (vertical) de cajas de caracteres. Reciben este nombre debido a que en cada una de estas cajas puede figurar únicamente un carácter (letra, número o símbolo) dibujado con un color determinado. La pantalla puede mostrar únicamente texto, pues la unidad mínima de

trabajo es la caja de caracteres y no se pueden cambiar los contenidos de los pixels en forma libre. Ver la Fig. 8.2.

- **Modo gráfico.** Se puede modificar el estado de cualquiera de los pixels que componen la pantalla, representando cualquier imagen dentro de las limitaciones de la resolución y cantidad de colores posibles. Ver la Fig. 8.3.

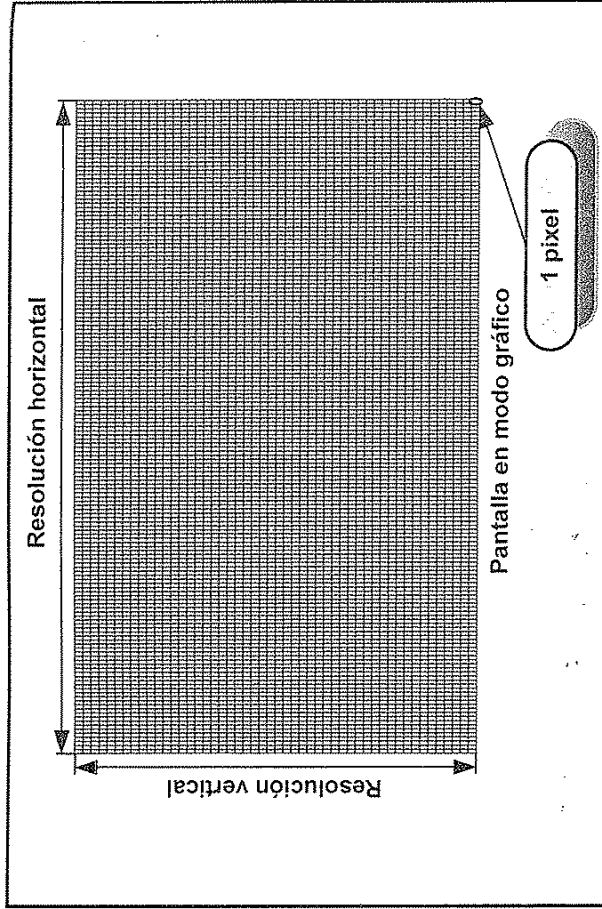


Fig. 8.3. Diagramas de la pantalla en modo gráfico, las resoluciones y los pixels.

Modo Texto

El modo texto tiene menores exigencias y requerimientos para las tarjetas de video así como para el procesador con respecto al modo gráfico, esto se debe a que la memoria de video solamente debe almacenar la matriz conformada por las cajas de caracteres que componen una pantalla de texto. Una pantalla típica de este modo de trabajo está compuesta de 80 columnas x 25 filas, es decir, de 2000 cajas de caracteres, las cuales pueden contener uno de los 256 caracteres posibles.

Utilizando un byte de memoria de video por cada caja de carácter es suficiente para cubrir las 256 posibilidades por caja de carácter (código ASCII). Como cada caja de carácter debe ir acompañada de su estado y, si correspondiera, de su información de color, se agrega otro byte adicional por cada una para almacenar estos datos.

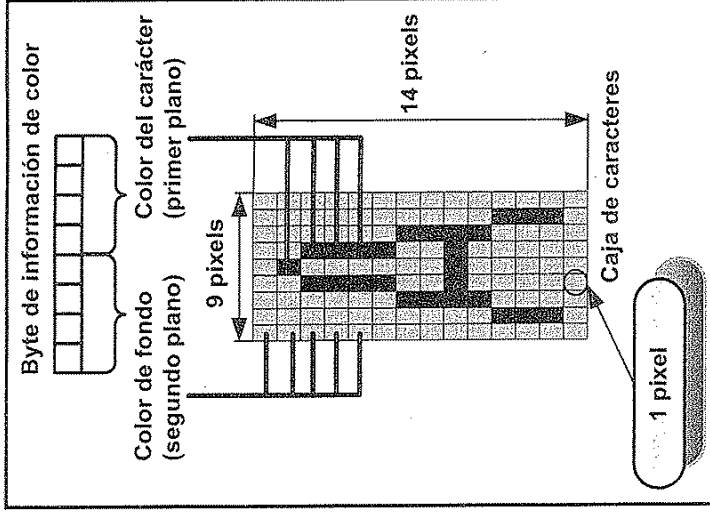


Fig. 8.4. Información de color por cada carácter en modo texto.

El byte que almacena la información de color está compuesto por dos partes (dos nibbles): los primeros cuatro bits contienen el color con el cual se dibuja el carácter y los últimos cuatro bits el color de fondo, es decir, el que rellena la caja de carácter que queda sin dibujar (ver la Fig. 8.4 y la Tabla 8.1). Por ejemplo, si este byte contiene el valor 00011111 binario = 1F hexadecimal, esa caja de carácter tendrá el texto de color blanco (F) sobre fondo azul (1).

Los modos de texto están limitados a 16 colores, por razones de compatibilidad. Esta limitación se debe a que para transferir la información del color, se combinan los tres colores primarios (rojo, verde y azul) posibilitando un máximo de 8 colores (2^3) que combinados con dos intensidades posibles (brillante o débil) posibilitan un máximo de 16 colores (8×2).

Es por ello que con menos de 4 KB (80 columnas x 25 filas x 2 bytes por caja = $4000 / 1024 = 3,91$ KB) de memoria de video se puede administrar una pantalla en modo texto de 80×25 (ver la Tabla 8.2).

Las tablas de definiciones de los pixels, que conforman cada uno de los caracteres de cada modo de texto que poseen las tarjetas de video, se encuentran guardadas en una memoria ROM. La tarjeta se encarga de transformar esa matriz de cajas de

caracteres en una matriz de pixels, buscando la definición de cada carácter en la ROM y luego transmitiendo la información final al monitor.

Tabla 8.1. Los 16 colores del modo texto.

Binario	Hexa-decimal	Decimal	Color Estándar	Binario	Hexa-decimal	Decimal	Color Intensificado
0000	0		Negro	1000	8	8	Gris
0001	1		Azul	1001	9	9	Azul claro
0010	2		Verde	1010	A	10	Verde claro
0011	3		Cyan	1011	B	11	Turquesa claro
0100	4		Rojo	1100	C	12	Rojo claro
0101	5		Magenta	1101	D	13	Magenta claro
0110	6		Marrón	1110	E	14	Amarillo
0111	7		Gris	1111	F	15	Blanco

Tabla 8.2. Modos de texto en los que pueden trabajar las tarjetas de video modernas.

Estándar	Resolución	Caja de carácter	Colores
EGA	40 x 25	8 x 14	16
VGA	40 x 25	8 x 16	16
EGA	80 x 25	8 x 14	16
VGA	80 x 25	9 x 16	16
EGA	80 x 25	18 x 14	2
VGA	80 x 25	9 x 16	2
VGA	80 x 30	8 x 16	16
EGA	80 x 43	8 x 11	16
VGA	80 x 50	8 x 8	16
VGA	80 x 60	8 x 8	16
VGA	132 x 25	8 x 14 ó 9 x 14	16
VGA	132 x 30	8 x 16 ó 9 x 16	16
VGA	132 x 43	8 x 11 ó 9 x 11	16
VGA	132 x 60	8 x 8 ó 9 x 8	16

Cuando trabajamos con la línea de comandos de Windows, Linux y Unix, estamos utilizando el modo texto de la tarjeta de video.

Modo Gráfico y Profundidades de Colores

El modo gráfico es mucho más exigente que el modo texto, pues en la memoria de video se guarda la información de color correspondiente a cada uno de los pixels

que componen la pantalla. Esto implica que a mayor posibilidad de colores por cada pixel, se requerirá mayor cantidad de memoria de video para una misma resolución. De igual modo, si aumentamos la resolución, también lo hace el requerimiento de memoria.

Existen muchas combinaciones de resoluciones y cantidades de colores para trabajar en este modo. Cada tarjeta de video está preparada para trabajar con un conjunto de estas combinaciones. Una limitación a las mismas, además del soporte por parte de la tarjeta de video, es la cantidad de memoria que se encuentra instalada en la misma.

La cantidad de colores que se pueden mostrar en forma simultánea en una pantalla es un subconjunto seleccionado de una paleta de colores con la cual puede trabajar la tarjeta de video. Por ejemplo, en un modo de 256 colores simultáneos, éstos se seleccionan de una paleta de 262.144 colores. En la pantalla pueden coexistir entonces, solamente esos 256 que se eligieron.

La paleta de colores se puede cambiar en cualquier momento sin necesidad de volver a dibujar toda la pantalla. Un ejemplo sería utilizar toda la gama de azules en pantalla y poder ir cambiando la paleta para la gama de otros tonos, como verde, turquesa o celeste y lograr un efecto visual interesante. Esta técnica se emplea mucho en los videojuegos.

Por ejemplo, para representar 256 colores, se necesitan 8 bits (1 byte) por pixel, que representan 2^8 posibilidades = 256. En la Tabla 8.3 se resumen la cantidad de bits y bytes necesarios para cada cantidad de colores.

Para saber la cantidad de memoria requerida para mostrar una determinada resolución con cierta cantidad de colores, se debe multiplicar la resolución horizontal por la vertical, para conseguir la cantidad de pixels totales que tiene que presentar la pantalla. Este número deberá ser multiplicado por la cantidad de bits que se necesitan para mostrar la cantidad de colores deseada por pixel y luego dividido por 8 para representar el valor en bytes. En la Tabla 8.3 podemos encontrar la cantidad de bits necesarios para cada cantidad de colores.

En la Tabla 8.4 podemos ver un resumen de todos los modos de video estandarizados con sus nombres, sus resoluciones y sus relaciones de aspecto, incluyendo las normas correspondientes a la televisión de alta definición (HDTV).

Memoria de Video

Para saber la cantidad de memoria requerida para mostrar una determinada resolución con cierta cantidad de colores se debe multiplicar la resolución horizontal por la vertical para conseguir la cantidad de pixels totales que debe presentar la pantalla. Este número deberá ser multiplicado por la cantidad de bits que se necesitan para mostrar la cantidad de colores deseada por pixel y luego

dividido por 8 para representar el valor en bytes. En la Tabla 8.3 podemos encontrar el número de bits necesarios para cada cantidad de colores.

Tabla 8.3. Número de bits para cierta cantidad de colores.

Bits		Colores	
1	1/8	2	
2	1/4	4	
4	1/2	16	
8	1	256	
15	1 7/8	32.768 (32 K)	
16	2	65.536 (64 K)	Llamada también High Color o Color de Alta Densidad.
24	3	16.777.216 (16,7 millones)	Llamada también True Color o Color Verdadero.
32	4	4.294.967.296 (4.295 millones)	Capacidad de trabajar con la gama completa de colores CMYK. También llamada True Color o Color Verdadero. La mayoría de los sistemas operativos modernos utilizan 32 bits de profundidad de color de otro modo. Usan 24 bits para la definición de 16,7 millones de colores y los 8 bits adicionales (un byte) para definir la transformación alpha de un color, es decir, su nivel de transparencia, el cual puede tener 2 ⁸ = 256 niveles (de totalmente opaco a transparente).

Tabla 8.4. Resoluciones para las diferentes normas.

Nombre	Resolución horizontal	Resolución vertical	Relación de aspecto
CGA	320	200	16:10
QVGA	320	240	4:3
VGA	640	480	4:3
NTSC	720	480	3:2
PAL	768	576	4:3
SVGA	800	600	4:3
WVGA	854	480	16:9
XGA	1024	768	4:3
HD720 o 720i	1280	720	16:9
SXGA	1280	1024	5:4
SXGA+	1400	1050	4:3
UXGA	1600	1200	4:3
WSXGA+	1680	1050	16:10
HD1080 (1080i y 1080p)	1920	1080	16:9
WUXGA	1920	1200	16:10
QXGA	2048	1536	4:3
WQXGA	2560	1600	16:10
QSXGA	2560	2048	5:4

Por ejemplo, si se quiere una resolución de 1024 x 768 pixels y se desean 16,7 millones de colores se necesitarán 24 bits de color para mostrar esos colores

(224= 16,7 millones), 1024 x 768 x 24 = 786.432 pixels x 24 bits = 18.874.368 bits; 18.874.368 / 8 = 2.359.296 bytes = 2,25 MB.

Tabla 8.5. Memoria de video necesaria en megabytes (MB) para las diferentes resoluciones y cantidad de colores.

Resolución	Cantidad de colores			
	16	256	65.536	16.777.216
Horizontal	Bits x pixel	4	8	16
	Bytes x pixel	0,5	1	2
Vertical				
640	0,15	0,29	0,59	0,88
800	0,23	0,46	0,92	1,37
1024	0,38	0,75	1,50	2,25
1152	0,47	0,95	1,90	2,85
1280	0,59	1,17	2,34	3,52
1280	0,63	1,25	2,50	3,75
1600	0,78	1,56	3,13	4,69
1600	0,92	1,83	3,66	5,49
1920	0,99	1,98	3,96	5,93
1920	1,10	2,20	4,39	6,59
1920	1,32	2,64	5,27	7,91
2048	1,50	3,00	6,00	9,00

Como se puede observar en la Tabla 8.5, la cantidad de colores simultáneos ha aumentado respecto de los 256 originales de las primeras tarjetas VGA. Las tarjetas que manejan resoluciones a partir de 640 x 480 y 256 colores simultáneos, resolución que requiere un mínimo de 512 KB (1/2 MB) de memoria de video, se conocieron como Súper VGA.



Vale aclarar que no solamente se necesita cierta cantidad de memoria, sino también el BIOS de la tarjeta de video debe soportar estos modos gráficos. Además, se debe tener en cuenta que el monitor debe poder trabajar con los mismos, como veremos más adelante.

En las tarjetas de video aceleradoras de gráficos 3D, la memoria de video encuentra otra utilidad: la de almacenar temporalmente las texturas que se utilizan para revestir las figuras tridimensionales en tiempo real y así lograr una mayor velocidad en este complejo proceso. Por esta razón, aparecen tarjetas con más de 512 MB de memoria de video, de los cuales se emplea lo que queda sin utilizar para almacenar los pixels que se deben mostrar en pantalla como almacenamiento de texturas (básicamente como un gran buffer). Es por ello que si bien en la

Tabla 8.5 figuran 12 MB de memoria de vídeo necesarios para la mayor profundidad de colores en 2048 x 1536, hay tarjetas con mucha más memoria.

Arquitectura Básica

Para comprender el funcionamiento de las tarjetas de vídeo modernas, primero vamos a analizar la arquitectura básica de las mismas, tomando como base una tarjeta VGA sin capacidades de aceleración gráfica 2D ni 3D, para entender los pilares fundamentales y luego avanzaremos hacia estos temas con el esquema conceptual ya comprendido.

Las primeras tarjetas VGA y Súper VGA básica carecen de capacidades de procesamiento. Esto significa que el procesador principal de la PC es quien se debe encargar de efectuar todos los cambios en la memoria de vídeo para que éstos se realicen en la pantalla. Por ejemplo, para dibujar una línea, el procesador deberá realizar todo el trabajo de calcular los puntos en la pantalla y las direcciones de memoria en donde deben aparecer puntos de color blanco, lo mismo si se desea dibujar un círculo, rellenar un polígono cerrado, etc. Esto implica un gran trabajo del procesador para efectuar los trabajos gráficos y le quita tiempo para efectuar otras tareas.



Estos adaptadores se encargaban simplemente de traducir el código en la memoria de vídeo a una señal que entienda el monitor, lo que se llama refresco de pantalla.

La arquitectura básica de una tarjeta VGA o Súper VGA (ver la Fig. 8.5) está compuesta por:

- La memoria RAM de vídeo (ya explicada anteriormente).
- El controlador gráfico.
- El secuenciador.
- El controlador de vídeo.
- El controlador de atributos.
- El serializador.
- El conversor digital-analógico.
- Los registros.
- El BIOS de vídeo.

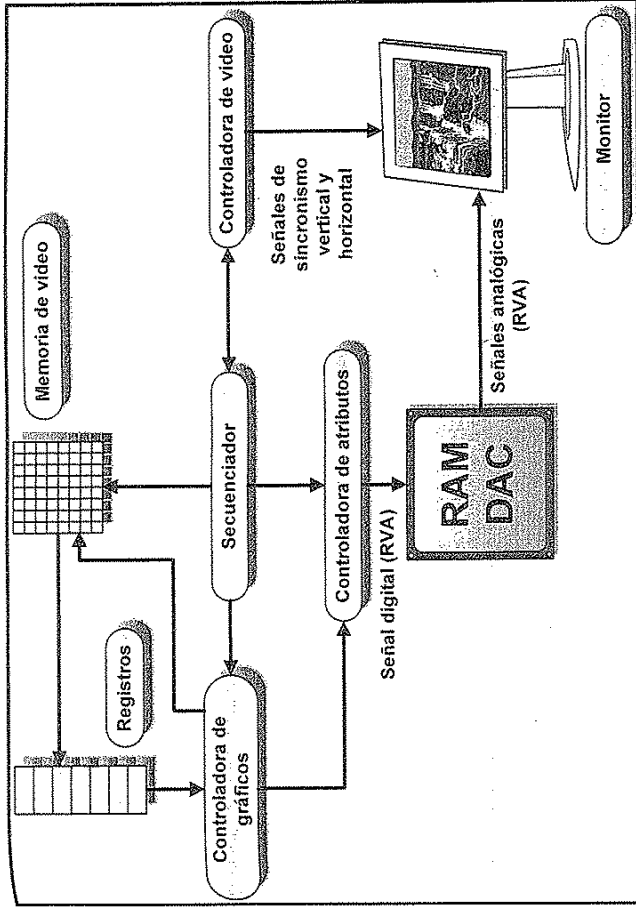


Fig. 8.5. Arquitectura básica de una tarjeta VGA o Súper VGA antigua.

Señal Analógica

Las tarjetas VGA convencionales tienen una salida analógica, al contrario de sus predecesores que utilizaban una señal digital para llevar información al monitor. En lugar de utilizar señales sí o no (encendido o apagado; 1 ó 0) para cada color, el nivel de cada tono puede variar en forma continua (ver la Fig. 8.6). A través de esta tecnología, la cantidad de colores que se pueden obtener es teóricamente infinita.

En las placas VGA, los niveles de señal para cada color principal estaban limitados a 64 valores posibles. Por lo cual, se podía obtener un máximo de $64^3 = 262.144$ tonos. En las Súper VGA, a 256, limitados entonces a un máximo de $256^3 = 16.777.216$ tonos.

El Conversor Digital-Analógico (RAM DAC)

El conversor digital-analógico (DAC - Digital Analog Converter) es el encargado de generar una señal analógica a partir de una digital. Los conversores programables se encuentran en la tarjeta de vídeo y se componen por tres conversores, uno para cada color: rojo, verde y azul.

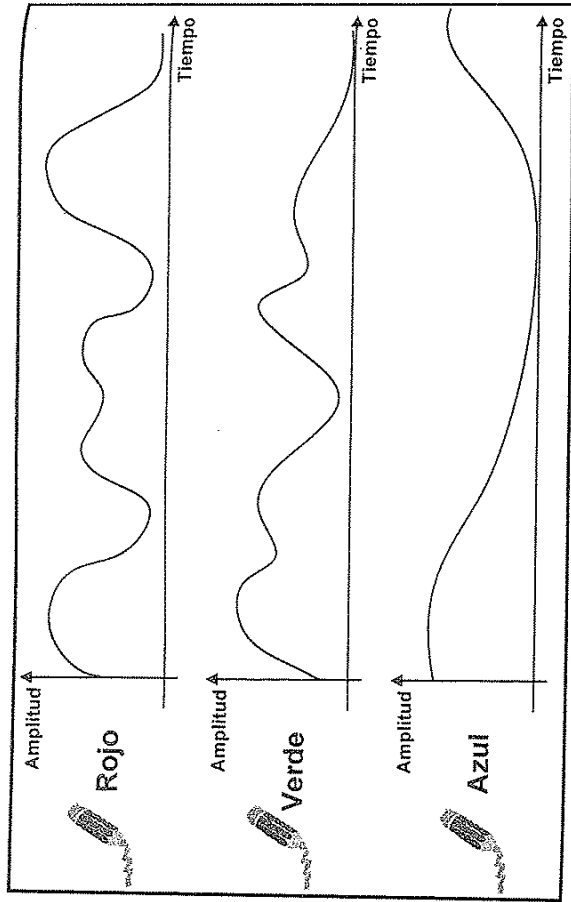


Fig. 8.6. El color se representa mediante tres señales analógicas de amplitud variable para cada color primario.

Los RAM DAC, también conocidos como *Palette DAC* (DAC de Paleta) poseen una memoria RAM incorporada en el conversor para la información correspondiente a la paleta de colores, por lo cual son más veloces que los DAC que deben ir a buscar esta información en chips RAM fuera del conversor. Aquellos que son integrados se encuentran incorporados dentro del procesador de la tarjeta de video y consiguen un rendimiento muy superior a los que se ubican en un chip aparte, fuera del procesador.

En el caso del VGA primitivo, cada DAC tiene una entrada de datos de 6 bits, lo que le permite distinguir entre $2^6 = 64$ tonos para cada uno de los colores primarios (rojo, verde y azul). Estos conversores incluyen una tabla de búsqueda de 64 entradas en las cuales se almacenan los niveles posibles de cada uno de los tres colores. Cuando reciben cada código de 6 bits, el conversor correspondiente toma el nivel de color de su tabla y genera una señal analógica. Este DAC era de 18 bit (3×6 bits). Las tarjetas de video modernas trabajan con DAC de 32 bits, que permiten generar 4.096 millones de colores.

La generación de la señal es el último proceso, por lo tanto, las señales digitales están disponibles antes del ingreso en los conversores para posibilitar la emulación de señales compatibles con los antiguos modos de trabajo de EGA y CGA. Es por

ello que hasta la más moderna tarjeta de video es compatible hacia atrás con sus predecesores.

La información del color llega al monitor por tres conductores separados, constituyendo tres informaciones independientes. La amplitud de la señal determina la tonalidad de los colores. A mayor amplitud, mayor intensidad y viceversa. Por ejemplo: si todas las señales están a máxima amplitud, se obtendrá un punto blanco que estará representado por la siguiente expresión: 100 % rojo + 100 % verde + 100 % azul = blanco.

Las tarjetas de video de mayor velocidad utilizan circuitos RAM DAC que trabajan a velocidades de reloj de 400 MHz o mayores, ofreciendo un rendimiento excelente para aplicaciones que impliquen una exigencia grande en el procesamiento de gráficos 3D en tiempo real, muy altas resoluciones y profundidades de color. En este tipo de tarjetas, se suele incluir la velocidad de reloj del RAM DAC como una de las características principales de las mismas, siendo las más comunes 300; 350; 400; 450 MHz y superiores para las de muy alto rendimiento.



La resolución, profundidad de colores y frecuencia de refresco son los tres factores que, al incrementarse, aumentan los requerimientos de velocidad del RAM DAC.

El Controlador Gráfico

Éste se encarga de administrar el acceso desde el microprocesador a la memoria de video, mediante la cual se modifica la información en pantalla. El acceso a la memoria de video en estas placas es más complicado que un simple acceso a memoria, se deben utilizar cuidadosamente los registros del controlador gráfico y respetar las reglas de escritura y lectura de la memoria de video. Cuando este controlador está activado, realiza operaciones lógicas con la información que ingresa y sale, cuando se lo desactiva, estas operaciones deberán ser efectuadas por el procesador central.

El Controlador de Atributos

El controlador de atributos administra la paleta de colores, posibilitando de esta manera que cualquier atributo de color contenga uno de los colores de todos los posibles. También controla otros efectos en modo texto: cursor, parpadeo, borde de color, desplazamiento horizontal suave del pixel, subrayado, etc.

El Controlador de Video

El controlador de video es el encargado de producir la señal de video que será enviada al monitor. También es el responsable de regular los tiempos y generar las

señales de sincronismo horizontal y vertical para el correcto funcionamiento del monitor.

Por ejemplo, algunos programas que modifican los registros en dicha sección son los que vienen acompañados de la tarjeta de video, permitiendo seleccionar entre diferentes modos de operación o frecuencias de barrido. Si al seleccionar alguna de las opciones de estas utilidades la imagen se achica, salta el horizontal o se siente un silbido proveniente del monitor, hay que volver inmediatamente al modo anterior para no dañar el monitor. Esto se produce debido a que el monitor no soporta algunos modos con los cuales sí puede trabajar la tarjeta de video.

El Serializador

El serializador se encarga de leer la información de la memoria de video y enviarla por un solo conductor en forma codificada.

El Secuenciador

Se encarga de controlar los tiempos y coordinar el funcionamiento, a su vez administra el plano al que se accederá cuando se pretenda escribir en la memoria de video.

Los Registros de la Tarjeta de Video

Los registros son puertos de E/S en los cuales se puede escribir o leer información. Es una forma de seleccionar el modo de operación de la tarjeta en forma directa, sin pasar por el BIOS de la misma, que es un proceso mucho más lento, aunque se utilice *ROM Shadow*.

Las tarjetas de video modernas tienen decenas de registros con diferentes funcionalidades. Debido a que si se comportaran como puertos de E/S tradicionales, ocuparían demasiadas direcciones de E/S, para acceder a los registros, es necesario realizar dos operaciones: primero se debe seleccionar el registro que se quiere leer o escribir y luego se lee o modifica el mismo.

Hay registros que controlan el estado general de la tarjeta y otros que están relacionados con cada uno de los componentes de la placa ya descritos:

- **Controladora de video.** Modifican cómo la placa genera la señal de video.
- **Secuenciador.** Incluyen funciones de control de tiempos de la tarjeta, seleccionan los planos de escritura, el juego de caracteres activo, etc.
- **Controladora gráfica.** Incluyen funciones de relleno de áreas, encendido de pixels, transferencias de bloques de memoria y otras operaciones especiales de procesamiento de la memoria de video.

El BIOS de Video

Esta memoria ROM se encarga de las operaciones básicas de video, especialmente de aquéllas referidas al modo texto y de efectuar operaciones como la modificación de los registros y la selección del modo de trabajo de la tarjeta.

A lo largo de los años, las tarjetas evolucionaron constantemente, se fueron agregando más y más modos de trabajo y se hizo necesario establecer un nuevo estándar para que todos los programas pudieran trabajar de la misma manera con los modos que ofrecían las placas, independientemente del fabricante.

VESA (*Video Electronics Standards Association* - Asociación de Normas Electrónicas de Video) es un consorcio creado en Abril de 1989 que reúne a un grupo de fabricantes de tarjetas de video para establecer modos de trabajo estándar. VESA estableció un modo de operación estándar que permite a los programas detectar y utilizar las extensiones de la tarjeta.

El primer estándar de VESA fue el modo 800×600 de 16 colores y se fueron agregando todos los modos siguientes. Todas las versiones de este estándar, como todo en las PC, son compatibles hacia atrás, lo cual significa que se le agregan modos de operación, manteniendo los existentes.

Las tarjetas de video modernas incluyen el soporte al estándar VESA en el BIOS de video.

Factores Que Influyen en la Velocidad

A medida que se fueron aumentando las resoluciones y la cantidad de colores, se necesitaba cada vez mayor manejo de información y más cantidad de memoria de video, lo cual implica que el procesador deberá trabajar cada vez más para transferir grandes bloques de información desde éste hacia la memoria de video. Para ver la diferencia, debemos recordar que para 640×480 y 256 colores se requieren solamente 307 KB (0,29 MB) de memoria de video, lo cual significa que modificar la pantalla completa implicaría escribir solamente 307 KB, mientras que para 1600×1200 y 32 bits de profundidad, se requieren 7,32 MB; casi 25 veces más.

Las aplicaciones hacen uso cada vez más intensivo de video, son cada vez mayores las resoluciones y la cantidad de colores y esto exige mucho más a las tarjetas de video para que trabajen con mayor velocidad para no transformarse en un cuello de botella para los sistemas y aplicaciones modernas.

Hay numerosos factores que influyen en la velocidad de la tarjeta de video y en consecuencia, del subsistema gráfico de una PC:

- El bus de expansión al cual se conecta la tarjeta.
- Las tecnologías de la memoria de video que utiliza la tarjeta.