



MEMORIA

Programación II



Temas

- Subsistema de Memoria
- Organización de Memoria Principal
- Memoria Cache



➔ Subsistema de memoria

Repaso: Componentes de la computadora

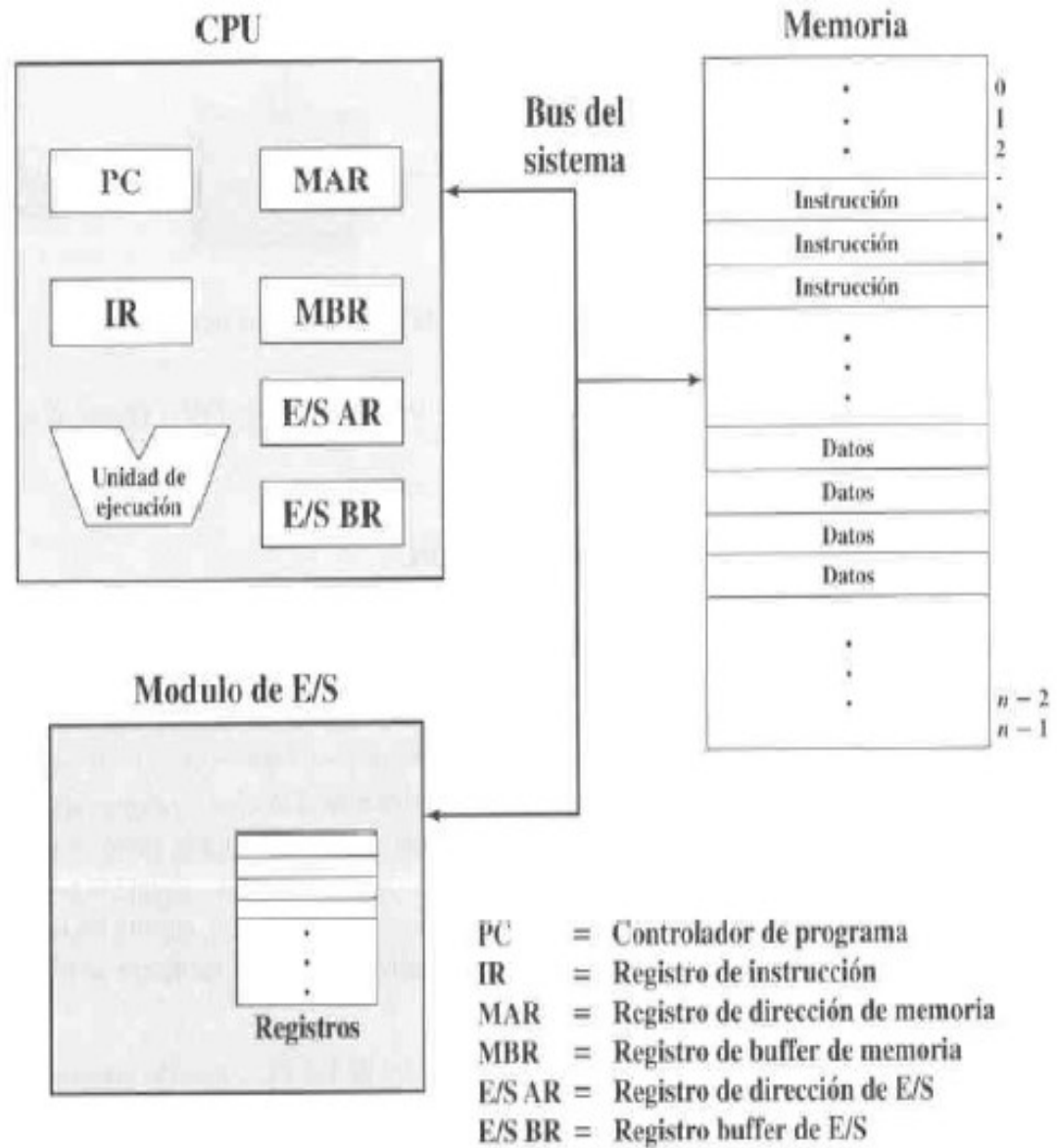


Figura 3.2. Componentes del computador: esquema de dos niveles.

Buses

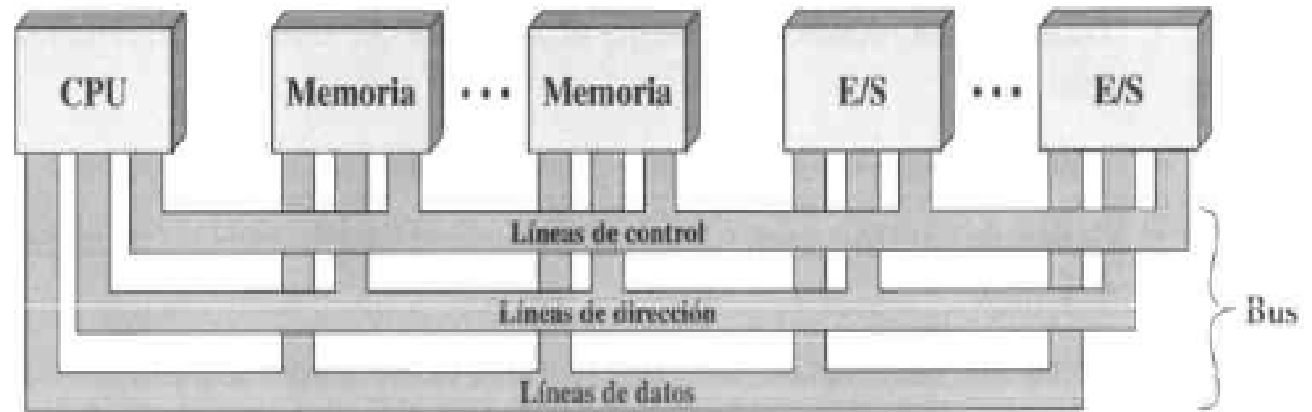


Figura 3.16. Esquema de interconexión mediante un bus.



Contexto actual (1)

- ▶ Velocidad del procesador: se duplica cada 18 meses (sin variar su precio) la cantidad de instrucciones ejecutadas por segundo.
- ▶ Memoria: se cuadruplica su tamaño cada 36 meses (al mismo precio). Velocidad aumenta a razón de un 10% anual.

Contexto actual (2)



- A medida que aumenta la brecha entre las velocidades del procesador y de la memoria, las distintas arquitecturas buscan tender un puente sobre esta brecha.
- Una computadora típica suele tener distintos tipos de memoria, desde una rápida y cara (registros) hasta una barata y lenta (discos).

Contexto actual (3)



- La interacción entre los diferentes tipos de memoria se aprovecha de forma tal que se logra un comportamiento, por parte de la computadora, equivalente al que tendría con una memoria única, grande y rápida, cuando en realidad tiene distintos tipos de memoria trabajando en forma coordinada.

Jerarquía de memorias



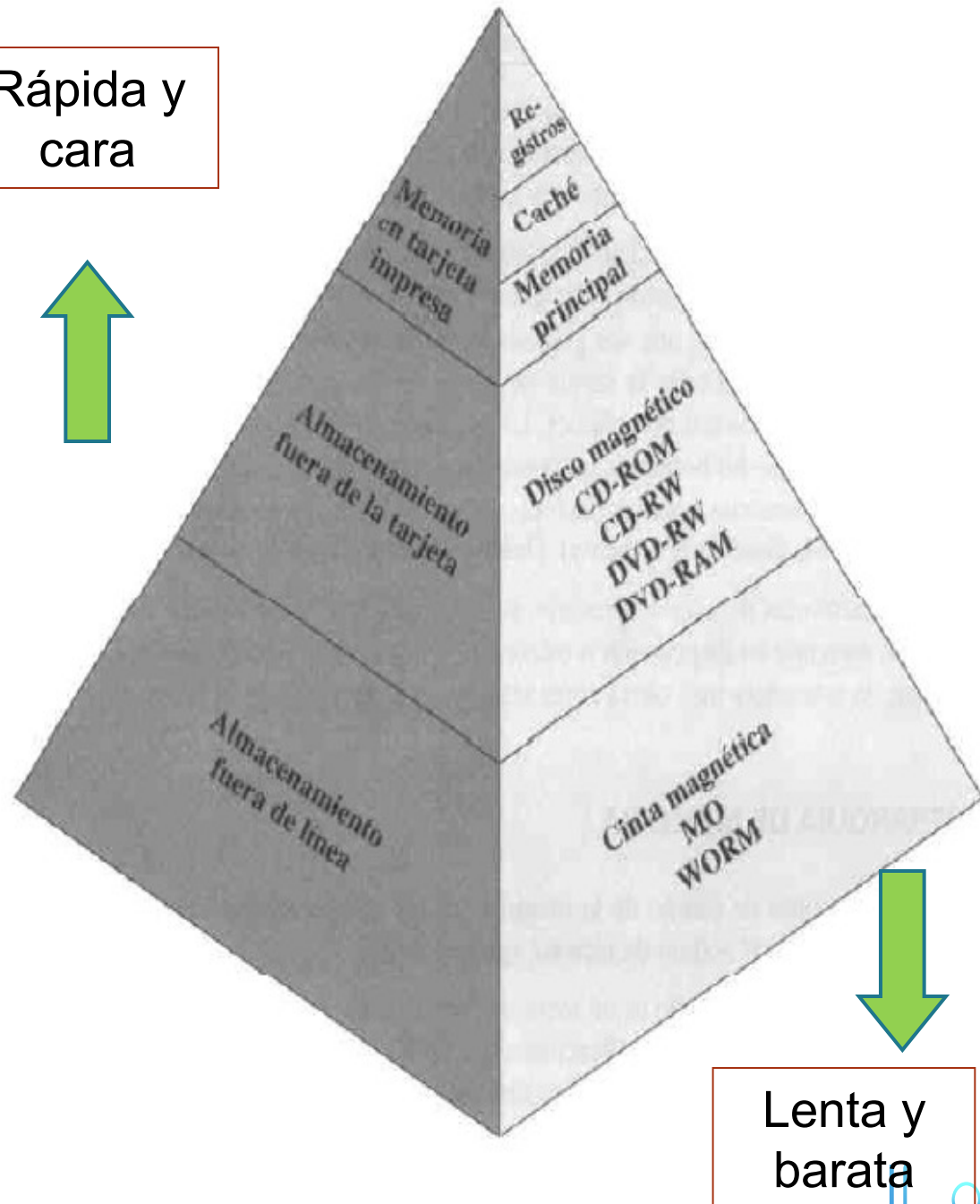
- La forma en que se organizan estos distintos tipos de memoria es lo que se conoce como jerarquía de memoria.
- En la cima de la jerarquía están los registros.
- En la base, las memorias secundarias (discos magnéticos) y de almacenamiento “off line” (CD, DVD, cintas).

Jerarquía de memorias

Rápida y cara

Al bajar en la jerarquía se observa:

- a) Disminuye el costo por bit
- b) Aumenta la capacidad
- c) Aumenta el tiempo de acceso
- d) Disminuye la frecuencia de acceso a la memoria por parte del procesador



Jerarquía de memorias



Tipos de memoria	Tiempos de acceso	Tamaño típico
Registros	1 ns	1KB
Caché	5-20 ns	1MB
Mem. Principal	60-80 ns	1GB
Discos	10 ms	160GB

Memoria - Características



► Duración de la información:

- Memorias volátiles: RAM
- Memorias no volátiles: discos, cintas
- Memorias permanentes: ROM, EPROM

► Modo de acceso:

- Acceso por palabra: memoria principal
- Acceso por bloque: discos, caché

Memoria - Características



► Velocidad

Memorias semiconductoras:

Tiempo de acceso: tiempo máximo que transcurre desde que se inicia la operación de lec/esc hasta obtener/almacenar el dato.

Tiempo de ciclo:

Tiempo mínimo que tiene que haber entre dos operaciones sucesivas sobre la memoria

$$t_{\text{ciclo}} > t_{\text{acceso}}$$

Memoria - Accesos



➤ Métodos de acceso

- **Acceso aleatorio (random):** el tiempo para acceder a una locación dada es independiente de la secuencia de accesos anteriores y es constante. Ejemplo la memoria principal.
- **Acceso secuencial:** el acceso debe hacerse en una secuencia lineal específica. Variable. Ejemplo son las unidades de cinta.

Memoria - Accesos



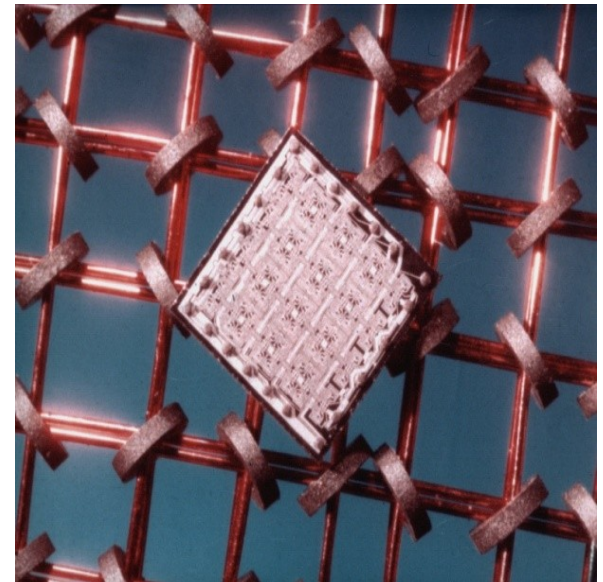
- **Acceso directo:** los bloques ó registros individuales tienen una dirección única que se basa en la localización física. Variable. Ejemplo los discos magnéticos.
- **Acceso asociativo:** es una memoria de acceso aleatorio que permite hacer una comparación de ciertas posiciones de bits dentro de una palabra buscando que coincidan con ciertos valores. Así una palabra se recupera por una porción de su contenido y no por su dirección. Las memoria caché pueden utilizar acceso asociativo.



➔ Organización de Memoria Principal

Memoria

En las primeras computadoras la forma más común de acceso aleatorio para la memoria principal consistía en una matriz de pequeños anillos ferromagnéticos llamados núcleos. La memoria principal recibía a menudo el nombre de núcleo (Core)



Organización interna de la memoria



Una celda de memoria es capaz de almacenar un bit de información. Por lo general, varias celdas se organizan en forma de arreglo.

En general la celda tiene 3 terminales funcionales capaces de llevar una señal eléctrica:

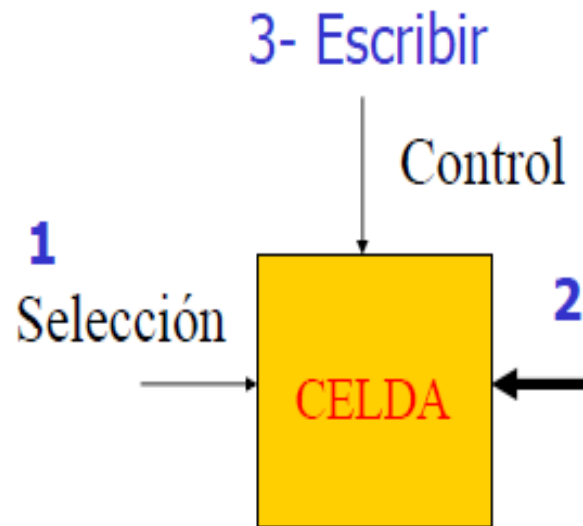
- ❖ **Selección:** selecciona una celda de memoria
- ❖ **Control:** especifica lectura ó escritura
- ❖ **Escritura/Lectura** de datos

Organización de la memoria

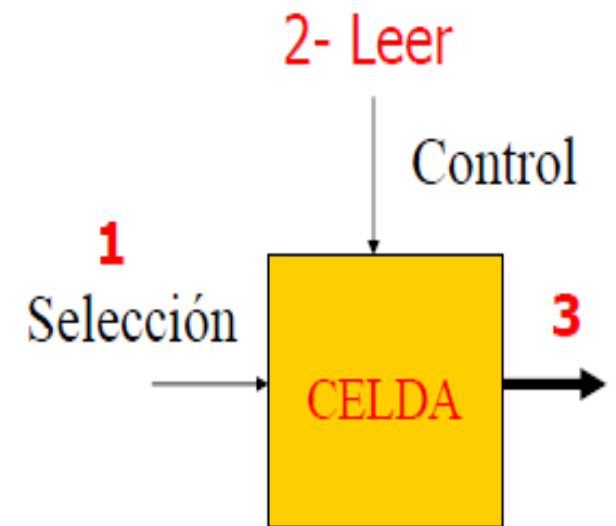


Elemento básico:
“celda” de
memoria
semiconductora

Tienen tres
terminales para
transmitir señales
eléctricas

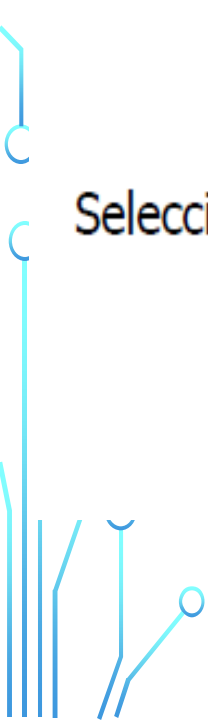
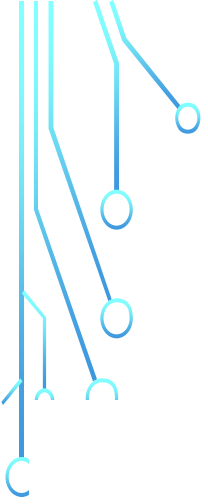


Escritura de Datos

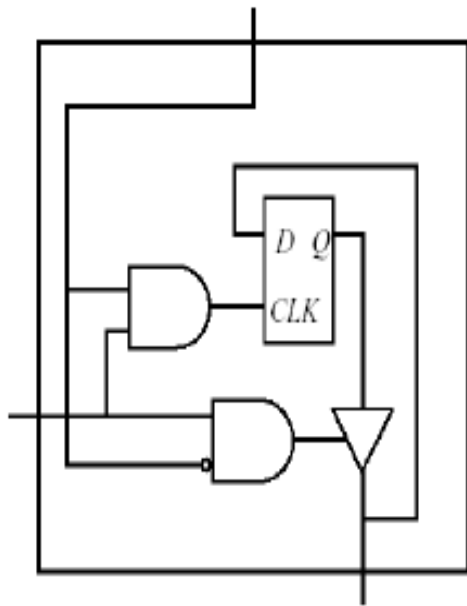


Lectura de Datos

Celda de memoria



Lectura/Escritura



Entrada y salida de datos

Selección

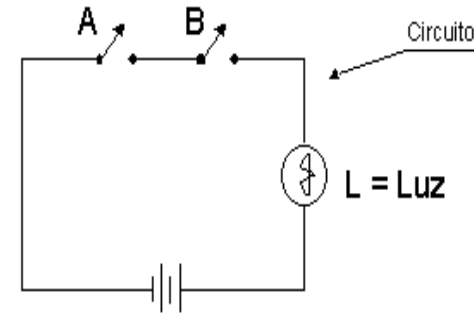
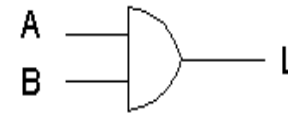


Tabla de Verdad

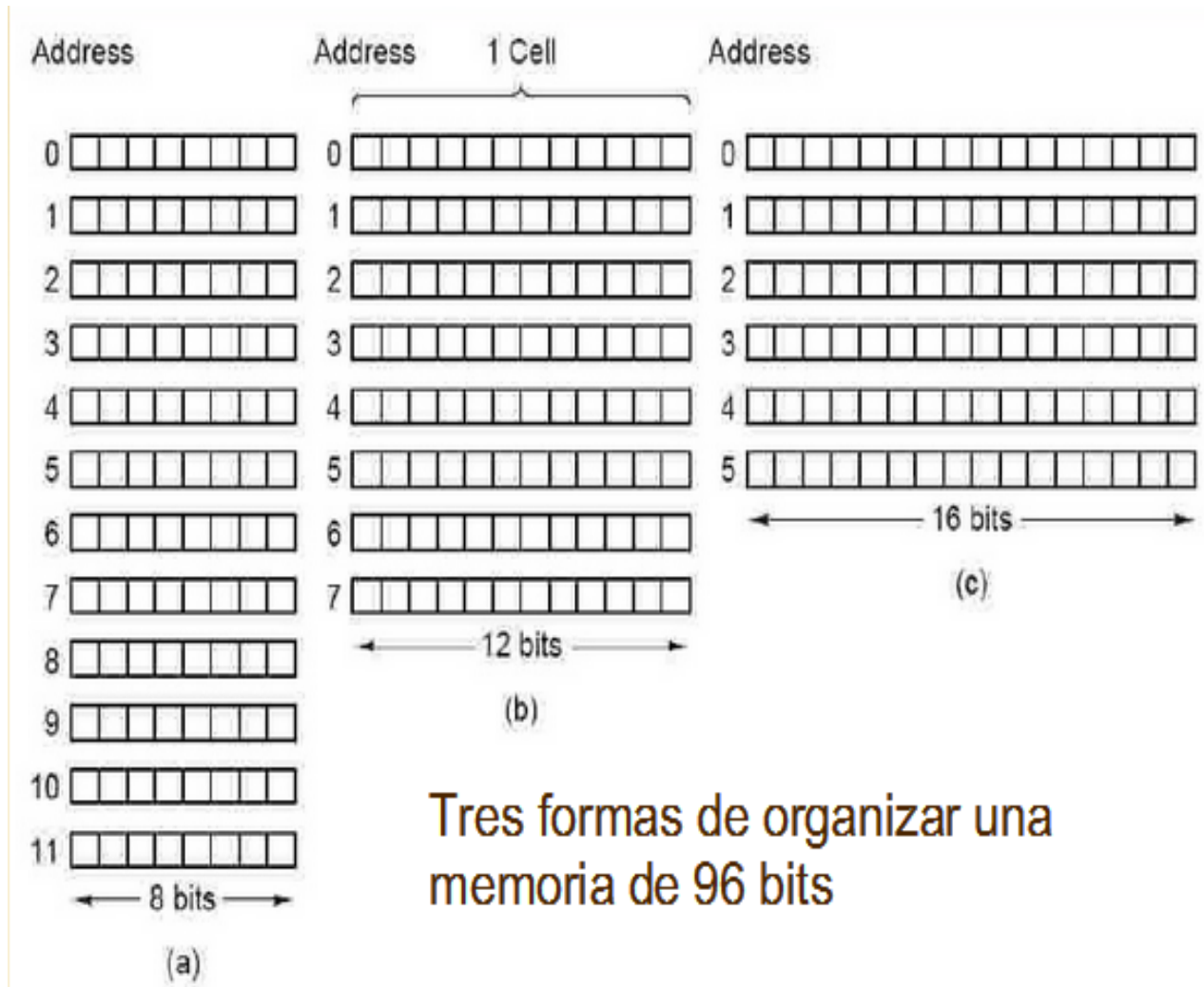
A	B	L
0	0	0
0	1	0
1	0	0
1	1	1



Compuerta Lógica



Organización de la memoria principal



Tres formas de organizar una memoria de 96 bits

Organización de la memoria



Cuando el tamaño de la memoria coincide con el número de bits por chip de memoria podemos ver que por ejemplo una memoria de 256K palabras de 8 bits se necesitan 18 bits ($2^{18} = 256K = 262144\text{bits}$) para direccionar desde un medio externo (ejemplo bus de direcciones)

Tamaño de la memoria



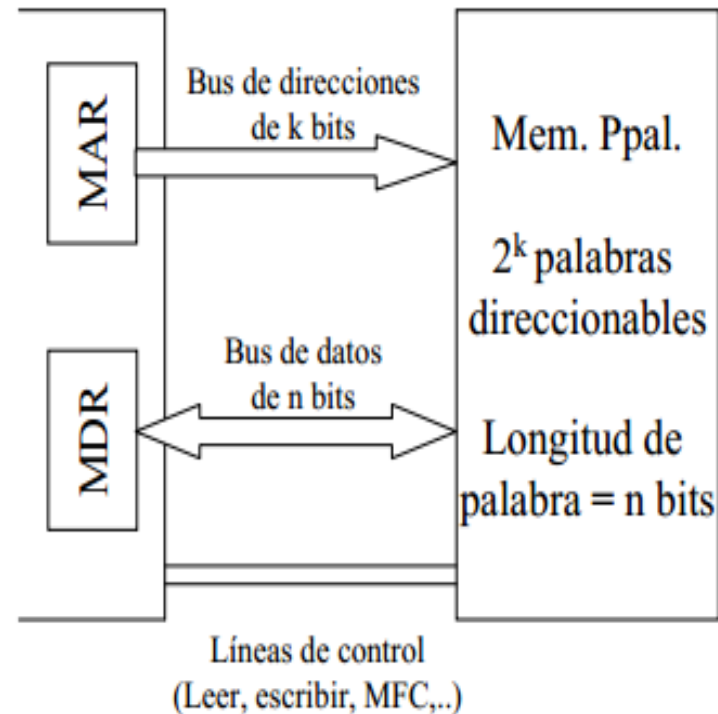
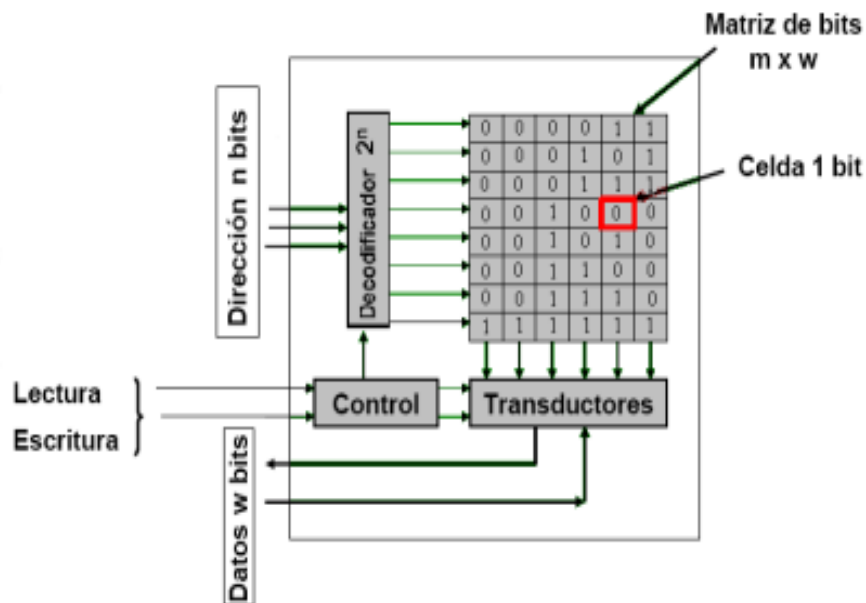
La capacidad de memoria viene dada por el bus de direcciones que establece el máximo número de posiciones direccionables por el computador.

Si se tienen n bits para el bus de direcciones, se podrá acceder hasta un máximo de 2^n posiciones.

Organización de la memoria



- La transferencia de datos entre la MP y la CPU se realiza gracias a dos registros:
 - MAR (Memory Address Register): delimita el tamaño de la memoria.
 - MDR (Memory Data Register): delimita el tamaño de los datos.



Organización de la memoria



- Organización por palabras:
 - La memoria se organiza en grupos de “k” bits llamados palabras
 - Para especificar una palabra se necesitan “n” bits
 - Cada palabra tiene asignada una dirección comprendida entre 0 y 2^n-1
- Organización por bytes:
 - A bytes individuales se le asignan direcciones independientes.
 - Una palabra compuesta por varios bytes puede ser accedida de manera individual a cada uno de ellos.
 - Para un mismo tamaño de memoria se necesita un mayor número de bits a la hora de poder direccionar.
 - Problemas:
 - Cómo relacionar las direcciones de las palabras con las direcciones de las unidades direccionables (bytes).
 - Cómo ordenar el contenido de una palabra en un conjunto consecutivo de unidades direccionables.

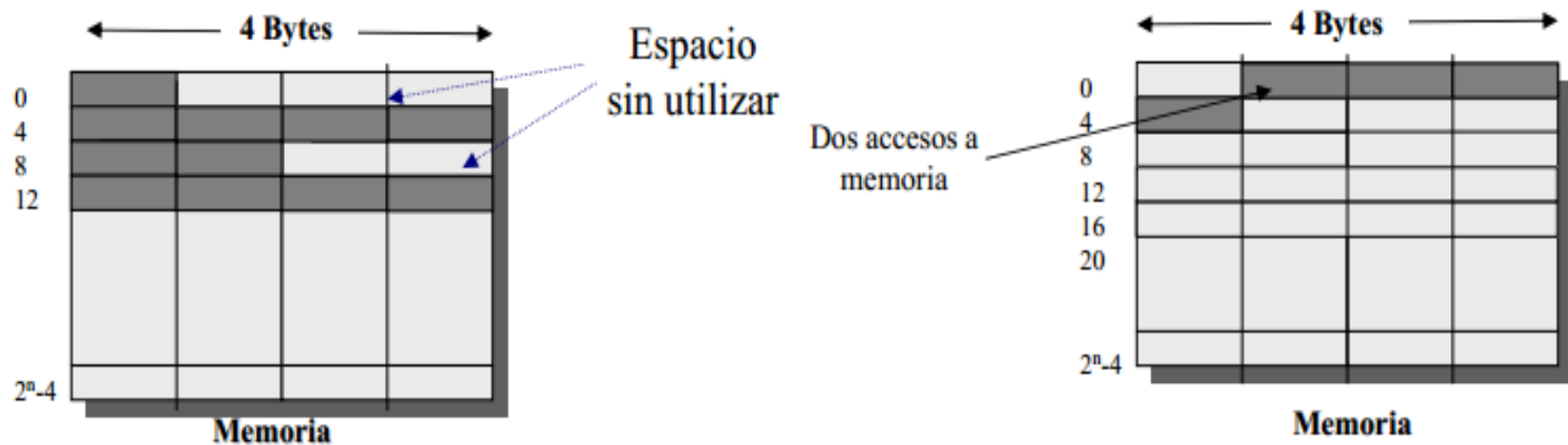
Organización de la memoria



- La memoria esta formada por varias celdas
- Cada celda tiene una dirección asociada
- Todas las celdas de una memoria tienen la misma cantidad de bits.
- Las celdas pueden ser de 1, 8, 12, 16...64 bits

Relación de direcciones: palabras-bytes

- Se obliga a que los datos empiecen en determinadas direcciones de memoria principal.
- Se dice que un dato de n bytes ubicado en la dirección D está alineado si: $D \bmod n = 0$.
 - Datos de 1 byte pueden ubicarse en cualquier posición de memoria.
 - Datos de 2 bytes pueden ubicarse en direcciones pares.
 - Datos de 4 bytes pueden ubicarse en direcciones múltiplos de 4.
- Ventaja: para acceder a una palabra donde los datos están alineados, únicamente se requiere un acceso a memoria.
- Desventaja: quedan espacios de memoria sin utilizar. Si se tiene un dato de 8 bits almacenado en la dirección 0 seguido de uno de 32 bits, se hace necesario dejar un espacio no útil de 24 bits.



Palabras alineadas

Palabras no alineadas



● Tipos de Memoria

RAM



RAM (Random Access Memory). Aleatorio significa que se puede acceder a cualquier celda de memoria en el mismo tiempo, independientemente de la posición en la estructura de la memoria.

Hay dos tipos:

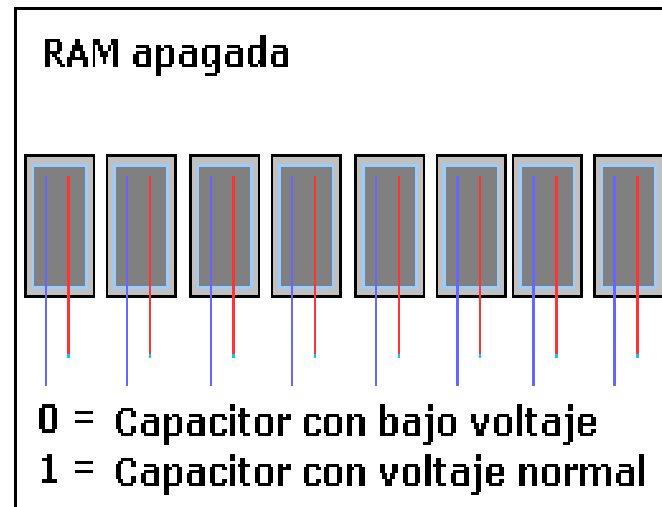
Basada en flip flops: memoria estática (SRAM). Se usa para memoria caché.

Basada en transistores: memoria dinámica (DRAM). Cargas almacenadas en transistores (capacitores) *El acceso a las ROM también es de éste tipo. Se usa para memoria principal*

ANIMACION DEL FUNCIONAMIENTO INTERNO DE LA MEMORIA RAM



- 1) La celda de memoria se carga de una corriente eléctrica alta cuándo indica el valor **1**.
- 2) La celda de memoria se carga de una corriente eléctrica baja cuándo indica el valor **0**.
- 3) Al apagar la computadora, las cargas desaparecen y por ello toda la información se pierde.
- 4) Este tipo de celdas tienen un fenómeno de recarga constante ya que tienden a descargarse, independientemente si la celda almacena un 0 ó un 1, esto se le llama "refrescar la memoria", solo sucede en memorias RAM

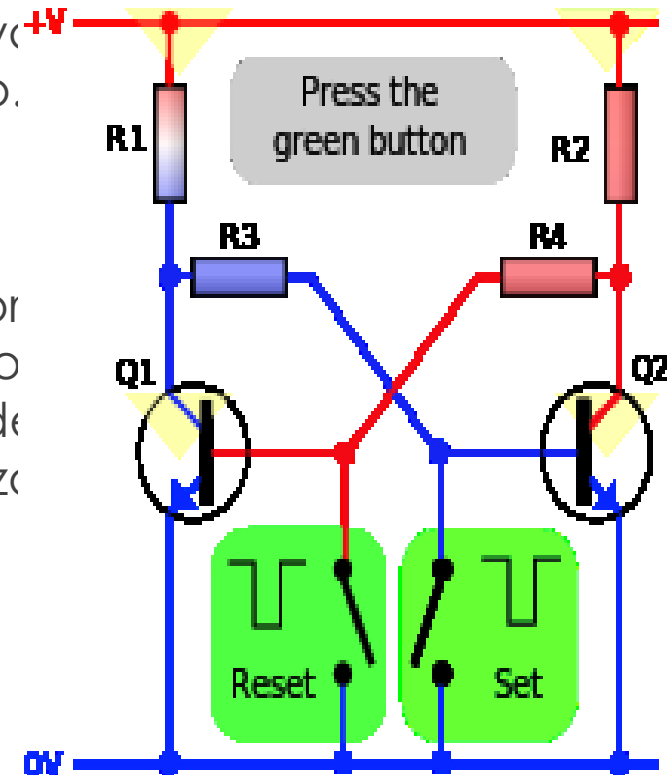


RAM estática o SRAM



El almacenamiento en RAM estática se basa en circuitos lógicos denominados **flip-flop**, que retienen la información almacenada en ellos mientras hay energía suficiente para hacer funcionar el dispositivo.

Un **Flip/flop** es un circuito oscilador (multivibrador) capaz de permanecer en uno de dos estados posibles durante un tiempo indefinido en ausencia de perturbaciones. El paso de un estado a otro se realiza variando sus entradas.



RAM estática o SRAM



La RAM estática no requiere ser actualizada y es normalmente mucho más rápida, mas cara y mas densa (celdas más pequeñas = más celdas por unidad de superficie) que la RAM dinámica (el tiempo de ciclo de la SRAM es de 8 a 16 veces más rápido)

ambién es más cara, por lo que se reserva generalmente para su uso en la memoria de acceso aleatorio(caché).

RAM estática o SRAM

Un chip de RAM estática puede almacenar tan sólo una cuarta parte de la información que puede almacenar un chip de RAM dinámica de la misma complejidad. ¿Por qué?



RAM dinámica o DRAM



DRAM almacenan la información en circuitos integrados que contienen condensadores, que pueden estar cargados o descargados . Almacena más información que SRAM en la misma superficie.

Los “capacitores” son más chicos que los flip flop.

❖ Como éstos pierden su carga en el transcurso del tiempo, se debe incluir los circuitos necesarios para "refrescar" los chips de RAM cada pocos milisegundos, para impedir la pérdida de su información

❖ Usada como memoria principal

RAM dinámica o DRAM



- Algunas memorias dinámicas tienen la lógica del refresco en la propia pastilla, dando así gran capacidad y facilidad de conexión a los circuitos. Estas pastillas se denominan casi estáticas.
- Mientras la RAM dinámica se refresca, el procesador no puede leerla. Si intenta hacerlo en ese momento, se verá forzado a esperar. Como son relativamente sencillas, las RAM dinámicas suelen utilizarse más que las RAM estáticas, a pesar de ser más lentas.

TIPOS DE MODULOS DE RAM



Los módulos que se utilizan en la actualidad



DDR2 RAM: tienen 240 pines. Los zócalos no son compatibles con DDR RAM. La muesca está situada dos milímetros hacia la izquierda con respecto a DDR RAM. Se comercializan pares de módulos de 2 Gb (2x2= 4 GB). Pueden trabajar a velocidades entre 400 y 800 MHz.

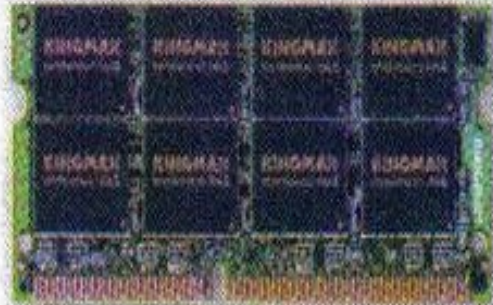


DDR RAM: sucesora de la memoria SDRAM, tiene un diseño similar pero con una sólo muesca y 184 contactos. Ofrece una velocidad entre 200 y 600 MHz. Se caracteriza por usar un mismo ciclo de rejoj para hacer dos intercambios de datos a la vez.



Rambus: puede ofrecer velocidades entre 600 y 1.066 MHz. Tiene 184 contactos. Algunos de estos módulos disponen de una cubierta de aluminio (dispersador de calor) que protege los chips de memoria de un posible sobrecalentamiento. Debido a su alto coste, su utilización no se ha extendido mucho.

TIPOS DE MODULOS DE RAM



SODIMM: el tamaño de estos módulos es más reducido que el de los anteriores ya que se emplean sobre todo en ordenadores portátiles. Se comercializan módulos con capacidades de 512 MB y 1 GB. Los hay de 100, 144 y 200 contactos.



MEMORIAS RIMM acrónimo de Rambus Inline Memory Module, designa a los módulos de memoria RAM que utilizan una tecnología denominada RDRAM, desarrollada por Rambus Inc. A pesar de tener la tecnología RDRAM niveles de rendimiento muy superiores a la tecnología SDRAM y las primeras generaciones de DDR RAM, debido al alto costo de esta tecnología no han tenido gran aceptación en el mercado de PCs. Su momento álgido tuvo lugar durante el periodo de introducción del Pentium 4 para el cual se diseñaron las primeras placas base, pero Intel ante la necesidad de lanzar equipos más económicos decidió lanzar placas base con soporte para SDRAM y más adelante para DDR RAM desplazando esta última tecnología a los módulos RIMM del mercado.

DRAM vs. SRAM



- RAM dinámica (DRAM)
 - Consumo mínimo.
 - Capacidad de almacenamiento comparativamente alta.
 - Costo por bit bajo.
 - Tiempo de acceso alto (lento), debido al circuito de regeneración de carga.
 - Si construimos el banco de memoria utilizando RAM dinámica, no aprovechamos la velocidad del procesador.
- RAM estática (SRAM)
 - Alto consumo relativo.
 - Capacidad de almacenamiento comparativamente baja.
 - Costo por bit alto.
 - Tiempo de acceso bajo (es mas rápida).
 - Si construimos el banco de memoria utilizando RAM estática, el costo y el consumo de la computadora son altos.

Las nuevas tecnologías en RAM



Enhanced DRAM (ESDRAM)

Para superar algunos de los problemas de latencia inherentes con los módulos de memoria DRAM standard, se incluye una cantidad pequeña de SRAM directamente en el chip, eficazmente creando un caché en el chip. Permite tiempos de latencia más bajos y funcionamientos de 200 mhz. La SDRAM oficia como un caché dentro de la memoria. 7. Una de las desventajas de estas memorias es que su valor es 4 veces mayor al de la memoria DRAM.

ROM



ROM (Memoria de solo lectura). Se utiliza para la microprogramación así como para almacenar subrutinas de uso frecuente, programas del sistema y tablas de funciones.

La ventaja es que estos datos están en forma permanente en la memoria principal y no es necesario traerlos desde otro dispositivo.

Deventaja: la etapa de inserción de datos tiene unos costos fijos relativamente grandes, tanto sea una o miles de ROM. Además no se permite un fallo. Si uno de los bits es erróneo debe desecharse la tirada completa de chips de memoria ROM

PROM



PROM (Programable Read Only Memory)

Es una memoria digital donde el valor de cada bit depende del estado de un que puede ser quemado una sola vez.

Por esto la memoria puede ser programada (pueden ser escritos los datos) una sola vez a través de un dispositivo especial, un programador PROM.

Estas memorias son utilizadas para grabar datos permanentes en cantidades menores a las ROMs, o cuando los datos deben cambiar en muchos o todos los casos.



PROM D23128C en la placa base de una Sinclair ZX Spectrum

EPROM



EPROM (**E**rasable **P**rogramable **R**ead **O**nly **M**emory) Está formada por celdas de FAMOS (Floating Gate Avalanche-Injection Metal-Oxide Semiconductor) o "transistores de puerta flotante", cada uno de los cuales viene de fábrica sin carga, por lo que son leídos como 1 (por eso, una EPROM sin grabar se lee como FF en todas sus celdas).

Se lee y se escribe como la PROM pero antes de escribir se borra todas las celdas de almacenamiento mediante luz ultravioleta.

Este proceso puede hacerse muchas veces y puede durar hasta 20 minutos.

Es más costosa que la PROM

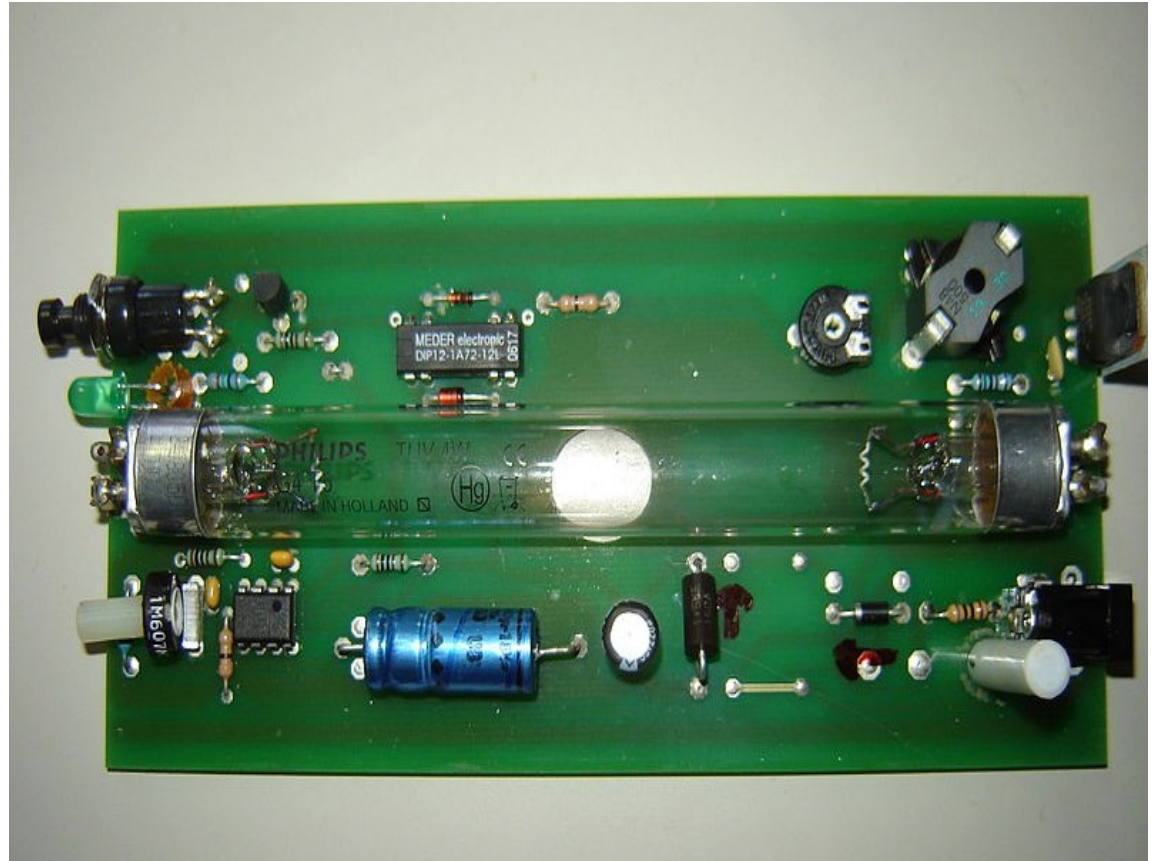


Una EPROM de 32KB (256Kbit).

Lámpara borradora de EPROM



Ventana
de la
EPROM
para
borrar



EEPROM



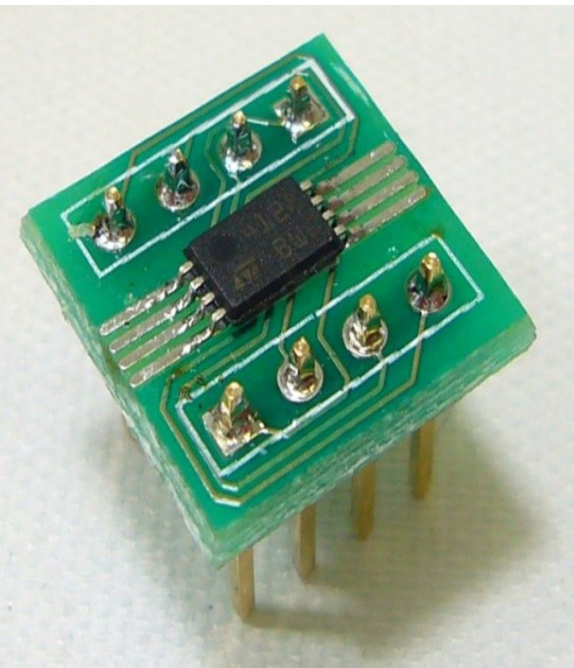
EEPROM (**E**lectrically **E**rasable **P**rogramable **R**ead **O**nly **M**emory) (ROM programable y borrada eléctricamente). Es un tipo de memoria ROM que puede ser programada, borrada y reprogramada eléctricamente, a diferencia de la EPROM que ha de borrarse mediante un aparato que emite rayos ultravioleta.

Son usadas para almacenar información programable de usuario, como por ejemplo:

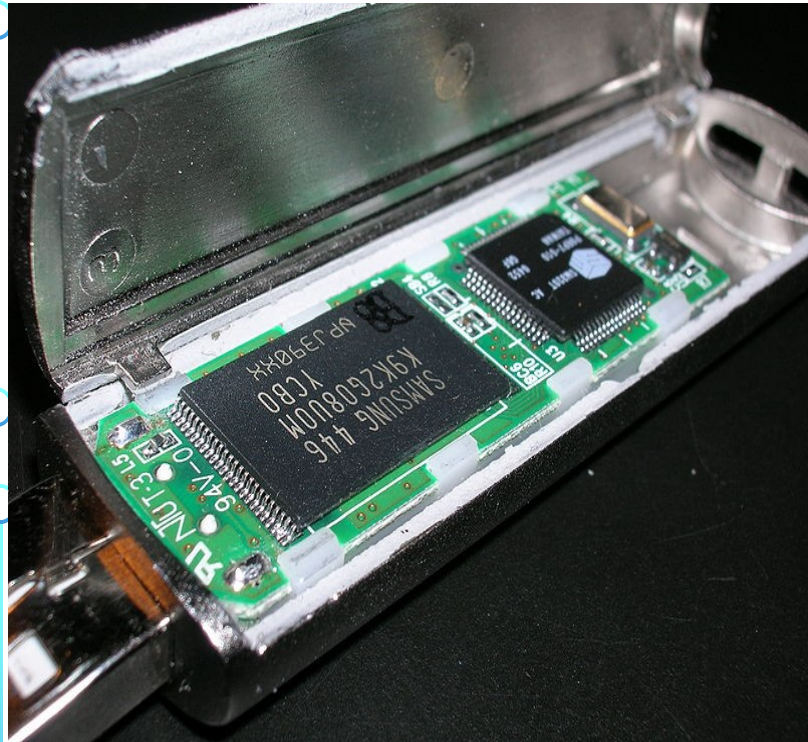
- Información de programación VCR
- Información de programación de CD
- Información de usuario de productos instalados en el equipo

Es menos densa que la EPROM y más costosa.

Se puede escribir en cualquier momento sin borrar su contenido anterior. Solo se actualiza el byte/es direccionados.



Memoria Flash



- Se denominan **Flash** por la velocidad a la que pueden reprogramarse. En cuanto a coste y funcionalidad están entre las EPROM y las EEPROM.
- Usan borrado eléctrico.
- Se puede borrar toda rápidamente o solo bloques.
- El microchip está organizado de manera que cada una de sus secciones de celdas se borra de un solo golpe o flash.
- No permiten borrar a nivel de byte.
- Usan un solo transistor por bit y son de densidad similar a las EPROM.

Una **memoria USB** . El chip de la izquierda es la memoria flash. El **controlador** está a la derecha.

Bibliografía recomendada



- Stallings Willam (2005). “Organización y arquitectura de Computadores”. **Capítulo 5: Memoria Interna**. 7ma Ed. Pearson Prentice Hall. Madrid. España.

- Link sugerido:

<http://www.pctechguide.com/14Memory.htm>



- **PRINCIPIOS BASICOS DE MEMORIA CACHE**

Memoria Cache orígenes

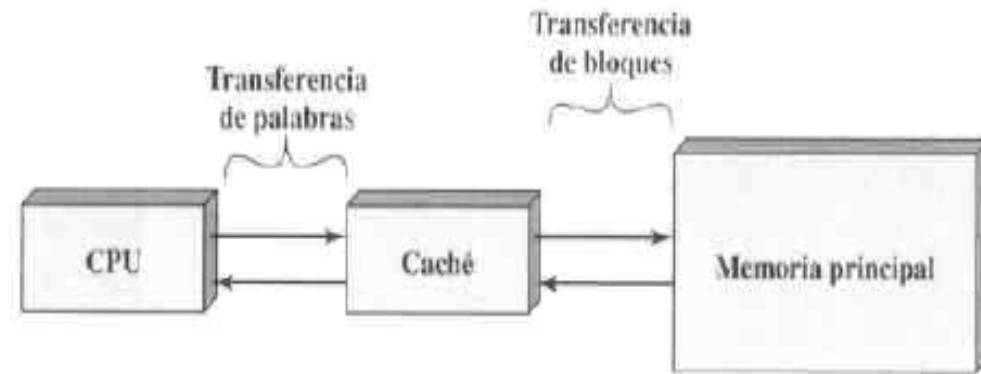


- Históricamente las CPU han sido más rápidas que las memorias.
- El aumento de circuitos que es posible incluir en un chip
 - ❖ Diseñadores de CPU lo usaron para hacerla más veloz (ej. pipeline).
 - ❖ Los diseñadores de memoria lo usaron para aumentar la capacidad del chip (más memoria, más grandes decodificadores).

Memoria Caché - Objetivo



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



Memorias caché y principal –
Tanenbaum pag 112 – 7ma edición

Cache principio de funcionamiento



La CPU "emite" una solicitud de lectura a la memoria (bus de direcciones, bus de control) pasan muchos ciclos de reloj antes que reciba la palabra que necesita, por el bus de datos.

Cache principio de funcionamiento



- En todos los ciclos de instrucción, la CPU accede a memoria al menos una vez, para buscar la instrucción y muchas veces accede a buscar operandos.
- La velocidad a la cual la CPU ejecuta instrucciones está limitada por el tiempo del ciclo de memoria.

¿Porque Memoria Cache?



- El problema no es tecnológico sino económico.

Se pueden construir memorias tan rápidas como la CPU, pero para obtener la máxima velocidad tiene que estar dentro del chip de la CPU llegar a la memoria por el bus del sistema es "lento".

¿Porque Memoria Cache?



➤ Solución

Técnicas para combinar una cantidad pequeña de memoria rápida con una cantidad grande de memoria lenta, para obtener la velocidad de memoria "casi" rápida.

Cache Concepto



La idea general es que cuando se hace referencia a una palabra, ella y alguna de las vecinas se traen de la memoria grande y lenta a la caché, para que en el siguiente acceso la palabra buscada se encuentre en el caché.

Cache: aciertos y fallos



La efectividad de la caché se expresa a través de la frecuencia de aciertos: es decir el número de veces que la caché acierta direcciones.

Un acierto de caché sucede cuando los datos que necesita el procesador están almacenados en la caché entonces la CPU obtiene los datos a alta velocidad.

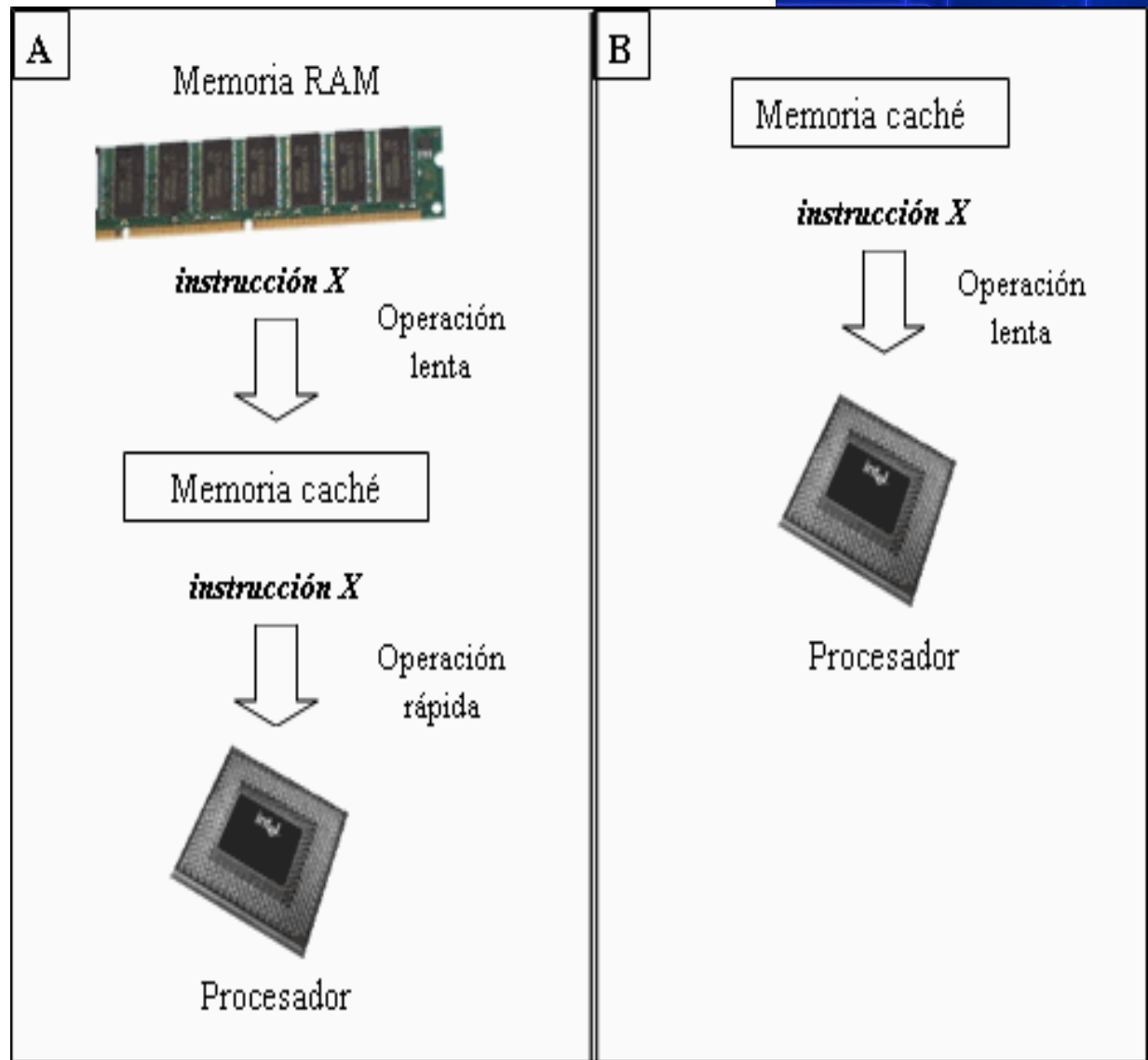
Cache: aciertos y fallos

Un fallo de caché ocurre cuando los datos buscados no se encuentran en la caché.

En este caso la CPU tiene que obtenerlos de la memoria principal, a una velocidad menor.



Funcionamiento de la Memoria Caché



Si el procesador desea buscar la instrucción *X* (situada en cierta posición de la RAM), intentará primero hallar dicha instrucción en la caché. Si la encuentra allí, el acceso será rápido (caso B). Si no la encuentra, se añade un retraso extra al buscar la instrucción en la RAM y copiarla en la caché (caso A).

Principios de la memoria Cache



El uso de la memoria caché se sustenta en dos principios ó

propiedades que exhiben

los programas:

1. Principio de localidad espacial de referencia

➤ cuando se accede a una palabra de memoria, es "muy probable" que el próximo acceso sea en la vecindad de la palabra anterior.

➤ 2. Principio de localidad temporal de referencia cuando se accede a una posición de memoria, es "muy probable" que un lapso de "tiempo

corto", dicha posición de memoria sea accedida nuevamente.

Cache-Localidad espacial



Se sustenta en:

- Ejecución secuencial del código
- Tendencia de los programadores a hacer próximas entre sí variables relacionadas
- Acceso a estructuras tipo matriz ó pila

Cache- Localidad temporal



Se sustenta en:
Formación de ciclos o bucles en el código

Subrutinas (Procedimientos o Funciones)

Pilas

Ejemplo de localidades

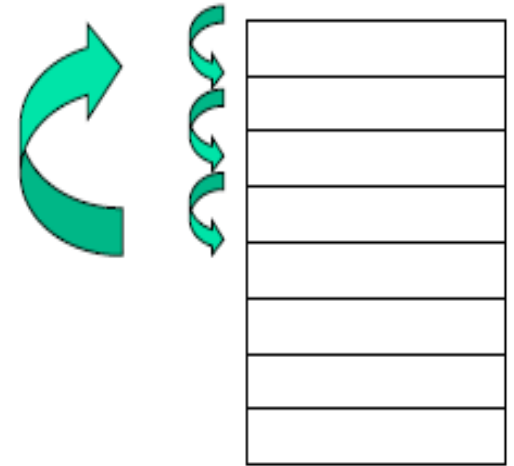


Estas 2 sentencias exhiben los dos principios antes mencionados:

```
for i=1 to i=10, do  
  A[i]:=0;
```

En cada ciclo se consulta cuanto vale i .

Cada asignación $A[i]:=0$ almacena un 0 en un elemento del arreglo (el siguiente).



Cache multinivel



Con el aumento de densidad de integración ha sido posible tener una caché en el mismo chip del procesador (on-chip). Comparada con la que es accedida por un bus externo, reduce los tiempos, dado que elimina los accesos al bus. (caché L1)

¿Conviene tener además una caché externa (off-chip)?

En la mayoría de los casos si, y se la denomina L2, la idea es retardar el acceso a la DRAM o Rom cuando no se encuentra un valor en la cache L1

Cache multinivel



Con el avance de la tecnología pudo llevarse también la caché L2 al procesador, con lo cual aparece un tercer nivel denominado L3, que parece mejorar los tiempos de acceso.

En algunos casos se realiza una diferenciación entre cache de datos y de instrucciones en L1

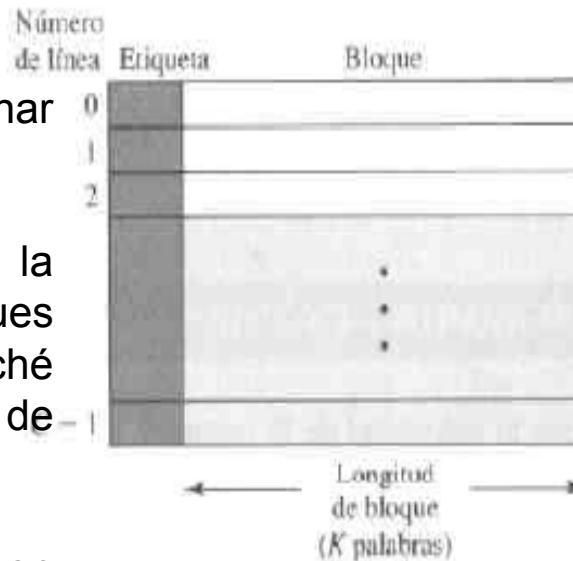
Memoria caché - Tamaño

Por ejemplo la **cache** puede almacenar **64KB**.

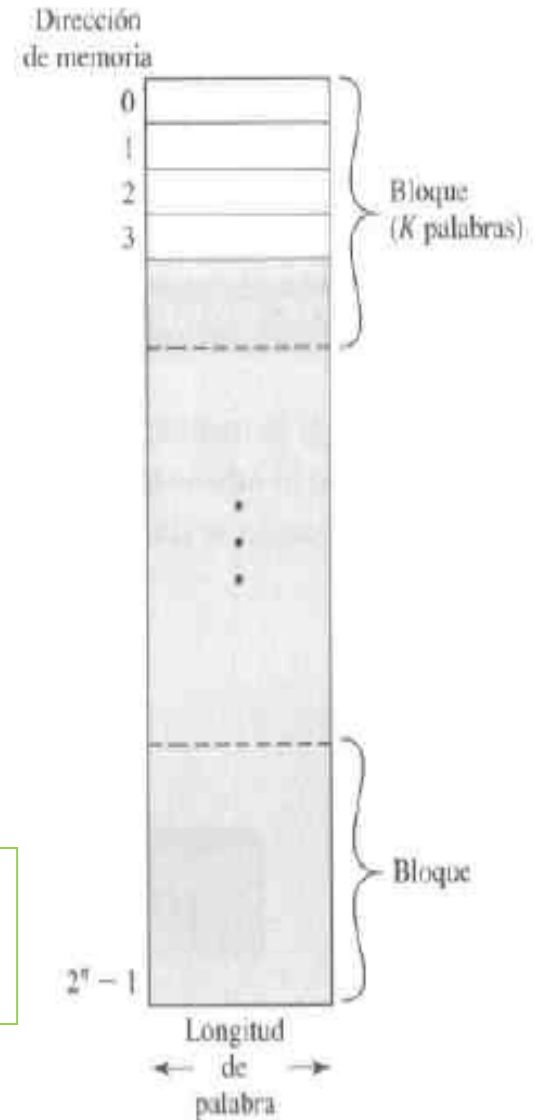
Los datos se transfieren entre la memoria principal y la cache en bloques de 4bytes. Esto significa que la caché está organizada en **16K** = 2^{14} líneas de **4 bytes** cada una.

La **memoria principal** es de **16MB** con cada byte directamente direccionable mediante una dirección de 24 bits (2^{24} = 16MB).

Así pues el hecho de realizar la correspondencia se puede considerar que la **memoria principal** consta de **4MB** bloques de **4 bytes** cada uno.



(a) Caché



(b) Memoria principal

Procesador	Tipo	Año de introducción	Caché L1 ^a	Caché L2	Caché L3
IBM 360/55	Gran computador	1968	16 a 32 KB	—	—
PDP-11/70	Minicomputador	1975	1 KB	—	—
VAX 11/780	Minicomputador	1978	16 KB	—	—
IBM 3033	Gran computador	1978	64 KB	—	—
IBM 3090	Gran computador	1985	128 a 26 KB	—	—
Interl 80486	PC	1989	8 KB	—	—
Pentium	PC	1993	8 KB/8 KB	256 a 512 KB	—
PowerPC 601	PC	1993	32 KB	—	—
PowerPC 620	PC	1996	32 KB/32 KB	—	—
PowerPC 64	PC/servidor	1999	32 KB/32 KB	256 KM 1 MB	2 MB
IBM S/390 G4	Gran computador	1997	32 KB	256 KB	2 MB
IBM S/390 G6	Gran computador	1999	256 KB	8 MB	—
Pentium 4	PC/servidor	2000	8 KB/8 KB	256 KB	—
IBM SP	Servidor de gama alta/Supercomputador	2000	64 KB/32 KB	8 MB	—
CRAY MTA ^b	Supercomputador	2000	8 KB	2 MB	—
Itanium	PC/Servidor	2001	16 KB/16 KB	96 KB	4 MB
SGI Origin 2001	Servidor de gama alta	2001	32 KB/32 KB	4 MB	—
Itanium 2	PC/Servidor	2002	32 KB	256 KB	6 MB
IBM POWER5	Servidor de gama alta	2003	64 KB	1,9 MB	36 MB
CRAY XD-1	Supercomputador	2004	64 KB/64 KB	1 MB	—

Tamaños de las cache en algunos procesadores – Stallings pag 116

Referencias de la tabla:

- a) Caché de instrucciones y caché de datos
- b) Solo caché de instrucciones

Evolución de la caché en los Intel – Stallings pag 130

Problema	Solución	Primer procesador en incluirla
Memoria externa más lenta que el bus del sistema.	Añadir caché externa usando tecnología de memoria más rápida.	386
El aumento de velocidad de los procesadores hace que el bus se convierta en un cuello de botella para el acceso a caché.	Trasladar la caché externa al mismo chip (caché <i>on-chip</i>), funcionando a la misma velocidad que el procesador.	486
La caché es pequeña debido a la disponibilidad de superficie <i>on-chip</i> limitada.	Añadir una caché L2 externa con tecnología más rápida que la empleada en la memoria principal.	486
Se produce contención cuando las unidades de precaptación de instrucciones y de ejecución necesitan acceder simultáneamente a la caché. En este caso, la unidad de precaptación se bloquea mientras la unidad de ejecución accede a los datos.	Crear cachés separadas de datos y de instrucciones.	Pentium
El incremento de velocidad del procesador hace que el bus externo sea un cuello de botella para el acceso a la caché L2.	Crear un bus externo específico o «puerta trasera» (BSB, <i>back-side bus</i>) que funcione a más velocidad que el bus principal. El BSB se dedica a la caché L2.	Pentium Pro
	Llevar la caché L2 al chip del procesador.	Pentium II
Algunas aplicaciones que trabajan con grandes bases de datos deben tener acceso rápido a cantidades ingentes de datos. Las cachés <i>on-chip</i> son demasiado pequeñas.	Añadir una caché L3 externa.	Pentium III
	Integrar la caché L3 <i>on-chip</i> .	Pentium 4

Bibliografía recomendada



- Stallings Willam (2005). “Organización y arquitectura de Computadores”. **Capítulo 4: Memoria Caché**. 7ma Ed. Pearson Prentice Hall. Madrid. España.

- Link sugerido:

<http://rabiapc.blogspot.com.ar/2011/01/memoria-cache-que-es-la-memoria.html>

