Memoria Interna

Introducción

Las memorias presentan la más amplia diversidad de tipos, tecnología, estructura, prestaciones y costo, de entre todos los componentes de la computadora. No existe una tecnología óptima para satisfacer las necesidades, por lo que una computadora está equipada con una *jerarquía de subsistemas de memoria*, que puede ser interna (accesible directamente por el procesador) o externa (accesible por el procesador mediante módulos de entrada/salida). Se revisarán las características clave de las memoria, los subsistemas de la memoria principal de semiconductores (ROM, DRAM y SRAM) y la memoria cache.

Conceptos básicos

Clasificación por ubicación

- *CPU* Constituida por los registros del CPU o la memoria para la unidad de control del CPU.
- *Memoria interna* se identifica como la memoria principal.
- *La memoria externa* consta de dispositivos de almacenamiento periféricos, como discos y cintas, que son usados por el CPU a través de controladores de E/S, conocida también com *secundaria*.

Clasificación por Capacidad

- *Tamaño de la palabra.*
- *Número de palabras.*

Para memorias internas la capacidad se expresa en bytes (1 byte=8 bits) o en *palabras*. Las longitudes de palabra son de 8, 16 y 32 bits. Para las memorias externas se expresa en bytes.

Clasificación por Unidad de transferencia (Palabra o Bloque)

Para memorias internas, la unidad de transferencia es igual al número de líneas de entrada/salida de datos del módulo de memoria.

Conceptos:

- *Palabra.* Es la unidad "natural" de organización de la memoria. El tamaño de la palabra suele coincidir con el número de bits utilizados para representar números y con la longitud de las instrucciones.

- *Unidad direccionable.* Para muchos sistemas la unidad direccionable es la palabra, aunque pueden direccionar a nivel de bytes.

- *Unidad de transferencia.* Para la memoria principal es el número de bits que se leen o escriben en memoria a la vez. Para la memoria externa, los datos se transfieren en unidades más grandes que la palabra denominada bloques.

Método de acceso

- *Acceso secuencial.* La memoria se organiza en unidades de datos llamadas registros. El acceso se realiza en secuencia. El tiempo de acceso a un registro dado es variable. Las unidades de cinta son de acceso secuencial

- *Acceso directo.* Los registro o bloques individuales tienen una dirección única basada en su dirección física. El acceso se lleva a cabo mediante un acceso directo a una vecindad dada, seguido de una búsqueda secuencial, bien contando, o esperando hasta alcanzar la posición final. El tiempo de acceso es variable. Las unidades de disco son de éste tipo.

- *Acceso aleatorio (random)* Cada posición direccionable de memoria tiene un único mecanismo de acceso cableado físicamente. El tiempo de acceso es constante e independiente de la secuencia de accesos previos. La memoria principal es de acceso aleatorio.

- *Asociativa.* Es del tipo de acceso aleatorio y que permite hacer una comparación de ciertas posiciones de bits dentro de una palabra para buscar que coincida con unos valores dados, haciéndolo para todas las palabras simultáneamente, por lo que una palabra se recupera empleando una porción de su contenido en vez de su dirección. El tiempo de recuperación de un dato es constante, ya que cada posición tiene su propio mecanismo de direccionamiento. Las memorias cache pueden emplear este tipo de acceso.

Por prestaciones

- *Tiempo de acceso.* Para memorias de acceso aleatorio es el tiempo que transcurre desde el instante en el que se presenta una dirección a la memoria hasta que el dato, o ha sido memorizado o está disponible para su uso. Para memorias de otro tipo, es el tiempo que se tarda en situar el mecanismo de lectura/escritura en la posición deseada.
- *Tiempo de ciclo de memoria.* Aplica para memorias de acceso aleatorio y es el tiempo de acceso y algún tiempo adicional requerido para iniciar un segundo acceso a memoria.

- *Velocidad de transferencia.* Es la velocidad a la que se pueden transferir datos a, o desde, una unidad de memoria. En memorias de acceso aleatorio coincide con el inverso del tiempo de ciclo. En otras memorias se tiene:

$$T_N = T_A + \frac{N}{R}$$

donde

$$T_N = \text{tiempo medio de escritura o de lectura de } N \text{ bits}$$

$$T_A = \text{tiempo de acceso medio}$$

$$N = \text{número de bits}$$

$$R = \text{Velocidad de transferencia, en bits por segundo (bps)}$$

Otras Características clave

Dispositivo físico

- *Semiconductor.* Emplean tecnologías LSI o VLSI.
- *Soporte Magnético.* Usadas en discos y cintas.

Dispositivo físico

- *Volátil/No Volátil.* En memorias volátiles la información se va perdiendo o desapareciendo cuando se desconecta la alimentación. En las memorias NO volátiles la información permanece sin deteriorarse una vez grabada, por ejemplo en la memorias de superficie magnética u óptica.
- *Borrable/No Borrable.* Las memorias no-borrables no pueden modificarse. Las memorias semiconductras de este tipo se conocen como ROM.

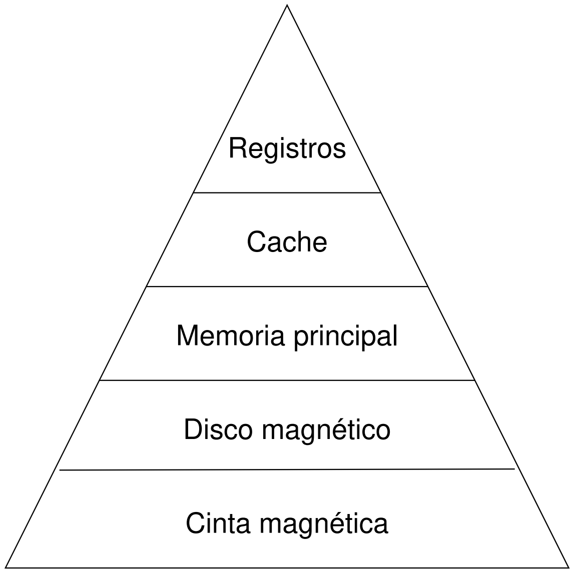
Restricciones de diseño en la memoria de una computadora

- *¿Cuánta?* Para una cierta capacidad, probablemente se desarrollarán aplicaciones que la utilicen.
- *¿Qué velocidad?* Para conseguir las prestaciones óptimas, la memoria debe seguir al CPU.
- *¿De qué costo?* Para un sistema práctico, el costo debe ser razonable con relación a los otros componentes.

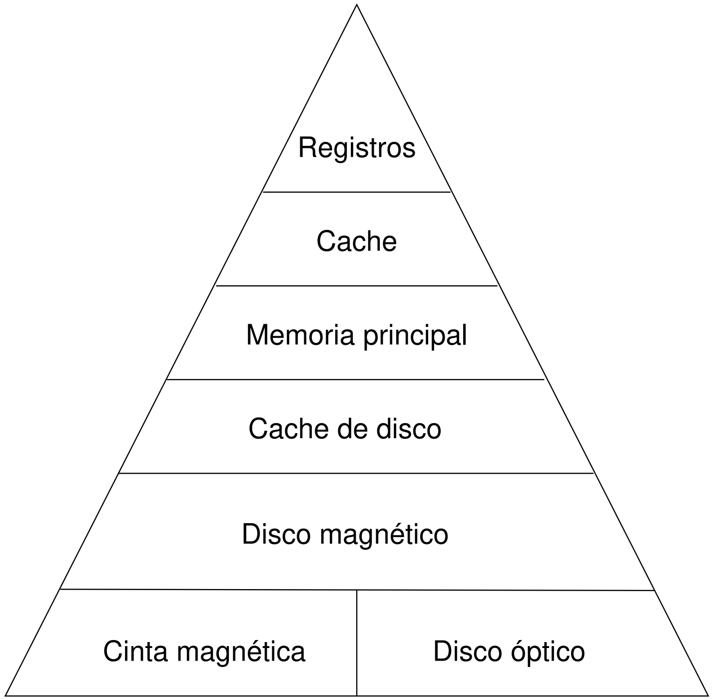
Se emplean diversas tecnologías para realizar los sistemas de memoria, en donde se cumplen las siguientes relaciones:

- A menor tiempo de acceso, mayor costo por bit.
- A mayor capacidad, menor costo por bit.
- A mayor capacidad, mayor tiempo de acceso.

Jerarquía de memoria



Jerarquía de memoria tradicional



Jerarquía de memoria moderna

Jerarquía de memoria

Para satisfacer las prestaciones requeridas y tener una gran capacidad, se emplea una jerarquía de memoria, en donde no se usa un solo componente de memoria, sino varios. Cuando se desciende en la jerarquía se tiene lo siguiente:

1. Disminución del costo por bit.
2. Incremento de la capacidad.
3. Incremento del tiempo de acceso.
4. *Disminución de la frecuencia de accesos a la memoria por parte del CPU.*

Memoria Principal Semiconductora

Tipo de memoria	*Clase*	*Borrado*	*Mecanismos de escritura*	*Volatilidad*
Memoria de acceso aleatorio (RAM)	Lectura/Escritura	Eléctricamente por bytes	Eléctricamente	Volátil
Memoria de sólo lectura (ROM)	Memoria de	No	Mediante máscaras	No volátil
ROM programable (PROM)	sólo lectura	posible	Eléctricamente	
PROM borrable (EPROM)	Memoria de	Luz ultravioleta, chip completo		
Memoria FLASH	sobre todo	Eléctricamente por bloques	Eléctricamente	No volátil
PROM borrable eléctricamente (EEPROM)	lectura	Eléctricamente por bytes		

Variantes de memoria RAM

- *Estáticas* Se forman con biestables ("flip-flop"). Son generalmente algo más rápidas que las dinámicas.
- *Dinámicas* Están hecha con celdas que almacenan los datos como cargas en condensadores, requieren refrescos periódicos para mantener memorizadas los datos, por lo que requieren de circuitería para el refresco. Una celda es más simple que una estática y por tanto más pequeña. Tienden a ser las preferidas para memorias grandes.

Aplicaciones de la memoria ROM

- Microprogramación (usada en la unidad de control del CPU).
- Subrutinas de biblioteca para funciones de uso frecuente.
- Programas del sistema.
- Tablas de funciones.

Los problemas de construir una ROM son:

- La étapa de inserción de datos implica costos fijos relativamente grandes, independiente del número de ROMs que se vayan a fabricar.
- No se toleran fallas.

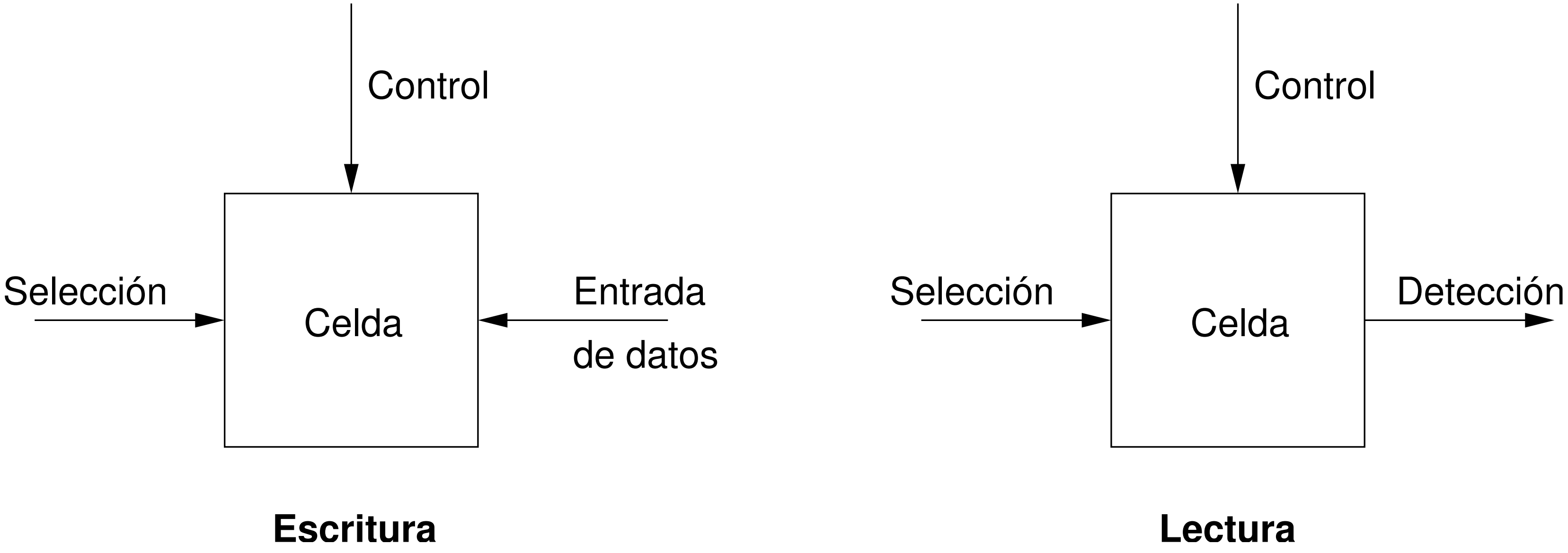
Lo anterior, debido al cableado de datos en el circuito integrado.

Organización

Propiedades de la memoria semiconductora:

- Presentan dos estados estables, usados para representar el 1 y 0 binarios.
- Puede escribirse en ellas (al menos una vez) para fijar su contineido.
- Puede leerse para detectar su estado.

Funcionamiento de una celda de memoria



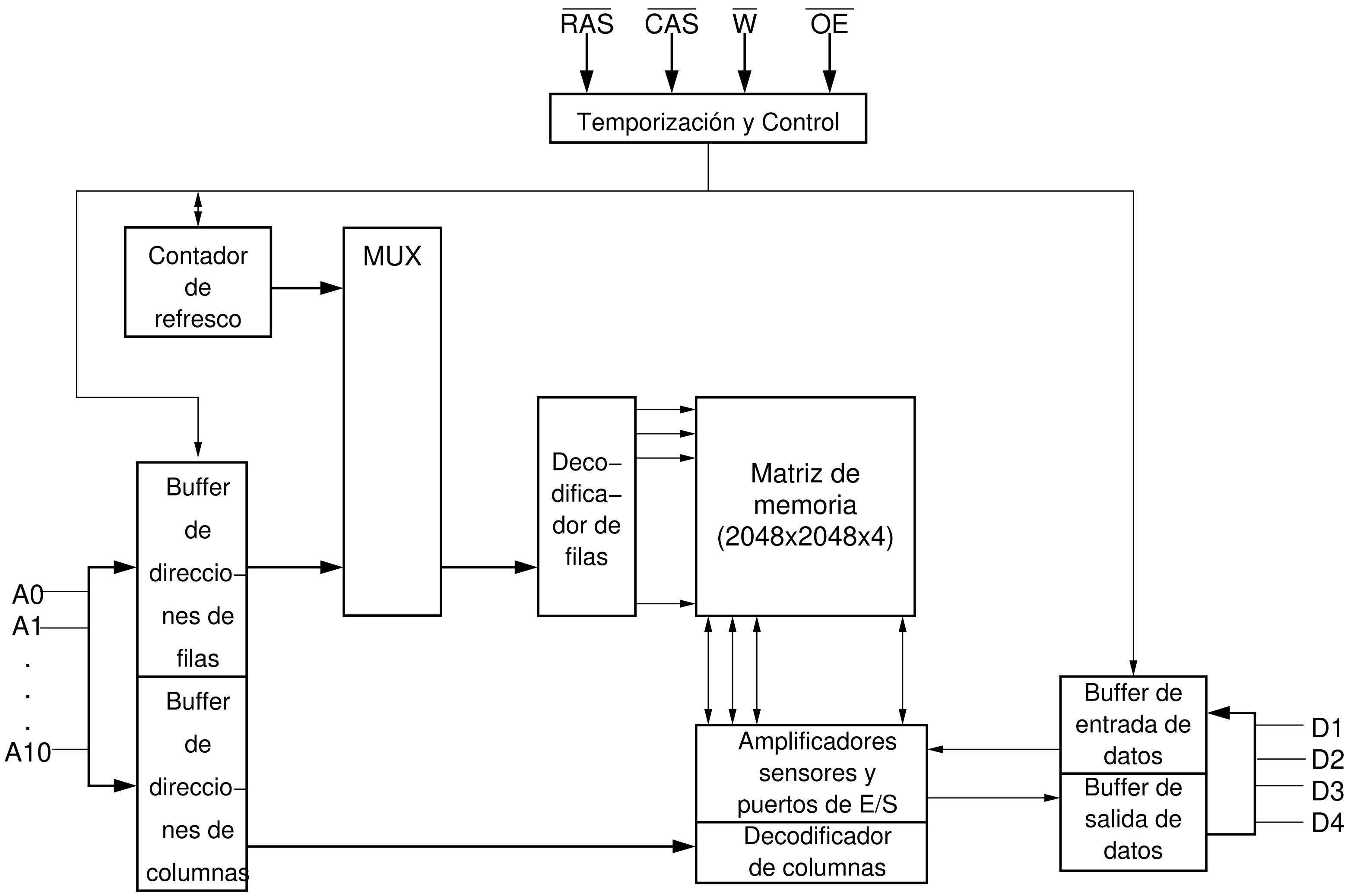
Lógica del chip

La matriz se organiza en W palabras de B bits cada una.

En la siguiente figura se muestra la organización de una DRAM típica de 16 Megabits ($4M \times 4$)

- Se escriben o leen 4 bits a la vez.
- La matriz de memoria está estructurada en 4 matrices cuadradas de 2048×2048 elementos.
- Se ocupan 11 líneas de direcciones para seleccionar 1 entre 2048 filas usando un decodificador y una señal de dirección de fila (RAS), y las otras 11 líneas de dirección y la señal (CAS) seleccionan una columna. Con lógica de selección externa se multiplexan las líneas de dirección.

DRAM típica de 16 Megabits (4M x 4)



Encapsulado de los chips

EPROM encapsulada con 8 Mbits organizados en $1M \times 8$. Una palabra por chip con 32 terminales.

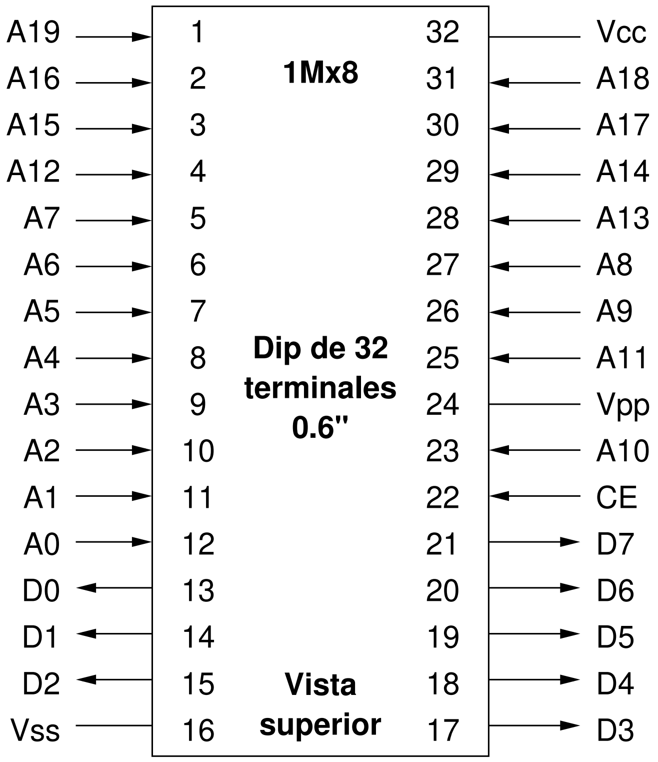
Sus señales son:

- La dirección de la palabra a la que se accede. Para 1M de palabras, se ocupan 20 ($2^{20} = 1M$) terminales (A0-A19).
- Dato a leer con 8 líneas (D0-D7).
- Una línea de alimentación del chip (Vcc).
- Una terminal de tierra (Vss).

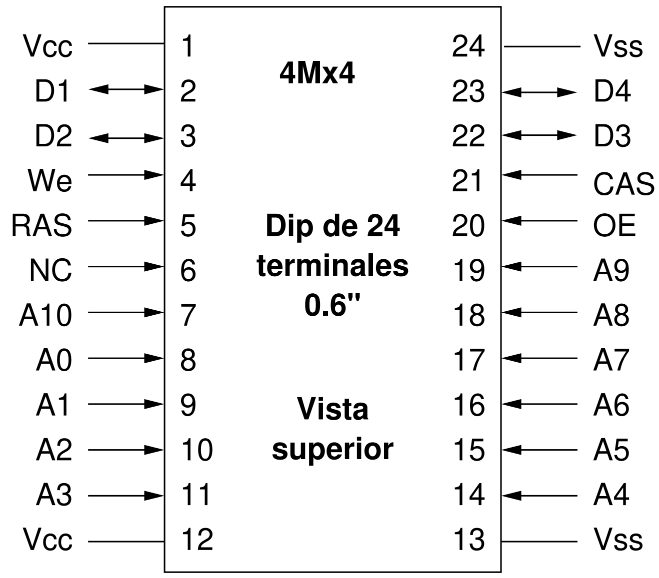
- Terminal de habilitación del chip (CE chip enable), ya que puede haber varios conectados al mismo bus de direcciones, donde CE indica si la dirección es válida o no. Queda activado por lógica cuya entrada son los bits de orden más alto del bus de direcciones (bits de direcciones superiores al A19).

- Tensión de programación (Vpp) que se suministra durante la programación.

Terminales y señales Típicas de chips encapsulados



EPROM de 8 Mbits



DRAM de 16 Mbits

Terminales de una DRAM típica de 16 Mbits

Organizada en $4M \times 4$. Sus terminales de datos son de entrada/salida.

- Habilitación de escritura (WE Write Enable) indica operación de escritura.
- Habilitación de salida (OE Output Enable) indica operación de lectura.
- 11 terminales para especificar 4M de combinaciones fila/columna que están multiplexadas ($2^{11} \times 2^{11} = 2^{22} = 4M$)

Organización en Módulos

Si un chip de RAM contiene 1 bit por palabra, se ocuparán al menos un número de chips igual al número de bits por palabra, por ejemplo si se quiere un módulo de memoria de 256K palabras de 8 bits, se puede emplear 8 chips de $256K \times 1$ bit.

Una memoria de 1M de palabras de 8 bits usando chips de 256×1 bit, requiere de 20 líneas de direcciones y 32 módulos. Los 18 bits menos significativos se conectan a los 32 módulos y los 2 bits de orden más alto son entradas a un decodificador (2-4) para seleccionar un grupo (de 8 módulos) usando una señal de habilitación. Se forman 4 grupos de 8 módulos.

Corrección de errores

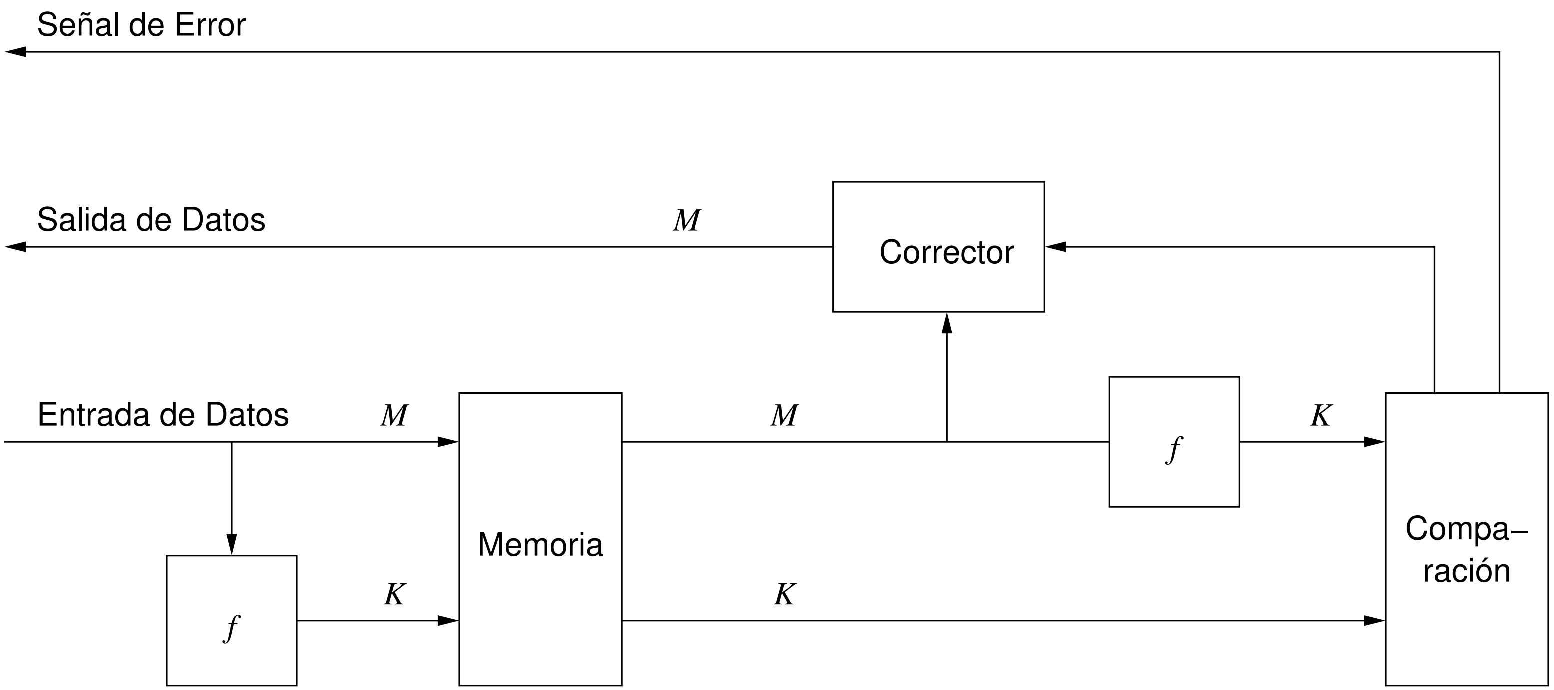
Clasificicación de errores

- Duros (hard). Es un defecto físico permanente en donde la celda o las celdas afectadas no pueden almacenar datos de forma segura, quedándose en 0 o 1, o conmutando erróneamente entre 0 y 1. Se originan por funcionamiento en condiciones adversas, defectos de fabricación o desgaste.
- Blandos (soft). Es un evento aleatorio no destructivo que altera el contenido de una o más celdas, sin dañar la memoria. Se debe a problemas de la fuente de alimentación o partículas alfa, las cuales son generadas por emisión radiactiva ya que hay núcleos radiactivos en casi todos los materiales.

La mayoría de los sistemas modernos incluyen lógica para detectar y corregir errores.

Códigos correctores de errores

Cuando se van a escribir datos en memoria, se realiza un cálculo con los datos (f) para producir un código. Se almacena tanto los datos como el código. El tamaño real de una palabra de datos de M bits y con un código de longitud K bits es $M + K$ bits.



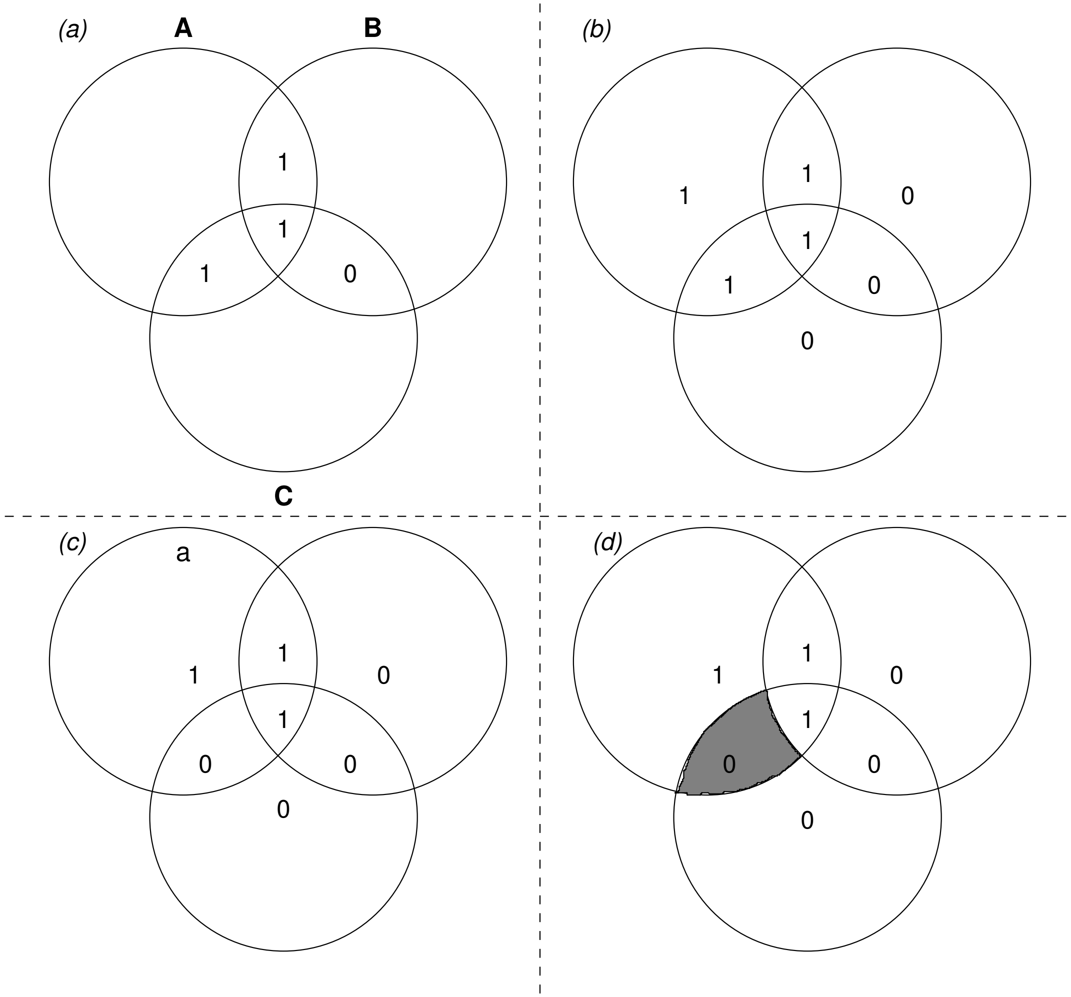
Cuando se va a leer una palabra se utiliza el código para detectar y posiblemente corregir el error. Se genera un nuevo código de K bits a partir de los M bits de datos, que se compara con los bits de código generados en la escritura. La comparación puede producir uno de tres posibles resultados:

- No se detectan errores. Los bits de datos captados se envían al exterior.
- Se detecta un error y es posible corregirlo. A un corrector se dan los bits de datos y los bits de corrección de error, lo que genera un conjunto reducido de M bits, que serán enviados fuera.
- Se detecta un error, pero no es posible corregirlo. Se informa lo anterior.

Código de *Hamming*

Es el código corrector de errores más sencillo ideado por Richard Hamming de los laboratorios Bell. La siguiente figura ilustra el uso empleando diagramas de Venn, sobre palabras de 4 bits ($M = 4$). Al intersectarse tres círculos se tienen 7 compartimentos. Se asignan los 4 bits de dato a los compartimentos interiores. Los demas se llenan con los bits de paridad. El bit de paridad se elige de tal manera que el número total de unos en su círculo sea par. Si un error cambia uno de los bits de datos se encuentra en la intersección de los círculos que tienen error en la paridad.

Código corrector de errores de Hamming



Código para detectar y corregir errores de un sólo bit en palabras de 8 bits.

La comparación bit a bit se hace mediante la OR-exclusiva de las dos entradas de K bits. El resultado se conoce como *palabra de síndrome*. La palabra de síndrome tiene una longitud de K bits y un rango de 2^K valores ($0 \leq K \leq 2^K - 1$). Con un valor 0 se indica que no hubo error, dejando $2^K - 1$ valores para indicar que bit fue el erróneo. Como el error puede ocurrir en los M bits o K bits se debe cumplir:

$$2^K - 1 \geq M + K$$

Para 8 bits de dato se requieren 4 bits de comprobación.

Conviene generar el síndrome con las siguientes características:

- Si el síndrome contiene sólo ceros, no se ha detectado error.
- Si el síndrome contiene sólo un bit puesto a 1, ha ocurrido un error en uno de los cuatro bits de comprobación. No se requiere corrección.
- Si el síndrome contiene más de un bit puesto a 1, entonces el valor numérico de dicho síndrome indica la posición del bit de dato erróneo. Se invierte dicho bit para corregirlo.

Distribución

Los bits de dato y de comprobación se distribuyen en una palabra de 12 bits. Los bits de comprobación se ponen en las posiciones de bit cuyos números son potencia de dos. Cada bit de comprobación opera sobre todo bit de datos cuyo número de posición contiene un 1 en la correspondiente posición de columna, es decir, la posición de bit n es comprobada por aquellas posiciones de bit C_i tal que $\sum i = n$.

Bit	1	2	3	4	5	6	7	8	9	10	11	12
MSB	0	0	0	0	0	0	0	1	1	1	1	1
	0	0	0	1	1	1	1	0	0	0	0	1
	0	1	1	0	0	1	1	0	0	1	1	0
LSB	1	0	1	0	1	0	1	0	1	0	1	0
Compr	C1	C2		C4				C8				
Dato			M1		M2	M3	M4		M5	M6	M7	M8

Ejemplo

Los bits de comprobación quedan calculados como sigue:

$$C1 = M1 \oplus M2 \oplus M4 \oplus M5 \oplus M7$$
$$C2 = M1 \oplus M3 \oplus M4 \oplus M6 \oplus M7$$
$$C4 = M2 \oplus M3 \oplus M4 \oplus M8$$
$$C8 = M5 \oplus M6 \oplus M7 \oplus M8$$

Por lo tanto, suponiendo una palabra de entrada de 8 bits con el contenido 00111001, siendo $M1$ el bit en la posición más a la derecha, se tienen los siguiente cálculos:

$$C1 = 1 \oplus 0 \oplus 1 \oplus 1 \oplus 0 = 1$$
$$C2 = 1 \oplus 0 \oplus 1 \oplus 1 \oplus 0 = 1$$
$$C4 = 0 \oplus 0 \oplus 1 \oplus 0 = 1$$
$$C8 = 1 \oplus 1 \oplus 0 \oplus 0 = 0$$

Suponiendo que en el tercer bit de datos hay un error cambiando de 0 a 1, cuando se recalculan los bits de comprobación se tiene:

$C1 = 1 \oplus 0 \oplus 1 \oplus 1 \oplus 0 = 1$
$C2 = 1 \oplus \cdot 1 \oplus 1 \oplus 1 \oplus 0 = 0$
$C4 = 0 \oplus \cdot 1 \oplus 1 \oplus 0 = 0$
$C8 = 1 \oplus 1 \oplus 0 \oplus 0 = 0$

Cuando se comparan los nuevos bits de comprobación con los antiguos, se genera la palabra de síndrome:

	C8	C4	C2	C1
	0	1	1	1
$\oplus$	0	0	0	1
	0	1	1	0

El resultado 0110 indica que la posición del bit 6, que contiene al tercer bit de dato es errónea.