



**Universidad
Europea de Madrid**

LAUREATE INTERNATIONAL UNIVERSITIES

MEMORIA

MEMORIA SEMICONDUCTORA

© Todos los derechos de propiedad intelectual de esta obra pertenecen en exclusiva a la Universidad Europea de Madrid, S.L.U. Queda terminantemente prohibida la reproducción, puesta a disposición del público y en general cualquier otra forma de explotación de toda o parte de la misma.

La utilización no autorizada de esta obra, así como los perjuicios ocasionados en los derechos de propiedad intelectual e industrial de la Universidad Europea de Madrid, S.L.U., darán lugar al ejercicio de las acciones que legalmente le correspondan y, en su caso, a las responsabilidades que de dicho ejercicio se deriven.

Índice

Presentación	4
Las memorias de los computadores	6
Jerarquía de memoria	8
Características de las memorias I	10
La ubicación	10
Unidad de transferencia	10
Características de las memorias II	12
Método de acceso	12
Características de las memorias III	14
Prestaciones	14
Memoria principal semiconductora	16
Organización de la memoria I	18
Organización de la memoria II	20
Corrección de errores	21
Resumen	24

Presentación

Recordemos un poco cómo es la memoria de un computador. Este componente, al que Von Neumann dio unas atribuciones muy importantes, es además uno de los componentes más críticos en el diseño de un computador.

En sistemas operativos, a este componente se le denomina *recurso crítico* y en arquitectura se le denomina *vital*, por lo que vamos a verlo con detalle.

Antes de empezar a ver sus características más importantes, vamos a hacer un repaso de cosas que ya hemos visto en cursos pasados, pero seguro que no está mal en emplear unos minutos en recordar.

En este tema vamos a refrescar la idea de memoria, haciendo especial mención a la **memoria semiconductora**, o lo que es lo mismo, la **memoria de lectura y escritura eléctrica**.

La memoria es un componente fácil de entender. Dado que su funcionamiento es crítico para el buen rendimiento de la computadora (la eficiencia de los computadores actuales, hoy por hoy, depende de la memoria), este componente ha evolucionado lo suficiente como para dedicarle un buen tiempo a su estudio y comprender bien su funcionamiento completo.

En este tema, ampliaremos y repasaremos conocimientos que ya tenemos. Organizaremos el tema en los siguientes puntos principales:

- Cómo organizan los computadores actuales la memoria.
- Tipos de memoria.



Las memorias de los computadores

La memoria es el origen y el destino de la gran mayoría de las instrucciones que ejecuta un computador actual. Un buen funcionamiento de esta hace que el computador tenga un rendimiento óptimo. Recordemos que en la ejecución de una **instrucción**, puede haber hasta **cuatro accesos a memoria**. Además, en memoria es donde se sitúan los **programas en ejecución**, así como los datos que estos manejan.



La situación ideal sería poner a un computador la **mayor cantidad de memoria** posible para permitir que nunca

se nos quedase cortos a la hora de ejecutar muchos programas simultáneamente; que sea **rápida**, para que haya poca demora en la carga y almacenamiento de los datos; y por supuesto que esta sea **barata**, para que el precio del computador sea competitivo.

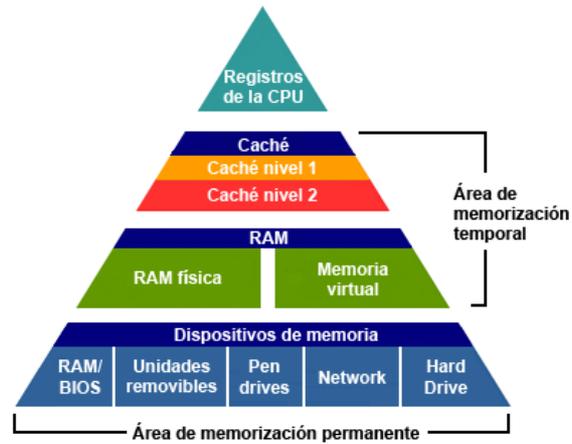
La mayor de parte de estas condiciones son inversamente proporcionales entre sí. El precio y el tamaño suelen ser incompatibles, así mismo el tamaño y la rapidez. ¿Cómo solucionar entonces estos problemas?

Dado que ninguna tecnología es óptima para satisfacer las necesidades de memoria de un computador, la solución que adoptan hoy los computadores actuales es diversificar en los tipos, tecnología, estructura, prestaciones y coste para la construcción de una memoria. O, mejor dicho, de un **conjunto de memorias de un computador**. Es decir, no hay un solo tipo de memoria en un computador, sino que hay toda una jerarquía de memoria con tecnologías, estructura, funcionamiento, capacidad, etc., muy diferentes. Algunos internos (directamente accesible), y otros externos (accesibles mediante entrada/salida).

Jerarquía de memoria

La pregunta es, por tanto, ¿cuánta? ¿cuán rápida? ¿de qué coste... hacemos la memoria de un ordenador?

El diseñador desearía utilizar tecnologías de memoria que proporcionen gran capacidad, rapidez y coste por bit bajo.



Jerarquía de memoria

De esta forma, se elige una estructura piramidal, donde verticalmente hemos expresado la rapidez y el coste (más arriba, más rápida y más cara), mientras que en horizontal hemos representado el tamaño: a más ancho, más tamaño. Con esta estructura expresamos en vertical la rapidez y el coste (más arriba, más rápido y más cara), mientras que en horizontal tenemos el tamaño: a más ancho, más tamaño.

La clave del éxito de esta organización está en la disminución de la frecuencia de acceso a los niveles de memoria inferiores. Se espera que el microprocesador utilice muchísimo más frecuentemente los datos almacenados en extractos superiores que inferiores. Por el contrario, se utilizarán los extractos inferiores para dar alta capacidad de almacenaje de datos, a pesar de que el acceso a los mismos sea lento.

La base para la validez de este esquema es el principio conocido como localidad de las referencias. En el curso de la ejecución de un programa, las referencias a memoria por parte del procesador, tanto para instrucciones como para datos, tienden a estar agrupadas (bucles iterativos y subrutinas).

Instrucciones agrupadas

Las operaciones sobre tablas o matrices conllevan accesos a un conjunto de datos agrupados. En períodos de tiempo largos, las agrupaciones en uso cambian, pero en períodos de tiempo cortos, el procesador trabaja principalmente con grupos fijos de referencias a memoria.

Gran capacidad, rapidez y coste por bit bajo

Lo que nos lleva a algunas contradicciones, pues:

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

Características de las memorias I

Si pensásemos en dotar al computador de una memoria, o de un conjunto de memorias, tendríamos que conocer bien las características de ésta. Pero, ¿cuáles son las características que permitirán diferenciar unas memorias de otras? Son muchas, algunas son obvias, como la rapidez, el tamaño o precio. Estas características normalmente vienen derivadas de características más técnicas. De ellas, las más destacables son las que se exponen a continuación.

La ubicación

Es el aspecto más visible de la memoria y tiene que ver con dónde está colocada. Es decir, a qué buses está conectada.

Memoria interna	Identifica a menudo con la memoria principal. Está conectada a los buses de sistema principal y es direccionable directamente por el microprocesador. Ejemplos de esta memoria son la memoria principal del computador o los registros de la CPU.
Memoria externa	Consta de dispositivos de almacenamiento periféricos, tales como discos y cintas. No son accesibles directamente por la CPU, sino que hay que hacer una petición a través de controladores de E/S. Ejemplo son los discos duros, CD-ROM, etc.

Unidad de transferencia

La unidad de transferencia describe cuantos datos (bytes, palabras, etc.) son transmitidos desde la memoria por cada unidad de acceso. Es decir, cuánta es la información mínima que se puede leer o escribir de la memoria por cada acceso a la misma.

La palabra	Es la unidad "natural" de organización de la memoria típica, como la memoria principal de un computador. Suele coincidir con el número de bits utilizados para representar números y con la longitud de las instrucciones. Las memorias que tienen unidad de transferencia de tipo palabra son la memoria principal del ordenador (RAM) o la caché.
------------	---

El bloque	Es una unidad mucho mayor que la palabra. Su tamaño es normalmente de entre 256 bytes hasta las 256Kbyte. Por tanto, su información no suelen ser datos simples, sino más bien bloques de información elaborada. Ejemplo de transferencia a nivel de bloque es el disco duro o la tarjeta de red.
-----------	---

Al hablar de unidad de transferencia, hay que diferenciar dos características: unidad direccionable y unidad de transferencia.

Características de la unidad de transferencia

- **Unidad direccionable:** mayoritariamente es la palabra. Sin embargo, algunos de ellos permiten direccionar a nivel de bytes. En cualquier caso, la relación entre la longitud A de una dirección y el número N de unidades direccionables, es $2^A=N$.
- **Unidad de transferencia:** es el número de bits que se leen o escriben en memoria a la vez. No tiene por qué coincidir con una palabra o con una unidad direccionable. Para la memoria externa, los datos se transfieren normalmente en unidades más grandes que la palabra, denominadas bloques.

Características de las memorias II

Método de acceso

Responde a cómo se acceden a los datos guardados en memoria. No se accede a todas las memorias del mismo modo. El método de acceso típico, mediante una dirección de memoria, es solo uno de los muchos métodos existentes.





Las memorias de este tipo, son un híbrido entre el acceso secuencial y el aleatorio. Los bloques individuales (o registros) 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. De nuevo el tiempo de acceso es variable. Las unidades de disco (HD) son ejemplos de este modo de acceso. La dirección de un dato en el disco se especifica mediante tres números exactos (cabezacilindro-sector), pero esta posición consta de muchos datos, por lo que la lectura de cada dato se realiza en un orden secuencial.



Es una memoria del tipo de acceso aleatorio que permite hacer una comparación de ciertas posiciones de bits dentro de una palabra buscando que coincidan con unos valores dados, y hacer esto para todas las palabras simultáneamente. Una palabra es por tanto recuperada basándose en una porción de su contenido en lugar de su dirección. El tiempo de recuperación de un dato es una constante independiente de la posición o de los patrones de acceso anteriores. Es decir, la información en una memoria de tipo asociativo, no tiene nada que ver con su posición, sino que a cada dato se le asigna una etiqueta. De modo que la recuperación de la información, no se hace por dirección, sino por etiqueta. Es como si se guardasen datos del tipo "peso=35", "edad=18", etc. Y la recuperación del dato se hace por "¿edad?", "¿peso?" y en ningún momento la posición que esta ocupa en la memoria es relevante. Un ejemplo de esta memoria es la memoria caché de un computador actual.

Características de las memorias III

Prestaciones

Con prestaciones nos referimos a la velocidad y a la capacidad de transmisión de datos. Para medirlas vamos a contabilizar las siguientes características.

Tiempo de acceso	<p>Es el tiempo que tarda en realizarse una operación de escritura o de lectura. El tiempo que transcurre desde el instante en el que se presenta una dirección a la memoria hasta que el dato ha sido memorizado o está disponible para su uso.</p> <p>Para memorias no aleatorias, es el que se tarda en situar el mecanismo de lectura/escritura en la posición deseada.</p>
Tiempo de ciclo	<p>Es el tiempo de acceso y el tiempo que se requiere antes de que pueda iniciarse un segundo acceso a memoria.</p> <p>Este tiempo adicional puede que sea necesario para que finalicen las transiciones en las líneas de señal o para regenerar los datos en el caso de lecturas destructivas.</p>
Velocidad de transferencia	<p>Es la velocidad a la que se pueden transferir datos a, o desde, una unidad de memoria.</p> <p>Para memorias de acceso aleatorio coincide con el inverso del tiempo de ciclo.</p>



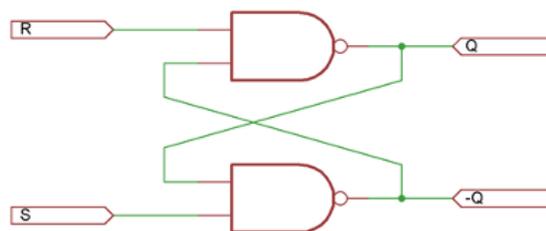
Memoria principal semiconductor

En computadores antiguos, la forma más común de almacenamiento de acceso aleatorio para la memoria principal consistía en una matriz de pequeños anillos ferro-magnéticos denominados núcleos.

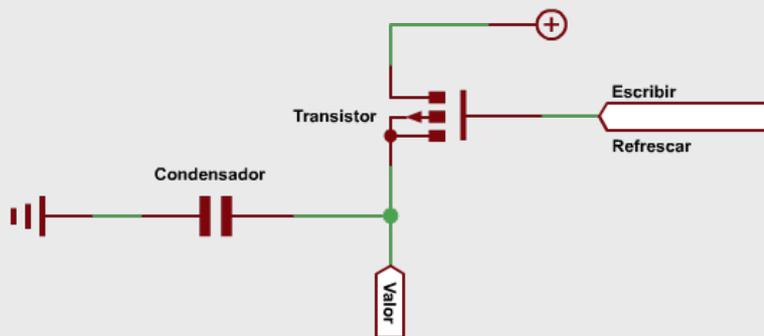
 [Memoria de núcleo de ferrita](#)
Documentos

Es por esto que la memoria principal recibía a menudo el nombre de núcleo. Pero la mejora de la microelectrónica, y sus ventajas, acabó con las memorias de núcleos. Hoy es casi universal el uso de memorias semiconductoras para la memoria principal. Una [memoria semiconductor](#) y la información se guardan a través de algún sistema eléctrico (corrientes eléctricas). Básicamente hay dos tipos.

RAM dinámica	Con siglas DRAM. Está hecha con celdas que almacenan los datos como cargas en condensadores. La presencia o ausencia de carga en un condensador se interpretan como el 1 o el 0 binarios. Ya que los condensadores tienen una tendencia natural a descargarse, las RAM dinámicas requieren refrescos periódicos para mantener memorizados los datos. Son más pequeñas físicamente y de menor coste por bit, pero más lentas.
RAM estática	De siglas SRAM. Los valores binarios se almacenan utilizando configuraciones de puertas formando biestables (<i>flip-flop</i>). Una RAM estática retendrá sus datos en tanto se mantenga alimentada, no necesita refresco periódico. Son más rápidas que las dinámicas, pero más caras y de mayor consumo eléctrico, por lo que necesitan refrigeración.



Esquema de un bit de memoria RAM Estática.

Memoria semiconductora

Ejemplo de núcleo de memoria semiconductora, concretamente un bit dinámico implementado mediante un transistor y un condensador

Pequeñas físicamente

A pesar de que en la imagen son de tamaño similar, recuérdese que una puerta AND como una de las dos que se necesitan en el flip-flop R-S, esta implementado por 4 transistores como el de la imagen.

Organización de la memoria I

Seguro que ya hemos visto algún zócalo de memoria RAM de las que se le pone al ordenador.

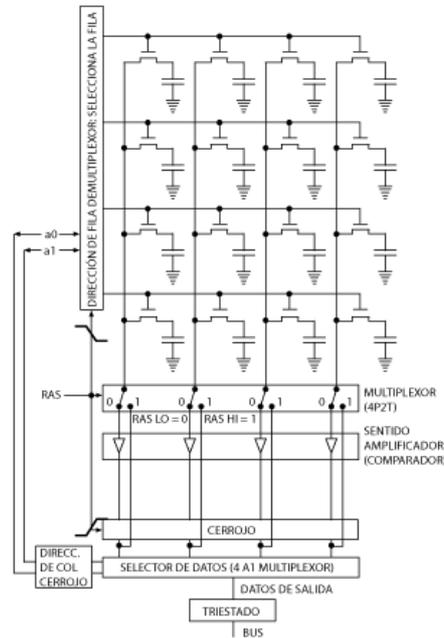
Todas tienen una característica común: a excepción de los pen drive, en cuya elaboración se antepone el tamaño a la eficiencia o disipación de temperatura, hay muy pocos módulos comerciales que estén formados por un único circuito integrado. Por lo general, están formados por varios módulos más pequeños que se unen para formar un circuito más grande.

Con organización conocemos cómo está integrado por dentro. La manera más simple de pensar cómo está organizado un circuito de memoria es imaginar una matriz enorme cuya vertical se organiza en función de los datos, mientras su horizontal queda organizada por el tamaño de cada dato.

Por ejemplo, si tenemos una memoria de 1Mgb (actualmente considerada muy pequeña), supondremos que disponemos de una matriz de un millón de celdas en vertical por 0 celdas en horizontal. Esta sería una buena solución, pero el problema está en el selector de celda.

Ahora que sabemos cómo es un decodificador y cómo se implementa, podemos determinar que este decodificador, si no imposible de construir, sería extremadamente grande, costoso y caro. Por lo que la solución sería hacer integrados de memoria más pequeños hasta que los costes de fabricación sean asequibles.

Pero hay otra solución, darle al chip de la memoria una estructura tridimensional.



Esquema de memoria lineal dinámica organizada en una columna (cuatro direcciones) de 4 bits cada dirección.

Un circuito más grande

Cada uno de estos circuitos funciona con un chip de memoria independiente pero ¿cómo funciona cada una de estos circuitos? ¿de qué capacidades existen? La respuesta es muy simple, hay muchas variedades, pero no hay todas las posibilidades.

Zócalo de memoria RAM

En la imagen, puedes comprobar el aspecto de un zócalo de memoria RAM.

**Selector de celda**

Recordemos que una memoria se puede leer y escribir con la misma velocidad, con independencia de la celda que se seleccione, lo que origina que haya que hacer un decodificador de 30 bits de entrada por un millón de salidas.

Organización de la memoria II

En una estructura tridimensional, la matriz de datos no es una matriz bidimensional en la que en vertical tenemos los elementos y en horizontal el tamaño de cada elemento. Sino que pasamos a una estructura donde, en vertical tenemos $\sqrt[3]{N}$ celdas, donde N es el número de elementos almacenados en la memoria, en vertical tenemos también $\sqrt[3]{N}$ celdas y, en la tercera dimensión, tenemos el número de bits que forman el elemento guardado en la memoria.

Así, para una memoria de 1Mgb como la descrita anteriormente, la matriz de datos tendría un tamaño de $1024 \times 1024 \times 8$, lo que implica que se necesitarían dos decodificadores de 1024 elementos de salida. Este decodificador es mucho más pequeño que el anterior (no es la mitad, ni la cuarta parte, sino varias potencias inferior), con lo que hablamos de decodificadores alcanzables y de coste reducido.

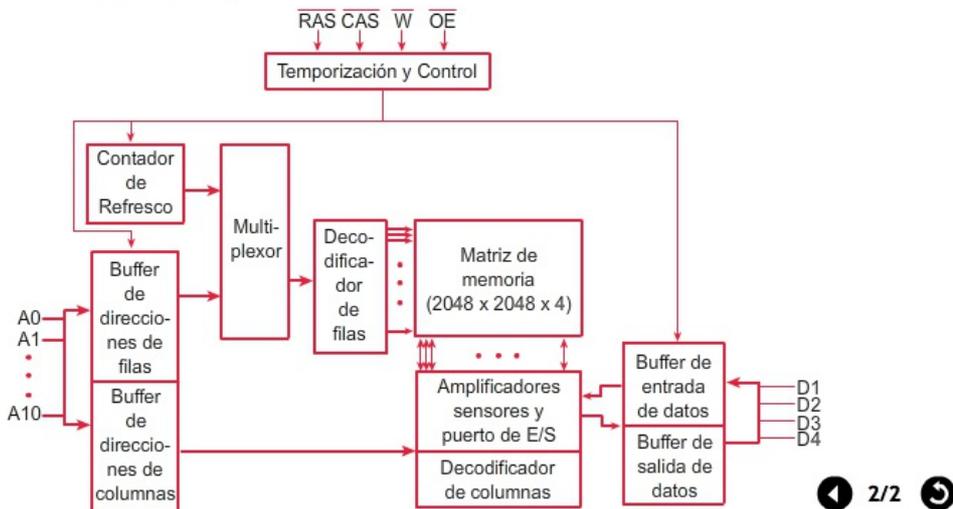
Al tener la información con esta estructura, las direcciones de memoria, de 20bits para el caso anterior, se convierten en 10bits (la mitad) para selección fila (verticalidad) y 10bits para seleccionar columna (horizontalidad).

Si extrapolamos esta idea podemos pensar que la dirección de un dato almacenado en memoria consta de N/2 bits para selección fila y N/2 para selecciona columna.

Si para terminar, hacemos el bus de direcciones de tamaño N/2, y ponemos dos señales extra llamadas RAS (Row Address Strobe o Row Address Select) y CAS (Column Address Strobe o Column Address Select). Así, por ejemplo, podemos direccionar 1Mgb de datos con solo 10+2 bits de direcciones.

1/2 

En la imagen, tenemos un ejemplo de memoria de 4Mgb de datos de 4 bits cada dato. En el dibujo se aprecian con facilidad los registros de direcciones (MAR), los de datos (MBR), las señales de RAS y CAS, los decodificadores de filas y columnas, las 11 líneas de direcciones (a la izquierda), las 4 líneas de datos (por la derecha) y dos señales extra W (Write = Escritura/lectura) y OE, conocida como One Chip-select (= activada, a veces se denota como CS).



Corrección de errores

Las memorias semiconductoras, a pesar de ser dispositivos muy estables, también tienen fallos.

Estos pueden clasificarse en fallos duros y errores blandos.

Obviamente, los errores de memoria son un problema grave ya que introducen aleatoriedad al correcto funcionamiento de un ordenador sin un patrón predecible, por lo que son muy difíciles de detectar y aún más de corregir. Por este motivo, la mayoría de los sistemas de memoria modernos incluyen lógica para detectar y corregir errores. Entre los sistemas más estandarizados están los siguientes.

Bit de paridad

Este método sirve para detectar pequeños fallos. Lo que se hace es ampliar en un bit el dato almacenado en memoria. Este bit no pertenece al dato, pero sirve para almacenar un bit, que contendrá un 0 ó 1 en función de si establecemos paridad par o impar. Este bit completa el número de unos presente en el dato, provocando, junto con los unos ya existentes en el número, una suma par (o impar) de unos.

Es un método fácil de implementar, pero que solo puede detectar un fallo. Es decir, si en el número se intercambian dos valores (se produce un fallo en dos bits distintos) el bit de paridad da el valor almacenado como correcto.

Código Hamming	<p>Este es uno de los códigos más utilizados ya que además de ser detector de errores, también es corrector.</p> <p>El código Hamming amplía el número de bits almacenados en memoria en K bits, siendo K el número que cumple $2^k > N$, y N el número de bits que se almacena en memoria. Para entender el código Hamming, te recomendamos consultar el documento que se enlaza a continuación.</p>
----------------	--

 [Código Hamming](#)
Documentos

Sistemas más estandarizados

Estos códigos están implementados internamente en la memoria, siendo transparentes para el usuario. A pesar de que el código Hamming es más seguro, el bit de paridad está siendo muy utilizado. Pero no por su facilidad, sino porque la tasa de fallos en una memoria es muy baja, lo que propicia que este método sea suficiente para establecer un nivel de seguridad aceptable.

Por el contrario, los códigos Hamming son utilizados en transmisiones de datos pues dado que el índice de fallos es muy elevado, permite aumentar la seguridad de los datos y la rapidez, ya que, al poder corregir el dato con errores, no es necesaria la retransmisión del mismo.

Fallo duro

Un **fallo duro** es un defecto físico permanente, las celdas de memoria afectadas no pueden almacenar datos de manera segura, quedándose ancladas a 0 ó a 1, o conmutando aleatoriamente. Los errores duros suelen estar causados por funcionamiento en condiciones adversas, defectos de fabricación y desgaste.

Error blando

Un **error blando** es un evento aleatorio no destructivo que altera el contenido de una o más celdas de almacenamiento, sin dañar a la memoria. Los errores blandos pueden deberse a problemas de la fuente de alimentación o ruido electromagnético.

Resumen

La memoria es un elemento muy simple en su definición, pero se convierte en un elemento muy complejo al querer responder efectivamente a varias premisas que son inversamente proporcionales entre sí: el precio, la eficiencia y el tamaño.

Estas premisas provocan la necesidad de crear una jerarquía de memoria, la necesidad de diversificar la tecnología utilizada y de especificar el almacenamiento de datos en niveles.

En este tema, hemos comprobado cómo se implementa la memoria en un ordenador, cómo son internamente los módulos de memoria y cómo estos se protegen contra fallos con los sistemas correctores de errores.

Los puntos destacados que hemos revisado en este tema son:

- Jerarquía de memoria.
- Características claves de los sistemas de memoria de computadores.
- Organización de la memoria.
- Sistemas correctores de fallos en la memoria.