# La Memoria

Desde un punto de vista genérico, la memoria es la parte del ordenador que se encarga de almacenar los datos que intervienen en el proceso. Sin embargo, dentro del sistema informático existen diferentes tipos de memoria, que vamos a ir desglosando:

- Registros: Los registros son pequeñas porciones de memoria que se encuentran integradas en el procesador y que, por lo tanto, funcionan a la misma velocidad que éste.
- Caché: Su funcionamiento es similar al de la memoria principal, que comentaremos a continuación, pero su tamaño es mucho menor y su acceso mucho más rápido.
- La idea es que, cuando el procesador necesita acceder a un dato, se copia a la caché todo el bloque que lo contiene. Así, si se producen accesos sucesivos (para leer o escribir) al mismo dato o a sus datos circundantes, el tiempo de acceso se reduce drásticamente.
- Memoria principal: También se llama Memoria de acceso aleatorio o Memoria RAM (del inglés, Random-Access Memory) porque en los primeros ordenadores era la única memoria que permitía acceder a los datos sin seguir un orden previo.

  Este es el lugar donde deben encontrarse tanto las instrucciones como los datos para que el procesador pueda utilizarlos. Su contenido se organiza en posiciones de memoria que están identificadas de forma individual por una dirección única.

- Dispositivos de almacenamiento externo: Son dispositivos que permiten almacenar grandes volúmenes de información.

  Su principal característica es que no es volátil, es decir, no necesitan un suministro continuo de corriente eléctrica para mantener la información que contienen.

  Existen dispositivos construidos a partir de tecnologías muy diferentes, como las unidades magnéticas (HDD, del inglés Hard Disk Drive), las ópticas (CD/DVD, del inglés Compact Disc/Digital Versatile Disc) o las flash (SSD, del inglés Solid State Drive).

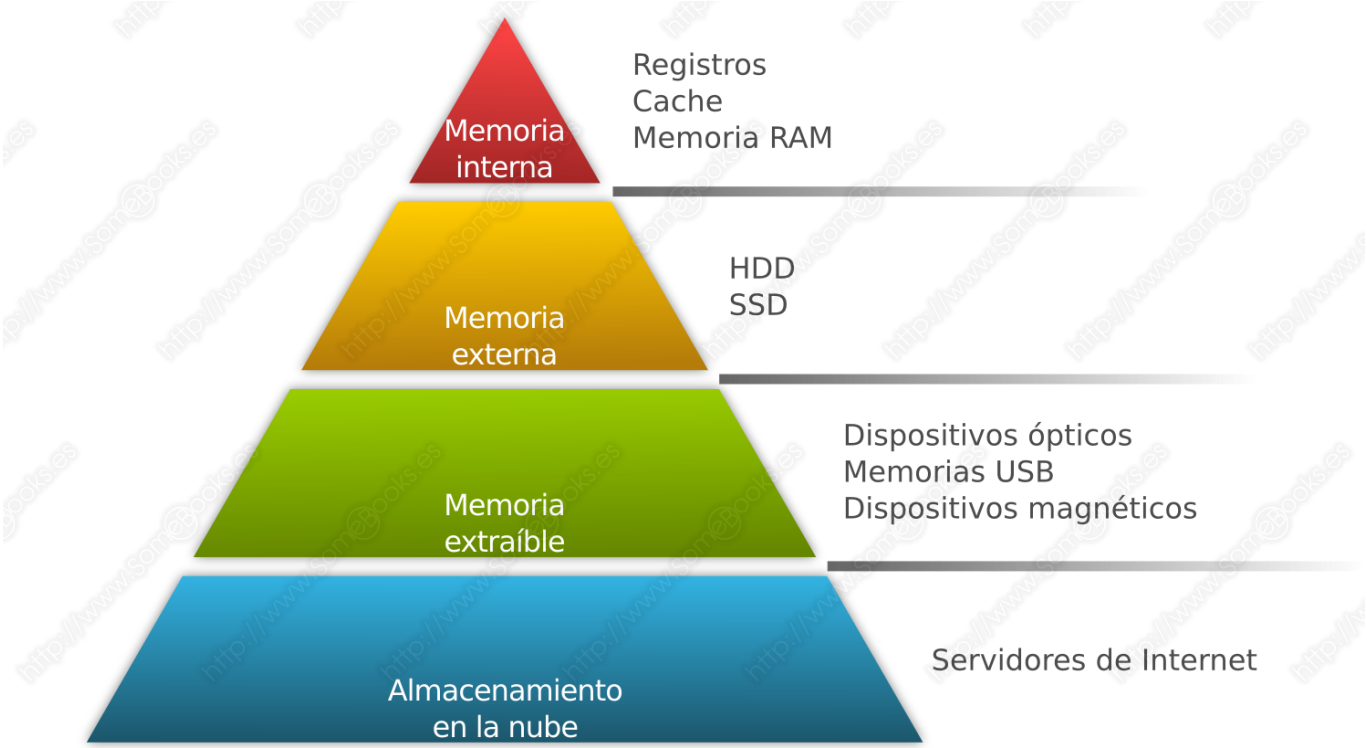
- Dispositivos de almacenamiento secundario o extraíble: Son dispositivos de gran capacidad, destinados fundamentalmente a la realización de copias de seguridad.

  Suelen ser dispositivos de acceso secuencial, como las cintas DAT (del inglés Digital Audio Tape).

> ☞ *Tanto la memoria principal, como la memoria caché y los registros son memoria volátil. Es decir, estos tipos de memoria sí necesitan constantemente alimentación eléctrica para recordar su contenido.*

En principio, existen tres datos fundamentales que debemos tener en cuenta cuando nos referimos a la memoria: su cantidad, su velocidad y el coste por unidad de almacenamiento (por ejemplo, el coste por byte).

Si representamos gráficamente los diferentes tipos de memoria en función de su tamaño medio, obtenemos una jerarquía de la memoria con forma piramidal:




Si tomamos como punto de partida la imagen anterior, podemos afirmar que, según se desciende en la jerarquía, se cumplen las siguientes características:

- Disminuye el coste por byte.
- Aumenta la capacidad.
- Aumenta el tiempo de acceso.
- Disminuye la frecuencia con la que se accede a la memoria.

Con el fin de abaratar costes, siempre se ha tratado sustituir los tipos de memoria más cara, pequeña y rápida por otros más baratos, grandes y lentos. El objetivo final es perder la menor cantidad de rendimiento posible.

El truco para conseguirlo está en aumentar la frecuencia con la que se accede a los dispositivos más rápidos en detrimento de los más lentos. La estrategia a seguir se basa en el principio de cercanía de referencias que, básicamente, consiste en que mientras se está ejecutando un determinado programa en el procesador, las referencias que se hagan a posiciones de memoria tenderán a estar agrupadas.

Esto permite acercar los datos al procesador, de modo que la cantidad de acceso a los niveles de memoria inmediatamente inferiores se vea considerablemente reducido.

Estas técnicas se aplican en diferentes niveles de la jerarquía de memoria:

- Usando memoria caché entre el procesador y la memoria RAM
- Las técnicas de memoria virtual, que pretenden simular una mayor cantidad de memoria RAM de la que existe realmente, usando el disco como almacenamiento de apoyo

- La caché de disco, que utiliza parte de la memoria RAM para guardar temporalmente los datos que deben transferirse al disco, lo que permite que las escrituras se agrupen, ahorrar accesos cuando un dato se escribe varias veces o recuperarlo más rápidamente si lo volvemos a usar poco después de escribir.

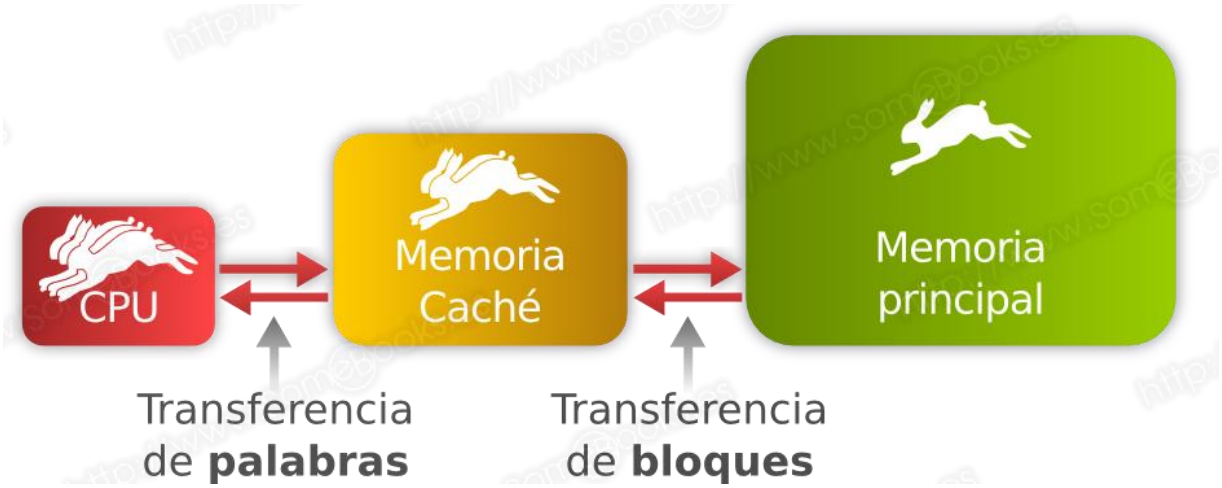
A modo de ejemplo, vamos a explicar con más detalle el funcionamiento de la memoria caché.

### *Memoria caché*

Ya hemos comentado que la memoria principal funciona a una velocidad muy inferior a la del procesador. Sin embargo, éste debe acceder a la memoria para obtener cada una de las instrucciones que debe ejecutar (muchas veces, tendrá que volver para obtener los datos involucrados en la instrucción). Resulta evidente la carga que supone esta situación para el rendimiento del procesador.

Para resolverlo, los diseñadores recurren al principio de cercanía. La idea consiste en colocar, entre el procesador y la memoria principal, una memoria de poco tamaño y gran velocidad, a la que llamamos memoria caché.

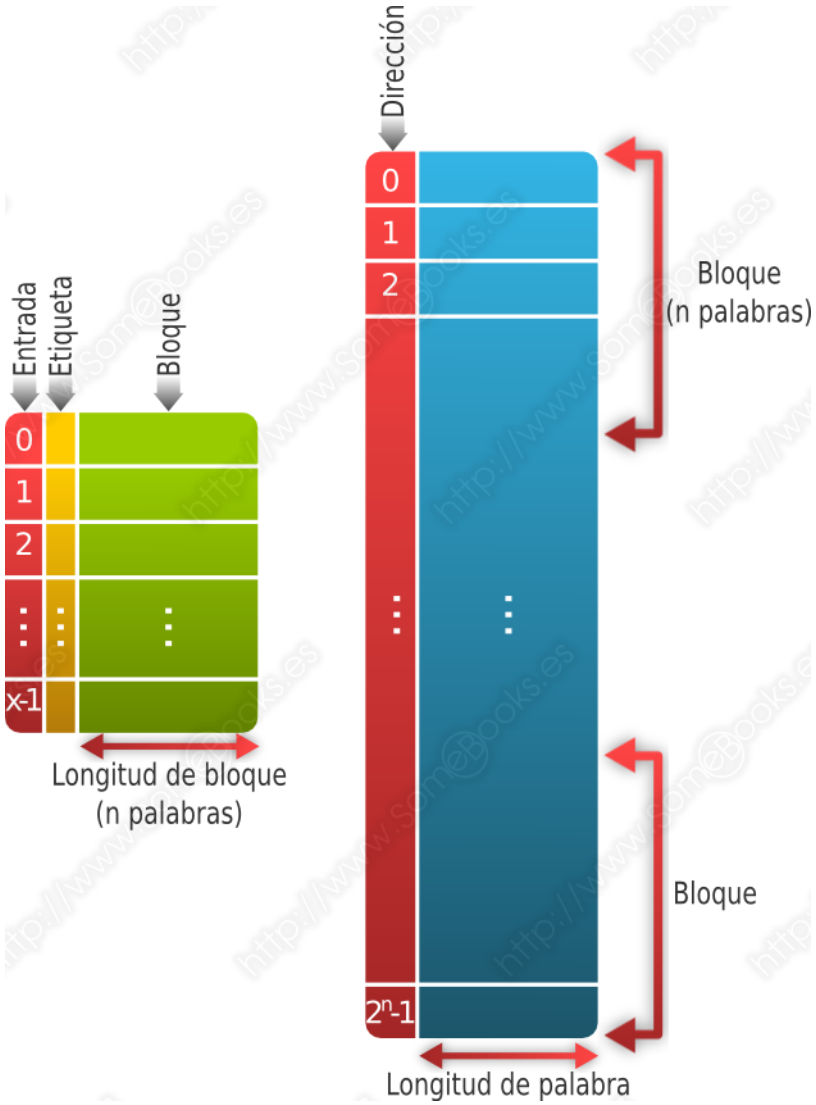
La idea es que, cada vez que el procesador solicite un dato de la memoria principal, se busque primero en la memoria caché. Si no se encuentra, se lee de la memoria principal el bloque completo que contiene el dato solicitado y se guarda en la caché (lógicamente, también se envía al procesador el dato que solicitó).




Según el principio de cercanía de referencias, es muy probable que, la próxima vez que el procesador solicite un nuevo dato, éste pertenezca al mismo bloque del dato anterior, con lo que se podrá devolver de forma casi inmediata.

Según el principio de cercanía de referencias, es muy probable que, la próxima vez que el procesador solicite un nuevo dato, éste pertenezca al mismo bloque del dato anterior, con lo que se podrá devolver de forma casi inmediata.

> ☞ *Por cuestiones de eficiencia, la memoria caché se gestiona completamente a nivel hardware y es invisible para el software del ordenador.*




Cuando se diseña un nuevo procesador, el equipo de diseño debe enfrentarse a varias preguntas que definirán los principios en los que estará basado el funcionamiento de la memoria caché. Estas son algunas de ellas:

¿Qué tamaño debe tener la memoria caché? En este sentido, debemos saber que no será necesaria una memoria caché muy grande para que su impacto sobre el rendimiento sea notable.

¿Qué tamaño debe tener cada bloque de memoria? Para responder a esta pregunta es muy importante otro concepto denominado tasa de aciertos. Es decir, el porcentaje de veces en el que se encontrará en memoria caché el dato que se está buscando.

Es fácil deducir que, a medida que escojamos bloques de memoria más grandes el principio de cercanía de referencias hará que los datos más alejados del que produjo la lectura del bloque, tengan menos probabilidades de ser utilizados. Además, al ser bloques más grandes, cabrán menos bloques en la caché. Esto deriva en una memoria caché saturada con datos que tienen pocas posibilidades de ser utilizados y la tasa de aciertos descenderá.

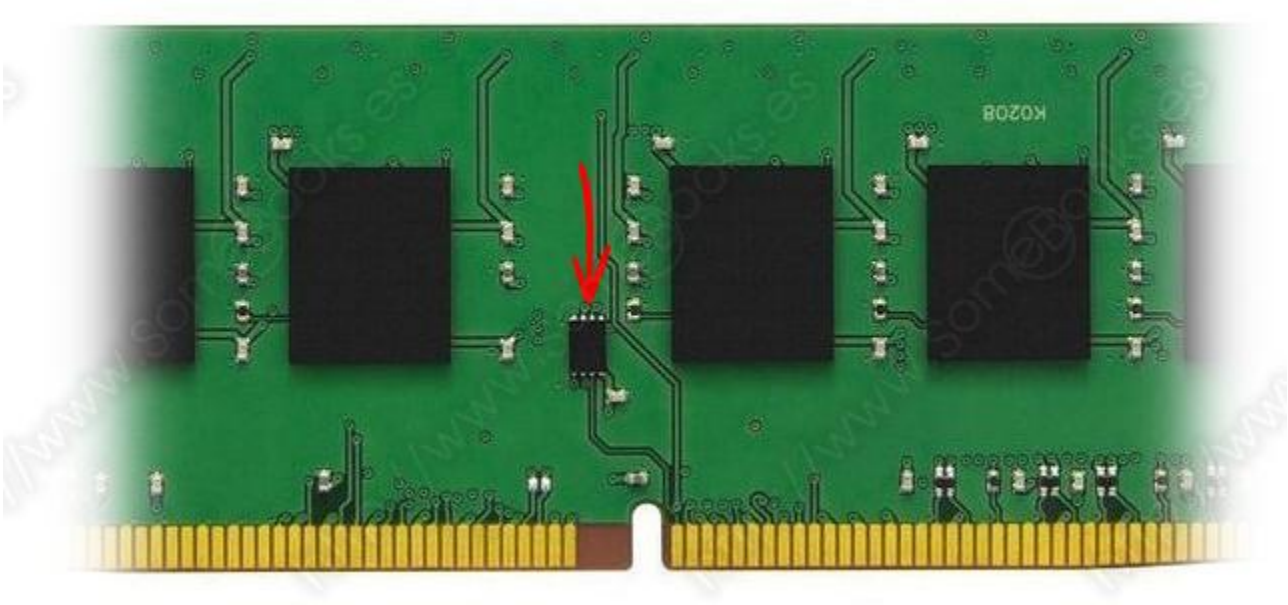
Por el contrario, si se eligen bloques demasiado pequeños, pueden quedar fuera del bloque algunos datos que, según el principio de cercanía de referencias, tengan una probabilidad elevada de ser referenciados. Esto también hará que descienda la tasa de aciertos.

Por consiguiente, el tamaño óptimo del bloque será el que ofrezca una mayor tasa de aciertos.

### *La memoria RAM desde el punto de vista físico*

Desde un punto de vista físico, la memoria RAM suele presentarse en forma de módulos. Éstos consisten, fundamentalmente, en pequeñas tarjetas de circuito impreso en cuya superficie se sueldan los chips de memoria.

El circuito impreso obtiene electricidad y se comunica con la placa base a través de una serie de contactos situados en uno de su laterales. Para lograrlo, utiliza el protocolo SPD (del inglés, Serial Presence Detect). Un pequeño microprocesador situado sobre la superficie del módulo, que suele recibir el mismo nombre, informa a la BIOS sobre las características del módulo (fabricante, velocidad, tamaño, voltaje, etc) y ésta configura el controlador para que la comunicación se produzca de forma óptima.

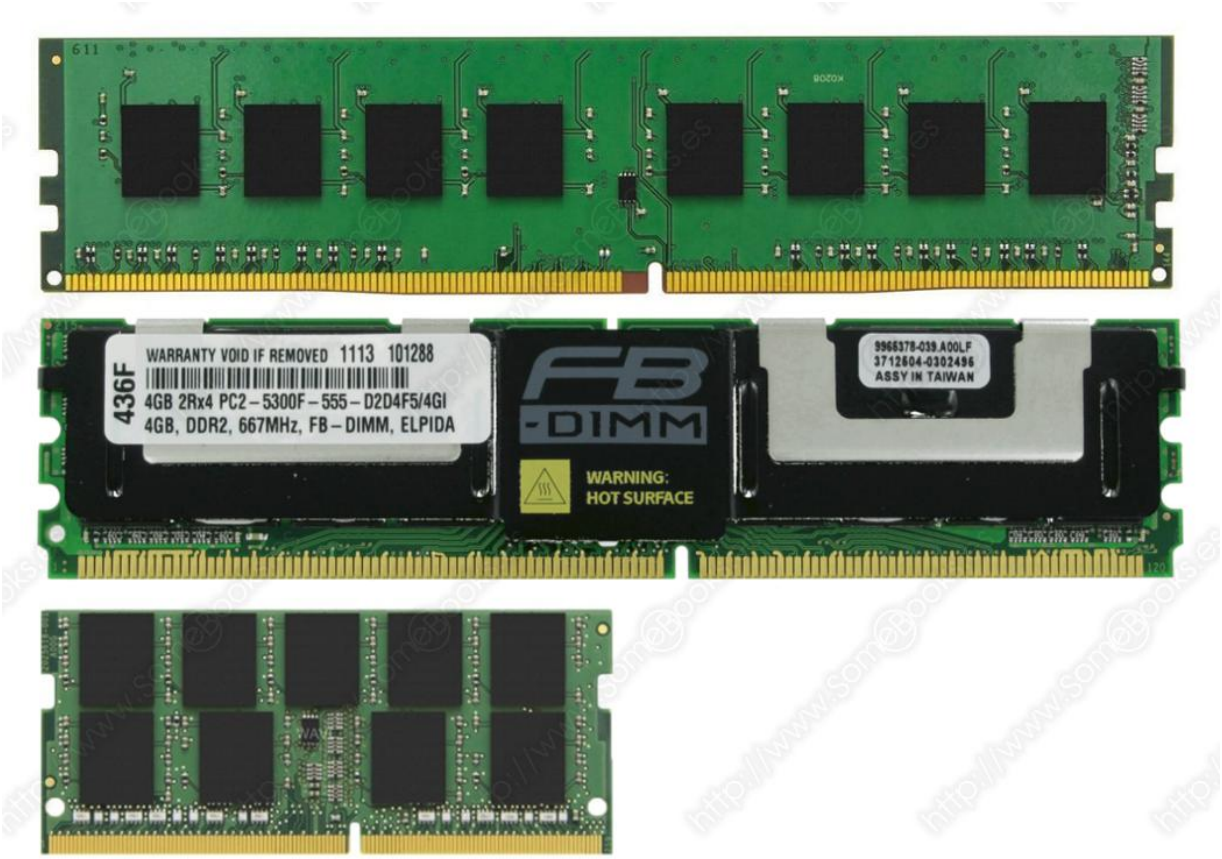


Los chips de memoria pueden aparecer en una cara del módulo o en las dos y suelen utilizar tecnología DRAM (Dynamic Random Access Memory, es decir, memoria dinámica de acceso aleatorio).

Lo más frecuente es que los módulos de memoria utilicen el reloj del sistema para mantenerse sincronizados, por lo que reciben el nombre de SDRAM (Synchronous Dynamic Random Access Memory).

Para que puedan combinarse módulos de diferentes fabricantes con distintas placas base, durante su diseño y fabricación deben seguirse de forma rigurosa las normas del Joint Electron Device Engineering Council (JEDEC). En este sentido, los módulos actuales siguen tres factores de forma diferentes:

- DIMM (del inglés Dual In-line Memory Module), que se utiliza fundamentalmente en ordenadores de sobremesa. Sus dimensiones son de 133,35 X 30 milímetros y un grosor máximo de 4 milímetros (el grueso del circuito debe ser de 1 milímetro), aunque en algunos módulos varía ligeramente la altura
- FB-DIMM (del inglés Fully-Buffered DIMM), destinado a servidores. Como las DIMM normales, tienen unas dimensiones de 133,35 x 30 milímetros, pero 5,1 milímetros de grosor máximo.
- SO-DIMM (del inglés Small Outline DIMM), usado principalmente en ordenadores portátiles. Su tamaño es de 67,6 x 30 milímetros y un grosor máximo de 3,8 milímetros.




Los módulos de memoria actuales utilizan tecnología DDR (del inglés Double Data Rate), que tiene la capacidad de transmitir datos dos veces en cada pulso de reloj, una vez durante la subida y otra durante la bajada de la señal. De esta forma, se duplica la velocidad de la memoria sin modificar la frecuencia de reloj del sistema.

En este sentido, existen tres conceptos que nos ayudan a medir la velocidad de una memoria:

- **El tiempo de acceso**: Representa el tiempo que tarda la memoria en devolver un dato almacenado en su interior. Se mide en nanosegundos.
- **La latencia**: Representa el tiempo que tarda la memoria en situarse sobre una posición particular.
- **La tasa de transferencia**: la cantidad de información que puede intercambiarse, en cada segundo, entre el módulo y el controlador de memoria de la placa base, integrado en el puente norte (North Bridge).

La memoria DDR ha ido evolucionando a lo largo del tiempo y han aparecido varias generaciones:

- DDR: Su conector dispone de 184 terminales, divididos en dos secciones por una muesca que impide su conexión incorrecta y funcionan a 2,5 voltios. Soportan capacidades máximas por módulo de 1GB y frecuencias de reloj entre 200 y 400 MHz. Su tiempo de acceso se encuentra entre 5 y 7,5 ns y su latencia entre 2,5 y 4 ns.

- DDR2: Tienen un conector de 240 terminales, divididos también en dos secciones por una muesca y funcionan a 1,8 v. Su latencia es superior que en DDR pero alcanza frecuencias de reloj entre 533 y 800 MHz. Su tiempo de acceso se encuentra entre 5 y 6 ns y su latencia entre 4 y 6 ns.

- DDR3: Como DDR2, su conector tiene 240 terminales, divididos también en dos secciones por una muesca, pero funcionan a 1,5 v. Admite rangos de frecuencia entre 800 y 2400 MHz.


- DDR4: En este caso, el conector dispone de 288 terminales, divididos en dos secciones por una muesca, aunque, en este caso, funcionan a 1,2 v. Es la más eficiente en consumo energético y puede funcionar en rangos de frecuencia entre 2.133 y 3333 Mhz (en el momento de escribir este documento).


Es importante saber que la muesca que divide el conector de los módulos de memoria cambia de posición de una generación a otra. Así, no podrán usarse por error en placas base que no los soporten.

Otro dato importante es la nomenclatura, que suele formarse con el tipo de memoria seguido de un guión y un número que indica la frecuencia máxima a la que pueden funcionar. Además, se suelen incluir las letras PC seguidas del número de generación, un guión y la tasa de transferencia máxima, medida en Megabytes.

Así, un módulo de memoria podría quedar definido, por ejemplo, como DDR4 3333 PC4-27700.