

Clase 6

Memorias

6.1 Introducción

Hoy en día, no importa cuánta memoria tenga la computadora, nunca parece ser suficiente. No hace “mucho tiempo” era común escuchar que una PC tuviera 16 MB de memoria, hoy en día esto es impensable. Sin embargo, debido a las exigencias actuales, la mayoría de los sistemas requiere al menos 2 GB para ejecutar las aplicaciones básicas, y de hasta 16 GB o más, si se necesita un desempeño óptimo cuando se utilizan programas de diseño gráfico, edición de video, virtualización e incluso es recomendable para ciertos videojuegos.

El desarrollo más evidente de los equipos de cómputo se ha dado en las últimas tres décadas.

Bill Gates acerca de la capacidad de memoria de las computadoras menciona:

“...en la época de Windows 98 decía que 128MB eran suficientes, para Windows XP dije: 512MB son suficientes, con Windows 7 decía que 2048MB y ahora, con Windows 10, utilizo 32GB...”

La necesidad de memoria aumenta a medida que las computadoras se vuelven más potentes y el software se vuelve más poderoso.

En las clases anteriores hemos tenido en cuenta las memorias internas y externas. En esta clase nos enfocaremos en su análisis, incluyendo las jerarquías de memoria, la estructura interna de la memoria RAM y la memoria cache.

6.2 La memoria RAM

En una computadora, a la memoria principal se le suele llamar RAM (por sus siglas en inglés, que se derivan de random access memory, memoria de acceso aleatorio).

En la actualidad, los productos que se ofrecen en el mercado poseen diferentes capacidades que varían de forma considerable. Día a día, éstas van en aumento de forma instantánea debido principalmente al abaratamiento constante de los chips de memoria, al aumento de la velocidad de acceso y a la creación de nuevos sistemas operativos vastos en el manejo de memorias de capacidad cada vez mayor.

Uno de los valores importantes de la memoria RAM es su capacidad que, tal como mencionamos, hoy en día ronda entre los 8GB y los 32GB.

Otro de los parámetros importantes en una memoria es su velocidad de acceso, la cual mide el tiempo transcurrido desde que la CPU le solicita la información que contiene una celda cualquiera hasta que ésta puede ser leída (o escrita).

Los tiempos de acceso de las memorias actualmente se miden en unidades de nanosegundos ($1 \text{ ns} = 0,000000001 \text{ s}$).

6.2.1 DRAM

La memoria principal (RAM) de prácticamente todos los computadores está compuesta por memorias de tipo DRAM semiconductoras.

La memoria DRAM o "Dynamic RAM" es un tipo de memoria volátil, es decir que una vez que le quitamos la alimentación los datos en ella desaparecen. La palabra dinámica (Dynamic) proviene de que dicho tipo de memoria necesita un "Ciclo de Refresco de Datos" cada "x" cantidad de tiempo para que estos permanezcan (aun cuando tenemos nuestro PC encendido y obviamente la DRAM con alimentación).

Este requerimiento proviene de que la memoria DRAM un capacitor, además de un transistor, para almacenar un bit. Los capacitores pueden mantener la carga por cierta cantidad de tiempo y luego empiezan a descargarse, aún conectados a la fuente de alimentación. Para que todo permanezca como estaba, y los bits almacenados, tienen que ser recargados nuevamente.

Es así que, cada fila de una DRAM. debe ser accedida en un cierto intervalo de tiempo, como por ejemplo 2 milisegundos, o la información de la DRAM puede perderse. Este requerimiento significa que el sistema de memoria ocasionalmente no está disponible porque está enviando una señal para refrescar cada chip.

Es esto lo que caracteriza a la DRAM, siendo una memoria más barata pero más lenta que otros tipos de memoria existente como la SRAM (que se utiliza para la cache).

Sin embargo, la DRAM es el tipo de memoria elegido en la actualidad para la memoria principal y esto tiene su justificación en su principal ventaja:

Ya que la DRAM sólo requiere un transistor por bit de datos, los chips pueden ser mucho más densos, permitiendo almacenar más datos. Esto mantiene menor su tamaño y también las hace más baratas de producir.

6.2.2 La memoria DDR

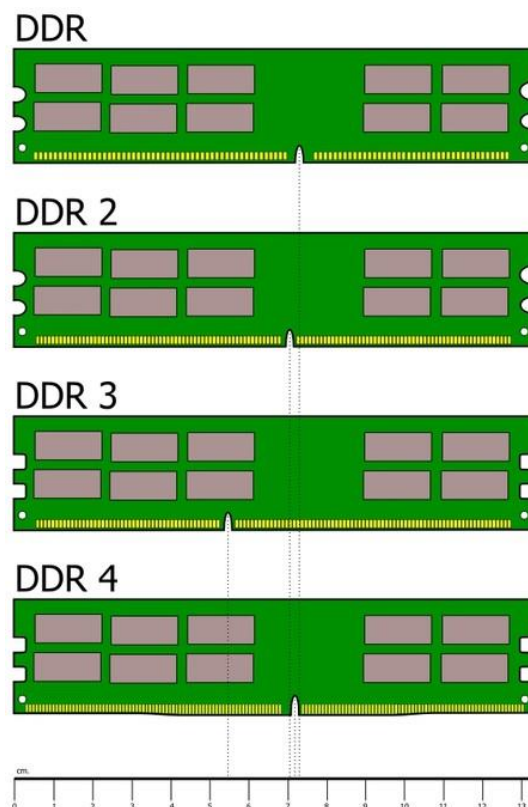
Las memorias de tipo DDR (Double Data Rate) se caracterizan por ser capaces de llevar a cabo dos operaciones en cada ciclo de reloj, a diferencia de las de tipo SDR (Single Data Rate), que solo ejecutaban una operación de lectura o escritura.

Para hacerlo posible los chips DDR se activan dos veces en cada ciclo de la señal de reloj, bien por nivel (alto o bajo), bien por flanco (de subida o bajada).

Las memorias RAM actuales son todas del tipo DDR, en particular actualmente nos encontramos en la versión DDR4.

La memoria RAM DDR4 es la evolución del estándar DDR3 y, por tanto, los módulos de memoria compatibles con esta tecnología son capaces de trabajar a una frecuencia de reloj más alta. También consumen menos y permiten la fabricación de módulos con más capacidad.

Las memorias DDR4 no son compatibles con las DDR3, ni estas con las DDR2 o la DDR; y por lo tanto las placas madre tienen que soportar específicamente el tipo de memoria. Para evitar insertar de manera incorrecta un módulo no correspondiente (que puede quemar el módulo de memoria pues utilizan distinto voltaje) se las diferencia mediante una muesca, como se muestra a continuación:

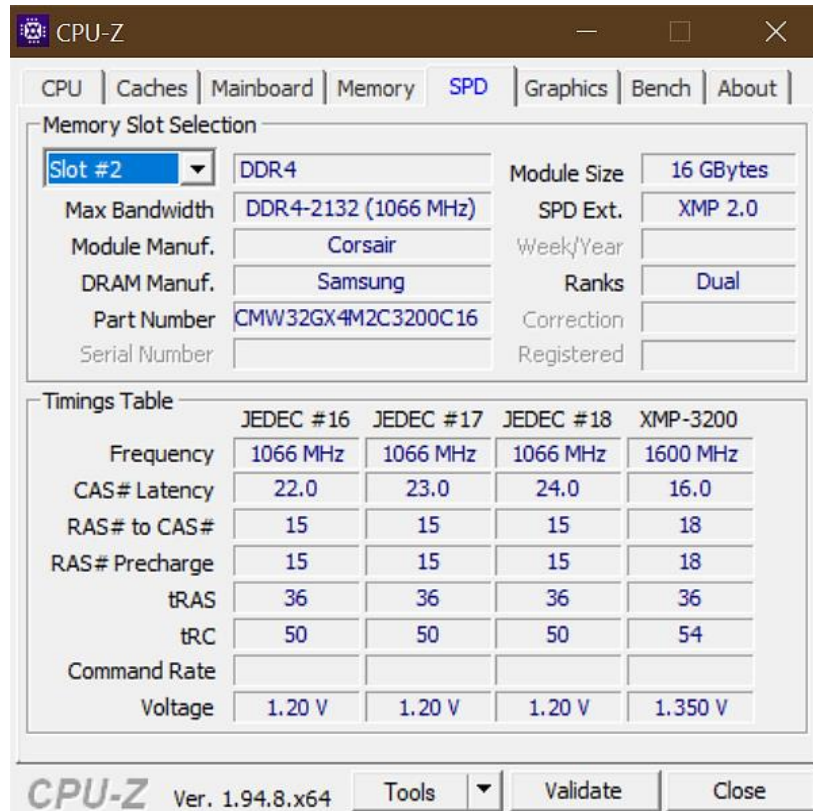


Diferencias en la conexión al motherboard entre las versiones de RAM DDR

6.2.3 La latencia CAS y la frecuencia de la memoria RAM

Cuando vamos a analizar la velocidad de un módulo de memoria RAM tenemos varios valores a tener en cuenta.

Con el software CPU-Z podremos obtener información acerca de nuestros módulos de memoria, por ejemplo:



En este caso observamos que se trata de un módulo de memoria DDR4 de 16GB de capacidad (en la celda Module Size).

Si observamos la sección "Timings Table" nos encontramos con una serie de columnas con información. Estas se refieren a los aspectos de performance o velocidad de la memoria, y la última columna de la derecha es la que representa la mayor velocidad del módulo.

En ese caso tenemos varios parámetros, de los cuales vamos a analizar los dos "más importantes", la frecuencia y la latencia CAS:

Frecuency (Frecuencia): La frecuencia de una memoria RAM indica la velocidad de reloj con la que trabaja ésta. En este caso indica 1600Mhz (Megahertz, millones de ciclos por segundo), que al ser una memoria DDR se traducen en 3200Mhz.

En otras palabras, el número de la frecuencia indica cuántos ciclos puede manejar la RAM cada segundo. En el ejemplo, la RAM ejecutará 3200 millones de ciclos por segundo.

Cada lectura y escritura de datos se hará en una cantidad determinadas de ciclos de reloj.

Por lo tanto, a mayor frecuencia de la memoria, se podrán realizar mayor cantidad de operaciones por segundo, lo que beneficiará al usuario, porque la ejecución de los programas será más fluida.

CAS Latencia (latencia CAS): La latencia es el tiempo que transcurre entre una petición y su respuesta. En este caso, la palabra latencia hace referencia al retraso o tiempo que tarda el procesador en conseguir acceder a un dato en la memoria RAM. Por lo tanto se desea que la latencia CAS sea lo menor posible.

La latencia CAS (CL) se mide en cantidad de ciclos de reloj de la RAM, y por lo tanto no puede considerarse de forma aislada, y hay una fórmula que convierte el tiempo CL en nanosegundos y se basa en la velocidad de transferencia de la RAM.

En una memoria RAM cuya información técnica esté escrita como RAM DDR4 3200 MHz CL 16, CL 16 indica la latencia de la misma. Los números de la latencia CAS muestran cuántos ciclos de reloj tardará la memoria RAM en responder a una petición de datos de parte del procesador.

El cálculo a realizar para saber cuánto tiempo se tardará en acceder a una palabra de memoria (64 bit) nos devuelve dicho tiempo en nanosegundos, y es el siguiente:

$$\frac{CL}{Frecuencia} \cdot 2000$$

En el ejemplo nos quedaría:

$$\frac{16}{3200} \cdot 2000 = 10 \text{ ns}$$

Dicho valor es la frecuencia real del módulo de memoria.

Como vemos para bajar la latencia de la memoria podemos aumentar la frecuencia o disminuir la latencia CAS.

Por ejemplo, un módulo con frecuencia de 3000Mhz y latencia de 14ns tendrá una latencia real de:

$$\frac{14}{3000} \cdot 2000 = 9,33 \text{ ns}$$

Y por lo tanto este módulo de 3000Mhz tardará menos en acceder a un dato que la el módulo de 3200Mhz.

Sin embargo obtener una baja latencia CAS no es sencillo, los módulos de baja latencia suelen ser caros, por lo cual los fabricantes

comúnmente optan por tratar de aumentar la frecuencia de la memoria.

Pronto tendremos disponibles las memorias de tipo DDR5, con frecuencias de hasta 6400Mhz.

6.3 Memoria ROM

La memoria ROM comúnmente se encuentra grabada en chips por el fabricante de hardware y tradicionalmente se la conoce con el nombre de firmware. También la encontramos en las PC's en el tipo de memoria en el que está grabado el BIOS de la placa madre o motherboard.

En un principio esta memoria era de sólo lectura (Read Only Memory), por lo que no podía ser modificada. Hoy en día esta memoria puede ser modificada, generalmente para ser actualizada, pero dicho proceso no se realiza fácilmente, para evitar modificaciones accidentales. Además, la memoria ROM no se altera por cortes de corriente.

Como dijimos, en esta memoria se almacena el BIOS, que contiene instrucciones para el inicio del sistema, su configuración y los valores correspondientes a las rutinas de arranque.

Estos nuevos tipos de circuitos reciben el nombre genérico de ROM (por sus siglas en inglés, que significan read only memory, memoria de sólo lectura) y se fabrican en varias configuraciones (PROM, EPROM, EEPROM, flash), de acuerdo con la mayor o menor eficacia para regrabarlas.

Actualmente la memoria flash es la elegida para implementar la memoria ROM. Dicha memoria permite la lectura y escritura de múltiples posiciones de memoria en la misma operación. Gracias a esto, la tecnología flash, permite velocidades de funcionamiento muy superiores frente a la antigua tecnología EEPROM, que sólo permitía actuar sobre una única celda de memoria en cada operación de programación.

Se trata de la tecnología empleada en las memorias USB o pendrive, unidades de estado sólido, firmware y las actuales BIOS.

6.3.1 El firmware

En términos generales, los programas que conviven en una memoria tipo ROM se conocen en inglés como firmware, y representan un intermedio entre los programas normales (software) y los circuitos electrónicos (hardware).

El firmware de un componente es un tipo de software que es capaz de proporcionar un control de bajo nivel específico para el componente en cuestión. Este software puede ser capaz de proporcionar un entorno de operación para las funciones más complejas del componente, o bien puede actuar como el propio sistema operativo interno del componente. En definitiva, el firmware de un componente es el encargado de decirle al mismo cómo ha de funcionar o cómo comportarse cuando funciona en consonancia con otros dispositivos.

Sin embargo, la separación entre hardware y software ha ido desapareciendo poco a poco debido, sobre todo, a la producción de nuevas tecnologías, conocidas como hardware programable, que reciben el nombre genérico, las cuales consisten en circuitos configurables por el usuario y que permiten una flexibilidad que antes no existía.

Sin ir más lejos, hoy en día nos encontramos con teclados, mouse, discos rígidos, gamepads, auriculares, Smart TV's, cuyo firmware es actualizado de manera sencilla, generalmente para incluir nuevas funcionalidades o solucionar errores.

Incluso hay dispositivos que permiten guardar las configuraciones en el firmware del dispositivo, como por ejemplo el mouse Logitech g502, entre tantos otros.



El mouse Logitech G502 dispone de un modo de memoria integrada (desactivado por defecto) que permite almacenar perfiles de configuración en la memoria ROM del dispositivo, para utilizar dicha configuración en cualquier equipo sin necesidad de instalar software adicional.

Un claro ejemplo de lo que es un firmware avanzado es la BIOS de nuestras placas base. Gracias a ellas, la placa puede hacer funcionar nuestro PC, pero también gracias a ella, podemos configurar muchos aspectos de nuestros sistemas, de manera diferente a lo que sería la configuración de serie del mismo.



Ejemplos de dos tipos de BIOS de PC

Como contrapunto, tenemos softwares más sencillos, como el que se usa para controlar nuestras unidades de almacenamiento. Generalmente, estos softwares no suelen ser modificables por los usuarios y los suelen proporcionar los fabricantes de los propios dispositivos.

Incluso los discos de estado sólido se actualizan en su firmware, por ejemplo, el disco SSD Kingston V300 requirió una actualización de firmware por un error que indicaba que al disco le quedaba poco tiempo de vida útil cuando en realidad era prácticamente nuevo.



Como norma, el firmware se suele almacenar en un chip de memoria no volátil. En la actualidad, se usan las memorias NAND Flash por su facilidad a la hora de actualizar el software que contienen.

6.4 La memoria cache

Generalmente cuando se hace referencia a la memoria cache se está hablando del tipo de memoria temporal y de muy rápido acceso que se encuentra en la CPU.

La memoria caché es considerada como un sistema de almacenamiento de alta velocidad, y su capacidad o cantidad es relativamente pequeña (normalmente del orden de los MegaBytes) debido al alto costo y al mayor espacio que ocupa comparada con la RAM.

La cache se ubica regularmente dentro del mismo circuito integrado de la CPU y se utiliza para almacenar datos que se intercambian entre la CPU y la memoria principal.

Dado que el procesador es mucho más rápido que la memoria, se trata de que el mismo no quede “esperando” a la RAM. Por ese motivo principal existe la memoria caché en las CPU. Es la memoria más rápida del sistema y se utiliza para acelerar el acceso a memoria de nuestro procesador, para que el procesador no tenga que acceder con tanta frecuencia a la RAM del sistema.

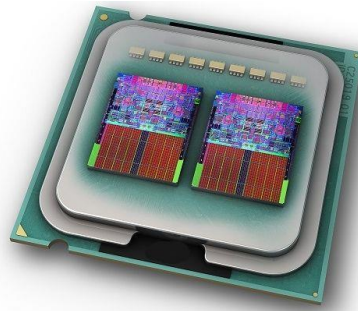
En esta memoria se almacenan los datos e instrucciones más utilizados por el sistema para acceder a ellos de una manera más rápida que si se accediese a la RAM.

La utilización de memoria cache tiene dos ventajas:

- La principal ventaja es reducir la latencia, ya que se descarga información continuamente desde la memoria RAM, con anticipación, para que el procesador siempre esté nutrido de datos y pueda maximizar su rendimiento.
- La segunda ventaja es que la frecuencia de la cache puede acercarse mucho a la que tiene el procesador. Menor latencia y más velocidad implica mayor rendimiento interno. La memoria caché se diseñó principalmente para resolver la contradicción que implica necesitar memoria RAM, pero no depender de su velocidad, puesto que la CPU consigue muchos mayores ciclos por segundo de forma interna, logrando velocidades de lectura y escritura que una RAM no puede alcanzar.

La memoria cache es una memoria del tipo SRAM (static RAM), y a diferencia de la DRAM se caracteriza por no utilizar capacitores para almacenar los bits, pero sí utiliza más transistores, con lo cual la memoria se hace más cara dado que puede almacenar menos información en el mismo espacio. Además, es más rápida porque no

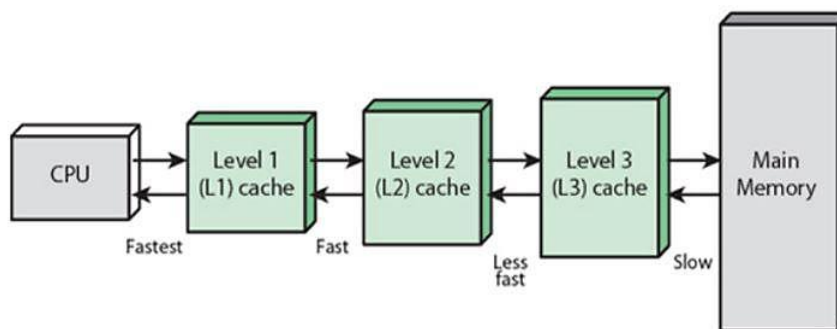
necesita refrescar la información continuamente como en el caso de la DRAM.



La memoria cache se encuentra dentro del circuito integrado de la CPU

6.4.1 Estructura de la memoria cache

Con el tiempo, la estructura de la memoria cache fue evolucionando y actualmente encontramos tres tipos diferentes de memoria caché que se encuentran incorporada en el procesador: L1, L2 y L3. En el siguiente esquema, por simplicidad, se muestran por fuera de la CPU:



Cada una de estas memorias tiene diferentes tamaños y velocidades, siendo la L1 la de menor tamaño y velocidad igual a la de la CPU, mientras que la L3 es la de mayor tamaño pero menor velocidad. Sin embargo, cualquiera de estas caches es más rápidas que la memoria RAM.

Básicamente el funcionamiento de las memorias cache es el siguiente:

- Cuando se ejecuta un programa en el PC, los datos van a la memoria RAM, y desde ésta a la caché L3, luego a la L2 y finalmente a la L1.
- Mientras el programa esté en ejecución, el procesador irá a buscar la información que necesita el procesador primero a la memoria más rápida, la caché L1.
- Si no la encuentra ahí, se irá a la L2.
- Si no la encuentra en la cache L2, buscará en la L3

- Y si no está en ninguna, entonces la buscará en la RAM.

6.4.2 Principios de localidad

El uso de la memoria caché se sustenta en dos principios o propiedades que exhiben los programas:

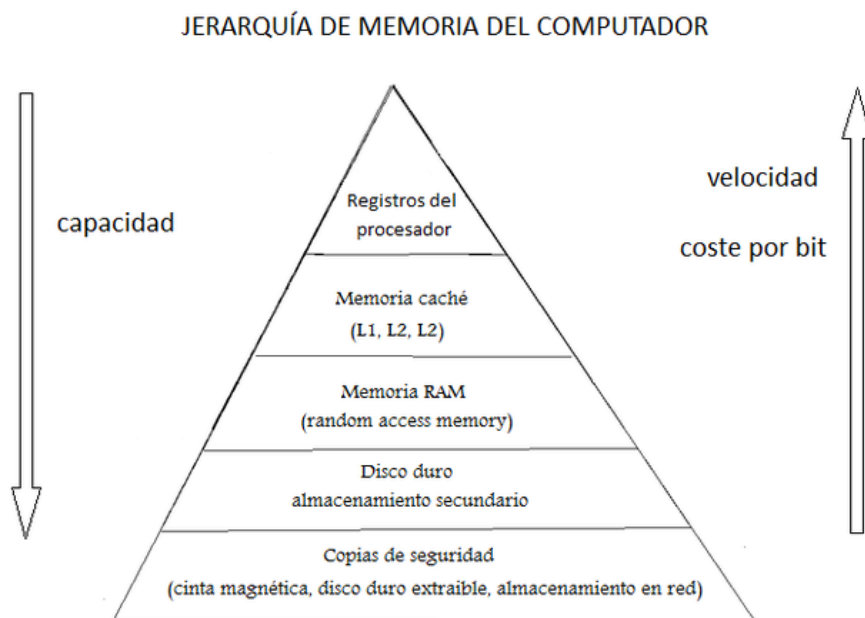
Localidad temporal (localidad en el tiempo): Si se accede a una dirección de memoria, esta dirección será utilizada de nuevo en un corto intervalo de tiempo.

Localidad espacial (localidad en el espacio): Si se accede a una dirección de memoria, las direcciones de memoria que se encuentran próximas a ella serán utilizadas en un corto intervalo de tiempo.

6.5 La jerarquía de memoria

Desde los primeros días de la informática, tanto los programadores como los usuarios en general han deseado cantidades ilimitadas de memoria rápida. Sin embargo, como hemos tratado de desarrollar en esta clase, esto no es posible, al menos no a bajo costo y sin aumentar el espacio físico que ocupan las memorias.

La jerarquía de memorias establece un orden, una organización en niveles, que tienen las memorias de una computadora.



Como podemos observar, en la cúspide de la pirámide se encuentran los registros del procesador (registro de instrucciones, contador de

programa, acumulador de la UAL, registros de datos, etc.). Estos son los más veloces.

Luego nos encontramos con la cache L1, tan rápida como el procesador, pero de poca capacidad, y cara de producir.

De esa manera observamos que a medida que descendemos disminuye la velocidad de la memoria, pero aumenta su capacidad de almacenamiento. También disminuye el costo por cada bit de almacenamiento.

En la jerarquía inferior a la memoria RAM nos encontramos con la memoria secundaria, es decir, el disco rígido o el disco de estado sólido SSD.

Finalmente tenemos las unidades extraíbles, como ser los pendrives.

Es así que los niveles que componen la jerarquía de memoria habitualmente son:

Nivel 0: Registros del microprocesador

Nivel 1: Memoria caché

Nivel 2: Memoria primaria (RAM)

Nivel 3: Memorias secundarias: Disco duro o disco SSD

Nivel 4: Memorias extraíbles (consideradas las más lentas)

Podríamos incorporar un quinto nivel, el de las redes, incluyendo el almacenamiento en red LAN, WLAN o bien el almacenamiento en la nube.



El disco de estado sólido Samsung EVO 980 PRO es actualmente uno de los exponentes más rápidos de la memoria secundaria de una computadora.