Tema 15: Memorias

Objetivos:

- Introducción.
- Características de las memorias.
- Jerarquía de las memorias.

Introducción

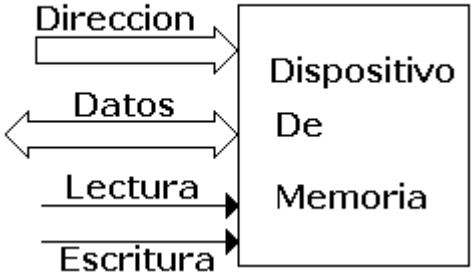
Breve Historia

- En la década de los años 30 se empleaban tarjetas perforadas como memorias y posteriormente se emplearon relés electromagnéticos.
- El computador ENIAC (1946) empleaba válvulas electrónicas de vacio para construir biestables.
- A comienzos de los 50 se usaron las líneas de retardo de mercurio con 1 Kbit por línea.
- El UNIVAC I (1951) fue la primera unidad comercial de banda magnética.
- El IBM 650 (1954) usó por primera vez como memoria un tambor magnético con una capacidad de 120 Kbits.
- El MIT (1953) dispuso de la primera memoria operativa de ferrita.
- Fue IBM en 1968 quien diseñó la primera memoria basada en semiconductores.

Introducción

- La memoria es un bloque fundamental de cualquier sistema computador cuya misión consiste en almacenar datos e instrucciones.
- Básicamente se pueden realizar dos operaciones básicas con las memorias: escritura o almacenamiento y lectura.
  - –La operación de escritura consiste en grabar la información deseada en una cierta dirección que debe ser indicada.
  - –La operación de lectura consiste en recuperar la información existente en una cierta dirección que debe ser indicada.
- Un dispositivo de memoria se puede asemejar inicialmente a una caja negra a la que se indica:
  - – La operación que se desea realizar sobre ella (leer/escribir).
  - – Dirección sobre la que se desea realizar dicha operación.
  - – Conjunto de líneas para colocar los datos.

Habitualmente la línea de lectura y escritura se reduce a una sola indicando de forma binaria la operación que se desea realizar



Introducción

- Idealmente una memoria debe cumplir:
  - Tener una capacidad ilimitada.
  - Servir o recoger los datos a la mayor velocidad posible.
  - Todo ello al menor coste posible.

•Desgraciadamente es imposible obtener velocidad y capacidad máximas por razones de volumen y de coste.

•Esto obliga a organizar la memoria empleando diferentes tipos de memoria según su localización dentro del sistema:

–Memoria Interna del procesador: Formada normalmente por unos pocos registros de poca capacidad pero de muy rápido acceso por parte de la unidad procesadora.

–Memoria principal: Es una memoria relativamente grande y rápida que se utiliza para el almacenamiento de las instrucciones y datos en ejecución en ese momento. En la actualidad suelen ser memorias basadas en semiconductores

–Memoria externa o secundaria: Se trata de dispositivos periféricos cuyo soporte físico suele ser : discos magnéticos, discos ópticos, cintas …etc…, este tipo de memoria suele ser de gran capacidad comparada con las anteriores a costa de lentitud y se utiliza para almacenar datos y programas que no se estén empleando en ese momento

Introducción

CARACTERÍSTICAS DE LA MEMORIA

1. **Tamaño o capacidad de almacenamiento.**
2. **Tiempo que se tarda en acceder a la información (VELOCIDAD).**
   - Tiempo de acceso (T_A).
   - Tiempo de ciclo de memoria (T_C).
   - Frecuencia de acceso (F_A).
3. **Coste por bit.**
4. **Alterabilidad.**
5. **Permanencia de la información.**
   - Lectura destructiva.
   - Volatilidad
   - Almacenamiento Estático/Dinámico
6. **Métodos de acceso.**
   - Acceso Aleatorio
   - Acceso secuencial
   - Acceso Directo
   - Acceso Asociativo

Características de la memoria

1-.Tamaño o capacidad de almacenamiento

- Es la cantidad de información binaria que puede almacenar un dispositivo de memoria y depende del número de posiciones o direcciones que disponga así como de la longitud de cada una de ellas
  - 1 Byte = 8 bits.
  - 1 Kilobyte (1 KB)= 2^{10} bytes = 1024 bytes.
  - 1 Megabyte (1MB) = 2^{10} K = 2^{20} bytes.
  - 1 Gigabyte (1GB) = 2^{30} M = 2^{30} bytes.
  - 1 Terabyte (1TB) = 2^{40} G = 2^{40} bytes.
  - 1 Petabyte (1TB) = 2^{50} T = 2^{50} bytes.

Características de la memoria

- La nomenclatura utilizada consiste en indicar en primer lugar el número de posiciones o direcciones de memoria y en segundo lugar el tamaño de cada una de esas direcciones:
  - 256 X 16 indica que la memoria dispone de 256 posiciones y cada una de esas posiciones es capaz de almacenar 16 bits de información.
  - 2Kb = 2*1024 bits = 2048 bits.
  - 3KB = 3*1024 bits = 3072 bits.
  - 8 KB = 8*1024 * 8 = 65536 bits.
  - 256X16 = 4096 bits.
- También se deben tener en cuenta las siguientes agrupaciones:
  - nibble = 4 bits.
  - Byte = 8 bits.
  - Palabra = 16 bits.
  - Doble palabra = 32 bits.
  - Cuádruple palabra = 64 bits.

Características de la memoria

2. Tiempo que se tarda en acceder a la información (VELOCIDAD).

La velocidad de un dispositivo de memoria se evalúa mediante tres parámetros:

- Tiempo de acceso (T_A).
- Tiempo de ciclo de memoria (T_C).
- Frecuencia de acceso (F_A).
- **Tiempo de acceso:** El tiempo de acceso de lectura representa el tiempo que ha de pasar desde que se solicita una información hasta que está disponible dicha información. De forma análoga se define el tiempo de acceso para la escritura en memoria.

  En memorias secudarias éste tiempo está fuertemente influido por el tiempo de posicionado de cabezas

Tipo de Memoria	Tiempo de acceso (t_A)
Semiconductor Bipolar	10^{-8} seg
Semiconductor MOS	10^{-7} seg
Discos magnéticos	10^{-3} seg
Discos ópticos	10^{-2} seg

Características de la memoria

- **Tiempo de ciclo de memoria (T_C):** Es el tiempo mínimo que debe de transcurrir entre dos accesos consecutivos a memoria. Esta característica se refiere principalmente a memorias en las que tras cada acceso se debe restaurar la información debido a que ésta se va degradando, en estos casos normalmente $T_C > T_A$, y el tiempo de restablecimiento T_{RES} será la diferencia entre ambos.
- **Frecuencia de acceso (F_A):** Es el número de accesos a memoria que se pueden realizar en un segundo $$f_A = \frac{1}{T_C}$$

3. **Coste por bit:** Es el precio que se paga por cada unidad mínima de información (bit) en un determinado dispositivo de memoria.

4. **Alterabilidad:** Esta característica hace referencia a si se puede alterar o no elñ contenido de una .

   - Las memorias cuyo contenido no puede ser modificado se denominan memorias ROM (**R**ead **O**nly **M**emory).
   - Las memorias cuyo contenido sí puede ser modificado se denominan memorias RWM (**R**ead **W**rite **M**emory), Las memorias RAM serán de este tipo

Características de la memoria

–Algunos tipos de memoria ROM son programables por el usuario, en ese caso se denominan PROM (**P**rogramable **ROM**) y otras PROM pueden ser borradas y vueltas a programar con otro contenido denominadas EPROM (**E**rasable **PROM**)

–El borrado se puede hacer de forma eléctrica (EEPROM) o haciendo incidir una luz ultravioleta (UVEPROM)

–EL tiempo de acceso de escritura de las PROM es muy superior al de lectura, además de que el proceso de borrado en muchos casos obliga a extraer el dispositivo o a añadir circuitería adicional , siendo estas las principales diferencias de éste tipo de memoria con las RAM.

Características de la memoria

5. **Permanencia de la información:** Existen tres parámetros de las memorias que pueden redundar en la destrucción de la información que almacenan: Lectura destructiva, Volatilidad, Almacenamiento Estático/Dinámico.
   - **Lectura destructiva :** Hay memorias en las que una operación de lectura ocasiona la destrucción de la información leída por lo que para evitar su pérdida hay que volver a reescribirlas inmediatamente después de haber sido leídas.
     - Memorias de lectura destructiva (DRO : Destructive ReadOut).
     - Memorias de lectura no destructiva (NDRO).
   - **Volatilidad :** Este parámetro hace referencia a la pérdida de información al cesar el suministro de electricidad en el dispositivo de memoria.
     - Memoria volátil : Aquella que pierde su información al cesar el suministro de corriente eléctrica.
   - **Almacenamiento Estático/Dinámico :**En las memorias dinámicas la información se va degradando por lo que necesitan ser refrescadas de forma periódica mientras que en las estáticas no se produce este problema
     - Por otro lado, en las dinámicas se consigue mayor densidad de almacenamiento y menor consumo energético

Características de la memoria

6. **Métodos de acceso:** En cualquier operación de lectura o escritura, en primer lugar es necesario localizar la dirección a la que se desea acceder
   - **Acceso Aleatorio :** Cuando se puede acceder a su información en cualquier orden.
   - El tiempo de acceso independiente de la posición a la que se desea acceder.

     A este tipo de memorias se les denomina RAM (Random Access Memory) y en ellas cada posición tiene una única dirección.

     Ejemplo : Memoria principal del computador.
   - **Acceso Secuencial :** El acceso a la información se realiza de forma secuencial. Partiendo de la posición actual del cabezal de lectura/escritura, se va avanzando por la memoria hasta llegar a la información deseada

     El tiempo de acceso no es independiente de la posición a la que se desea acceder.

     A este tipo de memorias se les denomina SAM (Sequential Access Memory).

     Ejemplo : Cinta magnética

Características de la memoria

- **Acceso Directo :** La información está dividida en bloques a los que se accede de forma aleatoria y dentro de cada bloque, el acceso a la información se hace de forma secuencial.

  El tiempo de acceso no es independiente de la posición a la que se desea acceder.

  Ejemplo : Unidades de disco.
- **Acceso Asociativo :** Se trata de memorias de acceso aleatorio, pero a diferencia de éstas, en lugar de realizar el acceso a la dirección especificada, se hace proporcionando el valor de un campo a buscar.

  En este tipo de memoria, cada posición consta de dos campos: Etiqueta + Dato y para buscar una información se indica la etiqueta a buscar, obteniendo como salida el campo Dato.

  A este tipo de memoria se le denominan CAM (Memorias por acceso por contenido)

  El tiempo de acceso es independiente de la posición a la que se desea acceder.

Jerarquía de la memoria

•Idealmente la memoria de un procesador sería la más rápida, de mayor capacidad y menor coste.

•En la realidad la mayor rapidez de una memoria implica mayor coste y menor capacidad.

•También una memoria grande lleva consigo una velocidad de acceso reducida.

•Se debe llegar a un compromiso entre los tres Parámetros coste – Velocidad – Capacidad.

•En la práctica lo que se hace es emplear una jerarquía de memorias consistente en distribuir la información manejar por el procesador en diversos tipos de memoria tal que:

– Cerca del procesador tendremos memorias de rápido acceso a costa de perder capacidad.

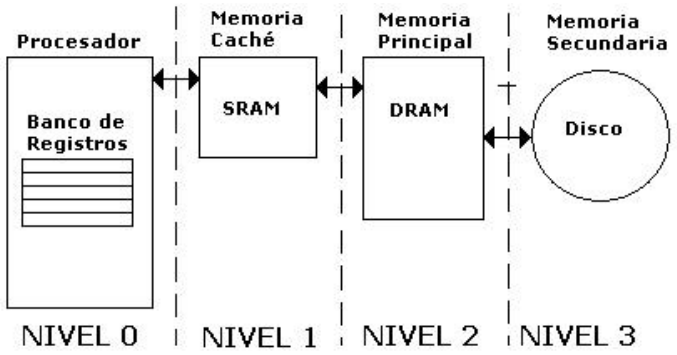
– La memoria de gran capacidad (Lenta) la separamos del procesador conteniendo la totalidad del código y de los datos

–La memoria cercana se carga de forma dinámica con la información que va a precisar el procesador y que la recibe de la memoria lejana

Jerarquía de la memoria

•Un posible esquema jerárquico podría ser el siguiente:

- A medida que pasamos de un nivel de jerarquía a otro La capacidad aumenta.
  1. La velocidad de acceso disminuye,
  2. La frecuencia con la que la CPU accede a esa memoria disminuye.
  3. El coste por unidad de información disminuye.



•Si el procesador no encuentra la información que necesita en el banco de registros, explora la caché, luego la principal y finalmente la secundaria aumentando en cada paso el tiempo de acceso y ralentizando el sistema

Jerarquía de la memoria

• Para conseguir un sistema lo más óptimo posible, debemos procurar situar los datos e instrucciones de uso más frecuente o inmediato en los niveles de jerarquía mas cercanos al procesador dejando el resto en los mas alejados que coinciden con los de mayor capacidad.

• Esto supone:

– ***Predecir*** que datos/instrucciones van a ser usadas en un futuro inmediato para ser trasladadas al nivel jerárquico más cercano al procesador.

– ***Predecir*** que datos/instrucciones presentes en dicho nivel ya no son necesarios allí.

•Para dichas predicciones se utiliza el “PRINCIPIO DE LOCALIDAD”. Este principio establece que, en un determinado instante, sólo se accede a una parte relativamente pequeña de todo el programa en curso.

•En cualquier programa se pueden encontrar dos tipos de localidad.

–***Localidad Temporal*** :”*Si un elemento de información de la memoria (código o dato) ha sido accedido o referenciado, volverá a ser referenciado pronto”.* Un ejemplo claro de esto son los bucles habitualmente utilizados en programación.

–***Localidad Espacial*** :*”Cuando la CPU acaba de referenciar un elemento, tiende a referenciar, seguidamente, elementos de direcciones próximas”*

•Si lo programas no siguen estas simples reglas de localidad, no se puede aplicar la jerarquía de la memoria lo que ralentizará el sistema.

Jerarquía de la memoria

- Todo el trasiego de información entre los diferentes niveles jerárquicos de memoria se controla mediante programas residentes del Sistema Operativo (S.O.) y que están basados en unos algoritmos que siguen las reglas de localidad.
- En la arquitectura de un computador, generalmente pueden encontrarse cuatro niveles jerárquicos en la memoria:
  1. ***Nivel 0 (Superior):*** Registros internos de procesador.
  2. ***Nivel 1:*** Memoria Caché.
  3. ***Nivel 2 :*** Memoria Inferior.
  4. ***Nivel 3 (Inferior):*** Memoria secundaria.

1. **Registros internos del procesador :** Se trata de un conjunto de registros agrupados en un banco de registros dentro del propio procesador .

   Su número es reducido pero de acceso muy rápido.

2. **Memoria Caché :** Se trata de dispositivos semiconductores de pequeña capacidad pero rápidos.

   Generalmente se tratan de memorias SRAM (RAM Estáticas) fabricadas con tecnología bipolar y normalmente su precio es elevado.

Jerarquía de la memoria

3. **Memoria Principal :** Es una memoria relativamente grande y veloz, habitualmente construida con tecnología MOS y de tipo dinámico (DRAM) por lo que requerirá de refresco para su correcto funcionamiento
4. **Memoria Secundaria o externa :** Se trata de una memoria lenta ya que habitualmente son elementos periféricos de entrada/salida de gran capacidad y muy bajo coste por bit. En su gran mayoría se trata de dispositivos magnéticos u ópticos

Terminología referente a las Jerarquías de la memoria

1. **Nivel jerárquico superior :** Es el nivel más cercano al procesador y por lo tanto es el que debe contener la memoria más rápida y pequeña.
2. **Nivel jerárquico inferior :** Es el más alejado del procesador conteniendo la memoria más grande pero más lenta.
3. **Acierto (Hit) :** Se produce cuando la información requerida por el procesador se encuentra en el nivel superior.
4. **Fallo (Miss) :** Se produce cuando la información requerida por el procesador NO se encuentra en el nivel superior, en cuyo caso se deberá acceder a niveles inferiores para trasladarla al nivel superior.
5. **Tasa de aciertos (P_A):** Es la cantidad de accesos a memoria que han obtenido un acierto.
6. **Tasa de fallos :** Es la cantidad de accesos a memoria que NO han obtenido un acierto.
7. **Tiempo de acierto (T_A) :** Es el tiempo necesario para acceder al nivel superior de la memoria. En él también está incluido el tiempo preciso para determinar si se ha producido un acierto o un fallo.

Terminología referente a las Jerarquías de la memoria

8. **Penalización de fallos :** Es el tiempo que le cuesta al procesador acceder a un dato de un nivel jerárquico inferior y se divide en el tiempo de búsqueda del dato en el nivel inferior, el tiempo necesarios para trasladarlo al nivel jerárquico superior y la entrega al procesador.
9. **Tiempo de acceso medio ($T_{A\ medio}$) :** El tiempo de acceso medio se calcula teniendo en cuenta que en caso de acierto, el tiempo empleado es el correspondiente al tiempo de acceso al nivel superior, mientras que en los fallos se tarda la suma de los tiempos de acceso al nivel superior e inferior.

Terminología referente a las Jerarquías de la memoria

- **Ejemplo :** Se dispone de un sistema con 2 niveles jerárquicos de memoria, en el nivel superior disponemos de una caché de 256 KB y T_{A1} = 15 ns, En el inferior tenemos la memoria principal de 256 MB con T_{A2} = 150 ns.

  Si la tasa de aciertos en la caché es del 90% calcular el tiempo de acceso medio.

$$T_{A_{medio}} = P_A \bullet T_{A_1} + (1 - P_A) \bullet (T_{A_1} + T_{A_2}) =$$

$$= 0.9 \bullet 15 + 0.1 \bullet 165 = 13.5 + 16.5 = 30 \text{ nsg.}$$