

Primer-Parcial-2oCuatri-Temas6-y...



luksiim



Estructura de Computadores



2º Grado en Ingeniería Informática



Escuela Técnica Superior de Ingeniería Informática
Universidad Politécnica de Valencia

1/6

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

ING BANK NV se encuentra adherido al Sistema de Garantía de Depósitos Holandés con una garantía de hasta 100.000 euros por depositante. Consulta más información en [ing.es](https://www.ing.es)

“Mi imperio romano
es el Bizum que me
debe mi amigo el rata.”



Cuenta NoCuenta

Perfecta para hacer todos
tus Bizums... ¡y pedirlos!

Cuéntame más



Tu aventura en el agua comienza aquí:

 Ocean Republik

Primer Parcial – Segundo Cuatrimestre

Tema 6 - Jerarquía de memoria

Concepto de Jerarquía de Memoria

Principio de Localidad

Es un principio empírico, derivado del diseño y comportamiento de los programas

Localidad Temporal

Tendencia a volver a referenciar el mismo dato o instrucción

Localidad Espacial

Tendencia a referenciar palabras en direcciones próximas, justifica la organización en bloques o páginas que actúan como unidad de transferencia

La Memoria Cache

Es una memoria rápida y de poca capacidad, situada entre la memoria principal y el procesador

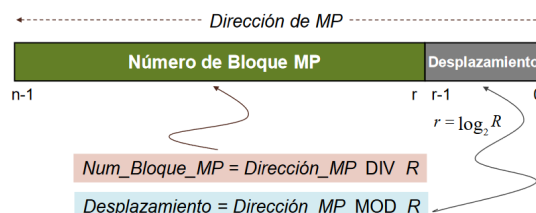
Primero se busca siempre la información en la memoria cache (SRAM) y, si no se encuentra aquí se accede a la memoria principal (DRAM). La unidad de transferencia entre la cache y la memoria principal es el bloque

Cálculo del Número de Bloque

Espacio de direccionamiento de MP = 2^n bytes

Tamaño de bloque: $R = 2^r$ bytes

Nº total de bloques = Tamaño espacio direccionamiento / tamaño bloque



SURF • WINDSURF • WINGFOIL • KAYAK • PADDLE SURF YOGA

¡Apúntate por la Web!

¡Apúntate por la App!



WUOLAH

Clasificación de los accesos

Al hacer un acceso, la CPU puede encontrar o no la información:

- **Acierto (hit):** la información se encuentra en la cache
- **Fallo (miss):** la información no se encuentra en la cache

Tasa de aciertos y Tiempo de acceso a datos

La tasa de aciertos (H) es la proporción de aciertos

$$H = \text{número de aciertos} / \text{número de accesos}$$

$$H = (\text{número de accesos} - \text{número de fallos}) / \text{número de accesos}$$

Tasa de fallos: $1 - H$ (proporción de fallos)

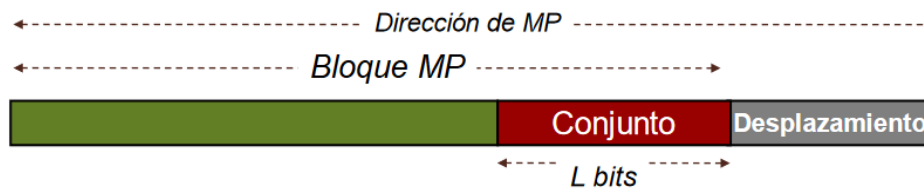
Tiempo medio de acceso a los datos:

$$T_{\text{medio}} = H * T_{\text{acierto}} + (1 - H) * T_{\text{fallo}}$$

Funciones de correspondencia

Como objetivo tiene decidir en qué línea de la cache debe almacenarse cierto bloque de memoria principal.

A las clases o recipientes de la MC se les denomina conjuntos.



Para calcular el número de conjuntos: 2^L

Tipos de correspondencia

Directa

Un bloque se almacena siempre en la misma línea (una línea solo en cada conjunto)

Etiqueta – Línea - Desplazamiento

Asociativa por conjuntos de n vías

Un bloque se almacena siempre en cualquiera de las n líneas (vías) contenidas en el conjunto

Etiqueta – Conjunto – Desplazamiento

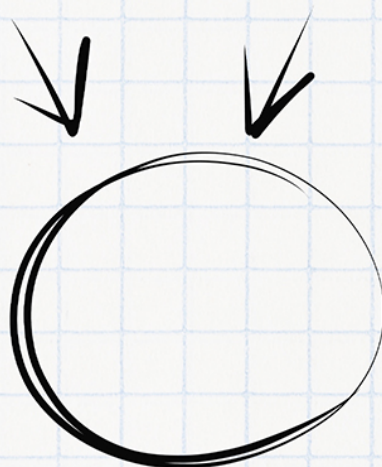
$$\text{Número de conjuntos} = \text{líneas en la cache} / \text{número de vías}$$

Imagínate aprobando el examen

Necesitas tiempo y concentración

Planes	 PLAN TURBO	 PLAN PRO	 PLAN PRO+
 Descargas sin publi al mes	10 	40 	80 
 Elimina el video entre descargas			
 Descarga carpetas			
 Descarga archivos grandes			
 Visualiza apuntes online sin publi			
 Elimina toda la publi web			
 Precios Anual <input type="checkbox"/>	0,99 € / mes	3,99 € / mes	7,99 € / mes

Ahora que puedes conseguirlo,
¿Qué nota vas a sacar?



WUOLAH

Estructura de Computadores



Comparte estos flyers en tu clase y consigue más dinero y recompensas



Banco de apuntes de la

- 1** Imprime esta hoja
- 2** Recorta por la mitad
- 3** Coloca en un lugar visible para que tus compis puedan escanar y acceder a apuntes

- 4** Llévate dinero por cada descarga de los documentos descargados a través de tu QR



Totalmente asociativa

Un bloque se puede almacenar en cualquier línea de la memoria cache
Etiqueta – Desplazamiento

Tipos de fallos (exclusivos)

Fallos de arranque

Se producen la primera vez que se referencia un bloque (no ha estado aun en la cache)

Fallos de capacidad

Se producen debido a la limitación del tamaño de la cache

Fallos de conflicto (o colisión)

Se producen cuando el conjunto esta lleno pero la cache no, solo en caches de correspondencia directa y asociativa por conjuntos,

Políticas de Lectura y Escritura

Lectura

Acierto de Lectura

El dato está en la cache, por lo tanto, se lleva el dato a la CPU

Fallo de Lectura

El dato no se encuentra en la cache, por lo que hay que traer el bloque desde la memoria principal. La memoria principal al leer el bloque ofrece primero la palabra que causo el fallo (critical word first)

Escritura

Acierto de escritura

Escritura directa (write through): se escribe el dato en la cache y en la memoria principal.

Escritura posterior (write back): Se escribe en la cache pero no en la memoria principal, cuando se reemplaza el bloque se escribe en el nivel inferior. Se requiere un bit adicional (dirty bit)

Fallo de escritura

Ubicación/reserva en escritura (write-allocate): Gestión del fallo similar a un fallo de escritura, se trae el bloque y se escribe el dato

No ubicación/reserva en escritura (no-write-allocate): Se escribe el dato en el nivel inferior, solo se utiliza con escritura directa

Algoritmos de reemplazo

Se aplican cuando un bloque debe almacenarse en un conjunto que se encuentra lleno

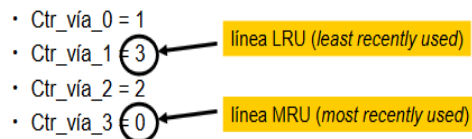
LRU (least recently used)

Se reemplazará el menos recientemente usado, se utiliza un contador por línea para mantener el orden de referencia

Tamaño del contador: $n = \log_2$ (Número de vías)

- **Fallo en la cache:** Si el conjunto esta lleno. Eliminar la línea con valor de contador máximo, para el resto de las líneas sumar 1
- **Acierto:** Para todos los que tengan valor inferior a la línea referenciada sumar 1
- **En cualquier caso:** valor de la línea referenciada = 0

Cada vez que se menciona o reemplaza uno, a ese se le tiene cambiar el bit a 0 y a todos los demás les sumas 1.



que

FIFO (first in first out)

Bits adicionales (misma cantidad que LRU)

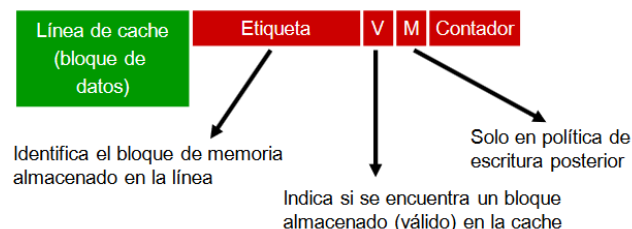
Aleatorio (random)

No es necesario añadir bits

Información de control: Directorio

Información adicional al bloque almacenado en una línea, su volumen depende de la configuración cache.

- Bits de etiqueta
- Bit de valido, bit de modificado (write back)
- Bits de contador del algoritmo de reemplazo



55

QUE NO TE LO CUENTEN, TEN UN DÍA DIFERENTE

Laser Game, Escape Rooms en realidad virtual, experiencias VR...

ALFAZONE

Cálculo del volumen del directorio

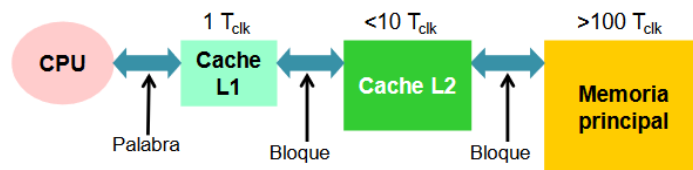
Para calcular el volumen serio: **etiqueta + valido + modificado + contador**

Mejorar el rendimiento

- **Reducir el tiempo de acierto:** cuestiones tecnológicas
- **Disminuir la tasa de fallos:** organizaciones más eficientes
- **Reducir el tiempo par resolver el fallo:** caches multinivel con 2 y hasta 3 niveles de cache, también buses mas anchos entre niveles de memoria cache

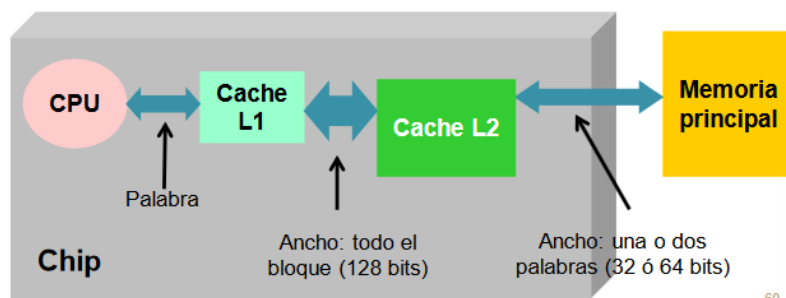
Memorias Cache Multinivel

La cache L2 es mucho mas grande y bastante mas lenta que la cache L1, las dos caches tienen organizaciones independientes



Ancho del bus en caches multinivel

Si la cache L2 esta integrada en el procesador, el ancho de bus L2-L1 puede ser todo el bloque. El ancho de bus entre la memoria y las caches es de una o dos palabras (transferencia por bloques).



Escanéame

WUOLAH

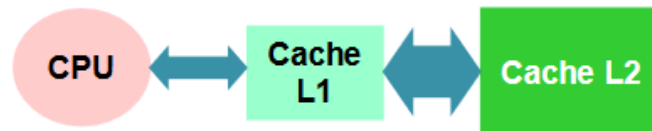
Penalización en el acceso a la cache L1

Secuencialmente

Se detecta un fallo en L1 y entonces se accede a L2

En Paralelo

Se accede a L1 y a L2 al mismo tiempo, si hay acierto en L1 se aborta la operación en L2, en caso de fallo, la penalización de L1 es el tiempo de acceso a L2.



Rendimiento y Parámetros

El diseño de una memoria cache contempla el número de niveles, la capacidad, la asociatividad, el tamaño de bloque, las políticas de lectura y escritura, las políticas de reemplazo.

El rendimiento de un diseño de memoria cache depende de las características que ejecuta el procesador, del ancho de banda demandado por el procesador y del ancho de banda de la memoria principal.

Memoria Virtual

El uso de memoria virtual permite que los programas dispongan de un espacio de direcciones mayor que el de la memoria física.

- En un momento dado, **la mayor parte de la MV direccionable se encuentra en disco**
- Para que los programas y datos sean accesibles por el procesador **deben encontrarse en memoria**
- El sistema operativo se encarga de realizar las gestiones oportunas

Desde la visión del programador, no se preocupa de la cantidad de memoria física del computador

Direcciones Virtuales y Físicas

Es necesario obtener la dirección física a partir de la dirección virtual para acceder al objeto, esto se realiza a base de páginas de memoria: la MV se divide en páginas mientras que la MP se divide en marcos de página, cada marco alberga una página.

Dirección Virtual

Dirección de un objeto utilizada por el procesador, el objeto puede estar en memoria principal (MP) o en memoria secundaria (MS)

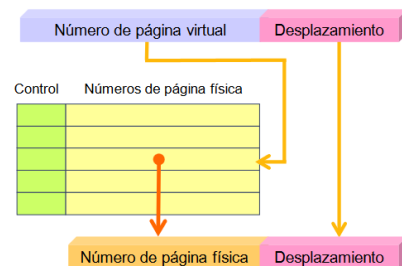
Dirección Física

Dirección de un objeto en memoria principal (MP), son las direcciones emitidas por el bus de direcciones

Traducciones

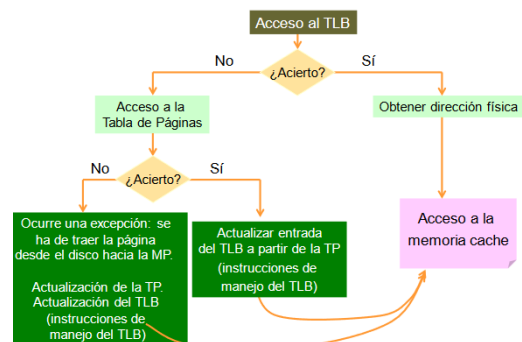
Se hacen mediante una tabla de páginas.

El número de página virtual sirve de índice para acceder a la tabla y recuperar el número de página física.



Traducción rápida de direcciones: el TLB

Hacen falta 2 accesos a memoria (ver la línea y luego la dirección real), se usa una TLB (translation lookaside buffer), cache para direcciones accedidas recientemente



Tema 7 - Adaptadores e Interfaces de Entrada/Salida

El sistema de Entrada – Salida

Permite comunicar al sistema (UCP+Mem) con el exterior por medio de dispositivos periféricos

Elementos de una Unidad de E/S

Periféricos

Dispositivo electrónico o electromecánico que permita la comunicación con el exterior o con otros dispositivos. Hay diferentes tipos:

- De interacción con humanos
- De interacción con otros dispositivos
- De almacenamiento
- De comunicación

Adaptador (controlador de E/S – interfaz de E/S)

Es un dispositivo electrónico que permite conectar el periférico al bus del sistema. Hace la conversión entre la tecnología propia del periférico y las señales del bus.

Constituye la interfaz visible para el programador, traduce las ordenes que se le envían desde los programas en acciones sobre el periférico, proporciona información sobre el estado del periférico y permite la transferencia de los datos.

Principales funciones del adaptador E/S

Comunicación con el procesador

Crea una interfaz al bus del sistema con registros

Comunicación con el dispositivo periférico

Crea una interfaz externa que son un conjunto de señales que conectan con el periférico (cables y conectores)

Control y temporización

Sincronización, prueba de estado vs interrupciones

Transferencia de datos

Almacenamiento temporal de datos (data buffering)

Transferencia por programa vs Acceso directo a memoria (ADM)

Control de errores

Acceso a los periféricos desde los programas

QUE NO TE LO CUENTEN, TEN UN DÍA DIFERENTE

Laser Game, Escape Rooms en realidad virtual, experiencias VR...

ALFAZONE

Visión de programa de los periféricos

Registros

Cada adaptador tiene su rango de direcciones diferentes para evitar la colisión, cada periférico es accesible a través de un conjunto de registros.

Interfaz

Cada periférico tiene su interfaz, en general distinta de la de otros en cuanto a número y uso de los registros

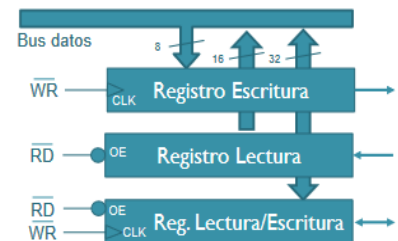
Direcciones

Cada registro de la interfaz tiene una dirección con un espacio de direccionamiento del procesador

Registros de Interfaz

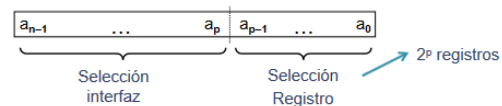
De tamaños 8, 16 y 32 bits, con modos de acceso de lectura (RD), escritura (WR) o ambos. Escribir en un registro de lectura no tiene ningún efecto, leer en un registro de escritura no da información útil.

El valor de un registro puede estar estructurado o no. Si está estructurado, cada bit o grupo de bits del registro tiene significado propio, independiente del resto. En general, entre los bits útiles de un registro puede haber otros indefinidos.



Direccionamiento de la Interfaz

La dirección inicial se denomina dirección base (DB) de la interfaz, los registros de cada interfaz ocupan un rango de direcciones consecutivas



Escanéame

WUOLAH

Esquemas de direccionamiento de las interfaces

Memory-Mapped Input/Output *Modelo MIPS*

Mapa de direccionamiento único compartido por la memoria y los adaptadores de E/S. El acceso a los registros se hace con instrucciones Load/Store

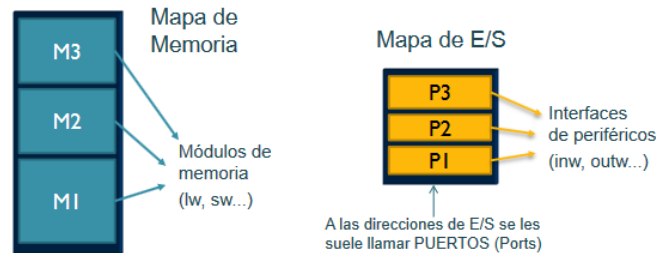
Interfaces de periférico (P1, P2, P3), para tratar los periféricos, cada interfaz abarca unas pocas direcciones y cada registro de una interfaz tiene un uso concreto.

Módulos de memoria (M1, M2, M3), para almacenar instrucciones y datos, gran capacidad, todas las palabras de la memoria tienen el mismo uso



Input/Output-Mapped I/O *Modelo Intel*

Mapas de direccionamiento separados para la memoria y la E/S, el acceso a la memoria se hace con instrucciones Load/Store y el acceso a los periféricos se hace con instrucciones específicas de tipo

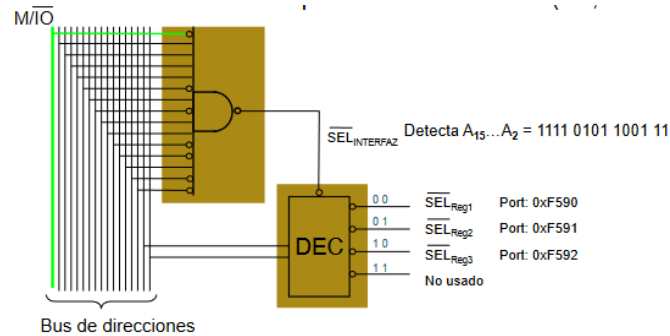


Input/Output

Selección y uso de la interfaz

Una línea en el bus permite distinguir entre los dos espacios de direccionamiento

- $M/\overline{IO} = 1$: Espacio de memoria
- $M/\overline{IO} = 0$: Espacio de Entrada/Salida



Selección y operación de los registros

Los registros tienen señal de escritura (CLK, flanco de reloj), de lectura (OE^*) o ambas y entrada/salida de datos paralela conectada al bus de datos.

Dos líneas de bus definen la operación: RD^* para la lectura y WR^* para escritura

