

# MEMORIA CACHE

## CUARTA UNIDAD



# Sumario

- I. Descripción general del sistema memoria de la computadora
  - i. Características de los sistemas de memoria
  - ii. La jerarquía de memoria
- II. Principios de caché de memoria
- III. Elementos de Diseño de caché
  - i. Las direcciones de caché
  - ii. Tamaño de la caché
  - iii. Asignación de funciones
  - iv. Algoritmos de sustitución
  - v. Política de escritura. (\*)
  - vi. Tamaño de la línea (\*)
  - vii. Número de cachés (\*)
- IV. Pentium 4 – Organización de Cache
- V. ARM - Organización de caché

# ¿ La Biblioteca y la Memoria Cache ?

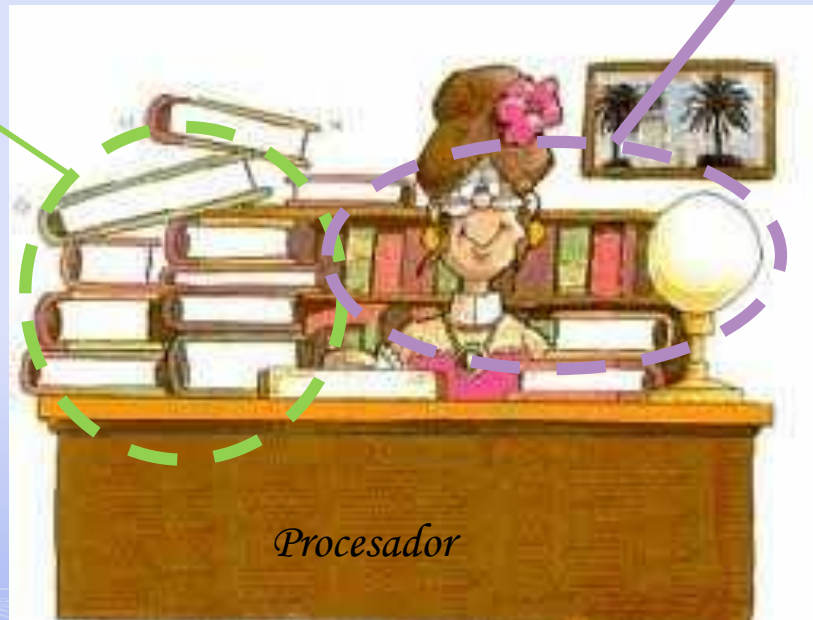
Memoria  
Principal  
(DRAM)



Memoria Cache L3

Memoria Cache L1

Memoria Cache L2



Procesador



Procesos





# I. Descripción general del sistema memoria de la computadora

## i. Características de los sistemas de memoria

### **Ubicación**

Internos (e.g. Registro de procesador, Memoria Principal, memoria cache)

Externos (e.g. discos ópticos, discos magnéticos, cintas)

### **Capacidad**

Numero de palabras

Numero de bytes

### **Unidad de Transferencia**

Bloque

Palabra

### **Método de Acceso**

Secuencial

Directo

Aleatorio

Asociativo

### **Rendimiento**

Tiempo de acceso

Tiempo de vida

Tasa de transferencia

### **Tipo Físico** Semiconductor Magnético

Óptico

Magneto-óptico

### **Características Físicas** Volátil / no volátil

Borrable /no Borrable

### **Organización**

Módulos de memoria

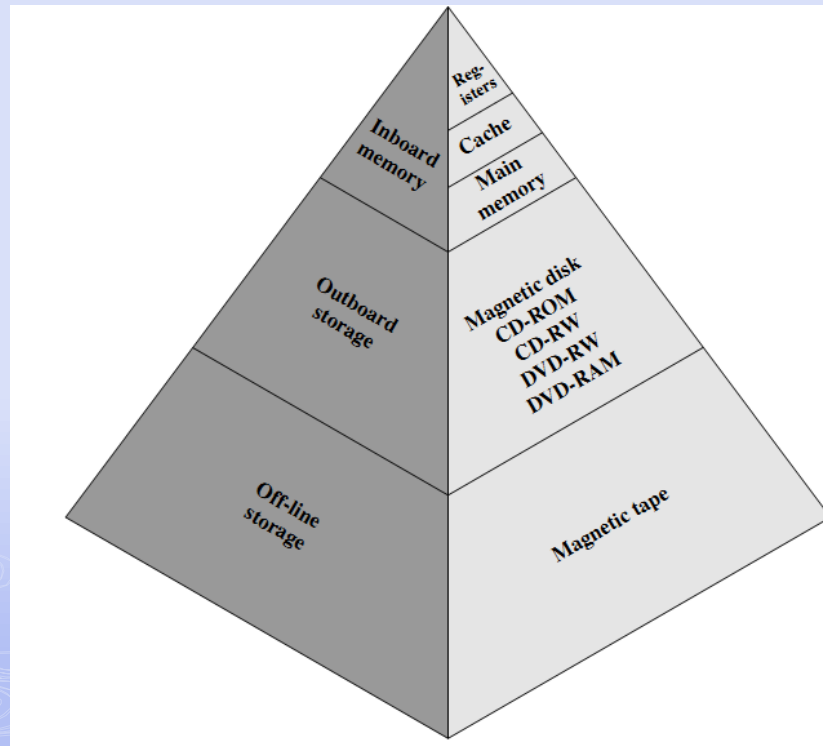
# I. Descripción general del sistema memoria de la computadora

## ii. Jerarquía de Memoria

Un máximo rendimiento procesador sucede cuando ejecuta instrucciones y no tiene que hacer una pausa en espera de instrucciones de sus Operandos.

Las relaciones que se tienen son:

- Acceso más rápido el tiempo, mayor costo por bit
- Mayor capacidad, menor costo por bit
- Mayor capacidad, más lento el tiempo de acceso



# I. Descripción general del sistema memoria de la computadora

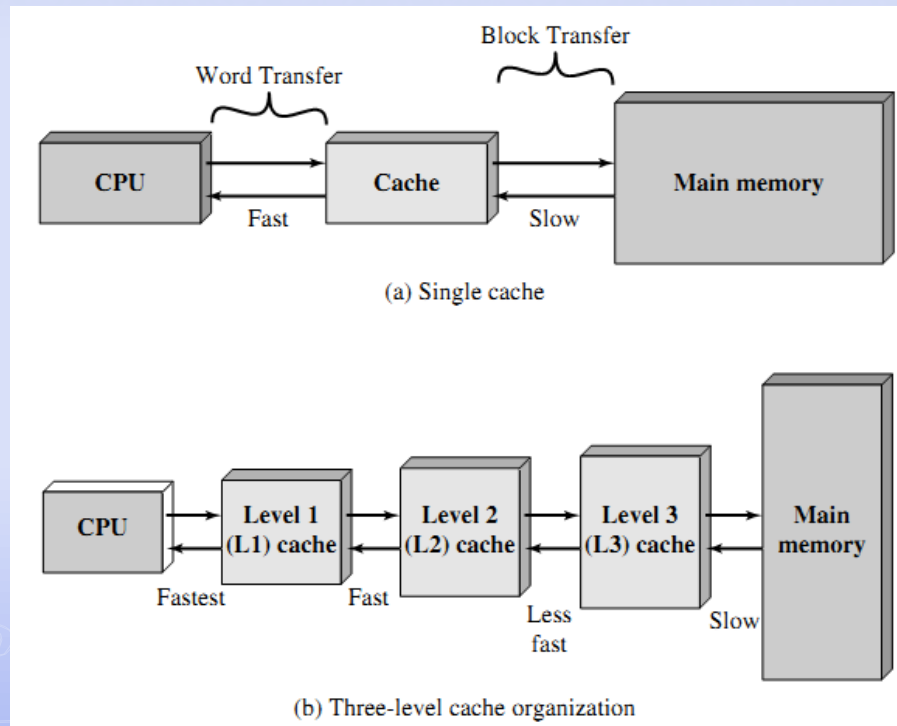
## i. Jerarquía de Memoria



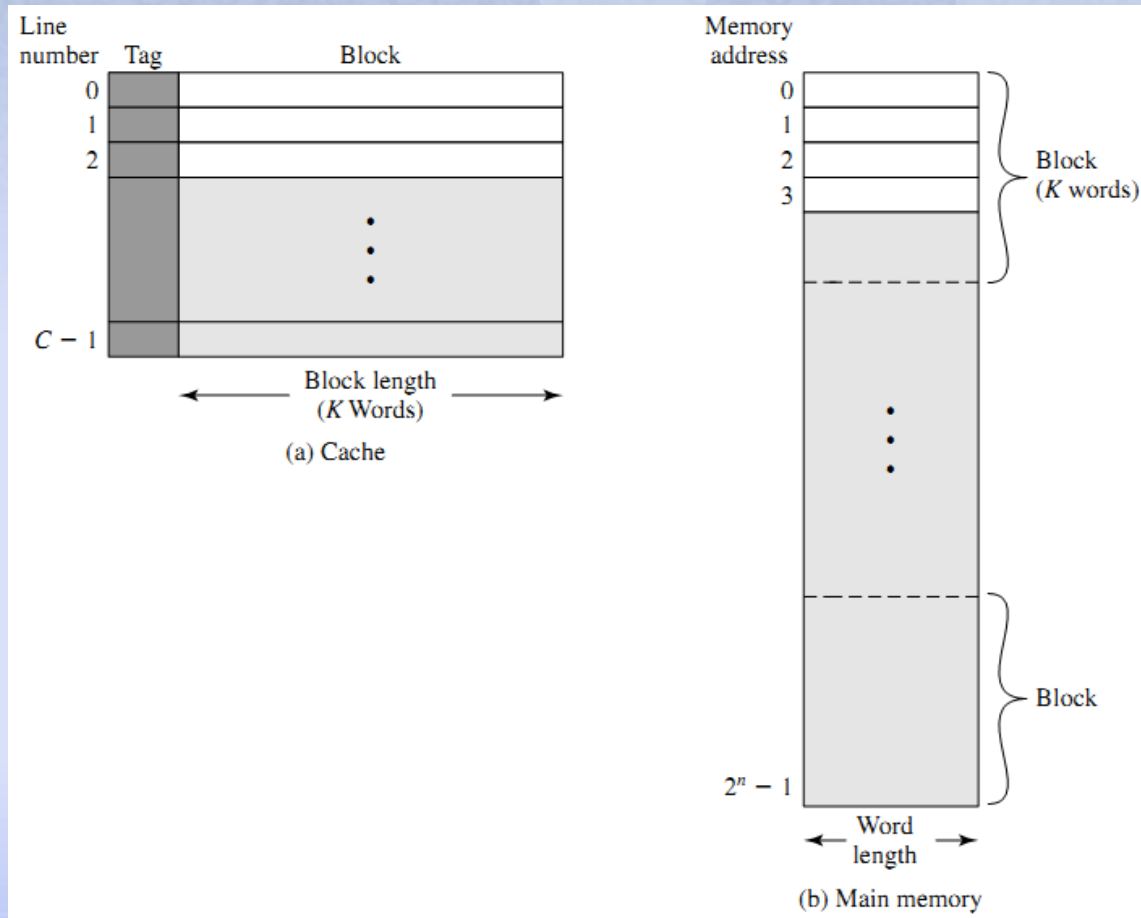


## II. Principios de la Memoria Cache

La memoria caché contiene una copia de partes de la memoria principal. Cuando el procesador intenta leer una palabra de la memoria, se realiza una comprobación para determinar si la palabra está en la caché. Si es así, la palabra se entrega al procesador. Si no, un bloque de memoria principal, que consta de un número fijo de palabras, se lee en la memoria caché y, a continuación la palabra se entrega al procesador.



## II. Principios de la Memoria Cache

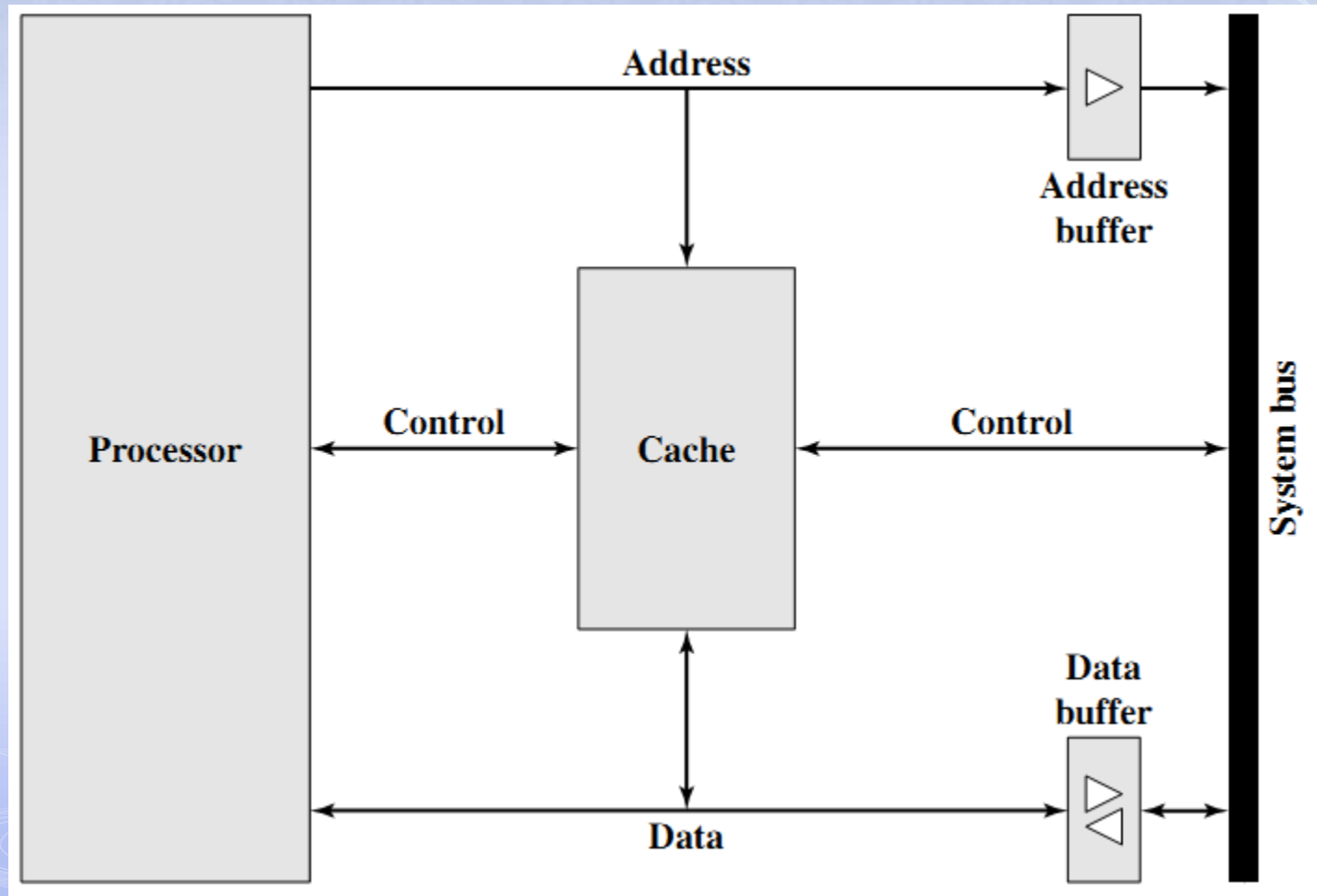




## II. Principios de la Memoria Cache

- En esta organización, la memoria caché del procesador se conecta a través de los datos, el control y las líneas de dirección.
- Los datos y las líneas de dirección también se unen a los datos y registros de dirección, que se conectan a un bus de sistema desde el que se alcanza la memoria principal.
- Cuando se produce un acierto de caché, los datos y registros de dirección están desactivados y la comunicación es sólo entre el procesador y la memoria caché, sin tráfico en el bus del sistema.
- Cuando se produce un error de caché, la dirección deseada se carga en el bus del sistema y los datos son devueltos a través del buffer de datos tanto a la memoria caché y el procesador.
- En otras organizaciones, la memoria caché está físicamente interpone entre el procesador y la memoria principal para todos los datos, direcciones y líneas de control.

## II. Principios de la Memoria Cache



### III. Elementos del Diseño de una Cache

**Cache Addresses**

Logical

Physical

**Cache Size****Mapping Function**

Direct

Associative

Set Associative

**Replacement Algorithm**

Least recently used (LRU)

First in first out (FIFO)

Least frequently used (LFU)

Random

**Write Policy**

Write through

Write back

Write once

**Line Size****Number of caches**

Single or two level

Unified or split



# III. Elementos del Diseño de una Cache

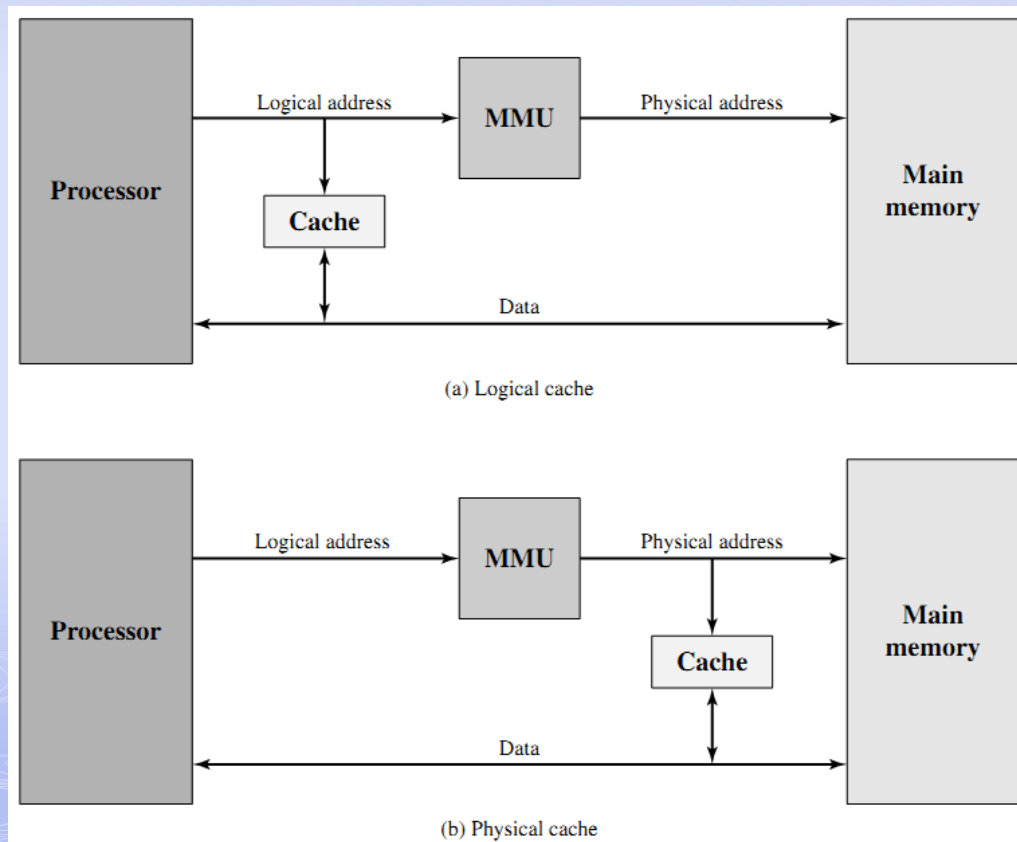
## i. Direcciones de la Cache

- En esencia, la memoria virtual es un servicio que permite a los programas hacer frente a la memoria desde un punto de vista lógico, sin tener en cuenta la cantidad de memoria principal dispone de preparados físicamente se utiliza memoria virtual, los campos de dirección de las instrucciones de la máquina contiene las direcciones virtuales.
- Para leer y escribir en la memoria principal, una unidad de gestión de hardware de memoria (MMU) traduce cada dirección virtual a una dirección física en la memoria principal.
- Cuando se utilizan direcciones virtuales, el diseñador del sistema puede optar por colocar la memoria caché entre el procesador y MMU, o entre la MMU y principal de memoria.

# III. Elementos del Diseño de una Cache

## i. Direcciones de la Cache

- Una caché lógica, también conocido como una caché virtual, almacena los datos el uso de direcciones virtuales. El procesador accede a la caché directamente, sin pasar a través de la MMU. A las tiendas físicas que utilizan datos de la caché de memoria física principal direcciones.



# III. Elementos del Diseño de una Cache

## ii. Tamaño de Cache

- Nos gustaría que el tamaño de la caché sea lo suficientemente pequeño para que el coste global medio por bit sea próximo a la memoria principal y lo suficientemente grande como para que el tiempo total de acceso promedio sea cercano.
- Cuanto mayor sea el caché, mayor será el número de puertas que se ocupan del resultado de la cache.
- Debido a que el rendimiento de la caché es muy sensible a la naturaleza de la carga de trabajo, es imposible llegar a un solo tamaño "óptimo" caché.



# III. Elementos del Diseño de una Cache

## ii. Tamaño de Cache

Processor	Type	Year of Introduction	L1 Cache <sup>a</sup>	L2 Cache	L3 Cache
IBM 360/85	Mainframe	1968	16 to 32 kB	—	—
PDP-11/70	Minicomputer	1975	1 kB	—	—
VAX 11/780	Minicomputer	1978	16 kB	—	—
IBM 3033	Mainframe	1978	64 kB	—	—
IBM 3090	Mainframe	1985	128 to 256 kB	—	—
Intel 80486	PC	1989	8 kB	—	—
Pentium	PC	1993	8 kB/8 kB	256 to 512 KB	—
PowerPC 601	PC	1993	32 kB	—	—
PowerPC 620	PC	1996	32 kB/32 kB	—	—
PowerPC G4	PC/server	1999	32 kB/32 kB	256 KB to 1 MB	2 MB
IBM S/390 G4	Mainframe	1997	32 kB	256 KB	2 MB
IBM S/390 G6	Mainframe	1999	256 kB	8 MB	—
Pentium 4	PC/server	2000	8 kB/8 kB	256 KB	—
IBM SP	High-end server/ supercomputer	2000	64 kB/32 kB	8 MB	—
CRAY MTA <sup>b</sup>	Supercomputer	2000	8 kB	2 MB	—
Itanium	PC/server	2001	16 kB/16 kB	96 KB	4 MB
SGI Origin 2001	High-end server	2001	32 kB/32 kB	4 MB	—
Itanium 2	PC/server	2002	32 kB	256 KB	6 MB
IBM POWER5	High-end server	2003	64 kB	1.9 MB	36 MB
CRAY XD-1	Supercomputer	2004	64 kB/64 kB	1 MB	—
IBM POWER6	PC/server	2007	64 kB/64 kB	4 MB	32 MB
IBM z10	Mainframe	2008	64 kB/128 kB	3 MB	24–48 MB

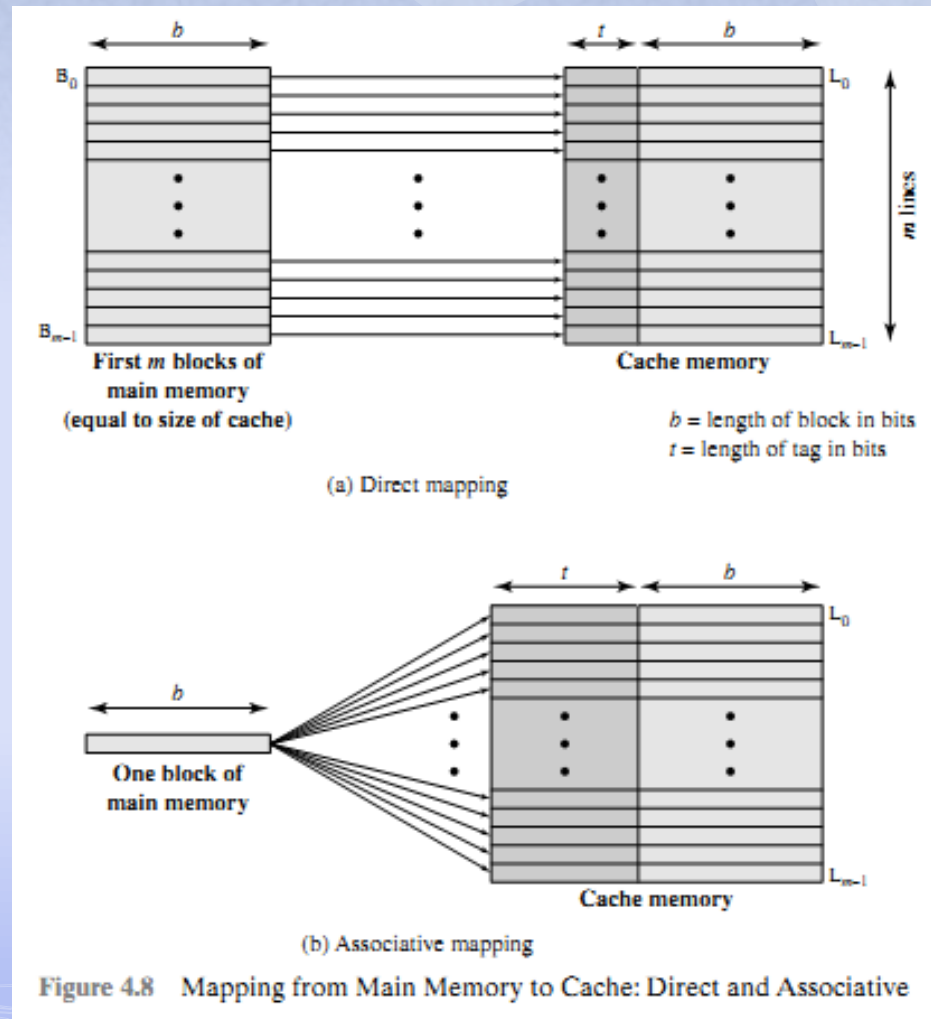
# III. Elementos del Diseño de una Cache

## iii. Asignación de Funciones

- Debido a que hay un menor número de líneas de caché que los bloques de memoria principal, un algoritmo es necesaria para el mapeo de los bloques de memoria principal en las líneas de caché.
- Un medio es necesaria para determinar qué bloque de memoria principal en la actualidad ocupa una línea de caché. La elección de la función dicta la forma en la cartografía de la memoria caché es organizada. Tres técnicas se pueden utilizar: directa, asociativa, y establecer asociación.

# III. Elementos del Diseño de una Cache

## iii. Asignación de Funciones





# III. Elementos del Diseño de una Cache

## iv. Algoritmos de reemplazo

Para el mapeo directo, sólo hay una línea posible para cualquier bloque particular, y otra opción es posible. Para las técnicas asociativas y asociativa de conjunto, un algoritmo de reemplazo es necesario.

(LRU) Cuando una línea se hace referencia, su bit de uso se establece en 1 y el bit de uso de la otra línea en conjunto que se establece en 0

# III. Elementos del Diseño de una Cache

## V. Política de Escritura

*“Resumen de media hoja”*

# III. Elementos del Diseño de una Cache

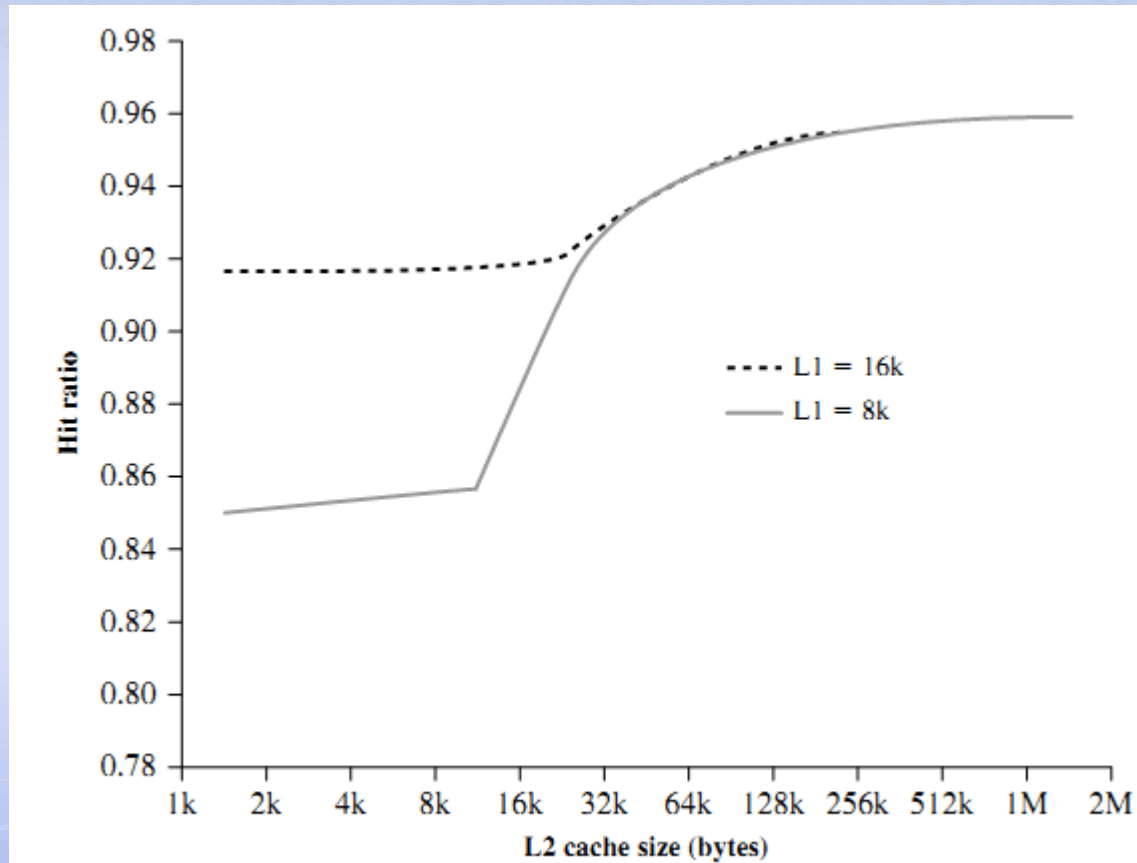
## V. Tamaño de la Línea

*“Resumen de media hoja”*



# III. Elementos del Diseño de una Cache

## Vi. Numero de Caches



**Figure 4.17** Total Hit Ratio (L1 and L2) for 8-Kbyte and 16-Kbyte L1

## IV. Organización de la Cache en una Pentium 4

Problem	Solution	Processor on which Feature First Appears
External memory slower than the system bus.	Add external cache using faster memory technology.	386
Increased processor speed results in external bus becoming a bottleneck for cache access.	Move external cache on-chip, operating at the same speed as the processor.	486
Internal cache is rather small, due to limited space on chip	Add external L2 cache using faster technology than main memory	486
Contention occurs when both the Instruction Prefetcher and the Execution Unit simultaneously require access to the cache. In that case, the Prefetcher is stalled while the Execution Unit's data access takes place.	Create separate data and instruction caches.	Pentium
Increased processor speed results in external bus becoming a bottleneck for L2 cache access.	Create separate back-side bus that runs at higher speed than the main (front-side) external bus. The BSB is dedicated to the L2 cache.	Pentium Pro
	Move L2 cache on to the processor chip.	Pentium II
Some applications deal with massive databases and must have rapid access to large amounts of data. The on-chip caches are too small.	Add external L3 cache.	Pentium III
	Move L3 cache on-chip.	Pentium 4

# IV. Organización de Cache ARN

- La organización de caché ARM ha evolucionado con la arquitectura general de la familia ARM.
- Los modelos utilizados ARM usa un caché L1 unificado, mientras que todos los modelos posteriores utilizan una fracción de la instrucción / caché de datos.
- Todos los diseños de ARM utilizar un caché de conjunto asociativo, con el grado de asociatividad y el tamaño de la línea variable.

Table 4.6 ARM Cache Features

Core	Cache Type	Cache Size (kB)	Cache Line Size (words)	Associativity	Location	Write Buffer Size (words)
ARM720T	Unified	8	4	4-way	Logical	8
ARM920T	Split	16/16 D/I	8	64-way	Logical	16
ARM926EJ-S	Split	4-128/4-128 D/I	8	4-way	Logical	16
ARM1022E	Split	16/16 D/I	8	64-way	Logical	16
ARM1026EJ-S	Split	4-128/4-128 D/I	8	4-way	Logical	8
Intel StrongARM	Split	16/16 D/I	4	32-way	Logical	32
Intel Xscale	Split	32/32 D/I	8	32-way	Logical	32
ARM1136-JF-S	Split	4-64/4-64 D/I	8	4-way	Physical	32