

Arquitectura y Organización de Computadoras I

El sistema de memoria

MSc. Ing. Ticiano J. Torres Peralta
Pu. Pablo Rodríguez Rey
Ing. Pablo Toledo
Tomas Consoli

2024 – Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Repaso – Medidas de Información

La unidad básica es el dígito binario (1 o 0), y se llama un bit. Desde el bit, podemos construir las demás medidas comunes que se usan para representar cantidades.

Nibble (half-byte, tetrade, semi-octet, quadbit, quartet): 4 bits.

Byte: 8 bits.

Palabra* (Word): la unidad natural de datos que maneja una arquitectura. Para una arquitectura de 32 bits, es 32 y para una de 64 bits, 64.

Half (Halfword): la mitad de una palabra.

*Por cuestiones de que procesadores tienden a retener una gran compatibilidad en revesa con viejas generaciones, la definición de una palabra se puede complicar. Por ejemplo, es común que un procesador que maneja tanto un ISA de 32 bits como una de 64 bits, defina el word como 32 bits y un longword como 64bits.

2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Repaso – Medidas de Información

System of Units (SI)			Binary Numeral					% Difference
Factor	Name	Symbol	Factor	Name	Symbol	# of Bytes		
10^3	kilobyte	KB	2^{10}	kibibyte	KiB	1,024	2.4%	
10^6	megabyte	MB	2^{20}	mebibyte	MiB	1,048,576	4.9%	
10^9	gigabyte	GB	2^{30}	gibibyte	GiB	1,073,741,824	7.4%	
10^{12}	terabyte	TB	2^{40}	tebibyte	TiB	1,099,511,627,776	10.0%	
10^{15}	petabyte	PB	2^{50}	pebibyte	PiB	1,125,899,906,842,624	12.6%	
10^{18}	exabyte	EB	2^{60}	exbibyte	EiB	1,152,921,504,606,846,976	15.3%	
10^{21}	zettabyte	ZB	2^{70}	zebibyte	ZiB	1,180,591,620,717,411,303,424	18.1%	
10^{24}	yottabyte	YB	2^{80}	yobibyte	YiB	1,208,925,819,614,629,174,706,176	20.9%	

2024

Primer Cuatrimestre

Repaso – Medidas de Información

Type	Storage size	Value range
char	1 byte	-128 to 127 or 0 to 255
unsigned char	1 byte	0 to 255
signed char	1 byte	-128 to 127
int	2 or 4 bytes	-32,768 to 32,767 or -2,147,483,648 to 2,147,483,647
unsigned int	2 or 4 bytes	0 to 65,535 or 0 to 4,294,967,295
short	2 bytes	-32,768 to 32,767
unsigned short	2 bytes	0 to 65,535
long	8 bytes or (4bytes for 32 bit OS)	-9223372036854775808 to 9223372036854775807
unsigned long	8 bytes	0 to 18446744073709551615

La memoria

Ya se menciono que existen dos tipos de memoria, la principal y la secundaria. En verdad, hay mas tipos y todas hacen lo mismo, guardan información. Entonces “¿Por que hay tantos tipos de memoria en la computadora?”

Durante los años, mejoras en la capacidad de procesamiento del procesador han superado ampliamente la capacidad del sistema de memoria en poder transferir datos al procesador. Entonces, siempre se están desarrollan nuevas tecnologías en el sistema de memoria con el objetivo de igualar la capacidades del procesador.

Para que el sistema de memoria no se vuelva un cuello de botella, se generaron soluciones para que el mismo pueda rápidamente entregarle datos al procesador. Estas se pueden organizar en lo que se conoce como: la jerarquía de memoria.

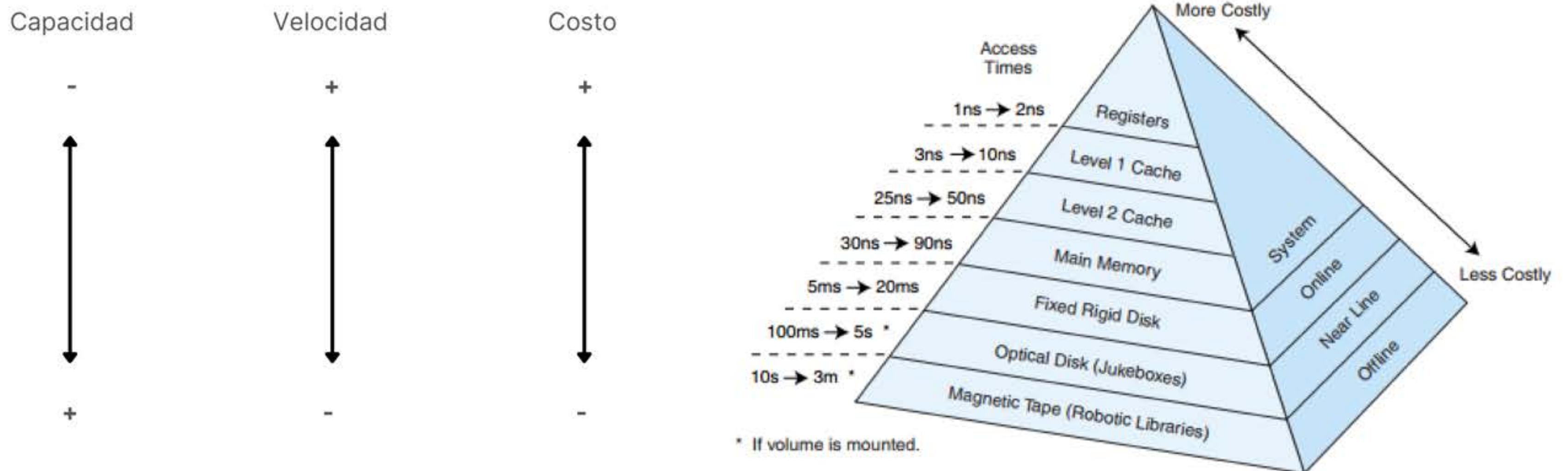
2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

La memoria



2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Memoria Secundaria

La memoria secundaria es la memoria mas barata y mas lenta de la computadora. Esta memoria no es volátil y es la memoria utilizada para almacenar cualquier tipo de información en las computadoras, como por ejemplo, el sistema operativo, aplicaciones y archivos.

Las memorias secundarias son periféricos que se enchufan en un bus, generalmente el PCIe o el SATA. A veces, se los llama dispositivos de bloques (Block Devices) porque leen y escriben datos en bloques típicamente entre 512B y 4KiB (4096B).

Hoy, tecnologías populares de memoria secundarias incluyen el HDD (Hard Disk Drive – Disco Duro) y el SSD (Solid-State Disk – Disco de Estado Solido).

2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

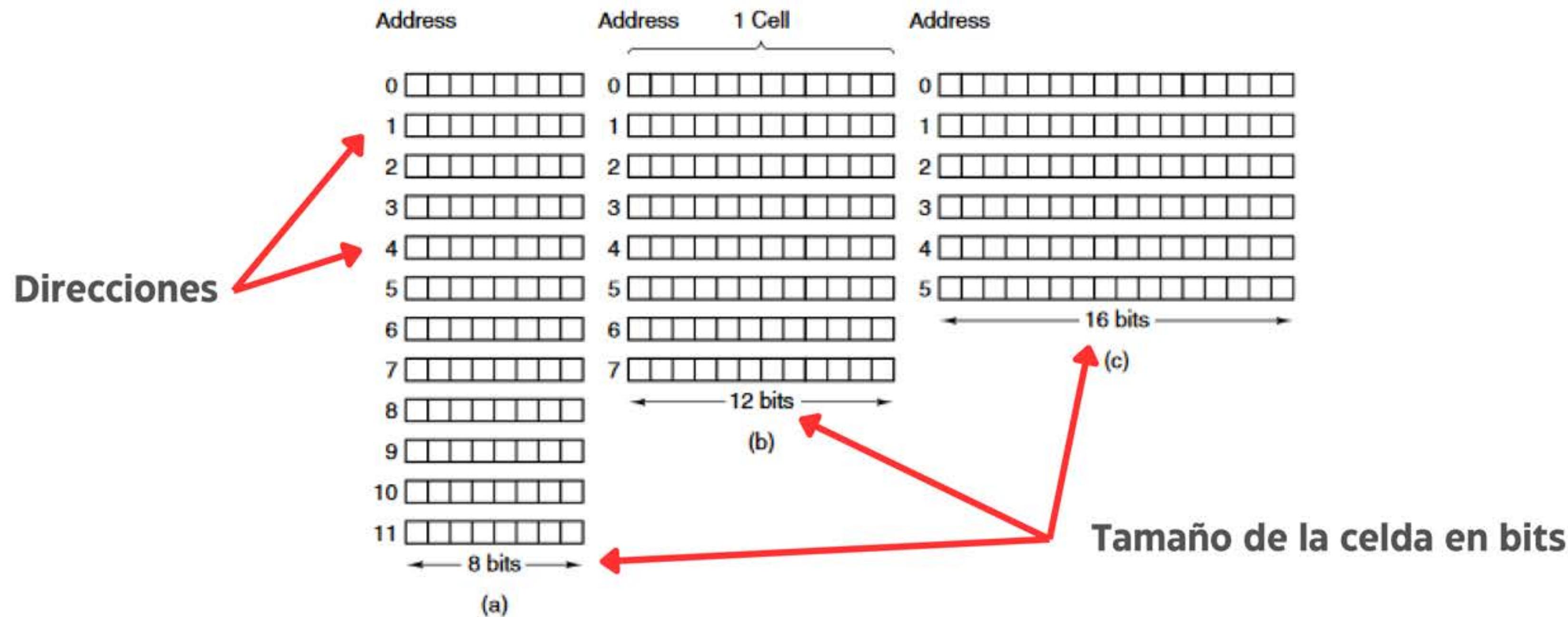
Memoria principal

La porción mas chica de información que se puede acceder en una memoria es lo que contiene una celda. Hoy, lo mas común es que una celda sea de un Byte, y es la razón porque el Byte y sus multiples es la medida mas utilizada en la informática.

Cada celda tienen asociado un numero que la identifica. Este numero, llamado dirección, es usado por programas para acceder a la información que contiene la celda en cuestión.

Por diseño, las direcciones de memoria siempre empiezan en 0 y están ordenadas de forma consecutiva hasta el tamaño máximo de la misma. Por ejemplo, una memoria de 4GiB (2^{32}) va de dirección 0x00000000 a 0xFFFFFFFF (0 a 4.294.967.295 en decimal).

Memoria principal



Memoria principal

Es normal que una memoria tenga implementado mas de un canal de comunicación para datos. Por lo general hay un mínimo de dos, uno donde el acceso es por byte (por celda) y otro donde el acceso es por palabra (64 bits hoy en día).

Por ejemplo, si leo de memoria dirección 0xFAB93CD0 por el canal que lee de un byte, accedo a la información de esa única celda, pero si uso el canal que lee de una palabra, leo la información en 0xFAB93CD0, 0xFAB93CD1, 0xFAB93CD2, 0xFAB93CD3, 0xFAB93CD4, 0xFAB93CD5, 0xFAB93CD6, y 0xFAB93CD7 (8 Bytes or 64 bits)

Es una memoria **volátil**, que significa que la información no perdura si se apaga el sistema.

2024

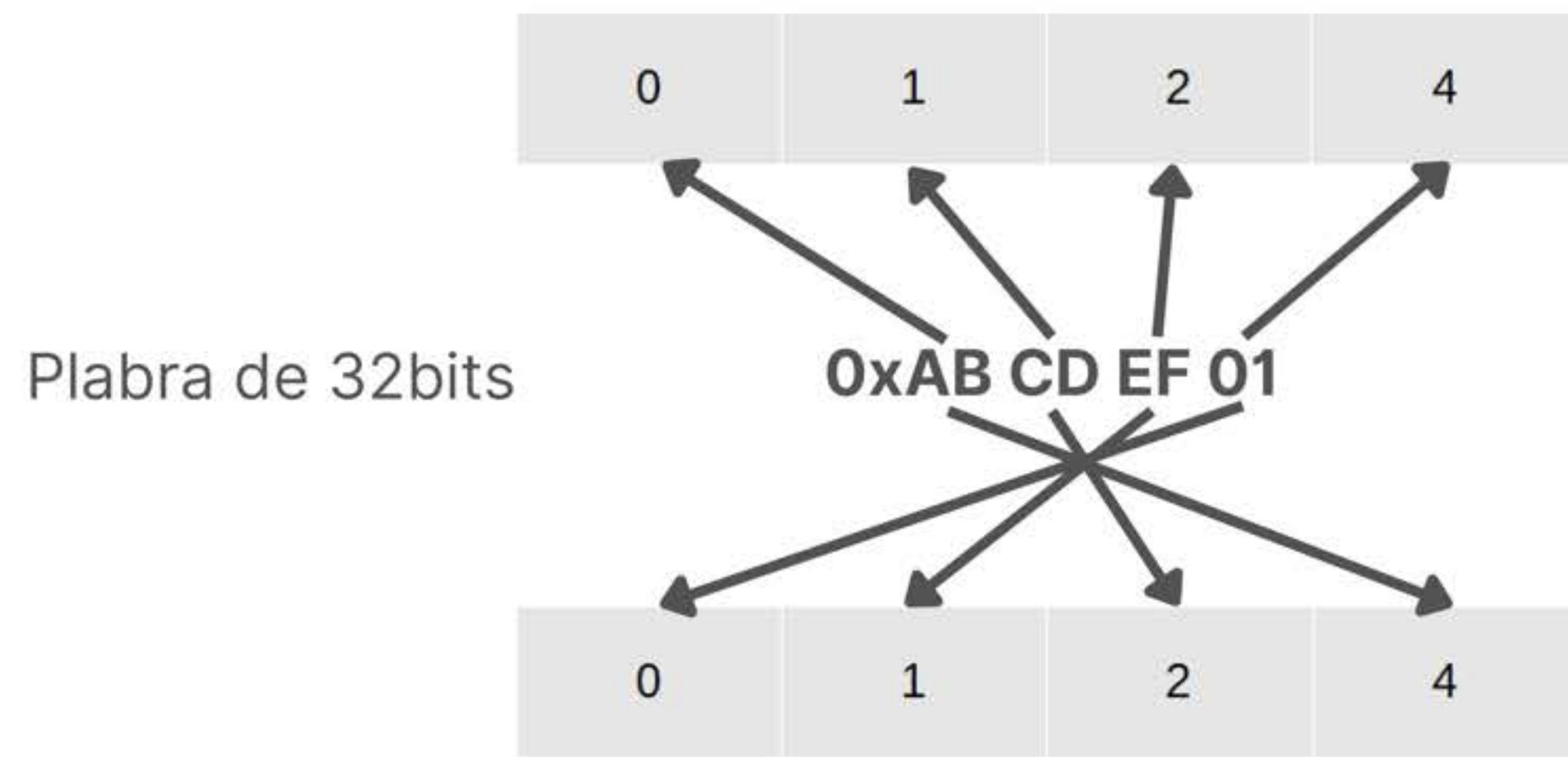
Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Memoria principal

Ya que se puede leer y escribir palabras en el sistema de memoria, ¿Como ordenamos los Bytes de la palabra en memoria?



2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
f_{ac}e_ct
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Memoria principal

Las dos opciones mostradas son posibles:

BigEndian: ordenamiento de bytes donde la numeración de celdas empieza en el byte mas significativo de la palabra.

LittleEndian: ordenamiento de bytes donde la numeración de celdas empieza en el byte menos significativo de la palabra.

Es importante considerar que si uno quiere transferir datos entre una maquina big endian y otra little endian, tienen que haber un proceso en el medio que invierte el orden.

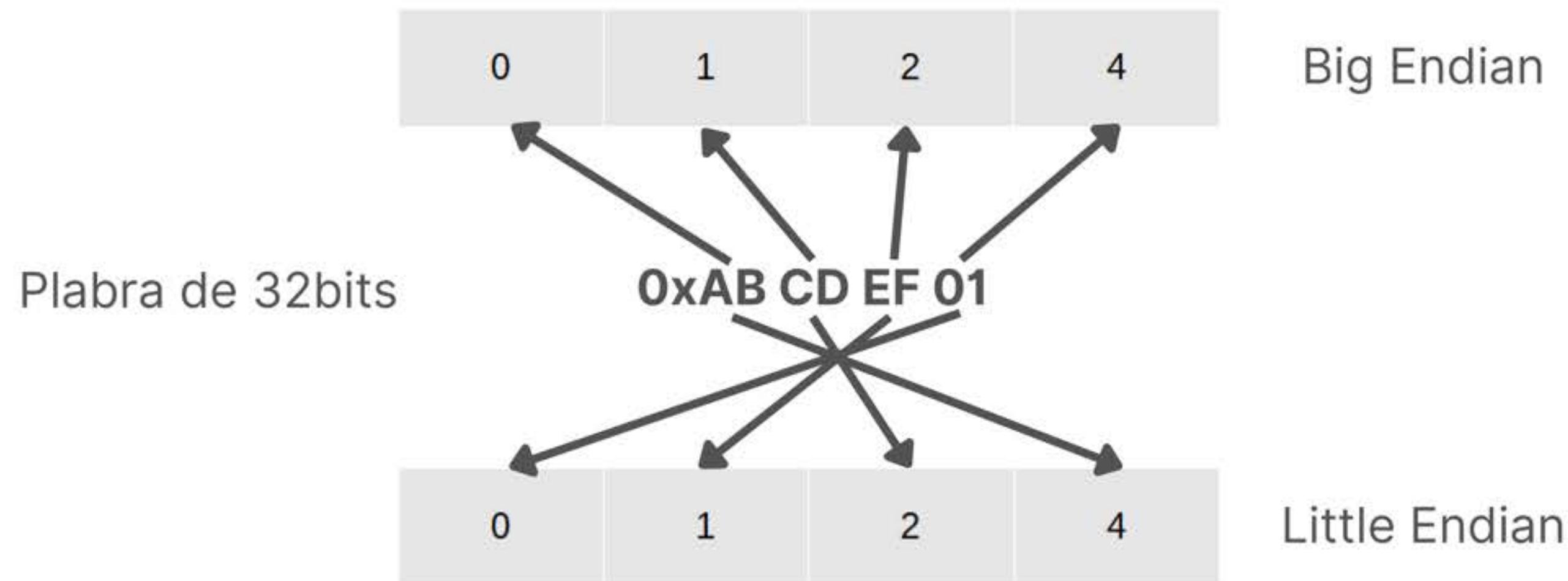
2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Memoria principal



Cache

Históricamente, el procesador siempre fue mas rápido que la memoria, y con el avance de la tecnología, esta disparidad aumento aun mas. Además, diseñadores de memoria (principal) normalmente usan las nuevas tecnologías para priorizar la capacidad y no la velocidad.

Esto causa un problema muy serio, porque cuando el procesador ejecuta una instrucción y necesita datos que están en memoria, tiene que esperar varios ciclos para recibirlas.

Para reducir este problema, por ejemplo, se podría crear memoria mas rápida que esté adentro del procesador (transferir por bus es lento), pero nos generaría un problema económico, aumentando el tamaño físico y el precio.

2024

Primer Cuatrimestre

Cache

Una solución mas viable, que toma un poco de los dos mundos, es tener una pequeña memoria rápida (cara) y una grande memoria mas lenta (barata). Esto permite que el sistema tenga la rapidez de la memoria rápida con la capacidad de la memoria lenta por un precio moderado. Esta pequeña memoria rápida es conocida como **cache**.

La idea básica de la cache es simple: tener los objetos mas usados en ella. De esta manera, si el procesador lo necesita, lo busca primero en cache y solo lo pide de memoria (principal) si no lo encuentra ahí.

La cache es una memoria volátil, como la memoria principal.

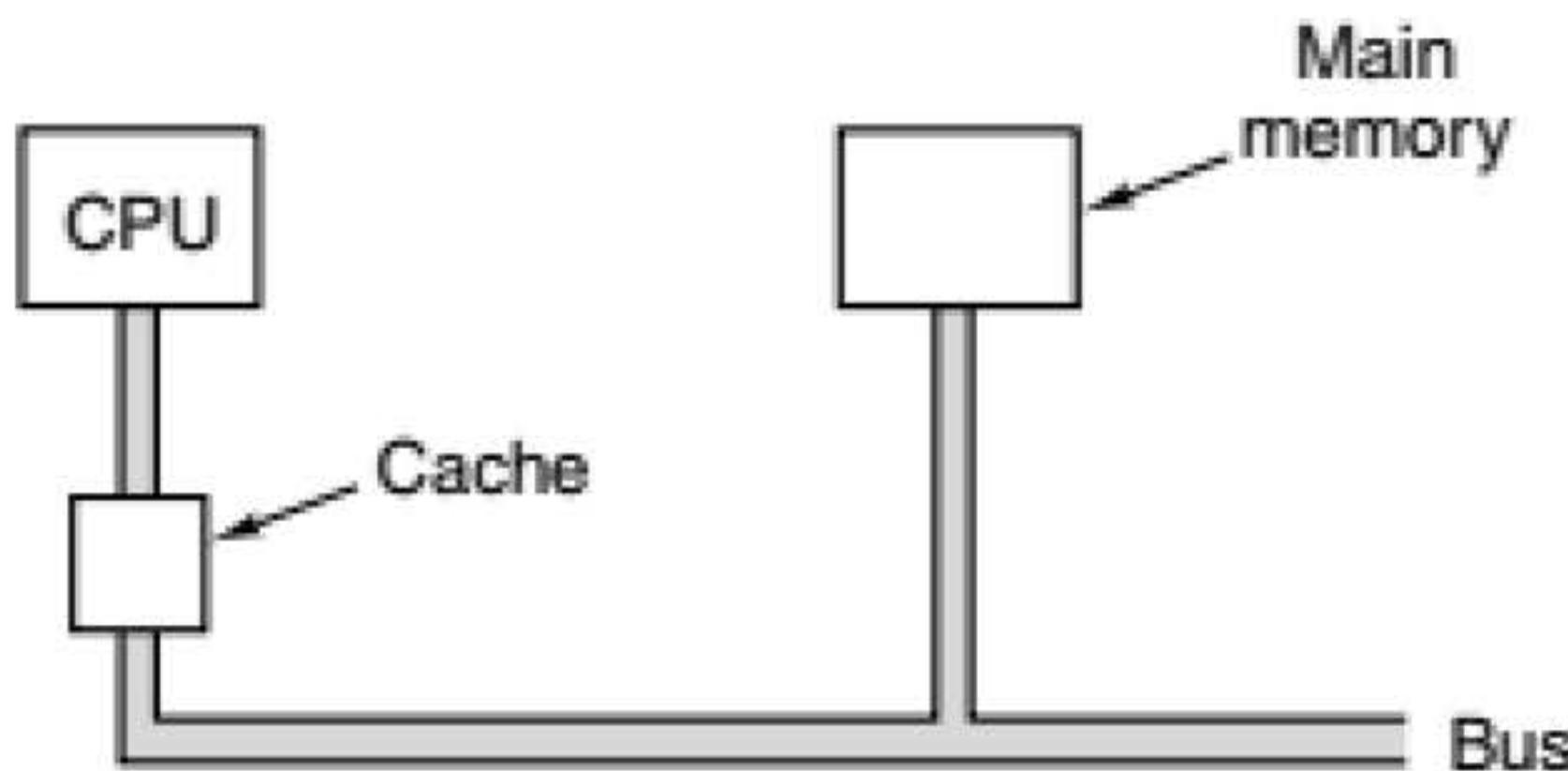
2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA
facet

Cache



2024
Primer Cuatrimestre

Cache

Para definir, si el procesador encuentra el objeto en cache, se llama un **acuerdo (cache hit)**, si no, se llama un **fallo (cache miss)**.

Hay diferentes teorías, algoritmos y tipos de cache que buscan maximizar los aciertos y minimizar los fallos. Hoy en día, las cache tienen una tasa de aciertos de alrededor del 90%.

2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA
facet

Cache

Internamente, la cache esta compuesta por varias líneas (**Cache Lines**) donde cada una contienen un bloque de memoria, ósea, varias palabras.

A diferencia de la memoria principal, la cache no es accedida por direcciones a bytes, si no por lo que se refiere a contenido. Por esto, a veces las caches son conocidas como **Memorias de Direcciónamiento por Contenido (Content Addressable Memory – CAM)**.

Factores que afectan el funcionamiento de la cache:

- El tamaño del cache
- La póliza de mapeo del cache (Cache Mapping Scheme)
- La póliza de remplazo del cache
- La póliza de escritura del cache

Cache

No hay una sola cache, puede haber varios niveles según el diseño y las necesidades del sistema.

En procesadores modernos existen típicamente dos niveles de cache privadas (L1 y L2) y un nivel de cache compartida (L3).

La cache no tiene que ser muy grande para funcionar bien. Tampoco queremos hacerla muy grande por cuestiones de costo (y hacer todo el sistema de memoria de cache sería imposiblemente caro).

Tamaños típicos hoy en día son:

L1 : 8KB a 64KB

L2: 256KB a 512KB

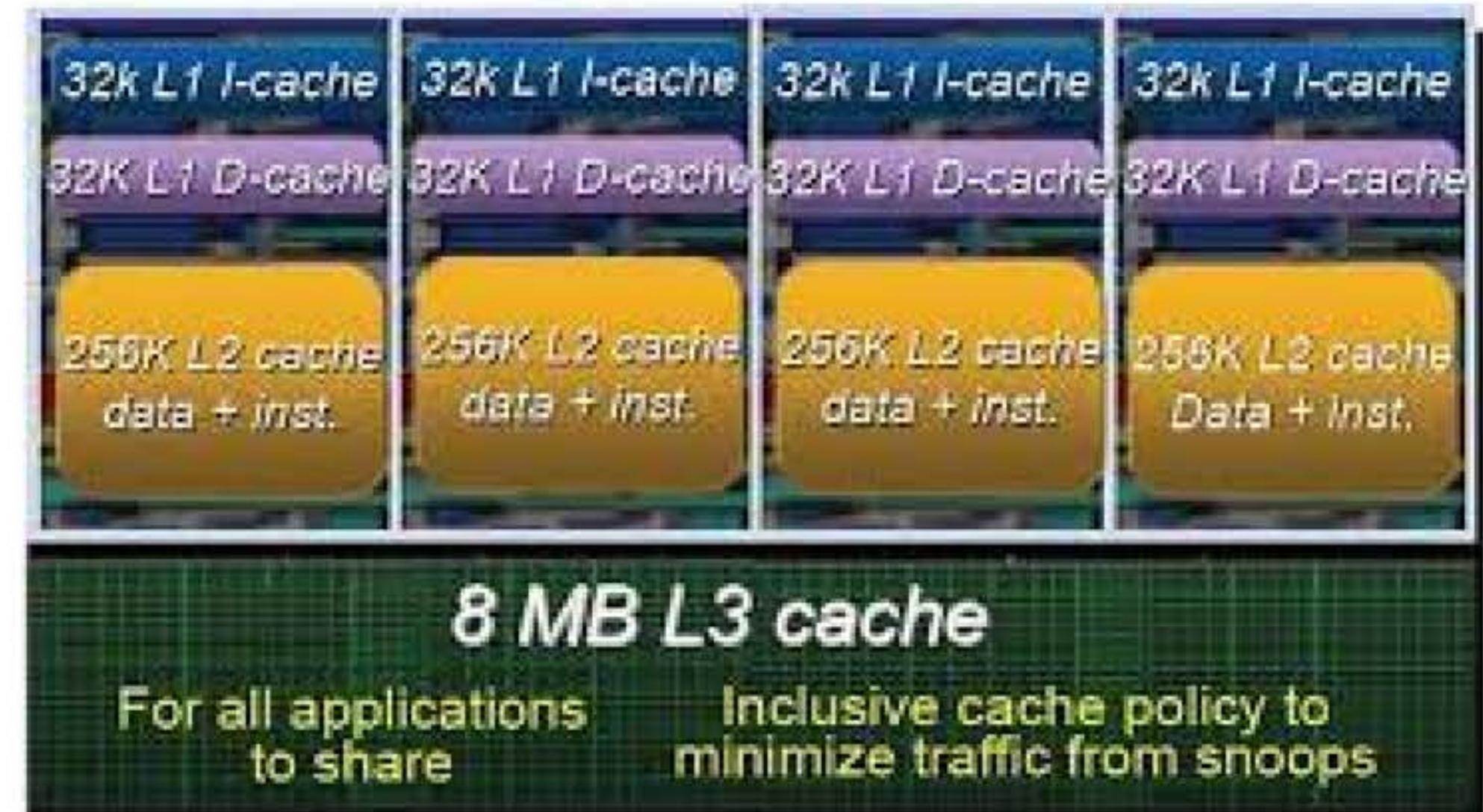
L3: 8MB a 32MB

Cache

Como pueden ver del diagrama:

- L1 esta dividida en dos, una para datos (L1 D-cache) y una para instrucciones (L1 I-cache).
- L2 es mas grande que L1 y guarda tanto instrucciones como datos.
- L3 es aun mas grande y es compartida entre todos los cores. Se puede usar para rápidamente transferir información entre cada core.

Niveles de cache para la arquitectura Nehalem de Intel



2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Cache

Cuando el procesador busca una instrucción o un dato, primero genera una dirección de memoria y pero primero busca si se encuentra en la cache. Si es que el dato está, la locación que tienen ese dato en cache es diferente a su dirección de memoria.

Entonces, para que la cache sea funcional, el procesador tiene que saber como encontrar esa información en la misma, ósea, como mapear una dirección de memoria a una línea de cache.

Esto nos genera dos preguntas:

¿Como sabe el procesador si la palabra esta en la cache?

¿Como sabe el procesador como encontrarla?

2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Cache

El procesador convierte la dirección de memoria a una locación en el cache. Esto se llama, la **póliza de mapeo del cache (Cache Mapping Scheme)**.

Para lograr esto, se usa parte de la dirección de memoria para dividir la memoria en grupos. Estos grupos contienen bloques de palabras, donde cada bloque es lo que se mapea a una línea de cache. Tambien, se usa un bit que indica que la línea de cache esta valida.

Dentro de las posibles pólizas de mapeo existen:

- Direct Mapped Cache
- Fully Assosiative Cache
- Set Assositve Cache

Cache

Dentro de las pólizas de mapeo, hay que decidir cuando un bloque de información debe ser remplazado. Esto, se conoce como la **póliza de remplazo**. Hay dos algoritmos populares:

- **LRU (Least Recently Used)**: Este algoritmo mantiene una historia de los accesos a los bloques en la cache. El bloque utilizado menos reciente es remplazado.
- **FIFO (First In, First Out)**: En este algoritmo, el bloque que estuvo mas tiempo en la cache es remplazado.

Existen otros algoritmos, pero indiferente de cual se usa, el concepto de la póliza de remplazo es tratar de minimizar lo que se llama **trashing**. Trashing ocurre cuando un bloques es constantemente eliminado y re insertado al cache.

Cache

¿Que pasa cuando un dato es procesado en el CPU y se necesita guardar el resultado? ¿Se escribe en el cache? ¿Se escribe a memoria? ¿Las dos cosas?

Cuando el procesador decide escribir un resultado a memoria principal, el resultado puede ser redireccionado a la cache con el concepto de que probablemente sea re utilizado pronto.

Los bloques que fueron modificados se los llama **bloques sucios (dirty blocks)**. Si embargo, eventualmente la memoria principal tienen que ser actualizada con esta nueva información. La decisión de cuando se hace esta actualización se conoce como la **póliza de escritura del cache**.

2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Cache

Hay prácticamente dos tipos de pólizas:

Write-through: el bloque es escrito a cache y a memoria principal en simultaneo. Esto significa que cada escritura de cache es seguida por una escritura a memoria principal, que es mucho mas lento el proceso y consecuentemente ralentizado todo el sistema. Pero garantiza que los bloques en cache y memoria son los mismos.

Write-back: el bloque es escrito a memoria principal solo cuando es eliminado de la cache. Esto reduce los accesos a memoria y por ende el sistema funciona mas rápido. Pero hay que tener cuidado porque ahora el bloque en cache y el mismo de memoria no contienen la misma información. En sistemas con varios niveles de cache, el bloque eliminado es escrito al próximo nivel de cache y eventualmente a memoria principal.

2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Registros

El registro esta en el tope de la jerarquía de memoria. Puede guardar muy pocos datos, por lo general entre una a pocas palabras.

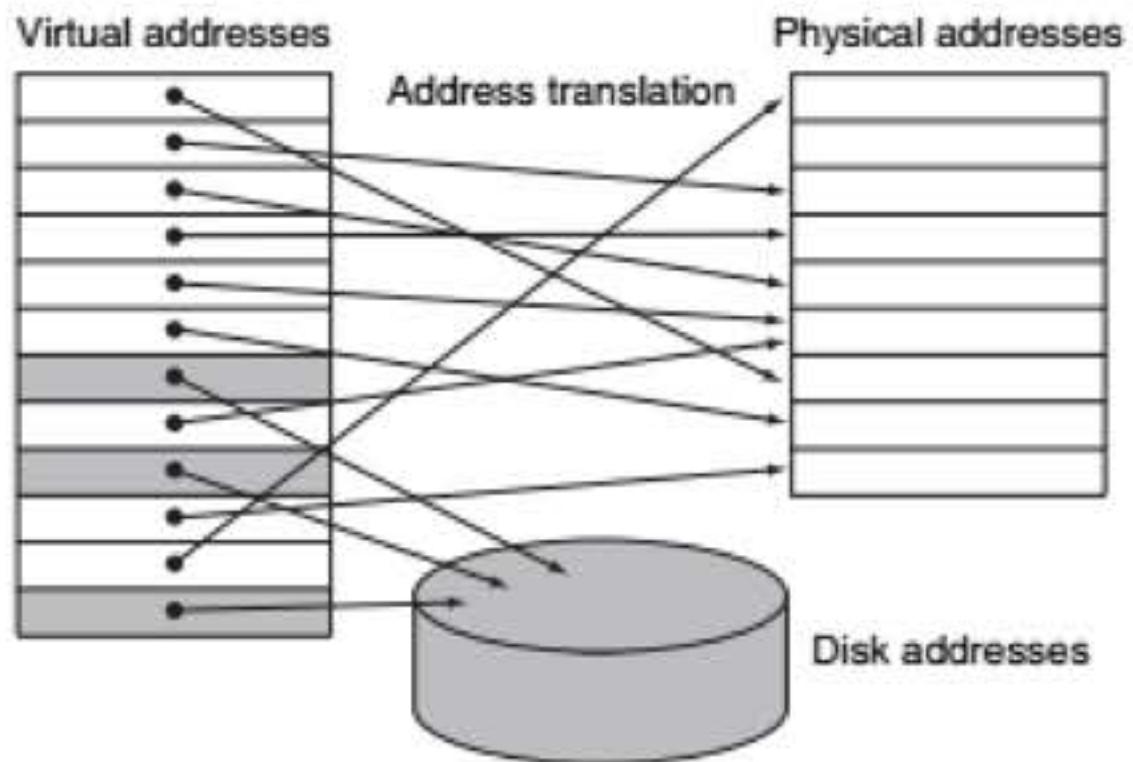
Son memorias temporales que contienen los datos que el procesador esta utilizando actualmente. Estos se encuentran adentro del procesador y los veremos en mas detalle mas adelante.

Es una memoria volátil.

Memoria Virtual

Memoria virtual es una técnica, que generalmente se usa a nivel del sistema operativo, que crea un mapa lógico de direcciones que se asocia a direcciones físicas en el sistema. Esta asociación puede pasar con la memoria principal del sistema como también con memoria secundaria.

Es común que en esta técnica, la memoria principal se piense como un “cache” de la memoria secundaria.



Memoria Virtual

Una motivación grande por el uso de memoria virtual, es que el espacio de memoria se puede definir mas grande que el espacio físico o real que provee la memoria principal.

¿Por que?:

En viejos tiempos, estaba a la responsabilidad del programador de dividir su programa en piezas mutuamente exclusivas para poder ser cargadas o descargadas de memoria. Estas piezas obviamente tenían que ser lo suficientemente chicas par entrar de forma completa en memoria.

Con memoria virtual, se permite que un programador pueda crear un programa que exceda el tamaño de la memoria principal y es la decisión del sistema operativo como distribuirla entre memoria principal y secundaria.

2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Memoria Virtual

Otra motivación importante es el incremento de seguridad que provee. En primer lugar, el sistema operativo le puede asignar un espacio virtual a cada programa. Este espacio virtual es exclusivo y solo accesible por el programa en cuestión.

En la traducción del espacio virtual al espacio físico, el sistema operativo pueden renforzar temas de seguridad y protección.

2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Memoria Virtual

El proceso de traducción entre el espacio virtual y el físico se llama **mapeo de dirección (Address Mapping)** o **traducción de dirección (Address Translation)**.

El concepto que cada programa tenga su propio espacio virtual de direcciones permite lo que se llama **reasignación (realocation)**, que permite que un programa sea cargado a cualquier lugar de memoria física.

Sistemas hoy reasignan un programa a bloques de tamaño fijo llamados **paginas (pages)**, facilitando la tarea del sistema operativo para encontrar bloques contiguos de memoria para cargar el programa. Con esto, solo tiene que encontrar la cantidad adecuada de paginas para el programa.

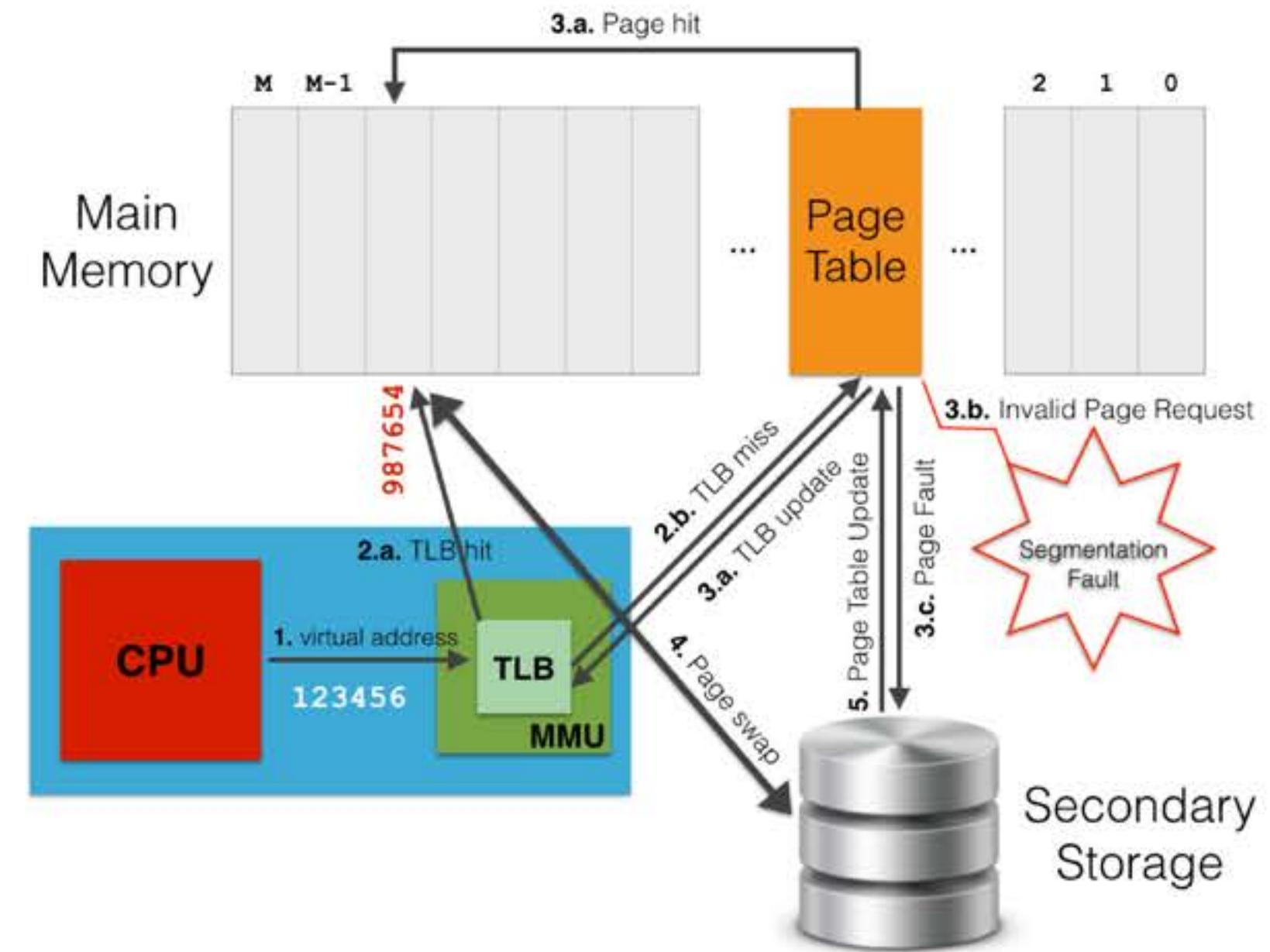
2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Memoria Virtual



2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Tecnologías de Memoria

Principalmente hay cuatro, que se usan en los diferentes niveles de la jerarquía de memoria.

- SRAM (Static Random Access Memory): caché y memorias cercanas al procesador.
- DRAM (Dynamic Random Access Memory): memoria principal.
- Flash Memory: memoria secundaria como SSD (Solid State Drive).
- Disco Magnético: memoria secundaria como HDD (Hard Disk Drive).

Memory technology	Typical access time	\$ per GiB in 2012
SRAM semiconductor memory	0.5–2.5 ns	\$500–\$1000
DRAM semiconductor memory	50–70 ns	\$10–\$20
Flash semiconductor memory	5,000–50,000 ns	\$0.75–\$1.00
Magnetic disk	5,000,000–20,000,000 ns	\$0.05–\$0.10

2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Tecnologías de Memoria

Random Access Memory (RAM) o memoria de acceso aleatorio, son memorias que pueden ser leídas o escritas...de forma aleatoria. El nombre puede ser un poco engañoso porque lo que es hoy, prácticamente cualquier tipo de memoria (no solo la RAM) tienen la capacidad de lectura y escrituras aleatorias. Cuando uno se refiere a la RAM de un sistema, en lo general se está refiriendo a la memoria principal del mismo.

Viene a dos variedades generales:

- Estática (Static RAM o SRAM)
- Dinámica (Dynamic RAM o DRAM)

2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

SRAM

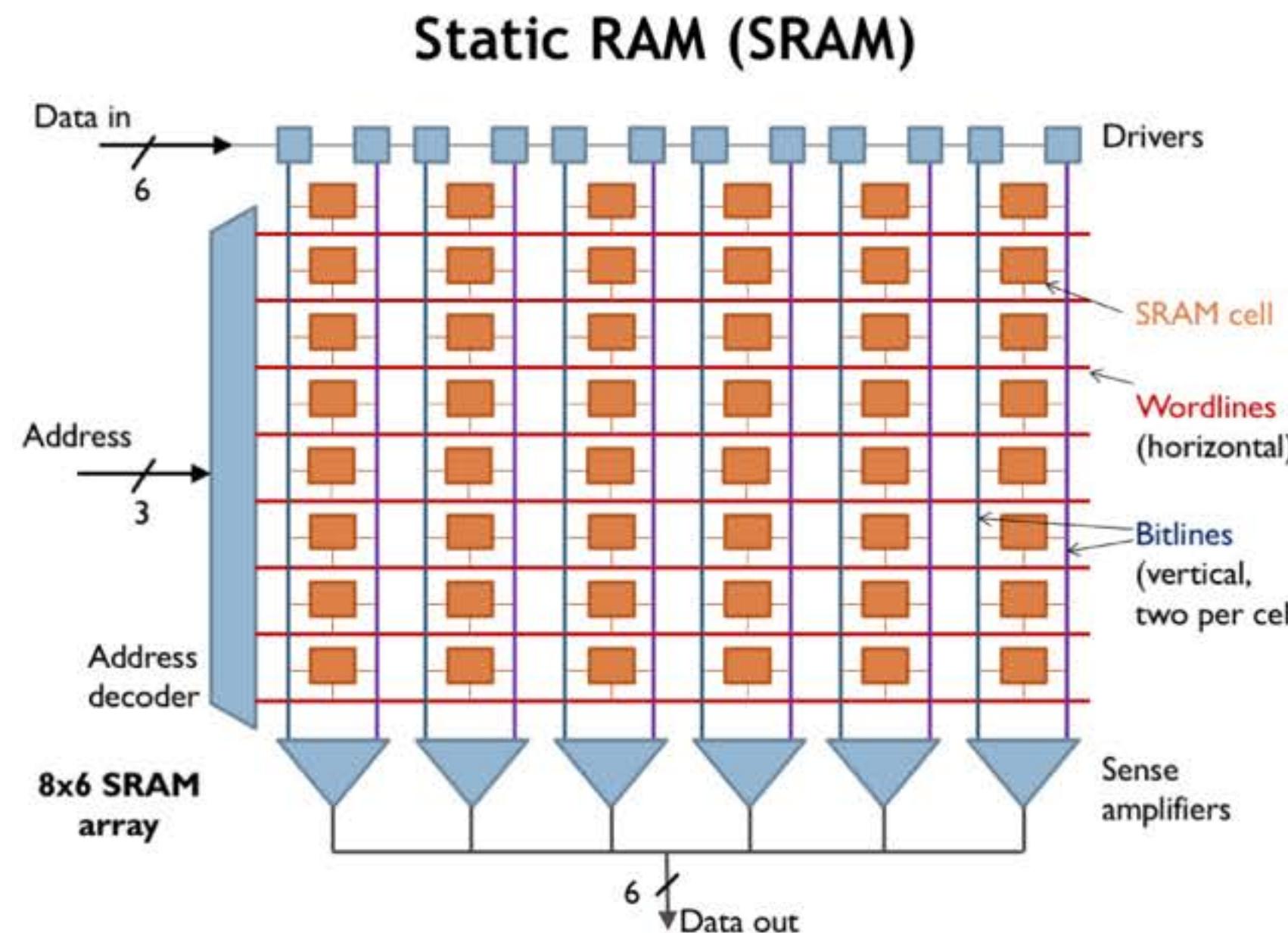
Son circuitos integrados formados por arreglos de circuitos de memoria (cada uno guarda un bit de información) con un puerto de acceso para hacer lecturas o escrituras. El tiempo de acceso de cualquier celda es fijo, aunque el tiempo de lectura y el de escritura pueden diferir. Este tiempo de acceso es muy cercano al ciclo del reloj.

Son memorias bi-estables, que significa que los dos estados (el 0 y el 1) son estables y no requieren un mecanismo para refrescar la información guardada. Para lograr esto se usan típicamente entre seis a ocho transistores para formar un bit de memoria.

Requiere una carga mínima de energía para funcionar y retener su información. Aunque se coloque el sistema en standby, si perdura la energía, perdura la información.

Sabían existir como chips separados, pero hoy en día se las encuentra integradas al procesador o dispositivo en cuestión.

SRAM



Aquí se puede ver un arreglo de 8x6, ósea, 8 direcciones de 6 bits, con una entrada de direcciones (3 bits), una entrada de datos (6 bits), y una salida de datos (6 bits).

Aquí la celda esta definida como 1 bit con una palabra de 6 bits. Solo se puede direccionar palabra por palabra y por ende leer/escribir de la misma forma.

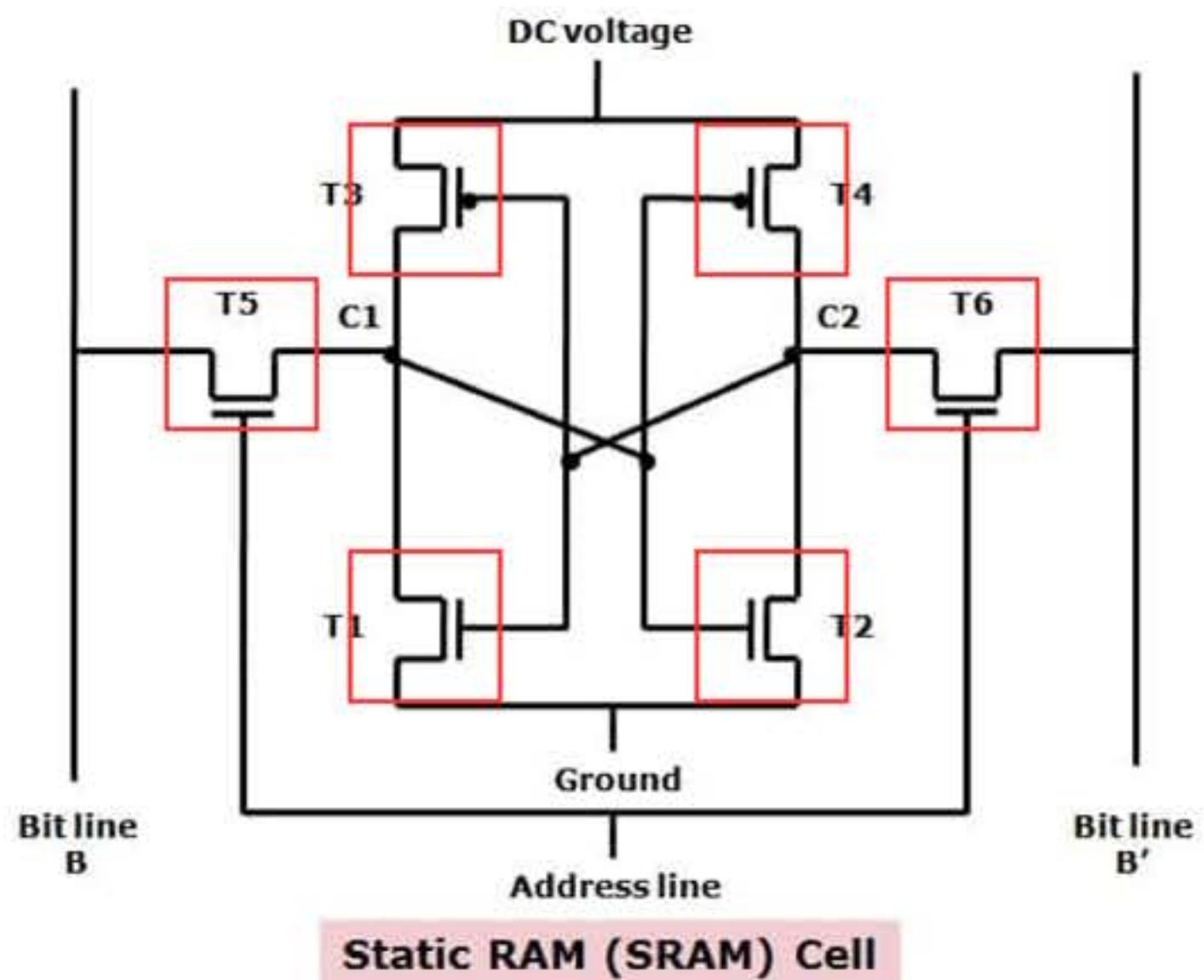
2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

SRAM



La esquemática de circuito de una celda de 1 bit de SRAM a nivel electronica. Se pueden observar 6 transistores.

DRAM

A diferencia de SRAM, donde el dato es guardado entre los transistores, con DRAM, se usa la carga de un capacitor (descargado es 0 mientras cargado es un 1). Este capacitor esta conectado a un solo transistor para poder hacer una lectura/escritura. No puede mantener su carga por mucho tiempo y debe ser periódicamente refrescada. Este mecanismos de refresco es lo que le da el nombre de memoria dinámica.

Como solo el estado 0 es estable (el 1 tienen que ser refrescado), se conocen como memorias mono-estable.

Al usar pocos transistores (a diferencia de SRAM), la DRAM es mucho mas barata y mas densa (mas memoria en un espacio físico definido).

Operaciones de lectura/escritura a DRAM tienen que ser sincronizadas al mecanismo de refresco y por ende los circuitos son mas complejos que causan que sean memoria mas lentas.

2024

Primer Cuatrimestre

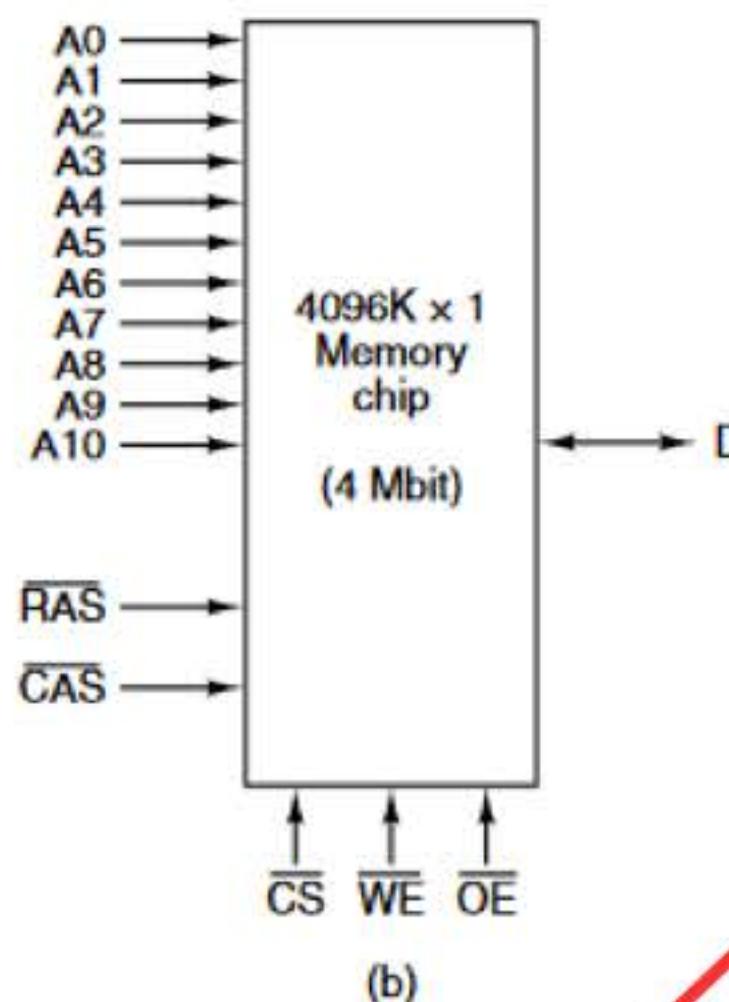


UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

DRAM

Aquí pueden ver un chip de DRAM de 4096Ki x 1, ósea, 4096Ki de direcciones de celdas de 1 bit por un total de 4 Mabit.

Contienen un bus de direccionamiento multiplexado de 11 bits para poder direccionar un total de 2^{22} bits (4×2^{20} , o 4 Mabit).



Se puede observar que cada celda esta compuesta por un transistor y un capacitor.

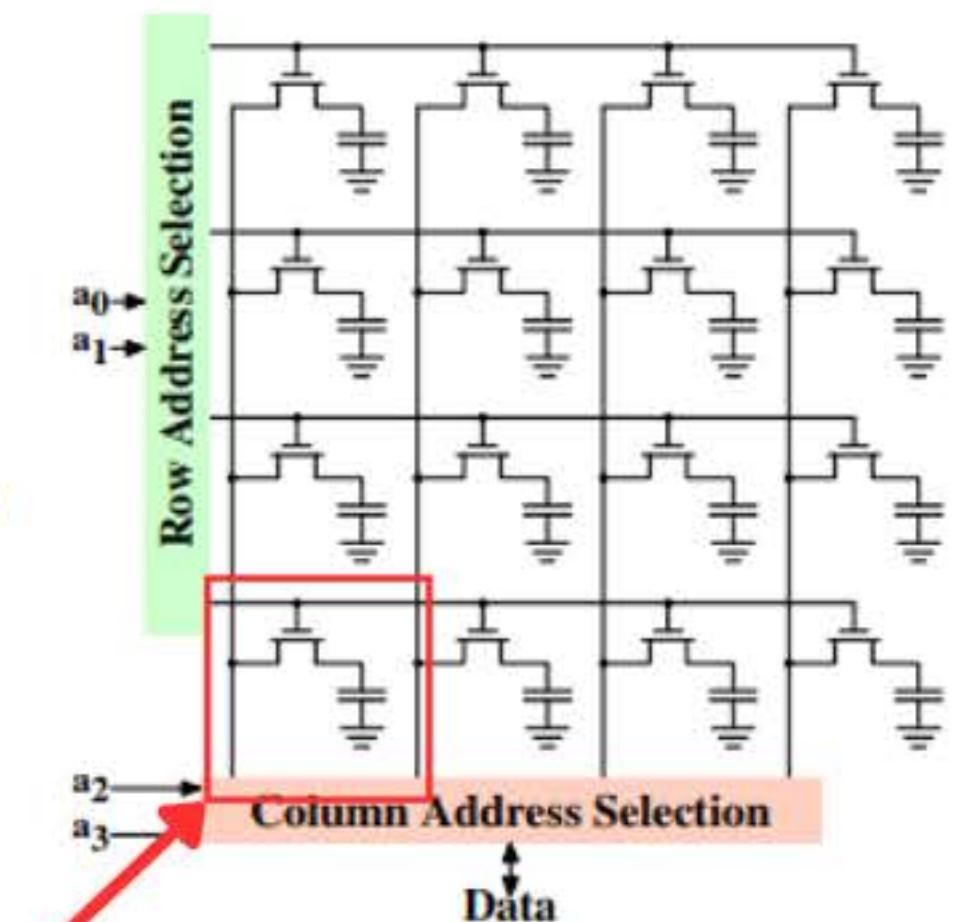
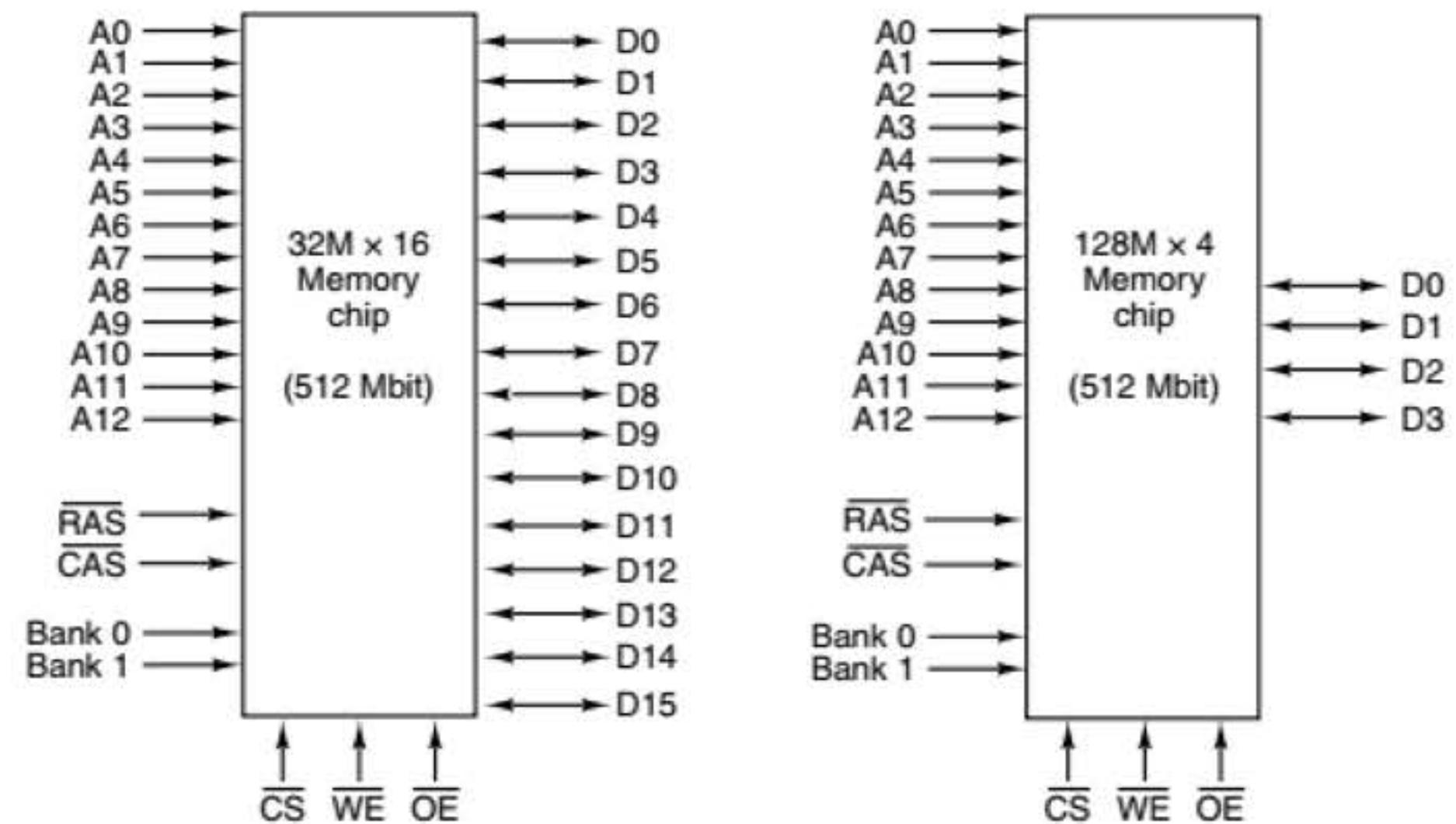


Figure 2.7: Dynamic RAM Schematic

DRAM

Dos chips de DRAM, uno de 32Mi x 16 bits y una de 128Mi x 4 bit por un mismo total de 512Mibits.



Tipos de DRAM

Fast Page Mode (FPM) DRAM: Internamente las celdas están organizadas de forma de matriz y cada celda se la dirige con dos decodificadores. Uno se hace cargo de seleccionar la fila 'RAS', mientras el otro la columna 'CAS'.

Usando la estrategia de dirigir primero la fila y después la columna también permitía multiplexar el bus de direccionamiento, ósea, se usa el mismo bus para 'RAS' y 'CAS'.

Usa señales de control que definen cuando tiene que responder a una operación, en otras palabras, corre de forma asíncrona al reloj principal del sistema.

Tipos de DRAM

Extended Data Output (EDO) DRAM: remplazo FPM DRAM y permitía comenzar una nueva operación de memoria mientras todavía se estaba resolviendo la previa operación. No reducía su latencia (que tan rápido acceda la memoria), pero si aumentaba su ancho de banda (más accesos por segundo).

Las EDO RAM también trabajan de forma asincrónica con el procesador. Esto funcionaba bien cuando la velocidad del procesador era mucho mas lento que la de la DRAM.

Tipos de DRAM

Cuando los procesadores aumentaron mucho mas la velocidad, la necesidad de tener DRAM mas rápida llevo a la creación del **Synchronous (S) DRAM**. SDRAM es una memoria híbrida entre SRAM y DRAM y esta sincronizada al reloj principal del sistema.

La gran ventaja de sincronizarla al procesador es que se podía eliminar circuitos complejos de control que manejaban cuando los chips de DRAM tenían que responder. Esto permitía maximizar la velocidad de transferencia de información entre la memoria principal y el procesador.

La próxima innovación después del SDRAM fue la **Double Data Rate (DDR) SDRAM**. Es el tipo de memoria mas usada hoy en día (actualmente tenemos DDR generación 5). La misma puede hacer una operación de memoria tanto en el flanco ascendente del reloj como en el flanco descendente, efectivamente duplicando su velocidad.

2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Tipos de DRAM

Standard	Clock rate (MHz)	M transfers per second	DRAM name	MB/sec /DIMM	DIMM name
DDR	133	266	DDR266	2128	PC2100
DDR	150	300	DDR300	2400	PC2400
DDR	200	400	DDR400	3200	PC3200
DDR2	266	533	DDR2-533	4264	PC4300
DDR2	333	667	DDR2-667	5336	PC5300
DDR2	400	800	DDR2-800	6400	PC6400
DDR3	533	1066	DDR3-1066	8528	PC8500
DDR3	666	1333	DDR3-1333	10,664	PC10700
DDR3	800	1600	DDR3-1600	12,800	PC12800
DDR4	1066–1600	2133–3200	DDR4-3200	17,056–25,600	PC25600

2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

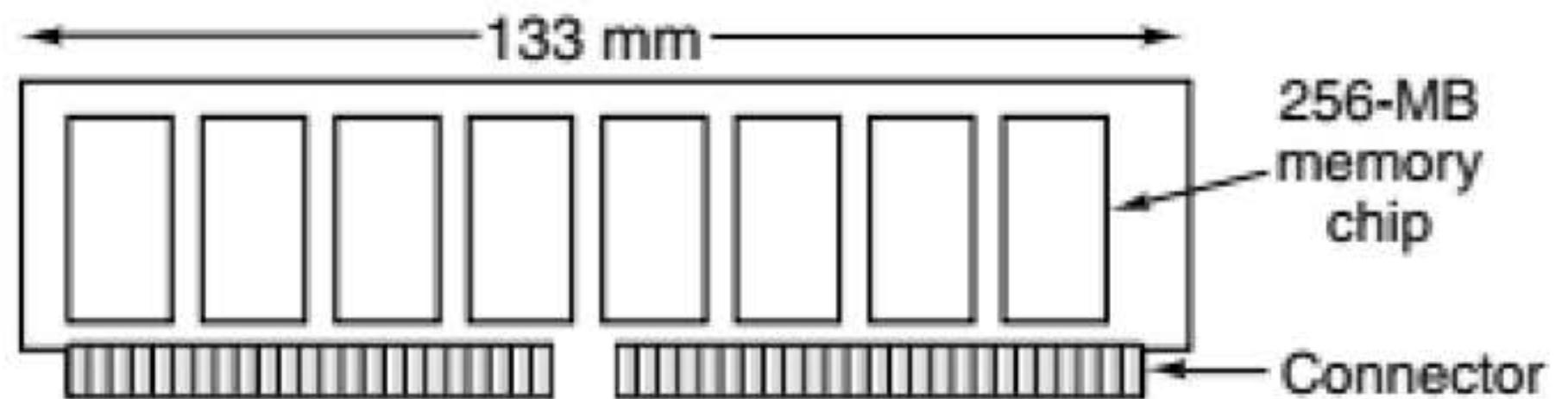
Tipos de DRAM

Year introduced	Chip size	\$ per GiB	Total access time to a new row/column	Average column access time to existing row
1980	64 Kibibit	\$1,500,000	250 ns	150 ns
1983	256 Kibibit	\$500,000	185 ns	100 ns
1985	1 Mebibit	\$200,000	135 ns	40 ns
1989	4 Mebibit	\$50,000	110 ns	40 ns
1992	16 Mebibit	\$15,000	90 ns	30 ns
1996	64 Mebibit	\$10,000	60 ns	12 ns
1998	128 Mebibit	\$4,000	60 ns	10 ns
2000	256 Mebibit	\$1,000	55 ns	7 ns
2004	512 Mebibit	\$250	50 ns	5 ns
2007	1 Gibibit	\$50	45 ns	1.25 ns
2010	2 Gibibit	\$30	40 ns	1 ns
2012	4 Gibibit	\$1	35 ns	0.8 ns

Tipos de DRAM

Hasta los años 1990, las memorias eran fabricadas en chips únicos. Las computadoras tenían varios enchufes donde se podían agregar más chips de entre 1Kb y 1Mb. Desde los 90s, los chips de memoria se empezaron a organizar de otra forma. Típicamente se agrupaba 8 o 16 chips en una placa y todo vendido como una sola unidad, o modulo.

El modulo viene en dos formas **SIMM (Single Inline Memory Module)** y **DIMM (Dual Inline Memory Module)** dependiendo si había conectores en un solo lado (Single) o los dos lados (Dual) de la placa.



Tipos de DRAM

Los módulos de RAM pueden venir con ECC y sin (no-ECC). ECC significa **Error-Correcting Code**, y su utilidad es corregir errores en caso que los datos se corrompan. Estos módulos son mas caros y solo se usan en placas madres de servidores donde es esencial evitar estos tipos de errores.

Cada modulo tienen una serie de chips que proveen la memoria. Estos chips de 8-bits están organizados en múltiples de 8 para módulos no-ECC y 9 para ECC. Esta combinación de chips permite la lectura de x8 en paralelo, y forman los 64bits que requiere el bus (72bits para los bus con ECC).

Organizarlo por 8 (o 9) chips también permite aumentar la capacidad del modulo, porque cada grupo de 8 proporcionalmente aumentaría su capacidad (por ejemplo, si 8 chips dan 512MiB (64MiB x 8) de memoria, 16 chips dan 1GiB). Cada grupo de ocho se conoce como un **banco de memoria (memory rank)**, un modulo con 8 chips tienen un solo banco de memoria mientras un modulo de 16, dos bancos de memoria. En el caso de DDR, podía tener un máximo de 32 chips (36 para ECC), o cuatro bancos de memoria.

2024

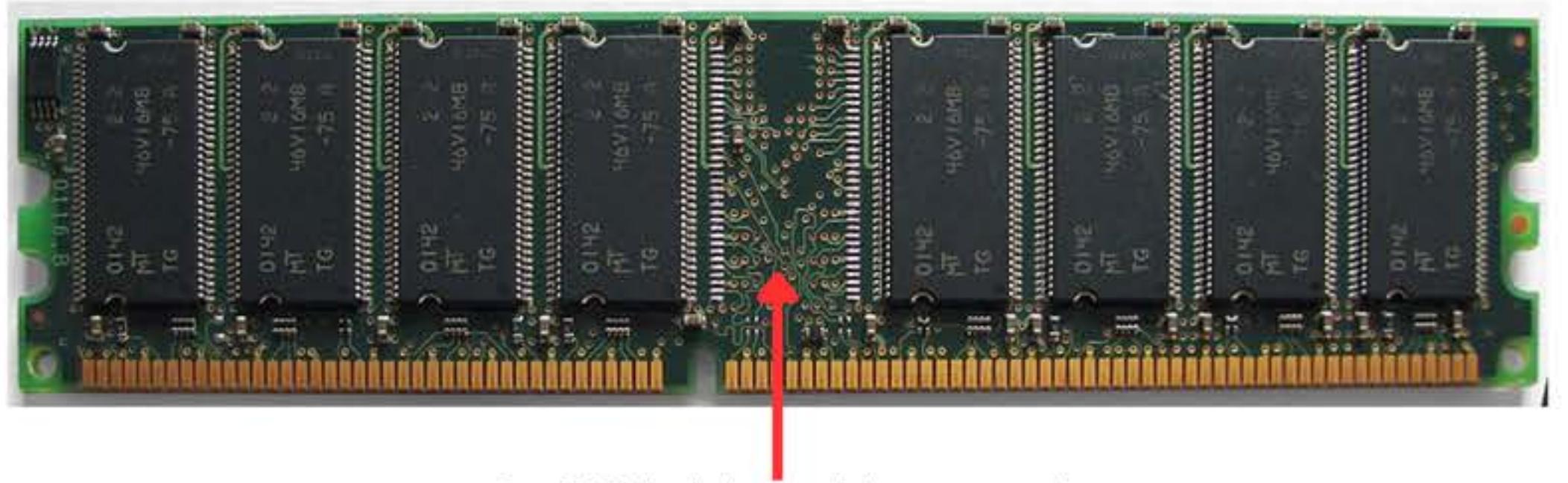
Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

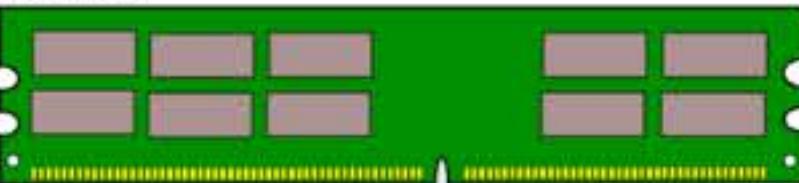
Tipos de DRAM

DDR SDRAM es el principal tipo de RAM en mercado. En si, de lo que es 2023, la versión mas reciente es DDR5. Desafortunadamente, cada nueva versión es incompatible con la previa, obligando a actualizar el resto de los componentes si se desea usar RAM mas nueva.

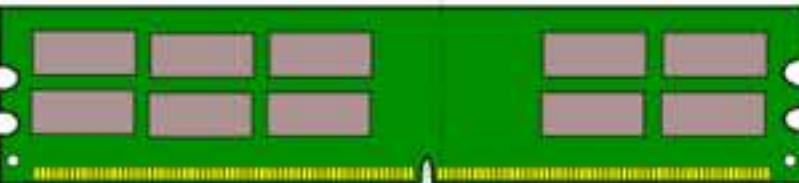


Con ECC habría un chip mas aquí.

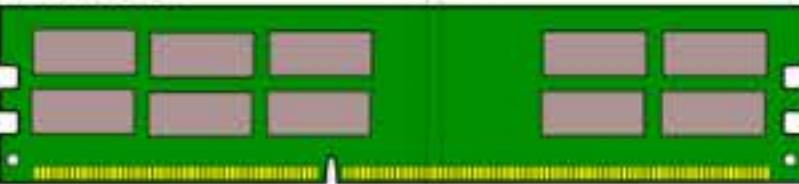
DDR



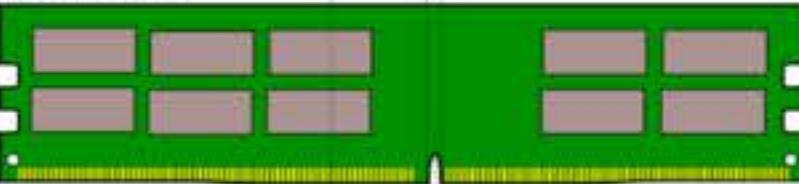
DDR 2



DDR 3



DDR 4



2024

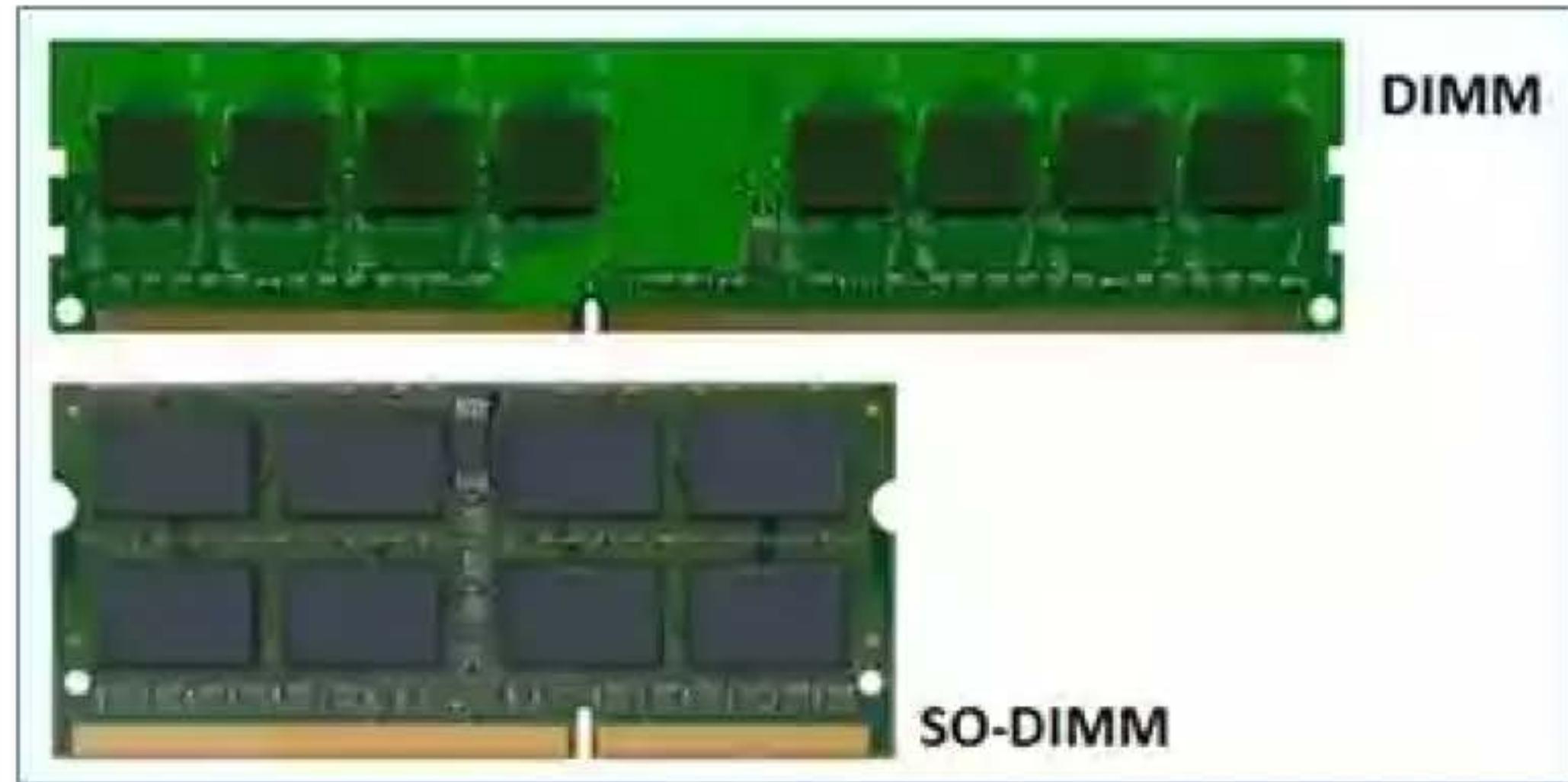
Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Tipos de DRAM

Existen DIMMs mas pequeñas, utilizados generalmente para notebook, llamadas **SO-DIMM (Small Outline DIMM)**.



2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Flash Memory

Estas son típicamente algún tipo de EEPROM (Electrically Erasable Programmable Read-Only Memory). Las PROM originales (Programmable Read-Only Memory) eran memoria que se las podía "programar" o escribir una sola vez y quedaban consecuentemente de solo lectura. La variante EEPROM tienen la capacidad de poder ser borrada y reprogramada a través de señales eléctricas.

La operación de escritura gasta la EEPROM, lo que los fabricantes siempre en las especificaciones aclaran cuál es la cantidad máxima de escrituras que pueden tener las celdas de este tipo de memoria. Por estas razones, los productos de esta tecnología tienen controladores con algoritmos inteligentes que esparcen las escrituras, con el propósito de evitar que la misma celda sea escrita repetidamente. Esto incrementa mucho la vida útil pero a un costo de velocidad de escritura.

Los SSD de hoy usan tecnologías más modernas y rápidas que las EEPROM pero con conceptos similares.

2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Discos Magnéticos

Los discos magnéticos están compuestos por una colección de platos que rotan a una velocidad entre 5400 y 15.000 rpm. Estos platos tiene sus dos lados cubiertos por un material magnético que se usa para guardar información, similar a como funciona el cassettes.

La superficie del plato esta dividida en miles de círculos concéntricos, llamados pistas, y cada pista esta dividida en sectores, donde se guarda la información. Cada pista también puede tener miles de sectores, y cada sector puede guardar entre 512 y 4096 bytes de información (depende del disco y del sistema operativo). El sector es la cantidad minima de información que se puede leer o escribir a la vez (concepto similar a la celda de memoria principal).

Para acceder a la información, el sistema operativo toma tres pasos.

1. Primero pone el lector sobre la pista adecuada. Esto se llama búsqueda (seek).
2. Cuando encontró la pista, el lector tienen que esperar es sector deseado.
3. Transferir la información guardada en el sector. Repetir para todos los sectores necesarios.

El tiempo de acceso puede variar dependiendo en el tamaño del sector, la velocidad de rotación y la densidad de la pista.

2024

Primer Cuatrimestre

Otros tipos de Memoria

Existen otro tipo de familia de memoria llamadas **ROM (Read-Only Memory)**. Su concepto es que sea una memoria de solo lectura, aunque su acceso a la información sigue siendo de forma aleatoria. Este tipo de memoria es no-volátil.

La necesidad de tener un componente de memoria que siempre tenga los mismos datos, a pesar que se apague la maquina, trajo al desarrollo de esta tecnología. Por ejemplo, el **BIOS (Basic Input Output System)**, que se hace cargo de arrancar el hardware y el sistema operativo de la maquina cuando uno la prende, es una memoria ROM.

Durante los años salieron varias tecnologías de ROM:

- ROM: la original, salía programada desde fabrica.
- PROM: programable con circuitos especializados pero solo una vez, consecuentemente quedaba de solo lectura.
- EPROM: se puede re-programar usando un circuito especializado que utiliza luz ultravioleta.
- EEPROM: se puede re-programar con señales eléctrica.
- Memoria Flash: la evolución del EEPROM.

2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Resumen Grafico

Type	Category	Erasure	Byte alterable	Volatile	Typical use
SRAM	Read/write	Electrical	Yes	Yes	Level 2 cache
DRAM	Read/write	Electrical	Yes	Yes	Main memory (old)
SDRAM	Read/write	Electrical	Yes	Yes	Main memory (new)
ROM	Read-only	Not possible	No	No	Large-volume appliances
PROM	Read-only	Not possible	No	No	Small-volume equipment
EPROM	Read-mostly	UV light	No	No	Device prototyping
EEPROM	Read-mostly	Electrical	Yes	No	Device prototyping
Flash	Read/write	Electrical	No	No	Film for digital camera

2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA

Lecturas

Lecturas recomendadas:

Tanenbaum, 2013: Capítulos 2.2.2, 2.2.3, 2.2.5, 2.3.1

Tanenbaum, 2000: Capítulos 2.2.2, 2.2.3, 2.2.5, 2.3.1

2024

Primer Cuatrimestre



UNIVERSIDAD NACIONAL DE TUCUMÁN
facet
FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA