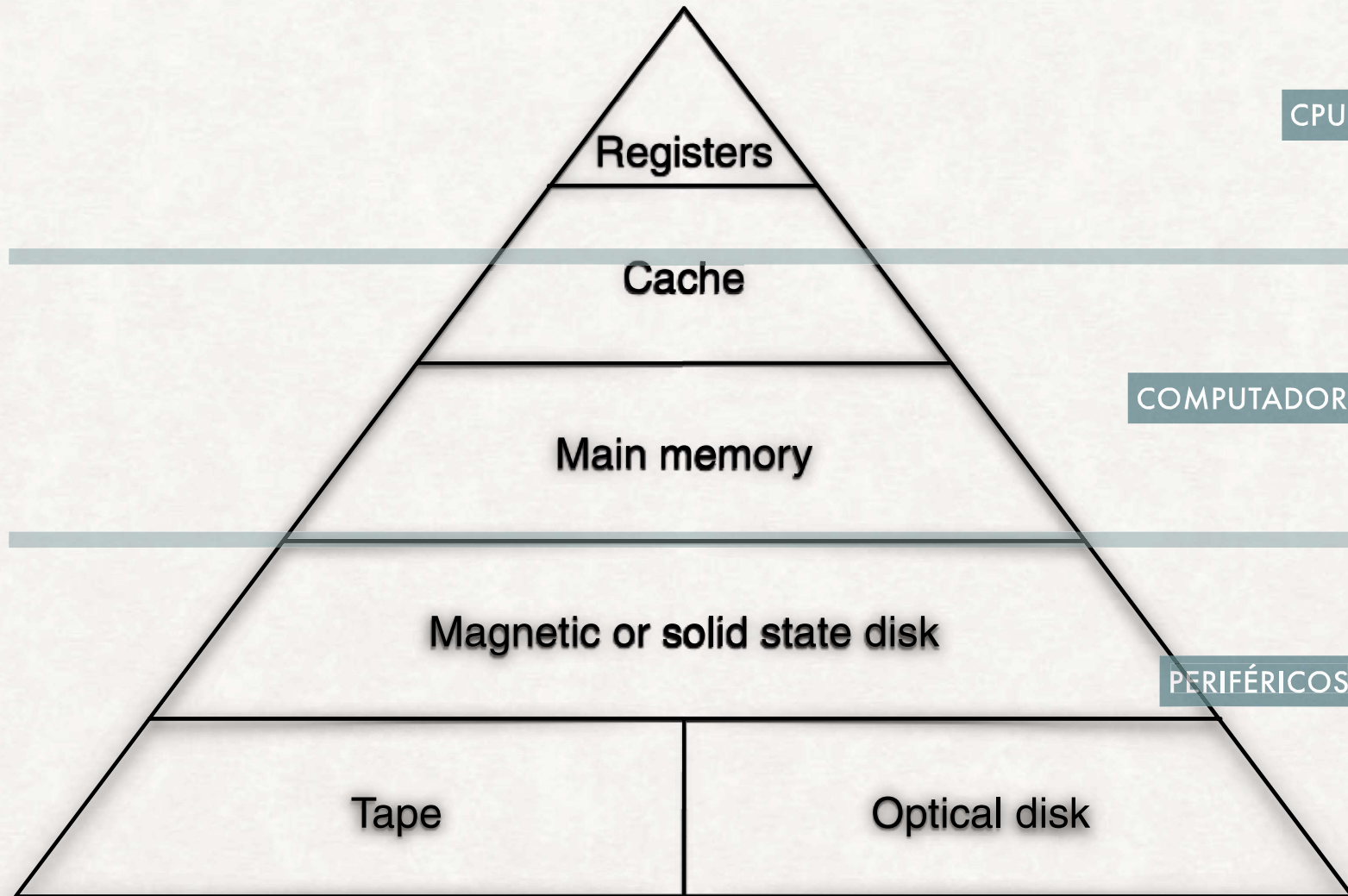


TEMA 2
DISPOSITIVOS DE
MEMORIA

JERARQUÍA DE MEMORIA

CINCO NIVELES:

RAPIDEZ
DE ACCESO



CPU

Registers

Cache

COMPUTADOR

Main memory

Magnetic or solid state disk

PERIFÉRICOS

Tape

Optical disk

REGISTROS INTERNOS DE LA CPU

TIPOS DE REGISTROS

- Toda CPU cuenta con un conjunto de *registros internos* con diversos propósitos:
 - Acumulador: operando y almacén de resultados de la ALU.
 - Registros genéricos de datos (en algunas CPUs cualquiera de ellos puede actuar de acumulador).
 - Registros de direcciones (también llamados registros índice).
 - Contador de programa (PC) (también llamado *instruction pointer (IP)*).
 - Apuntador de pila (stack pointer SP).
 - Registro de indicadores de condición de la ALU (conocido como *Program Status Word (PSW)*, *Condition Code Register (CCR)*, etc.).
 - Registro de estado (control de interrupciones, nivel de privilegio, etc.).
 - Registro de instrucción (usualmente no accesible desde programa).

REGISTROS INTERNOS DE LA CPU

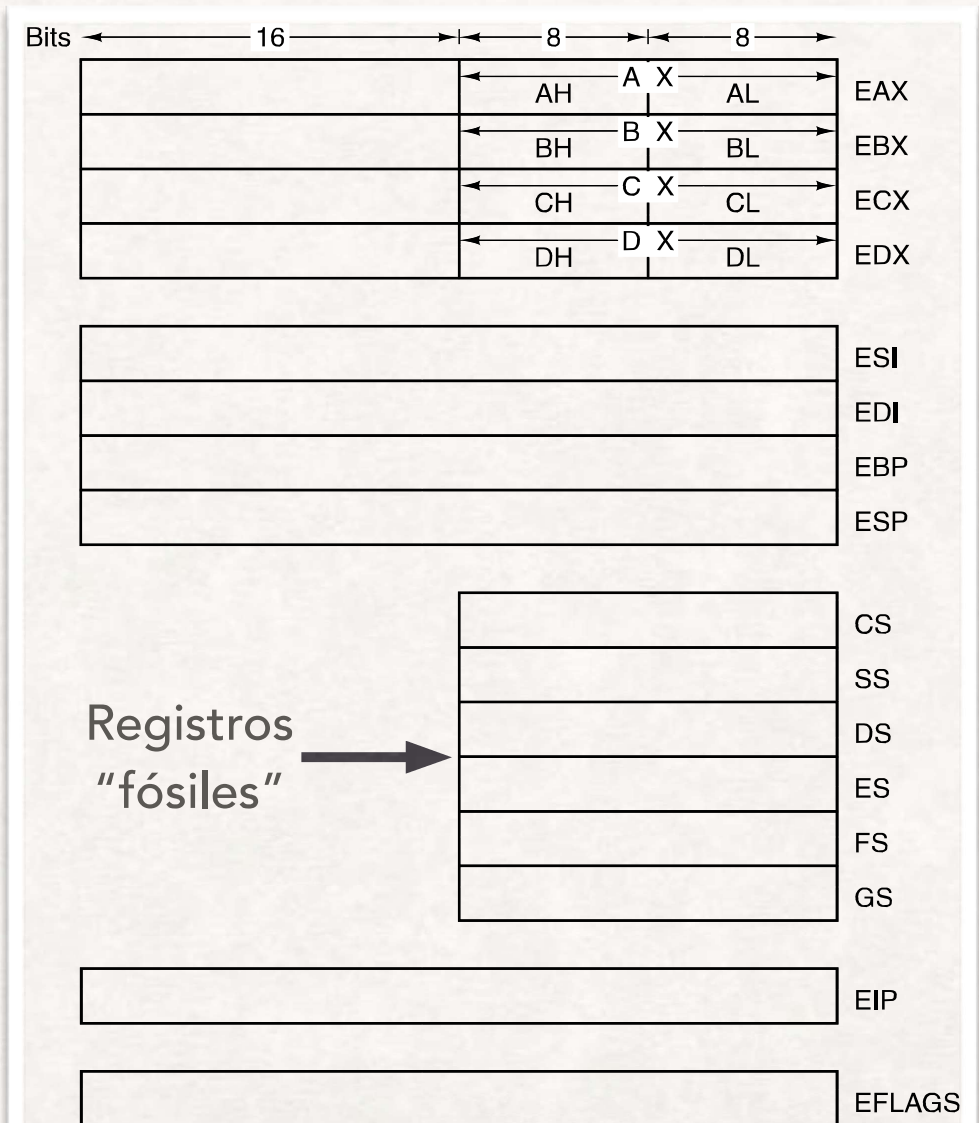
EJEMPLO: IA-32 (INTEL 80386 Y POSTERIORES)

AMD64/EM64T/x86-64:

Los registros "E" de IA-32 se extienden a 64 bits, con prefijo "R", y se añaden nuevos registros R8 a R15.

AMD Opteron (2003)

Intel Xeon "Nocona" (2004)



REGISTROS INTERNOS DE LA CPU

EJEMPLO: ARMv7

Register	Alt. name	Function
R0–R3	A1–A4	Holds parameters to the procedure being called
R4–R11	V1–V8	Holds local variables for the current procedure
R12	IP	Intraprocedure call register (for 32-bit calls)
R13	SP	Stack pointer
R14	LR	Link register (return address for current function)
R15	PC	Program counter

16 registros, de ellos 12 de propósito general.

REGISTROS INTERNOS DE LA CPU

EJEMPLO: FAMILIA M68000

Registros de datos

	31	15	7	0
D0				
D1				
D2				
D3				
D4				
D5				
D6				
D7				

Registros de direcciones

	31	15	0
A0			
A1			
A2			
A3			
A4			
A5			
A6			
A7		(USP)	
		(SSP)	



	31	0
PC		

MEMORIA CACHÉ

APLICACIÓN Y CARACTERÍSTICAS

- Su función es acelerar los accesos a la memoria principal, actuando como un almacenamiento intermedio.
- Se construye con dispositivos de memoria de acceso mucho más rápido, pero el mayor coste limita significativamente su tamaño.
- Se estructura en *líneas*, cada una conteniendo un bloque de datos. Las transferencias entre caché y memoria principal se realizan por líneas.
- Varios niveles (típicamente tres). L1 y L2 suelen situarse *on chip*. L3 puede ser *on chip* u *off chip* según el caso.
- Pueden ser *separados* (uno para instrucciones y otro para datos) o *unificados*.
- Si el chip contiene varios núcleos, éstos comparten el caché L3.

MEMORIA CACHÉ

EJEMPLO: INTEL "SANDY BRIDGE" (2011)

- Nivel 1: 64 KB por núcleo (32 KB datos + 32 KB instrucciones), *on chip*.
- Nivel 2: 256 KB por núcleo, *on chip*.
- Nivel 3: compartido entre todos los núcleos, *on chip*.
 - Core i3, i5, i7: de 1 a 8 MB (también compartido por la GPU).
 - Xeon E3, E5: de 3 a 20 MB.
- Tamaño de línea: 64 bytes.

MEMORIA PRINCIPAL

RAM Y ROM

- **RAM: *Random Access Memory***
 - Acceso aleatorio en contraposición a secuencial. En realidad *toda* la memoria principal (sea RAM o ROM) es de acceso aleatorio.
 - Se usa para almacenamiento de código y datos de los programas en ejecución.
 - Propiedades fundamentales:
 - Permite la **lectura y escritura** de cada elemento.
 - Es **volátil** (pierde su contenido al retirarle la alimentación).
- **ROM: *Read-Only Memory***
 - En realidad, muchos tipos permiten la escritura, pero con importantes limitaciones respecto a la lectura (la cual se realiza del mismo modo que con la RAM).
 - Su propiedad fundamental: es **no volátil**.
 - Suele contener el código necesario para realizar la configuración inicial al encender el computador y cargar el núcleo (*kernel*) del sistema operativo.

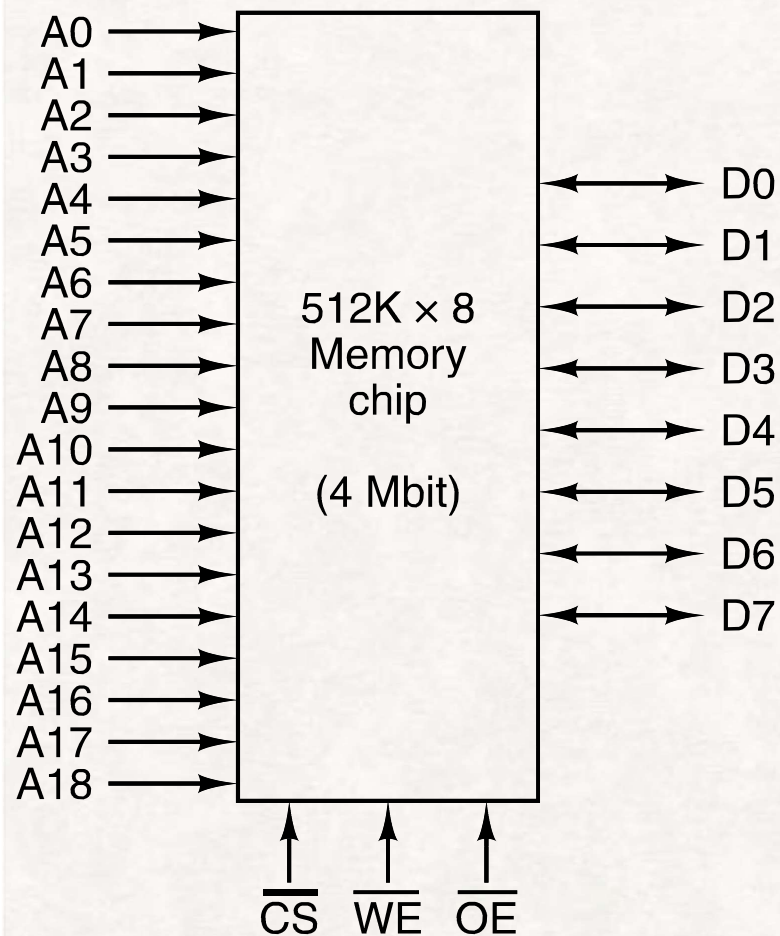
MEMORIA RAM

ESTÁTICA O DINÁMICA

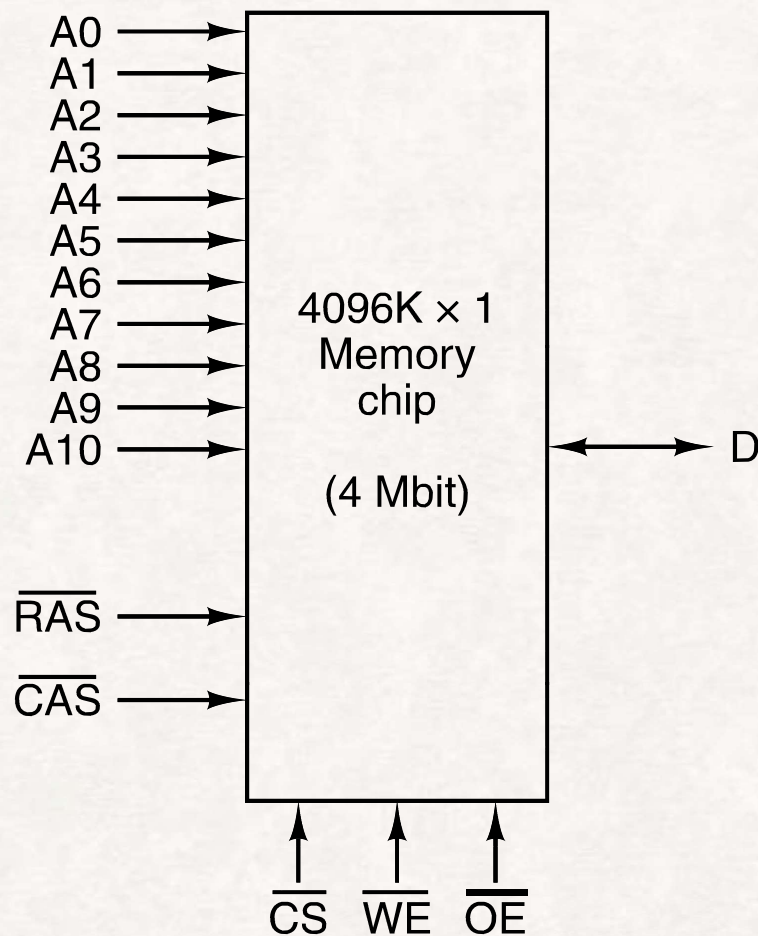
- **RAM estática (SRAM):**
 - Cada bit se almacena en un *flip-flop* o biestable, lógica TTL.
 - Se estructuran como un vector de registros de una determinada anchura de bits.
 - Pros: extremadamente rápida (t_{acc} muy bajo), acceso muy sencillo, alta fiabilidad.
 - Contras: baja densidad, alto precio, alto consumo de energía (durante los accesos).
 - Uso: registros de la CPU, memoria caché.
- **RAM dinámica (DRAM):**
 - Cada bit se almacena como la carga de un condensador: sólo se necesita un transistor y un condensador por bit.
 - Se estructuran como una matriz. Cada fila se puede dividir en grupos de una determinada anchura de bits.
 - Necesita un *refresco* externo (típicamente cada 64 ms) para recargar los condensadores, lo cual requiere soporte hardware específico, e implica un consumo constante.
 - Gran densidad de almacenamiento; uso generalizado como memoria principal.
 - Acceso más complejo, en dos fases: selección de fila y selección de columna. Además pueden contener varios *bancos* superpuestos.

MEMORIA RAM

ORGANIZACIÓN INTERNA Y ACCESO (I)



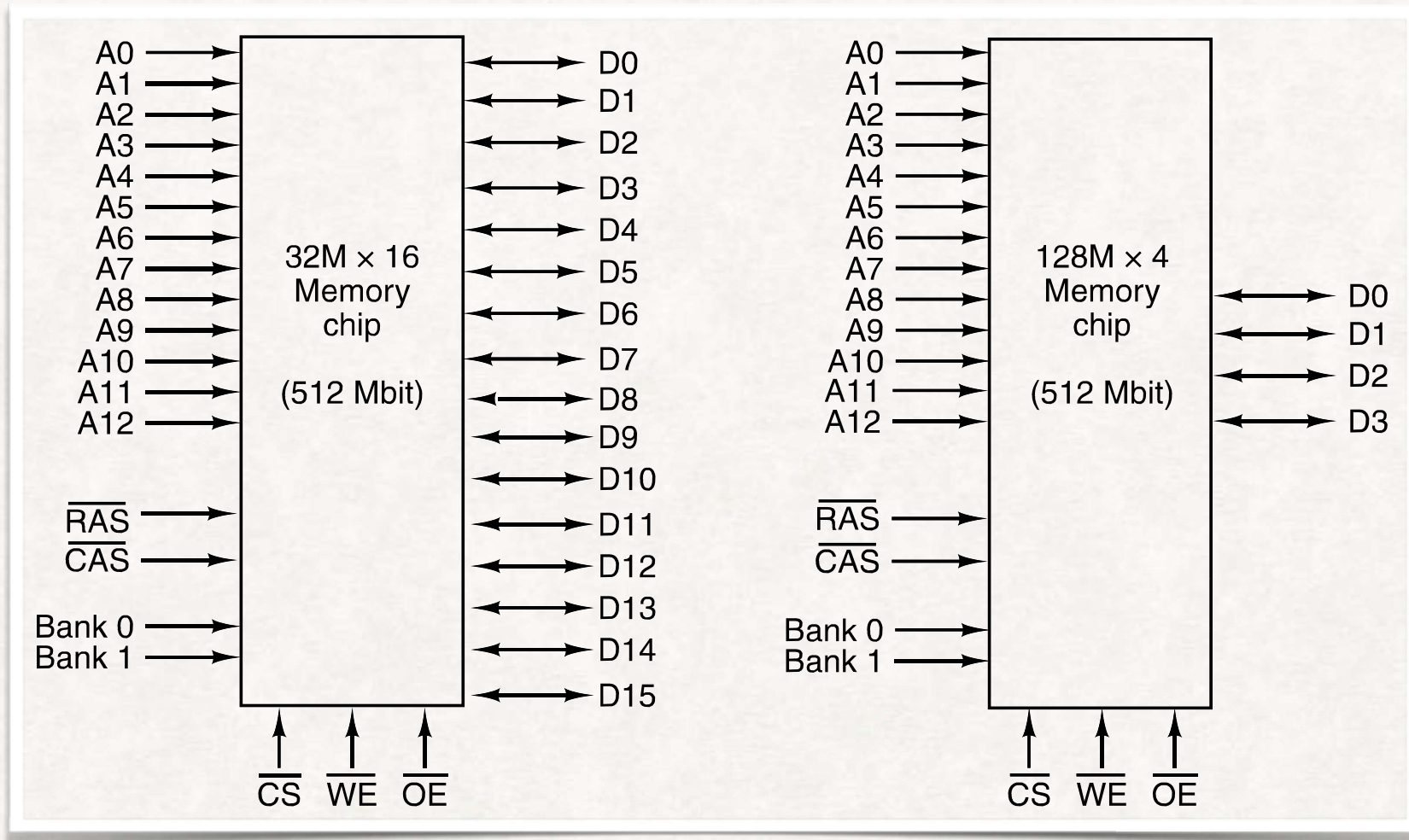
Vector de 512 K (2^{19}) bytes
19 líneas de direccionamiento



Matriz de 2048x2048 ($2^{11} \times 2^{11}$) bits
11 líneas para \overline{RAS} y 11 para \overline{CAS}

MEMORIA RAM

ORGANIZACIÓN INTERNA Y ACCESO (II)



4 bancos de 8192×1024 ($2^{13} \times 2^{10}$) \times 16 bits
13 líneas para $\overline{\text{RAS}}$ y 10 líneas para $\overline{\text{CAS}}$
2 líneas para selección de banco

4 bancos de 8192×4096 ($2^{13} \times 2^{12}$) \times 4 bits
13 líneas para $\overline{\text{RAS}}$ y 12 líneas para $\overline{\text{CAS}}$
2 líneas para selección de banco

MEMORIA RAM

TIPOS DE DISPOSITIVOS DRAM

- **Fast Page Mode (FPM):** Permite la selección secuencial de columnas consecutivas en la misma fila (manteniendo \overline{RAS}). Funcionamiento asíncrono.
- **Extended Data Output (EDO):** Permite iniciar un nuevo ciclo de acceso antes de terminar el anterior. Más ancho de banda con la CPU. Funcionamiento asíncrono.
- **Synchronous DRAM (SDRAM):** Funcionamiento sincronizado con CLK. La CPU le indica cuántos ciclos de acceso ha de realizar y la pone en marcha. Mayor ritmo de transferencia de datos.
- **Double Data Rate (DDR, DDR2, DDR3,...) SDRAM:** Realiza cada transferencia sincronizada tanto con el flanco de subida como el de bajada de CLK. Cada generación DDR opera con frecuencias más altas de CLK.
 - DDR3: 1067 MHz, 2133 MT/s, hasta 17067 MB/s.
 - DDR4: 1600 MHz, 3200 MT/s, hasta 25600 MB/s.

MEMORIA ROM

TIPOS DE MEMORIA ROM

- **ROM preprogramada:** contenido grabado en silicio.
- **PROM (Programmable ROM):** Programable eléctricamente (con un dispositivo específico) sólo una vez.
- **EPROM (Erasable PROM):** Borrable mediante exposición a luz UV para ser reprogramada. Sólo resisten cierto número de ciclos de borrado y reprogramación.
- **EEPROM (Electrically Erasable PROM):** Borrable y reprogramable (byte a byte) "*in place*" mediante señales específicas.
- **Flash ROM:** Tipo de EEPROM borrable y reprogramable por bloques, mucho más rápida y económica que la EEPROM.

MEMORIA FLASH

TIPOS Y APLICACIONES

- **Flash NOR:**
 - Lectura igual que una RAM (acceso aleatorio).
 - Escritura (reprogramación) también acceso aleatorio, sólo se pueden cambiar bits 1 a 0.
 - Borrado por bloques. El borrado pone todos los bits del bloque a 1.
 - Tamaños típicos de bloque: 64, 128, 256 KB.
 - Aplicación: memoria principal no volátil (típicamente para el *firmware*).
- **Flash NAND:**
 - Se estructuran en bloques compuestos por páginas. Ejemplo: bloque de 512 KB formado por 128 páginas de 4 KB.
 - Lectura y escritura (reprogramación sólo 1→0) por páginas.
 - Borrado sólo por bloques; pone todos los bits de todas las páginas del bloque a 1.
 - Mayor densidad de almacenamiento y menor coste que la NOR.
 - Borrado, lectura secuencial y escritura secuencial más rápida que la NOR.
 - Aplicación: memoria secundaria (periféricos de almacenamiento): SSD, USB stick, etc.

MEMORIA FLASH

LÍMITE DE CICLOS DE BORRADO-REESCRITURA

- Como toda EEPROM, las memorias Flash tienen un límite en el nº de ciclos de borrado/escritura:
 - SLC NAND: 100.000 borrados por bloque.
 - MLC NAND: 1.000 a 10.000.
 - SLC NOR: 100.000 a 1.000.000
 - MLC NOR: 100.000.
 - Ejemplos:
 - Samsung SSD 850 PRO (MLC): 6.000
 - Samsung 850 EVO (TLC): 1.000
 - Crucial MX300 (TLC): 3.000.

SLC: single-level cell (1 bit/célula). MLC: 2 bits/célula. TLC: 3 bits/célula. QLC: 4 bits/célula
SLC: más rápidas y caras; MLC, TLC, QLC: más lentas y económicas, y mayor tasa de error.

MEMORIA FLASH

ESTRATEGIAS PARA ALARGAR VIDA ÚTIL Y PRESTACIONES

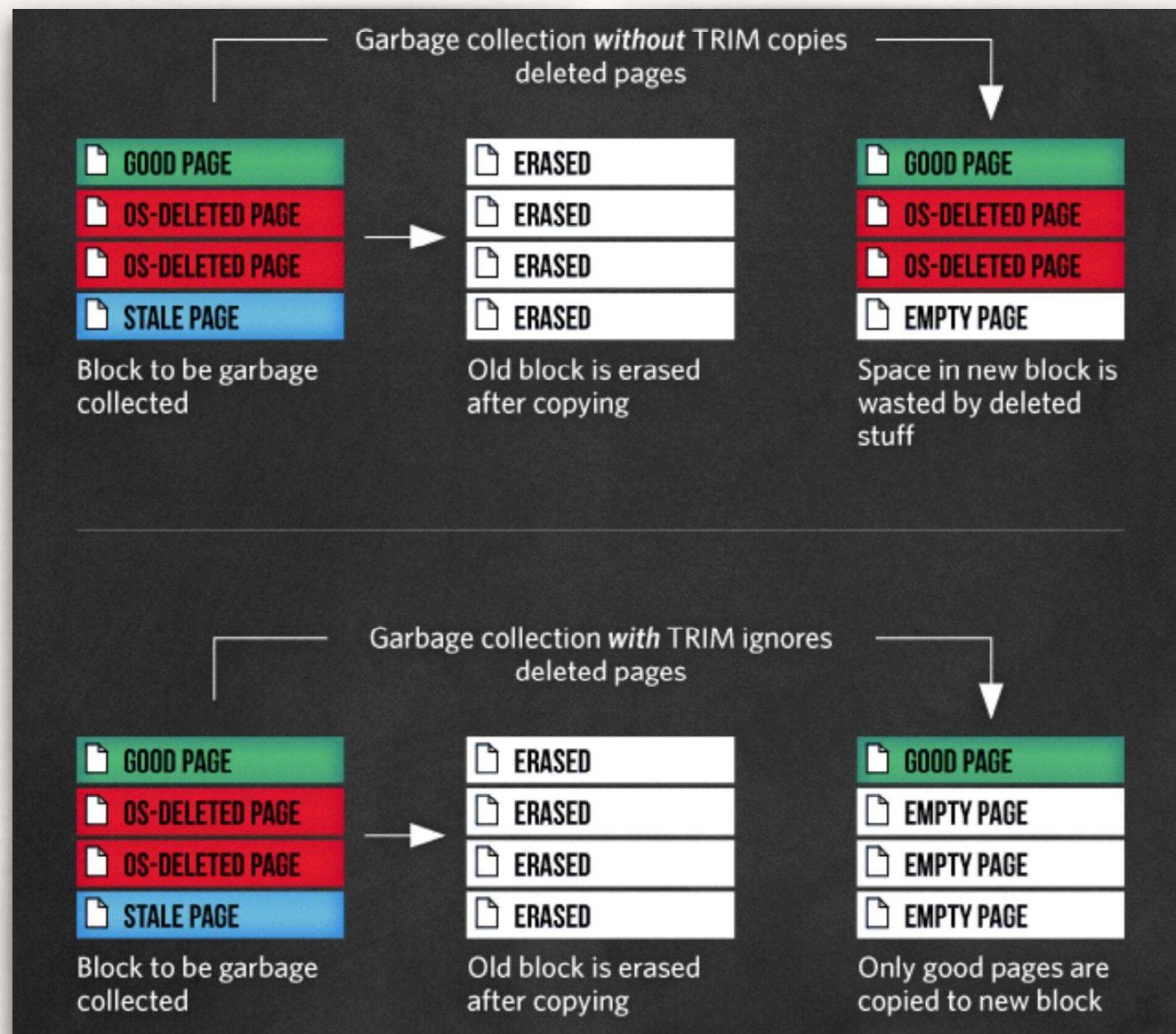
- **Wear leveling:** evitar sobrescribir repetidamente los mismos bloques.
- **Over-provisioning:** incluir bloques extra para reemplazar bloques inutilizados.
- **Garbage collection** (en realidad, recolección de páginas en uso): el controlador del SSD comprueba continuamente qué bloques tienen una mezcla de páginas en uso y en desuso (cuyo contenido, modificado, se copió a otra página vacía), copia las páginas en uso a otros bloques, y entonces borra el bloque que ya sólo contiene páginas en desuso.
- **TRIM** (comando ATA, SCSI): Ayuda a la "garbage collection". Cuando se elimina un archivo, *el sistema operativo* usa TRIM para marcar en desuso las páginas que ocupaba, por lo que ya no es necesario copiarlas. (Cuando se elimina un archivo la memoria Flash no puede saber por sí misma que las páginas que ocupaba ya no están en uso.)

MEMORIA FLASH

FUNCIONAMIENTO DEL TRIM

Sin TRIM: cuando se elimina un archivo, la memoria Flash no puede saber que las páginas que ocupaba ya no están en uso y por tanto ha de copiarlas durante la recolección.

Con TRIM: cuando se elimina un archivo, el S.O. indica a la memoria Flash que las páginas que ocupaba ya no están en uso, por lo que ya no necesita copiarlas.



(Origen: Ars Technica)

MAPEO DE MEMORIA

EN EL ESPACIO DE DIRECCIONES DE LA CPU

- Para que la CPU pueda transferir datos con los dispositivos que constituyen la memoria principal, éstos deben ubicarse en el espacio de direcciones de la CPU.
- El resultado es un “mapa” del espacio de direcciones en el que se especifica qué regiones (rangos de direcciones) están ocupadas por qué dispositivos RAM y ROM.
- Dicho mapa se implementa mediante un circuito lógico que toma como entradas líneas del bus de direcciones y del bus de control de la CPU, y cuya salida consiste en las señales de *chip select* (\overline{CS}) para los dispositivos implicados: esta función se denomina **descodificación de direcciones**.

MAPEO DE MEMORIA

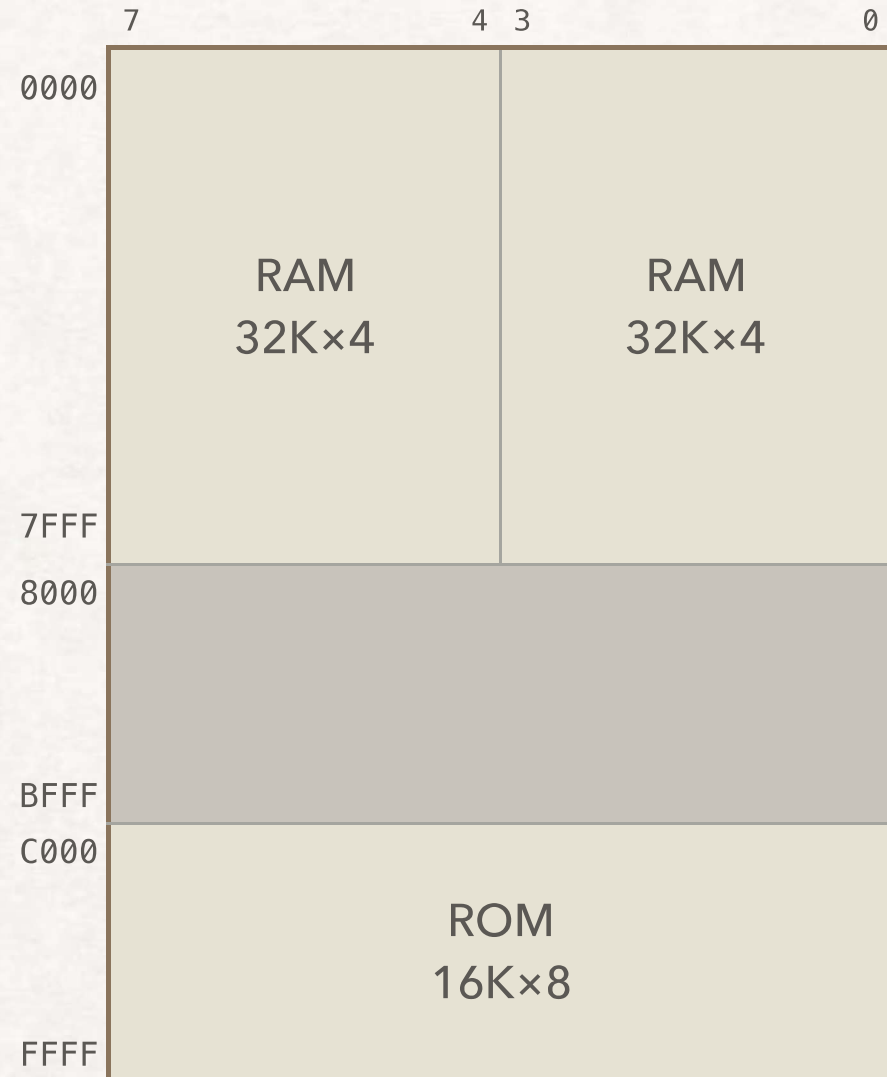
EJEMPLO 1: CON CPU Z-80 (1/2)

El Z-80 dispone de 16 líneas de direcciones y 8 de datos: espacio de 2^{16} posiciones de 8 bits (64 KB).

Queremos mapear 32 KB de RAM en la región inicial y 16 KB de ROM en la región final empleando:

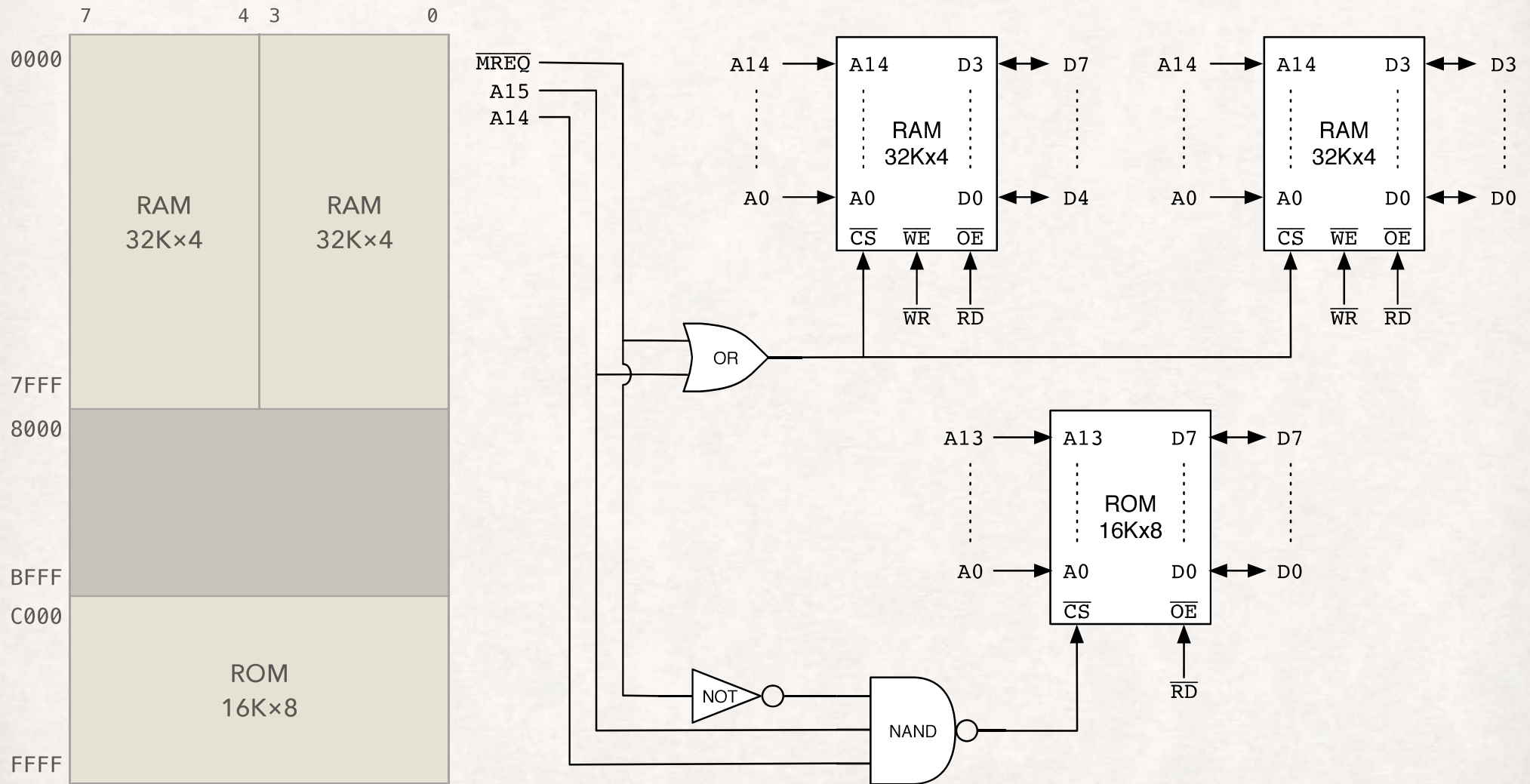
- Chips RAM 32K×4 (128 kbits).
- Chip ROM 16K×8 (128 kbits).

Mapa resultante:



MAPEO DE MEMORIA

EJEMPLO 1: CON CPU Z-80 (2/2)



MAPEO DE MEMORIA

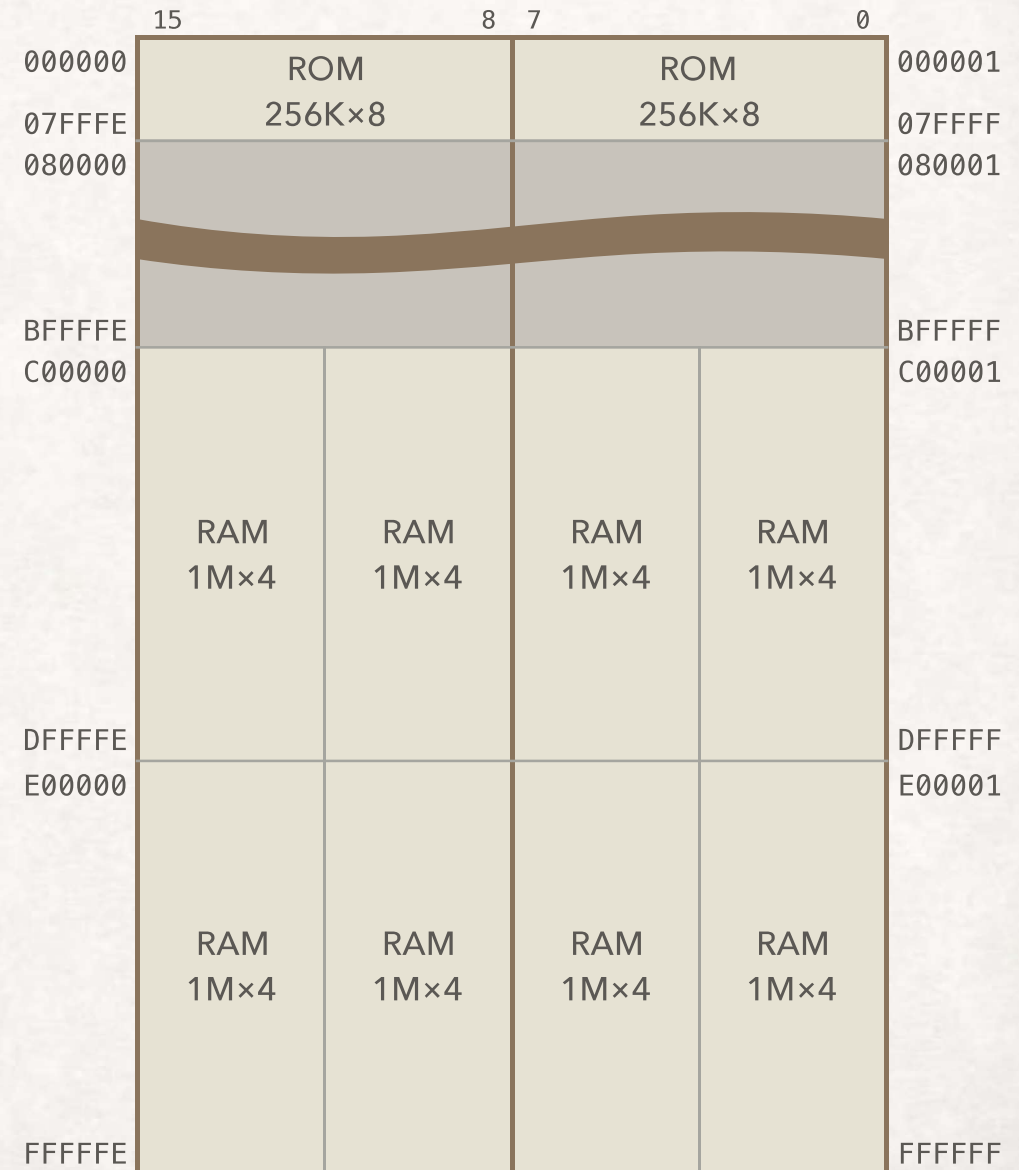
EJEMPLO 2: CON CPU MC68000 (1/3)

El 68000 dispone de 23 líneas de direcciones y 16 de datos, diferenciando entre MSB y LSB con las señales \overline{UDS} y \overline{LDS} : espacio de 2^{23} words (8 Mwords) con 2 bytes/word (16 MB).

Queremos mapear 512 KB de ROM en la región inicial y 4 MB de RAM en la región final empleando:

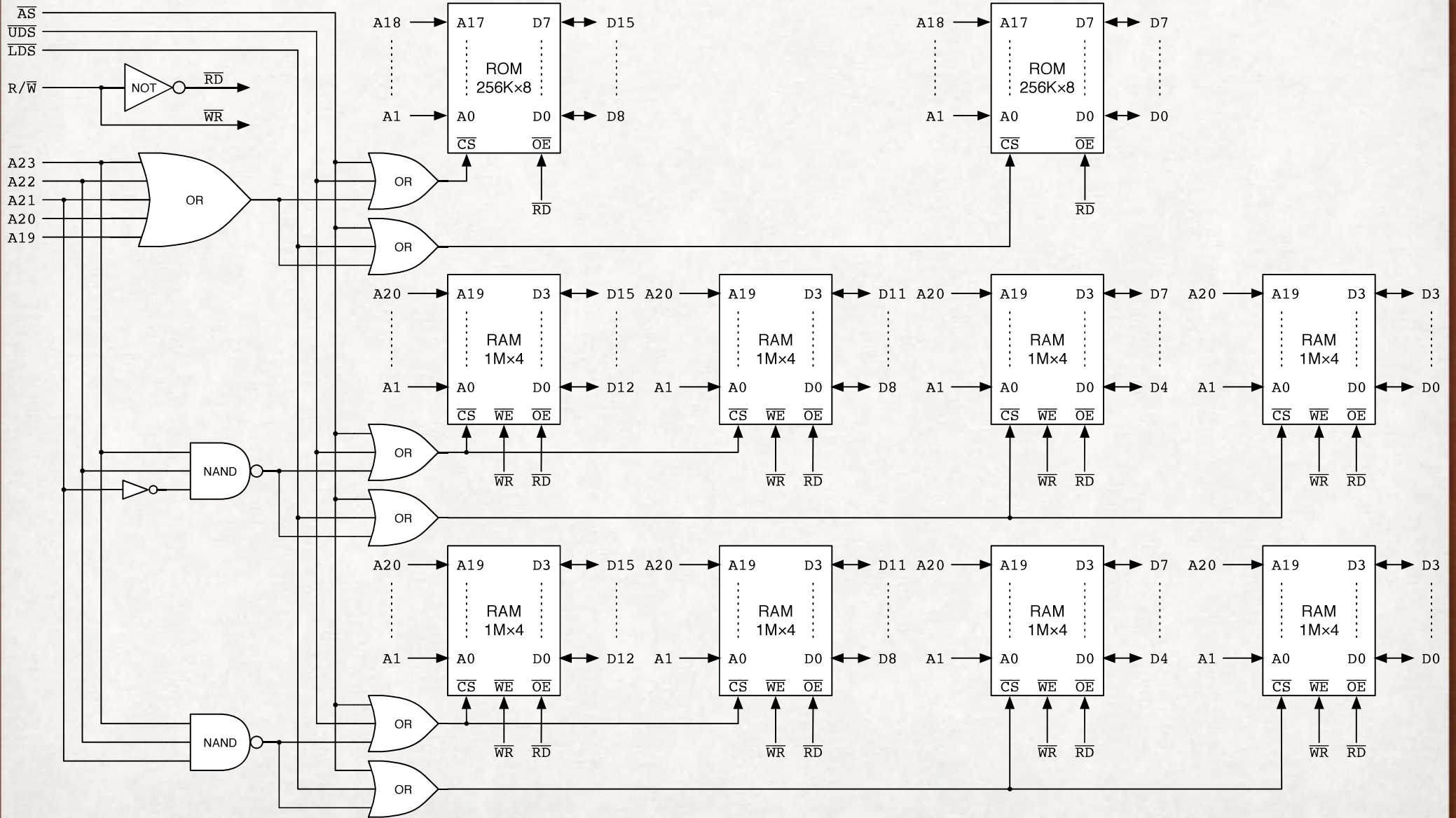
- Chips RAM 1M×4 (4 Mbits).
- Chips ROM 256K×8 (2 Mbits).

Mapa resultante:



MAPEO DE MEMORIA

EJEMPLO 2: CON CPU MC68000 (2/3)



MAPEO DE MEMORIA

EJEMPLO 2: CON CPU MC68000 (3/3)

