EE321

Computer Architecture

Chap. 05: Memory

Dr. Abdelhakim Khouas

Email : akhouas@hotmail.com

ab.khouas@univ-boumerdes.dz

IGEE (ex. INELEC)

University M'hamed Bougara of Boumerdes

Course chapters

1. Review of Digital Design
2. Top Level of Computer
3. Central Processing Unit (CPU)
4. Control Unit
5. Memory
6. Instruction Set and Addressing Modes

Lecture Objective

Understand structure and function of different types of memories

- Memory organization
- Memory types (ROM, SRAM, DRAM)
- Memory chip
- Memory map
- Memory interfacing
- Cache memory

Readings

Textbook

Computer Organization and Structure, Designing for Performance, By William Stallings, 8th edition

Sections

- Chapter 4, section: 4.1 and 4.2
- Chapter 5, sections: 5.1

Lecture Outline

1. Memory System Overview
2. Memory Organization
3. Memory Types
    1. ROM
    2. SRAM
    3. DRAM
4. Memory Chip
5. Memory map and Memory Interfacing
6. Byte Ordering (Endianness)
7. Cache Memory
8. External Memory

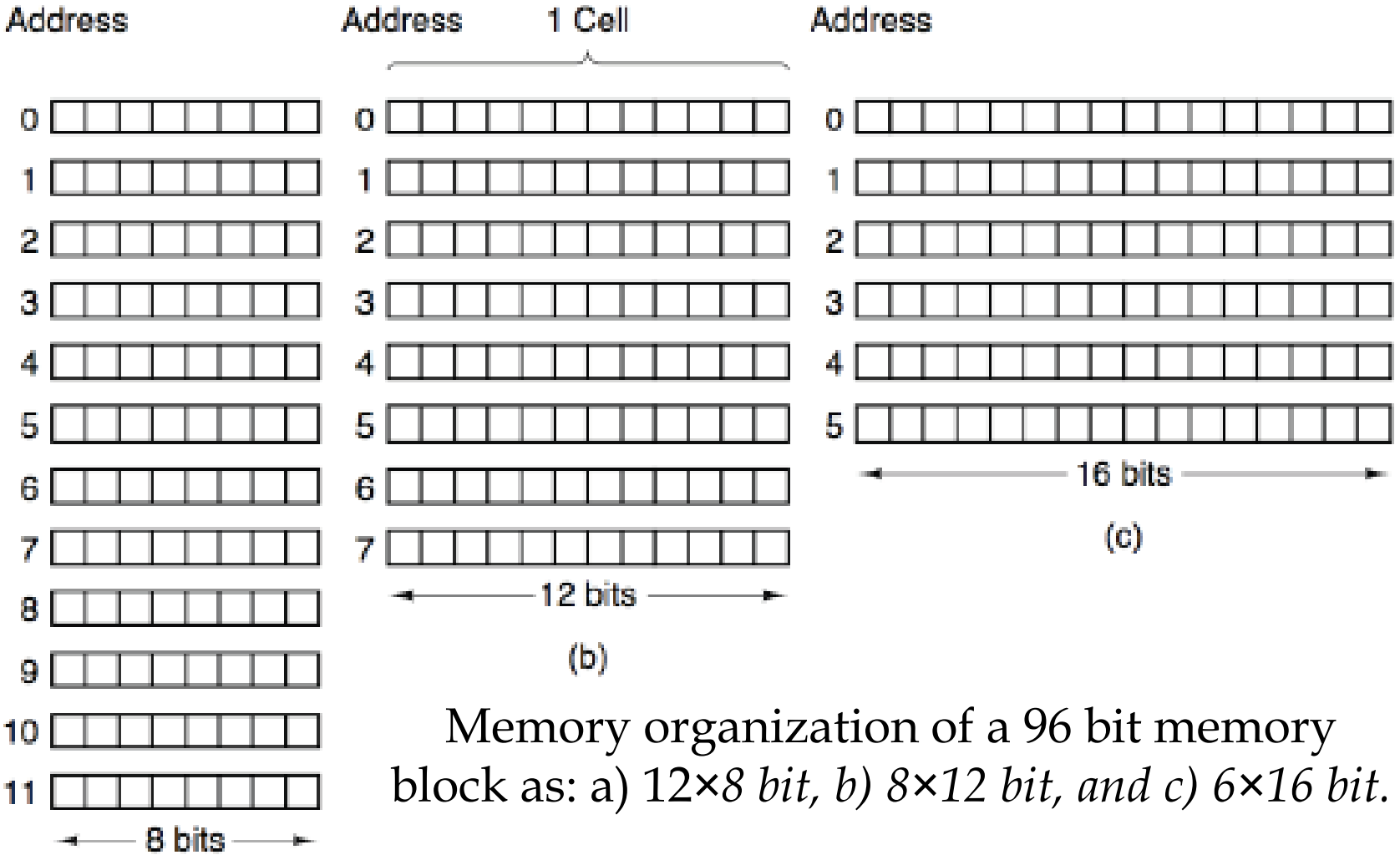
1. Memory System Overview

Memory Definition

Memory is data storage module

- Memory is one of the four main computer parts
- Memory can be M-bit serial/parallel access (no need for address)
- Memory can be organized as an array of locations, each location holds n bits and has an address
  - Depth: number of addressable locations
  - Width: number of bits per location (word size)

1. Memory System Overview



Memory organization of a 96 bit memory block as: a) 12×*8 bit, b) 8×12 bit, and c) 6×16 bit.*

1. Memory System Overview

Memory Characteristics

- Capacity: Typically expressed in terms of bytes or words. Common word lengths are 8, 16, and 32 bits
- Cost per bit
- Physical Characteristics: Volatile/non-volatile and Erasable/non-erasable

1. Memory System Overview

Memory Characteristics (cont'd)

- Access time (latency): time it takes to perform a read or write operation
- Transfer rate: This is the rate at which data can be transferred into or out of a memory unit
- Physical types: The most common types are semiconductor memory, magnetic memory used for disk and tape, and optical memory

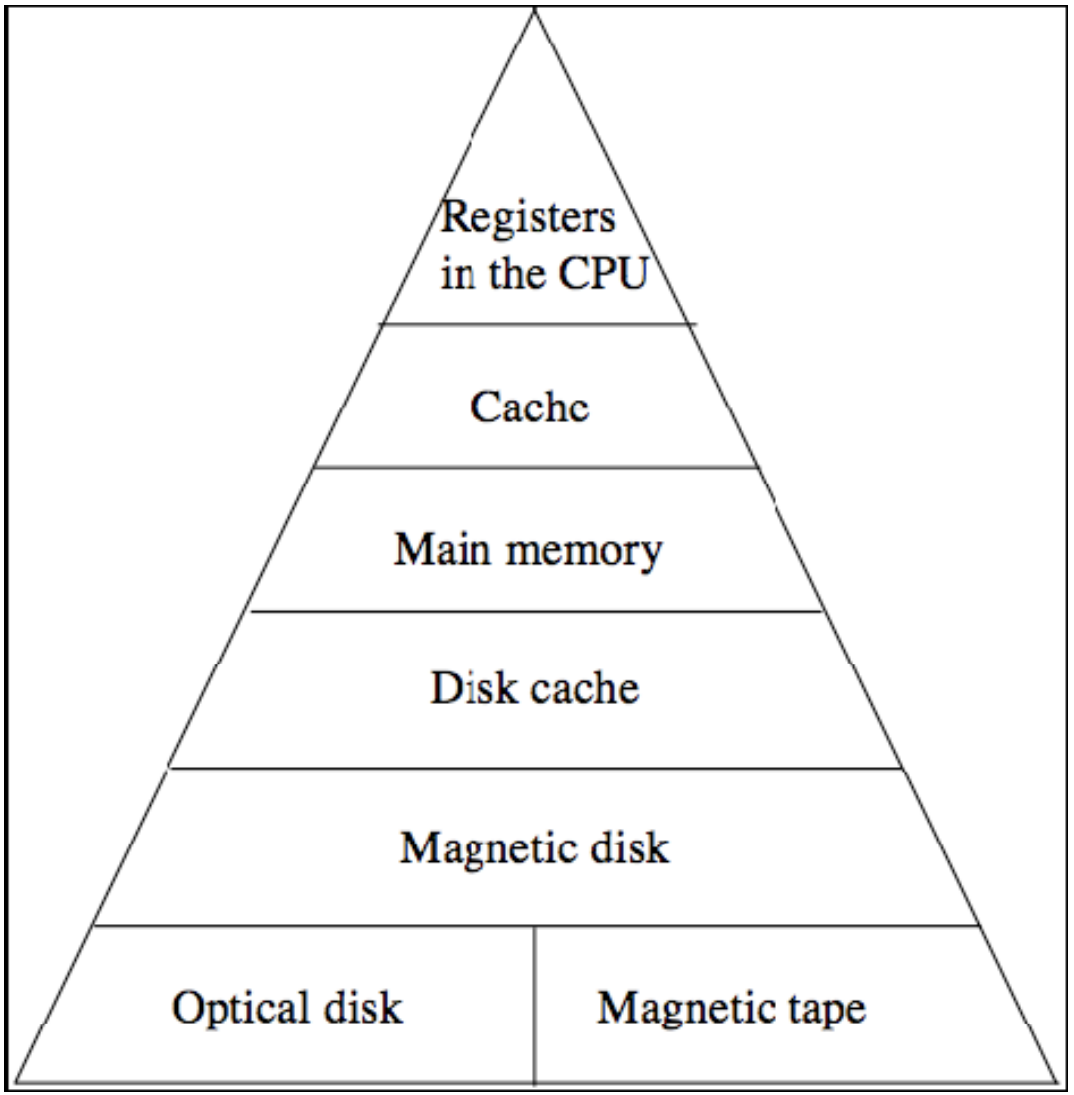
1. Memory System Overview

Memory Features (ideal memory)

- Fast access time
- High transfer rate
- Non-volatile
- Low power
- High capacity
- Low cost per bit (high density)
- In-system re-writable
- Fully-bit erasable

1. Memory System Overview

Memory Hierarchy



1. Memory System Overview

As we go down the memory hierarchy, following occur:

- Decreasing cost per bit: registers are the most expensive, optical disks/magnetic tapes are the cheapest
- Increasing capacity
- Increasing access time: registers are the fastest, optical disks/magnetic tapes are the slowest
- Decreasing frequency of access of the memory by the processor: references to registers should be more frequent than that for main memory or hard disk

1. Memory System Overview

It would be nice to use only the fastest memory, but because that is the most expensive memory, we trade off access time for cost by using more of the slower memory

- The design challenge is to organize the data and programs in memory so that the accessed memory words are usually in the faster memory

2. Memory Organization and Function

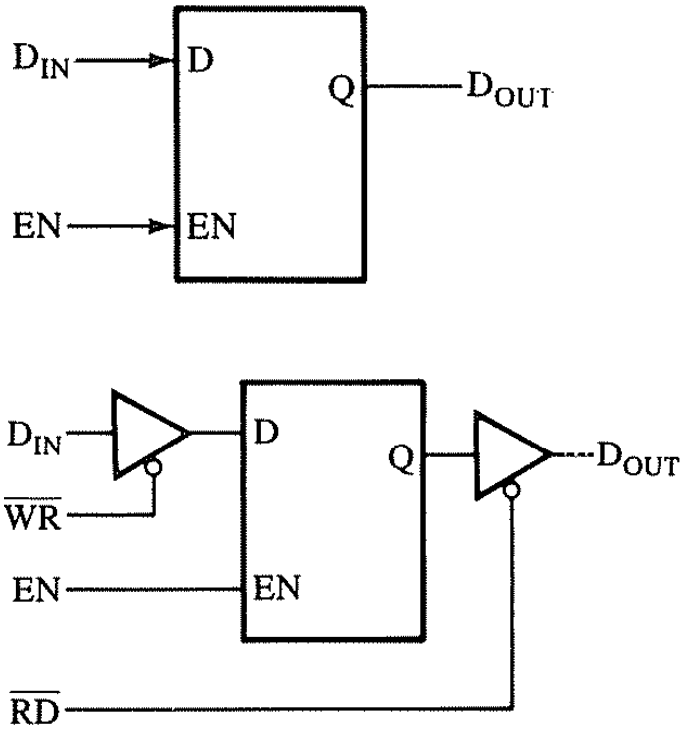
Memory module can be:

- Latch, Flip-Flop, Register
- Memory chip (IC): ROM, EPROM, EEPROM, FLASH, SRAM, DRAM
- Optical disc
- Magnetic disk
- Magnetic tape

2. Memory Organization and Function

Flip-Flop or Latch memory cell

- Latch or Flip-Flop is 1-bit memory celll
- In a simple memory cell, the stored bit is always available on the output
- When the memory cell is connected to a data bus, we use tristate buffer at the input and output of the cell

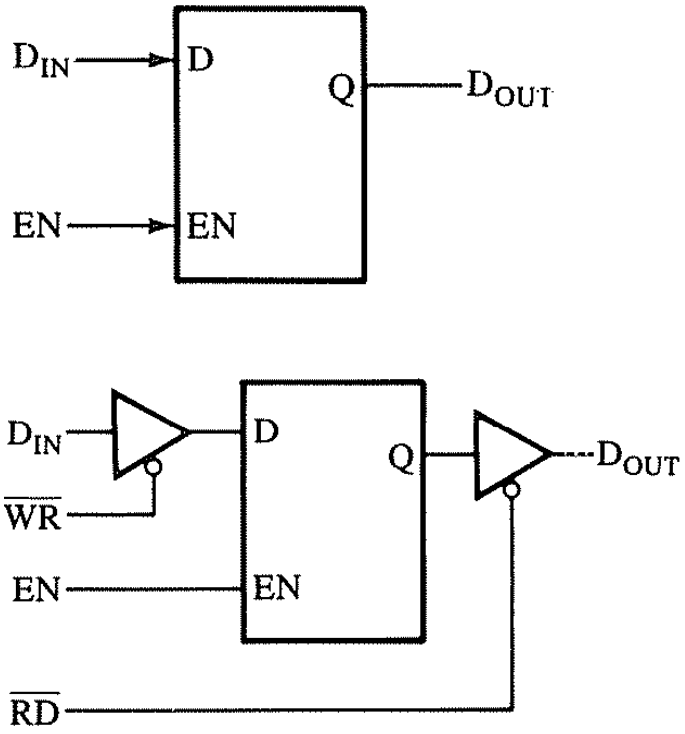


Source: The Z80 Microprocessor, by R. Gaonkar

2. Memory Organization and Function

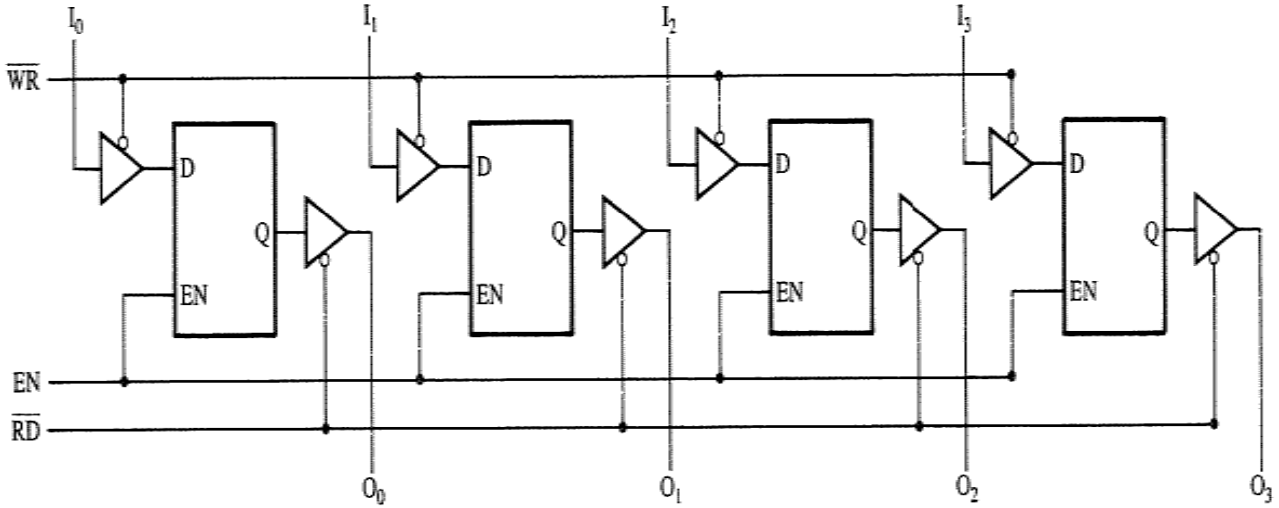
Flip-Flop or Latch memory cell

- The active low write signal (/WR) is used to enable the input tristate buffer to write into the memory cell (read from data bus)
- The active low read signal (/RD) is used to enable the output tristate buffer to read from the memory cell (write to the data bus)

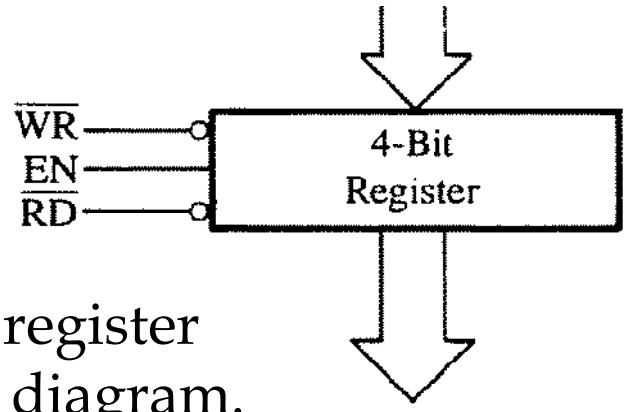


Source: The Z80 Microprocessor, by R. Gaonkar

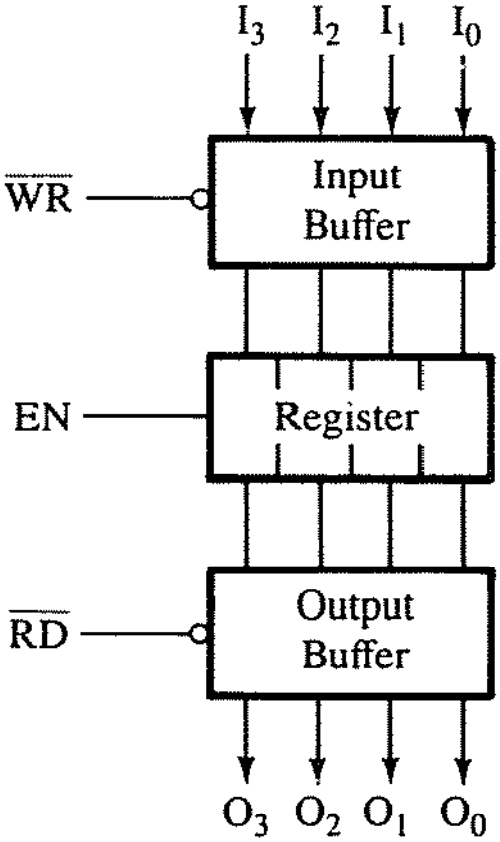
2. Memory Organization and Function



4 memory cells used as 4-bit register.



4-bit register block diagram.



4-bit register block diagram.

Source: The Z80 Microprocessor, by R. Gaonkar

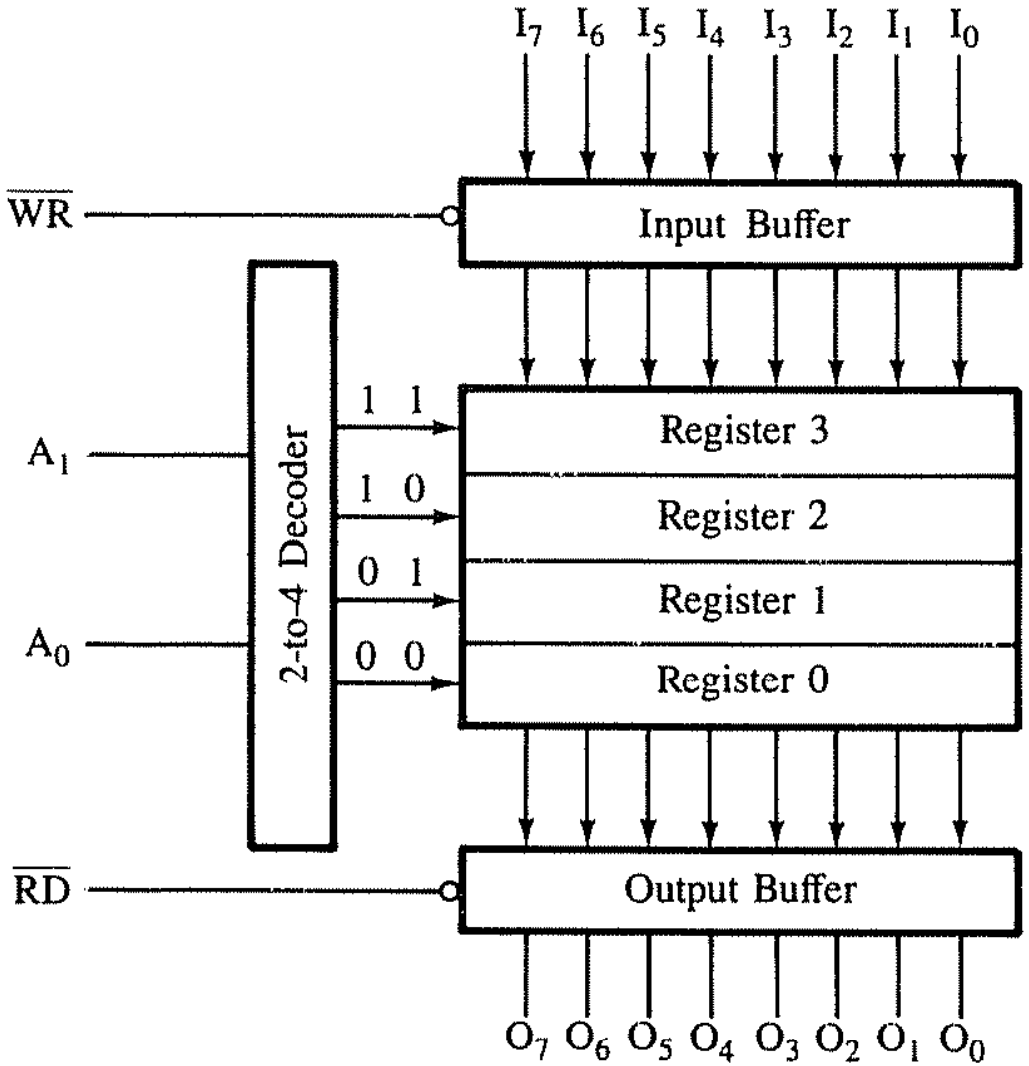
2. Memory Organization and Function

Set of registers

- To read form or write into any one of the register, a specific register should be enabled
- The registers are identified by 0,1,2, ...etc.
- We use a simple decoder to enable the registers
- The decoder requires new input lines called address lines
- N address lines can carry 2^N registers

2. Memory Organization and Function

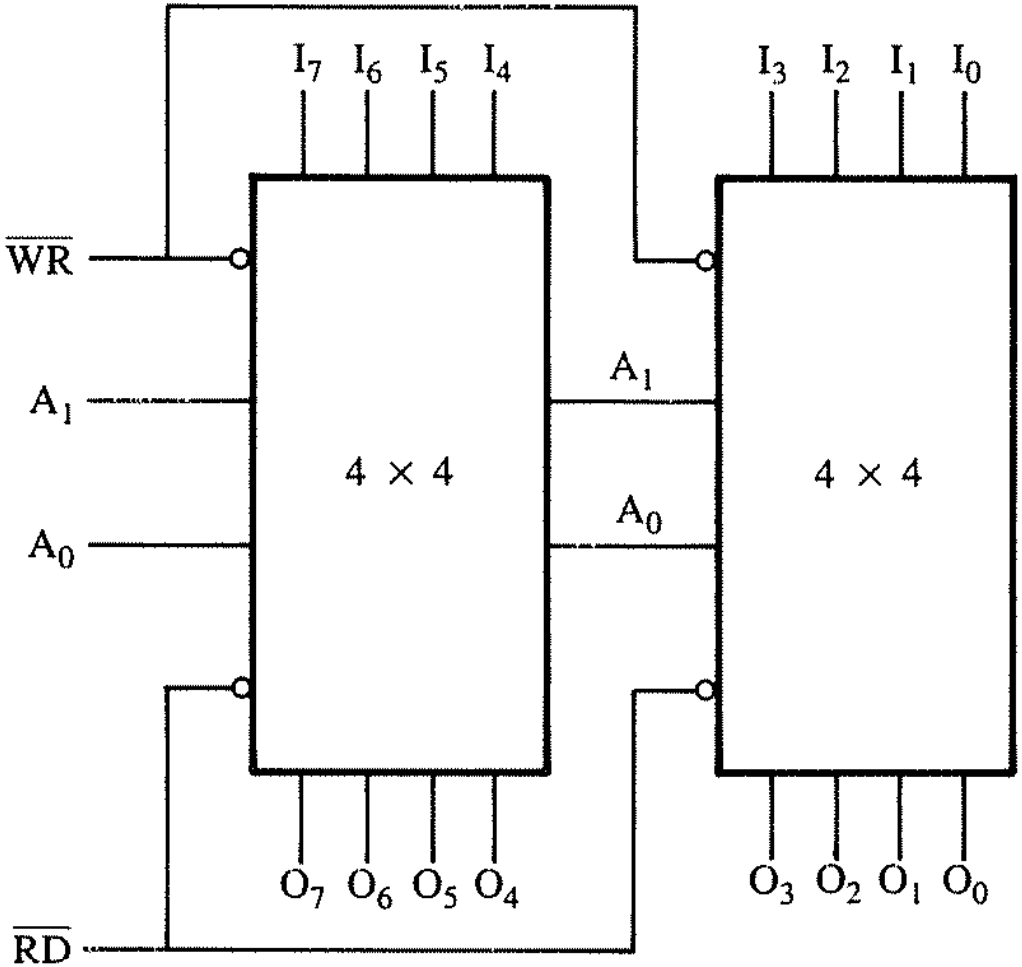
Example: 4x8-bit memory using one chip of 4x8-bit memory



Source: The Z80 Microprocessor, by R. Gaonkar

2. Memory Organization and Function

Example: 4x8-bit memory using two chips of 4x4-bit memory



Source: The Z80 Microprocessor, by R. Gaonkar

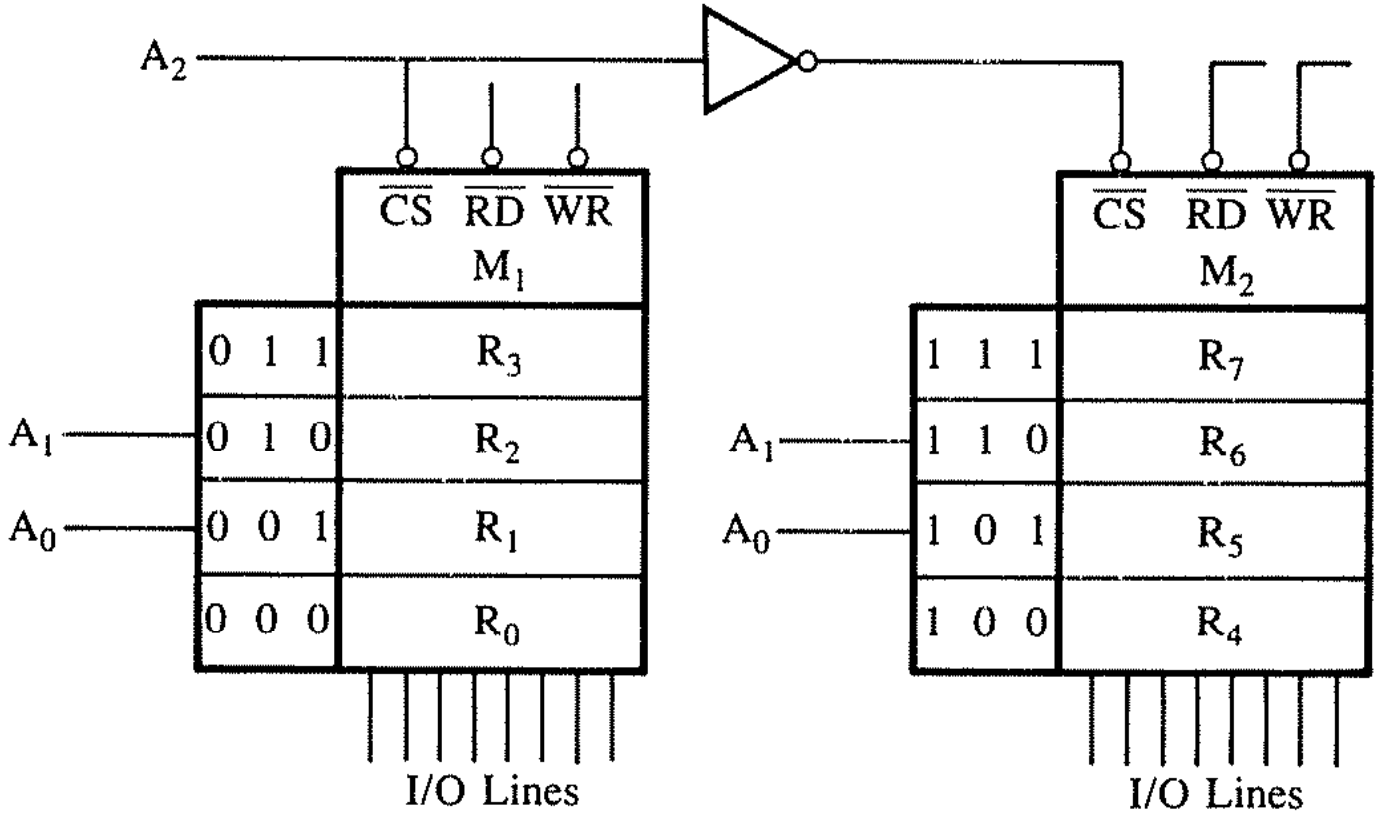
2. Memory Organization and Function

How to design 8x8-bit memory using two 4x8-bit memory chips?

- 8x8-bit memory needs 3 address lines A2,A1,A0
- 4x8-bit memory has 2 address lines A0 and A1
- A0 and A1 are connected to both memory chips
- RD and WR signals are connected to both chips
- A2 is used to select one memory chip each time
- The memory requires a new input signal called Chip Select (CS) or Chip Enable (CE)

2. Memory Organization and Function

Example: 8x8-bit memory using two 4x8-bit chip



Source: The Z80 Microprocessor, by R. Gaonkar

2. Memory Organization and Function

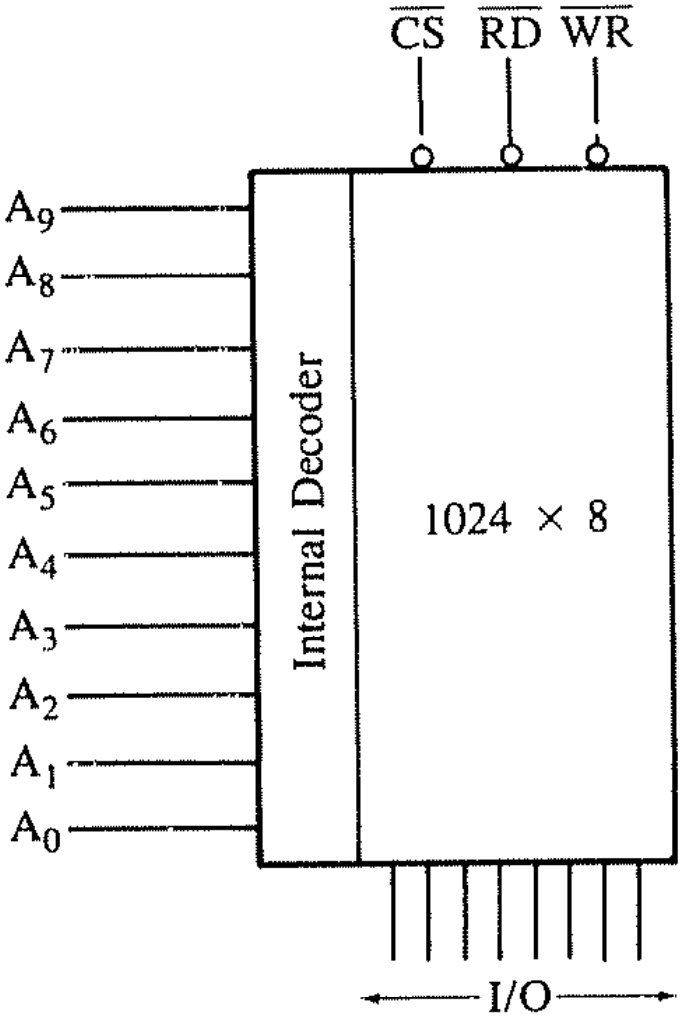
Memory chip requirements

- Address lines to identify a memory location
  - High order address is used to select memory chip
  - Low order address is used to select location in the memory
- Data lines
- Enable signal (clock)
- Chip Select (CS or CE) signal to selects specific chip
- Control signal Read RD to enables the output buffer
- Control signal Write WR to enable the input buffer for R/W memory only

2. Memory Organization and Function

Example: Memory chip

- ROM or RAM?
- Number of bits per location?
- Size of address bus?
  - Number of locations?
- Size of the memory?
- Function of internal decoder?
- Read signal and write signals?
- Chip select signal?

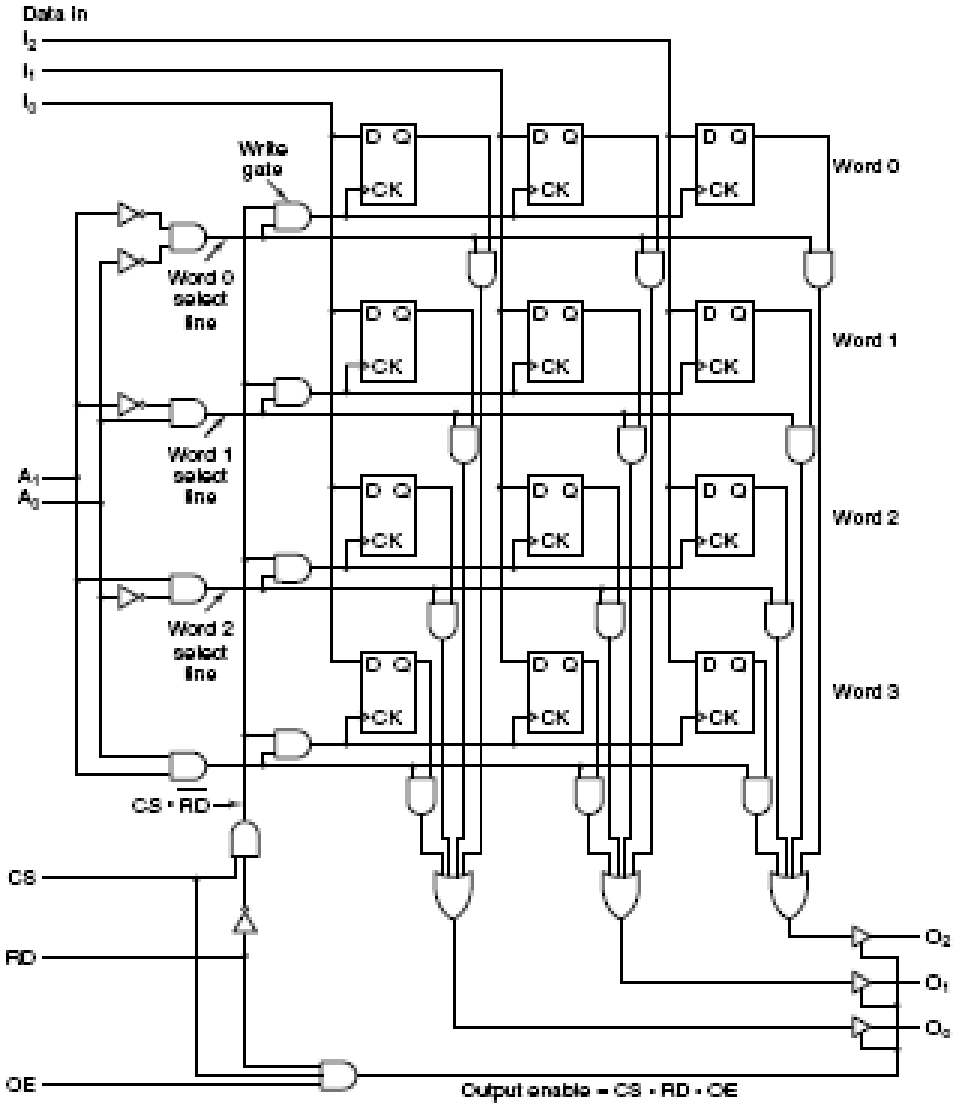


Source: The Z80 Microprocessor, by R. Gaonkar

2. Memory Organization and Function

Example: Logic diagram for 4x3bit

- I_2, I_1, I_0: Data in
- O_2, O_1, O_0: Data out
- A_1, A_0: Address bus
- CS: Chip Select
- RD: Read/Write signal
  - RD=1 for read
  - RD=0 for write
- OE: Output Enable Signal
- CLK: Clock signal



Source: Computer Organization and Architecture, by W. Stallings

2. Memory Organization and Function

Memory Read Cycle

1. CPU places the address of the memory location where a byte is to be stored. The interfacing logic of the memory chip decodes the address and selects the memory location to be written into
2. CPU Send the read control signal to enable the output buffer of the memory
3. The memory chip places the data on the data bus
4. CPU reads the data

2. Memory Organization and Function

Memory Write Cycle

1. CPU places the address of the memory location where a byte is to be stored. The interfacing logic of the memory chip decodes the address and selects the memory location to be written into
2. CPU places the data on the data bus, and sends the write control signal to enable the input buffer of the memory
3. Memory chip reads and stores the data

2. Memory Organization and Function

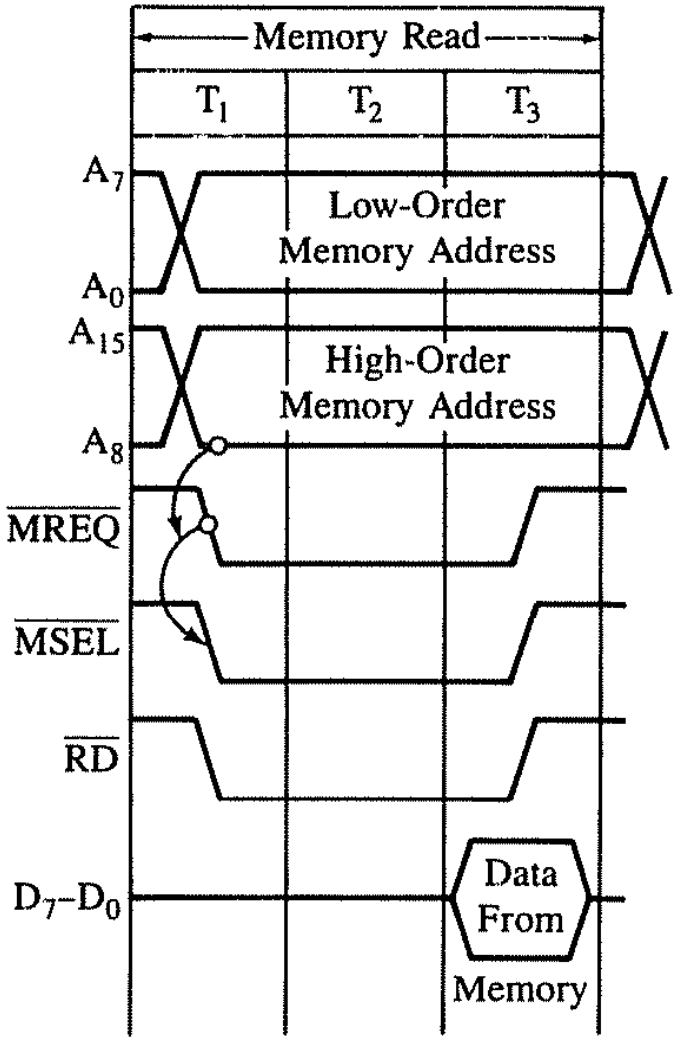
Z80 Memory Read Cycle

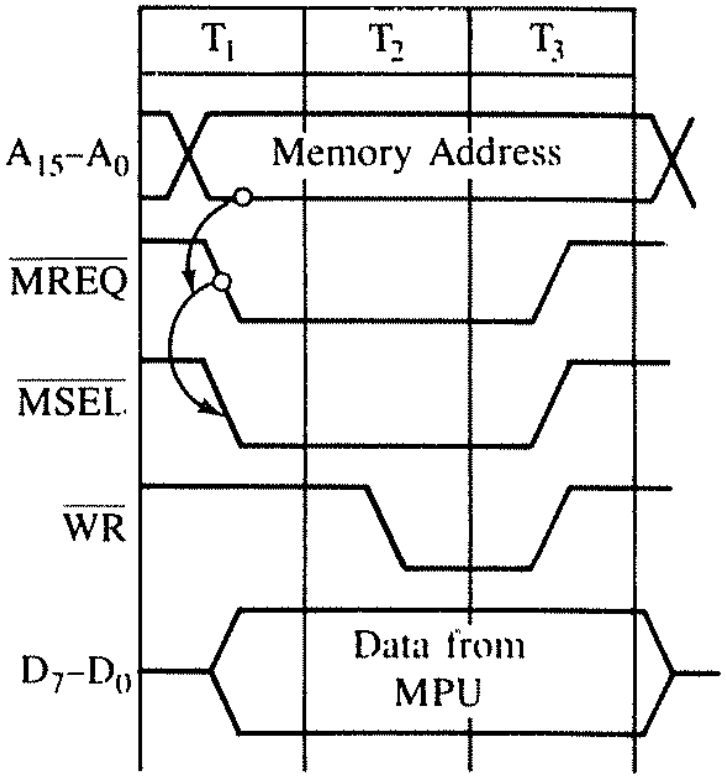
1. T1: Z80 places a 16-bit address on its address bus and asserts /MREQ and /RD
2. T2: Memory chip places the data on the data bus
3. T3: Z80 reads the data from data bus

Z80 Memory Read Cycle

1. T1: Z80 places a 16-bit address and 8-bit data on its buses and asserts /MREQ
2. T2: Z80 asserts /WR signal
3. T3: Memory chip reads the data from data bus

2. Memory Organization and Function





Z80 memory read and write timing diagram

Source: The Z80 Microprocessor, by R. Gaonkar

3. Memory Types

1) Read Only Memory (ROM)

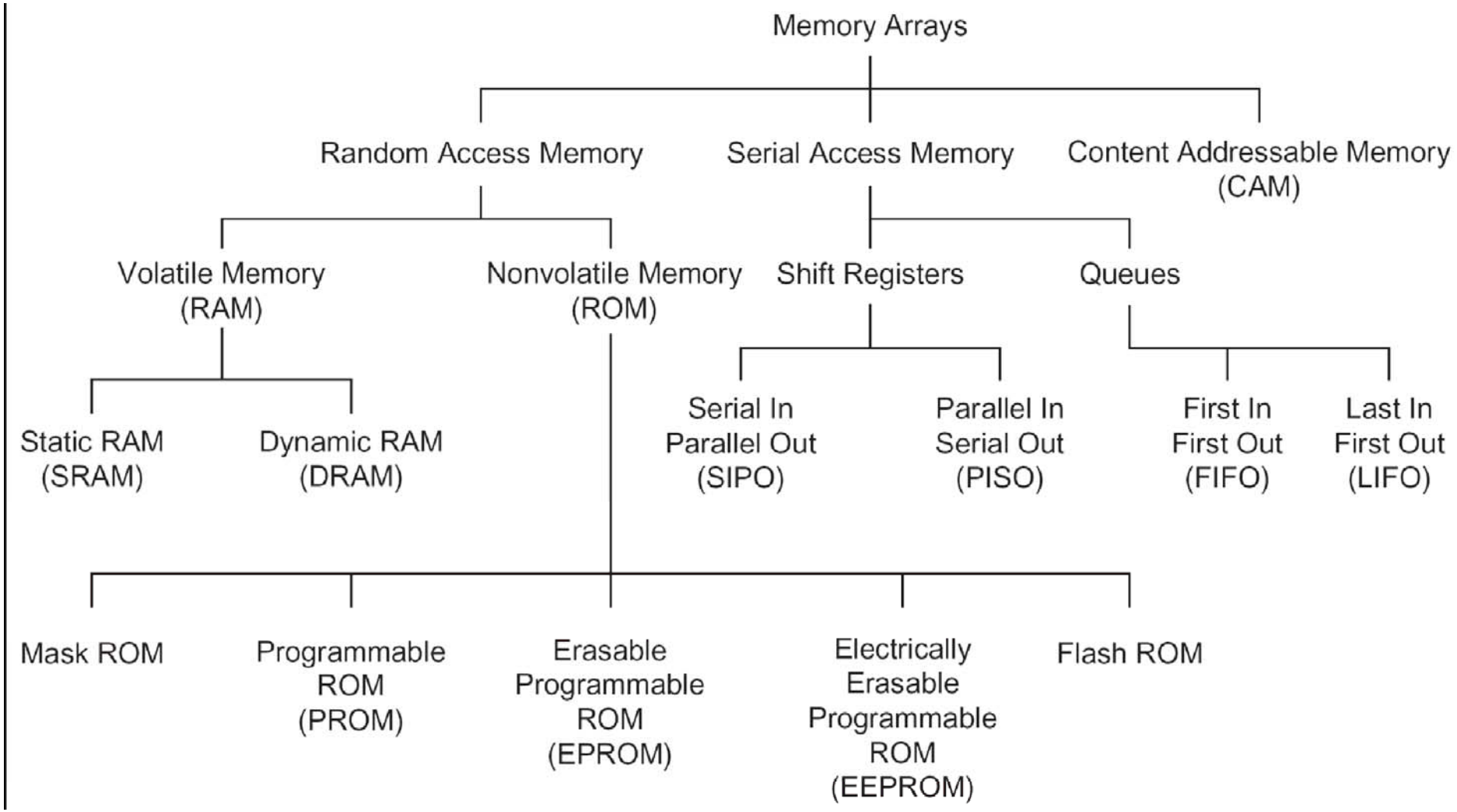
Contains a permanent pattern of data that cannot be changed easily and rapidly

- It is non-volatile

2) Random Access Memory (RAM)

- It is possible both to read data from the memory and to write new data into the memory easily and rapidly
  - It is volatile

3. Memory Types



Source: Memory Design, by K. R. Veveka

3. Memory Types

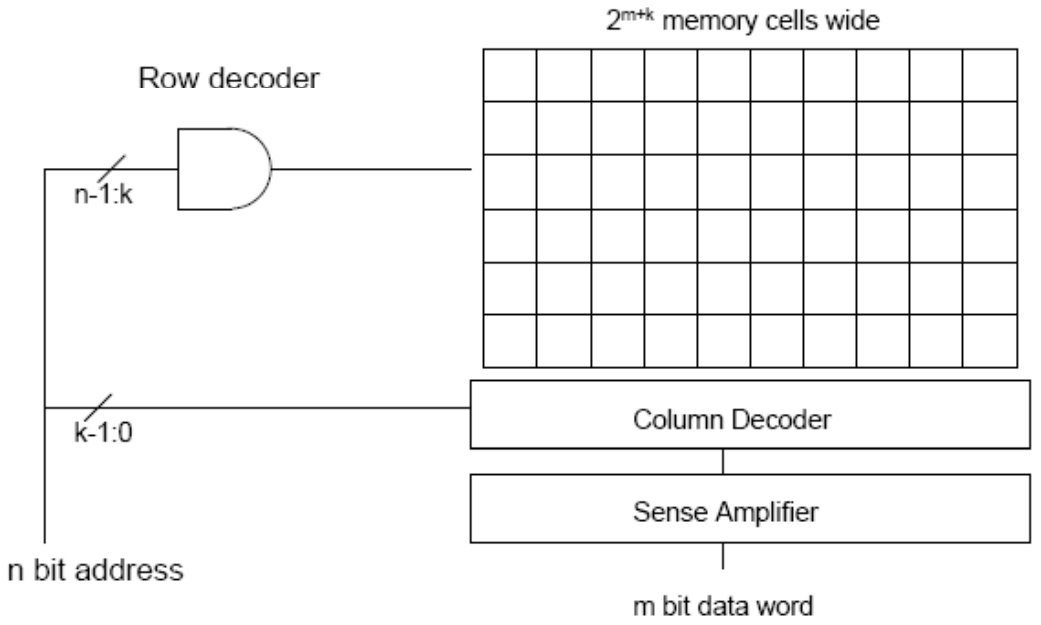
1-bit Memory Cell

- The basic element of a semiconductor memory is the memory cell
- Each memory cell exhibits two stable states (0 and 1)
- The main operations are:
  - Set state 0 (write 0 value)
  - Set state 1 (write 1 value)
  - Read state (read content of memory cell)

3. Memory Types

Memory structure

- Address lines
- Row decoder
- Column decoder
- Memory array
  - Word lines (WL)
  - Bit lines (BL)
- Sense amplifier
- Data lines



Source: Memory Design, by J. A. Chandy

3. Memory Types

Example: 8x2bit memory structure

3.1 ROM

Read Only Memory (ROM)

- Permanent storage
  - Nonvolatile
- Used in Microprogrammed Control Unit and systems programs (BIOS)

Types of ROM

- Maskable ROM,
- One-time programmable ROM (fuse ROM)
- EPROM: Erasable Programmable ROM
- EEPROM: Electrically Erasable Programmable ROM
  - Flash memory

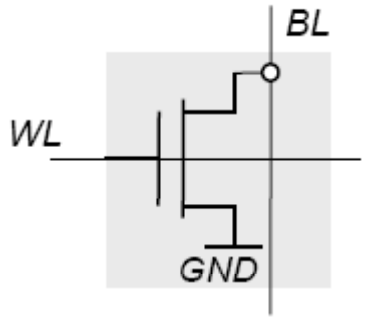
3.1 ROM

Maskable ROM

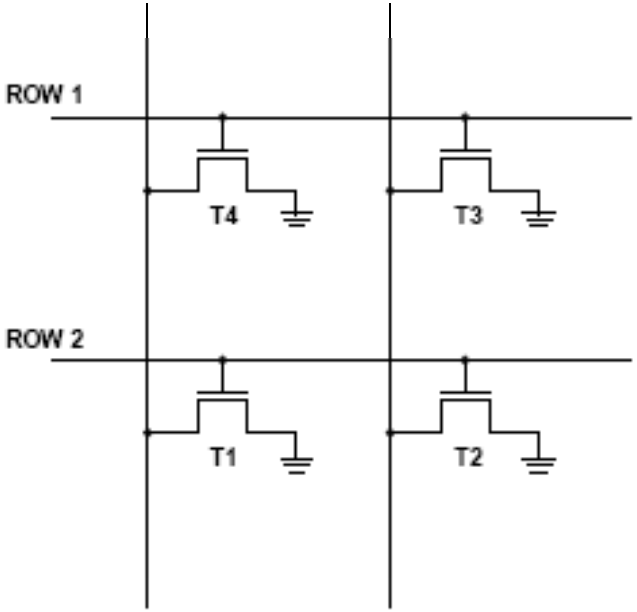
- Data is 'wired in' during fabrication at a chip manufacturer's plant
  - Very expensive for small runs
  - It is manufactured for high production runs (more than 10k)
  - Can you buy a ROM? Why?

3.1 ROM

Example: 4x3bit Maskable ROM using NMOS NOR



- Depletion NMOS transistor as pull-up resistance
- NMOS transistors in parallel (as NOR gate)
- If the NMOS transistor is connected to WL, the stored value is 0, otherwise it is 1



Source: Memory Design, by J. A. Chandy

3.1 ROM

One-time Programmable (OTP) ROM

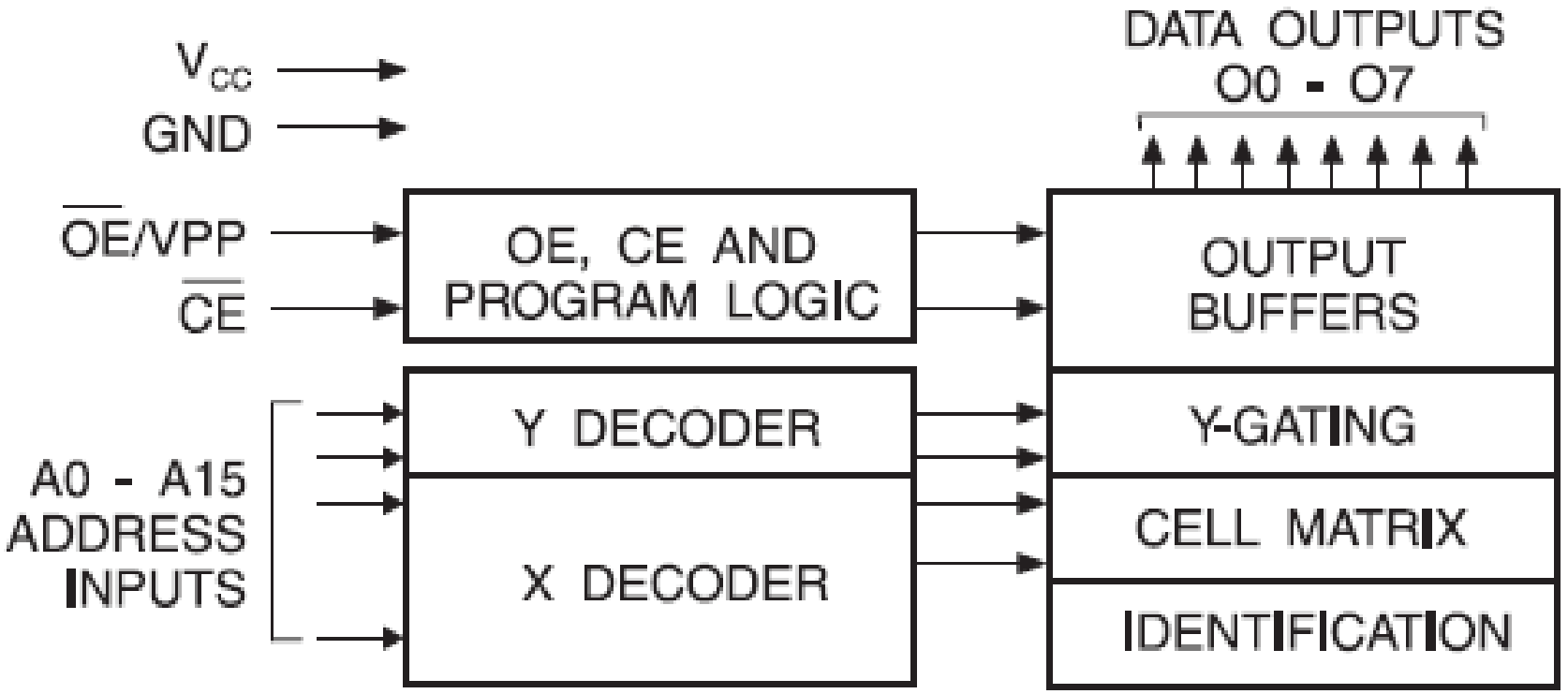
- In OTP ROM, data can be written once by the user
- We use anti-fuse for programmable connection
- Useful for small production runs

Example: 64Kx8bit Atmel OTP EPROM

- Fast read access time: 45 ns/Byte
- Rapid programming time: 100μs/Byte (typical)

3.1 ROM

Example: 64Kx8bit Atmel OTP EPROM

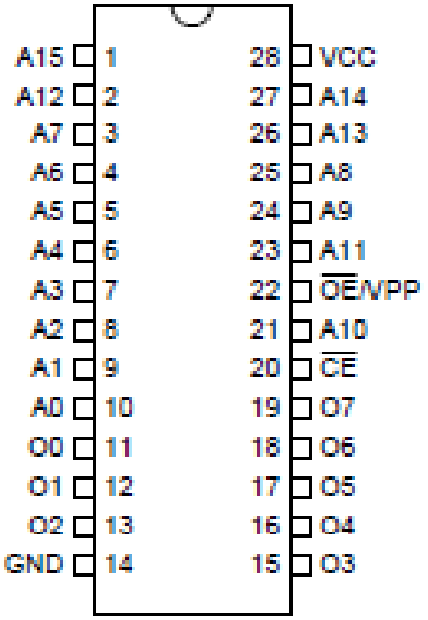


Block diagram

Example: 64Kx8bit Atmel OTP EPROM

Pin Name	Function
A0 to A15	Addresses
O0 - O7	Outputs
$\overline{CE}$	Chip Enable
$\overline{OE}$/VPP	Output Enable/VPP
NC	No Connect

PDIP, SOIC Top View

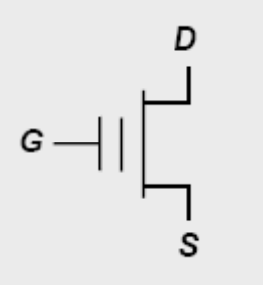


Pin configuration

3.1 ROM

Erasable Programmable ROM (EPROM)

- In EPROM, programming is similar to a PROM, but the ROM can be erased by exposing to Ultra Violet (UV) light
  - FGMOS (FAMOS) transistor
  - One transistor per memory cell
  - Entire memory is erased
  - Off-system programming and de-programming

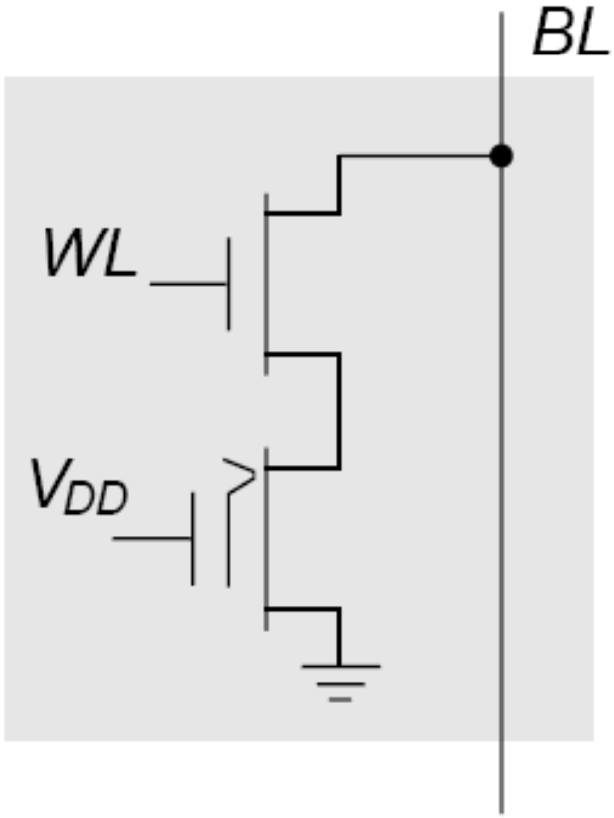


Floating-gate MOS (FGMOS) transistor

3.1 ROM

Electrically Erasable PROM (EEPROM)

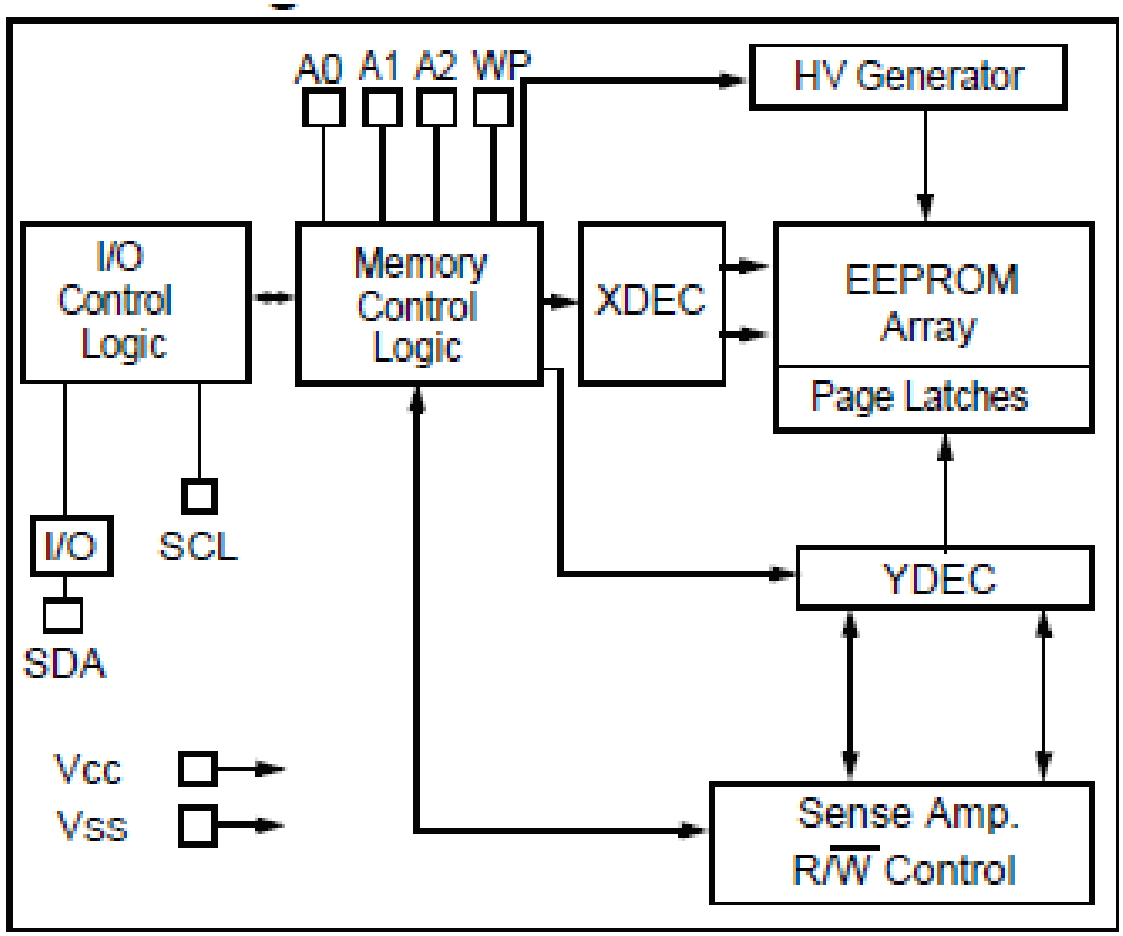
- EEPROM can be written to many times while remaining in a system
  - FLOTOX transistor
  - Writes require several hundred microseconds per byte
  - Two transistors for each cell



1-bit cell EEPROM

3.1 ROM

Example: 4kx8bit I2C Serial EEPROM



Block diagram

Example: 4kx8bit I2C Serial EEPROM

Name	PDIP	SOIC	SOIJ	TSSOP	Rotated TSSOP	DFN[1]	TDFN[1]	MSOP	SOT-23	CS	Description
A0	1	1	1	1	3	1	1	1	—	—	Chip Address Input
A1	2	2	2	2	4	2	2	2	—	—	Chip Address Input
A2	3	3	3	3	5	3	3	3	—	—	Chip Address Input
Vss	4	4	4	4	6	4	4	4	2	2	Ground
SDA	5	5	5	5	7	5	5	5	3	5	Serial Address/Data I/O
SCL	6	6	6	6	8	6	6	6	1	4	Serial Clock
WP	7	7	7	7	1	7	7	7	5	3	Write-Protect Input
Vcc	8	8	8	8	2	8	8	8	4	1	+1.7V to 5.5V Power Supply

Pin function table

3.1 ROM

Flash EEPROM

- One transistor per memory cell
- Similar to EEPROM in using electrical erase, but it supports block erasures
- Fast erasures

3.2 DRAM

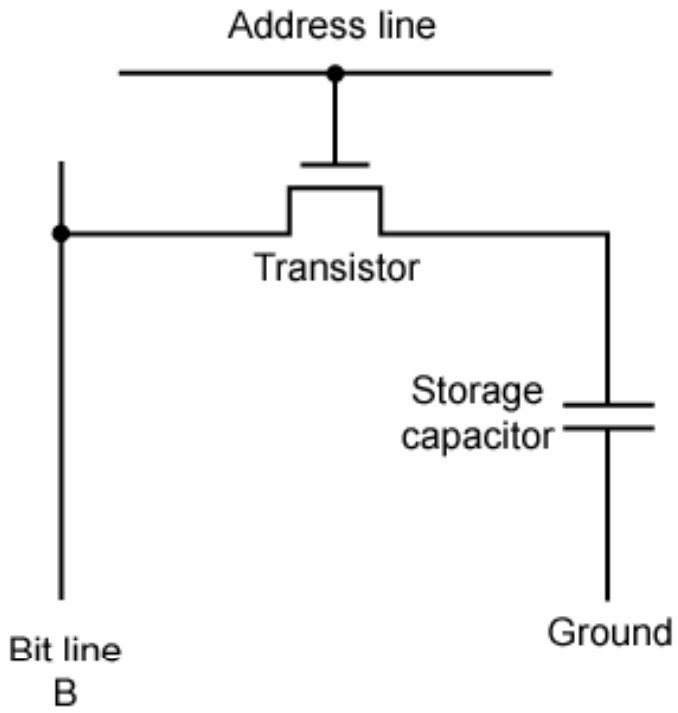
Random Access Memory (RAM)

- Misnamed as all semiconductor memory chips are random access
- Read/Write
- Volatile
- Temporary storage
- Static RAM (SRAM) and dynamic RAM (DRAM)

3.2 DRAM

DRAM

- Bits stored as charge in capacitors
- Charges leak
  - Need refresh circuits
- Simpler construction
- Less expensive (1tr./bit)
- Slower
- Used for main memory
- Essentially analogue
  - Level of charge determines value



3.2 DRAM

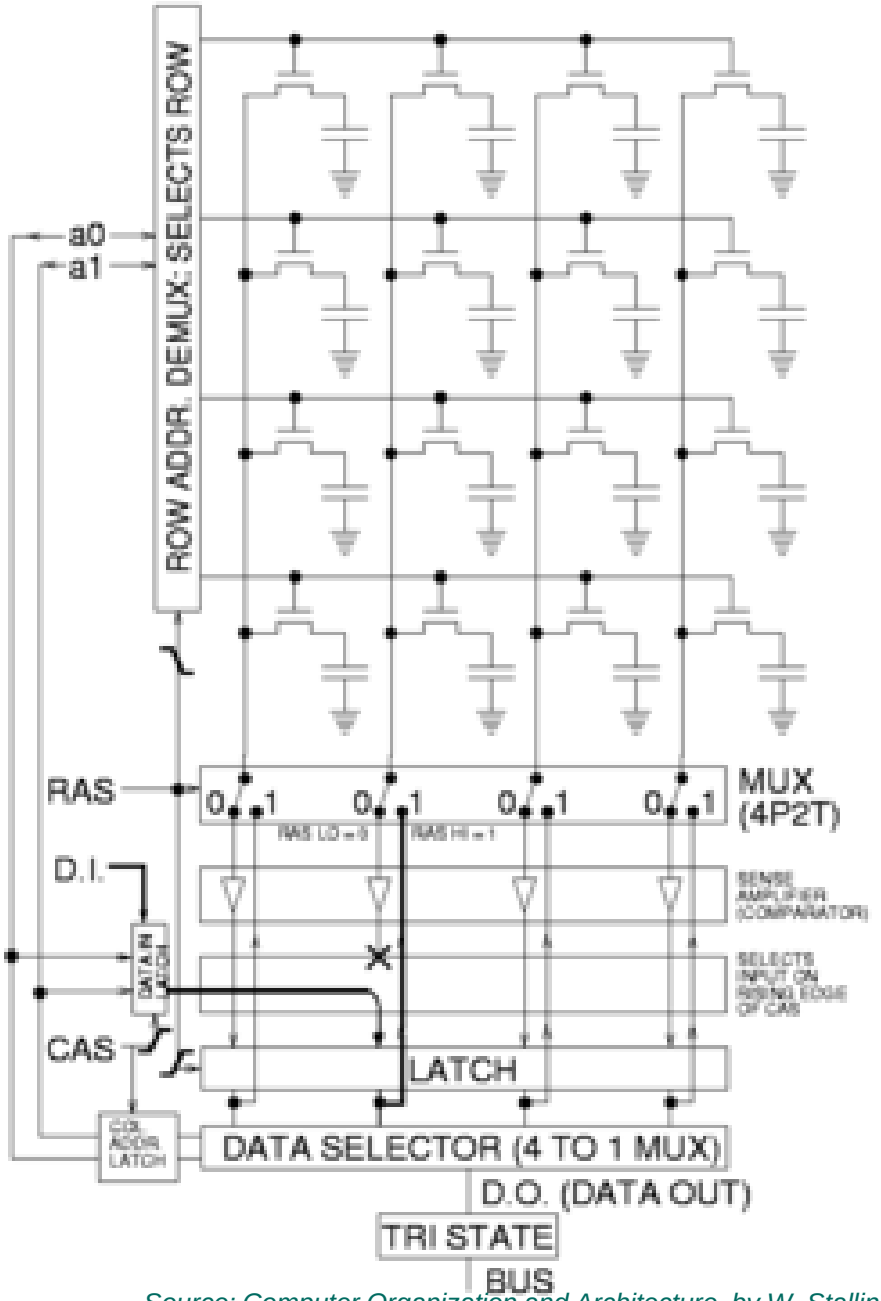
DRAM Operation

- Address line active when bit read or written
  - Transistor switch closed (current flows)
- Write
  - Voltage to bit line High for 1 low for 0
  - Then signal address line transfers charge to capacitor
- Read
  - Address line selected transistor turns on
  - Charge from capacitor fed via bit line to sense amplifier
  - Compares with reference value to determine 0 or 1
  - Capacitor charge must be restored

3.2 DRAM

RAS: Row Address Selector (used for row decoder, MUX and Latch)
CAS: Column Address Selector
DI: Data In
DO: Data Out
Write process ?
Read process ?

16x1-bit DRAM arranged as 4x4 square with row and column decoders.

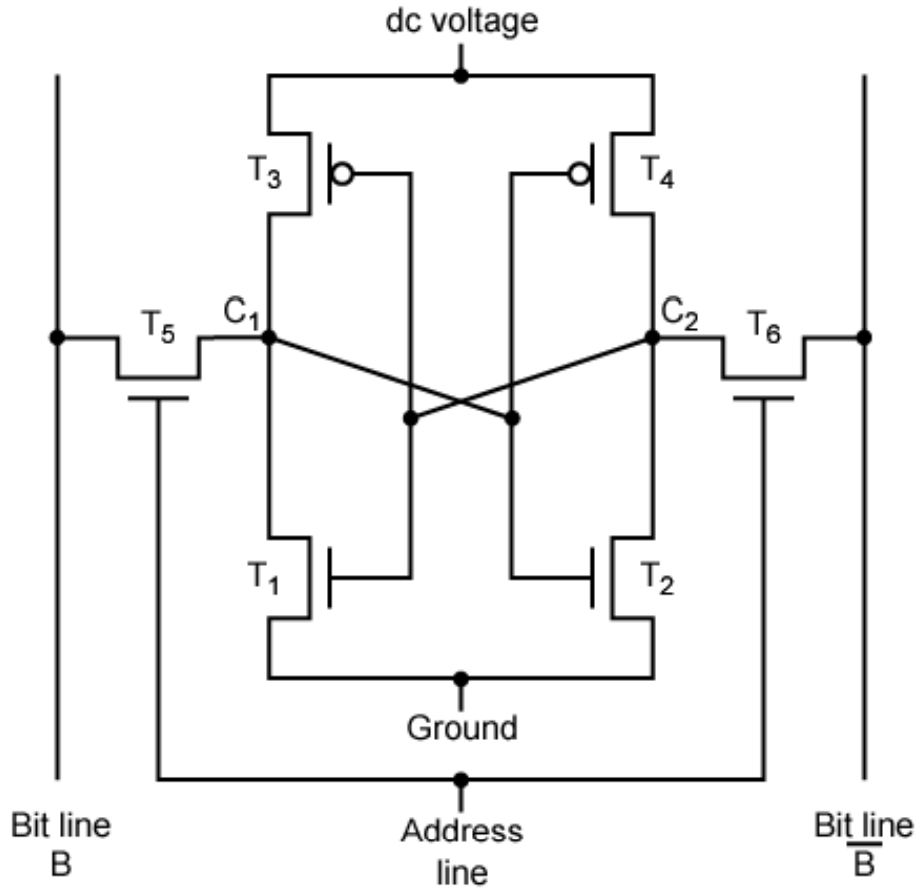


Source: Computer Organization and Architecture, by W. Stallings

3.3 SRAM

SRAM

- Bits stored as on/off switches
- No charges to leak
  - Does not need refresh circuits
- More complex construction
- More expensive (6tr./bit)
- Faster
- Used for cache memory

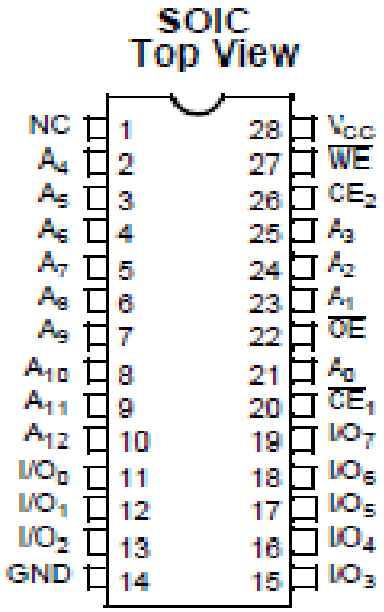


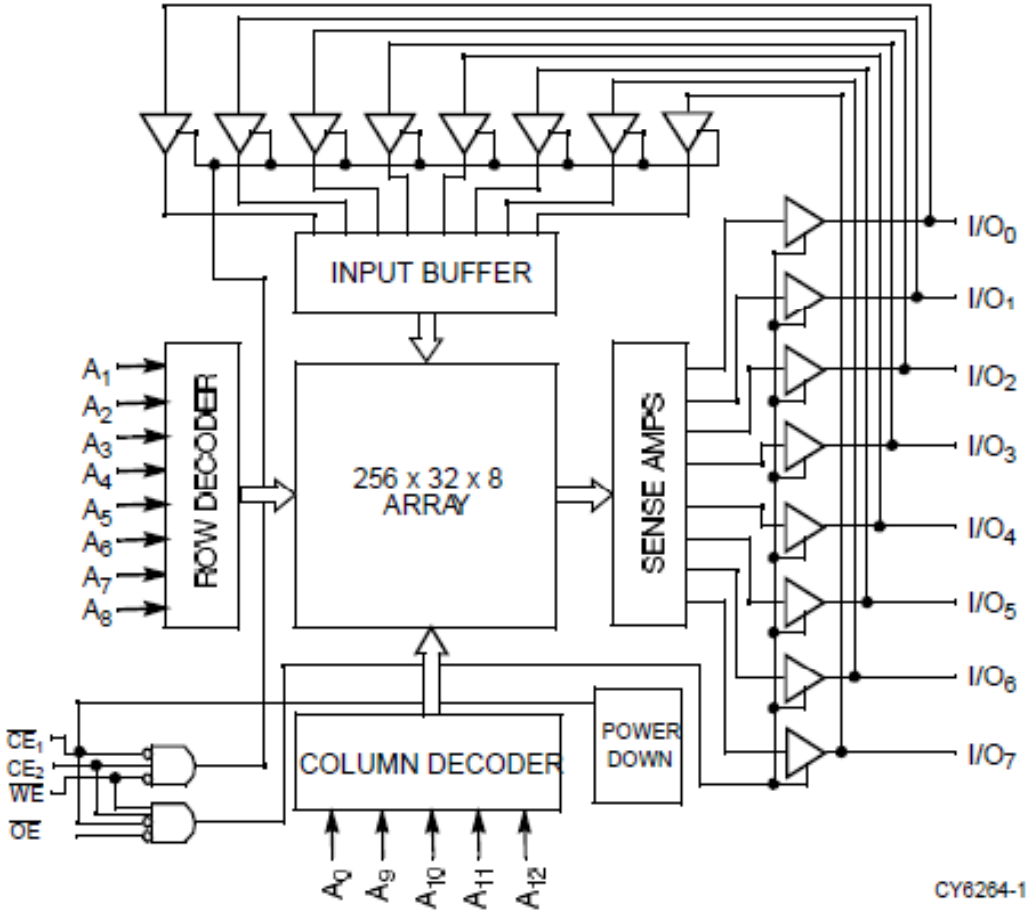
6T-SRAM cell.

Source: Computer Organization and Architecture, by W. Stallings

3.3 SRAM

Example: 8Kx8bit Cypress Static RAM (CY6264)





Pin configuration and logic block diagram

3.3 SRAM

SRAM vs. DRAM

- Both are volatile
- SRAM is faster
  - Used for cache memory
- DRAM is smaller and cheaper
- DRAM is more dense
  - Used for main memory
- DRAM needs refresh

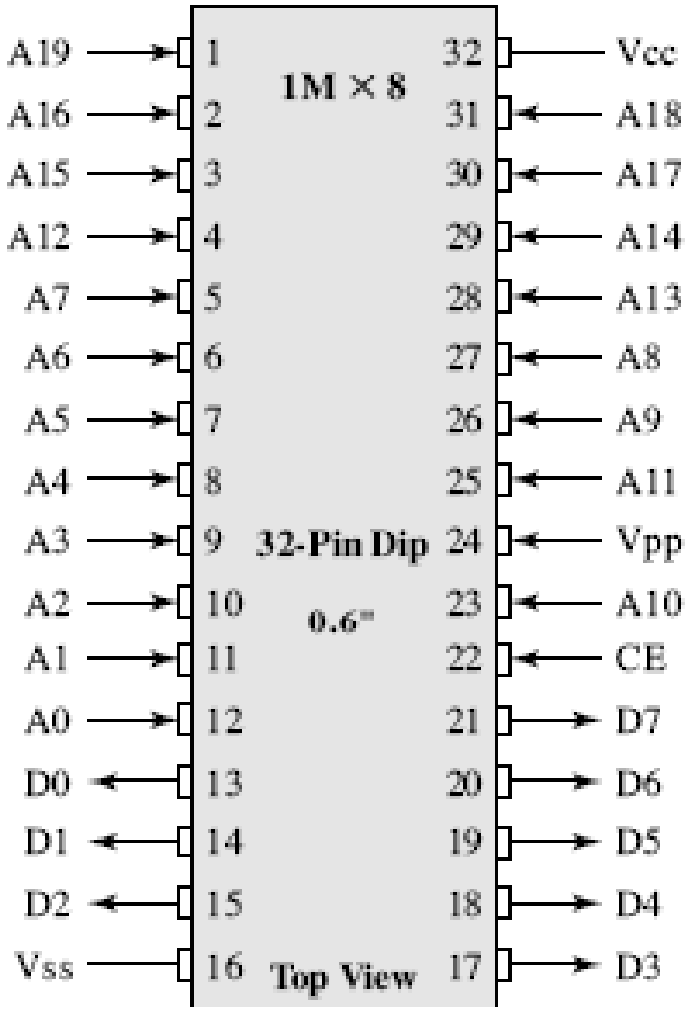
3. Memory Types

<table>
<tr><th>Memory Type</th><th>Category</th><th>Erasure</th><th>Write Mechanism</th><th>Volatility</th></tr>
<tr><td>Random-access memory (RAM)</td><td>Read-write memory</td><td>Electrically, byte-level</td><td>Electrically</td><td>Volatile</td></tr>
<tr><td>Read-only memory (ROM)</td><td rowspan="2">Read-only memory</td><td rowspan="2">Not possible</td><td>Masks</td><td rowspan="6">Nonvolatile</td></tr>
<tr><td>Programmable ROM (PROM)</td><td rowspan="4">Electrically</td></tr>
<tr><td>Erasable PROM (EPROM)</td><td rowspan="3">Read-mostly memory</td><td>UV light, chip-level</td></tr>
<tr><td>Electrically Erasable PROM (EEPROM)</td><td>Electrically, byte-level</td></tr>
<tr><td>Flash memory</td><td>Electrically, block-level</td></tr>
</table>

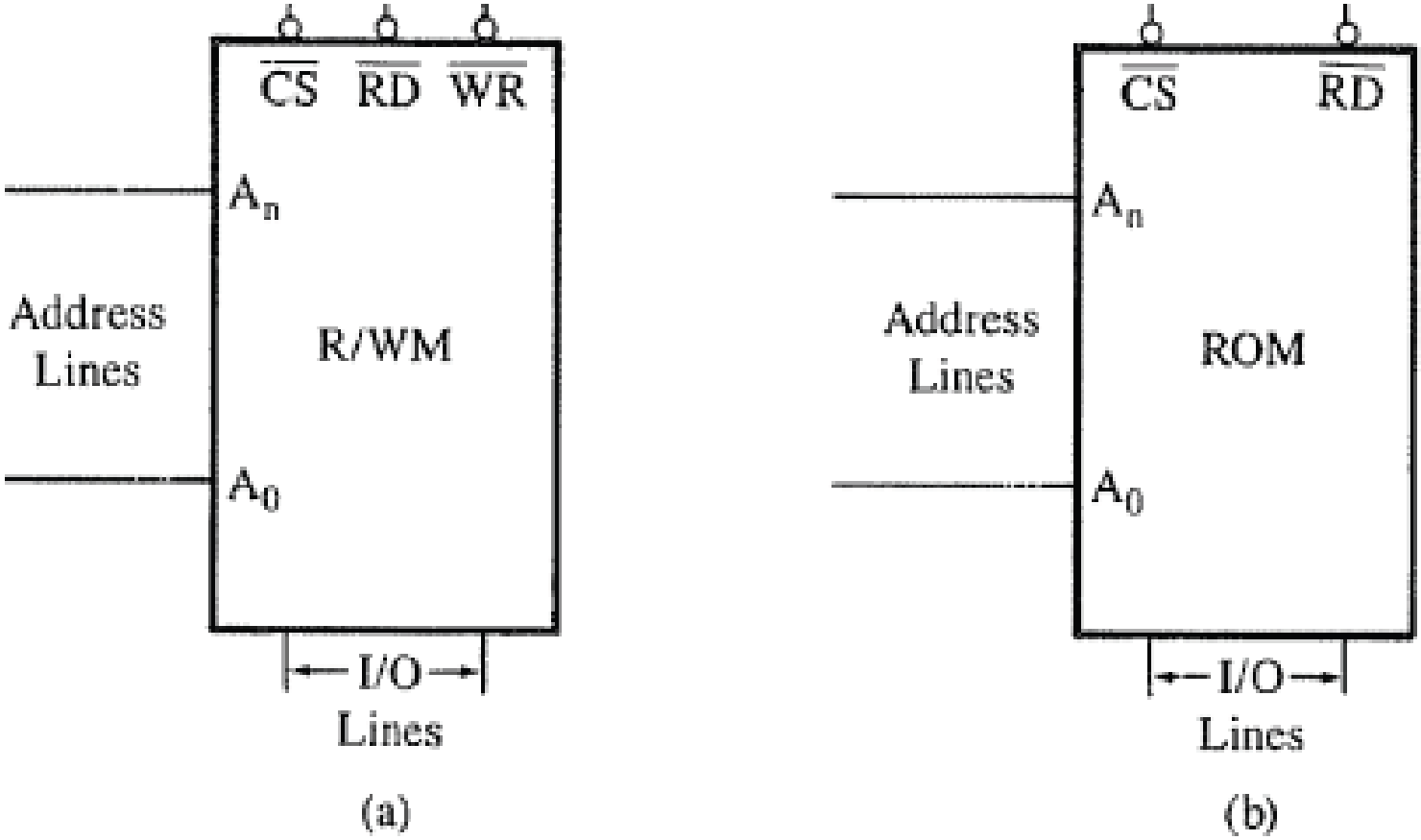
Major types of semiconductor memories

4. Memory Chip

As with other integrated circuit products, semiconductor memory comes in packaged chips. Each chip contains an array of memory cells

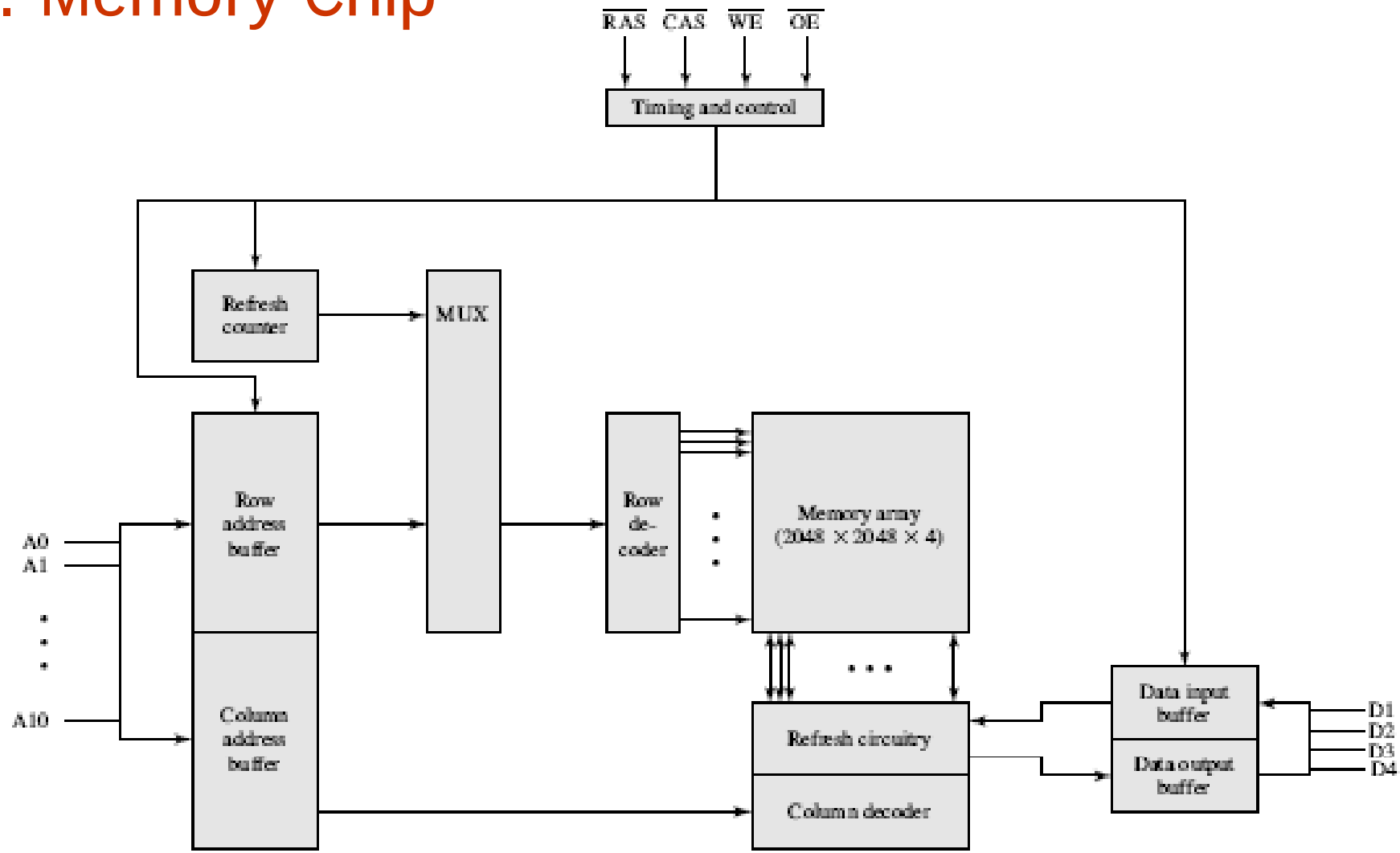


4. Memory Chip



Typical Memory Chip: a) RAM, and b) ROM.

4. Memory Chip



Example: Typical 16 Mbits DRAM (4Mx4bits)

Source: Computer Organization and Architecture, by W. Stallings

4. Memory Chip

Example: 16 Mbits DRAM (4Mx4bits)

- Note that there are only 11 address lines (A0–A10),
  - We expect 22
- The 22 required address lines are multiplexed onto the 11 address lines
  - First 11 address signals fo row address of the array
  - Other 11 address signals for the column address
- Row address select (RAS) and column address select (CAS) signals are used to provide timing
- This is done to save on the number of pins

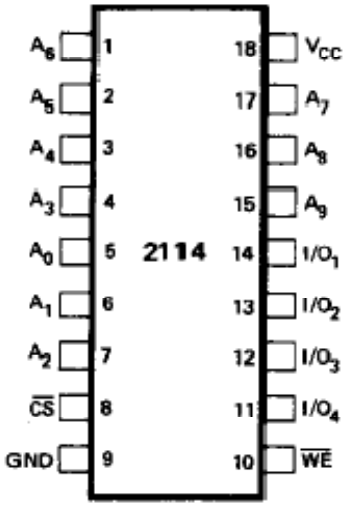
4. Memory Chip

Example: 16 Mbits DRAM (Cont'd)

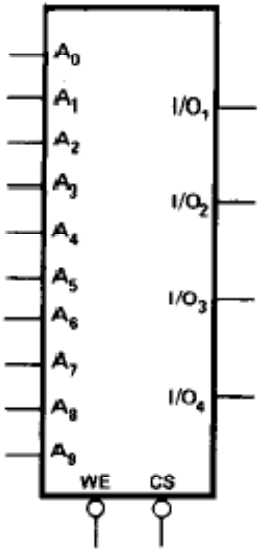
- To address the chip, first row is selected by putting n-bit number on the address pins, and the RAS signal is asserted
- Then, a column number is put on the address pins and CAS is asserted
- The chip responds by reading or writing the data

4. Memory Chip

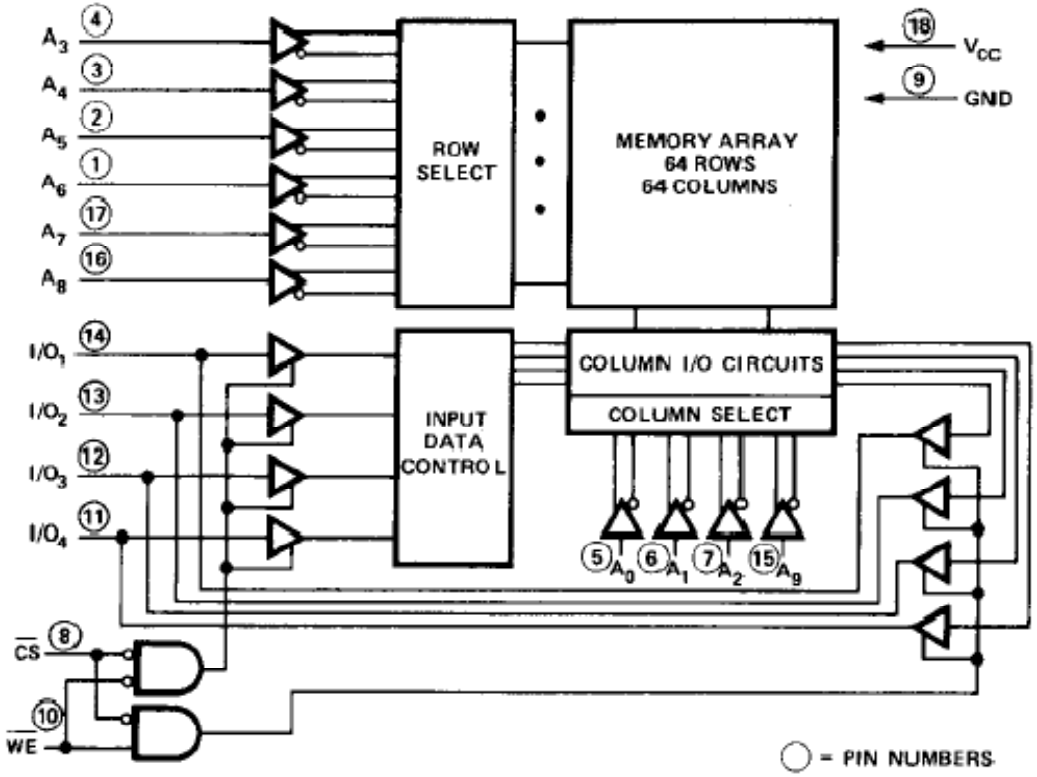
PIN CONFIGURATION



LOGIC SYMBOL



BLOCK DIAGRAM



PIN NAMES

A_0–A_9	ADDRESS INPUTS	V_{CC} POWER (+5V)
$\overline{WE}$	WRITE ENABLE	GND GROUND
$\overline{CS}$	CHIP SELECT	
I/O_1–I/O_4	DATA INPUT/OUTPUT	

1Kx4-bit SRAM (part number 2114N)

Source: Computer Organization and Architecture, by W. Stallings

4. Memory Chip

Memory Module Organization

- There are several ways to organize memory modules with respect to the way the memory chips are connected to the CPU

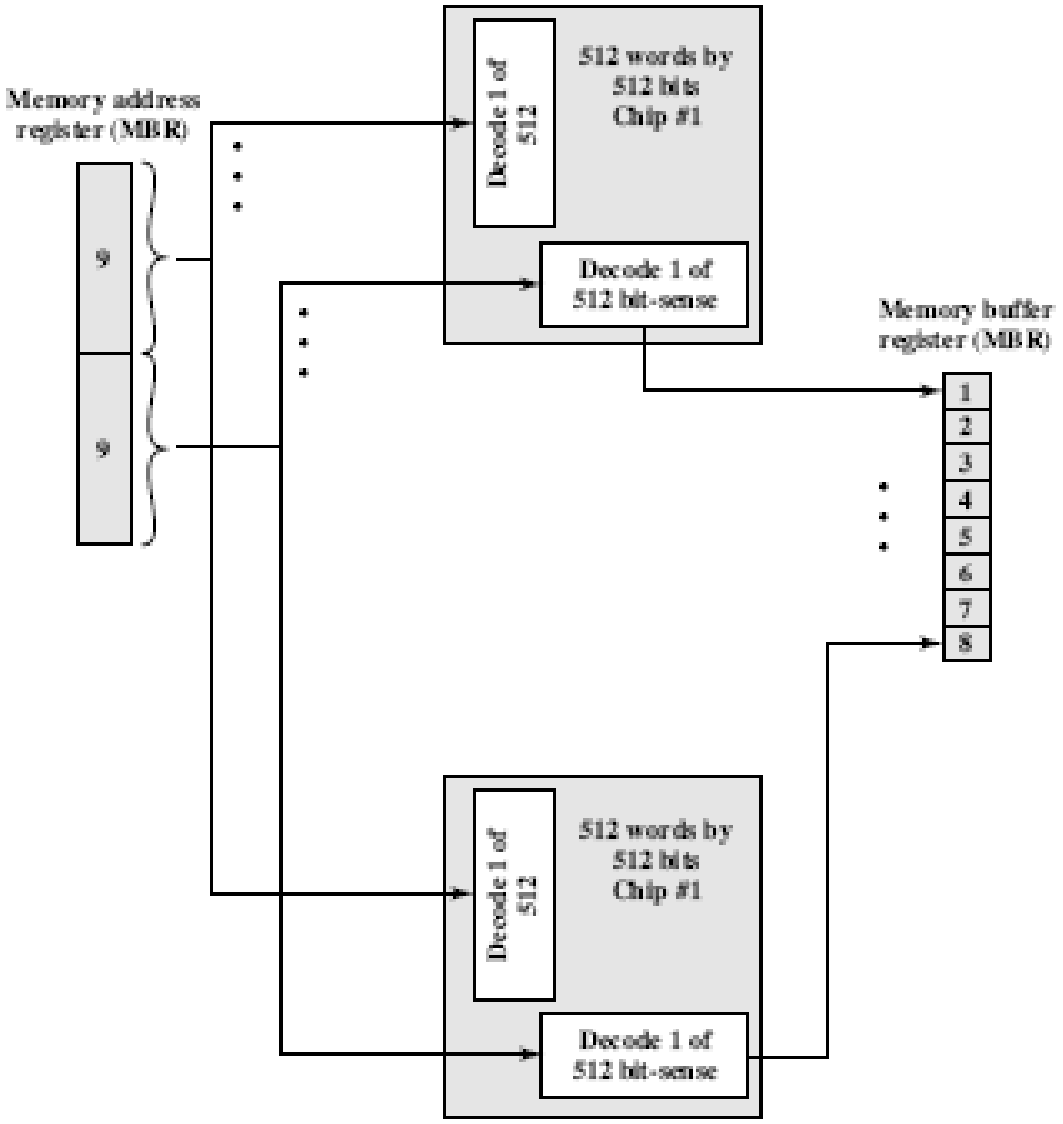
Example

- If a RAM chip contains only 1 bit per word, then clearly we will need at least a number of chips equal to the number of bits per word

4. Memory Chip

Memory Module Organization Example:

- 256 KBytes memory module organized as eight (8) 256Kx1-bit chips



Source: Computer Organization and Architecture, by W. Stallings

4. Memory Chip

Memory Module Organization

In a PC, a group of chips, typically 8 or 16, is mounted on a tiny printed circuit board (PCB) and sold as a unit called SIMM or DIMM

- SIMM (Single Inline Memory Module) card has a row of connectors on one side of the board
- DIMM (Dual Inline Memory Module) card has a row of connectors on both sides of the board

- 1GBytes DIMM card contains 16 chips of 64MBytes
  - Why not one single 1GB memory chip?

4. Memory Chip

Memory Module Organization

- Synchronous RAM (SDRAM) exchanges data with the processor synchronized to an external clock signal and running at the full speed of the bus without imposing wait states
- Double-Data-Rate SDRAM (DDR SDRAM) can send data twice per clock cycle, once on the rising edge of the clock pulse and once on the falling edge
  - 100-200MHz and 2.5V for DDR, 200-400MHz and 1.8V for DDR2, and 400-800MHz and 1.5V for DDR3

5. Memory Map/Interfacing

Memory Map

- Typically, in a computer with N-bit address lines, it is capable of addressing 2^N memory locations
- The memory range is divided by the programmer (or hardware designer) into sections suited to specific roles (e.g. program section, data section, stack section, ROM, EEPROM, RAM, etc.)
- This memory range division is called Memory Map

5. Memory Map/Interfacing

Memory Map: Z80 board example

Location	Description
0000 3FFF	Monitor EEPROM U24 (?? KB)
4000 5DFF	User RAM U26, User Stack
5E00 5FFF	Monitor Stack
6000 7FFF	Expansion (already interfaced)
8000 FFFF	Unused

Memory Interfacing

The primary function of memory interfacing is to allow the CPU to read from and write into a location of a given memory chip

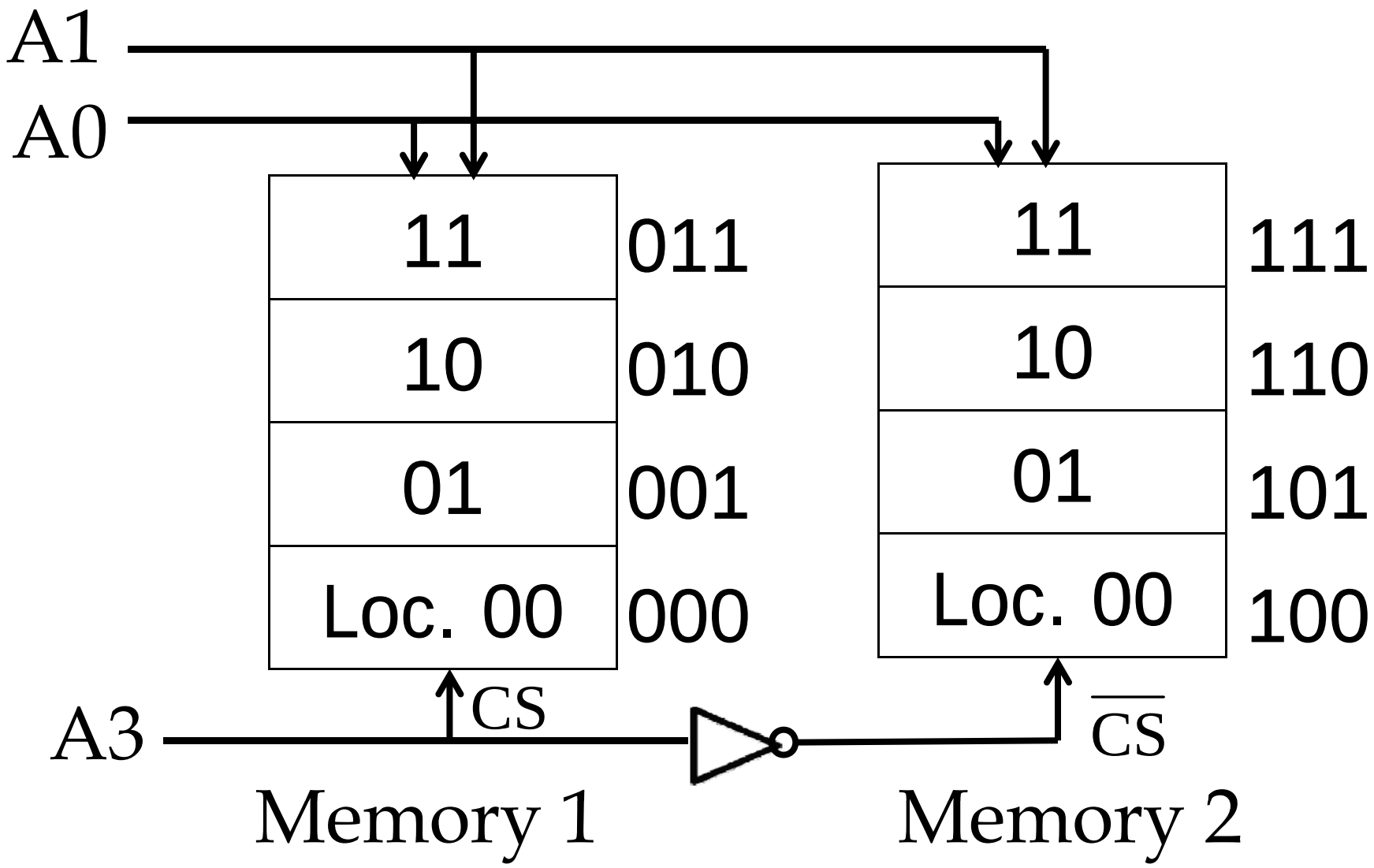
- Be able to select the chip using the chip select (CS) or chip enable (CE) signal
  - Memory address decoding
- Identify the location using the supplied address
- Enable the appropriate buffer using the Read Enable (RD) or the Write Enable (WR) signals

Memory Interfacing

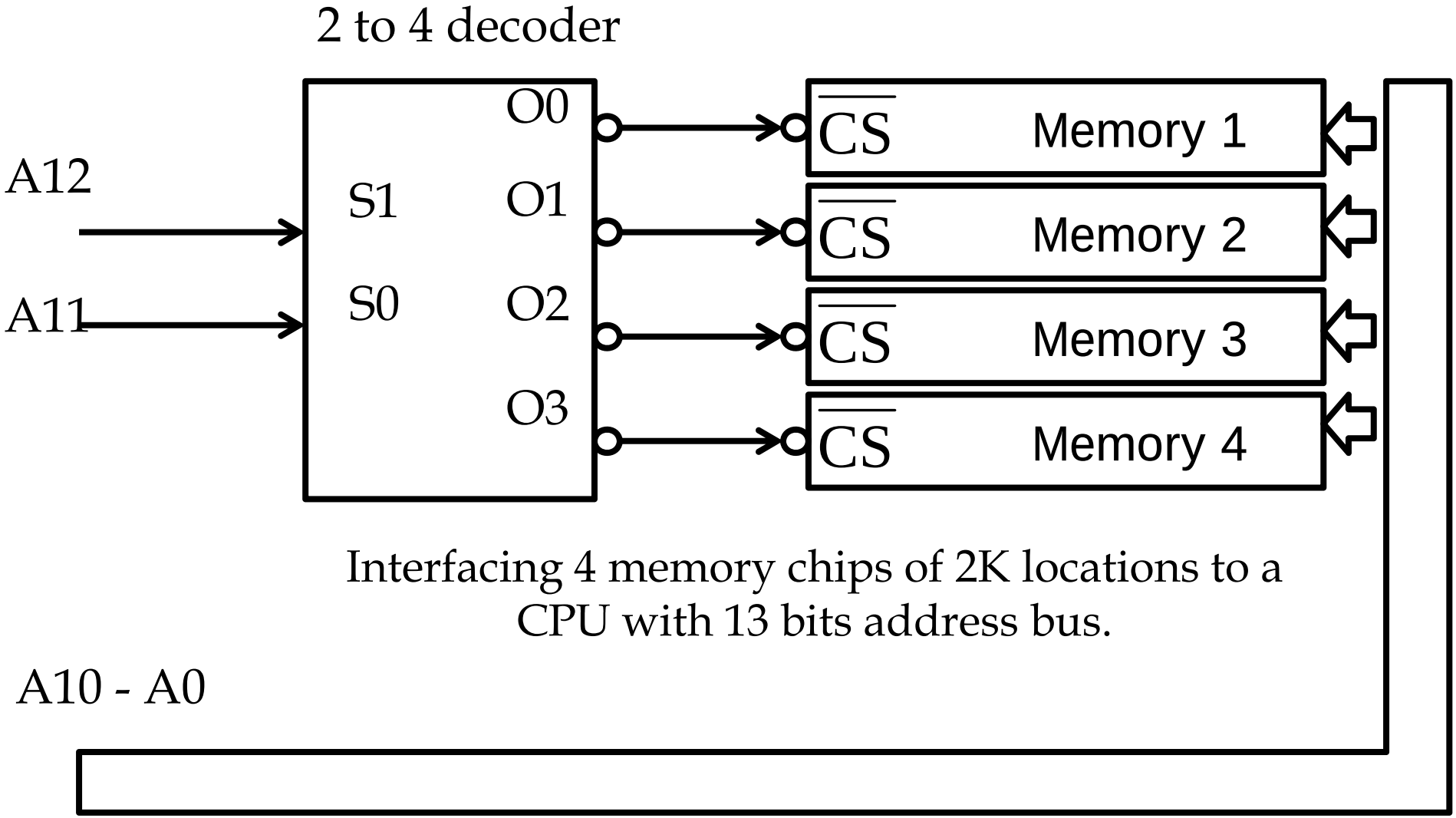
To interface memory with CPU:

- Connect the data lines of the memory chip to the data bus
- Connect the required address lines of the memory chip to the address bus
- Decode the remaining address lines to generate chip select signal (address decoding)
- Connect the read and write memory signals

5. Memory Map/Interfacing

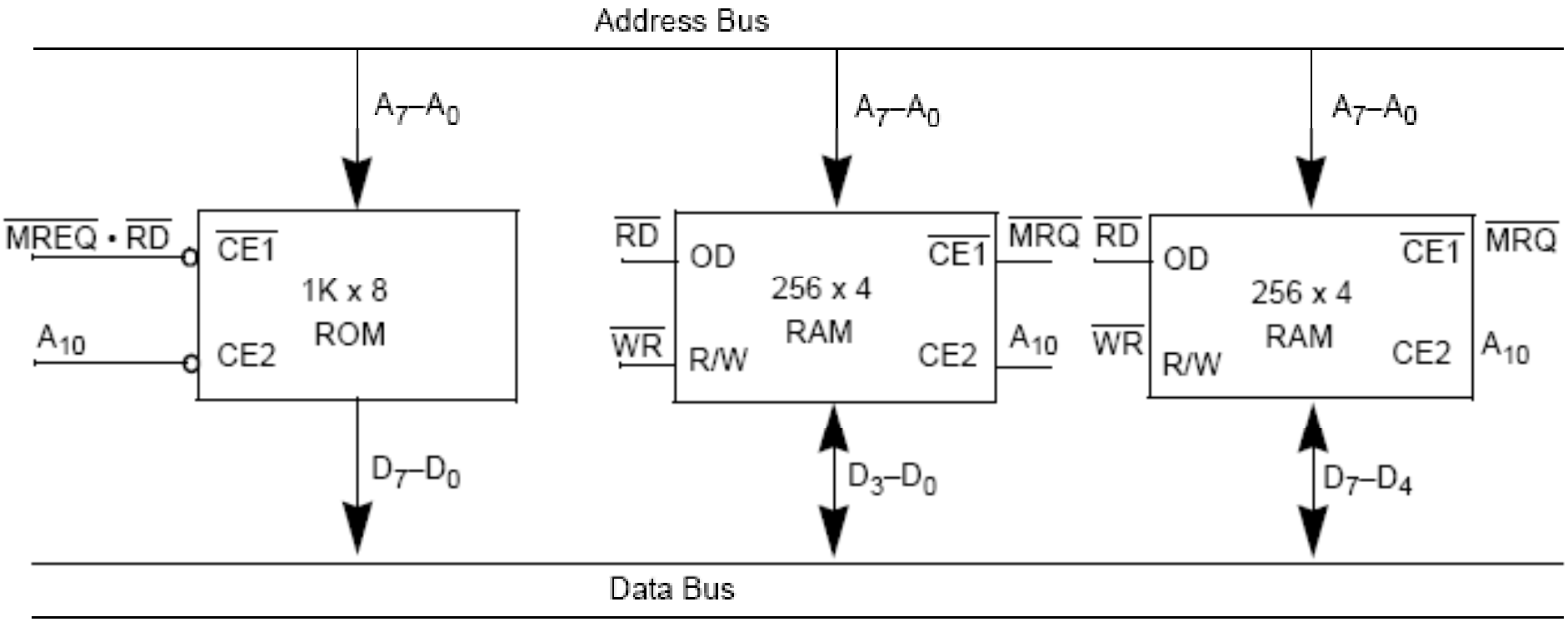


5. Memory Map/Interfacing



Interfacing 4 memory chips of 2K locations to a CPU with 13 bits address bus.

5. Memory Map/Interfacing



Z80-Memory Interfacing Example

MREQ: Memory Request Signal
Memory Map : ROM from 00...00 to 011...11 (000H to 8FFH)
RAM from 1XX0..0 to 1XX1..1 (800H to 9FFH if XX=00), (E00H to FFFH if XX=11) ==> Foldback memory (i.e. Multiple addresses same location)

6. Byte Ordering

Data stored in multiple locations may be ordered from left to right or right to left

- Example: 32 bit hex value 12345678H is stored in a byte-addressable memory at location 10H

	14h
12	13h
34	12h
56	11h
78	10h

Little-endian (e.g Intel X86 processors)

	14h
78	13h
56	12h
34	11h
12	10h

Big-endian (e.g Motorola and Sparc processors)

7. Cache Memory

In general, it is likely that most future accesses to main memory by the processor will be to locations recently accessed

- So the cache automatically retains a copy of some of the recently used words from the DRAM
- If the cache is designed properly, then most of the time the processor will request memory words that are already in the cache

7. Cache Memory

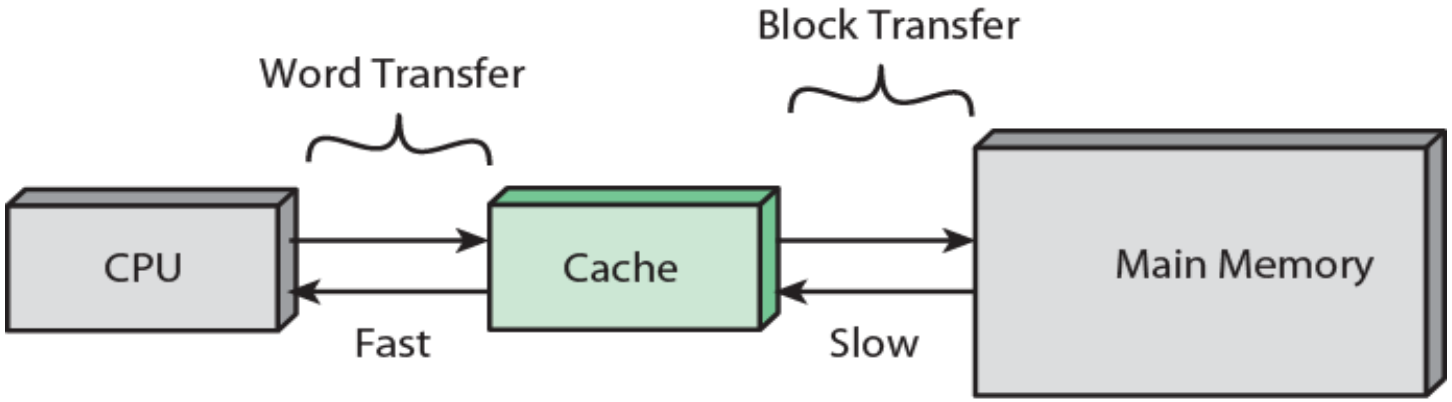
What is Cache Memory?

Cache memory is small amount of memory which:

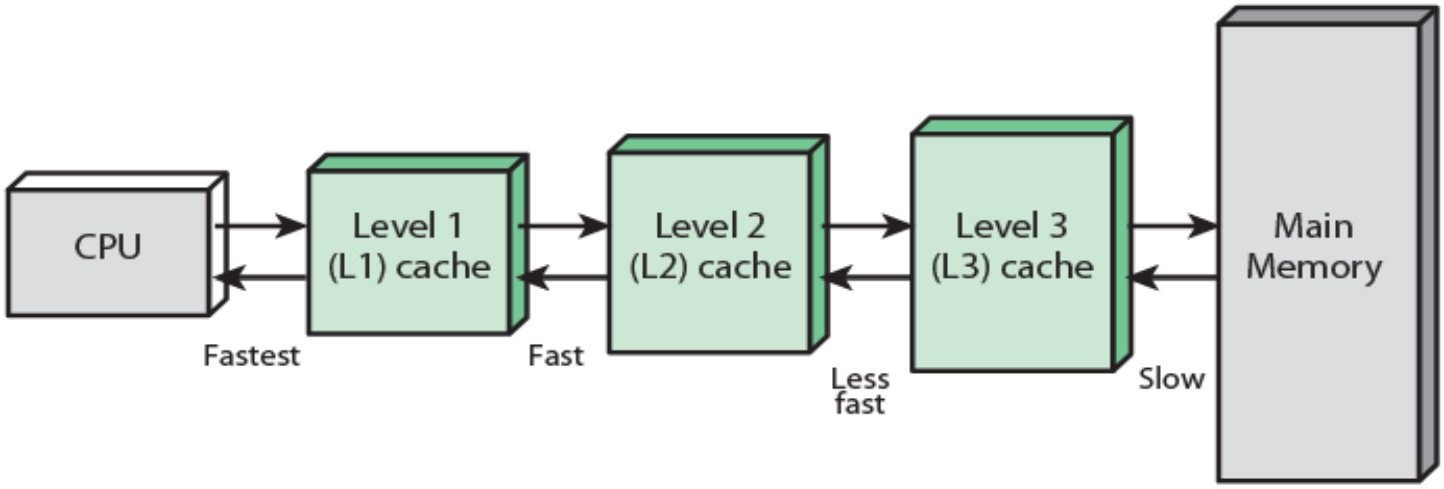
- is very fast
- sits between normal main memory and CPU
- may be located on CPU chip, CPU package or computer boaord

7. Cache Memory

Single cache and 3-level cache organization



(a) Single cache



(b) Three-level cache organization

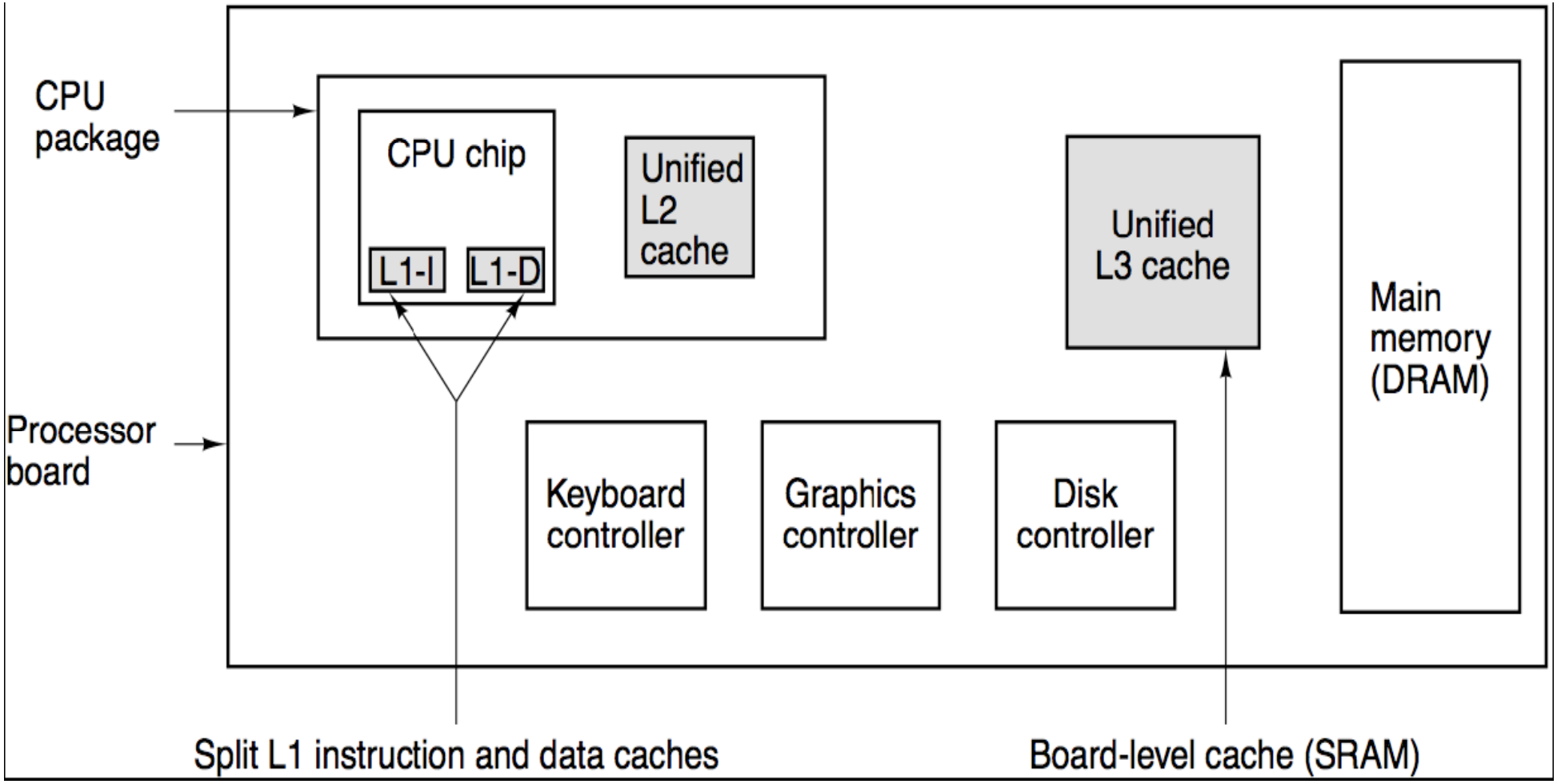
Source: Computer Organization and Architecture, by W. Stallings

7. Cache Memory

Cache Levels

- The CPU chip itself contains a small Level-1 Instruction cache (L1-I) and a small Level-1 Data cache (L1-D).
- Then there is the Level-2 (L2), which is not on the CPU chip. It may be included in the CPU package, next to the CPU chip and connected to it by a high-speed path. This cache is generally unified.
- The Level-3 cache (L3) is on the processor board and consists of few megabytes of SRAM

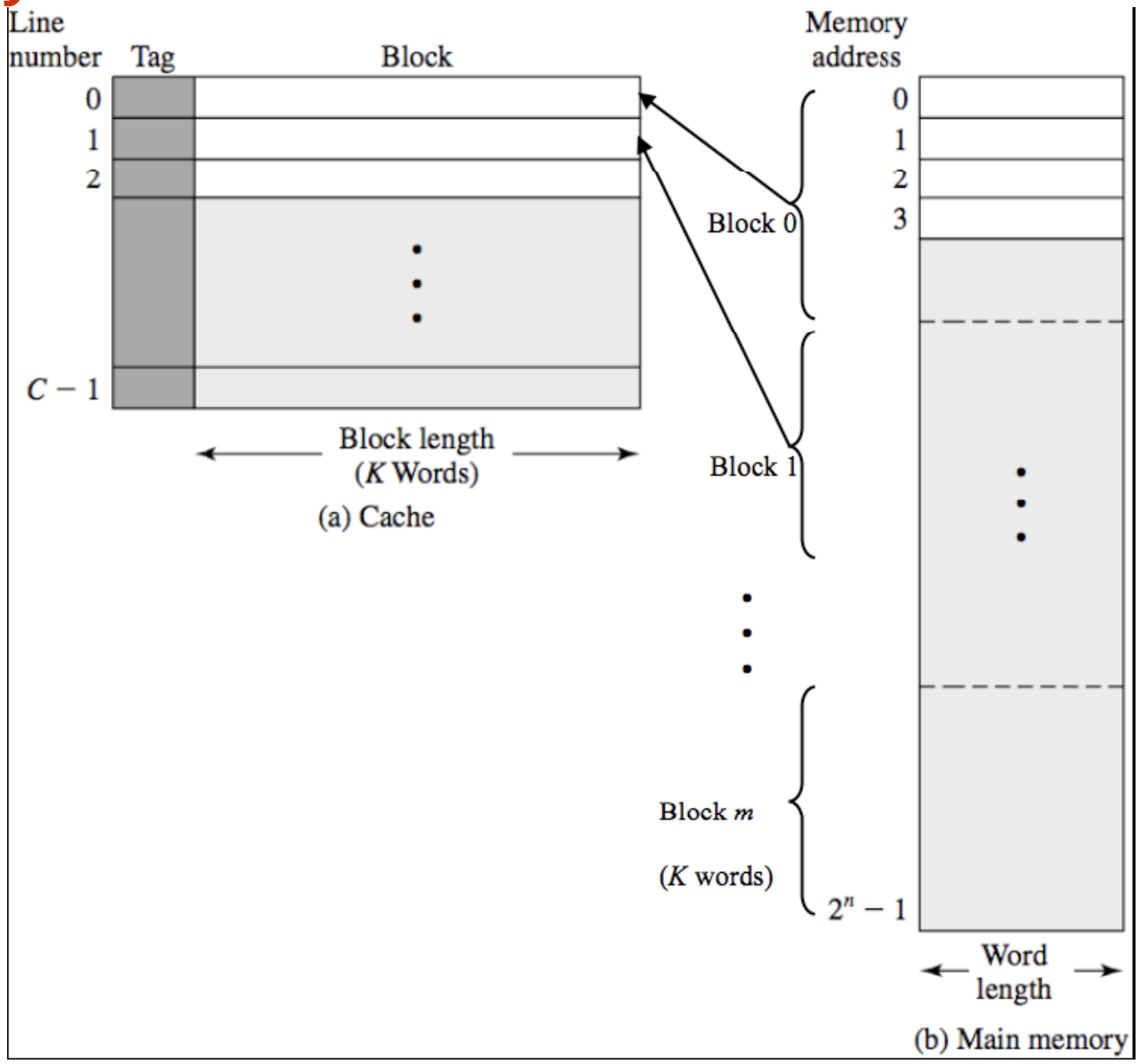
7. Cache Memory



Example of Memory Cache Organization

Source: Computer Organization and Architecture, by W. Stallings

7. Cache Memory



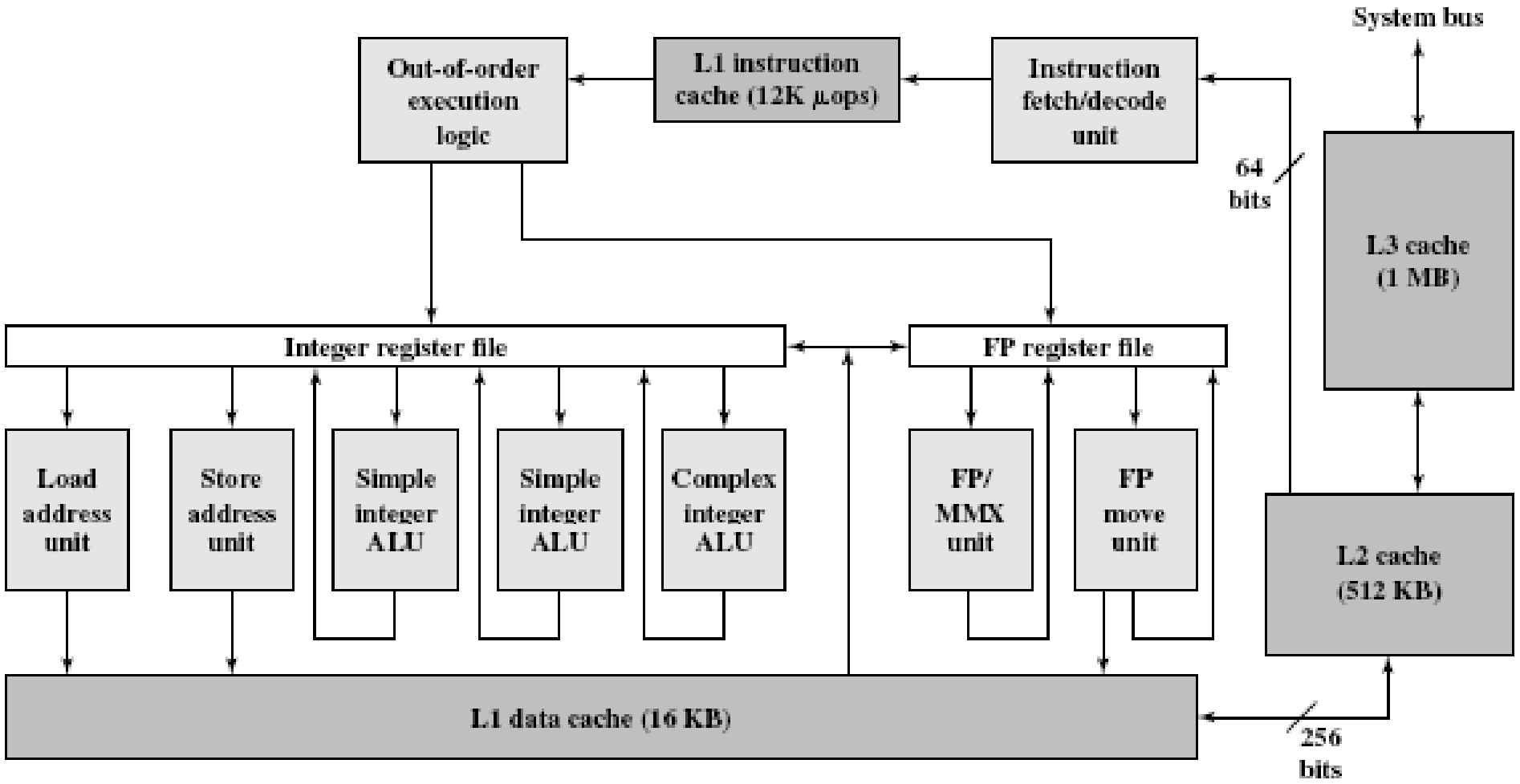
Direct-Mapped Cache

7. Cache Memory

Cache Operation - Overview

- CPU requests contents of memory location
- Check cache for this data
- If present, get from cache (fast)
- If not present, read required block from main memory to cache
- Then deliver from cache to CPU
- Cache includes tags to identify which block of main memory is in each cache slot

7. Cache Memory



Pentium 4 Block Diagram with Caches

Source: Computer Organization and Architecture, by W. Stallings

8. External Memory

Magnetic Disk

- Hard Drive Disk (HDD), magnetic tape

Optical Disc

- Compact Disc (CD): 700MB capacity
- Digital Versatile Disc (DVD): 4.7 GB capacity
- Blu Ray: 25-50-100-128 GB capacity

Solid State Storage (Integrated Circuits)

- Flash memory (compact flash, USB flash drive, memory stick, secure digital (SD) card)
- Solid State Disk (SSD)

Summary

Computer Memory System Overview

Internal Main Memory

- RAM and ROM
- DRAM and SRAM
- Memory Chip and Packaging
- Memory Map and Interfacing

Cache Memory

External Memory