Chapter 5 Internal Memory

William Stallings, Computer Organization and Architecture, 9th Edition

Objectives

- How are main memory structured?
- Whether main memory may cause errors?
- How many types of memory?
- After studying this chapter, you should be able to:
  - Present an overview of the principle types of semiconductor main memory.
  - Understand the operation of a basic code that can detect and correct singlebit errors in 8-bit words.
  - Summarize the properties of contemporary advanced DRAM organizations.

Contents

- 5.1 Semiconductor Main Memory
- 5.2 Error Correction
- 5.3 Advanced Dram Organization

Semiconductor- Chất bán dẫn (silic, germanium) là vật liệu trung gian giữa chất dẫn điện và chất cách điện. Chất bán dẫn chỉ hoạt động như một chất dẫn điện ở một điều kiện nào đó. Chất bán dẫn được dùng để tạo ra các transistor (transfer-resistor).

5.1- Semiconductor Main Memory

- Organization
- Semiconductor Memory Types
- Dynamic RAM and Static RAM
- Types of ROM
- Chip Logic
- Chip Packaging
- Module Organization
- Interleaved Memory

Organization

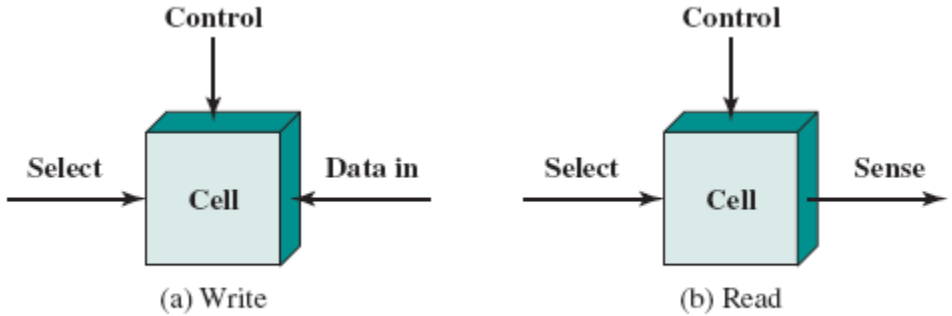


Figure 5.1 Memory Cell Operation

- Basic element of a semiconductor memory is the memory cell.
- Cell properties:
  - 1-They exhibit two stable (or semistable) states, which can be used to represent binary 1 and 0.
  - 2- They are capable of being written into (at least once), to set the state.
  - 3- They are capable of being read to sense the state

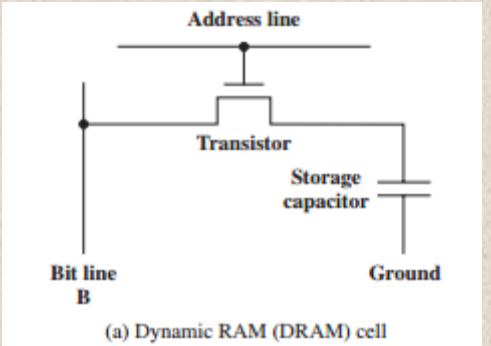
Semiconductor Memory Types

Table 5.1 Semiconductor Memory Types

Memory Type	Category	Erasure	Write Mechanism	Volatility
Random-access memory (RAM)	Read-write memory	Electrically, byte-level	Electrically	Volatile
Read-only memory (ROM)	Read-only memory	Not possible	Masks	Nonvolatile
Programmable ROM (PROM)			Electrically	
Erasable PROM (EPROM)	Read-mostly memory	UV light, chip-level		
Electrically Erasable PROM (EEPROM)		Electrically, byte-level		
Flash memory		Electrically, block-level		

All of the memory types that we will explore in this chapter are random access. That is, individual words of memory are directly accessed through wired-in addressing logic.

Dynamic RAM (DRAM)



(a) Dynamic RAM (DRAM) cell

Figure 5.2 Typical Memory Cell Structures

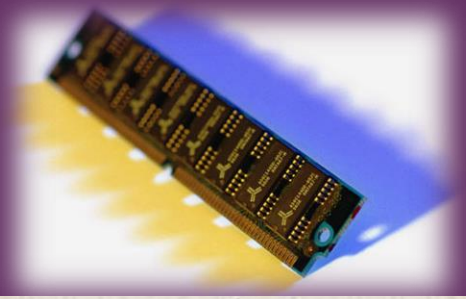
- RAM technology is divided into two technologies:
  - Dynamic RAM (DRAM)
  - Static RAM (SRAM)

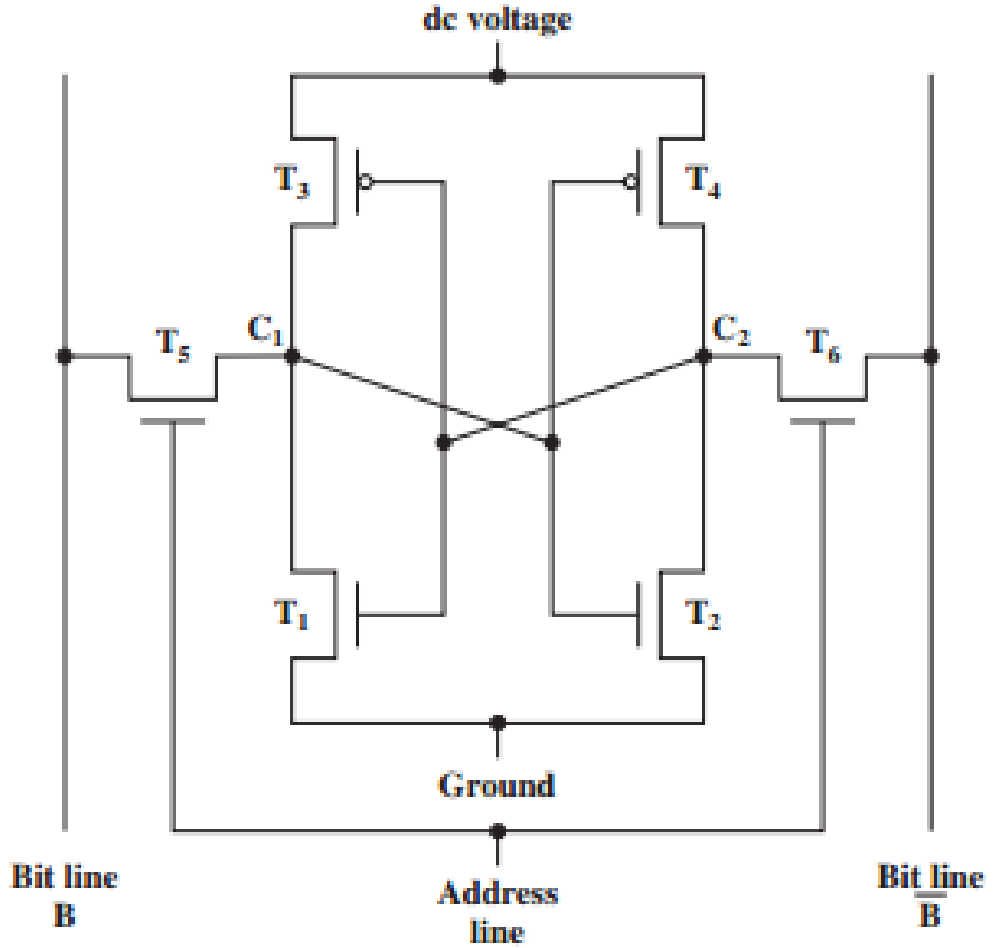
How Dram cell works? Read by yourself.

- DRAM
  - Made with cells that store data as charge on capacitors (tụ điện)
  - Presence or absence of charge in a capacitor is interpreted as a binary 1 or 0
  - Requires periodic charge refreshing to maintain data storage
  - The term *dynamic* refers to tendency of the stored charge to leak away, even with power continuously applied

Static RAM (SRAM)

- Digital device that uses the same logic elements used in the processor
- Binary values are stored using traditional flip-flop logic gate configurations
- Will hold its data as long as power is supplied to it





(b) Static RAM (SRAM) cell

Figure 5.2 Typical Memory Cell Structures

SRAM versus DRAM

- **Both volatile:** Power must be continuously supplied to the memory to preserve the bit values

- **Dynamic cell**
  - Simpler to build, smaller
  - (smaller cells = more cells per unit area)
  - Less expensive
  - Requires the supporting refresh circuitry
  - Tend to be favored for large memory requirements
  - Used for main memory

- **Static**
  - Faster
  - Used for cache memory (both on and off chip)

SRAM

DRAM

Read Only Memory (ROM)

- Contains a permanent pattern of data that cannot be changed or added to
- No power source is required to maintain the bit values in memory
- Data or program is permanently in main memory and never needs to be loaded from a secondary storage device
- Data is actually wired into the chip as part of the fabrication process
  - Disadvantages of this:
    - No room for error, if one bit is wrong the whole batch of ROMs must be thrown out
    - Data insertion step includes a relatively large fixed cost

Programmable ROM (PROM)

- Less expensive alternative
- Nonvolatile and may be written into only once
- Writing process is performed electrically and may be performed by supplier or customer at a time later than the original chip fabrication
- Special equipment is required for the writing process
- Provides flexibility and convenience
- Attractive for high volume production runs

Read-Mostly Memory

EPROM

- Erasable programmable read-only memory
- Erasure process can be performed repeatedly
- More expensive than PROM but it has the advantage of the multiple update capability

EEPROM

- Electrically erasable programmable read-only memory
- Can be written into at any time without erasing prior contents
- Combines the advantage of non-volatility with the flexibility of being updatable in place
- More expensive than EPROM

Flash Memory

- Intermediate between EPROM and EEPROM in both cost and functionality
- Uses an electrical erasing technology, does not provide byte-level erasure
- Microchip is organized so that a section of memory cells are erased in a single action or “flash”

Typical 16 Mb DRAM (4M x 4)

MUltipleXer is a device that selects one of several input signals and forwards the selected input into a single line

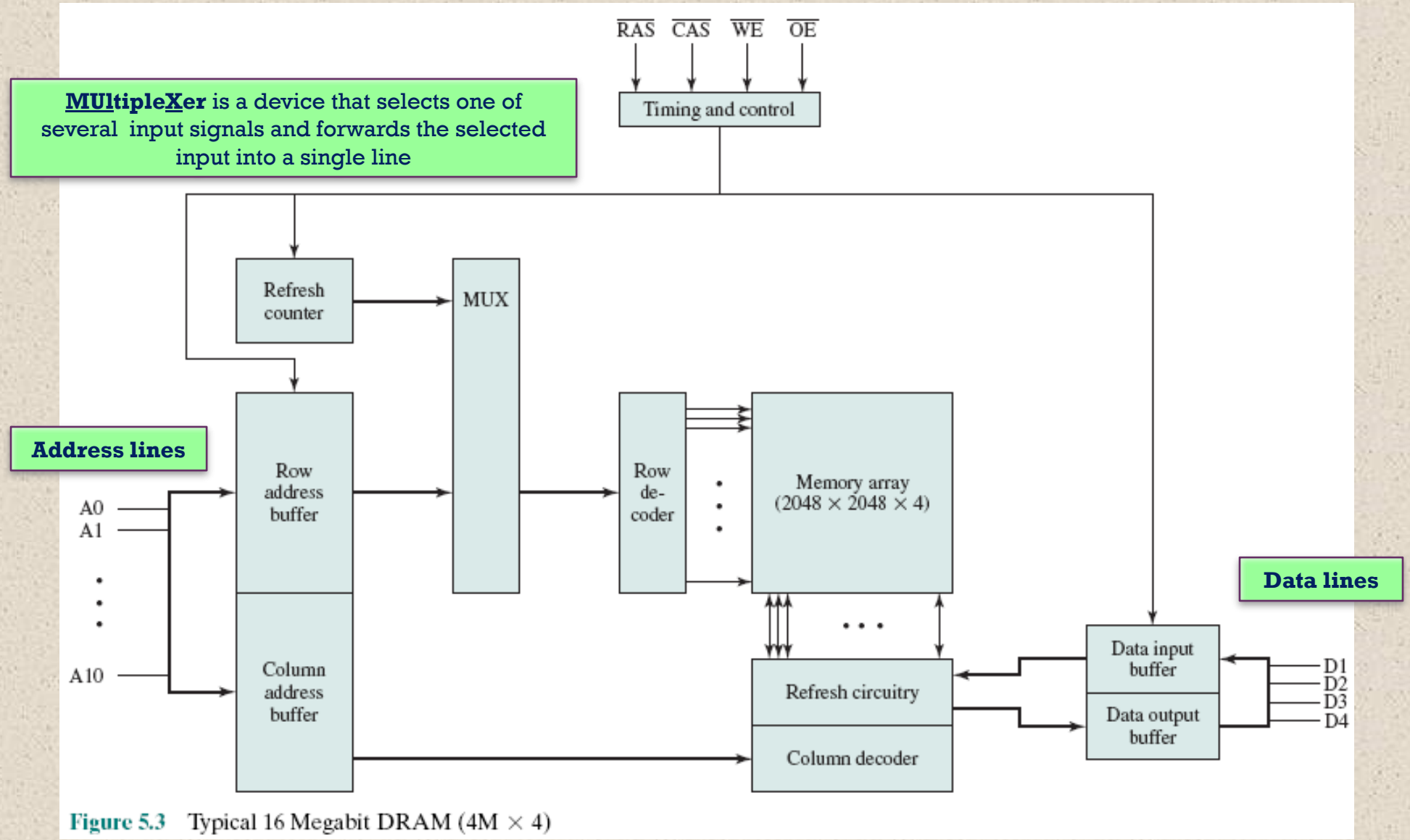
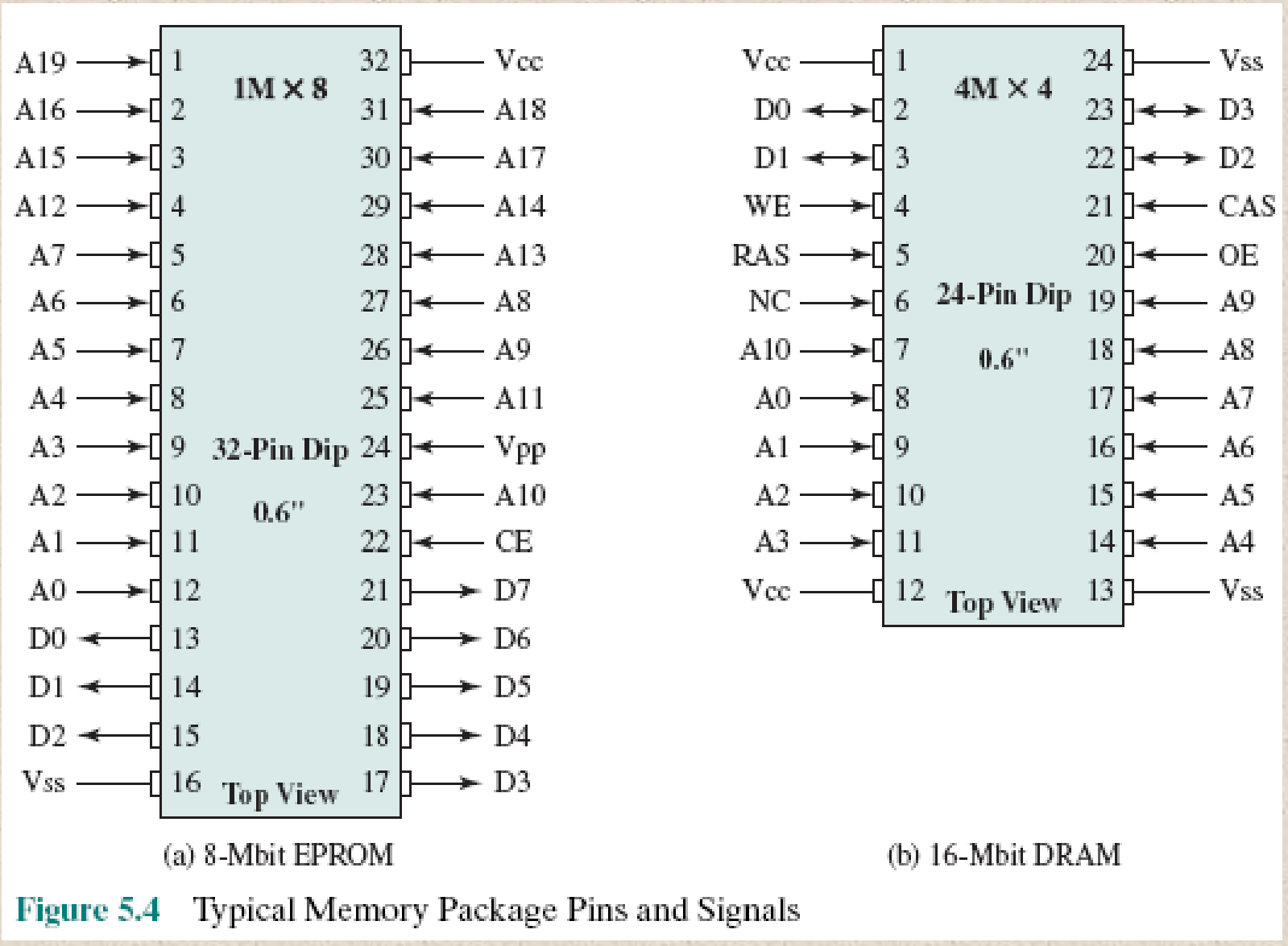


Figure 5.3 Typical 16 Megabit DRAM (4M × 4)

Chip Packaging



(a) 8-Mbit EPROM

(b) 16-Mbit DRAM

Figure 5.4 Typical Memory Package Pins and Signals

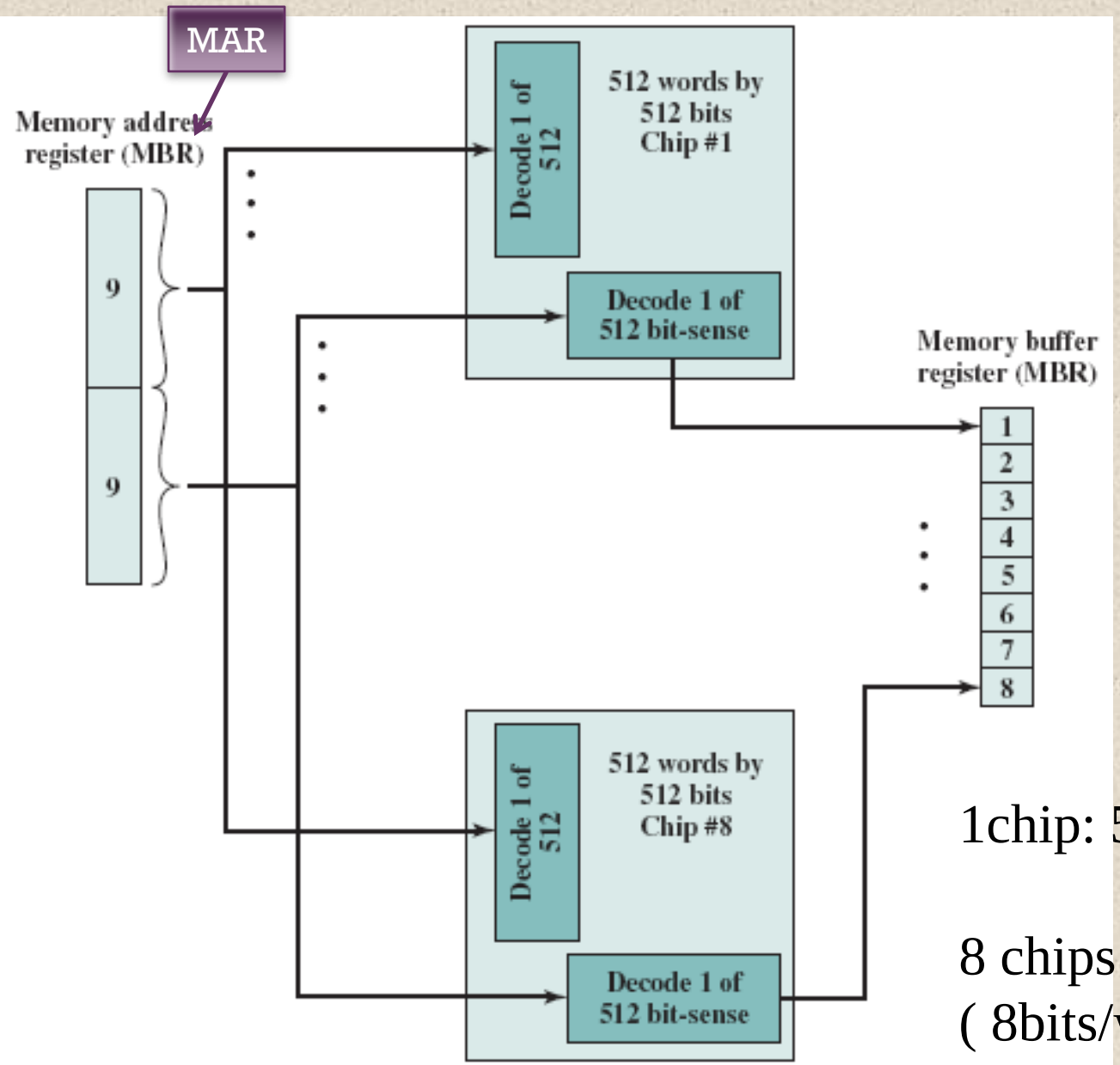


Figure 5.5 256-KByte Memory Organization

Figure 5.5

256-KByte Memory Organization

1chip: 512*512= 2^{18} bits
=256kb

8 chips → 256KB
(8bits/word)

1MByte Module Organization

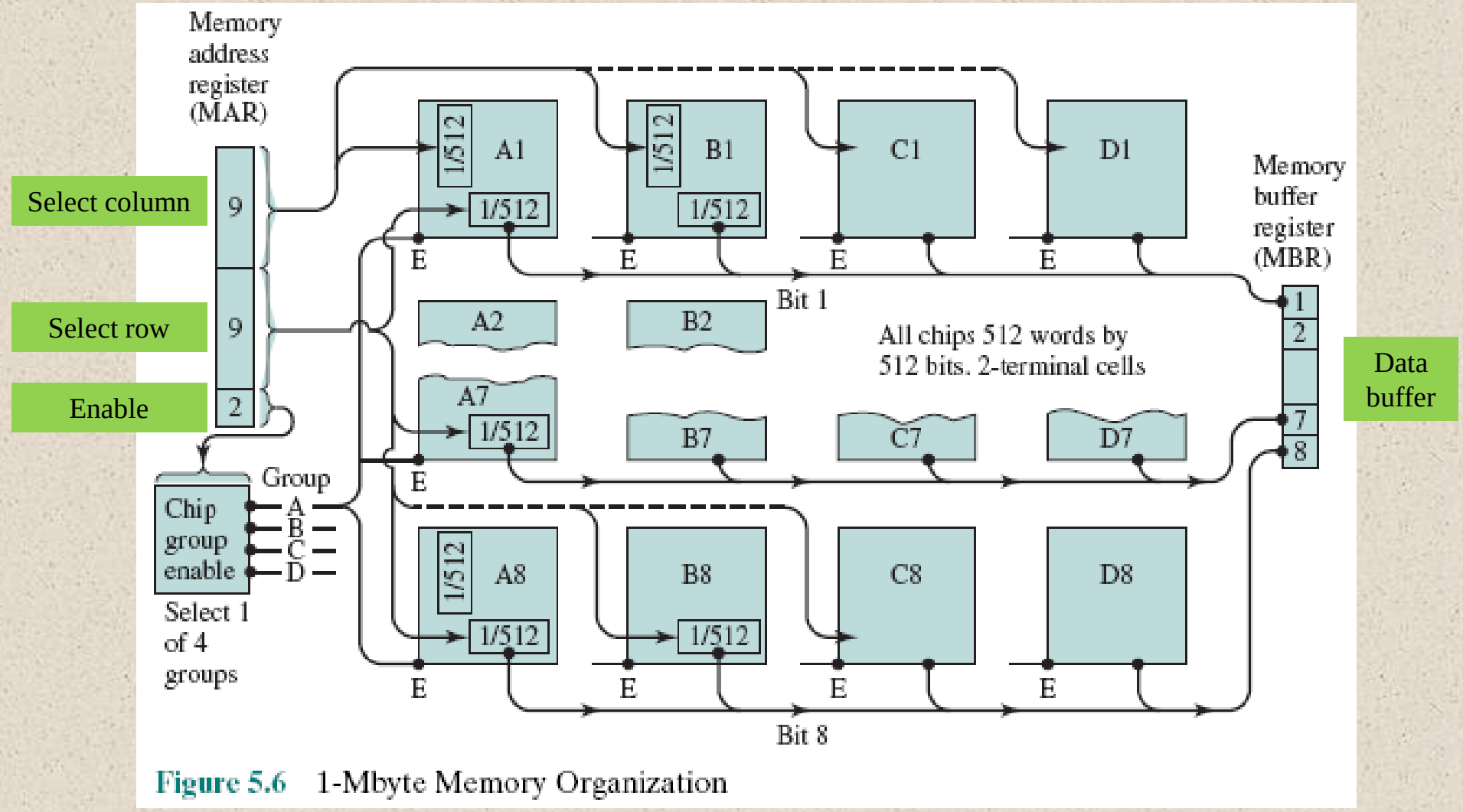
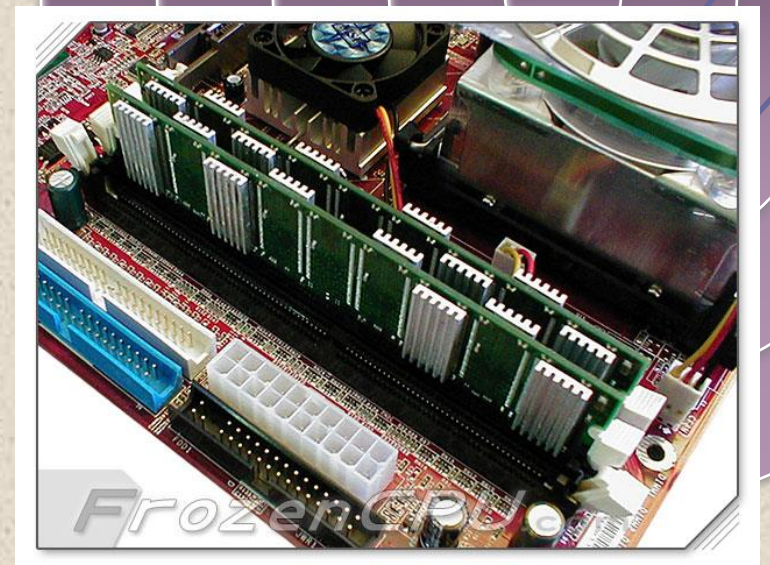


Figure 5.6 1-Mbyte Memory Organization

E: enable, signal permits the chip operating or not

Interleaved Memory



- Composed of a collection of DRAM chips
- Grouped together to form a *memory bank*
- Each bank is independently able to service a memory read or write request
- *K* banks can service *K* requests simultaneously, increasing memory read or write rates by a factor of *K*
- If consecutive words of memory are stored in different banks, the transfer of a block of memory is speeded up

5.2- Error Correction

- **Hard Failure**
  - Permanent physical defect (khuyết tật).
  - Memory cell or cells affected cannot reliably store data but become stuck at 0 or 1 or switch erratically between 0 and 1
  - **Can be caused by:**
    - Harsh (khắc nghiệt) environmental abuse(sự ngược đãi)
    - Manufacturing defects
    - Wear (hao mòn)
- **Soft Error**
  - Random, non-destructive event that alters the contents of one or more memory cells
  - No permanent damage to memory
  - **Can be caused by**:
    - Power supply problems
    - Alpha particles

Alpha particles: Phenomenon in which 2 protons and 2 neutrons bound together into a particle identical to a helium nucleus (Wiki for more details).

Error Correcting Code (ECC) Function

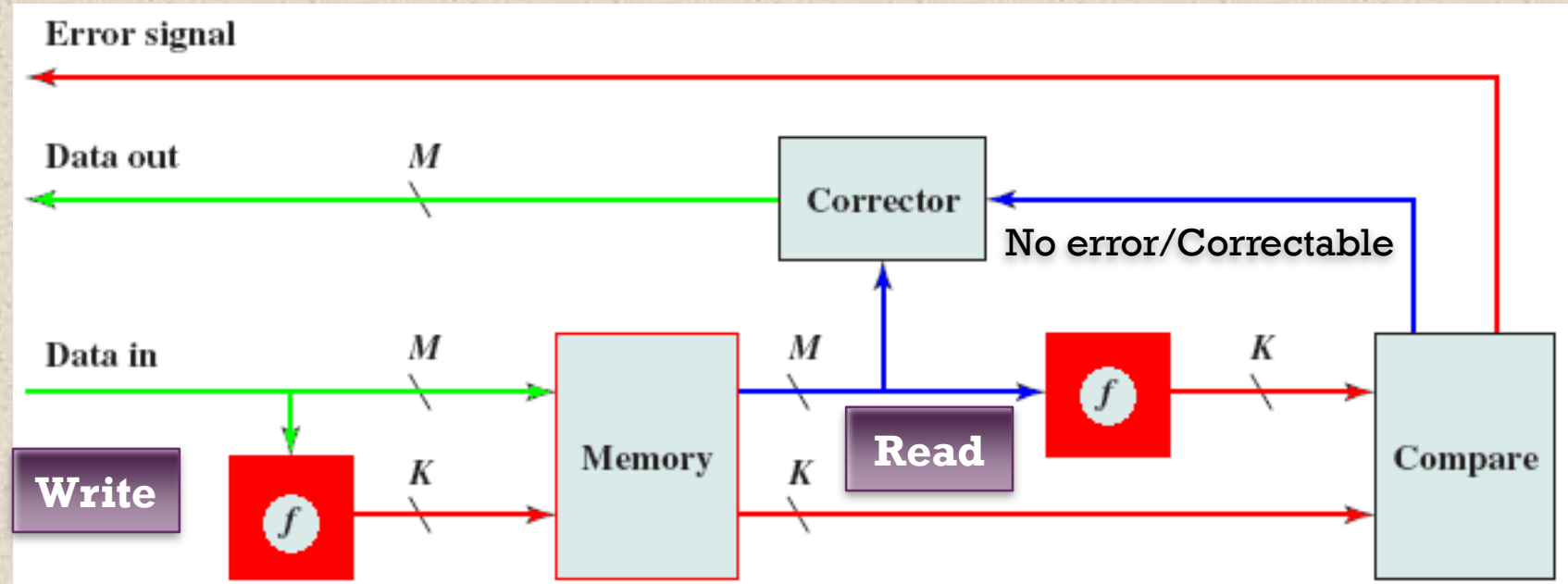


Figure 5.7 Error-Correcting Code Function

- No errors are detected. The fetched data bits are sent out.
- An error is detected, and it is possible to correct the error. The data bits plus **error correction** bits are fed into a corrector, which produces a corrected set of *M* bits to be sent out.
- An error is detected, but it is not possible to correct it. This condition is reported.

Next slide: An example for ECC function.

ECC Function: Examples

- The XOR operation is ussually used in ECC functions
- The most simple data for checking is the original data → A copy of original data is written to memory . 8-bit data: 00001111, ECC data: 00001111 → Memory must be increased to double size
- XORs some bits of M-bit original data to K-bit ECC will decrease memory size.
- Examples:

8 bits → 3 bits: **01010110 → 101**

8 bits → 2 bits: **01010110 → 00**

8 bits → 1 bits: **01010110 → 0**

- Main memory bank usually includes 9 chips. Why?

Hamming Error Correcting Code

Richard Hamming at Bell Laboratories

Parity bit (P) =1 if number of 1s is odd. Based on parity bit, data can be corrected.

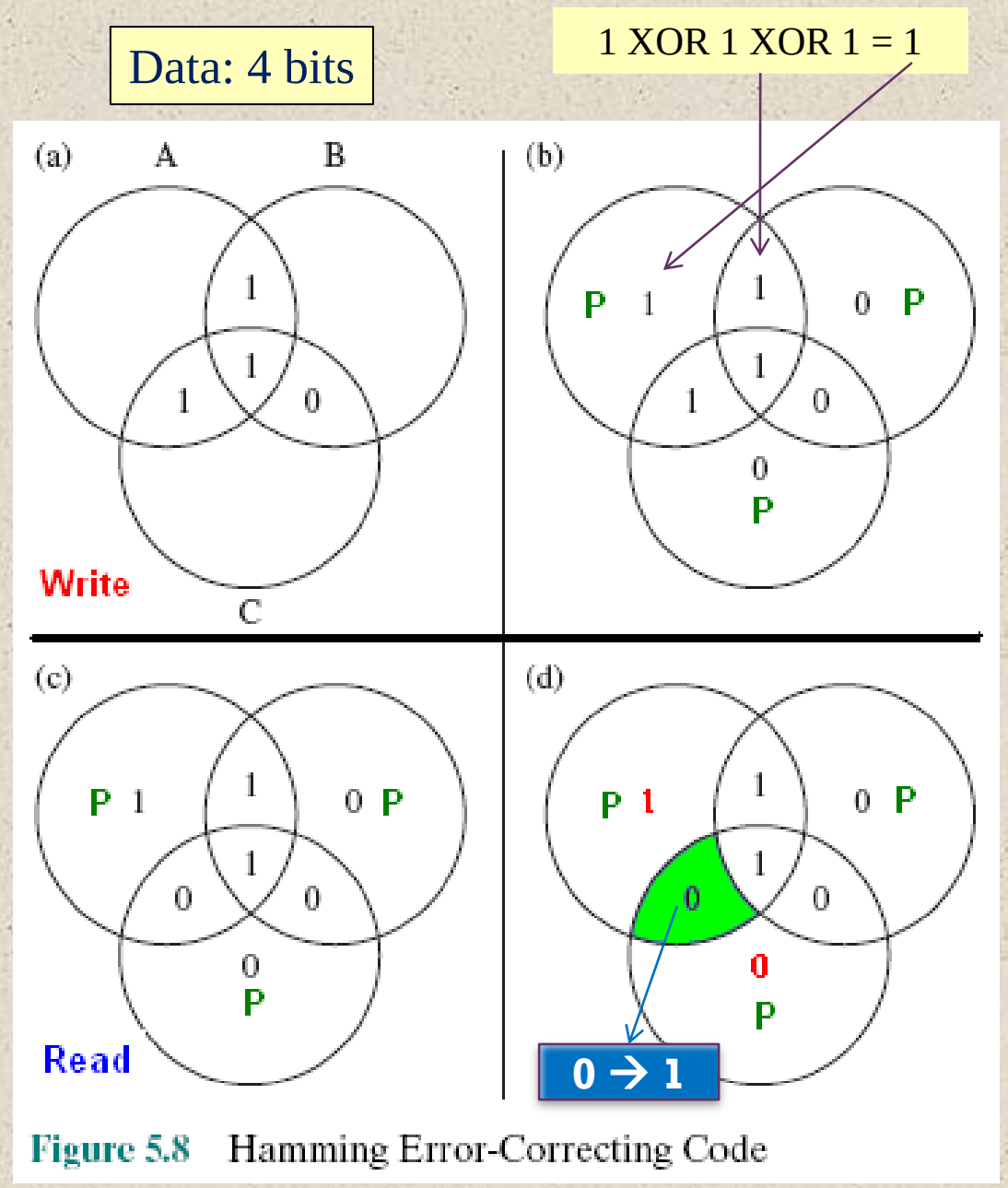


Figure 5.8 Hamming Error-Correcting Code

Increase in Word Length with ECC

Table 5.2 Increase in Word Length with Error Correction

	Single-Error Correction		Single-Error Correction/ Double-Error Detection	
Data Bits	**Check Bits**	**% Increase**	**Check Bits**	**% Increase**
8	4	50	5	62.5
16	5	31.25	6	37.5
32	6	18.75	7	21.875
64	7	10.94	8	12.5
128	8	6.25	9	7.03
256	9	3.52	10	3.91

Data 4 bits (2^2) → At least 3 bit ECC (2+1)

Data 8 bits (2^3) → At least 4 bit ECC (3+1)

Layout of Data Bits and Check Bits

Bit Position	12	11	10	9	8	7	6	5	4	3	2	1
Position Number	1100	1011	1010	1001	1000	0111	0110	0101	0100	0011	0010	0001
Data Bit	D8	D7	D6	D5		D4	D3	D2		D1		
Check Bit					C8				C4		C2	C1

Check positions: 2^3 2^2 2^1 2^0

Figure 5.9 Layout of Data Bits and Check Bits

(Algorithm for computing Ci bit is pre-defined)

Check Bit Calculation

Error

Bit position	12	11	10	9	8	7	6	5	4	3	2	1
Position number	1100	1011	1010	1001	1000	0111	0110	0101	0100	0011	0010	0001
Data bit	D8	D7	D6	D5		D4	D3	D2		D1		
Check bit					C8				C4		C2	C1
Word stored as	0	0	1	1	0	1	0	0	1	1	1	1
Word fetched as	0	0	1	1	0	1	1	0	1	1	1	1
Position Number	1100	1011	1010	1001	1000	0111	0110	0101	0100	0011	0010	0001
Check Bit					0				0		0	1

ECC write: 0111

ECC read: 0001

0111 XOR 0001 != 0 → Error

Figure 5.10 Check Bit Calculation

Hamming SEC-DED Code

Single-**E**rror **C**orrecting/**D**ouble-**E**rror **D**etecting

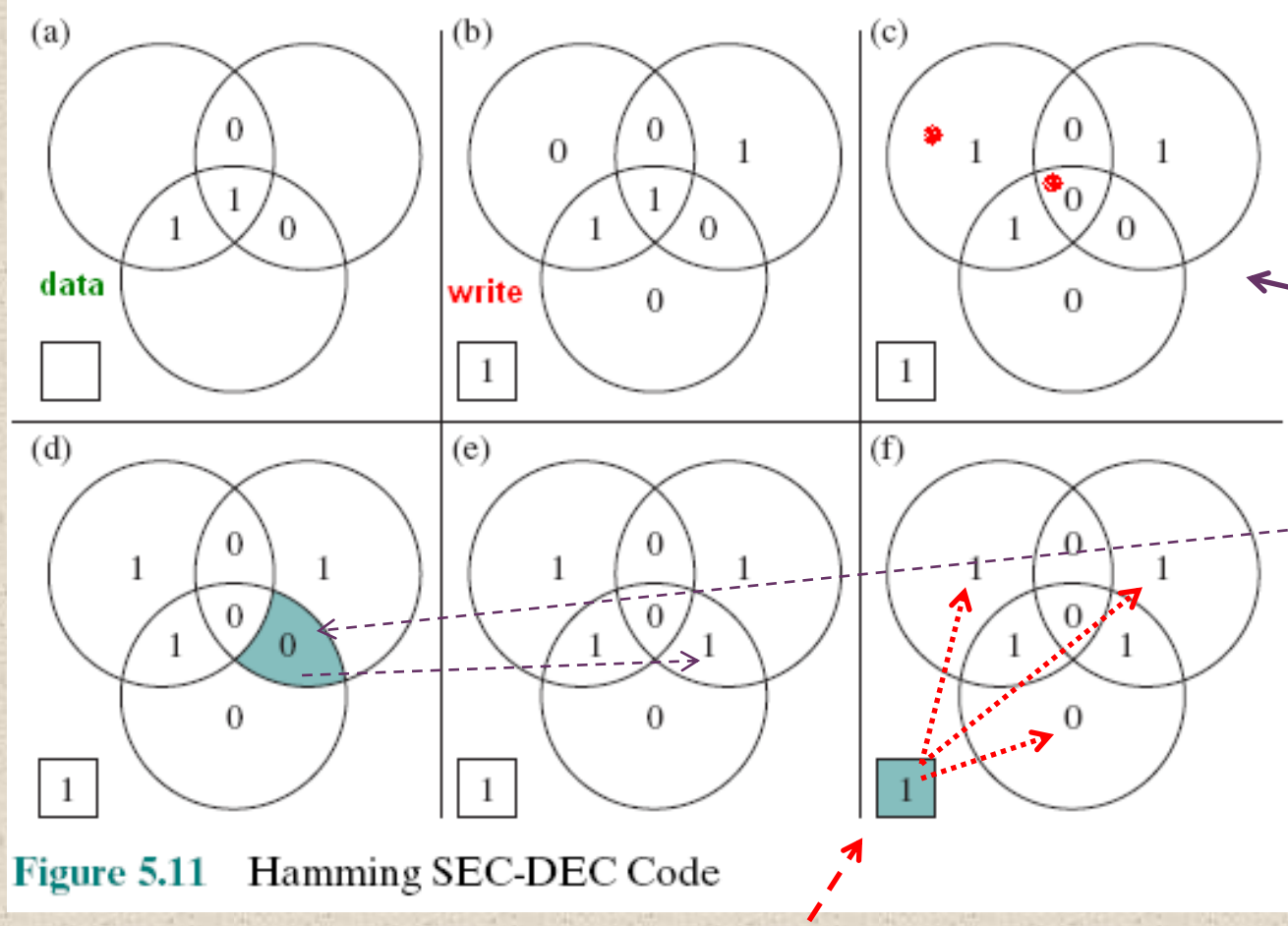


Figure 5.11 Hamming SEC-DEC Code

The sequence shows that if two errors occur (Figure 5.11c), the checking procedure goes astray – chệch hướng (d) and worsens the problem by creating a third error (e).

To overcome the problem, an eighth bit is added that is set so that the total number of 1s in the diagram is even. The extra parity bit catches the error (f).

Performance Comparison

DRAM Alternatives

Table 5.3

	Clock Frequency (MHz)	Transfer Rate (GB/s)	Access Time (ns)	Pin Count
SDRAM	166	1.3	18	168
DDR	200	3.2	12.5	184
RDRAM	600	4.8	12	162

Table 5.3 Performance Comparison of Some DRAM Alternatives

5.3- Advanced DRAM Organization

SDRAM

DDR-DRAM

RDRAM

- One of the most critical system bottlenecks when using high-performance processors is the interface to main internal memory
- The traditional DRAM chip is constrained both by its internal architecture and by its interface to the processor's memory bus
- A number of enhancements to the basic DRAM architecture have been explored:

	Clock Frequency (MHz)	Transfer Rate (GB/s)	Access Time (ns)	Pin Count
SDRAM	166	1.3	18	168
DDR	200	3.2	12.5	184
RDRAM	600	4.8	12	162

Table 5.3 Performance Comparison of Some DRAM Alternatives

Synchronous DRAM (SDRAM)

One of the most widely used forms of DRAM

Exchanges data with the processor synchronized to an external clock signal and running at the full speed of the processor/memory bus without imposing (long while) wait states

With synchronous access the DRAM moves data in and out under control of the system clock

- The processor or other master issues the instruction and address information which is latched by the DRAM
- The DRAM then responds after a set number of clock cycles
- Meanwhile the master can safely do other tasks while the SDRAM is processing

SDRAM

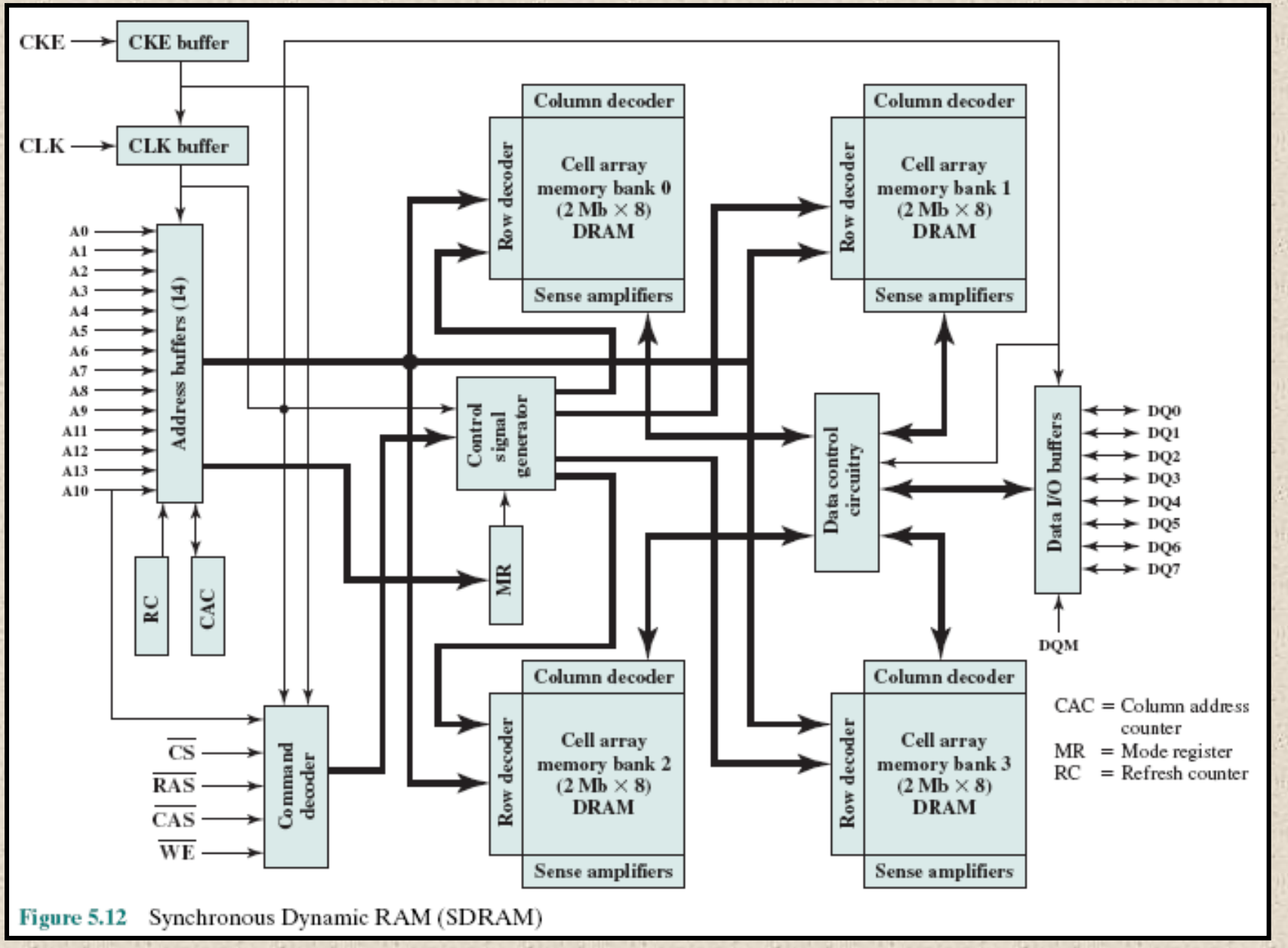


Figure 5.12 Synchronous Dynamic RAM (SDRAM)

SDRAM Pin Assignments

Table 5.4 SDRAM Pin Assignments

A0 to A13	**Address inputs**
CLK	Clock input
CKE	Clock enable
$\overline{CS}$	Chip select
$\overline{RAS}$	Row address strobe
$\overline{CAS}$	Column address strobe
$\overline{WE}$	Write enable
DQ0 to DQ7	Data input/output
DQM	Data mask

SDRAM Read Timing

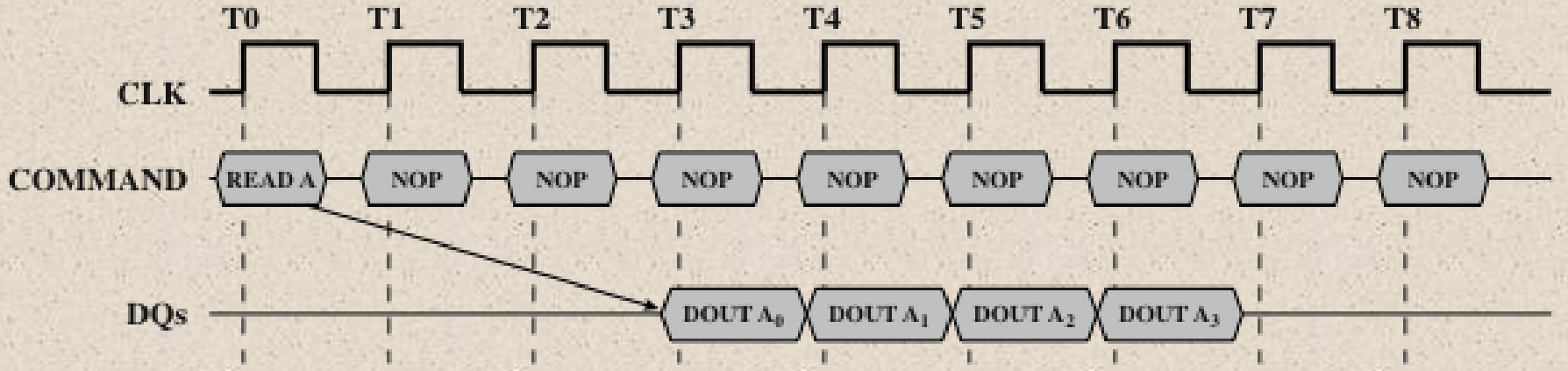


Figure 5.13 SDRAM Read Timing (Burst Length = 4, $\overline{\text{CAS}}$ latency = 2)

RDRAM

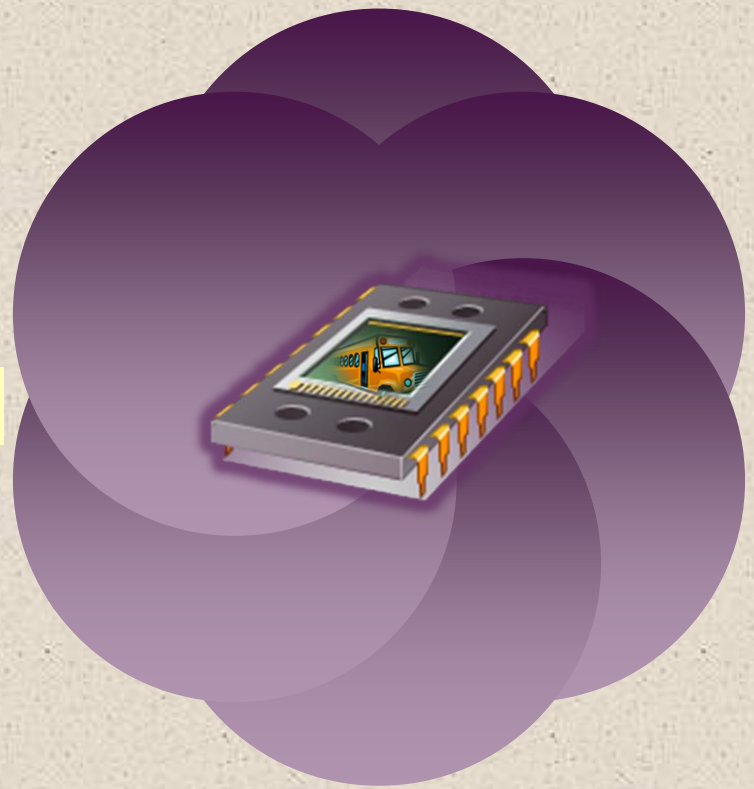
Developed by Rambus

Rambus Dynamic Random Access Memory

Bus delivers address and control information using an asynchronous block-oriented protocol

- Gets a memory request over the high-speed bus
- Request contains the desired address, the type of operation, and the number of bytes in the operation

Protocol: pre-defined rule



Adopted by Intel for its Pentium and Itanium processors

Bus can address up to 320 RDRAM chips and is rated at 1.6 GBps

Has become the main competitor to SDRAM

Chips are vertical packages with all pins on one side

- Exchanges data with the processor over 28 wires no more than 12 centimeters long

RDRAM Structure

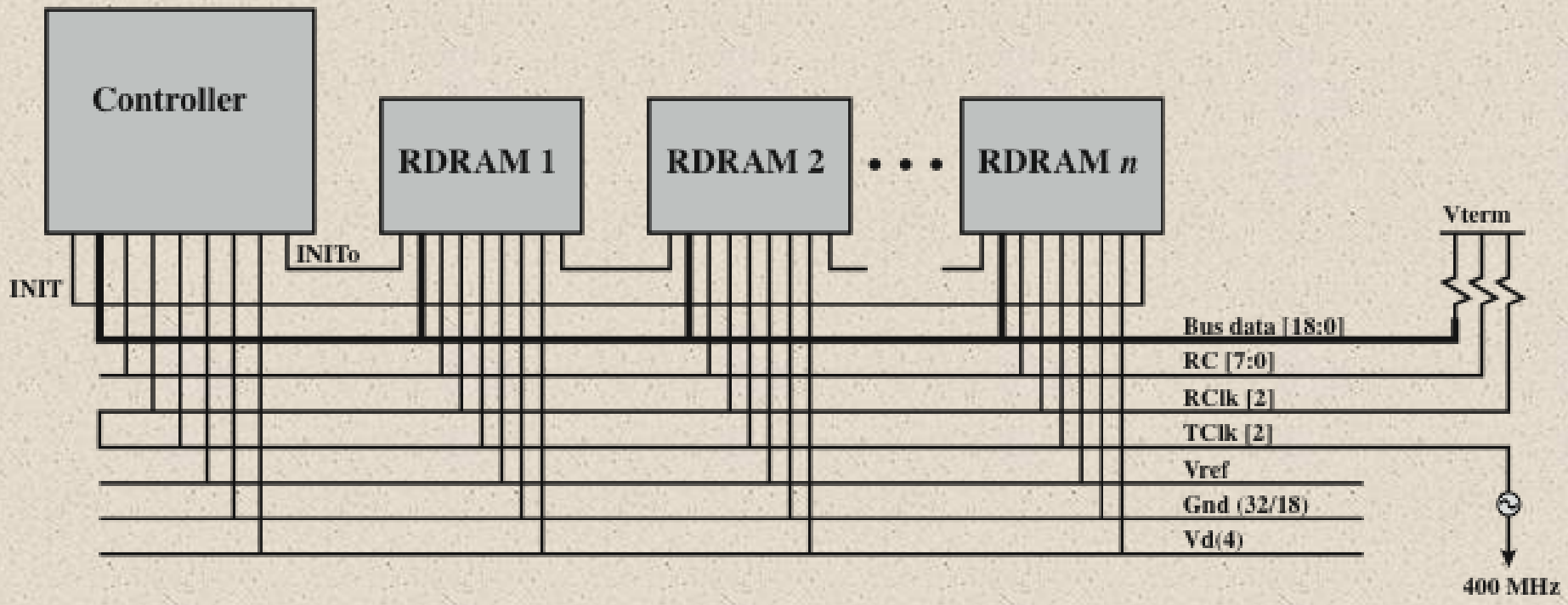
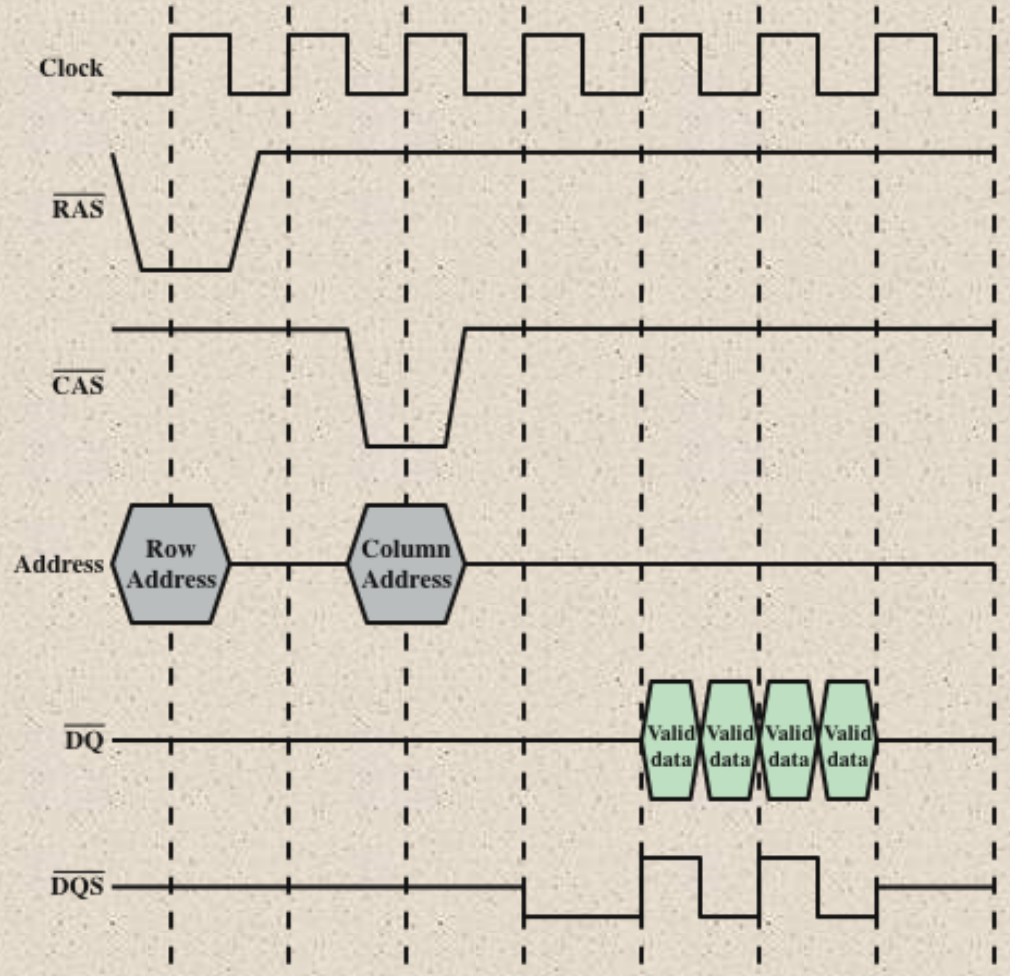


Figure 5.14 RDRAM Structure

Double Data Rate SDRAM (DDR SDRAM)

- SDRAM can only send data once per bus clock cycle
- Double-data-rate SDRAM can send data twice per clock cycle, once on the rising edge of the clock pulse and once on the falling edge
- Developed by the JEDEC Solid State Technology Association (Electronic Industries Alliance's semiconductor-engineering-standardization body)

DDR SDRAM Read Timing



RAS = row address select
CAS = column address select
DQ = data (in or out)
DQS = DQ select

Figure 5.15 DDR SDRAM Read Timing

Cache DRAM (CDRAM)

- Developed by Mitsubishi
- Integrates a small SRAM cache onto a generic DRAM chip
- SRAM on the CDRAM can be used in two ways:
  - It can be used as a true cache consisting of a number of 64-bit lines
    - Cache mode of the CDRAM is effective for ordinary random access to memory
  - Can also be used as a buffer to support the serial access of a block of data

Exercises

- 5.1 What are the key properties of semiconductor memory?
- 5.2 What are two interpretations of the term random-access memory?
- 5.3 What is the difference between DRAM and SRAM in terms of application?
- 5.4 What is the difference between DRAM and SRAM in terms of characteristics such as speed, size, and cost?
- 5.5 Explain why one type of RAM is considered to be analog and the other digital.
- 5.6 What are some applications for ROM?
- 5.7 What are the differences among EPROM, EEPROM, and flash memory?
- 5.8 Explain the function of each pin in Figure 5.4b. 182 CHAPTER 5 / INTERNAL MEMORY
- 5.9 What is a parity bit?
- 5.10 How is the syndrome for the Hamming code interpreted?
- 5.11 How does SDRAM differ from ordinary DRAM?

Summary

Chapter 5

Internal Memory

- Semiconductor main memory
  - Organization
  - DRAM and SRAM
  - Types of ROM
  - Chip logic
  - Chip packaging
  - Module organization
  - Interleaved memory
- Error correction
  - Hard failure
  - Soft error
- Hamming code
- Advanced DRAM organization
  - Synchronous DRAM
  - Rambus DRAM
  - DDR SDRAM
  - Cache DRAM