AAiT

Computer Architecture & Organization

Chapter 12

Internal Memory

AAiT

Memory Cell Operation

- In earlier computers, the most common form of random-access storage for computer main memory employed an array of doughnut-shaped ferromagnetic loops referred to as *cores.* Hence, main memory was often referred to as core.

- Today, the use of semiconductor chips for main memory is almost universal
- The basic element of a **semiconductor memory** is the memory cell

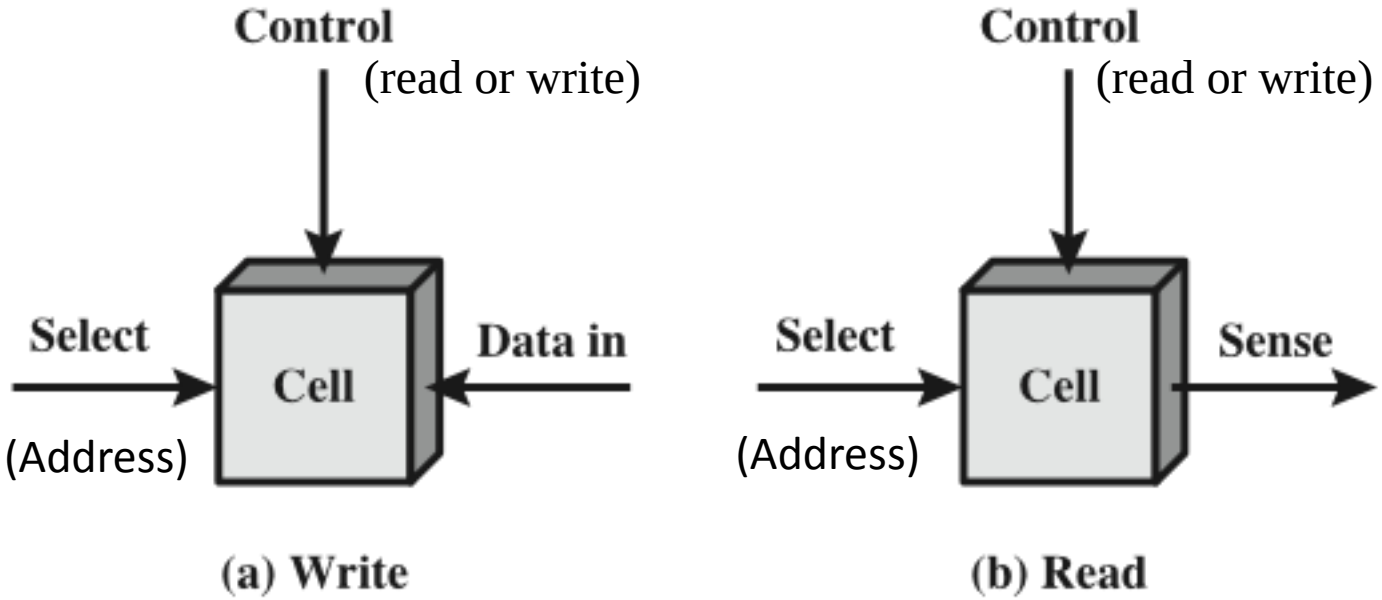


Figure 5.1 Memory Cell Operation

Semiconductor Memory Types

RAM:

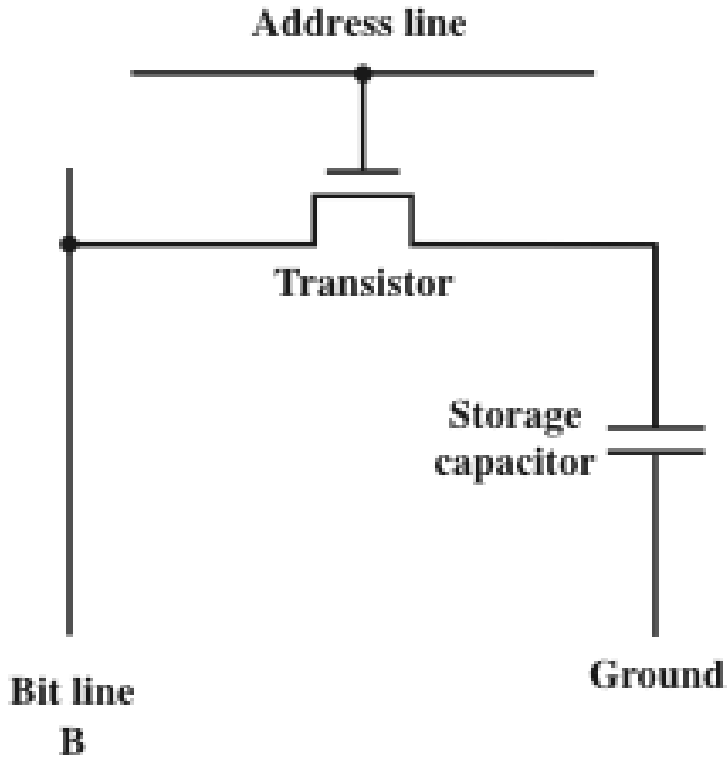
- both read and write is easy using electrical signals.
- Actually ROMs listed in this table are also random-access, misuse of term.

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)				
Erasable PROM (EPROM)		UV light, chip-level	Electrically	
Electrically Erasable PROM (EEPROM)	Read-mostly memory	Electrically, byte-level		
Flash memory		Electrically, block-level		

Table 5.1 Semiconductor Memory Types

Dynamic RAM (DRAM)

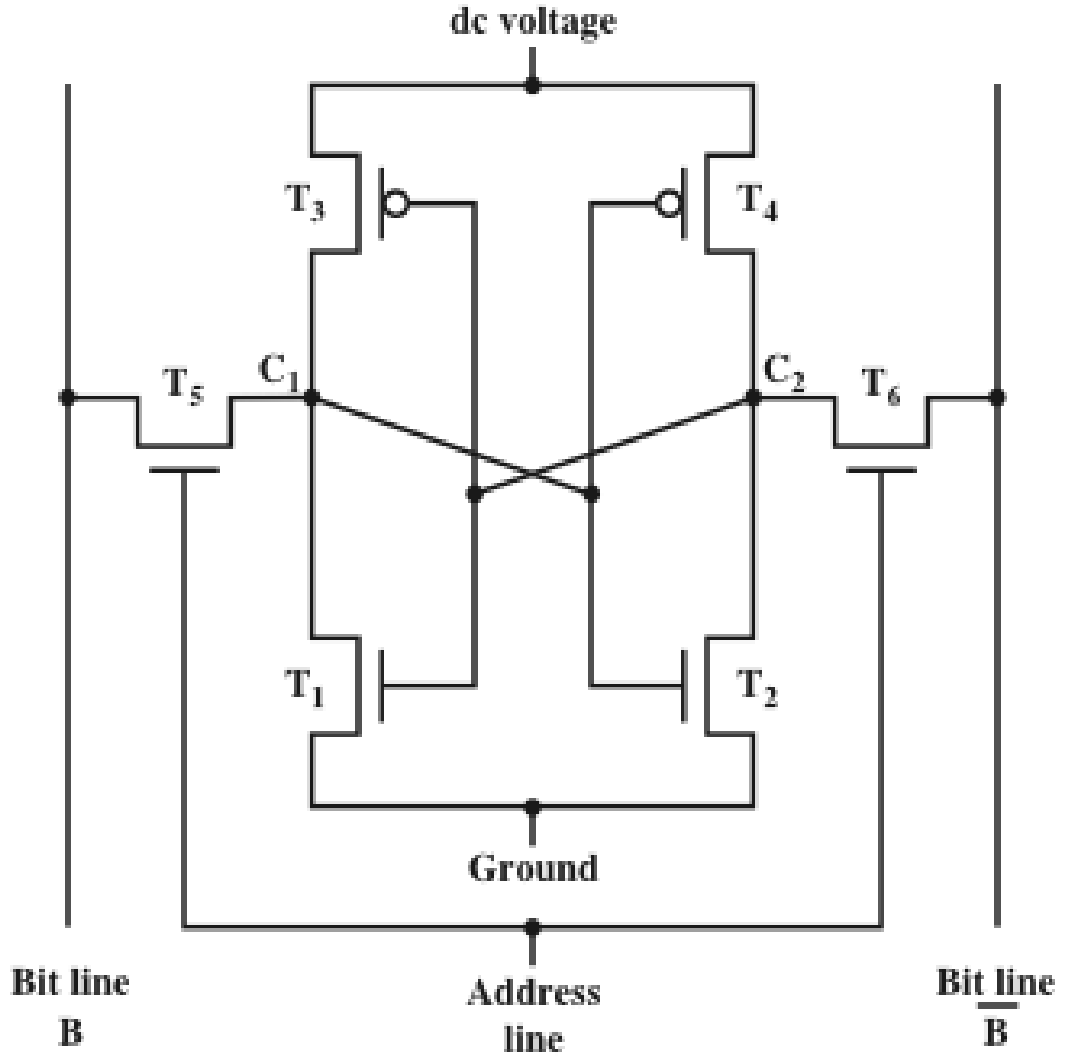
- RAM technology is divided into two technologies:
  - Dynamic RAM (DRAM)
  - Static RAM (SRAM)
- DRAM
  - Made with cells that store data as charge on capacitors
  - Presence or absence of charge in a capacitor is interpreted as a binary 1 or 0
  - Requires periodic charge refreshing to maintain data storage
  - The term *dynamic* refers to tendency of the stored charge to leak away, even with power continuously applied



(a) Dynamic RAM (DRAM) cell

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
- Unlike the DRAM, no refresh is needed to retain data
- For a write operation, the desired bit value is applied to line B, while its complement is applied to line B'.
- For a read operation, the bit value is read from line B



(b) Static RAM (SRAM) cell

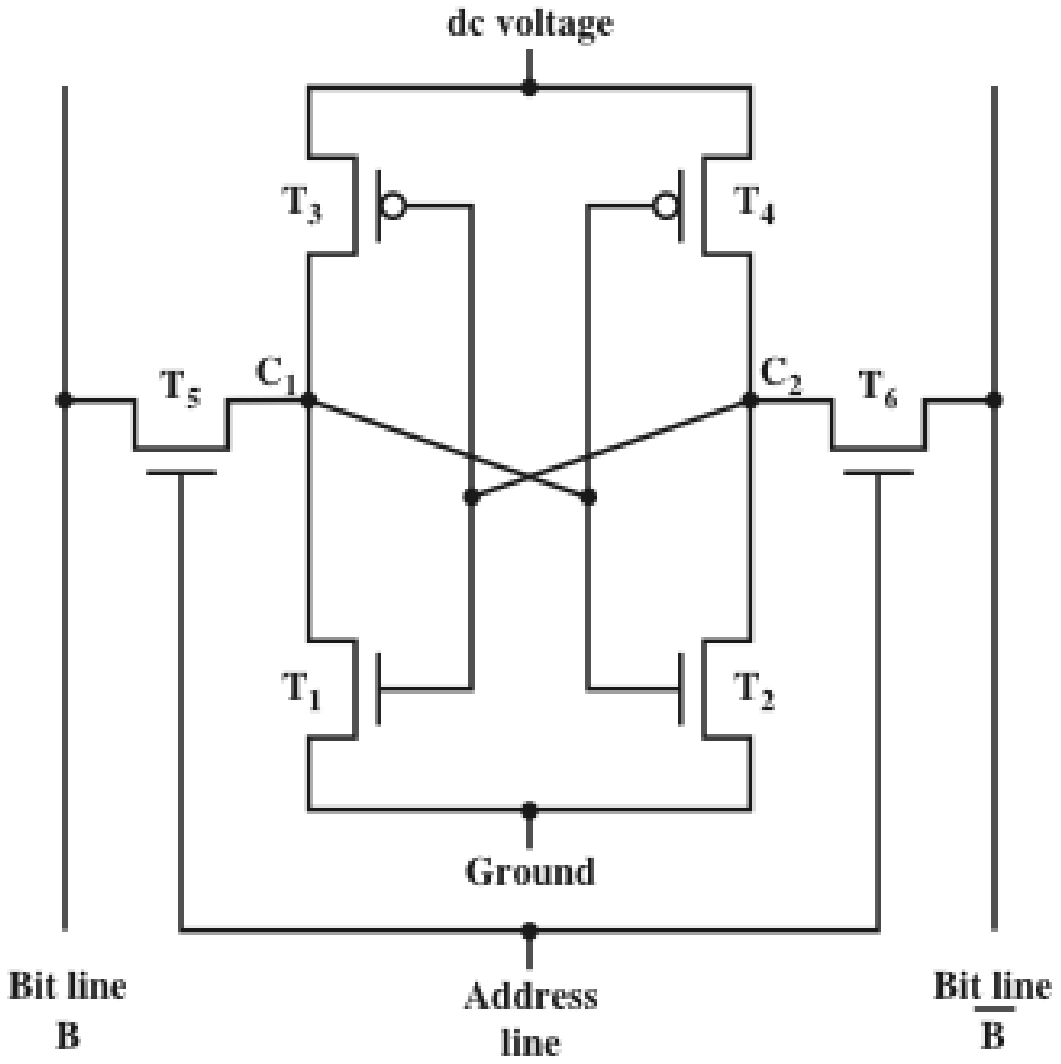
Figure 5.2b

Typical Memory Cell Structures

Exercise

- What is the state of each transistor, Bit lines, points C1 & C2
  a) Near the end of writing 1 to the SRAM cell?
  b) During reading a zero bit from the SRAM cell

	T1	T2	T3	T4	T5	T6	B	B’	C1	C2
a	off	on	on	off	on	on	1	0	1	0
b	on	off	off	on	on	on	0	1	0	1



SRAM versus DRAM

- Both volatile
  - Power must be continuously supplied to the memory to preserve the bit values
- Dynamic cell
  - Simpler to build, smaller
  - More dense (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)

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
- Must be erased before writing
- Erasure may take 20 minutes
- Erasure uses UV light
- 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)

- row address select (RAS)
- column address select (CAS)
- the write enable (WE)
- output enable (OE)

Refreshing:

- Disable the DRAM chip while all data cells are refreshed
- For each row, the output lines from the refresh counter are supplied to the row decoder and the RAS line is activated.
- The data are read out and written back into the same location.
- This causes each cell in the row to be refreshed.

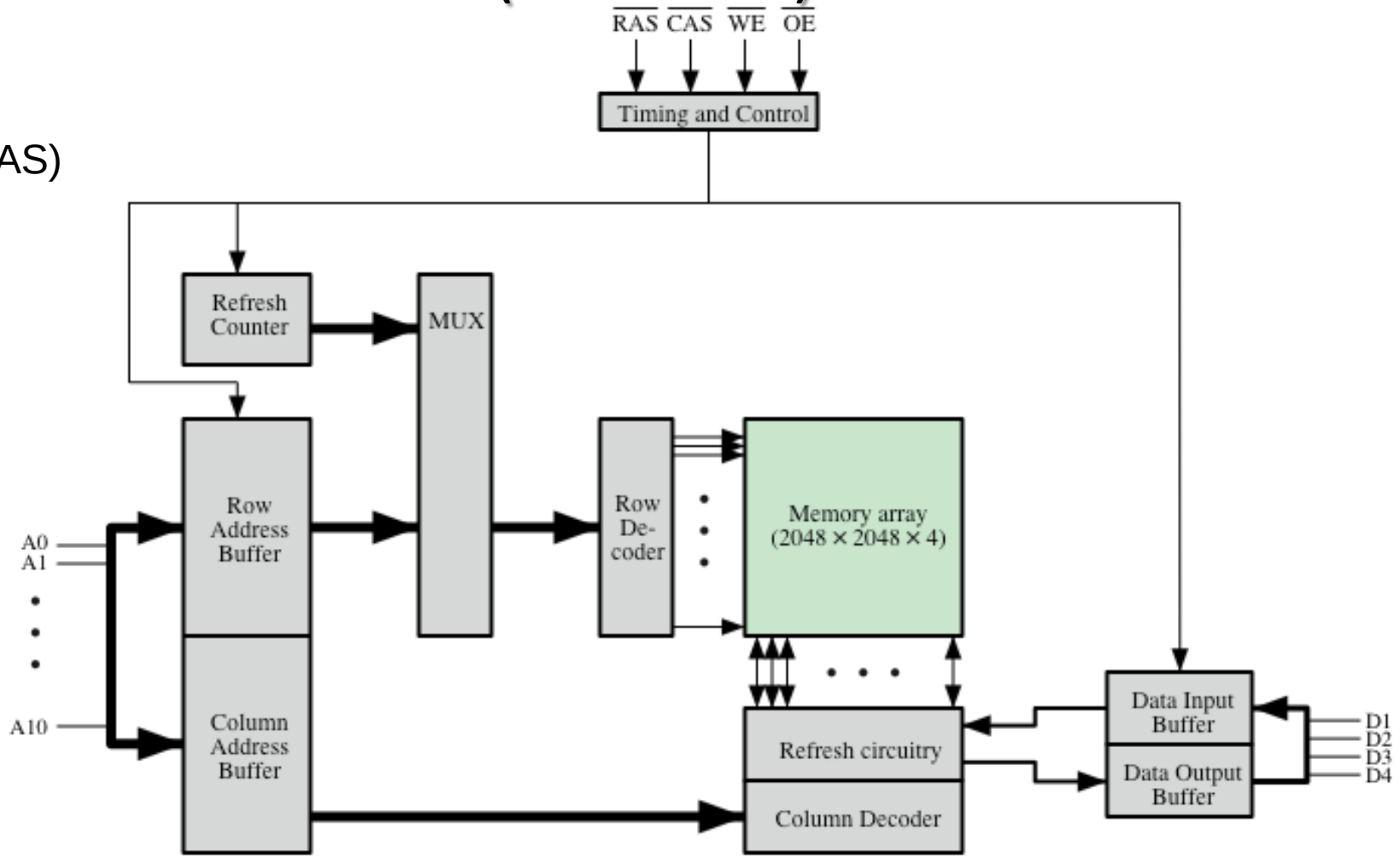
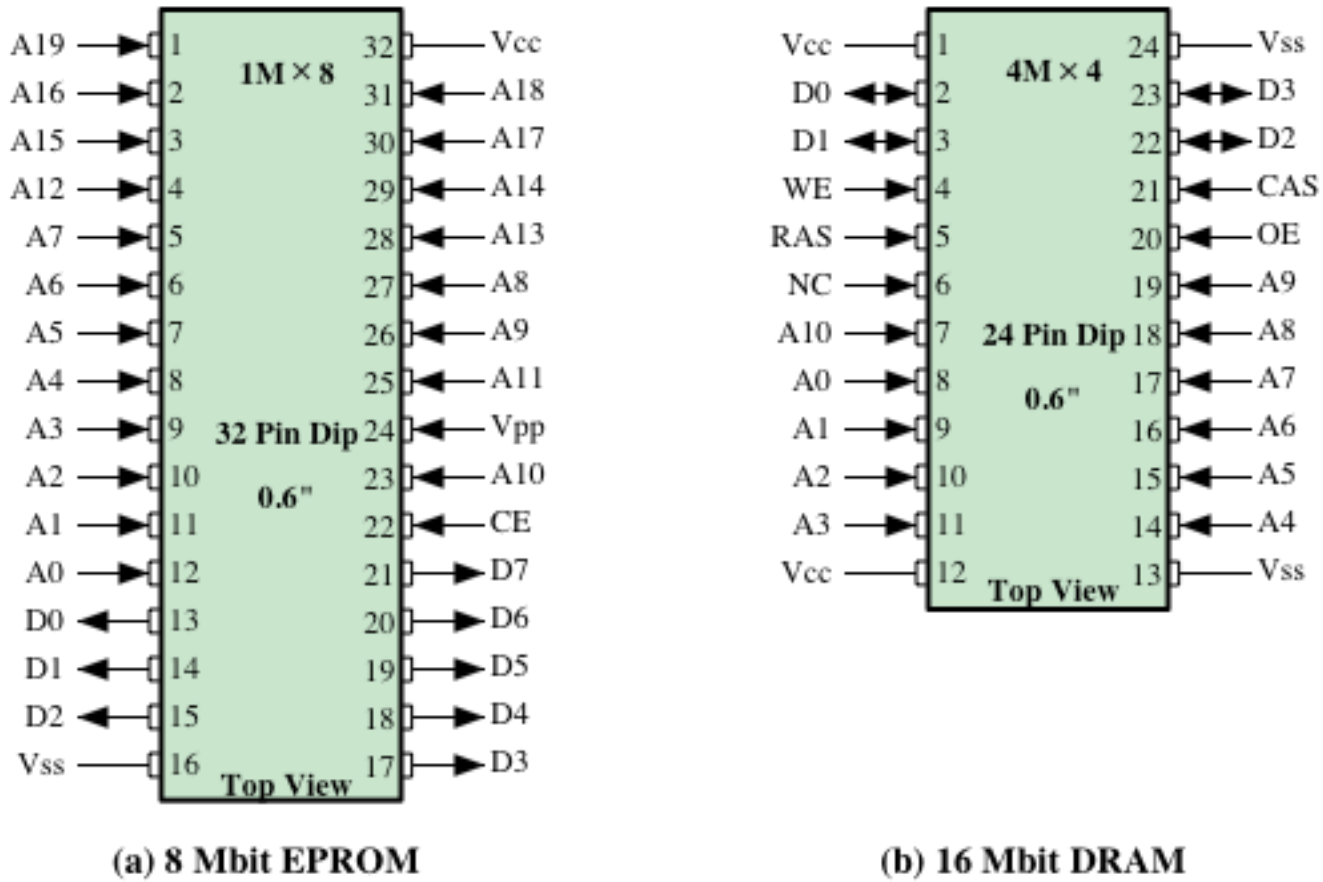


Figure 5.3 Typical 16 Megabit DRAM (4M × 4)

Chip Packaging



(a) 8 Mbit EPROM

(b) 16 Mbit DRAM

Figure 5.4 Typical Memory Package Pins and Signals

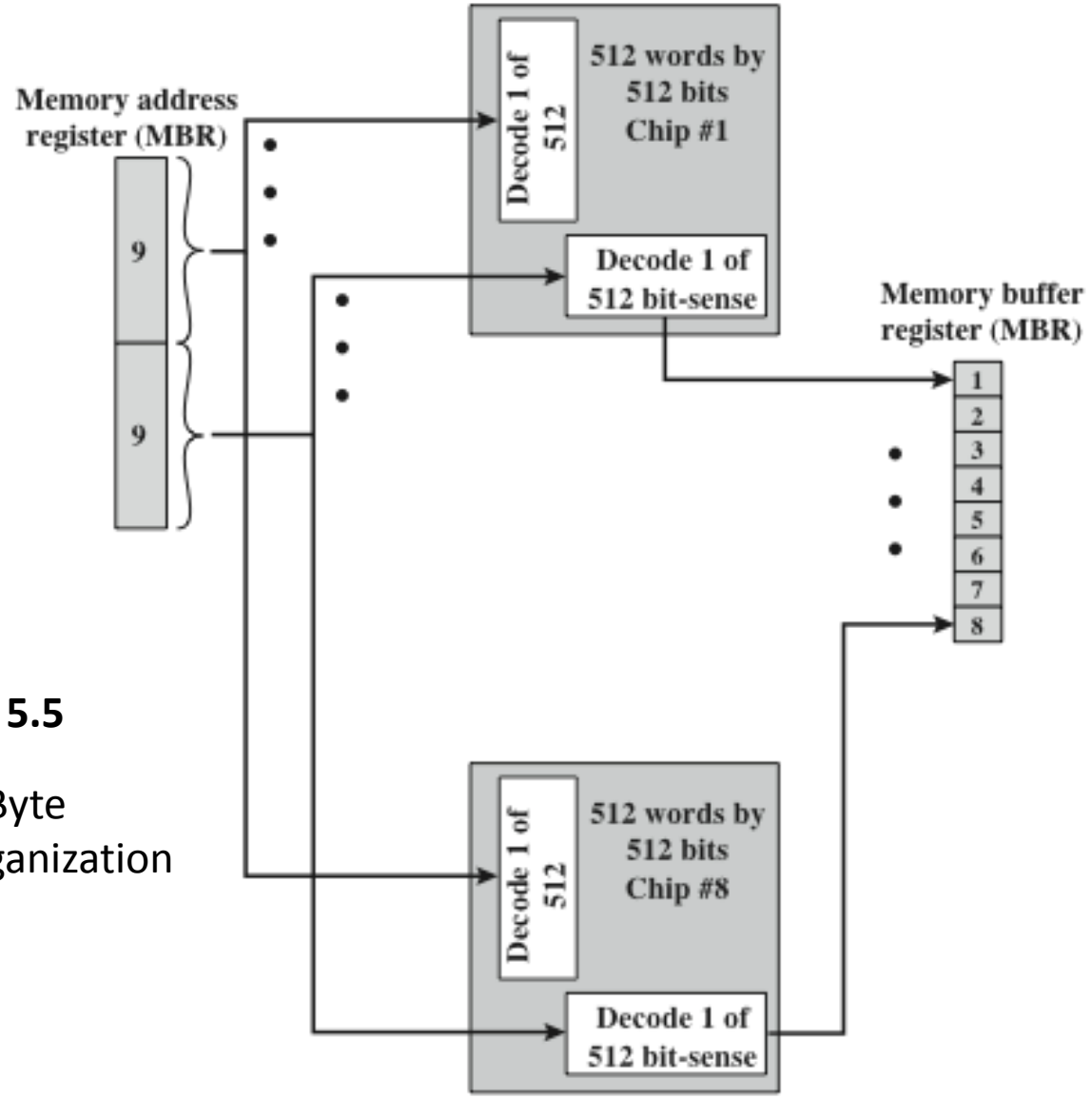


Figure 5.5

256-KByte Memory Organization

Figure 5.5 256-KByte Memory Organization

- 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 (8 in this example)

- In the case in which larger memory is required say 1M word by 8 bits per word, we have four columns of such chips, each column containing 256K words arranged as in Figure 5.5.
- For 1M word, 20 address lines are needed.
- The 18 least significant bits are routed to all 32 modules.
- The high-order 2 bits are input to a group select logic module that sends a chip enable signal to one of the four columns of modules.

Error Correction

- Hard Failure
  - Permanent physical defect
  - 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 environmental abuse
    - Manufacturing defects
    - Wear
- 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

Error Correcting Code Function

- If an *M*-bit word of data is to be stored and the code is of length *K* bits, then the actual size of the stored word is *M* + *K bits.*
- When the previously stored word is read out, the code is used to detect and possibly correct errors.
- If correctable error is detected, The data bits plus **error correction** bits are fed into a corrector.

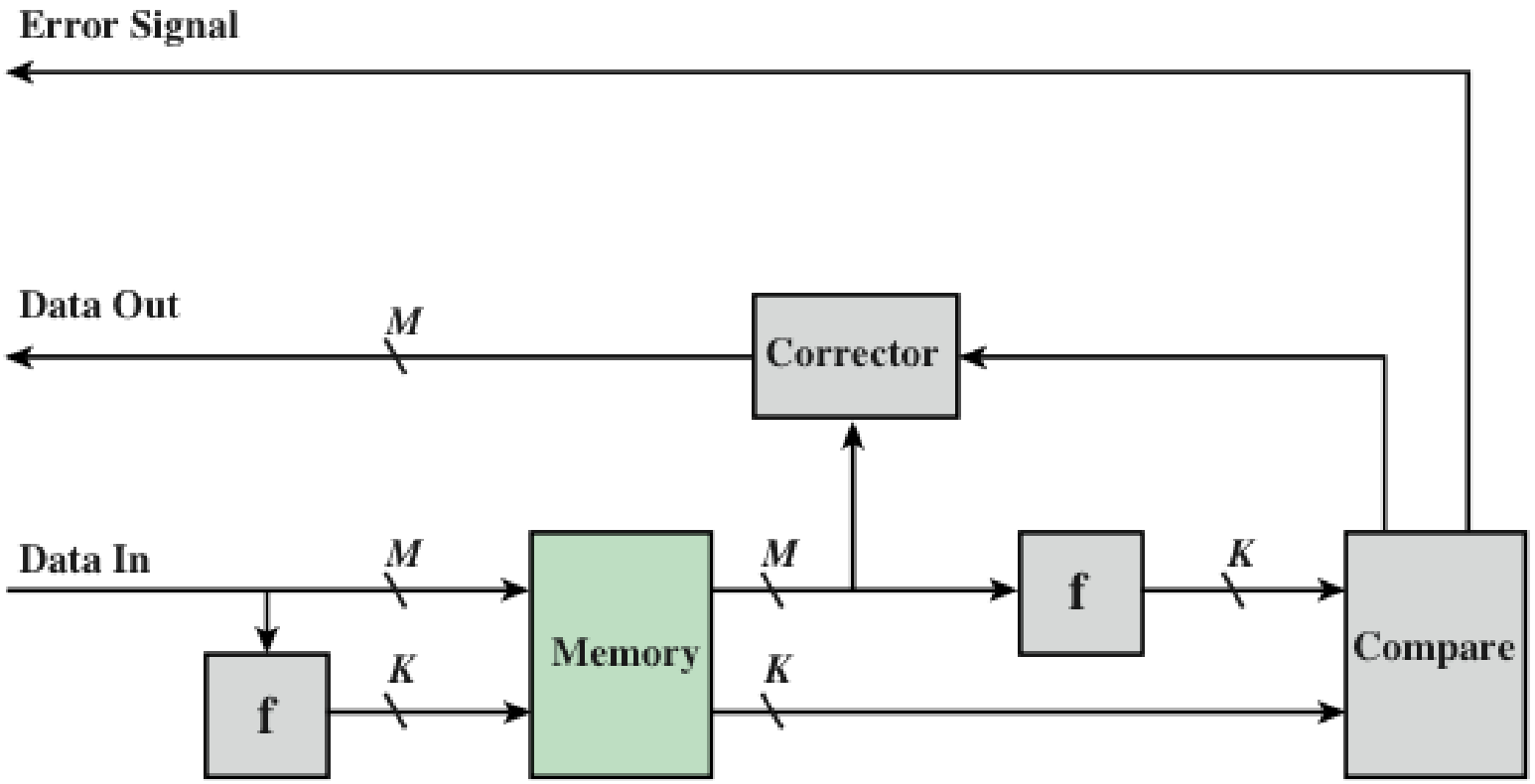


Figure 5.7 Error-Correcting Code Function

Hamming Error Correcting Code

- Assuming M=4, We assign the 4 data bits to the inner compartments
- (Figure5.8a). The remaining compartments are filled with what are called *parity bits.*
- Each parity bit is chosen so that the total number of 1s in its circle is even
- Thus, because circle A includes three data 1s, the parity bit in that circle is set to 1.
- Now, if an error changes one of the data bits (Figure 5.8c), it is easily found.
- Discrepancies are found in circle A and circle C but not in circle B. Only one of the seven compartments is in A and C but not B..

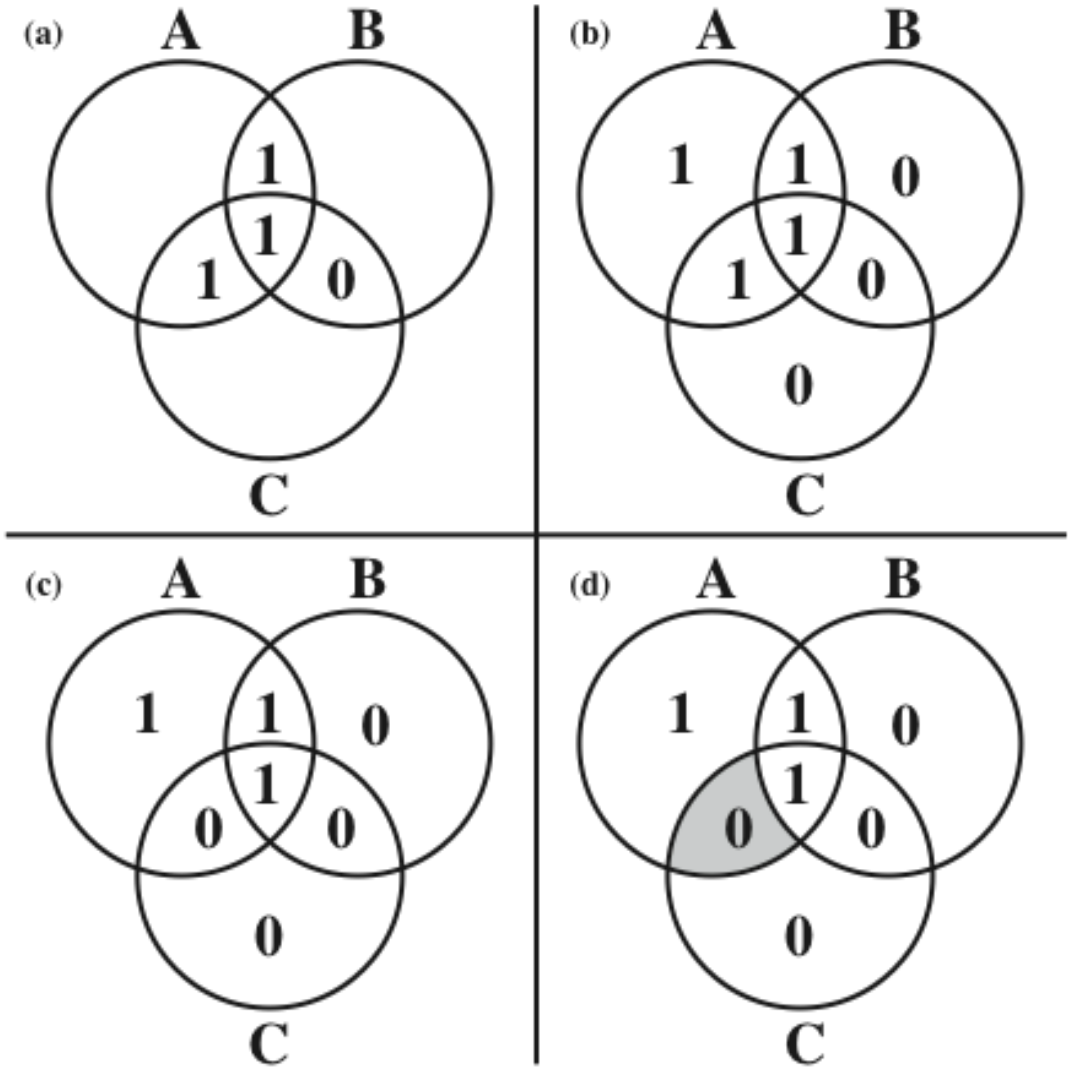


Figure 5.8 Hamming Error-Correcting Code

How long the code must be

- In Figure 5.7, the comparison logic receives as input two *K*-bit values.
- A bit-by-bit comparison is done by taking the exclusive-OR of the two inputs. The result is called the ***syndrome*** *word.*
- The syndrome word is therefore *K* bits wide and has a range between 0 and $2^K - 1$.
- The value 0 indicates that no error was detected, leaving $2^K - 1$ values to indicate, if there is an error, which bit was in error.
- Now, because an error could occur on any of the *M* data bits or *K* check bits, we must have $2^K - 1 \geq M + K$
- This inequality gives the number of bits needed to correct a single bit error in a word containing M data bits.

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

Layout of Data Bits and Check Bits

- For convenience, we would like to generate a 4-bit syndrome for an 8-bit data (4+8=12) word with the following characteristics:
  - If the syndrome contains all 0s, no error has been detected.
  - If the syndrome contains one and only one bit set to 1, then an error has occurred in one of the 4 check bits. No correction is needed.
  - If the syndrome contains more than one bit set to 1, then the numerical value of the syndrome indicates the position of the data bit in error. This data bit is inverted for correction.
- To achieve these characteristics, the data and check bits are arranged into a 12-bit word as depicted in Figure 5.9
- Those bit positions whose position numbers are powers of 2 are designated as 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

Figure 5.9 Layout of Data Bits and Check Bits

Check Bit Calculation

- Each check bit operates on every data bit whose position number contains a 1 in the same bit position as the position number of that check bit.

C1 has 1's at first bit position
➔ acts on D1, D2, D4, D5, D7
➔C1= $D1 \oplus D2 \oplus D4 \oplus D5 \oplus D7$
= $1 \oplus 0 \oplus 1 \oplus 1 \oplus 0 = 1$
…. Do same for C2, C3 & C4
Gives the hamming code C4C3C2C1 = 0111

Alternatively XOR of position numbers for which the data bits equals1 gives the hamming code
$\equiv 0011 \oplus 0111 \oplus 1001 \oplus 1010 = 0111$

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

Figure 5.10 Check Bit Calculation

Exercise

- Suppose now that data bit 3, in bit position 6, sustains an error and is changed from 0 to 1. How can you correct the error?

Ans.

- The resulting block is 001101101111, with a Hamming code of 0111.
- An XOR of the Hamming code and all of the bit position values for nonzero data bits results in 0110
- The nonzero result detects an error and indicates that the error is in bit position 6.

Advanced DRAM Organization

- 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:

DDR-DRAM

	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 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

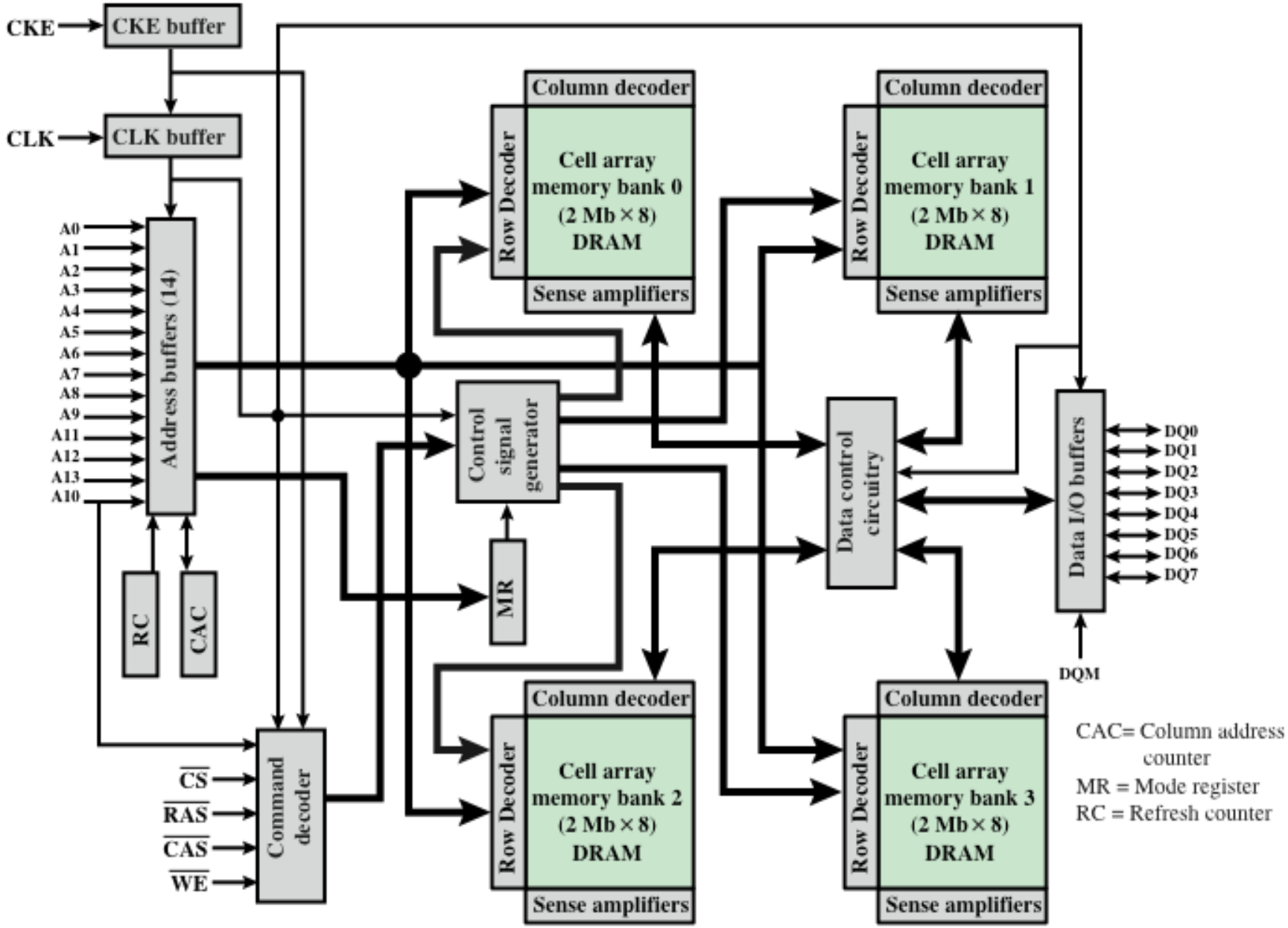
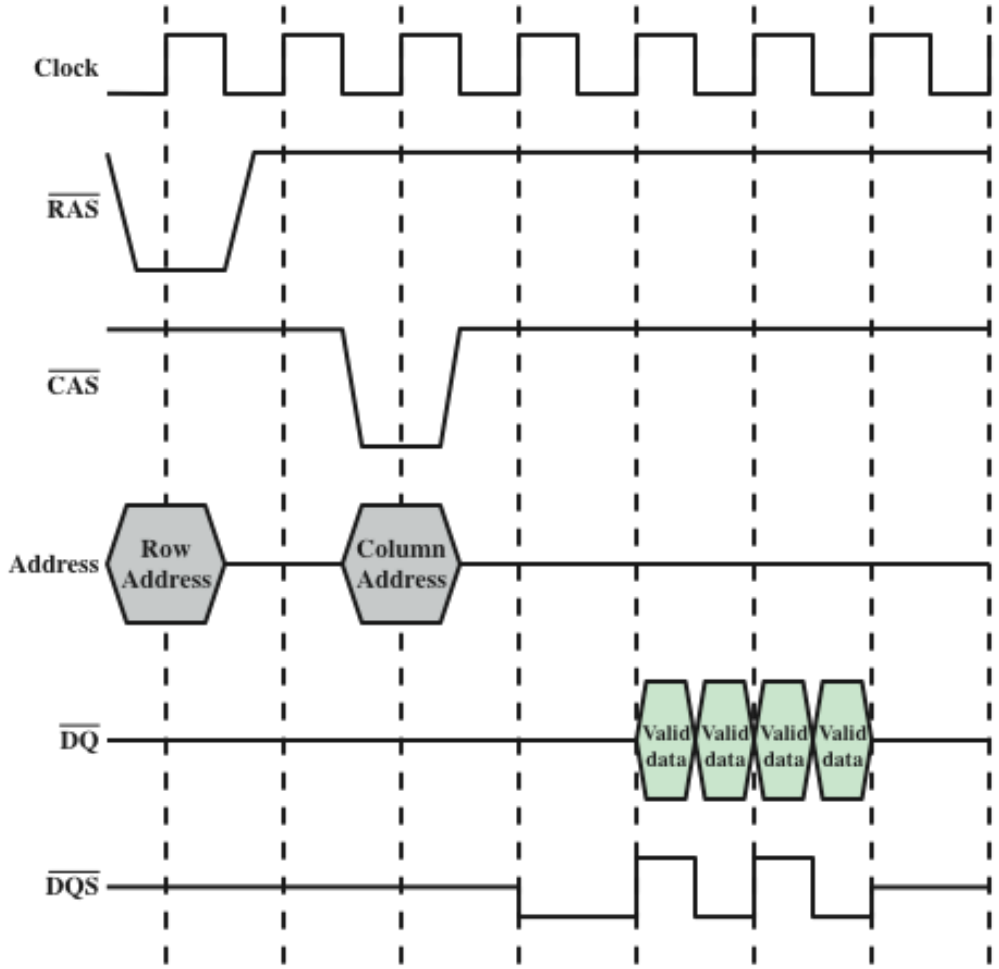


Figure 5.12 Synchronous Dynamic RAM (SDRAM)

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

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
  - DDR SDRAM
  - Cache DRAM