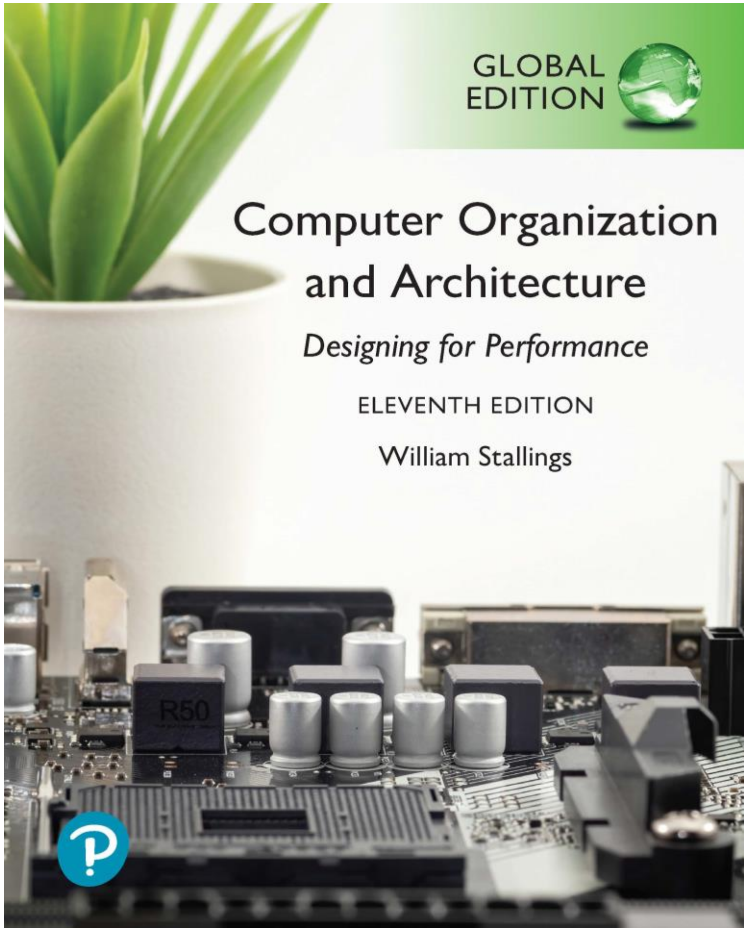
Computer Organization and Architecture
Designing for Performance

11th Edition, Global Edition



Chapter 6

Internal Memory

Figure 6.1
Memory Cell Operation

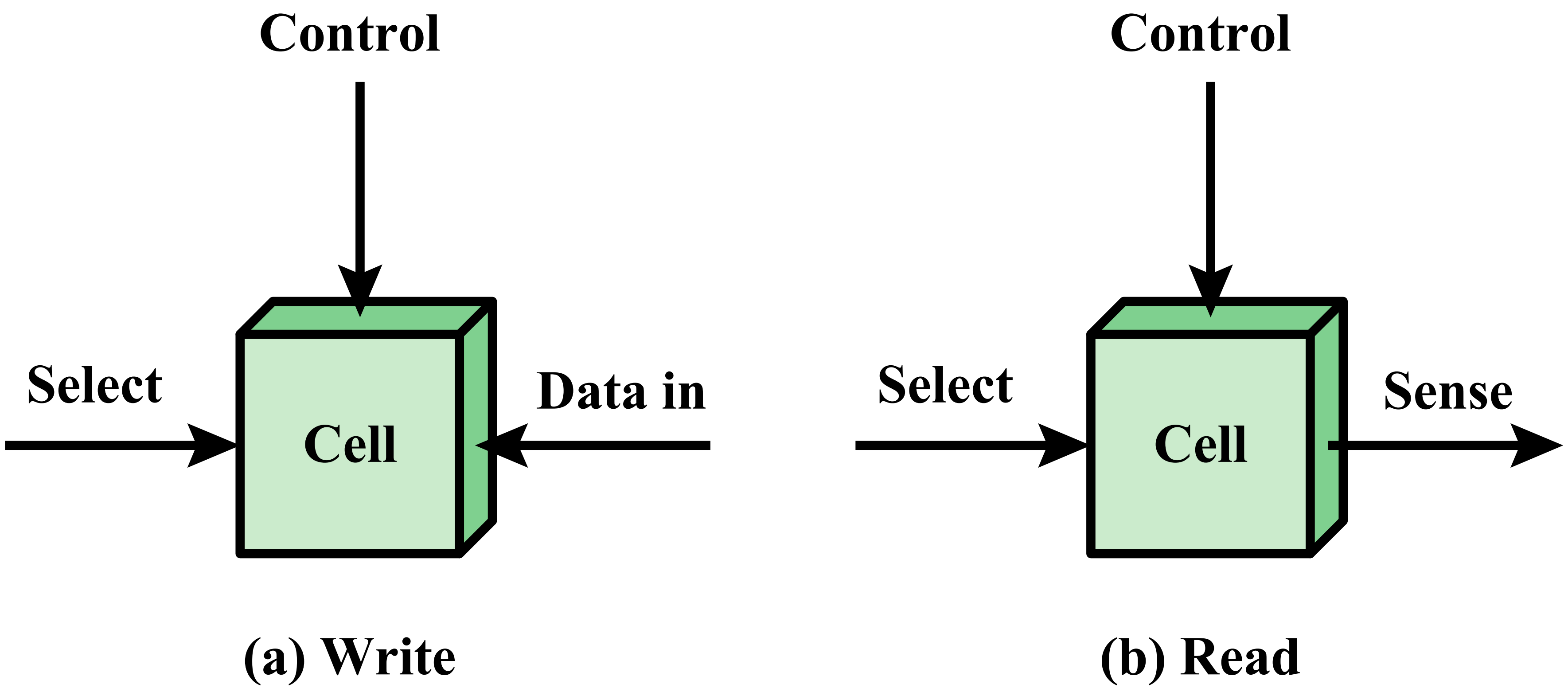


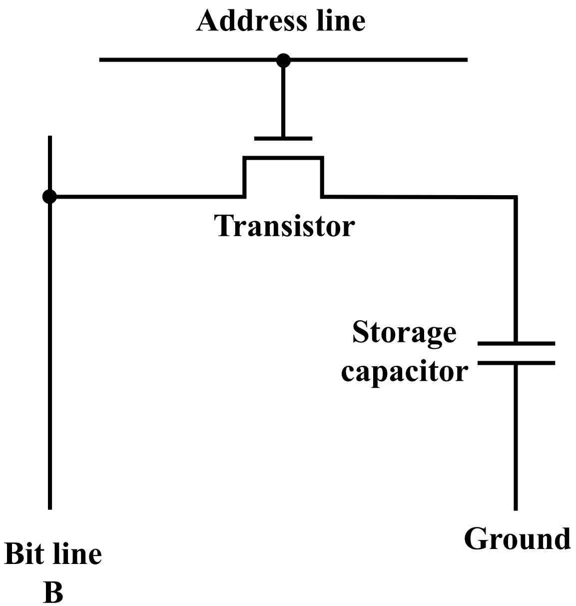
Table 6.1
Semiconductor Memory Types

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		

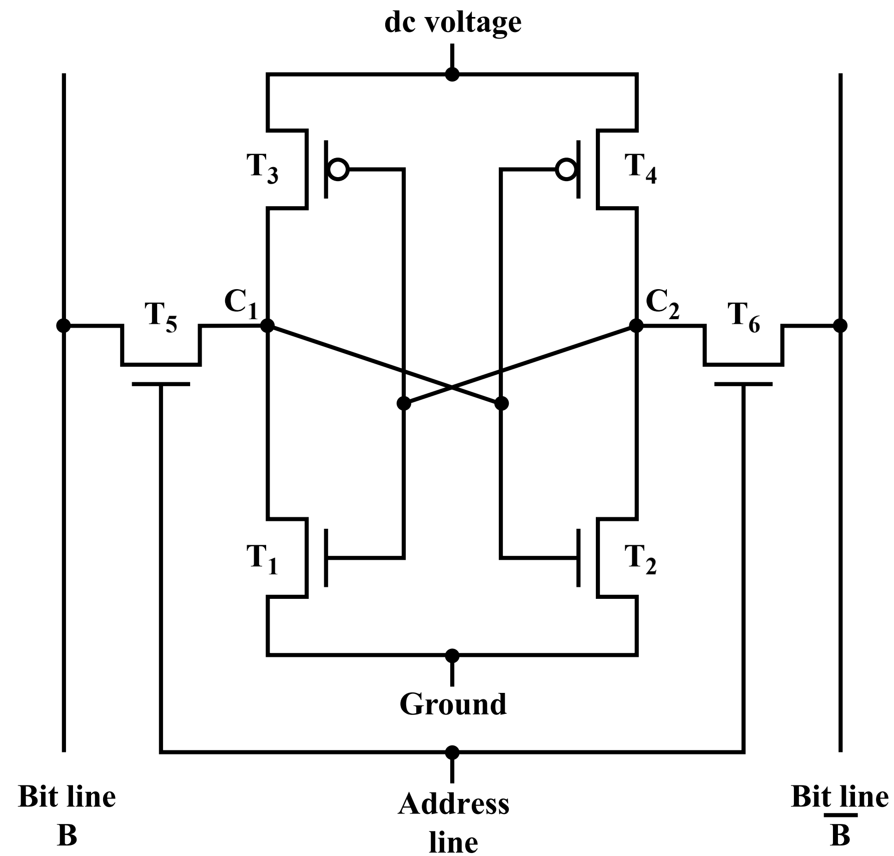
Dynamic RAM (DRAM)

- 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

Figure 6.2
Typical Memory Cell Structures



(a) Dynamic RAM (DRAM) cell



(b) Static RAM (SRAM) 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

SRAM versus DRAM

SRAM

- 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

DRAM

- 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	EEPROM	Flash Memory
Erasable programmable read-only memory	Electrically erasable programmable read-only memory	Intermediate between EPROM and EEPROM in both cost and functionality
Erasure process can be performed repeatedly	Can be written into at any time without erasing prior contents	Uses an electrical erasing technology, does not provide byte-level erasure
More expensive than PROM but it has the advantage of the multiple update capability	Combines the advantage of non-volatility with the flexibility of being updatable in place	Microchip is organized so that a section of memory cells are erased in a single action or "flash"
	More expensive than EPROM	

Figure 6.3
Typical 16-Mbit DRAM (4M × 4)

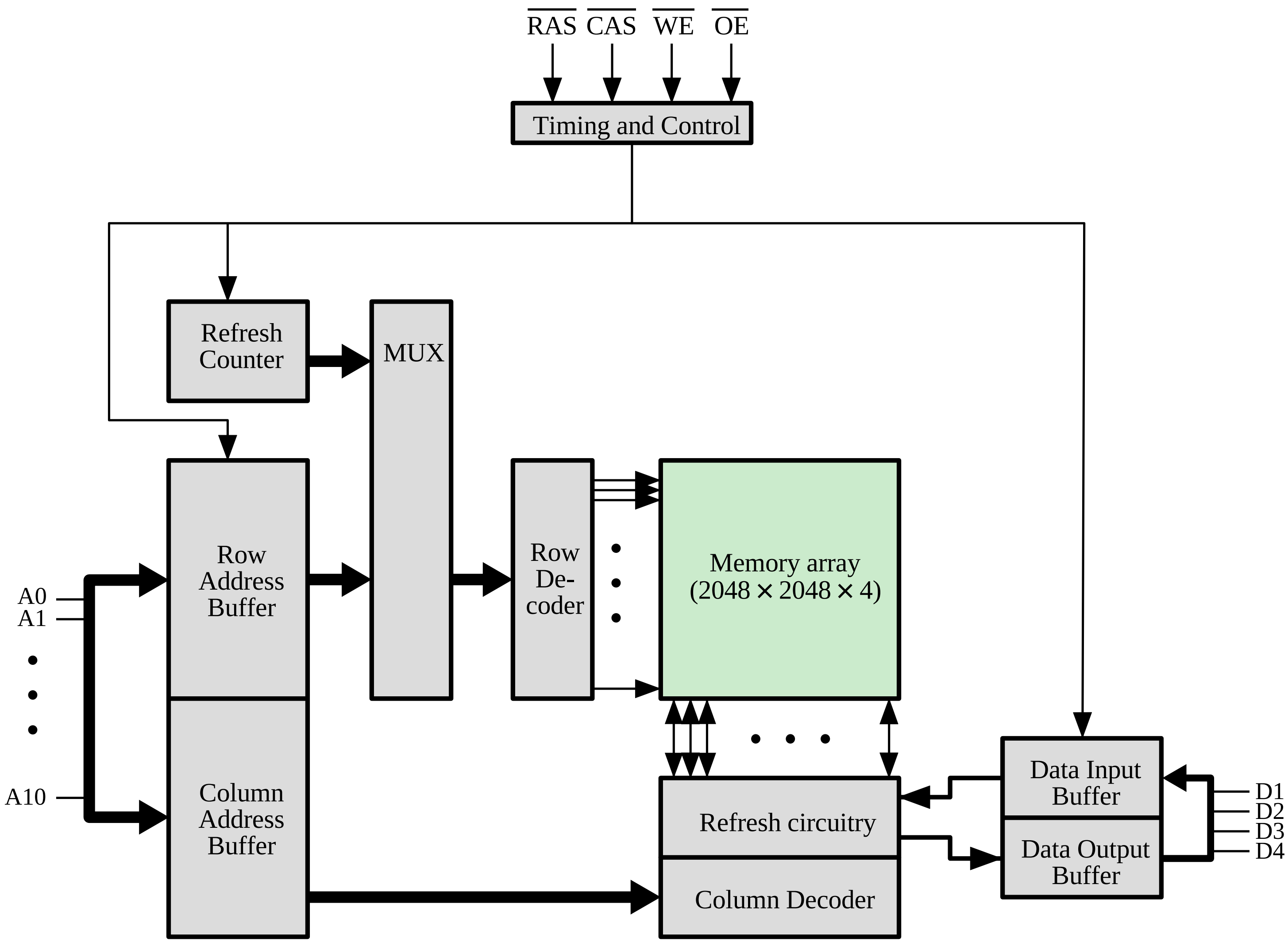
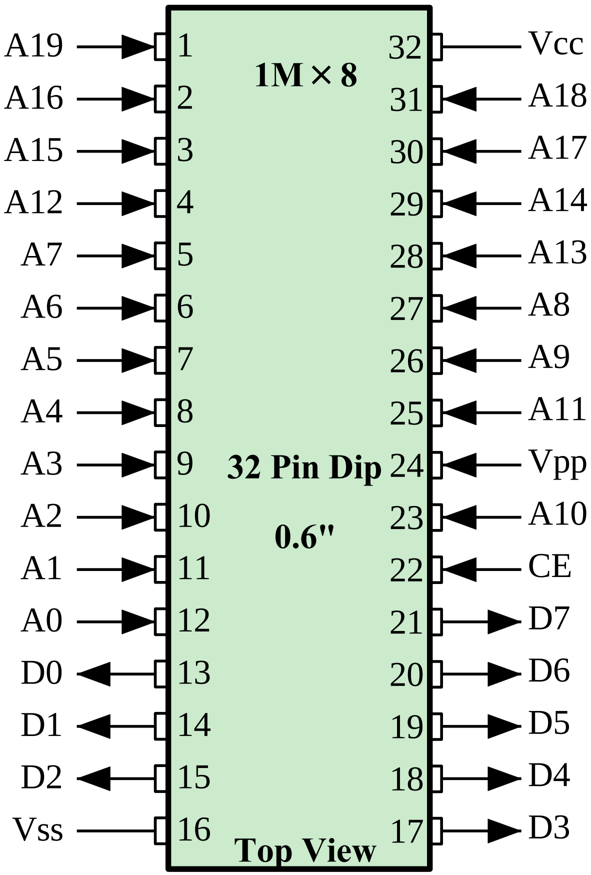
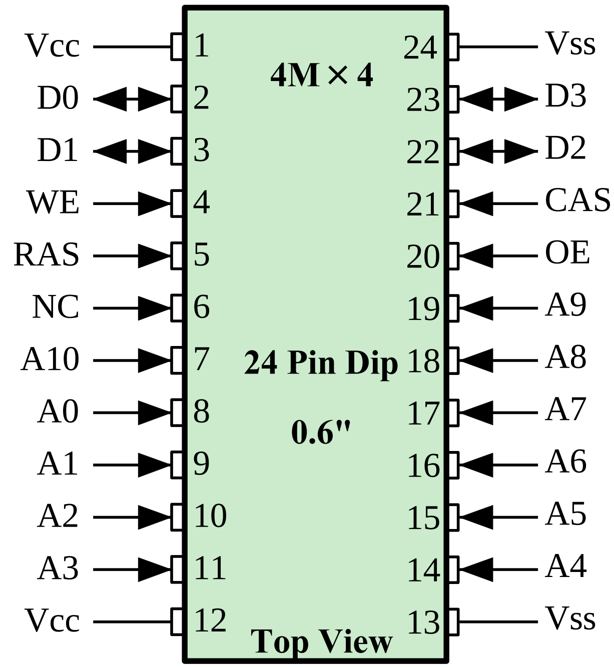


Figure 6.4
Typical Memory Package Pins and Signals



(a) 8 Mbit EPROM



(b) 16 Mbit DRAM

Figure 6.5
256-KByte Memory Organization

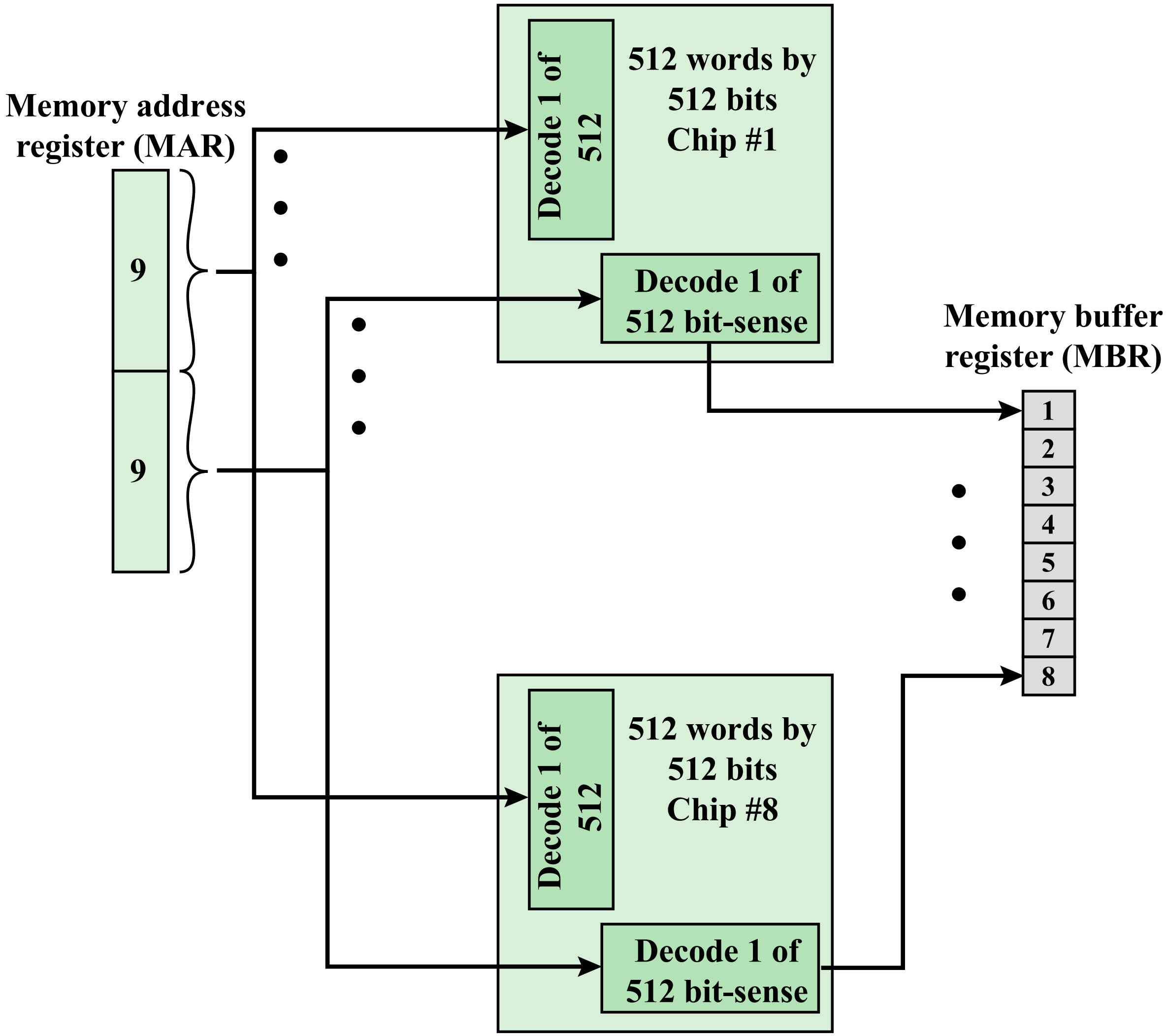
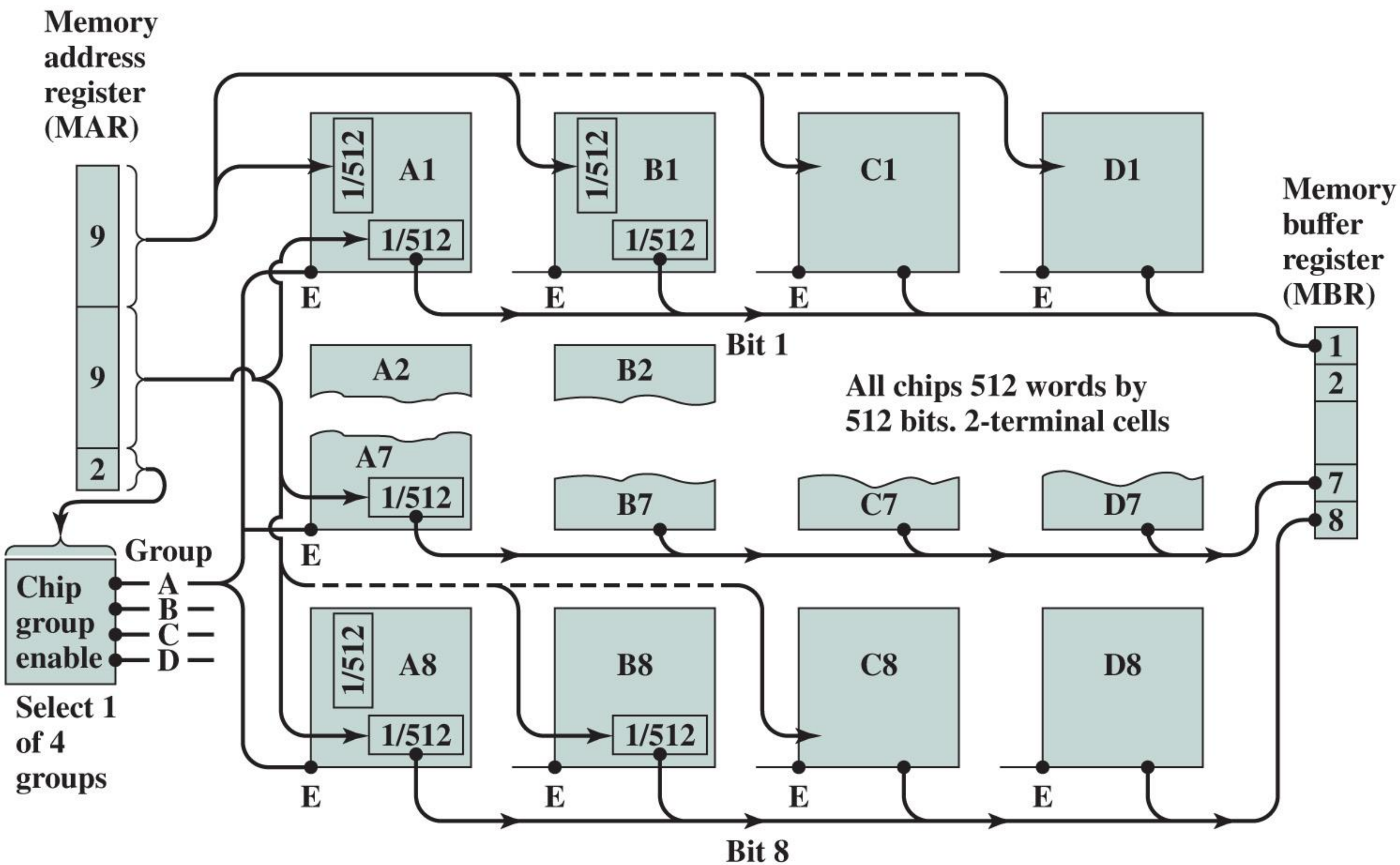
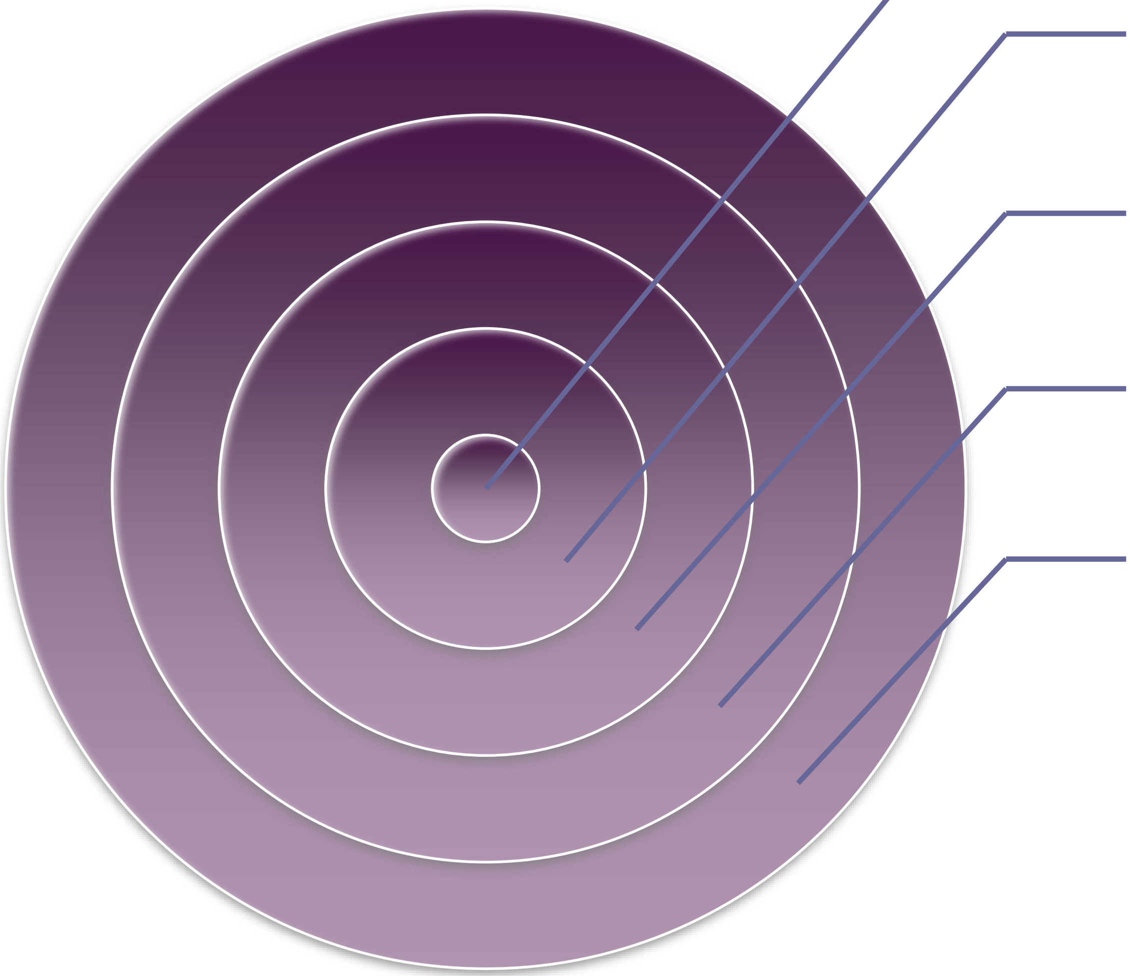


Figure 6.6
1-MB Memory Organization



Interleaved Memory

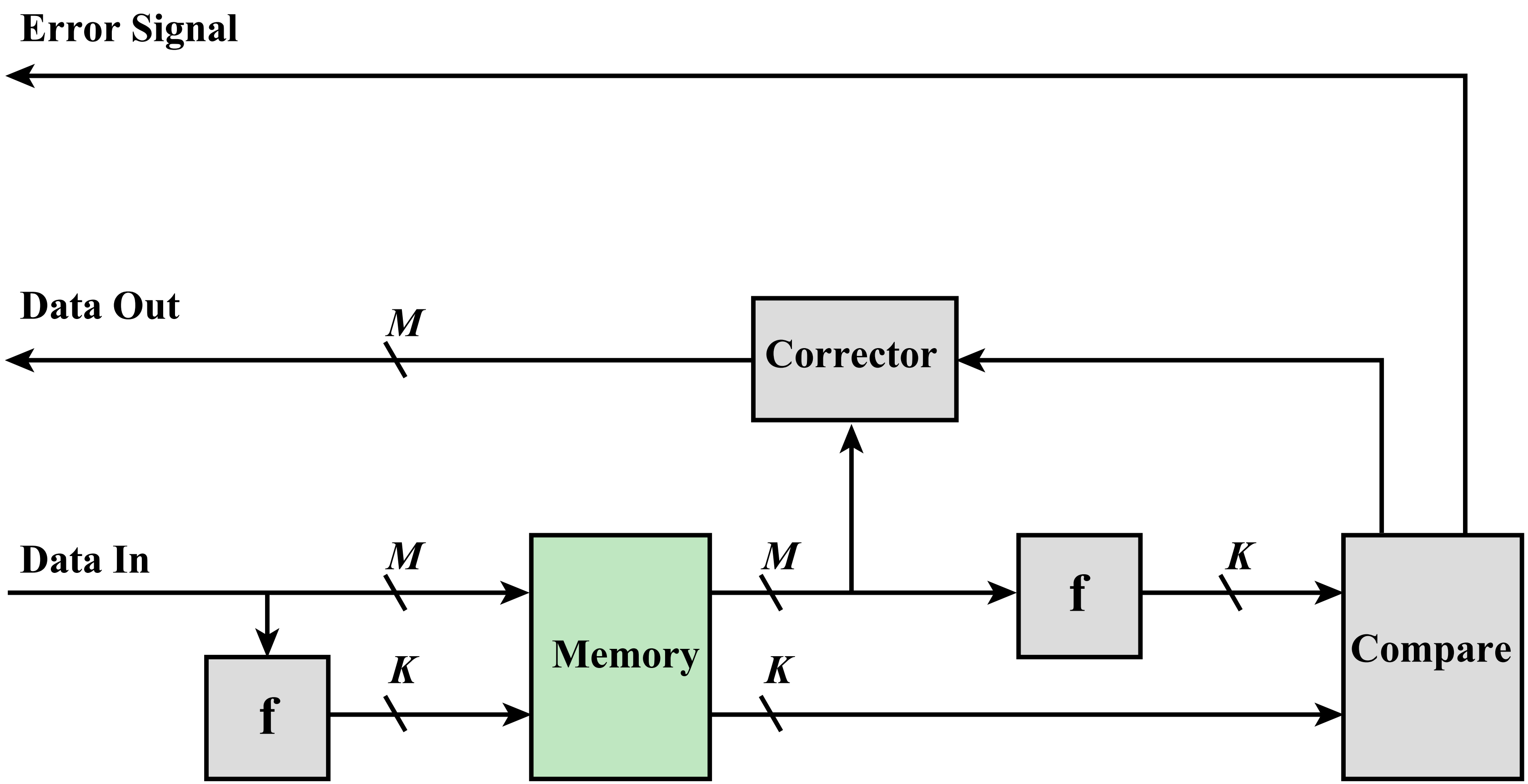


- Composed of a collection of DRAM chips
- Grouped together to form a *memory bank*
- Each bank is independently able to service a memory read or write request
- *K* banks can service *K* requests simultaneously, increasing memory read or write rates by a factor of *K*
- If consecutive words of memory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

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

Figure 6.7
Error-Correcting Code Function



Both the code and the data are stored. Thus, 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*.

Figure 6.8
Hamming Error-Correcting Code

The simplest of the error-correcting codes is the **Hamming code** devised by Richard Hamming at Bell Laboratories. Figure 6.8 uses Venn diagrams to illustrate the use of this code on 4-bit words ($M = 4$). With three intersecting circles, there are seven compartments. We assign the 4 data bits to the inner compartments (Figure 6.8a). 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 (Figure 6.8b). 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 6.8c), it is easily found. By checking the parity bits, 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. The error can therefore be corrected by changing that bit.

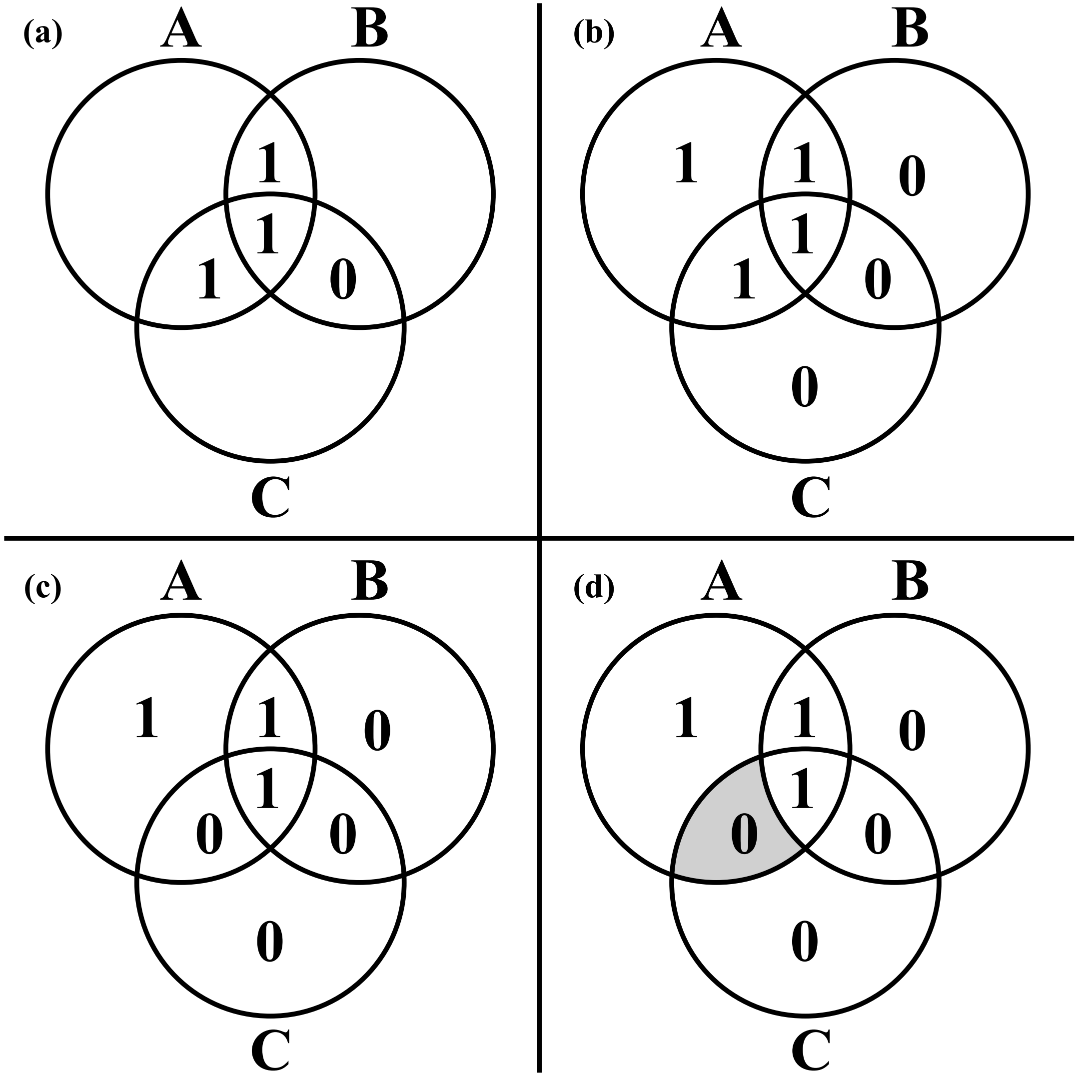


Table 6.2 Increase in Word Length with Error Correction

	Single-Error Correction		Single-Error Correction/ Double-Error Detection	
Data Bits	Check Bits	% Increase	Check Bits	% Increase
8	4	50.0	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

Figure 6.9
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

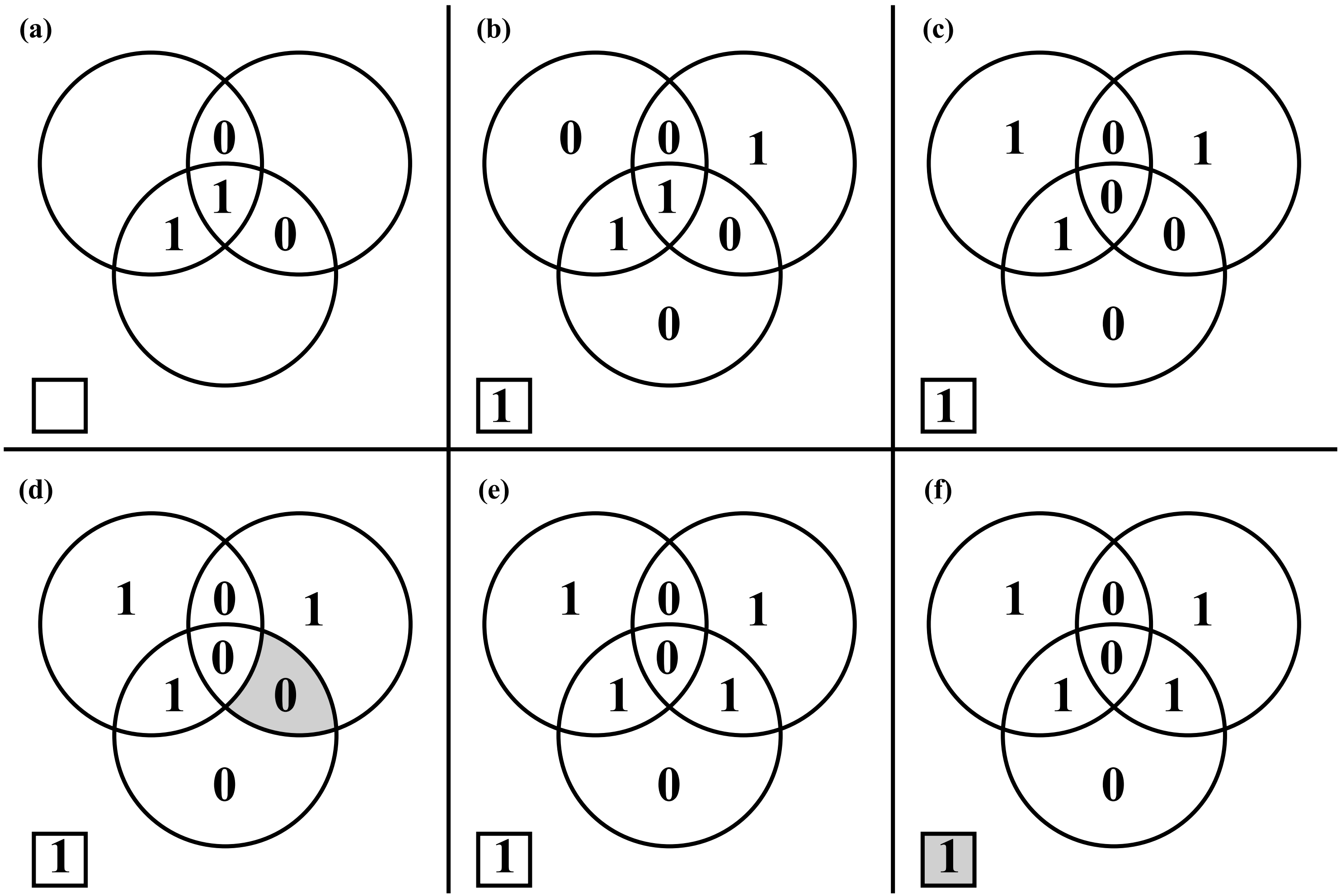
Pearson

Figure 6.10
Check Bit Calculation

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

Pearson

Figure 6.11
Hamming SEC-DEC Code



Pearson

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
  - The schemes that currently dominate the market are SDRAM and DDR-DRAM

SDRAM

DDR-DRAM

RDRAM

Synchronous DRAM (SDRAM)

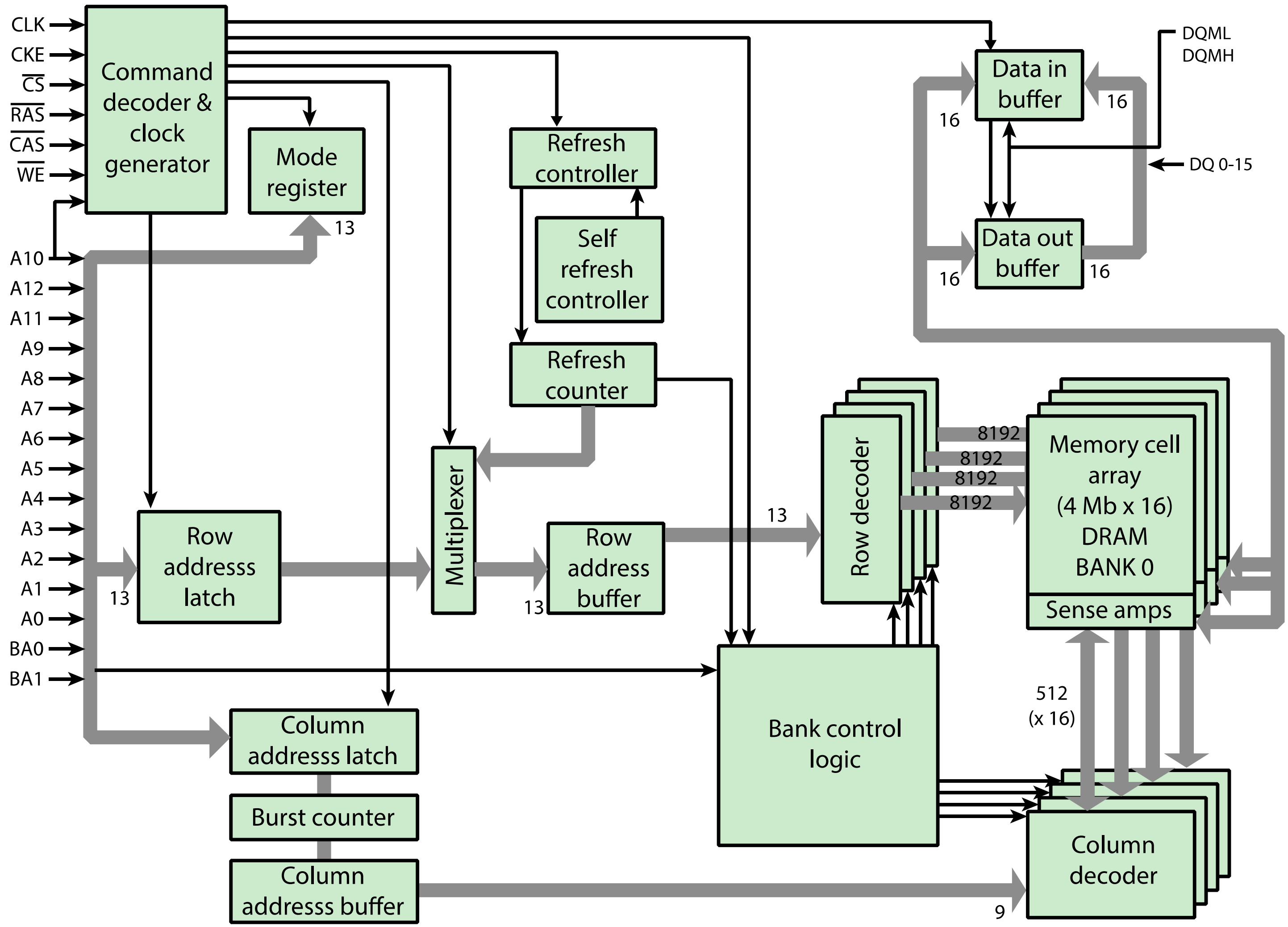
One of the most widely used forms of DRAM

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 6.12
256-Mb Synchronous Dynamic RAM (SDRAM)



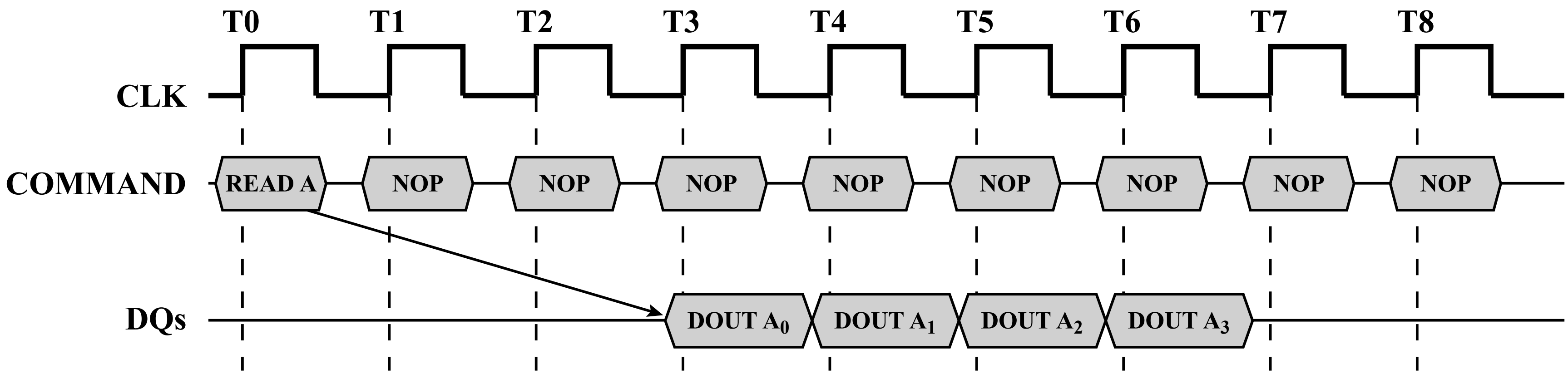
Pearson

Table 6.3
SDRAM Pin Assignments

A0 to A13	Address inputs
BA0, BA1	Bank address lines
CLK	Clock input
CKE	Clock enable
CS	Chip select
RAS	Row address strobe
CAS	Column address strobe
WE	Write enable
DQ0 to DQ7	Data input/output
DQM	Data mask

Figure 6.13

SDRAM Read Timing (burst length = 4, $\overline{CAS}$ latency = 2)



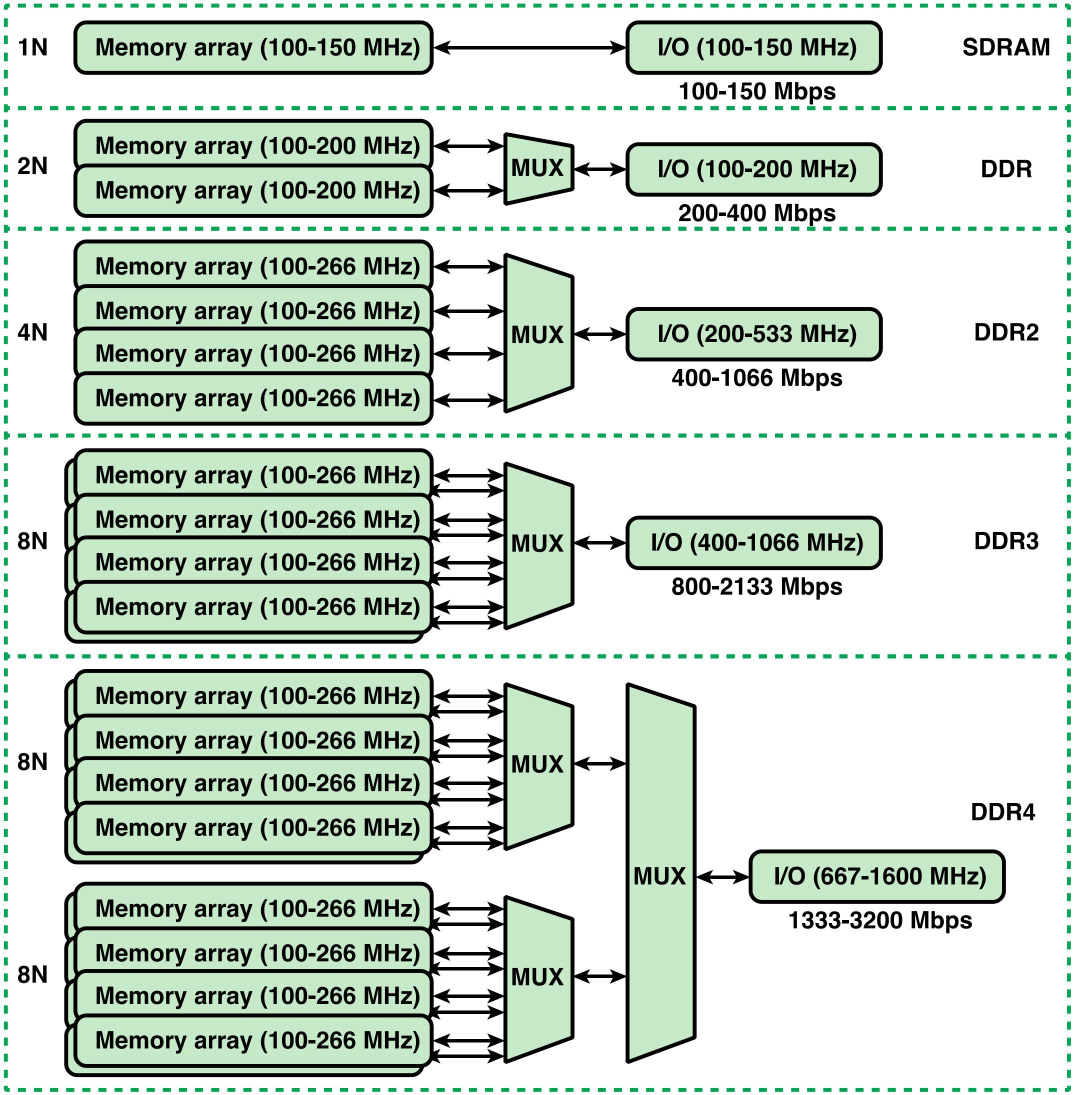
Double Data Rate SDRAM (DDR SDRAM)

- Developed by the JEDEC Solid State Technology Association (Electronic Industries Alliance’s semiconductor-engineering-standardization body)
- Numerous companies make DDR chips, which are widely used in desktop computers and servers
- DDR achieves higher data rates in three ways:
  - First, the data transfer is synchronized to both the rising and falling edge of the clock, rather than just the rising edge
  - Second, DDR uses higher clock rate on the bus to increase the transfer rate
  - Third, a buffering scheme is used

Table 6.4
DDR Characteristics

	DDR1	DDR2	DDR3	DDR4
Prefetch buffer (bits)	2	4	8	8
Voltage level (V)	2.5	1.8	1.5	1.2
Front side bus data rates (Mbps)	200—400	400—1066	800—2133	2133—4266

Figure 6.14
DDR Generations



Embedded DRAM (eDRAM)

- eDRAM is a DRAM integrated on the same chip or MCM of an application-specific integrated circuit (ASIC) or microprocessor
- For a number of metrics, eDRAM is intermediate between on-chip SRAM and off-chip DRAM
  - For the same surface area, eDRAM provides a larger size memory than SRAM but smaller than off-chip DRAM
  - eDRAM’s cost-per-bit is higher when compared to equivalent stand-alone DRAM chips used as external memory, but it has a lower cost-per-bit than SRAM
  - Access time to eDRAM is greater than SRAM but, because of its proximity and the ability to use wider busses, eDRAM provides faster access than DRAM
- Fundamentally eDRAMs use the same designs and architectures as DRAM

Figure 6.15
IBM z13 Storage Control (SC) Chip Layout

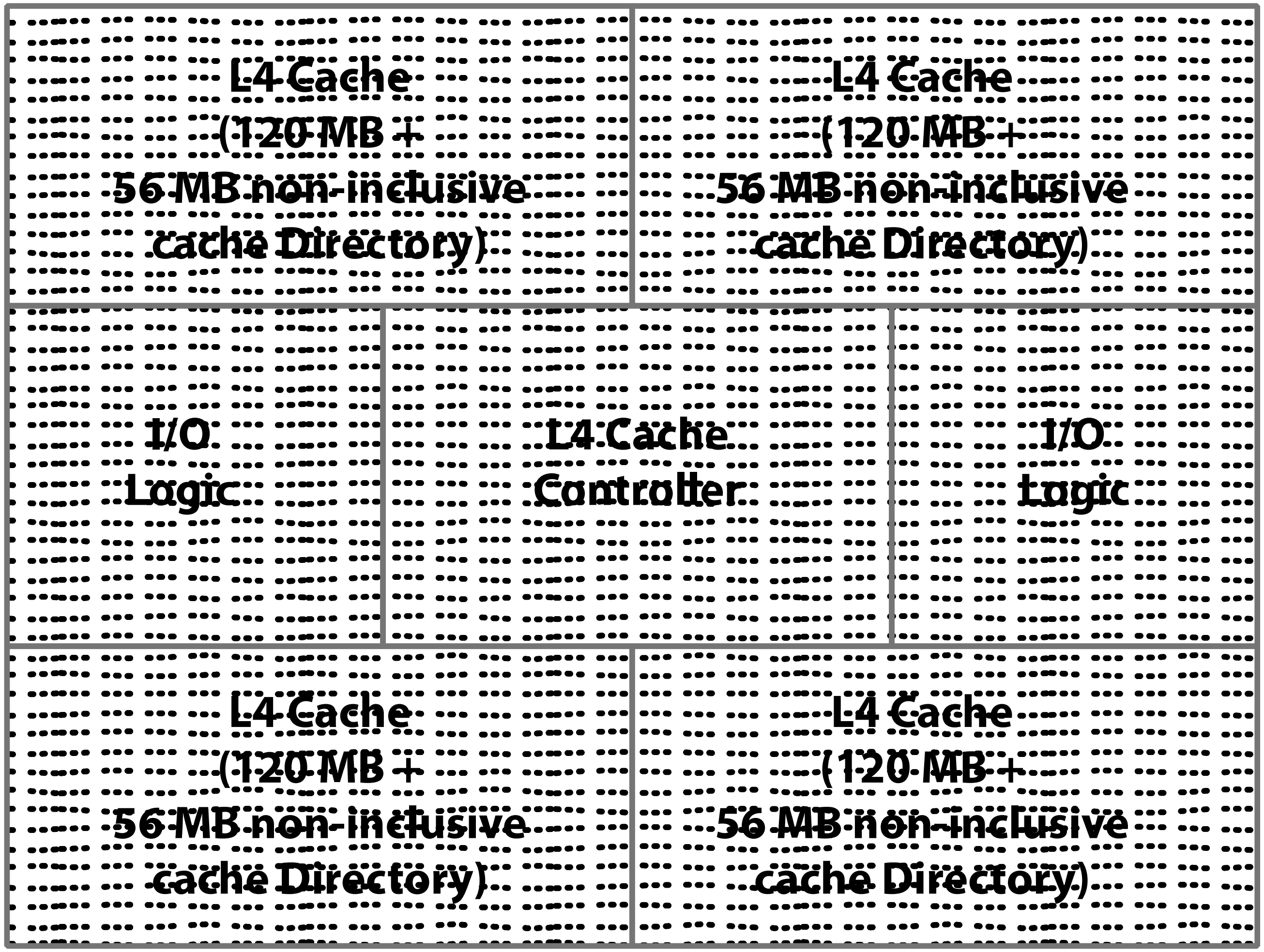
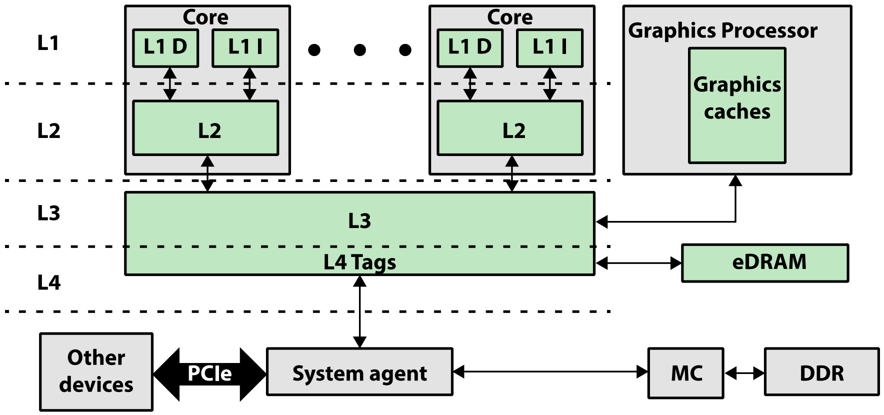
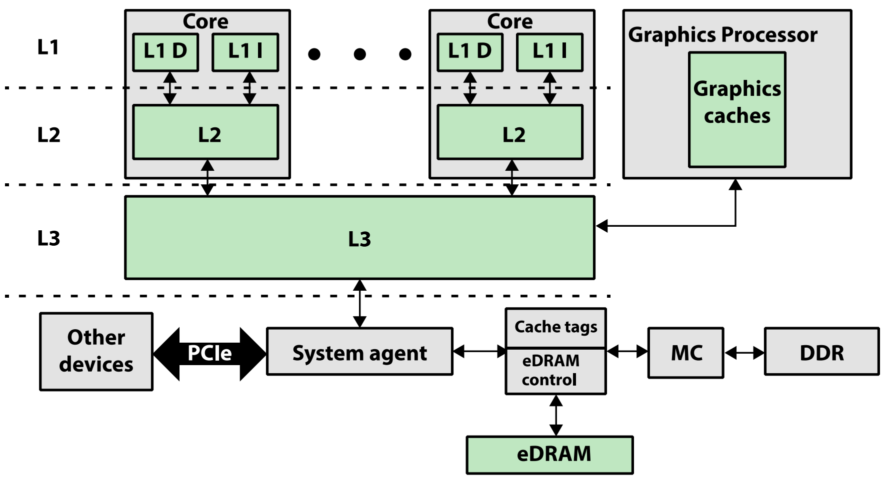


Figure 6.16
Use of eDRAM in Intel Core Systems



(a) Original use of eDRAM



(b) More recent use of eDRAM

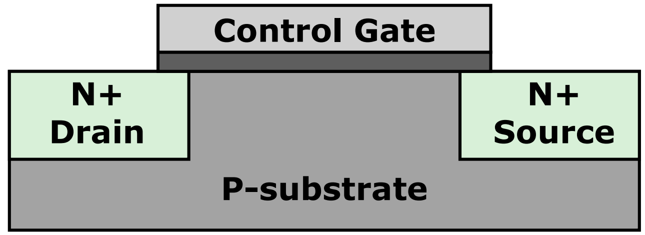
MC = memory controller

Pearson

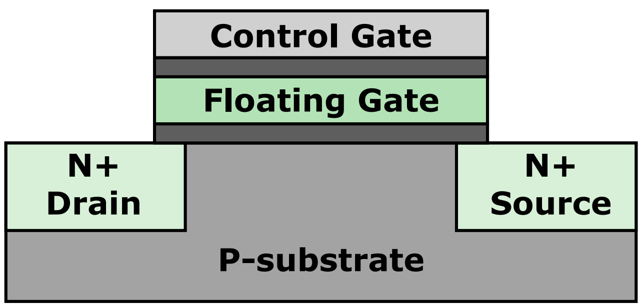
Flash Memory

- Used both for internal memory and external memory applications
- First introduced in the mid-1980's
- Is intermediate between EPROM and EEPROM in both cost and functionality
- Uses an electrical erasing technology like EEPROM
- It is possible to erase just blocks of memory rather than an entire chip
- Gets its name because the microchip is organized so that a section of memory cells are erased in a single action
- Does not provide byte-level erasure
- Uses only one transistor per bit so it achieves the high density of EPROM

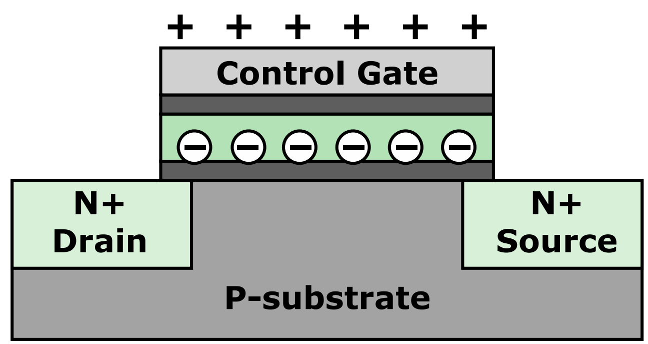
Figure 6.17
Flash Memory Operation



(a) Transistor structure

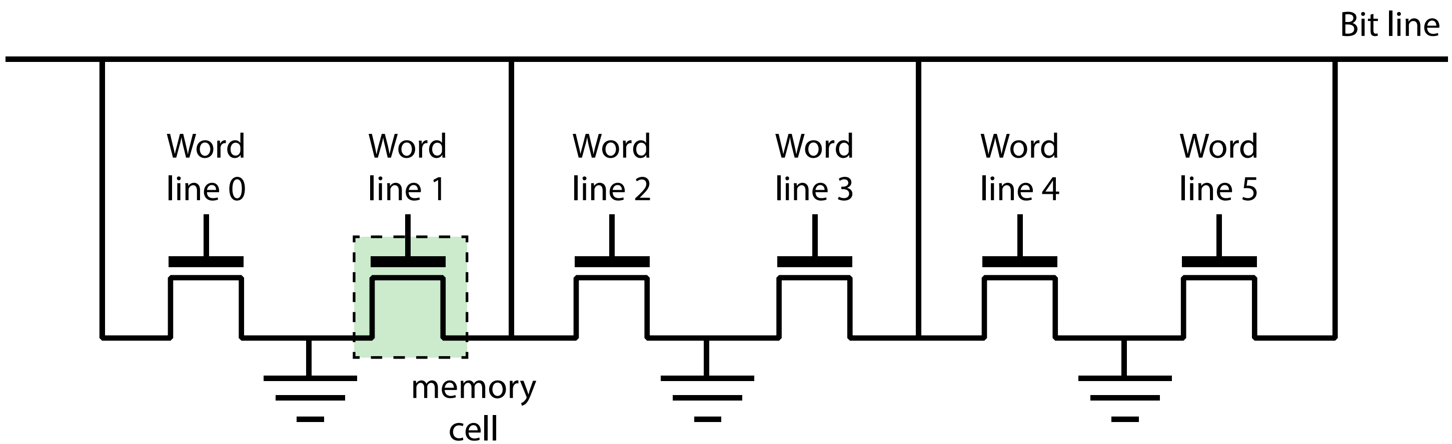


(b) Flash memory cell in one state

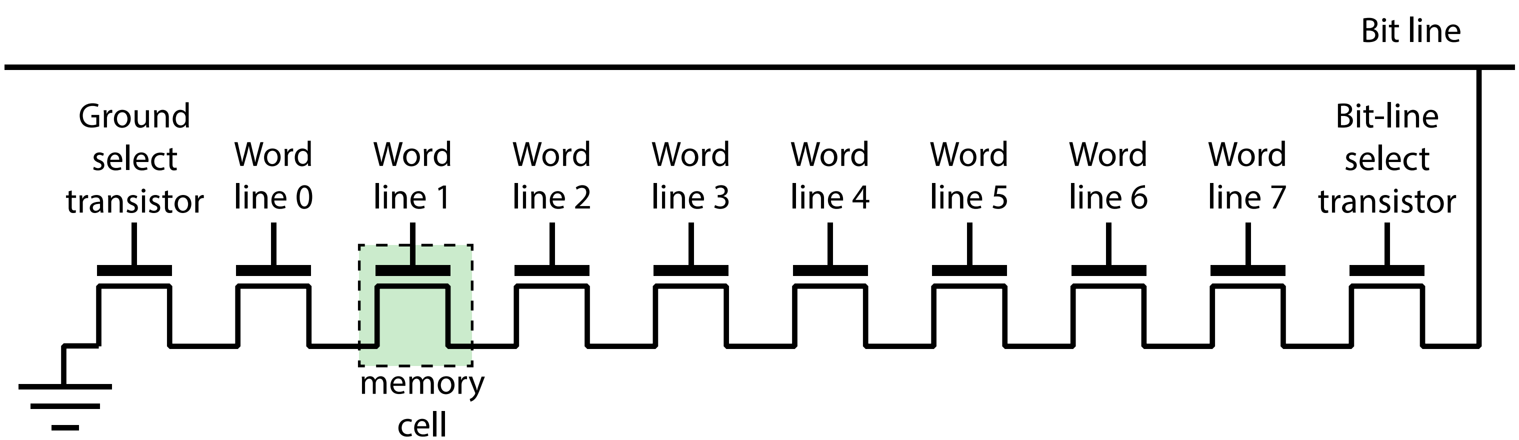


(c) Flash memory cell in zero state

Figure 6.18
Flash Memory Structures

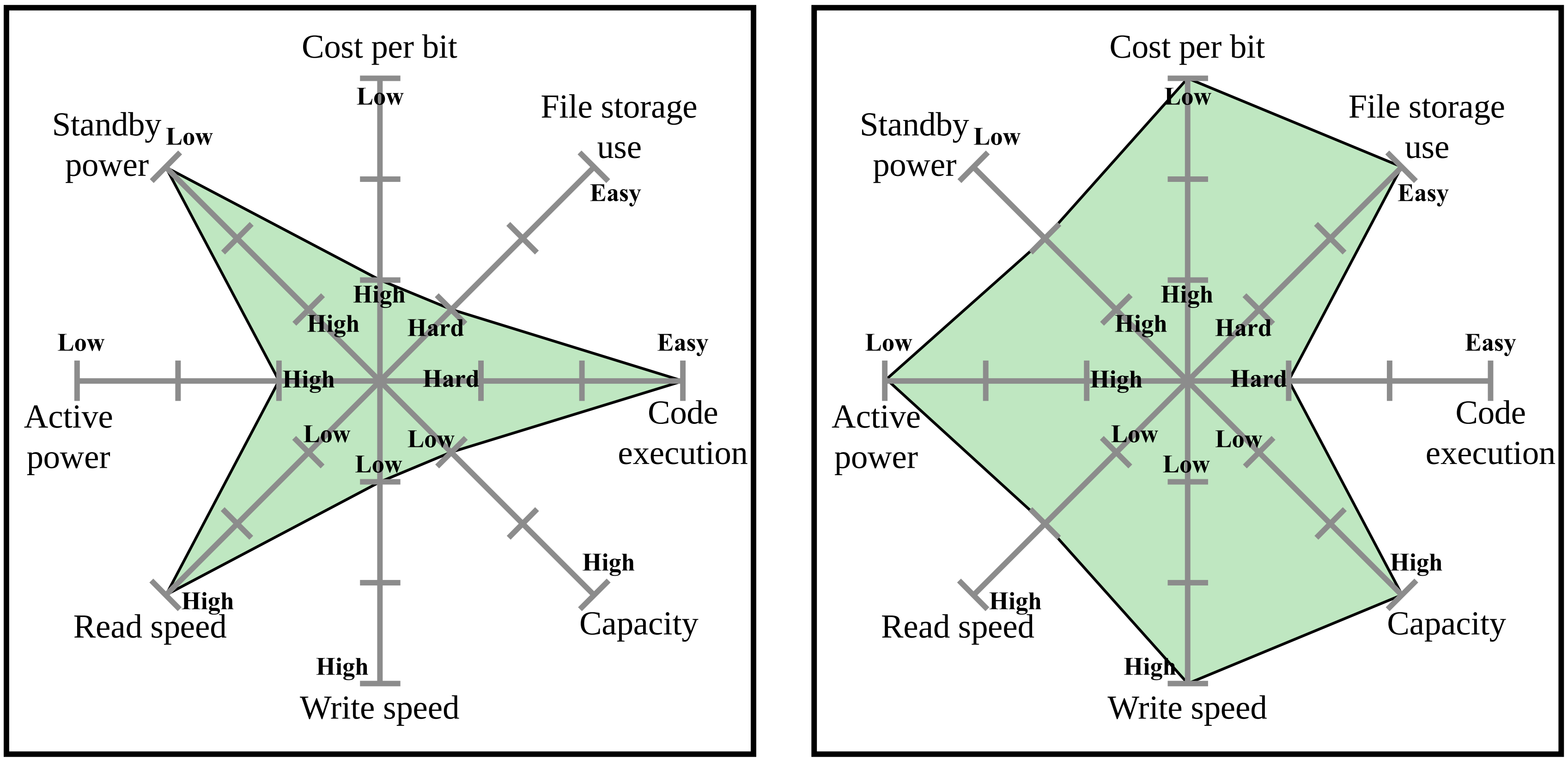


(a) NOR flash structure



(b) NAND flash structure

Figure 6.19
Kiviat Graphs for Flash Memory



(a) NOR

(b) NAND

Figure 6.20
Nonvolatile RAM within the Memory Hierarchy

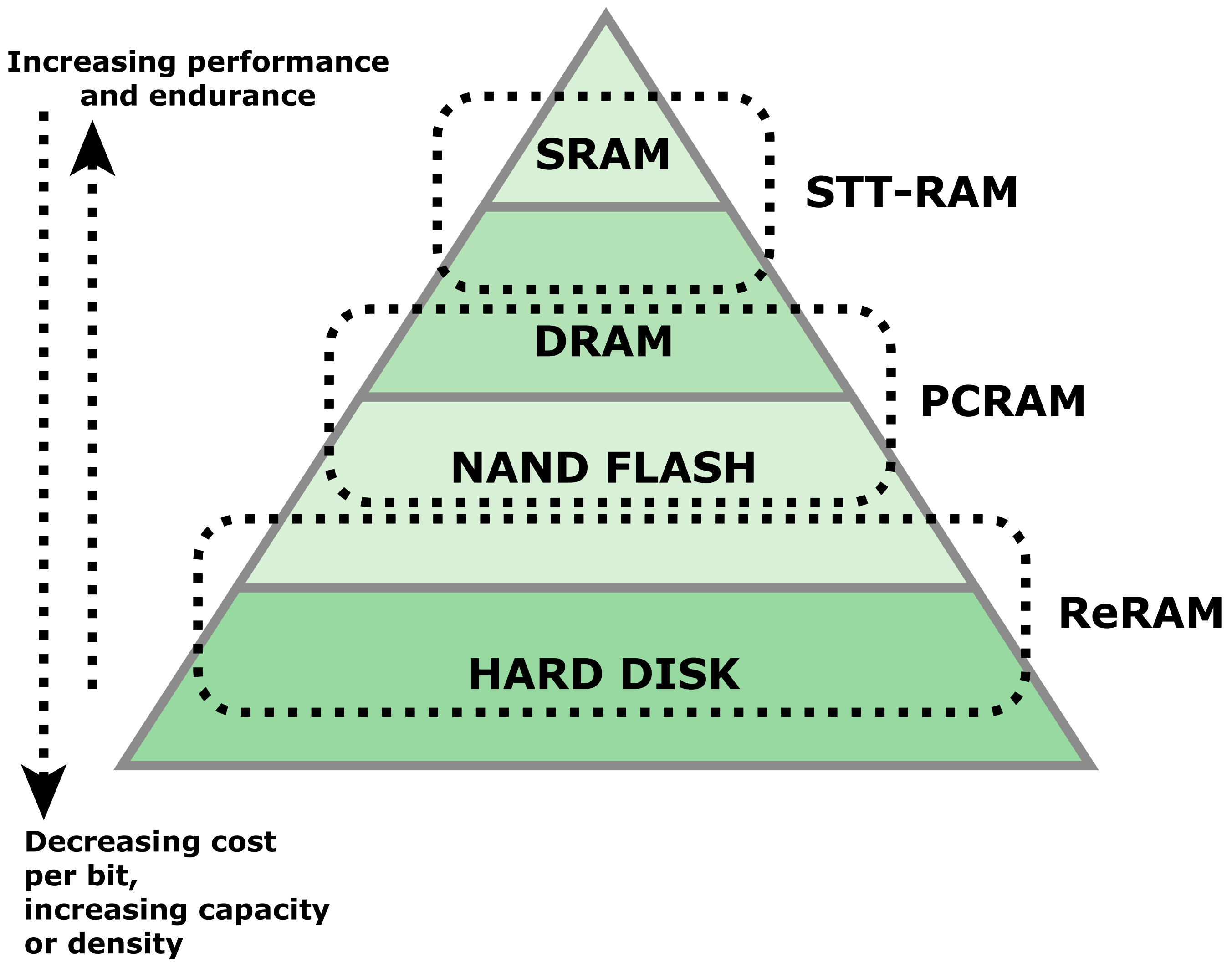
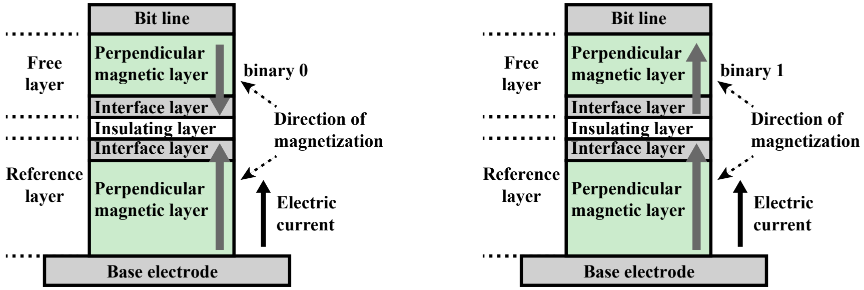
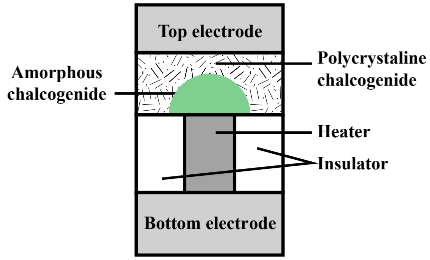


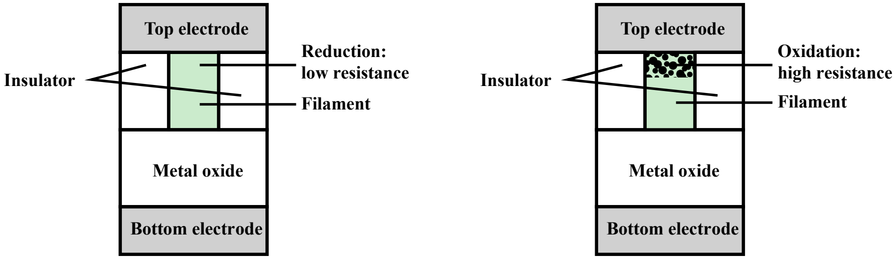
Figure 6.21
Nonvolatile RAM Technologies



(a) STT-RAM



(b) PCRAM



(c) ReRAM

Summary

Internal Memory

Chapter 6

- Semiconductor main memory
  - Organization
  - DRAM and SRAM
  - Types of ROM
  - Chip logic
  - Chip packaging
  - Module organization
  - Interleaved memory
- Error correction
- eDRAM
  - IBM z13 eDRAM cache structure
  - Intel core system cache structure
- DDR DRAM
  - Synchronous DRAM
  - DDR SDRAM
- Flash memory
  - Operation
  - NOR and NAND flash memory
- Newer nonvolatile solid-state memory technologies
  - STT-RAM
  - PCRAM
  - ReRAM