Computer Organization and Architecture

1

Basic Computer Networks

NLU-FIT

Chapter 05

INTERNAL MEMORY

KEY POINTS

2

Basic Computer Networks

NLU-FIT

- The two basic forms of semiconductor random access memory are dynamic RAM (DRAM) and static RAM (SRAM). SRAM is faster, more expensive, and less dense than DRAM, and is used for cache memory. DRAM is used for main memory.
- Error correction techniques are commonly used in memory systems.These involve adding redundant bits that are a function of the data bits to form an error-correcting code. If a bit error occurs, the code will detect and, usually, correct the error.
- To compensate for the relatively slow speed of DRAM, a number of advanced DRAM organizations have been introduced. The two most common are synchronous DRAM and RamBus DRAM. Both of these involve using the system clock to provide for the transfer of blocks of data.

Computer Organization and Architecture

5.1. Semiconductor Main Memory

5.1.1. Organization

- The basic element of a semiconductor memory is the memory cell.
- Although a variety of electronic technologies are used, all semiconductor memory cells share certain properties:
  - They exhibit two stable (or semistable) states, which can be used to represent binary 1 and 0.
  - They are capable of being written into (at least once), to set the state.
  - They are capable of being read to sense the state.

5.1.1. Organization

- Most commonly, the cell has three functional terminals capable of carrying an electrical signal. Figure 5.1

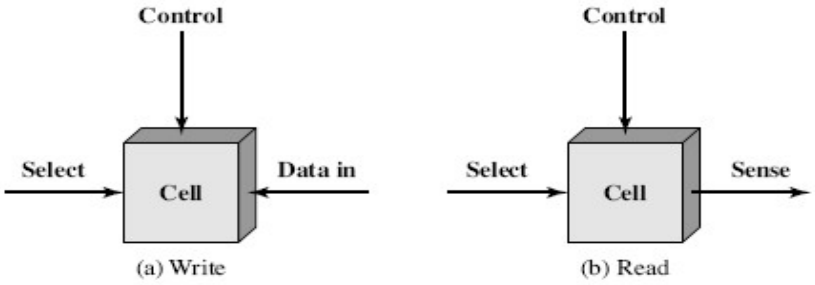


Figure 5.1. Memory Cell Operation

5.1.1. Organization

- The select terminal, as the name suggests, selects a memory cell for a read or write operation.
- The control terminal indicates read or write.
- For writing, the other terminal provides an electrical signal that sets the state of the cell to 1 or 0. For reading, that terminal is used for output of the cell's state.

5.1.2. DRAM and SRAM

- Lists the major types of semiconductor memory.

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		

5.1.2. DRAM and SRAM

- A dynamic RAM (DRAM) is made with cells that store data as charge on capacitors.
  - The presence or absence of charge in a capacitor is interpreted as a binary 1 or 0.
  - Because capacitors have a natural tendency to discharge, dynamic RAMs require periodic charge refreshing to maintain data storage.
  - The term dynamic refers to this tendency of the stored charge to leak away, even with power continuously applied.
- A static RAM (SRAM) will hold its data as long as power is supplied to it.

5.1.2. DRAM and SRAM

- **SRAM VERSUS DRAM**
  - Both static and dynamic RAMs are volatile; that is, power must be continuously supplied to the memory to preserve the bit values.
  - A dynamic memory cell is simpler and smaller than a static memory cell.Thus, a DRAM is more dense (smaller cells more cells per unit area) and less expensive than a corresponding SRAM.
  - On the other hand, a DRAM requires the supporting refresh circuitry.

5.1.2. DRAM and SRAM

- For larger memories, the fixed cost of the refresh circuitry is more than compensated for by the smaller variable cost of DRAM cells. Thus, DRAMs tend to be favored for large memory requirements.
- A final point is that SRAMs are generally somewhat faster than DRAMs.
- Because of these relative characteristics, SRAM is used for cache memory (both on and off chip), and DRAM is used for main memory.

5.1.3. Types of ROM

- As the name suggests, a read-only memory (ROM) contains a permanent pattern of data that cannot be changed.
- A ROM is nonvolatile; that is, no power source is required to maintain the bit values in memory.
- While it is possible to read a ROM, it is not possible to write new data into it.
- For a modest-sized requirement, the advantage of ROM is that the data or program is permanently in main memory and need never be loaded from a secondary storage device.
- Written during manufacture.

5.1.3. Types of ROM

- **Programmable ROM (PROM). Like the** ROM
  - The PROM is nonvolatile and may be written into only once.
  - For the PROM, the writing process is performed electrically and may be performed by a supplier or customer at a time later than the original chip fabrication.
  - Special equipment is required for the writing or “programming” process.

5.1.3. Types of ROM

- Another variation on read-only memory is the read-mostly memory, which is useful for applications in which read operations are far more frequent than write operations but for which nonvolatile storage is required.
- There are three common forms of read-mostly memory:
  - EPROM
  - EEPROM
  - and flash memory.

5.1.3. Types of ROM

- The optically **erasable programmable read-only memory (EPROM) is read** and written electrically, as with PROM.
  - ✓However, before a write operation, all the storage cells must be erased to the same initial state by exposure of the packaged chip to ultraviolet radiation.
- A more attractive form of read-mostly memory is **electrically erasable programmable read-only memory (EEPROM).** This is a read-mostly memory that can be written into at any time without erasing prior contents; only the byte or bytes addressed are updated.
  - ✓The write operation takes considerably longer than the read operation,

5.1.3. Types of ROM

15

Basic Computer Networks

NLU-FIT

- Another form of semiconductor memory is **flash memory (so named because** of the speed with which it can be reprogrammed).
- First introduced in the mid-1980s, flash memory is intermediate between EPROM and EEPROM in both cost and functionality.
- Like EEPROM, flash memory uses an electrical erasing technology. An entire flash memory can be erased in one or a few seconds, which is much faster than EPROM.
- In addition, it is possible to erase just blocks of memory rather than an entire chip.

5.1.3. Types of ROM

16

Basic Computer Networks

NLU-FIT

- An important application of ROMs is microprogramming
- Other potential applications include
  - Permanent storage
    - ✓Nonvolatile
  - Library subroutines for frequently wanted functions
  - Systems programs (BIOS)

5.1.4. Chip Logic

- As with other integrated circuit products, semiconductor memory comes in packaged. Each chip contains an array of memory cells.
- The array is organized into W words of B bits each. For example, a 16-Mbit chip could be organized as 1M 16-bit words.
- Figure 5.3 shows a typical organization of a 16-Mbit DRAM. In this case, 4 bits are read or written at a time.
- Logically, the memory array is organized as four square arrays of 2048 by 2048 elements. Various physical arrangements are possible.

5.1.4. Chip Logic

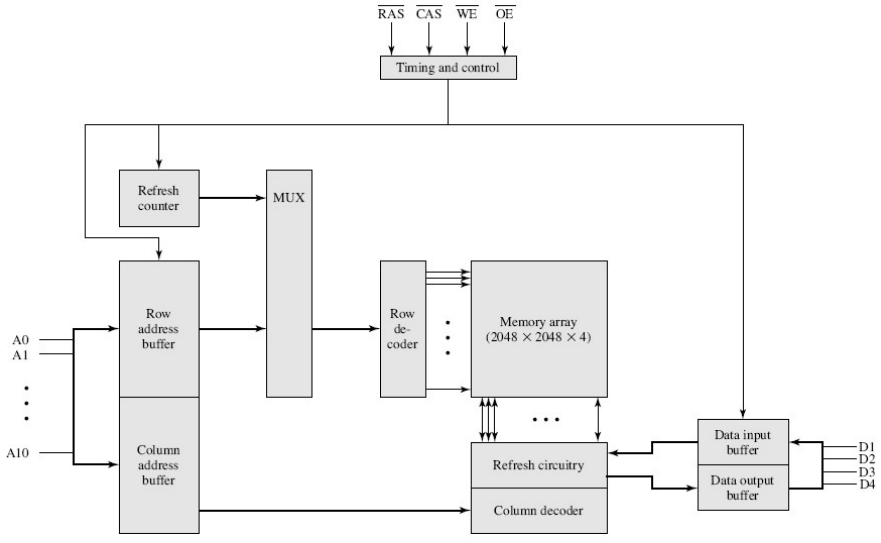


Figure 5.3. Typical 16 Megabit DRAM (4M x 4)

5.1.4. Chip Logic

- Address lines supply the address of the word to be selected.
  - In example, 11 address lines are needed to select one of 2048 rows.These 11 lines are fed into a row decoder, which has 11 lines of input and 2048 lines for output.
  - An additional 11 address lines select one of 2048 columns of 4 bits per column. Four data lines are used for the input and output of 4 bits to and from a data buffer.

5.1.4. Chip Logic

- Note that there are only 11 address lines (A0–A10), half the number you would expect for a 2048x2048 array.
- This is done to save on the number of pins.
- The 22 required address lines are passed through select logic external to the chip and multiplexed onto the 11 address lines.
  - First, 11 address signals are passed to the chip to define the row address of the array
  - and then the other 11 address signals are presented for the column address.
  - These signals are accompanied by row address select (RAS) and column address select (CAS) signals to provide timing to the chip.
  - The write enable (WE) and output enable (OE) pins determine whether a write or read operation is performed.
  - Two other pins are ground (Vss) and a voltage source (Vcc).

5.1.5. Chip Packaging

- An integrated circuit is mounted on a package that contains pins for connection to the outside world.
- Figure 5.4a shows an example EPROM package, which is an 8-Mbit chip organized as 1Mx8.
- The package includes 32 pins, which is one of the standard chip package sizes.
- The pins support the following signal lines:

5.1.5. Chip Packaging

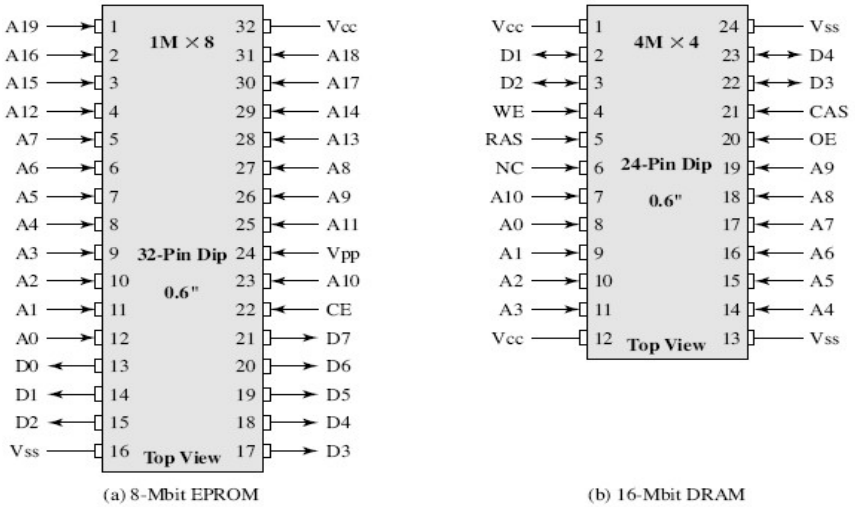


Figure 5.4 Typical Memory Package Pins and Signals

5.1.5. Chip Packaging

- The address of the word being accessed. For 1M words, a total of 20 (2^{20} =1M) pins are needed (A0–A19).
- The data to be read out, consisting of 8 lines (D0–D7).
- The power supply to the chip (Vcc).
- A ground pin (Vss).
- A chip enable (CE) pin.
  - ✓Because there may be more than one memory chip, each of which is connected to the same address bus, the CE pin is used to indicate whether or not the address is valid for this chip.
- A program voltage (Vpp) that is supplied during programming (write operations).

5.1.5. Chip Packaging

- A typical DRAM pin configuration is shown in Figure 5.4b, for a 16-Mbit chip organized as 4M x 4.
- There are several differences from a ROM chip. Because a RAM can be updated, the data pins are input/output.
  - The write enable (WE) and output enable (OE) pins indicate whether this is a write or read operation.
  - Because the DRAM is accessed by row and column, and the address is multiplexed, only 11 address pins are needed to specify the 4M row/column combinations (2^{11}x 2^{11}= 2^{22}= 4M).
  - The functions of the row address select (RAS) and column address select (CAS) pins were discussed previously.
  - Finally, the no connect (NC) pin is provided so that there are an even number of pins.

5.1.6. Module 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.
- As an example, Figure 5.5 shows how a memory module consisting of 256K 8-bit words could be organized.
  - For 256K words, an 18-bit address is needed and is supplied to the module from some external source (e.g., the address lines of a bus to which the module is attached).
  - The address is presented to 8 256K 1-bit chips, each of which provides the input/output of 1 bit.

5.1.6. Module Organization

- This organization works as long as the size of memory equals the number of bits per chip.
- In the case in which larger memory is required, an array of chips is needed.
- Figure 5.6 shows the possible organization of a memory consisting of 1M word by 8 bits per word.
  - In this case, we have four columns of chips, each column containing 256K words arranged as in Figure 5.5.

5.1.6. Module Organization

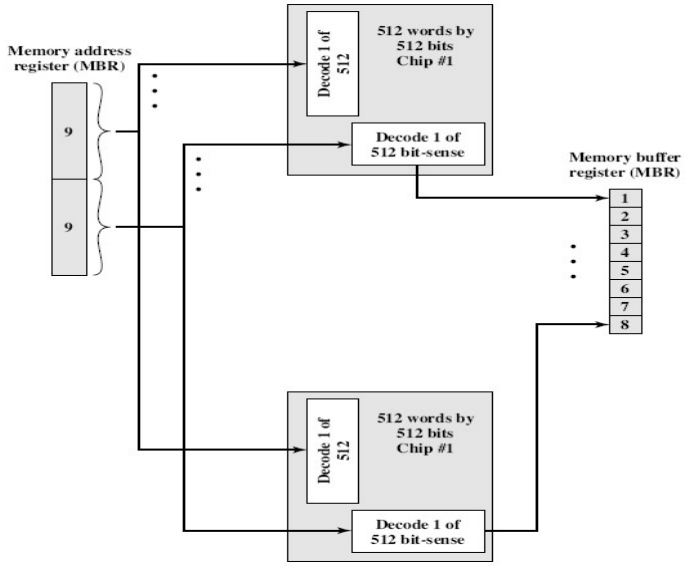
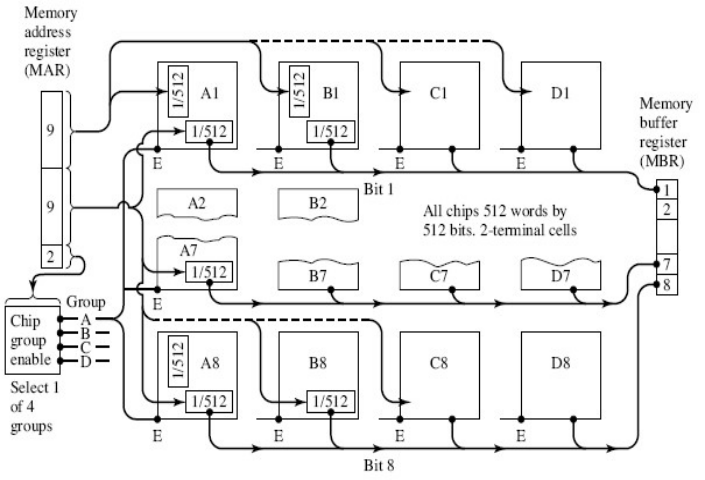


Figure 5.5. 256-KByte Memory Organization

5.1.6. Module Organization

Figure 5.6. 1-Mbyte Memory Organization



- 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.

5.2. Advanced DRAM Organization

29

Basic Computer Networks

NLU-FIT

- One of the most critical system bottlenecks when using high-performance processors is the interface to main internal memory.
- This interface is the most important pathway in the entire computer system.
- The basic building block of main memory remains the DRAM chip, as it has for decades; until recently, there had been no significant changes in DRAM architecture since the early 1970s.

5.2. Advanced DRAM Organization

30

Basic Computer Networks

NLU-FIT

- In recent years, a number of enhancements to the basic DRAM architecture have been explored, and some of these are now on the market.
- The schemes that currently dominate the market are SDRAM,DDR-DRAM,and RDRAM.

	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

Performance Comparison of Some DRAM Alternatives

5.2. Advanced DRAM Organization

Synchronous DRAM

- One of the most widely used forms of DRAM is the synchronous DRAM (SDRAM)
- Unlike the traditional DRAM, which is asynchronous, the SDRAM 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.

5.2. Advanced DRAM Organization

- In a typical DRAM, the processor presents addresses and control levels to the memory, indicating that a set of data at a particular location in memory should be either read from or written into the DRAM.
- After a delay, the access time, the DRAM either writes or reads the data.
- During the access-time delay, the DRAM performs various internal functions, such as activating the high capacitance of the row and column lines, sensing the data, and routing the data out through the output buffers.
- The processor must simply wait through this delay, slowing system performance.

5.2. Advanced DRAM Organization

33

Basic Computer Networks

NLU-FIT

- 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.
- CPU does not have to wait, it can do something else

5.2. Advanced DRAM Organization

34

Basic Computer Networks

NLU-FIT

- There is now an enhanced version of SDRAM, known as double data rate SDRAM (DDR-SDRAM) that overcomes the once-per-cycle limitation.
- SDRAM is limited by the fact that it can only send data to the processor once per bus clock cycle.
- DDRSDRAM can send data to the processor twice per clock cycle, once on the rising edge of the clock pulse and once on the falling edge.

5.2. Advanced DRAM Organization

Rambus DRAM

- RDRAM,developed by Rambus has been adopted by Intel for its Pentium and Itanium processors.
- It has become the main competitor to SDRAM.
- RDRAM chips are vertical packages, with all pins on one side.
- The special RDRAM bus delivers address and control information using an asynchronous block-oriented protocol.
- After an initial 480 ns access time, this produces the 1.6 GBps data rate.

Computer Organization and Architecture

- **Chapter 6: External Memory (*Reference*)**

- ***Reference**: Computer Organization and Architecture Designing for Performance (8th Edition), William Stallings, Prentice Hall, Upper Saddle River, NJ 07458.*