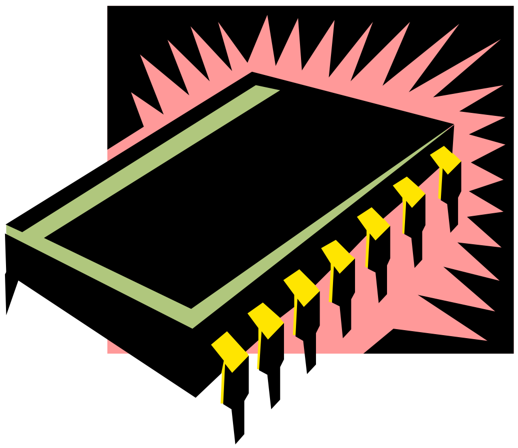
MEMORY



Memory

- Sequential circuits all depend upon the presence of memory.
  - A flip-flop can store one bit of information.
  - A register can store a single “word,” typically 32 or 64 bits.
- Memory allows us to store even larger amounts of data.

  Sequential Memories

  Shift Registers

  Charge Coupled Devices (CCD)

  - Read Only Memory (ROM)
  - ROM
  - PROM
  - EPROM
  - EAROM
  - Random Access Memory (RAM)
    - Static RAM (SRAM)
    - Dynamic RAM (DRAM)

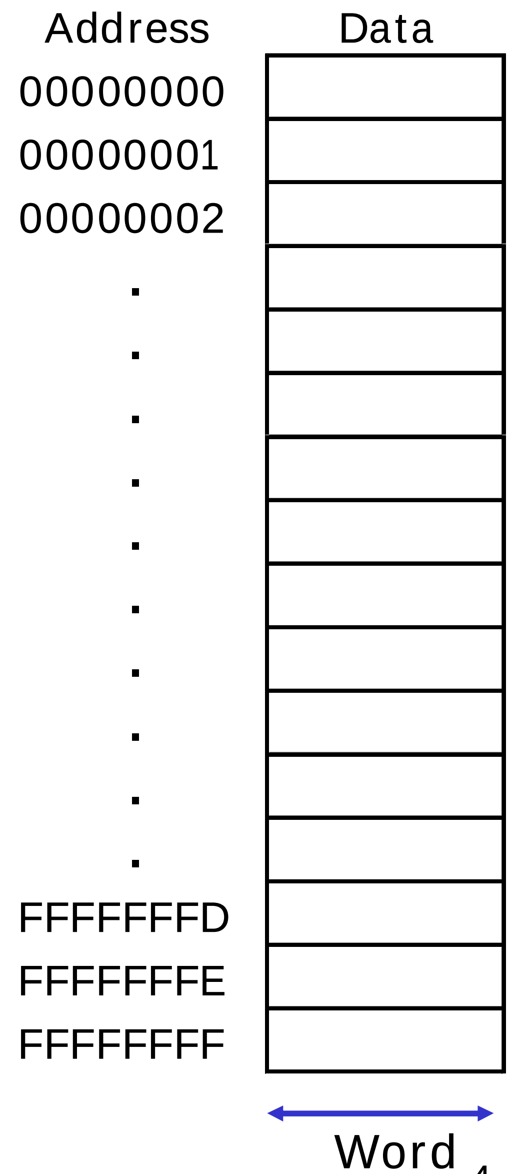
Advantages

- Small Size
- High speed
- Better Reliability
- Low Cost
- Ease of Expansion of memory

- One flip flop is one bit memory cell.

Picture of Memory

- You can think of memory as being one big array of data.
  - The address serves as an array index.
  - Each address refers to one word of data.
- You can read or modify the data at any given memory address, just like you can read or modify the contents of an array at any given index.



Memory Signal Types

- Memory signals fall into three groups
  - Address bus - selects one of memory locations
  - Data bus

    n number of data input lines

    - Read: the selected location's stored data is put on the data bus
    - Write (RAM): The data on the data bus is stored into the selected location
  - Control signals - specifies what the memory is to do
    - Control signals are usually active low
    - Includes read/write line and chip select line
    - Most common signals are:
      - CS: Chip Select; must be active to do anything
      - OE: Output Enable; active to read data
      - WR: Write; active to write data

Memory Address, Location and Size

- Location - the smallest selectable unit in memory
  - Has 1 or more data bits per location.
  - All bits in location are read/written together
  - Cannot manipulate single bits in a location
- For k address signals, there are 2^k locations in memory device
- Each location contains an n bit word
- Memory size is specified as
  - #loc x bits per location
    - 2^{24} x 16 RAM - 2^{24} = 16M words, each 16 bits long
    - 24 address lines, 16 data lines
  - #bits
    - The total storage capacity is 2^{24} x 16 = 2^{28} bits

Size matters!

- Memory sizes are usually specified in numbers of bytes (1 byte= 8 bits).
- The 2^{28}-bit memory on the previous page translates into:

$$2^{28} \text{ bits } / \text{ } 8 \text{ bits per byte} = 2^{25} \text{ bytes}$$

- With the abbreviations below, this is equivalent to 32 megabytes.

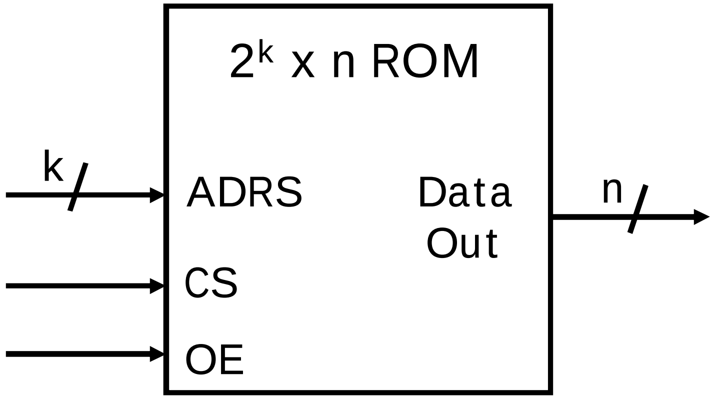
	Prefix	Base 2	Base 10
K	Kilo	2^{10} = 1,024	10^{3} = 1,000
M	Mega	2^{20} = 1,048,576	10^{6} = 1,000,000
G	Giga	2^{30} = 1,073,741,824	10^{9} = 1,000,000,000

Memory

- Parameters for classification of memory devices:

1. Principle of operation
2. Physical characteristics
3. Mode of access
4. Technology used for fabrication

Read-only memory (ROM)



- Non-volatile
  - If un-powered, its content retains
- Read-only
  - normal operation cannot change contents

- k-bit ADRS specifies the address or location to read from
- A Chip Select, CS, enables or disables the RAM
- An Output Enable, OE, turns on or off tri-state output buffers
- Data Out will be the n-bit value stored at ADRS

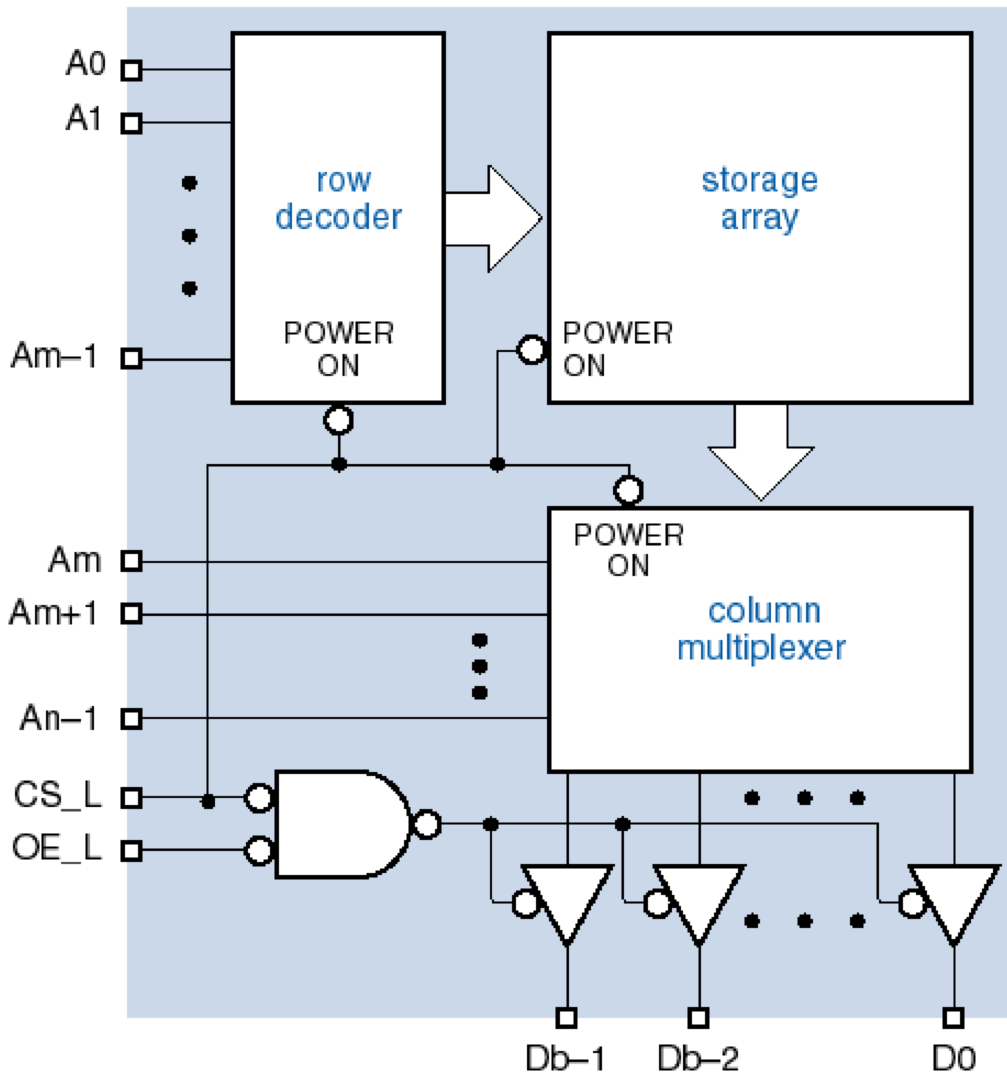
ROM Programming

- Content loading (programming) done many ways depending on device type
  - Programmed ROM (PROM): contents loaded at the factory
    - hardwired - can't be changed
    - embedded mass-produced systems
  - OTP (One Time Programmable): Programmed by user
  - UVPROM: reusable, erased by UV light
  - EEPROM: Electrically erasable; clears entire blocks with single operation

ROM Usage

- ROMs are useful for holding data that never changes.
  - Arithmetic circuits might use tables to speed up computations of logarithms or divisions.
  - Many computers use a ROM to store important programs that should not be modified, such as the system BIOS.
  - Application programs of embedded systems, PDAs, game machines, cell phones, vending machines, etc., are stored in ROMs

ROM Structure



Memories and functions

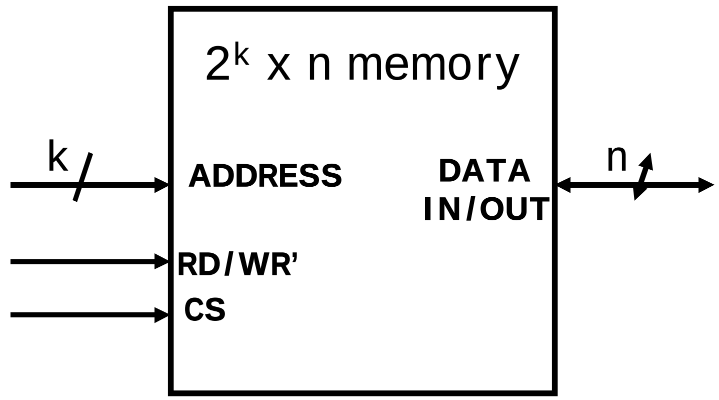
- ROMs are actually combinational devices, not sequential ones!
  - You can store arbitrary data into a ROM, so the same address will always contain the same data.
  - You can think of a ROM as a combinational circuit that takes an address as input, and produces some data as the output.
- A ROM table is basically just a truth table.
  - The table shows what data is stored at each ROM address.
  - You can generate that data combinationally, using the address as the input.

Address $A_2A_1A_0$	Data $V_2V_1V_0$
000	000
001	100
010	110
011	100
100	101
101	000
110	011
111	011

Introduction to RAM

- Random-access memory, or RAM, provides large quantities of temporary storage in a computer system.
  - Memory cells can be accessed to transfer information to or from any desired location, with the access taking the same time regardless of the location
- Volatility
  - Most RAMs lose their memory when power is removed
  - NVRAM = RAM + battery
  - Or use EEPROM
- SRAM (Static RAM)
  - Memory behaves like latches or flip-flops
- DRAM (Dynamic Memory)
  - Memory lasts only for a few milliseconds
  - Must “refresh” locations by reading or writing

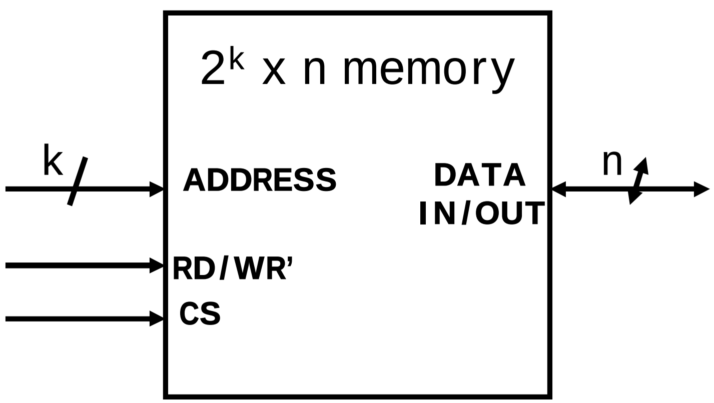
Block diagram of RAM



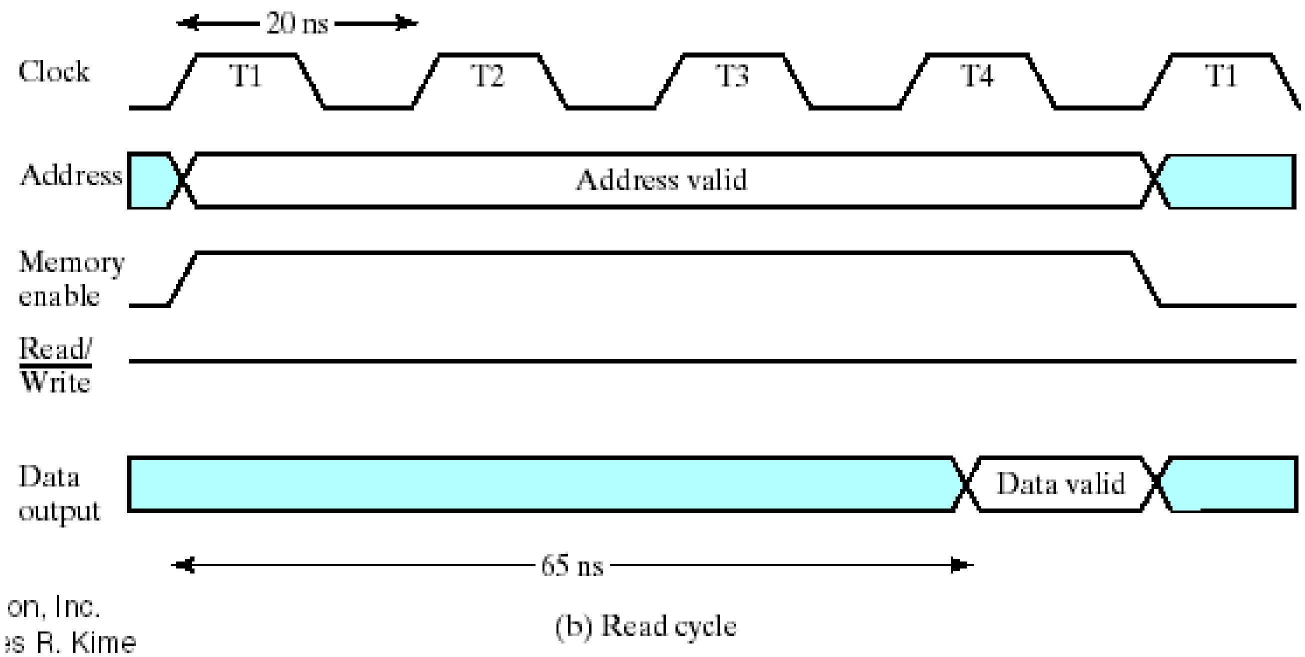
- This block diagram introduces the main interface to RAM.
  - A Chip Select, CS, enables or disables the RAM.
  - ADRS specifies the address or location to read from or write to.
  - RD/WR' selects between reading from or writing to the memory.
    - To read from memory, RD/WR' should be set to 1.
      DATA IN/OUT will be the n-bit value stored at ADRS.
    - To write to memory, we set RD/WR' to 0.
      DATA IN/OUT is the n-bit value to save in memory.

Reading RAM

- To *read* from this RAM, the controlling circuit must:
  - Enable the chip by ensuring CS = 1.
  - Select the read operation, by setting RD/WR’ = 1.
  - Send the desired address to the ADRS input.
  - The contents of that address appear on DATA IN/OUT after a little while.



Reading RAM

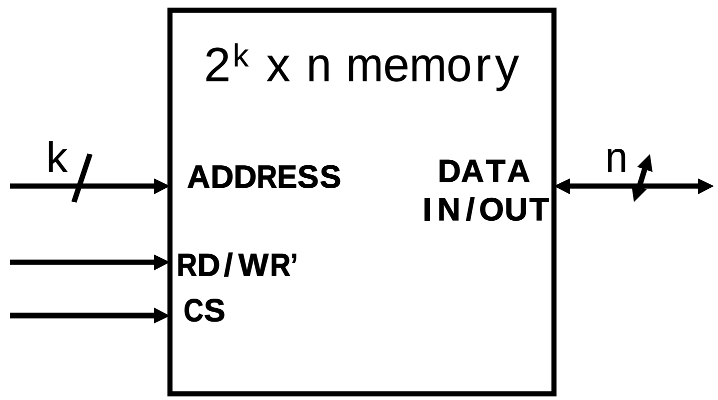


(b) Read cycle

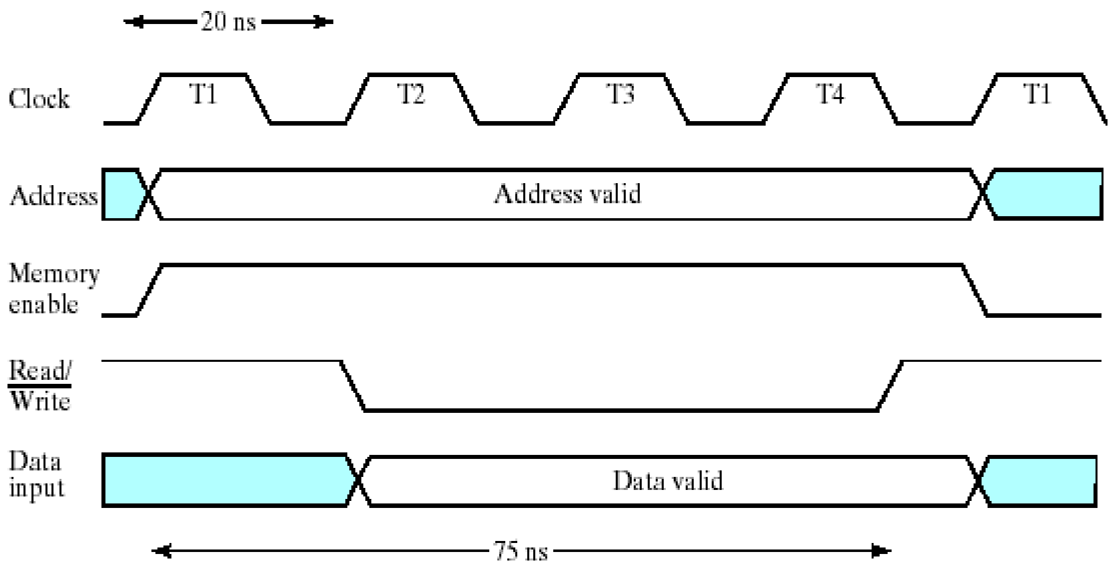
- 50 MHz CPU – 20 ns clock cycle time
- Memory access time= 65 ns
  - Maximum time from the application of the address to the appearance of the data at the Data Output

Writing RAM

- To *write* to this RAM, you need to:
  - Enable the chip by setting CS = 1.
  - Select the write operation, by setting RD/WR' = 0.
  - Send the desired address to the ADRS input.
  - Send the word to store to the DATA IN/OUT.



Writing RAM



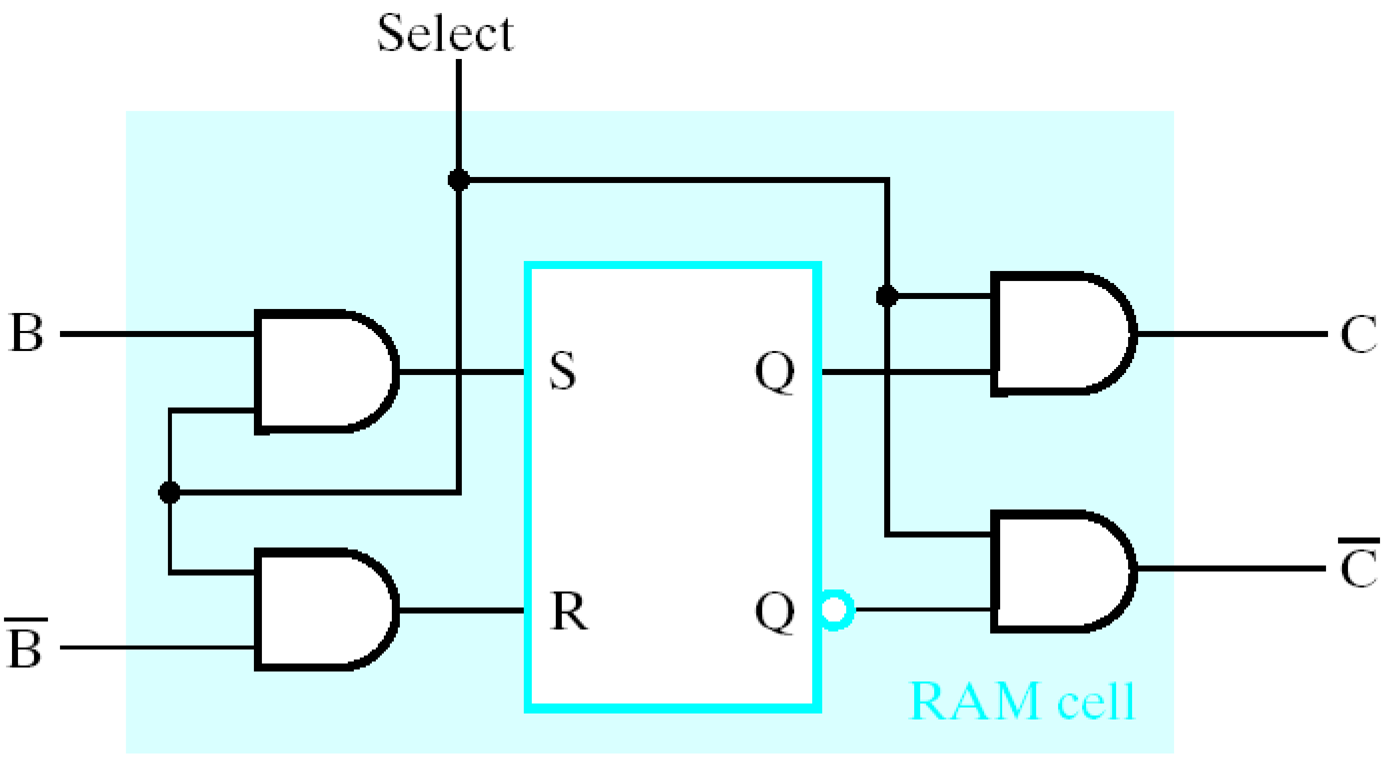
(a) Write cycle

- 50 MHz CPU – 20 ns clock cycle time
- Write cycle time= 75 ns
  - Maximum time from the application of the address to the completion of all internal memory operations to store a word

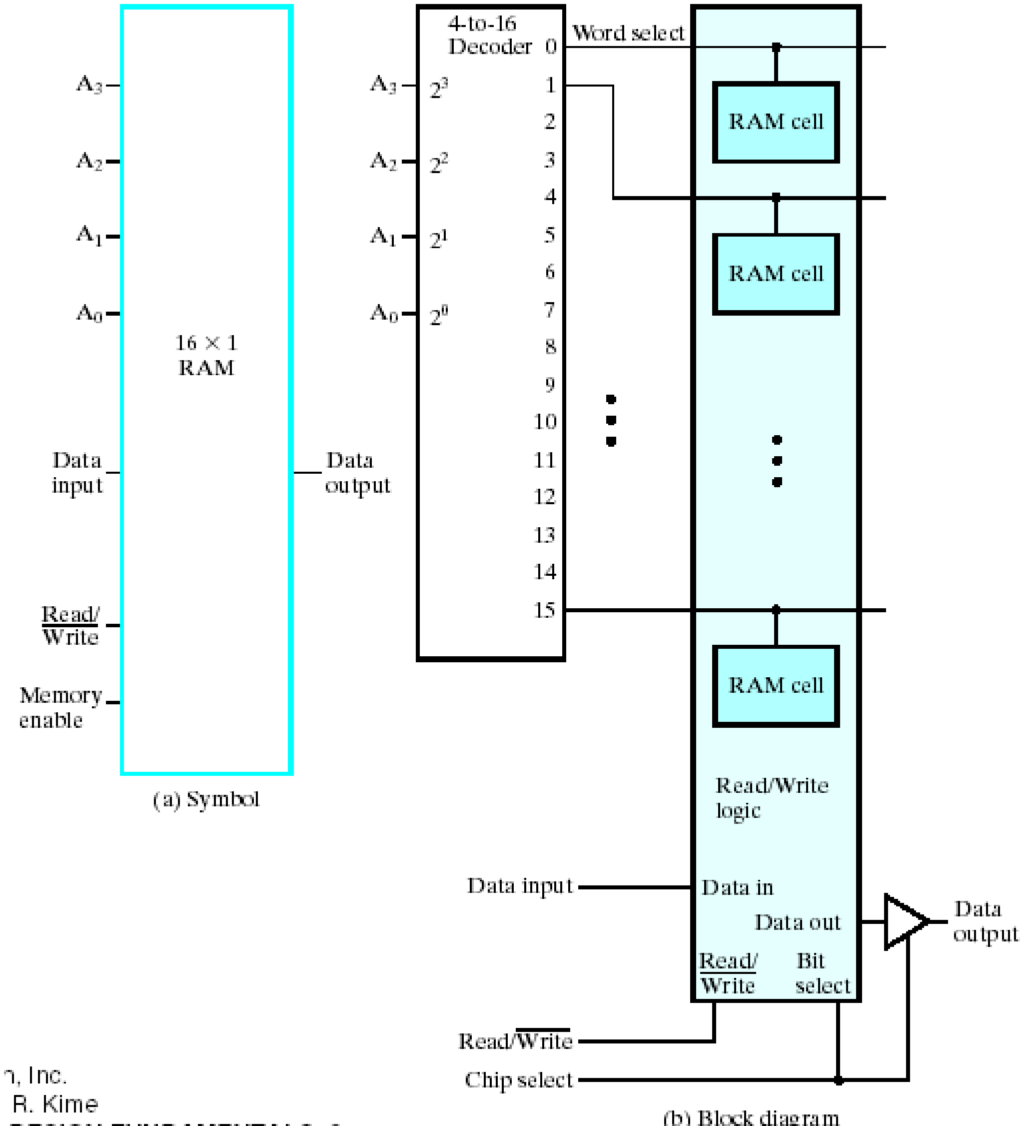
Static memory

- How can you implement the memory chip?
- There are many different kinds of RAM.
  - We'll start off discussing static memory, which is most commonly used in caches and video cards.
  - Later we mention a little about dynamic memory, which forms the bulk of a computer's main memory.
- Static memory is modeled using one *latch* for each bit of storage.
- Why use latches instead of flip flops?
  - A latch can be made with only two NAND or two NOR gates, but a flip-flop requires at least twice that much hardware.
  - In general, smaller is faster, cheaper and requires less power.
  - The tradeoff is that getting the timing exactly right is a pain.

RAM Cell with SR Latch

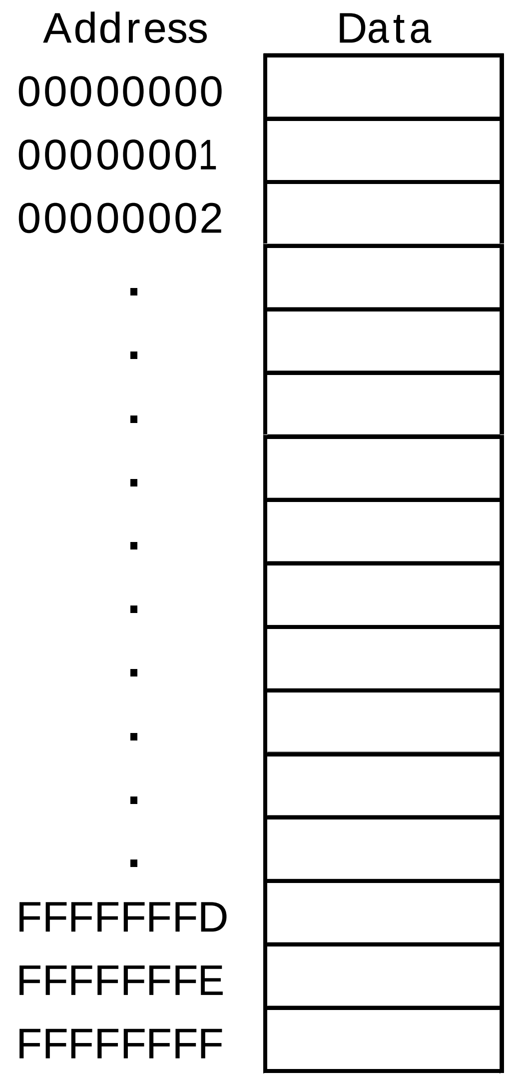


16-Word by 1-bit RAM Chip



Typical memory sizes

- Some typical memory capacities:
  - PCs usually come with 128-256MB RAM.
  - PDAs have 8-64MB of memory.
  - Digital cameras and MP3 players can have 32MB or more of storage.
- Many operating systems implement virtual memory, which makes the memory seem larger than it really is.
  - Most systems allow up to 32-bit addresses. This works out to 2^{32}, or about four billion, different possible addresses.
  - With a data size of one byte, the result is apparently a 4GB memory!
  - The operating system uses hard disk space as a substitute for “real” memory.



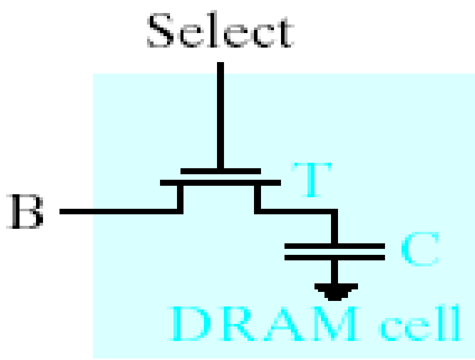
RAM Summary

- A RAM looks like a bunch of registers connected together, allowing users to select a particular address to read or write.
- Much of the hardware in memory chips supports this selection process:
  - Chip select inputs
  - Decoders
  - Tri-state buffers
- By providing a general interface, it's easy to connect RAMs together to make "longer" and "wider" memories.
- Next, we'll look at some other types of memories

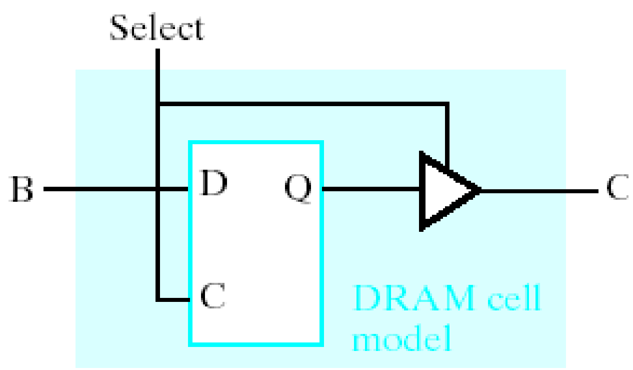
Dynamic memory

- Dynamic memory is built with capacitors.
  - A stored charge on the capacitor represents a logical 1.
  - No charge represents a logic 0.
- However, capacitors lose their charge after a few milliseconds. The memory requires constant refreshing to recharge the capacitors. (That's what's "dynamic" about it.)
- Dynamic RAMs tend to be physically smaller than static RAMs.
  - A single bit of data can be stored with just one capacitor and one transistor, while static RAM cells typically require 4-6 transistors.
  - This means dynamic RAM is cheaper and denser—more bits can be stored in the same physical area.

DRAM Cell



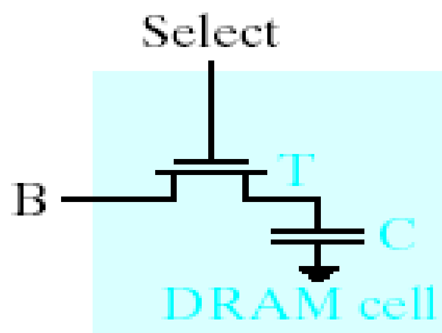
(a)



(h)

- DRAM cell: One transistor and one capacitor
  - 1/0 = capacitor charged/discharged
- SRAM cell: Six transistors – Costs 3 times more (cell complexity)
- Cost per bit is less for DRAM – reason for why large memories are DRAMs

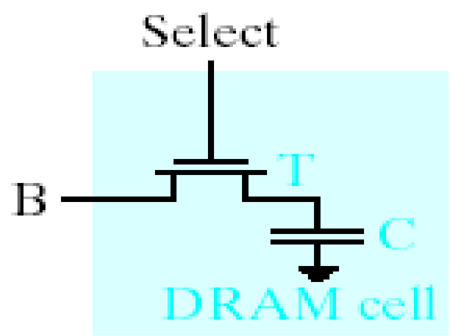
DRAM Cell Read



(a)

- Precharge bit line to VDD/2.
- Take the select line HIGH.
- Detect whether current flows into or out of the cell
- Cell contents are destroyed by the read!
    - Must write the bit value back after reading.

DRAM Cell Write



(a)

- Take the select line HIGH.
- Set the bit line LOW or HIGH to store 0 or 1.
- Take the select line LOW.
    - The stored charge for a 1 will eventually leak off.
    - Typical devices require each cell to be refreshed once every 4 to 64 msec.

Dynamic vs. static memory

- In practice, dynamic RAM is used for a computer's main memory, since it's cheap and you can pack a lot of storage into a small space.
  - These days you can buy 256MB of memory for as little as $60.
  - You can also load a system with 1.5GB or more of memory.
- The disadvantage of dynamic RAM is its speed.
  - Transfer rates are 800MHz at best, which can be much slower than the processor itself.
  - You also have to consider latency, or the time it takes data to travel from RAM to the processor.
- Real systems augment dynamic memory with small but fast sections of static memory called caches.
  - Typical processor caches range in size from 128KB to 320KB.
  - That's small compared to a 128MB main memory, but it's enough to significantly increase a computer's overall speed.

ROMs vs. RAMs

- There are some important differences between ROM and RAM.
  - ROMs are “non-volatile”—data is preserved even without power. On the other hand, RAM contents disappear once power is lost.
  - ROMs require special (and slower) techniques for writing, so they’re considered to be “read-only” devices.
- Some newer types of ROMs do allow for easier writing, although the speeds still don’t compare with regular RAMs.
  - MP3 players, digital cameras and other toys use CompactFlash, Secure Digital, or MemoryStick cards for non-volatile storage.
  - Many devices allow you to upgrade programs stored in “flash ROM.”