CSCI 4717/5717
Computer Architecture

Topic: Internal Memory Details

Reading: Stallings, Sections 5.1 & 5.3

Basic Organization
Memory Cell Operation

- Represent two stable/semi-stable states representing 1 and 0
- Capable of being written to at least once
- Capable of being read multiple times

Semiconductor Memory Types

- Random Access Memory (RAM)
- Read Only Memory (ROM)
- Programmable Read Only Memory (PROM)
- Eraseable Programmable Read Only Memory (EPROM)
- Electronically Eraseable Programmable Read Only Memory (EEPROM)
- Flash Memory

Random Access Memory

- Misnomer (Last week we learned that the term Random Access Memory refers to accessing individual memory locations directly by address)
- RAM allows reading and writing (electrically) of data at the byte level
- Two types
  - Static RAM
  - Dynamic RAM
- Volatile

Read Only Memory (ROM)

- Sometimes can be erased for reprogramming, but might have odd requirements such as UV light or erasure only at the block level
- Sometimes require special device to program, i.e., processor can only read, not write
- Types
  - EPROM
  - EEPROM
  - Custom Masked ROM
  - OTPROM
  - FLASH

ROM Uses

- Permanent storage – nonvolatile
- Microprogramming
- Library subroutines
- Systems programs (BIOS)
- Function tables
- Embedded system code

EPROM

- Written to only with a programmer.
- Erased with ultraviolet light
- Positive
  - non-volatile storage without battery
  - can write to it, but only with aid of programmer
- Negative
  - programmer requirements
  - Expensive
  - locations must be erased before writing

EEPROM

- Written to with either programmer or the processor (electrically)
- Erased with either a programmer or the processor (byte-by-byte electrically)
- Positive
  - non-volatile memory without batteries
  - programmable a single-location at a time
- Negative
  - Expensive
  - only smaller sizes available
  - extremely slow write times (10 mS vs. 100 to 200 nS)

Custom masked ROM

- You send the ROM manufacturer your data and they mask it directly to the ROM
- Use only when you are selling large volume of a single product
- Positive
  - becomes cheaper to use for approximately more than 2000 parts
  - components come from chip manufacturer already programmed and tested taking out a manufacturing step
- Negative
  - costs several thousand dollars for custom mask
  - software changes are costly
  - cannot be reprogrammed

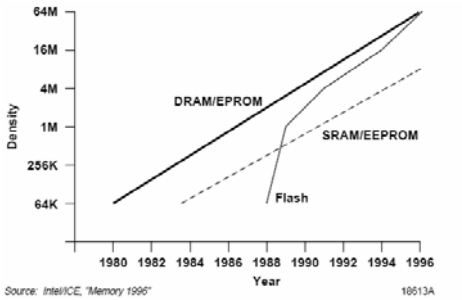
OTPROM

- Uses fuses that are burned to disconnect a logic 1 and turn it to a logic 0.
- Written to by you using a programmer similar to EPROM
- Once it's written to, the data is in there forever.
- Positive
  - cheaper than EPROM due to cheaper packaging
  - more packaging options than EPROM due to less constraints like erasure window
  - standard "off-the-shelf" component
  - cheaper than Custom masked ROM up to about 10,000 devices
- Negative – to reprogram, have to throw out the chip - Should only be used for stable design

FLASH

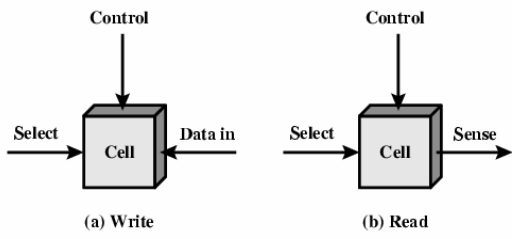
- These memories are basically EEPROMs except that erasure occurs at the block level in order to speed up the write process
- Nonvolatile
- This makes FLASH work like a fast, solid state hard drive
- Positive
  - non-volatile
  - higher desities than both SRAM and DRAM
- Negative
  - process of storing data is at a block level (and slower)
  - data cell must be erased before writing data to it

Flash Desity Comparison



Source: Griffin, J., Matas, B., de Suberbasaux, C., "Memory 1996" , Integrated Circuit Engineering Corporation, Scottsdale, AZ, on-line: http://smithsonianchips.si.edu/ice/cd/MEM96/TITLE.PDF

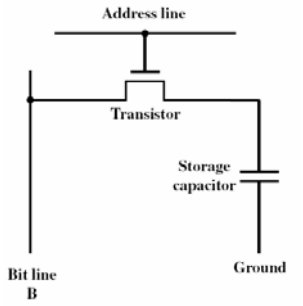
Memory Cell Operation



Dynamic RAM (DRAM)

- Bits stored as charge in capacitors
- Simpler construction
- Smaller per bit
- Less expensive
- Slower than SRAM (maintenance and read overhead explained later)
- Typical application is main memory
- Essentially analogue -- level of charge determines value

DRAM Structure (Figure 5.2 from textbook)



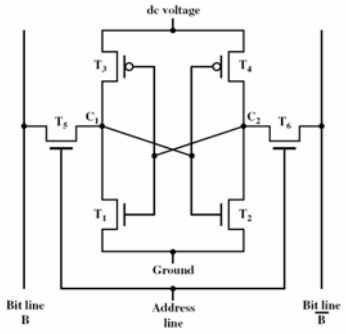
DRAM Operation (Figure 5.2a continued)

- Address line active when bit read or written
- Logic '1' closes transistor switch (i.e., current flows)
- Write
  - Voltage to bit line – High for 1 low for 0
  - Signal address line – Transfers charge to capacitor
- Read
  - Address line selected – transistor turns on
  - Charge from capacitor fed via bit line to sense amplifier
  - Compares with reference value to determine 0 or 1

Static RAM (SRAM)

- Essentially uses flip-flops to store charge (transistor circuit)
- As long as power is present, transistors do not lose charge (no refresh)
- Very fast (no sense circuitry to drive nor charge depletion)
- Can be battery-backed – A small battery is piggy-backed to the RAM chip an allows data to remain even when power is removed
- More complex construction
- Larger per bit
- More expensive
- Used for Cache RAM because of speed and no need for large volume

SRAM Operation (Figure 5.2b from textbook)



SRAM Operation (Figure 5.2b continued)

- Transistor arrangement gives stable logic state
- State 1
  - C1 high, C2 low
  - T1 T4 off, T2 T3 on
- State 0
  - C2 high, C1 low
  - T2 & T3 off, T1 & T4 on
- Address line transistors
  - T5 & T6 act as switches connecting cell
- Write – apply value to B & compliment to B
- Read – value is on line B

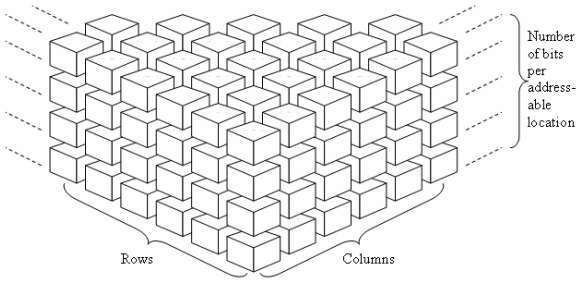
SRAM vs. DRAM

- Both volatile – Power needed to preserve data
- DRAM
  - Simpler to build, smaller
  - More dense
  - Less expensive
  - Needs refresh
  - Larger memory units
- SRAM
  - Faster
  - Used for cache

DRAM Organization Details (by example)

- A 16Mbit chip can be organised as a 2048 x 2048 x 4 bit array
- This arrangement reduces the number of address pins
- Multiplex row address and column address 11 pins to address (2^{11}=2048)
- Adding one more pin doubles range of values so multiplies capacity by four

DRAM Organization Details (continued)



DRAM Process

- Total number of address lines is half that of the total needed for the addressable locations
- A single addressable memory location has the address divided in half, e.g., the MSB half representing the row address and the LSB half representing the column address. This saves on pins.

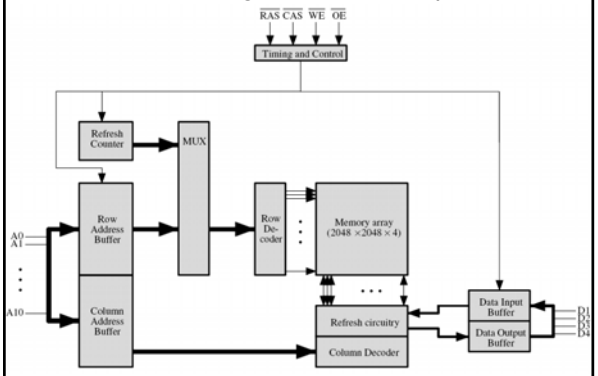
DRAM Process (continued)

- ^RAS (row address select) strobes the row address in to its buffer or latch while ^CAS (column address select) strobes the column address into its buffer or latch.
- Note: one more pin on the address quadruples the size of the matrix (doubles rows and doubles columns for an increase by factor of four)
- To make 16 bit wide data bus, you'll need four of these modules

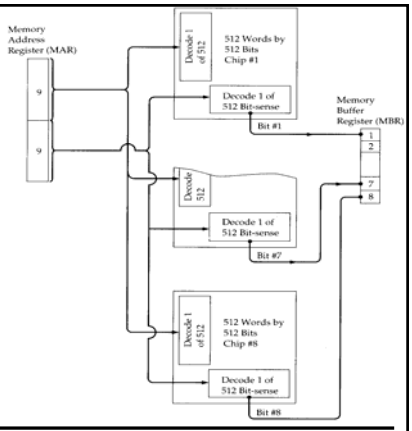
DRAM Refresh

- Two things discharge a DRAM capacitor
  - Data read
  - Leakage current
- Need refreshing even when powered
- Refresh circuit included on chip – Even with added cost, still cheaper than SRAM cost
- Refresh process involves disabling chip, then reading data and writing it back
- Performed by counting through "rows"
- Takes time – Slows down apparent performance

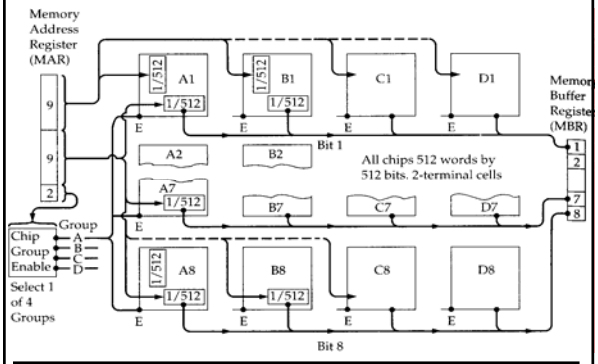
DRAM Organization Example



Module Organization: Using multiple memories in parallel to increase data bus width



Module Organization: Using chip selects to increase the number of words



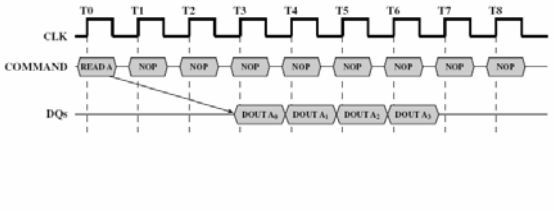
Advanced DRAM Organization

- SRAM Cache was the traditional way to improve performance of the DRAM
- Basic DRAM is unchanged since first RAM chips
- Enhanced DRAM
  - Contains small SRAM as well
  - SRAM acts as cache holding last line read
- Cache DRAM (CDRAM)
  - Larger SRAM added
  - Acts as either cache or serial buffer

Synchronous DRAM (SDRAM)

- Access is synchronized with an external clock
- Address is presented to RAM
- RAM finds data (CPU waits in conventional DRAM)
- Since SDRAM moves data in time with system clock, CPU knows when data will be ready
- CPU does not have to wait, it can do something else
- Burst mode allows SDRAM to set up stream of data and fire it out in block
- DDR-SDRAM sends data twice per clock cycle (leading & trailing edge)

SDRAM Sample Timing



RAMBUS or RDRAM

- Suggests transfer rates from 1.6 to 10.7 GBytes per second.
- Subsystem consists of the memory array, the RAM controller, and a well-defined bus
- Bus definition includes all components including the microprocessor and any other devices that may use it
- Vertical package (all pins on one side) called Rambus in-line memory modules (RIMMs)
- Adopted by Intel for Pentium & Itanium

Bus definition

- Data exchange over 28 wires
- Different definitions require bus lengths less than 12 cm long (some definitions are longer up to 25 cm long)
- Bus addresses up to 320 RDRAM chips
- Communication protocol is packet-based
- Implements pipelined operation overlapping command and data
- 800 to 1200 MHz operation
- Inititial access time = 480ns
- After that, 1.6 GBps