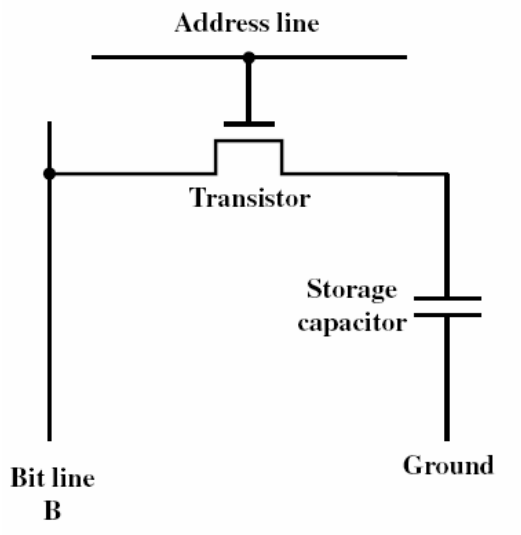
DYNAMIC RAM (DRAM)

POINTS OF INTEREST

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

SCHEMATIC



OPERATION

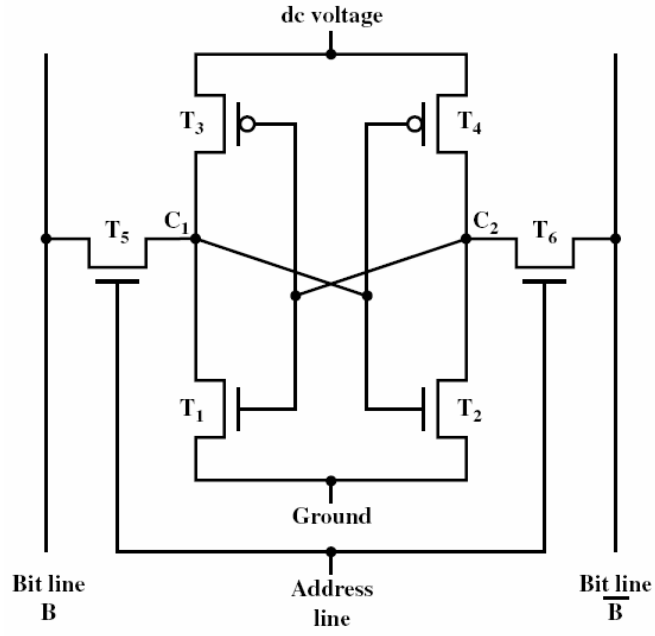
- Address line active when bit read or written
- Logic '1' closes transistor switch (i.e., current flows)
- Write
  - Voltage to bit line – High for 1 low for 0
  - Signal address line – Controls transfer of charge to capacitor
- Read
  - Address line selected – transistor turns on
  - Charge from capacitor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

STATIC RAM (SRAM)

POINTS OF INTEREST

- Uses latches 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 without power (Not possible with DRAM)
- More complex construction → larger and more expensive
- Used for Cache RAM because of speed and no need for large volume or high density

SCHEMATIC



OPERATION

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

DRAM ORGANIZATION

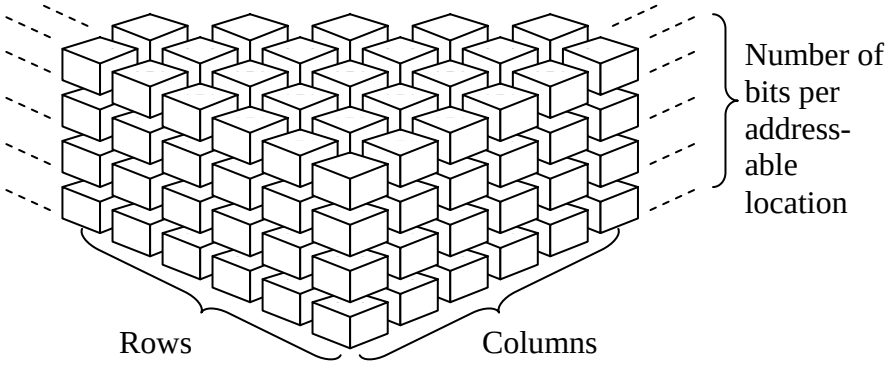
ADDRESSING EXAMPLE: Assume we have a memory device with 32 pins: 8 data, 1 chip select, 1 write enable, 1 read enable, 1 power, 1 electrical ground. That leaves 32 – 13 = 19 pins for addressing giving us 2^{19} = 512 K addresses or memory locations. Not very much.

What if we gave the memory device its address in two stages? During the first stage, the upper half of the address is presented and stored in the memory device. The second stage presents the lower half of the address is presented and the memory device can return the data.

Address lines:	Address (upper half)	Address (lower half)	
Data lines:	**No valid data**	**No valid data**	RAM returns data

This could effectively double the bits of our address. The above example now becomes 19 × 2 = 38 address bits giving us a memory space of 2^{38} = 256 Gig.

DRAM ADDRESSING 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.
- $\overline{\text{RAS}}$ (row address select) strobes the row address in to its buffer or latch.
- $\overline{\text{CAS}}$ (column address select) strobes the column address into its buffer or latch.

RETURN TO ADDRESSING EXAMPLE: This means that two additional pins are taken away from our earlier example leaving us with 17 pins → 17 × 2 = 34 address bits giving us a memory space of 2^{34} = 16 Gig.

- 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)
- If DRAM has 4-bit data bus, to make 16 bit wide data bus, need four modules wired in parallel on the bus.

YEAH, BUT… doesn't it slow up the memory process by having two cycles?

FAST PAGE MODE (FPM)

If the processor needs a block from memory, the first half of the address should be the same for all addresses, right? In this case, start the memory process with the first half of the address, then only use the last half for subsequent retrievals. For example, FPM for a memory block of size 4 would look like the following.

Address:	Hi Half Blck Addr	Lo Half Addr. 0		Lo Half Addr. 1		Lo Half Addr. 2		Lo Half Addr. 3	
Data:	**No valid data**	**No valid data**	Data Word 0	**No valid data**	Data Word 1	**No valid data**	Data Word 2	**No valid data**	Data Word 3

The data of a single row is referred to as a "page".

EXTENDED DATA-OUT (EDO)

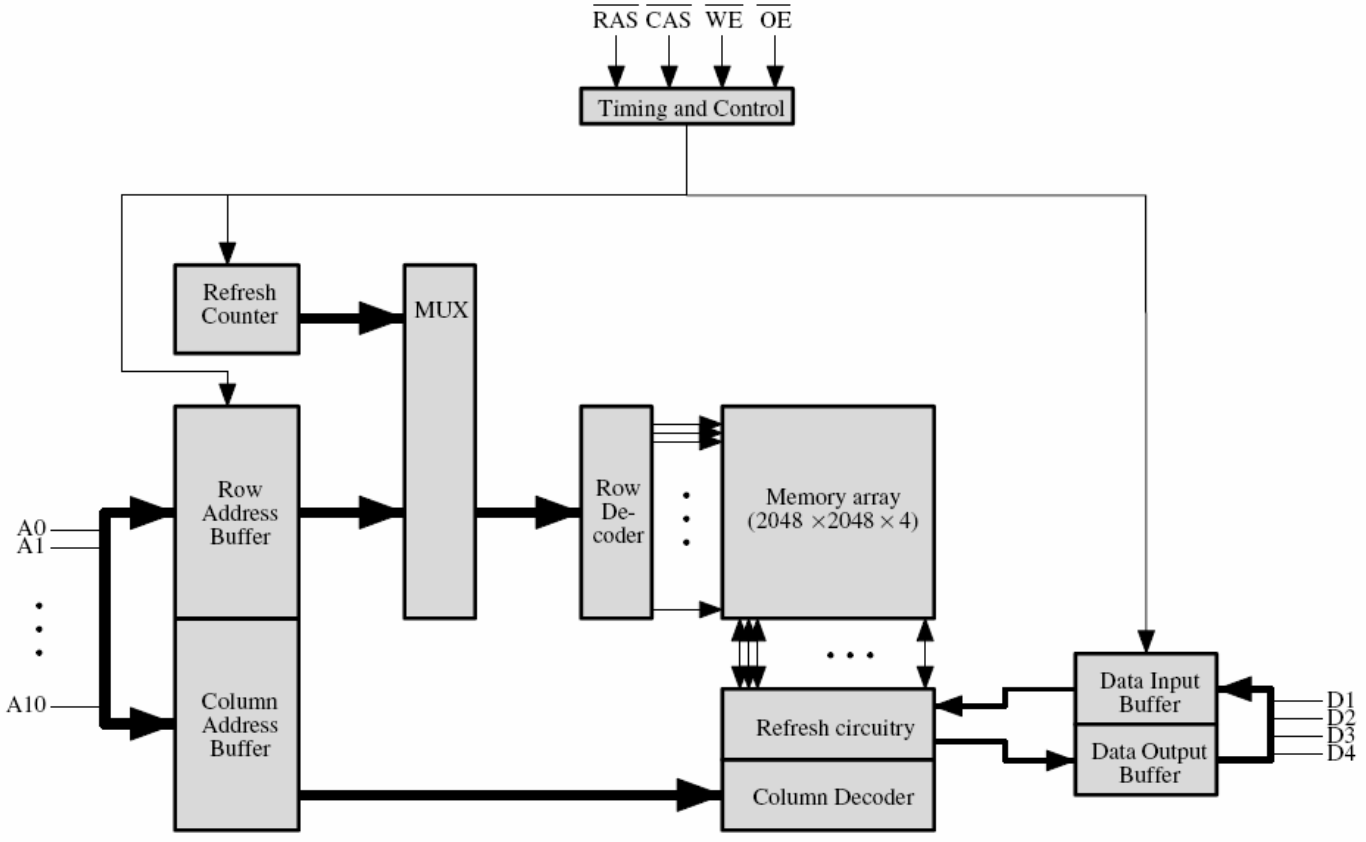
EDO allows the processor to overlap the data read cycle with the write for the next column address.

Address:	Hi Half Block Addr.	Lo Half Addr. 0	Lo Half Addr. 1	Lo Half Addr. 2	Lo Half Addr. 3	
Data:	**No valid data**	**No valid data**	Data Word 0	Data Word 1	Data Word 2	Data Word 3

EDO results in a savings of approximately 10 ns for each read within a single page.

REFRESHING A DRAM

- Two things discharge a DRAM capacitor
  - Data read
  - Leakage current
- Need refreshing even when powered and idle (once every few milliseconds)
- 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



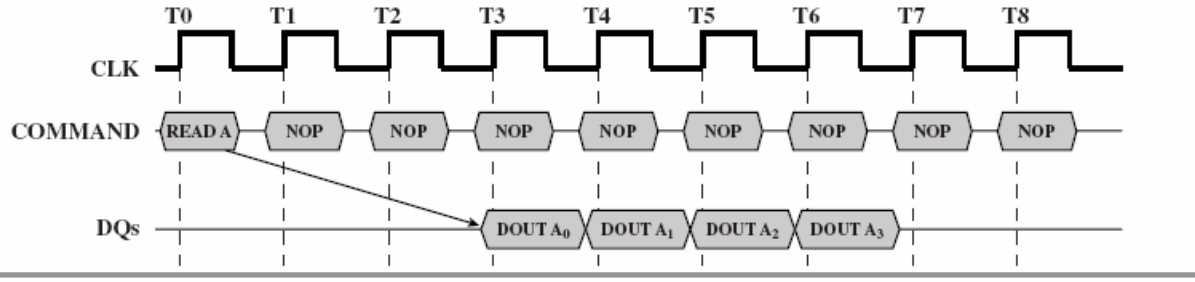
SPEEDING UP A DRAM

Basic DRAM is unchanged since first RAM chips, so alternate techniques are used to improve performance.

- Use FPM or EDO to improved access time
- Can speed effective access time of by adding a cache between processor and DRAM (CACHE DRAM or CDRAM)

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)



RAMBUS or RDRAM

Characteristics:

- 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 Suggests transfer rates from 1.6 to 10.7 GBytes per second.

DON'T FORGET CHIP SELECTS TO ALLOW FOR MULTIPLE MEMORIES