Memory

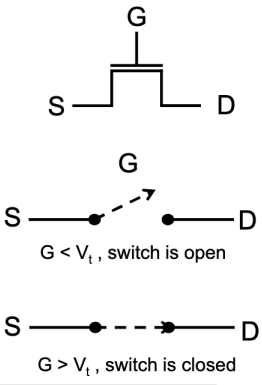
- Address n bits, 2^n locations in memory
- Memory stores m bits, m * 2^n storage

RAM

- fast, limited, volatile, no program code
  - SRAM (static - retains value as long as have power)
  - DRAM (dynamic - capacitor; slowly lose info unless refreshed)

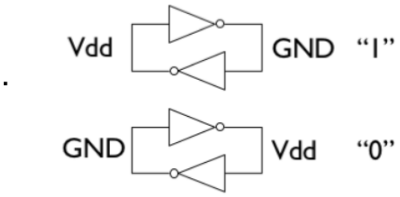
MOSFET

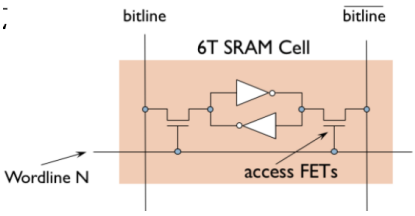
- 3 terminals: Source, Gate, Drain
  - Gate voltage $> V_t$, Source connected to Drain
  - Gate voltage $\leq V_t$, transistor is open switch



SRAM Cell

- 2 inverters connected in a loop
  - 2 access transistors control reading and writing the cell
  - `WL` is enable
    - READ `WL = 1`, data in cell will appear on bitline `BL`
    - WRITE `WL = 1`, drive BL and $\overline{BL}$ to strong values, when WL → 0, value retained

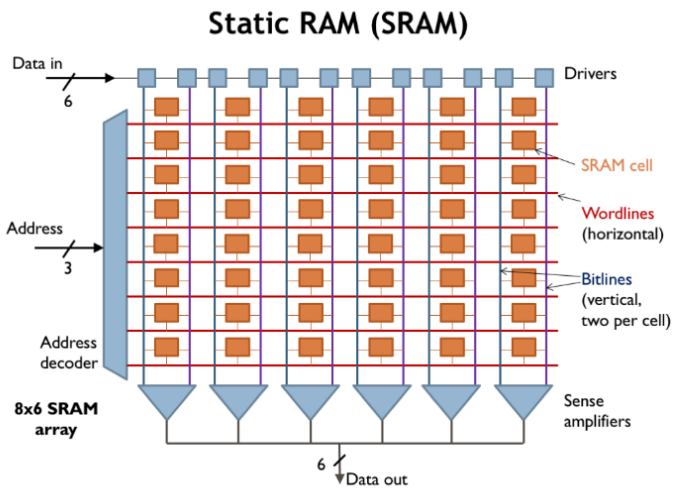




SRAM Organization

- SRAM with 8 locations, 6 bits each
- 3 bit address one-hot decoded to 8 bit values
  - 1 wordline is selected
- Drivers (top) push current on BL and NBL

- sense amplifier at bottom boosts weak current from each cell



DRAM

- stores value using 1C
  - sufficient charge = 1, else 0
  - 1 access transistor
- WRITE - set BL to Vdd/GND, activate WL - charge/discharge the capacitor
- READ - pre-charge BL to **Vdd/2** and activate WL
  - if C stored Vdd, BL voltage increase
  - If C stored GND, BL voltage decrease
  - sense amplifier boosts this small difference

DRAM READ

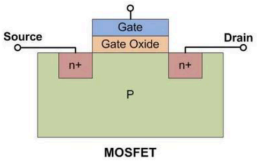
- destructive
  - needs write back after read
- DRAM needs recharge (cells lose charge overtime)
  - DRAM self refresh - re-write values to capacitors so they don't lose all their charge
  - few ms
- additional circuitry for *self refresh* and *writing back read values*
- 20x denser than SRAM but 10x slower

ROM

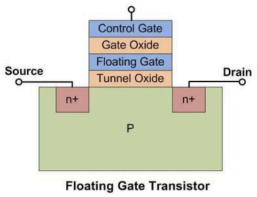
- meant for program code - shouldn't change once dev in use
  - used to be program once by manufacturer and cannot change

Flash Cell

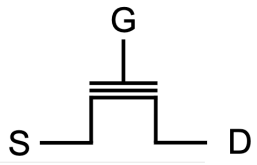
- FGFET (floating-gate transistor)
  - like MOSFET, but with 2 gates
- floating gate layer is electrically insulated
  - any electron on this gate are trapped; charge takes month to leak
- floating gate blocks usual FET operation
  - if charge exists on FG, prevents current flow from Src to Drain even if control gate voltage $> V_t$



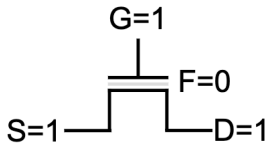
MOSFET

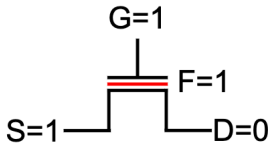


Floating Gate Transistor



- A charged floating gate (electrons trapped) represents one binary state (typically a "0")
- An uncharged floating gate represents the other state (typically a "1")
- READ
  - set both Src and Gate to Vdd/1
    - no charge on FG, regular MOSFET and read 1
    - yes charge on FG, read 0 at drain
    - *cells with charge store 0*
- Changing val of cell
  - to drain a charged cell (i.e. change a 0 to a 1), need to erase the cell
  - process is slow





Flash Blocks & Wearout

- to speed up erasing, cells are grouped into blocks, erased together
  - block size 8KB ~ 256KB
  - when we do write, single bit 0→1, we must erase the **entire block** and write it all again ?????
- flash memory suffers from wearout if used too much
  - write to a cell *too many* times, it no longer stores charge

NAND NOR Flash

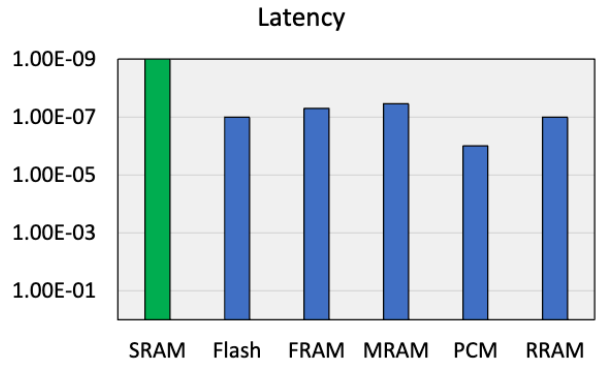
- 2 ways to arrange FGFET
  - connect in **parallel NOR flash**
    - supports fast read, slow writes
    - byte-granularity accesses
    - good for code (supports execute in place)
  - connect in **series NAND flash**
    - slow reads, fast writes

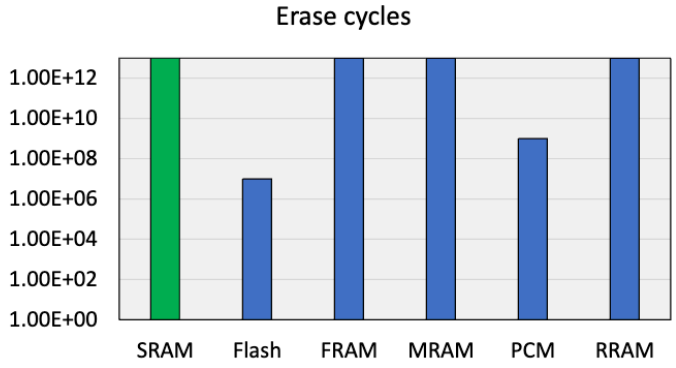
- higher density than NOR
- better for storage (USB drives)

	NOR	NAND
Read	0.1 uS	30 uS
Erase	500 ms	3.5 ms
Block size	64KB to 256KB	8KB to 32KB
Bad blocks	0%	~1-2%
Capacity	64MB - 2GB	1GB - 16GB
Cost	$$	$
Erase cycles	10^4 - 10^5	10^5 - 10^6

Other ROM

- non volatile memory
  - forroelectric ram, magnestoresistive RAM, PCM...





Memory	Location	Latency	Small	Medium	Large
SRAM	On-chip	1 – 10 ns	1 KB – 256 KB	256 KB – 1 MB	1 MB – 16 MB
NOR Flash	On-chip	100 ns			
NOR Flash	Off-chip	250 ns			
NAND Flash	Off-chip	10,000 ns	2 MB – 16 MB	16 MB - 256 MB	256 MB – 64 GB
DRAM	Off-chip	100 ns	-	32 MB – 1 GB	1 GB – 16 GB