10.1 Introduction

◆ Design issue
  - Area efficiency $\rightarrow$ cost per bit
  - Access time $\rightarrow$ speed
  - Power consumption $\rightarrow$ low-power
10.1 Introduction

**Semiconductor memory types**

- **Semiconductor Memories**
  - **Volatile**
    - Dynamic RAM (DRAM)
    - Static RAM (SRAM)
  - **Non-volatile**
    - Mask (Fuse) ROM
    - Programmable ROM (PROM)
      - Erasable PROM (EPROM)
      - Electrically Erasable PROM (EEPROM)
      - Flash Memory
      - Ferroelectric RAM (FRAM)
      - Magnetoresistive RAM (MRAM)
      - Resistive RAM (RRAM)
      - Phase-change RAM (PCRAM)
      - Spin Torque Transfer RAM (STT)
# Characteristic Summary of Memory Devices

<table>
<thead>
<tr>
<th>Memory type</th>
<th>DRAM</th>
<th>SRAM</th>
<th>UV EPROM</th>
<th>EEPROM</th>
<th>Flash</th>
<th>FRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data volatility</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Data refresh operation</td>
<td>Required</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Cell structure</td>
<td>1T-1C</td>
<td>6T</td>
<td>1T</td>
<td>2T</td>
<td>1T</td>
<td>1T-1C</td>
</tr>
<tr>
<td>Cell size ($F^2$) (F: min. feature size)</td>
<td>6~8</td>
<td>80~100</td>
<td></td>
<td></td>
<td>4~5(NAND)</td>
<td>9~10(NOR)</td>
</tr>
<tr>
<td>Cell density</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Power consumption</td>
<td>High</td>
<td>High/low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Read speed (latency)</td>
<td>~50 ns</td>
<td>~10/70 ns</td>
<td>~50 ns</td>
<td>~50 ns</td>
<td>~50 ns</td>
<td>~100 ns</td>
</tr>
<tr>
<td>Write speed</td>
<td>~40 ns</td>
<td>~5/40 ns</td>
<td>~10 μs</td>
<td>~5 ms</td>
<td>~(10 μs-1 ms)</td>
<td>~100 ns</td>
</tr>
<tr>
<td>Endurance</td>
<td>High</td>
<td>High</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Cost</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>In-system writability</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Power supply</td>
<td>Single</td>
<td>Single</td>
<td>Single</td>
<td>Multiple</td>
<td>Single</td>
<td>Single</td>
</tr>
<tr>
<td>Application example</td>
<td>Main memory</td>
<td>Cache/PDAs</td>
<td>Game machines</td>
<td>ID card</td>
<td>Memory card</td>
<td>solid-state disk</td>
</tr>
</tbody>
</table>

Equivalent Circuits of Memory Cells (1)

(a) DRAM

(b) SRAM
Equivalent Circuits of Memory Cells(2)

(c) Mask ROM        (d) EPROM             (e) FRAM
Conceptual RAM Array Organization

M columns
N rows
1 Kb memory = 1024b = 2^10
M = 5 6 4
N = 5 4 6
10.2 Dynamic Random Access Memory

Typical configuration of DRAM chip
(1.6Gbps 4Gb 30nm LPDDR3 w/ 8 banks)

chip size $\leftrightarrow$ performance

The number of cells per word and bit lines

Pin assignment
## Definition and Function of DRAM Pins

<table>
<thead>
<tr>
<th>Pin name</th>
<th>Definition</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLK</td>
<td>Clock input</td>
<td>Reference system clock for the operation and data communication</td>
</tr>
<tr>
<td>CKE</td>
<td>Clock Enable</td>
<td>Control the clock input</td>
</tr>
<tr>
<td>CS</td>
<td>Chip Select</td>
<td>Activate the DRAM device from a memory cluster</td>
</tr>
<tr>
<td>RAS</td>
<td>Row Address Strobe</td>
<td>Latch row address and start the cell core operation</td>
</tr>
<tr>
<td>CAS</td>
<td>Column Address Strobe</td>
<td>Latch column address and start the data communication operation</td>
</tr>
<tr>
<td>WE</td>
<td>Write Enable</td>
<td>Activate the write operation</td>
</tr>
<tr>
<td>A0 to A14</td>
<td>Address input</td>
<td>Select a data bit</td>
</tr>
<tr>
<td>DQ0 to DQ15</td>
<td>Data input and output</td>
<td>Communicate data with external devices</td>
</tr>
<tr>
<td>DQMU/DQML</td>
<td>DQ Mask for Upper (Lower) Byte</td>
<td>Mask byte data from the operations</td>
</tr>
<tr>
<td>$V_{DD}/V_{SS}$</td>
<td>Power pins</td>
<td>Power for DRAM core and peripheral circuits</td>
</tr>
<tr>
<td>$V_{DDQ}/V_{SSQ}$</td>
<td>Power pins</td>
<td>Power for DQ circuits</td>
</tr>
<tr>
<td>NC</td>
<td>No connection</td>
<td></td>
</tr>
</tbody>
</table>
Historical Evolution of DRAM Cell(1)

- **Four-transistor DRAM cell**
  - Operations are similar to SRAM cell
  - Two storage nodes
  - Periodically refresh is required
  - Non-destructive read operation

- **Three-transistor DRAM cell**
  - One storage node
  - One Tr. each for "read" and "write"
  - Non-destructive read operation
  - Two bit lines and two word lines
    - (additional contacts → increase area)
Historical Evolution of DRAM Cell(2)

- **Two-transistor DRAM cell**
  - Explicit storage cap.
  - Destructive read operation (share with the bit line)
  - Two bit lines and one word line

- **One-transistor DRAM cell**
  - Industry-standard DRAM cell
  - Destructive read operation (share with the bit line)
  - Charge restoring operation required
DRAM Cell Types

- With only one transistor and one capacitor
  - Smallest area of the all DRAM cells.
  - Destructive “read” operation
    - major effort: large cap. cell with minimized area

(a) DRAM cell with a stacked cap.
(b) DRAM cell with a trench cap.
Operation of Three-Transistor DRAM Cell

- Typical 3-T DRAM cell and voltage waveforms
Precharge Events

Base on two-phase non-overlapping clock scheme

\( \Phi_1 \): precharge phase
\( \Phi_2 \): active phase

Precharge signal PC goes up
MP1 and MP2 are activated
\( C_2 \) and \( C_3 \) are charged up
(Steady-state values)
Write “1” and Read “1” Operations

- **Write “1”**
  - DATA is low → $D_{in}$ remains high
  - M1 turns on (WS is high) → $C_2$ shared with $C_1$
  - $C_1$ charge up to high (M2 is conducting)

- **Read “1”**
  - RS is high → M3 turns on
  - M2 and M3 create conducting path from $C_3$ to GND
  - $C_3$ discharges
  - Non-destructive read operation

\[
\frac{1}{2} V_{DD} = 0.5
\]

\[
0.5 \times 100 C_1 + 0.990 C_1 = 0.490 C_1
\]

\[
\frac{100 C_1}{101 C_1} = \frac{100}{101} \Rightarrow V = 5 \text{ mV}
\]
Write “0” and Read “0” Operations

- **Write “0”**
  - DATA is high \(\rightarrow\) \(D_{in}\) goes low
  - M1 turns on (WS is high)
  - \(C_1\) discharges
    - (M2 turns off)

- **Read “0”**
  - RS is high \(\rightarrow\) M3 turns on
  - No conducting path
  - \(C_3\) does not discharge
Operation of One-Transistor DRAM Cell

- One explicit storage cap. and one access transistor
  - The most widely used storage structure
  - Bit lines are folded and precharged to half-$V_{DD}$
    - Improve noise-immunity & reduce power consumption
  - Operation: “read”, “write”, “refresh”
1-T DRAM Structure

- DRAM cell array with control circuits
- Latch amplifier to sense the small signal difference
- Bit lines and sensing nodes set to half-$V_{DD}$ through equalizer
DRAM Read Operation

- $C_S$ shared with $C_{BL}$ (=initially half $V_{DD}$)
  \[ \Delta V = \frac{C_S}{C_{BL} + C_S} \frac{V_{DD}}{2} \]

- $C_S : V_{DD} \rightarrow \frac{1}{2}V_{DD} + \Delta V$ (destructive)
- BL and BLB voltage difference amplified
- BLB $\rightarrow$ GND, BL $\rightarrow$ $V_{DD}$
  storage node is recovered (restoring)
- Column switch is enabled by column decoder (BL $\rightarrow$ BL_IO, BLB $\rightarrow$ BL_IOB)
- Read Amp. amplifies the voltage difference
  - $V_{PP} = V_{DD} + V_{th}$ for full charge restoration
  - PSA and PSAB are sequentially activated to reduce charge injection and short circuit current

DRAM Write Operation

- Identical sequence to normal read operation
- Strong write driver (buffer) to drive BL_IO, BL_IOB line cap. faster than read operation
- Column switch is selected by column decoder
- Bit line and cell data changed
Asynchronous DRAM Mode(1)

- single bit access (different row and column addresses)
- Operation uses address multiplexing scheme ($\overline{RAS}$ and $\overline{CAS}$)
  - reduce the chip package size
- $\overline{RAS}$ pull down $\rightarrow$ operation start
- Falling edge of $\overline{CAS}$ $\rightarrow$ data (from same word line) selected
- $\overline{RAS}$, $\overline{CAS}$ precharge before new data access
- $t_{RAC}$: memory read latency, time to read data from falling of $\overline{RAS}$
- Length of word line is determined by refresh cycle constraint
Asynchronous DRAM Mode (2)

- **page access**
  - keep the row address
  - read cell of same row address
  - faster read operation

- **extended data-out (EDO)**
  - new column address is captured at rising edge of CAS
  - Read data maintain during precharge time
  - Fastest read operation
Synchronous DRAM Mode

- **Four bit burst read**
  - Read frequency improve with use of the system clock
  - At falling edge, control signal and addresses become active
  - Pipelined based on clock to improve throughput
  - Use both of edges to improve bandwidth (Dual Data Rate)

- **Serial mode read**
  - Use small signal swing and clock recovery scheme to maximize the frequency
  - Send control input as packet
  - Send out data in a serial form
Leakage Currents in DRAM Cells

- Contact and bit line share by two adjacent cells

\[ I_{\text{leakage}} = I_{\text{sub}} + I_{\text{tunneling}} + I_j + I_{\text{cell-to-cell}} \]

- \( I_{\text{sub}} \): leakage through cell transistor
- \( I_{\text{tunneling}} \): tunneling through thin dielectric
- \( I_j \): junction leakage at storage node
- \( I_{\text{cell-to-cell}} \): leakage across the field oxide

- \( I_{\text{sub}} \) depend on \( V_{\text{th}} \)
- Increase \( V_{SB} \) to reduce \( I_{\text{sub}} \)

- \( I_{\text{tunneling}} \) is a serious issue because thickness of dielectric is reduced to increase cell cap.
Refresh Operation

- **ROR (RAS-only refresh) refresh**
  - Read and restore operation
  - Does not send data out
  - similar to normal read operation

- **CBR (CAS-before-RAS) refresh**
  - row address generated by on-chip counter
  - performed periodically

- **Self refresh**
  - period set according to operating condition
  - row address and control signal generated by internal circuit
Logic level of system board and memory chip are different required to convert logic levels input/output buffers

Input buffers

-inverter type
-latch type
-differential amp type
## Characteristic Comparison of Input Buffers

<table>
<thead>
<tr>
<th></th>
<th>Buffer type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inverter</td>
</tr>
<tr>
<td>Logic threshold determination ($V_{IH}$ and $V_{IL}$)</td>
<td>By $W_P/W_N$ ratio</td>
</tr>
<tr>
<td>Speed</td>
<td>Slow</td>
</tr>
<tr>
<td>Standby current</td>
<td>Small</td>
</tr>
<tr>
<td>Sensitivity to $V_{DD}$ and temperature</td>
<td>Large</td>
</tr>
<tr>
<td>Noise immunity</td>
<td>Bad</td>
</tr>
<tr>
<td>Constraint</td>
<td>None</td>
</tr>
</tbody>
</table>
DRAM Input/Output Circuits(2)

- Memory output buffers
  - Need to drive large cap.
  - Keep a high-impedance when chip is not selected
    ➡️ to prevent interference of output

PMOS pull-up structure

NMOS pull-up structure
To select cell from $2^{2M}$ memory array, M address bits are needed.

Practically, M transistors in series is impossible.

Decoding scheme is composed of pre and main decoder.
DRAM Decoder (2)

- $V_{PP}$ for full restoration
- Output of predecoder boosted by level shifter
- Self-bootstrapped driver for transferring to highly cap. without signal degradation

- Voltage of C when main decoder is selected

$$V_C = V_{PP} + \Delta V = V_{DD} - V_{TN} + \frac{C_{MN2}}{C_{MN2} + C_{Parasitic}} V_{PP}$$
Voltage Sense Amplifiers

- To detect signal difference on data lines
  - Current-mirror differential
    - Popular and good common-mode rejection ratio
    - Large area and large power consumption
  - Full CMOS latch type
    - High speed, small area and low power
    - Precharge signal required
    - Operation cannot be reversed
  - Semilatch type
    - Between current-mirror type and full CMOS latch type
Internal Voltage Regulator Circuit

- Lowering voltage to reduce power consumption
- $V_{\text{INT}}$ (internal voltage generator)
  - reduce operating current
Half $V_{DD}$ Voltage Generator

- Folded bit line structure with half $V_{DD}$ sensing scheme
  - Improve noise immunity and low power consumption
  - Reduce electric field across thin dielectric

![Bias circuit and driver circuit](image)

Simulated output waveforms
Negative Substrate Bias Voltage Generator

- Subthreshold current is major source of charge decay
  - Negative voltage substrate $\rightarrow$ increase threshold voltage
  - $\rightarrow$ reduce load cap. of bit line

\[ \frac{\Delta V_T}{\Delta V_{SB}} \propto \sqrt{V_{SB}} \]
Negative Substrate Bias Voltage Generator(2)

Circuit diagram

Timing diagram

Simulated waveforms
10.3 Static Random Access Memory (1)

- Stored data can be retained indefinitely
- Simple latch with two stable operating points
- Two access switches to connect 1-bit SRAM
- Poly resistor load inverter structure is more compact cell size (resistor stack on top of cell)
- Load R trade off: low power ↔ wider noise margin, high speed

Symbols:
- Symbolic representation
- Generic topology of SRAM
- Resistive-load SRAM
10.3 Static Random Access Memory (2)

- **Depletion-load NMOS SRAM**
  - Six-transistor
    (one poly and one metal layer)
  - Cell size relatively small
  - Static characteristics and noise margins better than resistive-load cell
  - Static power consumption

- **Full CMOS SRAM**
  - Most popular
  - Lowest static power
  - Superior noise margins and switching speed
Full CMOS SRAM Cell(1)

- Very small static power dissipation (limited by leakage current)
- High noise immunity (large noise margin)
- Ability to operate at lower supply
- Disadvantage: cell area slightly larger, latch-up phenomena
Full CMOS SRAM Cell(2)

Layout of CMOS SRAM cell

Layout of a 4-bit X 4bit SRAM array, consisting of 16 CMOS SRAM cells
Two basic requirements which dictate W/L ratio

- Non-destructive data read operation
- Modify stored data during write phase

Read Operation (0 stored)

- M3 and M1 conduct some current
- $V_{C_C}$ drops slightly and $V_1$ increases
- $V_{1,max} \leq V_{T,2}$ not to turn on M2
- M3 in Saturation and M1 in linear

\[
\frac{k_{n,3}}{2} (V_{DD} - V_1 - V_{T,n})^2 = \frac{k_{n,1}}{2} (2(V_{DD} - V_{T,n})V_1 - V_1^2)
\]

\[
k_{n,3} < \left(\frac{W}{L}\right)_3 < \frac{2(V_{DD} - 1.5V_{T,n})V_{T,n}}{(V_{DD} - 2V_{T,n})^2}
\]
Write 0 operation (initially, 1 was stored at node 1)

- $V_1$ must be reduced below $V_{T,2}$ → M2 turns off, V2 rises and V1 falls
- When $V_1 = V_{T,n}$, M3 in linear & M5 in saturation

$$\frac{k_{p,5}}{k_{n,3}} < \frac{2(V_{DD} - 1.5V_{T,n})V_{T,n}}{(V_{DD} + V_{T,p})^2}$$

$$\frac{W}{L} < \frac{\mu_n}{\mu_p} \cdot \frac{2(V_{DD} - 1.5V_{T,n})V_{T,n}}{(V_{DD} + V_{T,p})^2}$$
Memory Structure of SRAM

- Word line selected by row address
- Cell data kept during read operation
- Boosted voltage not required
- Address multiplexing scheme is not used (fast access time than DRAM)
- Depend on applications
  - ultra low power: load transistor turns off during read operation
  - high speed: remains on
Operation of SRAM

- **Read operation**
  - Word line enable
  - One bit line discharge
    (voltage change of bit line is very small)
  - Sense amp. detect the voltage difference on bit line
  - Multi-stage amp. is used to improve read speed

- **Write operation**
  - Word line selected by row address
  - Write buffer write data into cell
  - Write buffer has larger current driving capability than cell
  - Write is faster than read
SRAM Read Operation

- TTL level converts into CMOS level signal
- Internal voltage regulator to reduce power dissipation to improve reliability
Leakage Currents in SRAM Cells

- Major portion of standby current
- Standby power is key parameter for low power design
- High threshold

Reduction of leakage ↔ degradation of performance

\[ I_j : \text{junction current} \]
\[ \text{data "1" to substrate} \]
\[ I_{\text{nsub}} \text{ and } I_{\text{psub}} : \text{subthreshold leakage} \]
\[ \text{turn off NMOS and PMOS} \]
\[ I_{\text{tunneling}} : \text{tunneling current} \]
\[ \text{cross thin gate oxide} \]
SRAM Read/Write Circuits

- Current-mode sense amp widely used in SRAM
  - improve signal sensing speed independent of bit line cap.
- Signal line connect to source of latch transistor
- Current difference appears on DL and DL
- Open-loop gain
  \[ \text{Gain}_{\text{open-loop}} = \frac{g_m(m3) \times g_m(m4)}{g_m(m1) \times g_m(m2)} \]
- Current-mode sense amp: Drawback- larger power consumption
SRAM Cell at Low Supply Voltage

◆ SRAM cell susceptible to variabilities
  ▪ Due to minimum device size to minimize area
  ▪ Threshold voltage variation covered in Ch. 3 plus layout induced threshold voltage variation
  ▪ PMOS pair (M5, M6) in SRAM cell- different $V_T$ due to NBTI
  ▪ NMOS pair (M1, M2) in SRAM cell- different $V_T$ due to PBTI

◆ Static noise margin (SNM)
  ▪ A noise tolerant voltage before the stored data flip
  ▪ Equivalent ckt to measure SNM
  ▪ 6-T SRAM cell at low supply voltage degrades SNM

$V_n$: DC noise, SNM: min. DC noise which flips the state of SRAM cell during read operation
SNM Variation due to DC Noise

- How to measure SNM graphically

- SNM: The length of side of the smaller nested square in the two openings of butterfly curve
- Before two $V_n$ s are fed: SNM=$V_S$
- After $V_n$ s are fed: stable point A and unstable point B meets at D
- More $V_n$ s are applied: one common point C & the stored bits are flipped
SRAM Cell Writability

★ Write-trip point
  - A metric for writability
  - Max. bit line voltage to flip the state of the SRAM cell
  - Primarily determined by the pull-up ratio of SRAM cell
  (Ex: \( \frac{W}{L}_{5}/\frac{W}{L}_{3} \))

★ Variability tolerant 6T SRAM cell
  - Trade off bw. read stability and writability
  - \( M_3 \) & \( M_4 \) \( \uparrow \): SNM \( \uparrow \), writability \( \downarrow \)
8T SRAM Cell

- No secondary power supplies
- Decouples the SRAM cell nodes from the bit line which enables balancing the read & write modes
- Read operation doesn’t affect the stored data
- 6T cell has the worst SNM in read operation where the pass gate transistor increases the voltage at the ‘0’ stored
10.4 Nonvolatile Memory

- Simple combinational Boolean network
- Only one word line selected at a time
- Active transistors exist at cross point
- Dynamic ROM
  - use periodic precharge signal to reduce static power
Layout of NOR ROM Array(1)

- Initially, NMOS at every row-column intersection
- ‘1’-bits are realized by omitting drain or source connection or gate electrode of corresponding NMOS

Layout example of a NOR ROM array
In reality, metal column lines laid out directly on top of diffusion column to reduce horizontal dimension.

Layout of the 4-bit X 4-bit NOR ROM array (pp. 46)
Implant-mask Programmable NOR ROM

- Every two rows share a common ground connection
- Every metal to diffusion contact shared by two adjacent devices
4-bit x 4-bit NOR ROM Array

- Based on implant-mask programming
- Raised threshold voltage $> V_{OH}$ → "1"-bit
- Non-implanted → "0"-bit
- Higher core density (smaller silicon area per stored bit)
4-bit x 4-bit NAND ROM Array

- Bit line: depletion-load NAND gate
- Deactivated transistor $\Rightarrow$ "1"-bit
- Shorted or on transistor $\Rightarrow$ "0"-bit
Implant-mask layout of NAND ROM

- Lowered threshold voltage \(< 0\text{V}\) → “0”-bit
- Much more compact than NOR ROM
- Access time is slower than NOR ROM
Design of Row and Column Decoders (1)

- Select a particular memory location in array
- Row address decoder example

<table>
<thead>
<tr>
<th>$A_1$</th>
<th>$A_2$</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$R_3$</th>
<th>$R_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Design of Row and Column Decoders (2)

- ROM array and row decoder (two adjacent NOR arrays)
Row Decoder for NAND ROM

- Lower voltage for logic "0"
- Realized using same layout strategy as memory array
Column Decoder(1)

- Using NOR address decoder and NMOS pass transistor
- Only one pass transistor turned on at a time
- $2^M(M+1)$ transistors required
Column Decoder(2)

- Binary selection tree decoder
- NOR address decoder not needed
  - Reduce the number of transistors significantly
- But, long data access time
Example 10.1(1)

- Analyze the access time of a 32-kbit NOR ROM array

\[ \mu_n C_{ox} = 20 \mu A / V^2 \]

\[ C_{ox} = 3.47 \mu F / cm^2 \]

*Poly sheet resistance = 20Ω/square*
Example 10.1(2)

- Assume 7 row address bits and 8 column address bits (128 rows and 256 columns)
- Calculate row resistance and capacitance

![Diagram of row circuit]

\[ C_{\text{row}} = C_{\text{ox}} \cdot W \cdot L = 10.4 \, fF / \text{bit} \]
\[ R_{\text{row}} = (\# \, \text{of} \, \text{squares}) \times (\text{Poly sheet resistance}) = 60 \, \Omega / \text{bit} \]

- Calculate row access time

\[ t_{\text{row}} \approx 0.38 \cdot R_{T} \cdot C_{T} = 15.53 \, \text{ns} \]

\[ R_{T} = \sum_{\text{all columns}} R_{i} = 15.36 \, k\Omega \]
\[ C_{T} = \sum_{\text{all columns}} C_{i} = 2.66 \, pF \]
Example 10.1(3)

- A more accurate delay: Elmore time constant for RC ladder circuits

\[ t_{\text{row}} = \sum_{k=1}^{256} R_{jk} C_k = 20.52 \text{ns} \quad \text{where} \quad R_{jk} = \sum_{j=1}^{k} R_j \]

- Calculate column access time

\[ C_{\text{column}} = 128 \times (C_{\text{gd,driver}} + C_{\text{db,driver}}) \approx 1.5 \text{pF} \]

where \( C_{\text{gd,driver}} + C_{\text{db,driver}} = 0.0118 \text{pF} / \text{word line} \)
Example 10.1(4)

- To calculate column access time, consider the worst-case signal propagation delay $\tau_{PHL}$ for below inverter

\[ t_{column} = 18\text{ns} \quad \text{(using eq. 6.18: } t_{PHL}) \]

\[ t_{access} = t_{row} + t_{column} = 38.5\text{ns} \]
10.5 Flash Memory

- One transistor with floating gate
- Memory cell can have two states (two threshold)
- Electron accumulated at the floating gate $\rightarrow$ higher threshold $\rightarrow$ "1" state
- Electron removed from the floating gate $\rightarrow$ lower threshold $\rightarrow$ "0" state

Hot electron injection mechanism
(Data programming)

Fowler-Nordheim tunneling mechanism
(Data erasing)
Equivalent Capacitive-Coupling Circuit

- $V_{FG}$ by capacitive coupling after $V_{CG}$ & $V_D$ applied
  
  \[
  V_{FG} = \frac{Q_{FG}}{C_{total}} + \frac{C_{FC}}{C_{total}} V_{CG} + \frac{C_{FD}}{C_{total}} V_D
  \]

  \[
  C_{total} = C_{FC} + C_{FS} + C_{FB} + C_{FD}
  \]

- min. $V_{CG}$ to turn on the control gate transistor
  
  \[
  V_T(CG) = \frac{C_{total}}{C_{FC}} V_T(FG) - \frac{Q_{FG}}{C_{FC}} - \frac{C_{FD}}{C_{FC}} V_D
  \]

  \[
  \Delta V_T(CG) = -\frac{\Delta Q_{FG}}{C_{FC}}
  \]

$Q_{FC}$ : charge stored at floating gate
$C_{total}$ : total cap.
$C_{FC}$ : cap. between floating and control gate
$C_{FS}$, $C_{FB}$ and $C_{FD}$ : cap. between floating gate and source, bulk and drain
$V_{CG}$ and $V_D$ : voltage at control gate and drain
$V_T(FG)$ : threshold voltage to turn on the floating gate transistor
I-V Characteristic of Flash Memory

- Low and high threshold voltages for control gate voltage

Diagram showing the I-V characteristic of a Flash Memory with control gate voltage on the x-axis and transistor current on the y-axis. The diagram illustrates the difference in current for a low threshold cell (Data "0") and a high threshold cell (Data "1").
NOR Flash Memory Cell

- Bias conditions and configuration of NOR Cells
- F-N tunneling mechanism for erase operation
- Hot-electron injection mechanism for programming operation
## Bias Conditions of NOR Cell

<table>
<thead>
<tr>
<th>Signal</th>
<th>Operation</th>
<th>Erase</th>
<th>Programming</th>
<th>Read</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bit line 1</td>
<td>Open</td>
<td>6V</td>
<td>1V</td>
<td></td>
</tr>
<tr>
<td>Bit line 2</td>
<td>Open</td>
<td>0V</td>
<td>0V</td>
<td>0V</td>
</tr>
<tr>
<td>Source line</td>
<td>12V</td>
<td>0V</td>
<td>0V</td>
<td>0V</td>
</tr>
<tr>
<td>Word line 1</td>
<td>0V</td>
<td>0V</td>
<td>0V</td>
<td>0V</td>
</tr>
<tr>
<td>Word line 2</td>
<td>0V</td>
<td>12V</td>
<td>5V</td>
<td></td>
</tr>
<tr>
<td>Word line 3</td>
<td>0V</td>
<td>0V</td>
<td>0V</td>
<td></td>
</tr>
</tbody>
</table>
NAND Flash Memory Cell

- Cross-section view and configuration of NAND cells

- F-N tunneling mechanism for erase
- F-N tunneling mechanism for program
- Slower programming and read speed but smaller area than NOR cell structure
# Bias Conditions of NAND Cell

<table>
<thead>
<tr>
<th>Signal</th>
<th>Erase</th>
<th>Programming</th>
<th>Read</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bit line 1</td>
<td>Open</td>
<td>0V</td>
<td>1V</td>
</tr>
<tr>
<td>Bit line 2</td>
<td>Open</td>
<td>0V</td>
<td>1V</td>
</tr>
<tr>
<td>Select line 1</td>
<td>Open</td>
<td>5V</td>
<td>5V</td>
</tr>
<tr>
<td>Word line 1</td>
<td>0V</td>
<td>10V</td>
<td>5V</td>
</tr>
<tr>
<td>Word line 2</td>
<td>0V</td>
<td>10V</td>
<td>5V</td>
</tr>
<tr>
<td>Word line 3</td>
<td>0V</td>
<td>10V</td>
<td>5V</td>
</tr>
<tr>
<td>Word line 4</td>
<td>0V</td>
<td>10V</td>
<td>5V</td>
</tr>
<tr>
<td>Word line 5</td>
<td>0V</td>
<td>20V</td>
<td>0V</td>
</tr>
<tr>
<td>Word line 6</td>
<td>0V</td>
<td>10V</td>
<td>5V</td>
</tr>
<tr>
<td>Word line 7</td>
<td>0V</td>
<td>10V</td>
<td>5V</td>
</tr>
<tr>
<td>Word line 8</td>
<td>0V</td>
<td>10V</td>
<td>5V</td>
</tr>
<tr>
<td>Select line 2</td>
<td>Open</td>
<td>0V</td>
<td>5V</td>
</tr>
<tr>
<td>Source line</td>
<td>Open</td>
<td>0V</td>
<td>0V</td>
</tr>
<tr>
<td>p-well 2</td>
<td>20V</td>
<td>0V</td>
<td>0V</td>
</tr>
<tr>
<td>n-sub</td>
<td>20V</td>
<td>0V</td>
<td>0V</td>
</tr>
</tbody>
</table>
Comparison between NOR and NAND

<table>
<thead>
<tr>
<th></th>
<th>NOR</th>
<th>NAND</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erase method</td>
<td>F-N tunneling</td>
<td>F-N tunneling</td>
</tr>
<tr>
<td>Programming method</td>
<td>Hot electron injection</td>
<td>F-N tunneling</td>
</tr>
<tr>
<td>Erase speed</td>
<td>Slow</td>
<td>Fast</td>
</tr>
<tr>
<td>Program speed</td>
<td>Fast</td>
<td>Slow</td>
</tr>
<tr>
<td>Read speed</td>
<td>Fast</td>
<td>Slow</td>
</tr>
<tr>
<td>Cell size</td>
<td>Large</td>
<td>Small</td>
</tr>
<tr>
<td>Scalability</td>
<td>Difficult</td>
<td>Easy</td>
</tr>
<tr>
<td>Application</td>
<td>Embedded system</td>
<td>Mass storage</td>
</tr>
</tbody>
</table>
Multilevel Cell Concept

- Effective memory density can be improved
- Possible state number limited by
  - Available charge range
  - Accuracy of programming and read operations
  - Disturbance of state over time

Threshold voltage distribution of 2bits/cell storage
Flash Memory Circuit

- On-chip charge pump used to generate programming voltage
- Chain of diode and cap. to charge or discharge each half cycle

\[ V_{out} = V_{in} + (\gamma V_{DD} - V_T(MN1)) + \cdots + (\gamma V_{DD} - V_T(MN_n)) \]
10.6 Ferroelectric Random Access Memory

- Hysteresis characteristic of a ferroelectric cap.

- Total charge varies as function of applied voltage
Structure and Operation of FRAM

◆ Similar to DRAM except plate line

- Step-sensing scheme

\[
\Delta V_1 = \frac{C_1}{C_1 + C_{BL}} V_{DD}
\]

\[
\Delta V_0 = \frac{C_0}{C_0 + C_{BL}} V_{DD}
\]

$C_1$ and $C_0$ : linearly modeled ferroelectric cap
Problems of FRAM

◆ Step-sensing scheme cause reliability issues
  ▪ Pulse sensing scheme also used with read speed penalty

◆ Fatigue
  ▪ Capacitance charge gradually degraded with repeated use

◆ Imprint
  ▪ Ferroelectric cap tends to stay at one state preferably when state maintained for a long time