Semiconductor Memories

Adapted from Chapter 12 of

*Digital Integrated Circuits: A Design Perspective*

Jan M. Rabaey et al.

Copyright 2003 Prentice Hall/Pearson

Outline

- Memory Classification
- Memory Architectures
- The Memory Core
- Periphery
- Reliability
- Case Studies
### Semiconductor Memory Classification

<table>
<thead>
<tr>
<th>Read-Write Memory</th>
<th>Non-Volatile Read-Write Memory</th>
<th>Read-Only Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Access</td>
<td>Non-Random Access</td>
<td>EPROM</td>
</tr>
<tr>
<td>SRAM</td>
<td>FIFO</td>
<td>E²PROM</td>
</tr>
<tr>
<td>DRAM</td>
<td>LIFO</td>
<td>FLASH</td>
</tr>
<tr>
<td>Shift Register</td>
<td>CAM</td>
<td>Mask-Programmed Programmable (PROM)</td>
</tr>
</tbody>
</table>

### Memory Timing: Definitions

- **Read cycle**
- **Write cycle**
- **Read access**
- **Write access**
- **Data valid**
- **Data written**
Memory Architecture: Decoders

- Intuitive architecture for $N \times M$ memory
  - Too many select signals: $N$ words == $N$ select signals
  - Decoder reduces the number of select signals
    - $K = \log_2 N$

Problem: ASPECT RATIO or HEIGHT >> WIDTH

Array-Structured Memory Architecture

- Amplify swing to rail-to-rail amplitude
- Selects appropriate word
Hierarchical Memory Architecture

Advantages:
1. Shorter wires within blocks
2. Block address activates only 1 block => power savings

Block Diagram of 4 Mbit SRAM
**Contents-Addressable Memory**

- Address Decoder
- I/O Buffers
- Commands
- Validity Bits
- Priority Encoder
- Data (64 bits)
- Comparand
- Mask
- CAM Array
- Control Logic
- Address Decoder
- R/W Address (9 bits)
- 29 words 3 64 bits

**Memory Timing: Approaches**

- DRAM Timing
  - Multiplexed Addressing
- SRAM Timing
  - Self-timed
**Memory core**

**Read-Only Memory Cells**

![Diagram of Memories](image)

- **Diode ROM**
- **MOS ROM 1**
- **MOS ROM 2**

**MOS OR ROM**

![Diagram of MOS OR ROM](image)

© Digital Integrated Circuits 2nd Edition Memories
MOS NOR ROM

MOS NOR ROM Layout

Programmimg using the Active Layer Only

Cell (9.5\lambda \times 7\lambda)
**MOS NOR ROM Layout**

Cell (11λ x 7λ)

Programming using the Contact Layer Only

- Polysilicon
- Metal1
- Diffusion
- Metal1 on Diffusion

**MOS NAND ROM**

All word lines high by default with exception of selected row

- Pull-up devices

© Digital Integrated Circuits 2nd Memories
MOS NAND ROM Layout

- No contact to VDD or GND necessary;
- Drastically reduced cell size
- Loss in performance compared to NOR ROM

Programmimg using the Metal-1 Layer Only

Cell (8\(\lambda\) x 7\(\lambda\))

NAND ROM Layout

- Programming using Implants Only

Cell (5\(\lambda\) x 6\(\lambda\))

Polysilicon
Diffusion
Metal1 on Diffusion
Threshold-altering implant
Equivalent Transient Model for MOS NOR ROM

Model for NOR ROM

- **Word line parasitics**
  - Wire capacitance and gate capacitance
  - Wire resistance (polysilicon)

- **Bit line parasitics**
  - Resistance not dominant (metal)
  - Drain and Gate-Drain capacitance

Equivalent Transient Model for MOS NAND ROM

Model for NAND ROM

- **Word line parasitics**
  - Similar to NOR ROM

- **Bit line parasitics**
  - Resistance of cascaded transistors dominates
  - Drain/Source and complete gate capacitance
Decreasing Word Line Delay

(a) Driving the word line from both sides

(b) Using a metal bypass

(c) Use silicides

Precharged MOS NOR ROM

PMOS precharge device can be made as large as necessary, but clock driver becomes harder to design.
Non-Volatile Memories

The Floating-gate transistor (FAMOS)

Floating-gate transistor (FAMOS)

Device cross-section

Schematic symbol

Floating-Gate Transistor Programming

Avalanche injection

Removing programming voltage leaves charge trapped

Programming results in higher $V_T$.
A “Programmable-Threshold” Transistor

\[ I_D \] vs. \[ V_{GS} \]

“0”-state \[ \rightarrow \] “ON” \[ \rightarrow \] “1”-state

\[ V_{WL} \]

FLOTOX EEPROM

Floating gate

Gate

Drain

Source

Floating gate

Gate

Drain

Source

FLOTOX transistor

Fowler-Nordheim I-V characteristic
**EEPROM Cell**

Absolute threshold control is hard
Unprogrammed transistor might be depletion
\[ \Rightarrow 2 \text{ transistor cell} \]

**Flash EEPROM**

Control gate
Floating gate
Thin tunneling oxide

Many other options …
Cross-sections of NVM cells

Flash

EPROM

Basic Operations in a NOR Flash Memory—Erase

© Digital Integrated Circuits 2nd

Courtesy Intel Memories
Basic Operations in a NOR Flash Memory—Write

Basic Operations in a NOR Flash Memory—Read
Characteristics of State-of-the-art NVM

Table 12-1 Comparison between nonvolatile memories ([Itoh01]).

\[ V_{DD} = 3.3 \text{ or } 5 \text{ V}; V_{PP} = 12 \text{ or } 12.5 \text{ V}. \]

<table>
<thead>
<tr>
<th>Cell—Nr. of Transistors</th>
<th>Area (ratio wrt EPROM)</th>
<th>Mechanism</th>
<th>External Power Supply</th>
<th>Program/Erase Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MASK ROM</strong></td>
<td>1 T (NAND) 0.35–5</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>EPROM</strong></td>
<td>1 T</td>
<td>UV Exposure</td>
<td>Hot electrons  [ V_{PP} ]</td>
<td>[ V_{DD} ] ~100</td>
</tr>
<tr>
<td><strong>EEEPROM</strong></td>
<td>2 T 3–5</td>
<td>FN Tunneling</td>
<td>FN Tunneling  [ V_{PP} (int) ]</td>
<td>[ V_{DD} ] [ 10^4–10^5 ]</td>
</tr>
<tr>
<td><strong>Flash Memory</strong></td>
<td>1 T 1–2</td>
<td>FN Tunneling</td>
<td>Hot electrons  [ V_{PP} ]</td>
<td>[ V_{DD} ] [ 10^4–10^5 ]</td>
</tr>
</tbody>
</table>

Read-Write Memories (RAM)

- **STATIC (SRAM)**
  - Data stored as long as supply is applied
  - Large (6 transistors/cell)
  - Fast
  - Differential

- **DYNAMIC (DRAM)**
  - Periodic refresh required
  - Small (1-3 transistors/cell)
  - Slower
  - Single Ended
6-transistor CMOS SRAM Cell

CMOS SRAM Analysis (Read)
**CMOS SRAM Analysis (Read)**

\[ CR = \frac{W_1}{L_1} \cdot \frac{W_2}{L_2} \]

**CMOS SRAM Analysis (Write)**

\[ k_n, M_6 \left( V_{DD} - V_{TN} \right) V_Q - \frac{V_Q^3}{2} = k_p, M_4 \left( V_{DD} - |V_{DSAT_p}| \right) \frac{V_{DSAT_p}^2}{2} \]

\[ V_Q = V_{DD} - V_{TN} - \sqrt{\left( V_{DD} - V_{TN} \right)^2 - 2 \frac{\mu_p}{\varepsilon_p} (V_{DD} - |V_{DSAT_p}|) \frac{V_{DSAT_p}^2}{2}} \]

\[ PR = \frac{W_4}{W_6} \cdot \frac{L_1}{L_6} \]
CMOS SRAM Analysis (Write)

\[ PR = \frac{W_4/L_4}{W_6/L_6} \]

6T-SRAM — Layout

© Digital Integrated Circuits 2nd Edition
Memories
**Resistance-load SRAM Cell**

Static power dissipation -- Want $R_L$ large
Bit lines precharged to $V_{DD}$ to address $t_p$ problem

---

**SRAM Characteristics**

Table 12-2  Comparison of CMOS SRAM cells used in 1-Mbit memory
(from [Takade91])

<table>
<thead>
<tr>
<th>Complementary CMOS</th>
<th>Resistive Load</th>
<th>TFT Cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of transistors</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Cell size</td>
<td>$58.2 , \mu m^2$</td>
<td>$40.8 , \mu m^2$</td>
</tr>
<tr>
<td></td>
<td>(0.7-\mu m rule)</td>
<td>(0.7-\mu m rule)</td>
</tr>
<tr>
<td>Standby current (per cell)</td>
<td>$10^{-15} , A$</td>
<td>$10^{-12} , A$</td>
</tr>
</tbody>
</table>
3-Transistor DRAM Cell

No constraints on device ratios
Reads are non-destructive
Value stored at node X when writing a “1” = \( V_{\text{WWL}} - V_{\text{Tn}} \)

3T-DRAM — Layout
**1-Transistor DRAM Cell**

Write: \( C_S \) is charged or discharged by asserting WL and BL.
Read: Charge redistribution takes place between bit line and storage capacitance.

\[
\Delta V = V_{BL} - V_{PRE} = \frac{C_S}{C_S + C_B} \frac{V_{DD}}{2} - V_T
\]

Voltage swing is small; typically around 250 mV.

---

**DRAM Cell Observations**

- 1T DRAM requires a sense amplifier for each bit line, due to charge redistribution read-out.
- DRAM memory cells are single ended in contrast to SRAM cells.
- The read-out of the 1T DRAM cell is destructive; read and refresh operations are necessary for correct operation.
- Unlike 3T cell, 1T cell requires presence of an extra capacitance that must be explicitly included in the design.
- When writing a “1” into a DRAM cell, a threshold voltage is lost. This charge loss can be circumvented by bootstrapping the word lines to a higher value than \( V_{DD} \).
**Sense Amp Operation**

- $V_{BL}$
- $V_{PRE}$
- $V(1)$
- $V(0)$
- $t$
- $DV(1)$

Sense amp activated
Word line activated

**1-T DRAM Cell**

Uses Polysilicon-Diffusion Capacitance
Expensive in Area
SEM of poly-diffusion capacitor 1T-DRAM

Advanced 1T DRAM Cells

Trench Cell

Stacked-capacitor Cell
Periphery

- Decoders
- Sense Amplifiers
- Input/Output Buffers
- Control / Timing Circuitry

Row Decoders

Collection of $2^m$ complex logic gates
Organized in regular and dense fashion

(N)AND Decoder

\[
WL_0 = \overline{A_6}A_1A_2A_3A_4A_5A_6A_7A_8A_9
\]

\[
WL_{511} = \overline{A_0}A_1A_2A_3A_4A_5A_6A_7A_8A_9
\]

NOR Decoder

\[
WL_0 = A_0 + A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 + A_8 + A_9
\]

\[
WL_{511} = A_0 + A_1 + A_2 + A_3 + A_4 + A_5 + A_6 + A_7 + A_8 + A_9
\]
Hierarchical Decoders

Multi-stage implementation improves performance

NAND decoder using 2-input pre-decoders

Dynamic Decoders

2-input NOR decoder 2-input NAND decoder
4-input pass-transistor based column decoder

Advantages: speed (\( t_{pd} \) does not add to overall memory access time)
Only one extra transistor in signal path
Disadvantage: Large transistor count

4-to-1 tree based column decoder

Number of devices drastically reduced
Delay increases quadratically with # of sections; prohibitive for large decoders
Solutions: buffers
progressive sizing
combination of tree and pass transistor approaches
Sense Amplifiers

\[ t_p = \frac{C \cdot \Delta V}{I_{av}} \]

make \( \Delta V \) as small as possible

Idea: Use Sense Amplifier

Differential Sense Amplifier

Directly applicable to SRAMs
**Differential Sensing — SRAM**

![SRAM Sensing Scheme](image)

> © Digital Integrated Circuits 2nd Memories

**Latch-Based Sense Amplifier (DRAM)**

![Latch-Based Sense Amplifier](image)

> Initialized in its meta-stable point with EQ
> Once adequate voltage gap created, sense amp enabled with SE
> Positive feedback quickly forces output to a stable operating point.

© Digital Integrated Circuits 2nd Memories
Single-to-Differential Conversion

How to make a good $V_{\text{ref}}$?

Open bitline architecture with dummy cells
**DRAM Read Process with Dummy Cell**

![Diagrams showing read process with dummy cell](image)

**Noise Sources in 1T DRam**

![Noise sources diagram](image)
**Transposed-Bitline Architecture**

(a) Straightforward bit-line routing

(b) Transposed bit-line architecture