## 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></td>
</tr>
<tr>
<td>SRAM</td>
<td>FIFO</td>
<td>EPROM</td>
</tr>
<tr>
<td>DRAM</td>
<td>LIFO</td>
<td>E²PROM</td>
</tr>
<tr>
<td></td>
<td>Shift Register</td>
<td>FLASH</td>
</tr>
<tr>
<td></td>
<td>CAM</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mask-Programmed Programmable (PROM)</td>
</tr>
</tbody>
</table>
Memory Timing: Definitions

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

Intuitive architecture for $N \times M$ memory
Too many select signals:
$N$ words $\Rightarrow$ $N$ select signals

Decoder reduces the number of select signals
$K = \log_2 N$
Array-Structured Memory Architecture

Problem: ASPECT RATIO or HEIGHT >> WIDTH

- Row Decoder
- Bit line
- Storage cell
- Word line
- Sense amplifiers / Drivers
- Column decoder
- Input-Output (M bits)

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

Advantages:
1. Shorter wires within blocks (word and bit lines)
2. Block address activates only 1 block => power savings
Memory Timing: Approaches

**DRAM Timing**
- Row Address
- Column Address
- RAS
- CAS
- RAS-CAS timing

**Multiplexed Addressing**

**SRAM Timing**
- Address transition initiates memory operation
- Self-timed
Read-Only Memory Cells

Diode ROM

MOS ROM 1

MOS ROM 2
MOS OR ROM


V_D\text{DD}

V_{bias}

Pull-down loads
MOS NOR ROM

- WL[0]
- WL[1]
- WL[2]
- WL[3]

- BL[0]
- BL[1]
- BL[2]
- BL[3]

Pull-up devices

V_{DD}

GND
PMOS precharge device can be made as large as necessary, but clock driver becomes harder to design.
MOS NAND ROM

All word lines high by default with exception of selected row
Non-Volatile Memories
The Floating-gate transistor (FAMOS)
Floating-Gate Transistor Programming

Avalanche injection

Removing programming voltage leaves charge trapped

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

![Graph showing the relationship between $I_D$ and $V_{GS}$, with two states: “0”-state and “1”-state. The graph includes labels for $V_{WL}$, $\Delta V_T$, “ON”, and “OFF”.](image-url)
**FLOTOX EEPROM**

FLOTOX transistor

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

Absolute threshold control is hard
Unprogrammed transistor might be depletion
⇒ 2 transistor cell
Flash EEPROM: ETOX device

Control gate

Floating gate

Thin tunneling oxide

$\Theta$ programming

$p$-substrate

$\Theta$ erase

$n^+$ source

$n^+$ drain
Basic Operations in a NOR Flash Memory: Write
Basic Operations in a NOR Flash Memory: Read
Basic Operations in a NOR Flash Memory: Erase
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
A 0 is written in the cell by setting $\overline{BL}$ to 1 and $BL$ to 0
CMOS SRAM Analysis (Read)

$BL$ and $\overline{BL}$ are precharged to $V_{DD}$
Resistance-load SRAM Cell

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

No constraints on device ratios
Reads are non-destructive
Value stored at node X when writing a “1” = $V_{WWL-V_{Tn}}$
Write: \( C_S \) is charged or discharged by asserting WL and BL.

Read: Charge redistribution takes place between bit line and storage capacitance

\[
\Delta V = V_{BL} - V_{PRE} = \left( V_X - V_{PRE} \right) \frac{C_S}{C_S + C_{BL}}
\]

Voltage swing is small; typically around 250 mV.
Sense Amp Operation

\[ V_{\text{PRE}} \]

\[ \Delta V(1) \]

Sense amp activated

Word line activated
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 $V_{DD}$.
Static CAM Memory Cell

Wired-NOR Match Line
Periphery

- Decoders
- Input/Output Buffers
- Sense Amplifiers
- Control / Timing Circuitry
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
Latch-Based Sense Amplifier (DRAM)

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.
Single-to-Differential Conversion

How to make a good $V_{ref}$?
Open bitline architecture with dummy cells
Semiconductor Memory Trends

Memory Size as a function of time: x 4 every three years
Trends in Memory Cell Area

The graph illustrates the trend in memory cell area over time. The x-axis represents the years from 1970 to 2000, while the y-axis represents the memory cell area in micrometers squared (μm²). The data points show the reduction in memory cell area with time, indicating the advancement in memory technology. The graph includes markers for different types of memory cells, such as SRAM, DRAM, Flash, full CMOS, poly-Si load, planar capacitor, and 3-D capacitor. The data is sourced from ISSCC.