Memory Design I

Professor Chris H. Kim

University of Minnesota
Dept. of ECE

chriskim@ece.umn.edu

Array-Structured Memory Architecture

---

[Diagram of array-structured memory architecture]

- Storage cell
- Word line
- Bit line
- Amplify swing to rail-to-rail amplitude
- Selects appropriate word
- Sense amplifiers / Drivers
- Column decoder
- Input-Output (M bits)
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 Shift Register CAM</td>
<td>FLASH</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mask-Programmed</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Programmable (PROM)</td>
</tr>
</tbody>
</table>

Read-Write Memories (RWM)

- Basic storage elements of semiconductor memory

SRAM: cell has gain, 6T, FAST, LOW POWER, logic compatible, differential

DRAM: cell has no gain, 1T, refresh, slow, DRAM process, single ended, DENSE
Memory Scaling Trend

- High density is the primary design goal for memories
- Low voltage operation is essential for low power

Itoh, IBM R&D, 2003

Memory Scaling Trend

- Long retention time → low Ioff
  - High Vt is required
- Fast access time → high Ion
  - High Vgs-Vt is required
- Vdd cannot be scaled down aggressively for low power consumption

Itoh, IBM R&D, 2003
Why SRAMs are Important

- Memories have better power efficiency compared to logic
- ~9.9B out of 10B transistors will be used for SRAMs
- Company with better SRAM design will dominate

\[
\sigma_V \propto \frac{1}{\sqrt{\text{Area}}} = \frac{1}{\sqrt{\text{WL}}}
\]

Taur, Ning

- Area is the number one concern → minimum sized devices
- Smaller devices have larger variation
- Delay variation, stability, leakage is a problem
- Central limit theorem doesn’t hold (\(\sigma/\mu\))
Positive Feedback: Bi-Stability

Positive Feedback

\[ V_{o1} = V_{i2} \]

\[ V_{i1} = V_{o2} \]

Meta-Stability

Meta-Stability

Gain should be larger than 1 in the transition region
SRAM Memory Cell

• NMOS access transistors
• Read and write uses the same port: need sufficient margins
• One wordline to access cell
• Two bit lines (BL, BLB) to carry the data
• Almost minimum size transistors for small cell area

SRAM Read Operation

• Both bit lines are precharged to Vdd
• Wordline is fired for one of the cells on bit line
• Cell pulls down either BL or BLB
• Sense amp regenerates the differential signal
• Data should not flip after read access
• Driver TR must be stronger than access TR
For high density, large number of cells share bitline and wordline
- Subarray organization for 32Kb: 128 WL’s, 256BL’s

- \( C_{\text{bitline}} \) is large due to large number of cells attached
- \( I_{\text{cell}} \) is small due to high density cells
- \( \Delta V_{\text{bitline}} \) has to be minimized for high speed
  - < 100mV bitline voltage difference generated by SRAM cell
  - Let the sense amplifier finish the job
  - Increased noise sensitivity, circuit complexity
SRAM Read Operation: Precharge

- Option (a)
  - Similar to dynamic logic precharge
  - Balance transistor to equalize bitline voltages
  - Short wordline pulse required to limit bitline swing

- Option (b)
  - Pseudo-NMOS type circuit
  - Bitline voltage clamped during read

- Option (c)
  - NMOS pullup instead of PMOS
  - Precharge levels are limited to $V_{dd} - V_t$
  - Can't operate at low $V_{dd}$
SRAM Cell Read Margin

- When cell is not accessed (WL=0)
  - Data is safely kept inside the cell
  - High noise margin
- When cell is accessed (WL=V_{dd})
  - Access transistor acts as a noise source
  - Data '0' is pulled up to V_{x}
  - Cell data can flip if V_{x} rises above V_{in}

Static Noise Margin

- Destructive read problem
- The size of the largest square enclosed in the butterfly curves = read static noise margin
CMOS SRAM Analysis (Read)

Cell beta ratio = \((W/L)_{\text{drv}} / (W/L)_{\text{access}}\)

- Increasing the size of the driver NMOS improves read margin
- But remember, area is the number one constraint in memory design
- Increasing cell size a not a good trade off
Techniques to Improve Read Margin

• High $V_t$ transistors
  – Internal node on low side needs to rise to $V_t$ or more
  – Virtually never happens when $V_t$ is larger than half $V_{dd}$
  – Cell is extremely stable at ultra-low power design point
  – Beta ratio constraint is relaxed \(\Rightarrow\) smaller driver and larger access TR can be used for faster read and write

Techniques to Improve Read Margin

• Boosted cell supply
  – Supply voltage of SRAM cell is higher than outside
  – Makes driver stronger than access, suppressing the rise in the low side
  – Effectively improves the beta ratio
  – Driver NMOS can be downsized, decreasing cell size
SRAM Write Operation

- Launch the write data on BL and BLB
- Word line signal is fired
- Low bit line value flips cell data
- Access TR must be stronger than PMOS load

CMOS SRAM Analysis (Write)

\[ PR = \frac{(W/L)_4}{(W/L)_6} \]

\[ k_nM_n \left( V_{DD} - V_{T_n} \right) V_Q - \frac{V_Q^2}{2} = k_pM_p \left( V_{DD} - |V_{T_p}| \right) V_{DSS} - \frac{V_{DSS}^2}{2} \]

\[ V_Q = V_{DD} - V_{T_n} - \sqrt{\left( V_{DD} - V_{T_n} \right)^2 - 2\frac{U_{n}}{\mu_n C_m} \left( V_{DD} - |V_{T_p}| \right) V_{DSS} - \frac{V_{DSS}^2}{2}} \]
SRAM Cell Write Margin

- Access transistor must be stronger than PMOS to pull the below the trip point (typical pull-up ratio ~ 1.5)
- To avoid cell size increase, correct pull-up ratio achieved by controlling $V_{tn}$ and $V_{tp}$

Techniques to Improve Write Margin

- Sizing: access TR vs. PMOS in latch
- Higher WL voltage for access TR
- Virtual VDD
6T-SRAM Layout Until 90nm

Compact cell
Bitlines: M2
Wordline: strapped in M3

6T-SRAM Layout From 65nm
6T versus 4T SRAM

6T SRAM Cell
Supply current is limited to the leakage current of transistors in the stable state

4T SRAM Cell
High degree of compactness
High power consumption

RAM Variations
• Many variations to the basic 6T SRAM cell
• More functionality, smaller cells
  – Dual read or single write cell
  – True multi-ported cell
  – Content addressable memory (CAM)
  – 4T memory cell
  – 3T memory cell
  – 2T memory cell
  – 1T DRAM cell
Dual Read or Single Write Cell

- Two wordlines, one for each access transistor
- Small increase in cell size
- Can either
  - read two different cells in one cycle
  - or write to one cell

Multi Ported Cell

- Each port has separate address
- Memory access bandwidth is twice (ideally)
- “Write through”: data written can be read by another port in the very same cycle
Content Addressable Memory (CAM)

- Some applications need to find out if anything in the memory matches a certain key value (e.g., tag).
- Special memory with XOR gates that compare cell value with the data on the bitlines.
- Precharged match line shared across words will stay high only when the entire word matches.
- Needs an encoder that will output the matching address.

Smaller RAM Cells

- Internal nodes don't go to Vdd.
- Cell won't work at low Vdd.
- High value stored is degraded.
- Effective strength of NMOS driver is reduced.
- Refresh is needed.
- Need 2 wordlines, read WL and write WL.
- Can have 1 or 2 bitlines (Read/Write).
- Not very small, since it has more wires.
**1-T 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.

\[ \Delta V = V_{BL} - V_{PRE} = V_{BIT} - V_{PRE} \frac{C_S}{C_S + C_{BL}} \]

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.
- 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

- **$V_{BL}$**: Voltage on bit line.
- **$V_{PRE}$**: Precharge voltage.
- **$V(1)$**: Voltage for a high bit.
- **$V(0)$**: Voltage for a low bit.
- **$t$**: Time.

- **$\delta V(1)$**: Change in voltage.
- **Sense amp activated**: Once the sense amp is activated, the bit line voltage changes.
- **Word line activated**: Once the word line is activated, the cell is ready to sense the data.

### 1-T DRAM Cell

- **Cross-section**: Shows the layout of the DRAM cell.
- **Layout**: Shows the physical arrangement of the components.

- **Uses Polysilicon-Diffusion Capacitance**
- **Expensive in Area**: The layout is complex and uses significant area.
Advanced 1-T DRAM Cells

Trench Cell

Stacked-capacitor Cell

Good References on RAM