Class 16: Memories

Topics:
1. Intro
2. Types of Memories
3. Components of a memory
4. Memory Array Organization
5. Memory Array Organization
6. Memory Array Organization
7. SRAM - Read / Write
8. 6T SRAM - Read
9. 6T SRAM - Write
10. 4T SRAM
11. 4T SRAM
12. Sense Amps
13. Sense Amps
Class 16: Memories
Types of Memories

Volatile - need power to retain memory
  • SRAMs - fast, somewhat dense, used for data/instructions
  • DRAMs - fast, most dense, used for data/instructions

• Non-volatile - retains memory without power
  • EEPROMs - slow, used for initial boot up data, erase bit/word a time
  • FLASH - denser than EEPROMs, still slow, erase in a flash, used for boot up sequences, initial code development of ROMs
  • ROMs - dense, fast, used once code is finalized
Class 16: Memories
Components of a memory (Martin c.11, Sedra c.13)

• Memory Array
  • contains n x m memory cells, pass/access transistors
• Row Decoders
  • decodes n-bits for which row to select
• Column Decoders
  • decodes m-bits for which column to select
• Sense Amplifiers
  • takes output of memory cell and sends “1” or “0” to I/O
# Class 16: Memories

## Memory Array Organization

1024 x 16 => 1024 words x 16 I/Os = 256 rows x 4 words/row x 16 I/Os

<table>
<thead>
<tr>
<th>B00</th>
<th>B01</th>
<th>B02</th>
<th>B03</th>
<th>B04</th>
<th>B05</th>
<th>B06</th>
<th>B07</th>
<th>B08</th>
<th>B09</th>
<th>B10</th>
<th>B11</th>
<th>B12</th>
<th>B13</th>
<th>B14</th>
<th>B15</th>
</tr>
</thead>
<tbody>
<tr>
<td>W00</td>
<td>W01</td>
<td>W02</td>
<td>W03</td>
<td>W04</td>
<td>W05</td>
<td>W06</td>
<td>W07</td>
<td>W08</td>
<td>W09</td>
<td>W10</td>
<td>W11</td>
<td>W12</td>
<td>W13</td>
<td>W14</td>
<td>W15</td>
</tr>
<tr>
<td>R00</td>
<td>R01</td>
<td>R02</td>
<td>R03</td>
<td>R04</td>
<td>R05</td>
<td>R06</td>
<td>R07</td>
<td>R08</td>
<td>R09</td>
<td>R10</td>
<td>R11</td>
<td>R12</td>
<td>R13</td>
<td>R14</td>
<td>R15</td>
</tr>
<tr>
<td>C00</td>
<td>C01</td>
<td>C02</td>
<td>C03</td>
<td>C04</td>
<td>C05</td>
<td>C06</td>
<td>C07</td>
<td>C08</td>
<td>C09</td>
<td>C10</td>
<td>C11</td>
<td>C12</td>
<td>C13</td>
<td>C14</td>
<td>C15</td>
</tr>
</tbody>
</table>

256 Rows

16 I/Os

4 Words / Rows

---

Joseph A. Elias, PhD
Class 16: Memories
Memory Array Organization (Martin p.439)

In the example below, this is a 4096 x 1 memory array

64 x 64 = 4096 data bits
64 rows = $2^6$, so 6 bits are needed to address all the rows
64 columns = $2^6$, so 6 bits are needed to address all the columns
R/W is read or write select bit
CS is chip select, whether or not the array is being accessed

Figure 11.1 A simplified block diagram of a 4096 by 1 RAM.
Class 16: Memories
Memory Array Organization (Martin c.11)

• Word Lines W0-W3 are outputs of the row decoder, one will be high for the cell being accessed thus the name access (or pass) transistors
• Bit Lines B0-B3 are outputs of the column decoder, one will be high for the cell being accessed
• State of bit lines depends on read or write mode

Figure 11.2 A simplified portion of an SRAM array.
Class 16: Memories
SRAM - Read / Write (Martin c.11)

Order of events for Read operation:
• CS is high
• R/W is high
• D and D are both low
• B and B are brought to SA
• Output of SA sent to I/O

Order of events for Write operation:
• CS is high
• R/W is low
• Tri-State Buffer is high impedance, no Dout
• Write buffer senses “1” or “0” on Data line
• If D is low, then D will be high
• B will be low, B is pre-charged high
• Cell reflects the data on Data line
To read a “D=0”
• W/L at 0V, both access transistors are off
• Pre-charge both bit lines high (either VDD or VDD/2)
• Correct W/L will go high
• If a “D=0 D=1” is stored, then W/L=H causes Q5 to pass 0V to Q2/Q4, VDD to Q1/Q3
• Charge flows Q4->Q6, thus charging the B bit line voltage
• Charge flows Q5->Q1, thus discharging the B bit line voltage
• Differential B to B voltage of 50-100mV is sensed at the sense amp; must be small so FF doesn’t flip

To read a “D=1”
• W/L at 0V, both access transistors are off
• Pre-charge both bit lines high (either VDD or VDD/2)
• Correct W/L will go high
• If a “D=1 D=0” is stored, then W/L=H causes Q5 to pass VDD to Q2/Q4, 0V to Q1/Q3
• Charge flows Q6->Q2, thus discharging the B bit line voltage
• Charge flows Q3->Q5, thus charging the B bit line voltage
• Differential B to B voltage of 50-100mV is sensed at the sense amp; must be small so FF doesn’t flop
Class 16: Memories
6T SRAM - Write (Martin c.11)

To write a “1”, initially at a “0”
• W/L at 0V, both access transistors are off
• Pre-charge one bit line high (D=B=VDD), the other low (D=B=0V)
• Correct W/L will go high
• Source (B) of Q5 goes to 0->(VDD-Vt), and drain (B) of Q6 goes to VDD->0V
• Positive feedback takes over, and cell stores a “1” on D

To write a “0”, initially at a “1”
• W/L at 0V, both access transistors are off
• Pre-charge one bit line high (B=VDD), the other to ground (B=0V)
• Correct W/L will go high
• Source (B) of Q5 goes to VDD->0V, and drain (B) of Q6 goes to 0->(VDD-Vt)
• Positive feedback takes over, and cell stores a “0” on D
Class 16: Memories
4T SRAM (Martin c.11)

- To save room, load transistors can be replaced by polysilicon resistors
- If there is a double poly flow, one can place these poly resistors over the NMOS transistors
- The constraint on the resistive loads is the maximum static power dissipation
Example 11.1

Assume the memory array must dissipate less than 50 mW and there are approximately 1 million cells in the array. This means each cell must dissipate less than 50 mW/10^6 = 0.05 µW, which in turn limits the d.c. current of each cell to 0.05 µW/3.3 V = 15 nA, assuming V_{DD} = 3.3 V. At any given time, one of the load resistors will have no current through it, and one will have a current

\[ I_{cell} = \frac{V_{DD}}{R_L} \]  

(11.1)

Thus we need

\[ \frac{V_{DD}}{R_L} < 1.5 \times 10^{-8} \]  

(11.2)

which implies

\[ R_L > \frac{V_{DD}}{1.5 \times 10^{-8}} = 500 \text{ MΩ} \]  

(11.3)

- Advantage: lower transistor count, much denser
- Disadvantage: no drive capability, only enough to overcome leakage to preserve data
- If size of array is > 1Mbits, then poly resistor may be in terraohms (1e12)
• Due to large parasitic capacitances, and need to have fast reads, differential sense amps are used
• When not being used, Sel is low and Q5 is off, thus no static current for low power dissipation
• When being accessed, Sel goes high, and Q5 is turned on
• Both Q1 and Q2 begin to conduct
• Depending on D and D, the one with the higher gate voltage will conduct more current
• The output of sense amp will be high if the current of Q1 is large compared to Q2

• It is desired to have Q1-Q4 operate in the active (saturation) region when differential voltage is 0
• Transistors Q5 is normally wider, and operates in the triode (linear) region
Read operation:
1) $\phi_p$ is brought high
   - pre-charge and equalization ckt is active
   - $B$ and $B$ are equal (Q7) to VDD/2
2) $\phi_p$ is brought low
   - $B$ and $B$ float for a short time
3) W/L goes high
   - cell connected to $B$ and $B$
   - $B$ and $B$ have differential voltage
4) After some time, $\phi_s$ is brought high
   - Q5/Q6 turned on
   - FF operating at an unstable point
   - Both inverters in transition region at VDD/2
   - Current to $C_b$ is $(g_{mn} + g_{mp})V_i$
   - This develops a voltage, which is fed to $C_b$