CHAPTER 16 MEMORY CIRCUITS

Chapter Outline
16.1 Latches and Flip-Flops
16.2 Semiconductor Memories: Types and Architectures
16.3 Random-Access Memory (RAM) Cells
16.4 Sense-Amplifier and Address Decoders
16.5 Read-Only Memory (ROM)
16.1 LATCHES AND FLIP-FLOPS

Logic Classifications

- Combinational circuits: the output depends only on the present value of the input
- Sequential circuits: the output depends not only on the present input values but also on the previous values
- Static sequential circuits: use positive feedback to provide two stable states (bistable)
- Dynamic sequential circuits: use the storage of charge on a capacitor

Latch

- Exhibits two stable operating points and one unstable operating point
- The latch needs to be triggered to change stage
- The latch together with the triggering circuitry forms a flip-flop
The SR Flip-Flop

- The latch is triggered by input S and R
  - \( S = 1 \) and \( R = 0 \) \( \rightarrow \) \( Q = 1 \)
  - \( S = 0 \) and \( R = 1 \) \( \rightarrow \) \( Q = 0 \)
  - \( S = 0 \) and \( R = 0 \) \( \rightarrow \) \( Q \) unchanged
  - \( S = 0 \) and \( R = 0 \) \( \rightarrow \) \( Q \) undefined (not allowed).

CMOS Implementation of SR Flip-Flop

- The flip-flop is set by \( S \) and \( R \) when the clock \( \phi \) is high
  - \( Q_5 \) and \( Q_6 \) should be able to pull the output node at least below the threshold of inverter \((Q_3, Q_4)\)
  - The period of the set signal should be long enough to cause regeneration to take over
- The state is latched when the clock \( \phi \) is low
- No static power dissipation (no conducting path between \( V_{DD} \) and ground except during switching)
**D Flip-Flop Circuits**

- Simple implementation of the D flip-flop:
  - \( \phi = 1 \) and \( \bar{\phi} = 0 \): the loop is open and \( Q \) is determined by \( D \)
  - \( \phi = 0 \) and \( \bar{\phi} = 1 \): the loop is closed and the flip-flop is in latch mode
  - Two-phase non-overlapping clock is required for D flip-flop operation
  - Major drawback: the output simply follows the signal on the D input line during \( \phi \)

- Master-slave D flip-flop:
16.2 SEMICONDUCTOR MEMORIES

Memory Types

- Random-access memory (RAM): access time is independent of the physical location of the stored info
- Sequential memories: data are available only in the same sequence in which the data were stored
- Read/write memory: permits data to be stored and retrieved at comparable speeds
- Read-only memory (ROM): only permits reading operation

Memory-Chip Organization

- The bits on a memory chip are either individually addressable (64M × 1) or addressable in groups (16M × 4)
- The increase in word line and bit line lengths slows down their transient response due to larger R and C
- Memory chip is partitioned into a number of blocks to improve the transient response

Memory-Chip Timing

- Memory access time: the time between the initiation of a read operation and the data at the output
- Memory cycle time: the minimum time allowed between two consecutive memory operations.
- Access time and cycle time are in the range of a few to few hundred nanoseconds
16.3 RANDOM-ACCESS MEMORY (RAM) CELLS

RAM Cells
- It is imperative to reduce the cell size for a large number of bits on a chip
- The power dissipation should be minimized for RAM cells
- There are two basic types of MOS RAMs:
  - Static (SRAM): utilize static latches as the storage cells. (≈ 6 transistors/cell)
  - Dynamic (DRAM): store binary data on capacitors and require periodic refreshing. (≈ 1T+1C/cell)
- Both static and dynamic RAMs are volatile

Static Memory Cell
- A typical static memory cell comprises a latch (two cross-coupled inverters) and two access transistors
- The access transistors are turned on when the word line is selected
- The complementary bit lines are connected to the latch when the cell is selected

![Static Memory Cell Diagram](image-url)
Read Operation of SRAM

- The complementary bit lines are precharged to an intermediate voltage between $V_{DD}$ and ground
- The bit lines are connected to the latches through access transistors when the cell is selected
- A differential voltage develops between complementary bit lines
- The voltage change $v_G$ and $v_B$ should be sufficiently small not to change the latch state during readout
- The read operation in an SRAM is nondestructive
- Readout delay is determined by the rise time of word line and time needed to develop the required voltage
- The dynamic operation can be approximated by treating $Q_1$ and $Q_5$ as $r_{ON1}$ and $r_{ON5}$
- The larger ($W/L)_5$, the faster the $v_B$ develops
- The smaller ($W/L)_5$, the smaller voltage change for $v_B$
Design Constraint for Read Operation

- The worst case scenario is to choose a precharge value of $V_{DD}$
- $Q_5$ operates in saturation region and $Q_1$ operates in triode region:
  \[ I_5 = \frac{1}{2} \mu_n C_{ox} \left( \frac{W}{L} \right)_5 (V_{DD} - V_m - V_0)^2 = I_1 = \mu_n C_{ox} \left( \frac{W}{L} \right)_1 (V_{DD} - V_m) V_0 - \frac{1}{2} V_0^2 \]
- The design constraint is specified by
  \[ V_0 = (V_{DD} - V_m) \left( 1 - \frac{1}{\sqrt{1 + \frac{(W/L)_5}{(W/L)_1}}} \right) \leq V_m \rightarrow \frac{(W/L)_5}{(W/L)_1} \leq \frac{1}{\left( 1 - \frac{V_m}{V_{DD} - V_m} \right)^2} - 1 \]
- The voltage change in $v_Q$ will be very small
- The $\overline{B}$ line is discharged by $I_5$ at the beginning:
  \[ \Delta V = I_5 \Delta t / C_{\overline{B}} \]
- The time needed to develop a voltage difference $\Delta V$ is
  \[ \Delta t = C_{\overline{B}} \Delta V / I_5 \]
- The $\overline{B}$ line will be finally discharged to 0 in steady state
- Sense amplifiers are usually used to speed up the read operation
Write Operation of SRAM

- The complementary bit-lines are respectively set to $V_{DD}$ and ground before the operation.
- The cell is selected by the word line and the cell is connected to the bit lines thru access transistors.
- $Q_5$ is in saturation and the current for charging capacitor equals to $I_5 - I_1$.
- $I_5$ decreases (due to reduced $v_{GS5}$ and body effect) and $I_1$ increases (due to reduced $v_{GS1}$) as $v_{\bar{Q}}$ increases.
- When $v_\bar{Q}$ and $v_\bar{Q}$ reach $V_{DD}/2$, the regenerative feedback initiates and cause the flip-flop to change state.
- Write delay is determined by the time for $v_\bar{Q}$ and $v_\bar{Q}$ to reach $V_{DD}/2$ and the delay time of the flip-flop.
- The charging (discharging) component of write delay is much smaller than the corresponding component in the read operation because only small capacitance $C_Q$ needs to be charged (discharged).
- The delay time in write operation is dominated by the word-line delay.

\[\text{NTU EE Electronics III} \quad 16-9\]
Design Constraint for Write Operation

- \( Q_4 \) operates in saturation region and \( Q_6 \) operates in triode region:

\[
I_s = \frac{1}{2} \mu_n C_{ox} \left( \frac{W}{L} \right) (V_{DD} - |V_{in}|)^2 = I_6 = \mu_n C_{ox} \left( \frac{W}{L} \right) \left[ (V_{DD} - V_{in}) V_Q - \frac{1}{2} V_Q^2 \right]
\]

- The design constraint is specified by

\[
V_Q = (V_{DD} - V_{in}) \left( 1 - \frac{\mu_p (W / L)_4}{\mu_n (W / L)_6} \right) \leq V_{in} \Rightarrow \frac{(W / L)_4}{(W / L)_6} \leq \frac{\mu_p}{\mu_n} \left[ 1 - \left( 1 - \frac{V_{in}}{V_{DD} - V_{in}} \right)^2 \right]
\]

Design Procedure of the SRAM Cell

- The aspect ratios of the PMOS and NMOS devices in the latch are determined
- Choose the aspect ratio of the access transistors carefully
  - Read operation specifies the upper limit of the device size
  - Write operation specifies the lower limit of the device size
Dynamic RAM Cell

- One-transistor cell: one transistor and one capacitor per cell
- Only one bit line is used in DRAMs
- When the cell is storing a 1, the capacitor is charged to $V_{DD} - V_t$
- When the cell is storing a 0, the capacitor is discharged to 0V
- Refresh operation must be performed every 5 to 10 ms

Read Operation

- The bit lines are precharged to $V_{DD}/2$
- The word line is raised to $V_{DD}$ and all the access transistors in the selected row are conductive
- A small voltage difference will be detected in the bit lines
- The read operation is destructive

$$C_S V_{CS} + C_B \frac{V_{DD}}{2} = (C_B + C_S) \left( \frac{V_{DD}}{2} + \Delta V \right)$$

$$\Delta V = \frac{C_S}{C_B + C_S} \left( V_{CS} - \frac{V_{DD}}{2} \right) \approx \frac{C_S}{C_B} \left( V_{CS} - \frac{V_{DD}}{2} \right) \text{ for } C_B >> C_S$$

Write Operation

- The data bit to be written is set to $V_{DD}$ or ground and $C_S$ will be charged or discharged to $V_{DD} - V_t$ or 0V
- All other cells in the selected row perform write operation at the same time
- Refresh operation: a read operation followed by a write operation for stored data
16.4 SENSE AMPLIFIERS AND ADDRESS DECODERS

Operation of Sense Amplifier

- A sense amplifier is formed by a latch with two cross-coupled CMOS inverters
- $Q_5$ and $Q_6$ connect the sense amplifier to $V_{DD}$ and ground when data-sensing action is required
- Precharge and equalization circuit is controlled by $\phi_p$ prior to a read operation ($Q_8$ and $Q_9$ precharge the bit lines to $V_{DD}/2$; transistor $Q_7$ helps speed up this process by equalizing the voltages on the two lines)
- Complete read operation:
  - The precharge and equalization circuit is activated by $\phi_p$ and complementary bit lines are set to $V_{DD}/2$
  - Word line goes up and the cell is connected to bit lines
  - The sense amplifier turned on by $\phi_s$ as an adequate difference voltage signal is developed between the bit lines by the storage cell
  - The data stored in the cell is refreshed as the sense amplifier pulls the bit lines to $V_{DD}$ and ground
Dynamic Operation of Sense Amplifier

- When the sense amplifier is activated, both inverters are biased in transition region ($V_{IN} = V_{DD}/2$)
- With input signal $v_i$, the resulting output signal $(g_{mp} + g_{mn})v_i$ is fed back to the input (positive feedback)
- The bit line voltage rises/decays exponentially with a time constant of $C_B/G_m$ under small-signal operation
- As the voltage deviates from $V_{DD}/2$, the it tends to saturate at $V_{DD}$ or ground (large-signal operation)

Read-1 operation: $v_B = \frac{V_{DD}}{2} + \Delta V(1)e^{\left(\frac{G_m}{C_B}\right)t}$

Read-0 operation: $v_B = \frac{V_{DD}}{2} - \Delta V(0)e^{\left(\frac{G_m}{C_B}\right)t}$
An Alternative Sense Amplifier

- A differential amplifier with a current mirror as the active load
- The charging/discharging current is defined by the bias current $I$
- The voltage difference developed by the sense amplifier: $\Delta V = \frac{I \Delta t}{C}$
- The voltage required for complete current switching: $\Delta V = \sqrt{2} V_{OV}$

Alternative Precharging Arrangement
Differential Operation in Dynamic RAMs

- Each bit line is split into two identical halves
- Each half-line is connected to half the cells in the column and an additional dummy cell
- The dummy cell serves as the other half of a differential DRAM cell
- During precharge phase, two dummy cells are precharged to $V_{DD}/2$
- Differential signal $\Delta V(1)$ or $\Delta V(0)$ is detected by the sense amplifier when it is enabled

"NTUEE  Electronics III" 16-15
Row-Address Decoder

- The decoder can be realized by NOR functions provided by the matrix structure → NOR decoder
- The word lines are precharged to $V_{DD}$ during precharge
- All the unselected word lines will be discharged
- No static power dissipation due to dynamic operation

\[
W_0 = \overline{A_0 A_1 A_2} = A_0 + A_1 + A_2
\]
\[
W_1 = \overline{A_0 A_1 A_2} = A_0 + A_1 + A_2
\]
\[
W_2 = \overline{A_0 A_1 A_2} = A_0 + A_1 + A_2
\]
\[
W_3 = \overline{A_0 A_1 A_2} = A_0 + A_1 + A_2
\]
\[
W_4 = \overline{A_0 A_1 A_2} = A_0 + A_1 + A_2
\]
\[
W_5 = \overline{A_0 A_1 A_2} = A_0 + A_1 + A_2
\]
\[
W_6 = \overline{A_0 A_1 A_2} = A_0 + A_1 + A_2
\]
\[
W_7 = \overline{A_0 A_1 A_2} = A_0 + A_1 + A_2
\]
**Column-Address Decoder**

- Column-address decoder is to connect one of the $2^N$ bit lines to the data I/O line of the chip.
- It exhibits the same function as a multiplexer.
- Pass-transistor logic decoder:
  - A NOR decoder + pass-transistor multiplexer
- Tree decoder:
  - Utilize smaller number of transistors
  - Speed decreases due to a relatively large number of series transistors in the signal path
Pulse-Generation Circuits

- The ring oscillator:
  - Generates periodic output waveforms
  - Odd number of stages in the loop
  - The oscillation frequency is $1/2Nt_p$

- A one-shot or monostable multivibrator circuit
  - A single output pulse with a predetermined width is provided when triggered
  - The width is determined by $Nt_p$
16.5 READ-ONLY MEMORY (ROM)

MOS ROM

- The circuit can be in a static or a dynamic form
- The data stored in the ROMs is determined at the time of fabrication

Mask-Programmable ROMs

- Mask programmable ROMs can avoid having custom design each ROM from scratch
- MOSFETs are included at all bit locations but only the gates of those transistors where 0’s are to be stored are connected to the word lines
Programmable ROM (PROM)
- The data can be programmed by the user, but only once
- A typical PROM uses polysilicon fuses to connect the emitter of each BJT to the digital line

Erasable Programmable ROM (EPROM)
- The data can be erased and reprogrammed as many times by the uses
- A floating-gate transistor can be used for EPROM at all bit locations as memory cells (stacked-gate cells)
- Programming process:
  - Apply high voltage to the select gate and between the source and the drain
  - Hot electrons are injected into the floating gate and become trapped
  - After the programming process, the programmed transistors exhibit high threshold voltage
  - The process of charging is self-limiting as the charge accumulation increases
Erasure process:
- Illuminate the cell with UV light for a specified duration
- The UV light provides the trapped electrons sufficient energy to transported back to the substrate