Lecture 24
CMOS Logic Gates and Digital VLSI – II

In this lecture you will learn:

- Static CMOS Logic Gates
- FET Scaling
- CMOS Memory, SRAM and DRAM
- CMOS Latches, and Registers (Flip-Flops)
- Clocked CMOS
- CCDs

CMOS Logic: General Architecture

**Pull-up network:**
- Consists of only PFETs
- Charges the output to HIGH

**Pull-down network:**
- Consists of only NFETs
- Discharges the output to LOW
CMOS NAND Gate

When both A and B are HIGH, output is LOW
When either A or B is LOW, output is HIGH

In designing the pull-down network, see when the output ought to be LOW and then arrange the NFETs accordingly.

The pull-up network will have the complimentary topology.

CMOS NOR Gate

When both A and B are LOW, output is HIGH
When either A or B is HIGH, output is LOW

In designing the pull-down network, see when the output ought to be LOW and then arrange the NFETs accordingly.

The pull-up network will have the complimentary topology.
CMOS Gates: Pull Down Network Design

If A and B are both HIGH, output will be LOW

If either A or B is HIGH, output will be LOW

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

CMOS Gates: Pull Up Network Design

In designing the pull-up network, see when the output ought to be HIGH and then arrange the PFETs accordingly. The pull-down network will have the complimentary topology.

If either A or B is LOW, output will be HIGH

If A and B are both LOW, output will be HIGH

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ A \cdot B = A + B \]

De Morgan’s Law

\[ \overline{A + B} = \overline{A} \cdot \overline{B} \]
CMOS Gates: Pull Up and Pull Down Network Design

Pull up and pull down networks are “complimentary” of each other. Hence the name “Complementary MOS” or CMOS!

CMOS Gates: More Complex Logic Gates

Suppose we need to design a logic gate for:

\[ X = (A + B) \cdot C \]

If ((A or B) and C) are HIGH, the output will be LOW
FET Scaling – Inverter

The mobility of electrons in NFETs is generally almost twice that of the holes in PFETs.

One would want the current drives for charging and discharging the output and, consequently, the rise and fall times for the output to be identical (i.e. one would want the NFETs and the PFETs to have the same current drives).

Therefore the W/L ratio of the PFET is chosen to be twice that of the NFET in an inverter.

Fabricated FET Inverter: Dual Well CMOS Technology

![Fabricated FET Inverter Diagram](image)
FET Scaling – NAND Gate

As in the inverter case, one scales the PFETs by $2\alpha$

A HIGH output has to be discharged through the two NFETs in series

Two FETs in series, with the same gate voltage, are like one FET that is twice as long

Therefore, in order to keep the same current drive in discharging a HIGH output in the NAND gate as in the simple inverter, one needs to scale the NFETs by $2\alpha$ each

---

FET Scaling – NOR Gate

As in the inverter case, one scales the NFETs by $1\alpha$

A LOW output has to be charged through the two PFETs in series

Two FETs in series, with the same gate voltage, are like one FET that is twice as long

Therefore, in order to keep the same current drive in charging a LOW output in the NOR gate as in the simple inverter, one needs to scale the PFETs by $4\alpha$ each
The Transmission Gate

A transmission gate allows the logical value to pass from the input to the output only if the gate is OPEN (i.e. the control signal $B$ is HIGH and the switch above is closed).

If the gate is closed (i.e. the control signal $B$ is LOW and the switch above is open) the input and the output are disconnected from each other.
A transmission gate allows the logical value to pass from the input to the output only if the gate is OPEN (meaning the control signal B is HIGH).

If the gate is closed (meaning the control signal B is LOW) the input and the output are disconnected from each other.

**Case I:** Input is sitting at HIGH, the output is sitting at LOW, and the gate opens.

Although both the NFET and the PFET will pass the current, the NFET will cut-off when the output node is still $V_{TN}$ below logical HIGH value (~$V_{DD}$).

So the PFET is required to charge the output to the HIGH value (~$V_{DD}$).

**Case II:** Input is sitting at LOW, the output is sitting at HIGH, and the gate opens.

Although both the NFET and the PFET will pass the current, the PFET will cut-off when the output node is still $V_{TP}$ above the logical LOW value (~0).

So the NFET is required to discharge the output to the LOW value (~0).
Two inverters can be used to realize a bistable memory element.

Both of the following states are allowed:

- \( L \rightarrow H \)
- \( H \rightarrow L \)

Just drawn differently.

Two stable states.
Two inverters can be used together with NFET gates to realize a 6 FET SRAM cell.

Static Random Access Memory (SRAM)

SRAM is fast

Used for implementing fast caches in microprocessors or fast memories in electronic instruments.

The indicated values show the voltages when a logical 1 is being written in the cell shown.
Intel Core-i7 (4 Core) with 8MB L3 Cache (SRAM)

Intel Core-i7 (6 Core) with 12MB L3 Cache (SRAM)

1.17 Billion FETs
Dynamic Random Access Memory (DRAM)

A DRAM cell can be implemented using just one FET and one capacitor.

Dynamic Random Access Memory (DRAM)

The DRAM cell (0.5 µm²) uses stacked cylindrical capacitors with hemispherical silicon grains (HSG) in a capacitor (SAMSUNG).

Types of DRAM Capacitors

- Stack capacitor
- Multi-line (ML)
- Cylinder (CMA - 10)
- Trench capacitor (1M/AMS)
- Planar capacitor (WL - MF)
- Plate capacitor

DRAM is much slower than SRAM
Capacitor can discharge via leakage currents (needs periodic refresh)
Reading a stored bit destroys the stored bit – every read must be followed by a write
Used for implementing large memory in computers
CMOS Latch

- Data passes through from the input ($D$) to the output ($Q$) when the $CLK$ is HIGH (i.e. the latch is transparent)
- Data at the output ($Q$) is latched (and held in place) when the $CLK$ is LOW

CLK = HIGH

CLK = LOW
There are many different ways to realize a latch in CMOS. This is one example...

Suppose the CLK goes HIGH, the input gate opens, and the input data is written into the latch. As long as the CLK is HIGH, $D$ and $Q$ are the same (i.e. the latch is transparent).
CMOS Latch: Operation

When the CLK goes LOW, the input gate closes, and the data written into the latch is held in place (i.e. latched).

CMOS Data Register or a Flip-Flop

- When the CLK goes from LOW to HIGH, input data (D) is transferred to the output (Q) and held in place.

- A data register or a flip-flop can be realized by using two master/slave latches.
Pipelined CMOS Digital Design

FETs and Charge-Coupled Devices (CCDs)

A 24-MP CCD imaging chip

Willard S. Boyle and George E. Smith
(Bell Labs)
(2009 Nobel Prize in Physics)
MOS Structures and Charge-Coupled Devices (CCDs)

Photo-Generation

Charge Separation

FETs and Charge-Coupled Devices (CCDs)

Challenge:
How to get photo-generated charge out of each individual pixel in a simple cost-effective way??
Solution:
Charge-coupled MOS structures can be used as shift-registers to move charge!