Lecture 23
CMOS Logic Gates and Digital VLSI – I

In this lecture you will learn:

- Digital Logic
- The CMOS Inverter
- Charge and Discharge Dynamics
- Power Dissipation
- Digital Levels and Noise

A NFET Inverter

<table>
<thead>
<tr>
<th>A</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Digital Signal Levels

Valid logic levels:
- \( V_{IL} \) = Maximum valid logical LOW input
- \( V_{IH} \) = Minimum valid logical HIGH input
- \( V_{OL} \) = Maximum valid logical LOW output
- \( V_{OH} \) = Minimum valid logical HIGH output

\( V_{IL} \), \( V_{IH} \) and \( V_{OL} \), \( V_{OH} \) are determined by the unity gain points on the transfer curve (Otherwise amplification can corrupt the logic levels as they propagate in a chain)

Gain is necessary to realize logic gates …!!!
Noise Margins

And when noise is present......

One must have:

\[ V_{OL} + \text{noise} < V_{IL} \]

\[ V_{OH} - \text{noise} > V_{IH} \]

Noise margins:

\[ NM_L = V_{IL} - V_{OL} \]

\[ NM_H = V_{OH} - V_{IH} \]

Noise and device variations sets the minimum \( V_{DD} \) one can use

The Ideal Inverter Transfer Curve

A perfectly symmetric curve with a near-vertical transition is an ideal transfer curve because:

- Noise margins can be made very large
- Logical HIGH voltage can be made very small (because the noise margins are so large) resulting smaller power dissipation
Problem:
The input/output characteristics are not symmetric.

Noise margins are good but not excellent.

The Load Capacitance

$C_L$ is the capacitance of the subsequent CMOS stage(s) as well as of the interconnects.
A NFET Inverter: Charging Dynamics

Problem:
When the output is LOW, charging of the output to HIGH is slow because charging current is not uniform.

\[ I_{OUT} = \frac{V_{DD} - V_{OUT}}{R} = C_L \frac{dV_{OUT}}{dt} \]

\[ \Rightarrow V_{OUT}(t) = V_{DD} \left[ 1 - e^{-t/RC_L} \right] \]

A NFET Inverter: Charging and Discharging Dynamics

\( A \rightarrow \overline{A} \)

<table>
<thead>
<tr>
<th>A</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Problem:
When the input is HIGH, and the output is LOW, current keeps flowing through the FET and the resistor forever!!
This is an example of static power dissipation – extremely bad!

What Do We Want ..... Ideally?
Output going from HIGH to LOW
(Constant current discharging)

Output going from LOW to HIGH
(Constant current charging)
If $V_{TN}$ and $V_{DD} + V_{TP}$ are close to each other, the transition region can be made narrow and sharp.

→ The noise margins can be very wide!!

Charging and discharging by (non-ideal) FET current sources is better ........
A CMOS Inverter: No Static Power Dissipation

When the input goes HIGH or LOW, power is only dissipated during the time when the output makes the transition – after this period, there is no power dissipation.

There is no static power dissipation (ideally!), only dynamic power dissipation!!

Current

Rise Times, Fall Times and Propagation Delays

\[ t_r = \text{Rise time b/w 10\% to 90\% of the total upward swing} \]

\[ t_f = \text{Fall time b/w 90\% to 10\% of the total downward swing} \]

\[ t_{PHL} = \text{Propagation delay for H→L b/w 50\% points} \]

\[ t_{PLH} = \text{Propagation delay for L→H b/w 50\% points} \]
When the output is LOW, initial charging of the output to HIGH is done with a uniform current supplied by the PFET in saturation:

\[
I_{OUT} = C_L \frac{dV_{OUT}}{dt}
\]

\[
\Rightarrow k_p \left( V_{IN} - V_{DD} - V_{TP} \right)^2 = C_L \frac{dV_{OUT}}{dt}
\]

\[
\Rightarrow k_p \frac{1}{2} \left( V_{DD} + V_{TP} \right)^2 = C_L \frac{dV_{OUT}}{dt}
\]

\[
\Rightarrow V_{OUT}(t) = \frac{k_p}{2C_L} \left( V_{DD} + V_{TP} \right)^2 t
\]

Condition for the PFET to be in saturation:

\[
V_{DS} < V_{GS} - V_{TP}
\]

\[
\Rightarrow V_{OUT} - V_{DD} < V_{IN} - V_{DD} - V_{TP}
\]

\[
\Rightarrow V_{OUT} < V_{IN} - V_{TP} = -V_{TP}
\]

When \( V_{OUT} \) becomes larger than \(-V_{TP} \) then the PFET goes into the linear region....

For times \( 0 < t < t_1 \) when the PFET (M2) is in saturation:

\[
V_{OUT}(t) = \frac{k_p}{2C_L} \left( V_{DD} + V_{TP} \right)^2 t
\]

\[
\Rightarrow t_1 = -\frac{2C_L}{k_p \left( V_{DD} + V_{TP} \right)^2} = -\frac{2V_{TP}}{(V_{DD} + V_{TP})^2} \tau
\]

\[
\tau = \frac{C_L}{k_p \left( V_{DD} + V_{TP} \right)}
\]
**A CMOS Inverter: Charging Dynamics**

When $V_{OUT}$ becomes larger than $-V_{TP}$ then the PFET goes into the linear region, and from then onwards:

$$I_{OUT} = C_L \frac{dV_{OUT}}{dt}$$

$$\Rightarrow k_p \left( V_{IN} - V_{DD} - V_{TP} - \frac{(V_{OUT} - V_{DD})}{2} \right) (V_{OUT} - V_{DD}) = C_L \frac{dV_{OUT}}{dt}$$

$$\Rightarrow -k_p \left( V_{DD} + V_{TP} + \frac{(V_{OUT} - V_{DD})}{2} \right) (V_{OUT} - V_{DD}) = C_L \frac{dV_{OUT}}{dt}$$

One can obtain faster charging compared to a resistor in place of a PFET!

---

**Second Page**

$$-k_p \left( V_{DD} + V_{TP} + \frac{(V_{OUT} - V_{DD})}{2} \right) (V_{OUT} - V_{DD}) = C_L \frac{dV_{OUT}}{dt}$$

$$\Rightarrow \int_{-V_{TP}}^{V_{OUT}} \frac{dV_{OUT}}{V_{OUT} + V_{TP} + \frac{(V_{OUT} - V_{DD})}{2} (V_{OUT} - V_{DD})} = \frac{k_p}{C_L} \int_{t_1}^{t} dt = \frac{k_p t}{C_L}$$

$$\Rightarrow V_{OUT}(t > t_1) = V_{DD} \left( 1 - e^{-\frac{t-t_1}{\tau}} \right) - 2V_{TP} \frac{e^{-\frac{(t-t_1)}{\tau}}}{1 + e^{-\frac{(t-t_1)}{\tau}}}$$

$$\tau = \frac{C_L}{k_p (V_{DD} + V_{TP})}$$
A CMOS Inverter: Charging Dynamics

One can obtain faster charging compared to a resistor:

\[
\tau_r = \frac{2C_L}{k_p(V_{DD}+V_{TP})} \left\{ \frac{-V_{TP}}{V_{DD}+V_{TP}} - \frac{1}{2} \ln \left[ \frac{0.1V_{DD}}{2(V_{DD}+V_{TP})-0.1V_{DD}} \right] \right\}
\]

\[
\propto \frac{L^2}{\mu_pV_{DD}} \propto \tau_t \quad \text{if } C_L \propto C_{gs}
\]

In reality, the load capacitance is not just due to the next FET gate - it also includes the interconnect capacitances.

A CMOS Inverter: Discharging Dynamics

One can obtain faster discharging compared to a resistor:

\[
\tau_f = \frac{2C_L}{k_n(V_{DD}-V_{TN})} \left\{ \frac{V_{TN}}{V_{DD}-V_{TN}} - \frac{1}{2} \ln \left[ \frac{0.1V_{DD}}{2(V_{DD}-V_{TN})-0.1V_{DD}} \right] \right\}
\]

\[
\propto \frac{L^2}{\mu_nV_{DD}} \propto \tau_f \quad \text{if } C_L \propto C_{gs}
\]

In reality, the load capacitance is not just due to the next FET gate - it also includes the interconnect capacitances.
A CMOS Inverter: Charging and Discharging Dynamics

How to charge and discharge faster?
- Decrease \( L \)
- Increase charging current
  \[ \rightarrow \text{Increase supply voltage } V_{DD} \]

At what price?

Assuming \( C_L \) is not dominated by interconnect capacitance (not generally true), the only way to increase \( k_n \) and \( k_p \) of FETs and at the same time decrease \( C_{gs} \) is to decrease the FET length \( L \):

\[
t_f \propto \frac{L^2}{\mu \cdot V_{DD}}
\]

Intel FET Gate length Trends

Smaller transistor provides:
- Higher performance
- Lower power
- Lower cost per FET

Mark Bohr, Intel (2014)
Q: How much energy is dissipated (in the PFET and the wires) in charging the capacitor to HIGH from LOW?

A: Irrespective of how it is charged, the net energy dissipation in charging a capacitor equals the energy stored in the capacitor after charging!

\[ E_D = \frac{1}{2} C_L V_{DD}^2 \]

Q: How much energy is dissipated (in the NFET and the wires) in discharging the capacitor from HIGH to LOW?

A: Irrespective of how it is discharged, the net energy dissipation in discharging a capacitor equals the energy stored in the capacitor before discharging!

\[ E_D = \frac{1}{2} C_L V_{DD}^2 \]

Total energy dissipation in one charge and discharge cycle:

\[ E_D = C_L V_{DD}^2 \]
**Thermodynamics, Entropy, Information, and Computation**

**Question:** How much energy does it require to compute or process one bit of information?

The question was answered by Rolf W. Landauer (1927-1999) (IBM)

Any thermodynamically irreversible operation that manipulates information increases entropy, and an associated amount of energy is unavoidably dissipated as heat.

The minimum amount of energy needed to process or compute one bit of information equals:

$$KT \log(2)$$

$$KT \log(2) \approx 17.9 \text{ meV at room temperature}$$

For the smallest CMOS inverter Intel has:

$$E_D = C_L V_{DD}^2 \approx 30 \text{ eV}$$

$$C_L = 2 \times C_{gs} = 2 \times 2 \times \frac{2}{3} \epsilon_{ox} \frac{W}{L} \approx 2 \times 10^{-17} \text{ Farads}$$

$$V_{DD} \approx 0.5 \text{ V}$$

This is almost ~1600 times larger than the fundamental thermodynamic limit ...!!!

So there is plenty of room for improvement...!!!

**Dynamic Power Dissipation in CMOS Chips**

Total energy dissipation in one charge and discharge cycle per FET:

$$E_D = C_{gs} V_{DD}$$

Total energy dissipation in one charge and discharge cycle if $$N_{FET}$$ FETs in the chip are active:

$$E_D = N_{FET} C_{gs} V_{DD}^2$$

Total power dissipation (energy dissipation per second) if $$N_{FET}$$ FETs are active:

$$P_D = N_{FET} f_{CLK} C_{gs} V_{DD}^2$$

Number of cycles per second ~ $$f_{CLK}$$
Dynamic Power Dissipation in CMOS Chips

Clock speed = 3.16 GHz  
Gate length: 45 nm = .045 \mu m  
Number of FETs in the chip = 410 \times 10^6  
Power supply voltage ~ 2 V  

\[ P_D = N_{FET} f_{CLK} C_{gs} V_{DD}^2 \]

Fraction of active FETs at any instant on the average (~0.4%)

\[ N_{FET} = 0.004 \times 410 \times 10^6 = 1.64 \times 10^6 \]
\[ f_{CLK} = 3.1 \times 10^9 \]
\[ t_{ox} = 10 \AA \text{ (equivalent low-\kappa thickness)} \]
\[ W = 2 \mu m \]
\[ L = 0.045 \mu m \]
\[ C_{gs} = \frac{2 \varepsilon_{ox}}{3} W L = 2 \text{ femto - Farads} \]
\[ V_{DD} = 2 \text{ V} \]

Our power dissipation estimate:
\[ P_D = N_{FET} f_{CLK} C_{gs} V_{DD}^2 \]
\[ = 42 \text{ Watts!!} \]

Actual published number:
\[ P_T = 65 \text{ Watts!!} \]
\[ = P_D + P_S \]

Intel FET Gate length Trends

Smaller transistor provides:
- Higher performance
- Lower power
- Lower cost per FET
Recent trends:
1) Number of FETs/chip keeps increasing
2) Clock frequency is not increasing
3) Number of switching operation per sec is limited by our ability to remove heat from the chip (which at the moment is ~100 W/cm²)
4) And most FETs remain inactive most of the time in modern chips