4.10 The CMOS
Digital Logic Inverter

Reading Assignment: pp. 336-346

Complementary MOSFET (CMOS) is the predominant technology for constructing IC digital devices (i.e., logic gates).

Q:

A:

HO: The CMOS Inverter
HO: The CMOS Transfer Function
HO: Peak CMOS Current

HO: The CMOS Model
HO: CMOS Propagation Delay
The CMOS Inverter

Consider the complementary MOSFET (CMOS) inverter circuit:

In this circuit:

\[ i_{DP} = i_{DN} = i_D \]
\[ K_n = K_p = K \]
\[ V_{tp} = V_{tn} = V_T \]
\[ (V_{DD} > 2V_T) \]

Q: Why do we call it "Complementary"?

A: Because the device consists of an NMOS and PMOS transistor, each with equal \( K \) and equal but opposite \( V_T \).

Q: What makes the CMOS inverter so great?

A: Let's analyze the circuit and find out!
First, let’s consider the case where the input voltage is at the perfect “high” state \( v_I = V_{DD} \).

For this case, it is readily apparent that:

\[ v_{GSN} = V_{DD} \quad \text{and} \quad v_{GSP} = 0.0 \, V \]

Hence, we can conclude:

\[ v_{GSN} = V_{DD} > V_{tn} \quad \rightarrow \quad Q_N \text{ has an induced channel!} \]

and:

\[ v_{GSP} = 0.0 \, V > V_{tp} \quad \rightarrow \quad Q_P \text{ has no induced channel!} \]

Thus, we can conclude that \( Q_P \) is in cutoff, and \( Q_N \) is either in saturation or triode.

Let’s ASSUME that \( Q_N \) is in triode, so we ENFORCE the condition that:

\[ i_D = K_n \left[ 2(v_{GSN} - V_{tn})v_{DSN} - v_{DSN}^2 \right] \]
Note that:

\[ v_{GSN} = v_I = V_{DD} \quad \text{and} \quad v_{DSN} = V_O \]

Therefore:

\[ i_D = K_n \left[ 2 \left( v_{GSN} - V_m \right) v_{DSN} - V_{DSN}^2 \right] \]

\[ = K \left[ 2 \left( V_{DD} - V_t \right) v_O - V_O^2 \right] \]

Now, we actually KNOW that \( Q_P \) is in cutoff, so we likewise ENFORCE:

\[ i_D = 0.0 \]

Equating these two ENFORCED conditions, we find that:

\[ i_D = 0 = K \left[ 2 \left( V_{DD} - V_t \right) v_O - V_O^2 \right] \]

Solving, we find that the output voltage must be zero!

\[ v_O = v_{DSN} = 0.0 \ \text{V} \]

Thus, \( v_{DSN} = 0 < v_{GSN} - V_m = V_{DD} - V_t \).

\( Q_N \) is indeed in the triode mode!
So, let's summarize. When $v_I = V_{DD}$:

1. $Q_P$ is in cutoff and $Q_N$ is in triode
2. $i_D = 0.0$
3. $v_O = 0.0$

Now, let's consider the case where the input voltage is at the perfect “low” state $v_I = 0.0$.

For this case, it is readily apparent that:

$v_{GSN} = 0.0$ and $v_{GSP} = -V_{DD}$

$v_{GSN} = 0.0$ and $v_{GSP} = -V_{DD}$

$v_I = 0.0$ and $i_D = 0.0$

$v_O = 0.0$ and $v_{DSN} = -V_{DD}$
Hence, we can conclude:

\[ V_{GSN} = 0.0 < V_{tn} \rightarrow Q_N \text{ has no induced channel!} \]

and:

\[ V_{GSP} = -V_{DD} < V_{tp} \rightarrow Q_P \text{ has an induced channel!} \]

Thus, we can conclude that \( Q_N \) is in cutoff, and \( Q_P \) is either in saturation or triode.

Let’s ASSUME that \( Q_P \) is in triode, so we ENFORCE the condition that:

\[
\begin{align*}
    i_D &= K_p \left[ 2(V_{GSP} - V_{tp})V_{DSP} - V_{DSP}^2 \right] \\
    \text{Note that:} \\
    V_{GSP} &= V_I - V_{DD} = -V_{DD} \quad \text{and} \quad V_{DSP} = V_O - V_{DD} \\
    \text{Therefore:} \\
    i_D &= K_p \left[ 2(V_{GSP} - V_{tp})V_{DSP} - V_{DSP}^2 \right] \\
    &= K \left[ 2(V_I - V_{DD})(V_O - V_{DD}) - (V_O - V_{DD})^2 \right]
\end{align*}
\]

Now, we actually KNOW that \( Q_N \) is in cutoff, so we likewise ENFORCE:

\[ i_d = 0.0 \]
Equating these two ENFORCED conditions, we find that:

\[ i_D = 0.0 = K \left[ 2(V_I - V_{DD})(V_O - V_{DD}) - (V_O - V_{DD})^2 \right] \]

Solving, we find that the output voltage must be \( V_{DD} \)!

\[ v_O = V_{DD} \]

Thus, \( v_{DSP} = 0 > v_{GSP} - V_{tp} = V_I - V_{DD} \).

\( Q_P \) is indeed in the triode mode!

So, let’s summarize. When \( v_I = 0.0 \):

1. \( Q_N \) is in cutoff and \( Q_P \) is in triode
2. \( i_D = 0.0 \)
3. \( v_O = V_{DD} \)

\[ \begin{align*}
\text{Conditions} & \quad \text{Result} \\
V_{DD} & \quad \text{Input voltage} \\
Q_P & \quad \text{Transistor} \\
\downarrow & \quad \text{Connect} \\
0 & \quad \text{Output voltage} \\
Q_N & \quad \text{Transistor} \\
\downarrow & \quad \text{Connect} \\
0 & \quad \text{Input voltage} \\
\end{align*} \]
So, the overall behavior of the CMOS inverter is displayed in this table:

<table>
<thead>
<tr>
<th>$v_I$</th>
<th>$v_O$</th>
<th>$i_D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>$V_{DD}$</td>
<td>0.0</td>
</tr>
<tr>
<td>$V_{DD}$</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Look at what this means! The CMOS inverter provides lots of ideal inverter parameters:

\[
V_{OL} = 0.0 \text{ V} \quad \text{(ideal!)}
\]

\[
V_{OH} = 5.0 \text{ V} \quad \text{(ideal!)}
\]

And since $i_D$ is zero for either state, the static power dissipation is likewise zero:

\[
P_{D \ static} = 0.0 \quad \text{(ideal!)}
\]

This is one of the most attractive features of CMOS digital logic.
The CMOS Transfer Function

Now, instead of determining the output $v_O$ of a CMOS inverter for just two specific input voltages ($v_I = 0$ and $v_I = V_{DD}$), we can determine the value of $v_O$ for any and all input voltages $v_I$—in other words, we can determine the CMOS inverter transfer function $v_O = f(v_I)$!

Determining this transfer function is a bit laborious, so we will simply present the result (the details are in your book):
Look at how close this transfer function is to the ideal transfer function!

The transition region for this transfer function is very small; note that:

1. $V_{IL}$ is just a bit less than $V_{DD}/2$

2. $V_{IH}$ is just a bit more than $V_{DD}/2$

In fact, by taking the derivative of the transfer function, we can determine the two points on the transfer function (i.e., $V_{IL}$ and $V_{IH}$) where the slope is equal to -1.0. I.E.:

$$\nu_I \text{ where } \frac{d\nu_O}{d\nu_I} = -1.0$$

Taking this derivative and solving for $\nu_I$, we can determine explicit values for $V_{IL}$ and $V_{IH}$ (again, the details are in your book):

$$V_{IL} = \frac{1}{8}(3V_{DD} + 2\nu_r)$$

$$V_{IH} = \frac{1}{8}(5V_{DD} - 2\nu_r)$$
Now, recall earlier we determined that the CMOS inverter provides ideal values for $V_{OL}$ and $V_{OH}$:

$$
V_{OL} = 0.0
$$
$$
V_{OH} = V_{DD}
$$

Thus, we can determine the noise margins of a CMOS inverter:

$$
\mathcal{NM}_{L} = V_{IL} - V_{OL} = \frac{1}{8}(3V_{DD} + 2V_{r}) - 0.0
$$
$$
= \frac{1}{8}(3V_{DD} + 2V_{r})
$$

and:

$$
\mathcal{NM}_{H} = V_{OH} - V_{IH} = V_{DD} - \frac{1}{8}(5V_{DD} - 2V_{r})
$$
$$
= \frac{1}{8}(3V_{DD} + 2V_{r})
$$

Therefore, the two noise margins are equal, and thus we can say that the noise margin for a CMOS inverter is:

$$
\mathcal{NM}_{L} = \mathcal{NM}_{H} = \frac{1}{8}(3V_{DD} + 2V_{r})
$$
**Peak CMOS Current**

**Q:** What do you mean, "peak CMOS current"? I thought that the drain current of a CMOS inverter was $i_D=0$??

**A:** The drain current $i_D$ is zero specifically when $v_I = 0$ or $v_I = V_{DD}$. But, consider when $v_I$ is some value between 0 and $V_{DD}$.
Note it is apparent from the transfer function that:

1. If \( V_t < v_I < V_{DD}/2 \), then \( Q_N \) is in saturation and \( Q_P \) is in triode.

2. If \( v_I = V_{DD}/2 \), then \( Q_N \) and \( Q_P \) are both in saturation.

3. If \( V_{DD}/2 < v_I < (V_{DD} - V_t) \), then \( Q_N \) is in triode and \( Q_P \) is in saturation.

Note that for each of these three cases, a conducting channel is present in both transistors \( Q_N \) and \( Q_P \).

The drain current \( i_D \) is therefore non-zero !!!

Note that the peak current \( i_D^{peak} \) occurs when \( v_I = V_{DD}/2 \).
A: The answer is rather obvious! The peak current occurs when \( v_I = \frac{V_{DD}}{2} \). For that situation, we know that both transistor \( Q_N \) and \( Q_P \) are in saturation—and we know the current through a MOSFET when in saturation is:

“K times the excess gate voltage squared”

For this case, \( v_{GSN} = v_I = \frac{V_{DD}}{2} \), thus:

\[
\begin{align*}
    i_D^{\text{peak}} &= K_n (v_{GSN} - V_t)^2 \\
    &= K (V_{DD}/2 - V_t)^2
\end{align*}
\]

If we wish to minimize the dynamic power dissipation \( P_D \), then we need to minimize this current value (e.g., minimize \( K \), or maximize \( V_t \)).
The CMOS Model

Note that all our analysis of a CMOS inverter has been for the case where the output is connected to an open circuit, for example:

Q: What happens if we connect the CMOS inverter output to something?
A: Note that now we have a very different circuit (e.g., \( i_{dp} \neq i_{dn} \)). We must use **CMOS model** to analyze the circuit!

Let's again look at the case with an open circuit at the output:

In this case, \( Q_p \) is in Triode and \( Q_n \) is in Cutoff.

In other words, the channel in the NMOS device is not conducting, but the channel in the PMOS device is conducting.

Moreover, since \( \nu_{DSP} \) is small (i.e., \( \nu_{DSO} = \nu_o - \nu_{DD} = 0 \) \(<\) really small!!!), we can use the channel resistance approximation:

\[
\frac{-\nu_{DSP}}{i_{dp}} \approx r_{DSP} = \frac{-1}{2K(\nu_{GSP} - \nu_p)}
\]

In other words, the channel in the PMOS device acts like a resistor with resistance \( r_{DSP} \) ! Since \( \nu_{GSP} = -\nu_{DD} \), we find:

\[
r_{DSP} = \frac{-1}{2K(-\nu_{DD} - \nu_p)} = \frac{1}{2K(\nu_{DD} + \nu_p)} = \frac{1}{2K(\nu_{DD} - \nu_t)}
\]
We can thus **model** the CMOS inverter as:

Note for open output circuit, we get the correct answer from this model:

\[ i_D = 0 \quad \text{and} \quad v_O = V_{DD} \]

Now, let's see what happens if the output is **not** open circuited—let's place a load **resistor** at the output!
From the model circuit, we see that:

\[ i_D = i_{DP} = \frac{V_{DD} - 0}{r_{DSP} + R} = \frac{V_{DD}}{r_{DSP} + R} \]

\[ i_{DN} = 0 \]

\[ v_O = i_O R = V_{DD} \left( \frac{R}{r_{DSP} + R} \right) \]

Note, if \( R >> r_{DSP} \), then \( v_O \) will be slightly less than \( V_{DD} \)!

Note these are approximate values of \( i_{DP} \) and \( v_O \), but if we solved the CMOS circuit directly (with no approximations), we would get answers very close to these.

In other words, the CMOS model is an accurate approximation provided that \( v_{DSP} \) is small enough.

**Q:** What happens if the input voltage to the CMOS inverter is high (i.e., \( v_I = V_{DD} \))??

In that case, PMOS device is in cutoff and the NMOS device is in triode.

Therefore the CMOS model for this condition is:
Note that $r_{DSN} = r_{DSP}$ in for the two CMOS models!

The CMOS Model for $v_I = V_{DD}$.

$$r_{DSN} = \frac{1}{2K(V_{GSN} - V_{m})}$$

$$= \frac{1}{2K(V_{DD} - V_{t})}$$
CMOS Propagation Delay

The CMOS model can likewise be used to estimate the propagation delay of a CMOS inverter.

To see how, consider a CMOS inverter with its output at low level $v_O=0.0$ (i.e., its input is $v_I=5.0$). The voltage across the output capacitance $C$ is likewise zero:

\[ V_{DD} \quad r_{DS} \quad C \quad v_O = 0 \quad - \]

**Q:** Output capacitance?! What on Earth is that?

**A:** The output capacitance of a CMOS inverter is simply a value that represents the total capacitance associated with the inverter output. This includes the internal capacitances of the MOSFET devices, the wiring capacitance, and the capacitance of the device that the output is connected to!

Figure 10.6 (p. 958) - Schematic showing all capacitances associated with the output of a CMOS inverter.
Now, say the input of the inverter instantaneously changes to a low level $v_I=0.0$. Ideally, the output would likewise instantaneously change to a high level $v_C=V_{DD}$. However, because of the output capacitance, the output voltage will not instantaneously change state, but instead will change "slowly" with time.

Specifically, using the MOSFET model, we can determine the output voltage as a function of time:

From KCL:

$$\frac{V_{DD} - v_O(t)}{r_{DSP}} = C \frac{dv_O(t)}{dt}$$

resulting in the differential equation:

$$V_{DD} - v_O(t) - r_{DSP} C \frac{dv_O(t)}{dt} = 0$$

Solving this differential equation, using the initial condition $v_C(t)=0.0$, we get:

$$v_O(t) = V_{DD} \left(1 - e^{-t/\tau}\right)$$

where $\tau$ is the time constant $\tau = r_{DSP} C$. 
The time constant $\tau$ is therefore the approximate time it takes for the output voltage to change after the input has changed. As such, the value $\tau$ is **approximately** the propagation delay $t_p$ of the CMOS inverter!

$$t_p \approx r_{DS_{P}} C = \frac{C}{2K(V_{DD} - V_t)}$$

**Look** at what this means! We can **reduce** the propagation delay of a CMOS inverter by either:

1. Increasing $K$
2. Decreasing $V_t$
3. Decreasing $C$
Q: But wait! Didn’t you say earlier that if we increase $K$ or decrease $V_t$, we will increase peak current $i_{D}^{\text{max}}$, and thus increase the dynamic power dissipation $P_D$?

A: True! That is the dilemma that we face in all electronic design—increasing the speed (i.e., decreasing the propagation delay) typically results in higher power dissipation, and vice versa.

Our only option for decreasing the propagation delay without increasing the power dissipation is to decrease the output capacitance $C$!

Output capacitance $C$ can be reduced only by decreasing the size of the MOSFET devices—making the devices smaller likewise make them faster!