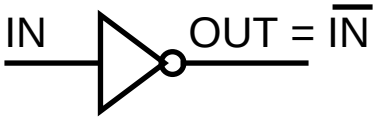
Chapter 3

CMOS Inverter and Multiplexer

3.1 Basic characterization of the CMOS inverter

An inverter is the simplest logic gate which implement the logic operation of negation. A logic symbol and the truth/operation table is shown in Figure 3.1. Two logic symbols, ‘0’ and ‘1’ are represented by



IN	OUT	V_{IN}	V_{OUT}
0	1	V_L	V_H
1	0	V_H	V_L

Figure 3.1: An inverter: logic symbol and the truth/operation table

two voltages ‘V_L’ and ‘V_H’.

A CMOS inverter is an ingenious circuit which is built form a pair of nMOS and pMOS transistors operating as complementary switches as illustrated in Figure3.2. The main advantage of a CMOS inverter over

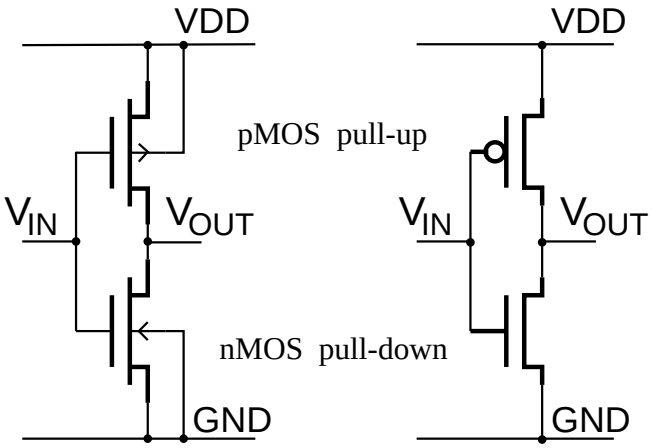


Figure 3.2: Schematic diagrams of a CMOS inverter

many other solutions is that it is built exclusively out of transistors operating as switches, without any other passive elements like resistors or capacitors.

From Figure 3.2 note that the pMOS (pull-up transistor) is connected between V_{DD} and the output node, V_{OUT}, whereas the nMOS (pull-down transistor) is connected between the output node, V_{OUT}, and the ground, GND.

The principle of operation is as follows (refer also to the right part of Figure 3.3).

- for small values of the input voltage, V_{IN}, the nMOS transistor is switched off, whereas the pull-up pMOS transistor is switched on and connects the output mode to V_{DD}
- for large values of the input voltage, V_{IN}, the pMOS transistor is switched off, whereas the pull-down nMOS transistor is switched on and connects the output mode to GND = 0V.

A better inside into the working of the CMOS inverter can be obtain by looking at its transfer and current characteristics presented in Figure 3.3.

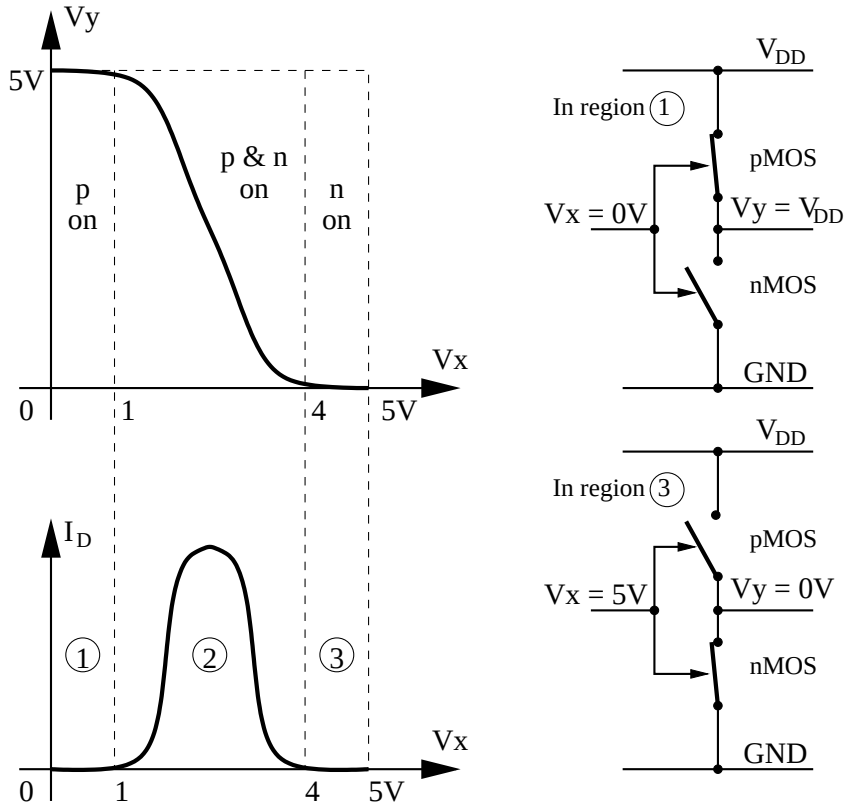


Figure 3.3: The transfer (top-left) and current (bottom-left) characteristics of a CMOS inverter

The **transfer characteristic** presents the output voltage V_{OUT} versus the input voltage V_{IN}. Note that when the input voltage increase from 0V to 5V the output voltage decreases from 5V to 0V.

The **current characteristic** presents the current flowing through the transistors between V_{DD} and GND also versus the input voltage V_{IN}.

From the above characteristics we can observe the existence of three basic regions of operations denoted 1, 2, 3 in Figure 3.3.

- In **region 1** when $0 \leq V_{IN} < V_{T_N}$
  the nMOS transistor is cut off, the pMOS switch is closed and
  $V_{OUT} = V_{DD} \quad I_D = 0$
- In **region 3** when $V_{IN} > V_{DD} - V_{T_P}$
  the pMOS transistor is cut off, the nMOS switch is closed and
  $V_{OUT} = 0\text{V} \quad I_D = 0$

The fact that in regions 1 and 3 NO current flows between V_{DD} and GND, is very attractive because there is no power dissipation at this stages. This very fact is the reason that all digital circuitry is now build in the CMOS technology.

In **region 2** when $V_N < V_{IN} < V_P$

- the transistor remains only for a short period of time, when the input voltage switches between V_L and V_H.
- In this region there is non-zero current flowing between V_{DD} and GND, and some power dissipation, which is converted into heat.

Note that the same current flows through the pMOS and nMOS transistors, that is,

$$I_{Dp} = I_{Dn} \tag{3.1}$$

3.1.1 Recommended relative size of pMOS and nMOS transistors

In order to build a **symmetrical inverter** the midpoint of the transfer characteristic must be centrally located, that is,

$$V_{IN} = \frac{1}{2} V_{DD} = V_{OUT} \tag{3.2}$$

For that condition both transistors are expected to work in the saturation mode. Now, if we combine eqn (3.1) with eqns (3.2) and (2.7) we can write the following:

$$g_n(0.5V_{DD} - V_{Tn})^2 = g_p(0.5V_{DD} - V_{Tp})^2 \tag{3.3}$$

If the threshold voltages are equal, the above condition for a symmetrical inverter is reduced to the condition of equality of their transconductance parameters defined in eqn (2.5), that is:

$$g_n = g_p \tag{3.4}$$

Hence we finally have:

$$\left(\frac{W}{L}\right)_p = \frac{\mu_n}{\mu_p}\left(\frac{W}{L}\right)_n \approx 2.5\left(\frac{W}{L}\right)_n \tag{3.5}$$

Therefore, in order to preserve symmetry of I/O voltages, the pMOS transistor should be 2.5 times wider than the nMOS of the same length. This ratio is the result of different mobility of positive and negative carriers as in eqn (2.3).

3.2 Layout of the CMOS Inverter

A circuit layout of a CMOS inverter can be obtain by joining appropriately the pMOS and nMOS circuits presented in Figure 2.12.

This layout does not take into account the different sizes of the pMOS and nMOS transistors require to have a symmetrical transient behaviour of the inverter. We need also intermediate metal path to connect different types of diffusions.

In addition, taking into account more complex circuitry derived from the inverter concept, we usually arrange diffusions in two parallel horizontal paths.

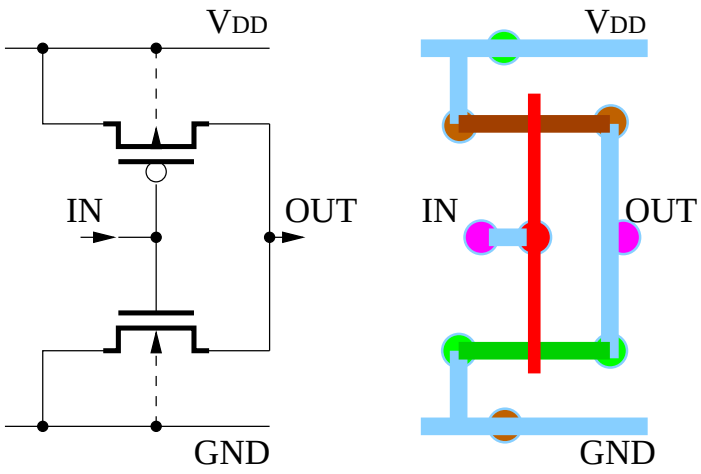


Figure 3.4: A schematic and a stick diagram of a CMOS inverter

As a result we can create a more realistic layout of a CMOS inverter. We start with a modified schematic and with a stick diagram as presented in Figure 3.4.

Note the horizontally oriented transistors, the vertical polysilicon path forming the transistor gates and connecting them together. Two different types of diffusion are connected using a vertical metal paths.

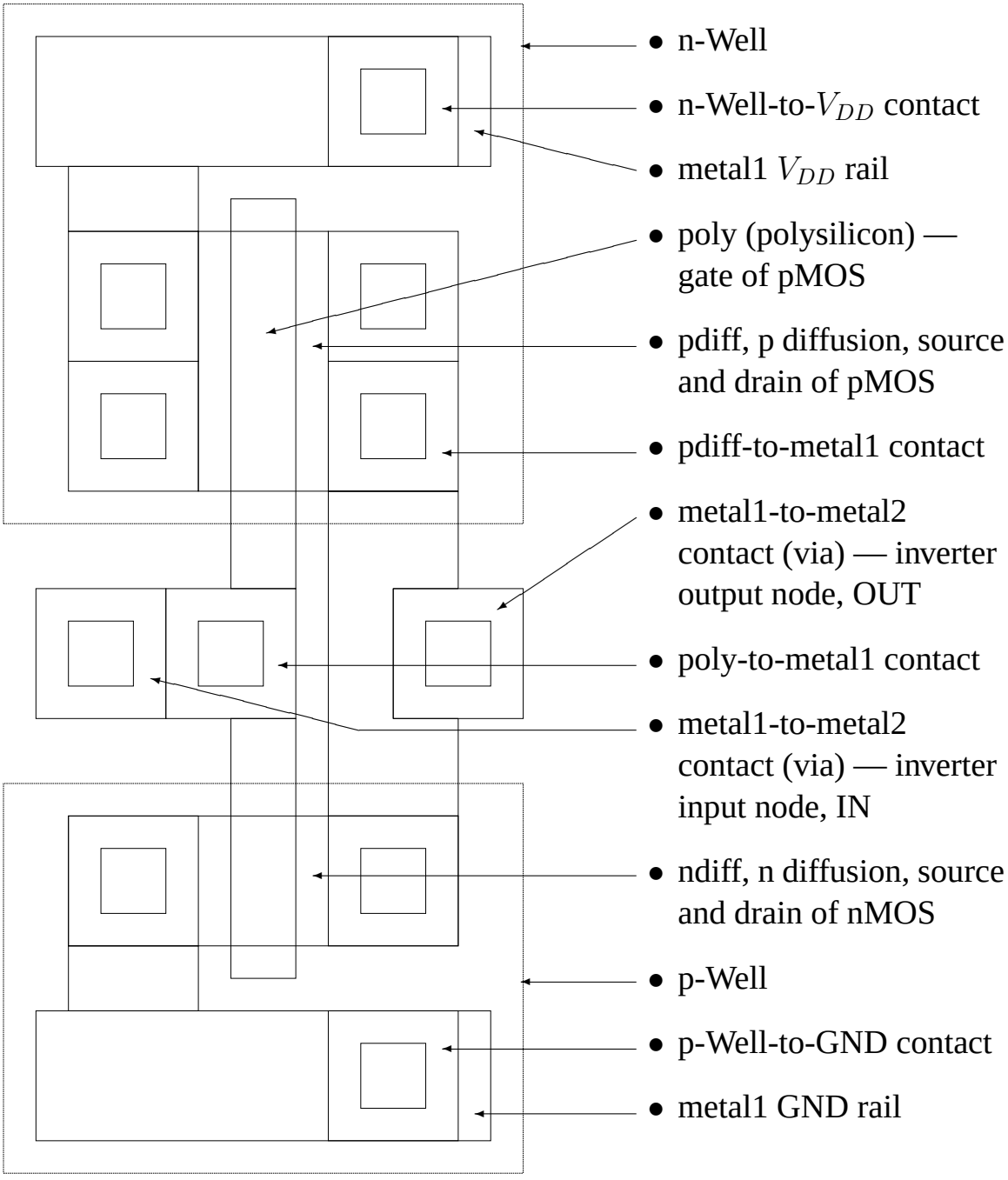


Figure 3.5: A realistic layout of a CMOS inverter

The stick diagram can now be converted into a realistic, but still a bit simplified circuit layout presented in Figure 3.5.

Next to the inverter layout of Figure 3.5 we list its 13 components, most of which can be also found in the schematic and the stick diagram presented in Figure 3.4. Note that a relatively large part of the layout is taken by contacts connecting circuit elements existing in different layers.

Comparing layouts of Figures 2.12 and 3.5 we note the following differences.

- The area around a transitor is referred to as a well. The pMOS is inside n-Well and the nMOS inside the p-Well. Often one of these wells is actually the substrate of a required type. The well is an area created in a substrate to house transistors of the appropriete type. Further details will be discussed in ch. ??
- Note that contacts have now more detailed layout with a small $2\lambda \times 2\lambda$ square inside the $4\lambda \times 4\lambda$ area. We will discuss details of contacts in the next section.
- Note that in relative term, if contacts are of the size $4 \times 4\lambda$, then the pMOS has the ratio $W/L = 8/2$ and the nMOS has $W/L = 4/2$. In other words, the pMOS is twice as wide as the nMOS.
- For the first time we indicated a possibility of having two metal layers. Note that the metal1-to-metal2 contacts, also known as the vias are used for the input-output terminals. The significance of this will be explained later on.
- Note that we have well-to-power contacts connecting the n-Well to V_{DD} and the p-Well to GND, respectively.

Finally, in Figure 3.6, we present an inverter layout created by the Mentor Graphics IC Station in the AMI0.5μ technology. This is an inverter ready for fabrication.

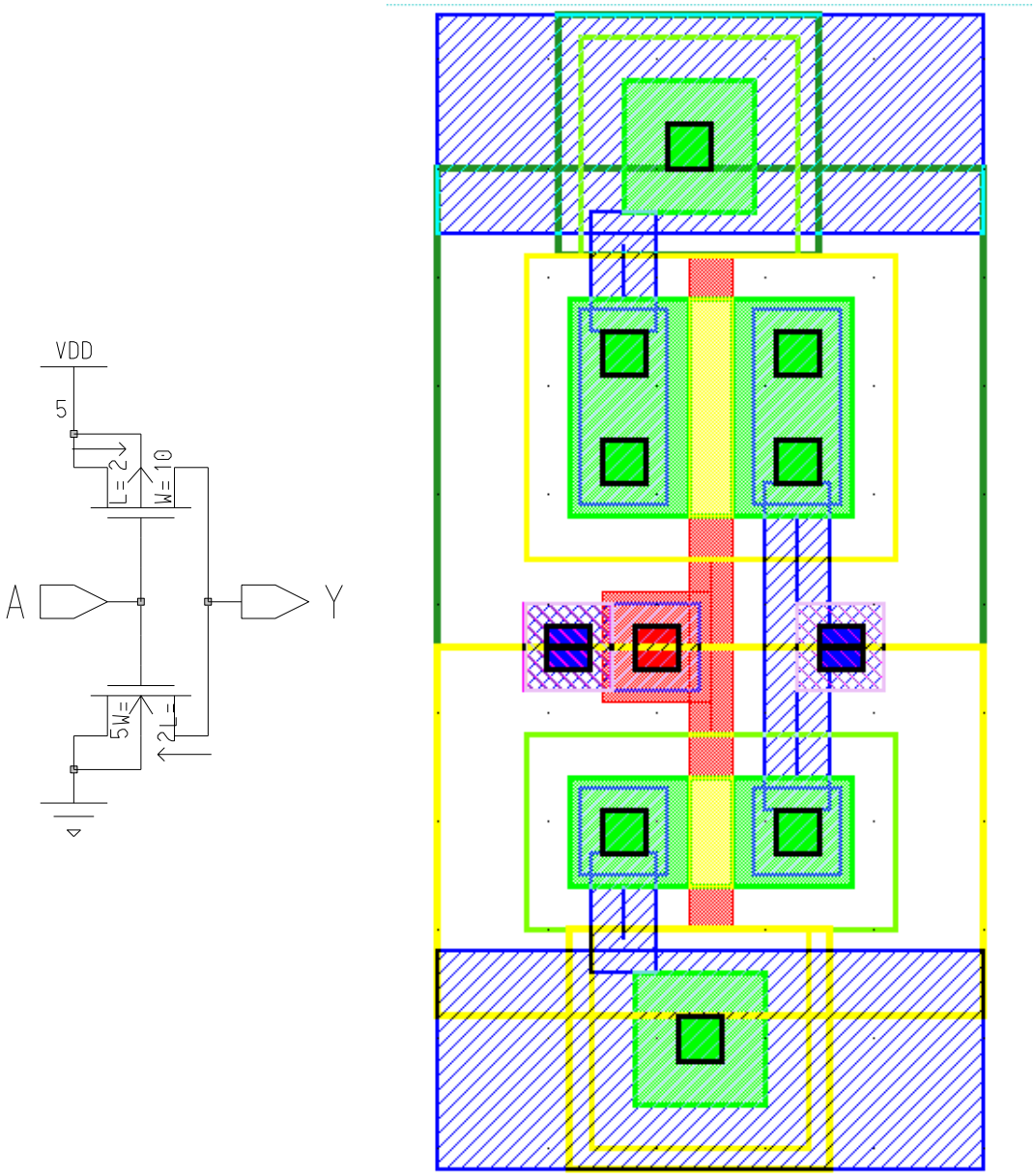


Figure 3.6: An inverter schematic created by the Design Architect and the layout created by the IC Station in the AMI0.5μ technology.

Comparing the complete layout of an inverter of Figure 3.6 with the simplified layout from Figure 3.5 we emphasize that the complete layout includes the detailed information about the masks required during the **fabrication process**. These masks can be inferred from the simplified layout but are not directly present there.

In particular, fabrication of the **diffusion** areas involves three masks, namely:

- The **active mask** which is a sum of the p diffusion and n diffusion areas.
- The **p+ mask** includes the p diffusion part of the active mask.
- The **n+ mask** includes the n diffusion part of the active mask.

As a result, the **substrate contacts** involve the following masks:

- The central $2\lambda \times 2\lambda$ **contact cut mask.**
- The **metal1 mask**.
- The $6\lambda \times 6\lambda$ **active mask.**
- The $10\lambda \times 10\lambda$ **p+/n+ mask.**
- The $12\lambda \times 11\lambda$ **p/n well mask.**

3.3 Transient properties of the CMOS inverter

In this section we will investigate basic transient properties of the CMOS inverter, that is, its dynamic behavior during switching the input signals from low-to-high or high-to-low voltages and associated power dissipation.

3.3.1 Propagation delay

Let us consider a CMOS inverter driven by a voltage pulse. Typical input/output waveforms are shown in Figure 3.7. Basic characterisation

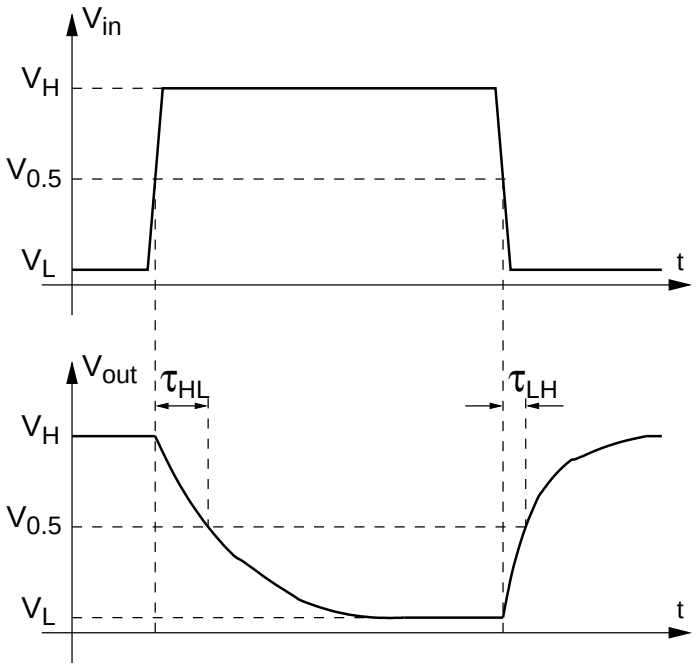


Figure 3.7: Input/output waveforms for a CMOS inverter.

of the dynamic behavior of an inverter is given by its two **propagation delay times**, τ_{HL} and τ_{LH} as illustrated in Figure 3.7. Note that these propagation times are specified with respect to the mid voltage $V_{0.5}$:

$$V_{0.5} = \frac{V_L + V_H}{2} \tag{3.6}$$

The propagation delay times, τ_{HL} (τ_{LH}) specifies the input-to-output time delay during the high-to-low (low-to-high) transition of the output voltage.

Often, it is convenient to refer to the **average propagation delay**, τ_p which specifies the average time required for the input signal to propagate through the inverter:

$$\tau_p = \frac{\tau_{HL} + \tau_{LH}}{2} \tag{3.7}$$

Similarly, we can define the **fall time**, τ_F, and the **rise time**, τ_R, as the time required for the output voltage to change between $V_{90\%}$ and $V_{10\%}$, where

$$V_{10\%} = V_L + 0.1(V_H - V_L) \ , \quad V_{90\%} = V_L + 0.9(V_H - V_L)$$

The physical reason for the propagation time delay is the existence of the **parasitic capacitances** associated with a MOS transistor. We can combine all such capacitances into an **equivalent load capacitance**, C_{ld}, as illustrated in Figure 3.8.

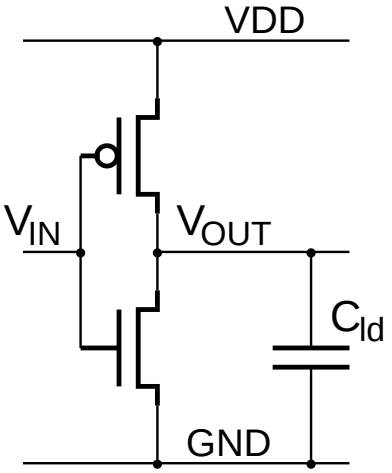


Figure 3.8: A CMOS inverter with an equivalent load capacitance.

In order to find the components of the load capacitance, C_{ld}, we should refer first to Figure 2.8 which illustrate a crossectional and a top view of a MOS transistor.

In addition we have to refer to a circuit layout of a CMOS inverter as in Figure 3.6. From these figures we can infer that load capacitance, C_{ld}, is formed from the following parasitic and lumped capacitances:

- $C_{gs,n}$, $C_{gd,n}$, $C_{gs,p}$, $C_{gd,p}$ — the parasitic capacitances between the gate and the source or drain of nMOS and pMOS transistors due to overlapping the gate and diffusion regions,
- $C_{db,n}$ and $C_{db,p}$ — the parasitic junction capacitances between the drains of the transistors and the relevant substrate,
- C_g — the gate capacitance over the gate area as discussed in sec. 2.3.2
- C_{int} — the equivalent lumped capacitance of the interconnection between the output node of the CMOS inverter and the input node of the circuitry being driven.

As illustrated by the transfer and current characteristics from Figure 3.3, during transition between low and high input voltages, there is a current flowing through the transistors forming the inverter. A part of this current charges and discharges the load capacitance which is responsible for propagation delay.
If we approximate the current flowing through the load capacitance by its average value, I_{avg}, then the propagation time can be estimated as:

$$\tau_p = \frac{C_{ld} \cdot \Delta V}{I_{avg}} \tag{3.8}$$

where ΔV indicates the voltage change across the load capacitance, that is, the change of the output voltage.

Direct calculation of the propagation time is rather complex, but from the eqn (3.8) we can conclude that one way of reducing the propagation time is to increase the average current flowing through the transistors to the load capacitance. We can do it by increasing the transistor transconductance parameter, g_c, that is, by increasing the W/L ratio as specified in eqn (2.5).
For an inverter INV01 from the AMI05 library we can find out that the value of the load capacitance of the inverter without the interconnecting lumped capacitance is in the order of

$$C_{ld} = 0.01\mathrm{pF} = 10\mathrm{fF}$$

Now, we can use the Mentor Graphics tools to build a test circuit consisting of two inverters as in Figure 3.9. The second inverter is a load for the first inverter. If we simulate this circuit we obtain the waveforms as in Figure 3.9. From the waveforms we can estimate that the propagation time is in the order of:

$$\tau_p = 100\mathrm{ps} = 0.1\mathrm{ns}$$

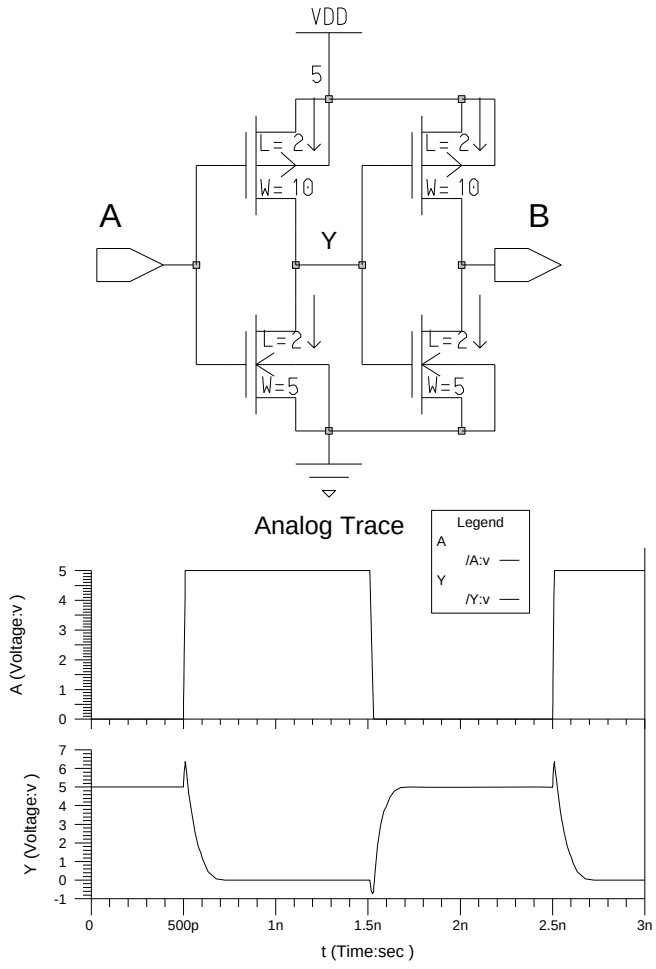


Figure 3.9: Simulation waveforms for a CMOS inverter driving another CMOS inverter.

3.3.2 Switching power dissipation

As it has been already discussed a CMOS inverted dissipate statically a negligible amount of power. However, during switching between low and high voltage level there is a non-negligible power dissipation due to charging up and down the load capacitance. This average power dissipated can be estimated as

$$P_{diss} = C_{ld} \cdot V_{DD}^2 \cdot f = E_{diss} \cdot f \tag{3.9}$$

where

C_{ld} is the load capacitance

V_{DD} is the supply voltage,

f is the switching frequency, and

$E_{diss} = C_{ld} \cdot V_{DD}^2$ is a **dissipated energy per signal transition**.

Note that the power dissipated is independent of the time delays and is proportional to the load capacitance, switching frequency and to the square of the supply voltage, V_{DD}. One of the common ways of reducing the power dissipation is through the reduction of the supply voltage.

For an inverter INV01 from the AMI05 library the dissipated energy is in the range of:

$$E_{diss} = 2\mathrm{pJ}$$

For the clock frequency, $f = 500$ MHz power dissipated by the inverter is

$$P_{diss} = 2 \cdot 10^{-12} \cdot 5 \cdot 10^{8} = 1\mathrm{mW}$$

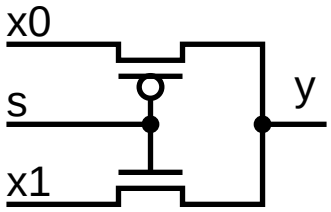
Comparatively, we can find out from the data book that a static power dissipation for the same inverter is only:

$$P_{stat} = 60\mathrm{pW}$$

The ratio of these two power is approximately equal to $17 \cdot 10^6$.

3.4 A “weak” multiplexer

From a simple switching model of a CMOS inverter as presented in Figure 3.3 we can observe that an input signal switches the output node, V_y, between the supply node, V_{DD}, and the ground node, GND. We can generalized this solution and switch the output node between two independent signals, creating a basic 2-to-1 multiplexer as in Figure 3.10.



$$y = \begin{cases} x_0 & \text{if} \quad s = 0 \\ x_1 & \text{if} \quad s = 1 \end{cases}$$

Figure 3.10: A 2-to-1 multiplexer based on a nMOS–pMOS pair.

A 2-to-1 multiplexer is a basic building block of the “**switch logic**”. The concept of the switch logic is that logic circuits are implemented as combination of switches, rather than logic gates. A 2-to-1 multiplexer has the structure as presented in Figure 3.10.

MOS transistors used in a way as in Figure 3.10 are referred to as pass transistors because they pass a signal from the source to drain.

The multiplexer from Figure 3.10 operates in such a way that when the select signal s is **low**, only the **pMOS** pass transistor is “on” and $y = x_0$, and, conversely, when the signal s is **high**, only the **nMOS** pass transistor is “on” and $y = x_1$.

The problem with the above circuit is that when an input signal, x_0 or x_1, is being passed to the output, the circuit **degrades** the signal levels, namely,

- the **low level** of the x_0 signal passing through the **pMOS** transistor,
- the **high level** of the x_1 signal passing through the **nMOS** transistor,

For that reason the above 2-to-1 multiplexer is called a “weak” multiplexer.

In order to get realistic simulation results a weak multiplexer drives a CMOS inverter as in Figure 3.11.

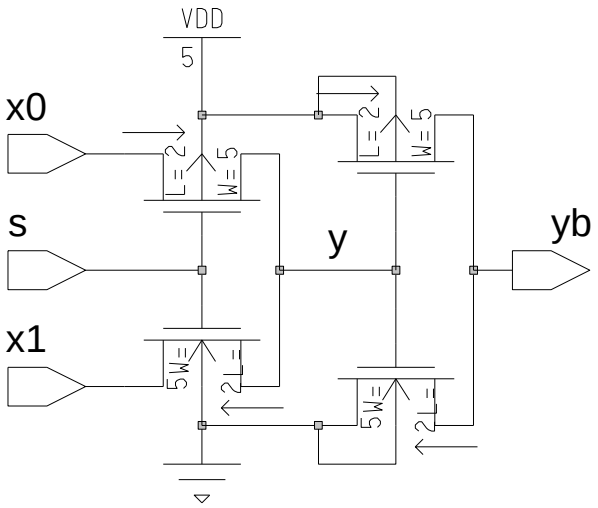


Figure 3.11: A weak multiplexer driving a CMOS inverter.

The inverter is needed to provide required power amplification, that is, to restore signal levels corrupted in the weak multiplexer.

Simulation waveforms are given in Figure 3.12.

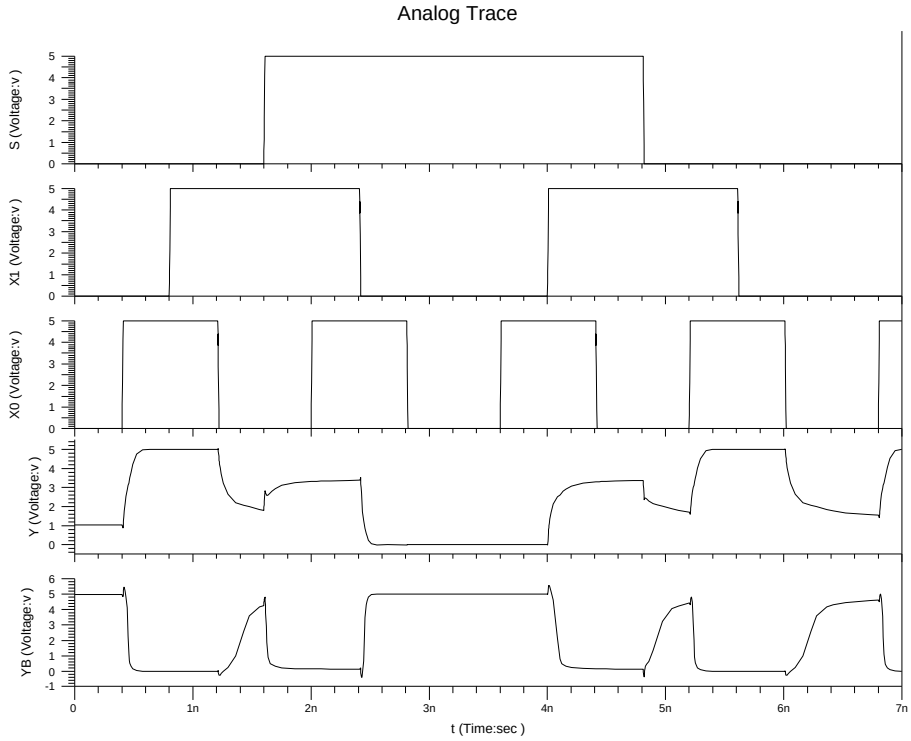


Figure 3.12: Simulation waveforms for a weak multiplexer driving a CMOS inverter.

The shortest distance between edges of two signals, say, between rising edges of x0 and x1 is set up to be 400ps.

Note that the output from the multiplexer, y, has not only corrupted levels, but also changes relatively slowly. In particular, note that when:

- s = 0V the input signal x0 is passed to the output through the nMOS which is switched on, pMOS being switched off. In particular, when x0 = 0V, y = +1V.
- s = 0V and x0 changes from 0V to +5V at 0.4ns then the rise time for the output $\tau_R \approx 100ps$. The equivalent fall time at the inverter output, yb, $\tau_F \approx 50ps$, which is a significant improvement.
- s = 0V and x0 changes from +5V back to 0V at 1.2ns then the multiplexer output goes slowly down so that after 400ps it reaches +1.8V. The inverter restores partly the signal level, but also with a significant rise time.
- s goes from 0V to +5V at 1.6ns the nMOS transistor is being switched off while the pMOS is switched on. Now the input signal x1 will be passed through to the output y. In particular, at 1.6ns the signal y starts to rise from the low corrupted level to the high corrupted level attaining 3.2V at 1.8ns.
- s = +5V and x1 changes from +5V to 0V at 2.4ns, this transition is passed to the input with a relatively short fall time, $\tau_F \approx 100ps$. As at the equivalent transition at 0.4ns the rise time at the inverter output is improved to $\tau_R \approx 50ps$.
- Finally, observe the events at 4ns, 4.8ns, 5.2ns, 6ns and 6.8ns and compare them with those described above.

In conclusion we say that the weak multiplexer is an interesting circuit due to its simplicity, however it is relatively slow and must be followed by an inverter or and equivalent power amplifying circuit.