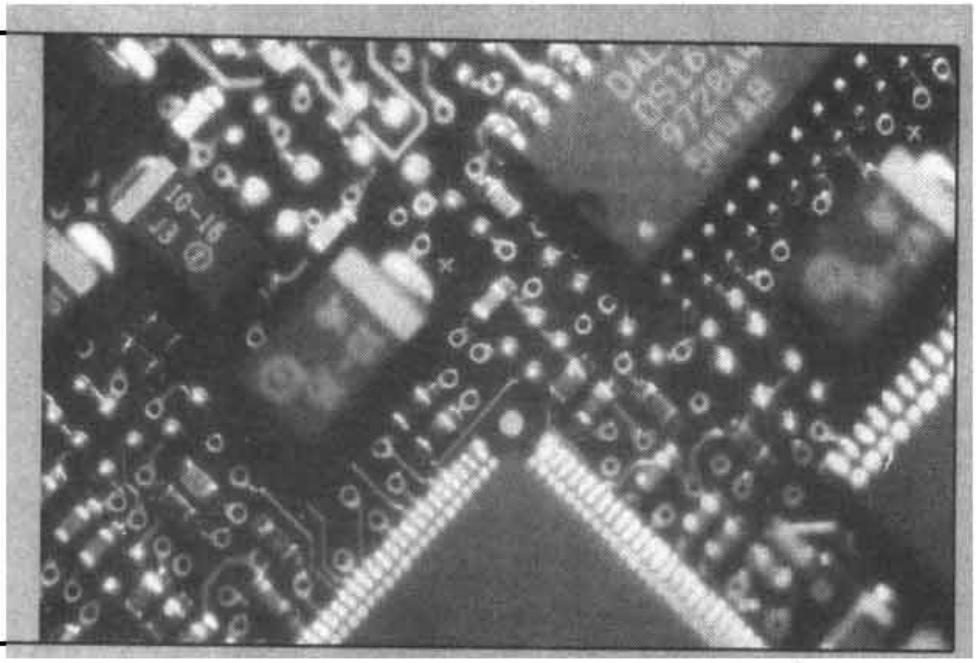


PART

III

DIGITAL ELECTRONICS

Part II of the text dealt with analog electronic circuits. Part III deals with digital electronics, another important category of electronics.

Chapter 16 examines field-effect transistor digital circuits. MOSFET digital circuits have revolutionized digital electronics, with the CMOS technology producing high-density, low-power digital circuits. Initially, we analyze basic NMOS and CMOS inverters, and then we develop NMOS and CMOS logic gates. Finally in this chapter, we analyze FET shift registers and flip-flops.

Bipolar digital circuits are considered in Chapter 17. We initially examine emitter-coupled logic, which is primarily used in specialized high-speed applications. Then, because transistor–transistor logic (TTL) circuits were the mainstay of logic design for many years, we analyze basic TTL and low-power Schottky TTL circuits, in order to obtain a good comparison between the FET and bipolar digital technologies.

CHAPTER

16

MOSFET Digital Circuits

16.0 PREVIEW

This chapter presents the basic concepts of MOSFET digital integrated circuits, which is the most widely used technology for the fabrication of digital systems. The small transistor size and low power dissipation of CMOS circuits allows for a high level of integration for logic and memory circuits. We initially examine NMOS logic circuits, which contain only n-channel transistors, and then complementary MOS, or CMOS, logic circuits, which contain both n-channel and p-channel transistors. JFET logic circuits are very specialized and are therefore not considered here.

The discussion of NMOS logic circuits will serve as an introduction to the analysis and design of digital circuits. This technology deals with only one type of transistor (n-channel) and therefore makes the analysis more straightforward than dealing with two types of transistors in the same circuit. This discussion will also serve as a baseline to point out the advantages of CMOS technology.

The CMOS inverter is the basis of CMOS logic gates. We will analyze the inverter dc voltage transfer characteristics and will determine the power dissipation in the CMOS inverter, demonstrating the principal advantage of CMOS circuits over NMOS circuits. The CMOS inverter leads to the basic CMOS NOR and NAND logic gates. We also look at more advanced clocked CMOS logic circuits, which eliminate almost half of the transistors in a conventional CMOS logic design while maintaining the lower power advantage of the CMOS technology.

In addition to the basic logic gates that are discussed, we consider additional logic circuits such as flip-flops, shift registers, and adders. Finally, a whole class of digital systems, called memories, is considered. Static memory cells, dynamic memory cells, and read-only memory cells are analyzed. Sense amplifiers and read/write circuitry are briefly discussed.

16.1 NMOS INVERTERS

The inverter is the basic circuit of most MOS logic circuits. The design techniques used in NMOS logic circuits are developed from the dc analysis results for the MOS inverter. Extending the concepts developed from the inverter to NOR and NAND gates is then direct. Alternative inverter load elements are

compared in terms of power consumption, packing density, and transfer characteristics. The transient analysis and switching characteristics of the inverters give an indication of the propagation delay times of NMOS logic circuits.

16.1.1 n-Channel MOSFET Revisited

We studied the structure, operation, and characteristics of MOS transistors in Chapter 5. In this section, we will quickly review the n-channel MOSFET characteristics, emphasizing specific properties important in digital circuit design.

A simplified n-channel MOSFET is shown in Figure 16.1(a). The body, or substrate, is a single-crystal silicon wafer, which is the starting material for circuit fabrication and provides physical support for the integrated circuit. The active transistor region is the surface of the semiconductor and comprises the heavily doped n^+ source and drain regions and p-type channel region. The channel length is L and the channel width is W. Normally, in any given fabrication process, the channel length is the same for all transistors, while the channel width is variable.

Figure 16.1(b) shows a more detailed view of the n-channel MOSFET. This figure demonstrates that the actual device geometry is more complicated than that indicated by the simplified cross section.

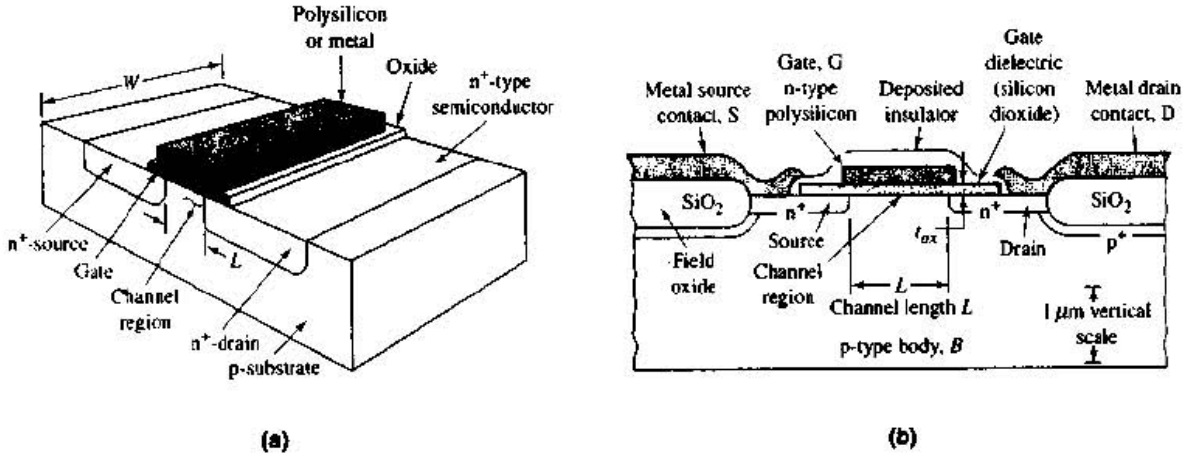


Figure 16.1 (a) n-channel MOSFET simplified view and (b) n-channel MOSFET detailed cross section

Figure 16.2(a) shows the simplified circuit symbols for the n-channel enhancement- and depletion-mode devices. When we explicitly consider the body or substrate connection, we will use the symbols shown in Figure 16.2(b).

In an integrated circuit, all n-channel transistors are fabricated in the same p-type substrate material. The substrate is connected to the most negative potential in the circuit, which for digital circuits, is normally at ground potential or zero volts. However, the source terminal of many of the transistors will not be at zero volts, which means that a reverse-biased pn junction will exist between the source and substrate.

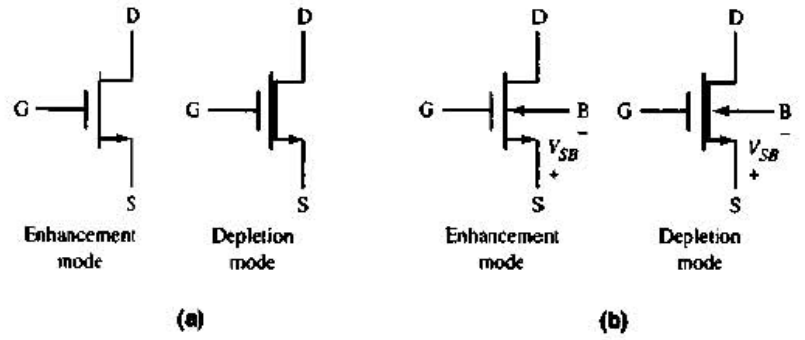


Figure 16.2 (a) Simplified circuit symbols for n-channel MOSFETs and (b) circuit symbols, showing substrate or body terminal

When the source and body terminals are connected together, the threshold voltage, to a first approximation, is independent of the applied voltages. However, when the source and body voltages are not equal, as when transistors are used for active loads, for instance, the threshold voltage is a function of difference between these voltages. We can write

$$\begin{aligned} V_{TN} &= V_{TNO} + \frac{\sqrt{2\varepsilon\varepsilon_s N_a}}{C_{ox}}\left[\sqrt{2\phi_{fp} + V_{SB}} - \sqrt{2\phi_{fp}}\right] \\ &= V_{TNO} + \gamma\left[\sqrt{2\phi_{fp} + V_{SB}} - \sqrt{2\phi_{fp}}\right] \end{aligned} \tag{16.1}$$

where V_{SB} is the source-to-body voltage, and V_{TNO} is the threshold voltage for zero source-to-body voltage or $V_{SB} = 0$. The parameter N_a is the p-type substrate doping concentration, ε_s is the semiconductor permittivity, C_{ox} is the oxide capacitance per unit area, ϕ_{fp} is a potential related to the substrate doping concentration, and γ is the body-effect coefficient.

Example 16.1 **Objective:** Determine the threshold voltage change due to a source-to-body voltage.

Consider a silicon n-channel MOSFET with the following parameters: $N_a = 1 \times 10^{16}\,\text{cm}^{-3}$, t_{ox} (oxide thickness) = 500 Å, and $\phi_{fp} = 0.347\,\text{V}$.

Solution: The oxide capacitance is

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} = \frac{(3.9)(8.85 \times 10^{-14})}{500 \times 10^{-8}} = 6.9 \times 10^{-8}\,\text{F/cm}^2$$

The change in threshold voltage is therefore

$$\begin{aligned} \Delta V_{TN} = V_{TN} - V_{TNO} &= \frac{\sqrt{2\varepsilon\varepsilon_s N_a}}{C_{ox}}\left[\sqrt{2\phi_{fp} + V_{SB}} - \sqrt{2\phi_{fp}}\right] \\ &= \frac{\sqrt{2(1.6 \times 10^{-19})(11.7)(8.85 \times 10^{-14})(1 \times 10^{16})}}{6.9 \times 10^{-8}}\left[\sqrt{0.694 + V_{SB}} - \sqrt{0.694}\right] \\ &= 0.834\left[\sqrt{0.694 + V_{SB}} - \sqrt{0.694}\right] \end{aligned}$$

For this case, the body-effect coefficient is $\gamma = 0.834\,\text{V}^{1/2}$. The threshold voltage change resulting from a source-to-body voltage V_{SB} is shown in Figure 16.3.

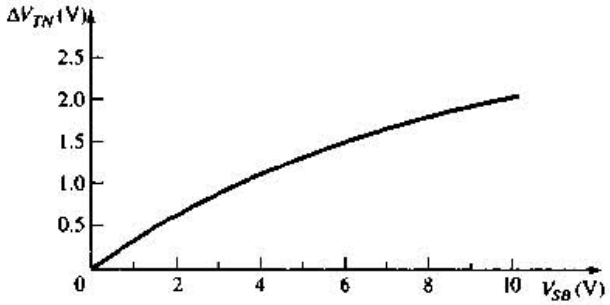


Figure 16.3 Change in threshold voltage versus source-to-body voltage for n-channel MOSFET in Example 16.1

Comment: The threshold voltage change with a change in V_{SB} will alter the current–voltage characteristics of the device and can alter the output voltage of an inverter.

The current–voltage characterisics of the n-channel MOSFET are functions of both the electrical and geometric properties of the device. When the transistor is biased in the nonsaturation region, for $v_{GS} \geq V_{TN}$ and $v_{DS} \leq (v_{GS} - V_{TN})$, we can write

$$i_D = K_n\left[2(v_{GS} - V_{TN})v_{DS} - v_{DS}^2\right] \tag{16.2(a)}$$

In the saturation region, for $v_{GS} \geq V_{TN}$ and $v_{DS} \geq (v_{GS} - V_{TN})$, we have

$$i_D = K_n(v_{GS} - V_{TN})^2 \tag{16.2(b)}$$

The transition point separates the nonsaturation and saturation regions and is the drain-to-source saturation voltage, which is given by

$$v_{DS} = v_{DS}(\text{sat}) = v_{GS} - V_{TN} \tag{16.3}$$

The term $(1 + \lambda v_{DS})$ is sometimes included in Equation (16.2(b)) to account for channel length modulation and the finite output resistance. In most cases, it has little effect on the operating characteristics of MOS digital circuits. In our analysis, the term λ is assumed to be zero unless otherwise stated.

The parameter K_n is the NMOS transistor conduction parameter and is given by

$$K_n = \left(\frac{1}{2}\mu_n C_{ox}\right)\left(\frac{W}{L}\right) = \frac{k_n'}{2}\frac{W}{L} \tag{16.4}$$

The electron mobility μ_n and oxide capacitance C_{ox} are assumed to be constant for all devices in a particular IC.

The current–voltage characteristics are directly related to the channel width-to-length ratio, or the size of the transistor. In general, in a given IC, the length L is fixed, but the designer can control the channel width W.

Since the MOS transistor is a majority carrier device, the switching speed of MOS digital circuits is limited by the time required to charge and discharge the capacitances between device electrodes and between interconnect lines and ground. Figure 16.4 shows the significant capacitances in a MOSFET. The capacitances C_{sb} and C_{db} are the source-to-body and drain-to-body n^+p junction capacitances. The total input gate capacitance, to a first approximation, is a constant equal to

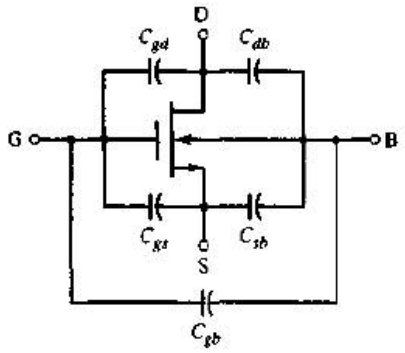


Figure 16.4 n-channel MOSFET and device capacitances

$$C_g = WLC_{ox} = WL\left(\frac{\varepsilon_{ox}}{t_{ox}}\right) \tag{16.5}$$

where C_{ox} is the oxide capacitance per unit area, and is a function of the oxide thickness. The parameter C_{ox} also appears in the expression for the conduction parameter.

Small Geometry Effects

The current–voltage relationships given by Equations (16.2(a)), (16.2(b)), and (16.3) are first-order approximations that apply to "long" channel devices. The tendency in device design is to make the devices as small as possible, which means the channel length is being reduced to values on the order of 0.25 μm or less. The corresponding channel widths are also being reduced. As the channel length is reduced, several effects alter the current–voltage characteristics. First, the threshold voltage becomes a function of the geometry of the device and is dependent on the channel length. This effect must be taken into account in the design of the transistor. Second, carrier velocity saturation reduces the saturation-mode current below the current value predicted by Equation (16.2(b)). The current is no longer a quadratic function of gate-to-source voltage, and tends to become a linear function of voltage. Channel length modulation means that the current tends to be larger than that predicted by the ideal equation. Third, the electron mobility is a function of the gate voltage so that the current tends to be smaller than the predicted value as the gate-to-source voltage increases. All of these effects complicate the analysis considerably.

We can, however, determine the basic operation and behavior of MOSFET logic circuits by using the first-order equations. We will use these first-order equations in our design of logic circuits. To determine the effect of small device size, a computer simulation may be performed in which the appropriate device models are incorporated in the simulation.

16.1.2 NMOS Inverter Transfer Characteristics

Since the inverter is the basis for most logic circuits, we will describe the NMOS inverter and will develop the dc transfer characteristics for three types of inverters with different load devices. This discussion will introduce voltage transfer functions, noise margins, and the transient characteristics of FET digital circuits.

NMOS Inverter with Resistor Load

Figure 16.5(a) shows a single NMOS transistor connected to a resistor to form an inverter. The transistor characteristics and load line are shown in Figure 16.5(b), along with the parametric curve separating the saturation and nonsaturation regions. We determine the voltage transfer characteristics of the inverter by examining the various regions in which the transistor can be biased.

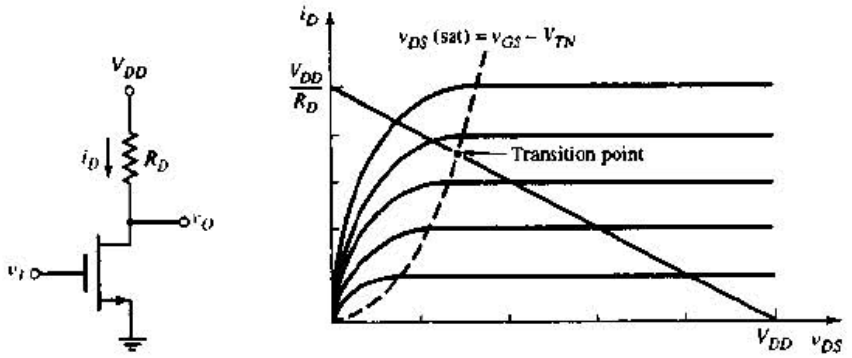


Figure 16.5 (a) NMOS inverter with resistor load and (b) transistor characteristics and load line

When the input voltage is less than or equal to the threshold voltage, or $v_I \leq V_{TN}$, the transistor is cut off, $i_D = 0$, and the output voltage is $v_O = V_{DD}$. The maximum output voltage is defined as the logic 1 level. As the input voltage becomes just greater than V_{TN}, the transistor turns on and is biased in the saturation region. The output voltage is then

$$v_O = V_{DD} - i_D R_D \tag{16.6}$$

where the drain current is given by

$$i_D = K_n(v_{GS} - V_{TN})^2 = K_n(v_I - V_{TN})^2 \tag{16.7}$$

Combining Equations (16.6) and (16.7) yields

$$v_O = V_{DD} - K_n R_D(v_I - V_{TN})^2 \tag{16.8}$$

which relates the output and input voltages as long as the transistor is biased in the saturation region.

As the input voltage increases, the Q-point of the transistor moves up the load line. At the transition point, we have

$$V_{Ot} = V_{It} - V_{TN} \tag{16.9}$$

where V_{Ot} and V_{It} are the drain-to-source and gate-to-source voltages, respectively, at the transition point. Substituting Equation (16.9) into (16.8), we determine the input voltage at the transition point from

$$K_n R_D(V_{It} - V_{TN})^2 + (V_{It} - V_{TN}) - V_{DD} = 0 \tag{16.10}$$

As the input voltage becomes greater than V_{It}, the Q-point continues to move up the load line, and the transistor becomes biased in the nonsaturation region. The drain current is then

$$i_D = K_n\left[2(v_{GS} - V_{TN})v_{DS} - v_{DS}^2\right] = K_n\left[2(v_I - v_{TN})v_O - v_O^2\right] \tag{16.11}$$

Combining Equations (16.6) and (16.11) yields

$$v_O = V_{DD} - K_n R_D\left[2(v_I - V_{TN})v_O - v_O^2\right] \tag{16.12}$$

which relates the input and output voltages as long as the transistor is biased in the nonsaturation region.

Figure 16.6 shows the voltage transfer characteristics of this inverter for three resistor values. Also shown is the line, given by Equation (16.9), which separates the saturation and nonsaturation bias regions of the transistor. The figure shows that the minimum output voltage, or the logic 0 level, for a high input decreases with increasing load resistance, and the sharpness of the transition region between a low input and a high input increases with increasing load resistance.

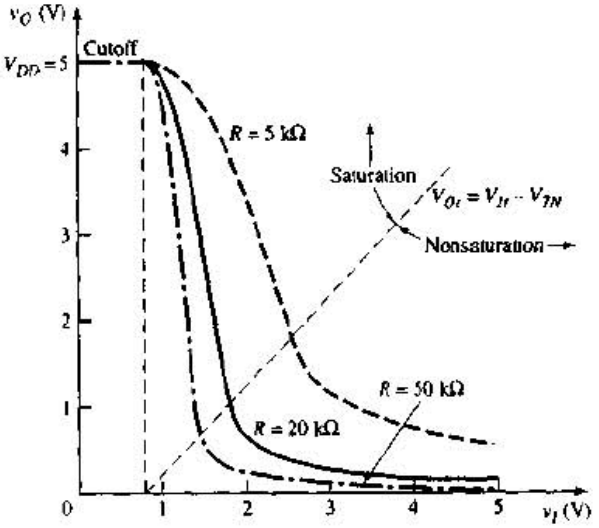


Figure 16.6 Voltage transfer characteristics, NMOS inverter with resistor load, for three resistor values

It should be noted that a large resistance is difficult to fabricate in an IC. A large resistor value in the inverter will limit current and power consumption as well as provide a small V_{OL} value. But it would also require a large chip area if fabricated in a standard MOS process. To avoid this problem MOS transistors can be used as load devices, replacing the resistor, as discussed in subsequent paragraphs.

Example 16.2 **Objective:** Determine the transition point and minimum output voltage of an NMOS inverter with resistor load.

Consider the circuit in Figure 16.5(a) with parameters $V_{DD} = 5\,\text{V}$ and $R_D = 20\,\text{k}\Omega$. The transistor parameters are $V_{TN} = 0.8\,\text{V}$ and $K_n = 0.2\,\text{mA/V}^2$.

Solution: The input voltage at the transition point is found from Equation (16.10). We have

$$(0.2)(20)(V_{it} - 0.8)^2 + (V_{it} - 0.8) - 5 = 0$$

which yields

$$V_{it} - 0.8 = 1 \quad \text{or} \quad V_{it} = 1.8\,\text{V}$$

The output voltage at the transition point is

$$V_{Ot} = V_{it} - V_{TN} = 1.8 - 0.8 = 1\,\text{V}$$

When v_I is high at $v_I = 5\,\text{V}$, the output voltage is found from Equation (16.12). We find

$$v_O = 5 - (0.2)(20)\left[2(5 - 0.8)v_O - v_O^2\right]$$

which yields the output low level as

$$v_O = V_{OL} = 0.147\,\text{V}$$

Only the negative root of the quadratic has physical significance because the positive root yields an output voltage greater than the supply voltage V_{DD}.

Comment: The level of V_{OL} is less than the threshold voltage V_{TN}; therefore, if the output of this inverter is used to drive a similar inverter, the driver transistor of the load inverter would be cut off and its output would be high, which is the desired condition.

Test Your Understanding

[Note: In the following exercise, assume $k_n' = 35\,\mu\text{A/V}^2$ for all NMOS transistors.]

16.1 Consider the NMOS inverter with resistor load in Figure 16.5(a) biased at $V_{DD} = 5\,\text{V}$. Assume transistor parameters of $W/L = 5$ and $V_{TN} = 0.8\,\text{V}$. (a) Find the value of R_D such that $v_O = 0.15\,\text{V}$ when $v_I = 5\,\text{V}$. (b) Using the results of part (a), determine the transition point for the driver transistor. (Ans. (a) $R_D = 44.8\,\text{k}\Omega$ (b) $V_{it} = 1.8\,\text{V}$, $V_{Ot} = 1.0\,\text{V}$)

NMOS Inverter with Enhancement Load

An n-channel enhancement-mode MOSFET with the gate connected to the drain can be used as a load device in an NMOS inverter. Figure 16.7(a) shows such a device. For $v_{GS} = v_{DS} \leq V_{TN}$, the drain current is zero. For

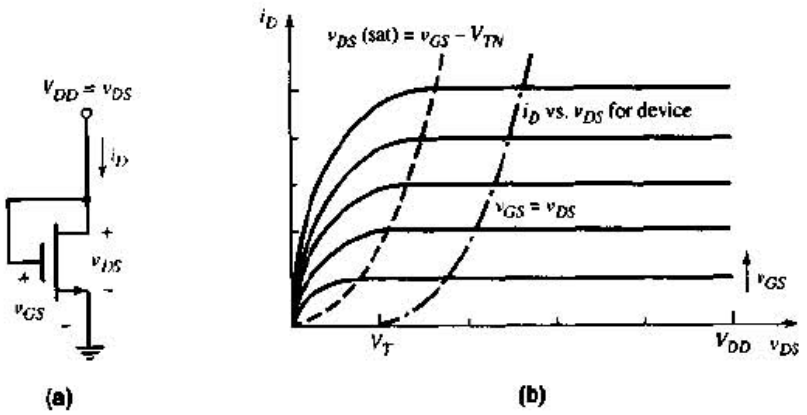


Figure 16.7 (a) n-channel MOSFET connected as saturated load device and (b) current–voltage characteristics of saturated load device

$v_{GS} = v_{DS} > V_{TN}$, a nonzero drain current is induced in the device. We can see that the following condition is satisfied:

$$v_{DS} > (v_{GS} - V_{TN}) = (v_{DS} - V_{TN}) = v_{DS}(\text{sat}) \tag{16.13}$$

A transistor with this connection always operates in the saturation region when not in cutoff.

The drain current is

$$i_D = K_n(v_{GS} - V_{TN})^2 = K_n(v_{DS} - V_{TN})^2 \tag{16.14}$$

We continue to neglect the effect of the output resistance and the λ parameter. The i_D versus v_{DS} characteristic is shown in Figure 16.7(b), which indicates that this device acts as a nonlinear resistor.

Figure 16.8(a) shows an NMOS inverter with the enhancement load device. The driver transistor parameters are denoted by V_{TND} and K_D, and the load transistor parameters are denoted by V_{TNL} and K_L. The substrate connections are not shown. In the following analysis, we neglect the body effect and we assume all threshold voltages are constant. These assumptions do not seriously affect the basic analysis, nor the inverter characteristics.

The driver transistor characteristics and the load curve are shown in Figure 16.8(b). When the inverter input voltage is less than the driver threshold voltage, the driver is cut off and the drain currents are zero. From Equation (16.14), we have

$$i_{DL} = 0 = K_L(v_{DSL} - V_{TNL})^2 \tag{16.15}$$

From Figure 16.8(a), we see that $v_{DSL} = V_{DD} - v_O$, which means that

$$v_{DSL} - V_{TNL} = V_{DD} - v_O - V_{TNL} = 0 \tag{16.16(a)}$$

The maximum output voltage is then

$$v_{O,\max} \equiv V_{OH} = V_{DD} - V_{TNL} \tag{16.16(b)}$$

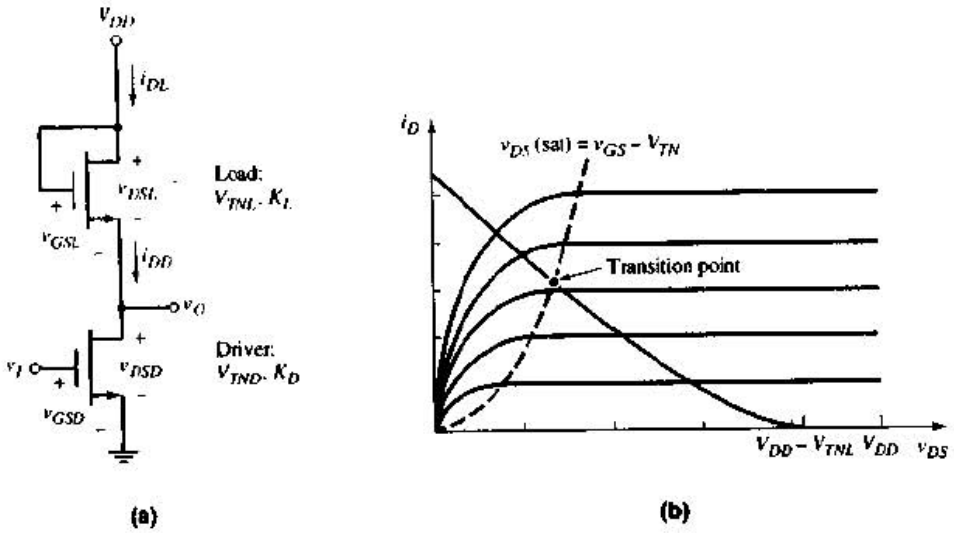


Figure 16.8 (a) NMOS inverter with saturated load and (b) driver transistor characteristics and load curve

For the enhancement-load NMOS inverter, the maximum output voltage, which is the logic 1 level, does not reach the full V_{DD} value. This cutoff point is shown in the load curve in Figure 16.8(b).

As the input voltage becomes just greater than the driver threshold voltage V_{TND}, the driver transistor turns on and is biased in the saturation region. In steady-state, the two drain currents are equal since the output will be connected to the gates of other MOS transistors. We have $i_{DD} = i_{DL}$, which can be written as

$$K_D(v_{GSD} - V_{TND})^2 = K_L(v_{GSL} - V_{TNL})^2 \tag{16.17}$$

Equation (16.17) is expressed in terms of the individual transistor parameters. In terms of the input and output voltages, the expression becomes

$$K_D(v_I - V_{TND})^2 = K_L(V_{DD} - v_O - V_{TNL})^2 \tag{16.18}$$

Solving for the output voltage yields

$$v_O = V_{DD} - V_{TNL} - \sqrt{\frac{K_D}{K_L}}(v_I - V_{TND}) \tag{16.19}$$

As the input voltage increases, the driver Q-point moves up the load curve and the output voltage decreases linearly with v_I.

At the driver transition point, we have

$$v_{DSD}(\text{sat}) = v_{GSD} - V_{TND}$$

or

$$V_{Oi} = V_{Ii} - V_{TND} \tag{16.20}$$

Substituting Equation (16.20) into (16.19), we find the input voltage at the transition point, which is

$$V_{Ii} = \frac{V_{DD} - V_{TNL} + V_{TND}\left(1 + \sqrt{\frac{K_D}{K_L}}\right)}{1 + \sqrt{\frac{K_D}{K_L}}} \tag{16.21}$$

As the input voltage becomes greater than V_{Ii}, the driver transistor Q-point continues to move up the load curve and the driver becomes biased in the nonsaturation region. Since the driver and load drain currents are still equal, or $i_{DD} = i_{DL}$, we now have

$$K_D\left[2(v_{GSD} - V_{TND})v_{DSD} - v_{DSD}^2\right] = K_L(v_{DSL} - V_{TNL})^2 \tag{16.22}$$

Writing Equation (16.22) in terms of the input and output voltages produces

$$K_D\left[2(v_I - V_{TND})v_O - v_O^2\right] = K_L(V_{DD} - v_O - V_{TNL})^2 \tag{16.23}$$

Obviously, the relationship between v_I and v_O in this region is not linear.

Figure 16.9 shows the voltage transfer characteristics of this inverter for three K_D-to-K_L ratios. The ratio K_D/K_L is the aspect ratio and is related to the width-to-length parameters of the driver and load transistors.

The line, given by Equation (16.20), separating the driver saturation and nonsaturation regions is also shown in the figure. We see that the minimum

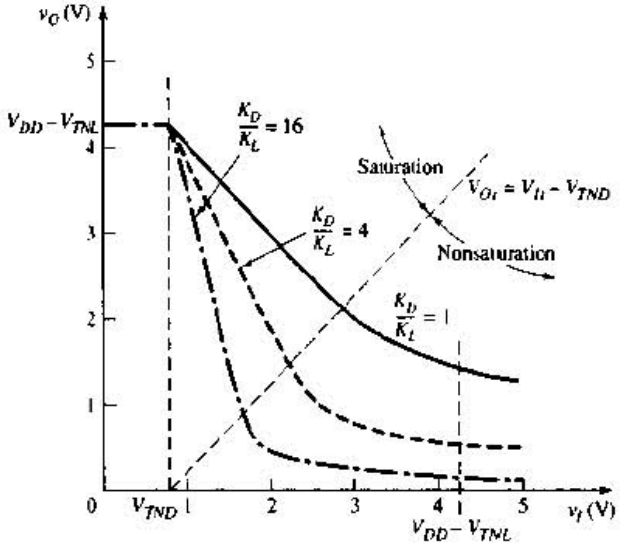


Figure 16.9 Voltage transfer characteristics, NMOS inverter with saturated load, for three aspect ratios

output voltage, or the logic 0 level, for a high input decreases with an increasing K_D/K_L ratio. As the width-to-length ratio of the load transistor decreases, the effective resistance increases, which means that the general behavior of the transfer characteristics is the same as for the resistor load. However, the high output voltage is

$$V_{OH} = V_{DD} - V_{TNL}$$

When the driver is biased in the saturation region, we find the slope of the transfer curve, which is the **inverter gain**, by taking the derivative of Equation (16.19) with respect to v_I. We see that

$$dv_O/dv_I = -\sqrt{K_D/K_L}$$

When the aspect ratio is greater than unity, the inverter gain magnitude is greater than unity. A logic circuit family with an inverter transfer curve that exhibits a gain greater than unity for some region is called a **restoring logic family**. Restoring logic is so named because logic signals that are degraded for some reason in one circuit can be restored by the gain of subsequent logic circuits.

Design Example 16.3 **Objective:** Design the aspect ratio K_D/K_L to produce a specified low output voltage, and determine the power dissipation in the inverter with enhancement load for a minimum W/L ratio for the load transistor. (Neglect the body effect.)

Consider the inverter shown in Figure 16.8(a) biased at $V_{DD} = 5\,\text{V}$. The transistor parameters are: $V_{TND} = V_{TNL} = 0.8\,\text{V}$ and $k_n' = 35\,\mu\text{A/V}^2$. Determine K_D/K_L such that $v_O = 0.10\,\text{V}$ when v_I = Logic 1 = 4.2 V, and determine $(W/L)_D$ and the power dissipation in the inverter for $(W/L)_L = 0.5$ and $v_I = 4.2\,\text{V}$.

Solution: For $v_I = 4.2\,\text{V}$, the driver transistor is biased in the nonsaturation region. Using Equation (16.23), we have

$$\frac{K_D}{K_L}\left[2(4.2 - 0.8)(0.1) - (0.1)^2\right] = (5 - 0.1 - 0.8)^2$$

which yields

$$\frac{K_D}{K_L} = 25.1$$

Since

$$\frac{K_D}{K_L} = \frac{(W/L)_D}{(W/L)_L}$$

then

$$(W/L)_D = 12.6$$

when

$$(W/L)_L = 0.5$$

The power dissipated in the inverter is $P = i_D V_{DD}$, and the drain current can be found from the load transistor, as follows:

$$i_D = K_L(V_{DD} - v_O - V_{TNL})^2 = \frac{k_n'}{2}\left(\frac{W}{L}\right)_L (V_{DD} - v_O - V_{TNL})^2$$

Therefore,

$$i_D = \left(\frac{35}{2}\right)(0.5)(5 - 0.1 - 0.8)^2 = 147\,\mu\text{A}$$

The power dissipation is

$$P = i_D V_{DD} = (147)(5) = 735\,\mu\text{W}$$

Comment: In the NMOS inverter with enhancement load, producing a relatively low output voltage V_{OL} requires a large difference in the sizes of the driver and load transistors. The load transistor size cannot be substantially reduced, so the power consumption also cannot be substantially reduced from the 735 μW value. If an IC contained a modest 100,000 inverters and all inverters were conducting, the total required current to the IC would be 14.7 A and the total power dissipated would be 73.5 W! We thus see the need to drastically reduce the power dissipation in each inverter.

Test Your Understanding

[Assume $k_n' = 35\,\mu\text{A/V}^2$.]

16.2 The enhancement-load NMOS inverter shown in Figure 16.8(a) is biased at $V_{DD} = 5\,\text{V}$. The threshold voltages are $V_{TND} = V_{TNL} = 1\,\text{V}$, and the width-to-length ratios are $(W/L)_D = 16$ and $(W/L)_L = 2$. (a) *Find* v_O *when: (i)* $v_I = 0$, *and* (ii) $v_I = 4\,\text{V}$. (b) Calculate the power dissipated in the inverter when $v_I = 4\,\text{V}$. (Ans. (a) $v_O = 4\,\text{V}$, $v_O = 0.30\,\text{V}$ (b) $P = 2.4\,\text{mW}$)

D16.3 Consider the NMOS inverter with enhancement load, as shown in Figure 16.8(a), biased at $V_{DD} = 5\,\text{V}$. The transistor threshold voltages are $V_{TND} = V_{TNL} = 0.8\,\text{V}$. Design the width-to-length ratios such that the output voltage is 0.2 V and the inverter power dissipation is $750\,\mu\text{W}$ when $v_I = 4.2\,\text{V}$. (Ans. $(W/L)_L = 0.536$, $(W/L)_D = 6.49$)

NMOS Inverter with Depletion Load

Depletion-mode MOSFETs can also be used as load elements in NMOS inverters. Figure 16.10(a) shows the NMOS inverter with depletion load. The gate and source of the depletion-mode transistor are connected together. The driver transistor is still an enhancement-mode device. As before, the driver transistor parameters are $V_{TND}(V_{TND} > 0)$ and K_D, and the load transistor parameters are $V_{TNL}(V_{TNL} < 0)$ and K_L. Again, the substrate connections are not shown. The fabrication process for this inverter is slightly more complicated than for

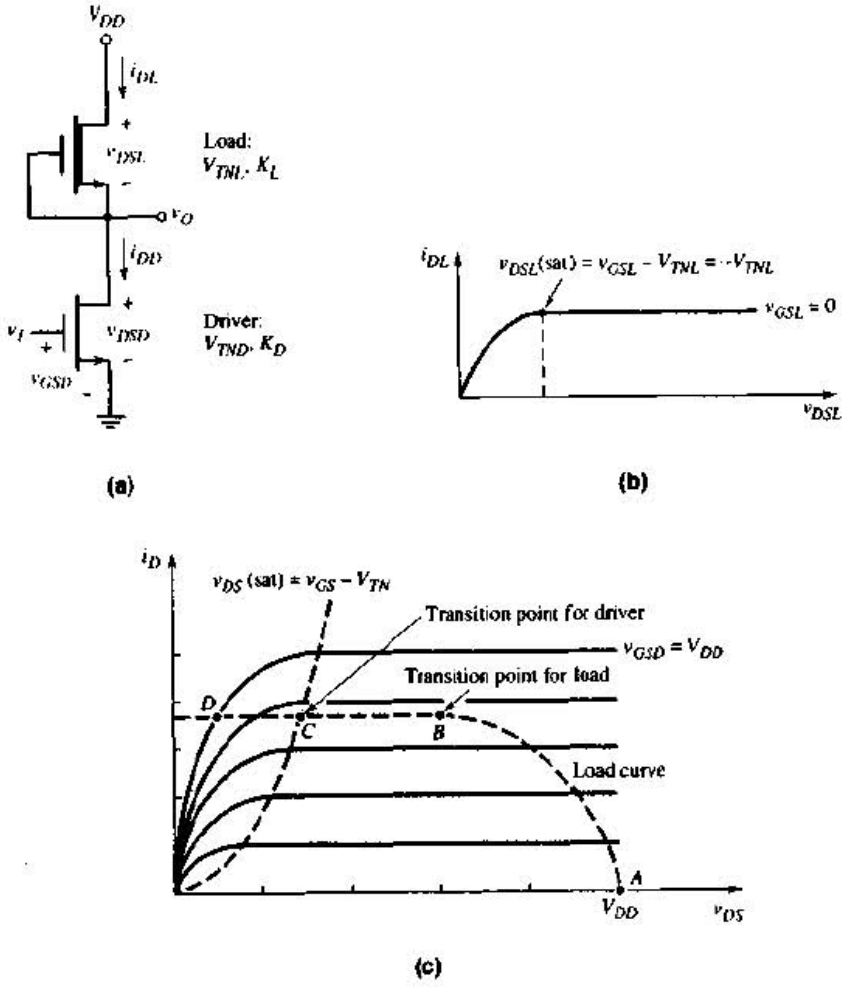


Figure 16.10 (a) NMOS inverter with depletion load, (b) current–voltage characteristic of depletion load, and (c) driver transistor characteristics and load curve

the enhancement-load inverter, since the threshold voltages of the two devices are not equal. However, as we will see, the advantages of this inverter make the extra processing steps worthwhile. This inverter has been the basis of many microprocessor and static memory designs.

The current-voltage characteristic curve for the depletion load, neglecting the body effect, is shown in Figure 16.10(b). Since the gate is connected to the source, $v_{GSL} = 0$, and the Q-point of the load is on this particular curve.

The driver transistor characteristics and the ideal load curve are shown in Figure 16.10(c). When the inverter input is less than the driver threshold voltage, the driver is cut off and the drain currents are zero. From Figure 16.10(b), we see that for $i_D = 0$, the drain-to-source voltage of the load transistor must be zero; therefore, $v_O = V_{DD}$ for $v_I \leq V_{TND}$. An advantage of the depletion-load inverter over the enhancement-load inverter is that the high output voltage, or the logic 1 level, is at the full V_{DD} value.

As the input voltage becomes just greater than the driver threshold voltage V_{TND}, the driver turns on and is biased in the saturation region; however, the load is biased in the nonsaturation region. The Q-point lies between points A and B on the load curve shown in Figure 16.10(c). We again set the two drain currents equal, or $i_{DD} = i_{DL}$, which means that

$$K_D[v_{GSD} - V_{TND}]^2 = K_L[2(v_{GSL} - V_{TNL})v_{DSL} - v_{DSL}^2] \tag{16.24}$$

Writing Equation (16.24) in terms of the input and output voltages yields

$$K_D[v_I - V_{TND}]^2 = K_L[2(-V_{TNL})(V_{DD} - v_O) - (V_{DD} - v_O)^2] \tag{16.25}$$

This equation relates the input and output voltages as long as the driver is biased in the saturation region and the load is biased in the nonsaturation region.

There are two transition points for the NMOS inverter with a depletion load: one for the load and one for the driver. These are points B and C, respectively, in Figure 16.10(c). The transition point for the load is given by

$$v_{DSL} = V_{DD} - V_{Ot} = v_{GSL} - V_{TNL} = -V_{TNL} \tag{16.26(a)}$$

or

$$V_{Ot} = V_{DD} + V_{TNL} \tag{16.26(b)}$$

Since V_{TNL} is negative, the output voltage at the transition point is less than V_{DD}. The transition point for the driver is given by

$$v_{DSD} = v_{GSD} - V_{TND}$$

or

$$V_{Ot} = V_{It} - V_{TND} \tag{16.27}$$

When the Q-point lies between points B and C on the load curve, both devices are biased in the saturation region, and

$$K_D(v_{GSD} - V_{TND})^2 = K_L(v_{GSL} - V_{TNL})^2 \tag{16.28(a)}$$

or

$$\sqrt{\frac{K_D}{K_L}}(v_I - V_{TND}) = -V_{TNL} \tag{16.28(b)}$$

Equation (16.28(b)) demonstrates that the input voltage is a constant as the Q-point passes through this region. This effect is also shown in Figure 16.10(c); the load curve between points B and C lies on a constant v_{GSD} curve. (This characteristic will change when the body effect is taken into account.)

For an input voltage greater than the value given by Equation (16.28(b)), the driver is biased in the nonsaturation region while the load is biased in the saturation region. The Q-point is now between points C and D on the load curve in Figure 16.10(c). Equating the two drain currents, we obtain

$$K_D\left[2(v_{GSD} - V_{TND})v_{DSD} - v_{DSD}^2\right] = K_L(v_{GSL} - V_{TNL})^2 \qquad \textbf{(16.29(a))}$$

which becomes

$$\frac{K_D}{K_L}\left[2(v_I - V_{TND})v_O - v_O^2\right] = (-V_{TNL})^2 \qquad \textbf{(16.29(b))}$$

This equation implies that the relationship between the input and output voltages are not linear in this region.

Figure 16.11 shows the voltage transfer characteristics of this inverter for three values of K_D/K_L. Also shown are the locus of transition points for the load and driver transistors as given by Equations (16.26(b)) and (16.27), respectively.

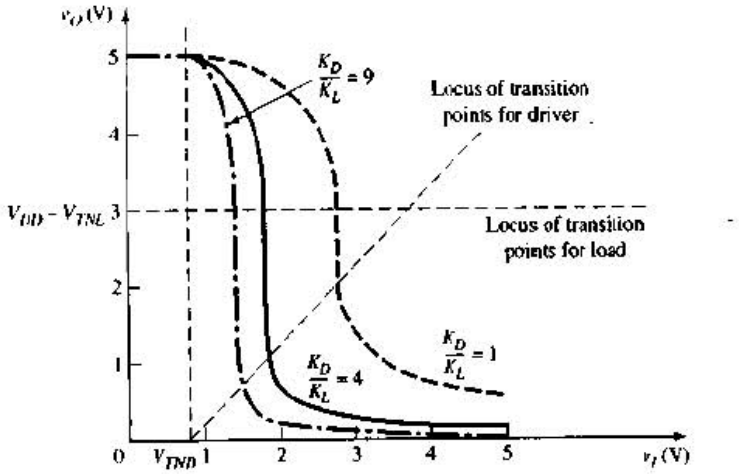
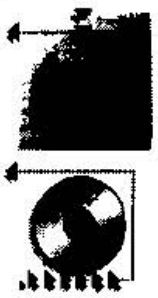


Figure 16.11 Voltage transfer characteristics, NMOS inverter with depletion load, for three aspect ratios

Design Example 16.4 Objective: Design the aspect ratio K_D/K_L to produce a specified low output voltage, and determine the power dissipation in the inverter with depletion load for a minimum W/L ratio for the load transistor.

Consider the inverter in Figure 16.10(a) biased at $V_{DD} = 5\,\text{V}$. The transistor parameters are: $V_{TND} = 0.8\,\text{V}$, $V_{TNL} = -2\,\text{V}$, and $k_n' = 35\,\mu\text{A/V}^2$. Determine K_D/K_L such that $v_O = 0.10\,\text{V}$ when $v_I = 5\,\text{V}$. Determine $(W/L)_D$ and the power dissipation in the inverter for $(W/L)_L = 0.5$.



Solution: For $v_I = 5\,\text{V}$, we assume the driver transistor is biased in the nonsaturation region and the load is in the saturation region. Using Equation (16.29(b)), we have

$$\frac{K_D}{K_L}[2(5-0.8)(0.1)-(0.1)^2] = [-(-2)]^2$$

which yields

$$\frac{K_D}{K_L} = 4.82$$

Since

$$\frac{K_D}{K_L} = \frac{(W/L)_D}{(W/L)_L}$$

then

$$(W/L)_D = 2.41$$

when

$$(W/L)_L = 0.5$$

The power dissipated in the inverter is $P = i_D V_{DD}$, and the drain current can be found from the load transistor, as follows:

$$i_D = K_L(-V_{TNL})^2 = \frac{k_n'}{2}\left(\frac{W}{L}\right)_L(-V_{TNL})^2 = \left(\frac{35}{2}\right)(0.5)[-(-2)]^2 = 35\,\mu\text{A}$$

The power dissipation is therefore

$$P = i_D V_{DD} = (35)(5) = 175\,\mu\text{W}$$

Comment: A relatively low output voltage V_{OL} can be produced in the NMOS inverter with depletion load, even when the load and driver transistors are not vastly different in size. The power dissipation in this inverter is also substantially less than in the enhancement-load inverter since the aspect ratio is smaller.

Design Consideration: The static analysis of the three types of NMOS inverters clearly demonstrates the advantage of the depletion load inverter. The size of the driver transistor is smaller for a given load device size to produce a given low output state. This allows a greater number of inverters to be fabricated in a given chip area. In addition, since the power dissipation is less, more inverters can be fabricated on a chip for a given total power dissipation.

Test Your Understanding

[Assume $k_n' = 35\,\mu\text{A/V}^2$.]

***16.4** The depletion load NMOS inverter shown in Figure 16.10(a) is biased at $V_{DD} = 5\,\text{V}$. the transistor parameters are: $V_{TND} = 0.7\,\text{V}$, $V_{TNL} = -1.5\,\text{V}$, $(W/L)_D = 6$, and $(W/L)_L = 2$. (a) Determine v_O for $v_I = 5\,\text{V}$. (b) Find the transition points for the driver and the load. (c) Calculate the power dissipation in the inverter when $v_I = 5\,\text{V}$. (Ans. (a) $v_O = 0.0881\,\text{V}$ (b) Load: $v_{It} = 1.57\,\text{V}$, $v_{Ot} = 3.5\,\text{V}$. Driver: $v_{It} = 1.57\,\text{V}$, $v_{Ot} = 0.87\,\text{V}$ (c) $P = 394\,\mu\text{W}$)

D16.5 Consider the depletion load inverter in Figure 16.10(a) biased at $V_{DD} = 5\,\text{V}$. The threshold voltages are $V_{TND} = 0.8\,\text{V}$ and $V_{TNL} = -2\,\text{V}$. Design the inverter such that the maximum power dissipation is $350\,\mu\text{W}$ and the output voltage is $0.05\,\text{V}$ when $v_I = 5\,\text{V}$. (Ans. $(W/L)_L = 1$, $(W/L)_D = 9.58$)

16.1.3 Noise Margin

The word "noise" means transient, unwanted variations in voltages or currents. In digital circuits, if the magnitude of the noise at a logic node is too large, logic errors can be introduced into the system. However, if the noise amplitude is less than a specified value, called the **noise margin**, the noise signal will be attenuated as it passes through a logic gate or circuit, while the logic signals will be transmitted without error.

Noise signals are usually generated outside the digital circuit and transferred to logic nodes or interconnect lines through parasitic capacitances or inductances. The coupling process is usually time dependent, leading to dynamic conditions in the circuit. In digital systems, however, the noise margins are usually defined in terms of static voltages.

Noise Margin Definition

For static noise margins, the type of noise usually considered is called series-voltage noise. Figure 16.12 shows two inverters in series in which the output of the second is connected back to the input of the first. Also included are series-voltage noise sources δV_L and δV_H. This type of noise can be developed by inductive coupling. The input voltage levels are indicated by H (high) and L (low). The noise amplitudes δV_L and δV_H can be different, and the polarities may be such as to increase the low output and reduce the high output. The noise margins are defined as the maximum values of δV_L and δV_H at which the inverters will remain in the correct state.

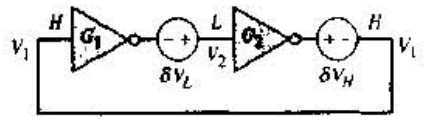


Figure 16.12 Two-inverter flip-flop, including series-voltage noise sources

The actual definitions of the noise margins NM_L and NM_H are not unique. In addition other types of noise, other than series-voltage source noise, may be present in the system. Dynamic noise sources also complicate the issue. However, in this text, in order to provide some measure of noise margin in a logic circuit, we will use the unity-gain approach to determine the logic threshold levels V_{IL} and V_{IH} and the corresponding noise margins.

Figure 16.13 shows a general voltage transfer function for an inverter. The expected logic 1 and logic 0 output voltages of the inverter are V_{OH} and V_{OL}, respectively. The parameters V_{IH} and V_{IL}, which determine the noise margins, are defined as the points at which

$$\frac{dv_O}{dv_I} = -1 \tag{16.30}$$

For $v_I \leq V_{IL}$, the inverter gain magnitude is less than unity, and the output changes slowly with a change in the input voltage. Similarly, for $v_I \geq V_{IH}$, the output again changes slowly with input voltage since the gain magnitude is less than unity. However, when the input voltage is in the range $V_{IL} < v_I < V_{IH}$, the gain magnitude is greater than one, and the output signal changes rapidly.

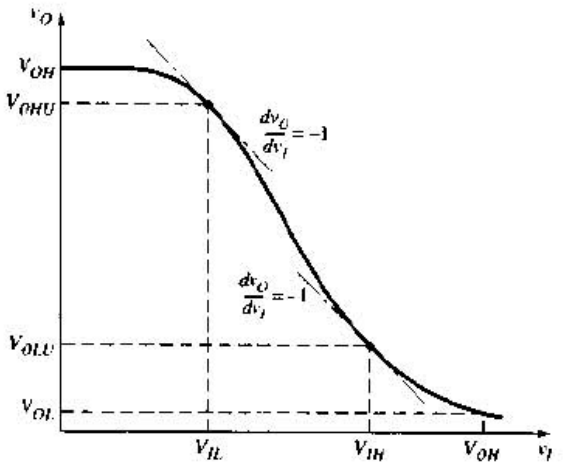


Figure 16.13 Generalized inverter voltage curve and defined voltage limits V_{IL} and V_{IH}

This region is called the **undefined range.** If the input voltage is inadvertently pushed into this range by a noise signal, the output may change logic states, and a logic error could be introduced into the system. The corresponding output voltages at the unity-gain points are denoted V_{OHU} and V_{OLU}, where the last subscript U signifies the unity-gain values.

The noise margins are defined as

$$\text{NM}_L = V_{IL} - V_{OLU} \tag{16.31(a)}$$

and

$$\text{NM}_H = V_{OHU} - V_{IH} \tag{16.31(b)}$$

We will see how these noise margin definitions correspond to the flip-flop conditions just discussed.

Since an inverter with a resistor load is rarely used in practice, we will determine the noise margins for NMOS inverters with enhancement and depletion loads.

Enhancement-Load Inverter

The general voltage transfer characteristic for the NMOS inverter with enhancement load is shown in Figure 16.14. Normally, $K_D > K_L$ and the gain magnitude for $v_I > V_{TND}$ is greater than one. Since the slope for this ideal curve is discontinuous at $v_I = V_{TND}$, we define the parameter $V_{IL} \equiv V_{TND}$. This is the threshold voltage of the driver. In this case, the output voltage V_{OHU} corresponds to V_{OH}.

At voltage V_{IH}, the driver is biased in the nonsaturation region. The relationship between the input and output voltages is given by Equation (16.23). Taking the derivative with respect to v_I, we obtain

$$\begin{aligned} &2K_D v_O + 2K_D v_I \frac{dv_O}{dv_I} - 2K_D V_{TND} \frac{dv_O}{dv_I} - 2K_D v_O \frac{dv_O}{dv_I} \\ &= -2K_L(V_{DD} - v_O - V_{TNL}) \frac{dv_O}{dv_I} \end{aligned} \tag{16.32}$$

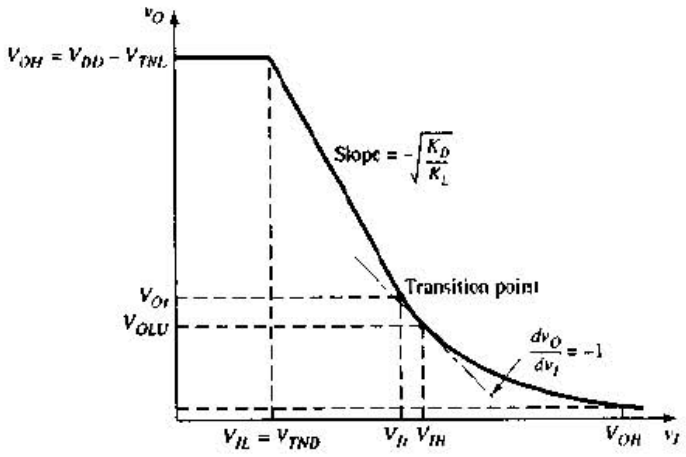


Figure 16.14 Voltage transfer characteristic, NMOS inverter with saturated load, and defined voltage limits V_{IL} and V_{IH}

Setting the derivative equal to -1 yields

$$K_D v_O - K_D v_I + K_D V_{TND} + K_D v_O = +K_L(V_{DD} - v_O - V_{TNL}) \tag{16.33}$$

Solving for v_O, we obtain

$$v_O = V_{OLU} = \frac{(V_{DD} - V_{TNL}) + \dfrac{K_D}{K_L}(v_I - V_{TND})}{1 + 2\dfrac{K_D}{K_L}} \tag{16.34}$$

Finally, combining Equations (16.34) and (16.23) and solving for v_I, we have

$$v_I = V_{IH} = V_{TND} + \left(\frac{V_{DD} - V_{TNL}}{\dfrac{K_D}{K_L}}\right)\left(\frac{1 + 2\dfrac{K_D}{K_L}}{\sqrt{1 + 3\dfrac{K_D}{K_L}}} - 1\right) \tag{16.35}$$

Example 16.5 **Objective:** Determine the noise margins of an inverter with enhancement load.

Consider the inverter shown in Figure 16.8(a) with the parameters given in Example 16.3. We have that $V_{OH} = 4.2\,\text{V}$, $V_{OL} = 0.10\,\text{V}$, and $K_D/K_L = 25.1$.

Solution: We know that $V_{IL} = V_{TND} = 0.8\,\text{V}$ and $V_{OHU} = V_{OH} = 4.2\,\text{V}$. The value of V_{IH} is, from Equation (16.35),

$$V_{IH} = 0.8 + \left(\frac{5 - 0.8}{25.1}\right)\left[\frac{1 + 2(25.1)}{\sqrt{1 + 3(25.1)}} - 1\right] = 1.61\,\text{V}$$

The output voltage corresponding to V_{IH} is, from Equation (16.34),

$$v_O = V_{OLU} = \frac{(5 - 0.8) + (25.1)(1.61 - 0.8)}{1 + 2(25.1)} = 0.479\,\text{V}$$

The noise margins are

$$\text{NM}_L = V_{IL} - V_{OLU} = 0.8 - 0.479 = 0.321\,\text{V}$$

and

$$\text{NM}_H = V_{OHU} - V_{IH} = 4.2 - 1.61 = 2.59\,\text{V}$$

Comment: This example shows that the two noise margins are not necessarily equal. In addition, the output voltage corresponding to V_{IH} is less than the threshold voltage of a driver transistor driven by the inverter. This means that as long as v_I remains in the range $V_{IH} \le v_I \le V_{OH}$, no logic error will be transmitted through a digital system.

The results of this example can be compared to the flip-flop conditions in Figure 16.12. Figure 16.15 shows two inverters in series, including series-voltage noise sources. If the input is high at $V_{OHU} = 4.2\,\text{V}$ and if a noise source of $\delta V_H = \text{NM}_H = 2.59\,\text{V}$ is included, the input to inverter G_1 is 1.61 V corresponding to V_{IH}, which is the minimum value corresponding to a logic 1 level. With an input of $V_{IH} = 1.61\,\text{V}$ to G_1, the output is 0.479 V. If a noise source of $\delta V_L = \text{NM}_L = 0.321\,\text{V}$ is present, then input to G_2 is 0.8 V corresponding to V_{IL}, which is the minimum value corresponding to a logic 0 level. With an input of $V_{IL} = 0.8\,\text{V}$ to G_2, the output is 4.2 V. If the output is connected back to the input, the resulting flip-flop configuration is in a stable state, although at the edge of switching. If the noise sources δV_H and δV_L increase slightly above the noise margins, the flip-flop will switch states.

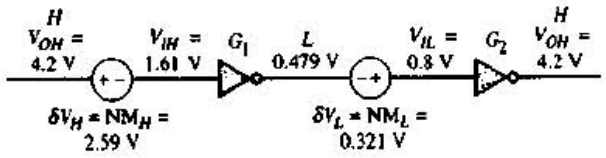


Figure 16.15 Two inverters and two series-voltage noise sources with the noise margins from Example 16.5

Test Your Understanding

16.6 For the transistors in the NMOS inverter with enhancement load, the parameters are $V_{TN} = 0.85\,\text{V}$ and $K_D/K_L = 16$. Let $V_{DD} = 5\,\text{V}$. Determine V_{IL}, V_{IH}, NM_L, and NM_H. (Ans. $V_{IH} = 1.81\,\text{V}$, $V_{IL} = 0.85\,\text{V}$, $\text{NM}_L = 0.259\,\text{V}$, $\text{NM}_H = 2.34\,\text{V}$)

Depletion-Load Inverter

The general transfer characteristic of the NMOS inverter with depletion load is shown in Figure 16.16. The point V_{IL} occurs when the load is biased in the nonsaturation region and the driver is biased in the saturation region. The relationship between the input and output voltages is given by Equation (16.25). Taking the derivative with respect to v_I, we obtain

$$2K_D[v_I - V_{TND}] = K_L\left[2(-V_{TNL})\left(-\frac{dv_O}{dv_I}\right) - 2(V_{DD} - v_O)\left(-\frac{dv_O}{dv_I}\right)\right] \tag{16.36}$$

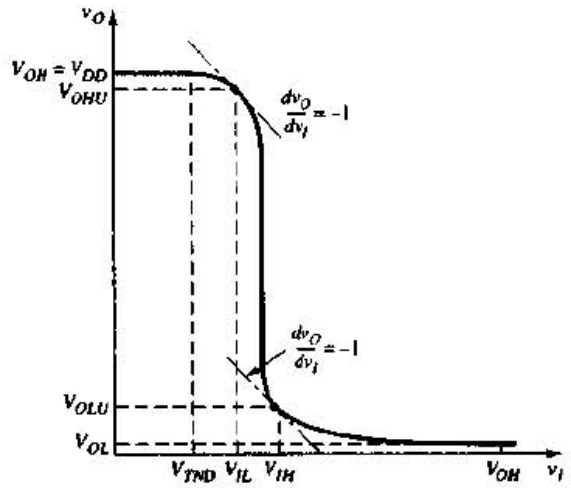


Figure 16.16 Voltage transfer characteristic, inverter with depletion load, and the defined voltage limits V_{IL} and V_{IH}

Setting the derivative equal to -1 and solving for v_O yields

$$v_O = V_{OHU} = (V_{DD} + V_{TNL}) + \left(\frac{K_D}{K_L}\right)(v_I - V_{TND}) \tag{16.37}$$

Combining Equations (16.37) and (16.25), we then have

$$v_I = V_{IL} = V_{TND} + \frac{(-V_{TNL})}{\sqrt{\left(\frac{K_D}{K_L}\right)\left(1 + \frac{K_D}{K_L}\right)}} \tag{16.38}$$

The point V_{IH} occurs when the load is biased in the saturation region and the driver is biased in the nonsaturation region. The relationship between the input and output voltages is given by Equation (16.29(b)). Taking the derivative with respect to v_I, we find

$$\frac{K_D}{K_L}\left[2(v_I - V_{TND})\frac{dv_O}{dv_I} + 2v_O - 2v_O\frac{dv_O}{dv_I}\right] = 0 \tag{16.39}$$

Setting the derivative equal to -1 and solving for v_O yields

$$v_O = V_{OLU} = \frac{(v_I - V_{TND})}{2} \tag{16.40}$$

Combining Equations (16.40) and (16.29(b)), we then have

$$v_I = V_{IH} = V_{TND} + \frac{2(-V_{TNL})}{\sqrt{3\left(\frac{K_D}{K_L}\right)}} \tag{16.41}$$

The noise margins are then determined from Equations (16.31(a)) and (16.31(b)). The calculation of noise margins in an inverter with depletion load is considered in Exercise 16.7.

Test Your Understanding

16.7 For the transistors in the NMOS inverter with depletion load, the parameters are: $V_{TND} = 1\,V$, $V_{TNL} = -1.7\,V$, and $K_D/K_L = 5$. Let $V_{DD} = 5\,V$. Determine V_{IL}, V_{IH}, NM_L and NM_H. (Ans. $V_{IL} = 1.31\,V$, $V_{IH} = 1.88\,V$, $NM_L = 0.87\,V$, $NM_H = 2.97\,V$)

16.1.4 Body Effect

Up to this point, we have neglected the body effect and assumed that all threshold voltages are constant. Figure 16.17 shows enhancement-load and depletion-load NMOS inverters with the substrates of all transistors tied to ground. A nonzero source-to-body voltage will then exist in the load devices. In fact, the source terminal of the depletion load can increase to V_{DD}. The threshold voltage given by Equation (16.1) must be used in the circuit calculations for the load transistor. This significantly complicates the equations for the voltage transfer calculations, making them very cumbersome for hand analyses.

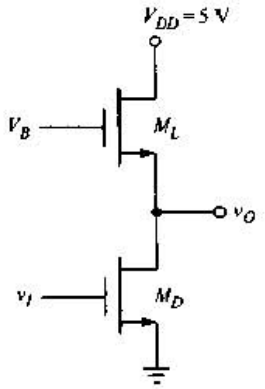


Figure 16.17 NMOS inverters, showing substrate connections to ground potential: (a) enhancement-load inverter and (b) depletion-load inverter

Example 16.6 **Objective:** Determine the change in the high output voltage of an NMOS inverter with enhancement load, taking the body effect into account.

Consider the NMOS inverter with enhancement load in Figure 16.17(a). The transistor parameters are $V_{TNDO} = V_{TNLO} = 0.8\,V$ and $K_D/K_L = 16$. Assume the inverter is biased at $V_{DD} = 5\,V$, assume the body effect coefficient is $\gamma = 0.90\,V^{1/2}$, and let $\phi_{fp} = 0.365\,V$.

Solution: When $v_I < V_{TNDO}$, the driver is cut off and the output goes high. From Equation (16.16(b)), the maximum output voltage is

$$v_{O,max} = V_{OH} = V_{DD} - V_{TNL}$$

where V_{TNL} is, from Equation (16.1),

$$V_{TNL} = V_{TNLO} + \gamma\left[\sqrt{2\phi_{fp} + V_{SB}} - \sqrt{2\phi_{fp}}\right]$$

From Figure 16.17(a), we see that $V_{SB} = v_O$; therefore, Equation (16.16(b)) can be written

$$v_{O,\max} = V_{DD} - \left\{V_{TNLO} + \gamma\left[\sqrt{2\phi_{fp} + v_{O,\max}} - \sqrt{2\phi_{fp}}\right]\right\}$$

Defining $v_{O,\max} \equiv V_{OH}$, we have

$$V_{OH} - 4.97 = -0.90\sqrt{0.73 + V_{OH}}$$

Squaring both sides and rearranging terms yields

$$V_{OH}^2 - 9.29V_{OH} + 24.9 = 0$$

Consequently, the maximum output voltage, or the logic 1 level, is

$$V_{OH} = 3.19\,\text{V}$$

Comment: Neglecting the body effect, the logic 1 output level is

$$V_{OH} = V_{DD} - V_{TNL} = 5 - 0.8 = 4.2\,\text{V}$$

The body effect, then, can significantly influence the logic high state of the NMOS inverter with enhancement load. These results also impact the inverter noise margins.

The source and body terminals of the depletion load device in the NMOS inverter shown in Figure 16.17(b) are not at the same potential when the output goes high. However, when the driver is cut off, the drain-to-source voltage of the depletion device must be zero in order that $v_{O,\max} = V_{OH} = V_{DD}$.

Computer Simulation: A computer analysis of the inverters in Figure 16.17 was performed, neglecting the body effect and taking the body effect into account. The threshold voltage of the depletion load device is $V_{TNLO} = -2\,\text{V}$ and the ratio K_D/K_L of the depletion load inverter is 4.82.

The body effect changes the voltage transfer characteristics of both the enhancement load and depletion load inverters. Figure 16.18(a) shows the voltage transfer characteristics for the enhancement-load inverter. The circuit and transistor parameters are the same as given in this example. For $v_I = 0$, the output voltage is 3.15 V when the

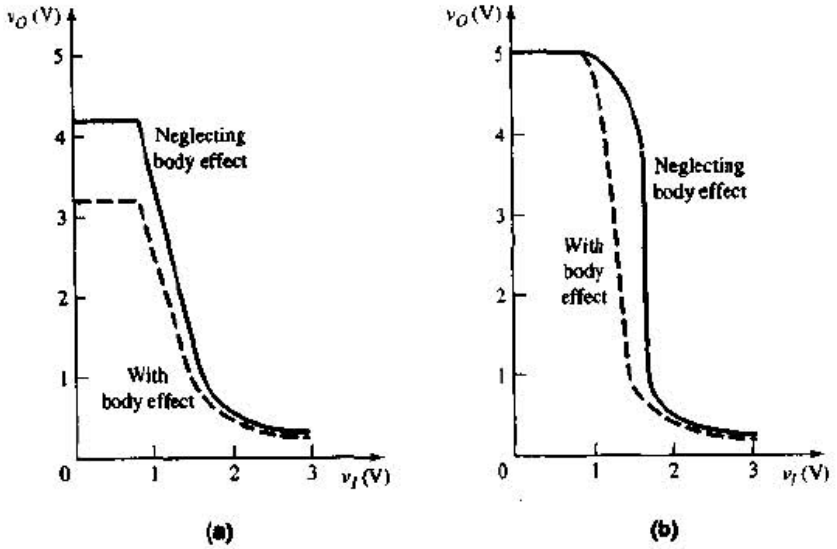


Figure 16.18 Voltage transfer characteristics of NMOS inverters with and without the body effect (a) enhancement load and (b) depletion load

body effect is taken into account. This compares favorably with the 3.19 V from the hand analysis.

Figure 16.18(b) shows the voltage transfer characteristics for the depletion-load inverter. As discussed, the output voltage is 5 V in the high state, which is independent of the body effect. However, the characteristics during the transition region are a function of the body effect.

16.1.5 Transient Analysis of NMOS Inverters

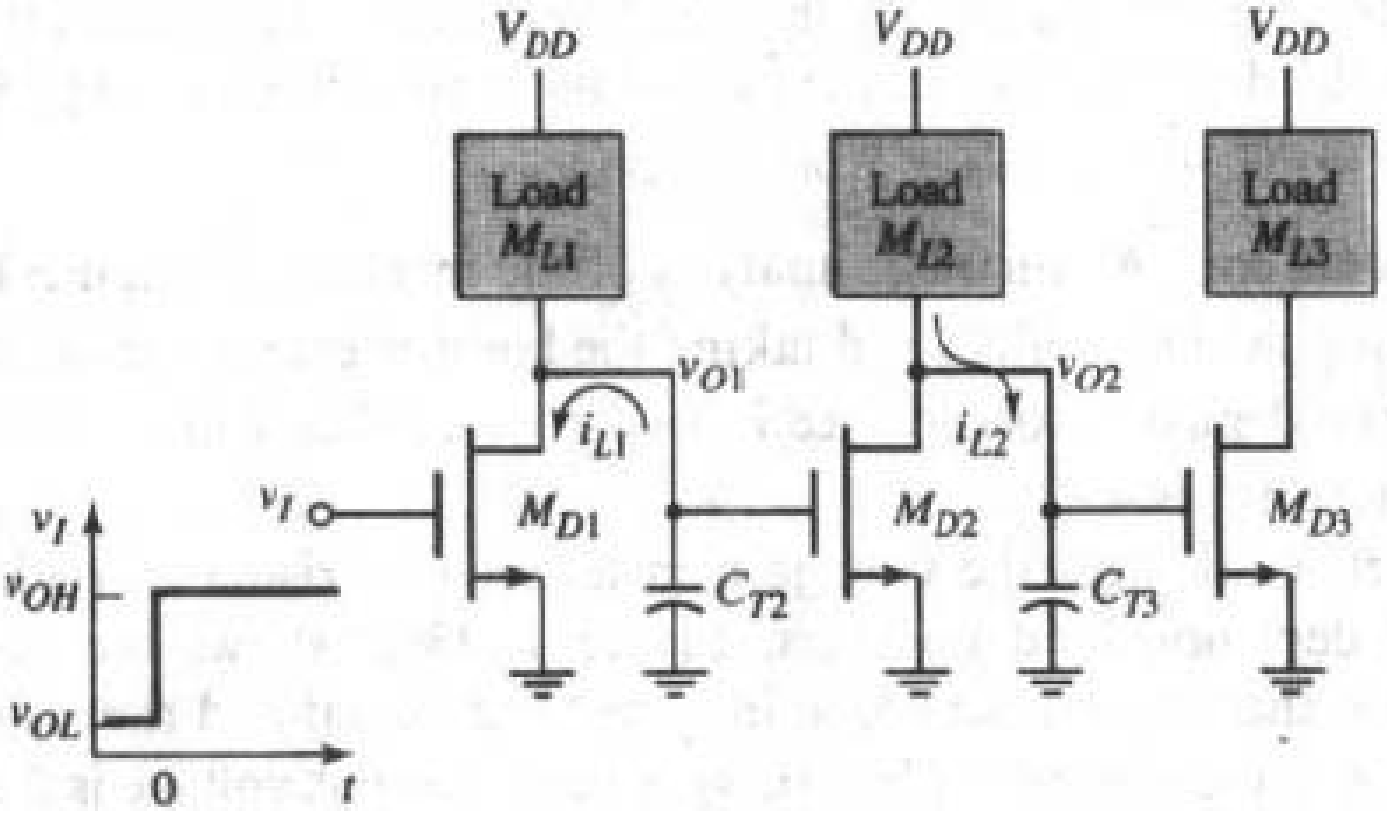
Figure 16.19 shows three NMOS inverters in cascade with generalized load devices. The output of the first inverter sees an effective capacitance looking into the gate of M_{D2}, and the output of the second inverter sees an effective capacitance looking into the gate of M_{D3}. The capacitances are C_{T2} and C_{T3}, respectively, and they include the transistor input capacitances, as well as any parasitic capacitances due to the interconnect lines between the inverter stages.

Figure 16.19 Three generalized NMOS inverters in cascade, for calculating transient effects

If input v_I is switched from high to low at time $t = 0$, driver M_{D1} cuts off, and output v_{O1} begins to go high. The effective load capacitance C_{T2} must then be charged by a current through load device M_{L1}. As v_{O1} increases, M_{D2} turns on and v_{O2} begins to go low. For v_{O2} to go low, the effective load capacitance C_{T3} must discharge. The discharge current i_{L2} is the difference between the driver current in M_{D2} and the load current in M_{L2}.

The rate at which capacitance C_{T2} charges is a function of the current–voltage characteristics of the load device. Figure 16.20 shows the characteristics of the driver transistor M_{D1}, with superimposed load curves for the three basic load components. These load curves neglect the body effect.

The load devices are assumed to be scaled such that $I_{D,max}$ is the same for each device. The constant current over a wide range of v_{DS} provided by the depletion load implies that this type of inverter will switch a capacitive load more rapidly than the other two types of inverter configurations. The rate at which the voltage across a load capacitance changes is a direct function of the current through the capacitance.


V_{DD}
V_{DD}
V_{DD}
Load M_{L1}
Load M_{L2}
Load M_{L3}
v_{O1}
v_{O2}
i_{L1}
i_{L2}
v_I
M_{D1}
M_{D2}
M_{D3}
C_{T2}
C_{T3}
v_I
v_{OH}
v_{OL}
0
t

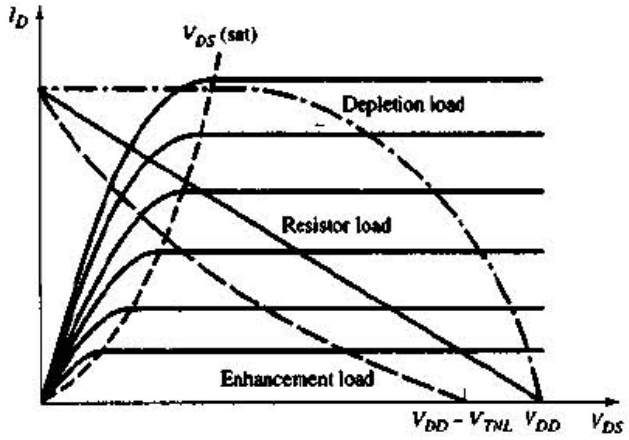


Figure 16.20 Driver transistor characteristics and load curves for the three types of NMOS inverters

Figure 16.21 shows the enhancement- and depletion-load inverter load curves that take the body effect into account. For the enhancement-load inverter, the resulting high output voltage is less than that when the body effect is neglected, as previously determined. Also, the current in the inverter is less over the entire voltage range, which implies that the switching times will be longer. For the depletion-load inverter, the resulting current is not a constant over a wide voltage range, and it is less than that when the body effect is neglected. Therefore, the time required to charge the load capacitance is longer.

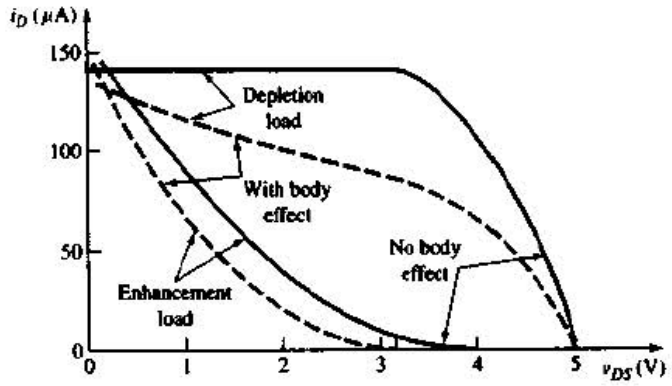


Figure 16.21 Load curves, enhancement-load and depletion-load NMOS inverters, with and without body effect

Figure 16.22 shows the switching characteristics of NMOS inverters with depletion-loads as determined from a PSpice analysis. A series of inverters, such as shown in Figure 16.19, was used. The width-to-length ratio of the load devices was $(W/L)_L = 1$ and that of the driver transistors was $(W/L)_D = 4$. The effective load capacitances were assumed to be 0.5 pF, which is larger than would normally be encountered in an IC.

Figure 16.22(a) shows the input (v_{O1}) and output (v_{O2}) voltage characteristics when the body effect is neglected. The fall time is relatively short, since the load capacitance discharges through the larger driver transistor. The rise

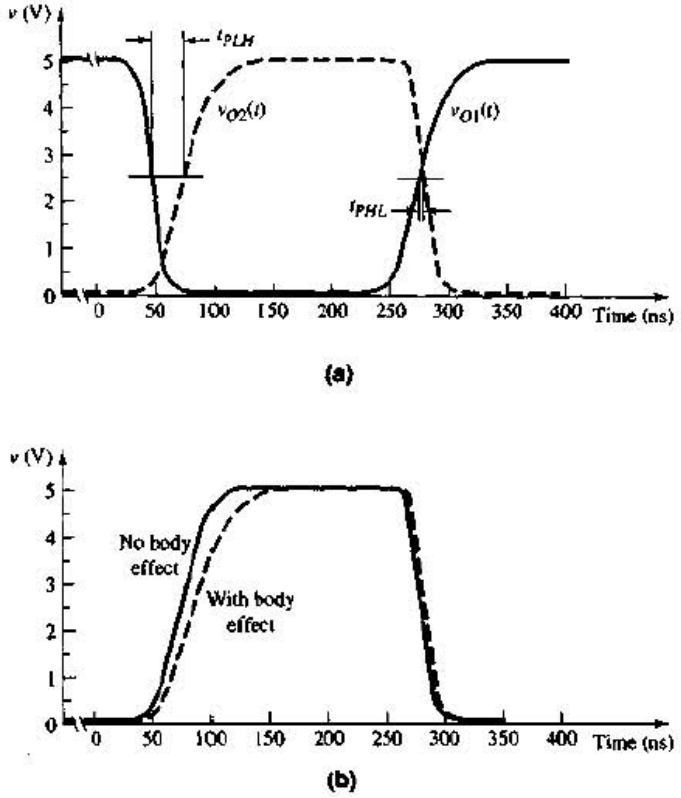


Figure 16.22 (a) Switching characteristics of an NMOS inverter with depletion load and (b) switching characteristics with and without body effect

time is longer, since the load capacitance is charged by the current through the smaller load transistor. The propagation delay times are shown in the figure.

The rise and fall times of the inverter with and without the body effect are shown in Figure 16.22(b). The rise time is longer when the body effect is taken into account, since the current in the load device is smaller. This was shown by the load curve in Figure 16.21. The fall time is not affected by the body effect, since the substrate of the driver is connected to the source terminal, which is at ground potential.

16.2 NMOS LOGIC CIRCUITS

NMOS logic circuits are formed by combining driver transistors in parallel, series, or series–parallel combinations to produce a desired output logic function.

16.2.1 NMOS NOR and NAND Gates

The NMOS NOR logic gate contains additional driver transistors connected in parallel. Figure 16.23 shows a two-input NMOS NOR logic gate with a depletion load. If $A = B =$ logic 0, then both M_{DA} and M_{DB} are cut off and $v_O = V_{DD}$. If $A =$ logic 1 and $B =$ logic 0, then M_{DB} is cut off and the NMOS inverter configuration with M_L and M_{DA} is the same as previously considered.

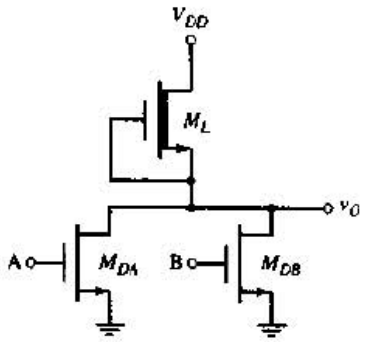


Figure 16.23 Two-input NMOS NOR logic gate with depletion load

and the output voltage goes low. Similarly, if A = logic 0 and B = logic 1, we again have the same inverter configuration.

If $A = B$ = logic 1, then both M_{DA} and M_{DB} turn on and the two driver transistors are effectively in parallel. The value of the output voltage now changes slightly. Figure 16.24 shows the NOR gate when both input voltages are a logic 1. From our previous analysis, we can assume that the two driver transistors are biased in the nonsaturation region and the load device is biased in the saturation region. We then have

$$i_{DL} = i_{DA} + i_{DB}$$

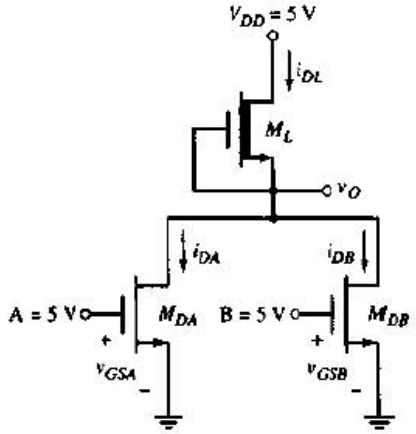


Figure 16.24 Two-input NMOS NOR logic gate for Example 16.7

which in general terms can be written

$$K_L[v_{GSL} - V_{TNL}]^2 = K_{DA}\left[2(v_{GSA} - V_{TNA})v_{DSA} - v_{DSA}^2\right] + K_{DB}\left[2(v_{GSB} - V_{TNB})v_{DSB} - v_{DSB}^2\right] \quad (16.42)$$

If we assume the two driver transistors are identical, then the driver conduction parameters and threshold voltages are also identical, or $K_{DA} = K_{DB} \equiv K_D$ and $V_{TNA} = V_{TNB} \equiv V_{TND}$. Noting that $v_{GSL} = 0$, $v_{GSA} = v_{GSB} = V_{DD}$, and $v_{DSA} = v_{DSB} = v_O$, we can write Equation (16.42) as

$$[-V_{TNL}]^2 = 2\left(\frac{K_D}{K_L}\right)[2(V_{DD} - V_{TND})v_O - v_O^2] \tag{16.43}$$

Equation (16.43) shows that when both drivers are conducting, the effective width-to-length ratio of the composite driver transistor doubles. This means that the output voltage becomes slightly smaller when both inputs are high.

Example 16.7 **Objective:** Determine the low output voltage of an NMOS NOR circuit.

Consider the NOR circuit in Figure 16.24 biased at $V_{DD} = 5\,\text{V}$. Assume that $k_n' = 35\,\mu\text{A/V}^2$. Also assume the width-to-length ratios of the load and driver transistors are $(W/L)_L = 1$ and $(W/L)_D = 4$, respectively. Let $V_{TND} = 0.8\,\text{V}$ and $V_{TNL} = -2\,\text{V}$. Neglect the body effect.

Solution: If, for example, A = logic 1 = 5 V and B = logic 0, then M_{DB} is cut off. The output voltage is determined from Equation (16.29(b)), which is

$$\frac{K_D}{K_L}[2(v_I - V_{TND})v_O - v_O^2] = (-V_{TNL})^2$$

or

$$\left(\frac{4}{1}\right)[2(5 - 0.8)v_O - v_O^2] = (2)^2$$

The output voltage is found to be

$$v_O = 0.121\,\text{V}$$

If both inputs go high, then

$$A = B = V_{DD} = 5\,\text{V}$$

and the output voltage can be found using Equation (16.43), which is

$$[-V_{TNL}]^2 = 2\left(\frac{K_D}{K_L}\right)[2(V_{DD} - V_{TND})v_O - v_O^2]$$

or

$$(2)^2 = 2\left(\frac{4}{1}\right)[2(5 - 0.8)v_O - v_O^2]$$

The output voltage is found to be

$$v_O = 0.060\,\text{V}$$

Comment: An NMOS NOR gate must be designed to achieve a specified V_{OL} output voltage when only one input is high. This will give the largest logic 0 value. When more than one input is high, the output voltage is smaller than the specified V_{OL} value, since the effective width-to-length ratio of the composite driver transistor increases.

The NMOS NAND logic gate contains additional driver transistors connected in series. Figure 16.25 shows a two-input NMOS NAND logic gate with a depletion load. If both $A = B =$ logic 0, or if either A or B is a logic 0, at least

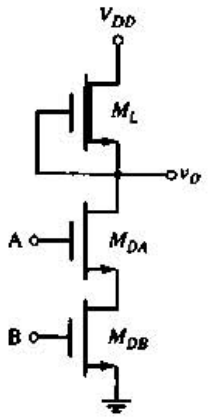


Figure 16.25 Two-input NMOS NAND logic gate with depletion load

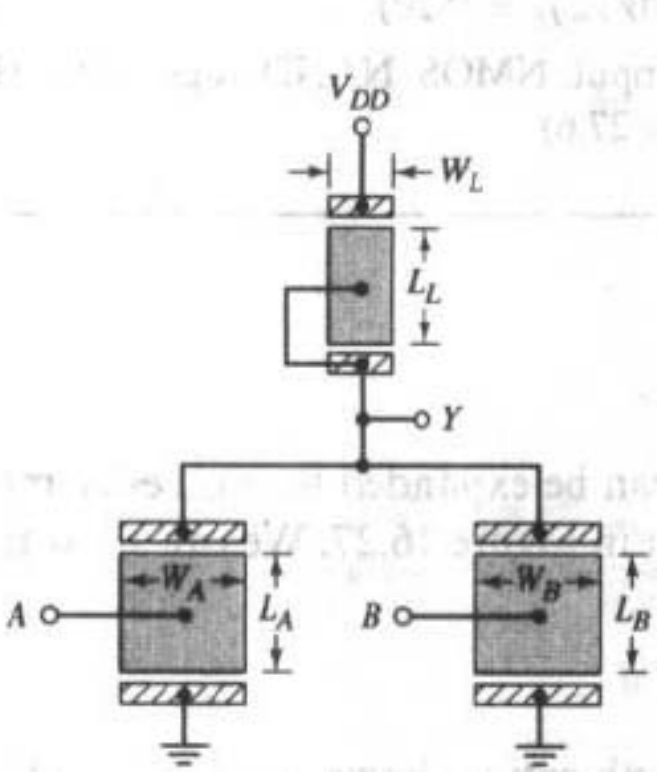
one driver is cut off, and the output is high. If both $A = B =$ logic 1, then the composite driver of the NMOS inverter conducts and the output goes low.

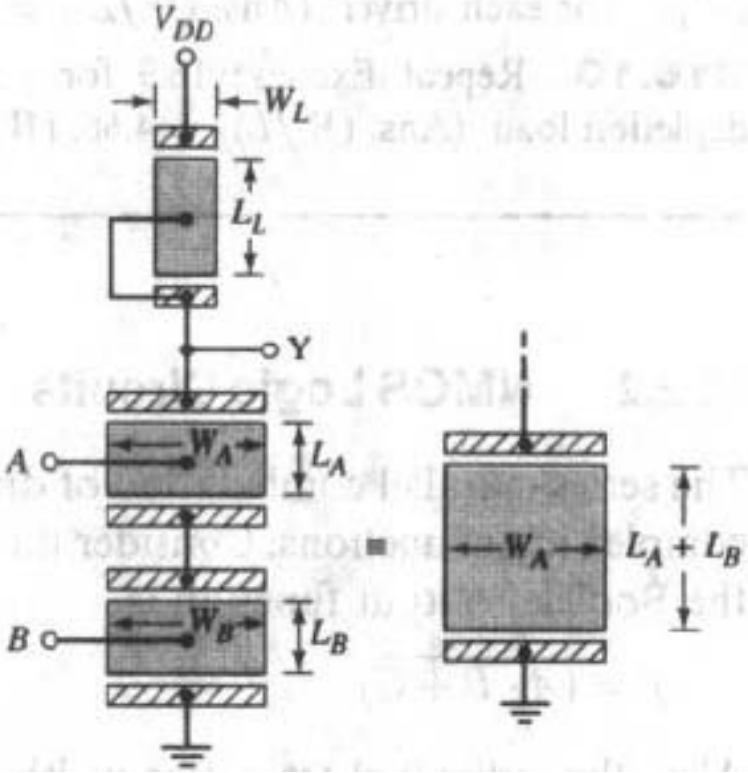
Since the gate-to-source voltages of M_{DA} and M_{DB} are not equal, determining the actual voltage V_{OL} of a NAND gate is difficult. The drain-to-source voltages of M_{DA} and M_{DB} must adjust themselves to produce the same current. In addition, if the body effect is also included, the analysis becomes even more difficult. Since the two driver transistors are in series, a good approximation assumes that the width-to-length ratio of the drivers must be twice that of a single driver in an NMOS inverter to achieve a given V_{OL} value.

The composite width-to-length ratios of the driver transistors in the two-input NMOS NOR and NAND gates are shown schematically in Figure 16.26. For the NOR gate, the effective *width* doubles; for the NAND gates, the effective *length* doubles.

(a) (b)

Figure 16.26 Composite width-to-length ratios of driver transistors in two-input NMOS logic configurations (a) NOR and (b) NAND


V_{DD}
W_L
L_L
Y
A
W_A
L_A
B
W_B
L_B


V_{DD}
W_L
L_L
Y
A
W_A
L_A
B
W_B
L_B
W_A
$L_A + L_B$

Example 16.8 **Objective:** Determine the low output voltage of an NMOS NAND circuit.

Consider the NAND circuit in Figure 16.25 biased at $V_{DD} = 5\,V$. Assume $k_n' = 35\,\mu A/V^2$. Also assume the width-to-length ratio of the load transistor is $(W/L)_L = 1$. Let $V_{TND} = 0.8\,V$ and $V_{TNL} = -2\,V$. Neglect the body effect.

Solution: From a PSpice analysis for $A = B =$ logic 1 $= 5\,V$, the output voltage is 0.060 V when the width-to-length ratio of each driver transistor is $(W/L)_D = 16$.

This result correlates very well with the results of Example 16.7. For the two-input NOR gate, the effective width of the composite driver doubles, or $(W/L)_C = 2 \times 4 = 8$, which results in an output voltage of 0.060 V. For the two input NAND gate, the effective length of the composite driver doubles, or $(W/L)_C = (1/2) \times 16 = 8$, which also results in an output voltage of 0.060 V.

Comment: If an N-input NMOS NAND logic gate were to be fabricated, then the width-to-length ratio of the drivers would need to be N times that of a single driver in an NMOS inverter to achieve a given value of V_{OL}. The increase in the required area of the driver transistors in a NAND logic gate means that logic gates with more than three or four inputs are not attractive.

Test Your Understanding

[Note: In the following exercises, assume $k_n' = 35\,\mu A/V^2$ for all NMOS transistors.]

***16.8** Consider the two-input NMOS NOR logic gate shown in Figure 16.23 biased at $V_{DD} = 10\,V$. Let $(W/L)_L = 2$, $(W/L)_D = 10$, $V_{TND} = 1.5\,V$, and $V_{TNL} = -3\,V$. Neglect the body effect. (a) Determine V_{OL} when: (i) $A =$ logic 1, $B =$ logic 0, and (ii) $A = B =$ logic 1. (b) Calculate the power dissipation in the circuit when: (i) $A =$ logic 1, $B =$ logic 0, and (ii) $A = B =$ logic 1. (Ans. (a) $V_{OL} = 0.107\,V$, $V_{OL} = 0.0531\,V$ (b) $P = 3.15\,mW$)

D16.9 Design a three-input NMOS NOR logic gate with depletion load such that $V_{OL}(\text{max}) = 0.12\,V$. Let $V_{DD} = 5\,V$, $V_{TND} = 0.8\,V$, and $V_{TNL} = -1.4\,V$. The maximum power dissipation in the circuit must be 0.8 mW. Determine (W/L) for the load and (W/L) for each driver. (Ans. $(W/L)_L = 4.66$, $(W/L)_D = 9.20$)

D16.10 Repeat Exercise 16.9 for a three-input NMOS NAND logic gate with depletion load. (Ans. $(W/L)_L = 4.66$, $(W/L)_D = 27.6$)

16.2.2 NMOS Logic Circuits

The series–parallel combination of drivers can be expanded to synthesize more complex logic functions. Consider the circuit in Figure 16.27. We can show that the Boolean output function is

$$f = \overline{(A \cdot B + C)}$$

Also, the individual transistor width-to-length ratios shown produce an effective K_D/K_L ratio of 4 for an effective single inverter when only M_{DA} and M_{DB} are conducting, or only M_{DC} is conducting. The actual complexity of the Boolean function is limited since the required width-to-length ratios of individual transistors may become unreasonably large.

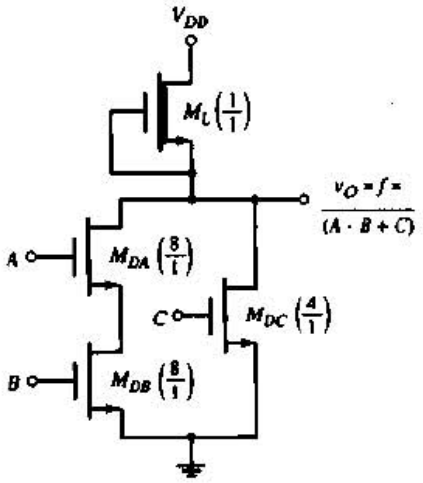


Figure 16.27 NMOS logic circuit example

Two additional logic functions are the exclusive-OR and exclusive-NOR. Figure 16.28 shows a circuit configuration that produces the exclusive-OR function. If $A = B =$ logic 1, a path exists from the output to ground through drivers M_{DA} and M_{DB}, and the output goes low. Similarly, if $A = B =$ logic 0, which means that $\bar{A} = \bar{B} =$ logic 1, a path exists from the output to ground through the drivers $M_{D\bar{B}}$ and $M_{D\bar{A}}$, and the output gain goes low. For all other input logic signal combinations, the output is isolated from ground so the output goes high.

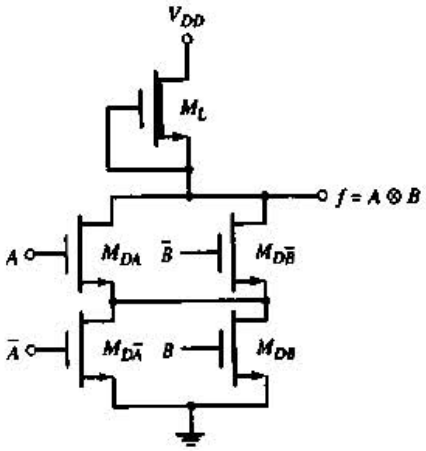
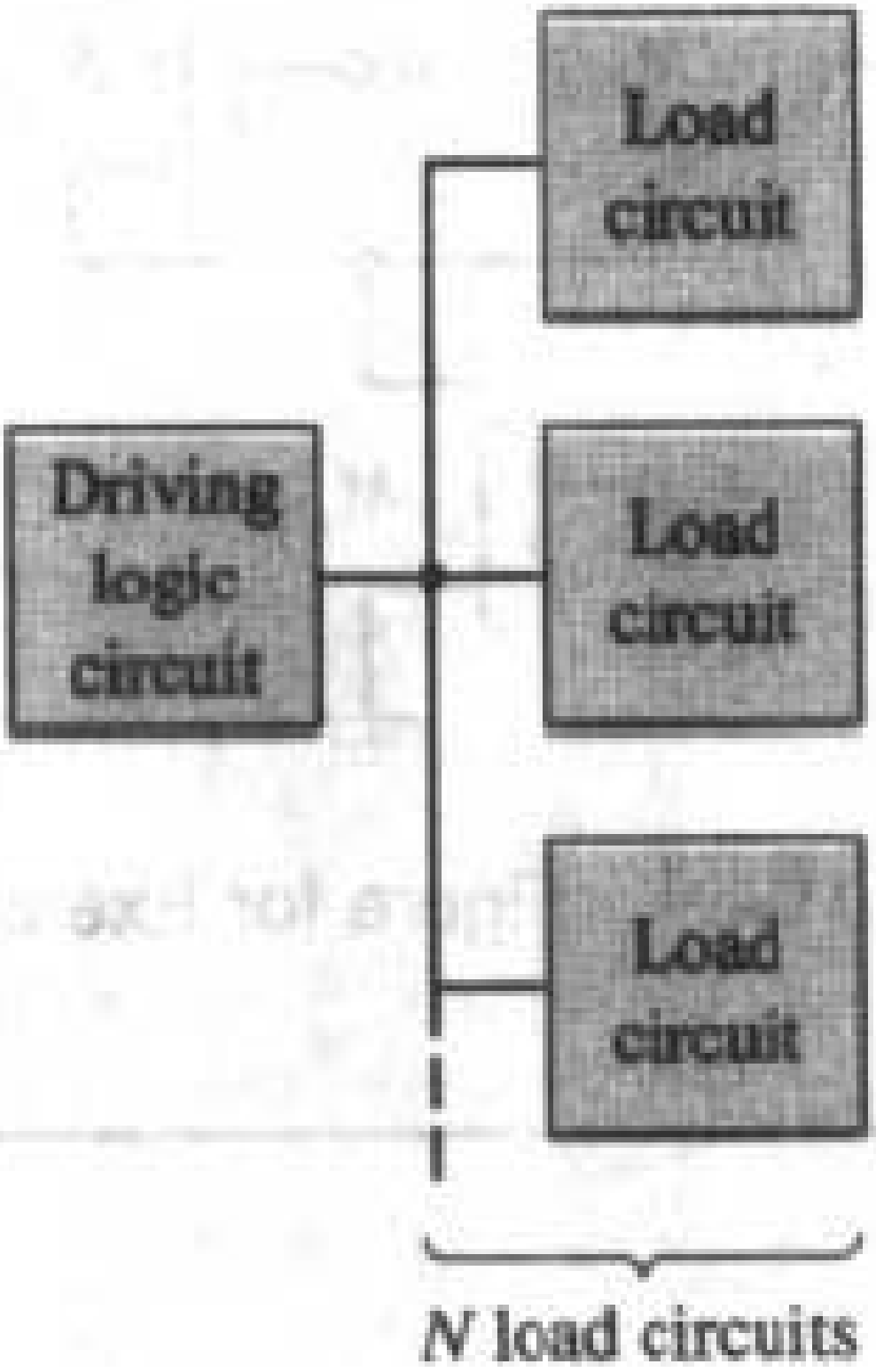


Figure 16.28 NMOS exclusive-OR logic gate

Figure 16.29 Logic circuit driving N load circuits

16.2.3 Fanout

An NMOS inverter or NMOS logic gate must be capable of driving more than one load, as shown in Figure 16.29. It is assumed that each load is identical to the driver logic circuit. The number of identical-load circuits connected to the output of a driver logic circuit is defined as the **fanout**. For MOS logic circuits,


Load circuit
Driving logic circuit
Load circuit
Load circuit
N load circuits

the inputs to the load circuits are the oxide-insulated gates of the MOS transistors; therefore, the static loading caused by multiple driver loads is so small that the dc transfer curve is essentially identical to a no-load condition. The dc characteristics of MOS logic circuits are unaffected by the fanout to other MOS logic inputs. However, the load capacitance due to a large fanout seriously degrades the switching speed and propagation delay times. Consequently, maintaining the propagation delay time below a specified maximum value determines the fanout of MOS digital circuits.

Test Your Understanding

[Assume $k_n' = 35\,\mu A/V^2$.]

***16.11** Consider the NMOS logic circuit in Figure 16.30. Let $V_{TN} = 0.7\,V$ for each transistor and assume all driver transistors are identical. (a) If $(W/L)_L = 0.5$, determine (W/L) for the drivers such that $V_{OL}(\max) = 0.15\,V$. (b) Determine the maximum power dissipation in the logic circuit. (Ans. (a) $(W/L)_D = 13.6$ (b) $P = 753\,\mu W$)

16.12 Repeat Exercise 16.11 for the NMOS logic circuit in Figure 16.31, except assume that the threshold voltage of the load device is $V_{TN} = -1.2\,V$. (Ans. (a) $(W/L)_D = 1.14$ (b) $P = 63\,\mu W$)

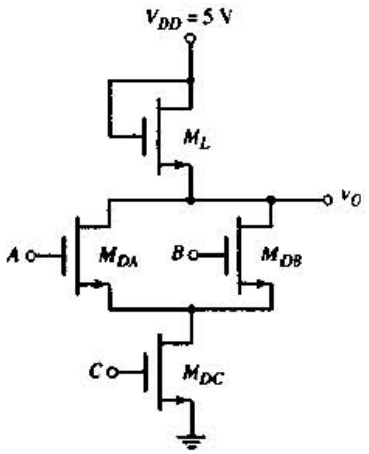


Figure 16.30 Figure for Exercise 16.11

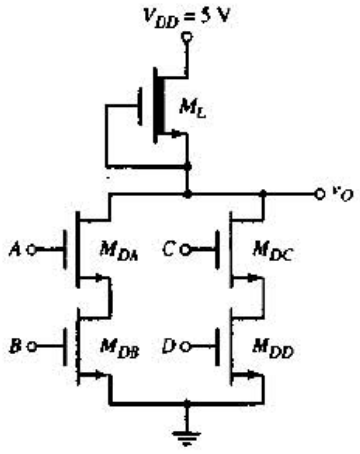


Figure 16.31 Figure for Exercise 16.12

16.3 CMOS INVERTER

Complementary MOS, or CMOS, circuits contain both n-channel and p-channel MOSFETs. As we will see, the power dissipation in CMOS logic circuits is much smaller than in NMOS circuits, which makes CMOS very attractive. We briefly review the characteristics of p-channel transistors, and will then analyze the CMOS inverter, which is the basis of most CMOS logic circuits. We will examine the CMOS NOR and NAND gates and other basic CMOS logic circuits, covering power dissipation, noise margin, fanout, and switching characteristics.

16.3.1 p-Channel MOSFET Revisited

Figure 16.32 shows a simplified view of a p-channel MOSFET. The p- and n-regions are reversed from those in an n-channel device. Again, the channel length is L and the channel width is W. Usually in any given fabrication process, the channel length is the same for all devices, so the channel width W is the variable in logic circuit design.

Figure 16.32 Simplified cross section of p-channel MOSFET

Figure 16.33(a) shows the simplified circuit symbol for the p-channel enhancement-mode device. When the body or substrate connection is needed, we will use the symbol shown in Figure 16.33(b). Usually, the p-channel depletion-mode device is not used in CMOS digital circuits; therefore, it is not addressed here.

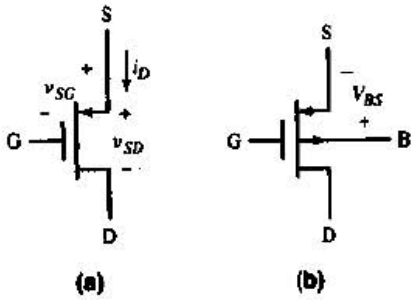
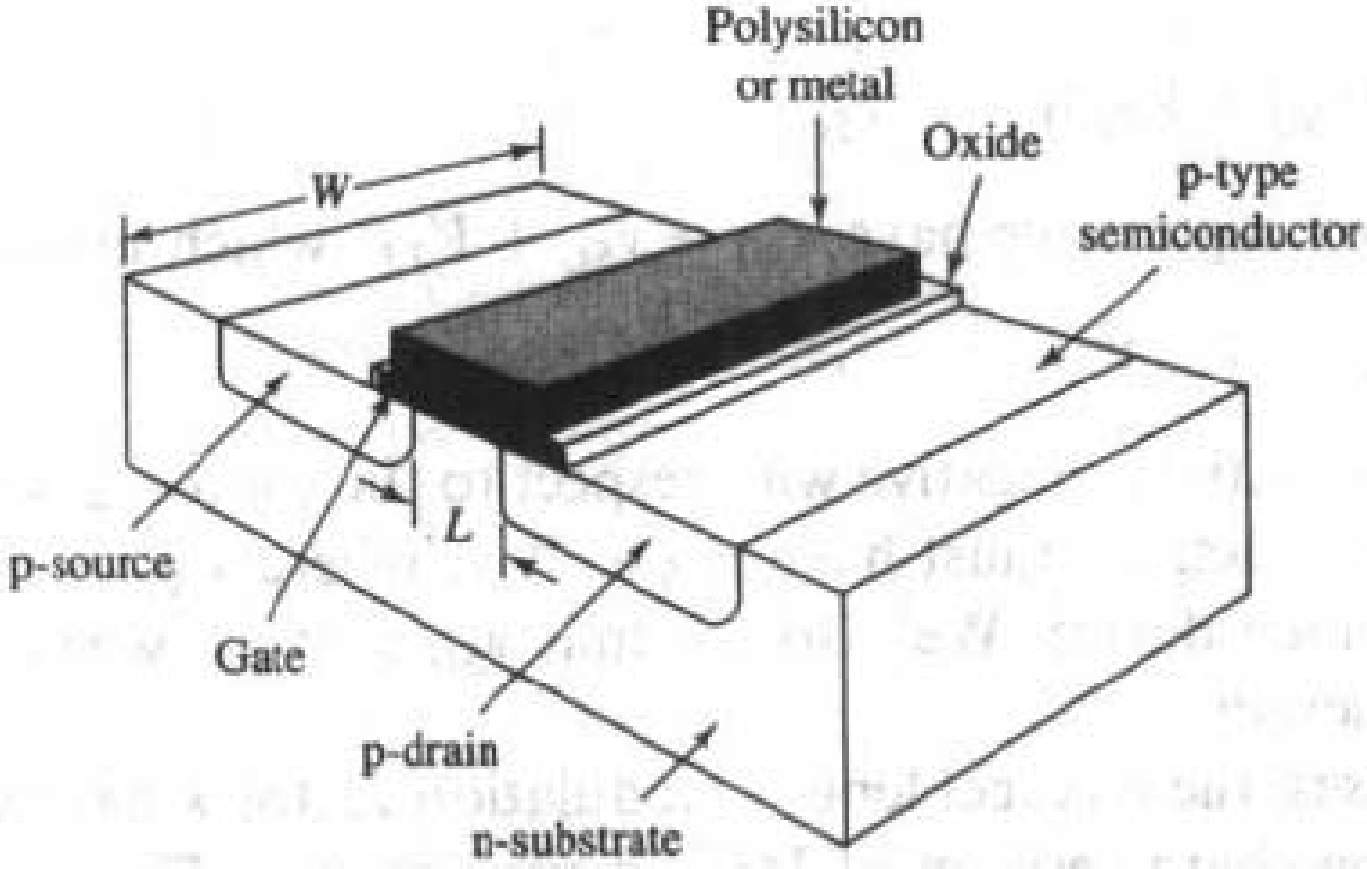


Figure 16.33 (a) Simplified circuit symbol, p-channel enhancement-mode MOSFET and (b) circuit symbol showing substrate connection

Normally, in an integrated circuit, more than one p-channel device will be fabricated in the same n-substrate so the p-channel transistors will exhibit a body effect. The n-substrate is connected to the most positive potential. The source terminal may be negative with respect to the substrate; therefore, voltage V_{BS} may exist between the body and the source. The threshold voltage is

$$\begin{aligned} V_{TP} &= V_{TPO} - \frac{\sqrt{2\varepsilon\varepsilon_s N_d}}{C_{ox}}\left[\sqrt{2\phi_{fn} + V_{BS}} - \sqrt{2\phi_{fn}}\right] \\ &= V_{TPO} - \gamma\left[\sqrt{2\phi_{fn} + V_{BS}} - \sqrt{2\phi_{fn}}\right] \end{aligned} \tag{16.44}$$


Polysilicon
or metal
Oxide
p-type
semiconductor
W
L
p-source
Gate
p-drain
n-substrate

where V_{TPO} is the threshold voltage for zero body-to-source voltage, or $V_{BS} = 0$. The parameter N_d is the n-substrate doping concentration and ϕ_{fn} is a potential related to the substrate doping. The parameter γ is the body effect coefficient.

The current–voltage characteristic of the p-channel MOSFET are functions of both the electrical and geometric properties of the device. When the transistor is biased in the nonsaturation region, we have $v_{SD} \leq v_{SG} + V_{TP}$. Therefore,

$$i_D = K_P[2(v_{SG} + V_{TP})v_{SD} - v_{SD}^2] \tag{16.45(a)}$$

In the saturation region, we have $v_{SD} \geq v_{SG} + V_{TP}$, which means that

$$i_D = K_P(v_{SG} + V_{TP})^2 \tag{16.45(b)}$$

The gate potential is negative with respect to the source. For the p-channel transistor to conduct, we must have $v_{GS} < V_{TP}$, where V_{TP} is negative for an enhancement-mode device. We also see that $v_{SG} > |V_{TP}|$ when the p-channel device is conducting.

In most cases, the channel length modulation factor λ has very little effect on the operating characteristics of MOS digital circuits. Therefore, the term λ is assumed to be zero unless otherwise stated.

The transition point, which separates the nonsaturation and saturation bias regions, is given by

$$v_{SD} = v_{SD}(\text{sat}) = v_{SG} + V_{TP} \tag{16.46}$$

The parameter K_P is the conduction parameter and is given by

$$K_P = \left(\frac{1}{2}\mu_p C_{ox}\right)\left(\frac{W}{L}\right) = \frac{k_p'}{2}\frac{W}{L} \tag{16.47}$$

As before, the hole mobility μ_p and oxide capacitance C_{ox} are assumed to be constant for all devices. The hole mobility in p-channel silicon MOSFETs is approximately one-half the electron mobility μ_n in n-channel silicon MOSFETs. This means that a p-channel device width must be approximately twice as large as that of an n-channel device in order that the two devices be electrically equivalent (that is, that they have the same conduction parameter values).

Small Geometry Effects

The same small geometry effects apply to the p-channel devices as we discussed for the n-channel devices in Section 16.1.1. As with the NMOS inverters and logic circuits, we can use Equations (16.45(a)), (16.45(b)), and (16.46) as first-order equations in the design of NMOS logic circuits. The basic operation and behavior of CMOS logic circuits can be predicted using these first-order equations.

16.3.2 DC Analysis of the CMOS Inverter

The **CMOS inverter**, shown in Figure 16.34, is a series combination of a p-channel and an n-channel MOSFET. The gates of the two MOSFETs are connected together to form the input and the two drains are connected together

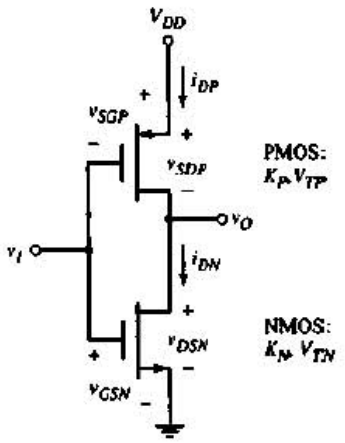


Figure 16.34 CMOS inverter

to form the output. Both transistors are enhancement-mode devices. The parameters of the NMOS are denoted by K_N and V_{TN}, where $V_{TN} > 0$, and the parameters of the PMOS are denoted by K_P and V_{TP}, where $V_{TP} < 0$.

Figure 16.35 shows a simplified cross section of a CMOS inverter. In this process, a separate p-well region is formed within the starting n-substrate. The n-channel device is fabricated in the p-well region and the p-channel device is fabricated in the n-substrate. Although other approaches, such as an n-well in a p-substrate, are also used to fabricate CMOS circuits, the important point is that the processing is more complicated for CMOS circuits than for NMOS circuits. However, the advantages of CMOS digital logic circuits over NMOS circuits justify their use.

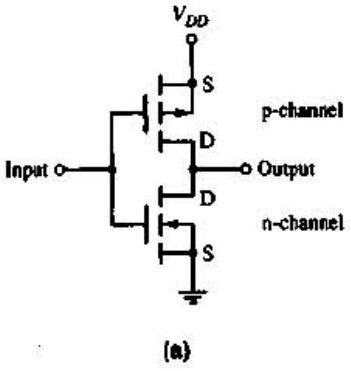
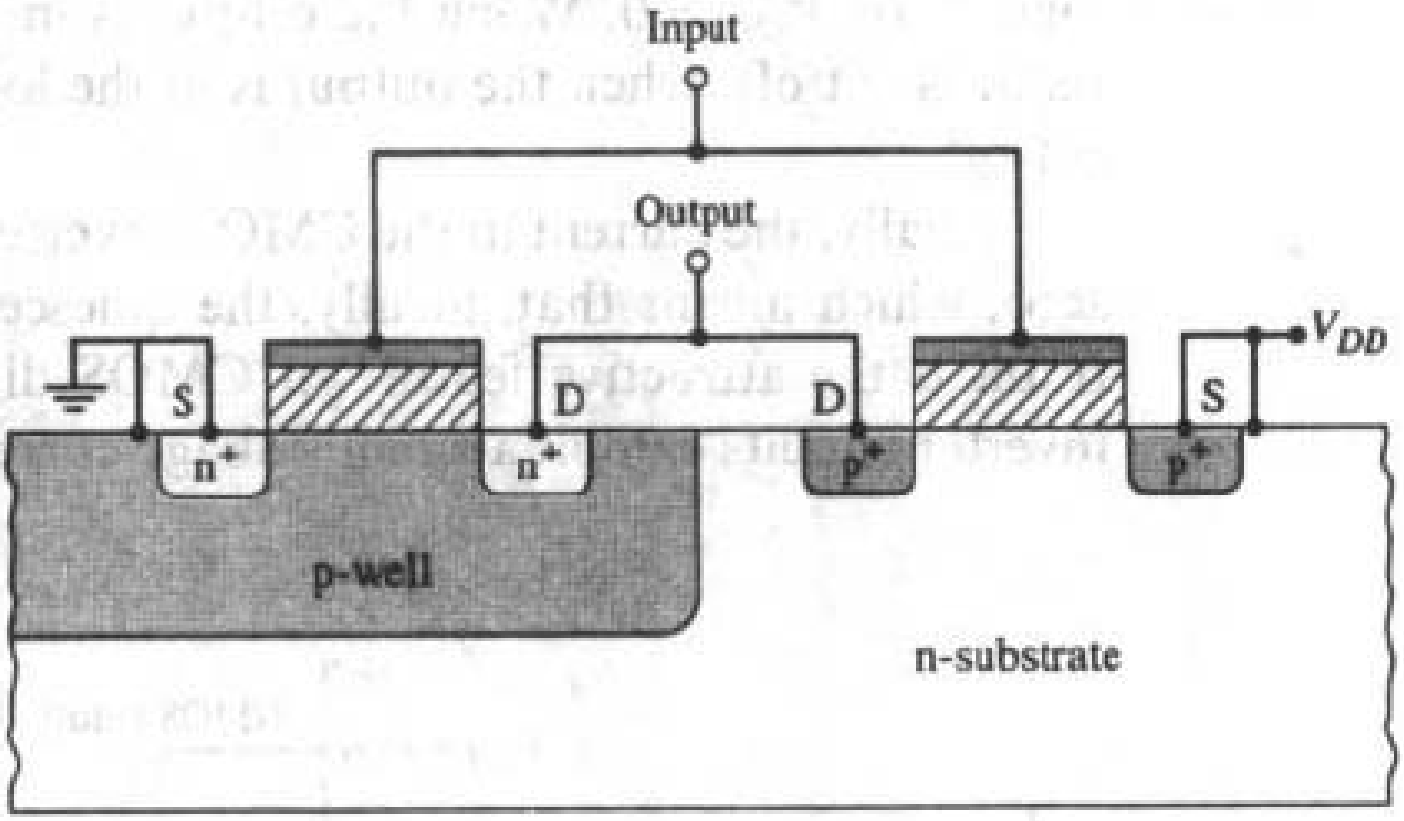


Figure 16.35 Simplified cross section, CMOS inverter

Voltage Transfer Curve

Figure 16.36 shows the transistor characteristics for both the n- and p-channel devices. We can determine the voltage transfer characteristics of the inverter by evaluating the various transistor bias regions. For $v_I = 0$, the NMOS device is cut off, $i_{DN} = 0$, and $i_{DP} = 0$. The PMOS source-to-gate voltage is V_{DD}, which



(b)

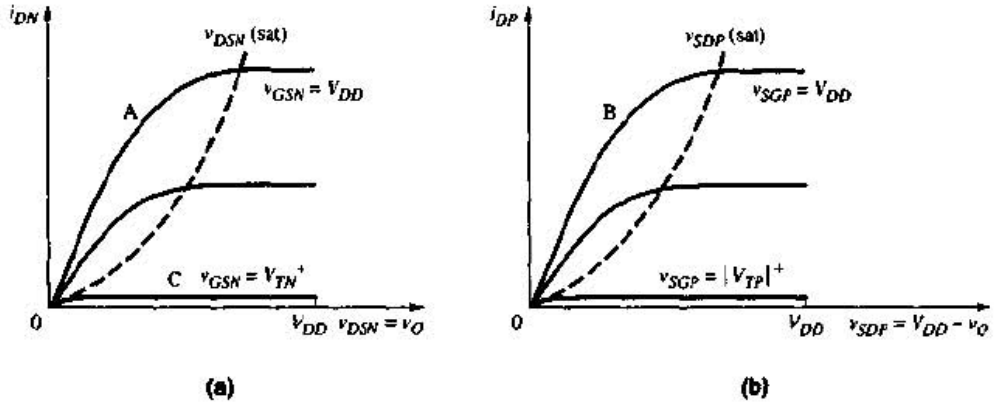


Figure 16.36 Current–voltage characteristics, (a) NMOS transistor and (b) PMOS transistor

means that the PMOS is biased on the curve marked B in Figure 16.36(b). Since the only point on the curve corresponding to $i_{DP} = 0$ occurs at $v_{SDP} = 0 = V_{DD} - v_O$, the output voltage is $v_O = V_{DD}$. This condition exists as long as the NMOS transistor is cut off, or $v_I \leq V_{TN}$.

For $v_I = V_{DD}$, the PMOS device is cut off, $i_{DP} = 0$, and $i_{DN} = 0$. The NMOS gate-to-source voltage is V_{DD} and the NMOS is biased on the curve marked A in Figure 16.36(a). The only point on the curve corresponding to $i_{DN} = 0$ occurs at $v_{DSN} = v_O = 0$. The output voltage is zero as long as the PMOS transistor is cut off, or $v_{SGP} = V_{DD} - v_I \leq |V_{TP}|$. This means that the input voltage is in the range $V_{DD} - |V_{TP}| \leq v_I \leq V_{DD}$.

Figure 16.37 shows the voltage transfer characteristics generated thus far for the CMOS inverter. The more positive output voltage corresponds to a logic 1, or $V_{OH} = V_{DD}$, and the more negative output voltage corresponds to a logic 0, or $V_{OL} = 0$. When the output is in the logic 1 state, the NMOS transistor is cut off; when the output is in the logic 0 state, the PMOS transistor is cut off.

Ideally, the current in the CMOS inverter in either steady-state condition is zero, which means that, ideally, the quiescent power dissipation is zero. This result is the attractive feature of CMOS digital circuits. In actuality, CMOS inverter circuits exhibit a small leakage current in both steady-state conditions,

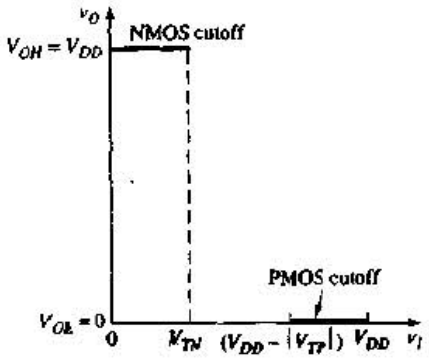


Figure 16.37 CMOS inverter output voltage for input voltage in either high state or low state

due to the reverse-biased pn junctions. However, the power dissipation may be in the nanowatt range rather than in the milliwatt range of NMOS inverters. Without this feature, VLSI would not be possible.

When the input voltage is just greater than V_{TN}, or

$$v_I = v_{GSN} = V_{TN}^+$$

the NMOS begins to conduct and the Q-point falls on the curve marked C in Figure 16.36(a). The current is small and $v_{DSN} \cong V_{DD}$, which means that the NMOS is biased in the saturation region. The PMOS source-to-drain voltage is small, so the PMOS is biased in the nonsaturation region. Setting $i_{DN} = i_{DP}$, we can write

$$K_N[v_{GSN} - V_{TN}]^2 = K_P[2(v_{SGP} + V_{TP})v_{SDP} - v_{SDP}^2] \tag{16.48}$$

Relating the gate-to-source and drain-to-source voltages in each transistor to the inverter input and output voltages, respectively, we can rewrite Equation (16.48) as follows:

$$K_N[v_I - V_{TN}]^2 = K_P[2(V_{DD} - v_I + V_{TP})(V_{DD} - v_O) - (V_{DD} - v_O)^2] \tag{16.49}$$

Equation (16.49) relates the input and output voltages as long as the NMOS is biased in the saturation region and the PMOS is biased in the nonsaturation region.

The transition point for the PMOS is defined from

$$v_{SDP}(\text{sat}) = v_{SGP} + V_{TP} \tag{16.50}$$

Using Figure 16.38, Equation (16.50) can be written

$$V_{DD} - V_{OPt} = V_{DD} - V_{IPt} + V_{TP} \tag{16.51(a)}$$

or

$$V_{OPt} = V_{IPt} - V_{TP} \tag{16.51(b)}$$

where V_{OPt} and V_{IPt} are the PMOS output and input voltages, respectively, at the transition point.

The transition point for the NMOS is defined from

$$v_{DSN}(\text{sat}) = v_{GSN} - V_{TN} \tag{16.52(a)}$$

or

$$V_{ONt} = V_{INt} - V_{TN} \tag{16.52(b)}$$

where V_{ONt} and V_{INt} are the NMOS output and input voltages, respectively, at the transition point.

On the basis that V_{TP} is negative for an enhancement-mode PMOS, Equations (16.51(b)) and (16.52(b)) are plotted in Figure 16.38. We determine the input voltage at the transition points by setting the two drain currents equal to each other when both transistors are biased in the saturation region. The result is

$$K_N(v_{GSN} - V_{TN})^2 = K_P(v_{SGP} + V_{TP})^2 \tag{16.53}$$

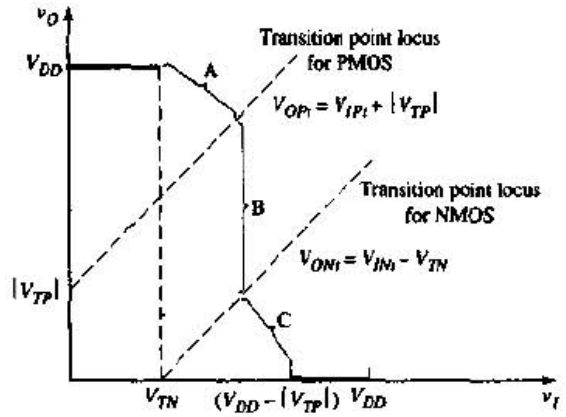


Figure 16.38 Regions of the CMOS transfer characteristics indicating NMOS and PMOS transistor bias conditions. The NMOS device is biased in the saturation region in areas A and B and in the nonsaturation region in area C. The PMOS device is biased in the saturation region in areas B and C and in the nonsaturation region in area A.

With the gate-to-source voltages related to the input voltage, Equation (16.53) becomes

$$K_N(v_I - V_{TN})^2 = K_P(V_{DD} - v_I + V_{TP})^2 \tag{16.54}$$

For this ideal case, the output voltage does not appear in Equation (16.54), and the input voltage is a constant, as long as the two transistors are biased in the saturation region.

Voltage v_I from Equation (16.54) is the input voltage at the PMOS and NMOS transition points. Solving for v_I, we find that

$$v_I = v_{It} = \frac{V_{DD} + V_{TP} + \sqrt{\dfrac{K_N}{K_P}}V_{TN}}{1 + \sqrt{\dfrac{K_N}{K_P}}} \tag{16.55}$$

For $v_I > V_{It}$, the NMOS is biased in the nonsaturation region and the PMOS is biased in the saturation region. Again equating the two drain currents, we have

$$K_N\left[2(v_{GSN} - V_{TN})v_{DSN} - v_{DSN}^2\right] = K_P(v_{SGP} + V_{TP})^2 \tag{16.56}$$

Also, relating the gate-to-source and drain-to-source voltages to the input and output voltages, respectively, modifies Equation (16.56) as follows:

$$K_N\left[2(v_I - V_{TN})v_O - v_O^2\right] = K_P(V_{DD} - v_I + V_{TP})^2 \tag{16.57}$$

Equation (16.57) relates the input and output voltages as long as the NMOS is biased in the nonsaturation region and the PMOS in the saturation region. Figure 16.39 shows the complete voltage transfer curve.

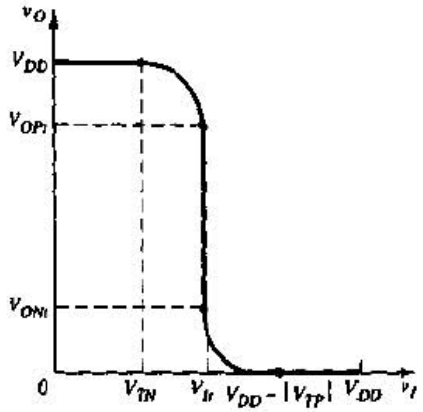


Figure 16.39 Complete voltage transfer characteristics, CMOS inverter

---

Example 16.9 **Objective:** Determine the critical voltages on the voltage transfer curve of a CMOS inverter.

Consider a CMOS inverter biased at $V_{DD} = 5\,V$ with transistor parameters of $K_N = K_P$ and $V_{TN} = -V_{TP} = 1\,V$. Then consider another CMOS inverter biased at $V_{DD} = 10\,V$ with the same transistor parameters.

Solution: For $V_{DD} = 5\,V$, the input voltage at the transition points is, from Equation (16.55),

$$V_{It} = \frac{5 + (-1) + \sqrt{1} \cdot 1}{1 + \sqrt{1}} = 2.5\,V$$

The output voltage at the transition point for the PMOS is, from Equation (16.51(b)),

$$V_{OPt} = V_{IPt} - V_{TP} = 2.5 - (-1) = 3.5\,V$$

and the output voltage at the transition point for the NMOS is, from Equation (16.52(b)),

$$V_{ONt} = V_{INt} - V_{TN} = 2.5 - 1 = 1.5\,V$$

For $V_{DD} = 10\,V$ and the same transistor parameters, the critical voltages are:

$$V_{It} = 5\,V \qquad V_{OPt} = 6\,V \qquad V_{ONt} = 4\,V$$

Comment: The two voltage transfer curves are shown in Figure 16.40. The figure depicts another advantage of CMOS technology, that is, CMOS circuits can be biased over a wide range of voltages.

Test Your Understanding

16.13 The CMOS inverter in Figure 16.34 is biased at $V_{DD} = 10\,V$, and the transistor threshold voltages are $V_{TN} = 2\,V$ and $V_{TP} = -2\,V$. Sketch the voltage transfer curve and show the critical points, as in Figure 16.39, for: (a) $K_N/K_P = 1$, (b) $K_N/K_P = 0.5$, and (c) $K_N/K_P = 2$. (Ans. (a) $V_{It} = 5\,V$, $V_{OPt} = 7\,V$, $V_{ONt} = 3\,V$ (b) $V_{It} = 5.51\,V$, $V_{OPt} = 7.51\,V$, $V_{ONt} = 3.51\,V$ (c) $V_{It} = 4.49\,V$, $V_{OPt} = 6.49\,V$, $V_{ONt} = 2.49\,V$)

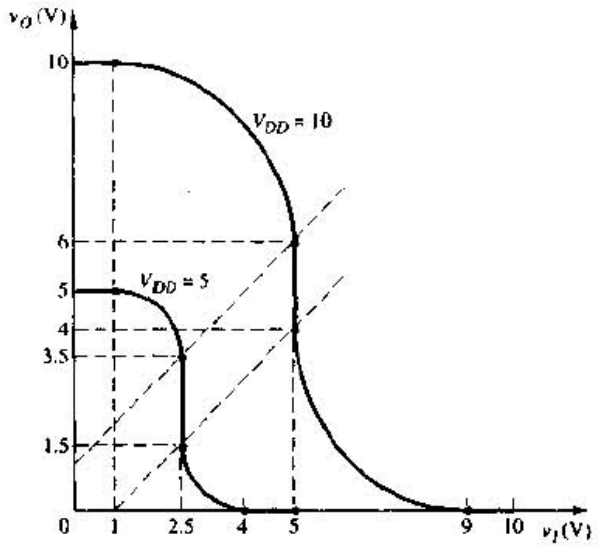


Figure 16.40 Voltage transfer characteristics, CMOS inverter biased at either $V_{DD} = 5$V or $V_{DD} = 10$V

CMOS Inverter Currents

When the CMOS inverter input voltage is either a logic 0 or a logic 1, the current in the circuit is zero, since one of the transistors is cut off. When the input voltage is in the range $V_{TN} < v_I < V_{DD} - |V_{TP}|$, both transistors are conducting and a current exists in the inverter.

When the NMOS transistor is biased in the saturation region, the current in the inverter is controlled by v_{GSN} and the PMOS source-to-drain voltage adjusts such that $i_{DP} = i_{DN}$. This condition is demonstrated in Equation (16.48). We can write

$$i_{DN} = i_{DP} = K_N(v_{GSN} - V_{TN})^2 = K_N(v_I - V_{TN})^2 \qquad \textbf{(16.58(a))}$$

Taking the square root yields

$$\sqrt{i_{DN}} = \sqrt{i_{DP}} = \sqrt{K_N}\,(v_I - V_{TN}) \qquad \textbf{(16.58(b))}$$

As long as the NMOS transistor is biased in the saturation region, the square root of the CMOS inverter current is a linear function of the input voltage.

When the PMOS transistor is biased in the saturation region, the current in the inverter is controlled by v_{SGP} and the NMOS drain-to-source voltage adjusts such that $i_{DP} = i_{DN}$. This condition is demonstrated in Equation (16.56). Using Equation (16.57), we can write that

$$i_{DN} = i_{DP} = K_P(V_{DD} - v_I + V_{TP})^2 \qquad \textbf{(16.59(a))}$$

Taking the square root yields

$$\sqrt{i_{DN}} = \sqrt{i_{DP}} = \sqrt{K_P}\,(V_{DD} - v_I + V_{TP}) \qquad \textbf{(16.59(b))}$$

As long as the PMOS transistor is biased in the saturation region, the square root of the CMOS inverter current is also a linear function of the input voltage.

Figure 16.41 shows plots of the square root of the inverter current for two values of V_{DD} bias. These curves are quasi-static characteristics in that no

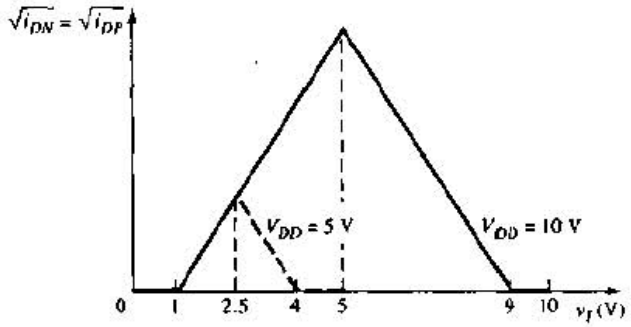


Figure 16.41 Square root of inverter current versus input voltage, CMOS inverter biased at either $V_{DD} = 5$ V or $V_{DD} = 10$ V

current is diverted into a capacitive load. At the inverter switching point, both transistors are biased in the saturation region and both transistors influence the current. At the switching point, the actual current characteristic does not have a sharp discontinuity in the slope. The channel length modulation parameter λ also influences the current characteristics at the peak value. However, the curves in Figure 16.41 are excellent approximations.

Test Your Understanding

16.14 Consider a CMOS inverter biased at $V_{DD} = 5$ V, with transistor threshold voltages of $V_{TN} = +0.8$ V and $V_{TP} = -0.8$ V. Calculate the peak current in the inverter for: (a) $K_N = K_P = 50\,\mu A/V^2$, and (b) $K_N = K_P = 200\,\mu A/V^2$. (Ans. (a) $i_D(\max) = 145\,\mu A$ (b) $i_D(\max) = 578\,\mu A$)

16.3.3 Power Dissipation

In the quiescent or static state, in which the input is either a logic 0 or a logic 1, power dissipation in the CMOS inverter is virtually zero. However, during the switching cycle from one state to another, current flows and power is dissipated. The CMOS inverter and logic circuits are used to drive other MOS devices for which the input impedance is a capacitance. During the switching cycle, then, this load capacitance must be charged and discharged.

In Figure 16.42(a), the output switches from its low to its high state. The input is switched low, the PMOS gate is at zero volts, and the NMOS is cut off. The load capacitance C_L must be charged through the PMOS device. Power dissipation in the PMOS transistor is given by

$$P_P = i_L v_{SD} = i_L(V_{DD} - v_O) \tag{16.60}$$

The current and the output voltage are related by

$$i_L = C_L \frac{dv_O}{dt} \tag{16.61}$$

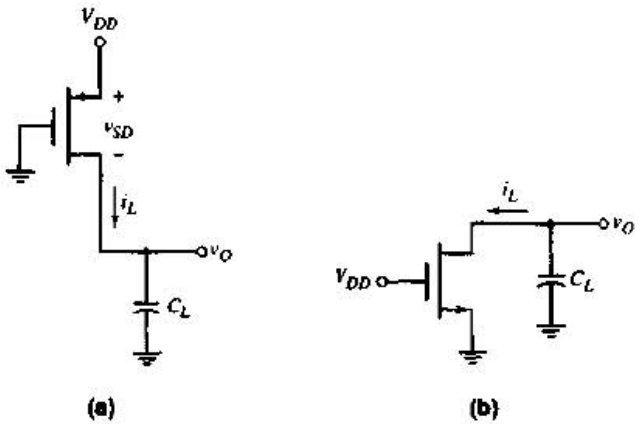


Figure 16.42 CMOS inverter when the output switches (a) low to high and (b) high to low

The energy dissipated in the PMOS device as the output switches from low to high is

$$\begin{aligned} E_P &= \int_0^\infty P_P dt = \int_0^\infty C_L(V_{DD} - v_O)\frac{dv_O}{dt}dt \\ &= C_L V_{DD}\int_0^{V_{DD}} dv_O - C_L\int_0^{V_{DD}} v_O dv_O \end{aligned} \tag{16.62}$$

which yields

$$E_P = C_L V_{DD} v_O\Big|_0^{V_{DD}} - C_L\frac{v_O^2}{2}\Big|_0^{V_{DD}} = \frac{1}{2}C_L V_{DD}^2 \tag{16.63}$$

After the output has switched high, the energy stored in the load capacitance is $(\frac{1}{2})C_L V_{DD}^2$. When the inverter input goes high, the output switches low, as shown in Figure 16.42(b). The PMOS device is cut off, the NMOS transistor conducts, and the load capacitance discharges through the NMOS device. All the energy stored in the load capacitance is dissipated in the NMOS device. As the output switches from high to low, the energy dissipated in the NMOS transistor is

$$E_N = \tfrac{1}{2}C_L V_{DD}^2 \tag{16.64}$$

The total energy dissipated in the inverter during one switching cycle is therefore

$$E_T = E_P + E_N = \tfrac{1}{2}C_L V_{DD}^2 + \tfrac{1}{2}C_L V_{DD}^2 = C_L V_{DD}^2 \tag{16.65}$$

If the inverter is switched at frequency f, the power dissipated in the inverter is

$$P = fE_T = fC_L V_{DD}^2 \tag{16.66}$$

Equation (16.66) shows that the power dissipated in a CMOS inverter is directly proportional to the switching frequency and to V_{DD}^2. The drive in digital IC design is toward lower supply voltages, such as 3 V or less.

Example 16.10 **Objective:** Calculate the power dissipation in a CMOS inverter.

Consider a CMOS inverter with a load capacitance of $C_L = 2\,\text{pF}$ biased at $V_{DD} = 5\,\text{V}$. The inverter switches at a frequency of $f = 100\,\text{kHz}$.

Solution: From Equation (16.66), power dissipation in the CMOS inverter is

$$P = fC_L V_{DD}^2 = (10^5)(2 \times 10^{-12})(5)^2 \Rightarrow 5\,\mu\text{W}$$

Comment: Previously determined values of static power dissipation in NMOS inverters were on the order of 500 μW; therefore, power dissipation in a CMOS inverter is substantially smaller. In addition, in most digital systems, only a small fraction of the logic gates change state during each clock cycle; consequently, the power dissipation in a CMOS digital system is substantially less than in an NMOS digital system of similar complexity.

The power dissipation is proportional to V_{DD}^2. In some digital circuits, such as digital watches, the CMOS logic circuits are biased at $V_{DD} = 1.5\,\text{V}$, so the power dissipation is substantially reduced.

Test Your Understanding

16.15 A CMOS inverter is biased at $V_{DD} = 3\,\text{V}$. The inverter drives an effective load capacitance of $C_L = 0.5\,\text{pF}$. Determine the maximum switching frequency such that the power dissipation is limited to $P = 0.10\,\mu\text{W}$. (Ans. $f = 22.2\,\text{kHz}$)

16.3.4 Noise Margin

Figure 16.43 shows the general voltage transfer function of a CMOS inverter. The parameters V_{IH} and V_{IL} determine the noise margins and are defined as the points at which

$$\frac{dv_O}{dv_I} = -1 \tag{16.67}$$

which is the same as for the NMOS inverters. For $v_I \leq V_{IL}$ and $v_I \geq V_{IH}$, the gain is less than unity and the output changes slowly with input voltage. However, when the input voltage is in the range $V_{IL} < v_I < V_{IH}$, the inverter gain is greater than unity, and the output signal changes rapidly with a change in the input voltage. This is the undefined range.

Point V_{IL} occurs when the NMOS is biased in the saturation region and the PMOS is biased in the nonsaturation region. The relationship between the input and output voltages is given by Equation (16.49). Taking the derivative with respect to v_I yields

$$2K_N[v_I - V_{TN}] = K_P\left[-2(V_{DD} - v_O) - 2(V_{DD} - v_I + V_{TP})\frac{dv_O}{dv_I} - 2(V_{DD} - v_O)\left(-\frac{dv_O}{dv_I}\right)\right] \tag{16.68}$$

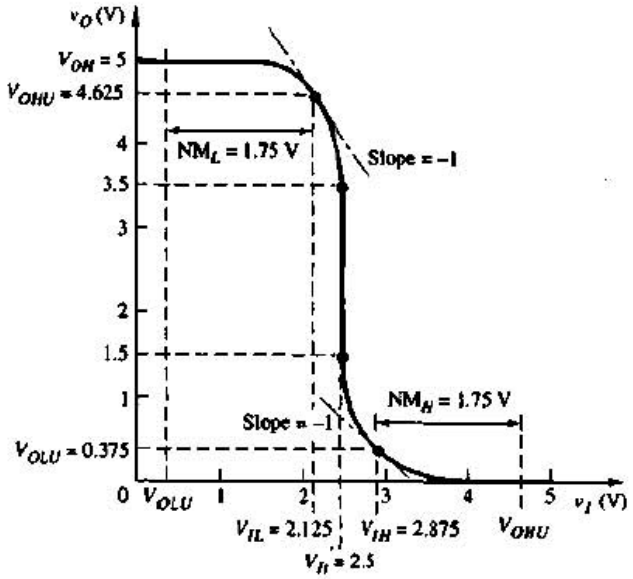


Figure 16.43 CMOS inverter voltage transfer characteristics with defined noise margins

Setting the derivative equal to −1, we have

$$K_N[v_I - V_{TN}] = -K_P[(V_{DD} - v_O) - (V_{DD} - v_I + V_{TP}) + (V_{DD} - v_O)] \tag{16.69}$$

Solving for v_O produces

$$v_O = V_{OHU} = \frac{1}{2}\left\{\left(1 + \frac{K_N}{K_P}\right)v_I + V_{DD} - \left(\frac{K_N}{K_P}\right)V_{TN} - V_{TP}\right\} \tag{16.70}$$

Combining Equations (16.70) and (16.49), we see that voltage V_{IL} is

$$v_I = V_{IL} = V_{TN} + \frac{(V_{DD} + V_{TP} - V_{TN}}{\left(\frac{K_N}{K_P} - 1\right)}\left[2\sqrt{\frac{\frac{K_N}{K_P}}{\frac{K_N}{K_P} + 3}} - 1\right] \tag{16.71}$$

If $K_N = K_P$, Equation (16.71) becomes indefinite, since a zero would exist in both the numerator and the denominator. However, when $K_N = K_P$, Equation (16.70) becomes

$$v_O = V_{OHU(K_N=K_P)} = \tfrac{1}{2}\{2v_I + V_{DD} - V_{TN} - V_{TP}\} \tag{16.72}$$

Substituting Equation (16.72) into Equation (16.49) yields a voltage V_{IL} of

$$v_I = V_{IL(K_N=K_P)} = V_{TN} + \tfrac{3}{8}(V_{DD} + V_{TP} - V_{TN}) \tag{16.73}$$

for $K_N = K_P$.

Point V_{IH} occurs when the NMOS is biased in the nonsaturation region and the PMOS is biased in the saturation region. The relationship between the

input and output voltages is given by Equation (16.57). Taking the derivative with respect to v_I yields

$$K_N\left[2(v_I - V_{TN})\frac{dv_O}{dv_I} + 2v_O - 2v_O\frac{dv_O}{dv_I}\right] = 2K_P(V_{DD} - v_I + V_{TP})(-1) \tag{16.74}$$

Setting the derivative equal to -1, we find that

$$K_N[-(v_I - V_{TN}) + v_O + v_O] = -K_P[V_{DD} - v_I + V_{TP}] \tag{16.75}$$

The output voltage v_O is then

$$v_O = V_{OLU} = \frac{v_I\left(1 + \frac{K_N}{K_P}\right) - V_{DD} - \left(\frac{K_N}{K_P}\right)V_{TN} - V_{TP}}{2\left(\frac{K_N}{K_P}\right)} \tag{16.76}$$

Combining Equations (16.76) and (16.57), yields voltage V_{IH} as

$$v_I = V_{IH} = V_{TN} + \frac{(V_{DD} + V_{TP} - V_{TN})}{\left(\frac{K_N}{K_P} - 1\right)}\left[\frac{2\frac{K_N}{K_P}}{\sqrt{3\frac{K_N}{K_P} + 1}} - 1\right] \tag{16.77}$$

Again, if $K_N = K_P$, Equation (16.77) becomes indefinite, since a zero would exist in both the numerator and the denominator. However, when $K_N = K_P$, Equation (16.76) becomes

$$v_O = V_{OLU(K_N = K_P)} = \tfrac{1}{2}\{2v_I - V_{DD} - V_{TN} - V_{TP}\} \tag{16.78}$$

Substituting Equation (16.78) into Equation (16.57) yields a voltage V_{IH} of

$$v_I = V_{IH(K_N = K_P)} = V_{TN} + \tfrac{5}{8}(V_{DD} + V_{TP} - V_{TN}) \tag{16.79}$$

Example 16.11 **Objective:** Determine the noise margins of a CMOS inverter.

Consider a CMOS inverter biased at $V_{DD} = 5\,\text{V}$. Assume the transistors are matched with $K_N = K_P$ and $V_{TN} = -V_{TP} = 1\,\text{V}$.

Solution: From Equation (16.57), the input voltage at the transition points, or the inverter switching point, is 2.5 V. Since $K_N = K_P$, V_{IL} is, from Equation (16.73)

$$V_{IL} = V_{TN} + \tfrac{3}{8}(V_{DD} + V_{TP} - V_{TN}) = 1 + \tfrac{3}{8}(5 - 1 - 1) = 2.125\,\text{V}$$

Point V_{IH} is, from Equation (16.79)

$$V_{IH} = V_{TN} + \tfrac{5}{8}(V_{DD} + V_{TP} - V_{TN}) = 1 + \tfrac{5}{8}(5 - 1 - 1) = 2.875\,\text{V}$$

The output voltages at points V_{IL} and V_{IH} are determined from Equations (16.72) and (16.78), respectively. They are

$$V_{OHU} = \tfrac{1}{2}[2V_{IL} + V_{DD} - V_{TN} - V_{TP}]$$
$$= \tfrac{1}{2}[2(2.125) + 5 - 1 + 1] = 4.625\,\text{V}$$

and

$$V_{OLU} = \tfrac{1}{2}[2V_{IH} - V_{DD} - V_{TN} - V_{TP}]$$
$$= \tfrac{1}{2}[2(2.875) - 5 - 1 + 1] = 0.375\,\text{V}$$

The noise margins are therefore

$$\text{NM}_L = V_{IL} - V_{OLU} = 2.125 - 0.375 = 1.75\,\text{V}$$

and

$$\text{NM}_H = V_{OHU} - V_{IH} = 4.625 - 2.875 = 1.75\,\text{V}$$

Comment: The results of this example are shown in Figure 16.43. Since the two transistors are electrically identical, the voltage transfer curve and the resulting critical voltages are symmetrical. Also, $(V_{OH} - V_{OHU}) = 0.375\,\text{V}$, which is less than $|V_{TP}|$, and $(V_{OLU} - V_{OL}) = 0.375\,\text{V}$, which is less than V_{TN}. As long as the input voltage remains within the limits of the noise margins, no logic error will be transmitted through the digital system.

Test Your Understanding

***16.16** A CMOS inverter is biased at $V_{DD} = 10\,\text{V}$. The transistor parameters are: $V_{TN} = 2\,\text{V}$, $V_{TP} = -2\,\text{V}$, $K_N = 200\,\mu\text{A/V}^2$, and $K_P = 80\,\mu\text{A/V}^2$. (a) Sketch the voltage transfer curve. (b) Determine the critical voltages V_{IL} and V_{IH}, and the corresponding output voltages. (c) Calculate the noise margins NM_L and NM_H. (Ans. (b) $V_{IL} = 3.39\,\text{V}$, $V_{IH} = 4.86\,\text{V}$ (c) $\text{NM}_L = 2.59\,\text{V}$, $\text{NM}_H = 4.57\,\text{V}$)

***16.17** Repeat Exercise 16.16 for a CMOS inverter biased at $V_{DD} = 5\,\text{V}$ with transistor parameters of: $V_{TN} = +0.8\,\text{V}$, $V_{TP} = -2\,\text{V}$, and $K_N = K_P = 100\,\mu\text{A/V}^2$. (Ans. (b) $V_{IL} = 1.63\,\text{V}$, $V_{IH} = 2.18\,\text{V}$ (c) $\text{NM}_L = 1.35\,\text{V}$, $\text{NM}_H = 2.55\,\text{V}$)

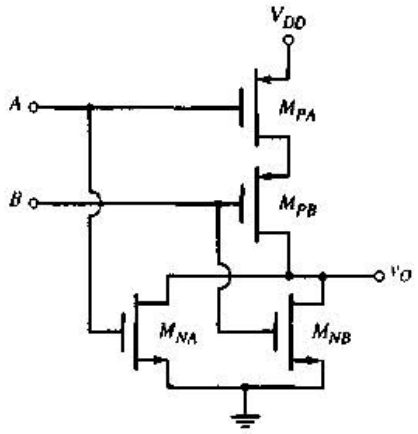
16.4 CMOS LOGIC CIRCUITS

Large-scale integrated CMOS circuits are used extensively in digital systems, including watches, calculators, and microprocessors. We will look at the basic CMOS NOR and NAND gates, and will then analyze more complex CMOS logic circuits. Since there is no clock signal applied to these logic circuits, they are referred to as **static CMOS logic** circuits.

16.4.1 Basic CMOS NOR and NAND Gates

In the basic or classical CMOS logic circuits, the gates of a PMOS and an NMOS are connected together, and additional PMOS and NMOS transistors are connected in series or parallel to form specific logic circuits. Figure 16.44(a) shows a two-input CMOS NOR gate. The NMOS transistors are in parallel and the PMOS transistors are in series.

If $A = B =$ logic 0, then both M_{NA} and M_{NB} are cut off, and the current in the circuit is zero. The source-to-gate voltage of M_{PA} is V_{DD} but the current is zero; therefore, v_{SD} of M_{PA} is zero. This means that the source-to-gate voltage



(a)

A	B	v_O
0	0	V_{DD}
V_{DD}	0	0
0	V_{DD}	0
V_{DD}	V_{DD}	0

(b)

Figure 16.44 (a) Two-input CMOS NOR logic circuit and (b) truth table

of M_{PB} is also V_{DD}. However, since the current is zero, then v_{SD} of M_{PB} is also zero. The output voltage is therefore $v_O = V_{DD}$ = logic 1.

If the input signals are A = logic 1 = V_{DD} and B = logic 0 = 0 V, then the source-to-gate voltage of M_{PA} is zero, and the current in the circuit is again zero. The gate-to-source voltage of M_{NA} is V_{DD} but the current is zero, so v_{DS} of M_{NA} is zero and $v_O = 0$ = logic 0. This result also holds for the other two possible input conditions, since at least one PMOS is cut off and at least one NMOS is in a conducting state. The NOR logic function is shown in the truth table of Figure 16.44(b).

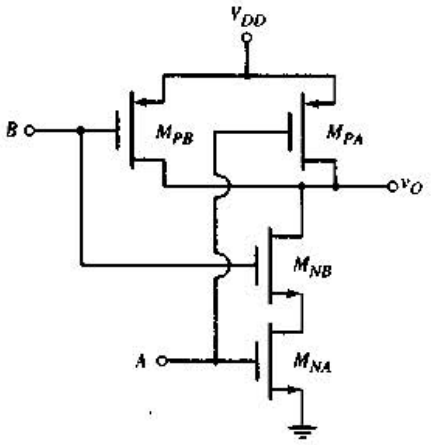
A two-input CMOS NAND logic gate is shown in Figure 16.45(a). In this case, the NMOS transistors are in series and the PMOS transistors are in parallel. If $A = B$ = logic 0, the two NMOS devices are cut off and the current in the circuit is zero. The source-to-gate voltage of each PMOS device is V_{DD}, which means that both PMOS transistors are in a conducting state. However, since the current is zero, v_{SD} for both M_{PA} and M_{PB} is zero and $v_O = V_{DD}$. This result applies if at least one input is a logic 0.

If the input signals are $A = B$ = logic 1 = V_{DD}, then both PMOS transistors are cut off, and the current in the circuit is zero. With A = logic 1, M_{NA} is in a conducting state; however, since the current is zero, then v_{DS} of M_{NA} is zero. This means that the gate-to-source voltage of M_{NB} is also V_{DD} and M_{NB} is also in a conducting state. However, since the current is zero, then v_{DS} of M_{NB} is zero, and v_O = logic 0 = 0 V. The NAND logic function is shown in the truth table in Figure 16.45(b).

In both the CMOS NOR and NAND logic gates, the current in the circuit is essentially zero when the inputs are in any quiescent state. Only very small reverse-bias pn junction currents exist. The quiescent power dissipation is therefore essentially zero. Again, this is the primary advantage of CMOS circuits.

To obtain symmetrical switching times or propagation delay times for the high-to-low and low-to-high output transitions, the effective conduction parameters of the composite PMOS and composite NMOS devices must be equal. For the CMOS NOR logic gate in Figure 16.44(a), we can write

$$K_{CN} = K_{CP} \tag{16.80}$$



(a)

A	B	v_O
0	0	V_{DD}
V_{DD}	0	V_{DD}
0	V_{DD}	V_{DD}
V_{DD}	V_{DD}	0

(b)

Figure 16.45 (a) Two-input CMOS NAND logic circuit and (b) truth table

where K_{CN} is the effective conduction parameter of the two parallel NMOS devices and K_{CP} is the effective conduction parameter of the two series PMOS transistors.

The effective channel *width* of the parallel NMOS devices is twice the individual width; similarly, the effective channel *length* of the series PMOS devices is twice the individual length. Equation (16.80) is then

$$\frac{k_n'}{2}\left(\frac{2W}{L}\right)_N = \frac{k_p'}{2}\left(\frac{W}{2L}\right)_P \tag{16.81}$$

since $k_n' \approx 2k_p'$, Equation (16.81) becomes

$$2\left(\frac{2W}{L}\right)_N = \left(\frac{W}{2L}\right)_P \tag{16.82(a)}$$

or

$$\left(\frac{W}{L}\right)_P = 8\left(\frac{W}{L}\right)_N \tag{16.82(b)}$$

This equation states that in the two-input CMOS NOR gate, the width-to-length ratio of the PMOS transistors must be approximately eight times that of the NMOS devices in order to provide the current required for symmetrical switching.

Figure 16.46 shows the voltage transfer and current characteristics of the two-input CMOS NOR logic gate from a PSpice analysis. Figure 16.46(a) shows the results when the two inputs are tied together; the circuit then behaves as a composite inverter. When the width-to-length ratios of the PMOS devices are eight times those of the NMOS devices, the transfer characteristics are symmetrical as expected.

Also shown are the voltage transfer and current characteristics if the p-channel width-to-length ratios are only twice as large as those of the n-channel devices. The resulting switching characteristics are not symmetrical and the maximum current is significantly reduced, implying that the switching times would be longer for this case.

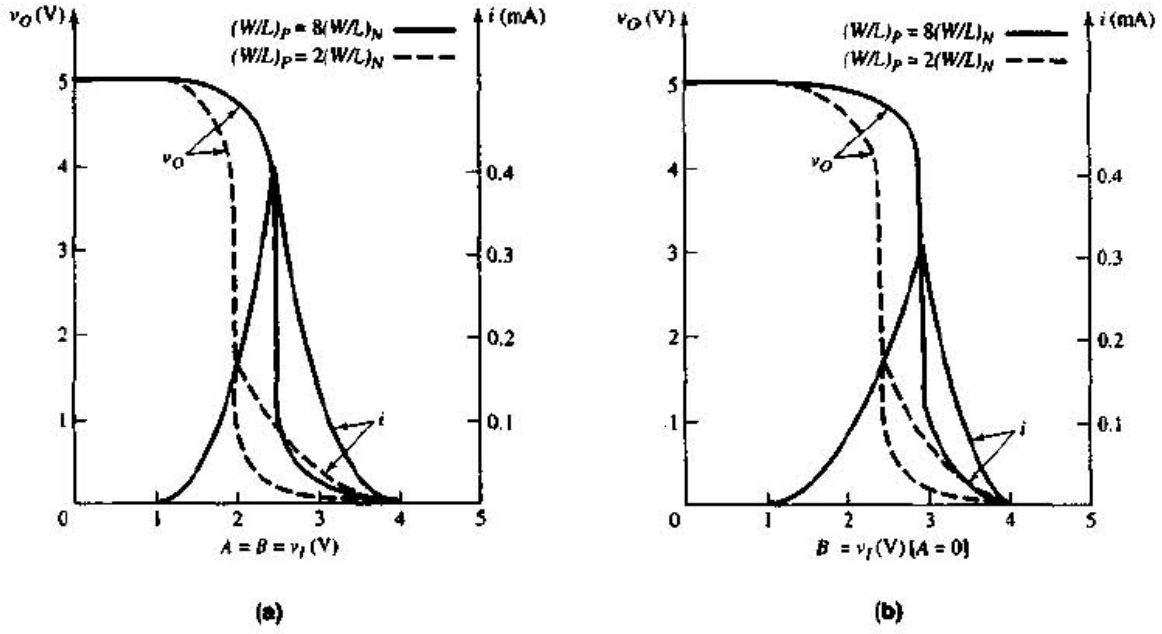


Figure 16.46 Voltage transfer characteristics, two-input CMOS NOR logic circuit for various width-to-length ratios: (a) $A = B = v_I$ and (b) $A = 0$, $B = v_I$

Figure 16.46(b) shows the transfer characteristics for input A held low and input B varied from zero to V_{DD}. This corresponds to only one input changing states while the other input remains constant. In this case, transistor M_{PA} is on and transistor M_{NA} is cut off. The resulting circuit is essentially equivalent to a simple CMOS inverter. When the width-to-length ratios of the PMOS devices are twice as large as those of the NMOS devices, the transfer characteristics are almost symmetrical, as we would expect for the CMOS inverter. When the PMOS width-to-length ratios are increased, the voltage transfer characteristic is no longer symmetrical. However, the maximum current is larger since the PMOS devices are larger, and this shortens the switching times. These results show that we obtain good transfer characteristics when $(W/L)_P = 8(W/L)_N$ for the two-input CMOS NOR circuit.

Test Your Understanding

[Note: In the following Exercises, let $k_n' = 35\,\mu A/V^2$ and $k_p' = 17.5\,\mu A/V^2$.]

16.18 Consider the two-input CMOS NOR gate in Figure 16.44(a). Assume $V_{TN} = 1\,V$, $V_{TP} = -1\,V$, and $V_{DD} = 5\,V$. (a) For A = logic 0 and $(W/L)_P = 8(W/L)_N$, plot the voltage transfer curve v_O versus B. Show all critical voltages. (b) If A = logic 0, determine $(W/L)_P$ and $(W/L)_N$ such that the peak current in the logic circuit is 50 μA. (Ans. (a) $V_{Il} = 2.76\,V$, $V_{OPl} = 3.76\,V$, $V_{ONl} = 1.76\,V$ (b) $(W/L)_N = 0.923$, $(W/L)_P = 8(W/L)_N = 7.38$)

D16.19 In the two-input CMOS NAND gate in Figure 16.45(a), determine the relationship between the (W/L) ratios of the n-channel and p-channel transistors such that the composite conduction parameters of the PMOS and NMOS devices are equal. (Ans. $(W/L)_N = 2(W/L)_P$)

D16.20 Design a three-input CMOS NOR logic gate such that the effective conduction parameters of the composite PMOS and NMOS transistors are equal. Determine $(W/L)_P/(W/L)_N$, where (W/L) is the width-to-length ratio of the individual PMOS and NMOS transistors. (Ans. $(W/L)_P = 18(W/L)_N$)

D16.21 Repeat Exercise 16.20 for a three-input CMOS NAND logic gate. (Ans. $(W/L)_N = 4.5(W/L)_P$)

16.4.2 Complex CMOS Logic Circuits

Just as with NMOS logic designs, we can form complex logic gates in CMOS, which avoids connecting large numbers of NOR, NAND, and inverter gates to implement the logic function. There are formal methods that can be used to implement the logic circuit. However, we can use the knowledge gained in the analysis and design of the NOR and NAND circuits.

Design Example 16.12 **Objective:** Design a CMOS logic circuit to implement a particular logic function.

Implement the logic function $Y = AB + C(D + E)$ in a CMOS design. The signals A, B, C, D, and E are available.

Design Approach: The general CMOS design is shown in Figure 16.47, in which the inputs are applied to both the PMOS and NMOS networks. We may start the design by considering the NMOS portion of the circuit. To implement a basic OR (NOR) function, the n-channel transistors are in parallel (Figure 16.44) and to implement a basic AND (NAND) function, the n-channel transistors are in series (Figure 16.45). We will consider whether the function or its complement is generated at the end of the design.

Solution: NMOS Design: In the overall function, we note the logic OR between the functions AB and $C(D + E)$, so that the NMOS devices used to implement AB will be in parallel with the NMOS devices used to implement $C(D + E)$. There is a logic AND between the inputs A and B, so that the NMOS devices with these inputs will be in

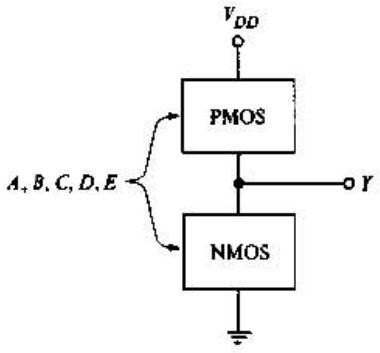


Figure 16.47 General CMOS design

series. Finally, the NMOS devices with the D and E inputs will be in parallel and this combination will be in series with the NMOS device with the C input. The NMOS implementation of the function is shown in Figure 16.48.

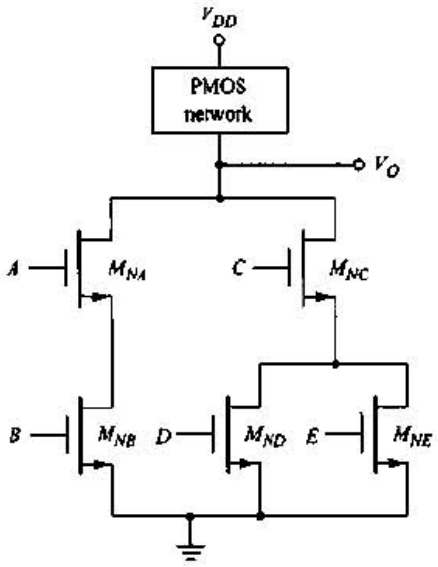


Figure 16.48 NMOS design for Example 16.12

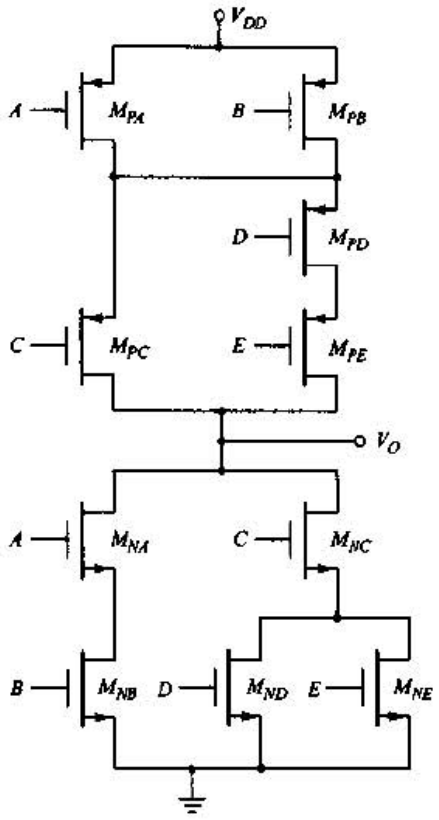


Figure 16.49 Complete CMOS design for Example 16.12

Solution: PMOS Design: The arrangement of the PMOS devices is complementary to that of the NMOS devices. PMOS devices that perform the basic OR function are in series and PMOS devices that perform the basic AND function are in parallel. We then see that the PMOS devices used to implement AB will be in series with the devices used to implement $C(D + E)$. The two PMOS devices with the A and B inputs will be in parallel. The two PMOS devices with the D and E inputs will be in series and in turn will be in parallel with the PMOS device with the C input. The completed circuit is shown in Figure 16.49.

Final Solution: By considering various inputs, we may note that the output signal of the circuit shown in Figure 16.49 is actually the complement of the desired signal. We may then simply add a CMOS inverter to the output to obtain the desired function.

Comment: As mentioned, there are formal ways in which to design circuits. However, in many cases, these circuits can be designed by using the knowledge and intuition gained from previous work. The width-to-length ratios of the various transistors can be determined as we have done in previous examples.

Test Your Understanding

D16.22 Design the width-to-length ratios of the transistors in the static CMOS logic circuit of Figure 16.49 so that the composite conduction parameters are the same. Assume the minimum W/L ratio of an n-channel device is 1 and assume $\mu_n = 2\mu_p$. (Ans. For M_A–M_E: 2, 2, 2, 1, 1. For P_A–P_E: 2, 2, 2, 4, 4.)

D16.23 Design a static CMOS logic circuit that implements the logic function $Y = (\overline{ABC + DE})$.

16.4.3 Fanout and Propagation Delay Time

Fanout

The term *fanout* refers to the number of load gates of similar design connected to the output of a driver gate. The maximum fanout is the maximum number of load gates that may be connected to the output. Since the CMOS logic gate will be driving other CMOS logic gates, the quiescent currrent required to drive the other CMOS gates is essentially zero. In terms of static characteristics, the maximum fanout is virtually limitless.

However, each additional load gate increases the load capacitance that must be charged and discharged as the driver gate changes state, and this places a practical limit on the maximum allowable number of load gates. Figure 16.50 shows a constant current charging a load capacitance. The voltage across the capacitance is

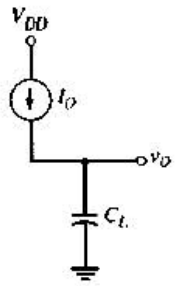


Figure 16.50 Constant-current source charging a load capacitor

$$v_O = \frac{1}{C_L}\int I_O dt = \frac{I_O t}{C_L} \tag{16.83}$$

The load capacitance C_L is proportional to the number N of load gates and to the input gate capacitance of each load. The current I_O is proportional to the conduction parameter of the driver transistor. The switching time is therefore

$$t \propto \frac{N(W \cdot L)_L}{\left(\frac{W}{L}\right)_D} \tag{16.84}$$

where the gate capacitance is directly proportional to the gate area of the load $(W \cdot L)_L$, and the conduction parameter of the driver transistor is proportional to the width-to-length ratio. Equation (16.84) can be rewritten as

$$t \propto N(L_L L_D)\left(\frac{W_L}{W_D}\right) \tag{16.85}$$

The propagation delay time, which is proportional to the switching time, increases as the fanout increases. The propagation delay time could be reduced by increasing the size of the driver transistor. However, in any given driver logic circuit and load logic circuit, the sizes of the devices are generally fixed. Consequently, the maximum fanout is limited by the maximum acceptable propagation delay time.

Propagation delay times are typically measured with a specified load capacitance. The average propagation delay time of a two-input CMOS NOR gate (such as an SN74HC36) is 25 ns, measured with a load capacitance of $C_L = 50\,\text{pF}$. Since the input capacitance is $C_I = 10\,\text{pF}$, a fanout of five would produce a 50 pF load capacitance. A fanout larger than five would increase the load capacitance, and would also increase the propagation delay time above the specified value.

Propagation Delay Time

Although the propagation delay time of the CMOS inverter can be determined by analytical techniques, it can also be determined by computer simulation. This is especially true when more complex CMOS logic circuits are considered. Using the appropriate transistor models in the simulation, the transient response can be produced. Obtaining an accurate transient response depends on using the correct transistor parameters. Some computer simulation problems in the end-of-chapter problems deal with propagation delay times. However, we will not go into detail here.

16.5 CLOCKED CMOS LOGIC CIRCUITS

The CMOS logic circuits considered in the previous section are called static circuits. One characteristic of a static CMOS logic circuit is that the output node always has a low-resistance path to either ground or V_{DD}. This implies that the output voltage is well defined and is never left floating.

Static CMOS logic circuits can be redesigned with an added clock signal while at the same time eliminating many of the PMOS devices. In general, the PMOS devices must be larger than NMOS devices. Eliminating as many PMOS devices as possible reduces the required chip area as well as the input capacitance. The low-power dissipation of the CMOS technology, however, is maintained.

Clocked CMOS circuits are dynamic circuits that generally precharge the output node to a particular level when the clock is at a logic 0. Consider the circuit in Figure 16.51. When the clock signal is low, or CLK = logic 0, M_{N1} is cut off and the current in the circuit is zero. Transistor M_{P1} is in a conducting state, but since the current is zero, then v_{O1} charges to V_{DD}. A high input to the CMOS inverter means that $v_O = 0$. During this phase of the clock signal, the gate of M_{P2} is precharged.

During the next phase, when the clock signal goes high, or CLK = logic 1, transistor M_{P1} cuts off and M_{N1} is biased in a conducting state. If input A = logic 0, then M_{NA} is cut off and there is no discharge path for voltage v_{O1}; therefore, v_{O1} remains charged at $v_{O1} = V_{DD}$. However, if CLK = logic 1 and A = logic 1, then both M_{N1} and M_{NA} are biased in a conducting state, providing a discharge path for voltage v_{O1}. As v_{O1} is pulled low, output signal v_O goes high.

The quiescent power dissipation in this circuit is essentially zero, as it was in the standard CMOS circuits. A small amount of power is required to precharge output v_{O1}, if it had been pulled low during the previous half clock cycle.

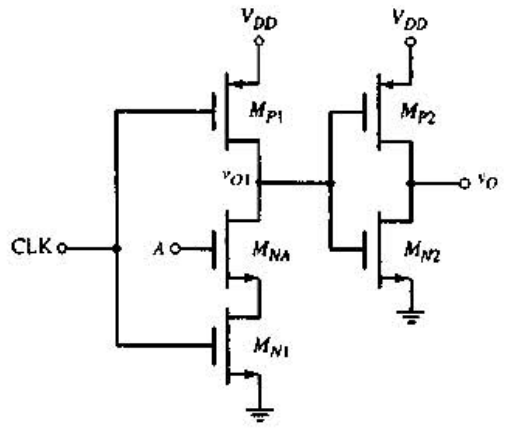


Figure 16.51 Simple clocked CMOS logic circuit

The single NMOS transistor M_{NA} in Figure 16.51 can be replaced by a more complex NMOS logic circuit. Consider the two circuits in Figure 16.52. When CLK = logic 0, then M_{N1} cuts off and M_{P1} is in its conducting state in both circuits; then, v_{O1} is charged to $v_{O1} = V_{DD}$ and $v_O = 0$. For the circuit in Figure 16.52(a), when CLK = logic 1, voltage v_{O1} is discharged to ground or pulled low only when $A = B$ = logic 1. In this case, v_O goes high. The circuit in Figure 16.52(a) performs the AND function. Similarly, the circuit in Figure 16.52(b) performs the OR function.

The advantage of the precharge technique is that it avoids the use of extensive pull-up networks: Only one PMOS and one NMOS transistor are required. This leads to an almost 50 percent savings in silicon area for larger circuits, and a reduction in capacitance resulting in higher speed. In addition, the static or quiescent power dissipation is essentially zero, so the circuit maintains the characteristics of CMOS circuits.

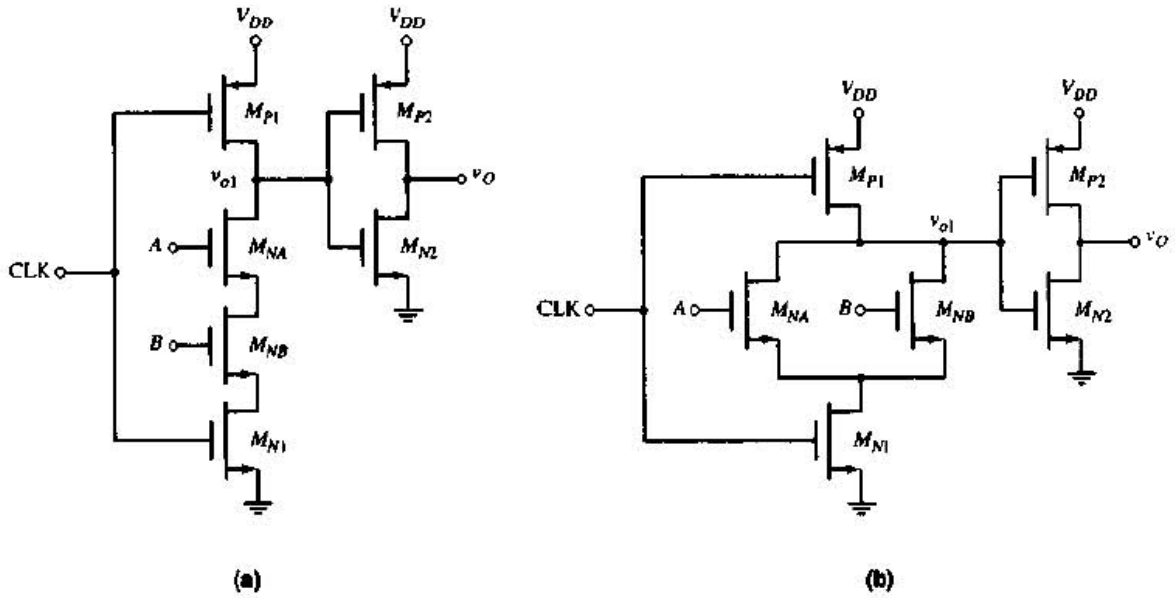


Figure 16.52 Clocked CMOS logic circuit: (a) AND function and (b) OR function

The AND and OR logic transistors M_{NA} and M_{NB} in Figures 16.52(a) and 16.52(b) can be replaced by a generalized logic network as indicated in Figure 16.53. The box marked f is an NMOS pull-down network that performs a particular logic function $f(X)$ of n variables, where $X = (x_1, x_2, \ldots, x_n)$. The NMOS circuit is a combination of series–parallel interconnections of n transistors. When the clock signal goes high, the CMOS inverter output is the logic function $f(X)$.

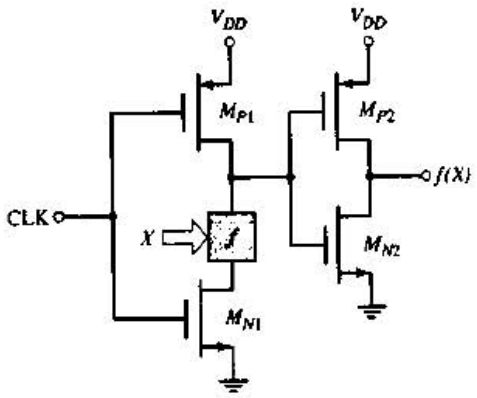


Figure 16.53 Generalized CMOS clocked logic circuit

The set of X inputs to the logic circuits f is derived from the outputs of other CMOS inverters and clocked logic circuits. This means that when CLK = logic 0, the outputs of all CMOS inverters are a logic 0 during the precharge cycle. As a result, all n variables $X = (x_1, x_2, \ldots, x_n)$ are a logic 0 during the precharge cycle. During this time, all NMOS transistors are cut off, which guarantees that output v_{O1} can be precharged to V_{DD}. There can then be only one possible transition at each node during the evaluation phase. The output of the CMOS buffer may change from a 0 to a 1.

An example of a cascaded domino CMOS circuit is shown in Figure 16.54. During the precharge cycle, in which CLK = logic 0, nodes 1 and 3 are charged high and nodes 2 and 4 are low. Also during this time, the inputs A, B, and C are all a logic 0. During the evaluation phase, in which CLK = logic 1, if $A = C$ = logic 1 and B = logic 0, then node 1 remains charged high, f_1 = logic 0, and node 3 discharges through M_{NC} causing f_2 to go high. However, if, during the evaluation phase, $A = B$ = logic 1 and C = logic 0, then node 1 is pulled low causing f_1 to go high, which in turn causes node 3 to go low and forces node 4 high. This chain of actions thus leads to the term **domino circuit**.

Test Your Understanding

D16.24 Design a clocked CMOS domino logic circuit, such as shown in Figure 16.53, to generate an output $f(X) = A \cdot B \cdot C + D \cdot E$.

D16.25 Sketch a clocked CMOS logic circuit that realizes the exclusive OR function.

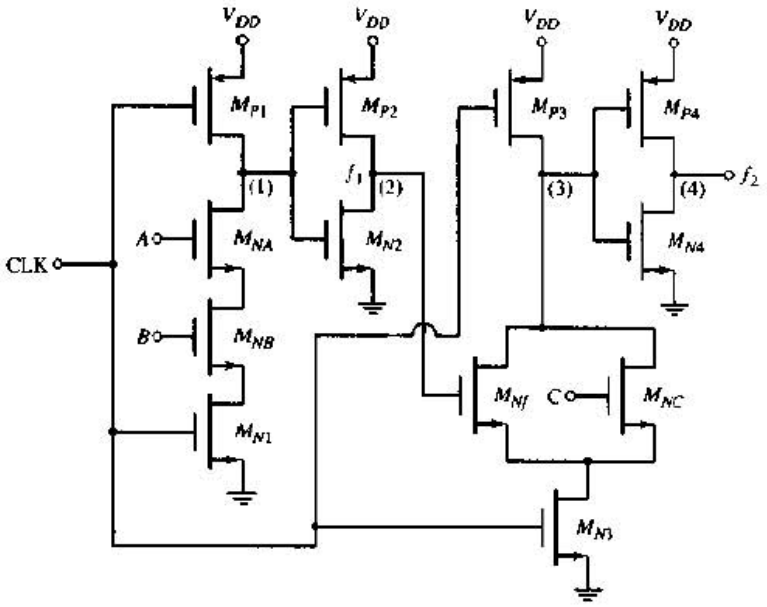


Figure 16.54 Cascaded clocked or domino CMOS logic circuit

16.6 TRANSMISSION GATES

Transistors can act as switches between driving circuits and load circuits. Transistors used to perform this function are called transmission gates. We will examine NMOS and CMOS transmission gates, which can also be configured to perform logic functions.

16.6.1 NMOS Transmission Gate

The NMOS enhancement-mode transistor in Figure 16.55(a) is a transmission gate connected to a load capacitance C_L, which could be the input gate capacitance of a MOS logic circuit. In this circuit, the transistor must be bilateral, which means it must be able to conduct current in either direction. This is a natural feature of MOSFETs. Terminals a and b are assumed to be equivalent, and the bias applied to the transistor determines which terminal acts as the drain and which terminal acts as the source. The substrate must be connected to the most negative potential in the circuit, which is usually ground.

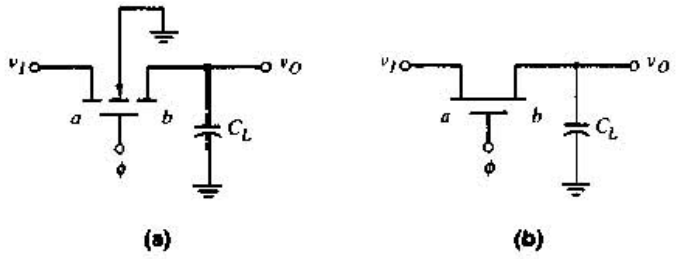


Figure 16.55 (a) NMOS transmission gate, showing substrate connection, and (b) simplified diagram

Figure 16.55(b) shows a simplified circuit symbol for the **NMOS transmission gate** that is used extensively.

We assume that the NMOS transmission gate is to operate over a voltage range of zero-to-V_{DD}. If the gate voltage ϕ is zero, then the n-channel transistor is cut off and the output is isolated from the input. The transistor is essentially an open switch.

If $\phi = V_{DD}$, $v_I = V_{DD}$, and v_O is initially zero, then terminal a acts as the drain since its bias is V_{DD}, and terminal b acts as the source since its bias is zero. Current enters the drain from the input, charging up the capacitor. The gate-to-source voltage is

$$v_{GS} = \phi - v_O = V_{DD} - v_O \tag{16.86}$$

As the capacitor charges and v_O increases, the gate-to-source voltage decreases. The capacitor stops charging when the current goes to zero. This occurs when the gate-to-source voltage v_{GS} becomes equal to the threshold voltage V_{TN}. The maximum output voltage occurs when $v_{GS} = V_{TN}$, therefore, from Equation (16.86), we have

$$v_{GS}(\text{min}) = V_{TN} = V_{DD} - v_O(\text{max}) \tag{16.87(a)}$$

or

$$v_O(\text{max}) = V_{DD} - V_{TN} \tag{16.87(b)}$$

where V_{TN} is the threshold voltage taking into account the body effect.

Equation (16.87(b)) demonstrates one disadvantage of an NMOS transmission gate. A logic 1 level degrades, or attenuates, as it passes through the transmission gate. However, this may not be a serious problem for many applications.

Figure 16.56 shows the quasi-static output voltage versus input voltage of the NMOS transmission gate. As seen in the figure, when $v_I = V_{DD}$, the output voltage is $v_O = V_{DD} - V_{TN}$ as we have discussed. For input voltages in the range $v_I < V_{DD} - V_{TN}$, the figure demonstrates that $v_O = v_I$. In this range of input voltages, the gate-to-source voltage is still greater than the threshold voltage. However, in steady-state, the current must be zero through the capacitor. In this case, the current becomes zero when the drain-to-source voltage is zero, or when $v_O = v_I$.

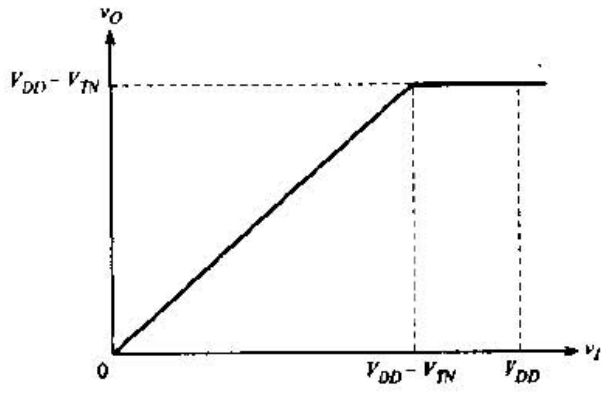


Figure 16.56 Output voltage versus input voltage characteristics of the NMOS transmission gate

Now consider the situation in which $\phi = V_{DD}$, $v_I = 0$, and $v_O = V_{DD} - V_{TN}$ initially. Terminal b then acts as the drain and terminal a acts as the source. The gate-to-source voltage is

$$v_{GS} = \phi - v_I = V_{DD} - 0 = V_{DD} \tag{16.88}$$

The value of v_{GS} is a constant, and the capacitor discharges as current enters the NMOS transistor drain. The capacitor stops discharging when the current goes to zero. Since v_{GS} is a constant at V_{DD}, the drain current goes to zero when the drain-to-source voltage is zero, which means that the capacitor completely discharges to zero. This implies that a logic 0 is transmitted unattenuated through the NMOS transmission gate.

Using an NMOS transmission gate in a MOS circuit may introduce a dynamic condition. Figure 16.57 shows a cross section of the NMOS transistor in the transmission gate configuration. If $v_I = \phi = V_{DD}$, then the load capacitor charges to $v_O = V_{DD} - V_{TN}$. When $\phi = 0$, the NMOS device turns off and the input and output become isolated.

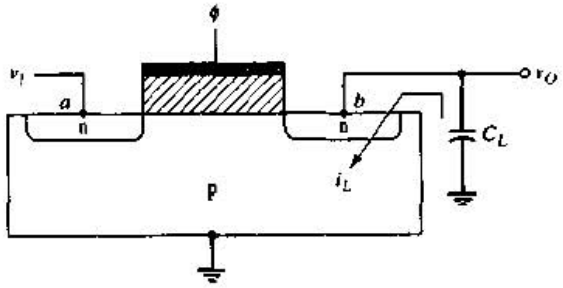


Figure 16.57 NMOS transmission gate with cross section of NMOS transistor

The capacitor voltage reverse biases the pn junction between terminal b and ground. A reverse-biased pn junction leakage current begins to discharge the capacitor, and the circuit does not remain in a static condition. This circuit is now dynamic in that the high output does not remain constant with time.

Example 16.13 **Objective:** Estimate the rate at which the output voltage v_O in Figure 16.57 decreases with time.

Assume the capacitor is initially charged to $v_O = 4\,\text{V}$. Let $C_L = 1\,\text{pF}$ and assume the reverse-biased pn junction leakage current is a constant at $i_L = 1\,\text{nA}$.

Solution: The voltage across the capacitor can be written as

$$v_O = -\frac{1}{C_L}\int i_L dt$$

where the minus sign indicates that the current is leaving the positive terminal of the capacitor. Since i_L is a constant, we have

$$v_O = -\frac{i_L}{C_L}t + K_1$$

where $K_1 = v_O(t = 0) = 4\,\text{V}$ is the initial condition. Therefore,

$$v_O = 4 - \frac{i_L}{C_L}t$$

The rate at which the output voltage decreases is

$$\frac{dv_O}{dt} = -\frac{i_L}{C_L} = -\frac{10^{-9}}{10^{-12}} = -1000\,\text{V/s} \Rightarrow -1\,\text{V/ms}$$

Therefore, in this example, the capacitor would completely discharge in 4 ms.

Comment: Even though the NMOS transmission gate may introduce a dynamic condition into a circuit, this gate is still useful in clocked logic circuits in which a clock signal is periodically applied to the NMOS transistor gate. If, for example, the clock frequency is 25 kHz, the clock pulse period is 40 μs, which means that the output voltage would decay by no more than 1 percent.

Example 16.14 **Objective:** Determine the output of an NMOS inverter driven by a series of NMOS transmission gates.

Consider the circuit shown in Figure 16.58. The NMOS inverter is driven by three NMOS transmission gates in series. Assume the threshold voltages of the n-channel transmission gate transistors and the driver transistor are $V_{TN} = +0.8\,\text{V}$, and the threshold voltage of the load transistor is $V_{TNL} = -1.5\,\text{V}$. Let $K_D/K_L = 3$ for the inverter. Determine v_O for $v_I = 0$ and $v_I = 5\,\text{V}$.

Solution: The three NMOS transmission gates in series act as an AND/NAND function. If $v_I = 0$ and $A = B = C =$ logic $1 = 5\,\text{V}$, the gate capacitance to driver M_D becomes completely discharged, which means that $v_{O1} = v_{O2} = v_{O3} = 0$. Driver M_D is cut off and $v_O = 5\,\text{V}$.

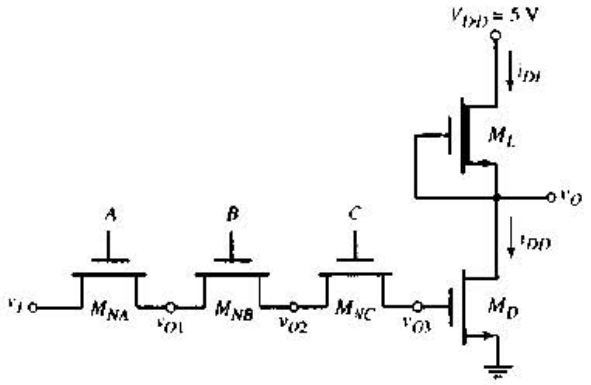


Figure 16.58 NMOS inverter driven by three NMOS transmission gates in series

If $v_I = 5\,\text{V}$ and $A = B = C =$ logic $1 = 5\,\text{V}$, the three transmission gates are biased in their conducting state, and the gate capacitance of M_D becomes charged. For transistor M_{NA}, the current becomes zero when the gate-to-source voltage is equal to the threshold voltage, or, from Equation (16.87(b)),

$$v_{O1} = V_{DD} - V_{TN} = 5 - 0.8 = 4.2\,\text{V}$$

Transistors M_{NB} and M_{NC} also cut off when the gate-to-source voltages are equal to the threshold voltage; therefore,

$$v_{O2} = v_{O3} = V_{DD} - V_{TN} = 5 - 0.8 = 4.2\,\text{V}$$

This result shows that the drain-to-source voltages of M_{NB} and M_{NC} are also zero. A threshold voltage drop is lost in the first transmission gate, but additional threshold voltage drops are not lost in subsequent NMOS transmission gates in series.

For a voltage of $v_{O3} = 4.2\,\text{V}$ applied to the gate of M_D, the driver is biased in the nonsaturation region and the load is biased in the saturation region. From $i_{DD} = i_{DL}$, we have

$$K_D\left[2(v_{O3} - V_{TN})v_O - v_O^2\right] = K_L[-V_{TNL}]^2$$

The output voltage is found to be

$$v_O = 0.112\,\text{V}$$

If any one of the transmission gate voltages, A or B or C, switches to a logic 0, then v_{O3} will begin to discharge through a reverse-biased pn junction in the transmission gates, which means that v_O will increase with time.

Comment: In this example, the inverter is again in a dynamic condition; that is, when any transmission gate is cut off, the output voltage changes with time. However, this type of circuit can be used in clocked digital systems.

Test Your Understanding

16.26 The threshold voltage of the NMOS transmission gate transistor in Figure 16.55(a) is $V_{TN} = 1\,\text{V}$. Determine the quiescent output voltage v_O for: (a) $v_I = \phi = 5\,\text{V}$; (b) $v_I = 3\,\text{V}$, $\phi = 5\,\text{V}$; (c) $v_I = 4.2\,\text{V}$, $\phi = 5\,\text{V}$; and (d) $v_I = 5\,\text{V}$, $\phi = 3\,\text{V}$. (Ans. (a) $v_O = 4\,\text{V}$ (b) $v_O = 3\,\text{V}$ (c) $v_O = 4\,\text{V}$ (d) $v_O = 2\,\text{V}$)

D16.27 Consider the NMOS inverter with enhancement load driven by an NMOS transmission gate in Figure 16.59. The threshold voltage of each n-channel transistor is $V_{TN} = 2\,\text{V}$. Neglect the body effect. Design K_D/K_L such that $v_O = 0.5\,\text{V}$ when: (a) $v_I = 8\,\text{V}$, $\phi = 10\,\text{V}$, and (b) $v_I = \phi = 8\,\text{V}$. (Ans. (a) $K_D/K_L = 9.78$ (b) $K_D/K_L = 15$)

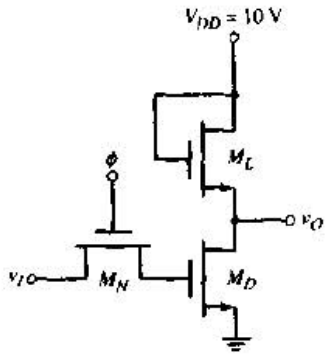


Figure 16.59 Figure for Exercise 16.27

16.6.2 NMOS Pass Networks

As integrated circuit technology advances, one emphasis is on increased circuit density. The maximum number of circuit functions per unit area is determined either by power dissipation density or by the area occupied by transistors and related devices.

One form of NMOS circuit logic that minimizes power dissipation and maximizes device density is called **pass transistor logic**. Pass transistor circuits use minimum-sized transistors, providing high density and high operating speed. The average power dissipation is due only to the switching power consumed by the driver circuits in charging and discharging the pass transistor control gates and driving the pass network inputs.

In this section, we present a few examples of NMOS pass transistor logic circuits. Consider the circuit in Figure 16.60. To determine the output response, we examine the conditions listed in Table 16.1 for the possible states of the input signals A and B. We assume that a logic 1 level is V_{DD} volts. In states 1 and 2, transmission gate M_{N2} is biased in its conducting state. For state 1, $\bar{A}$ = logic 1 is transmitted to the output so f = logic 1′, where the logic 1′ level is $(V_{DD} - V_{TN})$. The logic 1 level is attenuated by one threshold voltage drop. For state 2, A = logic 0 is transmitted unattenuated to the output. In states 3 and 4, transmission gate M_{N1} is biased in its conducting state. The A = logic 0 for state 3 is transmitted unattenuated to the output, and A = logic 1 for state 4 is attenuated during transmission; therefore, f = logic 1′. The output is thus the exclusive-NOR function.

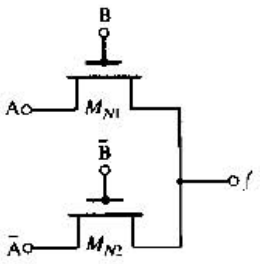


Figure 16.60 Simple NMOS pass logic network

Table 16.1 Input and ouput states for the circuit in Figure 16.60

State	A	B	$\bar{A}$	$\bar{B}$	M_{N1}	M_{N2}	f
1	0	0	1	1	off	on	1′
2	1	0	0	1	off	on	0
3	0	1	1	0	on	off	0
4	1	1	0	0	on	off	1′

Another example of an NMOS pass transistor logic circuit is shown in Figure 16.61. The ouput response as a function of the input gate controls A and B is shown in Table 16.2. This circuit is a multiplexer: that is, for a specific set of gate controls, the input signals P_i are individually passed to the output. By using both normal and inverted forms of A and B, four inputs can be controlled with just two variables.

A potential problem of NMOS pass transistor logic is that the output may be left floating in a high impedance state and charged high. Consider the circuit shown in Figure 16.62. If, for example, $\bar{B} = C$ = logic 0 and A = logic 1, then f = logic 1′, which is the logic 1 level attenuated by V_{TN}. When A is switched to logic 0, the output should be low, but there may not be a discharge path to ground, and the output may retain the logic 1′ stored at the output capacitance.

The NMOS pass network must be designed to avoid a high impedance output by passing a logic 0 whenever a 0 is required at the output. A logic network that performs the logic function $f = A + \bar{B} \cdot C$, as indicated in Figure 16.62, is shown in Figure 16.63. The complementary function $\bar{f} = \bar{A} \cdot (B + \bar{C})$ attached at the output node drives the output to a logic 0 whenever $f = 0$.

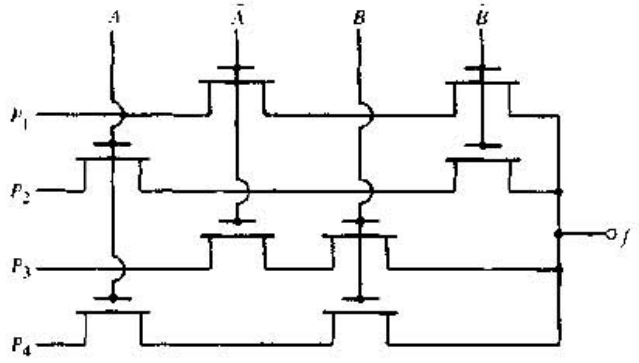


Figure 16.61 NMOS pass logic network example

Table 16.2 Input and output states for the circuit in Figure 16.61

State	A	B	$\bar{A}$	$\bar{B}$	f
1	0	0	1	1	P_1
2	1	0	0	1	P_2
3	0	1	1	0	P_3
4	1	1	0	0	P_4

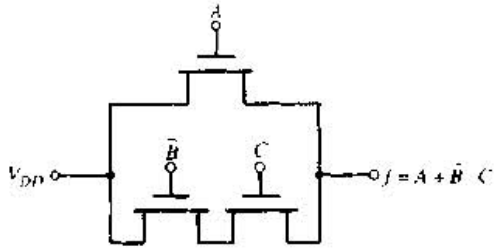


Figure 16.62 NMOS pass logic network with a potential problem

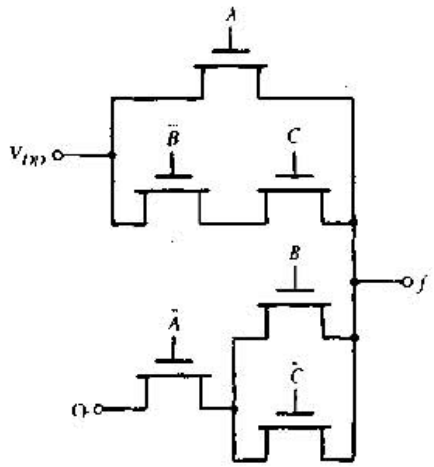


Figure 16.63 NMOS pass logic network with complementary function in parallel

Test Your Understanding

D16.28 Design an NMOS pass network to perform the exclusive-OR function.

16.6.3 CMOS Transmission Gate

A **CMOS transmission gate** is shown in Figure 16.64(a). The parallel combination of NMOS and PMOS transistors, with complementary gate signals, allows the input signal to be transmitted to the output without the threshold voltage attenuation. Both transistors must be bilateral; therefore, the NMOS substrate is connected to the most negative potential in the circuit and the PMOS substrate is connected to the most positive potential (usually, ground and V_{DD}, respectively). Figure 16.64(b) shows a frequently used simplified circuit symbol for the CMOS transmission gate.

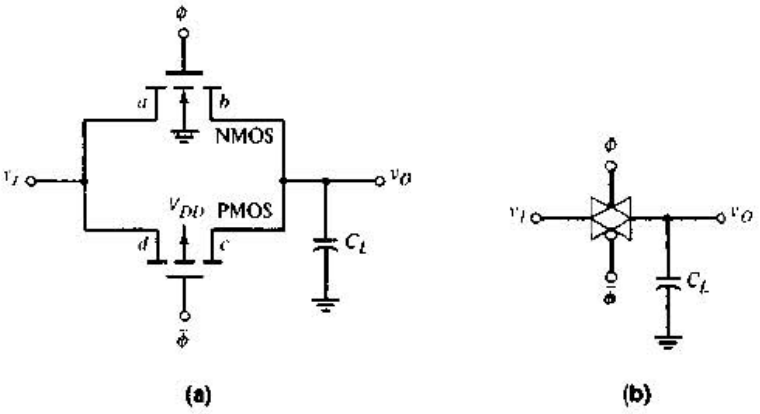


Figure 16.64 (a) CMOS transmission gate and (b) simplified circuit symbol

We again assume that the transmission gate is to operate over a voltage range of zero-to-V_{DD}. If the control voltages are $\phi = 0$ and $\bar{\phi} = V_{DD}$, then both the NMOS and PMOS transistors are cut off and the output is isolated from the input. In this state, the circuit is essentially an open switch.

If $\phi = V_{DD}$, $\bar{\phi} = 0$, $v_I = V_{DD}$, and v_O is initially zero, then for the NMOS device, terminal a acts as the drain and terminal b acts as the source, whereas for the PMOS device, terminal c acts as the drain and terminal d acts as the source. Current enters the NMOS drain and the PMOS source, as shown in Figure 16.65(a), to charge the load capacitor. The NMOS gate-to-source voltage is

$$v_{GSN} = \phi - v_O = V_{DD} - v_O \tag{16.89(a)}$$

and the PMOS source-to-gate voltage is

$$v_{SGP} = v_I - \bar{\phi} = V_{DD} - 0 = V_{DD} \tag{16.89(b)}$$

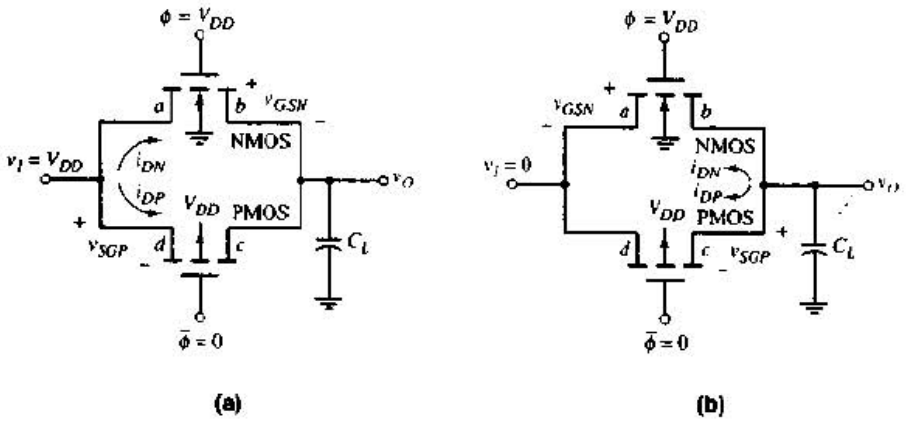


Figure 16.65 Currents and gate–source voltages in CMOS transmission gate for: (a) input high condition and (b) input low condition

As with the NMOS transmission gate, when $v_O = V_{DD} - V_{TN}$, the NMOS transistor cuts off and $i_{DN} = 0$ since $v_{GSN} = V_{TN}$. However, since the source-to-gate voltage of the PMOS device is a constant at $v_{SGP} = V_{DD}$, the PMOS transistor continues to conduct. The drain current i_{DP} goes to zero when the PMOS source-to-drain voltage goes to zero, or $v_{SDP} = 0$. This means that the load capacitor C_L continues to charge through the PMOS device until the output and input voltages are equal, or in this case, $v_O = v_I = 5$ V.

Consider what happens if $\phi = V_{DD}$, $\bar{\phi} = 0$, $v_I = 0$, and $v_O = V_{DD}$ initially. For the NMOS device, terminal a acts as the source and terminal b acts as the drain, whereas for the PMOS device, terminal c acts as the source and terminal d acts as the drain. Current enters the NMOS drain and the PMOS source, as shown in Figure 16.65(b), to discharge the capacitor. The NMOS gate-to-source voltage is

$$v_{GSN} = \phi - v_I = V_{DD} - 0 = V_{DD} \tag{16.90(a)}$$

and the PMOS source-to-gate voltage is

$$v_{SGP} = v_O - \bar{\phi} = v_O - 0 = v_O \tag{16.90(b)}$$

When $v_{SGP} = v_O = |V_{TP}|$, the PMOS device cuts off and i_{DP} goes to zero. However, since $v_{GSN} = V_{DD}$, the NMOS transistor continues conducting and capacitor C_L completely discharges to zero.

Using a CMOS transmission gate in a MOS circuit may introduce a dynamic condition. Figure 16.66 shows the CMOS transmission gate with simplified cross sections of the NMOS and PMOS transistors. If $\phi = 0$ and $\bar{\phi} = V_{DD}$, then the input and output are isolated. If $v_O = V_{DD}$, then the NMOS substrate-to-terminal b pn junction is reverse biased and capacitance C_L can discharge, as it did in the NMOS transmission gate. If, however, $v_O = 0$, then the PMOS terminal c-to-substrate pn junction is reverse biased and capacitance C_L can charge to a positive voltage. This circuit is therefore dynamic in that the output high or low conditions do not remain constant with time.

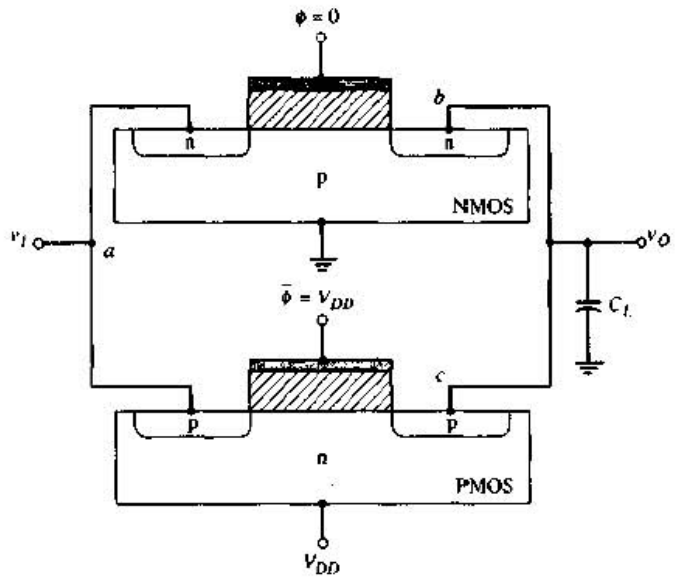


Figure 16.66 CMOS transmission gate showing cross sections of NMOS and PMOS transistors

Test Your Understanding

16.29 Consider the CMOS transmission gate in Figure 16.64(a). Assume transistor parameters of $V_{TN} = +0.8\,\text{V}$ and $V_{TP} = -1.2\,\text{V}$. When $\phi = 5\,\text{V}$, input v_I varies with time as $v_I = 0.5t\,\text{V}$ for $0 \le t \le 10\,\text{s}$. Let $v_O(t = 0) = 0$ and assume $C_L = 1\,\text{pF}$. Determine the range of times that the NMOS and PMOS devices are conducting or cut off. (Ans. NMOS conducting, $0 \le t < 8.4\,\text{s}$; PMOS conducting, $2.4 < t \le 10\,\text{s}$)

16.6.4 CMOS Pass Networks

CMOS transmission gates may also be used in pass network logic design. CMOS pass networks use NMOS transistors to pass 0's, PMOS transistors to pass 1's, and CMOS transmission gates to pass a variable to the output. An example is shown in Figure 16.67. One PMOS transistor is used to transmit a logic 1, while transmission gates are used to transmit a variable that may be either a logic 1 or a logic 0. We can show that for any combination of signals, a logic 1 or logic 0 is definitely passed to the output.

16.7 SEQUENTIAL LOGIC CIRCUITS

In the logic circuits that we have considered in the previous sections, such as NOR and NAND logic gates, the output is determined only by the instantaneous values of the input signals. These circuits are therefore classified as combinational logic circuits.

Another class of circuits is called **sequential logic circuits**. The output depends not only on the inputs, but also on the previous history of its inputs.

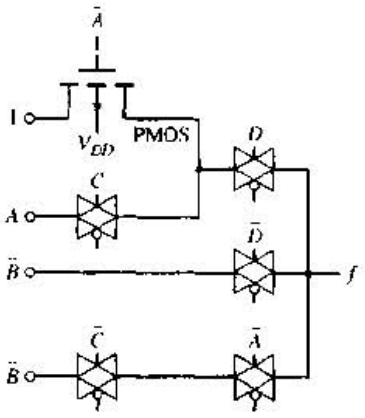


Figure 16.67 CMOS pass logic network

This feature gives sequential circuits the property of memory. Shift registers and flip-flops are typical examples of such circuits. We will also briefly consider a full-adder circuit. The characteristic of these circuits is that they store information for a short time until the information is transferred to another part of the system.

In this section, we introduce a basic shift register and the basic concept of a flip-flop. These circuits can become very complex and are usually described with logic diagrams. We will also introduce a CMOS full adder circuit in terms of its logic diagram and then provide the transistor implementation of this logic function. Additional information can be found in more advanced texts.

16.7.1 Dynamic Shift Registers

A **shift register** can be formed from transmission gates and inverters. Figure 16.68 shows a combination of NMOS transmission gates and NMOS depletion-load inverters. The clock signals applied to the gates of the NMOS transmission gates must be complementary, nonoverlapping pulses. The effective capacitances at the gates of M_{D1} and M_{D2} are indicated by the dotted connections to C_{L1} and C_{L2}.

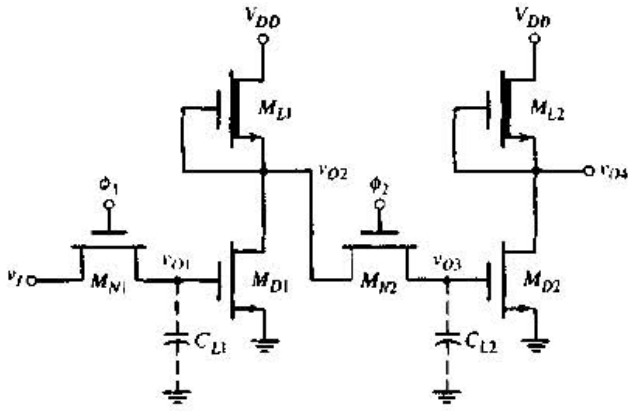


Figure 16.68 Dynamic shift register with NMOS inverters and transmission gates

If, for example, C_{L1} is initially uncharged when $v_{O1} = 0$ and if $v_I = V_{DD}$ when $\phi_1 = V_{DD}$, then a logic $1' = V_{DD} - V_{TN}$ voltage should exist at v_{O1} at the end of clock pulse ϕ_1. The capacitance of C_{L1} charges through M_{N1} and the driving circuit of v_I. The effective RC time constant must be sufficiently small to achieve this charging effect. As v_{O1} goes high, v_{O2} goes low, but the low is not transmitted through M_{N2} as long as ϕ_2 remains low.

Figure 16.69 is used to determine the operation of this circuit and the voltages at various times. For simplicity, we assume that $V_{DD} = 5\,\text{V}$ and $V_{TN} = 1\,\text{V}$ for the NMOS drivers and transmission gate transistors.

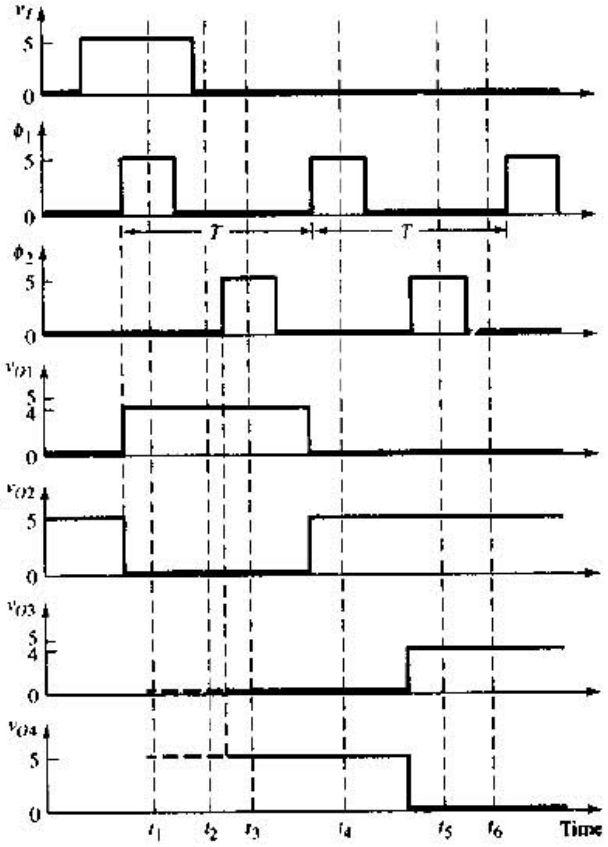


Figure 16.69 NMOS shift register voltages at various times

At $t = t_1$, $v_I = \phi_1 = 5\,\text{V}$, v_{O1} charges to $v_{DD} - V_{TN} = 4\,\text{V}$, and v_{O2} goes low. At this time, M_{N2} is still cut off, which means that the values of v_{O3} and v_{O4} depend on the previous history. At $t = t_2$, ϕ_1 is zero, M_{N1} is cut off, but v_{O1} remains charged. At $t = t_3$, ϕ_2 is high, and the logic 0 at v_{O2} is transmitted to v_{O3}, which forces v_{O4} to 5 V. The input signal $v_I = 5\,\text{V}$ at $t = t_1$ has thus been transmitted to the output; therefore, $v_{O4} = v_I = 5\,\text{V}$ at $t = t_3$. The input signal is transmitted, or *shifted*, from the input to the output during one clock cycle, making this circuit one stage of a shift register.

At $t = t_4$, $v_I = 0$, and $\phi_1 = 5\,\text{V}$, so that $v_{O1} = 0$ and $v_{O2} = 5\,\text{V}$. Since $\phi_2 = 0$, M_{N2} is cut off, and v_{O2} and v_{O3} are isolated. At $t = t_5$, $\phi_2 = 5\,\text{V}$, so that v_{O3} charges to $V_{DD} - V_{TN} = 4\,\text{V}$, and v_{O4} goes low (logic 0). At $t = t_6$,

both NMOS transmission gates are cut off, and the two inverters remain in their previous states. It is important that ϕ_1 and ϕ_2 do not overlap, or the signal would propagate through the whole chain at once and we would no longer have a shift register.

In the dynamic condition of NMOS transmission gates, the high output voltage across the output capacitance does not remain constant with time; it discharges through the transmission gate transistor. This same effect applies to the shift register in Figure 16.68. For example, from Figure 16.69, at $t = t_2$, $v_{O1} = 4\,\text{V}$, $\phi_1 = 0$, and M_{N1} is cut off. Voltage v_{O1} will start to decay and v_{O2} will begin to increase. To prevent logic errors from being introduced into the system, the clock signal period T must be small compared to the effective RC discharge time constant. The circuit in Figure 16.68 is therefore called a **dynamic shift register**.

A dynamic shift register formed in a CMOS technology is shown in Figure 16.70. Operation of this circuit is very similar to that of the dynamic NMOS shift register, except for the voltage levels. For example, when $v_I = \phi_1 = V_{DD}$, then $v_{O1} = V_{DD}$ and $v_{O2} = 0$. When ϕ_2 goes high, then v_{O3} goes to zero, $v_{O4} = V_{DD}$, and the input signal is shifted to the output during one clock period.

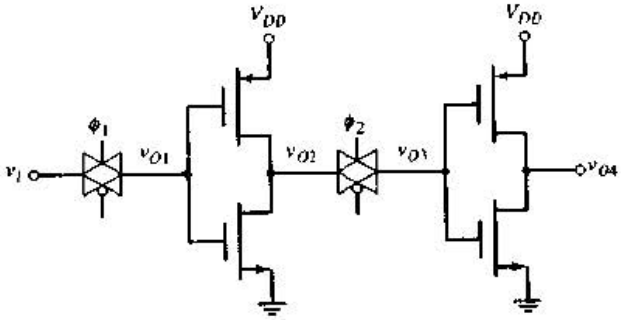


Figure 16.70 CMOS dynamic shift register

16.7.2 R–S Flip-Flop

Flip-flops are bistable circuits usually formed by cross-coupling two NOR gates. Figure 16.71 shows an R–S flip-flop using NMOS NOR logic gates with depletion loads. As shown, M_1, M_2, and M_3 form one NOR gate, and M_4, M_5, and M_6 form the second. The outputs of the two NOR circuits are connected back to the inputs of the opposite NOR gates.

If we assume that S = logic 1 and R = logic 0, then M_1 is biased in its conducting state and output $\overline{Q}$ is forced low. The inputs to both M_4 and M_5 are low, so output Q goes high to a logic 1 = V_{DD}. Transistor M_2 is then also biased in a conducting state. The two outputs Q and $\overline{Q}$ are complementary and, by definition, the flip-flop is in the set state when Q = logic 1 and $\overline{Q}$ = logic 0.

If S returns to logic 0, then M_1 turns off, but M_2 remains turned on so $\overline{Q}$ remains low and Q remains high. Therefore, when S goes low, nothing in the circuit can force a change and the flip-flop stores this particular logic state.

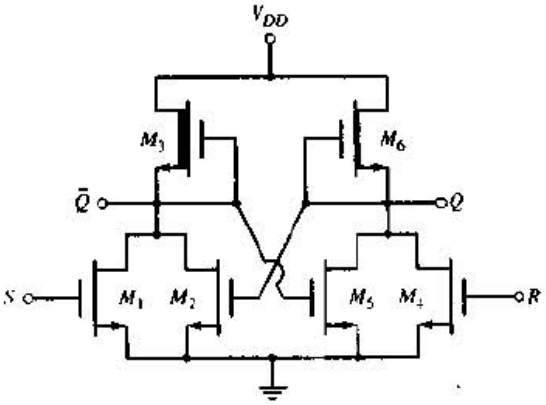


Figure 16.71 NMOS R–S flip-flop

When R = logic 1 and S = logic 0, then M_4 turns on so output Q goes low. With $S = Q$ = logic 0, then both M_1 and M_2 are cut off and $\overline{Q}$ goes high. Transistor M_5 turns on, keeping Q low when R goes low. The flip-flop is now in the reset state.

If both S and R inputs were to go high, then both outputs Q and $\overline{Q}$ would go low. However, this would mean that the outputs would not be complementary. Therefore, a logic 1 at both S and R is considered to be a forbidden or nonallowed condition. If both inputs go high and then return to logic 0, the state of the flip-flop is determined by whichever input goes low last. If both inputs go low simultaneously, then the outputs will flip into one state or the other, as determined by slight imbalances in transistor characteristics.

Figure 16.72 shows an R–S flip-flop using CMOS NOR logic gates. The outputs of the two NOR gates are connected back to the inputs of the opposite NOR gates to form the flip-flop.

If S = logic 1 and R = logic 0, then M_{N1} is turned on, M_{P1} is cut off, and $\overline{Q}$ goes low. With $\overline{Q} = R$ = logic 0, then both M_{N3} and M_{N4} are cut off, both

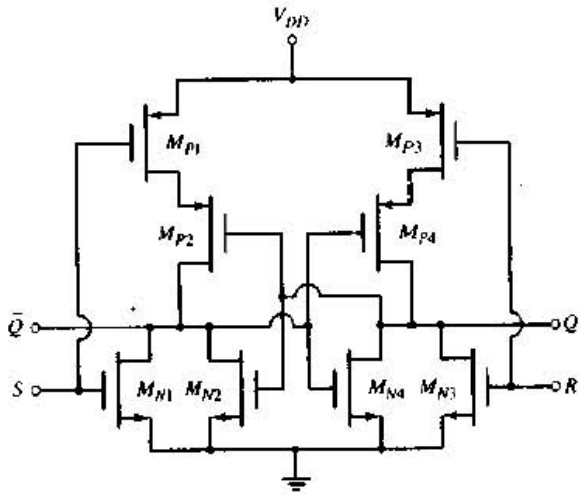


Figure 16.72 CMOS R–S flip-flop

M_{P3} and M_{P4} are biased in a conducting state so that the output Q goes high. With Q = logic 1, M_{N2} is biased on, M_{P2} is biased off, and the flip-flop is in a set condition. When S goes low, M_{N1} turns off, but M_{N2} remains conducting, so the state of the flip-flop does not change.

When S = logic 0 and R = logic 1, then output Q is forced low, output $\overline{Q}$ goes high, and the flip-flop is in a reset condition. Again, a logic 1 at both S and R is considered to be a forbidden or a nonallowed condition, since the resulting outputs are not complementary.

16.7.3 D Flip-Flop

A **D-type flip-flop** is used to provide a delay. The logic bit on the D input is transferred to the output at the next clock pulse. This flip-flop is used in counters and shift registers. The basic circuit is similar to the CMOS dynamic shift register in Figure 16.70, except that additional circuitry makes the D flip-flop a static circuit.

Consider the circuit in Figure 16.73. The CMOS inverter composed of M_{N2} and M_{P2} is driven by a CMOS transmission gate composed of M_{N1} and M_{P1}. A second CMOS inverter, M_{N3} and M_{P3}, is connected in a feedback configuration. If v_I = high, then v_{O1} goes high when the transmission gate is conducting, and output v_O, which is the input to the feedback inverter, goes low.

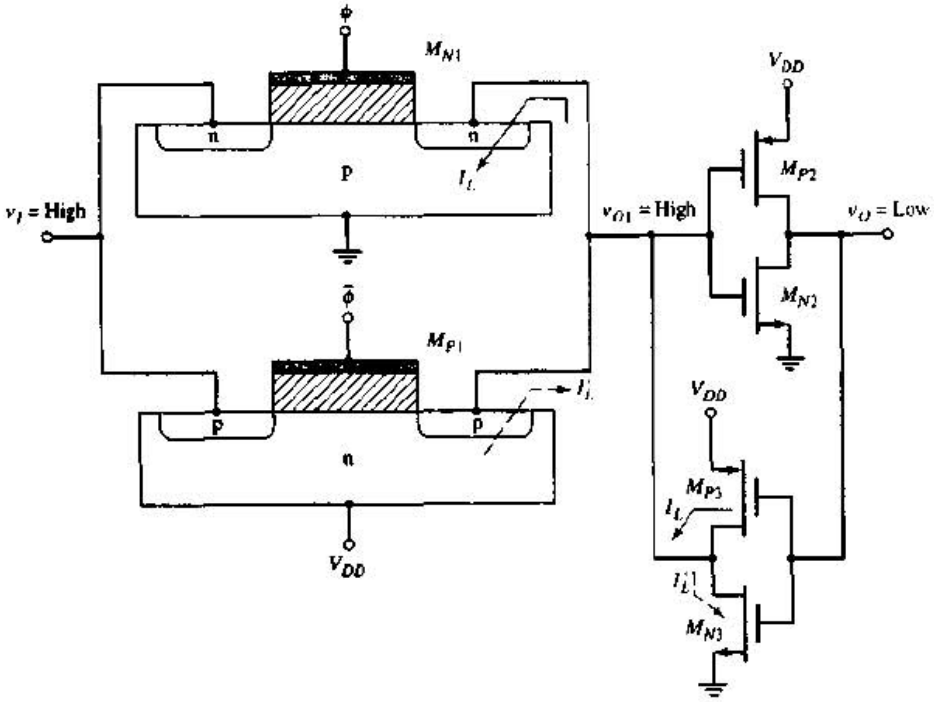


Figure 16.73 CMOS D-type flip-flop

When the CMOS transmission gate turns off, the pn junction in the M_{N1} transmission gate transistor is reverse biased. In this case, however, voltage v_{O1} is not simply across the gate capacitance of inverter M_{N2}–M_{P2}. Transistor M_{P3}

is biased in a conducting state, so the reverse-biased pn junction leakage current I_L is supplied through M_{P3}, as indicated in Figure 16.73. Since this leakage current is small, the source-to-drain voltage of M_{P3} will be small, and v_{O1} will remain biased at essentially V_{DD}. The circuit will therefore remain in this static condition.

Similarly, when v_{O1} is low and v_O is high, the pn junction in the M_{P1} transmission gate transistor is reverse biased and transistor M_{N3} is biased on. Transistor M_{N3} sinks the pn junction leakage current I_L', and the circuit remains in this static condition until changed by a new input signal through the transmission gate.

The circuit shown in Figure 16.74 is a master–slave configuration of a D flip-flop. When clock pulse ϕ is high, transmission gate TG1 is conducting, and data D goes through the first inverter, which means that $Q' = \overline{D}$. Transmission gate TG2 is off, so data stops at Q'. When clock pulse ϕ goes low, then TG3 turns on, and the master portion of the flip-flip is in a static configuration. Also when ϕ goes low, TG2 turns on, the data are transmitted through the slave portion of the flip-flop, and the output is $Q = \overline{Q'} = D$. The data present when ϕ is high are transferred to the output of the flip-flop during the negative transition of the clock pulse. The various signals in the D flip-flop are shown in Figure 16.75.

Additional circuitry can be added to the D flip-flop in Figure 16.74 to provide a set and reset capability.

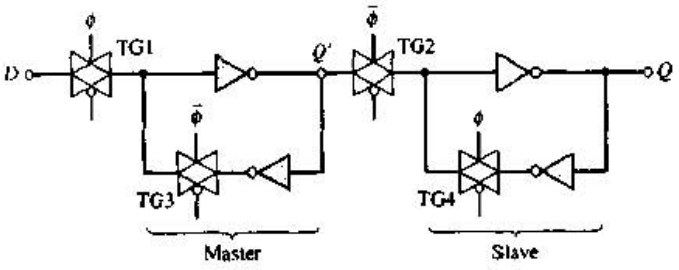


Figure 16.74 CMOS master–slave D flip-flop

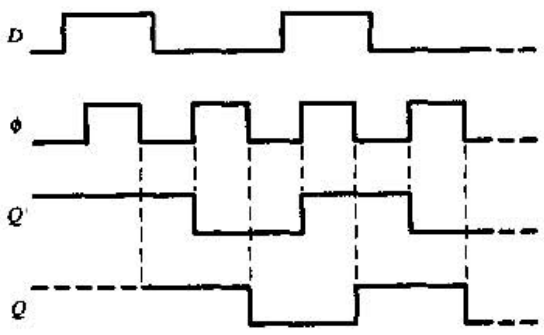


Figure 16.75 D flip-flop signals at various times

16.7.4 CMOS Full-Adder Circuit

One of the most widely used building blocks in arithmetic processing architectures is the one-bit full-adder circuit. We will first consider the logic diagram from the Boolean function and then consider the implementation in a conventional CMOS design.

Assuming that we have two input bits to be added plus a carry signal from a previous stage, the sum-out and carry-out signals are defined by the following two Boolean functions of three input variables A, B, and C.

$$\begin{aligned}\text{Sum-out} &= A \oplus B \oplus C \\ &= ABC + A\overline{B}\,\overline{C} + \overline{A}\,\overline{B}C + \overline{A}B\overline{C}\end{aligned} \tag{16.91(a)}$$

$$\text{Carry-out} = AB + AC + BC \tag{16.91(b)}$$

The logic diagrams for these functions are shown in Figure 16.76. As we have seen previously, the implementation at the transistor level can be done with fewer transistors than would be used if all the NOR and NAND gates were actually connected as shown in the logic diagram.

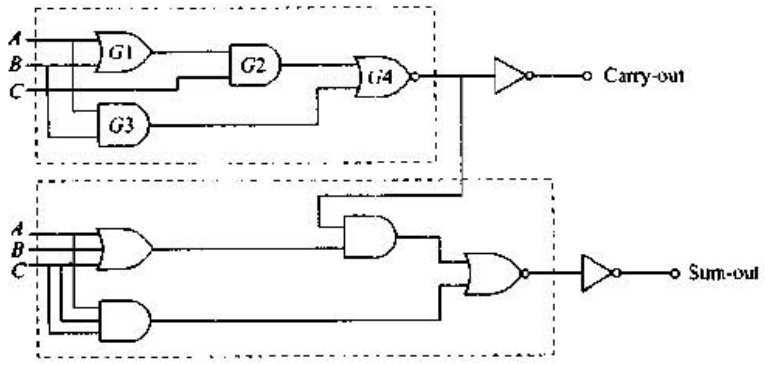


Figure 16.76 Gate configuration of the one-bit full adder

Figure 16.77 is a transistor-level schematic of the one-bit full-adder circuit implemented in a conventional CMOS technology. We can understand the basic design from the logic diagram. For example, we may consider the NMOS portion of the carry-out signal. We see that transistors M_{NA1} and M_{NB1} are in parallel, to perform the basic OR function, and these transistors are in series with transistor M_{NC1}, to perform the basic AND function. These three transistors form the NMOS portion of the design of the two gates labeled G_1 and G_2 in Figure 16.76. We also have transistors M_{NA2} and M_{NB2} in series, to perform the basic AND function of gate G_3. This set of two transistors is in parallel with the previous three transistors, and this configuration performs the basic OR function of gate G_4. This output signal goes through an inverter to become the final carry-out signal.

We can go through the same discussion for the design of the NMOS portion of the sum-out signal. The PMOS design is then the complement of the NMOS design. As mentioned, the total number of transistors in the final design is considerably less than would have occurred if the basic OR and AND gates shown in the logic diagram were actually incorporated in the design.

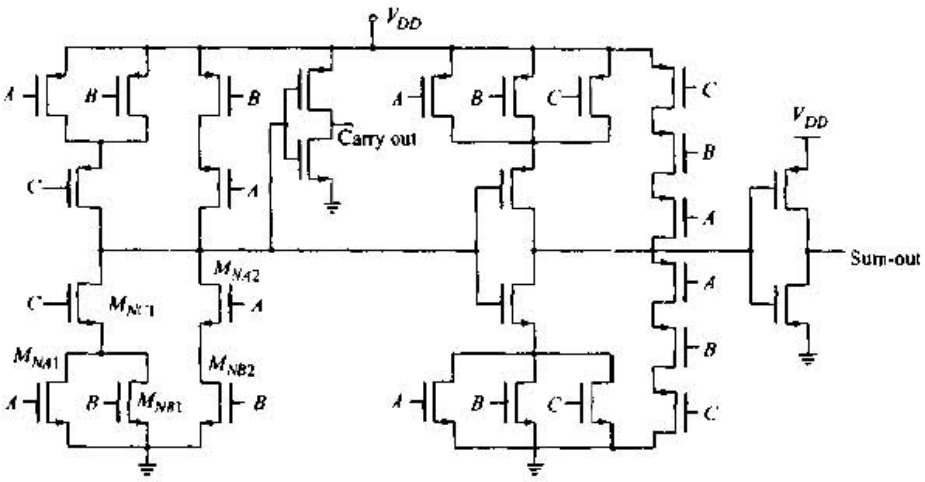


Figure 16.77 Transistor configuration of the CMOS one-bit full adder

16.8 MEMORIES: CLASSIFICATIONS AND ARCHITECTURES

In the previous sections of this chapter, various logic circuits were considered. Combinations of gates can be used to perform logic functions such as addition, multiplication, and multiplexing. In addition to these combinatorial logic functions, digital computers require some method of storing information. Semiconductor circuits form one type of memory, considered in this chapter, and define a class of digital electronic circuits that are just as important as the logic gates.

A memory cell is a circuit, or in some cases just a single device, that can store a bit of information. A systematic arrangement of memory cells constitutes a memory. The memory must also include peripheral circuits to address and write data into the cells as well as detect data that are stored in the cells.

In this section, we define the various types of semiconductor memories, discuss the memory organization, and briefly consider address decoders. In the next section, we analyze in detail some of the basic memory cells and briefly discuss sense amplifiers.

16.8.1 Classifications of Memories

Two basic types of semiconductor memory are considered. The first is the **random access memory** (RAM), a read write memory, in which each individual cell can be addressed at any particular time. The access time to each cell is virtually the same. Implicit in the definition of the RAM is that both the read and write operations are permissible in each cell with also approximately the same access time. Both static and dynamic RAM cells are considered.

A second class of semiconductor memory is the **read-only memory** (ROM). The set of data in this type of memory is generally considered to be fixed, although in some designs the data can be altered. However, the time required to write new data is considerably longer than the read access time of the memory cell. A ROM may be used, for example, to store the instructions of a system operating program.

A volatile memory is one that loses its data when power is removed from the circuit, while nonvolatile memory retains its data even when power is removed. In general, a random access memory is a volatile memory, while read-only memories are nonvolatile.

Random Access Memories

Two types of RAM are the static RAM (SRAM) and dynamic RAM (DRAM). A static RAM consists of a basic bistable flip-flop circuit that needs only a dc current or voltage applied to retain its memory. Two stable states exist, defined as logic 1 and logic 0. A dynamic RAM is an MOS memory that stores one bit of information as charge on a capacitor. Since the charge on the capacitor decays with a finite time constant (milliseconds), a periodic refresh is needed to restore the charge so that the dynamic RAM does not lose its memory.

The advantage of the SRAM is that this circuit does not need the additional complexity of a refresh cycle and refresh circuitry, but the disadvantage is that this circuit is fairly large. In general, a SRAM requires six transistors. The advantage of a DRAM is that it consists of only one transistor and one capacitor, but the disadvantage is the required refresh circuitry and refresh cycles.

Read-Only Memories

There are two general types of ROM. The first is programmed either by the manufacturer (mask programmable) or by the user (programmable, or PROM). Once the ROM has been programmed by either method, the data in the memory are fixed and cannot be altered. The second type of ROM may be referred to as an alterable ROM in that the data in the ROM may be reprogrammed if desired. This type of ROM may be called an EPROM (erasable programmable ROM), EEPROM (electrically erasable PROM), or flash memory. As mentioned, the data in these memories can be reprogrammed although the time involved is much longer than the read access time. In some cases, the memory chip may actually have to be removed from the circuit during the reprogramming process.

16.8.2 Memory Architecture

The basic memory architecture has the configuration shown in Figure 16.78. The terminal connections may include inputs, outputs, addresses, and read and write controls. The main portion of the memory involves the data storage. A RAM memory will have all of the terminal connections mentioned, whereas a ROM memory will not have the inputs and the write controls.

A typical RAM architecture, shown in Figure 16.79, consists of a matrix of storage bits arranged in an array with 2^M columns and 2^N rows. The array may be square, in which case M and N are equal. This particular array may be only one of several on a single chip. To read data stored in a particular cell within the array, a row address is inputted and decoded to select one of the row lines. All of the cells along this row are activated. A column address is also inputted and decoded to select one of the columns. The one particular memory cell at

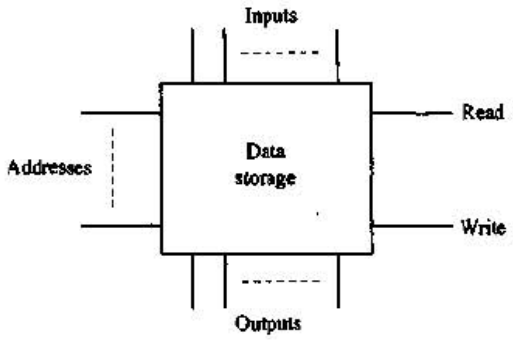


Figure 16.78 Schematic of a basic memory configuration

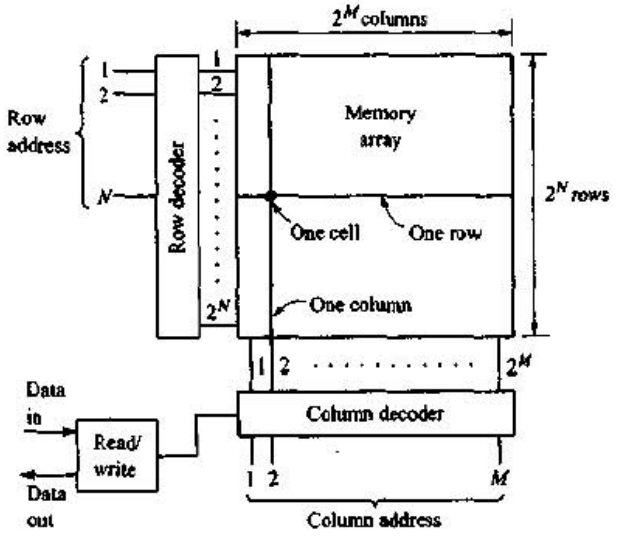


Figure 16.79 Basic random access memory architecture

the intersection of the row and column addressed is then selected. The logic level stored in the cell is routed down a bit line to a sense amplifier.

Control circuits are used to enable or select a particular memory array on a chip and also to select whether data are to be read from or written into the memory cell. Memory chips or arrays are designed to be paralleled so that the memory capacity can be increased. The additional lines needed to address parallel arrays are called **chip select signals**. If a particular chip or array is not selected, then no memory cell is addressed in that particular array. The chip select signal controls the tristate output of the data-in and data-out buffers. In this way, the data-in and data-out lines to and from several arrays may be connected together without interfering with each other.

16.8.3 Address Decoders

The row and column decoders in Figure 16.79 are essential elements in all memories. Access time and power consumption of memories may be largely determined by the decoder design. Figure 16.80 shows a simple decoder with a two-bit input. The decoder uses NAND logic circuits, although the same type

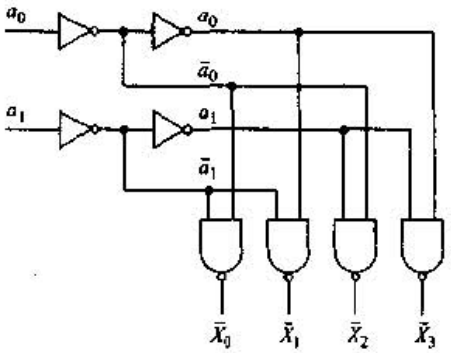


Figure 16.80 Simplified decoder with two-bit input

of decoder may be implemented in NOR gates. The input word goes through input buffers that generate the complement as well as the signal.

Another example of the direct implementation of a decoder is shown in Figure 16.81. Figure 16.81(a) shows a pair of NMOS input buffer-inverters, and Figure 16.81(b) shows a five-input NOR logic address decoder circuit using NMOS enhancement-mode drivers and a depletion load. A pair of input-buffer inverters is required for each input address line. The input signal is then required to drive only an inverter, while the buffer-inverter pair can be designed to drive the remainder of the logic circuits. The output of the NOR decoder goes high only when all inputs are a logic 0. The NOR gate in Figure 16.81(b) would decode the address word 00110 and select the sixth row or column for a read or write operation.

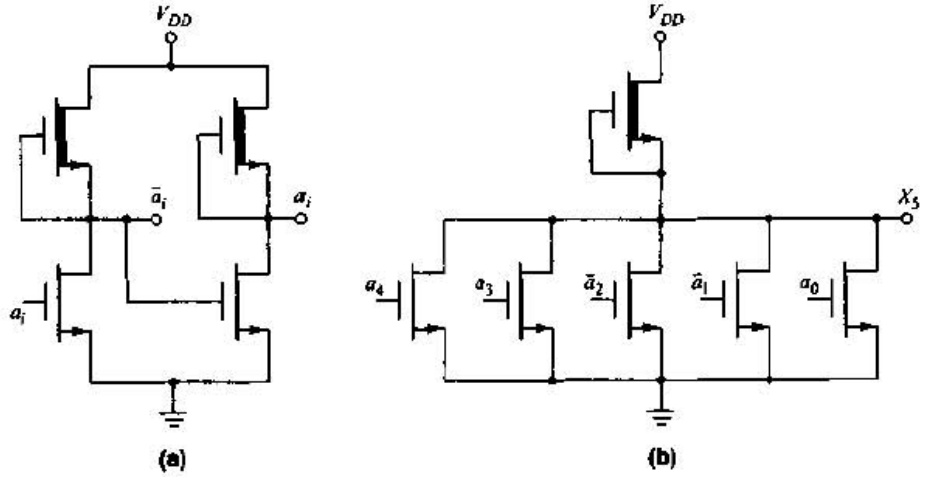


Figure 16.81 (a) Input buffer-inverter pair; (b) five-input NOR logic address decoder

As the size of the memory increases, the length of the address word must increase. For example, a 64-K (where 1 K = 1024 bits) memory whose cells are arranged in a square array would require an 8-bit word for the row address and another 8-bit word for the column address. As the word size increases, the decoder becomes more complex, and the number of transistors and power dissipation may become large. In addition, the total capacitance of MOS deco-

der transistors and interconnect lines increase so that propagation delay times may become significant. The number of transistors required to design a decoder may be reduced by using a two-stage decoder using both NOR and NAND gates. These circuits may be found in more advanced textbooks on digital circuits.

Test Your Understanding

16.30 A NOR logic address decoder, such as shown in Figure 16.81(b), is used in both the row and column address decoders in a memory arranged in a square array. Calculate the number of decoder transistors required for a (a) 1-K, (b) 4-K, and (c) 16-K memory. (Ans. 384, 896, 2048 plus buffer transistors.)

16.9 RAM MEMORY CELLS

In this section, we consider two designs of an NMOS static RAM (SRAM), one design of a CMOS static RAM, and one design of a dynamic RAM (DRAM). We also consider examples of sense amplifiers and read/write circuitry. This section is intended to present the basic concepts used in memory cell design. More advanced designs can again be found in advanced texts on digital circuits.

16.9.1 NMOS SRAM Cells

A static RAM cell is designed by cross-coupling the inputs and outputs of two inverters. In the case of an NMOS design, the load devices may be either depletion-mode transistors or polysilicon resistors, as shown in Figure 16.82. In either case, the inputs and outputs of the two inverters are cross-coupled to form a basic flip-flop. If transistor M_1 is turned on, for example, the output Q

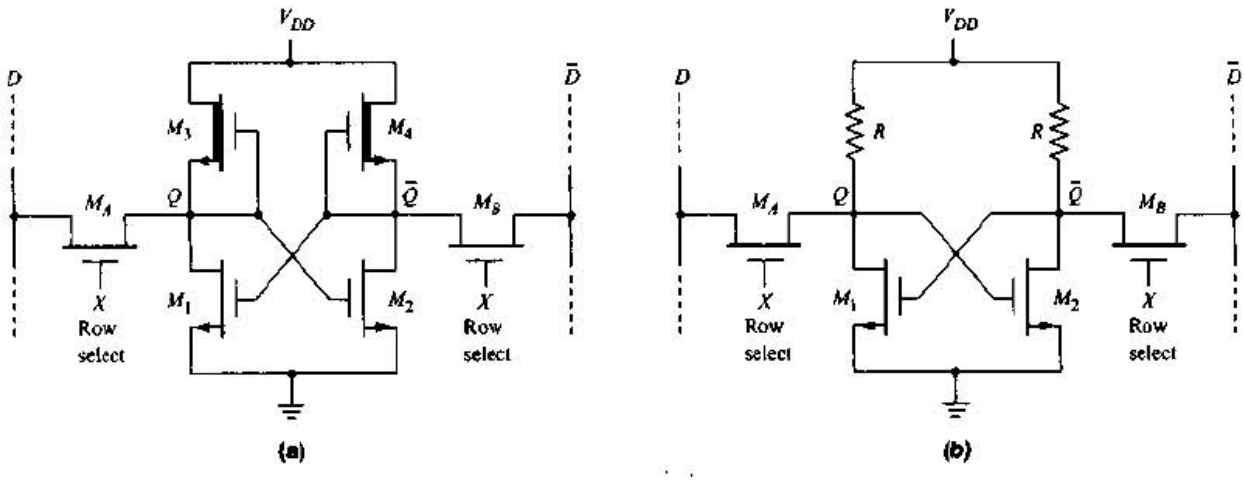


Figure 16.82 Static NMOS RAM cells with (a) depletion loads and (b) polysilicon resistor loads

is low, which means that transistor M_2 is cut off. Since M_2 is cut off, the output $\overline{Q}$ is high, ensuring that M_1 is turned on. Thus, we have a static situation as long as the bias voltage V_{DD} is applied to the circuit.

To access (read or write) the data contained in the memory cell, two NMOS transmission gate transistors, M_A and M_B, connect the memory cell to the complementary bit lines. When the word line signal or row select signal is low, both transmission gate transistors are cut off and the memory cell is isolated or in a standby condition. The data stored in the cell remain stored as long as power is applied to the cell. When the row select or word line signal goes high, the memory cell is then connected to the complementary data lines so that the data in the cell can be read or new data can be written into the cell.

One critical parameter in the design of RAM cells is power dissipation. As we will see in the following example, this is one situation in which incorporating a high-valued resistor as a load device improves the design. A lightly doped polysilicon load resistor is formed by ion implantation, which can accurately dope the polysilicon to produce the designed resistance value.

Example 16.15 **Objective:** Determine the currents, voltages, and power dissipation in two NMOS SRAM cells. The first design uses a depletion-load device and the second design uses a resistor-load device.

Assume the following parameters: $V_{DD} = 3\,\text{V}$ and $k_n' = 60\,\mu\text{A/V}^2$; driver transistors: $V_{TND} = 0.5\,\text{V}$ and $(W/L)_D = 2$; load devices: $V_{TNL} = -1.0\,\text{V}$, $(W/L)_L = 1/2$, and $R = 2\,\text{M}\Omega$.

Solution: With Depletion Load: Assume M_2 is cut off in the circuit in Figure 16.82(a) so that $\overline{Q} = V_{DD} = 3\,\text{V}$. M_1 is on in the nonsaturation region and M_3 is on in the saturation region. The drain current in M_1 and M_3 is then

$$i_D = \frac{k_n'}{2} \cdot \left(\frac{W}{L}\right)_L (V_{GSL} - V_{TNL})^2 = \frac{60}{2} \cdot \left(\frac{1}{2}\right)(0 - (-1))^2$$

or

$$i_D = 15\,\mu\text{A}$$

The power dissipated in the circuit is then

$$P = i_D \cdot V_{DD} = (15)(3) = 45\,\mu\text{W}$$

The logic 0 value of the Q output is found from

$$i_D = \frac{k_n'}{2} \cdot \left(\frac{W}{L}\right)_D [2(V_{GSD} - V_{TND})V_{DSD} - V_{DSD}^2]$$

or

$$15 = \frac{60}{2} \cdot (2)[2(3 - 0.5)Q - Q^2]$$

which yields

$$Q = 50.5\,\text{mV}$$

Solution: With Resistor Load: Again assume M_2 is cut off in the circuit in Figure 16.82(b) so that $\overline{Q} = V_{DD} = 3\,\text{V}$. Again M_1 is on in the nonsaturation region. The drain current is found from

$$\frac{V_{DD} - Q}{R} = \frac{k_n'}{2} \cdot \left(\frac{W}{L}\right)_D [2(V_{GSD} - V_{TND})Q - Q^2]$$

or

$$\frac{3 - Q}{2} = \frac{60}{2} \cdot (2)[2(3 - 0.5)Q - Q^2]$$

[Note that dividing by megohms on the left agrees with microamperes on the right.]

We find

$$Q \cong 5\,\text{mV}$$

The drain current is then found:

$$i_D = \frac{V_{DD} - Q}{R} = \frac{3 - 0.005}{2} \cong 1.5\,\mu\text{A}$$

The power dissipated in the circuit is then

$$P = i_D \cdot V_{DD} = (1.5)(3) = 4.5\,\mu\text{W}$$

Comment: We see that the SRAM with the resistive load dissipates 10 times less power than the SRAM with the depletion-load device. Thus, for a given allowed power dissipation per chip, the memory with the resistive load could be 10 times larger than that using the depletion load device.

Since the value of the load resistance R is, in general, very large, the memory must be designed so that the resistor R is not required to be a pull-up device. We will see this type of design later. The resistors can actually be fabricated on top of the NMOS transistors by a double-polysilicon technology, so that the cell with resistor load devices can be very compact, resulting in a high-density memory.

Test Your Understanding

D16.31 A 16-K NMOS static RAM cell using a resistor load is to be designed. Each cell is to be biased at $V_{DD} = 2.5\,\text{V}$. Assume transistor parameters as described in Example 16.15. The entire memory is to dissipate no more than 125 mW in standby. Design the value of R in each cell to meet this specification. (Ans. $R = 0.82\,\text{M}\Omega$)

16.9.2 CMOS SRAM Cells

The basic six-transistor CMOS SRAM cell is shown in Figure 16.83. The inputs and outputs of the two CMOS inverters are cross-coupled so that the circuit will be in one of two static conditions. For example, if $\overline{Q}$ is low, then M_{N1} is cut off so that Q is high, which in turn means that M_{P2} is cut off, ensuring that $\overline{Q}$ remains low. The two NMOS transmission gate transistors again connect the basic memory cell to the complementary data lines.

The traditional advantages of CMOS technology include low static power dissipation, superior noise immunity to either bipolar or NMOS, wide

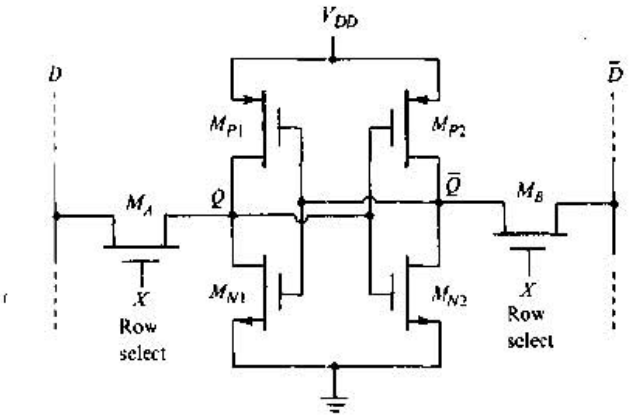


Figure 16.83 A CMOS static RAM cell

operating temperature range, sharp transfer characteristics, and wide voltage supply tolerance.

CMOS is inherently lower power than NMOS, since conducting paths between power and ground do not arise when the circuit is in one logic state or the other. In standard CMOS, the p- and n-channel devices in the memory cell and in the periphery circuits are in series and on at the same time only during switching. Current is, therefore, drawn only during switching. This makes SRAMs and CMOS extremely low power in standby, when there are only surface, junction, and channel leakage currents.

A more complete circuit of the CMOS static RAM is shown in Figure 16.84, which includes PMOS data line pull-up transistors on the complementary bit lines. If all word line signals are zero, then all pass transistors are

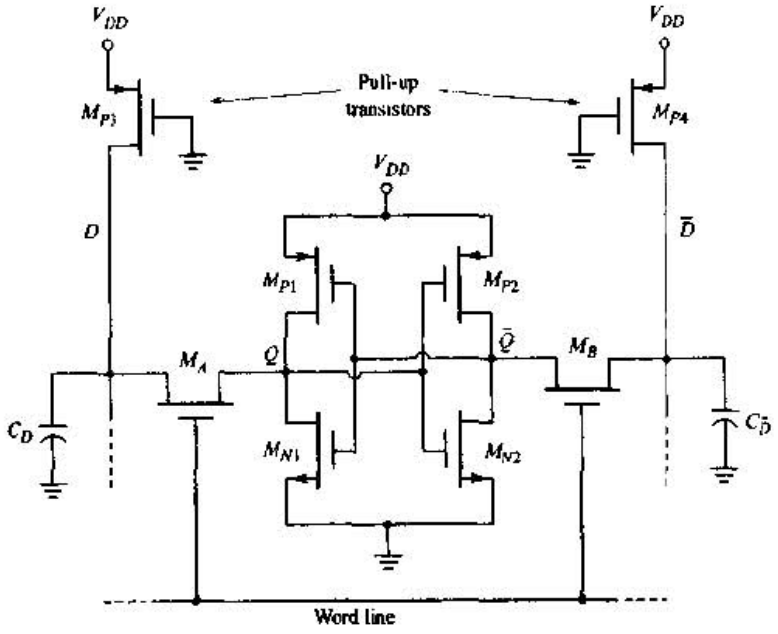


Figure 16.84 CMOS RAM cell including PMOS pull-up transistors

turned off. The two data lines with the relatively large column capacitances are charged up by the column pull-up transistors, M_{P3} and M_{P4}, to the full V_{DD} voltage.

To determine the (W/L) ratios of the transistors in a typical CMOS SRAM cell, two basic requirements must be taken into consideration. First, the read operation should not destroy the information stored in the cell, and second, the cell should allow for the modification of the data stored during a write operation. Consider a read operation in which a logic 0 ($Q = 0$ and $\overline{Q} = V_{DD}$) is stored in the cell. The voltage levels in the cell and on the data lines just prior to the read operation are shown in Figure 16.85. Transistors M_{P1} and M_{N2} are turned off while transistors M_{N1} and M_{P2} are biased in the nonsaturation region.

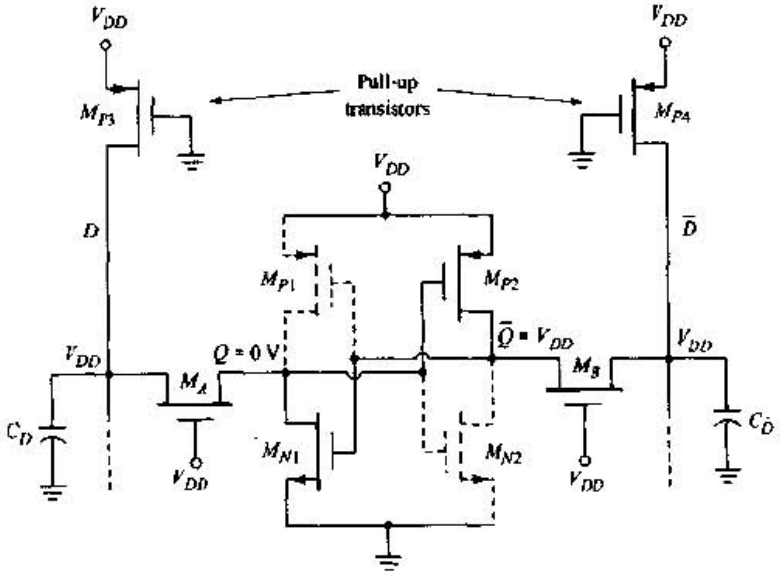


Figure 16.85 Voltage levels and "on" transistors in CMOS RAM cell at the beginning of the read cycle

Immediately after the word select signal is applied to the pass transistors M_A and M_B, the voltage on the $\overline{D}$ data line will not change significantly, since the pass transistor M_B is actually not conducting and no current flows. On the opposite side of the cell, current will flow through M_A and M_{N1} so that the voltage on the D data line will drop and the voltage Q will increase above its initial zero value. The key design point is that Q must not become larger than the threshold voltage of M_{N2}, so that M_{N2} remains cut off during the read phase. This will ensure that there is not a change in the data stored in the cell.

At the initial time the cell is addressed, we can assume that the D bit line remains at approximately V_{DD}, since the line capacitance cannot change instantaneously. The pass transistor M_A is biased in the saturation region and the transistor M_{N1} is biased in the nonsaturation region. Setting the drain currents through M_A and M_{N1} equal, we have

$$K_{nA}(V_{DD} - Q - V_{TN})^2 = K_{n1}[2(V_{DD} - V_{TN})Q - Q^2] \tag{16.92}$$

Setting $Q = Q_{max} = V_{TN}$ as our design limit, then from Equation (16.92), we find the relation between the transistor width-to-length ratios to be

$$\frac{(W/L)_{nA}}{(W/L)_{n1}} < \frac{2(V_{DD}V_{TN}) - 3V_{TN}^2}{(V_{DD} - 2V_{TN})^2} \tag{16.93}$$

Assuming that $V_{DD} = 3\,\text{V}$ and $V_{TN} = 0.5\,\text{V}$, we find that $(W/L)_{nA}/(W/L)_{n1} < 0.56$. So the width-to-length of the pass transistor should be approximately one-half that of the NMOS device in the memory cell. By symmetry, the same condition applies to the transistors M_{N2} and M_B.

We now need to consider the write operation. Assume that a logic 0 is stored and we want to write a logic 1 into the memory cell. Figure 16.86 shows the initial voltage levels in the CMOS SRAM cell when the cell is first addressed at the beginning of the write cycle. Transistors M_{P1} and M_{N2} are initially turned off, and M_{N1} and M_{P2} are biased in the nonsaturation region. The cell voltages are $Q = 0$ and $\overline{Q} = V_{DD}$ just before the pass transistors are turned on. The data line D is held at V_{DD} and the complementary data line $\overline{D}$ is forced to a logic 0 value by the write circuitry. We may assume that $\overline{D} = 0\,\text{V}$ for analysis purposes. The voltage Q will remain below the threshold voltage of M_{N2} because of the condition given by Equation (16.93). Consequently, the voltage at Q is not sufficient to switch the state of the memory cell. To switch the state of the cell, the voltage at $\overline{Q}$ must be reduced below the threshold voltage of M_{N1}, so that M_{N1} will turn off. When $\overline{Q} = V_{TN}$, then M_B is biased in the nonsaturation region and M_{P2} is biased in the saturation region. Equating drain currents, we have

$$K_{p2}(V_{DD} + V_{TP})^2 = K_{nB}\left[2(V_{DD} - V_{TN})V_{TN} - V_{TN}^2\right] \tag{16.94(a)}$$

which can be written in the form

$$\frac{K_{p2}}{K_{nB}} < \frac{2(V_{DD}V_{TN}) - 3V_{TN}^2}{(V_{DD} + V_{TP})^2} \tag{16.94(b)}$$

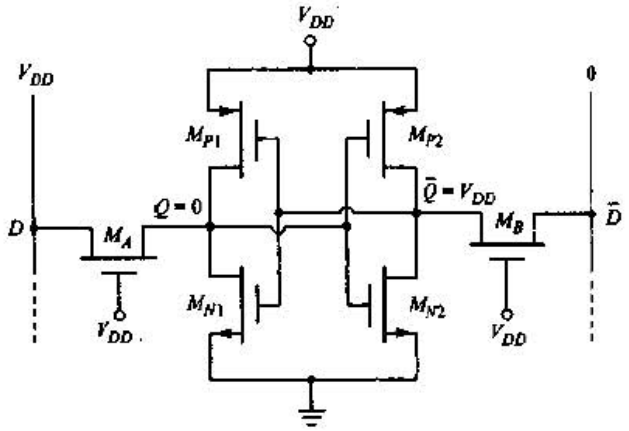


Figure 16.86 Voltage levels in the CMOS RAM at the beginning of a write cycle

Considering the width-to-length ratios, we find

$$\frac{(W/L)_{p2}}{(W/L)_{nB}} < \frac{k_n'}{k_p'} \cdot \frac{2(V_{DD}V_{TN}) - 3V_{TN}^2}{(V_{DD} + V_{TP})^2} \tag{16.95}$$

Assuming that $V_{DD} = 3\,\text{V}$, $V_{TN} = 0.5\,\text{V}$, $V_{TP} = -0.5\,\text{V}$, and $(k_n'/k_p') = (\mu_n/\mu_p) = 2$, we find that $(W/L)_{p2}/(W/L)_{nB} < 0.72$.

From previous results, if we assume that the width-to-length of the pass transistor is one-half that of the NMOS in the memory cell, and if we assume that the width-to-length of the PMOS in the memory cell is 0.7 that of the pass transistor, then the width-to-length of the PMOS in the cell should be approximately 0.35 that of the NMOS in the memory cell.

Test Your Understanding

D16.32 A six-transistor CMOS SRAM cell is biased at $V_{DD} = 2.5\,\text{V}$. The transistor parameters are $V_{TN} = +0.4\,\text{V}$, $V_{TP} = -0.4\,\text{V}$, and $(\mu_n/\mu_p) = 2.5$. Determine the relative width-to-length ratios such that Equations (16.92) through (16.95) are satisfied in terms of read/write requirements.

16.9.3 SRAM Read/Write Circuitry

An example of a read/write circuit at the end of a column is shown in Figure 16.87. We may consider the write portion of the circuit as shown in Figure 16.88(a). We may note that if the column is not selected, then M_3 is cut off and the two data lines are held at their precharged value of V_{DD}. When $X = Y = 1$, then the one-bit cell shown is addressed. If $\overline{W} = 1$ then the write cycle is deselected and both M_1 and M_2 are cut off. For $\overline{W} = 0$ and $D = 1$, M_1 is cut off and M_2 is turned on so that the $\overline{D}$ data line is pulled low while the D data line remains high. The logic 1 is then written into the cell. For $\overline{W} = 0$ and $D = 0$, the D data line is pulled low and the $\overline{D}$ data line is held high so that logic 0 is written into the cell.

Figure 16.88(b) shows the NMOS cross-coupled sense amplifier that is in the complete circuit of Figure 16.87. This circuit does not generate an output signal, but rather amplifies the small difference in the data bit lines. Suppose that a logic 1 is to be read from the memory cell. When the cell is addressed, the D bit line is high and the $\overline{D}$ bit line voltage begins to decrease. This means that when the M_3 transistor turns on, the M_2 transistor turns on harder than M_1 so that the $\overline{D}$ bit line voltage is pulled low and the M_1 transistor will eventually turn off.

Figure 16.88(c) shows the differential amplifier that senses the output of the memory cell. Note that this sense amplifier is connected to the bit lines through a couple of pass transistors, as seen in Figure 16.87. If the input signal to the pass transistors is also a function of the column select signal, then this configuration enables the use of one main sense amplifier to read the data out of several columns, one at a time. When the clock signal is zero, the M_3 transistor in the differential amplifier is cut off and the common source node

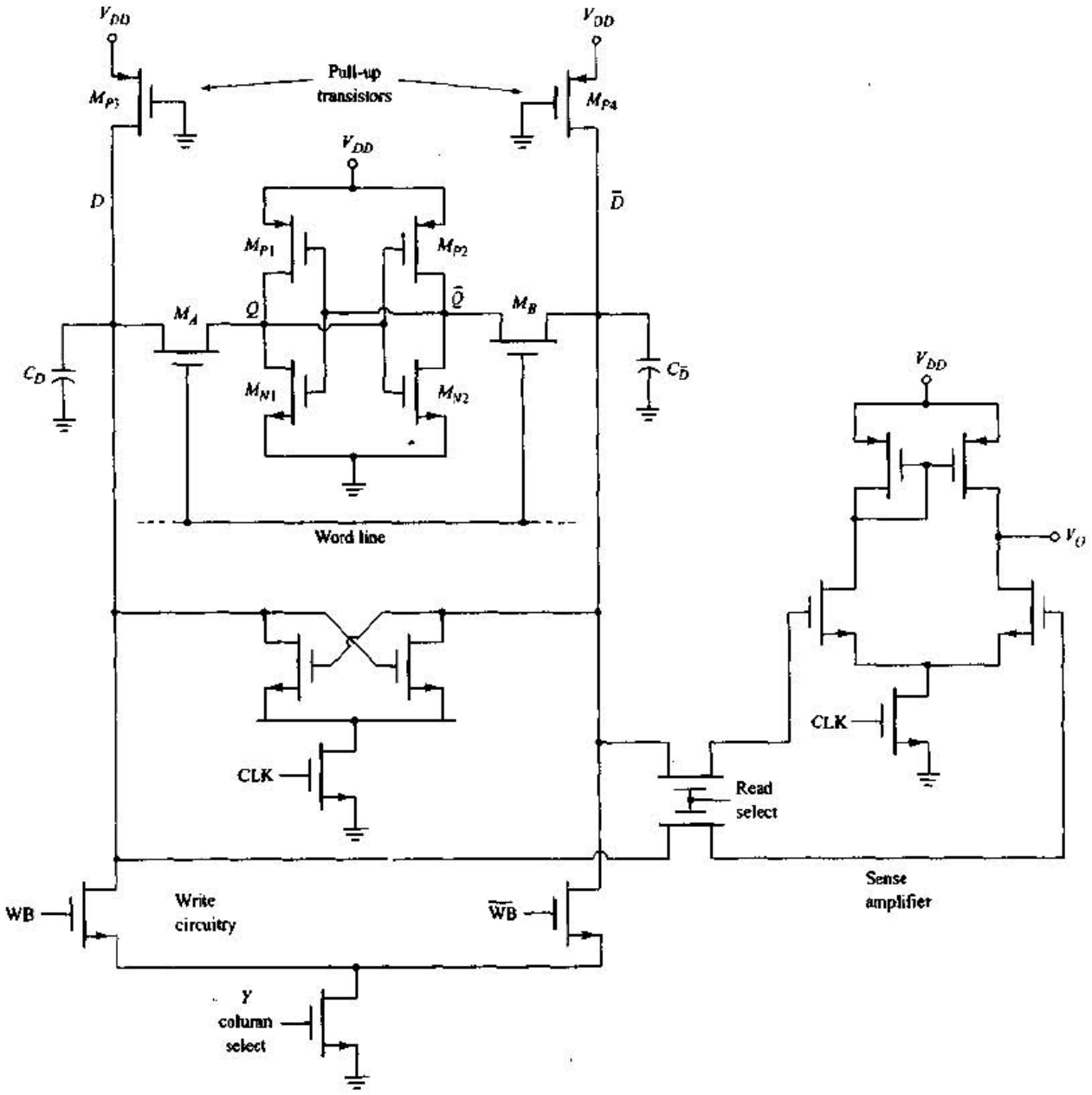


Figure 16.87 Complete circuit diagram of a CMOS RAM cell with write and read circuitry

of M_1 and M_2 is pulled high, which means the output voltage is pulled high. When a memory cell is selected and the clock goes high, M_3 turns on. If a logic 1 level is to be read, then D remains high and the $\overline{D}$ line voltage decreases. This means that the M_2 transistor will turn off and the output voltage remains high. If a logic 0 is to be read, then the D line voltage decreases and $\overline{D}$ remains high. The transistor M_1 will turn off while M_2 is turned on so that the output voltage goes low.

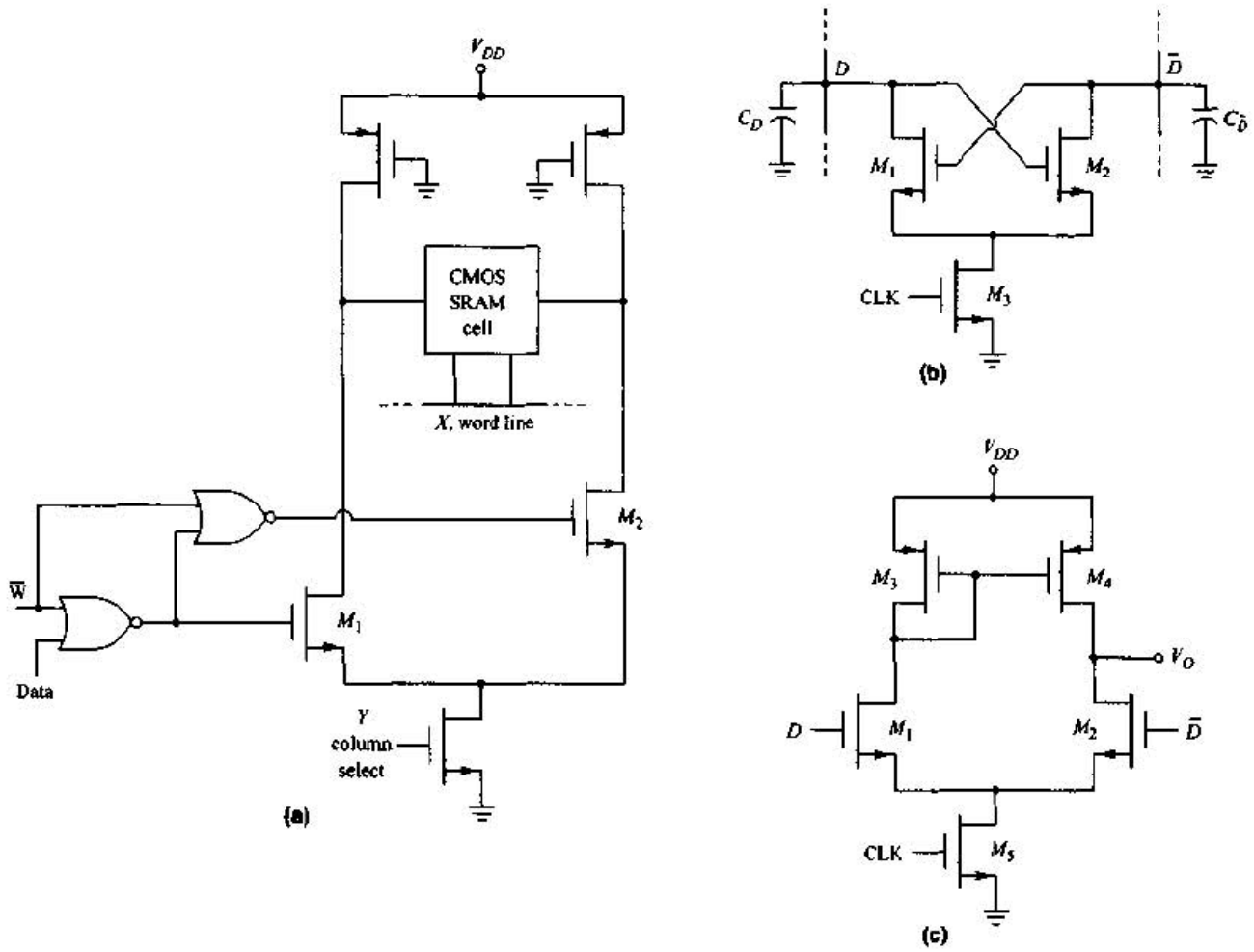


Figure 16.88 (a) Write circuitry associated with CMOS RAM cell; (b) cross-coupled NMOS sense amplifier; (c) CMOS differential sense amplifier

16.9.4 Dynamic RAM (DRAM) Cells

The CMOS RAM cell just considered requires six transistors and five lines connecting each cell, including the power and ground connections. A substantial area, then, is required for each memory cell. If the area per cell could be reduced, then higher-density RAM arrays would be possible.

In a dynamic RAM cell, a bit of data is stored as charge on a capacitor, where the presence or absence of charge determines the value of the stored bit. Data stored as charge on capacitors cannot be retained indefinitely, since leakage currents will eventually remove the stored charge. Thus the name *dynamic* refers to the situation in which a periodic refresh cycle is required to maintain the stored data.

One design of a DRAM cell is the one-transistor cell that includes a pass transistor M_S plus a storage capacitor C_S, shown in Figure 16.89. Binary information is stored in the form of zero charge on C_S (logic 0) and stored charge on C_S (logic 1). The cell is addressed by turning on the pass transistor via the word line signal WL and charges are transferred into or out of C_S on the bit line BL.

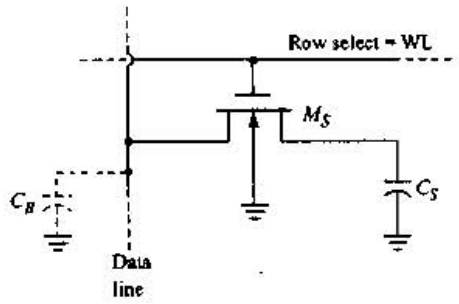


Figure 16.89 One-transistor dynamic RAM cell

The storage capacitor is isolated from the rest of the circuit when M_S is off, but the stored charge on C_S decreases because of the leakage current through the pass transistor. This effect was discussed in detail in Section 16.6 during the analysis of the NMOS pass transistor. As a result of this leakage, the cell must be refreshed regularly to restore its original condition.

An example of a sense amplifier to detect the charge stored in the memory cell is shown in Figure 16.90. On one side of the amplifier is a memory cell that either stores a full charge or is empty, depending on the binary value of the data. On the other side of the amplifier is a reference cell with a reference or dummy storage capacitor C_R that is one-half the value of the storage capacitor. The charge on C_R will then be one-half the logic 1 charge on C_S. A cross-coupled dynamic latch circuit is used to detect the small voltage differences and to restore the signal levels. The capacitors C_D and C_{DR} represent the relatively large parasitic bit line and reference bit line capacitances.

In the standby mode, the bit lines on both sides of the sense amplifier are precharged to the same potential. During the read cycle, both the WL and D–WL address signals go high allowing the charges in the cells to be redistributed along the bit lines. After the charge equalization and since the charge in the dummy cell is half the full charge, then $v_1 < v_2$ when the memory cell is empty or a logic 0, and $v_1 > v_2$ when the memory cell is full or a logic 1. The sense amplifier detects and amplifies the voltage difference between the bit lines, and will latch at the logic level stored in the basic memory cell.

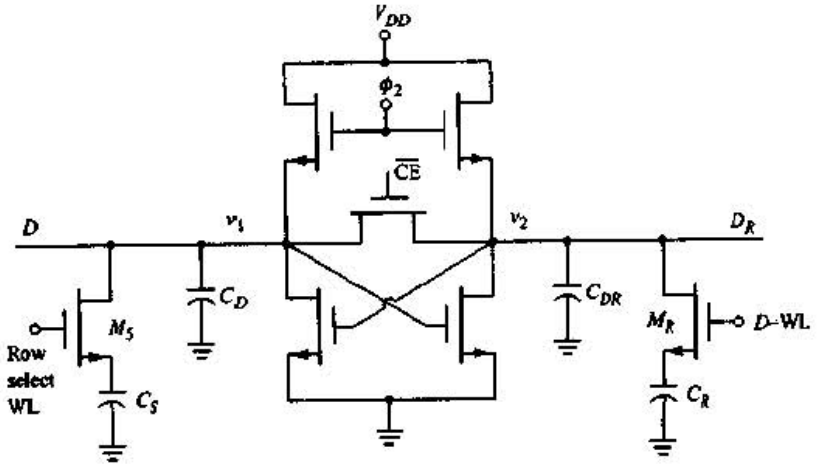


Figure 16.90 Sense amplifier configuration for dynamic RAM cell

Test Your Understanding

16.33 A one-transistor DRAM cell is composed of a 0.05 pF storage capacitor and an NMOS transistor with a 0.5 V threshold voltage. A logic 1 is written into the cell when both the data line and row-select line are raised to 3 V. Sensing circuitry permits the stored charge to decay to 50 percent of its original value. Refresh occurs every 1.5 ms. Determine the maximum allowed leakage current that can exist.

16.10 READ-ONLY MEMORY

We consider several examples of read-only memories in this section. The intent is again to provide an introduction to this type of memory. In the case of EPROMs and EEPROMs, the development effort has been directed toward the characteristics of the basic memory cell.

16.10.1 ROM and PROM Cells

We consider two types of ROMs. The first example is a mask-programmed ROM, in which contacts to devices are selectively included or excluded in the final manufacturing process to obtain the desired memory pattern. Figure 16.91 shows an example of an NMOS 16 × 1 mask-programmed ROM.

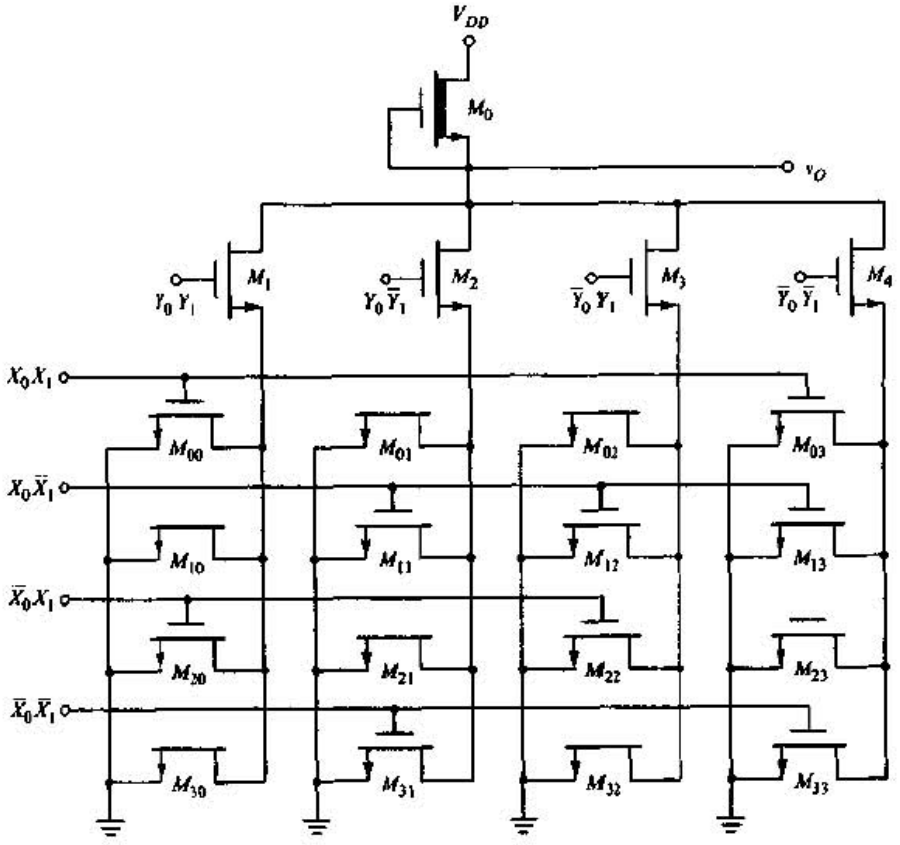


Figure 16.91 An NMOS 16 × 1 mask-programmable ROM

Enhancement-mode NMOS transistors are fabricated in each of the 16 cell positions (the substrate connections are omitted for clarity). However, gate connections are fabricated only on selected transistors. The transistors M_1–M_4 are column-select transistors and M_0 is a depletion-mode load device.

The inputs X_0, X_1, Y_0, and Y_1 are the row- and column-select signals. If, for example, $X_0 = \overline{X}_1 = \overline{Y}_0 = Y_1 = 1$, then the M_{12} transistor is addressed. Transistors M_{12} and M_3 turn on with this address, forcing the output to a logic 0. If the address changes, for example, to $\overline{X}_0 = X_1 = \overline{Y}_0 = \overline{Y}_1 = 1$, then the transistor M_{23} is addressed. However, this transistor does not have a gate connection and consequently never turns on, so the output is a logic 1.

The mask-programmed memory discussed is only a 16 × 1-bit ROM, while a more useful memory would contain many more bits. Memories can be organized in any desired manner, such as a 2048 × 8 for a 16-K memory. This ROM is a nonvolatile memory, since the data stored are not lost when power is removed.

The second example of a ROM is a user-programmed ROM. The data pattern is defined by the user after the final manufacture rather than during the manufacture. One specific type is shown schematically in Figure 16.92. A small fuse is in series with each emitter and can be selectively "blown" or left in place by the user. If, for example, the fuse in Q_{00} is left in place and this transistor is addressed by $X_0 = X_1 = Y_0 = Y_1 = 1$, then Q_{00} turns on, raising the data line voltage at the emitter of Q_{00}. The inverter N_1 is enabled, making the output a logic 0. If the fuse is blown in this transistor, then the input to the inverter is a logic 0, so the output is a logic 1.

The polysilicon fuse in the emitter of an npn bipolar transistor has a fairly low resistance, so with the fuse in place and at low currents, there is very little voltage drop across the fuse. When the current through the fuse is increased to the 20 to 30 mA range, the heating of the polysilicon fuse causes the temperature to increase. The silicon oxidizes, forming an insulator that effectively opens the path between the data line and the emitter. The bipolar ROM circuit with the fuses either in place or "blown" form a permanent ROM that is not alterable and is also nonvolatile.

16.10.2 EPROM and EEPROM Cells

An EPROM transistor is shown in Figure 16.93. The device has a double gate, with gate 1 being a "floating gate" that has no electrical contact. Gate 2 is used for cell selection, taking the role of the single gate of an MOS transistor.

Operation of this EPROM cell relies on being able to store charge on the floating gate. Initially, we assume no charge on the floating gate so that with gate 2, drain, and source grounded, the potential of gate 1 is also zero. As the voltage on gate 2 increases, the gate 1 voltage rises also, but at a lower rate as determined by the capacitive divider. The net effect of this is to effectively raise the threshold voltage of this MOSFET as seen from gate 2. However, when the gate 2 voltage is raised sufficiently (approximately twice the normal threshold voltage), a channel forms. Under these conditions, the device provides a stored logic 0 when used in the NOR array.

To write a logic 1 into this cell, both gate 2 and drain are raised to about 25 V while the source and substrate remain at ground potential. A relatively large drain current flows because of normal device conduction characteristics.

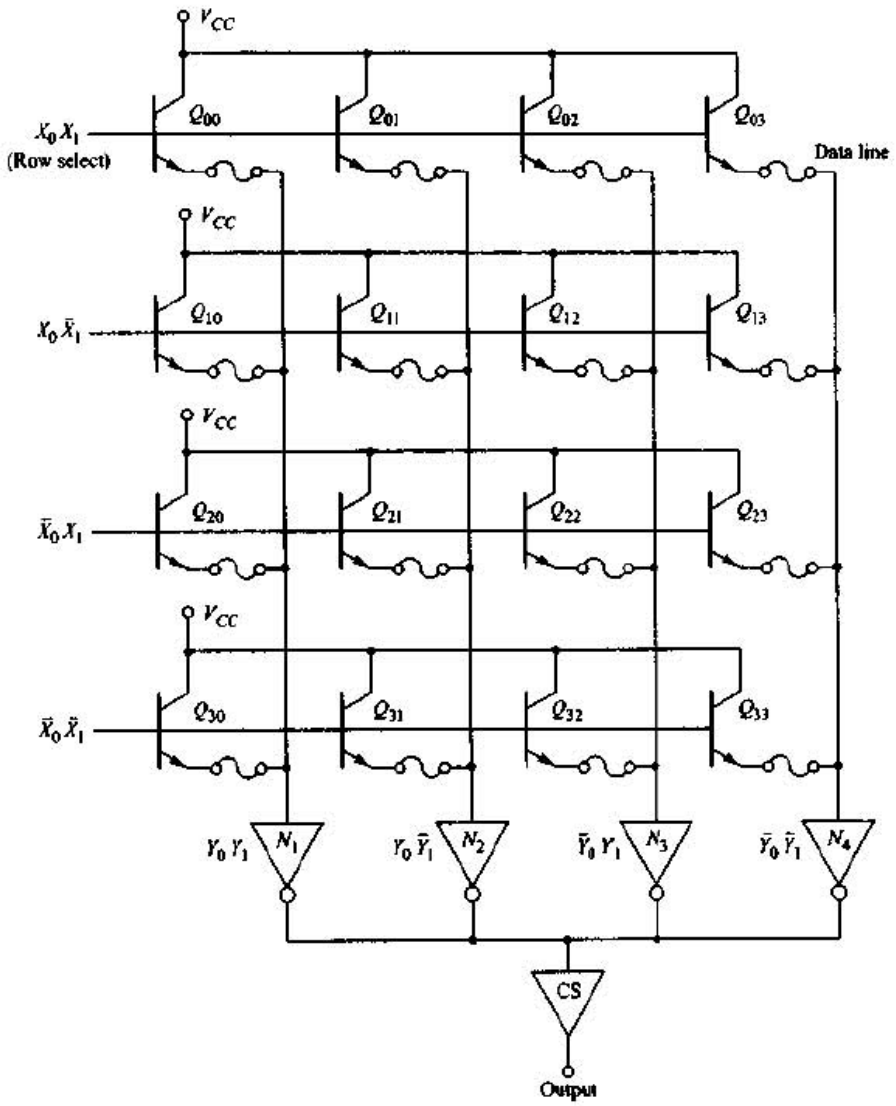


Figure 16.92 A bipolar fuse-linked user-programmable ROM

In addition, the high field in the drain–substrate depletion region results in avalanche breakdown of the drain–substrate junction, with a considerable additional flow of current. The high field in the drain depletion region accelerates electrons to high velocity such that a small fraction traverse the thin oxide and become trapped on gate 1. When the gate 2 and drain potentials are reduced to zero, the negative charge on gate 1 forces its potential to approximately −5 V. If the gate 2 voltage for reading is limited to +5 V, then a channel never forms. Thus a logic 1 is stored in the cell.

Gate 1 is completely surrounded by silicon dioxide (SiO_2), an excellent insulator, so charge can be stored for many years. Data can be erased, however, by exposing the cells to strong ultraviolet (UV) light. The UV radiation generates electron–hole pairs in the SiO_2 making the material slightly conductive. The negative charge on the gate can then leak off, restoring the transistor to its original uncharged condition. These EPROMs must be assembled in

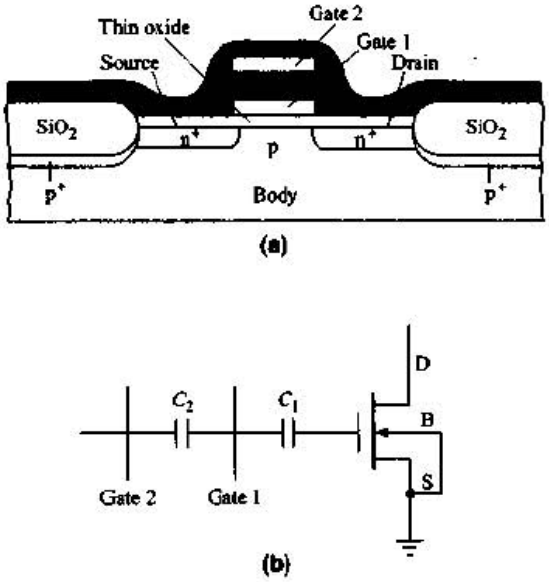


Figure 16.93 (a) Cross section of erasable programmable ROM; (b) equivalent circuit

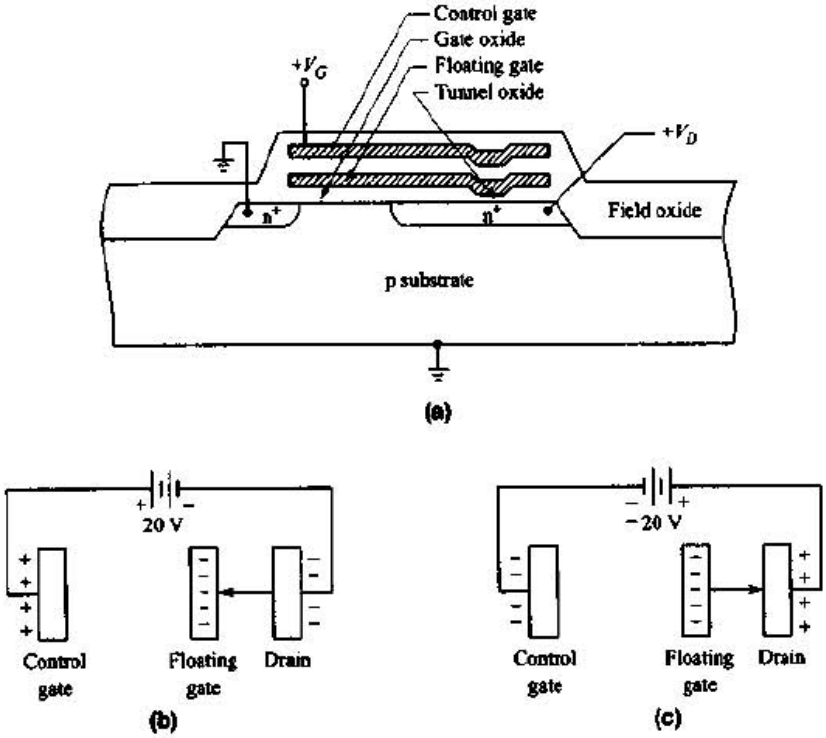


Figure 16.94 (a) Cross section of a floating-gate electrically erasable programmable ROM; (b) charging the floating gate; (c) discharging the floating gate

packages with transparent covers so the silicon chip may be exposed to UV radiation. One disadvantage is that the entire memory must be erased before any reprogramming can be done. In general, reprogramming must also be done on specialized equipment; therefore, the EPROM must be removed from the circuit during this operation.

In the EEPROM, each individual cell can be erased and reprogrammed without disturbing any other cell. The most common form of EEPROM is also a floating gate structure; one example is shown in Figure 16.94(a). The memory transistor is similar to an n-channel MOSFET, but with a physical difference in the gate insulator region. Charge may exist on the floating gate that will alter the threshold voltage of the device. If a net positive charge exists on the floating gate, the n-channel MOSFET is turned on, whereas if zero or negative charge exists on the floating gate, the device is turned off.

The floating gate is capacitively coupled to the control gate with the tunnel oxide thickness less than 200 Å. If 20 V is applied to the control gate while keeping $V_D = 0$, electrons tunnel from the n^+-drain region to the floating gate as demonstrated in Figure 16.94(b). This puts the MOSFET in the enhancement mode with a threshold voltage of approximately 10 V, so the device is effectively off. If zero volts is applied to the control gate and 20 V is applied to the drain terminal, then electrons tunnel from the floating gate to the n^+-drain terminal as demonstrated in Figure 16.94(c). This leaves a net positive charge on the floating gate that puts the device in the depletion mode with a threshold voltage of approximately −2 V, so the device is effectively on. If all voltages are kept to within 5 V during the read cycle, this structure can retain its charge for many years.

16.11 SUMMARY

- In this chapter, NMOS and CMOS digital logic circuits were analyzed and designed. These included basic logic gates, shift registers, flip-flops, and memories.
- The discussion of NMOS logic circuits served as an introduction to the analysis and design of digital logic circuits. Since this technology deals with only one type of transistor (n-channel), the analysis and design is straightforward.
- The NMOS inverter is the basis of NMOS logic circuits. The quasi-static voltage transfer characteristics of NMOS inverters with resistive load, enhancement load, and depletion load were generated. The transfer characteristics were designed to provide appropriate logic 0 values by designing the width-to-length ratios of the transistors. The impact of the body effect on the transfer curves and logic values was analyzed. The noise margin of the NMOS inverter is defined as the point where the magnitude of the voltage gain is unity.
- The basic NMOS NOR and NAND logic gates were analyzed. More sophisticated logic functions can be implemented by combining driver transistors in particular series and parallel combinations. The width-to-length ratios of the driver transistors were designed to produce a composite conduction parameter to produce a specified logic 0 value.
- The CMOS inverter is the basis of the CMOS logic circuits. The quasi-static voltage transfer characteristics were generated. For the CMOS circuit, the quiescent power dissipation is essentially zero when the input is in either logic state. The extremely low static power dissipation is the primary advantage of the CMOS technology. The

switching power dissipation is given by $P = fC_L V_{DD}^2$, where f is the switching frequency, C_L is the effective load capacitance, and V_{DD} is the supply voltage. The tendency in CMOS design is toward lower supply voltages on CMOS digital logic circuits because of the squared term in the power equation.

- The basic CMOS NOR and NAND logic gates were analyzed. In the classical CMOS design, the gates of a PMOS and NMOS are connected together. CMOS logic circuits are usually designed to provide equal current drive in the NMOS pull-down and PMOS pull-up portions of the circuit. Transistor width-to-length ratios were designed to provide equal composite conduction parameters in the NMOS and PMOS circuits.
- More sophisticated logic functions can be implemented in the classical CMOS technology. NMOS transistors in series implement the basic AND function and NMOS transistors in parallel implement the basic OR function. The combination of the PMOS transistors is the complement of the NMOS design. By combining transistors in a particular series or parallel combination, more complex logic functions can be implemented.
- Since the mobility of carriers in the PMOS transistor is smaller than that in the NMOS transistor, PMOS devices must be approximately twice as large as NMOS devices to provide the same current drive. Therefore, a savings in chip area as well as reduced capacitance can be achieved by eliminating as many PMOS transistors as possible. Clocked CMOS logic circuits achieve this goal. A generalized NMOS logic circuit is inserted between clocked PMOS and NMOS devices. The advantage of low static power dissipation is maintained.
- Sequential logic circuits are a class of circuits whose output depends not only on the inputs, but is also a function of the previous history of the inputs. Shift registers, flip-flops, and a full one-bit adder were analyzed in this section. Dynamic shift registers are formed with transmission gates and inverters. Both NMOS and CMOS designs were analyzed. A flip-flop can be implemented by cross-coupling two NOR gates. This bistable circuit can remain in either stable state indefinitely, as long as power is applied. A full one-bit CMOS adder was analyzed at both the gate and the transistor level.
- A whole classification of circuits called memories was considered. Typically, an array of memory cells is organized in a square matrix to form a memory. A cell is addressed via row and column decoders and data are read from the cell or written into a cell through data lines.
- A random-access memory (RAM) cell is a circuit or device that can store one bit of information, and whose information can be written (stored) or retrieved (read) with essentially the same access time. A static RAM (SRAM) retains its data as long as power is applied, whereas a dynamic RAM (DRAM) loses its stored data over time by leakage currents. The DRAM data must be refreshed.
- Three SRAM designs were considered. In the two NMOS designs, static power is dissipated in the cell, which limits the size of the memory because of the total chip power limitation. A CMOS SRAM was designed. The primary advantage of essentially no static power dissipation is again the primary advantage of CMOS technology. The size of a CMOS memory is limited primarily by chip area requirements. An example of the peripheral read/write circuitry required was considered.
- Read-only memory (ROM and PROM) contains fixed data that are implemented by the manufacturer (mask programmed) or by the user (user programmed). In both cases, the data cannot be altered. In the case of a mask-programmed ROM, for example, the gates of MOSFETs may be fabricated or may be deliberately left off in a cell depending on whether a logic 1 or logic 0 is to be stored. For a user-programmed ROM, a fuse in a particular memory cell can be left in place or "blown," depending on whether a logic 1 or logic 0 is to be stored.

- Erasable read-only memory (EPROM and EEPROM) cells contain MOSFETs with floating gates. The floating gates can be either charged or left uncharged by the user depending on whether a logic 1 or logic 0 is to be stored. The charge on the floating gate can be altered so that the data in the ROM can be erased and reprogrammed. The writing of new data, however, takes a relatively long time compared to the read access time.

CHECKPOINT

After studying this chapter, the reader should have the ability to:

✔ Analyze the transfer characteristics of NMOS inverters, including the determination of noise margins. (Section 16.1)

✔ Design an NMOS logic circuit to perform a specific logic function. (Section 16.2)

✔ Analyze the transfer characteristics of the CMOS inverter, including the determination of switching power and noise margins. (Section 16.3)

✔ Design a CMOS logic circuit to perform a specific logic function. (Section 16.4)

✔ Design a clocked CMOS logic circuit to perform a specific logic function. (Section 16.5)

✔ Design an NMOS or CMOS pass network to perform a specific logic function. (Section 16.6)

✔ Design an NMOS or CMOS RAM cell and design a simple sense amplifier. (Section 16.9)

REVIEW QUESTIONS

1. Explain qualitatively what is meant by the body effect in an NMOS device and discuss its effect on the threshold voltage of the NMOS transistor.
2. Sketch the quasi-static voltage transfer characteristics of an NMOS inverter with resistive load. Discuss the various intervals in terms of transistor bias. What is the effect on the transfer curve of increasing the transistor W/L ratio?
3. Sketch the quasi-static voltage transfer characteristics of an NMOS inverter with enhancement load. Discuss the various intervals in terms of transistor bias. Why doesn't the output ever reach the V_{DD} value? What is the effect on the transfer curve of changing the transistor W/L ratios?
4. Sketch the quasi-static voltage transfer characteristics of an NMOS inverter with depletion load. Discuss the advantage of the depletion-load inverter compared to the other two NMOS inverter designs. What is the effect on the transfer curve of changing the transistor W/L ratios?
5. Define the noise margin of an NMOS inverter.
6. What is the impact of the body effect on the NMOS inverter voltage transfer characteristics of each of the inverter designs?
7. Sketch an NMOS three-input NOR logic gate. Describe its operation. Discuss the condition under which the maximum logic 0 value is obtained.
8. Sketch an NMOS three-input NAND logic gate. Describe its operation. Discuss the effect of changing the driver transistor W/L ratios.
9. Discuss how more sophisticated (compared to the basic NOR and NAND) logic functions can be implemented in a single NMOS logic circuit.
10. Sketch the quasi-static voltage transfer characteristics of a CMOS inverter. Discuss the various intervals in terms of transistor bias. What is the effect on the transfer

curve of changing the transistor W/L ratios? What is the advantage of the CMOS inverter compared to an NMOS inverter?

11. Sketch the quasi-static current versus input voltage of a CMOS inverter. Discuss the various intervals in terms of transistor bias.
12. Discuss the difference between the static power dissipation and switching power dissipation in a CMOS inverter.
13. Discuss the parameters that affect the switching power dissipation in a CMOS inverter.
14. Define the noise margin in a CMOS inverter.
15. Sketch a CMOS three-input NOR logic gate. Describe its operation. Determine relative transistor W/L ratios to obtain equal NMOS and PMOS composite conduction parameters.
16. Sketch a CMOS three-input NAND logic gate. Describe its operation. Determine relative transistor W/L ratios to obtain equal NMOS and PMOS composite conduction parameters.
17. Discuss how more sophisticated (compared to the basic NOR and NAND) logic functions can be implemented in a single CMOS logic circuit.
18. Discuss the basic principles of a clocked CMOS logic circuit. Discuss why, in general, PMOS transistors must be larger in size than NMOS transistors.
19. Sketch an NMOS transmission gate and describe its operation. If the input and gate voltages are both V_{DD}, determine the maximum output voltage. Why can't the output voltage reach V_{DD}?
20. Consider three NMOS transmission gates in series or in cascade. If the input voltage and each gate voltage is V_{DD}, determine the output voltage. Discuss why three threshold voltage drops are *not* lost between the input and output.
21. Sketch a CMOS transmission gate and describe its operation. For this circuit, discuss why the quasi-static output voltage is always equal to the quasi-static input voltage.
22. Discuss what is meant by pass transistor logic.
23. If an NMOS or CMOS transmission gate is turned off (an open switch), discuss why the output voltage is, in general, not stable.
24. Sketch an NMOS dynamic shift register and describe its operation.
25. Sketch a CMOS R–S flip flop and describe its operation. Why must the input condition $R = S = 1$ be avoided?
26. Describe the basic architecture of a semiconductor random-access memory.
27. Discuss the differences between SRAM and DRAM cells. Discuss advantages and disadvantages of each design.
28. Sketch an NMOS SRAM cell and describe its operation. Discuss any disadvantages of this design.
29. Sketch a CMOS SRAM cell and describe its operation. Discuss any advantages and disadvantages of this design. Describe how the cell is addressed.
30. Describe the voltage levels in a CMOS SRAM cell during a read operation. Describe any limitations in voltage changes in the cell during this read cycle.
31. Describe the voltage levels in a CMOS SRAM cell during a write operation. Assume a logic 0 is initially stored and a logic 1 is to be written into the cell.
32. Sketch a one-transistor DRAM cell and describe its operation. What makes this circuit dynamic?
33. Describe a mask-programmed MOSFET ROM memory.
34. Describe the basic operation of a floating gate MOSFET and how this can be used in an erasable ROM.

PROBLEMS

[Note: In the following problems, unless otherwise stated, assume: $k_n' = 80\,\mu A/V^2$, $k_p' = 35\,\mu A/V^2$, $\lambda = 0$ for all transistors; $V_{TNO} = 0.8\,V$ for all n-channel enhancement-mode transistors; and $V_{TPO} = -0.8\,V$ for all p-channel enhancement-mode transistors. Neglect the body effect unless otherwise stated. The temperature is 300 °K.]

Section 16.1 NMOS Inverters

16.1 Consider an NMOS transistor with parameters: $K_n = 0.2\,mA/V^2$, $V_{TNO} = 0.8\,V$, $N_a = 8 \times 10^{15}\,cm^{-3}$, $t_{ox} = 450\,\text{Å}$, and $\phi_{fp} = 0.343\,V$. (a) Determine the change in threshold voltage from $V_{SB} = 1\,V$ and $V_{SB} = 2\,V$. (b) If $V_{GS} = 2.5\,V$ and $V_{DS} = 5\,V$, find the transistor current for $V_{SB} = 0$, $V_{SB} = 1\,V$, and $V_{SB} = 2\,V$.

RD16.2 The load resistor in the NMOS inverter in Figure 16.5(a) is 40 kΩ. The circuit is biased at $V_{DD} = 5\,V$. (a) Redesign the width-to-length ratio of the driver transistor such that $v_O = 0.10\,V$ when $v_I = 5\,V$. (b) Using the results of part (a), find the driver transition point and the maximum power dissipated in the inverter circuit.

16.3 For the circuit in Figure 16.5(a), assume the transistor conduction parameter is $K_n = 50\,\mu A/V^2$. (a) Plot the voltage transfer characteristics for $0 \le v_I \le 5\,V$ and for $R_D = 20\,k\Omega$. Mark the values of v_I and v_O at the transition point. (b) Repeat part (a) for $R_D = 200\,k\Omega$.

RD16.4 Redesign the inverter in Figure 16.5(a) such that the maximum power dissipated is no greater than 1 mW, and the output voltage is 0.2 V when the input voltage is 5 V. Determine the load resistance R_D and the transistor width-to-length ratio.

16.5 An NMOS inverter with saturated load is shown in Figure 16.8(a). Let $V_{DD} = 10\,V$, $V_{TNL} = V_{TND} = 2\,V$, $K_D = 200\,\mu A/V^2$, and $K_L = 50\,\mu A/V^2$. Calculate the transition point and determine v_O when $v_I = 8\,V$. Sketch the voltage transfer characteristics.

16.6 An NMOS inverter with saturated load is shown in Figure 16.8(a). The bias is $V_{DD} = 3\,V$ and the transistor threshold voltages are 0.5 V. (a) Find the ratio K_D/K_L such that $v_O = 0.25\,V$ when $v_I = 3\,V$. (b) Repeat part (a) for $v_I = 2.5\,V$. (c) If $W/L = 1$ for the load transistor, determine the power dissipation in the inverter for parts (a) and (b).

RD16.7 Consider the NMOS inverter with saturated load in Figure 16.8(a). Let $V_{DD} = 3\,V$ and let the threshold voltages be 0.5 V. (a) Redesign the circuit such that the power dissipated in the circuit is 400 μW and the output voltage is 0.10 V when the input voltage is a logic 1. Determine the driver transition point. (b) Determine the noise margin for this inverter.

16.8 The NMOS inverter with saturated load in Figure 16.8(a) operates with a supply voltage of V_{DD}. The MOSFETs have threshold voltages of $V_{TN} = 0.2V_{DD}$. Determine $(W/L)_D/(W/L)_L$ such that $V_O = 0.08V_{DD}$. Neglect the body effect.

16.9 The enhancement-load transistor in the NMOS inverter in Figure P16.9 has a separate bias applied to the gate. Assume transistor parameters of $K_n = 1\,mA/V^2$ for M_D, $K_n = 0.4\,mA/V^2$ for M_L, and $V_{TN} = 1\,V$ for both transistors. Using the appropriate logic 0 and logic 1 input voltages, determine V_{OH} and V_{OL} for: (a) $V_B = 4\,V$, (b) $V_B = 5\,V$, (c) $V_B = 6\,V$, and (d) $V_B = 7\,V$.

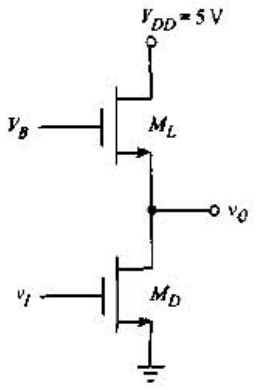


Figure P16.9

16.10 For the depletion-load NMOS inverter circuit in Figure 16.10(a), assume: $V_{DD} = 5\,V$, $V_{TNL} = -2\,V$, $V_{TND} = 0.8\,V$, $K_L = 100\,\mu A/V^2$, and $K_D = 500\,\mu A/V^2$. (a) Find the transition points for the load and driver transistors. (b) Calculate the

value of v_O for $v_I = 5\,\text{V}$. (c) Calculate i_D when $v_I = 5\,\text{V}$. (d) Sketch the voltage transfer characteristics for $0 \leq v_I \leq 5\,\text{V}$.

16.11 In the depletion-load NMOS inverter circuit in Figure 16.10(a), let $V_{TND} = 0.5\,\text{V}$ and $V_{DD} = 3\,\text{V}$, $K_L = 50\,\mu\text{A/V}^2$, and $K_D = 500\,\mu\text{A/V}^2$. Calculate the value of V_{TNL} such that $v_O = 0.10\,\text{V}$ when $v_I = 3\,\text{V}$.

RD16.12 Consider the NMOS inverter with depletion load in Figure 16.10(a). Let $V_{DD} = 3\,\text{V}$, and assume $V_{TNL} = -1.0\,\text{V}$ and $V_{TND} = 0.5\,\text{V}$. (a) Redesign the circuit such that the maximum power dissipated in the circuit is $150\,\mu\text{W}$ and the minimum output voltage is 0.10 V when the input voltage is a logic 1. Determine the transition points for the driver and load transistors. (b) Determine the noise margin for this inverter.

D16.13 The NMOS inverter with depletion load is shown in Figure 16.10(a). The bias is $V_{DD} = 2.5\,\text{V}$. The transistor parametes are $V_{TND} = 0.5\,\text{V}$ and $V_{TNL} = -1\,\text{V}$. The width-to-length ratio of the load device is $W/L = 1$. (a) Design the driver transistor such that $v_O = 0.05\,\text{V}$ when the input is a logic 1. (b) What is the power dissipated in the circuit when $v_I = 2.5\,\text{V}$?

16.14 Calculate the power dissipated in each inverter circuit in Figure P16.14 for the following input conditions: (a) Inverter a: (i) $v_I = 0.5\,\text{V}$, (ii) $v_I = 5\,\text{V}$; (b) Inverter b: (i) $v_I = 0.25\,\text{V}$, (ii) $v_I = 4.3\,\text{V}$; (c) Inverter c: (i) $v_I = 0.03\,\text{V}$, (ii) $v_I = 5\,\text{V}$.

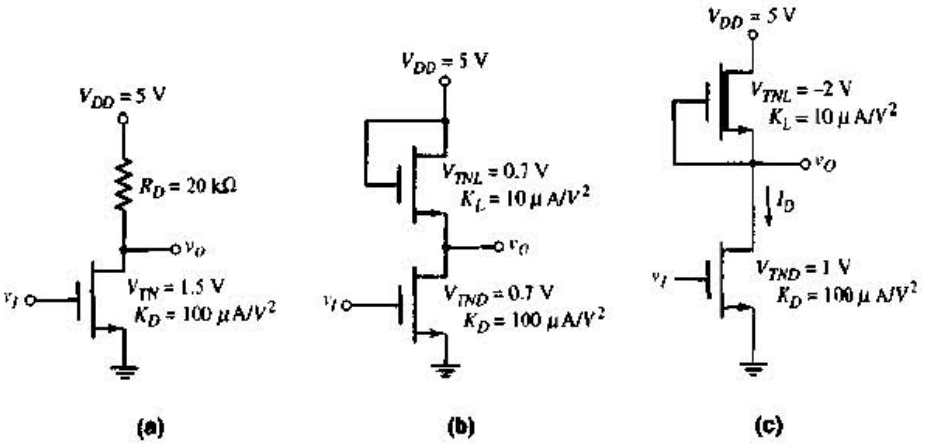


Figure P16.14

16.15 For the two inverters in Figure P16.15, assume width-to-length ratios of $W/L = 1$ for the load devices and $W/L = 10$ for the driver devices. Determine the values of v_I and v_{O2} if $v_{O1} = V_{IH}$. What is the value of V_{IH}?

16.16 Consider the circuit in Figure P16.16. The parameters of the driver transistors are $V_{TND} = 0.8\,\text{V}$ and $W/L = 4$, and those of the load transistors are $V_{TNL} = -2\,\text{V}$ and $W/L = 1$. (a) Find the values of v_I and v_{O2} if $v_{O1} = V_{IH}$. (b) Repeat part (a) if $v_{O1} = V_{IL}$. (c) What are the values of V_{IH} and V_{IL} in parts (a) and (b)?

16.17 For the two transistors in the NMOS inverter with saturated load in Figure 16.17(a), assume the parameters are as described in Problem 16.1, except that $K_D = 200\,\mu\text{A/V}^2$ and $K_L = 20\,\mu\text{A/V}^2$. Let $V_{DD} = 5\,\text{V}$. (a) Determine the output voltage when $v_I = 0$ for: (i) neglecting the body effect, and (ii) taking the body effect into account. (b) Compare the results of part (a) with a computer simulation analysis.

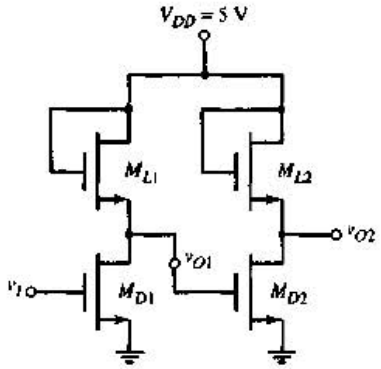


Figure P16.15

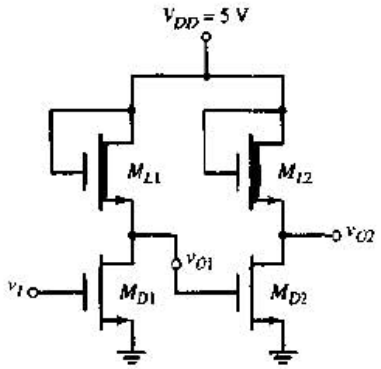


Figure P16.16

16.18 Consider the NMOS inverter with depletion load in Figure 16.17(b). Assume that the circuit and transistor parameters are the same as those given and determined in Example 16.4. Assume the body effect coefficient for the load transistor is $\gamma = 0.35\,\text{V}^{1/2}$. From a computer simulation, plot the load curve for: (a) neglecting the body effect, and (b) taking the body effect into account.

Section 16.2 NMOS Logic Circuits

16.19 Consider the circuit with a depletion load device shown in Figure P16.19. (a) Let $v_X = 5\,\text{V}$ and $v_Y = 0.20\,\text{V}$. Determine K_D/K_L such that $v_O = 0.20\,\text{V}$. (b) Using the results of part (a), determine v_O when $v_X = v_Y = 5\,\text{V}$. (c) If the width-to-length ratio of the depletion device is $W/L = 1$, determine the power dissipation in the logic circuit for the input conditions listed in parts (a) and (b).

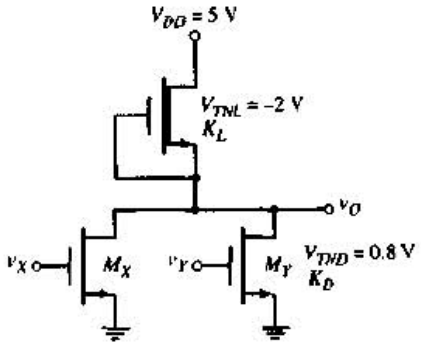


Figure P16.19

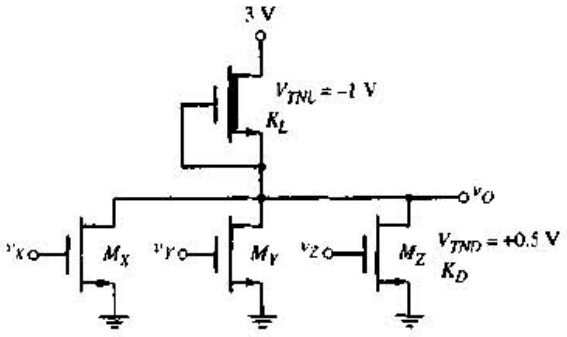


Figure P16.20

D16.20 Consider the three-input NOR logic gate in Figure P16.20. The transistor parameters are $V_{TNL} = -1\,\text{V}$ and $V_{TND} = 0.5\,\text{V}$. The maximum value of v_O in its low state is to be 0.1 V. (a) Determine K_D/K_L. (b) The maximum power dissipation in the NOR logic gate is to be 0.1 mW. Determine the width-to-length ratios of the transistors. (c) Determine v_O when $v_X = v_Y = v_Z = 3\,\text{V}$.

16.21 The transistor parameters for the circuit in Figure P16.21 are: $V_{TN} = 0.8\,\text{V}$ for all enhancement-mode devices, $V_{TN} = -2\,\text{V}$ for the depletion-mode devices, and $k_n' = 60\,\mu\text{A/V}^2$ for all devices. The width-to-length ratios of M_{L2} and M_{L3} are 1, and those for

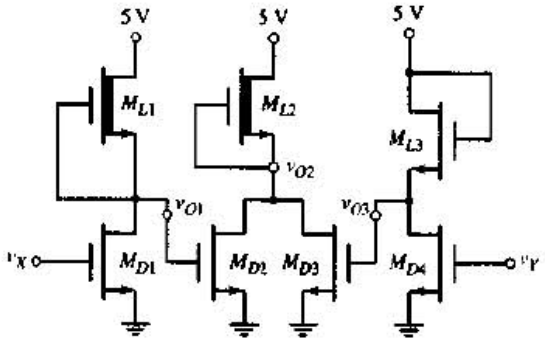


Figure P16.21

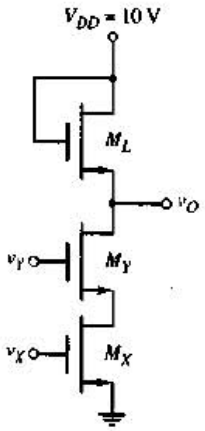


Figure P16.22

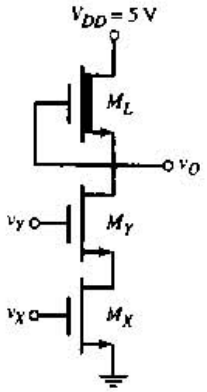


Figure P16.23

M_{D2}, M_{D3}, and M_{D4} are 8. (a) For $v_X = 5\,\text{V}$, output v_{O1} is 0.15 V, and the power dissipation in this inverter is to be no more than 250 μW. Determine $(W/L)_{ML1}$ and $(W/L)_{MD1}$. (b) For $v_X = v_Y = 0$, determine v_{O2}.

16.22 In the NMOS circuit in Figure P16.22, the transistor parameters are: $(W/L)_X = (W/L)_Y = 9$, $(W/L)_L = 1$, and $V_{TN} = 0.8\,\text{V}$ for all transistors. (a) Determine v_O when $v_X = v_Y = 9.2\,\text{V}$. What are the values of v_{GSX}, V_{GSY}, v_{DSX}, and v_{DSY}? (b) Repeat part (a) for $\gamma = 0.5$.

16.23 In the NMOS circuit in Figure P16.23, the transistor parameters are: $(W/L)_X = (W/L)_Y = 4$, $(W/L)_L = 1$, $V_{TNX} = V_{TNY} = 0.8\,\text{V}$, and $V_{TNL} = -1.5\,\text{V}$. (a) Determine v_O when $v_X = v_Y = 5\,\text{V}$. (b) What are the values of v_{GSX}, v_{GSY}, v_{DSX}, and v_{DSY}? Repeat part (a) for $\gamma = 0.5$.

16.24 Find the logic function implemented by the circuit in Figure P16.24.

16.25 Find the logic function implemented by the circuit in Figure P16.25.

D16.26 The Boolean function for a carry-out signal of a one-bit full adder is given by

$$\text{Carry-out} = A \cdot B + A \cdot C + B \cdot C$$

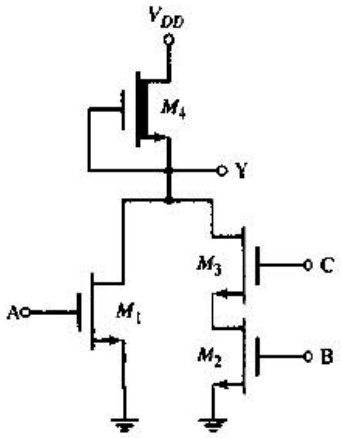


Figure P16.24

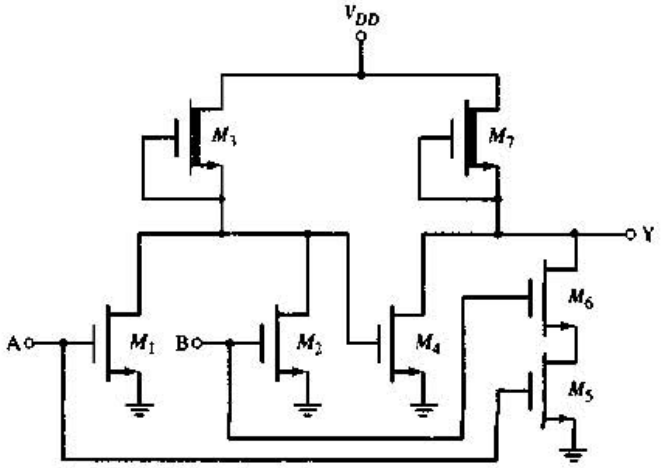


Figure P16.25