CHAPTER 13

CMOS Digital Logic Circuits

Introduction 1061

13.1 Digital Logic Inverters 1062

13.2 The CMOS Inverter 1089

13.3 Dynamic Operation of the CMOS Inverter 1098

13.4 CMOS Logic-Gate Circuits 1110

13.5 Implications of Technology Scaling: Issues in Deep-Submicron Design 1122

Summary 1132

Problems 1134

IN THIS CHAPTER YOU WILL LEARN

1. How the operation of the basic element in digital circuits, the logic inverter, is characterized by such parameters as noise margins, propagation delay, and power dissipation, and how it is implemented by using one of three possible arrangements of voltage-controlled switches (transistors).
2. That the three most significant metrics in digital IC design are speed of operation, power dissipation, and silicon area, and that each design is in effect a trade-off among the three metrics.
3. How and why CMOS has become the dominant technology for digital IC design.
4. The structure, circuit operation, static and dynamic performance analysis, and the design of the CMOS inverter.
5. The synthesis and design optimization of CMOS logic circuits.
6. The implications of technology scaling (Moore's law) over 40 years and continuing, and some of the current challenges in the design of deep-submicron ($L < 0.25$ μm) circuits.

Introduction

This chapter does three things: It introduces the basic element of digital circuits, the logic inverter; it presents a relatively detailed study of the CMOS inverter and of CMOS logic-circuit design; and it provides a perspective on the astounding phenomenon of technology scaling (Moore's law) and the opportunities and challenges of deep-submicron ($L < 0.25$ μm) IC design.

Our study of the inverter in Section 13.1 provides the foundation for the study of digital electronics in the remainder of the chapter and in the next two chapters. Without getting into circuit implementation detail, Section 13.1 introduces all the parameters and metrics used in digital IC design. As well, it provides an overview of digital IC technologies and logic-circuit families. In this way, it provides the basis for appreciating how and why CMOS has emerged the dominant technology in digital IC design. The section concludes with a discussion of the various styles of digital system design: from small-scale and medium-scale integrated-circuit (SSI and MSI) packages assembled on printed-circuit boards to systems assembled using very-large-scale integrated (VLSI) circuits such as microprocessors, memory, and custom and semicustom ICs.

Sections 13.2 and 13.3 provide a comprehensive and thorough study of the CMOS inverter. Section 13.4 builds on this material and presents the basic CMOS logic-gate circuits as well as a general approach for the CMOS implementation of arbitrary logic functions. We also consider the design optimization of the resulting circuits.

The chapter concludes with a retrospective and a prospective look at Moore's law and the technology scaling that has continued over the last 40 years and shows no signs of stopping. This leads naturally to a discussion of the phenomena that take place in deep-submicron ($L < 0.25\ \mu m$) MOSFETs and how to modify the model we studied in Chapter 5 to take account of these phenomena. This section should serve as a bridge between this introductory course and more advanced study of digital IC design.

This chapter provides a self-contained study of CMOS logic circuits, the bread and butter of digital IC design. We will build on this foundation in our study of the more specialized topics in the next two chapters.

13.1 Digital Logic Inverters

The logic inverter is the most basic element in digital circuit design; it plays a role parallel to that of the amplifier in analog circuits. In this section we provide an introduction to the logic inverter and to digital circuit design.

13.1.1 Function of the Inverter

As its name implies, the logic inverter inverts the logic value of its input signal. Thus, for a logic-0 input, the output will be a logic 1, and vice versa. In terms of voltage levels, consider the inverter shown in block form in Fig. 13.1. Its implementation will ensure that when v_I is low (close to 0 V), the output v_O will be high (close to V_{DD}), and vice versa.

13.1.2 The Voltage-Transfer Characteristic (VTC)

To quantify the operation of the inverter, we utilize its voltage-transfer characteristic (VTC). We have already introduced the concept of the VTC and utilized it to characterize the operation of basic MOSFET amplifiers in Section 5.4.2. Figure 13.2 shows such a circuit, together with its VTC. Observe that the circuit in fact implements the inverter function: For a logic-0 input, v_I is close to 0 V and specifically lower than the MOSFET threshold voltage V_{tn}, the transistor will be off, $i_D = 0$, and $v_O = V_{DD}$, which is a logic 1. For a logic-1 input, $v_I = V_{DD}$, the transistor will be conducting and operating in the triode region (at point D on the VTC), and the output voltage will be low (logic 0).

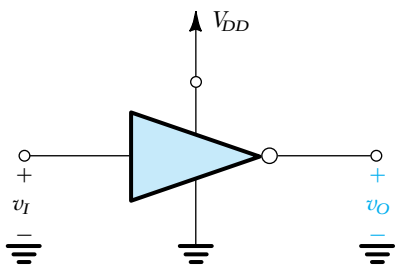


Figure 13.1 A logic inverter operating from a dc supply V_{DD}.

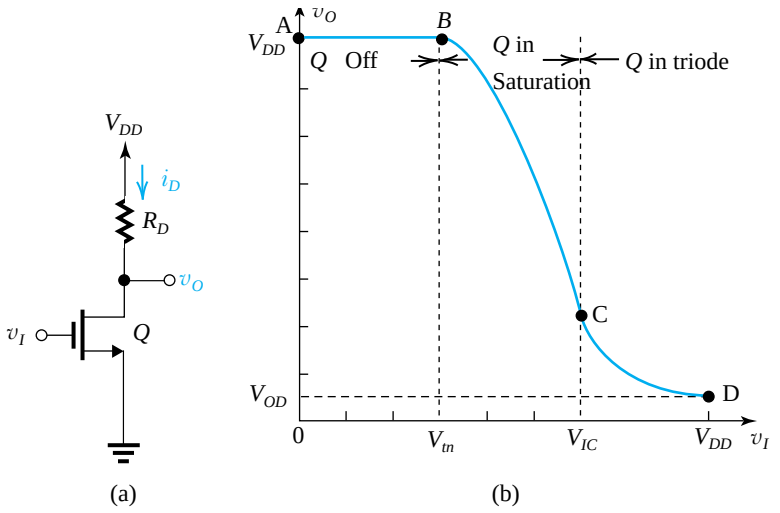


Figure 13.2 The simple resistively loaded MOS amplifier can be used as a logic inverter when operated in cut-off $(v_I < V_{tn})$ and in triode $(v_I > V_{IC})$. The output high level is V_{DD} and the low level is V_{OD}.

Thus to use this amplifier as a logic inverter, we utilize its extreme regions of operation. This is exactly the opposite to its use as a signal amplifier, where it would be biased at the middle of the transfer characteristic segment BC and the signal kept small enough to restrict operation to a short, almost linear, segment of the transfer curve. Digital applications, on the other hand, make use of the gross nonlinearity exhibited by the VTC.

With these observations in mind, we show in Fig. 13.3 a possible VTC of a logic inverter. For simplicity, we are using three straight lines to approximate the VTC, which is usually a nonlinear curve such as that in Fig. 13.2. Observe that the output high level, denoted V_{OH}, does not depend on the exact value of v_I as long as v_I does not exceed the value labeled V_{IL}; when v_I exceeds V_{IL}, the output decreases and the inverter enters its amplifier region of

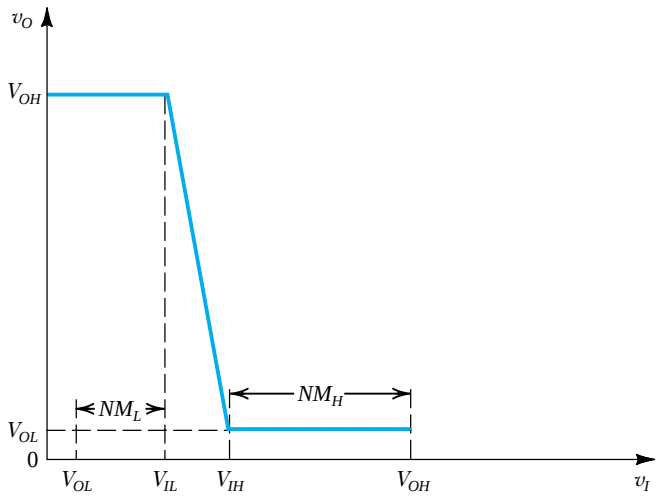


Figure 13.3 Voltage transfer characteristic of an inverter. The VTC is approximated by three straight-line segments. Note the four parameters of the VTC (V_{OH}, V_{OL}, V_{IL}, and V_{IH}) and their use in determining the noise margins (NM_H and NM_L).

operation, also called the **transition region**. It follows that V_{IL} is an important parameter of the inverter VTC: It is the *maximum value that* v_I *can have while being interpreted by the inverter as representing a logic 0*.

Similarly, we observe that the output low level, denoted V_{OL}, does not depend on the exact value of v_I as long as v_I does not fall below V_{IH}. Thus V_{IH} is an important parameter of the inverter VTC: It is the *minimum value that* v_I *can have while being interpreted by the inverter as representing a logic 1*.

13.1.3 Noise Margins

The insensitivity of the inverter output to the exact value of v_I within allowed regions is a great advantage that digital circuits have over analog circuits. To quantify this insensitivity property, consider the situation that occurs often in a digital system where an inverter (or a logic gate based on the inverter circuit) is driving another similar inverter, as shown in Fig. 13.4

Here we assume that a noise or interference signal v_N is somehow coupled to the interconnection between the output of inverter G_1 and the input of inverter G_2 with the result that the input of G_2 becomes

$$v_{I2} = v_{O1} + v_N \tag{13.1}$$

where the noise voltage v_N can be either positive or negative. Now consider the case $v_{O1} = V_{OL}$; that is, inverter G_2 is driven by a logic-0 signal. Reference to Fig. 13.3 indicates that in this case G_2 will continue to function properly as long as its input v_{I2} does not exceed V_{IL}. Equation (13.1) then indicates that v_N can be as high as $V_{IL}–V_{OL}$ while G_2 continues to function properly. Thus, we can say that inverter G_2 has a **noise margin for low input**, NM_L, of

$$NM_L = V_{IL} - V_{OL} \tag{13.2}$$

Similarly, if $v_{O1} = V_{OH}$, the driven inverter G_2 will continue to see a high input as long as v_{I2} does not fall below V_{IH}. Thus, in the high-input state, inverter G_2 can tolerate a negative v_N of magnitude as high as $V_{OH} - V_{IH}$. We can thus state that G_2 has a **high-input noise margin**, NM_H, of

$$NM_H = V_{OH} - V_{IH} \tag{13.3}$$

In summary, four parameters, V_{OH}, V_{OL}, V_{IH}, and V_{IL}, define the VTC of an inverter and determine its noise margins, which in turn measure the ability of the inverter to tolerate.

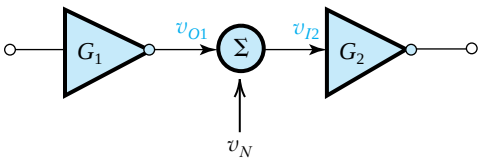


Figure 13.4 Noise voltage v_N is coupled to the interconnection between the output of inverter G_1 and the input of inverter G_2.

Table 13.1 Important Parameters of the VTC of the Logic Inverter (Refer to Fig. 13.3)

V_{OL}:	Output low level
V_{OH}:	Output high level
V_{IL}:	Maximum value of input interpreted by the inverter as a logic 0
V_{IH}:	Minimum value of input interpreted by the inverter as a logic 1
NM_L:	Noise margin for low input = $V_{IL} - V_{OL}$
NM_H:	Noise margin for high input = $V_{OH} - V_{IH}$

variations in the input signal levels. In this regard, observe that changes in the input signal level within the noise margins are *rejected* by the inverter. Thus noise is not allowed to propagate further through the system, a definite advantage of digital over analog circuits. Alternatively, we can think of the inverter as *restoring* the signal levels to standard values (V_{OL} and V_{OH}) even when it is presented with corrupted input signal levels (within the noise margins). As a summary, useful for future reference, we present a listing and definitions of the important parameters of the inverter VTC in Table 13.1.

The formal definitions of the threshold voltages V_{IL} and V_{IH} are given in Fig. 13.5. Observe that V_{IL} and V_{IH} are defined as the VTC points at which the slope is –1 V/V. As v_I exceeds V_{IL}, the magnitude of the inverter gain increases and the VTC enters its transition region. Similarly, as v_I falls below V_{IH}, the inverter enters the transition region and the magnitude of the gain increases. Finally, note that Fig. 13.5 shows the definition of another important point on the VTC; this is point M at which $v_O = v_I$. Point M is loosely considered to be the midpoint of the VTC and thus the point at which the *inverter switches from one state to the other*. Point M plays an important role in the definition of the time delay of the inverter, as we shall see shortly.

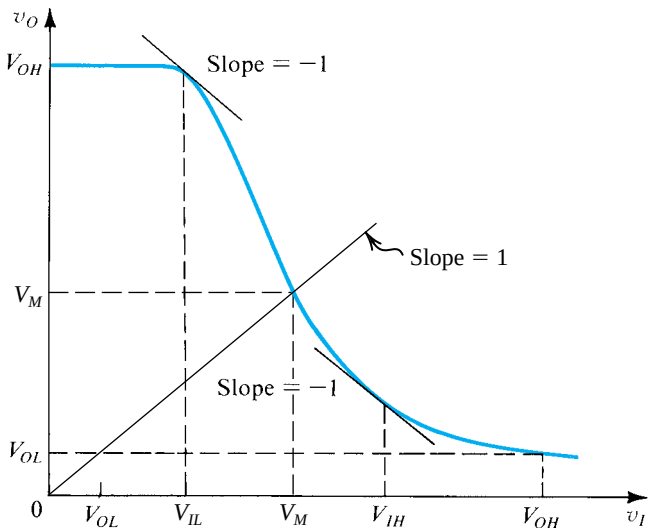


Figure 13.5 Typical voltage transfer characteristic (VTC) of a logic inverter, illustrating the definition of the critical points.

13.1.4 The Ideal VTC

The question naturally arises as to what constitutes an ideal VTC for an inverter. The answer follows directly from the preceding discussion: An ideal VTC is one that maximizes the output signal swing and the noise margins. For an inverter operated from a power supply V_{DD}, maximum signal swing is obtained when

$$V_{OH} = V_{DD}$$

and

$$V_{OL} = 0$$

To obtain maximum noise margins, we first arrange for the transition region to be made as narrow as possible and ideally of zero width. Then, the two noise margins are equalized by arranging for the transition from high to low to occur at the midpoint of the power supply, that is, at $V_{DD}/2$. The result is the VTC shown in Fig. 13.6, for which

$$V_{IL} = V_{IH} = V_M = V_{DD}/2$$

Observe that the sharp transition at $V_{DD}/2$ indicates that if the inverter were to be used as an amplifier, its gain would be infinite. Again, we point out that while the analog designer's interest would be focused on the transition region of the VTC, the digital designer would prefer the transition region to be as narrow as possible, as is the case in the ideal VTC of Fig. 13.6. Finally, we will see in Section 13.2 that inverters implemented using CMOS technology come very close to realizing the ideal VTC

13.1.5 Inverter Implementation

Inverters are implemented using transistors (Chapters 5 and 6) operating as **voltage-controlled switches**. The simplest inverter implementation. is shown in Fig. 13.7(a). The switch is

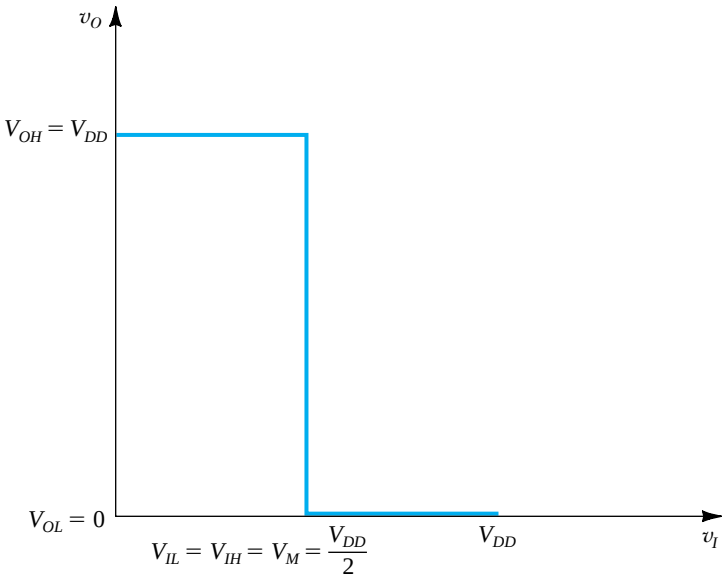


Figure 13.6 The VTC of an ideal inverter.

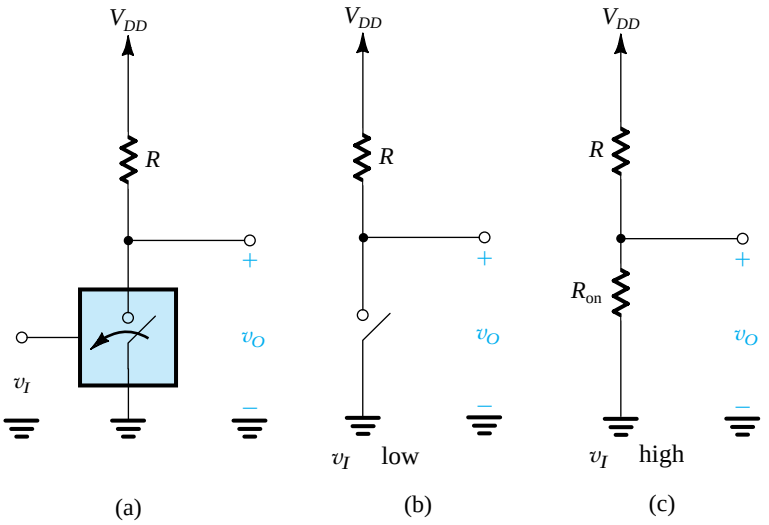


Figure 13.7 **(a)** The simplest implementation of a logic inverter using a voltage-controlled switch; **(b)** equivalent circuit when v_I is low; **(c)** equivalent circuit when v_I is high. Note that the switch is assumed to close when v_I is high.

controlled by the inverter input voltage v_I: When v_I is low, the switch will be open and $v_O = V_{DD}$, since no current flows through R. When v_I is high, the switch will be closed and, assuming an ideal switch, v_O will be 0.

Transistor switches, however, as we know from Chapters 5 and 6, are not perfect. Although their **off resistances** are very high and thus an open switch closely approximates an open circuit, the "on" switch has a finite closure or **"on" resistance**, R_{on}. The result is that when v_I is high, the inverter has the equivalent circuit shown in Fig. 13.7(c), from which V_{OL} can be found.[1]

$$V_{OL} = V_{DD} \frac{R_{on}}{R + R_{on}}$$

We observe that the circuit in Fig. 13.2(a) is a direct implementation of the inverter in Fig. 13.7. In this case, R_{on} is equal to r_{DS} of the MOSFET evaluated at its operating point in the triode region with $V_{GS} = V_{DD}$.

EXERCISE

D13.1 Design the inverter in Fig. 13.2(a) to provide $V_{OL} = 0.1$ V and to draw a supply current of 50 μA in the low-output state. Let the transistor be specified to have $V_t = 0.5$ V, $\mu_n C_{ox} = 125\ \mu A/V^2$, and $\lambda = 0$. The power supply $V_{DD} = 2.5$ V. Specify the required values of W/L and R_D. How much power is drawn from V_{DD} when the switch is open? Closed?
Hint: Recall that for small v_{DS},

[1] If a BJT is used to implement the switch in Fig. 13.7(a), its equivalent circuit in the closed position includes in addition to the resistance $R_{on} = R_{CEsat}$, an offset voltage of about 50 mV to 100 mV (see Fig. 6.19c). We shall not pursue this subject any further here, since the relatively long delay time needed to turn off a saturated BJT has caused the use of BJT switches operated in saturation to all but disappear from the digital IC world.

$$r_{DS} \simeq 1 \Big/ \left[(\mu_n C_{ox}) \left(\frac{W}{L} \right) (V_{GS} - V_t) \right]$$

Ans. 2; 48 kΩ; 0; 125 μW

More elaborate implementations of the logic inverter exist, and we show two of these in Fig. 13.8(a) and 13.9. The circuit in Fig. 13.8(a) utilizes a pair of **complementary switches**, the **"pull-up" (PU) switch** connects the output node to V_{DD}, and the **"pull-down" (PD) switch** connects the output node to ground. When v_I is low, the PU switch will be closed and the PD switch open, resulting in the equivalent circuit of Fig. 13.8(b). Observe that in this case R_{on} of PU connects the output to V_{DD}, thus establishing $V_{OH} = V_{DD}$. Also observe that no current flows and thus no power is dissipated in the circuit. Next, if v_I is raised to the logic-1 level, the PU switch will open while the PD switch will close, resulting in the equivalent circuit shown in Fig. 13.8(c). Here R_{on} of the PD switch connects the output to ground, thus establishing $V_{OL} = 0$. Here again no current flows, and no power is dissipated. The superiority of this inverter implementation over that using the single pull-down switch and a resistor (known as a **pull-up resistor**) should be obvious: With $V_{OL} = 0$ and $V_{OH} = V_{DD}$, the signal swing is at its maximum possible, and the power dissipation is zero in both states. This circuit constitutes the basis of the CMOS inverter that we will study in Section 13.3.

Finally, consider the inverter implementation of Fig. 13.9. Here a double-throw switch is used to steer the constant current I_{EE} into one of two resistors connected to the positive supply V_{CC}. The reader is urged to show that if a high v_I results in the switch being connected to R_{C1}, then a logic inversion function is realized at v_{O1}. Note that the output voltage is independent of the switch resistance. This *current-steering* or *current-mode* logic arrangement is the basis of the fastest available digital logic circuits, called emitter-coupled logic (ECL), which we shall study in Section 14.4. In fact, ECL is the only BJT logic-circuit type that is currently employed in new designs and the only one studied in this book.

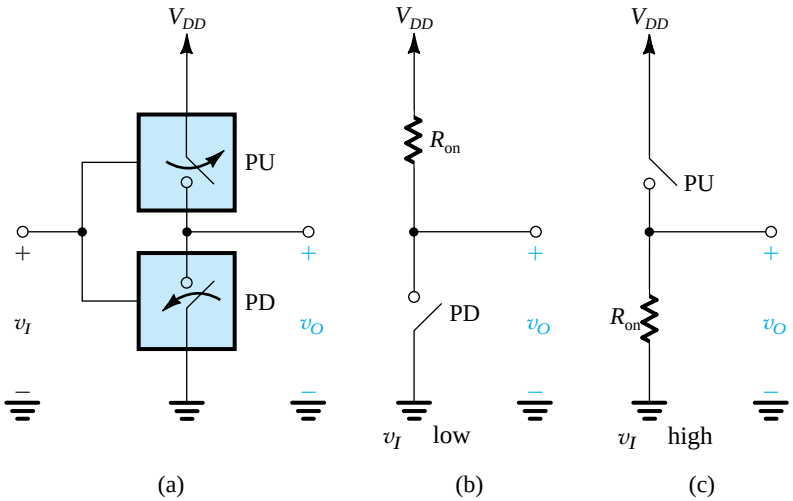


Figure 13.8 A more elaborate implementation of the logic inverter utilizing two complementary switches. This is the basis of the CMOS inverter that we shall study in Section 13.2.

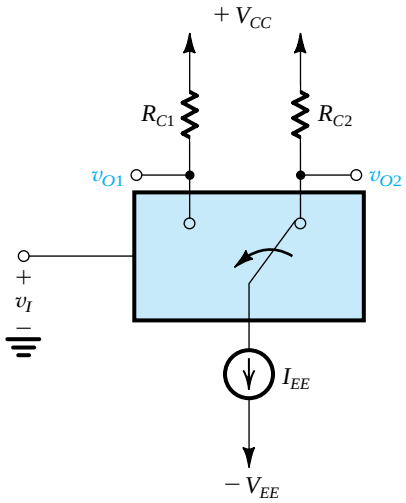


Figure 13.9 Another inverter implementation utilizing a double-throw switch to steer the constant current I_{EE} to R_{C1} (when v_I is high) or R_{C2} (when v_I is low). This is the basis of the emitter-coupled logic (ECL) studied in Chapter 14.

EXERCISE

13.2 For the current-steering circuit in Fig. 13.9, let $V_{CC} = 5$ V, $I_{EE} = 1$ mA, and $R_{C1} = R_{C2} = 2\ \text{k}\Omega$. What are the high and low logic levels obtained at the outputs?
Ans. $V_{OH} = 5$ V; $V_{OL} = 3$ V

Example 13.1 Resistively Loaded MOS Inverter

For the simple MOS inverter in Fig. 13.2(a):

(a) Derive expressions for V_{OH}, V_{OL}, V_{IL}, V_{IH}, and V_M. For simplicity, neglect channel-length modulation (i.e., assume $\lambda = 0$). Show that these inverter parameters can be expressed in terms of V_{DD}, V_t, and $(k_n R_D)$. The latter parameter has the dimension of V^{-1}; and to simplify the expressions, denote $k_n R_D \equiv 1/V_x$.

(b) Show that V_x can be used as a design parameter for the inverter circuit. In particular, find the value of V_x that results in $V_M = V_{DD}/2$.

(c) Find numerical values for all parameters and for the inverter noise margins for $V_{DD} = 1.8$ V, $V_t = 0.5$ V, and V_x set to the value found in (b).

(d) For $k_n' = 300\ \mu\text{A}/\text{V}^2$ and $W/L = 1.5$, find the required value of R_D and use it to determine the average power dissipated in the inverter, assuming that the inverter spends half of the time in each of its two states.

(e) Comment on the characteristics of this inverter circuit vis-à-vis the ideal characteristics as well as on its suitability for implementation in integrated-circuit form.

Example 13.1 *continued*

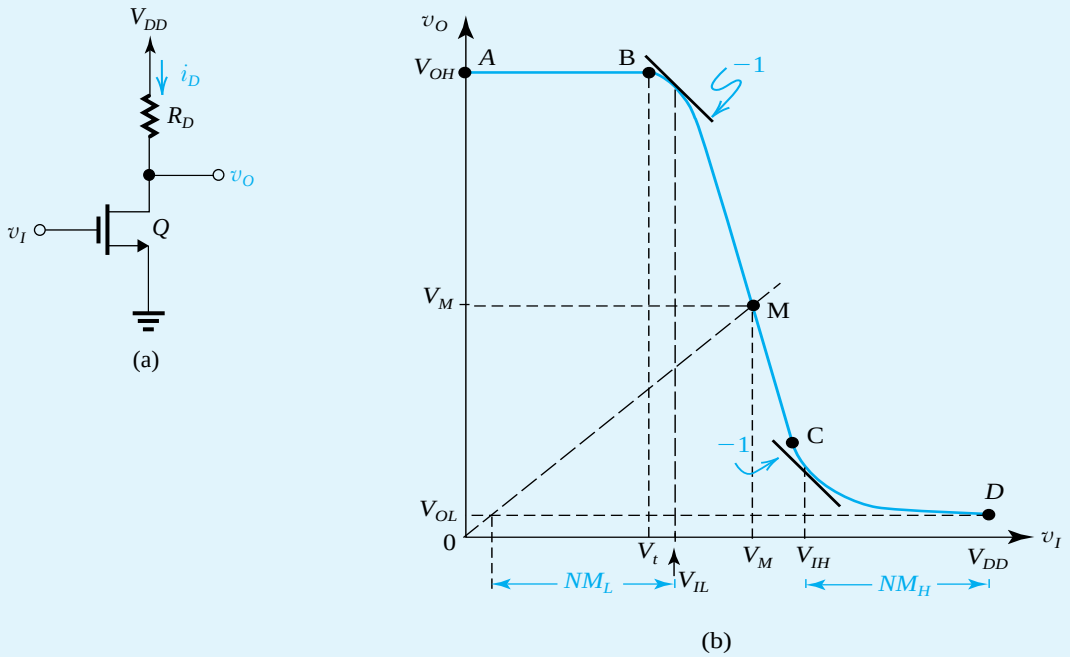


Figure 13.10 The resistively loaded MOS inverter and its VTC (Example 13.1).

Solution

(a) Refer to Fig. 13.10. For $v_I < V_t$, the MOSFET is off, $i_D = 0$, and $v_O = V_{DD}$. Thus

$$V_{OH} = V_{DD} \tag{13.4}$$

As v_I exceeds V_t, the MOSFET turns on and operates initially in the saturation region. Assuming $\lambda = 0$,

$$i_D = \frac{1}{2}k_n(v_I - V_t)^2 R_D$$

and

$$v_O = V_{DD} - R_D i_D = V_{DD} - \frac{1}{2}k_n R_D (v_I - V_t)^2$$

substituting $k_n R_D = 1/V_x$, the BC segment of the VTC is described by

$$v_O = V_{DD} - \frac{1}{2V_x}(v_I - V_t)^2 \tag{13.5}$$

To determine V_{IL}, we differentiate Eq. (13.5) and set $dv_O/dv_I = -1$,

$$\frac{dv_O}{dv_I} = -\frac{1}{V_x}(v_I - V_t)$$

$$-1 = -\frac{1}{V_x}(V_{IL} - V_t)$$

which results in

$$V_{IL} = V_t + V_x \tag{13.6}$$

To determine the coordinates of the midpoint M, we substitute $v_O = v_I = V_M$ in Eq. (13.5), thus obtaining

$$V_{DD} - V_M = \frac{1}{2V_x}(V_M - V_t)^2 \tag{13.7}$$

which can be solved to obtain

$$V_M = V_t + \sqrt{2(V_{DD}-V_t)V_x + V_x^2} - V_x \tag{13.8}$$

The boundary of the saturation-region segment BC, point C, is determined by substituting $v_O = v_I - V_t$ in Eq. (13.5) and solving for v_O to obtain

$$V_{OC} = \sqrt{2V_{DD}V_x + V_x^2} - V_x \tag{13.9}$$

and

$$V_{IC} = V_t + \sqrt{2V_{DD}V_x + V_x^2} - V_x \tag{13.10}$$

Beyond point C, the transistor operates in the triode region, thus

$$i_D = k_n\left[(v_I - V_t)v_O - \frac{1}{2}v_O^2\right]$$

and the output voltage is obtained as

$$v_O = V_{DD} - \frac{1}{V_x}\left[(v_I - V_t)v_O - \frac{1}{2}v_O^2\right] \tag{13.11}$$

which describes the segment CD of the VTC. To determine V_{IH}, we differentiate Eq. (13.11) and set $dv_O/dv_I = -1$:

$$\frac{dv_O}{dv_I} = -\left(\frac{1}{V_x}\right)\left[(v_I - V_t)\frac{dv_O}{dv_I} + v_O - v_O\ \frac{dv_O}{dv_I}\right]$$

$$-1 = -\frac{1}{V_x}[-(V_{IH} - V_t) + 2v_O]$$

which results in

$$V_{IH} - V_t = 2v_O - V_x \tag{13.12}$$

Substituting in Eq. (13.11) for v_I with the value of V_{IH} from Eq. (13.12) results in an equation in the value of v_O corresponding to $v_I = V_{IH}$, which can be solved to yield

$$v_O\big|_{v_I = V_{IH}} = 0.816\sqrt{V_{DD}V_x} \tag{13.13}$$

which can be substituted in Eq. (13.12) to obtain

$$V_{IH} = V_t + 1.63\sqrt{V_{DD}V_x} - V_x \tag{13.14}$$

Example 13.1 *continued*

To determine V_{OL} we substitute $v_I = V_{OH} = V_{DD}$ in Eq. (13.11):

$$V_{OL} = V_{DD} - \frac{1}{V_x}\left[(V_{DD} - V_t)V_{OL} - \frac{1}{2}V_{OL}^2\right] \tag{13.15}$$

Since we expect V_{OL} to be much smaller than $2(V_{DD} - V_t)$, we can approximate Eq. (13.15) as

$$V_{OL} \simeq V_{DD} - \frac{1}{V_x}(V_{DD} - V_t)V_{OL}$$

which results in

$$V_{OL} = \frac{V_{DD}}{1 + [(V_{DD} - V_t)/V_x]} \tag{13.16}$$

It is interesting to note that the value of V_{OL} can alternatively be found by noting that at point D, the MOSFET switch has a closure resistance r_{DS},

$$r_{DS} = \frac{1}{k_n(V_{DD} - V_t)} \tag{13.17}$$

and V_{OL} can be obtained from the voltage divider formed by R_D and r_{DS},

$$V_{OL} = V_{DD}\frac{r_{DS}}{R_D + r_{DS}} = \frac{V_{DD}}{1 + R_D/r_{DS}} \tag{13.18}$$

Substituting for r_{DS} from Eq. (13.17) gives an expression for V_{OL} identical to that in Eq. (13.16).

(b) We observe that all the inverter parameters derived above are functions of V_{DD}, V_t, and V_x only. Since V_{DD} and V_t are determined by the process technology, the only design parameter available is $V_x \equiv 1/k_nR_D$. To place V_M at half the supply voltage V_{DD}, we substitute $V_M = V_{DD}/2$ in Eq. (13.7) to obtain the value V_x must have as

$$V_x\big|_{V_M = V_{DD}/2} = \frac{(V_{DD}/2 - V_t)^2}{V_{DD}} \tag{13.19}$$

(c) For $V_{DD} = 1.8$ V and $V_t = 0.5$, we use Eq. (13.19) to obtain

$$V_x\big|_{V_M = 0.9\text{ V}} = \frac{(1.8/2 - 0.5)^2}{1.8} = 0.089\text{ V}$$

From Eq. (13.4): $V_{OH} = 1.8$ V

From Eq. (13.16): $V_{OL} = 0.12$ V

From Eq. (13.6): $V_{IL} = 0.59$ V

From Eq. (13.14): $V_{IH} = 1.06$ V

$$NM_L = V_{IL} - V_{OL} = 0.47\text{ V}$$

$$NM_H = V_{OH} - V_{IH} = 0.74\text{ V}$$

(d) To determine R_D, we use

$$k_nR_D = \frac{1}{V_x} = \frac{1}{0.089} = 11.24$$

Thus,

$$R_D = \frac{11.24}{k_n'(W/L)} = \frac{11.24}{300 \times 10^{-6} \times 1.5} = 25\ \text{k}\Omega$$

The inverter dissipates power only when the output is low, in which case the current drawn from the supply is

$$I_{DD} = \frac{V_{DD} - V_{OL}}{R_D} = \frac{1.8 - 0.12}{25\ \text{k}\Omega} = 67\ \mu\text{A}$$

and the power drawn from the supply during the low-output interval is

$$P_D = V_{DD} I_{DD} = 1.8 \times 67 = 121\ \mu\text{W}$$

Since the inverter spends half of the time in this state,

$$P_{D\text{average}} = \frac{1}{2} P_D = 60.5\ \mu\text{W}$$

(e) We now can make a few comments on the characteristics of this inverter circuit in comparison to the ideal characteristics:

1. The output signal swing, though not equal to the full power supply, is reasonably good: $V_{OH} = 1.8$ V, $V_{OL} = 0.12$ V.
2. The noise margins, though of reasonable values, are far from the optimum value of $V_{DD}/2$. This is particularly the case for NM_L.
3. Most seriously, the gate dissipates a relatively large amount of power. To appreciate this point, consider an IC chip with a million inverters (a small number by today's standards): Its power dissipation will be 61 W. This is too large, especially given that this is "static power," unrelated to the switching activity of the gates (more on this later).

We consider this inverter implementation to be entirely unsuitable for IC fabrication because each inverter requires a load resistance of 25 kΩ, a value that needs a large chip area (see Appendix A). To overcome this problem, we investigate in Example 13.2 the replacement of the passive resistance R_D with an NMOS transistor.

EXERCISES

D13.3 In an attempt to reduce the required value of R_D, to 10 kΩ, the designer of the inverter in Example 13.1 decides to keep the parameter V_x unchanged but increases W/L. What is the new value required for W/L? Do the noise margins change? What does the power dissipation become?
Ans. 3.75; no; 151 μW

D13.4 In an attempt to reduce the required value of R_D to 10 kΩ, the designer of the inverter in Examples 13.1 decides to change V_x while keeping W/L unchanged. What new value of V_x is needed? What do the noise margins become? What does the power dissipation become?
Ans. $V_x = 0.22$ V; $NM_L = 0.46$ V, $NM_H = 0.49$ V; 139 μW

Example 13.2 The Saturated NMOS-Load Inverter

To overcome the problem associated with the need for a large resistance R_D in the circuit of Fig. 13.10(a), studied in Example 13.1, R_D can be replaced by a MOSFET. One such possibility is the circuit shown in Fig. 13.11(a), where the load is an NMOS transistor Q_2 operated in the saturation region (by connecting its drain to its gate). Although not shown on the diagram, the body terminal of Q_2 is connected to the lowest-voltage node, which is ground.

(a) Neglecting the body effect in Q_2 and assuming $\lambda_1 = \lambda_2 = 0$, determine the inverter parameters V_{OH}, V_{OL}, V_{IL}, V_{IH}, and V_M. Express the results in terms of V_{DD}, V_t (where $V_{t1} = V_{t2} = V_t$), and $k_r \equiv \sqrt{k_{n1}/k_{n2}}$.

(b) For $V_{DD} = 1.8$ V, $V_t = 0.5$ V, $(W/L)_1 = 5$, and $(W/L)_2 = \frac{1}{5}$, find numerical values for all parameters and for the noise margins.

(c) If $k_n' = 300\ \mu\text{A/V}^2$, find the average power dissipated in the inverter, assuming that it spends half the time in each of its two states.

(d) Qualitatively describe how the body effect in Q_2 affects the noise margins.

(e) Comment on the characteristics of this inverter implementation vis-à-vis the ideal characteristics. How suitable is this circuit for implementation in IC form?

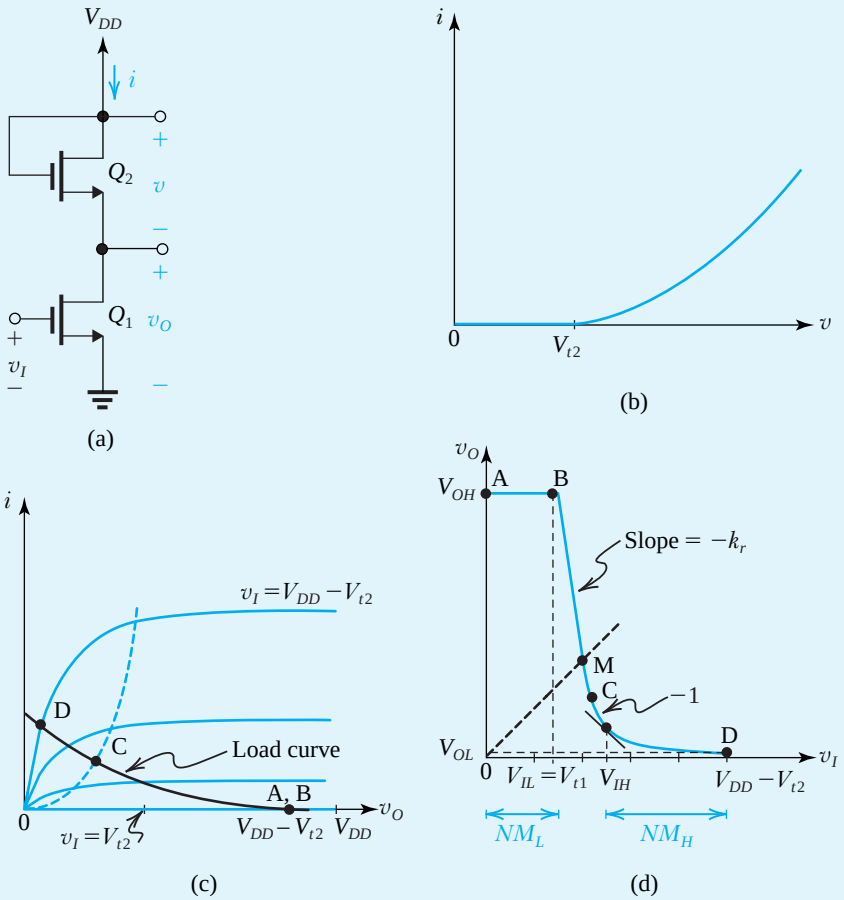


Figure 13.11 **(a)** Enhancement-load MOS inverter; **(b)** load curve; **(c)** construction to determine VTC; **(d)** the VTC.

Solution

(a) The inverter VTC can be determined graphically by superimposing the load curve, which is the i–v characteristic of the diode-connected transistor Q_2, shown in Fig. 13.11(b), on the i–v characteristics of Q_1. As we have done in the graphical analysis of MOSFET circuits in Section 5.4, we shift the load curve horizontally by V_{DD} and flip it around the vertical axis, as shown in Fig. 13.11(c). The resulting VTC is shown in Fig. 13.11(d).

For $v_I < V_{t1}$, Q_1 will be off, which forces the current in Q_2 to be zero. Transistor Q_2, although it will be conducting a zero current, will have a voltage drop of V_{t2}. This is a result of its i–v characteristic shown in Fig. 13.11(b). Thus the output voltage v_O will not reach V_{DD} but will be at $V_{DD} - V_{t2}$, that is,[2]

$$V_{OH} = V_{DD} - V_t \tag{13.20}$$

As v_I exceeds V_{t1}, Q_1 turns on and initially operates in saturation, thus

$$i_{D1} = \frac{1}{2}k_{n1}(v_I - V_{t1})^2$$

Since Q_2 operates in saturation at all times,

$$i_{D2} = \frac{1}{2}k_{n2}(V_{DD} - v_O - V_{t2})^2$$

Equating i_{D1} and i_{D2} and substituting $V_{t1} = V_{t2} = V_t$, and $\sqrt{k_{n1}/k_{n2}} = k_r$, gives

$$v_O = V_{DD} + (k_r - 1)V_t - k_r v_I \tag{13.21}$$

which is the equation for segment BC of the VTC in Fig. 13.11(d). It is interesting to observe that the relationship between v_O and v_I is linear and that the slope of this straight line is $-k_r$.

Since the slope of the VTC changes from zero to $-k_r$ at point B, it is reasonable to consider point B to be the determinant of V_{IL}; thus,

$$V_{IL} = V_{t1} = V_t \tag{13.22}$$

To obtain V_M we substitute $v_I = v_O = V_M$ in Eq. (13.21); thus,

$$V_M = \frac{V_{DD} + (k_r - 1)V_t}{k_r + 1} \tag{13.23}$$

We next determine the coordinates of point C at which Q_1 enters the triode region by substituting in Eq. (13.21) $v_O = v_I - V_t$. The result is

$$V_{IC} = \frac{V_{DD} + k_r V_t}{k_r + 1} \tag{13.24}$$

[2]To see this point more clearly, consider the usual situation of a capacitance C_L between the output node of the inverter and ground. Assume that initially v_I was high and v_O was low. Now let v_I go low. Q_1 cuts off, and Q_2 provides a current that charges C_L up. As v_O increases, the current provided by Q_2 decreases until v_O reaches $V_{DD} - V_{t2}$, at which point the current supplied by Q_2 reaches zero. Thus the charging process terminates and v_O stabilizes at $V_{DD} - V_{t2}$.

Example 13.2 *continued*

and

$$V_{OC} = \frac{V_{DD} - V_t}{k_r + 1} \tag{13.25}$$

Comparing Eqs. (13.24) and (13.23), we make the comforting observation that $V_{IC} > V_M$, confirming our implicit assumption that M lies on the linear segment of the VTC.

For $v_I > V_{IC}$, Q_1 operates in the triode region; thus,

$$i_{D1} = k_{n1}\left[(v_I - V_{t1})v_O - \frac{1}{2}v_O^2\right]$$

Meanwhile, Q_2 still operates in saturation. Equating their currents results in

$$2k_r^2\left[(v_I - V_t)v_O - \frac{1}{2}v_O^2\right] = (V_{DD} - V_t - v_O)^2 \tag{13.26}$$

Although this equation can be used to determine V_{IH}, the effort involved to do this symbolically is too great. We will instead find V_{IH} numerically; V_{OL}, however, can be determined by substituting in Eq. (13.26) $v_I = V_{OH} = V_{DD} - V_t$ and $v_O = V_{OL}$,

$$2k_r^2\left[(V_{DD} - 2V_t)V_{OL} - \frac{1}{2}V_{OL}^2\right] = (V_{DD} - V_t - V_{OL})^2 \tag{13.27}$$

Since we expect V_{OL} to be much smaller than $2(V_{DD} - 2V_t)$ and $(V_{DD} - V_t)$, we can approximate Eq. (13.27) as follows:

$$2k_r^2(V_{DD} - 2V_t)V_{OL} \simeq (V_{DD} - V_t)^2$$

Thus,

$$V_{OL} \simeq \frac{(V_{DD} - V_t)^2}{2k_r^2(V_{DD} - 2V_t)} \tag{13.28}$$

We observe that all the inverter parameters are functions of three quantities only: V_{DD}, V_t, and k_r. Since the first two are determined by the process technology, the only design parameter is k_r, which determines the steepness of the transition region.

(b) Given $V_{DD} = 1.8$ V, $V_t = 0.5$ V, $(W/L)_1 = 5$, and $(W/L)_2 = \frac{1}{5}$, we first determine k_r as

$$k_r = \sqrt{\frac{k_{n1}}{k_{n2}}} = \sqrt{\frac{(W/L)_1}{(W/L)_2}} = \sqrt{\frac{5}{1/5}} = 5$$

From Eq. (13.20): $V_{OH} = 1.3$ V

From Eq. (13.28): $V_{OL} = 0.04$ V

From Eq. (13.22): $V_{IL} = 0.5$ V

From Eq. (13.23): $V_M = 0.63$ V

To determine V_{IH} we utilize Eq. (13.26) together with setting $dv_O/dv_I = -1$. The result is

$$V_{IH} = 0.75 \text{ V}$$

Thus,

$$NM_L = V_{IL} - V_{OL} = 0.5 - 0.04 = 0.46 \text{ V}$$

$$NM_H = V_{OH} - V_{IH} = 1.3 - 0.75 = 0.55 \text{ V}$$

(c) The inverter dissipates power only when $v_O = V_{OL}$. In this case, the current drawn from the supply is

$$I_{DD} = i_{D2} = \frac{1}{2}k_{n2}(V_{DD} - V_{OL} - V_t)^2$$

Thus,

$$I_{DD} = \frac{1}{2} \times 300 \times \frac{1}{5} \times (1.8 - 0.04 - 0.5)^2$$
$$= 47.6 \ \mu\text{A}$$

and,

$$P_D = V_{DD} I_{DD} = 1.8 \times 47.6 = 85.7 \ \mu\text{W}$$

Since the inverter is in the low-output state for half the time,

$$P_{D\text{average}} = \frac{1}{2} \times 85.7 = 42.9 \ \mu\text{W}$$

(d) Since the body of Q_2 is connected to ground, its source-to-body voltage V_{SB} is

$$V_{SB} = v_O$$

Now, since the threshold voltage is given by

$$V_{t2} = V_{t0} + \gamma[\sqrt{V_{SB} + 2\phi_f} - \sqrt{2\phi_f}] \tag{13.29}$$

we see that V_{t2} will increase with v_O. This is of immense concern, since V_{t2} will be at its largest value for $v_O = V_{OH} = V_{DD} - V_{t2}$. Thus, V_{OH} will be lower than the value calculated above. This reduces the output signal swing and NM_H.

(e) We now can make the following comments on the characteristics of this inverter implementation:

1. The fact that V_{OH} is lower than V_{DD} by V_{t2} and that V_{t2} can be large because of the body effect imposes a major disadvantage on this NMOS-load inverter.
2. The noise margins are much lower than the ideal values of $V_{DD}/2$. Also, V_M is far from the power-supply midpoint.
3. The sharpness of the transition of the VTC increases with the value of k_r. Increasing k_r, however, has the effect of increasing the silicon area (see Exercise 13.6).
4. Like the resistively-loaded MOS inverter considered in Example 13.1, the NMOS-loaded inverter dissipates a large amount of power.

Since the circuit utilizes NMOS transistors exclusively, it is certainly suitable for implementation in IC form. As we will discuss shortly, all-NMOS technology was at one time (1970s) the technology of choice for the implementation of microprocessor chips. Its high power dissipation, however, has caused its demise in favor of CMOS technology.

EXERCISES

13.5 Repeat part (b) of Example 13.3 for the case $(W/L)_1 = 3$ and $(W/L)_2 = \frac{1}{3}$. Specifically, find the values of V_{OH}, V_{OL}, V_{IL}, V_{IH}, V_M, NM_H, and NM_L.
Ans. 1.3 V; 0.12 V; 0.5 V; 0.87 V; 0.7 V; 0.43 V; 0.38 V

13.6 Consider the inverter in Fig. 13.11(a) with $(W/L)_1 = k_r$ and $(W/L)_2 = 1/k_r$. Show that if the minimum dimension (i.e. length or width) of each of the two transistors is denoted d, the inverter silicon area is $2k_r d^2$.

13.1.6 Power Dissipation

Digital systems are implemented using very large numbers of logic gates. For space and other economic considerations, it is desirable to implement the system with as few integrated-circuit (IC) chips as possible. It follows that one must pack as many logic gates as possible on an IC chip. At present, one million gates or more can be fabricated on a single IC chip in what is known as **very-large-scale integration (VLSI)**. To keep the power dissipated in the chip to acceptable limits (imposed by thermal considerations), the power dissipation per gate must be kept to a minimum. Indeed, a very important performance measure of the logic inverter is the power it dissipates.

The inverter of Fig. 13.7 dissipates no power when v_I is low and the switch is open. In the other state, however, the power dissipation is approximately V_{DD}^2/R and can be substantial, as we have seen in Examples 13.1 and 13.2. This power dissipation occurs even if the inverter is not switching and is thus known as **static power dissipation**.

The inverter of Fig. 13.8 exhibits no static power dissipation, a definite advantage. Unfortunately, however, another component of power dissipation arises when a capacitance exists between the output node of the inverter and ground. This is always the case, for the devices that implement the switches have internal capacitances, the wires that connect the inverter output to other circuits have capacitance, and, of course, there is the input capacitance of whatever circuit the inverter is driving. Now, as the inverter is switched from one state to another, current must flow through the switch(es) to charge (and discharge) the load capacitance. These currents give rise to power dissipation in the switches, called **dynamic power dissipation**.

An expression for the dynamic power dissipation of the inverter of Fig. 13.8 can be derived as follows. Consider first the situation when v_I goes low. The pull-down switch P_D

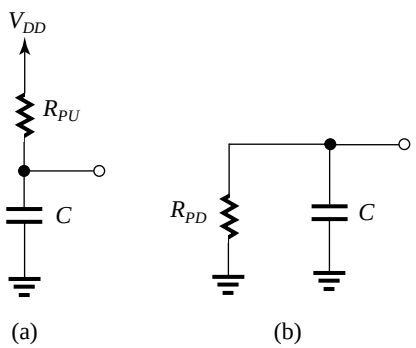


Figure 13.12 Equivalent circuits for calculating the dynamic power dissipation of the inverter in Figure 13.8: **(a)** When v_I is low; **(b)** When v_I is high.

turns off and the pull-up switch P_U turns on. In this state, the inverter can be represented by the equivalent circuit shown in Fig. 13.12(a). Capacitor C will charge through the on-resistance of the pull-up switch, and the voltage across C will increase from 0 to V_{DD}. Denoting by $i_D(t)$ the charging current supplied by V_{DD}, we can write for the instantaneous power drawn from V_{DD} the expression

$$p_{DD}(t) = V_{DD} i_D(t)$$

The energy delivered by the power supply to charge the capacitor can be determined by integrating $p_{DD}(t)$ over the charging interval T_c,

$$\begin{aligned} E_{DD} &= \int_0^{T_c} V_{DD} i_D(t)\, dt \\ &= V_{DD} \int_0^{T_c} i_D(t)\, dt \\ &= V_{DD} Q \end{aligned}$$

where Q is the charge delivered to the capacitor during the charging interval. Since the initial charge on C was zero,

$$Q = CV_{DD}$$

Thus,

$$E_{DD} = CV_{DD}^2 \tag{13.30}$$

Since at the end of the charging process the energy stored on the capacitor is

$$E_{\text{stored}} = \frac{1}{2} CV_{DD}^2 \tag{13.31}$$

we can find the energy dissipated in the pull-up switch as

$$E_{\text{dissipated}} = E_{DD} - E_{\text{stored}} = \frac{1}{2} CV_{DD}^2 \tag{13.32}$$

This energy is dissipated in the on-resistance of switch P_U and is converted to heat.

Next consider the situation when v_I goes high. The pull-up switch P_U turns off and the pull-down switch P_D turns on. The equivalent circuit in this case is that shown in Fig. 13.12(b). Capacitor C is discharged through the on-resistance of the pull-down switch, and its voltage changes from V_{DD} to 0. At the end of the discharge interval, there will be no energy left on the capacitor. Thus all of the energy initially stored on the capacitor, $\frac{1}{2}CV_{DD}^2$, will be dissipated in the pull-down switch,

$$E_{\text{dissipated}} = \frac{1}{2} CV_{DD}^2 \tag{13.33}$$

This amount of energy is dissipated in the on-resistance of switch P_D and is converted to heat.

Thus in each cycle of inverter switching, an amount of energy of $\frac{1}{2}CV_{DD}^2$ is dissipated in the pull-up switch and $\frac{1}{2}CV_{DD}^2$ is dissipated in the pull-down switch, for a total energy loss per cycle of

$$E_{\text{dissipated}}/\text{cycle} = CV_{DD}^2 \tag{13.34}$$

If the inverter is switched at a frequency of f Hz, the dynamic power dissipation of the inverter will be

$$P_{\text{dyn}} = fCV_{DD}^2 \tag{13.35}$$

This is a general expression that does not depend on the inverter circuit details or the values of the on-resistance of the switches.

The expression in Eq. (13.35) indicates that to minimize the dynamic power dissipation, one must strive to reduce the value of C. However, in many cases C is largely determined by the transistors of the inverter itself and cannot be substantially reduced. Another important factor in determining the dynamic power dissipation is the power-supply voltage V_{DD}. Reducing V_{DD}, reduces P_{dyn} significantly. This has been a major motivating factor behind the reduction of V_{DD} with every technology generation (see Table 7.A.1). Thus, while the 0.5-μm CMOS process utilized a 5-V power supply, the power-supply voltage used with the 0.13-μm process is only 1.2 V.

Finally, since P_{dyn} is proportional to the operating frequency f, one may be tempted to reduce P_{dyn} by reducing f. This, however, is not a viable proposition in light of the desire to operate digital systems at increasingly higher speeds. This point will be discussed next.

EXERCISES

13.7 Find the dynamic power dissipation of an inverter operated from a 1.8-V supply and having a load capacitance of 100 fF. Let the inverter be switched at 100 MHz.
Ans. 32.4 μW

13.8 A particular inverter circuit initially designed in a 0.5-μm process is fabricated in a 0.13-μm process. Assuming that the capacitance C will scale down in proportion to the minimum feature size (more on this later) and that the power supply will be reduced from 5 V to 1.2 V, by what factor do you expect the dynamic power dissipation to decrease? Assume that the switching frequency f remains unchanged.
Ans. 66.8

13.1.7 Propagation Delay

A very important measure of the performance of a digital system, such as a computer, is the maximum speed at which it is capable of operating. Although many factors come into play in determining the operating speed of a system, a core factor is the speed of operation of the basic logic inverter utilized in its implementation. This in turn is characterized by the time it takes the inverter to respond to a change at its input. To be more precise, consider an inverter fed with the ideal pulse shown in Fig. 13.13(a). The resulting output signal of the inverter is shown in Fig. 13.13(b). We make the following two observations.

1. The output signal is no longer an ideal pulse. Rather, it has rounded edges; that is, the pulse takes some time to fall to its low value and to rise to its high value. We speak of this as the pulse having finite fall and rise times. We will provide a precise definition of these shortly.
2. There is a time delay between each edge of the input pulse and the corresponding change in the output of the inverter. If we define the "switching point" of the output as the time at

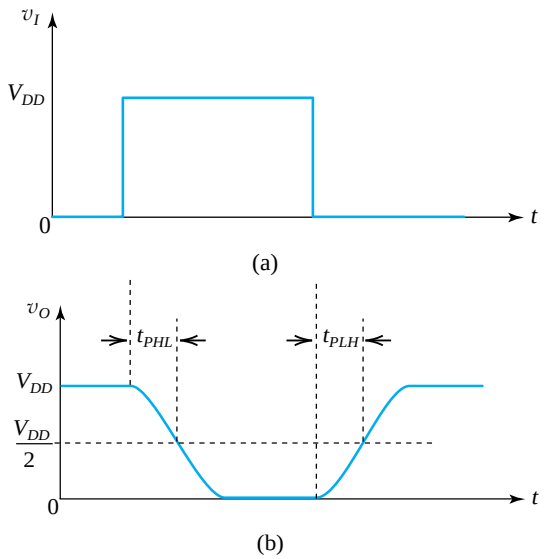


Figure 13.13 An inverter fed with the ideal pulse in (**a**) provides at its output the pulse in (**b**). Two delay times are defined as indicated.

which the output pulse passes through the half-point of its excursion, then we can define the propagation delays of the inverter as indicated in Fig. 13.13(b). Note that there are two propagation delays, which are not necessarily equal: the propagation delay for the output going from high to low, t_{PHL}, and the propagation delay for the output going from low to high, t_{PLH}. The inverter propagation delay t_P is defined as the average of the two,

$$t_P \equiv \frac{1}{2}(t_{PLH} + t_{PHL}) \tag{13.36}$$

Having defined the inverter propagation delay, we now consider the *maximum* switching frequency of the inverter. From Fig. 13.13(b) we can see that the *minimum* period for each cycle is

$$T_{\text{min}} = t_{PHL} + t_{PLH} = 2t_P \tag{13.37}$$

Thus the **maximum switching frequency** is

$$f_{\text{max}} = \frac{1}{T_{\text{min}}} = \frac{1}{2t_P} \tag{13.38}$$

At this point the reader is no doubt wondering about the cause of the finite propagation time of the inverter. It is simply a result of the time needed to charge and discharge the various capacitances in the circuit. These include the MOSFET capacitances, the wiring capacitance, and the input capacitances of all the logic gates driven by the inverter. We will have a lot more to say about these capacitances and about the determination of t_P in later sections. For the time being, however, we make two important points:

1. A fundamental relationship in analyzing the dynamic operation of a circuit is

$$I\Delta t = \Delta Q = C\Delta V \tag{13.39}$$

That is, a current I flowing through a capacitance C for an interval Δt deposits a charge ΔQ on the capacitor, which causes the capacitor voltage to increase by ΔV.

2. A thorough familiarity with the time response of single-time-constant (STC) circuits is of great help in the analysis of the dynamic operation of digital circuits. A review of this subject is presented in Appendix E. For our purposes here, we remind the reader of the key equation in determining the response to a step function:

Consider a step-function input applied to an STC network of either the low-pass or high-pass type, and let the network have a time constant τ. The output at any time t is given by

$$y(t) = Y_\infty - (Y_\infty - Y_{0+})e^{-t/\tau} \tag{13.40}$$

where Y_∞ is the final value, that is, the value toward which the response is heading, and Y_{0+} is the value of the response immediately after $t = 0$. This equation states that the output at any time t is equal to the difference between the final value Y_∞ and a gap whose initial value is $Y_\infty - Y_{0+}$ and that is shrinking exponentially.

Example 13.3

Consider the inverter of Fig. 13.7(a) with a capacitor C connected between the output node and ground. If at $t = 0$, v_I goes low, and assuming that the switch opens instantaneously, find the time for v_O to reach $\frac{1}{2}(V_{OH} + V_{OL})$. This is the low-to-high propagation time, t_{PLH}. Calculate the value of t_{PLH} for the case $R = 25\ \text{k}\Omega$ and $C = 10$ fF.

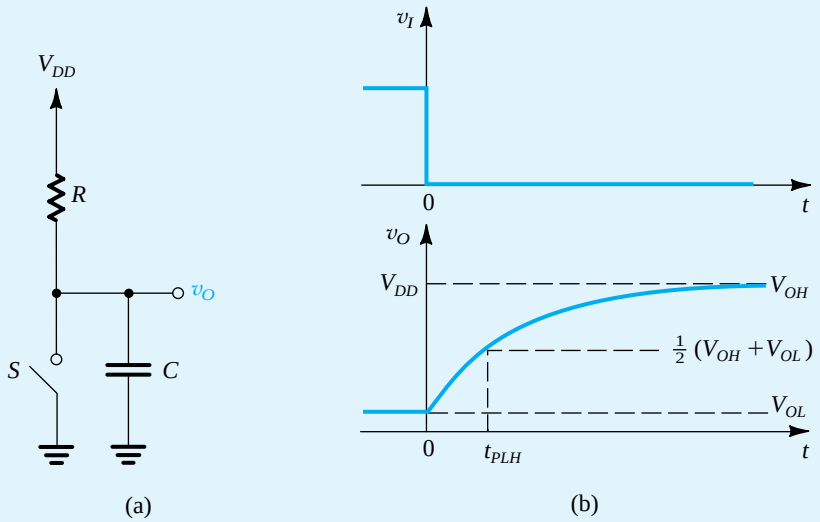


Figure 13.14 Example 13.3: **(a)** The inverter circuit after the switch opens (i.e., for $t \geq 0+$). **(b)** Waveforms of v_I and v_O. Observe that the switch is assumed to operate instantaneously. v_O rises exponentially, starting at V_{OL} and heading toward V_{OH}.

Solution

Before the switch opens, $v_O = V_{OL}$. When the switch opens at $t = 0$, the circuit takes the form shown in Fig. 13.14(a). Since the voltage across the capacitor cannot change instantaneously, at $t = 0+$ the output will still be V_{OL}. Then the capacitor charges through R, and v_O rises exponentially toward V_{DD}. The

output waveform will be as shown in Fig. 13.14(b), and its equation can be obtained by substituting in Eq. (13.39): $v_O(\infty) = V_{OH} = V_{DD}$ and $v_O(0+) = V_{OL}$. Thus,

$$v_O(t) = V_{OH} - (V_{OH} - V_{OL})e^{-t/\tau}$$

where $\tau = CR$. To find t_{PLH}, we substitute

$$v_O(t_{PLH}) = \frac{1}{2}(V_{OH} + V_{OL})$$

Thus,

$$\frac{1}{2}(V_{OH} + V_{OL}) = V_{OH} - (V_{OH} - V_{OL})e^{-t_{PLH}/\tau}$$

which results in

$$t_{PLH} = \tau \ln 2 = 0.69\tau$$

Note that this expression is independent of the values of V_{OL} and V_{OH}. For the numerical values given,

$$\begin{aligned} t_{PLH} &= 0.69RC \\ &= 0.69 \times 25 \times 10^3 \times 10 \times 10^{-15} \\ &= 173 \text{ ps} \end{aligned}$$

EXERCISES

13.9 A capacitor C whose initial voltage is 0 is charged to a voltage V_{DD} by a constant-current source I. Find the time t_{PLH} at which the capacitor voltage reaches $(V_{DD}/2)$. What value of I is required to obtain a 10-ps propagation delay with C = 10 fF and V_{DD} = 1.8 V?
Ans. $t_{PLH} = CV_{DD}/2I$; 0.9 mA

13.10 For the inverter of Fig. 13.8(a), let the on-resistance of P_U be 20 kΩ and that of P_D = 10 kΩ. If the capacitance C = 10 fF, find t_{PLH}, t_{PHL}, and t_P.
Ans. 138 ps; 69 ps; 104 ps

We conclude this section by showing in Fig. 13.15 the formal definition of the propagation delay of an inverter. As shown, an input pulse with finite (nonzero) **rise and fall times** is applied. The inverted pulse at the output exhibits finite rise and fall times (labeled t_{TLH} and t_{THL}, where the subscript T denotes transition, LH denotes low to high, and HL denotes high to low). There is also a delay time between the input and output waveforms. The usual way to specify the propagation delay is to take the average of the high-to-low propagation delay, t_{PHL}, and the low-to-high propagation delay, t_{PLH}. As indicated, these delays are measured between the 50% points of the input and output waveforms. Also note that the **transition times** are specified using the 10% and 90% points of the output excursion $(V_{OH} - V_{OL})$.

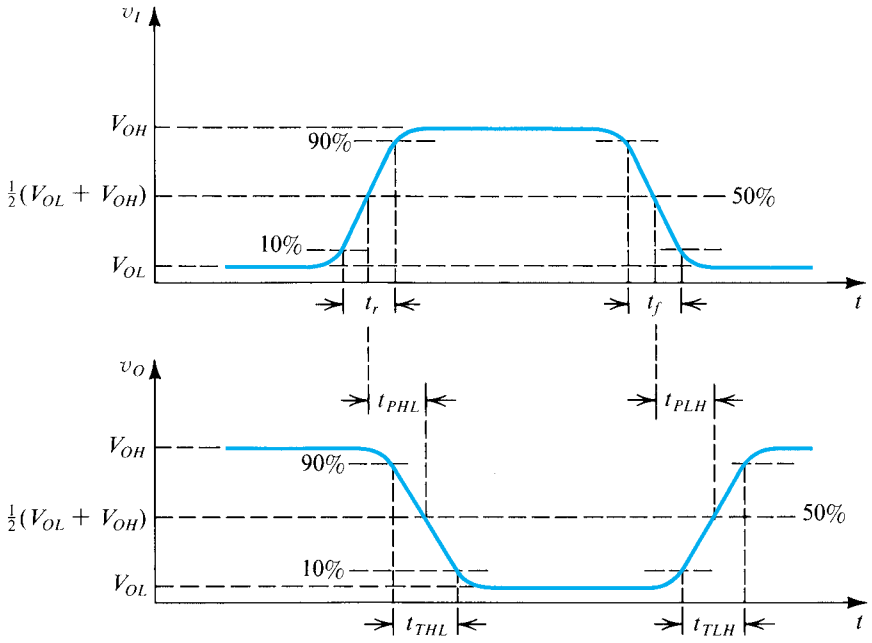


Figure 13.15 Definitions of propagation delays and transition times of the logic inverter.

EXERCISE

13.11 A capacitor $C = 100$ fF is discharged from a voltage V_{DD} to zero through a resistance $R = 2$ kΩ. Find the fall time t_f of the capacitor voltage.

Ans. $t_f \simeq 2.2CR = 0.44$ ns

13.1.8 Power–Delay and Energy–Delay Products

One is usually interested in high-speed operation (low t_P) combined with low power dissipation. Unfortunately, these two requirements are often in conflict: Generally, if the designer of an inverter attempts to reduce power dissipation by, say, decreasing the supply voltage V_{DD}, or the supply current, or both, the current-driving capability of the inverter decreases. This in turn results in longer times to charge and discharge the load and parasitic capacitances, and thus the propagation delay increases. It follows that a figure of merit for comparing logic-circuit technologies is the **power–delay product** (PDP) of the basic inverter of the given technology, defined as

$$PDP \equiv P_D t_P \tag{13.41}$$

where P_D is the power dissipation of the inverter. Note that the *PDP* is an energy quantity and has the units of joules. The lower the *PDP*, the more effective the inverter and the logic circuits based on the inverter are.

For CMOS logic circuits, which is the digital IC technology of primary interest to us here, the static power dissipation of the inverter is zero,[3] and thus P_D is equal to P_{dyn} and given by Eq. (13.35),

$$P_D = fCV_{DD}^2$$

Thus for the CMOS inverter,

$$PDP = fCV_{DD}^2 t_P \tag{13.42}$$

If the inverter is operated at its maximum switching speed given by Eq. (13.38), then

$$PDP = \frac{1}{2}CV_{DD}^2 \tag{13.43}$$

From our earlier discussion of dynamic power dissipation we know that $\frac{1}{2}CV_{DD}^2$ is the amount of energy dissipated during each charging or discharging event of the capacitor, that is, for each output transition of the inverter. Thus, the *PDP* has an interesting physical interpretation: *It is the energy consumed by the inverter for each output transition.*

Although the *PDP* is a valuable metric for comparing different technologies for implementing inverters, it is *not* useful as a design parameter for optimizing a given inverter circuit. To appreciate this point, observe that the expression in Eq. (13.43) indicates that the *PDP* can be minimized by reducing V_{DD} as much as possible while, of course, maintaining proper circuit operation. This, however, would not necessarily result in optimal performance, for t_P will increase as V_{DD} is reduced. The problem is that the *PDP* expression in Eq. (13.43) does not in fact have information about t_P. It follows that a better metric can be obtained by multiplying the energy per transition by the propagation delay. We can thus define the **energy–delay product** *EDP* as

$$EDP \equiv \text{Energy per transition} \times t_P$$

$$= \frac{1}{2}CV_{DD}^2 t_P \tag{13.44}$$

We will utilize the *EDP* in later sections.

13.1.9 Silicon Area

In addition to minimizing power dissipation and propagation delay, another objective in the design of digital VLSI circuits is the minimization of silicon area per logic gate. Smaller area requirement enables the fabrication of a larger number of gates per chip, which has economic and space advantages from a system-design standpoint. Area reduction occurs in three different ways: through advances in processing technology that enable the reduction of the minimum device size, through advances in circuit-design techniques, and through careful chip layout. In this book, our interest lies in circuit design, and we shall make frequent

[3]The exception to this statement is the power dissipation due to leakage currents and subthreshold conduction in the MOSFETs, discussed in Section 13.5.3.

comments on the relationship between the circuit design and its silicon area. As a general rule, the simpler the circuit, the smaller the area required. As will be seen shortly, the circuit designer has to decide on device sizes. Choosing smaller devices has the obvious advantage of requiring smaller silicon area and at the same time reducing parasitic capacitances and thus increasing speed. Smaller devices, however, have lower current-driving capability, which tends to increase delay. Thus, as in all engineering design problems, there is a trade-off to be quantified and exercised in a manner that optimizes whatever aspect of the design is thought to be critical for the application at hand.

13.1.10 Digital IC Technologies and Logic-Circuit Families

The chart in Figure 13.16 shows the major IC technologies and logic-circuit families that are currently in use. The concept of a logic-circuit family perhaps needs a few words of explanation. The basic element of a logic-circuit family is the inverter. A family would include a variety of logic-circuit types made with the same technology, having a similar circuit structure, and exhibiting the same basic features. Each logic-circuit family offers a unique set of advantages and disadvantages. In the conventional style of designing systems, one selects an appropriate logic family (e.g., TTL, CMOS, or ECL) and attempts to implement as much of the system as possible using circuit modules (packages) that belong to this family. In this way, interconnection of the various packages is relatively straightforward. If, on the other hand, packages from more than one family are used, one has to design suitable *interface circuits*. The selection of a logic family is based on such considerations as logic flexibility, speed of operation, availability of complex functions, noise immunity, operating-temperature range, power dissipation, and cost. We will discuss some of these considerations in this chapter and the next two. To begin with, we make some brief remarks on each of the four technologies listed in the chart of Fig. 13.16.

CMOS Although shown as one of four possible technologies, this is not an indication of digital IC market share: CMOS technology is, by a very large margin, the most dominant of all the IC technologies available for digital-circuit design. Although early microprocessors were made using NMOS logic (based on the inverter circuit we studied in Example 13.2), CMOS has completely replaced NMOS. There are a number of reasons for this development, the most important of which is the much lower power dissipation of CMOS circuits. CMOS has also replaced bipolar as the technology of choice in digital-system design and has

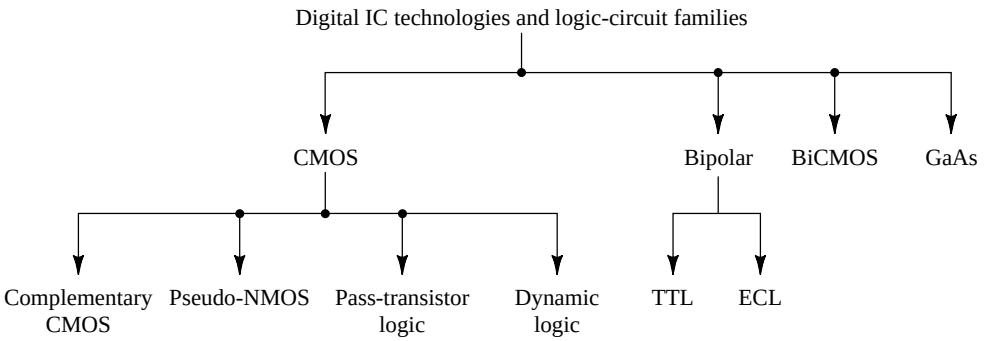


Figure 13.16 Digital IC technologies and logic-circuit families.

made possible levels of integration (or circuit-packing densities) and a range of applications, neither of which would have been possible with bipolar technology. Furthermore, CMOS continues to advance, whereas there appear to be few innovations at the present time in bipolar digital circuits. Some of the reasons for CMOS displacing bipolar technology in digital applications are as follows.

1. CMOS logic circuits dissipate much less power than bipolar logic circuits and thus one can pack more CMOS circuits on a chip than is possible with bipolar circuits.
2. The high input impedance of the MOS transistor allows the designer to use charge storage as a means for the temporary storage of information in both logic and memory circuits. This technique cannot be used in bipolar circuits.
3. The feature size (i.e., minimum channel length) of the MOS transistor has decreased dramatically over the years, with some recently reported designs utilizing channel lengths as short as 32 nm. This permits very tight circuit packing and, correspondingly, very high levels of integration. A microprocessor chip reported in 2009 had 2.3 billion transistors.

Of the various forms of CMOS, complementary CMOS circuits based on the inverter studied in Section 13.2 are the most widely used. They are available both as **small-scale integrated** (**SSI**) circuit packages (containing 1–10 logic gates) and **medium-scale integrated** (**MSI**) circuit packages (10–100 gates per chip) for assembling digital systems on printed-circuit boards. More significantly, complementary CMOS is used in **very-large-scale-integrated (VLSI)** logic (with millions of gates per chip) and memory-circuit design. In some applications, complementary CMOS is supplemented by one (or both) of two other MOS logic circuit forms. These are pseudo-NMOS, so-named because of the similarity of its structure to NMOS logic, and pass-transistor logic, both of which will be studied in Chapter 14.

A fourth type of CMOS logic circuit utilizes dynamic techniques to obtain faster circuit operation, while keeping the power dissipation very low. Dynamic CMOS logic, studied in Chapter 14, represents an area of growing importance. Lastly, CMOS technology is used in the design of memory chips, as will be detailed in Chapter 15.

Bipolar Two logic-circuit families based on the bipolar junction transistor are in some use at present: TTL and ECL. Transistor–transistor logic (TTL or T^2L) was for many years the most widely used logic-circuit family. Its decline was precipitated by the advent of the VLSI era. TTL manufacturers, however, fought back with the introduction of low-power and high-speed versions. In these newer versions, the higher speeds of operation are made possible by preventing the BJT from saturating and thus avoiding the slow turnoff process of a saturated bipolar transistor. These nonsaturating versions of TTL utilize the Schottky diode discussed in Section 4.7 and are called Schottky TTL or variations of this name. Despite all these efforts, TTL is no longer a significant logic-circuit family and will not be studied in this book. However, the interested reader can find significant amounts of material on TTL on the CD accompanying this book and on the book's website.

The other bipolar logic-circuit family in present use is emitter-coupled logic (ECL). It is based on the current-switch implementation of the inverter shown in Fig. 13.9. The basic element of ECL is the differential BJT pair studied in Chapter 8. Because ECL is basically a current-steering logic, and, correspondingly, also called **current-mode logic** (CML), in which saturation is avoided, very high speeds of operation are possible. Indeed, of all the commercially available logic-circuit families, ECL is the fastest. ECL is also used in VLSI circuit design when very high operating speeds are required and the designer is willing to accept higher power dissipation and increased silicon area. As such, ECL is considered an important specialty technology and will be discussed in Chapter 14.

BiCMOS BiCMOS combines the high operating speeds possible with BJTs (because of their inherently higher transconductance) with the low power dissipation and other excellent characteristics of CMOS. Like CMOS, BiCMOS allows for the implementation of both analog and digital circuits on the same chip. (See the discussion of analog BiCMOS circuits in Chapter 7.) At present, BiCMOS is used to great advantage in special applications, including memory chips, where its high performance as a high-speed capacitive-current driver justifies the more complex process technology it requires. A brief discussion of BiCMOS is provided in Chapter 14.

Gallium Arsenide (GaAs) The high carrier mobility in GaAs results in very high speeds of operation. This has been demonstrated in a number of digital IC chips utilizing GaAs technology. It should be pointed out, however, that GaAs remains an "emerging technology," one that appears to have great potential but has not yet achieved such potential commercially. As such, it will not be studied in this book. Nevertheless, considerable material on GaAs devices and circuits, including digital circuits, can be found on the CD accompanying this book and on the book's website.

13.1.11 Styles for Digital-System Design

The conventional approach to designing digital systems consists of assembling the system using standard IC packages of various levels of complexity (and hence integration). Many systems have been built this way using, for example, TTL SSI and MSI packages. The advent of VLSI, in addition to providing the system designer with more powerful off-the-shelf components such as microprocessors and memory chips, has made possible alternative design styles. One such alternative is to opt for implementing part or all of the system using one or more *custom VLSI* chips. However, custom IC design is usually economically justified only when the production volume is large (greater than about 100,000 parts).

An intermediate approach, known as *semicustom design,* utilizes *gate-array* chips. These are integrated circuits containing 100,000 or more unconnected logic gates. Their interconnection can be achieved by a final metallization step (performed at the IC fabrication facility) according to a pattern specified by the user to implement the user's particular functional need. A more recently available type of gate array, known as a **field-programmable gate array** (**FPGA**), can, as its name indicates, be programmed directly by the user. FPGAs provide a very convenient means for the digital-system designer to implement complex logic functions in VLSI form without having to incur either the cost or the "turnaround time" inherent in custom and, to a lesser extent, in semicustom IC design.

13.1.12 Design Abstraction and Computer Aids

The design of very complex digital systems, whether on a single IC chip or using off-the-shelf components, is made possible by the use of different levels of design abstraction, and the use of a variety of computer aids. To appreciate the concept of design abstraction, consider the process of designing a digital system using off-the-shelf packages of logic gates. The designer consults data sheets (in data books or on websites) to determine the input and output characteristics of the gates, their fan-in and fan-out limitations, and so on. In connecting the gates, the designer needs to adhere to a set of rules specified by the manufacturer in the data sheets. The designer does not need to consider, in a direct way, the circuit inside the gate package. In effect, the circuit has been abstracted in the form of a functional block that can be used as a component. This greatly simplifies system design. The digital-IC designer follows a similar process. Circuit blocks are designed,

characterized, and stored in a library as **standard cells**. These cells can then be used by the IC designer to assemble a larger subsystem (e.g., an adder or a multiplier), which in turn is characterized and stored as a functional block to be used in the design of an even larger system (e.g., an entire processor).

At every level of design abstraction, the need arises for simulation and other computer programs that help make the design process as automated as possible. Whereas SPICE is employed in circuit simulation, other software tools are utilized at other levels and in other phases of the design process. Although digital-system design and design automation are outside the scope of this book, it is important that the reader appreciate the role of design abstraction and computer aids in digital design. They are what make it humanly possible to design a billion-transistor digital IC. Unfortunately, analog IC design does not lend itself to the same level of abstraction and automation. Each analog IC to a large extent has to be "handcrafted." As a result, the complexity and density of analog ICs remain much below what is possible in a digital IC.

Whatever approach or style is adopted in digital design, some familiarity with the various digital-circuit technologies and design techniques is essential. This chapter and the next two aim to provide such a background.

13.2 The CMOS Inverter

In this section we study the inverter circuit of the most widely used digital IC technology: CMOS. The basic CMOS inverter is shown in Fig. 13.17. It utilizes two MOSFETs: one, Q_N, with an *n* channel and the other, Q_P, with a *p* channel. The body of each device is connected to its source, and thus no body effect arises. As will be seen shortly, the CMOS circuit realizes the conceptual inverter implementation studied in the previous section (Fig. 13.8), where a pair of switches are operated in a complementary fashion by the input voltage v_I.

13.2.1 Circuit Operation

We first consider the two extreme cases: when v_I is at logic-0 level, which is 0 V; and when v_I is at logic-1 level, which is V_{DD} volts. In both cases, for ease of exposition we shall consider the *n*-channel device Q_N to be the driving transistor and the *p*-channel device Q_P to be

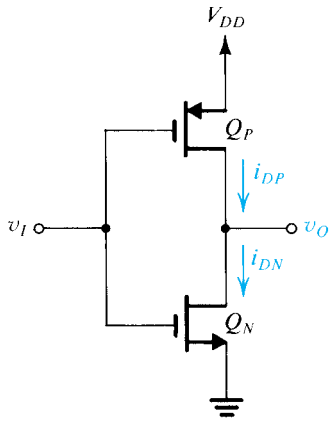


Figure 13.17 The CMOS inverter.

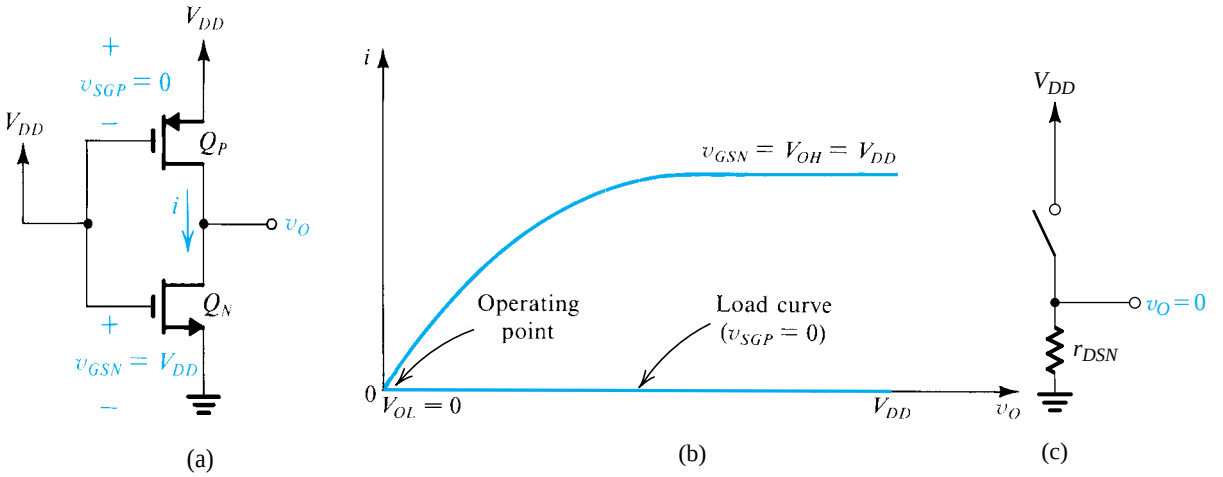


Figure 13.18 Operation of the CMOS inverter when v_I is high: **(a)** circuit with $v_I = V_{DD}$ (logic-1 level, or V_{OH}); **(b)** graphical construction to determine the operating point; **(c)** equivalent circuit.

the load. However, since the circuit is symmetric, this assumption is obviously arbitrary, and the reverse would lead to identical results.

Figure 13.18 illustrates the case when $v_I = V_{DD}$, showing the i_D–v_{DS} characteristic curve for Q_N with $v_{GSN} = V_{DD}$. (Note that $i_D = i$ and $v_{DSN} = v_O$.) Superimposed on the Q_N characteristic curve is the load curve, which is the i_D–v_{SD} curve of Q_P for the case $v_{SGP} = 0$ V. Since $v_{SGP} < |V_t|$, the load curve will be a horizontal straight line at zero current level. The operating point will be at the intersection of the two curves, where we note that the output voltage is zero and the current through the two devices is also zero. This means that the power dissipation in the circuit is zero. Note, however, that although Q_N is operating at zero current and zero drain-source voltage (i.e., at the origin of the i_D–v_{DS} plane), the operating point is on a steep segment of the i_D–v_{DS} characteristic curve. Thus Q_N provides a low-resistance path between the output terminal and ground, with the resistance obtained using Eq. (5.13b) as

$$r_{DSN} = 1 \Big/ \left[k_n' \left(\frac{W}{L} \right)_n (V_{DD} - V_{tn}) \right] \tag{13.45}$$

Figure 13.18(c) shows the equivalent circuit of the inverter when the input is high. This circuit confirms that $v_O \equiv V_{OL} = 0$ V and that the power dissipation in the inverter is zero.

The other extreme case, when $v_I = 0$ V, is illustrated in Fig. 13.19. In this case Q_N is operating at $v_{GSN} = 0$; hence its i_D–v_{DS} characteristic is a horizontal straight line at zero current level. The load curve is the i_D–v_{SD} characteristic of the p-channel device with $v_{SGP} = V_{DD}$. As shown, at the operating point the output voltage is equal to V_{DD}, and the current in the two devices is still zero. Thus the power dissipation in the circuit is zero in both extreme states.

Figure 13.19(c) shows the equivalent circuit of the inverter when the input is low. Here we see that Q_P provides a low-resistance path between the output terminal and the dc supply V_{DD}, with the resistance given by

$$r_{DSP} = 1 \Big/ \left[k_p' \left(\frac{W}{L} \right)_p (V_{DD} - |V_{tp}|) \right] \tag{13.46}$$

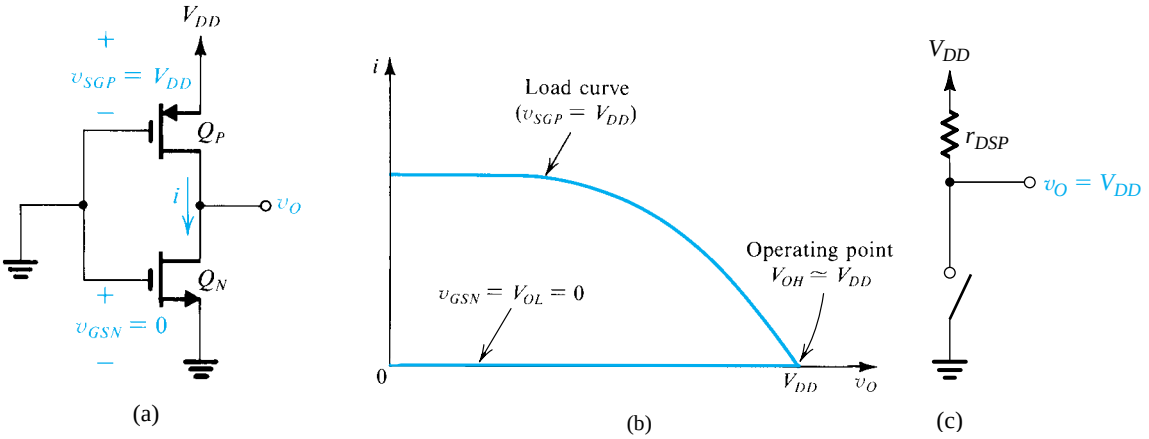


Figure 13.19 Operation of the CMOS inverter when v_I is low: **(a)** circuit with $v_I = 0$ V (logic-0 level, or V_{OL}); **(b)** graphical construction to determine the operating point; **(c)** equivalent circuit.

The equivalent circuit confirms that in this case $v_O \equiv V_{OH} = V_{DD}$ and that the power dissipation in the inverter is zero.

It should be noted, however, that in spite of the fact that the quiescent current is zero, the load-driving capability of the CMOS inverter is high. For instance, with the input high, as in the circuit of Fig. 13.18, transistor Q_N can sink a relatively large load current. This current can quickly discharge the load capacitance, as will be seen shortly. Because of its action in sinking load current and thus pulling the output voltage down toward ground, transistor Q_N is known as the **pull-down device**. Similarly, with the input low, as in the circuit of Fig. 13.19, transistor Q_P can source a relatively large load current. This current can quickly charge up a load capacitance, thus pulling the output voltage up toward V_{DD}. Hence, Q_P is known as the **pull-up device**. The reader will recall that we used this terminology in connection with the conceptual inverter circuit of Fig. 13.8.

From the above, we conclude that the basic CMOS logic inverter behaves as an ideal inverter. In summary:

1. The output voltage levels are 0 and V_{DD}, and thus the signal swing is the maximum possible. This, coupled with the fact that the inverter can be designed to provide a symmetrical voltage-transfer characteristic, results in wide noise margins.
2. The static power dissipation in the inverter is zero (neglecting the dissipation due to leakage currents) in both of its states. This is because no dc path exists between the power supply and ground in either state.
3. A low-resistance path exists between the output terminal and ground (in the low-output state) or V_{DD} (in the high-output state). These low-resistance paths ensure that the output voltage is 0 or V_{DD} independent of the exact values of the W/L ratios or other device parameters. Furthermore, the low output resistance makes the inverter less sensitive to the effects of noise and other disturbances.
4. The active pull-up and pull-down devices provide the inverter with high output-driving capability in both directions. As will be seen, this speeds up the operation considerably.

5. The input resistance of the inverter is infinite (because $I_G = 0$). Thus the inverter can drive an arbitrarily large number of similar inverters with no loss in signal level. Of course, each additional inverter increases the load capacitance on the driving inverter and slows down the operation. Shortly, we will consider the inverter switching times.

13.2.2 The Voltage-Transfer Characteristic

The complete voltage-transfer characteristic (VTC) of the CMOS inverter can be obtained by repeating the graphical procedure, used above in the two extreme cases, for all intermediate values of v_I. In the following, we shall calculate the critical points of the resulting voltage-transfer curve. For this we need the i–v relationships of Q_N and Q_P. For Q_N,

$$i_{DN} = k_n' \left(\frac{W}{L}\right)_n \left[(v_I - V_{tn})v_O - \frac{1}{2}v_O^2\right] \qquad \text{for } v_O \le v_I - V_{tn} \tag{13.47}$$

and

$$i_{DN} = \frac{1}{2}k_n' \left(\frac{W}{L}\right)_n (v_I - V_{tn})^2 \qquad \text{for } v_O \ge v_I - V_{tn} \tag{13.48}$$

For Q_P,

$$i_{DP} = k_p' \left(\frac{W}{L}\right)_p \left[(V_{DD} - v_I - |V_{tp}|)(V_{DD} - v_O) - \frac{1}{2}(V_{DD} - v_O)^2\right]$$

$$\text{for } v_O \ge v_I + |V_{tp}| \tag{13.49}$$

and

$$i_{DP} = \frac{1}{2}k_p' \left(\frac{W}{L}\right)_p (V_{DD} - v_I - |V_{tp}|)^2 \qquad \text{for } v_O \le v_I + |V_{tp}| \tag{13.50}$$

The CMOS inverter is usually designed to have $V_{tn} = |V_{tp}| = V_t$. Also, although this is not always the case, we shall assume that Q_N and Q_P are matched; that is, $k_n'(W/L)_n = k_p'(W/L)_p$. It should be noted that since μ_p is 0.25 to 0.5 times the value of μ_n, to make $k'(W/L)$ of the two devices equal, the width of the p-channel device is made two to four times that of the n-channel device. More specifically, the two devices are designed to have equal lengths, with widths related by

$$\frac{W_p}{W_n} = \frac{\mu_n}{\mu_p} \tag{13.51}$$

This will result in $k_n'(W/L)_n = k_p'(W/L)_p$, and the inverter will have a symmetric transfer characteristic and equal current-driving capability in both directions (pull-up and pull-down).

With Q_N and Q_P matched, the CMOS inverter has the voltage transfer characteristic shown in Fig. 13.20. As indicated, the transfer characteristic has five distinct segments corresponding to different combinations of modes of operation of Q_N and Q_P. The vertical segment BC is obtained when both Q_N and Q_P are operating in the saturation region. Because we are neglecting the finite output resistance in saturation, that is, assuming $\lambda_N = \lambda_P = 0$, the inverter gain in this region is infinite. From symmetry, this vertical segment occurs at $v_I = V_{DD}/2$ and is bounded by $v_O(\text{B}) = V_{DD}/2 + V_t$, at which value Q_P enters the triode region and $v_O(\text{C}) = V_{DD}/2 - V_t$, at which value Q_N enters the triode region.

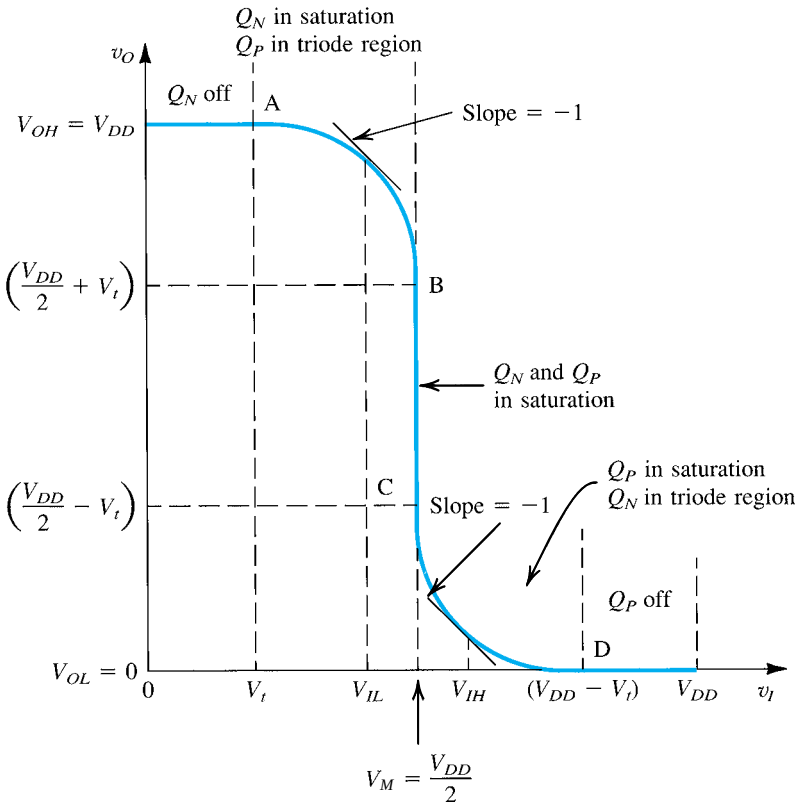


Figure 13.20 The voltage-transfer characteristic of the CMOS inverter when Q_N and Q_P are matched.

The reader will recall from Section 13.1.3 that in addition to V_{OL} and V_{OH}, two other points on the transfer curve determine the noise margins of the inverter. These are the maximum permitted logic-0 or "low" level at the input, V_{IL}, and the minimum permitted logic-1 or "high" level at the input, V_{IH}. These are formally defined as the two points on the transfer curve at which the incremental gain is unity (i.e., the slope is −1 V/V).

To determine V_{IH}, we note that Q_N is in the triode region, and thus its current is given by Eq. (13.47), while Q_P is in saturation and its current is given by Eq. (13.50). Equating i_{DN} and i_{DP}, and assuming matched devices, gives

$$(v_I - V_t)v_O - \frac{1}{2}v_O^2 = \frac{1}{2}(V_{DD} - v_I - V_t)^2 \tag{13.52}$$

Differentiating both sides relative to v_I results in

$$(v_I - V_t)\frac{dv_O}{dv_I} + v_O - v_O\frac{dv_O}{dv_I} = -(V_{DD} - v_I - V_t)$$

in which we substitute $v_I = V_{IH}$ and $dv_O/dv_I = -1$ to obtain

$$v_O = V_{IH} - \frac{V_{DD}}{2} \tag{13.53}$$

Substituting $v_I = V_{IH}$ and for v_O from Eq. (13.53) in Eq. (13.52) gives

$$V_{IH} = \frac{1}{8}(5V_{DD} - 2V_t) \tag{13.54}$$

V_{IL} can be determined in a manner similar to that used to find V_{IH}. Alternatively, we can use the symmetry relationship

$$V_{IH} - \frac{V_{DD}}{2} = \frac{V_{DD}}{2} - V_{IL}$$

together with V_{IH} from Eq. (13.54) to obtain

$$V_{IL} = \frac{1}{8}(3V_{DD} + 2V_t) \tag{13.55}$$

The noise margins can now be determined as follows:

$$\begin{aligned} NM_H &= V_{OH} - V_{IH} \\ &= V_{DD} - \frac{1}{8}(5V_{DD} - 2V_t) \\ &= \frac{1}{8}(3V_{DD} + 2V_t) \end{aligned} \tag{13.56}$$

$$\begin{aligned} NM_L &= V_{IL} - V_{OL} \\ &= \frac{1}{8}(3V_{DD} + 2V_t) - 0 \\ &= \frac{1}{8}(3V_{DD} + 2V_t) \end{aligned} \tag{13.57}$$

As expected, the symmetry of the voltage-transfer characteristic results in equal noise margins. Of course, if Q_N and Q_P are not matched, the voltage-transfer characteristic will no longer be symmetric, and the noise margins will not be equal.

13.2.3 The Situation When Q_N and Q_P Are Not Matched

In the above we assumed that Q_N and Q_P are matched; that is, in addition to $V_{tn} = |V_{tp}|$, the transconductance parameters k_n and k_p are made equal by selecting W_p/W_n according to Eq. (13.51). The result is a symmetrical VTC that switches at the midpoint of the supply; that is, $V_M = V_{DD}/2$. The symmetry, as we have seen, equalizes and maximizes the noise margins.

The price paid for obtaining a perfectly symmetric VTC is that the width of the *p*-channel device can be three to four times as large as that of the *n*-channel device. This can result in a relatively large silicon area which, besides being wasteful of silicon real estate, can also result in increased device capacitances and a corresponding increase in the propagation delay of the inverter. It is useful, therefore, to inquire into the effect of not matching Q_N and Q_P. Toward that end we derive an expression for the switching voltage V_M as follows.

Since at M, both Q_N and Q_P operate in saturation, their currents are given by Eqs. (13.48) and (13.50). Substituting $v_I = v_O = V_M$, and equating the two currents results in

$$V_M = \frac{r(V_{DD} - |V_{tp}|) + V_{tn}}{r + 1} \tag{13.58}$$

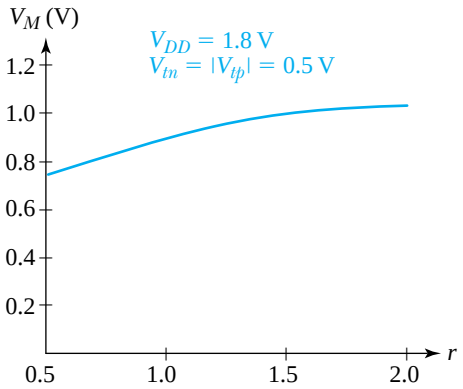


Figure 13.21 Variation of the inverter switching voltage, V_M, with the parameter $r = \sqrt{k_p/k_n}$.

where

$$r = \sqrt{\frac{k_p}{k_n}} = \sqrt{\frac{\mu_p W_p}{\mu_n W_n}} \tag{13.59}$$

where we have assumed that Q_N and Q_P have the same channel length L, which is usually the case with L equal to the minimum available for the given process technology. Note that the matched case corresponds to $r = 1$. For $|V_{tp}| = V_{tn}$, and $r = 1$, Eq. (13.58) yields $V_M = V_{DD}/2$, as expected. For a given process, that is, given values for V_{DD}, V_{tn}, and V_{tp}, one can plot V_M versus the matching parameter r. Such a plot, for a 0.18-μm process, is shown in Fig. 13.21. We make the following two observations:

1. V_M increases with r. Thus, making $k_p > k_n$ shifts V_M toward V_{DD}. Conversely, making $k_p < k_n$ shifts V_M toward 0.
2. V_M is not a strong function of r. For the particular case shown, lowering r by a factor of 2 (from 1 to 0.5), reduces V_M by only 0.13 V.

Observation 2 implies that if one is willing to tolerate a small reduction in NM_L, substantial savings in silicon area can be obtained. This point is illustrated in Example 13.4.

Example 13.4

Consider a CMOS inverter fabricated in a 0.18-μm process for which $V_{DD} = 1.8$ V, $V_{tn} = |V_{tp}| = 0.5$ V, $\mu_n = 4\mu_p$, and $\mu_n C_{ox} = 300\ \mu A/V^2$. In addition, Q_N and Q_P have $L = 0.18\ \mu m$ and $(W/L)_n = 1.5$.

(a) Find W_p that results in $V_M = V_{DD}/2 = 0.9$ V. What is the silicon area utilized by the inverter in this case?

(b) For the matched case in (a), find the values of V_{OH}, V_{OL}, V_{IH}, V_{IL}, and the noise margins NM_L and NM_H. For $v_I = V_{IH}$, what value of v_O results? This can be considered the worst-case value of V_{OL}. Similarly, for $v_I = V_{IL}$, find v_O that is the worst-case value of V_{OH}. Now, use these worst-case values to determine more conservative values for the noise margins.

(c) For the matched case in (a), find the output resistance of the inverter in each of its two states.

Example 13.4 *continued*

(d) If $\lambda_n = |\lambda_p| = 0.2\ \text{V}^{-1}$, what is the inverter gain at $v_I = V_M$. If a straight line is drawn through the point $v_I = v_O = V_M$ with a slope equal to the gain, at what values of v_I does it intercept the horizontal lines $v_O = 0$ and $v_O = V_{DD}$? Use these intercepts to estimate the width of the transition region of the VTC.

(e) If $W_p = W_n$, what value of V_M results? What do you estimate the reduction of NM_L (relative to the matched case) to be? What is the percentage savings in silicon area (relative to the matched case)?

(f) Repeat (e) for the case $W_p = 2W_n$. This case, which is frequently used in industry, can be considered to be a compromise between the minimum-area case in (e) and the matched case.

Solution

(a) To obtain $V_M = V_{DD}/2 = 0.9\ \text{V}$, we select W_p according to Eq. (13.51),

$$\frac{W_p}{W_n} = \frac{\mu_n}{\mu_p} = 4$$

Since $W_n/L = 1.5$, $W_n = 1.5 \times 0.18 = 0.27\ \mu\text{m}$. Thus,

$$W_p = 4 \times 0.27 = 1.08\ \mu\text{m}$$

For this design, the silicon area is

$$A = W_nL + W_pL = L(W_n + W_p)$$
$$= 0.18(0.27 + 1.08) = 0.243\ \mu\text{m}^2$$

(b)
$$V_{OH} = V_{DD} = 1.8\ \text{V}$$
$$V_{OL} = 0\ \text{V}$$

To obtain V_{IH} we use Eq. (13.54),

$$V_{IH} = \frac{1}{8}(5V_{DD} - 2V_t) = \frac{1}{8}(5 \times 1.8 - 2 \times 0.5) = 1\ \text{V}$$

To obtain V_{IL} we use Eq. (13.55),

$$V_{IL} = \frac{1}{8}(3V_{DD} + 2V_t) = \frac{1}{8}(3 \times 1.8 + 2 \times 0.5) = 0.8\ \text{V}$$

We can now compute the noise margins as

$$NM_H = V_{OH} - V_{IH} = 1.8 - 1.0 = 0.8\ \text{V}$$

$$NM_L = V_{IL} - V_{OL} = 0.8 - 0 = 0.8\ \text{V}$$

As expected, $NM_H = NM_L$, and their value is very close to the optimum value of $V_{DD}/2 = 0.9\ \text{V}$.

For $v_I = V_{IH} = 1\ \text{V}$, we can obtain the corresponding value of v_O by substituting in Eq. (13.53),

$$v_O = V_{IH} - \frac{V_{DD}}{2} = 1 - \frac{1.8}{2} = 0.1\ \text{V}$$

Thus, the worst-case value of V_{OL}, that is, $V_{OL\max}$, is 0.1 V, and the noise margin NM_L reduces to

$$NM_L = V_{IL} - V_{OL\max} = 0.8 - 0.1 = 0.7\ \text{V}$$

From symmetry, we can obtain the value of v_O corresponding to $v_I = V_{IL}$ as

$$v_O = V_{DD} - 0.1 = 1.7\ \text{V}$$

Thus the worst-case value of V_{OH}, that is, $V_{OH\min}$, is 1.7 V, and the noise margin NM_H reduces to

$$NM_H = V_{OH\min} - V_{IH} = 1.7 - 1 = 0.7 \text{ V}$$

Note that the reduction in the noise margins is slight.

(c) The output resistance of the inverter in the low-output state is

$$r_{DSN} = \frac{1}{\mu_n C_{ox}(W/L)_n(V_{DD} - V_{tn})}$$

$$= \frac{1}{300 \times 10^{-6} \times 1.5(1.8 - 0.5)} = 1.71 \text{ k}\Omega$$

Since Q_N and Q_P are matched, the output resistance in the high-output state will be equal, that is,

$$r_{DSP} = r_{DSN} = 1.71 \text{ k}\Omega$$

(d) If the inverter is biased to operate at $v_I = v_O = V_M = 0.9$ V, then each of Q_N and Q_P will be operating at an overdrive voltage $V_{OV} = V_M - V_t = 0.9 - 0.5 = 0.4$ V and will be conducting equal dc currents I_D of

$$I_D = \frac{1}{2}\mu_n C_{ox}\left(\frac{W}{L}\right)_N V_{OV}^2$$

$$= \frac{1}{2} \times 300 \times 1.5 \times 0.4^2$$

$$= 36 \text{ } \mu\text{A}$$

Thus, Q_N and Q_P will have equal transconductances:

$$g_{mn} = g_{mp} = \frac{2I_D}{V_{OV}} = \frac{2 \times 36}{0.4} = 0.18 \text{ mA/V}^2$$

Transistors Q_N and Q_P will have equal output resistances r_o,

$$r_{on} = r_{op} = \frac{|V_A|}{I_D} = \frac{1}{|\lambda| I_D} = \frac{1}{0.2 \times 36} = 139 \text{ k}\Omega$$

We can now compute the voltage gain at M as

$$A_v = -(g_{mn} + g_{mp})(r_{on} \| r_{op})$$

$$= -(0.18 + 0.18)(139 \| 139) = -25 \text{ V/V}$$

When the straight line at M of slope –25 V/V is extrapolated, it intersects the line $v_O = 0$ at $[0.9 + 0.9/25] = 0.936$ V and the line $v_O = V_{DD}$ at $(0.9 - 0.9/25) = 0.864$ V. Thus the width of the transition region can be considered to be $(0.936 - 0.864) = 0.072$ V.

(e) For $W_p = W_n$, the parameter r can be found from Eq. (13.59),

$$r = \sqrt{\frac{\mu_p W_p}{\mu_n W_n}} = \sqrt{\frac{1}{4} \times 1} = 0.5$$

The corresponding value of V_M can be determined from Eq. (13.58) as

$$V_M = \frac{0.5(1.8 - 0.5) + 0.5}{0.5 + 1} = 0.77 \text{ V}$$

Example 13.4 *continued*

Thus V_M shifts by only –0.13 V. Without recalculating V_{IL} we can estimate the reduction in NM_L to be approximately equal to the shift in V_M, that is, NM_L becomes $0.8 - 0.13 = 0.67$ V. The silicon area for this design can be computed as follows:

$$A = L(W_n + W_p) = 0.18(0.27 + 0.27)$$
$$= 0.0972 \ \mu m^2$$

This represents a 60% reduction from the matched case!

(f) For $W_p = 2W_n$,

$$r = \sqrt{\frac{1}{4} \times 2} = \frac{1}{\sqrt{2}} = 0.707$$

$$V_M = \frac{0.707(1.8 - 0.5) + 0.5}{0.707 + 1} = 0.83 \text{ V}$$

Thus, relative to the matched case, the shift in V_M is only –0.07 V. We estimate that NM_L will decrease from 0.8 V by the same amount; thus NM_L becomes 0.73 V. In this case, the silicon area required is

$$A = L(W_n + W_p) = 0.18(0.27 + 0.54)$$
$$= 0.146 \ \mu m^2$$

which represents a 40% reduction relative to the matched case!

EXERCISES

13.12 Consider a CMOS inverter fabricated in a 0.13-μm process for which $V_{DD} = 1.2$ V, $V_{tn} = -V_{tp} = 0.4$ V, $\mu_n/\mu_p = 4$, and $\mu_n C_{ox} = 430 \ \mu A/V^2$. In addition, Q_N and Q_P have $L = 0.13$ μm and $(W/L)_n = 1.0$.
(a) Find W_p that results in $V_M = 0.6$ V.
(b) For the matched case in (a), find the values of V_{OH}, V_{OL}, V_{IH}, V_{IL}, NM_H, and NM_L.
(c) For the inverter in (a), find the output resistance in each of its two states.
(d) For a minimum-size inverter for which $(W/L)_p = (W/L)_n = 1.0$, find V_M.
Ans. (a) 0.52 μm; (b) 1.2 V, 0 V, 0.65 V, 0.55 V, 0.55 V, 0.55 V, (c) 2.9 kΩ, 2.9 kΩ; (d) 0.53 V

D13.13 A CMOS inverter utilizes $V_{DD} = 5$ V, $V_{tn} = |V_{tp}| = 1$ V, and $\mu_n C_{ox} = 2\mu_p C_{ox} = 50 \ \mu A/V^2$. Find $(W/L)_n$ and $(W/L)_p$ so that $V_M = 2.5$ V and so that for $v_I = V_{DD}$, the inverter can sink a current of 0.2 mA with the output voltage not exceeding 0.2 V.
Ans. $(W/L)_n \simeq 5$; $(W/L)_p \simeq 10$

13.3 Dynamic Operation of the CMOS Inverter

As explained in Section 13.1.7, the speed of operation of a digital system (e.g., a computer) is determined by the propagation delay of the logic gates used to construct the system. Since the inverter is the basic logic gate of any digital IC technology, the propagation delay of the

inverter is a fundamental parameter in characterizing the technology. In the following, we analyze the switching operation of the CMOS inverter to determine its propagation delay. We shall do this by utilizing a two-step process.

1. Replace all the capacitances in the circuit: that is, the various capacitances associated with Q_N and Q_P, the capacitance of the wire that connects the output of the inverter to other circuits, and the input capacitance of the logic gates the inverter drives, by a single equivalent capacitance C connected between the output node of the inverter and ground.
2. Analyze the resulting capacitively loaded inverter to determine its t_{PLH} and t_{PHL}, and hence t_P.

We shall study these two separable steps in reverse order. Thus, in Section 13.3.1 we show how the propagation delay can be determined. Then, in Section 13.3.2, we show how to calculate the value of C.

13.3.1 Determining the Propagation Delay

Figure 13.22(a) shows a CMOS inverter with a capacitance C connected between its output node and ground. To determine the propagation delays t_{PHL} and t_{PLH}, we apply to the input an ideal pulse, that is, one with zero rise and fall times, as shown in Fig. 13.22(b). Since the circuit is symmetric, the analyses to determine the two propagation delays will be similar. Therefore, we will derive t_{PHL} in detail and extrapolate the result to determine t_{PLH}.

Just prior to the leading edge of the input pulse (i.e., at $t = 0-$), the output voltage is equal to V_{DD} and capacitor C is charged to this voltage. At $t = 0$, v_I rises to V_{DD}, causing Q_P to turn off and Q_N to turn on. From then on, the circuit is equivalent to that shown in Fig. 13.22(c), with the initial value of $v_O = V_{DD}$. Thus, at $t = 0+$, Q_N will operate in the saturation region and will supply a relatively large current to begin the process of discharging C. Figure 13.22(d) shows the trajectory of the operating point of Q_N as C is discharged. Here we are interested in the interval t_{PHL} during which v_O reduces from V_{DD} to $V_{DD}/2$. Correspondingly, the operating point of Q_N moves from E to M. For a portion of this time, corresponding to the segment EF of the trajectory, Q_N operates in saturation. Then at F, $v_O = V_{DD} - V_t$, and Q_N enters the triode region.

A simple approach for determining t_{PHL} consists of first calculating the average value of the current supplied by Q_N over the segment EM. Then, we use this average value of the discharge current to determine t_{PHL} by means of the charge balance equation

$$I_{\text{av}}\, t_{PHL} = C[V_{DD} - (V_{DD}/2)]$$

resulting in

$$t_{PHL} = \frac{CV_{DD}}{2I_{\text{av}}} \tag{13.60}$$

The value of I_{av} can be found as follows:

$$I_{\text{av}} = \frac{1}{2}[i_{DN}(\text{E}) + i_{DN}(\text{M})] \tag{13.61}$$

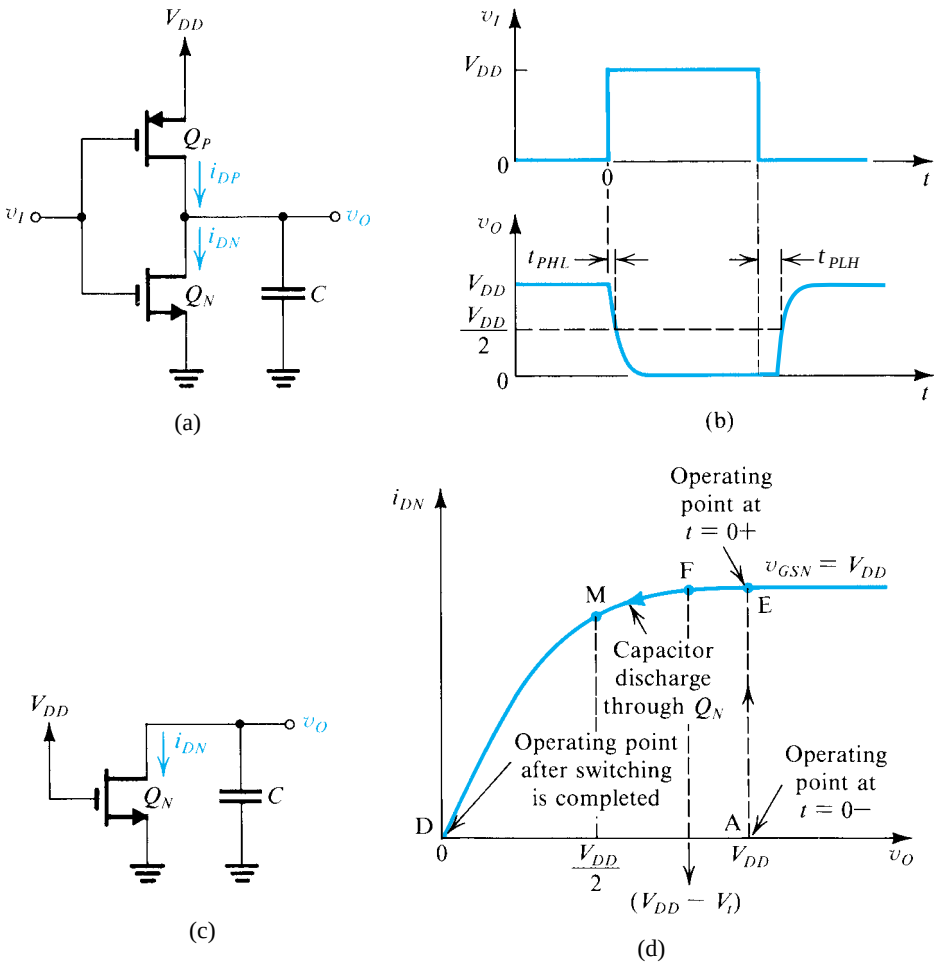


Figure 13.22 Dynamic operation of a capacitively loaded CMOS inverter: **(a)** circuit; **(b)** input and output waveforms; **(c)** equivalent circuit during the capacitor discharge; **(d)** trajectory of the operating point as the input goes high and C discharges through Q_N.

where

$$i_{DN}(\text{E}) = \frac{1}{2}k_n'\left(\frac{W}{L}\right)_n (V_{DD} - V_{tn})^2 \tag{13.62}$$

and

$$i_{DN}(\text{M}) = k_n'\left(\frac{W}{L}\right)_n \left[(V_{DD} - V_{tn})\left(\frac{V_{DD}}{2}\right) - \frac{1}{2}\left(\frac{V_{DD}}{2}\right)^2\right] \tag{13.63}$$

Note that we have assumed $\lambda_n = 0$. Combining Eqs. (13.60) to (13.63) provides

$$t_{PHL} = \frac{\alpha_n C}{k_n'(W/L)_n V_{DD}} \tag{13.64}$$

where α_n is a factor determined by the relative values of V_t and V_{DD};

$$\alpha_n = 2\Big/\left[\frac{7}{4} - \frac{3V_{tn}}{V_{DD}} + \left(\frac{V_{tn}}{V_{DD}}\right)^2\right] \tag{13.65}$$

The value of α_n falls in the range of 1 to 2.

An expression for the low-to-high inverter delay, t_{PLH}, can be written by analogy to the t_{PHL} expression in Eq. (13.64),

$$t_{PLH} = \frac{\alpha_p}{k_p'\left(\frac{W}{L}\right)_p V_{DD}} \tag{13.66}$$

where

$$\alpha_p = 2\Big/\left[\frac{7}{4} - \frac{3|V_{tp}|}{V_{DD}} + \left|\frac{V_{tp}}{V_{DD}}\right|^2\right] \tag{13.67}$$

Finally, the propagation delay t_P can be found as the average of t_{PHL} and t_{PLH},

$$t_P = \frac{1}{2}(t_{PHL} + t_{PLH})$$

Examination of the formulas in Eqs. (13.64) to (13.67) enables us to make a number of useful observations:

1. As expected, the two components of t_p can be equalized by selecting the (*W/L*) ratios to equalize k_n and k_p, that is, by matching Q_N and Q_P.
2. Since t_p is proportional to *C*, the designer should strive to reduce *C*. This is achieved by using the minimum possible channel length and by minimizing wiring and other parasitic capacitances. Careful layout of the chip can result in significant reduction in such capacitances.
3. Using a process technology with larger transconductance parameter k' can result in shorter propagation delays. Keep in mind, however, that for such processes C_{ox} is increased, and thus the value of *C* increases at the same time (more on this later).
4. Using larger *W/L* ratios can result in a reduction in t_P. Care, however, should be exercised here also, since increasing the size of the devices increases the value of *C*, and thus the expected reduction in t_P might not materialize. Reducing t_p by increasing *W/L*, however, is an effective strategy when *C* is dominated by components not directly related to the size of the driving device (such as wiring or fan-out devices).
5. A larger supply voltage V_{DD} results in a lower t_P. However, V_{DD} is determined by the process technology and thus is often not under the control of the designer. Furthermore, modern process technologies in which device sizes are reduced require lower V_{DD} (see Table 7.A.1). A motivating factor for lowering V_{DD} is the need to keep the dynamic power dissipation at acceptable levels, especially in very-high-density chips. We will have more to say on this point shortly.

These observations clearly illustrate the conflicting requirements and the trade-offs available in the design of a CMOS digital integrated circuit (and indeed in any engineering design problem).

An Alternative Approach The formulas derived above for t_{PHL} and t_{PLH} underestimate the delay values for inverters implemented in deep-submicron technologies. This arises because of the velocity saturation effect, which we shall discuss briefly in Section 13.5. There

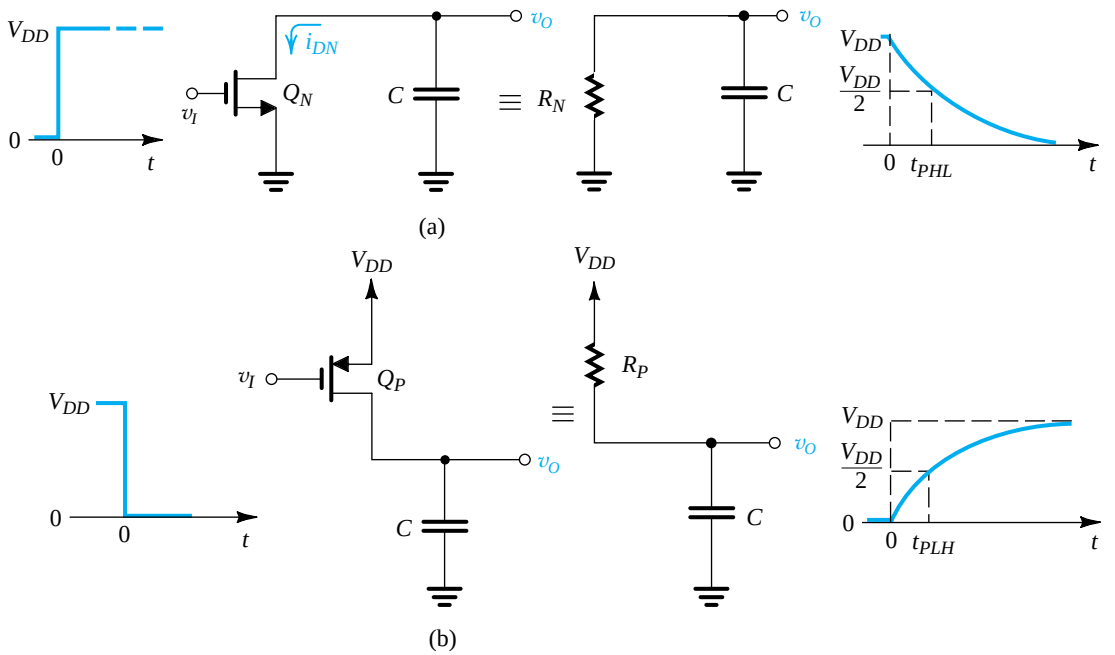


Figure 13.23 Equivalent circuits for determining the propagation delays (**a**) t_{PHL} and (**b**) t_{PLH} of the inverter.

we will see that velocity saturation results in lower MOSFET currents in the saturation region, and hence in increased delay times. To deal with this problem, we present a very simple alternative approach to estimating the inverter propagation delay.

Figure 13.23 illustrates the alternative approach. During the discharge delay t_{PHL}, Q_N is replaced by an equivalent resistance R_N. Similarly, during the charging delay t_{PLH}, Q_P is replaced by an equivalent resistance R_P. It is easy to show that

$$t_{PHL} = 0.69 R_N C \tag{13.68}$$

and

$$t_{PLH} = 0.69 R_P C \tag{13.69}$$

Empirical values have been found for R_N and R_P,

$$R_N = \frac{12.5}{(W/L)_n} \ \text{k}\Omega \tag{13.70}$$

$$R_P = \frac{30}{(W/L)_p} \ \text{k}\Omega \tag{13.71}$$

Furthermore, it has been found that these values apply for a number of CMOS fabrication processes including 0.25 μm, 0.18 μm, and 0.13 μm (see Hodges et al., 2004).

Example 13.5

For the 0.25-μm process characterized by $V_{DD} = 2.5$ V, $V_{tn} = -V_{tp} = 0.5$ V, $k_n' = 3.5k_p' = 115\ \mu A/V^2$, find t_{PLH}, t_{PHL}, and t_P for an inverter for which $(W/L)_n = 1.5$ and $(W/L)_p = 3$, and for $C = 10$ fF. Use both the approach based on average currents and that based on equivalent resistances, and compare the results obtained. If to save on power dissipation the inverter is operated at $V_{DD} = 2.0$ V, by what factor does t_P change?

Solution

(a) Using the average current approach, we determine from Eq. (13.65),

$$\alpha_n = \frac{2}{\frac{7}{4} - \frac{3 \times 0.5}{2.5} + \left(\frac{0.5}{2.5}\right)^2} = 1.7$$

and using Eq. (13.64),

$$t_{PHL} = \frac{1.7 \times 10 \times 10^{-15}}{110 \times 10^{-6} \times 1.5 \times 2.5} = 41.2 \text{ ps}$$

Since $|V_{tp}| = V_{tn}$,

$$\alpha_p = \alpha_n = 1.7$$

and we can determine t_{PLH} from Eq. (13.66) as

$$t_{PLH} = \frac{1.7 \times 10 \times 10^{-15}}{(110/3.5) \times 10^{-6} \times 3 \times 2.5} = 72.1 \text{ ps}$$

The propagation delay can now be found as

$$t_P = \frac{1}{2}(t_{PHL} + t_{PLH})$$

$$= \frac{1}{2}(41.2 + 72.1) = 56.7 \text{ ps}$$

(b) Using the equivalent resistance approach, we first find R_N from Eq. (13.70) as

$$R_N = \frac{12.5}{1.5} = 8.33 \text{ k}\Omega$$

and then use Eq. (13.68) to determine t_{PHL},

$$t_{PHL} = 0.69 \times 8.33 \times 10^3 \times 10 \times 10^{-15} = 57.5 \text{ ps}$$

Similarly we use Eq. (13.71) to determine R_P,

$$R_P = \frac{30}{3} = 10 \text{ k}\Omega$$

and Eq. (13.69) to determine t_{PLH},

$$t_{PLH} = 0.69 \times 10 \times 10^3 \times 10 \times 10^{-15} = 69 \text{ ps}$$

Thus, while the value obtained for t_{PHL} is higher than that found using average currents, the value for t_{PLH} is about the same. Finally, t_P can be found as

$$t_P = \frac{1}{2}(57.5 + 69) = 63.2 \text{ ps}$$

which a little higher than the value found using average currents.

Example 13.5 *continued*

To find the change in propagation delays obtained when the inverter is operated at $V_{DD} = 2.0$ V, we have to use the method of average currents. (The dependence on the power-supply voltage is absorbed in the empirical values of R_N and R_P.) Using Eq. (13.65), we write

$$\alpha_n = \frac{2}{\frac{7}{4} - \frac{3 \times 0.5}{2} + \left(\frac{0.5}{2}\right)^2} = 2.1$$

The value of t_{PHL} can now be found by using Eq. (13.64):

$$t_{PHL} = \frac{2.1 \times 10 \times 10^{-15}}{110 \times 10^{-6} \times 1.5 \times 2} = 63.6 \text{ ps}$$

Similarly, the value of $\alpha_p = \alpha_n = 2.1$ can be substituted in Eq. (13.66) to obtain

$$t_{PLH} = \frac{2.1 \times 10 \times 10^{-15}}{(110/3.5) \times 10^{-6} \times 3 \times 2} = 111.4 \text{ ps}$$

and t_P can be calculated as

$$t_P = \frac{1}{2}(63.6 + 111.4) = 87.5 \text{ ps}$$

Thus, as expected, reducing V_{DD} has resulted in increased propagation delay.

Before leaving the subject of propagation delay, we should emphasize that hand analysis using the simple formulas above should *not* be expected to yield precise results. Rather, its value is in obtaining design insight. Precise results can always be obtained using SPICE and Multisim simulations (see examples in Appendix B and the extensive material on the CD and the website). However, it is never a good idea to use simulation if one does not know beforehand approximate values of the expected results.

EXERCISES

13.14 For a CMOS inverter fabricated in a 0.18-μm process with $V_{DD} = 1.8$ V, $V_{tn} = -V_{tp} = 0.5$ V, $k_n' = 4k_p' = 300\ \mu\text{A/V}^2$ and having $(W/L)_n = 1.5$ and $(W/L)_p = 3$, find t_{PHL}, t_{PLH}, and t_P when the equivalent load capacitance $C = 10$ fF. Use the method of average currents.
Ans. 24.7 ps; 49.4 ps; 37 ps

D13.15 For a CMOS inverter fabricated in a 0.13-μm process, use the equivalent-resistances approach to determine $(W/L)_n$ and $(W/L)_p$ so that $t_{PLH} = t_{PHL} = 50$ ps when the effective load capacitance $C = 20$ fF.
Ans. 3.5; 8.3

13.3.2 Determining the Equivalent Load Capacitance C

Having determined the propagation delay of the CMOS inverter in terms of the equivalent load capacitance C, it now remains to determine the value of C. For this purpose, a thorough understanding of the various capacitances in a MOS transistor is essential, and we urge the reader to review the material in Section 9.2.1.

Figure 13.24 shows the circuit for determining the propagation delay of the CMOS inverter formed by Q_1 and Q_2. Note that we are showing the inverter driving a similar inverter formed by transistors Q_3 and Q_4. This reflects a practical situation and will help us explain how to determine the contribution of a driven inverter to the equivalent capacitance C at the output of the inverter under study (that formed by Q_1 and Q_2).

Indicated in Fig. 13.24 are the various transistor capacitances that connect to the output node of the Q_1–Q_2 inverter. Also shown is the **wiring capacitance** C_w, which represents the capacitance of the wire or **interconnect** that connects the output of the Q_1–Q_2 inverter to the input of the Q_3–Q_4 inverter. Interconnect capacitances have become increasingly dominant as the technology has scaled down. In fact, some digital IC designers hold the view that interconnect poses a greater limitation on the speed of operation than the transistors themselves. We will discuss this topic briefly in Section 13.5.

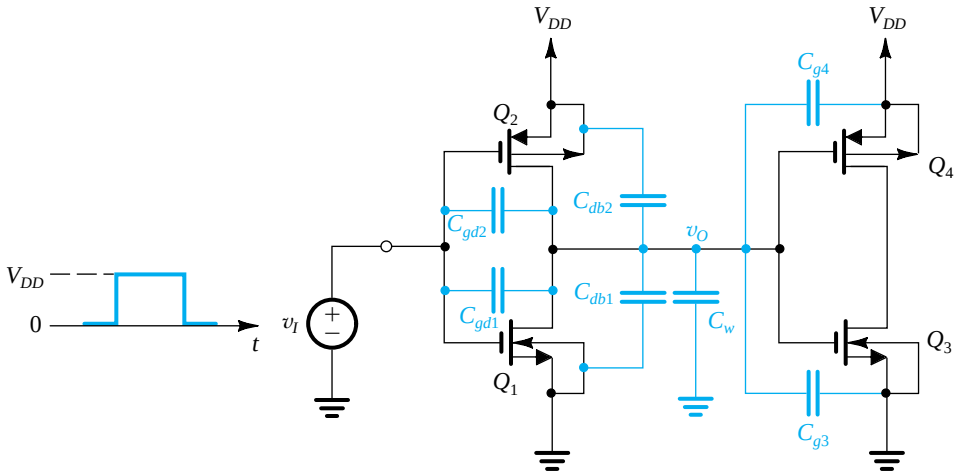


Figure 13.24 Circuit for analyzing the propagation delay of the inverter formed by Q_1 and Q_2, which is driving a similar inverter formed by Q_3 and Q_4.

A glance at the circuit in Fig. 13.24 should be sufficient to indicate that a pencil-and-paper analysis is virtually impossible. That, of course, is the reason we opted for the simplification of replacing all these capacitances with an equivalent capacitance C. Before we consider the determination of C, it is useful to observe that during t_{PLH} or t_{PHL}, the output of the first inverter changes from 0 to $V_{DD}/2$ or from V_{DD} to $V_{DD}/2$, respectively. It follows that the second inverter remains in the same state during each of our analysis intervals. This observation will have an important bearing on our estimation of the equivalent input capacitance of the second inverter. Let's now consider the contribution of each of the capacitances in Fig. 13.24 to the value of the equivalent load capacitance C:

1. The gate–drain overlap capacitance of Q_1, C_{gd1}, can be replaced by an equivalent capacitance between the output node and ground of $2C_{gd1}$. The factor 2 arises because of the Miller effect (Section 9.4.4). Specifically, refer to Fig. 13.25 and note that as v_I goes high and v_O goes low by the same amount, the change in voltage across C_{gd1} is twice that amount. Thus the output node sees in effect twice the value of C_{gd1}. The same applies for the gate–drain overlap capacitance of Q_2, C_{gd2}, which can be replaced by a capacitance $2C_{gd2}$ between the output node and ground.

2. Each of the drain–body capacitances C_{db1} and C_{db2} has a terminal at a constant voltage. Thus for the purpose of our analysis here, C_{db1} and C_{db2} can be replaced with equal capacitances between the output node and ground. Note, however, that the formulas given in Section 9.2.1 for calculating C_{db1} and C_{db2} are small-signal relationships, whereas the analysis here is obviously a large-signal one. A technique has been developed for finding equivalent large-signal values for C_{db1} and C_{db2} (see Hodges et al., (2004) and Rabaey et al., (2003)).
3. Since the second inverter does not switch states, we will assume that the input capacitances of Q_3 and Q_4 remain approximately constant and equal to the total gate capacitance $(WLC_{ox} + C_{gsov} + C_{gdov})$. That is, the input capacitance of the load inverter will be

$$C_{g3} + C_{g4} = (WL)_3 C_{ox} + (WL)_4 C_{ox} + C_{gsov3} + C_{gdov3} + C_{gsov4} + C_{gdov4} \quad (13.72)$$

4. The last component of C is the wiring capacitance C_w, which simply adds to the value of C.

Thus, the total value of C is given by

$$C = 2C_{gd1} + 2C_{gd2} + C_{db1} + C_{db2} + C_{g3} + C_{g4} + C_w \quad (13.73)$$

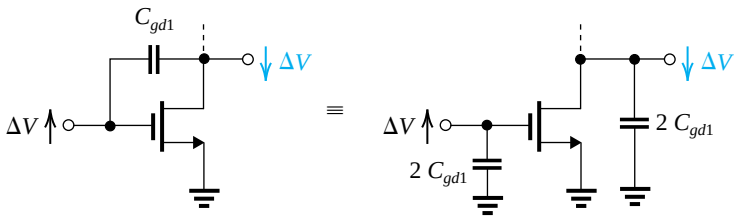


Figure 13.25 The Miller multiplication of the feedback capacitance C_{gd1}.

Example 13.6

Consider a CMOS inverter fabricated in a 0.25-μm process for which $C_{ox} = 6$ fF/μm^2, $\mu_n C_{ox} = 115$ μA/V^2, $\mu_p C_{ox} = 30$ μA/V^2, $V_{tn} = -V_{tp} = 0.5$ V, and $V_{DD} = 2.5$ V. The W/L ratio of Q_N is 0.375 μm/0.25 μm, and that for Q_P is 1.125 μm/0.25 μm. The gate–source and gate–drain overlap capacitances are specified to be 0.3 fF/μm of gate width. Further, the effective (large-signal) values of drain–body capacitances are $C_{dbn} = 1$ fF and $C_{dbp} = 1$ fF. The wiring capacitance $C_w = 0.2$ fF. Find t_{PHL}, t_{PLH}, and t_P when the inverter is driving an identical inverter.

Solution

First, we determine the value of the equivalent capacitance C using Eqs. (13.72) and (13.73),

$$C = 2C_{gd1} + 2C_{gd2} + C_{db1} + C_{db2} + C_{g3} + C_{g4} + C_w$$

where

$$C_{gd1} = 0.3 \times W_n = 0.3 \times 0.375 = 0.1125 \text{ fF}$$

$$C_{gd2} = 0.3 \times W_p = 0.3 \times 1.125 = 0.3375 \text{ fF}$$

$$C_{db1} = 1 \text{ fF}$$
$$C_{db2} = 1 \text{ fF}$$
$$C_{g3} = 0.375 \times 0.25 \times 6 + 2 \times 0.3 \times 0.375 = 0.7875 \text{ fF}$$
$$C_{g4} = 1.125 \times 0.25 \times 6 + 2 \times 0.3 \times 1.125 = 2.3625 \text{ fF}$$
$$C_w = 0.2 \text{ fF}$$

Thus,

$$C = 2 \times 0.1125 + 2 \times 0.3375 + 1 + 1 + 0.7875 + 2.3625 + 0.2 = 6.25 \text{ fF}$$

Next we use Eqs. (13.64) and (13.65) to determine t_{PHL},

$$\alpha_n = \frac{2}{\frac{7}{4} - \frac{3 \times 0.5}{2.5} + \left(\frac{0.5}{2.5}\right)^2} = 1.7$$

$$t_{PHL} = \frac{1.7 \times 6.25 \times 10^{-15}}{115 \times 10^{-6} \times (0.375/0.25) \times 2.5} = 24.6 \text{ ps}$$

Similarly, we use Eqs. (13.66) and (13.67) to determine t_{PLH},

$$\alpha_p = 1.7$$

$$t_{PLH} = \frac{1.7 \times 6.25 \times 10^{-15}}{30 \times 10^{-6} \times (1.125/0.25) \times 2.5} = 31.5 \text{ ps}$$

Finally, we determine t_P as

$$t_P = \frac{1}{2}(24.6 + 31.5) = 28 \text{ ps}$$

EXERCISES

13.16 Consider the inverter specified in Example 13.6 when loaded with an additional 0.1-pF capacitance. What will the propagation delay become?
Ans. 437 ps

13.17 In an attempt to decrease the area of the inverter in Example 13.6, $(W/L)_p$ is made equal to $(W/L)_n$. What is the percentage reduction in area achieved? Find the new values of C, t_{PHL}, t_{PLH}, and t_P. Assume that C_{dbp} does not change significantly.
Ans. 50%; 4.225 fF; 16.6 ps; 21.3 ps; 19 ps

13.18 For the inverter of Example 13.6, find the maximum frequency at which it can be operated.
Ans. 17.9 GHz

13.3.3 Inverter Sizing

In this section we address the question of selecting appropriate (W/L) ratios for the two transistors Q_N and Q_P in an inverter. Our reasoning can be summarized as follows.

1. To minimize area, the length of all channels is usually made equal to the minimum length permitted by the given technology.

2. In a given inverter, if our interest is strictly to minimize area, $(W/L)_n$ is usually selected in the range 1 to 1.5. The selection of $(W/L)_p$ relative to $(W/L)_n$ has influence on the noise margins and t_{PLH}. Both are optimized by matching Q_P and Q_N. This, however, is usually wasteful of area and equally important can increase the effective capacitance C, so that although t_{PLH} is made equal to t_{PHL}, the value of both can be higher than in the case without matching (see Problem 13.40). Thus, selecting $(W/L)_p = (W/L)_n$ is a possibility, and $(W/L)_p = 2(W/L)_n$ is a frequently used compromise.
3. Having settled on an appropriate ratio of $(W/L)_p$ to $(W/L)_n$, we still have to select $(W/L)_n$ to reduce t_P and thus allow higher speeds of operation. Any increase in $(W/L)_n$ and proportionally in $(W/L)_p$ will of course increase area, and hence the inverter contribution to the value of the equivalent capacitance C. To be more precise we express C as the sum of an intrinsic component C_{int} contributed by Q_N and Q_P of the inverter, and an extrinsic component C_{ext} resulting from the wiring and the input capacitance of the driven gates,

$$C = C_{\text{int}} + C_{\text{ext}} \tag{13.74}$$

Increasing $(W/L)_n$ and $(W/L)_p$ of the inverter by a factor S relative to that of a minimum size inverter for which $C_{\text{int}} = C_{\text{int0}}$ results in

$$C = SC_{\text{int0}} + C_{\text{ext}} \tag{13.75}$$

Now, if we use the equivalent-resistances approach to compute t_P and define an equivalent inverter resistance R_{eq} as

$$R_{\text{eq}} = \frac{1}{2}(R_N + R_P) \tag{13.76}$$

then,

$$t_P = 0.69R_{\text{eq}}C \tag{13.77}$$

Further, if for the minimum-size inverter R_{eq} is R_{eq0}, increasing $(W/L)_n$ and $(W/L)_p$ by the factor S reduces R_{eq} by the same factor:

$$R_{\text{eq}} = R_{\text{eq0}}/S \tag{13.78}$$

Combining Eqs. (13.77), (13.78), and (13.75), we obtain

$$t_P = 0.69\left(\frac{R_{\text{eq0}}}{S}\right)(SC_{\text{int0}} + C_{\text{ext}})$$

$$t_P = 0.69\left(R_{\text{eq0}}C_{\text{int0}} + \frac{1}{S}\,R_{\text{eq0}}C_{\text{ext}}\right) \tag{13.79}$$

We thus see that scaling the W/L ratios does *not* change the component of t_P caused by the capacitances of Q_N and Q_P. It does, however, reduce the component of t_P that results from capacitances external to the inverter itself. It follows that one can use

Eq. (13.79) to decide on a suitable scaling factor S that keeps t_P below a specified maximum value, keeping in mind of course the effect of increasing S on silicon area.

EXERCISE

13.19 For the inverter analyzed in Example 13.6:
(a) Find the intrinsic and extrinsic components of C.
(b) By what factor must $(W/L)_n$ and $(W/L)_p$ be increased to reduce the extrinsic part of t_P by a factor of 2?
(c) Estimate the resulting t_P.
(d) By what factor is the inverter area increased?
Ans. (a) 2.9 fF, 3.35 fF; (b) 2; (c) 20.5 ps; (d) 2

13.3.4 Dynamic Power Dissipation

The negligible static power dissipation of CMOS has been a significant factor in its dominance as the technology of choice in implementing high-density VLSI circuits. However, as the number of gates per chip steadily increases, the dynamic power dissipation has become a serious issue. The dynamic power dissipated in the CMOS inverter is given by Eq. (13.35), which we repeat here as

$$P_{\text{dyn}} = fCV_{DD}^2 \tag{13.80}$$

where f is the frequency at which the gate is switched. It follows that minimizing C is an effective means for reducing dynamic-power dissipation. An even more effective strategy is the use of a lower power-supply voltage. As we have mentioned, CMOS process technologies now utilize V_{DD} values of 1 V or less. These newer chips, however, pack much more circuitry on the chip (as many as 2.3 billion transistors) and operate at higher frequencies (microprocessor clock frequencies above 3 GHz are now available). The dynamic power dissipation of such high-density chips can be over 100 W.

In addition to the dynamic power dissipation that results from the periodic charging and discharging of the inverter load capacitance, there is another component of power dissipation that results from the current that flows through Q_P and Q_N during every switching event. Figure 13.26 shows this inverter current as a function of the input voltage v_I for a matched inverter. We note that the current peaks at $V_M = V_{DD}/2$. Since at this voltage both Q_N and Q_P operate in saturation, the peak current is given by

$$I_{\text{peak}} = \frac{1}{2}\mu_n C_{ox}\left(\frac{W}{L}\right)_n \left(\frac{V_{DD}}{2} - V_{tn}\right)^2 \tag{13.81}$$

The width of the current pulse will depend on the rate of change of v_I with time; the slower the rising edge of the input waveform, the wider the current pulse and the greater the energy drawn from the supply. In general, however, this power component is usually much smaller than P_{dyn}.

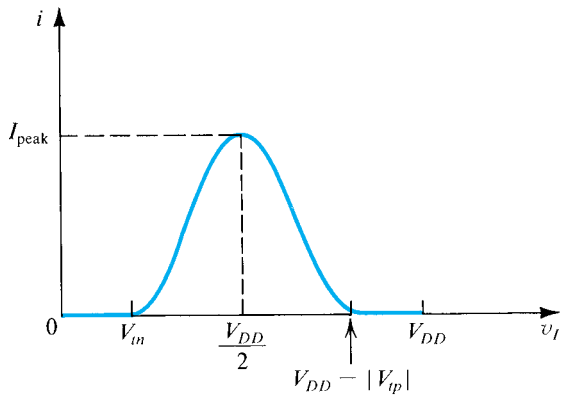


Figure 13.26 The current in the CMOS inverter versus the input voltage.

EXERCISE

13.20 Find the dynamic power dissipation of the inverter analyzed in Example 13.6 when operated at a 1-GHz frequency. If this inverter is switched at its maximum possible operating frequency, what is the value of the power–delay product?
Ans. 39 μW ; 19.5 fJ

13.4 CMOS Logic-Gate Circuits

In this section, we build on our knowledge of inverter design and consider the design of CMOS circuits that realize combinational-logic functions. In combinational circuits, the output at any time is a function only of the values of input signals at that time. Thus, these circuits do not have memory and do not employ feedback. Combinational-logic circuits are used in large quantities in a multitude of applications; indeed, every digital system contains large numbers of combinational-logic circuits.

13.4.1 Basic Structure

A CMOS logic circuit is in effect an extension, or a generalization, of the CMOS inverter: The inverter consists of an NMOS pull-down transistor, and a PMOS pull-up transistor, operated by the input voltage in a complementary fashion. The CMOS logic gate consists of two networks: the **pull-down network** (**PDN**) constructed of NMOS transistors, and the **pull-up network** (**PUN**) constructed of PMOS transistors (see Fig. 13.27). The two networks are operated by the input variables, in a complementary fashion. Thus, for the three-input gate represented in Fig. 13.27, the PDN will conduct for all input combinations that require a low output ($Y = 0$) and will then pull the output node down to ground,

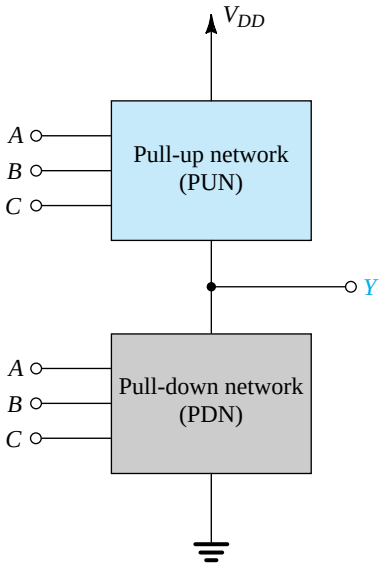


Figure 13.27 Representation of a three-input CMOS logic gate. The PUN comprises PMOS transistors, and the PDN comprises NMOS transistors.

causing a zero voltage to appear at the output, $v_Y = 0$. Simultaneously, the PUN will be off, and no direct dc path will exist between V_{DD} and ground. On the other hand, all input combinations that call for a high output ($Y = 1$) will cause the PUN to conduct, and the PUN will then pull the output node up to V_{DD}, establishing an output voltage $v_Y = V_{DD}$. Simultaneously, the PDN will be cut off, and again, no dc current path between V_{DD} and ground will exist in the circuit.

Now, since the PDN comprises NMOS transistors, and since an NMOS transistor conducts when the signal at its gate is high, the PDN is activated (i.e., conducts) when the inputs are high. In a dual manner, the PUN comprises PMOS transistors, and a PMOS transistor conducts when the input signal at its gate is low; thus the PUN is activated when the inputs are low.

The PDN and the PUN each utilizes devices in parallel to form an OR function, and devices in series to form an AND function. Here, the OR and AND notation refer to current flow or conduction. Figure 13.28 shows examples of PDNs. For the circuit in Fig. 13.28(a), we observe that Q_A will conduct when A is high ($v_A = V_{DD}$) and will then pull the output node down to ground ($v_Y = 0$ V, $Y = 0$). Similarly, Q_B conducts and pulls Y down when B is high. Thus Y will be low when A is high *or* B is high, which can be expressed as

$$\overline{Y} = A + B$$

or equivalently

$$Y = \overline{A + B}$$

The PDN in Fig. 13.28(b) will conduct only when A and B are both high simultaneously. Thus Y will be low when A is high *and* B is high,

$$\overline{Y} = AB$$

or equivalently

$$Y = \overline{AB}$$

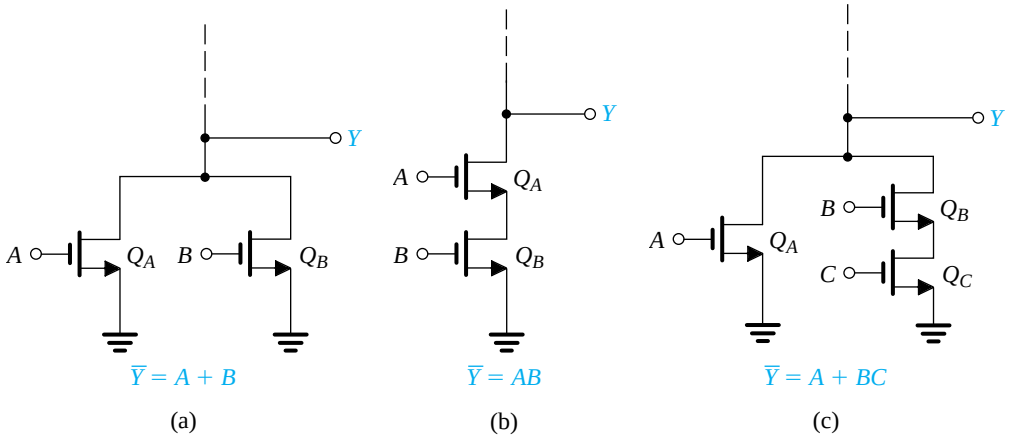


Figure 13.28 Examples of pull-down networks.

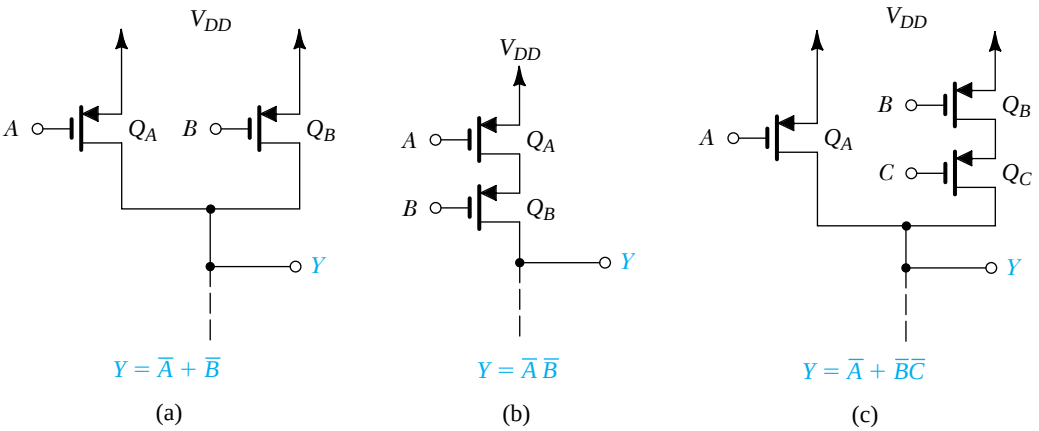


Figure 13.29 Examples of pull-up networks.

As a final example, the PDN in Fig. 13.28(c) will conduct and cause Y to be 0 when A is high *or* when B *and* C are both high, thus

$$\overline{Y} = A + BC$$

or equivalently

$$Y = \overline{A + BC}$$

Next consider the PUN examples shown in Fig. 13.29. The PUN in Fig. 13.29(a) will conduct and pull Y up to V_{DD}($Y = 1$) when A is low *or* B is low, thus

$$Y = \overline{A} + \overline{B}$$

The PUN in Fig. 13.29(b) will conduct and produce a high output ($v_Y = V_{DD}$, $Y = 1$) only when A *and* B are both low, thus

$$Y = \overline{A}\,\overline{B}$$

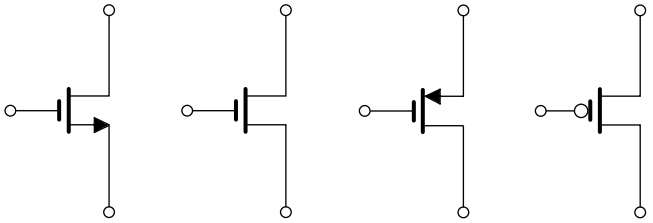


Figure 13.30 Usual and alternative circuit symbols for MOSFETs.

Finally, the PUN in Fig. 13.29(c) will conduct and cause Y to be high (logic 1) if A is low *or* if B *and* C are both low; thus,

$$Y = \overline{A} + \overline{B}\overline{C}$$

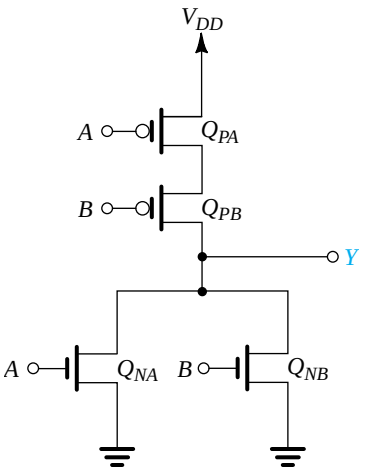
Having developed an understanding and an appreciation of the structure and operation of PDNs and PUNs, we now consider complete CMOS gates. Before doing so, however, we wish to introduce alternative circuit symbols, that are almost universally used for MOS transistors by digital-circuit designers. Figure 13.30 shows our usual symbols (left) and the corresponding "digital" symbols (right). Observe that the symbol for the PMOS transistor with a circle at the gate terminal is intended to indicate that the signal at the gate has to be low for the device to be activated (i.e., to conduct). Thus, in terms of logic-circuit terminology, the gate terminal of the PMOS transistor is an *active low* input. Besides indicating this property of PMOS devices, the digital symbols omit any indication of which of the device terminals is the source and which is the drain. This should cause no difficulty at this stage of our study; simply remember that for an NMOS transistor, the drain is the terminal that is at the higher voltage (current flows from drain to source), and for a PMOS transistor the source is the terminal that is at the higher voltage (current flows from source to drain). To be consistent with the literature, we shall henceforth use these modified symbols for MOS transistors in logic applications, except in locations where our usual symbols help in understanding circuit operation.

13.4.2 The Two-Input NOR Gate

We first consider the CMOS gate that realizes the two-input NOR function

$$Y = \overline{A + B} = \overline{A}\overline{B} \tag{13.82}$$

We see that Y is to be low (PDN conducting) when A is high or B is high. Thus the PDN consists of two parallel NMOS devices with A and B as inputs (i.e., the circuit in Fig. 13.28a). For the PUN, we note from the second expression in Eq. (13.82) that Y is to be high when A and B are both low. Thus the PUN consists of two series PMOS devices with A and B as the inputs (i.e., the circuit in Fig. 13.29b). Putting the PDN and the PUN together gives the CMOS NOR gate shown in Fig. 13.31. Note that extension to a higher number of inputs is straightforward: For each additional input, an NMOS transistor is added in parallel with Q_{NA} and Q_{NB}, and a PMOS transistor is added in series with Q_{PA} and Q_{PB}.



$Y = \overline{A + B}$

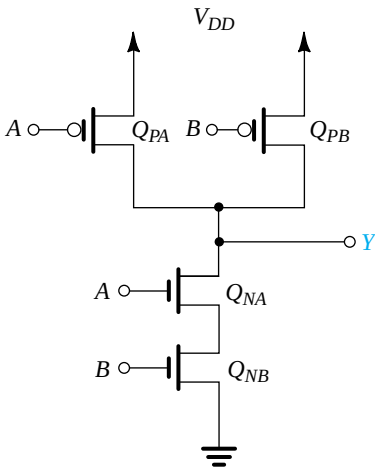
Figure 13.31 A two-input CMOS NOR gate.

13.4.3 The Two-Input NAND Gate

The two-input NAND function is described by the Boolean expression

$$Y = \overline{AB} = \bar{A} + \bar{B} \tag{13.83}$$

To synthesize the PDN, we consider the input combinations that require Y to be low: There is only one such combination, namely, A and B both high. Thus, the PDN simply comprises two NMOS transistors in series (such as the circuit in Fig. 13.28b). To synthesize the PUN, we consider the input combinations that result in Y being high. These are found from the second expression in Eq. (13.83) as A low or B low. Thus, the PUN consists of two parallel PMOS transistors with A and B applied to their gates (such as the circuit in Fig. 13.29a). Putting the PDN and PUN together results in the CMOS NAND gate implementation shown in Fig. 13.32. Note that extension to a higher number of inputs is straightforward: For each



$Y = \overline{AB}$

Figure 13.32 A two-input CMOS NAND gate.

additional input, we add an NMOS transistor in series with Q_{NA} and Q_{NB}, and a PMOS transistor in parallel with Q_{PA} and Q_{PB}.

13.4.4 A Complex Gate

Consider next the more complex logic function

$$Y = \overline{A(B + CD)} \tag{13.84}$$

Since $\bar{Y} = A(B + CD)$, we see that Y should be low for A high and simultaneously either B high or C and D both high, from which the PDN is directly obtained. To obtain the PUN, we need to express Y in terms of the complemented variables. We do this through repeated application of DeMorgan's law, as follows:

$$\begin{aligned} Y &= \overline{A(B + CD)} \\ &= \bar{A} + \overline{B + CD} \\ &= \bar{A} + \bar{B}\ \overline{CD} \\ &= \bar{A} + \bar{B}(\bar{C} + \bar{D}) \end{aligned} \tag{13.85}$$

Thus, Y is high for A low or B low and either C or D low. The corresponding complete CMOS circuit will be as shown in Fig. 13.33.

13.4.5 Obtaining the PUN from the PDN and Vice Versa

From the CMOS gate circuits considered thus far (e.g., that in Fig. 13.33), we observe that the PDN and the PUN are dual networks: Where a series branch exists in one, a parallel branch exists in the other. Thus, we can obtain one from the other, a process that can be simpler than having to synthesize each separately from the Boolean expression of the function. For instance, in the circuit of Fig. 13.33, we found it relatively easy to obtain the PDN, simply because we already had $\bar{Y}$ in terms of the uncomplemented inputs. On the other hand, to obtain the PUN, we had to manipulate the given Boolean expression to express Y as a function of the complemented variables, the form convenient for synthesizing PUNs. Alternatively, we could have used this duality property to obtain the PUN from the PDN. The reader is urged to refer to Fig. 13.33 to convince herself that this is indeed possible.

It should, however, be mentioned that at times it is not easy to obtain one of the two networks from the other using the duality property. For such cases, one has to resort to a more rigorous process, which is beyond the scope of this book (see Kang and Leblebici, 1999).

13.4.6 The Exclusive-OR Function

An important function that often arises in logic design is the exclusive-OR (XOR) function,

$$Y = A\bar{B} + \bar{A}B \tag{13.86}$$

We observe that since Y (rather than $\bar{Y}$) is given, it is easier to synthesize the PUN. We note, however, that unfortunately Y is not a function of the complemented variables only (as we

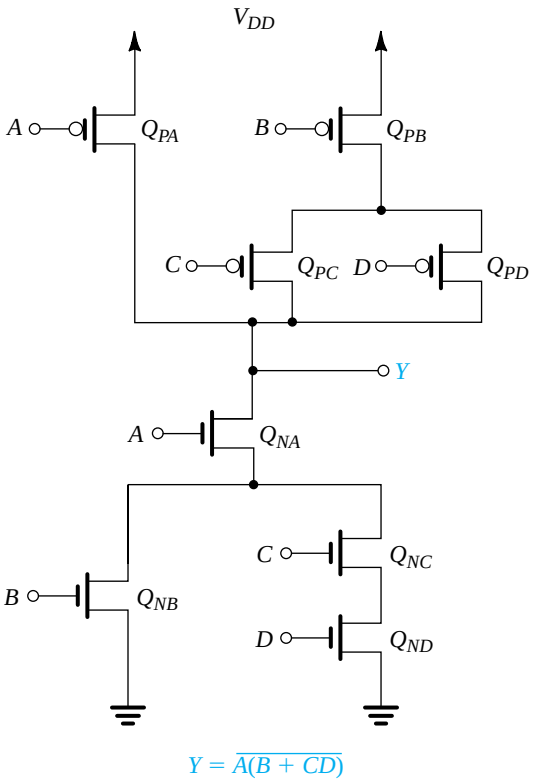


Figure 13.33 CMOS realization of a complex gate.

would like it to be). Thus, we will need additional inverters. The PUN obtained directly from Eq. (13.86) is shown in Fig. 13.34(a). Note that the Q_1, Q_2 branch realizes the first term ($A\bar{B}$), whereas the Q_3, Q_4 branch realizes the second term ($\bar{A}B$). Note also the need for two additional inverters to generate $\bar{A}$ and $\bar{B}$.

As for synthesizing the PDN, we can obtain it as the dual network of the PUN in Fig. 13.34(a). Alternatively, we can develop an expression for $\bar{Y}$ and use it to synthesize the PDN. Leaving the first approach for the reader to do as an exercise, we shall utilize the direct synthesis approach. DeMorgan's law can be applied to the expression in Eq. (13.86) to obtain $\bar{Y}$ as

$$\bar{Y} = AB + \bar{A}\bar{B} \tag{13.87}$$

The corresponding PDN will be as in Fig. 13.34(b), which shows the CMOS realization of the exclusive-OR function except for the two additional inverters. Note that the exclusive-OR requires 12 transistors for its realization, a rather complex network. Later, in Section 14.2, we shall show a simpler realization of the XOR employing a different form of CMOS logic.

Another interesting observation follows from the circuit in Fig. 13.34(b). The PDN and the PUN here are *not* dual networks. Indeed, duality of the PDN and the PUN is not a necessary condition. Thus, although a dual of PDN (or PUN) can always be used for PUN (or PDN), the two networks are not necessarily duals.

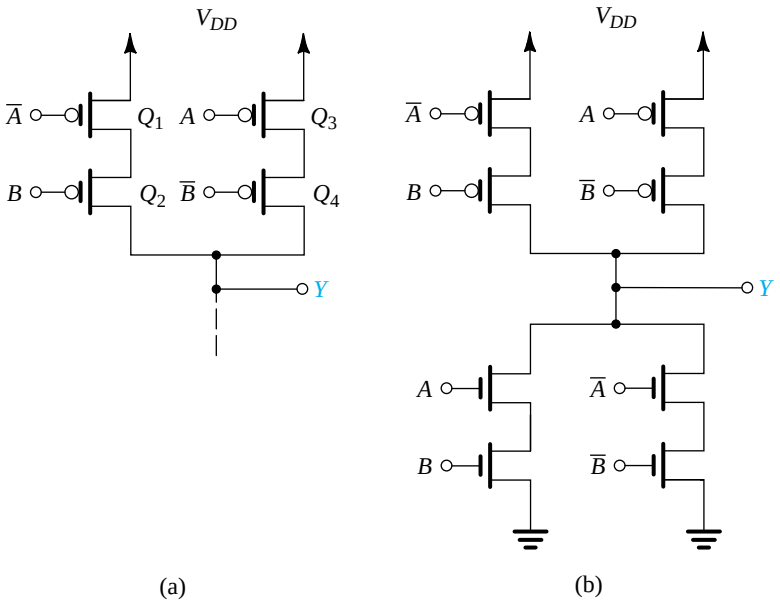


Figure 13.34 Realization of the exclusive-OR (XOR) function: **(a)** The PUN synthesized directly from the expression in Eq. (13.86). **(b)** The complete XOR realization utilizing the PUN in (a) and a PDN that is synthesized directly from the expression in Eq. (13.87). Note that two inverters (not shown) are needed to generate the complemented variables. Also note that in this XOR realization, the PDN and the PUN are not dual networks; however, a realization based on dual networks is possible (see Problem 13.47).

13.4.7 Summary of the Synthesis Method

1. The PDN can be most directly synthesized by expressing $\overline{Y}$ as a function of the *uncomplemented* variables. If complemented variables appear in this expression, additional inverters will be required to generate them.
2. The PUN can be most directly synthesized by expressing Y as a function of the *complemented* variables and then applying the uncomplemented variables to the gates of the PMOS transistors. If uncomplemented variables appear in the expression, additional inverters will be needed.
3. The PDN can be obtained from the PUN (and vice versa) using the duality property.

13.4.8 Transistor Sizing

Once a CMOS gate circuit has been generated, the only significant step remaining in the design is to decide on W/L ratios for all devices. These ratios usually are selected to provide the gate with current-driving capability in both directions equal to that of the basic inverter. For the basic inverter design, denote $(W/L)_n = n$ and $(W/L)_p = p$, where n is usually 1 to 1.5 and, for a matched design, $p = (\mu_n/\mu_p)\,n$; although often $p = 2n$ and for minimum area $p = n$. Thus, we wish to select individual W/L ratios for all transistors in a logic gate so that the PDN should be able to provide a capacitor discharge current *at least* equal to that of an NMOS transistor with $W/L = n$, and the PUN should be able to

provide a charging current *at least* equal to that of a PMOS transistor with $W/L = p$. This will guarantee a *worst-case* gate delay equal to that of the basic inverter.[4]

In the preceding description, the idea of "worst case" should be emphasized. It means that in deciding on device sizing, we should find the input combinations that result in the lowest output current and then choose sizes that will make this current equal to that of the basic inverter. Before we consider examples, we need to address the issue of determining the current-driving capability of a circuit consisting of a number of MOS devices. In other words, we need to find the *equivalent* W/L *ratio* of a network of MOS transistors. Toward that end, we consider the parallel and series connection of MOSFETs and find the equivalent W/L ratios.

The derivation of the equivalent W/L ratio is based on the fact that the on resistance of a MOSFET is inversely proportional to W/L (see Eqs. 13.70 and 13.71). Thus, if a number of MOSFETs having ratios of $(W/L)_1$, $(W/L)_2$, . . ., are connected in series, the equivalent series resistance obtained by adding the on-resistances will be

$$\begin{aligned} R_{\text{series}} &= R_{N1} + R_{N2} + \cdots \\ &= \frac{\text{constant}}{(W/L)_1} + \frac{\text{constant}}{(W/L)_2} + \cdots \\ &= \text{constant}\left[\frac{1}{(W/L)_1} + \frac{1}{(W/L)_2} + \cdots\right] \\ &= \frac{\text{constant}}{(W/L)_{\text{eq}}} \end{aligned}$$

resulting in the following expression for $(W/L)_{\text{eq}}$ for transistors connected in series:

$$(W/L)_{\text{eq}} = \frac{1}{\dfrac{1}{(W/L)_1} + \dfrac{1}{(W/L)_2} + \cdots} \tag{13.88}$$

Similarly, we can show that the parallel connection of transistors with W/L ratios of $(W/L)_1$, $(W/L)_2$, . . . , results in an equivalent W/L of

$$(W/L)_{\text{eq}} = (W/L)_1 + (W/L)_2 + \cdots \tag{13.89}$$

As an example, two identical MOS transistors with individual W/L ratios of 4 result in an equivalent W/L of 2 when connected in series and of 8 when connected in parallel.[5]

As an example of proper sizing, consider the four-input NOR in Fig. 13.35. Here, the worst case (the lowest current) for the PDN is obtained when only one of the NMOS transistors is conducting. We therefore select the W/L of each NMOS transistor to be equal to that of the NMOS transistor of the basic inverter, namely, n. For the PUN, however, the worst-case situation (and indeed the only case) occurs when all inputs are low and the four series PMOS transistors are conducting. Since the equivalent W/L will be one-quarter of that of

[4]This statement assumes that the total effective capacitance C of the logic gate is the same as that of the inverter. In actual practice, the value of C will be larger for a gate, especially as the fan-in is increased.
[5]Another way of thinking about this is as follows: Connecting MOS transistors in series is equivalent to adding the lengths of their channels while the width does not change; connecting MOS transistors in parallel does not change the channel length but increases the width to the sum of the W's.

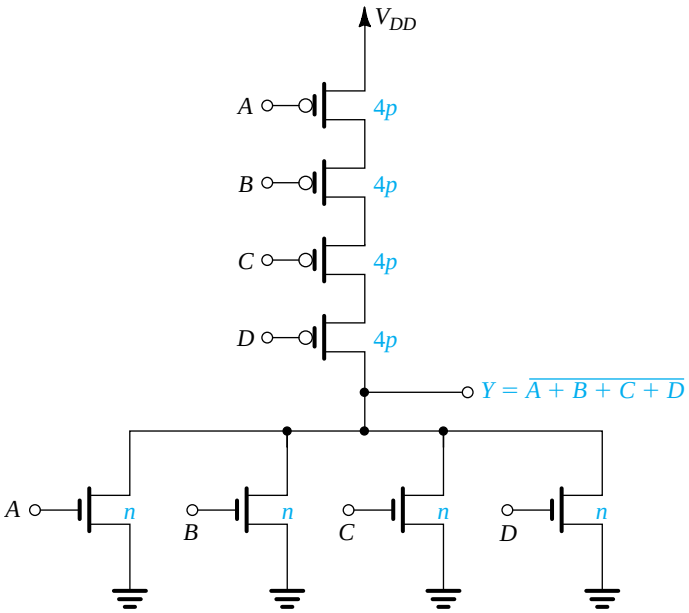


Figure 13.35 Proper transistor sizing for a four-input NOR gate. Note that n and p denote the W/L ratios of Q_N and Q_P, respectively, of the basic inverter.

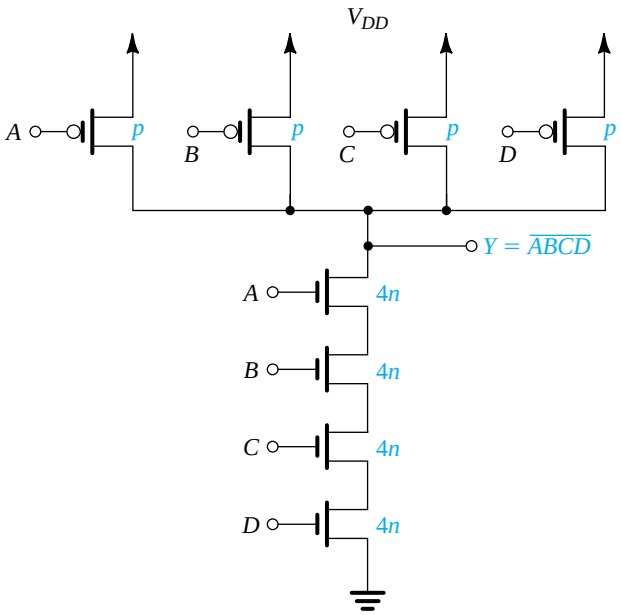


Figure 13.36 Proper transistor sizing for a four-input NAND gate. Note that n and p denote the W/L ratios of Q_N and Q_P, respectively, of the basic inverter.

each PMOS device, we should select the W/L ratio of each PMOS transistor to be four times that of Q_P of the basic inverter, that is, $4p$.

As another example, we show in Fig. 13.36 the proper sizing for a four-input NAND gate. Comparison of the NAND and NOR gates in Figs. 13.35 and 13.36 indicates that

because p is usually two to three times n, the NOR gate will require much greater area than the NAND gate. For this reason, NAND gates are generally preferred for implementing combinational logic functions in CMOS.

Example 13.7

Provide transistor W/L ratios for the logic circuit shown in Fig. 13.37. Assume that for the basic inverter $n = 1.5$ and $p = 5$ and that the channel length is 0.25 μm.

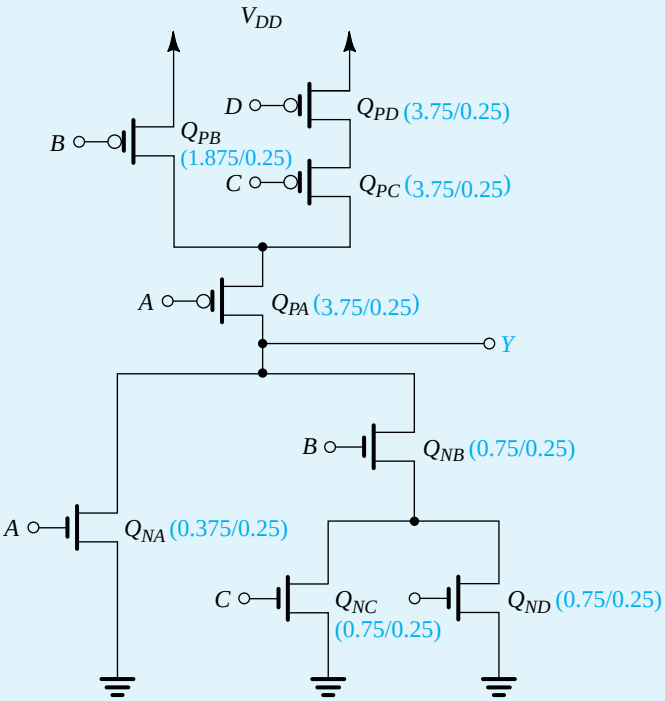


Figure 13.37 Circuit for Example 13.7.

Solution

Refer to Fig. 13.37, and consider the PDN first. We note that the worst case occurs when Q_{NB} is on and either Q_{NC} or Q_{ND} is on. That is, in the worst case, we have two transistors in series. Therefore, we select each of Q_{NB}, Q_{NC}, and Q_{ND} to have twice the width of the n-channel device in the basic inverter, thus

$$Q_{NB}: \; W/L = 2n = 3 = 0.75/0.25$$

$$Q_{NC}: \; W/L = 2n = 3 = 0.75/0.25$$

$$Q_{ND}: \; W/L = 2n = 3 = 0.75/0.25$$

For transistor Q_{NA}, select W/L to be equal to that of the n-channel device in the basic inverter:

$$Q_{NA}: \; W/L = n = 1.5 = 0.375/0.25$$

Next, consider the PUN. Here, we see that in the worst case, we have three transistors in series: Q_{PA}, Q_{PC}, and Q_{PD}. Therefore, we select the W/L ratio of each of these to be three times that of Q_P in the basic inverter, that is, $3p$, thus

$$Q_{PA}: \ W/L = 3p = 15 = 3.75/0.25$$

$$Q_{PC}: \ W/L = 3p = 15 = 3.75/0.25$$

$$Q_{PD}: \ W/L = 3p = 15 = 3.75/0.25$$

Finally, the W/L ratio for Q_{PB} should be selected so that the equivalent W/L of the series connection of Q_{PB} and Q_{PA} should be equal to p. It follows that for Q_{PB} the ratio should be $1.5p$,

$$Q_{PB}: \ W/L = 1.5p = 7.5 = 1.875/0.25$$

Figure 13.37 shows the circuit with the transistor sizes indicated.

13.4.9 Effects of Fan-In and Fan-Out on Propagation Delay

Each additional input to a CMOS gate requires two additional transistors, one NMOS and one PMOS. This is in contrast to other forms of MOS logic, where each additional input requires only one additional transistor (see Section 14.1). The additional transistor in CMOS not only increases the chip area but also increases the total effective capacitance per gate and in turn increases the propagation delay. The size-scaling method described earlier compensates for some (but not all) of the increase in t_p. Specifically, by increasing device size, we are able to preserve the current-driving capability. However, the capacitance C increases because of both the increased number of inputs and the increase in device size. Thus t_p will still increase with fan-in, a fact that imposes a practical limit on the fan-in of, say, the NAND gate to about 4. If a higher number of inputs is required, then "clever" logic design should be adopted to realize the given Boolean function with gates of no more than four inputs. This would usually mean an increase in the number of cascaded stages and thus an increase in delay. However, such an increase in delay can be less than the increase due to the large fan-in (see Problem 13.56).

An increase in a gate's fan-out adds directly to its load capacitance and, thus, increases its propagation delay.

Thus although CMOS has many advantages, it does suffer from increased circuit complexity when the fan-in and fan-out are increased, and from the corresponding effects of this complexity on both chip area and propagation delay. Later, in Sections 14.1 and 14.2, we shall study some simplified forms of CMOS logic that attempt to reduce this complexity, although at the expense of forgoing some of the advantages of basic CMOS.

EXERCISES

13.21 For a process technology with $L = 0.18$ μm, $n = 1.5$, $p = 3$, give the sizes of all transistors in (a) a four-input NOR and (b) a four-input NAND. Also, give the relative areas of the two gates.

Ans. (a) NMOS devices: $W/L = 0.27/0.18$, PMOS devices: $2.16/0.18$;
(b) NMOS devices: $W/L = 1.08/0.18$, PMOS devices: $0.54/0.18$;
NOR area/NAND area = 1.5

13.22 For the scaled NAND gate in Exercise 13.21, find the ratio of the maximum to minimum current available to (a) charge a load capacitance and (b) discharge a load capacitance.
Ans. (a) 4; (b) 1

⊕ 13.5 Implications of Technology Scaling: Issues in Deep-Submicron Design

As mentioned in Chapter 4, and in a number of locations throughout the book, the minimum MOSFET channel length has been continually reduced over the past 40 years or so. In fact, a new CMOS fabrication technology has been introduced every 2 or 3 years, with the minimum allowable channel length reduced by about 30%, that is, to 0.7 the value in the preceding generation. Thus, with every new **technology generation**, the device area has been reduced by a factor of $1/(0.7 \times 0.7)$ or approximately 2, allowing the fabrication of twice as many devices on a chip of the same area. This astounding phenomenon, predicted more than 40 years ago by Gordon Moore,[6] has become known as **Moore's law**. It is this ability to pack an exponentially increasing number of transistors on an IC chip that has resulted in the continuing reduction in the cost per logic function.

Figure 13.38 shows the exponential reduction in MOSFET channel length (by a factor of 2 every 5 years) over a 40 year period, with the dots indicating some of the prominent **technology generations**, or **nodes**. Thus, we see the 10-μm process of the early 1970s, the submicron $(L < 1\ \mu\text{m})$ processes of the early 1990s, and the deep-submicron $(L < 0.25\ \mu\text{m})$ processes of the last decade, including the current 45-nm process. A microprocessor chip fabricated in a 45-nm CMOS process and having 2.3 billion transistors was announced in 2009. Deep-submicron (DSM) processes present the circuit designer with a host of new opportunities and challenges. It is our purpose in this section to briefly consider some of these.

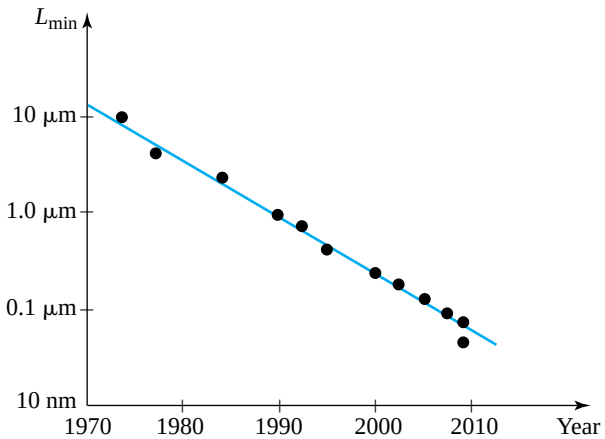


Figure 13.38 The MOSFET channel length has been reduced by a factor of 2 every about 5 years. This phenomenon, known as Moore's law is continuing.

[6]Gordon Moore is one of the pioneers of the semiconductor industry and a cofounder of Intel.

13.5.1 Scaling Implications

Table 13.2 provides a summary of the implications of scaling the device dimensions by a factor $1/S$, where $S > 1$. As well, we assume that V_{DD} and V_t are scaled by the same factor. Although the scaling of V_{DD} has occurred for a number of technology nodes (e.g., from 5 V for the 0.5-µm process down to 1.2 V for the 0.13-µm process), V_t has been reduced but not by the same factor. Thus the assumption in row 2 of Table 13.2 is not entirely correct. Nevertheless, our interest here is to gain a general appreciation for the effects of scaling.

Table 13.2 provides the relationships for the various transistor and inverter parameters in order to show how the resulting scale factors are obtained. We thus see that the device area scales by $1/S^2$; the oxide capacitance C_{ox}, and the transconductance parameters k_n' and k_p' scale by S; and the MOSFET gate capacitance scales by $1/S$. It is important to note that the component of the inverter propagation delay due to the transistor capacitances (i.e., excluding the wiring capacitance) scales by $1/S$; this very useful result of scaling implies that the circuit can be operated at S times the frequency; that is, the speed of operation increases by a factor S. Equally important, the dynamic power dissipation scales by $1/S^2$. This, of course, is a major motivating factor behind the scaling of V_{DD}. Another motivating factor is the need to keep the electric fields in the MOSFETs within acceptable bounds.

Although the dynamic power dissipation is scaled by $1/S^2$, the power per unit area remains unchanged. Nevertheless, for a number of reasons, as the size and complexity of digital IC chips continue to increase, so does their power dissipation. Indeed power dissipation has now become the number-one issue in IC design. The problem is further exacerbated by the static power dissipation, arising from both subthreshold conduction and diode leakage currents, that plagues deep-submicron CMOS devices. We will discuss this issue shortly.

Table 13.2 Implications of Device and Voltage Scaling

	Parameter	Relationship	Scaling Factor
1	W, L, t_{ox}		$1/S$
2	V_{DD}, V_t		$1/S$
3	Area/Device	WL	$1/S^2$
4	C_{ox}	ε_{ox}/t_{ox}	S
5	k_n', k_p'	$\mu_n C_{ox}, \mu_p C_{ox}$	S
6	C_{gate}	WLC_{ox}	$1/S$
7	t_P (intrinsic)	$\alpha C/k'V_{DD}$	$1/S$
8	Energy/Switching cycle (intrinsic)	CV_{DD}^2	$1/S^3$
9	P_{dyn}	$f_{\max} CV_{DD}^2 = \dfrac{CV_{DD}^2}{2t_P}$	$1/S^2$
10	Power density	P_{dyn}/Device area	1

EXERCISES

13.23 By what factor does the power–delay product *PDP* change if an inverter is fabricated in a 0.13 μm technology rather than a 0.25-μm technology? Assume $S \simeq 2$.
Ans. *PDP* decreases by a factor of 8.

13.24 If V_{DD} and V_t are kept constant, which entries in Table 13.2 change and to what value?
Ans. t_P now scales by $1/S^2$; the energy/switching cycle now scales by $1/S$ only; P_{dyn} now scales by S; and the power density now scales by S^3 (a major problem).

13.5.2 Velocity Saturation

The short channels of MOSFETs fabricated in deep-submicron processes give rise to physical phenomena not present in long-channel devices, and thus to changes in the MOSFET i–v characteristics. The most important of these **short-channel** effects is **velocity saturation**. Here we refer to the drift velocity of electrons in the channel of an NMOS transistor (holes in PMOS) under the influence of the longitudinal electric field established by v_{DS}. In our derivation of the MOSFET i–v characteristics in Section 5.1, we assumed that the velocity v_n of the electrons in an n-channel device is given by

$$v_n = \mu_n E \tag{13.90}$$

where E is the electric field given by

$$E = \frac{v_{DS}}{L} \tag{13.91}$$

The relationship in Eq. (13.90) applies as long as E is below a critical value E_{cr} which falls in the range 1 V/μm to 5 V/μm. For $E > E_{cr}$, the drift velocity saturates at a value v_{sat} of approximately 10^7 cm/s. Figure 13.39 shows a sketch of v versus E. Although the change from a linear to a constant v is gradual, we shall assume for simplicity that v saturates abruptly at $E = E_{cr}$.

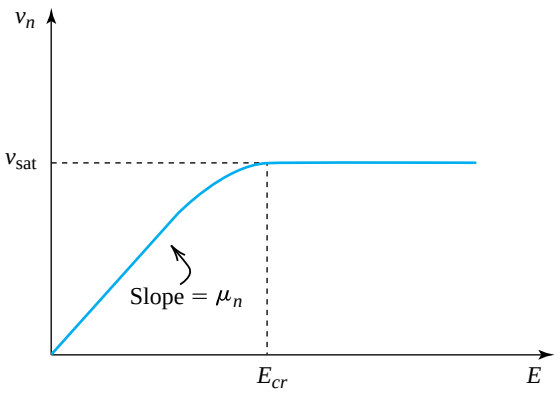


Figure 13.39 The velocity of electrons in the channel of an NMOS transistor reach a constant velocity $v_{\text{sat}} \simeq 10^7$ cm/s when the electric field E reaches a critical value E_{cr}. A similar situation occurs for *p*-channel devices.

The electric field E in a short-channel MOSFET can easily exceed E_{cr} even though V_{DD} is low. If we denote the value of v_{DS} at which velocity saturation occurs by V_{DSsat}, then from Eq. (13.91),

$$E_{cr} = \frac{V_{DSsat}}{L} \tag{13.92}$$

which when substituted in Eq. (13.90) provides

$$v_{sat} = \mu_n\left(\frac{V_{DSsat}}{L}\right) \tag{13.93}$$

or alternatively,

$$v_{DSsat} = \left(\frac{L}{\mu_n}\right)v_{sat} \tag{13.94}$$

Thus, V_{DSsat} is a device parameter.

EXERCISE

13.25 Find V_{DSsat} for an NMOS transistor fabricated in a 0.25-μm CMOS process with $\mu_n = 400\ \text{cm}^2/\text{V}\cdot\text{s}$. Let $L = 0.25\ \mu\text{m}$ and assume $v_{sat} = 10^7\ \text{cm/s}$.
Ans. 0.63 V

The i_D–v_{DS} Characteristics The $i_D - v_{DS}$ equations of the MOSFET can be modified to include velocity saturation as follows. Consider a long-channel NMOS transistor operating in the triode region with v_{GS} set to a constant value V_{GS}. The drain current will be

$$i_D = \mu_n C_{ox}\left(\frac{W}{L}\right)v_{DS}\left[(V_{GS} - V_t) - \frac{1}{2}v_{DS}\right] \tag{13.95}$$

where we have for the time being neglected channel-length modulation. We know from our study in Section 5.1 that i_D will saturate at

$$v_{DS} = V_{OV} = V_{GS} - V_t \tag{13.96}$$

and the saturation current will be

$$i_D = \frac{1}{2}\mu_n C_{ox}\left(\frac{W}{L}\right)(V_{GS} - V_t)^2 \tag{13.97}$$

This will also be the case in a short-channel device as long as the value of v_{DS} in Eq. (13.96) is lower than V_{DSsat}. That is, as long as

$$V_{OV} < V_{DSsat}$$

the current i_D will be given by Eqs. (13.95) and (13.97). If, on the other hand,

$$V_{OV} > V_{DSsat}$$

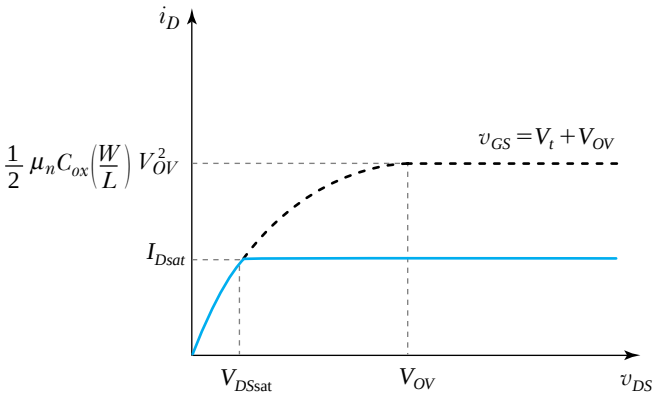


Figure 13.40 Velocity saturation causes the i_D–v_{DS} characteristic to saturate at V_{DSsat}. This early saturation results in a current I_{Dsat} that is lower than the value for a long-channel device.

then velocity saturation kicks in at $v_{DS} = V_{DSsat}$ and i_D saturates at a value I_{Dsat}, as shown in Fig. 13.40. The value of I_{Dsat} can be obtained by substituting $v_{DS} = V_{DSsat}$ in Eq. (13.95),

$$I_{Dsat} = \mu_n C_{ox}\left(\frac{W}{L}\right)V_{DSsat}\left(V_{GS} - V_t - \frac{1}{2}V_{DSsat}\right) \tag{13.98}$$

This expression can be simplified by utilizing Eq. (13.94) to obtain

$$I_{Dsat} = WC_{ox}v_{sat}\left(V_{GS} - V_t - \frac{1}{2}V_{DSsat}\right) \tag{13.99}$$

Replacing V_{GS} in Eq. (13.98) with v_{GS}, and incorporating the channel-length modulation factor $(1 + \lambda v_{DS})$, we obtain a general expression for the drain current of an NMOS transistor operating in velocity saturation,

$$i_D = \mu_n C_{ox}\left(\frac{W}{L}\right)V_{DSsat}\left(v_{GS} - V_t - \frac{1}{2}V_{DSsat}\right)(1 + \lambda v_{DS}) \tag{13.100}$$

which applies for

$$v_{GS} - V_t \geq V_{DSsat} \quad \text{and} \quad v_{DS} \geq V_{DSsat} \tag{13.101}$$

Figure 13.41 shows a set of i_D–v_{DS} characteristic curves and clearly delineates the three regions of operation: triode, saturation, and velocity saturation.

Equation (13.100) indicates that in the velocity-saturation region, i_D is linearly related to v_{GS}. This is a major change from the quadratic relationship that characterizes operation in the saturation region. Figure 13.42 makes this point clearer by presenting a graph for i_D versus v_{GS} of a short-channel device operating at $v_{DS} > V_{DSsat}$. Observe that for $0 < v_{GS} - V_t \leq V_{DSsat}$, the MOSFET operates in the saturation region and i_D is related to v_{GS} by the familiar quadratic equation (Eq. 13.97). For $v_{GS} - V_t \geq V_{DSsat}$, the transistor enters the velocity-saturation region and i_D varies linearly with v_{GS} (Eq. 13.100).

Short-channel PMOS transistors undergo velocity saturation at the same value of v_{sat} (approximately 10^7 cm/s), but the effects on the device characteristics are less pronounced than in the NMOS case. This is due to the lower values of μ_p and the correspondingly higher values of E_{cr} and V_{DSsat}.

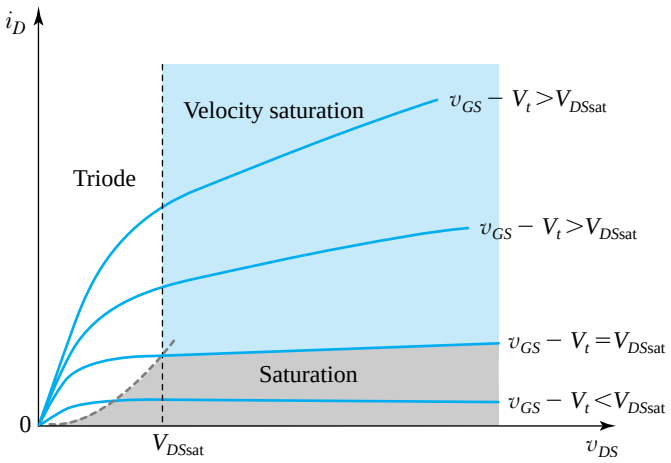


Figure 13.41 The i_D–v_{DS} characteristics of a short-channel MOSFET. Note the three different regions of operation: triode; saturation; and velocity saturation.

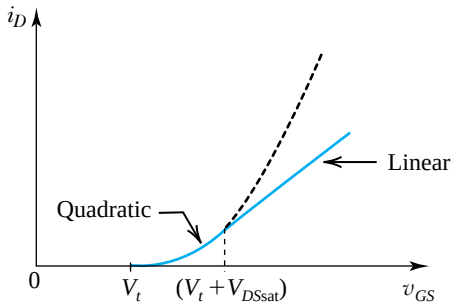


Figure 13.42 The i_D–v_{GS} characteristic of a short-channel NMOS transistor operating at $v_{DS} > V_{DSsat}$. Observe the quadratic and the linear portions of the characteristic. Also note that in the absense of velocity saturation, the quadratic curve would continue as shown with the broken line.

Example 13.8

Consider MOS transistors fabricated in a 0.25-μm CMOS process for which $V_{DD} = 2.5$ V, $V_{tn} = -V_{tp} = 0.5$ V, $\mu_n C_{ox} = 115\ \mu A/V^2$, $\mu_p C_{ox} = 30\ \mu A/V^2$, $\lambda_n = 0.6\ V^{-1}$, and $|\lambda_p| = 0.1\ V^{-1}$. Let $L = 0.25$ μm and $(W/L)_n = (W/L)_p = 1.5$. Measurements indicate that for the NMOS transistor, $V_{DSsat} = 0.63$ V, and for the PMOS device, $|V_{DSsat}| = 1$ V. Calculate the drain current obtained in each of the NMOS and PMOS transistors for $|V_{GS}| = |V_{DS}| = V_{DD}$. Compare with the values that would have been obtained in the absence of velocity saturation. Also give the range of v_{DS} for which i_D is saturated, with and without velocity saturation.

Solution

For the NMOS transistor, $V_{GS} = 2.5$ V results in $V_{GS} - V_{tn} = 2.5 - 0.5 = 2$ V, which is greater than V_{DSsat}. Also, $V_{DS} = 2.5$ V is greater than V_{DSsat}; thus both conditions in Eq. (13.101) are satisfied, and

Example 13.8 *continued*

the NMOS transistor will be operating in the velocity-saturation region, and thus i_D is given by Eq. (13.100):

$$i_D = 115 \times 10^{-6} \times 1.5 \times 0.63 \times \left(2.5 - 0.5 - \frac{1}{2} \times 0.63\right) \times (1 + 0.06 \times 2.5) = 210.6\ \mu\text{A}$$

If velocity saturation were absent, the current would be

$$\begin{aligned} i_D &= \frac{1}{2}(\mu_n C_{ox})\left(\frac{W}{L}\right)_n (v_{GS} - V_{tn})^2 (1 + \lambda v_{DS}) \\ &= \frac{1}{2} \times 115 \times 10^{-6} \times 1.5 \times (2.5 - 0.5)^2 \times (1 + 0.06 \times 2.5) \\ &= 396.8\ \mu\text{A} \end{aligned}$$

Thus, velocity saturation reduces the current level by nearly 50%! The saturation current, however, is obtained over a larger range of v_{DS}; specifically, for $v_{DS} = 0.63$ V to 2.5 V. (Of course, the current does not remain constant over this range because of channel-length modulation.) In the absence of velocity saturation, the current saturates at $V_{OV} = V_{GS} - V_t = 2$ V, and thus the saturation current is obtained over the range $v_{DS} = 2$ V to 2.5 V.

For the PMOS transistor, we see that since $|V_{GS}| - |V_t| = 2$ V and $|V_{DS}| = 2.5$ V are both larger that $|V_{DSsat}| = 1$ V the device will be operating in velocity saturation, and i_D can be obtained by adapting Eq. (13.100) as follows:

$$\begin{aligned} i_D &= (\mu_p C_{ox})\left(\frac{W}{L}\right)_p |V_{DSsat}|\left(|V_{GS}| - |V_{tp}| - \frac{1}{2}|V_{DSsat}|\right)(1 + |\lambda_p||V_{DS}|) \\ &= 30 \times 10^{-6} \times 1.5 \times 1 \times \left(2.5 - 0.5 - \frac{1}{2} \times 1\right)(1 + 0.1 \times 2.5) \\ &= 84.4\ \mu\text{A} \end{aligned}$$

Without velocity saturation, we have

$$\begin{aligned} i_D &= \frac{1}{2}(\mu_p C_{ox})\left(\frac{W}{L}\right)_p (|V_{GS}| - |V_{tp}|)^2 (1 + |\lambda_p||V_{DS}|) \\ &= \frac{1}{2} \times 30 \times 10^{-6} \times 1.5 \times (2.5 - 0.5)^2 (1 + 0.1 \times 2.5) \\ &= 112.5\ \mu\text{A} \end{aligned}$$

Thus velocity saturation reduces the current by 25% (which is less than in the case of the NMOS transistor), and the saturated current is obtained over the range $|V_{DS}| = 1$ V to 2.5 V. In the absence of velocity saturation, the saturated i_D would have been obtained for $|V_{DS}| = 2$ V to 2.5 V.

EXERCISE

13.26 Repeat the problem in Example 13.8 for transistors fabricated in a 0.13-μm CMOS process for which $V_{DD} = 1.2$ V, $V_{tn} = -V_{tp} = 0.4$ V, $\mu_n C_{ox} = 430\ \mu\text{A/V}^2$, $\mu_p C_{ox} = 110\ \mu\text{A/V}^2$, $\lambda_n = |\lambda_p| = 0.1\ \text{V}^{-1}$. Let $L = 0.13$ μm, $(W/L)_n = (W/L)_p = 1.5$, V_{DSsat} (NMOS) = 0.34 V, and V_{DSsat} (PMOS) = 0.6 V.

Ans. NMOS: $I_D = 154.4\ \mu A$, compared to $231.2\ \mu A$ without velocity saturation; saturation is obtained over the range $v_{DS} = 0.34$ V to 1.2 V, compared to $v_{DS} = 0.8$ V to 1.2 V in the absence of velocity saturation. PMOS: $I_D = 55.4\ \mu A$ compared to $59.9\ \mu A$, and $|v_{DS}| = 0.6$ V to 1.2 V compared to 0.8 V to 1.2 V.

Effect on the Inverter Characteristics The VTC of the CMOS inverter can be derived using the modified i_D–v_{DS} characteristics of the MOSFETs. The results, however, indicate relatively small changes from the VTC derived in Section 13.2 using the long-channel equations (see Rabaey et al., 2003 and Hodges et al., 2004), and we shall not pursue this subject here. The dynamic characteristics of the inverter, however, are significantly impacted by velocity saturation. This is because the current available to charge and discharge the equivalent load capacitance C is substantially reduced.

A Remark on the MOSFET Model The model derived above for short-channel MOSFETs is an approximate one, intended to enable the circuit designer to perform hand analysis to gain insight into circuit operation. Also, the model parameter values are usually obtained from measured data by means of a numerical curve-fitting process. As a result, the model applies only over a restricted range of terminal voltages.

Modeling short-channel MOSFETs is an advanced topic that is beyond the scope of this book. Suffice it to say that sophisticated models have been developed and are utilized by circuit simulation programs such as SPICE (see Appendix B). Circuit simulation is an essential step in the design of integrated circuits. However, it is not a substitute for initial hand analysis and design.

13.5.3 Subthreshold Conduction

In our study of the NMOS transistor in Section 5.1, we assumed that current conduction between drain and source occurs only when v_{GS} exceeds V_t. That is, we assumed that for $v_{GS} < V_t$ no current flows between drain and source. This, however, turns out not to be the case, especially for deep-submicron devices. Specifically, for $v_{GS} < V_t$ a small current i_D flows. To be able to see this **subthreshold conduction**, we have redrawn the i_D–v_{GS} graph of Fig. 13.42, utilizing a logarithmic scale for i_D, as shown in Fig. 13.43. Observe that at low values of v_{GS}, the relationship between $\log i_D$ and v_{GS} is linear, indicating that i_D varies exponentially with v_{GS},

$$i_D = I_S e^{v_{GS}/nV_T} \tag{13.102}$$

where I_S is a constant, $V_T = kT/q$ is the thermal voltage $\simeq 25$ mV at room temperature, and n is a constant whose value falls in the range 1 to 2, depending on the material and structure of the device.[7]

Subthreshold conduction has been put to good use in the design of very-low-power circuits such as those needed for electronic watches. Generally speaking, however, subthreshold conduction is a problem in digital IC design. This is so for two reasons.

[7]This relationship is reminiscent of the i_C–v_{BE} relationship of a BJT (Chapter 6). This is no coincidence, for the subthreshold conduction in a MOSFET is due to the lateral bipolar transistor formed by the source and drain diffusions with the substrate acting as the base region (see Fig. 5.1).

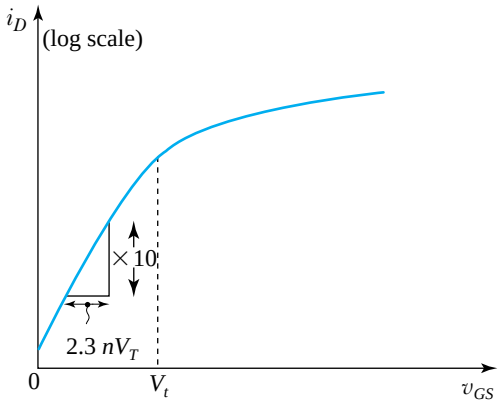


Figure 13.43 The i_D–v_{GS} characteristic of a short channel MOSFET. To show the details of subthreshold conduction a logarithmic scale is needed for i_D.

1. The nonzero current that flows for $v_{GS} = 0$ (see Fig. 13.43) causes the CMOS inverter to dissipate static power. To keep this **off current** as low as possible, V_t of the MOSFET is kept relatively high. This indeed is the reason why V_t has not been scaled by the same factor as that used for the channel length. Although the off current is low (10 pA to 100 pA) and the power dissipation per inverter is small, the problem becomes serious in chips with a billion transistors!
2. The nonzero current of a normally off transistor can cause the discharge of capacitors in dynamic MOS circuits. As we shall see in the next two chapters, dynamic logic and memory circuits rely on charge storage on capacitors for their proper operation. Thus, subthreshold conduction can disrupt the operation of such circuits.

EXERCISE

13.27 (a) Refer to Fig. 13.43 and to Eq. (13.102). Show that the inverse of the slope of the straight line representing subthreshold conduction is given by $2.3nV_T$ V per decade of current change.
(b) If measurements indicate $n = 1.22$ and $i_D = 100$ nA at $v_{GS} = 0.21$ V, find i_D at $v_{GS} = 0$.
(c) For a chip having 500 million transistors, find the current drawn from the 1.2-V supply V_{DD} as a result of subthreshold conduction. Hence estimate the resulting power dissipation.
Ans. (b) 0.1 nA; (c) 50 mA, 60 mW

13.5.4 Wiring—The Interconnect

The logic gates on a digital IC chip are connected together by metal wires[8] (see Appendix A). As well, the power supply V_{DD} and ground are distributed throughout the chip by metal wires. Technology scaling into the deep-submicron range have caused these wires to behave

[8]These are strips of metal deposited on an insulating surface on top of the chip. In modern digital ICs, as many as eight layers of such wiring are utilized.

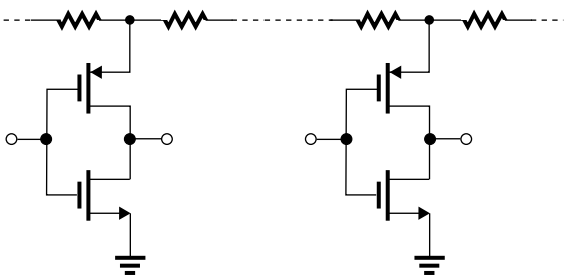


Figure 13.44 The power-supply line in a deep submicron IC has non-zero resistance. The *IR* drops along the V_{DD} line cause the voltages delivered to various circuits to differ.

not simply as wires! Specifically, the narrow wires typical of deep-submicron technologies exhibit nonzero resistance. The result is an *IR* drop on the V_{DD} line resulting in somewhat different voltages being delivered to different parts of the chip, as shown in Fig. 13.44. This can have deleterious effects on the operation of the overall circuit.

Since chips fabricated in deep-submicron technologies can have hundreds of millions of gates, the wire connection between gates can be long. The resulting narrow and long **interconnect** lines have not only nonzero resistance but also capacitance to ground, as shown in Fig. 13.45. The resistance and capacitance of an interconnect line can cause a propagation delay approaching that of the logic gate itself. As well, the capacitance between adjacent wires can cause the signals on one wire to be coupled to the other, which can cause erroneous operation of logic circuits.

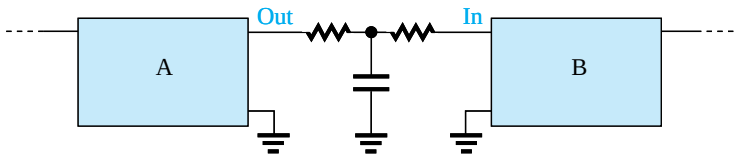


Figure 13.45 The interconnect (wire) between two circuit blocks, A and B, on an IC chip has finite resistance and a capacitance to ground.

In short, the circuit designer of modern deep-submicron digital ICs has to concern herself not only with the logic-circuit design but also with the wiring or interconnect issues. Indeed, advanced textbooks on digital IC design devote entire chapters to this topic (see Rabaey et al., 2003, and Hodges et al., 2004). Our intent here is simply to point out that interconnect has become an important issue in digital IC design.

Summary

- The digital logic inverter is the basic building block of digital circuits, just as the amplifier is the basic building block of analog circuits.
- The static operation of the inverter is described by its voltage-transfer characteristic (VTC). The VTC determines the inverter noise margins; refer to Fig. 13.5 and to Table 13.1 for the definitions of important VTC points and the noise margins. In particular, note that $NM_H = V_{OH} - V_{IH}$ and $NM_L = V_{IL} - V_{OL}$, and refer to the ideal VTC in Fig. 13.6.
- The inverter is implemented using transistors operating as voltage-controlled switches. There are three possible arrangements, shown in Figs. 13.7, 13.8, and 13.9. The arrangement in Fig. 13.8 results in a high-performance inverter and is the basis for the CMOS inverter studied in Section 13.2.
- An important performance parameter of the inverter is the amount of power it dissipates. There are two components of power dissipation: static and dynamic. The first is the result of current flow in either the 0 or 1 state or both. The second occurs when the inverter is switched and has a capacitor load C. Dynamic power dissipation $P_{\text{dyn}} = fCV_{DD}^2$.
- The speed of operation of the inverter is characterized by its propagation delay, t_P. Refer to Fig. 13.15 for the definitions of t_{PLH} and t_{PHL}, and note that $t_P = \frac{1}{2}(t_{PLH} + t_{PHL})$. The maximum frequency at which an inverter can be switched $f_{\max} = 1/2t_P$.
- A metric that combines speed of operation and power dissipation is the power–delay product, $PDP = P_D t_P$. The lower the *PDP*, the more effective the logic-circuit family is. If dynamic power is dominant, such as in CMOS, $PDP = CV_{DD}^2$, which is the energy drawn from the supply for a 0-to-1 and a 1-to-0 transition. (i.e., one switching cycle).
- Besides speed of operation and power dissipation, the silicon area required for an inverter is the third significant metric in digital IC design.
- Predominantly because of its low power dissipation and because of its scalability, CMOS is by far the most dominant technology for digital IC design. This situation is expected to continue for many years to come.
- Table 13.3 provides a summary of the important characteristics of the CMOS inverter.
- Digital ICs usually utilize the minimum channel length of the technology available. Thus for the CMOS inverter, Q_N and Q_P have $L = L_{\min}$. If matching is desired, W_p/W_n is selected equal to μ_n/μ_p. at the expense of increased area and capacitance. For minimum area, $W_p = W_n$. Also, a frequently used compromise is $W_p = 2W_n$.
- For minimum area, $(W/L)_n$ is selected equal to 1. However, to reduce t_P especially when a major part of C is extrinsic to the inverter, $(W/L)_n$ and correspondingly $(W/L)_p$ can be increased.
- A CMOS logic gate consists of an NMOS pull-down network (PDN) and a PMOS pull-up network (PUN). The PDN conducts for every input combination that requires a low output. Since an NMOS transistor conducts when its input is high, the PDN is most directly synthesized from the expression for the low output $(\overline{Y})$ as a function of the uncomplemented inputs. In a complementary fashion, the PUN conducts for every input combination that corresponds to a high output. Since a PMOS conducts when its input is low, the PUN is most directly synthesized from the expression for a high output (Y) as a function of the complemented inputs.
- CMOS logic circuits are usually designed to provide equal current-driving capability in both directions. Furthermore, the worst-case values of the pull-up and pull-down currents are made equal to those of the basic inverter. Transistor sizing is based on this principle and makes use of the equivalent W/L ratios of series and parallel devices (Eqs. 13.88 and 13.89).
- Refer to Table 13.2 for the implications of scaling the dimension of the MOSFET and V_{DD} and V_t by a factor $1/S$.
- In devices with short channels $(L < 0.25\ \mu\text{m})$ velocity saturation occurs. Its effect is that i_D saturates early, and its value is lower than would be the case in long-channel devices (see Figs. 13.40, 13.41 and 13.42, and Eq. 13.100).

Table 13.3 Summary of Important Characteristics of the CMOS Logic Inverter

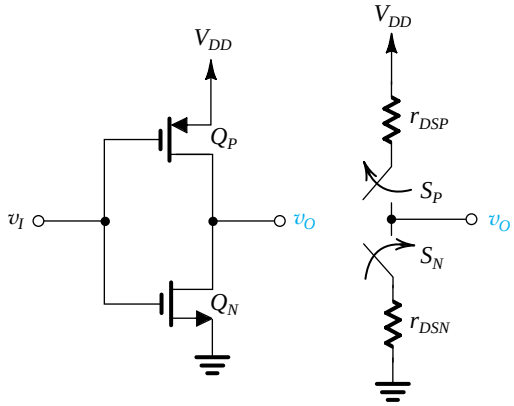
Inverter Output Resistance

- When v_O is low (current sinking):

$$r_{DSN} = 1\Big/\left[k_n'\left(\frac{W}{L}\right)_n (V_{DD} - V_{tn})\right]$$

- When v_O is high (current sourcing):

$$r_{DSP} = 1\Big/\left[k_p'\left(\frac{W}{L}\right)_p (V_{DD} - |V_{tp}|)\right]$$



Inverter VTC and Noise Margins

$$V_M = \frac{r(V_{DD} - |V_{tp}|) + V_{tn}}{1 + r} \quad \text{where} \quad r = \sqrt{\frac{k_p'(W/L)_p}{k_n'(W/L)_n}}$$

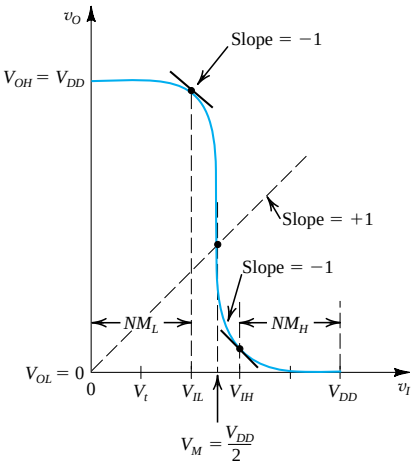
For matched devices, that is, $\mu_n\left(\frac{W}{L}\right)_n = \mu_p\left(\frac{W}{L}\right)_p$:

$$V_M = V_{DD}/2$$

$$V_{IL} = \tfrac{1}{8}(3V_{DD} + 2V_t)$$

$$V_{IH} = \tfrac{1}{8}(5V_{DD} - 2V_t)$$

$$NM_H = NM_L = \tfrac{1}{8}(3V_{DD} + 2V_t)$$



Propagation Delay (Fig. 13.22)

Using average currents:

$$t_{PHL} \simeq \frac{\alpha_n C}{k_n'(W/L)_n V_{DD}} \quad \text{where} \quad \alpha_n = 2\Big/\left[\frac{7}{4} - \frac{3V_{tn}}{V_{DD}} + \left(\frac{V_{tn}}{V_{DD}}\right)^2\right]$$

$$t_{PLH} \simeq \frac{\alpha_p C}{k_p'(W/L)_p V_{DD}} \quad \text{where} \quad \alpha_p = 2\Big/\left[\frac{7}{4} - \frac{3|V_{tp}|}{V_{DD}} + \left(\frac{|V_{tp}|}{V_{DD}}\right)^2\right]$$

Using equivalent resistances (Fig. 13.23):

$$t_{PHL} = 0.69 R_N C \quad \text{where} \quad R_N = \frac{12.5}{(W/L)_n}\ \text{k}\Omega$$

$$t_{PLH} = 0.69 R_P C \quad \text{where} \quad R_P = \frac{30}{(W/L)_p}\ \text{k}\Omega$$

SIM Problems identified by this icon are intended to demonstrate the value of using SPICE simulation to verify hand analysis and design, and to investigate important issues such as gate noise margins and propagation delays. Instructions to assist in setting up PSpice and Multisim simulations for all the indicated problems can be found in the corresponding files on the CD. Note that if a particular parameter value is not specified in the problem statement, you are to make a reasonable assumption. * difficult problem; ** more difficult; *** very challenging and/or time-consuming; D: design problem.

Section 13.1: Digital Logic Inverters

13.1 A particular logic inverter is specified to have $V_{IL} = 1.2$ V, $V_{IH} = 1.5$ V, $V_{OL} = 0.2$ V, and $V_{OH} = 2.5$ V. Find the high and low noise margins, NM_H and NM_L.

13.2 The voltage-transfer characteristic of a particular logic inverter is modeled by three straight-line segments in the manner shown in Fig. 13.3. If $V_{IL} = 2.0$ V, $V_{IH} = 2.5$ V, $V_{OL} = 0.5$ V, and $V_{OH} = 5$ V, find:

(a) The noise margins
(b) The value of V_M
(c) The voltage gain in the transition region

13.3 For a particular inverter design using a power supply V_{DD}, $V_{OL} = 0.1V_{DD}$, $V_{OH} = 0.8V_{DD}$, $V_{IL} = 0.4V_{DD}$, and $V_{IH} = 0.6V_{DD}$. What are the noise margins? What is the width of the transition region? For a minimum noise margin of 1 V, what value of V_{DD} is required?

13.4 A logic circuit family that used to be very popular is transistor-transistor logic (TTL). The TTL logic gates and other building blocks are available commercially in small-scale integrated (SSI) and medium-scale-integrated (MSI) packages. Such packages can be assembled on printed-circuit boards to implement a digital system. The device data sheets provide the following specifications of the basic TTL inverter (of the SN7400 type):

Logic-1 input level required to ensure a logic-0 level at the output: MIN (minimum) 2 V
Logic-0 input level required to ensure a logic-1 level at the output: MAX (maximum) 0.8 V
Logic-1 output voltage: MIN 2.4 V, TYP (typical) 3.3 V
Logic-0 output voltage: TYP 0.22 V, MAX 0.4 V
Logic-0-level supply current: TYP 3 mA, MAX 5 mA
Logic-1-level supply current: TYP 1 mA, MAX 2 mA
Propagation delay time to logic-0 level (t_{PHL}): TYP 7 ns, MAX 15 ns
Propagation delay time to logic-1 level (t_{PLH}): TYP 11 ns, MAX 22 ns

(a) Find the worst-case values of the noise margins.
(b) Assuming that the inverter is in the 1-state 50% of the time and in the 0-state 50% of the time, find the average static power dissipation in a typical circuit. The power supply is 5 V.
(c) Assuming that the inverter drives a capacitance $C_L = 45$ pF and is switched at a 1-MHz rate, use the formula in Eq. (13.35) to estimate the dynamic power dissipation.
(d) Find the propagation delay t_P.

13.5 Consider an inverter implemented as in Fig. 13.7(a). Let $V_{DD} = 5$ V, $R = 1.8$ kΩ, $R_{on} = 200$ Ω, $V_{IL} = 1$ V, and $V_{IH} = 2$ V.

(a) Find V_{OL}, V_{OH}, NM_H, and NM_L.
(b) The inverter is driving N identical inverters. Each of these load inverters, or **fan-out** inverters as they are usually called, is specified to require an input current of 0.2 mA when the input voltage (of the fan-out inverter) is high and zero current when the input voltage is low. Noting that the input currents of the fan-out inverters will have to be supplied through R of the driving inverter, find the resulting value of V_{OH} and of NM_H as a function of the number of fan-out inverters N. Hence find the maximum value N can have while the inverter is still providing an NM_H value approximately equal to its NM_L.
(c) Find the static power dissipation in the inverter in the two cases: (i) the output is low, and (ii) the output is high and driving the maximum fan-out found in (b).

13.6 For a logic-circuit family employing a 3-V supply, suggest an ideal set of values for V_M, V_{IL}, V_{IH}, V_{OL}, V_{OH}, NM_L, NM_H. Also, sketch the VTC. What value of voltage gain in the transition region does your ideal specification imply?

13.7 For a particular logic-circuit family, the basic technology used provides an inherent limit to the small-signal low-frequency voltage gain of 50 V/V. If, with a 3.3-V supply, the values of V_{OL} and V_{OH} are ideal, but $V_M = 0.4V_{DD}$, what are the best possible values of V_{IL} and V_{IH} that can be expected? What are the best possible noise margins you could expect? If the actual noise margins are only 7/10 of these values, what V_{IL} and V_{IH} result? What is the large-signal voltage gain [defined as $(V_{OH} - V_{OL})/(V_{IL} - V_{IH})$]. (*Hint:* Use straight-line approximations for the VTC.)

***13.8** A logic-circuit family intended for use in a digital-signal-processing application in a newly developed hearing aid can operate down to single-cell supply voltages of 1.2 V. If for its inverter, the output signals swing between 0 and V_{DD}, the "gain-of-one" points are separated by less than $\frac{1}{3}V_{DD}$, and the noise margins are within 30% of one another, what ranges of values of V_{IL}, V_{IH}, V_{OL}, V_{OH}, NM_L, and NM_H can you expect for the lowest possible battery supply?

D 13.9 Design the inverter circuit in Fig. 13.2(a) to provide $V_{OH} = 2$ V, $V_{OL} = 0.1$ V, and so that the current

drawn from the supply in the low-output state is 20 μA. The transistor has $V_t = 0.5$ V, $\mu_n C_{ox} = 100\ \mu A/V^2$, and $\lambda = 0$. Specify the required values of V_{DD}, R_D, and W/L. How much power is drawn from the supply when the output is high? When the output is low?

13.10 For the current-steering circuit in Fig. 13.9, $V_{CC} = 3$ V, $I_{EE} = 1$ mA, find the values of R_{C1} and R_{C2} to obtain a voltage swing of 1.5 V at each output. What are the values realized for V_{OH} and V_{OL}?

D 13.11 Refer to the analysis of the resistive-load MOS inverter in Example 13.1 and utilize the expressions derived there for the various inverter parameters. Design the circuit to satisfy the following requirements: $V_{OH} = 2.5$ V; $V_{OL} = 0.1$ V, and the power dissipation in the low-output state = 125 μW. The transistor available has $V_t = 0.5$ V, $\mu_n C_{ox} = 100\ \mu A/V^2$, and $\lambda = 0$. Specify the required values of V_{DD}, R_D, and W/L. What are the values obtained for V_{IL}, V_M, V_{IH}, NM_L, and NM_H?

D 13.12 Refer to the analysis of the resistive-load MOS inverter in Example 13.1 and utilize the expressions derived there for the various inverter parameters. For a technology for which $V_t = 0.2V_{DD}$, it is required to design the inverter to obtain $V_M = V_{DD}/2$. In terms of V_{DD}, what is the required value of the design parameter V_x? What values are obtained for V_{OH}, V_{OL}, V_{IL}, V_{IH}, NM_H, and NM_L, in terms of V_{DD}? Give numerical values for the case $V_{DD} = 2.5$ V. Now, express the power dissipated in the inverter in its low-output state in terms of the transistor's W/L ratio. Let $k_n' = 100\ \mu A/V^2$. If the power dissipation is to be limited to approximately 100 μW, what W/L ratio is needed and what value of R_D corresponds?

13.13 Consider the saturated-load inverter of Fig. 13.11(a), analyzed in Example 13.2. From Eq. (13.20),

$$V_{OH} = V_{DD} - V_{t2}$$

where V_{t2} is given by

$$V_{t2} = V_{t0} + \gamma[\sqrt{V_{OH} + 2\phi_f} - \sqrt{2\phi_f}]$$

For $V_{t0} = 0.5$ V, $V_{DD} = 1.8$ V, $\gamma = 0.3\ V^{1/2}$, $2\phi_f = 0.8$ V, use an iterative process to determine V_{t2} and V_{OH}. By how much is V_{OH} reduced as a result of the body effect on Q_2?

13.14 Determining V_{IH} of the saturated-load inverter of Fig. 13.11(a) requires a rather tedious process (see Example 13.2). An approximate estimate of V_{IH} can be obtained by reference to the VTC shown in Fig. 13.11(d). Specifically, when the straight-line segment BC is extrapolated, it meets the horizontal axis at $(V_M + V_M/k_r)$, which is usually close to the value of V_{IH}. What is the approximate value obtained this way for the case analyzed in Example 13.2? How much does it differ from the value calculated the long way in Example 13.2?

D 13.15 It is required to design the saturated-load inverter in Fig. 13.11(a) for the case $V_{DD} = 2.5$ V, $V_t = 0.5$ V, $k_n' = 100\ \mu A/V^2$, and $\lambda = 0$. Design for $V_{OL} \simeq 0.05$ V. Utilize the expressions derived in Example 13.2, except for V_{IH} use the following approximate expression (see Problem 13.11):

$$V_{IH} \simeq V_M + \frac{V_M}{k_r}$$

Neglect the body effect in Q_2. Determine V_M, NM_L, and NM_H for your design. Also determine $(W/L)_1$ and $(W/L)_2$ assuming that $(W/L)_2 = 1/(W/L)_1$. What is the power dissipated in the inverter during its low-output state?

13.16 An IC inverter fabricated in a 0.25-μm CMOS process is found to have a load capacitance of 10 fF. If the inverter is operated from a 2.5-V power supply, find the energy needed to charge and discharge the load capacitance. If the IC chip has 1 million of these inverters operating at an average switching frequency of 1 GHz, what is the power dissipated in the chip? What is the average current drawn from the power supply?

13.17 Consider a logic inverter of the type shown in Fig. 13.8. Let $V_{DD} = 5$ V, and let a 1-pF capacitance be connected between the output node and ground. If the inverter is switched at the rate of 100 MHz, determine the dynamic power dissipation. What is the average current drawn from the dc power supply?

13.18 In a particular logic family, operating with a 3.3-V supply, the basic inverter draws (from the supply) a current of 40 μA in one state and 0 μA in the other. When the inverter is switched at the rate of 100 MHz, the average supply current becomes 150 μA. Estimate the equivalent capacitance at the output node of the inverter.

13.19 A collection of logic gates for which the static-power dissipation is zero, and the dynamic-power dissipation is 10 mW is operating at 50 MHz with a 5-V supply. By what fraction could the power dissipation be reduced if operation at 3.3 V were possible? If the frequency of operation is reduced by the same factor as the supply voltage (i.e., 3.3/5), what *additional* power can be saved?

13.20 A logic inverter is implemented using the arrangement of Fig. 13.8 with switches having $R_{on} = 1$ kΩ, $V_{DD} = 5$ V, and $V_{IL} = V_{IH} = V_{DD}/2$.

(a) Find V_{OL}, V_{OH}, NM_L, and NM_H.
(b) If v_I rises instantaneously from 0 V to +5 V and assuming the switches operate instantaneously—that is, at $t = 0$,

PU opens and PD closes—find an expression for $v_O(t)$, assuming that a capacitance C is connected between the output node and ground. Hence find the high-to-low propagation delay (t_{PHL}) for $C = 1$ pF. Also find t_{THL} (see Fig. 13.15).
(c) Repeat (b) for v_I falling instantaneously from +5 V to 0 V. Again assume that PD opens and PU closes instantaneously. Find an expression for $v_O(t)$, and hence find t_{PLH} and t_{TLH}.

13.21 In a particular logic family, the standard inverter, when loaded by a similar circuit, has a propagation delay specified to be 1.2 ns:

(a) If the current available to charge a load capacitance is half as large as that available to discharge the capacitance, what do you expect t_{PLH} and t_{PHL} to be?
(b) If when an external capacitive load of 1 pF is added at the inverter output, its propagation delays increase by 70%, what do you estimate the normal combined capacitance of inverter output and input to be?
(c) If without the additional 1-pF load connected, the load inverter is removed and the propagation delays were observed to decrease by 40%, estimate the two components of the capacitance found in (b) that is, the component due to the inverter output and other associated parasitics, and the component due to the input of the load inverter.

13.22 Consider an inverter for which t_{PLH}, t_{PHL}, t_{TLH}, and t_{THL} are 20 ns, 10 ns, 30 ns, and 15 ns, respectively. The rising and falling edges of the inverter output can be approximated by linear ramps. Also, for simplicity, we define t_{TLH} to be 0% to 100% (rather than 10% to 90%) rise time, and similarly for t_{THL}. Two such inverters are connected in tandem and driven by an ideal input having zero rise and fall times. Calculate the time taken for the output voltage to complete its excursion for (a) a rising input and (b) a falling input. What is the propagation delay for the inverter?

13.23 A particular logic gate has t_{PLH} and t_{PHL} of 50 ns and 70 ns, respectively, and dissipates 1 mW with output low and 0.5 mW with output high. Calculate the corresponding delay–power product (under the assumption of a 50% duty-cycle signal and neglecting dynamic power dissipation).

D ****13.24** We wish to investigate the design of the inverter shown in Fig. 13.7(a). In particular, we wish to determine the value for R. Selection of a suitable value for R is determined by two considerations: propagation delay and power dissipation.

(a) Show that if v_I changes instantaneously from high to low and assuming that the switch opens instantaneously, the output voltage obtained across a load capacitance C will be

$$v_O(t) = V_{OH} - (V_{OH} - V_{OL})e^{-t/\tau_1}$$

where $\tau_1 = CR$. Hence show that the time required for $v_O(t)$ to reach the 50% point, $\frac{1}{2}(V_{OH} + V_{OL})$, is

$$t_{PLH} = 0.69CR$$

(b) Following a steady state, if v_I goes high and assuming that the switch closes immediately and has the equivalent circuit in Fig. 13.7(c), show that the output falls exponentially according to

$$v_O(t) = V_{OL} + (V_{OH} - V_{OL})e^{-t/\tau_2}$$

where $\tau_2 = C(R \parallel R_{on}) \simeq CR_{on}$ for $(R_{on} \ll R)$. Hence show that the time for $v_O(t)$ to reach the 50% point is

$$t_{PHL} = 0.69CR_{on}$$

(c) Use the results of (a) and (b) to obtain the inverter propagation delay, defined as the average of t_{PLH} and t_{PHL} as

$$t_P \simeq 0.35CR \quad \text{for } R_{on} \ll R$$

(d) Show that for an inverter that spends half the time in the 0-state and half the time in the 1-state, the average static power dissipation is

$$P = \frac{1}{2}\frac{V_{DD}^2}{R}$$

(e) Now that the trade-offs in selecting R should be clear, show that, for $V_{DD} = 5$ V and $C = 10$ pF, to obtain a propagation delay no greater than 10 ns and a power dissipation no greater than 10 mW, R should be in a specific range. Find that range and select an appropriate value for R. Then determine the resulting values of t_p and P.

D 13.25 A logic-circuit family with zero static-power dissipation, normally operates at $V_{DD} = 5$ V. To reduce its dynamic-power dissipation operation at 3.3 V is considered. It is found, however, that the currents available to charge and discharge load capacitances also decrease. If current is (a) proportional to V_{DD} or (b) proportional to V_{DD}^2, what reductions in maximum operating frequency do you expect in each case? What fractional change in delay–power product do you expect in each case?

Section 13.2: The CMOS Inverter

13.26 Consider a CMOS inverter fabricated in a 0.25-μm CMOS process for which $V_{DD} = 2.5$ V, $V_{tn} = -V_{tp} = 0.5$ V, and $\mu_n C_{ox} = 3.5\mu_p C_{ox} = 115\ \mu A/V^2$. In addition, Q_N and Q_P have $L = 0.25$ μm and $(W/L)_n = 1.5$.

(a) Find W_p that results in $V_M = V_{DD}/2$. What is the silicon area utilized by the inverter in this case?

(b) For the matched case in (a), find the values of V_{OH}, V_{OL}, V_{IH}, V_{IL}, NM_L, and NM_H.
(c) For the matched case in (a), find the output resistance of the inverter in each of its two states.

SIM 13.27 For the technology specified in Problem 13.26, investigate the variation of V_M with the ratio W_p/W_n. Specifically, calculate V_M for (a) $W_p = 3.5W_n$ (the matched case); (b) $W_p = W_n$ (the minimum-size case); and (c) $W_p = 2W_n$ (a compromise case). For cases (b) and (c), estimate the approximate reduction in NM_L and silicon area relative to the matched case (a).

13.28 For a technology in which $V_{tn} = 0.2V_{DD}$, show that the maximum current that the inverter can sink while its low-output level does not exceed $0.1\ V_{DD}$ is $0.075\ k_n'(W/L)_n V_{DD}^2$. For $V_{DD} = 2.5$ V, $k_n' = 115\ \mu A/V^2$, find $(W/L)_n$ that permits this maximum current to be 0.5 mA.

13.29 A CMOS inverter for which $k_n = 10k_p = 100\ \mu A/V^2$ and $V_t = 0.5$ V is connected as shown in Fig. P13.29 to a sinusoidal signal source having a Thévenin equivalent voltage of 0.1-V peak amplitude and resistance of 100 kΩ. What signal voltage appears at node A with $v_I = +1.5$ V? With $v_I = -1.5$ V?

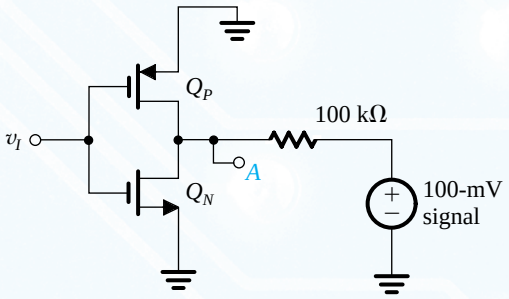


Figure P13.29

D 13.30 There are situations in which Q_N and Q_P of the CMOS inverter are deliberately mismatched to realize a certain desired value for V_M. Show that the value required of the parameter r of Eq. (13.59) is given by

$$r = \frac{V_M - V_{tn}}{V_{DD} - |V_{tp}| - V_M}$$

For a 0.18-μm process characterized by $V_{tn} = -V_{tp} = 0.5$ V, $V_{DD} = 1.8$ V, and $\mu_n = 4\mu_p$, find the ratio W_p/W_n required to obtain $V_M = 0.6V_{DD}$.

13.31 Consider the CMOS inverter of Fig. 13.17 with Q_N and Q_P matched and with the input v_I rising slowly from 0 to V_{DD}. At what value of v_I does the current flowing through Q_N and Q_P reach its peak? Give an expression for the peak current, neglecting λ_n and λ_p. For $k_n' = 300\ \mu A/V^2$, $(W/L)_n = 1.5$, $V_{DD} = 1.8$ V, and $V_{tn} = 0.5$ V, find the value of the peak current.

SIM 13.32 For a CMOS inverter fabricated in a 0.13-μm process with $V_{DD} = 1.2$ V, $V_{tn} = -V_{tp} = 0.4$ V, $k_n' = 4k_p' = 430\ \mu A/V^2$, and having $(W/L)_n = 1.5$ and $(W/L)_p = 3$, find t_{PHL}, t_{PLH}, and t_P when the equivalent load capacitance $C = 10$ fF. Use the method of average currents.

D 13.33 Consider a matched CMOS inverter fabricated in the 0.13-μm process specified in Problem 13.32. If $C = 20$ fF, use the method of average currents to determine the required (W/L) ratios so that $t_P \le 20$ ps.

13.34 For the CMOS inverter in Exercise 13.14 use the method of equivalent resistance to determine t_{PHL}, t_{PLH}, and t_P.

13.35 Use the method of equivalent resistance to determine the propagation delay of a minimum-size inverter, that is, one for which $(W/L)_n = (W/L)_p = 1$, designed in a 0.18-μm technology. The equivalent load capacitance $C = 10$ fF.

D 13.36 Use the method of equivalent resistance to design an inverter to be fabricated in a 0.18-μm technology. It is required that for $C = 10$ fF, $t_{PLH} = t_{PHL}$, and $t_P \le 40$ ps.

13.37 The method of average currents yields smaller values for t_{PHL} and t_{PLH} than those obtained by the method of equivalent resistances. Most of this discrepancy is due to the fact that the formula we derived for I_{av} does not take into account velocity saturation. As will be seen in Section 13.5.2, velocity saturation reduces the current significantly. Using the results in Example 13.5, by what factor do you estimate the current reduction to be in the NMOS transistor? Since t_{PLH} does not change, what do you conclude about the effect of velocity saturation on the PMOS transistor in this technology?

13.38 Find the propagation delay for a minimum-size inverter for which $k_n' = 3k_p' = 180\ \mu A/V^2$ and $(W/L)_n = (W/L)_p = 0.75\ \mu m/0.5\ \mu m$, $V_{DD} = 3.3$ V, $V_{tn} = -V_{tp} = 0.7$ V, and the capacitance is roughly 2 fF/μm of device width plus 1 fF/device. What does t_p become if the design is changed to a matched one? Use the method of average current.

13.39 A matched CMOS inverter fabricated in a process for which $C_{ox} = 3.7$ fF/μm^2, $\mu_n C_{ox} = 180\ \mu A/V^2$, $\mu_p C_{ox} = 45\ \mu A/V^2$, $V_{tn} = -V_{tp} = 0.7$ V, and $V_{DD} = 3.3$ V, uses $W_n = 0.75$ μm and $L_n = L_p = 0.5$ μm. The overlap capacitance and the effective drain–body capacitance per micrometer of gate width are 0.4 fF and 1.0 fF, respectively. The wiring capacitance is

$C_w = 2$ fF. If the inverter is driving another identical inverter, find t_{PLH}, t_{PHL}, and t_P. For how much additional capacitance load does the propagation delay increase by 50%?

D ***13.40** In this problem we investigate the effect of the selection of the ratio W_p/W_n on the propagation delay of an inverter driving an identical inverter, as in Fig. 13.24.

(a) Noting that except for C_w each of the capacitances in Eqs. (13.72) and (13.73) is proportional to the width of the relevant transistor, show that C can be expressed as

$$C = C_n\left(1 + \frac{W_p}{W_n}\right) + C_w$$

where C_n is determined by the NMOS transistors.

(b) Using the equivalent resistances R_N and R_P, show that for $(W/L)_n = 1$,

$$t_{PHL} = 8.625 \times 10^3 C$$

$$t_{PLH} = \frac{20.7 \times 10^3}{W_p/W_n} C$$

(c) Use the results of (a) and (b) to determine t_P in the case $W_p = W_n$, in terms of C_n and C_w.

(d) Use the results of (a) and (b) to determine t_P in the matched case: that is, when W_p/W_n is selected to yield $t_{PHL} = t_{PLH}$.

(e) Compare the t_P values in (c) and (d) for the two extreme cases:

(i) $C_w = 0$

(ii) $C_w \gg C_n$

What do you conclude about the selection of W_p/W_n?

13.41 An inverter whose equivalent load capacitance C is composed of 10 fF contributed by the inverter transistors, and 20 fF contributed by the wiring and other external circuitry, has been found to have a propagation delay of 60 ps. By what factor must $(W/L)_n$ and $(W/L)_p$ be increased so as to reduce t_P to 30 ps?

13.42 A CMOS microprocessor chip containing the equivalent of 1 million gates operates from a 5-V supply. The power dissipation is found to be 9 W when the chip is operating at 120 MHz, and 4.7 W when operating at 50 MHz. What is the power lost in the chip by some clock-independent mechanism, such as leakage and other static currents? If 70% of the gates are assumed to be active at any time, what is the average gate capacitance in such a design?

13.43 Repeat Problem 13.39 for an inverter for which $(W/L)_n = (W/L)_p = 0.75\ \mu m/0.5\ \mu m$. Find t_p and the dynamic power dissipation when the circuit is operated at a 250-MHz rate.

13.44 In this problem we estimate the inverter power dissipation resulting from the current pulse that flows in Q_N and Q_P when the input pulse has finite rise and fall times. Refer to Fig. 13.26 and let $V_{tn} = -V_{tp} = 0.5$ V, $V_{DD} = 1.8$ V, and $k_n = k_p = 450\ \mu A/V^2$. Let the input rising and falling edges be linear ramps with the 0-to-V_{DD} and V_{DD}-to-0 transitions taking 1 ns each. Find I_{peak}. To determine the energy drawn from the supply per transition, assume that the current pulse can be approximated by a triangle with a base corresponding to the time for the rising or falling edge to go from V_t to $V_{DD} - V_t$, and the height equal to I_{peak}. Also, determine the power dissipation that results when the inverter is switched at 100 MHz.

Section 13.4: CMOS Logic-Gate Circuits

D **13.45** Sketch a CMOS realization for the function $Y = \overline{A + B(C + D)}$.

D **13.46** A CMOS logic gate is required to provide an output $Y = \bar{A}BC + A\bar{B}C + AB\bar{C}$. How many transistors does it need? Sketch a suitable PUN and PDN, obtaining each first independently, then one from the other using the dual-networks idea.

D **13.47** Give two different realizations of the exclusive OR function $Y = A\bar{B} + \bar{A}B$ in which the PDN and the PUN are dual networks.

D **13.48** Sketch a CMOS logic circuit that realizes the function $Y = AB + \bar{A}\bar{B}$. This is called the **equivalence** or **coincidence function**.

D **13.49** Sketch a CMOS logic circuit that realizes the function $Y = ABC + \bar{A}\bar{B}\bar{C}$.

D **13.50** It is required to design a CMOS logic circuit that realizes a three-input, even-parity checker. Specifically, the output Y is to be low when an even number (0 or 2) of the inputs A, B, and C are high.

(a) Give the Boolean function $\bar{Y}$.
(b) Sketch a PDN directly from the expression for $\bar{Y}$. Note that it requires 12 transistors in addition to those in the inverters.
(c) From inspection of the PDN circuit, reduce the number of transistors to 10.
(d) Find the PUN as a dual of the PDN in (c), and hence the complete realization.

D **13.51** Give a CMOS logic circuit that realizes the function of three-input, odd-parity checker. Specifically,

the output is to be high when an odd number (1 or 3) of the inputs are high. Attempt a design with 10 transistors (not counting those in the inverters) in each of the PUN and the PDN.

D 13.52 Design a CMOS full-adder circuit with inputs A, B, and C, and two outputs S and C_0 such that S is 1 if one or three inputs are 1, and C_0 is 1 if two or more inputs are 1.

D 13.53 Consider the CMOS gate shown in Fig. 13.33. Specify W/L ratios for all transistors in terms of the ratios n and p of the basic inverter, such that the worst-case t_{PHL} and t_{PLH} of the gate are equal to those of the basic inverter.

D 13.54 Find appropriate sizes for the transistors used in the exclusive-OR circuit of Fig. 13.34(b). Assume that the basic inverter has $(W/L)_n = 0.27\ \mu m/0.18\ \mu m$ and $(W/L)_p = 0.54\ \mu m/0.18\ \mu m$. What is the total area, including that of the required inverters?

13.55 Consider a four-input CMOS NAND gate for which the transient response is dominated by a fixed-size capacitance between the output node and ground. Compare the values of t_{PLH} and t_{PHL}, obtained when the devices are sized as in Fig. 13.36, to the values obtained when all n-channel devices have $W/L = n$ and all p-channel devices have $W/L = p$.

13.56 Figure P13.56 shows two approaches to realizing the OR function of six input variables. The circuit in Fig. P13.56(b), though it uses additional transistors, has in fact less total area and lower propagation delay because it uses NOR gates with lower fan-in. Assuming that the transistors in both circuits are properly sized to provide each gate with a current-driving capability equal to that of the basic matched inverter, find the number of transistors and the total area of each circuit. Assume the basic inverter to have a $(W/L)_n$ ratio of 0.27 µm/0.18 µm and a $(W/L)_p$ ratio of 0.54 µm/0.18 µm.

***13.57** Consider the two-input CMOS NOR gate of Fig. 13.31 whose transistors are properly sized so that the current-driving capability in each direction is equal to that of a matched inverter. For $|V_t| = 1$ V and $V_{DD} = 5$ V, find the gate threshold in the cases for which (a) input terminal A is connected to ground and (b) the two input terminals are tied together. Neglect the body effect in Q_{PB}.

Section 13.5: Implications of Technology Scaling: Issues in Deep-Submicron Design

13.58 A chip with a certain area designed using the 10-µm process of the early 1970s contains 10,000 transistors. What does Moore's law predict the number of transistors to be on a chip of equal area fabricated using the 45-nm process of 2009?

13.59 Consider the scaling from a 0.18-µm process to a 45-nm process.

(a) Assuming V_{DD} and V_t are scaled by the same factor as the device dimensions $(S = 4)$, find the factor by which t_P, the maximum operating speed, P_{dyn}, power density, and PDP decrease (or increase)?
(b) Repeat (a) for the situation in which V_{DD} and V_t are scaled by a factor of only 2.

13.60 For a 0.18-µm technology, V_{DSsat} for minimum-length NMOS devices is measured to be 0.6 V and that for minimum-length PMOS devices 1.0 V. What do you estimate the effective values of μ_n and μ_p to be? Also find the values of E_{cr} for both device polarities.

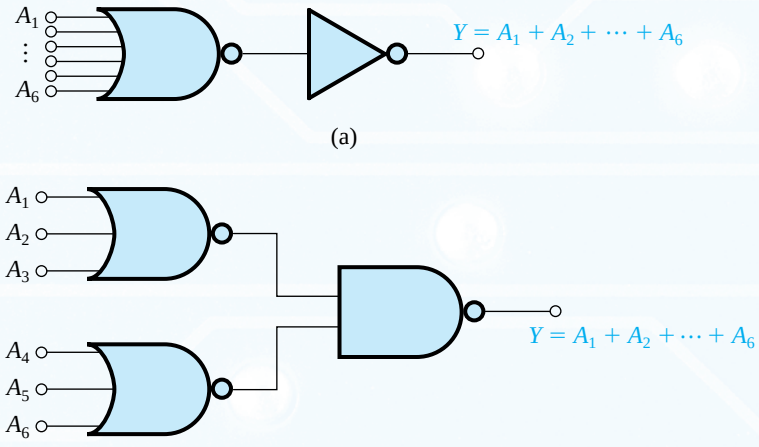


Figure P13.56

13.61 Consider NMOS and PMOS transistors with minimum channel length fabricated in a 0.13-μm CMOS process. If the effective values of μ_n and μ_p are 325 $cm^2/V \cdot s$ and 200 $cm^2/V \cdot s$, respectively, find the expected values of V_{DSsat} for both device polarities.

13.62 (a) Show that for short-channel NMOS transistor, the ratio of the current I_{Dsat} obtained at $v_{GS} = V_{DD}$ to the current obtained if velocity saturation were absent is given by

$$\frac{I_{Dsat}}{I_D} = \frac{2V_{DSsat}\left(V_{DD} - V_t - \frac{1}{2}V_{DSsat}\right)}{(V_{DD} - V_t)^2}$$

(b) Find the ratio in (a) for a transistor fabricated in a 0.13-μm process with $L = 0.13$ μm, $V_t = 0.4$ V, $V_{DSsat} = 0.34$ V, and $V_{DD} = 1.2$ V.

13.63 (a) Consider a CMOS inverter fabricated in a deep-submicron technology utilizing transistors with the minimum allowed channel length and having an equivalent load capacitance C. Let v_I rise instantaneously to V_{DD} and assume that Q_P turns off and Q_N turns on immediately. Ignoring channel-length modulation, that is, $\lambda = 0$, and assuming Q_N operates in the velocity-saturation region, show that

$$t_{PHL} = \frac{CV_{DD}}{2I_{Dsat}}$$

(b) Using the equivalent resistance of Q_N show that

$$t_{PHL} = 0.69C\ \frac{12.5 \times 10^3}{(W/L)_n}$$

(c) If the formulas in (a) and (b) are to yield the same result, find V_{DSsat} for the NMOS transistor for a 0.13-μm technology characterized by $V_{DD} = 1.2$ V, $V_t = 0.4$ V, and $\mu_n C_{ox} = 325\ \mu A/V^2$.

D 13.64 (a) For a CMOS inverter fabricated in a deep-submicron technology with $L_n = L_p$ = the minimum allowed channel length, it is required to select W_p/W_n so that $t_{PHL} = t_{PLH}$. This can be achieved by making I_{Dsat} of Q_N equal to I_{Dsat} of Q_P at $|v_{GS}| = V_{DD}$. Show that W_p/W_n is given by

$$\frac{W_p}{W_n} = \frac{\mu_n}{\mu_p}\frac{V_{DSsatn}}{|V_{DSsatp}|}\frac{V_{DD} - V_{tn} - \frac{1}{2}V_{DSsatn}}{V_{DD} - |V_{tp}| - \frac{1}{2}|V_{DSsatp}|}$$

(b) Find the required W_p/W_n for a 0.13-μm technology for which $\mu_n/\mu_p = 4$, $V_{DD} = 1.2$ V, $V_{tn} = -V_{tp} = 0.4$ V, $V_{DSsatn} = 0.34$ V, and $|V_{DSsatp}| = 0.6$ V.

D 13.65 The current I_S in the subthreshold conduction Eq. (13.102) is proportional to e^{-V_t/nV_T}. If the threshold voltage of an NMOS transistor is reduced by 0.1 V, by what factor will the static power dissipation increase? Repeat for a reduction in V_t by 0.2 V. What do you conclude about the selection of a value of V_t in process design?

13.66 An interconnect wire with a length L, a width W, and a thickness T has a resistance R given by

$$R = \rho\frac{L}{A} = \frac{\rho L}{TW}$$

where ρ is the resistivity of the material of which the wire is made. The quantity ρ/T is called the **sheet resistance** and has the dimension of ohms, although it is usually expressed as ohms/square or $\Omega/\square$ (refer to Fig. P13.66a).

(a) Find the resistance of an aluminum wire that is 10 mm long and 0.5 μm wide, if the sheet resistance is specified to be 27 $m\Omega/\square$.
(b) If the wire capacitance to ground is 0.1 fF/μm length, what is the total wire capacitance?
(c) If we can model the wire very approximately as an RC circuit as shown in Fig. P13.66(b), find the delay time introduced by the wire. (*Hint*: $t_{delay} = 0.69RC$.) (P.S. Only a small fraction of the interconnect on an IC would be this long!)

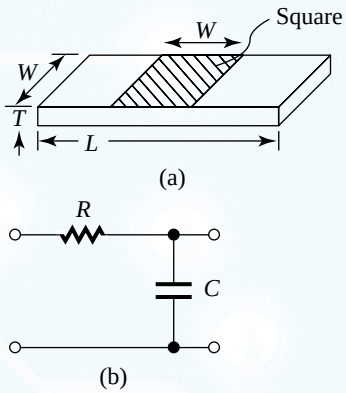


Figure P13.66