CHAPTER 7

BASIC DIGITAL BUILDING BLOCKS

7.0 INTRODUCTION

Digital microelectronic circuits dominate the applications of integrated circuits in terms of both sales volume and number of device types. Applications include subcomponents for computer systems and accessories, electronic games, hand-held calculators, digital instrumentation, and the ubiquitous microprocessor. Each of these applications uses circuits that are similar to those previously studied except that they are designed to operate with two-valued, or binary, inputs and outputs. Because these circuits require only two disjoint output voltage ranges to function correctly, practical circuits exist that can function reliably over wide ranges of voltages, currents, temperatures, and process characteristics. A relatively small number of basic digital circuits are used as building blocks. These building blocks are often interconnected to form complex circuits that perform tasks such as those just listed. Following an overview of the digital design process, these basic digital building blocks will be analyzed in this chapter.

MOS digital logic rather than bipolar digital logic is considered in the analysis presented here. The subject of small-scale circuits built from bipolar technology is adequately covered in many references. Most VLSI digital circuits are constructed using MOS technology. Exceptions are I^2L, GaAs, and bipolar gate arrays. These are limited applications compared to MOS VLSI circuits.

In digital circuits it is customary to drop the notations for a bulk connection on MOS transistors and to ignore the bulk effect on threshold voltages for simple logic analysis. When more accurate analysis is required, circuit simulators that include bulk effects are used. As a transition from the notation of previous

chapters (see Fig. 2.2-4), the arrows on the source terminals that indicate n- or p-channel polarity will be dropped for digital circuits. Instead, a bubble on the gate of a transistor indicates a p-channel transistor. The plain enhancement transistor symbol will be used for n-channel transistors, and the filled drain-to-source bar will continue to indicate depletion transistors.

7.1 DESIGN ABSTRACTION

One powerful resource for digital integrated circuit design is the capability to use several levels of abstraction for a complete design. These design levels, as shown in Fig. 7.1-1, can include functional blocks, register transfer descriptions, logic diagrams, circuit schematics, and geometrical circuit layout.

The functional block level is usually described by a block diagram showing the major subcomponents of a design. For a computer, these blocks might include processor, memory, input/output, and special functions such as memory cache or A/D interfaces. These blocks are usually described textually rather than with a formal language, although this is changing.[1] The functional block level is the highest description level from which a designer works.

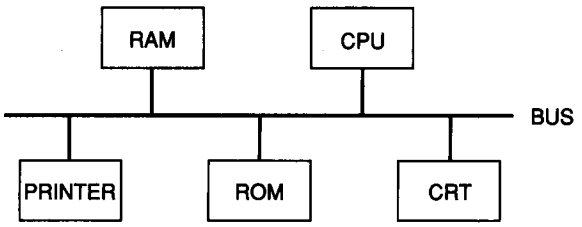
The next highest level, the register transfer level (RTL), consists of programlike statements describing the movement or processing of data between storage elements. One such language, the Instruction Set Processor Specification language,[2] has been used to describe many recent computers and other digital systems. The RTL level allows a formal definition of the operation of a digital processor system. Figure 7.1-1*b* shows a sequence of register transfer operations for a processor that adds the contents of two registers and stores the result in a third register.

The logic diagram is widely used to define the internal operation of blocks described at higher levels. A logic diagram contains well-defined logic functions such as NAND gates and latches, whose digital operation can be accurately described via Boolean logic.[3] For some styles of design, for example, gate array circuits, this is the lowest level of description required from a designer.

The circuit schematic is used to describe integrated circuits at the electrical level as in Fig. 7.1-1*d*. Circuit schematics may be provided for each logic block or may be used to describe portions of the circuit not directly describable by logic blocks—for example, MOS selector circuits or charge storage elements. The circuit level is characterized by discrete electronic components and their corresponding equations of operation. Within classical digital design, this level of detail is used infrequently. Because of the nature of MOS circuits, mixed descriptions are common, with classical logic functions described via logic symbols, and other functions—such as selectors—described by circuit schematics.

The geometrical layout level, depicted in Fig. 7.1-1*e*, most closely describes the physical realization of circuits and gates symbolized at the higher levels. Designer knowledge of this level is desirable because, to a large extent, the geometrical characteristics of circuit layout determine the performance and size of digital integrated circuits.

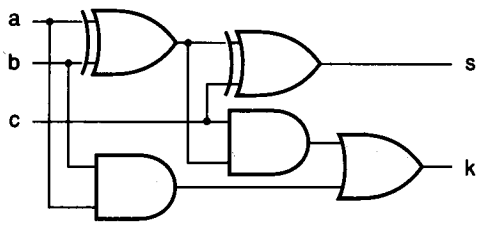
This chapter provides detailed information about several different types of logic circuits and their characteristics. In a later chapter we show that these simple



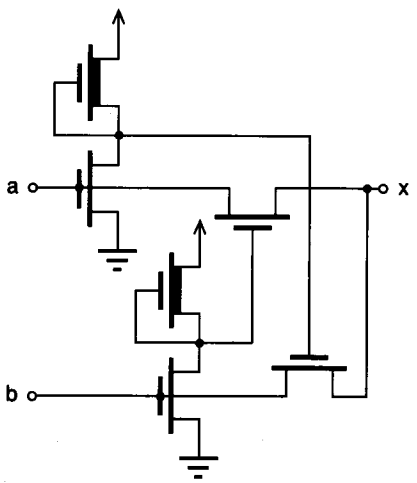
(a)

ADRBUS ← PC
PC ← PC + 1
IR ← DATABUS
ALUINA ← Reg (5)
ALUINB ← Reg (6)
REG (2) ← ALUOUT

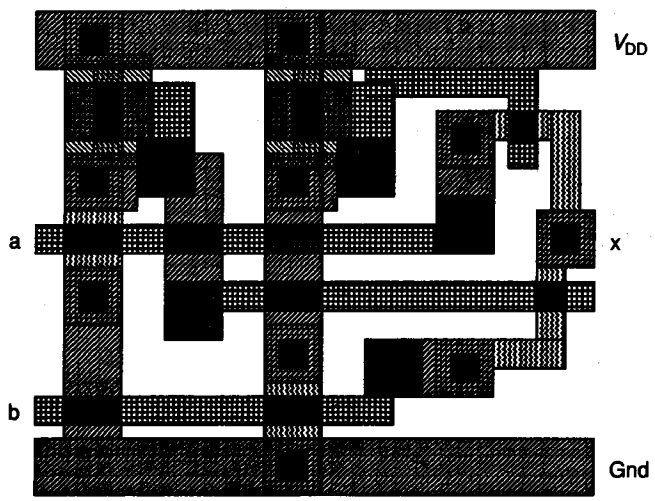
(b)



(c)



(d)



(e)

FIGURE 7.1-1
Levels of abstraction: (*a*) Functional blocks, (*b*) Register transfer level, (*c*) Logic diagram, (*d*) Circuit schematic, (*e*) Geometrical layout.

circuits can be composed into larger subsystems and that these subsystems can be used without specific attention to minute details of each transistor within a design. The capability to provide a hierarchy of increasing functionality with decreasing detail allows a digital circuit designer to create complex systems in a relatively short period of time.

For good digital system design, a designer must be aware of the characteristics and constraints of several levels of abstraction. The purpose of this chapter is to introduce concepts that are important for designing individual logic functions at the logic, circuit, and layout levels. A familiarity with the definitions and interaction of classical logic gates is assumed.

7.2 CHARACTERISTICS OF DIGITAL CIRCUITS

Digital circuits are almost always operated in a two-level, or binary, mode. This means that at steady state, each input and output is in one of two conditions. These conditions (or *states*) are often referred to as the true and false states, the high and low levels, or the 1 and 0 states, respectively. Because circuit outputs are generally voltages, these two states are characterized by two voltage ranges based on V_{IH} and V_{IL}, where $V_{IH} > V_{IL}$. V_{IH} and V_{IL} are called the high and low *logic thresholds*, respectively. It is agreed that if a nodal voltage V satisfies the inequality $V > V_{IH}$, then this node is in the high state; if $V < V_{IL}$, the node is in the low state. Nominally, the extreme limits of the allowable voltage ranges are ground and the supply voltage. If a nodal voltage V satisfies the inequality $V_{IL} < V < V_{IH}$, the state is indeterminate. Indeterminate states must be avoided in digital systems that have reached steady state. The designation of a state thus refers to a voltage range for a nodal voltage in a circuit. Because we are only interested in states, the exact value of the voltage within a range should not be of major concern. A procedure to determine values for V_{IL} and V_{IH} based on the dc voltage transfer characteristic will be given in Sec. 7.11.

7.2.1 Logic Level Standards

Although it might appear that the designer has total control in selecting V_{IH} and V_{IL}, this is not the case, for several practical reasons. First, each integrated circuit technology can generate certain high and low voltage levels more naturally than others. Second, the ability to interconnect distinct integrated circuits requires compatibility of electrical characteristics. Thus, standard voltage ranges and current capabilities for logic states have been defined for different technologies, and integrated circuit vendors from around the world attempt to adhere to these standards for their products. This standardization makes possible the design of complex digital systems that include ICs from different manufacturers. Moreover, it is common for systems manufacturers to insist that two or more sources (vendors) exist for all ICs used in a design. This eliminates sole dependence on the product of a single manufacturer. This *second-source* requirement provides additional incentive for integrated circuit manufacturers to adhere to standards.

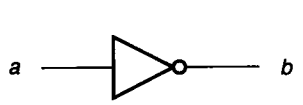
TABLE 7.2-1
Logic voltage specifications

Family	V_{CC}	V_{IL}	V_{IH}
TTL	5.0	0.8	2.0
ALSTTL[1]	5.0	0.8	2.0
ECL	−5.2	−1.5	−1.1
HCMOS[2]	4.5	0.9	3.2
MOS		No standard	
I^2L		No standard	

[1] Advanced Low-Power Schottky TTL
[2] High-speed CMOS

The recognized interface voltage levels for several different logic families are shown in Table 7.2-1.[4,5] Note that standard logic specifications are not provided for MOS and I^2L. Some technologies do not have standard logic voltage specifications because they are used only within large-scale integrated circuits. These large-scale circuits contain input and output interface circuitry to adhere to some widely accepted standard for small-scale logic, usually the TTL logic standard. Although adherence to standards is essential at the input and output terminals of most integrated circuits, such standards are not usually applied internally. Thus, interface circuitry is normally required at all inputs and outputs of an integrated circuit to maintain compatibility with other circuits.

The standard electrical interface characteristics of a logic family are determined essentially by the characteristics of the basic inverter. The symbol for the inverter along with its truth table are shown in Fig. 7.2-1. To investigate the characteristics of a logic family, consider the cascade of inverters shown in Fig. 7.2-2. Assume that all inverters are identical and that the voltage transfer characteristic of an inverter, loaded by a following stage, is as shown in Fig. 7.2-3. Assume the input of the first inverter of Fig. 7.2-2 is at a voltage level V_{i1}, where V_{i1} is in the range that characterizes a high level. The output of the first inverter, V_{o1}, serves as the input to the second stage and is designated as V_{i2} in Fig. 7.2-2. Because of the transfer characteristic of Fig. 7.2-3, $V_{o1} = V_{i2}$ will be in the range that characterizes a low level. Continuing in this manner, one obtains the sequence of input voltages, V_{ik}, where V_{ik} is the input voltage to the kth inverter stage. By repeated application of the inversions of Fig. 7.2-3, it is found that the sequence of high-level voltages, $V_{i1}, V_{i3}, \ldots$, is decreasing and converging to a limit, whereas the sequence of low-level voltages $V_{i2}, V_{i4}, \ldots$, is also converging, typically to a different limit, but in an increasing manner. The limits of these sequences, specifically, V_L and V_H, are termed *equilibrium values*.



(a)

a	b
0	1
1	0

(b)

FIGURE 7.2-1
Basic inverter: (*a*) Symbol, (*b*) Truth table.

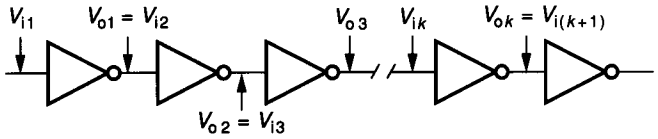


FIGURE 7.2-2
Inverter cascade.

If an inverter with the transfer characteristic of Fig. 7.2-3 is to be useful in a logic family, the low-level sequence must remain in the voltage range that characterizes one of the logic states, and the high-level sequence must remain in the voltage range that characterizes the second logic state. It should be apparent from this example that the shape of the inverter transfer characteristic completely determines the typical logic levels, V_L and V_H, of a logic family. Instead of specifying high and low voltage levels a priori and checking to see if these levels are satisfied for a given inverter, we can ask the related questions: If an inverter has a given transfer characteristic, can this inverter be used as the basis for a viable family of logic? If so, what are the high and low logic voltage levels? These questions are easily answered. We look for the answers by analyzing a pair of cascaded inverters in the next section.

7.2.2 Inverter-Pair Characteristics

First, we examine a pair of inverters in the inverter cascade of Fig. 7.2-2 in greater detail. Assume that the cascade is long enough so that at stage k, the circuit is at equilibrium; that is, $V_{ik} = V_{i(k+2)}$. An *inverter pair*, shown in Fig. 7.2-4*a*, can be thought of as the kth and $(k + 1)$th inverters in the cascade. The transfer characteristic for the inverter pair is shown in Fig. 7.2-4*b* where V_i and V_o are the input and output voltages, respectively. The slope of the transfer characteristic

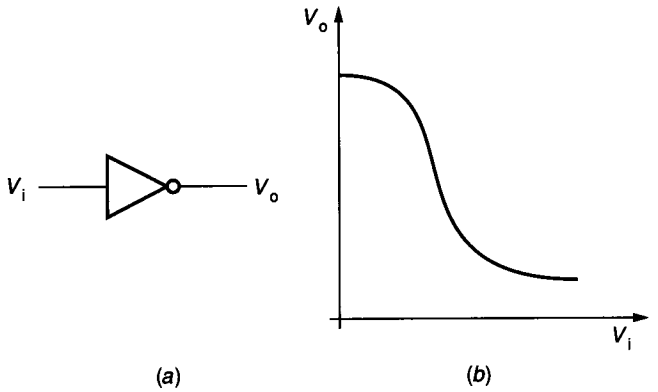


FIGURE 7.2-3
(a) Inverter, (b) Voltage transfer characteristic.

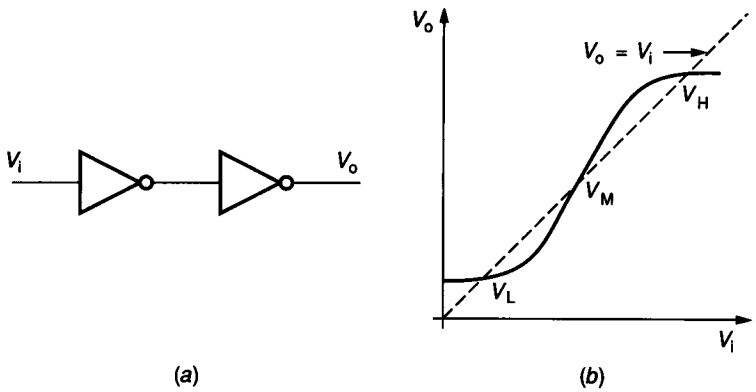


FIGURE 7.2-4
(*a*) Inverter pair, (*b*) Voltage transfer characteristic.

represents the voltage gain of the inverter pair. Note that for an input voltage of zero, the gain is less than one. As the input voltage increases, the gain increases to a value greater than one and then decreases to a value less than one as the input voltage reaches its maximum value. A logic family with an inverter-pair voltage transfer curve that exhibits a gain greater than unity for some region is called a *restoring* logic family. As a counter example, simple diode logic is a *nonrestoring* logic family. Restoring logic is so named because logic signals that are degraded in some way can be restored by the gain of subsequent logic. All widely used electronic logic circuits are of the restoring type.

The inverter-pair characteristic of Fig. 7.2-4*b* can be used to better understand the voltages representing the high and low logic states for a logic family. Note that the input and output voltages of an inverter pair must be equal for equilibrium voltages to exist in the inverter cascade of Fig. 7.2-2. Thus, the equilibrium voltages can be obtained by intersecting the straight line $V_o = V_i$ with the inverter-pair transfer characteristic of Fig. 7.2-4. Note that there are three points of intersection. The extreme intersection points are the high and low equilibrium values, designated by V_H and V_L, respectively. (V_H and V_L should not be confused with the logic threshold voltages V_{IH} and V_{IL} introduced at the beginning of this section.) The equilibrium voltages V_H and V_L represent stable values, as the voltage gain about these points is less than unity. Thus, any small input voltage perturbation about these equilibrium points will be diminished by the inverter pair. The middle intersection point in Fig. 7.2-4*b*, called the *switching threshold voltage* and denoted by V_M, represents an unstable value. Any small input voltage perturbation about V_M will be amplified by the (usually) large voltage gain about this point.

From the preceding analysis, it is apparent that two stable voltage ranges exist for an inverter pair when its characteristic exhibits less than unity gain in two regions. Yet the voltage transfer characteristic should exhibit greater than unity gain to restore degraded signals. Thus, an inverter forms the basis of a viable two-level logic family provided that its inverter-pair transfer characteristic crosses the

unity gain line at exactly three points: V_L, V_M, and V_H. The high and low logic thresholds, V_{IH} and V_{IL}, will be chosen so that $V_H > V_{IH} > V_M > V_{IL} > V_L$.

Additional useful information can be derived from the inverter-pair transfer characteristic. If this characteristic is monotonically increasing and $V_L < V_i < V_H$, then the sequence of high-level output voltages of an inverter cascade with V_i as input will monotonically increase toward V_H and the sequence of low-level output voltages will monotonically decrease toward V_L. For $V_i < V_L$ or $V_i > V_H$, the high-level output voltage sequence will monotonically decrease toward V_H and the low-level output voltage sequence will monotonically increase toward V_L. Thus, the high- and low-level sequences converge to V_H and V_L, respectively.

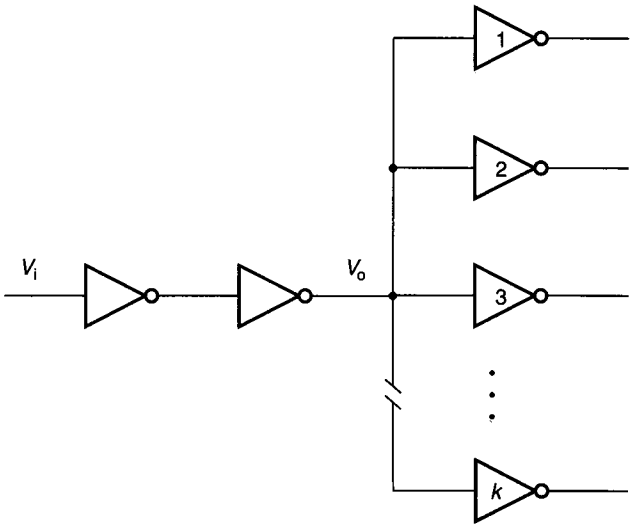
7.2.3 Logic Fan-out Characteristics

The preceding analysis for a cascade of inverters was performed with the assumption that each inverter was loaded by a single, identical inverter to obtain the inverter-pair transfer characteristic. If an inverter output must be capable of driving more than a single load, as shown in Fig. 7.2-5*a*, then the analysis must be performed again to establish high and low logic levels. Assume each inverter output drives k identical inverter inputs. A new inverter-pair transfer characteristic curve can be obtained as in Fig. 7.2-5*b* to show the effects of this loading. Note that increasing the loading on each stage decreases the high-level equilibrium point and increases the low-level equilibrium point, thereby changing the desired voltages for logic levels. However, for MOS logic, the static loading caused by driving multiple loads is so small as to be imperceptible in a transfer characteristic curve such as that of Fig. 7.2-5*b*. Whether a given inverter structure with a specified loading is satisfactory for a logic family depends on the inverter-pair transfer characteristic with loading. The number of identical inverter loads to which a circuit is connected is defined as the *fan-out* of that circuit. The second inverter of Fig. 7.2-5*a* has a fan-out of k. Usually logic level voltage specifications for a logic family are defined for a specific maximum fan-out.

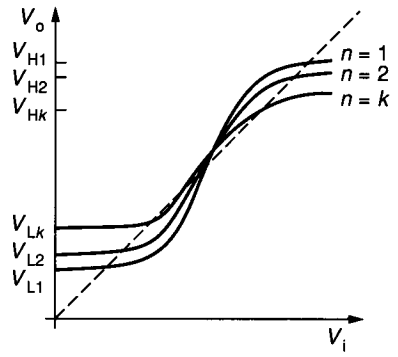
Inverter-pair dc transfer characteristics of some practical logic families are essentially unaffected by fan-out. Specifically, the inputs to logic functions fabricated with MOS technology are oxide-insulated gates of transistors that draw negligible quiescent input currents. Hence, dc characteristics of MOS logic functions are unaffected by fan-out to other MOS logic inputs. However, it will be shown later that the load capacitance attributed to a large fan-out seriously degrades the switching speed of MOS logic.

7.2.4 Digital Logic Analysis

Because the nodal voltages within a digital circuit can assume one of two distinct ranges, such circuits can be analyzed mathematically with Boolean algebra. This systematic way of characterizing digital systems is widely used by digital designers. It is assumed that readers of this book are familiar with the use of Boolean algebra to characterize digital systems. Relevant background information can be found in many references.[3,6]



(a)



(b)

FIGURE 7.2-5
Logic fan-out characteristics (*a*) Inverter with fan-out of k, (*b*) Inverter pair transfer characteristics with fan-out n.

Basic logic building blocks are often combined to realize complex logic functions. The most primitive set of logic functions consists of the AND, OR, and INVERT functions. This set is complete in the sense that any combinational logic function can be realized using only these three functions. However, the simplest electronic circuits used to realize logic functions always provide signal inversion (even diode logic circuits require signal amplification, and thus inversion, to restore voltage levels). Thus, basic electronic circuits provide the NOR rather than the OR and the NAND rather than the AND function. The NAND and the NOR functions are each individually complete in the sense defined previously. Either the NAND or the NOR logic circuit is the preferred structure in a given

technology, causing one of these logically complete functions to be the basic building block for a logic family. This will be described more fully in Sec. 7.4 and Sec. 7.6.

The basic logic circuits are typically termed *logic gates* or simply *gates* because of their ability to control or *gate* logic signal flow. (This term should not be confused with the gate terminal of the MOS transistor.) The characteristics and structure of integrated circuit implementations of these basic logic gates will be studied in the remainder of this chapter.

7.3 SINGLE-CHANNEL MOS INVERTERS

Single-channel MOS technology has been widely used to create large-scale integrated circuits. Early MOS digital circuits used PMOS technology because the natural gate threshold voltage of a p-channel transistor was more suitable for logic applications than was the gate threshold voltage of the n-channel transistor. With the advent of processing techniques such as ion implantation to control the threshold voltage, NMOS became the dominant single-channel technology. Most early microprocessors, including the Motorola 6800 and 6809, the Intel 8080 and 8086, and the Zilog Z80, were manufactured with NMOS processes.

NMOS technology is preferred over PMOS technology primarily because of the greater switching speed achievable with n-channel transistors. MOS transistors are majority carrier devices; thus, the current for an n-channel transistor is carried by electrons whereas that for a p-channel transistor is carried by holes. A silicon semiconductor has an electron mobility that is two to three times its hole mobility. This advantage in charge mobility allows NMOS circuits to operate faster than comparable PMOS circuits. Because the analysis of PMOS and NMOS logic is identical except for a reversal of voltage signs, the predominant device type, NMOS, is used for explanation of single-channel logic circuits in this text.

7.3.1 Basic Inverter

Two basic n-channel inverters are shown in Fig. 7.3-1. Each consists of a pull-down transistor, denoted by M1, and a pullup transistor or load device, denoted by M2. A device connected so as to pull the output voltage to the lower supply voltage—usually 0 V—is called a *pulldown* device; a device connected so as to pull the output voltage to the upper supply voltage—usually V_{DD}—is called a *pullup* device. The circuit of Fig. 7.3-1*a* is termed a *depletion-load* inverter because the pullup device, M2, is a depletion transistor. The circuit of Fig. 7.3-1*b* with an enhancement transistor M2 is termed an *enhancement-load* inverter. Note that these inverters are topologically identical to the analog inverters of Fig. 6.1-11*c* and *d*.

A typical plot of the transfer characteristic for each of these inverters is shown in Fig. 7.3-2, along with the inverter-pair transfer characteristics for devices sized according to the relationship $(W_1L_2)/(L_1W_2) = 4$, where the subscript 1 refers to M1 and the subscript 2 refers to M2. Figure 7.3-2*b* shows that the enhancement-load inverter is incapable of pulling the output any higher

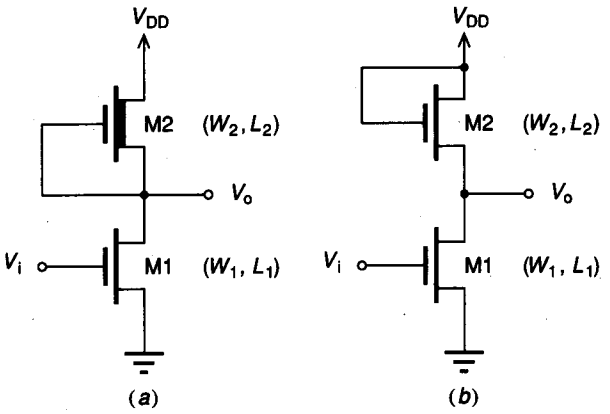


FIGURE 7.3-1
Basic NMOS inverters: (*a*) Depletion-load, (*b*) Enhancement-load.

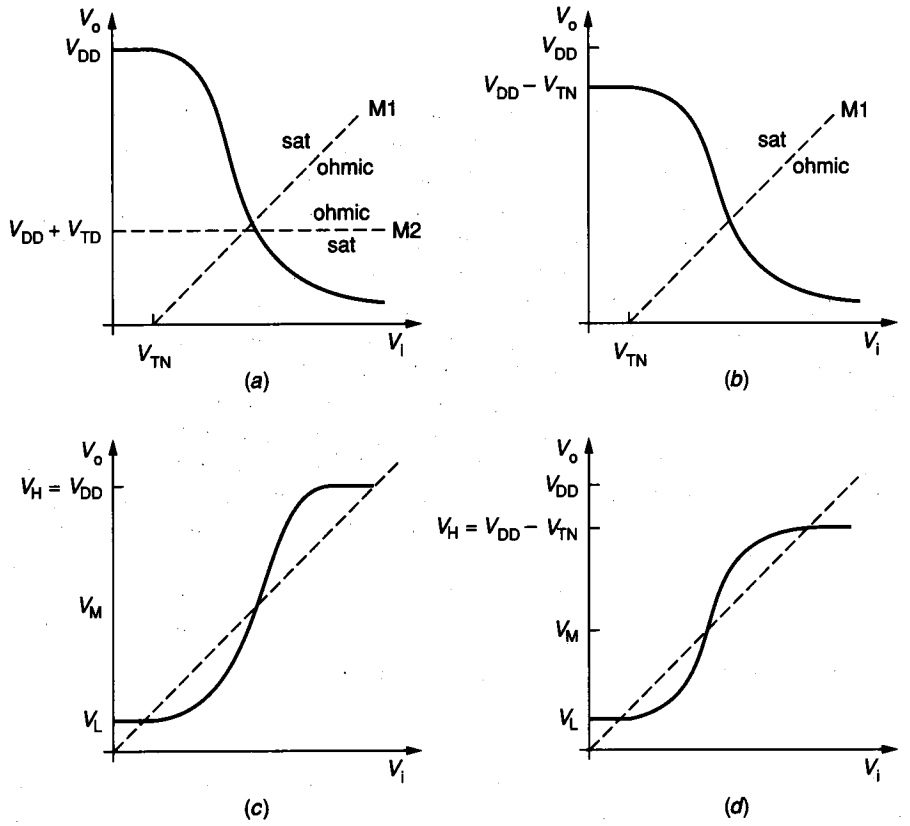


FIGURE 7.3-2
Transfer characteristics of NMOS inverters: (*a*) Depletion-load inverter, (*b*) Enhancement-load inverter, (*c*) Depletion-load inverter pair, (*d*) Enhancement-load inverter pair.

than one enhancement threshold voltage V_{TN} below V_{DD} when the input is at a low level, whereas the output voltage of the depletion-load inverter of Fig. 7.3-2*a* can be pulled to V_{DD}. These high-level output voltages are determined by the pullup transistor types and are nearly independent of device sizing. From the inverter-pair transfer characteristics, it can be seen that a better separation between high and low logic levels is possible with the depletion-load inverter. Good separation between logic voltage levels is important for proper operation of an inverter in the presence of noise voltages.

A layout of a depletion-load inverter with $(W_1L_2)/(L_1W_2) = 4$ is shown in Fig. 7.3-3. From this figure, it should be apparent that the area for the depletion-load transistor is four times as large as that of the input transistor. It should also be observed that inverter layout area is dominated by interconnection and spacing area rather than gate area.

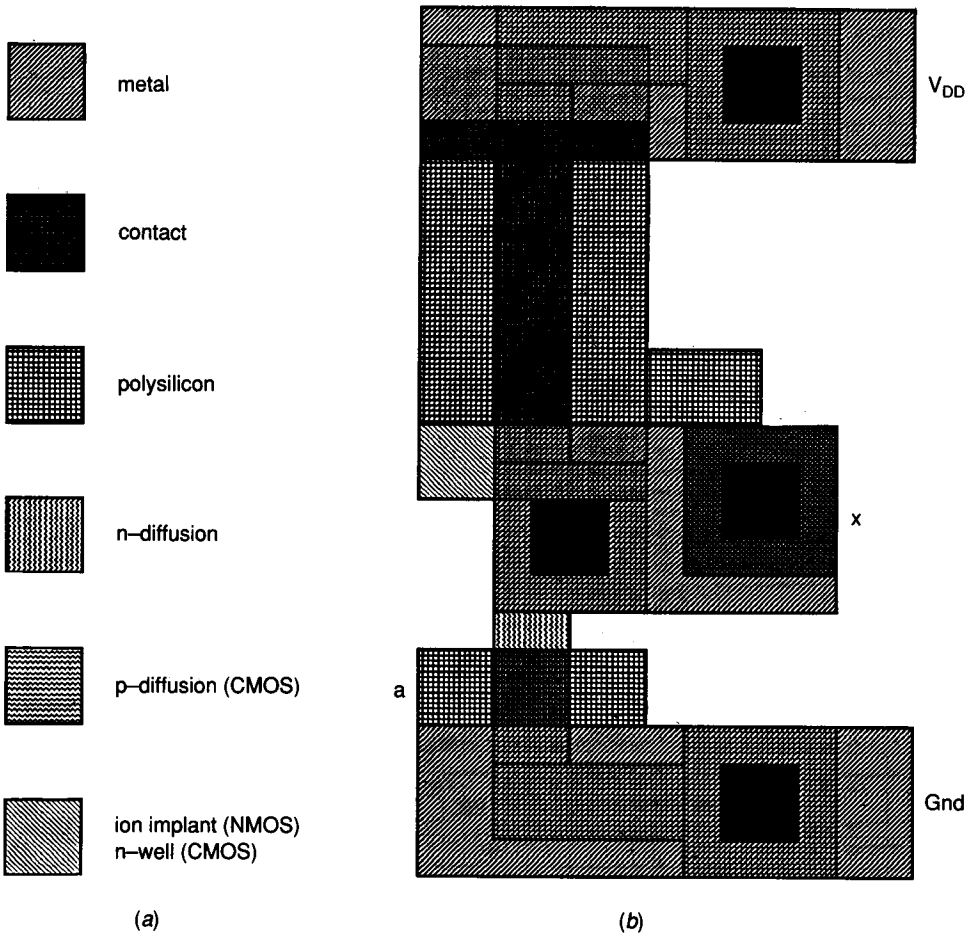


FIGURE 7.3-3
Depletion-load NMOS inverter. (a) Legend, (b) Layout with input *a* and output *x*.

7.3.2 Inverter Device Sizing

The question of how the two transistors forming an inverter should be practically sized naturally arises. Note that for the inverter circuits of Fig. 7.3-1, the highest voltage is determined by the pullup type rather than the pullup size. For example, the depletion-mode pullup acts as a resistor whose I-V characteristic has zero offset voltage from the origin. This causes the output voltage to be pulled to V_{DD} when the enhancement pulldown is off, irrespective of the relative sizes of the transistors. Because transistor sizing has minimal effect on V_H, relative sizing can be advantageously used to put V_M and V_L at acceptable levels.

Consider the depletion-load inverter of Fig. 7.3-1*a*. Because $V_H = V_{DD}$, it would be desirable to have the switching voltage V_M near $V_{DD}/2$ and V_L near 0 V. Observe from the voltage transfer characteristic (Fig. 7.3-2*a*) that M1 and M2 are typically in or near saturation for V_o near $V_{DD}/2$ using reasonable device sizes and process parameters. Hence, assuming that M1 and M2 are saturated, neglecting λ effects and assuming that $V_o = V_i = V_{DD}/2$, it follows from Eq. 3.1-4, upon equating drain currents, that

$$K'\frac{W_1}{L_1}\frac{[(V_{DD}/2) - V_{TN}]^2}{2} = K'\frac{W_2}{L_2}\frac{(V_{TD})^2}{2} \tag{7.3-1}$$

where V_{TN} and V_{TD} are the threshold voltages of the enhancement and depletion devices, respectively. Solving for the relationship between the design parameters, we obtain the expression

$$\frac{W_1L_2}{L_1W_2} = \left[\frac{V_{TD}}{(V_{DD}/2) - V_{TN}}\right]^2 \tag{7.3-2}$$

Defining the geometric device sizing parameter k by

$$k = \frac{W_1L_2}{L_1W_2} \tag{7.3-3}$$

it follows that although the designer has control of the four parameters W_1, W_2, L_1, and L_2, only a single degree of freedom as determined by the parameter k is actually available for controlling V_M. In a typical 5 V digital NMOS process with a desired voltage of 2.5 V for V_M, V_{TD} is set near -3 V and V_{TN} is set near 1 V. Solving Eq. 7.3-2 for the device size ratio gives

$$k = \frac{W_1L_2}{L_1W_2} = \left(\frac{-3}{2.5 - 1}\right)^2 = 4 \tag{7.3-4}$$

This inverter size ratio is important for digital logic circuits and is often termed the *4 : 1 inverter sizing rule*.

With the 4 : 1 size ratio and typical threshold voltages, it is easy to verify the initial assumption that M1 and M2 are at or near saturation when $V_o = V_i = V_{DD}/2$. For positive V_{TN}, the pulldown transistor is saturated for any $V_o = V_i$. For the pullup transistor with $V_{GS} = 0$, saturation occurs when $V_{DD} - V_o > -V_{TD}$. This condition is not satisfied for our example with $V_{DD} = 5$ V, $V_{TD} = -3$ V, and

$V_o = 2.5$ V; however, the shape of the *I–V* characteristic of Fig. 7.8-11 for the depletion-load device shows that the current is nearly constant for voltages near the saturation value. Because V_{DS} for the pullup transistor is within 0.5 V of the saturation value, the assumption of saturation current is reasonable.

Having used the device sizing ratio k to set V_M to $V_{DD}/2$, it is necessary to verify that V_L is near 0 V. To obtain V_L, assume that the inverter input is at V_H, causing $V_o = V_L$. Assume that M1 is in the ohmic region and M2 is saturated (Fig. 7.3-2*a*). Using Eq. 3.1-1 for the ohmic operating region of M1, again neglecting λ effects and equating drain currents, we obtain the expression

$$K'\frac{W_2}{L_2}\frac{(V_{TD})^2}{2} = K'\frac{W_1}{L_1}\left(V_i - V_{TN} - \frac{V_o}{2}\right)V_o \tag{7.3-5}$$

Substituting $V_o = V_L$ and $V_i = V_H = V_{DD}$, it follows from Eq. 7.3-5 with the approximation

$$\left(V_{DD} - V_{TN} - \frac{V_L}{2}\right) \approx (V_{DD} - V_{TN}) \tag{7.3-6}$$

that

$$V_L \approx \frac{W_2 L_1}{L_2 W_1}\frac{V_{TD}^2}{2(V_{DD} - V_{TN})} = \frac{1}{k}\frac{V_{TD}^2}{2(V_{DD} - V_{TN})} \tag{7.3-7}$$

Using the parameters specified earlier and our typical device sizing strategy, it follows that

$$V_L \approx \left(\frac{1}{4}\right)\frac{9}{2(5-1)} = \frac{9}{32}\text{ V} \tag{7.3-8}$$

The assumptions that M1 is ohmic and M2 is saturated are readily verified.

From the previous analysis, we can examine the operation of inverter logic more closely. Consider the flat segment of the transfer characteristic of Fig. 7.3-2*a* corresponding to $V_i < V_{TN}$. For this range of input voltages, the output voltage of a depletion-mode inverter will be at V_{DD}. If the output voltage $V_o = V_{DD}$ is used as the input voltage to a subsequent inverter, its output voltage will be less than V_{TN}, as shown by Eq. 7.3-8. In fact, for depletion-load inverters sized according to the 4 : 1 rule and with normal threshold voltages, any stable interconnection of inverters will have all logic voltages at either V_L or V_H for valid logic inputs.

The previous analysis showed that the 4 : 1 sizing rule gives attractive high and low logic levels that are approximately equidistant from V_M for an inverter. Because no dc loading is experienced when an inverter output drives many inverter inputs within MOS technology, these logic levels are maintained even in the presence of high fan-out. For this reason, the 4 : 1 sizing rule is widely followed in digital logic systems using depletion-load devices.

From a practical viewpoint, it is desirable to minimize area associated with layout. The 4 : 1 sizing ratio specifies only pullup/pulldown ratios and not the width and length of each transistor. Each process technology file specifies design rules for the minimum width and length of a transistor. Often M1 will be a minimum-

size transistor and M2 will have a minimum width resulting in a length for M2 given by

$$L_2 = \frac{4L_1W_2}{W_1} \tag{7.3-9}$$

An alternate construction for an inverter, still satisfying the minimum length and width rules but using less area, can be found (see Prob. 7.12). Subsequently, it will be shown that power and speed considerations may override the minimum area requirement in the determination of the appropriate device sizes.

Throughout the remainder of this book, performance comparisons will be made between a *reference inverter* and other logic circuits. It will be assumed that in any logic family, the sizing of a basic or reference inverter has been specified. The reference inverter used for NMOS will be a depletion-mode inverter formed from a minimum-size enhancement transistor and a minimum-width depletion-mode transistor with length given by Eq. 7.3-9. Although the following discussions are based on this reference inverter, equivalent results may be obtained for reference inverters of other sizes.

It might be instructive to question what improvements, if any, could be obtained by working with a sizing strategy other than 4 : 1. For smaller values of the device sizing parameter k, a modest reduction in area would be attained at the expense of a significant increase in V_L as given by Eq. 7.3-7. This reduces the separation between V_H and V_L, causing the inverter to be more susceptible to noise voltages. Larger values for k offer slight improvements in the signal swing, $V_H - V_L$, at the expense of increases in silicon area. More importantly, it will be shown later that larger k values seriously degrade the switching speed of an inverter.

Finally, it is worth emphasizing that threshold voltages are easily set through process changes. Thus, threshold voltages are chosen specifically to enhance the size and speed characteristics of logic devices. The values $V_{TN} = 1$ V, $V_{TD} = -3$ V and $V_{DD} = 5$ V used in deriving the 4:1 sizing rule are typical for processes used in building digital integrated circuits because they offer reasonable signal swings with practical silicon area requirements.

7.3.3 Enhancement-Load versus Depletion-Load Inverters

Following an analysis similar to the one used for the depletion-load inverter, a sizing rule can be derived for the enhancement-load inverter of Fig. 7.3-1*b*. This analysis would show that the same 4 : 1 sizing rule offers reasonable tradeoffs between signal swing and switching speed. The 4 : 1 sizing rule is thus widely used for single-channel logic circuits.

A comparison of depletion-load and enhancement-load inverters is in order. From a size viewpoint, the silicon area requirements of both types of inverters are nearly the same. From a signal-swing viewpoint, the depletion-load inverter is preferred because the output voltage of the enhancement-load inverter cannot be pulled higher than $V_{DD} - V_{TN}$. From a processing viewpoint, an extra process

step is required for the depletion implant, raising the cost of depletion-mode logic relative to enhancement-mode logic. In terms of switching speed, the depletion-mode inverter is considerably faster than an enhancement-mode inverter in going from a low output voltage to a high output voltage. The reason for this variation is the large difference in equivalent resistances for the two transistors. (Fig. 7.8-11 shows this graphically).

For modern single-channel circuits, the advantages of the depletion-mode inverter far outweigh the cost disadvantage. As a result, enhancement-mode inverters are used only in specialized applications.

Single-channel MOS inverters were analyzed in this section. The inverter device sizing ratio was defined and a ratio of $k = 4$ was found to set the logic threshold near $V_{DD}/2$. Corresponding values of $V_H = V_{DD}$ and $V_L < V_{TN}$ were found with this ratio. A reference inverter was defined as the basis for analyzing a logic family. Depletion-load inverters were found to be preferable to enhancement-load inverters for most applications.

7.4 NMOS NOR AND NAND LOGIC CIRCUITS

In this section NMOS logic gates will be analyzed as extensions of the reference inverter of the previous section. Simple NOR, NAND, and multi-input gates will be considered.

7.4.1 Basic NMOS NOR Logic Circuits

The positive logic NOR function, in which the output is low if any of the inputs are high, is easily realized with NMOS circuits by a parallel connection of two or more pulldown transistors in place of the single pulldown transistor of the inverter circuits of Fig. 7.3-1. Figure 7.4-1 shows a circuit that realizes the positive logic NOR function along with the corresponding logic truth table. Also shown are the Boolean logic equation and the distinctive-shape logic symbol.

The operation of the depletion-load NOR gate will now be considered from a qualitative viewpoint. The analysis of the enhancement-load NOR gate is similar and is left to the reader. If inputs a and b of this circuit are both at a voltage near 0 V, neither of the two parallel pulldown transistors will conduct, and the pullup transistor will pull the circuit output to a voltage near V_{DD}. If either the a or the b input is connected to a voltage near V_{DD}, the corresponding pulldown transistor will conduct, causing the output to be pulled near 0 V irrespective of the other parallel pulldown transistor. Of course, if both inputs are connected to voltages near V_{DD}, then both pulldown transistors conduct, pulling the output to a voltage near 0 V as well.

The device sizing problem for the NOR gate must be addressed. It will be assumed that the device sizing of the NOR gate is done so that for normal input conditions, the high and low logic voltages at the output will be at least as high and low, respectively, as those established for the reference inverter. Thus, a

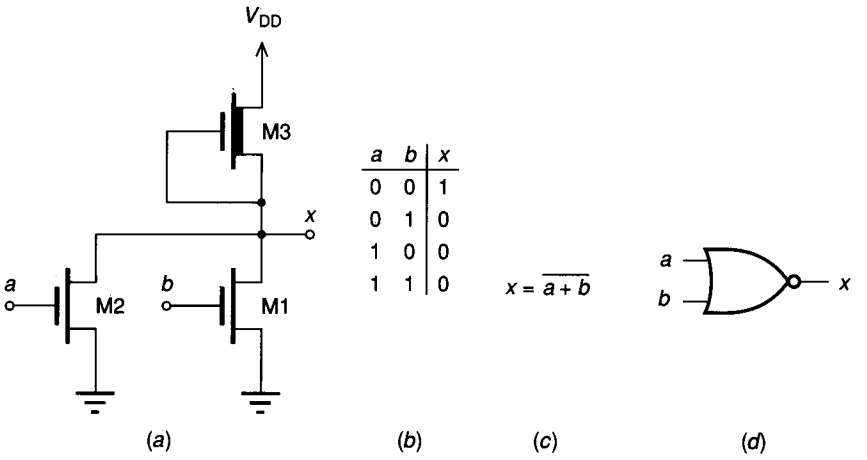


FIGURE 7.4-1
NMOS NOR gate with depletion pullup transistor: (*a*) Circuit, (*b*) Truth table, (*c*) Boolean equation, (*d*) Distinctive-shape symbol.

NOR gate will not cause a deterioration in logic signal levels. The same strategy is generally used to size all logic blocks in a family of logic.

By symmetry, it can be assumed that the pulldown transistors of Fig. 7.4-1, M1 and M2, are the same size. From the corresponding truth table, it is clear that there are only four possible logic conditions for the inputs. These will now be exhaustively considered. If both inputs are low, M1 and M2 are off, causing the output to go to V_{DD} irrespective of any device sizing. If the input to M1 is low and the input to M2 is high, M1 is off and hence allows no current to flow into its drain terminal. This makes the remainder of the NOR gate look like the reference inverter of Fig. 7.3-1*a* with M2 playing the role of the pulldown transistor and M3 the role of the load transistor. To maintain a low voltage level at least as low as that for the reference inverter, M2 and M3 must satisfy the 4 : 1 sizing rule. This results in the requirement

$$\frac{L_3 W_2}{W_3 L_2} \geq 4 \tag{7.4-1}$$

Likewise, if the input to M1 is high and the input to M2 is low, we obtain the requirement

$$\frac{L_3 W_1}{W_3 L_1} \geq 4 \tag{7.4-2}$$

These requirements are equivalent since it was assumed that M1 and M2 are the same size. Finally, if the inputs to both M1 and M2 are high, the M1–M2 combination behaves as a single transistor of width $2W_1$ and length L_1. To maintain the low voltage level, the combination must satisfy the expression

$$\frac{L_3 2W_1}{W_3 L_1} \geq 4 \tag{7.4-3}$$

From Eq. 7.3-7 we can determine that a sizing ratio greater than 4 : 1 causes a lower V_L. The requirements of Eqs. 7.4-1 and 7.4-2 dominate the requirement of Eq. 7.4-3, resulting in the following sizing rule for the NOR gate:

$$\frac{L_3W_2}{W_3L_2} = \frac{L_3W_1}{W_3L_1} = 4 \tag{7.4-4}$$

If each pulldown transistor for a NOR gate is of minimum size, as was the pulldown transistor for the reference inverter, then the load device for the NOR gate should be the same size as the load device for the reference inverter. With this sizing strategy, it is obvious that when both inputs to a NOR gate are high, the output is pulled even lower than the V_L level obtained for the reference inverter. In general, if the sizing rule for the reference inverter is k:1, the corresponding sizing rule for the N-input NOR gate of Sec. 7.4.3 is also k:1.

7.4.2 Basic NMOS NAND Logic Circuits

Figure 7.4-2 shows a circuit that realizes the positive NAND logic function along with its logic truth table, Boolean logic equation, and distinctive-shape logic symbol. For the NAND logic function, the pulldown transistors M1 and M2 are arranged in series rather than in parallel, as was the case for the NOR circuit. The operation of the depletion-load NAND gate will now be considered qualitatively. If either (or both) input a or b is near 0 V, the pulldown transistor corresponding to that input will not conduct. If either pulldown transistor does not conduct, the series path to ground is broken, and the pullup transistor M3 pulls the output

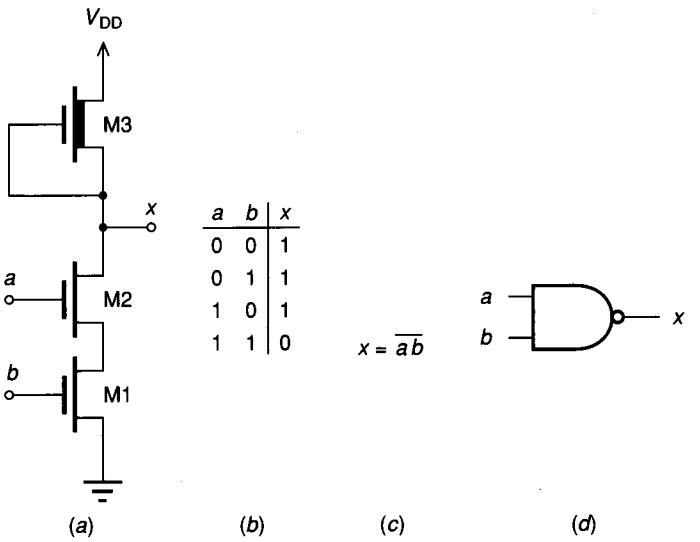


FIGURE 7.4-2
NMOS NAND gate with depletion pullup transistor: (*a*) Circuit, (*b*) Truth table, (*c*) Boolean equation, (*d*) Distinctive-shape symbol.

voltage near to V_{DD}. If both inputs a and b are set to V_{DD}, then both pulldown transistors conduct, pulling the output to a low voltage. This relationship between inputs a and b and output x realizes the NAND logic function.

Device sizing for the NAND gate must be addressed. Once again, it is assumed that the device sizing of the NAND gate is done so that for normal input conditions, the high and low logic voltages at the output will be at least as high and low, respectively, as those established for the reference inverter. With this choice, a NAND gate will not cause a deterioration in logic voltage levels.

The constraint that determines the sizing for the NOR gate corresponds to the weakest pulldown path that will force the output to the low state. For the NAND gate, the *only* pulldown path requires high input levels for both inputs a and b. This condition will determine the sizing for the NAND gate. From a quantitative viewpoint, if M1 and M2 are the same size and both inputs are high, the M1–M2 combination acts as a single transistor of length $2L_1$ and width W_1. To maintain the required output voltage for a low logic level, the 4 : 1 sizing rule can be applied. For the two-input NAND gate this results in

$$\frac{L_3 W_1}{W_3 2L_1} = 4 \tag{7.4-5}$$

or

$$\frac{L_3 W_1}{W_3 L_1} = \frac{L_3 W_2}{W_3 L_2} = 8 \tag{7.4-6}$$

as the sizing rule. In general, if the sizing rule for the reference inverter is k:1, the corresponding sizing rule for the N-input NAND gate of Sec. 7.4.3 is Nk:1.

Two approaches can be followed to achieve the 8:1 ratio for two-input NAND gates. In one approach, the two pulldown transistors are minimum-area devices ($L_1 = W_1 = L_2 = W_2 =$ minimum dimension), requiring the depletion pullup transistor to have the ratio $L_3/W_3 = 8$. In an alternate approach, the load device is sized like the depletion load for the reference inverter, and the two pulldown transistors are sized equally, with $W_1/L_1 = W_2/L_2 = 2$. The first approach is preferred because it provides minimum capacitive loading on preceding logic stages. It can be shown that the total transistor gate area requirement is slightly different for the two sizing strategies (see Prob. 7.14). Also note that, in either case, the minimum total gate area required for the NAND function is larger than that required for the NOR function.

7.4.3 Multi-Input NAND and NOR Logic Circuits

For simplicity, two-input gates were analyzed in the preceding section. Multi-input logic functions, with three or more inputs to a gate, are also widely used. An N-input NOR gate is easily formed by adding $N - 2$ enhancement transistors in parallel with the M1 and M2 transistors of Fig. 7.4-1. An N-input depletion-load NOR gate is shown in Fig. 7.4-3. Multi-input NOR gates that follow the same k:1 sizing rule given for the two-input NOR gate have the same V_L and V_H values as the reference inverter with a sizing ratio of k.

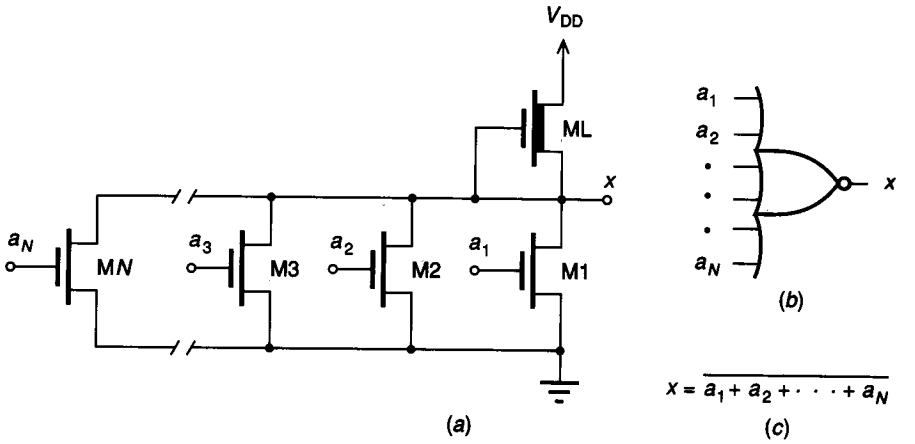


FIGURE 7.4-3
Multi-input NOR gate: (*a*) Circuit, (*b*) Distinctive-shape symbol, (*c*) Boolean equation.

Multi-input NAND gates are formed by adding enhancement transistors in series with the M1 and M2 pulldown transistors shown in Fig. 7.4-2. An N-input NAND gate requires the addition of $N - 2$ transistors in series, as shown in Fig. 7.4-4. Unfortunately, the ratio of the size of the load device to that of the pulldown devices increases with the number of inputs to maintain a good low logic voltage. This sizing ratio becomes Nk:1. An increase in this size ratio increases area and decreases switching speed. As a consequence, single-channel NAND gates with more than two or three inputs are seldom used. Rather, an equivalent multi-input NAND circuit is formed from the appropriate interconnection of inverters and multi-input NOR gates (see Prob. 7.17).

NAND and NOR logic gates are realizable directly as simple modifications to the reference inverter and are more widely used than AND or OR logic gates. When the AND or OR functions are required, they are realized by cascading a basic inverter after a NAND or a NOR gate, respectively.

The structure and sizing of two-input NOR and NAND gates in the NMOS technology were presented in this section. The analysis was generalized to N-input NAND and NOR gates.

7.5 COMPLEMENTARY MOS INVERTERS

The first MOS-based technology to be widely applied at the small-scale integration level was complementary MOS, or CMOS. However, because of the relative simplicity of processing and layout for single-channel MOS circuits, PMOS and then NMOS technologies were the first to achieve widespread use in large-scale integrated circuits. The few early large-scale CMOS circuits were reserved for applications that required the low power dissipation, stable temperature characteristics, or high noise immunity achievable with a CMOS process. Although such characteristics are desirable for all applications of integrated circuit technology, CMOS integrated circuits had disadvantages that limited their use in early VLSI designs.

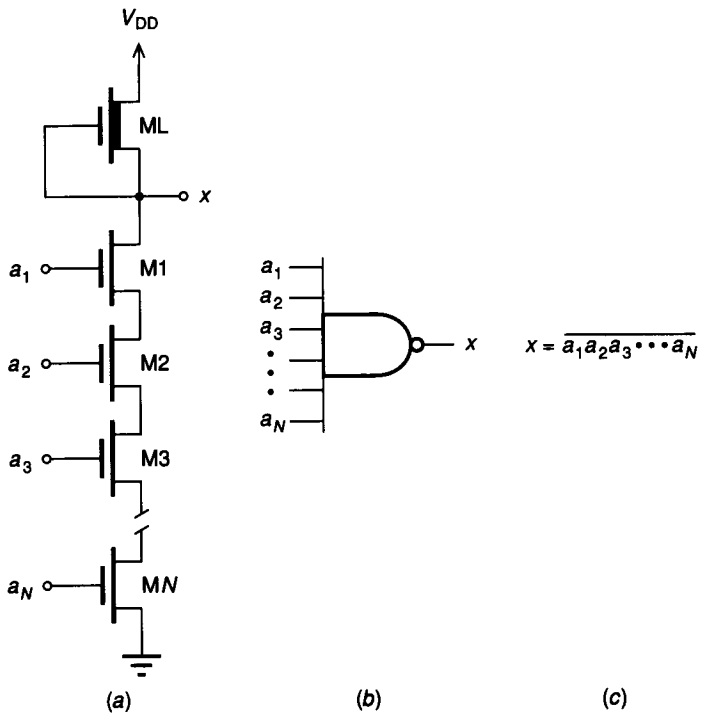


FIGURE 7.4-4
Multi-input NAND gate: (*a*) Circuit, (*b*) Distinctive-shape symbol, (*c*) Boolean equation.

The CMOS fabrication process is more complicated than NMOS fabrication, requiring additional steps to complete the processing of a wafer. Addition of p-channel transistors to a p-substrate process requires creation of an n-well area for the p-channel transistors. This reduces the achievable circuit density compared to an NMOS process. Similarly, a p-well area is required to add n-channel transistors to an n-substrate process. Also, traditional CMOS logic circuits require additional transistors compared to their NMOS counterparts. These disadvantages translate to increased fabrication costs with lower integrated circuit yield for CMOS.

Recent advances in fabrication methods for CMOS technology and in circuit design techniques for CMOS digital circuits have made CMOS more practical for large-scale applications. Also, the need to provide multiple-threshold n-channel devices for speed and power purposes has complicated the standard NMOS process. These factors have changed the relative merits of NMOS versus CMOS technologies for large-scale digital circuit design. Most integrated circuit manufacturers now use CMOS for new designs. For example, both the Intel 80386 and the Motorola 68030 microprocessors are fabricated with CMOS processes. Also, many integrated circuit manufacturers offer CMOS gate array circuits to provide large-scale integration of digital systems functions. In this section, basic CMOS inverter circuits are discussed.

7.5.1 A Basic CMOS Inverter

A CMOS inverter is similar to an NMOS inverter in that two transistors are required. In the CMOS inverter of Fig. 7.5-1*a*, the p-channel transistor M2 acts as a pullup device, and the n-channel transistor M1 acts as a pulldown device. In fact, for some CMOS processes (n-well process), the pulldown transistor is identical to its NMOS counterpart. The primary distinction between NMOS and CMOS inverters lies in the characteristics of the pullup transistors. For NMOS, either a passive enhancement pullup device or a passive depletion pullup device is used, with the depletion pullup preferred from a speed standpoint. In a CMOS process, an actively driven p-channel transistor is normally used as the pullup device. This p-channel pullup transistor is turned on by a gate signal that is opposite in polarity to the gate signal required to turn on the n-channel pulldown transistor. The digital symbol for the p-channel pullup device, shown in Fig. 7.5-1*b*, emphasizes this complementary polarity by showing the gate connection with a bubble to represent signal inversion. This symbol is preferred for digital CMOS circuits. As explained in Sec. 7.0, source designations are assumed understood with this notation. The use of opposite-polarity (complementary-polarity) transistors for load devices characterizes the CMOS inverter.

7.5.2 CMOS Inverter Logic Levels

Note that in the CMOS inverter of Fig. 7.5-1, the gates of both transistors are connected to the inverter input, unlike the comparable NMOS inverter circuit of Fig. 7.3-1. This circuit functions as an inverter because of the opposite transfer characteristics of the n-channel pulldown transistor and the p-channel pullup transistor. The threshold voltages, V_{TP} and V_{TN}, are established so that a voltage $V_i > V_{DD} + V_{TP} = V_{DD} - |V_{TP}|$ on the common gate node causes the n-channel device to conduct while it causes the p-channel device to turn off. (To reduce the potential for error, the notation $(-|V_{TP}|)$ will replace (V_{TP}) to ensure that

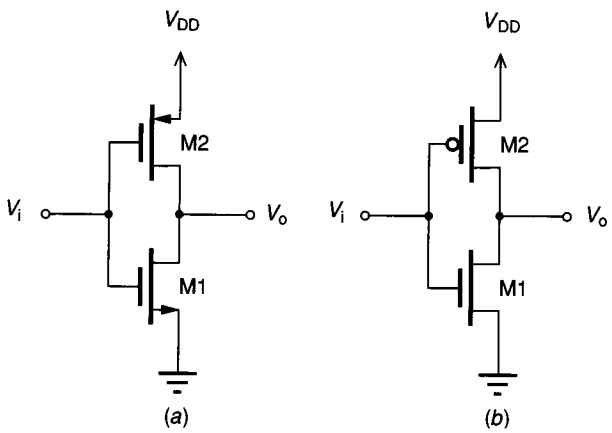


FIGURE 7.5-1
CMOS inverter circuit: (*a*) n-channel and p-channel designation, (*b*) Preferred notation for digital circuits.

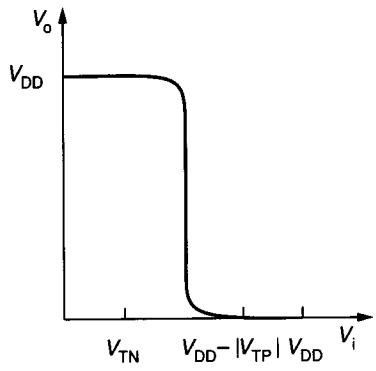


FIGURE 7.5-2
Voltage transfer characteristic for CMOS inverter.

a negative value is used in calculations with V_{TP}.) This results in an active pulldown capability at the output of the circuit. Because the pullup device is off, the pulldown transistor does not have to compete to establish a low-voltage condition on the output. At steady state, the output voltage will reach $V_o = 0$ V. On the other hand, a voltage $V_i < V_{TN}$ on the common gate node causes the n-channel device to turn off while it causes the p-channel device to conduct. This condition provides an active pullup transistor without competition from a pulldown device, allowing the inverter output to be pulled high. At steady state, $V_o = V_{DD}$ for this condition. Figure 7.5-2 shows a typical voltage transfer characteristic for a CMOS inverter.

For valid logic input voltages, note that when one transistor of a CMOS inverter is conducting, the complementary transistor is off. This condition is especially important for two reasons. First, it virtually eliminates static power dissipation in CMOS logic circuits because no current can flow from the voltage supply to ground through the inverter in steady state. Second, it allows a maximum logic voltage swing equal to the supply voltage V_{DD}.

Consider the steady state effect of a logic high voltage on the cascade of CMOS inverter stages of Fig. 7.5-3. Assume that a logic high voltage $V_{i1} > V_{DD} - |V_{TP}|$ is provided at the input to the first inverter. The output of this inverter will reach $V_{o1} = 0$ V because the pullup transistor is off for this input condition. Thus, the input to the second inverter is $V_{i2} = V_{o1} = 0$ V. This gives the condition $V_{i2} < V_{TN}$, causing the n-channel pulldown device of the

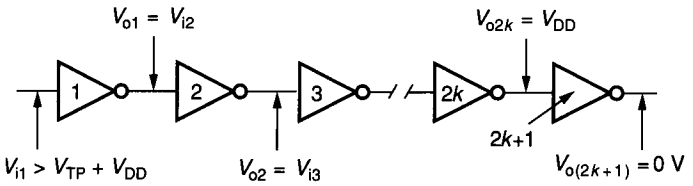


FIGURE 7.5-3
CMOS inverter cascade.

second inverter to be off. The output is then pulled to $V_{o2} = V_{DD}$ by the p-channel transistor. An extension of this analysis to the kth inverter pair shows that $V_{TN} > V_L = 0$ V and $V_{DD} - |V_{TP}| < V_H = V_{DD}$. This provides an ideal voltage swing of

$$V_H - V_L = V_{DD} - 0 \text{ V} = V_{DD} \tag{7.5-1}$$

The best possible voltage swing for a single-polarity inverter output is equal to the difference between the upper and lower supply voltages to the inverter—in this case, V_{DD} and 0 V. It should be obvious that an analysis starting with a logic low voltage of $V_{il} < V_{TN}$ for the CMOS inverter cascade will yield the same values for V_H and V_L.

7.5.3 Inverter Device Sizing

Device sizing for a CMOS inverter will now be considered. The ability of either transistor in a CMOS inverter to act as a low-resistance device with the complementary transistor turned off allows the design of ratioless inverters. A *ratioless* logic device is one in which, from a logic-level viewpoint, the steady state output voltage levels are independent of the ratio of the pullup and pulldown transistor sizes. This is in contrast to the single-channel inverters of Sec. 7.3, which are known as *ratio* logic devices because the pullup/pulldown size ratio k determines one of the equilibrium voltages V_L. The independence of logic voltage levels and transistor sizes allows other considerations to determine the relative sizes of the pullup and pulldown transistors in a CMOS inverter.

The primary effect of the sizes of the pullup and pulldown transistors is on the equivalent resistance of the transistors in the conducting state. Thus, sizing can be used to provide approximately equal capability to source or sink load current; this equality is termed *symmetric output drive*. In contrast, NMOS inverters provide asymmetric output drive because of the difference in pullup and pulldown resistance needed to achieve useful logic levels. The symmetric drive capability of CMOS allows comparable transition times for output voltages irrespective of the direction of the transition.

It is interesting to determine the relative sizes of n-channel and p-channel transistors required to achieve symmetric output drive for a CMOS inverter. An n-channel transistor for either an NMOS or an n-well CMOS process can be fabricated with similar characteristics. The minimum channel lengths and widths of transistors in both processes are set primarily by the resolution of the fabrication process. The transconductance parameter K'_N of an n-channel transistor is about two or three times greater than the transconductance parameter K'_P of a p-channel transistor; a factor of 2.5 is used in this analysis. This factor is caused primarily by the difference in majority carrier mobility for the p- and n-channel transistors described previously. The transconductance parameter difference influences the effective pullup and pulldown resistances of the p-channel and n-channel transistors. The equivalent resistances for the n-channel transistor R_N and for the p-channel transistor R_P are directly proportional to L and inversely proportional to K' and W as indicated by equations

$$R_N \propto \frac{L_N}{W_N K'_N} \tag{7.5-2}$$

and

$$R_P \propto \frac{L_P}{W_P K'_P} \tag{7.5-3}$$

Simple models for the n-channel and p-channel transistors composed of an ideal switch and the effective pullup and pulldown resistances are shown in Fig. 7.5-4 (also see Sec. 3.1-9). The switch is open or closed corresponding to the gate-to-source voltage that turns the transistor off or on, respectively. This simple model is useful for predicting current drive capability and approximating propagation delays in digital applications. Because the proportionality constants for Eqs. 7.5-2 and 7.5-3 are comparable, it follows that

$$R_N \approx \left(\frac{L_N}{W_N K'_N}\right)\left(\frac{W_P K'_P}{L_P}\right) R_P \tag{7.5-4}$$

For symmetric output drive, it is required that $R_N = R_P$. If $K'_N = 2.5K'_P$, then according to Eq. 7.5-4, the devices must be sized with

$$\frac{L_N W_P}{W_N L_P} = \frac{K'_N}{K'_P} = 2.5 \tag{7.5-5}$$

With this sizing, the n-channel and p-channel transistors have symmetric *I–V* characteristics. The layout for a CMOS inverter with these characteristics is given in Fig. 7.5-5. Note that the p-channel transistor width is 2.5 times the width of the n-channel transistor. In practice, symmetric drive capability is often sacrificed by the use of minimum-size transistors to reduce layout area.

In this section a basic CMOS inverter was analyzed. It was found that the output logic voltage levels of $V_L = 0$ V and $V_H = V_{DD}$ were ideal. Because the CMOS inverter is a ratioless logic circuit, device geometries can be used to obtain symmetric output drive.

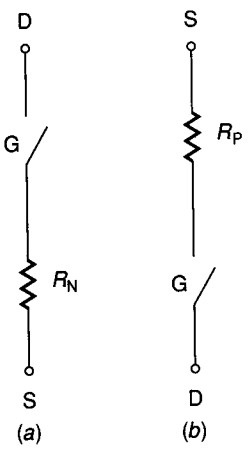


FIGURE 7.5-4
Simple transistor models for digital applications: (*a*) n-channel, (*b*) p-channel.

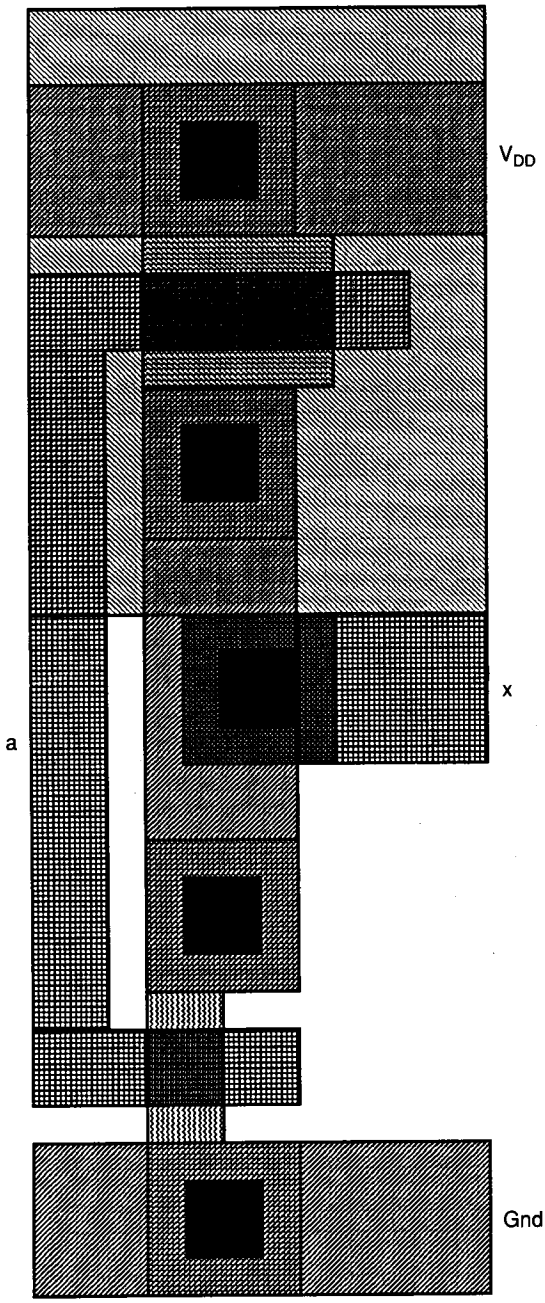


FIGURE 7.5-5
Layout for CMOS inverter with input a and output x.

7.6 CMOS LOGIC GATES

Classical CMOS logic gates are described in this section. As their characteristics are developed, a comparison with the corresponding NMOS logic gates will be given. Two-input NOR and NAND logic gates are used to develop the unique characteristics of CMOS logic. A discussion of multi-input CMOS gates will conclude the section.

7.6.1 CMOS NOR Logic Gate

Once the reference inverter circuit for a particular logic family is defined, the design of circuits to implement other logic functions can be based on that inverter circuit. Figure 7.6-1 shows the circuit schematic for a CMOS two-input positive logic NOR gate based on the CMOS inverter of Fig. 7.5-1. This circuit is obviously different from the NMOS NOR logic circuit of Fig. 7.4-1 because four transistors are required to implement the logic function rather than three. The two pulldown transistors, M1 and M2, are in parallel as they were for the corresponding NMOS logic gate. However, two pullup transistors, M3 and M4, are required in a series connection to complete the CMOS NOR circuit. The layout of a two-input CMOS NOR gate with all minimum-size transistors is shown in Fig. 7.6-2. From this figure, it can be observed that greater silicon area is required compared to an equivalent layout of the NMOS NOR gate of Fig. 7.4-1 given a common basic process resolution. The slight decrease in pullup transistor area for the CMOS gate—the longer depletion pullup of NMOS is unnecessary—is more than offset by the additional circuit connections, the second pullup transistor, and the n-well isolation required for the CMOS pullup transistors.

The operation of the CMOS NOR gate of Fig. 7.6-1 can be explained as follows. When both inputs a and b are below V_{TN}, the parallel n-channel pulldown transistors are off and the series p-channel pullup transistors conduct. This condition provides an active pullup through the series transistors for the output of the CMOS gate. There is no competition from the pulldown transistors.

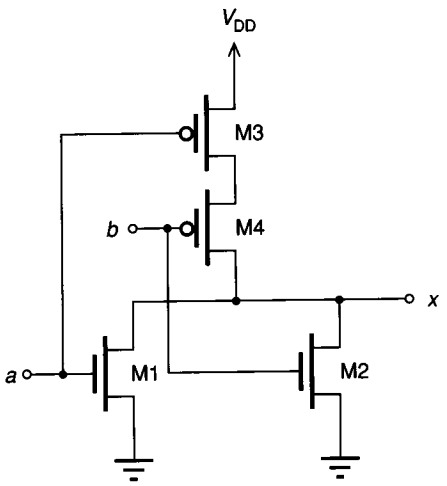


FIGURE 7.6-1
Two-input CMOS NOR gate.

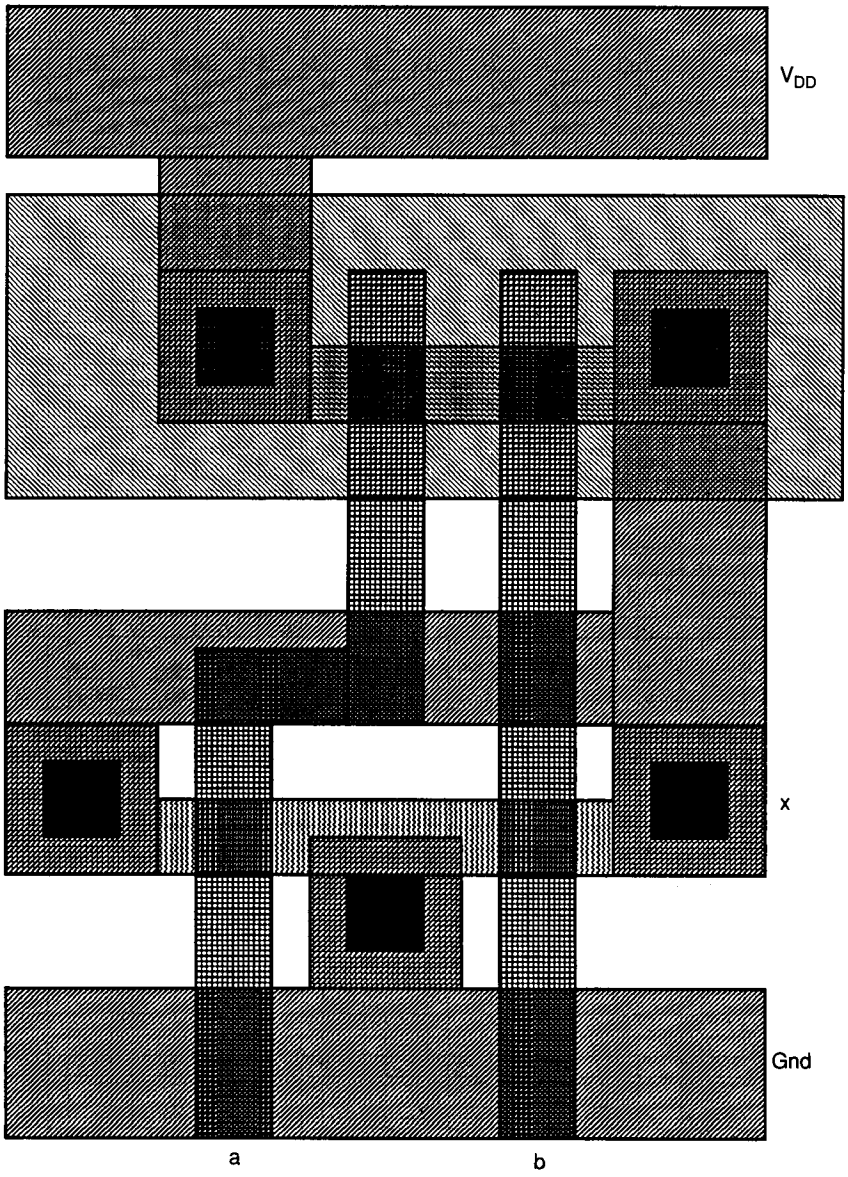


FIGURE 7.6-2
CMOS NOR gate layout with inputs a and b and output x.

Using the resistive model of Sec. 7.5 for the p-channel transistors, it follows that

$$R_{UP} = R_{P3} + R_{P4} \propto \frac{2L_P}{W_P K'_P} \tag{7.6-1}$$

Thus, for equally sized pullup transistors, the equivalent pullup resistance is twice that of either pullup alone.

If input a is connected to a voltage greater than $V_{DD} - |V_{TP}|$, p-channel pullup transistor M3 is off and n-channel pulldown transistor M1 conducts. Now the series path to V_{DD} is broken, and the output node is pulled to 0 V by transistor M1. This gives a pulldown resistance

$$R_{N1} \propto \frac{L_1}{W_1 K'_N} \tag{7.6-2}$$

A corresponding analysis for input b high, which turns M4 off, shows that the output node is pulled to 0 V by transistor M2. Now the pulldown resistance is

$$R_{N2} \propto \frac{L_2}{W_2 K'_N} \tag{7.6-3}$$

If both inputs a and b are above $V_{DD} - |V_{TP}|$, then both transistors M3 and M4 are off and transistors M1 and M2 both conduct. This parallel connection results in a pulldown resistance that is equal to the parallel combination of R_{N1} and R_{N2}, thus ensuring that the output node is pulled to 0 V. Because the output is pulled low if input a or b or both are high, this circuit realizes the NOR logic function.

For the CMOS NOR gate, as for the CMOS inverter, the steady state output voltage is set by a ratioless connection of conducting transistors. Thus, the dc logic voltages are ideal; that is, the output voltage is pulled to either V_{DD} or 0 V. Because dc logic levels are not affected by the relative sizes of the pullup and pulldown transistors, these transistors can be sized to obtain the desired output drive characteristics, as was done previously for the CMOS inverter. Because of the presence of two inputs (and four logic input conditions) for the NOR gate, it is not possible to maintain symmetric output drive capabilities for all input conditions. A common approach is to size the devices so that the worst-case drive capability is as good as that of the reference inverter. If this strategy is adopted, then M3 and M4 can each be sized for half the effective pullup resistance of the reference inverter. If M1 and M2 are both of minimum size, then M3 and M4 would both have $W/L = 5$ to maintain this worst-case drive capability. In practice, minimum-size transistors are often used for both pullups and pulldowns to conserve area, resulting in asymmetric output drive.

A simple resistive model of the two-input NOR gate based on the device model of Fig. 7.5-4 is shown in Fig. 7.6-3. If all devices are of minimum size, then the pullup resistance is $2R_P = 5R_N$ for $K'_N = 2.5K'_P$. The pulldown resistance is R_N for a single NOR gate input high or $R_N/2$ for both inputs high. Thus, a maximum input-dependent asymmetry of 10:1 exists for this CMOS NOR structure based on minimum-size gates.

7.6.2 CMOS NAND Logic Gate

The circuit for a CMOS two-input positive logic NAND gate is also easily obtained once the reference inverter circuit has been defined. A CMOS NAND circuit is shown in Fig. 7.6-4 with its corresponding layout in Fig. 7.6-5. This circuit is called the *dual* of the CMOS NOR circuit because the two pullup transistors are connected in parallel rather than in series and the two pulldown transistors

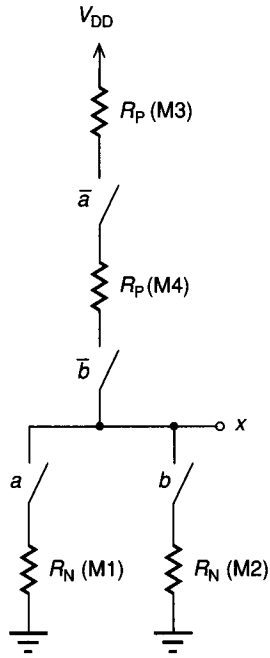


FIGURE 7.6-3
Simple resistive model of CMOS NOR gate.

are connected in series rather than in parallel. Once again, four transistors are required for a CMOS two-input NAND circuit, as compared with only three transistors for the similar NMOS circuit of Fig. 7.4-2.

The operation of the CMOS NAND circuit is similar to the CMOS NOR circuit except for the logic function realized. The interchange of the parallel and series connections for the pullup and pulldown paths changes the output voltage generated for specific input conditions, resulting in a realization of the NAND function. If both inputs a and b are below V_{TN}, the series n-channel pulldown

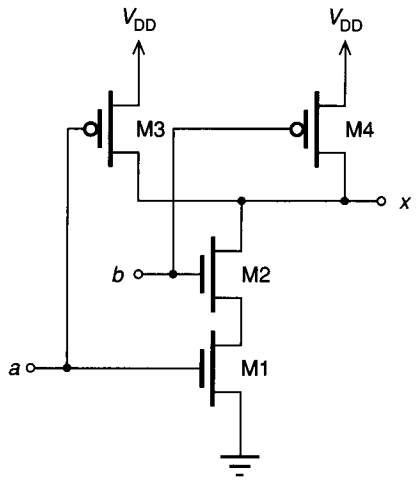


FIGURE 7.6-4
Two-input CMOS NAND gate.

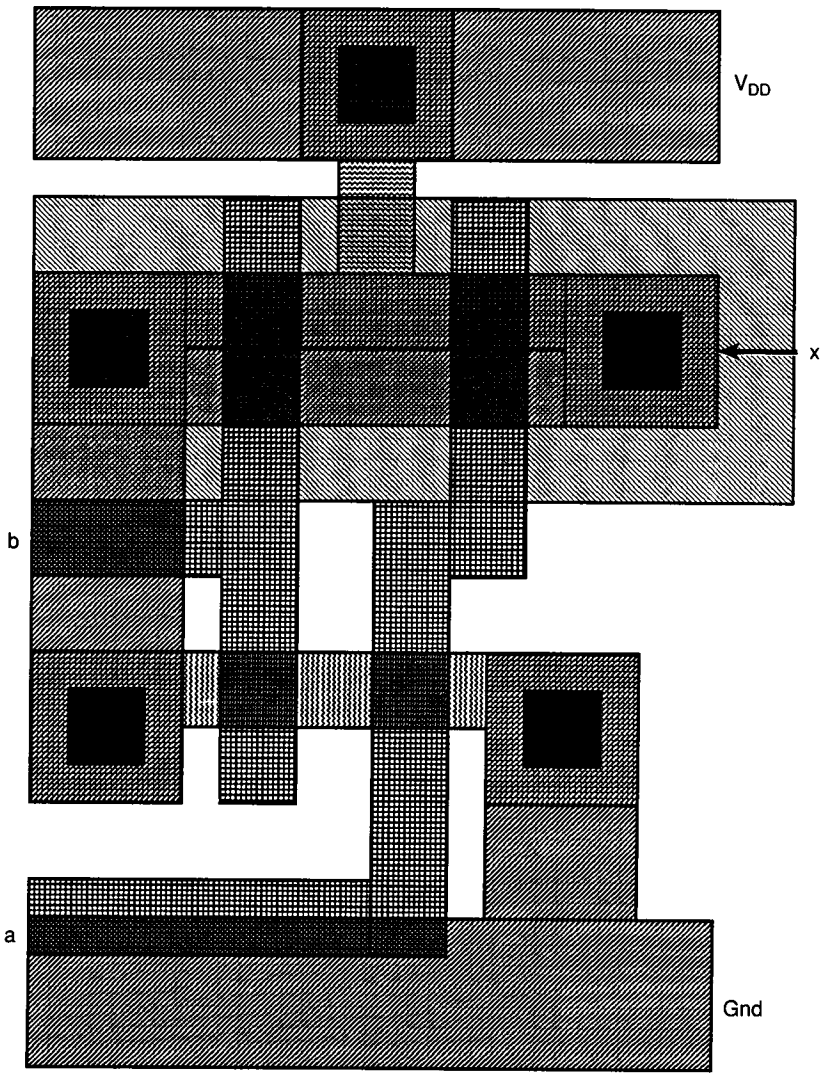


FIGURE 7.6-5
CMOS NAND gate layout with inputs a and b and output x.

transistors M1 and M2 are off while the parallel p-channel pullup transistors M3 and M4 conduct. This results in an output voltage level of V_{DD}. If only input a is below V_{TN}, then transistor M1 is off while transistor M3 conducts. Thus, the output voltage is still pulled to V_{DD}. If only input b is below V_{TN}, transistor M2 is off and transistor M4 conducts, again setting the output to V_{DD}. If both inputs a and b are above $V_{DD} - |V_{TP}|$, then transistors M3 and M4 are off and transistors M1 and M2 conduct. This condition sets the output voltage to 0 V. This input-output voltage relation realizes the positive NAND logic function.

For the NAND gate, the series path to ground uses n-channel transistors, and the parallel path to V_{DD} uses p-channel transistors. Based on minimum-size

transistors and a 2.5:1 transconductance advantage of n-channel over p-channel transistors, the series resistance of two n-channel transistors to ground will be nearly matched by a single p-channel pullup to V_{DD}. To demonstrate this, consider the NAND gate model of Fig. 7.6-6 based on the simple resistive models of Fig. 7.5-4. Let all transistors corresponding to Fig. 7.6-6 be of minimum size. For this circuit, the effective pulldown resistance is $2R_N$ when both inputs are high. The effective pullup resistance is either R_P or $R_P/2$ depending on whether only one or both inputs are low, respectively. Noting that $R_P = 2.5R_N$ for $K'_N = 2.5K'_P$, the worst-case input-dependent asymmetry is less than 2:1, which is much better than that obtained for the corresponding NOR circuit. This near-symmetry makes the NAND gate the preferred CMOS logic form. Note that a CMOS NOR gate requires greater silicon area to achieve either a lower resistance p-channel path or a higher resistance n-channel path to reduce the worst-case input-dependent asymmetry.

7.6.3 Multi-Input CMOS Logic Gates

Classical multi-input CMOS logic gates are formed by adding an n-channel pulldown and a p-channel pullup transistor for each additional input. For the NOR gate, the pulldown transistor is added in parallel with the other n-channel transistors, and the pullup transistor is added in series with the other p-channel transistors. A three-input CMOS NOR gate is shown in Fig. 7.6-7. The three-input CMOS NAND gate of Fig. 7.6-8 is formed by inserting a pulldown transistor in series with the n-channel transistors and adding a pullup transistor in parallel with the p-channel transistors.

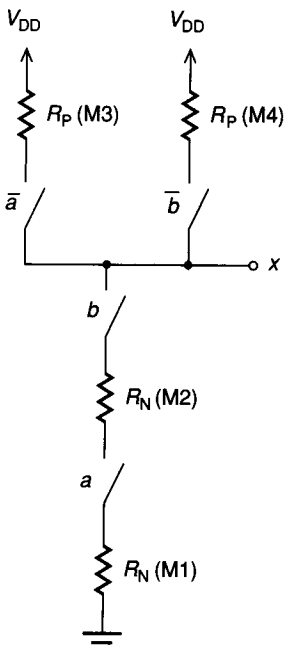


FIGURE 7.6-6
Simple resistive model of CMOS NAND gate.

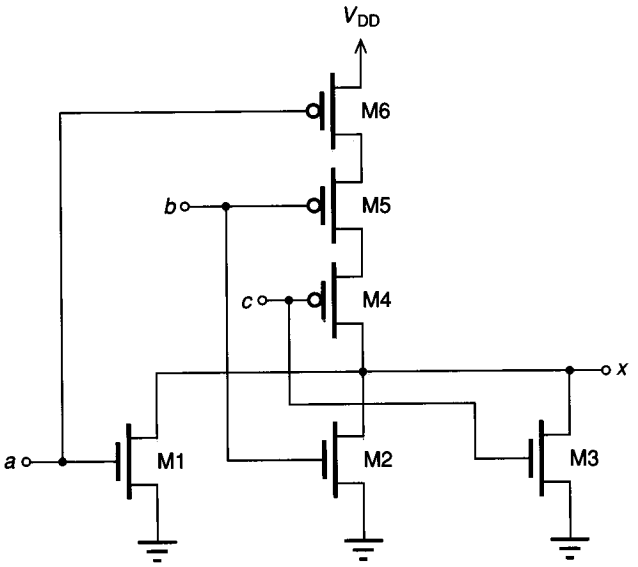


FIGURE 7.6-7
Three-input CMOS NOR gate.

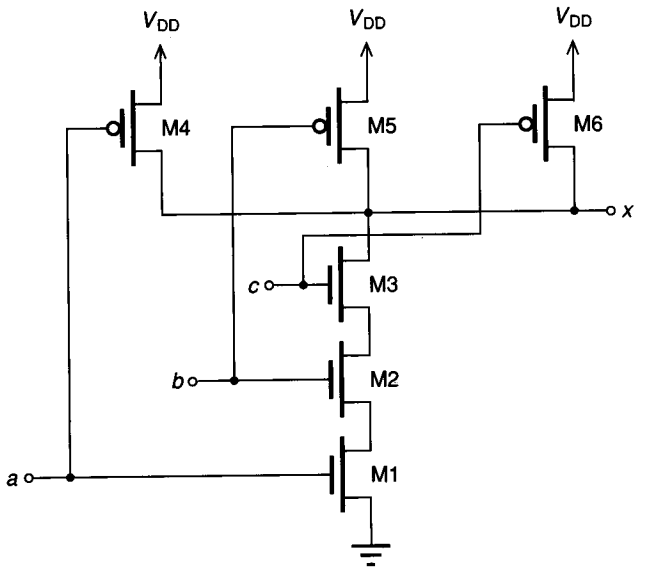


FIGURE 7.6-8
Three-input CMOS NAND gate.

Two characteristics of classical multi-input CMOS logic gates limit their use in VLSI circuits. Let N be the number of inputs to a multi-input logic gate. Let M_{NMOS} be the number of transistors required to form an N-input gate in NMOS or PMOS, and let M_{CMOS} be the number of transistors needed to form an N-input gate in CMOS. The number of transistors required to form NAND or NOR gates in either an NMOS or PMOS technology is

$$M_{NMOS} = N + 1 \tag{7.6-4}$$

whereas the number of transistors necessary to form classical NAND or NOR gates in a CMOS technology is

$$M_{CMOS} = 2N \tag{7.6-5}$$

The CMOS logic circuits introduced in this section are designated *classical* CMOS logic because of their relatively long history of use. Classical multi-input CMOS logic gates require more transistors than their NMOS counterparts for any number of inputs. As the number of inputs to a multi-input logic gate grows, a CMOS gate requires approximately twice as many transistors as an NMOS gate. To circumvent this disadvantage, clever circuit structures are commonly used for multi-input gates in CMOS. One of these techniques, domino CMOS, is discussed in a subsequent section. Alternatively, NMOS-like structures are sometimes used in CMOS, with a p-channel pullup with its gate grounded used to emulate the depletion-load device of NMOS. Like NMOS logic, this pseudo-NMOS logic has the disadvantage of static power dissipation.

A second disadvantage of classical multi-input CMOS gates arises from the requirement for a series transistor path from the output to one of the power supply nodes. For a NAND gate, this series path is the pulldown path. For a NOR gate, the series path is the pullup path. As the number of inputs to a classical CMOS gate grows, it becomes difficult to size the transistors for proper output drive through the series path. Remember that the multi-input NMOS NOR gate of Sec. 7.4 did not have this disadvantage.

Classical CMOS logic gates and their characteristics were described in this section. As with the CMOS inverter, ratioless CMOS logic gates provide ideal logic voltage levels and allows device sizing to be used for near symmetrical output drive. As a disadvantage, classical CMOS logic gates require almost twice as many transistors as their NMOS counterparts.

7.7 TRANSMISSION GATES

Because a MOS transistor is an excellent switching device, as demonstrated in Sec. 5.1, it is possible to connect this transistor in series with a logical signal to either pass or inhibit the signal. A MOS transistor connected in this way is called a *pass transistor* or *transmission gate* because it passes or transmits signals under control of its gate terminal. This connection has several advantages and some disadvantages compared to other methods of controlling a logic signal. In this section, characteristics of pass transistors are analyzed for NMOS circuits. Then the transmission gate structure for CMOS circuits is studied.

7.7.1 NMOS Pass Transistor

Figure 7.7-1 shows an input signal V_i connected through an n-channel transistor to the input V_a of a standard inverter circuit. A transistor M1 connected in this manner is called a *pass transistor*. This same circuit connection is important in dynamic storage circuits, to be discussed in Sec. 9.6. If the gate voltage, V_G, of the pass transistor is held at 0 V, and if the drain and source voltages are constrained between 0 V and V_{DD}, an enhancement pass transistor can never conduct because a positive gate-to-source voltage cannot be established. Because of the high off-resistance of an enhancement transistor, the input V_i is effectively disconnected from the inverter circuit.

If, on the other hand, V_G is held at V_{DD}, the pass transistor will conduct to equalize the voltages at its source and drain. This is explained by considering the alternatives with $V_G = V_{DD}$. First, assume that V_i, the input voltage to the pass transistor, is held at a low logic level: $V_i = 0$ V. If the input terminal of the pass transistor is considered the source, the gate-to-source voltage is V_{DD} and the pass transistor conducts to bring the drain terminal voltage V_a to 0 V. Note that no current can be supplied from the inverter input, an oxide-insulated gate.

Next assume that the input is held at $V_i = V_{DD}$. The pass transistor may appear to be off because both V_G and V_i are at the same voltage. However, because an enhancement transistor is symmetrical with respect to the drain and source terminals, the V_a terminal of the pass transistor can act as the source. If the voltage at the inverter input $V_a = 0$ V, the gate-to-source voltage of the pass transistor is initially V_{DD}, the transistor will conduct, and the inverter input V_a will be pulled to a higher voltage. Note that as V_a increases, the pass transistor gate-to-source voltage is reduced. When V_a reaches $V_{DD} - V_{TN}$, the pass transistor will cease to conduct. Thus, the pass transistor can only pull the inverter input to one threshold voltage drop below the pass transistor gate voltage. This connection of the pass transistor with the V_i and V_G terminals held at V_{DD} is the same circuit configuration as the enhancement pullup load for the MOS inverter discussed in Secs. 6.1 and 7.3.

One last alternative must be considered. If both V_G and the input voltage V_i of the pass transistor are at V_{DD} and the inverter input $V_a = V_{DD}$, the pass transistor will not conduct. However, because the logic levels at the source and drain terminals of the pass transistor are already equivalent, the logical effect is the same as if the pass transistor were conducting. This analysis is summarized by the following two equations.

$$V_G = 0 \text{ V} \rightarrow \text{Pass transistor is off} \tag{7.7-1}$$

$$V_G = V_{DD} \rightarrow \text{Pass transistor conducts} \tag{7.7-2}$$

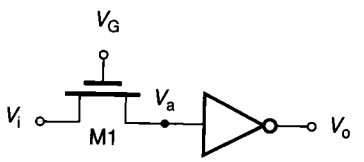


FIGURE 7.7-1
Input signal connected to an inverter through a pass transistor.

At this point the reader might infer that a pass transistor could be used to logically AND two input signals. One signal could be applied to the pass transistor input terminal, and the other could be applied to the pass transistor gate terminal. The output of the pass transistor would be pulled high only if both the gate terminal *and* the input terminal of the pass transistor were high. This is correct; however, the converse is not true. The output of the pass transistor will be pulled low if its input terminal is low, but not if its gate terminal is low. According to Eq. 7.7-1, when the gate is at 0 V, the pass transistor is off and cannot force the output to a low voltage to satisfy the AND functional requirements. Thus, this connection does not function as an AND gate. The assumption by a designer that a pass transistor can function as an AND gate would be a simple but fatal (to circuit operation) mistake.

A pass transistor used as a logic switch has important advantages in terms of integrated circuit layout constraints. First, the pass transistor consists of a single transistor and therefore requires less area than a logic gate. Even the simplest logic gate, an inverter, requires two transistors. Additionally, a pass transistor is a three-terminal device, whereas an inverter is a four-terminal device if one counts power and ground. A requirement for fewer interconnections is a major advantage when integrated circuit layout constraints are considered. For many applications, the pass transistor can be a minimum-size device, further reducing layout area. Finally, a pass transistor requires no dc power—a significant advantage.

A typical use of pass transistors to create a 4-to-1 selector circuit is shown in Fig. 7.7-2. A circuit that can connect one of N different inputs to an output is called an N-to-1 *selector circuit*. Here 8 pass transistors replace an equivalent of 21 NMOS transistors or 32 CMOS transistors if classical logic gates are used. This selector circuit can realize all 16 logic functions of the two inputs a and

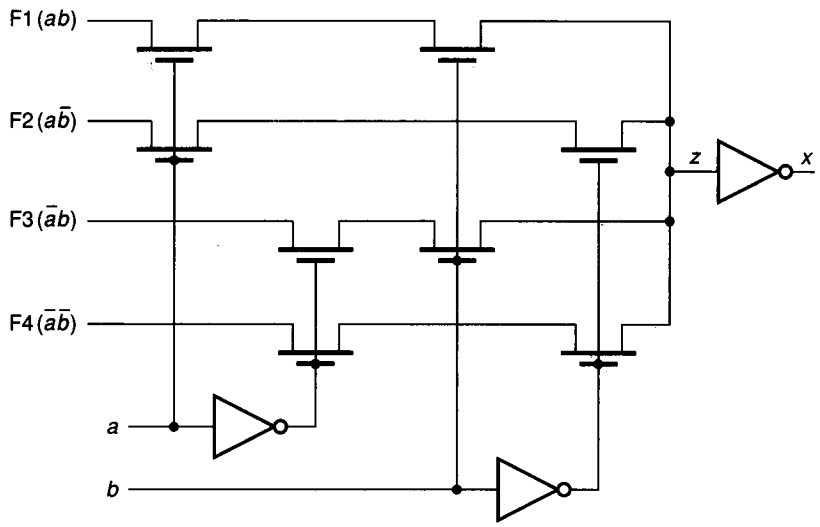


FIGURE 7.7-2
Pass transistor selector circuit.

b. For example, if F1–F4 are all set to 0, then the output z is 0. If F1 is set high and F2–F4 are set low, then the AND of a and b is given at z. If F2 and F3 are set high while F1 and F4 are set low, then the exclusive-OR function of a and b is realized at z. If F1–F3 are set high and F4 is set low, then a OR b is given at z. The 16 possible combinations of 0 or 1 at the four inputs F1–F4 yield the 16 possible logic functions of a and b. This selector circuit function block is popular within microprocessor ALUs, where both logic and arithmetic functions must be realized. Note that power dissipation, number of devices, and interconnection requirements are considerably reduced from an equivalent logic gate implementation. This is but one example of the simplicity available through the use of pass transistor logic.

A series string of pass transistors connected to control the logical path of an input signal is an appealing circuit configuration in terms of area and interconnection constraints compared to other alternatives. For example, such a string is frequently used to selectively propagate the carry in a multibit binary adder. However, at least two significant problems are encountered with a series string of pass transistors. The first problem arises because of design imposed constraints on signal propagation delays. Figure 7.7-3 shows a series string of pass transistors connecting V_i to V_o with all gates held high and an approximate equivalent circuit using a lumped circuit element model. R represents equivalent resistance between the source and drain of a pass transistor. (Bulk effects increase the equivalent resistance here because the source terminal is not grounded.) C represents the capacitance to ground for each stage; the value is determined by the gate capacitance and the drain and source diffusion capacitance of each pass transistor. It can be shown that the signal propagation delay from V_i to V_o is proportional to the square of the number of identical stages.

A simplified explanation for the square-law delay for a series string of pass transistors can be given in terms of the resistance, capacitance, and associated RC time constants of the lumped circuit element model. A single pass transistor stage exhibits a delay that is a function of the series resistance R and the capacitance to ground C. This delay is proportional to the time constant $\tau = RC$. When a second

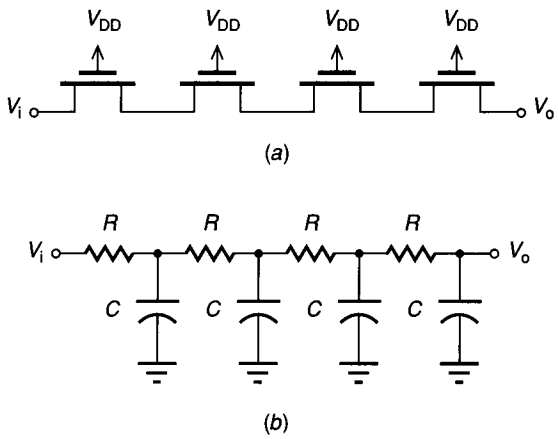


FIGURE 7.7-3
(*a*) Series string of pass transistors, (*b*) Lumped-element equivalent circuit.

pass transistor is added in series with the first and only the additional capacitance to ground is considered, the RC time constant ($\tau = R2C$), and therefore the transmission delay, is doubled. If the additional series resistance is considered but the additional capacitance to ground is neglected ($\tau = 2RC$), the delay is still doubled. The combination of twice the resistance and twice the capacitance ($\tau = 2^2RC$) causes the total delay to nearly quadruple. In general, the delay is proportional to N^2RC where N is the number of series pass transistors. For this reason, long series connections of pass transistors are usually segmented by the addition of inverters at intervals of about four pass transistors. The inverters buffer the signal and break the square-law delay effect of the series pass transistor cascade, resulting in smaller overall signal delay.

A second problem is encountered if pass transistors are cascaded with the output of one pass transistor connected to the gate of the next pass transistor as shown in Fig. 7.7-4. This problem is related to the threshold voltage drop from the gate of a pass transistor to its output terminal. As described previously for enhancement pullup transistors, the source voltage can only be pulled to a value V_{TN} less than the gate terminal voltage. If both the gate and drain terminals of a pass transistor are at a voltage V_{DD}, then the source terminal is pulled no higher than $V_{DD} - V_{TN}$. Succeeding circuits still reliably interpret this voltage as a high logic level. However, if the output (source terminal) of a pass transistor is connected to the gate of a second pass transistor, as in the series pass transistor cascade of Fig. 7.7-4, the output of the second pass transistor can only be pulled to a voltage of $V_{DD} - 2V_{TN}$. In general, for N pass transistors cascaded source to gate, the voltage at the last source terminal V_o can only be pulled to

$$V_o < V_{DD} - NV_{TN} \tag{7.7-3}$$

This voltage is too low to be recognized reliably as a high logic level if $N \geq 2$. For this reason, source-to-gate cascades of pass transistors represent an electrical rule violation and are avoided in practice.

7.7.2 CMOS Transmission Gate

The NMOS pass transistor described previously is an ideal element for performing many logic and control functions and is widely used in NMOS designs. However, is the pass transistor a useful device within CMOS designs? To help answer this question, Fig. 7.7-5 shows two CMOS inverters joined using a typical

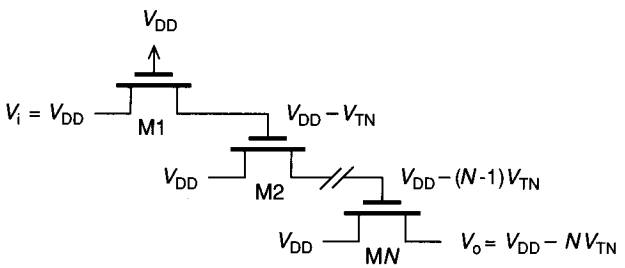


FIGURE 7.7-4
Improper cascaded connection of pass transistors.

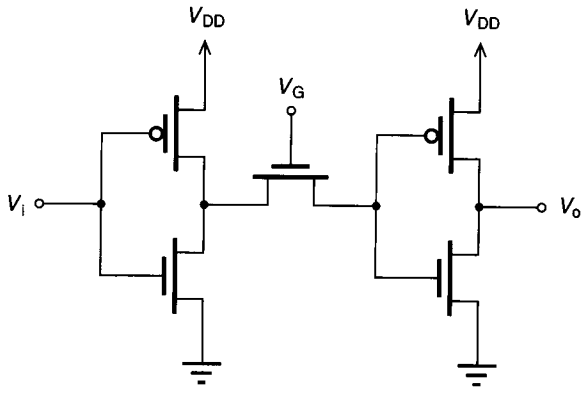


FIGURE 7.7-5
Pass transistor connecting two CMOS inverters.

n-channel pass transistor. Certainly, if the pass transistor gate voltage $V_G = 0$ V, the pass transistor isolates the two CMOS inverters, just as in the case of NMOS inverters. If the pass transistor gate voltage $V_G = V_{DD}$ and the output of the first inverter is at 0 V, the pass transistor will pull the input of the second inverter to 0 V. Once again, this is similar to the NMOS case and is satisfactory. One other condition must be considered. If both the output of the first inverter, that is, the drain terminal of the pass transistor, and the pass transistor gate are at V_{DD}, then the source terminal of the pass transistor can be pulled no higher than $V_{DD} - V_{TN}$. This voltage is sufficient to turn on the n-channel pulldown transistor of the second inverter, but it may not completely turn off the corresponding p-channel pullup transistor.

A gate terminal voltage greater than $V_{DD} - |V_{TP}|$ is required to turn off a p-channel pullup transistor. When the gate voltage drops below $V_{DD} - |V_{TP}|$, the p-channel transistor begins to conduct. If the p-channel pullup transistor conducts because of mismatched V_{TN} and V_{TP} threshold voltages, for example, static power is dissipated in the gate, and the effective noise margin is reduced. Thus, the standard n-channel pass transistor configuration is undesirable for driving a CMOS gate.

The pass transistor is such a useful circuit element in digital designs that a CMOS equivalent is used to accomplish a similar effect. The CMOS circuit is not called a pass transistor but, rather, is called a *transmission gate* because more than a single transistor is required. Figure 7.7-6 shows a CMOS transmission

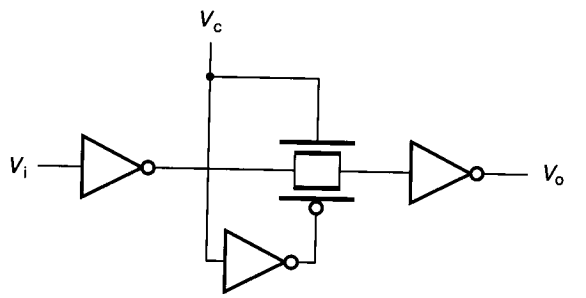


FIGURE 7.7-6
Transmission gate connecting two CMOS inverters.

gate connecting two inverters. The single n-channel pass transistor is replaced by parallel n-channel and p-channel transistors driven by opposing logic levels. Unfortunately, unless the control signal V_c for the transmission gate is available as a double-rail logic signal, an inverter is required in addition to the parallel n-channel and p-channel transistors.

Operation of a CMOS transmission gate circuit can be explained as follows. If the control signal to the transmission gate (shown in Fig. 7.7-6) is $V_c = 0$ V, the gate terminal of the n-channel transistor is also at 0 V and the gate terminal of the p-channel transistor is at V_{DD} because of the control signal inverter; thus, both transistors are off. If the control signal is $V_c = V_{DD}$, then appropriate voltages are generated to cause both the n-channel and the p-channel transistors of the transmission gate to conduct. Under this condition the two parallel transistors of the transmission gate function in a complementary manner to connect the first and second inverters of the signal path in Fig. 7.7-6. When the output of the first inverter is at 0 V, the n-channel transistor of the transmission gate pulls the input to the second inverter to 0 V. When the output of the first inverter is at V_{DD}, the p-channel transistor of the transmission gate pulls the input to the second inverter all the way to V_{DD}. This circumvents the threshold voltage drop problem encountered with a single enhancement pass transistor for CMOS logic. Thus, the CMOS transmission gate provides voltage levels that are compatible with other CMOS logic gates. Unfortunately, the cost of extra silicon layout area for transistors and interconnection is high, rendering the CMOS transmission gate a less useful device than the comparable NMOS pass transistor.

Both pass transistors and transmission gates were described in this section. These building blocks are most advantageous where they can select or control the flow of a logic signal. Care must be exercised in cascading these devices to prevent large signal propagation delays.

7.8 SIGNAL PROPAGATION DELAYS

Earlier sections of this chapter dealt with steady state analysis of MOS circuits that perform digital logic functions. Now an additional parameter—time—is injected into the analysis. Real logic circuits require nonzero but finite time for signals to propagate from input to output. Models to analyze and predict the propagation delay are crucial for practical logic design.

Delays encountered in digital circuitry are composed of two principle components: *gate delay* and *interconnection delay*. Logic gate delay, the time required for a signal to propagate from the input of a logic gate to the output of the same gate, is an important parameter in determining the capabilities of a logic family such as TTL or NMOS logic. Historically, logic gate delay has been the major limiting factor in setting clock rates, and therefore the computational speed, of single-chip digital integrated circuits. Digital systems, comprising many integrated circuit chips, require analysis of interconnection delay in addition to the logic gate delay of the circuits themselves. Digital system interconnection delays are those arising from integrated circuit package connections, printed circuit board connections, and chassis back-plane connections. As integrated circuits

have been manufactured with reduced device sizes, internal gate-to-gate interconnection delays have increased in importance relative to logic gate delays. As device sizes reach submicron dimensions, internal interconnection delays dominate the gate delays. The characterization of signal propagation delays for logic gates loaded by a single, identical logic gate with minimal interconnection is addressed in this section. In a subsequent section delays caused by nonhomogeneous logic gate characteristics, logic fanout, interconnection capacitance, and off-chip loads are analyzed.

7.8.1 Ratio-Logic Model

To address the problem of signal propagation delay, the simple case of Fig. 7.8-1*a*, where the output of one inverter drives the input to a second, identical inverter, will be considered initially. For this analysis, depletion-load inverters such as the one analyzed in Fig. 7.3-1*a*, and redrawn in Fig. 7.8-1*b*, will be used. Figure 7.8-2*a* shows an ideal voltage step of amplitude V_{DD}. Assume that this voltage step is applied to the input of the first inverter of Fig. 7.8-1*a* at $t = t_0$. For the output of the first inverter, the response of Fig. 7.8-2*b* is ideal, but the actual response will resemble that of Fig. 7.8-2*c*. In this case, the signal propagation delay is the difference between the time of the input transition, t_0, and the time that the output is recognized as a valid logic low voltage, t_1. This delay is caused by parasitic capacitances in the MOSFETs, as discussed in Sec. 3.1, and interconnection capacitance. The slight initial overshoot in the actual response is caused by the gate-drain overlap capacitance of M1. The overall delay time, however, is dominated by the effects of the capacitive load caused by the second inverter including the interconnection capacitance. A closed-form mathematical expression for the output waveform of Fig. 7.8-2*c* is unwieldy because of the

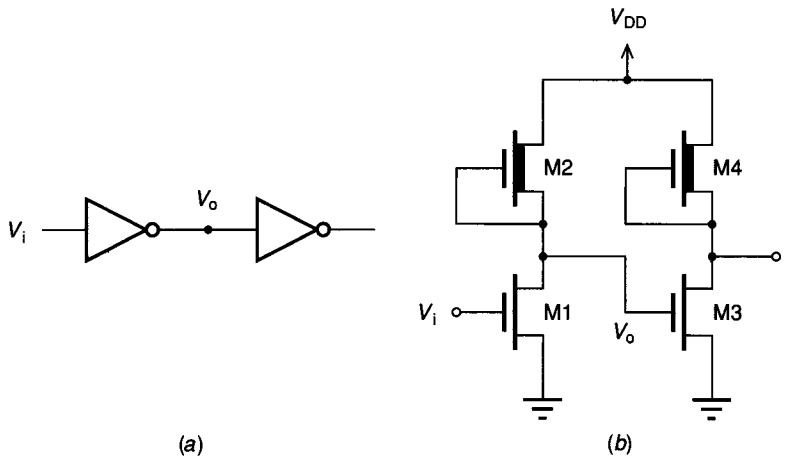


FIGURE 7.8-1
Single inverter driving a second, identical inverter: (*a*) Logic diagram, (*b*) Circuit diagram.

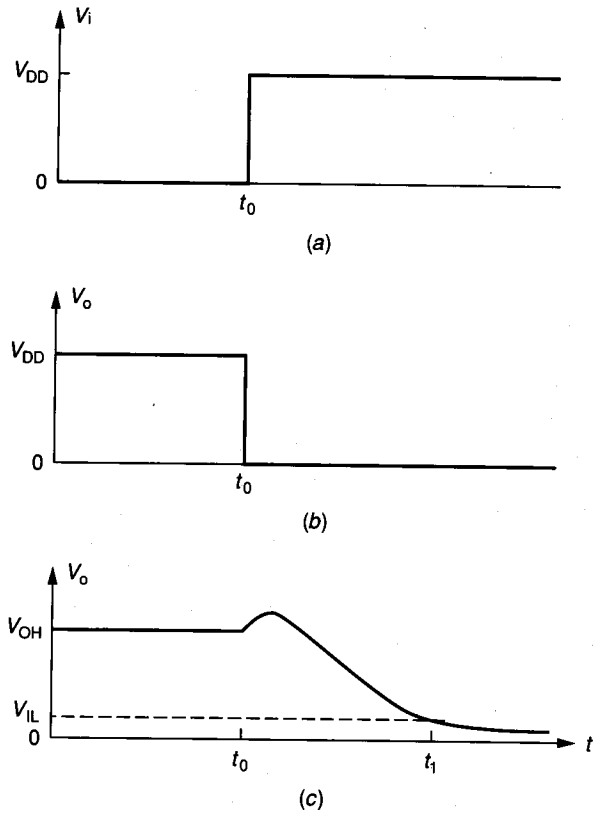


FIGURE 7.8-2
Voltage step input and inverter response: (*a*) Ideal voltage step, (*b*) Ideal response, (*c*) Typical response.

nonlinear nature of the device *I–V* characteristics, the voltage dependence of all parasitic capacitors, and the changes in operating region for the MOSFETs during the transition. A closed-form solution requires the simultaneous solution of a set of nonlinear partial differential equations.

For hand analysis and insight into design, a quick and simple method of approximating the response of this circuit is needed for both high-to-low and low-to-high output transitions. This initial analysis neglects interconnection capacitance and is based on the simple switch resistor model of Fig. 7.5-4, with the addition of gate capacitance. Figure 7.8-3*a* shows a symbolic convention that emphasizes the resistive models for transistors M1 and M2. This can be partitioned further to the resistance-capacitance inverter model of Fig. 7.8-3*b*. Note that M2 is modeled by a resistor without a switch since the depletion transistor with $V_{GS} = 0$ V is always on. M1 is modeled by C_G, S1, and R_1. Switch S1 is open for V_i low and closed for V_i high. The resistance R_2 models the pullup transistor resistance, and the resistance R_1 models the pulldown transistor resistance. The capacitance C_G is the input capacitance to the inverter. It will be demonstrated later that C_G is approximately equal to $C_{ox}WL$ where W and L are the width and length of the pulldown transistor (M1 or M3 in Fig. 7.8-1*b*). For this analysis, the equivalent input capacitance to the reference inverter in a logic family is termed C_G. From

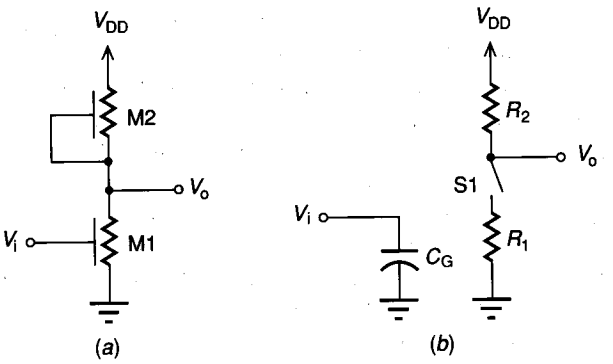


FIGURE 7.8-3
Inverter models for delay calculations: (*a*) Symbolic representation, (*b*) Resistance-capacitance inverter model.

a dc viewpoint, it follows from the model of Fig. 7.8-3*b* that the high output voltage level is V_{DD} and the low output voltage level is

$$V_L = \frac{V_{DD}R_1}{R_1 + R_2} \tag{7.8-1}$$

Logic circuits where the dc output voltage is determined by the ratio of two series resistors are termed *ratio logic* circuits. As was stated in Eqs. 7.5-2 and 7.5-3, the equivalent resistance of a MOS transistor in the model of Fig. 7.8-3*b* is proportional to its length L divided by its width W. Thus, the relationship between resistance R_2 and resistance R_1 can be approximated as

$$R_2 = \frac{L_2W_1}{W_2L_1}R_1 = kR_1 \tag{7.8-2}$$

With the 4 : 1 sizing rule of Sec. 7.3, $k = 4$ and $R_2 = 4R_1$. With this sizing rule and the model of Fig. 7.8-3*b*, it follows from Eq. 7.8-1 that the high and low logic levels for the output are V_{DD} and $V_{DD}/5$, respectively. For $V_{DD} = 5$ V, this gives a high logic level of 5 V and a low logic level of 1 V. Although the high logic level is the same as that obtained by the more exact analysis of Sec. 7.3, the low logic level differs from the more exact value of 9/32 V obtained previously with $V_{DD} = 5$ V. This difference in dc voltage levels is not significant for an approximate delay analysis.

To provide a simplified model for delay calculations, approximate values for resistances R_1 and R_2 of Fig. 7.8-3*b* must be found. Observe that when V_i is high, S1 is closed and M1 is in the ohmic region for much of the high-to-low output transition. From Eq. 3.1-8 with $V_{GS} = V_i = V_{DD}$, a good approximation for the resistance R_1 near $V_o = V_{DS} = 0$ V is the small signal equivalent resistance

$$R_{ss} = \frac{L_1}{K'W_1(V_{DD} - V_{TN})} \tag{7.8-3}$$

(Note that Eq. 7.8-3 demonstrates the proportionality factor described for Eqs. 7.5-2 and 7.5-3.) This resistance is a good reference point but does not consider the effects of the pullup resistance or the nonlinear characteristics of the enhancement transistor over the actual output voltage range.

A better model for delay estimation is found by considering the large signal equivalent resistance. The equivalent resistance at the start of the output transition is found by dividing the initial voltage across M1 by the initial current through M1. The pullup transistor M2 contributes no current at this point because it has $V_{DS} = 0$ V. Assuming $V_i = V_{DD}$, the equivalent resistance is given by

$$R_1 = \frac{2V_{DD}L_1}{K'W_1(V_{DD} - V_{TN})^2} \approx 2R_{ss} \tag{7.8-4}$$

using the approximation that $V_{DD} - V_{TN} \approx V_{DD}$. As the output voltage falls, the resistance of the enhancement transistor approaches R_{ss}, but the depletion pullup transistor begins to supply current, thereby increasing the effective resistance of the inverter output. Analysis of the large signal equivalent resistance at an output voltage near V_{TN} with the effects of pullup transistor M2 considered gives

$$R_1 = \frac{V_o L_1}{K'W_1\left[(V_{DD} - V_{TN} - V_o/2)V_o - V_{TD}^2/2k\right]} \tag{7.8-5}$$

It can be shown that this expression also reduces to $R_1 \approx 2R_{ss}$ with typical parameters for the reference inverter. For the approximate propagation delay analysis in this section, it will be assumed that R_1 is given by Eq. 7.8-4.

For a low-to-high output transition, transistor M2 starts in the saturated region and finishes in the ohmic region. With typical parameter values for the 4:1 reference inverter, the large signal equivalent resistance of M2 for a low-to-high output transition varies from about $13R_{ss}$ to $5R_{ss}$ as V_o increases from 0 V to V_{DD}. An average equivalent resistance of $8R_{ss}$ is a reasonable compromise. Note that considering the 4:1 resistance ratio for the reference inverter, the equivalent pullup resistance is

$$R_2 \approx 4R_1 = 2kR_{ss} \tag{7.8-6}$$

It should be noted that for this approximate analysis, significant errors of $\pm 50\%$ or more can be anticipated using the simple model of Fig. 7.8-3*b*. These errors are justifiable for a quick approximate analysis. The alternative when better accuracy is required is to use circuit simulation programs such as SPICE to account for the nonlinear characteristics of MOS transistors and the effect of nonzero rise time for the input voltage. If $L_1 = W_1$, $L_2 = 4W_2$, $K' = 30\ \mu A/V^2$, $V_{TN} = 1$ V, $V_{TD} = -3.5$ V, and $V_{DD} = 5$ V, then

$$R_1 \approx 16.6\ k\Omega \quad \text{and} \quad R_2 \approx 66.4\ k\Omega \tag{7.8-7}$$

With approximate models for the transistor resistances in hand, the analysis now turns to the input capacitance. Because M1 of Fig. 7.8-1*b* is ohmic for $V_i = V_{DD}$ and cutoff for $V_i = 0$ V, it follows from Fig. 3.1-19*b* that the parasitic input capacitance at either logic level is approximately $C_{ox}W_1L_1$. During transitions, M1 enters the saturation region momentarily, causing the input capacitance to drop to approximately $(2/3)C_{ox}W_1L_1$, as indicated in Fig. 3.1-19*b*. Because the capacitance change is relatively small and M1 is normally in the saturation region for only part of the output transition time, the change is neglected in the model of Fig. 7.8-3*b*. Thus, C_G is approximated as

$$C_G \approx C_{ox}WL \tag{7.8-8}$$

To estimate propagation delays, the equivalent circuit model of Fig. 7.8-3*b* is simplified to the model of Fig. 7.8-4*a* by adding switch $\overline{S1}$. This model isolates the effects of R_1 and R_2 to simplify the analysis further. Remember that the effects of the pullup transistor during a high-to-low transition were included when R_1 was approximated earlier.

The high-to-low and low-to-high transition times for an inverter loaded by an identical inverter will now be determined. Because the input to the load inverter looks like a capacitor of value C_G, the equivalent circuit for a low-to-high step input voltage is as shown in Fig. 7.8-4*b*. Note that this input causes a high-to-low output transition. Because this model is a simple *RC* circuit with an initial voltage on C_G, it follows that a major portion of a high-to-low output transition, designated t_{HL}, occurs in two *RC* time constants, where $R = R_1$ and $C = C_G$ (for an ideal *RC* circuit, the 90% to 10% transition requires 2.2 time constants). Thus, the high-to-low output transition is approximated here by

$$t_{HL} \approx 2R_1C_G \tag{7.8-9}$$

Through a similar analysis of the circuit in Fig. 7.8-4*c*, the low-to-high output transition time is approximated by

$$t_{LH} \approx 2R_2C_G \tag{7.8-10}$$

For ratio-logic circuits, R_1 and R_2 of Fig. 7.8-4*a* are related by the device-sizing parameter, k. From Eqs. 7.8-2, 7.8-9, and 7.8-10, it follows that

$$t_{LH} = kt_{HL} \tag{7.8-11}$$

For the 4:1 sizing rule,

$$t_{LH} = 4t_{HL} \tag{7.8-12}$$

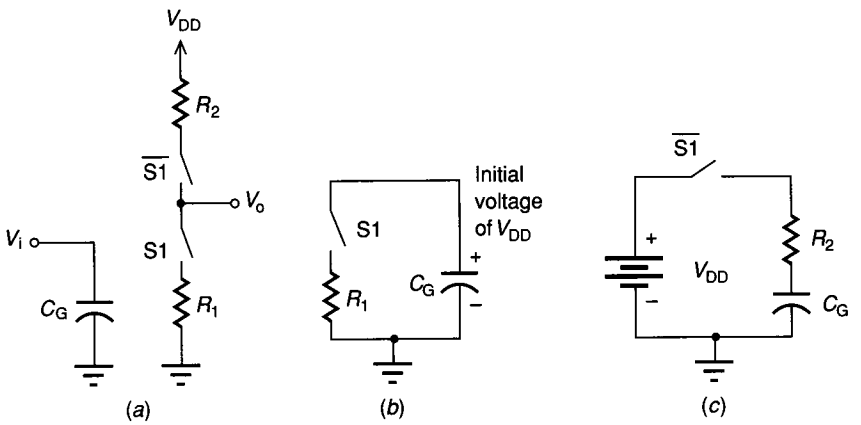


FIGURE 7.8-4
Equivalent circuits for inverter delay analysis: (*a*) Simplified *RC* inverter model, (*b*) Equivalent circuit for high-to-low output transition, (*c*) Equivalent circuit for low-to-high output transition.

7.8.2 Process Characteristic Time Constant

Each process will have geometrical design rules that limit the minimum size for a transistor. The process will also have electrical design rules that specify the desired supply voltage V_{DD}, the transconductance parameter K', the threshold voltage V_T, and the gate capacitance per unit area C_{ox}. From these parameters, a characteristic time constant for the process can be determined. This time constant, designated as τ_P, and defined as $R_{ss}C_G$ is useful for comparing delay characteristics of different processes. For both NMOS and CMOS technologies the values of K' and V_T for the n-channel enchancement transistor are used; and for PMOS technologies the values of K' and V_T for the p-channel enchancement transistor are used. From Eqs. 7.8-3 and 7.8-8,

$$\tau_P = R_{ss}C_G = \frac{L_1}{K'W_1(V_{DD} - V_{TN})} C_{ox}W_1L_1 \tag{7.8-13}$$

This reduces to

$$\tau_P = \frac{L_1^2 C_{ox}}{K'(V_{DD} - V_{TN})} \tag{7.8-14}$$

Assume the typical minimum transistor dimension is 2 μ, $K' = 45$ $\mu A/V^2$, $C_{ox} = 1$ fF/μ^2, $V_{TN} = 1$ V, and $V_{DD} = 5$ V. Using these values in Eq. 7.8-14, the process characteristic time constant is

$$\tau_P = 0.02 \text{ ns} \tag{7.8-15}$$

It must be noted that τ_P is not a measure of expected circuit delay. The value of τ_P depends only on process geometrical and electrical parameters and is thus independent of a particular circuit implementation. Therein lies the usefulness of τ_P. From Eqs. 7.8-4, 7.8-6, 7.8-9, and 7.8-10, it can be observed that t_{HL} and t_{LH} can be expressed in terms of τ_P as $t_{HL} = 4\tau_P$ and $t_{LH} = 16\tau_P$ for a minimum-size inverter.

7.8.3 Inverter-Pair Delay

As was the case in the analysis of logic levels in Sec. 7.2, the inverter pair plays an important part in logic gate delay analysis. Because the high-to-low and low-to-high transition times are asymmetric, neither transition time is adequate to characterize a logic family. If the signal propagation delay from V_i to V_o for a pair of identical cascaded inverters shown in Fig. 7.8-5 is considered, then both

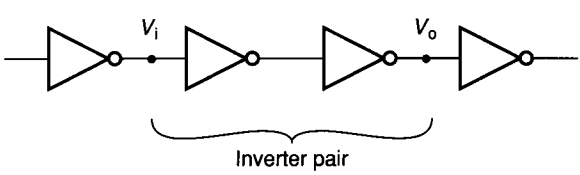


FIGURE 7.8-5
Cascade of identical inverters.

high-to-low and low-to-high transitions contribute to the delay. Thus, the inverter-pair delay, designated t_{ipd}, is defined as the sum of a high-to-low transition and a low-to-high transition time:

$$t_{ipd} = t_{HL} + t_{LH} \tag{7.8-16}$$

This is often expressed in terms of the device sizing ratio as

$$t_{ipd} = (1 + k)t_{HL} \tag{7.8-17}$$

or from Eqs. 7.8-4, 7.8-9, and 7.8-13 in terms of τ_P:

$$t_{ipd} = 4(1 + k)\tau_P \tag{7.8-18}$$

The inverter-pair delay is a key parameter in characterizing the speed of operation of a digital logic family. It represents a fundamental lower bound on the clock period of a synchronous system because a signal must be able to experience at least one high-to-low and one low-to-high transition during a system clock cycle. It will be seen later that the system clock period is typically much longer than the inverter-pair delay because of the contribution of interconnection delays and multiple levels of logic.

A physical interpretation of the inverter-pair delay deserves consideration. If either a high-to-low or a low-to-high input transition is applied to a cascade of inverters with identical loads, as in Fig. 7.8-5, then the delay associated with propagation of the signal through any two inverters (i.e., an inverter pair) is the inverter-pair delay. Assume the inverter cascade of Fig. 7.8-5 is designed in the typical NMOS process used for Eq. 7.8-15. Further assume that the gate of the pulldown device for each inverter is the minimum size of 2 $\mu \times 2$ μ, it follows from Eqs. 7.8-15 and 7.8-18 that the inverter-pair delay is

$$t_{ipd} = 4(1 + 4)0.02 \text{ ns} = 0.4 \text{ ns} \tag{7.8-19}$$

Note that based on the simplified device model used in these calculations, the response time for the reference inverter pair is very fast.

Three important observations must be made at this point. First, the present analysis gives unrealistically small delay times because it neglects interconnection capacitance. Second, the low-to-high transition delay for ratio logic is about k times as long as the high-to-low transition delay, as indicated by Eq. 7.8-11. This introduces significant asymmetry in the two transitions and is of concern in many applications, particularly if an output is loaded by a large capacitance. A typical response of the reference inverter to a fast square-wave excitation in a ratio-logic circuit is given in Fig. 7.8-6. This figure clearly shows the asymmetric rise and fall times for a ratio inverter circuit. The asymmetry occurs because the active pulldown device, M1, has much lower resistance than the passive pullup device, M2. One might be tempted to improve the pullup characteristics of M2 to decrease t_{LH} by resizing M2 for lower equivalent resistance. Unfortunately, this solution is not feasible because it would cause a degradation in logic signal levels, as can be seen from Eq. 7.8-1. The asymmetry in transition times is thus inherent in ratio-logic systems, as indicated by Eq. 7.8-11.

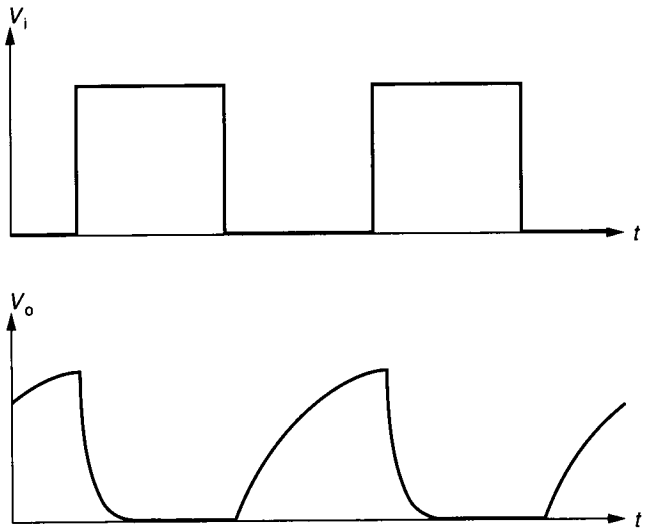


FIGURE 7.8-6
Asymmetric response for a ratio-logic inverter driven with a square wave.

A comparison of results obtained from the approximate delay analysis with those obtained from a detailed circuit simulation shows some differences.[7] These are attributable primarily to the terms R_1 and R_2 in Eqs. 7.8-9 and 7.8-10. These differences are due, in part, to the fact that the pullup and pulldown devices are not linear throughout their entire output transitions. Also, the previous analysis estimated a stage delay based on an ideal voltage step input. For a cascade of MOS inverters, the ideal step input applies only to the first stage; subsequent stages typically have an input rise time almost as long as the delay of the preceding stage. This partially invalidates the earlier assumption that the inverter input voltage was at V_{DD} during the entire high-to-low output transition. Thus, if the delay for a cascade of inverters is computed as the sum of single inverter-stage responses with ideal step inputs, further error is introduced.

Two good ways exist to improve the accuracy of a simplified inverter-delay analysis without increasing its complexity. First, the inverter-pair delay t_{ipd} (or even t_{HL} and t_{LH}) for a reference inverter can be experimentally measured for a specific process or determined from an accurate computer simulation. These delay parameters can then be thought of as design parameters. Overall delays can be expressed directly in terms of these delay parameters of the reference inverter. Alternatively, the resistance values R_1 in Eq. 7.8-4 and R_2 in Eq. 7.8-6 can be modified to compensate for errors in t_{HL} and t_{LH}. It should be remembered that the approximate analysis is useful primarily as a quick estimate of circuit performance. Attempts to create a simple, accurate delay analysis are undermined by many factors in practical circuits.[8]

The third observation is that R_1 and R_2 from Eqs. 7.8-4 and 7.8-6 are dependent on device length-to-width ratios but not on the actual lengths and widths of a transistor. As device sizes shrink—for example, from 2 μ to 1 μ—the

transistor resistances remain relatively constant. The gate area of M3, the input device of the second stage in Fig. 7.8-1*b*, provides the parasitic capacitance load to the first stage. In a linearly scaled process, the gate area of M3 decreases with the square of the feature size, while the gate oxide thickness decreases directly with device size (see constant field scaling, Sec. 1.1). For these approximations, the total capacitive loading decreases linearly with feature size. It thus follows that the inverter-pair delay, which depends on effective transistor resistance and parasitic capacitive loading, decreases approximately linearly with feature size. This decrease offers the potential for significant system speed improvements as devices become smaller. The process characteristic time constant τ_P for the scaled-down process shows this speed increase.

The previous analysis was for a single inverter loaded by a single, identical inverter stage. If an input signal transition must ripple through a cascade of identical logic circuits, a first-order approximation of the total delay for the cascade assumes that the start of the transitions on successive stages is delayed until the transition of the preceding stage is complete. This allows the total delay to be simply computed as the sum of the individual stage delays. If the cascade delay is denoted as t_{cas}, the number of identical stages is N, and the average stage delay is one-half the inverter-pair delay t_{ipd}, then an approximation to the total delay is

$$t_{cas} = \frac{Nt_{ipd}}{2} \tag{7.8-20}$$

It will be shown in a subsequent section that capacitive loading from interconnections and fan-out causes delays larger than that predicted by Eq. 7.8-20.

7.8.4 Superbuffers

The asymmetric output delay of a ratio logic circuit is particularly undesirable when a highly capacitive bus or a large number of secondary device inputs must be driven. The slow pullup capability can seriously limit clock speeds for a system. One partial solution to this problem is a special circuit configuration called a *superbuffer*. Figures 7.8-7 and 7.8-8 show circuits for a noninverting super-

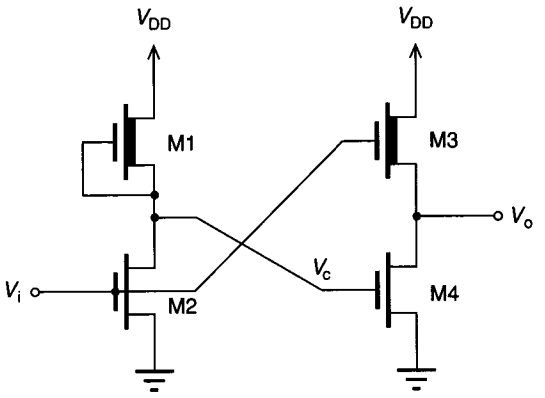


FIGURE 7.8-7
Noninverting superbuffer circuit.

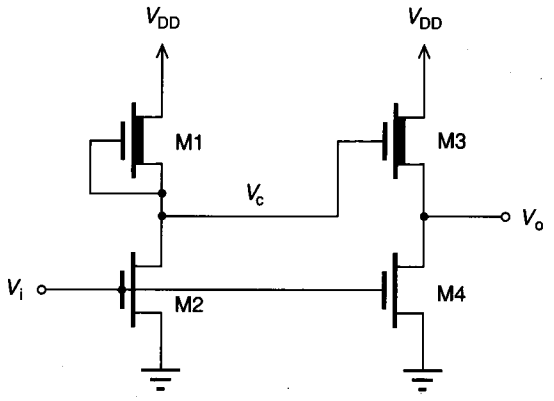


FIGURE 7.8-8
Inverting superbuffer circuit.

buffer and an inverting superbuffer, respectively. The output stage of each of these circuits is modified from the standard ratio-logic inverter in that the pullup transistor has its gate connected to an active logic signal. This connection increases the current sourcing ability of the pullup device, allowing faster drive for capacitive loads.

As can be seen from Figs. 7.8-7 and 7.8-8, a superbuffer consists of a standard inverter stage connected to a second inverter stage with an active pullup. The first inverter stage provides the input signal and its logical complement to drive the transistors of the output inverter stage. Both the input signal and its complement are necessary to drive the pulldown transistor and pullup transistor of the output stage at complementary times.

The logical operation of the noninverting superbuffer of Fig. 7.8-7 will now be examined. Assume that the output stage with transistors M3 and M4 has the 4:1 pullup/pulldown ratio that was used for the reference inverter. In this case, both transistors may remain unchanged sizewise from the reference inverter circuit. If the input to the noninverting superbuffer is low, then the output should also be low. Because the M1-M2 circuit is an inverter, the intermediate voltage V_c will be high when the superbuffer input is a logic low. Because this intermediate voltage drives the gate of pulldown transistor M4 to a logic high voltage, the output of the superbuffer is pulled low. The gate of the pullup transistor M3 is tied directly to the superbuffer input; thus, its gate is connected to a low voltage. If the superbuffer output is low, pullup transistor M3 has its gate and source each connected to a low voltage. This provides a gate-to-source voltage of 0 V, and the circuit is therefore equivalent to the standard depletion-load inverter when the superbuffer input is low. Based on this analysis, the superbuffer output stage functions as a standard depletion-load inverter stage when the superbuffer output is low.

The logical operation of the noninverting superbuffer will now be examined in response to a high input. This high input at the first inverter stage will drive its output, V_c, to a low level. The signal V_c is connected to the gate of pulldown transistor M4, causing it to turn off. The pullup transistor M3 has its gate connected directly to the superbuffer input, which is high. Initially, before

the superbuffer output is pulled high, the source terminal of the output pullup transistor, M3, will be at a low voltage. This provides a gate-to-source voltage for pullup transistor M3 that is approximately equal to a logic high voltage minus a logic low voltage, or nearly V_{DD}. For a typical NMOS process for digital circuits, it is easily shown that a gate-to-source voltage of 0 V for a depletion-mode transistor allows a current that is nearly equal to the current for a similarly sized enhancement transistor with a gate-to-source voltage of $V_{DD} - V_{TN}$. Increasing the gate-to-source voltage to V_{DD} for a depletion-mode transistor is roughly equivalent to doubling the gate-to-source voltage for an enhancement transistor. This provides improved pullup capability during the low-to-high transition of the superbuffer output.

The actively driven pullup transistor M3 improves the pullup characteristic of the superbuffer as just explained; the following analysis quantifies this improvement. For integrated digital circuitry with depletion-mode pullup transistors, the magnitude of V_{TD} is typically chosen as a large percentage of the supply voltage—for example, $-V_{TD} \approx 0.7V_{DD}$. For a standard depletion-mode pullup transistor with a gate-to-source voltage of 0 V, the transistor is in the ohmic region whenever the drain-to-source voltage is less than $0.7V_{DD}$. If the gate-to-source voltage is greater than $0.3V_{DD}$, the depletion-mode transistor is always in the ohmic region because the drain-to-source voltage is limited to V_{DD}. For a transistor in the ohmic region with constant drain-to-source voltage, its equivalent resistance is inversely proportional to the effective gate-to-source voltage, $V_{GS} - V_{TN}$ (see Eq. 3.1-8). If the effective gate-to-source voltage is doubled, the effective resistance of the pullup transistor is halved.

In the analysis of the preceding paragraph, it was shown that the equivalent resistance of the active pullup transistor of a superbuffer is initially half that of a standard inverter pullup for a low-to-high output transition. From Eq. 7.8-10, the delay t_{LH} is directly proportional to the effective resistance of the pullup transistor. Thus, the new low-to-high delay becomes

$$t'_{LH} = \frac{1}{2}t_{LH} = 2t_{HL} \tag{7.8-21}$$

The equivalent inverter-pair delay for a superbuffer output stage with the standard pullup/pulldown ratio becomes

$$t'_{ipd} = t_{HL} + t'_{LH} = 3t_{HL} \tag{7.8-22}$$

This means that a superbuffer output stage can drive a heavy capacitive load almost twice as fast as a standard inverter of the same size ($3t_{HL}$ versus $5t_{HL}$). A similar analysis applies to the inverting superbuffer of Fig. 7.8-8.

7.8.5 NMOS NAND and NOR Delays

Many times, a designer has a choice between using a NAND circuit or a NOR circuit to realize a particular logic function. For example, any combinational logic function can be realized in a sum-of-products form with a NAND-NAND logic gate combination or as a product-of-sums form with a NOR-NOR logic gate

structure.[6] Figure 7.8-9 shows the implementation of a simple logic function, the exclusive-OR, with both a NAND-NAND and a NOR-NOR realization. Because either structure can be used, it is useful to ask if one structure has advantages over the other.

In Sec. 7.4, it was determined that the typical pullup/pulldown sizing ratio for a NOR gate was about 4:1 and that a typical pullup/pulldown ratio for a two-input NAND gate was about 8:1. Assuming the input devices for the NAND and the NOR gates are sized the same as for the reference inverter, it follows from the models of Fig. 7.8-4 that the reduced circuits of Fig. 7.8-10*a* and *b* are useful to estimate the logic gate delays for two-input NAND and two-input NOR gates. The delay here refers to the worst-case delay for the NOR gate (when only one of the pulldown transistors switches on). Note that if two or more NOR gate pulldown transistors switch on simultaneously, the effective resistance, and therefore the delay, will be reduced. The reduced models of Fig. 7.8-10 show only a single pulldown path for each NOR gate and lump the series pulldown transistors of a NAND gate into a single, equivalent resistor. The circuit of Fig. 7.8-10*c* is used to compare the delay for a multi-input NOR gate with the delay for the two-input gates. The parameter values for all three cases are listed in Fig. 7.8-10*d*.

Based on the analysis methods of this section, it is easy to see that the delay for the propagation of a digital signal through two cascaded NOR gates is the same as for an inverter pair, provided that both NOR gates are loaded by a single, equivalent input. This NOR-pair delay is just t_{ipd}. The delay for two cascaded NAND gates is greater because of the high-resistance pullup device required by the 8 : 1 pullup/pulldown ratio and because of the two series transistors in the pulldown path. Because the resistances for both the pullup path and pulldown path are doubled, the NAND-pair delay is $2t_{ipd}$. Because the NAND-pair delay is double the NOR-pair delay, it is obvious that the NOR configuration is preferable to the NAND configuration from a delay viewpoint. It is further noted from Fig. 7.8-10*c* and *d* that the delay for a multi-input NOR gate is the same as that for the two-input NOR gate. Multi-input NMOS NOR gates are widely used; NMOS NAND gates with more than two inputs are rare.

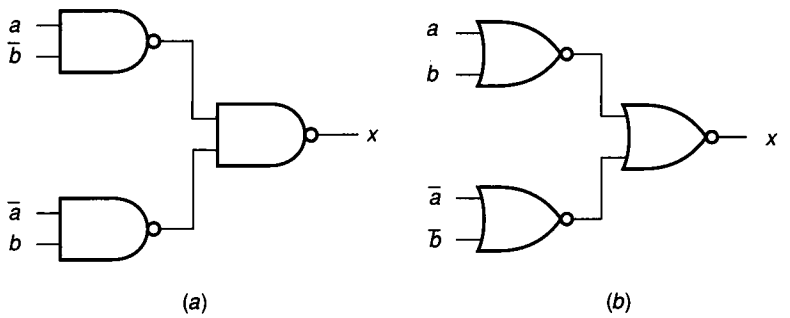
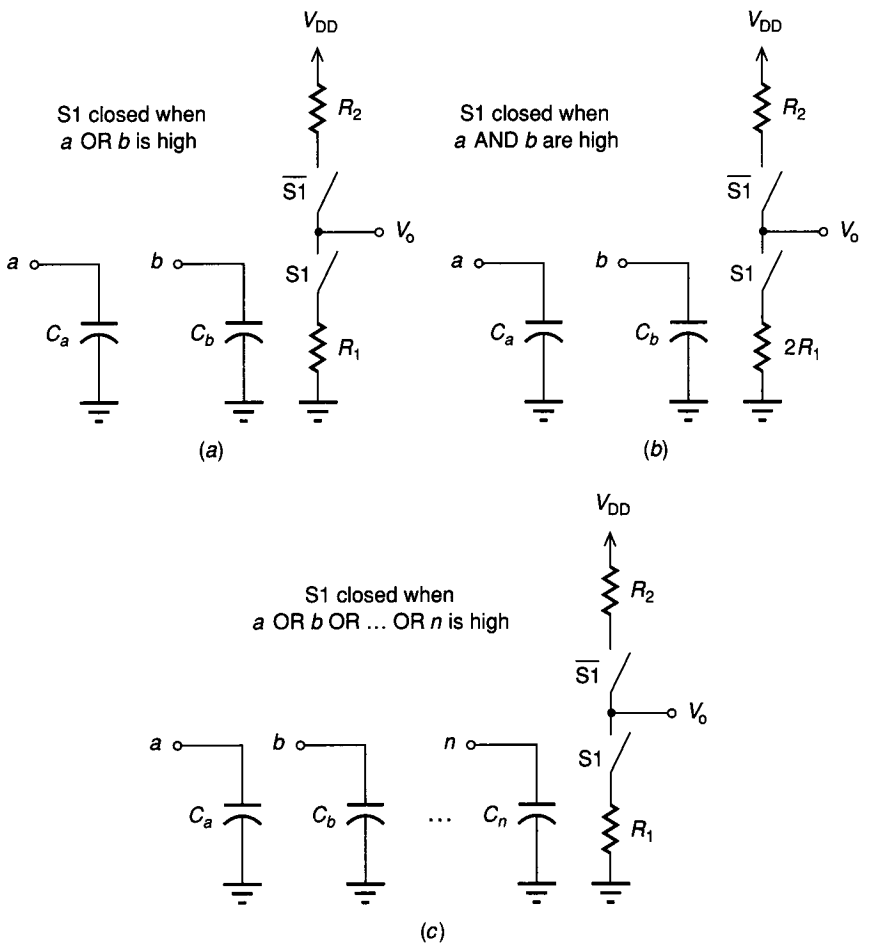


FIGURE 7.8-9
Equivalent circuits to obtain the exclusive-OR function: (*a*) NAND-NAND circuit, (*b*) NOR-NOR circuit.



	NOR	NAND	Multi-input NOR
R_1	16.7 k Ω	16.7 k Ω	16.7 k Ω
R_2	$4R_1 = 66.7$ k Ω	$8R_1 = 133.4$ k Ω	$4R_1 = 66.7$ k Ω
C_a	$C_{ox}W_a L_a$	$C_{ox}W_a L_a$	$C_{ox}W_a L_a$
C_b	$C_{ox}W_b L_b$	$C_{ox}W_b L_b$	$C_{ox}W_b L_b$
C_n	—	—	$C_{ox}W_n L_n$

(d)

FIGURE 7.8-10
Delay models for multi-input gates: (*a*) Delay model for two-input NOR, (*b*) Delay model for two-input NAND, (*c*) Delay model for multi-input NOR, (*d*) Resistance and capacitance values.

7.8.6 Enhancement versus Depletion Loads

It was shown earlier in this chapter that the depletion load is preferred to the enhancement load from a signal-swing viewpoint. Both the enhancement- and depletion-load logic gates are ratio-logic circuits. It was further stated that the same 4 : 1 sizing rule is generally used for both types of logic gates. However, within a single–power supply logic family, the dynamic performance of the two gates is substantially different.

Figure 7.8-11 shows the *I–V* characteristics for three pullup devices: an enhancement-mode transistor with its gate tied to its drain, a depletion-mode transistor with its gate tied to its source, and a linear resistor. Any one of these devices can be used with an enhancement pulldown transistor to form an inverter. All three devices can be designed to have the same equivalent large signal resistance *(V/I)* at a point *X* by suitable choice of width-to-length ratios and threshold voltages for the transistors. Let the devices be sized so that point *X* corresponds to the voltage and current for the pullup of an inverter stage when its output is low. The three devices are indistinguishable from a steady state viewpoint when used as static load devices operating at point *X*.

Now consider the dynamic operation of an inverter circuit. When the pull-down stage of the inverter is turned off, the pullup device should quickly bring the output voltage to a logic high value. This requires the lowest possible resistance for the pullup during the low-to-high transition. The *I–V* curves of Fig. 7.8-11 are particularly instructive here. The linear resistor provides a current that is directly proportional (Ohm's law) to the voltage across the load resistor, $V_{DD} - V_o$, where V_o is the output voltage of the inverter. The enhancement-mode transistor provides an equivalent resistance that tends to infinity as the output voltage rises to $V_{DD} - V_{TN}$, whereas the equivalent resistance of the depletion-mode transistor tends to a value much smaller than the linear resistor value as V_o nears V_{DD}. In the calculation of t_{LH} in Eq. 7.8-10, an equivalent resistance for the pullup

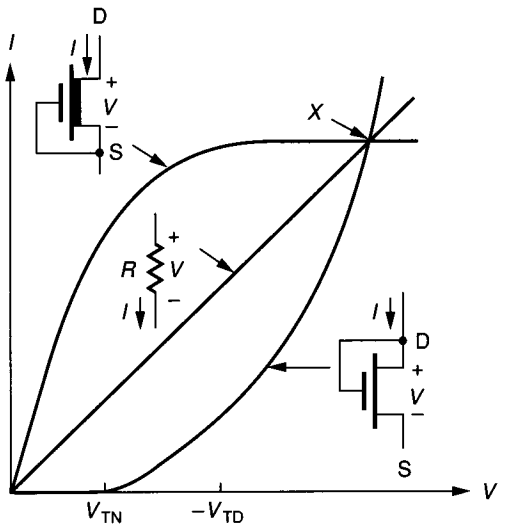


FIGURE 7.8-11
I-V curves for three load devices: depletion pullup, linear resistor, and enhancement pullup.

device determines the delay time. Thus, from the viewpoint of signal rise time, the depletion-mode transistor is a better choice than either the linear resistor or the enhancement-mode transistor. Because the output signal fall time is almost independent of the pullup device type, essentially all modern MOS ratio logic is designed with depletion pullups to reduce total gate delay times.

7.8.7 CMOS Logic Delays

So far, the delay analysis of NMOS ratio logic has been emphasized. As described in Sec 7.5, the basic CMOS inverter is not a ratio-logic device because the output is actively pulled to one logic level or the other depending on the input. Fortunately, the basic concepts of the preceding delay analysis apply equally well to CMOS logic. The unique delay characteristics of CMOS logic as compared to NMOS logic are developed in this section.

The initial delay analysis of the reference CMOS inverter is based on one inverter driving a second, identical inverter, as shown in both logic and circuit diagram form in Fig. 7.8-12. Several observations about this circuit are important here. An obvious difference from the NMOS inverter circuit of Fig. 7.8-1 is that the output of the first CMOS inverter must drive the gates of two transistors: one for the n-channel pulldown transistor and one for the p-channel pullup transistor. Both gates provide capacitive loading that slows the transition of logic signal values. Analysis of the gate capacitance for the n-channel pulldown transistor is identical to that for the NMOS inverter expressed by Eq. 7.8-8. This gate capacitance is denoted by C_{GN} and is given as

$$C_{GN} = C_G \approx C_{ox} W_N L_N \tag{7.8-23}$$

The gate capacitance of the p-channel transistor, denoted by C_{GP}, is similar to that of the gate of the n-channel transistor based on the model of Fig. 3.1-19. When the p-channel transistor is off, that is, the input voltage to the gate is a logic high, the capacitance is $C_{ox} W_P L_P$. In saturation, that is, when the input

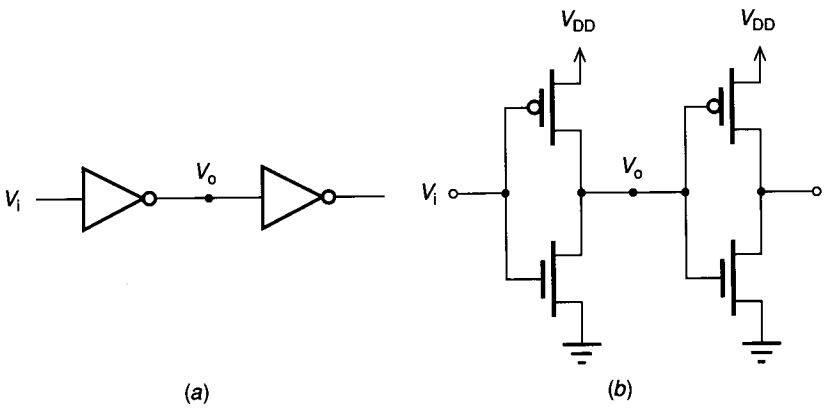


FIGURE 7.8-12
CMOS inverter driving a second, identical CMOS inverter: (*a*) Logic diagram, (*b*) Circuit diagram.

voltage to the gate is a logic low, the p-channel transistor capacitance is again $C_{ox}W_PL_P$. When the p-channel transistor is in the ohmic region, the gate capacitance is about 2/3 of its value in saturation. It is reasonable to assume that the p-channel transistor is in the ohmic region for only a short part of signal transition. For this reason, the gate capacitance of the p-channel transistor is approximated as

$$C_{GP} \approx C_{ox}W_PL_P \tag{7.8-24}$$

Because the capacitances for the two transistors of the second inverter are effectively in parallel (one to ground and the other to V_{DD}), the capacitive load seen by the first inverter is

$$C_{GC} = C_{GN} + C_{GP} \tag{7.8-25}$$

A simplified model for the CMOS inverter circuit of Fig. 7.8-12 is given in Fig. 7.8-13. This model is based on the simple resistive model of Fig. 7.8-3 for the driving transistors and the capacitance values that were just discussed. When a low-to-high step input is applied to the first inverter, the equivalent circuit of Fig. 7.8-13*b* is applicable. When a high-to-low step input is used, the equivalent circuit of Fig. 7.8-13*c* is appropriate. Approximating the 10% to 90% signal rise time by two time constants for the ideal *RC* circuit models of Fig. 7.8-13*b* and *c*, the rising and falling transition delays are given by the equations

$$t_{LH} = 2R_2C_{GC} \tag{7.8-26}$$

and

$$t_{HL} = 2R_1C_{GC} \tag{7.8-27}$$

The value of R_1 is determined using an analysis similar to that used for the NMOS inverter pulldown and is given by Eq. 7.8-4. Since the p-channel pullup

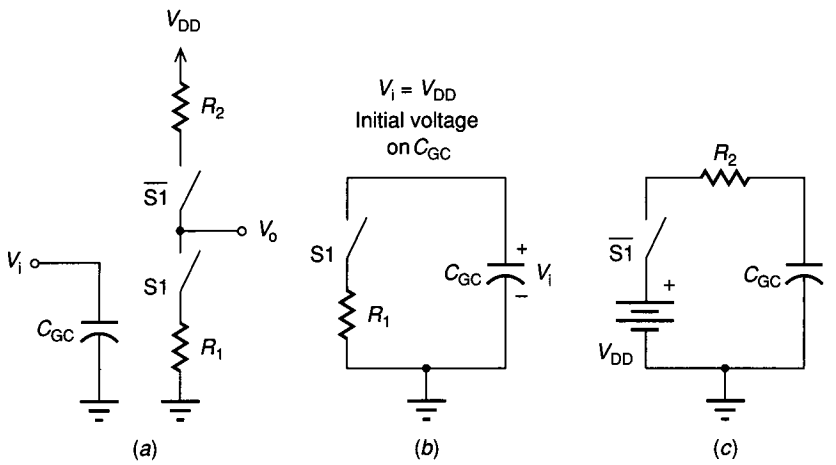


FIGURE 7.8-13
Equivalent circuits for CMOS delay analysis: (*a*) Simplified *RC* inverter model, (*b*) Equivalent circuit for high-to-low output transition, (*c*) Equivalent circuit for low-to-high output transition.

transistor is an enhancement transistor, its equivalent resistance R_2 is also given by Eq. 7.8-4 with the appropriate substitution of p-channel values for K', L, W, and V_T. If the transistors are sized for symmetrical output drive with equivalent pullup and pulldown resistances, the delay is

$$t_{LH} = t_{HL} = 2R_2C_{GC} = 2R_1C_{GC} \tag{7.8-28}$$

With the simplified RC model for a CMOS inverter as the basis, analysis of CMOS NAND and NOR gates is easy. In general, each logic gate output must drive the gate capacitance of two transistors for every connection to another CMOS logic element. The equivalent pullup and pulldown resistance for a logic gate output depends on process-dependent characterisitics for n-channel and p-channel transistors, the width-to-length ratios of the individual transistors, and the series connection required for the pulldown section of a NAND gate or the pullup section of a NOR gate.

The following example demonstrates the propagation delay analysis for CMOS inverters. Subsequently the CMOS inverter-pair delay is compared with the equivalent inverter-pair delay for NMOS inverters.

Example 7.8-1. Consider a cascade of CMOS inverters in an n-well process. The n-channel pulldown transistor is a minimum-size device in a 2 μ technology. Assume the values of K' for the n-channel and p-channel transistors are 45 μA/V^2 and 15 μA/V^2 respectively. The p-channel pullup is sized for symmetrical output drive capability. The thresholds are $V_{TN} = 1$ V and $V_{TP} = -1$ V, and the supply voltage is $V_{DD} = 5$ V. The gate-oxide capacitance for both transistor types is $C_{ox} = 1$ fF/μ^2.

Determine the inverter-pair delay for this CMOS inverter in terms of absolute time and in terms of τ_P from Eq. 7.8-14.

Solution. Because the n-channel pulldown is of minimum size, it has $W_N = L_N = 2$ μ. The p-channel pullup must have a width-to-length ratio of 3 to compensate for the relative values of K' assumed for the p- and n-channel transistors. Choose $L_P = L_N = 2$ μ and $W_P = 3L_P = 6$ μ.

From Eqs. 7.8-23, 7.8-24, and 7.8-25,

$$C_{GC} = C_{GN} + C_{GP} = C_{ox}(W_NL_N + W_PL_P) = C_{ox}L_N(W_N + W_P)$$

Then

$$C_{GC} = 1 \text{ fF}/\mu^2 \times 2\ \mu(2\ \mu + 6\ \mu) = 0.016 \text{ pF}$$

From Eq. 7.8-4,

$$R_N = \frac{2 \times 2\ \mu}{45\ \mu\text{A/V}^2 \times 2\ \mu(5 \text{ V} - 1 \text{ V})} = 11.1 \text{ k}\Omega$$

and

$$R_P = \frac{2 \times 2\ \mu}{15\ \mu\text{A/V}^2 \times 6\ \mu(5 \text{ V} - 1 \text{ V})} = 11.1 \text{ k}\Omega$$

From Eqs. 7.8-26 and 7.8-27,

$$t_{LH} = 2R_PC_{GC} = 2 \times 11.1 \text{ k}\Omega \times 0.016 \text{ pF} = 0.355 \text{ ns}$$

and from Eq. 7.8-28, for symmetrical output drive,

$$t_{HL} = 2R_N C_{GC} = t_{LH} = 0.355 \text{ ns}$$

The inverter-pair delay is

$$t_{ipd} = t_{LH} + t_{HL} = 0.355 + 0.355 = 0.71 \text{ ns}$$

From Eq. 7.8-13, the process characteristic time constant is

$$\tau_P = R_{ss} C_{GN} = 5.56 \text{ k}\Omega \times 0.004 \text{ pF} = 0.022 \text{ ns}$$

Thus, $t_{ipd} = 32\tau_P$.

The inverter-pair delay for CMOS inverters determined in the preceding example should be compared with the inverter-pair delay for the NMOS inverters of Eq. 7.8-18. For NMOS with a 4:1 size ratio, $t_{ipd} = 20\tau_P$. The two inverter-pair delays demonstrate a 50% decrease in raw circuit speed for CMOS as compared with NMOS. The analysis in terms of τ_P removes geometric dependencies from the comparison. This analysis does, however, depend on the relative values of K' for the n- and p-channel transistors but does not depend on the choice of symmetric drive for the CMOS inverter.

7.8.8 Interconnection Characteristics

The analysis up to this point has been concentrated on logic gates and process characteristics to analyze signal propagation delay. This section includes a brief introduction to the interconnection capacitance and resistance that results from the connection of the output of one gate to the input of another gate. This discussion is limited to on-chip interconnections. Such interconnections are possible using one of several layers such as one or more layers of metal, polysilicon, or diffusion.

The proper interconnection medium depends on the physical properties of the layers and on circuit topological constraints. The metal layer is the most flexible because it does not interact directly with either of the other two layers to create transistors. Each of the interconnection layers exhibits parasitic capacitance to substrate (ground) that slows signal propagation. The value of this capacitance per unit area is given for an NMOS process in Table 2A.4 of Appendix 2A and for a CMOS process in Table 2B.4 of Appendix 2B. The capacitance for metal and polysilicon is considered in terms of the classical case of parallel plates separated by a dielectric. As a first-order approximation, the metal and polysilicon capacitances are independent of voltage and geometrical shape. Capacitance densities from metal to substrate or polysilicon on field oxide to substrate are typically less than C_{ox} by a factor of 10 to 20. Capacitance from diffusion to substrate consists of two primary components: bottom capacitance per unit area and sidewall capacitance per unit length. The sidewall capacitance is given per unit length for convenience because the diffusion region is considered to be of constant depth. Diffusion capacitance is caused by the reverse-biased diode junction between diffusion and substrate and is voltage-dependent. As a simple approximation, the voltage dependence is eliminated by choosing a voltage that provides a nominal

capacitance value. For minimum-width geometries, diffusion capacitance per unit area is typically less than C_{ox} by a factor of 5 to 10. Even though the capacitances per unit area for metal, polysilicon, and diffusion are factors of 5 to 20 less than C_{ox}, the area associated with interconnection is usually much larger than the transistor gate area. As a consequence, these interconnection loading effects may represent the dominant capacitive loading on many nodes, particularly as device geometries scale down to the 1 μ range and below.

The interconnection layers exhibit resistance to current flow as given by the values in Table 2A.4 of Appendix 2A for an NMOS process and the values in Table 2B.4 of Appendix 2B for a CMOS process. The resistances of polysilicon and diffusion are typically greater than the resistance of metal by three orders of magnitude. For short interconnections between adjacent devices, the resistance of the interconnection may be safely ignored in comparison to the effective transistor resistances. For longer interconnections, polysilicon and diffusion present large resistances that cannot be ignored. For this reason, metal is used for long interconnections.

Logic building block characteristics and associated delays can be estimated as soon as a logic diagram or circuit diagram with device sizing is complete. Interconnection delays depend so heavily on circuit layout that their effect is often neglected until layout is available. This is an unfortunate situation because interconnection capacitance is a significant delay factor for digital logic circuits. A rough rule of thumb with which to consider interconnection delays for minimum-size digital circuits prior to circuit layout can be derived from the following premises.

1. Assume that the average interconnection capacitance per unit area is one-tenth that of C_{ox}.
2. Assume that the local interconnection area is 10 times the gate area.

With these assumptions, interconnections can be modeled by doubling the effective capacitance of each driven gate. Although this is a crude approximation, it is substantially better than ignoring interconnection delays until after layout is available. For a particular design style and technology, this approximation can be improved with measurements from previous, similar designs. For example, many gate array manufacturers provide average interconnection capacitance estimates based on anticipated die size. Even better estimates of interconnection capacitance for critical nodes can be obtained if a floor plan of the circuit is available to describe the relative placements of digital building blocks within the die area.

Many aspects of signal propagation delay have been examined in this section. The delay is a combination of gate delay and interconnection delay. Ratio logic was found to exhibit asymmetric rising and falling delay times that differ by the pullup/pulldown resistance ratio. A process charcteristic time constant was defined to allow unbiased delay comparisons of different processes. Inverter-pair delay was introduced to capture the effects of both rising and falling delays. Then a special circuit configuration called a superbuffer was introduced to minimize the asymmetric delay characteristics of ratio logic. Next, delay analysis of both

NMOS and CMOS gates was presented, followed by an example to compare the two technologies with respect to delay characteristics. Finally, the electrical characteristics of interconnections were introduced.

7.9 CAPACITIVE LOADING CONSIDERATIONS

In the preceding section, consideration was given to signal propagation delays for logic gates that were loaded by single, identical logic gates. A significant problem in large-scale integrated circuit design for digital circuits is driving the relatively large capacitive loads caused by high gate fanout, interconnections, and off-chip connections. As stated in Sec. 7.2, both signal-level degradation and propagation delay are considered when specifying the maximum fanout for a logic circuit. Because of the extremely high input resistance of MOS devices, minimal signal-level degradation occurs even in driving a large number of gates. The primary fanout consideration is the input capacitance of successive logic circuits and their interconnections. Now the analysis of the previous section is expanded to consider the effects of heavy capacitive loading and to investigate circuit techniques to minimize the associated increase in signal propagation delay.

7.9.1 Capacitive Loading

Several factors contribute to capacitive loading of the output of a logic gate. These include inputs to other gates, interconnection routing or buses, bonding pads, and external loads. Regardless of the cause, each adds parasitic capacitance and contributes cumulatively and nearly linearly to the overall delay. If the total capacitive loading at the output node of a logic gate caused by these factors is found to be C_T, then the propagation delay time constant can be approximated by the expression

$$\tau_T = R_T C_T \tag{7.9-1}$$

where R_T is the equivalent charging or discharging resistance. If the capacitance C_T is driven by a reference inverter with pulldown resistance R_T and gate capacitance C_G, the average propagation delay is

$$t_{dly} = \frac{t_{apd} C_T}{C_G} \tag{7.9-2}$$

where t_{apd} is the average logic stage propagation delay of the logic family defined in terms of inverter-pair delay by

$$t_{apd} = \frac{t_{ipd}}{2} \tag{7.9-3}$$

From Eq. 7.8-18, this can be written in terms of the process characteristic time constant τ_P for an NMOS reference inverter with pullup/pulldown ratio k as

$$t_{dly} = \frac{2(1 + k)\tau_P C_T}{C_G} \tag{7.9-4}$$

7.9.2 Logic Fan-out Delays

In many situations, the output of a logic stage is required to drive more than one equivalent gate input. The response time is slowed because of the parasitic capacitance of the additional inputs. If the fan-out, that is, the number of equivalent reference inverter loads to be driven, is f, then the total capacitive load is fC_G where C_G is the capacitive load of a single reference inverter. Replacing C_T of Eq. 7.9-2 with fC_G, the average stage delay for a single stage with a fan-out of f is

$$t_{stage} = t_{apd} f \tag{7.9-5}$$

With these observations, the delay along a homogeneous signal path in a digital integrated circuit can quickly be approximated. Assume a signal passes through N levels of logic with an equivalent fan-out of f_i at the ith stage. Then the total path delay is given as

$$t_{path} = t_{apd} \sum_{i=1}^{N} f_i \tag{7.9-6}$$

For the analysis so far all stages are identical to a reference inverter, and interconnection capacitance has been neglected.

It is often convenient to decompose the total capacitive loading of Eq. 7.9-1 into that caused by MOS gate loading plus that caused by other factors, such as bus or interconnection loading. Assume a node is loaded by the equivalent of f reference inverter inputs and capacitive loading C_I from interconnections. From the interconnection area and the process-dependent capacitance per unit area (see Appendices 2A and 2B), C_I can be determined. If C_G is the input capacitance of the reference inverter, then the interconnection load C_I is equivalent to that of m reference inverter loads, where

$$m = \frac{C_I}{C_G} \tag{7.9-7}$$

It follows that the average propagation delay of this node due to both fan-out and interconnection loading is given by

$$t_{node} \approx (m + f) t_{apd} \tag{7.9-8}$$

If a signal must propagate through a sequence of N stages, where the output drive of each stage is equivalent to that of the reference inverter, it follows from Eqs. 7.9-6 and 7.9-8 that the signal propagation delay through this N-stage path can be approximated by

$$t_{path} \approx t_{apd} \sum_{i=1}^{N} (m_i + f_i) \tag{7.9-9}$$

where f_i is the equivalent number of reference inverter loads and m_i is the equivalent of interconnection loads on the ith node. Including the interconnection

loading is important in most circuits and will dominate the actual gate loading for long connections or for buses. Interconnection loading is particularly significant in submicron structures.

Equation 7.9-2 has a straightforward and useful extension. Assume that an inverter is to drive a total capacitive load C_T. The drive capability for this inverter is improved by increasing the widths of both its pullup and pulldown transistors by a factor of θ over the corresponding widths for the reference inverter. Note that this improved inverter has an input capacitance θC_G. Then it can be shown that the equivalent propagation delay for this inverter is given by

$$t_{inv} = \frac{t_{apd} C_T}{\theta C_G} \tag{7.9-10}$$

Although it could be the case that all inputs to all logic gates are equal in size to that of the reference inverter, in many situations it is advantageous to use logic circuits where the input devices have nonhomogeneous sizes. For example, this might occur where transistors are individually sized to reduce the delay in driving capacitive loads.

Equation 7.9-10 can be extended to obtain the signal propagation delay of paths that contain logic gates with varying drive capabilities. Under the assumption that the reference inverter device sizing ratio k is used for all gates in the cascade, and that the drive capability of the output of the ith gate in the cascade is θ_i times that of the reference inverter (i.e., the corresponding resistances in the model of Fig. 7.8-4 are $1/\theta_i$ times those of the reference inverter), it can be shown that the signal propagation delay through an N-stage path can be approximated by

$$t_{path} \approx t_{apd} \sum_{i=1}^{N} \frac{m_i + f_i}{\theta_i} \tag{7.9-11}$$

At the expense of some layout area, the increase in drive capability can be obtained by increasing the width of the driving transistors while keeping their length and device sizing ratio k constant.

In summary, a simple method of obtaining an approximation to the signal propagation delay along a path in digital circuits based on the average stage delay has been presented. Interconnections and other parasitic loading factors are easily included in the calculation once layout is determined. The approximation may have a significant error of $\pm 50\%$ or more. The simple analysis presented here is, however, good enough for at least two purposes. First, it allows an estimate of circuit speed for use in comparing alternative designs prior to implementation. Second, it is usually adequate to determine the critical paths that must be analyzed in detail and possibly modified to improve performance. If more accurate timing information is required, a timing analysis program or circuit simulator such as SPICE can be used. It is often prudent to identify the critical delay paths in a system and perform a detailed analysis of those paths to obtain a more precise estimate of the system delay.

7.9.3 Distributed Drivers

It can be seen from Eq. 7.9-2 that the delay associated with driving a large capacitive load from a minimum-size inverter increases linearly with the load capacitance C_T. This linear dependence is particularly troublesome because situations often arise where the total load capacitance may be as much as $100C_G$ to $10{,}000C_G$ (see Table 7.9-1). The corresponding increase in delay by a factor of 100 to 10,000 is seldom acceptable. The following question naturally arises: Is there a faster way to drive a large capacitive load? At first glance, Eq. 7.9-11 suggests that if θ_i is made larger by widening the pullup and pulldown transistors to reduce their equivalent resistances, then the delay can be reduced. (Normally the pullup/pulldown ratio k is maintained to ensure valid logic voltage levels.) However, these changes cause heavier capacitive loading on previous stages, perhaps negating the net performance gain. The following example shows that, in terms of propagation delay, it may be better to distribute the load than to increase the drive of a single stage when an output is driving a high-fanout load.

Example 7.9-1. Assume that a minimum-size inverter drives a set of 10 other minimum-size inverters, as shown in Fig. 7.9-1*a*. Estimate the propagation delay from V_i to V_c if the single inverter drives the 10 inverters directly, and if the single inverter drives two other minimum-size inverters that each drive five inverters, as shown in Fig. 7.9-1*b*. Neglect interconnection capacitance and the inversion of the logic signal.

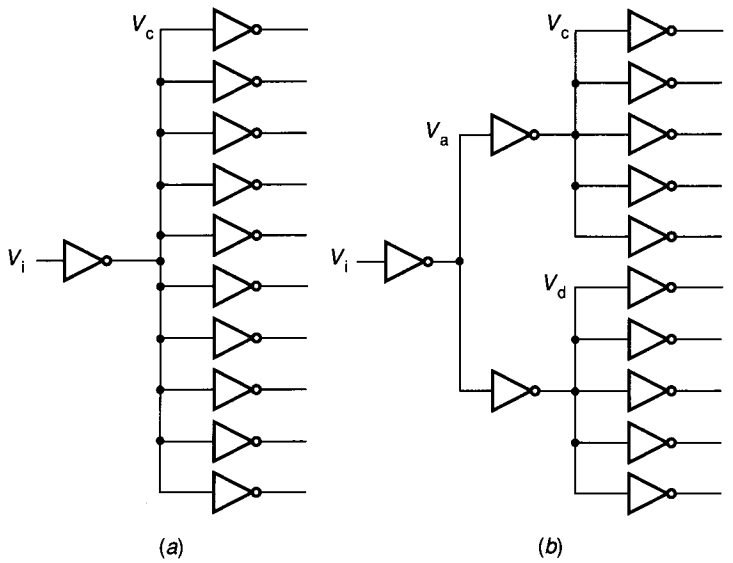


FIGURE 7.9-1
Distributed drivers versus concentrated driver: (*a*) 1:10 capacitive load, (*b*) 1:2:5 capacitive load.

Solution. Assuming that the minimum-size inverter has a pullup/pulldown ratio of $k = 4$, the average stage delay required to drive a single minimum-size inverter is t_{apd}. It follows from Eq. 7.9-2 that if the inverter drives 10 identical inverters, the total gate capacitance will be increased by a factor of 10, and the total delay becomes $10t_{apd}$.

If the minimum-size inverter drives two stages that, in turn, drive five stages each, the propagation delay will be the sum of the two delays. The average delay to drive two identical inverters will be twice the delay for a single inverter, or $2t_{apd}$. The average delay for the second stage to drive five identical stages will be five times the delay for a single inverter, or $5t_{apd}$. The combined delay from input to output with the intermediate stage will be $7t_{apd}$, which is less than the $10t_{apd}$ required for driving the 10 inputs directly.

The example shows that it may be better to include intermediate stages than to drive a heavy capacitive load from a single stage. Table 7.9-1 provides a comparison of some of the capacitive loading requirements that must be addressed for the typical process listed in Table 2A.4 of Appendix 2A. Although interconnection loading may represent capacitive loading equivalent to 100 or more reference inverters, pad loading and off-chip loading are often even larger. The divide-and-conquer strategy presented in the example will not work for a single large load. Methods of driving large, concentrated capacitive loads will be considered in the following section.

7.9.4 Driving Off-Chip Loads

Digital integrated circuits generally use the smallest possible transistors to implement logical functions. The small size is important to maximize the number of circuits per unit area and therefore minimize silicon area and cost. Ultimately, logic circuits must provide results of internal circuit operations to the outside world. These logic signals must overcome both the inherent loading effects of the bus or interconnection path from the source of the output signal to an external pin and the loading effects of some other circuit to which this output signal acts as an input. Additionally, it is frequently desirable to generate signals that are compatible, relative to logic voltage levels, with some other logic family, such as Advanced Low-Power Schottky TTL (ALSTTL), which provides current as well as capacitive loading. Other logic families are handled by increasing the

TABLE 7.9-1
Typical capacitive loading

Load	C_T	C_T/C_G
Single reference inverter (3 μ × 3 μ)	0.0063 pF	1
Ten reference inverters	0.063 pF	10
4 mm × 4.5 μ metal bus	0.450 pF	71
Standard output pad (100 μ × 100 μ)	0.250 pF	40
Oscilloscope probe	10.0 pF	1587
Memory chip address pin	5.0 pF	794

width of the driving transistors to increase current capability and by varying the pullup/pulldown ratio k to match external logic voltages. To illustrate the former case, the following example demonstrates the capacitive loading effects of driving a simple output bonding pad from a minimum-size inverter.

Example 7.9-2. Consider driving a metal output bonding pad from a minimum-size inverter in the NMOS process summarized in Table 2A.4 of Appendix 2A. Calculate the capacitance ratio between the capacitance of the bonding pad and the input gate capacitance of a minimum-size inverter. Using this ratio, estimate the delay to drive the output pad in terms of the reference delay t_{apd}.

Solution. The size of an output bonding pad is typically 100 $\mu \times 100$ μ. From the NMOS process electrical characteristics

$$C_{pad} = 0.025 \text{ fF}/\mu^2 \times 10,000 \ \mu^2 = 0.25 \text{ pF}$$

Note that this matches the value listed in Table 7.9-1. The dimensions of an input gate for a minimum-size inverter in this process are 3 $\mu \times 3$ μ.

$$C_G = 0.7 \text{ fF}/\mu^2 \times 9 \ \mu^2 = 0.0063 \text{ pF}$$

Thus, the ratio of C_{pad} to C_G is

$$C_{ratio} = \frac{C_{pad}}{C_G} = 39.7$$

From Eq. 7.9-2 the average propagation delay of the reference inverter driving the output pad is

$$t_{dly} = 39.7 t_{apd}$$

From Eqs. 7.8-14, 7.8-18, and 7.9-3, the value for the average gate delay for the NMOS process of Appendix 2A with $k = 4$ is $t_{apd} = 0.63$ ns. The average propagation delay in driving the bonding pad is thus 25 ns.

The following example demonstrates the effect of adding an external capacitive load to the bonding pad just considered.

Example 7.9-3. Consider the case of a standard oscilloscope probe connected directly to the output bonding pad of Example 7.9-2. Determine the approximate average propagation delay that will result.

Solution. The total load capacitance will be the sum of the output pad capacitance and the oscilloscope probe capacitance. Table 7.9-1 indicates that an oscilloscope probe provides 10 pF of capacitive load. Thus,

$$C_{load} = C_{pad} + C_{probe} = 0.25 \text{ pF} + 10 \text{ pF} = 10.25 \text{ pF}$$

The capacitance ratio will be

$$C_{ratio} = \frac{C_{load}}{C_G} = 1627$$

The average propagation delay is obtained from Eq. 7.9-2:

$$t_{dly} = 1627 t_{apd} = 1627 \times 0.63 \text{ ns} = 1025 \text{ ns}$$

Contrast this delay with the 0.63 ns average propagation delay of the reference inverter. Such a delay is obviously detrimental to high-speed operation of digital circuits, limiting clock frequency to less than $1/(2t_{dly}) = 488$ kHz! The oscilloscope probe capacitance is comparable in value to the total capacitance encountered in driving inputs on other integrated circuit chips (see Table 7.9-1).

7.9.5 Cascaded Drivers

It is obvious from the two examples of the preceding section that the signal delay encountered in driving off-chip loads directly from a minimum-size inverter is unacceptable. Fortunately, there are circuit configurations that reduce the effective delay in driving large capacitive loads. One good circuit configuration employs a cascade of inverters with increasing current-drive capability to minimize this delay.

Assume a signal is available at the output of a minimum-size inverter (reference inverter) and that it is to drive a load C_L. From Eq. 7.9-2 the average propagation delay associated with driving this load directly is

$$t_{dir} = \frac{t_{apd} C_L}{C_G} \tag{7.9-12}$$

where t_{apd} is the average logic stage delay and C_G is the input capacitance of the reference inverter. For any integer $n \geq 1$, define α by the expression

$$\alpha = \left(\frac{C_L}{C_G}\right)^{1/n} \tag{7.9-13}$$

Alternatively, n can be represented in terms of α as

$$n = \frac{\ln(C_L/C_G)}{\ln \alpha} \tag{7.9-14}$$

Consider now the alternative structure of Fig. 7.9-2 for driving a load C_L. This structure is composed of a cascade of n inverters (including the initial reference inverter) each sized by the 4 : 1 sizing rule and each with a drive capability that is α times as large as the previous stage. The width and length of the kth stage can be characterized by the equations

$$\begin{aligned} W_{dk} &= \alpha^{k-1} W_{d1} \\ L_{dk} &= L_{d1} \\ W_{uk} &= W_{dk} \\ L_{uk} &= 4L_{dk} \end{aligned} \tag{7.9-15}$$

where the device dimensions W_{dk} and L_{dk} correspond to the pulldown transistor and W_{uk} and L_{uk} correspond to the pullup transistor of the kth inverter structure in the cascade, as indicated in Fig. 7.9-2. It can be observed that the load on the kth stage C_{Lk} is related to the reference inverter input capacitance C_G by the expression

$$C_{Lk} = \alpha^k C_G \tag{7.9-16}$$

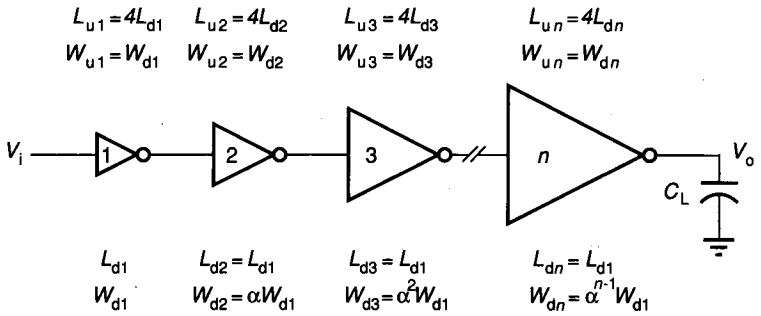


FIGURE 7.9-2
Cascaded drivers for a concentrated load.

From Eq. 7.9-2 it follows that the average propagation delay of the first inverter is αt_{apd}. It can be shown (Prob. 7.41) that the average propagation delay for each inverter in this geometric cascade is αt_{apd}. Hence, it follows from Eq. 7.9-11 with $f_i/\theta_i = \alpha$ and $m_i = 0$, that the total delay for the cascade is

$$t_{cas} = n\alpha t_{apd} \tag{7.9-17}$$

Let r be the ratio between the propagation delays of the direct-drive circuit and of the geometric cascade approach. From Eqs. 7.9-12 and 7.9-17 it follows that

$$r = \frac{t_{cas}}{t_{dir}} = \frac{n\alpha t_{apd}}{t_{apd}C_L/C_G} = \frac{n\alpha C_G}{C_L} \tag{7.9-18}$$

It is our goal to determine n and α to minimize r and thus minimize the propagation delay in driving the load. From Eq. 7.9-14 it follows that n can be eliminated from the expression for r to obtain the expression

$$r = \frac{\ln(C_L/C_G)}{C_L/C_G}\frac{\alpha}{\ln\alpha} \tag{7.9-19}$$

The terms involving capacitance are fixed by the load requirements. The goal is thus to determine α in Eq. 7.9-19 to minimize r. The second term on the right-hand side of Eq. 7.9-19 is plotted in Fig. 7.9-3. It is easy to see that $\alpha/\ln\alpha$ has a wide local minimum at $\alpha = e$ with value e. A plot of the number of stages n versus C_L/C_G for minimizing the delay with $\alpha = e$, as obtained from Eq. 7.9-14, is shown in Fig. 7.9-4. Plots of n versus C_L/C_G for $\alpha = 3$ and $\alpha = 5$ are also shown in Fig. 7.9-4.

It should be noted that because n is the number of cascade stages, n must assume an integer value greater than or equal to 1. Quantization of device geometries during layout precludes setting α to an exact ratio of e. In fact, α is usually set to a value greater than e to reduce the number of cascade stages required while still reducing the propagation delay significantly. As can be seen from Fig. 7.9-3, as long as α is between 2 and 4, the deviation from minimum delay is less than about 5%. For conservative design, the values for n and α can be selected to drive a load a little larger than C_L within the allotted delay time.

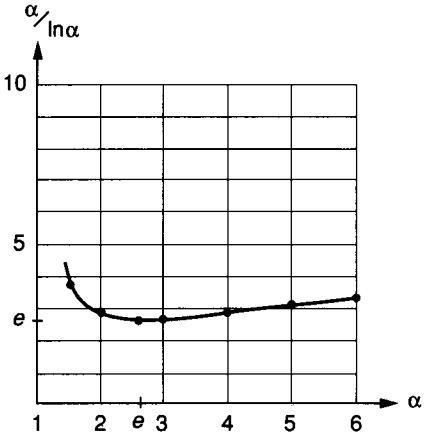


FIGURE 7.9-3
Plot of $\alpha/\ln\alpha$ versus α.

It is instructive to obtain an appreciation for how much benefit is practically derived from the cascaded driver approach. It can be shown from Eq. 7.9-17 that for small loading ratios, the speed improvements are small and the area overhead associated with the cascade may not be justified. For large capacitive load ratios, the speed improvement offered by the cascade is significant. For example, from Eq. 7.9-18 a seven-stage optimally sized cascade would drive a capacitive load that is approximately $1100C_G$ in 1.7% of the time required with direct drive!

Two final points deserve mention. First, if the number of inverter stages in the cascade is odd, the output signal is inverted. If inversion is unacceptable, a minimal delay increase occurs if the circuit is preceded by a reference inverter. Second, even though the speed improvements are significant for large n, the silicon area penalty is also quite high. The active silicon area grows as a geometric function of the number of stages. For example, the final stage in a seven-stage optimally weighted cascade requires $e^6 = 403$ times as much active area as a reference inverter. This is almost twice the area of all the preceding cascaded stages combined.

The following example compares the delay for three stages and the optimal number of stages driving the oscilloscope probe of Example 7.9-3.

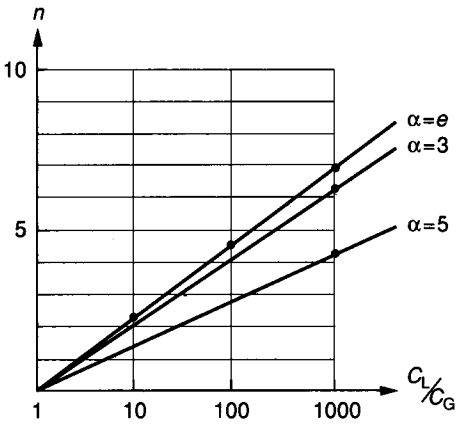


FIGURE 7.9-4
Plot of n versus C_L/C_G for $\alpha = 3, e$, and 5.

Example 7.9-4. Determine the number of stages required to drive the output pad of Example 7.9-3 with minimum delay. Calculate the delay for this minimum delay case and for the case where three stages sized according to Eq. 7.9-13 are used.

Solution. From Eq. 7.9-14 the optimal number of stages n is found with $\alpha = e$. From Example 7.9-3, $C_L/C_G = 1627$, so $n = \ln(1627) = 7.39$. This is rounded to seven stages, including the reference inverter at the signal source. The propagation delay for the seven stages is given by Eqs. 7.9-17 and 7.9-13 as

$$t_7 = 7 \times 2.88 \times t_{apd} = 20.16 t_{apd}$$

Note from Eq. 7.9-17 that each stage contributes equally to the delay.

If $n = 3$ is chosen, then from Eq. 7.9-13, $\alpha = 11.76$ for $C_L/C_G = 1627$. From Eq. 7.9-17, this gives a delay of

$$t_3 = 3 \times 11.76 t_{apd} = 35.3 t_{apd}$$

Note that the delay for three cascaded driver stages is only 75% more than the delay for seven cascaded driver stages. If this delay is acceptable, the area savings is significant. Even with just three cascaded driver stages, the delay is reduced to only 2% of the delay for the load driven directly from a reference inverter.

Standard output driver stages are needed to drive most output nodes for high-speed digital circuits. These circuits are called *pad drivers* and are usually available in libraries of standard circuits. Pad drivers are generally just a cascade of inverters with a geometrically increasing drive capability sized to give reduced delay. Figure 7.9-5 shows a plot of a Low-Power Schottky TTL-compatible output pad driver circuit. Figure 7.9-6 gives the circuit schematic showing the number of stages and the device sizing for each stage. Note the use of two superbuffers and of enhancement transistors for both the pullup and pulldown in the final stage. The use of an active enhancement pullup prevents static power dissipation in the unloaded pad driver output stage.

The analysis in this section has covered the various aspects of large capacitive loads. First, delays from capacitive loading and from logic fanout were demonstrated. Then the delays caused by driving large off-chip capacitances were analyzed. Two approaches to driving these loads were presented: distributed drivers and optimally cascaded drivers.

7.10 POWER DISSIPATION

One major limitation of MOS ICs is internal power dissipation. Although the power dissipation of MOS circuits is generally much less than that of bipolar integrated circuits performing the same function, it becomes a major factor limiting the size of VLSI MOS circuits. Electrical power dissipation in an integrated circuit is converted to heat that must be removed through the circuit's packaging. Integrated circuit packaging offers a resistance to heat removal. The heat flow through this resistance generates a temperature difference across the package analogous to the voltage difference caused by current flow through an electrical resistance. For a given integrated circuit package, a specified maximum temperature of the integrated circuit, and a specified ambient temperature, a maximum

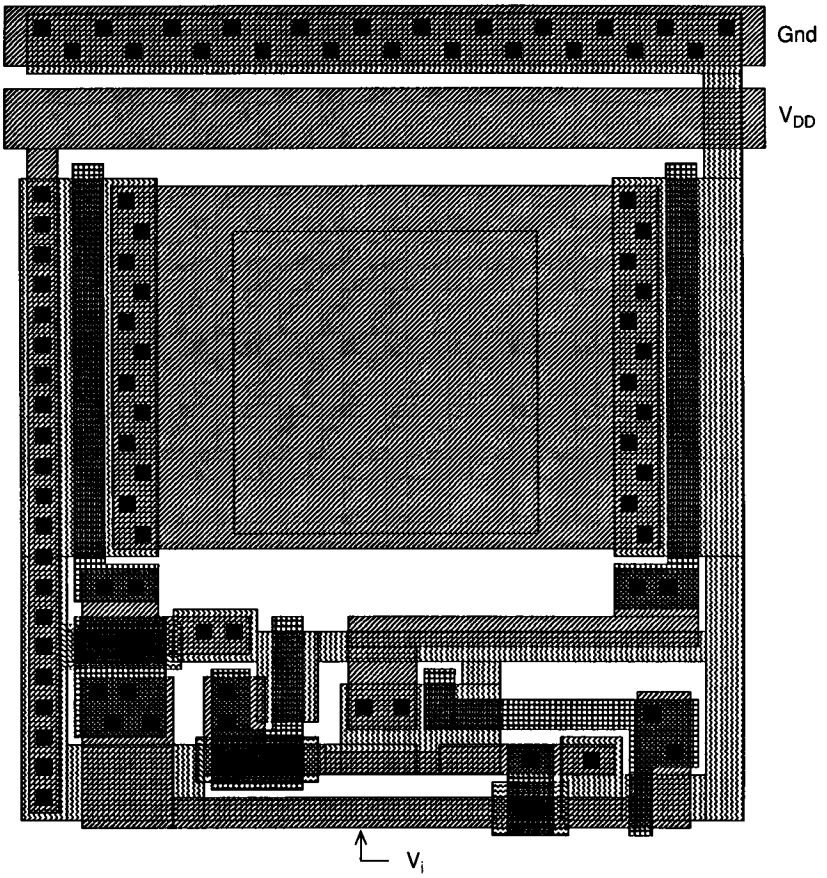


FIGURE 7.9-5
LSTTL output pad driver layout.

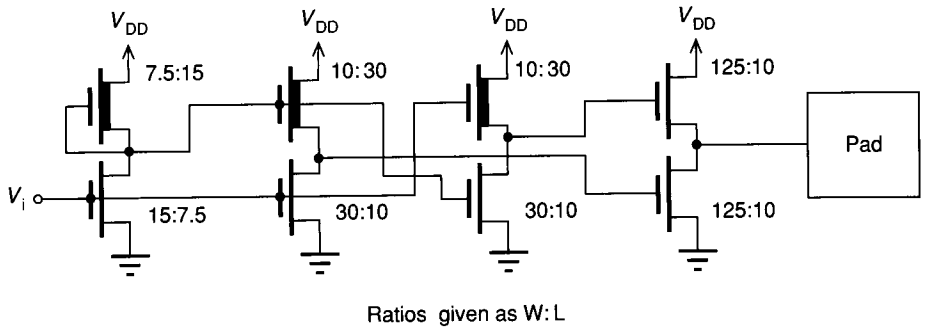


FIGURE 7.9-6
Circuit diagram for LSTTL output pad driver.

power dissipation (heat flow) can be specified. The thermal resistance θ_{JA} for a plastic 40-lead package is about 100°C/W, and a maximum integrated circuit junction temperature T_J of about 150°C is specified for reliability considerations. The thermal resistance θ_{JA} is the sum of junction to die, die to package, and package to ambient resistances.[9] If the maximum ambient temperature T_A is expected to be 50°C, then the maximum power dissipation is limited to

$$P_D = \frac{T_J - T_A}{\theta_{JA}} = \frac{150 - 50}{100} = 1 \text{ W} \tag{7.10-1}$$

Most plastic or ceramic integrated circuit packages can dissipate 1–2 W of power in an ambient temperature of 70°C. As a second, special consideration, low power dissipation is important in applications that must operate from batteries. In this section the main sources of power dissipation are identified, first for NMOS circuits and then for CMOS circuits.

7.10.1 NMOS Power Dissipation

To help explain integrated circuit power dissipation, consider the reference inverter of Fig. 7.10-1. When V_i is low, M1 is off, causing a supply current I_D of 0 A. Hence, the power dissipation is essentially zero! If, however, V_i is high (V_{DD}), then M1 is ohmic and M2 is saturated, so from Eq. 3.1-4

$$I_D = K'\left(\frac{W_2}{L_2}\right)\frac{V_{TD}^2}{2} \tag{7.10-2}$$

With $K' = 20\ \mu\text{A/V}^2$, $L_2 = 4W_2$, and $V_{TD} = -3.5$ V, the drain current is $I_D = 31\ \mu$A. Hence, the power dissipation of this single inverter with $V_i = V_{DD} = 5$ V is approximately 150 μW. Although the power dissipation of a single inverter is small, VLSI circuits contain thousands of inverters and gates. If, on the average, half of the inputs are in the low state and half are in the high state, then a circuit with N reference inverters will have an average *static power dissipation* of

$$P_{avg} = \left(\frac{N}{2}\right)150\ \mu\text{W} \tag{7.10-3}$$

and a worst-case power dissipation of

$$P_{wc} = (N)150\ \mu\text{W} \tag{7.10-4}$$

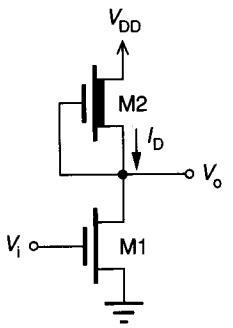


FIGURE 7.10-1
NMOS reference inverter.

From the NOR gate of Fig. 7.4-3, it can be concluded that if the pullup device ML is the same size as the pullup device of the reference inverter, then the power dissipation of the multi-input NOR is the same as that of the reference inverter. This is true irrespective of the number of inputs to the NOR gate.

Example 7.10-1. If a packaged IC can dissipate a maximum average power of 1 W, how many NMOS NOR gates can be accommodated? Assume the NOR gate pullup is sized to match the reference inverter.

Solution. From Eq. 7.10-3,

$$N = \frac{(1\text{ W})2}{150\ \mu\text{W}} = 13,333 \text{ NOR gates} \tag{7.10-5}$$

In the worst case, the circuit could be limited to about 6667 gates.

Note that if greater driving capability is required, for example, to drive highly capacitive loads as described in Sec. 7.9, then the W_2/L_2 ratio in Eq. 7.10-2 will increase, causing a corresponding increase in average static power dissipation. From Example 7.9-4, W_2/L_2 for the pullup device of the last stage of an optimally sized seven-stage pad driver is e^6 times as large as the width-to-length ratio of the load device of a reference inverter. Consequently, the last inverter alone in the seven-stage pad driver would have a power dissipation of 60 mW when it conducts! This is more than twice the static power dissipation of all preceding stages combined.

Logic circuits with inputs that change rapidly can dissipate considerable additional power over their static power requirements. This additional power is referred to as *dynamic power dissipation*. For example, immediately after a low-to-high input transition for the reference inverter of Fig 7.10-1, the output voltage will be near V_{DD} as the load capacitance starts to discharge. While the output is near V_{DD}, the pulldown device M1 is in the saturation region, causing an instantaneous drain current of

$$I_{D1} = K' \frac{W_1}{L_1} \frac{(V_{DD} - V_{TN})^2}{2} \tag{7.10-6}$$

With $V_{DD} = 5$ V, $K' = 20\ \mu\text{A/V}^2$, $W_1 = L_1$, and $V_{TN} = 1$ V, the drain current becomes 160 μA and the initial power dissipation is 800 μW. This is considerably higher than the static power dissipation of 150 μW found earlier. The actual power dissipation will vary in a continuous manner from 800 μW to 150 μW during an output high-to-low transition. With fast input signals to a digital integrated circuit, the dynamic power dissipation obtained from a time average of the instantaneous power dissipation minus the static power dissipation can be much larger than the static power dissipation and must be taken into consideration.

The two primary components of power dissipation for NMOS circuits are static power dissipation and dynamic power dissipation. For circuits with average logic signal changes less frequent than about 10 MHz, static power dissipation is dominant. For higher-frequency signal changes, dynamic power becomes the major consideration. Following is a quantitative analysis of dynamic power dissipation for CMOS circuits. The corresponding analysis for NMOS circuits is similar.

7.10.2 CMOS Power Dissipation

CMOS logic has long been used for circuits in applications requiring extremely low power, such as space probes, calculators, and wristwatches. Perhaps the most important asset of CMOS for today's digital logic designs is its low power dissipation and attendant minimal cooling requirements. Power dissipation in CMOS can be classified into three categories: static power dissipation, *dc switching power* occurring when both transistors conduct momentarily during a transition, and *ac switching power* lost while charging (discharging) capacitive loads. The combination of dc switching power and ac switching power is often called dynamic power dissipation.

Static power dissipation is negligible for CMOS logic circuits. The dc path from power to ground is always broken by an off transistor when the circuit is quiescent. With today's digital CMOS circuits, static power dissipation is not a consideration compared with dynamic power dissipation.

The second component of power dissipation, dc switching power, occurs during the time that the inputs to a logic gate are between the valid logic levels. During this time, both transistors conduct, providing a path from V_{DD} to ground. For an individual gate, the average power dissipated increases with switching time for its input logic signal; that is, the longer the input stays between valid logic levels, the longer the conductive path from V_{DD} to ground exists. The power lost to this cause is typically less than 10% of total power dissipation for digital CMOS circuits such as memories and microprocessors.

The ac switching power is the main component of power dissipation for digital CMOS circuits. This power loss can be analyzed by considering the circuit of Fig. 7.10-2*a*. This circuit models a capacitively loaded inverter output during a low-to-high transition initiated at time zero. Assume no charge on the capacitor when the switch is initially closed. Since this is a simple *RC* circuit, it is well known that the current *i(t)* from the battery is

$$i(t) = \left(\frac{V}{R}\right)e^{-t/RC} \tag{7.10-7}$$

The total energy supplied by the battery is given by

$$e = \int_0^\infty V\, i(t)dt = V^2C \tag{7.10-8}$$

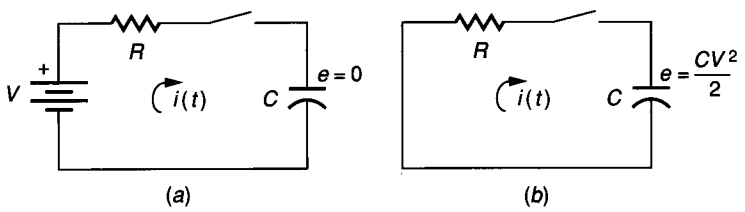


FIGURE 7.10-2
Equivalent circuits for ac power dissipation: (*a*) Charging circuit, (*b*) Discharging circuit.

Because the final energy stored on the capacitor is only (1/2) CV^2 then half the total energy supplied by the battery must be dissipated in the resistance while the capacitor is being charged. Now consider that the switch is opened momentarily, the battery is replaced by a short, and the switch is reclosed as shown in Fig. 7.10-2*b*. As the capacitor discharges, all its stored energy will be dissipated in the resistor. The total energy lost (dissipated in the resistor) during a capacitor charge/discharge cycle e_{cycle} must be CV^2.

For CMOS logic circuits, the capacitance in Eq. 7.10-8 represents the sum of the parasitic gate and interconnection capacitance. Note that Eq. 7.10-8 is independent of the charging and discharging resistances. The power dissipation is just the rate of use of energy. If the charge/discharge cycle occurs periodically with a period T, the average power dissipation is

$$P = \frac{e_{\text{cycle}}}{T} \tag{7.10-9}$$

In terms of operating frequency $f = 1/T$, the average power used is

$$P = e_{\text{cycle}} f = CV^2 f \tag{7.10-10}$$

This analysis shows that the average ac power dissipation for a CMOS circuit is proportional to the total capacitance, to the square of the supply voltage, and to the operating frequency. A similar analysis holds for NMOS circuits, except that the voltage swing is smaller because the lower logic level is above ground. For CMOS circuits, the ac power dissipation dominates, and both the ac power dissipation and the dc switching power increase with frequency.

Example 7.10-2. Consider an approximate limitation on the number of gates in a CMOS integrated circuit if the total power dissipation is limited to 1 W, the operating frequency is 10 MHz, and the supply voltage is 5 V.

Solution. It will be assumed that the ac power dissipation is dominant. From Example 7.8-1, a single gate load capacitance for a reference CMOS inverter is 0.016 pF. With the addition of interconnection capacitance, assume that the total capacitance driven by each logic gate is 0.032 pF. The total ac power dissipation is equal to the number of active gates N_a times the power dissipation of each gate. Under the assumptions given, the number of active gates to gve 1 W power dissipation is

$$N_a = \frac{P}{CV^2 f} = \frac{1}{0.032 \times 10^{-12} 5^2 10^7} = 125,000$$

In typical integrated circuits, only 10% to 25% of the gates switch during a given clock cycle. Using a conservative 25% figure, the total number of gates required to dissipate 1 W is

$$N_{\max} = \frac{N_a}{0.25} = 500,000$$

This can be compared with the required number of NMOS gates from Example 7.10-1 to show why CMOS is preferred in terms of power dissipation.

It is important to consider how future fabrication and operating conditions will affect power limitations on the number of gates in a CMOS integrated circuit. As CMOS circuits are reduced in size, it is expected that the operating frequency will increase and supply voltages and gate capacitance will decrease. If clock frequency is increased and power supply voltages and gate capcitance are reduced proportionally for a constant number of gates, the total CMOS power dissipation decreases rapidly because it is a linear function of frequency and capacitance and a squared function of voltage. The ability to operate faster and at reduced power is one of the driving forces for scaling down minimum device sizes.

In this section, static and dynamic power dissipation were introduced. Dynamic power dissipation includes both dc and ac switching power. Static power dissipation dominates for NMOS logic, whereas dynamic power dissipation dominates for CMOS logic.

7.11 NOISE IN DIGITAL LOGIC CIRCUITS

Electrical noise in integrated circuits represents an unwanted perturbation of desired signal voltages. Noise can arise from many sources and may or may not affect the operation of a circuit, depending on the magnitude of the noise disturbance and the sensitivity of the circuit to the disturbance. Noise disturbances may come from the external environment, as in electromagnetic interference or alpha particles, or may be internally generated from causes that can be controlled by the designer. Internal noise in digital integrated circuits comes from two primary sources: noise coupled from a common resistive path to power or ground, and noise that is capacitively coupled from another signal path. In this section the common causes of internal noise in digital integrated circuits are considered, the parameters used to specify digital integrated circuit immunity to noise are defined, and the noise immunity for NMOS and CMOS inverters is investigated.

7.11.1 Resistive Noise Coupling

Figure 7.11-1*a* shows a resistance R_c in the common ground path for two inverters. In practical circuits, this might be caused by a common diffusion path to a metal ground bus. The model of Fig. 7.11-1*b*, where each transistor is replaced by an equivalent resistance, is used for the following simplified analysis. If switch S1 is closed and switch S3 is open, the output voltage V_1 is determined as

$$V_1 = \frac{V_{DD}(R_1 + R_c)}{R_1 + R_2 + R_c} \tag{7.11-1}$$

Note that if $R_c = 0$ ohms, Eq. 7.11-1 reduces to Eq. 7.8-1. This output voltage corresponds to the low logic state. If R_c is nonzero, the output voltage is increased, thereby degrading the low logic voltage level.

A shift occurs in the output voltage V_1 if switch S3 of Fig. 7.11-1*b* is closed. This shift produces an unwanted noise voltage that, if large enough, can affect the logic value seen by subsequent stages. An equivalent circuit when both inverters

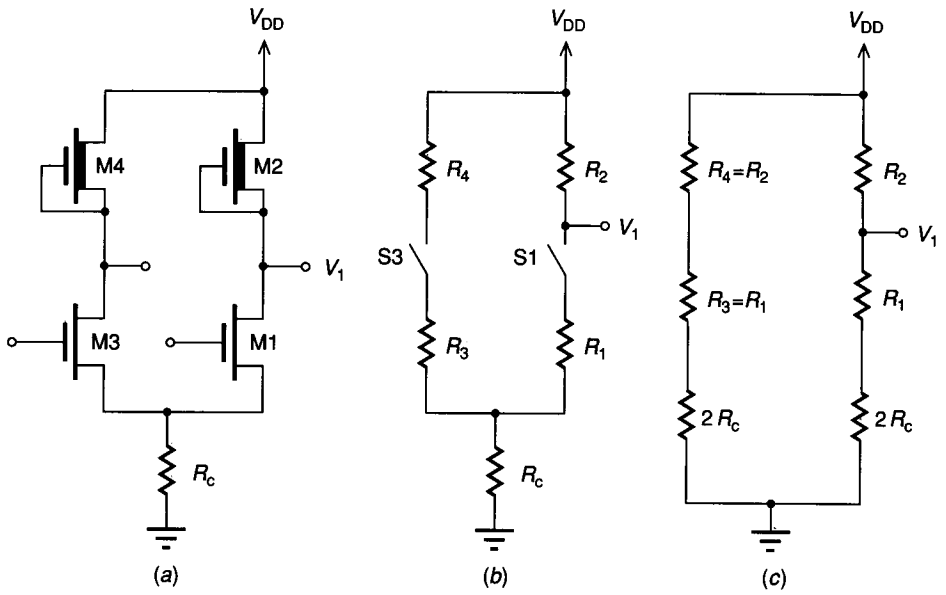


FIGURE 7.11-1
Resistive noise coupling: (*a*) Two inverters with common resistance to ground, (*b*) Resistive model for (*a*), (*c*) Model for matched inverters.

are identical and both switches are closed is given in Fig. 7.11-1*c*. Under these conditions, the output voltage V_1 is

$$V_1 = \frac{V_{DD}(R_1 + 2R_c)}{R_1 + R_2 + 2R_c} \tag{7.11-2}$$

The increase in output voltage depends on the relative resistance of R_c and the equivalent inverter output resistance.

The preceding example with simple inverters demonstrates resistive noise coupling, but in practice this case does not cause serious noise problems because logic gates are usually connected to low-resistance power and ground buses through relatively short paths. Resistance in a common supply path to two inverters is another source of resistive noise coupling. As a general rule, resistance in any segment of power and ground paths may cause unwanted noise voltages. For this reason, power and ground buses are almost always run in a low-resistance metal layer.

Although resistive noise coupling is not serious for the preceding simple case, two related conditions are important. If a large number of gates common to a power supply bus change state in concert, the combined effect may produce significant voltage shifts that detrimentally affect logic states. Second, output pad drivers and fast bus drivers draw considerable current and thus may generate noise voltages in other parts of the circuit from voltage drops even in a low-resistance power supply or ground path. The second condition is more easily detected than the first, but both are important considerations in integrated circuit design.

7.11.2 Capacitive Noise Coupling

Capacitive noise coupling is caused by physical relationships of signal paths. For integrated circuits, the two conditions that cause significant mutual capacitance are long adjacent signal paths, and overlapping but disjoint signal paths. In either case, the coupling capacitance will be called C_c. A general model of the coupling is given in Fig. 7.11-2*a*. (Standard Laplace transform notation is used in this analysis.) If the admittance to ground of the affected circuit at the point of coupling is $Y(s)$, and the effective noise signal is $V_s(s)$, the noise voltage $V_n(s)$ is given by

$$V_n(s) = \frac{V_s(s)sC_c}{Y(s) + sC_c} \tag{7.11-3}$$

Two cases will be considered to show the effects of capacitively coupled noise. First, the effect on the dynamic storage node of Fig. 7.11-2*b* will be considered. A dynamic storage node can be adequately represented as a capacitance C_d to ground with an admittance $Y_s(s) = sC_d$. From Eq. 7.11-3, the noise voltage is

$$V_n(s) = \frac{V_s(s)sC_c}{sC_d + sC_c}$$

or, after canceling the s term,

$$V_n = \frac{V_sC_c}{C_d + C_c} \tag{7.11-4}$$

The coupled noise voltage is just the noise voltage times the ratio of the coupling capacitor to the sum of the two capacitances. Because C_d is normally the small gate capacitance of a transistor, even a small coupling capacitance can be significant. If C_c is equal to C_d, a high-to-low voltage change in V_s can destroy

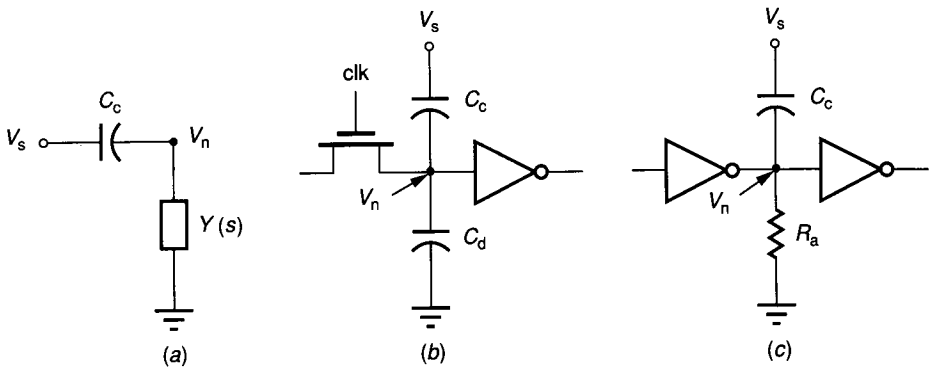


FIGURE 7.11-2
Capacitive noise coupling: (*a*) General noise model for capacitive noise coupling, (*b*) Dynamic storage node, (*c*) Actively driven node.

information on a dynamic storage node. It is important to keep the connection from the gating transistor to the dynamic storage node short and to ensure that no other signal lines cross the connection.

A second consideration is capacitive coupling to a driven logic node, as in Fig. 7.11-2*c*. Here the impedance to ground can be adequately modeled by the equivalent resistance R_a of the active output transistor of the first inverter. Again, from Eq. 7.11-3, the noise voltage is

$$V_n(s) = \frac{V_s(s)sC_c}{1/R_a + sC_c} = \frac{V_s(s)R_aC_cs}{1 + R_aC_cs} \tag{7.11-5}$$

For the normal condition of $|R_aC_cs| << 1$, Eq. 7.11-5 reduces to

$$V_n(s) = V_s(s)R_aC_cs \tag{7.11-6}$$

If $R_a = 10\ \text{k}\Omega$, $C_c = 0.1$ pF, and $f = 1.59$ MHz, then with $\omega = 2\pi f = 10$ Mrad/sec and $s = j\omega$

$$|V_n| = |V_s|/100$$

These values show that significant capacitive coupling is necessary to change the logic level for an actively driven node.

7.11.3 Definition of Noise Margins

Explanation of noise margins for digital logic gates requires the definition of several voltages in the voltage transfer characteristic of a logic circuit.

1. V_{IL}: the highest voltage reliably recognized as a logic low
2. V_{IH}: the lowest voltage reliably recognized as a logic high
3. V_{OL}: the nominal logic low voltage generated by a stage
4. V_{OH}: the nominal logic high voltage generated by a stage

V_{OL} and V_{OH} were introduced previously as V_L and V_H, the nominal output voltages of a reference inverter. V_{IL} and V_{IH} were introduced as the logic threshold voltages.

The voltages V_{IL} and V_{IH} require further explanation. As was shown in Sec. 7.2.2, the voltage transfer function of an inverter contains two regions with less-than-unity inverting gain separated by a region with greater-than-unity inverting gain. V_{IL} and V_{IH} are defined to denote the intersections of these regions, as shown in Fig. 7.11-3. The slope of the transfer characteristic is -1 at the two points representing an inverting gain of unity. Consider V_{IL} first. If the input voltage $V_i < V_{IL}$, the inverting gain of the inverter is less than unity and any small noise perturbation will be reduced in amplitude by the inverter. Otherwise, if $V_{IL} < V_i < V_M$, the input represents a low logic level, but any small noise perturbation is amplified as it is passed to the next stage. This condition may cause erroneous logic voltages to appear in subsequent stages. Thus, V_{IL} represents

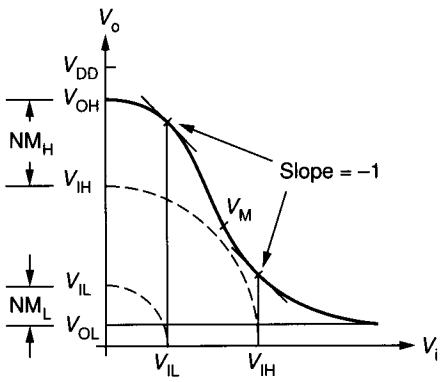


FIGURE 7.11-3
Generic dc voltage transfer characteristic for inverter.

the highest allowable input voltage level that is considered to reliably function as a logic low value. A similar analysis for V_{IH} reveals that it is the lowest allowable input voltage level that is considered to reliably function as a logic high value.

With suitable definitions for high and low logic voltages, the high and low noise margins can be defined. The low-level noise margin is

$$\text{NM}_\text{L} = V_{\text{IL}} - V_{\text{OL}} \tag{7.11-7}$$

and the high-level noise margin is

$$\text{NM}_\text{H} = V_{\text{OH}} - V_{\text{IH}} \tag{7.11-8}$$

NM_H and NM_L represent the largest noise voltages that can be sustained at an inverter input and still allow the inverter to function reliably. Noise margins for logic gates other than inverters are similarly defined based on their voltage transfer characteristics. Subsequent sections examine high and low noise margins for both NMOS and CMOS devices.

7.11.4 NMOS Noise Margins

The noise margins for an NMOS gate can be found from its voltage transfer characteristic. The voltages V_{OL} and V_{OH} were identified in Sec. 7.3 for the reference inverter as V_L and V_H. The depletion-load inverter and its voltage transfer characteristic of Fig. 7.11-4 will be used to help determine V_{IL} and V_{IH}. Note the straight lines that separate the ohmic and saturated regions for the enhancement and depletion transistors of the inverter. These help show that V_{IL} will be found with the depletion transistor in the ohmic region and the enhancement transistor in the saturation region. The voltage transfer characteristic in this region is given by

$$K'\frac{W_\text{N}}{L_\text{N}}\frac{(V_\text{i} - V_\text{TN})^2}{2} = K'\frac{W_\text{D}}{L_\text{D}}\left(-V_\text{TD} - \frac{V_\text{DD} - V_\text{o}}{2}\right)(V_\text{DD} - V_\text{o}) \tag{7.11-9}$$

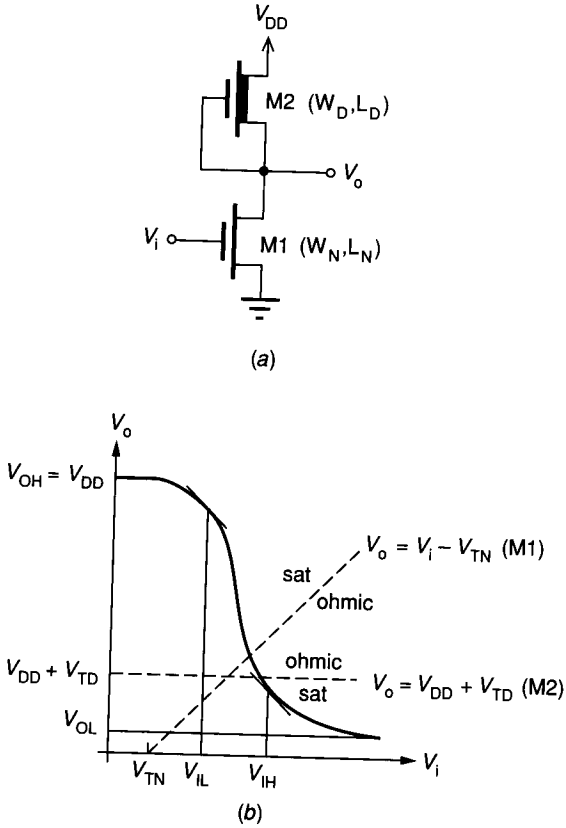


FIGURE 7.11-4
NMOS inverter noise model: (a) Circuit, (b) Voltage transfer characteristic.

Assuming an inverter with a pullup/pulldown ratio of k and solving Eq. 7.11-9 for V_i gives

$$V_i = V_{TN} + \sqrt{(1/k)(-2V_{TD} - V_{DD} + V_o)(V_{DD} - V_o)} \quad (7.11\text{-}10)$$

Taking the derivative with respect to V_o, setting it equal to -1, and solving for V_o gives

$$V_o = V_{DD} + V_{TD}\left(1 - \sqrt{k/(1+k)}\right) \quad (7.11\text{-}11)$$

The corresponding input voltage is

$$V_i = V_{IL} = V_{TN} - \frac{V_{TD}}{\sqrt{k(1+k)}} \quad (7.11\text{-}12)$$

Substituting typical values of $V_{TN} = 1$ V, $V_{TD} = -3.5$ V, $k = 4$, and $V_{DD} = 5$ V into Eqs. 7.11-11 and 7.11-12 gives

$$V_o = 4.63 \text{ V} \quad \text{and} \quad V_{IL} = 1.78 \text{ V} \quad (7.11\text{-}13)$$

To find V_{IH}, Fig. 7.11-4 indicates that the operating regions for the two transistors are reversed. The voltage transfer characteristic is now

$$K'\frac{W_N}{L_N}\left(V_i - V_{TN} - \frac{V_o}{2}\right)V_o = K'\frac{W_D}{L_D}\frac{V_{TD}^2}{2} \tag{7.11-14}$$

Solving Eq. 7.11-14 for V_i gives

$$V_i = V_{TN} + \frac{V_o}{2} + \frac{V_{TD}^2}{2kV_o} \tag{7.11-15}$$

Taking the derivative with respect to V_o, setting it equal to -1, and solving for V_o gives

$$V_o = \frac{-V_{TD}}{\sqrt{3k}} \tag{7.11-16}$$

The corresponding input voltage is

$$V_i = V_{IH} = V_{TN} - \frac{2V_{TD}}{\sqrt{3k}} \tag{7.11-17}$$

Substituting typical values into Eqs. 7.11-16 and 7.11-17 gives

$$V_o = 1.01 \text{ V and } V_{IH} = 3.02 \text{ V} \tag{7.11-18}$$

The noise margins can now be calculated. Using the results of Eqs. 7.11-13 and 7.11-18,

$$NM_L = V_{IL} - V_{OL} = 1.78 - 0.40 = 1.38 \text{ V} \tag{7.11-19}$$

and

$$NM_H = V_{OH} - V_{IH} = 5.00 - 3.02 = 1.98 \text{ V} \tag{7.11-20}$$

These noise margins are typical for NMOS logic gates, with the high noise margin greater than the low noise margin.

7.11.5 CMOS Noise Margins

The noise margins for a CMOS gate can be found from its voltage transfer function. The voltages $V_{OL} = 0$ V and $V_{OH} = V_{DD}$ were found in Sec. 7.5 for the CMOS reference inverter as V_L and V_H. The inverter and its voltage transfer characteristic of Fig. 7.11-5 will be used to help determine V_{IL} and V_{IH}. Note the straight lines that separate the ohmic and saturated regions for the n- and p-channel transistors of the inverter. These help show that V_{IL} will be found with the p-channel transistor in the ohmic region and the n-channel transistor in the saturation region. The voltage transfer characteristic in this region is given by

$$K_N'\frac{W_N}{L_N}\frac{(V_i - V_{TN})^2}{2} = K_P'\frac{W_P}{L_P}\left(V_i - V_{DD} - V_{TP} - \frac{V_o - V_{DD}}{2}\right)(V_o - V_{DD}) \tag{7.11-21}$$

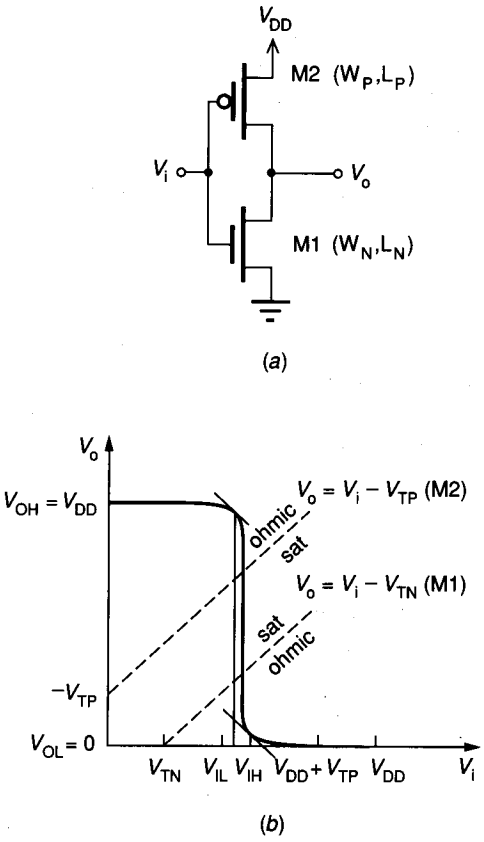


FIGURE 7.11-5
CMOS inverter noise model: (a) Circuit, (b) Voltage transfer characteristic.

Note that V_{TP} is negative for CMOS circuits. The substitution

$$x = \frac{K'_N}{K'_P}\frac{W_N}{L_N}\frac{L_P}{W_P}$$

and solving Eq. 7.11-21 for V_o gives

$$V_o = V_i - V_{TP} - \sqrt{(V_i - V_{TP})^2 - 2V_{DD}(V_i - V_{TP} - V_{DD}/2) - x(V_i - V_{TN})^2} \tag{7.11-22}$$

Letting $x = 1$ for symmetric output drive, taking the derivative with respect to V_i, setting it equal to -1, and solving for V_{IL} gives

$$V_{IL} = \frac{3V_{DD} + 3V_{TP} + 5V_{TN}}{8} \tag{7.11-23}$$

To find V_{IH}, Fig. 7.11-5 indicates that the operating regions for the two transistors are reversed. The voltage transfer characteristic is now

$$K'_N\frac{W_N}{L_N}\left(V_i - V_{TN} - \frac{V_o}{2}\right)V_o = K'_P\frac{W_P}{L_P}\frac{(V_i - V_{DD} - V_{TP})^2}{2} \tag{7.11-24}$$

Note that V_{TP} is negative for CMOS circuits. The substitution

$$y = \frac{K'_P}{K'_N}\frac{W_P}{L_P}\frac{L_N}{W_N}$$

and solving Eq. 7.11-24 for V_o gives

$$V_o = V_i - V_{TN} - \sqrt{(V_i - V_{TN})^2 - y(V_i - V_{DD} - V_{TP})^2} \tag{7.11-25}$$

Letting $y = 1$ for symmetric output drive, taking the derivative with respect to V_i, setting it equal to -1, and solving for V_{IH} gives

$$V_{IH} = \frac{5V_{DD} + 5V_{TP} + 3V_{TN}}{8} \tag{7.11-26}$$

The noise margins can now be calculated. Using typical values of $V_{TN} = 1$ V, $V_{TP} = -1$ V, $V_{DD} = 5$ V, and the results of Eqs. 7.11-23 and 7.11-26,

$$NM_L = V_{IL} - V_{OL} = 2.125 - 0 = 2.13 \text{ V} \tag{7.11-27}$$

and

$$NM_H = V_{OH} - V_{IH} = 5.00 - 2.875 = 2.13 \text{ V} \tag{7.11-28}$$

These noise margins are typical for CMOS logic gates. For symmetrical output drive, the noise margins are equal. Note that the sum of the CMOS noise margins (4.26 V) is much greater than the sum of the NMOS noise margins (3.36 V). CMOS logic is preferred in high-noise environments.

Primary causes of noise voltages in digital MOS circuits were described and analyzed in this section. Noise margins for digital circuits were defined, and these definitions were used to derive noise margins for both NMOS and CMOS inverters. Typical noise margins were calculated for each inverter type.

7.12 SUMMARY

Most digital designers accept a characterization of design abstraction that includes functional, register transfer level, logic, circuit, and layout descriptions as shown in Fig. 7.1-1 Practical digital design is sometimes accomplished by starting with cell layout and working up the abstraction hierarchy from layout to the functional level—often termed *bottom-up design*. At the other extreme, designs are defined at the functional level and then filled in with details at each lower level of abstraction until layout is created—termed *top-down design*. In fact, most design is accomplished using a combination of top-down and bottom-up design styles based on a prerequisite knowledge of basic digital building blocks.

The goal of this chapter has been to introduce basic digital building blocks at the logic, circuit, and layout levels. To accomplish this, salient characteristics of digital logic were described followed by explanations of NMOS and CMOS inverters, logic gates, and pass transistors. With knowledge of the primitive building blocks in hand, important electrical characteristics such as signal propagation delays, parasitic capacitance, power dissipation, and noise margins were detailed.

REFERENCES

1. M. Shahdad, R. Lipsett, E. Marschner, K. Sheehan, H. Cohen, R. Waxman, and D. Ackley: "VHSIC Hardware Description Language," *Computer*, vol. 18, no. 2, pp. 94–103, Feb. 1985.
2. Mario R. Barbacci: "Instruction Set Processor Specifications (ISPS): The Notation and Its Applications," *IEEE Trans. Comput.*, vol. C-30, no. 1, pp. 25–40, Jan. 1981.
3. W. I. Fletcher: *An Engineering Approach to Digital Design*, Prentice-Hall, Englewood Cliffs, N. J., 1980.
4. *ALS/AS Logic Circuits Data Book*, Texas Instruments, Dallas, 1983.
5. *High-Speed CMOS Logic Data Book*, Texas Instruments, Dallas, 1984.
6. F. J. Hill and G. R. Peterson: *Introduction to Switching Theory and Logical Design*, John Wiley & Sons, New York, 1981.
7. David A. Hodges and Horace G. Jackson, *Analysis and Design of Digital Integrated Circuits*, McGraw-Hill, New York, pp. 89–96, 1983.
8. Lance A. Glasser: "The Analog Behavior of Digital Integrated Circuits," *Proc. 18th Design Automation Conf.*, pp. 603–612, 1981.
9. S. M. Sze, Ed., *VLSI Technology*, McGraw-Hill, New York, 1983.

PROBLEMS

Section 7.1

7.1. Provide a multilevel description of a full adder having inputs A, B, and C and outputs Sum and Carry. Include a block diagram, an RTL-level description (use Boolean equations), and a logic diagram.

Section 7.2

7.2. Find V_L and V_H for the following piecewise-linear voltage transfer characteristic (VTC) of an inverter. Let the VTC intersect the points $(V_i, V_o) = (0,5.0)$, $(2.5,3.5)$, $(3.5,0.5)$, $(5.0,0.5)$.

7.3. What primary characteristic distinguishes restoring logic from nonrestoring logic?

7.4. Verify by using DeMorgan's theorem that the Boolean logic function $X = AB + AC + BC$ can be realized using only NAND logic gates. Show the NAND gate logic diagram for this function.

7.5. Assume that $V_{IL} = 1$ V, $V_{IH} = 4.5$ V, and $V_{DD} = 5$ V. Give the allowable voltage ranges for V_L and V_H for a practical logic device.

7.6. Given an inverter voltage transfer characteristic where $V_o = (2.5 + 2\cos(\pi V_i/5))$ V, find the values V_L, V_M, and V_H.

Section 7.3

7.7. Solve for the inverter ratio k with $V_{TD} = -4$ V, $V_{TN} = 1$ V, $V_{DD} = 5$ V, and $V_M = 2.5$ V. Assume both transistors are in saturation. Is this assumption valid?

7.8. If $V_{TD} = -3.5$ V, $V_{TN} = 1$ V, $V_{DD} = 5$ V, and $V_M = 2.5$ V, determine the inverter ratio k using the appropriate equations (ohmic or saturation) for both transistors of an inverter.

7.9. For a depletion-load inverter with $V_{TD} = -4$ V, $V_{TN} = 1$ V, $V_{DD} = 5$ V, and $k = 4$, calculate the value of V_L without the approximation of Eq. 7.3-6.

7.10. For a depletion-load inverter with $V_{TD} = -4$ V, $V_{TN} = 1$ V, $V_{DD} = 5$ V, and $k = 3$, calculate the approximate value of V_L.

7.11. Using an analysis similar to that used to derive Eq. 7.3-2, determine the sizing ratio for an enhancement-load inverter.

7.12. Assume that L and W are sized in integral multiples of dimension λ. Find values of L and W in terms of λ for both transistors of an inverter having a 4 : 1 pullup/pulldown ratio so that the total gate area is minimized.

Section 7.4

7.13. Calculate V_L for a two-input depletion-load NOR gate with one input high and with both inputs high. Assume $V_{TD} = -3.5$ V, $V_{TN} = 1$ V, $V_{DD} = 5$ V, and $k = 4$.

7.14. Assume that L and W are sized in integral multiples of dimension λ. Find values of L and W for all three transistors in a two-input depletion-load NAND gate to minimize the total gate area. Assume the pulldown transistors are sized equally and $k = 4$.

7.15. Analytically determine V_L for an n-input depletion-load NOR gate with all inputs high as n goes to infinity.

7.16. For a two-input NAND gate with $V_{TD} = -3.5$ V, $V_{TN} = 1$ V, and $V_{DD} = 5$ V, determine V_L if the pullup transistor is sized according to $k = 4$ instead of $k = 8$.

7.17. Show that a four-input NAND function can be realized with a multi-input NOR gate and inverters.

Section 7.5

7.18. Based on the VTC of Fig. 7.5-2, provide the VTC for a CMOS inverter pair consisting of two identical inverters.

7.19. Assume a 2 μ process with $K'_N = 2K'_P$ and a CMOS inverter with symmetrical output drive. Find the minimum L and W for the p-channel pullup if the n-channel pulldown is of minimum size. Next assume the p-channel pullup is of minimum size, and find the minimum L and W for the n-channel pulldown. The smallest transistor for a 2 μ process is 2 μ by 2 μ.

7.20. Assume the transistors are in the ohmic region with V_{DS} near 0 V. Determine the constant of proportionality for Eq. 7.5-2 and for Eq. 7.5-3.

Section 7.6

7.21. For a NOR gate with $K'_N = 2K'_P$ and with both n-channel pulldown transistors set to minimum size, find the L and W of equally sized p-channel pullups for symmetrical drive.

7.22. Assuming that $K'_N = 2K'_P$, compare the minimum total gate area in terms of dimension λ for a two-input NOR and a two-input NAND gate in CMOS if approximately symmetrical output drive is required for the worst-case condition (series resistance with all transistors on equal to parallel resistance with all but one transistor off).

7.23. Determine the transistor sizes for a three-input CMOS NAND gate with minimum-size pulldown transistors if symmetric output drive is desired for the worst-case pullup and pulldown conditions. Determine the maximum and minimum resistance considering all possible input combinations.

7.24. Compare the minimum total gate area for a four-input NAND gate in both NMOS and CMOS if $K'_N = 2K'_P$, using proper pullup/pulldown ratios for the NMOS gate.

Section 7.7

7.25. Assume a series string of NMOS pass transistors having a lumped-parameter model, as in Fig. 7.7-3, with $R = 10$ kΩ and $C = 0.1$ pF. Calculate the time constant for a single stage and for 10 series stages using the square-law delay approximation.

7.26. If five NMOS pass transistors are cascaded source to gate, find the highest output voltage from the fifth pass transistor if the gate of the first pass transistor is tied to 5 V. Assume $V_{TN} = 1$ V.

7.27. Consider a 4-to-1 selector circuit with select inputs a and b. Compare a solution using standard NMOS logic gates to the pass-transistor solution of Fig. 7.7-2. Give the number of transistors required for each solution.

7.28. Does a series cascade of CMOS transmission gates have the same square-law delay characteristic as a string of NMOS pass transistors? Why or why not?

Section 7.8

7.29. For $L = W = 2\ \mu$, $K' = 25\ \mu A/V^2$, $V_{TN} = 1$ V, and $V_{DD} = 5$ V, find R_{ss} from Eq. 7.8-3 and the exact value of R_1 from Eq. 7.8-4. If $V_{TD} = -4$ V and $k = 4$, find the value of R_1 from Eq. 7.8-5.

7.30. Verify that the 10%-to-90% rising transition time for the exponential voltage $V = 1 - e^{-t/\tau}$ is 2.2τ.

7.31. For the parameters of Prob. 7.29, find t_{LH} and t_{HL} if $C_{ox} = 0.75\ fF/\mu^2$.

7.32. Find the process characteristic time constant for the parameters of Prob. 7.29 if $C_{ox} = 0.75\ fF/\mu^2$.

7.33. With the parameters used to derive the process characteristic time constant for the preceding problem, use SPICE to calculate the inverter-pair delay by simulation.

7.34. If the dimensions and voltages of Prob. 7.29 are scaled down by a factor of 2, calculate the resulting process characteristic time constant and compare it with the unscaled value (remember to scale the gate oxide thickness).

7.35. Compare the current for an enhancement transistor with $V_{GS} = V_{DS} = V_{DD}$ to the current for an equally sized depletion transistor with $V_{GS} = 0$ and $V_{DS} = V_{DD}$. Now calculate the depletion transistor current with $V_{GS} = V_{DD}$. Use the parameters of Prob. 7.29

7.36. Calculate t_{ipd} for a three-input NOR gate, and compare it with t_{ipd} for a three-input NAND gate in NMOS with typical parameter values of Prob 7.29.

7.37. Compute the CMOS t_{ipd} for Ex. 7.8-1 if $K'_P = 15\ \mu A/V^2$ and the p-channel transistor is sized for symmetrical drive.

Section 7.9

7.38. If the process characteristic time constant is $\tau_p = 0.01$ ns and $C_G = 0.1$ pF for an NMOS inverter with $k = 4$, calculate the average delay to drive a 10 pF load.

7.39. Assume that a logic signal drives the root of a binary tree that has 16 outputs, each driving a single reference inverter. Calculate the path delay from the root to the load in terms of t_{apd}. Now calculate the delay if the same logic signal drives the 16 loads directly.

7.40. Repeat the preceding problem, but include the effects of interconnection loading by assuming that the interconnection capacitance for each driven node is twice the gate load capacitance.

7.41. For a geometrically sized cascade of four inverters with each inverter having twice the drive of the preceding stage, compute the average propagation delay in terms of t_{apd} for each of the four stages. Assume the final load is geometrically sized.

7.42. Calculate t_{dir} versus t_{cas} in terms of t_{apd} for a load equivalent to 500 reference inverters when driven (*a*) directly, (*b*) by an optimal cascade (calculate n and truncate to an integer), and (*c*) by a cascade of two drive stages.

Section 7.10

7.43. If $\theta_{JA} = 20°C/W$ because of special packaging, the maximum junction temperature is $T_J = 150°C$, and the ambient temperature is $T_A = 30°C$, find the allowable power dissipation for the integrated circuit.

7.44. Calculate the static power dissipation for an NMOS inverter in the on state with $k = 5$, $K' = 40\ \mu A/V^2$, and $V_{TD} = -4$ V. As long as the low logic voltage is below $V_{DD} + V_{TD}$, does the size of the enhancement pulldown affect the power dissipation? Why or why not?

7.45. Based on Eq. 7.10-6, what is the maximum instantaneous power dissipated in the inverter of the preceding problem? Assume $V_{TN} = 1$ V.

7.46. Show that the energy dissipated in the resistor of the circuit of Fig. 7.10-2*b* is $CV^2/2$ during the capacitor discharge.

7.47. A new CMOS microprocessor is advertised as having a complexity of 300,000 transistors, uses a 20 MHz clock, and operates from a 5 V supply. Assume the microprocessor is composed entirely of gates with an average of five transistors per gate and that the average gate loading capacitance is 0.1 pF. Calculate the dynamic power dissipated by this chip. Is this a reasonable value? If not, which assumptions of the problem are probably incorrect? Explain.

Section 7.11

7.48. For the circuit of Fig. 7.11-1, assume that $R_1 = 10\ k\Omega$, $R_2 = 40\ k\Omega$, and $R_c = 5\ k\Omega$. Calculate the voltage V_1 (*a*) when only switch S1 is closed, and (*b*) when both switches S1 and S3 are closed, if $V_{DD} = 5$ V.

7.49. A dynamic storage node has capacitance $C_n = 0.2$ pF and initial voltage $V_n = 4$ V. If a nearby signal line is coupled to the storage node by capacitance $C_c = 0.1$ pF, find the voltage on V_n after the signal line switches from 5 V to 1 V.

7.50. For $V_{TD} = -4$ V, $V_{TN} = 1$ V, $V_{DD} = 5$ V, and $k = 5$, find V_{IL}, V_{OL}, V_{IH}, V_{OH}, NM_L, and NM_H.

7.51. For an NMOS inverter, derive a simplified expression for V_{OL} based on V_{DD}, V_{TD}, V_{TN}, and k.

7.52. How do NM_L and NM_H behave as k increases without limit?

7.53. Verify Eqs. 7.11-11 and 7.11-12.

7.54. Verify Eqs. 7.11-16 and 7.11-17.

7.55. For a CMOS inverter, give the expressions to calculate NM_L and NM_H directly in terms of V_{DD}, V_{TN}, and V_{TP}.

7.56. For $V_{DD} = 5$ V, calculate the noise margins for a CMOS inverter under the following conditions

(*a*) $V_{TN} = 0.5$ V, $V_{TP} = -0.5$ V
(*b*) $V_{TN} = 1.5$ V, $V_{TP} = -1.5$ V
(*c*) $V_{TN} = 0.5$ V, $V_{TP} = -1.5$ V