# CHAPTER

# 3

# The CMOS Inverter

## Chapter Objectives

- Review MOSFET device structure and basic operation.
- Establish device models for MOSFET that are used for circuit-level design. Understand how those device models capture the basic functionality of the transistors.
- Analyze a static CMOS inverter circuit, including its voltage transfer function and transient performance.
- Apply the device models to compute both static and dynamic parameters for a static CMOS inverter using hand calculations.
- Understand the sources of power dissipation in the static CMOS inverter and compute power consumption.

## 3-1 CMOS Devices

In the last chapter, we enjoyed designing logic gates with perfect transfer characteristics and instantaneous transitions. While those switch-based gates are useful for learning how logical gates are implemented physically, they regrettably remain unrealizable in practice. Since we must use real transistors to implement these switches, the nonidealities of those transistors prevent us from creating ideal gates. Nevertheless, the transistors do approximate the ideal switch behavior, so transistor-based logic does provide a robust mechanism for digital computation.

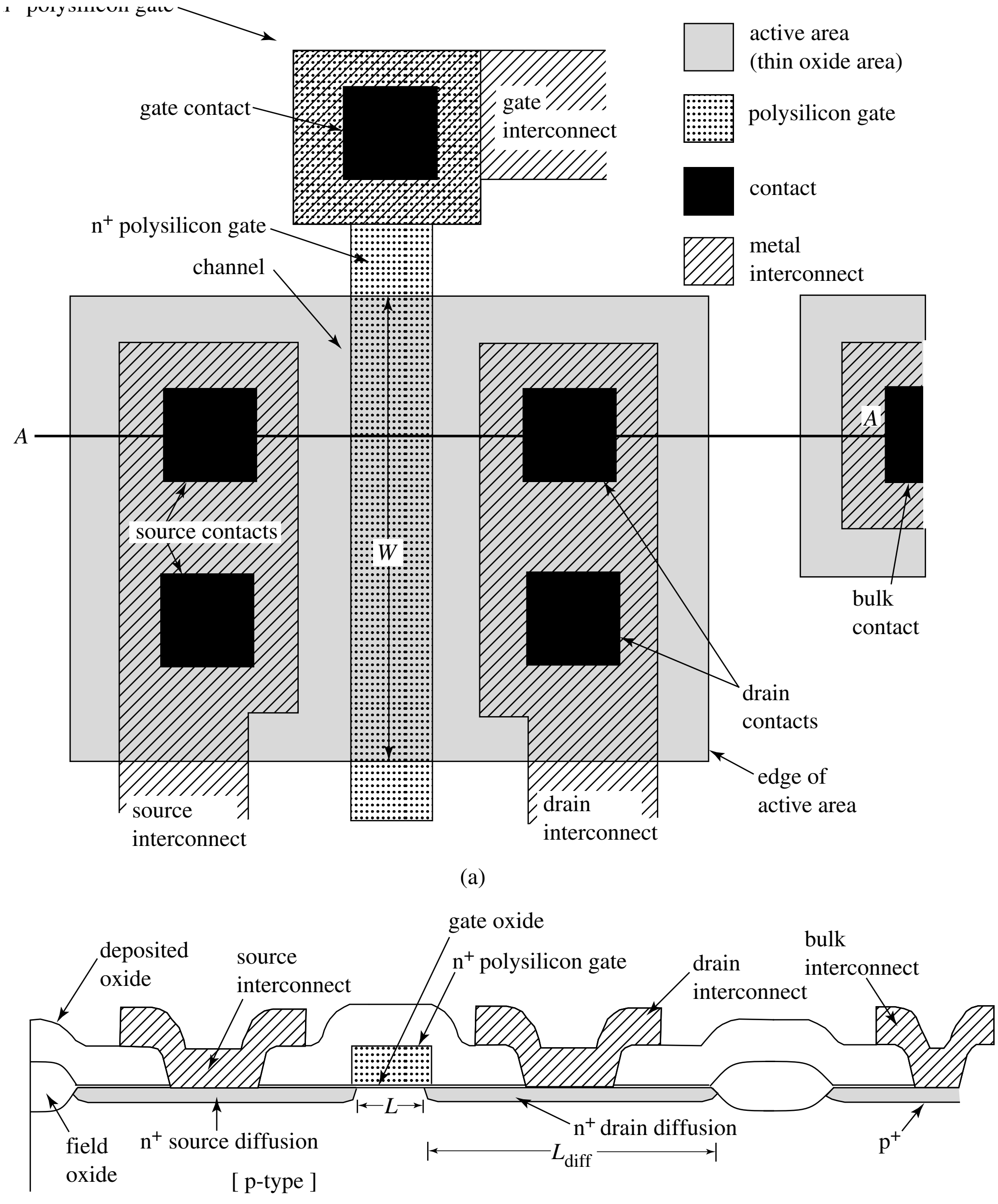
This chapter takes the next step into the circuit domain where transistors serve as switches. This allows us to develop a more realistic view of circuit behavior. We first examine the transistor devices themselves and then look at how their behavior is modeled for circuit analysis. Then we use the transistor device models to take a deep look at the CMOS inverter.

### 3-1-1 MOSFET Physical Structure

The **MOS field-effect transistor (MOSFET)** is the electronic device behind the explosive growth in digital electronics since 1970. We can deal with a MOSFET at several layers of hierarchy. We are most interested in this book with the circuit view of a MOSFET, so we will forgo a lengthy discussion of the underlying device physics. However, circuit designers do deal with more than just circuit schematics; they also must create the *layout* for their circuits. Layout refers to the drawing that represents how a circuit will be physically constructed. It usually consists of many overlapping polygons drawn in different colors and patterns. Each color/pattern combination represents a certain layer that will be fabricated eventually on the IC. Layout is thus a birds-eye view of these layers. While we will not delve too deeply into the device physics of the MOSFET, we will give a preview of the MOSFET structure by looking at its layout before we get to the device models for MOS transistors.

The layout and cross section of a modern n-channel MOSFET are shown in Figure 3–1(a) and (b), respectively. Again, the layout view shows the transistor from above. The **gate** of the transistor is a film of n$^+$ polycrystalline silicon (polysilicon or poly). Underlying the gate is the **gate oxide**, which is a layer of thermally grown silicon dioxide ($SiO_2$). The gate oxide is visible in Figure 3–1(b), but it is obscured by the poly in the layout view of Figure 3–1(a).

The MOSFET **channel** is formed under the polysilicon gate between the two n$^+$ regions, as shown in Figure 3–1(b). The channel is the location of all of the important



(a)

(b)

**Figure 3–1:** (a) Four-mask layout and (b) cross section of an integrated n-channel MOSFET.

action in a MOSFET. It appears in the layout as the region where the poly overlaps the diffusion. The gate oxide is not drawn in the layout but is instead inferred to exist over this channel region. The **channel length *L*** of the MOSFET is defined in Figure 3–1(b) as the gap between the $n^+$ diffusions or as the horizontal width of the poly over the diffusion in the layout. On the layout in Figure 3–1(a), the **channel width *W*** of the MOSFET is defined as the width of the active area overlaid by the poly. One $n^+$ diffusion is called the **source**, the other is called the **drain**. Since the source and drain are physically indistinguishable due to the MOSFET's symmetry, the potentials on the diffusions will be the basis for differentiating the source from the drain, as discussed in the next section. The source and drain diffusions have a length $L_{\text{diff}}$ and the same width $W$ as the channel in this transistor. Metal interconnections make ohmic contacts to the gate, source, and drain. In addition, the underlying p-type bulk region (the substrate or a deep p-type diffusion) is contacted with a fourth metal interconnection called the **bulk**. This substrate connection often is called a *tap*. Although the body connection is shown near the MOSFET in Figure 3–1, it does not necessarily need to be situated immediately next to the MOSFET, nor is there always a specific body connection associated with every transistor. Instead, the substrate is connected to the correct potential by multiple body (or well) taps that provide the proper body voltage for all of the transistors that share that substrate. Since pMOS devices, with $p^+$ source and drain diffusion, must reside in n-type silicon, they are placed in an n-well. The n-well is simply a deep (relative to the source/drain diffusion) region of n-type material that is large enough to accomodate the pMOS device(s) inside.

The complexity and capability of MOS integrated circuits have increased many orders of magnitude over the past several decades. Many of these advances resulted from the *scaling* (or shrinking) of CMOS technology. Throughout most of the history of CMOS scaling, the basic MOSFET has remained essentially the same. Modern MOSFETs, however, have adopted more complicated physical structures in response to emerging changes to traditional circuit-level characteristics. Since these changes result for small devices, they often are called *short-channel effects* or *deep-submicron effects* in reference to the short channel length of deeply scaled devices. In this book, we will use conventional models of *long-channel* transistors that do not include the short-channel effects. A deep understanding of the fundamental long-channel MOSFET will provide you with a firm foundation for incorporating short-channel behavior during additional study.

### 3-1-2 MOSFET Circuit Symbol and Terminal Characteristics

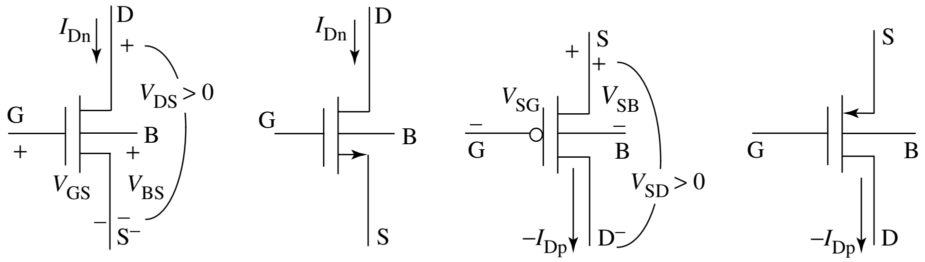
Before introducing the current–voltage characteristics of the MOSFET, it is helpful to have a symbol for this device. There are two common symbols for each MOSFET type. Figure 3–2 shows the circuit symbols and basic structure for an **n-channel MOSFET** or **nMOS** (shown in Figure 3–1) and for the complementary **p-channel MOSFET** or **pMOS** that has $p^+$ source and drain diffusions in an n-type bulk region. In this text, we will use the symbols without arrows for digital circuits. The cross sections for the two MOSFETS have been rotated 90° in order to illustrate the close correspondence between the symbols and the physical structure.

The MOSFETs are oriented as they often appear in circuit schematics: The drain for the n-channel MOSFET and the source for the p-channel MOSFET are on top—toward the positive voltage supply. Voltages between terminals are labeled by an ordered pair of subscripts, such as $V_{GS} = V_G - V_S$. By IEEE convention, the reference directions for currents are defined as positive *into* the device terminals. Since the drain current for the p-channel device is negative, it is convenient to use $-I_{Dp} > 0$ (which is positive) leaving the drain terminal.

As we mentioned before, the MOSFET is an inherently symmetrical device. It is therefore impossible to distinguish between source and drain based on physical definitions (e.g., the source and drain are interchangeable). We therefore use the potentials at the sides of the MOSFET channel as a basis for defininining the source and drain. For the nMOS in Figure 3–2(a), the drain and source terminals are selected so that $V_{DS} > 0$, which is equivalent to defining the source as the terminal with the *lower* potential. As shown in Figure 3–2(b), the labeling convention is opposite that for the p-channel MOSFET. Its source and drain are identified so that $V_{SD}>0$, which means its source is defined as the terminal with the *higher* potential.

The n-channel MOSFET's static (DC) terminal characteristics can be measured with voltage sources and ammeters. Since the MOSFET is a four-terminal device, we need to experiment and discover which voltages are most important. Let us assume that $V_{BS} = 0$ V. Since the voltage *difference*s between the drain and source and the gate and source are most important, we ground the source ($V_S = 0$ V) for convenience. Since the gate terminal is insulated, the only current of interest is that into the drain terminal $I_{Dn.}$

Figure 3–3(a) is a schematic of the circuit used to find the drain current's functional dependence on the two volt-



(a) n-channel MOSFET

(b) p-channel MOSFET

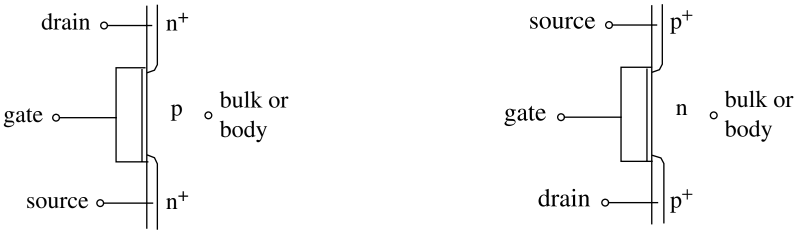


**Figure 3–2:** (a) -channel MOSFET symbols and structure and (b) p-channel symbols and structure.

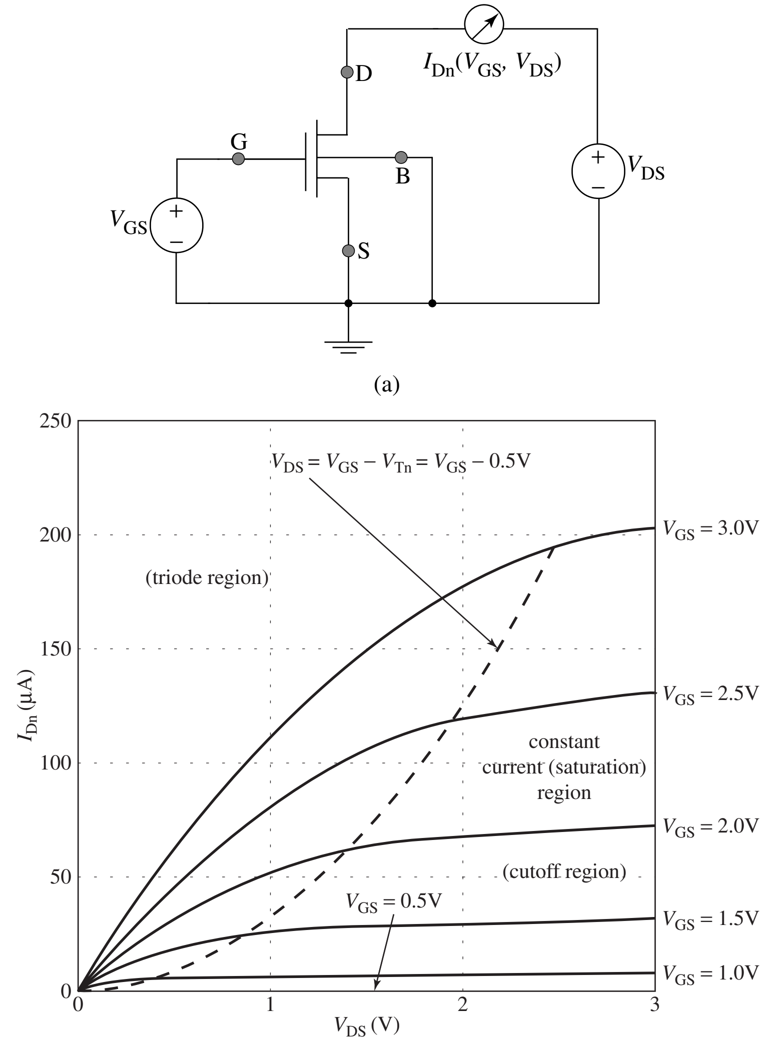
ages under our control: the gate–source voltage $V_{GS}$ and the drain–source voltage $V_{DS}$. The conventional way to graph the measured data is to plot a family of curves $I_{Dn}$ $(V_{GS}, V_{DS})$, with $V_{GS}$ as a parameter. The drain-current versus drain–source voltage curves for a selected set of gate–source voltages are called the MOSFET **drain characteristics**.

Figure 3–3(b) shows simulated drain characteristics for an n-channel MOSFET in a digital CMOS process. Although the voltage and current ranges can vary with the technology and device layout, the ranges in Figure 3–3(b) are typical for a 1 μ m technology (e.g., L = 1 μ m). In this graph, the drain current $I_D$ is measured as a function of drain–source voltage $V_{DS}$ for seven gate–source voltages.

Several observations can be made from studying the drain characteristics in Figure 3–3(b).

1. The MOSFET is **cutoff** ($I_{Dn} = 0$ A) for gate–source voltages that are less than some critical voltage, which is defined as the **threshold voltage** $V_{Tn}$ of the n-channel device. A typical value for the n-channel threshold voltage in a long-channel transistor is $V_{Tn} = 0.5$ V (modern devices may have lower threshold voltages). The $n^+$ source and drain are isolated electrically until sufficient voltage is applied to the gate to create an n-type inversion layer or **channel** between them.
2. The drain current is nearly independent of the drain–source voltage once $V_{DS} > V_{GS} - V_{Tn}$. In this region of operation (called **saturation)** the MOSFET behaves like a current source. Further study reveals that the drain current in saturation is proportional to the square of the gate–source voltage above threshold $(V_{GS} - V_{Tn})^2$. This can be seen from the ratios of drain current for $V_{GS} = 3$ V and for $V_{GS} = 2$ V. This constant-current behavior is exploited in both analog and digital circuit design.
3. The region where $I_{Dn}$ depends on both $V_{GS}$ and $V_{DS}$ is termed the **triode**, or linear region.

The p-channel MOSFET's drain characteristics are shown in Figure 3–4. In this case, the source and n-type bulk terminals are both connected to a 3 V supply. The effect of varying the drain voltage and gate voltage on the drain current can be investigated. The shape of the drain characteristics are identical to those of the n-channel MOSFET if $-I_{Dp}$ (which is positive) is plotted against the source–drain drop $V_{SD} = 3\ \text{V} - V_D$ as a function of the source–gate voltage $V_{SG} = 3\ \text{V} - V_G$. Notice that the subscripts on those voltages are backwards relative to the nMOS, so that $V_{SG} = V_S - V_G$. Do not let those reversed subscripts or negative signs fool you. They have to be that way to account for the fact that the pMOS source is at a higher potential (whereas nMOS is at a lower potential) and current flows out of the pMOS drain (and into the nMOS drain). A little practice helps to sort this all out. A



(a)

(b)

**Figure 3–3:** n-channel MOSFET drain characteristics: (a) test circuit and (b) measurements of drain current as function of drain–source voltage, with gate–source voltage as a parameter for a typical integrated device. $V_{GS} = 0, 0.5, 1, 1.5, 2, 2.5, 3$ V, for $0\text{ V} < V_{DS} < 3$ V.

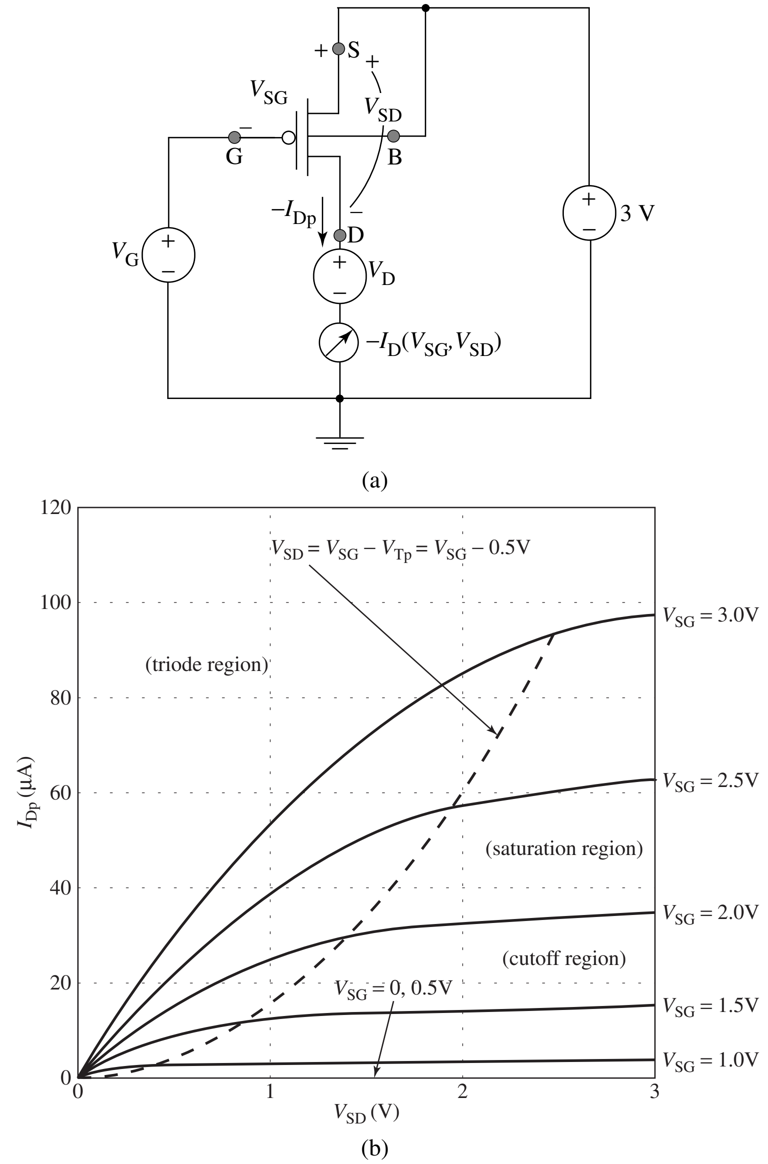
typical value for the p-channel threshold voltage in a long-channel technology is $V_{Tp} = -0.5$ V.

## 3-2 MOSFET Device Models for Use in Circuit Design

In the previous section, we developed the framework for device models for the MOSFET by showing the MOSFET structure and circuit symbols. In this section, we will present the device models that describe the current–voltage behavior of the MOSFETs, which we can use for circuit design. We will start by deriving the n-channel MOSFET device models and then adapt the analysis to the p-channel MOSFET.

### 3-2-1 Device Models for nMOSFETs

The drain-current equations for the cutoff, triode, and saturation regions of operation for an n-channel MOSFET are given in Eqs. (3.1) through (3.3).



(a)

(b)

**Figure 3–4:** (a) Test circuit and (b) measured drain characteristics of a p-channel MOSFET with the same geometric dimensions as the n-channel device used for Figure 4.3. The negative of the drain current $-I_{Dp}$ (> 0) is plotted against the source–drain drop $V_{SD}$, with the source–gate drop $V_{SG}$ as a parameter. The drain current for the p-channel is about half that of the n-channel device for identical geometries and complementary threshold voltages $V_{Tp} = -V_{Tn} = -0.5$ V.

For $V_{GS} \le V_{Tn}$ the nMOS is in cutoff, so its current is

$$I_D = 0 \text{ A} \tag{3.1}$$

For $(V_{GS} \ge V_{Tn}, V_{DS} \le V_{GS} - V_{Tn})$, the nMOS is in the triode region with current:

$$I_D = (W/L)\mu_n C_{ox} V_{DS}[V_{GS} - V_{Tn} - (V_{DS}/2)](1 + \lambda_n V_{DS}) \tag{3.2}$$

For $(V_{GS} \ge V_{Tn}, V_{DS} \ge V_{GS} - V_{Tn})$, the nMOS is in saturation:

$$I_D = (W/2L)\ \mu_n C_{ox}(V_{GS} - V_{Tn})^2(1 + \lambda_n V_{DS}) \tag{3.3}$$

The equations match the behavior that we observed in the nMOS drain characteristics in Figure 3–3. For the cutoff region, when the gate–source voltage is below the threshold voltage, the device conducts essentially zero current, as in Eq. (3.1). As the gate–source voltage rises, carriers in the p-type region under the gate oxide are repelled from the gate, creating a depletion region under the gate. Once the gate–source voltage exceeds the threshold voltage, the depletion region ceases to widen, and electrons from the source enter the channel region, producing a very thin n-layer at the Si-$SiO_2$ interface (which is called the channel). This channel acts as a conductor between the n-type source and drain terminals, and the voltage from the drain–source creates an electric field in order to drift the electrons through the channel. In the triode region, the current increases with the drain–source voltage according to Eq. (3.2). Once the drain-to-source voltage exceeds the gate overdrive, $V_{GS}$–$V_{Tn}$, the channel becomes pinched off at the drain end, so current becomes nearly constant with additional $V_{DS}$. This voltage is called $V_{DS_{SAT}}$, where $V_{DS} = V_{GS} - V_{Tn}$. Equation (3.3) captures the behavior of the current in saturation. The factor $(1 + \lambda_n V_{DS})$ in Eq. (3.3) models the remaining dependence of current on $V_{DS}$ in saturation, which results from channel-length modulation. Channel-length modulation refers to the movement of the pinch-off point near the drain, causing a change in the effective channel length. This description of the physics underlying the MOSFET current equations is about as brief as possible. For further investigation of the device physics, read module V.

The channel-length modulation term is included in Eq. (3.2) to prevent a discontinuity in the drain current when $V_{DS} = V_{DS_{SAT}} = V_{GS} - V_{Tn}$. Channel-length modulation can be neglected in the triode region for hand calculations, since the drain voltage $V$ is relatively small and the effect has no practical significance.

Since the bulk voltage does not appear explicitly in these equations, it is easy to forget that the nMOS transistor has a fourth terminal. The drain current actually does depend on the bulk voltage, but indirectly through the threshold voltage. That is a higher-order effect that we will not describe in this book. In this text, we will assume that $V_{Tn}$ is a technology-dependent constant and that $V_B$ for any nMOS device is tied to $V_{SS}$. Since $V_{SS}$ is the lowest voltage in the circuit, this ensures that the p and n-diodes between the bulk and source/drain never are forward biased (which could cause damage due to large currents).

Of the various terms in Eqs. (3.1) through (3.3), the gate oxide capacitance $C_{ox}$ can be found directly from measurements of the gate-oxide thickness $t_{ox}$. The other terms cannot be "measured" directly, but experimental values for $V_{Tn}$, $\mu_n$, and $\lambda_n$ can be derived by extracting the parameters from the measured drain characteristics. Although the equations are to some extent simplifications of reality, their grounding in the physics of the MOSFET provides us with insight into how it will behave under different scenarios. They are therefore a significant improvement over an alternative fitted curve (e.g., polynomial fit).

At first glance, the ratio $W/L$ in the device current equations can be read directly from the device layout. However, the channel length often is designed as the minimum allowed dimension in the fabrication technology. Since the channel length is in the denominator, small errors in $L$ can lead to substantial errors in $W/L$. Figure 3–5 illustrates the need to be precise in specifying the channel length for the MOSFET. The "as drawn" channel length $L_{mask}$ on the layout becomes the "as etched" length $L_{gate}$ of the polysilicon gate due to systematic lithography and etching effects. Finally, the ion-implanted source and drain regions diffuse under the polysilicon gate by a distance $L_D$, as shown in Figure 3–5. The actual channel length, as defined earlier in Figure 3–1, is given by

$$L = L_{gate} - 2L_D \tag{3.4}$$

So far, we have treated the gate–source voltage of the n-channel MOSFET as a parameter in the other equations. We can get a feel for how this parameter affects the device current by sweeping it at a given $V_{DS}$. Suppose we wish to plot $I_D$ as a function of $V_{GS}$ for $V_{GS}$ ranging from 0 to 3 V for the n-channel MOSFET. The circuit configuration for this measurement is shown in Figure 3–6. We will assume that $\mu_n = 215\ \text{cm}^2\text{V}^{-1}\text{s}^{-1}$, $t_{ox} = 150$ Å, $V_{Tn} = 0.5$ V, $L = 1\ \mu\text{m}$, and $W = 30\ \mu\text{m}$. Figure 3–7 shows the plot of the

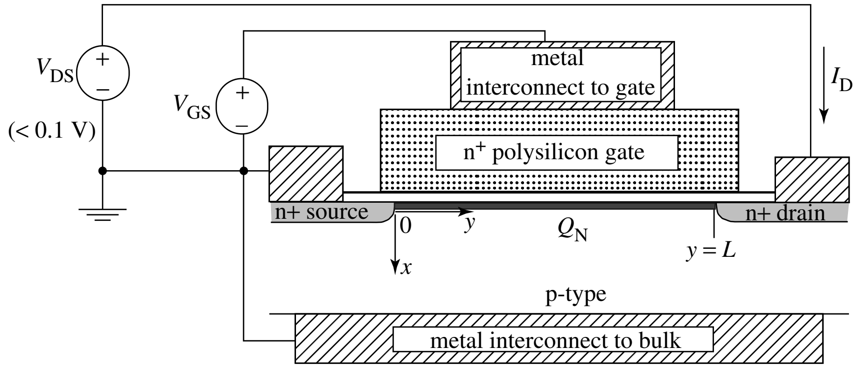


**Figure 3–5:** Definitions of the as-drawn gate length $L_{mask}$ on the mask, the "as-etched" polysilicon gate length $L_{gate}$, and the channel length $L = L_{gate} - 2L_D$ .

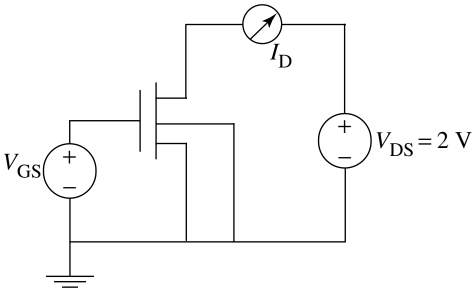


**Figure 3–6:** n-channel MOSFET circuit for current measurements.

equations with these parameters. The changing current clearly displays the behavior that we would expect based on the current equations. When $V_{GS}$ is less than $V_{Tn}$= 0.5V, there is no current. As it rises above $V_{GS}$, the nMOS enters saturation, because $V_{DS}$ (2 V) is larger than $V_{GS}$–$V_{Tn}$. The current increases quadratically with the turn on voltage $V_{GS}$–$V_{Tn}$, as the equation states. Finally, when $V_{GS}$–$V_{Tn}$ exceeds $V_{DS}$, the nMOS enters the triode region, and its current increases linearly with $V_{GS}$.

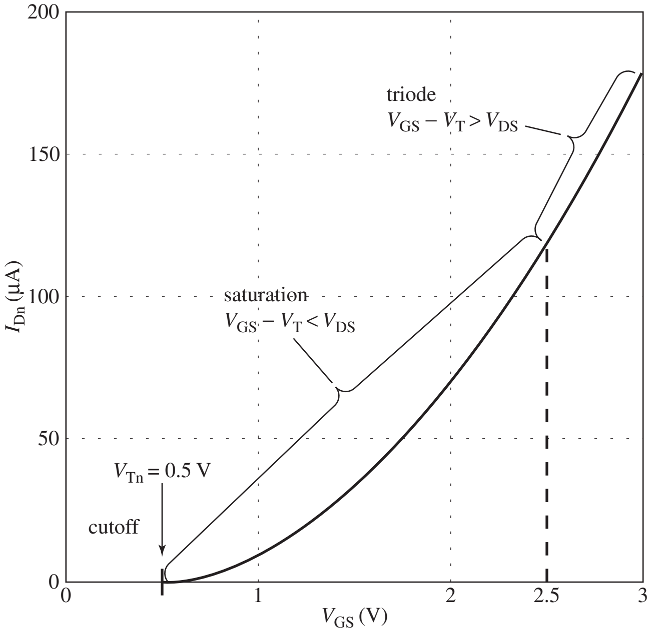


**Figure 3–7:** Plot of the $I_D$ versus $V_{GS}$ for a n-channel MOSFET with fixed $V_{DS}$ at 2 V.

While Figure 3–7 shows that the n-channel MOSFET definitely cannot be called an ideal switch, it does demonstrate switch-like behavior. When the input $V_{GS}$ to the switch stays below the threshold voltage, the transistor conducts no current. Once the input rises above $V_{Tn}$, the transistor begins to conduct current. It clearly cannot provide perfect conductivity (infinite current or 0 resistance) like an ideal switch, but it does at least behave functionally in the same fashion. After we look at device models for the pMOS transistor, we will use this switching behavior of the transistors to construct a realizable inverter using MOSFETs.

### 3-2-2 Device Models for pMOSFETs for Use in Circuit Design

In a CMOS technology, both n- and p-channel MOSFETs are fabricated on the same substrate, which is usually p-type. The pMOS transistors are placed in a well of n-type (n-well) silicon. The layout of a pMOS looks essentially the same as an nMOS layout, except that the source/drain diffusion sits in the n-well and is p-type. The tap ($n^+$) connects to the n-well, which is the body terminal of the pMOS transistor.

As described before, the pMOSFET behaves as something of an upside-down version of the nMOSFET. Since its source is at a higher potential than its drain, all of the reference voltages are negative (e.g., $V_{GS}$ and $V_{DS}$ are negative). Likewise, pMOS device thresholds are negative, including $V_{Tp}$. There is a straightforward way to alter the nMOS equations to use these negative values. First, switch the order of all of the subscripts (e.g., GS becomes SG, DS becomes SD, and SB becomes BS). Second, replace $V_{Tn}$ with $-V_{Tp}$. Finally, replace $I_D$ with $-I_D$. The current equations for the pMOS then changes:

For ($V_{SG} \le -V_{Tp}$), the pMOS is in cutoff, so its current is

$$I_D = 0 \text{ A} \tag{3.5}$$

For $(V_{SG} \ge -V_{Tp}, V_{SD} \le V_{SG} + V_{Tp})$, the pMOS is in triode:

$$-I_D = (W/L)\mu_p C_{ox}[V_{SG} + V_{Tp} - (V_{SD}/2)](1 + \lambda_p V_{SD})V_{SD} \tag{3.6}$$

For $(V_{SG} \ge -V_{Tp}, V_{SD} \ge V_{SG} + V_{Tp})$, the pMOS is in saturation:

$$-I_D = (W/2L)\ \mu_p C_{ox}(V_{SG} + V_{Tp})^2(1 + \lambda_p V_{SD}) \tag{3.7}$$

In these equations, $V_{Tp}$ is negative, unlike for nMOS. All of the other parameters have the same sign as for nMOS. Thus, $\mu_p$, $C_{ox}$, and $\lambda_p$ are positive. As with the nMOS, we will assume that $V_{Tp}$ is a technology constant. We also will assume that $V_B$ for a pMOS is always tied to $V_{DD}$, which prevents the p and n diodes between the source/drain and bulk to be forward biased.

These equations easily can yield mistakes if you misuse the negative signs. Notice that switching the subscripts simply makes the reference voltages positive, and the parameter substitutions ensure that the final current value is positive on the right side of the current equations. This is why $I_D$ needs to become negative on the left side to indicate that it flows out of the drain of the pMOS. This is equivalent to simply treating *all* of the pMOS reference voltages and parameters as though they were nMOS values (e.g., by taking absolute value of the reference voltages) and using them in the nMOS, as in Eqs. (3.1) through (3.3). Remember to invert the final current value to make it negative. You may find this trick of calculating the pMOS current as though it is an nMOS to be less confusing than using Eqs. (3.5) through (3.7).

#### Example 3-1: MOSFET Operating Regions

Consider the nMOS and pMOS transistors arranged in parallel in Figure 3–8. The gate voltage of the nMOS is tied to $V_{DD}$, and the gate voltage of the pMOS is tied to ground. Assume that the input to this configuration instantly switches from 0 to $V_{DD}$. As the output charges up from 0 to $V_{DD}$, in what regions of operation do the transistors function? Give the ranges of the output voltage, $V_{out}$, and the associated regions of operation. Assume that $(W/L)_n = (W/L)_p = 2\ \mu m/1\ \mu m$. Assume the following data:

$$\mu_n C_{ox} = 2\mu_p C_{ox} = 50\ \mu A/V^2$$

$$V_{Tn} = -V_{Tp} = 0.5V,\ L_n = L_p = 1\mu m$$

$$V_{DD} = 3V$$

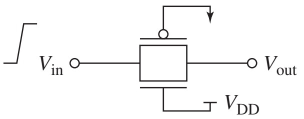


**Figure 3–8:** Parallel n-channel and p-channel MOS transistors used as a transmission gate (i.e., a switch) that is turned on. If the input switches instantly to $V_{DD}$, in what regions of operation do the transistors function as the output changes from 0 to $V_{DD}$?

#### SOLUTION

We will use the following steps to identify the regions of operation. First, we will draw an equivalent circuit. Second, we will identify the terminals of the device. Third, we will see if the device is in cutoff. Finally, for the regions where the device is not cutoff, we will test to see if the device is in triode or saturation.

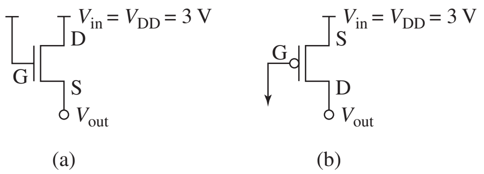


**Figure 3–9:** Equivalent circuits for the (a) nMOS and (b) pMOS transistors in Figure 3–8.

Let us begin the solution by looking at the nMOS transistor. Figure 3–9(a) shows the equivalent circuit for the nMOS after the input has switched. Since the drain is defined as the highest potential, the drain of the nMOS is at the $V_{in}$ side of the transistor once it switches to 3 V. The gate is at 3 V also, and the source is $V_{out}$. The next thing to check is if the transistor is on. For this, we compare $V_{GS}$ to $V_{Tn}$. For the nMOS in Figure 3–9(a), $V_{GS} = 3 - V_{out}$. The nMOS is in cutoff then if

$$3 - V_{out} \le V_{Tn}$$

Plugging in $V_{Tn} = 0.5$ V and solving for $V_{out}$ gives

$$V_{out} \ge 2.5$$

Thus, the nMOS is in **cutoff** if $V_{out}$ is over 2.5 V.

This means that the nMOS is on for the range of $V_{out}$ between 0 V and 2.5 V. To divide this range into the triode and saturation regions, we need to compare $V_{DS}$ to $V_{GS} - V_{Tn}$. We know from (3.2) that the device is in triode if $V_{DS} < V_{GS} - V_{Tn}$. For the nMOS in Figure 3–9(a), $V_{DS} = 3 - V_{out}$ and $V_{GS} - V_{Tn} = 2.5 - V_{out}$, so the device is never in triode. It is thus in **saturation** for the range of $V_{out}$ from 0 V to 2.5 V.

Now, we need to perform the same analysis for the pMOS transistor in Figure 3–8, and we will use the same four steps. First, we redraw the equivalent circuit for the pMOS after the input has changed, as shown in Figure 3–9(b). Next, we identify the terminals of the transistor. Since the drain of a pMOS is defined as the *lowest* voltage, the pMOS drain is at $V_{out}$, and its source is at 3 V.

This means that $V_{SG}$ for the pMOS is 3 V, which is greater than $-V_{Tp}$, so the pMOS will not be cutoff at all during this transition. Now we need only to distinguish between triode and saturation operation. The pMOS is in triode when $V_{SD} < V_{SG} - (-V_{Tp})$. For the pMOS, $V_{SD} = 3 - V_{out}$, and $V_{SG} - (-V_{Tp}) = 2.5$ V. Thus, we can write the following equation and solve for $V_{out}$ to see the range of the output for which the pMOS is in triode:

$$3 - V_{out} \le 2.5$$

$$V_{out} \ge 0.5$$

Thus, the pMOS is in **triode** when the output voltage is above 0.5 V and is in **saturation** when it is less than 0.5 V.

### 3-2-3 Small-Signal MOSFET Model

In analog applications, a DC operating point specified by the bias voltages $V_{GS}$, $V_{DS}$, and $V_{BS}$, is established so $V_{DS} > V_{DS_{SAT}} = V_{GS} - V_{Tn}$. Incremental voltages $v_{gs}$, $v_{ds}$, and $v_{bs}$ that are much smaller in magnitude than the bias voltages perturb the operating point. A *small-signal model* is a circuit representation of the response of the drain current of the MOSFET to these perturbations. Since this book is about digital circuits, we will not spend much time with the small-signal model of a MOSFET. However, we will need to use it occasionally (such as to calculate gain of a digital inverter), so we introduce the model in this section using the nMOS transistor as an example.

The total drain current $i_D = I_D + i_d$ is a function of the gate, drain, and bulk voltages in saturation. By summing the contributions from the perturbations $v_{gs}$, $v_{bs}$, and $v_{ds}$, we can write a first-order expansion for the drain current

$$I_D + i_d = I_D + \left.\frac{\partial i_D}{\partial v_{GS}}\right|_Q v_{gs} + \left.\frac{\partial i_D}{\partial v_{BS}}\right|_Q v_{bs} + \left.\frac{\partial i_D}{\partial v_{DS}}\right|_Q v_{ds} \quad (3.8)$$

where $Q$ is the DC operating point of the MOSFET, which is defined by $V_{GS}$, $V_{SB}$, and $V_{DS}$. The partial derivatives in Eq. (3.8) represent small-signal circuit elements and are given the following symbols.

$$i_d = g_m v_{gs} + g_{mb} v_{bs} + g_o v_{ds} \quad (3.9)$$

where

$$g_m = \left.\frac{\partial i_D}{\partial v_{GS}}\right|_Q, \; g_{mb} = \left.\frac{\partial i_D}{\partial v_{BS}}\right|_Q, \text{ and } g_o = \left.\frac{\partial i_D}{\partial v_{DS}}\right|_Q \quad (3.10)$$

The MOSFET **transconductance** $g_m$ connects the small-signal gate–source voltage $v_{gs}$ to the resulting incremental drain current. The **backgate transconductance** $g_{mb}$ represents the perturbation of the drain current by an incremental change in the bulk–source voltage. Finally, the

**output conductance** $g_o$ relates the incremental change in drain current due to a small-signal, drain–source voltage.

We can calculate these transconductances by evaluating the derivatives shown in Eq. (3.10). For the purposes of this book, we are interested only in the transconductance and the output conductance. A useful approximation to the equation for $g_m$ can be made by neglecting the contribution from the channel-length modulation, which will be a small error if $\lambda_n V_{DS} \ll 1$ :

$$g_m \cong \left(\frac{W}{L}\right)\mu_n C_{ox}(V_{GS} - V_{Tn}) = \sqrt{2\left(\frac{W}{L}\right)\mu_n C_{ox} I_D} \quad (3.11)$$

These results indicate that $g_m$ is equally sensitive to the bias drain current, the channel width-to-length ratio, and $\mu_n C_{ox}$. For a typical channel mobility 215 $cm^2V^{-1}s^{-1}$; oxide capacitance 2.3 $fF\mu m^{-2}$; bias drain current $I_D = 100\ \mu A$; and $(W/L) = 10$, we find that the transconductance $g_m = 316\ \mu S$ (S = A/V).

The drain conductance $g_o$ represents the dependence of the drain current on the drain–source voltage

$$g_o = \left.\frac{\partial i_D}{\partial v_{DS}}\right|_Q = \left(\frac{W}{2L}\right)\mu_n C_{ox}(V_{GS} - V_{Tn})^2\lambda_n \cong \lambda_n I_D \quad (3.12)$$

where we again make the approximation $\lambda_n V_{DS} \ll 1$ at the operating point to obtain a simpler result. The channel-length modulation parameter $\lambda_n$ has been found to vary inversely with the channel length

$$\lambda_n = \propto \frac{1}{L} \quad (3.13)$$

for a MOSFET with $L = 1\ \mu m$, and $\lambda_n = 0.06\ V^{-1}$. At a DC bias current $I_D = 100\ \mu A$, the inverse of the drain conductance (called the **output resistance**) is

$$(g_o)^{-1} = r_o = 150\ k\Omega \quad (3.14)$$

## 3-2-4 Small-Signal Circuit

We can assemble the small-signal model for steady–state perturbations on the operating point by properly connecting these three elements between the four terminals of the MOSFET. Kirchhoff's current law indicates that the three elements are connected in parallel at the drain node shown in Figure 3–10, since their contributions sum to give the small-signal drain current

$$i_d = g_m v_{gs} + g_{mb} v_{bs} + g_o v_{ds} \quad (3.15)$$

Note that the small-signal, steady–state gate current of the MOSFET is zero. In addition, the reverse-biased source–bulk junction is an open circuit in steady–state, as shown in Figure 3–10.

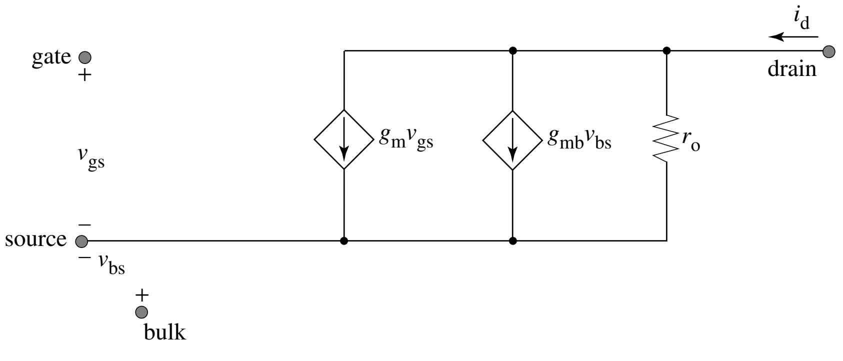


**Figure 3–10:** Steady–state small-signal model for the n-channel MOSFET in saturation.

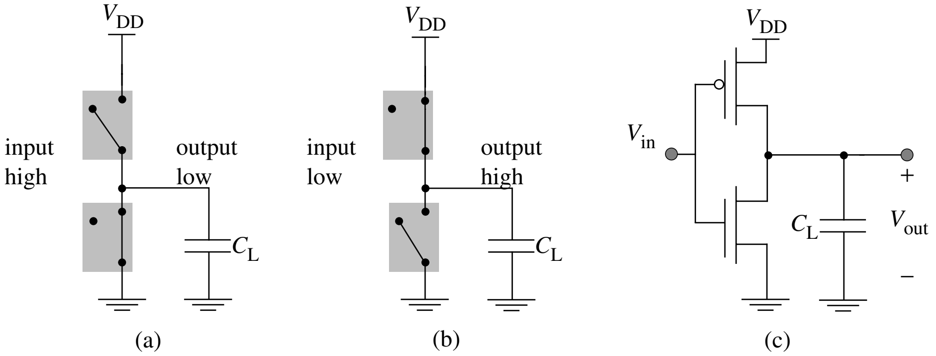


**Figure 3–11:** The CMOS inverter: (a) switch level representation $V_{in}$ = High, $V_{out}$ = Low; (b) switch level representation $V_{in}$ = Low, $V_{out}$ = High; and (c) circuit diagram for CMOS inverter.

## 3-3 CMOS Inverter Characteristics

We demonstrated the concept for a logical inverter in Chapter 2 using ideal switches, and we have seen how MOSFETs behave as approximations of those ideal switches. Now we are ready to use the nMOS and pMOS transistors to construct a CMOS inverter. Figure 3–11(a) and (b) recounts the inverter structure using ideal switches. Since voltage is the quantity that we use to map to logic levels, the input to the inverter is the voltage $V_{in}$. If the input voltage is high, the lower switch closes to discharge the capacitive load. If the input is low, we turn on the top switch to charge up the capacitive load. At no time (or for a very short time) are both switches on, which prevents DC current from flowing from the positive power supply to ground. A simple implementation of this switching concept is shown by shorting the gate terminals of the p- and n-channel devices shown in Figure 3–11(c).

Qualitatively, this circuit acts like the switching circuits, since the p-channel transistor has exactly the opposite characteristics of the n-channel transistor. Hence, when the input voltage is high (e.g., 3 V), the p-channel transistor is off (cutoff), and the n-channel transistor is on (triode). When the input voltage is low, the p-channel transistor is on (triode), and the n-channel transistor is off (cutoff). In the transition region, both transistors are saturated, and the circuit operates with a large voltage gain.

To understand how we develop the voltage transfer characteristic for the CMOS inverter, consider the transistor characteristics shown in Figure 3–12 for the n- and p-channel transistors.

Starting at point **(1)**, where the input voltage is equal to 0 V, we see the output voltage is equal to $V_{DD}$. As we increase the input voltage to point **(2)**, the n-channel transistor operates in its constant-current region while the p-channel transistor is in the triode region. As we further increase the voltage to point **(3)**, both n- and p-channel transistors are in their constant-current region, and we are in the high-gain region of the inverter. Further increasing the input voltage to point **(4)** puts the n-channel transistor in the triode region and the p-channel transistor in its constant-current region. Finally, when the input voltage is greater than $V_{DD} + V_{Tp}$, at point **(5)**, we find the p-channel transistor is below its threshold voltage (cutoff) and the n-channel transistor has 0 V corresponding to an output voltage equal to 0 V.

Qualitatively, the CMOS inverter has low power dissipation, since there is no DC current flowing in either logical state. Also, the speed of the inverter can be set with a constant current charging and discharging the load capacitor. These excellent characteristics, along with its robustness to noise (which we discuss in Section 3-3-2) have made CMOS the technology of choice for complex logic functions, as well as for semiconductor memory.

### 3-3-1 Voltage Transfer Characteristic

The voltage transfer characteristic of the CMOS inverter at the top of Figure 3–12 has some strong similarities to the ideal VTC, which we replicate in Figure 3–13 for your reference.

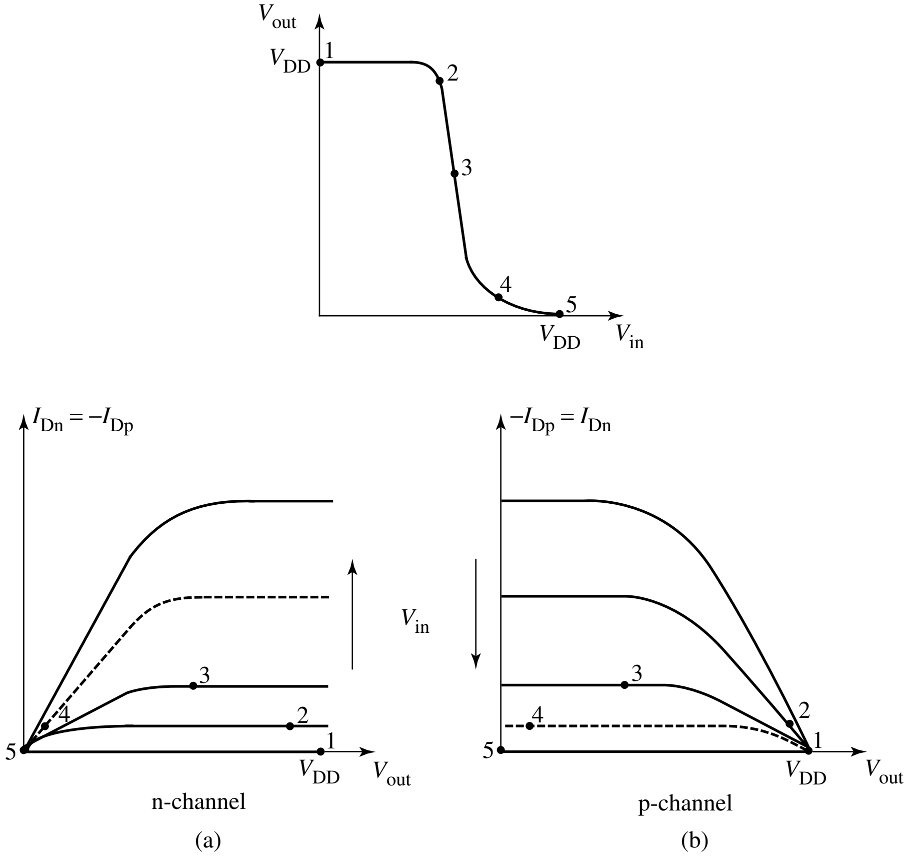


**Figure 3–12:** CMOS voltage transfer characteristic shown with (a) nMOS transistor characteristics and (b) pMOS transistor characteristics. Points 1 through 5 correspond to increasing $V_{in}$ from 0 V to $V_{DD}$. *Note*: Dashed characteristics on n- and p-channel curves are for equal $V_{in}$.

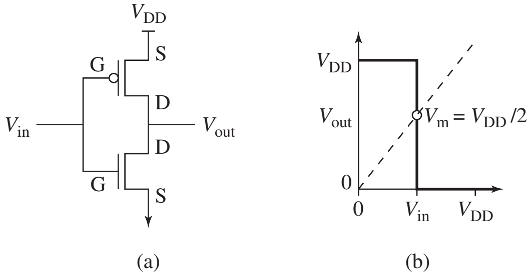


**Figure 3–13:** (a) The inverter with an input voltage, $V_{in}$, output voltage, $V_{out}$, and power supply voltage $V_{DD}$. (b) Ideal transfer characteristic for the inverter.

The CMOS-inverter VTC has the same basic shape as the ideal VTC, but its transitions are less abrupt. We can define a few key values on the real VTC that help us to quantify its non-idealities. Figure 3–14 shows a closer look at a typical CMOS-inverter VTC that has these key points labeled.

To start with, the ideal VTC will map any input voltage below the switching threshold $V_M$ to a logical **1**. The same is not true for the CMOS inverter, because the output voltage drops below $V_{DD}$ before the switching threshold. This occurs because the gain of the non-ideal inverter is not infinite (that is, the slope of the vertical part of the VTC is not straight up and down). We thus must be more careful about specifying valid input and output voltages. We define specific input and output voltages as follows.

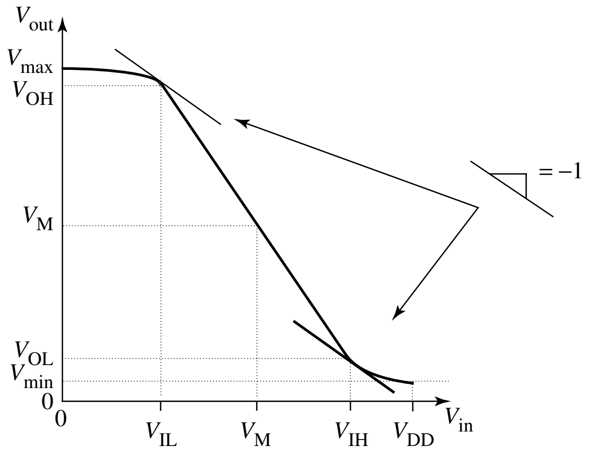


**Figure 3–14:** Typical inverter transfer characteristic. For $V_{in} = 0$ V, $V_{out} = V_{max}$; for $V_{in} = V_{IL}$, $V_{out} = V_{OH}$; for $V_{in} = V_M$, $V_{out} = V_M$; for $V_{in} = V_{IH}$, $V_{out} = V_{OL}$; for $V_{in} = V_{DD}$, $V_{out} = V_{min}$.

**$V_{IL}$ (voltage input low)**—lower input voltage where the slope of the transfer characteristic is equal to –1

**$V_{IH}$ (voltage input high)**—higher input voltage where the slope is equal to –1

**$V_{OH}$ (voltage output high)**—output voltage given an input voltage of $V_{IL}$

**$V_{OL}$ (voltage output low)**—output voltage given an input voltage of $V_{IH}$

**$V_M$ (voltage midpoint)**—input voltage at which the inverter yields an output voltage equal to the input voltage

When the input voltage is equal to 0 V, $V_{out} = V_{max}$ (**maximum output voltage**), which in most cases of interest is $V_{DD}$. For a small input voltage between 0 V and $V_{IL}$, the output voltage will be between $V_{max}$ and $V_{OH}$. $V_{OH}$ is the minimum output voltage for a valid logic **1**. As we further increase the input voltage, the output voltage rapidly falls through the transition region. When the input voltage is between $V_{IL}$ and $V_{IH}$, the output voltage is in the transition region where the logic level is undefined. Notice that the definition of $V_{IL}$ and $V_{IH}$ based on the slope of the VTC (slope = –1) provides a quantitative method of defining the transitional region. If the gain never becomes more negative than –1, the inverter will not properly reject noise. As the input voltage is increased to a value between $V_{IH}$ and $V_{DD}$, the output voltage is a low value between $V_{OL}$ and $V_{min}$ (**minimum output voltage**). $V_{OL}$ is the maximum output voltage for a valid logic **0**. When $V_{in}$ is equal to $V_{DD}$, we define $V_{out} = V_{min}$. In general, $V_{min}$ may not be 0 V, however for static CMOS circuits like this inverter, $V_{min} = 0$ V.

From this transfer characteristic, one can see that if $V_{in} < V_{IL}$, the output voltage is a valid logic **1**. Correspondingly, if $V_{in} > V_{IH}$, the output voltage is a valid logic **0**. If the input voltage is between $V_{IL}$ and $V_{IH}$, the logic output is undefined. This range of input voltages is referred to as the transition region.

## 3-3-2 Logic Levels and Noise Margin

As mentioned in the previous section, the inverter characteristic has three distinct regions: the low region where $V_{in} < V_{IL}$, the high region where $V_{in} > V_{IH}$, and the transition region where $V_{IL} < V_{in} < V_{IH}$. In Figure 3–15, we have one inverter driving a second inverter. The input and output voltage levels corresponding to each of these inverters are indicated. If we look at the region where the first inverter's output is connected to the second inverter's input, we see that $V_{OH1} > V_{IH2}$ and $V_{OL1} < V_{IL2}$.

We can define a **noise margin high**, $NM_H$ which ensures that a logic **1** output from the first inverter is interpreted as a logic **1** input to the second inverter. Similarly, we can define a **noise margin low**, $NM_L$ which ensures that a logic **0** output from the first inverter is interpreted as a logic **0** input to the second inverter. The expressions for both noise margins are given by

$$NM_H = V_{OH} - V_{IH} \tag{3.16}$$

$$NM_L = V_{IL} - V_{OL} \tag{3.17}$$

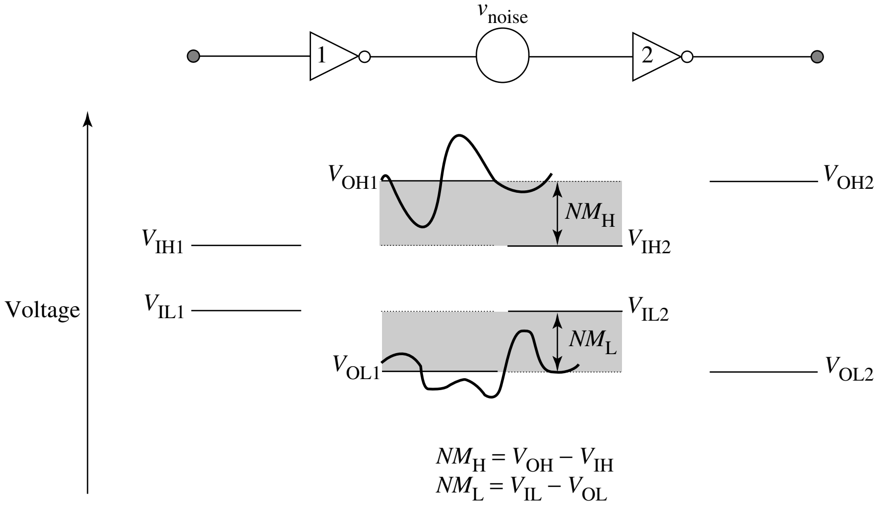


**Figure 3–15:** Chain of two inverters to demonstrate the concept of noise margin. A noise source that represents a variety of noise sources such as capacitive coupling is placed between the inverters.

In summary, we have defined several parameters that indicate the voltage transfer characteristics of the inverter. The conventional notation used to represent voltages in digital circuits is uppercase variable names and uppercase subscripts.

$V_{OH}$ — the minimum output voltage from an inverter which indicates a logic **1**

$V_{OL}$ — the maximum output voltage from an inverter which indicates a logic **0**

$V_{IH}$ — the minimum input voltage to an inverter to output a logic **0**

$V_{IL}$ — the maximum input voltage to an inverter to output a logic **1**

$V_M$ — the voltage at which the input and output voltages are equal

With these values, we have defined noise margins to ensure the proper logical output. The CMOS inverter that we have used as an example has excellent noise margins due to its high gain in the transition region. The strong robustness of static CMOS logic-to-noise contributes to the popularity of this logic implementation choice for digital circuits.

### 3-3-3 Transient Characteristics

In this section, we will determine the time required for an output to change state given that the input changes state. This time is called a **propagation delay**. Figure 3–16 shows a waveform describing a change in input voltage as a function of time and also shows how the output voltage of an inverter changes as a function of time. The **rise time** $t_R$ is defined as the time required for the input or output voltage to change from 10 percent of its high value to 90 percent of its high value. The **fall time** $t_F$ is defined as the time required for the input or output voltage to change from 90 percent of its high value to 10 percent of its high value.

Referring to the output waveform, we define the **propagation delay from high-to-low** $t_{PHL}$ as the delay between the 50 percent points of the input and output waveforms. Similarly, we define the **propagation delay from low-to-high** $t_{PLH}$ as the time between the 50 percent points of the input and output waveforms during this transition of the output. The total **propagation delay** $t_P$, is the average of the low-to-high and high-to-low delays given as

$$t_P = \frac{t_{PHL} + t_{PLH}}{2} \tag{3.18}$$

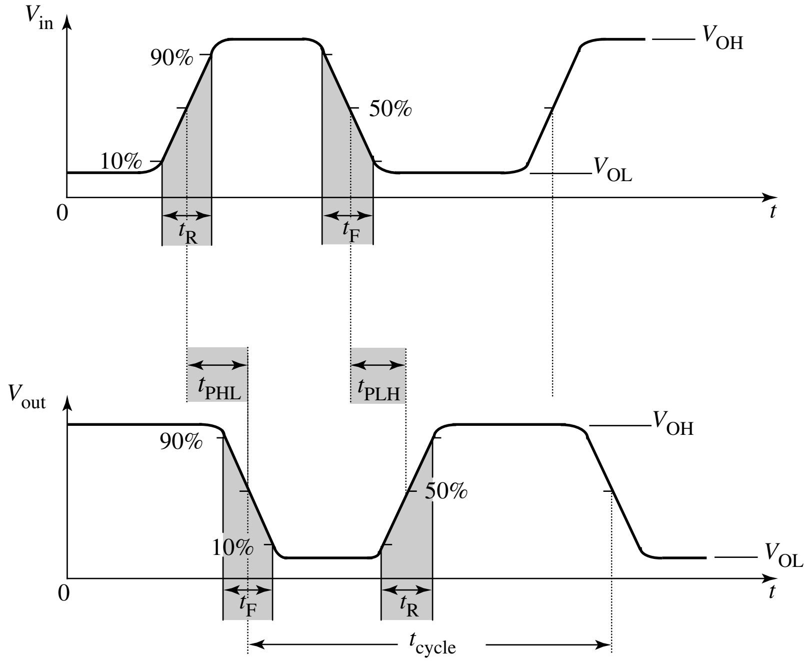


**Figure 3–16:** Input and output voltage as a function of time for an inverter. Definitions of $t_R$, $t_F$, $t_{PHL}$ and $t_{PLH}$ are depicted graphically.

## 3-4 CMOS Inverter Analysis

In the previous section, we presented a qualitative view of the CMOS inverter and its definitions describing its operation. In this section, we will perform a static analysis to analyze the inverter transfer characteristic, quantify logic levels, and determine noise margins. In addition, we will perform a transient analysis to calculate the propagation delay for a typical CMOS inverter. Finally, we will calculate the power dissipation of the CMOS inverter.

### 3-4-1 Simplified Transfer Characteristic for Hand Calculations

Although it is possible to calculate the transfer function and specific voltages $V_{IL}$, $V_{IH}$, $V_{OL}$, $V_{OH}$, and $V_M$ from the large-signal models for MOS transistors, it is algebraically complex. If we make some simplifications to the typical voltage transfer characteristic for the inverter, we find that the calculations become algebraically simple and reasonably accurate for hand analysis.

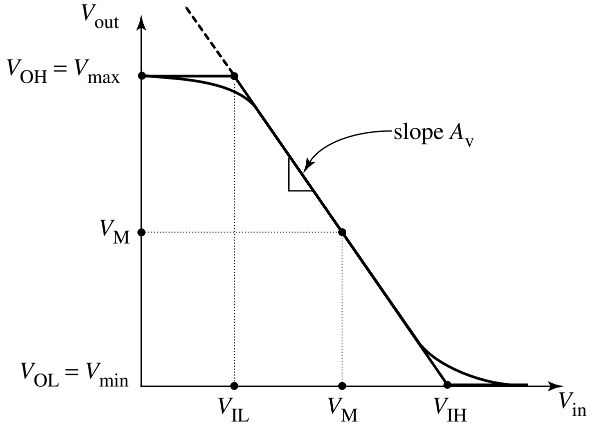


**Figure 3–17:** Idealized inverter transfer characteristic superimposed on a typical inverter transfer characteristic. *Note*: The idealized definitions for $V_{OH}$ and $V_{OL}$ are used in this diagram.

The inverter transfer function for hand calculations is shown in Figure 3–17. For this voltage transfer characteristic, we have redefined $V_{OH}$ as $V_{max}$ and $V_{OL}$ as $V_{min}$. The error introduced is small and negligible for hand analysis.

The critical points in this transfer function are found by first determining $V_{OH}$ and $V_{OL}$ for the circuit. For CMOS $V_{OH} = V_{max} = V_{DD}$ and $V_{OL} = V_{min} = 0$ V. $V_M$ is defined as the input voltage at which the input and output voltages are equal. Under this condition, both the pMOS and nMOS transistors usually are operating in their constant-current or saturation regions. We can use the current equations in a model of the circuit to find the voltage gain $A_v$ at $V_{in} = V_M$. After finding $A_v$ we can draw a straight line at $V_M$ with the slope $A_v$. The points at which this line intersects output voltages $V_{OH}$ and $V_{OL}$ provide a new definition for the input voltages $V_{IL}$ and $V_{IH}$ respectively, which is suitable for hand calculation.

It should be noted that for digital gates to be useful, the voltage gain at the midpoint must be greater than one in magnitude. In practice, the magnitude of the voltage gain is on the order of ten. Because of the relatively large voltage gain, the error in using these definitions for hand calculations is well within those required for an initial design.

## 3-4-2 Static Analysis of the CMOS Inverter

The goal of this analysis is to determine the logic levels and noise margins for a CMOS inverter. From the previous qualitative description, we know that $V_{OH} = V_{DD}$ and $V_{OL} = 0$ V.

### Determine Switching Threshold, $V_M$

The next step is to find the input voltage $V_M$ when the input and output voltages are equal. Under this condition, we will assume that both the n- and p-channel devices are operating in their constant-current region (it is important to plug in the numbers so that you can check this assumption later). The current for the nMOS is given by

$$I_{nD} = \left(\frac{W}{2L}\right)_n \mu_n C_{ox}(V_M - V_{Tn})^2(1 + \lambda_n V_M) \tag{3.19}$$

and the current for the p-channel device is given by

$$-I_{Dp} = \left(\frac{W}{2L}\right)_p \mu_p C_{ox}(V_{DD} - V_M + V_{Tp})^2(1 + \lambda_p(V_{DD} - V_M)) \tag{3.20}$$

If we let

$$k_n = \left(\frac{W}{L}\right)_n \mu_n C_{ox} \tag{3.21}$$

and

$$k_p = \left(\frac{W}{L}\right)_p \mu_p C_{ox} \tag{3.22}$$

and set $I_{Dn}$ equal to $-I_{Dp}$ (i.e., equate the magnitude of the currents), we can solve for $V_M$. We assume that the channel-length modulation terms can be ignored. In fact, for symmetrical transistors ($\lambda_n = \lambda_p$ and $V_M = V_{DD}/2$) the terms precisely cancel. The resulting equation for $V_M$ is

$$V_M = \frac{V_{Tn} + \sqrt{\frac{k_p}{k_n}}(V_{DD} + V_{Tp})}{1 + \sqrt{\frac{k_p}{k_n}}} \tag{3.23}$$

Looking at Eq. (3.23), we see that if $k_n \gg k_p$ then $V_M$ is approximately $V_{Tn}$. Physically, a large $k_n$ implies that the n-channel transistor can sink much more current than the p-channel transistor can provide, and hence, the trip point will be reduced. If $k_p \gg k_n$, then the opposite effect occurs and the trip point $V_M$ moves toward the positive power supply, and in the limit becomes equal to $V_{DD} + V_{Tp}$. While $V_M$ will not reach these extremes for reasonably sized devices, it is very useful to remember that a stronger (e.g., larger $W/L$) nMOS device moves $V_M$ towards 0, and a stronger pMOS device moves $V_M$ towards $V_{DD}$.

**Example 3-2: CMOS Inverter Static Analysis**

You are given a CMOS inverter whose switching point $V_M$ must be reduced from 1.5 V to 1.0 V. Due to layout constraints, the only adjustable parameter is the width of the n-channel transistor $W_n$. When $V_M = 1.5$ V, $W_n = 2$ µm. Find the new n-channel transistor width given the following data:

$$\mu_n C_{ox} = 2\mu_p C_{ox} = 50\ \mu\text{A/V}^2$$

$$V_{Tn} = -V_{Tp} = 0.5\text{V},\ L_n = L_p = 1\mu\text{m}$$

$$V_{DD} = 3\text{V}$$

**SOLUTION**

First find the width of the p-channel transistor. We would like to use Eq. (3.23), but it is only valid if both transistors are saturated, so we must check to see if this is the case. At $V_M = 1.5$ V, $V_{GSn} = 1.5$ V, so $V_{GSn} - V_{Tn} = 1.0$ V which means that the nMOS is on. Furthermore, since $V_{DSn} =$

1.5 V, $V_{GSn} - V_{Tn} < V_{DSn}$, so the nMOS is in saturation. For the pMOS, $V_{SG} = 1.5$ V, $V_{SD} = 1.5$ V, and $-V_{Tp} = 0.5$ V, so the pMOS also is saturated. Now we can safely use Eq. (3.23) with $V_M = 1.5$ V to find that $k_p = k_n$, so

$$\left(\frac{W}{L}\right)_p = \left(\frac{W}{L}\right)_n \cdot \frac{\mu_n C_{ox}}{\mu_p C_{ox}} = \frac{4}{1.0}$$

so $W_p = 4$ µm.

To find the n-channel width required to lower $V_M$ to 1 V, rearrange Eq. (3.23) and use $V_M = 1.0$ V.

$$V_M\left(1 + \sqrt{\frac{k_p}{k_n}}\right) = V_{Tn} + \sqrt{\frac{k_p}{k_n}}(V_{DD} + V_{Tp})$$

$$\sqrt{\frac{k_p}{k_n}} = \frac{V_M - V_{Tn}}{V_{DD} + V_{Tp} - V_M} = \frac{1 - 0.5}{3 - 0.5 - 1} = \frac{0.5}{1.5}$$

so $k_n = 9k_p$

$$\left(\frac{W}{L}\right)_n = 9\left(\frac{W}{L}\right)_p \frac{\mu_p C_{ox}}{\mu_n C_{ox}} = 9 \cdot \frac{4}{1.0} \cdot \frac{1}{2} = \frac{18}{1.0}$$

Therefore, the new n-channel transistor width required is 18 µm. Notice that the required width to move the switching threshold so low is quite large.

## Determine Noise Margins

The first step to finding the noise margins is to calculate $V_{IL}$ and $V_{IH}$. We determine the slope of the transfer characteristic at $V_{in} = V_M$, (i.e., voltage gain) and use it to project a line to intersect at $V_{out} = V_{min} = 0$ V to find $V_{in}$. Similarly, we project a line to intersect at $V_{out} = V_{max} = V_{DD}$ to find $V_{IL}$. To find the voltage gain when the input voltage is equal to $V_M$, we use the small-signal model of both MOS transistors. At that operating point, we find the transconductance and the output resistance of the n- and p-channel devices are given by

$$g_{mn} = k_n(V_M - V_{Tn}) \tag{3.24}$$

$$r_{on} = 1/(\lambda_n I_{Dn}) \tag{3.25}$$

$$g_{mp} = k_p(V_{DD} - V_M + V_{Tp}) \tag{3.26}$$

$$r_{op} = 1/(\lambda_p I_{Dp}) \tag{3.27}$$

The small-signal model for the CMOS inverter is shown in Figure 3–18. Backgate transconductance generators are open circuits, since we assume that $v_{bs} = 0$ V. The relationship between the output voltage and the input voltage is given by

$$A_v = \left(\frac{v_{out}}{v_{in}}\right) = -(g_{mn} + g_{mp})(r_{on} \| r_{op}) \tag{3.28}$$

In CMOS inverters, one generally uses the shortest channel-length transistor that the technology will provide in order to have the largest current possible for a given device size. For MOSFETs, the output resistance will be small if the channel length is small, and hence, the voltage gain of typical CMOS inverters is on the order of –10.

To find $V_{IL}$ and $V_{IH}$, we use the known slope $A_v$ and find

$$V_{IL} = V_M + \frac{(V_{DD} - V_M)}{A_v} \tag{3.29}$$

$$V_{IH} = V_M - \frac{V_M}{A_v} \tag{3.30}$$

These equations are extremely simple because we made the straight-line approximation that the voltage gain was constant between $V_{IL}$ and $V_{IH}$ as shown in Figure 3–19.

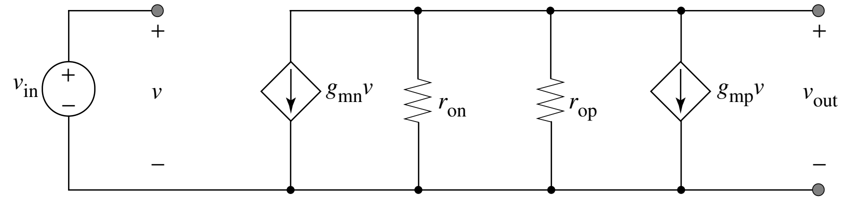


**Figure 3–18:** Small-signal model of CMOS inverter evaluated at $V_{in} = V_M$.

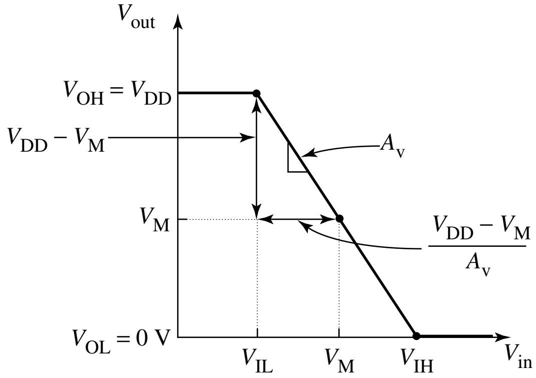


**Figure 3–19:** CMOS transfer characteristic illustrating the calculation of $V_{IL}$, using the hand-analysis approximation.

Their accuracy certainly is adequate for hand analysis.

The last quantities to calculate for the CMOS inverter static analysis are the noise margins. They are given by

$$NM_L = V_{IL} - V_{OL} = V_M + (V_{DD} - V_M)/A_v \qquad (3.31)$$

$$NM_H = V_{OH} - V_{IH} = V_{DD} - V_M + V_M/A_v \qquad (3.32)$$

### Design Example 3-3: CMOS Static Inverter

In this example, we are going to design a basic CMOS inverter. The inverter's transition voltage must be $V_M = 1.5$ V, the current through the n- and p-channel devices should be 100 μA at $V_{in} = V_M$, and $NM_L = NM_H \geq 1.35$ V. The power supply voltage $V_{DD}$ is 3 V. Specify to the process engineer the maximum channel length modulation λ that can be tolerated to meet all of the above design specifications. Assume $\lambda_n = \lambda_p$.

**MOS Device Data:**

$$\mu_n C_{ox} = 2\mu_p C_{ox} = 50\ \mu A/V^2$$

$$V_{Tn} = -V_{Tp} = 0.5V \quad \text{and} \quad L_n = L_p = 1\mu m$$

### SOLUTION

To begin this design, we need to find the $(W/L)$ ratios of the devices needed to carry 100 μA at $V_{in} = V_M$. Setting $V_M = 1.5$ V and assuming that λ is small for this calculation,

$$I_{Dn} = \left(\frac{1}{2}\left(\frac{W}{L}\right)_n \mu_n C_{ox}(V_M - V_{Tn})^2\right)$$

$$\left(\frac{W}{L}\right)_n = \frac{4.0}{1.0}$$

We will have to make $(W/L)_p$ twice as big as $(W/L)_n$, because $\mu_n = 2 \cdot \mu_p$, so

$$\left(\frac{W}{L}\right)_p = \frac{8.0}{1.0}$$

To have the required noise margins of 1.35 V, the minimum voltage gain needed can be found by using Eqs. (3.31) and (3.32).

$$NM_L = V_M + \frac{V_{DD}}{2A_v} \quad \text{and} \quad NM_H = V_{DD} - V_M + \frac{V_{DD}}{2A_v}.$$

So

$$A_v = \frac{V_{DD}}{2(NM_L - V_M)} = -10$$

Since $(W/L)_p = 2(W/L)_n$ and $\mu_n C_{ox} = 2\mu_p C_{ox}$, $k_n = k_p$. Using Eq. (3.31) and noting $\lambda_n = \lambda_p$, $g_{mn} = g_{mp}$ and $r_{on} = r_{op}$ for $V_M = 1.5$ V. We write

$$A_v = -(g_{mn} + g_{mp})(r_{on} \parallel r_{op}) = -(2g_{mn})\left(\frac{r_{on}}{2}\right)$$

Now

$$r_{on} = \frac{1}{\lambda_n I_D}.$$

Substituting for $r_{on}$ and solving for $\lambda_n$, we arrive at

$$\lambda_n \leq \frac{-g_{mn}}{A_v I_D}$$

We obtain a value for $g_{mn}$ by

$$g_{mn} = \sqrt{2\left(\frac{W}{L}\right)_n \mu_n C_{ox} I_{Dn}} = 0.28\text{mS}$$

Substituting $g_{mn} = 0.28$ mS, $A_v = -10$, and $I_D = 100$ μA into the expression for $\lambda_n$, we have

$$\lambda_n \leq 0.28\ V^{-1}$$

## 3-4-3 Propagation Delay

It is extremely difficult to calculate transient voltage waveforms accurately using only hand analysis. Therefore, as we did in the static analysis, we will idealize the transient responses so a simple hand analysis of the propagation delay may be performed. Figure 3–20 shows the idealized transient voltages that we will use for hand analysis.

The essential difference between the idealized and real versions is that we assume the input voltage makes its transition infinitely fast. We measure the propagation delay from the edge of the input voltage transition to the 50 percent point on the output voltage as indicated. The input signal transition is positioned at the 50 percent point of the actual input-voltage waveform. Computer simulation is required for more accurate analysis of the transient response.

A typical CMOS inverter may drive successive gates. The input capacitance of these gates as well as the parasitic capacitance due to the depletion capacitance of the drains of the inverters and wiring capacitance are lumped in a load capacitance $C_L$, as shown in Figure 3–21(a).

Figure 3–21(b) shows Inverter 1 driving two inverters labeled 2 and 3. In our analysis, we will calculate the propagation delay of Inverter 1. As stated earlier, we assume that the input transition time is infinitely fast. Looking at the high-to-low transition for $V_{out}$ for the inverter, we follow the trajectory of the nMOS current-voltage transistor characteristic shown in Figure 3–22(a).

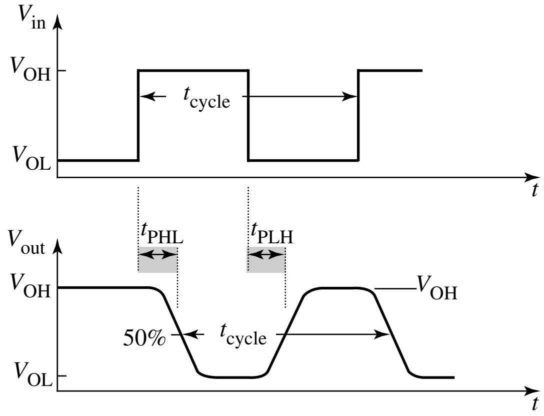
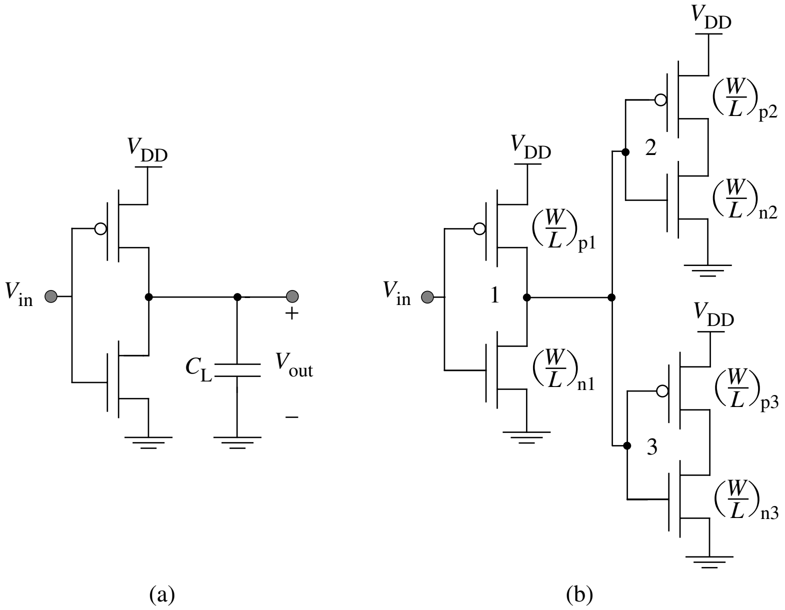


**Figure 3–20:** Idealized transient response of an inverter used for hand calculation.



(a) (b)

**Figure 3–21:** CMOS inverter (a) Driving a lumped load capacitance $C_L$. (b) driving additional inverters 2 and 3.

At $t = 0^-$, the input voltage is equal to zero and the output voltage is equal to $V_{OH}$. At $t = 0^+$, we assume the input voltage has made an instantaneous transition to $V_{OH}$. The output voltage then decreases along the trajectory shown. As discussed in Section 3-3-3, $t_{PHL}$ is defined as the time it takes the output voltage to discharge from $V_{OH}$ to $V_{OH}/2$. To discharge the capacitor, $C_L$, from $V_{OH}$ to $V_{OH}/2$, we need to remove an amount of charge $Q = C_L \Delta V$. Current is the flow of the charge over time, so we can write an equation for $t_{PHL}$ as

$$t_{PHL} = \frac{C_L \Delta V}{I_D} = \frac{C_L V_{OH}/2}{\frac{k_n}{2}(V_{OH} - V_{Tn})^2} \tag{3.33}$$

We have assumed that the nMOS transistor remains in the constant-current region throughout this transition, so the capacitor voltage is linearly discharged by its saturation current, as shown in Figure 3–22(b). Although this may not be strictly true for our definition of $V_{DS_{SAT}}$, it is quite adequate for this simplified hand analysis. We also have ignored the effects of channel-length modulation, which would cause the delay to decrease slightly due to the larger current.

The load capacitance $C_L$ consists of two major components.

1. The *total gate capacitance* $C_G$ that is the gate capacitance of the transistors being driven by the inverter.
2. A *parasitic capacitance* $C_P$ that results from the drain diffusions of the driving inverter and the wiring to the gates being driven.

In the following paragraphs, we will explore the method to calculate these two components of $C_L$. Referring to Figure 3–21, we see that the inverter being analyzed is driving two additional inverters. Its fan-out is therefore two. These two inverters are switching from the triode region through saturation to cutoff during either a high-to-low or low-to-high transition. To accurately calculate the total charge supplied to the load inverters, we must perform an analysis with non-linear charge storage elements. However, for a simple hand calculation, we can estimate a worst-case capacitance. The total gate capacitance of the load inverters is given by

$$C_G = \sum_{\text{(load)inverters}} C_{in} \tag{3.34}$$

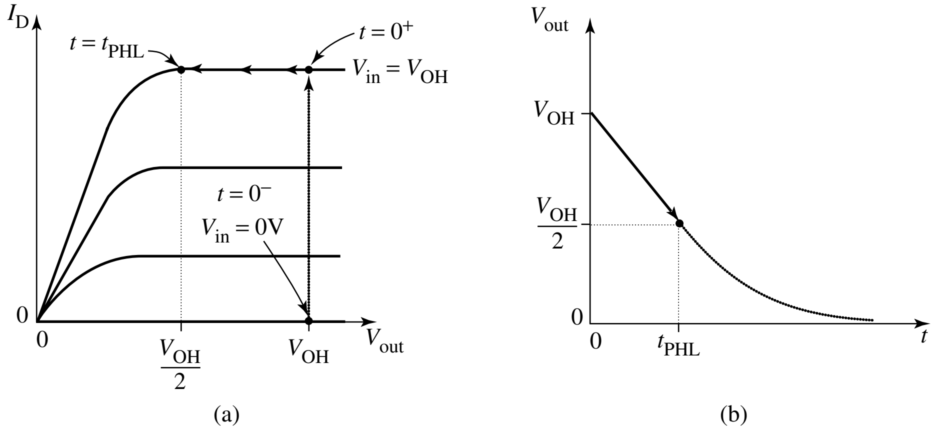
where the input-gate capacitance of a single inverter is approximated by

$$C_{in} \approx C_{ox}(W \cdot L)_p + C_{ox}(W \cdot L)_n \tag{3.35}$$

The total gate capacitance $C_G$ for the two inverters in Figure 3–21(b) can be estimated as

$$C_G = C_{ox}[(WL)_{p2} + (WL)_{n2} + (WL)_{p3} + (WL)_{n3}] \tag{3.36}$$

Refer to the layout and cross-section of an NMOS transistor in Figure 3–23(a) to calculate the parasitic capacitance, which is the second component of $C_L$, due to the drain diffusions of Inverter 1. The n$^+$ drain diffusion has a deple-



(a) (b)

**Figure 3–22:** (a) NMOS transistor characteristic showing the trajectory of the current-voltage relationship from $t = 0$ s to $t = t_{PHL}$. (b) Transient waveform for $V_{out}$.

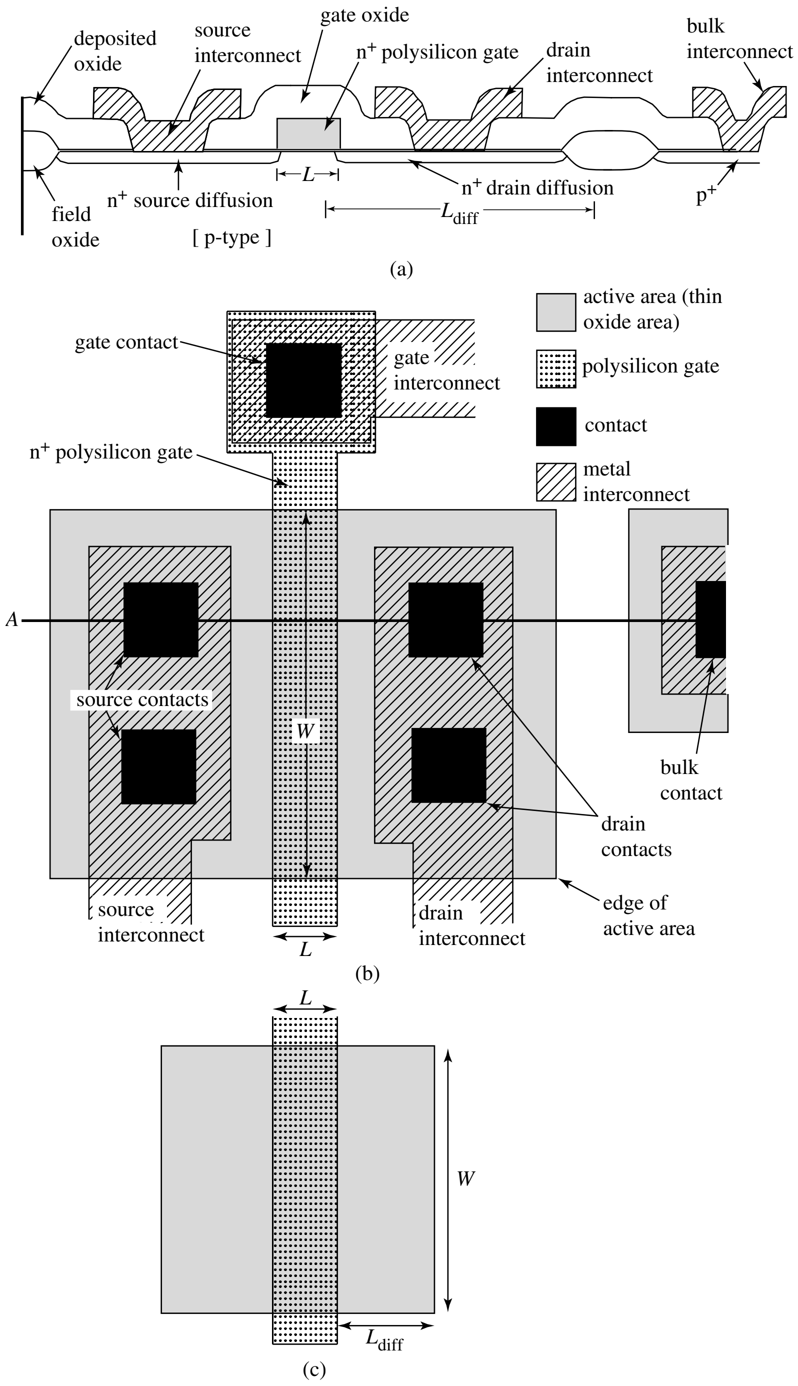


**Figure 3–23:** Origin of parasitic drain-bulk depletion capacitance. (a) nMOS cross section; (b) nMOS top view; (c) definition for dimensions $L_{\text{diff}}$ and $W$ used to calculate depletion capacitance.

tion-region capacitance that is determined primarily by the doping concentration in the p-type region under that diffusion. The doping concentration due to the field implant can be higher along the sides of the $n^+$ drain diffusion, than at the bottom of the diffusion. Because of this, we separately account for the perimeter capacitance along the edge of the drain diffusion. Although this capacitance is voltage dependent, for hand analysis we usually use a zero-bias capacitance value for our worst-case estimate.

Most MOS process technologies specify a value to the circuit designer for the **bottom junction capacitance per unit area, $C_{Jn}$**, and a **sidewall junction capacitance per unit length, $C_{JSWn}$**. Figure 3–23(b) shows a top view of the n-channel MOS transistor. Because there are design rules that limit the minimum size of the contact, the spacing between the contact and polysilicon gate, and the contact and outer region of the active region, there is a **minimum length of the drain diffusion** that we call **$L_{diff}$**. This important dimension (along with the device width) is sketched in Figure 3–23(c).

The area of the drain diffusion then is $W \times L_{diff}$. The sidewall perimeter of the device is equal to $W + 2L_{diff}$. It should be noted that the diffusion capacitance along the edge of the gate is not included in this calculation. This capacitance already is taken into account in the intrinsic MOS transistor model. A similar analysis applies to the pMOS transistors. This discussion shows that the total drain-bulk depletion capacitance $C_{DB}$ for an inverter is given by

$$C_{DB} = W_n L_{diffn}(C_{Jn}) + W_p L_{diffp}(C_{Jp}) + (W_n + 2L_{diffn})C_{JSWn} + (W_p + 2L_{diffp})C_{JSWp} \quad (3.37)$$

Finally, we should add a component to the parasitic capacitance that is due to the wiring between the drain of the inverter and input of the next gate. Figure 3–24 shows a cross section of a metal line connecting a polysilicon gate and running across both thermal and deposited oxide. The capacitance of this structure is given approximately by the permittivity of the oxide divided by the total dielectric thickness. To calculate the contribution of the wiring to the parasitic capacitance, multiply this capacitance per-unit-area times the width and length of the metal interconnect line.

Between tightly coupled logic gates, the metal wiring capacitance often is negligible compared to the drain–bulk depletion capacitance. However, it is important to realize that in large digital systems we often have metal busses that extend a significant distance across a chip. Under this condition, the wiring capacitance is the dominant capacitance that determines the speed of the transition. In fact, in modern integrated circuits, the ultimate speed of microprocessors often is determined by the delay caused by long wires. Wire capacitance (specifically due to capacitance between wires in the same routing layer) contributes even more significantly in more deeply scaled technologies due to the fact that the distance between wires is shrinking while their heights remain roughly constant.

The total parasitic capacitance in our example is given by

$$C_P = C_{DB} + C_{wire} \quad (3.38)$$

We add the values of $C_G$ and $C_P$ to obtain the total load capacitance $C_L$. Substituting $C_G + C_P$ for $C_L$, we find the value of $t_{PHL}$ to be

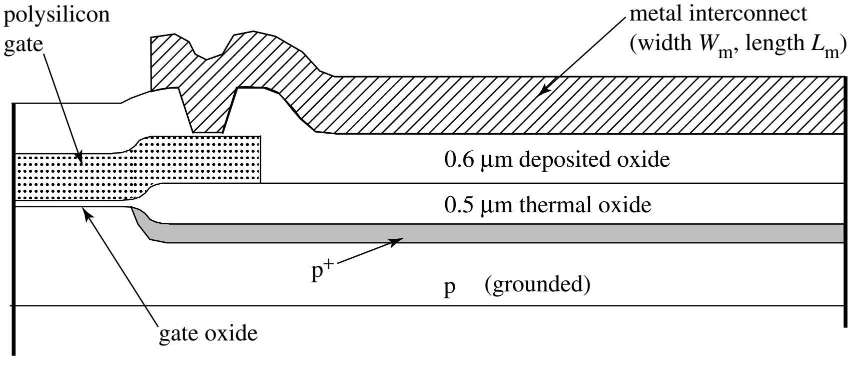


**Figure 3–24:** Cross section showing origin of parasitic wiring capacitance.

$$t_{\mathrm{PHL}} = \frac{(C_{\mathrm{G}} + C_{\mathrm{P}})V_{\mathrm{OH}}/2}{k_{\mathrm{n}}/2(V_{\mathrm{DD}} - V_{\mathrm{Tn}})^2} \tag{3.39}$$

Similarly, we can find a value for $t_{\mathrm{PLH}}$ to be

$$t_{\mathrm{PLH}} = \frac{(C_{\mathrm{G}} + C_{\mathrm{P}})V_{\mathrm{OH}}/2}{k_{\mathrm{p}}/2(V_{\mathrm{DD}} + V_{\mathrm{Tp}})^2} \tag{3.40}$$

It often is useful to design digital gates to make the propagation delay symmetrical; namely to make $t_{\mathrm{PHL}} = t_{\mathrm{PLH}}$. This removes the data dependency of the delay and minimizes the time that both transistors are on simultaneously. Because the mobility of holes in the p-channel transistor is approximately one-half that of electrons in the n-channel transistor, Eqs. (3.39) and (3.40) indicate that setting $(W/L)_{\mathrm{p}} = 2(W/L)_{\mathrm{n}}$ equalizes the rise and fall delays, assuming $V_{\mathrm{Tn}} = -V_{\mathrm{Tp}}$.

The average propagation delay $t_P$ can be estimated by averaging the transition times as shown in

$$t_{\mathrm{P}} = (t_{\mathrm{PHL}} + t_{\mathrm{PLH}})/2 \tag{3.41}$$

In large, complex digital systems, many individual gate-propagation delays are incurred by placing digital gates in series. In our example, the driving inverter had a fan-out of two. The average fan-out is higher than two, and an inverter with a fan-out of four commonly is used to provide a reference propagation delay called the FO4 delay. Typical systems have between twenty and fifty FO4 propagation delays for every major clock cycle, although the fastest modern processors may shrink this towards ten for some blocks. This implies that a microprocessor with a clock speed of 3 GHz must have single gate delays less than 30 picoseconds. These small delays are possible only in newer process technologies than the example technology that we present in this text.

## Example 3-4: CMOS Inverter Propagation Delay Analysis

A CMOS inverter driving two other identically sized inverters is shown in Figure 3–25. Calculate $t_{\mathrm{PHL}}$ and $t_{\mathrm{PLH}}$. Ignore wiring capacitance but include the parasitic drain-bulk depletion capacitance.

**MOS Device Data**

$\mu_{\mathrm{n}}C_{\mathrm{ox}} = 50\ \mu\mathrm{A/V}^2$

$\mu_{\mathrm{p}}C_{\mathrm{ox}} = 25\mu\mathrm{A/V}^2$

$V_{\mathrm{Tn}} = -V_{\mathrm{Tp}} = 0.5\mathrm{V}$

$C_{\mathrm{ox}} = 2.3\ \mathrm{fF}/\mu\mathrm{m}^2$

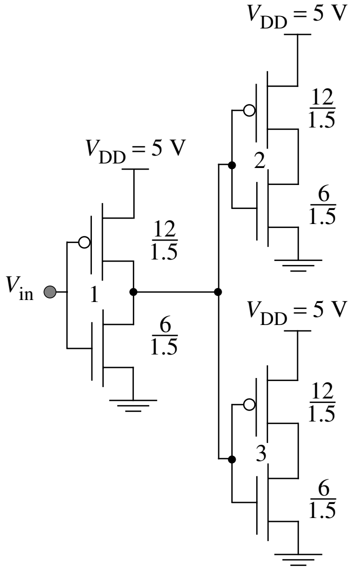


**Figure 3–25:** Example inverter with fan-out of two for computing propagation delay in Example 3-4.

$C_{\mathrm{Jn}} = 0.1\ \mathrm{fF}/\mu\mathrm{m}^2$

$C_{\mathrm{Jp}} = 0.3\ \mathrm{fF}/\mu\mathrm{m}^2$

$C_{\mathrm{JSWn}} = 0.5\ \mathrm{fF}/\mu\mathrm{m}$

$C_{\mathrm{JSWp}} = 0.35\ \mathrm{fF}/\mu\mathrm{m}$

$L_{\mathrm{diffn}} = L_{\mathrm{diffp}} = 6\ \mu\mathrm{m}$

### SOLUTION

Begin the solution by finding $C_{\mathrm{L}}$. Since Inverters 2 and 3 are identically sized, $C_{\mathrm{G}}$ is two times that of Inverter 2. From Eq. (3.34), we find

$$C_{\mathrm{G}} = 2C_{\mathrm{ox}}[(WL)_{\mathrm{p2}} + (WL)_{\mathrm{n2}}]$$

$$C_{\mathrm{G}} = 2\left(2.3\frac{\mathrm{fF}}{\mu\mathrm{m}^2}\right)[(12\ \mu\mathrm{m}\cdot 1.0\mu\mathrm{m}) + (6\ \mu\mathrm{m}\cdot 1.0\mu\mathrm{m})]$$

$$= 82.8\mathrm{fF}$$

Since the wiring capacitance is neglected, $C_{\mathrm{P}} = C_{\mathrm{DB}}$.

$$C_{\mathrm{P}} = C_{\mathrm{DB}} = W_{\mathrm{n1}}L_{\mathrm{diffn}}C_{\mathrm{Jn}} + W_{\mathrm{p1}}L_{\mathrm{diffp}}C_{\mathrm{Jp}}$$

$$(W_{\mathrm{n1}} + 2L_{\mathrm{diffn}})C_{\mathrm{JSWn}} + (W_{\mathrm{p1}} + 2L_{\mathrm{diffp}})C_{\mathrm{JSWp}}$$

$$C_{\mathrm{DB}} = \left[6\mu\mathrm{m}\cdot 6\mu\mathrm{m}\cdot 0.1\frac{\mathrm{fF}}{\mu\mathrm{m}^2}\right] + \left[12\mu\mathrm{m}\cdot 6\mu\mathrm{m}\cdot 0.3\frac{\mathrm{fF}}{\mu\mathrm{m}^2}\right]$$

$$+\left[(6\mu\text{m} + 2\cdot 6\mu\text{m})\cdot 0.5\frac{\text{fF}}{\mu\text{m}}\right]$$

$$+\left[(12\mu\text{m} + 2\cdot 6\mu\text{m})\cdot 0.35\frac{\text{fF}}{\mu\text{m}}\right]$$

$$C_{DB} = 3.6\text{ fF} + 21.6\text{ fF} + 9\text{ fF} + 8.4\text{fF} = 42.6\text{ fF}$$

So

$$C_L = C_G + C_{DB} = 82.8\text{ fF} + 42.6\text{ fF} = 125\text{ fF}.$$

From Eq. (3.39),

$$t_{PHL} = \frac{(C_G + C_P)(V_{DD}/2)}{\frac{k_n}{2}(V_{DD} - V_{Tn})^2} \quad\text{and}\quad k_n = \left(\frac{6}{1.0}\right)50\mu\text{A/V}^2,$$

so $t_{PHL}$ = 200 ps.

From Eq. (3.40),

$$t_{PLH} = \frac{(C_G + C_P)(V_{DD}/2)}{\frac{k_p}{2}(V_{DD} + V_{Tp})^2} \quad\text{and}\quad k_p = \left(\frac{12}{1.0}\right)25\ \mu\text{A/V}^2$$

so $t_{PLH}$ = 200 ps.

## Design Example 3-5: CMOS Input Inverter/Buffer

Design a 3-V CMOS inverter that is an input to an integrated circuit. The midpoint logic level of the signal driving the CMOS inverter is 0.9 V. The CMOS inverter must be able to drive a 6-μm wire that may traverse the entire 1-cm chip as well as two other inverters sized at $(W/L)_p = 2(W/L)_n = 10/1.0$. The specifications needed for the inverter require $t_P$ to be less than 5 ns and both noise margins to be at least 0.8 V. $\lambda_n = \lambda_p = 0.1\text{ V}^{-1}$, and the capacitance value of the wire is 0.035 fF/μm². Specify the channel widths needed to build such an inverter. Use the same device data as in Example 3-4. Verify your design with a SPICE simulation.

### SOLUTION

Start by finding the ratios of the n- and p-channel devices. After checking to ensure that both devices are in saturation, rearrange Eq. (3.23) to find

$$\sqrt{\frac{k_p}{k_n}} = \frac{V_M - V_{Tn}}{V_{DD} + V_{Tp} - V_M} = \frac{0.9 - 0.5}{3 - 0.5 - 0.9} = \frac{0.4}{1.6},$$
$$\rightarrow k_n = 16k_p.$$

Since $k_n \gg k_p$, $t_{PHL}$ will be negligible. This leaves $t_{PLH}$ to be determined by the $t_P$ specification. To find $t_{PLH}$, first determine the values of $C_G$ and $C_P$. From Example 3-4, $C_G$ = 83 fF. Since the size of the p-channel device is not yet known, we cannot accurately determine $C_P$. However, we do know it will be dominated by the long wire.

$$C_P \approx C_{\text{wire}} = 6\mu\text{m}\cdot 10,000\mu\text{m}\cdot 0.035\text{fF}/\mu\text{m}^2\cdot\frac{1}{1000}\frac{\text{pF}}{\text{fF}}$$
$$= 2.1\text{pF}$$

From Eq. (3.40) we determine

$$k_p \approx \frac{200\mu\text{A}}{\text{V}^2}$$

This implies that $(W/L)_p = 8.0/1.0$. To make $k_n = 16k_p$, we need $(W/L)_n = 16/2(W/L)_p$, so $(W/L)_n = 64/1.0$.

Finally, check to see if our inverter will meet the minimum noise-margin requirements. Since $V_M$ is closer to 0, the low noise margin is the only one likely to fail to meet the constraint. Since we know that the noise margin must be 0.8 V, we can solve Eq. (3.31) to find the gain that is required to ensure adequate $NM_L$, which shows that the gain must be $A_v \le -21$. Now we can find the voltage gain $A_v$ of our inverter, at $V_{in} = V_M$, to compare with this constraint.

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

$$g_{mn} = \sqrt{2k_n I_{Dn}}$$

where

$$I_{Dn}(V_{in} = V_M) = \frac{k_n}{2}(V_M - V_{Tn})^2 \approx 256\mu\text{A}$$

yielding

$$g_{mn} = 1.28\text{ mS and } g_{mp} = \frac{g_{mn}}{\sqrt{16}} = 0.32\text{mS}$$

Thus,

$$r_{on} = r_{op} = \frac{1}{\lambda I_{Dn}} = 39\text{k}\Omega\text{, so } A_v = -31.2.$$

Since the gain is less than -21, the noise margins are sufficient, and our hand analysis has given us a good approximation.

In digital-integrated circuit design, it is very important to use computer simulation tools to verify both static and transient analyses. Tools that accurately extract the gate and parasitic capacitance should be used. These capacitance values can be transferred into a circuit simulation program (such as SPICE) to find more accurate delay times for the particular digital gate being studied.

### 3-4-4 Power Dissipation

The increased use of portable electronics (such as cellular phones and notebook computers) has made power dissipation an important design metric in modern microelectronics. Portable devices that operate using a battery have a limited energy supply and thus have lifetimes that are constrained by their power consumption. Even ICs in systems that are plugged into a continuous power supply are becoming power constrained due to the difficulty of dissipating the heat that results from consuming power on a chip with many tightly packed transistors.

There are two primary components that make up the power dissipation in static CMOS digital gates. The first is **dynamic power**, which is needed to charge and discharge the load capacitance.

#### Dynamic Power

Figure 3–26 shows an inverter driving a capacitive load, which we will use to calculate the dynamic power consumption. We begin by noting that energy is drawn from the positive power supply $V_{DD}$ only when the inverter charges the load capacitor up to a logical '1'. Thus, on the high-to-low output transition, no energy is drawn from the power supply. For the low-to-high output transition, we can calculate the energy drawn from the supply as the integral of power over time. Thus, we can write that the energy drawn from the power supply to charge up $C_L$ is equal to

$$E_{V_{DD}} = \int_0^T P(t)dt = V_{DD}\int_0^T i_{V_{DD}}(t)dt \tag{3.42}$$

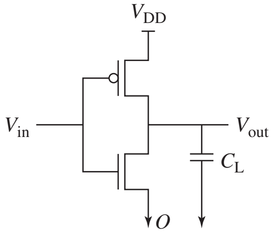


**Figure 3–26:** Inverter driving a capacitive load, $C_L$.

where $T$ is the clock period. Now, we know that the current coming from the power supply to charge the capacitor must be equal to the current in the pMOS transistor, and this current is nonlinear according to the MOS equations. However, the current from the power supply is also equal to the current flowing onto the capacitor. The charge on a capacitor equals the capacitance times the voltage across the cap ($Q = CV$, so long as $C$ is not a function of voltage, which we will assume), and current is just the flow of charge:

$$i(t) = \frac{d}{dt}q(t)$$

Using these facts, we can rewrite the current as

$$i_{V_{DD}}(t) = i_{cap}(t) = \frac{dq_{cap}}{dt} = C_L\frac{dV_{out}}{dt} \tag{3.43}$$

Now, plugging Eq. (3.43) into Eq. (3.42) gives:

$$E_{V_{DD}} = C_L V_{DD}\int_0^{V_{DD}} \frac{dV_{out}}{dt}dt = C_L V_{DD}\int_0^{V_{DD}} dV_{out}$$

$$E_{V_{DD}} = C_L V_{DD}^2 \tag{3.44}$$

The total energy pulled from the supply to charge up the capacitor is thus independent of the dimensions of the MOSFET devices and the length of the transition. Instead, it depends solely on the size of the capacitor and on the voltage swing.

We can use a similar approach to calculate the energy stored in the capacitor after the transition completes and the capacitor is fully charged to $V_{DD}$. Again, energy equals the integral of power:

$$E_{cap} = \int_0^T P(t)dt = \int_0^T V_{out} i_{V_{DD}}(t)dt \tag{3.45}$$

Making a similar substitution as before yields

$$E_{cap} = C_L\int_0^{V_{DD}} V_{out}dV_{out} = \frac{1}{2}C_L V_{DD}^2 \tag{3.46}$$

The energy stored in the capacitor is thus half of the total energy drawn from the power supply during the transition. Therefore, the energy *dissipated* by the p-channel transistor in charging the capacitor to $V_{DD}$ is equal to the total energy drawn from the supply less the amount remaining on the cap after it is charged. This subtraction yields $E_{diss}$=(1/2) $C_L V_{DD}{}^2$. Notice that the remaining energy on the capacitor will be dissipated during the high-to-low transition of the output.

We can use this analysis of the energy consumed during an inverter's operation to determine the average power consumption of that inverter. Consider that the inverter in Figure 3–26 switches on and off $f$ times/second. This means that it will charge the load capacitance $f$ times/second, and it will draw $C_L V_{DD}^2$ Joules of energy from the power supply each time. Since average power is just average energy divided by time, let us use $1/f$ as the time period of interest. This corresponds to the equivalent clock period of the circuit containing the inverter. If we divide the total energy per transition by the clock period, then the average dynamic power $P_D$ dissipated by the CMOS gate is equal to

$$P_D = f C_L V_{DD}^2 \tag{3.47}$$

This equation tells us the average power consumed by the inverter, assuming that it switches high then low once every cycle (e.g., every $1/f$ seconds). If the inverter has an input that is clocked at the frequency $f$ but that does not switch every cycle, then we must scale this power number by the rate at which the input transitions. Specifically, we can quantify this switching rate as the activity factor of the output switching from low-to-high, as $\alpha_{0\to 1}$. The average power consumption thus becomes

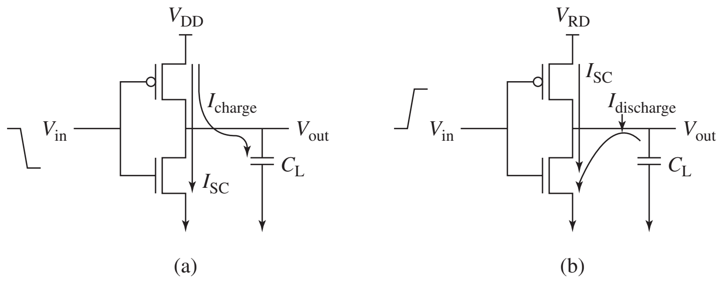
$$P_D = \alpha_{0\to 1} f C_L V_{DD}^2 \tag{3.48}$$

For a well designed CMOS circuit, the majority of the power consumption is attributed to dynamic power. Equation (3.48) shows the power consumed by an inverter driving a capacitive load. This same equation often is used to describe the power consumption of a larger digital CMOS circuit. In that case, $f$ is the clock frequency, and $V_{DD}$ remains the supply voltage. The other terms, however, become approximations of the actual values for the activity factor and the switched capacitance of the circuit. For example, the activity factor will change as different inputs are applied to a complicated circuit. Selecting an $\alpha_{0\to 1}$ to represent the average switching activity and calculating the capacitance as the average effective switched capacitance allows us to use Eq. (3.48) to find a good estimate of the total average power consumption.

## Short Circuit Power

The second type of power dissipated in a CMOS circuit happens during the time when the output of the circuit is undergoing its transition. During this time, both the nMOS and pMOS transistors are on, and current can flow from the power supply to ground. This type of power is called **short-circuit power**. Figure 3–27 shows an inverter driving a capacitive load. Depending on the direction of the output transition, a current will flow to charge or to discharge the capacitor. Regardless of the direction of the output transition, short circuit current will flow temporarily from power to ground while both transistors are on.

From a DC point of view, we can see that the short circuit current will flow in an inverter as long as its input voltage is between $V_{Tn}$ and $V_{DD}-|V_{Tp}|$, since both transistors will be in either the linear or saturation mode for this range of the input. The short circuit power can be defined as



(a) (b)

**Figure 3–27:** Inverter driving a capacitive load, $C_L$, with short circuit current illustrated during (a) a rising and (b) falling transition.

$$P_{SC} = V_{DD} I_{SCavg} \tag{3.49}$$

where $I_{SCavg}$ is the average short circuit current drawn during the transition. The average short circuit current depends strongly on the inverter design, since it will vary based on the duration of the input-voltage transition and on the output voltage.

Let us first analyze an inverter without an output load, so consider the inverter in Figure 3–27 with $C_L = 0$. This will make the dynamic component of power equal to zero, so all of the current that flows in the inverter will be short circuit current. Figure 3–28(a) shows a piecewise linear-input waveform driving this ideal inverter without any load capacitance along with the output voltage. Since there is no capacitance, the output voltage can settle immediately to its steady–state value, so it takes the same shape as the inverter's VTC (it is just stretched along the time axis depending on the duration of the input transition). Figure 3–28(b) shows the short circuit current of the inverter for this transition. The current remains at 0 when the input voltage is less than $V_{Tn}$ or greater than $V_{DD} + V_{Tp}$ because either the nMOS or pMOS is off. From the time point labeled $t_{VTn}$ to the time when the current reaches its maximum value, $t_{mid}$, the nMOS transistor is in saturation. Ignoring channel length modulation, we can write that the short circuit current from $t_{VTn}$ to $t_{mid}$ is given at any instant as

$$I_{SC} = k/2(V_{in}(t) - V_{Tn})^2 \tag{3.50}$$

Assuming that the pMOS is symmetric with the nMOS in its parameters, the current will mirror this behavior as it decreases from $t_{mid}$ to $t_{VTp}$. Let us take the rising transition of the input as an example and calculate the short circuit power. Assuming a linear relationship between the input and time during a rising transition (for example), as shown in Figure 3–28(a) as

$$V_{in}(t) = \frac{V_{DD} t}{t_{rf}} \tag{3.51}$$

where $t_{rf}$ is the total (e.g., 0 to 100 percent) rise time (fall time) of the linearized input signal. From this assumption, we can solve

$$t_{VTn} = V_{Tn}(t_{rf}/V_{DD}) \tag{3.52}$$

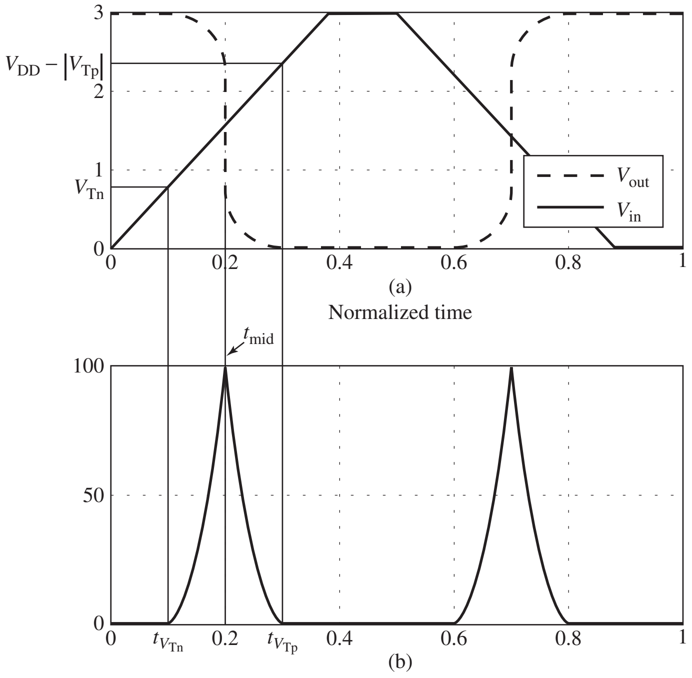


**Figure 3–28:** Inverter driving a capacitive load, $C_L$, with short circuit current illustrated during (a) a rising and (b) falling transition.

and

$$t_{\mathrm{mid}} = t_{\mathrm{rf}}/2\,. \tag{3.53}$$

The average short circuit current from $t_{V\mathrm{Tn}}$ to $t_{\mathrm{mid}}$ is simply the integral of current over this time period divided by $t_{\mathrm{mid}}$–$t_{V\mathrm{Tn}}$. Judging from Figure 3–28(a), there are four regions of current with this same average over an entire period of time $T$ (assuming symmetric nMOS and pMOS). We thus can write an expression for the average $I_{\mathrm{SC}}$ during a period ($T$) of the input signal (e.g., both a rise and a fall) as

$$I_{\mathrm{SCavg}} = \frac{4}{T}\int_{t_{V\mathrm{Tn}}}^{t_{mid}} I(t)dt = \frac{2k}{T}\int_{t_{VTn}}^{t_{mid}} (V_{in}(t) - V_{\mathrm{Tn}})^2 dt \tag{3.54}$$

We can use Eqs. (3.52) and (3.53) to rewrite the bounds of the integral and plug in Eq. (3.51) to solve for the mean current [1]:

$$\begin{aligned} I_{\mathrm{SCavg}} &= \frac{2k}{T}\int_{V_{\mathrm{T}}t_{\mathrm{rf}}/V_{\mathrm{DD}}}^{t_{\mathrm{rf}}/2} \left(\frac{V_{\mathrm{DD}}}{t_{\mathrm{rf}}}t - V_{\mathrm{T}}\right)^2 dt \\ &= \frac{1}{12}\frac{k}{V_{\mathrm{DD}}}(V_{\mathrm{DD}} - 2V_{\mathrm{T}})^3\frac{t_{\mathrm{rf}}}{T} \end{aligned} \tag{3.55}$$

Multiplying this expression by $V_{\mathrm{DD}}$ gives the equation for average short circuit power as [1]:

$$P_{\mathrm{SC}} = \frac{k}{12}(V_{\mathrm{DD}} - 2V_{\mathrm{T}})^3\frac{t_{\mathrm{rf}}}{T} \tag{3.56}$$

The short circuit power depends linearly on the frequency ($1/T$) and on the device parameters $k = (W/L)\mu C_{\mathrm{ox}}$ and $V_{\mathrm{T}}$. It also depends linearly on the rise or fall time of the input, which confirms the intuition that shorter rise and fall times will decrease $P_{\mathrm{SC}}$. Keep in mind that the result in Eq. (3.56) applies only to an ideal inverter with no capacitive load. Now, let us examine the impact of adding a capacitance at the output.

First, consider the case where the output capacitance is very large relative to the drive strength of the inverter in Figure 3–28(a). For this case, the input will transition, but the output will not change due to the large capacitance.

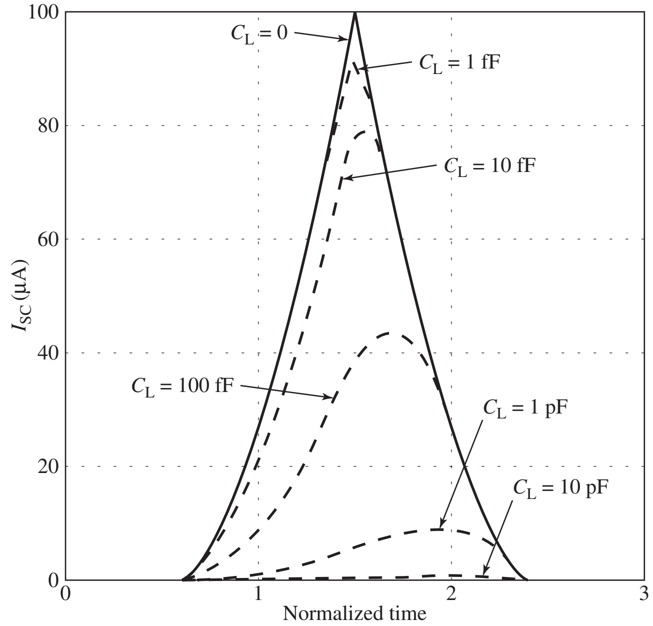
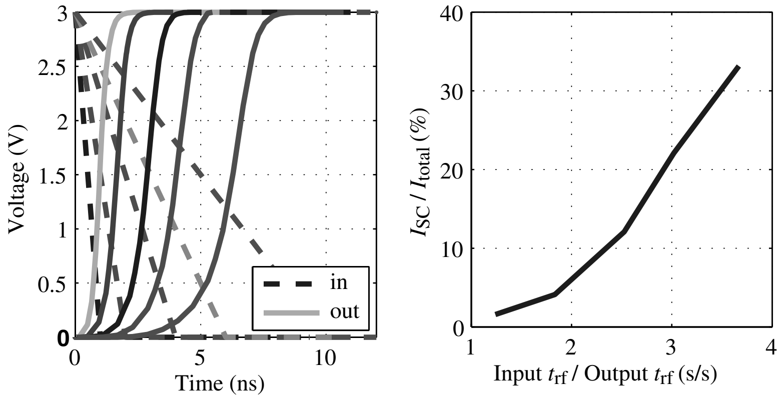


**Figure 3–29:** Inverter driving a capacitive load, $C_{\mathrm{L}}$, with short circuit current illustrated during rising and falling transition.



(a) (b)

**Figure 3–30:** (a) Varying input fall time for an inverter driving a capacitive load.(b) The percentage of short circuit current to total current versus the ratio of input to output rise(fall) times.

Thus, for a rising input, the input voltage will reach $V_{DD}$ while the output remains at $V_{DD}$. This biases the nMOS strongly in the saturation region, and it will begin to discharge the large capacitance. However, by this time, the pMOS already is cut off, since the input is at $V_{DD}$. Thus, due to the slow reaction of the output voltage, there will be zero short circuit current in this case. Therefore, the cases when $C_L$ equals zero or infinity set the bounds on the short circuit current. Generally, a larger $C_L$ will lead to less short circuit power (but more dynamic power). Figure 3–29 confirms this result for a few specific cases of $C_L$. A good design rule related to this observation is to ensure that the rise and fall times of a gate's input are equal to or less than the rise and fall times of the gate's output. For this condition, the short circuit power will be a small fraction of the total power dissipation [1]. Figure 3–30(a) shows the input and output waveforms of an inverter driving a capacitive load. As the input fall time increases, the short circuit current also increases. Figure 3–30(b) shows the ratio of short circuit current to total current versus the ratio of output rise (fall) time to input (rise) fall time. Since most designs keep the input and output $t_{rf}$ approximately the same, short circuit power largely can be ignored relative to dynamic power.

In general, the total CMOS power dissipation is quite low when compared to other digital technologies. However, as clock frequencies continually increase, the dynamic power in CMOS digital integrated circuits is becoming large to the point of becoming the limiting constraint. One important method to significantly reduce this power is to lower the power-supply voltage, since this reduces dynamic power dissipation quadratically. The power-supply voltage in modern technologies with gate lengths of 45 nm is around 1.0 V.

## Summary

In this chapter, we described the analysis and design of a simple, static CMOS inverter based on CMOS device models. We concentrated on simplified hand analysis of the static and transient characteristics of the CMOS inverter. We identified and quantified the primary sources of power dissipation in static CMOS logic. Important concepts that you should have mastered are

- Formal and simplified definitions of logic levels and noise margin from the voltage transfer characteristic.
- Calculation of propagation delay including the effect of load-gate capacitance and parasitic wiring and drain/source–bulk capacitances.
- The relationship between device parameters (e.g., $V_T$, $W/L$) and the voltage transfer characteristic and propagation delay for the static CMOS inverter.
- How to compute static parameters, delay, and power consumption using CMOS device models.

## References

**1.** H. Veendrick, "Short-Circuit Dissipation of Static CMOS Circuitry and Its Impact on the Design of Buffer Circuits," *IEEE Journal of Solid State Circuits*, Volume SC-19, Number 4, August 1984, pages 468-473.

**2.** L. A. Glasser and D. W. Dobberpuhl, *The Design and Analysis of VLSI Circuits*, Addison-Wesley, 1985, Chapter 4.

Includes a discussion of the sources of noise in MOS digital circuits.

**3.** D. A. Hodges and H. G. Jackson, *Analysis and Design of Digital Integrated Circuits, 2nd Ed.*, McGraw-Hill, 1988, Chapter 3.

Includes a discussion on nMOS logic at about the same level.

**4.** J. Rabaey, A. Chandrakasan, and B. Nikolic, *Digital Integrated Circuits: A Design Perspective*, Second Edition, Prentice Hall, 1996, Chapters 3, 5, and 6.

Chapters 5 and 6 have a more in-depth discussion of CMOS static, dynamic, pass transistor logic and several other CMOS logic styles. Chapter 4 points out performance limitations due to interconnect.

**5.** M. Shoji, *CMOS Digital Circuit Technology*, Prentice Hall, 1988, Chapters 2 through 5.

More advanced treatment of CMOS static and dynamic logic gates.

**6.** J. P. Uyemura, *Fundamentals of MOS Digital Integrated Circuits*, Addison-Wesley, 1988, Chapters 3 through 4, and 6.

Chapters 3 and 4 describe static and transient analysis of an MOS inverter. Chapter 6 analyzes combinational MOS logic circuits at about the same level.

**7.** N. H. E. Weste and K. Eshraghian, *Principles of CMOS VLSI Design, 2nd Ed.*, Addison-Wesley, 1993, Chapters 4 through 5.

More advanced treatment of the practical limitations to circuit performance and a more complete discussion of modern CMOS logic structures.

# Problems

**P3.1** Noise often comes from capacitive coupling between clock lines and our signal of interest, as shown in Figure P3-1. Given $V_{OH} = 2.7\,\text{V}$, $V_{OL} = 0.3\ \text{V}$, $V_{IH} = 1.8\ \text{V}$ and $V_{IL} = 1.3\ \text{V}$

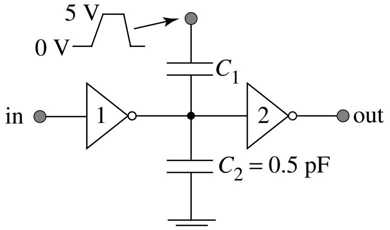


**Figure P3-1**

(a) Calculate $NM_H$ and $NM_L$.

(b) Find the maximum value for $C_1$ given the noise margins calculated in (a). (*Hint*: Recall that $\Delta V/V = C_1/(C_1 + C_2)$ .

(c) To improve the immunity to capacitive coupling we could increase the size of $C_2$ by 2X. What is the new value of $C_1$ which can be tolerated? (Note: This method of reducing coupling also increases propagation delay.)

**P3.2** Given an nMOS inverter with a pull-up resistor of 1kΩ (that replaces the pMOS)

(a) Find the *W/L* of the nMOS transistor such that $V_{OL} = 0.5$ V

(b) For the nMOS device size found in (a), calculate $NM_H$ and $NM_L$. Use the simplified hand calculation method that linearizes the gain.

(c) Repeat (a) and (b) for a pull-up resistor of 10 kΩ.

**P3.3** Repeat Problem P3.2 for a pMOS inverter with a pull-up resistor of 1 kΩ.

**P3.4** In this problem, compare an nMOS and pMOS inverter with a current source pull-up as shown in Figure P3-4. Size the n- and p-channel devices such that $V_M = 1.5$ V for both cases. Choose the smaller device size to be 2/1.0.

(a) For the nMOS case in Figure P3-4(a), calculate $V_{OH}$, $V_{OL}$, and $A_v$ given that the gate of the p-channel pull-up device is tied to $V_B = 2.0$ V.

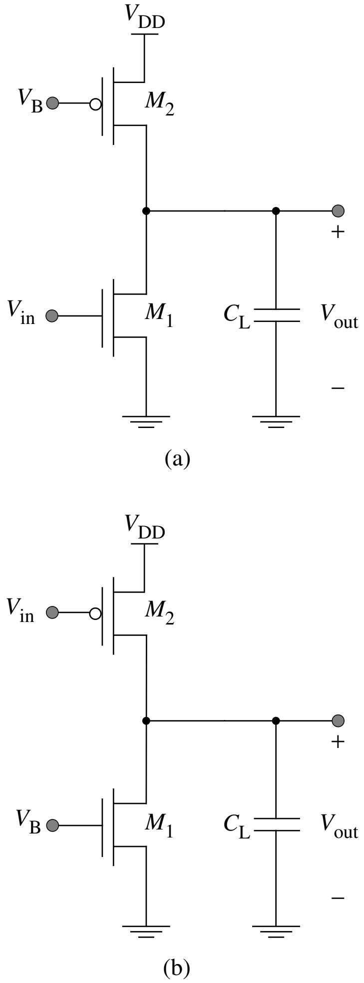


**Figure P3-4**

(b) For the pMOS case in Figure P3-4(b), calculate $V_{OH}$, $V_{OL}$ and $A_v$ given that the gate of the n-channel pull-down device is tied to $V_B = 1.5$ V.

**P3.5** For the nMOS inverter with current source pull-up shown in Figure P3-4(a) with $(W/L)_n = (W/L)_p = 4/1.0$ and $V_B = 2.0$ V

(a) Calculate $t_{PHL}$ and $t_{PLH}$ if the inverter is loaded only by an identical inverter. Include $C_{DB}$ for both the nMOS and pMOS transistors.

(b) Calculate $t_{PHL}$ and $t_{PLH}$ if the inverter is loaded with three identical inverters.

(c) Assume that the system must have a clock with period $T \geq 10t_p$, where $t_p$ is the average propagation delay. Find the maximum length of interconnect wires from the output of the inverter to the inputs of the next stages for a fan-out of 3 if we want the system clock to run at 50 MHz. Assume the width of the interconnect is 4 μm and the capacitance of the wires are 0.05 fF/μm$^2$.

**P3.6** An nMOS inverter with current source pull-up shown in Figure 3-4(a), has $(W/L)_n = 4/1.0$, and $(W/L)_p = 2/4$, and $V_B = 0$ V. Use the simplified hand-calculation method. Assume devices are in their constant-current region.

(a) Sketch the voltage transfer characteristic and label $V_{IL}$, $V_{IH}$, $V_M$, $V_{OH}$ and $V_{OL}$.

(b) If a 100 -fF load capacitor is connected to the output of the inverter, calculate the propagation delay $t_p$. Neglect $C_{DB}$ for this part.

(c) Calculate $t_p$ including $C_{DB}$.

(d) What is the static power consumed by this circuit?

(e) Calculate the device widths such that $C_{DB} = 100$ fF while maintaining the same $V_M$.

(f) What is $t_p$ for the device sizes calculated in part (e)?

**P3.7** To explore how the noise margin can change with device sizing in a CMOS inverter, calculate $V_M$, $NM_H$, and $NM_L$ when

(a) $(W/L)_n = 1$ and $(W/L)_p = 2$

(b) $(W/L)_n = 10$ and $(W/L)_p = 0.2$

(c) $(W/L)_n = 0.1$ and $(W/L)_p = 20$

**P3.8** In this problem, we will explore the accuracy of our simplified equation Eq. (3.33), to calculate $t_{PHL}$ for the CMOS inverter.

(a) Write a differential equation that is valid when the nMOS transistor is saturated that relates the device current and capacitor current in terms of $V_{out}(t)$.

(b) Write another differential equation that is valid when the nMOS transistor is in the triode region that relates the device current and capacitor current.

(c) Solve these equations and write an expression for $t_{PHL}$, that is the time it takes for the output voltage to drop from $V_{OH}$ to $V_{OH}/2$.

(d) For an nMOS transistor with $W/L = 4/1.0$ and $C_L = 100$ fF, solve for $t_{PHL}$ using the expression found in (c) and compare it with the value found from Eq. (5.30).

(e) Repeat (d) for $W/L = 20/1.0$ and $C_L = 500$ fF.

**P3.9** In this problem, you will size a CMOS inverter with process parameters:

$V_{Tn} = 0.5$ V, $V_{Tp} = -0.7$ V, $\mu_n = 50$ cm$^2$/Vs,

$\mu_p = 20$ cm$^2$/Vs, $t_{ox} = 20$ nm, $\lambda_n = \lambda_p = 0.05$ V$^{-1}$.

Assume equal channel lengths, $V_{DD} = 3$ V and all other process parameters are unchanged.

(a) Calculate the ratio $W_n/W_p$ such that $V_M = 1.5$ V

(b) When $V_{in} = V_M$ we want the current through the inverter to be 1mA. What are $W_n$ and $W_p$ assuming the channel length of both devices is 1.5 μm?

(c) Sketch and label the voltage transfer characteristic.

(d) What are $NM_L$ and $NM_H$?

**P3.10** The CMOS inverter which you have sized in Problem P3.9 must drive two identical inverters connected in parallel, as shown in Figure 3–21. Using process parameters from Problem P3.9,

(a) What is the component of load capacitance for the drain–bulk capacitance from your inverter?

(b) What is the component of load capacitance from the two additional inverters?

(c) Calculate $t_{PHL}$ and $t_{PLH}$.

**P3.11** With a greater emphasis on portable electronics, CMOS logic is being designed with lower power-supply voltages to reduce power dissipation. In this problem, we will investigate the changes in the voltage transfer function and the propagation delay when the power supply is reduced. You are given a CMOS inverter driving three identical inverters with $(W/L)_n = 2/1.0$ and $(W/L)_p = 4/1.0$.

(a) Sketch the voltage transfer function for $V_{DD} = 3$ V.

(b) Calculate $t_{PHL}$ and $t_{PLH}$.

(c) What is the dynamic power dissipation when the inverter is running at 50 MHz

(d) Repeat (a) through (c) with $V_{DD} = 1.8$ V.

**P3.12** In Problem P3.11 we saw that the power dissipation was significantly reduced with the reduction of the power supply voltage at the expense of noise margin and propagation delay. Modern processes are using reduced threshold voltages to reduce this problem. Repeat Problem P3.11 with $V_{Tn} = -V_{Tp} = 0.4$ V.

**P3.13** Given a minimum-geometry CMOS inverter with $(W/L)_n = 2/1.0$ and $(W/L)_p = 4/1.0$, what is the maximum fan-out possible while keeping $t_{PLH}$ and $t_{PHL} \leq 600$ ps? Calculate with

(a) $V_{DD} = 3$ V

(b) $V_{DD} = 1.8$ V

**P3.14** How many identical inverters with

(a) $(W/L)_n = 2/1.0$ and $(W/L)_p = 4/1.0$

(b) $(W/L)_n = 20/1.0$ and $(W/L)_p = 40/1.0$

(c) $(W/L)_n = 100/1.0$ and $(W/L)_p = 200/1.0$

can be driven by a single inverter, if the propagation delay must be less than 10x larger than that of the unloaded single inverter? For the unloaded inverter, neglect wiring capacitance but include $C_{DB}$.

**P3.15** For the device data and inverter sizes given in Example 3-4, what is the maximum wire length between the first inverter and inverters 2 and 3 for $t_p \leq 1\text{ns}$? Assume that $C_{wire} = 0.1$ fF/µm.

**P3.16** A figure of merit for a digital technology is the power-delay product, $P_D t_p$. Plot the power-delay product as a function of $V_{DD}$ for $1.25\text{ V} \leq V_{DD} \leq 3\text{ V}$. For this calculation, use a CMOS inverter with $(W/L)_n = 2/1.0$ and $(W/L)_p = 4/1.0$ driving three identical inverters (fan-out = 3). Assume a clock frequency of 50 MHz.

## Design Problems

For these design problems, the technology restricts $W > 2$ µm and $L > 1.0$ µm.

**D3.1** You are to design an nMOS inverter with resistor pull-up to drive a load capacitance of 0.3 pF. The $RC$ time constant must be less than 1 ns. Size the nMOS transistor such that $V_M = 1.5$ V and $NM_L$ and $NM_H \geq 0.8$ V. Minimize the power dissipation.

(a) For the initial design neglect $C_{DB}$.

(b) Include $C_{DB}$ for the refined design.

(c) Verify your design in SPICE

**D3.2** Repeat the design in Problem D3.1 with a pMOS inverter with resistor pull-up.

**D3.3** The goal of this problem is to determine the number of identical inverters that a CMOS inverter can drive, given that they are far away from the driving inverter.

(a) Calculate the component of load capacitance from a single inverter load, $C_G$. The device sizes for the inverter are $(W/L)_n = 4/1.0$ and $(W/L)_p = 8/1.0$

(b) The load inverters are located 1000 µm from the driving inverter. The connecting wire is 2-µm wide aluminum and lies on a deposited glass and field oxide layer. The total thickness of the dielectric layer is 1.0 µm and you can assume it behaves like a parallel capacitor. The permittivity of the dielectric is $3.9\varepsilon_o$. What is the parasitic capacitance resulting from driving one inverter? Include both wiring and drain–bulk capacitance.

(c) Calculate $t_{PHL}$ and $t_{PLH}$ for a single inverter load.

(d) The specifications of the system you are designing has a clock rate of 50 MHz, and during each clock phase, a signal may experience a maximum of 25 propagation delays. Using a 20 percent safety margin on the propagation delay, what is the fan-out of the inverter?

**D3.4** Repeat Problem D3.3 with the load inverters 1 cm from the driving inverter. Calculate the increase in power dissipation when the larger size inverter is used.