CHAPTER 3

DEVICE MODELING

In this chapter device models are discussed. Models for the MOSFET, diode, and BJT are presented. The models presented here are widely used for developing design equations, hand analysis, and initial computer simulations. Both dc models, which are useful for biasing and large signal analysis, and ac models, which are useful for small signal sinusoidal steady state analysis, are discussed.

3.0 MODELING

The fundamental goal in device modeling is to obtain the functional relationship among the terminal electrical variables of the device that is to be modeled. These electrical characteristics depend upon a set of parameters including both geometric variables and variables dependent upon the device physics. In the models discussed, all variables that appear in the equations will be identified as either process or design parameters. The circuit designer has control of the design parameters and judiciously sets these parameters during design. The process parameters are characteristic of the semiconductor process itself and are not at the control of the circuit designer once the process has been specified. The circuit designer may, however, interact with the process engineer to help specify the process that will be used for a particular design.

For most physical electrical devices, at best only a good approximation to the actual relationship of the electrical variables can be obtained. Tradeoffs are often made between the quality of the approximation and its complexity. The required accuracy and the intended use of the model are factors the engineer considers when making these tradeoffs. A very simple model is generally necessary to provide insight for design and facilitate symbolic hand manipulations whereas a more

accurate and correspondingly more sophisticated model is generally preferred for computer simulations of circuits employing these devices. Typically, the models are initially developed by analytically applying basic physical principles and then empirically modifying the resulting mathematical expressions to improve agreement between theoretical and experimental results. For the convenience of the circuit designer, it is often also required that the device models have electrical characteristics identical to those of relatively simple circuits composed of basic circuit components.

The starting point for modeling both the BJT and MOSFET discussed here will be a dc model. From this dc model, a linear small signal model and equivalent simplified ac and dc circuits will be derived. The models will then be expanded upon to provide better agreement between the theoretical and experimental results for use in computer simulation. This approach is summarized in Fig. 3.0-1.

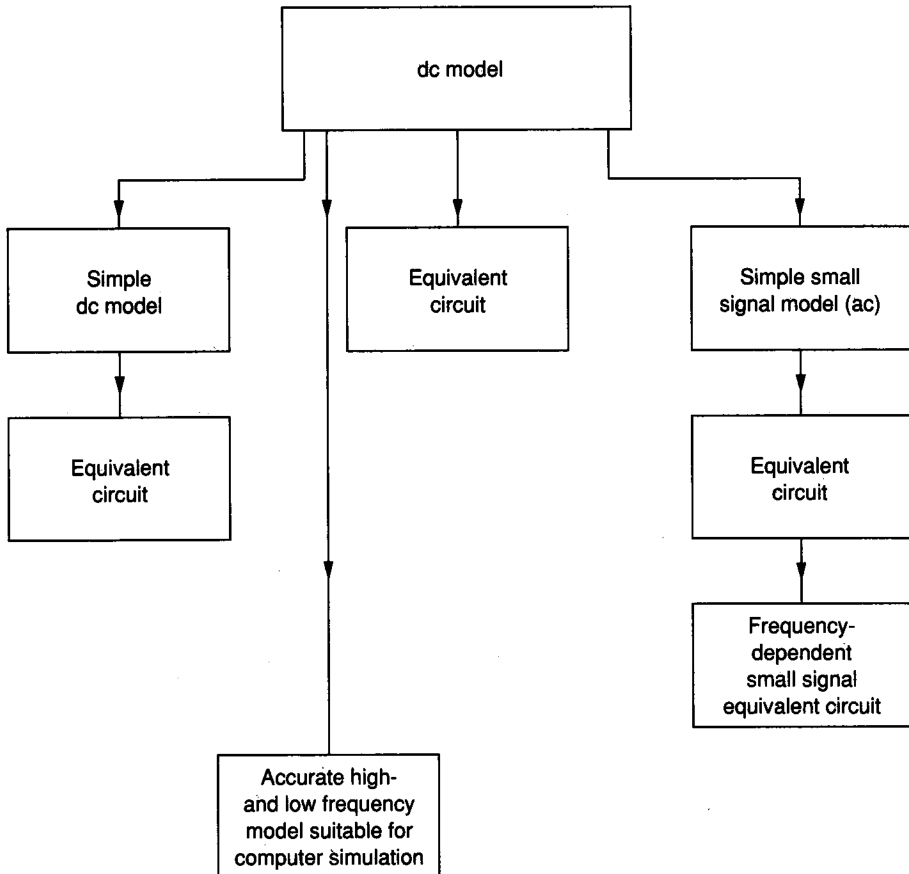


FIGURE 3.0-1
Approach to device modeling.

3.0.1 dc Models

The dc model of a device is a mathematical or numerical relationship that relates the actual terminal voltages and currents of the device at dc and low frequencies. The dc model should be valid over a large range of terminal voltages and currents. It is valid at frequencies where the difference between the actual and dc solution is deemed negligible for the problem under investigation.

For the MOSFET, diode, and bipolar transistor, the dc model is far from linear. This generally makes a dc analysis of circuits containing these devices somewhat tedious.

3.0.2 Small Signal Models

Most circuits perform the task for which they were intended only over a limited excitation range. This range is typically specified in terms of a maximum input signal excursion about some nominal point. Internal to the circuit, these input variations typically cause excursions around some nonzero dc operating point. Often these inputs are sinusoids of small amplitude compared to the supply voltages providing power to the circuit. An analysis of how these *small sinusoids* propagate through the circuit is termed small signal analysis or ac analysis. The points (nodal voltages and branch currents) about which the circuit operates are termed the *bias points* or quiescent points (*Q-points*). Although it might be desirable to have all Q-points at either zero volts or zero amps, it is not practical and generally not possible to do this.

For small signal applications circuits are often designed so that both the MOSFETs and BJTs behave as linear devices even though the devices themselves are actually very nonlinear. This is achieved by restricting signal excursions to a region sufficiently small to obtain approximately linear device behavior. It will be shown later that because of the high degree of nonlinearity in the dc models for the MOS and bipolar transistors, considerable simplification in the small signal analysis of circuits often results if the nonlinear large signal models for the transistors are replaced with small signal linear models that are valid for small excursions about the Q-point. These simpler models are also extremely useful for design purposes. The question of how small the signals must be to be considered *small signals* is often avoided by both authors and practicing engineers, thus raising concern about the validity of analyses based on these models. When addressing this problem, one must consider the circuit topology, device characteristics, and Q-point as well as the maximum tolerable distortion at the output. For some circuits the small signal analysis may be justifiable only for sinusoidal signals in the 10 mV range, whereas for others it may be good for signals in the 10 V range or larger.

To distinguish between small signal, quiescent, and large signal (total) values, the following convention will be adopted. An upper case variable, or when necessary an upper case variable with an upper case subscript or a numeral, will denote the instantaneous total variable value. A lower case variable (with or

without a subscript) will denote a small signal value and an upper case variable with a lower case subscript will denote the quiescent value. The relationship between these variables is given by

$$V_C = v + V_c \tag{3.0-1}$$

where for small signal analysis V_C is assumed to be periodic with period T and the quiescent value is defined by

$$V_c = \frac{1}{T}\int_0^T V_C(t)\,dt \tag{3.0-2}$$

The small signal variable is thus the time-varying component of V_C. It is often the case that for small v, the quiescent value V_c is nearly independent of the amplitude of v. Unless specifically noted to the contrary, it will be assumed in this text that the quiescent value is independent of the amplitude of v. The relationship between V_C, V_c, and v is depicted in Fig. 3.0-2 where V_P and V_1 are constants.

The electrical behavior of linear multiple terminal networks can typically be modeled in terms of one or more established sets of transfer function parameters, such as the h parameters, y parameters, or g parameters. It will be shown later in this chapter that the y parameter (admittance parameter) model of the linear

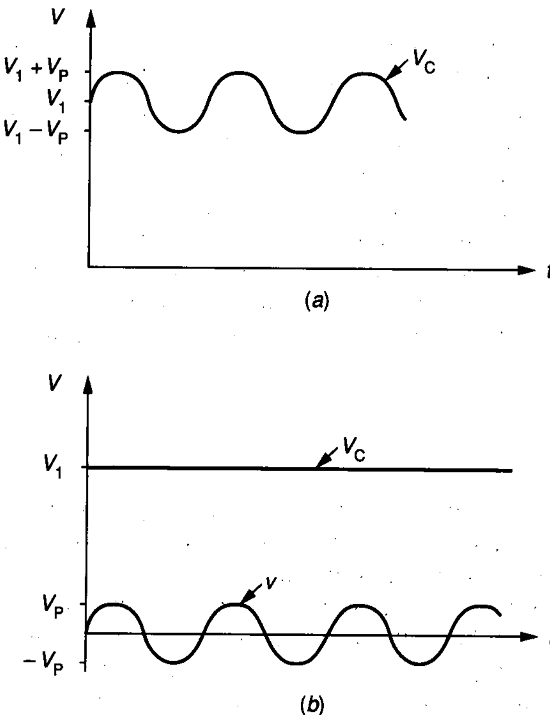


FIGURE 3.0-2
Relationship between total (V_C), small signal (v), and quiescent (V_c) values (a) total instantaneous value, (b) quiescent and small signal values.

small signal equivalent circuit of both the MOSFET and BJT exists. This model has been widely adopted for modeling these devices. The small signal linear model can usually be obtained very easily and systematically from the dc model of a device. A discussion of four-terminal and three-terminal small signal linear networks and how they can be derived from a nonlinear dc device model follows.

For the linear small signal four-terminal network shown in Fig. 3.0-3*a*, one terminal (terminal 4) is selected as a reference. With this reference and the assumption of linearity, it follows by definition that the *y*-parameters relate the terminal voltages and currents by the expressions

$$\begin{aligned} i_1 &= y_{11}v_1 + y_{12}v_2 + y_{13}v_3 \\ i_2 &= y_{21}v_1 + y_{22}v_2 + y_{23}v_3 \\ i_3 &= y_{31}v_1 + y_{32}v_2 + y_{33}v_3 \end{aligned} \tag{3.0-3}$$

where

$$y_{kj} = \left.\frac{i_k}{v_j}\right|_{v_m=0\ ,\ m\in\{1,\,2,\,3\},\ m\neq j} \tag{3.0-4}$$

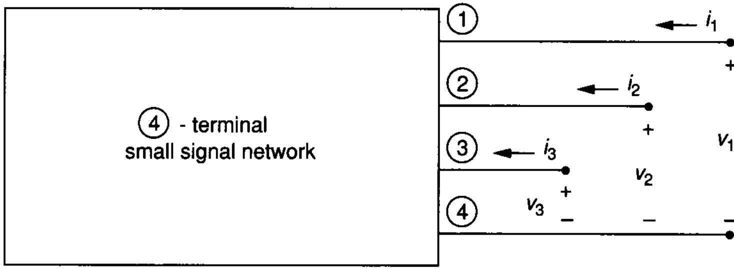
Since the small signal voltage and current variables are the time-varying part of the corresponding total terminal voltage and current variables in the parent network relative to the Q-point (see Fig. 3.0-3*b*), then it follows that the *y*–parameters can be obtained from the large signal variables and thus from the dc model by the expression

$$y_{kj} = \left.\frac{\partial I_k}{\partial V_j}\right|_{V_m=quiescent\ \ value,\ m\in\{1,\,2,\,3\}} \tag{3.0-5}$$

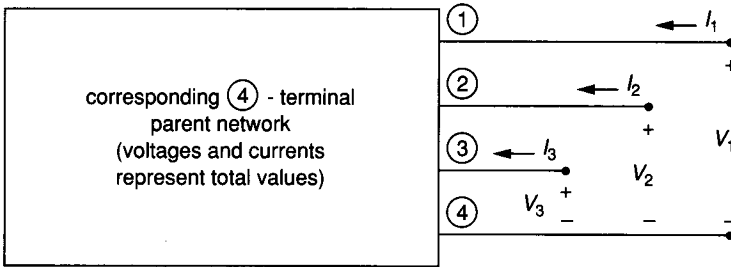
A small signal equivalent circuit of the multiport network is often found useful for circuit analysis and design. It is easy to verify that the circuit shown in Fig. 3.0-3*c* has the same *i*-*v* characteristics as the four-terminal network of Fig. 3.0-3*a*, which is characterized by Eqs. 3.0-3 and hence is electrically indistinguishable. In Fig. 3.0-3*c* and throughout this book, the same symbol is used to denote both resistance and conductance. It is assumed that from the context the reader will correctly distinguish these devices and that any conductor can be equivalently and interchangeably represented by a resistor of value equal to the reciprocal of that conductor.

It will be seen later that a four-terminal network is necessary to represent the small signal MOSFET whereas a three-terminal network is generally adequate for the BJT. The linear small signal model and the corresponding equivalent circuit for the three-terminal network can be readily obtained from the four-terminal network formulation discussed above (see Problem 3.1).

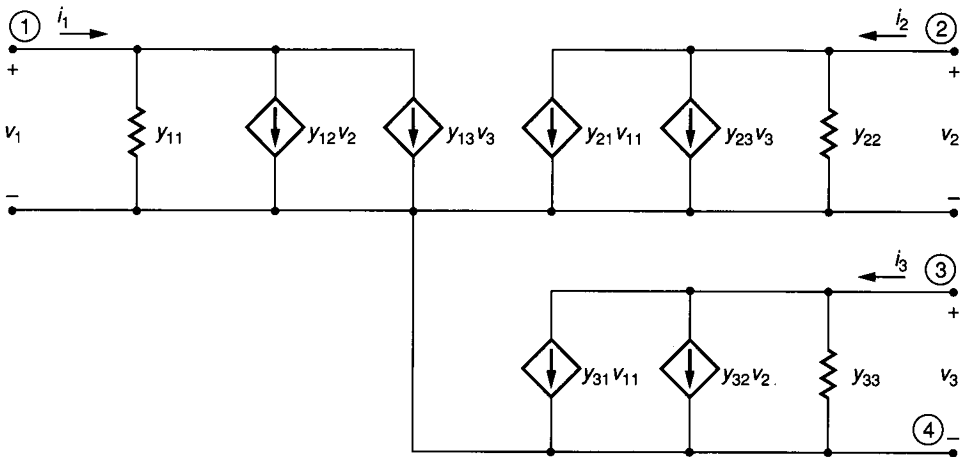
The following example is included to demonstrate mathematically how easily the small signal equivalent circuit can be obtained from the dc model. This same technique will be used for obtaining the small signal model of both the MOSFET and BJT later in this chapter.



(a)



(b)



(c)

FIGURE 3.0-3
General four-terminal network: (*a*) Small signal, (*b*) Parent network, (*c*) Equivalent circuit for small signal network.

Example 3.0-1. For the network shown in Fig. 3.0-4, the mathematical relationship between the variables V_1, V_2, I_1, and I_2 is given by the equations

$$V_1 = I_1 R_1$$

$$I_2 = \begin{cases} I_O e^{k_1 V_1} + V_2/R_2 & V_1 > 0 \\ k_2 V_1^2 + I_O + V_2/R_2 & V_1 < 0 \end{cases}$$

(*a*) Determine the dc model of the network.
(*b*) Determine the small signal model of the network for $V_1 > 0$.
(*c*) Determine the small signal model of the network for $V_1 < 0$.
(*d*) Draw the small signal equivalent circuit if the device is biased to operate at a quiescent value of $I_1 = 10$ mA and $V_2 = 4$ V. Assume $I_O = 5$ mA, $k_1 = 0.2$ V^{-1}, $k_2 = 3$ mA/V^2, $R_1 = 2$ kΩ, and $R_2 = 4$ kΩ.

Solution.
(*a*) The dc model is given by the expressions for V_1 and I_2.
(*b*) From Problem 3.1 it follows that the small signal voltage and current variables are related by

$$\begin{aligned} i_1 &= y_{11} v_1 + y_{12} v_2 \\ i_2 &= y_{21} v_1 + y_{22} v_2 \end{aligned} \qquad (3.0\text{-}6)$$

For $V_1 > 0$, it follows from (3.0-6) and the expressions for V_1 and I_2 with $V_1 > 0$ that

$$y_{11} = \left.\frac{\partial I_1}{\partial V_1}\right|_{\substack{V_1 = V_{1Q} \\ V_2 = V_{2Q}}} = \frac{1}{R_1}$$

$$y_{12} = \left.\frac{\partial I_1}{\partial V_2}\right|_{\substack{V_1 = V_{1Q} \\ V_2 = V_{2Q}}} = 0$$

$$y_{21} = \left.\frac{\partial I_2}{\partial V_1}\right|_{\substack{V_1 = V_{1Q} \\ V_2 = V_{2Q}}} = k_1 I_O e^{k_1 V_1}$$

$$y_{22} = \left.\frac{\partial I_2}{\partial V_2}\right|_{\substack{V_1 = V_{1Q} \\ V_2 = V_{2Q}}} = \frac{1}{R_2}$$

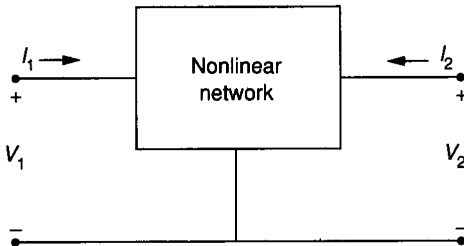


FIGURE 3.0-4
Example 3.0-1.

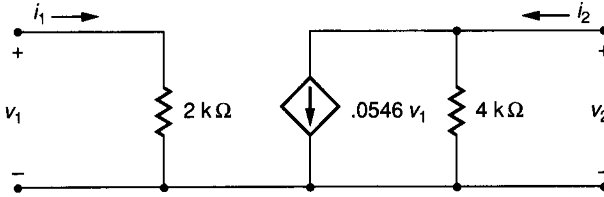


FIGURE 3.0-5
Solution of Example 3.0-1.

(*c*) From (3.0-6) and the expressions for V_1 and I_2, it follows that $y_{11} = 1/R_1, y_{12} = 0$, $y_{21} = 2k_2V_1$, and $y_{22} = 1/R_2$.

(*d*) Since $I_1 > 0$, it follows from the expression for I_2 that $V_1 > 0$ and hence from part (b) that $y_{11} = 1/(2\,\text{k}\Omega), y_{12} = 0, y_{22} = 1/(4\,\text{k}\Omega)$, and

$$y_{21} = (0.2\,\text{V}^{-1})(5\,\text{mA})e^{(0.2\,\text{V}^{-1})(10\,\text{mA})(2\,\text{k}\Omega)} = 54.6\,\text{mmho}$$

It thus follows from Problem 3.1 that the small signal equivalent circuit is as shown in Fig. 3.0-5.

3.0.3 Use of Device Models in Circuit Analysis

Electrical circuits typically are composed of one or more active devices, such as BJTs or MOSFETS, possibly along with some passive components, such as resistors or capacitors. The designer is generally interested in controlling the relationship between two or more electrical port variables. A typical two-port network is shown in Fig. 3.0-6 along with both the large signal and small signal port variables. The relationships of interest often include the large signal transfer characteristics, such as V_O versus V_I or V_I versus I_I, and the small signal transfer characteristics such as the voltage gain, $A_v = v_o/v_i$ or the input impedance, $Z_{in} = v_i/i_i$. These latter characteristics are often frequency dependent.

The analog IC designer is often simultaneously interested in both the small signal and large signal characteristics. The performance specifications the analog designer must meet are typically expressed in terms of the small signal characteristics, whereas knowledge of the large signal characteristics is necessary for biasing (setting the Q-point).

The dc transfer characteristics of a circuit are obtained by using the dc device models to characterize all devices in the network. A standard nodal analysis of the electrical network that includes the nonlinear device models is often made. It will be seen later that the dc analysis of even very simple circuits that contain BJTs or MOSFETs is often unwieldy. Relating to the unwieldy nature of the analysis of circuits containing nonlinear devices, two points deserve mention. First, most existing practical circuits are composed of very simple building blocks. The overall analysis can often be conveniently decomposed into the analysis of the individual blocks. This approach has evolved because most engineers cannot comprehend complicated nonlinear structures well enough to design structures that are not readily decomposable into simpler blocks. Second, to develop insight

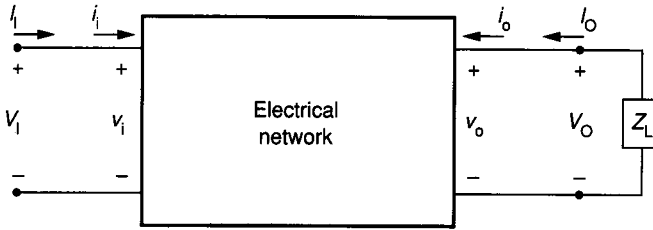


FIGURE 3.0-6
Typical electrical network structure.

into the dc performance of even simple networks, it is often necessary to make simplifying assumptions on the device models. Without these assumptions, the mathematical expressions relating the electrical parameters of interest are often so complicated that they obscure even the basic functionality of the circuit. The simplified dc analysis is, moreover, often sufficiently accurate to be useful for biasing purposes and can also provide useful insight into the small signal operation of a circuit.

The low-frequency small signal characteristics of a circuit are typically obtained in one of two ways. The first involves replacing all devices and elements with a linear small signal equivalent circuit in which the parameters in the small signal models are a function of the Q-point. (The small signal equivalents of all independent dc voltage sources are short circuits, and the small signal equivalents of all independent dc current sources are open circuits—see Problem 3.2.) The resulting circuit is linear and can be readily analyzed. The second method is based directly upon the dc transfer characteristics. If x_i and x_o are the small signal input and output variables of interest and X_I and X_O are the corresponding large signal variables, then these variables are related by the expression

$$\frac{x_o}{x_i} = \left.\frac{\partial X_O}{\partial X_I}\right|_{\text{Q-point}} \tag{3.0-7}$$

where the partial derivative of the large signal variable is evaluated at the Q-point. Equation 3.0-7 is functionally similar to Eq. 3.0-5. The results obtained from a direct small signal analysis will be identical to those obtained by differentiation of the large signal variables provided that the small signal models of the devices are obtained from the same dc model used to obtain the dc transfer characteristics.

There are, however, some distinct advantages offered by obtaining the small signal transfer characteristics directly from a small signal analysis rather than from differentiation of the dc variables. First, since the small signal equivalent circuit is linear, the analysis of the small signal equivalent circuit is typically less involved than the dc analysis. Second, more accuracy is *practically* attainable with the direct small signal analysis. The direct small signal analysis is more accurate because simplifying assumptions are often necessary in the dc analysis to maintain mathematical tractability for obtaining X_O and these assumptions make the required derivative of (3.0-7) less accurate. Only a modest increase in

complexity results if the more exact small signal device models are employed in the small signal analysis.

Finally, the frequency response of a network is expressed in terms of the small signal electrical parameters. The small signal equivalent network used for obtaining the frequency response is obtained by adding identifiable parasitic capacitors and devices to the small signal equivalent circuit and the small signal device models, respectively. Addition of the capacitors does not affect linearity, so the modelling is straightforward. No practical alternatives exist for obtaining the frequency response directly from the large signal transfer characteristics.

Example 3.0-2. Assume a two-port device is characterized by the equations

$$I_1 = 0$$

$$I_2 = hV_1^2(1 + \lambda V_2) \text{ for } V_1 > 0 \text{ and } V_2 > 0$$

This device is used in the circuit of Fig. 3.0-7.

(*a*) Obtain the small signal model of the two-port device for $V_1 > 0$ and $V_2 > 0$.

(*b*) Obtain the dc transfer characteristics of the circuit in the figure relating V_O to V_I if $V_{CC} = 5$ V, $h = 0.04$ mA/V^2, $\lambda = 0.1$ V^{-1} and $R_L = 4$ kΩ.

(*c*) Obtain parametrically the dc large signal transfer characteristics if the model is simplified by assuming $\lambda = 0$.

(*d*) Obtain v_o/v_i from the simplified dc large signal transfer characteristics of part (*c*) if the circuit is biased to operate at $V_I = 5$ V. Assume $V_{CC} = 5$ V, $\lambda = 0.1\text{V}^{-1}$, and $h = 0.04$ mA/V^2.

(*e*) Obtain v_o/v_i from the small signal analysis if the circuit is biased to operate at a Q-point of $V_I = 5$ V.

(*f*) If the device has an internal parasitic capacitance from node ② to node ④, determine the frequency response of the circuit of Fig. 3.0-7.

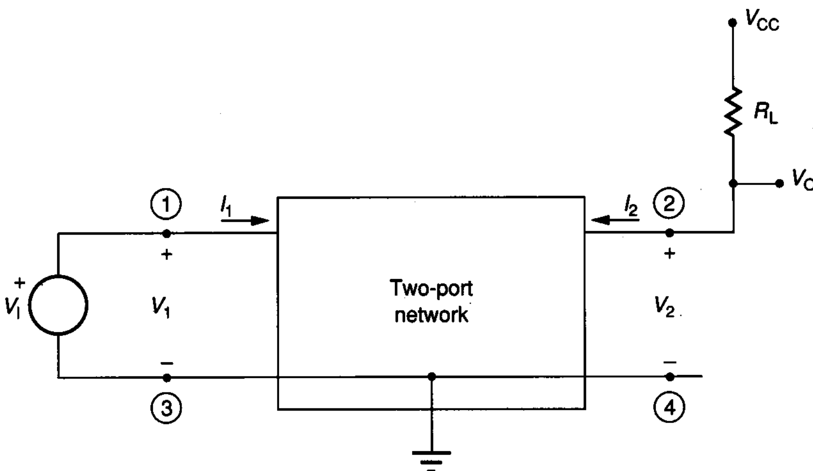


FIGURE 3.0-7
Circuit for Example 3.0-2.

Solution.

(*a*) From Eq. 3.0-5,

$$y_{21} = \left.\frac{\partial I_2}{\partial V_1}\right|_{\text{Q-pt}} = 2hV_I(1 + \lambda V_O)\Big|_{\text{Q-pt}}$$

$$y_{22} = \left.\frac{\partial I_2}{\partial V_2}\right|_{\text{Q-pt}} = \lambda h V_I^2\Big|_{\text{Q-pt}}$$

$$y_{12} = y_{11} = 0$$

We thus obtain the equivalent circuit of Fig. 3.0-8a.

(*b*) From the expression for I_2 and KVL (Kirchoff's Voltage Law) at the output, $V_O = V_{CC} - I_2R_L$, we obtain $V_O = V_{CC} - 0.16\ V_I^2(1 + \lambda V_O)$. A plot of the dc transfer characteristic appears in Fig. 3.0-8*b*. Note that this expression is nonlinear.

(*c*) If $\lambda = 0$, it follows from part (*b*) that $V_O = V_{CC} - (4\ \text{k}\Omega)hV_I^2$. Note that the simplifying assumption $\lambda = 0$ significantly simplifies the input-output relationship.

(*d*)

$$A_v = \frac{v_o}{v_i} = \left.\frac{\partial V_O}{\partial V_I}\right|_{V_I=5\text{ V}} = -2(4\ \text{k}\Omega)hV_I\Big|_{V_I=5\text{ V}}$$

If $V_I = 5\ V$, then

$$A_v = \left.\frac{\partial V_O}{\partial V_I}\right|_{V_I=5\text{ V}} = -1.6$$

(*e*) The small signal equivalent circuit is shown in Fig. 3.0-8*c*. Hence

$$\frac{v_o}{v_i} = \frac{-y_{21}}{y_{22} + g_L}$$

From part (*b*), if $V_I = 5$ V, then $V_O = 0.714$ V. Therefore from (*a*) $y_{21} = 0.429$ and $y_{22} = 0.1$. Since $g_L = 1/R_L = 0.25$ mmho, it follows that $v_o/v_i = -1.23$. Note that this more exact gain differs appreciably from that obtained by assuming $\lambda = 0$ in part (*d*).

(*f*) If a parasitic capacitor is added from node ② to node ④, we obtain the small signal equivalent circuit of Fig. 3.0-8*d*. It follows directly that

$$\frac{v_o}{v_i} = \frac{-y_{21}}{y_{22} + g_L + sC}$$

where s is the standard normalized frequency variable ($s = j\ \omega$). The circuit thus behaves as a first-order lowpass network (see the Bode plot of Fig. 3.0-8*e*) with a 3 dB cutoff frequency of

$$f_{3\text{dB}} = \frac{y_{22} + g_L}{2\pi C}$$

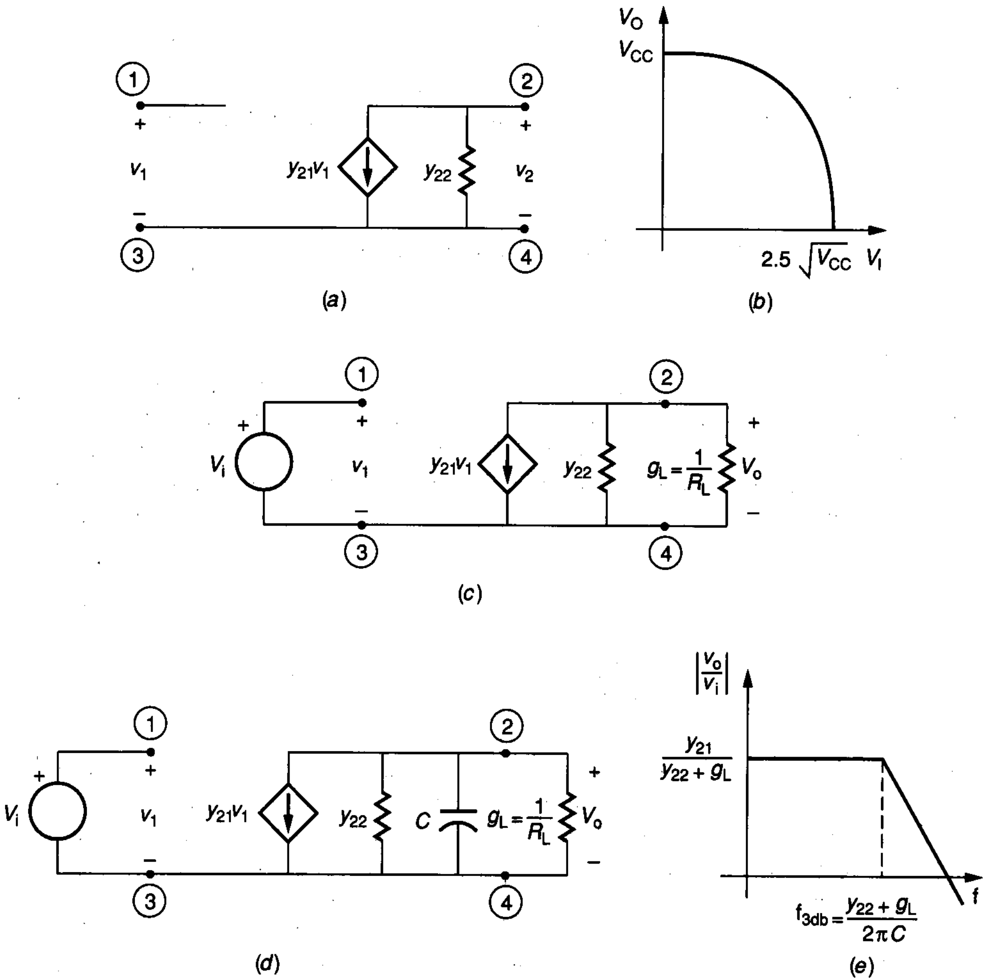


FIGURE 3.0-8
Solution of Example 3.0-2: (*a*) Small signal device model, (*b*) Large signal transfer characteristics, (*c*) Small signal equivalent circuit, (*d*) Small signal equivalent circuit including parasitic capacitor, (*e*) Magnitude of frequency response.

3.1 MOS MODELS

The n-channel and p-channel enhancement MOSFET devices along with the convention for the electrical variables established in Chapter 2 are shown in Fig. 3.1-1 along with a top view of the typical geometrical layout. The same electrical conventions and geometric gate definitions are followed for depletion devices.

The following model development will be restricted to the n-channel transistor. The p-channel development is identical with the exception of sign changes in some of the equations. Results only for the p-channel devices will be presented following the n-channel discussions.

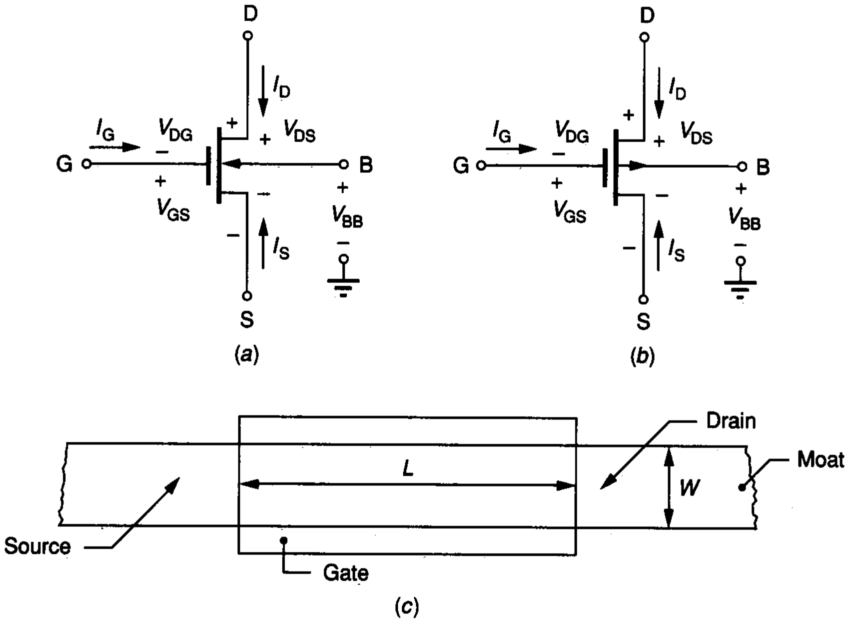


FIGURE 3.1-1
Convention for electrical variables of MOS transistors: (*a*) n-channel, (*b*) p-channel, (*c*) Top view of device geometry. Not shown is V_{BS}, the voltage from bulk (B) to source (S).

3.1.1 dc MOSFET Model

The experimentally obtained $I_D - V_{DS}$ characteristics as a function of V_{GS} for a typical MOSFET were shown in Fig. 2.2-3, and are repeated in Fig. 3.1-2. It is desired to obtain a mathematical model of the MOSFET that will accurately predict the experimental characteristics over a wide range of geometrical and process parameters as well as operating conditions.

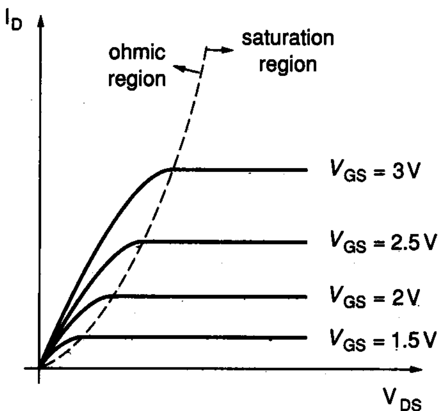


FIGURE 3.1-2
Typical output characteristics for n-channel MOSFET.

The starting point will be the dc model introduced by Sah[1] in 1964, which is given by

$$I_D = \begin{cases} \dfrac{K'W}{L}\left[(V_{GS} - V_T) - \dfrac{V_{DS}}{2}\right]V_{DS} & V_{GS} > V_T \\ 0 & \end{cases} \tag{3.1-1}$$

$$I_G = 0 \qquad\qquad V_{GS} < V_T \tag{3.1-2}$$

which was derived for small values (both positive and negative) of V_{DS}. Small values of V_{DS} correspond to the ohmic region of operation discussed in Chapter 2. In this equation, L = channel length, W = channel width, K' = transconductance parameter, and V_T = threshold voltage. L and W are design parameters and are depicted in Fig. 3.1-1*c*. Note that these refer to the *dimensions of the channel* (intersection of gate and moat). K' and V_T are process parameters and will be discussed later. Although both W and L are at the designer's control, only their ratio appears in the device model. The W/L ratio is the only geometrical design parameter available to the design engineer that affects the performance of MOS transistors. Assuming the parameters K' and V_T are constant, it can be shown (see Problem 3.4) from Eq. 3.1-1 that the device is electrically symmetric with respect to drain and source. The choice of which end of the channel to designate as source and drain is thus arbitrary. This electrical symmetry was anticipated because of the geometrical symmetry of the MOSFET as shown in Fig. 3.1-1*c*. It will be shown later that V_T is somewhat voltage dependent. The models discussed in this text that include the voltage dependence of V_T are not electrically symmetrical. Electrical symmetry is sacrificed in these models to reduce both model and computational complexity. A convention for the designation of the drain and source regions with the asymmetric model will be discussed later.

A plot of I_D versus V_{DS} for a MOSFET as given by Eq. 3.1-1 is shown in Fig. 3.1-3 for several values of V_{GS}. The portions of the curves that have negative derivatives are dashed because Sah's model is good only when the derivatives are positive. The region where (3.1-1) is valid is termed the *ohmic*, *linear*, or *active* region. The point where the partial derivative of I_D with respect to V_{DS} is zero is easily found (see Problem 3.5) from (3.1-1) to be

$$V_{DS} = V_{GS} - V_T \tag{3.1-3}$$

This value of V_{DS} causes the channel to pinch off, as mentioned in the qualitative discussion of MOSFET operation in Section 2.2.1. For $V_{DS} > V_{GS} - V_T$, the current remains practically constant (independent of V_{DS}) at the value obtained when the channel first pinched off. This value can thus be obtained by evaluating (3.1-1) at $V_{DS} = V_{GS} - V_T$ to obtain

$$I_D = \frac{K'W}{2L}(V_{GS} - V_T)^2 \tag{3.1-4}$$

which is good for $V_{DS} > V_{GS} - V_T$ and $V_{GS} > V_T$. The region of operation where (3.1-4) is valid is termed the *saturation region*. It should be noted from

(3.1-4) that I_D is independent of V_{DS}. The reader should be cautioned that the term *saturation region* for a MOSFET does not correspond to the saturation region for a BJT, as will be seen in Sec. 3.3.

For reasons that will become apparent later, the MOSFET is operated in the saturation region in most linear applications. It follows from (3.1-4) that when operating in the saturation region, the MOSFET is inherently a transconductance-type device with the input a voltage, V_{GS}, and the output a current, I_D.

The segment of the curve in the third quadrant of Fig. 3.1-3 where $V_{DS} < V_{BS}$ is also dashed. Equation 3.1-1 is no longer valid for $V_{DS} < V_{BS} - 0.5$ V because the drain substrate interface, which is actually a pn junction, becomes forward biased, causing considerable bulk current to flow. No additional details about the device model for values of $V_{DS} < V_{BS} - 0.5$ V will be given here because the devices are seldom operated in this mode. The MOSFET model based upon (3.1-1) and (3.1-4) is termed Sah's model. This is the simplest model that will be discussed here. In many situations, this model is quite tractable for hand calculations and adequate for the analytical portions of the design.

It can be shown both theoretically and experimentally that the drain current in the saturation region increases slightly in approximately a linear manner with V_{DS}. This is physically due to a slight shortening of the channel as V_{DS} is increased in the saturation region. Defining λ to be the coefficient that represents the linear dependence of I_D on V_{DS}, a more accurate expression for the drain current in the saturation region is given by

$$I_D = \frac{K'W}{2L}(V_{GS} - V_T)^2(1 + \lambda V_{DS}) \qquad V_{DS} > V_{GS} - V_T,\ V_{GS} > V_T \quad (3.1\text{-}5)$$

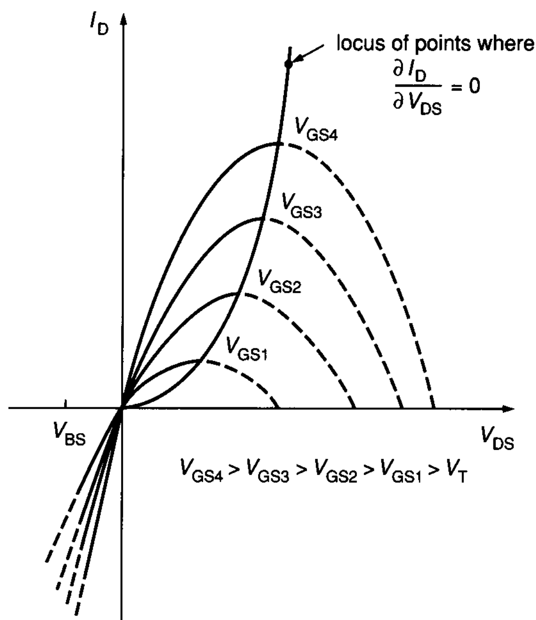


FIGURE 3.1-3
I_D–V_{DS} characteristics of MOSFET as predicted by Sah's equation.

The parameter λ (termed the channel length modulation), in Eq. 3.1-5 should not be confused with the unrelated parameter λ (discussed in Chapter 2) that was used to characterize the minimum feature size of a process. The coefficient λ is quite small for long devices but increases considerably for very short transistors. The effects of λ on MOSFET performance will be discussed later. The MOSFET model defined by (3.1-1) and (3.1-5), which differs from the model of Sah only through the parameter λ, is termed the Shichman-Hodges model[2]. Some authors also multiply the drain current in the ohmic region by the parameter $(1 + \lambda V_{DS})$. We have not done so because there is no physical justification to do so, since the model becomes more complicated in the ohmic region, and since λV_{DS} is so small in the ohmic region that it has little effect on I_D. By not making such a modification, however, we are creating a slight discontinuity in our device model at $V_{DS} = V_{GS} - V_T$ (see Problem 3.10). This discontinuity could cause problems in computer simulations, which require continuity of the functions as well as of the derivatives, but should not cause much of a problem for hand calculations since the device is seldom operated at the saturation–ohmic transition region. The discontinuity due to λ effects is depicted in Fig. 3.1-4. Note that all projections from the saturation region into Quadrant 2 intersect the V_{DS} axis at $-1/\lambda$. It should be emphasized that the discontinuity in the model has been introduced only for reducing model complexity and that models used in computer simulators, such as the program SPICE which will be discussed in Chapter 4, are invariably continuous at this transition.

The threshold voltage, V_T, is somewhat dependent upon the bulk–source voltage. This dependence can be anticipated since the bulk–channel voltage will affect the carriers in the depletion region under the gate, which in turn affect the voltage that must be applied to the gate to form the inversion layer. This dependence can be approximated by

$$V_T = V_{T0} + \gamma\left(\sqrt{\phi - V_{BS}} - \sqrt{\phi}\right) \tag{3.1-6}$$

where V_{BS} is the bulk–source voltage and V_{T0}, γ, and ϕ are process parameters:

$$V_{T0} = \text{threshold voltage for } V_{BS} = 0$$
$$\gamma = \text{bulk threshold parameter}$$
$$\phi = \text{strong inversion surface potential}$$

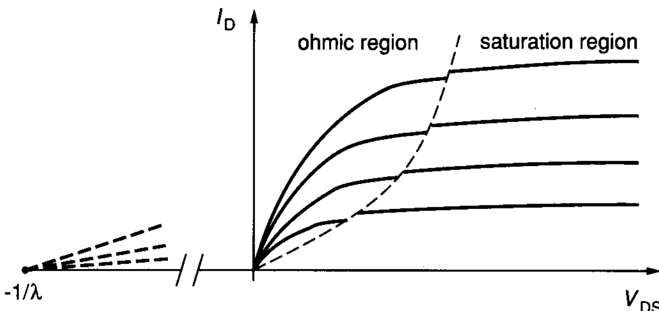


FIGURE 3.1-4
Discontinuity in model due to λ effects.

At this point it appears that the design engineer can control V_T with V_{BS} and thus have a second means (in addition to the W/L ratio) of controlling the characteristics of the MOS transistor. This V_{BS} dependence, which typically is controlled through the bulk voltage, V_{BB}, actually offers very little additional flexibility to the designer for two reasons. First, the threshold voltage is only weakly dependent upon V_{BS}. Second, since the bulks of all transistors are common in an NMOS process, the bulk voltage for the entire IC is established at a single value. Tailoring the characteristics of individual transistors is therefore impossible. For CMOS it is conceptually possible to have individual control of those MOSFETS formed in the tubs (wells) by independently controlling the tub voltages. For practical considerations, however, all tubs of the same type (p or n) are often tied together and connected to a common voltage, thus removing this potential flexibility in controlling individual device characteristics although individual tub potentials are established in some analog circuits. A typical plot of V_T versus V_{BS} appears in Fig. 3.1-5. Note that the change in V_T can be quite significant for large V_{BS}. The effect becomes even worse with larger γ.

Note also that V_T as modeled by (3.1-6) is not symmetric with respect to drain and source. This causes the expression for I_D to be asymmetric with respect to drain and source. This is somewhat disturbing because of the geometrical symmetry of the MOSFET. Models of the MOSFET that maintain the electrical symmetry are available[3], but the increased complexity makes them unsuitable for most hand manipulations. The approximation for V_T given in (3.1-6) gives reasonably close agreement between theoretical and experimental results while still maintaining an acceptable degree of mathematical tractability in hand calculations. The convention usually followed in making the drain and source designations is to label these regions so that $V_{DS} > 0$. With this convention, the drain and source designations may not remain fixed in some analyses.

For n-channel transistors the devices are termed enhancement if $V_{T0} > 0$ and depletion if $V_{T0} < 0$. p-channel MOSFETs are enhancement mode devices if $V_{T0} < 0$ and depletion type if $V_{T0} > 0$.

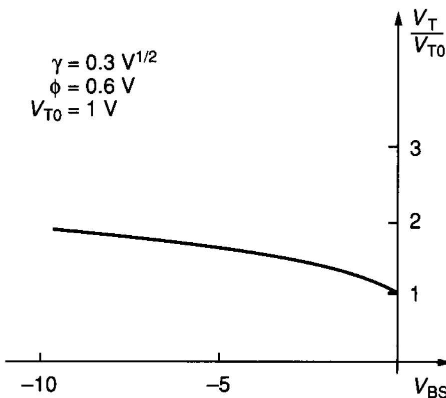


FIGURE 3.1-5
Effects of V_{BS} on threshold voltage.

The transconductance parameter, K', can be expressed in terms of other physical process parameters by

$$K' = \mu C_{ox} \tag{3.1-7}$$

where C_{ox} = capacitance density (capacitance per unit area) of the gate-channel capacitor, and μ = channel mobility.

A summary of the MOSFET model for both n-channel and p-channel devices that is adequate for most dc hand calculations appears in Table 3.1-1. For the purpose of subsequent calculations, process parameters for typical $3\,\mu$ NMOS and CMOS processes such as those described in Section 2.2.1 using the design rules of Tables 2A.2 and 2B.2 are given in Table 3.1-2. This simplified parameter set is compatible with the more detailed descriptions of Tables 2A.4 and 2B.4.

TABLE 3.1-1
Low-frequency MOSFET model

Expression	Condition	Eq.
n-channel MOSFET		
$I_G = 0$		(1)
$I_D = 0$	$V_{GS} < V_T$ (cutoff), $V_{DS} \geq 0$	(2)
$I_D = \dfrac{K'W}{L}\left(V_{GS} - V_T - \dfrac{V_{DS}}{2}\right) V_{DS}$	$V_{GS} > V_T$, $0 < V_{DS} < V_{GS} - V_T$ (ohmic)	(3)
$I_D = \dfrac{K'W}{2L}(V_{GS} - V_T)^2 (1 + \lambda V_{DS})$	$V_{GS} > V_T$, $V_{DS} > V_{GS} - V_T$ (saturation)	(4)
where $V_T = V_{T0} + \gamma(\sqrt{\phi - V_{BS}} - \sqrt{\phi})$		(5)
p-channel MOSFET		
$I_G = 0$		(6)
$I_D = 0$	$V_{GS} > V_T$ (cutoff), $V_{DS} \leq 0$	(7)
$I_D = -\dfrac{K'W}{L}\left(V_{GS} - V_T - \dfrac{V_{DS}}{2}\right) V_{DS}$	$V_{GS} < V_T$, $0 > V_{DS} > V_{GS} - V_T$ (ohmic)	(8)
$I_D = -\dfrac{K'W}{2L}(V_{GS} - V_T)^2 (1 - \lambda V_{DS})$	$V_{GS} < V_T$, $V_{DS} < V_{GS} - V_T$ (saturation)	(9)
where $V_T = V_{T0} - \gamma(\sqrt{\phi + V_{BS}} - \sqrt{\phi})$		(10)

Design parameters:
- W = channel width
- L = channel length

Process parameters:
- K' = transconductance parameter
- V_{T0} = threshold voltage for $V_{BS} = 0$
- γ = bulk threshold parameter
- ϕ = strong inversion surface potential
- λ = channel length modulation parameter

K', γ, ϕ, and λ are positive for both n-channel and p-channel transistors. n-channel transistors are termed enhancement if $V_{T0} > 0$ and depletion if $V_{T0} < 0$. p-channel transistors are termed enhancement if $V_{T0} < 0$ and depletion if $V_{T0} > 0$.

TABLE 3.1-2
Typical process parameters for 3 μ NMOS and CMOS processes

Parameter	n-channel	p-channel	units
K'	24	8	μ A/V^2
V_{TOE}	0.75	−0.75	V
V_{TOD}	−3	3	V
$\gamma_{(CMOS)}$	0.8	0.4	V$^{1/2}$
$\gamma_{(NMOS)}$	0.4		V$^{1/2}$
ϕ	0.6	0.6	V
λ	0.01	0.02	V^{-1}
θ (for short channels)	0.2	0.2	V^{-1}

Plots of I_D versus V_{DS} for Sah's model and the Shichman-Hodges model are compared in Fig. 3.1-6. In this plot, the Shichman-Hodges model includes the $(1 + \lambda V_{DS})$ multiplier in the ohmic region to maintain continuity at the transition.

Many algorithms exist for extracting model parameters from experimental data. Since the models do not perfectly match the physical devices, the various parameter extraction schemes will not yield identical parameter values. A parameter extraction scheme that gives reasonably good correlation between experimental and theoretical results in the neighborhood of the intended operating point should be adopted (see Problem 3.6).

The operating regions of the MOSFET in the I_D–V_{DS} plane are shown in Fig. 3.1-7. The near linear relationship between V_{DS} and I_D can be seen in the ohmic (linear) region. The dependence of I_D on V_{GS}, and essential independence from V_{DS}, should be apparent in the saturation region.

Equivalent circuits for the dc operation of the MOSFET are shown in Fig. 3.1-8. The first circuit is valid for small V_{DS} in the ohmic region. The model that corresponds to this circuit can be obtained from the expression for I_D in the ohmic

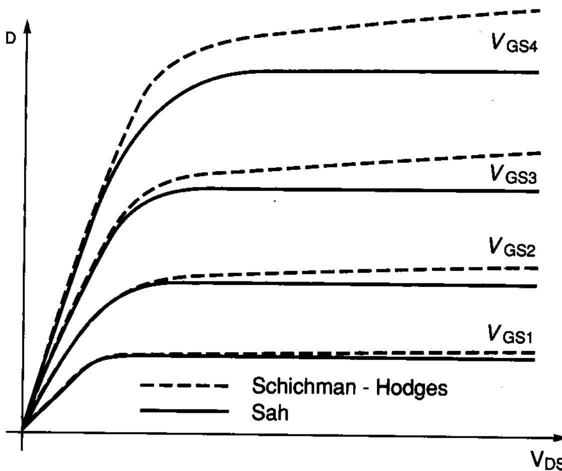


FIGURE 3.1-6
Comparison of Sah's model and the Shichman-Hodges model.

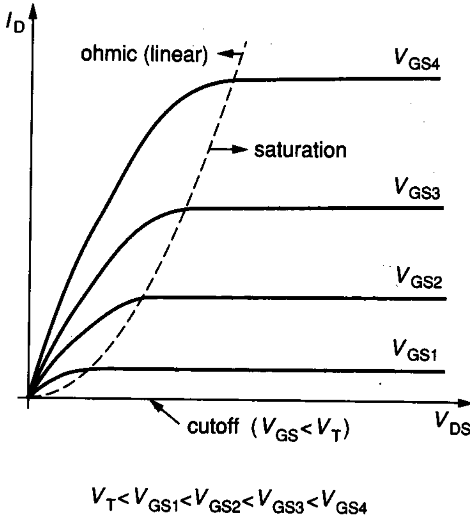


FIGURE 3.1-7
Regions of operation of MOSFET in I_D–V_{DS} plane.

region (Eq. 3.1-1) by assuming $V_{DS}/2$ is negligible compared to $(V_{GS} - V_T)$. With this assumption a linear relationship between I_D and V_{DS} ensues and results in an equivalent FET resistance of

$$R_{FET} \simeq \frac{L}{W K'(V_{GS} - V_T)} \tag{3.1-8}$$

As V_{DS} increases towards $(V_{GS} - V_T)$, it follows from (3.1-1) that the relationship between I_D and V_{DS} becomes nonlinear. For many applications the nonlinear

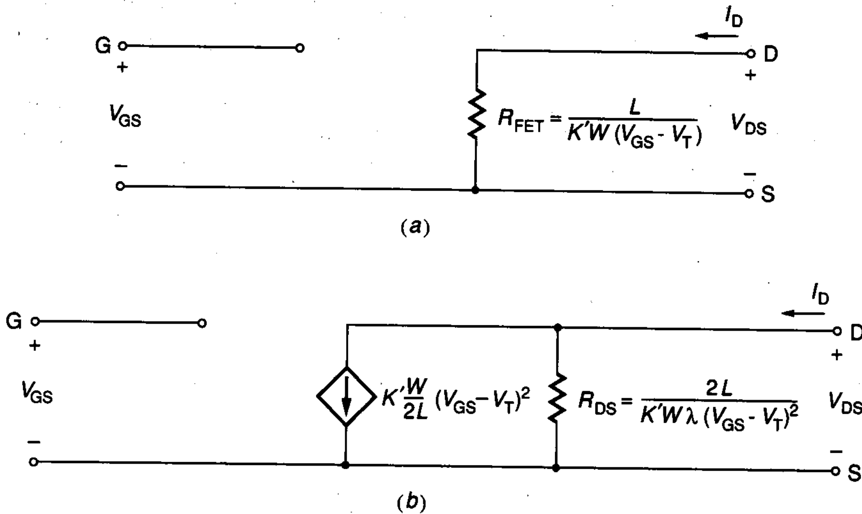


FIGURE 3.1-8
Equivalent circuits of MOSFET for dc operation: (*a*) Ohmic region (small V_{DS}), (*b*) Saturation region.

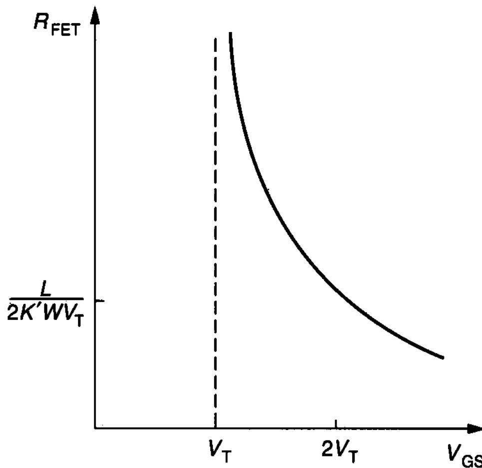


FIGURE 3.1-9
MOSFET resistance versus V_{GS} in the ohmic region.

effects, which generally cause harmonic distortion, are small provided $V_{DS} < (V_{GS} - V_T)/3$. The MOSFET is often used in the ohmic region as a Voltage Variable Resistor (VVR) by using V_{GS} as the controlling voltage. A plot of R_{FET} versus V_{GS} obtained from Eq. 3.1-8 is depicted in Fig. 3.1-9. Note that a very large adjustment range is possible but that R_{FET} is very sensitive to small changes in V_{GS} for values of V_{GS} close to V_T. In addition to the high sensitivity for V_{GS} close to V_T, two other concerns practically preclude utilizing this device as a very high impedance resistor in this region. First, the model we introduced in Table 3.1-1, which was used to derive (3.1-8), is not very accurate for small $(V_{GS} - V_T)$. A more accurate model in this region based upon what is called weak inversion operation is discussed in a later section. Second, the assumption that $|V_{DS}| < V_{GS} - V_T$ forces the allowable signal swing across the resistor to approach 0 as V_{GS} approaches V_T. This practically limits where this "large" resistor can be used.

At this point, it is assumed that the VVR is used in applications where $I_D > 0$. Extension to bidirectional currents can be made. Details can be found in Sec. 3.1.7.

The equivalent circuit of Fig. 3.1-8*b* is valid in the saturation region. This can be verified by observing that the voltage and current characteristics of the circuit agree with the MOSFET model in the saturation region defined by Eq. 3.1-5. The effects of the resistor R_{DS} in this equivalent circuit are negligible in most dc applications. This dc equivalent circuit, which is useful for biasing, is highly nonlinear and should not be confused with the small signal equivalent circuit, which will be discussed in the following section.

Example 3.1-1. Using the typical NMOS process parameters of Table 3.1-2,

(*a*) Determine the output voltage for the circuit shown in Fig. 3.1-10 if $R = 15$ kΩ.

(*b*) Determine the maximum value of R allowable to keep M1 operating in the saturation region.

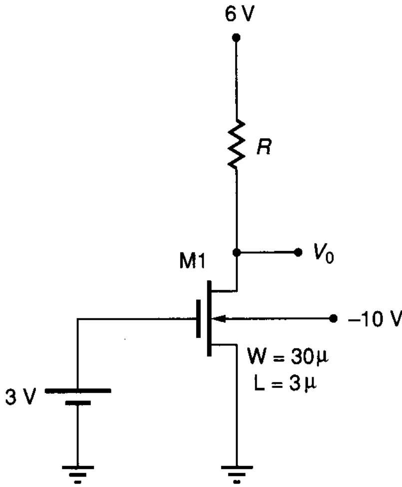


FIGURE 3.1-10
Example 3.1-1.

Solution.

(*a*) Initially the region of operation of M1 must be determined. To make this determination, it is necessary to obtain V_T. From (3.1-6) it follows that

$$V_T = 0.75\text{ V} + 0.4\text{ V}^{1/2}(\sqrt{0.6\text{ V} - (10\text{ V})} - \sqrt{0.6\text{ V}}) = 1.74\text{ V}$$

Note that this differs considerably from $V_{T0} = 0.75$ V! Since $V_{GS} = 3$ V, it can be concluded that M1 is in either the ohmic or saturation region. At this point, it is expedient to assume a state for M1, solve for V_O, and then either verify the assumption is valid or conclude that the device is operating in the alternate state.

With this approach, assume M1 is in the saturation region. Then, neglecting λ effects,

$$\begin{aligned} I_{D1} &= K'\frac{W}{2L}(V_{GS} - V_T)^2 \\ &= 24\ \mu\text{A/V}^2\left(\frac{30\ \mu}{3\ \mu}\right)\left(\frac{1}{2}\right)(3\text{ V} - 1.74\text{ V})^2 \\ &= 190.5\ \mu\text{A} \end{aligned}$$

Thus

$$V_O = 6\text{ V} - I_{D1}R = 3.14\text{ V}$$

It remains to verify the region of operation. Since $V_{DS} = V_O$, it follows that $V_{DS} > V_{GS} - V_T$. Thus, the initial assumption on the region of operation was valid and the value of $V_O = 3.14$ V is correct.

(*b*) The maximum value of R (R_{MAX}) for saturation region operation will be that value which makes $V_{DS} = V_{GS} - V_T$. Hence,

$$6\text{ V} - (190.5\ \mu\text{A})R_{MAX} = (3\text{ V} - 1.74\text{ V})$$

and thus

$$R_{MAX} = 24.9\text{ k}\Omega.$$

Example 3.1-2. Determine the size of M2 in Fig. 3.1-11 with minimum area so that $V_O = 3$ V. Assume M1 is minimum size. Use the NMOS process discussed in Sec. 2.2-1 along with the design rules of Appendix 2A (with a feature size $\lambda = 1.5\ \mu$) and the process parameters of Table 3.1-2. Neglect λ effects.

Solution. M1 will be considered first. Since $V_{BS} = 0$, it follows that $V_{T1} = V_{T0} = 0.75$ V. The region of operation for M1 must be determined. Since $V_{GS1} = 2$ V $> V_{T1}$, it can be concluded that M1 is not cutoff. Checking

$$\frac{V_{DS1}}{(V_{GS1} - V_{T1})} = \frac{3}{(2 - 0.75)} = 2.4 > 1$$

it can be concluded that M1 is operating in the saturation region. Thus,

$$I_{D1} = \frac{K'W_1}{2L_1}(V_{GS1} - V_{T1})^2$$

Considering now M2, $V_{BS2} = -3$ V (since $V_{BS1} = 0$ and $V_{S2} = V_O$). From Table 3.2-1, the nominal threshold voltage of M2, V_{T20}, is $V_{T0D} = -3$ V. Therefore,

$$V_{T2} = V_{T20} + \gamma\left[\sqrt{\phi - V_{BS}} - \sqrt{\phi}\right] = -2.55\text{ V}$$

To check the region of operation of M2, observe that $V_{GS2} = 0 > V_{T2}$, so M2 is in either the active or saturation region. Checking

$$\frac{V_{DS2}}{V_{GS2} - V_{T2}} = \frac{5}{[0 - (-2.55)]} = 1.96 > 1$$

Thus, it can be concluded that M2 is also operating in the saturation region. Hence

$$I_{D2} = \frac{K'W_2}{2L_2}(V_{GS2} - V_{T2})^2 = \frac{K'W_2}{2L_2}V_{T2}^2$$

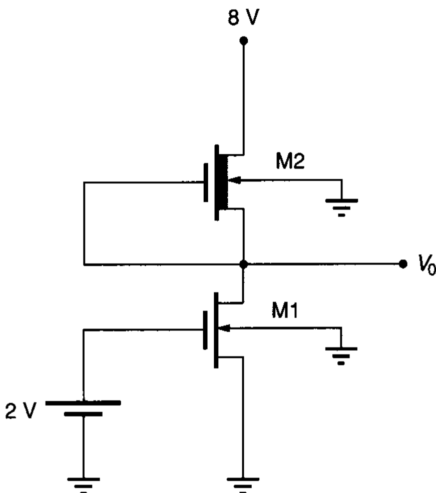


FIGURE 3.1-11
Example 3.1-2.

Since $I_{D1} = I_{D2}$, it follows from equating currents that

$$\frac{W_2/L_2}{W_1/L_1} = \left(\frac{V_{GS1} - V_{T1}}{V_{T2}}\right)^2 = 0.24$$

To minimize the area of M2, it is necessary to minimize the product of W_2 and L_2 while still satisfying the design rules of the process. From the design rules of Table 2A.2 with $\lambda = 1.5\ \mu$, it follows that $W_{min} = 3\ \mu$ and $L_{min} = 3\ \mu$. Since M1 is minimum size, $W_1/L_1 = 1$. If one were to pick $L_2 = L_{min}$, then solving the previous equation for W_2 yields $W_2 = 0.72\ \mu$, which violates the minimum width of the moat as specified in the design rules. Picking $W_2 = W_{min} = 3\ \mu$, we obtain $L_2 = 12.5\ \mu$, which is acceptable within the existing design rules.

Example 3.1-3. A p-well CMOS inverter is shown in Fig. 3.1-12. Assuming the typical process parameters of Table 3.1-2 were used for the process, that the substrate is connected to ground, and that the value of V_{DD} is 6 V, obtain an expression for and plot V_O versus V_I for $0\ V \leq V_I \leq 6\ V$.

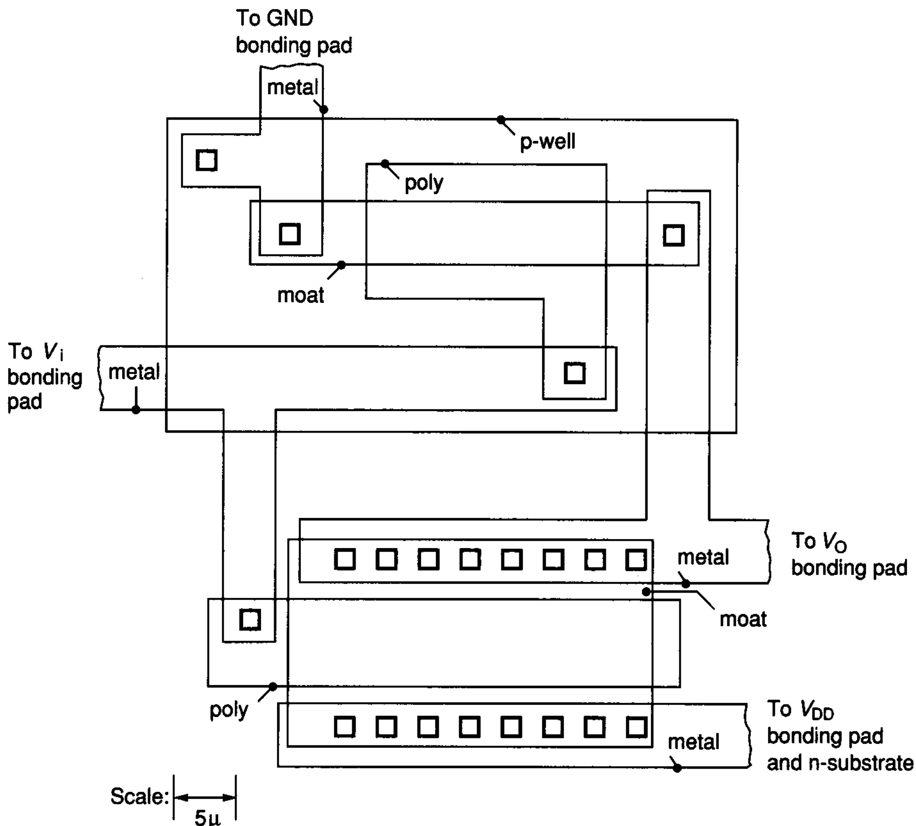


FIGURE 3.1-12
Example 3.1-3—CMOS inverter.

Solution. Initially one must obtain the circuit schematic along with device sizing information from Fig. 3.1-12. The equivalent circuit is shown in Fig. 3.1-13*a* along with the component sizing information obtained from the layout.

The threshold voltages of both M1 and M2 are needed to determine the regions of operation of the devices as well as for the analysis. Since the bulk of M2 is tied to the source of M2 and since the p-well (which also serves as the bulk for M1) is connected to the ground, it follows that $V_{BS} = 0$ for both M1 and M2. The threshold voltages for M1 and M2 are thus given respectively by V_{TON} and V_{TOP}, which can be obtained, along with the remaining process parameters, from Table 3.1-2. The balance of the analysis presented here is parameterized to make it more general; quantitative results follow directly by substituting the numerical parameters into the symbolic expressions.

The regions of operation for M1 and M2 must be obtained. Since V_I varies from 0 to 6 V, it can be readily concluded that both M1 and M2 will operate in at least two different operating regions in this circuit.

Consider first the case where $V_I < V_{TON}$. It follows that M1 is cutoff and thus from (2) of Table 3.1-1 that $I_{D1} = 0$. Since $V_{GS2} = V_I - V_{DD} < V_{TOP}$ for the parameter values listed in Table 3.1-2, it follows that M2 is in either the saturation or ohmic region. If M2 were in the saturation region, it would follow from (9) of Table 3.1-1 that the drain current of M2 would be

$$I_{D2} = -\frac{(K_2'W_2)}{2L_2}(V_{GS2} - V_{TOP})^2(1 - \lambda[V_O - V_{DD}])$$

However, since $I_{D2} = -I_{D1} = 0$ and since $V_{GS2} \neq V_{TOP}$ and $V_O \neq 1/\lambda + V_{DD}$, it must be concluded that M2 is in the ohmic region and that

$$I_{D2} = \frac{K'W_2}{L_2}\left([V_I - V_{DD} - V_{TOP}] - \frac{[V_O - V_{DD}]}{2}\right)(V_O - V_{DD}) = 0$$

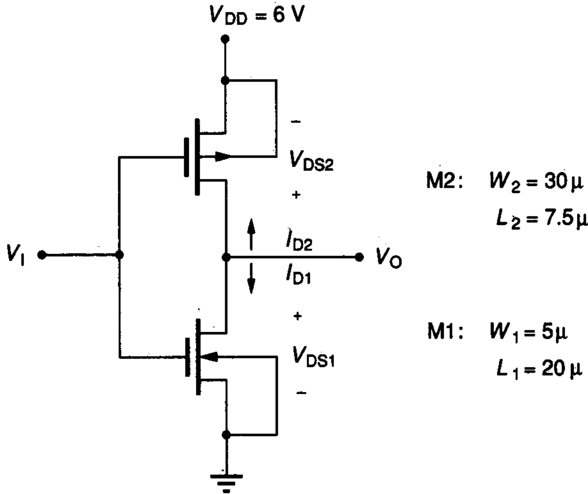
which will happen for all $V_I < V_{TON}$ only if $V_O = V_{DD}$. The portion of the transfer characteristics where M1 is cutoff and M2 is ohmic is shown as region ① in Fig. 3.1-13*b*.

If V_I is increased beyond V_{TON}, M1 will leave the cutoff region and immediately enter the saturation region. Since V_{DS2} is very small at this transition, M2 will remain in the ohmic region until V_O drops to $V_I - V_{TOP}$ (the condition where $V_{DS2} = V_{GS2} - V_{T2}$). The line $V_O = V_I - V_{TOP}$, which separates the ohmic and saturation regions for M2, is shown in the figure. Since M1 is in the saturation region and M2 is in the ohmic region, it follows from equating drain currents that

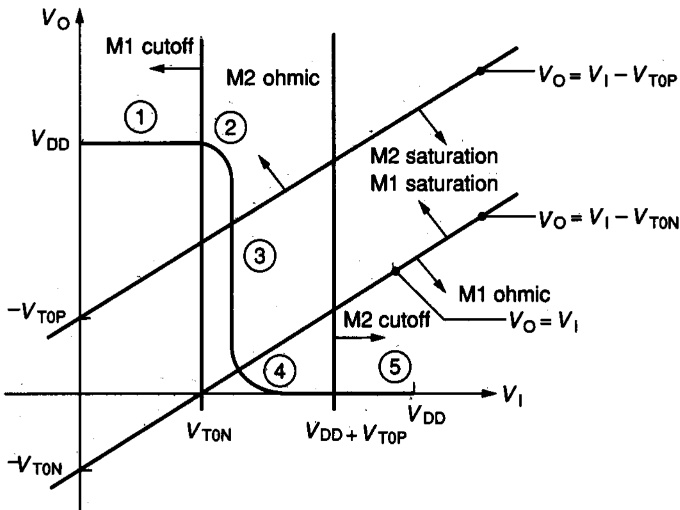
$$K_1'\frac{W_1}{2L_1}(V_I - V_{TON})^2(1 + \lambda V_O) = \frac{K_2'W_2}{L_2}\left\{\left([V_I - V_{DD} - V_{TOP}] - \frac{[V_O - V_{DD}]}{2}\right)(V_O - V_{DD})\right\}$$

This relationship is somewhat unwieldy for hand calculations but does provide the desired relationship between V_O and V_I. It is shown as region ② in Fig. 3.1-13*b*.

If V_I is increased further, M2 will enter saturation and M1 will remain in the saturation region provided $V_{DS1} > V_{GS1} - V_{T1}$ (i.e., provided $V_O \geq V_I - V_{TON}$). A plot of $V_O = V_I - V_{TON}$ is also shown on Fig. 3.1-13*b*. With both M1 and M2



(a) Equivalent circuit



(b) Transfer characteristics

FIGURE 3.1-13
Solution of Example 3.1-3.

in saturation it follows by equating drain currents that V_O and V_I are related by the expression

$$K_1'\frac{W_1}{2L_1}(V_I - V_{TON})^2(1 + \lambda V_O) = K_2'\frac{W_2}{2L_2}(V_I - V_{DD} - V_{TOP})^2[1 + \lambda(V_O - V_{DD})]$$

This relationship between V_O and V_I is shown as region ③ in Fig. 3.1-13*b*.

Increasing V_I further causes a further decrease in V_O, forcing M1 into the ohmic region and driving M2 deeper into saturation. In this region, which is depicted

as region ④ in Fig. 3.1-13*b*, and which is valid until M2 enters cutoff, the relationship between V_O and V_I is given by

$$\frac{K_1' W_1}{L_1}\left[(V_I - V_{TON}) - \frac{V_O}{2}\right]V_O = K_2' \frac{W_2}{2L_2}(V_I - V_{DD} - V_{TOP})^2[1 + \lambda(V_O - V_{DD})]$$

For $V_I > V_{DD} + V_{TOP}$, M2 is in cutoff, forcing I_{D2} and, correspondingly, V_O to 0. This is shown as region ⑤ in Fig. 3.1-13*b*.

It should be noted that if the value of V_{DD} were sufficiently reduced or the values of V_{TON} or V_{TOP} were changed by a large enough amount, some of the five regions of operation indicated in Fig. 3.1-13*b* would not occur (see Problem 3.7).

3.1.2 Small Signal MOSFET Model

The small signal model of the MOS transistor will now be obtained. From Sec. 3.0.2 it can be concluded that the small signal model can be obtained directly from the dc model summarized in Table 3.1-1. Since there are three regions of operation identified in the dc model, there are different small signal models for the MOSFET corresponding to each of these regions. In the cutoff region the drain current is essentially zero, resulting in a trivial small signal model. Since the MOSFET is typically not biased for small signal operation in the ohmic region because of performance limitations, the small signal model for operation in this region will not be developed here (but see Problem 3.8). Most small signal applications employ the MOSFET biased in the saturation region. In the saturation region, it follows from (4) of Table 3.1-1 that

$$I_D = \frac{K'W}{2L}(V_{GS} - V_T)^2(1 + \lambda V_{DS})$$

and (3.1-9)

$$I_G = I_B = 0$$

It remains to find y_{kj}, k, and $j \in \{1, 2, 3\}$ as defined in Sec. 3.0. Convention has resulted in the selection of the source node as the reference. For notational convenience, g, d, and b rather than 1, 2, and 3 will be used to designate the gate, drain, and bulk nodes of the MOSFET, respectively. The parameter y_{dg}, often termed g_m, is generally the dominant parameter in the model. From (3.0-5) and (3.1-9),

$$y_{dg} = g_m = \left.\frac{\partial I_D}{\partial V_{GS}}\right|_{\substack{V_{DS}=V_{DSQ}\\ V_{BS}=V_{BSQ}\\ V_{GS}=V_{GSQ}}} = \left.\frac{K'W}{L}(V_{GS} - V_T)(1 + \lambda V_{DS})\right|_{\substack{V_{DS}=V_{DSQ}\\ V_{BS}=V_{BSQ}\\ V_{GS}=V_{GSQ}}} \qquad (3.1\text{-}10)$$

Upon relating $V_{GS} - V_T$ to the quiescent drain current and noting that the λ effects are typically negligible in this model, it follows that

$$g_m \simeq \sqrt{\frac{2K'W}{L}}\sqrt{|I_{DQ}|} \qquad (3.1\text{-}11)$$

TABLE 3.1-3
Nonzero small signal model parameters for MOSFET

$$g_m = \sqrt{\frac{2K'W}{L}}\sqrt{|I_{DQ}|}$$

$$g_{mb} = \eta g_m$$

$$g_{ds} \simeq \lambda\,|I_{DQ}| \qquad (r_{ds} = 1/g_{ds})$$

where $$\eta = \frac{\gamma}{2\sqrt{\phi - V_{BSQ}}}$$

The other parameters can readily be obtained (see Problem 3.18) from Eqs. 3.1-9 and 3.0-5. The nonzero parameters are summarized in Table 3.1-3, and the small signal model equivalent circuit appears in Fig. 3.1-14. It should be noted that if $v_{bs} \neq 0$, any ac signals that appear on the bulk will also modulate the drain current. Although g_{mb} is typically considerably less than g_m, it will be seen in later chapters that g_{mb} is one of the major factors that limits the performance of many circuits. Since the current $g_m v_{gs}$ typically dominates the drain current, it should be apparent that the small signal MOSFET is inherently a good transconductance amplifier. Its performance directly as a current amplifier, voltage amplifier, or transresistance amplifier would be quite poor.

Example 3.1-4. For the circuit of Fig. 3.1-15, obtain
(*a*) The quiescent output voltage.
(*b*) The small signal steady state output voltage.
(*c*) The total output voltage.
Use the NMOS process parameters of Table 3.1-2.

Solution.
(*a*) Following an approach similar to that of Example 3.1-1, it can be concluded that the MOSFET is operating in the saturation region with $V_T = 1.085$ V. The quiescent drain current is approximately (by neglecting λ effects) $I_{DQ} \simeq 44.0\,\mu$A. The quiescent output voltage is 3.60 V. If λ effects are included, one obtains $I_{DQ} = 46.79\,\mu$A and $V_{OQ} = 3.321$ V.

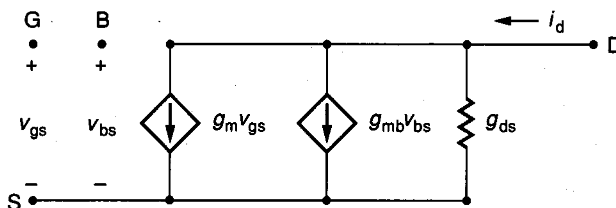


FIGURE 3.1-14
Small signal model equivalent circuit for MOSFET.

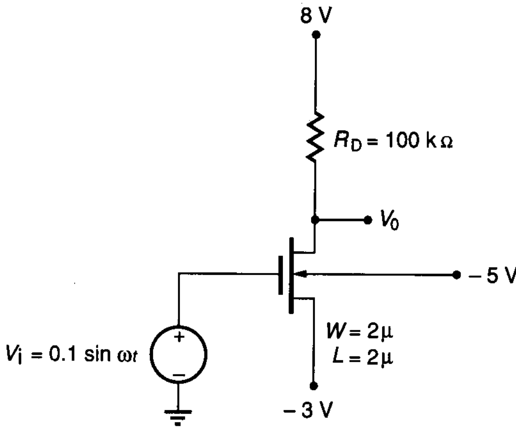


FIGURE 3.1-15
Example 3.1-4.

(*b*) The small signal equivalent circuit shown in Fig. 3.1-16 can be readily obtained from Fig. 3.1-14 and Fig. 3.1-15. From Table 3.1-3 it follows that

$$g_m = \sqrt{2\frac{K'W}{L}}\sqrt{|I_{DQ}|} = 4.60 \times 10^{-5} \text{ mho}$$

$$g_{ds} = 4.68 \times 10^{-7} \text{ mho}$$

$$g_{mb} = 5.71 \times 10^{-6} \text{ mho}$$

Since $v_{bs} = 0$, the small signal voltage gain is $-g_m/g_{ds} = -4.39$, and hence the small signal output voltage is

$$v_o(t) = -.439 \sin \omega t$$

(*c*) The total output voltage is the sum of the small signal and quiescent values, which is simply

$$V_o(t) = 3.32 \text{ V} - .439 \sin \omega t$$

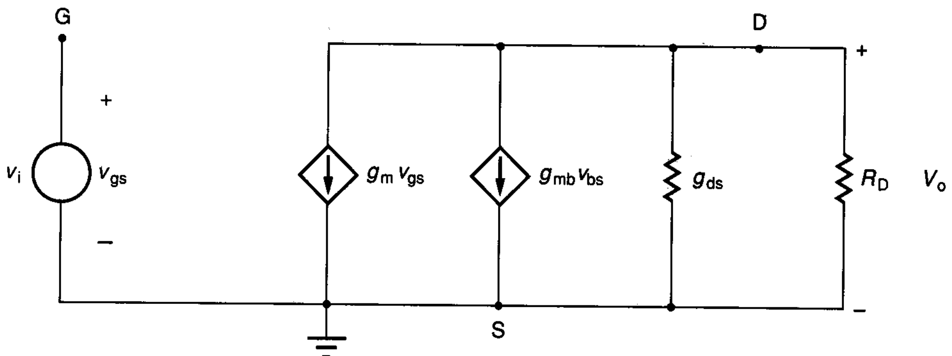


FIGURE 3.1-16
Solution of Example 3.1-4.

3.1.3 High Frequency MOSFET Model

At high frequencies both the dc and small signal models of the MOSFET introduced in the previous sections are generally considered inadequate. These limitations are to a large extent attributable to the unavoidable parasitic capacitances inherent in existing MOS structures. These parasitic capacitances can be divided into two groups. The first group is composed of those parasitic capacitors formed by sandwiching an insulating dielectric of fixed geometric dimensions between two conductive regions. The capacitance of these types of devices remains essentially constant for local changes in the voltage applied to the plates of the capacitor. Assuming the area of the normally projected intersection of the capacitor plates is A and that the distance between the plates is constant with thickness d, then this capacitance is given by the expression

$$C = \epsilon \frac{A}{d} \tag{3.1-12}$$

where ϵ is the permittivity of the dielectric material separating the plates. Often it is more convenient to combine ϵ/d into a single parameter, C_d, called the capacitance density. Following this convention,

$$C = C_d A \tag{3.1-13}$$

C_d is thus a process parameter and A is a design parameter.

The second group is composed of the capacitors formed by the separation of charge associated with a pn junction. The depletion region associated with the semiconductor junction serves as the dielectric. These junction capacitors are quite voltage dependent. They are typically expressed in terms of the process parameter C_{j0}, which denotes the junction capacitance density at zero volts bias. The capacitance of these devices can be approximated by

$$C = \frac{C_{j0} A}{(1 - V_F/\phi_B)^n} \tag{3.1-14}$$

where A is the junction area, V_F is the dc forward bias voltage of the pn junction, ϕ_B is the barrier potential (alternately, built-in potential), and n is a constant depending upon the type of junction. This expression is reasonably good for $-\infty < V_F < \phi_B/2$. These junctions are seldom operated under forward bias in MOS integrated circuits and thus V_F is normally negative. The model for $V_F > \phi_B/2$ is, however, discussed in Sections 3.2.3 and 3.3.3.

The relationship between the process parameter ϕ_B and process-controllable parameters can be found in Chapter 4. ϕ_B is typically in the 0.7 V range and will be assumed to be 0.7 V unless otherwise specified. The constant n is used to denote the type of junction. For step-graded and linearly graded junctions, n assumes the values 1/2 and 1/3, respectively. A plot of the junction capacitance density versus the forward bias voltage is shown in Fig. 3.1-17.

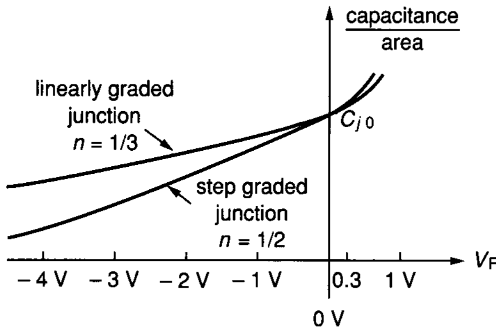


FIGURE 3.1-17
Voltage-dependent capacitance of pn junctions.

The parasitic capacitors that dominate the high-frequency behavior of MOS transistors are shown in Fig. 3.1-18. The capacitors C_{BD1}, C_{BC1}, and C_{BS1} are all voltage-dependent junction capacitors. The remaining capacitors are parallel plate capacitors governed by (3.1-13). Some of the parasitic capacitors are actually distributed devices. In addition, some of the capacitor values are operating region dependent. To simplify analysis, a model of the MOSFET in each of the three operating regions containing only lumped nodal capacitors will be presented. Comments about each of the parasitic capacitors follow.

C_{ox} represents the capacitance density of the gate/oxide/channel capacitor. The capacitance density appeared earlier in the expression for K' (Eq. 3.1-7). If ε_{SiO_2} is the relative permittivity of the silicon dioxide dielectric, which is assumed to be of thickness t_{ox}, then

$$C_{ox} = \frac{\varepsilon_0 \varepsilon_{SiO_2}}{t_{ox}} \tag{3.1-15}$$

where ε_0 is the permittivity of free space.

C_{GDO} and C_{GSO} are the gate–drain and gate–source overlap capacitors, respectively. Their existence can be attributed to the unwanted diffusion under the gate of some of the impurities used to create the drain–source regions. In addition to contributing to the parasitic capacitance, this diffusion causes a small

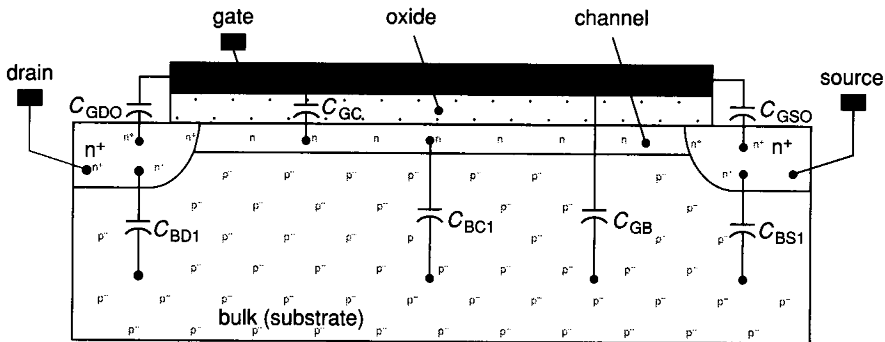


FIGURE 3.1-18
Parasitic capacitors in MOS transistors, shown for n-channel device.

decrease in the effective length of the transistors. If L_D represents the distance of the lateral moat diffusion under the gate, then both of these parasitic capacitors are nearly rectangular with length W and width L_D, resulting in a capacitance of

$$C_{GDO} = C_{GSO} = C_{ox}WL_D \tag{3.1-16}$$

C_{GC} represents the gate-to-channel capacitance. In the cutoff region the channel is not formed, so $C_{GC} = 0$. In the ohmic region the channel is quite uniform, extending from drain to source under the entire gate region. In this region, the total gate-to-channel capacitance is given by

$$C_{GC} = C_{ox}WL \tag{3.1-17}$$

Actually, the length of the channel obtained by subtracting $2L_D$ from the "drawn length" L would be more accurate, but this additional notational complexity is often not justified for hand analysis. Since the channel is a distributed rather than a lumped circuit element connecting the source to drain, and since no node corresponding to the channel appears in the MOSFET model, it is typically assumed that half of C_{GC} appears directly from gate to source and the other half appears from gate to drain to simplify calculations when the device is operating in the ohmic region. In the saturation region, the majority of the channel area acts as an ohmic extension of the source region. Typically $\frac{2}{3}C_{ox}WL$ is modeled as a lumped element between gate and source, and the remaining $\frac{1}{3}C_{ox}WL$ is neglected.

C_{GB} represents the gate–bulk capacitance. Far in the cutoff region (i.e., no depletion region in the channel), the inversion layer (alternately channel) does not exist, so the bulk region extends to the bottom side of the gate oxide layer and $C_{GB} = C_{ox}WL$. C_{GB} becomes voltage dependent and decreases with bias when operating in cutoff with a depletion region present in the channel. In both the saturation and ohmic regions, the existence of the inversion layer makes C_{GB} essentially zero. A small gate–bulk capacitance associated with gate material overlapping the bulk with the thick field oxide as a dielectric does exist. This is highly layout dependent and will be associated with layout parasitics; it is therefore not included in the model of Fig. 3.1-18.

C_{BC1} is the bulk-to-channel junction capacitance. In the cutoff region, the inversion layer does not exist, causing C_{BC1} to vanish. In both the ohmic and saturation regions, a junction capacitance, from the bulk to channel exists, with value approximated by (3.1-14). To maintain lumped element modeling, the $\frac{1}{2}$:$\frac{1}{2}$ and $\frac{2}{3}$:0 rules used previously for C_{GC} are often used in the ohmic and saturation regions, respectively, to distribute this parasitic capacitance between the drain and source.

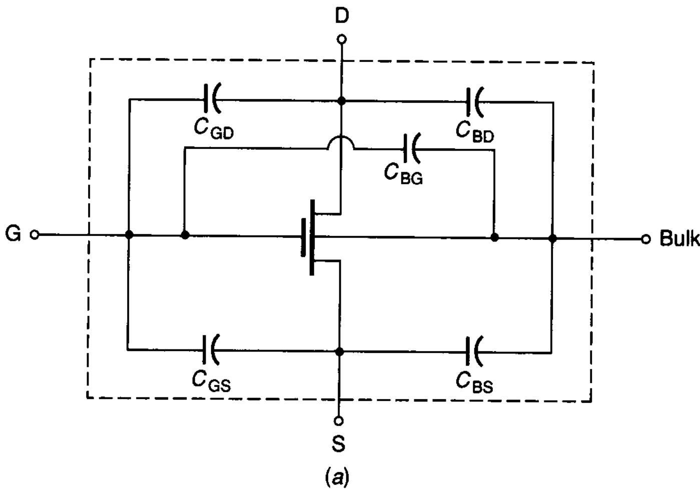
C_{BS1} and C_{BD1} are the capacitances of the bulk/source and bulk/drain junctions. These are often approximated with (3.1-14) using the source and drain areas and the process parameter C_{BM0}, which represents the zero bias bulk–moat capacitance density.

Other parasitic capacitors are often included in computer simulations but will not be considered in this section. All junction capacitances have, to this point, been assumed to be determined by the lateral junction area. For short channel

transistors (less than 3–5 μ in length) the lateral capacitance associated with the edge of the junctions becomes significant. These are descriptively termed *sidewall capacitances*. Another significant source of parasitic capacitance that typically plagues both small and large devices is the capacitance associated with the layout-dependent interconnections. Additional details about parasitic capacitors in MOS devices can be found in Chapter 4.

The lumped parasitic capacitors considered in Fig. 3.1-18 are shown in Fig. 3.1-19*a*. Note that the distributed capacitors C_{GC} and C_{BC1} have been lumped and split between the drain and source nodes. The values for these capacitors in the three regions of operation discussed above are listed in Fig. 3.1-19*b*.

Conspicuously absent from the model of Fig. 3.1-19*a* is a parasitic capacitance between the drain and source. It can, however, be seen from Fig. 3.1-18 that no physical mechanism exists to create such a capacitor.



(*a*)

	Region		
	Cutoff	Ohmic	Saturation
C_{GD}	$C_{OX}WL_D$	$C_{OX}WL_D + \frac{1}{2}WLC_{OX}$	$C_{OX}WL_D$
C_{GS}	$C_{OX}WL_D$	$C_{OX}WL_D + \frac{1}{2}WLC_{OX}$	$C_{OX}WL_D + \frac{2}{3}WLC_{OX}$
C_{BG}	$C_{OX}WL$	0	0
C_{BD}	C_{BD1}	$C_{BD1} + \frac{C_{BC1}}{2}$	C_{BD1}
C_{BS}	C_{BS1}	$C_{BS1} + \frac{C_{BC1}}{2}$	$C_{BS1} + \frac{2}{3}C_{BC1}$

(*b*)

FIGURE 3.1-19
Parasitic capacitors: (*a*) Lumped model, (*b*) values, (*c*) Small signal equivalent circuit.

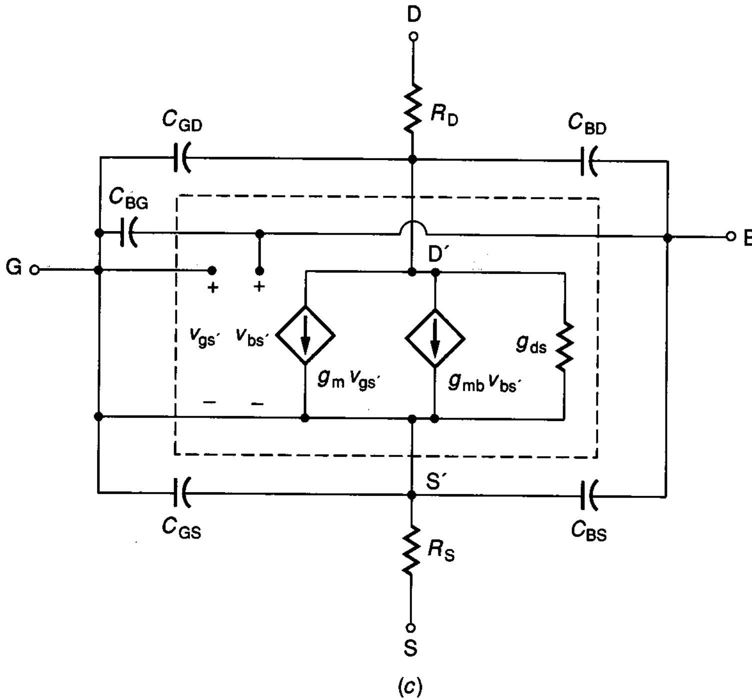


FIGURE 3.1-19
(*continued*)

The small signal equivalent circuit of the MOSFET including the capacitive parasitics is shown in Fig. 3.1-19*c*. The resistors R_D and R_S have been included to account for the ohmic resistance in the drain and source regions from the physical drain and source connections to the actual drain (D′) and source of the device. For large physical drain and source areas, these resistors become quite large. Techniques discussed in Chapter 2 can be used to determine these values from the sheet resistance and physical dimensions of the drain and source regions. For minimum size drain and source regions, these resistances will be in the 50 Ω range.

The careful reader may be concerned that the rather qualitative approach to modeling the parasitic capacitances associated with the MOSFET may seriously limit use of this model to *accurately* predict high-frequency MOSFET performance. Although the concern is well-founded, because of the inability to accurately control the parasitic capacitances during processing and because of the operating region and voltage dependence of these capacitances, practical designs *must not* have important specifications directly dependent upon these parasitic capacitances. The designer thus designs "away from" these capacitances and consequently a crude model of these parasitics is usually sufficient for the designer to verify that these parasitics will not significantly impact the key performance specifications in a design.

Example 3.1-5. Calculate all of the parasitic capacitors included in Fig. 3.1-19*a* for M1 in the circuit of Fig. 3.1-20. Use the CMOS process parameters of Table 2B.4 in Chapter 2. Assume the areas of the sources of both M1 and M2 are equal to 100 μ^2 and that the drain areas of M1 and M2 are each 60 μ^2. Neglect junction sidewall capacitances. (Note that this circuit has a large voltage gain at the Q-point established, and thus the Q-point is highly sensitive to changes in either bias voltages or device sizes.)

Solution. The region of operation for M1 must initially be determined along with the operating point. Assume initially that M1 and M2 are both operating in saturation. With this assumption, it follows upon equating drain currents that

$$\frac{K_N' W_1}{2L_1}(V_I - V_{TN})^2(1 + \lambda_N V_O) = K_P' \frac{W_2}{2L_2}(V_I - V_{DD} - V_{TP})^2(1 - \lambda_P[V_O - V_{DD}])$$

Inserting the given excitation voltages, dimensions, and process parameters from Table 2B.4 into this equation, it follows that $V_O = 5$ V. It is readily verified that the assumption initially made that both transistors are operating in the saturation region is valid.

From Table 2B.4 and (3.1-16) it follows that

$$C_{ox} = 0.7\ \text{fF}/\mu^2$$

$$C_{GDO} = (0.7\ \text{fF}/\mu^2)(0.45\ \mu)W_1 = 1.58\ \text{fF}$$

$$C_{GSO} = (0.7\ \text{fF}/\mu^2)(0.45\ \mu)W_1 = 1.58\ \text{fF}$$

$$C_{BD} = (0.33\ \text{fF}/\mu^2)(60\ \mu^2) = 19.8\ \text{fF}$$

$$C_{BS} = (0.33\ \text{fF}/\mu^2)(100\ \mu^2) = 33\ \text{fF}$$

$$C_{BC} = (0.33\ \text{fF}/\mu^2)(5\ \mu)(10\ \mu) = 16.5\ \text{fF}$$

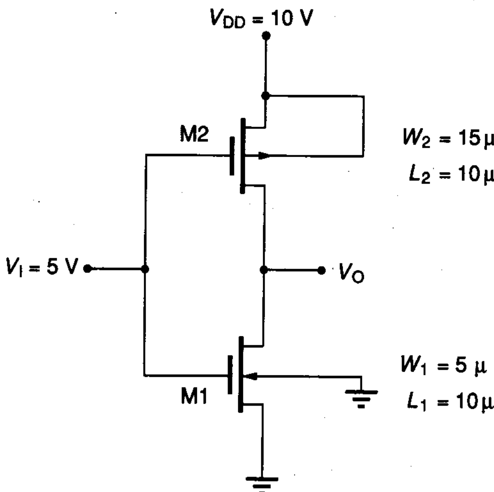


FIGURE 3.1-20
Example 3.1-5.

Thus from the third column of Fig. 3.1-19*b*,

$$C_{GD} = 1.58\,\text{fF}$$

$$C_{GS} = 24.9\,\text{fF}$$

$$C_{BG} = 0$$

$$C_{BD} = 19.8\,\text{fF}$$

$$C_{BS} = 44.0\,\text{fF}$$

Note the small size of these parasitic capacitors relative to those associated with discrete electronic circuitry. Although they are small, it will be seen later that the effects of these parasitics are significant at high frequencies.

3.1.4 Measurement of MOSFET Model Parameters

Typical values of process parameters are generally available at the start of a design. Knowledge of the process parameters themselves, as well as of statistical and/or worst case information, is necessary during the design stage so that tradeoffs between performance, complexity, and yield can be considered by the designer. Experimentally measured process parameters also serve as an interface between the fabrication and design groups. A check to verify that the process parameters fall within previously agreed-upon bounds is typically made after fabrication and prior to extensive circuit testing by the designers. These measurements are made over a large enough number of dies to establish confidence in the measurements.

Automated device characterization equipment is readily available for measuring many of the parameters that are used to characterize devices. In the absence of this equipment, some of the more important parameters can be measured with reasonable accuracy using standard laboratory instruments. Techniques for measuring the parameters K', V_{T0}, γ, and λ that appear in the model of the MOSFET previously introduced are discussed in this section. Since models are not perfect, it is advisable to measure model parameters in the same regions where the device will be operated and near the intended operating point if possible. The following discussion assumes that the reader is familiar with the basic operation of an operational amplifier. For those lacking background in this area, most electronics texts have a brief, self-contained section discussing operational amplifiers (e.g., References 4 and 5).

Measurement of λ. Consider the expression for the drain current of the MOSFET in the saturation region as given in Table 3.1-1:

$$I_D = \frac{K'W}{2L}(V_{GS} - V_T)^2(1 + \lambda V_{DS}) \qquad \text{for} \quad V_{DS} > V_{GS} - V_T, V_{GS} > V_T \tag{3.1-18}$$

This equation is plotted in Fig. 3.1-21 for different values of $(V_{GS} - V_T)$ with constant V_{BS}. Projection of these curves into quadrant 2 shows that all curves

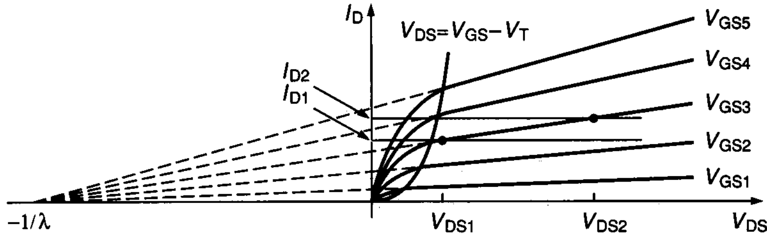
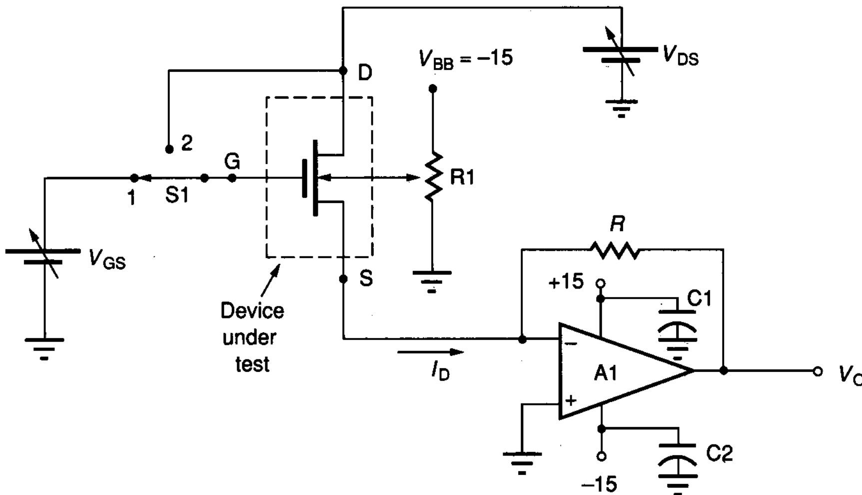


FIGURE 3.1-21
Measurement of λ.

intersect the V_{DS} axis at $V_{DS} = -1/\lambda$. The parameter λ can thus be determined by fixing V_{GS} and V_{BS}, thus fixing $(V_{GS} - V_T)$, and determining the $-V_{DS}$ axis intercept of I_D versus V_{DS}. If I_{D1} and I_{D2} are the currents corresponding to two distinct values of V_{DS}, V_{DS1}, and V_{DS2} (see Fig. 3.1-21), then it readily follows that

$$\lambda = \frac{I_{D2} - I_{D1}}{I_{D1}V_{DS2} - I_{D2}V_{DS1}} \tag{3.1-19}$$

The circuit of Fig. 3.1-22 with S1 in position 1 is useful for measuring λ. V_{GS} and V_{BS} are fixed so that $V_{GS} > V_T$. The operational amplifier (op amp) is used to establish the relationship $V_O = -I_D R$ while maintaining a source voltage of nearly 0 V. If V_{DS1} and V_{DS2} are the two distinct values of V_{DS} required in



A 356 op amp is a reasonable choice for A1. C1 and C2, which may be needed for stability, should be as close to the supply terminals of the op amp as possible to prevent oscillation. A value of 1µF is reasonable for these capacitors.

FIGURE 3.1-22
Circuit for measuring MOSFET parameters. V_{GS}, V_{DS}, and V_{BB} should be of opposite polarity when measuring p-channel parameters.

(3.1-19) and V_{O1} and V_{O2} respectively are the output voltages of the op amp for these inputs, it follows that

$$\lambda = \frac{V_{O2} - V_{O1}}{[V_{O1}V_{DS2} - V_{O2}V_{DS1}]} \tag{3.1-20}$$

Measurement of V_T and V_{T0}. With V_{BS} and V_{DS} fixed, the values of I_D (I_{D1} and I_{D2}) corresponding to two distinct values of V_{GS} (V_{GS1} and V_{GS2}) can be measured. If the device is operating in saturation for both values of V_{GS}, a routine calculation shows

$$V_T = \frac{V_{GS2}(I_{D1}/I_{D2})^{1/2} - V_{GS1}}{(I_{D1}/I_{D2})^{1/2} - 1} \tag{3.1-21}$$

If V_{BS} is set to 0 V, it follows by definition that the expression in (3.1-21) yields V_{T0}. The circuit of Fig. 3.1-22 with S1 in position 1 can again be used to measure V_T.

Measurement of K'. Once λ and V_T have been measured, it follows from (3.1-18) that K' can be determined from a measurement of I_D for fixed values of V_{GS} and V_{DS} from the expression

$$K' = \frac{2LI_D}{W(V_{GS} - V_T)^2(1 + \lambda V_{DS})} \tag{3.1-22}$$

The circuit of Fig. 3.1-22 is again useful for this measurement.

Another scheme is often used for measuring K' and V_T. Since λV_{DS} is typically much less than 1, it follows from (3.1-18) that

$$I_D \simeq \frac{K'W}{2L}(V_{GS} - V_T)^2 \qquad \text{for} \quad V_{GS} > V_T, V_{DS} > V_{GS} - V_T$$

To guarantee operation in the saturation region for enhancement devices, it is convenient to tie the gate to the drain by setting switch S1 to position 2 and for depletion devices to establish a large dc bias on V_{DS} with S1 in position 1, yielding

$$I_D = \frac{K'W}{2L}(V_{GS} - V_T)^2 \qquad V_{GS} > V_T \tag{3.1-23}$$

A typical plot of $\sqrt{I_D}$ versus V_{GS} is shown in Fig. 3.1-23 for both enhancement and depletion MOSFETs for several different values of V_{BS}. In all cases the slopes of the straight lines are $[(K'W)/(2L)]^{1/2}$, enabling a ready determination of K'. Note the bending of the curves for low values of V_{GS}. This bending is due to subthreshold currents and is not included in the models discussed up to this point. A discussion of subthreshold operation appears in Section 3.1.6. The drain current must be large enough so that the slope is not affected by the subthreshold effects when measuring V_T. The horizontal axis intercept of the projection of the straight lines is the threshold voltage. The intercept corresponding to $V_{BS} = 0$ is V_{T0}. The circuit of Fig. 3.1-22 is useful for measuring V_T and K' following this approach.

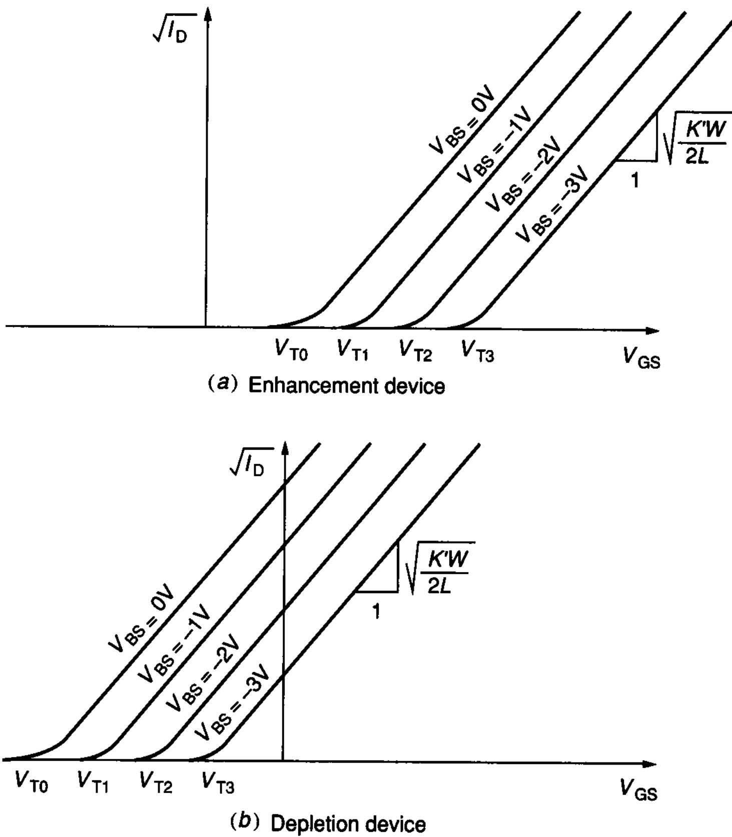


FIGURE 3.1-23
Alternate method of measuring K' and V_T: (*a*) Enhancement device, (*b*) Depletion device.

Measurement of γ. From Table 3.1-1 recall that

$$V_{\mathrm{T}} = V_{\mathrm{T0}} + \gamma\left(\sqrt{\phi - V_{\mathrm{BS}}} - \sqrt{\phi}\right) \tag{3.1-24}$$

Since methods of measuring both V_{T} and V_{T0} have been presented, both γ and ϕ can be determined by obtaining values of V_{T} from two distinct nonzero values of V_{BS}. Details about obtaining these parameters from either of these two distinct values or by applying some minimum mean squared error approximation from several measured values is left to the reader.

For large values of V_{BS}, the threshold voltage is only weakly dependent upon ϕ. In that case, using a typical value of ϕ, ϕ_{typ}, a good approximation to γ can be obtained with a single large value of V_{BS} from the expression

$$\gamma \simeq \frac{V_{\mathrm{T}} - V_{\mathrm{T0}}}{\sqrt{\phi_{\mathrm{typ}} - V_{\mathrm{BS}}} - \sqrt{\phi_{\mathrm{typ}}}} \tag{3.1-25}$$

Information about the parasitic capacitance parameters is often required for accurate time domain analysis of MOS circuits. The parasitic capacitors discussed in the previous section are quite small for minimum-sized MOSFETs (well under 1 pF). Some were also seen to be voltage dependent. Due primarily to their small size, it is nearly impossible to measure directly these capacitances with any acceptable degree of accuracy using standard laboratory equipment. Special test circuits are often included on an IC for the purpose of obtaining information about the parasitic capacitances. These test circuits may contain either geometrically large devices so that the total parasitic capacitance is large enough to measure easily or they may contain more complicated circuit structures (such as ring oscillators) from which these parasitics can be calculated indirectly from measurements of other characteristics of the circuits.

Other measurement considerations. The methods presented above provide simple measurement of some of the key parameters of a MOSFET. Questions about such problems as accuracy and Q-point were avoided. Regardless of what measurement technique is used, it is advisable to measure these parameters as close to the intended device operating point (Q-point) as possible because model inaccuracies can cause significant global variations in these parameters whereas local variations are usually quite small. Also, measurements at several data points in the neighborhood of the Q-point, followed by data smoothing or curve fitting, usually give better results than single-point or two-point measurements. This is easily achieved if computer-controlled measurement equipment is available. The extension of these measurement methods to smoothing or curve fitting is straightforward and is left to the reader.

Finally, the question of how accurately one has measured these parameters and how much accuracy is needed naturally arises. Unfortunately, the functional form of the model presented here is not perfect, and thus the parameters, along with the functional form of the model, only approximate the device performance. Because of this imperfect fit, it is inherent that different measurement algorithms will give slightly different values for these parameters. This problem is analogous to the problem of trying to fit a second-order polynomial to data points obtained from a third-order polynomial. Nevertheless, the model for the MOSFET is quite good, and the measured parameters are useful for predicting device performance. In applications where matching or ratio matching of parameters is particularly important, the same parameter measurement algorithms should be employed for each to avoid introducing systematic errors.

3.1.5 Short Channel Devices

Downward trends in device scaling have caused minimum MOSFET device lengths to decrease from the 10 μ range of the mid-1970s to the 1 μ range of the mid-1980s. Device lengths in the 0.5 μ range or below are projected by the early 1990s. The models that we discussed earlier in this chapter are quite good for channel lengths longer than 5 μ. The models gradually deteriorate for shorter channel lengths. Transistors with channel lengths less than 3 to 5 μ are termed

short channel devices. With short channel devices the ratio between lateral and vertical dimensions is reduced. Geometrically, the channel region changes from a rather uniform, thin right-rectangular region to a much more irregular structure. The effects of the transition region from the drain and source diffusions to the channel become significant, causing need for increased complexity in the device model.

Short channel transistors offer some significant advantages over larger transistors but also have serious limitations. The major advantages are the reduced area requirements and improvements in speed that are attainable with circuits employing short channel transistors. The improvements in speed are mostly attributable to the reduced input capacitance associated with the smaller devices. The major limitation is a deterioration in the output impedance characteristics. Good matching of short channel transistor characteristics is also more difficult to achieve.

The typical output characteristics of a short channel transistor are compared with that of a longer channel device in Fig. 3.1-24. A simple empirical model of the MOSFET which is essentially an extension of the long channel model summarized in Table 3.1-1 is given by the following equations:

$$I_G = 0 \tag{3.1-26}$$

$$I_D = \begin{cases} 0 & V_{GS} < V_T \text{ (cutoff)} \\ \dfrac{K'W_{\text{eff}}}{L_{\text{eff}}}\left(V_{GS} - V_T - \dfrac{V_{DS}}{2}\right)V_{DS}\left[1 + \lambda\left(\dfrac{1+\theta L_{\text{eff}}}{\theta L_{\text{eff}}}\right)V_{DS}\right] & V_{GS} > V_T,\ 0 < V_{DS} < V_{GS} - V_T \text{ (ohmic)} \quad (3.1\text{-}27) \\ \dfrac{K'W_{\text{eff}}}{2L_{\text{eff}}}(V_{GS} - V_T)^2\left[1 + \lambda\left(\dfrac{1+\theta L_{\text{eff}}}{\theta L_{\text{eff}}}\right)V_{DS}\right] & V_{GS} > V_T,\ V_{DS} > V_{GS} - V_T \text{ (saturation)} \quad (3.1\text{-}28) \end{cases}$$

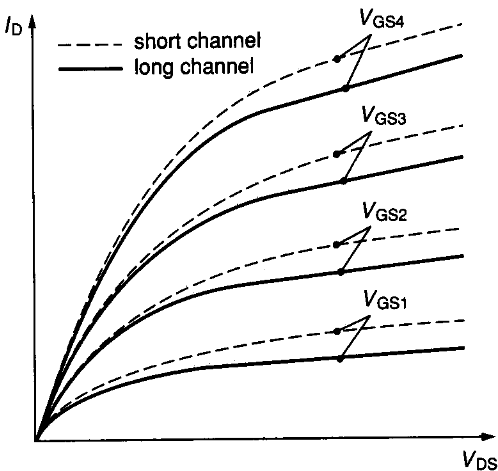


FIGURE 3.1-24
Short channel effects on MOSFET performance. Solid curves for 5 μ device, dashed curves for 1 μ device.

The parameter θ has units $(\text{length})^{-1}$. A typical value of θ may be 0.2 μ^{-1}. The parameters W_{eff} and L_{eff} denote the effective channel width and length respectively. These are given by

$$W_{\text{eff}} = W - W_{\text{R}} \tag{3.1-29}$$

$$L_{\text{eff}} = L - 2L_{\text{D}} - L_{\text{R}} \tag{3.1-30}$$

where W and L are the drawn width and length, W_{R} and L_{R} are constants representing width and length reduction due to processing, and L_{D} is the lateral diffusion of the source or drain under the gate. W_{eff} and L_{eff} could also be used to model W and L in the model of the long channel MOSFET presented earlier and are, in fact, included in most good models used for simulation. The use of W_{eff} and L_{eff} for hand analysis of large structures is not so important since the difference between these values and the drawn width and length is relatively small. For short channel devices, however, the relative effects of the width and length reduction and the lateral diffusions become more pronounced. Short channel effects should be considered for channel lengths less than 5 μ. This model is still usable in hand calculations. It should be noted that a size adjust of the appropriate mask is often done so that the effective length and/or width ideally agrees with the drawn length and/or width. The designer should be aware of what size adjusts, if any, will be used at mask generation.

More accurate models of the short channel effects are available.[6–7,19] These models are much more involved and not amenable to hand calculations but do give reasonable results when used for computer simulations. The model in Reference 7 is incorporated into the popular circuit simulator, SPICE. For the reasons discussed below, a simple model such as that of (3.1-27) and (3.1-28) is often adequate for hand analysis of circuits utilizing short channel transistors.

Were it not for the additional notational complexity, the model presented in Table 3.1-1 could be replaced with the short channel model introduced in this section. Note, in particular, that for large L the short channel model is essentially equivalent to that for the longer transistors. It can be concluded that the short channel model can be appropriately used for short as well as long device sizes if desired. The empirical short channel model parameter θ was included in Table 3.1-2.

We close this section by comments about where short channel devices are used. Short channel devices are widely used in digital logic, where a net improvement in speed and reduction in circuit area are readily achievable in spite of the performance limitations of the short channel MOSFET. In analog applications, where good matching characteristics and high output impedance are often critical, short channel transistors find fewer applications in spite of the improved frequency response. When analog circuits are built in processes where short channel lengths are available, most devices will be somewhat larger than minimum size; some performance benefits in terms of device matching will, however, be derived in these processes because of the more stringent lithography requirements that must be maintained in the small feature size processes.

3.1.6 Subthreshold Operation

In the dc model of the MOSFET introduced in Sec. 3.1.1, the drain current for positive V_{DS} was assumed to be zero for $V_{GS} < V_T$ and nonzero for $V_{GS} > V_T$ as indicated by Eq. 3.1-1. In physical devices such an abrupt transition is not anticipated and does not occur experimentally. The drain current is, however, much smaller for $V_{GS} < V_T$ than for $V_{GS} > V_T$ and hence in most applications the assumption that $I_D = 0$ for $V_{GS} < V_T$ is justifiable.

Applications do exist, however, where it is crucial that current levels be extremely small. These include, but are not limited to, biomedical applications such as pacemakers and other implantable devices that must operate for several years with small nonrechargeable batteries. Useful circuits with extremely low supply currents in which $V_{GS} < V_T$ find applications in such situations. If $V_{GS} > V_T$, the devices are said to be operating in strong inversion. If $V_{GS} < V_T$, the devices are said to be operating in weak inversion, or equivalently, in the subthreshold region. At room temperature, the transition between strong inversion and weak inversion actually occurs around $V_{GS} \simeq V_T + 100$ mV. The expression

$$V_{GS} = V_T + 2nV_t \tag{3.1-31}$$

where n is a constant between 1 and 2, can be used to predict the transition at other temperatures.[8] The term V_t is equal to kT/q where k is Boltzmann's constant ($k = 1.387 \times 10^{-23}$ V · C/°K), T is the device temperature in degrees Kelvin, and q is the charge of an electron ($q = 1.6 \times 10^{-19}$ C). At room temperature, $V_t = 26$ mV.

A plot of I_D versus V_{GS} for the MOSFET with $V_{DS} = V_{GS}$ is shown in Fig. 3.1-25 for a wide range of V_{GS} values. As can be seen from this figure, the current in weak inversion is much smaller than the current in strong inversion, as previously stated. The behavior is, however, well characterized far into weak inversion and varies exponentially rather than quadratically with V_{GS}. The potential for extremely low power dissipation should be apparent.

Example 3.1-6.

(*a*) Compare the power dissipation required for a single device operating with $V_{GS} = 3V_T$ driven from a 5 V source to the power dissipation required for the same device operating from a 1 V supply biased at $V_{GS} = 0.85V_T$. Assume the device has characteristics as shown in Fig. 3.1-25.

(*b*) If ideal batteries with a total stored energy of 1 watt-hr are used to power the two devices, what is the battery life of each?

Solution.

(*a*) From Fig. 3.1-25*a*, it follows that with $V_{GS} = 3V_T$, $I_{D1} \simeq 150\ \mu$A, and from Fig. 3.1-25*b*, with $V_{GS} = 0.85V_T$, $I_{D2} \simeq 1$ nA. The power dissipation ratio is thus $(5I_{D1})/(I_{D2}) = 7.5 \times 10^5$.

(*b*) The battery life with the 5 V system would be

$$T = 1 \text{ watt-hr}/(150\ \mu\text{A}\cdot 5\text{ V}) = 1333 \text{ hr} \simeq 8 \text{ weeks}$$

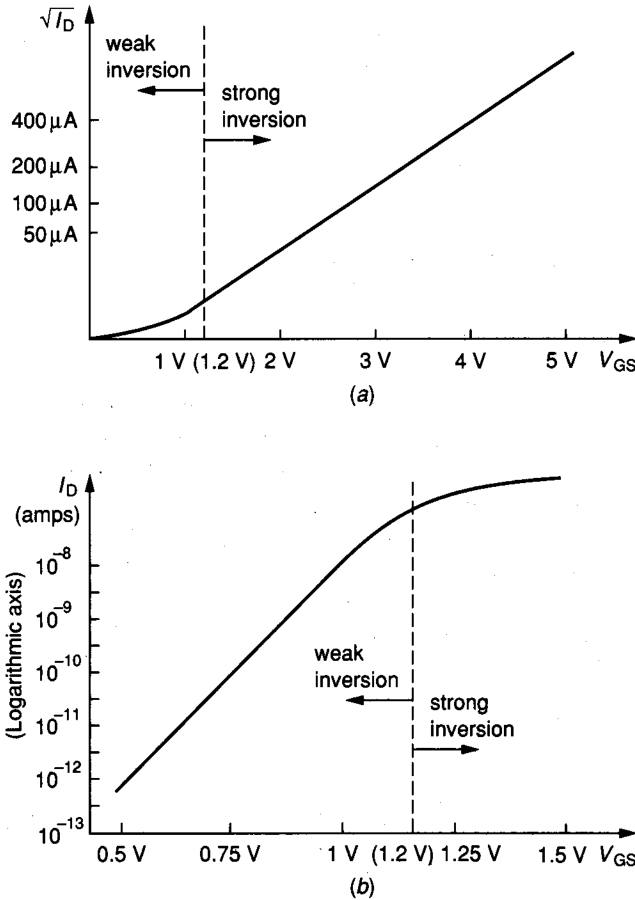


FIGURE 3.1-25
Typical I_D–V_{DS} characteristics for MOSFET operating in weak inversion: (*a*) Quadratic vertical axis, (*b*) Logarithmic vertical axis ($V_T = 1\ V$, $V_{DS} = V_{GS}$).

and the battery life with the 1 V system would be

$$T = 1 \text{ watt-hr}/[(1 \text{ nA}) \cdot 1 \text{ V}] \simeq 10^9 \text{ hr} \simeq 114{,}000 \text{ years}$$

It should be apparent that the subthreshold system may be practical whereas the strong inversion system would be impractical in many applications.

A dc model for the MOSFET operating in weak inversion is needed for both design and simulation. The following model is useful in weak inversion.[9]

$$\left.\begin{aligned} I_G &= 0 \\ I_D &= \frac{W}{L} I_{DO} e^{-V_{BS}[(1/(nV_t))-(1/V_t)]} \left(1 - e^{-V_{DS}/V_t}\right) e^{(V_{GS}-V_T)/(nV_t)} \end{aligned}\right\} \quad (3.1\text{-}32)$$

where V_T is the threshold voltage and n and V_t are as defined in Eq. 3.1-31. The constants I_{DO} and n are process parameters. Typical values for these parameters are $I_{DO} \simeq 20$ nA and $n = 2$.

A typical plot of I_D versus V_{DS} is shown in Fig. 3.1-26. Note that for $V_{DS} > 3V_t$, the term e^{-V_{DS}/V_t} in Eq. 3.1-32 becomes negligible, and one obtains the equivalent of the strong inversion saturation region of operation. Often $V_{BS} = 0$. Under the assumption that $V_{BS} = 0$ and $V_{DS} > 3V_t$, Eq. 3.1-32 simplifies considerably, to

$$\left.\begin{aligned} I_G &= 0 \\ I_D &\simeq \frac{W}{L} I_{DO} e^{(V_{GS}-V_T)/(nV_t)} \end{aligned}\right\} \qquad (3.1\text{-}33)$$

It is often the case, however, that power supply restrictions will require operation of subthreshold devices with smaller values of V_{DS}.

The parameter I_{DO} is related to the transconductance parameter and is approximately given by the expression

$$I_{DO} \simeq \frac{K'2(nV_t)^2}{e^2}$$

This approximation for I_{DO} provides for continuity in the device model of (3.1-33) and (3.1-4) and in its derivative, $(\partial I_D/\partial V_{GS})$, at the strong inversion–weak inversion interface defined in (3.1-31)—see Problem 3.30.

There are several practical limitations of devices operating in weak inversion that deserve noting. First, the frequency response of devices operating far into weak inversion is poor. This can be qualitatively seen by observing that the parasitic device capacitances are geometrically determined and hence nearly equal to those discussed in the strong inversion model. The maximum current available to charge and discharge these capacitors is significantly less, causing a significant deterioration of frequency response. Second, the drain and source substrate currents associated with the reverse-biased moat–substrate junction are not necessarily negligible compared to subthreshold drain currents. Third, the linearity

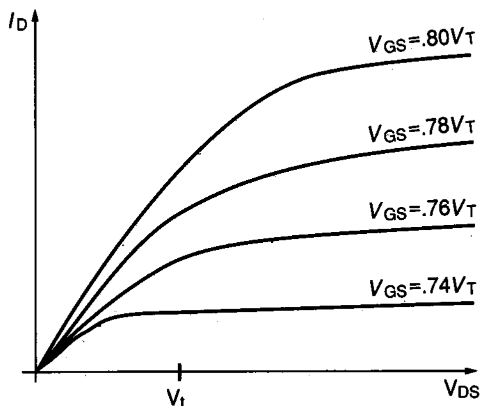


FIGURE 3.1-26
Typical I_D–V_{DS} characteristics for MOSFET operating in subthreshold.

is quite poor for $V_{DS} < 3V_t$, making linear designs more challenging. Fourth, the deterioration of matching characteristics of MOS transistors with decreasing drain currents further complicates linear design.

Because of these practical limitations, it is often desirable to operate near the transition region between strong inversion and weak inversion, where some of the benefits of reduced power associated with weak inversion can be derived but where these other limitations are not too problematic. Whereas diffusion current dominates weak inversion operation and drift current dominates strong inversion operation, both mechanisms interact in the transition region and thus complicate the modeling problem. A discussion of the modeling problem in this region is beyond the scope of this text and continues to receive attention in the technical literature. Even weak inversion models used in simulation programs such as SPICE have major limitations. Nevertheless, some designers do design useful circuits that operate in this transition region.

3.1.7 Operation in the Third Quadrant of the I_D–V_{DS} Plane

Although MOS devices are typically operated with $I_D > 0$ and $V_{DS} > 0$, (Quadrant 1 in Fig. 3.1-7), there are applications where the device is biased to operate in Quadrant 3 of the I_D–V_{DS} plane for some inputs. For notational convenience, a MOS transistor without drain and source designations is shown in Fig. 3.1-27, along with plots of I_1 versus V_{12} for different values of V_{32} and fixed V_B. The normal convention for nodal designations if the device were operating in Quadrant 1 would be ① → drain, ② → source, and ③ → gate. For operation in Quadrant 3, the drain and source designations would be reversed, based upon the device models discussed earlier in this chapter.

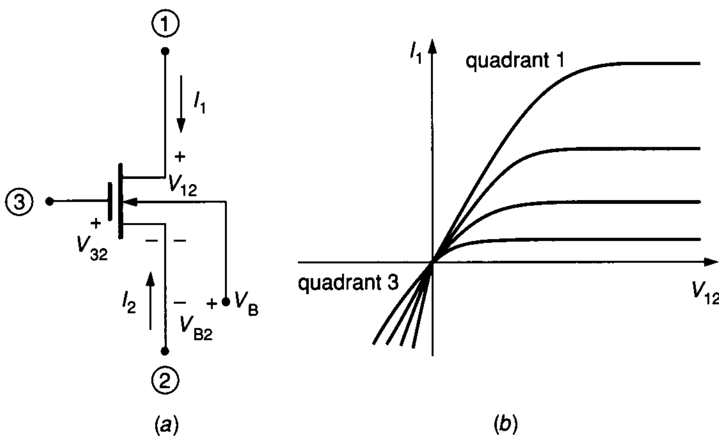


FIGURE 3.1-27
Operation of the MOSFET in the third quadrant: (*a*) Terminal convention, (*b*) I–V characteristics.

It should be apparent that the $I-V$ characteristics do not display symmetry from Quadrant 1 to Quadrant 3. The asymmetry becomes more pronounced with decreasing reverse bulk bias. It should also be apparent that, based upon the device model introduced earlier in this chapter, the drain–source designations are not arbitrary.

To obtain a qualitative appreciation for how the device behaves in Quadrant 3, assume that $0 < V_{32} < V_T$ and that subthreshold currents are negligible. By the previous model, we are tempted to conclude that the device is cutoff since "$V_{GS} < V_T$" and hence conclude that $I_1 = 0$. If, however, the voltage at terminal ① in Fig. 3.1-27 were made negative relative to terminal ③ and if V_{13} is less than $-V_T$, it can be argued that the gate–channel voltage near terminal ① is greater than V_T, causing the formation of an inversion layer which would, in turn, cause a positive current I_2 to flow. This is what happens in an actual MOSFET. This biasing strategy is consistent with that discussed earlier in this chapter, provided the nodal designations ① → source, ② → drain, and ③ → gate are made.

For consistency with the model summarized in Table 3.1-1, the MOS device will be modeled in Quadrant 3 by using the model of Table 3.1-1 with the agreement that the designations *source* and *drain* will be made so that $V_{DS} > 0$. This means that in some applications, the drain and source designations change during normal operation of the device.

As an alternative to allowing the drain and source designations to change so that $V_{DS} \geq 0$ at all times, the designations can be arbitrarily fixed and the device model in Quadrant 3 can be defined to be equal to that which would be attained were the designations changed. Following this approach, the equivalent fixed drain–source designation model (for the n-channel transistor) summarized in Table 3.1-4 is obtained. The parameter λ has been included in the ohmic region expressions in this model to maintain continuity in the device characteristics. The regions of operation in the $V_{DS}-V_{GS}$ plane are depicted in Fig. 3.1-28 for a fixed bulk bias.

Example 3.1-7. The device of Fig. 3.1-29 is to be used as an active grounded resistor. Verify that the $I-V$ characteristics are continuous and differentiable at the origin ($V = 0$).

Solution. Since for small V, this depletion device is operating in either the forward ohmic or reverse ohmic regions, it follows from Table 3.1-4 that

$$I = K'\frac{W}{L}\left(-V_{Ti} - \frac{V}{2}\right)V(1 + \alpha_i \lambda V) \qquad i = 1,2$$

where $i = 1$ for $V > 0$ and $i = 2$ for $V < 0$, where $\alpha_1 = +1$ and $\alpha_2 = -1$, and where $V_{DS} = V$.

To verify continuity, we must show that

$$\lim_{V \to 0^+} I = \lim_{V \to 0^-} I$$

From the expression for I,

$$\lim_{V \to 0^+} I = 0 = \lim_{V \to 0^-} I.$$

TABLE 3.1-4
Low-frequency fixed-terminal MOSFET model in four quadrants.

$$
I_D = \begin{cases}
0 & V_{GS} < V_{T1},\ V_{DS} > V_{GS} - V_{T2} & \text{cutoff} \\
\dfrac{K'W}{2L}(V_{GS} - V_{T1})^2(1 + \lambda V_{DS}) & V_{GS} > V_{T1},\ V_{DS} > V_{GS} - V_{T1} & \text{forward saturation} \\
\dfrac{K'W}{L}\left(V_{GS} - V_{T1} - \dfrac{V_{DS}}{2}\right)V_{DS}(1 + \lambda V_{DS}) & 0 < V_{DS} < V_{GS} - V_{T1},\ V_{GS} > V_{T1} & \text{forward ohmic} \\
\dfrac{K'W}{L}\left(V_{GS} - V_{T2} - \dfrac{V_{DS}}{2}\right)V_{DS}(1 - \lambda V_{DS}) & V_{GS} > V_{T2},\ V_{DS} < 0 & \text{reverse ohmic} \\
-\dfrac{K'W}{2L}(V_{GS} - V_{DS} - V_{T2})^2(1 - \lambda V_{DS}) & V_{GS} < V_{T2},\ V_{DS} < V_{GS} - V_{T2} & \text{reverse saturation}
\end{cases}
$$

$$V_{T^*}{}^{a} = V_{T0} + \gamma\left(\sqrt{\phi - V_{BB}} - \sqrt{\phi}\right)$$

$$V_{T1} = V_{T0} + \gamma\left(\sqrt{\phi - V_{BS}} - \sqrt{\phi}\right)$$

$$V_{T2} = V_{T0} + \gamma\left(\sqrt{\phi - V_{BS} + V_{DS}} - \sqrt{\phi}\right)$$

[a] V_{T^*} characterizes the V_{DS} transition in Fig. 3.1-28.

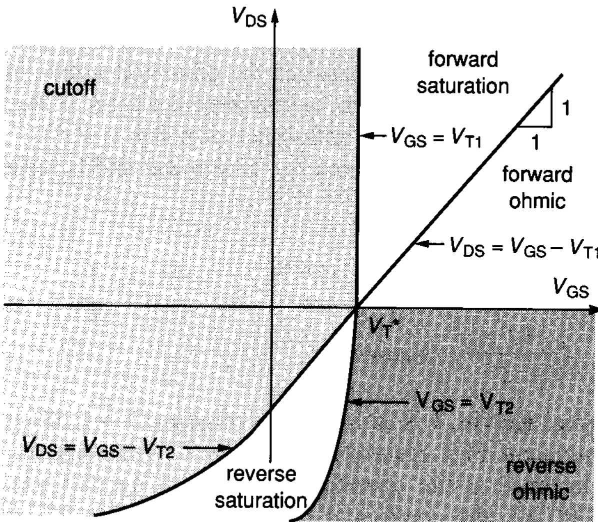


FIGURE 3.1-28
Model extension regions of operation of MOSFET (fixed V_{BS}).

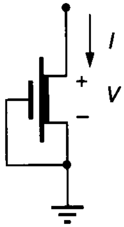


FIGURE 3.1-29
Grounded simulated resistor of Example 3.1-7.

To verify differentiability, we must show

$$\lim_{V\to 0^+} \frac{\partial I}{\partial V} = \lim_{V\to 0^-} \frac{\partial I}{\partial V}$$

From the expression for I, it follows that

$$\frac{\partial I}{\partial V} = -\frac{K'W}{2L}\left[V^2\left(3 + 2\frac{\partial V_{Ti}}{\partial V}\right)(\alpha_i\lambda) + V\left(2 + 4\alpha_i\lambda V_{Ti} + \frac{2\partial V_{Ti}}{\partial V}\right) + 2V_{Ti}\right]\Bigg|_{V_{DS}=0\,V}$$

and hence that

$$\lim_{V\to 0} \frac{\partial I}{\partial V} = -\frac{K'W}{L}V_{Ti}\Bigg|_{V_{DS}=0\,V} \qquad V = \frac{K'W}{L}\left(V_{T0} + \gamma\left[\sqrt{\phi - V_{BB}} - \sqrt{\phi}\right]\right)$$

for $i = 1,2$. The model is therefore differentiable at the origin.

3.1.8 Modeling Noise Sources in MOSFETs

Two mechanisms are the primary contributors to the presence of noise in MOSFETs. One is thermal noise associated with the carriers in the channel. The second is flicker noise associated with the trapping and releasing of electrons in the Si–SiO_2 interface region. These noise sources contribute to the total drain current and can thus be modeled as a current source between the drain and source in either the large signal or small signal device model, as indicated in Fig. 3.1-30*a*, where this current source combines the effects of both types of noise.

The thermal noise current is white noise, which has zero mean and is most easily characterized by its spectral density:

$$S_{IW} = \begin{cases} \dfrac{4kT}{R_{FET}} & \text{ohmic region} \\ \dfrac{8kTg_m}{3} & \text{saturation} \end{cases} \qquad (3.1\text{-}34)$$

where T is temperature in degrees Kelvin, k is Boltzmann's constant ($k = 1.381 \times 10^{-23}$ J/°K), R_{FET} is the equivalent FET resistance given by (3.1-8), and g_m is the small signal transconductance at the operating point given by (3.1-11). At room temperature (T = 300°K), the coefficient $8kT/3$ equals 1.1×10^{-20} V · A · sec.

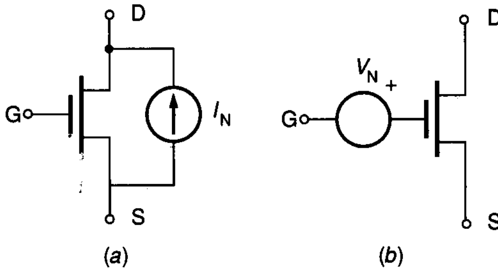


FIGURE 3.1-30
Model of noise sources in the MOSFET: (*a*) Current source in output, (*b*) Input-referred voltage source for small signal operation in the saturation region.

It can be seen that the units of S_{IW} are $A^2 \cdot$ sec. The RMS noise current will be obtained from the spectral density later in this section. Both current and voltage spectral densities are used to characterize noise, and these are often termed the *power spectral densities*. Although it may appear from the saturation region expression in (3.1-34) that the small signal MOSFET model plays a role in characterizing the noise of the MOSFET in the saturation region due to the presence of g_m, this may be somewhat misleading. It can be shown that the expression for the thermal noise contains a $\sqrt{V_{GS} - V_T}$ term, which also appears in the small signal g_m expression of (3.1-11), thus motivating the use of the g_m term in the noise current Eq. 3.1-34. Actually, the thermal noise current is just that associated with the channel source resistance of the MOSFET when operating in the saturation region (see Problems 3.22 and 3.23). Thus, in the ohmic regions, the thermal noise current is identical to that of a resistor of value R_{FET}, and in the saturation region it is equal to that in a resistor of value $3/(2g_m)$.

The flicker noise current in both the saturation and ohmic regions is characterized by the spectral density

$$S_{If} = \frac{2K_f K' I_{DQ}}{C_{ox} L^2 f} \qquad (3.1\text{-}35)$$

where K_f is the flicker noise coefficient, I_{DQ} is the quiescent current, f is frequency, and K', C_{ox}, and L are the MOSFET model parameters introduced in Sec. 3.1.1. K_f is typically[10] about 3×10^{-24} $V^2 \cdot F$. The flicker noise is often termed $1/f$ noise because of the $1/f$ dependence in (3.1-35).

The white noise and $1/f$ noise sources are statistically uncorrelated, and hence the spectral current density of the noise current source, I_N in Fig. 3.1-30*a*, can be obtained by adding S_{IW} and S_{If} to obtain

$$S_N = S_{IW} + S_{If} \qquad (3.1\text{-}36)$$

It is well known that the RMS noise current source in the frequency band $[f_1, f_2]$ can be obtained from the spectral density and is given by

$$I_{NB} = \sqrt{\int_{f_1}^{f_2} S_N \, df} \qquad (3.1\text{-}37)$$

It should be pointed out that if we define the RMS white noise and flicker noise currents by the square root of the integrals of the spectral densities

$$I_{WB} = \sqrt{\int_{f_1}^{f_2} S_{IW}\, df} \tag{3.1-38}$$

and

$$I_{FB} = \sqrt{\int_{f_1}^{f_2} S_{If}\, df} \tag{3.1-39}$$

then these currents add in the RMS sense by the equation

$$I_{NB} = \sqrt{I_{WB}^2 + I_{FB}^2} \tag{3.1-40}$$

The current I_{NB} is the contribution to I_N in Fig. 3.1-30 in the frequency band $[f_1, f_2]$.

The relationship between small signal voltages and/or currents in a circuit and the total output noise voltage or current is often of interest because it characterizes the impact noise has on the performance of a circuit. Two commonly used indicators of this relationship are the signal-to-noise ratio (SNR) and the dynamic range. Since the noise signal amplitudes are usually sufficiently small to be considered small signals, the small signal model of the MOSFET is often used to analyze the noise performance of circuits. This analysis is usually much simpler than a noise analysis based on the large signal models of the MOSFETs.

It can be readily shown that if the MOSFET in Fig. 3.1-30*a* is replaced by its small signal model, then an *equivalent* small signal model that characterizes the noise performance can be obtained by replacing the noise current source from drain to source with a noise voltage source in series with the gate, as denoted by V_N in Fig. 3.1-30*b*. This is termed an *input-referred* noise source. The spectral density of the input-referred noise voltage source relates to the noise spectral density of the noise current source by the expression

$$S_{VN} = \frac{S_{IN}}{g_m^2} \tag{3.1-41}$$

Correspondingly, the input-referred noise voltage source relates to the noise current source by the expression

$$V_N = \frac{I_N}{g_m} \tag{3.1-42}$$

where g_m is the small signal transconductance gain of the MOSFET. In the common case where the MOSFET is operating in the saturation region, the parameter g_m is given by (3.1-11). The input-referred model is useful for simplifying the analysis of some circuits.

Two points about the input-referred noise model of Fig. 3.1-30*b* deserve mention. First, the reader must keep in mind that this is *merely* a small signal

equivalent circuit characterization. The mechanisms that contribute to noise are inherently associated with the drain current. Second, whereas the noise model of Fig. 3.1-30*a* was valid for either a large signal or small signal model of the MOSFET, the input-referred model of Fig. 3.1-30*b* was developed under the assumption of small signal operation of the MOSFET. The equivalence of the circuit in Fig. 3.1-30*b* as characterized by (3.1-42) is not valid for large signal MOSFET operation.

Example 3.1-8. A simple transconductance amplifier is shown in Fig. 3.1-31.

(*a*) Calculate the output noise current spectral density, the input-referred voltage spectral density and the RMS output thermal noise current, the RMS output flicker noise current, the output RMS noise current, and the input-referred RMS noise voltage in the flat frequency band from 100 Hz to 1 MHz if the small signal input, v_i, is zero. Assume $W = 5\ \mu$, $L = 5\ \mu$, $V_{gg} = 2$ V, $K_f = 3 \times 10^{-24}$, $V^2 \cdot F$. Use the typical CMOS parameters of Table 2B.4 in Appendix 2B.

(*b*) Repeat for $W = 200\ \mu$, $L = 5\ \mu$, and $V_{gg} = 1.25$ V.

Solution.

(*a*) Observe that the MOSFET is operating in the saturation region since $V_{DS} = V_{DD} > V_{GS} - V_T$. From (3.1-34) and (3.1-35), both g_m and I_{DQ} are needed.

$$I_{DQ} = \frac{K'W}{2L}(V_{GSQ} - V_T)^2 = 18.75\ \mu\text{A}$$

$$g_m = \sqrt{\frac{2K'W}{L}}\sqrt{I_{DQ}} = 30.0\ \mu\text{A/V}$$

Thus, from (3.1-34), (3.1-35), and (3.1-36),

$$S_{IW} = 3.3 \times 10^{-25}\ \text{A}^2 \cdot \text{sec}$$

$$S_{If} = \frac{1.54 \times 10^{-19}}{f}\ \text{A}^2 \cdot \text{sec}$$

$$S_N = 3.3 \times 10^{-25} + \frac{1.54 \times 10^{-19}}{f}\ \text{A}^2 \cdot \text{sec}$$

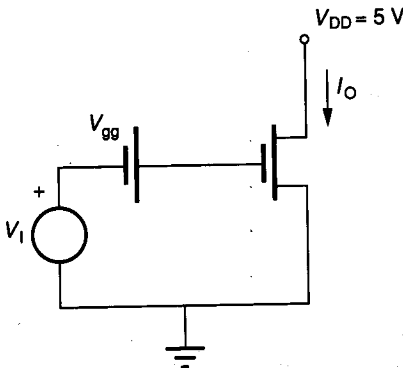


FIGURE 3.1-31
Example 3.1-8.

From (3.1-38), (3.1-39), and (3.1-40),

$$I_{\mathrm{WB}} = \sqrt{\int_{10^2}^{10^6} S_{\mathrm{IW}}\, df} = 0.489 \text{ nA (RMS)}$$

$$I_{\mathrm{FB}} = \sqrt{\int_{10^2}^{10^6} S_{\mathrm{If}}\, df} = 1.190 \text{ nA (RMS)}$$

$$I_{\mathrm{NB}} \doteq \sqrt{I_{\mathrm{WB}}^2 + I_{\mathrm{FB}}^2} = 1.29 \text{ nA (RMS)}$$

and from (3.1-42)

$$V_{\mathrm{N}} = \frac{I_{\mathrm{N}}}{g_{\mathrm{m}}} = 43\ \mu\text{V (RMS)}$$

(*b*) Repeating with the new device sizes,

$$I_{\mathrm{DQ}} = 120\ \mu\text{A}$$

$$g_{\mathrm{m}} = 48 \text{ mA/V}$$

$$S_{\mathrm{IW}} = 5.28 \times 10^{-24} \text{ A}^2 \cdot \text{sec}$$

$$S_{\mathrm{If}} = (9.87 \times 10^{-19}/f) \text{ A}^2 \cdot \text{sec}$$

$$I_{\mathrm{WB}} = 2.29 \text{ nA (RMS)}$$

$$I_{\mathrm{FB}} = 3.01 \text{ nA (RMS)}$$

$$I_{\mathrm{NB}} = 3.78 \text{ nA (RMS)}$$

$$V_{\mathrm{N}} = 7.88\ \mu\text{V (RMS)}$$

Note that in the previous example the $1/f$ noise is strongly dominant in the first part whereas the relative contributions are comparable in the second part. Device sizes and bias current affect both the noise current and the relative significance of the two types of noise. At higher frequencies the thermal noise effects dominate whereas the flicker noise, because of its $1/f$ type spectral density, dominates at lower frequencies.

Example 3.1-9. Obtain an expression in terms of the quiescent excess gate bias, $V_{\mathrm{GS}} - V_{\mathrm{T}}$, that shows where the spectral density of the $1/f$ noise crosses over that of the thermal noise. Assume operation in the saturation region.

Solution. Equating spectral densities from Eqs. 3.1-34 and 3.1-35, we obtain

$$\frac{8kT g_{\mathrm{m}}}{3} = \frac{K_{\mathrm{f}} 2K' I_{\mathrm{DQ}}}{f C_{\mathrm{ox}} L^2}$$

From (3.1-4) and (3.1-11), this simplifies to

$$f = \frac{3}{8} \frac{K' K_{\mathrm{f}}}{kT C_{\mathrm{ox}}} \left(\frac{V_{GS} - V_T}{L^2} \right)$$

Using typical values of these parameters from Table 2B.4, we obtain

$$f \simeq \left(9.5\,\frac{\mu^2}{\mathrm{V}}\right)\left(\frac{V_{\mathrm{GS}} - V_{\mathrm{T}}}{L^2}\right)\ \mathrm{MHz}$$

From Example 3.1-9 it can be observed that the relative impact of the flicker noise and thermal noise are strongly dependent on the quiescent operating point and the device size.

3.1.9 Simple MOSFET Models for Digital Applications

Many digital applications involve large numbers of gates and transistors, which make hand analysis using the device models, presented in Sec. 3.1.1 and summarized in Table 3.1-1, totally impractical. A very simple model that makes hand analysis tractable is needed for these applications. Such a model will, of necessity, not provide highly accurate results. We are willing to accept the tradeoffs between model accuracy and simplicity in these applications. Models that are widely used for this purpose are shown in Fig. 3.1-32.

In the circuit of Fig. 3.1-32*a*, the switch S1 is open if $V_{\mathrm{GS}} < V_{\mathrm{H}}/2$ and closed if $V_{\mathrm{GS}} > V_{\mathrm{H}}/2$ where V_{H} is the logical *high* voltage in the digital logic family and where the logical *low* voltage is assumed to be near zero. The values of C_μ and R are given by

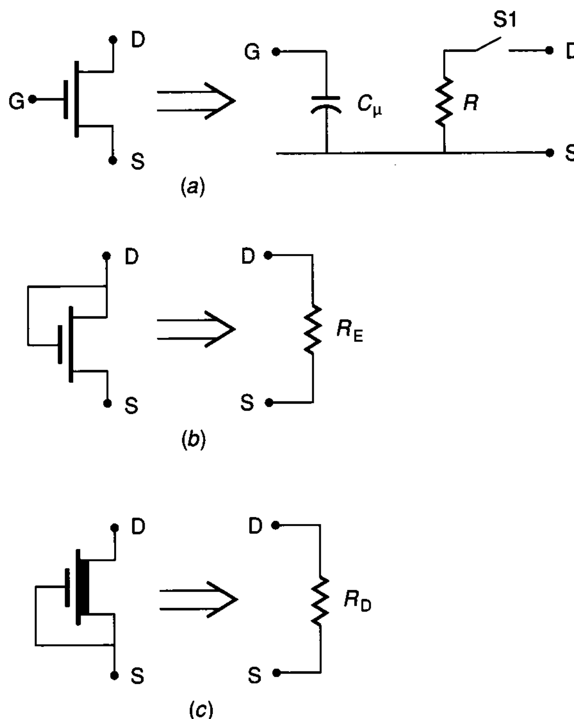


FIGURE 3.1-32
Models for the MOSFET used for simple analysis of digital circuits: (*a*) Three-terminal MOSFET, (*b*) Two-terminal enhancement, (*c*) Two-terminal depletion.

$$C_\mu = C_{ox} W L \tag{3.1-43}$$

$$R = \frac{L}{K'W(V_H - V_T)} \tag{3.1-44}$$

where V_T is the threshold voltage of the MOSFET. The two-terminal enhancement and depletion configurations of Figs. 3.1-32*b* and *c* are characterized by

$$R_E \doteq \frac{L}{K'W(V_H - V_{TE})} \tag{3.1-45}$$

$$R_D = \frac{L}{K'W\,|V_{TD}|} \tag{3.1-46}$$

where V_{TE} and V_{TD} denote the threshold voltages of the enhancement and depletion transistors.

The model of Fig. 3.1-32*a* is motivated by the observation that in most digital circuits, the inputs to the logic gates are typically inputs to the gates of transistors, which swing between the high and low logic levels. From the results presented in Section 3.1.3, it follows that the gate input port is capacitive with a value near $C_{ox} W L$ throughout most of the input signal swing. When the input is low, the output port looks nearly like an open circuit; and when the gate of the MOSFET is driven to V_H, which is typically much larger than V_T, the output voltage is sufficiently low through most of the high to low transition to force operation in the ohmic region. The ohmic region impedance which is modeled by R in Fig. 3.1-32 is characterized by (3.1-8), which is repeated in (3.1-44). In applications where the gate is permanently connected to the drain or source as shown in Fig. 3.1-32*b* and *c*, the approximation of (3.1-8) is less justifiable but still widely used (see Problem 3.17). These models are used for simple dc and timing analysis in Chapter 7 with modest modifications of the resistances given in (3.1-45) and (3.1-46) to obtain closer agreement between theoretical and experimental results.

Example 3.1-10. Using the simple models of Fig. 3.1-32, obtain the dc transfer characteristics, V_O versus V_I, for the circuit of Fig. 3.1-33. Use the typical process parameters of Table 3.1-2.

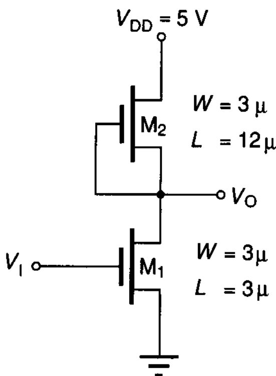


FIGURE 3.1-33
Simple digital inverter.

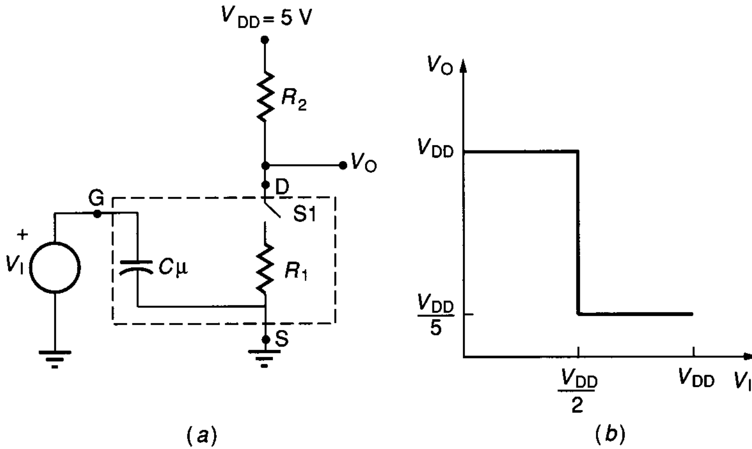


FIGURE 3.1-34
Solution of Example 3.1-10: (*a*) Equivalent circuit, (*b*) Plotting of results.

Solution. It follows from the models of Fig. 3.1-32 that the equivalent circuit can be drawn as shown in Fig. 3.1-34*a*. Assuming the logic high, V_H, is 5 V, it follows from Table 3.1-2 and Eqs. 3.1-44 and 3.1-45 that $R_2 = 39.2$ kΩ and $R_1 = 9.8$ kΩ. For $V_I < V_{DD}/2$, switch S1 is open, so $V_O = V_{DD} = 5$ V. For $V_I > V_{DD}/2$, the switch S1 is closed and the resistive voltage divider yields an output of $V_O = V_{DD}/5 = 1$ V. These idealized results are plotted in Fig. 3.1-34*b*.

3.2 DIODE MODELS

The pn junction diode with the convention used here for the electrical variables is shown in Fig. 3.2-1. The junction diode is characterized by a region of n-type material adjacent to a region of p-type material. A depletion region is formed at the interface. With a sufficiently large forward bias, considerable current will flow, but very little current flows under reverse bias. Such a device, which passes current in one direction and blocks it in the opposite direction, has many practical applications.

3.2.1 dc Diode Model

The diode depicted in Fig. 3.2-1 is characterized at low frequencies by the equation

$$I = I_S\left(e^{\frac{V}{nV_t}} - 1\right) \tag{3.2-1}$$

where the parameters I_S, n, and V_t characterize the device. The parameters in the equation are designated by

I_S = saturation current

n = emission coefficient

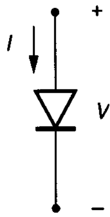


FIGURE 3.2-1
Variable convention for a diode.

and the parameter V_t is given by the equation

$$V_t = \frac{kT}{q} \tag{3.2-2}$$

where k is Boltzmann's constant ($k = 1.381 \times 10^{-23}$ V·C/°K), T is absolute temperature in degrees Kelvin, and q is the charge of an electron ($q = 1.6 \times 10^{-19}$ C). At room temperature (300 °K), $V_t = 26$ mV. Equation 3.2-1 is often called the "diode equation."

The emission coefficient takes on a value of about 1 for silicon. The saturation current can be expressed as

$$I_S = J_S A \tag{3.2-3}$$

where J_S is a process parameter equal to the saturation current density and A is the cross-sectional area of the junction.

The $I-V$ characteristics of a typical diode are shown in Fig. 3.2-2. The current scale for $I < 0$ is considerably smaller than that for $I > 0$ to allow for plotting of the entire diode characteristics, in a meaningful manner, on one set of axes.

The breakdown that occurs for a large reverse bias is not predicted by the diode equation. In discrete diodes this may be quite large (often −100 V or higher). Since the knee is typically very sharp, the breakdown can be gainfully used in voltage-reference applications. Diodes that are intentionally designed to operate in breakdown are called zener diodes. The breakdown voltage for zener diodes is intentionally reduced. Discrete zener diodes with breakdown voltage specifications from the low volt range to over 20 V are readily available.

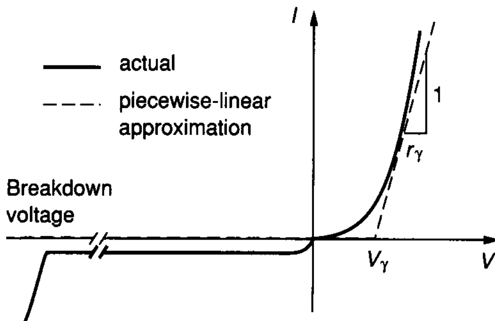


FIGURE 3.2-2
I–V characteristics of a diode.

In IC applications, diodes are often parasitic devices formed in the process of making other devices, as can be seen from the descriptions of the typical CMOS and bipolar processes in Chapter 2. As such, the process is tailored to optimize performance of the MOSFET (in a MOS process) or the BJT (in a bipolar process) rather than the diode. The reverse breakdown voltage of the base-emitter junction in a bipolar process is typically in the 7 V range, making this diode a useful parasitic zener diode. The reverse breakdowns of the parasitic diodes in the MOS processes and for the base-collector junction in a bipolar process are typically much larger. The latter junctions are seldom operated intentionally in the breakdown region.

The diode equation simplifies considerably for $V >> nV_t$. Under forward biases greater than 0.25 V, essentially no accuracy is lost in modeling the diode by

$$I = I_S e^{\frac{V}{nV_t}} \tag{3.2-4}$$

Likewise, for $V < -0.25\ V$, the diode is accurately modeled by

$$I = -I_S \tag{3.2-5}$$

The piecewise-linear model indicated by the dashed curves in Fig. 3.2-2 is often used for biasing and/or large signal applications. The piecewise-linear model is mathematically characterized by the equations

$$\left.\begin{array}{ll} V = I r_\gamma + V_\gamma & \text{for } I > 0 \\ I = 0 & \text{for } V < V_\gamma \end{array}\right\} \tag{3.2-6}$$

where r_γ is the *on* resistance of the diode and V_γ is the diode cut-in voltage.

The equivalent circuit corresponding to this piecewise-linear model is shown in Fig. 3.2-3. The value of r_γ is obtained by measuring the slope of the $I-V$ characteristics near the desired point of operation. V_γ is the V-axis intercept of the corresponding tangent line; r_γ is typically in the tens of ohms range while $V_\gamma \simeq 0.7$ V. An ideal diode is one in which $r_\gamma = 0$ and $V_\gamma = 0$. In many applications, little accuracy is lost by assuming $r_\gamma = 0$.

Example 3.2-1. Assume $I_S = 10^{-14}$A, T=300°K, and $n = 1$.
(*a*) Show that $V \simeq 0.7$ V for diode currents in the 1 mA to 100 mA range.
(*b*) Determine V_γ and r_γ at $I = 1$ mA and $I = 100$ mA.

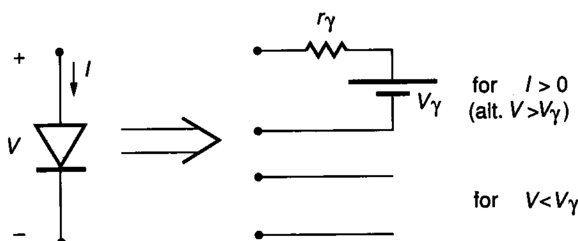


FIGURE 3.2-3
Piecewise-linear diode model.

Solution.

(*a*) Referring to the variable convention of Fig. 3.2-1, assume that $V > 0.25$ V (this must be later verified). It thus follows from Eq. 3.2-4 that

$$V = nV_t \ln\left(\frac{I}{I_S}\right)$$

At about room temperature, we thus have for 1 mA $\leq I \leq$ 100 mA

$$0.658 \text{ V} \leq V \leq 0.778 \text{ V}$$

The initial assumption that $V > 0.25$ V is verified and the conclusion that $V_\gamma \simeq 0.7$ V follows.

(*b*) From part (*a*), it suffices to consider the model of (3.2-4) at $I = 1$ mA and $I = 100$ mA. Upon differentiation of (3.2-4) we obtain

$$r_\gamma = \frac{nV_t}{I} \tag{3.2-7}$$

which at 1 mA and 100 mA becomes 26 Ω and 0.26 Ω respectively. The corresponding V_γs are given by

$$V_\gamma = V - Ir_\gamma$$

where V and I are the coordinates of the operating point. It follows from part (*a*) and the values obtained for r_γ that V_γ is 0.632 V for $I = 1$ mA and V_γ is 0.752 V for $I = 100$ mA. In practical applications, V_γ is often approximated by 0.7 V, independent of I.

3.2.2 Small Signal Diode Model

The diode is seldom used in low-frequency small signal applications and, as such, a small signal model need not be developed. Since the small signal model of the diode is, however, very simple, it will be mentioned in passing. For reverse bias, the low-frequency small signal model is an open circuit. Under forward bias, it is modeled by a single resistor of value determined by (3.2-7), evaluated at the dc operating point. The diode is occasionally used in high-frequency small signal applications as a capacitor. When used as a capacitor, the diode is often termed a vari-cap or a varactor diode. These diode capacitances are strongly dependent upon the quiescent voltage. A discussion of the high-frequency diode model follows.

3.2.3 High-Frequency Diode Model

At high frequencies, the capacitance formed by the depletion region in the pn junction becomes significant. This capacitance is modeled by a capacitor connected between the terminals of the diode of value

$$C = \begin{cases} \dfrac{C_{j0}A}{\left(1 - \dfrac{V}{\phi_B}\right)^n} & V < \dfrac{\phi_B}{2} \quad (3.2\text{-}8) \\[2ex] 2^n C_{j0}A\left[\dfrac{2nV}{\phi_B} + (1-n)\right] & V > \dfrac{\phi_B}{2} \quad (3.2\text{-}9) \end{cases}$$

where the parameters C_{j0}, A, ϕ_B, and n are as defined for the parasitic junction capacitances of the MOSFET in Section 3.1.3. Note that for $V < \phi_B/2$, this parasitic capacitance agrees with that of the MOSFET given by (3.1-14). It can be readily shown (see Problem 3.31) that the value for $V > \phi_B/2$ is the continuous and differentiable linear extension of the reverse-bias value across the boundary defined by $V = \phi_B/2$.

3.3 BIPOLAR MODELS

The npn and pnp bipolar transistors and the convention for the electrical variables established in Chapter 2 are shown in Fig. 3.3-1, along with a top view of a typical geometrical layout of a bipolar transistor. The following model development will

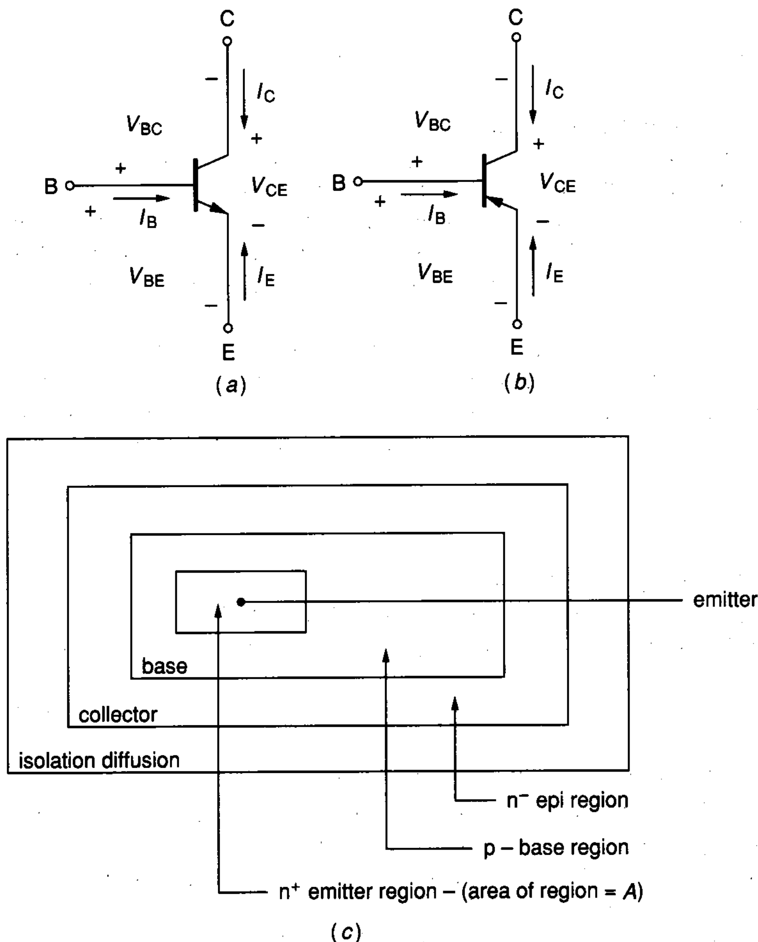


FIGURE 3.3-1
Convention for electrical variables of BJT transistors: (*a*) npn, (*b*) pnp, (*c*) Top view of vertical npn transistor geometry.

be restricted to the npn transistor. The pnp development is identical with the exception of sign changes in some of the equations. Results only for the pnp devices will be presented following the npn discussion.

3.3.1 dc BJT Model

The starting point will be the dc model introduced by Ebers and Moll in 1954[11]. In this model the relationship between the terminal variables is given by

$$I_C = I_S(e^{(V_{BE}/V_t)} - 1) - \frac{I_S}{\alpha_R}(e^{(V_{BC}/V_t)} - 1) \tag{3.3-1}$$

$$I_E = \frac{-I_S}{\alpha_F}(e^{(V_{BE}/V_t)} - 1) + I_S(e^{(V_{BC}/V_t)} - 1) \tag{3.3-2}$$

where the four parameters I_S, α_R, α_F, and V_t characterize the device. The parameters in these equations are designated by

I_S = transport saturation current

α_R = large signal reverse current gain of common base configuration

α_F = large signal forward current gain of common base configuration

and

$$V_t = \frac{kT}{q} \tag{3.3-3}$$

where k is Boltzmann's constant ($k = 1.381 \times 10^{-23}$ V· C/°K), T is absolute temperature in degrees Kelvin, and q is the charge of an electron ($q = 1.602 \times 10^{-19}$ C). At room temperature (300 °K), $V_t \simeq 26$ mV. This is the same V_t that appeared in the diode model and the subthreshold MOSFET model.

The two Ebers-Moll equations along with KVL and KCL applied to the transistor itself

$$I_B = -I_C - I_E \tag{3.3-4}$$

$$V_{CE} = V_{BE} - V_{BC} \tag{3.3-5}$$

provide four independent equations relating the six terminal variables, $I_C, I_B, I_E, V_{CE}, V_{BC}$, and V_{BE}. The BJT is characterized by these four equations.

It should be noted that there is complete functional symmetry in the device model with respect to the collector and emitter terminals. This is comforting since the npn transistor can be thought of as a sandwich of a p-type region between two n-channel layers. Since the shape and impurity concentrations of the collector and emitter differ, however, one would not expect these terminals to be interchangeable. The lack of geometrical symmetry causes large differences in α_F and α_R.

The parameters α_F and α_R are determined by impurity concentrations and junction depths, and as such are process parameters. The parameter V_t is a function of two physical constants and temperature and thus cannot be considered

a design parameter. The parameter V_t should not be confused with the threshold voltage for a MOSFET, which uses the symbol V_T. Because of the temperature dependence, V_t can be considered an environmental parameter. The transport saturation current can be expressed as

$$I_S = J_S A \tag{3.3-6}$$

where A is the area of the emitter and J_S, the transport saturation current density, is a process parameter. The parameter J_S can be further expressed as

$$J_S = q\overline{D}_n n_i^2 / Q_B \tag{3.3-7}$$

where $\overline{D}_n$ is the average effective electron diffusion constant, n_i is the intrinsic carrier concentration in silicon ($n_i = 1.45 \times 10^{10}$ atoms/cm^3 at 300 °K), and Q_B is the number of doping atoms in the base per unit area.

From the model presented, it can be seen that the emitter area, A, is the only design parameter available to the design engineer for the vertical bipolar transistor. The shape of the emitter, the size and shape of the base, and the size and shape of the collector have no effect on the dc model of the BJT just introduced.

Typical values of $\overline{D}_n$ and Q_B are 13 cm^2/sec and 7.25×10^{12} atoms/cm^2, respectively. α_R is typically between 0.3 and 0.8. α_F is generally close to unity, falling in the range of 0.9 to 0.999. Typical values for the bipolar parameters that will be used in examples in this text are given in Table 3.3-1.

Labeling of the emitter and collector terminals is arbitrary at this point. Convention has been established, however, for making this distinction. The typical values for α_R and α_F given in the previous paragraph follow this convention. Specifically, the emitter and collector designations are established so that $\alpha_F > \alpha_R$.

An equivalent circuit for the bipolar transistor based upon the Ebers-Moll model is sometimes useful. The circuit of Fig. 3.3-2 serves as such an equivalent circuit where the currents in the two diodes satisfy the standard diode equations

$$I_F = \frac{I_S}{\alpha_F}\left(e^{(V_{BE}/V_t)} - 1\right) \tag{3.3-8}$$

$$I_R = \frac{I_S}{\alpha_R}\left(e^{(V_{BC}/V_t)} - 1\right) \tag{3.3-9}$$

TABLE 3.3-1
Typical process parameters for vertical npn transistor in bipolar technology

Parameter	Value	
J_S	$= 6 \times 10^{-10}$	μA/mil^2
α_F	$= 0.99$	($\beta_F = 100$)
α_R	$= 0.3$	($\beta_R = 0.43$)
V_{AF}	$= 200$ V	
V_{AR}	$= 200$ V	

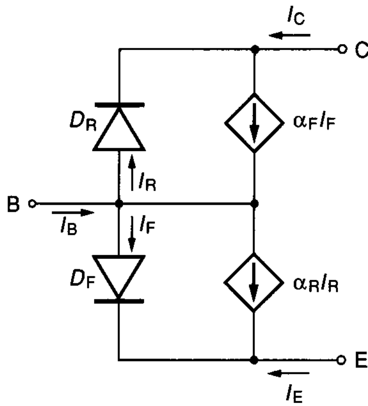


FIGURE 3.3-2
Equivalent circuit of BJT based upon Ebers-Moll model.

Several different regions of operation for the BJT are readily identifiable. Those that are of the most interest are depicted in the V_{BC}–V_{BE} plane shown in Fig. 3.3-3. The forward active, reverse active, and saturation regions are defined by the dashed curves. The cutoff region is defined to be the third quadrant of the V_{BC}–V_{BE} plane. No particular designation is made of the lightly shaded region. Most applications involve operation in one of the four well-defined regions, with possible excursions through the lightly shaded region in digital circuits during switching transitions. For power dissipation reasons, operation in the reverse and forward active regions is generally restricted further to the hatched regions and in the saturation region to the heavily shaded region (see Problem 3.40). It will be shown that major simplifications in the Ebers-Moll equations are possible in all except the lightly shaded region of Fig. 3.3-3. These simplifications are crucial for hand analysis because of the unwieldiness of the Ebers-Moll Eqs. 3.3-1 and 3.3-2 in even relatively simple problems (see Problem 3.38). Simplified models for the forward and reverse active, saturation, and cutoff regions will now be discussed.

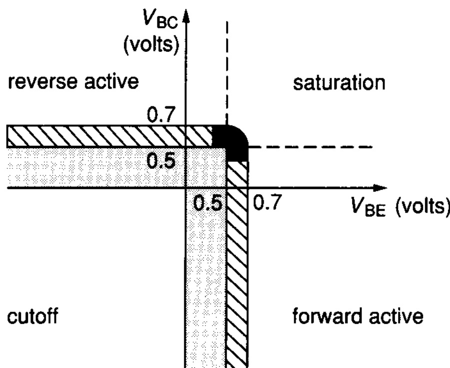


FIGURE 3.3-3
Regions of operation of npn silicon bipolar transistor.

FORWARD AND REVERSE ACTIVE REGIONS OF OPERATION. The BJT is typically operated in the forward active region in most linear applications. If $V_{BE} > 0.5$ V and $V_{BC} < 0.3$ V, then it can be shown that (3.3-1) and (3.3-2) simplify to

$$I_C = I_S e^{(V_{BE}/V_t)} \tag{3.3-10}$$

$$I_E = -\frac{I_S}{\alpha_F} e^{(V_{BE}/V_t)} \tag{3.3-11}$$

Note that "=" was used rather than "≃" in these equations. A simple calculation (see Problem 3.39) will show that the terms dropped in obtaining (3.3-10) and (3.3-11) are several orders of magnitude smaller than the terms that were retained!

It is often more convenient to use I_B and I_C rather than I_C and I_E as the dependent variables when modeling the BJT. Defining β_F by

$$\beta_F = \frac{\alpha_F}{1 - \alpha_F} \tag{3.3-12}$$

it follows from (3.3-10) and (3.3-11) that I_C and I_B can be written as

$$I_C = I_S e^{(V_{BE}/V_t)} \tag{3.3-13}$$

$$I_B = \frac{I_S}{\beta_F} e^{(V_{BE}/V_t)} \tag{3.3-14}$$

Note that in the forward active region the BJT acts as a good current amplifier from the base to the collector with the gain relationship

$$I_C = \beta_F I_B \tag{3.3-15}$$

This should be contrasted to the MOSFET, which is inherently a good transconductance amplifier.

Similarly, if β_R is defined as

$$\beta_R = \frac{\alpha_R}{1 - \alpha_R} \tag{3.3-16}$$

then the Ebers-Moll equations in the reverse active region ($V_{BC} > 0.5$ V, $V_{BE} < 0.3$ V) simplify to

$$I_E = I_S e^{(V_{BC}/V_t)} \tag{3.3-17}$$

$$I_B = \frac{I_S}{\beta_R} e^{(V_{BC}/V_t)} \tag{3.3-18}$$

As before, it follows that the base-to-emitter current gain in the reverse active region is given by

$$I_E = \beta_R I_B \tag{3.3-19}$$

Actually, it can be shown both theoretically and experimentally that the collector current of the BJT in the forward active region increases slightly in

approximately a linear manner with V_{CE}. Defining V_{AF} to be the coefficient that represents the linear dependence on V_{CE}, the following improvement on (3.3-13) can be made:

$$I_C \simeq I_S e^{(V_{BE}/V_t)}\left(1 + \frac{V_{CE}}{V_{AF}}\right) \tag{3.3-20}$$

Correspondingly, this type of linear dependence in the reverse active region can be characterized by the parameter V_{AR}, resulting in the following improvement on (3.3-17):

$$I_E \simeq I_S e^{(V_{BC}/V_t)}\left(1 - \frac{V_{CE}}{V_{AR}}\right) \tag{3.3-21}$$

The parameters V_{AF} and V_{AR} are process parameters and are referred to as the forward and reverse Early voltages, respectively. V_{AF} and V_{AR} for the BJT serve the same role as the parameter λ (actually $1/\lambda$) in the MOSFET, as can be seen by comparing Eqs. 3.3-20 and 3.1-5.

Many IC applications require base currents in the 0.1 μA to 1 mA range. It is shown in Example 3.3-1 that in the forward active region the base–emitter voltage is typically in the 0.5 V to 0.75 V range. For hand manipulations it is often sufficient to obtain the base–emitter voltage to within 0.1 V ($V_{BE} = 0.6$ V is often used) in this region.

Example 3.3-1. Show that at around room temperature the base–emitter voltage is between 0.5 V and 0.75 V for base currents in the 0.1 μA to 1 mA range for transistors with 1 mil^2 emitter areas and reverse-biased base–collector junctions. Use the process parameters of Table 3.3-1.

Solution. From Eqs. 3.3-6 and 3.3-14 it follows that

$$I_B \simeq \frac{AJ_S}{\beta_F} e^{(V_{BE}/V_t)}$$

With $A = 1\ \text{mil}^2$ and $J_S = 6 \times 10^{-10}\ \mu\text{A/mil}^2$, we can write

$$0.1\ \mu\text{A} \leq 6 \times 10^{-10} e^{(V_{BE}/.026\ \text{V})} \leq 1\ \text{mA}$$

It thus follows that

$$0.5\ \text{V} < V_{BE} < 0.75\ \text{V} \tag{3.3-22}$$

It can be shown that even with substantial changes in temperature, J_S, or A, the changes in V_{BE} for the current range considered in this example are modest.

An even simpler model is often used for characterizing the BJT in the forward and reverse active regions. As indicated in Example 3.3-1, the base–emitter voltage is about 0.6 V in the forward active region for a wide range of base currents. The model

$$V_{BE} = 0.6\ \text{V} \tag{3.3-23}$$

$$I_C = \beta_F I_B \tag{3.3-24}$$

is often adequate for biasing applications and operating point calculations when the device is operating in the forward active region. In the reverse active region the simplified model becomes

$$V_{BC} = 0.6 \text{ V} \tag{3.3-25}$$

$$I_E = \beta_R I_B \tag{3.3-26}$$

The reader should be cautioned, however, not to attempt to use the value of V_{BE} or V_{BC} as given by (3.3-23) or (3.3-25) directly in the Ebers-Moll equations to obtain I_C or I_B. Small errors in these junction voltages will result in large errors in the currents because of the exponential relationship of the model.

SATURATION REGION. It will be seen later that most (although not all) digital applications involve switching the BJT between the saturation and cutoff regions. When the device is operating in the saturation region with large positive base and collector currents, it is operating near the forward active region. The base–emitter voltage will be near 0.7 V and the base–collector voltage near 0.5 V. The device can thus be modeled by two voltage sources, $V_{BE} = 0.7$ V and $V_{CE} = 0.2$ V, when operating in this region. This collector–emitter voltage in saturation is often termed V_{CESAT}. When the device is operating in saturation with a large positive base current and a large negative collector current, the device is operating near the reverse active region. The base–collector voltage will be near 0.7 V and the base–emitter voltage near 0.5 V. In this case the device will be modeled by two voltage sources, $V_{BC} = 0.7$ V and $V_{CE} = -0.2$ V. This simple modeling approach for the saturation region will be adequate for most hand manipulations. The reader again must be cautioned that these are approximate junction voltage values and thus cannot be used to obtain I_C and I_E in the Ebers-Moll equations since small changes or errors in V_{BC} and V_{BE} cause large changes in the corresponding currents. When the collector and base currents must be accurately determined for devices operating in the saturation region, they are often obtained from equations describing the network itself (KVL or KCL) rather than directly from the Ebers-Moll equations. This situation is illustrated in the following example.

Example 3.3-2. Determine good approximations (± 5% accuracy is close enough) for the base current, collector current, V_{BE}, and V_{CE} for the circuit of Fig. 3.3-4

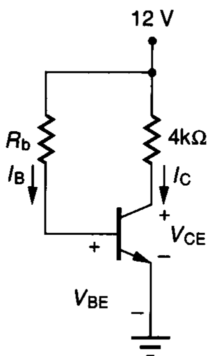


FIGURE 3.3-4
Example 3.3-2.

if (*a*) R_b = 500 kΩ, (*b*) R_b = 100 kΩ. Also obtain an accurate expression for V_{BE}.
Assume the transistor has an emitter area of 4 mil^2, $\beta_F = 100$, $\beta_R = 0.2$, $J_S = 6 \times 10^{-10}\ \mu A/mil^2$, and T = 300 °K.

Solution. Assume initially that I_B is in the 0.1 μA to 1 mA range. From Example 3.3-1 it can be concluded that $V_{BE} \simeq 0.6$ V and consequently the current through R_b, which is also the base current, can be approximated by $I_b \simeq (11.4\ V)/R_b$.

(*a*) With R_b = 500 kΩ, it follows that $I_B \simeq 0.0228$ mA, which satisfies the criterion established in Example 3.3-1 on I_B and thus justifies the initial assumption $V_{BE} \simeq 0.6$ V. If it is now assumed (a fact that must later be verified) that the device is operating in the forward active region, it follows from (3.3-15) that $I_C \simeq 2.28$ mA. With $I_C \simeq 2.28$ mA it follows that $V_{CE} = 12-(4\ k\Omega)(2.28\ mA) = 2.88$ V. With V_{BE} in the 0.6 V range, it thus can be concluded that $V_{BC} < 0$, thus verifying the initial assumption of operation in the forward active region. It remains to accurately determine V_{BE}. Since I_C is now approximately known, it follows from (3.3-13) with V_t = 26 mV that

$$V_{BE} = (.026\ V)\ln(I_C/I_S)$$
$$= (.026\ V)\ln[2.28\ mA/(6 \times 10^{-10} \cdot 4\ \mu A)] = 0.717\ V$$

(*b*) With R_b = 100 kΩ, it follows that $I_B \simeq (11.4\ V)/(100\ k\Omega) = 0.114$ mA. If we were to assume, as in part (*a*), that the transistor were operating in the forward active region, we would conclude that I_C = 11.4 mA. Consequently, $V_{CE} = -33.6$ V, yielding a value for V_{BC} of 34.2 V, which is in strong violation of the assumption $V_{BC} < 0.3$ V that was necessary to obtain (3.3-15). It can thus be concluded that the collector current is too large for operation in the forward active region and thus the device is operating in saturation near the forward active region. From the above modeling in the saturation region it can be concluded that

$$V_{BE} \simeq 0.7\ V$$

$$I_B \simeq \frac{12\ V - 0.7\ V}{100\ k\Omega} = 0.113\ mA$$

$$V_{CE} \simeq 0.2\ V$$

$$I_C \simeq \frac{12\ V - 0.2\ V}{4\ k\Omega} = 2.95\ mA$$

CUTOFF. When the BJT is operating in cutoff, the collector and base currents are very small compared to those which typically flow in the active and saturation regions. In many applications it is adequate to model the device with $I_C = 0$ and $I_B = 0$ when operating in the cutoff region. The relative magnitude of the currents in the BJT when operating in cutoff to those when operating in the forward active region are shown in Example 3.3-3.

BJT MODEL SUMMARY. A summary of the simplified dc model for the BJT in the four major regions of operation—forward active, reverse active, saturation, and cutoff—appears in Table 3.3-2. This simplified model should be adequate for most hand calculations except in the seldom used transition region (between cutoff and forward or reverse active) indicated by the lightly shaded region in

TABLE 3.3-2
Simplified dc model for bipolar transistors

npn			
$I_C = J_S A e^{(V_{BE}/V_t)}[1 + (V_{CE}/V_{AF})]$	(1)	Forward active	$V_{BE} > 0.5$ V
$I_B = \dfrac{J_S}{\beta_F} A e^{(V_{BE}/V_t)}$	(2)		$V_{BC} < 0.3$ V
$I_E = J_S A e^{(V_{BC}/V_t)}[1 - (V_{CE}/V_{AR})]$	(3)	Reverse active	$V_{BC} > 0.5$ V
$I_B = \dfrac{J_S}{\beta_R} A e^{(V_{BC}/V_t)}$	(4)		$V_{BE} < 0.3$ V
$I_B = I_C = 0$	(5)	Cutoff	$(V_{CE} < 0,\ V_{BC} < 0)$
$V_{BE} = 0.7$ V, $V_{CE} = 0.2$ V	(6)	Forward saturated	
$V_{BC} = 0.7$ V, $V_{CE} = -0.2$ V	(7)	Reverse saturated	
pnp			
$I_C = -J_S A e^{-(V_{BE}/V_t)}[1 - (V_{CE}/V_{AF})]$	(8)	Forward active	$V_{BE} < -0.5$ V
$I_B = -\dfrac{J_S A}{\beta_F} e^{-(V_{BE}/V_t)}$	(9)		$V_{BC} > -0.3$ V
$I_E = -J_S A e^{-(V_{BC}/V_t)}[1 + (V_{CE}/V_{AR})]$	(10)	Reverse active	$V_{BC} < -0.5$ V
$I_B = -\dfrac{J_S A}{\beta_R} e^{-(V_{BC}/V_t)}$	(11)		$V_{BE} > -0.3$ V
$I_B = I_C = 0$	(12)	Cutoff	$(V_{CE} > 0,\ V_{BC} > 0)$
$V_{BE} = -0.7$ V, $V_{CE} = -0.2$ V	(13)	Forward saturated	
$V_{BC} = -0.7$ V, $V_{CE} = +0.2$ V	(14)	Reverse saturated	

Design parameters: A
Process parameters: $J_S, \beta_F, \beta_R, V_{AR}, V_{AF}$
Environmental parameter: V_t
All design, process, and environmental parameters are positive for both npn and pnp transistors.

Fig. 3.3-3. For operation in this transition region, the more complicated Ebers-Moll equations 3.3-1 and 3.3-2 are needed.

For biasing or operating point calculations, an even simpler model is often adequate. This model is summarized in Table 3.3-3. An equivalent circuit appears in Fig. 3.3-5. This model can be used in place of the model of Table 3.3-2 to reduce computational complexity with minimal loss of accuracy in most biasing applications. This simplified model is generally justified if the base current can be reasonably accurately determined from the circuit in which the BJT is embedded (as in Example 3.3-2) rather than from the exponential relationship between I_B and V_{BE} (as in Problem 3.38*b*).

Early voltage effects can be readily included in the Ebers-Moll equations 3.3-1 and 3.3-2 by multiplying I_C by $(1 + V_{CE}/V_{AF})$ for $V_{CE} > 0$ or multiplying I_E by $(1 - V_{CE}/V_{AR})$ for $V_{CE} < 0$. Since the BJT is inherently a good current amplifier, it is common to look at the collector current versus V_{CE} for different

TABLE 3.3-3
Simplified model of BJT used for biasing and operating point calculations

Region	npn		pnp	
Forward active	$I_C = \beta_F I_B$	(1)	$I_C = \beta_F I_B$	(11)
	$V_{BE} = 0.6$ V	(2)	$V_{BE} = -0.6$ V	(12)
Reverse active	$I_E = \beta_R I_B$	(3)	$I_E = \beta_R I_B$	(13)
	$V_{BC} = 0.6$ V	(4)	$V_{BC} = -0.6$ V	(14)
Forward saturated	$V_{BE} = 0.7$ V	(5)	$V_{BE} = -0.7$ V	(15)
	$V_{CE} = 0.2$ V	(6)	$V_{CE} = -0.2$ V	(16)
Reverse saturated	$V_{BC} = 0.7$ V	(7)	$V_{BC} = -0.7$ V	(17)
	$V_{CE} = -0.2$ V	(8)	$V_{CE} = 0.2$ V	(18)
Cutoff	$I_C = 0$	(9)	$I_C = 0$	(19)
	$I_B = 0$	(10)	$I_B = 0$	(20)

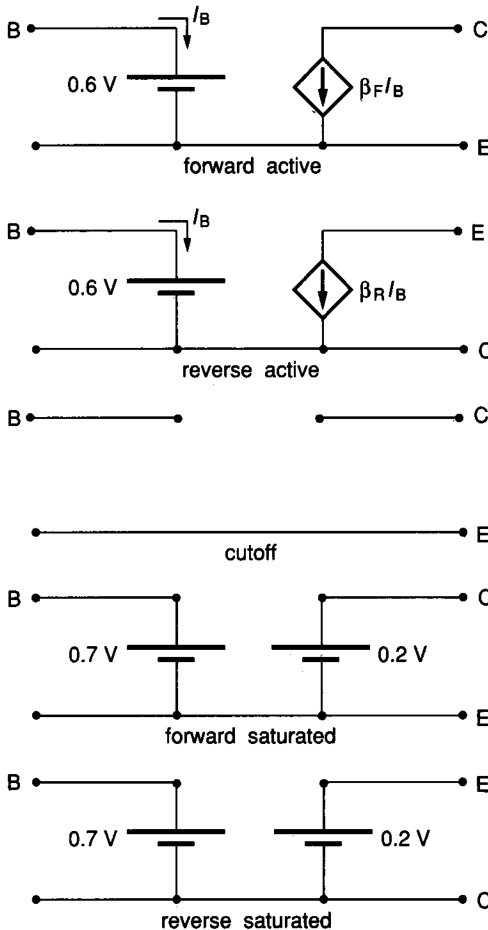


FIGURE 3.3-5
Equivalent circuits of npn BJT for biasing and operating point calculations.

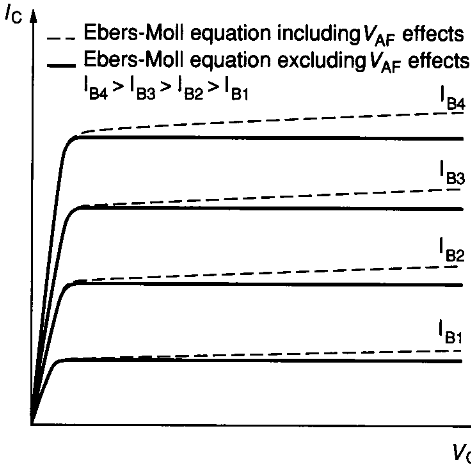


FIGURE 3.3-6
Typical output characteristics of npn BJT based upon Ebers-Moll model.

values of the base current, I_B. Such a plot describes what are termed the *output characteristics*. A plot of the output characteristics for a typical BJT is shown in Fig. 3.3-6 for operation in Quadrant 1 of the I_C–V_{CE} plane. The same basic relationship (except curves are spaced closer together) is obtained in Quadrant 3. Results are shown with and without Early voltage effects. The regions of operation of the BJT in the I_C–V_{CE} plane are shown in Fig. 3.3-7. Comparison of the nomenclature used to define the regions of operation of the BJT in Fig. 3.3-7 to that used to identify the corresponding regions for the MOSFET of Fig. 3.1-7 shows that the term *saturation* is used to define *nonanalogous* regions for the BJT and MOSFET. As mentioned when the terminology for the operating regions of the MOSFET was discussed, the reader should be aware of this nomenclature to avoid possible confusion.

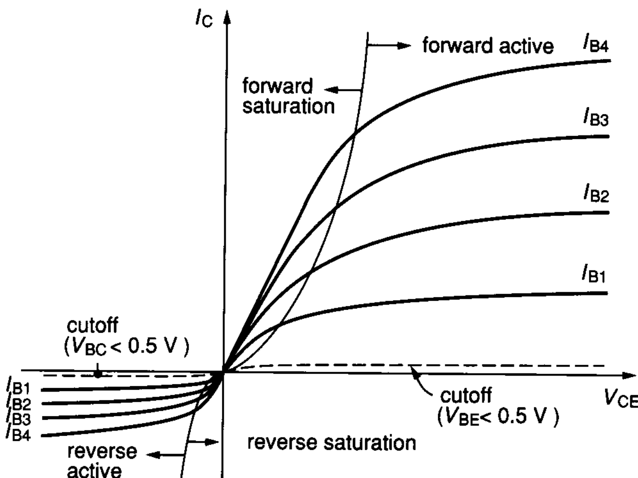


FIGURE 3.3-7
Regions of operation for BJT in I_C–V_{CE} plane.

Example 3.3-3. Compare, using the Ebers-Moll model, the collector and base currents in the forward active region to those in cutoff for a transistor with 1 mil^2 emitter area if $V_{BE} = 0.66$ V, $V_{BC} = -10$ V in the forward active region, and $V_{BE} = -0.3$ V, $V_{BC} = -10$ V in the cutoff region. Use the typical process parameters of Table 3.3-1.

Solution. From (3.3-1) and (3.3-2), I_C and I_B in cutoff are 1.4×10^{-9} μA and 1.406×10^{-9} μA respectively. In the forward active region, it follows from the same equations that $I_C = 63.4$ μA and $I_B = 0.641$ μA. The relative "smallness" of the device currents in the cutoff region should be apparent from these calculations.

3.3.2 Small Signal BJT Model

The small signal model of the bipolar transistor will now be obtained. From Sec. 3.0.2 it can be concluded that the small signal model can be obtained directly from the dc model introduced in the last section. Since there are four regions of operation that have been previously identified in the dc model, there is a different small signal model of the BJT corresponding to each of these regions. The BJT is biased to operate in the forward active region for most small signal applications. The small signal model developed here will be restricted to this region of operation. In the forward active region the collector and base currents can be obtained from

$$I_C = J_S A e^{(V_{BE}/V_t)} \left(1 + \frac{V_{CE}}{V_{AF}}\right) \tag{3.3-27}$$

$$I_B = \frac{J_S A}{\beta_F} e^{(V_{BE}/V_t)} \tag{3.3-28}$$

Since the BJT is modeled as a three-terminal device, it remains to find y_{11}, y_{12}, y_{21}, and y_{22}, as defined in Sec. 3.0 to obtain the small signal BJT model. Convention has resulted in the selection of the emitter node as the reference. For notational convenience b and c rather than 1 and 2 will be used to denote the base and collector nodes of the BJT, respectively. The parameter y_{cb}, often termed g_m, is generally the dominant parameter in the model. From (3.0-5) and (3.3-27) it follows that

$$y_{cb} = g_m = \left.\frac{\partial I_C}{\partial V_{BE}}\right|_{\substack{V_{CE}=V_{CEQ} \\ V_{BE}=V_{BEQ}}} = \frac{J_S A}{V_t} e^{(V_{BEQ}/V_t)} \left(1 + \frac{V_{CEQ}}{V_{AF}}\right) \tag{3.3-29}$$

Evaluating (3.3-29) at the Q-point and observing that $V_{CEQ}/V_{AF} << 1$, it follows from (3.3-27) and (3.3-29) that

$$g_m = \frac{I_{CQ}}{V_t} \tag{3.3-30}$$

The balance of the y parameters can be readily obtained from (3.0-5), (3.3-27), and (3.3-28). These parameters are summarized in Table 3.3-4. Also listed in this table is the equivalent h parameter model. The h parameter model has been included since it is also widely used. The small signal model equivalent circuits appear in Fig. 3.3-8.

TABLE 3.3-4
Small signal model parameters for the BJT

y Parameters	*h* Parameters
$y_{cb} = g_m = \dfrac{I_{CQ}}{V_t}$	$h_{fe} = \beta_F$
$y_{be} = g_\pi = \dfrac{g_m}{\beta_F}$	$h_{ie} = \dfrac{\beta_F V_t}{I_{CQ}}$
$y_{bc} = 0$	$h_{re} = 0$
$y_{cc} = g_o = \dfrac{g_m V_t}{V_{AF}} = \dfrac{I_{CQ}}{V_{AF}}$	$h_{oe} = \dfrac{I_{CQ}}{V_{AF}}$

Fundamental parameter relationships

$$r_\pi = 1/g_\pi = h_{ie}$$

$$r_o = 1/g_o = 1/h_{oe}$$

$$\beta_F = g_m r_\pi$$

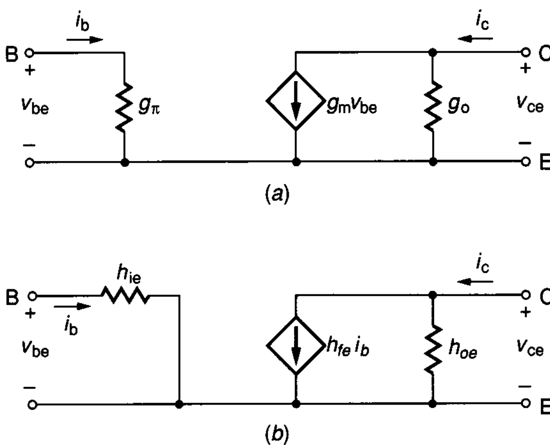


FIGURE 3.3-8
Small signal model of bipolar transistor: (*a*) *y* parameter model (Note: $r_\pi = 1/g_\pi, r_o = 1/g_o$ and $\beta_F = g_m r_\pi$), (*b*) *h* parameter model.

Example 3.3-4. For the circuit shown in Fig. 3.3-9, determine (*a*) the quiescent output voltage, (*b*) the small signal steady state output voltage, and (*c*) the total output voltage. Use the typical process parameters of Table 3.3-1 to characterize the BJT.

Solution

(*a*) To obtain the Q-point it suffices to use the equivalent circuits of Fig. 3.3-5. If it is assumed that the BJT is operating in the forward active region, the circuit of Fig. 3.3-10*a* is obtained. From Ohm's law, it follows that

$$I_{BQ} = (12\text{ V} - 0.6\text{ V})/(600\text{ k}\Omega) = 19\ \mu\text{A}$$

so that

$$I_{CQ} = \beta_F I_{BQ} = 1.9\text{ mA}$$

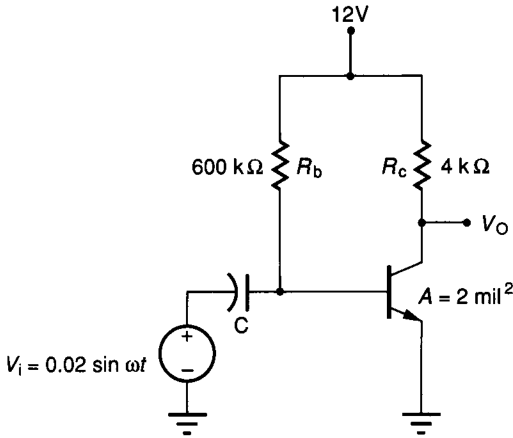


FIGURE 3.3-9
Example 3.3-4.

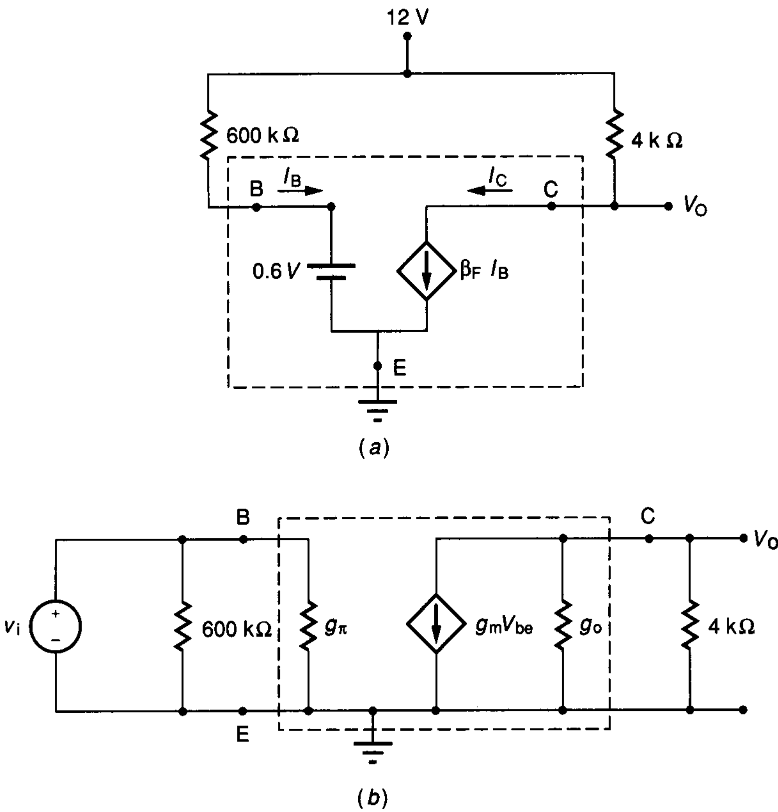


FIGURE 3.3-10
Solution of Example 3.3-4: (*a*) Equivalent circuit for biasing, (*b*) Small signal equivalent circuit.

Hence the quiescent output voltage, V_{OQ}, is given by $V_{OQ} = 12\ \text{V} - I_{CQ}R_c = 4.4$ V. With $V_O = 4.4$ V it follows that the assumption of operation in the forward active region is valid.

(*b*) Assuming C is large enough to act as a short circuit for ac signals, it follows from the model of Fig. 3.3-8 that the circuit can be drawn as in Fig. 3.3-10b. From Table 3.3-4 it follows that the small signal transistor *g* parameters at room temperature ($V_t = 26$ mV) are

$$g_m = 73.1 \text{ mmho}$$

$$g_\pi = 0.731 \text{ mmho}$$

$$g_o = 9.5\ \mu\text{mho}$$

From Ohm's law

$$v_o = (-g_m v_i)\left(\frac{1}{g_o + 1/(4\text{k}\Omega)}\right)$$

$$v_o/v_i = -282$$

The small signal steady state output voltage thus becomes

$$v_o(t) = 5.64 \sin \omega t$$

(*c*) The total output voltage is the sum of the quiescent and small signal steady state values; hence

$$V_O(t) = 4.4\ \text{V} - (5.64 \sin \omega t)\ \text{V}$$

The minus sign appears in the sum to maintain the correct phase relationship between V_O and I_I in this example.

3.3.3 High-Frequency BJT Model

At high frequencies both the dc and small signal models of the BJT introduced in the previous sections are generally considered inadequate. These limitations are primarily attributable to the unavoidable parasitic capacitances inherent in existing bipolar integrated circuits.

These capacitances are of two types. The first type is composed of pn junction capacitances. These capacitances, which are voltage dependent, can be modeled as in the MOSFET and diode cases by (3.1-14) and (3.2-8) provided the junction is not forward biased by any more than $\phi_B/2$ or about 0.35 V. Since the BJT is typically operated with at least one junction forward biased, it is necessary to model the parasitic junction capacitances under forward bias.

For forward biases greater than $\phi_B/2$, it follows from (3.2-9) that the charge storage capacitance can be approximated by[12]

$$C_j = 2^n C_{j0} A \left[2n \frac{V_F}{\phi_B} + (1 - n) \right] \tag{3.3-31}$$

where V_F is the forward bias on the junction. This linear approximation is a continuous extension of (3.1-14) across $V_F = \phi_B/2$ that also agrees in slope at the transition point (see Problem 3.31). The model of the BJT junction capacitances is identical to that of the pn junction diode given in (3.2-8) and (3.2-9).

The cross sections of a vertical npn and a lateral pnp transistor are shown in Fig. 3.3-11. The parasitic junction capacitors for all junctions are identified. Note that the vertical structure has parasitic base–collector, base–emitter, and collector–substrate capacitances, whereas the lateral structure has base–emitter, base–collector, and base–substrate parasitics.

The second type of capacitance is due to majority carrier charge accumulation in the base region, which occurs under forward bias near the emitter junction and introduces a second parasitic between the base and emitter. This capacitance, which is also voltage dependent, can be modeled by

$$C_{AC} = \frac{q t_f I_{CQ}}{kT} = t_f g_m \tag{3.3-32}$$

where t_f is the forward base transit time and I_{CQ} is the quiescent collector current. The process parameter t_f is related to the base width, W_B, and the electron diffusion constant, D_n, by the expression

$$t_f = \frac{W_B^2}{2D_n} \tag{3.3-33}$$

The capacitance C_{AC} is added to the base–emitter junction capacitance to obtain the total base–emitter parasitic capacitance. Typical values of t_f range from 0.1 to 1 ns for vertical npn transistors and from 20 to 40 ns for lateral pnp transistors.[13] The parasitic capacitors for both the vertical and lateral BJT are summarized in Table 3.3-5.

A high-frequency small signal equivalent circuit for the BJT is shown in Fig. 3.3-12. In addition to the parasitic capacitors discussed above, four resistive parasitics are shown. These resistive parasitics are present at all frequencies. The resistance R_C, R_E, and R_B represent the ohmic resistance between the metal contact and the junctions in the transistor. Layouts that minimize the distance between the contacts and junctions minimize these resistors. Note that a "primed"

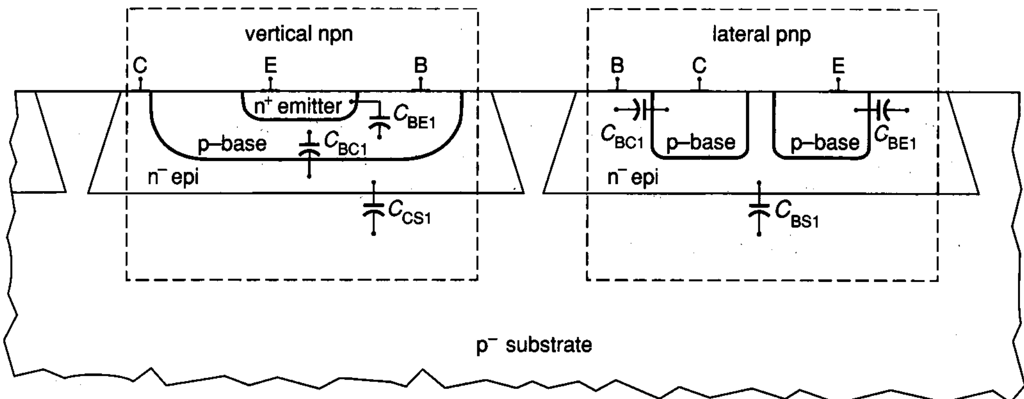


FIGURE 3.3-11
Parasitic junction capacitors in vertical npn and lateral pnp transistors.

TABLE 3.3-5
Parasitic capacitance for the BJT

	Transistor type		
	Alternate notation	**Lateral**	**Vertical**
C_{BE}	C_π	$C_{BE1} + C_{AC}$	$C_{BE1} + C_{AC}$
C_{BC}	C_μ	C_{BC1}	C_{BC1}
C_{BS}		C_{BS1}	0
C_{CS}		0	C_{CS1}

notation has been used to denote the E, B, and C terminals of the effective BJT. The distinction between the "primed" terminals and the actual terminals is necessary since the series ohmic resistances in the emitter, base, and collector leads were neglected when the model of the BJT was derived.

R_E is in the 1 Ω range and is essentially voltage independent. This low value is due to the shallow depth and high doping density of the emitter region. Although the value is small, the reader must be cautioned that even small resistors in the emitter circuit can significantly affect circuit performance (see Problem 3.44). The resistors R_C and R_B are typically in the 100 Ω range. They are both somewhat voltage dependent.

The resistor r_μ, which appears from the base to collector, is large in value and can be approximated by[13]

$$r_\mu \simeq 10\beta_F/g_o \tag{3.3-34}$$

In many applications r_μ can be ignored.

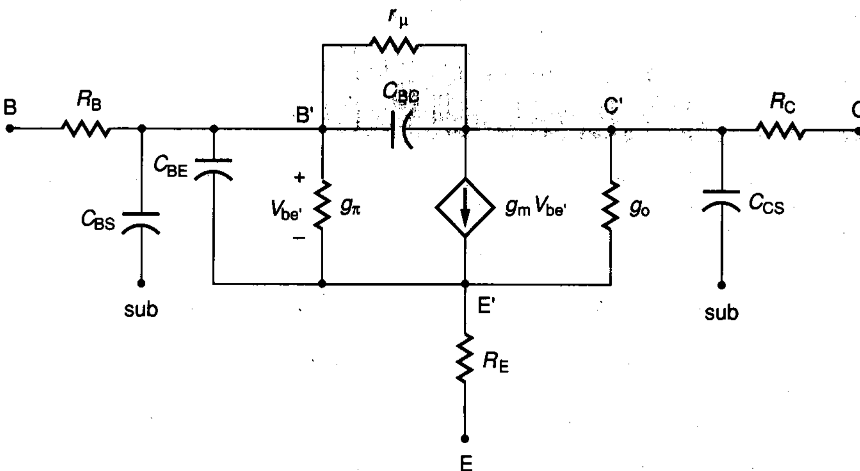


FIGURE 3.3-12
High-frequency small signal equivalent circuit of BJT.

3.3.4 Measurement of BJT Model Parameters

As stated in Section 3.1.4, experimentally measured process parameters are of considerable use to the circuit designer. As in the MOS case, automated measurement equipment is available for bipolar process characterization. In the absence of such equipment, several of the more important BJT device parameters can be measured with reasonable accuracy in the laboratory using standard laboratory equipment. Techniques for measuring J_S, β_F, β_R, V_{AF}, and V_{AR} will be discussed in this section.

From Table 3.3-2, the expressions for the collector and base currents of the npn transistor in the forward active region are

$$I_C = J_S A e^{(V_{BE}/V_t)}\left(1 + \frac{V_{CE}}{V_{AF}}\right) \tag{3.3-35}$$

and

$$I_B = \frac{J_S}{\beta_F} A \exp\left(\frac{V_{BE}}{V_t}\right) \tag{3.3-36}$$

These expressions will serve as a guide for measuring J_S, β_F, and V_{AF}. Measurements of the parameters should be made with I_C close to the intended operating point if possible.

Measurement of V_{AF}. Equation 3.3-35 (extended by Eq. 3.3-1 for small V_{CE}) is plotted in Fig. 3.3-13 for several different values of V_{BE}. Note that all extended curves intersect the V_{CE} axis at $V_{CE} = -V_{AF}$. The parameter V_{AF} can thus be determined by fixing V_{BE} and measuring the V_{CE} axis intercept. If I_{C1} and I_{C2} are the collector currents corresponding to the two different collector–emitter voltages V_{CE1} and V_{CE2} respectively, as depicted in Fig. 3.3-13, then

$$V_{AF} = \frac{I_{C2}V_{CE1} - I_{C1}V_{CE2}}{I_{C1} - I_{C2}} \tag{3.3-37}$$

The circuit of Fig. 3.3-14 with S1 closed can be used for making this measurement. Initially set V_{CE} at V_{CE2} (10 V would be reasonable) and adjust $V_1 = V_{BE}$ so that the desired collector current, I_{C2}, flows ($I_C = -V_O/R$). Leav-

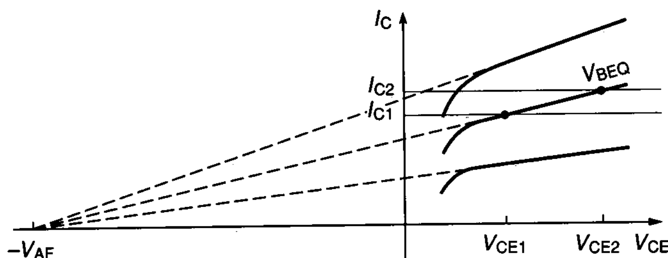
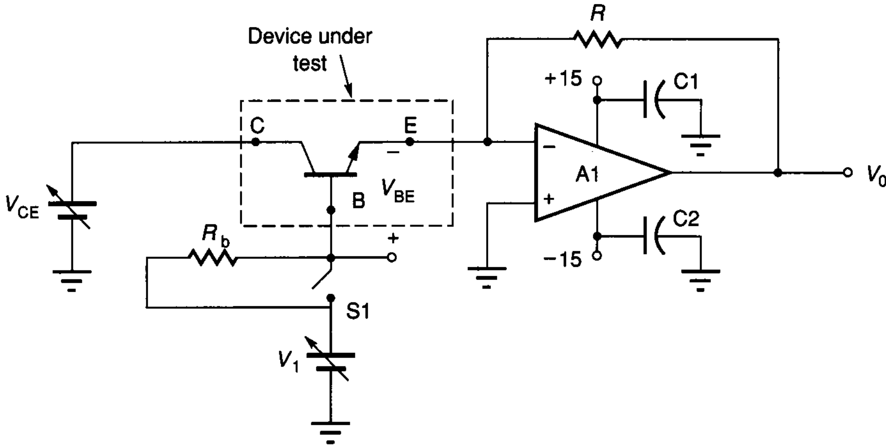


FIGURE 3.3-13
Projection of $I-V$ characteristics of BJT into second quadrant.



A 356 op amp is a reasonable choice for A1. C1 and C2, which may be needed for stability, should be as close as possible to the supply terminals of the op amp to prevent oscillation. A value of 1μF for these capacitors is reasonable. Note: Care should be taken to avoid ever making $V_1 > V_{CE}$ with S1 closed, as this will permanently damage the transistor since no mechanism is provided to limit current flow when the base-collector junction is forward biased.

FIGURE 3.3-14
Circuit for measuring bipolar transistor parameters.

ing V_{BE} fixed ($V_{BE} = V_{BEQ}$), reduce V_{CE} (2 V would be reasonable) to V_{CE1} and measure the current I_{C1}. V_{AF} now follows from (3.3-37). Reasonably wide separation between V_{CE1} and V_{CE2} should be maintained to reduce the sensitivity to the measured voltages and currents.

Measurement of J_S. The previous circuit with S1 closed can be used for measuring $I_S = J_S A$. By measuring the temperature in °K, V_t can be calculated from Eq. 3.3-3. If I_{C2} and V_{CE2} are the initial collector current and collector–emitter voltage used in the measurement of V_{AF}, then it follows from (3.3-35) that it remains to accurately measure $V_{BE} = V_1$ to obtain I_S from the expression

$$I_S = \frac{I_{C2}\exp(-V_{BE}/V_t)}{1 + V_{CE2}/V_{AF}} \tag{3.3-38}$$

Once I_S is obtained, J_S can be determined by dividing by the emitter area, A.

For measuring J_S little accuracy would be lost by either neglecting V_{AF} or using a typical value of V_{AF} in (3.3-38). Accurate measurement of V_{BE} is, however, required.

Measurement of β_F. The circuit of Fig. 3.3-14 with S1 open can be used to measure β_F. Adjust V_1 to obtain the desired collector current, $I_C = -V_O/R$, and adjust V_{CE} to obtain the desired collector–emitter voltage. Measure $I_B = (V_1 - V_{BE})/R_b$. From (3.3-34) and (3.3-35) it thus follows that

$$\beta_F = \frac{I_C/I_B}{1 + V_{CE}/V_{AF}} \tag{3.3-39}$$

As in the case of the measurement of J_S, little accuracy is lost if V_{AF} is neglected or replaced by its nominal value. Since β_F is quite current dependent, it is *particularly important* that it be measured at a value close to the intended operating point.

Methods of measuring β_R and V_{AR} parallel those discussed for β_F and V_{AF} and are left to the reader, as are modifications for measuring parameters of pnp transistors. Information about the parasitic capacitors is generally required for high-frequency applications. Because of their small size, direct measurement with an acceptable degree of accuracy using standard laboratory equipment is not possible. Special test circuits, which either have exceptionally large test devices or which can be used to measure these parasitics indirectly through an investigation of the frequency response of the circuit, are often used.

3.4 PASSIVE COMPONENT MODELS

Discrete passive components are quite easy to model. Resistors and capacitors can generally be modeled by ideal resistors and capacitors respectively. The major limitations are manufacturing tolerances and temperature deviations—both of which can be reduced to acceptable levels in most applications through judicious component selection/specification.

Monolithic resistors and capacitors are far from ideal. They are typically both temperature and voltage dependent. The practical range of values is seriously limited by area constraints. Large resistor or capacitor values are impractical. Process deviations preclude accurate control of absolute component values. Parasitic effects are often quite significant. Relative accuracy (ratioing) between passive components is, however, often quite good.

The following figures of merit are used to characterize passive components.

Resistors

1. Sheet resistance
2. Resistance density
3. Temperature coefficient of resistance
4. Voltage coefficient of resistance
5. Absolute accuracy
6. Relative (ratio) accuracy

Capacitors

1. Capacitance density
2. Temperature coefficient of capacitance
3. Voltage coefficient of capacitance
4. Absolute accuracy
5. Relative (ratio) accuracy

The sheet resistance, temperature coefficients, and voltage coefficients were discussed in Chapter 2. The resistance density is generally a function of the process parameters and layout design rules. The capacitance density is a process parameter. The absolute accuracy is a measure of how accurately the actual resistor and capacitor values can be controlled during processing.

The relative (ratio) accuracy is a measure of how closely two resistors or capacitors can be matched. The ratio accuracy is affected by component placement on a die, device geometry, the physical size of the components, and the nominal relative values of the components themselves.

The parasitic effects were not listed above in the figures of merit primarily because of the difficulty in obtaining a single figure with which to meaningfully quantify the parasitic effects. Nevertheless, parasitic effects include contact resistance and distributed capacitances for resistors. Contact resistance, overlap stray capacitances, and edge capacitances are considered as parasitic effects for capacitors.

3.4.1 Monolithic Capacitors

Any structure in which a voltage-induced separation of charge occurs can serve as a capacitor. Some of the structures used for capacitors, along with their characteristics[14] are shown in Table 3.4-1. In a MOS process, the most common capacitors are formed by sandwiching a thin oxide layer between two conductive polysilicon layers. These capacitors are nearly independent of applied voltage and can be modeled as ideal capacitors. The major limitation is the large parasitic capacitor that is always formed between the lower plate and the substrate. This parasitic limits how these capacitors can be used although clever design techniques, such as stray-insensitive SC filters,[15–16] often evolve to minimize these effects.

When the luxury of the double polysilicon layers is not present in MOS processes, metal–poly, metal–diffusion (lower plate formed by a diffused region in the substrate), or poly–diffusion capacitors with an SiO_2 dielectric are used. These capacitors typically have a lower capacitance density and/or increased voltage dependence and/or a less conductive lower plate than the double poly capacitors.

In the bipolar process, the most desirable common capacitors are metal–diffusion capacitors with an SiO_2 dielectric. The heavily doped emitter diffusion is used for the lower diffusion plate. The characteristics are quite good, but an additional mask step is required for forming the dielectric region. Alternatives include the voltage-dependent junction capacitances formed by either the B–C or B–E junction. These capacitors are modeled by (3.1-14) and (3.3-31). The B–E junction offers reasonable capacitance density at the expense of a limited reverse breakdown voltage (typically 5 to 7 V). The B–C junction reverse breakdown voltage is quite high (typically in the 30 V range), but the capacitance density is quite low. All junction capacitors are limited by requirements that the junctions remain reverse biased (actually, not forward biased by more than a few tenths of a volt).

TABLE 3.4-1
Characteristics of monolithic capacitors

Capacitor type	Process	Dielectric	Absolute accuracy	Ratio accuracy	Voltage characteristic	Temperature	Comments
Poly–Poly	MOS	SiO_2	±20%	±0.06%	−5 ppm/V	25 ppm/°C	Most popular MOS, best characteristic
Poly–Diffusion	MOS	SiO_2	±10%	±0.06%	−20 ppm/V	25 ppm/°C	Lower plate potential often fixed
Metal–Diffusion (with thin oxide)	MOS/Bipolar	SiO_2	±10%	±0.06%	−20 ppm/V	25 ppm/°C	Most desirable bipolar
Moat–Substrate	MOS	Si					Voltage dependent
Base–Collector	Bipolar	Si					Must be reverse biased, low density
Base–Emitter	Bipolar	Si					Must be reverse biased
Conductor–Conductor	Thin film	Varies					Good characteristics
Screened	Thick film	Varies					Chip capacitors often preferred

3.4.2 Monolithic Resistors

Considerably more options exist for monolithic resistors. Some monolithic resistors are passive devices and others contain active devices. Major tradeoffs must be made between linearity, area, biasing complexity, and temperature characteristics in monolithic resistors. Table 3.4-2 lists the characteristics of some of the structures that are used for resistors.

In standard MOS processes, the most ideal resistors are merely strips of polysilicon. Diffusion strips are also used for resistors but exhibit an undesirable nonlinear relationship between voltage and current. Ion implants offer some advantages over depositions for the introduction of impurities to control absolute resistance values in diffused resistors. Thin film resistors with excellent characteristics are added in some specialized processes. For each of these types of resistors, a serpentine pattern is often used to improve packing density. The major limitations of these resistors are the low resistance densities, which limit the total resistance to quite small values; the high deviations in resistance due to process variations; and large temperature coefficients.

In bipolar processes epitaxial strips or diffusion strips are commonly used for resistors. These devices are quite linear. The base diffusion is often used because of its reasonably high sheet resistance. A base-diffused resistor is shown in Fig. 3.4-1a. To prevent forward biasing of the "base–collector" junction, a contact is needed to the epitaxial layer. This will be typically connected to the most positive power supply voltage used for the circuit.

It can be argued that the resistance of the base-diffused resistor could be increased if the depth of the p-base diffusion could be decreased. The depth of this diffusion, however, is generally determined to optimize performance of the BJTs themselves. An alternative is to place an n^+ emitter diffusion in the p base region. This masking step already exists and will result in a significant increase of the sheet resistance of the underlying p diffusion. Such a device, which is termed a *pinch* resistor, is shown in Fig. 3.4-1*b*. Contact must be made to both the n^+ emitter diffusion and the n^- epitaxial region. These regions are typically both connected to the most positive power supply voltage used for the circuit. Although the resistance increases significantly due to this pinching, the variance in emitter and base diffusion depths due to process variations makes the tolerances of pinch resistors quite wide. They also exhibit an increased voltage dependent nonlinearity and are limited in voltage range to circumvent breakdown of the reverse-biased base–emitter junction. Other types of pinch resistors (e.g., epitaxial pinch) can also be made.

Several *active resistors* are shown in Fig. 3.4-2. These active resistors often offer considerable reductions in area requirements compared to passive resistors at the expense of increased nonlinearity and/or reduced signal swing and/or complicated biasing requirements. A more detailed discussion of the active resistor structures appears in Chapter 5.

The circuit of Fig 3.4-2*a* is merely a MOSFET biased to operate in the ohmic region. From (3) of Table 3.1-1, the relationship between I_D and V_{GS} is

$$I_D = \frac{K'W}{L}\left(V_{GS} - V_T - \frac{V_{DS}}{2}\right) V_{DS} \tag{3.4-1}$$

TABLE 3.4-2
Characteristics of MOS Resistors

Device	Characterizing equation	Ideal resistance†
Poly Strip	$V = I R_{\square} \frac{L}{W}$	$R = R_{\square} \frac{L}{W}$
Diffusion	$V = I R_{\square} \frac{L}{W}$	$R = R_{\square} \frac{L}{W}$
MOSFET (Ohmic region, Fig. 3.4-2a)	$I = \frac{K'W}{L}[(V_{GS} - V_T) - \frac{V}{2}]V$	$R = \frac{L}{K'W(V_{GS} - V_T)}$
MOSFET (Saturated region, Fig. 3.4-2b)	$I = \frac{K'W}{2L}(V_{GS} - V_T)^2(1 + \lambda V)$	$R_{SS} = \frac{2L}{\lambda K'W(V_{GS} - V_T)^2}$
MOS depletion (Fig. 3.4-2c)	$I = \begin{cases} \frac{K'W}{L}(-V_T)V & V < -V_T \\ \frac{K'W}{2L}V_T^2(1 + \lambda V) & V > V_T \end{cases}$	$R = \frac{L}{K'W \mid V_T \mid}$ $R_{SS} = \frac{2L}{K'W V_T^2 \lambda}$
MOS Enhancement (Fig. 3.4-2d)	$I = \frac{K'W}{2L}(V - V_T)^2(1 + \lambda V)$	$R_{SS} = \frac{L}{K'W(V_Q - V_T)}$
Linearity Compensated[18] (Fig. 3.4-2e)	$I = \begin{cases} \frac{2K'W}{L}(-V_T)V & V < -V_T \\ \frac{K'W}{L}V_T^2(1 + \lambda V) & V > -V_T \end{cases}$	$R = \frac{L}{2K'W \mid V_T \mid}$ $R_{SS} = \frac{L}{K'W V_T^2 \lambda}$
Bootstrapped Budak[17] (Fig. 3.4-2f)	$I = \frac{K'W}{L}(V_C - V_T - V\frac{[\theta + 1]}{2})(1 - \theta)V$ $V < V_C - V_T$	$R = \frac{L}{K'W(1 - \theta)}$

If $V_{DS} << (V_{GS} - V_T)$, then, as seen in Eq. 3.4-1, the relationship between I_D and V_{DS} is linear (assuming V_{GS} is independent of V_{DS}), resulting in an equivalent resistance of

$$R_{eq} \simeq \frac{L}{K'W(V_{GS} - V_T)} \tag{3.4-2}$$

This approximation is quite good for $V_{DS} < 0.5(V_{GS} - V_T)$. A model that includes the nonlinearity is

$$R_{eq} \simeq \frac{L}{K'W[(V_{GS} - V_T) - V_{DS}/2]} \tag{3.4-3}$$

The resistance density (resistance per unit area) of the MOS resistor will now be calculated. Assume d_1 is the minimum moat width, d_2 the minimum

Resistance density‡	Temperature characteristic	Absolute accuracy	Relative accuracy	Comments
$\dfrac{R_{\square}}{d_4(d_s + d_4)}$	1500 ppm/°C	±30%	±2%	Linear, low resistance density
$\dfrac{R_{\square}}{d_3(d_s + d_4)}$	1500 ppm/°C	±35%	±2%	Somewhat voltage dependent
$\dfrac{1}{K'(V_{GS} - V_T)d_1(d_1 + 2d_2 + d_3)}$				To minimize distortion, $V < (V_{GS} - V_T)/2$
$\dfrac{1}{\lambda \frac{K'}{2}(V_{GS} - V_T)^2 d_1(d_1 + 2d_2 + d_3)}$				Small signal impedance only, high impedance values
$\dfrac{1}{K' \mid V_T \mid d_1(d_1 + 2d_2 + d_3)}$				Popular load device
$\dfrac{1}{\lambda \frac{K'}{2} V_T^2 d_1(d_1 + 2d_2 + d_3)}$				Quite nonlinear
$\dfrac{1}{K'(V_Q - V_T)d_1(d_1 + 2d_2 + d_3)}$				Quite nonlinear, good resistance density
$\dfrac{1}{2K' \mid V_T \mid d_1(d_1 + 2d_2 + d_3)}$				Major improvement in density
$\dfrac{1}{\lambda K' V_T^2 d_1(d_1 + 2d_2 + d_3)}$				
Depends upon how θV is realized				Good linearity potential

†R denotes large signal impedance; R_{SS} denotes small signal impedance only.

‡Assuming a minimum size layout, see Fig. 3.4-3 d_1, d_2, d_3, and d_4 as defined in Sec. 3.4.2.

required overlap of poly over moat, and d_3 the minimum poly–poly spacing as defined by the design rules of the process. If a large serpentine MOS device is constructed with minimum moat width and with separate polysilicon strips for each diffusion strip in the serpentine and if the incremental area required by the corners is neglected, then the diagram of Fig. 3.4-3 can be used to calculate the resistance density of this device. The section of length h has a resistance of

$$R = \frac{h}{K'd_1(V_{\text{GS}} - V_{\text{T}})} \tag{3.4-4}$$

and an area of

$$A = h(d_3 + d_1 + 2d_2) \tag{3.4-5}$$

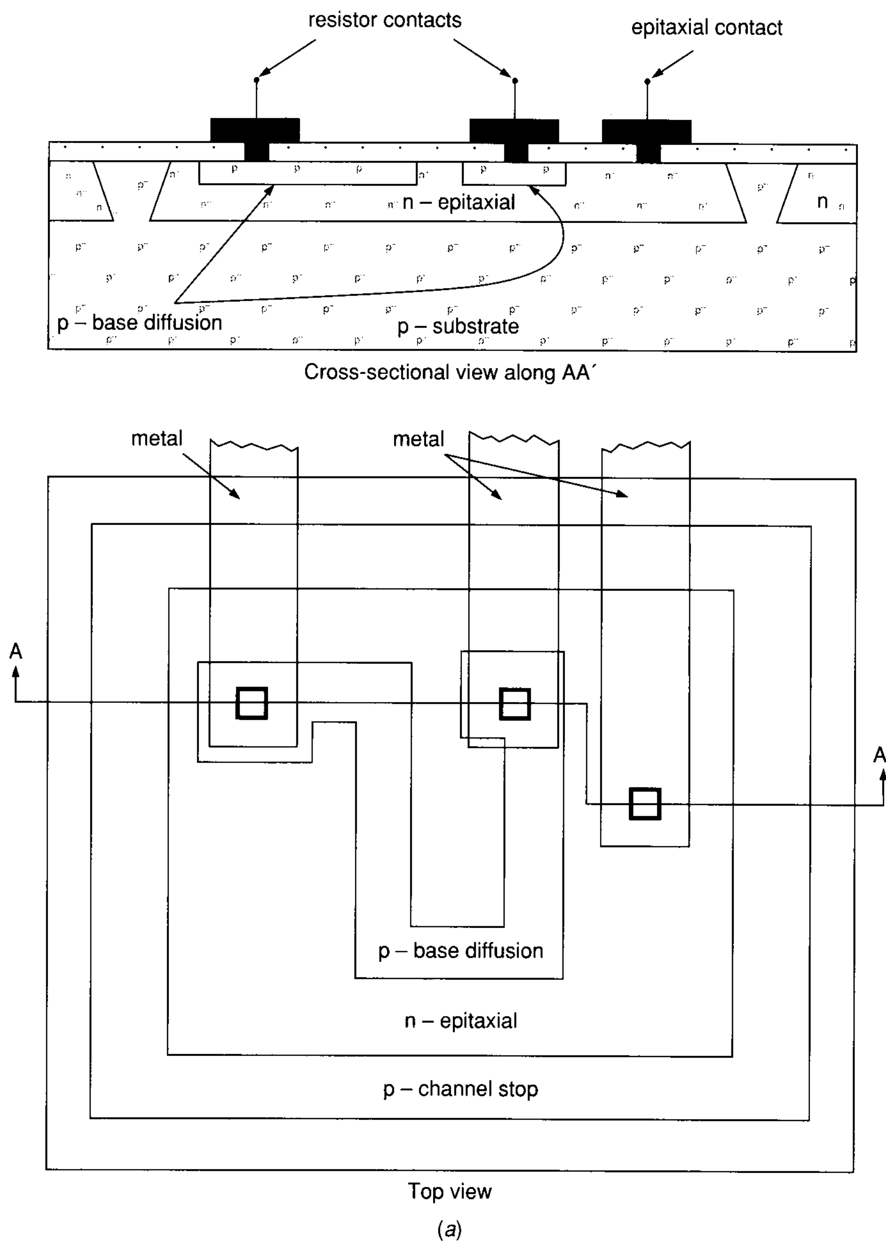


FIGURE 3.4-1
Monolithic resistors: (*a*) Base diffused, (*b*) Base-pinch.

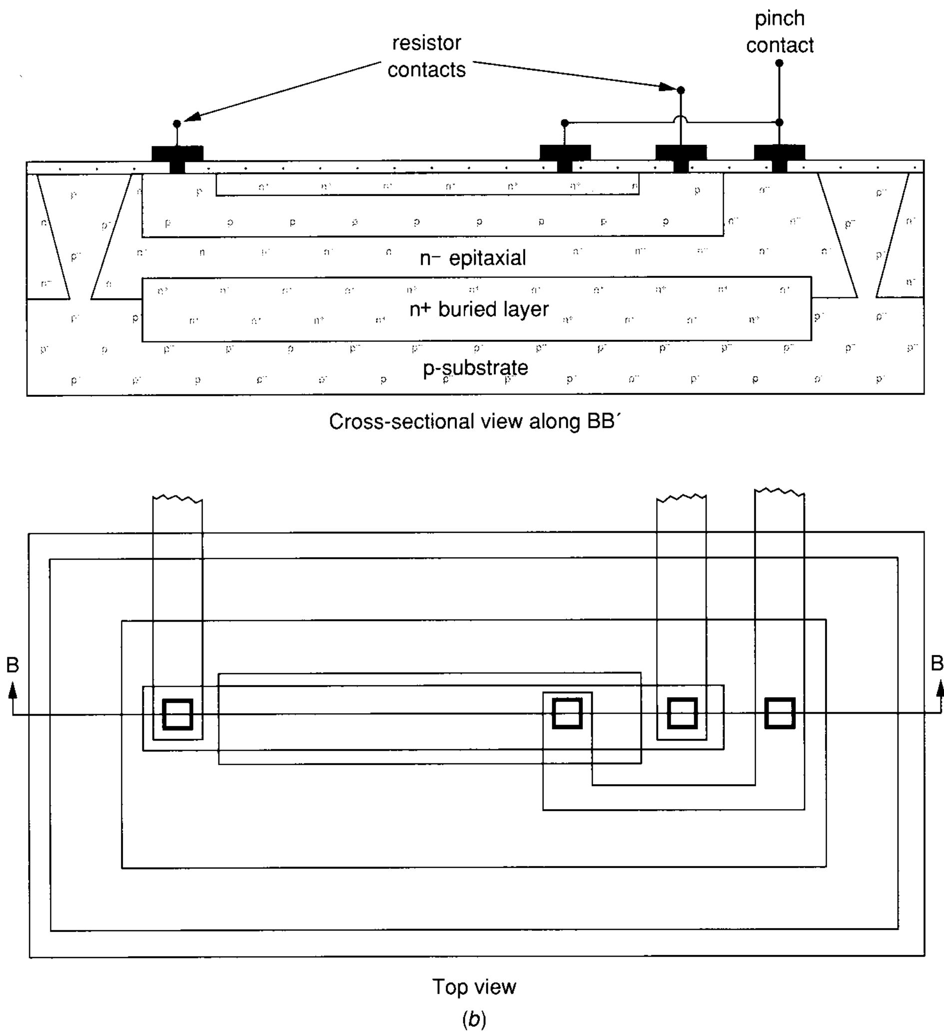


FIGURE 3.4-1
(*continued*)

so a resistance density of

$$R_d = \frac{1}{K'(V_{GS} - V_T)(d_1 + 2d_2 + d_3)d_1} \tag{3.4-6}$$

Although making V_{GS} close to V_T can make R_d very high, it is generally impractical due to process variations to directly make $V_{GS} - V_T < 1$ V. Some improvements in resistance density can be realized if the polysilicon strips are merged into one large rectangular region. In this case, the minimum diffusion spacings will determine how close the serpentined strips can be placed. If the MOSFET is a depletion device, it is particularly convenient to make $V_{GS} = 0$ as shown in Fig. 3.4-2*c*, thus eliminating the need for a voltage source.

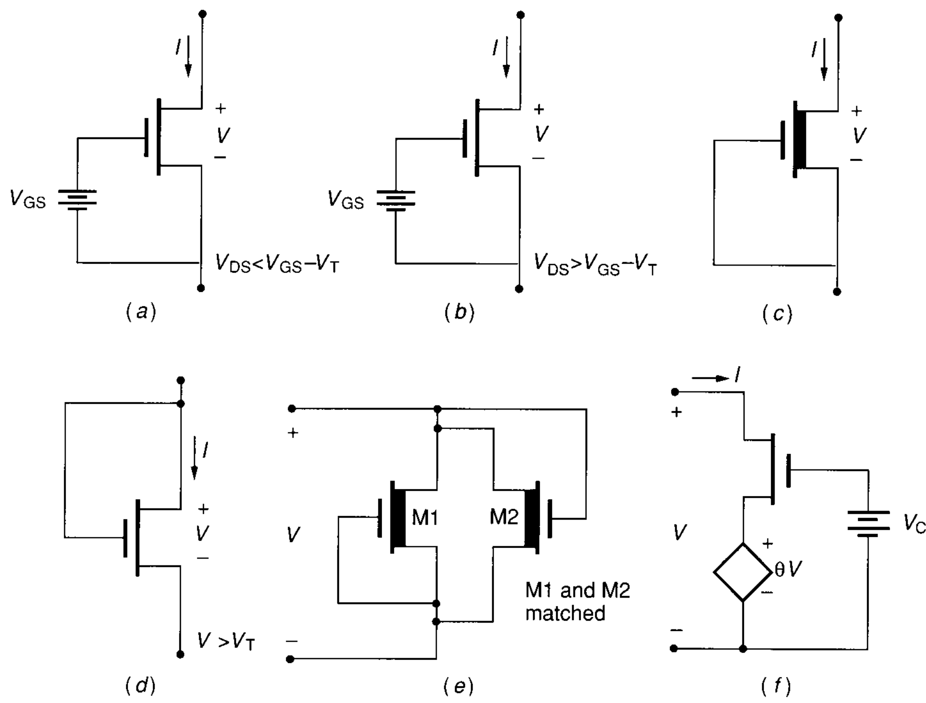


FIGURE 3.4-2
Active MOS resistors.

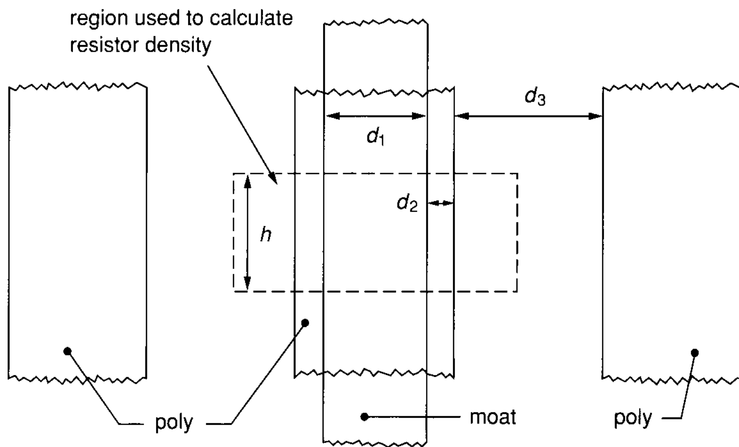


FIGURE 3.4-3
Calculation of resistance density of MOS resistor in ohmic region.

Example 3.4-1. Compare the area required to make a 1 MΩ resistor using a minimum size POLY I string with that using a MOS depletion resistor operating in the saturation region with $V_{GS} = 0$. Use the design rules and process parameters of the NMOS process discussed in Appendix 2A of Chapter 2. Assume the POLY I is uncovered and the feature size of the process is 3 μ (i.e., $\lambda = 1.5\ \mu$).

Solution. The resistance density of the POLY I string can be readily obtained (see Problem 3.55) and is given by

$$R_{dp} = \frac{R_{\square}}{d_4(d_3 + d_4)} \tag{3.4-7}$$

where d_3 is the minimum poly spacing, d_4 is the minimum poly width, and $R_{\square}$ is the sheet resistance of POLY I. From Tables 2A.2 and 2A.4 of Appendix A of Chapter 2, $d_3 = 3\ \mu, d_4 = 3\ \mu$, and $R_{\square} = 25\ \Omega/\square$. Substituting into (3.4-7), it follows that $R_{dp} = 1.39\ \Omega/\mu^2$.

If the MOSFET shown in Fig. 3.4-2*c* is biased to operate in the saturation region, then, from (4) of Table 3.1-1, the drain current is given by

$$I_D = \frac{K'W}{2L}(V_{GS} - V_T)^2(1 + \lambda V_{DS})$$

The small signal impedance can be obtained by differentiation (assuming V_{GS} and V_{BS} constant) with respect to V_{DS} to obtain

$$R_{ss} = \frac{2L}{[\lambda K'W(V_{GS} - V_T)^2]} \tag{3.4-8}$$

Defining d_1–d_3 as in Fig. 3.4-3, it follows that the resistance density for a large serpentined structure is

$$R_{ds} = \frac{2}{[\lambda K'd_1(V_{GS} - V_T)^2(d_1 + 2d_2 + d_3)]} \tag{3.4-9}$$

From Tables 2A.2 and 2A.4 of Chapter 2, $\lambda = 0.01\ \mathrm{V}^{-1}$, $K' = 25\ \mu\mathrm{A/V}^2$, $V_T = -3$ V, $d_1 = 3\ \mu$, $d_2 = 3\ \mu$, and $d_3 = 3\ \mu$. Substituting into (3.4-9) and setting $V_{GS} = 0$, we obtain $R_{ds} = 74\ \mathrm{k}\Omega/\mu^2$. It follows that the area ratio is the ratio of the resistance densities; that is, the ratio of R_{dp} to R_{ds}, which is about 67,000 to 1. The area required to make the 1 MΩ poly resistor is about $9 \times 10^5\ \mu^2$, whereas the active resistor requires an area of around 13 μ^2.

Note that for reasonable values of $V_{GS} - V_T$, the λ in the denominator of (3.4-9) makes the resistance density of the MOS resistor in the saturation region considerably higher than the resistance density of the same device biased to operate in the ohmic region. Small signal resistances in the 1 MΩ range become practical with the MOS resistor biased to operate in the saturation region.

Several other MOS active resistors are shown in Fig. 3.4-2. Those of Fig. 3.4-2*c* and *d* have proven the most popular in NMOS applications due to their simplicity. The structures of Fig. 3.4-2*e* and *f* offer improvements in linearity at the expense of increased area and/or complexity. The characteristics of these active loads are summarized in Table 3.4-2.

A popular bipolar active resistor is shown in Fig. 3.4-4. The equivalent resistance and resistance density will now be calculated.

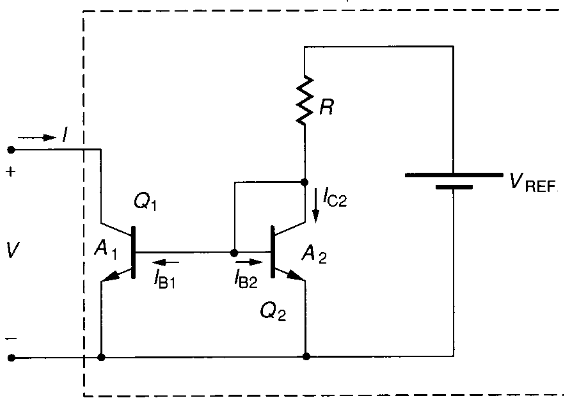


FIGURE 3.4-4
Active bipolar resistor.

Assuming Q_2 is operating in the active region and that β_F is large enough so that $(I_{B1} + I_{B2}) << I_{C2}$, it follows from (2) of Table 3.3-3 that

$$I_{C2} \simeq (V_{REF} - 0.6\text{ V})/R \tag{3.4-10}$$

and from (1) of Table 3.3-2 (neglecting V_{AF}, which is justifiable since $V_{CE2} \simeq 0.6\text{ V} << V_{AF}$),

$$I_{C2} \simeq J_S A_2 \exp\left(\frac{V_{BE2}}{V_t}\right) \tag{3.4-11}$$

Equating the two expressions for I_{C2}, it follows that

$$\exp\left(\frac{V_{BE2}}{V_t}\right) \simeq \frac{V_{REF} - 0.6\text{ V}}{RJ_S A_2} \tag{3.4-12}$$

Again, from (1) of Table 3.3-2, it follows that

$$I = J_S A_1\left(1 + \frac{V}{V_{AF}}\right)\exp\left(\frac{V_{BE1}}{V_t}\right) \tag{3.4-13}$$

Since $V_{BE1} = V_{BE2}$ it follows from (3.4-12) and (3.4-13) that

$$I = \frac{A_1\left(1 + \frac{V}{V_{AF}}\right)(V_{REF} - 0.6\text{ V})}{A_2 R} \tag{3.4-14}$$

The small signal impedance follows from differentiation with respect to V and is given by

$$R_{eq} = \frac{V_{AF} A_2 R}{A_1(V_{REF} - 0.6\text{ V})} \tag{3.4-15}$$

It can be shown that for large R, the area of R dominates that of Q_1 and Q_2; hence the resistance density is approximately m times that of R where

$$m = \left(\frac{V_{AF}}{V_{REF}}\right)\left(\frac{A_2}{A_1}\right) \tag{3.4-16}$$

It can be seen that this active resistor can easily result in an area savings of from 10 to 1000 over that required for the resistor R. The high output impedance of the circuit of Fig. 3.4-4 will be exploited in the discussion of current mirrors in Chapter 5.

Both active and passive resistors are used in IC designs. When area required for passive resistors becomes excessive, the designer should consider the active resistor alternative but must bear in mind that the active resistors are typically nonlinear and often require special biasing considerations.

3.5 SUMMARY

In this chapter, models for MOSFETs, diodes, BJTs, and passive components have been developed. Included are large signal dc models, small signal models, and high-frequency models of varying degrees of complexity. Small signal models were obtained by differentiation of the large signal dc models, and high-frequency models were obtained by identifying the relevant parasitic capacitances inherent in the processes in which the devices were fabricated. This multiple-model approach allows the user to select the simplest model acceptable for a given application. The most notable differences between the MOSFET and the BJT are (1) square law contrasted to exponential dc transfer characteristics, and (2) different input impedances, characterized by a nearly infinite input impedance for the MOSFET contrasted to a small input impedance for the BJT. The models developed here will be used for analysis and design throughout later chapters and will be expanded for use in computer simulations in Chapter 4.

REFERENCES

1. C. T. Sah, "Characteristics of the Metal-Oxide-Semiconductor Transistor," *IEEE Trans. Electron. Devices*, vol. ED-11, pp. 324–345, July 1964.
2. H. Shichman and D. A. Hodges, "Modeling and Simulation of Insulated-Gate Field-Effect Transistor Switching Circuits," *IEEE J. Solid State Circuits*, vol. SC-3, pp. 285–289, Sept. 1968.
3. R. Muller and T. Kamins, *Device Electronics for Integrated Circuits*, Wiley, New York, 1977, Chapter 7.
4. J. Millman and C. Halkias, *Integrated Electronics: Analog and Digital Circuits and Systems*, McGraw-Hill, New York, 1972.
5. A. Sedra and K. Smith, *Microelectronic Circuits*, Holt, Rinehart & Winston, New York, 1982.
6. C. Duvurry, "A Guide to Short Channel Effects in MOSFETS," *TI Engineering Journal*, Texas Instruments Inc., pp. 52–56, July–August 1984.
7. A. Vladimirescu and S. Liu, "The Simulation of MOS Integrated Circuits Using SPICE2," Memorandum No. UCB/ERL M80/7, Electronics Research Laboratory, University of California, Berkeley, Feb. 1980.
8. L. Stotts, "Introduction to Implantable Biomedical IC Design," *IEEE Circuits and Devices Magazine*, pp. 12–18 January 1989.
9. Y. Tsividis and P. Antognetti, eds., *Design of MOS VLSI Circuits for Telecommunications*, Prentice-Hall, Englewood Cliffs, New Jersey, 1985, Chapter 4.
10. P. Gray, D. Hodges, and R. Brodersen, eds., *Analog MOS Integrated Circuits*, IEEE Press, New York, 1980, p. 31.
11. J. J. Ebers and J. L. Moll, "Large Signal Behavior of the Junction Transistor," *Proc. IRE*, Vol. 42, pp. 1761–1772, Dec. 1954.

12. I. Getreu, *Modeling the Bipolar Transistor*, Tektronix Inc., Beaverton, Oregon, 1976.
13. P. Gray and R. Meyer, *Analysis and Design of Analog Integrated Circuits*, Wiley, New York, 1977.
14. P. E. Allen and D. R. Holberg, *CMOS Analog Circuit Design*, Holt, Rinehart & Winston, New York, 1987.
15. R. Gregorian and G. C. Temes, *Analog MOS Integrated Circuits for Signal Processing*, Wiley, New York, 1986.
16. P. E. Allen and E. Sánchez-Sinencio, *Switched Capacitor Circuits*, Van Nostrand Reinhold, New York, 1984.
17. K. Nay and A. Budak, "A Voltage-Controlled Resistance with Wide Dynamic Range and Low Distortion," *IEEE Trans. on Circuits and Systems*, vol. CAS-30, pp. 770–772, October 1983.
18. K. Peterson and R. L. Geiger, "CMOS OTA Structures with Improved Linearity," *Proc. 27th Midwest Symposium on Circuits and Systems*, Morgantown, WV, pp. 63–66, June 1984.
19. B. J. Sheu, D. L. Scharfetter, and P. K. Ko, "SPICE2 Implementation of BSIM", UCB/ERL M85/42, Engineering Research Laboratory, University of California, Berkeley, May 1985.

PROBLEMS

Section 3.0

3.1. A linear three-terminal network can be characterized in terms of the y-parameters by the equations

$$i_1 = y_{11}v_1 + y_{12}v_2$$

$$i_2 = y_{21}v_1 + y_{22}v_2$$

Obtain the corresponding small signal equivalent circuit.

3.2. Show that the small signal Thevenin equivalent circuits of both of the circuits in Fig. P3.2 are resistors of value R using the definitions of Section 3.0-3.

3.3. Develop a small signal model for a device characterized by the following equations:

$$I_1 = 5(V_1 - 4)$$

$$I_2 = -25(V_1 - 2)^3\left(1 + \frac{V_2}{5}\right)$$

Section 3.1

3.4. Verify that if a MOSFET is modeled by Eq. (3.1-1), the designation of "source" and "drain" is arbitrary (i.e., the same I–V characteristics are obtained if the node initially labeled "drain" is called the "source" and the node initially labeled "source" is called the "drain").

3.5. Derive an expression for the MOSFET for the locus of points in the I_D–V_{DS} plane where

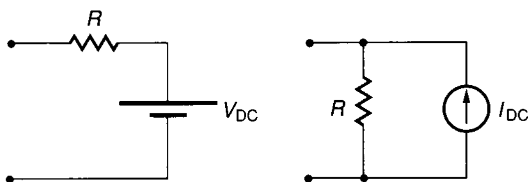


FIGURE P3.2
Circuits for Problem 3.2.

$$\frac{\partial I_D}{\partial V_{DS}} = 0$$

Assume the device is modeled by Eq. 3.1-1.

3.6. Assume a MOS device is to be modeled in the saturation region by the equation

$$I_D = \frac{K'W}{2L}(V_{GS} - V_T)^2 \tag{P1}$$

but that the device is actually characterized by the equation

$$I_D = \frac{K'W}{2L}(V_{GS} - V_T)^2\left(1 + \frac{V_{DS}}{20}\right) \tag{P2}$$

where V_T is assumed constant in both equations. Also assume that the unknown parameters K' and V_T are to be experimentally obtained from (P1) by measuring I_D at $V_{GS} = 2.0$ V and again at $V_{GS} = 2.2$ V with $V_{DS} = 2$ V and then solving simultaneously the two versions of (P1) to obtain the two unknowns. Further assume that $W = 5\ \mu, L = 15\ \mu$, and the measured values of I_D are 6.336 μA and 8.624 μA at 2.0 and 2.2 V respectively.

(*a*) What values will be obtained for the parameters K' and V_T?

(*b*) What percentage error between experimental and theoretical results will be obtained in using Eq. P1 and the parameters obtained in (*a*) to predict I_D if the device is to operate at $V_{GS} = 2.1$ V and $V_{DS} = 2.5$ V?

(*c*) What percentage error between experimental and theoretical results will be obtained in using Eq. P1 and the parameters obtained in (*a*) to predict I_D if the device is to operate at $V_{GS} = 6$ V and $V_{DS} = 15$ V?

(*d*) How can correlation between theoretical and experimental results be improved?

3.7. The transfer characteristics for the CMOS inverter shown in Fig. P3.7 are plotted parametrically in Fig. 3.1-13*b* for the case that $V_{DD} > V_{TON} - V_{TOP}$. Plot parametrically the transfer characteristics of this circuit if $V_{DD} < V_{TON} - V_{TOP}$ and identify the mode of operation of each transistor in each region.

3.8. The small signal model of the MOSFET is used most often for the analysis of MOS circuits in which the transistors are operating in the saturation region.

(*a*) Derive the small signal low frequency model for the MOS transistor operating in the ohmic region with *Q*-point I_{DQ}, V_{GSQ}, V_{DSQ}, and V_{BSQ}.

(*b*) What does this reduce to if $V_{DSQ} = 0$ V?

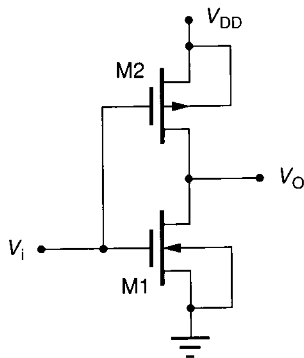


FIGURE P3.7
Circuit for Problem 3.7.

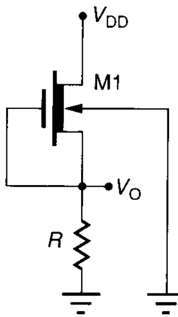


FIGURE P3.9
Circuit for Problem 3.9.

3.9. One easy way to include λ and γ effects in the calculation of V_O for the circuit shown in Fig. P3.9 is to set up an iteration defined for $N = 1$ by

$$V_{T1} = V_{T0}$$

$$I_{D1} = \frac{K'W}{2L}(V_{T1})^2$$

$$V_{O1} = I_{D1}R$$

and for $N > 1$,

$$V_{TN} = V_{T0} + \gamma\left(\sqrt{\phi + V_{O,N-1}} - \sqrt{\phi}\right)$$

$$I_{DN} = \frac{K'W}{2L}(V_{TN})^2[1 + \lambda(V_{DD} - V_{O,N-1})]$$

$$V_{ON} = I_{DN}R$$

How many iterations are required for I_{DN} to converge to within 0.01% of its actual value if M1 is fabricated in the NMOS process of Table 3.1-2 and if $W = 3\,\mu$, $L = 12\,\mu$, $R = 20\,\text{k}\Omega$, and $V_{DD} = 6$ V?

3.10. Obtain an expression for and plot the maximum deviation over the ohmic region, in percentage, from the current as predicted by (3.1-1) from that obtained if (3.1-1) is multiplied by $(1 + \lambda V_{DS})$ to maintain continuity at the ohmic region–saturation region interface.

3.11. Assume the MOSFET in Fig. P3.11 is characterized by the CMOS model parameters of Table 3.1-2.

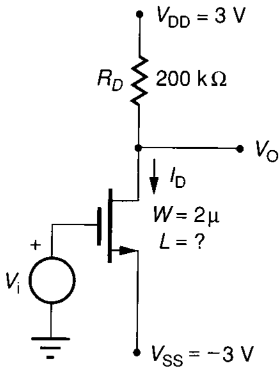


FIGURE P3.11
MOSFET for Problem 3.11.

(*a*) Determine L so that $A_V = v_o/v_i = -25$. What is the Q-point (V_{OQ}, I_{DQ}) with this value of L?

(*b*) Determine L so that $V_{OQ} = OV$. What is $A_V = v_o/v_i$ with this value of L?

3.12. Obtain the dc model for the devices shown in Fig. P3.12. Make a qualitative comparison of the performance of these structures.

3.13. The output characteristics for a MOSFET with $V_{BS} = 0$ are shown in Fig. P3.13.

(*a*) Determine as many parameters as possible from the set

$$\{K', V_{T0}, \gamma, \lambda, W, L, K'W/L, \phi, C_{ox}\}$$

(*b*) If the device is biased to operate at the point Q, determine as many parameters as possible of the set $\{g_m, g_{ds}, g_{mb}\}$.

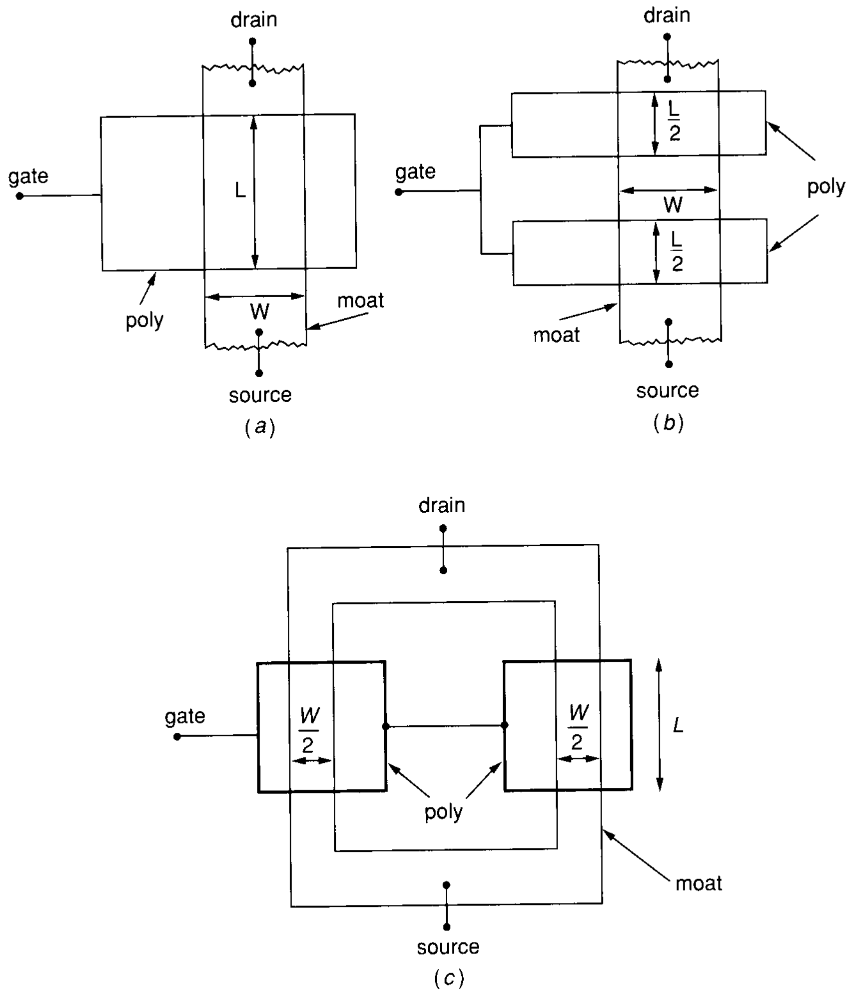


FIGURE P3.12
Devices for Problem 3.12.

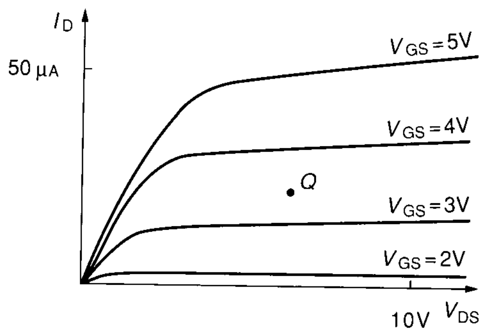


FIGURE P3.13
Graph for Problem 3.13.

3.14. Plot g_m, g_{ds}, and g_{mb} versus I_{DQ} on the same axis for I_{DQ} in the interval [100 nA, 10 mA] for a MOSFET with $W = L = 5\ \mu$ designed in the CMOS process characterized in Table 3.1-2.

3.15. Determine the small signal model of the MOSFET operating in weak inversion saturation. Compare this model to that of the same device if it operates in strong inversion.

3.16. A simple single transistor amplifier is shown in Fig. P3.16.

(*a*) Using the typical CMOS process parameters of Table 3.1-2, determine W/L of M so that $I_{DQ} = I_{BIAS} = 20\ \mu A$, if $V_{SS} = -2$ V and $V_i = 0$ V.

(*b*) Determine the small signal voltage gain $A_V = v_o/v_i$ for the value of W/L obtained in (*a*).

(*c*) If the subthreshold parameters I_{DO} and n of (3.1-32) are $I_{DO} = 20$ nA and $n = 2$, determine W/L of M so that $I_{DQ} = I_{BIAS} = 20\ \mu A$ at room temperature if $V_{SS} = -0.1$ V and $V_i = 0$ V.

(*d*) Determine the small signal voltage gain $A_V = v_o/v_i$ for the value of W/L obtained in (*c*) and compare these results with that obtained for operation in strong inversion.

3.17. Obtain an expression for and plot I versus V for the enhancement load circuit of Fig. P3.17 based upon the MOSFET model of Table 3.1-1, and compare these transfer characteristics with those used for the simple model of Eq. 3.1-45.

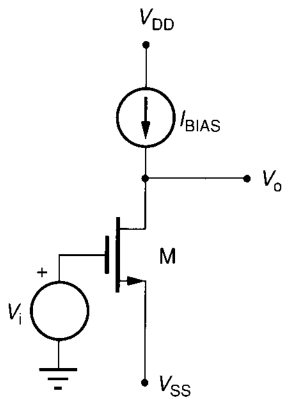


FIGURE P3.16
Problem 3.16.

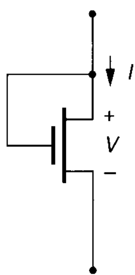


FIGURE P3.17
Circuit for Problem 3.17.

3.18. Derive the small signal model for the MOSFET operating in the saturation region from the dc model of the MOSFET given by Eq. 3.1-5. Verify that these results are equivalent to those given in Table 3.1-3.

3.19. If $I_B = 50\ \mu A$, then M1 in Fig. P3.19 is operating in strong inversion for $V_i = 0$. If I_B remains constant, how much must the W/L ratio of M1 be increased to force M1 into subthreshold? Assume $I_{DO} = 20$ nA, $n = 2$, $K' = 24\ \mu A/V^2$, and $V_T = 0.75$ V.

3.20. Consider a single MOS transistor. If $V_{BS} = 0$ and $V_{DS} > 3V_t$, determine the value of V_{GS} at the intersection of the strong inversion drain saturation current of (4) in Table 3.1-1 with the weak inversion current. How does this relate to the transition region of Eq. 3.1-31? Assume $I_{DO} = 20$ nA, $n = 2$ in Eq. 3.1-32, $K' = 24\ \mu A/V^2$, and $\lambda = 0$.

3.21. Obtain the output voltage for the circuit layout of Color Plate 10 if $V_{DD} = 5$ V and $V_I = 2$ V. Assume the devices were fabricated in the NMOS process summarized in Appendix 2A.

3.22. In the saturation region of operation, most of the channel region is characterized by an inversion layer, which is an extension of the source as depicted in Fig. 2.2-2*f*. This represents a series impedance in the source. Assuming the voltage drop is sufficiently small that Eq. 3.1-8 applies, calculate the series source–channel impedance, R_{sc}. *Note*: Due to a tapering of the channel near the drain, the source–channel impedance is actually about 50% larger than that predicted by (3.1-8).

3.23. If γ and λ effects are neglected, a MOSFET of width W and length L is equivalent to the series connection of the two devices as shown in Fig. P3.23 where ϵ is any real number which satisfies the expression $0 < \epsilon \leq L$.

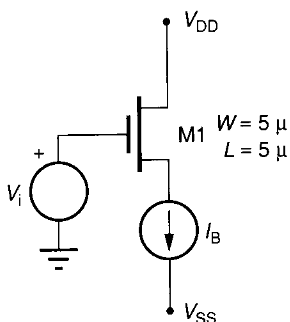


FIGURE P3.19
Circuit for Problem 3.19.

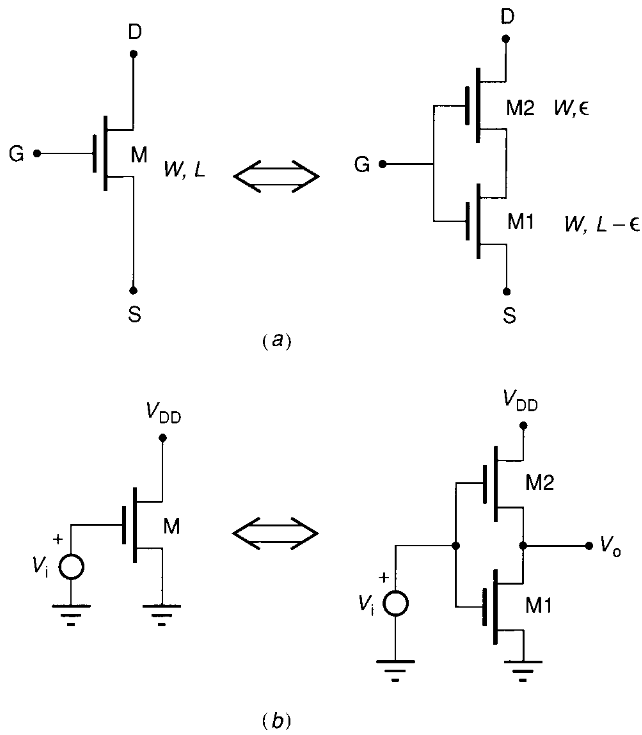


FIGURE P3.23
Circuits for Problem 3.23.

(*a*) If the MOSFET M is operating in the saturation region, show that in the equivalent circuit of Fig. P3.23*a*, M1 is operating in the ohmic region and M2 in the saturation region.

(*b*) If the equivalent MOSFET is connected as shown if Fig. P3.23*b* and M is operating in the saturation region, derive an expression for V_o in terms of V_i.

(*c*) Calculate the equivalent impedance of M1 in Fig. P3.23*b* assuming V_i is sufficiently small that (3.1-8) applies.

3.24. For the circuit of Fig. P3.24,

(*a*) Calculate parametrically the output thermal noise current in the frequency band 10 kHz $\leq f \leq$ 200 kHz and the output flicker noise current in the same

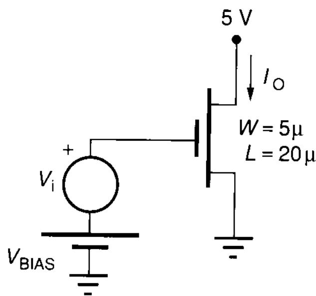


FIGURE P3.24
Circuit for Problem 3.24.

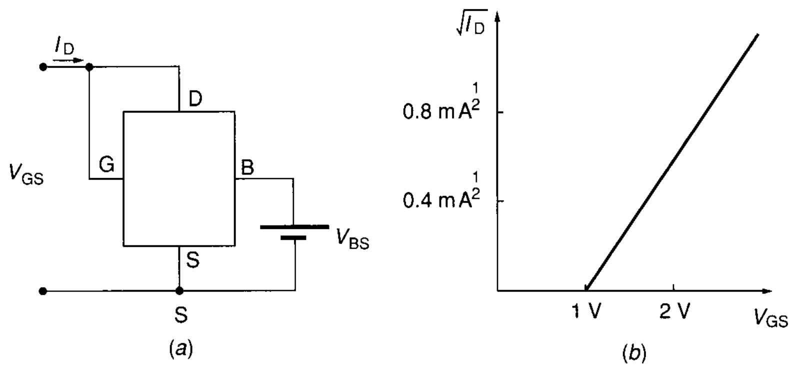


FIGURE P3.25
Circuit and graph for Problem 3.25.

band if $K_f = 3 \times 10^{-24}$ $V^2 \cdot F$ at the dc operating point determined by $V_i = 0$.

(*b*) Calculate parametrically the equivalent input referred offset voltage.

(*c*) Calculate the signal to noise ratio of the output current if $V_{BIAS} = 2$ V and $V_i = 0.25 \sin 10^5 t$ over the frequency band given in (*a*). Neglect λ and γ effects and assume the remaining process parameters given in Appendix 2B.

3.25. A plot of V_{GS} versus $\sqrt{I_D}$ for an NMOS transistor is shown in Fig. P3.25. It has $\mu C_{ox} = 60\ \mu A/V^2$, $V_{BS} = 0$, and the remainder of the characterization parameters as indicated in the NMOS design rules discussed in Chapter 2.

(*a*) Is the device a depletion or enhancement transistor?
(*b*) What mode of operation is the device in?
(*c*) What is W/L?
(*d*) What is the threshold voltage?
(*e*) What will the horizontal axis intercept if V_{BS} is changed to -9 V?

3.26. The pn junction between an n^+ moat diffusion and substrate is used to form a capacitor. Determine the value of this capacitor with a reverse bias of 6 V if the area of the n^+ moat diffusion is 16 mil^2. Assume the process parameters of the NMOS process of Appendix 2A.

3.27. Compare the minimum area required to build a first-order lowpass filter with a 3 db cutoff frequency of 2 kHz using the two techniques shown in Fig. P3.27*a* and *b*

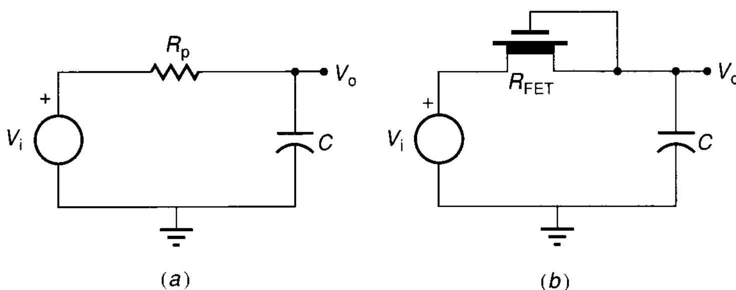


FIGURE P3.27
Circuits for Problem 3.27.

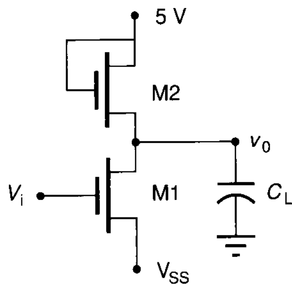


FIGURE P3.28
Circuit for Problem 3.28.

respectively. Assume R_p is made of a poly resistor with $R_\square = 30\ \Omega/\square$. R_{FET} is biased to operate in the ohmic region ($K' = 15\mu A/V^2$, $V_{T0} = -3.5V$, $\gamma = 0$, $\phi = 0$, $\lambda = 0$), the capacitance density is 0.2 pF/mil^2, and both R_p and R_{FET} are serpentined with minimum feature width of 5 μ, minimum feature spacing of 5 μ, and a gate–moat overlap of 2.5 μ. Derive all equations needed.

3.28. The voltage V_{SS} can be used to establish the quiescent drain currents with $V_i = 0$. Obtain an expression for the 3 db bandwidth of the circuit in Fig. P3.28 in terms of I_{DQ} and the model parameters of M1 and M2. Neglect all parasitic capacitors in M1 and M2.

3.29. Determine the small signal equivalent circuit for the circuit shown in Fig. P3.29.

3.30. Assume a MOSFET is characterized in strong and weak inversion by Eqs. (3.1-4) and (3.1-33) respectively.

(*a*) Show that if the transconductance gain is to be continuous at the strong inversion–weak inversion interface, then the transition must occur at $V_{GS} = V_T - 2nV_t$.

(*b*) Show that if I_D is to be continuous at the transition determined in (*a*), then I_{DO} and K' must be related by the equation

$$I_{DO} = \frac{K'2(nV_t)^2}{e^2}$$

Section 3.2

3.31. Show that the forward-biased pn junction capacitance model of Eq. 3.2-9 is a continuous and differentiable extension of the reverse-bias model of Eq. 3.2-8 at the transition $V = \phi_B/2$.

3.32. Assume a depletion region capacitor is characterized by $C_{ox} = 250$ fF/mil^2, $\phi = 0.6$ V, $n = 1/2$, and $A = 6$ mil^2. Determine the bias voltage required to reduce the capacitance to half of what it is at 0 V bias.

3.33. Accurately determine I in Fig. P3.33. Assume the diode is characterized by the BE diffusion of the bipolar process of Appendix 2C with a junction area of 500 μ^2. Assume operation at $T = 30°$ C.

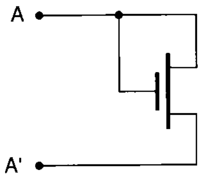


FIGURE P3.29
Circuit for Problem 3.29.

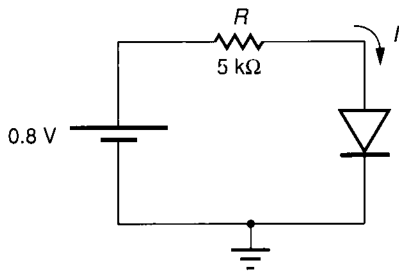


FIGURE P3.33
Circuit for Problem 3.33.

3.34. Assume the vertical npn base–collector junction of the process of Appendix 2C is to be used as a varactor diode.
(*a*) Determine the area of the base needed to generate a capacitance of 50 pF with a reverse bias of 3 V.
(*b*) Plot the capacitance of this diode versus reverse bias for $-5 \leq V_{BC} \leq 0$.
(*c*) Repeat part (*a*) if the BE junction is used instead of the BC junction in the same process.

3.35. A diode can be fabricated in the bipolar process of Appendix 2C in several ways. Give all ways that the diode can be fabricated and compare the performance characteristics from both dc and ac viewpoints. Include in your discussion a characterization of all relevant parasitics.

3.36. A "diode-connected transistor" is shown in Fig. P3.36*a*. Compare the performance of this to that of the BE and BC diodes (Figs. P3.36*b* and *c*) in the bipolar process of Appendix 2C.

3.37. The circuit shown in Fig. P3.37 is proposed as a rectifier. Compare the dc performance of this to that of the pn junction and that of the diode-connected transistor of Problem 3.36. How does the performance of this circuit as a rectifier change with *W* and *L*?

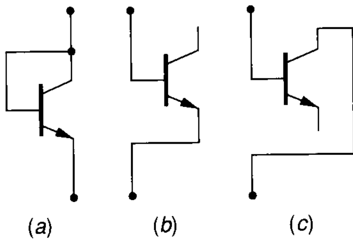


FIGURE P3.36
Diodes for Problem 3.36.

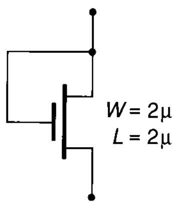


FIGURE P3.37
Circuit for Problem 3.37.

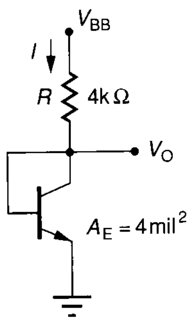


FIGURE P3.38
Circuit for Problem 3.38.

Section 3.3

3.38. Using Fig. P3.38, obtain V_O to within ±1% if (*a*) V_{BB} = 12 V and (*b*) V_{BB} = 0.9 V. Assume the BJT is characterized by the parameters of Table 3.3-1 and modeled by the Ebers-Moll Eqs. 3.3-1 and 3.3-2.

3.39. Determine the maximum error, in percentage, that results from using (3.3-10) and (3.3-11) instead of the more complicated Ebers-Moll equations of (3.3-1) and (3.3-2) in the region $V_{BE} > 0.5$ V, $V_{BC} < 0.3$ V. Assume the BJT is at room temperature and characterized by the parameters of Table 3.3-1.

3.40. Calculate the power that would be dissipated in a BJT if A_E = 4 mil^2 and the device is operating with (*a*) V_{BE} = 0.6 V and V_{BC} = −5 V, and (*b*) V_{BE} = 1.6 V and V_{BC} = −5 V. Assume the BJT is characterized by the process parameters of Table 3.3-1.

3.41. Plot the transconductance gain, g_m, versus bias current (collector or drain) for a MOSFET with $W = L = 10\ \mu$ and a BJT with $A_E = 100\ \mu^2$. Use the typical process parameters of Tables 3.1-2 and 3.3-1.

3.42. Rewrite the Ebers-Moll equations of (3.3-1) and (3.3-2) using V_{BE} and V_{CE} as the independent variables in terms of the parameters I_S, β_F, β_R, and V_T.

3.43. At high frequencies, a small signal gate current flows into the gate of the MOSFET. Compare the unity small signal–short circuit current gain frequency (of $A_I = i_o/i_i$)

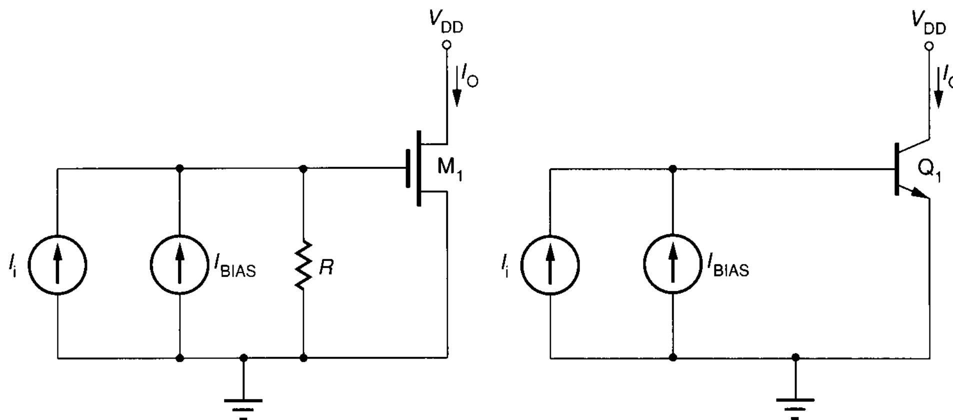


FIGURE P3.43
Circuit for Problem 3.43.

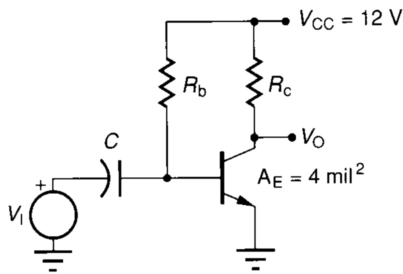


FIGURE P3.44
Circuit for Problem 3.44.

for the MOSFET to that of the BJT (see Fig. P3.43). Assume both devices are in their high-gain operating region, are of minimum size, are biased at $I_{OQ} = 200\,\mu A$, and are characterized by the process of Tables 2A and 2C of the Appendices of Chapter 2. Assume R is large.

3.44. (*a*) For the circuit shown in Fig. P3.44, bias the circuit (determine R_b) so that $V_{OQ} = 2$ V when $R_c = 500\ \Omega$. Determine the small signal voltage gain $A_v = v_o/v_i$ if it is assumed C is large. Use the typical process parameters of Table 3.3-1.
(*b*) Repeat part (*a*) if the parasitic emitter resistance is $R_E = 2\ \Omega$. Compare the results with those obtained in (*a*).
(*c*) Repeat parts (*a*) and (*b*) if $R_C = 10\ k\Omega$ and $V_{OQ} = 6$ V and compare with the results obtained in parts (*a*) and (*b*).

3.45. Assume the BJT of Fig. P3.45 is biased with R_b so that Q1 operates in the forward active region. What will be the tolerance in the value of V_{BE} due to processing? What will be the wafer-level tolerance in the value of V_{BE} due to wafer-level parameter variations? Assume the BJT process in which Q1 is fabricated is characterized by Table 2C.4 of Appendix 2C and that wafer-level variations (matching) are characterized by footnote 1 of that table.

3.46. The depletion region capacitance of the BE junction of an npn transistor with a rectangular emitter that is 0.3 mil × 6 mil is 0.63 pF at zero volts reverse bias. A circular transistor of the same test bar has the same emitter area. What is the depletion region capacitance of this device for a 6 V reverse bias at room temperature if the built-in potential, ϕ_B, is 0.70 V?

3.47. For the transistor shown in Fig. P3.47, $I_s = 2 \times 10^{-15}$ A and the emitter area is 1.5 mil^2. With $V = 1$ V, the current I is 52 μA, and with $V = 21$ V, I is 56 μA.
(*a*) Determine V_{AF}.
(*b*) Determine β_F.
(*c*) Determine V_{BE} at room temperature (300 °K).

3.48. If we plot ln I_C (vertical) versus V_{BE} (horizontal) for a bipolar transistor operating

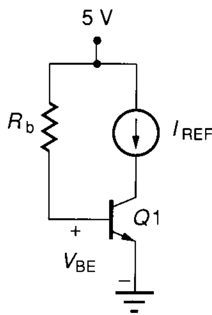


FIGURE P3.45
Circuit for Problem 3.45.

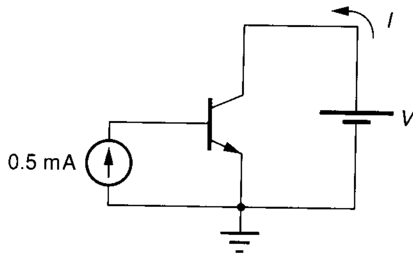


FIGURE P3.47
Transistor for Problem 3.47.

in the forward-active region, how can the area of the emitter be determined from the plot? What can be determined from the slope? Neglect Early voltage effects.

3.49. For the circuit shown in Fig. P3.49, assume the collector current is 10 mA, the emitter is a square with sides of 1 mil, the base is a square with sides of 3 mil, and the collector area is defined by a square with sides of 6 mil. For the following, assume that no changes are made in the circuit schematic.

(*a*) What will be the collector current if the sides of the base are increased to 4 mil?

(*b*) What will be the collector current if the emitter sides are decreased to 0.5 mil?

(*c*) What will be the collector current if the collector sides are increased to 7 mil?

3.50. A plot of ln I_C versus V_{BE} for a BJT transistor, with emitter area 4 mil^2, and a base area of 10 mil^2, operating in the forward active region is shown by the solid line in Fig. P3.50. Determine as many of the following model parameters as possible: J_S, T, V_T, β_F, β_R, α_F.

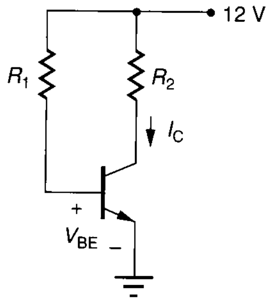


FIGURE P3.49
Circuit for Problem 3.49.

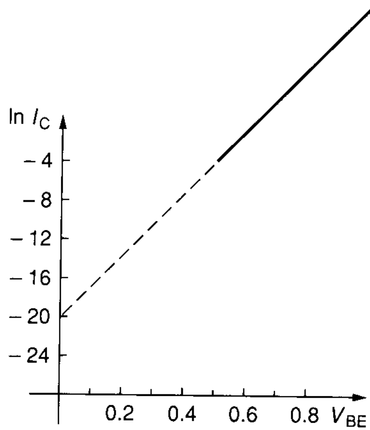


FIGURE P3.50
Graph for Problem 3.50.

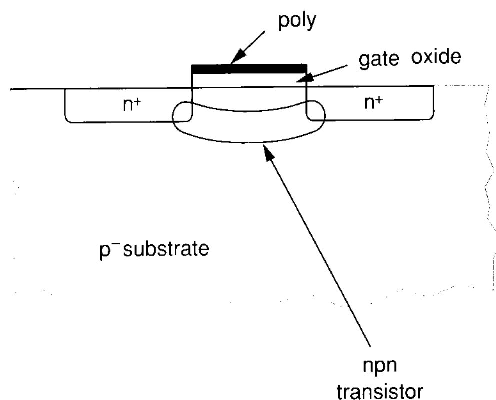


FIGURE P3.51
NMOS structure for Problem 3.51.

3.51. The NMOS structure shown in Fig. P3.51 is proposed as an npn transistor to be used with n-channel MOSFET devices. Give two good reasons why this device may not be practical.

3.52. Assume the two circuits shown in Fig. P3.52 are biased (with V_{B1} and V_{B2} respectively) so that $V_{OQ} = 5$ V. Compare the small signal voltage gains of the two circuits.

3.53. Calculate the value of C_{JE} and C_{JC} for the npn vertical transistor characterized by the process of Table 2C.4 in Appendix 2C. Assume the emitter is square with dimensions 20 μ × 20 μ, the base is square with sides of 50 μ, and the collector is square with sides of 100 μ. Compare these results to the values listed in Table 2C.5.

3.54. Calculate the value of C_{JE}, C_{JC}, and the base–substrate capacitance for a lateral pnp transistor with 30 μ × 30 μ square collector and emitter areas separated by a distance of 15 μ if they lie symmetrically inside an isolated square epitaxial region that is 110 μ on a side. Assume the process parameters of Table 2C.4 of the Appendix 2C.

3.55. From the equivalent circuit of Fig. 3.3-8, rigorously define the h parameters and derive an expression for the h parameters in the small-signal BJT model (a) in terms of the y parameters g_m, g_π, and g_o, and (b) in terms of the operating point and parameters used in the Ebers-Moll model.

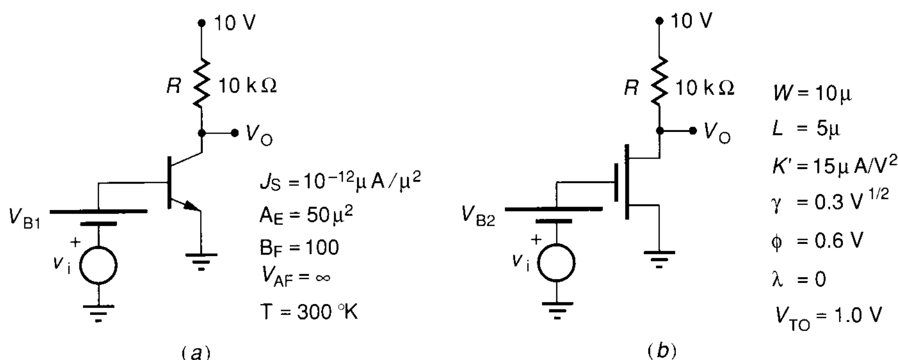


FIGURE P3.52
Circuits for Problem 3.52.

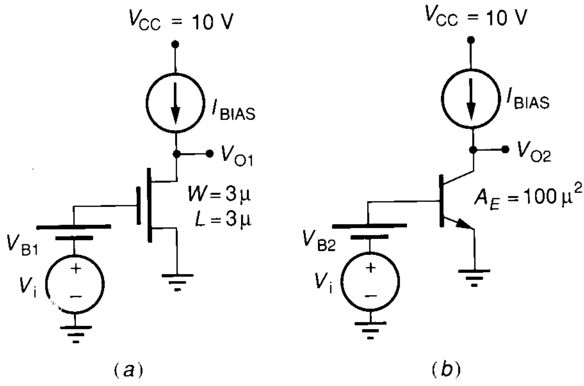


FIGURE P3.56
Amplifiers for Problem 3.56.

3.56. Plot the small signal voltage gains, v_{O1}/v_i and v_{O2}/v_i, versus bias current for the two amplifiers shown in Fig. P3.56. Assume that V_{B1} and V_{B2} are adjusted so that $V_{O1Q} = V_{O2Q} = 5$ V. Use the typical process parameters of Tables 3.1-2 and 3.3-1.

3.57. Two different schemes for biasing the BJT are shown in Fig. P3.57. Assume the BJT is characterized by the process parameters of Table 3.3-1.
(*a*) Determine the value of R_B and V_B necessary to bias the outputs at 5 V.
(*b*) Determine the change in V_{O1} and V_{O2} that will be expected for each circuit if V_{CC} varies around 10 V by ± 0.1 V.
(*c*) Determine the change in V_{O2} that will be expected if V_B varies about the value obtained in part (*a*) by ± 0.1 V.
(*d*) What can be said about the sensitivity of the quiescent output voltage to the biasing voltages V_{CC} and V_B for the two circuits?

Section 3.4

3.58. If the sheet resistance of polysilicon is $R_{\square}$, the minimum poly width is d_4, and the minimum poly spacing is d_3, determine the resistance density of a large, minimum-feature-size, serpentined poly resistor in terms of d_3, d_4, and $R_{\square}$. Neglect any area associated with corners and contacts.

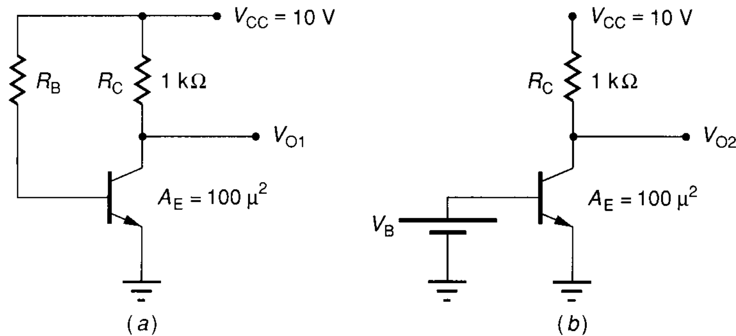


FIGURE P3.57
Circuits for Problem 3.57.