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.

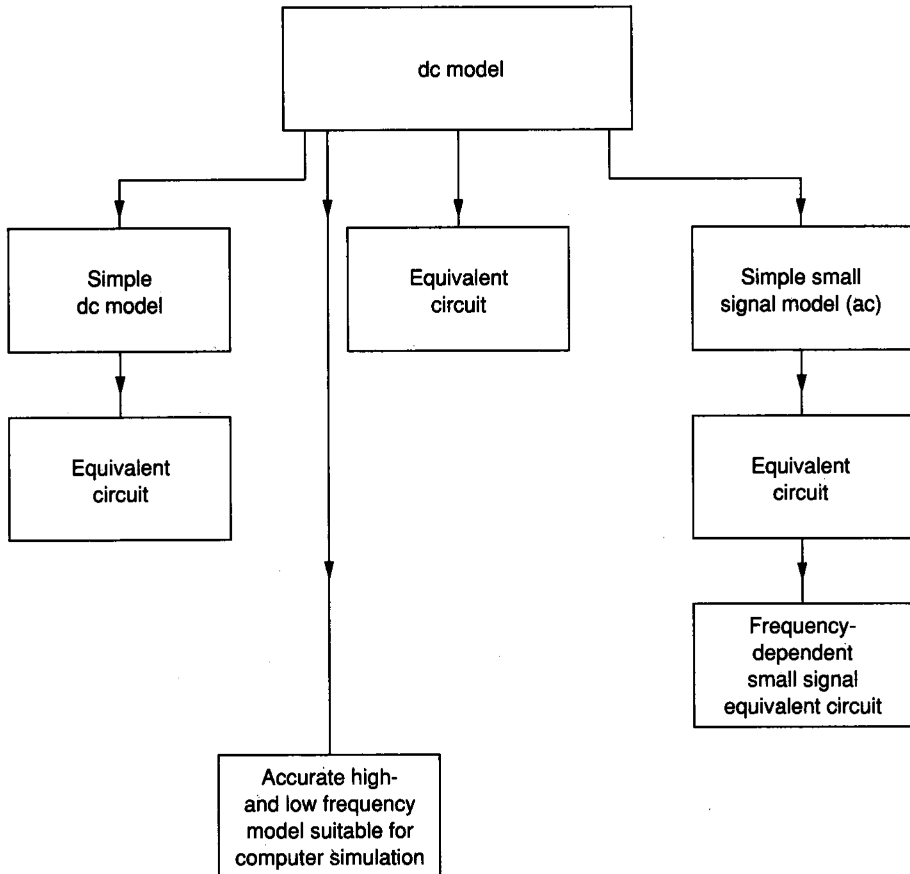


FIGURE 3.0-1
Approach to device modeling.

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

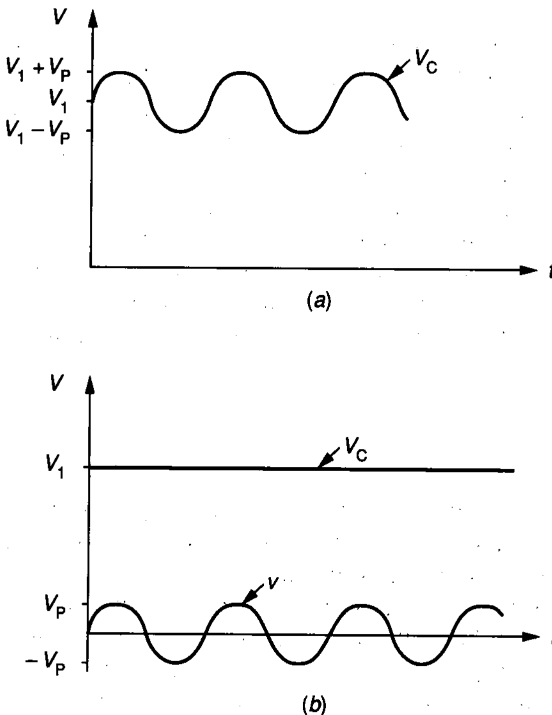


FIGURE 3.0-2
Relationship between total (V_C), small signal (v), and quiescent (V_c) values (*a*) total instantaneous value, (*b*) quiescent and small signal values.

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. 3.0-3a, 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

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

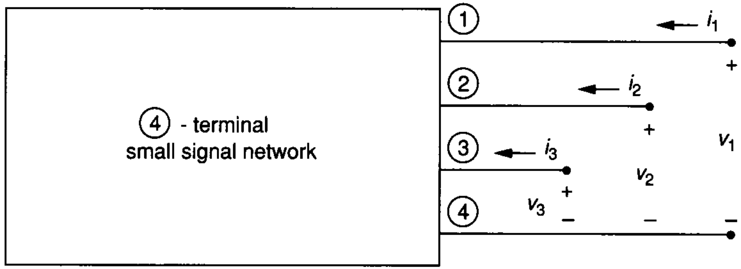
Since the small signal voltage and current variables are the time-varying part of the corresponding total terminal voltage and current variables in the parent network relative to the Q-point (see Fig. 3.0-3b), then it follows that the y–parameters can be obtained from the large signal variables and thus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

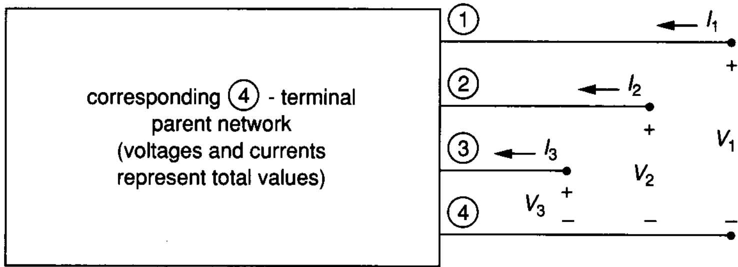
$$y_{kj} = \left.\frac{\partial I_k}{\partial V_j}\right|_{V_m=quiescent\ \ value,\ m\in\{1,\,2,\,3\}} \tag{3.0-5}$$

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. 3.0-3c has the same i-v characteristics as the four-terminal network of Fig. 3.0-3a, which is characterized by Eqs. 3.0-3 and hence is electrically indistinguishable. In Fig. 3.0-3c 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.





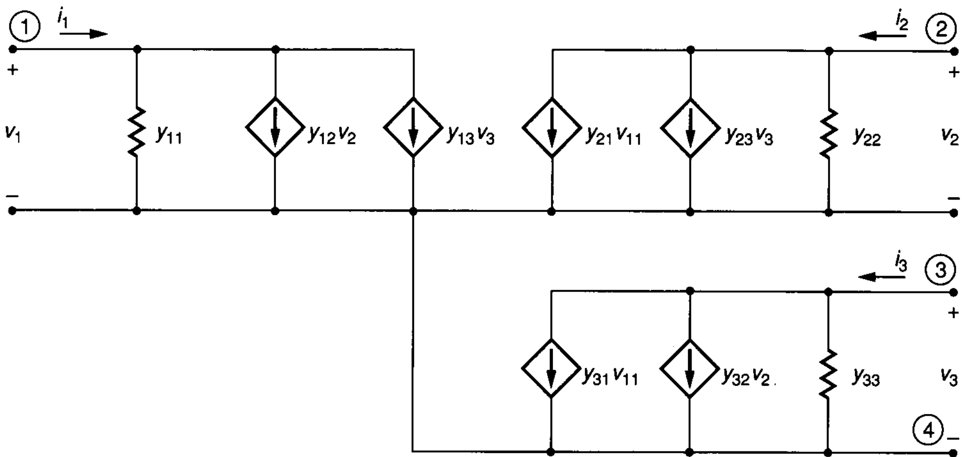


FIGURE 3.0-3
General four-terminal network: (*a*) Small signal, (*b*) Parent network, (*c*) Equivalent circuit for small signal network.

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

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

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

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

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

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

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

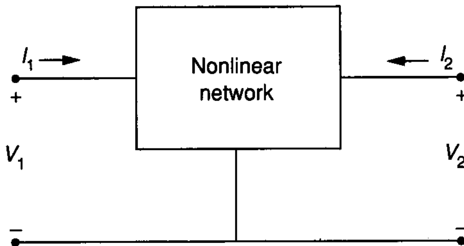


FIGURE 3.0-4
Example 3.0-1.

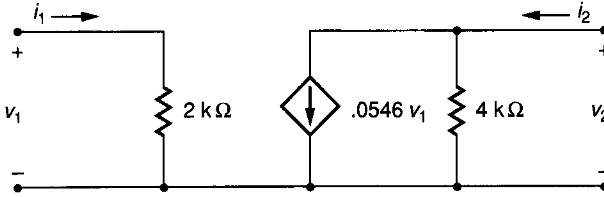


FIGURE 3.0-5
Solution of Example 3.0-1.

(*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

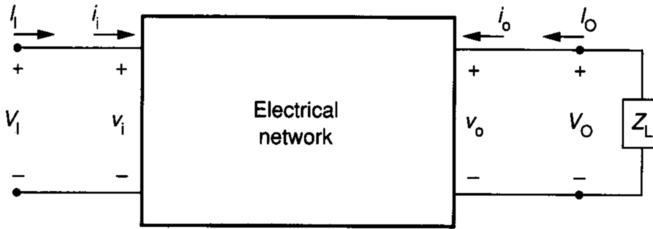


FIGURE 3.0-6
Typical electrical network structure.

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, $\lambda = 0.1$ 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.

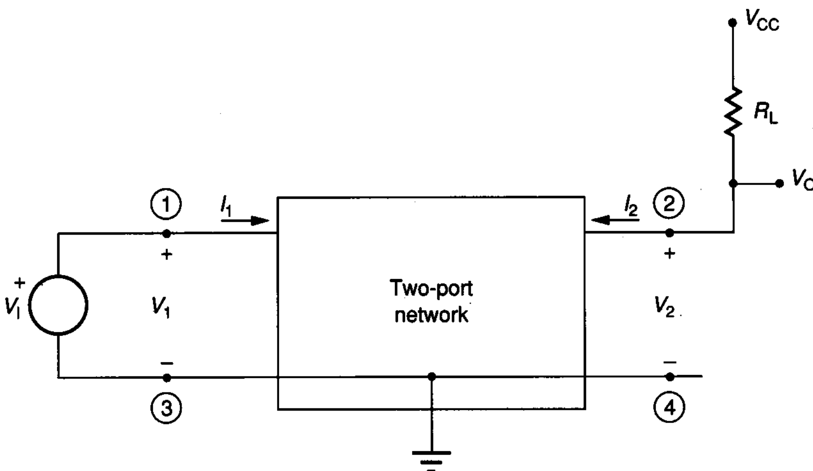


FIGURE 3.0-7
Circuit for Example 3.0-2.

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

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

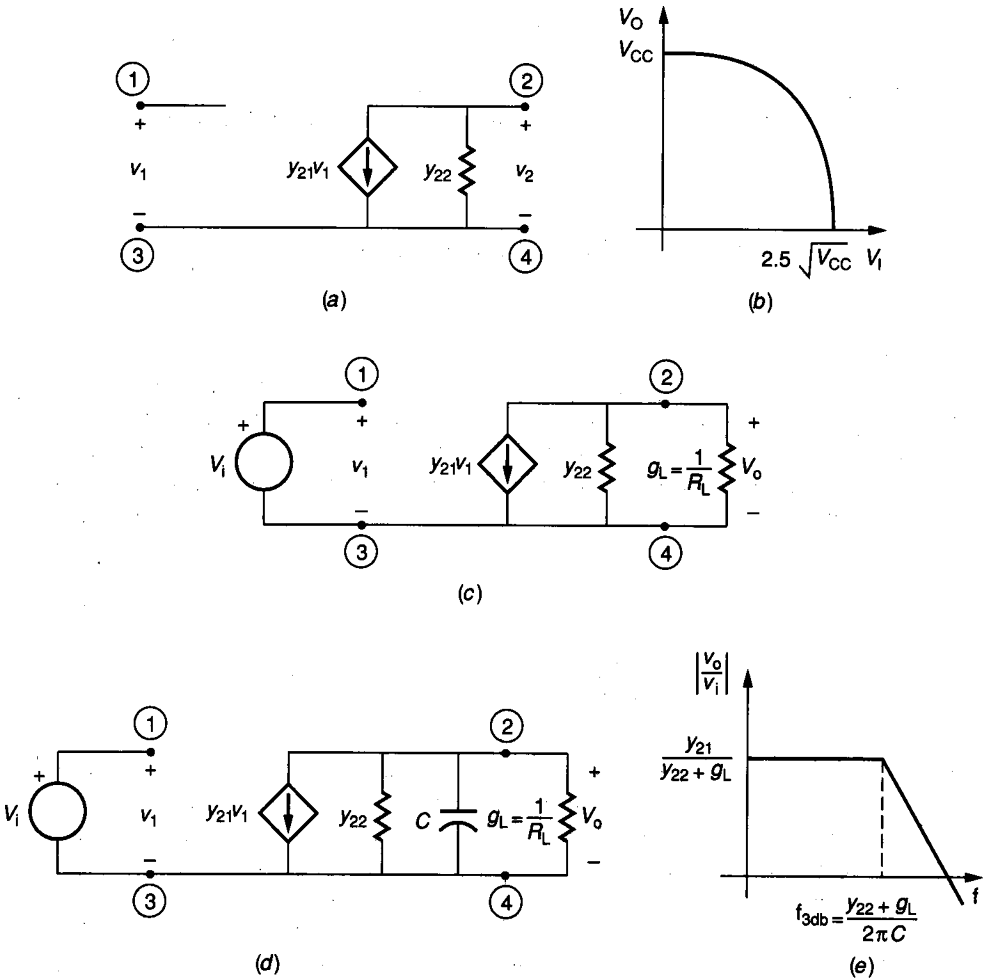


FIGURE 3.0-8
Solution of Example 3.0-2: (*a*) Small signal device model, (*b*) Large signal transfer characteristics, (*c*) Small signal equivalent circuit, (*d*) Small signal equivalent circuit including parasitic capacitor, (*e*) Magnitude of frequency response.

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.

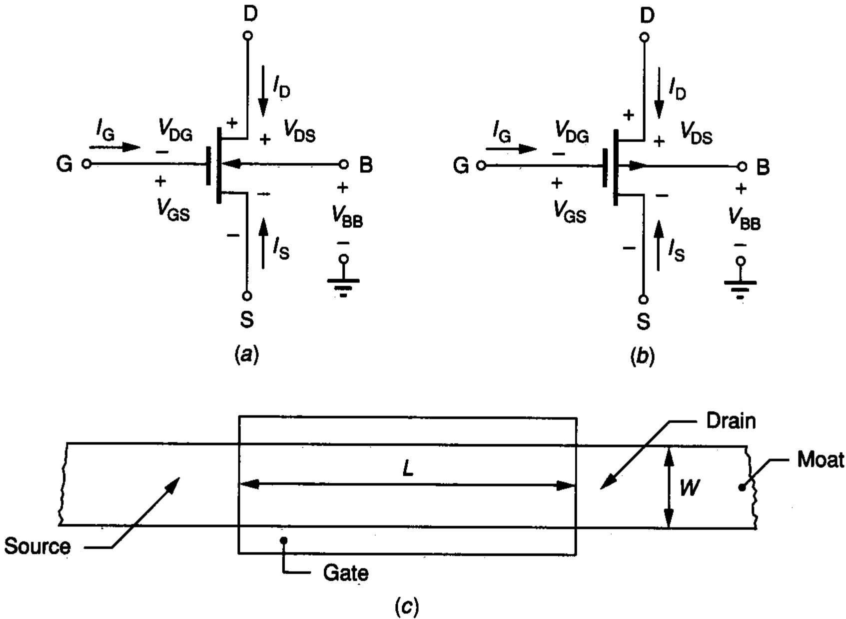


FIGURE 3.1-1
Convention for electrical variables of MOS transistors: (*a*) n-channel, (*b*) p-channel, (*c*) Top view of device geometry. Not shown is V_{BS}, the voltage from bulk (B) to source (S).

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.

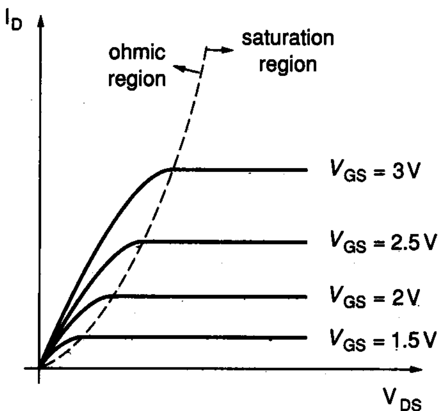


FIGURE 3.1-2
Typical output characteristics for n-channel MOSFET.

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

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

$$I_G = 0 \qquad V_{GS} < V_T \qquad (3.1\text{-}2)$$

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

$$V_{DS} = V_{GS} - V_T \qquad (3.1\text{-}3)$$

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

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

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

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

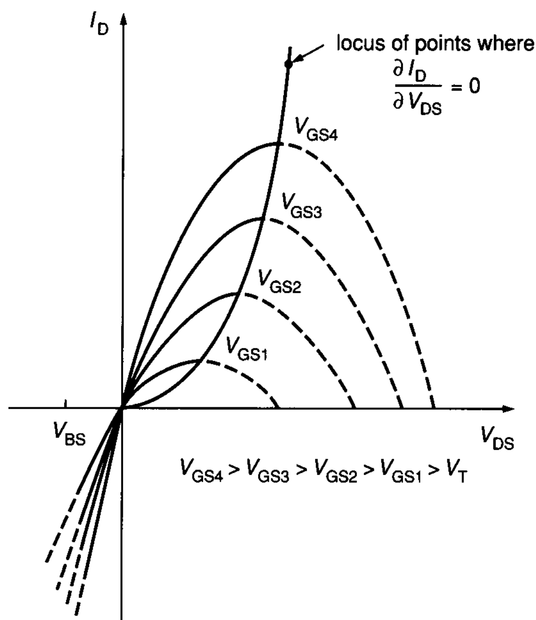


FIGURE 3.1-3
I_D–V_{DS} characteristics of MOSFET as predicted by Sah's equation.

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} = threshold voltage for $V_{BS} = 0$

γ = bulk threshold parameter

ϕ = strong inversion surface potential

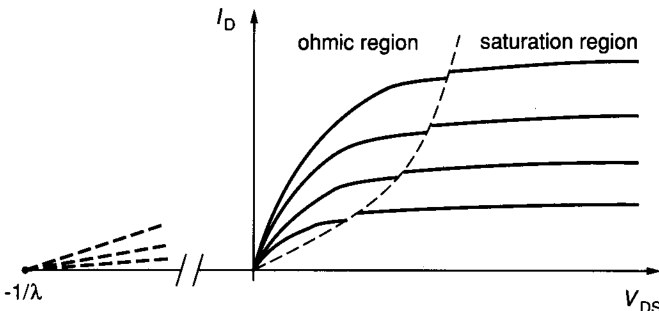


FIGURE 3.1-4
Discontinuity in model due to λ effects.

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 $V_{T0} > 0$.

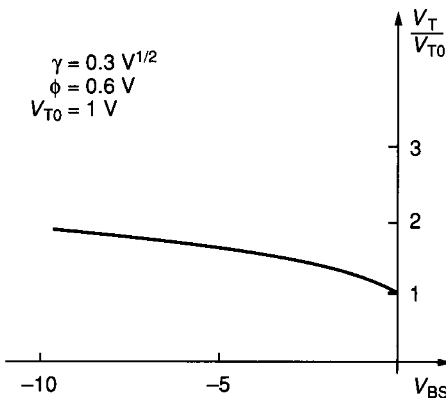


FIGURE 3.1-5
Effects of V_{BS} on threshold voltage.

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 μ 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

n-channel MOSFET		
$I_G = 0$		(1)
$I_D = 0$	$V_{GS} < V_T$ (cutoff), $V_{DS} \geq 0$	(2)
$I_D = \frac{K'W}{L}\left(V_{GS} - V_T - \frac{V_{DS}}{2}\right) V_{DS}$	$V_{GS} > V_T$, $0 < V_{DS} < V_{GS} - V_T$ (ohmic)	(3)
$I_D = \frac{K'W}{2L}(V_{GS} - V_T)^2 (1 + \lambda V_{DS})$	$V_{GS} > V_T$, $V_{DS} > V_{GS} - V_T$ (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)
$I_D = -\frac{K'W}{L}\left(V_{GS} - V_T - \frac{V_{DS}}{2}\right) V_{DS}$	$V_{GS} < V_T$, $0 > V_{DS} > V_{GS} - V_T$ (ohmic)	(8)
$I_D = -\frac{K'W}{2L}(V_{GS} - V_T)^2 (1 - \lambda V_{DS})$	$V_{GS} < V_T$, $V_{DS} < V_{GS} - V_T$ (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

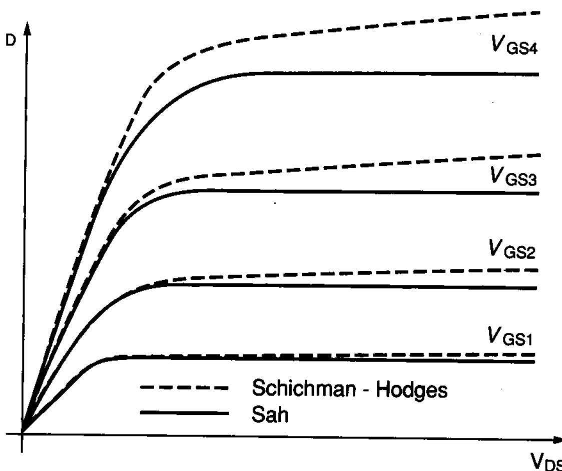


FIGURE 3.1-6
Comparison of Sah's model and the Shichman-Hodges model.

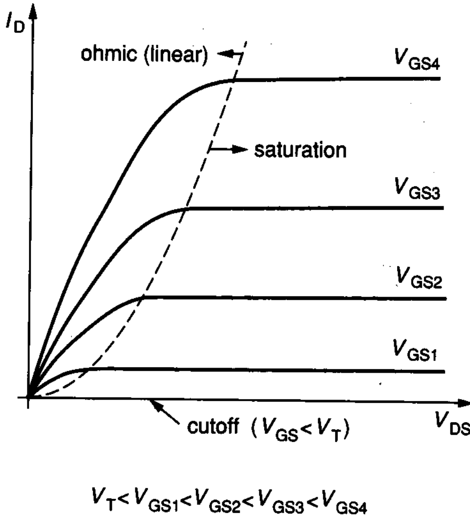


FIGURE 3.1-7
Regions of operation of MOSFET in I_D–V_{DS} plane.

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

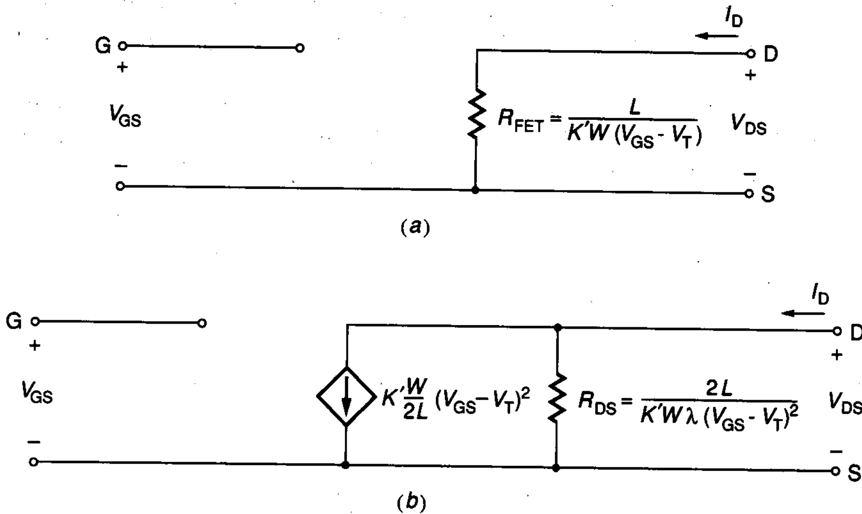


FIGURE 3.1-8
Equivalent circuits of MOSFET for dc operation: (*a*) Ohmic region (small V_{DS}), (*b*) Saturation region.

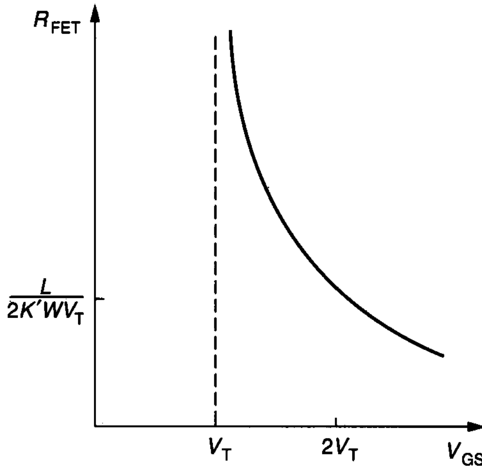


FIGURE 3.1-9
MOSFET resistance versus V_{GS} in the ohmic region.

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.

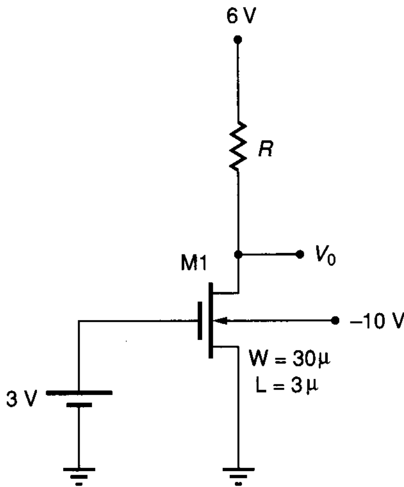


FIGURE 3.1-10
Example 3.1-1.

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

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

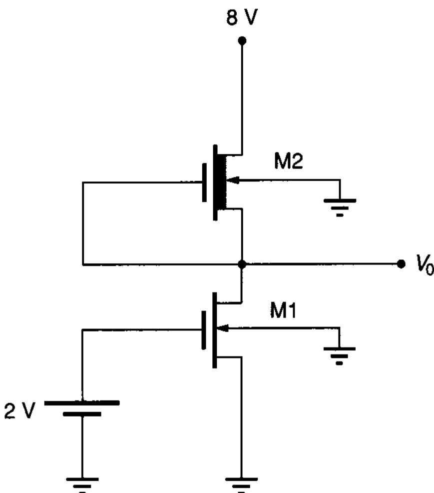


FIGURE 3.1-11
Example 3.1-2.

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 $0\text{ V} \leq V_I \leq 6\text{ V}$.

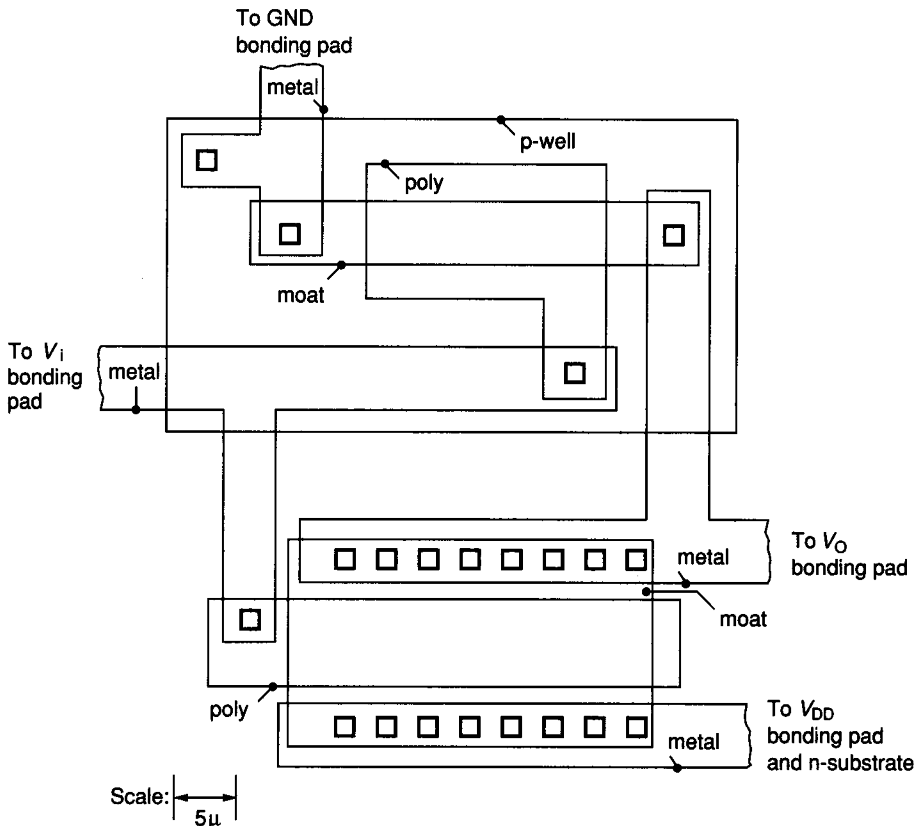


FIGURE 3.1-12
Example 3.1-3—CMOS inverter.

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

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

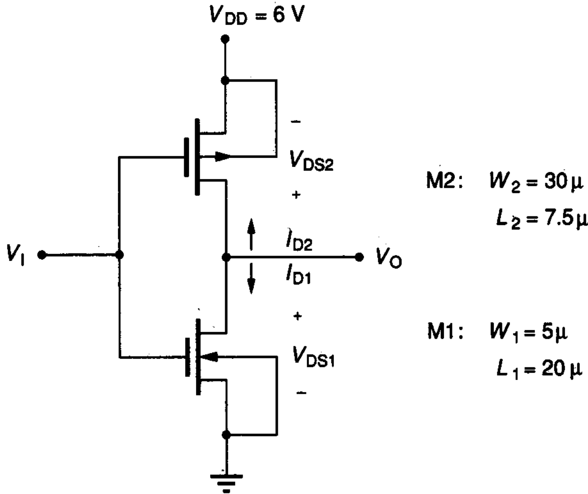
which will happen for all $V_I < V_{TON}$ only if $V_O = V_{DD}$. The portion of the transfer characteristics where M1 is cutoff and M2 is ohmic is shown as region ① in Fig. 3.1-13*b*.

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

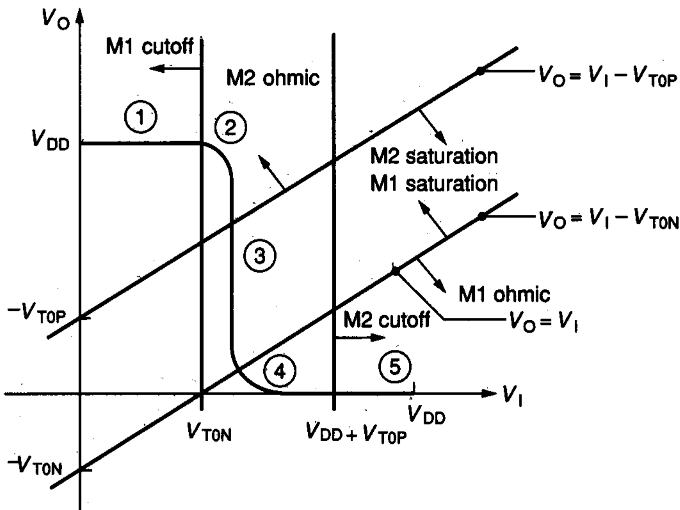
$$K_1'\frac{W_1}{2L_1}(V_I - V_{TON})^2(1 + \lambda V_O) = \frac{K_2'W_2}{L_2}\left\{\left([V_I - V_{DD} - V_{TOP}] - \frac{[V_O - V_{DD}]}{2}\right)(V_O - V_{DD})\right\}$$

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. 3.1-13b, 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. 3.1-13b.

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. 3.1-13b 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),

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

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

$$g_m \simeq \sqrt{\frac{2K'W}{L}}\sqrt{|I_{DQ}|} \tag{3.1-11}$$

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.

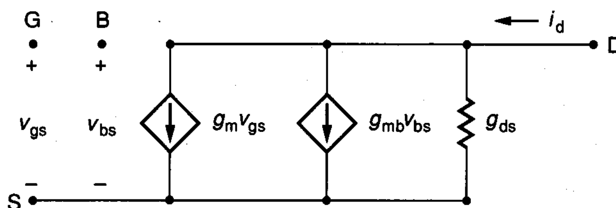


FIGURE 3.1-14
Small signal model equivalent circuit for MOSFET.

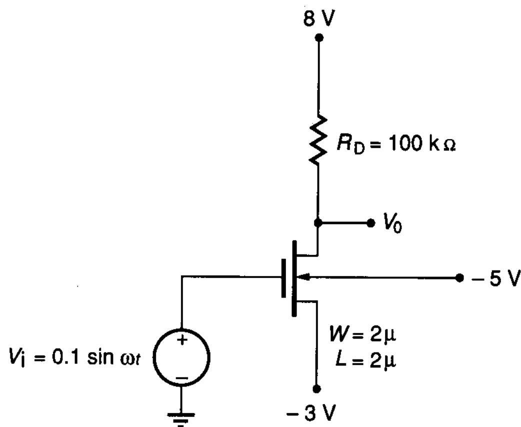


FIGURE 3.1-15
Example 3.1-4.

(*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

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

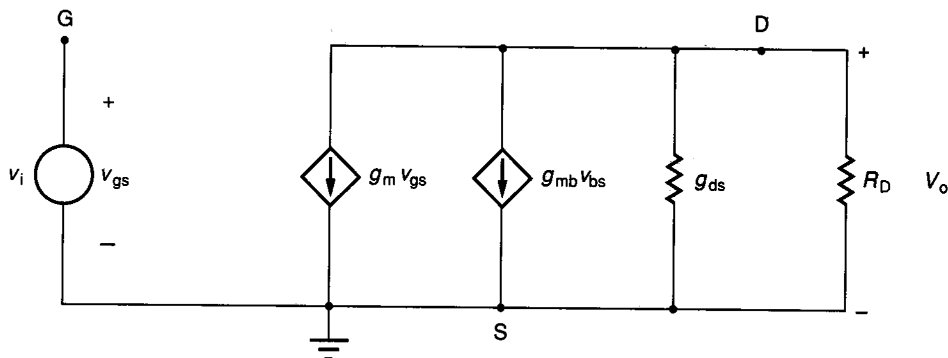


FIGURE 3.1-16
Solution of Example 3.1-4.

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.

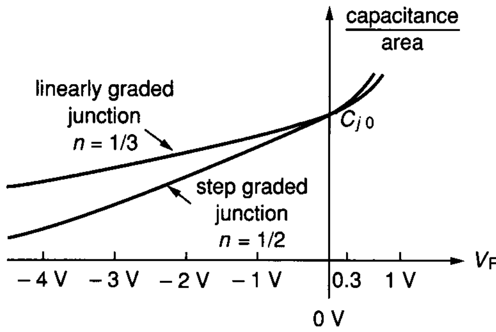


FIGURE 3.1-17
Voltage-dependent capacitance of pn junctions.

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

$$C_{ox} = \frac{\epsilon_0 \epsilon_{SiO_2}}{t_{ox}} \tag{3.1-15}$$

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

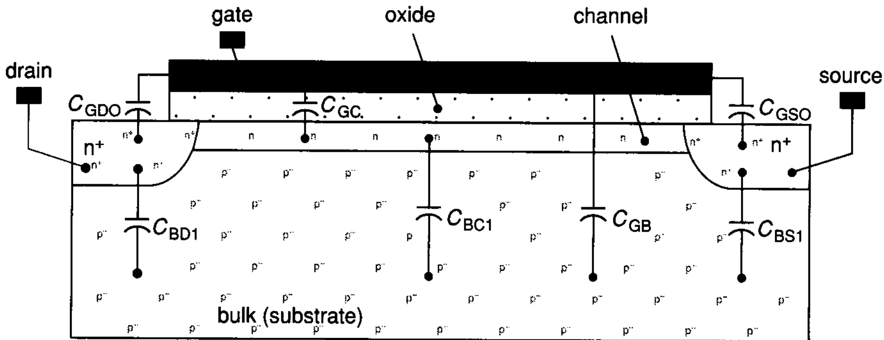


FIGURE 3.1-18
Parasitic capacitors in MOS transistors, shown for n-channel device.

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

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

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

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

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.

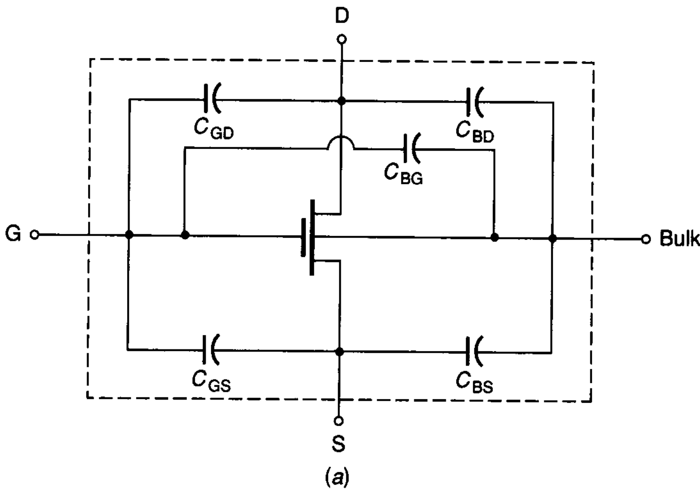
C_{BS1} and C_{BD1} are the capacitances of the bulk/source and bulk/drain junctions. These are often approximated with (3.1-14) using the source and drain areas and the process parameter C_{BM0}, which represents the zero bias bulk–moat capacitance density.

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}	$C_{OX} W L_D$	$C_{OX} W L_D + \frac{1}{2} WLC_{OX}$	$C_{OX} W L_D$
C_{GS}	$C_{OX} W L_D$	$C_{OX} W L_D + \frac{1}{2} WLC_{OX}$	$C_{OX} W L_D + \frac{2}{3} WLC_{OX}$
C_{BG}	$C_{OX} W L$	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.

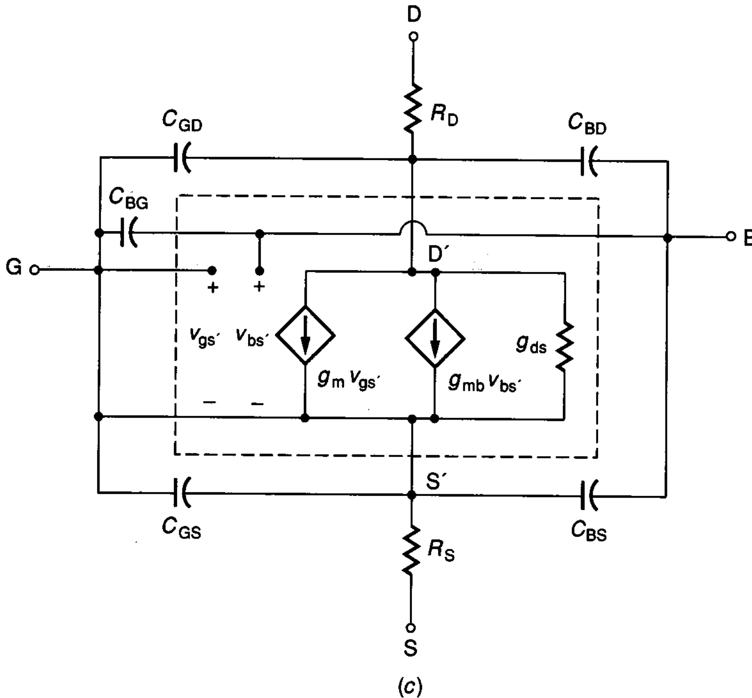


FIGURE 3.1-19
(*continued*)

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

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

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

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

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

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

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

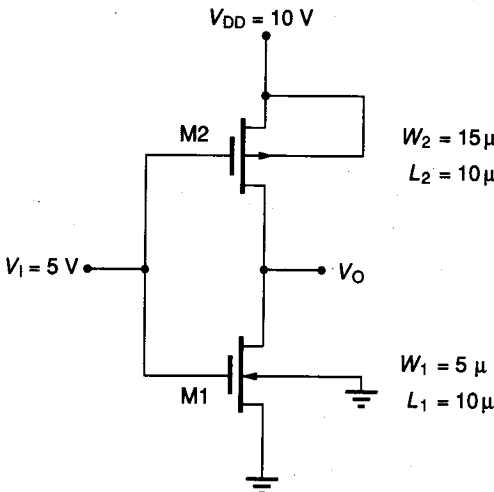


FIGURE 3.1-20
Example 3.1-5.