ANALOG INTEGRATED
CIRCUIT DESIGN

DAVID A. JOHNS

KEN MARTIN

ANALOG INTEGRATED CIRCUIT DESIGN

David Johns

Ken Martin

University of Toronto

Contents

sometimes used in digital circuits, where the circle indicates that a low voltage on the gate turns the transistor on, as opposed to a high voltage for an n-channel transistor (Fig. 1.7(a)). The symbols of Fig. 1.8(d) or Fig. 1.8(e) might be used in larger circuits where many transistors are present, to simplify the drawing somewhat. They will not be used in this text.

Basic Operation

The basic operation of MOS transistors will be described with respect to an n-channel transistor. First, consider the simplified cross sections shown in Fig. 1.9, where the source, drain, and substrate are all connected to ground. In this case, the MOS transistor operates similarly to a capacitor. The gate acts as one plate of the capacitor, and the surface of the silicon, just under the thin insulating SiO_2, acts as the other plate.

If the gate voltage is very negative, as shown in Fig. 1.9(a), positive charge will be attracted to the channel region. Since the substrate was originally doped p^-, this negative gate voltage has the effect of simply increasing the channel doping to p^+,

(a)

(b)
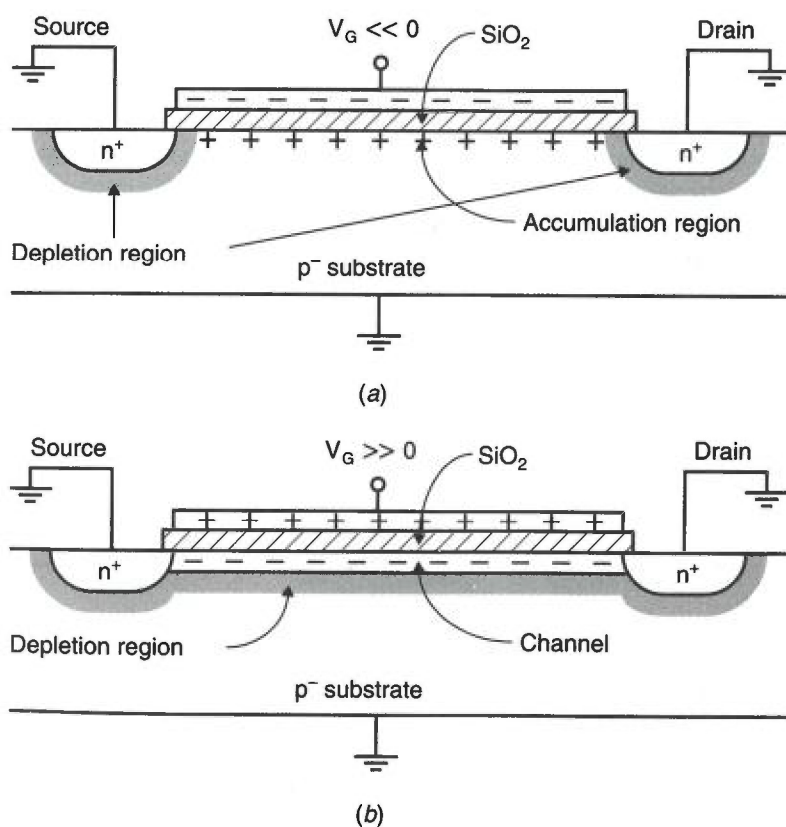
Fig. 1.9 An n-channel MOS transistor. (a) $V_G \ll 0$, resulting in an accumulated channel (no current flow); (b) $V_G \gg 0$, and the channel is present (current flow possible from drain to source).

resulting in what is called an *accumulated channel*. The n^+ source and drain regions are separated from the p^+-channel region by depletion regions, resulting in the equivalent circuit of two back-to-back diodes. Thus, only leakage current will flow even if one of the source or drain voltages becomes large (unless the drain voltage becomes so large as to cause the transistor to break down).

In the case of a positive voltage being applied to the gate, the opposite situation occurs, as shown in Fig. 1.9(*b*). For small positive gate voltages, the positive carriers in the channel under the gate are initially repulsed and the channel changes from a p^- doping level to a depletion region. As a more positive gate voltage is applied, the gate attracts negative charge from the source and drain regions, and the channel becomes an n region with mobile electrons connecting the drain and source regions.[5] In short, a sufficiently large positive gate-source voltage changes the channel beneath the gate to an n region, and the channel is said to be *inverted*.

The gate-source voltage, for which the concentration of electrons under the gate is equal to the concentration of holes in the p^- substrate far from the gate, is commonly referred to as the *transistor threshold voltage* and denoted V_{tn} (for n-channel transistors). For gate-source voltages larger than V_{tn}, there is an n-type channel present, and conduction between the drain and the source can occur. For gate-source voltages less than V_{tn}, it is normally assumed that the transistor is off and no current flows between the drain and the source. However, it should be noted that this assumption of zero drain-source current for a transistor that is off is only an approximation. In fact, for gate voltages around V_{tn}, there is no abrupt current change, and for gate-source voltages slightly less than V_{tn}, small amounts of *subthreshold current* can flow, as discussed in Section 1.3.

When the gate-source voltage, V_{GS}, is larger than V_{tn}, the channel is present. As V_{GS} is increased, the density of electrons in the channel increases. Indeed, the carrier density, and therefore the charge density, is proportional to $V_{GS} - V_{tn}$, which is often called the *effective gate-source voltage* and denoted V_{eff}. Specifically, define

$$V_{eff} \equiv V_{GS} - V_{tn} \qquad (1.54)$$

The charge density of electrons is then given by

$$Q_n = C_{ox}(V_{GS} - V_{tn}) = C_{ox}V_{eff} \qquad (1.55)$$

Here, C_{ox} is the gate capacitance per unit area and is given by

$$C_{ox} = \frac{K_{ox}\varepsilon_0}{t_{ox}} \qquad (1.56)$$

where K_{ox} is the relative permittivity of SiO_2 (approximately 3.9) and t_{ox} is the thickness of the thin oxide under the gate. A point to note here is that (1.55) is only accurate when both the drain and the source voltages are zero.

---

5. The drain and source regions are sometimes called diffusion regions or junctions for historical reasons. This use of the word *junction* is not synonymous with our previous use, in which it designated a pn interface of a diode.

To obtain the total gate capacitance, (1.56) should be multiplied by the effective gate area, WL, where W is the gate width and L is the effective gate length. These dimensions are shown in Fig. 1.10. Thus the total gate capacitance, C_{gs}, is given by

$$C_{gs} = WLC_{ox} \tag{1.57}$$

and the total charge of the channel, Q_{T-n}, is given by

$$Q_{T-n} = WLC_{ox}(V_{GS} - V_{tn}) = WLC_{ox}V_{eff} \tag{1.58}$$

The gate capacitance, C_{gs}, is one of the major load capacitances that circuits must be capable of driving. Gate capacitances are also important when one is calculating *charge injection,* which occurs when a MOS transistor is being turned off because the channel charge, Q_{T-n}, must flow from under the gate out through the terminals to other places in the circuit.

Next, if the drain voltage is increased above 0 V, a drain-source potential difference exists. This difference results in current flowing from the drain to the source.[6] The relationship between V_{DS} and the drain-source current, I_D, is the same as for a resistor, assuming V_{DS} is small. This relationship is given [Sze, 1981] by

$$I_D = \mu_n Q_n \frac{W}{L} V_{DS} \tag{1.59}$$

where $\mu_n \cong 0.06 \text{ m}^2/\text{Vs}$ is the mobility of electrons near the silicon surface, and Q_n is the charge concentration of the channel per unit area (looking from the top down). Note that as the channel length increases, the drain-source current decreases, whereas this current increases as either the charge density or the transistor width increases. Using (1.58) and (1.59) results in

$$I_D = \mu_n C_{ox} \frac{W}{L}(V_{GS} - V_{tn})V_{DS} = \mu_n C_{ox} \frac{W}{L} V_{eff} V_{DS} \tag{1.60}$$
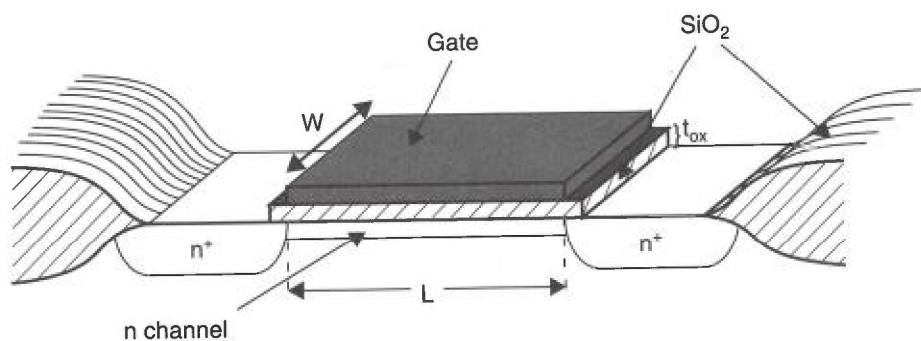
Fig. 1.10   The important dimensions of a MOS transistor.

---

6. The current is actually conducted by negative carriers (electrons) flowing from the source to the drain. Negative carriers flowing from source to drain results in a positive current from drain to source, I_{DS}.

where it should be emphasized that this relationship is only valid for drain-source voltages near zero (i.e., V_{DS} much smaller than V_{eff}).

As the drain-source voltage increases, the channel charge concentration decreases at the drain end. This decrease is due to the smaller gate-to-channel voltage difference across the thin gate oxide as one moves closer to the drain. In other words, since the drain voltage is assumed to be at a higher voltage than the source, there is an increasing voltage gradient from the source to the drain, resulting in a smaller gate-to-channel voltage near the drain. Since the charge density at a distance x from the source end of the channel is proportional to $V_G - V_{ch}(x) - V_{tn}$, as $V_G - V_{ch}(x)$ decreases, the charge density also decreases.[7] This effect is illustrated in Fig. 1.11.

Note that at the drain end of the channel, we have

$$V_G - V_{ch}(L) = V_{GD} \tag{1.61}$$

For small V_{DS}, we saw from (1.60) that I_D was linearly related to V_{DS}. However, as V_{DS} increases, and the charge density decreases near the drain, the relationship becomes nonlinear. In fact, the linear relationship for I_D versus V_{DS} flattens for larger V_{DS}, as shown in Fig. 1.12.
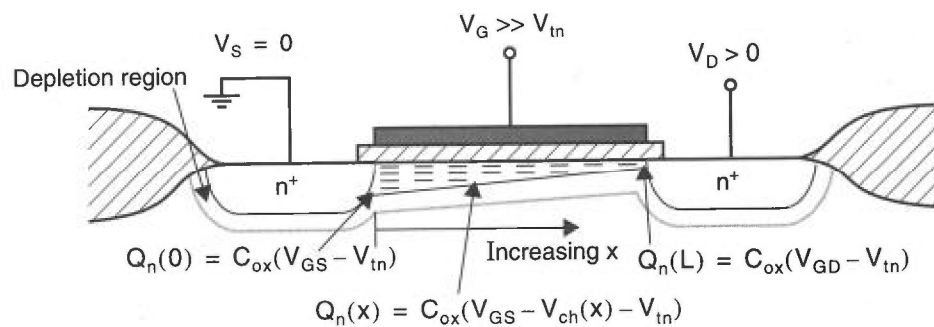
Fig. 1.11 The channel charge density for $V_{DS} > 0$.
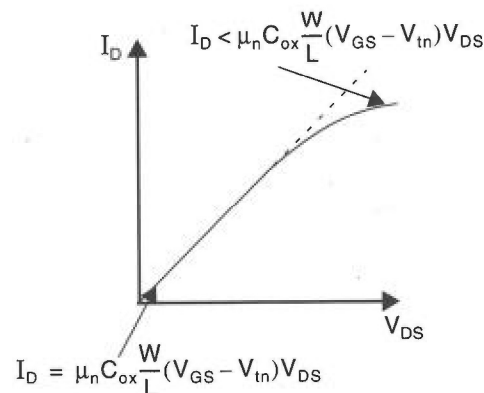
Fig. 1.12 For V_{DS} not close to zero, the I_D versus V_{DS} relationship is no longer linear.

---

7. $V_G - V_{CH}(x)$ is the gate-to-channel voltage drop at distance x from the source end, with V_G being the same everywhere in the gate, since the gate material is highly conductive.

As the drain voltage is increased, at some point the gate-to-channel voltage at the drain end will decrease to the threshold value V_{tn} — the minimum gate-to-channel voltage needed for n carriers in the channel to exist. Thus, at the drain end, the channel becomes *pinched off,* as shown in Fig. 1.13. This pinch-off occurs at $V_{GD} = V_{tn}$, since the channel voltage at the drain end is simply equal to V_D. Thus, pinch-off occurs for

$$V_{DG} > -V_{tn} \tag{1.62}$$

Denoting $V_{DS\text{-}sat}$ as the drain-source voltage when the channel becomes pinched off, we can substitute $V_{DG} = V_{DS} - V_{GS}$ into (1.62) and find an equivalent pinch-off expression

$$V_{DS} > V_{DS\text{-}sat} \tag{1.63}$$

where $V_{DS\text{-}sat}$ is given[8] by

$$V_{DS\text{-}sat} = V_{GS} - V_{tn} = V_{eff} \tag{1.64}$$

The electron carriers travelling through the pinched-off drain region are velocity saturated, similar to a gas under pressure travelling through a very small tube. If the drain-gate voltage rises above this critical pinch-off voltage of $-V_{tn}$, the charge concentration in the channel remains constant (to a first-order approximation) and the drain current no longer increases with increasing V_{DS}. The result is the current-voltage relationship shown in Fig. 1.14 for a given gate-source voltage. In the region of operation where $V_{DS} > V_{DS\text{-}sat}$, the drain current is independent of V_{DS} and is called the *active region.*[9] The region where I_D changes with V_{DS} is called the *triode region.* When MOS transistors are used in analog amplifiers, they almost always are biased in the active region. When they are used in digital logic gates, they often operate in both regions.
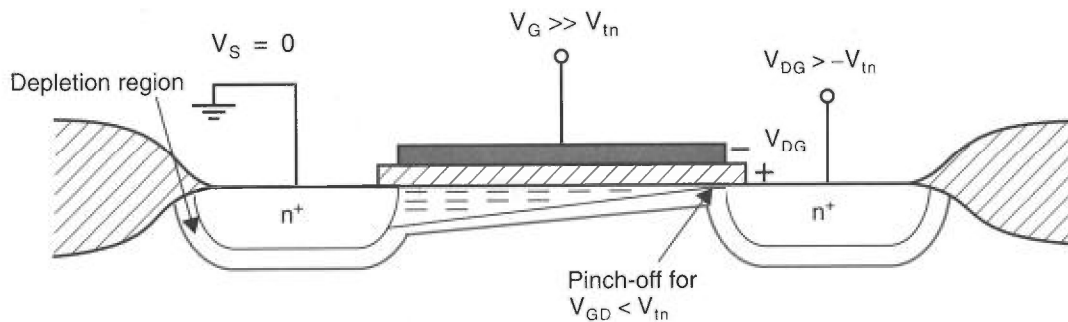
Fig. 1.13   When V_{DS} is increased so that $V_{GD} < V_{tn}$, the channel becomes pinched off at the drain end.

8. Because of the body effect, the threshold voltage at the drain end of the transistor is increased, resulting in the true value of $V_{DS\text{-}sat}$ being slightly lower than V_{eff}.
9. Historically, the active region was called the saturation region, but this led to confusion because in the case of bipolar transistors, the saturation region occurs for small V_{CE}, whereas for MOS transistors it occurs for large V_{DS}. The renaming of the saturation region to the active region is becoming widely accepted.

$$I_D = \mu_n C_{ox} \frac{W}{L}\left[(V_{GS} - V_{tn})V_{DS} - \frac{V_{DS}^2}{2}\right] \qquad\qquad I_D = \frac{\mu_n C_{ox}}{2}\frac{W}{L}(V_{GS} - V_{tn})^2$$

$$I_D \cong \mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{tn})V_{DS}$$

Fig. 1.14 The I_D versus V_{DS} curve for an ideal MOS transistor. For $V_{DS} > V_{DS\text{-}sat}$, I_D is approximately constant.

Before proceeding, it is worth discussing the terms *weak, moderate,* and *strong inversion.* As just discussed, a gate-source voltage greater than V_{tn} results in an inverted channel, and drain-source current can flow. However, as the gate-source voltage is increased, the channel does not become inverted (i.e., n-region) suddenly, but rather gradually. Thus, it is useful to define three regions of channel inversion with respect to the gate-source voltage. In most circuit applications, noncutoff MOS-FET transistors are operated in strong inversion, with $V_{eff} > 100$ mV (many prudent circuit designers use a minimum value of 200 mV). As the name suggests, strong inversion occurs when the channel is strongly inverted. It should be noted that all the equation models in this section assume strong inversion operation. Weak inversion occurs when V_{GS} is approximately 100 mV or more below V_{tn} and is discussed as subthreshold operation in Section 1.3. Finally, moderate inversion is the region between weak and strong inversion.

Large-Signal Modelling

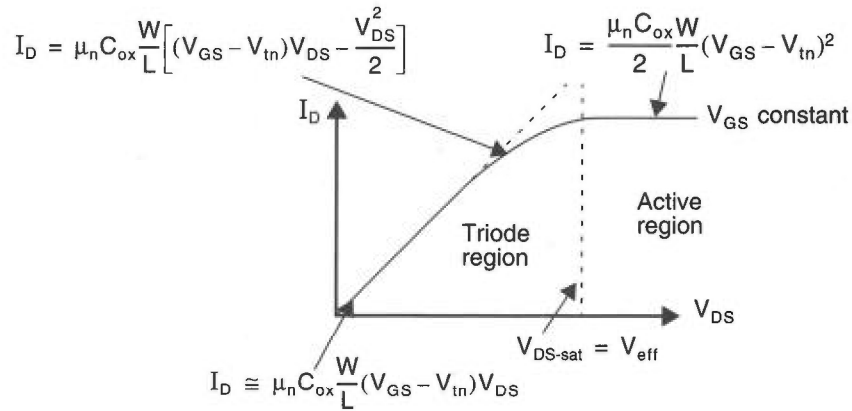The *triode region equation* for a MOS transistor relates the drain current to the gate-source and drain-source voltages. It can be shown (see Appendix) that this relationship is given by

$$I_D = \mu_n C_{ox}\left(\frac{W}{L}\right)\left[(V_{GS} - V_{tn})V_{DS} - \frac{V_{DS}^2}{2}\right] \qquad (1.65)$$

As V_{DS} increases, I_D increases until the drain end of the channel becomes pinched off, and then I_D no longer increases. This pinch-off occurs for $V_{DG} = -V_{tn}$, or approximately,

$$V_{DS} = V_{GS} - V_{tn} = V_{eff} \qquad (1.66)$$

Right at the edge of pinch-off, the drain current resulting from (1.65) and the drain current in the active region (which, to a first-order approximation, is constant with

respect to V_{DS}) must have the same value. Therefore, the *active region equation* can be found by substituting (1.66) into (1.65), resulting in

$$I_D = \frac{\mu_n C_{ox}}{2}\left(\frac{W}{L}\right)(V_{GS} - V_{tn})^2 \tag{1.67}$$

For $V_{DS} > V_{eff}$, the current stays constant at the value given by (1.67), ignoring second-order effects such as the finite output impedance of the transistor. This equation is perhaps the most important one that describes the large-signal operation of a MOS transistor. It should be noted here that (1.67) represents a squared current-voltage relationship for a MOS transistor in the active region. In the case of a BJT transistor, an exponential current-voltage relationship exists in the active region.

As just mentioned, (1.67) implies that the drain current, I_D, is independent of the drain-source voltage. This independence is only true to a first-order approximation. The major source of error is due to the channel length shrinking as V_{DS} increases. To see this effect, consider Fig. 1.15, which shows a cross section of a transistor in the active region. A pinched-off region with very little charge exists between the drain and the channel. The voltage at the end of the channel closest to the drain is fixed at $V_{GS} - V_{tn} = V_{eff}$. The voltage difference between the drain and the near end of the channel lies across a short depletion region often called the *pinch-off region*. As V_{DS} becomes larger than V_{eff}, this depletion region surrounding the drain junction increases its width in a square-root relationship with respect to V_{DS}. This increase in the width of the depletion region surrounding the drain junction decreases the effective channel length. In turn, this decrease in effective channel length increases the drain current, resulting in what is commonly referred to as *channel-length modulation*.

To derive an equation to account for channel-length modulation, we first make use of (1.11) and denote the width of the depletion region by x_d, resulting in

$$\begin{aligned} x_d &\cong k_{ds}\sqrt{V_{D\text{-}ch} + \Phi_0} \\ &= k_{ds}\sqrt{V_{DG} + V_{tn} + \Phi_0} \end{aligned} \tag{1.68}$$

where

$$k_{ds} = \sqrt{\frac{2K_s\varepsilon_0}{qN_A}} \tag{1.69}$$
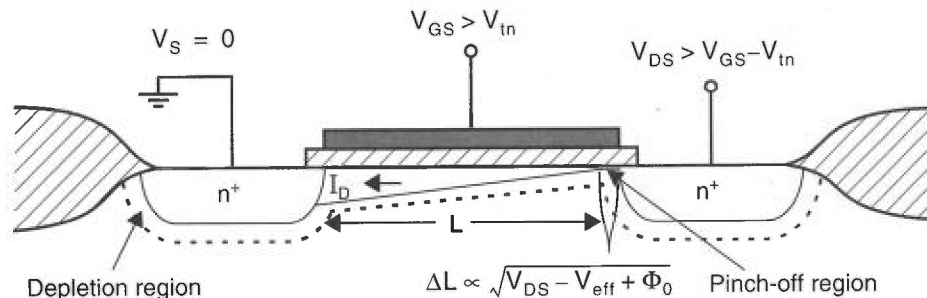
Fig. 1.15   Channel length shortening for $V_{DS} > V_{eff}$.

and has units of $m/\sqrt{V}$. Note that N_A is used here since the n-type drain region is more heavily doped than the p-type channel (i.e., $N_D \gg N_A$). By writing a Taylor approximation for I_D around its operating value of $V_{DS} = V_{GS} - V_{tn} = V_{eff}$, we find I_D to be given by

$$I_D = I_{D\text{-sat}} + \left(\frac{\partial I_D}{\partial L}\right)\left(\frac{\partial L}{\partial V_{DS}}\right)\Delta V_{DS} \cong I_{D\text{-sat}}\left(1 + \frac{k_{ds}(V_{DS} - V_{eff})}{2L\sqrt{V_{DG} + V_{tn} + \Phi_0}}\right) \tag{1.70}$$

where $I_{D\text{-sat}}$ is the drain current when $V_{DS} = V_{eff}$, or equivalently, the drain current when the channel-length modulation is ignored. Note that in deriving the final equation of (1.70), we have used the relationship $\partial L/\partial V_{DS} = -\partial x_d/\partial V_{DS}$. Usually, (1.70) is written as

$$I_D = \frac{\mu_n C_{ox}}{2}\left(\frac{W}{L}\right)(V_{GS} - V_{tn})^2[1 + \lambda(V_{DS} - V_{eff})] \tag{1.71}$$

where λ is the output impedance constant (in units of V^{-1}) given by

$$\lambda = \frac{k_{ds}}{2L\sqrt{V_{DG} + V_{tn} + \Phi_0}} = \frac{k_{ds}}{2L\sqrt{V_{DS} - V_{eff} + \Phi_0}} \tag{1.72}$$

Equation (1.71) is accurate until V_{DS} is large enough to cause second-order effects, often called *short-channel effects*. For example, (1.71) assumes that current flow down the channel is not *velocity-saturated* (i.e., increasing the electric field no longer increases the carrier speed). Velocity saturation commonly occurs in new technologies that have very short channel lengths and therefore large electric fields. If V_{DS} becomes large enough so short-channel effects occur, I_D increases more than is predicted by (1.71). Of course, for quite large values of V_{DS}, the transistor will eventually break down.

A plot of I_D versus V_{DS} for different values of V_{GS} is shown in Fig. 1.16. Note that in the active region, the small (but nonzero) slope indicates the small dependence of I_D on V_{DS}.
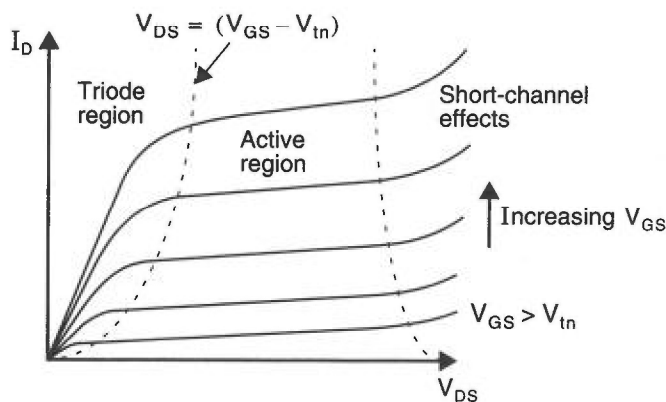

Fig. 1.16 I_D versus V_{DS} for different values of V_{GS}.