THE FIELD EFFECT TRANSISTOR (FET) – INTRODUCTION

FETs are a class of semiconductors that operate on an entirely different principle than BJTs. As the name field-effect implies, conduction is controlled by an electric field established by a voltage applied to the control terminal (the gate). However, unlike the BJTs, with the FET, current is conducted by only one type of carrier (electrons or holes) depending on the type of FET (n channel or p channel) hence, the FET is also referred to as the *unipolar transistor*. *Both BJTs and FETs, are equally important and each offers distinct advantages and has unique areas of application.]*

FETs are used in most computer circuits (ICs) because of several important advantages they have over BJTs. Compared to BJTs, FET transistors can be made quite small hence, occupying a small silicon area on an IC chip. Their manufacturing process is also relatively simple. Digital logic and memory functions can be implemented with circuits that exclusively use MOSFETs. For these reasons, most very-large-scale integrated (VLSI) circuits are made at present using MOS technology. examples include microprocessor and memory chips. MOS technology has also been applied extensively in the design of analog integrated circuits.
A particular kind of FET, the enhancement-type **MOSFET** is the most widely used field-effect transistor.

STRUCTURE AND PHYSICAL OPERATION OF THE ENHANCEMENT-TYPE METAL OXIDE SEMICONDUCTOR FIELD-EFFECT TRANSISTOR (MOSFET)

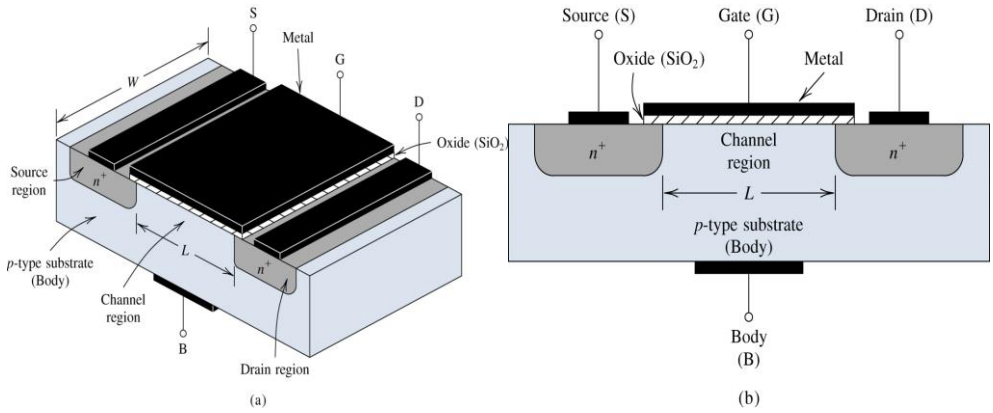


Fig.1 physical structure of enhancement type n-MOSFET. Physical structure of the enhancement-type NMOS transistor: (a) perspective view; (b) cross section. Typically L = 1 to 10 μm, W = 2 to 500 μm, and the thickness of the oxide layer is in the range of 0.02 to 0.1 μm.

Fig.1 shows the physical structure of the n-channel enhancement-type MOSFET. The transistor is fabricated on a p-type substrate (which is a single-crystal silicon wafer) that provides physical support for the device. Two heavily doped n-type regions, indicated by n^+ **source** and n^+ **drain** regions, are created in the substrate. A thin layer of silicon dioxide (SiO_2), is grown on the surface of the substrate, covering the area between the source and drain regions. Metal is deposited on top of the oxide layer to form the gate electrode of the device. Metal contacts are also made to the source region, the drain region and the substrate, also known as the body. The substrate forms pn junctions with the source and drain regions. In normal operation, these pn junctions are kept reverse-biased at all times. The two pn junctions (source-body and drain-body) are cut-off by connecting the body to the source. Thus the substrate (body) will be considered as having no effect on device operation and the Mosfet will be treated as a three-terminal device. The gate (G), the Source (S) and the drain (D).

OPERATION WITH NO GATE VOLTAGE:

With no bias voltage applied to the gate, two back-to-back diodes exist in series between drain and source. These diodes prevent current conduction from drain to source when a voltage V_{DS} is applied. The path between drain and source has a very high resistance of the order of $10^{12}\Omega$

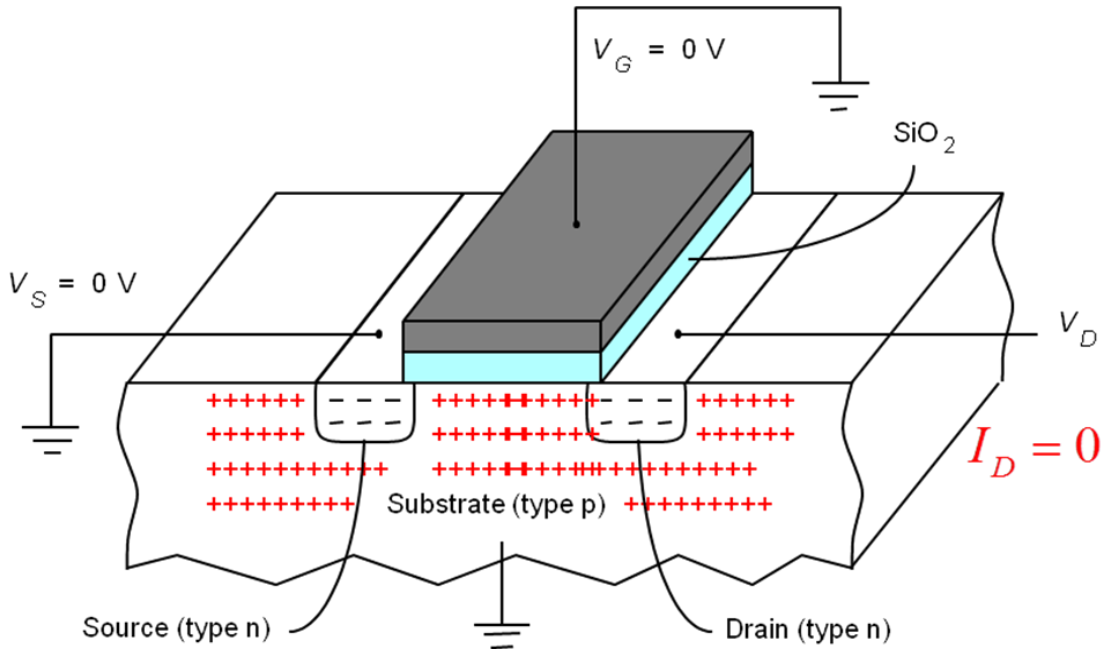


Figure 2: Mosfet operation with no gate voltage

CREATING A CHANNEL FOR CURRENT FLOW:

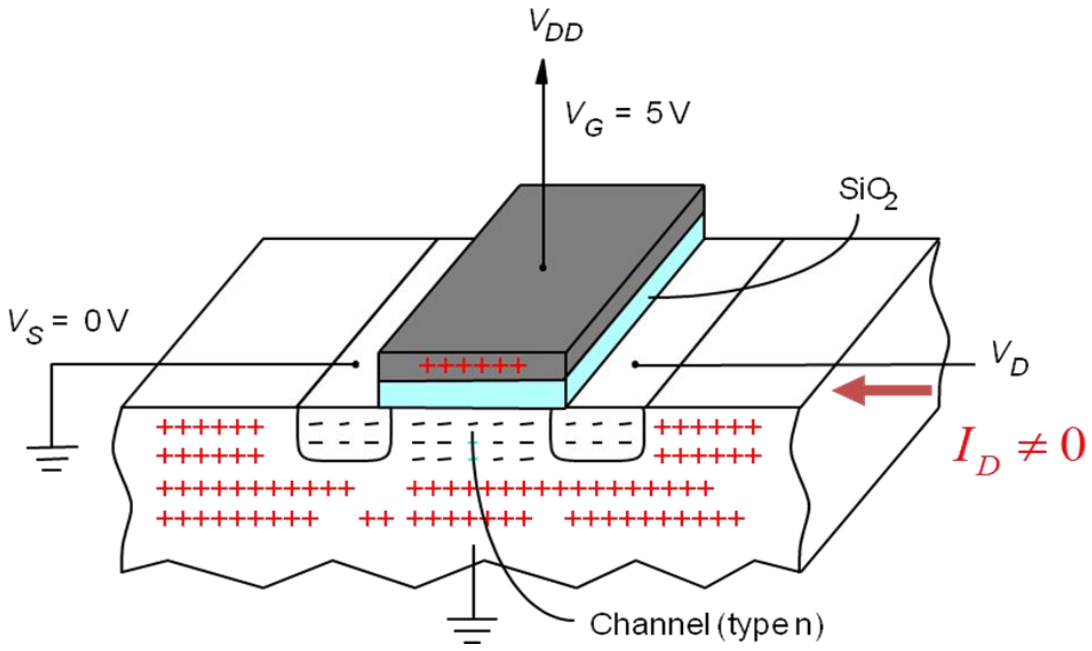


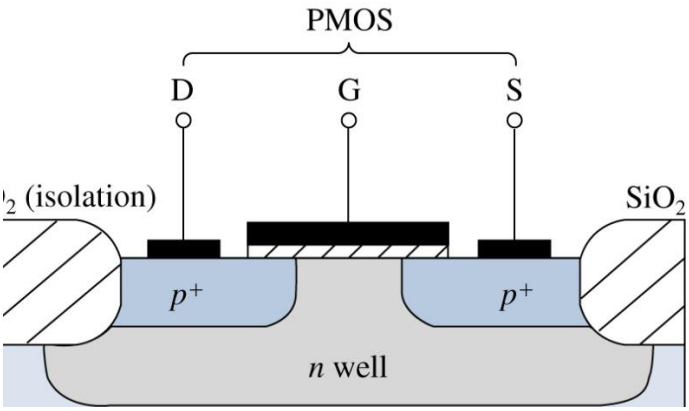
Fig. 3 Creating an n-channel

Consider the next situation depicted in fig. 3. We have grounded the source and drain and applied a positive voltage to the gate. *(Since the source is grounded the gate voltage appears in effect between gate and source and thus is denoted V_{GS}).* The positive gate voltage causes in the first instance the free holes to be repelled from the region of the substrate under the gate (the channel region). These holes are pushed downward into the substrate, leaving behind a carrier-depletion region. The positive gate voltage also attracts electrons from the n^+ source and drain into the channel region.

Now if, a voltage is applied between drain and source, current flows through this induced n-region, carried by the mobile electrons. The induced n region forms a channel for current flow from source to drain. Thus, the MOSFET of figure 3 is called an n-channel MOSFET or an NMOS transistor. [Note that an n channel MOSFET is created in a p type substrate. The induced channel is also called an inversion layer]. The value of V_{GS} at which a sufficient number of mobile electrons accumulate in the channel region to form a conducting channel is called the threshold voltage (Vt). In NMOS Vt is positive and lies between 1 and 3 V.

The positive gate voltage cause positive charges to accumulate on the top plate of the capacitor (the gate electrode) and the corresponding negative charge of the bottom plate is formed by the electrons in the induced channel. The charges which accumulate on the gate cannot cross into the channel region because of the insulating effect of the oxide layer. An electric field thus develops in the vertical direction. It is this field that controls the amount of charge in the channel, and thus it determines the channel conductivity and, in turn, the current that will flow through the channel when a voltage V_{DS} is applied.

The PMOS transistor



A p-channel enhancement-type MOSFET (PMOS transistor) is fabricated on an n-type substrate with p+ regions for the drain and source, and holes as charge carriers. The device operates just like the n-channel device except that V_{DS} and V_{GS} are negative and Vt is also negative. The current i_D enters the source terminal and leaves through the drain.

PMOS versus NMOS

- NMOS devices can be made smaller
- They operate faster
- NMOS requires lower supply voltages
- Both NMOS and PMOS are utilized in the design of CMOS

Complementary MOS technology (CMOS) employs transistors of both polarities.

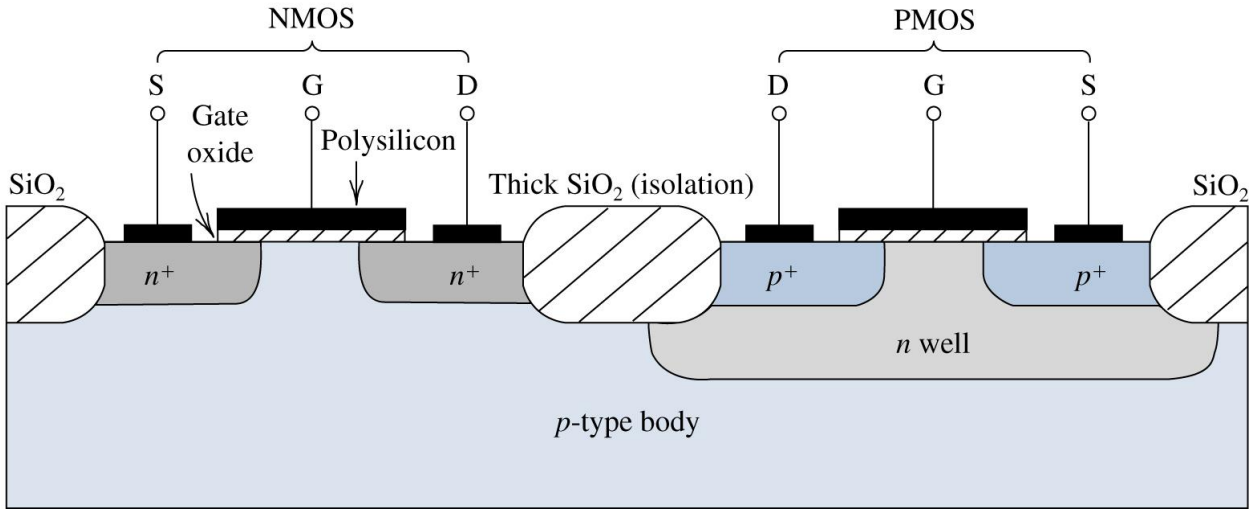


Fig. 4. Cross section of a CMOS integrated circuit. Note that the NMOS is formed in a separate p-type region, known as a p well. Another arrangement is also possible in which a p-type body is used and the p device is formed in an n well.

APPLYING A SMALL V_{DS}

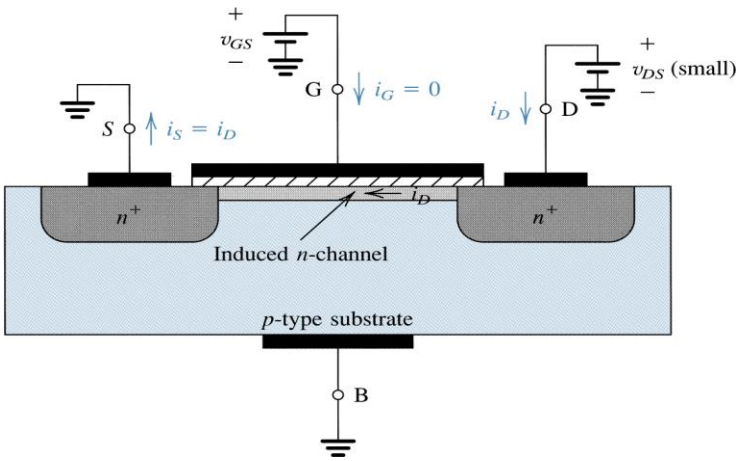


Figure 5: An NMOS transistor with $v_{GS} > V_t$ and with a small v_{DS} applied

Consider the NMOS shown in figure 5, applying a positive voltage V_{DS} between drain and source, causes a current i_D to flow through the induced n-channel. (Current is carried by free electrons flowing from source to drain). The magnitude of i_D depends on the density of electrons in the channel, which in turn depends on the magnitude of V_{GS}.

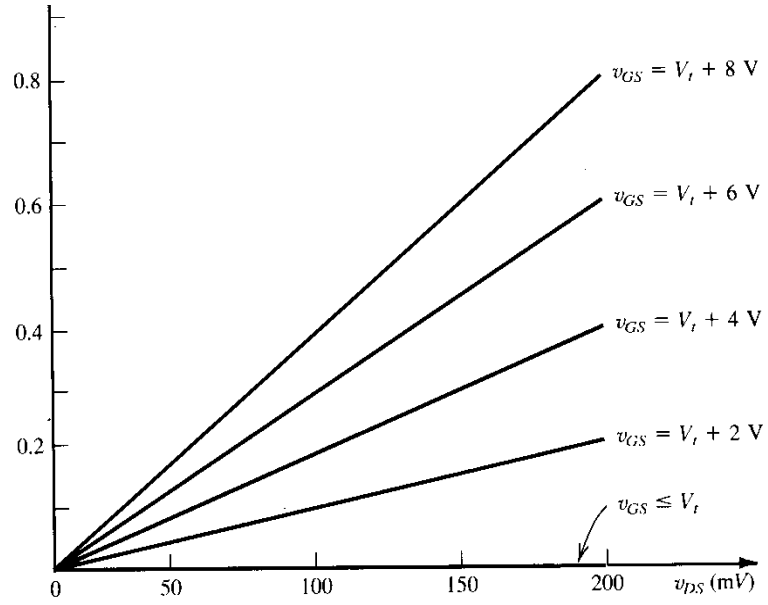


Fig. 5.4 The i_D–v_{DS} characteristics of the MOSFET in Fig. 5.3. Note that v_{DS} is small. The device operates as a linear resistor whose value is controlled by v_{GS}.

Figure 5.4. Shows a graph of i_D against V_{DS} for various values of V_{GS}. We observe that the MOSFET is operating as a linear resistance whose value is controlled by V_{GS}. The resistance is infinite for $V_{GS} \leq V_t$ and its value decreases as V_{GS} exceeds V_t.

For $V_{GS} = Vt$, the channel is just induced, and the current conducted i_D is negligibly small. As V_{GS} exceeds Vt more electrons are attracted into the channel and the result is a channel of increased conductance. The conductance of the channel is proportional to the excess gate voltage (V_{GS} - Vt). It follows that i_D will be proportional to V_{GS} - Vt and, of course, the voltage V_{DS} that causes i_D to flow.

From the description above, one can say that for the MOSFET to conduct, a channel must be induced. Then increasing V_{GS} above the threshold voltage Vt enhances the channel, hence the names enhancement-mode operation and enhancement-type MOSFET. Finally, we note that the current that leaves the source terminal (i_S) is equal to the current that enters the drain terminal i_D, and the gate current $i_G = 0$

OPERATION AS V_{DS} IS INCREASED:

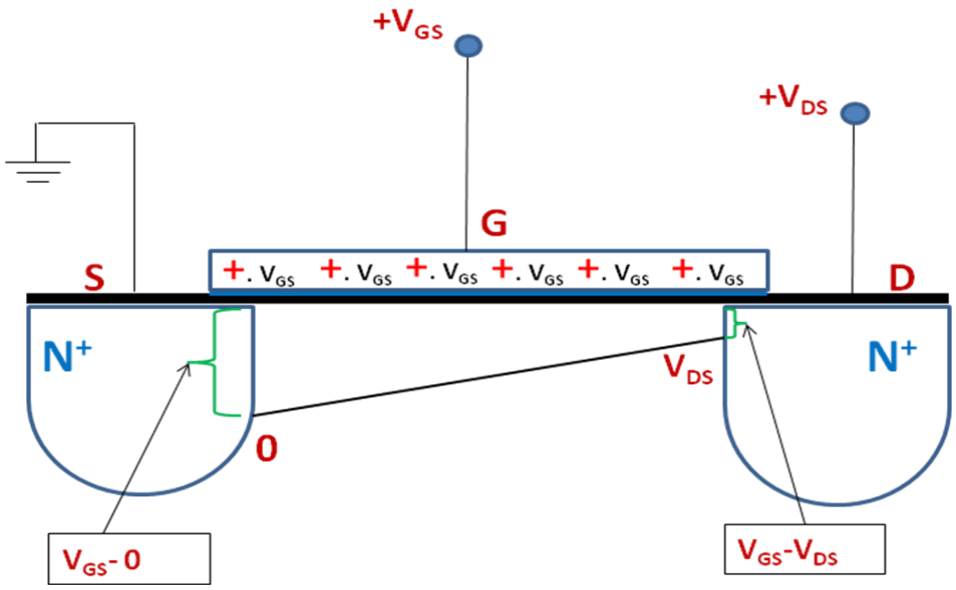


Fig. 6a Operation as V_{DS} is Increased

let V_{GS} be held constant and greater than V_t and therefore a channel is formed, as in figure 6. V_{DS} appears as a voltage drop across the length of the channel. That is as we travel along the channel from source to drain, the voltage (measured relative to the source) increases from 0 to V_{DS}. The voltage between the gate and points along the channel decreases from V_{GS} (at the source end) to V_{GS} - V_{DS} at the drain end. Since the channel depth depends on this voltage, we find that the channel is no longer of uniform depth; rather, the channel takes the tapered form. For example, let's take V_{GS} to be 10 V and V_{DS} to be 5 V. at the source end of the mosfet, the channel depth would be equivalent to 10 – 0 = 10, whilst at the drain end the channel depth would be equivalent to 10 – 5 = 5. As V_{DS} is increased, the channel becomes more tapered and its resistance increases correspondingly.

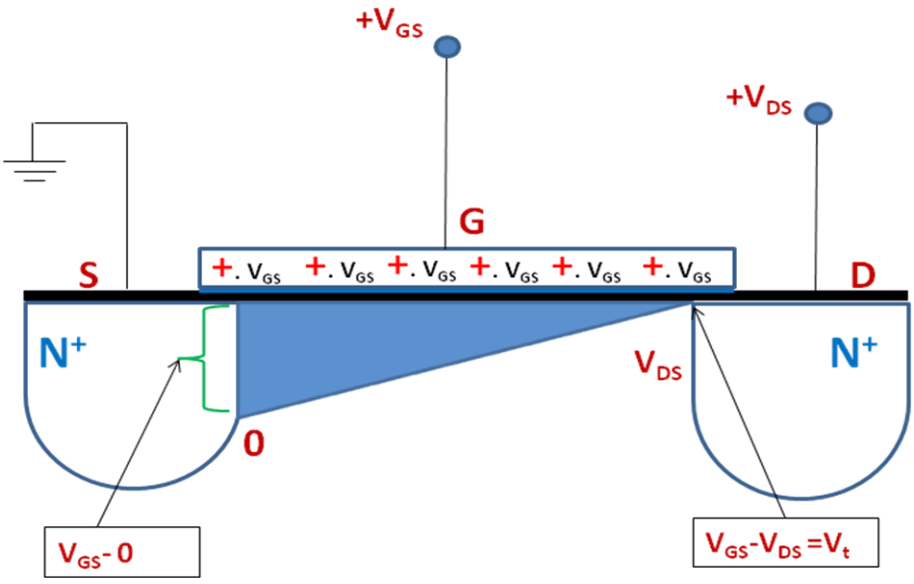


Figure 6b: Pinched-off channel

Eventually at V_{GS} - V_{DS} = V_t (or V_{DS} = V_{GS} - V_t) the channel depth at the drain end decreases to almost zero and the channel is said to be pinched-off. Increasing V_{DS} beyond this value has little effect (theoretically, no effect) on the channel shape, and the current through the channel remains constant at the value reached for V_{DS} (at which, V_{GS} - V_{DS} = V_t). The drain current thus saturates at this value, and the MOSFET is said to have entered the saturation region of operation.

The voltage V_{DS} at which saturation occurs is denoted V_{DS} sat

$$\mathbf{V_{DS\ sat} = V_{GS} - V_t}$$

Obviously for every value of $V_{GS} \geq V_t$, there is a corresponding value of V_{DS} sat. The device operates in the saturation region if $V_{DS} \geq V_{DS}$ sat.
The region of the i_D - V_{DS} characteristic obtained for $V_{DS} < V_{DS}$ sat , sat is called the triode region.

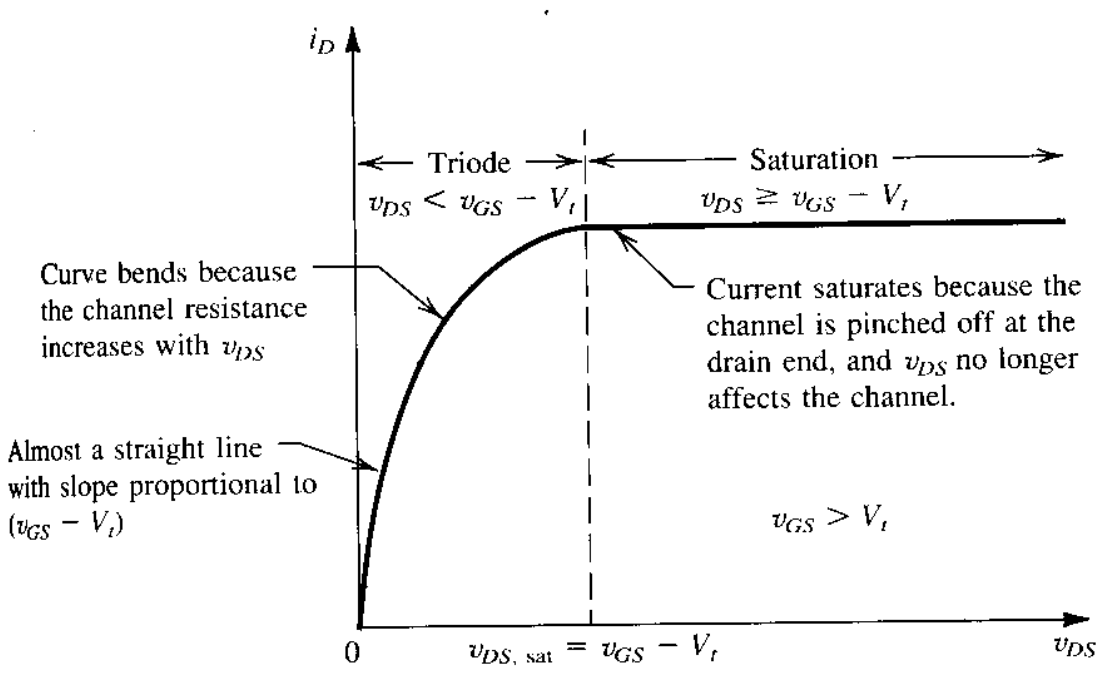
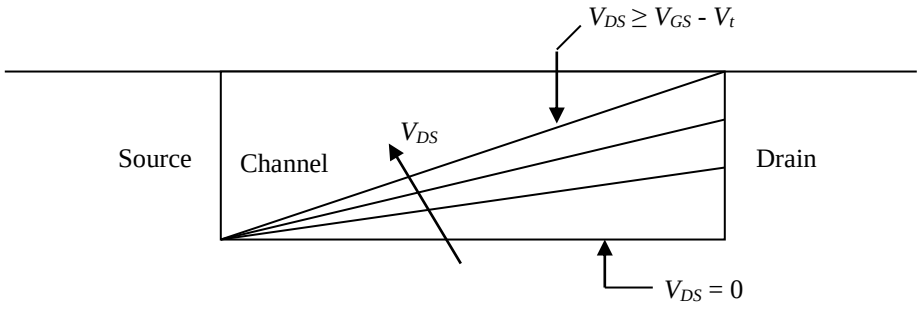


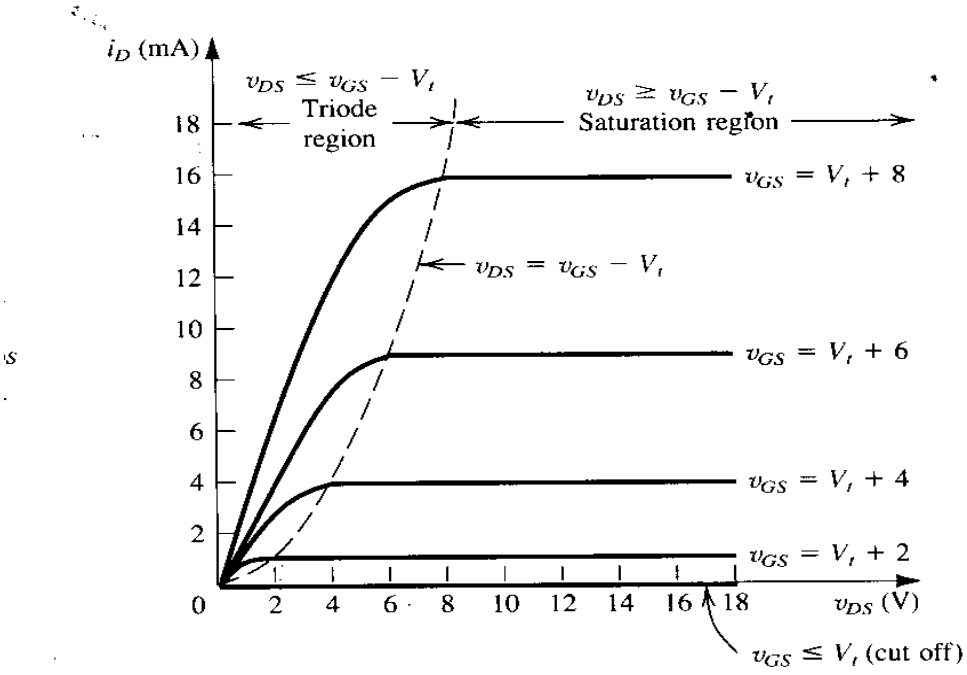
Figure 7

To help further in visualizing the effect of V_{DS}, we show in fig. 8 sketches of the channel as V_{DS} is increased as U_{GS} remains constant.



Increasing V_{DS} causes the channel to acquire a tapered shape. Eventually, as V_{DS} reaches $V_{GS} - V_t$, the channel is pinched off at the drain end. Increasing V_{DS} above $V_{GS} - V_t$ has little effect (theoretically no effect) on the channel's shape.

EQUATIONS FOR DEVICE OPERATION



(b)

The i_D - V_{DS} characteristics, shown in fig.5.10b. are a family of curves, each measured at a constant $V_{GS.}$ There are 3 distinct regions of operation: the cut off, the triode region and the saturation region.

Saturation region - amplifier operation

Cut off
Triode - switch operations

The device is cut off when $U_{GS} < V_t$.

TRIODE REGION OF OPERATION (NMOS)

To operate the MOSFET in the triode region we must first induce a channel

$V_{GS} \geq V_t$ (Induced channel)

and keep V_{DS} small enough so that the channel remain continuous. This is achieved by ensuring that the gate-to-drain voltage is

$V_{GD} > V_t$ (continuous channel). ……*

This condition can be stated explicitly in terms of V_{DS} by writing:

$$\mathbf{V_{DS} < V_{GS} - V_t \text{ (continuous channel)}\ldots\ldots\ldots **}$$

Either equation ***** or ****** Can be used to ascertain triode - region operation. In words, the n-channel enhancement-type MOSFET operates in the triode region when U_{GS} is greater than V_t and the drain voltage is lower than the gate voltage by at least V_t volts. Or put simply; When the potential difference between gate and drain is greater than V_t, the MOSFET is in triode region

In the triode region, the i_D-V_{DS} characteristics can be approximately described by the relationship:

$$\mathbf{i_D = K[2(V_{GS} - V_t) V_{DS} - V_{DS}^2] \ldots\ldots\ldots(1)}$$

in which K is a device parameter given by;

$$\mathbf{K = \tfrac{1}{2}\, \mu_n C_{ox}(W/L) \quad (Units\ A/V^2).}$$

μ_n - physical constant known as the Electron Mobility. Its value in this case applies for the electrons in the induced n-channel

C_{ox} - Oxide capacitance; is the capacitance per unit area of the gate-to-body capacitor for which oxide layer serves as dielectric.

L - length of channel
W - width

(Since for a given fabrication process the quantity $\mathbf{\tfrac{1}{2}\, \mu_n C_{ox}}$ *is a constant (approximately* $10\mu A/V^2$ *for the standard NMOS process with a 0.1 -*μ*m oxide thickness), the aspect ratio of the device (W/L) determines the conductivity parameter K).*

If V_{DS} is sufficiently small so that we can neglect V^2_{DS} term in equation 1. We obtain for the i_D - V_{DS} characteristic near the origin the following relationship;

$$\mathbf{i_D = 2K\,(V_{GS} - V_t)\, V_{DS}}$$

this linear relation represents the operation of the MOS transistor as a linear resistance r_{DS}

$$\mathbf{r_{DS} = \frac{V_{DS}}{i_D} = [2K\,(V_{GS} - Vt)]^{-1}}$$

Whose value is controlled by V_{GS}.

SATURATION REGION OF OPERATION (NMOS)

To operate the MOSFET in the saturation region, a channel has to be induced;

$$\mathbf{V_{GS} \geq V_t \text{ (induced channel)}}$$

And pinched of at the drain end by raising V_{DS} to a value that results in the gate to drain voltage falling below Vt,

$$\mathbf{V_{GD} \leq V_t \text{ (pinched of channel)}}$$

This condition can be expressed explicitly in terms of V_{DS} as;

$$\mathbf{V_{DS} \geq V_{GS} - V_t \text{ (pinched of channel)}}$$

Thus, the n-channel enhancement type MOSFET operates in the saturation region when U_{GS} is greater than V_t. and the drain voltage does not fall below the gate voltage by more than Vt. Volts. Or when the potential difference between gate and drain becomes equal to or less than V_t, the MOSFET enters saturation region.
The boundary between the triode region and the saturation region is characterized by;

$$\mathbf{V_{DS} = V_{GS} - Vt.}$$

Substituting this into equation (1), gives the saturation value of the current i_D as;

$$\mathbf{i_D = K(V_{GS} - Vt)^2 \ldots\ldots\ldots\ldots\ldots\ldots 3}$$

thus, in saturation, the MOSFET provides a drain current whose value is independent of the voltage V_{DS} and is determined by the gate voltage V_{GS} according to the square law relationship shown in the equation (3) above.

EQUATIONS FOR A P-CHANNEL MOSFET

For a p-channel device the threshold voltage Vt. Is negative. To induce a channel we apply a gate voltage that is more negative than Vt.

$V_{GS} = \leq$ Vt. (induced channel)

And apply a drain voltage that is more negative than the source voltage (i.e. V_{DS} is negative or equivalently V_{SD} is positive). The current i_D flows out of the drain terminal. To operate in the triode region V_{DS} must satisfy;

$$V_{DS} \geq V_{GS} - Vt. \text{ (continuous channel)}$$

That is the drain voltage must be higher than the gate voltage by at least |Vt.|. The current i_D is given by the same equation as for NMOS,

$$\mathbf{i_D = K[2(V_{GS} - Vt)\, V_{DS} - V_{DS}^2]}$$

where V_{GS}, Vt. And V_{DS} are negative and K is given by;

$$\mathbf{K = \tfrac{1}{2}\, \mu_p C_{ox}(W/L)}$$

Where μ_p is the mobility of holes in the induced p-channel. Typically,

$$\mathbf{\mu_p = \tfrac{1}{2}\mu_n}$$

with the result that for the same W/L ratio, a PMOS transistor has half the value of K as the NMOS.

To operate in saturation V_{DS} must satisfy;

$V_{DS} \leq V_{GS} - Vt.$ (pinched of channel)

The current i_D is given by the same equation used for NMOS;

$$\mathbf{i_D = K(V_{GS} - Vt)^2}$$

Where V_{GS} and Vt. Are negative.

Channel length modulation

When we say i_D is independent of V_{DS} in saturation and therefore the output resistance at the drain is infinite, we are talking about an ideal case. In practice, increasing V_{DS} above V_{DS}sat moves the channel pinch-off point slightly away from the drain toward the source. Thus the effective channel length is reduced. The slight linear dependence of i_D on V_{DS} in the saturation region can be analytically accounted for by incorporating the function $1 + \lambda V_{DS}$ in the i_D equation as follows:

$$\mathbf{i_D = K(V_{GS} - Vt.)^2(1 + \lambda V_{DS})}$$

where $V_A = 1/\lambda$

the channel length modulation makes the output resistance in saturation finite. Thus incremental output resistance in saturation, ro is;

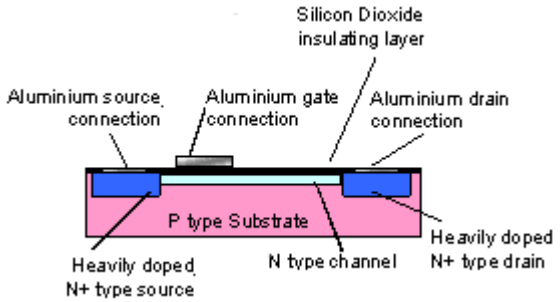
$$\mathbf{r_o = [\lambda K(V_{GS} - V_{t.})^2]^{-1}}$$

$$\mathbf{r_o = [\lambda I_D]^{-1}}$$

Where I_D is the current corresponding to the particular value of U_{GS} for which r_o is being evaluated. Sometimes $r_o = V_A/I_D$.

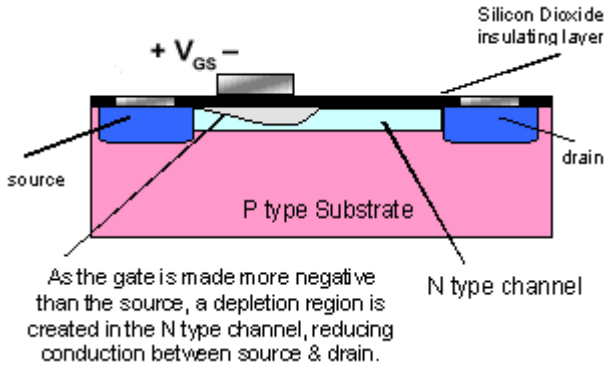
NB. r_{DS} is the drain-to-source resistance in the triode region for small V_{DS}.

THE DEPLETION TYPE MOSFET



The depletion mode MOSFET shown as an N channel device (P channel is also available) in Fig is more usually made as a discrete component, i.e. a single transistor rather than IC form. In this device a thin layer of N type silicon is deposited just below the gate−insulating layer, and forms a conducting channel between source and drain.

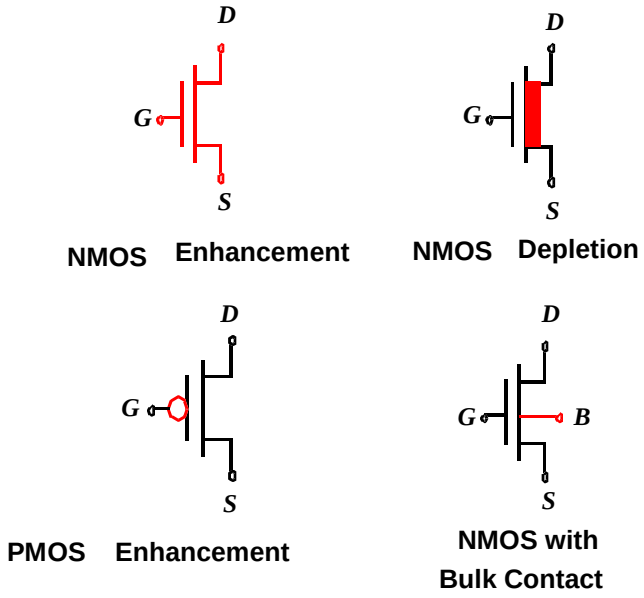
In other words, there is no need to induce a channel, unlike the case of the enhancement MOSFET. Therefore when the gate source voltage V_{GS} is zero, current (in the form of free electrons) can flow between source and drain. The channel depth and hence its conductivity can be controlled by V_{GS} as in the enhancement type device. Applying a positive V_{GS}, enhances the channel by attracting the electrons into it, while a negative V_{GS} repels the electrons from the channel; and thus the channel becomes shallower and its conductivity decreases.



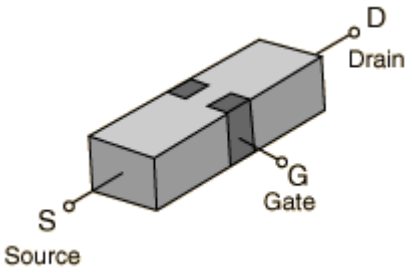
The negative V_{GS} is said to deplete the channel of its charge carriers and this mode of operation is called depletion mode. As the magnitude of V_{GS} is increased in the negative direction, a value is reached at which the channel is completely depleted of charge carriers and i_D is reduced to zero even though V_{DS} may still be applied. This negative value of V_{GS} is the threshold voltage of the n-channel depletion type MOSFET.
The current-voltage characteristics of the depletion-type MOSFET are described by the equations given in the previous section for the enhancement device except that, for an n-channel depletion device, Vt. Is negative. Another parameter for the depletion MOSFET is the value of drain current obtained in saturation with V_{GS}=0. It is denoted, I_{DSS}. It can be shown that

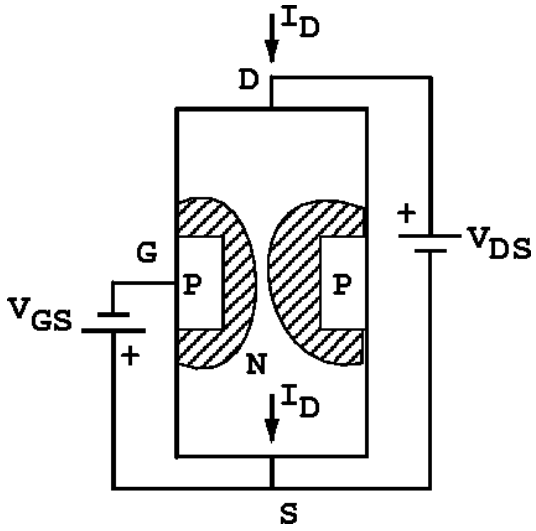
$$\mathbf{I_{DSS} = KVt^2}$$

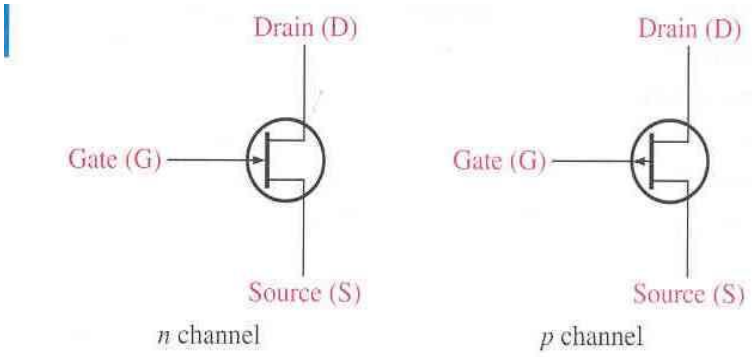
MOS transistor types and symbols



THE JUNCTION FIELD EFFECT TRANSISTOR (JFET)







(a) basic structure of n-channel JFET (b) Circuit symbol for the n-channel JFET (c) Circuit symbol for the p-channel JFET

PHYSICAL OPERATION:

In a junction field-effect transistor, or JFET, the controlled current passes from source to drain, or from drain to source as the case may be. The controlling voltage is applied between the gate and source. Note how the current does not have to cross through a PN junction on its way between source and drain: the path (called a *channel*) is an uninterrupted block of semiconductor material. In the image just shown, this channel is an N-type semiconductor. P-type channel JFETs are also manufactured:

Lets consider an n-channel JFET with V_{GS}=0, the application of a voltage V_{DS} causes current to flow from the drain to the source. When a negative V_{GS} is applied the depletion region of the gate-channel junction widens and the channel becomes correspondingly narrower. Thus the channel resistance increases and the current i_D (for a given V_{DS}) decreases. Because V_{DS} is small, the channel is almost of uniform width. The JFET is simply operating as a resistance whose value is controlled by V_{GS}. If we keep increasing V_{GS} in the negative direction, a value is reached at which the depletion region occupies the entire channel. At this value of V_{GS} the channel is completely depleted of charge carriers (electrons); the channel has in effect disappeared. This value of V_{GS} is therefore the threshold voltage of the device Vt., which is obviously negative for an n-channel JFET.

Current-Voltage Characteristics

For a JFET the saturation region is usually called the pinch-off region, and the threshold voltage is usually called the pinch-off voltage and is denoted by, Vp, thus Vp=Vt. Manufacturers usually specify the value of drain current at the onset of saturation for U_{GS}=0, denoted I_{DSS} instead of the conductance parameter K. It can be easily shown that;

$$\mathbf{I_{DSS} = KVt^2 = KVp^2}$$

The JFET characteristics can be described by the same equations used for MOSFETs. Specifically, re labeling Vt. As Vp, we can write;

The n-channel JFET will be cut off for; $V_{GS} <\_ Vp$ where Vp is negative.

To turn the device on, we apply a gate-to-source voltage U_{GS}.

$$\mathbf{Vp < V_{GS} <\_ 0}$$

And a positive drain-to source voltage V_{DS}. The JFET operates in the triode region for;

$$\mathbf{V_{DS} <\_ V_{GS} - Vp}$$

In which case the drain current is given by;

$$\mathbf{i_D = K[2(V_{GS} - Vp)V_{DS} - V_{DS}^2]}$$

this can also be expressed below if Vp^2 is extracted from the square brackets and we replace KVp^2 by I_{DSS}.

$$\mathbf{i_D = I_{DSS}[2(1 - V_{GS}/Vp)(V_{DS}/-Vp) - (V_{DS}/Vp)^2]}$$

The JFET operates in saturation (pinch-off) for; $V_{DS} \geq V_{GS} - Vp$

In words, for the JFET to operate in pinch-off, the drain voltage must be greater than the gate voltage by at least |Vp|. In pinch-off, the drain current is given by

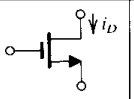
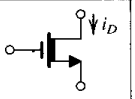
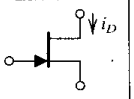
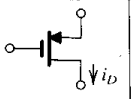
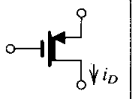
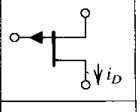
$$\mathbf{i_D = I_{DSS}(1 - V_{GS}/Vp)^2(1 + \lambda V_{DS})}$$

Where $\lambda \equiv 1/V_A$ is a positive constant included to account for the dependence of i_D on V_{DS} in a pinch-off.

The P-channel JFET

The equations are the same, however, Vp is positive, ($0 <\_ U_{GS} <\_ Vp$). V_{DS} is negative λ and V_A are negative and the current i_D flows out of the drain terminal. To operate the p-channel JFET in pinch-off, $V_{DS} <\_ U_{GS} - Vp$, which in words means that the drain voltage must be lower than the gate voltage by at least |Vp|. Otherwise with $U_{DS} \geq U_{GS} - Vp$, the p-channel JFET operates in the triode region.

Table 5.1 SUMMARY OF THE CHARACTERISTICS OF FIELD-EFFECT TRANSISTORS

	n-Channel			*p*-Channel		
	Enhancement MOSFET	Depletion MOSFET	JFET	Enhancement MOSFET	Depletion MOSFET	JFET
Circuit Symbol						
V_t	+	−	−	−	+	+
K	$\frac{1}{2}\mu_n C_{OX}(W/L)$		I_{DSS}/V_P^2	$\frac{1}{2}\mu_p C_{OX}(W/L)$		I_{DSS}/V_P^2
To turn transistor on	$v_{GS} > V_t$			$v_{GS} < V_t$		
v_{DS}	+			−		
To operate in the triode region	$v_{DS} \leq v_{GS} - V_t$			$v_{DS} \geq v_{GS} - V_t$		
To operate in the saturation region	$v_{DS} \geq v_{GS} - V_t$			$v_{DS} \leq v_{GS} - V_t$		
$\lambda = 1/V_A$	+			−		
In triode region	$i_D = K[2(v_{GS} - V_t)v_{DS} - v_{DS}^2]$					
In saturation region	$i_D = K(v_{GS} - V_t)^2(1 + \lambda\, v_{DS})$					
r_o	$\lvert V_A\rvert/I_D$					