Lecture-13
Metal Oxide Semiconductor FET (MOSFET)

Introduction: Metal-oxide-semiconductor field-effect transistors (MOSFETs) have played a major role in the development of complex large scale integrated circuits. In particular, they have provided the basis for most large scale integrated digital circuits. The MOSFET operation depends on the conductance modulation of the channel of carriers that are induced by an applied gate voltage. Modulation is achieved by the variation in the carrier surface density. In the JFET on the other hand, conductance modulation is accomplished by the variation of the cross-sectional area of the channel formed from the *pn* junction. FETs are also known as unipolar devices because conduction is by means of one type of carrier only. One very important application of MOSFETS is in the arrangement known as a complementary metal oxide semiconductor system (CMOS). The CMOS forms, at present, the mainstream of high density digital system design technology.
There are two types of MOS transistors. The *depletion* MOSFET has a behavior similar to that of the JFET; at zero voltage and a fixed drain voltage, the current is a maximum and then decreases with the applied gate potential as in Fig.6 of LN-11. The second kind of device, called the *enhancement* MOSFET, exhibits no current at zero gate voltage and the magnitude of the output current increases with an increase in the magnitude of the gate potential. Both types can exist in either the *p*-channel or *n*-channel variety. We consider the characteristics of an *n*-channel for the subsequent discussions.

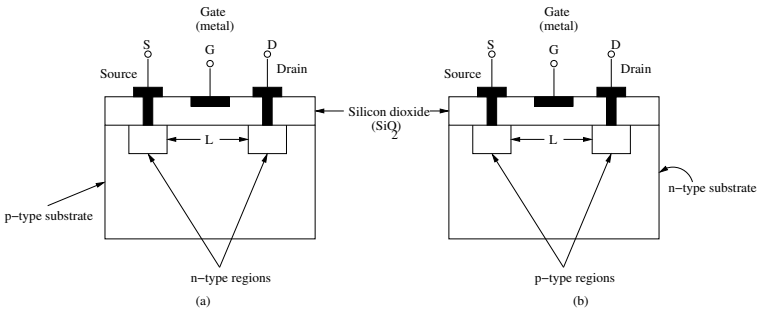

Figure 1: Enhancement (MOSFET) structures: (a) *n*-channel; (b) *p*-channel

The Enhancement MOS Structure: A simplified form of the structure of an *n*-channel enhancement MOSFET is shown in the Fig.1(a) and that for a *p*-channel device, in the Fig.1(b). The devices depicted in the Fig.1 are commonly referred

to as *NMOS* and *PMOS transistors*. As indicated in the Fig.1(a), the two *n*-type regions embedded in the *p*-type substrate (the body) are the source and drain electrodes. The region between source and drain is the channel, which is covered by the thin silicon dioxide (SiO_2) layer. The gate is formed by the metal electrode played over the oxide layer. At present, MOSFET fabrication technology utilizes a polysilicon[1] conducting layer for the gate rather than the metal gate displayed in the Fig.1. The physical principles which govern MOSFET operation, however, are the same for both types of the gate.

The metal are of the gate, in conjunction with the insulating dielectric oxide layer and the semiconductor channel, form a parallel-plate capacitor. The insulating layer of the silicon dioxide is the reason why this device is also called the *insulated-gate field-effect transistor* (IGFET). This layer results in an extremely high input resistance (10^{10} to 10^{15} Ω) for the MOSFET.

Physical Behavior of the Enhancement MOSFET: In the Fig.2(a) we show an NMOS transistor in which the source and substrate are grounded and the drain-to-source voltage V_{DS} is set to zero. The positive voltage applied to the gate establishes an electric field which is directly perpendicularly through the oxide. This field will end on "induced" negative charges near the semiconductor surface, as shown in the Fig.2(a). Since the *p*-type substrate contains very few electrons, the positive surface charges are primarily electrons obtained from the *n*-type source-to-drain. These mobile negative charges, which are minority carriers in the *p*-type substrate, form an "inversion layer". Such an inversion layer is formed only if V_{GS} exceeds a threshold voltage V_T.[2] The induced chargers beneath the oxide constitute an *n*-channel. As the voltage on the gate increases beyond V_T, the number of induced negative charges in the semiconductor increases. Consequently, the conductivity of the channel increases. Application of a positive potential between the drain and source produces a current in the induced channel between drain and source. Thus the drain current is enhanced by the positive gate voltage and the device is called *enhancement-type MOSFET*.

Let us consider the situation where V_{DS} is increased from zero with V_{GS} maintained at a constant positive value greater than V_T (that is, $V_{GS}-V_T > 0$). For small values of V_{DS} ($V_{DS} < V_{GS}-V_T$), an increase in V_{DS} is accompanied by an increase in drain current I_D. The behavior of the MOSFET is that of resistance, and this region is referred to as the *ohmic region*. As V_{DS} increases, the drop across the gate oxide at the drain side of the channel $V_{GD} = V_{DS} - V_{GS}$ decreases. This reduced potential difference lowers the field across the drain end of the dielectric, which results in

[1]Polysilicon refers to doped silicon in which the individual parts of the crystalline structure are randomly oriented in space. The behavior of polysilicon is similar that of a metal

[2]In this topic the threshold voltage should not be confused with the volt-equivalent of temperature

fewer inversion charges in this region portion of induced channel. The channel is being "pinched off," and I_D increases much more slowly with respect to increases in V_{DS} than in the ohmic region near the origin. Ideally once pinch-off is achieved, a further increase in V_{DS} produces no change in I_D and current saturation exists. This *saturation region* is similar in nature to velocity saturation in the JFET. The value of I_D attained in saturation depends on the value of V_{GS}. Increases in $V_{GS} > V_T$ result in increasing saturation values of I_D.

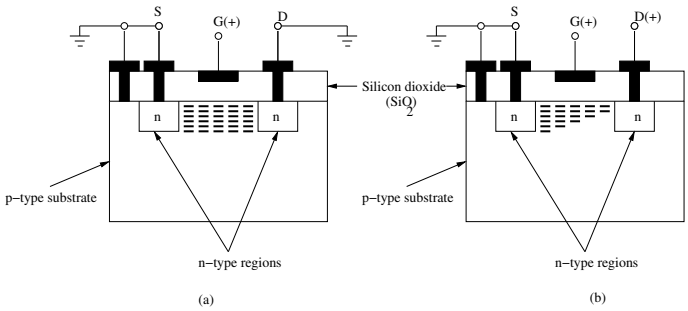

Figure 2: Biased NMOS enhancement transistor showing induced channel with (a) $V_{DS} = 0$ and (b) $V_{DS} > 0$

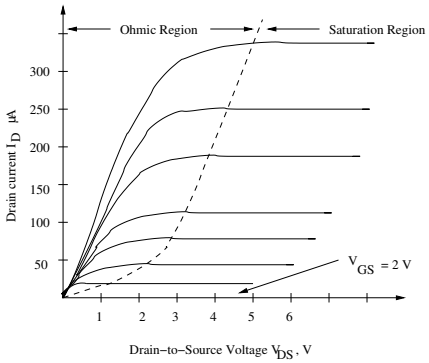

Figure 3: Enhancement NMOS output characteristics

The Enhancement MOSFET Volt-Ampere Characteristics: Manufacturers of IC MOS transistors do nor provide curves of the volt-ampere characteristics. If needed or desired, these curves are generated from the analytical expressions for MOSFET behavior in each region of operation.

Analytical Expressions for the volt-ampere characteristics: An inversion channel exists between source and drain, with $V_{DS} = 0$, only if $V_{GS} > V_T$. For

$V_{GS} < V_T$, there are no mobile carriers at the drain end of the channel and $I_D = 0$. Thus V_T is analogous to the pinch-off voltage in a JFET. The condition that $V_{GS} < V_T$ and $I_D = 0$ signifies that the MOSFET is a *cut off* and corresponds to an open switch.

1. **Ohmic Region:** As described in the previous section, for $V_{GS} > V_T$, the channel conductivity is controlled by V_{DS} in the *ohmic* also called *nonsaturation* or *triode* region. More precisely, the ohmic region is defined by $V_{GS} - V_T > V_{DS}$ (or $V_{GD} = V_{GS} - V_{DS} > V_T$). Theoretical analysis[3] of the ohmic region leads to the result that the drain characteristic is given by

$$I_D = k\left(\frac{W}{L}\right)[2(V_{GS} - V_T)V_{DS} - V_{DS}^2] \tag{1}$$

   where L is the channel length, W the channel width (perpendicular to L) and k is the process parameter in $\mu A/V^2$. The process parameter $k = \mu_n C_o/2$, where μ_n is the electron mobility and C_o is the gate capacitance per unit area (and equals ϵ/T_{ox}, the ratio of permittivity and thickness of the oxide layer). Of note is that V_T also depends on C_o as well as the doping densities of the n-type drain and source and p-substrate.

2. **Saturation Region:** Ideally, I_D is constant and independent of V_{DS} in the saturation region for which $V_{GS} - V_T < V_{DS}$ (but greater than zero). The value of I_D depends only on the effective control voltage $V_{GS} - V_T$ as given below,

$$I_D = k\left(\frac{W}{L}\right)(V_{GS} - V_T)^2 \equiv I_{DS} \tag{2}$$

   where the subscript S added to I_D denotes that the drain current in the saturation region is under consideration. The dividing line between the ohmic region and saturation regions is given by $V_{GS} - V_T = V_{DS}$. Substitution of this value of V_{DS} in Eqn.(1) results in Eqn.(2). The dashed curve in Fig.3, which indicates that boundary between the ohmic and saturation regions, is given by

$$I_D \equiv k\left(\frac{W}{L}\right)V_{DS}^2 \tag{3}$$

Several observations concerning the expressions in the Eqn.(1) and (2) are noteworthy. First, the *aspect ratio* W/L is an important parameter as it serves as a scale factor for the drain current. Thus two (or more) MOSFETs having the same value of V_T but with different current capabilities can be fabricated on the same chip by

[3]Derivation of these expressions will be done in the later lectures

using two (or more) different values of W/L. Second, the parameter k has typical values which lie in the range of 10 to 50 $\mu A/V^2$ in present commercial NMOS processes. Consequently high values of I_D (several milliamperes) are obtainable only in devices with high W/L ratio that is devices which consume a large area.
The MOSFET transfer characteristic is plot of I_D versus V_{GS} at constant V_{DS} in the saturation region. The curve in Fig.4 is the transfer characteristic for the MOSFET given in Fig.3

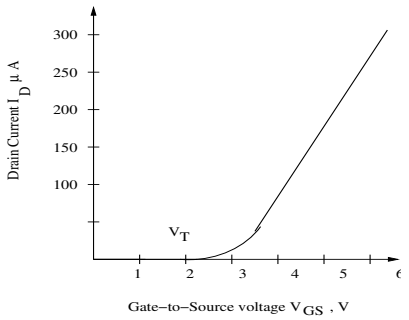

Figure 4: Transfer Characteristics of NMOS enhancement transistor in Fig.3

The volt-ampere characteristics in the Fig.3 are for an ideal MOSFET. in reality, I_D increases slightly with V_{DS} in saturation region. The cause of this "channel-length modulation," an effect analogous to the base-width modulation in the BJT. As shown in the Fig.5, if actual characteristics are extended back into the second quadrant, they all meet at $V_{DS} = -1/\lambda$. Because of the similarity with Early effect in BJTs, the quantity $1/\lambda$ is also referred to as the *Early Effect*. Typical values of λ are in the range of 0.01 to 0.03 V^{-1}. To account for the channel-length modulation, Eqn.(2) is modified by the factor $(1 + \lambda V_{DS})$ as given by,

$$I_D = k\left(\frac{W}{L}\right)(V_{GS} - V_T)^2(1 + \lambda V_{DS}) \tag{4}$$

The effect of the term $1 + \lambda V_{DS}$ is usually negligible in digital circuits but can be important in analog circuits.

Comparison of PMOS and NMOS Transistors: Historically p-channel enhancement transistors were used first in MOS system because they were more easily produced with greater yields and reliability than n-channel devices. Improvement in fabrication methods have led to the dominance of NMOS transistors. The reasons for this is described as, the hole mobility in silicon at normal filed intensities is about 500 $cm^2/(V.s)$. Thus, for devices having the same dimensions (1) the current in the PMOS transistor is less than half of that in an NMOS device and (2) the

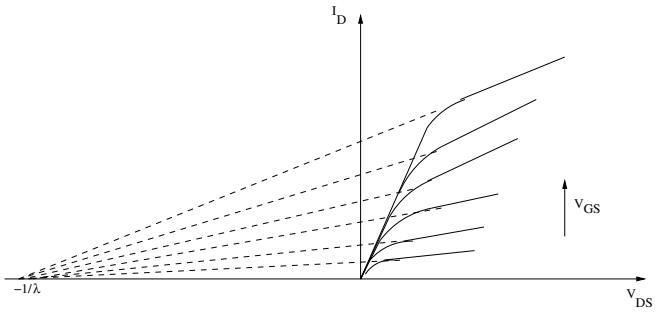

Figure 5: Extension of output characteristics of NMOS transistor showing effect of channel-length modulation

ON resistance of a p-channel MOSFET is nearly three times that for an n-channel MOSFET. Alternatively, to achieve the same values of current and ON resistance as in an NMOS transistor, the W/L ration of an PMOS device must be increased to account for the lower hole mobility. This results in the PMS devices requiring nearly three times the area of an equivalent NMOS transistor. Thus NMOS circuits are smaller than PMOS circuits of the same complexity. The higher packaging density of the n-channel MOS also makes it faster in switching applications due to the smaller junction areas. For all the reasons stated in this paragraph, NMOS devices are used almost exclusively.

The Depletion MOSFET: A second type of MOS transistor can be made if, between the n-type regions for drain and source, a narrow n channel is embedded into the substrate. Let us consider the operation of such n-channel structure as shown in the Fig.6. The minus signs in the Fig.6 are intended to indicate free electrons in the channel near the interface with the oxide layer. With $V_{DS} = 0$, negative gate voltage induces positive charge into the channel. The recombination of induced positive charge with the existing negative charge in the channel causes a depletion of majority carriers. this action accounts for the designation "depletion MOSFET". If the gate voltage is made more negative, majority carriers can be virtually depleted, and in effect, the channel is eliminated. Under these circumstances, the drain current is zero. The least negative value of V_{GS} for which the channel is depleted of majority carriers is called the *threshold voltage* V_T (analogous to the pinch-off voltage in a JFET).

With $V_{GS} = 0$, application of a positive V_{DS} produces an appreciable drain current denoted by I_{DSS}. As V_{GS} decreases toward the threshold, the drain current decreases. At fixed V_{GS}, increasing values of V_{DS} cause the drain current to saturate as the channel becomes pinched off. The reasons for this are similar to the causes of saturation in enhancement devices. Note in the Fig.6(b) that because of the voltage

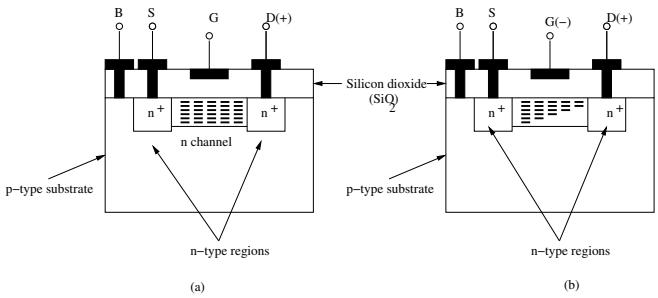

Figure 6: Structure of an n-channel depletion-mode MOSFET with (a) $V_{GS} = 0$ and (b) $V_T < V_{GS} < 0$

drop along the channel due to I_D, the region of the channel nearest the drain is depleted more than is the region in the vicinity of the source. this phenomenon is analogous to pinch-off occurring in a JFET at the drain end of the channel.

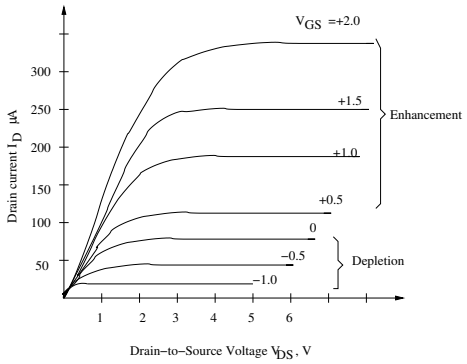

Figure 7: Output Characteristics of an NMOS transistor

A MOSFET of the type just described may also be operated in an enhancement mode. It is only necessary to apply a positive gate voltage so that negative charges are induced into the n-type channel. The additional negative charges induced into the channel ("enhances") the number of majority carriers already present. Thus, for positive V_{GS}, the drain current I_D is greater than I_{DSS}. The transfer function for this device at $V_{DS} = 5\ V$ is depicted in Fig.8.

MOSFET Circuit Symbols: Four commonly used circuit symbols for n-channel MOSFETs are depicted in Fig.9. the symbols in Figs.9(a) and Fig.9(b) can be used either for enhancement or depletion devices. The circuit symbol in Fig.9(c) is used only for the enhancement-mode device. If the body, or substrate

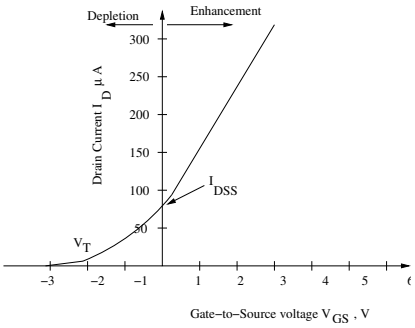

Figure 8: Transfer Characteristics of an NMOS transistor

connection is not indicated, it is assumed that the substrate is either connected to the source terminal or that B is tied to the most negative potential. this connection reverse-biases the pn junctions formed by the drain and source regions and the substrate. Where both enhancement and depletion devices are employed in the same circuit, we distinguish depletion MOSFETs by the use of the symbol given in Fig.9(d). Most often we use Fig.9(b) as the n-channel MOSFET circuit symbol with standard substrate connections implied.

The positive sense of all terminal currents is into the device. Thus, for an n-channel MOSFET, I_D is positive and I_S is negative. Since $I_G = 0$, $I_D = I_S$. The voltage drop between drain and source is designated by V_{DS}; V_{GS} is used to indicate the voltage drop from gate-to-source. Both quantities are positive for n-channel enhancement MOSFETs. Depletion-mode operation requires negative values of V_{GS} and positive values of V_{DS}.

For p-channel MOSFETs, the circuit symbols shown in the Fig.9 are used with direction of the arrow reversed.

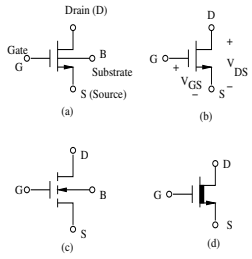

Figure 9: Circuit Symbols for an NMOS transistor