INTRODUCTION TO MOS TECHNOLOGY

1. The MOS transistor

The most basic element in the design of a large scale integrated circuit is the transistor. For the processes we will discuss, the type of transistor available is the Metal-Oxide-Semiconductor Field Effect Transistor (MOSFET). These transistors are formed as a ``sandwich'' consisting of a semiconductor layer, usually a slice, or wafer, from a single crystal of silicon; a layer of silicon dioxide (the oxide) and a layer of metal. These layers are patterned in a manner which permits transistors to be formed in the semiconductor material (the ``substrate''); a diagram showing a typical (idealized) MOSFET is shown in Figure ⌐. Silicon dioxide is a very good insulator, so a very thin layer, typically only a few hundred molecules thick, is required. Actually, the transistors which we will use do not use metal for their gate regions, but instead use polycrystalline silicon (poly). Polysilicon gate FET's have replaced virtually all of the older devices using metal gates in large scale integrated circuits. (Both metal and polysilicon FET's are sometimes referred to as IGFET's --- insulated gate field effect transistors, since the silicon dioxide under the gate is an insulator. We will still continue to use the term MOSFET to refer to polysilicon gate FET's.)
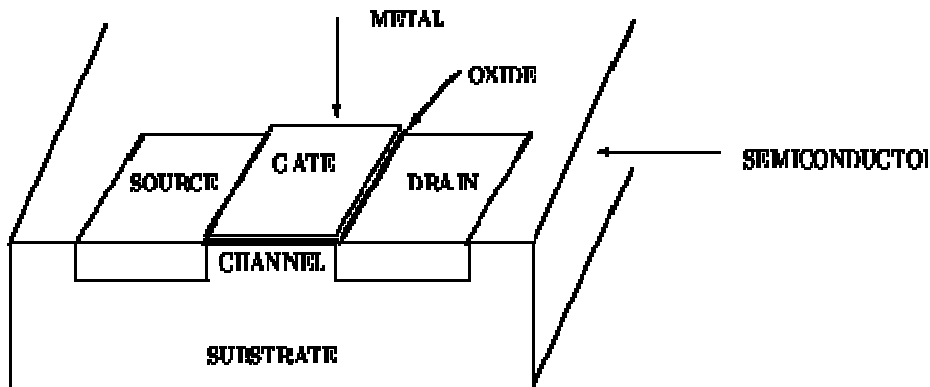
Figure: 1.1 MOS Transistor

The transistor consists of three regions, labeled the ``source'', the ``gate'' and the ``drain''. The area labeled as the gate region is actually a ``sandwich'' consisting of the underlying substrate material, which is a single crystal of semiconductor material (usually silicon); a thin insulating layer (usually silicon dioxide); and an upper metal layer. Electrical charge, or current, can flow from the source to the drain depending on the charge applied to the gate region. The semiconductor material in the source and drain region are ``doped'' with a different type of material than in the region under the gate, so an NPN or PNP type structure exists between the source and drain region of a MOSFET. Figure 1.1 shows a cross section of both types of MOSFET. In Figure 1.1 (a), the source and drain regions are doped with N type material and the substrate doped with P type material. Such a transistor is called an N channel MOSFET. If they were doped with P type material, and the substrate doped with N type material as in Figure 1.1 (b), the device would be called a P channel MOSFET.
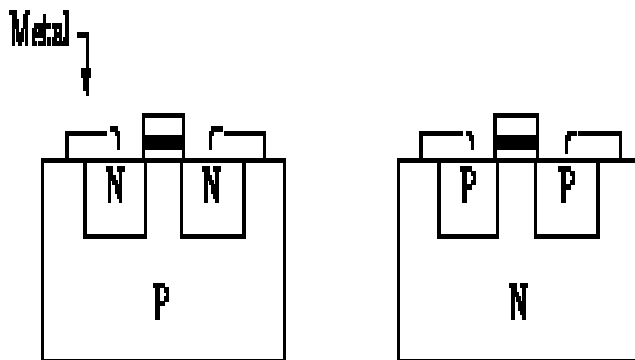
Figure: 1.1(a) N channel MOSFET   **Figure:** 1.1(b) P channel MOSFET

The source and drain regions are quite similar, and are labeled depending on to what they are connected. The source is the terminal, or node, which acts as the source of charge carriers; charge carriers leave the source and travel to the drain. In the case of an N channel MOSFET, the source is the more negative of the terminals; in the case of a P channel device, it is the more positive of the terminals. The area under the gate oxide is called the ``channel''.

The MOSFET can operate as a very efficient switch for current flowing between the source and drain region of the device. For the simplest type of MOSFET, the ``enhancement mode MOSFET'', which acts as a ``normally open'' switch, the operation of the device can be described qualitatively with reference to Figure.1.3
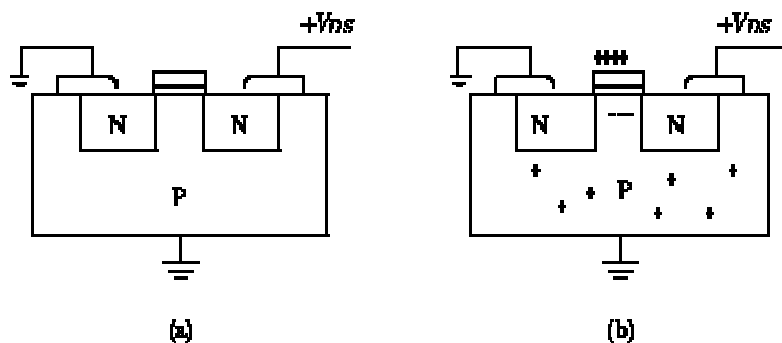

Figure:1.3 enhancement mode

Figure 1.3 (a) shows an N-channel MOSFET with the source and drain connected to power (V_{DS}) and ground (V_{SS}); the substrate, or body of the device, is also connected to ground. In this case, there is a reverse biased PN junction between at least one of the N wells and the substrate, so no current can flow through the substrate. In particular, there will be no current flow in the channel region under the gate of the transistor, and therefore no current will flow between the source and drain of the device. Under these conditions, the MOSFET is turned *off*.

Figure1.3 (b) shows the same N-channel MOSFET with a positive charge applied to the gate of the device. Under these circumstances, if the gate is given a sufficiently large charge, negative charge carriers (electrons) will be attracted from the bulk of the substrate material into the channel region immediately below the oxide under the gate. When more electrons are attracted into this region than there are positive charge carriers (holes) in the channel, then the channel effectively behaves as an N type region, and current can flow between the source and the drain. When this happens, the MOSFET is turned *on*. Note that a certain minimum charge must be applied to the gate to overcome the excess of holes already in the channel region because of the P type doping in the substrate. This means that the switch is not turned on immediately, rather there must be some minimum amount of charge applied to the gate before the transistor is switched on. The voltage which must be applied to the gate before the transistor allows current to flow between the source and drain is called the ``threshold voltage'', designated as V_{th}.

This type of transistor is called an N channel enhancement mode MOSFET. (It is called N channel because the conduction in the channel is due to N type charge carriers; it is said to be an ``enhancement mode'' device because the channel conduction is enhanced by a charge applied to the gate.) Figure1.4 shows a set of typical characteristic curve for the current I_{DS} between the drain and source of a MOSFET as a function of the voltage V_{DS} for a range of gate voltages, V_{GS}.
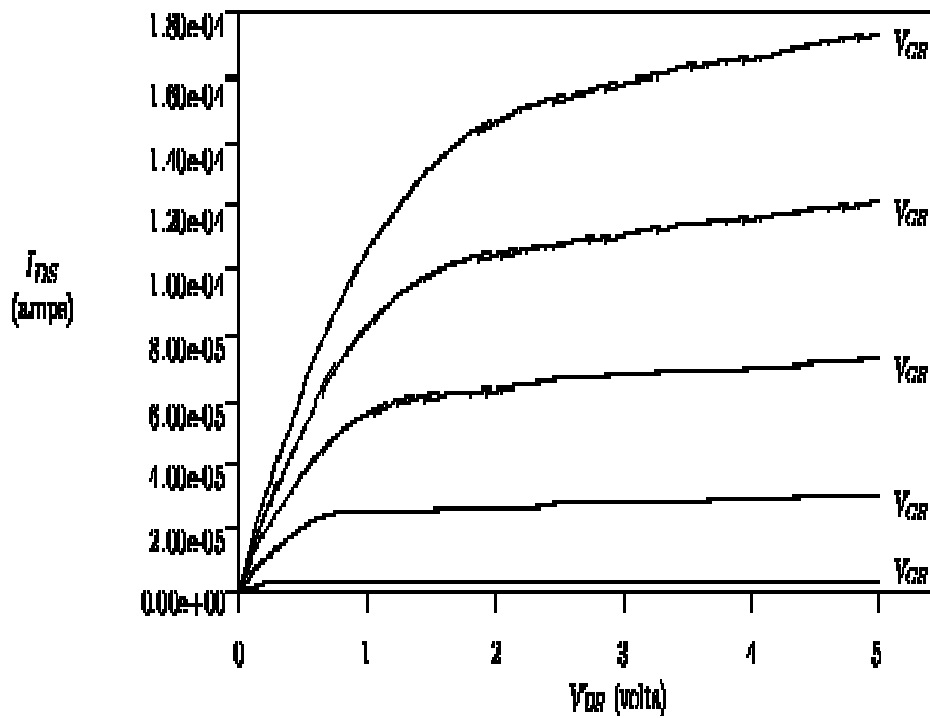
Figure:1.4 characteristic curve enhancement mode

A second type of MOSFET can also be constructed; this type of device is commonly used in purely NMOS designs, but is not used in CMOS designs. (Presently, we only have access to CMOS processes.) This type of MOSFET, the ``depletion mode MOSFET'', acts as a ``normally closed'' switch. Its behavior can qualitatively be explained with reference to Figure 1.5 which shows an N channel depletion mode MOSFET.
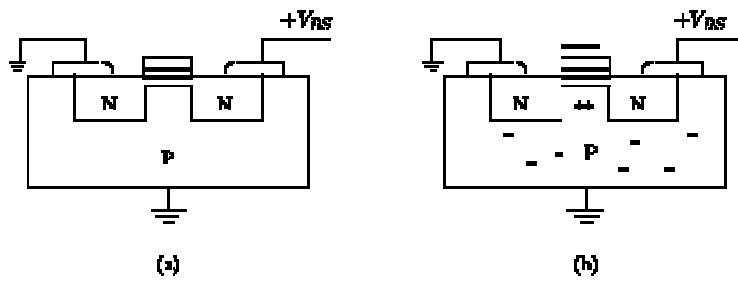
Figure: 1.5 (a) N channel depletion mode MOSFET **Figure: 1.5** (b) P channel depletion mode MOSFET

In the depletion mode MOSFET, a thin layer of semiconductor material immediately beneath the gate oxide is permanently doped with the same type material as the source and drain regions (but different from the bulk of the substrate semiconductor material). This thin layer allows conduction to occur in the channel region when no charge is applied to the gate. If a negative charge is applied to the gate, then the negative charge carriers in the thin N-doped region immediately beneath the gate oxide will be repelled from this region, leaving no free charge carriers, and conduction will cease. In the depletion mode MOSFET, a charge (with the same polarity as the drain dopant) applied to the gate turns the transistor *off*.

Depletion mode MOSFETs find their most common use not as switches but as resistors. As a permanently ``on" transistor, the device has a high resistance compared with the doped semiconductor material itself, and the resistance is readily variable by modifying the size of the transistor. (At fabrication time, the resistance can be modified by varying the number of ions which are implanted in the gate region of the device). The commonly used circuit symbols for N- and P- channel enhancement and depletion mode MOSFETs are shown in Figure 1.6. Figure 1.6 (a) shows the commonly used circuit symbols for P- and N- channel enhancement mode MOSFETs; the corresponding circuit symbols for depletion mode devices are shown in Figure 1.6 (b).
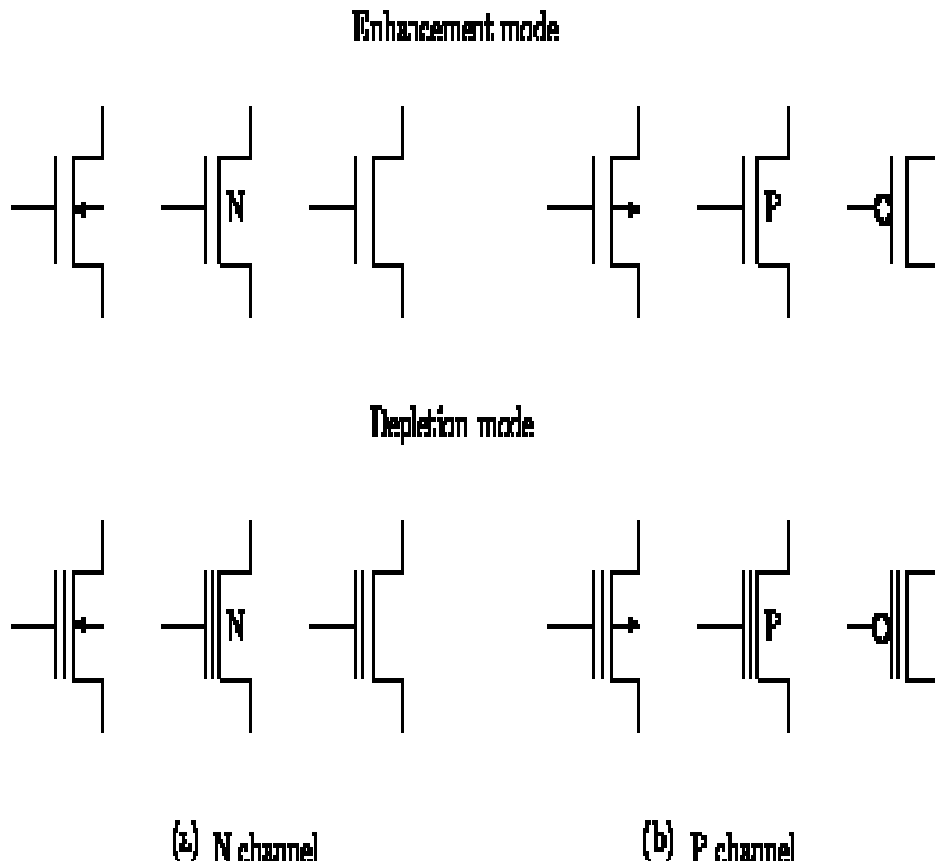
Enhancement mode

Depletion mode

(a) N channel   (b) P channel

Figure: 1.6 N- and P- channel enhancement and depletion mode MOSFETs

Both enhancement and depletion mode transistors are used in many of today's microelectronic circuits. The most popular circuit technology using both enhancement and depletion mode devices is the conventional NMOS technology. In this technology, depletion mode transistors are mainly as resistors, and enhancement mode transistors are used as switches. Figure 1.7 shows a typical inverter implemented in this technology, together with its switch equivalent. Also shown is a plot of the output of a typical example of such an inverter for a given input pulse. (The input pulse has a rise and fall time of 0.5 ns.)
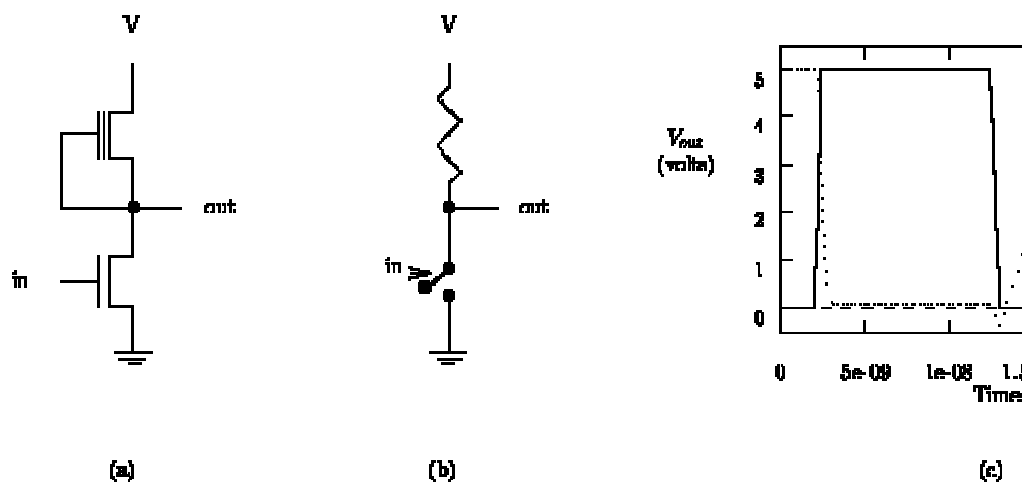
Figure: 1.7 An NMOS inverter

The gate of the depletion mode transistor is connected to its drain, to keep the transistor permanently turned on. The depletion mode transistor is used as a ``pull-up'' resistor, and the enhancement mode transistor is used as a switch to ``pull down'' the output when the switch is turned on. Note that in this technology, the resistance of the permanently turned on depletion mode transistor must be large compared with the ``on'' resistance of the enhancement mode transistor, but small compared with the ``off'' resistance of the transistor. This type of logic is often called a ``ratioed logic'', since the ratio of the pull-up resistance to the pull-down resistance effectively determines the voltage at which the output of the device changes state. Typically, $R_{pu} \approx 4R_{pd}$. The large resistive pull-up transistor causes three particular problems with this technology:

1. The depletion mode transistor must be made large (*i.e.*, long and thin) to create the large ``on'' resistance.
2. When driving a capacitive output load such as the gate of another transistor, the charging time (proportional to $R_{dep}C$) will be long compared to the discharging time (proportional to $R_{enh}C$). This effect is clearly evident in Figure 1.7 (c).
3. The device consumes DC power whenever the enhancement mode pull down device is turned on, due to the resistive losses in the pull-up transistor.

The third problem becomes more serious as feature sizes for transistors decrease, because the number of such resistors per unit area increases, and the devices may

well, resulting in device failure due to overheating.