MOS Transistor Characteristics

The MOS transistor is a majority-carrier device in which the current in a conducting channel between the source and drain is controlled by a voltage applied to the gate. In an nMOS transistor, the majority carriers are electrons; in a pMOS transistor, the majority carriers are holes. The behavior of MOS transistors can be understood by first examining an isolated MOS structure with a gate and body but no source or drain. Figure 2.2 shows a simple MOS structure. The top layer of the structure is a good conductor called the ***gate***. The middle layer is a very thin **insulating film** of SiO2 called the ***gate oxide***. The bottom layer is the doped silicon **body**. The figure shows a p-type body in which the carriers are holes. The body is grounded and a voltage is applied to the gate. The gate oxide is a good insulator so almost zero current flows from the gate to the body.

In Figure 2.2(a) , a negative voltage is applied to the gate, so there is negative charge on the gate. The mobile positively charged holes are attracted to the region beneath the gate. This is called the **accumulation** mode. In Figure 2.2(b), a small positive voltage is applied to the gate, resulting in some positive charge on the gate. The holes in the body are repelled from the region directly beneath the gate, resulting in a **depletion region** forming below the gate. In Figure 2.2(c), a higher positive potential exceeding a critical threshold voltage Vt is applied, attracting more positive charge to the gate. The holes are repelled further and some free electrons in the body are attracted to the region beneath the gate. This conductive layer of electrons in the p-type body is called the **inversion layer**.

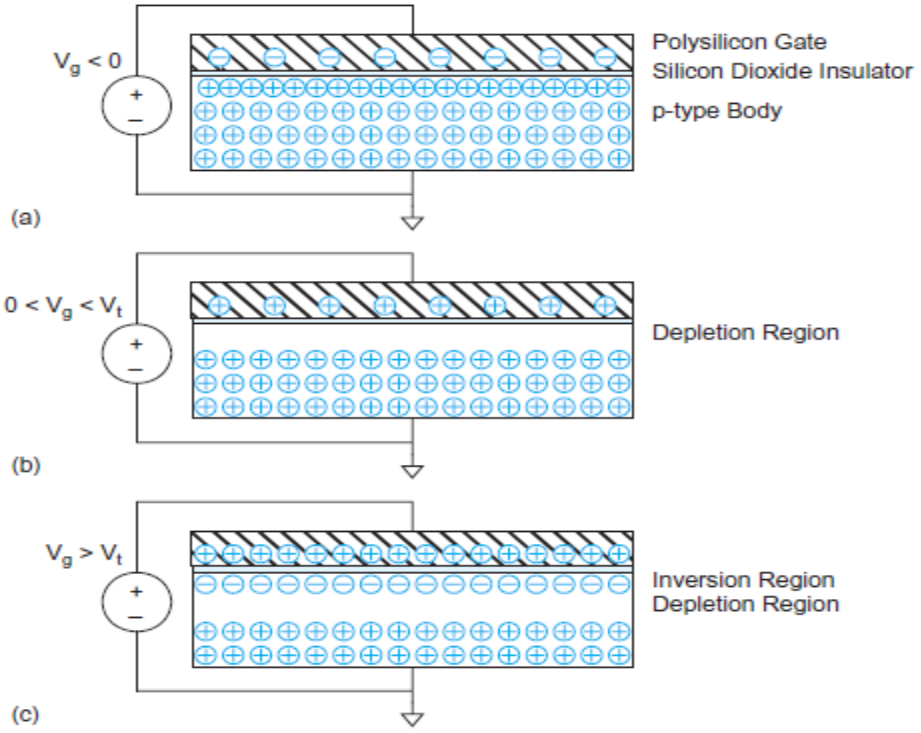
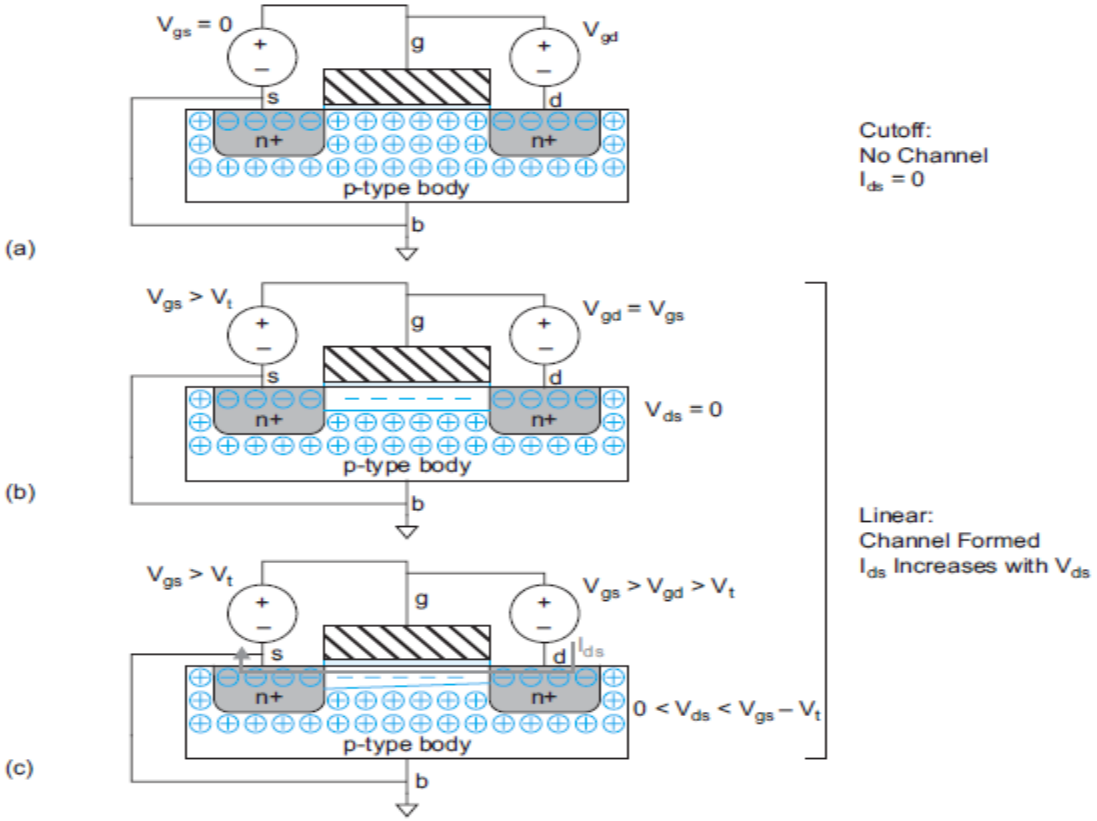


FIGURE 2.2 MOS structure demonstrating (a) accumulation, (b) depletion, and (c) inversion

Figure 2.3 shows an nMOS transistor. The transistor consists of the MOS stack between two n-type regions called the *source* and *drain*. In Figure 2.3(a), the gate-to-source voltage ***Vgs*** is less than the threshold voltage. The source and drain have free electrons. The body has free holes but no free electrons. Suppose the source is grounded. The junctions between the body and the source or drain are zero-biased or reverse-biased, so little or no current flows. We say the transistor is OFF, and this mode of operation is called ***cutoff***. It is often convenient to approximate the current through an OFF transistor as zero, especially in comparison to the current through an ON transistor. Remember, however, that small amounts of current leaking through OFF transistors can become significant, especially when multiplied by millions or billions of transistors on a chip. In Figure 2.3(b), the gate voltage is greater than the threshold voltage. Now an inversion region of electrons (majority carriers) called the ***channel*** connects the source and drain, creating a conductive path and turning the transistor **ON**. The number of carriers and the conductivity increases with the gate voltage. The potential difference between drain and source is ***Vds = Vgs - Vgd*** **. If *Vds* = 0 (i.e., *Vgs* = *Vgd*)**, there is no electric field tending to push current from drain to source. When a small positive potential *Vds* is applied to the drain (Figure 2.3(c)), current *Ids* flows through the channel from drain to source. This mode of operation is termed ***linear,*** **resistive, triode, nonsaturated, or unsaturated**; the current increases with both the drain voltage and gate voltage. If **Vds** becomes sufficiently large that **Vgd < Vt** , the channel is no longer inverted near the drain and becomes pinched off (Figure 2.3(d)). However, conduction is still brought about by the drift of electrons under the influence of the positive drain voltage. As electrons reach the end of the channel, they are injected into the depletion region near the drain and accelerated toward the drain. Above this drain voltage the current Ids is controlled only by the gate voltage and ceases to be influenced by the drain. This mode is called **saturation**.



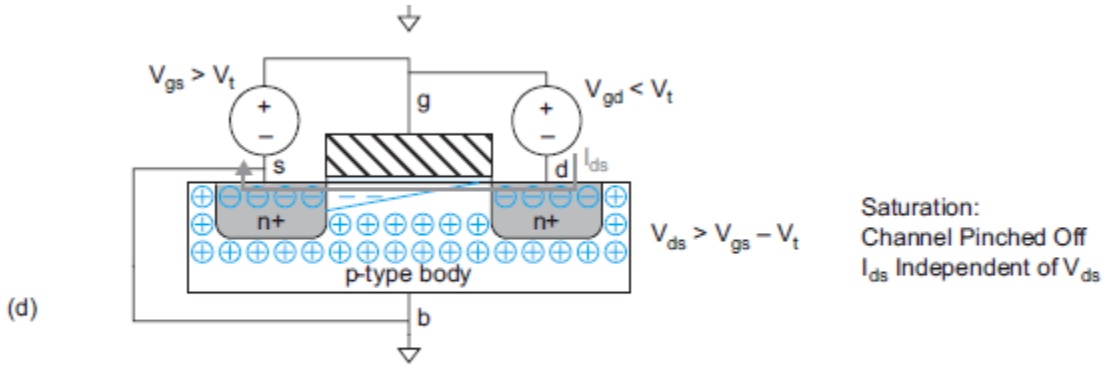


FIGURE 2.3 nMOS transistor demonstrating cutoff, linear, and saturation regions of operation