Metal-Oxide-Semiconductor-Field-Effect-Transistor

Metal-oxide-semiconductor-field-effect-transistors (MOSFETS) are the most widely utilized semiconductor transistors in contemporary technology. MOSFETS are four-terminal devices consisting of a source, drain, gate and ground. When a voltage is applied to the gate, current is allowed to flow from the source to the drain by the field effect. MOSFETS can be used to amplify electrical signals and are networked to form circuit logic. The miniaturization of integrated circuits has directly resulted from advances in fabrication technology which allow for smaller MOSFET lengths and device widths.

Structure and Operation

A 2-D schematic of an n-channel MOSFET device is shown in Figure 1. The width of the MOSFET is orthogonal to the image.

MOSFETS incorporate a MOS Capacitor that is next to two anti-doped regions. When no voltage is applied to the gate, the switch is off, and current cannot flow between the source and drain contacts. This is because free electrons from the source immediately recombine with holes once they reach the p-doped region. Put another way, when the switch is off, current cannot flow between the n-doped regions no matter the voltage applied because at one side there will be a pn
junction in reverse bias.

When voltage is applied to the gate of the MOS capacitor, it creates an electric field which brings free electron charge carriers to the contact surface between the insulator and semiconductor. Initially this creates a depletion region, and then an inversion layer forms where free electrons dominate. This inversion layer is a conduction channel between the n-doped regions, which allows current to flow, and the switch is on. The gate voltage required for this effect is called the threshold voltage, because before this voltage no current can flow. This electric field effect on the semiconductor charge carriers is the MOSFET's namesake.

**Pinch-Off Effect**

The inversion layer in a MOSFET that is conducting is not symmetrical. An illustration of the inversion layer in this case is in Figure 2. In the inversion layer, the current density is highest near the drain and lower near the source due to the asymmetrical shape.

![Inversion layer in a conducting MOSFET](https://eng.libretexts.org/Bookshelves/Materials_Science/Supplemental_Modules_(Materials_Science)/Semiconductors/Metal-...

The slanted shape of the inversion layer is due to the fact that charge carriers along the channel experience a combination of voltage potentials from the gate, source, and drain. At the drain, there is higher voltage than at the source, so the electrons immediately under the gate next to the drain experience less influence from the gate; they are partially shielded. If the voltage difference between the source and drain is high enough, the inversion layer will pinch off before the drain, even in the presence of a greater-than-threshold gate voltage. This is the saturation voltage, and it is a limiting factor in the current a MOSFET can carry.

**Fabrication**

An illustration of MOSFET fabrication process is presented in figure 2.
In making a MOSFET, first a lightly doped p-type Silicon wafer is used. $\text{SiO}_2$ is grown over this and $\text{Si}_3\text{N}_4$ is then deposited over that. The device region is protected by a layer of photo-resist, and then the surface is heavily doped with boron through the $\text{Si}_3\text{N}_4$ and $\text{SiO}_2$ into the wafer surface. This creates heavily doped p channel stop which prevent conduction between devices. Next, the $\text{SiO}_2$ is etched away and a layer of field oxide is deposited, and more boron is doped in. This is to modify the doping concentration in the channel. Next, the field oxide is removed over the active device area and the gate oxide is grown in the center. A poly silicon layer is deposited on top of the gate and heavily doped to be conductive. Finally, arsenic atoms are doped to form the n+ drain and source regions. The gate itself shields the doping so that the regions are defined by the gate. This minimizes over-lap capacitance between the n+ regions and the gate.

Questions

1. Q1: Why is the insulator present? What would happen to the device if it were absent?
2. Q2: If the gate voltage is nonzero but below the threshold voltage, and the source to drain voltage is zero, what does the inversion layer look like?
3. Q3: How does the device fail if it gets too hot? Will it be pinned open or closed?
1. If the metal gate made direct contact to the semiconductor, when voltage was applied it would conduct current directly from the source, and the device would not operate properly.

2. The inversion layer will not be present. There will be a depletion region at the oxide contact surface, but because it is below the threshold voltage, not enough carriers are drawn to the surface to create an inversion layer. The depletion layer will be flat along the length of the device because there is no horizontal voltage gradient to slant it.

3. The device will be pinned open. When the MOSFET is too hot, thermally excited carriers will dominate and the depletion zone between the source and drain will dissipate, allowing current flow.

References