5 MOS Junction and Devices

5.1 MOS junction

A junction in which a metal and a Si semiconductor are connected via an insulator film (in this case, an oxide film) is called a metal oxide semiconductor (MOS) junction. Figure 5.1 shows the junction configuration and energy band structure of a MOS junction with a P-type Si semiconductor as an example.

ϕ_M : Electron work function in metal

E_{fM} : Fermi level in metal

χ_{ox} : Electron affinity in oxide film

ε_{ox} : Permittivity of oxide film

ϕ_S : Electron work function in semiconductor

E_{fS} : Fermi level in semiconductor

χ_S : Electron affinity in semiconductor

E_C : Lowest energy level of conduction band

E_V : Highest energy level of valence band

$E_G = E_C - E_V$: Band gap

E_i : Fermi level in case of intrinsic semiconductor

$$E_i \approx \frac{1}{2}(E_C + E_V)$$

$\Delta E_f = E_i - E_{fS}$: Fermi energy difference

ε_S : Permittivity of semiconductor

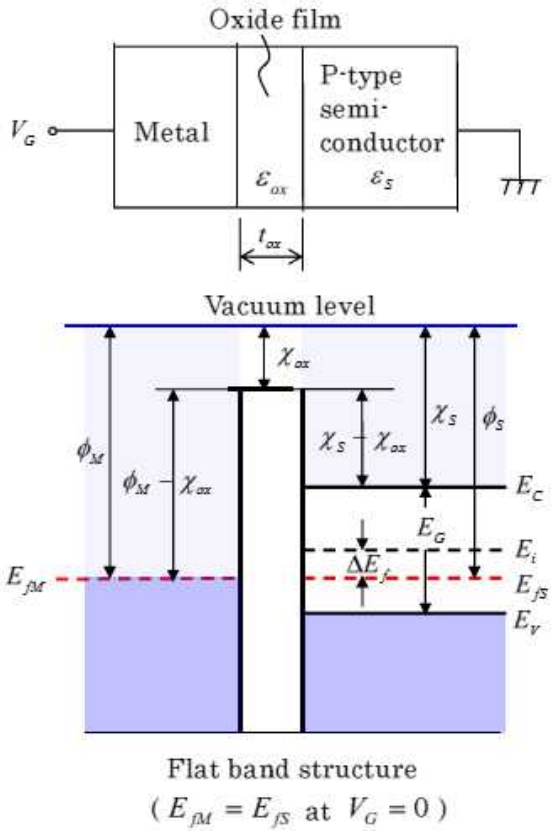


Figure 5.1 Junction configuration and energy band structure of a MOS Junction (metal- P-type Si semiconductor MOS junction, flat band structure)

For the analysis, it is first assumed here that the potential energies of the metal and semiconductor before the junction are equal, i.e. that the work

functions and Fermi energy levels are equal ($\phi_M = \phi_S$, $E_{fM} = E_{fS}$). As a result, there is no energy transfer between the metal and semiconductor during the junction, and as a result the energy band structure remains flat without any change (curve) (called a flat band structure). For the analysis, the junction area is assumed to be the unit area. The conduction electron (minority carrier) density n and the electron hole (majority carrier) density p in the semiconductor at thermal equilibrium are given as follows (see section 1.5, equations (1.17) and (1.22))

$$\begin{aligned} n &= N_C \exp\left(-\frac{E_C - E_{fS}}{k_B T}\right) \\ &= N_C \exp\left(-\frac{E_i - E_{fS}}{k_B T}\right)\exp\left(-\frac{E_C - E_i}{k_B T}\right) = n' \exp\left(-\frac{\Delta E_f}{k_B T}\right) \\ n' &= N_C \exp\left(-\frac{E_C - E_i}{k_B T}\right), \quad N_C = 2\left(\frac{2\pi m_e^* k_B T}{h^2}\right)^{\frac{3}{2}}, \quad \Delta E_f = E_i - E_{fS} \end{aligned} \tag{5.1}$$

$$\begin{aligned} p &= N_V \exp\left(-\frac{E_{fS} - E_V}{k_B T}\right) \\ &= N_V \exp\left(-\frac{E_i - E_V}{k_B T}\right)\exp\left(-\frac{E_{fS} - E_i}{k_B T}\right) = p' \exp\left(\frac{\Delta E_f}{k_B T}\right) \\ p' &= N_V \exp\left(-\frac{E_i - E_V}{k_B T}\right), \quad N_V = 2\left(\frac{2\pi m_h^* k_B T}{h^2}\right)^{\frac{3}{2}} \end{aligned} \tag{5.2}$$

where E_i is given by the following equation, corresponding to the Fermi energy level of an intrinsic semiconductor.

$$E_i = \frac{1}{2}(E_C + E_V) \tag{5.3}$$ (same as Eq. (1.28))

In this case, n' , p' correspond to the carrier density of an intrinsic semiconductor, and the following relationship holds.

$$n_i = n' = p' = N_C \exp\left(-\frac{E_C - E_i}{k_B T}\right) = N_V \exp\left(-\frac{E_i - E_V}{k_B T}\right) \tag{5.4}$$

Next, the following cases are described in turn, where the semiconductor side is grounded and the voltage V_G is applied to the metal side as shown in Fig. 5.1.

When $V_G \leq 0$ ... accumulation state

When $V_G \geq 0$ and V_G is small ... depletion state

When $V_G >> 0$ ... inverted state (weak inversion and strong inversion)

(1) When $V_G \leq 0$ ... accumulation state

Fig. 5.2 shows the appearance of the junction charge and the energy band structure when $V_G \leq 0$ is applied.

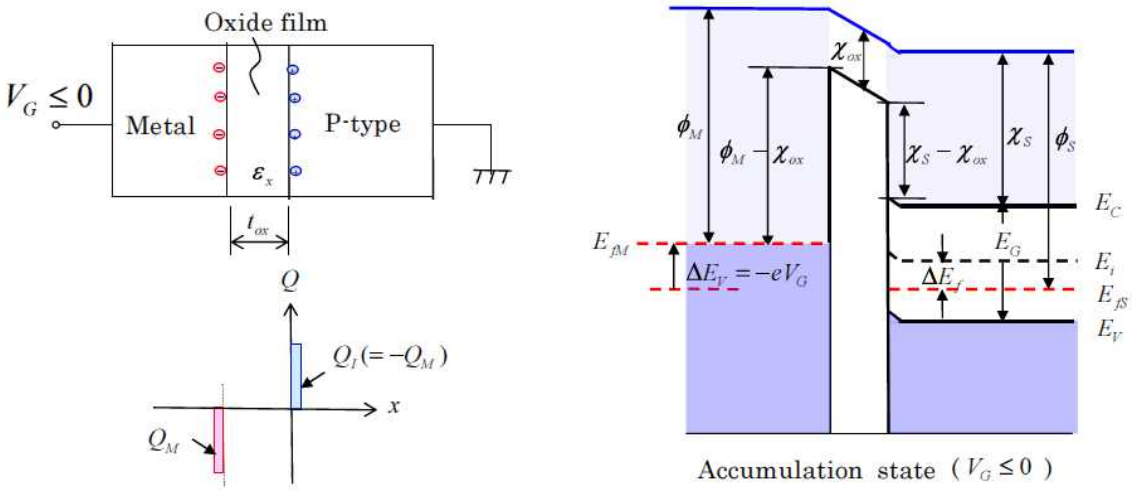


Figure 5.2 Appearance of the junction charge and energy band structure (When $V_G \leq 0$ ... accumulation state)

Current does not flow through the junction and the applied voltage V_G is applied wholly between the parallel plate capacitor (capacitance per unit area $C_{ox} = \varepsilon_x / t_{ox}$), which is made up of an oxide film. This causes a surface charge Q_M (≤ 0) due to electrons to appear on the metal side junction surface and a surface charge Q_I ($= -Q_M$) (≥ 0) due to electron holes to

appear on the semiconductor side junction surface. Depending on the applied voltage V_G, the metal side Fermi energy level changes by $\Delta E_V = -eV_G (\geq 0)$ (higher direction in this case). The energy band structure changes accordingly. This state is called the accumulation state.

(2) When $V_G \geq 0$ and V_G is small ... depletion state

Figure 5.3 shows (a) the appearance of the charge and (b) the electric field $\mathrm{E}(x)$, potential $V(x)$ and energy $E_P(x)$ at the MOS junction in the case when a voltage is applied where $V_G \geq 0$ and V_G is small.

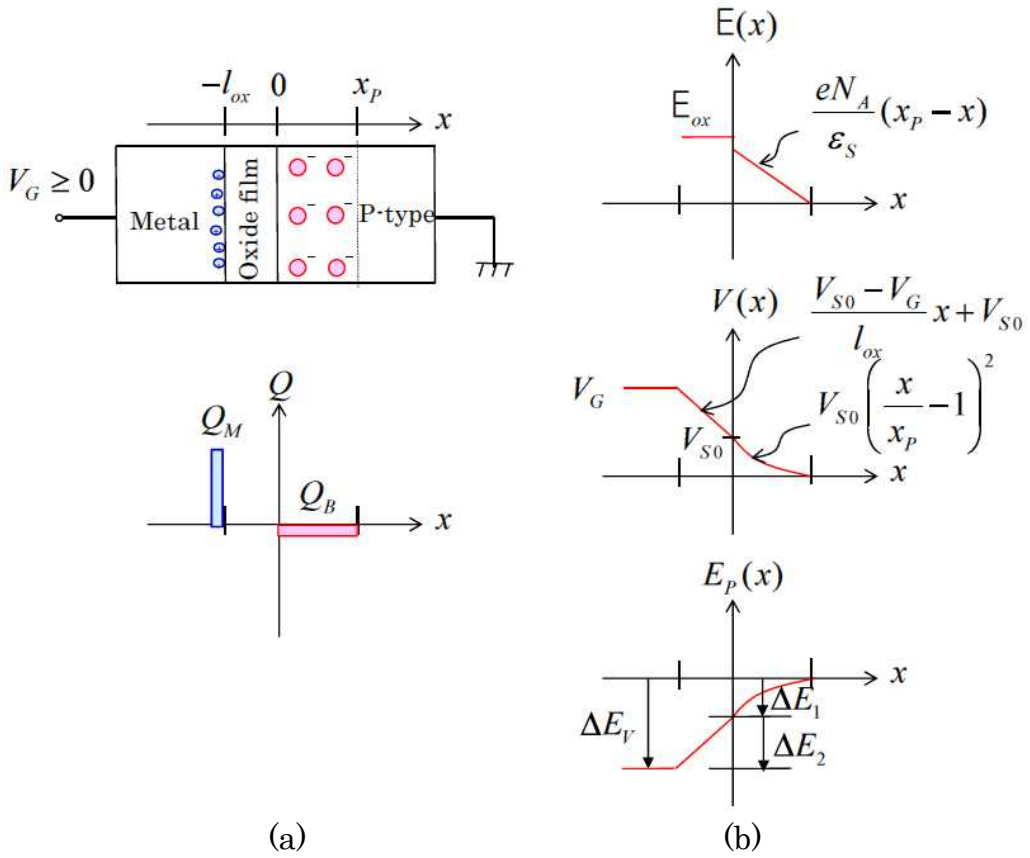


Figure 5.3 (a) Appearance of charge at the MOS junction and (b) electric field $\mathrm{E}(x)$,potential $V(x)$,energy $E_P(x)$
(When $V_G \geq 0$ and V_G is small ... depletion state)

In this case, a surface charge Q_M (≥ 0) due to electron holes is generated in the metal-side junction and a corresponding charge Q_B $(= -Q_M)$ (≤ 0) is

generated in the depletion layer of the semiconductor by ionised acceptors. When the ionised acceptor density is N_A and the depletion layer thickness is x_P, Q_M and Q_B can be related by the following equation.

$$Q_B = -eN_A x_P = -Q_M \tag{5.5}$$

Assuming that the electric field and voltage are given as functions of position x, and letting $\mathsf{E}(x)$ denote the electric field and $V(x)$ the potential at x, the following 1-dimensional Poisson equation holds.

$$-\frac{d^2V(x)}{dx^2} = \frac{d\mathsf{E}(x)}{dx} = \frac{-eN_A}{\varepsilon_S} \qquad (0 \le x \le x_P) \tag{5.6}$$

Here the analysis is carried out by assuming that the influence of charge due to carriers (conduction electrons, electron holes) in the depletion layer ($0 \le x \le x_P$) can be neglected (depletion approximation) Note *5,

Note*5

Strictly speaking, the analysis must be based on a model that can take into account the effects of conduction electrons and electron holes. This is especially true for the inversion state, where the conduction electron density increases. At present, however, the main focus is on understanding the band structure of MOS junctions, and analysis is carried out using a model based on the depletion approximation.

From the boundary conditions, $\mathsf{E}(x_P) = 0$ and $V(x_P) = 0$ at $x = x_P$, $\mathsf{E}(x)$ and $V(x)$ are obtained at $0 \le x \le x_P$ as follows.

$$\mathsf{E}(x) = \frac{eN_A}{\varepsilon_S}(x_P - x) \tag{5.7}$$

$$V(x) = \frac{eN_A}{2\varepsilon_S}x^2 - \frac{eN_A}{\varepsilon_S}xx_P + \frac{eN_A}{2\varepsilon_S}x_P{}^2 = V_{S0}\left(\frac{x}{x_P} - 1\right)^2 \tag{5.8}$$

where $V_{S0} = V(0) = \dfrac{eN_A x_P^{\,2}}{2\varepsilon_S}$ (5.9)

In the oxide film ($-l_{ox} \le x \le 0$), the following Poisson equation holds.

$$-\frac{d^2V(x)}{dx^2} = \frac{d\mathsf{E}(x)}{dx} = 0 \qquad (-l_{ox} \le x \le 0) \tag{5.10}$$

The boundary conditions are (I) $V(0) = V_{S0} = \dfrac{eN_A x_P^{\,2}}{2\varepsilon_S}$ at $x = 0$ and (II) $V(-l_{ox}) = V_G$ at $x = -l_{ox}$, from the continuity of the potential. From this, $V(x)$ can be obtained in $-l_{ox} \le x \le 0$ as follows.

$$V(x) = \frac{V_{S0} - V_G}{l_{ox}} x + V_{S0} \tag{5.11}$$

And from this, $\mathsf{E}(x)$ is the following constant value.

$$\mathsf{E}(x) = -\frac{\partial V(x)}{\partial x} = \frac{V_G - V_{S0}}{l_{ox}} = \mathsf{E}_{ox} \tag{5.12}$$

Depending on the applied voltage $V_G (\ge 0)$, the energy on the metal side changes (decreases) by $\Delta E_V = -eV_G (\le 0)$. When the energy difference in the depletion layer is ΔE_1 and the energy difference in the oxide film is ΔE_2, ΔE_V is given by

$$\Delta E_V = -eV_G = \Delta E_1 + \Delta E_2 \tag{5.13}$$

where $\Delta E_1 = -eV_{S0}$ (5.14)

From above equations (5.13), (5.14)

$$\Delta E_2 = -e(V_G - V_{S0}) \tag{5.15}$$

The energy band structure changes (curves) accordingly. In doing so, however, the electron affinity and the energy gap are conserved everywhere. Figure 5.4 shows the appearance of the junction charge and the energy band structure when $V_G \geq 0$ and V_G is small. When V_G is small and $|\Delta E_1| \leq \Delta E_f$, the conduction electron density n is very small, so that the negative charge in the semiconductor is mostly due to ionised acceptors in the depletion layer.

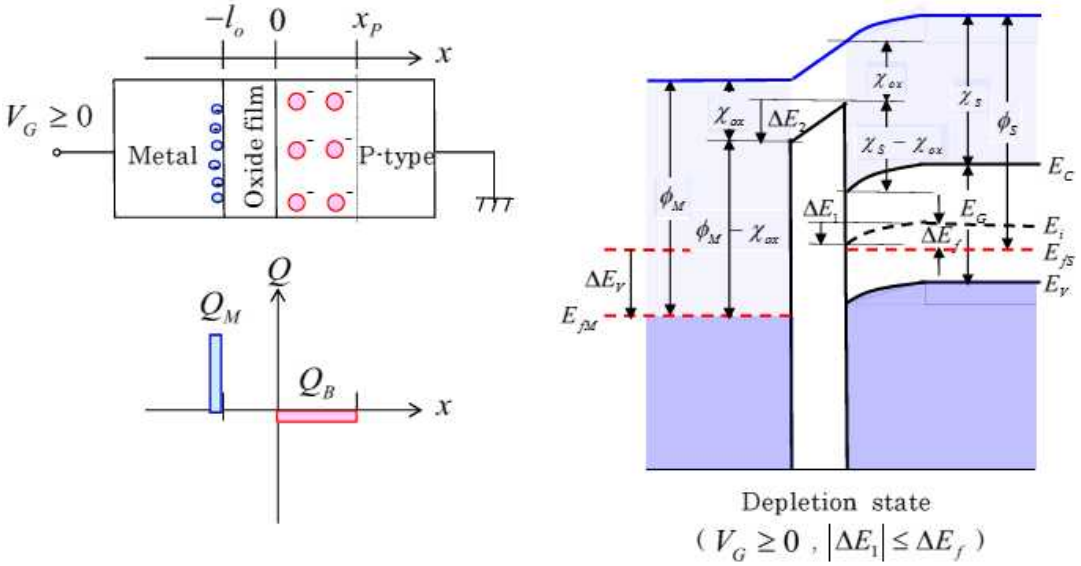


Figure 5.4 Appearance of junction charge and energy band structure (When $V_G \geq 0$ and V_G is small ... depletion state)

(2) When $V_G >> 0$ ... inverted state (weak and strong inversion)

As V_G is further increased positively, $|\Delta E_1|$ also increases and eventually a region appears where $\Delta E_f \leq |\Delta E_1|$. In this region, the conduction electron density n becomes larger than the electron hole density p, and the original P-type is inverted to N-type. This state is called the inverted state. When inversion occurs, the negative charge corresponding to the positive charge $Q_M (\geq 0)$ due to electron holes appearing on the metal junction surface becomes the charge Q_B due to the acceptor ion in the depletion layer and the

charge Q_I due to the conduction electrons appearing on the semiconductor junction surface in the inverted state. The following relationship exists between these charges.

$$Q_B + Q_I = -Q_M \tag{5.16}$$

In the inverted state, the depletion layer expansion almost stops. Let x_p at this time be x_{pmax}. The subsequent increase in negative charge in the semiconductor is due to conduction electrons appearing at the junction surface. However, in the range $\Delta E_f \leq |\Delta E_1| \leq 2\Delta E_f$ the conduction electron density n is close to that of an intrinsic semiconductor and is still not very large. This state is called weak inversion. Figure 5.5 shows the appearance of the junction charge and the energy band structure for the weak inversion case at $V_G >> 0$.

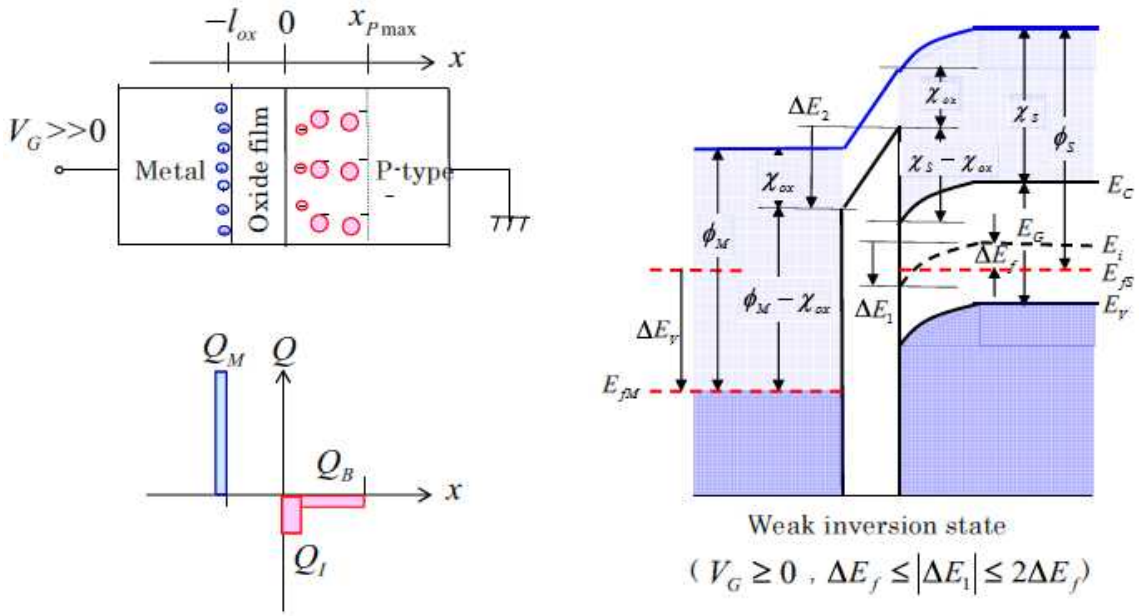


Figure 5.5 Appearance of junction charge and energy band structure (When $V_G >> 0$ ... weak inversion state)

Applying an even larger positive bias voltage ($V_G >> 0$) on the metal side, $|\Delta E_1|$ also increases further. Eventually, when $2\Delta E_f \leq |\Delta E_1|$, a conduction

electron density n appears which is greater than the electron hole density $p(\approx N_A)$ in the original P-type semiconductor, i.e. $n \geq N_A$, and a channel is formed by conduction electrons on the junction surface. This state is called strong inversion. Figure 5.6 shows the appearance of the charge and energy band structure of the junction when strongly inverted at $V_G >> 0$.

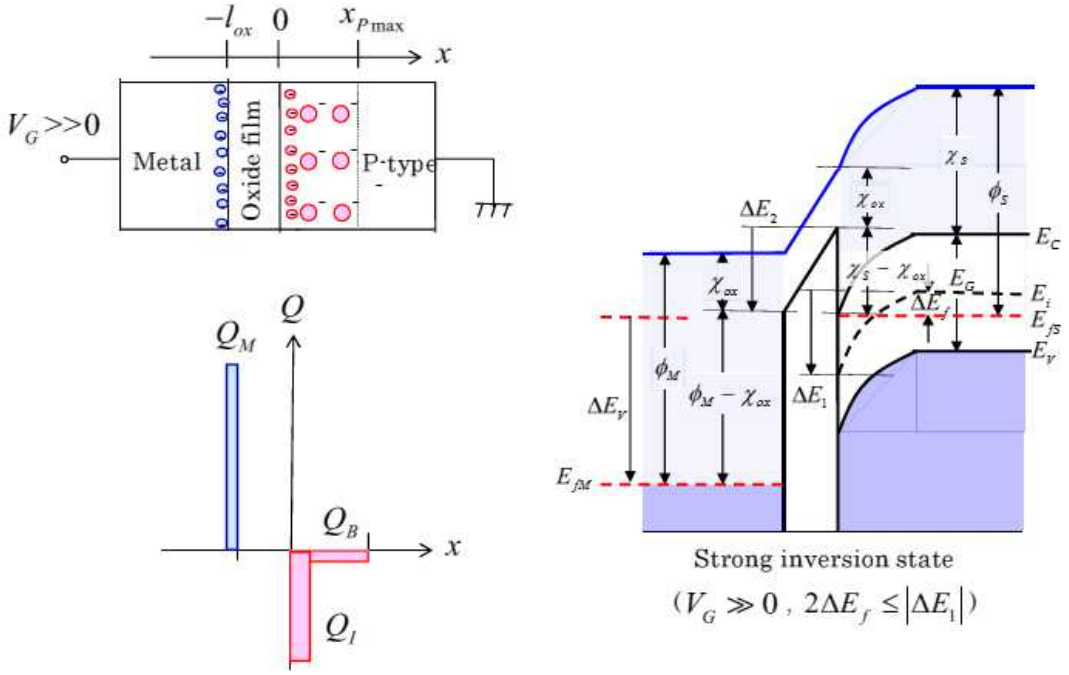


Figure 5.6 Appearance of junction charge and energy band structure (When $V_G >> 0$ ... strong inversion state).

The negative charge Q_B due to the depletion layer is given by

$$Q_B = -eN_A x_P \tag{5.17}$$

In the depletion state x_p has the following relationship with V_{S0} (from equation (5.9))

$$V_{S0} = V(0) = \frac{eN_A x_P^{\,2}}{2\varepsilon_S} \longrightarrow x_P = \left(\frac{2\varepsilon_S V_{S0}}{eN_A} \right)^{\frac{1}{2}} \tag{5.18}$$

From this

$$Q_B = -eN_A x_P = -eN_A \left(\frac{2\varepsilon_S V_{S0}}{eN_A} \right)^{\frac{1}{2}} = -\left(2e\varepsilon_S N_A V_{S0} \right)^{\frac{1}{2}} \tag{5.19}$$

If V_G is increased and a strong inversion state ($2\Delta E_f \leq |\Delta E_1|$) is reached, Q_B is approximately at the following constant value.

$$Q_B \approx -eN_A x_{P\max} = -eN_A \left(\frac{2\varepsilon_S}{eN_A} 2\frac{\Delta E_f}{e} \right)^{\frac{1}{2}} = -2\left(\varepsilon_S N_A \Delta E_f \right)^{\frac{1}{2}} \tag{5.20}$$

where the value of x_{Pmax} is obtained by letting V_{S0} be $2\frac{\Delta E_f}{e}$ in equation (5.18) for x_P. Note that the relationship between V_{S0} and $2\frac{\Delta E_f}{e}$ is obtained from the following relationship for the strongly inverted state.

$$2\Delta E_f = \left|\Delta E_1\right| = eV_{S0} \longrightarrow V_{S0} = 2\frac{\Delta E_f}{e} \tag{5.21}$$

Figure 5.7 shows Q_M, Q_B and Q_I for ΔE_1. As V_G is increased, ΔE_1 increases and the strong inversion state is reached, the charge Q_B due to the ionised acceptor remains almost unchanged and the subsequent negative charge on the semiconductor is predominantly Q_I due to the conduction electrons. At the start of the strong inversion state, i.e. when $V_{S0} = 2\frac{\Delta E_f}{e}$, the charges Q_M, Q_B and Q_I have the following relationship.

$$\left.\begin{aligned} Q_B &\approx -eN_A x_{P\max} \\ Q_M &= C_{ox}(V_G - V_{S0}) = C_{ox}(V_G - 2\frac{\Delta E_f}{e}) \\ Q_I &= -Q_M - Q_B \end{aligned}\right\} \tag{5.22}$$

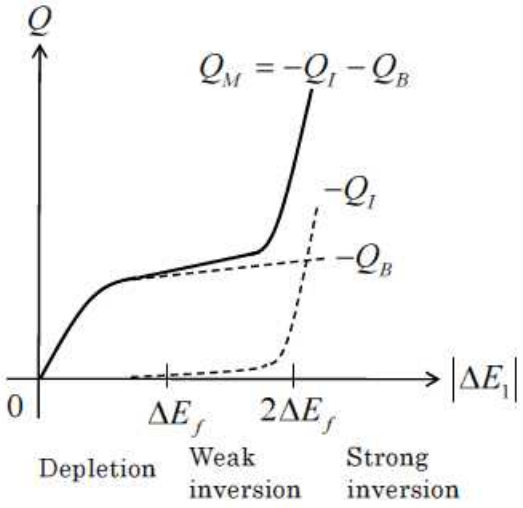


Figure 5.7 Q_M , Q_B and Q_I for ΔE_1

From equation (5.22),

$$Q_I = -C_{ox}(V_G - 2\frac{\Delta E_f}{e}) + eN_A x_{P\max} = -C_{ox}(V_G - V_{th}) \tag{5.23}$$

where. $V_{th} = 2\frac{\Delta E_f}{e} + \frac{eN_A x_{P\max}}{C_{ox}}$ (5.24)

Furthermore, substituting $C_{ox} = \frac{\varepsilon_{ox}}{l_{ox}}$, $x_{P\max} = \frac{2}{e}\left(\frac{\varepsilon_S \Delta E_f}{N_A}\right)^{\frac{1}{2}}$ (obtained from equation (5.20)), V_{th} is given by

$$V_{th} = 2\frac{\Delta E_f}{e} + 2\left(\varepsilon_S N_A \Delta E_f\right)^{\frac{1}{2}} \frac{l_{ox}}{\varepsilon_{ox}} \tag{5.25}$$

V_{th} is called the threshold voltage, the value of which is determined by the physical property values and structure of the material. Figure 5.8 shows the Q_I versus V_G characteristic.

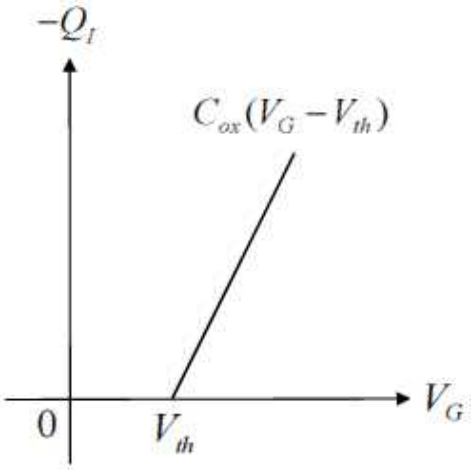


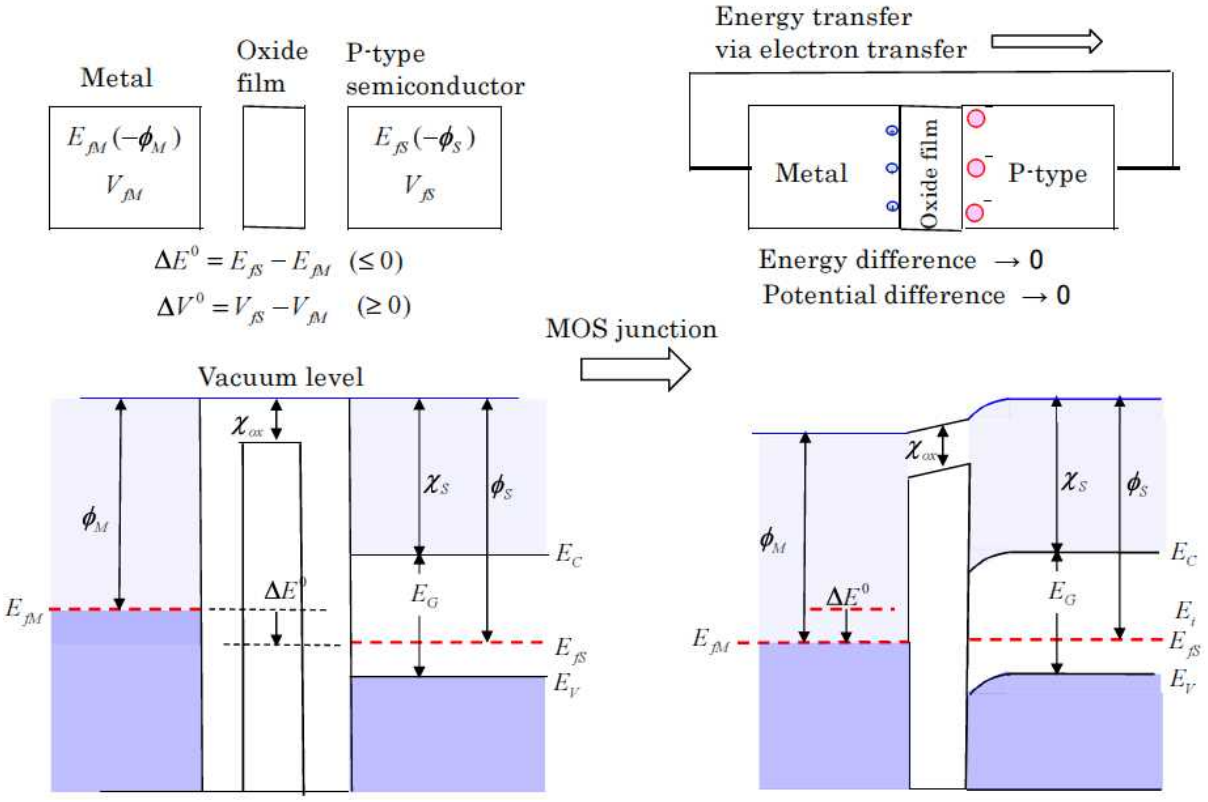
Figure 5.8 Q_I versus V_G characteristic

When V_G is increased, Q_I does not occur at $V_G \leq V_{th}$, and at $V_{th} \leq V_G$, Q_I increases in proportion to the increasing width of V_G. That is, an n-channel is formed by conduction electrons on the P-type semiconductor surface of the MOS junction, and the (negative) charge Q_I of the channel is controlled (linearly) by varying V_G.

Next, consider the general case where the potential energies originally possessed by the metal and semiconductor before joining are not equal. As an example, we consider the case where the original energy of the metal and the (P- type) semiconductor is $\phi_M \leq \phi_S (E_{fM} \geq E_{fS})$. Figure 5.9 shows the energy band structure of this metal-semiconductor MOS junction (a) before and (b) after the junction. In this case, the energy originally possessed by the metal for electrons is higher than that of the semiconductor by $\Delta E^0 (= E_{fM} - E_{fS})$, so that energy is transferred from the metal to the semiconductor via electron transfer at the junction Note *6.

Note *6.

In the case of MOS junctions, electrons cannot actually move through the oxide film. However, the metal-semiconductor connection is operated by an external circuit during bias wiring, and electron transfer can be considered to take place at this time.



(a) before junction (b) after junction

Figure 5.9 Junction configuration and energy band structure of the MOS junction (for $\phi_M \leq \phi_S (E_{fM} \geq E_{fS})$)

The transfer of electrons causes a positive charge due to electron holes to appear on the metal junction surface and a negative charge due to ionised acceptors to appear on the semiconductor junction surface. The positive and negative charges are equal (the sum of the charges is zero). This creates a potential difference ΔV_0 (positive on the metal side) between the metal and the semiconductor, which cancels out the originally existing potential difference $\Delta V^0 (= -\frac{\Delta E^0}{e})$ (negative on the metal side) and makes the potentials equal. The ΔV_0 creates an energy difference $\Delta E_0 (= -e\Delta V_0)$ (negative on the metal side), which cancels out the originally existing energy difference ΔE^0 (positive on the metal side), and the energies (Fermi energy levels) become equal, resulting in a state of thermal equilibrium. When the

Fermi energy level changes, the energy band structure changes (curves) accordingly.

Consider that in this MOS junction, the semiconductor side is grounded and a voltage of $-\Delta V_0$ is applied to the metal side. The potential difference ΔV_0 had been created at the junction by the electron transfer is cancelled out and the energy band structure becomes the same as before the junction, as shown in Figure 5.10, resulting in a flat band structure.

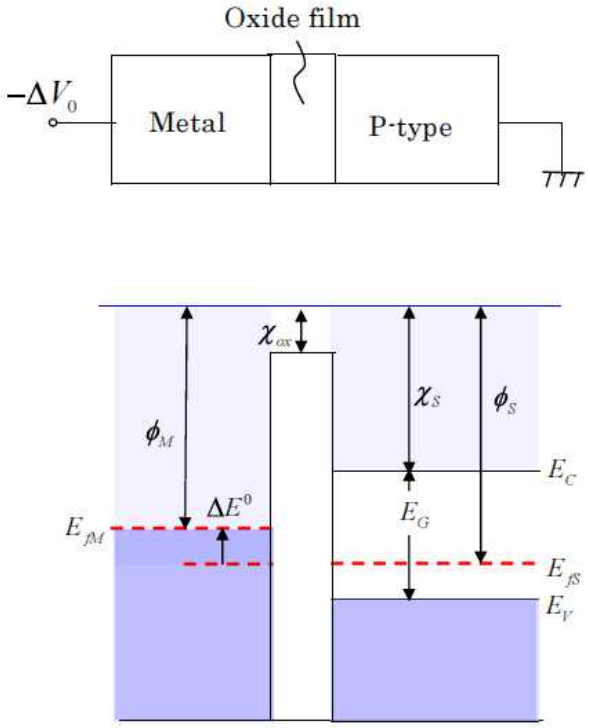


Figure 5.10 Flatbanding of MOS junction by applying $-\Delta V_0$

Applying $-\Delta V_0$ on the metal side corresponds to applying 0 volts in a flat band structure. This means that applying a voltage V_G on the metal side is equivalent to applying $V_G + \Delta V_0$ in the flat band structure. From this it can be said that the V_G versus Q_I characteristic in the general case of non-flatband is obtained by replacing V_G by $V_G + \Delta V_0$ in equation (5.23) given for the flatband structure. That is, the following equation is obtained.

$$Q_I = -C_{ox}(V_G + \Delta V_0 - V_{th}) = -C_{ox}(V_G - V'_{th}) \tag{5.26}$$

$$V_{th}' = V_{th} - \Delta V_0$$
$$= 2\frac{\Delta E_f}{e} + 2\left(\varepsilon_S N_A \Delta E_f\right)^{\frac{1}{2}} \frac{l_{ox}}{\varepsilon_{ox}} - \frac{1}{e}\left(E_{fM} - E_{fS}\right) \tag{5.27}$$

From equation (5.26), in the general case of non-flatband, it can be shown that $V_{th} \rightarrow V_{th}'(= V_{th} - \Delta V_0)$ as in Figure 5.11. According to equation (5.27), it can be said that the desired value of V_{th}' can be obtained by choosing the physical properties and structure of the material.

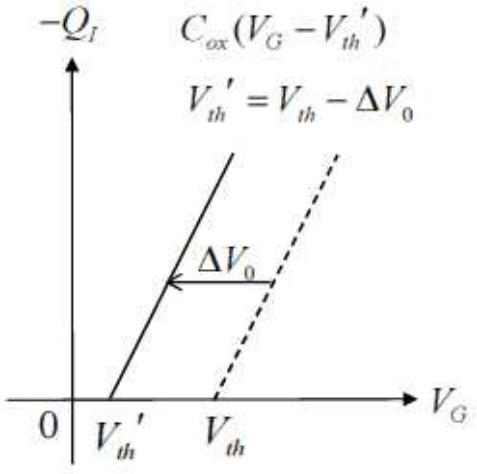


Figure 5.11 V_{th}' at a non-flatband (general case) MOS junction

5.2 MOS junction FET (MOSFET)

Figure 5.12 shows a structural model of an n-channel MOS junction FET (nMOSFET) constructed using a P- type Si semiconductor. The FET is composed of a P- type Si semiconductor (p-Si) with source (S; Source), drain (D; Drain) and gate (G; Gate) terminals on the substrate, as shown in the figure. The S and D terminals are connected to n^+-Si, which has a PN junction with a P-type semiconductor, while the G terminal is connected by a MOS junction across an oxide film. The channel direction of the semiconductor is the y-axis, the right end of the S terminal is the origin ($y = 0$), the left end of the D terminal is $y = L_g$ (channel length L_g) and the

gate width is W_g.

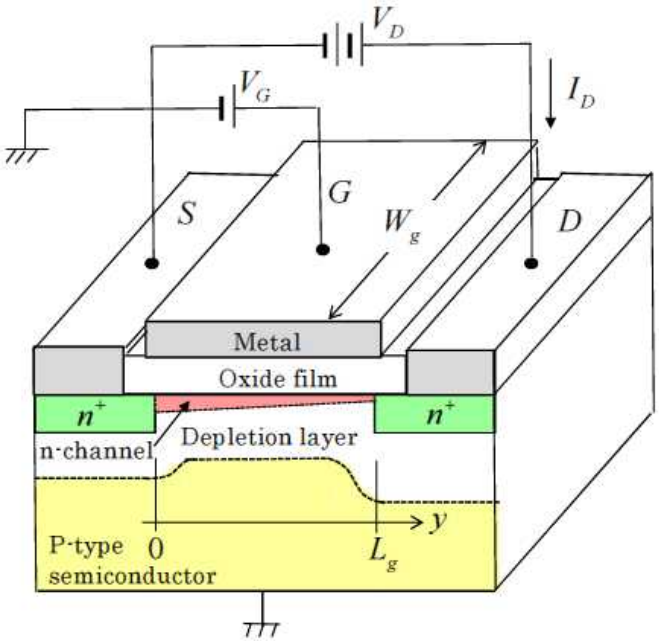


Figure 5.12 Structural model of an n-channel MOS junction FET (nMOSFET)

Consider grounding the S terminal and applying a gate voltage V_G ($V_G \geq V_{th}$) and a small (linear region) drain voltage V_D so that the drain current I_D flows. As in the JFET case, the potential $V(y)$ in the channel is given as a function of position y by the voltage drop. Then the inversion charge $Q_I(y)$ at y is given by

$$Q_I(y) = -C_{ox}(V_G - V_{th} - V(y)) \tag{5.28}$$

The drain current I_D (positive in the direction of flow from drain to source) can be regarded as a drift current due to carriers (conduction electrons) in the channel and is given by

$$I_D = -W_g \mu_e Q_I(y) \frac{dV}{dy} \tag{5.29}$$

From this,

$$I_D dy = W_g \mu_e C_{ox}(V_G - V_{th} - V)dV \tag{5.30}$$

Integrating the left-hand side with $y = 0 \rightarrow L_g$ and the right-hand side with $V = 0 \rightarrow V_D$ $(V(0) \rightarrow V(L_g))$, I_D is obtained as follows.

$$I_D = \frac{W_g \mu_e C_{ox}}{L_g}\left((V_G - V_{th})V_D - \frac{1}{2}V_D^2\right) \tag{5.31}$$

The left-hand integral uses the fact that I_D is constant regardless of location due to the continuity of the current. The equation is valid in $0 \leq V_D \leq V_G - V_{th}$. As V_D increases and $V_D = V_P = V_G - V_{th}$ (where V_P is the pinch-off voltage), the inverted charge at the drain end $(y = L_g)$ becomes $Q_I = 0$, where it pinches off. From equation (5.29), for a finite current I_D to flow when $Q_I = 0$, $\frac{dV}{dy} \rightarrow \infty$ must occur there, and the integration must be done within a range where this does not occur. When $V_D \geq V_P$, the excess voltage is used to expand the depletion layer after the channel disappearance so that, approximately, the configuration of the channel under the gate remains unchanged. From this, I_D becomes a constant value at $V_D \geq V_P$ (the same as in the Shockley model for JFETs). The maximum current $I_{D_{max}}$ for each V_G is obtained when $V_D = V_P = V_G - V_{th}$ and is given by

$$I_{D\max} = \frac{1}{2}\frac{W_g \mu_e C_{ox}}{L_g}(V_G - V_{th})^2 = \frac{1}{2}\frac{W_g \mu_e C_{ox}}{L_g}V_P^2 \tag{5.32}$$

The characteristic of $I_{D_{max}}$ versus V_P given by the above equation is called the pinch-off curve. Figure 5.13 shows the $I_D - V_D$ characteristics of the

MOSFET. The figure also shows the pinch-off curve of $I_{D_{max}}$ vs. V_p. The left side of the pinch-off curve is called the linear region and the right side the saturated region.

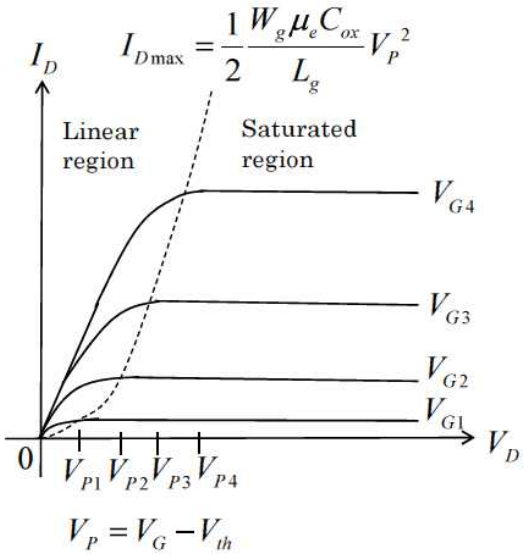


Figure 5.13 $I_D - V_D$ characteristics of MOSFETs

The transconductance g_m in the saturated region is given by

$$g_m = \frac{\partial I_{D\max}}{\partial V_G} = \frac{W_g \mu_e C_{ox}}{L_g}(V_G - V_{th}) \tag{5.33}$$

The gate-source capacitance C_{gs} is given below as the capacitance C_{ox} per unit area of the oxide film section multiplied by the area of the gate section.

$$C_{gs} = L_g W_g C_{ox} \tag{5.34}$$

Figure 5.14 shows a small-signal RF equivalent circuit of a MOSFET in its simplest representation (not including parasitic components).

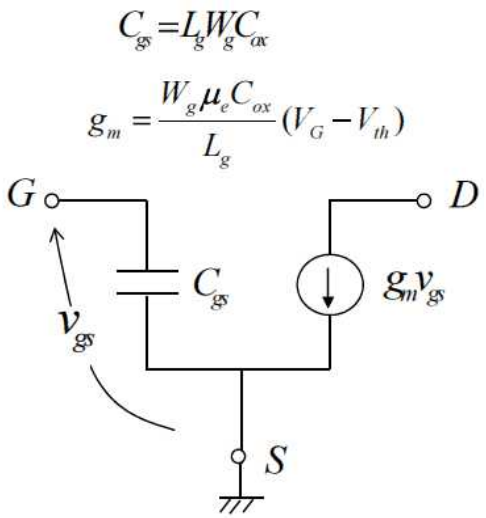


Figure 5.14 Small-signal RF equivalent circuit of a MOSFET (not including parasitic components)

From the equivalent circuit, the source ground current gain β can be expressed as

$$\beta=\frac{g_m}{j\omega C_{gs}}=\frac{1}{j\omega L_gW_gC_{ox}}\frac{W_g\mu_eC_{ox}}{L_g}(V_G-V_{th})=\frac{\mu_e}{j\omega L_g^2}(V_G-V_{th}) \tag{5.35}$$

The cut-off frequency f_T at which $|\beta|=1$ is the following.

$$f_T=\frac{\mu_e}{2\pi L_g^2}(V_G-V_{th}) \tag{5.36}$$

From the above equation, a shorter gate length L_g is necessary to increase β and f_T. By the way, if the device structure is the same, β and f_T can be increased by increasing the gate bias voltage V_G (increasing the drain bias current I_D).