MOS CAPACITOR AND MOSFET

Dr. Mohammed M. Farag

Faculty of Engineering
Alexandria University

Ideal MOS Capacitor

Semiconductor Devices – Physics and Technology
Chapter 5

MOS Capacitor Structure

- The metal-oxide-semiconductor field-effect transistor (MOSFET) is composed of a MOS capacitor and two p-n junctions placed immediately adjacent to the MOS capacitor
- The MOS device 3D view and cross section are shown in Figure, *d* is the thickness of the oxide and V_G is the applied voltage on the metal plate (gate).

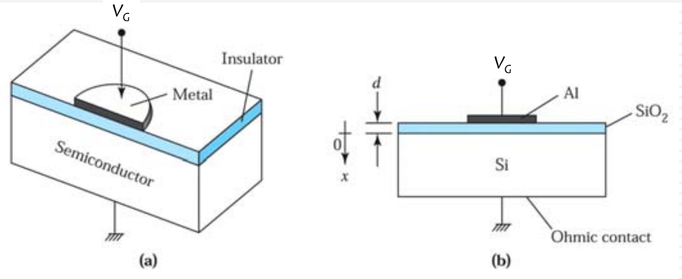


Figure 5.1
© John Wiley & Sons, Inc. All rights reserved.

Band Diagram of an Ideal MOS

- The energy band diagram of an ideal p-type semiconductor MOS at V= 0 is shown in χ
- The work function is the energy difference between the Fermi level and the vacuum level ($q\phi_m$ for the metal and $q\phi_s$ for the semiconductor
- Electron affinity $q\chi$ is the energy difference between the conduction band and vacuum level in semiconductor
- Oxide electron affinity $q\phi_B$ is the energy barrier between the metal and oxide
- $q\psi_B$ is the energy difference between the Fermi level E_f and intrinsic Fermi level E_i

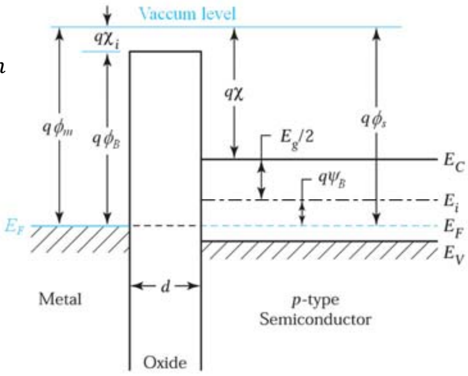


Figure 5.2
© John Wiley & Sons, Inc. All rights reserved.

Ideal MOS

- An ideal MOS is defined as follows: (a) At zero applied bias, the energy difference between the metal work function $q\phi_m$ and the semiconductor work function $q\phi_s$ is zero, or the work function difference $q\phi_{ms}$ is zero:

$$q\phi_{ms} \equiv (q\phi_m - q\phi_s) = q\phi_m - \left(q\chi + \frac{E_g}{2} + q\psi_B\right) = 0\,, \quad (1)$$

- The energy band is flat (flat-band condition) when
  - There is no applied voltage to the gate
  - The only charges that exist in the capacitor under any biasing conditions are those in the semiconductor and those with equal but opposite sign on the metal surface adjacent to the oxide, i.e. no charges in the SiO_2
  - There is no carrier transport through the oxide under direct current (dc)-biasing conditions, or the resistivity of the oxide is infinite
  - This is the ideal MOS capacitor

Ideal MOS Biasing

- When an ideal MOS capacitor is biased with a gate voltage, three cases can exist at the semiconductor surface:
  - Accumulation ($V_G < 0$): excess positive carriers (holes) will be accumulated at the SiO_2-Si interface and no current flows.
  - Depletion ($0 < V_G < V_T$): excess negative carriers will be induced at the SiO_2-Si interface but the number of majority carriers is still larger
  - Inversion ($V_G > V_T$): number of excess negative minority carriers will exceed the number of majority carriers near the sio_2-Si interface

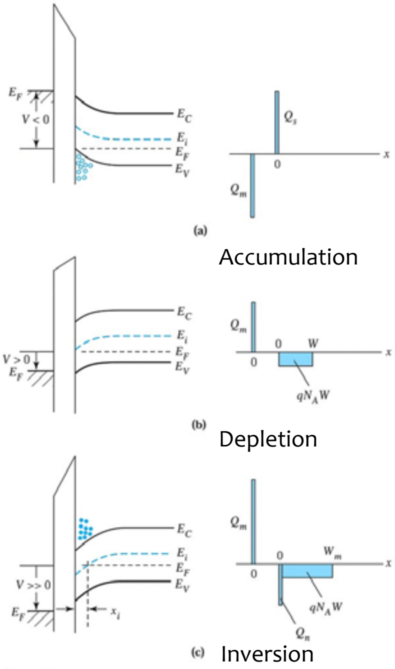
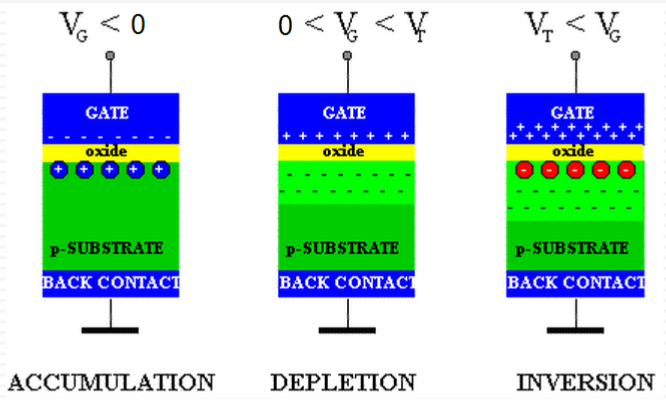


Figure 5.3

Ideal MOS Biasing (2)



Accumulation

- For a p-type semiconductor, when a negative voltage ($V_G < 0$) is applied to the metal plate, excess positive carriers (holes) will be induced at the SiO_2-Si interface and the energy bands near the semiconductor surface are bent upward
- For an ideal MOS capacitor, no current flows in the device regardless of the value of the applied voltage, therefore, the Fermi level in the semiconductor remains constant
- The carrier density in the semiconductor depends exponentially on the energy difference $E_i - E_F$:

$$p_p = n_i e^{(E_i - E_F)/kT} \qquad (2)$$

- The hole concentration increases as (E_i-E_f) increases with further decreasing V_G
- The corresponding charge distribution is shown in the previous Figure where Q_s is the positive charge per unit area in the semiconductor and Q_m is the negative charge per unit area in the metal where $|Q_m| = Q_s$

Depletion

- When a small positive voltage ($V_G > 0$) is applied to the gate, the energy bands near the semiconductor surface are bent downward and the majority carriers (holes) are depleted
- The space charge per unit area, Q_{sc}, in the semiconductor is equal to qN_AW, where W is the width of the depletion region

Inversion

- When a larger positive voltage is applied, the energy bands bend downward even more so that the intrinsic level E_i at the surface crosses over the Fermi level, as shown in Figure C
- As a result, the positive gate voltage starts to induce excess negative carriers (electrons) at the SiO_2-Si interface
- The electron concentration in the semiconductor depends exponentially on the energy difference $E_F - E_i$, and is given by

$$n_p = n_i e^{(E_F - E_i)/kT} \qquad (3)$$

- In this case, the electron concentration n_p at the interface is larger than n_i, and the hole concentration is less than n_i
- The number of electrons (minority carriers) at the surface is greater than holes (majority carriers);

Inversion (2)

- □ Initially, the surface is in a weak inversion condition since the electron concentration is small
- □ As the bands are bent further, the conduction band E_C edge comes close to the Fermi level E_F
- □ The case of strong inversion occurs when the electron concentration near the SiO_2-Si interface is equal to the substrate doping level of the holes N_A
- □ After this point most of the additional negative charges in the semiconductor consist of the charge Q_n in a very narrow n-type inversion layer $0 \le x \le x_i$, where x_i is the width of the inversion region

Strong Inversion

- □ Once strong inversion occurs, the surface depletion-layer width reaches a maximum
- □ This is because when the bands are bent downward far enough for strong inversion to occur, even a very small increase in band bending corresponding to a very small increase in depletion-layer width results in a large increase in the charge Q_n in the inversion layer
- □ Thus, under a strong inversion condition the charge per unit area Q_s in the semiconductor is the sum of the charge Q_n in the inversion layer and the charge Q_{sc} in the depletion region:

$$Q_s = Q_n + Q_{sc} = Q_n - qN_A W_m \quad (4)$$

where W_m is the maximum width of the surface depletion region

Electrostatic Potential

- In Figure, we define a potential ψ at any point x relative to the equilibrium position of E_i
- Ψ represents the band bending
  - Ψ=0 in the bulk of the semiconductor (x>>0)
  - $\psi = \psi_s$ at the semiconductor surface (x=0), the surface potential
- We can express electron and hole concentrations as a function of ψ

$$n_p = n_i e^{q(\psi - \psi_B)/kT}, \qquad (5)$$

$$p_p = n_i e^{q(\psi_B - \psi)/kT},$$

where ψ is positive when the band is bent downward

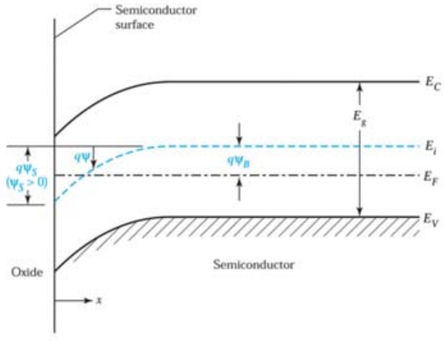


Figure 5.4
© John Wiley & Sons, Inc. All rights reserved.

Electrostatic Potential and Poisson's Equation

- At the surface, the densities are $n_s = n_i e^{q(\psi_s - \psi_B)/kT}$, (6)

$$p_s = n_i e^{q(\psi_B - \psi_s)/kT}.$$

- We can describe different operating regions with the aid of ψ_s
  - $\psi_s < 0$ Accumulation of holes (bands bend upward)
  - $\psi_s = 0$ Flat-band condition
  - $\psi_B > \psi_s > 0$ Depletion of holes (bands bend downward)
  - $\psi_s = \psi_B$ Midgap with $n_s = n_p = n_i$ (intrinsic concentration)
  - $\Psi_s > \psi_B$ Inversion (bands bend downward)
- The potential ψ as a function of distance x can be obtained by solving the one-dimensional Poisson's equation:

$$\frac{d^2\psi}{dx^2} = \frac{-\rho_s(x)}{\varepsilon_s}, \qquad (7)$$

where $\rho_s(x)$ is the charge density per unit volume at position x and ε_s, is the dielectric permittivity

Solving Poisson's Equation

$$\frac{d^2\psi}{dx^2} = -\frac{\rho_s(x)}{\epsilon_s}$$

We already know that $\rho_s(x) = q(N_D^+ - N_A^- + p_p - n_p)$ (I)

Away from the surface, ρ_s=0 $\Rightarrow N_D^+ - N_A^- = n_0 - p_0$ (II)

Also we know that $p_p = n_i e^{\frac{q}{kT}(\psi_B - \psi)} = p_0 e^{\frac{-q\psi}{kT}}$

$, n_p = n_i e^{\frac{q}{kT}(\psi - \psi_B)} = n_0 e^{\frac{q\psi}{kT}}$ (III)

where $p_0 = n_i e^{\frac{q\psi_B}{kT}}, n_0 = n_i e^{\frac{-q\psi_B}{kT}}$

Sub (II), (III) in (I): $\rho_s(x) = q\left(p_0(e^{\frac{-q\psi}{kT}} - 1) - n_0(e^{\frac{q\psi}{kT}} - 1)\right)$

$$\therefore \frac{d^2\psi}{dx^2} = -\frac{q}{\epsilon_s}\left(p_0(e^{\frac{-q\psi}{kT}} - 1) - n_0(e^{\frac{q\psi}{kT}} - 1)\right)$$

Also we remember that $\mathcal{E} = -\frac{d\psi}{dx}$

Solving Poisson's Equation (2)

$$\Rightarrow \frac{d^2\psi}{dx^2} = -\frac{d\mathcal{E}}{dx} = -\frac{d\mathcal{E}}{d\psi}\frac{d\psi}{dx} = \mathcal{E}\frac{d\mathcal{E}}{d\psi}$$

$$\therefore \mathcal{E}\frac{d\mathcal{E}}{d\psi} = -\frac{q}{\epsilon_s}\left(p_0(e^{\frac{-q\psi}{kT}} - 1) - n_0(e^{\frac{q\psi}{kT}} - 1)\right)$$

Integrate the last equation to get the electric field $\mathcal{E}$

$$\int_0^{\mathcal{E}} \mathcal{E}\,d\mathcal{E} = \int_0^{\psi} -\frac{q}{\epsilon_s}\left(p_0(e^{\frac{-q\psi}{kT}} - 1) - n_0(e^{\frac{q\psi}{kT}} - 1)\right) d\psi$$

$$\Rightarrow \mathcal{E}^2 = (\frac{kT}{q})^2\left(\frac{q^2 p_0}{2\epsilon_s kT}\right)\left[(e^{\frac{-q\psi}{kT}} + \frac{q}{kT}\psi - 1) + \frac{n_0}{p_0}(e^{\frac{q\psi}{kT}} - \frac{q}{kT}\psi - 1)\right]$$

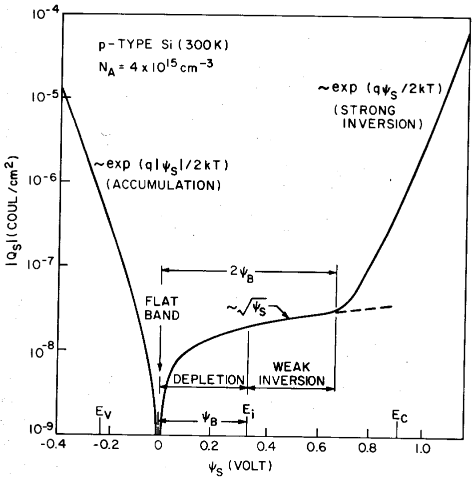
Let $\beta = \frac{q}{kT}$, $L_D = \sqrt{\frac{kT\epsilon_s}{p_0 q^2}} = \sqrt{\frac{\epsilon_s}{q p_0 \beta}}$ (Debye length)

$$F\left(\beta\psi, \frac{n_0}{p_0}\right) = \left[(e^{-\beta\psi} + \beta\psi - 1) + \frac{n_0}{p_0}(e^{\beta\psi} - \beta\psi - 1)\right]^{1/2}$$

$$\Rightarrow \mathcal{E} = \pm\frac{\sqrt{2}kT}{qL_D}F\left(\beta\psi, \frac{n_0}{p_0}\right)$$

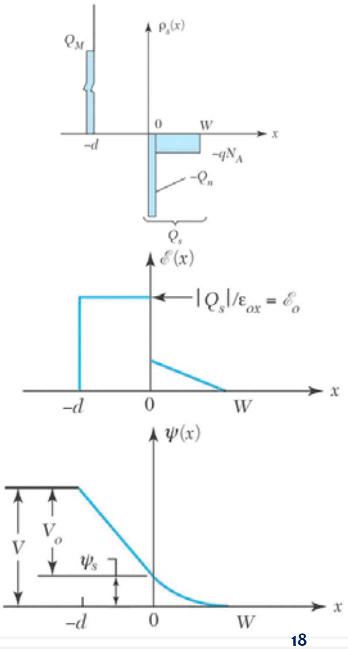
Surface Charge

- Use Gauss's law to find surface charge per unit area
  $$Q_s = -\epsilon_s \mathcal{E}_s$$
  - When ψ_s is negative (accumulation), the first term $e^{-\beta\psi_s}$ dominates the electric field equation
  - At ψ_s= 0 (flat band)there is no space charge
  - When ψ_s is positive, initially the linear term $\beta\psi$ in the electric field equation dominates as a result of the exposed, immobile dopants
  - In depletion region, as ψ gets larger, the term $\frac{n_0}{p_0}e^{\beta\psi} \to 1$, the term $e^{\beta\psi}$ dominates the equation



- $$\left[\left(e^{-\beta\psi}+\beta\psi-1\right)+\frac{n_0}{p_0}\left(e^{\beta\psi}-\beta\psi-1\right)\right]^{1/2}$$

Charge, Field, and Potential



- The electric field does not penetrate the metal
- It is constant across the oxide as there are no charges in the oxide
- The electric field in the semiconductor drops linearly, as we would expect
- Charge on metal = induced surface charge in semiconductor
  $$Q_M = qN_AW - Q_n = -Q_S$$
- The potential is constant in the metal
- It drops linearly across the oxide
- The potential also drops across the depletion region of the semiconductor

The Depletion Region

- In a depletion region with width *W*, the charge in the semiconductor is approximately given by $\rho_s = -qN_A$
- Integrating Poisson's equation gives the potential ψ_s as a function of distance *x* in the depletion region, where the boundary conditions are $(\mathcal{E} = -\frac{d\psi}{dx} = 0\ \&\ \psi = 0\ \ at\ x = w)$

$$\psi = \frac{qN_AW^2}{2\epsilon_s}(1-\frac{x}{W})^2 \qquad (8)$$

- The electrostatic potential $\psi = \psi_s$ at *x*=0:

$$\Rightarrow \psi_s = \frac{qN_AW^2}{2\epsilon_s} \qquad (9)$$

- The surface is inverted whenever ψ_s is larger than ψ_B
- In strong inversion, the electron concentration at the surface is equal to the substrate impurity concentration

$$n_s = N_A = n_i e^{\frac{q\psi_B}{kT}} \quad \Rightarrow \psi_B = \frac{kT}{q}\ln(\frac{N_A}{n_i})$$

The Depletion Region (2)

- Using the previous equation ,we get

$$\psi_s(inv) = 2\psi_B = 2\frac{kT}{q}\ln(\frac{N_A}{n_i}) \qquad (10)$$

- This equation states that a potential ψ_B is required to bend the energy bands down to the intrinsic condition at the surface ($E_i = E_F$), and bands must then be bent downward by another ψ_B at the surface to obtain the strong inversion
- Use the p-n junction approximation to calculate the depletion width $W = \left[\frac{2\varepsilon_s\psi_s(N_A+N_D)}{qN_AN_D}\right]^{1/2} \approx \left[\frac{2\varepsilon_s\psi_s}{qN_A}\right]^{1/2}$ for p-type MOS
- The depletion layer reaches a maximum when the surface is strongly inverted and the maximum width of the surface depletion region W_m is given by substituting $\psi_s = 2\psi_B$ in eqn (8)

$$W_m = \sqrt{\frac{2\varepsilon_s\psi_s(inv)}{qN_A}} = 2\sqrt{\frac{\varepsilon_s\psi_B}{qN_A}} \qquad (11)$$

MOS Capacitance

- In the absence of any work function differences, the applied voltage will be divided on the oxide and semiconductor

$$V = V_o + \psi_s, \qquad (13)$$

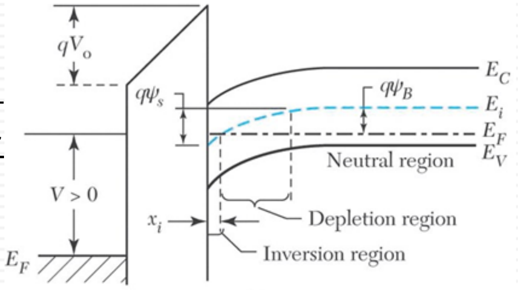
where V_o is the potential across the oxide and is given by

$$V_o = \mathscr{E}_o d = \frac{|Q_s|\, d}{\varepsilon_{ox}} \equiv \frac{|Q_s|}{C_o} \qquad (14)$$

where $\mathcal{E}_o$ is the field in the oxide, $C_o = \varepsilon_{ox}/d$) is the oxide capacitance per unit area

$$|Q_S| = qN_A W = qN_A \sqrt{\frac{2\varepsilon_S \psi_S}{qN_A}}$$

$$= \sqrt{2qN_A \varepsilon_S \psi_S}$$



MOS Capacitance

- The total capacitance C_o f the MOS capacitor is a series combination of the oxide capacitance C_o and the semiconductor depletion-layer capacitance C_j

$$C = \frac{C_o C_j}{(C_o + C_j)} \ \mathrm{F/cm^2}, \qquad (15)$$

where $Cj = \varepsilon_s/W$, the same as for an abrupt p-n junction

- From eqs. (9), (13), (14) we can eliminate *W* and obtain C by sub in eqn (15)

$$\frac{C}{C_o} = \frac{1}{\sqrt{1 + \frac{2\varepsilon_{ox}^2 V}{qN_A \varepsilon_s d^2}}}, \qquad (16)$$

which states that the capacitance decreases with increasing metal-plate voltage while the surface is being depleted

- When the applied voltage is negative, there is no depletion region, and we have an accumulation of holes at the semiconductor surface and thus the total capacitance is close to the oxide capacitance ε_{ox}/d

MOS Capacitance (2)

- When strong inversion occurs, the width of the depletion region will not increase with increasing the applied voltage
- This condition takes place at a metal-plate voltage that causes the surface potential ψ_s to reach $2\psi_B$
- Substitute $\psi_s = 2\psi_B$ in Eq. 13 and note that the corresponding charge per unit area is qN_AW_m yields the metal-plate voltage at the case of strong inversion
- This voltage is called the threshold voltage V_T:

$$\boxed{V_T = \frac{qN_AW_m}{C_o} + \psi_s(inv) \cong \frac{\sqrt{2\varepsilon_s qN_A(2\psi_B)}}{C_o} + 2\psi_B\ .} \qquad (17)$$

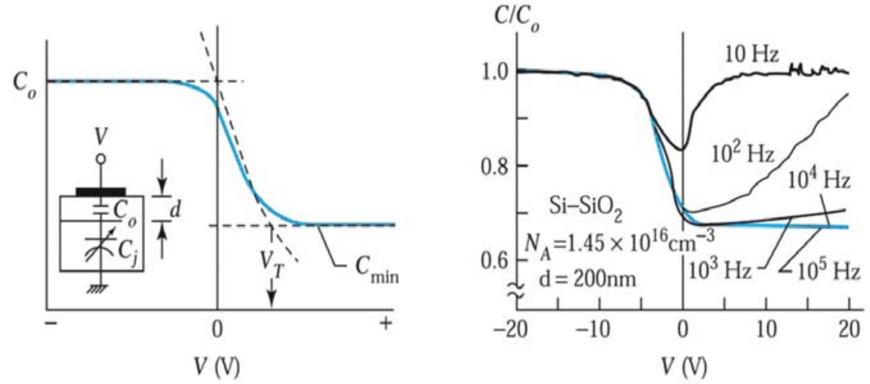
MOS Capacitance (3)

- Once the strong inversion takes place, the total capacitance will remain at a minimum value given by Eq 15 with $C_j = \varepsilon_s/W_m$,

$$C_{\min} = \frac{\varepsilon_{ox}}{d + \left(\varepsilon_{ox}/\varepsilon_s\right)W_m} \qquad (18)$$

- After inversion is reached, the small signal capacitance depends on whether the measurements are made at high (typically~1MHz) or low (typically- 1-100Hz) frequency, where "high“ and "low“ are w.r.t the generation-recombination rate of the minority carriers in the inversion layer
- If signal applied to make measurement is too fast, inversion layer carriers can’t respond and do not contribute to the charge
- Slowly varying signals allow time for minority carriers to be generated, drift across depletion region, or recombine
- Majority carriers in the accumulation region respond much faster

Ideal MOS C-V Characteristics



HF C-V Characteristics

LF C-V Characteristics

Figure 5.7
© John Wiley & Sons, Inc. All rights reserved.

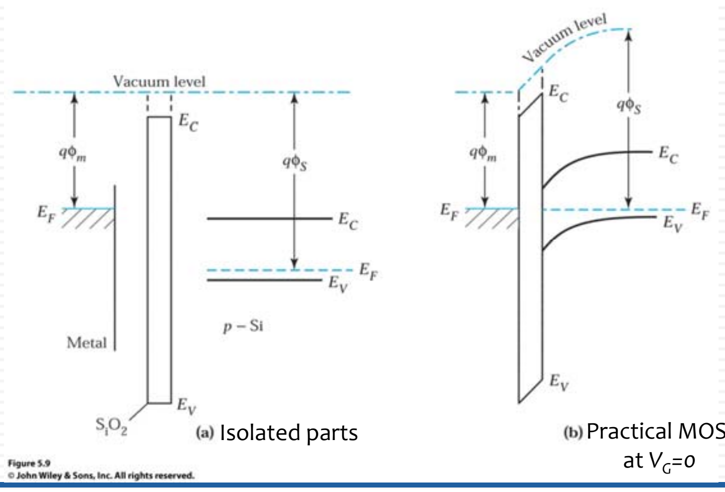
Practical MOS Capacitor

The Work Function Difference

- ☐ The metal-SiO_2-Si is the most studied MOS capacitance
- ☐ The electrical characteristics of the SiO_2-Si system approach those of the ideal MOS capacitor
- ☐ However, for commonly used metal electrodes, the work function difference $q\phi_{ms}$ is generally not zero, and there are various charges inside the oxide or at the SiO_2-Si interface that will affect the ideal MOS characteristics
- ☐ The work function of a semiconductor $q\phi_s$ and consequently the work function $q\phi_{ms}(= q\phi_m - q\phi_s)$ varies with the doping concentration
- ☐ The ϕ_{ms} can vary over a 2V range depending on the electrode materials (Aluminum, p^+Polysilicon, or n^+ Polysilicon) and the silicon doping concentration

Energy Band Diagram of Practical MOS Capacitance

- ☐ In the isolated situation shown in Figure (a) , all bands are flat; this is the flat-band condition



(a) Isolated parts

(b) Practical MOS at V_G=0

Figure 5.9
© John Wiley & Sons, Inc. All rights reserved.

Energy Band Diagram of Practical MOS Capacitance (2)

- At thermal equilibrium, the Fermi level must be a constant and the vacuum level must be continuous
- To accommodate the work function difference, the semiconductor bands bend downward, as shown in Figure
- Thus, the metal is positively charged and the semiconductor surface is negatively charged at thermal equilibrium
- To achieve the ideal flat-band condition of the ideal MOS capacitor, we have to apply a voltage equal to the work function difference $q\phi_{ms}$ to the metal
- The flat-band voltage is negative where $V_{FB} = \phi_{ms}$
- In ideal MOS equations the voltage *V* must be replaced with (*V*-V_{FB})

$$V = V_{FB} + V_o + \psi_s \qquad (13`)$$

- The new expression of the threshold voltage V_T will be

$$V_T = V_{FB} + \frac{\sqrt{2\varepsilon_s q N_A (2\psi_B)}}{C_0} + 2\psi_B \qquad (17`)$$

Interface Traps and Oxide Charges

- In addition to the work function difference, the MOS capacitor is affected by charges in the oxide and traps at the SiO_2-Si interface
- They are the interface trapped charge, fixed-oxide charge, oxide-trapped charge, and mobile ionic charge
- The basic classification of these traps and charges are shown in the Figure

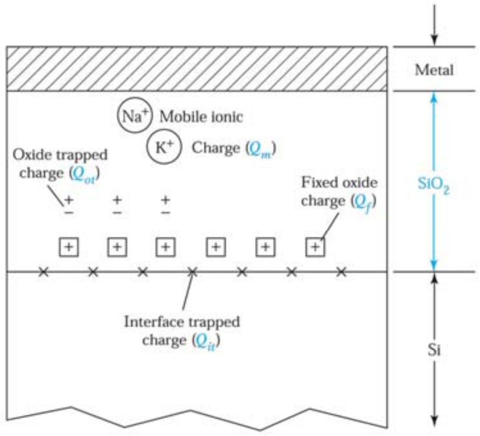
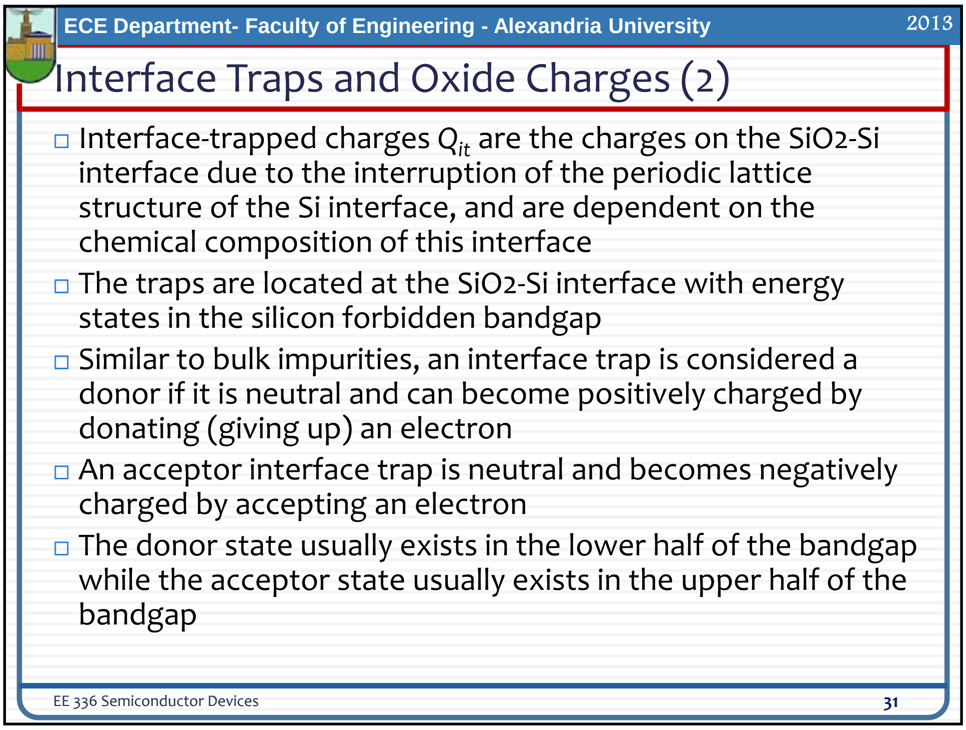
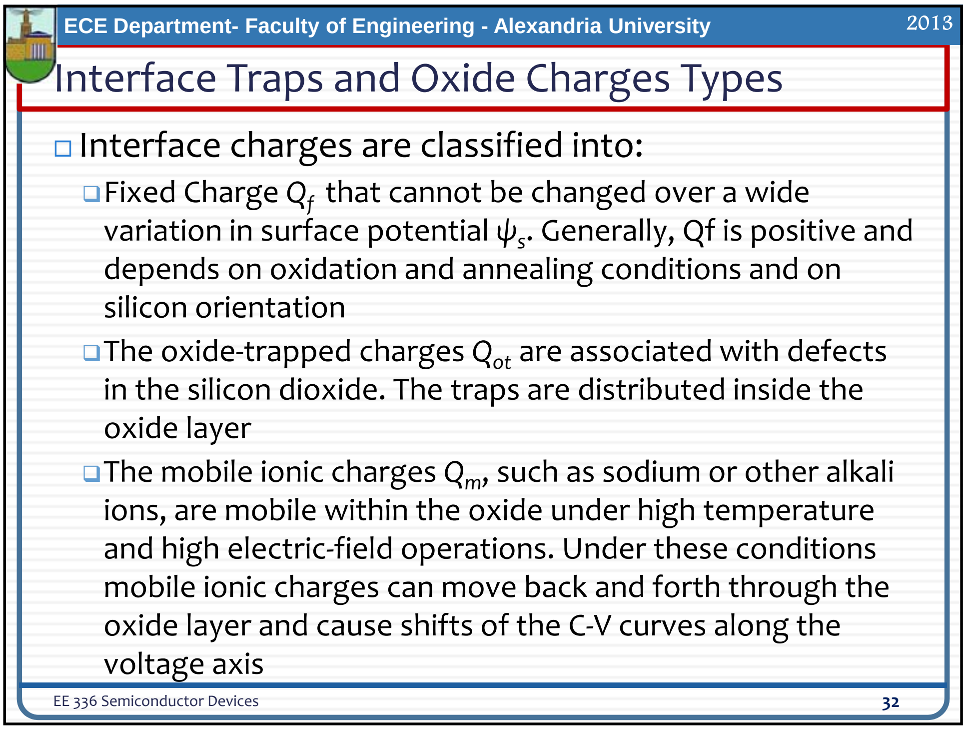


Figure 5.10
© John Wiley & Sons, Inc. All rights reserved.



Interface Traps and Oxide Charges (2)

- Interface-trapped charges Q_{it} are the charges on the SiO2-Si interface due to the interruption of the periodic lattice structure of the Si interface, and are dependent on the chemical composition of this interface
- The traps are located at the SiO2-Si interface with energy states in the silicon forbidden bandgap
- Similar to bulk impurities, an interface trap is considered a donor if it is neutral and can become positively charged by donating (giving up) an electron
- An acceptor interface trap is neutral and becomes negatively charged by accepting an electron
- The donor state usually exists in the lower half of the bandgap while the acceptor state usually exists in the upper half of the bandgap



Interface Traps and Oxide Charges Types

- Interface charges are classified into:
  - Fixed Charge Q_f that cannot be changed over a wide variation in surface potential ψ_s. Generally, Qf is positive and depends on oxidation and annealing conditions and on silicon orientation
  - The oxide-trapped charges Q_{ot} are associated with defects in the silicon dioxide. The traps are distributed inside the oxide layer
  - The mobile ionic charges Q_m, such as sodium or other alkali ions, are mobile within the oxide under high temperature and high electric-field operations. Under these conditions mobile ionic charges can move back and forth through the oxide layer and cause shifts of the C-V curves along the voltage axis

Measuring Interface Traps and Oxide Charges

- Figure shows an interface-trap system consisting of both acceptor states and donor states, in which the states above a neutral level E_o are of the acceptor type and those below E_o are of the donor type
- To calculate the trapped charge, we can assume that at room temperature, the occupancy takes on the value of 0 and 1 above and below E_F
- With these assumptions, the interface-trapped charges Q_{it} can now be easily calculated by:

$$Q_{it} = -q\int_{E_0}^{E_F} D_{it}\,dE \qquad E_F \text{ above } E_0,$$

$$= +q\int_{E_F}^{E_0} D_{it}\,dE \qquad E_F \text{ below } E_0,$$

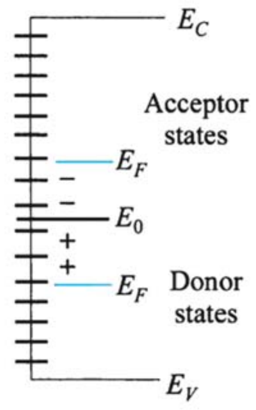


Figure 5.11
© John Wiley & Sons, Inc. All rights reserved.

Measuring Interface Traps and Oxide Charges

where D_{it} is the interface trap density and Q_{it} is the effective net charges per unit area (i.e., C/cm^2)

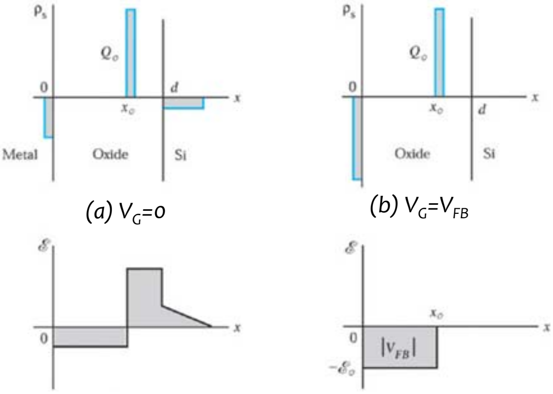
- The interface trap levels are distributed across the energy bandgap and the interface trap density is given by

$$D_{it} = \frac{1}{q}\frac{dQ_{it}}{dE} \qquad \text{Number of traps / cm}^2\text{ - eV} \qquad (20)$$

- The method used to determine D_{it} experimentally is to measure the change of Q_{it} w.r.t the change of E_F or surface potential ψ_s
- When a gate voltage is applied, the Fermi level moves up or down with respect to the interface-trap levels and a change of charge in the interface traps occurs
- This change of charge affects the MOS capacitance and alters the ideal MOS curve

Oxide Charges and V_{FB}

- Assume a positive sheet charge per unit area, Q_o, within the oxide, as shown in Figure
- This positive sheet charge will induce negative charges partly in the metal and semiconductor
- The resulting field distribution, obtained from integrating Poisson's equation is shown in Figure (a) for $q\phi_{ms} = 0$, $V_G=0$
- To reach the flat-band condition, we must apply a negative voltage to the metal as shown in Figure (b)



(a) $V_G=0$ (b) $V_G=V_{FB}$

Oxide Charges and V_{FB}

- As the negative voltage increases, more negative charges are put on the metal and thereby the electric-field distribution shifts downward until the electric field at the semiconductor surface is zero
- Under this condition, the area contained under the electric-field distribution corresponds to the flat-band voltage V_{FB}

$$V_{FB} = -\mathcal{E}_o x_o = -\frac{Q_o}{\varepsilon_{ox}} x_o = -\frac{Q_o}{C_o}\frac{x_o}{d} \qquad (21)$$

- The flat-band voltage is dependent on both the density of the sheet charge Q_o and its location x_o within the oxide
  - If the charge Q_o is located at the metal-SiO_2 interface $x_o=0$ → $V_{FB}=0$
  - If the charge Q_o is located at the Si-SiO_2 interface $x_o=d$ → $V_{FB} = -\frac{Q_0}{C_0}$

Oxide Charges and V_{FB} (2)

- For the more general case of an arbitrary space charge distribution within the oxide, the flat-band voltage is given by

$$\boxed{V_{FB} = \frac{-1}{C_o}\left[\frac{1}{d}\int_0^d x\rho(x)\,dx\right],} \qquad (23)$$

where $\rho(x)$ is the volume charge density in the oxide

- If the value of the work function difference $q\phi_{ms}$ is not zero:

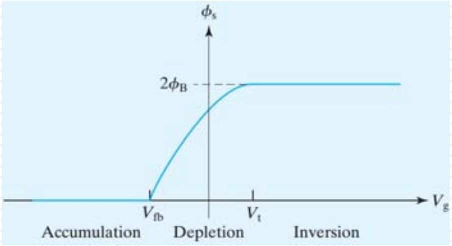
$$V_{FB} = \phi_{ms} - \frac{1}{C_o}\left[\frac{1}{d}\int_0^d x\rho(x)dx\right]$$

Review of Basic MOS Capacitor Theory and C-V Characteristics

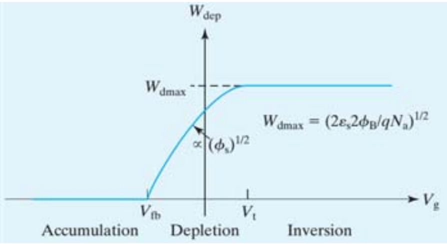
Modern Semiconductor Devices for Integrated Circuits
Chapter 5

Review: Basic MOS Theory

- The surface potential, ψ_s, is zero at V_{FB} and approximately zero in the accumulation region
- As V_G increases from V_{FB} into the depletion regime, ψ_s increases from zero towards $2\psi_B$
- When ψ_s reaches $2\psi_B$, the surface electron concentration becomes so large that the surface is considered inverted
- The V_G at that point is called V_T, the threshold voltage



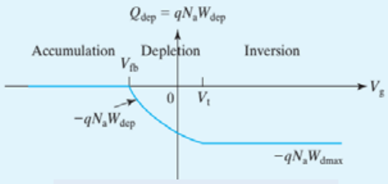
Surface Potential versus gate voltage



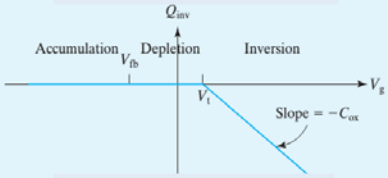
Depletion width versus gate voltage

Charge Components

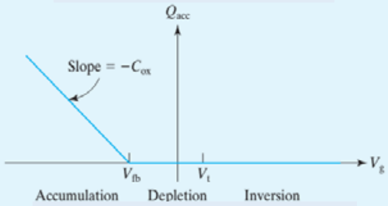
- Figure reviews the three charge components in the substrate
- The depletion charge Q_{dep} is constant in the inversion region because W_{dep} is a constant there
- $Q_{inv} = -C_{ox}(V_G - V_T)$ appears in the inversion region
- Q_{acc} shows up in the accumulation region



(a) Depletion layer charge



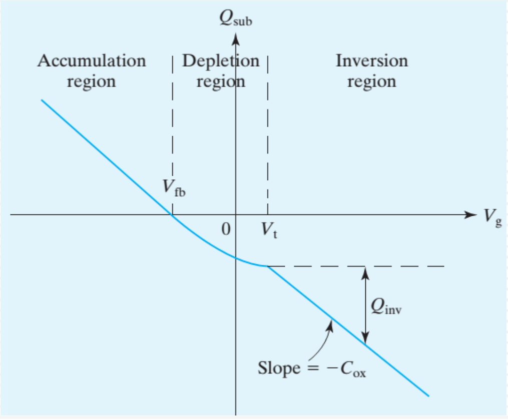
(b) Inversion layer charge



(c) Accumulation layer charge

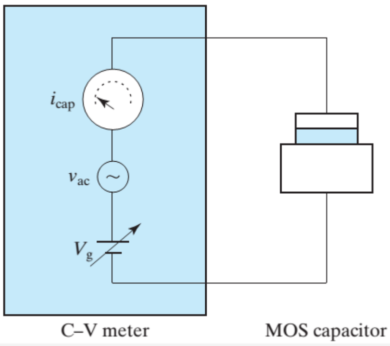
Total Substrate Charge

- Figure shows the total substrate charge, Q_{sub}
- Q_{sub} in the accumulation region is made of accumulation charge
- Q_{sub} is made of Q_{dep} in the depletion region
- In the inversion region, there are two components, Q_{dep} that is a constant and Q_{inv} that is equal to $-C_{ox}(V_G - V_T)$



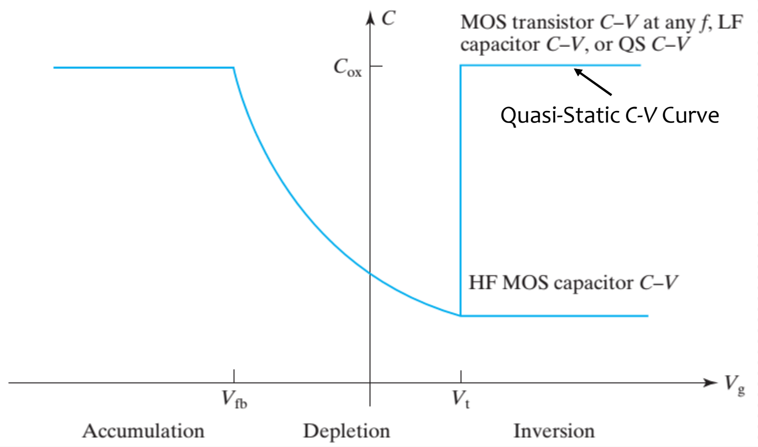
MOS C-V Characteristics

- The capacitance–voltage (*C–V*) measurement is a powerful and commonly used method of determining the gate oxide thickness, substrate doping concentration, threshold voltage, and flat-band voltage
- The *C–V* curve is usually measured with a *C–V* meter which applies a DC bias voltage, *Vg* , and a small sinusoidal signal (*1 kHz–10 MHz*) to the MOS capacitor and measures the capacitive current with an AC ammeter. The capacitance is calculated from $i_{cap}/v_{ac} = \omega C$



$$C \equiv \frac{dQ_g}{dV_g} = -\frac{dQ_{sub}}{dV_g}$$

The Quasi-Static MOS C–V Characteristics



Quasi-Static C-V Curve

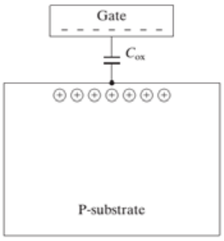
MOS Capacitor in All Bias Regions

- ☐ Capacitance eqs:

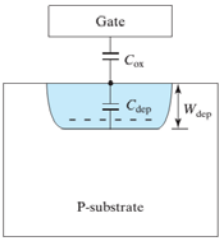
$$C_{dep} = \frac{\varepsilon_s}{W_{dep}}$$

$$\frac{1}{C} = \frac{1}{C_{ox}} + \frac{1}{C_{dep}}$$

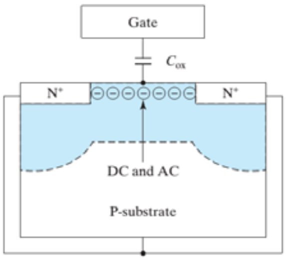
$$\frac{1}{C} = \sqrt{\frac{1}{C_{ox}^2} + \frac{2(V_g - V_{fb})}{qN_a\varepsilon_s}}$$



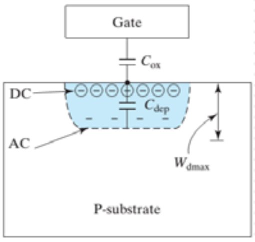
(a) Accumulation



(b) Depletion



(C) Inversion in MOSFET



(d) Inversion in MOS Capacitance

MOS Capacitor in All Bias Regions (2)

- In the accumulation region, the MOS capacitor is just a simple capacitor with capacitance C_{ox} as shown in Figure
- In the depletion region, the MOS capacitor consists of two capacitors in series: the oxide capacitor, C_{ox}, and the depletion-layer capacitor, C_{dep}. Under the AC small-signal voltage, W_{dep} expands and contracts slightly at the AC frequency. Therefore, the AC charge appears at the bottom of the depletion layer
- Fig (c) shows that an inversion layer exists at the Si–SiO2 interface. In response to the AC signal, Q_{inv} increases and decreases at the AC frequency. The inversion layer plays the role of the bottom electrode of the capacitor. Therefore, C reverts to C_{ox} in the inversion region
- This *C–V* curve is called the quasi-static *C–V* because Q_{inv} can respond to the AC signal as if the frequency is low (<100 Hz)

High-Frequency MOS Capacitor C-V Curve

- If the p-n junctions are not present, the p-type substrate is an inefficient supplier of electrons. It produces electrons through thermal generation at a very slow rate and Q_{inv} cannot respond to the AC signal and remains constant at its DC value
- Instead, the AC signal causes ψ_s to oscillate around $2\psi_B$ and causes W_{dep} to expand and contract slightly around W_{dmax}
- This change of W_{dep} can respond at very high frequencies because it only involves the movement of the abundant majority carriers and, consequently, the AC charge exists at the bottom of the depletion region
- The result is a saturation of C at V_T as illustrated by the lower curve as show in the MOS C-V characteristics slide.
- This curve is known as the capacitor *C–V* or the high-frequency MOS capacitor *C–V* (HF *C–V*).

MOSFET Fundamentals

Metal-Oxide Semiconductor Field-Effect Transistor

- ☐ The n-channel MOSFET shown in Figure is a four terminal device consisting of a p-type semiconductor substrate in which two n^+ regions, the source and drain, are formed
- ☐ The metal plate on the oxide is called the gate and the fourth contact is the substrate
- ☐ The basic device parameters are the channel length L, width Z, the oxide thickness d, the junction depth r_j, and the substrate doping N_A

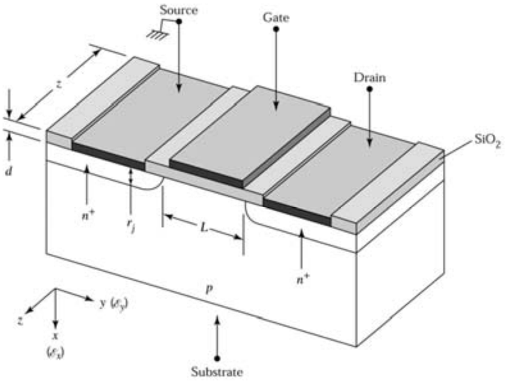


Figure 5.21
© John Wiley & Sons, Inc. All rights reserved.

MOSFET Operation

- Modes of operation are:
  - Cut-off mode
  - Linear mode
  - Saturation mode
- Cut-off Region:
  - when no voltage is applied to the gate, the source-to-drain electrodes correspond to two p–n junctions connected back to back
  - For $V_G<V_T$ (in accumulation and depletion mode), there is no charge carriers that can conduct current between source and drain
    → $I_D=0$

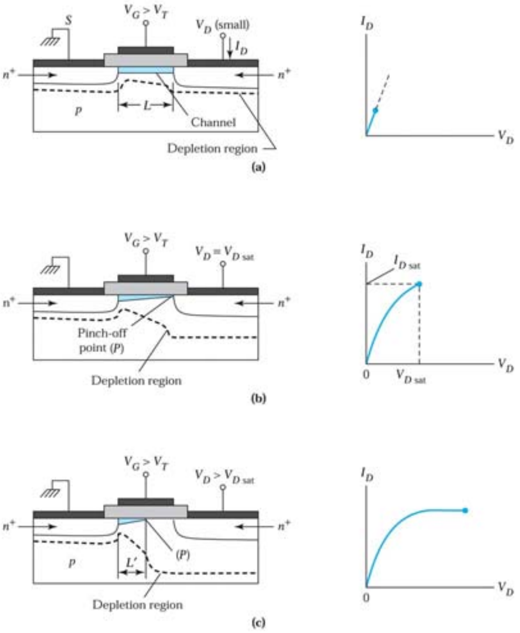


Figure 5.22
© John Wiley & Sons, Inc. All rights reserved.

Linear Region

- If $V_G>V_T$, inversion occurs at the semiconductor surface
- If a small drain voltage is applied, electrons will flow from the source to the drain through the conducting channel
- Thus, the channel acts as a resistor, and the drain current I_D is proportional to the drain voltage
- The conductance of this channel can be modulated by varying the gate voltage

Saturation Region

- When the drain voltage increases to V_{dsat} (the thickness of the inversion layer x_i near $y= L$ is reduced to zero); this is called the pinch-off point P
- Beyond the pinch-off point, the drain current remains essentially the same, because for $V_D> V_{dsat}$, at point P the voltage V_{dsat} remains the same
- Thus, the number of carriers arriving at point P from the source or the current flowing from the drain to the source remains the same
- This is the saturation region, since ID is a constant regardless of an increase in the drain voltage
- The major change is the decrease of *L* to the value *L*` shown in Figure
- Carrier injection from P into the drain depletion region is similar to that of carrier injection from an emitter-base junction to the base-collector depletion region of a bipolar transistor

MOSFET *I-V* Characteristics

- We will derive the MOSFET I-V characteristics under the following ideal conditions:
  a) The gate structure corresponds to an ideal MOS capacitor, that is, there are no interface traps, fixed-oxide charges, or work function differences
  b) Only drift current is considered
  c) Carrier mobility in the inversion layer is constant
  d) Doping in the channel is uniform
  e) Reverse-leakage current is negligibly small
  f) The transverse field created by the gate voltage ($\varepsilon_x$in the *x*-direction, which is perpendicular to the current flow) in the channel is much larger than the longitudinal field created by the drain voltage (ε_y in the y-direction, which is parallel to the current flow)

Linear region

- The total charge induced in the semiconductor per unit area, Q_s, at a distance y from the source as shown in Figure is given from eqs 13, 14:

  $Q_s(y) = -[V_G - \psi_s(y)]C_o$, (26)
- Since Q_s is the sum of the charge in the inversion layer, Q_n, and the charge in surface depletion region, Q_{sc}, we can obtain Q_n as

$$\begin{aligned} Q_n(y) &= Q_s(y) - Q_{sc}(y), \\ &= -[V_G - \psi_s(y)]C_o - Q_{sc}(y) \end{aligned}$$

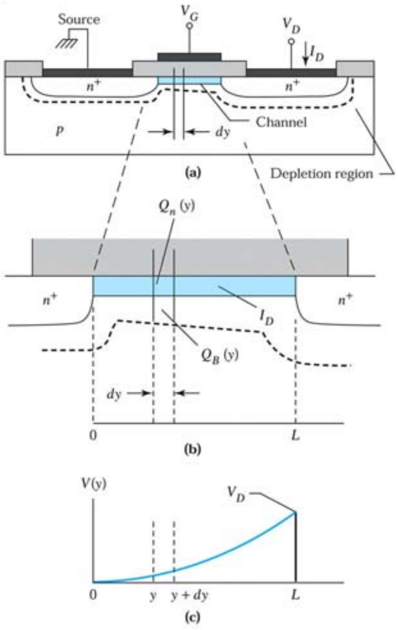


Figure 5.23
© John Wiley & Sons, Inc. All rights reserved.

Linear region (2)

- The surface potential $\psi_s(y)$ at inversion can be approximated by $2\psi_B+V(y)$, where $V(y)$, as shown in Figure, is the reverse bias between the point y and the source electrode
- The charge within the surface depletion region $Q_{sc}(y)$ was given previously as

$$Q_{sc}(y) = -qN_A W_m \cong -\sqrt{2\varepsilon_s q N_A [2\psi_B + V(y)]}$$

$$\rightarrow Q_n(y) \cong -[V_G - V(y) - 2\psi_B]C_o + \sqrt{2\varepsilon_s q N_A [2\psi_B + V(y)]}$$

- The conductivity of the channel at position y can be approximated by $\sigma(x) = qn(x)\mu_n(x)$
- For a constant mobility, the channel conductance is then given by

$$g = \frac{Z}{L}\int_0^{x_i} \sigma(x)\,dx = \frac{Z\mu_n}{L}\int_0^{x_i} qn(x)\,dx \qquad (31)$$

Linear region (3)

- ☐ The integral $\int_0^{x_i} qn(x)\,dx$ corresponds to the total charge per unit area in the inversion layer and is therefore equal to $|Q_n|$

→ $$g = \frac{Z\mu_n}{L}|Q_n| \qquad (32)$$

- ☐ The channel resistance of an elemental section *dy* is:

$$dR = \frac{dy}{gL} = \frac{dy}{Z\mu_n |Q_n(y)|}, \qquad (33)$$

and the voltage drop across the elemental section is

$$dV = I_D dR = \frac{I_D dy}{Z\mu_n |Q_n(y)|} \qquad (34)$$

where I_D is the drain current, which is independent of *y*

Linear region (4)

- ☐ Substituting eqn (29) into eqn 34 and integrating from the source (*y=0, V=0*) to the drain (*y=L, V=V*$_D$) yield

$$I_D \approx \frac{Z}{L}\mu_n C_o \left\{ \left(V_G - 2\psi_B - \frac{V_D}{2} \right) V_D - \frac{2}{3}\frac{\sqrt{2\varepsilon_s q N_A}}{C_o}\left[(V_D + 2\psi_B)^{3/2} - (2\psi_B)^{3/2} \right] \right\}$$

where $$\boxed{V_T = \frac{\sqrt{2\varepsilon_s q N_A (2\psi_B)}}{C_o} + 2\psi_B.}$$

- ☐ For small and very small V_D, the above equation can be reduced to:

$$\boxed{I_D \cong \frac{Z}{L}\mu_n C_o \left(V_G - V_T - \frac{V_D}{2} \right) V_D \quad \text{for} \quad V_D < (V_G - V_T).}$$

$$\boxed{I_D \cong \frac{Z}{L}\mu_n C_o (V_G - V_T) V_D \quad \text{for} \quad V_D << (V_G - V_T),}$$

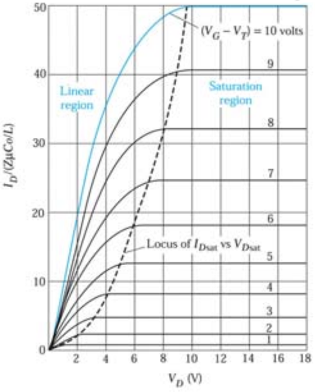


Figure 5.24

Linear region (4)

- By plotting I_D versus V_G (for a given small V_D), the threshold voltage can be deduced from the linearly extrapolated value at the V_g axis
- In the linear region, the channel conductance g_D and the transconductance g_m are given as

$$\boxed{g_D \equiv \left.\frac{\partial I_D}{\partial V_D}\right|_{V_G \ \text{constant}} \cong \frac{Z}{L}\mu_n C_o (V_G - V_T - V_D),}$$

$$\boxed{g_m \equiv \left.\frac{\partial I_D}{\partial V_G}\right|_{V_D \ \text{constant}} \cong \frac{Z}{L}\mu_n C_o V_D.}$$

Saturation Region

- When the drain voltage is increased to a point that the charge $Q_n(y)$ in the inversion layer at $y = L$ becomes zero (pinch-off), the number of mobile electrons at the drain are reduced drastically
- The drain voltage and the drain current at this point are called V_{Dsat} and I_{Dsat}, respectively
- For drain voltages larger than V_{Dsat}, we have the saturation region
- We can obtain the value of V_{Dsat} from Eq. 29 under the condition $Q_n(L) = 0$:

$$V_{Dsat} \cong V_G - 2\psi_B + K^2\left(1 - \sqrt{1 + \frac{2V_G}{K^2}}\right), \quad (40)$$

where $K \equiv \dfrac{\sqrt{\varepsilon_s q N_A}}{C_o}$

Saturation Region (2)

- The saturation current can be obtained by substituting Eq. 40 into Eq. 35

$$I_{Dsat} \cong \left(\frac{Z\mu_n C_o}{2L}\right)(V_G - V_T)^2.$$

- The threshold voltage V_T in the saturation region for low substrate doping and thin oxide layers is the same as that from Eq. 37
- At higher doping levels, V_T becomes V_G dependent
- For an idealized MOSFET in the saturation region, the channel conductance is zero, and the transconductance can be obtained from Eq. 41:

$$g_m \equiv \left.\frac{\partial I_D}{\partial V_G}\right|_{V_D \text{ constant}} = \frac{Z\mu_n \varepsilon_{ox}}{dL}(V_G - V_T).$$

The Subthreshold Region

- When the gate voltage is below the threshold voltage and the semiconductor surface is weakly inverted, the corresponding drain current is called the subthreshold current
- The subthreshold region is important when the MOSFET is used as a low-voltage, such as a switch in digital ICs, because it describes how the switch turns on and off
- In the subthreshold region, the drain current is dominated by diffusion instead of drift and is derived in the same way as the collector current in a bipolar transistor with homogeneous base doping. If we consider the MOSFET as an n-p-n(source-substrate-drain) bipolar transistor, we have

$$I_D = -qAD_n \frac{\partial n}{\partial y} = -qAD_n \frac{n(0) - n(L)}{L}, \qquad (43)$$

The Subthreshold Region (2)

where *A* is the channel cross section of the current flow and *n(0)* and *n(L)* are the electron densities in the channel

at the source and drain, respectively

- The electron densities are given by

$$n(0) = n_i e^{q(\psi_s - \psi_B)/kT},$$
$$n(L) = n_i e^{q(\psi_s - \psi_B - V_D)/kT}, \qquad (44)$$

- Substituting eq (44) in eq (43) gives

$$I_D = \frac{qAD_n n_i e^{-q\psi_B/kT}}{L}\left(1 - e^{-qV_D/kT}\right)e^{q\psi_s/kT} \qquad (45)$$

- The surface potential ψ_s is approximately $V_G - V_T$
- Therefore, the drain current will decrease exponentially when V_G becomes less than V_T

$$\boxed{I_D \sim e^{q(V_G - V_T)/kT}.} \qquad (46)$$

The Subthreshold Region (3)

- A typical measured curve for the subthreshold region is shown
- Note the exponential dependence of I_D on $(V_G - V_T)$ for $V_G < V_T$
- An important parameter in this region is the subthreshold swing, *S*, which is defined as $\ln 10[dV_G/d(\ln I_D)]$
- This parameter quantifies how sharply the transistor is turned off by the gate voltage and is given by the gate-voltage change needed to induce a drain-current change of one order of magnitude
- To reduce the subthreshold current to a negligible value, we must bias the MOSFET a half-volt or more below VT

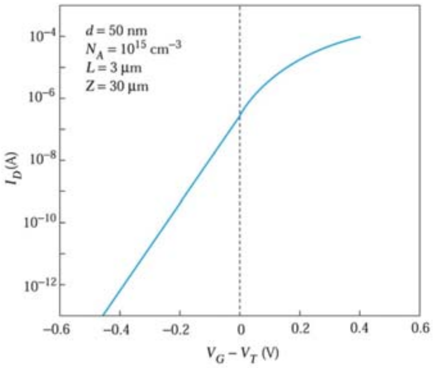


Figure 5.25

Types of MOSFET

- There are 4 types of MOSFET based on the substrate type (n or p) and the depletion region existence at $V_G=0$
- If at $V_G=0$ a channel already exists, the MOSFET is called Depletion MOSFET, else it is called enhancement MOSFET
- The most common is the enhancement MOSFET type

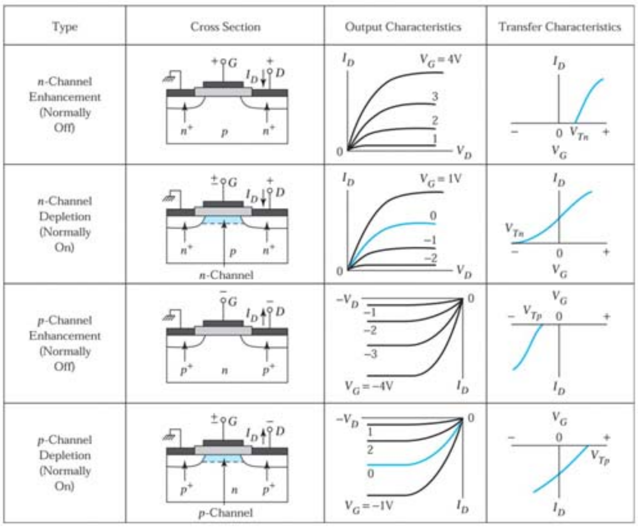
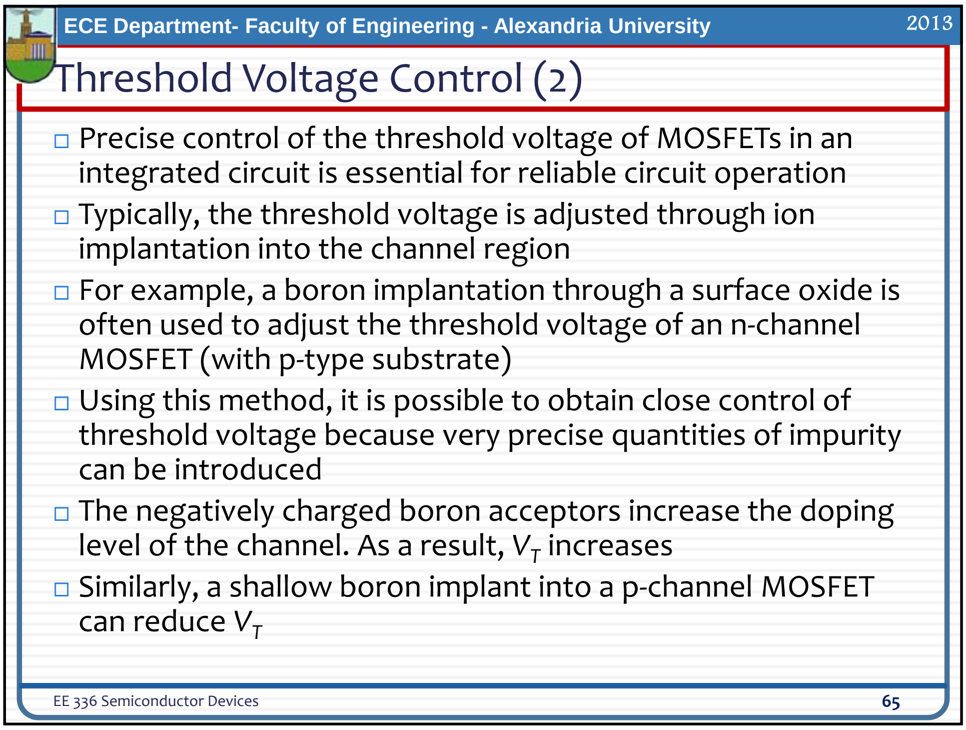


Figure 5.26
© John Wiley & Sons, Inc. All rights reserved.

Threshold Voltage Control

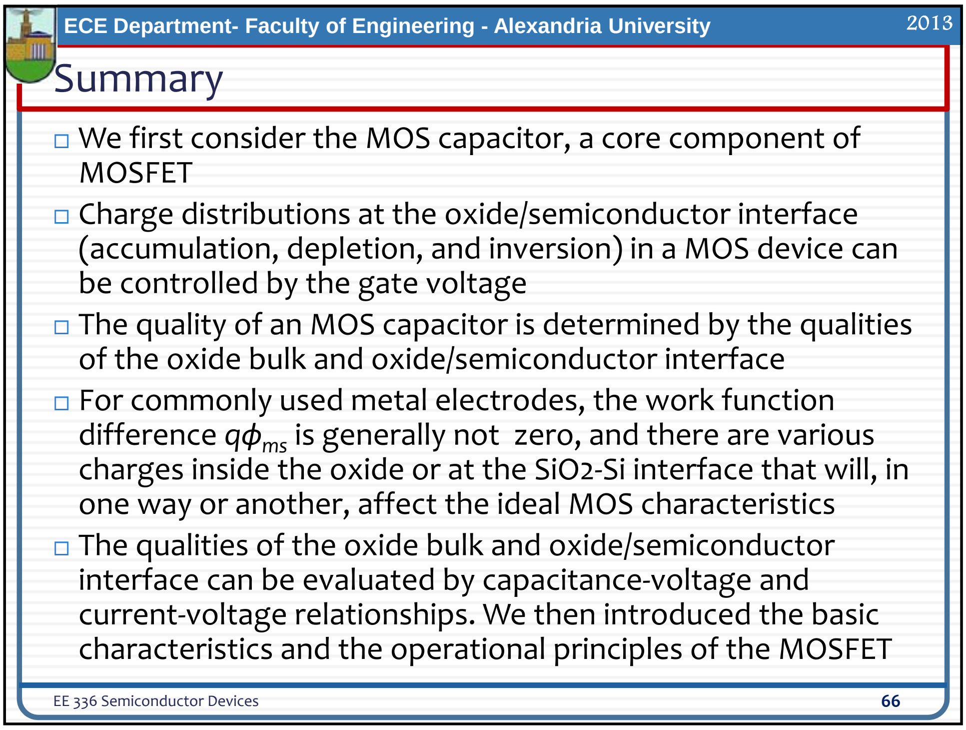
- One of the most important parameters of the MOSFET is the threshold voltage
- The ideal threshold voltage is given in Eq (37)
- However, when we incorporate the effects of the fixed-oxide charge and the difference in work function, there is a flat-band voltage shift
- Additionally, substrate bias can also influence the threshold voltage. When a reverse bias is applied between the substrate and the source, the depletion region is widened and the threshold voltage required to achieve inversion must be increased to accommodate the larger Q_{sc}

$$\boxed{V_T \approx V_{FB} + 2\psi_B + \frac{\sqrt{2\varepsilon_s q N_A (2\psi_B + V_{BS})}}{C_o},} \qquad (47)$$



Threshold Voltage Control (2)

- Precise control of the threshold voltage of MOSFETs in an integrated circuit is essential for reliable circuit operation
- Typically, the threshold voltage is adjusted through ion implantation into the channel region
- For example, a boron implantation through a surface oxide is often used to adjust the threshold voltage of an n-channel MOSFET (with p-type substrate)
- Using this method, it is possible to obtain close control of threshold voltage because very precise quantities of impurity can be introduced
- The negatively charged boron acceptors increase the doping level of the channel. As a result, V_T increases
- Similarly, a shallow boron implant into a p-channel MOSFET can reduce V_T



Summary

- We first consider the MOS capacitor, a core component of MOSFET
- Charge distributions at the oxide/semiconductor interface (accumulation, depletion, and inversion) in a MOS device can be controlled by the gate voltage
- The quality of an MOS capacitor is determined by the qualities of the oxide bulk and oxide/semiconductor interface
- For commonly used metal electrodes, the work function difference $q\phi_{ms}$ is generally not zero, and there are various charges inside the oxide or at the SiO2-Si interface that will, in one way or another, affect the ideal MOS characteristics
- The qualities of the oxide bulk and oxide/semiconductor interface can be evaluated by capacitance-voltage and current-voltage relationships. We then introduced the basic characteristics and the operational principles of the MOSFET

Summary (2)

- The MOSFET is formed when a source and a drain are placed adjacent to the MOS capacitor
- Output current (i.e., drain current) is controlled by varying the gate and drain voltages
- The threshold voltage is the main parameter that determines the on-off characteristics of an MOSFET
- The threshold voltage can be adjusted by choosing suitable substrate doping, oxide thickness, substrate bias, and gate materials