Introduction

The metal-oxide-semiconductor field-effect transistor consists of two p-n junctions either side of a MOS diode which acts as the gate.

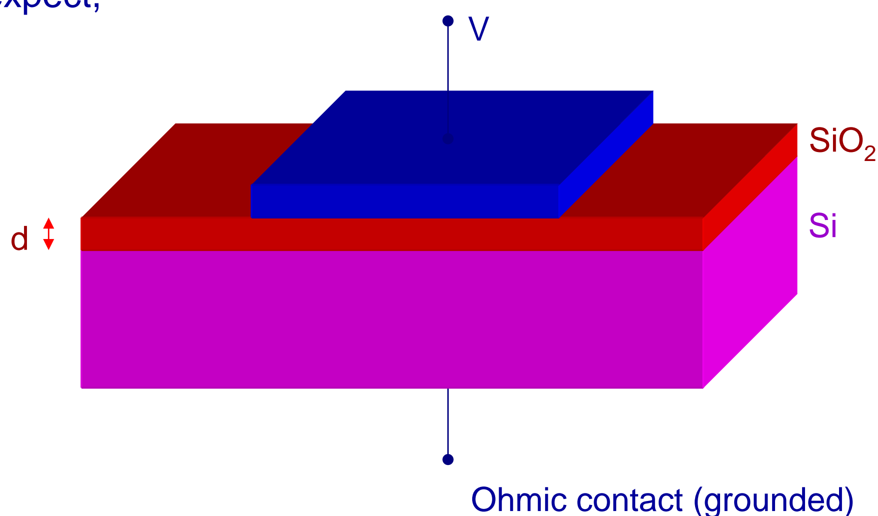
MOSFET devices make up around 90% of the electronics industry due to their low power and their small size (compared with bipolar transistors).

They are the basis for MOS memory structures such as Flash memory and CMOS logic structures.

Before we can understand the MOSFET we need to look closely at the MOS diode itself…

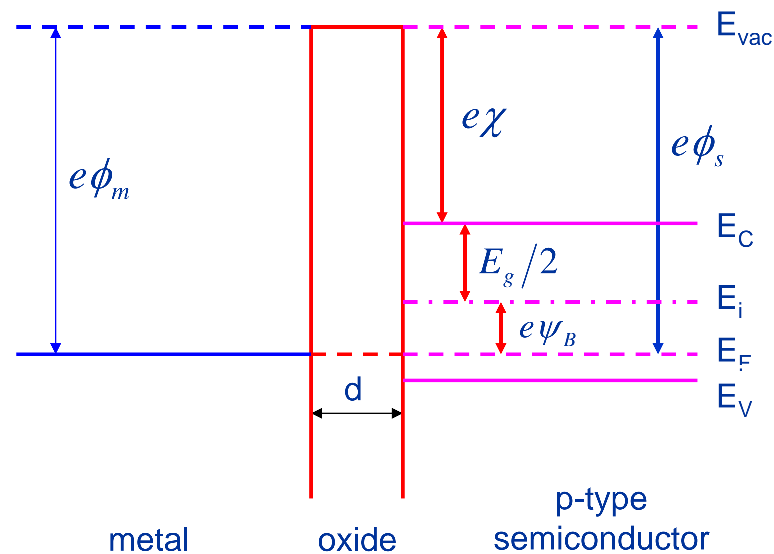
The MOS diode

The metal-oxide-semiconductor (MOS) diode has the following structure as you would expect,



V is the voltage applied to the metal contact and is positive if the applied bias is positive with respect to the Ohmic contact.

Band diagram



At thermal equilibrium, with no bias applied the Fermi levels of the metal and semiconductor are aligned.

Ideal MOS diode

An ideal MOS diode is defined as follows,

1) At zero applied bias, the energy difference between the metal and the semiconductor work functions is zero (the flat-band condition),

$$e\phi_{ms} = e\phi_m - e\phi_s = e\phi_m - \left(e\chi + \frac{E_g}{2} + e\psi_B \right) = 0$$

2) The only charges which exist in the system at any bias are those in the semiconductor and those with equal and opposite sign on the metal surface.

3) There is no carrier transport through the oxide under dc-biasing conditions.

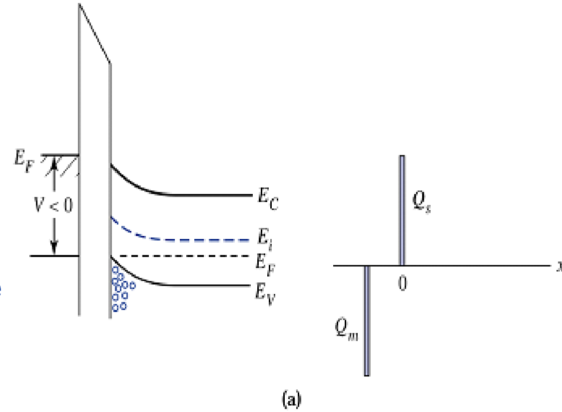
For the purposes of our discussion we will consider a p-type semiconductor MOS diode.

MOS diode under bias accumulation

Negative bias (V<0)

No current flows so semiconductor Fermi level constant. The bands bend upwards. Excess holes are induced at the semiconductor-oxide interface.

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



Since E_i-E_F is increased at the semiconductor surface, holes accumulate near the interface. This is known as the *accumulation* case.

$$|Q_m| = Q_s$$

Semiconductor Devices, 2/E by S. M. Sze
Copyright © 2002 John Wiley & Sons. Inc. All rights reserved.

MOS diode under bias depletion

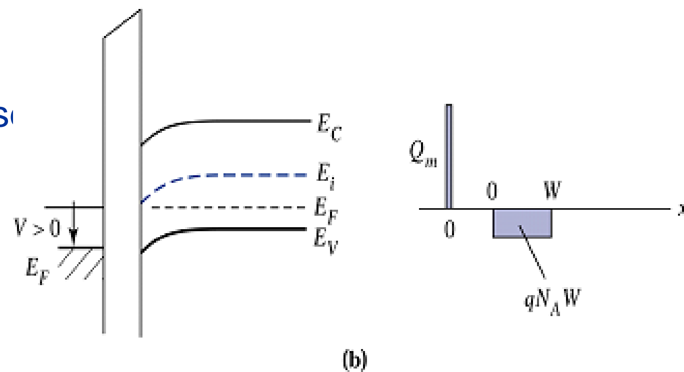
Small positive bias (V>0)

When a small positive bias is applied to the MOS diode, the bands bend downward, and the majority carriers are depleted.

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

This is known as the *depletion* cas

$$Q_{sc} = -eN_A W$$



Semiconductor Devices, 2/E by S. M. Sze
Copyright © 2002 John Wiley & Sons. Inc. All rights reserved.

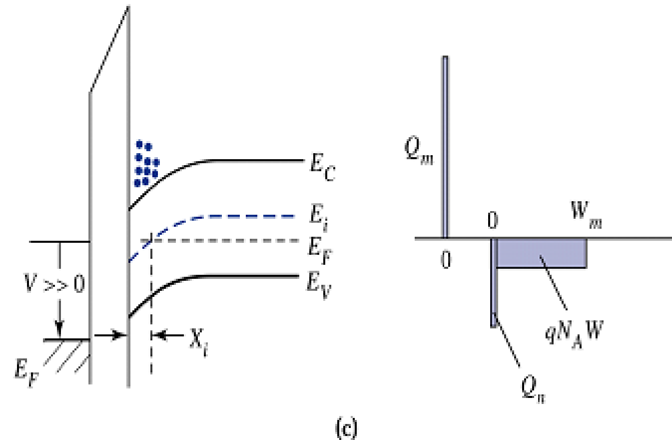
MOS diode under bias inversion

Large positive bias (V>0)

When a large positive bias is applied to the MOS diode, the energy bands bend downward, even more so that the intrinsic Fermi level at the surface crosses the Fermi level.

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

This is known as the *inversion* case. Where the minority carrier concentration (electrons) at the surface exceeds the majority carrier concentration (holes).



Semiconductor Devices, 2/E by S. M. Sze

Copyright © 2002 John Wiley & Sons. Inc. All rights reserved.

MOS diode under bias

strong inversion

Weak inversion begins as soon as $(E_F - E_i) > 0$, and the minority carrier concentration increases exponentially.

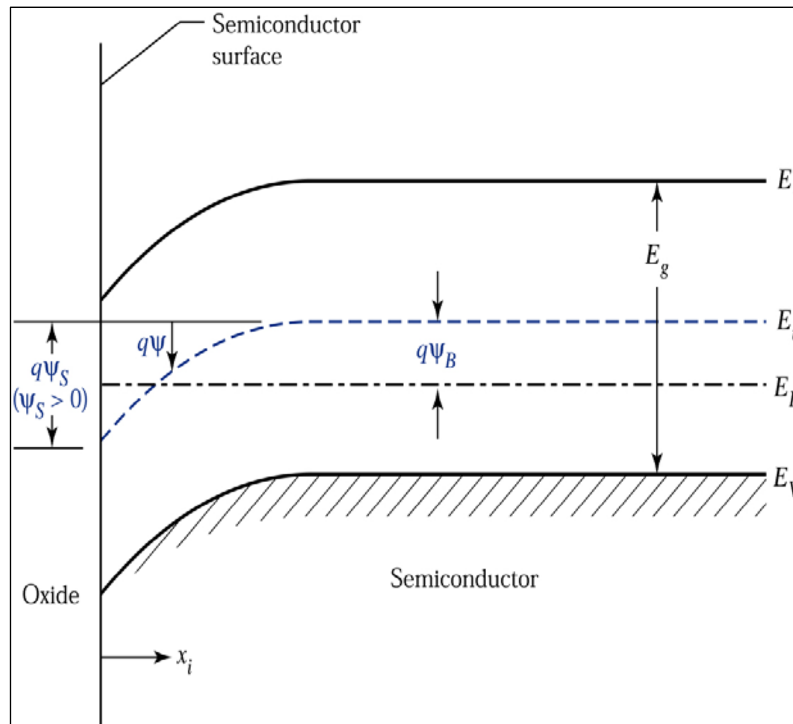
Strong inversion occurs when the minority carrier concentration is equal to the substrate doping level, (although convenient, this is a largely arbitrary criterion).

Beyond this point most of the additional negative charges consist of the charge Q_n in a very narrow *n*-type inversion layer of width x_i at the oxide-semiconductor interface (x_i ~1-10 nm << W).

The depletion width is at it maximum (W_m) at strong inversion since a small increase in band bending result in a very large increase in Q_n.

$$Q_s = Q_n + Q_{sc} = Q_n - eN_A W_m$$

Surface depletion region



Semiconductor Devices, 2/E by S. M. Sze Copyright © 2002 John Wiley & Sons. Inc. All rights reserved.

Electrostatic potential

Band-bending is described in terms of a quantity which we will call the electrostatic potential. This is defined as zero in the bulk and is a measure of the intrinsic Fermi level position with respect to the bulk intrinsic Fermi level.

At the semiconductor surface,

$$\psi = \psi_s \quad \text{Surface potential}$$

Since the electron and hole concentrations are given by,

$$n_p = n_i e^{(E_F - E_i)/kT} \quad \& \quad p_p = n_i e^{(E_i - E_F)/kT}$$

We can express them too as a function of the electrostatic potential,

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

Positive when bands bent downwards $\quad E_i$-E_F

Accumulation to Inversion

The carrier concentrations at the surface are therefore,

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

The following regions of surface potential are therefore identified,

$\psi_s < 0$	Accumulation of holes (bands bent upward)
$\psi_s = 0$	Flat-band condition
$0 < \psi_s < \psi_B$	Depletion of holes (bands bent downward)
$\psi_s = \psi_B$	Intrinsic condition ($n_s = n_p = n_i$)
$\psi_s > \psi_B$	Inversion (bands bent downward)

Depletion width

We can think of the surface depletion region as a one sided n^+-p abrupt junction where the built-in potential is replaced by the surface potential,

$$W = \sqrt{\frac{2\varepsilon_s \psi_s}{eN_A}}$$

Our earlier convenient-but-arbitrary definition stated that at the onset of strong inversion the electron concentration at the surface is equal to the substrate impurity concentration ($n_s = N_A$),

$$n_i e^{e(\psi_s(inv) - \psi_B)/kT} = n_i e^{e\psi_B/kT}$$

$$\psi_s(inv) \cong 2\psi_B = \frac{2kT}{e} \ln\left(\frac{N_A}{n_i}\right)$$

Band diagram

Depletion width

When the surface is strongly inverted the depletion width is at its maximum,

$$W_m = \sqrt{\frac{2\varepsilon_s \psi_s(inv)}{eN_A}} \cong \sqrt{\frac{2\varepsilon_s 2\psi_B}{eN_A}}$$

$$W_m = \sqrt{\frac{4\varepsilon_s kT \ln(N_A/n_i)}{e^2 N_A}}$$

C-V Characteristics

In the absence of any work function differences, the applied bias will appear partly across the oxide and partly across the semiconductor,

$$V = V_o + \psi_s$$

The potential across the oxide is given by,

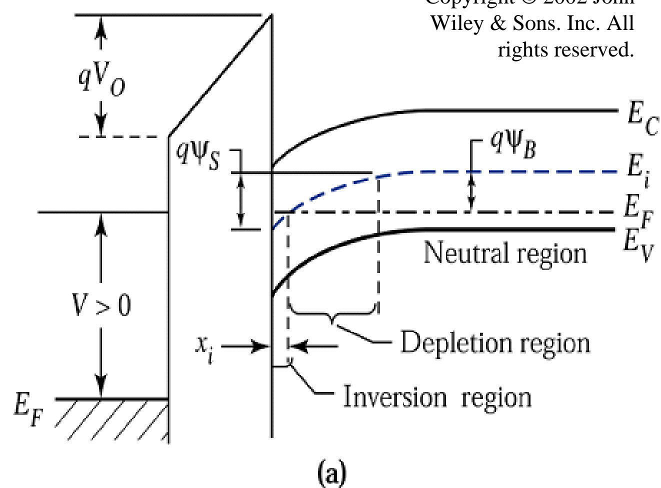
$$V_o = F_o d$$

$$V_{ox} = \frac{|Q_s| d}{\varepsilon_{ox}}$$

$$C_{ox} = \varepsilon_{ox}/d$$

$$V_{ox} = \frac{|Q_s|}{C_{ox}}$$

Semiconductor Devices, 2/E by S. M. Sze Copyright © 2002 John Wiley & Sons. Inc. All rights reserved.



C-V Characteristics

The total capacitance of the MOS diode is a series combination of the oxide capacitance and the semiconductor depletion-layer capacitance,

$$C = \frac{C_o C_j}{(C_o + C_j)}$$

where,

$$C_j = \frac{\varepsilon_s}{W}$$

$$C_o = \frac{\varepsilon_{ox}}{d}$$

When the applied bias is negative, there is no depletion layer and we have accumulation of holes at the surface of the semiconductor. In this case, the total capacitance is close to that of the oxide.

C-V Characteristics

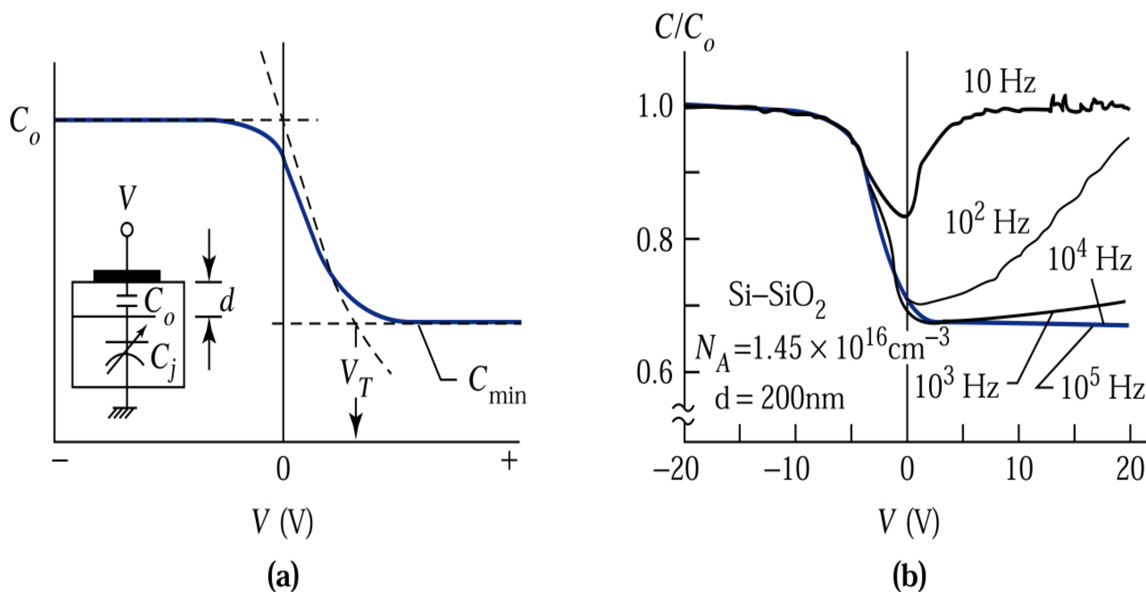


Figure 6.7. (*a*) High-frequency MOS *C-V* curve showing its approximated segments (dashed lines). Inset shows the series connection of the capacitors. (*b*) Effect of frequency on the *C-V* curve.[2]

Threshold voltage

When the semiconductor surface is strongly inverted, the width of the depletion region is at its maximum. This occurs at a plate voltage that causes the surface potential to reach its inversion value. This is known as the threshold voltage,

$$V = V_o + \psi_s$$

$$V_T = V_o + \psi_s(inv)$$

$$V_T = \frac{eN_A W_m}{C_o} + \psi_s(inv) \cong \frac{\sqrt{2\varepsilon_s e N_A (2\psi_B)}}{C_o} + \psi_B$$

Beyond strong inversion, the capacitance will remain at a minimum value,

$$C_{\min} = \frac{\varepsilon_{ox}}{d + (\varepsilon_{ox}/\varepsilon_s)W_m} \qquad \text{since,} \qquad C_j = \frac{\varepsilon_s}{W_m}$$

Real Si-SiO_2 MOS diodes

The work function difference between the metal and the semiconductor is generally not zero as in our idealised case, and certainly isn't for the most commonly used metals in the metal-Si-SiO_2 system…

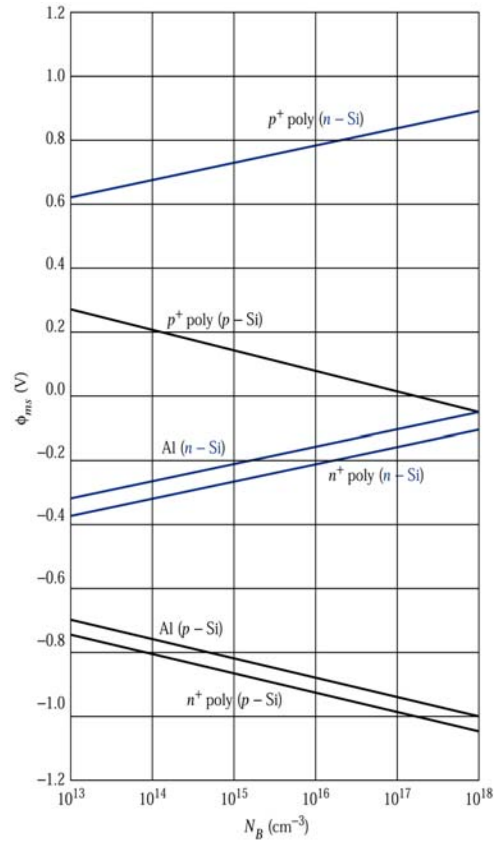
$\phi_m = 3.95eV$ n^+-polysilicon

$\phi_m = 4.1eV$ Aluminium

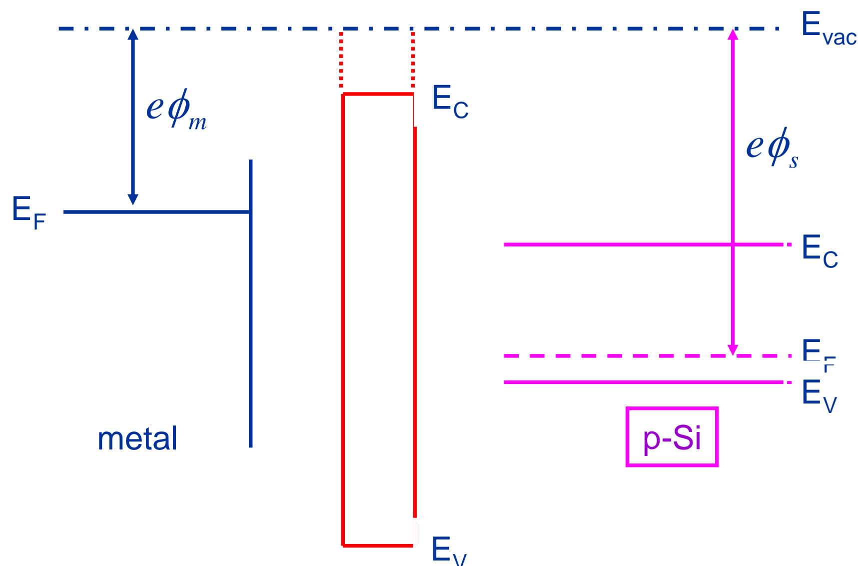
$e(\phi_m - \phi_s) < 0$ For all doping levels of Si

How will this affect our band diagram?

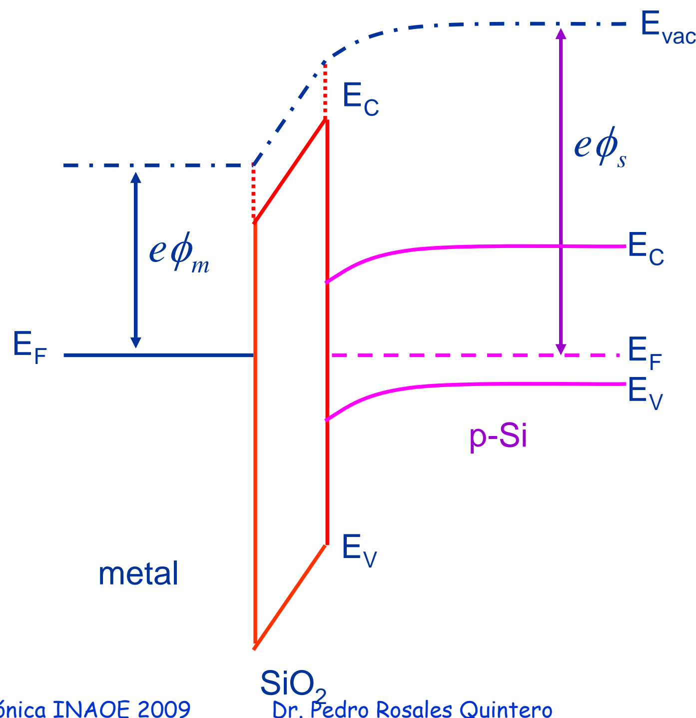
Figure 6.8. Work function difference as a function of background impurity concentration for Al, *n*+-, and *p*+ polysilicon gate materials.



Si-SiO_2 band diagram



Si-SiO$_2$ band diagram

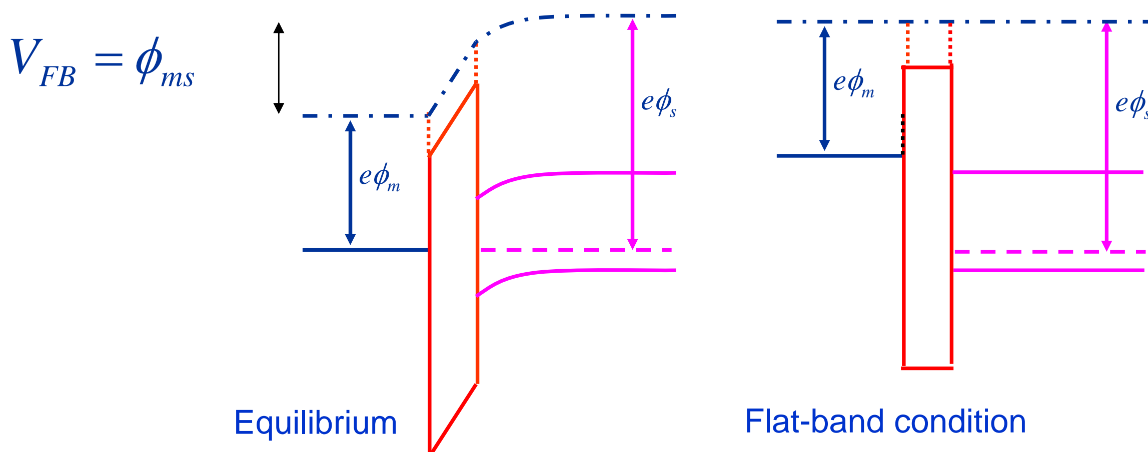


Flat-band voltage

At thermal equilibrium, the bands are bent downwards so the semiconductor surface is negatively charged, and the metal positively charged.

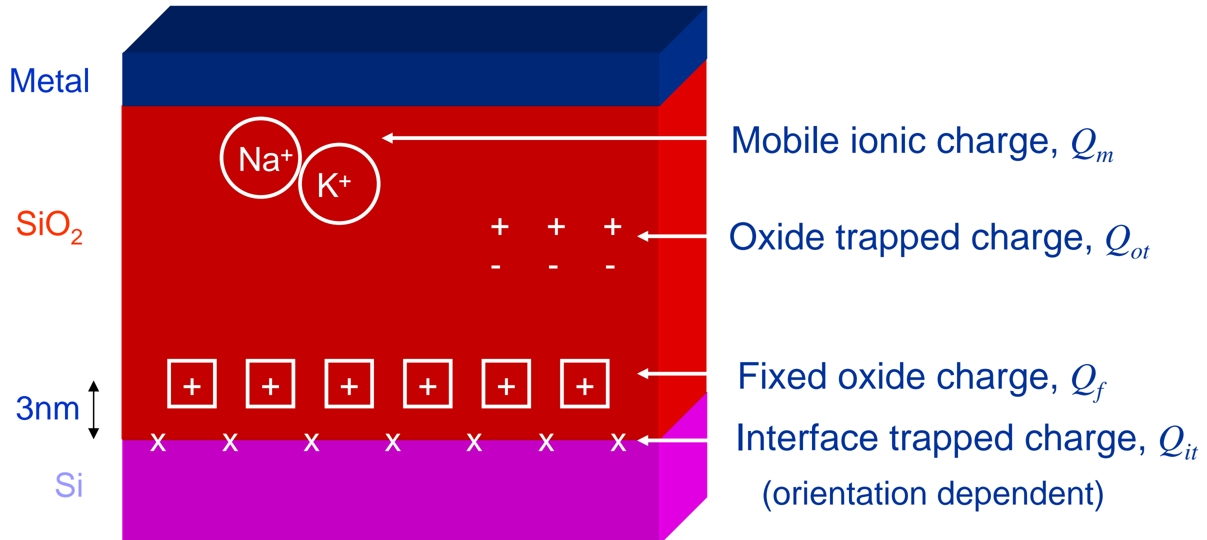
In order to get back to the flat-band condition discussed for our ideal MOS diode we need to apply a negative bias to the metal (the flat-band voltage),

$$V_{FB} = \phi_{ms}$$

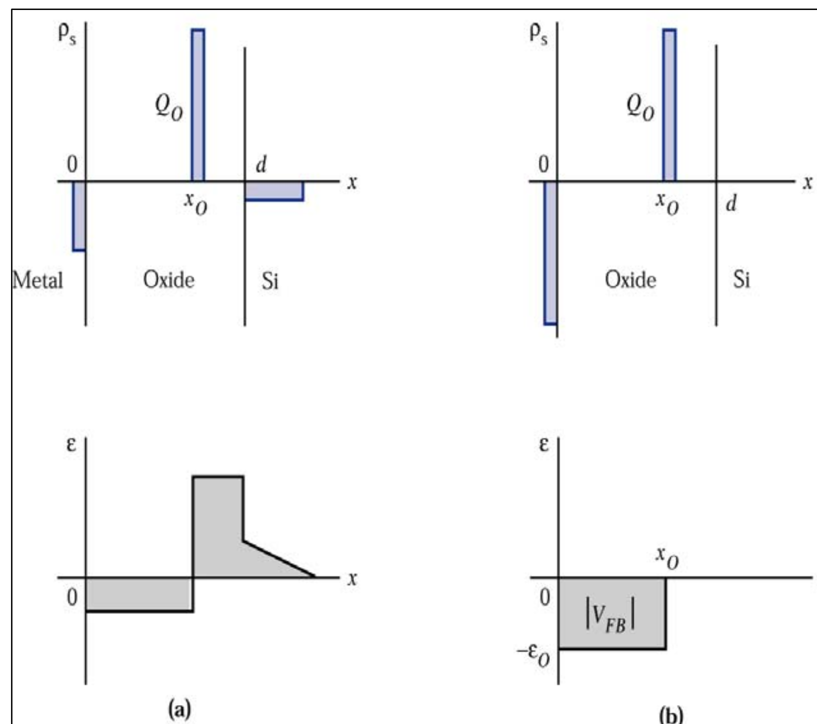


Traps & charges

Real MOS diodes are also affected by charges trapped within the oxide and the interface…



Oxide charges & the flat-band voltage



Semiconductor Devices, 2/E by S. M. Sze Copyright © 2002 John Wiley & Sons. Inc. All rights reserved.

Oxide charges & the flat-band voltage

Consider a positive charge sheet per unit area, Q_o within the oxide, inducing negative charges on both the metal and the semiconductor.

The flat-band condition is reached by increasing the charges on the metal (by applying a negative voltage, V_{FB}) to reduce the electric field distribution at the semiconductor surface to zero.

The area under the electric field profile at this point gives the flat-band voltage,

$$V_{FB} = -F_o x_o = -\frac{Q_o}{\varepsilon_{ox}} x_o = -\frac{Q_o}{C_o}\frac{x_o}{d}$$

Oxide charges & the flat-band voltage

For an arbitrary space charge distribution within the oxide we can integrate the volume charge density,

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

For a real diode in which the work function difference is nonzero and the interface-trapped charge is negligible, the flat-band voltage is given by,

$$V_{FB} = \phi_{ms} - \frac{(Q_f + Q_m + Q_{ot})}{C_o}$$

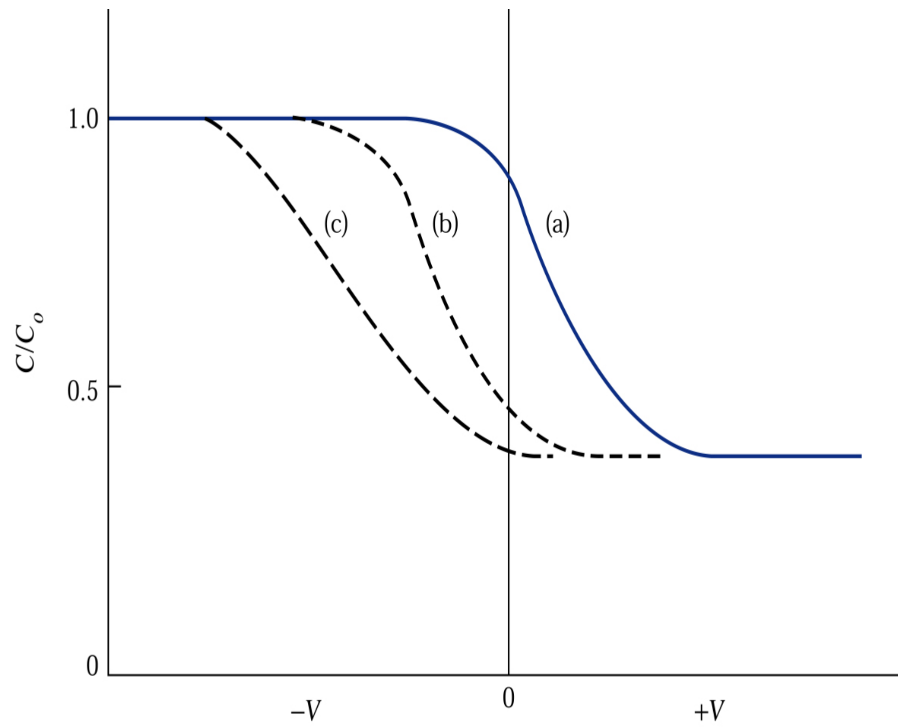


Figure 6.12. Effect of a fixed oxide charge and interface traps on the *C-V* characteristics of an MOS diode.