The Transistor: First proposed by Lilienfeld in 1930 (but he could never really get it to work right because of surface states)

1930

Field-effect transistor:
Using a gate C, Lilienfeld thought that it should be possible to modulate the current from A to B.

This is conceptually very similar to the vacuum triode, which was used as the amplifier at the time.
In 1947, Bardeen and Brattain invented the Ge point contact transistor. They wanted to make the Field-Effect Transistor, but ended up with a Bipolar Transistor, and got the Nobel Prize anyway. Schockley then developed the bipolar junction transistor.
Field effect devices: The electric field of a gate or grid is used to modulate the number of charges (i.e., electron current) moving from the source to the drain.

In a triode, the charges are electrons accelerated through a vacuum.
\[
\frac{d\mathcal{E}}{dx} = \frac{\rho}{K_S \varepsilon_0} \equiv -\frac{qN_A}{K_S \varepsilon_0} \quad (0 \leq x \leq W)
\]
Charge balance of a MOS capacitor. $-Q = +Q$
\[ E_i(\text{surface}) - E_i(\text{bulk}) = 2[E_F - E_i(\text{bulk})] \]

\[ p_s = n_i e^{[E_i(\text{surface}) - E_F]/kT} = n_i e^{[E_F - E_i(\text{bulk})]/kT} = n_{\text{bulk}} = N_D \]
Energy band diagram

Applied dc voltage

$V_G = 0$

$V_G < 0$

$V_T > V_G > 0$

$V_G > V_T$

Charge diagram

Name     Flat band  Accumulation  Depletion  Inversion

Exposed acceptors
Electrons
Substitute the definition of $E_i$ and $E_f$

$$p_{\text{bulk}} = n_i e^{[E_i(\text{bulk}) - E_F]/kT} = N_A \quad \text{... if } N_A \gg N_D$$

$$n_{\text{bulk}} = n_i e^{[E_F - E_i(\text{bulk})]/kT} = N_D \quad \text{... if } N_D \gg N_A$$

$$\phi(x) = \frac{1}{q} [E_i(\text{bulk}) - E_i(x)]$$

$$\phi_S = \frac{1}{q} [E_i(\text{bulk}) - E_i(\text{surface})]$$

$$\phi_F = \frac{1}{q} [E_i(\text{bulk}) - E_F]$$

(a)

(b)
\[
\phi_F = \begin{cases} 
\frac{kT}{q} \ln\left(\frac{N_A}{n_i}\right) & \ldots \text{p-type semiconductor} \\
-\frac{kT}{q} \ln\left(\frac{N_D}{n_i}\right) & \ldots \text{n-type semiconductor}
\end{cases}
\]

\(\phi_s = 2\phi_F\) at the depletion–inversion transition point
Derivation of the depletion width of a MOS capacitor (on p-doped semiconductor)

\[ \rho = q(p - n + N_D - N_A) \approx -qN_A \quad (0 \leq x \leq W) \]

\( N_A \gg N_D \)

\[ \frac{d\mathcal{E}}{dx} = \frac{\rho}{K_S \varepsilon_0} \approx -\frac{qN_A}{K_S \varepsilon_0} \quad (0 \leq x \leq W) \]

Integrate

\[ \mathcal{E}(x) = -\frac{d\phi}{dx} = \frac{qN_A}{K_S \varepsilon_0} (W - x) \quad (0 \leq x \leq W) \]

Integrate again

\[ \phi(x) = \frac{qN_A}{2K_S \varepsilon_0} (W - x)^2 \quad (0 \leq x \leq W) \]

Depletion width into the semiconductor

Note that this is the permittivity of the oxide
As is the case in typical Schottky diodes and p-n diodes, the depletion width changes as a function of the square root of the dopant concentration.

This assumes zero bias:

\[ \phi_S = \frac{qN_A}{2K_S \varepsilon_0} W^2 \]

\[ W_T = \left[ \frac{2K_S \varepsilon_0}{qN_A} (2\phi_F) \right]^{1/2} \]
Charge distribution with different applied voltages

Note that the charge associated with inversion resides in an extremely narrow channel immediately adjacent to the oxide.
Definition of surface potential and band bending

\[
E_i(\text{surface}) - E_i(\text{bulk}) = 2[E_F - E_i(\text{bulk})]
\]

\[
p_s = n_i e^{[E_i(\text{surface}) - E_F]/kT} = n_i e^{[E_F - E_i(\text{bulk})]/kT} = n_{\text{bulk}} = N_D
\]

\[
\phi(x) = \frac{1}{q} \left[E_i(\text{bulk}) - E_i(x)\right]
\]

\[
\phi_s = \frac{1}{q} \left[E_i(\text{bulk}) - E_i(\text{surface})\right]
\]

\[
\phi_F = \frac{1}{q} \left[E_i(\text{bulk}) - E_F\right]
\]
Influence of a gate voltage

\[ V_G = \Delta \phi_{\text{semi}} + \Delta \phi_{\text{ox}} \]

\[ \Delta \phi_{\text{semi}} = \phi(x = 0) = \phi_s \]

\[ \varepsilon_{\text{ox}} = -\frac{d\phi_{\text{ox}}}{dx} = \text{constant} \]

\[ \Delta \phi_{\text{ox}} = \int_{-x_0}^{0} \varepsilon_{\text{ox}} dx = x_0 \varepsilon_{\text{ox}} \]

The voltage drop through the oxide must be considered.

The dielectric displacements must be equal at the interface between the oxide and the semiconductor.

\[ (D_{\text{semi}} - D_{\text{ox}}) \big|_{\text{O-S interface}} = Q_{\text{O-S}} \]
\[ D_{ox} = D_{\text{semi}}|_{x=\xi} \]

\[ \varepsilon_{ox} = \frac{K_S}{K_O} \varepsilon_S \]

\[ \Delta \phi_{ox} = \int_{-x_0}^{0} \varepsilon_{ox} \, dx = x_o \varepsilon_{ox} \]

We start here, and start substituting:

\[ \Delta \phi_{ox} = \frac{K_S}{K_O} x_o \varepsilon_S \]

Since:

\[ V_G = \Delta \phi_{\text{semi}} + \Delta \phi_{ox} \]

\[ V_G = \phi_S + \frac{K_S}{K_O} x_o \varepsilon_S \]

Next, we substitute the depletion derivation

\[ \varepsilon_S = \left[ \frac{2qN_A}{K_c \varepsilon_0} \phi_S \right]^{1/2} \]

\[ V_G = \phi_S + \frac{K_S}{K_O} x_o \sqrt{\frac{2qN_A}{K_S \varepsilon_0}} \phi_S \quad (0 \leq \phi_S \leq 2\phi_F) \]
P: An MOS-C is maintained at $T = 300$ K, $x_o = 0.1$ $\mu$m, and the silicon doping is $N_A = 10^{15}$/cm$^3$. Compute:

(a) $\phi_F$ in $kT/q$ units and in volts

(b) $W$ when $\phi_s = \phi_F$

(c) $\varepsilon_s$ when $\phi_s = \phi_F$

(d) $V_G$ when $\phi_s = \phi_F$

S: (a)

$$\frac{\phi_F}{kT/q} = \ln\left(\frac{N_A}{n_i}\right) = \ln\left(\frac{10^{15}}{10^{16}}\right) = 11.51$$

$$\phi_F = 11.51 \times \frac{kT}{q} = (11.51)(0.0259) = 0.298 \text{ V}$$

(b) Using Eq. (16.15),

$$W = \left[\frac{2K_S\varepsilon_0}{qN_A} \phi_F\right]^{1/2} = \left[\frac{2(11.8)(8.85 \times 10^{-14})(0.298)}{(1.6 \times 10^{-19})(10^{15})}\right]^{1/2} = 0.624 \mu$m

(c) Evaluating Eq. (16.12) at $x = 0$ yields $\varepsilon_s$. Thus

$$\varepsilon_s = \frac{qN_A}{K_S\varepsilon_0} W = \frac{(1.6 \times 10^{-19})(10^{15})(6.24 \times 10^{-5})}{(11.8)(8.85 \times 10^{-14})} = 9.56 \times 10^3 \text{ V/cm}$$

(d) Substituting into Eq. (16.26) gives

$$V_G = \phi_F + \frac{K_S}{K_O} x_0 \varepsilon_s \quad \ldots \varepsilon_s \text{ evaluated at } \phi_F$$

$$= 0.298 + \frac{(11.8)(10^{-5})(9.56 \times 10^3)}{3.9} = 0.587 \text{ V}$$
Typical C-V measurement system
Derivation of the capacitance of a MOS

\[ C_o = \frac{K_o \varepsilon_0 A_G}{x_o} \quad \text{(oxide capacitance)} \]

\[ C_s = \frac{K_s \varepsilon_0 A_G}{W} \quad \text{(semiconductor capacitance)} \]

\[ C(\text{depl}) = \frac{C_o C_s}{C_o + C_s} = \frac{C_o}{1 + \frac{K_o W}{K_s x_o}} \]

\[ C(\text{inv}) \approx C_o \quad \text{for } \omega \to 0 \]

\[ C(\text{inv}) = \frac{C_o C_s}{C_o + C_s} = \frac{C_o}{1 + \frac{K_o W_T}{K_s x_o}} \quad \text{for } \omega \to \infty \]
Assuming a delta-depletion formulation, the capacitance can be expressed in simpler terms:

\[
C = \begin{cases} 
C_0 & \text{acc} \\
\frac{C_0}{1 + \frac{K_0 W}{K_s x_0}} & \text{depl} \\
C_0 & \text{inv} \ (\omega \to 0) \\
\frac{C_0}{1 + \frac{K_0 W_T}{K_s x_0}} & \text{inv} \ (\omega \to \infty)
\end{cases}
\]

\[W = \frac{K_s}{K_0} x_0 \left[ \sqrt{1 + \frac{V_G}{V_\delta}} - 1 \right]\]

where

\[V_\delta = \frac{q}{2} \frac{K_s x_0^2}{K_0^2 \varepsilon_0} N_A \quad \ldots \text{p-bulk device} \]

(for n-bulk \(N_A \to -N_D\))

\[C = \frac{C_0}{\sqrt{1 + \frac{V_G}{V_\delta}}} \quad \text{(depletion biases)}\]
<table>
<thead>
<tr>
<th>Bias Condition</th>
<th>Capacitance (a–g)</th>
<th>Charge Diagram (1–5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accumulation</td>
<td>g</td>
<td>3</td>
</tr>
<tr>
<td>Depletion</td>
<td>c</td>
<td>1</td>
</tr>
<tr>
<td>Inversion</td>
<td>a,b,c</td>
<td>4</td>
</tr>
<tr>
<td>Flat band</td>
<td>f</td>
<td>2</td>
</tr>
<tr>
<td>Depl/inv transition</td>
<td>d</td>
<td>5</td>
</tr>
</tbody>
</table>

depletion

accumulation

inversion
When the channel is turned off, this is a back-to-back diode and only the saturation current is flowing.

Once there is a conducting channel (inversion takes place) this changes to a resistor…

As the voltage on the drain is increased, the current to the drain is also increased. Note the saturation (pinch-off) voltage, though.
Note that the charge associated with inversion resides in an extremely narrow channel immediately adjacent to the oxide
As the drain voltage is increased, the inversion voltage is reduced.
The saturation current can be controlled by adjusting the gate voltage.
Ideal threshold voltage dependence: The threshold voltage should only depend on the number of dopants and the dielectric constant. However, there are extra charges in the oxide and on the surface, which make our lives somewhat more complicated.

\[
V_T = 2\phi_F + \frac{K_S x_0}{K_O} \sqrt{\frac{4qN_A}{K_S \varepsilon_0}} \phi_F \quad \ldots \text{ideal } n\text{-channel} \quad (p\text{-bulk}) \text{ devices}
\]

\[
V_T = 2\phi_F - \frac{K_S x_0}{K_O} \sqrt{\frac{4qN_D}{K_S \varepsilon_0}} (-\phi_F) \quad \ldots \text{ideal } p\text{-channel} \quad (n\text{-bulk}) \text{ devices}
\]
Impurities in oxide and at interface provide extra charges which need to be considered.

Q_m Mobile ionic charge
Q_ot Oxide trapped charge
Q_f Oxide fixed charge
Q_it Interface trap charge

V = V_{FB} = -\frac{Q_i}{C_i}
As the dopant concentration is changed, the workfunction difference is also changed.
Threshold voltage

Interfacial and trapped charges

Metal semiconductor workfunction difference

Charges from depleted dopants

\[ V_T = \Phi_{ms} - \frac{Q_i}{C_i} - \frac{Q_d}{C_i} + 2\phi_F \]
The threshold voltage is dependent on the dopant concentration too. This is a way the MOSFET $V_T$ can be tuned.
Crossection through the gate of a narrow-channel MOSFET
Figure 17.5 Visualization of surface scattering at the Si–SiO$_2$ interface.

\[
\bar{\mu}_n = \frac{\int_{0}^{x_c(y)} \mu_n(x, y)n(x, y) \, dx}{\int_{0}^{x_c(y)} n(x, y) \, dx}
\]