Field-Effect Transistors (FETs)
Introduction to Electronic Devices

6 Field-Effect-Transistors (FETs)

6.1 Introduction
6.2 Basic Device Structures
6.3 The MOS structure
  6.3.1 Accumulation
  6.3.2 Depletion
  6.3.3 Weak Inversion
  6.3.4 Strong Inversion
  6.3.5 Carrier Distribution for an ideal MOS structure
  6.3.6 Potential Distribution for an ideal MOS structure
  6.3.7 The ideal MOS structure

6.4 The silicon oxide MOS structure
  6.4.1 The Work Function Difference
  6.4.2 Interface Traps and Oxide Charges
  6.4.3 The MOSFET Threshold voltage
6.5 Metal Oxide Semiconductor Field Effect Transistor (MOSFET)

6.5.1 Basic characteristics of MOS Field Effect Transistor

6.5.2 MOSFETs modeling
   6.5.2.1 Gradual Channel Approximation
   6.5.2.2 I/V characteristic in the linear region
   6.5.2.3 I/V characteristic in the saturation region
   6.5.2.4 The output characteristic
   6.5.2.5 The transfer characteristic
   6.5.2.6 Conductance and Transconductance
   6.5.2.7 The Sub threshold Region

6.5.3 Types of MOSFETs

6.5.4 Threshold voltage control

6.5.5 Scaling of MOSFETs
   5.5.5.1 Threshold Voltage Roll-off in the linear region
   5.5.5.2 Drain Induced Barrier Lowering in the saturation region
   5.5.5.3 Bulk punch through
   5.5.5.4 Scaling Rules

6.6 Silicon on Insulator (SOI)

6.7 Thin Film Transistors (TFTs)
   6.7.1 Amorphous silicon TFTs
   6.7.2 Poly silicon TFTs
   6.7.3 Organic TFTs

References
6 Field Effect Transistors (FETs)

The concept of the Field Effect Transistor (FET) was proposed by Lilienfeld (1930). The idea got practical after the pioneering work of Shockley in the early 1950. Today the Field-Effect Transistor is the most important electronic device used for microprocessors and semiconductor memories.

Throughout the last 50 years several field-effect transistor concepts have been developed and implemented. The most important group of field effect transistors is the class of metal insulator semiconductor field effect transistor (MISFET). In this case the gate is insulated from the channel of the transistor by an insulator. Out of this class the metal-oxide-semiconductor field-effect transistor (MOSFET) structure is by far the most important structure.

Another class of transistors, which belongs to the group of field effect transistors is the MESFET. The Metal-semiconductor field effect transistor (MESFET) structure is different from the MISFET structure. Instead of using a MOS structure to modulate the channel conductivity a metal semiconductor (Schottky barrier / Schottky diode) is used control the conductivity of the channel.

In the following the operating principle of the Metal-Oxide-Semiconductor Field-Effect Transistor will be discussed.
6.1 Introduction

The metal-oxide-semiconductor field-effect transistor (MOSFET) is composed of a MOS structure (MOS diode / MOS capacitor) and two pn-junctions placed immediately adjacent to the MOS structure.

A MOSFET is a **charge controlled device**. Charges have to be accumulated on the gate to control the device properties. As a consequence of voltage applied to the gate a channel is formed at the interface between dielectric and substrate.

![Schematic cross section of an enhanced-type NMOS transistor.](image)
6.1 Introduction

The induced channel of a MOSFET can be a n-channel (high concentration of electrons) or a p-channel (high concentration of holes). Therefore, a MOSFET is an **unipolar device**. Either electrons or holes contribute to the current flow. In terms of the real implementation the statement is not completely right, because pn-junctions are placed adjacent to the MOS structure.

The output current of the transistor is defined to be the current between the drain and source contact. The drain current is controlled by the gate bias. The FET is in the off-state, when only a few electrons (holes) move from the source to the drain. In the on-state electrons or holes are injected via the source and flow to the drain.

Therefore, the operating principle of a MOSFET is quite different from the behavior of a bipolar junction transistor (BJT). A BJT is a current/voltage controlled device, whereas the MOSFET is a charge controlled device. A MOSFET is a unipolar device, whereas BJTs or diodes are bipolar devices.
6.2 Basic Device Structures

The MOSFET is a charge controlled device, so that the charge on the gate should be maximized. The charge on the gate can be calculated by:

\[ Q_G = C_G V_G = \frac{\varepsilon_0 \cdot \varepsilon_{\text{dielectric}}}{d} \cdot V_G \]

The charge on the gate can be maximized by increasing the dielectric constant of the gate dielectric and decreasing the thickness of the gate dielectric. For example: The gate capacitance can be increased by changing the dielectric material from silicon oxide SiO\(_2\) to Aluminum oxide Al\(_2\)O\(_3\). Aluminum oxide exhibits a higher dielectric constant. The gate electrode and the channel form a plate capacitor or a MOS structure, so that the electrons (holes) are located in a thin region close to the dielectric interface.
6.3 The MOS structure

The MOS structure is the „heart“ of a MOSFET. In the following the operating principle of the MOS structure will be discussed before addressing the fundamentals of the MOSFET.

In general a MOS structure can be used to store charges. The concept of storing charge in a MOS structure is used for example to realize Charge Coupled Devices (CCD) and MOSFETs.

Cross section of a metal oxide semiconductor (MOS) structure.

Ref.: M.S. Sze, Semiconductor Devices
6.3 The MOS structure

The energy band diagram of a MOS structure using a p-type semiconductor is shown in the figure under thermal equilibrium (V=0). Before discussing the energy diagram the following functions and energies are introduced:

The **work function** $q\Phi$ is defined as the energy required to remove an electron from the Fermi level $E_F$ to a position outside the material (vacuum level). The work function can be defined for a semiconductor, a metal or an insulator.

The **electron affinity** $q\chi$ is the energy required to remove an electron from the bottom of the conduction band to the vacuum level.

Energy band diagram of an ideal MOS structure for V=0 (flat band condition).

Ref.: M.S. Sze, Semiconductor Devices
6.3 The MOS structure

\[ q\phi_{ms} \equiv q\phi_m - q\phi_s = q\phi_m - \left( q\chi + \frac{E_g}{2} + q\varphi_B \right) = 0 \]

At zero applied bias (V=0) the energy difference between the metal work function \( q\phi_m \) and the semiconductor work function \( q\phi_s \) is zero. As a consequence the work function difference \( q\phi_{ms} \) is zero. In this case the energy band diagram is flat. This case is called flat-band condition!

Before we discuss the influence of different gate voltages on the band energy diagram and the carrier distribution of a MOS structure, we will introduce the surface electrostatic potential (or simply surface potential). The gate voltage leads to a bending of the band diagram and we will describe the bending of the band by using the bulk and the surface potential.
6.3 The MOS structure

The surface potential $\varphi_S$ is defined as zero in the bulk of the semiconductor. At the semiconductor surface the electric potential is equal to the surface potential. If now a voltage is applied to MOS structure charges are displaced due to coulomb interaction. In the following it is assumed that the MOS structure acts as a plate capacitor. The formation of charges in the semiconductor leads to the formation of opposite charges on the metal electrode.

There is no carrier transport through the gate dielectric. Therefore, the MOS structure is in thermal equilibrium even though a voltage is applied to the structure.

Energy band diagram at the surface of a p-type semiconductor

Ref.: M.S. Sze, Semiconductor Devices
6.3 The MOS structure

In the following it is assumed that the semiconductor material is p-type doped.

5.3.1 Accumulation

If a negative voltage (V<0) is applied to the MOS structure, excess carriers (holes) will be accumulated at the semiconductor / dielectric interface. In this case, the energy band at the interface between the semiconductor and the dielectric is bent upwards. There is no current flow independent of the applied bias voltage to the MOS structure.

Energy band diagram and charge distribution of an ideal MOS structure in accumulation (V<0).

Ref.: M.S. Sze, Semiconductor Devices
6.3.1 Accumulation

The Fermi level in the semiconductor is constant independent of the applied bias voltage. The carrier concentration in the semiconductor can be described by:

\[ p_p = n_i \cdot \exp\left(\frac{E_i - E_F}{kT}\right) \]

As the difference between the intrinsic energy level and the Fermi level is increased close to the interface the hole concentration is distinctly increased and the electron concentration is decreased.

This case is called accumulation.
### 6.3.2 Depletion

For positive voltages \((V>0)\) applied to the gate electrode the region close to the interface of the gate dielectric and the channel will be depleted. The energy bands bend downwards and the majority carriers (holes) are depleted. Therefore, this case is called **depletion**.

Energy band diagram and charge distribution of an ideal MOS structure in depletion \((V>0)\).

Ref.: M.S. Sze, Semiconductor Devices
6.3.2 Depletion

For positive voltages ($V>0$) the energy bands bend downwards and the majority carriers (holes) are depleted. The carrier concentration is again given by:

$$p_p = n_i \cdot \exp\left(\frac{E_i - E_F}{kT}\right)$$

As the difference between the intrinsic energy level and the Fermi level is decreased closer to the interface of the dielectric and the semiconductor the hole concentration is distinctly decreased, whereas the electron concentration is increased.

Energy band diagram and carrier distribution of an ideal MOS structure in depletion ($V>0$).
6.3.3 Weak Inversion

If a larger positive voltage is applied to the MOS structure, the energy bands bend downwards even more so that the intrinsic energy at the surface crosses the Fermi level. The positive gate voltage starts to induce excess negative carriers (electrons) at the SiO₂ interface. The carrier concentration can be described by:

\[ n_p = n_i \cdot \exp \left( \frac{E_F - E_i}{kT} \right) \]

Energy band diagram and charge distribution of an ideal MOS structure in weak inversion (V>0).

Ref.: M.S. Sze, Semiconductor Devices
6.3.3 Weak Inversion

Therefore, the electron concentration is larger than the hole concentration at the interface.

This case is called inversion. Initially, the surface is in weak inversion, since the electron concentration is small. If the bias voltage applied to the gate is further increased the bands bend further and eventually the conduction band comes close to the Fermi level.

Energy band diagram and carrier distribution of an ideal MOS structure in weak inversion ($V>0$).

Energy:
- $E_C$
- $E_i$
- $E_F$
- $E_V$

Carrier distribution:
- $p_0$
- $N_A$
- $n_i$
- $n_0$

Inversion (weak Inversion)
6.3.4 Strong Inversion

If the applied bias voltage is further increased the MOS structure turns into strong inversion. Strong inversion occurs when the electron concentration at the surface is higher than the doping concentration in the bulk of the material. Most of the additional charges are located in a narrow inversion layer close to the interface of the dielectric and the semiconductor.

Under strong inversion the width of the depletion layer reaches its maximum. A very small increase of the band bending corresponds to a large increase of the charges in the inversion layer and a small increase of the carrier concentration in the depletion region. The charge balance can be described by:

\[ n_p = n_i \cdot \exp \left( \frac{E_F - E_i}{kT} \right) \]

Energy band diagram and charge distribution of an ideal MOS structure in strong inversion \((V >> 0)\).
6.3.4 Strong Inversion

Initially, the surface is in **weak inversion** since the electron concentration is smaller than the hole concentration in the bulk of the semiconductor. If the applied bias voltage on the gate is further increased, the band bends further and eventually the conduction band comes close to the Fermi level.

Energy band diagram and carrier distribution of an ideal MOS structure in strong inversion ($V>>0$).
6.3.5 Carrier Distribution for an ideal MOS structure

The majority and the minority carrier concentration in the bulk and at the surface can be expressed in terms of the bulk and the surface potential:

The carrier concentration in the bulk can be described by:

\[
\begin{align*}
\ni_p &= n_i \cdot \exp\left(\frac{q \cdot (\varphi - \varphi_b)}{kT}\right) \\
\ip &= n_i \cdot \exp\left(\frac{q \cdot (\varphi_b - \varphi)}{kT}\right)
\end{align*}
\]

The carrier concentration at the surface can be described by:

\[
\begin{align*}
\ns &= n_i \cdot \exp\left(\frac{q \cdot (\varphi_s - \varphi_b)}{kT}\right) \\
\ps &= n_i \cdot \exp\left(\frac{q \cdot (\varphi_b - \varphi_s)}{kT}\right)
\end{align*}
\]
6.3.6 Potential Distribution for an ideal MOS structure

Different cases of device operation for a MOS structure (p-type semiconductor) are listed in the following. The voltage applied to MOS structure, V, and surface and bulk potentials are listed for each operating case. Furthermore, the minority carrier concentration at the interface \( n_S \) is compared with the minority carrier concentration in the bulk \( n_B \).

<table>
<thead>
<tr>
<th>V&lt;0</th>
<th>Accumulation (of holes)</th>
<th>( \varphi_S&lt;0 )</th>
<th>( n_S&lt;n_B&lt;n_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>V=0</td>
<td>Flat-band conditions</td>
<td>( \varphi_S=0 )</td>
<td>( n_S=n_B&lt;n_i )</td>
</tr>
<tr>
<td>V&gt;0</td>
<td>Depletion (of holes)</td>
<td>( \varphi_B&gt;\varphi_S&gt;0 )</td>
<td>( n_B&lt;n_S&lt;n_i )</td>
</tr>
<tr>
<td>V&gt;&gt;0</td>
<td>Weak inversion</td>
<td>( \varphi_B=\varphi_S )</td>
<td>( n_S=n_i&gt;n_B )</td>
</tr>
<tr>
<td>V&gt;&gt;&gt;0</td>
<td>Strong inversion</td>
<td>( \varphi_B&lt;\varphi_S )</td>
<td>( n_S&gt;n_i&gt;n_B )</td>
</tr>
</tbody>
</table>
6.3.7 The ideal MOS structure

Initially the surface is in **weak inversion** since the electron concentration (p-type MOS structure) is small. With increasing band bending, eventually the conduction band edge comes close to the Fermi level. The onset of **strong inversion** occurs when the electron concentration is equal to the dopant (acceptors) concentration.

This can be achieved for relatively high positive voltages applied to the MOS structure. Under such conditions most of the charges are located in a narrow layer at the interface between the dielectric and the channel. Current transport occurs in this thin layer. The layer thickness ranges from 1-10nm. The layer is called the **inversion layer**. The inversion layer is much thinner than the width of the depletion layer.

Energy band diagram and charge distribution of an ideal MOS structure in weak inversion (V>>0).
6.3.7 The ideal MOS structure

In the following the charge distribution, the electric field distribution and the potential distribution of a MOS diode will be determined. The carrier distribution in the semiconductor can be approximated by the Delta-Depletion approximation:

Operating region: Weak inversion

\[ Q_s = Q_n + Q_{SCR} \]
\[ Q_{SCR} = -q N_A W \]
\[ Q_s = Q_n - q N_A W \]

Energy band diagram and Charge distribution of a MOS structure under weak inversion.

Ref.: M.S. Sze, Semiconductor Devices
6.3.7 The ideal MOS structure

The electric field and the potential distribution can be calculated by solving the Poisson equation:

\[
\frac{d^2 \varphi}{dx^2} = -\frac{\rho_s(x)}{\varepsilon_s}
\]

In order to describe the charges in the semiconductor the Depletion approximation is used. The influence of the inversion layer on the electric field and potential distribution is negligible even though the current transport occurs in the inversion layer.

\[
\rho_s = -qN_A
\]

We already used the depletion approximation to calculate the electric field and the potential distribution of the pn-junction (chapter 4). The potential distribution is calculated by

\[
\varphi = \varphi_s \cdot \left(1 - \frac{x}{W}\right)^2
\]
6.3.7 The ideal MOS structure

where the surface potential is given by:

\[ \Phi_s = \frac{qN_A W^2}{2 \varepsilon_s} \]

Surface Potential

Electric field and potential distribution of a MOS structure under weak inversion.

Ref.: M.S. Sze, Semiconductor Devices
6.3.7 The ideal MOS structure

The surface is getting “inverted” when $\varphi_S$ is larger than $\varphi_B$. In this case the MOS structure operates under weak inversion conditions. If the applied voltage is further increased the bands are bended more and the structure operates under strong inversion conditions. The criterion for strong inversion is that the electron concentration (minority carriers in the bulk) at the interface is larger than the majority carrier concentration (dopant concentration) in the bulk. The surface potential under strong inversion can be calculated by:

\[
\begin{align*}
    n_s &= N_A \\
    n &= n_i \exp\left(\frac{q \varphi_b}{kT}\right) \\
    n_s &= n_i \cdot \exp\left(\frac{q \cdot (\varphi_s - \varphi_b)}{kT}\right)
\end{align*}
\]

Conditions for strong inversion

Electron concentration in the bulk

Electron concentration at the surface
6.3.7 The ideal MOS structure

The surface potential for strong inversion can now be calculated by

\[ \varphi_s (\text{inversion}) \approx 2 \varphi_b = \frac{2kT}{q} \ln\left( \frac{N_A}{n_i} \right) \]

The bending of the bands reaches its maximum when the surface is strongly inverted.

\[ \varphi_s = \frac{qN_A W^2}{2 \varepsilon_s} \]

As a consequence the width of the depletion layer is maximized. The width of the depletion layer can be calculated by:

\[ W_{\text{max}} = \sqrt{\frac{2 \varepsilon_s \varphi_s (\text{inv.})}{qN_A}} \approx \sqrt{\frac{4 \varepsilon_s \varphi_B}{qN_A}} \]
6.3.7 The ideal MOS structure

Leading to the expression for the maximum width of the depletion region:

\[ W_{\text{max}} = 2\sqrt[4]{\frac{\varepsilon_s kT \ln \left( \frac{N_A}{n_i} \right)}{q^2 N_A}} \]

Maximum width of the depletion layer under strong inversion

Relationship between the maximum depletion layer width and the dopant concentration.

Ref.: M.S. Sze, Semiconductor Devices
6.3.7 The ideal MOS structure

The potential distribution for a MOS structure is given by

\[ V = V_o + \varphi_s \]

where the voltage drop across the gate oxide is given by

\[ V_o = F \cdot d = \frac{|Q_s| \cdot d}{\varepsilon_{ox}} \]

\(Q_s\) is the charge on the gate and \(\varepsilon_{ox}\) is the dielectric constant of the gate oxide. The voltage drop across the gate dielectric can be expressed in terms of the charge on the gate divided by the gate capacitance \(C_o\) or \(C_G\).

\[ V_o \equiv \frac{|Q_s|}{C_o} \]

![Electric field and potential distribution of a MOS structure under weak inversion.](image)

Ref.: M.S. Sze, Semiconductor Devices
6.3.7 The ideal MOS structure

The overall MOS structure can be described by the series connection of a junction capacitor and a gate capacitor,

\[ C = \frac{C_o C_j}{C_o + C_j} \]

where \( C_j = \frac{\varepsilon_S}{W} \). The equation for the MOS structure can be rewritten in the following form:

\[ C = \frac{C_0}{\sqrt{1 + \frac{2\varepsilon_{ox}^2 V}{qN_A\varepsilon_S d^2}}} \]

Capacitance of a MOS structure

Ref.: M.S. Sze, Semiconductor Devices

(High frequency) Capacitance Voltage (CV) curve of a MOS structure.
6.4 The silicon oxide MOS structure

In the following the basic concepts of the MOS structure will be applied to study the behavior of a real materials system. The behavior of a silicon oxide (thermal oxide) MOS structure will be discussed.

6.4.1 The Work Function Difference

Under flat-band conditions the energy difference between the metal work function $q\phi_m$ and the semiconductor work function is zero $q\phi_s$. As a consequence the work function difference $q\phi_{ms}$ is zero.

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

Energy band diagram of an ideal MOS structure for $V=0$ (flat band condition).

Ref.: M.S. Sze, Semiconductor Devices
6.4.1 The Work Function Difference

In order to achieve the flat band case the work function of the gate material and the semiconductor has to be equal, which means that the Fermi levels are identical for the two materials. However, this is only the case for particular combinations of materials. The work functions for different metals is different. Metals like gold and platinum have a high functions. The work function of chromium and aluminum is slightly lower. The work function of the semiconductor depends on the doping concentration of the semiconductor.

Work function difference as a function of the dopant concentration for aluminum and poly silicon (gate).

Ref.: M.S. Sze, Semiconductor Devices
6.4.1 The Work Function Difference

The work function difference of the two materials leads to a bending of the bands in the semiconductor. Free charges in the semiconductor are simply attracted or repelled by the charges on the gate. By doing that the MOS structure will reach thermal equilibrium. The Fermi level will be constant throughout the MOS structure.

But not only the bands in the semiconductor are bent. The vacuum level follows the band bending of the conduction and the valence band leading to a bending of the vacuum level.

The work function of the different materials is known, so that the band bending can be adjusted by choosing a particular material combination.

Energy band diagram of a MOS structure in thermal equilibrium.

Ref.: M.S. Sze, Semiconductor Devices
6.4.2 Interface Traps and Oxide Charges

In addition to the work function difference of the different materials the MOS structure is affected by charges in the oxide and traps located in the oxide or at the silicon oxide/silicon interface. This charges lead to an additional bending of the bands at the interface between the channel and the dielectric.

Different kinds of charges/traps can be classified:

• Interface trapped charges,
• Fixed oxide charges,
• Oxide trapped charges and
• Mobile ionic charges.

Interface traps and dielectric charges associated with thermal oxide.

Ref.: M.S. Sze, Semiconductor Devices
6.4.2 Interface Traps and Oxide Charges

**Interface-trapped charges** are formed at the silicon oxide / silicon interface. The properties at the interface depend on the bonding between the silicon lattice and the lattice of the dielectric.

**Fixed charges** are located within 3nm of the silicon oxide / silicon interface. The charges are fixed and cannot be recharged. The charges are typically positive. The underlying mechanism is not completely understood, but the concentration of fixed charges is related to the oxidation and annealing conditions of the MOS structure.

**Oxide trapped charges** are associated with defects in the silicon oxide itself. Most of the defects can be removed by annealing of the dielectric.

**Mobile ions** such as sodium or alkali ions start to get mobile in the dielectric at high temperatures (>100°C) and high electric fields. The mobile ions lead to stability problems of the MOS structure.
6.4.2 Interface Traps and Oxide Charges

The existence of a work function difference between the gate material and channel (semiconductor) on one side and the influence of the interface and trapped charges on the other side prevents that the energy band diagram is flat for $V_G=0$.

To reach the flat-band conditions a voltage has to be applied to the gate, which counteracts these effects. Therefore, we introduce a **flat band voltage**, which counteracts these effects.

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

Flat band voltage

6.4.3 The MOSFET Threshold Voltage

Silicon MOSFETs usually have gate contacts made of highly doped poly silicon. Doped polysilicon is highly conductive and it forms ohmic contacts with metal electrodes. On the next slide the energy band diagram of a MOS structure is shown for a p-type silicon and a n-type silicon substrate.
6.4.3 The MOSFET Threshold voltage

The poly silicon gate is highly n-type doped, so that the Fermi level is very close to (or even in the) conduction band. As a consequence the flat-band voltage for a p-type substrate can be as high as −1V. For a n-type substrate the flat-band voltage is much smaller. The flat-band voltage is caused by the relative difference of the Fermi levels if the influence of fixed charges and traps is ignored.

The band diagram of a PMOS and a NMOS FET is shown for flat-band condition.

Ref.: M. Shur, Introduction to Electronic Devices
6.4.3 The MOSFET Threshold voltage

If a voltage $V_G$ is applied to the MOS structure the flat-band voltage can be compensated, so that the Fermi level gets constant throughout the entire structure.

The band diagram of a MOS FET under positive bias conditions.

Ref.: M. Shur, Introduction to Electronic Devices
6.4.3 The MOSFET Threshold voltage

The voltage applied to the MOS structure can be described by

\[ qV_G + |V_{FB}| = +qV_o + q\varphi_S + q\chi_{gate} - q\chi_{semi} \]

It can be assumed that the electron affinity of the semiconductor is equal to the electron affinity of the gate material,

\[ q\chi_{gate} \approx q\chi_{semi} \]

so that the following equation can be derived:

\[ V_G = V_{FB} + V_o + \varphi_S \]

The threshold voltage is defined at the onset of strong inversion.

\[ V_T = V_G \left( \varphi_S = 2\varphi_B \right) \]

\[ V_T = V_{FB} + \sqrt{2\varepsilon_S qN_A \cdot 2\varphi_B} \left/ C_0 \right. + 2\varphi_B \]

Threshold voltage
6.5 Metal Oxide Semiconductor Field Effect Transistor (MOSFET)

The metal-oxide-semiconductor field-effect transistor (MOSFET) is composed of a MOS structure (MOS diode / MOS capacitor) and two pn-junctions placed immediately adjacent to the MOS structure. The MOSFET is the most important device for integrated circuits like microprocessor and semiconductor memories.

The MOSFET is an unipolar device. (At least the theory based on the transport of either electrons or holes). The practical implementation however includes pn-junctions placed adjacent to the MOS structure.

Schematic cross section of an enhanced-type NMOS transistor.

Ref.: M.S. Sze, Semiconductor Devices
6.5.1 Basic characteristics of MOS Field Effect Transistor

In the following the I/V characteristic of the MOSFET will be derived. The source contact of the MOSFET is used as a reference throughout the following discussion. If no voltage is applied to the drain and source contacts no current can flow besides the leakage current of the back to back connected diodes.

For positive voltages the MOS structure is inverted, so that an inversion layer (or channel) is formed at the interface between the dielectric and the substrate. As a consequence a large current can flow between the drain and source. The conductivity of the channel can be modulated by the applied gate voltage.

Two basic operation region can be distinguished for an MOSFET, the linear region and the saturation region.
6.5.1 Basic characteristics of MOS Field Effect Transistor

In the first step a positive voltage is applied to the gate so that an inversion layer is formed in the p-type substrate. In the next step a voltage is applied to the drain electrodes, while the source electrode is grounded. If the applied bias voltage is small the current flow between the drain and source is proportional to the conductivity of the channel. The channel acts as an resistor and the resistivity is modulated by the gate voltage. The drain current $I_D$ is proportional to the drain voltage. This behavior (region) is called the **linear region**.

Schematic cross section of an enhanced-type NMOS transistor under positive applied bias voltage and output curve in the linear region.

Ref.: M.S. Sze, Semiconductor Devices
6.5.1 Basic characteristics of MOS Field Effect Transistor

As a consequence the channel behaves like a resistor, which can be modulated by the gate voltage. The electric field in the channel can be assumed to be constant.

Schematic cross section of an enhanced-type NMOS transistor in the linear region and voltage drop across the channel.

Ref.: M.S. Sze, Semiconductor Devices
6.5.1 Basic characteristics of MOS Field Effect Transistor

When the drain voltage is increased the voltage eventually reaches the point, where the thickness of the inversion layer is reduced to zero. This point is called the **pinch-off** point. As a consequence the resistance of the channel can not be modulated by the applied drain voltage anymore. The drain current is getting saturated.

Schematic cross section of an enhanced-type NMOS transistor under positive applied bias voltage and output curve under pinch off conditions.

Ref.: M.S. Sze, Semiconductor Devices
6.5.1 Basic characteristics of MOS Field Effect Transistor

Beyond the pinch off point the drain current remains essentially constant. Therefore, the number of carriers flowing from the drain to the source is essentially independent of the drain voltage. This region is called the saturation region.

Schematic cross section of an enhanced-type NMOS transistor under positive applied bias voltage and output curve in the saturation region.

Ref.: M.S. Sze, Semiconductor Devices
6.5.2 MOSFETs modeling

We will now derive the I/V characteristic of a MOSFET under the following ideal conditions:

• The gate structure corresponds to an ideal MOS structure (No fixed or trapped charges in the dielectric and no difference of the work function).
• Only drift current is considered
• The carrier mobility in the inversion layer is constant
• The doping of the channel is uniform.
• The reverse-leakage is negligible
• The gradual channel approximation applies, which means that the transverse field created by the gate which is perpendicular to the channel is much large than the longitudinal electric field.
6.5.2 MOSFETs modeling

If the gate voltage $V_G$ is larger than the threshold voltage $V_T$ an inversion layer (channel) is formed at the interface of the dielectric and the substrate.

If the voltage applied to the drain is very small the concentration of carriers flowing along the channel is constant (the source electrode is connected to ground). As a consequence the description of the MOSFET can be reduced to a 1-dimensional model. For higher drain voltages the concentration of carriers is not constant throughout the channel, so that the field effect transistor may have to be described by a 2- or 3-dimensional model.

2- and 3- dimensional MOSFET models exists and they are of particular interest for the description of short channel MOSFETs. The description however is rather complex.

In our case the channel can be considered to be long and we will use the gradual channel approximation to reduce the description of a MOSFET to an 1-dimensional problem.
6.5.2.1 Gradual Channel Approximation

Schematic cross section of a MOSFET including the distribution of the electric field perpendicular of the insulator-semiconductor interface ($F_y$) and the electric field in the semiconductor at the insulator-semiconductor interface parallel to the interface ($F_x$).

Ref.: M. Shur, Introduction to Electronic Devices
6.5.2.1 Gradual Channel Approximation

The graduate channel approximation applies if

\[ \frac{\partial F_x}{\partial x} \ll \frac{\partial F_y}{\partial y} \approx \frac{\rho}{\varepsilon_S} \]

where \( F_y \) is the electric field perpendicular to the channel (along the MOS structure) and \( F_x \) (parallel to the channel) is the electric field along the channel.

The gradual channel approximations simply assumes that the dimensions of the transistor perpendicular to the channel are much larger than the dimensions in direction of the channel.

As a consequence the two-dimensional Poisson equation

\[ \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} = \frac{\rho}{\varepsilon_S} \]

Can be reduced to an 1-Dimensional Poisson equation.

\[ \frac{\partial F_y}{\partial y} \approx \frac{\rho}{\varepsilon_S} \]
### 6.5.2.2 I/V characteristic in the linear region

As the MOSFET is a charge controlled device the concentration of carriers in the channel can be described by

\[ qn_s = C_G \cdot (V_{GS} - V_T - V_y) \]

It is assumed that the drift current is the dominant current contribution and the mobility is constant throughout the material, so that the drift velocity of the carriers in the inversion layer is given by

\[ v_n = \mu_n F_y = \mu_n \frac{dV_y}{dy} \]

In the next step the current in the channel is calculated by

\[ I_d = Wq \mu_n \frac{dV_y}{dy} n_s \]

The drain current can now be calculated by

\[ I_d dy = W \mu_n C_G \cdot (V_{GS} - V_T - V_y) dV_y \]
6.5.2.2 I/V characteristic in the linear region

The drain current can be calculated after integration along the channel (from the source contact, \( y=0 \) to the drain contact, \( y=L \)).

\[
\int_{0}^{L} I_d \, dy = W \mu_n C_G \cdot \int_{0}^{V_{DS}} \left( V_{GS} - V_T - V_y \right) dV_y
\]

The integration leads to the final equation for the drain current in the linear region.

Linear region means that for small drain voltages the charge induced in the channel does not depend on the potential along the channel. The channel can be modulated by the gate voltage. Therefore, the conductivity of the channel can be modulated by the gate voltage and the drain current is proportional to the drain voltage.

\[
I_d = \mu_n C_G \frac{W}{L} \cdot \left( V_{GS} - V_T - \frac{V_{DS}}{2} \right) \cdot V_{DS}
\]

Drain current in the linear region

for \( V_{DS} \ll V_G - V_T \)
6.5.2.3 I/V characteristic in the saturation region

The strong inversion layer at the drain electrode is getting zero under **Pinch-off** conditions. Pinch off occurs for

\[ V_{DS} = V_{GS} - V_T \]

Substituting the drain-source voltage leads to the expression for the drain current in the saturation region.

\[ I_d = \mu_n C_G \frac{W}{2L} \cdot (V_{GS} - V_T)^2 \]

for \( V_{DS} \geq V_G - V_T \)

Drain current in the saturation region
6.5.2.4 The output characteristic

Summary: MOSFET in linear and saturation region

\[ I_d = \mu_n C_G \frac{W}{L} \left( V_{GS} - V_T - \frac{V_{DS}}{2} \right) \cdot V_{DS} \]

for \( V_{DS} \ll V_G - V_T \)

Drain current in the linear region

\[ I_d = \mu_n C_G \frac{W}{2L} \left( V_{GS} - V_T \right)^2 \]

for \( V_{DS} \geq V_G - V_T \)

Drain current in the saturation region

*Drain current characteristic of a PMOS FET. The output curves can be distinguished in respect to the linear, the pinch-off and the saturation region.*

Ref.: M.S. Sze, Semiconductor Devices
6.5.2.5 The transfer characteristic

Summary: MOSFET in linear and saturation region

Transfer characteristic of a MOSFET for different threshold voltages. The threshold voltage is defined at the onset of strong inversion, so that the threshold voltage is constant. However, the threshold voltages depend on the surface potential, which is effected by the applied gate voltage. As a consequence a subthreshold region is formed (next slide).

*Drain current characteristic of a PMOS FET for different threshold voltages.*
6.5.2.5 The transfer characteristic

Summary: MOSFET in linear and saturation region

All necessary information like the threshold voltage and the mobility can be extracted from the experimental data (output curves and transfer curves).

Drain current characteristic of a PMOS FET. The transfer curves can be distinguished in respect to the linear and the saturation region.

Ref.: M.S. Sze, Semiconductor Devices
6.5.2.6 Conductance and Transconductance

The channel conductance can be described by

\[ g_D \equiv \frac{\partial I_D}{\partial V_{DS}} = \mu_n C_G \frac{W}{2L} \cdot (V_{GS} - V_T) \]

Channel Conductance

The transconductance of in linear and the saturation region of the MOSFET is given by

\[ g_m \equiv \frac{\partial I_D}{\partial V_G} = \mu_n C_G \frac{W}{L} \cdot V_D \]

Transconductance in the linear region

\[ g_m \equiv \frac{\partial I_D}{\partial V_G} = \mu_n C_G \frac{W}{L} \cdot (V_G - V_T) \]

Transconductance in the saturation region
6.5.2.7 The Sub threshold Region

When the gate voltage is below the threshold voltage and the semiconductor surface is only weakly inverted, the corresponding drain current is called the sub threshold current. In the sub threshold region the drain current is proportional to

\[ I_D \propto \exp\left(\frac{q \cdot (V_G - V_T)}{kT}\right) \]

The sub threshold region is typically characterized by the sub threshold slope. The sub threshold slope should be as small as possible. The transition from the off to the on-state should be as „sharp“ as possible.

*Transfer characteristic and subthreshold region of a PMOS FET.*

Ref.: M.S. Sze, Semiconductor Devices
6.5.2.7 The Sub threshold Region

The sub threshold slope can be calculated by:

\[ S = \frac{\partial \left( \log(I_D) \right)}{\partial V_G} \]

In the case of a silicon MOSFET the sub threshold slope is typically in the range of 50-100mV/decade.

5.5.3 Types of MOSFETs

There are basically four types of MOSFETs, depending on the type of inversion layer. If the transistor is in the off-state for \( V_G \) equal to zero, we speak about enhancement transistors (NMOS and PMOS). The threshold voltage \( V_T \) has to be overcome before the channel starts to conduct. Depending on the doping of the substrate the threshold voltage can be positive or negative and the gate voltage has to be positive or negative to turn the transistor on.
6.5.3 Types of MOSFETs

If the transistor is already conducting for $V_G$ equal to zero we speak about a depletion transistor (NMOS and PMOS). The transistor is already depleted. For example in the case of a p-type substrate a n-type channel is already formed. The channel is formed by „physical“ charges. Therefore, the transistor already conducts for $V_G$. Depletion mode transistors have a threshold voltage, but the threshold voltage is shift to higher positive or negative voltages.

Cross section, output and transfer characteristic of the four different types of MOSFETs.

Ref.: M.S. Sze, Semiconductor Devices
6.5.4 Threshold voltage control

One of the most important parameters of a MOSFET transistor is the threshold voltage. The threshold voltage has to be uniform throughout all transistors on a chip and the threshold voltage has to be reproduceable. Otherwise it is very difficult to realize integrated circuits.

The threshold voltage is given by:

\[ V_T = V_{FB} + \sqrt{2\varepsilon_S q N_A \cdot 2\varphi_B} + 2\varphi_B \]

Threshold voltage

The bulk potential can be controlled very well by adjusting the doping concentration of the substrate. The flat band voltage is defined by the work function difference between the gate and the semiconductor and the interface and trapped charges in the dielectric. The work function of the gate and the channel material can be adjusted quite well. It is difficult to control/predict the influence of the defect and traps in the dielectric on the surface potential and the flat-band voltage.
6.5.4 Threshold voltage control

A further parameter which affects the threshold voltage is the substrate or backgate/bulk voltage. The backgate voltage is applied to the backside of the substrate. The backgate voltage leads to a widening of the depletion region so that the voltage required to achieve strong inversion has to be increased.

Therefore, the backgate voltage will affect the electric field distribution and the voltage drop across the dielectric. The influence of the backgate voltage on the threshold voltage can be considered in the following way:

\[
V_T = V_{FB} + \sqrt{\frac{2\epsilon_s q N_A \cdot (2\varphi_B + V_{BS})}{C_0}} + 2\varphi_B
\]

Threshold voltage

Schematic cross section of an enhanced-type NMOS transistor.
6.5.4 Threshold voltage control

Calculated threshold voltage of n-channel and p-channel MOSFETs as a function of the doping concentration. The gate materials was assumed to be poly silicon. Midgap indicates that the work functions of the gate are in the middle of the energy gap.

Ref.: M.S. Sze, Semiconductor Devices
6.5.4 Threshold voltage control

It is of major importance to control the threshold voltage of a MOSFET.

In general four major strategies exist to control the threshold voltage:

• Firstly, the threshold voltage can be controlled by the doping concentration at the interface between the dielectric and the channel. The doping concentration can be controlled very precisely by ion implantation.

• The second option is the control of the thickness of the gate dielectric. Change the thickness of the gate oxide has of course an influence on other device parameters like to threshold slope. In order to avoid the influence the thickness of the gate of the isolation transistor can be varied. The isolation transistor is a parasitic transistor which is formed in order to isolate the individual transistors from each other. The threshold voltage can be influenced by adjusting the thickness of the field (not the gate) oxide.

• The third option is the use of materials with different work function

• The fours option requires the control of the backgate/bulk voltage
6.5.5 Scaling of MOSFETs

Scaling down of MOSFETs is a continuous trend since the invention of the integrated circuits. Smaller devices enable higher density of transistors and therefore more functionality.

However, the reduction of the device dimensions leads not only to improved device performance. The down scaling of the device leads to several physical and technological problems like

• Threshold voltage Roll-off in the linear region
• Drain Voltage induced barrier lowering
• Bulk punch through
6.5.5.1 Threshold Voltage Roll-off in the linear region

So far we assumed that the gradual channel approximation applies. With decreasing channel length, however, we have to consider effects along the channel. Channel side effects start to affect the performance of the transistors. Channel side effects get more important as the channel length reaches the dimensions below 1\(\mu\)m. It can be observed that with decreasing channel length the threshold voltage is reduced. In the literature the effect is called „Threshold voltage roll-off“.

With decreasing channel length the threshold voltage moves back to zero. The effect can be explained by a „Charge sharing model“. Under such conditions the depletion region of the contacts and the depletion region of the channel overlap. The threshold voltage of a MOS structure can be described by:

\[
V_T = \frac{qN_A W_{\text{max}}}{C_o} + \varphi_s (\text{inv.}) = \frac{\sqrt{2\varepsilon_s qN_A (2\varphi_B)}}{C_o} + 2\varphi_B
\]

Threshold voltage

The influence of the drain and source depletion regions on the threshold voltage can be described by

\[
V_T = \frac{L + L'}{2} \cdot \frac{qN_A W_{\text{max}}}{C_o} + \varphi_s (\text{inv.})
\]

Threshold voltage
6.5.5.1 Threshold Voltage Roll-off in the linear region

For long transistor channels \( L \approx L' \), the influence of the depletion regions of the drain and source contacts on the formation of the depletion region in the channel can be neglected. For short channel transistors \( L > L' \), so that \( L' \) has an influence on the threshold voltage.

With increasing drain-source voltage the depletion region of the drain and source contacts is extended. As a consequence a stronger roll-off of the threshold voltage can be observed.

Schematic cross section of a MOSFET under charge sharing conditions.

Ref.: M.S. Sze, Semiconductor Devices
6.5.5.1 Threshold Voltage Roll-off in the linear region

Threshold voltage roll-off characteristics for CMOS field effect transistors.

Ref.: M.S. Sze, Semiconductor Devices
6.5.5.2 Drain Induced Barrier Lowering in the saturation region

The roll-off of the threshold voltage can be observed in the linear region of operation. In the saturation region the effect is called drain induced barrier lowering. As we are moving from the linear region of device operation towards the saturation region the voltage applied to the drain is increased. As a consequence the depletion region formed by the drain contact is wider than the depletion region formed by the source contact. In the case of a long channel MOSFET this does not have a significant influence on the device performance.

Effect of drain induced barrier lowering for a transistor in the saturation region. Calculated surface potential along the channel for a n-type MOSFET.

Ref.: M.S. Sze, Semiconductor Devices
6.5.5.2 Drain Induced Barrier Lowering in the saturation region

For short channel MOSFETs the effect is different. The increase of the drain-source voltage leads to a decrease of the barrier (surface potential). Therefore, we speak about „Drain induced barrier lowering“.

Subthreshold characteristic of a long channel MOSFET.

Subthreshold characteristic of a short channel MOSFET.

Ref.: M.S. Sze, Semiconductor Devices
6.5.5.3 Bulk punch through

The **bulk punch through** is another limiting factor of a short channel MOSFET. However, the term „punch-through“ is little bit misleading. Punch-through does not mean that the device is destroyed under punch-through conditions.

It has been already be discussed that the depletion region increases with increasing drain-source voltage. For short channel MOSFETs and high drain source voltages the depletion regions of the drain and source contacts merge. This leads to an additional current path parallel to the induced channel. The additional current contribution occurs as a leakage current. As a consequence the MOSFET cannot be turned off anymore for high drain-source voltages.

Ref.: M.S. Sze, Semiconductor Devices
6.5.5.4 Scaling Rules

With deceasing device dimensions of the MOSFET several device parameters can be improved like the power dissipation and the circuit delay. However, short channel effect complicate the scaling of the MOSFETs. In the following we will discuss some guidelines for the scaling of the MOSFETs.

One way to reduce problems related to problems associated to short channel effects is the scaling of all transistor dimensions at the same time. This scaling is called „constant field scaling“. Under such conditions all dimensions are scaled by the scaling factor kappa, $\kappa$.

### Scaling of MOSFETs by the scaling factor $\kappa$ and circuit parameter

Ref.: M.S. Sze, Semiconductor Devices
6.6 Silicon on insulators

For certain applications MOSFETs are fabricated on insulators rather than on semiconductor substrates. In the case of silicon the technology is called SOI (Silicon on insulator) technology. The SOI technology is more expensive but has several advantages. For example the latch-up problem can be avoided. The latch-up problem occurs for CMOS circuits. Due to the necessary diffusion or implantation of N and p wells in the substrates additional parasitic BJTs are formed. These BJTs can affect the performance of the circuit and special design rules have to be considered to avoid the formation of the parasitic BJTs. Furthermore, other parasitic effects occur between the components and the substrate, which complicate the circuit design. The use of SOI technology eliminates a lot of these problems. An insulation layer is either introduced in the silicon wafer or the silicon is directly grown on a insulating substrate.
6.7 Thin Film Transistors

Amorphous silicon and poly silicon are the standard materials for the manufacturing of thin film transistors (TFTs). The transistors are typically deposited on a neutral substrate like glass. Thin Film transistors are very important devices. TFTs are used as switches for LCDs (liquid crystal displays).

Important Parameters:
- Mobility
- Threshold Voltage
- On/off Ratio

Schematic structure of a bottom gate thin film transistor (TFT)
6.7 Field-Effect-Transistors

Two different classes of field effect transistors exist. The first class well known from microelectronics are inversion type of device. Here an inversion layer is formed when applying a voltage larger than the threshold voltage to the gate. In this case the semiconductor is doped and the channel is formed by the inversion of the semiconductor. Thin Film Transistors are fundamentally different. All thin film transistor devices are accumulation type the devices. Here the semiconductor is intrinsic. The channel of the transistor is simply formed by the accumulation of charges.

Inversion-type transistor (transistor in microelectronics) and Accumulation-type transistor (thin film transistor).
6.7 Thin Film Transistors

**Specification**
- Low cost substrates
- Large areas
- Low temperature (150-300°C)

**Materials**
- Amorphous,
- Nanocrystalline,
- and Poly silicon.

**Silicon thin film electronics**

**No Photolithography**
⇒ Printing technologies
- No Vacuum Systems
⇒ Processing at ambient condition
⇒ Processing at very lower temp.
⇒ Processing at room temperature

**Small Molecules Polymers**

**Organic thin film electronics**
6.7 Thin Film Transistors

- **Mobility**
  - Crystalline silicon
  - Poly silicon
  - Amorphous silicon
  - Small molecules
  - Hybrid materials
  - Polymers

- **CMOS technology**
  - CPU, memory products
  - Low Cost ICs, drivers
  - LCD displays

- **Displays, smart cards**
  - RF information tags

- **E paper, E ink**
6.7 Thin Film Transistors

Schematic cross section of a top gate (staggered) thin film transistor (TFT).

Amorphous silicon TFTs are realized as top or bottom gate structures! Silicon nitride is used as an gate dielectric.

Poly silicon and nanocrystalline TFTs with high mobility can only be realized as top gate structure. The gate dielectric has to be silicon oxide!

Organic and polymeric TFTs are realized as top or bottom gate structures!
6.7.1 Amorphous Silicon TFTs

**Advantages:** Inexpensive and reliable technology

**Applications:** Large area applications
Mainly Active Matrix Liquid Crystal Displays (AMLCDs)

**Disadvantages:** Relatively low (electron) mobility: \(~1\text{cm}^2/\text{Vs}\),
Stability,
Bias stress effects,
Performance is most likely not good enough for oLED displays

6.7.2 Poly Silicon TFTs

**Advantages:** High electron mobilities (close to single crystalline silicon),
High stability

**Applications:** High resolution projector displays,
Drivers for LCD displays and OLED displays

**Disadvantages:** Expensive,
High processing temperatures (>400°C),
High off currents
6.7.3 Organic Thin Film Transistors

Pentacene, C_{22}H_{14}:

- Aromatic hydrocarbons based on linear arranged benzene rings
- Tendency to form highly ordered films at low temperatures
- Electronic transport limit: >1 cm²/Vs (electrons / holes)
- Fabrication: Thermal Evaporation:
  - Substrate temperature: 60-70°C
- Not compatible with standard semiconductor processing

Schematic cross section of a bottom gate thin film transistor (TFT)
6.7.3 Organic Thin Film Transistors

Pentacene on thermal oxide

Atomic force micrographs of thermally evaporated pentacene films
6.7.3 Organic Thin Film Transistors

Electronic properties

![Graph showing the electronic properties of Organic Thin Film Transistors.]

**Linear region:** \( V_D < V_G - V_T \)

\[
I_D = C_G \frac{W}{L} \mu_{\text{eff}}^p V_D \left( V_G - V_T - \frac{V_D}{2} \right)
\]

\[
\mu_{\text{eff, lin}} = \frac{L}{W} \cdot \frac{1}{C_G} \frac{d}{dV_G} (I_D)
\]

**Saturation region:** \( V_D > V_G - V_T \)

\[
I_D = C_G \frac{W}{L} \mu_{\text{eff}}^p (V_G - V_T)^2
\]

\[
\mu_{\text{eff, sat}} = \frac{2L}{W} \cdot \frac{1}{C_G} \left( \frac{d}{dV_G} \sqrt{I_D} \right)
\]
References

(Price: US$100), Audience: under graduate students

(Price: US$115), Audience: under graduate students

(Price: US$25 per book), Audience: under graduate students