

MOSFETS - Basics

Introduction

More than 99 % of all ICs are MOSFETs used for random-access memory, flash memory, processors, ASICs (application-specific integrated circuits), and other applications.

In the year 2000, 10^6 MOSFETs per person per year were manufactured.

All MOSFETs are transistors that consist of metal (**M**) SiO_2 (oxide or **O**) and Si (semiconductor or **S**).

Si is a very stable material. Why? Because Si crystallizes in the **diamond structure**, the most stable structure known. Si is in the same column as C, directly below C in the periodic system of elements. SiO_2 is also a very stable material. SiO_2 is the native oxide of Si.

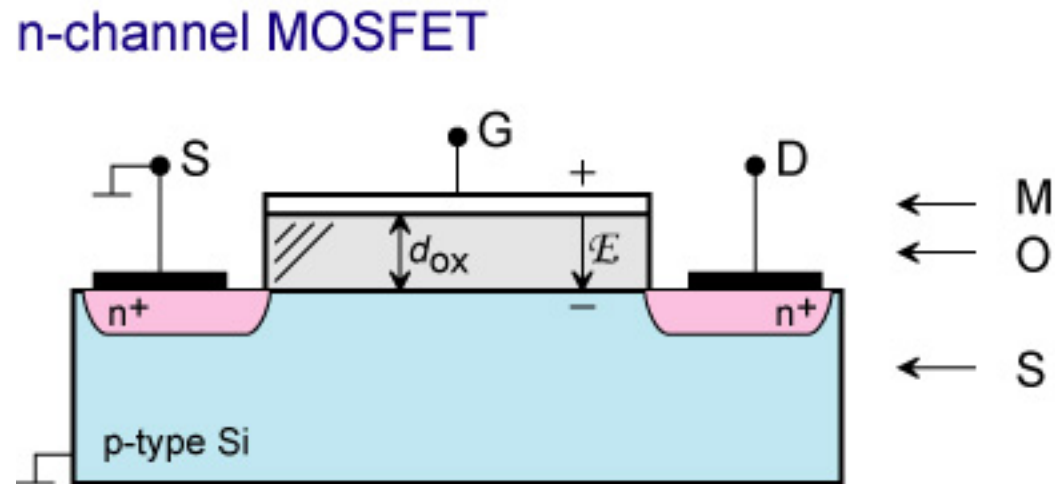
Oxidation reaction: $\text{Si} + \text{O}_2 \leftrightarrow \text{SiO}_2$

MISFET is the general name for **metal-insulator-semiconductor** structures. The most common MISFET is the Si MOSFET so the name MOSFET is much more frequently used than MISFET.

The basic principle of the MOSFET is that the source-to-drain current (SD current) is controlled by the **gate voltage**, or better, by the **gate electric field**. The electric field induces charge (field effect) in the semiconductor at the semiconductor-oxide interface.

Thus the MOSFET is a voltage-controlled current source.

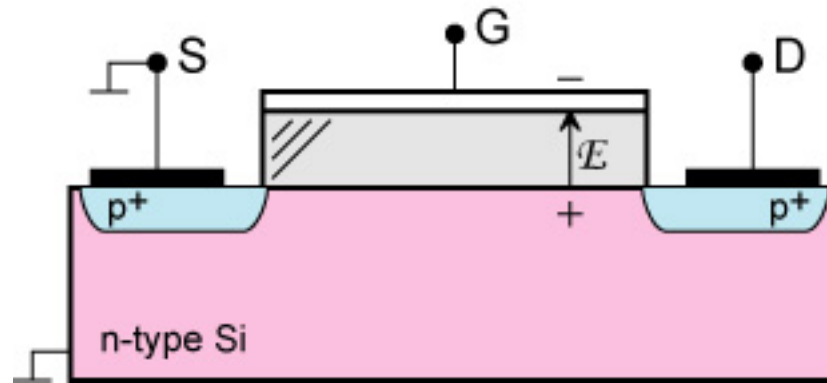
Structure of an n-channel Si MOSFET



An electron (n-type) channel is induced in the p-type semiconductor by positive charges on the gate.

Structure of a p-channel Si MOSFET

p-channel MOSFET



A hole (p-type) channel is induced in the n-type semiconductor by negative charges on the gate.

Thought experiment:

Consider an n-channel MOSFET and a p-channel MOSFET connected in series. Assume that the gates of the two transistors are connected. Thus, regardless of the type of charge on the gate, always one of the transistors is in the “off state” and one of them in the “on state”. This basic circuit thus consumes very little power.

MOSFET circuits consisting of an n-channel and p-channel MOSFET are **complementary MOS** or **CMOS** circuits.

Qualitative discussion of MOSFET operation

It will become clear that all FETs, *i. e.* JFETs, MESFETs, MISFETs, and MOSFETs have similar output characteristics.

We will discuss **n-channel MOSFET** here.

We differentiate between three different voltage regimes of V_{DS} , namely

- (1)** $V_{DS} = 0$,
- (2)** $V_{DS} > 0$, and
- (3)** $V_{DS} \gg 0$.

(1) V_{DS} is very small ($V_{DS} \approx 0$)

$$V_{GS} = 0$$

In this case, no DS current flows. Why? Because we have n^+pn^+ junctions, that is two back-to-back diodes, one of which will be in the reverse direction and therefore block the current flow.

$$V_{GS} > 0$$

We have a moderately positive gate voltage.

This mode is the **depletion mode** of operation. Holes in semiconductor are repelled by positive charge on the gate. The semiconductor is depleted of free holes and a **depletion layer** is created.

$$V_{GS} \gg 0$$

We have a strongly positive gate voltage.

This mode is the **inversion mode** of operation. Electrons are induced in the semiconductor near the oxide-semiconductor interface. An electron current flows from S to D. The magnitude of the gate voltage determines the magnitude of the SD current.



- (2) Small drain-source voltage ($V_{DS} > 0$)
 $V_{GS} \gg 0$ (Inversion mode)

The electric field in the oxide field is **highest** at the source end of the channel. Thus many electrons are induced near the source.

The electric field in the oxide field is **lowest** at the drain end of the channel. Thus few electrons are induced near the drain.

An increasing DS voltage has two effects:

- I_D increases
- Fewer electrons at the drain end of the channel
→ I_D -vs.- V_{DS} begins to have negative curvature.



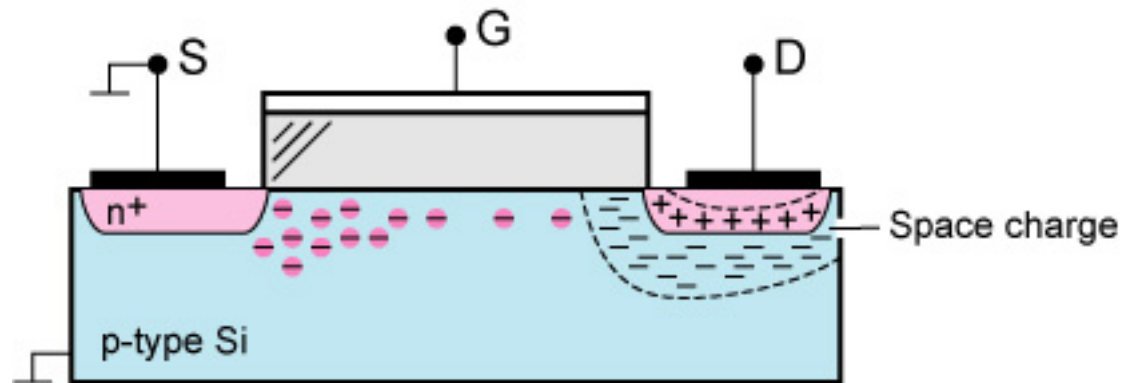
(3) Large drain-source voltage ($V_{DS} \gg 0$)

$V_{GS} \gg 0$ (Inversion mode)

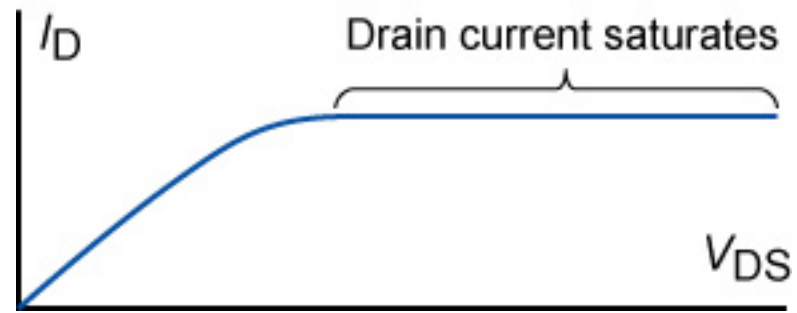
Electric field in the oxide is **highest** at the source end of the channel. Thus there are many electrons near source.

Electric field in the oxide is **very low or zero** at the drain end of the channel. Thus there are no free electrons near drain. The channel is **pinched off**.

Illustration:



Electrons traverse the space charge region of the reverse biased pn^+ junction.



The ideal MOS capacitor

MOS capacitor consists of

- M** Metal with work function Φ_M
- O** The oxide is SiO_2 , also called silicon dioxide or silica. SiO_2 has a large gap $E_g > 5\text{eV}$. SiO_2 is transparent for all visible wavelengths. SiO_2 is a great insulator with a very high breakdown field.
- S** The semiconductor is Si. Work function Φ_{Semi}

We consider here **p-type** Si.

Initially we assume that

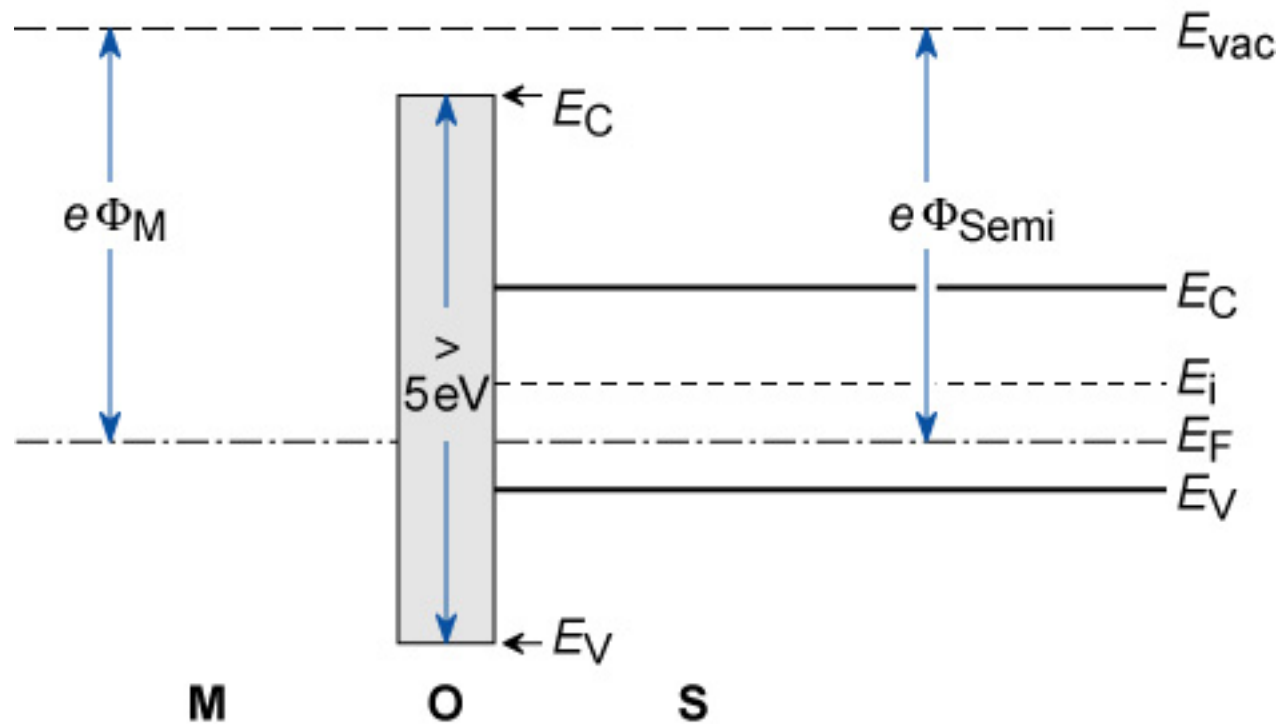
$$\Phi_M = \Phi_{\text{Semi}}$$

where Φ_M is the metal work function

and

Φ_{Semi} is the semiconductor work function.

- (1) $V_G = 0$ (Equilibrium)
 ($E_F = \text{constant throughout structure}$)



In this illustration you can consider the *source* to be *above the page* and the *drain* to be *below the page*. The metal is the gate.

(2) $V_G > 0$ (Accumulation)

The gate bias is **negative**.

This means that E_F at the gate “goes up”.

M and S have much higher conductivity than O.

→ Voltage between gate and channel drops mostly across the oxide

→ An electric field is generated in oxide.

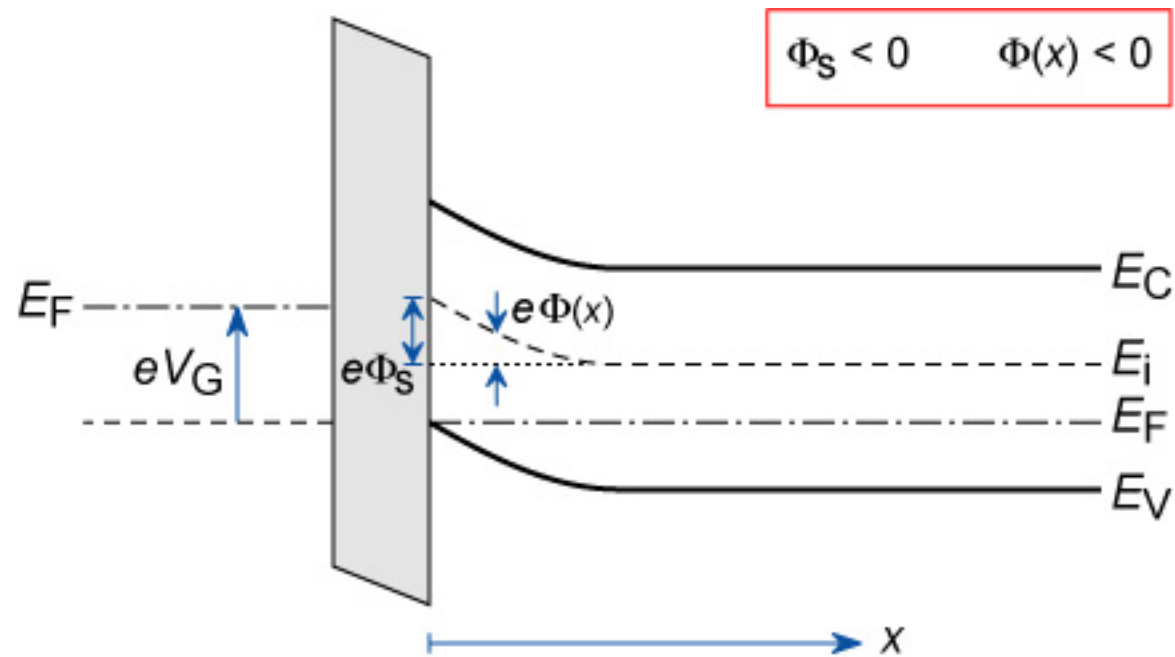
Electrostatic boundary condition of O-S interface:

$$\mathcal{D}_{\text{ox}} = \mathcal{D}_S$$

or

$$\varepsilon_{\text{ox}} E_{\text{ox}} = \varepsilon_S \mathcal{D}_S$$

Band diagram:



Fermi levels are different in M and S.

Fermi levels are constant within M and within S.

Recall that

$$p(x) = n_i e^{(E_i - E_F)/kT} \quad (1)$$

$p(x)$ increases near the surface → **Accumulation** (*i. e.* we have an accumulation of holes near the surface)

Definition of **surface potential** = Φ_S

Surface potential energy = $e \Phi_S$ = difference of bulk value of E_i and surface value of E_i

When bands bend upwards:

$$E_i^{\text{bulk}} - E_i^{\text{surface}} = e\Phi_S < 0 \quad (2)$$

When bands bend downwards:

$$E_i^{\text{bulk}} - E_i^{\text{surface}} = e\Phi_S > 0 \quad (3)$$

Under flatband conditions:

$$\Phi_S = 0 \quad (4)$$

Introducing dependence of the potential Φ on the position x :

$$e\Phi(x) = E_i^{\text{bulk}} - E_i(x) \quad (5)$$

$$\Phi_S = \Phi(x=0) \quad (6)$$

Using Eqs. (1) and (5), one obtains

$$\begin{aligned} p(x) &= n_i e^{(E_i - E_F) / kT} \\ &= n_i e^{[E_i^{\text{bulk}} - e\Phi(x) - E_F] / kT} \\ &= n_i e^{(E_i^{\text{bulk}} - E_F) / kT} e^{-e\Phi(x) / kT} \end{aligned}$$

$$p(x) = p_0 e^{-e\Phi(x) / kT} \quad (7)$$

Since $\Phi(x) < 0$, $p(x)$ increases close to the surface.

That is, we have **accumulation**.

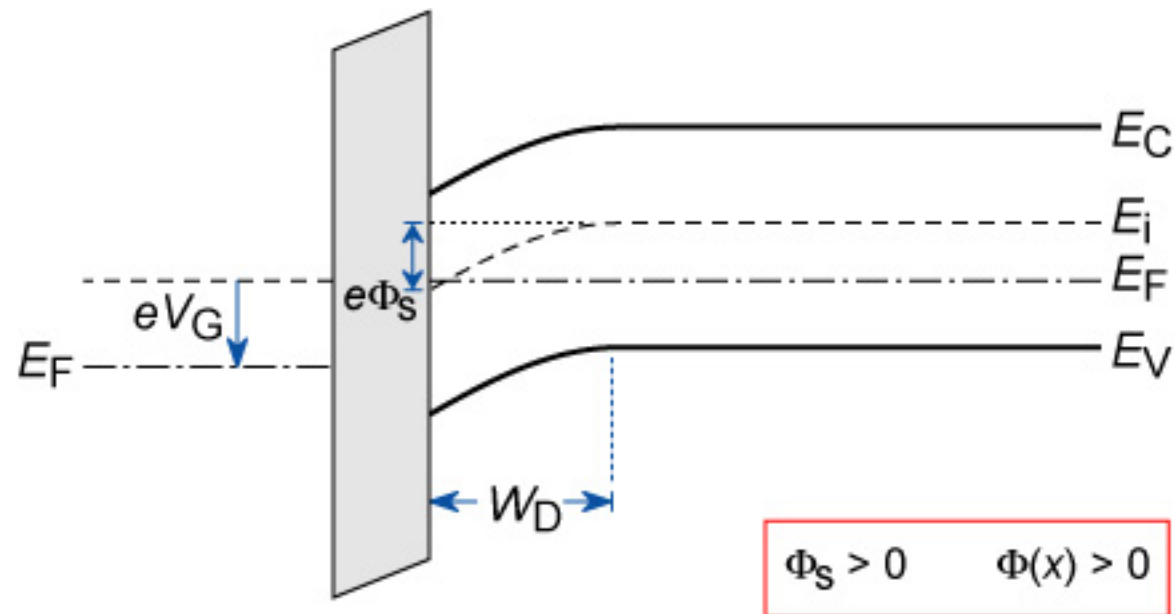
That is the result of Eq. (7) is consistent with band diagram.

(3) $V_G > 0$ (Depletion)

The gate bias is positive.

E_F “goes down” in the metal.

Band diagram:



Semiconductor is depleted near surface.

The depletion layer thickness follows from Poisson's equation:

$$W_D = \sqrt{\frac{2\varepsilon}{eN_A} \Phi_S} \quad (8)$$

E_F is near E_i at the surface.

→ Semiconductor is practically **intrinsic** at the surface.

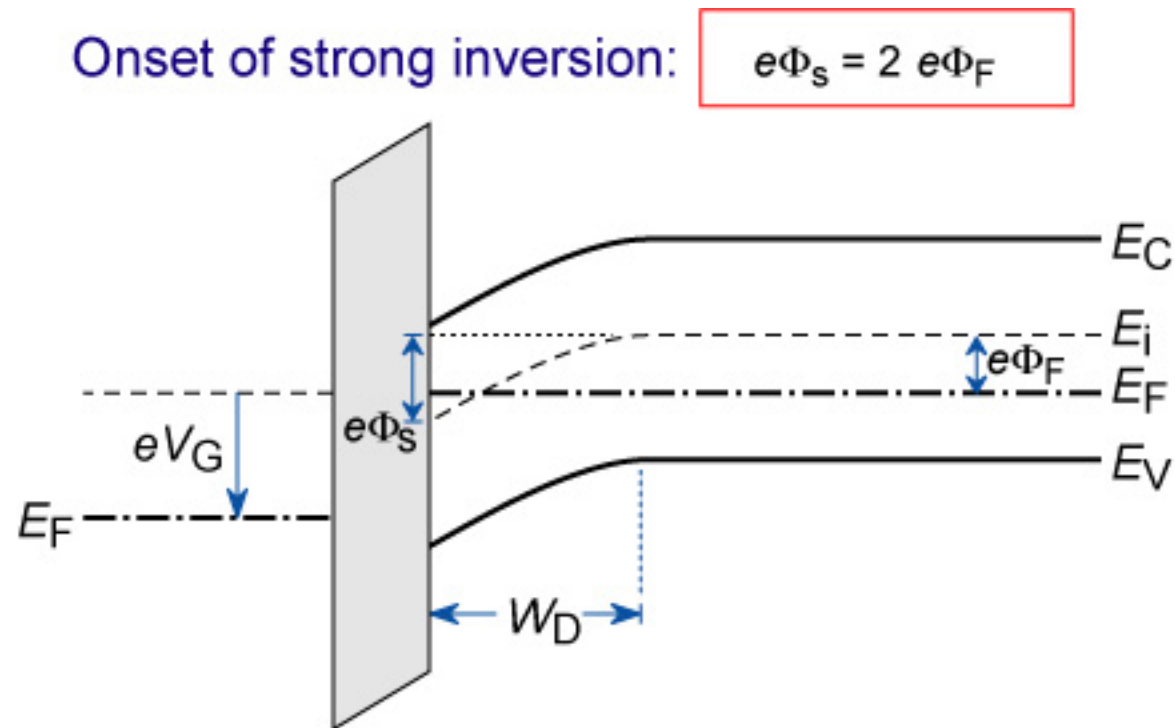
Recall Eq. (7): $p(x) = p_0 e^{-e\Phi(x)/kT}$

It is $\Phi(x) > 0 \rightarrow p < p_0$, that is, we have a depleted layer near the surface.

(4) $V_G \gg 0$. (Onset of strong inversion)

The gate bias is position.

E_F goes further down in metal.



Semiconductor is depleted of holes near surface.

E_F is closer to E_C than to E_V at the surface.

→ Semiconductor is **n-type** near surface

→ Conductivity type of semiconductor is **inverted**.

Criterion for the **onset of strong inversion**:

$$e\Phi_S = 2e\Phi_F \quad (\text{Onset of strong inversion}) \quad (9)$$

where

$$e\Phi_F = E_i^{\text{bulk}} - E_F^{\text{bulk}} \quad (10)$$

Onset of strong inversion means that the semiconductor is as strongly n-type at the surface as it is p-type in the bulk.

Using Boltzmann statistics

$$p = n_i e^{(E_i - E_F) / kT} \quad (11)$$

and Eqs. (9) and (10), one obtains

$$e\Phi_S = 2e\Phi_F = 2kT \ln \frac{N_A}{n_i} \quad (\text{Onset of strong inversion}) \quad (12)$$

At the onset of strong inversion, an n-channel begins to be formed at the semiconductor surface.

The depletion layer thickness at the onset of strong inversion is given by:

$$W_{D, \max} = \sqrt{\frac{2\varepsilon}{eN_A} 2\Phi_F} = \sqrt{\frac{2\varepsilon}{eN_A} 2 \frac{kT}{e} \ln \frac{N_A}{n_i}}$$

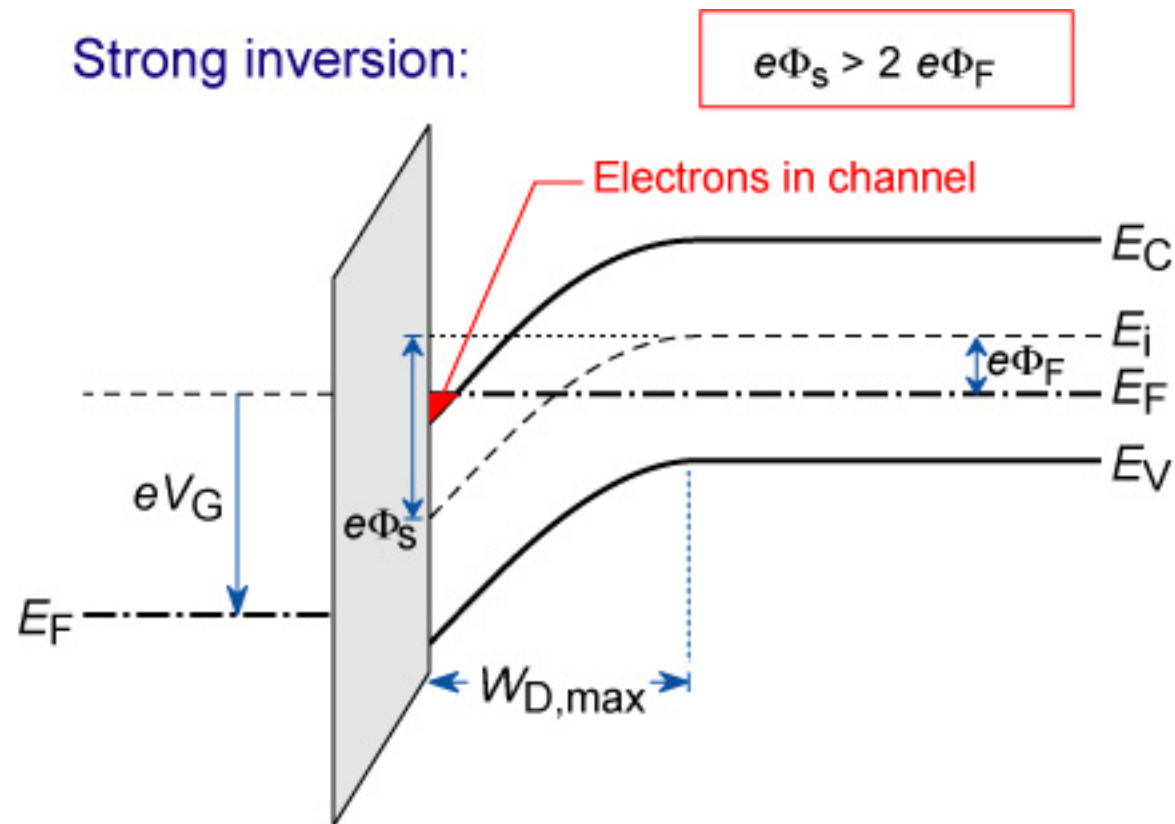
$$W_{D, \max} = 2 \sqrt{\frac{\varepsilon kT}{e^2 N_A} \ln \frac{N_A}{n_i}} \quad (13)$$

$W_{D, \max}$ is the maximum depletion layer thickness.

A further increase in V_G will result in **more inversion** rather than in **more depletion**.

(5) Let's go **beyond the onset** of strong inversion.

Strong Inversion



Beyond the onset of strong inversion, electrons are filled into the electron channel. The depletion layer thickness does not increase further, *i. e.* $W_D = W_{D, \max}$.

Example:

Find $W_{D, \max}$ for ideal Si MOS capacitors, one of them having a background doping of $N_A = 10^{16} \text{ cm}^{-3}$ and one of them having a background doping of $N_A = 10^{18} \text{ cm}^{-3}$.

Solution: It is

$$n_i = 10^{10} \text{ cm}^{-3}, \quad \epsilon_r = 9$$

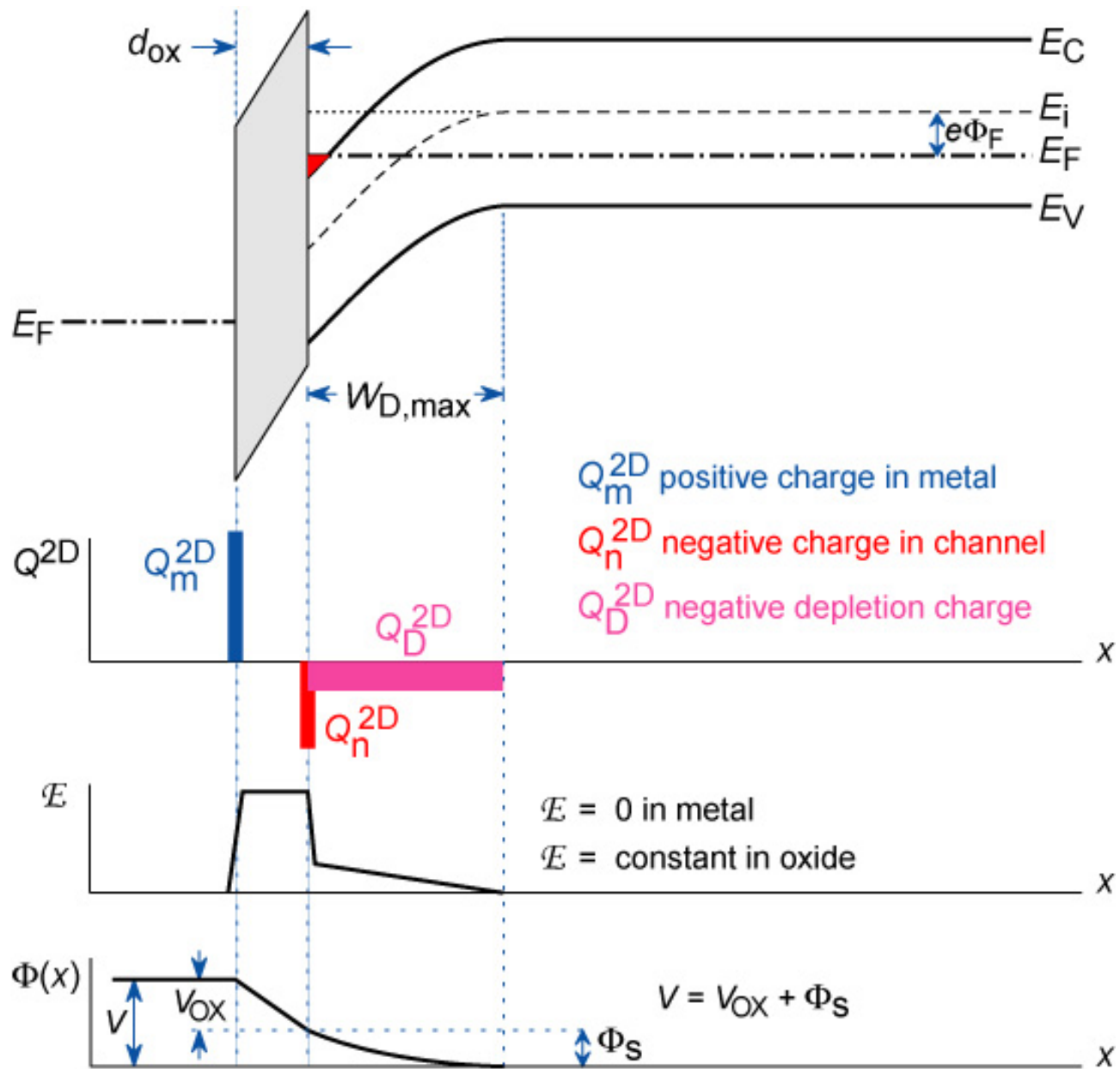
$$W_{D, \max} = 2 \sqrt{\frac{\epsilon kT}{e^2 N_A} \ln \frac{N_A}{n_i}}$$

$$N_A = 10^{16} \text{ cm}^{-3} \Rightarrow W_{D, \max} = 0.27 \mu\text{m} = 2700 \text{ \AA}$$

$$N_A = 10^{18} \text{ cm}^{-3} \Rightarrow W_{D, \max} = 0.03 \mu\text{m} = 300 \text{ \AA}$$

Band diagram, charge, field, and potential of ideal MOS capacitor

(see next page)



We use the **depletion approximation** for the acceptor charge in the semiconductor:

$$Q_D^{3D} = e N_A^- \quad \text{for } 0 \leq x \leq W_D \quad (14)$$

$$Q_D^{3D} = 0 \quad \text{for } x > W_D \quad (15)$$

where “ Q^{3D} ” is the “charge per unit volume”. Analogously, we will denote a charge per unit area as “ Q^{2D} ”.

Threshold voltage of ideal MOS capacitor

The **overall charge** of a MOS device (or any other device) is zero.

Charge neutrality thus requires:

$$Q_M^{2D} = -Q_S^{2D} = -\left(Q_D^{2D} + Q_n^{2D}\right) \quad (16)$$

where Q_M^{2D} = metal charge, Q_S^{2D} = semiconductor charge, Q_D^{2D} = depletion charge, and Q_n^{2D} = electron charge.

$$Q^{2D} = \text{Charge per unit area} \quad (17)$$

$$Q_D^{2D} = -eN_A W_D \quad (18)$$

Voltage across MOS capacitor

$$V = V_{\text{OX}} + \Phi_S \quad (19)$$

where

$$\begin{aligned} V_{\text{OX}} &= -E d_{\text{OX}} \\ &= -(Q_S^{2\text{D}} / \epsilon_{\text{OX}}) d_{\text{OX}} \\ &= -Q_S^{2\text{D}} / C_{\text{OX}}^{2\text{D}} \end{aligned} \quad (20)$$

$$V_{\text{OX}} = -\frac{Q_S^{2\text{D}}}{C_{\text{OX}}^{2\text{D}}} = \frac{Q_S^{2\text{D}}}{\epsilon_{\text{OX}} / d_{\text{OX}}} \quad (21)$$

Note that

$$C_{\text{OX}} = C_{\text{OX}}^{2\text{D}} A \quad (22)$$

where

C_{OX} = Capacitance

$C_{\text{OX}}^{2\text{D}}$ = Capacitance per unit area

A = Area

At threshold, an **electron channel** is **induced** at the O-S interface. This happens at the onset of **strong inversion**.

Onset of strong inversion:

$$Q_n^{2D} \approx 0 \quad (23)$$

$$Q_S^{2D} = -e N_A W_{D, \max} = Q_D^{2D} \quad (24)$$

$$\Phi_S = 2\Phi_F \quad (25)$$

Insertion of Eqs. (23) – (25) into Eqs. (19) and (21) yields

$$V_{th} = -\frac{Q_D^{2D}}{C_{OX}^{2D}} + 2\Phi_F = \frac{e N_A W_{D, \max}}{C_{OX}^{2D}} + 2\Phi_F \quad (26)$$

Thus the threshold voltage is the sum of a voltage drop in the oxide and in the semiconductor at the onset of strong inversion. Eq. (26) applies to the **ideal MOS structure**.

Capacitance of ideal MOS capacitor

Capacitance of oxide capacitor:

$$C_{\text{OX}}^{2\text{D}} = \frac{\epsilon_{\text{OX}}}{d_{\text{OX}}} \quad (27)$$

Capacitance of depletion layer:

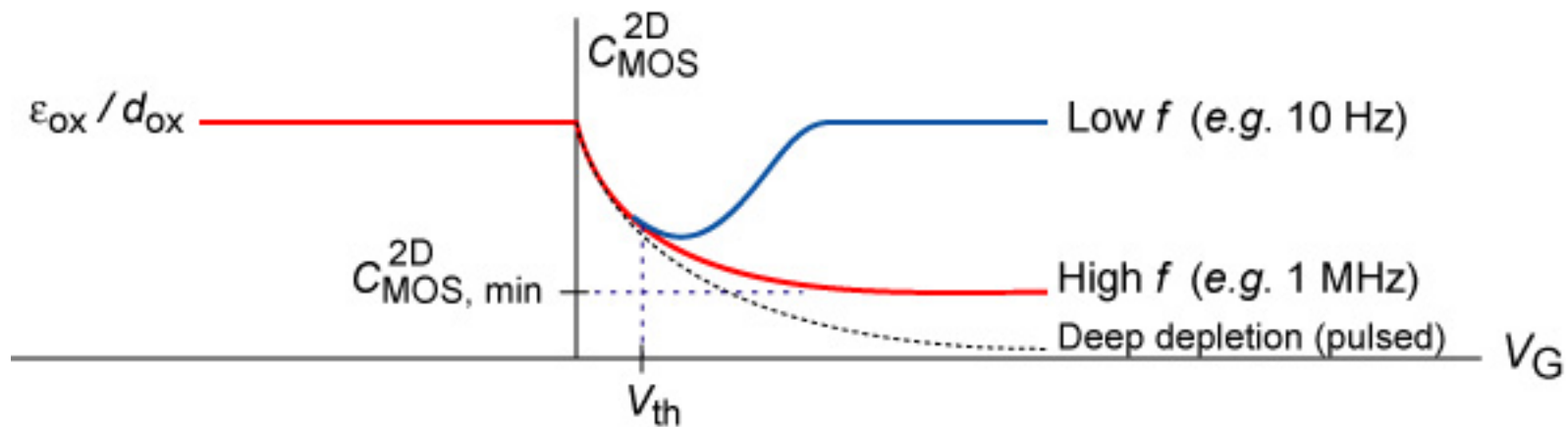
$$C_{\text{D}}^{2\text{D}} = \frac{\epsilon_{\text{S}}}{W_{\text{D}}} \quad (28)$$

We have two capacitors in series. Thus, the total capacitance is given by:

$$C_{\text{MOS}}^{2\text{D}} = \left(\frac{1}{C_{\text{OX}}^{2\text{D}}} + \frac{1}{C_{\text{D}}^{2\text{D}}} \right)^{-1} \quad (29)$$

$$C_{\text{MOS}}^{2\text{D}} = \left(\frac{d_{\text{OX}}}{\epsilon_{\text{OX}}} + \frac{W_{\text{D}}}{\epsilon_{\text{S}}} \right)^{-1} \quad (30)$$

The following illustration shows the $C_{\text{MOS}}^{2\text{D}}$ - vs. - V curve. Note that W_{D} depends on V .



$$C_{\text{MOS}, \text{min}}^{2\text{D}} = \left(\frac{d_{\text{OX}}}{\epsilon_{\text{OX}}} + \frac{W_{\text{D}, \text{max}}}{\epsilon_{\text{S}}} \right)^{-1} \quad (31)$$

Discussion of $C_{\text{MOS}}^{2\text{D}}$ - vs. - V curve

$$V_G > 0$$

Accumulation

Holes accumulate at the O-S interface

$$C_{\text{MOS}}^{2\text{D}} = \epsilon_{\text{OX}} / d_{\text{OX}}$$

$$V_G > 0$$

Depletion

Depletion layer thickness increases with V

$$V = V_{\text{th}}$$

Onset of strong inversion

$$W_D = W_{D, \text{max}}$$

$$C_{\text{MOS}, \text{min}}^{2\text{D}} = \left(\frac{d_{\text{OX}}}{\epsilon_{\text{OX}}} + \frac{W_{D, \text{max}}}{\epsilon_S} \right)^{-1}$$

$$V > V_{th}$$

(1) Low frequency

An inversion channel is formed at the O-S interface.

$$C_{MOS}^{2D} = \epsilon_{OX} / d_{OX}$$

(2) High frequency

Electron-hole pairs are generated too slowly to follow the ac signal of the measurement.

$$C_{MOS}^{2D} = C_{MOS, min}^{2D}$$

Note: Using short pulsed measurements, a state called “**deep depletion**” can be created. In the deep depletion mode, the depletion layer can be extended over what was referred to as $W_{D, max}$. The pulses need to be very short, to prevent the formation of an inversion channel.

Real MOS capacitor

Generally, there is a work function difference between metal and semiconductor. That is

$$\Phi_M \neq \Phi_S$$

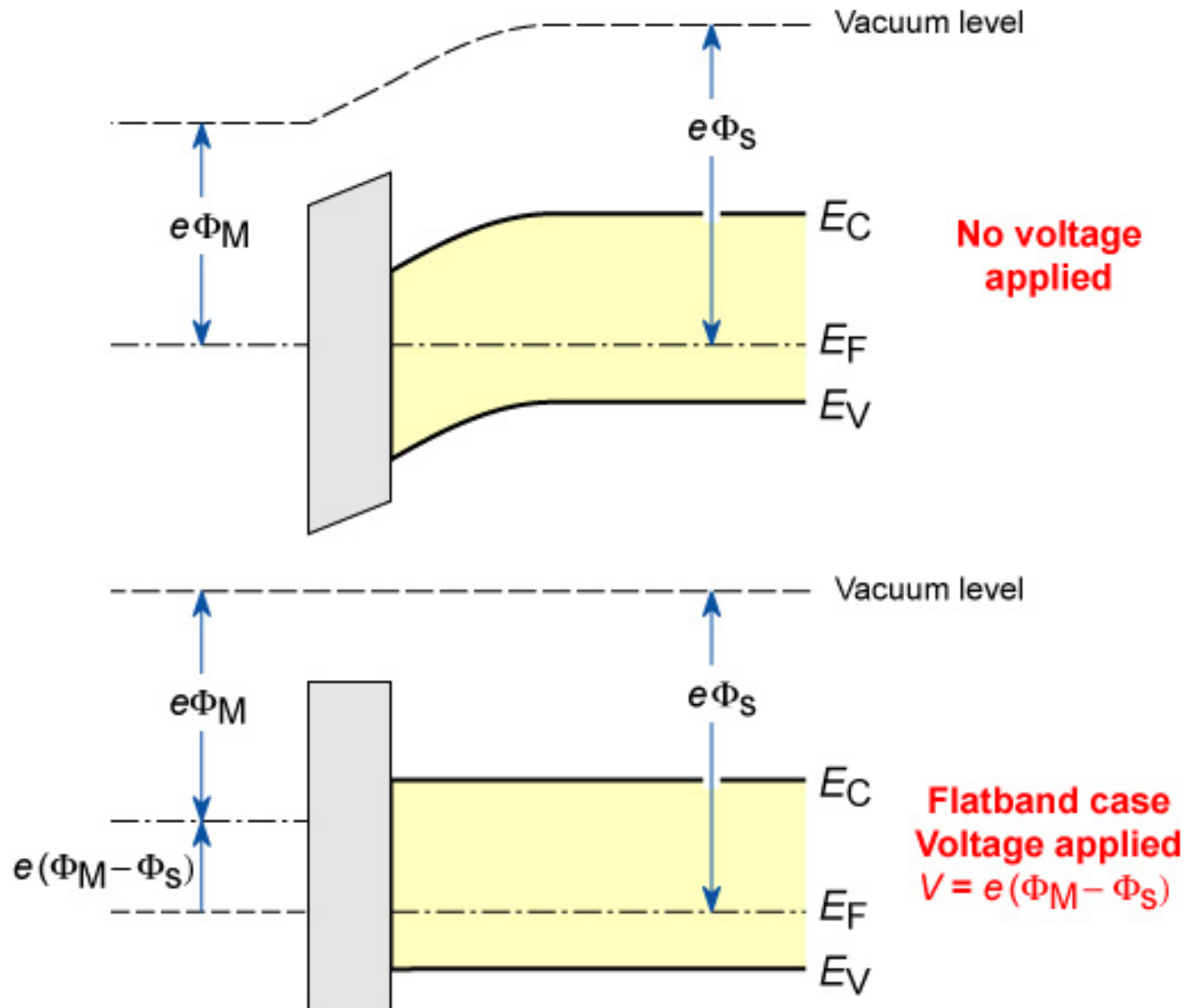
Work function difference

$$\Phi_{MS} = \Phi_M - \Phi_S \quad (32)$$

There are usually charges trapped in the oxide, for example Na⁺ (sodium) ions. The oxide charges produce a voltage:

$$V = Q_{OX}^{2D} / C_{OX}^{2D} \quad (33)$$

Band diagram (for $\Phi_{MS} \neq 0$ and $Q_{OX} \neq 0$)



Flatband voltage

Adding Eqs. (32) and (33) yields

$$V_{\text{FB}} = \Phi_{\text{MS}} - \frac{Q_{\text{OX}}}{C_{\text{OX}}} \quad (34)$$

where Q_{OX} is an effective positive charge at the O-S interface.

Threshold voltage

Eq. (26) holds for the **ideal** MOS structure. In the case of a **real** MOS structure, the effects of the work function difference and oxide charges must be included. The threshold voltage for the **real** MOS structure is given by:

$$V_{th} = \Phi_{MS} - \frac{Q_{OX}^{2D}}{C_{OX}^{2D}} - \frac{Q_D^{2D}}{C_{OX}^{2D}} + 2\Phi_F \quad (35)$$

MOSFET Operation

We will use again Shockley's **gradual channel approximation**.

Charge density in channel: Q_n^{2D}

Below threshold: $Q_n^{2D} = 0$

Above threshold: $|Q_n^{2D}| > 0$

Recall charge neutrality condition

$$\left| Q_M^{2D} \right| = Q_S^{2D} = Q_n^{2D} + Q_D^{2D} \quad (36)$$

$$Q_n^{2D} = Q_S^{2D} - Q_D^{2D} \quad (37)$$

Recall Gauss's law

$$\mathcal{E} = -\frac{1}{\varepsilon} \int_{-\infty}^x \rho(x) dx = -\frac{1}{\varepsilon} Q^{2D} \quad (38)$$

Using Gauss's law in Eq. (37) yields:

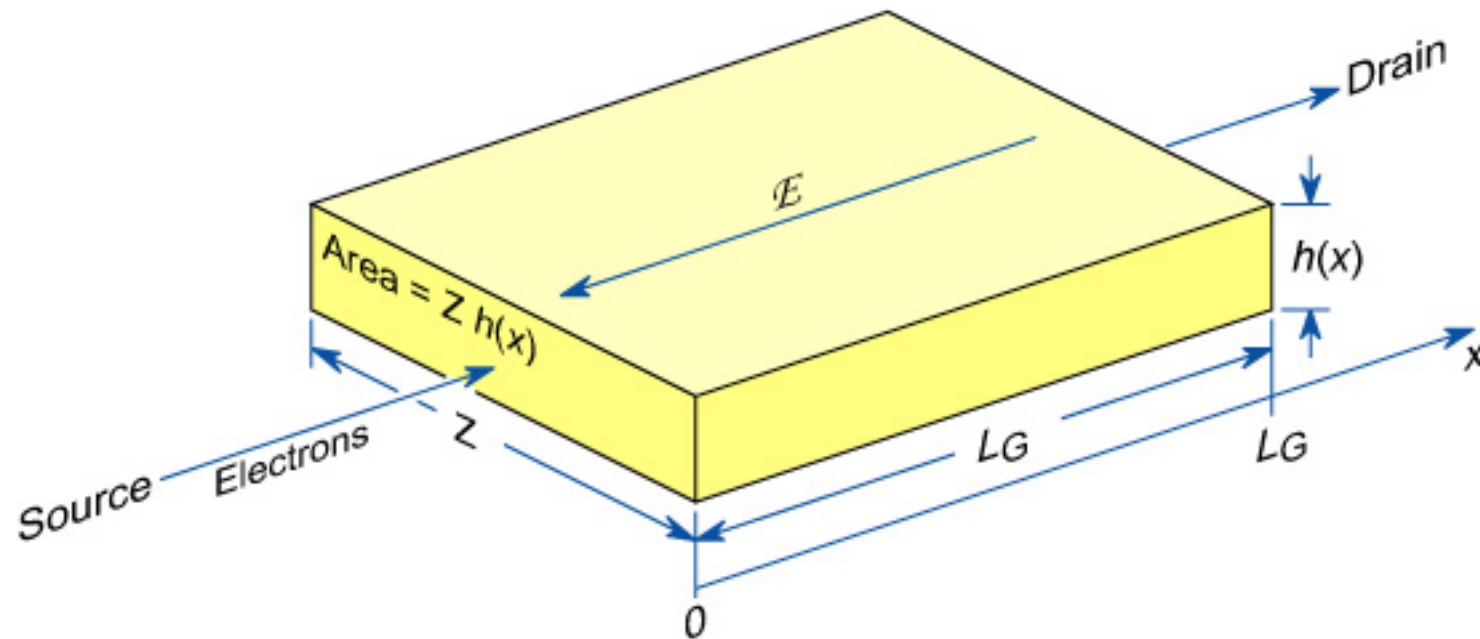
$$\begin{aligned} Q_n^{2D} &= -\varepsilon_{\text{OX}} \mathcal{E}_{\text{OX}}(x) + \varepsilon_{\text{OX}} \frac{V_{\text{th}}}{d_{\text{OX}}} \\ &= -\varepsilon_{\text{OX}} \frac{V_{\text{GS}} - V(x)}{d_{\text{OX}}} + \varepsilon_{\text{OX}} \frac{V_{\text{th}}}{d_{\text{OX}}} \end{aligned} \quad (39)$$

$$Q_n^{2D} = -\frac{\varepsilon_{\text{OX}}}{d_{\text{OX}}} [V_{\text{GS}} - V_{\text{th}} - V(x)] \quad (40)$$

where $V(x)$ is the channel-to-source voltage.

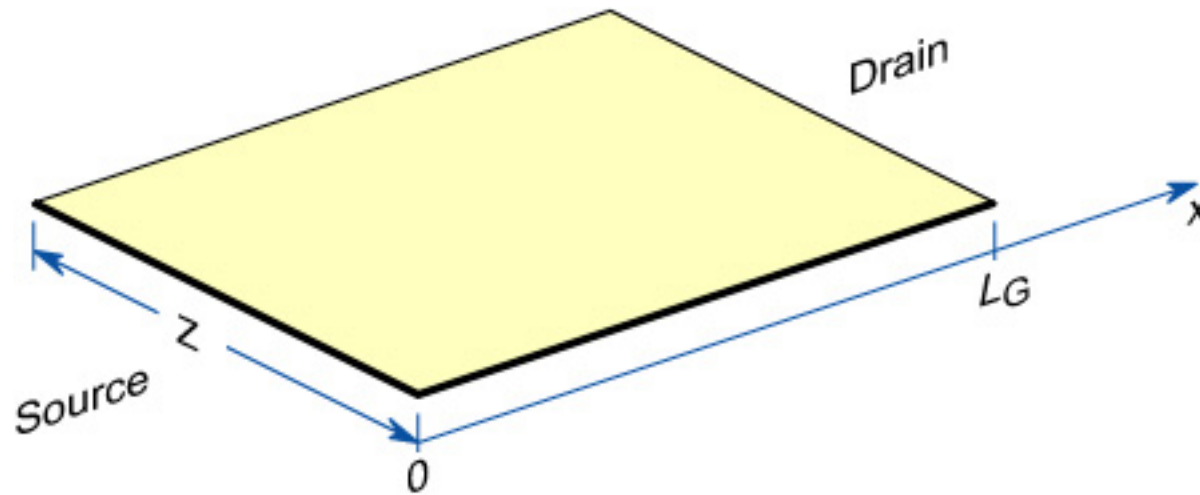
Illustration of sheet charge

(1) MESFET



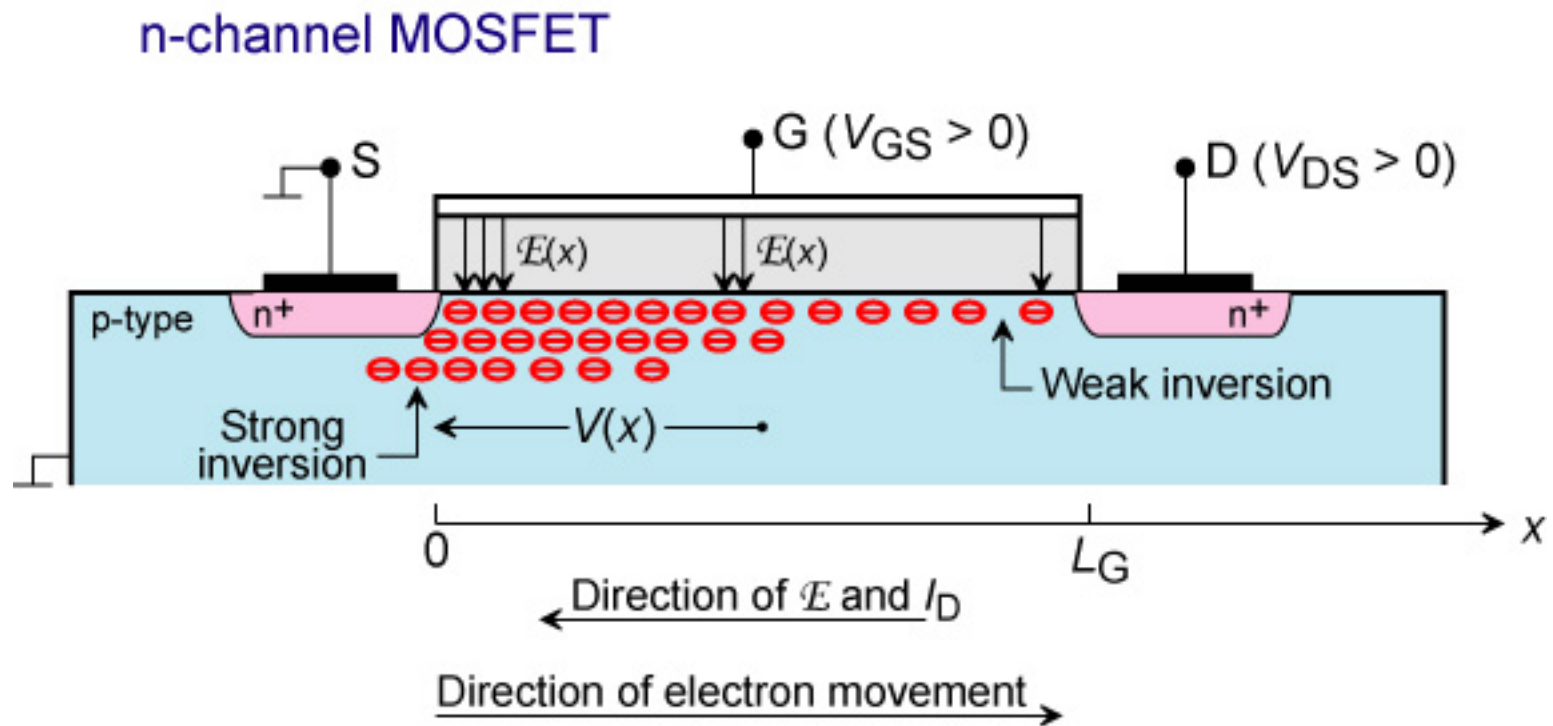
Carrier concentration:	$n(x)$	(cm^{-3})
Sheet carrier concentration:	$n^{2D}(x) = n(x) h(x)$	(cm^{-2})
Sheet charge density:	$Q_n^{2D} = e n(x) h(x)$	(C / cm^2)

(2) MOSFET $h(x) \rightarrow 0$ (Inversion layer)



Thickness of inversion layer	$h(x) \rightarrow 0$	
Sheet carrier concentration	$n^{2D}(x)$	(cm^{-2})
Sheet charge density	$Q_n^{2D} = e n^{2D}(x)$	(C / cm^2)

Gradual channel approximation



Recall from MESFET: $I_D = -e n v(x) h(x) Z$

However, in the MOSFET we have:

$$I_D = Q_n^{2D}(x) v(x) Z \quad (41)$$

$$= Q_n^{2D}(x) [-\mu_n E(x)] Z \quad (42)$$

$$= Q_n^{2D}(x) \mu_n \frac{dV(x)}{dx} Z \quad (43)$$

Insertion of Eq. (40) into Eq. (43):

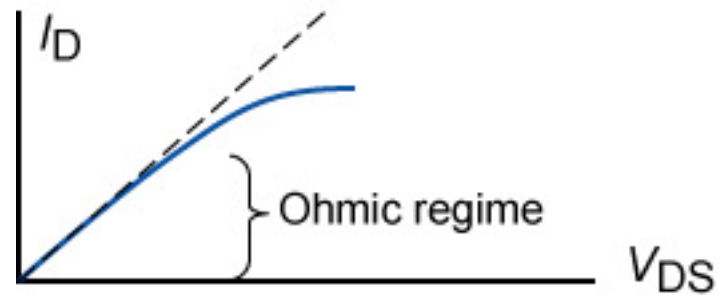
$$I_D = -\frac{\epsilon_{OX}}{d_{OX}} Z \mu_n [V_{GS} - V_{th} - V(x)] \frac{dV(x)}{dx} \quad (44)$$

Separation of variables and integration yields

$$I_D \int_0^{L_G} dx = -\frac{\epsilon_{OX} \mu Z}{d_{OX}} \int_0^{V_{DS}} (V_{GS} - V_{th} - V(x)) dV(x) \quad (45)$$

$$(-) I_D = \frac{\epsilon_{OX} \mu Z}{d_{OX} L_G} \left[(V_{GS} - V_{th}) V_{DS} - \frac{1}{2} V_{DS}^2 \right] \quad (46)$$

- ... current I_D is negative since it flows in the negative x direction. However, since the direction is not important to us, the “-” sign is put in parenthesis.
- ... first term in bracket (minuend) increases linearly with V_{DS} .
- ... second term in bracket (subtrahend) increases parabolically with V_{DS} . Second term is negative.

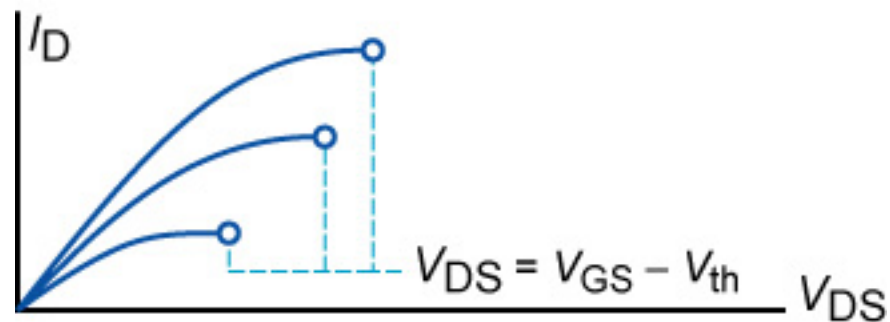


... for very small V_{DS} , the term $(\frac{1}{2}) V_{DS}^2$ can be neglected.

→ Then I_D depends linearly on V_{DS} .

→ This is the ohmic regime.

At $V_{DS} = V_{GS} - V_{th}$, the slope $dI_D / dV_{DS} = 0$.

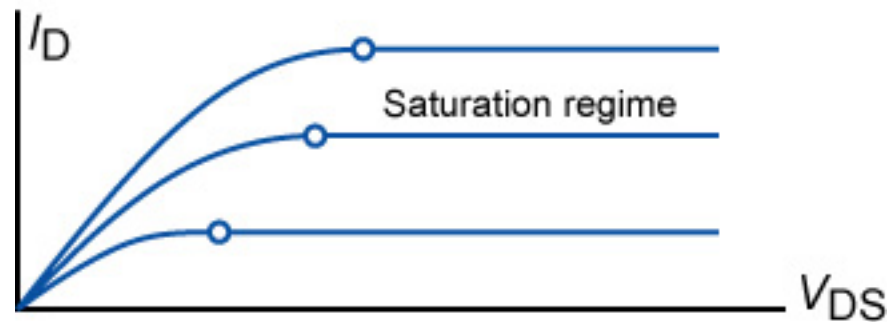


Onset of saturation:

$$V_{DS, \text{sat}} = V_{GS} - V_{th} \quad (47)$$

Insertion of Eq. (49) into Eq. (46) yields

$$I_{D, \text{sat}} = \frac{\epsilon_{OX} \mu Z}{d_{OX} L_G} \frac{1}{2} V_{DS, \text{sat}}^2 = \frac{\epsilon_{OX} \mu Z}{2 d_{OX} L_G} (V_{GS} - V_{th})^2 \quad (48)$$



Transconductance

$$g_{m, \text{sat}} = \frac{dI_{D, \text{sat}}}{dV_{GS}} = \frac{\epsilon_{OX} \mu Z}{d_{OX} L_G} (V_{GS} - V_{th}) \quad (49)$$

The equation shows that in order to get a high transconductance, the following properties are desirable:

- (1) High mobility
- (2) Thin oxide (in 2002: $d_{OX} \approx 40 \text{ \AA}$) or dielectric with high ϵ_r
- (3) Short gate length

Transconductance increases with gate width Z . But increasing Z , also increases the area of the FET.

Frequently used figure of merit:

g_m / Z = Transconductance per unit gate width.

Questions:

What are the units of g_m / Z ?

How does g_m depend on Z ?

How does the input capacitance (C_{GS}) depend on Z ?

What is the relevance of g_m / C_{GS} ?

How does g_m / C_{GS} depend on Z ?