Chapter 2

MOS Transistors

2.1 Structure of MOS transistors

We will discuss the structure of two MOS Field-Effect-Transistors (FETs) that are building blocks for all digital devices.
The **nMOS** transistor shown in Figure 2.1 (n-type, n-channel, enhancement mode field-effect transistor) is built on the p-type semiconductor substrate, which is usually acceptor-doped silicon.

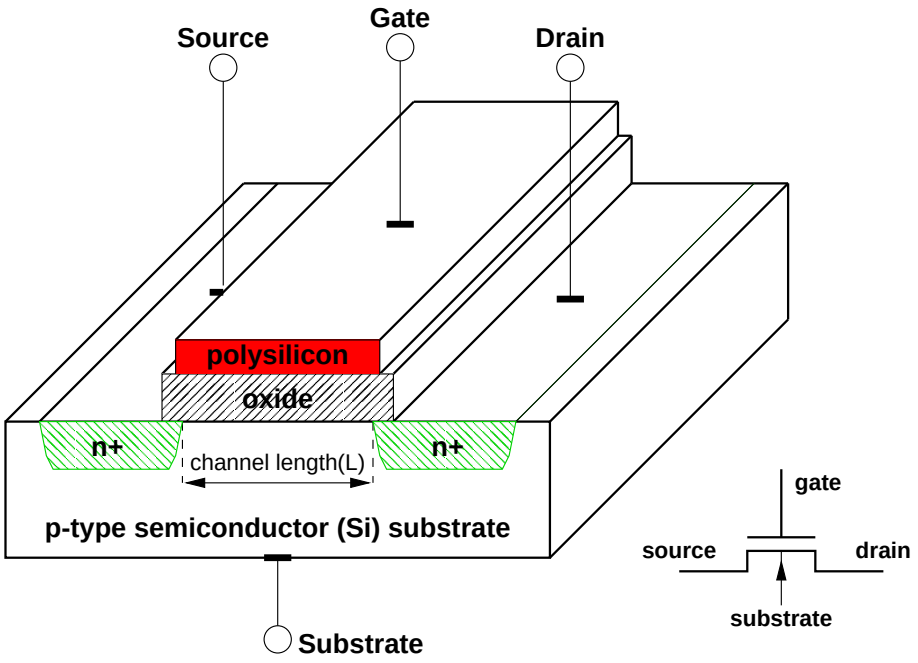

Figure 2.1: An internal structure of an nMOS transistor.

- Two **n+ diffusion** regions ('+' indicates the high degree of doping) form the source and drain of the transistor. The area in between forms a conducting channel. Potentially, electrons, negative carriers, will form the current in the channel.
- The gate, which is formed from a conductor, typically polysilicon, is insulated from the source-channel-drain structure (and from the substrate) by the layer of silicon dioxide.
- The voltage between the gate and the substrate induces the electric field which controls the flow of the carriers in the channel. This gives the rise to the name: field-effect transistor (FET).
- Transistor structure is completely **symmetrical** with respect to the source and drain. The role is defined by terminal voltages which establish the direction of the current (carriers) flow.

The **pMOS transistor** (p-type, p-channel) is a complementary structure to the nMOS transistor as depicted in Figure 2.2.

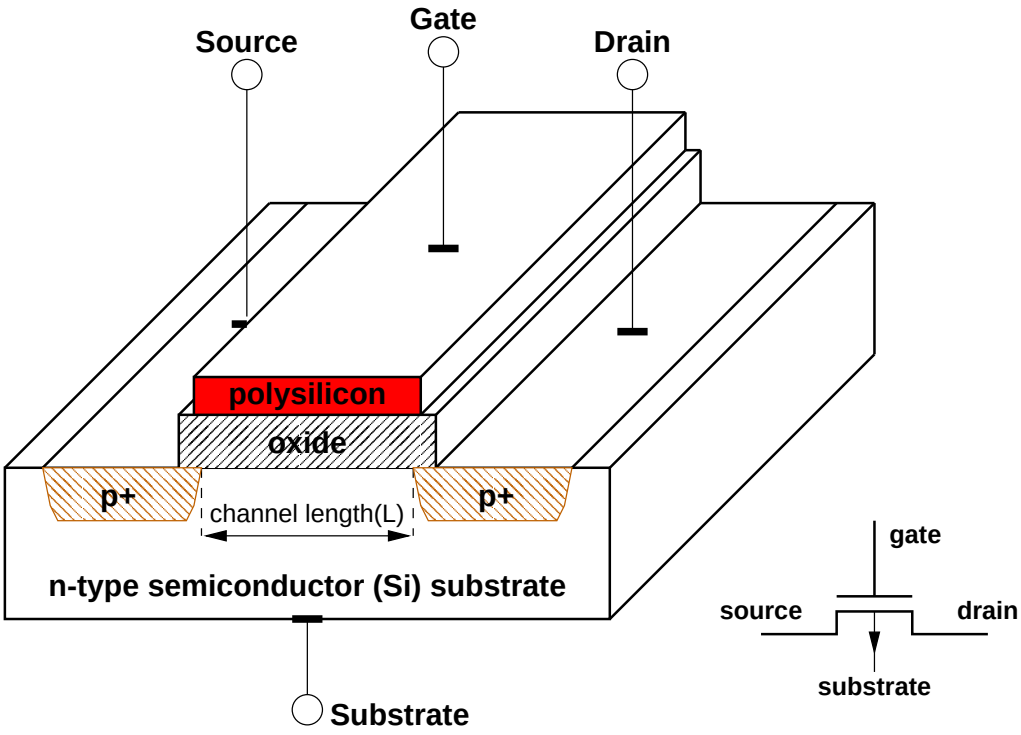

Figure 2.2: An internal structure of a pMOS transistor.

- The pMOS transistor is built on the n-type substrate which is donor-doped silicon.
- The source and drain of a pMOS transistor are now **p+ diffusion** regions.
- The carriers in the channel are now positive holes. As previously, their flow is controlled by the gate-substrate voltage.

MOS transistors described above are referred to as **enhancement mode** transistors.
There are also **depletion mode** transistors used mainly in the analog circuitry.

2.2 Operation of MOS transistors

We will describe operation of an enhancement-mode n-channel MOS field effect transistor (nMOS) as illustrated in Figure 2.1. The pMOS operates in the dual way. The basic principle of operation can be stated as follows.

> The flow of the current between the source and the drain is controlled by the electric field generated by the gate-substrate voltage.

In order for the drain-source current to exist there must be carriers existing in the area between the source and drain referred to as the conducting channel. We will examine first how the channel is created and then how the drain current depends on relevant voltage. Note that a MOS transistor is a four terminal device. In most cases, the substrate and the source of an nMOS are connected to the ground potential (GND) as in Figure 2.3.

Step 1: (Figure 2.3) $V_{GS} < V_T \implies I_D = 0$

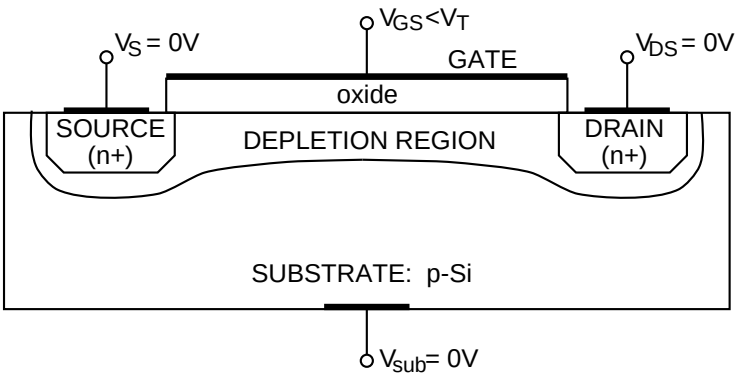

Figure 2.3: Formation of a depletion region in an enhancement-mode nMOS FET.

We start with case when the **gate voltage**, V_{GS} is smaller than the **threshold voltage**, V_T. The drain voltage, V_{DS} is in this situation irrelevant and can be zero.

The electric field induced by the gate voltage points down from the gate through the channel. This fields repels the majority carriers for the p-type substrate, that is, positive holes, from the channel hence forming a region depleted of carriers as shown in Figure 2.3.

As a result, due to the lack of free carriers, no current flows between the source and the drain at this stage, that is, $I_D = 0$.

Step 2: (Figure 2.4) $V_{GS} > V_T$ and $V_{DS} = 0 \implies I_D = 0$

In this case (Figure 2.4), when the gate voltage V_{GS} increases above

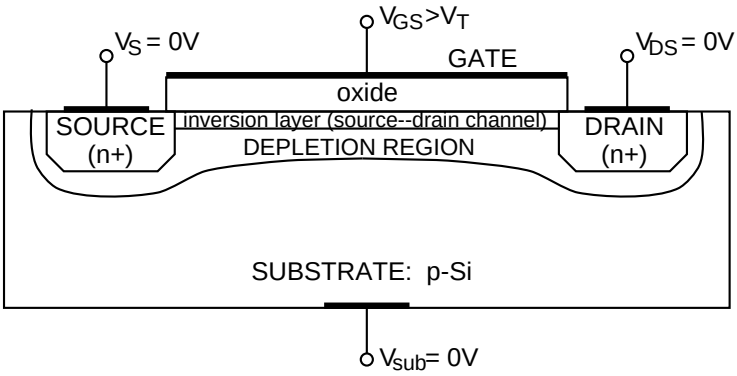

Figure 2.4: Formation of the conducting channel (**inversion layer**) in an enhancement-mode nMOS.

the threshold voltage V_T, then the electric field repels more holes from the channel area leaving an excess of electrons.
The field also pulls out electrons from the source and drain area which, by virtue of being the n+ regions, have excess of electrons.

As a result in the area between source and drain an **inversion layer** is created in which there is an excess of the negative carriers, that is, electrons. In other words a conducting channel has been formed between the source and drain.

Due to the fact that in this case we assume that the drain-source voltage $V_{DS} = 0$, thermal equilibrium exists in the channel region and the drain current, $I_D = 0$.

The **threshold voltage** V_T depends on a specific transistor configuration, that is, on a specific technology of fabrication of MOS transistors and usually is in a range of 0.5V.

Step 3: Linear region (Figure 2.5)

$$V_{GS} > V_T \text{ and } 0 < V_{DS} < V_{sat} \implies I_D > 0$$

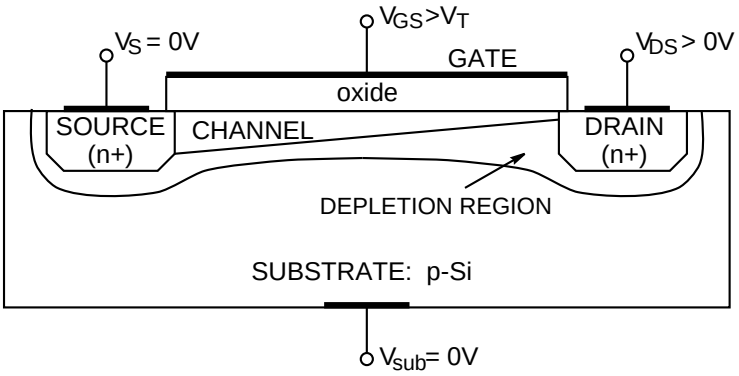

Figure 2.5: An nMOS transistor operating in the linear region

In this case, in the presence of free electrons in the conducting channel, when the drain-source voltage increases above zero, $V_{DS} > 0$, the drain-source current, I_D starts to flow.

When the V_{DS} voltage is relatively small, the transistor operates in the so-called linear region. In this region of operation the drain current I_D is a quadratic function of the source-drain voltage, V_{DS}. Descriptively it means that the increase of the drain current slows down when the source-drain voltage increases.

The channel depth at the drain end decreases with the increase of the source-drain voltage as illustrated in Figure 2.5. Equivalently we can say that the channel region acts as a voltage controlled resistor: the resistance increases when the source-drain voltage increases.

Step 4: pinch-off point (Figure 2.6)

$$V_{GS} > V_T \ , \ V_{DS} = V_{sat} \implies I_D > I_{sat}$$

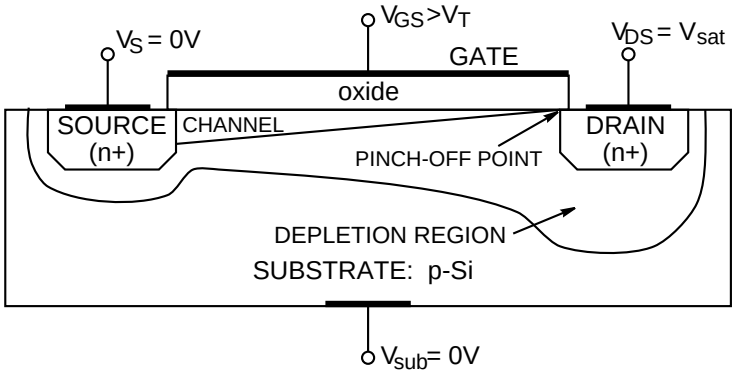

Figure 2.6: The pinch-off point for an nMOS transistor

When the source-drain voltage, V_{DS}, reaches a certain value, V_{sat}, the channel depth at the drain end is reduced to zero. This is called the pinch-off point. In other words, at the pinch-off point, $V_{DS} = V_{sat}$.

From now on, the further increase of the source-drain voltage does not result in an increase of the source-drain current. The transistor now operates in the saturation mode.

Step 5: Saturation mode (Figure 2.7)

$$V_G > V_T \ , \ V_D > V_{Dsat} \implies I_D = I_{Dsat}$$

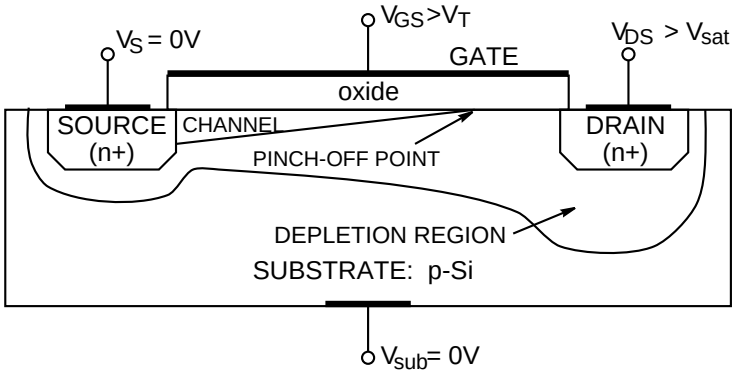

Figure 2.7: The nMOS transistor operating in the saturation mode

In the saturation mode, the depletion region adjacent to the drain is enlarged. Note that in the depletion region there are no free electric carriers and the area acts as a dielectric.

The source-drain current, I_D is now independent of the source-drain voltage, V_{DS}.

Electrons arriving from the source to the channel are injected into the depleted part of the channel and are accelerated towards the drain by the high electric field induced by the source-drain voltage.

Finally, it is important to remember that under no conditions there is a constant current flowing between the gate and other transistor terminals because the gate is insulated by a layer of SiO_2.

2.3 Geometric and material properties of a MOS transistor

2.3.1 Geometric configuration of a MOS transistor

Three components of the MOS transistor structure, namely, the gate, source and drain form a 3-D structure as illustrated in Figure 2.8.

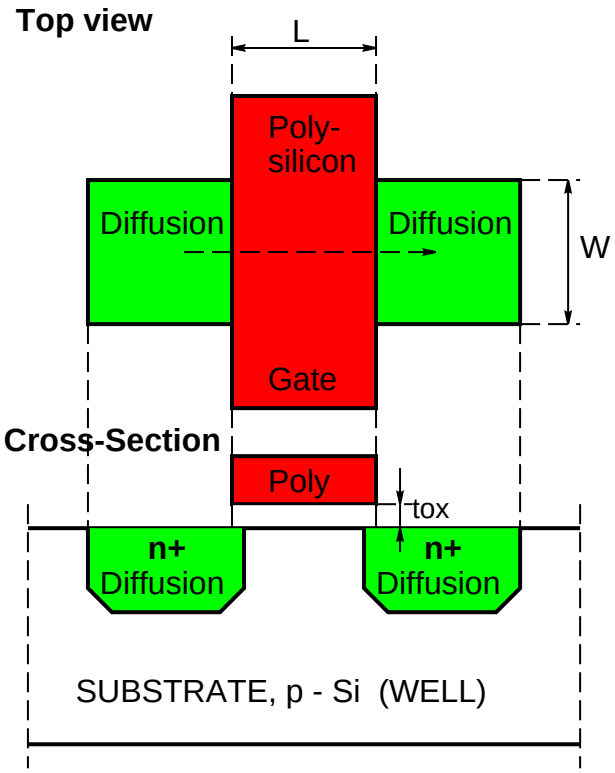

Figure 2.8: Basic geometric parameters of a MOS transistor.

The gate of the MOS transistor is usually made of polysilicon, which is formed from polycrystaline silicon and relatively good conductance.

The gate is insulated by the layer of the silicon dioxide, SiO_2, from a conducting channel existing between two diffusion areas which form the drain and the source of the transistor.

Diffusion areas (source and drain) are created inside a substrate (also known in some technological context as the **well**) of the opposite type, e.g. n+ diffusion inside the p substrate, where 'n+' indicates silicon highly doped with donors.

From the top and cross-sectional views of the MOS transistor presented in Figure 2.8 we note that three basic geometrical parameters of the transistor are the following:

- L and W – the **length** and **width** of the conducting channel between the source and drain,
- t_{ox} — thickness of the oxide layer between the gate and the diffusion/substrate areas.

2.3.2 The gate capacitance

The gate-oxide-channel structure forms a **capacitor**. The gate-oxide capacitance per unit area can be approximately calculated as:

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}} \tag{2.1}$$

where

$$\varepsilon_{ox} = 0.351\mathrm{pF/cm}$$

is the permittivity (a dielectric constant) of SiO_2. Note that the capacitance is inversely proportional to the thickness of the silicon dioxide layer.

Example

Let the oxide thickness be: $t_{ox} = 500\text{Å} = 500 \cdot 10^{-8}\text{cm} = 0.05\mu\text{m}$. Then

$$C_{ox} = \frac{0.351 \cdot 10^{-12}}{0.5 \cdot 10^{-5}} = 0.7 \cdot 10^{-7}\mathrm{F/cm^2} = 70\mathrm{nF/cm^2}$$

Note that the oxide thickness and the resulting gate capacitance per unit area are parameters specified by the **technological process** of fabrication of MOS transistors.

2.3.3 Mobility of carriers

Movement of carriers (electron and holes) can be characterised by their **mobility**. The mobility is a proportionality constant between applied electric field (in V/cm) and resulting velocity of the carriers (in cm/sec). The intrinsic values (for pure silicon) of the mobility for electrons and holes in the room temperature are

$$\mu_n = 1350\text{cm}^2/\text{V}\cdot\text{s}\ , \quad \text{(electrons)} \qquad \mu_p = 480\text{cm}^2/\text{V}\cdot\text{s} \quad \text{(holes)} \tag{2.2}$$

Values of the mobility in doped semiconductor are smaller, but the ratio

$$\frac{\mu_n}{\mu_p} \approx 2.5 \tag{2.3}$$

is preserved. The fact that holes are more sluggish than electron has some influence on relative sizes of nMOS and pMOS transistors.

2.3.4 Basic configurations of MOS transistors

The four terminals of MOS transistors, namely, the drain, source, gate and substrate are usually connected to the ground (GND) and supply voltages in the way as indicated in Figure 2.9.

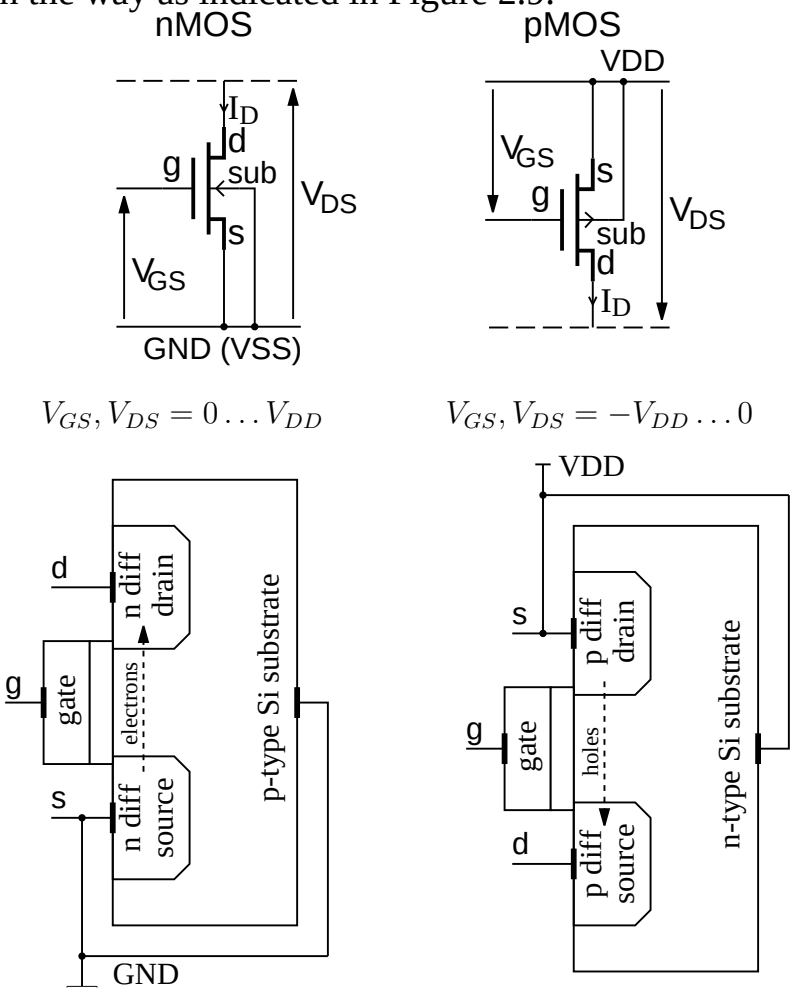

Figure 2.9: Basic configurations of MOS transistors

The nMOS transistor has its source and the p-type substrate connected to the ground terminal GND (V_{SS}) = 0V, whereas the pMOS transistor has its source and the n-type substrate connected to V_{DD} = 2–5V.

The V_{GS} and V_{DS} voltages are positive for the nMOS transistor and negative for the pMOS.

2.4 DC analysis of the MOS transistors

2.4.1 Transistor parameters

In this section we will discuss the relationship between constant (DC) voltages at the transistor terminals and the resulting **drain current** I_D. Apart from the voltages, the I_D current is also a function of

- the process parameters: V_T — the **threshold voltages** (V_{Tn} or V_{Tp}), and a **process transconductance**, k_c defined as follows

$$k_c = \mu_c \cdot C_{ox} \tag{2.4}$$

  where μ_c is an effective mobility of the carriers (μ_e or μ_p), and C_{ox} is the gate capacitance per unit area,

- the **width**, W, and **length**, L, of the channel between the source and the drain (see Figure 2.8).

The parameter which links the process transconductance, k_c, with the transistor dimension is called the (non-linear) **transistor transconductance** parameter, g_c, and is defined in the following way:

$$g_c = \frac{\mu_c \cdot C_{ox}}{2} \cdot \frac{W}{L} \tag{2.5}$$

It is also convenient to use a **gate voltage relative to the threshold voltage** defined as follows:

$$V_\Delta = V_{GS} - V_T \tag{2.6}$$

Note that all the above parameters, namely, k_c, g_c, V_Δ can be referred to a specific type on MOS transistor as, k_n, k_p, g_n, g_p, and $V_{\Delta_n}, V_{\Delta_p}$, respectively.

2.4.2 Current-Voltage relationships

With the above parameters, the relationships between the DC I_D current and relevant voltages can be summarised as in Table 2.1.

nMOS	**pMOS**
Cut-off region:	
$V_{GS} < V_{Tn}$	$V_{GS} > V_{Tp}$
$I_D = 0$	
Linear region:	
$V_{GS} \geq V_{Tn}, V_{DS} < V_{\Delta_n}$	$V_{GS} \leq V_{Tp}, V_{DS} > V_{\Delta_p}$
$I_D = g_c \cdot (2V_\Delta - V_{DS})\, V_{DS}$	
Saturation region:	
$V_{GS} \geq V_{Tn}, V_{DS} \geq V_{\Delta_n}$	$V_{GS} \leq V_{Tp}, V_{DS} \leq V_{\Delta_p}$
$I_D = g_c \cdot V_\Delta^2$	

Table 2.1: Fundamental DC relationships for MOS transistors

Note that the drain current I_D is proportional to the ratio W/L of the transistor channel size. For given process parameters and voltages,

- the wider the transistor channel, W, the larger I_D current, and
- the longer the transistor channel, L, the smaller I_D current.

Note also that the **saturation** occurs when

$$V_{DS} = V_{GS} - V_T = V_\Delta$$

At the saturation point the current expressions for the linear and saturation regions are identical what can be seen from the following derivation:

$$I_{sat} = g_c \cdot (2V_\Delta V_\Delta - V_\Delta^2) = g_c \cdot V_\Delta^2 \tag{2.7}$$

Example

Consider an nMOS transistor with the following parameters:
$\mu_n = 600$ cm^2/V·s ,
$C_{ox} = 7 \cdot 10^{-8}$ F/cm^2,
$V_{Tn} = 1$V,
$W = 20\mu$m, $L = 2\mu$m.
The transconductance parameter can be calculated as (watch out to use the consistent units):

$$g_c = 0.5 \cdot 600 \cdot 7 \cdot 10^{-8} \cdot \frac{20}{2} = 210 \cdot 10^{-6}\mathrm{A/V}^2 = 0.21\mathrm{mA/V}^2$$

2.4.3 Current-voltage characteristics of an nMOS transistor

Now we can use MATLAB (or similar package)to plot the current-voltage characteristics

$$I_D = I_D(V_{GS}, V_{DS})$$

as presenter in Table 2.1. The drain current is a function of two voltages, namely, the gate-source and drain-source voltages, therefore, could be represented as a surface in a three-dimensional space, V_{DS}, V_{GS}, I_D. Traditionally, however, this surface is represented by the family of curves as in Figure 2.10.

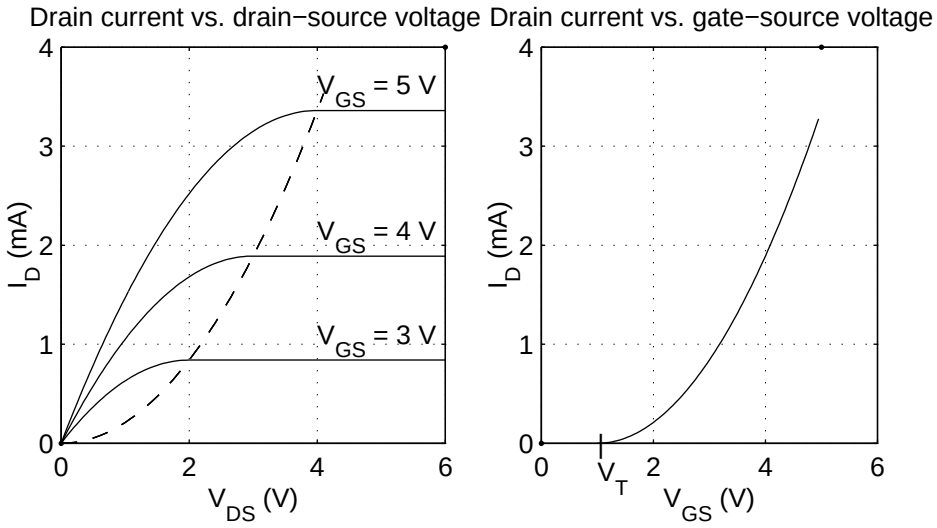

Figure 2.10: An example of MATLAB generated current-voltage characteristics of a MOS transistor. Left plot: the drain current I_D versus the drain-source voltage, V_{DS} for various values of the gate-source voltage, V_{GS}. Right plot: the drain current I_D versus the gain-source voltage, V_{GS}, in saturation.

The current-voltage characteristics were generated for the value of the transistor non-linear transconductance $g_c = 0.21\mathrm{mA/V}^2$ and the threshold voltage $V_T = 1\mathrm{V}$.
From the plots in Figure 2.10 and Table 2.1 you can identify: the threshold voltage, the linear region, the saturation region, the saturation voltage.

2.4.4 Switching model of MOS transistors

In digital circuits MOS transistors work in such a way that they switch between the off-state and saturation.

Therefore their switching DC model can be approximated by a **controlled switch** and a controlled current source (sink) as presented in Figure 2.11.

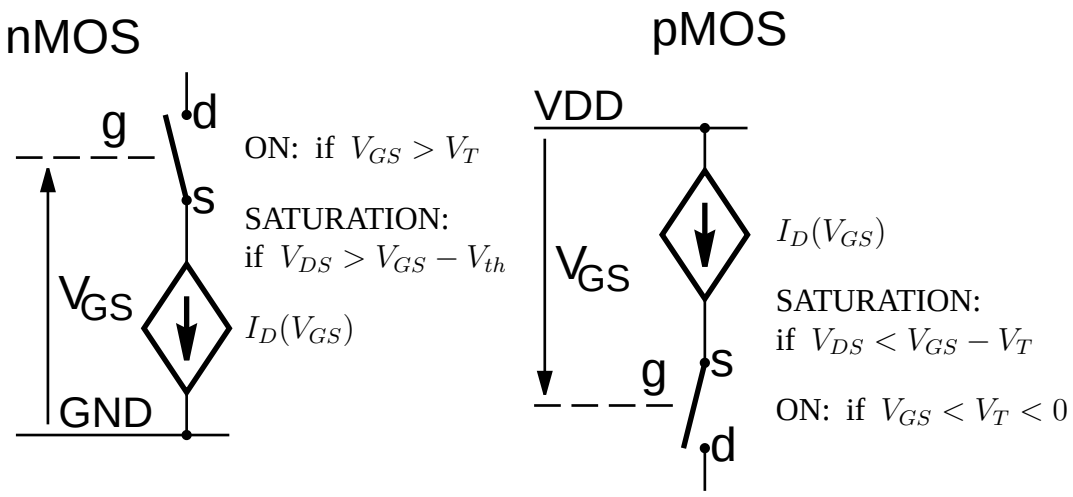

Figure 2.11: A simplified switching model of MOS transistors

In other words, in digital circuits, a MOS transistor is either in

- off-state, $V_{GS} \approx 0$V: an open switch in the off-state, or in
- saturation, $V_{GS} \approx V_{DD}$: a current source generating current $I_D = g_c \cdot V_\Delta^2$

Note that the gate is always electrically insulated from the source and the drain.

2.5 MOS transistors — topology and geometry of the circuit layout

From the designer view point we will be operating with three representations of MOS circuitry as presented in Figure 2.12. These three representations are:

- **Schematic diagrams**.
- **Stick diagrams** representing **topology** of the integrated circuit.
- **Circuit layouts** representing the exact **geometry** of the integrated circuit. The layouts are generalisation of the top view of a MOS transistor as in Figure 2.8 with additional connections required to build a complete circuit.

  The circuit geometry must have its dimensions specified precisely in micrometers (μm) or in relative units called λ–units.
  Ultimately, from the circuit layouts we extract **photolithographic masks** used in fabrication of integrated circuits.

Comparing three circuit representations from Figure 2.12 we note that:

- **Transistors** are represented by four-terminal symbols in the schematics. In the stick diagram and the circuit layout transistors are identified by crossing of the **red** path representing the gate, over the **green** (nMOS) or **brown**(pMOS) path representing the relevant diffusion. The drain and source terminals exist on both sides of the gate.
- Note that the sources of the transistors are connected to to either VDD (pMOS), or GND (nMOS). These two power rails are made of metal and are represented by **blue** paths in the stick diagrams and circuit layouts. A special **contact** must be made to connect diffusion to metal.

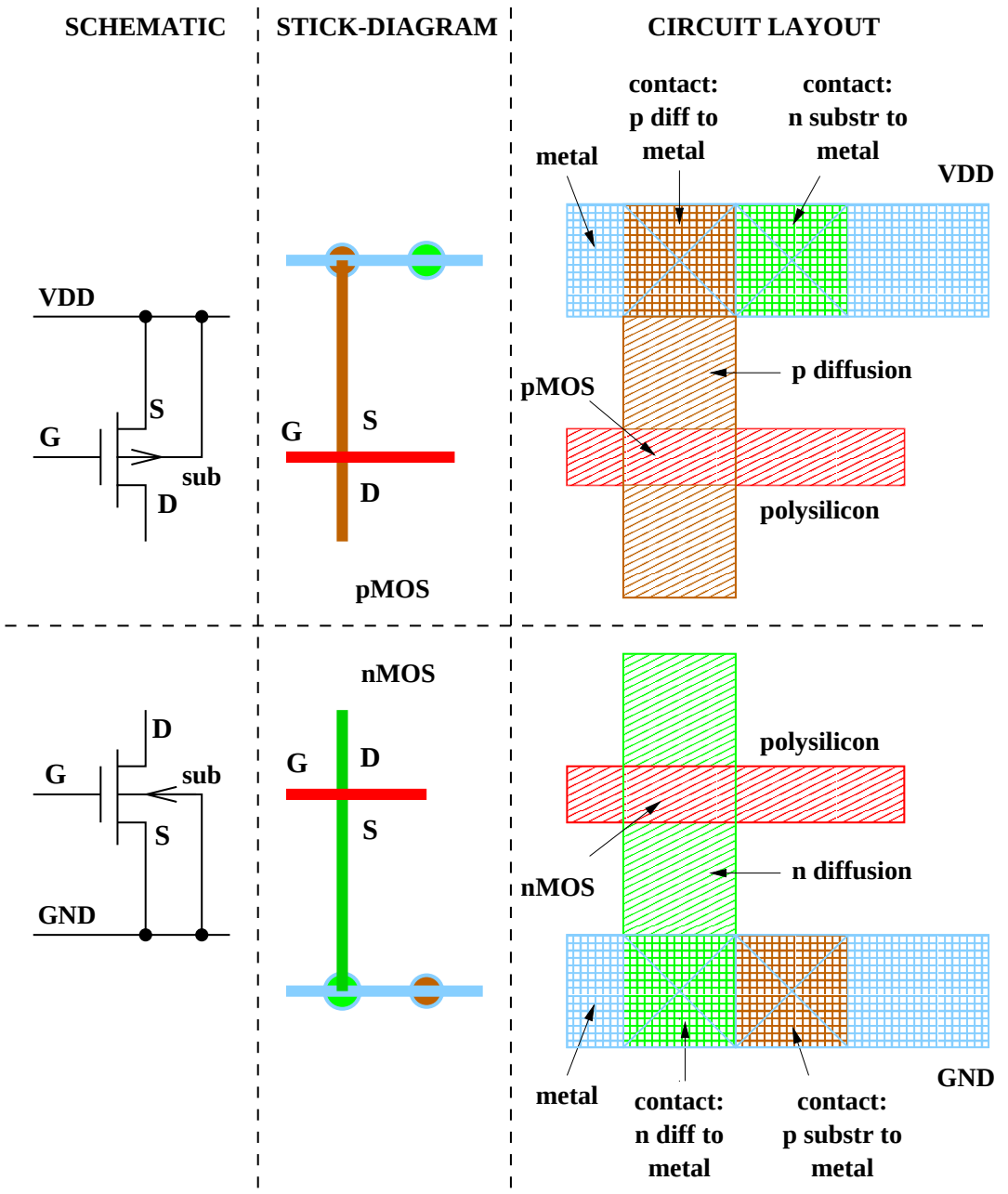

Figure 2.12: Three representations of MOS circutry: schematics, stick diagrams and circuit layouts.

- Finally, the forth terminal, namely the substrate must be also connected to the appropriate power rail, that is, to either VDD (pMOS, n substrate), or GND (nMOS, p substrate). Remember that pMOS transistors are created in the n-type substrate and nMOS transistors in the p-type substrate. We have to remember about the substrate contacts even if we use simplified three-terminal symbols of MOS transistors as in (Figure 2.13).

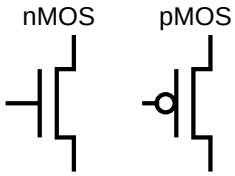

Figure 2.13: Three-terminal symbols of MOS transistors.

- In the example in Figure 2.12, transistors have the size:

$$W = 4\lambda\ , \quad L = 2\lambda$$

- The contacts usually have dimension $4\lambda \times 4\lambda$ and they occupied a significant portion of the circuit layout.