Digital Systems

2. Fundamentals of Digital Circuits
   - 2.1 Boolean Functions
   - 2.2 Representation of Boolean Functions
   - 2.3 Functions with 1 input variable
   - 2.4 Functions with 2 input variables
   - 2.5 Semiconductor Diodes
   - 2.6 The MOS Transistor
   - 2.7 Construction of simple gates using transistors

2. Fundamentals of Digital Circuits

Literature:

Waldschmidt, K.: Schaltungen der Datenverarbeitung, Teubner, 1980, ISBN 3-519-06108-2

Bauer, L, Goos, G.: Informatik, Springer Verlag, 1971, ISBN 0-38705303-4

Kohavi, Z.: Switching and Finite Automata Theory, McGraw-Hill, New York, 1978

Definition:

Let B={0,1} and n and m be natural numbers. A function

f: $B^n \rightarrow B^m$ is known as Boolean Function

(Named after the Mathematician George Boole, 1815-1864)

Definition:

The Truth Table of a Boolean function describes the values for all possible combinations of the input variables. It gives a complete and unique description of the function.

Example of a Truth Table:

Input Variables			Output Variables	
x_2	x_1	x_0	y_1	y_0
0	0	0	0	0
0	0	1	0	1
0	1	0	0	1
0	1	1	1	0
1	0	0	0	1
1	0	1	1	0
1	1	0	1	0
1	1	1	1	1

Definition:

The functional description of a Boolean function is an abstract representation of the input and output relationships of a circuit realising this function.

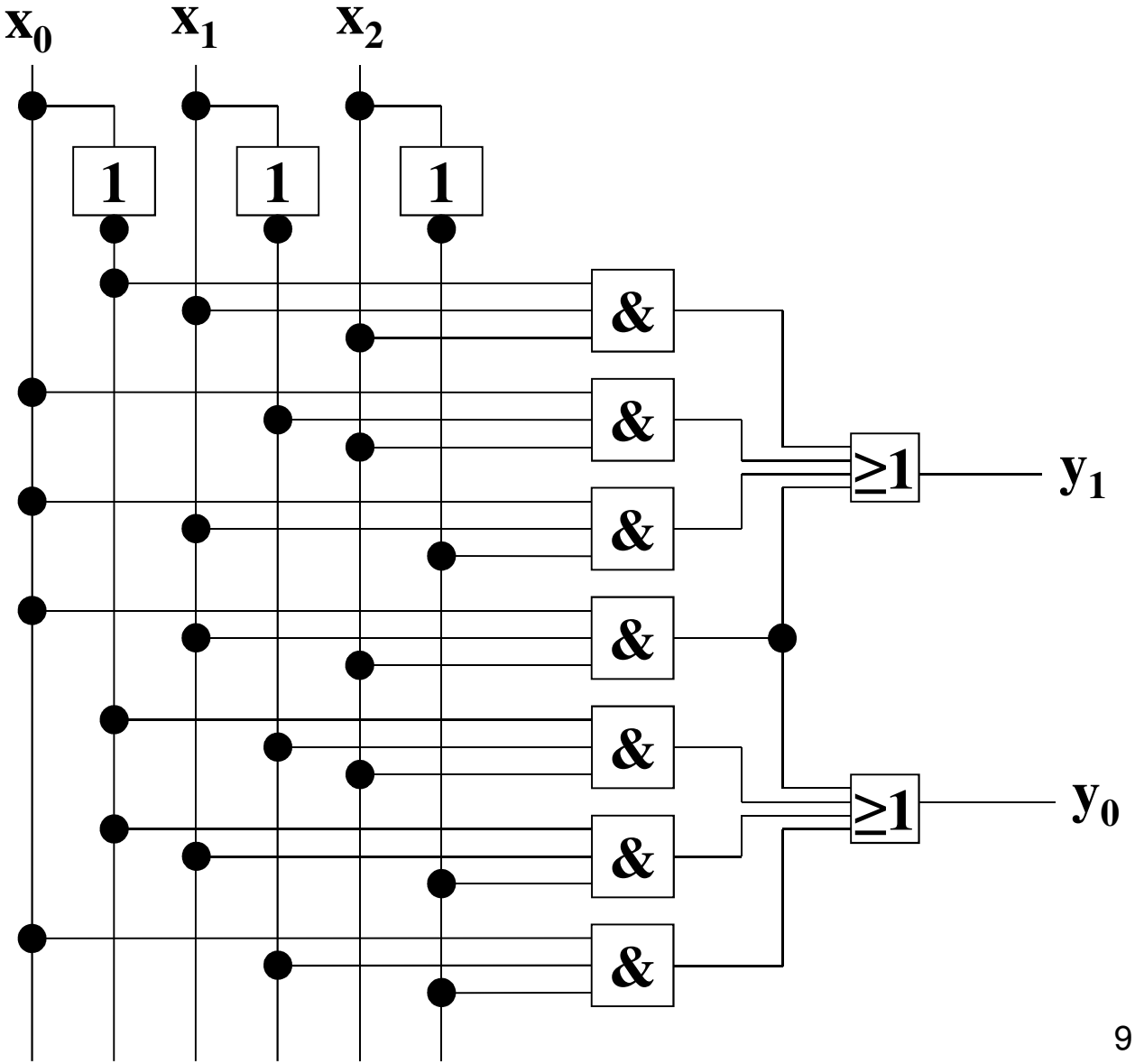
Example of a Functional Description:

Full Adder
Inputs: x2, x1, x0
Outputs: y1, y0
Characteristics: If an odd number of inputs is 1, then y0 is 1. Else y0 is 0.
If two or more inputs have the value 1, then y1 has the value 1; else y1 = 0.

Definition:

The description of a boolean function as a circuit diagram is a representation of the function, which is targeted to be a technical realization using a given circuit technology. It is complete but not however unique.

Example of a Circuit Diagram for a Function:



Definition:

The description of a boolean function as amount of function equations is a complete description of the function. For each output variable, there has to be a function equation. On the right side of the equation, there are only input variables.

Example of a Description with Function Equations:

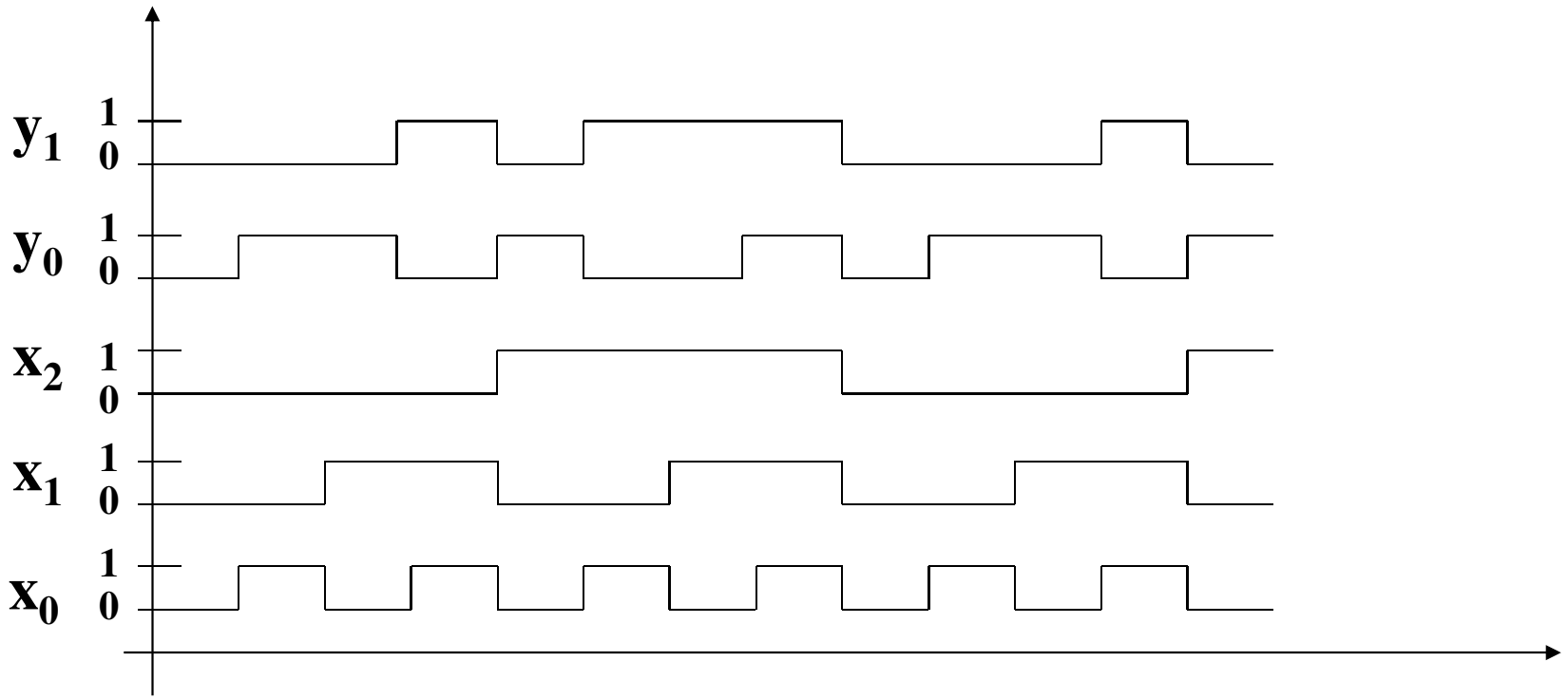
$$y_0 = \bar{x}_0 \bar{x}_1 x_2 + \bar{x}_0 x_1 \bar{x}_2 + x_0 \bar{x}_1 \bar{x}_2 + x_0 x_1 x_2$$

$$y_1 = \bar{x}_0 x_1 x_2 + x_0 \bar{x}_1 x_2 + x_0 x_1 \bar{x}_2 + x_0 x_1 x_2$$

Definition:

With an impulse diagram, the characteristics of a boolean function in the form of modification of virtual physical variable (symbolizing the value of the output variable) is represented in terms of dependency of the values of the input variables.

Example of a Description as Impulse Diagram:

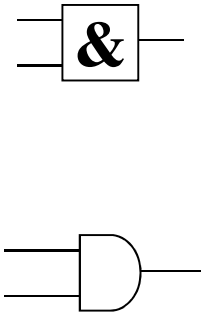


Boolean Function without Input Variables:

	x=0	x=1	Algebraic Representation
Null Function	**0**	**0**	**y=0**
Identity	**0**	**1**	**y=x**
Negation	**1**	**0**	**$y=\bar{x}$**
Ones Function	**1**	**1**	**y=1**

Boolean Function with two Variables:
1. AND Function

AND

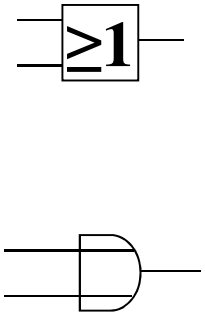


x_0	x_1	$x_0 \wedge x_1$
0	0	0
0	1	0
1	0	0
1	1	1

$$x_0 \wedge x_1 = x_0 \cdot x_1 = x_0 x_1$$

Boolean Function with Two Variables:
2. OR Function

OR

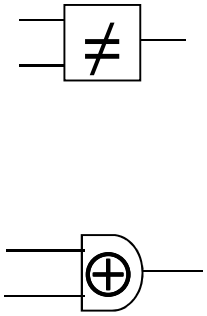


x_0	x_1	$x_0 \vee x_1$
0	0	0
0	1	1
1	0	1
1	1	1

$$x_0 \vee x_1 = x_0 + x_1$$

Boolean Functionen with Two Variables:
3. Exclusive OR

XOR

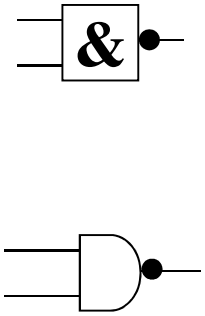


x_0	x_1	$x_0 \oplus x_1$
0	0	0
0	1	1
1	0	1
1	1	0

$$x_0 \oplus x_1$$

Boolean Function with Two Variables:
4. NAND Function, Sheffer Function

NAND

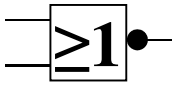


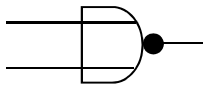
x_0	x_1	$\overline{x_0 \wedge x_1}$
0	0	1
0	1	1
1	0	1
1	1	0

$$\overline{x_0 \wedge x_1} = \overline{x_0 \cdot x_1} = \overline{x_0 x_1}$$

Boolean Function with Two Variables:
5. NOR Function, Pierce Function

NOR



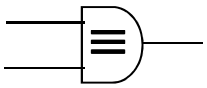


x_0	x_1	$\overline{x_0 \vee x_1}$
0	0	1
0	1	0
1	0	0
1	1	0

$$\overline{x_0 \vee x_1} = \overline{x_0 + x_1}$$

Boolean Function with Two Variables:
6. Equivalent Function

EQV

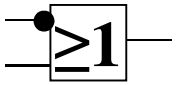


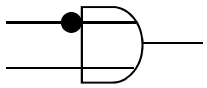
x_0	x_1	$x_0 \equiv x_1$
0	0	1
0	1	0
1	0	0
1	1	1

$$x_0 \equiv x_1 = x_0 \Leftrightarrow x_1$$

Boolean Function with Two Variables:
7. Implication

IMP



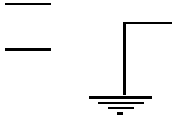


x_0	x_1	$x_0 \Rightarrow x_1$
0	0	1
0	1	1
1	0	0
1	1	1

$$x_0 \Rightarrow x_1 = \overline{x_0} + x_1$$

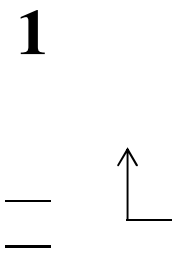
Boolean Function with Two Variables:
8. Null Function

0



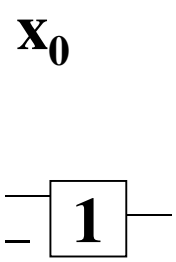
x_0	x_1	Null
0	0	0
0	1	0
1	0	0
1	1	0

Boolean Function with Two Variables:
9. One Function



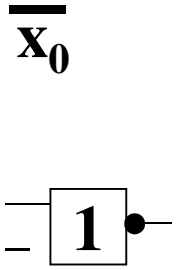
x_0	x_1	Ones
0	0	1
0	1	1
1	0	1
1	1	1

Boolean Function with Two Variables:
10. Identity for x_0



x_0	x_1	x_0
0	0	0
0	1	0
1	0	1
1	1	1

Boolean Function with Two Variables:
11. Negation for x_0

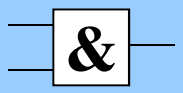
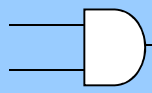
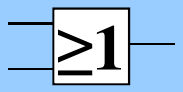
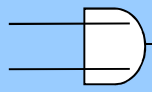
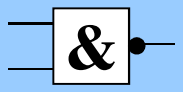
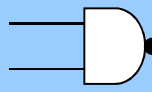
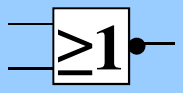
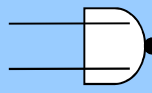
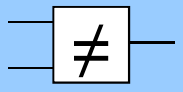
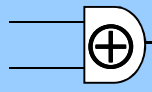
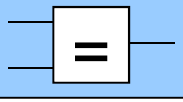
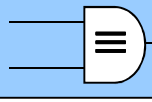
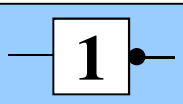
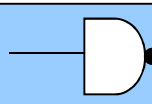


x_0	x_1	$\overline{x_0}$
0	0	1
0	1	1
1	0	0
1	1	0

Boolean Function with Two Variables:

12 – 16 more functions, e.g. Identity for x_1,

The Important Functions with Two Variables:

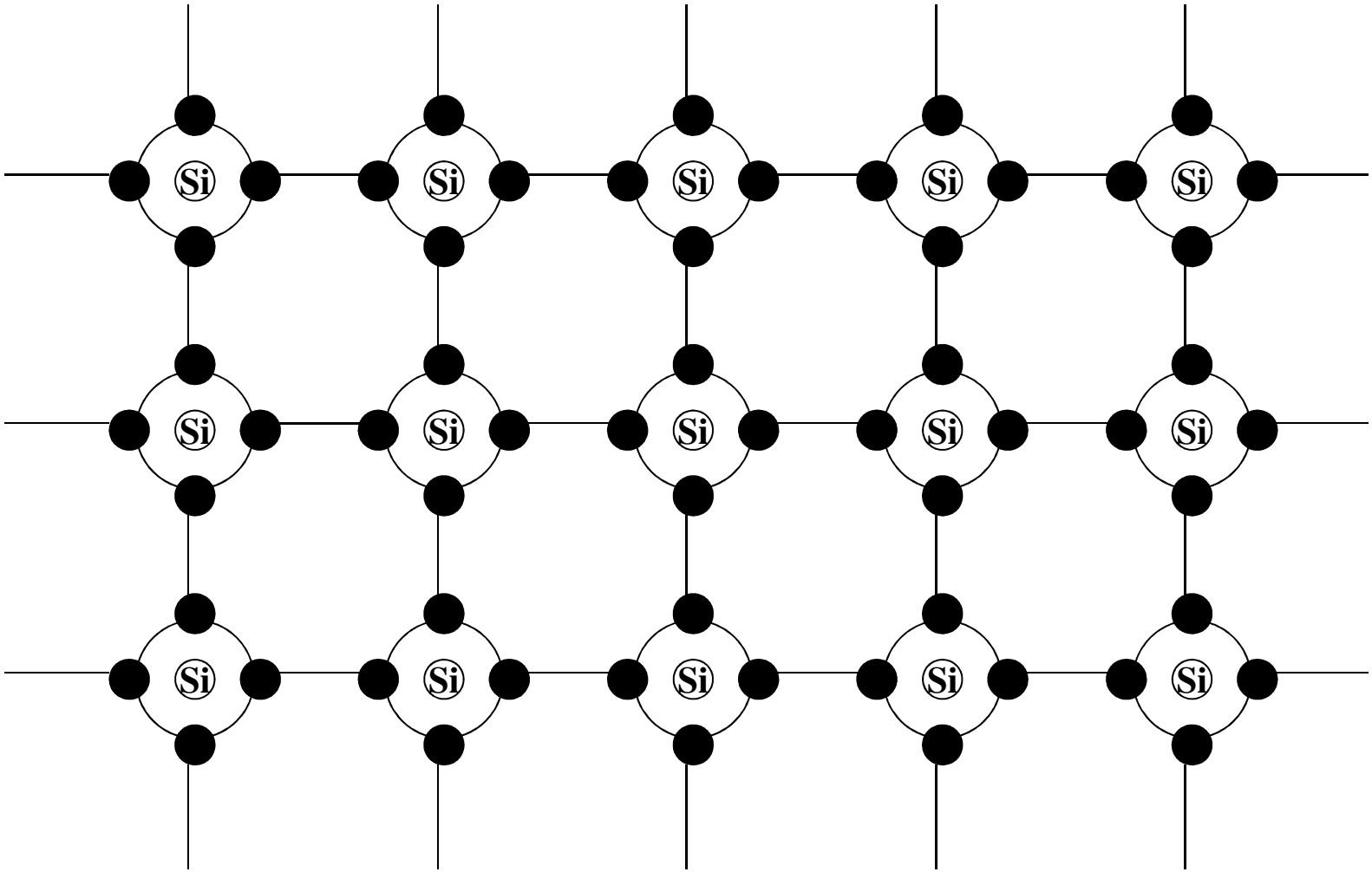
	Formel	Symbol	Altes Symbol
UND (AND)	$x_0 \wedge x_1$		
ODER (OR)	$x_0 \vee x_1$		
NAND	$\overline{x_0 \wedge x_1}$		
NOR	$\overline{x_0 \vee x_1}$		
XOR	$x_0 \oplus x_1$		
EQV	$x_0 \equiv x_1$		
NOT x_0	$\overline{x_0}$		

2.5 Semiconductor Diodes

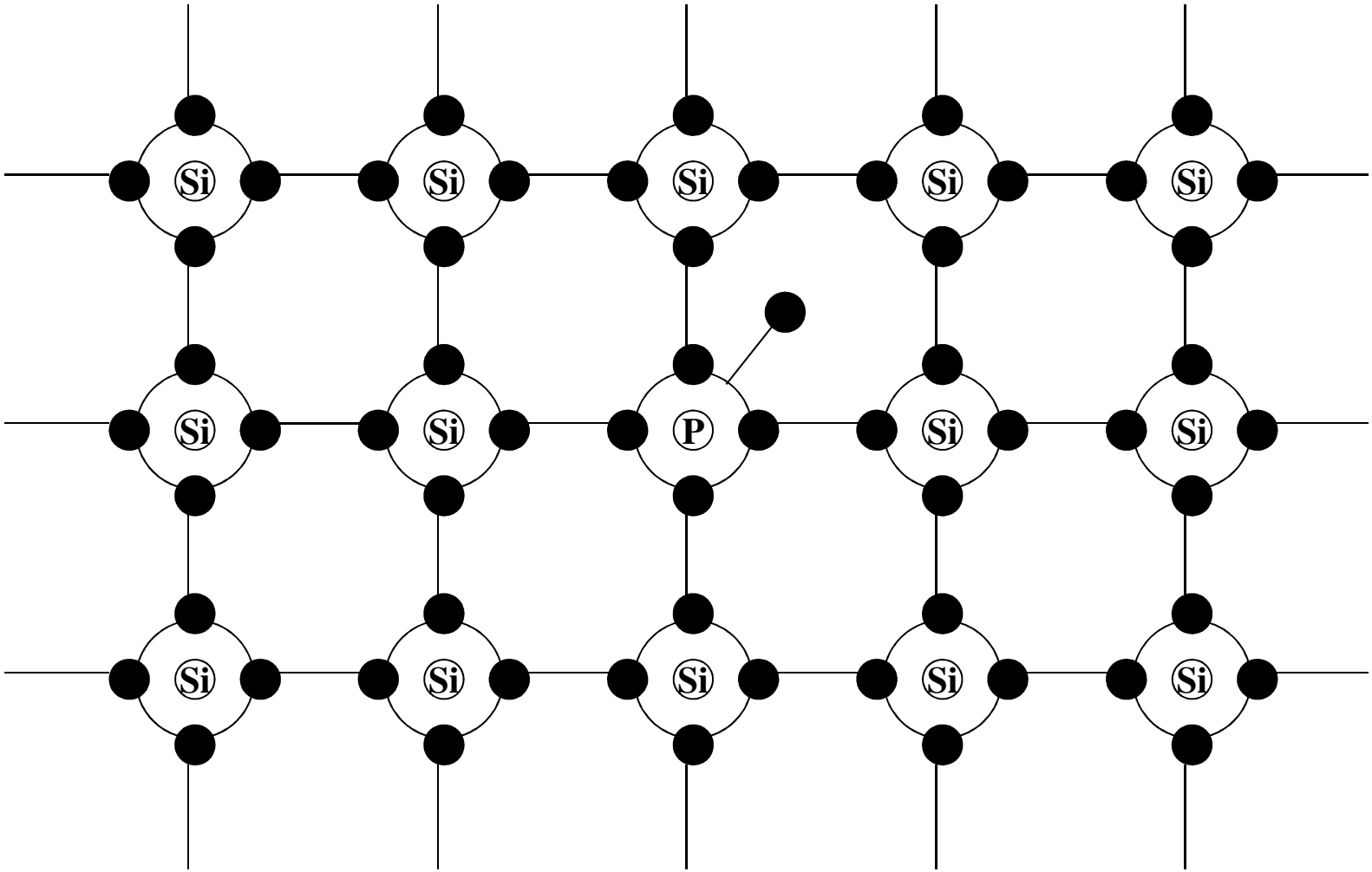
Semiconductor diodes are components, which made use of the conductivity properties of the **pn-junction**. They are most of the time manufactured using silicon. A pn-junction is the junction of positively doped (p-) silicon to the negatively doped (n-) silicon. Doping is the selective introduction of foreign atoms into the crystal structure of the silicon. Silicon is quadrivalent, and is mapped in a purest form of crystal structure, whereby every two electrons of neighbouring atoms form a connection with each other. One can imagine that like an uniform rectangular grid, as shown in the next slide. It comprises no free electrons and is therefore, also hardly conductive.

If one inserts into the grid now, an atom that has five valence foreign atoms, e.g. Phosphorous, there is actually one free electron in the entire grid that can form a connection.

Crystal Structure of Pure Silicon



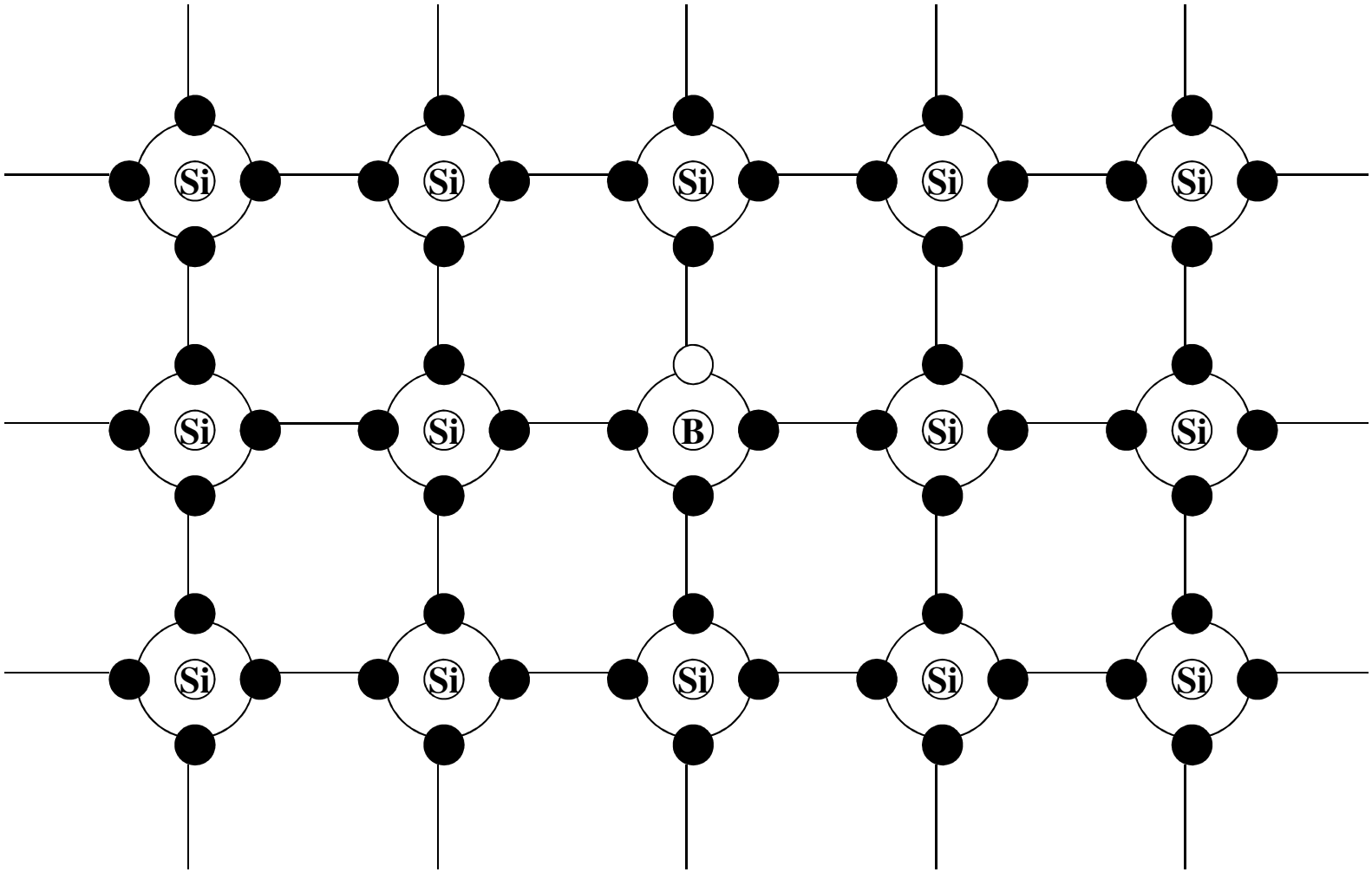
Crystal Structure of n-doped Silicon



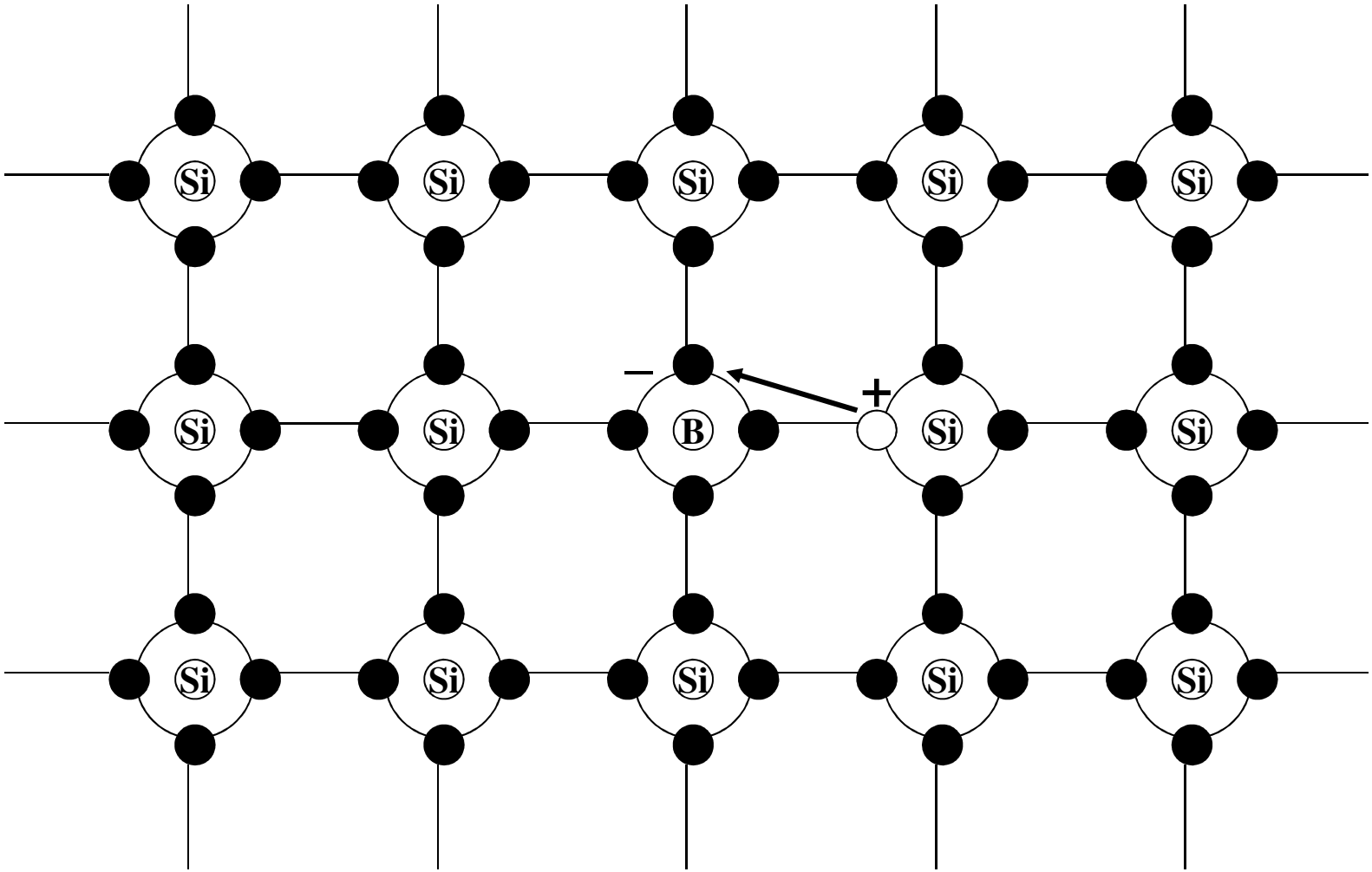
This free electron can now be used to conduct electric currrent. Through the doping of silicon with phosphorous, one can make the formerly bad conductor now to be conductive. This is the reason why silicon is known as **semiconductor**. The doping with Phosphorous generates free electrons in silicon. Since these are negatively charged, one refers to that as **n-doping**, occassionally also as a **n-semiconductor**.

One uses a three valence element, e.g. Boric, instead of Phosphorous, in order to obtain a **p-doping**. A corresponding crystal grid is shown in the following slide. There is however a qualitative difference. In p-doping material, electrons are missing in the crystal grid. The missing electron of an atom can be replaced by an electron from the neighboring atoms, when it is unhinged from its paired bonding. It creates a **positive charge**, a **hole** or **defect electron**. The foreign atom (Boric), which has now received an electron, becomes negatively charged, and stays however, fixed.

Crystal Structure of p-doped Silicon

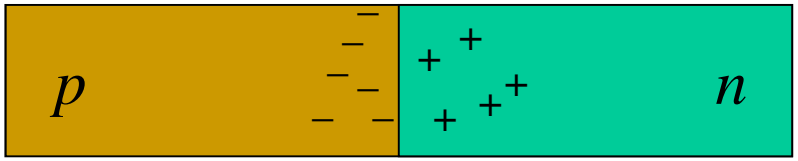


Movement of Positive Charged in p-doped Silicon

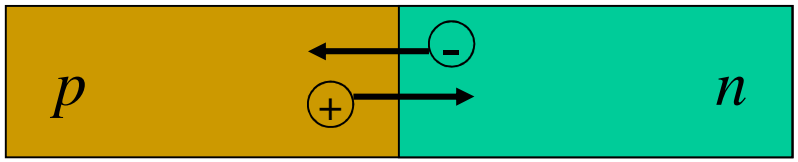


By the process of the replacement of a missing electron by a neigbouring electron, **the hole travels** now in the crystal grid. It can therefore, likewise be used to transport electrical current. However, the mobility of the holes in p-doped semiconductor is not as great as the mobility of the free electron in n-doped semiconductor. This is because the electron in p-doped material must first unhinged from its existing bonding. As a rule of a thumb, one can note that n-doped silicon is about three times more conductive than a p-doped.
If one leaves now a p-doped directly side by side to a n-doped, it generates a **pn-junction**. Because of the free electrons in the n-zone and the (freely mobile) holes in the p-zone, it creates a special reaction at the border: Free electrons diffuse in the p-zone and the holes in the n-zone and this is where they recombine. Therefore, it reduces the number of free charged carriers at the boundary layer. The free charged carriers at the boundary layer becomes a high impedance so-called depletion layer.

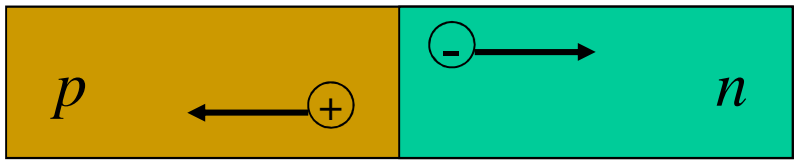
Through the diffusion of the electrons in the depletion layer, the fixed and positive ions (**so-called space charges**) remain, and through the recombination of the holes with the electrons, it creates fixed negative ions in the p-zone.



Between the positive space charge and the negative space charge, an electrical field is created. In the space charged zone, the free charged carriers faciliate the diffusion and they also facilitate the electrical field force in the opposite direction.

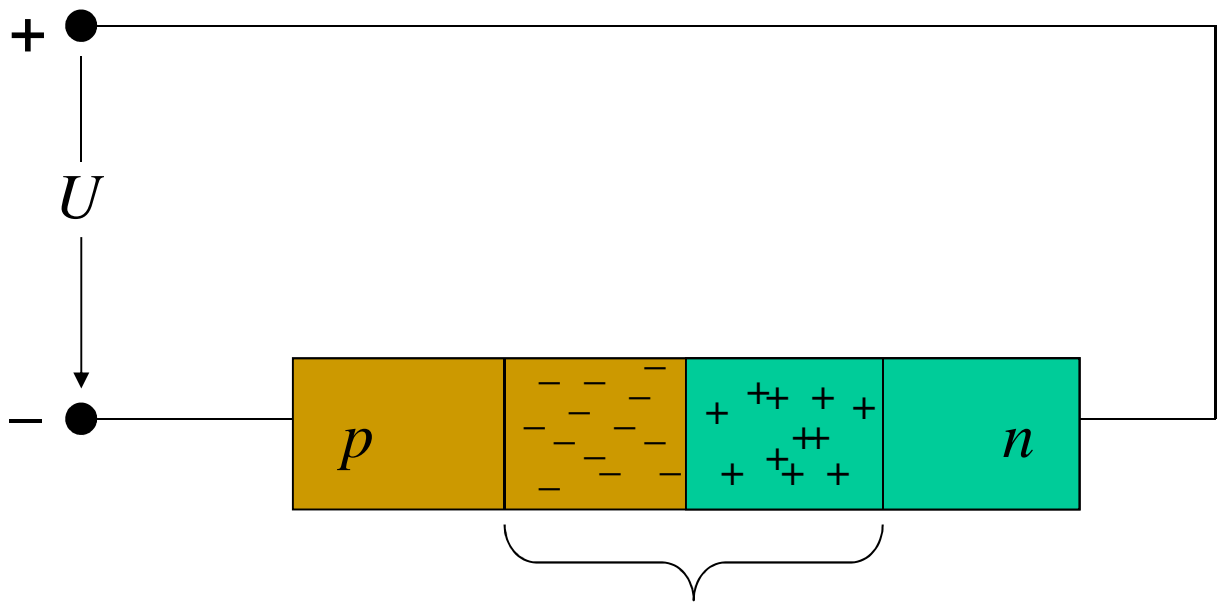


Diffusion Effect

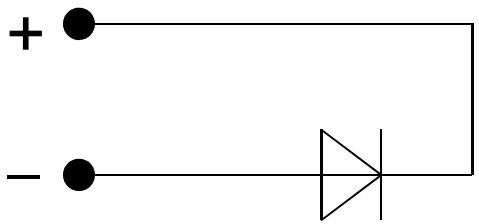


Feld Effect

It sets up, by itself, a dynamic equilibrium at the pn-junction, if the field effect and the diffusion effect are equally large. Then, between the positive space charge in the n-zone and the negative space charge in the p-zone, there exists a fixed voltage, which is the diffusion voltage, $U_{D.}$ This is approximately 0.75V in silicon. If one applies on the diode a DC voltage, it will – depending on the polarity – **conduct or unconduct**. If the minus polarity of the voltage supply source is applied at the p-zone and the plus polarity at the n-zone of the diode, the voltage in the space charge zone rises to U_D+$U.$ The field strength becomes larger and the zone of the free charged carriers becomes wider. It becomes a high impedance depletion zone that one says that the diode is polarized in the reverse bias.

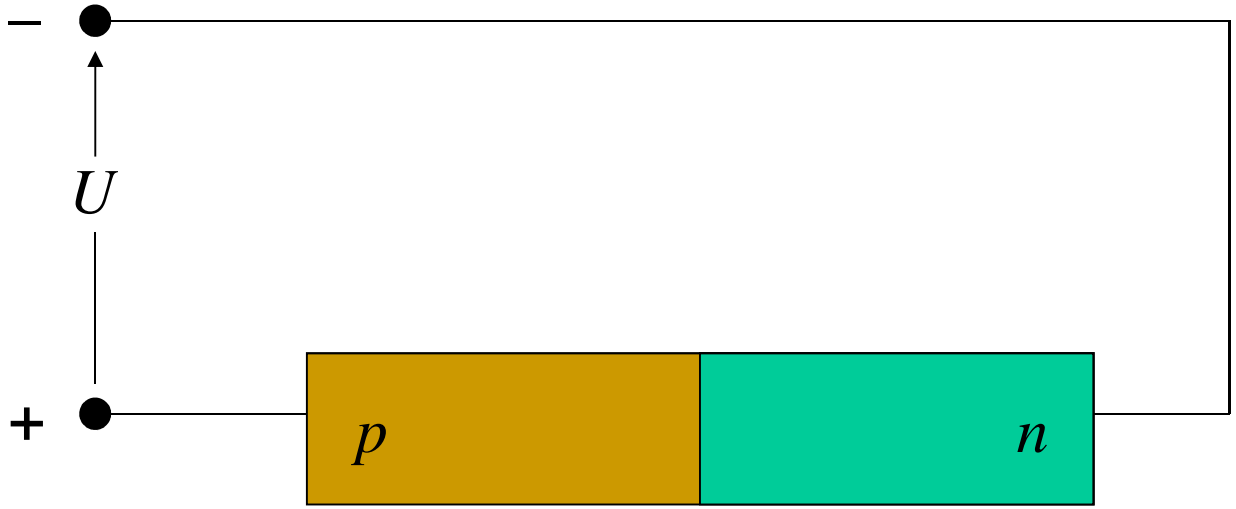


Space Charge Zone without
Free Electrons or Holes

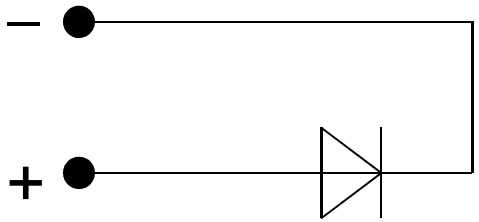


Vice versa, if the minus polarity of the voltage supply is applied at the n-zone and the plus polarity at the p-zone of the diode, the voltage sinks in the space charge zone to U_D-U. The zone of the free charged carriers becomes narrower and disappear completely, when $U>U_D$. (The value of U is also known as the **threshold voltage**). Therefore, the diode conducts because there are now sufficient free charge carriers for the entire way to and fro.

Through the combination of two pn-junctions, one can build bipolar transistors. Since these have no longer great importance in today's digital circuit techniques, they will not be handled here. Instead, we will concentrate on switching elements, that play the most significant role in the building of digital circuits today and also in the future. These are the MOS transistors.



No Space Charge Zone, Free Electrons and Holes Exist



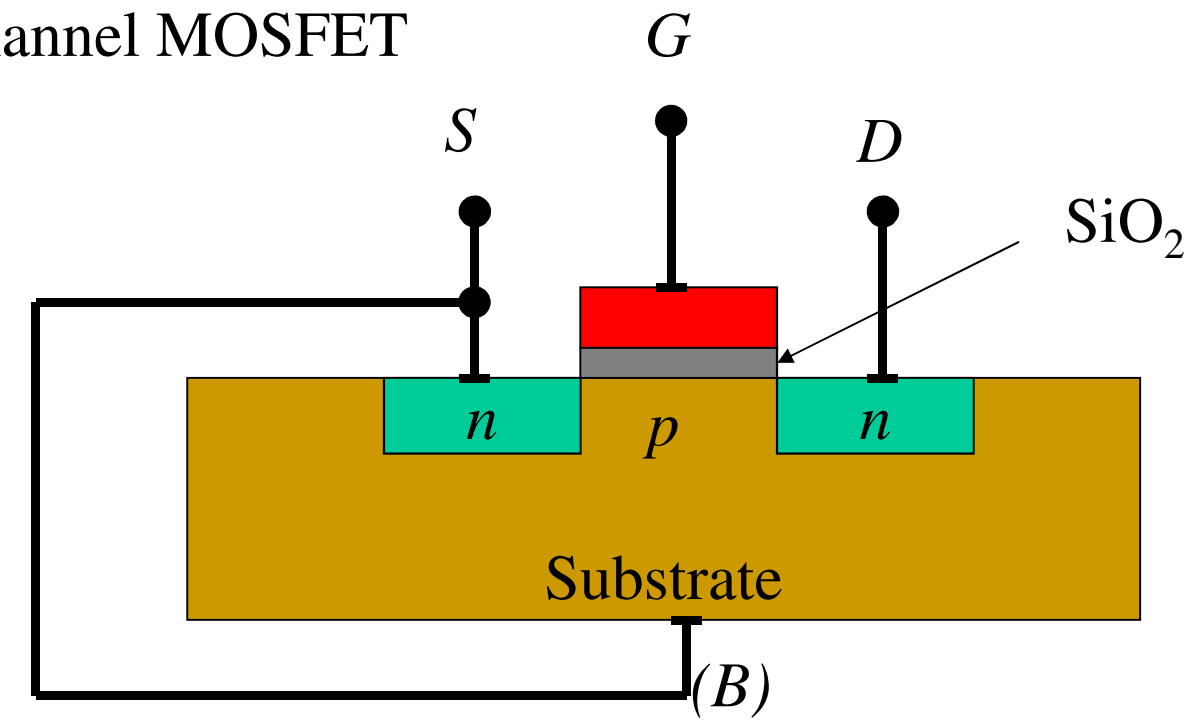
2.6 MOS-Transistors

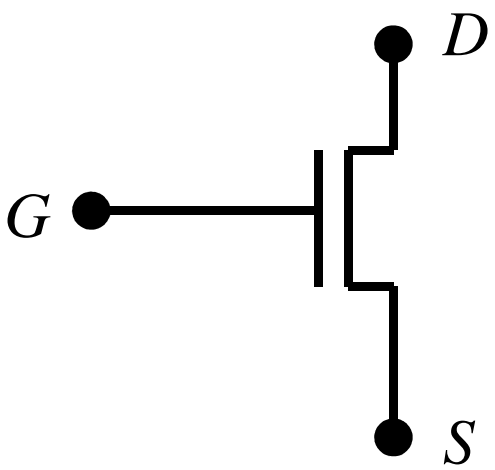
The abbreviation **MOSFET** stands for **Metal-Oxide-Semiconductor Field Effect Transistor**. Metal-Oxide-Semiconductor is built on basically as what is described in the layer sequence (today, one use poly-crystalline silicon instead of the normally used, metal). Field Effect Transistor means: The transistor effect will be determined by the generation of an electrical field through the application of voltage on the control electrode.

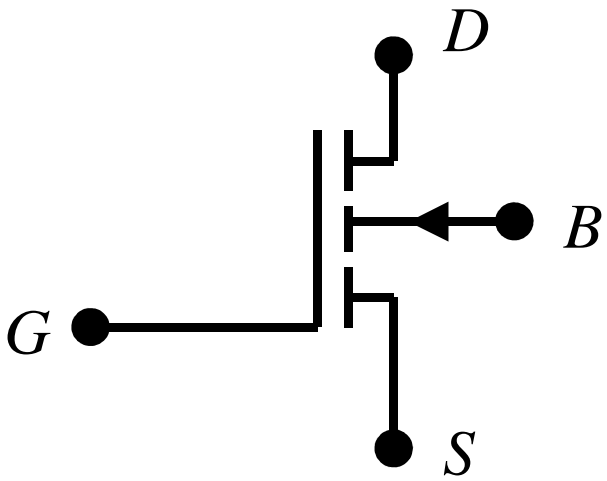
The three contacts of a FETs are known as **D (Drain), S (Source)** and **G (Gate)**. The gate is the **control electrode**, where one applies the voltage and thereby switching the connection between drain and source.

One differentiate MOSFETS by the doping characteristics of the semiconductor material, in which (using suitable interconnect) generates the conductive channel between drain and source. We begin with the **self conducting n-channel MOSFET**, **(enhencement mode n-channel MOSFET)** whose construction is seen in the following slide.

n-Channel MOSFET



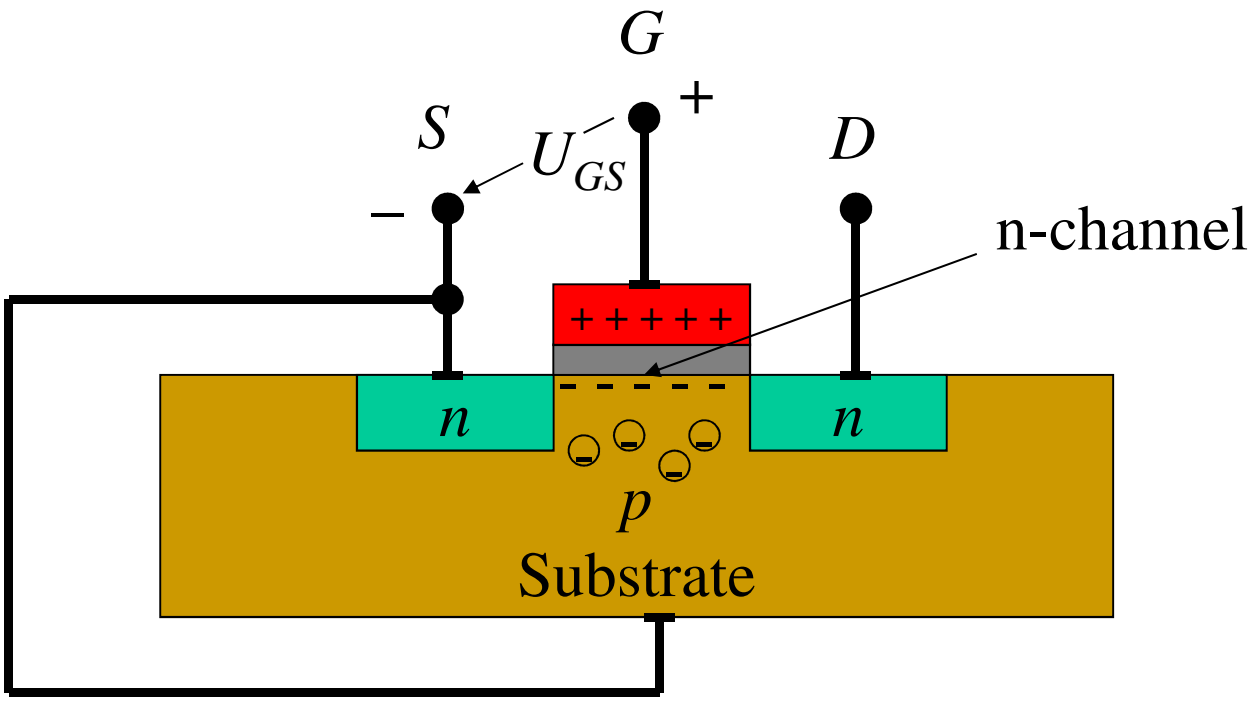




In the p-semiconductor, there is the referenced **Substrat,** whereby two highly doped n-zones are diffused as source and drain. They are connected respectively with the source and the drain contacts. On this substrate, an isolating layer of silicon is built between these two zones. It has a gate, which is isolated from the counterparts source, drain and substrate. Since the oxide film is very thin, the gate with the substrate forms a capacitor. The zone sequence source-substrate-drain is a npn-format. Due to the too large spacing between both of the n-zones, no bipolar transistor is formed.

Now, positive voltage is applied on the gate, opposite of the substrate, so that the holes, as moving charged carriers, are moved away from the gate and into the substrate. A negative space charge zone develops at the outer zone to the oxide. When that creates such a large electrical field that the free electrons can no longer diffuse into the substrate, a conductive layer made up of free charged carriers (electrons) forms around the space charge zone. This is known as the **n-Channel**.

Conductive Channel at Positive Gate-Substrate-Voltage



The voltage, from the one conductive channel is known as the **threshold voltage** (Threshold) $\boldsymbol{U_{th}}$ (in English V_{th}).

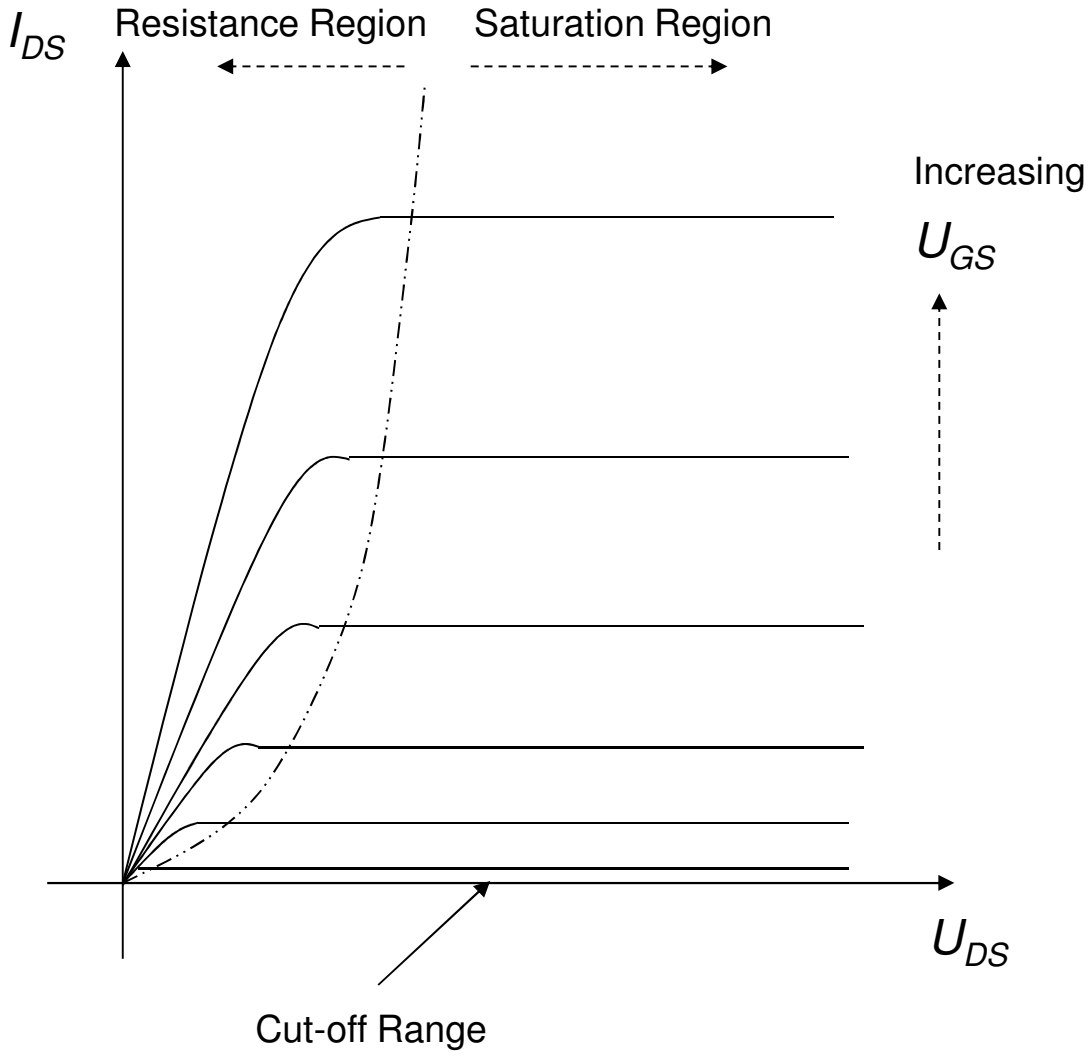
When such a conductive channel is being created, a drain current can therefore flow (from drain to source), when $U_{DS} > 0$V. The relationship of the drain current I_D from U_{DS} and U_{GS} is represented as a set of **characteristic curves** (next slide).

The substrate contact is connected to the source contact and a voltage potential of 0V is applied to it. This is used as a reference potential (Bulk, B) (which is also explicitly defined in the old german circuit symbol).

The cut-off range is when $U_{DS} > 0$V and $U_{GS} < U_{th}$. It cannot build a conducting channel. Since the drain-substrate-junction is represented as a reverse-biased connected diode, no current I_D flows.

In the working range $U_{GS} > U_{th}$, a conducting n-channel is created. Through this, the electrons flow out of the n-zone as drain current on the basis that the voltage is between drain and source.

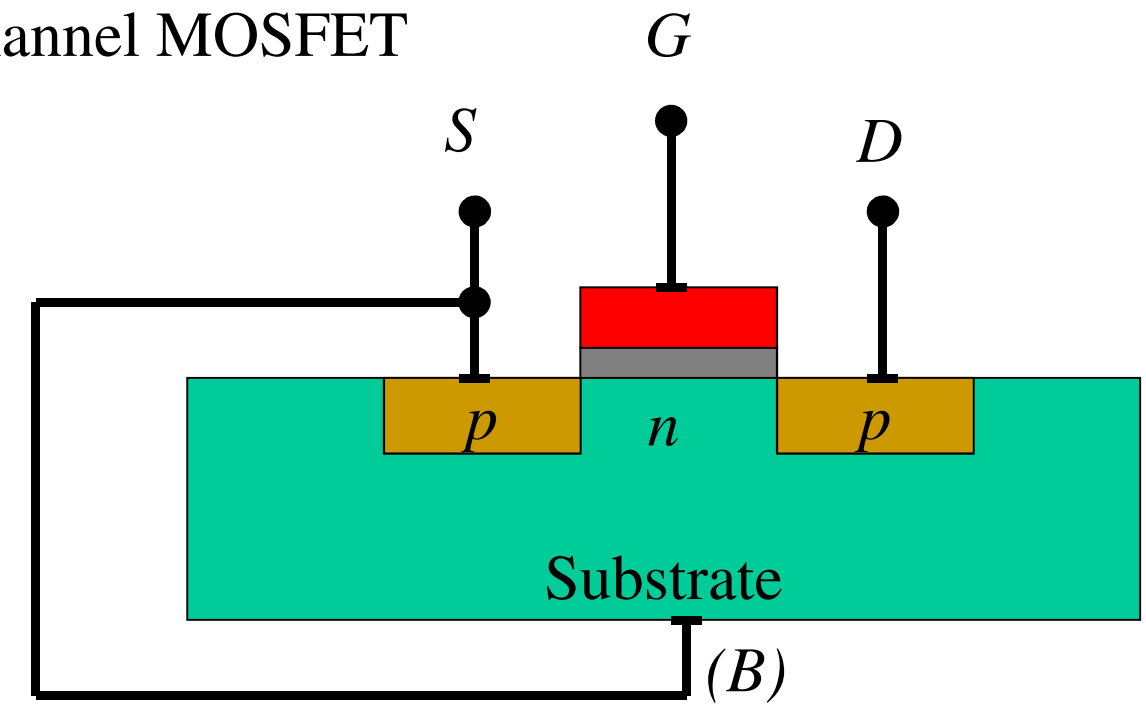
Characteristic Curves of a n-Channel Transistors

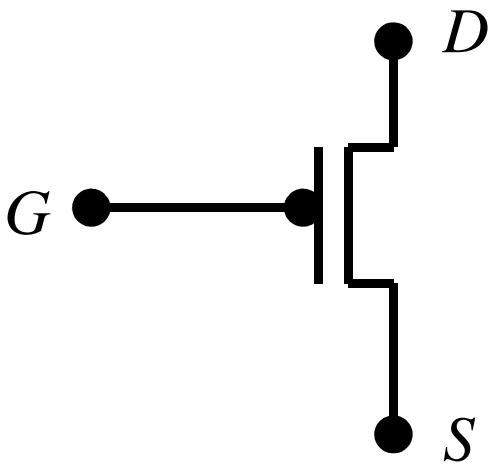


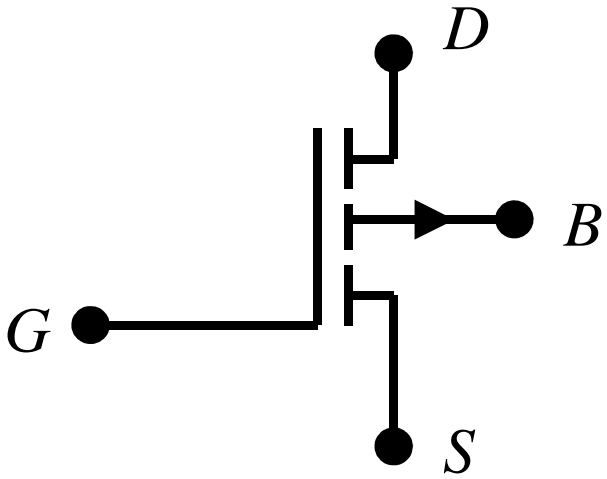
As long as $U_{DS} < U_{GS} - U_{th}$, the drain current I_D rises approximately proportional to the drain voltage U_{DS}. This is the **linear range of the characteristic curve (or resistance range)**. However, when $U_{DS} > U_{GS} - U_{th}$, the space charge zone at the drain-substrate-junction becomes larger (because it represents a diode in the reverse-biased) and the conducting channel becomes "constricted". The drain current I_D flows in the so-called **saturation range** (and not rising to any appreciable value, also when U_{DS} expands). Even with the constriction, a current still flows, since the channel exists up to a determined distance from the drain, and the electrons from there is drawn by the electrical field from the drain-source-voltage to the drain.

The p-channel MOSFET works in a similar way. Here, however, the gate is triggered negatively opposite of the source and the substrate. This, thereby creating an outer isolation layer of a conductive **p-channel**, where the drain current can flow over.

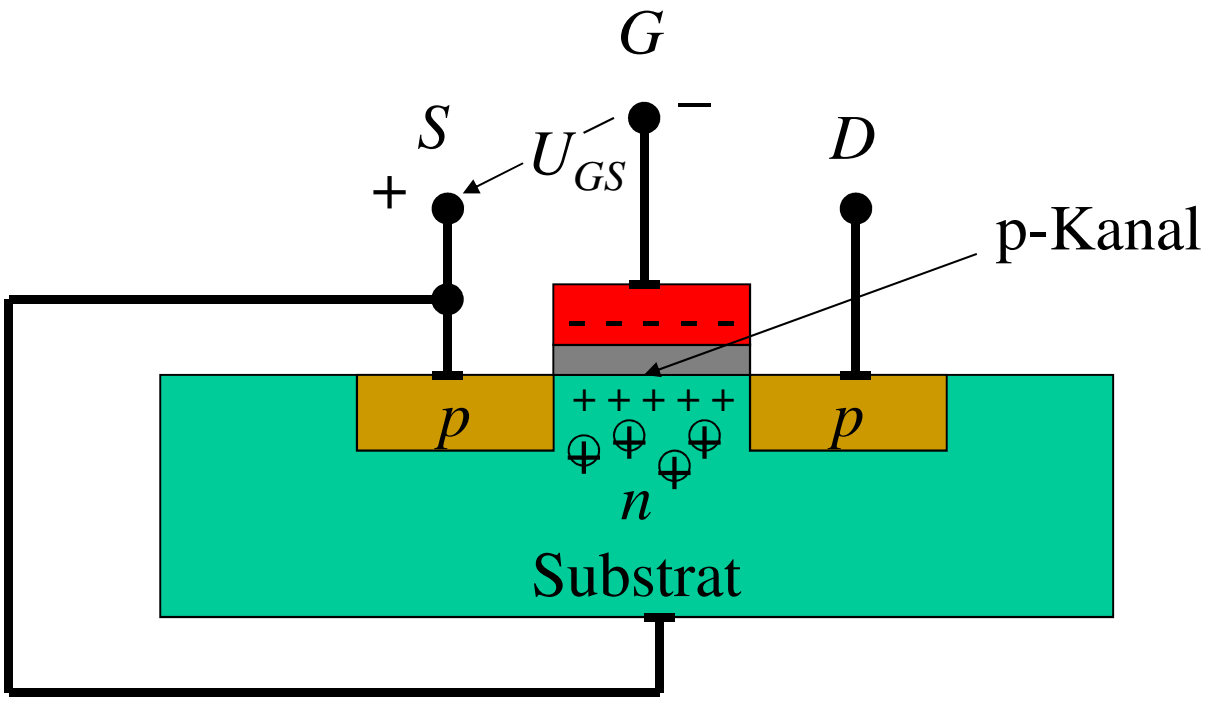
p-Channel MOSFET



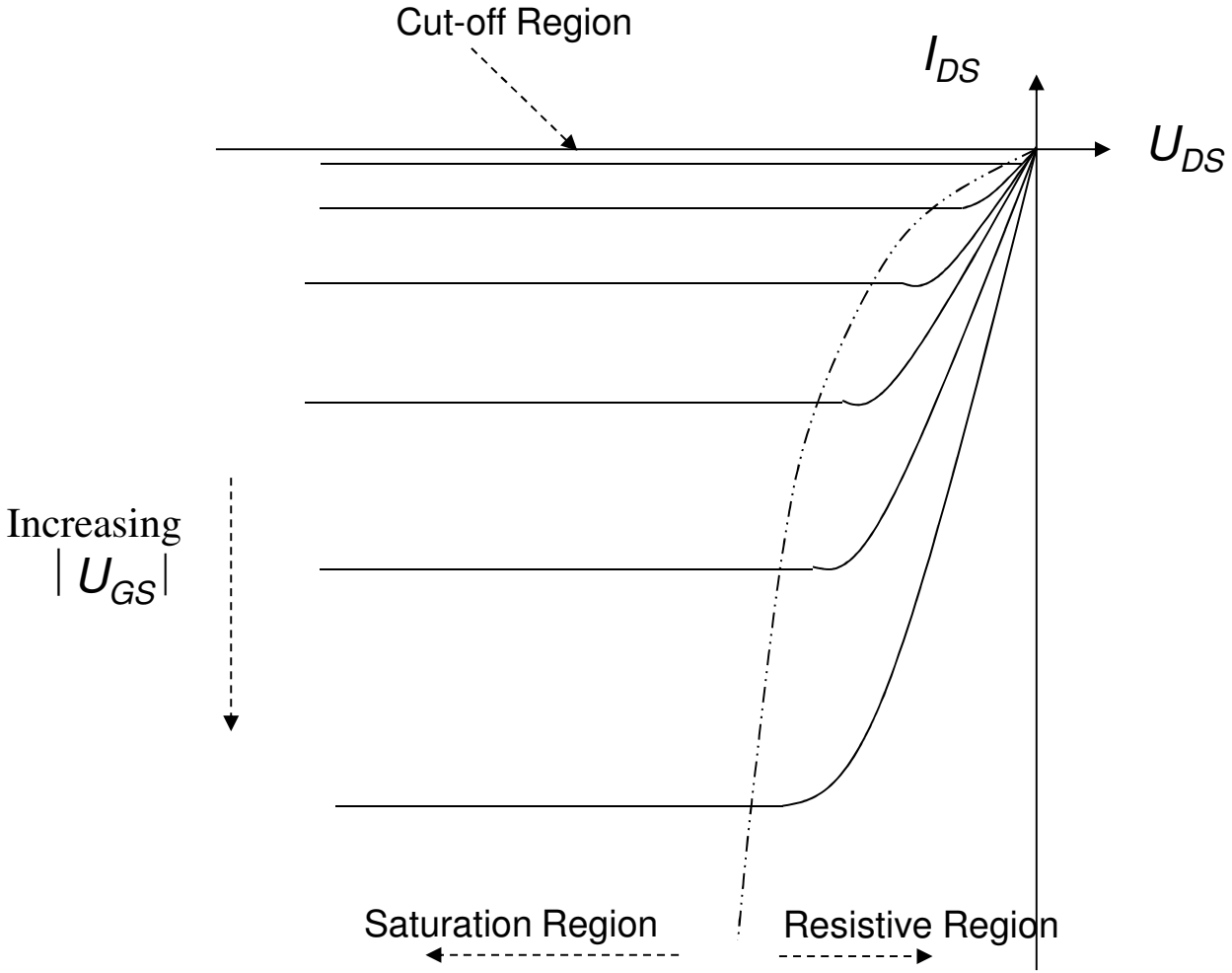




Conducting Channel with Negative Gate-Substrate-Voltage



Characteristic Curves of a p-Channel Transistor



2.7 Construction of Simple Gates using Transistors

MOSFETs can be used as switches. A n-channel transistor for example, connects drain and source, when a sufficiently high voltage is applied at its gate ($U_{GS}>U_{th}$). If the input voltage is low, the drain and the source are isolated ($U_{GS}<U_{th}$). The p-channel transistor connects when its input voltage is sufficiently small (negative) as compared to the source ($U_{GS}<V_{dd}-U_{th}$) and separated when it is not small enough ($U_{GS}>V_{dd}-U_{th}$).

Through this combination of both types of transistors, we can now build logical circuits. The simplest of such a circuit is the **Inverter**. The inverter has one input and one output. We identify the voltage at the input and output with logical values, e.g. the full input voltage V_{dd} with logical 1 (or TRUE) and the ground potential GND (0V) with logical 0 (or FALSE).

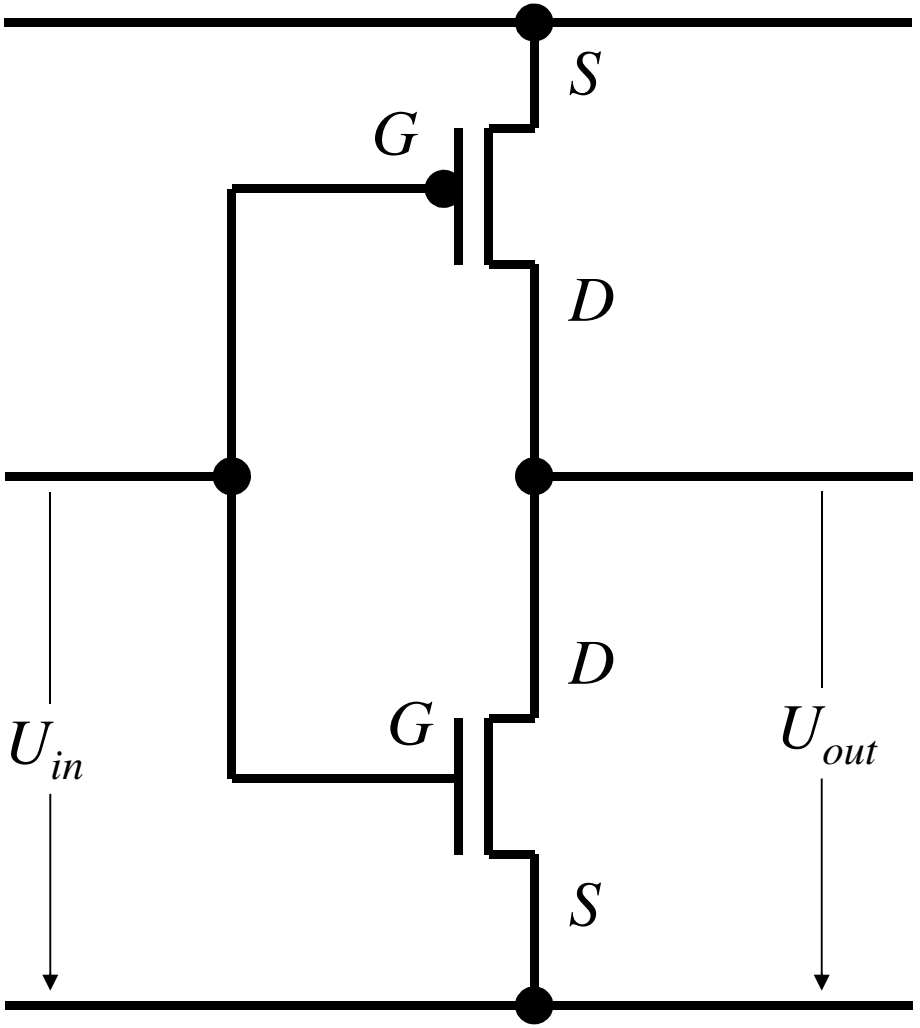
The inverter has then the following behaviour. If its input is 1, then its output shall be 0. If the input is 0, then its output should be 1. The following function table corresponds to its function:

In	out
0	1
1	0

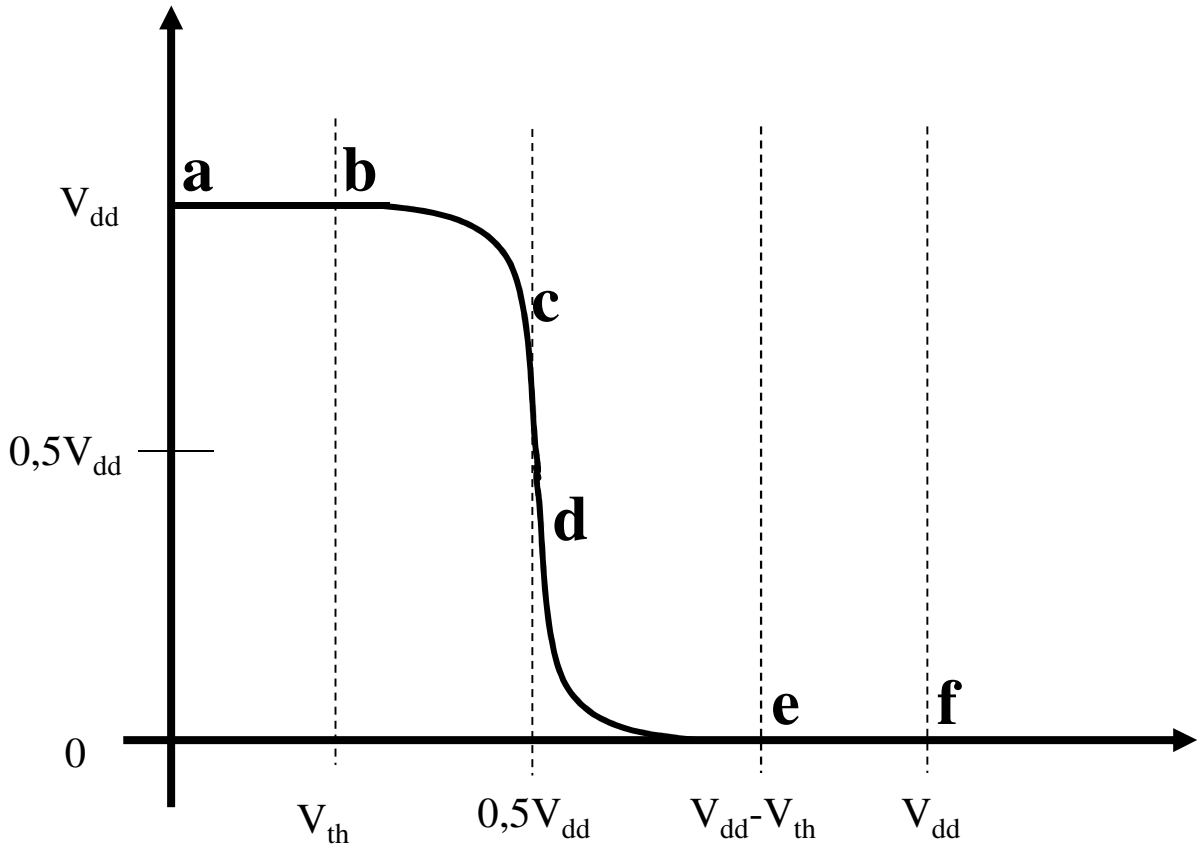
In voltage, it is written as shown in the table (whereby it is run from a supply voltage of 2.5V, which is commonly used today for integrated circuits):

U_{in}	U_{out}
0V	2,5V
2,5V	0V

Inverter

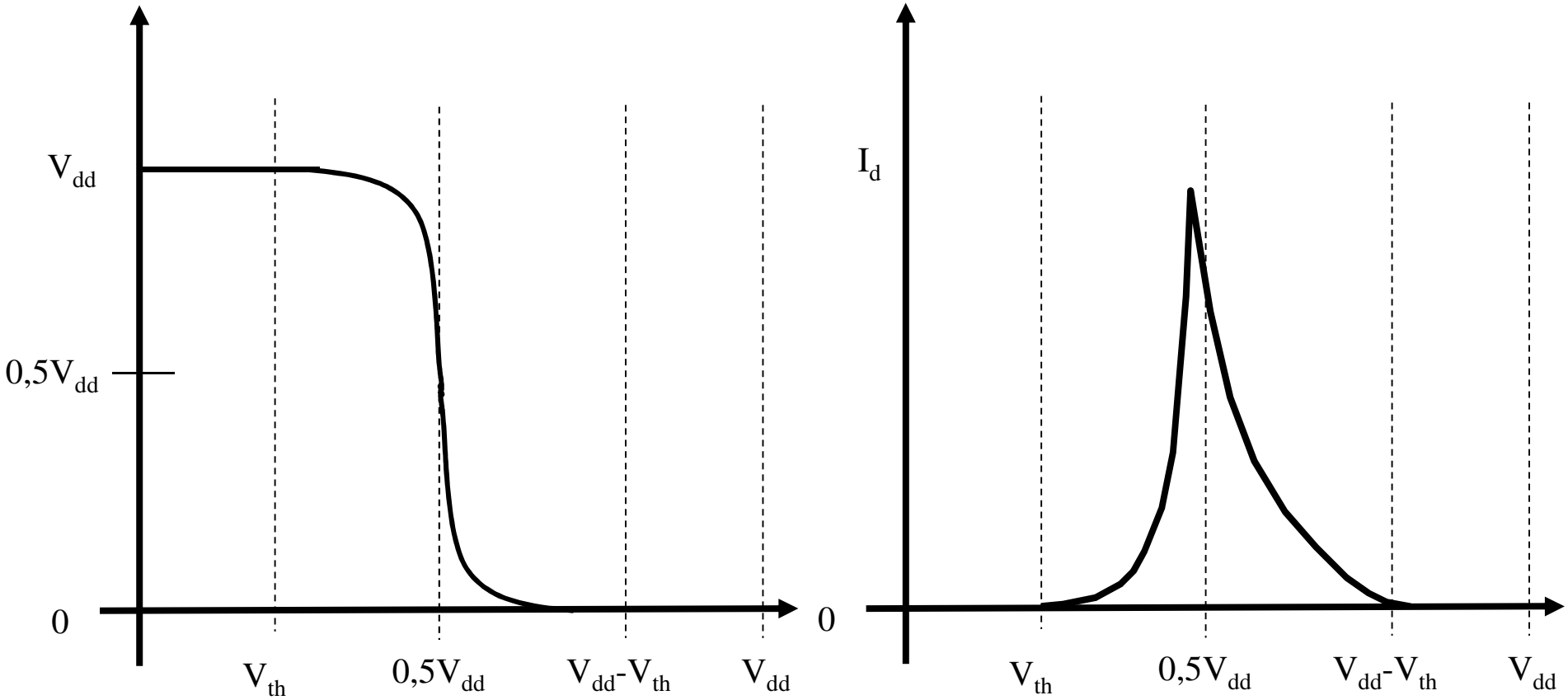


CMOS-Inverter
Output Voltage in relation to Input Voltage



We trace the voltage curve from left to right. From a to b, the lower transistor is at its cut-off region, while the upper transistor is in its resistive region. From b to c, the lower transistor is in saturation, while the transistor above continues to be in the resistive range. The resistance of the lower transistor is significantly larger than that of the above. From c to d, both of the transistors are in saturation region (here, the current is the highest). From d to e, the lower transistor enters its resistive region and the above transistor continues to be in saturation. Now, the lower transistor has only a smaller resistance as compared to the above. From point e onwards, the above transistor is cut-off. It can no longer allow current to flow.

CMOS-Inverter
Current Intake in Relation to Input Voltage



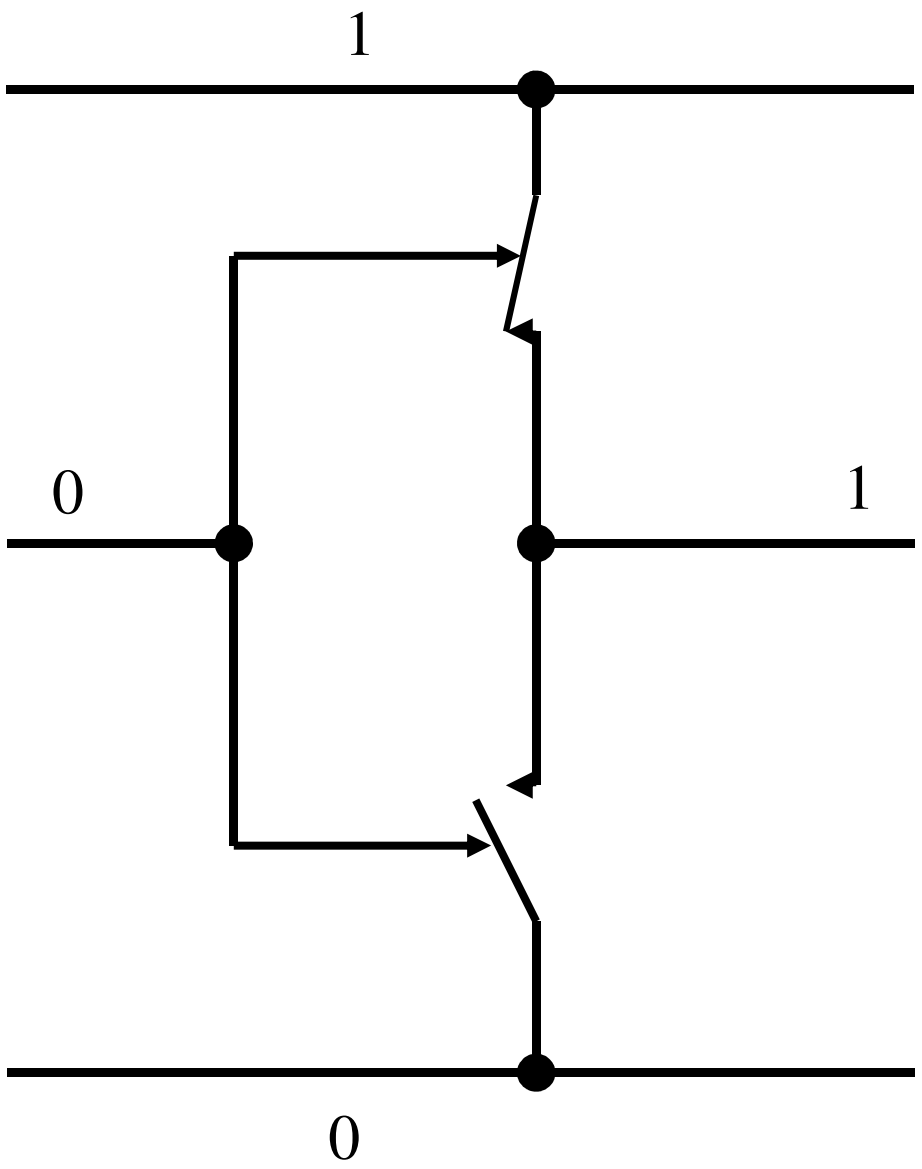
Important:

In static state of a CMOS circuit with levels $V_e<V_{th}$ or $V_e>V_{dd}-V_{th}$ there is only a very very small power dissipation.

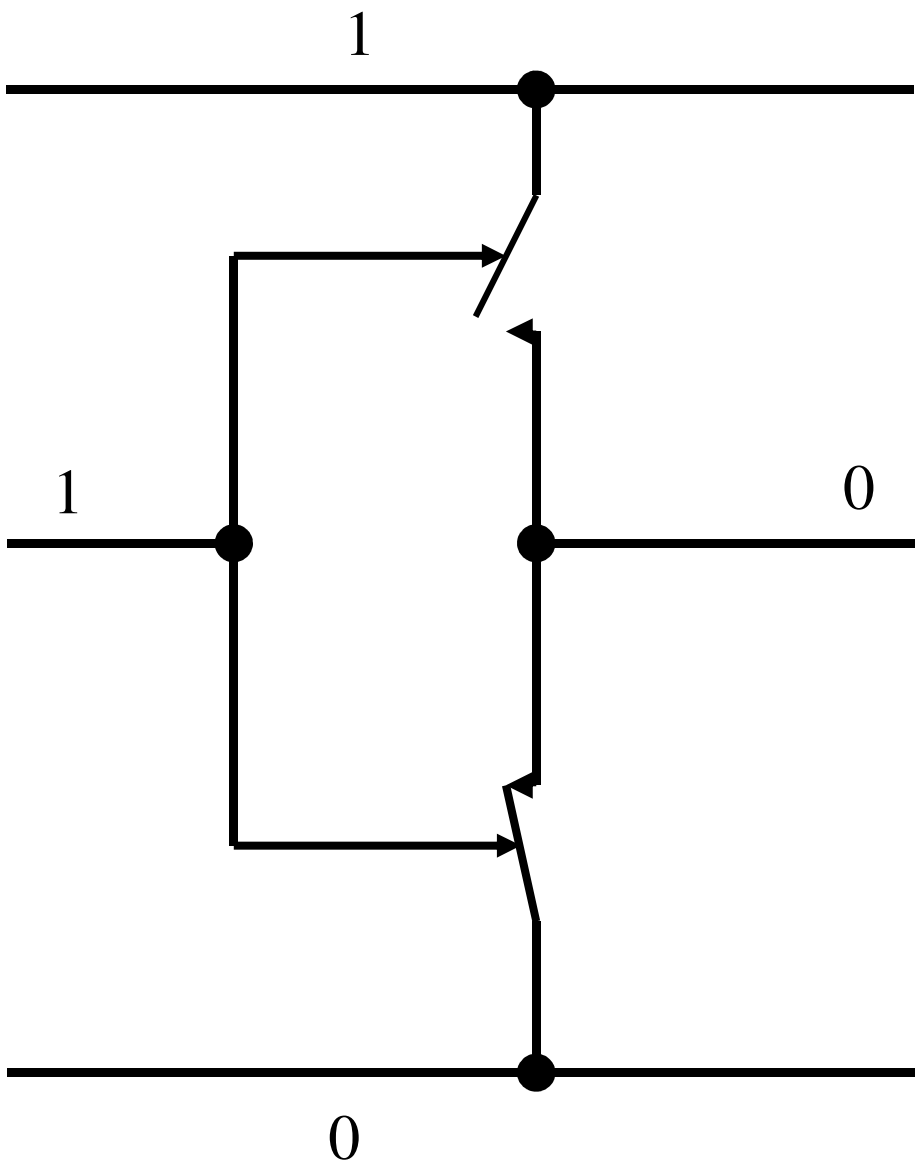
Dynamic CMOS-Power Dissipation

The power dissipation of a CMOS circuit at constant rising and
falling time of signals is proportional to frequency of the switching operation.

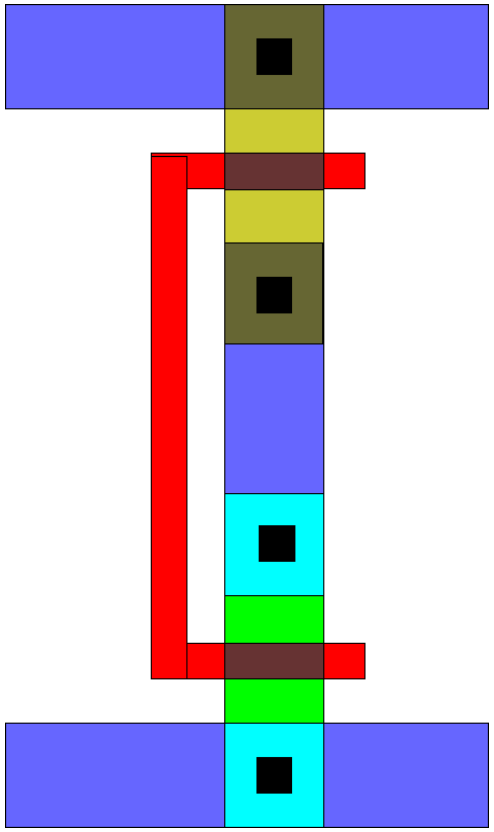
Inverter



Inverter



Top View of an Inverter on a Chip



One effect must be mentioned here. When one operates a n-MOS transistor in the saturation region, there exists only a channel, if the voltage between gate and source is greater than the threshold voltage ($U_{GS} > U_{th}$). That means that if the full power supply voltage is applied at the drain and the gate, and the source is open, it cannot set up a potential at the source, which is higher than $U_{GS} - U_{th}$. That means that a n-transistor is in fact suitable to transfer the GND potential but when transferring the full supply voltage, the supply voltage is further reduced by a value which is stated as the threshold voltage.

In the concept of logical values, it means that a 0 of a n-transistor can be well transmitted further, however, not a 1. At the output, a "bad" 1 would be generated. Therefore, the input voltage is decreased further by a value which is depicted as the threshold voltage.

A corresponding effect is created at the p-transistor. This is suitable for the forwarding of 1, i.e. at its output, a "good" 1 is generated. However, with 0 at the input, the output will generate a potential which corresponds to a GND potential plus the threshold voltage.

We take note of: **We want to use n- transistors in order to create a logical value 0 at the output of our circuit and p- transistors in order to create a logical value 1 at the output of our circuit.**

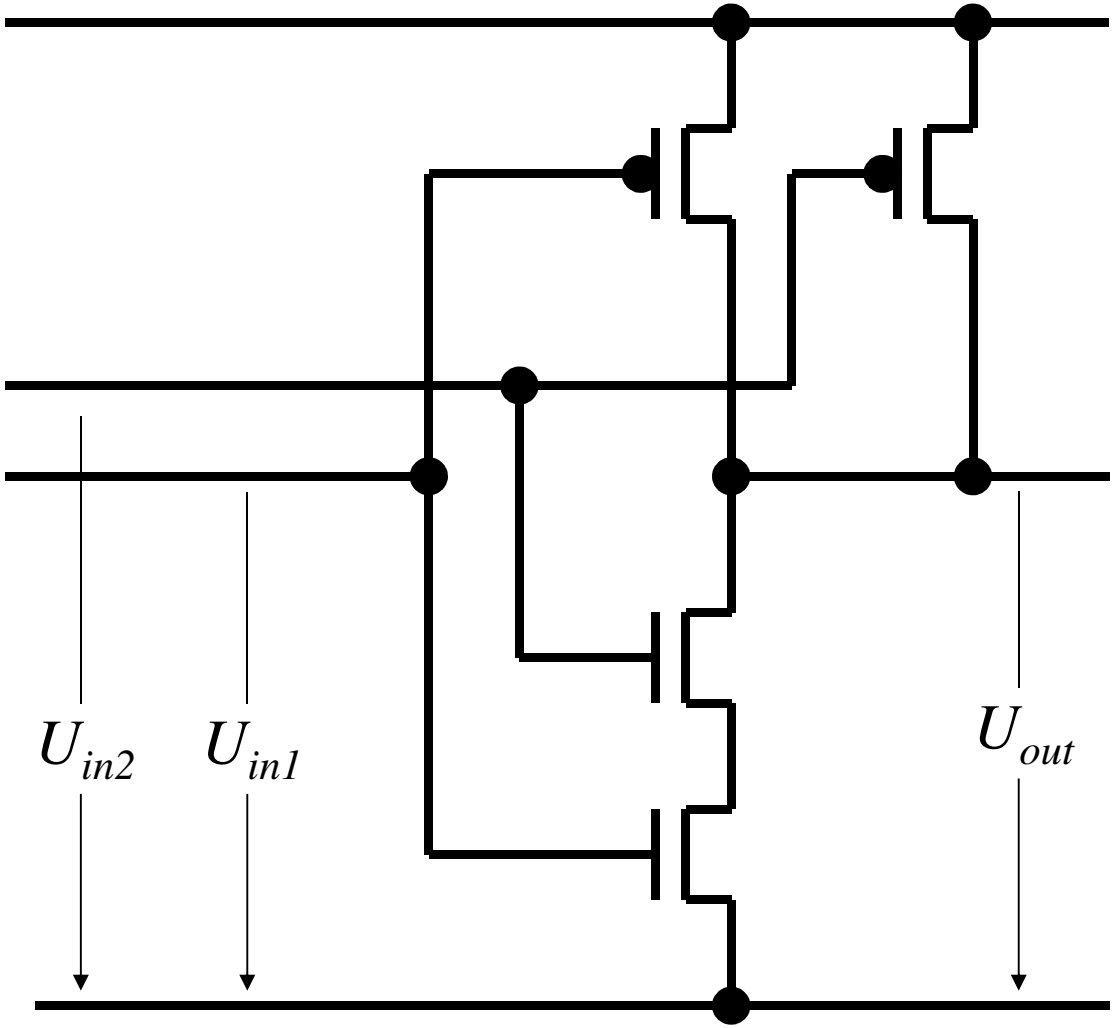
When we examine the properties of our inverter, we discover that it fulfills: With a 1 at the input, the n-transistor opens and is operating in the resistive region, i.e. at the output, a good 0 is generated. The p- transistor is disabled, since no negative gate voltage is applied in contrast to the substrate.

With a 0 at the input, the p-transistor opens and is operating in the resistive region, i.e. at the output, a good 1 is generated. The n-transistor is disabled, since no positive gate voltage is applied in contrast to the substrate.

The next function which we want to realise in a gate with MOS transistors is a NAND gate. The output of a NAND gate is 1, when both of the inputs are not 1. We must therefore, obtain at the output the GND potential when both inputs are 1(*). Furthermore, we must bring the supply voltage V_{dd} to the output, in case at least one of the inputs is 0(**).

How to achieve that? We connect two n-transistors in series and connect the source of the first transistor to GND. With that, the drain of the second transistor fulfills the first condition for the output(*). Furthermore, we connect two p-transistors in parallel, whose source are connected to V_{dd}. Once again, the drain is the output. With that, the second condition(**) for the output is fulfilled. We connect now the drain contacts of both of these paths and thus, the NAND gate is created.

NAND-Gate



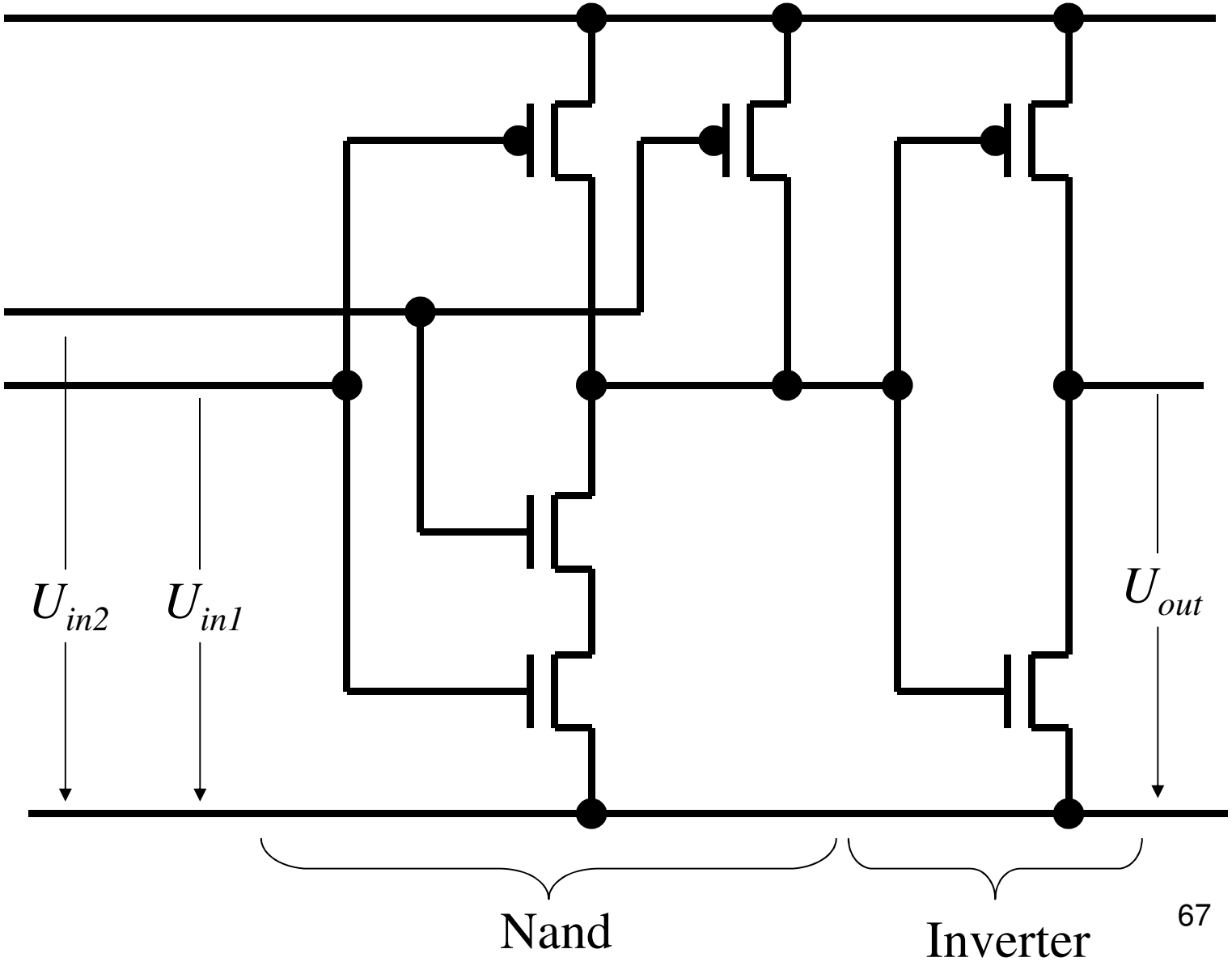
in1	in2	out
0	0	1
0	1	1
1	0	1
1	1	0

One observes that the n-transistors are simply needed again, in order to obtain the potential of logical 0 (GND) at the output. Likewise, we use the p- transistors only for transferring the logical 1 potential (V_{dd}). With that, the signals at the output do not either increase or decrease by the value of the threshold voltage as compared to the intended potential for the corresponding logical values. It is a "good" 0 and a "good" 1, which are observed at the output.

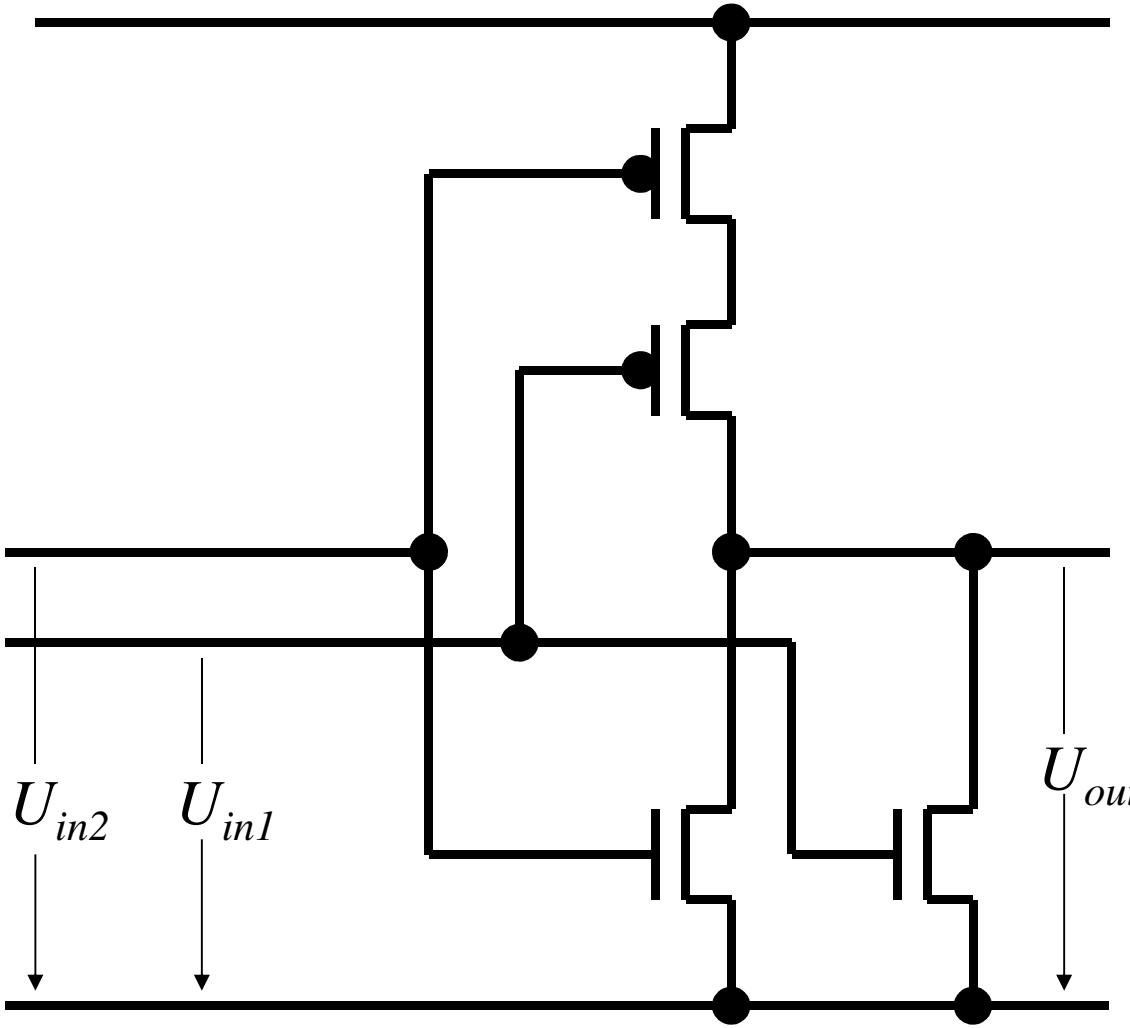
How can we build now an AND gate?
By connecting a NAND gate in series with an inverter. This circuit is seen on the next slide.

Does that work with fewer transistors?
Please try it yourself. Observe however, that the output should have a "good" signal. Why?

AND-Gate



Nor-Gate



in1	in2	out
0	0	1
0	1	0
1	0	0
1	1	0

OR-Gate