Education and Culture

Leonardo da Vinci

Pilot Project No: CZ/02/B/F/PP/134001

DynLAB

Course on Dynamics of multidisplicinary and Controlled Systems

Part II
Modelling of Semiconductor Devices

October 14, 2004

Christoph Clauss, Karl-Heinz Diener, Peter Trappe
Fraunhofer-Institut für Integrierte Schaltungen
Branch Lab Design Automation
01069 Dresden, Germany
http://www.eas.iis.fhg.de

Contents

Module 1

Introduction

To become competitive at markets, today, innovative system solutions for new products have to be based on key technologies as electronics, microelectronics, telecommunication and micro-electro-mechanical systems (MEMS). Actually, the technical progress in these key technologies proceeds rapidly. Therefore, classical education at universities has to be accomplished by additional learning activities, particularly focused on

- Life-long learning in companies and research institutions
- Assisting classical teaching courses to improve the learner's motivation
- Imparting short-lived knowledge distilled from latest R & D activities.

To cope with the enormous increase of knowledge day for day on the one hand, and the limited time to impart it on the other hand, multi-media-based learning facilities become more and more important. In particular, the Web-based education methodology offers new chances and challenges. Indeed, Internet-based techniques have set up the technical basis to meet the demands for flexibility in learning at different places and at any time as well as updating the short-lived knowledge.

In fact, simulation has proved as a very important method to perform virtual experiments to get deeper understanding of interdisciplinary processes, and to gain insight in complex systems aimed at verifying and optimizing new designs and developments. Therefore, simulation should be an integral part in classic teaching mainly relied on face-to-face courses as well as in engineering.

This widespread spectrum of applications is one of many other reasons to provide powerful simulation facilities.

In this sense, the interactive simulator Dynast http://virtual.cvut.cz/cacsd/msa/onlinetools.shtml developed at the Czech Technical University in Prague which can be used via the Internet is an appropriate simulation tool.

Dynast is the software basis of a Web-based course on Dynamics and Control of Multidisciplinary Engineering Systems (DynLAB). Partners in four countries are developing this course under the umbrella of the European Leonardo project. The multidisciplinary modeling and simulation strategy is strongly focused on a very general "multipole approach" that has been proved successfully for years.

Generally, the development of mixed electrical and non-electrical systems as well as electronic solutions for new product waves is mainly based on design-driven simulations. Therefore, the Dynast simulator has to be provided with a suit of models for semiconductor devices (e.g. diodes, MOS and Bipolar transistors). Semiconductor devices are the main item in modern circuitry, either in a complex net of discrete electronic devices (capacitors, inductors, resistors and transistors, etc.), or as integrated circuits (IC) on semiconductor chips. Basically, diodes and transistors exhibit a nonlinear current/voltage (I/V) characteristic. That is why already pretty simple circuits cannot be understood completely without a detailed analyze. This can be done by simulation with Dynast. Basically, four fields of using Dynast may be highlighted in the following:

1. The user (or learner) is mainly interested in analyzing a given and developed circuit, respectively. In this case, he needs stable device models capable to describe all essential effects to meet the circuits functionality. Therefore, a set of models for different demands (simple models until advanced models) should be available. The user's (learner or engineer) main interest is focused on how to specify the model's parameters to simulate the circuit under test successfully.

2. The user (mostly designers and engineers) has to find out why his circuit does not meet all characteristic parameters as specified. In this case, he has to identify what the weak points are, and how to cancel them. In this case, the electronic devices have to be included in the failure consideration. That requires more information on device parameters to be specified before simulation. In particular, the users are interested in their impact on the device's internal electronics. Often, the problems arisen in the circuit under development are due to thermal effects. Typically, power dissipation can shift temperature sensitive parameters of the semiconductor devices resulting in breaking down the circuit's functionality. To identify such thermal effects it is very important to have available device models with a thermal pin. Therefore, critical heating effects can be identified by simulations and minimized before manufacturing the circuits.

3. The user is compelled to include his own model (company specific models) into the Dynast simulation environment. In this case, he needs all relevant information to implement his models correctly.

4. Web-based course on Dynamics and Control of Multidisciplinary Engineering Systems has been set up for assisting classical teaching courses with virtual experiments to improve the learners' motivation to deal with complex physical phenomena more in detail. Moreover, it is aimed at performing professional training in companies and also for in-house qualification activities.

Based on appropriate advanced models, Dynast also offers the chance to provide engineers with leading edge as well as fast moving knowledge. Thus, Dynast will bridge the gap between the basics mainly imparted at universities and interdisciplinary expertise necessary for mastering the design processes in key technologies as electronics and micro-electro-mechanical systems (MEMS).

The course "Modelling of Semiconductor Devices" is indented to fulfill these entire goals.

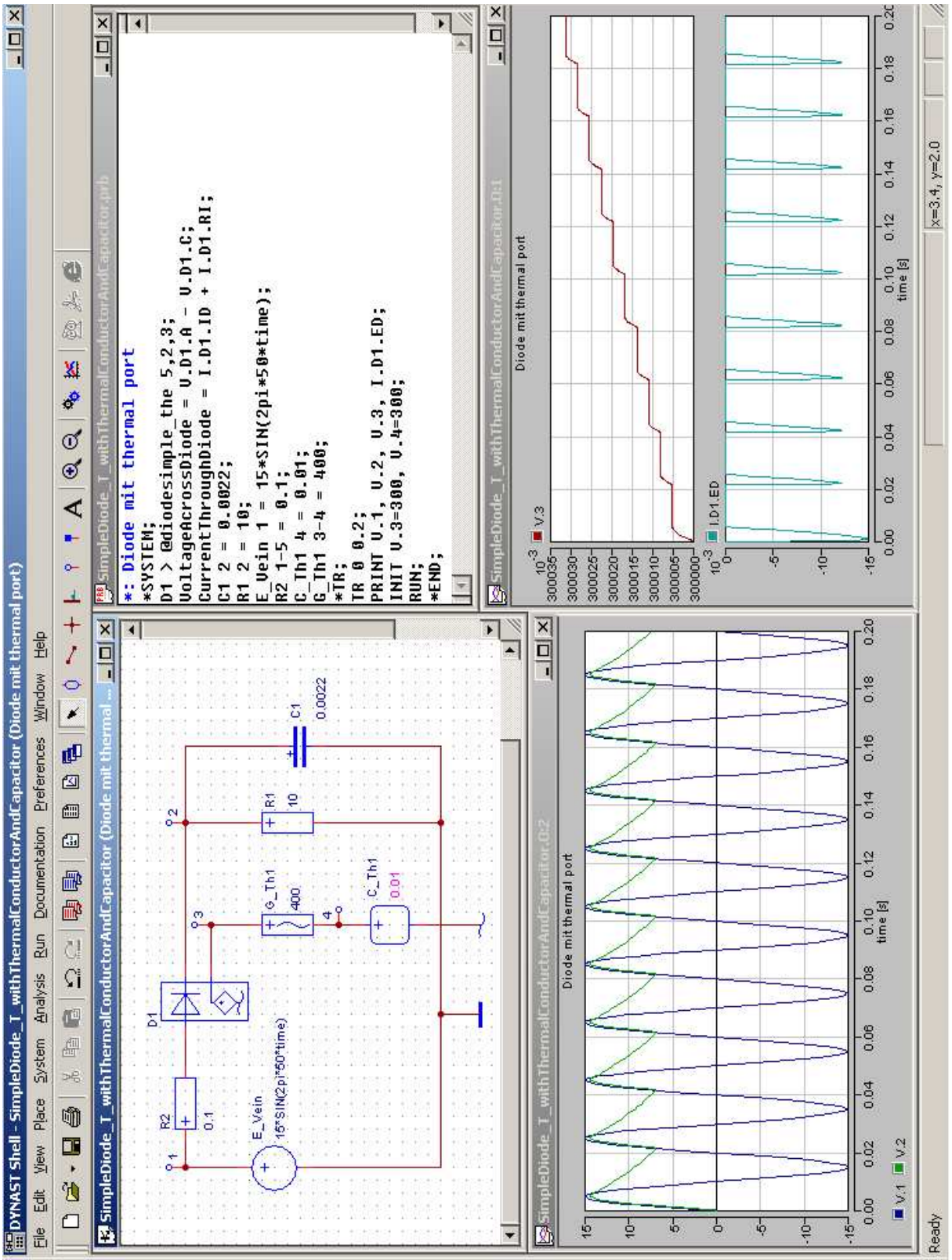


Figure 1.1: Dynast simulation: schematic entry, simulation control, and results

Module 2

Inserting a Device Model into DYNAST

Module overview. Granted that a user wants to create a new transistor model into the DYNAST simulator. In this module it is described step by step what has to be done to insert a transistor model if the formula is given.
The very simple model is chosen in view of transistor models that are commonly in use, especially in DYNAST.

Module objectives. *When you have completed this module you should be able to:*

1. Insert an own model into DYNAST
2. Use his model within simulation experiments

Module prerequisites. DYNAST installation, basic experience in using DYNAST

As an example we take an oversimplified model with the pins G (gate), D (drain) and S (source). The drain-source-current i_tr is equal to

```
i_tr = if vg>p1 then (vd -vs) p2 else (vd - vs)*p3
```

with vd - drain voltage, vs - source voltage, vg - gate voltage, and p1, p2, and p3 parameters. The gate current is zero.

The following steps have to be carried out:

1. Start the DYNAST shell

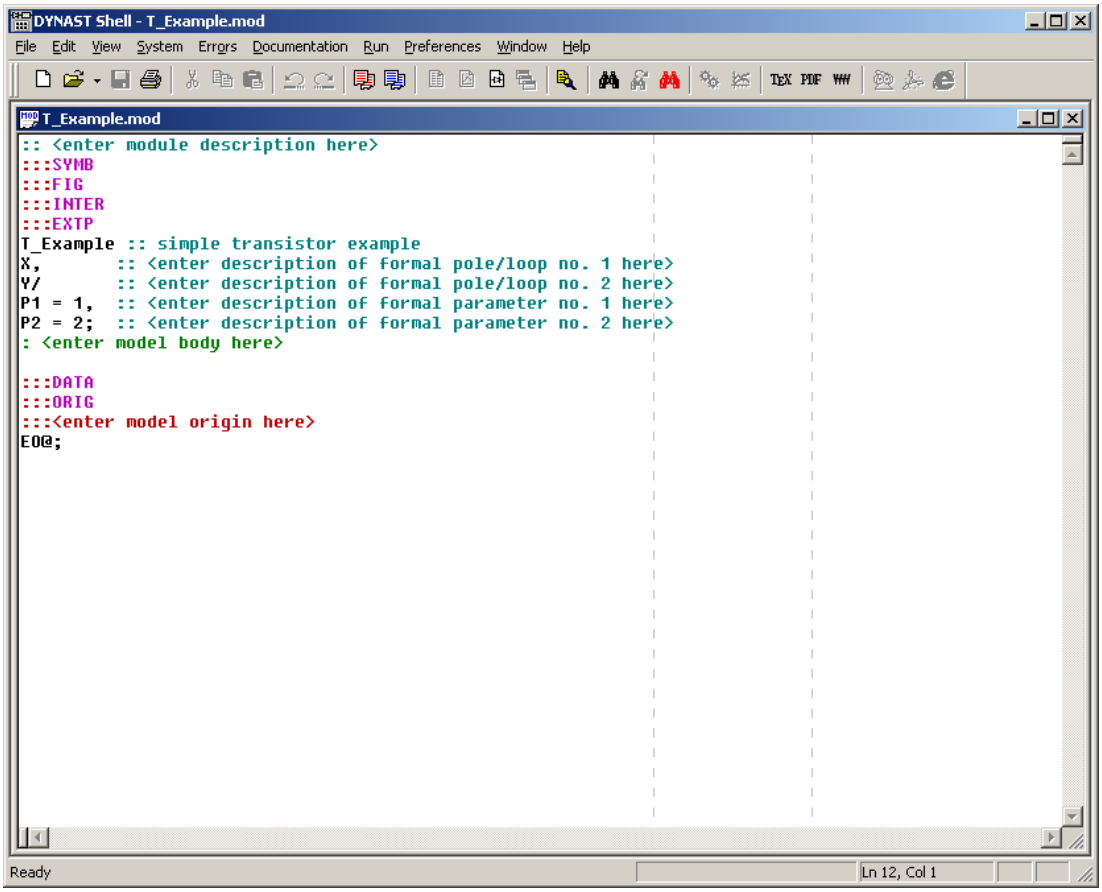
2. Create a model file

Press ***File*** > ***New***

In the following window choose :
File type: ***submodel text***
File name: T_Example
Model: simple transistor example
Choose the documentation template.

Press ***OK***

A window with a general template for a model appears where the text can be edited:



3. Fill in the Model File

The model file has to be filled in by

- Adding the pins (X, Y are replaced by the pin names)
- Adding the parameters (P1, P2 are replaced)
- Adding the transistor formula.

The pin currents are set by using an internal current source:

```
IG> J G = 0.0;
```

IG is the name of the current source, J is the current source identifier, G means, that the current is flowing from the pin G to the ground. The current value is zero. In such a way single pin currents can be set to certain values.

Since the i_tr current flows from the drain to the source pin, one current source can be taken for both, the drain and the source pin:

```
IDS> J D S = i_tr;
```

The result after filling in the model file is:

```
::<enter module description here>
:::SYMB
:::FIG
:::INTER
:::EXTP
T_Example :: simple transistor example
G,     :: Gate
D,     :: Drain
S /    :: Source
P1=1,  :: Threshold
P2 = 100,  :: high conductance
P3 = .01;  :: low~conductance

IG> J G = 0;

i_tr  =   (V.G>p1)  *  (V.D-V.S)*p2+
          (V.G<=p1) *  (V.D-V.S)*p3;

IDS> J D - S = i_tr;

:::DATA
:::ORIG
:::<enter model origin here>
EO@;
```

This step is concluded by pressing ***OK***

4. Create a symbol library

Press ***File*** > ***New***

In the following window choose :

File type: symbol library
File name: T_Lib

Press ***OK***

Then a window appears for graphical editing of a symbol of the transistor.

5. Design a symbol of the transistor

Draw a picture which you want to use for the transistor within the red line.

Create the ports and connect it to the port names of the model:

Press the **A Button**, put the connector to the correct place, and fill in the pin properties:

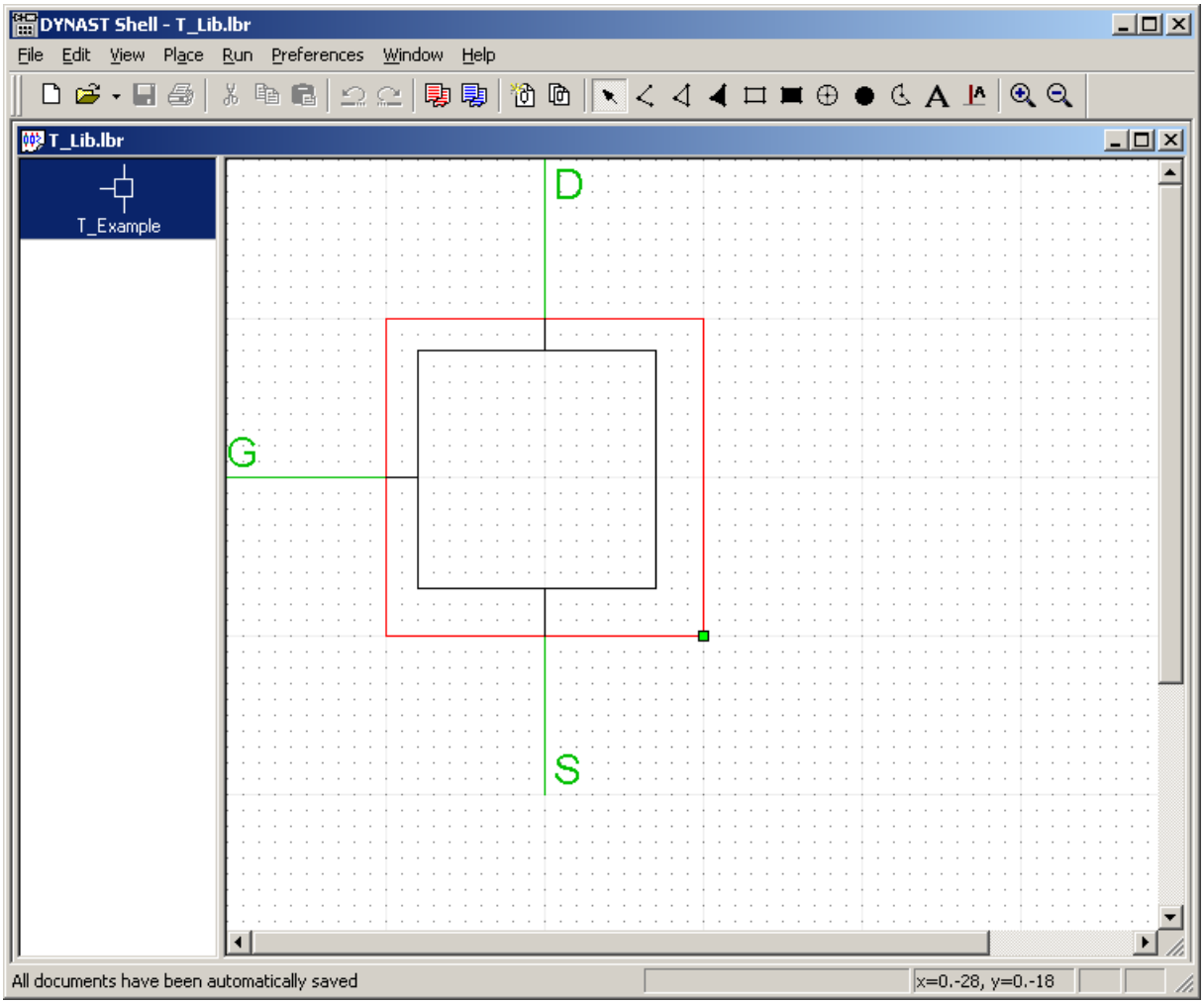
Pin Properties: **Name:** **D**
Type: electric

Press **OK**
Click the symbol schematic in the left upper corner twice.
Change the symbol properties:

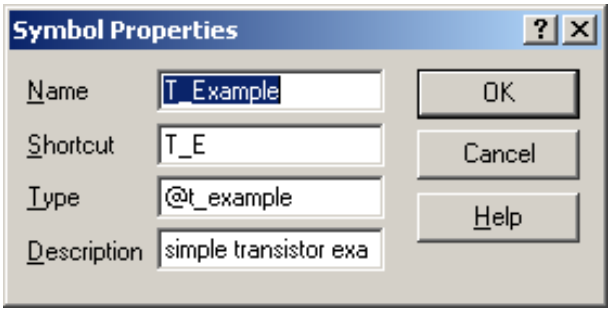
Symbol Properties: **Name:** T_Example
Shortcut: T_E
Type: @t_example
Description: simple transistor example

Close the window.

In our test case the symbol is a simple rectangle:



This is the window of the symbol properties:



6. Add comments to the transistor model

To complete the transistor modelling some comments should be added as it is adumbrated in the following model text:

```
:: simple transistor example
:::
::: Information for model documentation in latex formate ...
:::
:::INTER
:::EXTP
:::DATA
T_Example :: simple transistor example
G,    :: Gate
D,    :: Drain
S /   :: Source
P1 = 1,     :: Threshold
P2 = 100,   ::  high conductance
P3 = .01;   ::  low conductance

IG> J G = 0;

i_tr =  (V.G>p1)   * (V.D-V.S)*p2 +
        (V.G<=p1)  * (V.D-V.S)*p3;

ID> J D - S = i_tr;

EO@;
:::ORIG
:::Copyright in latex formate
```

Now the model is complete and can be used in any circuit!

7. Create a circuit model

To create a circuit model the following step has to be done:

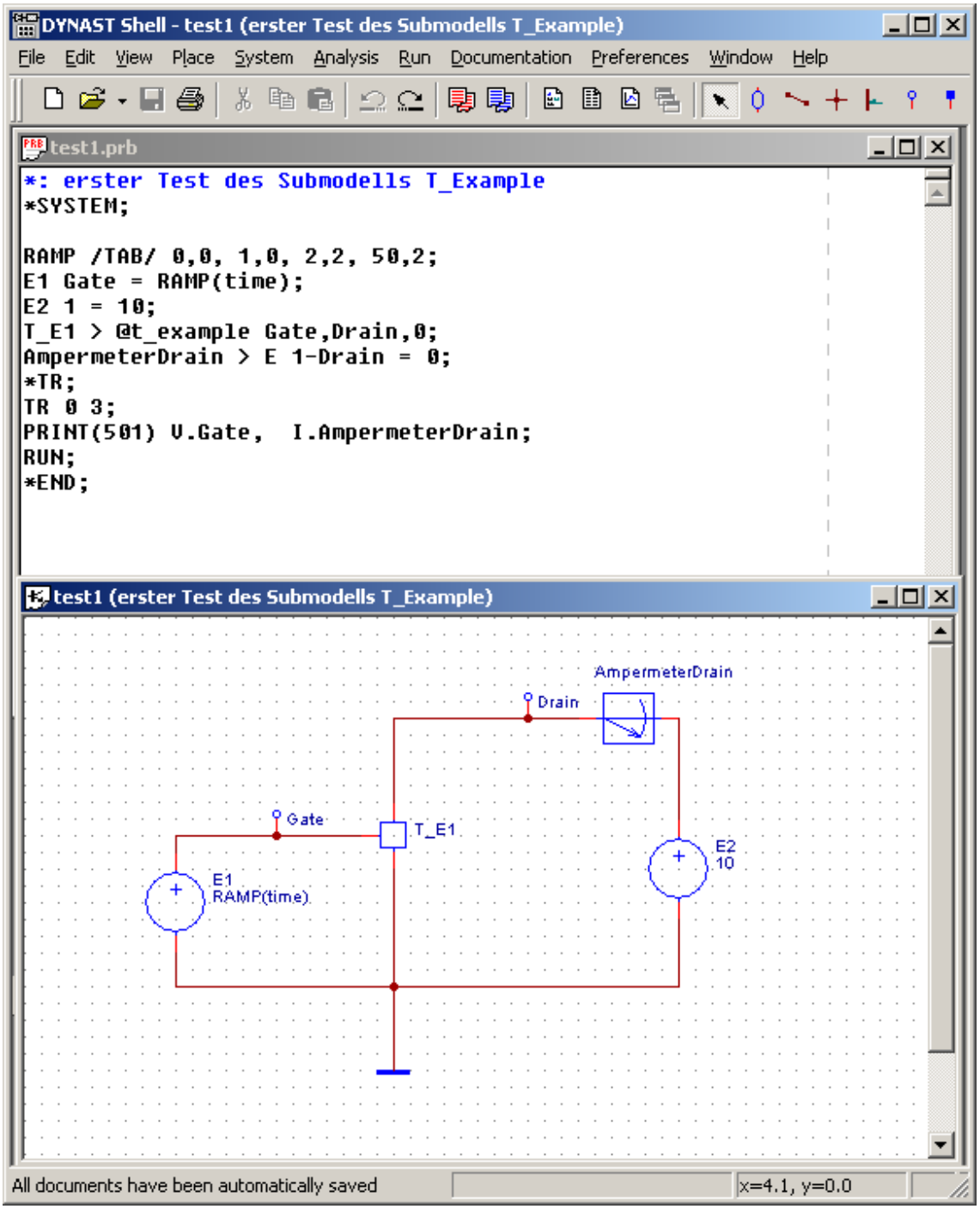
Press ***File*** > ***New***

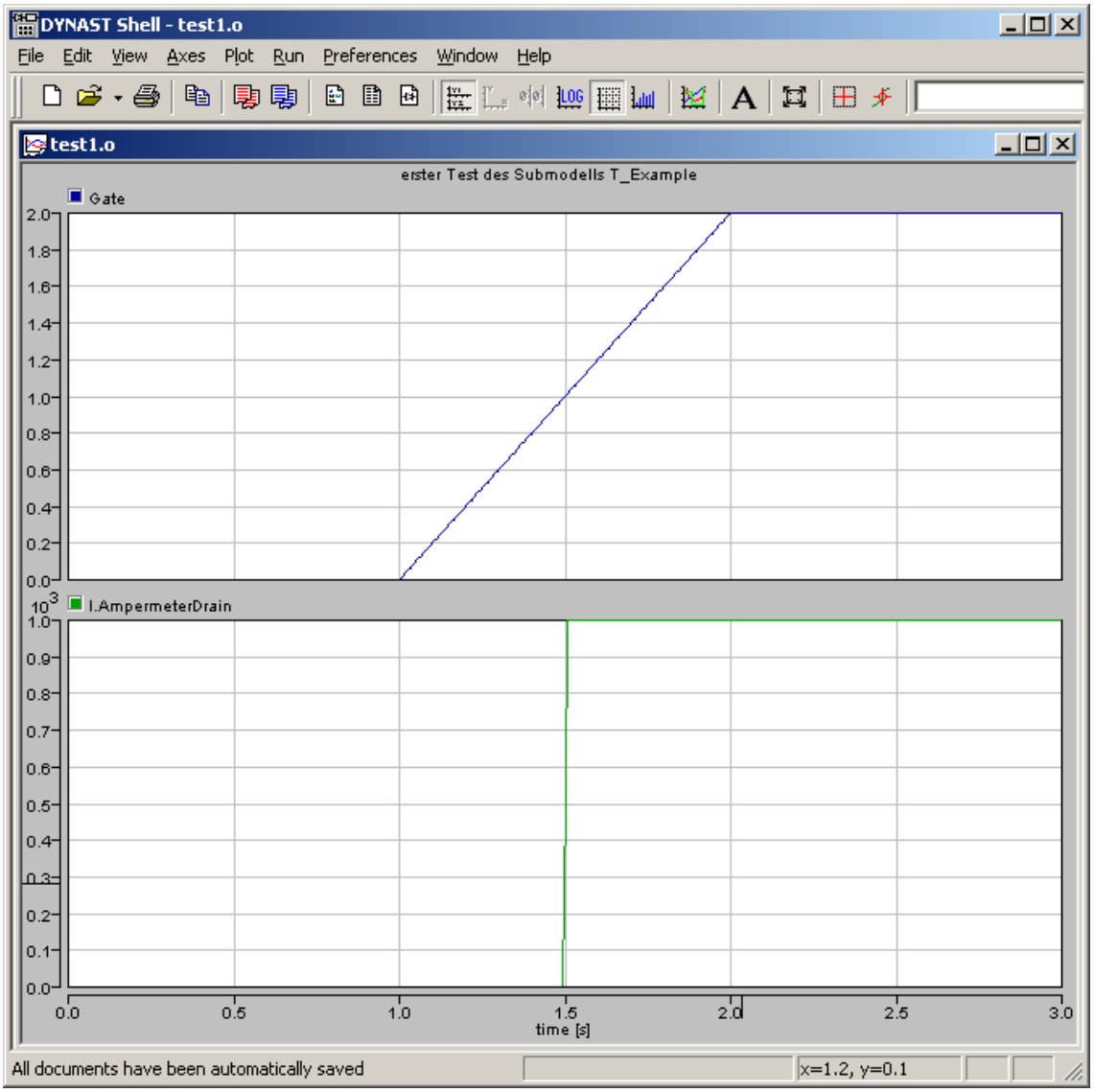
In the following window choose :
File type: ***problem text***
File name: test1
Title: erster Test ...

Choose Standard template.
Press ***OK***

Then a text window appears to be edited. Additionally a diagram window can be opened, in which the transistor and other electrical models can be placed by drag and drop, and connected.

The model text and the circuit scheme are e.g. like this:


DYNAST Shell - test1 (erster Test des Submodells T_Example)
File Edit View Place System Analysis Run Documentation Preferences Window Help
test1.prb
*: erster Test des Submodells T_Example
*SYSTEM;
RAMP /TAB/ 0,0, 1,0, 2,2, 50,2;
E1 Gate = RAMP(time);
E2 1 = 10;
T_E1 > @t_example Gate,Drain,0;
AmpermeterDrain > E 1-Drain = 0;
*TR;
TR 0 3;
PRINT(501) V.Gate, I.AmpermeterDrain;
RUN;
*END;
test1 (erster Test des Submodells T_Example)
AmpermeterDrain
Drain
Gate
T_E1
E1
RAMP(time)
E2
10
All documents have been automatically saved
x=4.1, y=0.0



The simulation results show that the drain current is jumping if the gate voltage becomes greater than one. It is caused by increasing the conductance from p3= 0.01 to p2=100 at the same position.
This example (netlist and simulator commands) can be found in the file
$InsertExercise2-1.prb$.

Module 3

Electronics of Semiconductor Structures

Module units

Module overview. The electronics of semiconductor devices is mainly based on mobile and fixed charges controlled by electric fields due to space charges within the semiconductor and supplied electrode potentials. Before analyzing the basics of the interaction between these quantities in a phenomenological manner, we have to be prepared on questions like

- What are the main differences between a semiconductor on the one hand and metallic conductors and insulators on the other hand?
- How to provide semiconductors purposefully with mobile carriers?
- How to make a semiconductor to a p -type or n -type conductor?
- What kind of charge carriers are available to carry a current in semiconductors?
- What makes the difference between free charges outside the semiconductor and such ones caught in a crystal lattice?
- Which phenomena are active to move carrier in semiconductors?
- How to find out the interaction between electric fields and charges in semiconductors?

This module is intended to give answers to these and other questions from a more phenomenological point of view. That means we are especially interested in providing modeling approaches rested on the basics of electronics as imparted in the curricula at the universities and advanced technical colleges. Moreover, it should be mentioned that the processing technology for semiconductor devices will be focused on dedicated results only without discussing details on how to perform such processes.

So, we establish the basis for a widespread understanding of the working principles of the most important semiconductor devices.

Module objectives. *After studying this module learners should:*

1. Understand the basics necessary for an analytically oriented description of the operation principles of semiconductor devices.
2. Be acquainted with dedicated approaches proven in device modeling for years.

Module prerequisites. Basics in semiconductor physics and electronics

3.1 Charge carriers in semiconductors

3.1.0 Constants of semiconductor electronics

Absolute dielectric constant of the empty space	$\varepsilon_0 = 8.86 \times 10^{-12} As/Vm$
Permittivity constant of Silicon (Si)	$\varepsilon_H = 10^{-12} As/Vcm$
Relative dielectric constant Si	$\varepsilon_{r_H} = 12$
Dielectric constant of Silicon oxide (SiO_2)	$\varepsilon_{ox} = 3 \times 10^{-13} As/Vcm$
Elementary charge	$q = 1.6 \times 10^{-19} As$
Free electron mass	$m_0 = 9.11 \times 10^{-31} kg$
Energy gap of Si	$W_g = 1.1\ eV$
Electron affinity of Si against vacuum	$W_{EA_{Si}} = 4.1\ eV$
Electron affinity of Si against SiO_2	$W_{EA_{SiO_2}} = 3.25\ eV$
Planck's constant h	$h = 6.63 \times 10^{34} Ws^2$
Boltzmann's constant$k$	$k = 1.38 \times 10^{-23} Ws/K$
Reference temperature	$T_0 = 300\ K$
Thermal voltage V_T	$V_T\,(300\ K) = 25.9\ mV$
Effective density of states in conduction band	$N_C \approx 10^{19} cm^{-3}$
Effective density of states in valence band	$N_V \approx 10^{19} cm^{-3}$
Intrinsic density for Si	$n_i\,(300\ K) = 1.5 \times 10^{-10} cm^{-3}$
Diffusion coefficient for electrons in Si	$D_n \approx 40\ cm^2/s$
Diffusion coefficient for holes in Si	$D_p \approx 10\ cm^2/s$
Averaged electron lifetime in Si	$\tau_n \approx 10^{-9}\ s$
Averaged hole lifetime in Si	$\tau_p \approx 10^{-7}\ s$
Space charge layer carrier life time	$\tau_s \approx 10^{-9}\ s$
Forward floating potential for Si diodes	$V_{FO} \approx 0.7\ V$

3.1.1 Electrons and holes in doped semiconductors

Generally, particle currents in semiconductors rely on two sorts of mobile carriers, the electrons and defect electrons (so-called holes). Electrons and holes carry the elementary charge $-q$ and $q = 1.6 \times 10^{-19} As$, respectively. Because these carriers are caught in a crystal lattice, they are interactions with the ionized atoms fixed in the crystal lattice.

That is why they are not "free", rather than "quasi-free". Basically, the carriers in crystals are influenced by electric fields (forces) externally imposed on the semiconductor, as well as by internal fields due to the ionized atoms of the lattice. Therefore, they exhibit characteristics which differ from them of fee carriers. To keep formally at the approach developed for supplied external forces (fields), we have to involve such internal interactions implicitly at least. Thus, we use the carrier's effective mass (m_n and m_p) instead of the their free carrier mass m_{0n} and m_{0p}, respectively. The effective mass can be derived from the energy band structure of semiconductors. Based on the wave-particle-dualism, globally, the band structure describes some characteristics of the "wave behavior" of quanta in semiconductors. In fact, semiconductors exhibit anisotropic band structures, i.e. the effective mass of electrons and holes are tensors.

Another important difference to free carriers concerns their energy. Generally, free carriers can accept unlimited energy levels, however, that is not true for carriers in crystals. Due to the internal interactions with the ionized lattice atoms, the electrons and holes in solids may only reach restricted values of energy. In particular, there is a material-specific energy band gap W_g(forbidden energy band) for electron and holes. It separates the conduction band with the lower band edge W_C from the valence band with the upper band edge W_V (see Figure 3.1), i.e. $W_g = W_C - W_V > 0$.

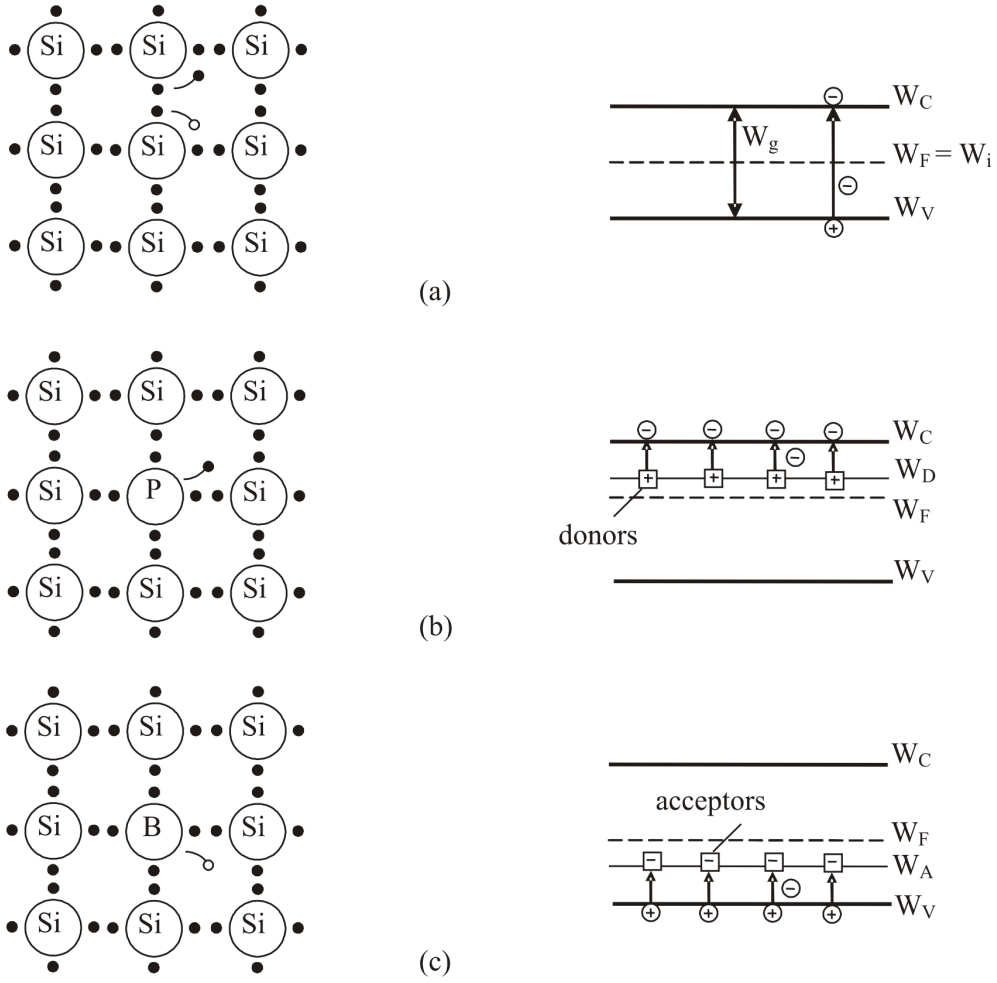


Figure 3.1: 2D lattice structure (schematically) [2]

At room temperature the forbidden energy gap between W_C and W_V is $W_g = 1.1\ eV$ for silicon Si. Basically, carriers in undoped semiconductors (intrinsic semiconductors) are generated in pairs by thermal exciting valence band electrons. Leaving the valence band and entering the conduction band, each of the thermally excited electrons generates a (mobile) hole in the valence band (see Figure 3.1(a)). Both, the electrons lifted into the conduction band and the defect electrons (holes) left behind the valence band will contribute to a current flow. Generally, this kind of intrinsic conductivity is negligible compared to the conductivity due to ionized impurities [1], [2].

In this case, impurity atoms with a valency of five (so-called donors like phosphorus (P) or arsenic (As)) and a valency of three (called acceptors e.g. boron (B)) replace purposefully some silicon (Si) atoms in the crystal lattice, respectively. If the donors and acceptors get ionized, each donor delivers an electron to the conduction band, and each acceptor will capture an electron from the valence band leaving a defect electron (hole) behind. Really, to be executed, these ionization processes only need the comparably small energy difference $\Delta W_D > W_C - W_D > 0$ respective $\Delta W_A > W_A - W_V > 0$ (see Figure 3.1(b), (c)).

In this way two important phenomena occur. On the one hand, excess electrons and holes will be delivered by ionized donors and acceptors, turning the semiconductor to an n -type or p -type conductor. On the other hand, the ionized positively charged donors and the negatively charged acceptors fixed in the lattice create space charges which have to be considered for the modeling of internal electronics in semiconductor devices. The latter item will be discussed in the next paragraphs in more details.

Normally, at room temperature all donors (densityN_D) and acceptors (concentration N_A) are ionized. Consequently, in an n -type semiconductor doped with donors N_D the electron density is approximately $n = N_D^+ \approx N_D$ where the intrinsic concentration n_i can be neglected ($n_i << N_D$).

For a p -type semiconductor the hole concentration becomes $p = N_A^- \approx N_A$ since $n_i << N_A$ (see Figure 3.8) [2].

3.1.2 Energy-band diagram

As mentioned above electrons and holes in semiconductors exhibit dedicated characteristics, like acceptance of only restricted energy levels, effective mass, etc. Some of them can be illustrated with the band structure of semiconductors and its energy-band diagram. Whereas the band structure (already mentioned above) shows important conclusions found out by quantum-mechanic analysis, generally, an energy-band diagram presents dedicated energy levels (among others the band edges W_C and W_V) against the location described with a position coordinate (x , y or z) of a Cartesian system. To illustrate the increase of electron energy in the energy-band diagram, usually the ordinate with an arbitrary reference level is used for that.

Before going in details, we should mention what the band model's positive features are. As we will demonstrate with examples below, in particular, the energy-band diagram is used to detect where space charges, electric fields and currents will occur within the semiconductor. Moreover, it is very useful to find out the right (internal) boundary conditions needed to calculate the potential and field distribution within the semiconductor device structures. In this way a problem of the complex contact electronics can be defused drastically for modeling purposes. Last but not least, based on it we are able to explain in a phenomenological manner the operating principle of complex semiconductor devices without a detailed analysis of the internal electrons.

Mostly, only the three energy levels (the band edges W_C, W_V and the Fermi-level W_F) as well as the potential energy of a fee unmoved electron (W_e), and additionally, the two energy differences (the electron affinity W_{EA} and work function W_{WF}) are sufficient to complete the energy-band model as demonstrated in the following.

From the band-structure we know that just energy levels near the band edges W_C and W_V are most interesting for the existence of electrons and holes in semiconductors. That is why just theses levels have to be pictured. As mentioned above, from them we find the energy gap $W_g = W_C - W_V$.

If we have doped semiconductors, we still need the position of the so-called Fermi-level W_F in the energy band diagram. The Fermi-edge W_F is an important formal parameter which characterizes the occupation probability of available energy states in the conduction and valence band with electrons or holes (see also paragraph 3.1.3). The position of the energy level W_F in the band model is dependent on the impurity concentration. Normally, the Fermi-level is within the forbidden band (see Figure 3.1). Such semiconductors are said to be non-degenerated. Otherwise, in heavily doped semiconductors ($N_D > 10^{19} cm^{-3}$ or $N_A > 10^{19} cm^{-3}$) the W_F level can dip into the conduction or valence band (degenerated semiconductors).

Characteristically, under thermal equilibrium (that means e.g. without currents in the semiconductor) the Fermi-edge W_F is constant versus the position coordinate throughout the energy-band model (e.g. $\frac{d}{dx} W_F = 0$). In this case, there is only one energy level W_F for electrons and holes. However, in an unbalanced thermo-dynamic state (e.g. with current flow due to supplied voltages) the Fermi-level is split-up into a quasi-Fermi-level for electrons W_{F_n}, and another one for holes W_{F_p}. Normally, both levels vary differently versus the position coordinate.

It should be mentioned, that the energy equivalent $q \cdot V$ of a supplied voltage V has to meet the condition $q \cdot V = W_{F\_n} - W_{F\_p}$.

More details on how to posit W_F in the energy-band diagram, and about W_{F_n} and W_{F_p} as well as their potential equivalents ψ_n and ψ_p will be given in the following paragraphs.

To find out locations where space charges and electric fields dominate the internal electronics in semiconductors; additionally, we need information on changing the electrostatic (macro) potential against the position coordinate. Therefore, the energy equivalent of the electrostatic (macro) potential, i.e. the level of potential energy is also pictured in the band diagram (see Figure 3.2).

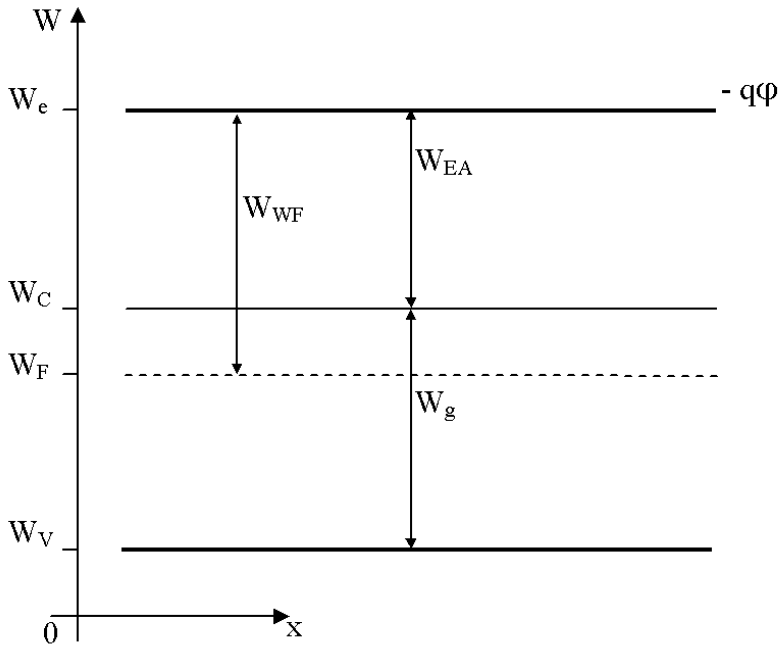


Figure 3.2: Energy-band diagram [1], [2]

Because the energy difference between the energy equivalent of the macro potential $W_e = -q \cdot \varphi > 0$ and the band edges W_C and W_V is only dependent on the crystal characteristics, the potential W_e and the band edges remain in parallel throughout the structure as long as the crystal's characteristics (e.g. the semiconductor material) do not change.

Regarding to W_e, we are able to make visible two essential energy differences, namely, the electron affinity W_{EA} and the work function W_{WF} (see Figure 3.2).

The energy necessary to free an electron from the conduction band edge W_C is called electron affinity W_{EA}. In other words, $W_{EA} = W_e - W_C > 0$, where $W_e = -q \cdot \varphi > 0$ is the potential energy of a resting and free electron outside of crystals.

In particular, for modeling electronic devices in Si technology, we need $W_{EA\_Si} = 4.1\ eV$ and $W_{EA\_SiO_2} = 3.25\ eV$. These energy values are necessary to emit an electron into vacuum and silicon oxide, respectively.

The energy difference between W_e (energy equivalent to the macro potential φ) and W_F (Fermi-level) is called work function W_{WF}, which is a material-specific parameter.

We know, for doped semiconductors W_F is dependent on the impurity concentration, and consequently W_{H_Si}, too. In particular, we should distinguish between $W_{H\_Si\_p}$ and $W_{H\_Si\_n}$ for p - and n -doped semiconductors.

Based on the energy quantities pictured in the energy band diagram and their relative position to each other we are able to distill consequences out of the energy-band diagram for metals, semiconductors and insulators.

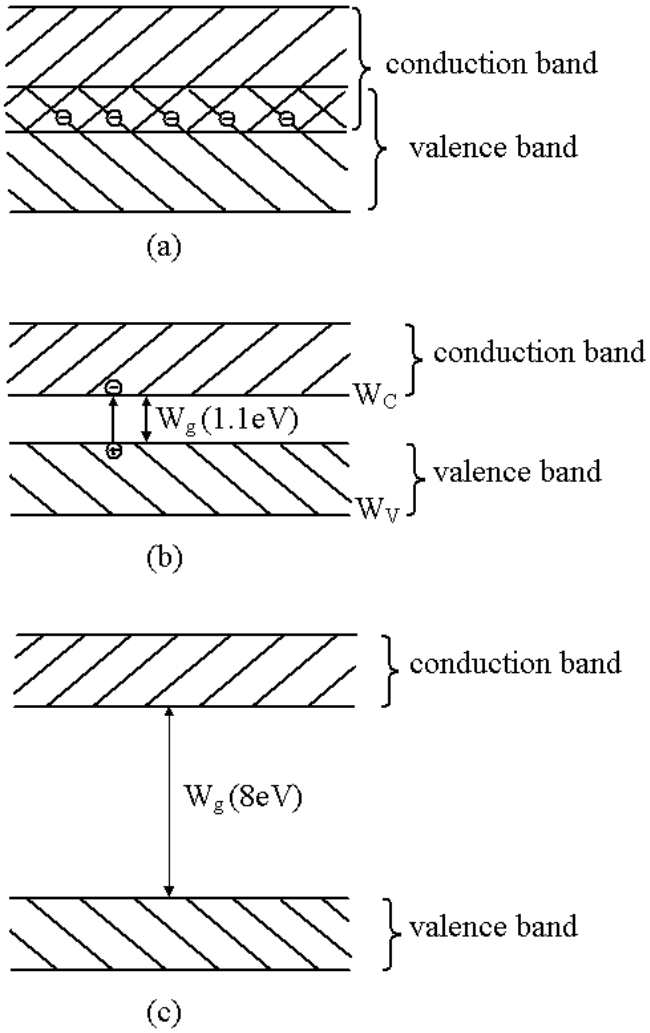


Figure 3.3: Energy band diagram: (a) metal, (b) semiconductor, (c) insulator [2]

As shown in Figure 3.3(a), metals have no forbidden band, because the valence band overlaps the conduction band. Consequently, in metals all electrons will contribute to the comparably high conductivity.

Contrarily, an insulator (see Figure 3.3(c)), is characterized by a large energy gap. Therefore, a relatively high energy is necessary to lift valence band electrons into the conduction band. That is why the thermal generation of carriers is extremely weak and at room temperature pretty unlikely, i.e. no mobile charges are available within the insulator.

The semiconductor (see Figure 3.3(b)) is neither as a good conductor as the metals are, nor an insulator. The particularity exploited for electronic devices is due to its small intrinsic conductivity, which can be prepared goal-oriented by doping.

To meet the expectations and to illustrate the explanation given above, the energy-band diagram of a p - and n -doped semiconductor under thermal equilibrium will be presented at first (see Figure 3.4).

Obviously, the main differences between these energy-band diagrams concern the relative position of the Fermi-level W_F within the forbidden band, and consequently, the work function W_{WF} of the p -and n -type semiconductor.

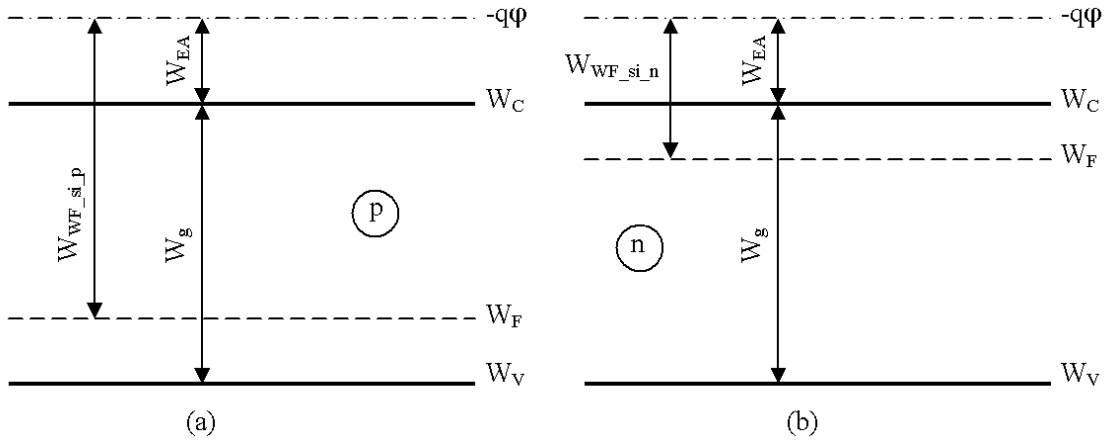


Figure 3.4: Energy band diagram of a p - and n -doped semiconductor [1], [2]

If we physically implement a p - and n -type semiconductor layer in the same crystal, a so-called pn -junction is formed. To construct its energy-band model we have to bring together the band-diagrams given in Figure 3.4(a) and (b), carefully. In doing so we should consider that W_F is constant throughout the structure. In fact, there are no interface and dipole charges at the stoichiometric junction where the conducting type is technologically changed from p -type to n -type or vice verse. That is why the electrostatic potential within the structure may not exhibit any discontinuities. As mentioned above, the energy equivalent to the electrostatic potential has to stay in parallel to the band edges. Consequently, careful combining the energy-band diagrams of a p - and n -type semiconductor layer results in energy-band diagram of a pn -junction as shown in Figure 3.5.

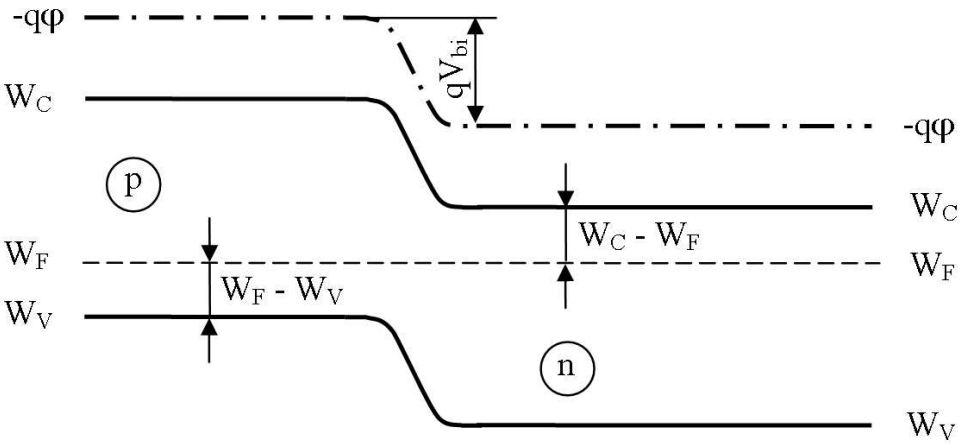


Figure 3.5: Energy-band diagram of a pn -junction [1], [2]

Without any analytic calculations we can learn from it, that within a small region around the junction from a p - to an n -type semiconductor (x_j) the electrostatic potential varies with the x -coordinate. Below (e.g. see paragraph 3.2) we will show, that this fact signalizes an electric field due to space charges in this region. Moreover, caused by this electric field an internal (built-in) potential difference V_{bi} occurs, which has to be considered for determining the potential and field distribution within the structure correctly.

Another important basic structure of semiconductor devices is the MOS-structure shown in Figure 3.6 for a p -doped semiconductor.

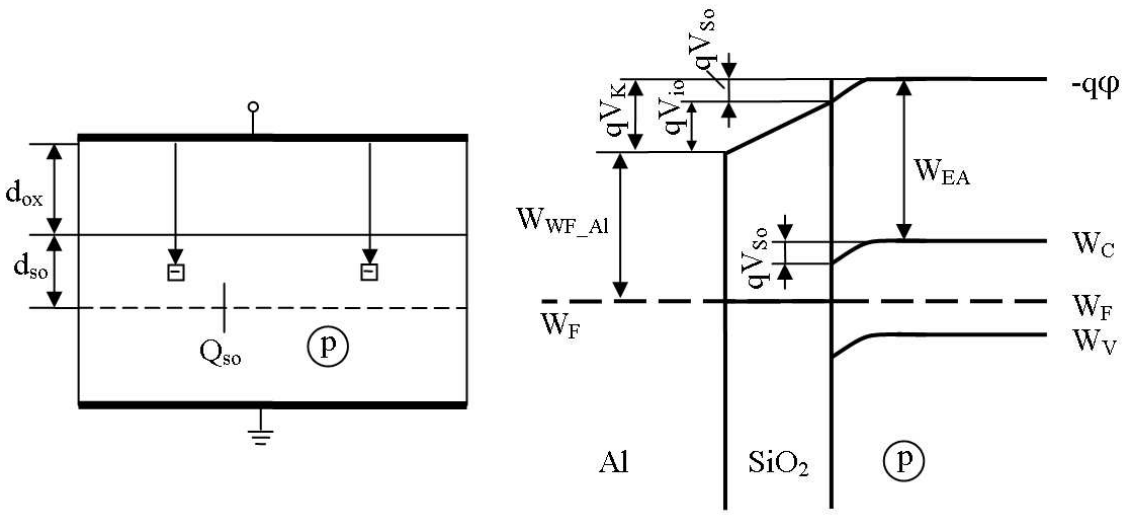


Figure 3.6: Energy-band diagram of a MOS-structure under thermal equilibrium

In our example we assume, that the gate is made from aluminum Al .

To come up with the energy band diagram of such a MOS-structure we need the band model for the gate metal (Al) and the p -doped semiconductor. Simply, the metal's energy-band model is given with the Fermi-level W_F and its work function W_{WF_Al}. Separated by a very thin SiO_2 layer a p -doped semiconductor completes the MOS-structure where the semiconductor's band model is well-known already.

Once again, we have to make sure that under thermal equilibrium conditions W_F is constant throughout the complete structure (see Figure 3.6).

Due to the difference in the work function of Si ($W_{WF\_Si\_p}$) and Al (W_{WF_Al}) we get a linear increase of the potential (potential energy) within the SiO_2 insulator layer. Supposed there are no interface charges located at the interface between the insulator and semiconductor, the electrostatic potential and its slope (the slope corresponds to the electric field as discussed in paragraph 3.2), may not exhibit any kind of discontinuities at the interface. Moreover, deep in the semiconductor, i.e. in a sufficient distance from the interface between oxide and semiconductor, the MOS-structure's energy-band diagram has to agree with that of a p -doped semiconductor. Therefore, we have to accept the band bending as shown in Figure 3.6. Finally, $W_{EA\_SiO_2}$ gives the energy difference to W_C, wherewith the energy-band diagram of the MOS structure can easily be completed, supposed all relevant rules given above are considered carefully.

To sum up what we can learn from this example, we should recognize the following:

Already under thermal equilibrium (i.e. without supplied voltages) we can identify a space charge at the interface between SiO_2 and Si. Roughly speaking, this space charge layer (Q_{s0} with the layer width d_{s0}) is due to complex phenomena of contact electronics and leads to a electric field built-in the oxide and also in the semiconductor. Consequently, a drop of potential across the oxide (V_{i0}) as well as the semiconductor (V_{s0}) will occur. The sum of these voltages is the so-called contact potential V_K, which plays an important role to define the correct boundary conditions for a detailed analysis of the potential and field distribution within the MOS structure with power supply.

3.1.3 Energy and potential dependency of charge carriers

To calculate the energy distribution of electrons and holes in semiconductors, at first, we need the density $D_C(W)$ and $D_V(W)$ of energy levels W in the conduction and valence band available to be occupied with electrons and holes, respectively. Indeed, a quantum-mechanics-based analysis of the semiconductor's band structure results in $D_C(W)$ and $D_V(W)$ with the so-called effective density of states in the conduction and valence band N_C and N_V. These are given by

$$N_C = 2 \cdot \left(\frac{2 \cdot \pi \cdot m_n \cdot k \cdot T}{h^2} \right)^{\frac{3}{2}} \approx 10^{19}\ cm^{-3} \tag{3.1}$$

and

$$N_V = 2 \cdot \left(\frac{2 \cdot \pi \cdot m_p \cdot k \cdot T}{h^2} \right)^{\frac{3}{2}} \approx 10^{19}\ cm^{-3} \tag{3.2}$$

where h is Planck's constant, k is Boltzmann's constant, T stands for the absolute temperature, and m_n, m_p is the effective mass of electrons and holes, respectively.

Secondly, we have to consider the Fermi-Dirac-statistics for the occupation of states with electrons and holes. Generally, the occupation probability $f(W)$ for a state with the energy W by an electron is

$$f(W) = \left(1 + e^{\frac{W - W_F}{k \cdot T}} \right)^{-1} \tag{3.3}$$

where W_F is the well-known Fermi-level (see paragraph 3.1.2). Consequently, for defect electrons (holes) we have to replace $f(W)$ by $1 - f(W)$.

Integrating $D_C(W) \cdot f(W)$ and $D_V(W) \cdot (1 - f(W))$ over the conduction and valance band, respectively, we find the energy distribution of electrons (density n_0) and holes (concentration p_0) under thermal equilibrium conditions. Finally, for non-degenerated semiconductors ($n_0 < N_C$ and $p_0 < N_V$) we get for electrons

$$n_0 = N_C \cdot e^{-\frac{W_C - W_F}{k \cdot T}}, \tag{3.4}$$

and

$$p_0 = N_V \cdot e^{-\frac{W_F - W_V}{k \cdot T}} \tag{3.5}$$

for holes. Using Equations (3.4) and (3.5) we find the intrinsic density n_i from the product

$$n_i^2 = n_0 \cdot p_0 = N_C \cdot N_V \cdot e^{-\frac{W_g}{k \cdot T}} \tag{3.6}$$

The intrinsic density n_i is a typical material parameter with a strong temperature dependency. At room temperature, the intrinsic density is $n_i(300\ K) = 1.5 \times 10^{-10} cm^{-3}$ in Si. For more information see paragraph 3.1.4.

For instance, at room temperatures in a p -type semiconductor is $p_{0p} \approx N_A >> n_{0p} = \frac{n_i^2}{p_{0p}}$. That is why in a p-type semiconductor the holes (p_{0p}) are called the majority carriers, and consequently, the electrons are the minority carriers. In an n -type semiconductor the electrons (n_{0n}) are the majority carriers and the holes (p_{0n}) are in the minority.

In paragraph 3.1.2 we have already learned that the band edges W_C and W_A run in parallel to the electrostatic potential as long as there are no any discontinuities (e.g. dipole-charges would cause potential discontinuities) occur. Moreover, we should remember that a single Fermi-level W_F for electrons and holes makes only sense in thermal equilibrium, i.e. without currents due to supplied voltages. Otherwise, we have to take into account a quasi-Fermi-level W_{F_n} for electrons and W_{F_p} for holes. Their potential equivalents are $W_{F\_n} = -q \cdot \psi_n$ and $W_{F\_p} = -q \cdot \psi_p$. Therewith and under consideration of Equation (3.6) we find the potential representation of the carrier density:

$$n = n_i \cdot e^{\frac{\varphi - \psi_n}{V_T}} \tag{3.7}$$

and

$$p = n_i \cdot e^{\frac{\psi_p - \varphi}{V_T}}. \tag{3.8}$$

where V_T is the so-called thermal voltage. It follows from

$$V_T = \frac{k \cdot T}{q} \qquad V_T\,(300\ K) \approx 25.9\ mV \tag{3.9}$$

As we will see below, the Equation (3.7) and (3.8) play an essential role for modeling the internal electronics of semiconductor devices.

3.1.4 Temperature dependency of the charge carrier density

Typically, semiconductors are strongly dependent on the temperature T. Apart from the electrostatic potential φ, all other quantities of the carrier densities n and p (see Equation (3.7) and(3.8)) are more or less strongly dependent on T. In particular the intrinsic density has an exponential temperature dependency

$$n_i\,(T) = \sqrt{N_C\,(T) \cdot N_V\,(T)} \cdot e^{-\frac{W_g\,(T)}{2 \cdot k \cdot T}}. \tag{3.10}$$

As measured by the exponential increasing of n_i with heating, the temperature influence on the energy gap (forbidden band) $W_g\,(T)$ as well as on the effective density of states in the conduction and valence band $N_C\,(T) \propto T^{3/2}$ and $N_V\,(T) \propto T^{3/2}$ is comparably weak (see Equations (3.1) and (3.2)). To come to acceptable results for our modeling purposes, we should discuss the temperature dependence of electrons and holes in an n -doped semiconductor under thermal equilibrium as given in Figure 3.7.

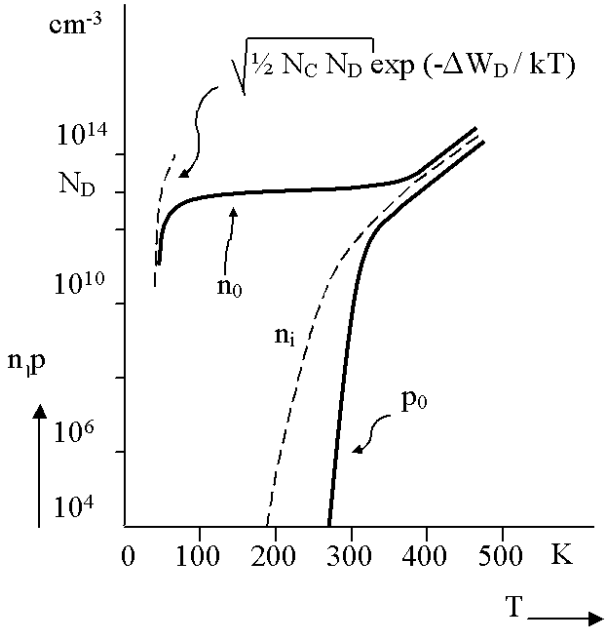


Figure 3.7: Temperature dependency of carriers in an n -doped semiconductor [2]

At very low temperatures the electron concentration n_{0n} is pretty small, however, it increases exponentially with rising temperatures. In this low temperature range (e.g. $T < 100\ K$) a lot of donors has not

been ionized yet, i.e. $p_{0n} << n_{0n} < N_D$ (see Figure 3.8).

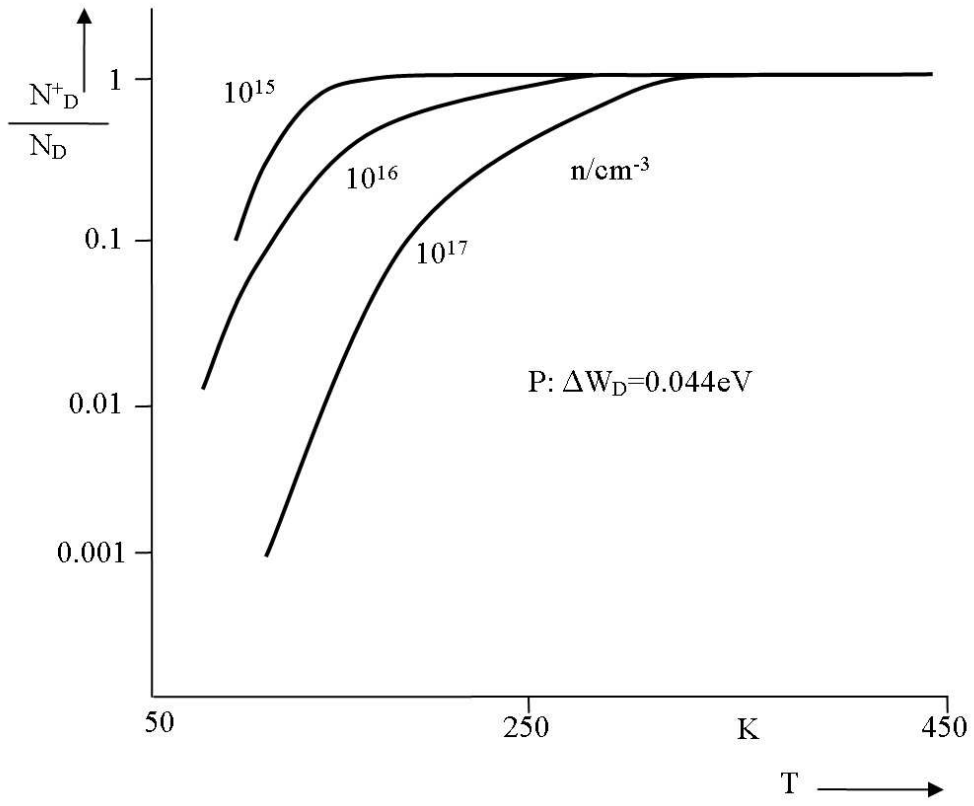


Figure 3.8: Temperature dependency of ionization [2]

At room temperatures the ionization process is completed, and we have $p_{0n} << n_{0n} \approx N_D$. For electronic application, just this temperature range is of greatest interest.

At high temperatures (e.g. $T > 350\ K$) the thermal generation of carrier pairs will dominate, and the semiconductor is increasingly losing its typical difference between minority and majority carriers.

3.1.5 Recombination

If the semiconductor is supplied with "external" energy (e.g. by heating) or it may absorb photons with an energy $h \cdot f > W_g$ (f: photon's frequency) and an appropriate impulse, valence band electrons will be lifted into the conduction band. Thus, pairs of electrons and holes will be generated and characterized by the rate G. This rate G describes the number of generated electron-hole pairs per unit time and volume. As generation processes play an important role in advanced electronic devices (laser, light emitting diodes (LED)), the opposite process to generation, called recombination appears in all semiconductor-based devices more or less.

The recombination rate R gives the number of electron-hole pairs recombined (disappeared) per unit time and volume. Both mechanisms the generation and the recombination are schematically illustrated in Figure 3.9.

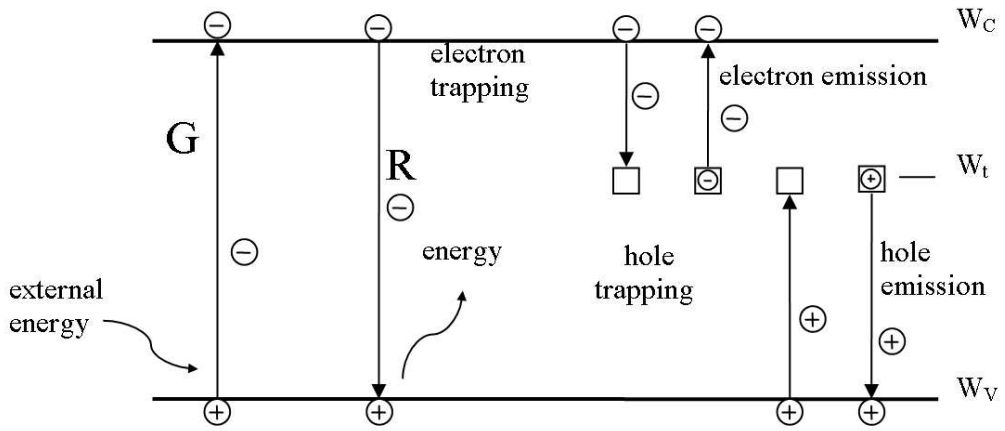


Figure 3.9: Generation and recombination in semiconductors [2]

Principally, there is a wide-variety of recombination mechanisms. However, at the moment we are only interested in the so-called indirect recombination performed via energy states within the band gap. Under steady state conditions electrons and holes are trapped at the same rate by such energy states (traps).

If anywhere in the semiconductor the carrier densities differ from their concentration under thermal equilibrium ($\Delta n = n - n_0$ and $\Delta p = p - p_0$), and moreover, these derivations are sufficiently small ($\Delta n << n_0$ and $\Delta p << p_0$) we obtain

$$R = \frac{\Delta n}{\tau_n} = \frac{\Delta p}{\tau_p} \tag{3.11}$$

for the recombination rate where τ_n and τ_p are the averaged electron and hole lifetime until to their recombination. The lifetime is a very complex parameter affected by the semiconductor material, the doping, the temperature and last but not least from the carrier concentration. For estimation purposes we use $\tau_n \approx 10^{-9}s$ and $\tau_p \approx 10^{-7}s$.

Self-Assessment Questions

Question 3.1.1:

What type of impurities has to be used to make an n-type semiconductor, donors or acceptors?

Question 3.1.2:

What energy levels are necessarily needed to pivture an energy-band diagram of a p-type semiconductor under thermal equilibrium conditions?

Question 3.1.3:

What is the meaning of the quantities n, n_i, φ, ψ_n and V_T of $n = n_i \cdot e^{\frac{\varphi - \psi_n}{V_T}}$?

Question 3.1.4:

What kind of termerature dependency is essential for the intrinsic density n_i?

3.2 Electric field and space charges in semiconductors

Basically, the operation principles of most electronic devices are mainly based on mobile and fixed charges as well as currents due to moved charge carries. To move mobile carriers in semiconductors, basically, two reasons have to be considered, the electric field and the diffusion, respectively. Generally, the electric and magnetic fields have also to be taken into account if we intend to model electronic device's functionality

on phenomenological basics. Therefore, the modeling process needs considering Maxwell's equations. Application-driven, we are particularly interested in solving modeling problems where the time rate of changing the magnetic field is not too high. That means we may assume that the electric field $\vec{E}$ does not have curls caused by fast varying the magnetic field with the time, i.e.

$$curl\ \vec{E} = \vec{0}. \tag{3.12}$$

In this time regime the vector $\vec{E}$ can be derived from an electrostatic potential field $\varphi(\vec{r})$ by the space derivation with respect to the position vector $\vec{r} = \vec{i} \cdot x + \vec{j} \cdot y + \vec{k} \cdot z$:

$$\vec{E} = -grad\ \varphi = -\left(\vec{i} \cdot \frac{\partial}{\partial x}\varphi + \vec{j} \cdot \frac{\partial}{\partial x}\varphi + \vec{k} \cdot \frac{\partial}{\partial z}\varphi\right). \tag{3.13}$$

$\vec{i}$, $\vec{j}$ and $\vec{k}$ are the unit vectors to the space coordinates of a Cartesian coordinate system.

Supposed the electric field $\vec{E}$ is curl-free, generally, it originates from and terminates on charges. Therefore, charges are said to be the sources of the electrostatic fields. For instance, it can be due to space charges and also charges on interface with material discontinuities as well as on electrodes. So, an electric field is built-up between electrodes supplied with different potentials. In this case, $\vec{E}$ originates from positive surface charges on the electrode connected to the positive pole of the power supply and terminates on the negative surface charges on the electrode connected to the negative pole. To detect the sources of electric fields in a space, Maxwell's system prescribes the vector operation divergence:

$$div\,\vec{D} = \frac{\partial}{\partial x}\left(\vec{D} \cdot \vec{i}\right) + \frac{\partial}{\partial x}\left(\vec{D} \cdot \vec{j}\right) + \frac{\partial}{\partial z}\left(\vec{D} \cdot \vec{k}\right) = \rho \tag{3.14}$$

where $\vec{D}$ denotes the vector of electric displacement and ρ is the space charge density.

Equation (3.14) is the differential form of Gauss' low

$$\oint_{(A)} \vec{D} \cdot d\vec{A} = \int_{(V)} div\vec{D} \cdot dV = \int_{(V)} \rho \cdot dV. \tag{3.15}$$

Basically, the vector element $d\vec{A}$ of the closed surface A with the interior volume V is directed outwards. The vectors $\vec{E}$ and $\vec{D}$ are not independent from each other, in fact, it yields

$$\vec{D} = \varepsilon \cdot \vec{E} = \varepsilon_0 \cdot \varepsilon_r \cdot \vec{E}. \tag{3.16}$$

where ε is the material's permittivity with the absolute dielectric constant $\varepsilon_0 = 8.86 \times 10^{-12} As/Vm$ of the empty space and the relative dielectric constant ε_r typical for the material's dielectric characteristic.

For Si we use the permittivity constant $\varepsilon_H = 10^{-12} As/Vcm$, whereas the dielectric constant for silicon oxide (SiO_2) is given by $\varepsilon_{ox} = 3 \times 10^{-13} As/Vcm$.

Generally, ε is a tensor, but in most applications we use the material's permittivity as scalar parameter.

Inserting Equation (3.16) with Equation (3.13) in Equation (3.14) we obtain

$$div\,\vec{D} = -\varepsilon \cdot div\,(grad\ \varphi) = -\varepsilon \cdot \left(\frac{\partial^2}{\partial x^2}\varphi + \frac{\partial^2}{\partial y^2}\varphi + \frac{\partial^2}{\partial z^2}\varphi\right) = \rho \tag{3.17}$$

and finally, from Equation (3.17) Poisson's equation in its well-known fashion:

$$\Delta\varphi = \frac{\partial^2}{\partial x^2}\varphi + \frac{\partial^2}{\partial y^2}\varphi + \frac{\partial^2}{\partial z^2}\varphi = -\frac{\rho}{\varepsilon}. \tag{3.18}$$

To be solved application-specifically, this second order partial differential equation needs boundary conditions, i.e. so-called boundary problems have to be defined. Depending on the kind of boundary conditions

to be met, basically, we distinguish between three basic boundary problems. To tackle these tasks is the main goal of the potential theory in mathematics. Actually, most of our modeling approaches demand solutions of the Dirichlet's problems where the potential on the boundary has to be specified before solving the Poisson's equation. Basically, the Dirichlet's problem can formally be solved by Green's function, which is typical for the geometry of the field space surrounded with boundaries provided with specified potentials.

Unfortunately, this approach does not meet our modeling demands sufficiently because the Green's function is unknown for most practical applications, and typically, the boundary problems for dedicated modeling problems in electronics require additional assumptions. In particular, there is a remarkable difference to electrostatics where the space charges and boundary potentials are specified and fixed before solving Poisson's equation. In our modeling tasks, often, we have to involve the strong interaction between charges and electric field. Because of charge neutrality, usually, no space charges appear before applying the electric field. Imposing an electric field on a semiconductor, space charges are coming out, and will interact with this field. For instance, this kind of modeling problems has to be solved for determining the width of the space charge layer of a *pn* -junction (e.g. see paragraph 4). Moreover, if mobile charges ought to be included in the modeling procedure, in many cases we have to consider their exponential dependency in terms of the potential φ to be determined. In such cases the problems become even nonlinear, and we have to find out physically-founded additional conditions to make these problems analytically solvable. Anyway, the calculation of electronic fields in semiconductor structures is really more complex than in the classic electrostatics. Below, we will demonstrate our approach on how to calculate the field and potential distribution in a *pn* -junction and a MOS structure, respectively.

Self-Assessment Question

Question 3.2:

What is the Poisson's differential equation needed for?

3.3 Charge carrier transport in semiconductors

3.3.0 Summary

Basically, the currents in semiconductors are composed of two components, namely, the convection current resulting from moved carriers, and the displacement current due to the time-variable changing the displacement vector field $\vec{D}$. Apart from the dynamics of the mobile carrier's assembling and rebuilding processes at very high frequencies (RF-range), mostly the entire device electronics can be analyzed under neglecting the displacement current. Indeed, holes and electrons can contribute to the current in semiconductors. Each of these carriers may be moved by an electric field and/or by diffusion due to a density gradient of the carriers. That is why we have to consider two transport equations (holes and electrons), and each with both current components, the field drift part and the diffusion component.

The electric field interacts with all charges in the semiconductor. Therefore, we have to take into account as space charges as the charge carriers. That makes the analysis non-linear and highly complex.

Based on Maxwell's equations only the total current density must be sources free. Consequently, this constrain has not to be applied to the electron and hole current. Therefore, for these carrier currents we may accept sources (carrier generation) and sinks (carrier recombination), if the carrier generation and recombination processes take place in pairs. This approach leads to two continuity equations (one for holes and one for electrons) which make sure the Maxwell's requirements.

The following paragraph is aimed at a phenomenological description of the carrier transport in semiconductors included interactions between carrier density and the electric field and electrostatic potential, respectively.

3.3.1 Field drift and diffusion currents

The carrier (or convection) current density $\vec{S}$ in semiconductors arises from mobile space charges of holes with the concentration $\rho_p = q \cdot p$ and electrons with the density $\rho_n = -q \cdot n$ moved with the velocities $\vec{v}_P$ and $\vec{v}_N$, respectively. The density vector $\vec{S}$ is given by

$$\vec{S} = \vec{S}_p + \vec{S}_n = \rho_p \cdot \vec{v}_P + \rho_n \cdot \vec{v}_N . \tag{3.19}$$

Generally, drift fields and/or diffusion processes are the reason for carrier moving velocities. Whereas the drift components (carrier velocity $\vec{v}_p$ or $\vec{v}_n$) come from an electric drift field $\vec{E}$, the diffusion parts (carrier velocities $\vec{v}_{p\_D}$ and $\vec{v}_{n\_D}$) originate from carrier density gradients (*grad p* and *grad n*).

For our modeling approaches we are especially interested in carrier drift due to relatively weak electric fields. At low electric fields the drift velocities $\vec{v}_p$ and $\vec{v}_n$ are proportional to the field $\vec{E}$:

$$\vec{v}_p = \mu_p \cdot \vec{E} \tag{3.20}$$

and

$$\vec{v}_n = -\mu_n \cdot \vec{E}. \tag{3.21}$$

The mobility μ_n for electrons and μ_p of holes is dependent on the temperature T and the total impurity concentration as shown in Figure 3.10, 3.11 and Figure 3.12.

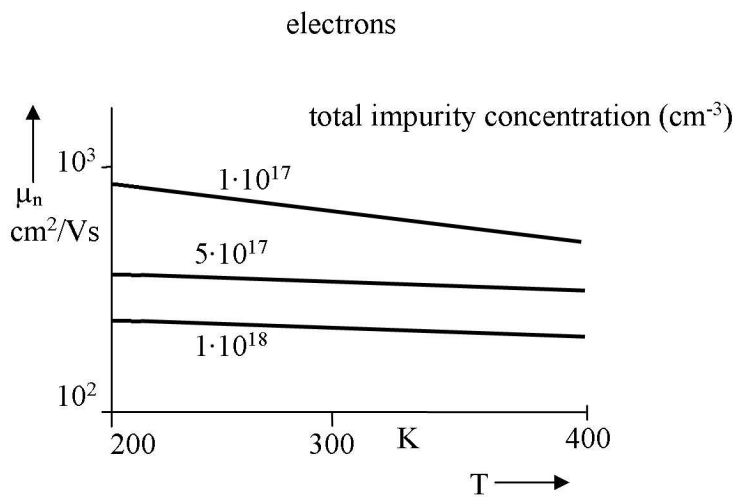


Figure 3.10: μ_n 's temperature dependency in Si [2]

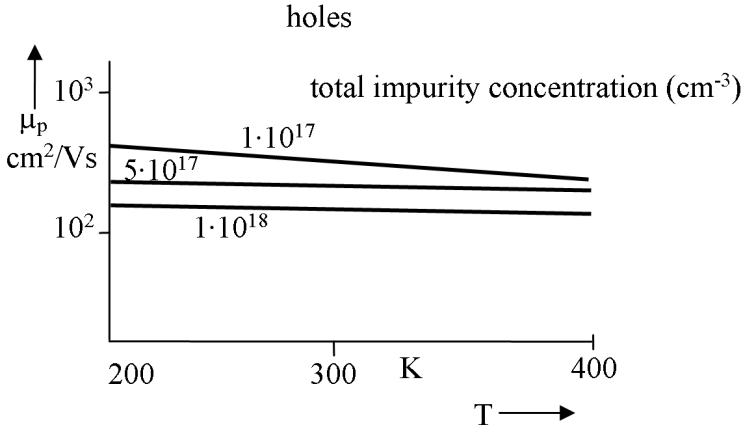


Figure 3.11: μ_p 's temperature dependency in Si [2]

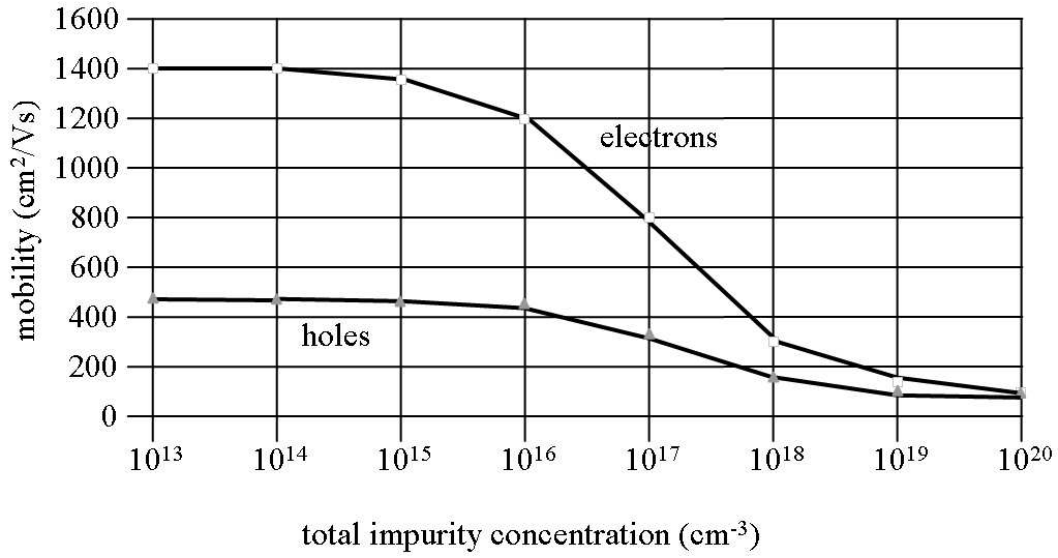


Figure 3.12: μ_n and μ_p in dependency on the total impurity concentration [5]

Because of the scattering mechanisms, the carrier mobilities decrease also at high electric fields. Therefore, the drift velocity of electrons and holes saturates, and the simple mobility models given by Equation (3.20) and Equation (3.21) need qualifying.

The second velocity components $\vec{v}_{p\_D}$ and $\vec{v}_{n\_D}$ are due to diffusion, and are given by

$$\vec{v}_{p_D} = -D_p \cdot \frac{grad\ p}{p} = -D_p \cdot grad\ \ln \frac{p}{n_i} \tag{3.22}$$

and

$$\vec{v}_{n_D} = -D_n \cdot \frac{grad\ n}{n} = -D_n \cdot grad\ \ln \frac{n}{n_i} \tag{3.23}$$

D_p and D_n are the diffusion coefficient for holes and electrons, respectively. For medium-doped silicon we have $D_p \approx 10\ cm^2/s$ and $D_n \approx 40\ cm^2/s$.

Using Equations (3.19) to (3.23) we obtain the transport equation for holes $\vec{S}_p$ and electrons $\vec{S}_n$:

$$\vec{S}_p = \rho_p \cdot \vec{v}_P = \rho_p \cdot (\vec{v}_p + \vec{v}_{p_D}) = q \cdot \left(\mu_p \cdot p \cdot \vec{E} - D_p \cdot grad\ p\right) \tag{3.24}$$

and

$$\vec{S}_n = \rho_n \cdot \vec{v}_N = \rho_n \cdot (\vec{v}_n + \vec{v}_{n_D}) = q \cdot \left(\mu_n \cdot n \cdot \vec{E} + D_n \cdot grad\, n\right) \tag{3.25}$$

If the density of mobile carriers is very high (e.g. in metallic conductors) only a relatively small electric field is needed to transport any demanded current. In such cases, the diffusion contributions to the total convection current can be neglected, and we find from Equation (3.24) and Equation (3.25)

$$\vec{S} = \vec{S}_p + \vec{S}_n = \rho_p \cdot \vec{v}_p + \rho_n \cdot \vec{v}_n = q \cdot (\mu_n \cdot n + \mu_p \cdot p) \cdot \vec{E} = \kappa \cdot \vec{E} \tag{3.26}$$

where κ is the conductor's conductivity given by

$$\kappa = q \cdot (\mu_n \cdot n + \mu_p \cdot p)\,. \tag{3.27}$$

Supposed the carrier densities and their mobilities are independent on the electric field, the right part of Equation (3.26) can be interpreted as "Ohm's law" for semiconductors whose conductivity is majority carrier controlled.

In general, the carrier mobility and the diffusion coefficient for holes and electrons are not independent from each other. In fact, the Einstein's relation yields

$$\frac{D_p}{\mu_p} = \frac{D_n}{\mu_n} = V_T = \frac{k \cdot T}{q} \tag{3.28}$$

where V_T is the thermal voltage well-known from Equation (3.9).

Using Equation (3.28) and $\vec{E} = -grad\ \varphi$ (see Equation (3.13)) the drift and diffusion part of the hole and electron current density can be combined in a single equation for each. Let us take the hole current as an example. We find

$$\vec{S}_p = -q \cdot \mu_p \cdot p \cdot grad \left(\varphi + V_T \cdot \ln \frac{p}{n_i}\right) = -q \cdot \mu_p \cdot p \cdot grad\ \psi_p \tag{3.29}$$

In an analog way we determine for the electron current density

$$\vec{S}_n = -q \cdot \mu_n \cdot n \cdot grad \left(\varphi - V_T \cdot \ln \frac{n}{n_i}\right) = -q \cdot \mu_n \cdot n \cdot grad\ \psi_n \tag{3.30}$$

where

$$\psi_p = \varphi + V_T \cdot \ln \frac{p}{n_i} \tag{3.31}$$

and

$$\psi_n = \varphi - V_T \cdot \ln \frac{n}{n_i} \tag{3.32}$$

is the quasi-Fermi potential for holes and electrons, respectively.

Basically, from Equation (3.29) and Equation (3.30) we learn that current flow compels the variation of the quasi-Fermi potentials along the local coordinates. Without any current is

$$\psi_p = \psi_n = \Psi_F \tag{3.33}$$

where Ψ_F is the potential equivalent of the Fermi-level.

From Equation (3.31) and Equation (3.32) we extract the hole and electron concentration in a semiconductor with an electric field and a carrier density gradient inside. We obtain the Equations (3.7) and (3.8)

The carrier's dependency on the electrostatic potential φ as well as the quasi-Fermi potential ψ_p, ψ_n is extremely helpful in analyzing the field and potential distribution in semiconductor structures by Poisson's equation.

3.3.2 Carrier flow and the continuity equations

From Maxwell's theory we know, the total current (density $\vec{S}_t$) is composed of a convection current part (density$\vec{S}$) and the displacement current (density$\vec{S}_D$)

$$\vec{S}_t = \vec{S} + \vec{S}_D. \tag{3.34}$$

Whereas $\vec{S}$ describes the charge transport by moved carriers (electron and holes), the displacement current is due to the time-variable changing the displacement vector $\frac{\partial}{\partial t}\vec{D}$. Based on Maxwell's equations, the total current $\vec{S}_t$ has to be free of sources. Using the divergence operator div () and considering the Equations (3.19) and (3.14) as well as the space charges for holes $\rho_p = q \cdot p$ and electrons $\rho_n = -q \cdot n$, formally, we get

$$div\ \vec{S}_t = div\ \vec{S} + div\ \vec{S}_D = div\ \left(\vec{S}_p + \vec{S}_n\right) + \frac{\partial}{\partial t}\left(\rho_p + \rho_n\right) = 0. \tag{3.35}$$

Principally, in semiconductors we have to take generation and recombination processes into account. As mentioned in paragraph 3.1.5, these mechanisms are executed pair-wise. From the phenomenological point of view, these processes will have an influence on the electron and hole current, however, they may not break Maxwell's equation on the continuity of $\vec{S}_t$, i.e. $div\ \vec{S}_t = 0$. Therewith, we can derive from Equation (3.35) the so-called continuity equations for the electron and hole current. Finally, we achieve

$$div\ \vec{S}_n = -q \cdot (G - R) + q\frac{\partial}{\partial t}n \tag{3.36}$$

for the electron current, and

$$div\ \vec{S}_p = q \cdot (G - R) - q \cdot \frac{\partial}{\partial t}p. \tag{3.37}$$

for the hole current where G and R stand for the generation and recombination rate.

With appropriate models for generation and recombination (e.g. see Equation(3.11)), and using the transport equations (see Equations (3.24) and (3.25)) as well as the field distribution (e.g. derived from the solution of Poisson's equation), the continuity equations (3.36) and (3.37) describe the dynamics (assembly and rebuilding) of the carrier distributions in semiconductors. We will use them to model the internal electronics, in particular the minority carrier distribution within bipolar devices.

Self-Assessment Questions

Q 3.3.1:

Which phenomena makes the electrons and holes move within a semivonductor?

Q 3.3.2:

Which components of carrier currents have to be considered in semiconductors?

Module 4

pn-Junction and Diodes

Module units

Module overview. In fact, the *pn* -junction is the most important elementary structure in bipolar semiconductor devices. In particular, it is used as the basic structure to implement the operation principles of diodes, *pnp* - and *npn* -transistors and relative semiconductor devices. Moreover, *pn* -junctions also play an important role in MOS-transistors (e.g. to contact the channel current at source and drain) as well as to decouple integrated devices in microelectronics (IC). Some of these applications will be discussed in detail in this course module.

From the didactics point of view, the learners are instructed to deal creatively with the approaches and facts imparted in previous modules. Moreover, some new ideas and thoughts are added, too.

Module objectives. *Among others, this module is intended to answer the questions:*

1. What happens at the junction where a p - and an n-type conducting region implemented in one single crystal get in touch?

2. Why does a so-called space charge layer appear at the stoichoimentic *pn* -junction?

3. What role plays the so-called built-in potential?

4. How to control the space charge layer and the space charge capacity?

5. What phenomenon is behind a minority carrier injection current?

6. What effects and phenomena dominate the dynamics of diodes in circuits?

7. How is the appropriate network models applied correctly?

Module prerequisites. knowledge, given in module 3, basics in electronics and electrical engineering

4.1 Basics of the operation principle

4.1.0 Summary

The key to understand the internal electronics of pn -junctions and diodes is to be acquainted with the field and potential distribution within these structures. Generally, this problem will be tackled by solving a boundary problem to Poisson's equation (see paragraph 3.2). For our purpose we may simplify this task by

- Taking into account only the ionized impurities within a depletion layer at the stoichiometric pn -junction.

- One dimensional (1D-)analysis regarding the position coordinates.

Additionally, we will show the main difference to a similar boundary problem of electrostatics.

Finally, we will deal with approaches aimed at calculating the space charge and minority carrier charges and the corresponding capacities controlled by the supplied voltage.

4.1.1 Field and potential distribution in diodes

The basic technological concept behind setting up pn -junctions is to diffuse or implant impurities of acceptor- or donor-type into an n - or p -type pre-doped semiconductor, respectively. Figure 4.1 shows a pn -junction made by diffusion of acceptors from the surface (acceptor's surface concentration $N_{A0} \approx 10^{18} cm^{-3}$) into the n -type semiconductor pre-doped with donor concentration $N_D \approx 10^{16} cm^{-3}$.

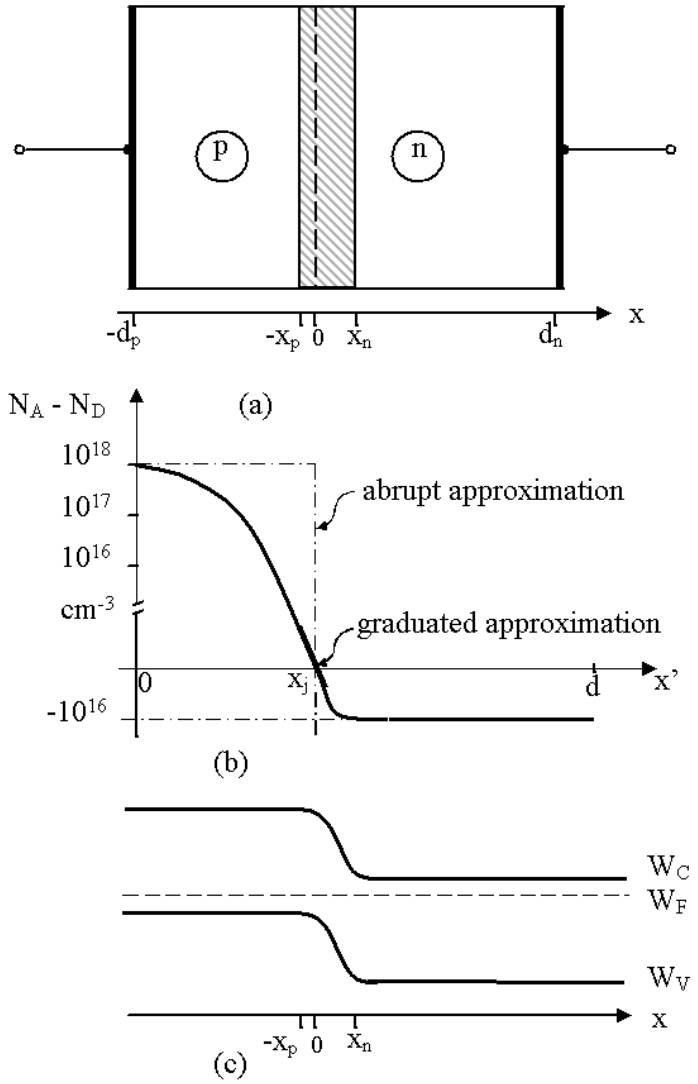


Figure 4.1: *pn* -junction: (a) structure, (b) impurity profile, (c) energy-band diagram [2]

So we get a p - and n -type region in the same crystal. The stoichiometric junction is obtained at x_j where the acceptor and donor impurity concentration are equal (see Figure 4.1). Near this junction from p - to n -type (or vice versa) enormous carrier densities gradients occur. For instance, in a p -type semiconductor the holes are the majority carriers, and consequently, the electrons are the minorities there. Logically, in an n -type semiconductor the opposite is true. That is why enormous differences in carrier concentration have to be balanced within the depletion layer by diffusion. Figure 4.2 illustrates these differences of carrier concentrations.

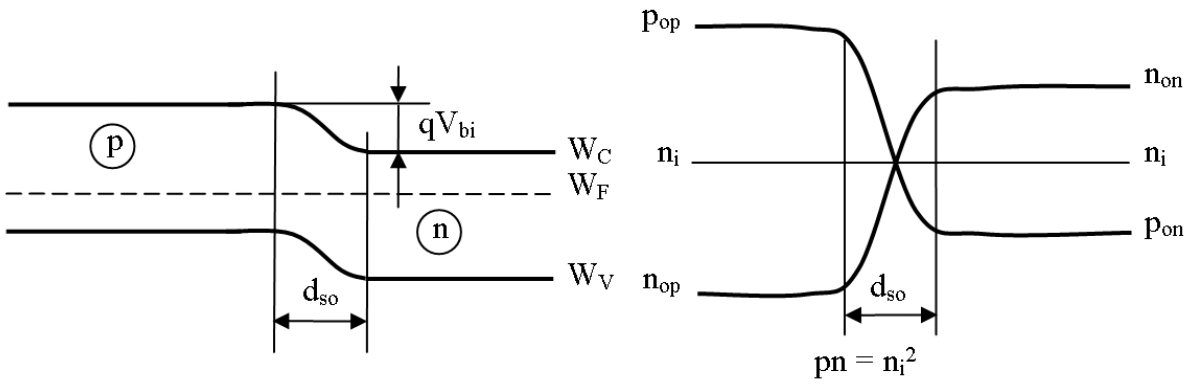


Figure 4.2: Distribution of carrier concentration across a *pn* -junction (schematically) and the energy-band diagram [1]

Simultaneously with carrier diffusion across the *pn* -junction, the ionized impurities (fixed in the crystal lattice) form a space charge layer at the junction. As discussed in paragraph 3.2, this space charge is the reason for the electric fields, which gives the necessary balance to the carrier diffusion. A balanced state is achieved, if the diffusion currents and the drift currents due to the electric field are equal.

Remember, as we depicted the energy-band diagram of a *pn* -junction in paragraph 3.1.2 (compare Figure 3.5 with Figure 4.1 and Figure 4.2), we have clearly detected that a space charge layer with an appropriate electric field has to build up at the *pn* -junction.

Now, we are going to analyze the field and potential distribution based on the so-called impurity approach by integrating Poisson's equation.

The starting point of our analysis is the space charge distribution within a space charge layer in $-x_p \leq x \leq x_n$ with the width $d_s = x_n + x_p$ as shown in Figure 4.3.
The space charge in this layer is modeled by considering ionized acceptors $(-q \cdot N_A)$ and donors $(q \cdot N_D)$, only. The neglecting of mobile carries in the total space charge is called impurity approach. Therefore, the space charge ρ is

$$\rho = q \cdot \begin{cases} -N_A & -x_p \leq x \leq -0 \\ \\ N_D & +0 \leq x \leq x_n \end{cases} \tag{4.1}$$

The p - and n -regions between the contacts at $x_{c_p} = -d_p$ and $x_{c_n} = d_n$, and outside of the space charge layer are electrically neutral, i.e. $\rho = 0$.

Basically, the field and potential distribution can be derived by the twofold integration of a 1D-Poisson's equation with the space charge defined in Equation (4.1)

$$\frac{d}{dx} E\left(-x_p \leq x \leq x_n\right) = -\frac{d^2}{dx^2} \varphi\left(-x_p \leq x \leq x_n\right) = \frac{\rho}{\varepsilon_H} \tag{4.2}$$

ε_H is the permittivity of the semiconductor, e.g. $\varepsilon_H = 10^{-12} As/Vcm$ for silicon.

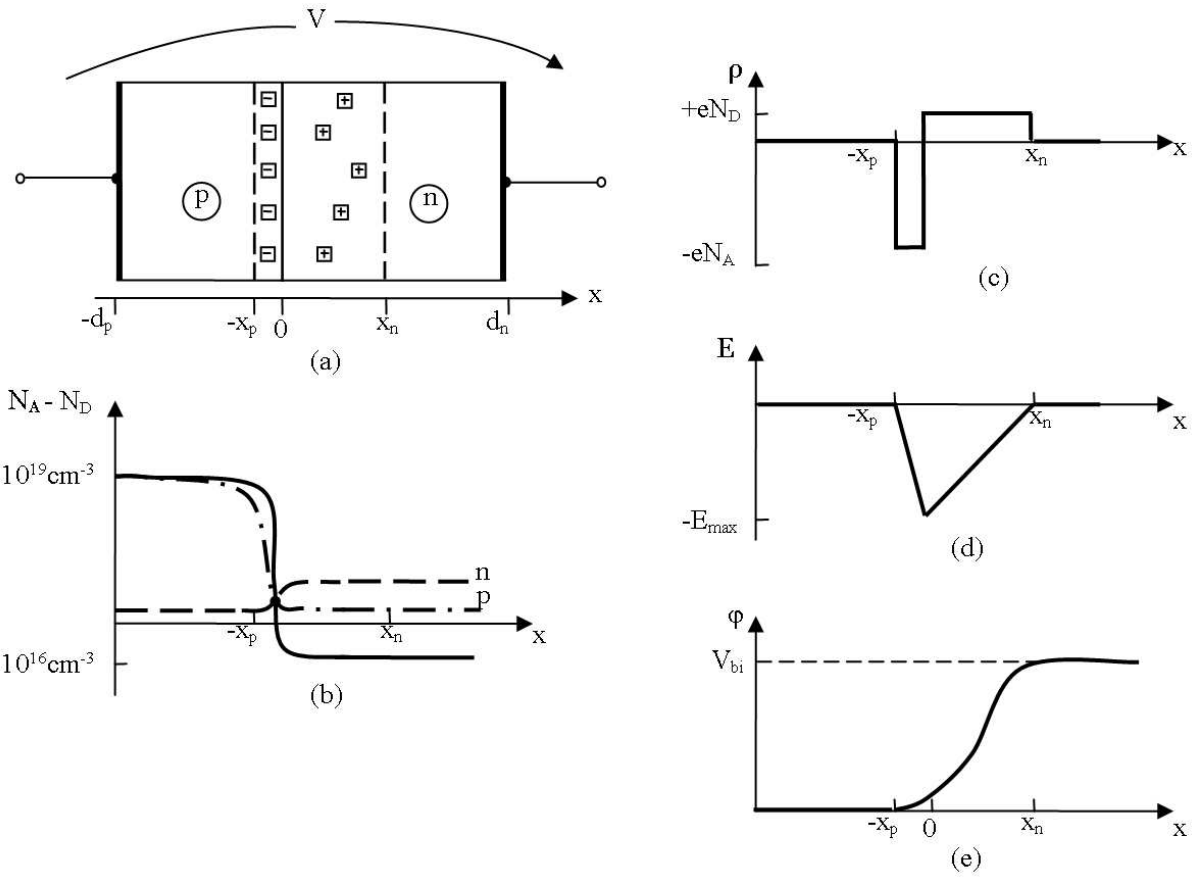


Figure 4.3: *pn* -junction: (a) cross section, (b) total impurity ($N_A - N_D$) and carrier density (p, n) distribution, (c) space charge distribution, (d) electric field, (e) potential distribution [2]

However, we are only interested in a solution adapted to the boundary conditions $\varphi(-x_p)$ and $\varphi(x_n)$. Actually, what we know are the potentials supplied to the contacts $\varphi\left(x_{c_p} = -d_p\right)$ and $\varphi\left(x_{c_n} = d_n\right)$. However, these potentials do not agree with $\varphi(-x_p)$ and $\varphi(x_n)$ demanded as boundary conditions for Equation(4.2). As mentioned before, we are free to define the reference potential ($\varphi = 0$). For instance, we will fix it at the left-hand side x_p of the space charge layer. So we receive the first boundary condition

$$\varphi(-x_p) = 0. \tag{4.3}$$

If we intend to supply a positive voltage $V = \varphi(-d_p) - \varphi(d_n) \geq 0$ to the contacts of the diode (*pn* -junction structure), we have to provide $\varphi(d_n) = -V$ at the cathode contact ($x_{c_n} = d_n$). To determine the second boundary condition $\varphi(x_n)$ we have to consider the following facts and insights learned in the energy-band discussion before (e.g. see 3.1.2):

First of all, we have to make sure the space charge neutrality in the neutral p - and n -type regions included the electrodes. There we define $p(-d_p) \approx N_A$ and $n(d_n) \approx N_D$.

Secondly, we should consider carefully the quasi-Fermi potentials which play an essential role, when a semiconductor is supplied with voltages where currents can flow. In such cases, the semiconductor structure does not hold the thermal equilibrium conditions anymore.

Particularly, the quasi-Fermi potentials ψ_p and ψ_n are assumed to be constant within the neutral p - and n -regions. This fact can be accepted because the current is transported by the majority carriers only. Consequently, only a negligible gradient of ψ_p and ψ_n is necessary to form any current as demanded (see Equations (3.29) and (3.30)).

To determine the second boundary condition for solving the Poisson's equation (4.2), we should remember that a supplied voltage gives a difference in the quasi-Fermi potentials.

In paragraph 3.1.2 we already stated $q \cdot V = W_{F_n} - W_{F_p}$. Using the equivalence between energy and potential, we achieve

$$\psi_p - \psi_n = V > 0. \tag{4.4}$$

Now, we are able to construct the missing boundary condition from Equation (3.31) and Equation(3.32). Finally, we obtain the second boundary condition

$$\varphi(x_n) = V_T \cdot \ln \frac{N_D \cdot N_A}{n_i^2} - (\psi_p - \psi_n) = V_{bi} - V \tag{4.5}$$

where V_{bi} is the so-called built-in potential (voltage) defined by

$$V_{bi} = V_T \cdot \ln \frac{N_D \cdot N_A}{n_i^2}. \tag{4.6}$$

V_{bi} represents the internal drop of potential across the space charge layer at the *pn* -junction (see Figure 4.2). Please note, it is not dependent on supplied voltages, however, V_{bi} is given with the doping concentrations and the intrinsic density n_i. We should mention that V_{bi} cannot be measured, because the contact potentials at the electrodes compensate the built-in potential V_{bi}, exactly. Otherwise, a *pn* -junction would supply energy due to V_{bi}. The built-in potential plays a similar role as the contact voltage V_K (see paragraph 3.1.2) of a MOS-structure.

The Poisson's equation (4.2) complemented with the two boundary conditions (see Equations (3.16) with (3.17)) defines a boundary value tasks, which can only be solved if the space charge layer was fixed. Unfortunately, that is not applicable in this case, because the width $d_s = x_n + x_p$ of the space charge layer is modulated by the applied voltage V . That is why we have to determine how x_p and x_n depend on V. For that purpose, we need two additional conditions.

Appropriate to the dedicated space charge in the semiconductor (see Equation (4.1) and $\rho = 0$ elsewhere) we may define that an electric field $E(-d_p \le x \le d_n) \cdot \vec{i} \neq \vec{0}$ exists only within the space charge layer $-x_p < x < x_n$, and consequently, the rest of the semiconductor is field-free:

$$E(-d_p \le x \le -x_p) = E(x_n \le x \le d_n) = 0. \tag{4.7}$$

Therewith, we have the two additional conditions to fix the position of the space charge layer's borders x_p and x_n in dependence on V.

At the stoichiometric junction at $x = 0$ where the space charge abruptly jumps from a negative to positive value, there are neither surface charges nor dipole charges. So, we have to take care for the consistency of the field and potential distribution. Now, the defined boundary value problem can be solved elementarily. Here are the outcomes (see Figure 4.3):

$$E(x) = -\frac{q}{\varepsilon_H} \cdot \begin{cases} N_A \cdot (x + x_p) & -x_p \le x \le 0 \\ N_D \cdot (x_n - x) & 0 < x \le x_n. \\ 0 & elsewhere \end{cases} \tag{4.8}$$

Within the space charge layer the electric field is linearly changing with $x$as shown in Figure 4.3. The space charge layer's borders are given by

$$x_p = \sqrt{\frac{2 \cdot \varepsilon_H \cdot N_D}{q \cdot N_A \cdot (N_D + N_A)} \cdot (V_{bi} - V)} \tag{4.9}$$

and

$$x_n = \sqrt{\frac{2 \cdot \varepsilon_H \cdot N_A}{q \cdot N_D \cdot (N_D + N_A)} \cdot (V_{bi} - V)}. \tag{4.10}$$

Consequently, the total width of the space charge $d_s = x_n + x_p$ is also dependent on the supplied voltage. Finally, we achieve the space charge layer modulation from

$$d_s(V) = \sqrt{\frac{2 \cdot \varepsilon_H \cdot (N_D + N_A)}{q \cdot N_A \cdot N_D} \cdot (V_{bi} - V)}. \tag{4.11}$$

The potential distribution has a quadratic dependency on the x -coordinate. We find

$$\varphi(-d_n \leq x \leq d_p - 0) = \frac{q}{2 \cdot \varepsilon_H} \cdot \begin{cases} 0 & -d_n \leq x < -x_p \\ N_A \cdot (x + x_p)^2 & -x_p \leq x \leq 0 \\ -N_D \cdot (x_n - x)^2 + (V_{bi} - V) & 0 < x \leq x_n \\ V_{bi} - V & x_n < x \leq d_p - 0 \end{cases} . \tag{4.12}$$

also pictured in Figure 4.3.

We should note, that a voltage $V > 0$ is a so-called forward bias, whereas voltages $V < 0$ are reveres voltages.

The reverse biasing enlarges the space charge stored within the space charge layer because $d_s = x_n + x_p$ increases. This phenomenon is essential for the reverse behavior of diodes and pn -junctions. Basically, it is characterized by extremely small reverse currents and the so-called space charge layer capacity, which dominates the dynamics.

Contrarily, forward voltages decrease the drop of potential $V_{bi} - V$ across the space charge at the junction. That is why remarkable forward currents (I/V-characteristic) can flow through the pn -junction. Thus, a minority carrier charge is stored within the neutral p - and n -regions effecting the device's dynamics due to the so-called diffusion capacity. Both phenomena will be discussed in detail next.

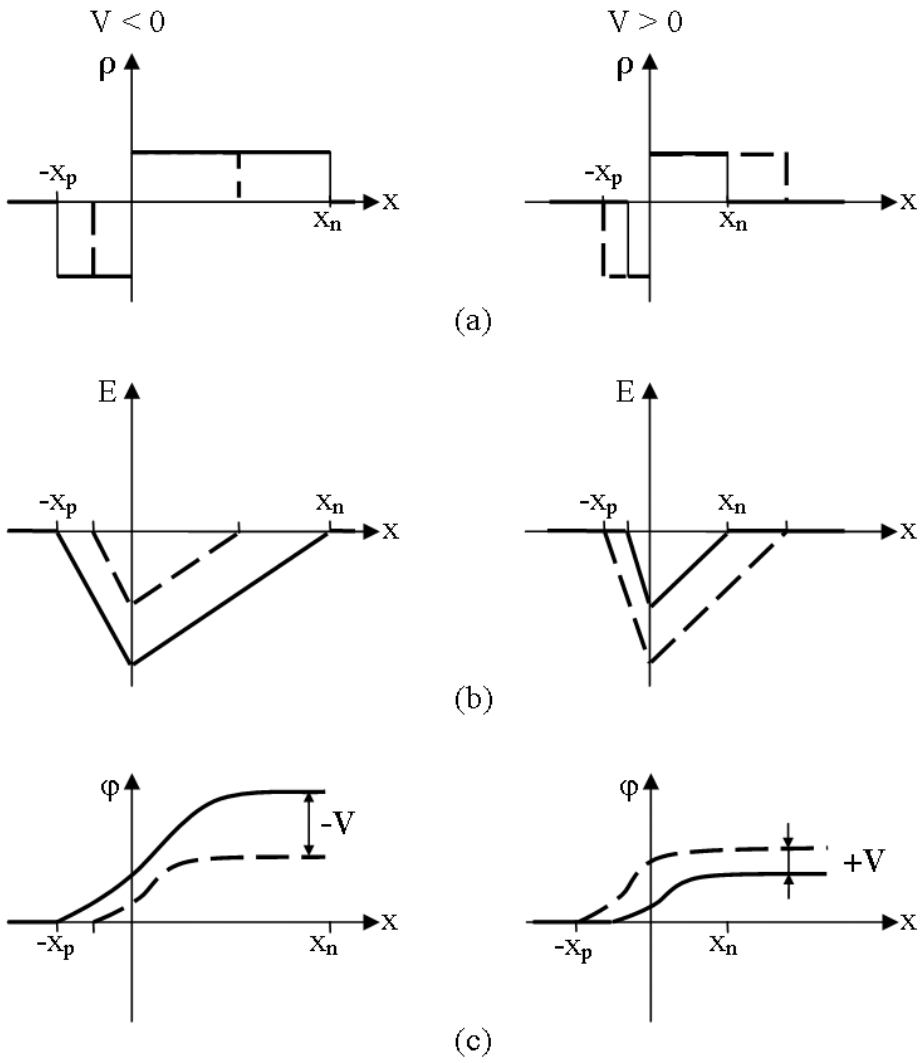


Figure 4.4: (a) space charge, (b) electric field, (c) potential distribution at the *pn* -junction for reverse and forward bias [2]

Self-Assessment Question

Question 4.1:

What quantities determine the widths of the space charge layer near the stochiometric *pn*-junction?

4.2 Space charge capacitance

From the basics in electronics we know, that capacitive effects are caused by charges on electrodes due to supplied voltages. Characteristically, there is a displacement vector field $\vec{D} = \varepsilon \cdot \vec{E}$ where the $\vec{D}$ -field lines have their sources in the positive surface charge on the positively biased electrode. On the other hand, they will end in the negative surface charge on another electrode. This is a simple field model to determine the capacity of structures with metallic electrodes embedded within a constant dielectric (ε). In this case we have to calculate the charge $Q > 0$ on the positive electrode which linearly depends on the supplied voltage V. Then, the capacity C is the proportionality factor between Q and V. Thus, we define

$$C = \frac{Q}{V}. \tag{4.13}$$

This linear and static charge model cannot be applied one-to-one to the space charge layer. The

space charge layer of a pn -junction consists of two separated regions with negative space charge (e.g. in$-x_p \leq x \leq 0$) and positive space charge (e.g. in$0 < x \leq x_n$), respectively. With x_n from Equation (4.10), for instance, we get the positive space charge per unit area Q''_s by

$$Q''_s = q \cdot N_D \cdot x_n\,(V) = \sqrt{2 \cdot q \cdot \varepsilon_H \cdot \frac{N_D \cdot N_A}{N_D + N_A} \cdot (V_{bi} - V)}. \tag{4.14}$$

As shown, the space charge $Q''_s > 0$ stored in $0 < x \leq x_n$ is non-linearly dependent on the supplied voltage. That is why we have to replace the simple capacity definition given in Equation (4.13) by the definition of so-called dynamic capacities

$$C_d = \frac{d}{dV} Q\,(V) > 0. \tag{4.15}$$

In our case an increasing $dQ''_s > 0$ of the space charge $Q''_s > 0$ is due to a decreasing $-dV$ of the supplied voltage V. So, we get the (positive) space charge layer capacity per unit area C''_s from

$$C''_s = -\frac{d}{dV} Q''_s\,(V) = \sqrt{\frac{q \cdot \varepsilon_H}{2 \cdot V_{bi}} \cdot \frac{N_D \cdot N_A}{N_D + N_A}} \cdot \left(1 - \frac{V}{V_{bi}}\right)^{-\frac{1}{2}} > 0. \tag{4.16}$$

Defining a technology-specific constant C''_{s0}

$$C''_{s0} = \sqrt{\frac{q \cdot \varepsilon_H}{2 \cdot V_{bi}} \cdot \frac{N_D \cdot N_A}{N_D + N_A}} \tag{4.17}$$

we obtain the space charge layer capacity C''_s in a fashion typically for network models

$$C''_s = C''_{s0} \cdot \left(1 - \frac{V}{V_{bi}}\right)^{-\frac{1}{2}} > 0. \tag{4.18}$$

In the strict sense, this model of the space charge layer capacity is valid for planar pn -junctions with abrupt doping profiles only. For more complex profiles (e.g. linear graduated profile (see Figure 4.1) obtained by diffusion and implantation) and geometries C''_s has to be calculated by solving the Poisson's equation numerically.

Although the exponent and the technology constant of Equation (4.18) have to be adapted to the dedicated process technology, principally, a space charge layer capacity C''_s may be modeled by

$$C''_s = C''_{s0} \cdot \left(1 - \tfrac{V}{V_{bi}}\right)^{-\alpha} > 0 \qquad with \qquad -\tfrac{1}{3} < \alpha < \tfrac{1}{2} \tag{4.19}$$

Figure 4.5 pictures the voltage dependency of the space charge capacity according to Equation(4.19).

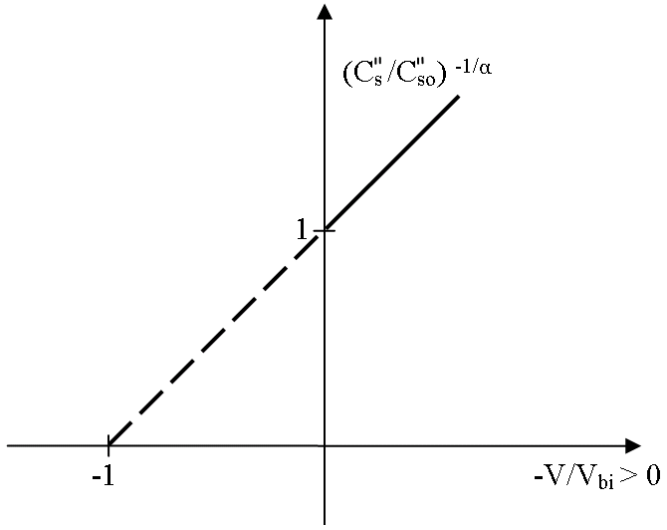


Figure 4.5: Space charge layer capacity C_s''/C_{s0}'' in dependency of the reverse voltage $-V/V_{bi}$

Self-Assessment Question

Question 4.2:

What makes the essential differences between the space charge layer capacity and the capacity of a plate capacitor (electrostatics)?

4.3 I/V-characteristic

To calculate the I/V-characteristic, actually, we need an answer on the question what minority carriers injection means. Then, we can find the minority carrier "storage charge" within the neutral p - and n -side as well as the minority carrier injection currents into these regions. These storage charges and injection currents together with the current contribution due to the recombination in the depletion layer are the essential inputs to determine the dynamic effects and the I/V-characteristic of pn -junctions and diodes.

In module 3.3 we have learned, that the total carrier current $I = I_p + I_n$ in semiconductors is composed of a hole current $I_p(x)$ as well as an electron current $I_n(x)$. Each of them has a drift and a diffusion part. If we assume steady state conditions, no displacement currents have to be considered. That is why the total carrier current I may not vary along the position coordinate x throughout the pn -structure, whereas the hole and the electron currents (I_p andI_n) are not subject to this constrain. As discussed in paragraph 3.3.2, they are influenced by carrier generation and recombination processes in the structure. This fact is used to come up with a simple current balance, set up at the space charge layer as follows:

As discussed in paragraph 4.1, the carriers reverse their role as majority and minority carriers within the space charge layer. Consequently, a similar hole and electron current distribution has to appear (see Figure 4.6).

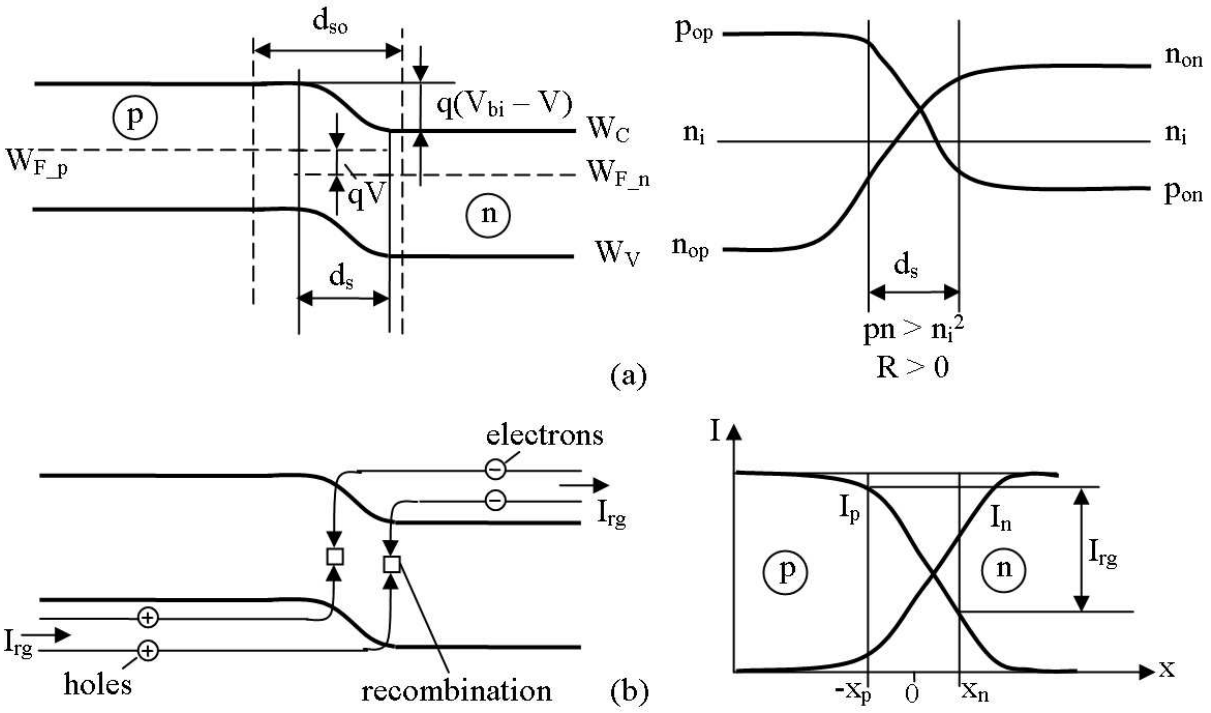


Figure 4.6: (a) Energy-band diagram and carrier distribution; (b) hole and electron current distribution in a pn -semiconductor structure for forward bias [1]

With approaching to the pn -junction, and particularly in the space charge layer, a recombination-based transfer from a hole to an electron current and vice versa takes place as shown in Figure 4.6.

The basic idea behind the development of the (stationary) I/V -characteristic is the continuity of the total carrier current $I = I_p(x) + I_n(x) = const.$ From Figure 4.6 we can take the current balance equation

$$I = I_p(-x_p) + I_n(-x_p). \tag{4.20}$$

Using the one-dimensional continuity equation for the hole current (see Equation(3.37)) under steady state conditions $\frac{d}{dx} I_p = -q \cdot A \cdot R$, and with help of Gauss' law, we find

$$I_p(-x_p) - I_p(x_n) = q \cdot A \cdot \int_{-x_p}^{x_n} R \cdot dx = I_{rg} \tag{4.21}$$

where A is the pn -junction's boundary area that the current perpendicularly flows through ($I_p = \vec{S}_p \cdot \vec{A} = S_p \cdot A$). With Equation (4.21) the Equation (4.20) becomes

$$I = I_{rg} + I_p(x_n) + I_n(-x_p). \tag{4.22}$$

I_{rg} is the recombination current within the space charge layer

$$I_{rg} = q \cdot A \cdot \int_{-x_p}^{x_n} R \cdot dx. \tag{4.23}$$

Based on the Equation (4.22) the development of the pn -junction's (or diodes) I/V -characteristic is led back to a hole and an electron minority carrier injection ($I_p(x_n)$ and$I_n(-x_p)$) into the n - and p -type regions as well as the recombination current I_{rg} in the space charge layer. Actually, the detailed analysis of these three current contributions is not trivial, and therefore, it should not be performed here in detail.

In fact, we will focus our thoughts on the basic ideas behind, and moreover, on discussing the outcomes of such analysis.

The recombination current I_{rg} has to be calculated based on the pretty complex Shockley-Read recombination model $R = R\left(n\left(\varphi,\ \psi_n\right),\ p\left(\varphi,\ \psi_p\right)\right)$. Without proof, we get

$$I_{rg} = q \cdot A \cdot \frac{n_i}{\tau_s} \cdot d_s\left(V\right) \cdot \frac{e^{-\frac{V}{2 \cdot V_T}}}{1 + e^{-\frac{V}{2 \cdot V_T}}} \cdot \left(e^{\frac{V}{V_T}} - 1\right) \tag{4.24}$$

where $\tau_s \approx 1ns$ is a carrier life time in the space charge layer, and d_s is the total width of the space charge layer (e.g. see Equation (4.11)). For small forward bias ($V \geq 0.5\ V$), I_{rg} becomes negligibly small. For reveres voltages ($V < 0$, i.e. negative bias) Equation (4.24) simply delivers

$$I_{rg} = -q \cdot A \cdot \frac{n_i}{\tau_s} \cdot d_s\left(V < 0\right) = I_R < 0. \tag{4.25}$$

The increasing of I_{rg} is due to the space charge layer extension by reverse biasing.

Basically, the minority carrier injection currents $I_p\left(x_n\right)$and $I_n\left(-x_p\right)$ can be derived from the minority carrier distribution in the neutral p -type and n -type region. Obviously, from the methodical point of view, these currents can be calculated in the same way.

For steady state conditions, the necessary minority carrier distributions are solutions of differential equations. Taking into account recombination, we may develop them from the continuity equations (3.36) and (3.37) with the transport equations (3.24) and (3.25). If the recombination can be neglected, we only need the transport equations (see (3.24)(3.25)). In particular, the latter approach is convenient if complex impurity profiles have to be considered and the recombination is really negligible, e.g. in latest technologies.

Actually, we are only interested in solutions adapted to realistic boundary conditions. The most important boundary condition concerns the so-called minority carrier injection at $-x_p$ and x_n of the space charge layer. Minority carrier injection only means that at $-x_p$ and x_n an increased minority carrier concentration occurs, not more. This up-rating of minority carriers is due to the forward bias. Its quantity is given by the Boltzmann's boundary conditions

$$p\left(x_n\right) = p_{0n} \cdot e^{\frac{V}{V_T}} \qquad \textit{and} \qquad n\left(-x_p\right) = n_{0p} \cdot e^{\frac{V}{V_T}} \tag{4.26}$$

as illustrated in Figure 4.7. The minority densities n_{0p} and p_{0n} have to be taken at $-x_p$ and x_n. This is important, if inhomogeneous impurity concentrations should be considered.

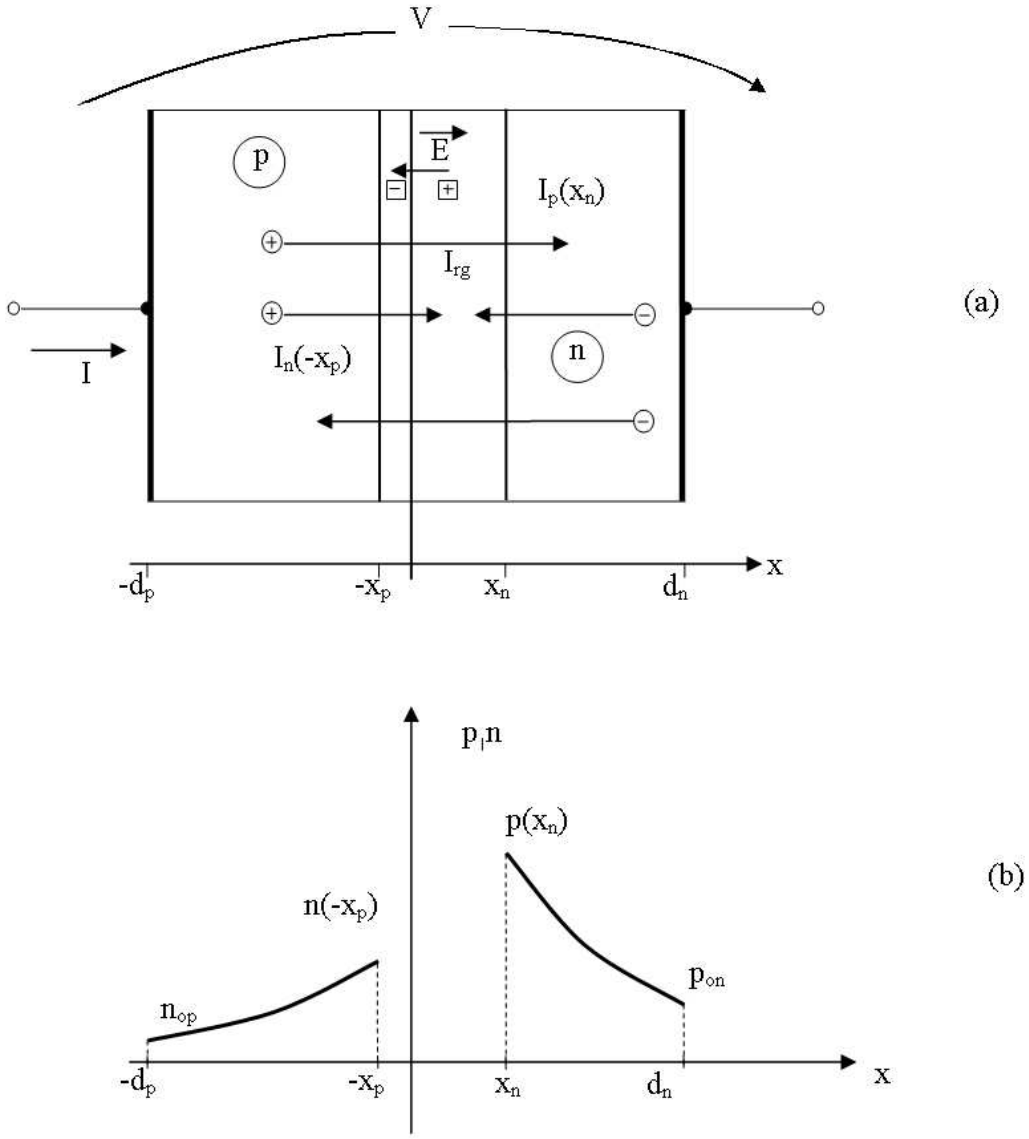


Figure 4.7: (a) Currents and (b) minority carrier distributions at a *pn* -junction [2]

Another important boundary condition concerns the minority concentrations at the contacts (electrodes). There, we may suppose an extremely high recombination velocity. Thus, we use the thermal equilibrium condition

$$(p_{0p} \cdot n_{0p})_{x=-d_p} = (p_{0n} \cdot n_{0n})_{x=d_n} = n_i^2. \tag{4.27}$$

Therewith, the essential ideas on how to determine the injection currents are mentioned.

Looking at latest technologies, actually, the minority carrier density model without recombination, but with complex impurity profiles meets the reality better than other models. Therefore, we will prefer this approach for our modeling purposes. The calculation of the minority injection currents result in

$$I_n\left(-x_p\right) = q \cdot A \cdot D_n \cdot \frac{n_i^2}{Z_A} \cdot \left(e^{\frac{V}{V_T}} - 1\right) \qquad \textit{and} \qquad I_p\left(x_n\right) = q \cdot A \cdot D_p \cdot \frac{n_i^2}{Z_D} \cdot \left(e^{\frac{V}{V_T}} - 1\right) \tag{4.28}$$

where Z_A and Z_D describe the number of acceptors and donors per unit area in the neutral p and n -region given by

$$Z_A = \int_{-d_n}^{-x_p} N_A(x) \cdot dx \qquad and \qquad Z_D = \int_{x_n}^{d_n} N_D(x) \cdot dx. \tag{4.29}$$

Finally, we obtain the demanded I/V -characteristic from the current balance (4.22) under consideration of the recombination current in the space charge layer (4.24) and the minority injection currents (4.28) with(4.29):

$$I = q \cdot A \cdot n_i^2 \cdot \left(\frac{D_n}{Z_A} + \frac{D_p}{Z_D} + \frac{d_s}{n_i \cdot \tau_s} \cdot \frac{e^{-\frac{V}{2 \cdot V_T}}}{1 + e^{-\frac{V}{2 \cdot V_T}}} \right) \cdot \left(e^{\frac{V}{V_T}} - 1 \right) \tag{4.30}$$

This I/V -characteristic is depicted in Figure 4.8.

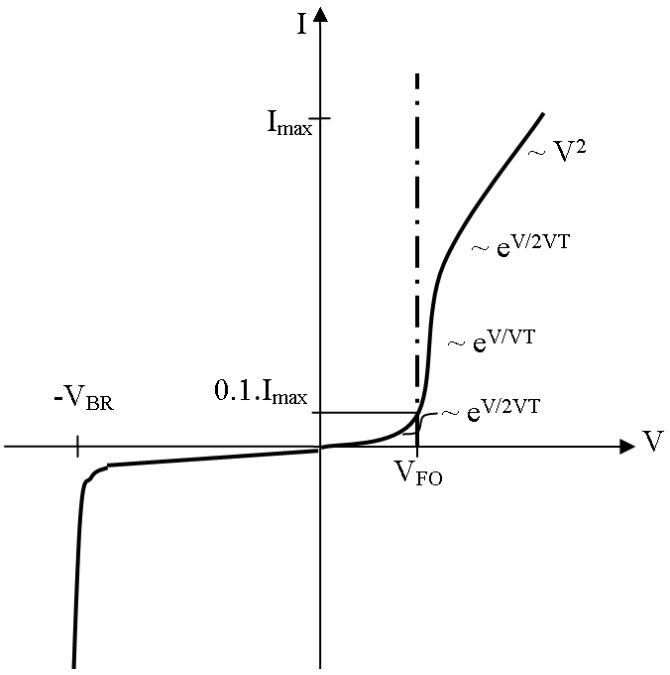


Figure 4.8: I/V -characteristic of a pn -junction (diode) [2]

Generally, the I/V -characteristic of a pn -junction can be given in the well-known form

$$I = I_S \cdot \left(e^{\frac{V}{m \cdot V_T}} - 1 \right) \qquad with \qquad 1 \leq m \leq 2 \tag{4.31}$$

In Equation (4.31) the parameter $I_S(V)$ stands for the saturation current which is dependent on V. Therefore, we may qualify I and I_S correspondingly. In particular, for devices based on Si -technology we get:

$$I = q \cdot A \cdot n_i^2 \cdot \begin{cases} -\frac{d_s}{\tau_s \cdot n_i} & V \leq -0.5\ V \\ \left(\frac{D_n}{Z_A} + \frac{D_p}{Z_D} + \frac{d_s}{n_i \cdot \tau_s} \cdot \frac{e^{-\frac{V}{2 \cdot V_T}}}{1 + e^{-\frac{V}{2 \cdot V_T}}} \right) \cdot \left(e^{\frac{V}{V_T}} - 1 \right) & -0.5\ V < V \leq 0.5\ V \\ \left(\frac{D_n}{Z_A} + \frac{D_p}{Z_D} \right) \cdot e^{\frac{V}{V_T}} & V > 0.5\ V \end{cases} . \tag{4.32}$$

Equation (4.32) needs completing for large negative ($V \leq -V_{BR}$) and higher voltages (e.g. $V > 1.5\ V_{FO}$). The I/V -characteristic for these voltage ranges is pictured in Figure 4.8, too. Indeed, large reverse voltages lead to breakdown and avalanche as well as punch-through effects resulting in an extremely rampant

current. On the other hand, higher positive voltages make high-injection effects, e.g. conductivity modulation, space charge limited currents, which have to be modeled with $m = 2$ in the exponent in Equation(4.31).

In particular, to model digital circuits, the I/V -characteristic of diodes is simplified as a switch that is "on" for $V \geq V_{FO}$ and "off" for $V < V_{FO}$. For Si diodes the forward floating potential V_{FO} where the current is 10 % of its maximum is $V_{FO} \approx 0.7\,V$.

Self-Assessment Questions

Question 4.3.1:

What current components have to be considered to calculate the I/V-characteristic of an pn-junction?

Question 4.3.2:

What stands the floating potential V_{FO} for?

4.4 Dynamics of pn -junction (diode)

The dynamic behavior of pn -junctions and diodes, especially, their dynamic reactions to high frequency (RF) signals or switch-on and switch-off "events" are mainly dominated by charge rebuilding processes within the device. For reverse biases, particularly, the charging and discharging of the space charge layer capacitance govern the dynamics. On the other hand, for forward voltages the rebuilding of minority carrier charges "stored" within the neutral p - and n -type regions outside the space charge layer is essential. In accordance with these facts we may describe such dynamic effects with appropriate capacitances.

Remember, the space charge layer capacitance has been discussed in paragraph 4.2 already. Now, we are going to deal with the so-called diffusion capacity which is important for forward voltages.

The basic idea behind this capacity rests on the minority carrier's transit (travel) time through the neutral p - and n -type regions.

Basically, a phenomenological description of the device dynamics can be developed from the continuity equations (see Equations (3.36) and (3.37)) for the minority carriers. Taking into account the holes (minority charges) in the n -type region $x_n \leq x \leq d_n$, after integrating Equation (3.37) we get the minority carrier charge contained in this n -type region. Practically, it can be assumed that the generation and recombination ($G = R = 0$) may be neglected within the neutral regions. Using the Gauss' low (3.15) we achieve from (3.37)

$$i_p\,(d_n,\ t) - i_p\,(x_n,\ t) = -\frac{d}{dt} Q_p\,(t)\,. \tag{4.33}$$

where Q_p is the minority charge "stored" in the neutral n -region. It is given by

$$Q_p\,(t) = q \cdot A \cdot \int_{x_n}^{d_n} p\,(x,\ t) \cdot dx > 0 \tag{4.34}$$

In an analogous way we receive from Equations (3.36)

$$i_n\,(-x_p,\ t) - i_n\,(-d_p,\ t) = -\frac{d}{dt} Q_n\,(t) \tag{4.35}$$

with the electron charge Q_n in the neutral p -region

$$Q_n\,(t) = -q \cdot A \cdot \int_{-d_p}^{-x_p} n\,(x,\ t) \cdot dx < 0. \tag{4.36}$$

The key point to model the device dynamics based on these minority carrier charges is to approach

$$i_n\left(-d_p,\ t\right) = -\frac{Q_n(t)}{\tau_{n\_p}} \qquad and \qquad i_p\left(d_n,\ t\right) = \frac{Q_p(t)}{\tau_{p\_n}} \tag{4.37}$$

That means the minority carrier currents at the contacts $i_n\left(-d_p,\ t\right)$ and $i_p\left(d_n,\ t\right)$ are assumed to be directly proportional to the minority carrier charge in the neutral p - and n -region. In Equation (4.37), the parameter τ_{n_p} and τ_{p_n} stand for the electron's and hole's travel (transit) time through the neutral p -region and n -region, respectively. Thus, we obtain from Equation (4.1) and (4.3)

$$i_p\left(x_n,\ t\right) = \frac{d}{dt}Q_p\left(t\right) + \frac{Q_p\left(t\right)}{\tau_{p\_n}}. \tag{4.38}$$

and

$$i_n\left(-x_p,\ t\right) = -\left(\frac{d}{dt}Q_n\left(t\right) + \frac{Q_n(t)}{\tau_{n\_p}}\right) \qquad Q_n < 0\ . \tag{4.39}$$

With the parameter λ defined by

$$\lambda = \frac{-Q_n}{Q_p} > 0 \tag{4.40}$$

we are able to combine the Equations (4.38) and (4.39). Finally, we achieve the charge controlling equation

$$i\left(t\right) = i_p\left(x_n,\ t\right) + \ i_n\left(-x_p,\ t\right) = \frac{d}{dt}Q\left(t\right) + \frac{Q\left(t\right)}{1+\lambda} \cdot \left(\frac{1}{\tau_{p\_n}} + \frac{\lambda}{\tau_{n\_p}}\right) \tag{4.41}$$

where $i\left(t\right)$ is the total diode current, and $Q\left(t\right)$ is the total minority carrier charge in both neutral regions

$$Q = Q_p - Q_n > 0. \tag{4.42}$$

It should be mentioned that for forward biases the recombination current in the space charge layer does not play any role, and therefore, it was neglected in Equation (4.41).

Additional to the key point referring the minority carrier's transit time, the second basic idea of the charge controlling approach is to state that the transit time parameters τ_{n_p} and τ_{p_n} as well as the charge ratio controlling constant λ can be calculated under steady state conditions.

Neglecting recombination, we determine the transit time parameters in the same way as proceeded for Equation (4.32). The results are

$$\tau_{n\_p} = \frac{1}{D_n} \cdot \int_{-d_p}^{-x_p} \frac{1}{N_A\left(x\right)} \cdot \int_{-d_p}^{x} N_A\left(x'\right) \cdot dx' \cdot dx \tag{4.43}$$

and

$$\tau_{p\_n} = \frac{1}{D_p} \cdot \int_{x_n}^{d_n} \frac{1}{N_D\left(x\right)} \cdot \int_{x_n}^{x} N_D\left(x'\right) \cdot dx' \cdot dx. \tag{4.44}$$

For homogenously doped regions (i.e. there is no electric drift field) we receive from Equation (4.11) and Equation (4.12)

$$\tau_{n\_p} = \frac{\left(d_p - x_p\right)^2}{2 \cdot D_n} \qquad and \qquad \tau_{p\_n} = \frac{\left(d_n - x_n\right)^2}{2 \cdot D_p}\ . \tag{4.45}$$

These results are typical for carrier transport by diffusion only.

Based on their definitions (see Equation(4.37)) using Equation (4.46) and Equation(4.32), finally we find

$$\lambda = \frac{-Q_n}{Q_p} = \frac{\tau_{n\_p}}{\tau_{p\_n}} \cdot \frac{I_n}{I_p} = \frac{\tau_{n\_p}}{\tau_{p\_n}} \cdot \frac{D_n \cdot Z_D}{D_p \cdot Z_A} > 0 \tag{4.46}$$

From Equation (4.41) under consideration of Equations (4.43) to (4.46) we should learn, that the dynamics of bipolar devices is mainly charge controlled.

However, in many applications the overall dynamics of circuits and electronic systems is not limited by the electronic devices. Often, the device dynamics is so high that the internal device electronics works always under steady state conditions. In this case, Equation (4.41) can be simplified. We get

$$Q = \tau_V \cdot I \qquad with \qquad \tau_V = \frac{1+\lambda}{\frac{1}{\tau_{p\_n}} + \frac{\lambda}{\tau_{n\_p}}} \quad . \tag{4.47}$$

This equation is the starting point for defining the so-called (dynamic) diffusion capacitance C_d.

$$C_d = \frac{d}{dV}Q = \frac{d}{dI}Q \cdot \frac{d}{dV}I \approx \tau_V \cdot \frac{I}{V_T} = \tau_V \cdot \frac{1}{r_d}. \tag{4.48}$$

To avoid any kind of misunderstanding, it has to be clear, that C_d is a capacitance by definition only. The essential characteristic of a physically real capacitance, a displacement field between specially separated charges is missing here. In fact, in the neutral regions at any place and time the minority carrier charge is balanced by a corresponding majority carrier charge.

The quantity $1/r_d$

$$\frac{1}{r_d} = g_d = \frac{d}{dV}I \approx \frac{I}{V_T} \tag{4.49}$$

may be interpreted as the device's dynamic input conductance. From this point of view, the total transit time parameter $\tau_V = C_d \cdot r_d$ is the essential device's characteristic time constant to describe the dynamics of *pn*-junctions and diodes under forward conditions [1].

Self-Assessment Question

Question 4.4:

What capacitance dominates the dynamics of the forward and reversed biased diode, respectively?

4.5 Large-signal network model

The architecture of an appropriate (large signal) network model can be distilled out of the physical implemented structure. Principally, such structures can be divided into a neutral p -and n -region with the pn -junction in between.

At first we focus our considerations on an appropriate network model for the "pure" pn -structure. Basically, we have to take into account the non-linear I/V-characteristic (see equation (4.32)). It is mapped into the dynamic input resistance or conductivity $r_d = 1/g_d$ (e.g. see Equation (4.49)). To model the device's dynamics we need the dynamic space charge layer capacity (C_s) as well as the diffusion capacity (C_d). Both capacities are connected in parallel to the dynamic input resistance. As discussed above, C_s works for reverse biases, whereas C_d is only active for forward voltages.

This model structure is based on an I/V-characteristic where the supplied voltage becomes effective for controlling the pn -junction without any drop of voltage across the neutral regions. That would only be acceptable for extremely high doped regions and relatively small currents. Otherwise, a drop of voltage

decreases the *pn* -junction's control voltage. In practice, we consider this effect by a constant resistor R_B connected in series as shown in Figure 4.9.

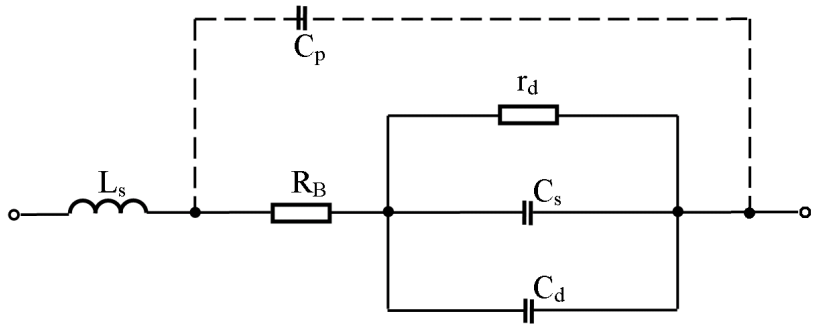


Figure 4.9: Network model of a *pn* -diode [1]

Moreover, for RF applications the network model has to be completed with a parasitic inductivity L_s and a stray capacity C_p due to connecting lines and packaging, respectively (see Figure 4.9).

Self-Assessment Question

Question 4.5:

What does the network model elements r_d, C_s, C_d represent? Have they to be connected in parallel or in series whithin an network model?

4.6 Thermal behavior

The temperature dependence of the I/V-characteristic dominates essentially the thermal behavior of the diode (*pn* -junction). Let us discuss the temperature dependence of the I/V-characteristic in two steps, at first for forward voltages, and after that for reverse biasing.

The methodical approach starts from a forward current $I_F(T_0) = I_{F_S}(T_0) \cdot e^{\frac{V}{V_T(T_0)}}$ given for a reference temperature T_0 (e.g. room temperature $T_0 = 300\ K$). Now, we are interested in the forward current $I_F(T)$ at the temperature $T = T_0 + \Delta T$ where $|\Delta T| << T_0$. Using Equation (4.30) and considering Equations (3.6) with (3.1) and (3.2) as well as Equation (3.9) we find

$$\frac{I_F(T)}{I_F(T_0)} = \frac{n_i^2(T)}{n_i^2(T_0)} \cdot \left(e^{\frac{V}{V_T(T)} - \frac{V}{V_T(T_0)}}\right) = \left(\frac{T}{T_0}\right)^3 \cdot \left(e^{\frac{W_g - q\cdot V}{k\cdot T\cdot T_0}\cdot(T-T_0)}\right) \approx e^{\frac{W_g - q\cdot V}{k\cdot T_0^2}\cdot(T-T_0)}. \tag{4.50}$$

To obtain the latter term in Equation (4.30) we exploited the relation $T_0 << |\Delta T|$, i.e., we used the acceptable approximations $(T/T_0)^3 \approx 1$ and$k \cdot T \cdot T_0 \approx k \cdot T_0^2$.

Equation (4.30) shows, with rising temperatures we have to take into account an exponential increase of the forward current.

In an analogous way we find an exponential temperature dependence for the reverse current, too.

$$\frac{I_R(T)}{I_R(T_0)} = \frac{n_i(T)}{n_i(T_0)} \cdot \frac{\tau_s(T_0)}{\tau_s(T)} \approx e^{\frac{W_g}{k\cdot T_0^2}\cdot(T-T_0)}. \tag{4.51}$$

Self-Assessment Question

Question 4.6:

What kind of temperature dependency has to be taken into account for the forward and the reverse biased diode?

4.7 Problems and examples

Exercise 4 1: pn-diode with thermally effects

A simulation example regarding to the diode with thermal effects can be found in

$DiodeExercise4-1.prb$

Simulate for a sinusoidal input signal E_Vein with an amplitude of 15 V and a frequency of 50 Hz the output voltage (node 2).

Plot the temperature of the diode (node 3) and the current through the diode (I.D1.ED).

Repeat the simulation with a large heat sink (C_th1=1). With this the temperature of the diode is nearly constant 300 K (Figure 4.12) and the maxima of the current is greater (Figure 4.13).

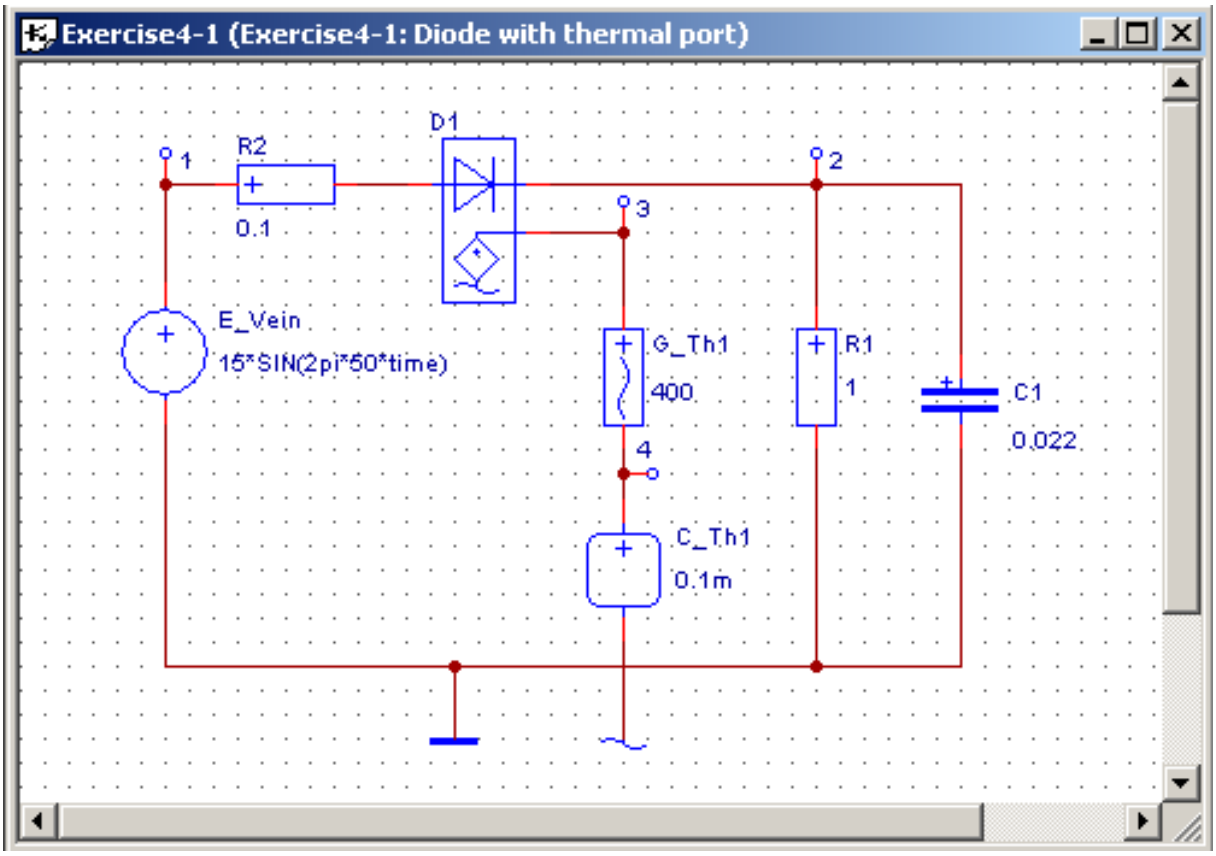


Figure 4.10: Schematic of a rectifier circuit with thermally effects

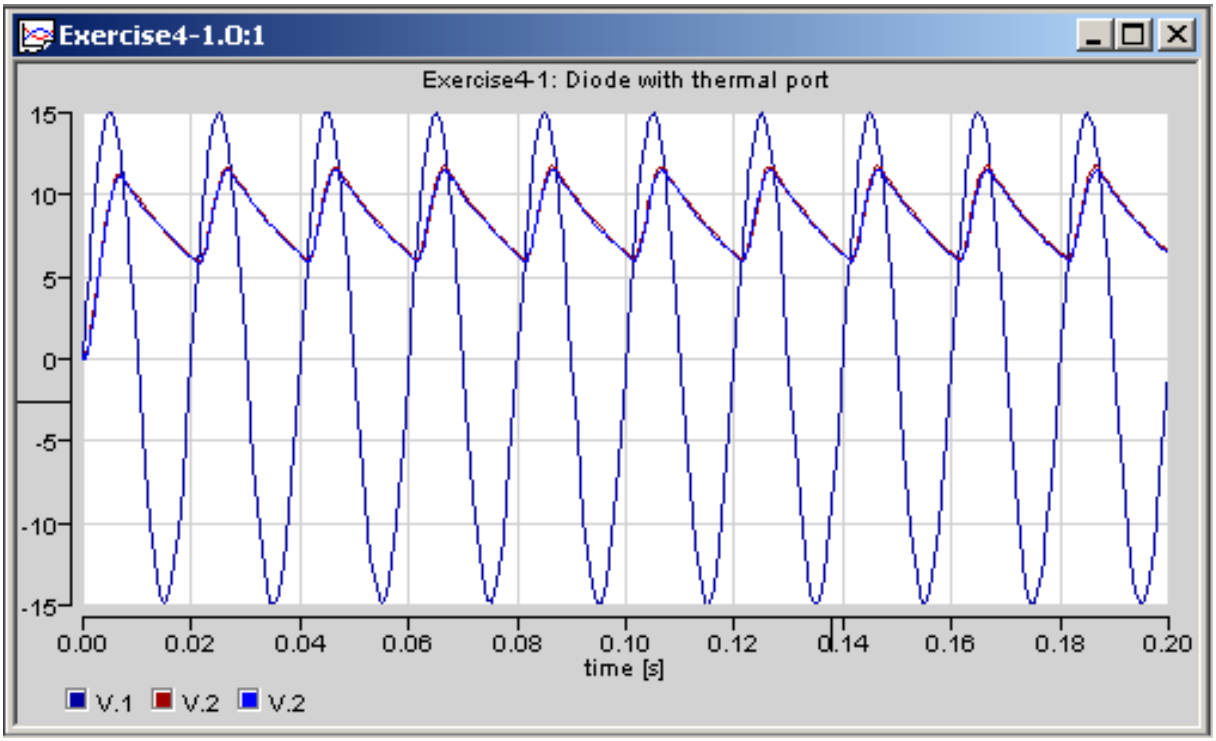


Figure 4.11: Input and output voltage of a rectifier circuit

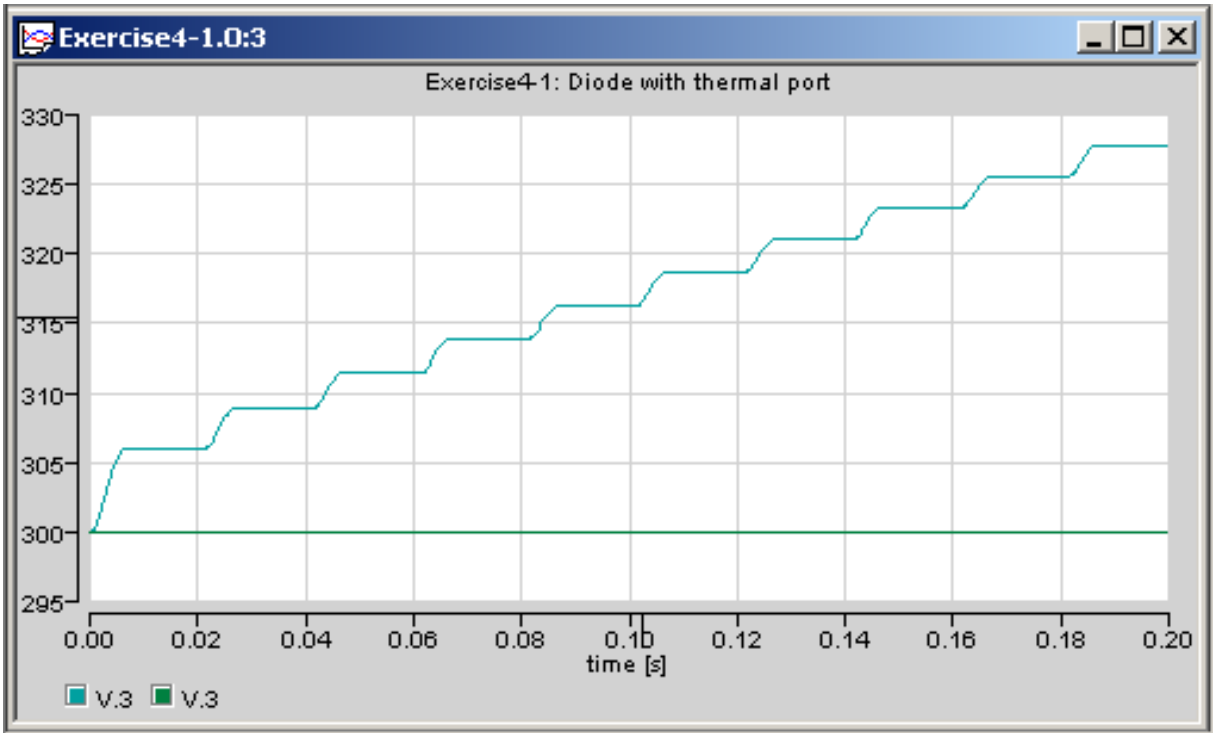


Figure 4.12: Temperature of the *pn* -diode (the green line gives the temperature with a very large heat sink C_Th1=1)

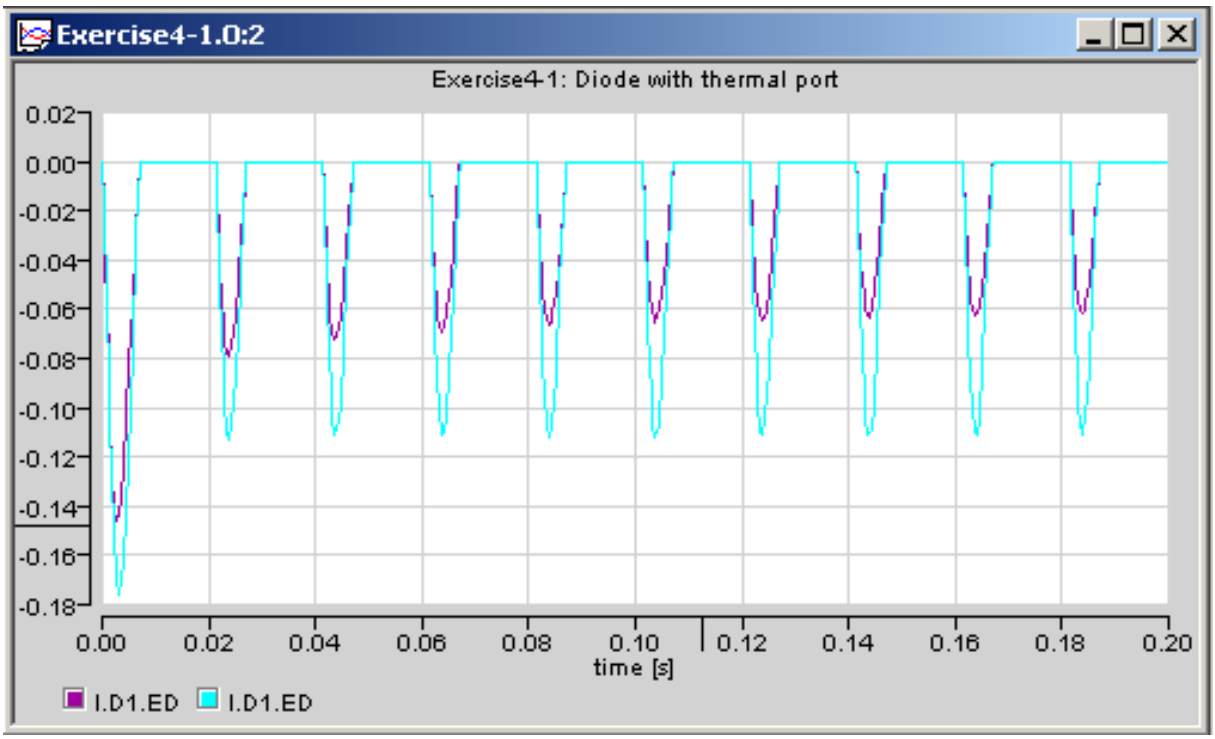


Figure 4.13: Current of the *pn* -diode (cyan: wit very large heat sink C_Th1=1)

Module 5

Bipolar Transistors

Module units

Module overview. The reader becomes familiar with:

- the structure of bipolar transistors
- the function principles of bipolar transistors
- the role of minority and majority carriers
- the density of electrons and holes
- the fundamental equations
- dynamic effects

Module objectives. *They are aimed at*

1. Understanding the function principle of bipolar transistors.
2. Being acquainted with the network models (Ebers-Moll, Gummel-Poon).
3. Applying the appropriate network models correctly.

Module prerequisites. Expertise imparted in module 3 and module 4

5.1 Function principles of bipolar transistors

5.1.0 I/V-characteristic

Bipolar transistors are three terminal devices consisting of p-n-p or n-p-n structures. The operation is based on minority carrier injection. But both carrier types (minorities and majorities) are involved in the operation. Bipolar transistors are current controlled devices.

The three terminals are emitter (E), base (B), and collector (C). Unless stated otherwise, the device structure considered here is that of a vertical npn transistor.

Simply spoken, the bipolar transistor consists of two opposite polarized diodes with a common very shallow region, the bases (B) (see Figure 5.1).

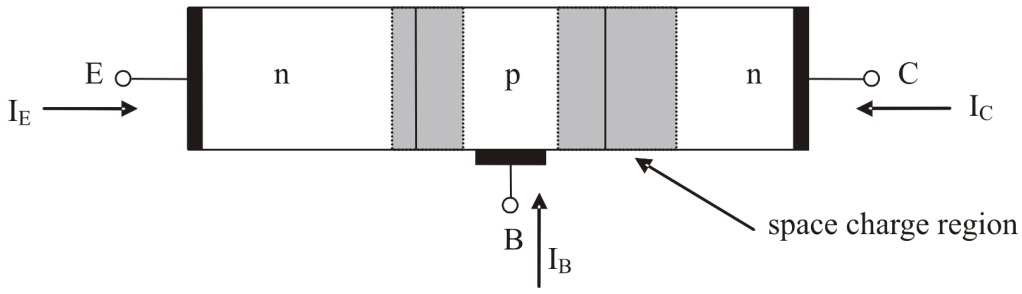


Figure 5.1: One-dimensional structure of the bipolar transistor

First we describe the transistor operation qualitatively. Normally the emitter-base junction is forward biased and the base-collector junction is reverse biased. Therefore, two principles of the diode play an important role for the operation of the bipolar transistor:

- Injection of minority carriers (electrons) in the quasi-neutral base through the forward biased emitter-base junction (emitter current).
- Collection of this minorities through the backward biased base-collector junction (the space charge regions are approximately free of mobile carriers).

The injected minorities can reach the base-collector junction by diffusion because of the narrow base. The loss of minorities in the base due to recombination is very small and is part of the base current. Therefore the collector current and the base current are

$$\begin{aligned} I_C &= -\alpha_f \cdot I_E \\ I_B &= -(1-\alpha_f) \cdot I_E \end{aligned} \tag{5.1}$$

where α_f is the common-base current gain and is somewhat smaller than unity.

To derive the fundamental equations for the computation of the currents in the bipolar transistor we have to solve the continuity equation (3.36) and the current density equation (3.24) in the neutral base region under the following simplifying assumptions:

- Current flow is one-dimensional and dominated by the diffusion component.
  The depletion layer edges of the emitter-base and base-collector junctions are abrupt and situated at $x = 0$ and $x = W_b$ respectively (s. Figure 5.2).
- The dope concentration N_a is constant.
- The region in between $(0 < x < W_b)$ is quasi-neutral and $p(x) \approx N_a + n(x)$, where $n(x)$ are the injected minority carriers into the base.

- Low injection into the base ($n(0) \ll N_a$).
- The generation-recombination part of the continuity equation can be written as $G - R = -(n - n_0)/\tau_{n_0}$ (n_0 - equilibrium concentration, τ_{n_0} - lifetime of electrons in the base) (see module 3.1.5).

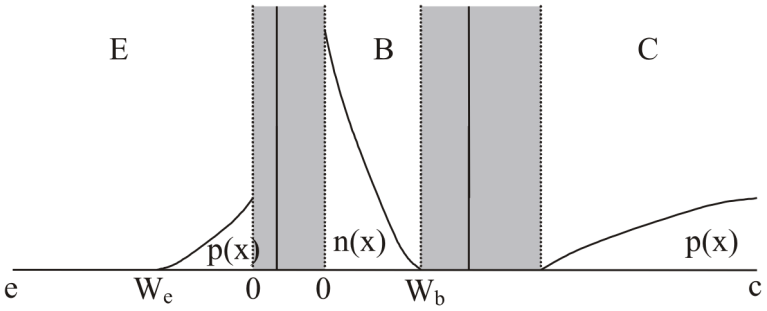


Figure 5.2: Minority carrier distributions in the emitter and base of an npn transistor

With all these assumptions, we get the access minority carrier concentration in the base (static case) (Figure 5.2)

$$n(x) - n_{0b} = q \cdot \frac{n_{ib}^2}{N_A} \cdot \frac{\sinh\left(\frac{W_b - x}{L_n}\right)}{\sinh\left(\frac{W_b}{L_n}\right)} \cdot \left(e^{\frac{V_{BE}}{V_T}} - 1\right) \tag{5.2}$$

and the minority current density

$$S_n(x) = -q \cdot D_n \frac{n_{ib}^2}{N_A \cdot L_n} \frac{\cosh\left(\frac{W_b - x}{L_n}\right)}{\sinh\left(\frac{W_b}{L_n}\right)} \cdot \left(e^{\frac{V_{BE}}{V_T}} - 1\right). \tag{5.3}$$

Appropriate formulas also apply to the injection of holes from the base into the emitter:

$$p(x) - p_{0e} = q \cdot \frac{n_{ie}^2}{N_D} \cdot \frac{\sinh\left(\frac{W_e - x}{L_p}\right)}{\sinh\left(\frac{W_e}{L_p}\right)} \cdot \left(e^{\frac{V_{BE}}{V_T}} - 1\right). \tag{5.4}$$

Because the doping levels in emitter and base are quite different ($N_D \gg N_A$), the quantities n_{ib}^2 and n_{ie}^2 are not the same, due to the effect of bandgap narrowing ($n_{ib}^2 < n_{ie}^2$).

$$S_p(x) = q \cdot D_p \frac{n_{ie}^2}{N_D \cdot L_p} \frac{\cosh\left(\frac{W_e - x}{L_p}\right)}{\sinh\left(\frac{W_e}{L_p}\right)} \cdot \left(e^{\frac{V_{BE}}{V_T}} - 1\right) \tag{5.5}$$

Multiplying the current densities by the emitter area gives for the total emitter current

$$I_{BE} = -\left(S_n(0) - S_p(0)\right) \cdot A_E = I_{0E} \cdot \left(e^{\frac{V_{BE}}{V_T}} - 1\right) \tag{5.6}$$

with the emitter saturation current

$$I_{0E} = q \cdot A_E \cdot \left[D_n \frac{n_{ib}^2}{N_A \cdot L_n} \frac{1}{\tanh\left(\frac{W_b}{L_n}\right)} + D_p \frac{n_{ie}^2}{N_D \cdot L_p} \frac{1}{\tanh\left(\frac{W_e}{L_p}\right)} \right]. \tag{5.7}$$

All minorities which reach the collector junction give the collector current

$$I_C = -S_n(W_b) \cdot A_E = I_{0C} \cdot \left(e^{\frac{V_{BE}}{V_T}} - 1 \right) \tag{5.8}$$

with the collector saturation current

$$I_{0C} = q \cdot A_E \cdot D_n \frac{n_{ib}^2}{N_A \cdot L_n} \frac{1}{\sinh\left(\frac{W_b}{L_n}\right)}. \tag{5.9}$$

With the Equations (5.8), (5.1), and the additional assumption for transistors with a rather thick base region, that the hole injection into the emitter can be neglected in relation to the electron injection into the base ($-S_n(0) \gg S_p(0)$), we get the common base current gain

$$\alpha_f = \frac{1}{\cosh\left(\frac{W_b}{L_n}\right)}. \tag{5.10}$$

5.1.1 The Ebers-Moll model

The Equation (5.1) with (5.6) - (5.10) represent the Ebers-Moll model for the normal forward mode (Figure 5.3), the simplest and eldest model of the bipolar transistor.

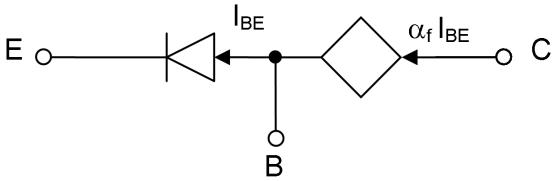


Figure 5.3: Ebers-Moll equivalent circuit diagram for the normal forward mode

Only a part of the injected emitter current arrives at the collector, the rest equals the base current. Because the injected current plays the dominant role in this model another name is injection model.

Due to the symmetrical structure of the transistor (see Figure 5.1) the same equations are valid for the reverse mode when the collector-base junction is forward biased. The complete static Ebers-Moll model for all modes of operation (forward and reverse) is shown in the following figure.

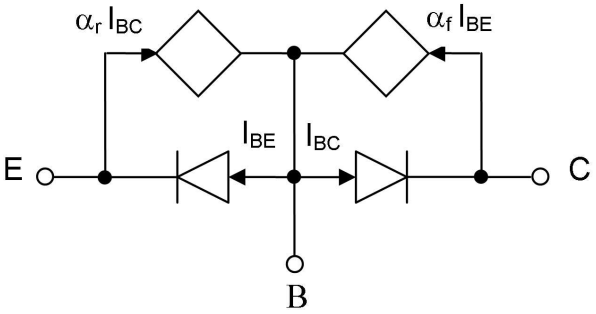


Figure 5.4: Ebers-Moll equivalent circuit diagram

α_f and α_r are the current gain in forward and reverse mode, respectively. They are smaller than 1 and normally $\alpha_f > \alpha_r$

5.1.2 Dynamical effects and capacitances

Dynamical effects are due to the charge storage in the transistor. Two effects must be considered (see module 4), the charge storage in the space-charge layers, modelled by the space-charge layer capacitances

$$C_{Te} = \frac{C_{je}}{(1-V_{BE}/V_{je})^{m_{je}}}$$
$$C_{Tc} = \frac{C_{jc}}{(1-V_{BC}/V_{jc})^{m_{jc}}} \tag{5.11}$$

and the diffusion capacitances which can be estimated from the minority carrier charge stored in the base region in forward and reverse mode, respectively (see the charge control principle in module 4.4):

$$C_{be} = \tau_f \cdot \frac{dI_{BE}}{dV_{BE}}$$
$$C_{bc} = \tau_r \cdot \frac{dI_{BC}}{dV_{BC}} \tag{5.12}$$

τ_f, τ_r are the integral transit times of electrons and holes through the transistor.

The complete dynamical large signal Ebers-Moll is shown in Figure 5.5.

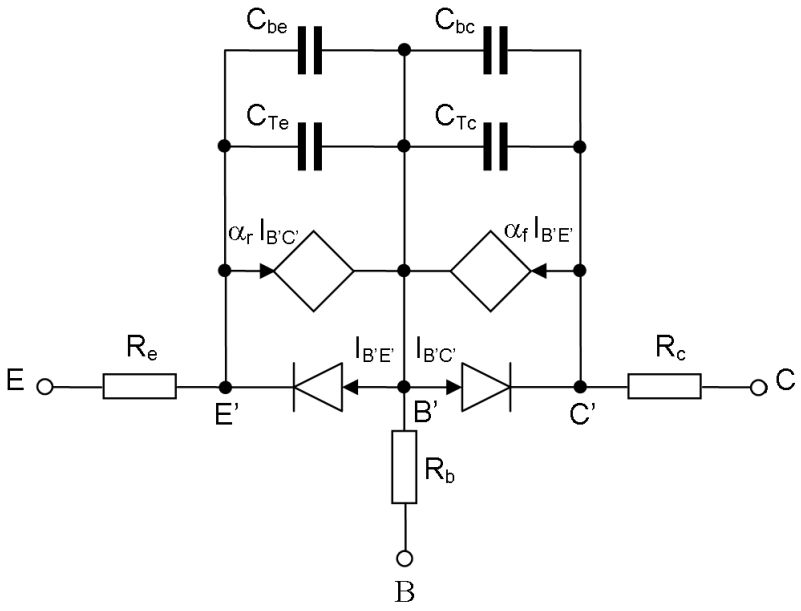


Figure 5.5: Ebers-Moll model with series resistances and capacities

Self-Assessment Questions

Question 5.1.1:

Which mechanisms play a fundamental role in the bipolar transistor?

Question 5.1.2:

Why is this transistor called bipolar transistor?

Question 5.1.3:

Which phenomena cause the base current?

Question 5.1.4:

What does the Ebers-Moll model describe?

Question 5.1.5:

Which effects cause the diffusion capacitances?

5.2 Advanced models of bipolar transistors

5.2.0 Real structures of bipolar transistor

The structure of the bipolar transistor in Figure 5.1 is a rough simplification, in order to be able to derive the transistor equations without complicated boundary conditions. Figure 5.6 and Figure 5.7 show the cross sections through today used, planar transistors. The mainly active regions are hatched represented. The technologically necessary areas outside of these regions bring additional, parasitic effects (bulk resistors, capacities, diodes).

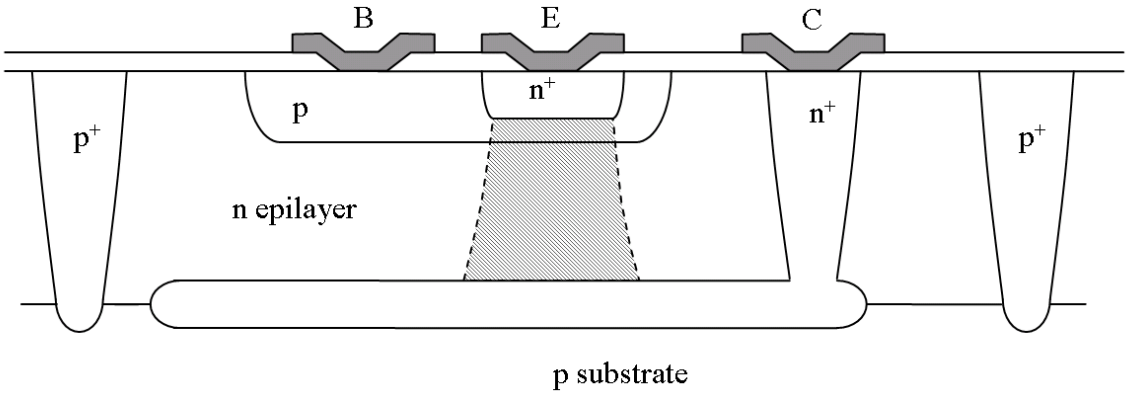


Figure 5.6: Cross-section of a vertical npn transistor

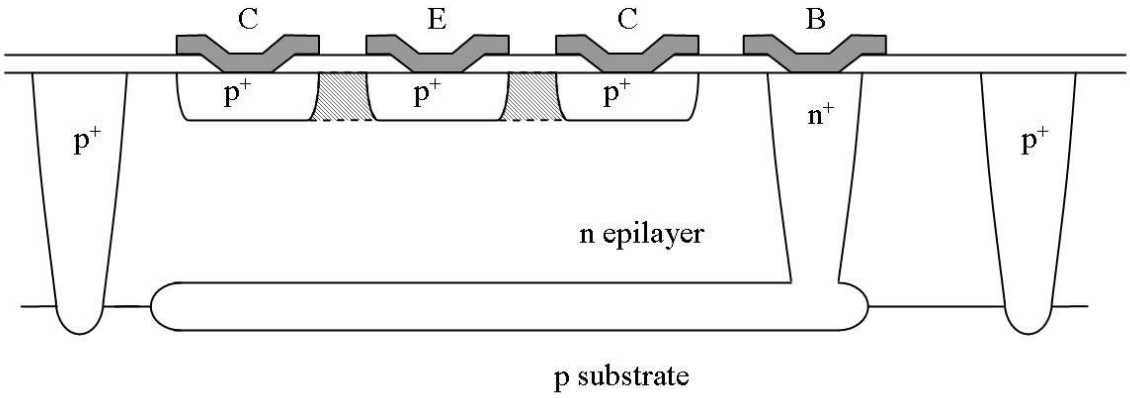


Figure 5.7: Cross-section of a lateral pnp transistor

The structural difference of the lateral pnp with respect to the vertical npn has consequences for the electrical behaviour:

- The lateral distances are usually longer than the vertical ones (esp. the base width), so the transit times are also longer. This reduces the attainable transit frequency compared to vertical transistors.
- The n epilayer in a lateral pnp is part of the base region and not part of the collector as in a vertical npn transistor. So the lightly doped epilayer gives rise to high injection effects in the base, but quasi-saturation does not occur and $R_{cv} \equiv 0$. However, extra charge storage does occur, but in the base region under the emitter. This is detrimental to high f_T values, too.

- Emitter and collector dope concentrations in lateral pnp transistors resemble those of the base in a vertical npn, so they are in the range $10^{18} - 10^{19}\ cm^{-3}$ instead of the usual emitter dope around $10^{20}\ cm^{-3}$. This makes the emitter less efficient because of increased electron injection from the base. It also increases the emitter series resistance and enhances current crowding under the emitter.
- The forward-biased emitter injects holes into the base under the emitter. Most of these holes are stored there, but part of them will be collected by the substrate, giving rise to a substrate current that reduces the gain.
- The recombination in the epilayer bulk is usually negligible, but the recombination at the oxide interface may be important and must be taken into consideration.

In comparison with vertical npn transistors lateral pnp transistors have a lower current gain (< 50), low maximum transit frequencies ($< 100\ MHz$), and lower collector currents. Despite all these differences, the same models as for the npn transistor can be used. The parameters must be adapted to the concrete device. But for more exact investigations the models must be supplemented.

We assumed the minorities in the base move only by their relatively slow diffusion due to the density gradient. The highest frequency, with which a transistor can be used, is determined however by the transit time and thus by the speed, with which the minorities in the base move. The charge carriers can be further accelerated by an additional electrical drift field in the base. Such field results from the impurity concentration profile in the base, represented in the picture (Figure 5.8). In the equilibrium, the majority charge carriers diffuse due to the impurity concentration profile in the direction of the collector and produce such an accelerating field for the minorities (see module 3.2). Such transistors were also called diffused base transistors.

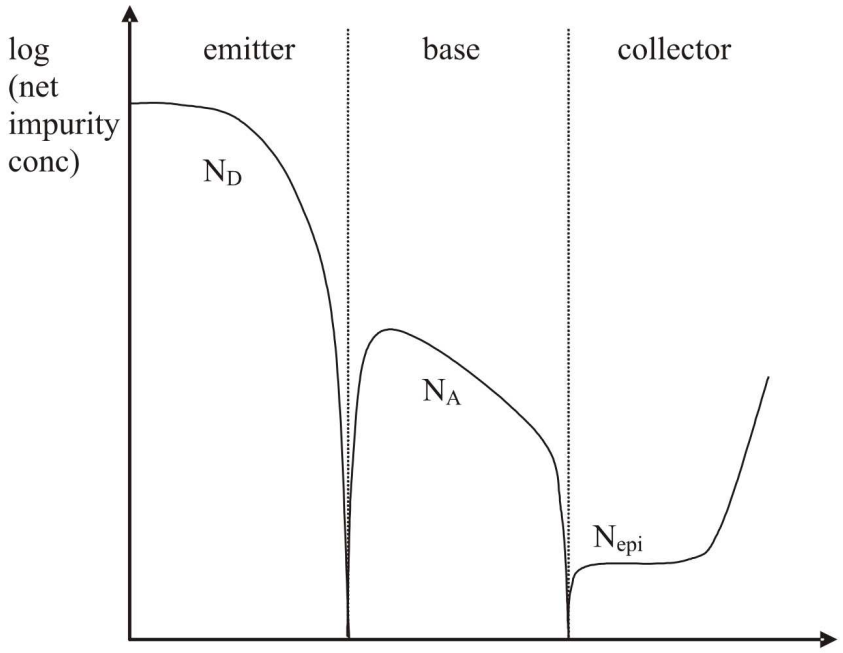


Figure 5.8: Typical doping profile in an npn transistor under the emitter contact

5.2.1 Transport model

For modern transistors the indicated Ebers Moll Injection model is not suitable. Due to very small base widths, the high drift field in the base, and a low recombination rate (high relaxation time) we have $\alpha_f \approx 1$.

With $i_f = \alpha_f \cdot I_{BE}$ and $i_r = \alpha_r \cdot I_{BC}$, and splitting of the diodes in an ideal diode (with i_f and i_r), and an nonideal ones (with $i_f \cdot (1/\alpha_f - 1)$ and $i_r \cdot (1/\alpha_r - 1)$), the equivalent circuit diagram Figure 5.4 will be changed to Figure 5.9.

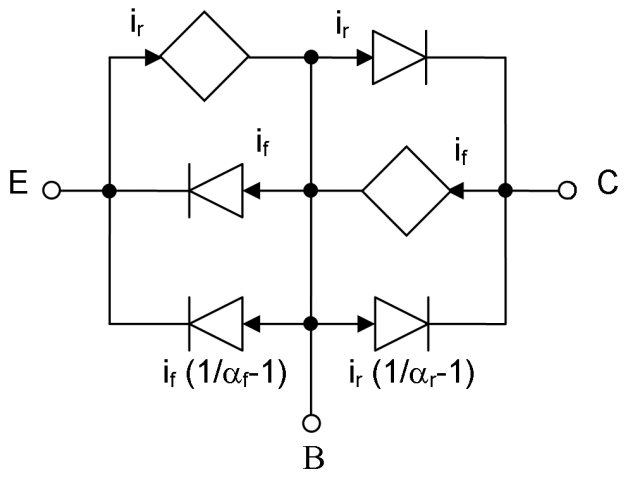


Figure 5.9: Development of the transport model

The ideal diodes have characteristics with the same saturation currents I_s

$$\begin{aligned} i_f &= I_s \cdot \left(e^{\frac{V_{BE}}{V_T}} - 1 \right) \\ i_r &= I_s \cdot \left(e^{\frac{V_{BC}}{V_T}} - 1 \right) . \end{aligned} \tag{5.13}$$

Because in the two current sources and the correspondent diodes the same current flows the following simplification to Figure 5.10 is possible ($\beta. = \alpha./(1-\alpha.)$ are the current gains in the common emitter configuration).

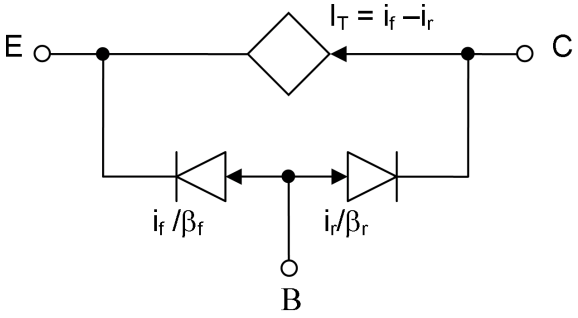


Figure 5.10: Equivalent circuit diagram of the transport model

This is the basic form of the so called Ebers-Moll transport model. With resistances and capacitances we get the complete model in Figure 5.11.

For a commercial transistor 2N4124 one receives the following characteristics compared with the substantially more exact Spice model, which is used in modern simulators. It can be seen that the transport model has some deviations opposite the real transistor:

- The current gain depends on the collector-emitter voltage and
- The current gain depends on the collector current.

Therefore, further improvements are necessary and will be described in the next chapter.

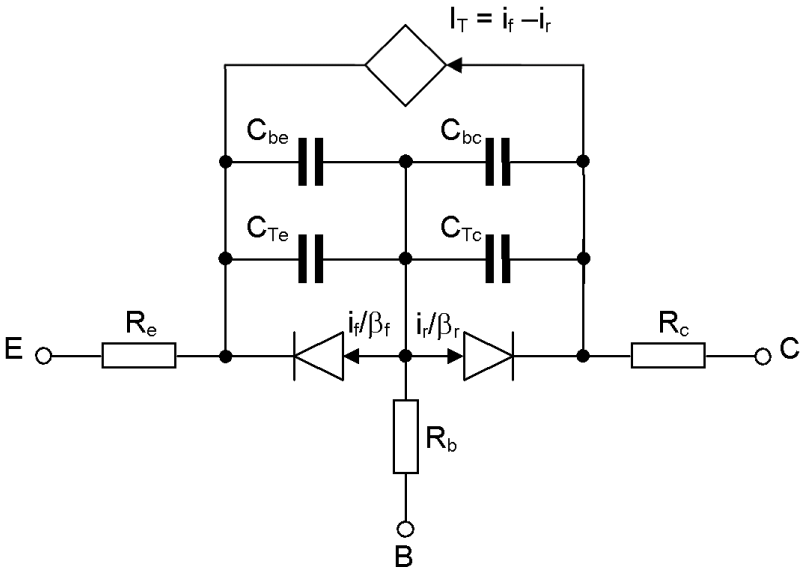


Figure 5.11: Transport model with capacities and series resistance

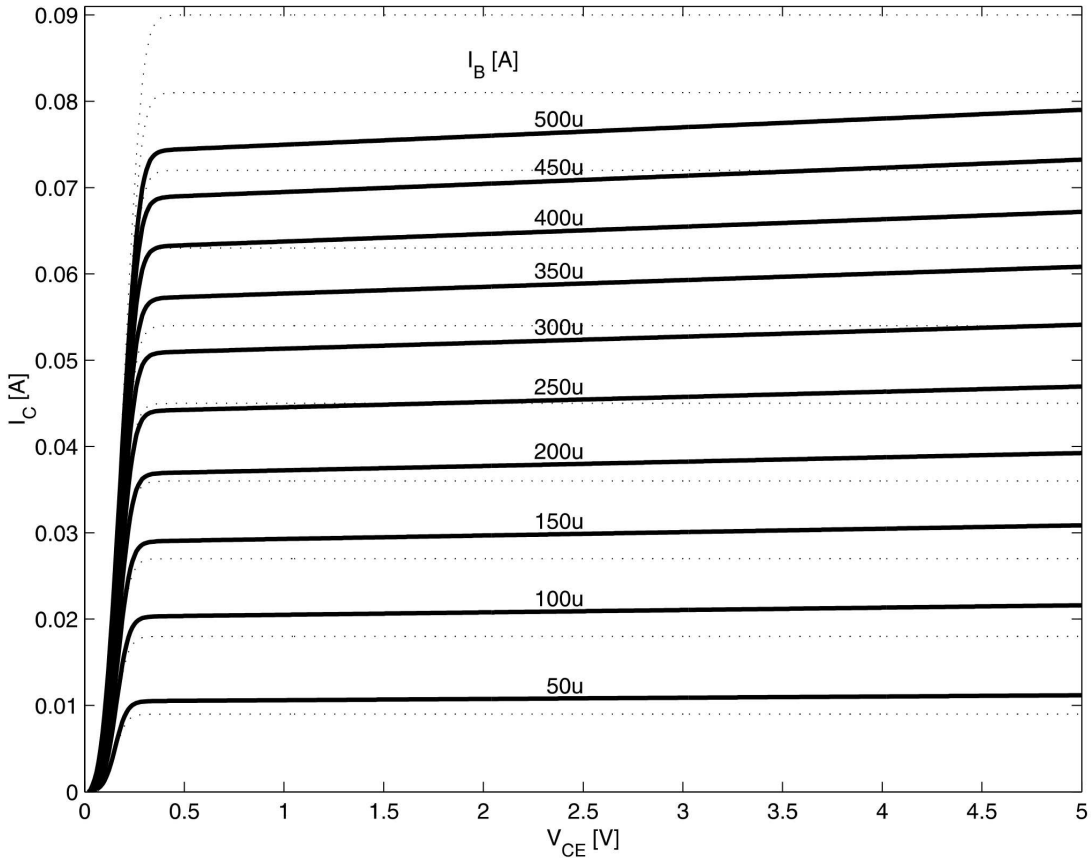


Figure 5.12: Comparison of the transport model (dotted line) with the real transistor 2N4124

5.2.2 Gummel-Poon model

5.2.2.0 Base charge

The continuity equation (3.36) in the one-dimensional npn transistor (Figure 5.1) and in static case is

$$S_n = -q \cdot \mu_n \cdot n \cdot \frac{d\psi_n}{dx}. \tag{5.14}$$

With the electron concentration $n = n_i \cdot e^{\frac{\varphi - \psi_n}{V_T}}$ (see Equation (3.7)) Equation (5.14)

will be integrated from the emitter contact to the collector contact

$$\int_e^c \frac{S_n}{q \cdot \mu_n \cdot n_i} \cdot \mathrm{e}^{-\frac{\varphi}{V_T}}\, dx = -V_T \cdot \left(\mathrm{e}^{-\frac{\psi_{ne}}{V_T}} - \mathrm{e}^{-\frac{\psi_{nc}}{V_T}}\right). \tag{5.15}$$

ψ_{ne}, ψ_{nc} are the quasi-Fermi levels on the emitter and collector, respectively. We neglect the recombination in the base region, so Sp=0 and $\psi_p(x) = 0$. The base-emitter voltage and the base-collector voltage are

$$\begin{aligned} V_{be} &= -\psi_{ne} \\ V_{bc} &= -\psi_{nc}. \end{aligned} \tag{5.16}$$

Neglect of generation and recombination also makes S_n constant. Therefore Equation (5.15) can be written as

$$S_n \cdot \int_e^c \frac{1}{q \cdot D_n \cdot n_i} \cdot \mathrm{e}^{-\frac{\varphi}{V_T}} dx = -\left(\mathrm{e}^{-\frac{\psi_{ne}}{V_T}} - \mathrm{e}^{-\frac{\psi_{nc}}{V_T}}\right). \tag{5.17}$$

Only the base region gives a significant contribution to the integral on the left hand side and with $\psi_p = 0$ and

$$p = n_i \cdot \mathrm{e}^{\frac{-\varphi + \psi_p}{V_T}} \tag{5.18}$$

we get

$$S_n = \frac{\mathrm{e}^{\frac{V_{BE}}{V_T}} - \mathrm{e}^{\frac{V_{BC}}{V_T}}}{\int_e^c \frac{p}{q \cdot D_n \cdot n_i^2}\, dx}. \tag{5.19}$$

With the total charge of one type (holes or electrons) in the active part of the transistor

$$Q_b = q \cdot \int_e^c A_E \cdot p\, dx, \tag{5.20}$$

its value at zero bias Q_{b0}, and the collector saturation current I_s(Eq. (5.8)) the denominator can be rewritten as

$$\int_e^c \frac{p}{q \cdot D_n \cdot n_i^2}\, dx = \frac{Q_b}{I_s \cdot Q_{b0}}. \tag{5.21}$$

Now, the general expression for the collector current becomes

$$I_C = \frac{Q_{b0}}{Q_b} \cdot I_s \cdot \left(e^{\frac{V_{BE}}{V_T}} - e^{\frac{V_{BC}}{V_T}}\right). \tag{5.22}$$

Splitting up into forward and reverse currents gives

$$I_C = I_f - I_r \tag{5.23}$$

with

$$\begin{aligned} I_f &= \frac{Q_{b0}}{Q_b} \cdot I_s \cdot \left(e^{\frac{V_{BE}}{V_T}} - 1 \right) \\ I_r &= \frac{Q_{b0}}{Q_b} \cdot I_s \cdot \left(e^{\frac{V_{BC}}{V_T}} - 1 \right) . \end{aligned} \tag{5.24}$$

The total base charge consists of

- Q_{b0} - the zero bias base charge
- Q_{Te} and Q_{Tc} - the depletion charges of the emitter an collector junctions, respectively
- Q_{be} and Q_{bc} - the stored charges of the minority carriers, injected respectively from the emitter and collector

$$Q_b = Q_{b0} + Q_{Te} + Q_{Tc} + Q_{be} + Q_{bc} \tag{5.25}$$

5.2.2.1 Stored Charges of Injected Minority Carriers

The stored charges Q_{be} and Q_{bc} can be related to the forward and reverse currents i_f and i_r by the charge-control principle

$$\begin{aligned} Q_{be} &= \tau_f \cdot I_f \\ Q_{bc} &= \tau_r \cdot I_r \end{aligned} \tag{5.26}$$

where the constants τ_f and τ_r are the forward and reverse base transit times. Substitution of (5.24) and (5.26) into (5.25) and under neglect of the junction charges gives

$$Q_b = Q_{b0} + \tau_f \cdot \frac{Q_{b0}}{Q_b} \cdot I_s \cdot \left(e^{\frac{V_{BE}}{V_T}} - 1 \right) + \tau_r \cdot \frac{Q_{b0}}{Q_b} \cdot I_s \cdot \left(e^{\frac{V_{BC}}{V_T}} - 1 \right) . \tag{5.27}$$

and with Equation (5.13) of the transport model

$$Q_b = Q_{b0} + \tau_f \cdot \frac{Q_{b0}}{Q_b} \cdot i_f + \tau_r \cdot \frac{Q_{b0}}{Q_b} \cdot i_r \tag{5.28}$$

5.2.2.2 Depletion Charges of the Junctions - Early Effect

The depletion charges of the emitter and collector junctions cause the rising of the collector current with the collector-emitter voltage in the saturation region. These depletion charges depend on their junction voltages. This effect is called the Early effect. The charges are defined as functions of V_{BE} and V_{BC}, respectively (C_{Te}, C_{Tc} s. Equation(5.11)):

$$\begin{aligned} Q_{Te} &= \int_0^{V_{BE}} C_{Te}\, dv \\ Q_{Tc} &= \int_0^{V_{BC}} C_{Tc}\, dv \end{aligned} \tag{5.29}$$

For the Gummel-Poon model usually the expressions are approximated by using the mean values over

the working range of the capacitances $\bar{C}_{Te}$, $\bar{C}_{Tc}$ or by defining so-called Early voltages $V_{Early*} = Q_{b0}/\bar{C}_{T*}$

$$\begin{aligned} Q_{Te} &= \bar{C}_{Te} \cdot V_{BE} = Q_{b0} \cdot \tfrac{V_{BE}}{V_{Earlyf}} \\ Q_{Tc} &= \bar{C}_{Tc} \cdot V_{BC} = Q_{b0} \cdot \tfrac{V_{BC}}{V_{Earlyr}} \; . \end{aligned} \tag{5.30}$$

So we have

$$Q_{b0} + Q_{Te} + Q_{Tc} = Q_{b0} + Q_{b0} \cdot \frac{V_{BE}}{V_{Earlyf}} + Q_{b0} \cdot \frac{V_{BC}}{V_{Earlyr}} = Q_{b0} \cdot (1 + q_1). \tag{5.31}$$

The abbreviation $1 + q_1$ is called the Early factor.

5.2.2.3 Implementation of the Base Charge in the Gummel-Poon Model

For the whole base charge Q_b we get with (5.28) and (5.31)

$$Q_b = Q_{b0} \cdot (1 + q_1) + \tau_f \cdot \frac{Q_{b0}}{Q_b} \cdot i_f + \tau_r \cdot \frac{Q_{b0}}{Q_b} \cdot i_r \tag{5.32}$$

The solution of this equation is

$$\frac{Q_b}{Q_{b0}} = \frac{1 + q_1}{2} \cdot \left[1 + \sqrt{1 + \frac{4}{1 + q_1} \cdot \left(\frac{\tau_f}{Q_{b0}} \cdot i_f + \frac{\tau_r}{Q_{b0}} \cdot i_r \right)} \right]. \tag{5.33}$$

Because it is not possible to measure the zero bias base charge, Q_{b0}/τ_f will be replaced by I_{kf}, the so-called knee-voltage in forward mode. It can be considered as the current value at which high injection in the base starts. It applies for the reverse mode $I_{kr} = Q_{b0}/\tau_r$ accordingly.

Normally the Early factor should be approximately 1, therefore $1 + q_1 \approx \frac{1}{1 - q_1}$. If we neglect the Early factor in the square root we get the usually used formula for the base charge

$$\frac{Q_b}{Q_{b0}} = \frac{1}{2} \cdot \left[\frac{1 + \sqrt{1 + 4 \cdot \left(\frac{i_f}{I_{kf}} + \frac{i_r}{I_{kr}} \right)}}{1 - \frac{V_{BE}}{V_{Earlyf}} - \frac{V_{BC}}{V_{Earlyr}}} \right]. \tag{5.34}$$

Self-Assessment Questions

Question 5.2.1:

What are the main differences between a vertical npn transistor and the lateral pnp transistor?

Question 5.2.2:

What describes the Ebers-Moll transport model?

Question 5.2.3:

Which problems are not modeled with the transport model?

Question 5.2.4:

What makes the essential difference between the Gummel-Poon model and the transport model?

Question 5.2.5:

Of which portions does the base charge consist?

Question 5.2.6:

What is the Early effect and how is it modeled in the Gummel-Poon model?

5.3 Special effects

- Quasi-saturation / the base widening
- Kirk effect (hot carriers)
- Avalanche multiplication
- Emitter-base current crowding
- Access phase shift of transfer current

5.4 Temperature effects

There are three reasons for the temperature dependencies of the behaviour of bipolar transistors (see also module 4.6):

- the intrinsic concentration
- the mobility and accordingly the diffusion constant
- the ionization of the impurity concentration in the base

The following equations are used in the Spice Gummel-Poon model. They are not derived here.

The temperature dependency of the transport saturation current is

$$I_s\left(T\right) = I_{s0} \cdot \left(\frac{T}{T_{nom}}\right)^{X_{TI}} \cdot e^{\left(\frac{T}{T_{nom}}-1\right)\cdot\frac{W_g(T)}{V_T}} \quad and \quad V_T = \frac{k \cdot T}{q}. \tag{5.35}$$

$W_g(T)$ is the temperature dependent bandgap voltage

$$W_g(T) = 1.16 - \frac{7.02e{-}4\cdot T^2}{T+1108} \quad for\ silicon. \tag{5.36}$$

The temperature dependency of the leakage saturation currents are

$$I_{s*}\left(T\right) = I_{s*0} \cdot \left(\frac{T}{T_{nom}}\right)^{\frac{X_{TI}}{n_*}-X_{tb}} \cdot e^{\left(\frac{T}{T_{nom}}-1\right)\cdot\frac{W_g(T)}{n_*\cdot V_T}} \quad *: e \ \ or \ \ c \ . \tag{5.37}$$

The temperature dependency of the current gain is

$$\beta_*\left(T\right) = \beta_{*0} \cdot \left(\frac{T}{T_{nom}}\right)^{X_{tb}} \quad *: f \ \ or \ \ r \ . \tag{5.38}$$

The temperature dependency of the junction capacitances is

$$C_{j*}\left(T\right) = K_{C*}(T) \cdot C_{j*0} \quad *: e, c \ \ or \ \ s \tag{5.39}$$

$$K_{C*}(T) = \frac{1+m_{j*}\cdot\left(4e{-}4\cdot(T-T_{ref})-\frac{V_{j*}\left(\frac{T_{nom}}{T_{ref}}-1\right)+K_{PB}(T_{nom})}{V_{j*}-K_{PB}(T_{nom})}\right)}{1+m_{j*}\cdot\left(4e{-}4\cdot(T_{nom}-T_{ref})-\frac{V_{j*0}\left(\frac{T_{nom}}{T_{ref}}-1\right)+K_{PB}(T_{nom})}{V_{j*0}-K_{PB}(T_{nom})}\right)} \quad *: e, c \ \ or \ \ s \tag{5.40}$$

$$K_{PB}(T) = -\frac{k \cdot T}{q} \cdot \left[3 \cdot \ln \frac{T}{T_{ref}} + \frac{q}{k} \cdot \left(\frac{W_g(T_{ref})}{T_{ref}} - \frac{W_g(T)}{T}\right)\right] \tag{5.41}$$

$$V_{j*}(T) = (V_{j*0} - K_{PB}(T_{nom})) \cdot \frac{T}{T_{nom}} + K_{PB}(T). \tag{5.42}$$

Self-Assessment Question

Question 5.4:

What are the reasons for the temperature dependency of the bipolar transistor?

5.5 The Spice Gummel-Poon Model

The Equations (5.23), (5.24), and (5.34) are the basic equations of the Gummel-Poon model. The structure of the model is derived from the transport model (Figure 5.13).

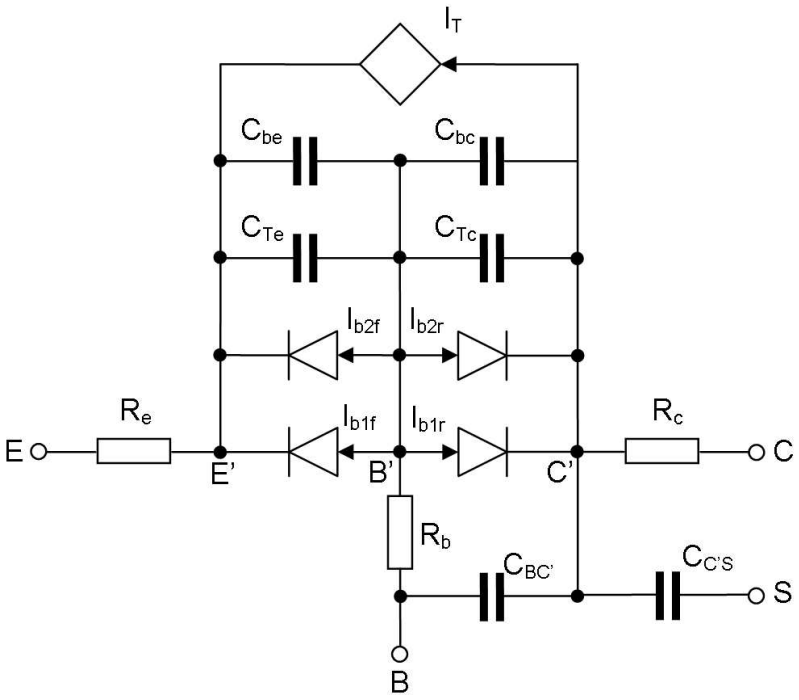


Figure 5.13: Equivalent circuit diagram of the Gummel-Poon model

$$\begin{aligned} i_f &= I_s \cdot \left(e^{\frac{V_{B'E'}}{n_f \cdot V_T}} - 1\right) \\ i_r &= I_s \cdot \left(e^{\frac{V_{B'C'}}{n_r \cdot V_T}} - 1\right) \end{aligned} \tag{5.43}$$

$$I_T = \frac{Q_{b0}}{Q_b} \cdot (i_f - i_r) \quad with \quad \frac{Q_b}{Q_{b0}} = \frac{1}{2} \cdot \left[\frac{1+\sqrt{1+4 \cdot \left(\frac{i_f}{I_{kf}} + \frac{i_r}{I_{kr}} \right)}}{1 - \frac{V_{B'E'}}{V_{Earlyf}} - \frac{V_{B'C'}}{V_{Earlyr}}} \right] \tag{5.44}$$

The base current is built by ideal components:

$$\begin{aligned} I_{b1f} &= \frac{i_f}{\beta_f} \\ I_{b1r} &= \frac{i_r}{\beta_r} \end{aligned} \tag{5.45}$$

and non-ideal base current components. These non-ideal components arise from recombination in the depletion region of forward biased junctions (leakage currents):

$$\begin{aligned} I_{b2f} &= I_{se} \cdot \left(e^{\frac{V_{B'E'}}{n_e \cdot V_T}} - 1 \right) \\ I_{b2r} &= I_{sc} \cdot \left(e^{\frac{V_{B'C'}}{n_c \cdot V_T}} - 1 \right) \end{aligned} \tag{5.46}$$

The parameters of the model for the current Equations (5.43) - (5.46) and (5.34) are (see also Figure 5.15):

I_s	n_f	n_r	β_f	β_r	V_{Earlyf}	I_{kf}	V_{Earlyr}	I_{kr}
I_{se}	I_{sc}	n_e	n_c.					

The diffusion capacitances are

$$\begin{aligned} C_{be} &= \tau_f \cdot \frac{d\, I_{BE}}{d V_{BE}} \\ C_{bc} &= \tau_r \cdot \frac{d\, I_{BC}}{d V_{BC}} \; . \end{aligned} \tag{5.47}$$

The forward transit time parameter τ_f is a function of $V_{B'C'}$ and i_f:

$$\tau_f = \tau_{f0} \cdot \left[1 + x_{\tau_f} \cdot \left(\frac{i_f}{i_f + I_{\tau_f}} \right)^2 \cdot e^{\frac{V_{B'C'}}{V_{\tau_f}}} \right] \tag{5.48}$$

This is a more empirical (curve fitting) function, than derivate from the physical theory of the bipolar transistor.

The space charge layer capacitances are

$$\begin{aligned} C_{Te} &= \frac{C_{je}}{(1 - V_{B'E'}/V_{je})^{m_{je}}} \\ C_{Tc} &= X_{cjc} \cdot C_{Tc0} \\ C_{BC'} &= (1 - X_{cjc}) \cdot C_{Tc0} \\ C_{C'S} &= \frac{C_{js}}{(1 - V_{C'S}/V_{js})^{m_{js}}} \; . \end{aligned} \quad with \quad C_{Tc0} = \frac{C_{jc}}{(1 - V_{B'C'}/V_{jc})^{m_{jc}}} \tag{5.49}$$

The parameters of the model for the capacitances are

τ_{f0} X_{τ_f} V_{τ_f} I_{τ_f} τ_r

C_{je0} V_{je} m_{je} C_{jc0} V_{jc} m_{jc} X_{cjc} C_{js0} V_{js} m_{js}.

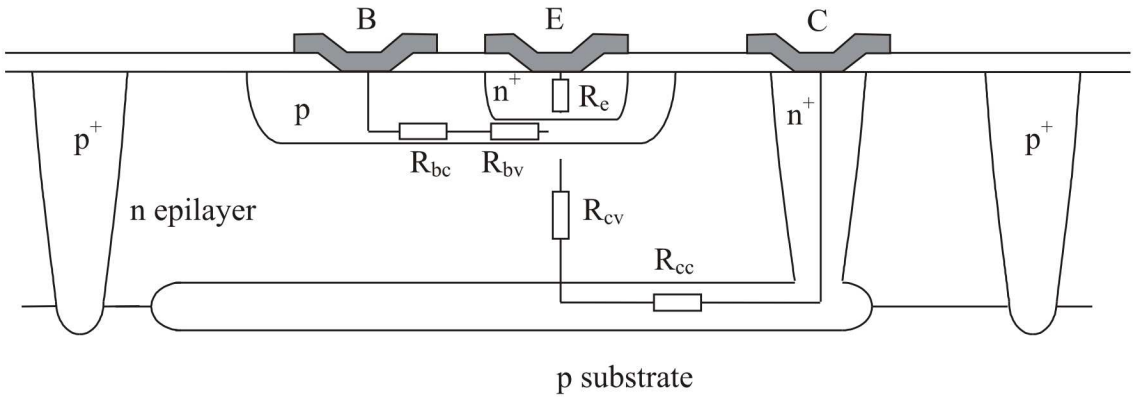


Figure 5.14: Series resistances in an npn-structure

The resistances R_e and R_c are linear and constant ($R_{cv} \approx 0$). But the base resistance has a constant part R_{bc} for the inactive region outside the emitter and a variable part R_{bv} for the base region under the emitter. The last part depends on the base current (emitter-base current crowding). The model equations are noted without derivation (see [4], pp76):

$$R_b = \begin{cases} R_{BM} + \frac{Q_{b0}}{Q_b} \cdot (R_B - R_{BM}) & if\ I_{RB} = 0 \\ R_{BM} + 3 \cdot \frac{(R_B - R_{BM}) \cdot (\tan z - z)}{(\tan z)^2 \cdot z} & if\ I_{RB} \neq 0 \end{cases} \tag{5.50}$$

$$z = \frac{\pi^2}{24} \cdot \frac{\sqrt{1 + \left(\frac{12}{\pi}\right)^2 \cdot \frac{I_B}{I_{RB}}} - 1}{\sqrt{\frac{I_B}{I_{RB}}}} .$$

The parameters of the model for the resistances are

R_B I_{RB} R_{BM} R_e R_c .

All parameters of the Spice Gummel-Poon model are listed in the following table (Figure 5.15):

	name	symbol	parameter	units	default
1	IS	I_{s0}	transport saturation current	A	1.0e-16
2	BF	β_{f0}	ideal maximum forward beta	-	100
3	NF	n_f	forward current emission coefficient	-	1
4	VAF	V_{Earlyf}	forward Early voltage	V	infinite
5	IKF	I_{kf}	corner for forward beta high current roll-off	A	infinite
6	ISE	I_{se0}	B-E leakage saturation current	A	0
7	NE	n_e	B-E leakage emission coefficient	-	1.5
8	BR	β_{r0}	ideal maximum reverse beta	-	1
9	NR	n_r	reverse current emission coefficient	-	1
10	VAR	V_{Earlyr}	reverse Early voltage	V	infinite
11	IKR	I_{kr}	corner for reverse beta high current roll-off	A	infinite
12	ISC	I_{sc0}	B-C leakage saturation current	A	0
13	NC	n_c	B-C leakage emission coefficient	-	2
14	RB	R_B	zero-bias base resistance	Ω	0
15	IRB	I_{RB}	current where base resistance falls halfway to its minimum	A	infinite
16	RBM	R_{BM}	minimum base resistance at high currents	Ω	RB
17	RE	R_e	emitter resistance	Ω	0
18	RC	R_c	collector resistance	Ω	0
19	CJE	C_{je}	B-E zero-bias depletion capacitance	F	0
20	VJE	V_{je}	B-E built-in potential	V	0.75
21	MJE	m_{je}	B-E junction exponential factor	-	0.33
22	TF	τ_f	ideal forward transit time	s	0
23	XTF	$X_{\tau f}$	coefficient for bias dependence of TF	-	0
24	VTF	$V_{\tau f}$	voltage describing VBC dependence of TF	V	infinite
25	ITF	$I_{\tau f}$	high-current parameter for effect on TF	A	0
26	PTF	$\varphi_{\tau f}$	excess phase at freq=1.0/(TF*2PI) Hz	deg	0
27	CJC	C_{jc}	B-C zero-bias depletion capacitance	F	0
28	VJC	V_{jc}	B-C built-in potential	V	0.75
29	MJC	m_{jc}	B-C junction exponential factor	-	0.33
30	XCJC	X_{cjc}	fraction of B-C depletion capacitance connectedto internal base node	-	1
31	TR	τ_r	ideal reverse transit time	s	0
32	CJS	C_{js}	zero-bias collector-substrate capacitance	F	0
33	VJS	V_{js}	substrate junction built-in potential	V	0.75
34	MJS	m_{js}	substrat junction exponential factor	-	0
35	XTB	X_{tb}	forward and reverse beta temperature exponent	-	0
36	EG	W_g	energy gap for temperature effect on IS	eV	1.11
37	XTI		temperature exponent for effect on IS	-	3
38	KF		flicker-noise coefficient	-	0
39	AF		flicker-noise exponent	-	1
40	FC		coefficient for forward-bias depletion capacitance formula	-	0.5
41	TNOM	T_{nom} in K	parameter measurement temperature	C°	27

Figure 5.15: Modified Gummel-Poon BJT parameter

Self-Assessment Question

Question 5.5:

Why is the base resistance not constant?

5.6 Problems and examples

For many standard transistors the parameters are available from semiconductor manufacturers. Generally, however, model parameters must be derived. The measurement techniques leading to the Spice parameters of bipolar transistors are presented in detail in [8]. The extraction of some parameters from simulation results is outlined in the following exercises.

Exercise 5-1: Measurement of the Early voltage

The simulation of the output characteristic of a bipolar transistor can be found in

$BipolarExercise5-1.prb$.

Derive the Early voltage from the simulation of the characteristic $I_C = f(V_{CE}, I_B)$. In Figure 5.16 you can see the schematic of the simulation circuit for the characteristic $I_C = f(V_{CE}, I_B)$ or $I_C = f(I_B, V_{CE})$.

The transistor has the following parameters (Q2N4124):

name	parameter	value	units
IS	transport saturation current	6.734f	A
BF	ideal maximum forward beta	495	-
VAF	forward Early voltage	74.03	V
IKF	corner for forward beta high current roll-off	69.35	A
ISE	B-E leakage saturation current	6.734f	A
NE	B-E leakage emission coefficient	1.28	-
BR	ideal maximum reverse beta	0.7214	-
RB	zero-bias base resistance	10	Ω
RC	collector resistance	1	Ω
CJE	B-E zero-bias depletion capacitance	4.493p	F
TF	ideal forward transit time	301.3p	s
XTF	coefficient for bias dependence of TF	2	-
VTF	voltage describing VBC dependence of TF	4	V
ITF	high-current parameter for effect on TF	0.4	A
CJC	B-C zero-bias depletion capacitance	3.638p	F
MJC	B-C junction exponential factor	0.3085	-
TR	ideal reverse transit time	238.3n	s

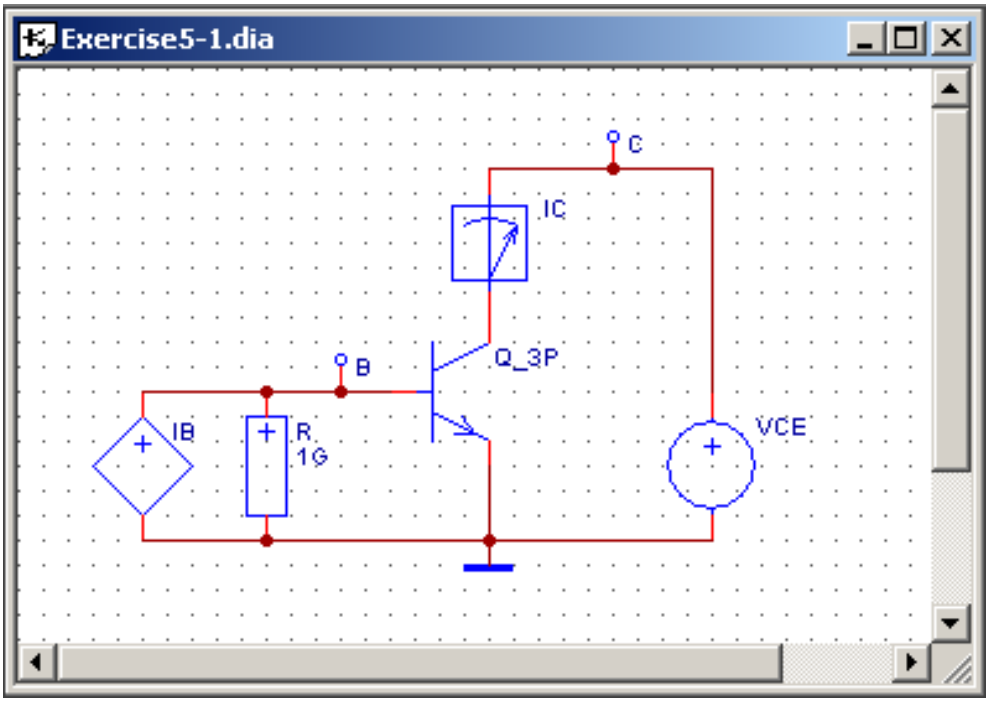


Figure 5.16: Schematic for the simulation of the output characteristic of an npn transistor

The results are shown in Figure 5.17. From this you can compute the forward Early voltage V_{Earlyf} and compare with the given parameter.

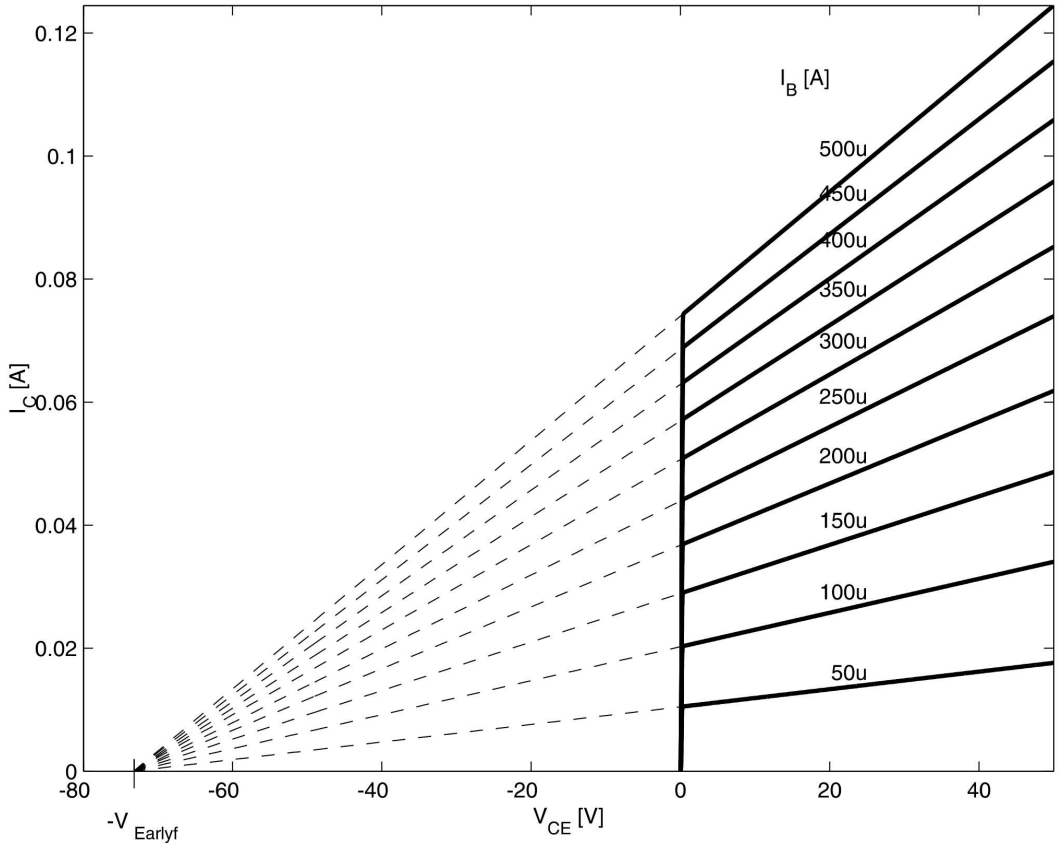


Figure 5.17: $I_C = f(V_{CE}, I_B)$ characteristics with the Early voltage

Exercise 5-2: Measurement of some dc parameters of a bipolar transistor

The simulation example (netlist and simulator commands) can be found in

$BipolarExercise5-2.prb$.

Extract the parameter R_B I_s n_f β_f I_{kf} I_{se} n_e from the dc simulation for the common base configuration in Figure 5.18. The voltage $V_{CB} = 5\,V$ should be constant.

In a first step you must compute the base resistance R_B (see Figure 5.19). The slope for high voltages is approximately $1/R_B$.

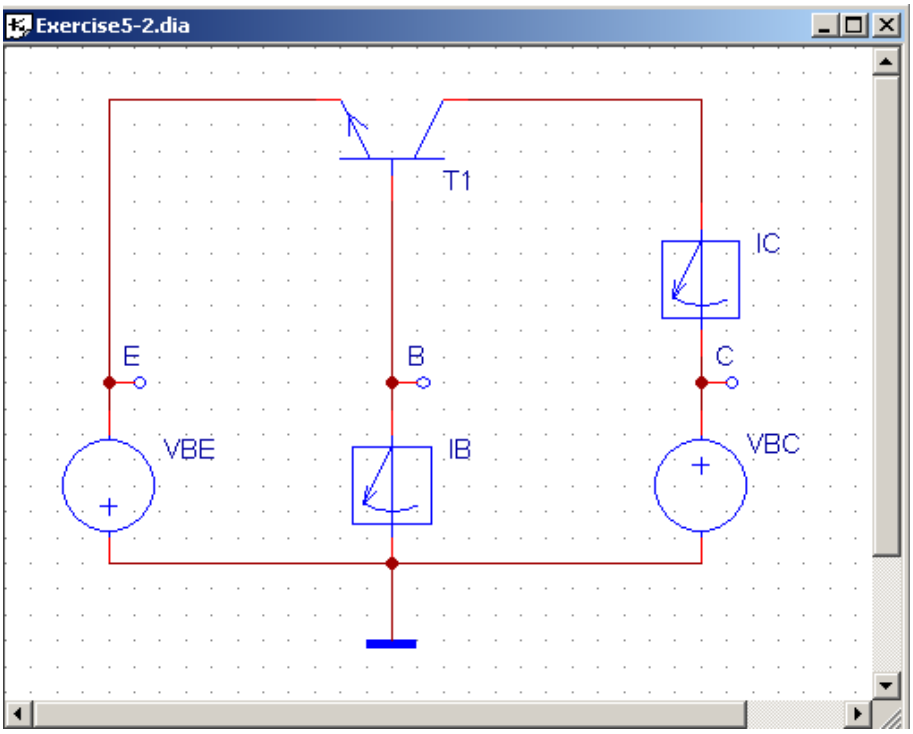


Figure 5.18: Circuit of a bipolar transistor in common base configuration for a dc simulation

For the other parameters the logarithm of the collector and base currents must be plotted as a function of the base emitter voltage (Figure 5.20, the so called Gummel plot):

- Extrapolation of the collector current to $V_{BE} = 0$ gives the parameter I_s.
- Extrapolation of the base current at low base emitter voltage (the asymptote I_{b2f}) gives the parameter I_{se} and at large base emitter voltage (the asymptote I_{b1f}) the value I_s/β_f and therefore the parameter β_f.
- The slopes of the asymptotes in this logarithmic diagram gives the emission coefficients n_f and n_e.
- The asymptotes of I_C for low values of $V_{B'E'}$ (low injection) and high values (high injection) intersect at the current value I_{kf}.

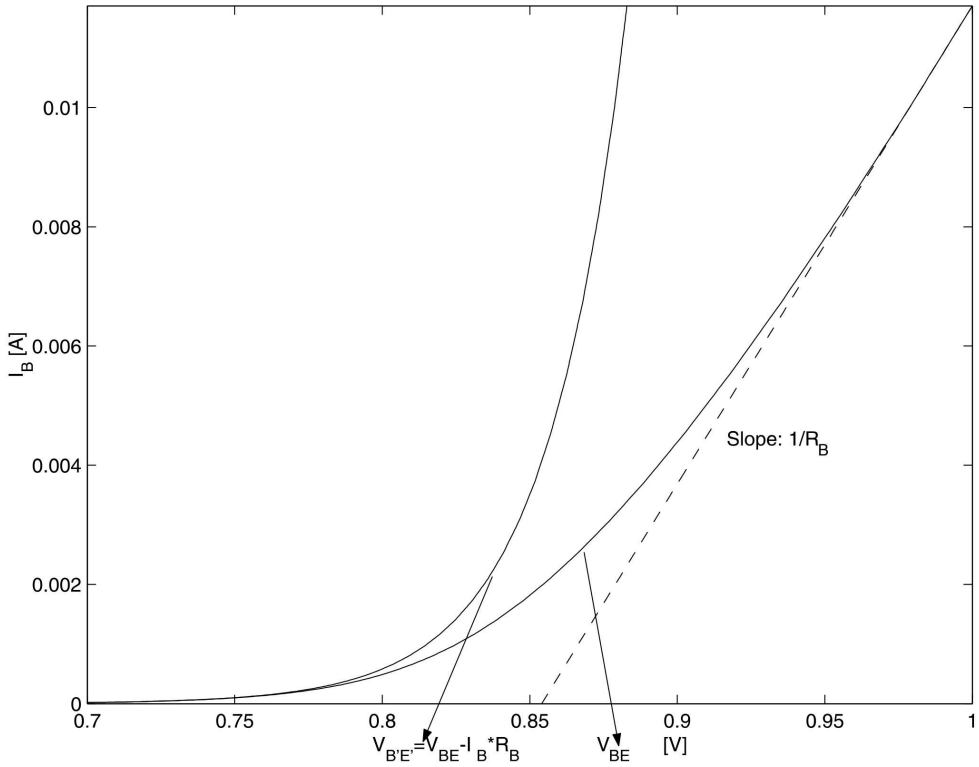


Figure 5.19: The base current $I_B = f(V_{B'E'})$ and the influence of the base resistance

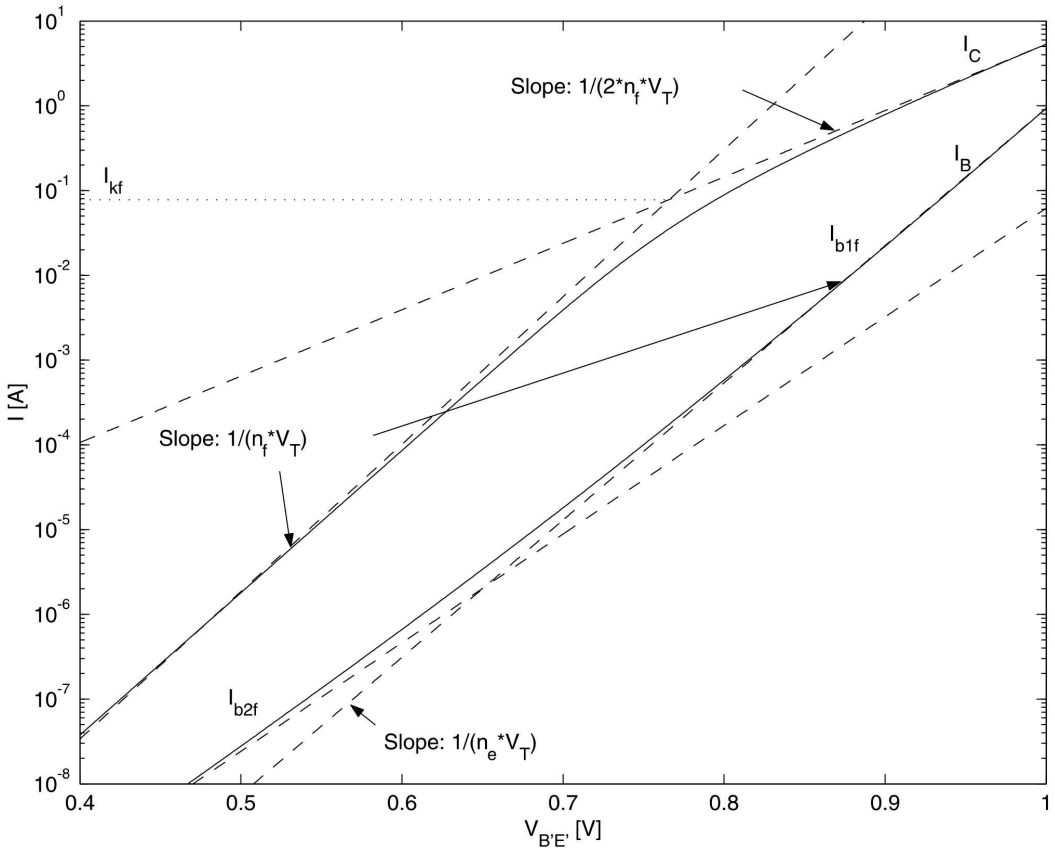


Figure 5.20: The Gummel plot $I_C = f(V_{B'E'})$ and $I_B = f(V_{B'E'})$

Exercise 5-3: Measurement of the ideal forward transit time (ac simulation)

Derive the ideal forward transit time τ_f from the ac simulation of the current gain i_C / i_B (Figure 5.22, $I_B = 20\,\mu A$ and $V_{CE} = 5\,V$):

$\tau_f = \frac{1}{2 \cdot \pi \cdot f_1}$

Set the parameters CJE=0 and CJC=0 and observe the influence on the frequency f_1.
The simulation of the ac characteristic can be found in

$BipolarExercise5-3.prb$

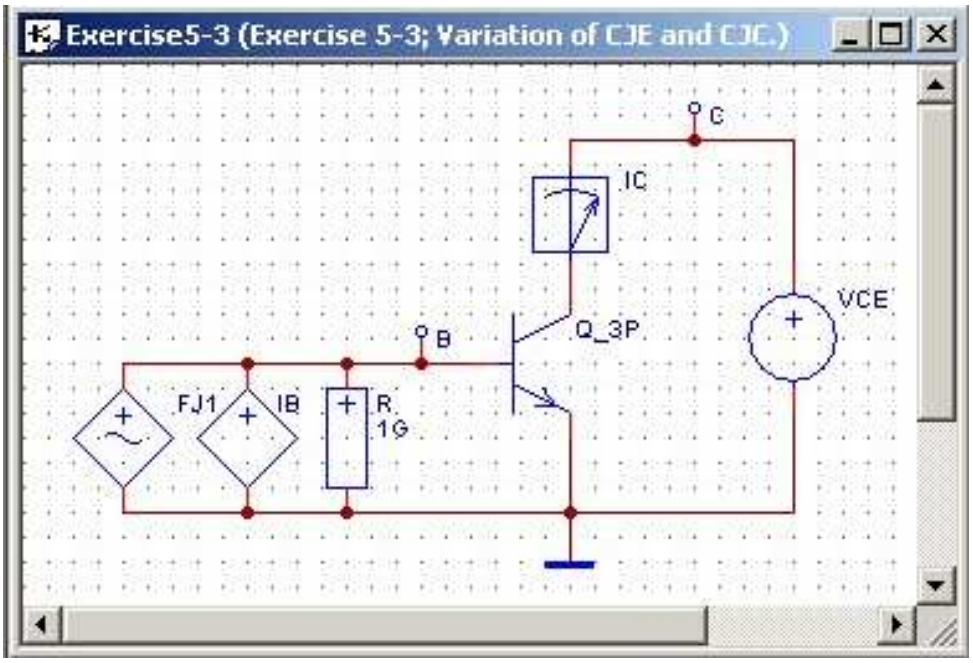


Figure 5.21: Circuit of a bipolar transistor in common emitter configuration for an ac simulation

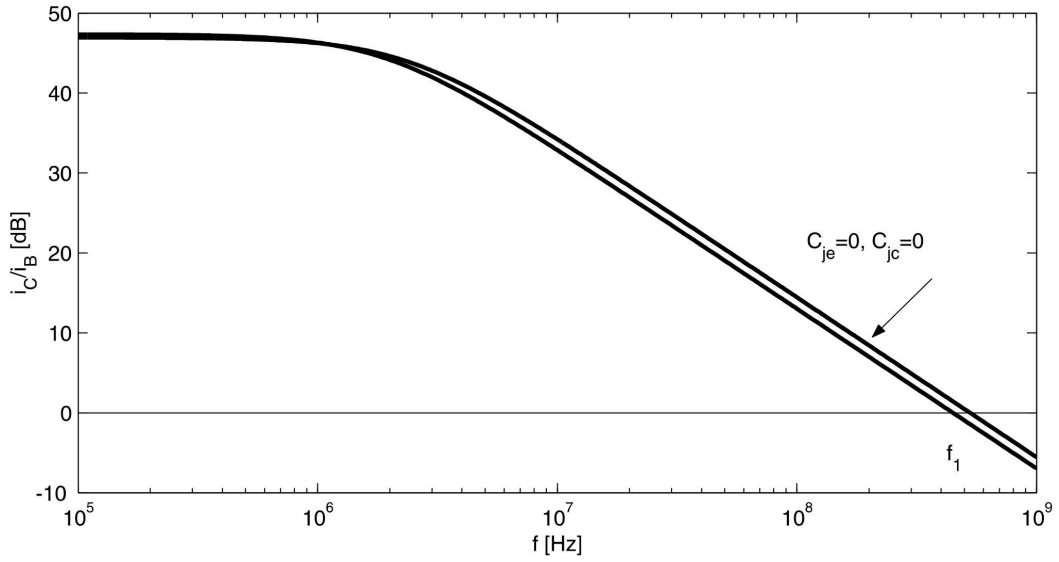


Figure 5.22: The frequency characteristic of the small signal current gain

Module 6

MOS field-effect transistors

Module units

Module overview. MOS field-effect transistors are very important electronic devices, which have founded a dedicated circuitry. For instance, field-effect transistors with an insulated gate electrode are widely used in very large scale integrated (VLSI) circuits, in particular, for applications in the digital technique. To understand the operation principles of MOS devices, some modeling background has to be presented. Starting from phenomenological explanations of the most important physical effects, the derivation of corresponding model expressions is given and sketched, respectively. Because circuit designers tend to view devices with an equivalent circuit, the development of a large signal network model is presented. Among others, stable network models of MOS devices capable for circuit simulations are needed, if the functionality, the performance, the dynamics, etc. of complex MOS circuits have to be checked by network simulations. To avoid too many references within this module, it should be explicitly mentioned, that the content is mainly based on [1], [2].

Module objectives.
Basically, this module is intended to help the model user to understand the concept and the physical meaning of the model equations and parameters. In this way, a better understanding of network models for MOS devices should be achieved.
In particular, this module is considered to answer questions like:

1. What means "field-effect"?

2. What is the contact potential, and how to determine it?

3. How to create an inversion channel?

4. What is the threshold voltage, and which meaning does it have?

5. What means "active channel" operation mode?

6. What is typical for the "pinch-off" operation mode?

7. What makes the difference between an n -channel and a p -channel transistor?

8. What is the MOS transistor's dynamics based on?

Module prerequisites. Knowledge learned in module 3 and module 4, basics in electrical engineering and electronics usually offered in universities' curricula

6.1 Energy-band diagram of the MOS transistor structure

6.1.0 Summary

In the paragraph 3.1.2 we have learned, what an energy-band diagram is, and how to use it for getting deep insights into the internal electronics without numerical analysis. Remember, one of the energy-band's features is to picture the internal electronics under thermal equilibrium conditions. For instance, due to putting together materials with different energy-band diagrams, an internal electric field will occur in such semiconductor structures. This is also the case in MOS structures where the so-called contact potential has to be taken into account. Moreover, from the energy-band diagram we can identify the necessary conditions that an inversion channel can be created. So, we can recognize, what field-effect means.

Consequently, these and other facts have to be considered if we analyze the operation principles of MOS transistors with an insulated control electrode.

6.1.1 MOS transistor structure under thermal equilibrium

To understand the basic electronics of MOS transistors (enhancement type), at first, we consider a MOS transistor with a homogeneously p -doped (acceptor density N_A) semiconductor (bulk) shown in Figure 6.1.

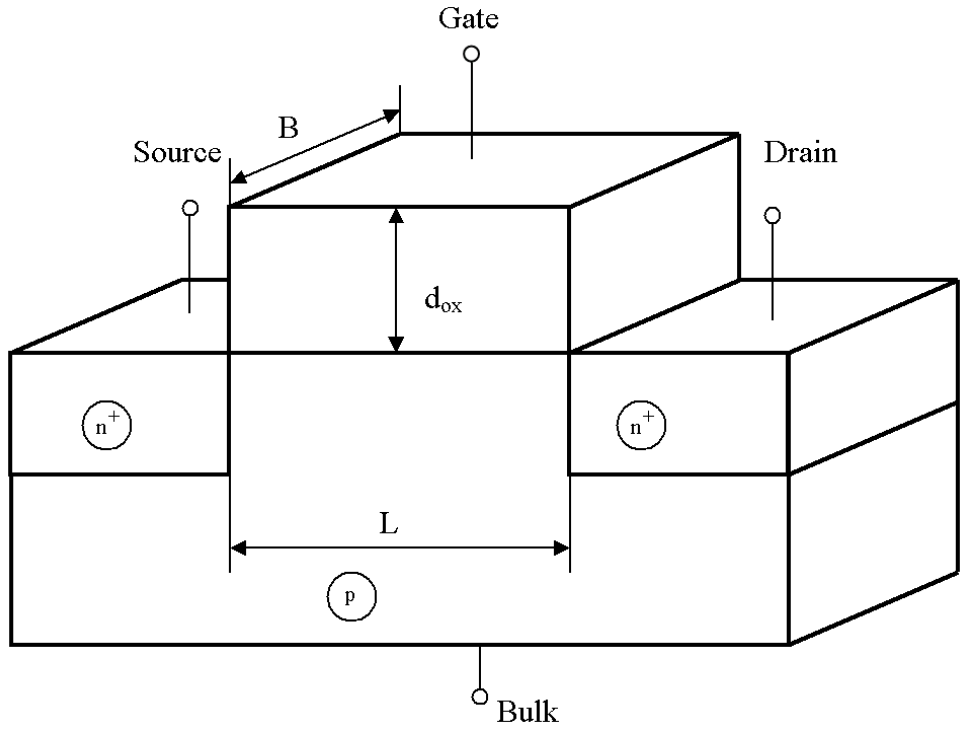


Figure 6.1: MOS transistor structure

In paragraph 3.1.2 we introduced the energy-band diagram of a MOS structure with a metallic gate (Al) under thermal equilibrium (see Figure 3.6). Now, we are going to repeat the construction again for an MOS structure with a highly-doped Si gate. To meet the thermal equilibrium conditions, by definition, we start from $V_{GB} = V_{GS} = V_{GD} = 0$. In this case the Fermi-level W_F is constant throughout the structure (see Figure 6.2). As shown in Figure 6.2, however, that is not the case for the band edges W_V and W_C as well as the intrinsic level W_i. These band deflections, particularly, at the interface between oxide and semiconductor bulk, indicate an internal electric field with a drop of voltage on the oxide (V_{io}) and also on the bulk (V_{so}).

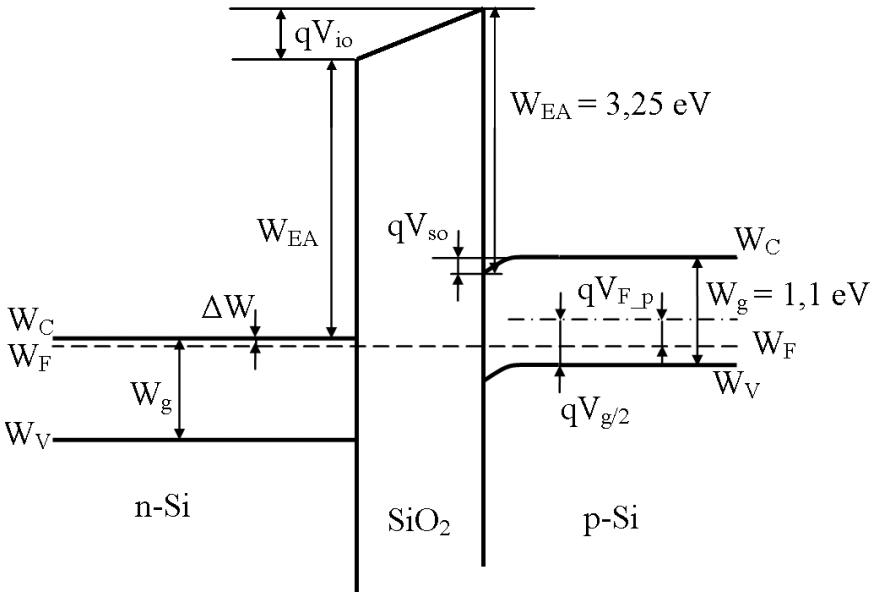


Figure 6.2: Energy-band diagram of a MOS structure under thermal equilibrium [1]

The total voltage drop is a built-in potential and is called contact potential V_K, where

$$V_K = V_{io} + V_{so}. \tag{6.1}$$

V_{io} and V_{so} are the potential differences across the gate insulator and the semiconductor under thermal equilibrium, respectively.

From Figure 3.6 we find

$$V_K = V_{EA} - V_{WF\_Al} + V_{F\_p} + V_{g/2} \tag{6.2}$$

for a metal gate electrode where $V_{EA} = \frac{W_{EA}}{q}$, $V_{WF\_Al} = \frac{W_{WF\_Al}}{q}$ and $V_{g/2} = \frac{W_g}{2 \cdot q}$ are the voltage equivalents to the electron affinity W_{EA}, the metal work function W_{WF}, and the half $W_g/2$ of the forbidden band W_g. The Fermi-potential V_{F_p} (potential difference to intrinsic potential) in the bulk is given by

$$V_{F\_p} = V_T \cdot \ln \frac{N_A}{n_i}. \tag{6.3}$$

As usually, V_T is the thermal voltage, and n_i is the intrinsic density of silicon.

The contact potential V_K of MOS structures plays a similar role as the built-in potential V_{bi} in a pn -junction. This will be clear if we determine V_K for a MOS structure with a heavily n^+ -doped polysilicon gate (donor density N_D). Then we get a Fermi potential V_{F_n}

$$V_{F\_n} = V_T \cdot \ln \frac{N_D}{n_i}. \tag{6.4}$$

For such structure and $\Delta W \approx 0$ (due to the high concentration of donors in the gate), from Figure 3.6 we get

$$V_K = V_{F\_n} + V_{F\_p} = \frac{W_g}{2 \cdot q} - \frac{\Delta W}{q} + V_{F\_p} \approx \frac{W_g}{2 \cdot q} + V_{F\_p} = V_{g/2} + V_{F\_p} \tag{6.5}$$

with

$$\frac{\Delta W}{q} = \frac{W_C - W_F}{q} = V_T \cdot \ln \frac{N_C}{N_D} << \frac{W_g}{2 \cdot q}. \tag{6.6}$$

As introduced and already explained in paragraph 3.1.3, $N_C \approx 10^{19}\ cm^{-3}$ is the effective density of states in the conduction band. Donor densities above N_C make the Fermi-level dip into the conduction band, and such semiconductor is called "degenerated", but this case will not play any role in modeling the functional principle of MOS transistors. From Equation (6.5) we recognize the similarity between V_K and V_{bi}.

6.1.2 Field-effect control

Because their numerous applications, we are particularly interested in modeling the electronics of n -channel MOS transistors (NMOS transistors). In such device, the inversion channel consists of electrons enhanced at the interface between the silicon oxide and the surface of a p -doped semiconductor (bulk).

For our modeling purposes, we may use the modeling approach called "strong" inversion approximation. This model works when a remarkable (strong) inversion channel of electrons has been "born". Basically, an inversion channel only appears, if the gate bulk voltage $V_{GB} > 0$ exceeds a typical threshold voltage V_t. To understand, what strong inversion means, and how to define this threshold voltage, we use the energy-band diagram again . For better understanding we will proceed step-by-step.

At first we supply a positive gate-bulk voltage $V_{GB} > 0$ to the gate.

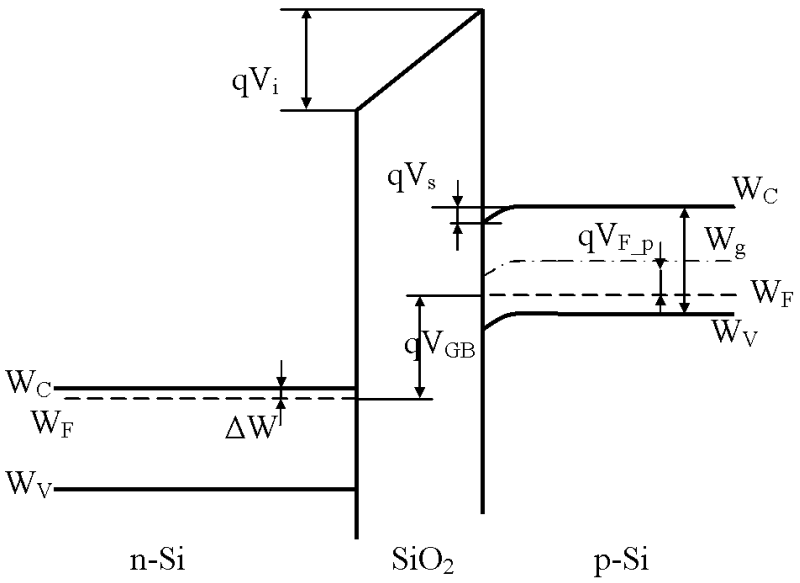


Figure 6.3: Energy-band diagram with gate voltage [1]

As shown in Figure 6.3, now, we have a difference in the Fermi-levels of the gate and the semiconductor. Under consideration of knowledge given in paragraph 3.1.2 we learn from Figure 6.3

$$V_i + V_s = V_{GB} + V_K \tag{6.7}$$

where V_i and V_s are the drop of voltage on the oxide and semiconductor, respectively, if $V_{GB} > 0$ has been supplied. V_K is the contact potential given by Equation(6.1).

A detailed analysis of the field and potential distribution in a MOS structure shows that beyond beginning strong inversion the surface potential V_s (corresponds to the drop of potential on the semiconductor) holds on its value $V_s \approx 2 \cdot V_{F\_p}$ (see Figure 6.4).

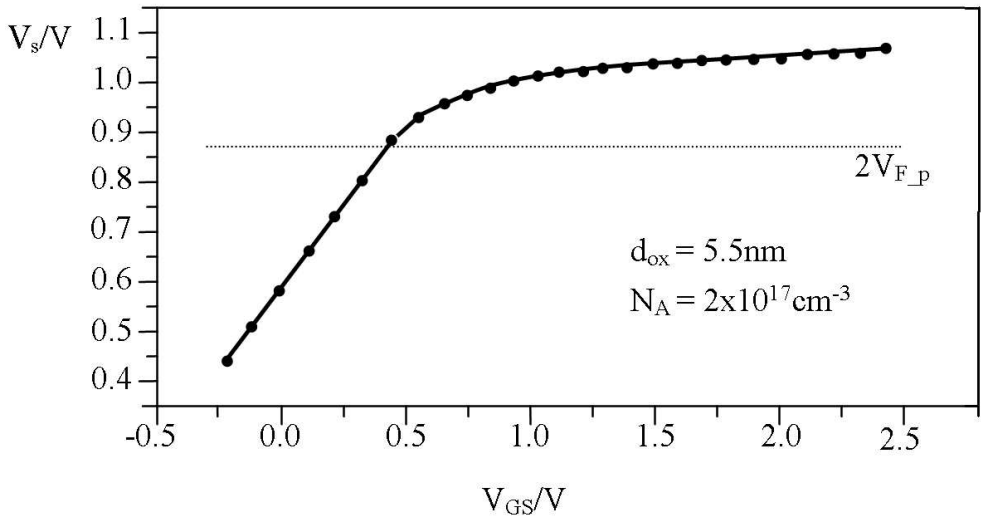


Figure 6.4: Surface potential as function of the gate source voltage [9]

Additionally to $V_{GB} \geq 0$, now, we supply a source-bulk voltage $V_{SB} \geq 0$ and hold on $V_{DB} = 0$. With $V_{SB} \geq 0$ the Fermi-level of the bulk is split-up in a quasi-Fermi level W_{F_p} for holes and such one for electrons W_{F_n}, respectively (see Figure 6.5).

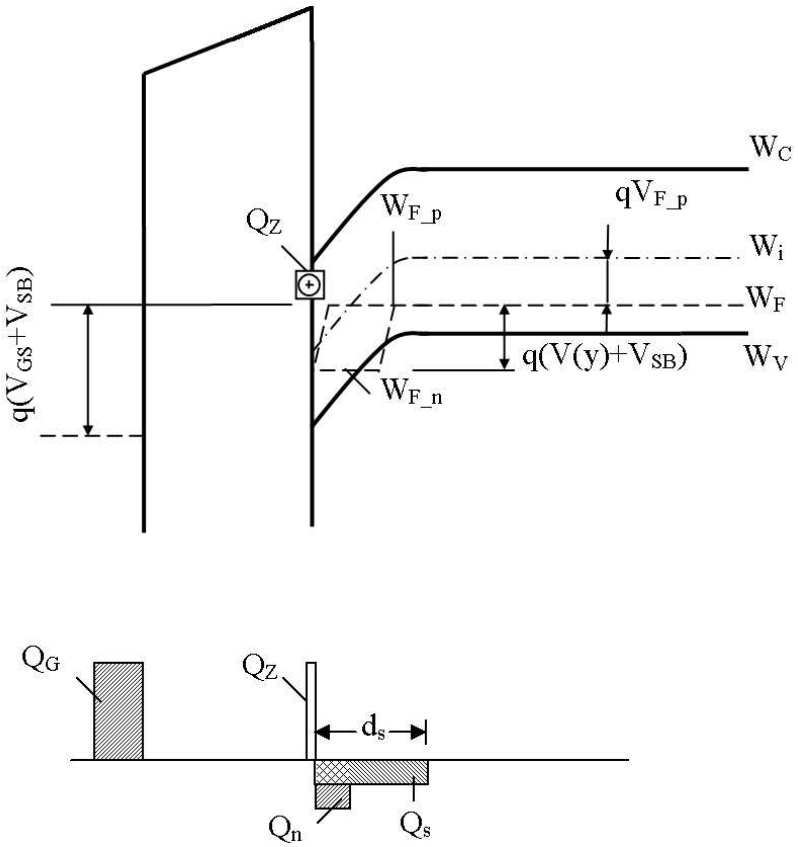


Figure 6.5: Energy-band diagram with gate and source bulk voltage [1]

So, the quasi Fermi-level of the source electrons and the Fermi-level of the bulk differs just about the externally supplied voltage $V_{SB} \geq 0$. As shown in the energy-band diagram in Figure 6.5, a similar situation can be recognized regarding $q \cdot V_{F\_p}$ and the quasi Fermi level of the electrons W_{F_n} in the inversion channel (see Figure 6.5).

We may not forget that the influenced channel electrons will shield the bulk against a further penetration of field from the oxide side. Therefore, the total voltage drop on the bulk V_s remains

$$V_s \approx 2 \cdot V_{F\_p} + V_{SB}. \tag{6.8}$$

As V_{GB}, V_s and V_K have been specified, now we are able to calculate the voltage drop on the oxide V_i:

$$V_i = V_{GB} - V_{SB} - 2 \cdot V_{F\_p} + V_K = V_{GS} - 2 \cdot V_{F\_p} + V_K, \tag{6.9}$$

where $V_{GS} = V_{GB} - V_{SB} > 0$ is the gate-source voltage. The oxide voltage V_i and the surface potential V_s are the essential controlling voltages (potentials) of this so-called "active" channel modus.

With the third step, finally, a drain-bulk voltage $V_{DB} \geq V_{SB} \geq 0$ is also supplied (see Figure 6.6). That means an electric drift field $\vec{E}_{ch}$ is impressed on the channel

$$\vec{E}_{ch} = \vec{j} \cdot E_{ch} \, (0 \leq y \leq L). \tag{6.10}$$

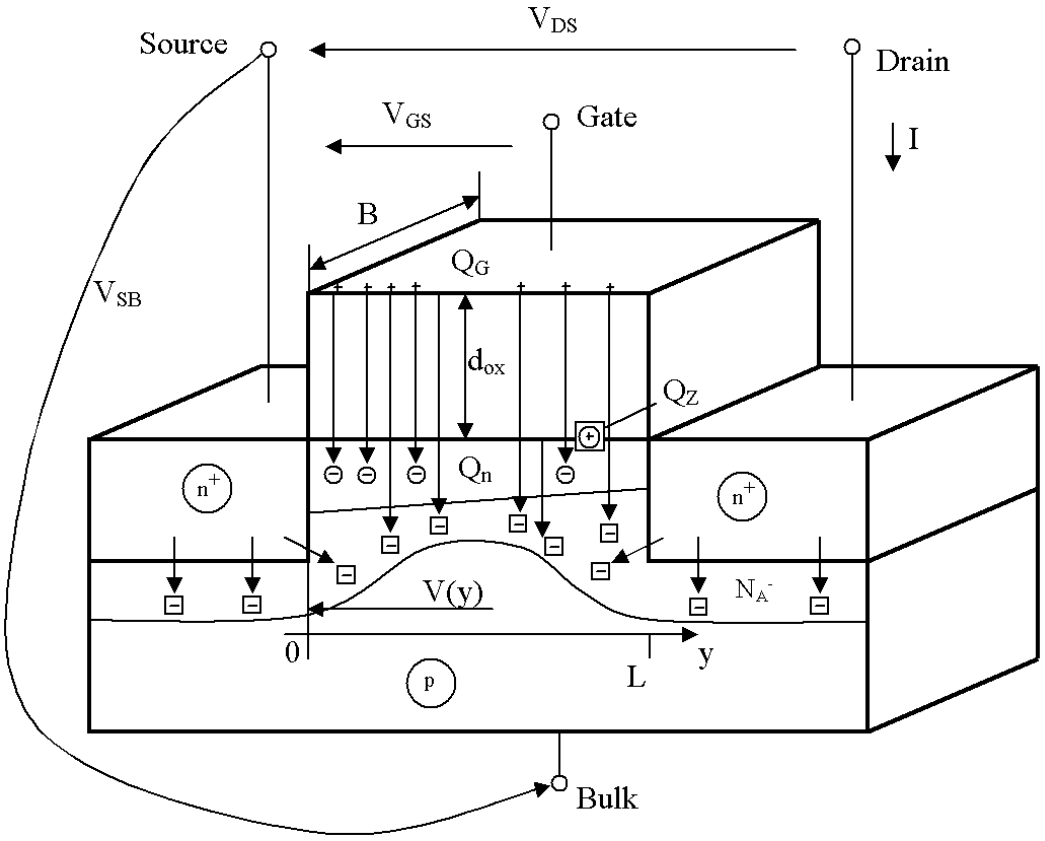


Figure 6.6: NMOS transistor with supplied terminal voltages

This electric field make the channel electrons move. Thereby, a current flow from drain to source is created, and necessarily, a channel potential $V(y)$ occurs.

Basically, the drift field $E_{ch}(y)$ is considered to be small in comparison to the transversal electric field in the oxide. Therefore, we may assume that the electric field in the oxide holds on having a transversal component only, but it will decline towards the drain.

Self-Assessment Question

Question 6.1:

Which quantity achieved the surface potential V_s at the beginning of strong inversion?

6.2 Control charges in the NMOS transistor

6.2.0 Summary

Basically, there are two items which make us studying the control charges in MOS transistors. The first reason is to become acquainted with the working principle of the field effect transistors. In particular, this concerns the control charges on the gate, the interface charges which are typical for the technology used, and also the space charge in bulk that governs the threshold voltage. On the other hand, the changing of charges due to supplied voltages, causes capacitive effects. Such capacitances have to be considered for modeling the dynamics of electronic devices. For this purpose, the charges in the space charge layers at the pn -junctions around the source and drain contact regions have to be involved in the analysis, too.

6.2.1 Field-effect based conductivity and charge control

Based on the voltage control and the charge neutrality we find

$$Q''_G + Q''_Z + Q''_s + Q''_n = 0. \tag{6.11}$$

Next, we will qualify the charge contributions Q''_G, Q''_s and Q''_Z needed for the determination of the channel charge Q''_n. In Figure 6.7 these charges are pictured for different gate-source voltages.

For the charge (per unit area) on the gate Q''_G we find

$$Q''_G\left(0 \le y \le L\right) = \varepsilon_{ox} \cdot \frac{V_i}{d_{ox}} = \varepsilon_{ox} \cdot \frac{V_{GS} - V\left(y\right) - 2 \cdot V_{F\_p} + V_K}{d_{ox}} > 0. \tag{6.12}$$

The charge (per unit area) in the space charge layer Q''_s is given by

$$Q''_s = -q \cdot N_A \cdot d_s < 0 \tag{6.13}$$

where the layer's depth d_s depends on the channel potential $V\left(y\right)$. For modeling approach purposes, we may neglect the impact of $V\left(y\right)$ on the bulk space charge. At the source is $V\left(y = 0\right) = 0$, and the bulk charge (per unit area) is (see Equation (6.13) and Equation(6.8))

$$Q''_s = -q \cdot N_A \cdot d_s = -q \cdot N_A \cdot \sqrt{\frac{2 \cdot \varepsilon_H}{q \cdot N_A} \cdot V_s} = -\sqrt{2 \cdot q \cdot N_A \cdot \varepsilon_H \cdot \left(V_{SB} + 2 \cdot V_{F\_p}\right)} < 0. \tag{6.14}$$

The interface state charge Q''_Z, is considered to be constant, e.g.

$$Q''_Z = 10^{-8} As/cm^2. \tag{6.15}$$

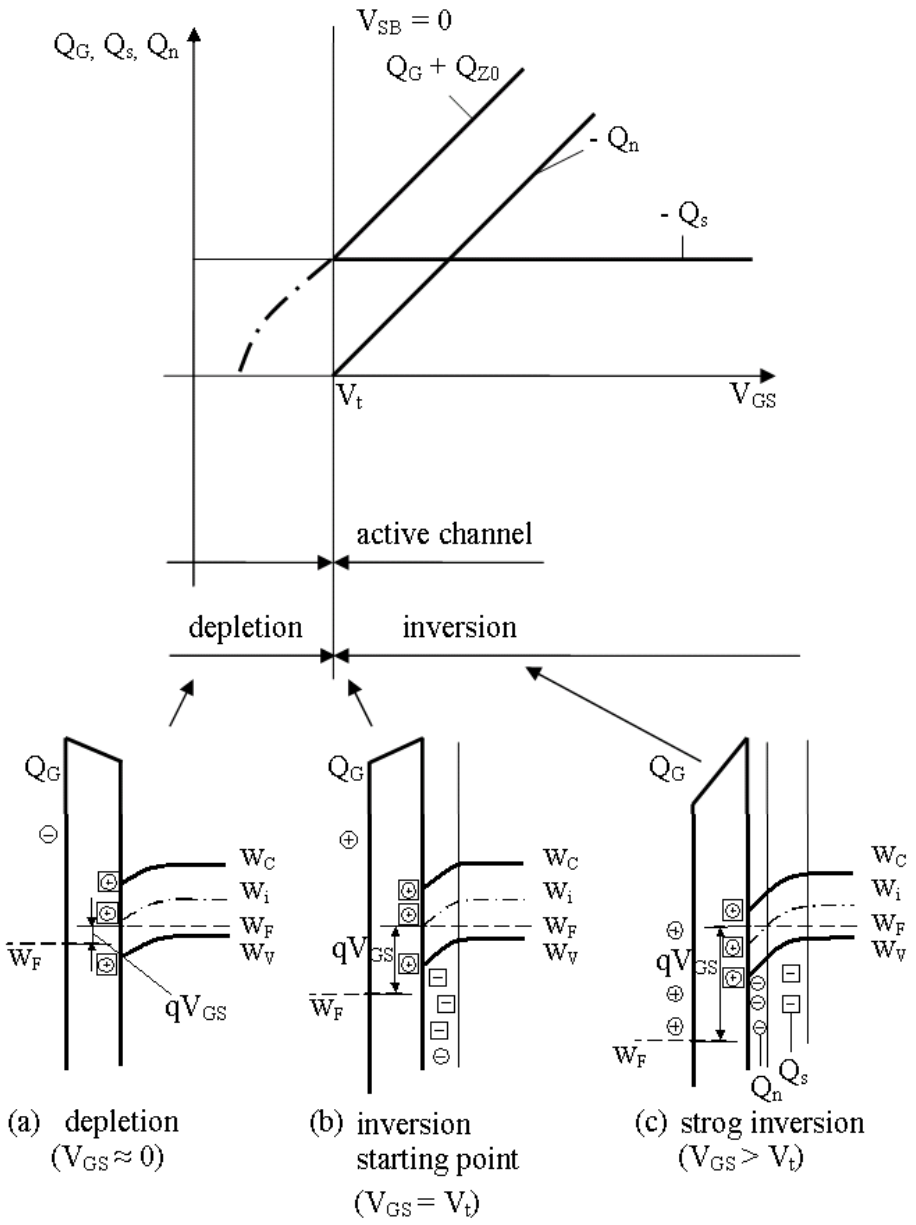


Figure 6.7: Control charges in a MOS structure [1]

Taking into account Equations (6.12) to (6.15), finally, from Equation (6.11) we receive the electron charge (per unit area) in the channel Q''_n. The result can be given by

$$Q''_n\left(V_{GS} \geq (V(y) + V_t) > 0\right) = -\frac{\varepsilon_{ox}}{d_{ox}} \cdot (V_{GS} - (V(y) + V_t)) \leq 0. \tag{6.16}$$

Thus, we find the threshold voltage $V_t > 0$

$$V_t = 2 \cdot V_{F\_p} + V_{FB} + \frac{d_{ox}}{\varepsilon_{ox}} \cdot \sqrt{2 \cdot q \cdot N_A \cdot \varepsilon_H \cdot (V_{SB} + 2 \cdot V_{F\_p})} > 0 \tag{6.17}$$

where the technology-based flat band voltage V_{FB} stands for

$$V_{FB} = -\left(V_K + \frac{d_{ox}}{\varepsilon_{ox}} \cdot Q''_Z\right). \tag{6.18}$$

The contact potential V_K has already been defined and discussed in paragraph 6.1.1. It is given by

Equation (6.2) and Equation (6.5), respectively.

From this point of view, the threshold voltage $V_t > 0$ can be understood as minimum gate-source voltage to create a strong inversion channel starting at the source. In other words, for gate-source voltages $0 < V_{GS} < V_t$ the inversion channel disappears.

Self-Assessment Question

Question 6.2:

Which charges have to be considered to calculate the inversion channel conductivity?

6.3 Operation Principles of MOS Transistors

6.3.0 Summary

Generally, the operation principle of MOS transistors with an insulated gate electrode is based on the control of a conducting channel at (or near) the interface between the semiconductor and the oxide. Basically, there are four types of MOS field-effect transistors. So, we distinguish the n - and p -channel enhancement types from the n - and p -channel depletion transistors. Whereas the depletion types are devices for special applications, the enhancement transistors have a great importance for discrete and integrated circuits. In particular, they are used together in integrated CMOS IC. That is why we have focused our thoughts and discussions on the enhancement transistors only.

6.3.1 Types of MOS transistors

The n -channel enhancement transistor (NMOS) in Figure 6.8(a) is fabricated on a p -type semiconductor (bulk) with heavily-doped n -regions as source (S) and drain (D) regions. A very thin silicon oxide layer insulates the control electrode (called gate G) from the semiconductor surface. If a high enough positive gate-source voltage, $V_{GS} > V_t > 0$, is supplied between gate and source, the so-called "field-effect" works. That means, the strong electric field originating from positive gate charges penetrates into the semiconductor and influences a conducting inversion channel at surface. In a p -type bulk it consists of electrons enhanced at the interface between oxide and the semiconductor surface. So, this channel connects source and drain and allows an electron current I, if a positive drain-source voltage V_{DS} is additionally supplied. This current can be controlled by $V_{GS} > V_t$, where $V_t > 0$ is the so-called threshold voltage. V_t is the minimum gate source voltage to create a strong inversion channel.

If we use an n -type substrate with p -type source and drain regions, we obtain a p -channel enhancement transistor (see Figure 6.8(b)). Fore these types the signs of all voltages and currents are reversed, but the operating principle is the same as of the n -channel enhancement transistor.

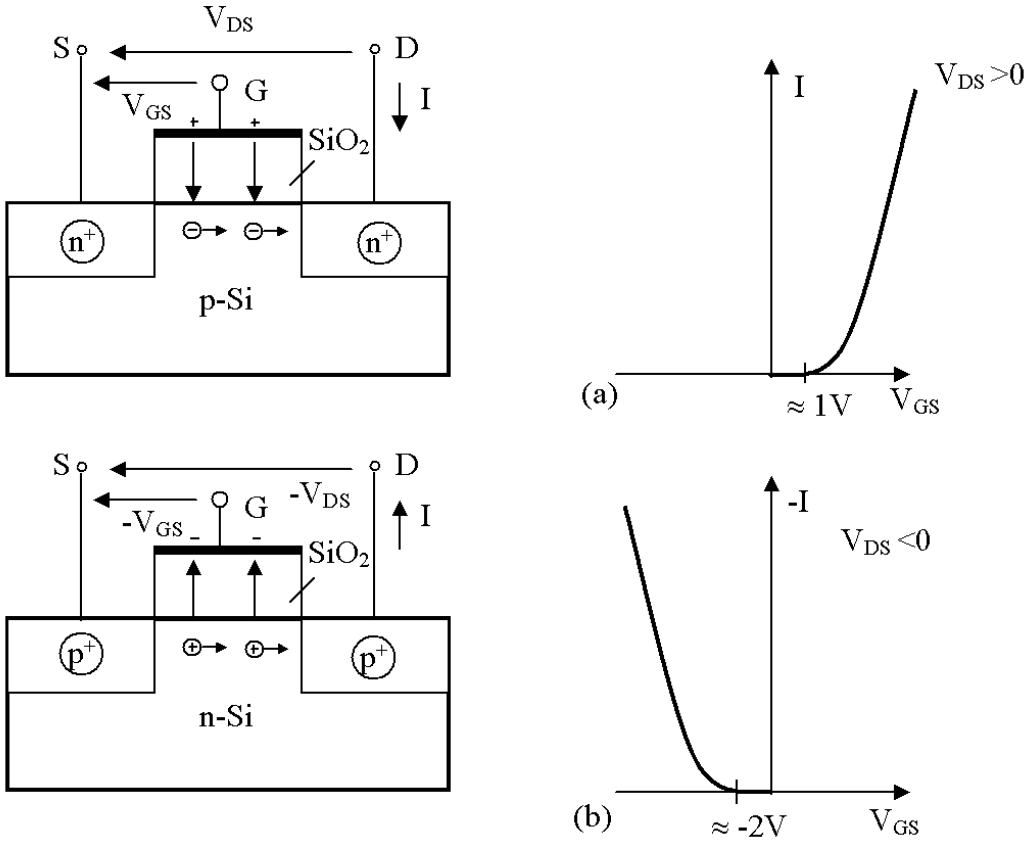


Figure 6.8: Different types of MOS transistors: (a) n -channel enhancement, (b) p -channel enhancement [2]

Self-Assessment Question

Question 6.3:

Is there an important difference referring to the operation principles of a p-channel and an n-channel enhancement MOS transistor?

6.4 I/V-characteristic for strong inversion

6.4.0 Summary

Based on the voltage control (see paragraph 6.1.2) we studied the forming of corresponding control charges (see paragraph 6.2.1) due to the supplied terminal voltages gate-bulk V_{GB}, source-bulk V_{SB} and drain-bulk V_{DB}. Thus, we have learned that for $V_{GS} = V_{GB} - V_{SB} \geq V_{th} > 0$ an inversion channel of electrons is enhanced at the interface between bulk and silicon oxide. If we choose $V_{DS} = V_{DB} - V_{SB} > 0$ an electric drift field is imposed on the channel which drives the electrons in the channel to drain, i.e. a channel current $I \geq 0$ flows from drain to source. This phenomenon will be discussed in detail, next.

Summarizing what we have learned up to now, we should recognize the following:

The inversion channel influenced by the electric field is extremely thin and only contains electrons in a very high concentration. It contacts the n^+ -doped source and drain regions which provide the channel with electrons and absorb them, respectively. Moreover, the source and drain are also necessary to pick-up the channel current I. These facts turn MOS transistors into the majority of carrier controlled devices, which are usually "fast" devices, i.e. the electronic device works nearly all times under steady state conditions. This characteristic feature will be discussed in detail later on.

6.4.1 Unipolar channel current

The current in an n -(inversion) channel is a pure electron drift current.

As developed in paragraph 6.2.1, the channel conductivity due to Q_n'' (influenced by the field-effect) depends on the channel potential $V(y)$ (see Equation(6.16)). Moreover, the potential's changing rate along the channel defines the electric drift field. This combination makes the MOS transistor's functional principle becomes quadratic-nonlinear. That makes a remarkable difference to the bipolar transistor. Actually, the electronics within the base of bipolar transistors is linear, however, the injection of minority carriers from the emitter and collector into the base is nonlinearly (exponentially) dependent on the control voltages supplied between emitter and base, and collector and basis, respectively.

The I/V -characteristic of an NMOS transistor can be obtained by the integration of the drift current equation

$$I = \mu_n \cdot B \cdot (-Q_n''(V)) \cdot \frac{d}{dy} V. \tag{6.19}$$

Under consideration of Equation (6.16) with Equation (6.17) and Equation(6.18), we obtain the I/V -characteristic from

$$I \cdot \int_0^L dy = \mu_n \cdot B \cdot \frac{\varepsilon_{ox}}{d_{ox}} \cdot \int_{V(0)=0}^{V(L)=V_{DS}} (V_{GS} - V_t - V) \cdot dV \tag{6.20}$$

with the threshold voltage V_t

$$V_t = 2 \cdot V_{F\_p} + V_{FB} + \frac{d_{ox}}{\varepsilon_{ox}} \cdot \sqrt{2 \cdot q \cdot N_A \cdot \varepsilon_H \cdot (V_{SB} + 2 \cdot V_{F\_p})} > 0 \tag{6.21}$$

and the flat-band voltage V_{FB}

$$V_{FB} = -\left(V_K + \frac{d_{ox}}{\varepsilon_{ox}} \cdot Q_Z''\right). \tag{6.22}$$

Due to the contact potential V_K (see Equation (6.2) and Equation(6.5)) and the interface charges Q'', the flat-band voltage V_{FB} becomes material and technology specific (see paragraph 6.1.1 and 6.2.1).

The integration in Equation (6.20) results in the NMOS transistor's I/V -characteristic, given by

$$I = \mu_n \cdot \frac{\varepsilon_{ox}}{d_{ox}} \cdot \frac{B}{L} \cdot \begin{cases} 0 & V_{GS} - V_t \le 0 \\ (V_{GS} - V_t) \cdot V_{DS} - \frac{V_{DS}^2}{2} & 0 < V_{DS} \le V_{GS} - V_t \\ \frac{(V_{GS} - V_t)^2}{2} & 0 < V_{GS} - V_t < V_{DS} \end{cases} . \tag{6.23}$$

The complete I/V -characteristic is pictured in Figure 6.9. On the left hand side, the transfer characteristic is presented, and on the right hand side the output characteristic is given.

As shown in Equation (6.21), the threshold voltage depends on the source bulk voltage V_{SB}. The effect based on this V_{SB} -dependency is called body-effect.

The modeling approach pursued up to now, has been relied on a strong inversion channel starting at source end ending at drain. This situation is the so-called "active channel" operation mode. It is characterized by the voltage relation $0 < V_{DS} \le V_{GS} - V_t$. If this relation reverses, i.e. $0 < V_{GS} - V_t < V_{DS}$, the channel is said to be "pinched-off".

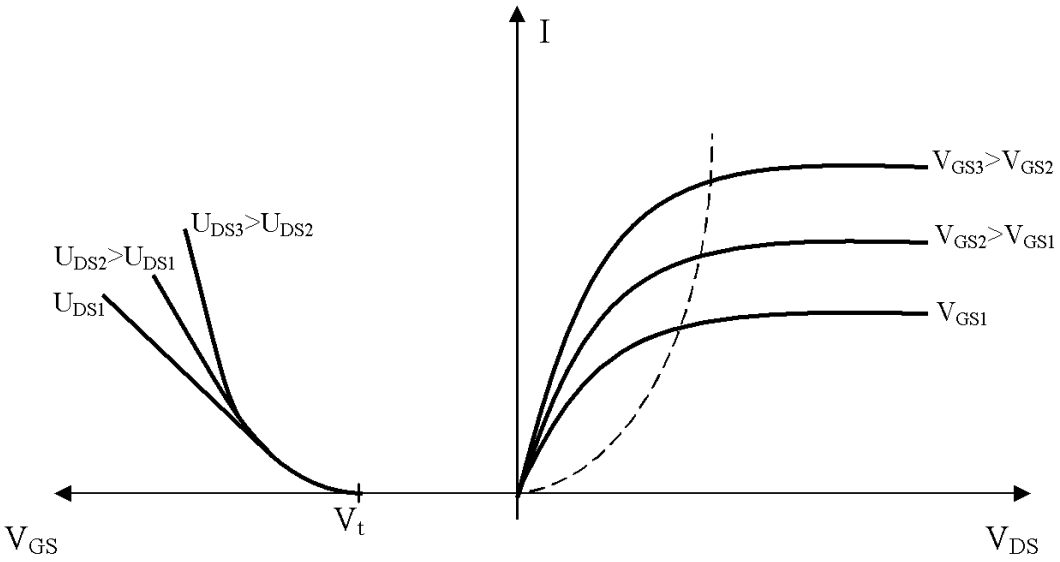


Figure 6.9: I/V -characteristics of NMOS transistors [2]

The simplest model to describe the current flow in the pinch-off mode is derived from the minimum drain-source voltage $V_{DS} = V_{GS} - V_t$ necessary to pinch-off the channel just at drain. For $0 < V_{GS} - V_t < V_{DS}$ this basic model defines the channel current to become constant and independent on V_{DS}. Its value is corresponding to I that comes out if we expand the active operation mode to its limit $V_{DS} = V_{GS} - V_t$.

This conflicts with the model assumptions which are based on weak channel fields. "Weak" means that the channel field can be neglected compared to the field originating from gate, which is necessary to create the channel. A realistic modeling procedure has to take the interaction between the channel and oxide field into account as soon as their strengths have achieved comparable magnitudes.

These facts are crucial. That is why we have to refine our model concerning the electronic in the pinch-off operation mode.

In our I/V -model (see Equation(6.23)) the out-put resistance is extremely high in the pinch-off mode. From Equation (6.23) we calculate the output conductance in the pinch-off mode $g_{out} = (dI/dV_{DS})_{V_{DS} > V_{GS} - V_t} = 0$, i.e., the output resistance tends to become infinite. Real devices exhibit pinch-off out-put resistances which are much more lower (see Figure 6.10) than infinite.

That is due to dedicated pinch-off effects to be mentioned, now. A phenomenological model relays on a new declaration of how to understand the channel length L .

Clearly, in the active operation mode L is a design (structure) parameter which denotes the distance between source and drain. However, in the pinch-off mode the channel length turns to become an electronic parameter controlled by V_{DS}. That means, in the pinch-off operation mode the channel length L has to be replaced by an electronic channel length $L_e = L - \Delta L_e\,(V_{DS}) > 0$. Actually, this channel shorting makes the out-put current arise with increasing $V_{DS} > 0$. Often, this approach is compared with the Early effect of bipolar transistors.

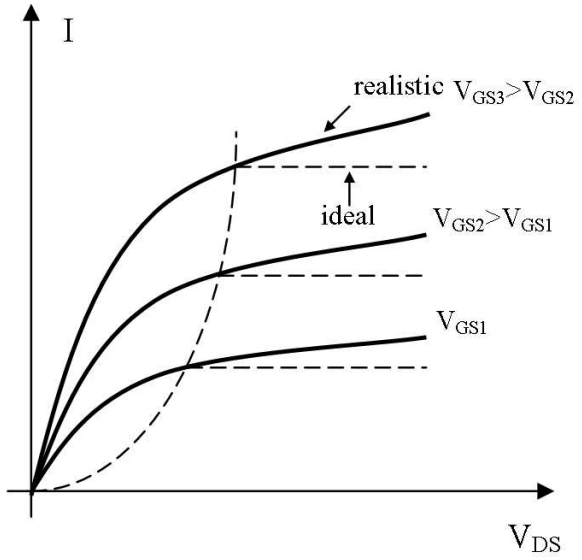


Figure 6.10: Adapted out-put-characteristics of NMOS transistors [2]

Additionally, for our modeling aims we should distinguish between the "projected" channel length L and an effective channel length L_{eff}. As explained above, originally, L denotes the distance measured from source (channel side) to drain (channel side), which is assumed to be completely influenced by the oxide field (see Figure 6.11). On the other hand, L_{eff} is a corrected channel length taking into account the lateral diffusion of source and drain contacts ($2 \cdot \Delta L_{lat}$) into the projected channel region leading to the reduction from L to $L_{eff} = L - 2 \cdot \Delta L_{lat} > 0$ L_{eff} .

Figure 6.11 depicts the effective channel lengths caused by lateral diffusion into the channel region.

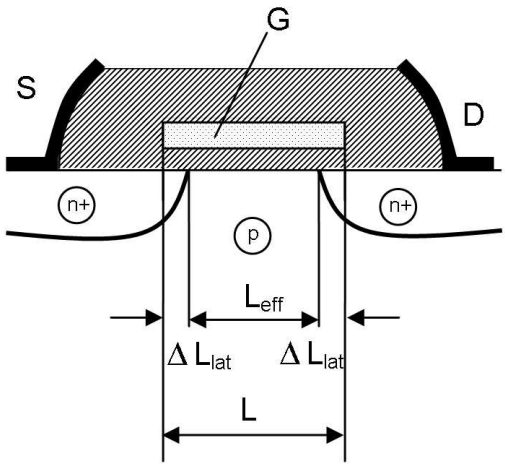


Figure 6.11: Channel shortening due to lateral diffusion [2]

6.4.2 Thermal effects

Up to now, our modeling process has been focused on I/V -characteristics at the (absolute) nominal temperature $T_{nom} = 300\ K$. To consider thermal effects, however, we need correspondingly refined models. Indeed, a set of modeling parameters necessary for describing the MOS transistor's I/V -characteristics is sensitively dependent on the operational temperature T (ϑ). Therefore, thermal effects occur if the operational temperature has been changed. For instance, such temperature changes originate from power dissipation generated in the transistors themselves, and also from external heating sources. In particular, in IC implementation of complex MOS circuits with a huge number of transistors, the transistor's

operational temperature can appreciably differ from T_{nom}. To obtain realistic and reliable simulations, we should be able to acquire the essential thermal effects.

Generally, may be stated, that currents of MOS transistors are much less dependent on temperature than currents of bipolar transistors.

Neglecting very low temperature effects due to the impurity freeze-out at very low temperatures, we will concentrate on the temperature dependency of the channel electrons' mobility as well as the threshold voltage. Moreover, there are some other thermal effects, which also have to be considered if an advanced model should be developed. The temperature dependency of leakage currents due to the, (bipolar) reverse currents of the reverse biased pn -junctions of the source and drain region and the channel current for weak inversion are examples of them.
The empiric decline of the channel electron's mobility in terms of temperature is very simple and given by

$$\mu_n(T) = \frac{\mu_{n_0}}{\sqrt{\left(\frac{T}{T_{nom}}\right)^3}}. \tag{6.24}$$

The parameter μ_{n_0} is the channel mobility for electrons at T_{nom}.

In fact, μ_{n_0} is not dependent on the temperature, however, μ_{n_0} is influenced by the electric drift field E in the channel. If necessary, we may include so-called high field effect in the approach. Then, we get the dedicated field model

$$\mu_n(T,\ E) = \frac{\mu_{n_0}(E)}{\sqrt{\left(\frac{T}{T_{nom}}\right)^3}}. \tag{6.25}$$

The quantity $\mu_{n_0}(E)$ still needs specifying in terms of E . The Figure 6.12 pictures the electron's mobility decline due to high field effects.

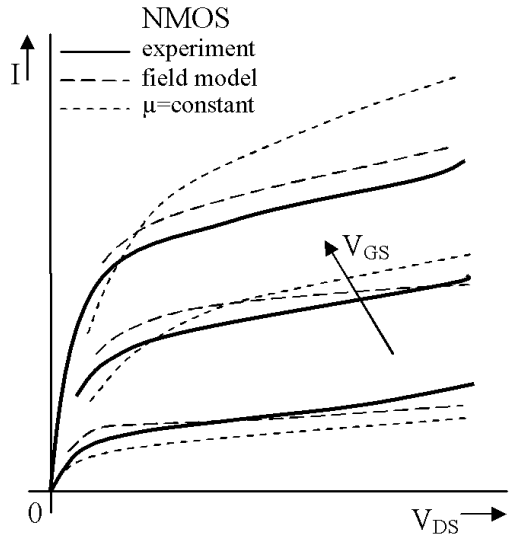


Figure 6.12: The decrease of the electron mobility [2]

The threshold voltage V_t is the most important parameter of the MOS transistor's I/V -characteristics. Generally, it controls the operation mode, and thus the working mode of MOS circuits. Therefore, reliable circuit simulations need V_t models suitable for reacting appropriate on temperature changes wherever they come from. To make V_t cope with fluctuating temperatures as demanded for simulations of real life applications, we start from Equation(6.21). If the temperature sensitive parameters and quantities are

highlighted we obtain

$$V_t(T) = V_{FB}(T) + 2 \cdot V_{F_p}(T) + \frac{d_{ox}}{\varepsilon_{ox}} \cdot \sqrt{2 \cdot q \cdot N_A \cdot \varepsilon_H \cdot \left(V_{SB} + 2 \cdot V_{F_p}(T)\right)} > 0. \tag{6.26}$$

The flat-band voltage $V_{FB}(T)$ (see Equation(6.18)) depends on temperature T due to $V_K(T)$ (see Equation (6.2) and Equation (6.5)) and is given by

$$V_{FB}(T) = -\left(\frac{d_{ox}}{\varepsilon_{ox}} \cdot Q''_Z + \begin{cases} V_{EA} - V_{WF\_Al} + V_{F\_p}(T) + V_{g/2}(T) & Al\ gate \\ V_{g/2}(T) + V_{F\_p}(T) & Si\ gate \end{cases} \right). \tag{6.27}$$

Thus, we can see that the $V_t's$ dependency on T is mainly due to the temperature sensitivity of semiconductor materials.

Self-Assessment Questions

Question 6.4.1:

What charges determine the threshold voltage V_t?

Question 6.4.2:

Which voltages control the output characteristic of an ideal NMOS transistor?

6.5 Large-signal network model

6.5.0 Summary

The knowledge indented to be imparted in this paragraph can be summarized as follows:

1. Based on the ideal I/V first an appropriate DC large signal model is developed.
2. Dynamic charge changing processes of the gate charge Q_G as well as the bulk charge Q_s may be seen as the physical reasons for capacitive effects within the NMOS transistor. Generally, to model these effects completely, in total six dynamic capacitances (C_{gd}, C_{gs}, C_{gb} relating to Q_G and C_{bd}, C_{bg}, C_{bs} regarding Q_s) have to be derived from Q_G and Q_s, respectively.
3. The dynamics of the internal transistor is assumed to be extremely high so that the control charge $Q_G(V_{GS}(t), V_{DS}(t), V_t(V_{SB}(t)))$ can follow the time-variable controlling voltages $V_{GS}(t)$, $V_{DS}(t)$ and $V_t(V_{SB}(t))$ without any delay. That is why the dynamic capacitances may be defined based on the capacitors of a steady state electric field. However, the control charges are nonlinearly dependent on the controlling voltages. That has to be considered for the definition of dynamic capacitance by applying differential quotients instead of ordinary quotients.
4. Based on our relatively simple threshold voltage model $V_t = V_t(V_{SB}(t))$ with a weak dependency on $V_{SB}(t)$, only the dynamic gate-drain capacitance C_{gd} and the dynamic gate-source capacitances C_{gs} remain at last. The rest of capacitances C_{gb}, C_{bd}, C_{bg} and C_{bs} can be neglected. From this point of view, the total oxide capacity $C_{ox} = L \cdot B \cdot \frac{\varepsilon_{ox}}{d_{ox}}$ seems to become electronically divided into the two parts C_{gd} and C_{gs}. For extremely strong inversion yields $C_{gd} = C_{gs} = \frac{1}{2} \cdot C_{ox}$, whereas in the pinch-off region we get $C_{gd} = 0$ and $C_{gs} = \frac{2}{3} \cdot C_{ox}$.
5. It should also be indicated that C_{gd} and C_{gs} are different in the active and pinch-off operation mode. In fact, in the pinch-off mode the oxide capacity C_{ox} has to be replaced by $C_{ox} \cdot \left(1 - \frac{\Delta L_e(V_{DS})}{L}\right)$, because the projected channel length L has been electronically shortened to $L - \Delta L_e(V_{DS})$.

Generally, it should be highlighted once again that all dynamic capacitances developed in this paragraph are necessary to model the dynamics of the pure NMOS transistor. Stray and overlay capacitors due to the technological implementation have to be analyzed separately. The necessary supplements will be discussed and considered with respect to the large signal network model to be developed below.

6.5.1 Model architecture

Basically, large-signal models are necessary for computer simulations of complex MOS integrated circuits (IC) and also for MOS transistor-based circuitry. Therefore, such models should comprise all essential functional aspects as well as parasitic effects if necessary. To avoid a priori limitation to dedicated application purposes, generally, we are keen developing a large-signal model to be used without any serious restrictions. Without loosing the universality, in particular, we are interested in large-signal models appropriate to be applied for direct current (DC) analysis, and also to be used for radio frequency (RF) applications. To make this complex task more transparent, we separate the development into determination of a large-signal non-linear DC model and the calculation of supplements. Mostly, such complements are capacitances necessary for simulation of the dynamic circuits.

Principally, the architecture of a non-linear network model can be derived from the physical implementation of such device. In Figure 6.13 is schematically shown, how to distil the model's architecture out of the structure of an integrated NMOS transistor.

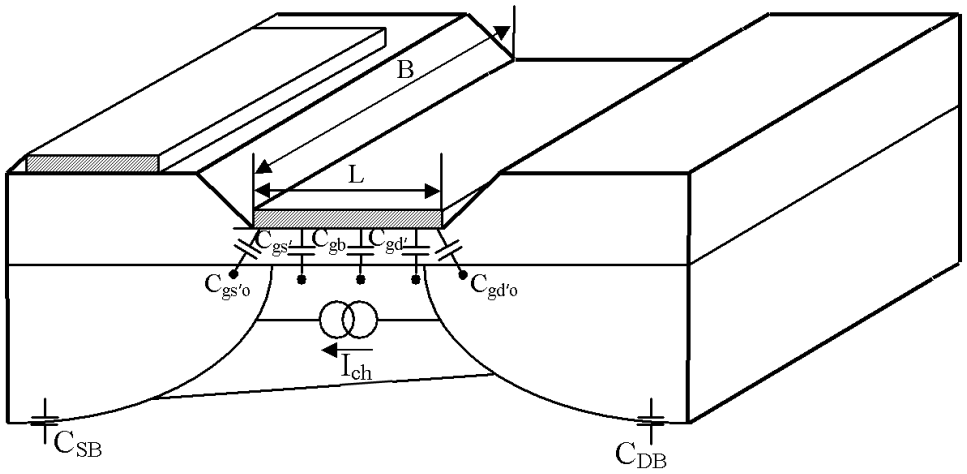


Figure 6.13: Physical implementation of NMOS transistor [2]

The outcome of the graphical extraction of model's architecture is pictured in Figure 6.14. To incorporate source and drain resistors which could become important for high currents applications we define an internal transistor model with internal nodes denoted with apostrophes. Then, the source and drain resistors connect the internal transistor model to the external source and drain terminal.

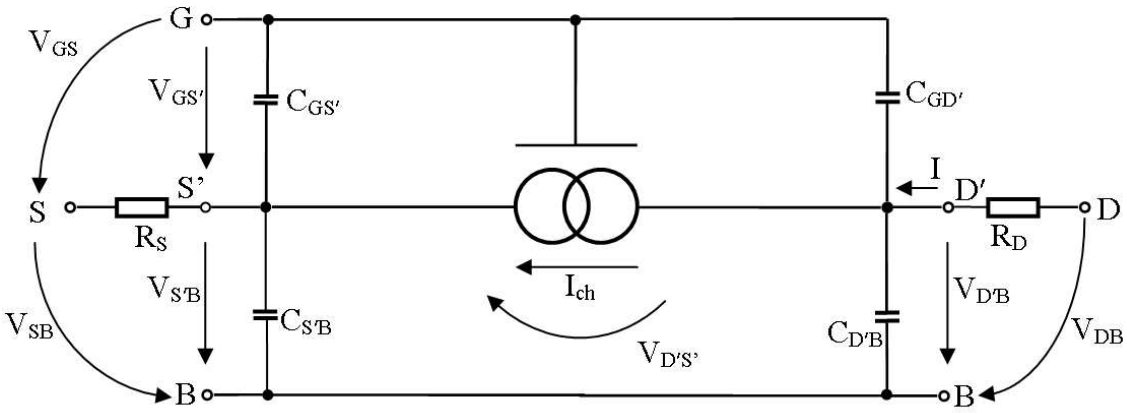


Figure 6.14: Architecture of the NMOS transistor's large-signal model

6.5.2 Large-signal DC model

The most important DC parameter is the current source I_{ch} representing the channel current controlled by the internal source-bulk, gate-source and drain-bulk voltages. To obtain acceptable results in modeling and simulation of complex MOS circuits, the ideal I/V -transistor model (see Equation (6.23)) needs refining. The demanded refinements should be concentrated on two items.

The first point concerns the definition of internal nodes. Consequently, the "pure" channel current has to be controlled by voltages ($V_{GS'}$, $V_{D'S'}$) regarding to the internal nodes. As mentioned above, the inversion channel has to be contacted with highly-doped source and drain contact regions. From the electronics point of view, these regions can be modelled by a reverse-biased pn -junction and Ohmic resistors R_S and R_D for the neutral p - and n -regions

As shown in Figure 6.14, the source and drain contact resistors R_S and R_D connect the (external) terminals source S and drain D to the internal nodes S' and drain D' , respectively. A more sophisticated modelling approach takes into account the contact resistance between the external line (e.g. made from Al) and the integrated resistors of source and drain contact regions. Generally, R_S and R_D are assumed to be independent on the temperature and constant. In fact, their resistances will become important, if RF-networks have to be simulated.

The second point of refinement is focussed on the "pinch-off" operation mode. The philosophy behind rests on physically founded approaches completed with empirical complements to achieve the model's performance and quality as demanded. Next, we demonstrate, how to adapt the ideal I/V -characteristic given by Equation (6.23) to the characteristics of real devices. Basically, there is a lot of different approaches to achieve appropriate device models.

A very simply possibility to adapt the ideal model's out-put characteristic to corresponding measurements is based on the replacement of the pre-factor β of Equation

$$\beta = \mu_n \cdot \frac{\varepsilon_{ox}}{d_{ox}} \cdot \frac{B}{L} \tag{6.28}$$

by the coefficient β^*

$$\beta^* = \beta \cdot (1 + \lambda \cdot V_{D'S'}) \,. \tag{6.29}$$

The constant λ (e.g. $\lambda \approx 0.01/V$) is used to provide the out-put characteristic with a definite rise in the pinch-off operation mode. To avoid any discontinuity in the I/V -characteristics, consequently, we have also to correct the transfer characteristic slightly. Thus, we receive

$$I_{ch} = \mu_n \cdot \frac{\varepsilon_{ox}}{d_{ox}} \cdot \frac{B}{L} \cdot (1 + \lambda \cdot V_{D'S'}) \cdot \begin{cases} 0 & V_{GS'} - V_{t'} \leq 0 \\ (V_{GS'} - V_{t'}) \cdot V_{D'S'} - \frac{V_{D'S}^2}{2} & 0 < V_{D'S'} \leq V_{GS} - V_{t'} \\ \frac{(V_{GS'} - V_{t'})^2}{2} & 0 < V_{GS'} - V_{t'} < V_{D'S'} \end{cases} . \tag{6.30}$$

with the threshold voltage

$$V_{t'} = -\left(V_K + \frac{d_{ox}}{\varepsilon_{ox}} \cdot Q''_Z\right) + 2 \cdot V_{F\_p} + \frac{d_{ox}}{\varepsilon_{ox}} \cdot \sqrt{2 \cdot q \cdot N_A \cdot \varepsilon_H \cdot (V_{S'B} + 2 \cdot V_{F\_p})} > 0 \tag{6.31}$$

6.5.3 Large signal model's capacitances

To be applied to the simulation of transient processes in MOS circuits (dynamic behavior) the DC model has to be provided with appropriate capacitors (see Figure 6.14).

The approach used here, results from the founded assumption that the dynamic behavior of the "pure" field-effect is much faster than transient processes of real MOS transistors. Basically, the same relation

is correct referring to the MOS circuitry to be driven by such transistors. Actually, the stray and parasitic capacitors due to transistors' physical implementations delay the transient behavior of real devices. Consequently, the complex model of a distributed non-linear RC -line used for modeling the dynamic electronics of the pure field effect can be replaced by the channel current source completed with the dynamic capacitors related to the dynamic charge and discharge processes. From the physical point of view, these recharging processes start out from the source and drain contacts where either the temporally needed channel electrons can immediately be delivered or an actual excess of electrons can be adopted without any remarkable delay. The adequate positive charge is influenced on the gate by the electric field between channel and gate electrode. The bulk charge can be updated by modulation of the depth d_s of the space charge layer below the channel. All these recharging reactions, the exchange of electrons between channel and its contacts, the influence of positive charges on the gate electrode as well as the modulation of bulk charge are very fast processes (taking a few ps only), which will not essentially adulterate the internal transistor's dynamics. Hence, the modulation of control charges is assumed to be performed under steady state conditions. In other words, any time-variable change in $V_{GS}(t)$, $V_{DS}(t)$ or $V_{SB}(t)$ immediately results in an adequate reaction in Q_G, Q_n and Q_s. Therewith, we may model the transistor dynamics based on $Q_\nu(t) = Q_\nu(V_{\nu 1}(t), ..., V_{\nu\mu}(t), ...,)$. These charges can be found from Q''_G, Q''_n by integration. Therewith, we get $Q_s = -(Q_G + Q_Z + Q_n) \leq 0$ for the bulk charge.

Assumed, at least one of the controlling voltages $V_{DS}(t)$, $V_{SB}(t)$, $V_{GS}(t)$ is changing fast enough, then the corresponding control charges will react with adequate recharging processes including fast changes of the electric field between spatially separated charges on the gate and channel as well as space charges of bulk, respectively. As well-known from linear capacitors, time-variable field changes lead to displacement currents. However, there is a remarkable difference to linear capacitors, because the control charges in NMOS transistors are nonlinearly dependent on the controlling voltages. That makes the control charge-based capacitances become dynamic elements. From the basics of electrical engineering we know the definition of dynamic capacitors:

$$C_{\nu\mu} = \frac{\partial}{\partial V_{\nu\mu}} Q_\nu(V_{\nu 1}, ..., V_{\nu\mu}(t), ...,) \geq 0. \tag{6.32}$$

Having Q_G in terms of their controlling voltages available, we define

$$C_{gd} = \left(\frac{\partial}{\partial V_{GD}} Q_G\right)_{V_{GS},\, V_{SB}=const.} = -\frac{\partial}{\partial V_{DS}} Q_G \geq 0. \tag{6.33}$$

Thus, we get

$$C_{gd} = -\frac{\partial}{\partial V_{DS}} Q_G = \frac{2}{3} \cdot C_{ox} \cdot \begin{cases} 1 - \frac{(V_{GS}-V_t)^2}{(2\cdot(V_{GS}-V_t)-V_{DS})^2} & V_{GS} - V_t \geq V_{DS} > 0 \\ \\ 0 & 0 < V_{GS} - V_t < V_{DS} \end{cases} \tag{6.34}$$

where C_{ox} stands for the oxide capacity

$$C_{ox} = B \cdot L \cdot \frac{\varepsilon_{ox}}{d_{ox}}. \tag{6.35}$$

We should note that gate-drain capacity C_{gd} varies between $C_{gd} = 0$for pinch-off, and $C_{gd} = 1/2 \cdot C_{ox}$ for a very strong inversion channel. The C_{gd} 's dependencies on the working point fixed by the controlling voltages $V_{GS} - V_t$ and V_{DS} are given in Figure 6.15 and Figure 6.16, respectively.

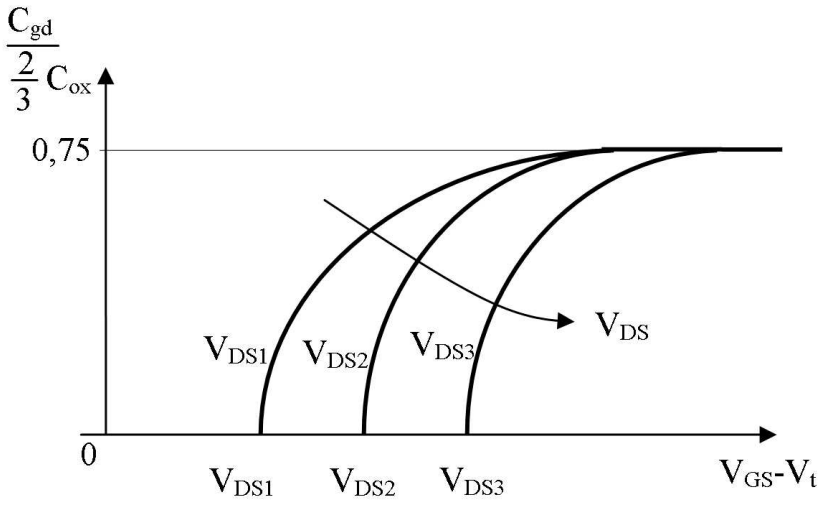


Figure 6.15: Dependency of the gate-drain capacity C_{gd} on the gate-source voltage $V_{GS} - V_t$ [1]

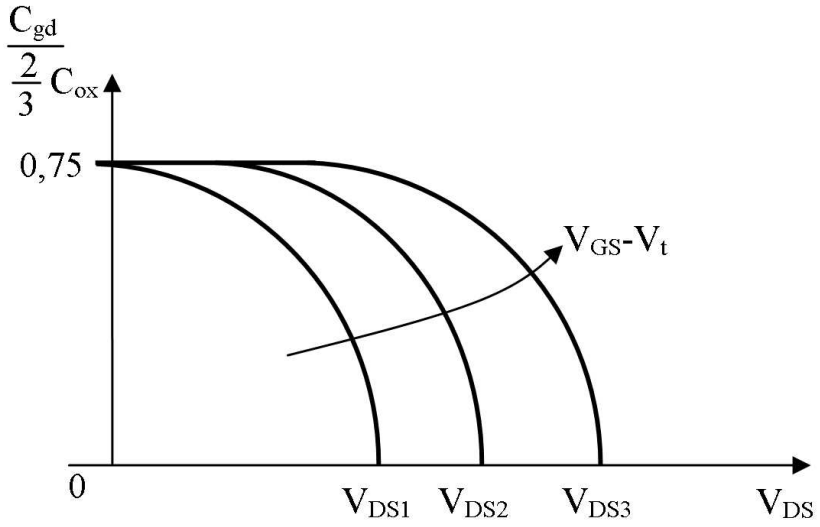


Figure 6.16: Dependency of the gate-drain capacity C_{gd} on the drain-source voltage V_{DS}[1]

Exploiting a similarly modeling approach, we obtain

$$C_{gb} \approx C_{bg} = 0 \tag{6.36}$$

and

$$C_{gs} = \frac{2}{3} \cdot C_{ox} \cdot \begin{cases} 1 - \frac{(V_{GS}-V_t-V_{DS})^2}{(2\cdot(V_{GS}-V_t)-V_{DS})^2} & V_{GS} - V_t \geq V_{DS} > 0 \\ \\ 1 & 0 < V_{GS} - V_t < V_{DS} \end{cases} . \tag{6.37}$$

For strong inversion $V_{GS} - V_t >> V_{DS} > 0$ Equation (6.37) delivers $C_{gs_{\min}} = \frac{1}{2} \cdot C_{ox}$ as the minimum of C_{gs}, whereas the maximum $C_{gs_{\max}} = \frac{2}{3} \cdot C_{ox}$ is reached for pinch-off.

Summarizing, we should state that only $C_{gd'}$ and $C_{gs'}$ will play an essential role, whereas the rest of dynamic capacitances can be neglected. In particular, this is true for the dynamic gate-bulk capacitances. These dynamic capacitances are indirectly involved in the recharging processes via the threshold voltage V_t.

As mentioned above, to model the dynamic behavior of real transistors, additionally, some stray capacitors have to be taken into account (see Figure 6.14). The physical origin of these capacitors can also be withdrawn from the device's physical implementation sketched in Figure 6.13. Generally, these capacitances could be determined by using the Green's function (pure geometry functions) of the space with the stray fields, however, most of them are not available. Mostly, an approximation based on the plate capacitor model with homogeneous field is sufficient for that purpose. Thus, the calculation problem is reduced to measuring and estimating the geometric parameters of these capacitors. Assumed these parasitic capacitance can be estimated acceptably, we will find $C_{gd'0}$ representing the stray fields at drain, and correspondingly $C_{gs'0}$ for the stray fields at source. It should be noted that these stray capacitances are considered to be neither dependent on the working point nor on the temperature.

With the control capacitances $C_{gd'}$ and $C_{gs'}$ from Equation (6.34) and Equation (6.37), finally, we obtain the total gate drain and gate source capacitances ($C_{GD'}$ and$C_{GS'}$) as needed for the large-signal network model. It is

$$C_{GD'} = C_{gd'0} + C_{gd'} \tag{6.38}$$

and in a similar way

$$C_{GS'} = C_{gs'0} + C_{gs'}. \tag{6.39}$$

As shown in Figure 6.1 the source and drain contact regions are n^+ -doped. With respect to the p -type bulk, each of them forms a pn -junction to insulate the contacts from bulk.

Thus, $V_{SB} \geq 0$ and also the drain-bulk voltages $V_{DB} \geq 0$ are reverse voltages of the source and drain contact regions, and therefore, the width of the space charge layers around the contacts will be extended in terms of $V_{SB} \geq 0$ and $V_{DB} \geq 0$. Therefore, the source and drain regions contribute to the network model, the (parasitic) space charge layer capacitors $C_{S'B}$ and $C_{D'B}$. Based on the electronics behind, we have to distinguish their space charge capacitances from the dynamic charge-based channel capacitances developed above.

To calculate the space charge capacitances $C_{S'B}$ and $C_{D'B}$, firstly, we determine the width of the space charge layer on the p -side of the pn -junction at source. After that, we are able to find the stored charge (per unit area) in the p -side of the space charge layer at source and drain Q''_{s_s} and Q''_{s_d} in terms of $V_{SB} \geq 0$ and $V_{DB} \geq 0$. Finally, we obtain the space charge capacitance (per unit area) C''_{SB} for source from

$$C''_{SB} = \frac{d}{dV_{SB}}\left(-Q''_{s_s}\right) = \sqrt{\frac{\varepsilon_H \cdot q \cdot N_A}{2 \cdot V_{bi}}} \cdot \left(1 + \frac{V_{SB}}{V_{bi}}\right)^{-\frac{1}{2}} > 0 \qquad \frac{V_{SB}}{V_{bi}} >> 1 \tag{6.40}$$

and correspondingly C''_{DB} for drain from

$$C''_{DB} = \frac{d}{dV_{DB}}\left(-Q''_{d_s}\right) = \sqrt{\frac{\varepsilon_H \cdot q \cdot N_A}{2 \cdot V_{bi}}} \cdot \left(1 + \frac{V_{DB}}{V_{bi}}\right)^{-\frac{1}{2}} > 0 \qquad \frac{V_{DB}}{V_{bi}} >> 1. \tag{6.41}$$

Actually, the expressions for the space charge capacitances C''_{SB} and C''_{DB} loose their validity for small reverse voltages ($\frac{V_{SB}}{V_{bi}} \approx 1$, $\frac{V_{DB}}{V_{bi}} \approx 1$), and particularly for forward voltages ($V_{SB} < 0$, $V_{DB} < 0$). For such cases, dedicated relations are needed.

It should be mentioned that C''_{SB} and C''_{DB} do not change their voltage dependences essentially, if the doping profile is not longer one-sided abrupt. For a linear graduated pn -junction for instance, we achieve instead of the exponent $-1/2$ in Equation (6.40) and Equation (6.41) the exponent $-1/3$. That is why an approximation

$$C''_S = C''_s \cdot \left(1 + \frac{V_S}{V_{bi}}\right)^{-m} > 0 \qquad \frac{V_S}{V_{bi}} >> 1 \tag{6.42}$$

is used often. Therein, the factor C''_s and the exponent m serve as fitting parameters.

Self-Assessment Questions

Question 6.5.1:

How can be found the network model's architecture?

Question 6.5.2:

What is the most important DC parameter in a large-signal network model of an NMOS transistor, and by which voltages is it controlled?

Question 6.5.3:

Which capacitances have to be taken into account to consider dynamic transistor effects?

6.6 Problems and examples

Exercise 6-1: I/V-characteristic of an NMOS-transistor

Simulate the I/V-characteristic of an NMOS-transistor. The schematic of the simulation circuit is shown in Figure 6.17. The gate voltage should be changed from 2V to 5V in stages of 0.5V.

The executable example (netlist and simulator commands) can be found in the $MosfetExercise6-1.prb$.

The parameters of the NMOS-transistor are

name	parameter	value	units
L	length of channel	2μ	m
W	width of channel	100μ	m
VTO	threshold voltage	2	V
KP	transconductance parameter	20μ	A/V^2
TOX	thin-oxide thickness	$100n$	M
NSUB	bubstrate doping	4e15	cm^{-3}
IS	bulk junction saturation current	2e-16	A
PB	bulk junction potential	0.8	V
CGDO	GD overlap capacitance	1.7n	F/m
CGSO	GS overlap capacitance	9n	F/m
RD	drain ohmic resistance	1m	Ω
RS	source ohmic resistance	1m	Ω

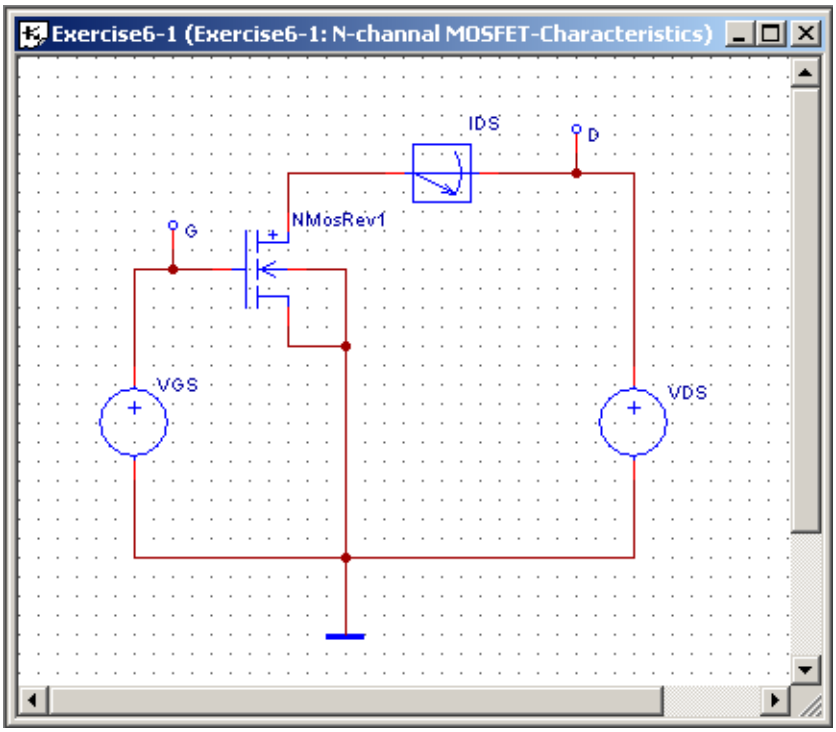
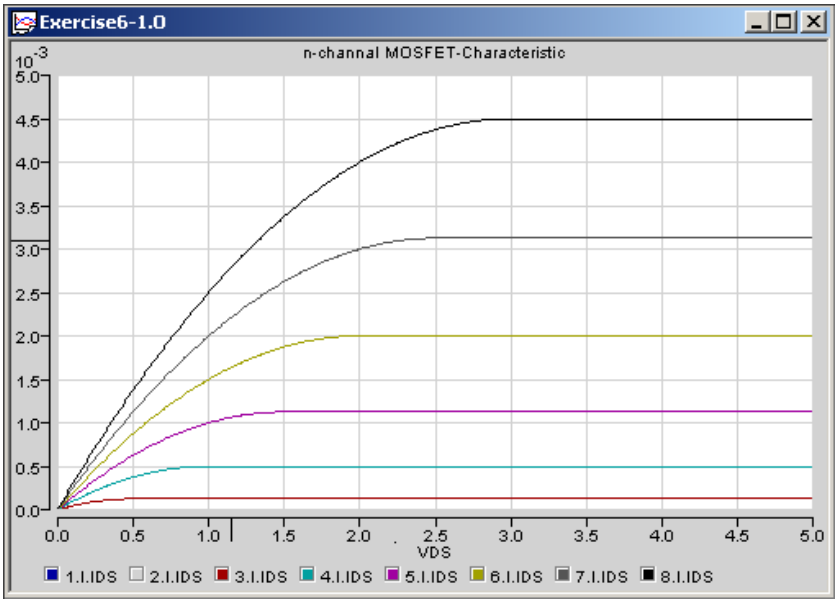


Figure 6.17: Schematic for Exercise 6-2



$V_{GS} = 5V$

$V_{GS} = 4.5V$

$V_{GS} = 4V$

$V_{GS} = 3.5V$

$V_{GS} = 3V$
$V_{GS} = 2.5V$

Figure 6.18: I/V-characteristic of an NMOS transistor

Exercise 6 2: Measurement of the capacitances of an NMOS transistor

The simulation example (netlist and simulator commands) can be found in the file $MosfetExercise6-2.prb$.

This exercise demonstrates the measurement of the gate capacitances in dependency of the drain-source voltage. After the DC analysis you must perform an AC analysis with only one frequency (e.g. fmeas=1MHz) and change the the drain-source voltage (see Figure 6.19).

The drain and source resistances are set to zero. Then the capacitances can be computed from the imaginary part of the appropriate currents:

$$C_{GD} = \frac{\mathrm{Im}(ID)}{2 \cdot \pi \cdot f_{meas}} \qquad C_{GS} = \frac{\mathrm{Im}(IS)}{2 \cdot \pi \cdot f_{meas}}$$

The results show that the gate-drain capacitance for $VDS = 0V$ has a value of $C_{GD} = C_{GDoverlap} + C_{ox}/2$ and with increasing drain voltage goes down to $C_{GD} = C_{GDoverlap}$ (see Equation (6.38) and (6.34)). The gate-source capacitance rises from $C_{GS} = C_{GSoverlap} + C_{ox}/2$ to $C_{GS} = C_{GSoverlap} + 2 \cdot C_{ox}/3$ (see Equation (6.39) and (6.37)).

Simulate the same example with $RD = RS = 1\Omega$. Compare the results and describe the influence of source and drain resistance on the results!

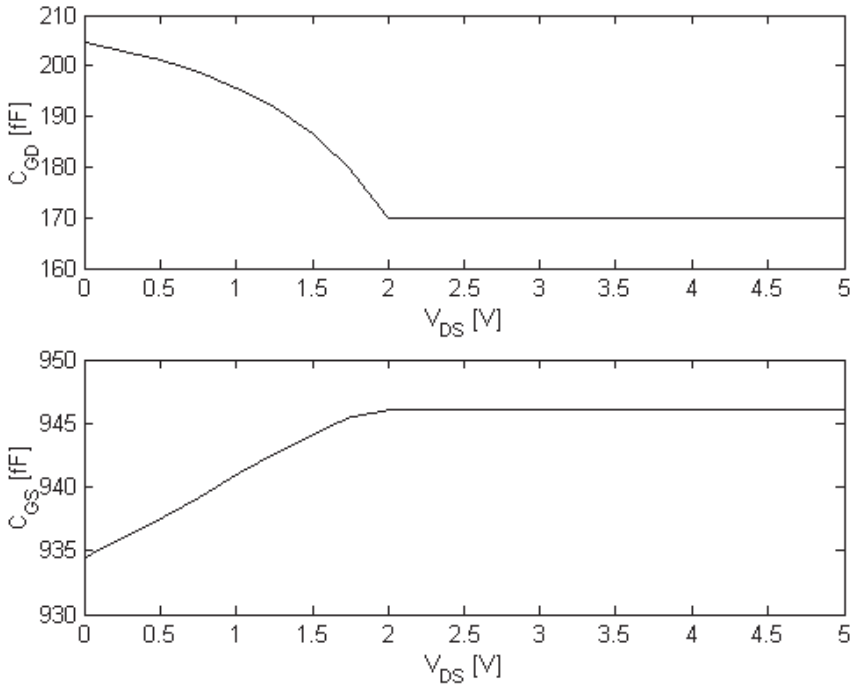


Figure 6.19: Gate-source and gate-drain capacitance of an NMOS transistor

Exercise 6 3: Transfer characteristic of a CMOS inverter circuit

The executable examples of a CMOS inverter can be found in the file $MosfetExercise6-3.prb$.

Simulate the CMOS inverter circuit in Figure 6.20. The models of both transistors are the very simple MOS-model mosnsimple and do not contain capacitances. Therefore the simulation result with $Ce = 0$ gives the static transfer characteristic of the inverter.

Describe the cause for the changed characteristic with $Ce = 0.2pF$ and determine the discharge current from it approximately and compare it with the drain currents of the transistors.

(The rise of the output signal in the range of 10 - 20 ns is caused by the ramping of the battery voltage EDD)

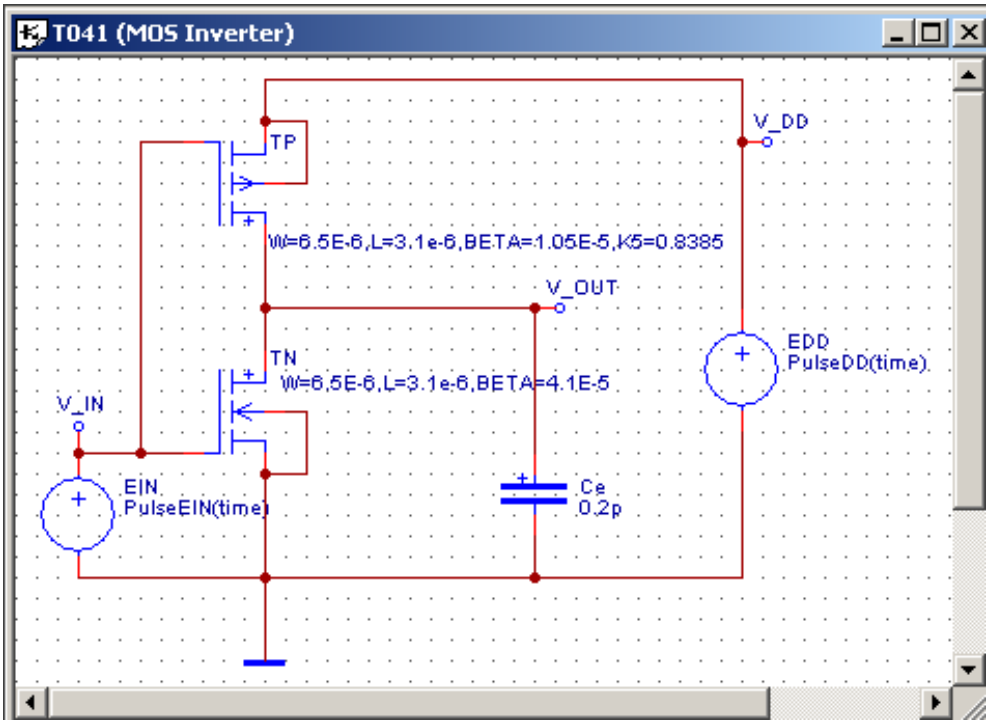


Figure 6.20: Schematic for exercise 6-3

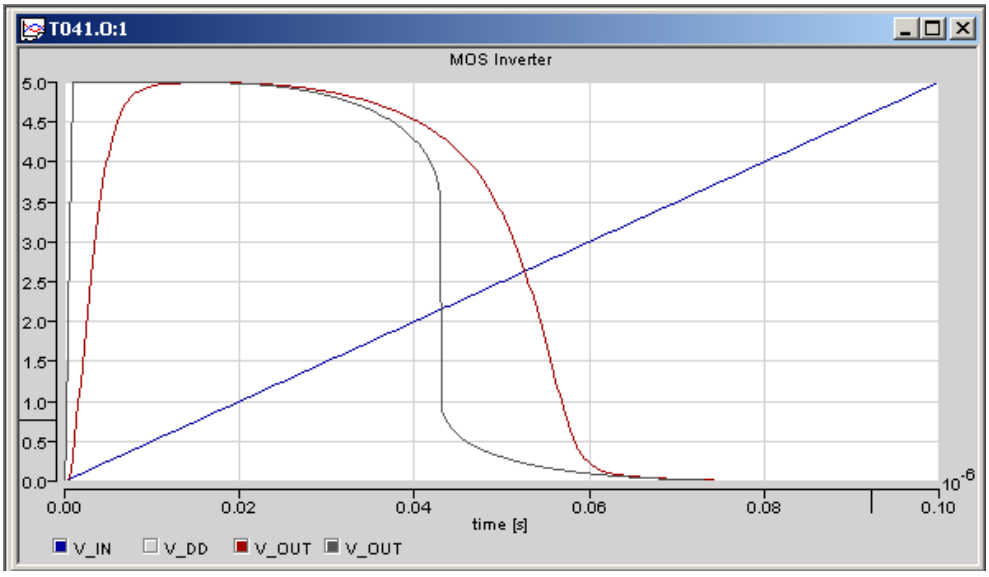


Figure 6.21: Transfer characteristics of a CMOS inverter

Module 7

References

[1] Möschwitzer, A.; Lunze, K.: Halbleiterelektronik. VEB Verlag Technik, Berlin, 1981

[2] Möschwitzer, A.: Semiconductor devices, circuits, and systems. Clarendon Press, Oxford, 1991

[3] Horowitz, P.; Hill, W.: The art of electronics. Cambridge University Press, 1989

[4] Graff, H.C.de; Klaassen, F.M.: Compact transistor modelling for circuit design. Springer-Verlag, 1990

[5] Fuller, L.: A Review of IC Fabrication Technology. Microelectronics Engineering, Rochester Institute of Technology, Course, Nov. 2003

[6] Liu, W.: MOSFET Models for SPICE Simulation including BSIM3v3 and BSIM 4, John Wily & Sons; Inc., New York Chichester Weinheim Brisbane Singapore Toronto, 2001

[7] Vladimirescu, A.: The Spice Book. John Wiley & Sons, Inc., New York, 1994

[8] Getreu, I.: Modeling the Bipolar Transistor. Tektronix Inc., Beaverton, OR, 1976

[9] Cheng, Y. Hu, Ch.: MOSFET MODELING & BSIM3 USER'S GIUDE, Kluwer Academic Publishers, Boston Dordrecht London, 1999

Module 8

Answers to Self-Assessment Questions

Module overview. This module contains the answers to the self-assessment questions which are asked in the course modules.

The questions are repeated here to use this module as an self-assessment module of the whole course.

to 3: Electronics of semiconductor structures

to 3.1: Charge carriers in semiconductors

Question 3.1.1:

What type of impurities has to be used to make an n-type semiconductor, donors or acceptors?

Answer 3.1.1:

Donors, like phosphorus.

Question 3.1.2:

What energy levels are necessarily needed to pivture an energy-band diagram of a p-type semiconductor under thermal equilibrium conditions?

Answer 3.1.2:

We need the valence and conduction band edges W_V and W_C, the Fermi-level W_F, the work function level W_{WF} as well as the electron affinity W_{EA}.

Question 3.1.3:

What is the meaning of the quantities n, n_i, φ, ψ_n and V_T of $n = n_i \cdot e^{\frac{\varphi - \psi_n}{V_T}}$?

Answer 3.1.3:

n: density of electrons in semiconductors
n_i: intrinsic density (material parameter)
φ:electrostatic potential
ψ_n:quasi-Fermi-potential for electrons
V_T: thermal temperature voltage ($V_T(T = 300K) = 25.9mV$)

Question 3.1.4:

What kind of termerature dependency is essential for the intrinsic density n_i?

Answer 3.1.4:

An exponential temperature dependency, proportional to $e^{-\frac{W_g}{k \cdot T/q}}$.

to 3.2: Electric field and space charges

Question 3.2:

What is the Poisson's differential equation needed for?

Answer 3.2:

It is used to determine the potential distribution due to space charges.

to 3.3: Charge carrier transport in semiconductors

Question 3.3.1:

Which phenomena makes the electrons and holes move within a semivonductor?

Answer 3.3.1:

These carriers move due to diffusion processes and electric field drifts.

Question 3.3.2:

Which components of carrier currents have to be considered in semiconductors?

Answer 3.3.2:

The carrier current is composed of mobile electrons and holes due to diffusion and drift fields.

to 4: *pn*-junctions

to 4.1: Basics of operation principles

Question 4.1:

What quantities determine the widths of the space charge layer near the stochiometric *pn*-junction?

Answer 4.1:

The impurity densities of donors N_D and acceptors N_A as well as the sum of built-voltage $V_{bi} > 0$ and the revers voltage $-V > 0$ on the *pn*-junction.

to 4.2: Space charge capacity

Question 4.2:

What makes the essential differences between the space charge layer capacity and the capacity of a plate capacitor (electrostatics)?

Answer 4.2:

The charges of a plate capacitor are surface charges on the plates. The total charge on an electrode is linearly dependent on the voltage of the capacitor.
Instead of two charged plates a space charge layer capacity is formed of two layers, one with positive and another with negative space charges. The total charge within the positive (or negative) layer is nonlinearly dependent on the voltage across both layers.

to 4.3: I/V-characteristic

Question 4.3.1:

What current components have to be considered to calculate the I/V-characteristic of an *pn*-junction?

Answer 4.3.1:

The minority carriers injection currents into the neutral *p*- and *n*-region (I_n and I_p) as well as the recombination/generation current in the space charge layer (I_{rg}).

Question 4.3.2:

What stands the floating potential V_{FO} for?

Answer 4.3.2:

$V_{FO} \approx 0.7V$ highlights the minimum forward voltage to be applied to "open" the *pn*-junction (diode).

to 4.4: Dynamics

Question 4.4:

What capacitance dominates the dynamics of the forward and reversed biased diode, respectively?

Answer 4.4:

The so-called diffusion capacitance C_d is essential for the dynamics of the forward biased diode, whereas the space charge capacitance C_s dominates the dynamics of the reversed biased *pn*-junction.

to 4.5: Large-signal network model

Question 4.5:

What does the network model elements r_d, C_s, C_d represent? Have they to be connected in parallel or in series whithin an network model?

Answer 4.5:

r_d is the dynamic diffusion resistance. C_d stands for the dynamic diffusion capacitance. C_s means the dynamic space charge layer capacitance. These elements have to be connected in parallel.

to 4.6: Thermal behavior

Question 4.6:

What kind of temperature dependency has to be taken into account for the forward and the reverse biased diode?

Answer 4.6:

In any case an exponential temperature dependency has to be considered.

to 5: Bipolar Transistor

to 5.1: Function principles of bipolar transistors

Question 5.1.1:

Which mechanisms play a fundamental role in the bipolar transistor?

Answer 5.1.1:

- Injection of minorities into the base
- Collection of the minorities through the backward biased base-collector junction

Question 5.1.2:

Why is this transistor called bipolar transistor?

Answer 5.1.2:

Because both types of carriers, minorities as well as majorities, play a role for the function of the transistor:
- the minorities in the base build the collector current
- the base current is build by majority carriers

Question 5.1.3:

Which phenomena cause the base current?

Answer 5.1.3:

- Injection from the base into the emitter region
- Recombination in the base region

Question 5.1.4:

What does the Ebers-Moll model describe?

Answer 5.1.4:

The dependence of the injected currents of the junction voltages.

Question 5.1.5:

Which effects cause the diffusion capacitances?

Answer 5.1.5:

The transit time of the carriers through the transistor, mainly of the minority carriers through the base.

to 5.2: Advanced models of bipolar transistors

Question 5.2.1:

What are the main differences between a vertical npn transistor and the lateral pnp transistor?

Answer 5.2.1:

- lower current gain
- lower maximum transit frequency

Question 5.2.2:

What describes the Ebers-Moll transport model?

Answer 5.2.2:

The dependence of the current of the junction voltages flowing from the emitter to the collector.

Question 5.2.3:

Which problems are not modeled with the transport model?

Answer 5.2.3:

The dependency of the current gains from the collector-emitter voltage and the currents.

Question 5.2.4:

What makes the essential difference between the Gummel-Poon model and the transport model?

Answer 5.2.4:

The base charge and its dependency of the currents and voltages are taken into account.

Question 5.2.5:

Of which portions does the base charge consist?

Answer 5.2.5:

The total base charge consists of
- Q_{b0} the zero bias base charge,
- Q_{Te} and Q_{Tc} the depletion charges and
- Q_{be} and Q_{bc} the stored charges of the injected minority carriers

Question 5.2.6:

What is the Early effect and how is it modeled in the Gummel-Poon model?

Answer 5.2.6:

The Early effect causes the rising of the collector current in the saturation region. It is modeled by the depletion charges in the total base charge.

to 5.4: Temperature effects

Question 5.4:

What are the reasons for the temperature dependency of the bipolar transistor?

Answer 5.4:

The temperature dependences of
- the intrinsic carrier concentration
- the mobility
- the ionisation of the impurities in the base

is the main reason.

to 5.5: The Spice Gummel-Poon model

Question 5.5:

Why is the base resistance not constant?

Answer 5.5:

The base resistance consists of a part for the active base region under the emitter. Because of the emitter-base current crowding the resistance depends on the base current.

to 6: MOS-field effect transistors

to 6.1: Energy-band diagram under thermal equlibrium

Question 6.1:

Which quantity achieved the surface potential V_s at the beginning of strong inversion?

Answer 6.1:

The surface potential becomes $V_s \approx 2 \cdot V_{F\_p}$, where $V_{F\_p} = V_T \cdot \ln \frac{N_A}{n_i}$ is given by the doping concentration in the bulk material.

to 6.2: Control charges in NMOS transistors

Question 6.2:

Which charges have to be considered to calculate the inversion channel conductivity?

Answer 6.2:

The channel conductivity is given by the channel charge (per unit area) $Q''_n = -(Q''_G + Q''_Z + Q''_S)$, i.e. we need the gate charge Q''_G, the interface charge Q''_Z and the charge of the space charge layer Q''_S (all charges per unit area).

to 6.3: Operation principes of MOS transistors

Question 6.3:

Is there an important difference referring to the operation principles of a p-channel and an n-channel enhancement MOS transistor?

Answer 6.3:

No, the operation principle of both types is the same.

to 6.4: I/V-Characteristic for stron inversion

Question 6.4.1:

What charges determine the threshold voltage V_t?

Answer 6.4.1:

V_t is mainly determined by the interface charge Q''_Z (unit per area) and charge of the space charge layer Q''_S (unit per area).

Question 6.4.2:

Which voltages control the output characteristic of an ideal NMOS transistor?

Answer 6.4.2:

The output characteristic of an ideal NMOS transistor is controlled by $0 < V_{GS} - V_t < V_{DS}$

to 6.5: Large-signal network model

Question 6.5.1:

How can be found the network model's architecture?

Answer 6.5.1:

The network model's architecture can be derived from the transistors physical implementation.

Question 6.5.2:

What is the most important DC parameter in a large-signal network model of an NMOS transistor, and by which voltages is it controlled?

Answer 6.5.2:

The most important DC parameter is the channel current I_{ch} controlled by the internal source-bulk voltage ($V_{S'B}$), the gate source and drain source voltages ($V_{GS'}$ and $V_{D'S'}$)

Question 6.5.3:

Which capacitances have to be taken into account to consider dynamic transistor effects?

Answer 6.5.3:

To consider the transistor's dynamics, we have to consider the internal space charge layer capacities of the source and drain contact regions $C_{S'B}$, $C_{D'B}$, the dynamic channel capacities $C_{gs'}$ and $C_{gd'}$ as well as stray capacities due to the device's physical implementation.