FUNDAMENTALS OF SEMICONDUCTOR PHYSICS AND DEVICES

Rolf Enderlein
Norman J. M. Horing

World Scientific

FUNDAMENTALS OF SEMICONDUCTOR PHYSICS AND DEVICES

FUNDAMENTALS OF SEMICONDUCTOR PHYSICS AND DEVICES

Rolf Enderlein
Humboldt-University Berlin
University of Sao Paulo

Norman J. M. Horing
Stevens Institute of Technology
Hoboken, NJ

World Scientific
Singapore • New Jersey • London • Hong Kong

Dedication

This book is dedicated to the memory of Adele and Werner Enderlein (parents of R.E.), and to the memory of Joseph and Esther Morgenstern (Morganstein)(grandparents of N.J. M. H.).

Preface

People come to technical books with a vast array of different needs and requirements, arising from their differing educational backgrounds, professional orientations and career objectives. This is particularly evident in the field of semiconductors, which stands at the juncture of physics, chemistry, electronic engineering, material science and mathematics. No longer just an academic discipline, this field is at the heart of an ongoing revolution in communications, computation and electronic device applications, that innovate many fields and change modern life in myriad ways, large and small. Its profound impact and further potential command interest and attention from all corners of the earth, and from a wide variety of students and researchers.

The clear need to address a broadly diversified and variously motivated readership has weighed heavily in the authors' considerations. It poses a pedagogical problem faced by many teachers of intermediate level courses on semiconductor physics. Generally speaking, every student has previously studied about half of the course material. The difficulty lies in the fact that each student's exposure is likely to have been to a *different* half, depending on which lower level courses and teachers they have had, and where the emphasis lay. To accommodate readers with varied backgrounds we start from first principles and provide fully detailed explanations and proofs, assuming only that the reader is familiar with the Schrödinger equation. This intensively tutorial treatment of the electronic properties of semiconductors includes recent fundamental developments and is carried through to the physical principles of device operation, to meet the needs of readers interested in engineering aspects of semiconductors, as well as those interested in basic physics. Clarity of explanation and breadth of exposure relating to the electronic properties of semiconductors, from first principles to modern devices, are our principal objectives in this frankly pedagogical book. We offer full mathematical derivations to strengthen understanding and discuss the physical significance of results, avoiding reliance on 'hand waving arguments alone.

To support the reader's introduction to the physics of semiconductors, we provide a thorough grounding in the basic principles of solid state physics, assuming no prior knowledge of the field on the part of the reader. An elementary discussion of the crystal structure, chemical nature and macroscopic properties characterizing semiconductors is given in Chapter 1. Moreover, we also include an extensive appendix to guide the reader through group theory and its applications in connection with the symmetry properties of semiconductors, which are of major importance. Beside spatially homogeneous bulk semiconductors, we undertake a full exposition of inhomogeneous semiconductor junctions and heterostructures because of their crucial role

The book has emerged from lectures which the authors presented for physics students at the Humboldt-University of Berlin, Germany, and the State University of Sao Paulo, Brazil, and for physics and engineering physics students at the Stevens Institute of Technology in Hoboken, New Jersey, U.S.A. Part of the book has similarities with the german book "Grundlagen der Halbleiterphysik" ("Fundamentals of Semiconductor Physics") which was written by one of us (R. E.) together with A. Schenk. We are thankful to Dr. Schenk (now at ETH Zürich) for allowing us to use part of his work in the present volume. In writing this book we have had excellent support from many of our colleagues at our own and other Universities. We are particularly thankful to Prof. Dr. J. Auth (Humboldt-University Berlin), Prof. Dr. F. Bechstedt (Friedrich-Schiller University Jena), Prof. Dr. W. A. Harrison (Stanford University), Prof. Dr. M. Scheffler (Fritz-Haber Institut, Berlin), Prof. Dr. J. R. Leite, Prof. Dr. A. Fazzio, and Prof. Dr. J. L. Alves (State University Sao Paulo), as well as to Prof. Dr. H. L. Cui, Prof. Dr. G. Rothberg, Mr. G. Lichtner (Stevens Institute of Technology), and Prof. Dr. G. Gumbs (Hunter College, CUNY, New York), who read parts of the manuscript and contributed helpful suggestions and critical remarks. The technical assistance of Mrs. Hannelore Enderlein is gratefully acknowledged.

Rolf Enderlein — Norman J.M. Horing

Sao Paulo — Hoboken, NJ

October 1996

Contents

1 Characterization of semiconductors **1**

1.1 Introduction 1

1.2 Atomic structure of ideal crystals 5

1.2.1 Crystal lattices 6

1.2.2 Point groups of equivalent directions and crystal classes 12

1.2.3 Space groups and crystal structures 14

1.2.4 Cubic semiconductor structures 16

1.2.5 Hexagonal semiconductor structures 22

1.3 Chemical nature of semiconductors. Material classes 28

1.3.1 Group IV elemental semiconductors 29

1.3.2 III-V semiconductors 30

1.3.3 II-VI semiconductors 31

1.3.4 Group VI elemental semiconductors 31

1.3.5 IV-VI semiconductors 32

1.3.6 Other compound semiconductors 32

1.4 Macroscopic properties and their microscopic implications . . 33

1.4.1 Electrical conductivity 34

1.4.2 Dependence of conductivity on the semiconductor state 35

1.4.3 Optical absorption spectrum and the band model of semiconductors 38

1.4.4 Electrical conductivity in the band model 41

1.4.5 The Hall effect and the existence of positively charged freely mobile carriers 45

1.4.6 Semiconductors far from thermodynamic equilibrium . 49

2 Electronic structure of ideal crystals **51**

2.1 Atomic cores and valence electrons 51

2.2 The dynamical problem 54

2.2.1 Schrödinger equation for the interacting core and valence electron system 54

2.2.2 Adiabatic approximation. Lattice dynamics 57

2.2.3 One-particle approximation. One-particle Schrödinger equation 66

2.3 General properties of stationary one-electron states in a crystal 82

2.3.1 Symmetry properties of the one-electron Schrödinger equation 82

2.3.2 Bloch theorem 85

2.3.3 Reciprocal vector space and the reciprocal lattice . . . 89

2.3.4 Relation between energy eigenvalues and quasi-wave-vector 94

2.4 Schrödinger equation solution in the nearly-free-electron approximation . . . 98
2.4.1 Non-degenerate perturbation theory . . . 100
2.4.2 Degenerate perturbation theory . . . 103
2.5 Band structure . . . 105
2.5.1 Brillouin zones . . . 105
2.5.2 Degeneracy of energy bands . . . 116
2.5.3 Critical points and effective masses . . . 119
2.5.4 Density of states . . . 123
2.5.5 Spin . . . 128
2.5.6 Calculational methods for band structure determination 133
2.6 Tight binding approximation . . . 140
2.6.1 Fundamentals . . . 140
2.6.2 TB theory of diamond and zincblende type semiconductors . . . 148
2.6.3 sp^3-hybrids, total energy and chemical bonding . . . 165
2.7 $\mathbf{k} \cdot \mathbf{p}$-method . . . 179
2.7.1 Fundamentals . . . 179
2.7.2 Valence bands of diamond structure semiconductors without spin-orbit interaction . . . 184
2.7.3 Luttinger-Kohn model . . . 189
2.7.4 Kane model . . . 200
2.8 Band structure of important semiconductors . . . 211
2.8.1 Silicon . . . 212
2.8.2 Germanium . . . 218
2.8.3 III-V semiconductors . . . 219
2.8.4 II-VI semiconductors . . . 221
2.8.5 IV-VI semiconductors . . . 224
2.8.6 Tellurium and selenium . . . 224

3 Electronic structure of semiconductor crystals with perturbations 225
3.1 Atomic structure of real semiconductor crystals . . . 226
3.1.1 Classification of perturbations . . . 226
3.1.2 Point perturbations . . . 227
3.1.3 Formation of point perturbations and their movement 235
3.1.4 Line and planar defects . . . 240
3.2 One-electron Schrödinger equation for point perturbations . . 241
3.2.1 Electron-core interaction . . . 242
3.2.2 Electron-electron interaction . . . 245
3.3 Effective mass equation . . . 252
3.3.1 Effective mass equation for a single band . . . 253
3.3.2 Multiband effective mass equation . . . 259

3.4 Shallow levels. Donor and acceptor states 265
3.4.1 Hydrogen model 266
3.4.2 Improvements upon the hydrogen model 271
3.5 Deep levels 281
3.5.1 General characterization of deep levels 281
3.5.2 Defect molecule model 285
3.5.3 Solution methods for the one-electron Schrödinger equation of a crystal with a point perturbation 293
3.5.4 Correlation effects 301
3.5.5 Results for selected deep centers 308
3.6 Clean semiconductor surfaces 334
3.6.1 The concept of clean surfaces 334
3.6.2 Atomic structure of clean surfaces 336
3.6.3 Electronic structure of crystals with a surface 354
3.6.4 Atomic and electronic structure of particular surfaces 371
3.7 Semiconductor microstructures 388
3.7.1 Heterojunctions 388
3.7.2 Microstructures: Fabrication, classifications, examples 396
3.7.3 Methods for electronic structure calculations 409
3.7.4 Electronic structure of particular microstructures . . . 420
3.8 Macroscopic electric fields 433
3.8.1 Effective mass equation and stationary electron states 434
3.8.2 Non-stationary states. Bloch oscillations 437
3.8.3 Interband tunneling 440
3.8.4 Photon assisted interband tunneling 442
3.9 Macroscopic magnetic fields 443
3.9.1 Effective mass equation in a magnetic field 444
3.9.2 Solution of the effective mass equation 452

4 Electron system in thermodynamic equilibrium 457
4.1 Fundamentals of the statistical description 457
4.2 Calculation of average particle numbers 460
4.2.1 Configuration-independent one-particle states 460
4.2.2 Configuration-dependent one-particle states 462
4.3 Density of states 469
4.3.1 Total electron concentration 469
4.3.2 Density of states of ideal semiconductors 470
4.3.3 Density of states of real semiconductors 474
4.4 Free carrier concentrations 477
4.4.1 Conservation of total electron number 477
4.4.2 Free carrier concentration dependence on Fermi energy. Law of mass action. 478
4.4.3 Intrinsic semiconductors 482

4.4.4 Extrinsic semiconductors 484
4.4.5 Compensation of donors and acceptors 489
4.4.6 More complex cases 492

5 Non-equilibrium processes in semiconductors 499
5.1 Fundamentals of the statistical description of non-equilibrium processes 500
5.2 Systematics of non-equilibrium processes in semiconductors . 505
5.2.1 Temporal inhomogeneity and spatial homogeneity . . 505
5.2.2 Spatial inhomogeneity and temporal homogeneity . . . 506
5.2.3 Space and time inhomogeneities 508
5.3 Generation and annihilation of free charge carriers 509
5.3.1 Generation processes 510
5.3.2 Unipolar annihilation of free charge carriers: capture at deep centers 511
5.3.3 Bipolar annihilation of carriers at deep centers 517
5.4 Drift current 523
5.5 Diffusion and annihilation of free carriers 527
5.6 Equilibrium of free carriers in inhomogeneously doped semiconductors 530

6 Semiconductor junctions in thermodynamic equilibrium 535
6.1 pn-junction 537
6.1.1 Establishment of thermodynamic equilibrium 539
6.1.2 Diffusion voltage 541
6.1.3 Spatial variation of the electric and chemical potentials: Schottky approximation 542
6.2 Heterojunctions 549
6.2.1 Equilibrium condition 550
6.2.2 Electrostatic potential. $GaAs/Ga_{1-x}Al_xAs$ heterojunction as an example 552
6.3 Metal-semiconductor junctions 557
6.3.1 Energy level diagram before establishing equilibrium . 557
6.3.2 Electrostatic potential 559
6.3.3 Schottky barrier 563
6.4 Insulator-semiconductor junctions 567
6.4.1 Thermodynamic equilibrium 567
6.4.2 Influence of interface states 570
6.4.3 Semiconductor surfaces 572

7 Semiconductor junctions under non-equilibrium conditions 573
7.1 pn-junction in an external voltage 574
7.1.1 Electrostatic potential profile 576

7.1.2 Mechanism of current transport through a pn-junction 577
7.1.3 Chemical potential profiles for electrons and holes . . 580
7.1.4 Dependence of current density on voltage 583
7.1.5 Bipolar transistor 585
7.1.6 Tunnel diode 593
7.2 pn-junction in interaction with light 595
7.2.1 Photoeffect at a pn-junction. Photodiode and photovoltaic element 595
7.2.2 Laser diode 599
7.3 Metal-semiconductor junction in an external voltage. Rectifiers 606
7.4 Insulator-semiconductor junction in an external voltage . . . 612
7.4.1 Field effect 612
7.4.2 Inversion layers 614
7.4.3 MOSFET 620

Appendices

A Group theory for applications in semiconductor physics 623
A.1 Definitions and concepts 623
A.1.1 Group definition 623
A.1.2 Concepts 624
A.2 Rigid displacements 627
A.2.1 Definition 627
A.2.2 Translations 628
A.2.3 Orthogonal transformations 629
A.2.4 Geometrical interpretation 631
A.2.5 Screw rotations and glide reflections 632
A.3 Translation, point and space groups 635
A.3.1 Lattice translation groups 635
A.3.2 Point groups 636
A.3.3 Space groups 654
A.4 Representations of groups 655
A.4.1 Introduction 655
A.4.2 Irreducible representations 661
A.4.3 Products of representations 667
A.5 Representations of the full rotation group 673
A.5.1 Vector representation of the rotation group and generators of infinitesimal rotations 674
A.5.2 Representations for dimensions other than three . . . 676
A.6 Spinor representations 682
A.6.1 Space-dependent spinors 682

A.6.2 Representation $\mathcal{D}_{\frac{1}{2}}$ 683
A.6.3 Irreducible spinor representations 684
A.6.4 Double group method 685
A.7 Projective representations 687
A.7.1 Factor systems 687
A.7.2 Definitions and theorems 689
A.7.3 Construction of projective representations 692
A.8 Time reversal symmetry 692
A.8.1 Time reversal operator 693
A.8.2 Additional degeneracies of energy eigenvalues 694
A.8.3 Additional selection rules for matrix elements 697
A.9 Irreducible representations of space groups 698
A.9.1 Representations of translation groups 698
A.9.2 Star of wavevectors 700
A.9.3 Small point groups and their projective representations 702
A.9.4 Representations of the full space group 704
A.9.5 Spinor representations of space groups 706
A.9.6 Implications of time reversal symmetry 707
A.9.7 Compatibility 712
A.10 Irreducible representations of small point groups 712
A.10.1 Character tables 712
A.10.2 Multiplication tables 731
A.10.3 Compatibility relations 734

B Corrections to the adiabatic approximation **737**

C Occupation number representation **741**

Bibliography **747**

Index **757**

Chapter 6

Semiconductor junctions in thermodynamic equilibrium

The distinguished role of semiconductors in the modern world stems from devices which can be fabricated from these materials and, like transistors and semiconductor lasers, have opened the door to the present computer and communication technologies. The operation of semiconductor devices depends largely on effects which occur at *semiconductor junctions*, in addition to the properties of homogeneous semiconductors which have been treated almost exclusively in this book thus far. A semiconductor junction is understood as a structure composed of two interconnected layers, of which at least one is a semiconductor material. Various semiconductor junctions are shown in Figure 6.1 schematically. In case (a) the two layers are made of the same semiconductor material, e.g. silicon, but one of the layers is p-doped and the other n-doped. This is called a *pn-junction*. In case (b) two layers of different semiconductor materials are connected. The doping type may, but need not be different. This structure is called a semiconductor *heterojunction*. We encountered such semiconductor heterostructures above in section 3.7. If a semiconductor layer is combined with a metal layer, then one has a *metal-semiconductor junction*, and if it is connected with an insulator layer it is an *insulator-semiconductor junction*. Formally, one can also include semiconductor surfaces here, which correspond to *vacuum-semiconductor junctions*.

Each of the semiconductor junctions identified above is associated with an important physical discovery. Several of these discoveries were even honored with Nobel prizes. Moreover, each of these junctions plays a role in electronic devices, and in most cases a semiconductor junction is crucial for operation of the device. Electric rectification at a metal-semiconductor junction, the so-called Schottky contact, was historically the first semiconductor junction to be used technologically. In this particular case, funda-

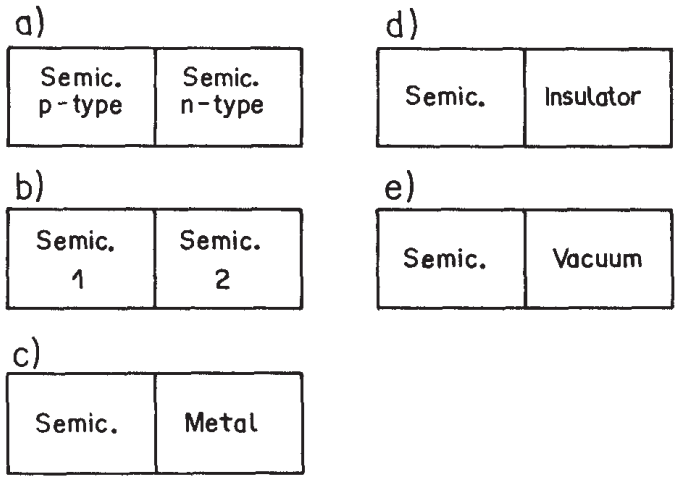

Figure 6.1: The various semiconductor junctions.

mental general laws of semiconductor junctions were first recognized, particularly by Schottky and Mott. It is thus astonishing that, in this case of the metal-semiconductor junction, a complete microscopic understanding is elusive even now. The *pn*-junction, more precisely the combination of a *pn*- and a *np*-junction, underlies the bipolar transistor, for which the inventors Shockley, Bardeen and Brattain received the Nobel prize for physics in 1956. Also, the tunnel diode, for which the Japanese physicist Esaki earned the Nobel prize in 1973, rests on the *pn*-junction. This device employs electron tunneling between the valence and conduction bands of such a junction. The unipolar field effect transistor (MOSFET) contains, beside a *pn*-junction, also an insulator-semiconductor junction as an active structural element. In experimental studies of the 2-dimensional electron gas in specially tailored silicon MOSFET's, the quantized Hall effect was discovered by von Klitzing. This accomplishment was honored with the Nobel prize for physics in 1985. The 2-dimensional electron gas and the quantized Hall effect may also be observed in semiconductor heterojunctions. Combining heterojunctions with *pn*-junctions, efficient light emitting diodes (LEDs) and semiconductor injection lasers may be fabricated. Heterojunctions also form the key structural elements of which the superlattices and quantum wells of section 3.7 are composed. The advent of these systems has opened a completely new and fruitful area in semiconductor physics and electronics, that of semiconductor microstructures.

In general terms, the operation of electronic devices based on semiconductors may be described as follows: The thermodynamic equilibrium of a semiconductor junction is disturbed by an external perturbation, for exam-

ple, by applying a voltage or by exposing the sample to light. The result of the perturbation are large changes of some properties of the junction, for example, its electric resistivity. These changes are the effects employed to advantage in semiconductor devices. To understand the perturbation of an equilibrium state of a semiconductor junction, one must first deal with the equilibrium state itself. This will be done in the present chapter. Non-equilibrium processes will be treated in Chapter 7. To initiate discussion of the equilibrium state, we start with the pn-junction.

6.1 pn-junction

pn-junctions are fabricated such that either the surface region of a homogeneously doped n (or p)-type sample, is p (or n)-doped by means of diffusion or ion implantation, or that a p (or n)-type layer is deposited on a n (or p)-type substrate by means of epitaxy. In the particular case of Si, neutron bombardment may also be used. Due to the nuclear reaction ${}^{30}_{14}Si + n \rightarrow {}^{31}_{14}Si + \gamma \rightarrow {}^{31}_{15}P + \beta$, a previously p-doped sample becomes n-doped in the surface region exposed to the neutrons.

In the following theoretical description of the pn-junction we assume homogeneously doped p- and n-regions left and right of the plane normal to the x-axis of a Cartesian coordinate system (see Figure 6.2a). The p-region extends to -∞ and the n-region to +∞. The transition from the p- to the n-region takes place abruptly at $x = 0$. We refer to this position as the 'nominal transition'. In the transverse plane normal to the x-axis, the two semi-infinite semiconductor samples are taken to extend to infinity in all directions. Thus, there are no dependencies on y and z, provided the semiconductor materials are homogeneous in the transverse plane, which we take to be the case. With this, the problem is effectively one dimensional. The carrier concentration distribution along the x-axis before equilibrium is established is shown schematically in Figure 6.2a. Conceptually, this corresponds to the instant at which the junction is formed, so that the carrier distribution has not yet had enough time to adjust to form a new equilibrium state. Considering that the p-region also has electrons which are minority carriers in it, and that the n-region also has holes as minority carriers there, one must distinguish the region in which these concentrations are meant. We indicate this by affixing a subscript to n and p. The concentrations in the n-region will be denoted by n_n and p_n, and those in the p-region by n_p and p_p. Thus, n_n and p_p are majority carrier concentrations, while n_p and p_n are minority carrier concentrations. The different values of the charge carrier concentrations in the n- and p-regions are related to different values E_{Fn} and E_{Fp} of the bulk Fermi energies in these regions, such that

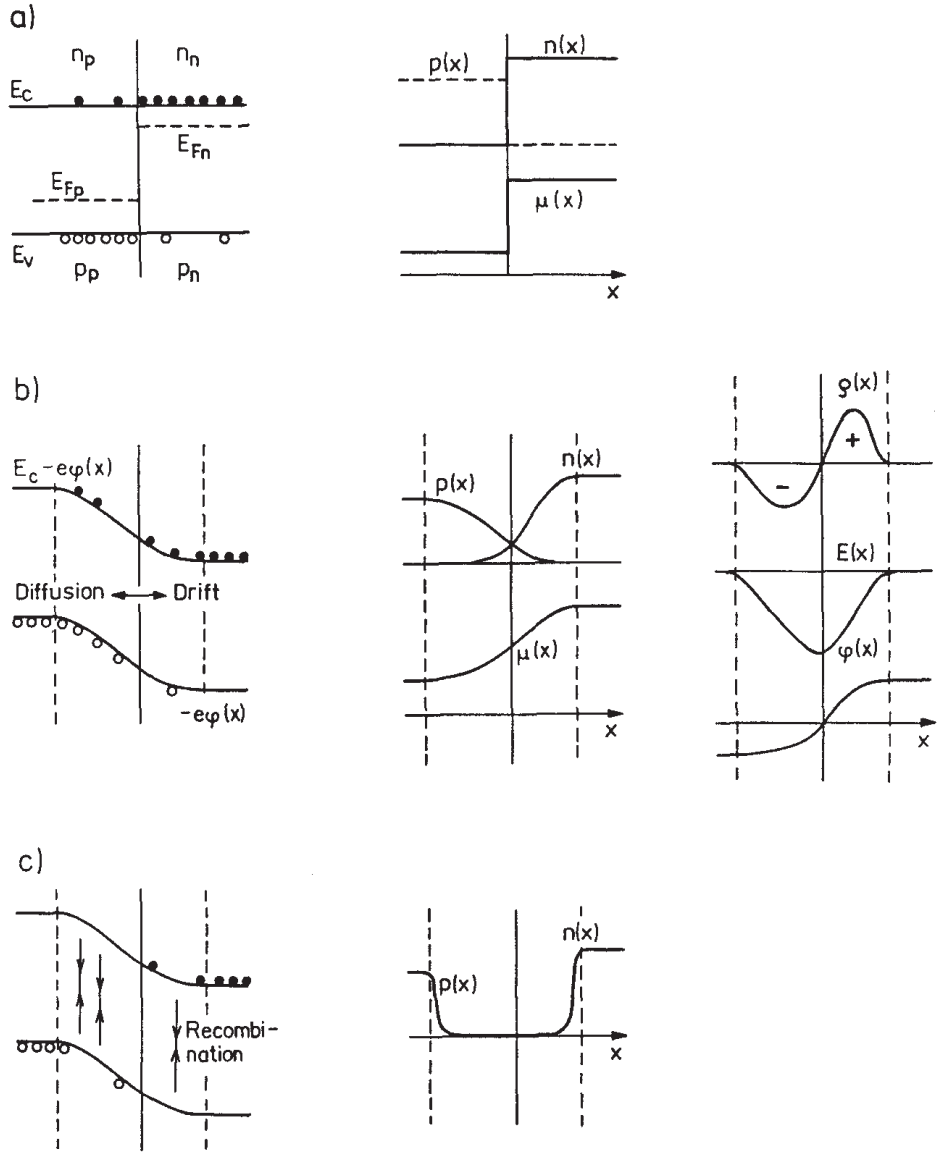

Figure 6.2: Non-equilibrium state of a pn-junction (a), from which the establishment of thermodynamic equilibrium develops (b,c).

$$n_n = N_c e^{(E_{Fn}-E_g)/kT}, \quad p_n = N_v e^{-E_{Fn}/kT}, \tag{6.1}$$

$$n_p = N_c e^{(E_{Fp}-E_g)/kT}, \quad p_p = N_v e^{-E_{Fp}/kT}. \tag{6.2}$$

6.1.1 Establishment of thermodynamic equilibrium

The situation shown in Figure 6.2a certainly does not represent an equilibrium state. The charge carrier concentrations are spatially varying functions $n(x)$ and $p(x)$, and because $n_p \ll n_n$ and $p_n \ll p_p$, strong gradients of $n(x)$ and $p(x)$ exist in the vicinity of the nominal transition $x = 0$. They cause electrons to diffuse to the left into the p-region, and holes to the right into the n-region (Figure 6.2b). Consequently, the n-region becomes positively charged, and the p-region negatively. In the terminology of Chapter 5, one can say that a space charge region is formed. It has an associated electric field $E(x)$ and an electric potential $\varphi(x)$. The dependence of these quantities on x is shown in Figure 6.2b (right column), illustrating results which will actually be obtained later by detailed analysis. For our immediate considerations, only their qualitative behavior is important. In Figure 6.2b the band edges are shown as position-dependent quantities. This representation is surprising initially, since the band edges are energy eigenvalues of the Schrödinger equation, and as such they do not depend on space coordinates. Such a representation in terms of position-dependent band edges is nevertheless meaningful in the present case, based on the fact that $\varphi(x)$ is a smoothly varying potential in the sense of section 3.3. Therefore, the effective mass equation can be applied. For the envelope function $X_c(x)$ of the conduction band electrons this equation reads

$$\left[-\frac{\hbar^2}{2m_c^*}\frac{d^2}{dx^2} + E_c - e\varphi(x)\right] X_c(x) = EX_c(x), \tag{6.3}$$

and for the envelope function $X_v(x)$ of the valence band holes it has the form

$$\left[-\frac{\hbar^2}{2m_v^*}\frac{d^2}{dx^2} + e\varphi(x)\right] X_v(x) = (-E)X_v(x). \tag{6.4}$$

According to equation (6.3), the electrons at the conduction band edge exhibit a potential energy $E_c - e\varphi(x)$, and the holes at the valence band edge exhibit a potential energy $e\varphi(x)$. These energy functions are depicted in Figure 6.2b. In this, one has to recognize that the energy $e\varphi(x)$ of the holes counts negatively, so that on the positive electron energy scale one must therefore plot $-e\varphi(x)$.

The electric field in the space charge region drives a current in the direction opposite to that of the diffusion current (Figure 6.2b). The concentrations

$$\frac{dn_s}{dU} = \frac{5}{4}\alpha U^{1/4}. \tag{7.119}$$

With $U \approx 5\ V$, this yields $(dn_s/dU) \approx 10^{13}\ cm^{-2}V^{-1}$. A voltage increase of 1 V results in about 10^{13} electrons being induced into the inversion layer per cm^2.

Electrons in an inversion layer have been the subject of many physical investigations. Like carriers at a semiconductor heterojunction (see section 6.2), these electrons are confined in a potential well. Just above the bottom ($\simeq 0.1\ eV$), the width of this well is still smaller than L. Thus, carriers of such energy are freely mobile only parallel to the well. One has a quasi 2-dimensional electron gas. In it, confinement effects, like the formation of subbands, can be observed. By means of an applied voltage, the density of the electron gas can be varied. In this context, the accumulation layer at the interface between SiO_2 and Si has been of particular interest. Historically, it was the 2D gas of this layer in which carrier confinement effects were studied first, and in which the quantized Hall effect was discovered.

7.4.3 MOSFET

Whether or not the charge density of a semiconductor inversion layer suffices to cause a resistance change of sufficient magnitude to function as a transistor, depends (among other things) mainly on the specific resistivity of the semiconductor material in the absence of an applied voltage. This must be as high as possible. The highest possible resistivity, or the smallest conductivity, of a semiconductor material is observed when it is in its intrinsic state. This fact is used to advantage in the most important type of field effect transistor, the so-called *MISFETs* (Metal Insulator Semiconductor Field Effect Transistor). Here, one exploits the fact that in a pn-junction a depletion region is formed wherein the charge carrier concentrations have intrinsic values. The p-region is embedded between two n-regions, as shown in Figure 7.22. If one applies a positive voltage between the left n-region (*source*) and the right n-region (*drain*) (Figure 7.22), then the left pn-junction is biased in the flow direction, and the right in the blocking direction. Between the two n-regions, one therefore has a p-region which is almost completely depleted of holes. In the theory developed above, this may be taken into account formally by adding a positive prevoltage U_V to the voltage U applied to the insulator-semiconductor junction from outside, i.e. by the replacement

$$U \rightarrow U + U_V. \tag{7.120}$$

For $U = 0$, the semiconductor region between source and drain, the so-called *channel*, has high intrinsic resistivity. If a positive voltage $U = U_G$ is applied between the bulk of the p-semiconductor (*substrate*) and the metal layer on

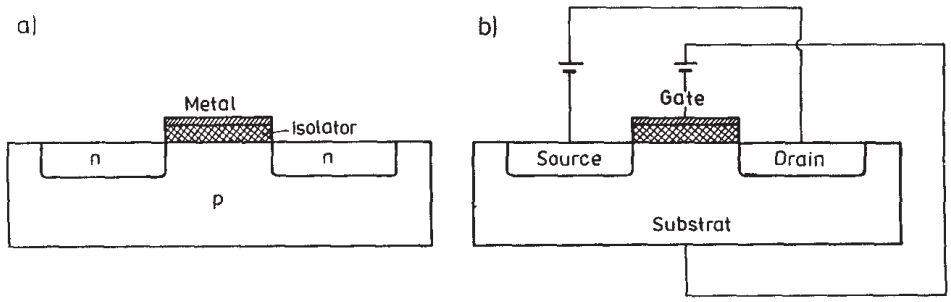

Figure 7.22: Structure of a MISFET (a). In part (b) the switching scheme of the MISFET is shown.

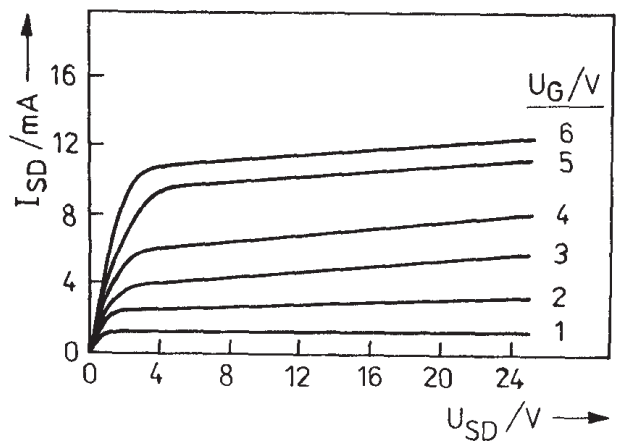

Figure 7.23: Influence of the gate voltage U_G on the I_{SD} – versus – U_{SD} current-voltage characteristics for a n-channel MOSFET of enhancement-type. *(After Völz, 1986.)*

top of the insulator (*gate*), large enough to create an inversion layer, then the channel becomes a good conductor. The current in the source-drain circuit can be tuned by the voltage U_G of the gate-substrate circuit since the electron density of the inversion layer depends on U_G. The voltage U_G is called *gate voltage*, and the minimum gate voltage U_G necessary to a achieve inversion is the *threshold voltage*. If one also includes a working resistor in the source-drain circuit and compares the power changes in this resistor with those in the gate-substrate circuit, one finds that the former exceeds the latter appreciably. In this sense, the MISFET operates as an amplifier, just like the bipolar transistor. Because of the insulating layer below the gate electrode, the MISFET has a much larger input resistance than the bipolar

transistor. The output resistance of the MISFET, i.e. that of the source-drain circuit, is small because of the accumulation layer between source and drain.

If the MISFET is realized using Si as the semiconductor material and SiO_2 as the insulator, one has the Metal Oxide Semiconductor Field Effect Transistor, abbreviated *MOSFET*, which is by far the most important MISFET. The basic requirement that the electrons of the inversion layer shall not be captured by interface states, are met by the MOSFET extremely well. Just as the base of a bipolar transistor can be made either of p- or n-material, one also has two possibilities in the case of the unipolar MOSFET – the charge carriers in the conducting channel can either be electrons, as has been assumed thus far, or holes. In the first case one speaks of a *n-channel MOSFET*, and in the second of *p-channel MOSFET*. The prevoltage U_V, which was introduced above only formally to simulate the intrinsic state, can really exist for various reasons, e.g., because of positively charged centers within the oxide arising during its formation. The prevoltage can take such large values that inversion exists even without an applied gate voltage. Then the transistor is already in its conducting state at zero gate voltage. To switch it into its blocking state, one must apply a negative gate voltage in the case of the n-channel MOSFET, and a positive gate voltage in the case of the p-channel MOSFET. One says that the transistor operates in the *depletion mode*. If the transistor is blocked without an applied voltage and changes into the conducting state by applying a positive (for n-channel) or negative (for p-channel) gate voltage, one has the *enhancement mode*. The current-voltage characteristics of the source-drain circuit of a n-channel-enhancement-mode MOSFET are shown in Figure 7.23 for different gate voltages. One recognizes how the source-drain current at a fixed value of U_{SD} increases with increasing gate voltage U_G. Somewhat unexpectedly, for a fixed gate voltage U_G, the current saturates at higher source-drain voltages U_{SD}. This is a consequence of the fact that the effective gate voltage in the vicinity of the drain electrode becomes smaller and smaller as the drain potential grows larger. This creates a 'pinch off' of the inversion layer, making the current stay constant.

The MOSFET represents the most important electronic component of digital circuits in microelectronics. Here it is used as an electrically controllable switch, meaning that its control function is reduced to two states only, one with maximum output power, corresponding to a binary '1', and another having minimum output power, corresponding to a binary '0'. Today, MOSFETs can be made as small that millions can be integrated in one single Si chip, more than of any other electronic component. It is this ultra-large scale integration (ULSI) technique which has made modern computer and communication technologies possible.