THE PHYSICAL THEORY OF TRANSISTORS

Leopoldo B. Valdes
MANAGER, RESEARCH AND DEVELOPMENT DEPARTMENT
RHEEM SEMICONDUCTOR CORPORATION
MOUNTAIN VIEW, CALIFORNIA

McGRAW-HILL BOOK COMPANY, INC.
New York Toronto London
1961

THE PHYSICAL THEORY OF TRANSISTORS

 Library of Congress Catalog Card Number 60-16733

66807

THE MAPLE PRESS COMPANY, YORK, PA.

THE PHYSICAL THEORY OF TRANSISTORS

McGRAW-HILL ELECTRICAL AND ELECTRONIC ENGINEERING SERIES

Frederick Emmons Terman, *Consulting Editor*
W. W. Harman and J. G. Truxal, *Associate Consulting Editors*

Ahrendt and Savant · Servomechanism Practice
Angelo · Electronic Circuits
Aseltine · Transform Method in Linear System Analysis
Bailey and Gault · Alternating-current Machinery
Beranek · Acoustics
Brenner and Javid · Analysis of Electric Circuits
Bruns and Saunders · Analysis of Feedback Control Systems
Cage · Theory and Application of Industrial Electronics
Cauer · Synthesis of Linear Communication Networks
Chirlian and Zemanian · Electronics
Clement and Johnson · Electrical Engineering Science
Cuccia · Harmonics, Sidebands, and Transients in Communication Engineering
Cunningham · Introduction to Nonlinear Analysis
Eastman · Fundamentals of Vacuum Tubes
Evans · Control-system Dynamics
Feinstein · Foundations of Information Theory
Fitzgerald and Higginbotham · Basic Electrical Engineering
Fitzgerald and Kingsley · Electric Machinery
Frank · Electrical Measurement Analysis
Geppert · Basic Electron Tubes
Glasford · Fundamentals of Television Engineering
Greiner · Semiconductor Devices and Applications
Hammond · Electrical Engineering
Hancock · An Introduction to the Principles of Communication Theory
Happel and Hesselberth · Engineering Electronics
Harman · Fundamentals of Electronic Motion
Harrington · Introduction to Electromagnetic Engineering
Harrington · Time-harmonic Electromagnetic Fields
Hayt · Engineering Electromagnetics
Hill · Electronics in Engineering
Johnson · Transmission Lines and Networks
Koenig and Blackwell · Electromechanical System Theory
Kraus · Antennas
Kraus · Electromagnetics
Kuh and Pederson · Principles of Circuit Synthesis
Ledley · Digital Computer and Control Engineering
LePage · Analysis of Alternating-current Circuits
LePage · Complex Variables and the Laplace Transform for Engineers
LePage and Seely · General Network Analysis

LEY, LUTZ, AND REHBERG · Linear Circuit Analysis
LYNCH AND TRUXAL · Introductory System Analysis
MILLMAN · Vacuum-tube and Semiconductor Electronics
MILLMAN AND SEELY · Electronics
MILLMAN AND TAUB · Pulse and Digital Circuits
MISHKIN AND BRAUN · Adaptive Control Systems
MOORE · Traveling-wave Engineering
PETTIT · Electronic Switching, Timing, and Pulse Circuits
PETTIT AND MCWHORTER · Electronic Amplifier Circuits
PFEIFFER · Linear Systems Analysis
REZA · An Introduction to Information Theory
REZA AND SEELY · Modern Network Analysis
ROGERS · Introduction to Electric Fields
RUDENBERG · Transient Performance of Electric Power Systems
RYDER · Engineering Electronics
SCHWARTZ · Information Transmission, Modulation, and Noise
SEELY · Electron-tube Circuits
SEELY · Electrical Engineering
SEELY · Introduction to Electromagnetic Fields
SEELY · Radio Electronics
SEIFERT AND STEEG · Control Systems Engineering
SISKIND · Direct-current Machinery
SKILLING · Electronic Transmission Lines
SKILLING · Transient Electric Currents
SPANGENBERG · Fundamentals of Electron Devices
SPANGENBERG · Vacuum Tubes
STEVENSON · Elements of Power System Analysis
STEWART · Fundamentals of Signal Theory
STORER · Passive Network Synthesis
STRAUSS · Wave Generation and Shaping
TERMAN · Electronic and Radio Engineering
TERMAN AND PETTIT · Electronic Measurements
THALER · Elements of Servomechanism Theory
THALER AND BROWN · Analysis and Design of Feedback Control Systems
THOMPSON · Alternating-current and Transient Circuit Analysis
TOU · Digital and Sampled-data Control Systems
TRUXAL · Automatic Feedback Control System Synthesis
VALDES · The Physical Theory of Transistors
WILLIAMS AND YOUNG · Electrical Engineering Problems

PREFACE

When a large technological achievement, like the invention of the transistor, takes place, we usually find that it is the result of a composite effort involving knowledge from various areas of science and engineering. In a very short time the new technology acquires its own identity, and it becomes impossible to identify it as a branch of one of the existing fields of knowledge. For example, although transistor technology is derived primarily from physics, chemistry, metallurgy, and electrical engineering, it cannot be described as being part of any one of these fields. University courses on the theory of transistors are just as likely to be taught in the physics department as in the electrical engineering department. In industry, electrical engineers grow semiconductor crystals and work on other metallurgical problems, whereas chemists measure the electrical characteristics of transistors.

This book is intended as a textbook for a graduate course in the hybrid field called transistor physics. It is planned for a course on the theory of diodes and transistors, with primary emphasis on the internal structure of these devices. To analyze the flow of current through semiconductor materials and to derive relationships between the electrical characteristics of transistors and their physical structure are the final objectives of this book. However, many other aspects of solid-state physics must be discussed in order to reach such goals.

It is presumed that the course will be taught in the physics or electrical engineering department of a university but that the students may have backgrounds from widely different fields. As a benefit to such students and to those individuals using the book for reference purposes, it has been deemed advisable to include some introductory material in the book. This material is also included in order to unify the various concepts which comprise the field of transistor physics, in the manner seen by someone who participated in the development of the field from almost its conception to the huge technological complex which it now comprises.

This book is divided into three parts. The first four chapters include introductory or review material. The second part, Chapters 5 to 13, is devoted to the more physical aspects of transistor theory and deals in generalities. The material in Chapters 14 to 18 is devoted to analyzing specific device structures and the electrical characteristics of diodes and transistors. Each part of the book may be treated as a separate and complete entity. Part 2, for example, contains the material which should be included in a solid-state physics course, and Part 3 deals with the more applied aspects which are required in an electrical engineering course.

The introductory material included in Part 1 has made it possible to keep at a minimum the number of prerequisites for reading this book. Prior courses on linear differential equations and on electromagnetic theory (where elementary vector analysis is used and the wave equation is solved in cylindrical and spherical coordinates) are recommended. Knowledge of atomic or solid-state physics, four-terminal-network theory, chemistry, and metallurgy will aid in understanding certain parts of the book but is not considered a prerequisite.

Although the elementary description of transistor theory given in Chapter 4 at first may appear superfluous, it is required as a framework for the general discussions to follow in Chapters 7, 9, 10, and 11. It is believed that the reader benefits by having a broad picture of the desired end results before embarking on a detailed study of the various intermediate mathematical steps needed in order to achieve the results.

The object of the material in Part 2 is to establish the physical concepts and develop the mathematical tools required in order to formulate later (in Part 3) the theory of semiconductor devices. For example, the transport of carriers through a semiconductor is treated in a general manner by simply solving the transport equations for several different geometrical and mathematical boundary conditions. *p-n* junction theory is treated as a physical phenomenon arising from a discontinuity in a single-crystal semiconductor, and its practical device aspects are neglected. Discussions of the thermoelectric, optical, magnetic, and surface properties of semiconductors (Chapters 12 and 13) are included in Part 2 for completeness, although it is recognized that some readers will prefer to go directly from Chapter 11 to Chapter 14 and to treat this material later.

Part 3 of this book sets out to analyze specific device structures. Equations are obtained for hole and electron current, equivalent-circuit parameters, and large-signal device properties in terms of the geometrical dimensions and the properties of the semiconductor material in each region of the transistor structure. The emphasis, however, is on laying a general groundwork in transistor theory rather than on solving a large number of specific problems. It is hoped that the student who completes a course based on this text will be in a position to analyze different and more complicated transistor structures than those treated here. The problems at the end of each chapter should aid in training the student for this purpose.

It is also proper to mention that the nonstudent readers of this book have not been overlooked. Those engaged in transistor research, development, or manufacturing should find this text a useful introduction to the field. Circuit development engineers, technical managers, and those who are not in the transistor field but who have a certain amount of scientific curiosity will derive some satisfaction from the introductory chapters. The separation of the various levels of technical complexity which has been achieved by dividing the book into three parts should help readers with different backgrounds and interests.

It is hoped that this book will not only make a contribution as a text but that it will also perform another important function. It is evident that the design theory of transistors is scattered throughout the technical literature. This presents a very serious problem to a beginner in the semiconductor field, not knowing which articles describe the more significant contributions to transistor theory. Yet, it is believed that any attempt to unify this knowledge demands familiarity with the practical problems of device design and fabrication. Even if this book does nothing more than make its readers aware of the need for a complete treatise on the design theory of transistors, it will be considered a success in this respect. It is hoped that the bibliographical references will help in this purpose.

Now is the time for a few personal notes. It would be impossible for me to acknowledge assistance from even a small fraction of the people with whom I have been associated in my professional career. However, it is impossible to escape mentioning J. A. Morton. He is responsible for the fact that I started in 1949 to work in the transistor group at Bell Telephone Laboratories and for much of my early education in transistor research and development. The engineering extension department of the University of California is responsible indirectly (by requesting that I teach a course in transistor theory) for making me aware of the need for a book of this type. For the final push, the time and facilities required to finish this task, I am indebted to the Rheem Semiconductor Corporation.

Finally, there is my wife, whose constant encouragement and cooperation were necessary for a project of this nature.

L. B. Valdes

CONTENTS

Preface vii

Part 1—Introductory Concepts

Chapter 1 Introduction to Diodes and Transistors 3

1-1. Electrical behavior of diodes 3
1-2. Diode structural considerations 4
1-3. External characteristics of transistors 7
1-4. Types of transistors 12
1-5. Photoelectrical transducers 15

Chapter 2 Crystal Structure and Imperfections 17

2-1. The crystalline structure 18
2-2. Imperfections in semiconductor crystals 22
2-3. Perfect crystals and chemical impurities 23
2-4. Lattice defects and dislocations 26

Chapter 3 Semiconductor Materials 30

3-1. Physical properties 30
3-2. Electrical conductivity 32
3-3. Intrinsic semiconductor 33
3-4. Hole and electron conduction 36
3-5. Extrinsic semiconductor 39
3-6. Hole and electron densities 43
3-7. Resistivity and lifetime 45
3-8. Large impurity concentration 47

Chapter 4 Physical Basis for Transistor Behavior 49

4-1. Flow of holes and electrons 49
4-2. Electron energy and hole and electron densities . . . 55
4-3. *p-n* junctions 57
4-4. Junction transistors 60

Part 2—Electronic Properties of Semiconductor Materials

Chapter 5 Electron Energies and Distributions 67
5-1. Review of atomic phenomena 67
5-2. Concept of energy bands 72
5-3. Energy-band theory of solids 76
5-4. Fermi-Dirac distribution law 83
5-5. Density of quantum states in allowed energy bands . . 87

Chapter 6 Fermi Level in Semiconductors 95
6-1. Effective density of energy states, electrons, and holes . 95
6-2. Fermi level and electrostatic potential in intrinsic semiconductors 98
6-3. Fermi level in n-type semiconductors 102
6-4. Fermi level in p-type and compensated semiconductors . 105
6-5. Effect of large donor or acceptor density 107
6-6. Temperature dependence of the Fermi level 110
6-7. Continuity of the Fermi level across a junction between dissimilar materials 112

Chapter 7 Conservation and Motion of Charge 117
7-1. Basic equations for semiconductors 117
7-2. Generation and recombination of carriers and the continuity equation 120
7-3. Current flow in semiconductors 122
7-4. The diffusion equation 124
7-5. Internal built-in fields in semiconductors 126

Chapter 8 Mobility and Lifetime of Electrons and Holes . . . 129
8-1. Lattice and impurity scattering mobilities 129
8-2. Effective masses and cyclotron resonances 133
8-3. Experimental mobility data 136
8-4. Recombination of minority carriers 142
8-5. Lifetime of electrons and holes 146
8-6. Nonrecombining electron and hole traps 149

Chapter 9 Current Flow in Semiconductors 154
9-1. Planar flow in semi-infinite geometry 154
9-2. Diffusion between planar source and sink 157
9-3. Flow of alternating current 158
9-4. Drift plus diffusion in planar geometry 160
9-5. Ambipolar diffusion 162
9-6. Diffusion in cylindrical coordinates 164
9-7. Radial diffusion in spherical coordinates 165

Chapter 10 The Theory of Planar p-n Junctions 170

10-1. Current flow in nonuniform semiconductors 170
10-2. Space charge in semiconductors. 176
10-3. Electrostatic potential in planar p-n junctions at thermal equilibrium 178
10-4. Planar p-n junctions with an external bias 183
10-5. Excess carriers and current flow through a depletion layer 188

Chapter 11 Current Flow in p-n Junctions 192

11-1. Planar p-n junction diodes 192
11-2. Reverse-biased p-n junction. 197
11-3. Reverse current breakdown and multiplication 203
11-4. p-i and n-i junctions. 209
11-5. Hemispherical p-n junction. 212

Chapter 12 Thermoelectric, Optical, and Magnetic Effects . . 221

12-1. Thermoelectric power in semiconductors 221
12-2. Optical properties of silicon and germanium 224
12-3. Optical generation of carriers 227
12-4. Hall effect in semiconductors 228
12-5. Suhl effect and other magnetoelectric phenomena . . 232

Chapter 13 Surface Properties. 236

13-1. Review of semiconductor surface properties 236
13-2. Theory of surface states. 238
13-3. Inversion layers and surface conductance. 244
13-4. Measurement of surface properties. 248
13-5. Recombination of minority carriers at the surface . . 252
13-6. Metal-Semiconductor contacts 254

Part 3—Properties of Semiconductor Devices

Chapter 14 Electrical Characteristics of Semiconductor Diodes . 259

14-1. Current flow in diodes 259
14-2. Small-signal diode impedance 261
14-3. Large-signal transient behavior. 267
14-4. High-current characteristics. 275

Chapter 15 Direct-current Theory of Junction Transistors . . 280

15-1. Transistor d-c flow equations 280
15-2. Carrier transport in the base region 284
15-3. Series ohmic resistances 290
15-4. Large-signal d-c behavior 296

Chapter 16 Small-signal Transistor Theory 301

16-1. Theory of four-terminal networks 301
16-2. Simple transistor equivalent circuit 304
16-3. Transistors current multiplication 307
16-4. Low-frequency equivalent circuit 314

Chapter 17 Transistors in High-frequency Amplifying and High-speed Switching Circuits. 323

17-1. Decrease of alpha at high frequencies 323
17-2. Equivalent-circuit parameters at high frequencies . . . 327
17-3. High-frequency small-signal circuit behavior 330
17-4. Transistors in switching circuits 334

Chapter 18 Special Structures and Other Factors Affecting Transistor Performance 341

18-1. Structural considerations for high-frequency transistors. 341
18-2. Esaki diode 347
18-3. Special transistor structures. 350
18-4. Noise in transistors 354
18-5. Power dissipation and operation at different ambient temperatures 356

Name Index 359

Subject Index 362

PART 1

INTRODUCTORY CONCEPTS

CHAPTER 1

INTRODUCTION TO DIODES AND TRANSISTORS

Prior to the invention of the transistor, electronic engineers had only one device capable of amplifying electrical signals over a wide frequency range. Therefore, a new era in the design of active circuits began when transistors became sufficiently reliable to offer the electronic vacuum tube some real competition. The success with which transistors have been able to perform many of the circuit functions previously reserved exclusively for electronic tubes can be determined by observing the amazing growth of the transistor industry during its first decade.

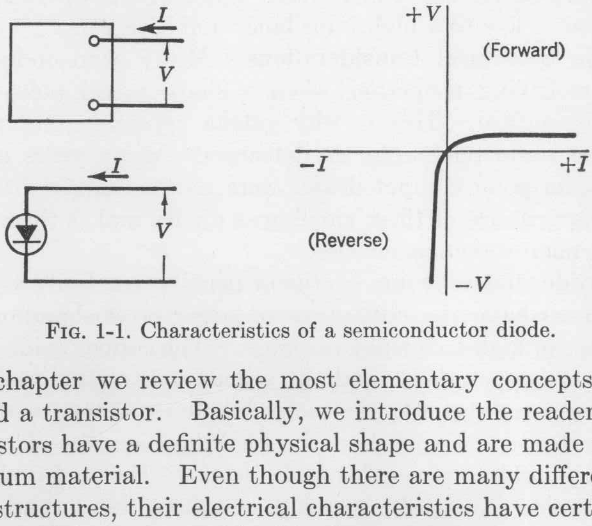

FIG. 1-1. Characteristics of a semiconductor diode.

In this chapter we review the most elementary concepts and define a diode and a transistor. Basically, we introduce the reader to the fact that transistors have a definite physical shape and are made from silicon or germanium material. Even though there are many different types of transistor structures, their electrical characteristics have certain features in common, which are treated in this chapter.

1-1. Electrical Behavior of Diodes. A diode is a nonlinear two-terminal device. The electrical impedance is low for one applied bias polarity and high when the bias polarity is reversed. Figure 1-1 shows the symbol used to represent a diode in circuits and the voltage-current characteristics of an ideal diode. The arrow in the diode symbol points in the direction of "easy flow" for current (low impedance).

The electrical characteristics of Fig. 1-1 are divided into forward and reverse regions. A forward bias always produces a large current at small applied voltages. In Fig. 1-1 this corresponds to positive voltages and currents. The reverse characteristics are obtained with a negative-bias polarity.

The current I flowing through a diode can be expressed as

$$I = I_S(e^{qV/kT} - 1) \tag{1-1}$$

where I_S is a constant for any one diode, $q/kT = 38.6$ at 27°C, and V is the applied voltage. It should be noted that, for forward voltages in excess of 50 mv, the current increases exponentially and multiplies itself by a factor of 10 every 60 mv. In the reverse direction the exponential term approaches zero, and the current equals $-I_S$.

There are three basic uses for semiconductor diodes. They may be used as power rectifiers or detectors, where the diode is used to convert alternating current to direct current. In mixers, modulators, and other low-level applications (where the applied signal is much less than 1 volt), the diodes are used simply as nonlinear elements in order to generate signals at other frequencies. Finally, the diodes may be used as on-off gates in switching circuits and in other applications where it is desired to switch from a low to a high impedance and vice versa.

1-2. Diode Structural Considerations. Many semiconductor materials show rectifying properties when a sharp metal point is pressed against their surface. This is why galena crystals were used in the early days of radio telegraphy as detectors. Many years later silicon and germanium point-contact diodes were used as mixers in radar sets. The small capacitance of these small-area diodes makes them very good detectors at microwave frequencies.

Copper oxide and selenium rectifiers usually are built as large-area devices. These have the advantage of larger power-handling capacity at the expense of high-frequency response. The copper oxide is obtained by oxidizing a copper sheet, and the selenium is melted and spread or evaporated on a metal sheet. In each of these rectifiers a second contact is needed in order to complete the electrical circuit. Copper oxide rectifiers use graphite or aquadag and a lead washer held under pressure against the oxide layer. In selenium rectifiers the second contact is produced by spraying the selenium surface with a low-melting-point alloy.

Junction Diodes. All the point-contact and large-area rectifiers described so far rely on the properties of the metal-to-semiconductor contact. On the other hand, modern silicon and germanium diodes are made from a single piece of semiconductor material which has a small concentration of chemical impurities. The desired rectification is obtained within the body of the semiconductor. The metallic contacts which are made to the semiconductor in order to complete the electrical circuit are ohmic (nonrectifying) in nature and have a very low resistance.

Silicon and germanium devices are those of primary concern to us. Figure 1-2 illustrates point-contact and large-area (junction) diodes which are fabricated with germanium material. Electrical connections are made through the metallic contacts. Rectification is achieved in the point-contact diode in the region of the germanium which is immediately adjacent to the tungsten point. In the large-area diode, rectification takes place in the center of the germanium block, along the junction plane between the regions which are labeled n-type and p-type.

p-n Junctions. The designations n-type and p-type refer to the type of current carrier which predominates in the semiconductor. In metals it is known that current is due to the flow of negatively charged *electrons*. However, in semiconductors it is possible also to have positive carriers, which are called *holes*. Thus, the majority of the current carriers in an n-type semiconductor are negative electrons. Similarly, there is an excess of positive holes in a p-type semiconductor.

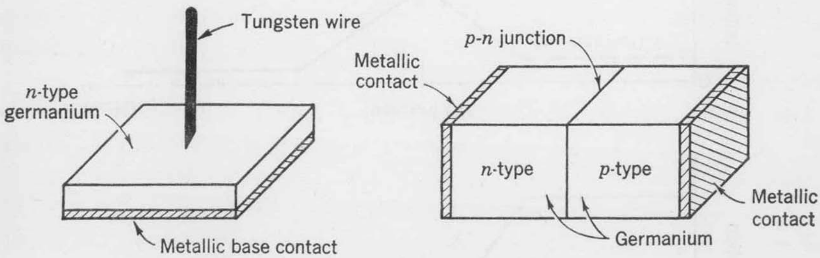

FIG. 1-2. Point-contact and junction diodes.

These positive and negative carriers are created by adding chemical impurities to the semiconductor at the time the material is produced. Very pure silicon or germanium is called an *intrinsic* semiconductor. When certain chemical impurities are present, it is called an *extrinsic* semiconductor. An impurity such as phosphorus, arsenic, or antimony makes the material n-type. An extrinsic p-type semiconductor is obtained by adding boron, aluminum, gallium, or indium as the chemical impurity. In either case the impurity concentration is very small; for example, there may be 1 atom of antimony for every 10^7 germanium atoms.

In order to have p-n junctions, the chemical composition of a single piece of semiconductor must vary as a function of position within the semiconductor. For example, the junction diode of Fig. 1-2 may be fabricated from a rod of germanium which has a large concentration of antimony atoms on the left-hand end and a large boron concentration on the right-hand side. Let us consider now that the antimony concentration decreases as a function of distance and approaches zero on the