An Introduction

TO THE THEORY AND PRACTICE OF TRANSISTORS

J. R. TILLMAN

F. F. ROBERTS

AN INTRODUCTION TO THE THEORY AND PRACTICE OF TRANSISTORS

BY

J. R. TILLMAN, D.Sc., A.R.C.S.

AND

F. F. ROBERTS, B.Sc.(Eng.), A.M.I.E.E.

LONDON

SIR ISAAC PITMAN & SONS, LTD.

Preface

NEWCOMERS to the subject of transistors and related semiconductor devices will find an overwhelming accumulation of papers describing the physical principles of semiconductors, the preparation of materials, the fabrication of devices, the internal electronics, the electrical representation of the devices and numerous applications. They will find also an increasing number of textbooks, most of which tend to be limited to two or three main divisions of the subject. Our aim in adding another book has been to present a wider treatment in one volume, mainly by picking out those parts which seem to us to be of most permanent value to a student of the subject, particularly in helping him to obtain a confident quantitative understanding of all the primary physical effects on which minority-carrier transistors depend for their electrical behaviour and an insight into the potentialities and limitations of the devices as circuit elements. It was not our intention to write a handbook for the designer of either devices or circuits.

Thus in Part I we deal with the electronic properties of homogeneous semiconductors, of isolated p–n junctions and of the interaction between the pair of junctions which make up the transistor. The theories of bulk and surface recombination have been given in some detail because these phenomena control the fundamental property of current gain. The role of surfaces is emphasized again, in analysing a model of a transistor combining an active base region with lateral diode sub-regions, with the object of showing why the electrical behaviour of practical transistors may not conform at all well with over-simplified equivalent circuits.

Part II begins with a chapter covering the preparation of monocrystals of germanium and silicon of controlled purity, and outlines the techniques used to produce junction devices. Major changes continue to occur in this part of the subject, as seen for example in the rapid adoption of diffusion techniques, and no description can remain fully up-to-date for long. Readers who are less interested in the fabrication than in the use of transistors can, on first reading, bypass this chapter. The two remaining chapters of Part II describe the properties of transistors and diodes and many of the basic circuits in which the two devices are used. These descriptions contain enough detail, in analytical form, to suggest how circuits may be designed and analysed in meeting the requirements set for equipments to be used in the communication, computer, entertainment and other industries; but the future may bring much more rationalization to circuit design.

We are conscious that we have omitted parts of the subject which some readers would have wished to see included. Thus we give no description of unipolar transistors or of the relative virtues of the different cross-sections proposed for them; there is however little sign yet that the unipolar transistor can compete commercially with minority-carrier transistors. The reliability of transistors—a subject of much importance to many industrial users—depends vitally on the quality of the encapsulation used, but we have not discussed the merits of the different types of sealing, nor the factors affecting reliability as a whole, in view of the limited amount of statistically significant information yet available. Our chapters on properties and applications refer mainly to the well established uniform-base transistor, although much recent circuit work, centred on the transistors with diffused bases (often Mesa types) which must be used for the highest frequencies and switching rates, has slight differences in approach. By contrast we have, largely for historical reasons, included brief sections on the point-contact transistor, despite its lack of importance commercially.

In writing this book we have drawn on many original and review papers, on some books in the field, and on our own experience and that of some of our colleagues. We can not make all acknowledgements separately, but some are made specifically in the text, as the occasion arises, by references to the source. We thank the Engineer-in-Chief of the Post Office for permission to reproduce the plates, to make use of some material in Sections 1.9, 4.3.3, 5.1.2 and 5.5, and for the opportunity, without which this book would not have been written, of working in this stimulating field for the past several years.

Dollis Hill, 1960.

J. R. T.
F. F. R.

Principal Letter Symbols

a	internal current gain
a_0	low-frequency value of a
a_I	average spacing of ionized impurities; value of a in inverse direction
a_N	value of a in normal direction
A	cross-sectional area of crystal or junction; amplification factor
A_{tn}, A_{tp}	capture cross-sections of traps
B	elastic bulk modulus; magnetic flux density
c	coefficients in general transistor equations
c_n	capture coefficient
C	capacitance
C_{De}, C_{D11}, C_{Dc}, C_{D22}	emitter and collector diffusion capacitances
C_{te}, C_{tc}	emitter and collector depletion layer capacitances
C_c	collector capacitance
$C_{b'c}$, $C_{b'e}$	capacitances in hybrid-pi equivalent circuit
C_0	initial or surface concentration
D	diffusion coefficient
D_a, D_d, D_n, D_p	diffusion coefficients of acceptors, donors, electrons, holes
e_r	energy level of state r
e_F, e_m, e_{Fi}	Fermi level, mid-gap level, intrinsic Fermi level
e_{Fn}, e_{Fp}	electron and hole imrefs (quasi Fermi levels)
e_n	emission coefficient
E	electric potential gradient; energy of electron
E_g, E_{gn}, E_{gv}	energy gap, non-vertical gap, vertical gap

E_C, E_V	energy levels of conduction and valence band edges
E_a, E_d	energy levels of acceptor and donor states
f	frequency
f_a	frequency at which $\mid a \mid = a_0/\sqrt{2}$
f_α	$(=f_{hfb})$ frequency at which $\mid \alpha \mid = \alpha_0/\sqrt{2}$
f_1	frequency at which $\mid \beta \mid (= \mid h_{fe} \mid) = 1$
f_{Max}	maximum frequency of oscillation
$F(\,)$	Fermi–Dirac function
g	generation rate; conductance
g_m	mutual conductance
h	Planck's constant; hybrid parameters of a 4-pole (*see* eqn. 5.4)
i	instantaneous value of current: varying component has l.c. subscript; average and instantaneous total values have u.c. subscript
I	average (d.c.) or r.m.s. value of current: r.m.s. value has l.c. subscript; total value has u.c. subscript
I_{EBO}, I_{CBO}	emitter and collector saturation (leakage) currents
I_S	saturation current of a p-n junction
J	current density
k	Boltzmann's constant; segregation coefficient (ratio of solubilities)
K	$= [(1 + \mathrm{j}\omega\tau_p)/\mathrm{D}_p\tau_p]^{1/2}$
l	mean free path (of electron)
L	diffusion distance; inductance
L_p	diffusion distance of holes
m	mass of free electron; grading factor
m_e, m_h	effective mass of electron or hole in crystal
M	effective mass of phonon; avalanche multiplication factor of p-n junction
n	density of electrons; turns ratio
n_i	intrinsic density of electrons or holes
N	noise factor
N_C, N_V	equivalent density of states in conduction and valence bands
N_a, N_d, N_I, N_t	volume densities of acceptors, donors, ionized impurities, traps
N_{ts}	surface density of surface traps

p	density of holes
p_n	thermodynamic equilibrium density of holes in n-type semiconductor
P	$= p - p_n$; power
q	electronic charge
Q_V	total charge of one sign in depletion layer at bias V
Q	circuit magnification factor
r_b, r_c, r_e	resistive components of simple T equivalent circuit (*see* Fig. 5. 24)
r_{bs}, r_{cs}, r_{es}	base, collector and emitter series resistances (*see* Chap. 3)
$r_{bb'}$, $r_{b'e}$, $r_{b'c}$	resistive components of hybrid-pi equivalent circuit
R_G, R_L	generator and load resistances
s	surface recombination velocity
S	thermal resistance
$d\mathbf{S}$	element of surface (vector)
t, t_1, t_2	time, storage time, decay (fall) time
t_f	mean free time
T, T_a, T_j	temperature, ambient and junction temperatures
U_c, U_{cn}, U_{cp}	capture rate, capture rates for electrons and holes
v	velocity; instantaneous value of voltage (*see* i for subscript conventions)
v_d	drift velocity
V	average (d.c.) or r.m.s. value of voltage (*see* I for subscript conventions)
V_A	avalanche, reference or Zener voltage
V_B	surface potential rise (of surface barrier)
V_F, V_R	forward and reverse voltages across p-n junctions
V_T	turnover voltage
V_{BB}, V_{CC}, V_{EE}	supply voltages to base, collector and emitter circuits
V_{CEsat}	collector-emitter voltage of saturated transistor
W	base width; width of zone

y	admittance parameters of a general transistor or 4-pole (*see* eqns. (3.48) and (5.3))
z	impedance parameters of a general transistor or 4-pole (*see* eqn. (5.2))
α	$(= -h_{fb})$ external current gain with common base
β	$(= h_{fe})$ external current gain with common emitter; base transmission factor (Sec. 4.2.2)
γ	emitter injection efficiency
ε, ε_0	absolute permittivity, permittivity of space
λ, λ_0	wavelength, threshold wavelength
μ, μ_n, μ_p	mobility, mobility of electrons or holes
μ_{ec}	Early feedback voltage ratio
μ	magnetic permeability (absolute)
ϕ	magnetic flux
ϕ_0, ϕ_s, ϕ_t	potentials characterizing surface recombination
Φ	phase-shift
ρ	resistivity
σ	conductivity
τ_n, τ_p	lifetimes of electrons and of holes
τ_1, τ_2	component transient lifetimes
ω	$(= 2\pi f)$ angular frequency

Introduction

Historical

Many materials having electrical conductivities within an order or two of 10^{-1} (ohm cm)$^{-1}$ have long been called, perhaps rather loosely, semiconductors; they were thus distinguished from metals, with conductivities of $\sim 10^5$ (ohm cm)$^{-1}$ on the one hand, and insulators, with conductivities of $\sim 10^{-10}$ (ohm cm)$^{-1}$ on the other. Their study received a powerful and immediate stimulus when transistor action was discovered in 1948. Physicists saw that it would inevitably lead to a major revision of some parts of the basic theory of semiconductors and hence to reasons for some earlier unexplained observations, and that it promised many more discoveries; electrical engineers saw a new active element to rival the thermionic valve.

The earlier history of semiconductors contained numerous experimental observations which confounded the theories current at the time without otherwise much increasing the understanding of the subject. Often the experimenters were ignorant of at least one of the principal conditions controlling their observations and confirmation in another laboratory could be a matter of chance. Three earlier discoveries do stand out, however. First, the electrical conductivity of many semiconductors increases with temperature over a range, often a large range, of temperature. Second, the conductivity increases with illumination. Third, the contact, between a metallic point and some at least of the semiconductors, is rectifying, a fact obscured for some time by a belief that the rectification was merely the manifestation of a non-ohmic behaviour in the bulk of the material.

The rectifiers attracted further attention when they proved useful in detecting radio signals. As a result a range of semiconducting materials, mostly minerals, were investigated; but reliable point-contact diodes were rare. On the other hand the copper oxide rectifier—an area rectifier useful, in different forms, at both power and broadcast frequencies—soon became, and has remained, a ommercial product. In its early days, improvements in its performance came mainly as the results of empirical changes and refinements in technology rather than as practical applications of the current theories. Later, the two elements silicon and selenium began to figure prominently. Selenium, the first material in which photoconductivity had been observed, was used to make area rectifiers efficient at both audio and power frequencies; the rectifiers were made by spraying a metallic layer on to a thin film of selenium, prepared from powder by melting and solidifying. Physical investigation of the units so made proved difficult, however, if only because the selenium was polycrystalline and the metal–selenium contact could not be defined sufficiently well to permit a sound analysis to be attempted. Pressure contacts of metallic points on silicon, though equally intractable in analysis, proved useful in detecting radio signals of wavelengths shorter than those easily detected by thermionic diodes. By the late 'thirties the metallurgy of silicon had been so improved that control of conductivity was possible and the experimenter had one less uncontrolled variable to contend with. The idea that the sign of the carriers responsible for the electrical conductivity, as deduced from observations of the Hall effect, was peculiar to each semiconductor, i.e. each semiconductor could be classified once and for all as either p-type (positively charged carriers) or n-type (negatively charged carriers) had to be modified; for, by suitable metallurgy, specimens of both p-type and n-type silicon could be made. The additives to silicon could decide the classification which was therefore not an unconditional property of silicon. Silicon point-contact rectifiers are still the most widely used low-level detectors and mixers in radar and other applications of microwaves.

Germanium, an element little used before 1940, resembles silicon chemically and crystallographically; both are in the fourth column of the periodic table and have the structure of diamond. Germanium, too, proves to be a semiconductor; its conductivity ultimately increases rapidly with temperature and increases with illumination. It did not at once compete with silicon in the making of diodes sensitive at microwave frequencies, but it did make possible diodes capable of withstanding reverse biases of the order of 100 volts without breaking down. It proved easier to handle than silicon, both chemically and metallurgically, and it quickly became the preferred semiconductor for many experimental studies of rectification, photoconductivity and the temperature dependence of conductivity and carrier density. Comprehensive quantitative agreement between theory and experiment still seemed far away, however. The surface of the semiconductor appeared to play an important part in determining the properties of the rectifying barrier found when point contacts were made, and it became a centre of attention. Attempts were made to modulate the conductivity of a thin slice of germanium by a strong external transverse field, but the results were always much less pronounced than expected. Could the surface be screening the interior from the effects of the field? One attempt to clarify the question led Bardeen and Brattain[1] to place the pointed ends of two metal whiskers close to one another on the surface of a piece of germanium. They found that when one whisker, later called the emitter, was biased in the direction of easy flow of current to the germanium (called the base), and the second whisker, later called the collector, was biased in the opposite direction, the current I_C passed by the second depended on the current passed by the first, I_E. Moreover the ratio dI_C/dI_E could exceed unity, provided, amongst other conditions, that the separation of the two points was not more than a few thousandths of an inch. Because the input impedance (emitter-to-base) was low and the output impedance (collector-to-base) was high (even with the interaction), power gains of up to 200 were observed.

A qualitative explanation of the action of the new structure—at first called a transistor, later a type-A transistor and still later a point-contact transistor—was soon given. Three essential operations contribute. There must be (*a*) injection by the emitter, into the base of germanium, of mobile charge carriers of polarity opposite to those (the majority carriers) responsible for most of the conductivity of the base region, which is usually n-type, (*b*) transmission of the injected carriers through the region of the base between emitter and collector, and (*c*) their collection. No fully satisfactory explanations of the physics of injection and, still less, of the process responsible for current gain were given, but they were immediately sought in many laboratories.

Shockley[2] took the next step in postulating the junction transistor and in predicting its performance; injection of minority carriers takes place at one n–p (or p–n) junction into a thin base region of p (or n) material and collection at a p–n (or n–p) junction, the whole unit being contained within a small monocrystalline piece of germanium. The coming of the n–p–n (or p–n–p) transistor marked, perhaps even more so than that of the point-contact transistor, the beginning of a new subject, transistor electronics. This book deals almost entirely with the discoveries, studies, analyses, technological processes and applications made or developed since the initial discoveries by scientists of the Bell Telephone Laboratories.

The subject of transistor electronics is now a very large one and frequent references will have to be made to other books and long review articles which concentrate on particular parts of it. Some of the parts which have yet to make any impact on electrical engineering are completely omitted. Thus the transistor action and other properties described in some detail for germanium and silicon occur in other materials, but few references will be made to those other materials despite the interest which some compounds, particularly those between elements of Groups III and V of the periodic table, are arousing as semiconductors.

Many circuit configurations have been devised using transistors, but relatively few have yet been fully examined or put on an engineering basis. Some are described in detail qualitatively and even quantitatively, but the majority are not. The distinction between those which are and those which are not does not automatically imply preferences. Circuit development initially followed the availability of transistors and the coming of laboratory samples of improved performance rather than of units designed to the users' requirements; we have attempted to describe only the simpler applications designed mainly around units in production.

Comparison with Thermionic Valves

Comparisons between the transistor and the thermionic valve have continued to be drawn since the discovery of the transistor. They can be made between the two structures, the two sets of physical effects responsible for their behaviour, the electrical properties, the elementary circuit applications and the manufacturing processes. Spangenberg[3] has made out a case for emphasizing the similarities of both the physical effects and the electrical properties.

The two structures have little in common, least of all in size and the materials used. The difference in size arises from the different speeds at which the charge carriers responsible for the wanted behaviour move between the electrodes *generating and receiving* them. Electrons attain speeds of the order of 5×10^8 cm/sec in receiving valves, taking perhaps 2 mμsec to cross the distance of about 0·5 cm from cathode to anode. The charge carriers in transistors traverse the uniform base region, of width about 0·003 cm between the emitter and collector, of a typical p-n-p alloy unit, in about 200 mμsec; the base thickness must be reduced to 0·0003 cm before the transit time falls to 2 mμsec. Because the base is so thin and, for many uses, the areas of the emitter and collector often need to be no larger than about 1 mm^2, the total mass of the effective volume of the transistor is very small indeed.

The differences in the physical processes are profound: although emission of charge carriers occurs in both devices, there is little detailed similarity in the two processes of emission from a heated cathode into a vacuum on the one hand and emission across a p–n junction (or the contact between a metal and a semiconductor) on the other. Chapter 3 analyses emission in transistors; except for "punch-through" conditions, the expressions developed bear little qualitative or quantitative resemblance to Richardson's law. The transmission of the charge carriers to the anode of the thermionic valve is controlled by the electric field set up by the voltages applied to the electrodes and by the space-charge of the carriers, and obeys very accurately the laws of motion of charged particles in electric fields. In contrast, electric fields played no part in the transmission of the charge carriers through the base region of the earlier junction transistors, diffusion completely predominating over drift (motion in an electric field). The realization of the need for smaller transit times has led to means for generating a drift field within the base, but not generally by way of the application of terminal voltages. Space-charge effects occur in transistors: the most obvious arises from the fixed charges in the layer depleted of mobile carriers at a junction, particularly under reverse bias, and it is largely independent of any flow of current.

The reception of current by the anode of a valve is, apart from any secondary emission, straightforward; the reception by the collector of a transistor can be more complex, current gain being possible by one process in point-contact transistors and by another in junction units. The temperature-dependence of the processes occurring in the two devices differ. In valves, space-charge effects make the performance almost independent of ambient temperature; but in the transistor two key mechanisms at least are temperature-dependent.

Each device has essentially three terminals (though frequently represented as a four-terminal network, it has one terminal common to both input and output) but the electrical properties differ very much. The input *conductance* of the valve, with the control grid as the input electrode, approaches zero; but it is the input *impedance* of the transistor which is low—very low when the base is the electrode which is earthed and the emitter is the input electrode, and moderately low when the base and emitter are interchanged. Comparisons between the properties of the output of the two elements are more difficult to draw, partly because those of thermionic valves differ as between triodes and pentodes, but there is some resemblance between junction transistors and pentodes. The transistor shows some reverse transmission even at very low frequencies, a property absent in thermionic valves. Both units in fact display properties not well matched in the other. The equivalent circuit of the transistor contains more elements than that of the valve, making circuit analysis more complex, but not so much so as to deter engineers with a sound knowledge of electric circuit theory.

The manufacturing processes used bear little resemblance to one another. Exactness of geometry enters both processes. In the valve it is achieved entirely by mechanical operations, in the point-contact transistor primarily so, but in the junction transistor metallurgy replaces some of the mechanics. The more important materials differ. The coating of the cathode of small receiving valves is a semiconductor, but its impurity content (free barium) may be as high as many parts in ten thousand; its crystalline state is only of secondary importance in deciding the initial performance of the valve. The semiconductor in a transistor must be much purer and monocrystalline. Indeed the perfection of the crystallinity in the transistor is much more appropriately compared with the perfection of the vacuum in a valve.

The greatest similarity between valve and transistor is seen in the performance of equipments using the two elements. The same overall results are possible, but not by the mere replacement of one element by the other. The additional components, the power supplies, and the methods of the circuit designers in utilizing the two active elements, to achieve the same electrical performance from amplifiers, oscillators and switches, often show marked differences.

For those equipments which can be designed around either valves or transistors, there is a firm basis for making engineering and economic comparisons. Already, in the few applications for which field results as well as detailed initial performance are available, the transistor is fulfilling its early promise of having important advantages over the thermionic valve. Savings in power consumption and heat dissipation and in size have always been apparent; in addition, the transistor now promises greater flexibility, longer life, only marginally poorer noise factor and, because its manufacture involves fewer operations, lower price in large-scale production.

The transistor has only just begun its invasion of radio engineering, in which the thermionic valve has long been the all-important active element; the invasion seems assured of substantial success. Some applications, e.g. to radio and radar transmitters, seem likely, however, to remain secure to the thermionic valve for a long time yet.

There has never been much doubt about the best approach to an understanding of the motion of electrons across the vacuum of a thermionic valve. No knowledge of wave mechanics is needed. An understanding of the motion of electrons in the cathode coating and, more particularly, of the emission by the cathode can involve the latter concepts, but most physicists and engineers are content to use a particle model throughout. The theory of the properties of semiconductors and of the motion of electrons in solids has come about only as the result of applications of wave and quantum mechanics. The results of the analyses made, e.g. of the electronic or band structure of solids and of the processes responsible for the scattering and trapping of electrons, can, however, be interpreted in terms of a particle model without demanding very much understanding of the analytical processes used. Moreover, some further deductions can be drawn by using only the particle model. It is not surprising, therefore, that explanations of transistor action and of the accompanying electrical properties of transistors are widely given and understood in terms of the particle model; they do not contain statements or steps which have proved unacceptable to those engineers or physicists who are more analytically minded. We have in general followed that practice. Readers who wish for a fuller understanding of the basic properties of semiconductors are advised to read the review of the band structure by Herman,[4] the review of the electrical properties of germanium and silicon by Harvey Brooks[5] and the later chapters of Shockley's book.[6] There they will find adequate references for further reading.

Terminology and Symbols

Terminology, graphical symbols and letter symbols for semiconductors and transistors have yet to be agreed. We will use what seems most likely to be agreed. Those terms which are clearly peculiar to the subject and whose meanings are not immediately obvious will be defined as they arise. Unfortunately a few terms whose meanings appear obvious can occasionally give rise to ambiguities. Thus "collector," if used alone, can mean one of several things for, say, a grown junction transistor; it therefore becomes necessary, sometimes, to distinguish between the collection junction, the collector region and the collector connexion. The parameters of transistors are dependent on the bias conditions, perhaps more so than those of thermionic valves; no property should be assumed invariant. Although several families of graphical symbols have been proposed for transistors, indications are that a set based on the symbol used early on by the Bell Telephone Laboratories will be preferred.

Because the letter symbols proposed by the Institute of Radio Engineers seem likely to be universally adopted, we have used them —at the price of discarding a few common usages.

References

1. BARDEEN, J. and BRATTAIN, W. H., "The transistor, a semiconductor triode," *Phys. Rev.*, **74**, 230 (1948).
2. SHOCKLEY, W., "The theory of p-n junctions in semiconductors and p-n junction transistors," *Bell Syst. Tech. J.*, **28**, 435 (1949).
3. SPANGENBERG, K. R., *Fundamentals of Electron Devices* (New York, McGraw-Hill, 1957).
4. HERMAN, F., "The electronic band structure of silicon and germanium," *Proc. Inst. Rad. Engrs.*, **43**, 1,703 (1955).
   See also BERNSTEIN, E. and EGLI, P., "Physics of semiconductors" in *Advances in Electronics and Electron Physics*, ed. L. Marton, Vol. **7**, p. 1 (New York, Academic Press, 1955).
5. BROOKS, H., "Theory of electrical properties of germanium and silicon," *Ibid.*, p. 85.
6. SHOCKLEY, W., *Electrons and Holes in Semiconductors* (Princeton, van Nostrand, 1950).

Contents

PAGE

PART 1

THE THEORETICAL BACKGROUND: THE PHYSICS OF SEMICONDUCTORS AND TRANSISTORS