ADVANCED MICROELECTRONICS

M. Reisch

High-Frequency Bipolar Transistors

Springer

Springer Series in
ADVANCED MICROELECTRONICS 11

Springer-Verlag Berlin Heidelberg GmbH

M. Reisch

High-Frequency Bipolar Transistors

Physics, Modeling, Applications

With 330 Figures

Springer

Series Editors:

Dr. Kiyoo Itoh
Hitachi LTD. Central Research Laboratory
1-280 Higashi-Koigakubo
Kokubunji-shi
Tokyo 185-8601
Japan

Professor Takayasu Sakurai
Center for Collaborative Research
University of Tokyo
7-22-1 Roppongi, Minato-ku
Tokyo 106-8558
Japan

Author:

Professor Dr. Michael Reisch
University of Applied Sciences
FH Kempten
Bahnhofstraße 61-63
87435 Kempten/Allgäu
Germany

E-mail: reisch@fh-kempten.de

ISBN 978-3-642-63205-1 ISBN 978-3-642-55900-6 (eBook)
DOI 10.1007/978-3-642-55900-6

Cataloging-in-Publication Data applied for
Bibliographic information published by Die Deutsche Bibliothek.
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data is available in the Internet at <http://dnb.ddb.de>.

http://www.springer.de

Originally published by Springer-Verlag Berlin Heidelberg New York in 2003
Softcover reprint of the hardcover 1st edition 2003

Typesetting: Camera ready by authors
Cover-design: design & production, Heidelberg
Printed on acid-free paper 62 / 3020 hu - 5 4 3 2 1 0 -

Preface

This book provides a rather comprehensive presentation of the physics and modeling of high-frequency bipolar transistors with particular emphasis given to silicon-based devices. I hope it will be found useful by those who do as well as by those who intend to work in the field, as it compiles and extends material presented in numerous publications in a coherent fashion.

I've worked on this project for years and did my best to avoid errors. Despite all efforts it is possible that "something" has been overlooked during copy-editing and proof-reading. If you find a mistake please let me know.

Michael Reisch

Kempten, December 2002

Notation

It is intended here to use the most widely employed notation, in cases where the standard textbook notation is different from the SPICE notation, the latter is used. In order to make formulas more readable, model parameters represented in SPICE by a series of capital letters are written here as one capital letter with the rest in the form of a subscript (e.g. X_{CJC} is used here instead of the `XCJC` used in the SPICE input). Concerning the use of lower-case and capital letters, the following rules are applied:

- Time-dependent large-signal quantities are represented by lower-case letters. The variables i, v and p therefore denote time-dependent current, voltage and power values.
- Stationary large-signal quantities are represented by capital letters. The variables I, V and P therefore denote constant values of current, voltage and power or effective values of the corresponding periodic time-dependent quantities.
- Subscript capital letters are used for large-signal quantities (such as i_{C}, I_{C}, v_{BE} and V_{BE}).
- Subscript lower-case letters are used for small-signal quantities (such as v_{be} and i_{c}).

This convention is also applied to the node and element names in equivalent circuits: node names in large-signal equivalent circuits are represented by capital letters (such as E and S), and node names in small-signal equivalent circuits are represented by lower-case letters (such as e and s). Elements of a large-signal equivalent circuit are denoted by capital letters (such as R), and elements of a small-signal equivalent circuit are denoted by lower-case letters (such as r). The symbol $r_{\mathrm{bb'}}$ therefore denotes the resistance that lies between nodes b and b$'$ of a small-signal equivalent circuit. In some frequently occuring cases we deviate from this convention in order to simplify the notation: for example, instead of $g_{\mathrm{b'e'}}$, the symbol g_π is used to denote the internal small-signal conductance of the emitter–base junction.

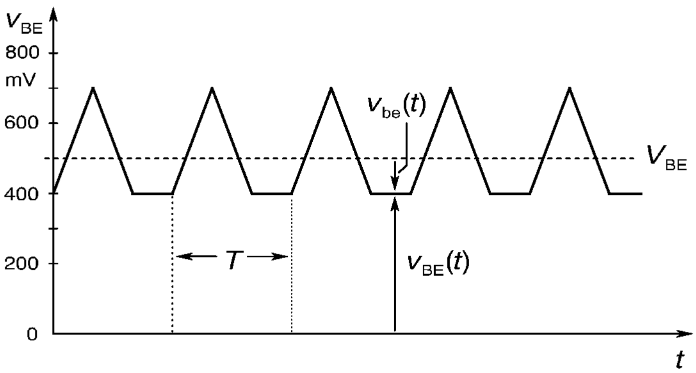
Overlines are used to denote time averages of periodic quantities. A time-dependent voltage $v(t)$ that is periodic with period T is therefore denoted by

$$\overline{v} \;=\; \frac{1}{T}\int_0^T v(t)\,\mathrm{d}t\ .$$

The subscript "eff" is used to denote effective values of periodic time-dependent quantities such as $V_{\mathrm{eff}} = V_{\mathrm{rms}}$, where

$$V_{\text{eff}} = \sqrt{\frac{1}{T}\int_0^T v^2(t)\,\mathrm{d}t} = \sqrt{\overline{v^2}}\,.$$

No subscript is used, if it is clear from the context that a variable denotes an effective value.



As an example, consider a time-dependent base–emitter voltage as depicted in the Figure. The large-signal quantity v_{BE} takes values in the interval between 400 mV and 700 mV as indicated in Table 1.

Table 1. Values of $v_{\mathrm{BE}}(t)$

Time t	$0 < t \leq T/3$	$T/3 < t \leq 2T/3$	$2T/3 < t \leq T$
$v_{\mathrm{BE}}/\mathrm{mV}$	$400 + 900\,\dfrac{t}{T}$	$400 + 900\left(\dfrac{2}{3} - \dfrac{t}{T}\right)$	400
$v_{\mathrm{be}}/\mathrm{mV}$	$-100 + 900\,\dfrac{t}{T}$	$-100 + 900\left(\dfrac{2}{3} - \dfrac{t}{T}\right)$	-100

The time average of the voltage is

$$\begin{aligned}\frac{\overline{v_{\mathrm{BE}}}}{\mathrm{mV}} &= \frac{1}{T}\left[400\int_0^T \mathrm{d}t + 900\int_0^{T/3}\frac{t}{T}\,\mathrm{d}t + 900\int_{T/3}^{2T/3}\left(\frac{2}{3}-\frac{t}{T}\right)\mathrm{d}t\right]\\ &= \frac{1}{T}\left(400\,T + 300\,\frac{T}{3}\right) = 500\,.\end{aligned}$$

and represents the dc component. The ac part $v_{\mathrm{be}}(t)$ is obtained by subtraction of the dc component of the base–emitter voltage $V_{\mathrm{BE}} = \overline{v}_{\mathrm{BE}}$ from its time-dependent large-signal value $v_{\mathrm{BE}}(t)$:

$$v_{\rm BE}(t) \;=\; V_{\rm BE} + v_{\rm be}(t)\;.$$

The ac part takes values in the interval between −100 mV and 200 mV and has a time average of zero. Since $\overline{v_{\rm be}} = 0$, one obtains

$$\overline{v_{\rm BE}^2} \;=\; \overline{(V_{\rm BE} + v_{\rm bc})^2} \;=\; \overline{V_{\rm BE}^2 + 2V_{\rm BE}v_{\rm bc} + v_{\rm be}^2} \;=\; V_{\rm BE}^2 + \overline{v_{\rm be}^2}\;,$$

i.e. the effective value $V_{\rm BE,eff}$ of $v_{\rm BE}$ is given by

$$V_{\rm BE,eff} \;=\; \sqrt{V_{\rm BE}^2 + \overline{v_{\rm bc}^2}} \;=\; \sqrt{V_{\rm BE}^2 + V_{\rm bc}^2}\;,$$

where $V_{\rm be} = \sqrt{\overline{v_{\rm be}^2}}$ denotes the effective value of the ac part. The value of $V_{\rm be}$ is determined from

$$\begin{aligned}\frac{V_{\rm be}^2}{\rm mV^2} \;&=\; \frac{1}{T}\int_0^{T/3}\left(900\,\frac{t}{T} - 100\right)^2 {\rm d}t + \frac{1}{T}\int_{T/3}^{2T/3}\left(1000 - 900\,\frac{t}{T}\right)^2 {\rm d}t \\ &\quad + \frac{1}{T}\int_{2T/3}^{T}(100)^2\,{\rm d}t \;=\; 10^4\end{aligned}$$

as 100 mV. The effective value of the large-signal base–emitter voltage is therefore larger than its dc component:

$$V_{\rm BE,eff} \;=\; \sqrt{V_{\rm BE}^2 + V_{\rm bc}^2} \;=\; \sqrt{(500\,{\rm mV})^2 + (100\,{\rm mV})^2} \;=\; 509.9\,{\rm mV}\;.$$

Complex Notation

Harmonic time-dependent quantities are represented in complex notation as follows:[1]

$$v(t) \;=\; {\rm Re}[\,\underline{v}(t)\,] \;=\; {\rm Re}\left(\underline{\hat{v}}\,{\rm e}^{{\rm j}\omega t}\right)\;,$$

where the complex amplitude $\underline{\hat{v}}$ has a magnitude $\hat{v} = |\underline{\hat{v}}|$. Complex time-dependent quantities are represented by underlines, the symbols $\underline{v}$ and $\underline{i}$ therefore correspond to

$$\underline{v} \;=\; \underline{\hat{v}}\,{\rm e}^{{\rm j}\omega t} \quad\text{and}\quad \underline{i} \;=\; \underline{\hat{i}}\,{\rm e}^{{\rm j}\omega t}\;.$$

The amplitude $\hat{v}$ is related to the effective value V by

$$V \;=\; \hat{v}/\sqrt{2}\,.$$

The use of complex effective quantities such as

$$\underline{V} \;=\; \underline{\hat{v}}/\sqrt{2}$$

[1] Underlines denote generally complex quantities. To simplify notation, however, not all complex quantities are underlined: impedances (symbol z, Z), admittances (symbol y, Y) or hybrid parameters (symbol h), for example, are understood as complex quantities without explicit underlining.

is convenient for power and noise calculations. If the instantaneous current is

$$i(t) \;=\; \mathrm{Re}[\,\underline{i}(t)\,] \;=\; \frac{1}{\sqrt{2}}\left(\underline{I}\,\mathrm{e}^{\mathrm{j}\omega t} + \underline{I}^{*}\,\mathrm{e}^{-\mathrm{j}\omega t}\right)$$

and the instantaneous voltage is

$$v(t) \;=\; \mathrm{Re}[\,\underline{v}(t)\,] \;=\; \frac{1}{\sqrt{2}}\left(\underline{V}\,\mathrm{e}^{\mathrm{j}\omega t} + \underline{V}^{*}\,\mathrm{e}^{-\mathrm{j}\omega t}\right) \;,$$

the instantaneous power is

$$\begin{aligned} p(t) \;&=\; v(t)i(t) \;=\; \mathrm{Re}\,(\underline{I}\,\underline{V}^{*}\,) + \mathrm{Re}\left(\underline{I}\,\underline{V}\,\mathrm{e}^{\mathrm{j}2\omega t}\right) \\ &=\; IV\cos(\varphi_i - \varphi_v) + IV\cos(2\omega t + \varphi_i + \varphi_v) \end{aligned}$$

while the average power is

$$P \;=\; \overline{p(t)} \;=\; \mathrm{Re}\,(\underline{I}\,\underline{V}^{*}) \;=\; IV\cos(\varphi_i - \varphi_v) \;,$$

if the complex effective values $\underline{V}$ and $\underline{I}$ are represented as $\underline{V} = V\,\mathrm{e}^{\mathrm{j}\varphi_v}$ and $\underline{I} = I\,\mathrm{e}^{\mathrm{j}\varphi_i}$. Transfer factors, impedances, admittances, etc. are defined in terms of averages and are independent of the notation used as is illustrated by the following example:

$$Z \;=\; \frac{\underline{V}}{\underline{I}} \;=\; \frac{\hat{\underline{v}}}{\hat{\underline{i}}} \;=\; \frac{\underline{v}}{\underline{i}} \;.$$

A parallel connection of several impedances Z_i can be represented by a single impedance with a value given by

$$Z = Z_1 \parallel Z_2 \parallel \cdots \parallel Z_n \;=\; \left(\frac{1}{Z_1} + \frac{1}{Z_2} + \cdots + \frac{1}{Z_n}\right)^{-1} \;.$$

The symbol "$\|$" denotes an operator that causes the addition of the reciprocals of the specified impedances, with subsequent reciprocation. The operation is performed before addition or subtraction, but after multiplication or division.

Vectors

In rectangular coordinates, we may define vectors in terms of their components, e.g.

$$\boldsymbol{a} \;=\; a_x\boldsymbol{e}_x + a_y\boldsymbol{e}_y + a_z\boldsymbol{e}_z \;,$$

using the unit vectors $\boldsymbol{e}_x$, $\boldsymbol{e}_y$ and $\boldsymbol{e}_z$, which specify what we mean by the x direction, y direction and z direction.

Table of Contents

Part I

INTRODUCTION

1 An Introductory Survey

This chapter briefly reviews the fundamentals of circuit theory, semiconductors, pn junctions, bipolar transistor physics and compact models for the analysis of large-signal, small-signal and noise behavior. Before going into details, it introduces the reader to the field at an elementary level. It should make the material presented in part II accessible to graduate students.

1.1 History

The invention of the transistor in 1947 – more than half a century ago – was an important milestone in the development of industrial civilization. Since that time the combined effort of thousands of physicists, chemists and engineers has allowed the basic principle of the transistor to be developed to the present state. Although the aim of this book is a description of the state of the art, it seems worthwhile to spend a few moments to look back on this fascinating development.

1.1.1 Early Developments

There was some work on semiconductors before the invention of the transistor, and around 1880 the unique features of semiconductors (photoconductivity, photovoltaic effect, negative temperature coefficient of specific resistance, nonlinear rectifying properties of semiconductor contacts) were known – the term semiconductor, however, was not coined until 1911. A deeper understanding of the behavior of semiconducting materials had to wait for the development of quantum theory and its application to electrons in solids in the early 1930s. Around 1945 semiconductors were already recognized as materials with a bandgap that could show both n-type or p-type conductivity, dependent on the nature of the impurities incorporated in the lattice; a theoretical investigation of the rectifying properties that "considered most of the essential features of modern low-level rectification theory and specifically pointed out the importance of minority carriers in determining the rectifying action" [1] was published by Davydov in 1938.

Before the invention of the transistor, two devices – the vacuum tube and the relay – were known as electrically controlled switches. While the first suffered from a significant standby power and limited lifetime, the second was slow, with switching times in excess of a millisecond. In the 1920s and 1930s it was already being speculated that solid-state electronic devices could be used as amplifiers and replace the bulky, voltage- and power-hungry vacuum tubes in use then.[1] Later in 1945, a research group headed by W. Shockley was established at Bell Telephone Laboratories to investigate semiconductors with the goal to develop a device that could replace the vacuum tube and the relay for amplifying and switching purposes.

1.1.2 The First Transistors

The first transistor[2] was realized by Brattain and Bardeen [3] by placing two closely adjacent electrodes[3] of the point-contact rectifier type on an n-type block of germanium. The schematic presented in the first publication is redrawn in Fig. 1.1 with slight modifications and shows what we would call today a small-signal amplifier with a pnp transistor in common-base configuration: The signal $v_{\mathrm{eb}}(t)$ is superimposed on the bias voltage V_{EB} and modu-

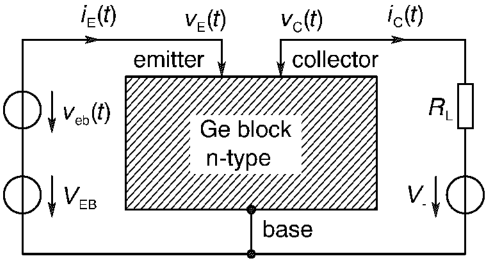


Fig. 1.1. Schematic of semiconductor triode (after [3])

lates the current $i_{\mathrm{E}}(t)$ that flows into the emitter and reaches the collector to a certain extent. From this cross section the naming of the base region becomes obvious. With such devices power gains of about 100 (20 dB) at an output power of 25 mW and frequency up to 10 MHz were achieved [4]. Due to the lateral spacing of the (small-size) contacts, these so-called point-contact transistors suffered from inherent limitations in power-handling capacity and operating frequency. Furthermore, their behavior was largely governed by surface

[1]The concept foreseen at that time was what is known today as the field-effect transistor, and numerous experiments were performed to observe the expected field effect. As explained by Bardeen, they all failed because of the large number of charged surface states that shield the semiconductor body from the effects of the external field.

[2]The naming of the transistor [2] is explained by the fact, that at the time the device was considered to be a "dual to the vacuum diode in that input was a current and the output was a voltage (in a vacuum tube the input is a voltage and the output is a current)" – if "a vacuum tube has transconductance its dual should have transresistance"; a slight shortening of the corresponding "transresistor" resulted in "transistor". This assumption, which emphasizes transistors as current-controlled devices, is not true for state-of-the-art transistors with a base current that may be neglected in many applications; such a device acts instead like a nonlinear voltage-controlled current source, i.e., as a pure transconductance element.

[3]For example gold wires with a thickness of approx. 50 µm forming gold contacts with a separation of approx. 50 – 250 µm; passing short current pulses (forming) through the contacts improved the electrical characteristics (probably due to alloying).

effects, resulting in poor reproducibility of electrical characteristics, electrical instability and high noise levels that were in addition affected by changes in temperature and humidity. Only small numbers of the point-contact transistor were fabricated in the 1950s.

Shockley soon recognized the important role of minority carriers for the operation of the device and towards the end of January 1948 formulated his theory of pn junctions [5–7]. Based on his analysis of minority-carrier injection at forward-biased junctions and collection at reverse biased junctions, he invented the junction transistor, which uses two pn junctions in close proximity: minority carriers injected at a forward-biased junction can diffuse through the intermediate layer and be collected by the second junction, which is operated in reverse bias. The current generated in the low-impedance emitter circuit therefore causes a current to flow in the high-impedance collector circuit, resulting in power gain [8]. The potentials of the new device were almost immediately recognized.[4]

After another two years spent learning how to grow single crystals of high-purity germanium by slowly pulling a seed crystal from the melt (Czochralski method), how to purify materials using the process of zone refining, and how to control the doping during crystal growth, in April 1950 a germanium crystal was grown with a thin p-type region sandwiched between two n-type regions [8]. The electrical characteristics of the npn transistor realized in this grown-junction technique were pretty stable and largely consistent with Shockley's theory; in addition to this, the noise level was considerably lower than in point-contact transistors. Nevertheless, still considerable leakage currents were observed; furthermore, the realization of thin base layers and finding their correct location to realize a base contact was a severe problem with this technique. A manufacturable and reliable device still had to be developed.

A considerable push in transistor development came from military programs. From the early 1950s the interest in low-weight, small-size and highly reliable electronic devices, which arose from the implementation of aircraft and rocket control systems with increasing complexity, helped to fund the development of the new technology at a rapid pace. Various approaches, such as the alloy transistor, formed by alloying for example indium to n-type germanium, were investigated, but the reproducible production of high-frequency

[4] Becker and Shive introduced their classic paper [9], which appeared in March of 1949, with the words: "A transistor differs advantageously from a vacuum tube in several important respects. It has no vacuum. It has no filament; consequently it consumes no filament power and requires no warm-up time. It is both smaller and lighter than any commercially available vacuum tube. Within the present limitations of their power handling capacity, noise, and frequency response, transistors can perform many of the tasks now performed by vacuum tube triodes. They have been successfully demonstrated in radio-frequency, intermediate-frequency, and audio-frequency amplifiers, oscillators, mixers, and pulse generators."

bipolar transistors was not possible with these techniques. This was achieved for the first time with the development of the impurity diffusion technique, which allowed for good control of base width and doping concentrations. The first germanium-diffused transistor was made by Lee using the mesa technique and showed a cutoff frequency of 500 MHz [8]. Furthermore, the diffusion process allowed for batch production with a substantial reduction in production costs.

1.1.3 Silicon Transistors

The first transistors were realized with germanium, a semiconductor with a bandgap ($W_\mathrm{g} \approx 0.67$ eV) considerably smaller than the bandgap of silicon ($W_\mathrm{g} \approx 1.12$ eV). The small bandgap of germanium results in a small voltage drop across germanium pn junctions under forward bias and therefore in reduced power dissipation. Another advantage of germanium with respect to silicon is its larger minority-carrier mobility, resulting in smaller transit time and larger speed of operation. Since the intrinsic carrier density, n_i, is much larger in germanium than in silicon, germanium junctions suffer from a substantial reverse current that increases rapidly with temperature. The reverse current of a comparable silicon pn junction is orders of magnitude smaller; this is a significant advantage of silicon bipolar transistors, which may be employed in switching applications over a much wider temperature range. Further advantages of silicon are the processing flexibility that results from dopants with different diffusion coefficients[5] and, in particular, the fact that an excellent insulator (SiO_2) is obtained by simple oxidation of silicon – of particular importance in this respect is the low density (down to $10^{10}\,\mathrm{cm}^{-2}$) of surface states at the $Si{-}SiO_2$ interface.

The advantages of silicon were quickly recognized. However, the problem with silicon was, that the chemical and metallurgical process of device formation required much higher temperatures[6] as in germanium. Nonetheless, in 1954, after application of the Czochralski method to the growth of single-crystal silicon, and the floating zone method to purify the resulting contaminated silicon crystal, the first manufacturable silicon transistor was realized with grown-junction technology at Texas Instruments.

The observation by Frosch and Derick [8] that a thin layer of silicon dioxide is able to mask the diffusion of certain donors and acceptors into the underlying silicon bulk material and the development of photolithography based on early observations of Andrus and Bond, paved the way to the invention of the planar process by Hoerni of the Fairchild Corporation in 1959. In this process pn junctions are formed by diffusion of dopants through an opening in a layer

[5]While in Ge all donors diffuse much faster than acceptors, in Si both slow- (arsenic, antimony) and fast-diffusing (boron, phosphorus) donors and acceptors are available.

[6]The melting point of Ge (960 °C) lies well below the melting point of silicon (1430 °C).

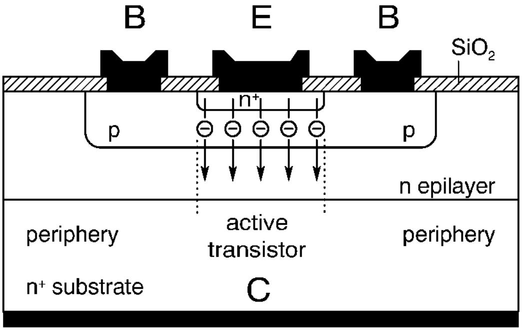


Fig. 1.2. Cross section (schematic) of a vertical npn bipolar transistor fabricated with planar technology

of thermal oxide, which masks the rest of the semiconductor. The planar process provided transistors with unprecedented reliability and allowed for the development of integrated circuits. An essential improvement of the planar process was the termination of the junction with thermally grown silicon dioxide, which, due to its low defect density, avoided the nonideal characteristics and reverse-bias leakage currents of earlier approaches. Planar devices showed stable characteristics and were much more reliable than previously produced mesa transistors. The realization of transistors with a low-doped collector region, proposed by Early in 1954 [10], became possible with the development of epitaxial growth techniques. In 1960 the first epitaxial layer was grown on a silicon substrate [11], and the first transistor with a low-doped epitaxial region on a heavily doped substrate and a cross section as depicted in Fig. 1.2 followed soon.

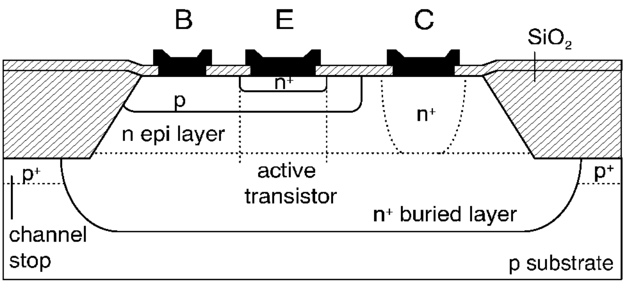


Fig. 1.3. Cross section (schematic) of vertical npn transistor as employed for the application in integrated circuits

1.1.4 Integrated Bipolar Transistors

The concept of integrated circuits , in discussion already in the early 1950s [4] and first realized by Kilby in 1958, became commercially viable with the first silicon planar integrated circuits realized by R. Noyce of Fairchild. In March 1961 Fairchild produced the first integrated circuits, which provided transistors, diodes, resistors and capacitors on one chip, isolated by reverse-biased pn junctions. During the next two decades the planar process evolved with the introduction of epitaxial collector regions above heavily doped buried layers, LOCOS isolation as a substitute for the pure pn junction isolation, and implantation of impurities for the realization of narrow base regions (mandatory for large cutoff frequencies), to mention just the most important developments. Reliable integrated high-frequency bipolar transistors with a cross section as depicted in Fig. 1.3 have been fabricated since the middle of the 1970s [12].

Continuous effort was spent on the reduction of vertical device dimensions aiming towards increased cutoff frequencies. Arsenic-doped emitters with shallow boron-doped bases realized by implantation allowed for the fabrication of thin emitter and base layers. Metal contacts , however, were quickly recognized to increase the base current when the emitter depth decreased, and thus to spoil the current gain. It was therefore an important observation that polysilicon contacts to monocrystalline emitter regions behave differently in this respect and yield current gains that are a factor of four to five larger than those of comparable metal-contacted emitters [13]. Such emitters allowed for the vertical scaling of the device. Further progress came with reduced lithog-

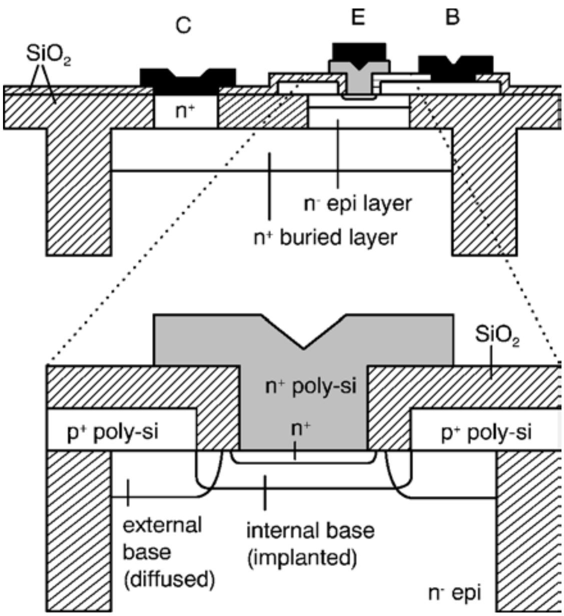


Fig. 1.4. Integrated high-frequency bipolar transistor with a self-aligned emitter-base configuration in double-poly technology and u-groove isolation (after [14])

raphy, but due to the large parasitics of the conventional transistor cross section, performance gain was limited. An important innovation that allowed for a significant reduction of the lateral device dimensions came with the concept of self-alignment, which was made possible by plasma-etching processes.

Plasma etching was first investigated as a possible alternative to wet-etching processes at the end of the 1960s. Significant advances during the 1970s and the observation that plasma-etching processes are anisotropic with a vertical etching rate that considerably exceeds the lateral etching rate made plasma etching a key process technology for the realization of devices in the micron and submicron regime. With this technique, a novel cross section, as depicted in Fig. 1.4, came into reach. Transistors of this type were first presented at the beginning of the 1980s [14–16] and represent the state of the art of silicon high-frequency bipolar transistors in production.

Table 1.1 shows the development of typical data used for the characterization of bipolar technologies. While junction capacitances are determined by the lateral device dimensions and thus by lithography and layout, parameters such as the forward transit time, the cutoff frequency or the collector–emitter breakdown voltage BV_{CEO} are essentially determined by the vertical doping profile and thus by physical and chemical processes such as implantation, diffusion or epitaxial growth. The reductions of the collector-substrate (cs) capacitance were made possible by novel isolation techniques, which replaced the early pn junction isolation by LOCOS and then by trench isolation.

Table 1.1. Development of characteristic bipolar parameters (typical data for IC)

Year of introduction	1979	1983	1986	1992	1999
Emitter stripe width [µm]	3	2.5	1.4	0.5	0.3
Base width [nm]	250	200	150	100	40
Epi thickness [µm]	2	1.5	1.2	0.8	0.4
BC capacitance [fF]	120	80	35	20	8
CS capacitance [fF]	370	200	60	8	5
CE breakdown, BV_{CEO} [V]	12	10	8	5	2.5
Cutoff frequency [GHz]	3	5	10	40	100
Forward transit time τ_{f}, [ps]	35	25	20	10	2.5

1979: Junction isolation, 1983: LOCOS isolation, 1986: self-aligned polysilicon emitter, 1992: trench isolation, scaled polysilicon emitter, 1999: SiGe heterostructure BJTs

1.1.5 Heterojunction Bipolar Transistors

Parallel to the described developments, progress in epitaxial deposition techniques made it possible to form thin base layers of strained SiGe on silicon substrates, opening the world of heterojunction bipolar transistors to silicon IC technology. The principle of heterojunction bipolar transistors has been known since the 1950s [17, 18]: Wide-bandgap emitters were recognized as a means to reduce minority-carrier injection into the emitter region, resulting in increased current gain and cutoff frequency. The realization of heterojunctions in silicon technology, however, failed since the formation of silicon alloys always results in a significant change of the lattice constant, an effect that prohibited the formation of single-crystal heterostructures. Due to this reason, heterojunction bipolar transistors were the exclusive domain of compound semiconductor technology, since the substitution of atoms in GaAs or InP with homologous ones generally results in a modification of the bandgap without affecting the lattice constant very much. Novel epitaxial techniques, however, made it possible to grow very thin layers of SiGe on silicon crystals, with the same lattice constant. This is not possible without strain, and it is possible to grow only thin layers without the formation of dislocations, which would relax the strain due to the different lattice constant [19–21]. Growing SiGe layers with a Ge content that varies continuously with position allows the development of transistors with a graded base, i.e. base regions in which the bandgap continuously decreases from the emitter to the collector. This acts like an inherent electric field and reduces the base transit time, resulting in increased cutoff frequency and a larger $B_{\mathrm{N}}V_{\mathrm{AF}}$-product[7] [22,23].

1.1.6 CAD, Device Modeling

Device theory had a strong impact on the development of the transistor. As a matter of fact the junction transistor was invented on theoretical considerations, and it is interesting to note that cutoff frequencies of 40 GHz were predicted as soon as 1958 [24] for devices with a base width of 150 nm and a base charge of $10^{13}\,\mathrm{cm}^{-3}$. This is roughly what has been achieved more than three decades later with self-aligned double-polysilicon devices.

With increasing complexity of bipolar processes, device modeling with the goal to reliably predict the behavior of integrated circuits by means of computer simulations became of vital importance. Before the work of Gummel [25], approaches to bipolar device modelling were based on the regional approach, which divides the transistor volume into quasi-neutral regions and space charge layers. Gummels iterative procedure allowed a numerical solution of the basic semiconductor equations without such assumptions. However, a reliable prediction of important device parameters, such as the value and temperature dependence of the current gain, required an improved understanding and description of heavy doping effects, recombination mechanisms and material parameters such as minority-carrier mobilities.

The Ebers–Moll model published in 1954 [26,27] provided a compact model for the description of the large-signal transient response of junction transistors. A compact model that was more appropriate for the description of the bipolar transistor and that considered high-level injection effects was developed in the late 1960s by Gummel and Poon [28] and has served as the workhorse in bipolar IC development for now more than two decades. Recent modeling approaches, such as HICUM or MEXTRAM, essentially are ex-

[7]The product of Early voltage and current gain is an important figure of merit (see Sect. 3.10) in analog circuit design and determines the output resistance of current mirrors.

tensions of the ideas underlying the Gummel–Poon model, aiming towards an improved description of high-level injection effects, temperature and self-heating effects, influence of the substrate, etc.

1.1.7 Applications

Bipolar transistors were almost immediately employed for the realization of amplifiers, oscillators, etc. The development of the first silicon integrated circuits then allowed for the realization of digital circuits. Various circuit techniques such as RTL, TTL and I^2L were invented for the implementation of the logic functions, with TTL as the work horse of digital electronics in the 1970s and early 1980s [29]. However, all these techniques suffer from saturation of the switching transistors, resulting in a substantial slow down of the switching speed. Progress in the development of MOSFETs and in particular the invention of the CMOS process made these techniques obsolete. Only the ECL (emitter-coupled logic) technique, which avoids saturation of the switching transistors, is still of some interest owing to its outstanding switching speed. Figure 1.5 shows the decrease of ECL gate delay over the years that has resulted in gate delays below 10 ps nowadays. Because of substantial static power consumption, ECL is only employed if unavoidable. Recently, owing to the advent of wireless communication techniques, integrated bipolar circuit techniques found wide-spread use in RF applications.

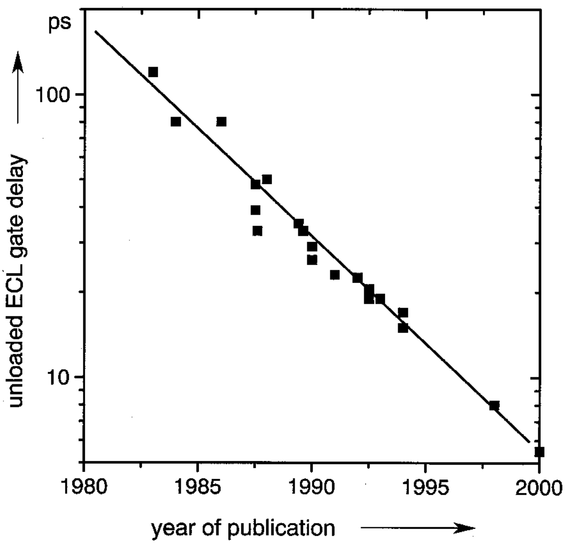


Fig. 1.5. Reported ECL gate delays versus year of publication [30–32]

1.2 Devices, Circuits, Compact Models

The efficient development of integrated circuits requires means for circuit analysis and simulation for design verification.

The most important tools for circuit analysis are Kirchhoff's laws, which apply to circuits of lumped elements, i.e., elements with a spatial extension, d, that is small in comparison with the wavelength, λ, of relevant signals [33,34]. If the signals propagate with velocity c, the requirement $d \ll \lambda$ defines a critical frequency,

$$f_{\mathrm{C}} = c/d \;;$$

if f_{C} is large in comparison with the frequency at which the circuit operates, each conducting element that connects two elements in the circuit may be considered as a node. To each node we may attribute a name, α, and a node potential, v_α (or $v(\alpha)$).

According to Kirchhoff's current law (KCL), the values of all currents that flow into a node add up to zero, while according to Kirchhoff's voltage law (KVL), all voltage drops around a closed loop add up to zero. Application of these laws to a given circuit, such as the amplifier shown in Fig. 1.6, provides a set of equations that can be used for the calculation of circuit behavior. In addition to this, so-called constitutive relations, which relate the current across a device to the voltage drop, are needed. Circuit elements with well-defined constitutive relations are used to develop equivalent circuits that represent the behavior of physical devices.

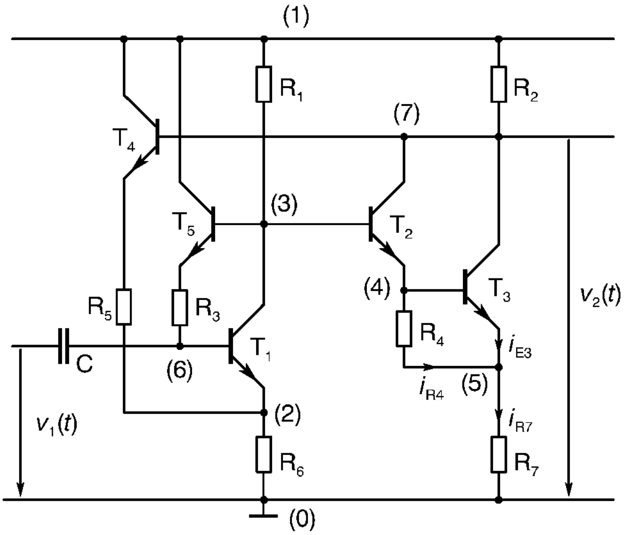


Fig. 1.6. Circuit plan of an electronic amplifier

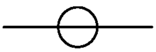
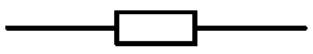
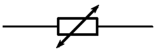
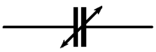
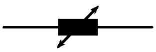
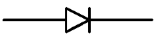
Symbol	Meaning	Constitutive relation
	independent voltage source	$v = v(t)$
	independent current source	$i = i(t)$
	linear resistance	$v(t) = R\, i(t)$
	linear capacitance	$i(t) = C\, \mathrm{d}v/\mathrm{d}t$
	linear inductance	$v(t) = L\, \mathrm{d}i/\mathrm{d}t$
	nonlinear resistance	$v(t) = v[i(t)]$
	nonlinear capacitance	$i(t) = \mathrm{d}q/\mathrm{d}t\, ;\ q(t) = q[v(t)]$
	nonlinear inductance	$v(t) = \mathrm{d}\phi/\mathrm{d}t\, ;\ \phi(t) = \phi[i(t)]$
	ideal diode	$i(t) = I_S\, [\, \exp(v/V_T)\text{-}1\,]$

Fig. 1.7. Circuit elements used in compact modeling

1.2.1 Circuit Elements

Circuit elements are the components of equivalent circuits[8] or compact models, which are used to describe the electrical behavior of real electronic devices. Figure 1.7 shows the elements required for our purpose, together with the symbols employed for their description.

A resistor is a two-terminal circuit element in which the branch voltage $v(t) = v[i(t)]$ is determined by the branch current $i(t)$ at time t. If this function is linear, $v(t) = R\,i(t)$, the resistor is called ohmic or linear.

A capacitor is a two-terminal network element with a stored charge, $q(t) = q[v(t)]$, determined by the branch voltage $v(t)$ at the same time, and a branch current, $i(t)$, that equals the time rate of change of $q(t)$,

$$i(t) = \mathrm{d}q/\mathrm{d}t\,. \tag{1.1}$$

The capacitor is called linear if $q(t) = C\,v(t)$ varies proportionally to $v(t)$. For linear capacitors (1.1) becomes

$$i(t) = C\,\mathrm{d}v/\mathrm{d}t\,. \tag{1.2}$$

[8]An equivalent circuit is understood as a graphical representation of a compact model; the two terms are used synonymously in this text.

A two-terminal element is called an inductor if its flux $\phi(t) = \phi[i(t)]$ at time t is a function of current $i(t)$ at the same instant. The inductor is called linear if $\phi(t) = L\,i(t)$ is proportional to $i(t)$. The voltage $v(t)$ across the inductor is given by Faraday's law ,

$$v(t) \;=\; \mathrm{d}\phi/\mathrm{d}t\,, \tag{1.3}$$

which in the case of a linear inductor simplifies to

$$v(t) \;=\; L\,\mathrm{d}i/\mathrm{d}t\,. \tag{1.4}$$

An independent voltage source is a two-terminal element with a pre-defined, possibly time-dependent, branch voltage that is independent of the branch current. An independent current source is a two-terminal element with a pre-defined, possibly time-dependent, branch current that is independent of the branch voltage. Controlled sources are current or voltage sources that deliver a current or voltage which is determined by another current or voltage. A source controlled by another current or voltage should formally be represented as a two-port, with an input port for the controlling quantity and an output port for the controlled quantity. In order to simplify equivalent circuit drawings, we will employ a simplified representation, as depicted in Fig. 1.8b.

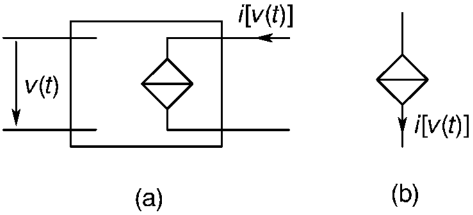


Fig. 1.8. (**a**) Full representation of a voltage-controlled current source as a two-port element, (**b**) simplified representation

1.2.2 Nonlinear Network Elements, Small-Signal Description

If the current across a linear resistor is plotted versus the applied voltage, a straight line with a slope given by the conductance of the resistor results (Fig. 1.9). The current–voltage characteristic of a nonlinear resistor is more complicated. In small-signal operation, i.e., if only the small deviations $v_\mathrm{S}(t) = v(t) - V_0$ and $i_\mathrm{S}(t) = i(t) - I_0$ from a given bias point (V_0, I_0) are considered, a linear approximation of the current–voltage characteristic may be employed, replacing the exact current–voltage characteristic with a first–order Taylor series expansion. This results in

$$i_\mathrm{s}(t) \;=\; i\,[\,V_0{+}v_\mathrm{s}(t)] - I_0 \;\approx\; g(V_0)v_\mathrm{s}(t)\;, \tag{1.5}$$

where

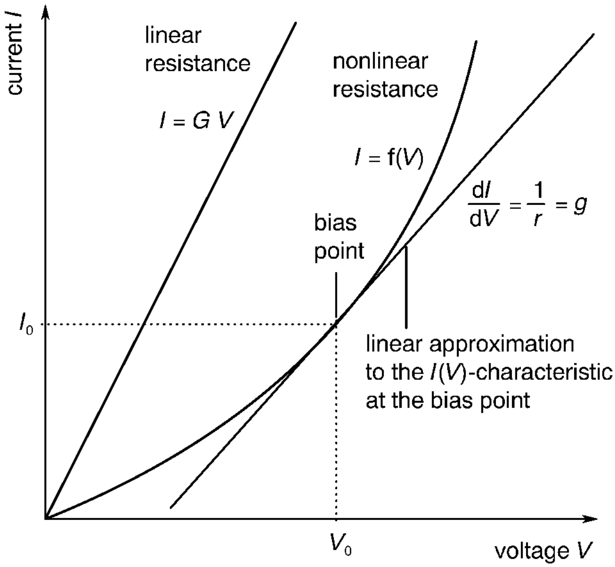


Fig. 1.9. Current-voltage characteristic of a linear and a nonlinear resistor

$$g(V_0) = \left.\frac{\mathrm{d}I}{\mathrm{d}V}\right|_{V_0} = \frac{1}{r(V_0)} \tag{1.6}$$

is the small-signal conductance of the nonlinear resistor at the bias point considered; its inverse $r(V_0)$ is the small-signal resistance. Thus the "small-signal current" $i_\mathrm{s}(t)$ is related to the "small-signal voltage" $v_\mathrm{s}(t)$ as in a linear resistor with conductance $g = g(V_0)$. In a small-signal equivalent circuit, the nonlinear resistor therefore is replaced by a linear resistor with conductance $g(V_0)$; if the small-signal voltage $v_\mathrm{s}(t)$ is applied to this circuit, the small-signal current $i_\mathrm{s}(t)$ results (Fig. 1.10b). As $g(V_0)$ corresponds to the slope of the nonlinear current–voltage characteristic, its value depends on the bias point chosen. For some purposes, such as the analysis of intermodulation

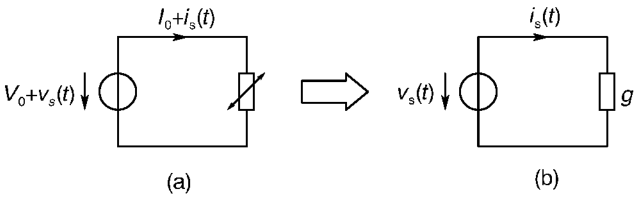


Fig. 1.10. (**a**) Large–signal and (**b**) small–signal description of a nonlinear resistor

distortions, higher-order terms in the Taylor series expansion of the current–voltage characteristic are important. Considering terms up to the third order of the Taylor series expansion yields, e.g.,

$$\begin{aligned} i(t) &\approx I(V_0) + \left.\frac{\mathrm{d}I}{\mathrm{d}V}\right|_{V_0} v_\mathrm{s}(t) + \frac{1}{2}\left.\frac{\mathrm{d}^2 I}{\mathrm{d}V^2}\right|_{V_0} v_\mathrm{s}^2(t) + \frac{1}{6}\left.\frac{\mathrm{d}^3 I}{\mathrm{d}V^3}\right|_{V_0} v_\mathrm{s}^3(t) \\ &= I(V_0) + g\, v_\mathrm{s}(t) + g_2\, v_\mathrm{s}^2(t) + g_3\, v_\mathrm{s}^3(t)\ . \end{aligned} \tag{1.7}$$

In a small-signal representation that considers the third-order nonlinear response, this current–voltage characteristic is represented by a linear resistor with conductance g in parallel with two voltage-controlled polynomial sources, as shown in Fig. 1.11.

In (1.7) the current is expressed as a power series of the applied voltage. An equivalent representation of the nonlinear current–voltage characteristic describes the small-signal voltage drop $v_\mathrm{s}(t)$ across the resistor as a function of the small-signal current $i_\mathrm{s}(t) = i(t) - I_0$ through the device,

$$v_\mathrm{s}(t) = r\, i_\mathrm{s}(t) + r_2\, i_\mathrm{s}^2(t) + r_3\, i_\mathrm{s}^3(t) + \cdots \tag{1.8}$$

If the coefficients g_α are known, it is possible to compute the r_β by series inversion (cf. [35], formula 3.6.25) with the result

$$r = 1/g\ , \quad r_2 = -g_2/g^3\ , \quad r_3 = (2g_2^2 - g g_3)/g^5 \tag{1.9}$$

for the first three terms.

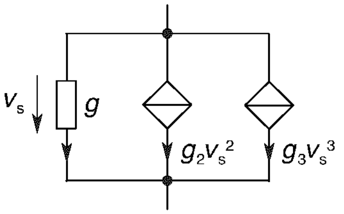


Fig. 1.11. Small-signal equivalent circuit of a nonlinear resistor considering the third-order signal response

Example 1.2.1 As an important example we consider the nonlinear current–voltage characteristic of a diode given by

$$I = I_\mathrm{S}\left[\exp\left(\frac{V}{NV_\mathrm{T}}\right) - 1\right]\ ,$$

where I_S, N, and V_T are parameters that do not depend on the applied voltage V. The small-signal conductance at the bias point V_0 follows through differentiation,

$$g = \frac{I_\mathrm{S}}{NV_\mathrm{T}} \exp\left(\frac{V_0}{NV_\mathrm{T}}\right) = \frac{I(V_0) + I_\mathrm{S}}{NV_\mathrm{T}}\ .$$

In forward bias with $I_0 = I(V_0) \gg I_\mathrm{S}$, we obtain, in particular,

$$g = I_0/NV_\mathrm{T}\ ,$$

in other words, the small-signal conductance increases in proportion to the bias current. The second- and third-order terms follow by subsequent differentiation, with the result

$$g_2 = \frac{g}{2NV_{\mathrm{T}}} \quad \text{and} \quad g_3 = \frac{g}{6(NV_{\mathrm{T}})^2} .$$

The nonlinear terms $g_2 v_{\mathrm{S}}(t)$ and $g_3 v_{\mathrm{S}}^2(t)$ will therefore only be negligible in comparison with g if the condition $|v_{\mathrm{S}}(t)/V_{\mathrm{T}}| \ll 1$ is fulfilled.

In complete analogy to the small-signal conductance, we may define the small-signal capacitance of a nonlinear capacitor as the change $\mathrm{d}q$ of stored charge associated with the change $\mathrm{d}v$ of applied voltage,

$$c(v) = \mathrm{d}q/\mathrm{d}v . \tag{1.10}$$

If the (bias-dependent) small-signal capacitance $c(v)$ is given, the charge that is stored on the capacitor charged to voltage v is

$$q(v) = \int_0^v c(V)\,\mathrm{d}V . \tag{1.11}$$

If a voltage $v(t) = V_0 + v_{\mathrm{S}}(t)$ is applied to the capacitor the current through the element is

$$i(t) = \mathrm{d}q/\mathrm{d}t = c[v(t)]\,\mathrm{d}v_{\mathrm{S}}/\mathrm{d}t . \tag{1.12}$$

Developing $c[v(t)]$ into a Taylor series up to second order,

$$c[v(t)] \approx c(V_0) + \left.\frac{\mathrm{d}c}{\mathrm{d}V}\right|_{V_0} v_{\mathrm{S}}(t) + \frac{1}{2}\left.\frac{\mathrm{d}^2 c}{\mathrm{d}V^2}\right|_{V_0} v_{\mathrm{S}}^2(t) = c + c_1 v_{\mathrm{S}}(t) + c_2 v_{\mathrm{S}}^2(t) ,$$

yields the small-signal current up to third order of $v_{\mathrm{S}}(t)$. Nonlinear terms represented by c_1 and c_2 contribute, for example, to the high-frequency intermodulation distortion (see Appendix A) in bipolar transistors.

1.2.3 Two-Ports

Since bipolar transistors have three terminals, as is illustrated in Fig. 1.12 a for an npn bipolar transistor, three terminal currents (I_{C}, I_{B} and I_{E}) together with three terminal voltages (V_{CE}, V_{BE} and V_{CB}) are used to describe their electrical characteristics. Due to Kirchhoff's laws only two currents and two voltages can be chosen independently, while the third of each derives from the relations

$$I_{\mathrm{E}} = I_{\mathrm{C}} + I_{\mathrm{B}} \quad \text{and} \quad V_{\mathrm{CB}} = V_{\mathrm{CE}} - V_{\mathrm{BE}} .$$

These four variables split up into two independent (controlling) variables and two dependent (controlled) variables. Under dc operating conditions the elec-

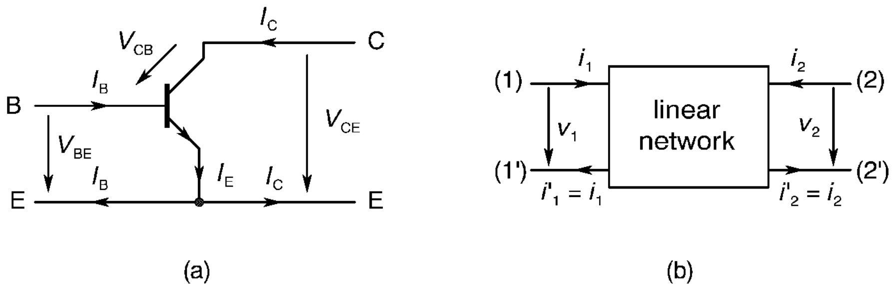


Fig. 1.12. A two-port. (**a**) npn bipolar transistor in common-emitter configuration considered as a two-port and (**b**) the general block symbol

trical behavior is completely described by the dependence of the two terminal currents I_1 and I_2 (e.g., base current I_B and collector current I_C) on the two terminal voltages V_1 and V_2 (e.g., V_{BE} and V_{CE}),

$$I_1 = I_1(V_1, V_2) \quad \text{and} \quad I_2 = I_2(V_1, V_2) \; .$$

Bipolar transistors may therefore be considered as two-ports, i.e., four-terminal networks with an input port and an output port, where both terminals of each port carry the same current (Fig. 1.12 and Appendix B).

Linear Small-Signal Analysis of Nonlinear Two-Ports

In a small-signal analysis one is interested, for example, in the changes $i_{s1}(t)$ and $i_{s2}(t)$ in the input and output currents that result from changes $v_{s1}(t)$ and $v_{s2}(t)$ in the input and output voltages. Under small-signal conditions these are related by the first-order differential:

$$\begin{aligned}
i_{s1}(t) &= \left(\frac{\partial I_1}{\partial V_1}\right)_{V_2} v_{s1}(t) + \left(\frac{\partial I_1}{\partial V_2}\right)_{V_1} v_{s2}(t) \\
&= y_{11}(0) v_{s1}(t) + y_{12}(0) v_{s2}(t) \\
i_{s2}(t) &= \left(\frac{\partial I_2}{\partial V_1}\right)_{V_2} v_{s1}(t) + \left(\frac{\partial I_2}{\partial V_2}\right)_{V_1} v_{s2}(t) \\
&= y_{21}(0) v_{s1}(t) + y_{22}(0) v_{s2}(t) \; .
\end{aligned}$$

The parameters $y_{\alpha\beta}(0)$ are the (dc) admittance parameters or y-parameters of the device. These are coefficients of a Taylor series expansion and therefore depend on the bias point. Capacitive or inductive effects are not considered in this description. For electrical excitation with a given angular frequency, ω, the electrical behavior may, however, be described using complex admittance parameters: if voltages

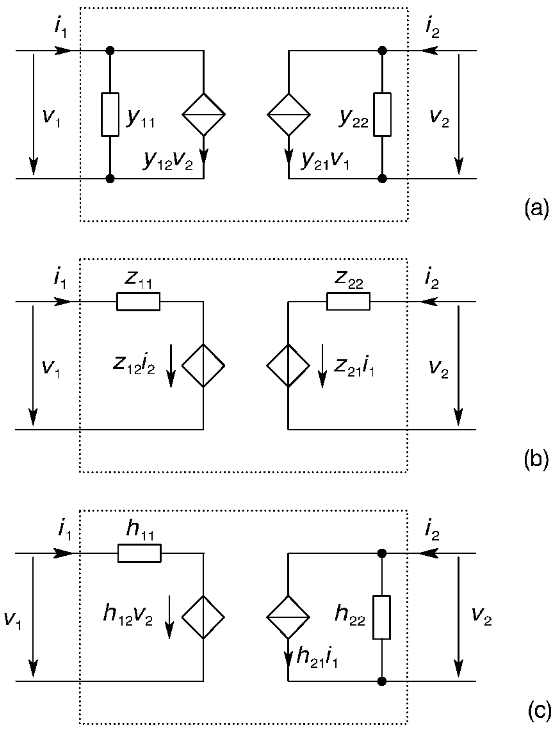


Fig. 1.13. Network representations of linear two-ports in (**a**) admittance form, (**b**) impedance form and (**c**) hybrid form [33, 36]

$$v_1(t) = V_1 + \mathrm{Re}\left(\underline{\hat{v}}_1\,\mathrm{e}^{\mathrm{j}\omega t}\right) \quad \text{and} \quad v_2(t) = V_2 + \mathrm{Re}\left(\underline{\hat{v}}_2\,\mathrm{e}^{\mathrm{j}\omega t}\right)$$

are applied to a transistor, the currents

$$i_1(t) = I_1 + \mathrm{Re}\left(\underline{\hat{i}}_1\,\mathrm{e}^{\mathrm{j}\omega t}\right) \quad \text{and} \quad i_2(t) = I_2 + \mathrm{Re}\left(\underline{\hat{i}}_2\,\mathrm{e}^{\mathrm{j}\omega t}\right) .$$

result under small-signal conditions, with complex phasors $\underline{i}_1$, $\underline{i}_2$, $\underline{v}_1$ and $\underline{v}_2$ which are coupled by a linear relationship[9],

$$\begin{pmatrix} \underline{i}_1 \\ \underline{i}_2 \end{pmatrix} = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix} \begin{pmatrix} \underline{v}_1 \\ \underline{v}_2 \end{pmatrix} . \tag{1.13}$$

The admittance parameters or y-parameters, $y_{\alpha\beta}$, are generally complex frequency-dependent quantities, and determine admittances and controlled current sources in an equivalent circuit of the two-port (Fig. 1.13a).

[9] In this notation, the first index denotes the line, and the second index the row of the admittance matrix in which the respective parameter is placed. Alternative representations use letter subscripts i, o, f and r according to the code

11	≡	i	input	12	≡	r	reverse transfer
21	≡	f	forward transfer	22	≡	o	output

Another choice of dependent and independent parameters leads to the so-called hybrid representation of the two-port characteristics. In this form the input voltage and the output current are chosen as dependent, while the input current and the output voltage are chosen as the independent variables, i.e., the characteristics are represented in the form $V_1 = V_1(I_1, V_2)$ and $I_2 = I_2(I_1, V_2)$. The corresponding small-signal characteristics give the small-signal voltage $v_{s1}(t)$ at the input and the small-signal current $i_{s2}(t)$ at the output that result from small changes $i_{s1}(t)$ and $v_{s2}(t)$ of the independent variables:

$$\begin{aligned} v_{s1}(t) &= \left(\frac{\partial V_1}{\partial I_1}\right)_{V_2} i_{s1}(t) + \left(\frac{\partial V_1}{\partial V_2}\right)_{I_1} v_{s2}(t) \\ &= h_{11}(0) i_{s1}(t) + h_{12}(0) v_{s2}(t) \\ i_{s2}(t) &= \left(\frac{\partial I_2}{\partial I_1}\right)_{V_2} i_{s1}(t) + \left(\frac{\partial I_2}{\partial V_2}\right)_{I_1} v_{s2}(t) \\ &= h_{21}(0) i_{s1}(t) + h_{22}(0) v_{s2}(t) \ . \end{aligned}$$

The parameters $h_{\alpha\beta}(0)$ are termed dc hybrid parameters or h-parameters and can be represented as slopes of different transistor characteristics. Under ac conditions complex frequency-dependent hybrid parameters $h_{\alpha\beta}$ must be used. The phasors $\underline{v}_1$ and $\underline{i}_2$ then are connected with the phasors $\underline{i}_1$ and $\underline{v}_2$ by

$$\begin{pmatrix} \underline{v}_1 \\ \underline{i}_2 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} \begin{pmatrix} \underline{i}_1 \\ \underline{v}_2 \end{pmatrix} . \tag{1.14}$$

The coefficients $h_{\alpha\beta}$ of the hybrid matrix are of particular interest if the input of the two-port is current-controlled. Impedance parameters or z-parameters describe the voltages at the input and output port as functions of the input and output current,

$$\begin{pmatrix} \underline{v}_1 \\ \underline{v}_2 \end{pmatrix} = \begin{pmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{pmatrix} \begin{pmatrix} \underline{i}_1 \\ \underline{i}_2 \end{pmatrix} . \tag{1.15}$$

The impedance parameters $z_{\alpha\beta}$ determine impedances and controlled voltage sources in an equivalent circuit of the two-port (Fig. 1.13b). For a description of the behavior of two-ports in series, the chain parameters $a_{\alpha\beta}$ or $ABCD$-parameters are helpful:

$$\begin{pmatrix} \underline{v}_1 \\ \underline{i}_1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \underline{v}_2 \\ -\underline{i}_2 \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \underline{v}_2 \\ -\underline{i}_2 \end{pmatrix} . \tag{1.16}$$

Each representation provides a complete description of the two-port behavior, i.e., the coefficients of various representations transform into one another.

Table 1.2. Transformation of two-port parameters

	Y	H	Z	A
Y	$\begin{matrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{matrix}$	$\begin{matrix} \frac{1}{h_{11}} & \frac{-h_{12}}{h_{11}} \\ \frac{h_{21}}{h_{11}} & \frac{\Delta_h}{h_{11}} \end{matrix}$	$\begin{matrix} \frac{z_{22}}{\Delta_z} & \frac{-z_{12}}{\Delta_z} \\ \frac{-z_{21}}{\Delta_z} & \frac{z_{11}}{\Delta_z} \end{matrix}$	$\begin{matrix} \frac{a_{22}}{a_{12}} & \frac{-\Delta_a}{a_{12}} \\ \frac{-1}{a_{12}} & \frac{a_{11}}{a_{12}} \end{matrix}$
H	$\begin{matrix} \frac{1}{y_{11}} & \frac{-y_{12}}{y_{11}} \\ \frac{y_{21}}{y_{11}} & \frac{\Delta_y}{y_{11}} \end{matrix}$	$\begin{matrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{matrix}$	$\begin{matrix} \frac{\Delta_z}{z_{22}} & \frac{z_{12}}{z_{22}} \\ \frac{-z_{21}}{z_{22}} & \frac{1}{z_{22}} \end{matrix}$	$\begin{matrix} \frac{a_{12}}{a_{22}} & \frac{\Delta_a}{a_{22}} \\ \frac{-1}{a_{22}} & \frac{a_{21}}{a_{22}} \end{matrix}$
Z	$\begin{matrix} \frac{y_{22}}{\Delta_y} & \frac{-y_{12}}{\Delta_y} \\ \frac{-y_{21}}{\Delta_y} & \frac{y_{11}}{\Delta_y} \end{matrix}$	$\begin{matrix} \frac{\Delta_h}{h_{22}} & \frac{h_{12}}{h_{22}} \\ \frac{-h_{21}}{h_{22}} & \frac{1}{h_{22}} \end{matrix}$	$\begin{matrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{matrix}$	$\begin{matrix} \frac{a_{11}}{a_{21}} & \frac{\Delta_a}{a_{21}} \\ \frac{1}{a_{21}} & \frac{a_{22}}{a_{21}} \end{matrix}$
A	$\begin{matrix} \frac{-y_{22}}{y_{21}} & \frac{-1}{y_{21}} \\ \frac{-\Delta_y}{y_{21}} & \frac{-y_{11}}{y_{21}} \end{matrix}$	$\begin{matrix} \frac{-\Delta_h}{h_{21}} & \frac{-h_{11}}{h_{21}} \\ \frac{-h_{22}}{h_{21}} & \frac{-1}{h_{21}} \end{matrix}$	$\begin{matrix} \frac{z_{11}}{z_{21}} & \frac{\Delta_z}{z_{21}} \\ \frac{1}{z_{21}} & \frac{z_{22}}{z_{21}} \end{matrix}$	$\begin{matrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{matrix}$

Table 1.2 shows the transformations for the most important sets; the abbreviations Δ_y, Δ_z, Δ_h and Δ_a denote the determinant of the respective coefficient matrix, for example, $\Delta_y = y_{11}y_{22} - y_{12}y_{21}$.

1.2.4 Device Modeling

For the computation of circuit properties, device models, i.e., mathematical descriptions of each circuit element, are required.

One approach is to use numerical table models based on a set of measurement points and numerical interpolation procedures. This approach calls for lots of data to be stored, a disadvantage that can be overcome by using empirical formulas fitted to the measured data to obtain a description of device characteristics in terms of mathematical formulas. Although this technique generally provides a device model in a rather short time, it is not widely used, since this approach does not provide any insight [37].

Physical device models, which are derived from the basic semiconductor equations, are therefore of more interest. In this approach one begins by determining the important physical processes and introduces suitable approximations in order to arrive at tractable closed-form expressions for circuit analysis [38]. As it is always possible to increase the accuracy of the description by increasing the model complexity, the appropriate model, that is the simplest one providing sufficient accuracy for the problem under study, has to be chosen in order to perform efficient simulations. IC designers who employ circuit simulation for design verification must therefore be aware of the properties and limitations of the models used.

Since analytical formulas for the device characteristics can only be obtained under idealizing assumptions, parameters used for circuit simulation are generally not calculated from process data. The usual approach is to fit "physically motivated" analytical functions to measured data. To illustrate this we consider the current voltage characteristic of the pn junction, which under idealized conditions can be calculated with the result [5]

$$I(V) = I_{\mathrm{S}} \left[\exp\left(\frac{V}{V_{\mathrm{T}}}\right) - 1 \right] .$$

The saturation current I_{S} is determined by geometric extensions, dopant concentrations, minority-carrier mobilities, contact properties and temperature (see Section 1.4). In order to adopt this ideal case to a real pn junction diode, an additional "empirical" parameter N is introduced, resulting in the expression

$$I(V) = I_{\mathrm{S}} \left[\exp\left(\frac{V}{N V_{\mathrm{T}}}\right) - 1 \right] .$$

Fitting I_{S} and N to measured data generally provides an accurate representation of the current voltage characteristic over several decades of current.

Physical device models should employ a set of analytical functions with continuous derivatives[10] in order to avoid numerical instabilities. These functions employ various model parameters which may be expressed in terms of process data and device dimensions at least for idealized situations. Therefore, it is possible to estimate the parameters from geometry and doping data and to forecast the effect of process modifications on the model parameters. Furthermore, the correlation between fluctuations of model parameters due to process tolerances can be calculated from physical device models.

[10]While the linear small-signal response is determined by the first-order derivatives of the device characteristics, harmonics and intermodulation products are determined by derivatives of higher order. This imposes substantial demands on model accuracy, since the models employed for such investigations have to yield second- and third-order derivatives with sufficient accuracy.

1.3 Semiconductors

Semiconductors[11] are materials with an energy gap, which is a region on the energy scale in which no electron states exist. This gap separates the valence band from the conduction band. Pure semiconductors at low temperatures behave as insulators – no electrons are present in the conduction band, while the valence band is completely filled. Conductivity is observed if electrons are removed from the valence band or introduced into the conduction band. This can be achieved by incorporation of acceptor and donor impurities and leads to p-type or n-type conduction. For doping concentrations of practical interest, the majority-carrier distribution can be controlled by adjusting the doping concentration, while the minority-carrier concentration is determined from the law of mass action.

Silicon and germanium are so-called elementary semiconductors, i.e. solids composed of identical atoms. Elementary semiconductors are located in group IV of the periodical system and crystallize in a diamond lattice. Compound semiconductors are covalent-bonded crystals that consist of different elements: binary semiconductors employ two different elements, ternary semiconductors three and quaternary semiconductors four.

1.3.1 Electrons and Holes

Electronic states in atoms are quantized and can be grouped to shells. Electrical and chemical properties are determined almost exclusively by the so-called valence electrons in the outer shell, since only these electrons have ionization energies in the range of a few electron volts. Electrons in inner shells have ionization energies that typically exceed 100 eV and are therefore tightly bound to "their" nuclei, since comparable energies are not available in typical electrical or chemical elementary processes. Atoms with a completely filled outer shell show large ionization energies and are chemically inert. For that reason we consider the nucleus together with the filled shells of an atom as an entity, the lattice ion, that is surrounded by the valence electrons.

In semiconductors, neighboring atoms are coupled by covalent bonds[12]. In silicon each lattice atom has four nearest neighbors and and shares a bonding orbital with each of these [42]. At 0 K these bonding orbitals are fully occupied with two electrons of different spin orientation. According to Pauli's exclusion principle no electron state may be occupied twice. This excludes

[11] Sections 1.3 to 1.7 provide a survey on elementary semiconductor physics and its applications to pn junctions and bipolar transistors. The reader who has not been exposed to this material before will find very readable presentations in [39] or [40] at the introductory level or in [41], which focuses on bipolar devices, but requires some prior knowledge.

[12] If two atoms come close to each other, the wave functions of the valence electrons will overlap; for certain distances this may cause a reduction of the total energy – then a covalent bond is formed.

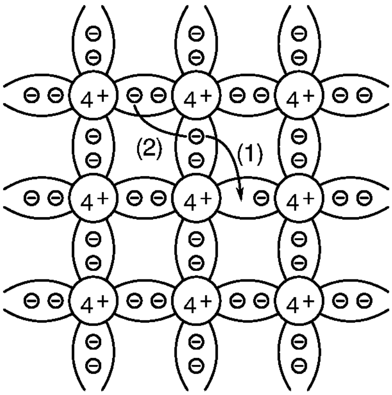


Fig. 1.14. Carrier transport in the valence band

current transport at 0 K, since electrons are hindered in moving to adjacent lattice sites. With increasing temperature, however, electrons can gain enough energy to leave bonding states and are excited into higher conduction band states, which are unoccupied in the ground state and therefore may carry a current. In addition to this, unoccupied bonding states appear (Fig. 1.14) that can be occupied by electrons in adjacent orbitals (1), which in turn leave unoccupied states behind (2).

In a semiconductor crystal, the orbitals of the valence electrons of adjacent atoms overlap and electrons may move from one lattice site to another: delocalized states appear. These states which are closely spaced on the energy scale may be grouped to (energy) bands: the valence band and the conduction band. Current, i.e., charge transport, in semiconductors is made possible by occupied states – electrons – in the conduction band and unoccupied states holes in the valence band.

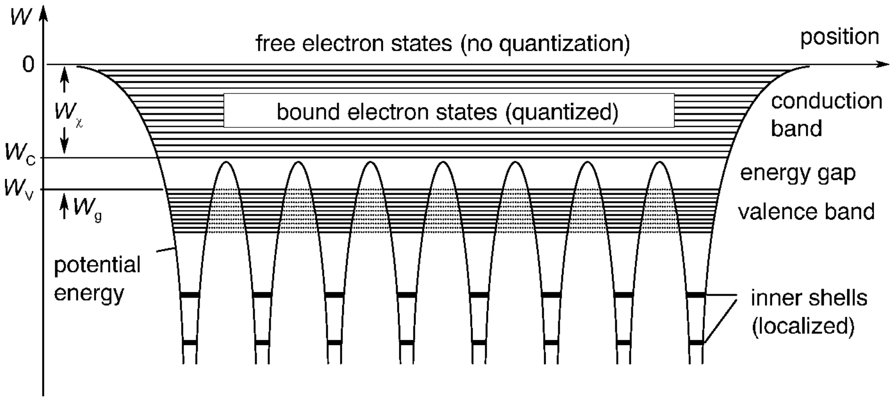


Fig. 1.15. Energy bands

In semiconductors the valence and conduction bands do not overlap and are separated by an energy gap or bandgap,

$$W_g = W_C - W_V \,. \tag{1.17}$$

This is the minimum energy that has to be transferred to an electron in the valence band in order to excite it to the conduction band. The energy required by an electron which is located at the conduction band edge, W_C, to be able to leave the crystal is denoted as the electron affinity, W_χ. Generally, the energy scale is chosen such that electrons bound to the crystal show negative energy values (Fig. 1.15).

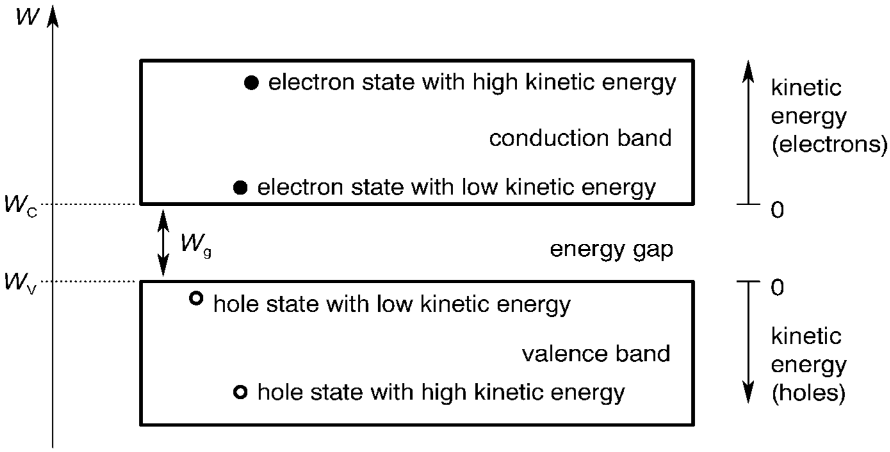


Fig. 1.16. Energy band scheme of a semiconductor if no electric field is present

Phenomena occurring in semiconductor devices are often illustrated with the one-dimensional band scheme depicted in Fig. 1.16. In such a scheme the location of the conduction-band minimum W_C (conduction-band edge) and valence-band maximum W_V (valence-band edge) on the energy scale is shown as a function of position. The kinetic energy of electrons in the conduction band and holes in the valence band is measured from the respective band edges. The location of the band edges can be interpreted as the potential energy of the mobile carriers. The potential energy will become position-dependent as soon as an electric field appears.

1.3.2 Thermal Equilibrium

In thermal equilibrium no current flows and the density of electrons and holes is determined solely by temperature and doping, independent of time. Generation of electron–hole pairs (i.e., the excitation of an electron from the valence to the conduction band) then compensates recombination of electron–hole pairs (i.e., the transition of electrons from conduction-band states to unoccupied valence-band states).

The electron density n is determined from the density of states and their probability of occupation, i.e., the Fermi distribution (see Sect. 2.2). As long as the Fermi energy W_F lies at least $3\,k_BT$ below the conduction-band edge[13], the electron density n_0 in thermal equilibrium is approximately

$$n_0 \approx N_C \exp\left(\frac{W_F - W_C}{k_B T}\right) , \tag{1.18}$$

where $k_B \approx 1.38 \times 10^{-23}\,\mathrm{J/K}$ denotes the Boltzmann constant, T the absolute temperature and N_C the effective density of states in the conduction band; the subscript 0 is used to indicate the state of thermal equilibrium. A similar expression is obtained for the hole density p_0 in thermal equilibrium:

$$p_0 \approx N_V \exp\left(\frac{W_V - W_F}{k_B T}\right) , \tag{1.19}$$

where N_V denotes the effective density of states in the valence band. Multiplication of n_0 and p_0 yields the mass-action law:

$$n_0\, p_0 = N_C N_V\, \mathrm{e}^{-W_g/k_B T} = n_i^2(T) , \tag{1.20}$$

according to which the product of the electron and hole densities equals the square of the intrinsic carrier density $n_i(T)$, independent of doping.

Table 1.3. Parameters of important semiconductors at $T = 300\,\mathrm{K}$ (from [43–45])

Semiconductor	ϵ_r	N_C ($\mathrm{cm^{-3}}$)	N_V ($\mathrm{cm^{-3}}$)	W_g (eV)	n_i ($\mathrm{cm^{-3}}$)
Ge	16.3	1.04×10^{19}	6.1×10^{18}	0.66	2.4×10^{13}
Si	11.8	2.86×10^{19}	3.1×10^{19}	1.12	1.02×10^{10}
GaAs	10.9	4.7×10^{17}	7×10^{18}	1.42	1.79×10^{6}

In intrinsic semiconductors (no doping) electrons in the conduction band must have been excited from the valence band, i.e., electrons and holes are generated in pairs, and therefore

$$n_0 = p_0 = n_i .$$

Using (1.18) and (1.19), the Fermi energy W_{Fi} of the intrinsic semiconductor is therefore as follows:

$$W_{Fi} = \frac{W_V + W_C}{2} + \frac{k_B T}{2} \ln\left(\frac{N_V}{N_C}\right) . \tag{1.21}$$

[13] The location of the Fermi energy on the energy scale can be altered by doping.

Since the second term on the right-hand side is of the order of several meV, the Fermi energy W_{Fi} in intrinsic semiconductors lies to a good approximation in the middle of the energy gap.

Doping

Doping, i.e., controlled incorporation of impurities into a semiconductor lattice, allows the ratio of the electron to the hole concentration to be changed. Impurities that can easily give an electron to the conduction band are termed donors, while impurities that can easily pick up an electron from the valence band are termed acceptors. As an example, the doping of silicon with boron (B) and arsenic (As) is considered.

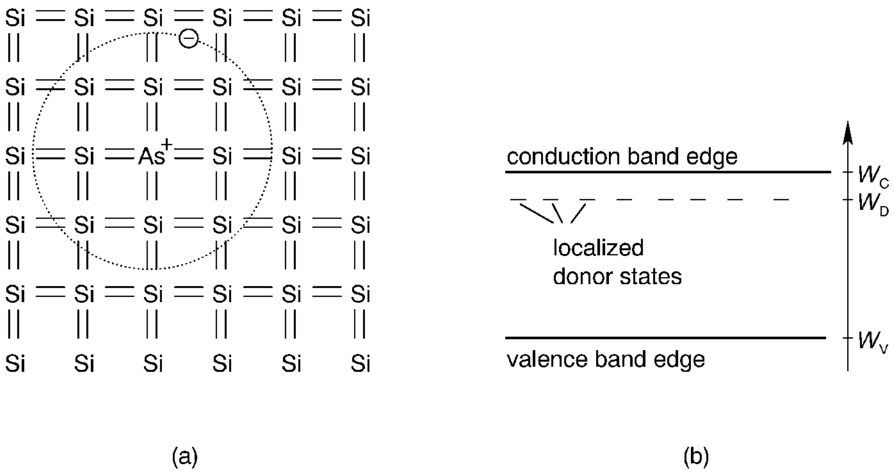


Fig. 1.17. Donors. (**a**) Incorporation of arsenic into the silicon lattice as an As^{+}ion with a weakly bound electron; (**b**) band scheme with localized donor states

Arsenic (As) has five valence electrons. If an arsenic atom substitutes a silicon atom in the lattice, only four valence electrons are required to fill the molecular orbitals shared with the nearest neighbors, i.e., arsenic is incorporated as an As^{+} ion. The remaining electron of the As atom behaves similar[14] to an electron bound in a hydrogen atom [47]. Due to dielectric screening the potential energy is, however, reduced to $V(r) = -e^2/(4\pi\epsilon r)$, where $\epsilon = \epsilon_0\epsilon_{\mathrm{r}}$ is the permittivity of the semiconductor and r denotes the separation of the electron from the positive charge center. Considering this in the formula for the ionization energy of the hydrogen atom and replacing the free electron

[14] The following consideration can only provide an order-of-magnitude estimation. For a more refined analysis of impurity states in semiconductors, see, e.g., [46].

mass m_e by the effective mass m_n^*, yields for the ionization energy of the donor state[15]

$$W_D - W_C = -\frac{m_n^* e^4}{32\pi^2\hbar^2\epsilon^2} = -\frac{1}{\epsilon_r^2}\frac{m_n^*}{m_e} W_H \, , \tag{1.22}$$

where $W_H = 13.6$ eV denotes the Rydberg energy. Due to the large value of the dielectric constant ϵ_r in semiconductors, the binding energy $W_C - W_D$ for typical donors such as As is in the range $10 - 50$ meV, orders of magnitude below the ionization energy W_H of the hydrogen atom. Such weakly bound electron states are considered in the band scheme as localized states below the conduction-band edge (Fig. 1.17). In weakly doped semiconductors, at $T = 0$ K all valence electrons are bound to the As atom, i.e. all donor states are occupied and the conductivity is zero. At room temperature most donors are ionized – the density of donor atoms introduced into the semiconductor lattice therefore allows to manipulate the conductivity.

Boron (B) has three valence electrons. Incorporating boron into a silicon lattice leaves one of the bonding orbitals filled incompletely – a localized unoccupied electron state results that is energetically close to the valence-band edge (Fig. 1.18). Electrons from the valence band are easily excited

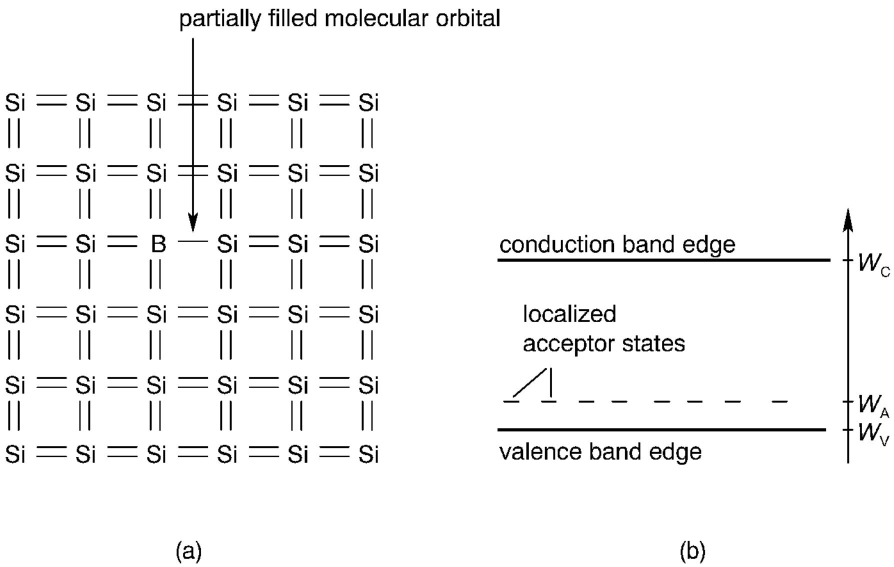


Fig. 1.18. Acceptors. (**a**) Incorporation of a boron atom into the silicon lattice; (**b**) one-dimensional band scheme showing localized acceptor states above the valence-band edge

[15]The Bohr radius is increased to $a_0 = 4\pi\hbar^2\epsilon/(m_n^* e^2)$ and is typically of the order of 10 nm. Since this value is large in comparison with the lattice constant, the description of the dielectric screening in terms of the macroscopic dielectric constant ϵ_r is justified a posteriori.

thermally to such acceptor states, leaving unoccupied states in the valence band, which can carry a hole current.

The lattice of compound semiconductors is built with atoms which have different numbers of valence electrons. In GaAs for example, Ga has three valence electrons and As five. If a Ga atom is substituted by an atom with two valence electrons (such as zinc) the impurity will act as an acceptor; in contrast if an As atom is replaced by an atom with six valence electrons (such as selenium) the impurity will act as a donor. Atoms with four valence electrons may act as donors or acceptors – dependent on whether they substitute a Ga or As atom. Such impurities are termed amphoteric.

Carrier Densities in Doped Semiconductors

Homogeneously doped semiconductors are neutral in thermal equilibrium, i.e.,

$$p_0 - n_0 + N_\mathrm{D}^+ - N_\mathrm{A}^- \; = \; 0 \,, \tag{1.23}$$

where N_D^+ denotes the density of ionized donors and N_A^- the density of ionized acceptors. In combination with the mass-action law, $n_0\,p_0 \; = \; n_\mathrm{i}^2$, the equilibrium carrier densities can be written as follows

$$n_0 \;=\; \frac{1}{2}\left[N_\mathrm{D}^+ - N_\mathrm{A}^- + \sqrt{(N_\mathrm{D}^+ - N_\mathrm{A}^-)^2 + 4\,n_\mathrm{i}^2}\right] \tag{1.24}$$

$$p_0 \;=\; \frac{1}{2}\left[N_\mathrm{A}^- - N_\mathrm{D}^+ + \sqrt{(N_\mathrm{A}^- - N_\mathrm{D}^+)^2 + 4\,n_\mathrm{i}^2}\right] . \tag{1.25}$$

Doping concentrations are generally chosen to be large in comparison with the intrinsic carrier density n_i for the temperature range of operation. Thus, in n-type semiconductors ($N_\mathrm{A}^- \approx 0$) the relation $N_\mathrm{D}^+ \gg n_\mathrm{i}$ is fulfilled, and the majority-carrier density $n_{\mathrm{n}0}$ and the minority-carrier density $p_{\mathrm{n}0}$ are, respectively,

$$n_{\mathrm{n}0} \approx N_\mathrm{D}^+ \quad \text{and} \quad p_{\mathrm{n}0} \approx n_\mathrm{i}^2/N_\mathrm{D}^+ \,. \tag{1.26}$$

In p-type semiconductors, $N_\mathrm{D}^+ \approx 0$ holds and the majority-carrier density $p_{\mathrm{p}0}$ and minority-carrier density $n_{\mathrm{p}0}$ are, respectively,

$$p_{\mathrm{p}0} \approx N_\mathrm{A}^- \quad \text{and} \quad n_{\mathrm{p}0} \approx n_\mathrm{i}^2/N_\mathrm{A}^- \,, \tag{1.27}$$

i.e., the majority-carrier densities $n_{\mathrm{n}0}$ and $p_{\mathrm{p}0}$ are determined by the doping concentration and are approximately independent of temperature, while the minority-carrier concentrations $n_{\mathrm{p}0}$ and $p_{\mathrm{n}0}$ show a strong temperature dependence. Equations (1.26) and (1.27) may only be used in a finite interval of temperature. At low temperature, carriers are trapped by the impurities and reduce the majority-carrier densities, while at high temperature the intrinsic carrier density may become comparable with or larger than the doping concentration (see Sect. 2.2). In the intrinsic case, i.e., for dopant concentrations N_{D} or N_{A} that are small in comparison with n_{i},

$$n_{\mathrm{n0}} \approx p_{\mathrm{n0}} \approx n_{\mathrm{i}}$$

is fulfilled. The semiconductor then is no longer n-type or p-type; pn junctions lose their rectifying property if used in this temperature range.

Fermi Energy in Doped Semiconductors

The Fermi energy is determined from

$$n_0 = N_{\mathrm{C}} \exp\left(\frac{W_{\mathrm{F}} - W_{\mathrm{C}}}{k_{\mathrm{B}}T}\right) , \quad p_0 = N_{\mathrm{V}} \exp\left(\frac{W_{\mathrm{V}} - W_{\mathrm{F}}}{k_{\mathrm{B}}T}\right)$$

and the neutrality condition (1.23). In p-type semiconductors the approximations $N_{\mathrm{D}}^{+} \approx 0$ and $n_{\mathrm{p0}} \approx 0$ hold, resulting in

$$W_{\mathrm{F}} \approx W_{\mathrm{V}} + k_{\mathrm{B}}T \ln(N_{\mathrm{V}}/N_{\mathrm{A}}^{-}) . \tag{1.28}$$

With increasing values of ionized-acceptor density N_{A}^{-}, the Fermi energy shifts to the valence-band edge. In n-type semiconductors $N_{\mathrm{A}}^{-} \approx 0$ and $p_{\mathrm{n0}} \approx 0$ are fulfilled, resulting in

$$W_{\mathrm{F}} \approx W_{\mathrm{C}} - k_{\mathrm{B}}T \ln(N_{\mathrm{C}}/N_{\mathrm{D}}^{+}) . \tag{1.29}$$

An increase in the density N_{D}^{-} of ionized donors causes a shift of the Fermi energy to the conduction-band edge.

1.3.3 Drift and Diffusion Currents

Current can be transported by majority carriers and by minority carriers. While majority carriers flow predominantly due to drift, minority carriers can also flow due to diffusion, which is caused by the thermal motion of the carriers in conjunction with a spatial dependence of the carrier concentration.

If a voltage is applied to a semiconductor, the resulting electric field will accelerate the carriers – holes in and electrons opposite to the direction of the electric-field vector. Since the potential energy of the carriers in the electric field depends on position, the band edges in the one-dimensional band scheme are no longer horizontal. An electron that is accelerated in the electric field without losing energy to the lattice simply transforms potential to kinetic energy without changing its total energy; its path has to be represented as a horizontal line in the one-dimensional band scheme (Fig. 1.19).

Due to scattering at lattice imperfections, the directed motion of the carriers that results from the acceleration in the electric field is randomized very often,

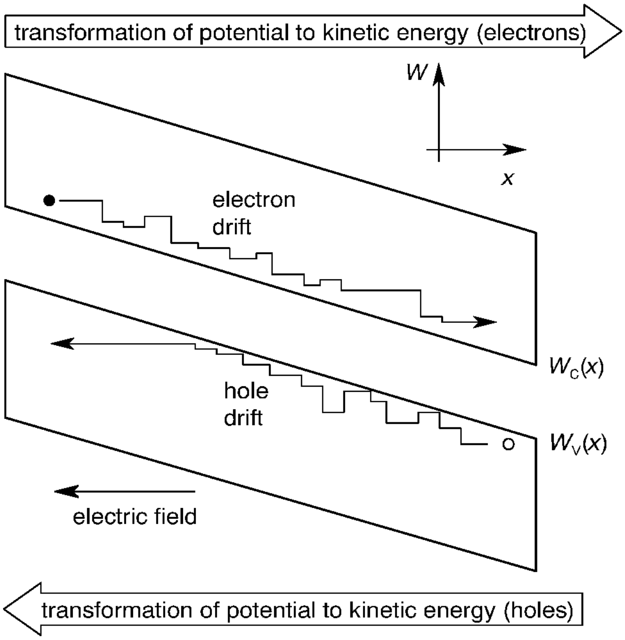


Fig. 1.19. Energy band scheme for a semiconductor in a homogeneous electric field

with the result that the average drift velocity assumes a finite value. Collisons with the lattice may change the particle energies; this has to be represented as vertical changes of the $W(x)$ curve in the one-dimensional band scheme (Fig. 1.19). Since the probability of such collisions increases rapidly with particle energy, only a few carriers will gain energies in the range of 1 eV or beyond – most $W(x)$ curves will appear in a band with a width of 100 mV at the band edges $W_C(x)$ and $W_V(x)$.

The drift velocities v_n and v_p for electrons and holes depend on the electric field according to

$$v_n = -\mu_n E \quad \text{and} \quad v_p = \mu_p E \; ; \tag{1.30}$$

the quantities μ_n and μ_p are the mobilities of electrons and holes. These quantities depend on temperature, impurity concentration and the electric field itself, thus causing a non-ohmic behavior at large values of applied voltage. The drift current densities J_n and J_p for electrons and holes for a given electric-field strength, E, vary in proportion to the carrier density:

$$\begin{array}{rcrcrcl} J_n & = & -env_n & = & e\mu_n nE & = & \sigma_n E \\ J_p & = & epv_p & = & e\mu_p pE & = & \sigma_p E \end{array} \; . \tag{1.31}$$

As long as the carrier density does not vary with position, the thermal motion of carriers will not cause a current, since the number of carriers transported out of a certain volume element will, on average, equal the number transported into it. This is no longer true if the carrier density varies with position. Then, the thermal motion of the carriers leads to a particle diffusion current from the high-density regions to the low-density regions. This current varies to the first order with the carrier concentration gradient, resulting in the following expressions for the diffusion current:

$$J_{\mathrm{n}} = eD_{\mathrm{n}}\frac{\partial n}{\partial x} \quad \text{and} \quad J_{\mathrm{p}} = -eD_{\mathrm{p}}\frac{\partial p}{\partial x}\,. \tag{1.32}$$

The diffusion coefficients D_{n} and D_{p} for electrons and holes are related to the respective mobilities by the Einstein relations,

$$D_{\mathrm{n}} = \mu_{\mathrm{n}} V_{\mathrm{T}} \quad \text{and} \quad D_{\mathrm{p}} = \mu_{\mathrm{p}} V_{\mathrm{T}}\,, \tag{1.33}$$

where $V_{\mathrm{T}} = k_{\mathrm{B}}T/e$ denotes the thermal voltage. The current equations for electrons and holes are obtained by combination of the expressions for drift and diffusion current densities:

$$J_{\mathrm{n}} = e\mu_{\mathrm{n}} nE + eD_{\mathrm{n}}\frac{\partial n}{\partial x} \quad \text{and} \quad J_{\mathrm{p}} = e\mu_{\mathrm{p}} pE - eD_{\mathrm{p}}\frac{\partial p}{\partial x}\,. \tag{1.34}$$

1.3.4 Generation and Recombination

In thermal equilibrium the rate at which electrons are excited from the valence band to the conduction band equals the rate at which electrons recombine with holes from the valence band, and the densities of electrons and holes stay constant. Deviations of electron and hole densities from their equilibrium values cause preferred generation or recombination, since the semiconductor tries to recover the equilibrium situation. The recombination rate R denotes the rate at which the electron (hole) density decreases due to recombination,[16]

$$\partial n/\partial t|_{\mathrm{rec}} \;=\; \partial p/\partial t|_{\mathrm{rec}} \;=\; -R\,,$$

while the generation rate G denotes the rate at which the electron (hole) density increases due to generation,

$$\partial n/\partial t|_{\mathrm{gen}} \;=\; \partial p/\partial t|_{\mathrm{gen}} \;=\; G\,.$$

The net recombination rate is defined as the difference $R - G$. In a doped semiconductor its value is approximately proportional to the deviation of the respective minority-carrier density from its equilibrium value, i.e.,

$$R - G \;=\; \frac{p_{\mathrm{n}} - p_{\mathrm{n0}}}{\tau_{\mathrm{p}}} \qquad \text{in n-type regions} \quad \text{and} \tag{1.35}$$

$$R - G \;=\; \frac{n_{\mathrm{p}} - n_{\mathrm{p0}}}{\tau_{\mathrm{n}}} \qquad \text{in p-type regions}\,, \tag{1.36}$$

[16]Electrons and holes will recombine at the same rate if trapping effects at the recombination centers may be neglected (see Sect. 2.6).

if light-induced generation–recombination processes and impact ionization are neglected. The minority-carrier lifetimes τ_n and τ_p for electrons and holes strongly depend on the doping concentration.

Since energy is required for a generation process and energy is set free in the course of a recombination process, a third particle (phonon, photon, electron or hole) must be involved in the generation or recombination process. Electron–hole pairs can, for example, be generated by light. This requires a photon of energy $h\nu > W_g$ to be absorbed within the semiconductor. The energy of this photon is then used to excite an electron from the valence band to the conduction band (Fig. 1.20).

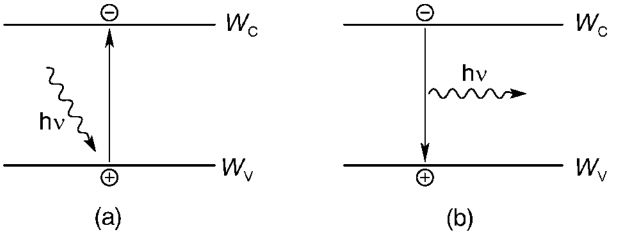


Fig. 1.20. (**a**) Generation and (**b**) recombination by absorption and emission of a photon

An electron from the conduction band may drop into an unoccupied state in the valence band, i.e., recombine with a hole and emit the resulting energy as a photon. This process is, however, not very likely in semiconductors with and indirect energy gap, such as silicon.

Impurities in semiconductors generally cause localized states in the energy gap. In these, electrons or holes may be trapped, resulting in a two-step generation or recombination process (Fig. 1.21).

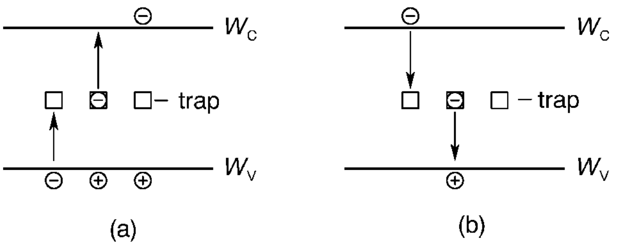


Fig. 1.21. Shockley–Read–Hall mechanism for the (**a**) generation and (**b**) recombination of an electron-hole pair

This mechanism was first investigated by Shockley, Read and Hall and is therefore termed Shockley–Read–Hall (SRH) recombination. The SRH mechanism works most efficiently if the impurity level is located in the middle of the energy gap (so-called deep impurities), i.e., donor or acceptor impurities (so-called shallow impurities) do not act as SRH centers. Recombination in modestly doped silicon regions is predominantly SRH recombination. In heavily doped semiconductor regions the majority-carrier density is sufficiently large in order to make so-called Auger processes the predominant recombination mechanism. In these processes, the energy set free by a recombining electron hole pair is transferred to a third carrier, which then has a large value of kinetic energy.

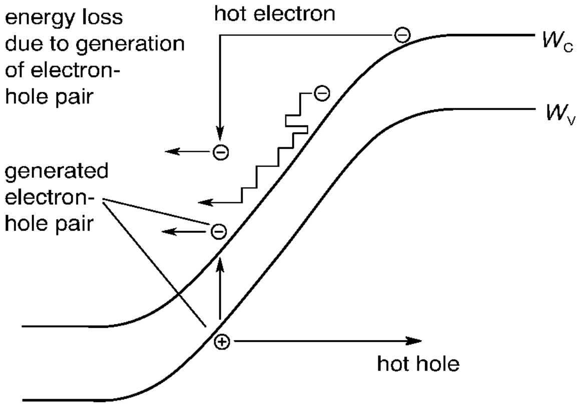


Fig. 1.22. Impact ionization in a reverse-biased pn junction

Electrons or holes with kinetic energies in excess of W_g are able to generate another electron–hole pair, a mechanism termed as impact ionization (Fig. 1.22). This phenomenon causes avalanche breakdown of pn junctions.

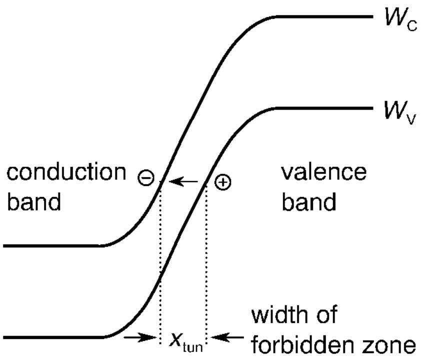


Fig. 1.23. Generation of an electron–hole pair by internal field emission (interband tunneling)

In heavily doped pn junctions electrons may tunnel from the valence band to the conduction band (Fig. 1.23). This mechanism, which is termed the Zener effect, internal field emission or interband tunneling, causes the generation of an electron–hole pair. The probability for this process exponentially depends on the width x_{tun} of the forbidden zone (Fig. 1.23) and therefore – since $|E|x_{\mathrm{tun}} \approx W_g$ – on the electric-field strength E.

1.3.5 Basic Semiconductor Equations

In the current equations (1.34)

$$J_\mathrm{n}(x,t) \;=\; e\mu_\mathrm{n} n(x,t)E(x,t) + eD_\mathrm{n}\,\partial n/\partial x \tag{1.37}$$

$$J_\mathrm{p}(x,t) \;=\; e\mu_\mathrm{p} p(x,t)E(x,t) - eD_\mathrm{p}\,\partial p/\partial x\ , \tag{1.38}$$

five field variables, $J_\mathrm{n}(x,t)$, $J_\mathrm{p}(x,t)$, $n(x,t)$, $p(x,t)$ and $E(x,t)$, appear, that require three additional equations – the continuity equations and the Poisson equation – for a unique solution.

The continuity equations for electrons and holes relate changes in particle densities to the current density and the net recombination rate. The continuity equation for electrons reads in one-dimensional form

$$\frac{\partial n}{\partial t} \;=\; \frac{1}{e}\frac{\partial J_\mathrm{n}}{\partial x} - (R-G)\ , \tag{1.39}$$

while the continuity equation for holes is given by

$$\frac{\partial p}{\partial t} \;=\; -\frac{1}{e}\frac{\partial J_\mathrm{p}}{\partial x} - (R-G)\ . \tag{1.40}$$

In addition to current and continuity equations an extra relation is necessary that allows the electric-field strength E or the electrostatic potential ψ,

$$E \;=\; -\partial\psi/\partial x\ , \tag{1.41}$$

to be calculated from the distribution of the electric charges. This job is done by the Poisson equation

$$\partial^2\psi/\partial x^2 \;=\; -\rho/\epsilon\ , \tag{1.42}$$

where $\rho = e(p-n+N_\mathrm{D}^+ - N_\mathrm{A}^-)$ denotes the charge density and $\epsilon = \epsilon_0\epsilon_\mathrm{r}$ is the permittivity of the semiconductor. These equations were established in 1950 by van Roosbroeck [48].

In the absence of an electric field ($E = 0$) the current and continuity equations can be combined to give the diffusion equations for electrons in p-type semiconductors and holes in n-type semiconductors:

$$\frac{\partial n_\mathrm{p}}{\partial t} \;=\; -\frac{n_\mathrm{p}-n_\mathrm{p0}}{\tau_\mathrm{n}} + D_\mathrm{n}\frac{\partial^2 n_\mathrm{p}}{\partial x^2} \tag{1.43}$$

$$\frac{\partial p_\mathrm{n}}{\partial t} \;=\; -\frac{p_\mathrm{n}-p_\mathrm{n0}}{\tau_\mathrm{p}} + D_\mathrm{p}\frac{\partial^2 p_\mathrm{n}}{\partial x^2}\ . \tag{1.44}$$

These are important for the study of minority-carrier transport.

1.4 PN Junctions

PN junctions are fundamental to all bipolar devices and a thorough understanding of their electrical behavior is mandatory for the understanding of bipolar transistor operation.

1.4.1 PN Junctions in Thermal Equilibrium

An abrupt pn junction is formed by a homogeneously doped n-type region with donor concentration N_D adjacent to a homogeneously doped p-type region with acceptor concentration N_A as depicted in Fig. 1.24.

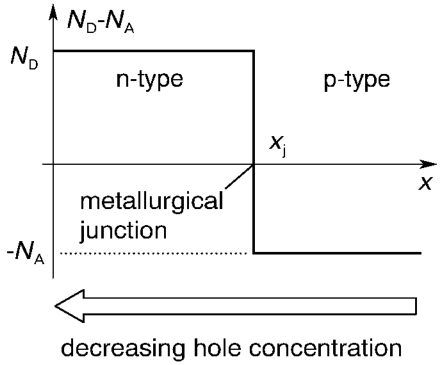


Fig. 1.24. Abrupt pn junction

Due to the gradient of the electron and hole densities at the junction, electrons will diffuse from the n-type to the p-type region, while holes will diffuse in the opposite direction. Thus, a positive space charge builds up on the donor-doped side and a negative space charge on the acceptor-doped side. The electric field $E(x)$ caused by this dipole layer will finally stop the diffusion current and confine electrons to the n-type region and holes to the p-type region.

Depletion Approximation

Figure 1.25a and b sketch the electron and hole densities in an abrupt junction in thermal equilibrium as a function of position x. Since the electron density $n(x)$ rapidly decreases within the space-charge region (a corresponding statement holds for the hole density $p(x)$), the depletion approximation has proven useful. This approximation introduces the space-charge-layer boundaries located at x_n and x_p to divide the pn diode into three regions: the depletion layer (or space charge layer) ($x_\mathrm{n} < x < x_\mathrm{p}$) and the n-type and p-type regions, which are both assumed to be electrically neutral. Within the space-charge layer the density of mobile carriers is assumed to be negligible, resulting the space charge density $\rho = eN_\mathrm{D}$ on the n-side and $\rho = -eN_\mathrm{A}$ on the p-side (Fig.

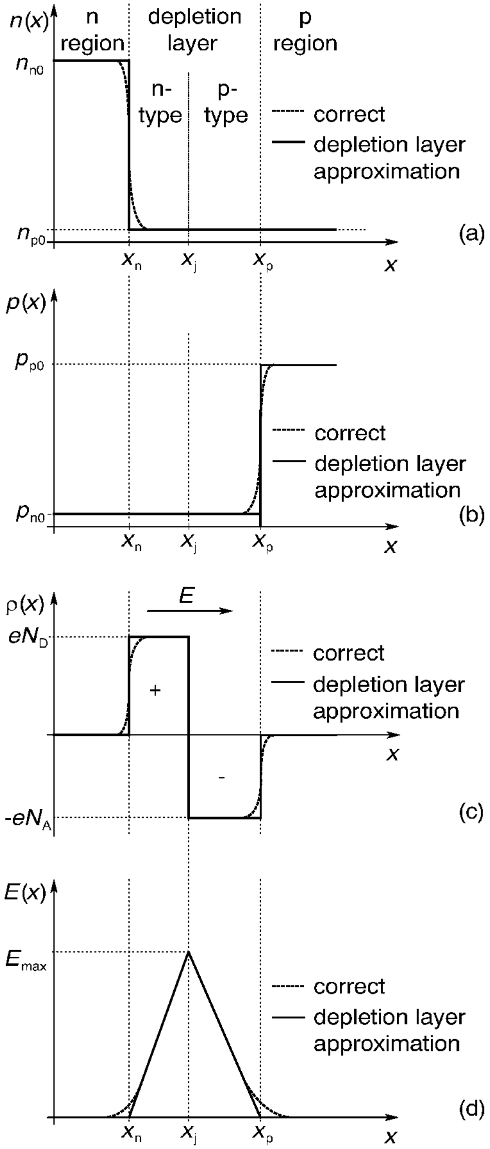


Fig. 1.25. Abrupt pn junction in thermal equilibrium. (**a**) Electron density, (**b**) hole density, (**c**) space charge density and (**d**) electric field

1.25c, complete ionization of impurities, $N_A \approx N_A^-$ and $N_D \approx N_D^+$, assumed for simplicity) a position-dependent electric field with maximum value at the metallurgical junction exists (Fig. 1.25d). In thermal equilibrium the Fermi energy is constant throughout all regions. According to Fig. 1.26, therefore, the identity

$$eV_J = W_g - k_B T \ln(N_C/N_D) - k_B T \ln(N_V/N_A)$$

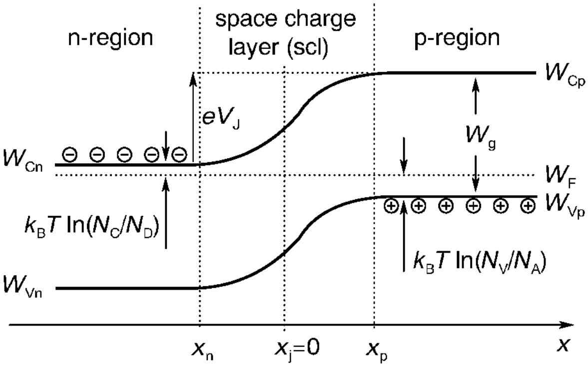


Fig. 1.26. Band scheme of a pn junction in thermal equilibrium

has to be fulfilled. Considering that $n_i^2 = N_C N_V \exp(-W_g/k_B T)$, this allows one to calculate the height V_J of the potential barrier, the so-called built-in voltage,

$$V_J \approx V_T \ln\left(N_A N_D/n_i^2\right) . \tag{1.45}$$

The locations of the depletion-layer edges x_n and x_p result from the potential difference ψ_j across the junction and the requirement of electrical neutrality

$$\int_{x_n}^{x_p} N(x)\,dx = 0 , \tag{1.46}$$

where $N(x) = N_D(x) - N_A(x)$ denotes the net doping concentration. With $\rho(x) = eN(x)$, a double integration of the Poisson equation (1.42) yields

$$\psi_j = -\frac{e}{\epsilon}\int_{x_n}^{x_p}\int_{x_n}^{x} N(x')\,dx'\,dx = \frac{e}{\epsilon}\int_{x_n}^{x_p} x\,N(x)\,dx , \tag{1.47}$$

where $\psi_j = V - V_J$ is the total potential difference in the presence of an external voltage, V. Equations (1.46) and (1.47) provide two equations for the determination of the two unknowns x_n and x_p. In the case of an abrupt pn junction, the neutrality condition (1.46) reads

$$x_n N_D + x_p N_A = 0 ,$$

whereas (1.47) transforms to

$$N_D x_n^2 + N_A x_p^2 = 2\epsilon(V_J - V)/e .$$

Thus, the extensions of the depletion layer into the n-type and p-type regions are

$$-x_n = \sqrt{\frac{2\epsilon N_A (V_J - V)}{e N_D (N_A + N_D)}} \quad \text{and} \quad x_p = \sqrt{\frac{2\epsilon N_D (V_J - V)}{e N_A (N_A + N_D)}} , \tag{1.48}$$

and the depletion layer width is

$$d_{\mathrm{j}} = x_{\mathrm{p}} - x_{\mathrm{n}} = \sqrt{\frac{2\epsilon}{e}\frac{N_{\mathrm{A}}+N_{\mathrm{D}}}{N_{\mathrm{A}}N_{\mathrm{D}}}(V_{\mathrm{J}}-V)}\ . \tag{1.49}$$

In asymmetric pn junctions with different doping concentrations on either side, the depletion layer will therefore extend predominantly into the low-doped region. The value of electric-field strength follows after simple integration of Poisson's equation (1.42),

$$E(x) = \frac{e}{\epsilon}\int_{x_{\mathrm{n}}}^{x} N(x')\,\mathrm{d}x'\ ; \tag{1.50}$$

its maximum value occurs at the metallurgical junction and is given by

$$E_{\mathrm{max}} = \sqrt{\frac{2eN_{\mathrm{A}}N_{\mathrm{D}}(V_{\mathrm{J}}-V)}{\epsilon(N_{\mathrm{A}}+N_{\mathrm{D}})}}\ . \tag{1.51}$$

Example 1.4.1 Consider an abrupt pn junction with a donor concentration $N_{\mathrm{D}} = 10^{18}\ \mathrm{cm}^{-3}$ and an acceptor concentration $N_{\mathrm{A}} = 10^{16}\ \mathrm{cm}^{-3}$ in silicon ($\epsilon_{\mathrm{r}} = 11.9$) at $T = 300\ \mathrm{K}$, and assume complete ionization of the impurities. With the intrinsic carrier density $n_{\mathrm{i}} = 1.02 \times 10^{10}\ \mathrm{cm}^{-3}$ and the thermal voltage $V_{\mathrm{T}} = k_{\mathrm{B}}T/e = 25.9\ \mathrm{mV}$, the built-in voltage is

$$V_{\mathrm{J}} = 25.9\ \mathrm{mV} \times \ln\left(\frac{10^{34}}{(1.02 \times 10^{10})^2}\right) = 834\ \mathrm{mV}\ .$$

Since

$$\frac{x_{\mathrm{j}} - x_{\mathrm{n}}}{x_{\mathrm{p}} - x_{\mathrm{j}}} = \frac{N_{\mathrm{A}}}{N_{\mathrm{D}}} = \frac{1}{100}\ ,$$

the depletion layer extends predominantly into the p-type region. In such so-called one-sided junctions, both the depletion layer width,

$$d_{\mathrm{j}} = \sqrt{\frac{2\epsilon(N_{\mathrm{A}}+N_{\mathrm{D}})V_{\mathrm{J}}}{eN_{\mathrm{A}}N_{\mathrm{D}}}} \approx \sqrt{\frac{2\epsilon V_{\mathrm{J}}}{eN_{\mathrm{A}}}} \approx 0.33\ \mu\mathrm{m}\ ,$$

and the maximum electric-field (here calculated at $V = 0$),

$$E_{\mathrm{max}} = \sqrt{\frac{2eN_{\mathrm{A}}N_{\mathrm{D}}V_{\mathrm{J}}}{\epsilon(N_{\mathrm{A}}+N_{\mathrm{D}})}} \approx \sqrt{\frac{2eN_{\mathrm{A}}V_{\mathrm{J}}}{\epsilon}} \approx 50\ \frac{\mathrm{kV}}{\mathrm{cm}}\ ,$$

are determined by the dopant concentration on the weakly doped side.

1.4.2 Forward-Biased PN Junction

Forward biasing a pn junction lowers the potential barrier between n-type and p-type regions. This allows electrons to diffuse at an increased rate from the n-type to the p-type region and holes vice versa.

Shockley Boundary Conditions

Without biasing, the ratio of hole density p_{n0} in the n-type region to the hole density p_{p0} in the p-type region is

$$\frac{p_{n0}}{p_{p0}} = \exp\left(\frac{W_{Vn} - W_{Vp}}{k_B T}\right) = \exp\left(-\frac{V_J}{V_T}\right) .$$

An analogous relation holds for the ratio of the electron density n_{p0} in the p-type region to the electron density n_{n0} in the n-type region:

$$\frac{n_{p0}}{n_{n0}} = \exp\left(\frac{W_{Cn} - W_{Cp}}{k_B T}\right) = \exp\left(-\frac{V_J}{V_T}\right) .$$

Application of a forward bias, V, to the pn junction reduces the height of the potential barrier and causes minority-carrier injection. This results in increased minority-carrier densities at the depletion-layer edges in accordance with the boundary conditions[17]

$$\frac{p_n(x_n)}{p_p(x_p)} = \exp\left(-\frac{V_J - V}{V_T}\right) \tag{1.52}$$

and

$$\frac{n_p(x_p)}{n_n(x_n)} = \exp\left(-\frac{V_J - V}{V_T}\right) . \tag{1.53}$$

At the depletion-layer edges the following generalized mass-action law holds:

$$n_p(x_p) p_p(x_p) = n_n(x_n) p_n(x_n) = n_i^2 \exp(V/V_T) . \tag{1.54}$$

Low-level injection means that the density of injected minority carriers is small in comparison with the corresponding majority-carrier density. With the approximations $p_p(x_p) \approx p_{p0} \approx N_A$ and $n_n(x_n) \approx n_{n0} \approx N_D$, (1.52) and (1.53) give the Shockley boundary conditions[18]:

$$p_n(x_n) \approx p_{n0} \exp\left(\frac{V}{V_T}\right) \quad \text{and} \quad n_p(x_p) \approx n_{p0} \exp\left(\frac{V}{V_T}\right) , \tag{1.55}$$

i.e., a forward bias ($V > 0$) causes the minority-carrier densities $n_p(x_p)$ and $p_n(x_n)$ at the depletion-layer edges to increase by a factor $\exp(V/V_T)$ in comparison with their equilibrium densities. Since the density of minority carriers will decrease versus their equilibrium values with increasing distance from the depletion layer, an electron diffusion current[19] will flow into the p-type region and a hole diffusion current will flow into the n-type region.

[17] These assume the carrier densities on both sides of the junction to be in equilibrium with one another. This is not strictly true under forward-bias conditions, where a net current flows as is discussed in further detail in Appendix C.

[18] Under high-level-injection conditions, the modified boundary conditions of Fletcher or Misawa have to be used (see Appendix C).

[19] The electric-field strength in the n-type and p-type regions is assumed to be negligible.

Screening

The Coulomb force exerted by injected minority carriers attracts majority carriers that neutralize the injected charge. This occurs rapidly in materials with high conductivity, σ. The characteristic time for this process is the dielectric relaxation time,

$$\tau_\epsilon = \epsilon/\sigma \,. \tag{1.56}$$

For doping concentrations in excess of 10^{17} cm^{-3}, as they are typical in the different regions of high-frequency bipolar transistors, τ_ϵ is smaller than 1 ps, i.e., these regions can be considered neutral in the whole frequency range of interest. The characteristic length for the screening of injected minorities is given by the Debye length,[20]

$$L_\mathrm{D} = \sqrt{\epsilon V_\mathrm{T}/en}\,, \tag{1.57}$$

where n denotes the majority-carrier density, i.e., n_n in an n-type semiconductor and p_p in a p-type semiconductor. In n-type silicon with $N_\mathrm{D} = 10^{17}$ cm^{-3}, the Debye length is, for example, $L_\mathrm{D} = 13$ nm, i.e., the neutralizing majorities are to a very good approximation at the same place as the injected minorities.

Current–Voltage Characteristics Under Low-Level Injection

The dc current is composed of the current due to electrons injected into the p-type region,

$$I_\mathrm{n} = eA_\mathrm{j}D_\mathrm{n}\,\mathrm{d}n_\mathrm{p}/\mathrm{d}x|_{x_\mathrm{p}}\,, \tag{1.58}$$

and the current due to holes injected into the n-type region (Fig. 1.4.2),

$$I_\mathrm{p} = -eA_\mathrm{j}D_\mathrm{p}\,\mathrm{d}p_\mathrm{n}/\mathrm{d}x|_{x_\mathrm{n}}\,. \tag{1.59}$$

The minority-carrier densities have to be determined from the respective diffusion equations, as is exemplified for the injection of holes into a n-type region of thickness d_n (Fig. 1.27). The diffusion equation for the excess hole density $\Delta p(x) = p_\mathrm{n}(x) - p_\mathrm{n0}$ in the n-type region reads

$$\frac{\mathrm{d}^2\Delta p}{\mathrm{d}x^2} = \frac{\Delta p(x)}{L_\mathrm{p}^2}\,, \tag{1.60}$$

where

$$L_\mathrm{p} = \sqrt{D_\mathrm{p}\tau_\mathrm{p}} \tag{1.61}$$

[20] In degenerate semiconductors, which do not obey Maxwell–Boltzmann statistics, this expression has to be modified.

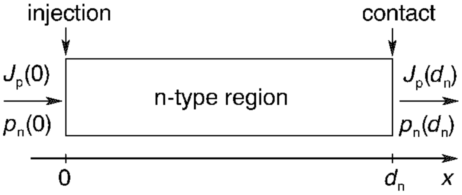


Fig. 1.27. Injection of holes into a n-type region of thickness d_n

denotes the diffusion length for holes in the n-type region. The general solution of (1.60) is

$$\Delta p(x) = \Delta p_+ \exp\left(\frac{x}{L_p}\right) + \Delta p_- \exp\left(-\frac{x}{L_p}\right) , \tag{1.62}$$

with constants Δp_+ and Δp_- that have to be determined from appropriate boundary conditions. Under low-level-injection conditions one of these is given by the Shockley boundary condition, which determines the excess hole density at the depletion-layer edge,

$$\Delta p(0) = p_{n0}\left[\exp\left(\frac{V}{V_T}\right) - 1\right] , \tag{1.63}$$

while the other depends on the contact properties (Fig. 1.30). In so-called long-base diodes, with a contact so far away from the depletion-layer edge that minority carriers cannot reach the contact, the boundary condition

$$\lim_{x\to\infty} \Delta p(x) = 0 \tag{1.64}$$

must be fulfilled, since the hole density approaches its equilibrium value as a result of recombination. This will only be the case if the width d_n of the n-type region is large in comparison with the hole diffusion length L_p. Since this condition is generally not fulfilled, minority carriers will reach the contact. In such, so-called short-base diodes, the mixed boundary condition

$$J_p(d) = -eD_p \left.\frac{d\Delta p}{dx}\right|_{d_n} = eS_p \Delta p(d_n) \tag{1.65}$$

is generally applied. The parameter S_p is the surface recombination velocity[21] for holes at the contact to the n-type region. From (1.62), (1.63) and (1.65) one obtains for the hole density

$$\Delta p(x) = \frac{\sinh\left(\dfrac{d_n - x}{L_p}\right) + \nu_p \cosh\left(\dfrac{d_n - x}{L_p}\right)}{\sinh\left(\dfrac{d_n}{L_p}\right) + \nu_p \cosh\left(\dfrac{d_n}{L_p}\right)} \Delta p(0) , \tag{1.66}$$

[21] See Sects. 2.6 and 3.3 for a discussion of this approach.

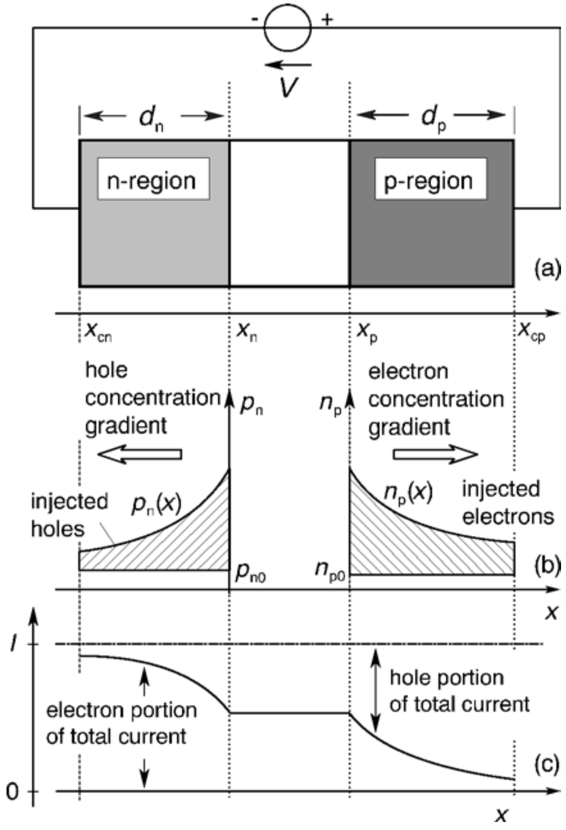


Fig. 1.28. Forward-biased pn junction

where $\nu_p = D_p/S_p L_p$. In a one-sided junction, where only the current due to holes injected into the n-type region counts[22], the exponential current voltage characteristic

$$I = I_S \left[\exp\left(\frac{V}{V_T}\right) - 1\right] \tag{1.67}$$

results from (1.59), with the saturation current

$$I_S = \frac{eA_j D_p p_{n0}}{L_p} \frac{\cosh(d_n/L_p) + \nu_p \sinh(d_n/L_p)}{\sinh(d_n/L_p) + \nu_p \cosh(d_n/L_p)} . \tag{1.68}$$

The diode current and therefore the saturation current $I_S = I_{Sc} + I_{Sr}$ can be split up into the term

$$I_{Sc} = \frac{eA_j D_p p_{n0}}{L_p} \frac{1}{\sinh(d_n/L_p) + \nu_p \cosh(d_n/L_p)} , \tag{1.69}$$

[22] Otherwise an analogous expression for the electron current injected into the p-type region has to be added.

which represents the current recombining at the contact, and the term

$$I_{\mathrm{Sr}} = \frac{1}{\tau_{\mathrm{p}}} \frac{eA_{\mathrm{j}} \int_0^{d_{\mathrm{p}}} \Delta p(x)\, \mathrm{d}x}{\exp(V/V_{\mathrm{T}}) - 1} = \frac{eA_{\mathrm{j}} D_{\mathrm{p}} p_{\mathrm{n0}}}{L_{\mathrm{p}}} \frac{\cosh(d_{\mathrm{n}}/L_{\mathrm{p}}) - 1 + \nu_{\mathrm{p}} \sinh(d_{\mathrm{n}}/L_{\mathrm{p}})}{\sinh(d_{\mathrm{n}}/L_{\mathrm{p}}) + \nu_{\mathrm{p}} \cosh(d_{\mathrm{n}}/L_{\mathrm{p}})} , \quad (1.70)$$

which is due to recombination in the p-type region. For values of d_{n} that are large in comparison with the hole diffusion length L_{p}, the current recombining at the contact vanishes in proportion to $\exp(-d_{\mathrm{n}}/L_{\mathrm{p}})$; under such conditions the diode current will be rather insensitive to the properties of the contact to the n-type region. The long-base diode is described in the limit $d_{\mathrm{n}}/L_{\mathrm{p}} \to \infty$, i.e., $\tanh(d_{\mathrm{n}}/L_{\mathrm{p}}) \to 1$; then

$$I_{\mathrm{S}} \to I_{\mathrm{Sr}} \to eA_{\mathrm{j}} D_{\mathrm{p}} p_{\mathrm{n0}} / L_{\mathrm{p}} ,$$

in accordance with the result derived for the long-base diode in most introductory textbooks.

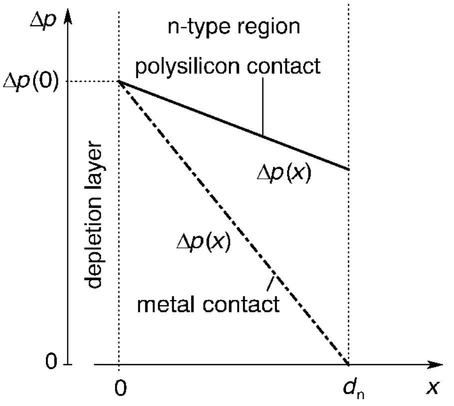


Fig. 1.29. Minority-carrier profiles in a homogeneously doped n-type region for a metal contact and contact with finite surface recombination velocity (polysilicon contact) if recombination in the volume is neglected

In the limit $d_{\mathrm{n}}/L_{\mathrm{p}} \ll 1$, the current due to recombination in the p-type region vanishes in proportion to $d_{\mathrm{n}}/L_{\mathrm{p}}$, i.e., for short p-type regions, the electron current will be dominated by the current recombining at the contact. With the approximations $\cosh(d_{\mathrm{n}}/L_{\mathrm{p}}) \approx 1$ and $\sinh(d_{\mathrm{n}}/L_{\mathrm{p}}) \approx d_{\mathrm{n}}/L_{\mathrm{p}}$, one obtains

$$I_{\mathrm{Sc}} \approx \frac{eA_{\mathrm{j}} D_{\mathrm{p}} p_{\mathrm{n0}}}{d_{\mathrm{n}} + D_{\mathrm{p}}/S_{\mathrm{p}}} \quad (1.71)$$

for the saturation current component that describes recombination at the contact and

$$I_{\mathrm{Sr}} \approx \frac{eA_{\mathrm{j}}D_{\mathrm{p}}p_{\mathrm{n0}}}{d_{\mathrm{n}}+D_{\mathrm{p}}/S_{\mathrm{p}}}\,\nu_{\mathrm{p}}\,\frac{d_{\mathrm{n}}}{L_{\mathrm{p}}} = I_{\mathrm{Sc}}\nu_{\mathrm{p}}\,\frac{d_{\mathrm{n}}}{L_{\mathrm{p}}} \tag{1.72}$$

for the saturation current component that describes volume recombination.

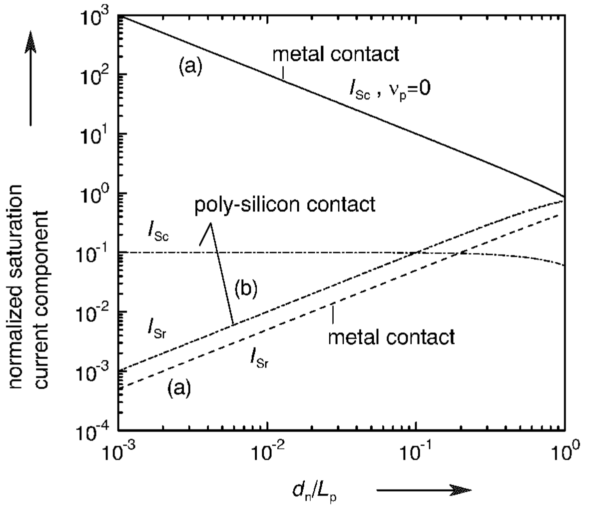


Fig. 1.30. Normalized values (in units of $eA_{\mathrm{j}}D_{\mathrm{p}}p_{\mathrm{n0}}/L_{\mathrm{p}}$) of saturation current components I_{Sc} and I_{Sr} for different recombination velocities at the contact. (**a**) Metal contact ($S_{\mathrm{p}} = \infty$, $\nu_{\mathrm{p}} = 0$), (**b**) polysilicon contact ($S_{\mathrm{p}} = 5000$ cm/s, $\nu_{\mathrm{p}} = 10$)

Simplified Analysis. The same result is obtained from the following simplified analysis, which is suitable if most of the injected minority current reaches the contact. Under this condition the hole current density J_{p} may be assumed to be constant, corresponding to a linear dependence of p on x. Combining

$$J_{\mathrm{p}} = eS_{\mathrm{p}}\Delta p(d) \quad \text{and} \quad J_{\mathrm{p}} = eD_{\mathrm{p}}\frac{\Delta p(0)-\Delta p(d)}{d_{\mathrm{n}}}$$

yields $\Delta p(0)/\Delta p(d) = 1+d_{\mathrm{n}}S_{\mathrm{p}}/D_{\mathrm{p}}$. By combining this result with the Shockley boundary condition (1.63), one obtains

$$I_{\mathrm{p}} = \frac{eA_{\mathrm{j}}S_{\mathrm{p}}}{1+d_{\mathrm{n}}S_{\mathrm{p}}/D_{\mathrm{p}}}\,\Delta p(0) = I_{\mathrm{Sc}}\left[\exp\left(\frac{V}{V_{\mathrm{T}}}\right)-1\right] ,$$

in which I_{Sc} is given by (1.71). The minority charge stored in the n-type region is

$$Q_{\mathrm{T}} = eA_{\mathrm{j}}d_{\mathrm{n}}\frac{\Delta p(0)+\Delta p(d)}{2} = \frac{eA_{\mathrm{j}}d_{\mathrm{n}}}{2}\left(1+\frac{1}{1+d_{\mathrm{n}}S_{\mathrm{p}}/D_{\mathrm{p}}}\right)\Delta p(0) ;$$

this allows the current due to recombination in the n-type layer to be estimated as

$$\frac{Q_{\mathrm{T}}}{\tau_{\mathrm{p}}} = \frac{eA_{\mathrm{j}}D_{\mathrm{p}}}{d_{\mathrm{n}}+D_{\mathrm{p}}/S_{\mathrm{p}}}\,\frac{d_{\mathrm{n}}}{L_{\mathrm{p}}}\left(\nu_{\mathrm{p}}+\frac{d_{\mathrm{n}}}{2L_{\mathrm{p}}}\right)\Delta p(0) \approx I_{\mathrm{Sr}}\left[\exp\!\left(\frac{V}{V_{\mathrm{T}}}\right)-1\right],$$

in which I_{Sr} is given by (1.72) if the term of order $(d_{\mathrm{n}}/L_{\mathrm{p}})^2$ is neglected.

Example 1.4.2 Figure 1.30 shows values of I_{Sc} and I_{Sr} in units of $eA_{\mathrm{j}}D_{\mathrm{p}}p_{\mathrm{n0}}/L_{\mathrm{p}}$ computed according to (1.69) and (1.70) for a metal contact ($S_{\mathrm{p}} \to \infty$, $\nu_{\mathrm{p}} = 0$) and a polysilicon contact ($S_{\mathrm{p}} \approx 5000$ cm/s, $\nu_{\mathrm{p}} \approx 10$; [49]). In both cases the term due to bulk recombination vanishes in proportion to $d_{\mathrm{n}}/L_{\mathrm{p}}$ for values of $d_{\mathrm{n}}/L_{\mathrm{p}} \ll 1$. The current component due to recombination at the contact dominates in the case of the metal contact for all values of $d_{\mathrm{n}}/L_{\mathrm{p}} < 1$, while in the case of the polysilicon contact $I_{\mathrm{Sc}} > I_{\mathrm{Sr}}$ if $d_{\mathrm{n}}/L_{\mathrm{p}} < 0.1$. If $d_{\mathrm{n}}/L_{\mathrm{p}} \ll 0.1$, the current will also be predominantly determined by the recombination at the contact, the n-type region is then called semi-transparent. This situation is found in the emitter regions of state-of-the-art high-frequency bipolar transistors.

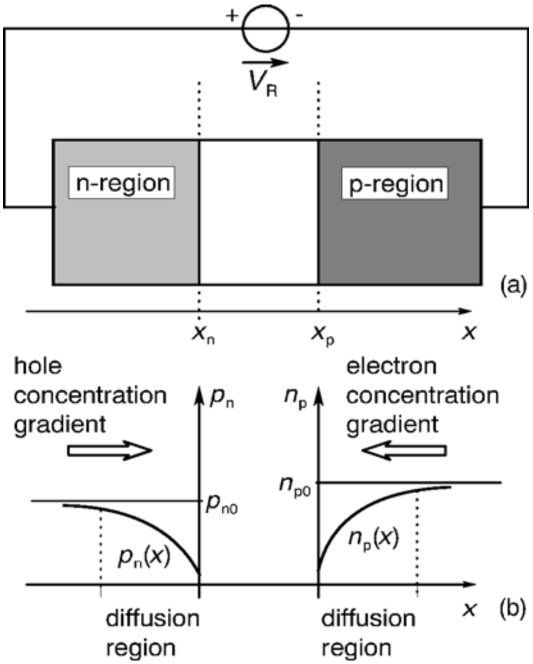


Fig. 1.31. Reverse-biased pn junction. (**a**) Cross section (schematic), (**b**) minority-carrier densities

1.4.3 Reverse-Biased PN Junction

From (1.67) one obtains $I = I_{\mathrm{S}}\,[\exp(V/V_{\mathrm{T}})-1] \to -I_{\mathrm{S}}$ in the limit $V \ll -V_{\mathrm{T}}$, i.e., the reverse-biased pn junction carries a reverse current equal to the saturation current I_S. Figure 1.31 explains the physical mechanism that causes this current.

In reverse bias ($V < 0$) the minority-carrier densities at the depletion layer edges are reduced with respect to their equilibrium values, i.e., $n_p(x_p) < n_{p0}$ and $p_n(x_n) < p_{n0}$. The minority-carrier densities therefore decrease towards the respective depletion-layer edge. Since thermal generation steadily produces minority carriers, a diffusion current will flow towards the depletion layer, causing the reverse saturation current.

If the reverse voltage applied to a pn junction exceeds a critical value, breakdown occurs, which may be either due to internal field emission or due to impact ionization. Both phenomena require the electric-field strength to reach a critical value and are therefore strongly related to the maximum value of electric field strength in the junction. Since this value is determined by the doping concentration and the applied voltage, the breakdown voltage will decrease if the doping concentrations increase.

1.4.4 Stored Charge

A change in the applied voltage v causes a change of electron and hole concentrations in the pn junction. The charge stored by the diode under given bias conditions is described in terms of the depletion capacitance and the diffusion capacitance.

Depletion Capacitance

The stored charge associated with the bias-dependent location of the space-charge-layer boundaries is described by the depletion capacitance. The bias dependence of this capacitance is usually written as[23]

$$c_j(V) = \frac{C_{J0}}{(1-V/V_J)^M} = \frac{dQ_j}{dV}, \tag{1.73}$$

with parameters C_{J0}, M and V_J adjusted to measured data. Integration then gives the charge stored in the depletion layer capacitance

$$Q_j = \int_0^V c_j(v)\, dv = \frac{C_{J0}\, V_J}{1-M}\left[1 - \left(1-\frac{V}{V_J}\right)^{(1-M)}\right]. \tag{1.74}$$

Equation (1.73) is motivated by results derived for the abrupt pn junction and the linear pn junction [50]. Table 1.4 gives equations for the computation of these parameters from doping data. In these equations N_A^- and N_D^+ denotes the densities of ionized acceptors and donors on either side of the

[23] This formula is not correct for large forward bias and shows a divergence for $V \to V_J$ (see Sect. 3.4).

abrupt junction, while a is the gradient of net doping density in the case of the linear junction.

Table 1.4. Capacitance parameters for abrupt and linear graded pn junctions

Junction	C_J	V_J	M
Abrupt	$A_j\sqrt{\dfrac{e\,\epsilon}{2V_J}\dfrac{N_A N_D}{N_A+N_D}}$	$V_T\ln\left(\dfrac{N_A N_D}{n_i^2}\right)$	1/2
Linear	$A_j\left(\dfrac{ea\epsilon^2}{12\,V_J}\right)^{1/3}$	$\dfrac{2V_T}{3}\ln\left(\dfrac{3a^2\epsilon V_J}{2en_i^3}\right)$	1/3

Diffusion Charge

Figure 1.32 shows the charge distribution in a forward-biased pn junction. To reduce the depletion-layer width, a hole charge Q_{jp} and an equal but opposite electron charge Q_{jn} are necessary. Due to minority-carrier injection into the n-type and p-type regions, the minority-carrier charges ΔQ_{pn} and ΔQ_{np}, respectively, occur. Since the injected minorities are screened by additional majority carriers, the additional charges $\Delta Q_{nn} = -\Delta Q_{pn}$ and $\Delta Q_{pp} = -\Delta Q_{np}$, respectively, are supplied through the terminals in order to neutralize the n-type and p-type regions. In dc operation, the diffusion charge, $Q_T = \Delta Q_{pp}+\Delta Q_{pn}$, is proportional to the current I;

$$Q_T = T_T I \, , \tag{1.75}$$

where T_T is termed the transit time. Under reverse-bias conditions the diffusion charge may be neglected. Due to the exponential increase of diode current with forward bias, the diffusion charge will dominate the stored charge for large forward bias (Fig. 1.33). The minority charge Q_T stored in the short-base diode is determined by the current component due to volume recombination with the saturation current I_{Sr} (1.70) and the hole lifetime τ_p

$$\begin{aligned} Q_T &= \tau_p I_{Sr}\left[\exp\left(\frac{V}{V_T}\right)-1\right] \\ &= \tau_p \frac{I_{Sr}}{I_{Sc}+I_{Sr}} I(V) \, , \end{aligned} \tag{1.76}$$

i.e., the transit time T_T is reduced with respect to the minority carrier lifetime τ_p by the factor $I_{Sr}/(I_{Sc}+I_{Sr})$, which is small in comparison to one. For small-signal analysis the (quasi-static) diffusion capacitance [38],

$$c_t = \frac{dQ_T}{dV} = T_T \frac{dI}{dV} \, , \tag{1.77}$$

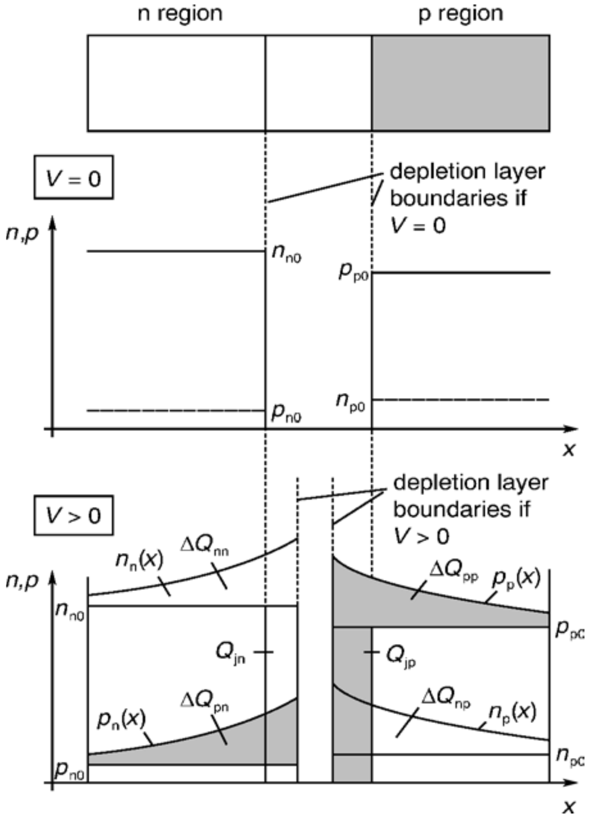


Fig. 1.32. Electron density $n(x)$ and hole density $p(x)$ in pn junction at $V = 0$ and with forward bias $(V > 0)$

is sometimes defined.[24] The value of the quasi-static diffusion capacitance generally differs from the small-signal capacitance,

$$\tilde{c}_{\mathrm{t}} = -\mathrm{j}\,\frac{\mathrm{d}y}{\mathrm{d}\omega}\bigg|_0 = \frac{g_{\mathrm{d}}\tau_{\mathrm{p}}}{2}\left[1 + F\left(\frac{d_{\mathrm{n}}}{L_{\mathrm{p}}}\right)\right], \tag{1.78}$$

which is obtained from an ac solution of the diffusion equation. The term $F(d_{\mathrm{n}}/L_{\mathrm{p}})$ depends on the thickness of the n-type layer and is derived and further discussed in Appendix C.

[24] The diffusion capacitance is not a real capacitance, which stores energy

$$\Delta W = \int_0^V V' c(V')\,\mathrm{d}V'$$

in the form of an electric field, since the charge stored on the diffusion capacitance is neutralized by majority carriers and does not affect the energy of the electrostatic field. Nevertheless, energy is stored on the diffusion capacitance, since injection of electrons in a p-type region introduces carriers in excited conduction-band states with a stored energy that is roughly determined from $\Delta W = W_{\mathrm{g}}\,Q_{\mathrm{T}}/e$. A thermodynamical description that is less coarse-grained can be found in [51].

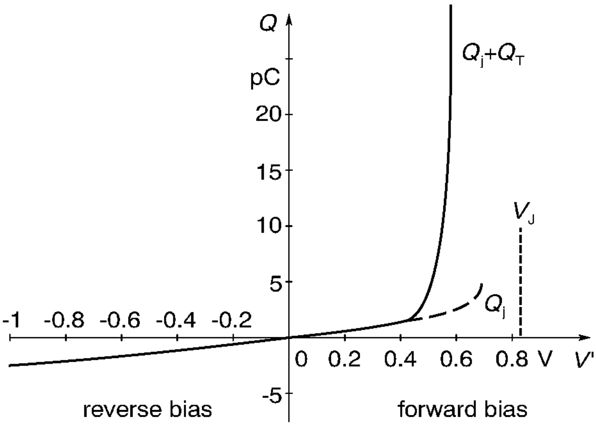


Fig. 1.33. Charge stored in a pn diode as a function of forward bias

1.4.5 Switching, Charge-Control Theory

The charge-control theory assumes that a current flowing into the diode recombines or causes a change in the hole charge in the diode:

$$i(t) = \frac{q_{\mathrm{T}}(t)}{T_{\mathrm{T}}} + c_{\mathrm{j}}(v')\,\frac{\mathrm{d}v'}{\mathrm{d}t} + \frac{\mathrm{d}q_{\mathrm{T}}}{\mathrm{d}t}\,. \tag{1.79}$$

Here i, q_{T} and v' denote the transient diode current, the transient diffusion charge and the voltage drop across the depletion layer. This equation directly derives from the continuity equation and therefore demands virtually no approximations. The corresponding equivalent circuit is shown in Fig. 1.34.

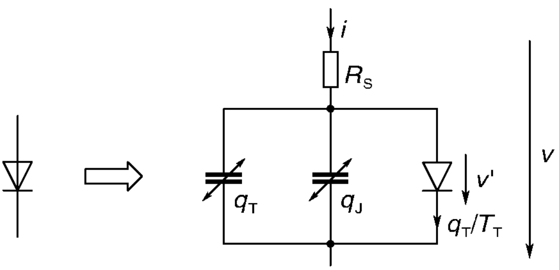


Fig. 1.34. Equivalent-circuit representation of the charge-control model of a pn diode

For a complete description, the bias dependence of the diffusion charge q_{T} has to be known. In charge-control theory the quasi-static approximation

$$q_{\mathrm{T}}(t) = q_{\mathrm{T}}[\,v'(t)\,] \;\approx\; T_{\mathrm{T}} I_{\mathrm{S}}\left\{\exp\left[\,v'(t)/N V_{\mathrm{T}}\,\right] - 1\right\} \tag{1.80}$$

is used, which applies the relationship found for stationary operation to the transient case. This approximation is a compromise between numerical effort and correctness (see Appendix C).

1.4.6 Epitaxial Diodes

Large breakdown voltages and small depletion capacitances require a lightly doped and therefore high-resistance substrate,[25] while small series resistances require a heavily doped semiconductor substrate. Therefore, the base–collector (bc) junction of vertical bipolar transistors is generally realized in a lightly doped epitaxial layer grown on a heavily doped low-resistance subcollector (the buried layer) of the same type.

Low–High Junctions

If a lightly doped region is in contact with a heavily doped region of the same type, a so-called low–high junction results. In the following a nn^+ junction as depicted in Fig. 1.35 will be considered.[26] Due to the gradient of the electron density in the vicinity of the junction, electrons will diffuse from the heavily doped region into the lightly doped region. The heavily doped region thus loses some of its electrons (resulting in a positive sheet of space charge that builds up in the vicinity of the junction), while the lightly doped region gains additional electrons (and becomes negatively charged). The electric field associated with the dipole layer thus formed finally compensates the electron diffusion. An equilibrium consideration similar to to that given in Sect. 1.4.1 yields the potential difference across the low–high junction:

$$V_{\mathrm{JH}} = V_{\mathrm{T}} \ln\left[n_{\mathrm{n}}(x_2)/n_{\mathrm{n}}(x_1)\right] \approx V_{\mathrm{T}} \ln\left[N_{\mathrm{D2}}/n(x_1)\right] . \tag{1.81}$$

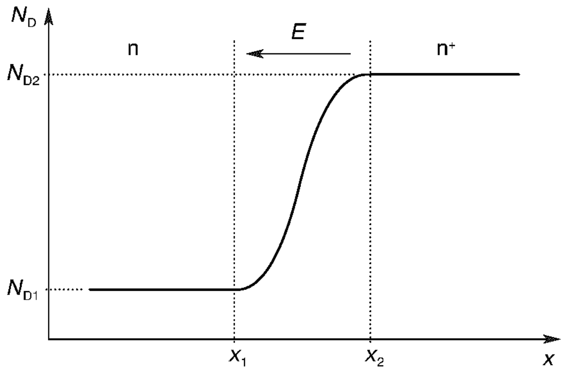


Fig. 1.35. Donor concentration $N_{\mathrm{D}}(x)$ as function of position for a nn^+ low high junction

[25] In pn junctions that are realized by the planar technique, the more heavily doped region always lies close to the surface. Only in epitaxial processes may lightly doped regions be formed on heavily doped ones.

[26] This situation is of interest here, since it serves as a simple model for the contact of the lightly doped epitaxial collector region to the buried layer.

In thermal equilibrium $n_\mathrm{n}(x_1) \approx N_\mathrm{D1}$ and $n_\mathrm{n}(x_2) \approx N_\mathrm{D2}$ and V_JH is determined solely by the dopant concentrations. Hole injection into the lightly doped region, however, may cause an increase in the electron density $n_\mathrm{n}(x_1)$. Assuming neutrality and complete ionization of the donor impurities at the space-charge-layer boundary x_1, i.e., $p_\mathrm{n}(x_1)+N_\mathrm{D1} = n_\mathrm{n}(x_1)$, the potential difference across the low–high junction is

$$V_\mathrm{JH} = V_\mathrm{T} \ln\left(\frac{N_\mathrm{D2}}{N_\mathrm{D1}+p_\mathrm{n}(x_1)}\right) ;$$

it determines the ratio of the hole densities on both sides of the junction,

$$p_\mathrm{n}(x_2) = p_\mathrm{n}(x_1) \exp\left(-\frac{V_\mathrm{JH}}{V_\mathrm{T}}\right) = \frac{p_\mathrm{n}(x_1)[N_\mathrm{D1}+p_\mathrm{n}(x_1)]}{N_\mathrm{D2}} .$$

If the thickness of the heavily doped region is large[27] in comparison with the diffusion length L_p2, the hole current that flows into the heavily doped region is given by $J_\mathrm{p} = eD_\mathrm{p2}[p_\mathrm{n}(x_2)-p_\mathrm{n02}]/L_\mathrm{p2}$. Using $p_\mathrm{n01} = n_\mathrm{i}^2/N_\mathrm{D1}$ and $p_\mathrm{n02} = n_\mathrm{i}^2/N_\mathrm{D2}$, this expression transforms to[28]

$$J_\mathrm{p} = e\frac{D_\mathrm{p2}}{L_\mathrm{p2}}\frac{N_\mathrm{D1}}{N_\mathrm{D2}}\left(\frac{p_\mathrm{n}(x_1)[N_\mathrm{D1}+p_\mathrm{n}(x_1)]}{N_\mathrm{D1}} - p_\mathrm{n01}\right) . \tag{1.82}$$

Under low-level-injection conditions $p_\mathrm{n}(x_1) \ll N_\mathrm{D1}$ holds. Introducing the effective surface recombination velocity at the low–high transition,

$$S_\mathrm{nn+} = \frac{D_\mathrm{p2}}{L_\mathrm{p2}}\frac{N_\mathrm{D1}}{N_\mathrm{D2}} , \tag{1.83}$$

the hole current becomes under these conditions

$$J_\mathrm{p} = eS_\mathrm{nn+}\left[p_\mathrm{n}(x_1)-p_\mathrm{n01}\right] . \tag{1.84}$$

Under high-level-injection conditions, the relation

$$J_\mathrm{p} = eS_\mathrm{nn+}\left(\frac{p_\mathrm{n}(x_1)[N_\mathrm{D1}+p_\mathrm{n}(x_1)]}{N_\mathrm{D1}} - p_\mathrm{n01}\right) , \tag{1.85}$$

has to be used.

[27] If the diffusion length L_p2 in the heavily doped region is not small in comparison with the thickness of the n^+ region, the contact properties will influence the value of the effective surface recombination velocity. A metal contact formed to an n^+ region of thickness d_n2 yields, for example [52],

$$S_\mathrm{nn+} = \frac{D_\mathrm{p2}}{L_\mathrm{p2}}\frac{N_\mathrm{D1}}{N_\mathrm{D2}} \coth\left(\frac{d_\mathrm{n2}}{L_\mathrm{p2}}\right) .$$

[28] In this approach bandgap narrowing has been neglected. Considering bandgap narrowing, the energy gap in the n^+ region will be smaller by ΔW_g than the bandgap in the n-region and $n_\mathrm{i2}^2 = n_\mathrm{i1}^2 \exp(\Delta W_\mathrm{g}/k_\mathrm{B}T)$; then (1.82) has to be replaced by

$$J_\mathrm{p} = e\frac{D_\mathrm{p2}}{L_\mathrm{p2}}\frac{N_\mathrm{D1}}{N_\mathrm{D2}}\left[\frac{p_\mathrm{n}(x_1)[N_\mathrm{D1}+p_\mathrm{n}(x_1)]}{N_\mathrm{D1}} - p_\mathrm{n01}\exp\left(\frac{\Delta W_\mathrm{g}}{k_\mathrm{B}T}\right)\right] .$$

Current–Voltage Characteristics

We consider a p$^+$nn$^+$ diode, with a thickness d of the n-type region that is small in comparison with the hole diffusion length L_p, and neglect recombination in the epitaxial region. Application of the approximations (1.71) and (1.83) yields the hole current that is injected into the epitaxial layer under low-level-injection conditions:

$$I_\mathrm{p,lo} \approx I_\mathrm{S}\left[\exp\left(\frac{V'}{V_\mathrm{T}}\right)-1\right], \quad \text{with} \quad I_\mathrm{S} = \frac{eA_\mathrm{j}D_\mathrm{p1}p_\mathrm{n01}}{d+D_\mathrm{p1}/S_\mathrm{nn^+}}. \tag{1.86}$$

A description of short epitaxial layers that also applies under high-level-injection conditions is given in Appendix C, where an implicit relation for the hole current density is derived. Under low-level-injection conditions, this result reduces to (1.86), whereas under high-level-injection conditions the approximation

$$I_\mathrm{p,hi} \approx I_\mathrm{SH}\exp\left(\frac{V'}{2V_\mathrm{T}}\right), \quad \text{with} \quad I_\mathrm{SH} = \frac{2eA_\mathrm{j}D_\mathrm{p1}n_\mathrm{i1}}{d}, \tag{1.87}$$

is found. Introducing the knee current, $I_\mathrm{KF} = I_\mathrm{SH}^2/I_\mathrm{S}$, allows the two asymptotic relations (1.86) and (1.87) to be combined, to the approximation

$$I_\mathrm{p} \approx \frac{I_\mathrm{p,lo}}{\sqrt{1+I_\mathrm{p,lo}/I_\mathrm{KF}}} \tag{1.88}$$

Figure 1.36 illustrates the corresponding current–voltage characteristic in the absence of series resistances.

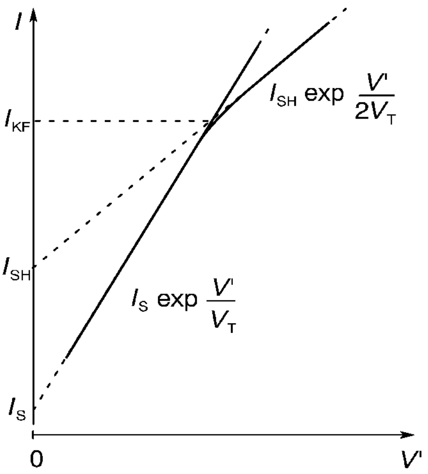


Fig. 1.36. Current–voltage characteristic of an epitaxial diode, illustrating the asymptotic behavior of low-level and high-level injection

1.5 Bipolar Transistor Principles

Bipolar junction transistors employ two closely spaced pn junctions in a semiconductor crystal. It is an npn transistor,[29] if these junctions have a common p-type region, otherwise it is a pnp transistor (Fig. 1.37). The three regions

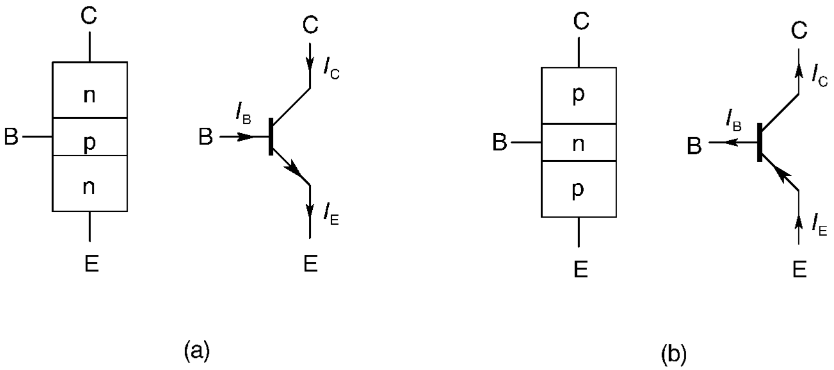


Fig. 1.37. Doping and circuit symbol for (**a**) npn and (**b**) pnp transistors

are known as emitter, base and collector of the transistor, each of these regions has a contact, making the bipolar transistor a three-terminal device. The base terminal serves as a control electrode; it allows the so-called transfer current to be controlled. In npn transistors, this transfer current is carried by electrons that flow from the emitter to the collector, resulting in a positive current from the collector to the emitter. With positive current directions[30] as indicated in Fig. 1.37, Kirchhoff's current law reads $I_E = I_C + I_B$.

1.5.1 Modes of Operation

Figure 1.38 shows the band scheme of an npn bipolar transistor if no external voltages are applied. The Fermi energy is constant throughout the device and there is no current flow.

Upon application of a voltage $V_{CE} > 0$, electrons would flow from emitter to collector if they were able to surmount the potential barrier formed by the base layer. As long as the emitter base (eb) junction is not forward biased, only few electrons will be able to do so, with the result of a negligibly small transfer current. Forward biasing the eb junction ($V_{BE} > 0$) reduces the height of the potential barrier and enables electrons to be injected into the base

[29] In the following we will focus on npn bipolar transistors, which provide better performance in comparison with pnp transistors due to the larger electron mobility.

[30] In this notation all currents are positive under forward operation of the transistor. Another frequently used convention defines the emitter current to be negative under forward operation.

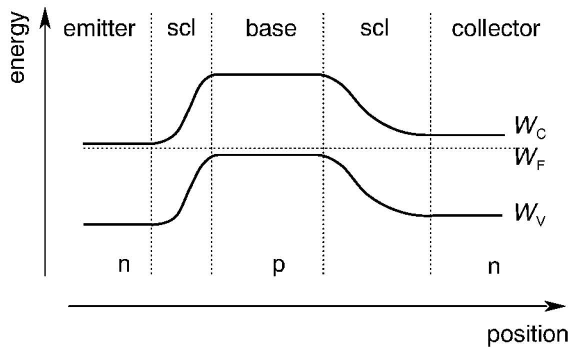


Fig. 1.38. Band scheme of an npn bipolar transistor in equilibrium

layer. These can either recombine in the base layer or diffuse to the base–collector (bc) junction. Choosing the thickness of the base layer to be small in comparison with the diffusion length of the electrons, however, allows most electrons to reach the bc space-charge layer. There, the electric field transports them to the collector region.[31] With a change of forward bias V_{BE}, the current of electrons injected into the base layer changes, and therefore the current of electrons that arrives at the bc space-charge layer (scl). By this mechanism the collector current I_{C}, is controlled by V_{BE}.

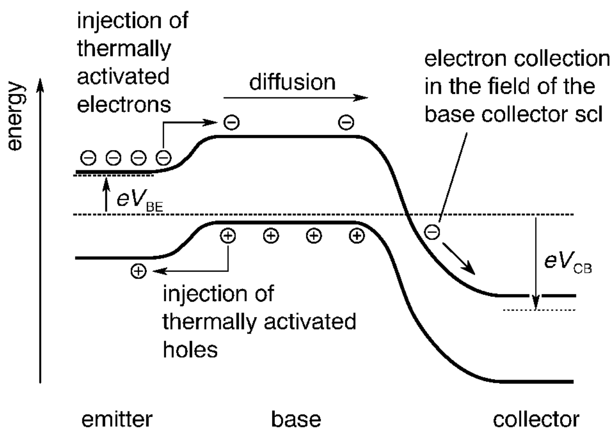


Fig. 1.39. Band scheme of an npn bipolar transistor in forward operation

Forward biasing the eb junction causes injection of holes into the emitter, where they recombine. These holes must be delivered by the base contact and form the base current. In order to control a large collector current with

[31] This explains the notation emitter and collector: the emitter emits the particles carrying the transfer current, the collector collects them after they have passed the base layer. The term base can be understood if the first transistor structure is considered; there, two point contacts (emitter and collector) were placed on top of a semiconductor crystal, which was termed the base (see Sect.1.1).

a small base current, the doping concentration in the emitter is chosen to be much larger than the doping concentration in the base layer. Under these conditions many more electrons are injected into the base region than holes are injected into the emitter region. The forward current gain,

$$B_{\mathrm{N}} = I_{\mathrm{C}}/I_{\mathrm{B}} \,, \tag{1.89}$$

defines the ratio of the collector and emitter currents in normal operation; its value should be much larger than one.

Besides forward operation, with $V_{\mathrm{CE}} > 0$ and electrons flowing from emitter to collector, reverse transistor operation, with $V_{\mathrm{CE}} < 0$, a forward-biased bc junction and electrons flowing from the collector to the emitter, can be observed. The ratio

$$B_{\mathrm{I}} = -I_{\mathrm{E}}/I_{\mathrm{B}} \tag{1.90}$$

defines the current gain for reverse operation. Its value is usually small in comparison with B_{N}, since transistors generally are optimized for forward operation, with an emitter doping concentration that is several orders of magnitude larger than the collector doping concentration.

Dependent on the polarities of V_{BE} and V_{BC} we may generally distinguish between the operating modes listed in Table 1.5.

Table 1.5. Modes of operation

V_{BE}	V_{BC}	Mode of operation (npn)
> 0	< 0	Forward
< 0	> 0	Reverse
< 0	< 0	Cutoff
> 0	> 0	Saturation

1.5.2 Transfer Current

In the following the one-dimensional transistor model depicted in Fig. 1.40 is considered. The space-charge-layer boundaries of the eb junction (with coordinates x_{eb}, x_{be}) and the bc junction (with coordinates x_{bc} and x_{cb}) divide the transistor volume into five regions: emitter (e), base (b), collector (c), the eb space-charge layer (eb) and the bc space-charge layer (bc).

In dc operation the base current I_{B} consists of a component, due to holes injected into the emitter, I_{BE}, a component, due to holes injected into the collector, I_{BC}, and a component, due to recombination of holes in the base region, I_{BB}:

$$I_{\mathrm{B}} \;=\; I_{\mathrm{BE}} + I_{\mathrm{BC}} + I_{\mathrm{BB}} \,. \tag{1.91}$$

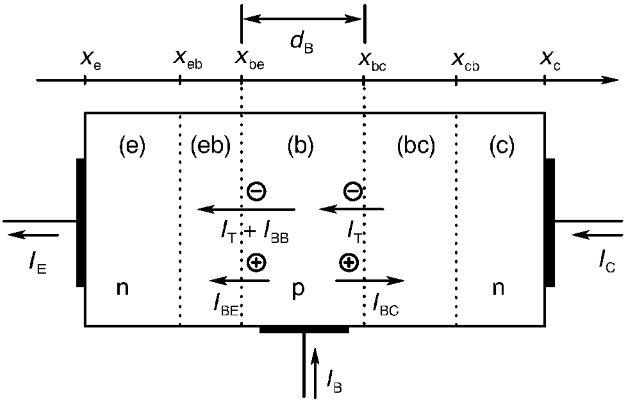


Fig. 1.40. One-dimensional transistor model

In forward operation ($V_{BC} < 0$) the current I_{BC} equals the reverse current of the bc diode and may generally be neglected. As long as the base width

$$d_B = x_{bc} - x_{be} \tag{1.92}$$

is small in comparison with the diffusion length for holes in the base region, we may neglect I_{BB}, and

$$I_B \approx I_{BE}$$

is determined by the current–voltage characteristic of the eb diode. The collector current I_C is composed of the transfer current I_T and the current I_{BC} of the bc diode,

$$I_C = I_T - I_{BC} .$$

Under forward operation $I_{BC} \approx 0$ holds, that is, the collector current approximately equals the transfer current,

$$I_C \approx I_T .$$

For an approximate calculation of the transfer current we make the following simplifying assumptions:

- The pn junctions are assumed to be abrupt; doping concentrations in the various regions are assumed to be constant.
- Series resistances are neglected.
- Generation and recombination in the base region and the space-charge layers are neglected.
- Low-level injection.
- One-dimensional current transport.

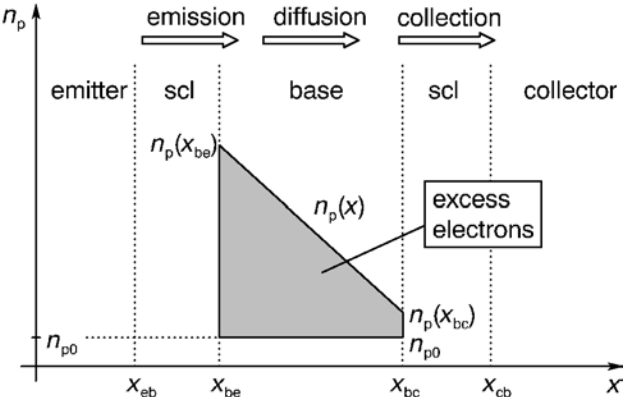


Fig. 1.41. minority-carrier distribution in a homogeneous base region

If there is no recombination in the base layer, the current density J_n of electrons traversing the base layer is independent of position. For the homogeneously doped base assumed, the transfer current will be due to diffusion:

$$J_\mathrm{n} = eD_\mathrm{n}\frac{\mathrm{d}n_\mathrm{p}}{\mathrm{d}x} = \text{const.}$$

Since e and D_n are constant, $\mathrm{d}n_\mathrm{p}/\mathrm{d}x$ must be constant, i.e., $n_\mathrm{p}(x)$ shows a linear dependence on x, and therefore

$$\frac{\mathrm{d}n_\mathrm{p}}{\mathrm{d}x} = \frac{n_\mathrm{p}(x_\mathrm{bc})-n_\mathrm{p}(x_\mathrm{be})}{x_\mathrm{bc}-x_\mathrm{be}}.$$

Under low-level injection conditions the values of $n_\mathrm{p}(x_\mathrm{be})$ and $n_\mathrm{p}(x_\mathrm{bc})$ are determined by the Shockley boundary conditions[32]:

$$n_\mathrm{p}(x_\mathrm{be}) = n_\mathrm{p0}\exp\left(\frac{V_\mathrm{BE}}{V_\mathrm{T}}\right) \quad \text{and} \quad n_\mathrm{p}(x_\mathrm{bc}) = n_\mathrm{p0}\exp\left(\frac{V_\mathrm{BC}}{V_\mathrm{T}}\right).$$

With $d_\mathrm{B} = x_\mathrm{bc} - x_\mathrm{be}$ and the emitter area A_je, the transfer current therefore is

$$I_\mathrm{T} = -A_\mathrm{je}J_\mathrm{n} = \frac{en_\mathrm{p0}D_\mathrm{n}A_\mathrm{je}}{d_\mathrm{B}}\left[\exp\left(\frac{V_\mathrm{BE}}{V_\mathrm{T}}\right) - \exp\left(\frac{V_\mathrm{BC}}{V_\mathrm{T}}\right)\right]. \tag{1.93}$$

The minus sign was introduced in order to obtain a positive value of the transfer current under forward operation. Substitution of n_p0 in (1.93) by n_i^2/p according to the mass-action law gives

$$I_\mathrm{T} = \frac{e^2A_\mathrm{jc}^2D_\mathrm{n}n_\mathrm{i}^2}{Q_\mathrm{B}}\left[\exp\left(\frac{V_\mathrm{BE}}{V_\mathrm{T}}\right) - \exp\left(\frac{V_\mathrm{BC}}{V_\mathrm{T}}\right)\right], \tag{1.94}$$

[32]These apply to forward-biased junctions under low-level-injection conditions but may lead to erroneous results if minority carriers are injected into a reverse-biased junction, owing to velocity saturation (see Sect. 3.2).

where the so-called base charge, $Q_B = eA_{je}pd_B$, is the charge of the holes in the base region. The value of Q_B is bias-dependent: for $V_{BE} \neq 0$ or $V_{BC} \neq 0$, its value will differ from the value Q_{B0} obtained for $V_{BE} = V_{BC} = 0$. Introducing the normalized base charge,

$$q_B = Q_B(V_{BE}, V_{BC})/Q_{B0} , \tag{1.95}$$

the transfer current becomes

$$I_T = \frac{I_S}{q_B}\left[\exp\left(\frac{V_{BE}}{V_T}\right) - \exp\left(\frac{V_{BC}}{V_T}\right)\right] = \frac{I_{CE} - I_{EC}}{q_B} , \tag{1.96}$$

where

$$I_S = \frac{eA_{je}D_n n_{p0}}{d_{B0}} = \frac{e^2 A_{je}^2 D_n n_i^2}{Q_{B0}} \tag{1.97}$$

is the transfer saturation current. The transfer current may therefore be written as the difference between the current

$$I_{CE} = I_S\left[\exp\left(\frac{V_{BE}}{V_T}\right) - 1\right] \tag{1.98}$$

controlled by the eb diode and the current

$$I_{EC} = I_S\left[\exp\left(\frac{V_{BC}}{V_T}\right) - 1\right] . \tag{1.99}$$

controlled by the bc diode. In forward operation $I_{EC} \approx 0$, while in reverse operation $I_{CE} \approx 0$. The normalized base charge q_B shows a complicated dependence on the applied voltages, which will be analyzed in more detail in Chap. 3.

Base Transit Time and Base Transport Factor

Assuming forward operation with $V_{BE} > 0$ and $V_{BC} = 0$ gives $n_p(x_{be}) \gg n_{p0}$ and $n_p(x_{bc}) = n_{p0}$. The magnitude of the excess charge due to the electrons traversing the base region is then as follows (Fig. 1.41):

$$|\Delta Q_{nB}| = eA_{je}d_B \, \frac{n_p(x_{be}) - n_{p0}}{2} = \tau_B \, I_{CE} , \tag{1.100}$$

where τ_B denotes the base transit time

$$\tau_B = d_B^2/2D_n . \tag{1.101}$$

This quantity is the average time needed by an electron to traverse the base region. For typical high-frequency bipolar transistors, the base transit time lies orders of magnitude below the minority-carrier lifetime in the base region ($\tau_B \ll \tau_n$); this justifies the assumption of negligible recombination within the base region. In general, the recombination of electrons traversing the base region can be taken into account by introduction of the base transport factor A_T, which defines the fraction of electrons injected into the base region that reach the bc junction. For a homogeneously doped base region, a solution of the diffusion equation yields, under forward operation conditions, the following [50]:

$$A_\mathrm{T} = \frac{J_\mathrm{n}(x_\mathrm{bc})}{J_\mathrm{n}(x_\mathrm{be})} = \left[\cosh\left(\frac{d_\mathrm{B}}{L_\mathrm{n}}\right)\right]^{-1} \approx 1 - \frac{d_\mathrm{B}^2}{2L_\mathrm{n}^2} = 1 - \frac{\tau_\mathrm{B}}{\tau_\mathrm{n}}, \tag{1.102}$$

i.e., the base transport factor is one to a good approximation, as long as the base transit time τ_B is small in comparison with the minority-carrier lifetime τ_n. Since electrons and holes recombine in pairs, the base current component I_BB due to recombination in the base region is

$$I_\mathrm{BB} = (1-A_\mathrm{T})I_\mathrm{CE} \tag{1.103}$$

under forward operation conditions. This current is attributed to the current of the eb diode in the model discussed in Sect. 1.6.

1.5.3 Current Gain

The realization of thin base regions requires shallow emitters, i.e., emitter regions with a thickness that is small in comparison with the minority-carrier diffusion length, if these regions are defined by implantation or diffusion processes. The hole current injected into the emitter region will then predominantly be due to recombination at the contact; for an approximate analysis, the recombination in the emitter volume may therefore be neglected.

Assuming an ideal metal contact at $x_\mathrm{e} = 0$, the excess hole density $\Delta p_\mathrm{n}(0)$ at the emitter contact will be zero, and

$$I_\mathrm{B} \approx I_\mathrm{BE} \approx \frac{eA_\mathrm{j}D_\mathrm{p}p_\mathrm{n0}}{d_\mathrm{E}} \exp\left(\frac{V_\mathrm{BE}}{V_\mathrm{T}}\right)$$

if $V_\mathrm{BE} \gg V_\mathrm{T}$. The corresponding relation for the transfer current in forward operation is

$$I_\mathrm{C} \approx \frac{eA_\mathrm{j}D_\mathrm{n}n_\mathrm{p0}}{d_\mathrm{B}} \exp\left(\frac{V_\mathrm{BE}}{V_\mathrm{T}}\right) .$$

Making use of $n_\mathrm{p0} = n_\mathrm{iB}^2/N_\mathrm{AB}$ and $p_\mathrm{n0} = n_\mathrm{iE}^2/N_\mathrm{DE}$ yields for the current gain

$$B_\mathrm{N} = \frac{I_\mathrm{C}}{I_\mathrm{B}} = \frac{D_\mathrm{n}}{D_\mathrm{p}}\frac{N_\mathrm{DE}}{N_\mathrm{AB}}\left(\frac{n_\mathrm{iB}}{n_\mathrm{iE}}\right)^2\frac{d_\mathrm{E}}{d_\mathrm{B}} . \tag{1.104}$$

Since $D_\mathrm{n} > D_\mathrm{p}$, the choice of $N_\mathrm{DE} \gg N_\mathrm{AB}$ allows values of the current gain $B_\mathrm{N} \gg 1$ to be realized. A reduction of the current gain is due to bandgap narrowing (see Sect. 2.7) in the emitter region; the reduced bandgap results

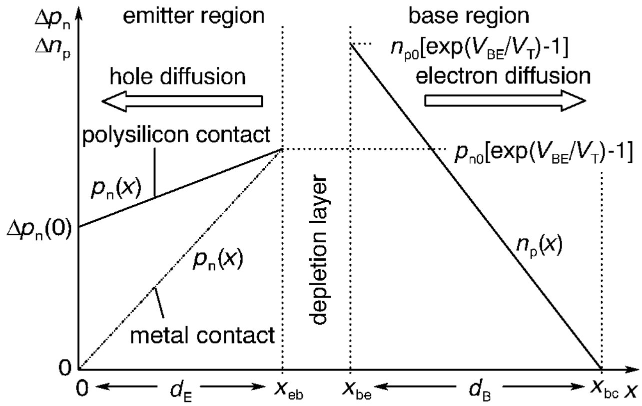


Fig. 1.42. Minority-carrier densities in a bipolar transistor with negligible recombination in the homogeneously doped base and emitter regions assuming low-level injection

in an intrinsic carrier concentration that is larger in the emitter region than in the base region ($n_{iE} > n_{iB}$).

The current gain decreases in proportion to the emitter depth d_E. Shallow emitters would therefore result in unacceptably small values of current gain. Since emitters with a direct metal contact might, in addition, impose a reliability problem due to metal spikes extending into the the monocrystalline silicon region, emitters with a polysilicon contact are used almost exclusively in modern high-frequency bipolar transistors. In such devices the monocrystalline emitter region is formed by diffusion of, for example, As out of the polysilicon contact. At comparable values of d_E, polysilicon emitters show a current gain that is several times larger than the current gain of corresponding metal-contacted devices [13]. The reason for this is the interface between poly- and monocrystalline silicon, which acts as a diffusion barrier to the holes, but shows only a finite surface recombination velocity S_p. The excess hole density $\Delta p_n(0)$ at the contact can be approximated by

$$\Delta p_n(0) \approx \frac{1}{eS_p}\frac{I_B}{A_{je}} \gg 0 .$$

The reduced slope of the minority-carrier distribution (1.42) then results in a smaller value of the diffusion current (see Sect. 1.4)

$$I_B = \frac{eA_{je}D_p p_{n0}}{d_E + D_p/S_p}\left[\exp\left(\frac{V_{BE}}{V_T}\right) - 1\right] ,$$

i.e., less holes diffuse into the emitter region and the current gain increases to

$$B_N = \frac{D_n}{D_p}\frac{N_{DE}}{N_{AB}}\left(\frac{n_{iB}}{n_{iE}}\right)^2 \frac{d_E + D_p/S_p}{d_B} . \tag{1.105}$$

For finite values of D_p/S_p, acceptable values of B_N may therefore be achieved even in the limit $d_E \to 0$.

1.5.4 Transistor Amplifiers and Switches

Bipolar transistors are three-terminal devices with output characteristics that are controlled by the input terminal. Figure 1.43 shows a bipolar transistor in

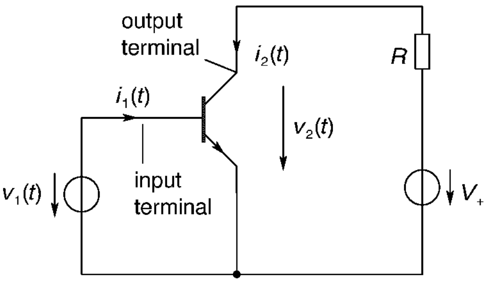


Fig. 1.43. Transistor applied for amplification or switching

common-emitter configuration, which could be applied for amplification and switching. The voltage source $v_1(t)$ connected to the input terminal controls the current $i_2(t)$ that flows into the output terminal, thus changing the voltage drop across R and therefore the output voltage $v_2(t)$. For low-frequency operation, the output voltage V_2 for a given input voltage V_1 is obtained graphically as the intercept of the load line with the output characteristic $I_2(V_2, V_1)$, as is illustrated in Fig. 1.44.

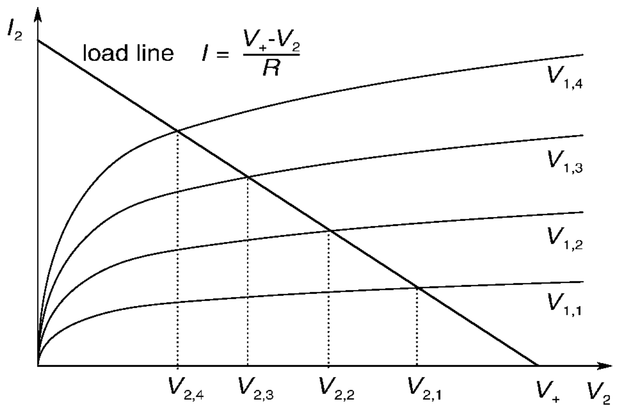


Fig. 1.44. Graphic determination of dc transfer characteristic

The (dc) voltage transfer characteristic is obtained from a plot of V_2 versus the corresponding values of V_1. Figure 1.45 schematically shows a typical voltage transfer characteristic. Small values of V_1 cause a small output current, and the output voltage is approximately equal to the supply voltage V_+. For input voltages in this range, the transistor can be considered as a switch in the OFF state. For large values of V_1 a large collector current that is limited by the load flows, and the output voltage drops to a small value, $V_{2,\text{on}}$. For large values of input voltage, the transistor may therefore be considered as a switch in the ON state.

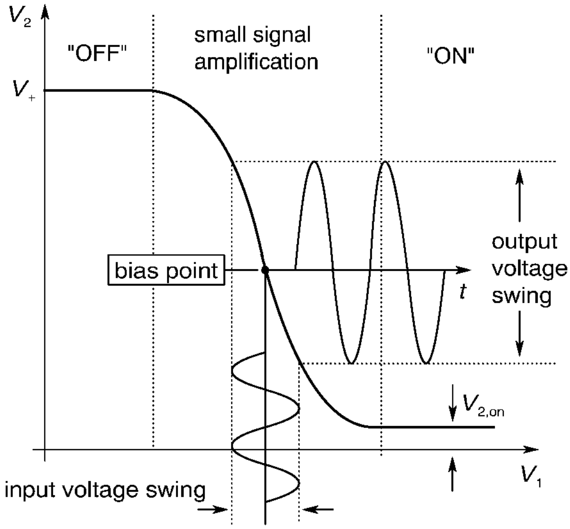


Fig. 1.45. DC voltage transfer characteristic of single transistor amplifier stage

If the transistor is used as a (small-signal) amplifier, a bias point (V_1, V_2) has to be chosen in the transition region between ON and OFF. Superposition of a small signal $\Delta v_1(t) = v_1(t) - V_1$ then gives, for the output voltage under low-frequency conditions,

$$v_2(t) = V_2 + \left.\frac{\mathrm{d}V_2}{\mathrm{d}V_1}\right|_{V_1} \Delta v_1(t) ,$$

where the transfer characteristic has been developed up to the first order in $\Delta v_1(t)$. The derivative $\mathrm{d}V_2/\mathrm{d}V_1$ equals the small-signal voltage transfer factor $\underline{H}_{\text{V0}} = \mathrm{d}V_2/\mathrm{d}V_1$, its magnitude is the voltage gain $A_{\text{V0}} = |\underline{H}_{\text{V0}}|$. The small-signal voltage transfer factor may be obtained from $\Delta V_2 = -R_\text{C}\Delta I_\text{C}$ and the total differential

$$\Delta I_\text{C} = \left(\frac{\partial I_\text{C}}{\partial V_\text{BE}}\right)_{V_\text{CE}} \Delta V_\text{BE} + \left(\frac{\partial I_\text{C}}{\partial V_\text{CE}}\right)_{V_\text{BE}} \Delta V_\text{CE} .$$

Since $\Delta V_\text{BE} = \Delta V_1$ and $\Delta V_\text{CE} = \Delta V_2$, the two equations may be combined to give

$$\underline{H}_{\text{V0}} = \frac{\Delta V_2}{\Delta V_1} = -R_\text{C} \frac{(\partial I_\text{C}/\partial V_\text{BE})_{V_\text{CE}}}{1 + R_\text{C}\,(\partial I_\text{C}/\partial V_\text{CE})_{V_\text{BE}}} .$$

Therefore, in order to obtain a large voltage gain, the transistor must have a large transconductance $\partial I_\text{C}/\partial V_\text{BE}$ and a small output conductance $\partial I_\text{C}/\partial V_\text{CE}$.

1.5.5 Leakage Currents

If a reverse bias is applied between any two terminals of a bipolar transistor, generally a small leakage current will be observed. Figure 1.46 illustrates the leakage current values which are generally specified and how they are measured.

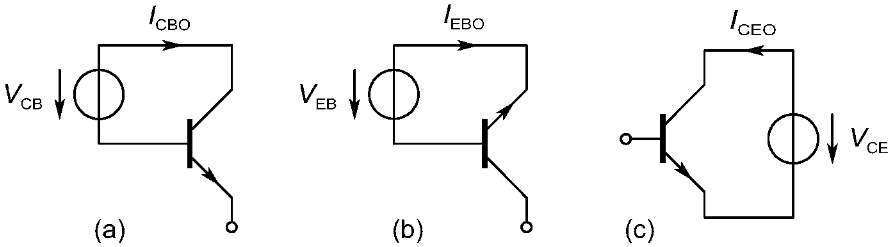


Fig. 1.46. Leakage currents in bipolar transistors. (**a**) Collector–base leakage current I_{CBO}, (**b**) emitter–base leakage current I_{EBO}, (**c**) collector–emitter leakage current I_{CEO}

The current I_{CBO} flows between collector and base if the emitter is left open. This current is the reverse current of the bc diode, i.e., in the elementary transistor model (see Sect. 1.6) $I_{CBO} = I_S/B_R$. The three characters CBO in the subscript indicate that a positive voltage has been applied between the terminals defined by the first (here C) and second character (here B), while the third terminal is left open (O).

The current I_{EBO} therefore defines the reverse current of the eb diode, measured with the collector left open. In the elementary transistor model (see Sect. 1.6), therefore, $I_{EBO} = I_S/B_F$. For the current flow between collector and base different values of leakage current are specified, dependent on the biasing of the base terminal.

The value of I_{CEO} is determined for the base left open. The primary source of this current is the reverse current of the bc diode. The electrons generated in the process flow to the collector terminal, while the electric field of the bc diode transports the holes to the base region. This situation corresponds to a base current that drives a transfer current $B_N I_{CBO}$, which flows in addition to the "primary" generation current I_{CBO}, such that $I_{CEO} = (B_N+1)I_{CBO}$.

The value of I_{CES} is determined as the collector current for a short circuit between emitter and base. For transistors that do not show punchthrough, one observes $I_{CES} \approx I_{CBO}$ (see p. 66 and Sect. 3.4). Besides these, sometimes leakage currents I_{CER} (measured with a shunt resistor between emitter and base terminal) and I_{CEV} (measured with a reverse-biased eb diode) are specified. All leakage currents show a strong dependence on temperature and generally are proportional to n_i^γ, where γ typically has a value between one and two.

1.5.6 Voltage Limits, Breakdown

The voltage that may be applied between any two terminals of the transistor is limited by the corresponding breakdown voltage of the junction or the maximum allowed current under forward bias.

Collector–Base Breakdown. Avalanche breakdown of the bc diode defines the maximum reverse voltage that may be applied between collector and base with the emitter left open. The corresponding breakdown voltage, BV_{CBO}, decreases with increasing values of collector doping and has values of the order of 10 V in integrated high-frequency bipolar transistors.

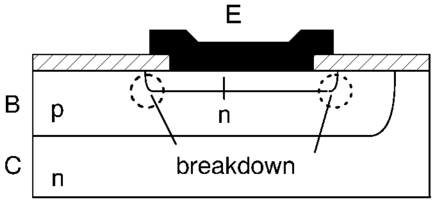


Fig. 1.47. Breakdown of the eb diode occurs generally in the sidewall diode

Emitter–Base Breakdown. The breakdown voltage BV_{EBO} of the eb diode is determined by the doping of the eb diode and decreases with increasing values of base doping. Integrated high-frequency bipolar transistors have values of BV_{EBO} in the range of a few volts, determined by internal field emission (Zener effect). This breakdown occurs typically in the sidewall diode (Fig. 1.47), since there the electric-field strength shows a maximum.

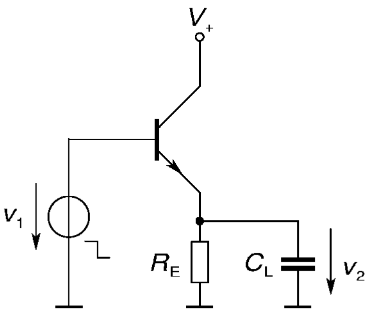


Fig. 1.48. Emitter follower with capacitive load

In practical applications one tries to avoid reverse-biased eb junctions. During switching transients, such a reverse biasing can, however, not generally be excluded, as is shown by the circuit example depicted in Fig. 1.48. Assume C_{L} to be charged for $t < 0$ to $v_1(0^-)-V_{\mathrm{BEon}}$; if v_1 decreases at $t = 0$ to $v_1(0^+)$, the instant reverse bias at the eb diode is $v_1(0^-) - v_1(0^+) - V_{\mathrm{BEon}}$, since the voltage v_2 remains constant during the switching transient. If this value is beyond BV_{EBO}, breakdown of the eb diode occurs and the capacitor discharges across the reverse-biased eb diode. Generally, this will lead to an increase in recombination centers, which cause a parasitic base current component and therefore a degradation of the current gain. If the power dissipated during the discharge exceeds a critical limit, the transistor will be destroyed.

Collector–Emitter Breakdown. The open-base breakdown voltage BV_{CEO} defines the maximum allowed voltage V_{CE} if the base is driven with a constant current; for $V_{\mathrm{CE}} > BV_{\mathrm{CEO}}$ the current in the output port can no longer be controlled by a current in the input port. Open-base breakdown can be caused by carrier multiplication due to impact ionization in the bc diode or due to the punchthrough effect.

Avalanche-induced breakdown is caused by a positive-feedback effect. For small values of V_{CE} under open-base conditions, one observes the leakage current $I_{\mathrm{CEO}} \approx (B_{\mathrm{N}} + 1) I_{\mathrm{CBO}}$, which is comparatively small. With increasing values of V_{CE}, the electric-field strength in the bc diode grows and carrier multiplication increases the current by a factor M_{n}. Since the holes generated in the avalanche process act as a base current, the resulting collector current is

$$I_{\mathrm{C}} = M_{\mathrm{n}} \left[\frac{B_{\mathrm{N}}'}{1 - B_{\mathrm{N}}'(M_{\mathrm{n}} - 1)} + 1 \right] I_{\mathrm{CBO}} ,$$

where $B_{\mathrm{N}}' \approx I_{\mathrm{BE}}/I_{\mathrm{T}}$ is the current gain that would be observed if carrier multiplication were negligible. This expression shows a divergence for $B_{\mathrm{N}}'(M_{\mathrm{n}} - 1) = 1$ corresponding to collector–emitter breakdown. If M_{n} obeys Miller's formula,

$$M_{\mathrm{n}} = \frac{1}{1 - (V_{\mathrm{CB}}/BV_{\mathrm{CBO}})^n} , \qquad \text{and thus} \quad 1 - \frac{1}{M_{\mathrm{n}}} = \left(\frac{V_{\mathrm{CB}}}{BV_{\mathrm{CBO}}} \right)^n ,$$

the breakdown condition $B_{\mathrm{N}}'(M_{\mathrm{n}} - 1) = 1$ gives

$$V_{\mathrm{CEO}} = BV_{\mathrm{CBO}} (B_{\mathrm{N}}' + 1)^{-1/n} + V_{\mathrm{BE}} . \tag{1.106}$$

The value of V_{CEO} is bias-dependent (see Sect. 3.4) and decreases with increasing values of current gain. The open-base breakdown voltage BV_{CEO} has to be defined as the minimum value of $V_{\mathrm{CEO}}(I_{\mathrm{C}})$ and is measured by forcing a current $I_{\mathrm{C}} = I_{\mathrm{E}}$ with open base.

If the collector–emitter breakdown occurs due to punchthrough, the base layer is so thin that the bc space-charge layer reaches through to the eb space-charge layer. Thermal emission of carriers across the reduced potential barrier then causes a significant increase in current. Generally, the base width is chosen in order to avoid punchthrough.

1.5.7 Some Differences of Bipolar Transistors and MOSFETs

Figure 1.49 points out some important differences between bipolar transistors and MOSFETs. • The input conductance of a MOSFET is negligibly small

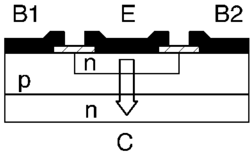
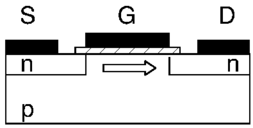
device	bipolar transistor	n-channel MOSFET
cross section (schematically)		
current flow	vertical	lateral
critical dimension for frequency limit	base width	channel length
noise (main causes)	shot noise thermal noise	thermal noise $1/f$-noise
current-voltage characteristics	exponential	quadratic (in saturation)
transconductance	$g_m = I_C/V_T$	$g_m = \sqrt{2\beta_n I_D}$

Fig. 1.49. A comparison of an npn bipolar transistor and an n-channel MOSFET

under NF operation – in contrast to a bipolar transistor, no bias current is needed at the input terminal. MOSFETs therefore allow one to use charge as a signal quantity – in addition to current and voltage, which can also be used in bipolar transistors. This enables MOS circuit concepts, such as dynamic logic, that have no counterpart in the bipolar world.

• Bipolar transistors, which are operated as switches, are generally driven into saturation and build up a significant minority charge in the collector region that slows down turn-off. Since MOSFETs do not suffer from minority-carrier storage, they generally are better suited for fast switches that are operated rail-to-rail.

• The exponential current–voltage characteristic of the bipolar transistor causes a comparatively small dependence of the threshold voltage

$$V_{\mathrm{BEon}} \approx V_{\mathrm{T}} \ln\left(\frac{I_{\mathrm{C}}}{I_{\mathrm{S}}}\right) \approx V_{\mathrm{T}} \ln\left(\frac{I_{\mathrm{C}} N_{\mathrm{A}} d_{\mathrm{B0}}}{e A_{\mathrm{je}} D_{\mathrm{n}} n_{\mathrm{i}}^2}\right)$$

on variations in geometry and doping concentration. A 10 % change in emitter area, for example, causes $V_{\mathrm{BEon}} \approx 0.8$ V to change by $\Delta V_{\mathrm{BEon}} \approx$

$V_T \ln(1.1) \approx 2.5\,\text{mV}$ at room temperature. The corresponding relative change is $\Delta V_{\text{BEon}}/V_{\text{BEon}} \approx 0.3\,\%$. MOSFETs with short channel length generally show considerably larger deviations: a 10 % change in channel length might cause a 10 % change in threshold voltage under unfavorable conditions. Differential amplifiers built with bipolar transistors therefore generally show offset voltages that lie significantly below those observed in MOSFET realizations. A further advantage of the bipolar transistor is the long-time stability of the threshold voltage – only small changes are observed, in contrast to threshold voltage shifts of degraded MOSFETs.

• Owing to their exponential transfer characteristic $I_C(V_{BE})$, bipolar transistors have generally a larger transconductance than MOSFETs, with their approximately linear or square-law transfer characteristic $I_D(V_{GS})$. In addition to this, the transconductance g_m will be strictly proportional to the collector current over several decades of I_C, a phenomenon that may be exploited in translinear circuits [53].

• The smaller curvature of the MOSFET transfer current characteristic generally causes smaller nonlinear distortion under large-signal operation.

• Current flow is generally vertical in bipolar transistors and lateral in MOSFETs. The critical dimensions determining device speed are base width and channel length, respectively. In bipolar transistors the device speed is determined by diffusion, implantation or deposition, while in integrated MOSFETs it is usually determined by a lithographic process. For this reason, high-speed bipolar transistors can be realized with a relaxed lithography.

• An increase in temperature generally causes an increase in the current carried by a bipolar transistor, while the current carried by a MOSFET decreases. Parallel operation of bipolar transistors can therefore be problematic, since the current distribution among the various transistors may become unstable.

• In contrast to majority-carrier devices such as the MOSFET, bipolar transistors show shot noise associated with the transport of carriers across the junction. Furthermore, in bipolar transistors a thermal noise current component associated with the base current exists that is not found in field-effect transistors, which generally show negligible gate noise. This principal advantage of MOSFETs is, however, alleviated by $1/f$ noise with a noise-corner frequency that usually lies significantly beyond values observed in bipolar transistors.

• In comparison with CMOS processes, a bipolar process is more sensitive to defects. The yield of highly integrated bipolar circuits will therefore generally be less than the yield obtained with a CMOS circuit of comparable complexity.

1.6 Elementary Large-Signal Models

In this section a subset of the Gummel–Poon model, which considers the most important effects, is presented and applied to the description of the current–voltage characteristics and switching transients.

1.6.1 The Elementary Transistor Model

A simple equivalent circuit[33] to describe the transistor characteristics is shown in Fig. 1.50. The model employs two ideal diodes ($N = 1$), to describe the

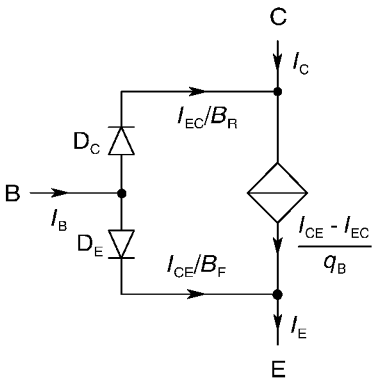


Fig. 1.50. Elementary transistor model

emitter–base (eb) and base–collector (bc) junctions, together with a voltage-controlled current source, which describes the coupling of the two pn junctions, i.e., the transfer current. Introducing the ideal forward current gain B_F, the current carried by the eb diode is written as

$$I_{DE} = \frac{I_{CE}}{B_F} = \frac{I_S}{B_F}\left[\exp\left(\frac{V_{BE}}{V_T}\right) - 1\right] , \tag{1.107}$$

whereas the current carried by the bc diode is written as

$$I_{DC} = \frac{I_{EC}}{B_R} = \frac{I_S}{B_R}\left[\exp\left(\frac{V_{BC}}{V_T}\right) - 1\right] , \tag{1.108}$$

with the ideal reverse current gain B_R.

In accordance with (1.96), the transfer current is expressed as the difference of a current I_{CE} controlled by V_{BE} and a current I_{EC} controlled by V_{BC}, divided by the normalized base charge q_B:

$$I_T = \frac{I_{CE} - I_{EC}}{q_B} . \tag{1.109}$$

[33]This equivalent circuit is sometimes called the nonlinear hybrid-π model [54].

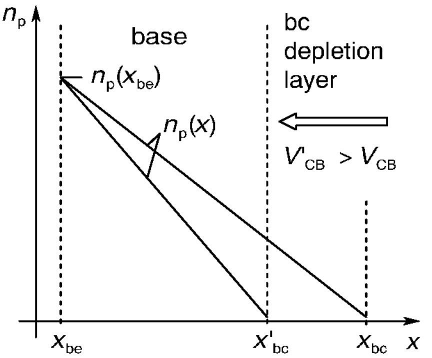


Fig. 1.51. Explanation of the Early effect, considering the bias-dependent minority-carrier distribution in the base region

Under low-level-injection conditions q_B takes account of the bias-dependent base width d_B. This so-called Early effect is roughly approximated by

$$\frac{1}{q_B} \approx 1 + \frac{V_{CB}}{V_{AF}} - \frac{V_{BE}}{V_{AR}} , \tag{1.110}$$

where V_{AF} is termed the forward Early voltage and V_{AR} the reverse Early voltage. The bias-dependent normalized base charge q_B takes account of the Early effect [55]; this is explained in Fig. 1.51, where the effect of the bias-dependent extension of the bc space-charge layer is shown: increasing V_{CB} causes a shift of x_{bc} to the left and therefore reduces the base width d_B. Since $n_p(x_{be}) = \text{const.}$ for $V_{BE} = \text{const.}$ and $n_p(x_{bc}) \approx 0$, this causes the slope of the minority-carrier distribution to increase in magnitude with the consequence of an increased transfer current. For basic calculations we will generally neglect the reverse Early effect, which considers the effect of the bias-dependent extension of the eb space-charge layer, and approximate (1.110) by

$$\frac{1}{q_B} \approx 1 + \frac{V_{CE}}{V_{AF}} . \tag{1.111}$$

The transfer current source then has the following current–voltage characteristic:

$$I_T \approx I_S \left(1 + \frac{V_{CE}}{V_{AF}}\right) \left[\exp\left(\frac{V_{BE}}{V_T}\right) - \exp\left(\frac{V_{BC}}{V_T}\right)\right] . \tag{1.112}$$

Elementary Ebers–Moll Model

Figure 1.52 shows the so-called Ebers–Moll model of the bipolar transistor, that has been widely used in the literature. Assuming $q_B = 1$ for simplicity, four parameters are needed to describe the elements of the Ebers–Moll model: the ideal forward current gain A_F in common-base configuration, the ideal reverse current gain A_R in common-base configuration, and the saturation

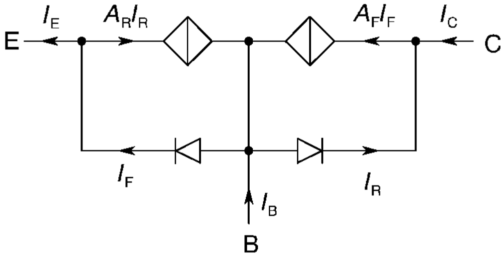


Fig. 1.52. Ebers–Moll model

currents I_{ES} and I_{CS} of the eb diode and the bc diode respectively. The currents I_F and I_R are modeled as

$$I_F = I_{ES}\left[\exp\left(\frac{V_{BE}}{V_T}\right) - 1\right] \quad \text{and} \quad I_R = I_{CS}\left[\exp\left(\frac{V_{BC}}{V_T}\right) - 1\right] .$$

I_F is the current that would flow across the eb junction if the collector-sided depletion layer were replaced by an ohmic contact. Correspondingly, I_R describes the current that would flow across the bc junction with an ohmic contact placed instead of the emitter-sided depletion layer. Since the Ebers–Moll model must yield the same current–voltage characteristics as the elementary transistor model if the Early effect is neglected ($V_{AF} \to \infty$), the current gains are related by

$$A_F = \frac{B_F}{B_F+1} \quad \text{and} \quad A_R = \frac{B_R}{B_R+1} ,$$

while the saturation currents have to obey

$$I_{ES} = I_S\left(1+\frac{1}{B_F}\right) = \frac{I_S}{A_F} \quad \text{and} \quad I_{CS} = I_S\left(1+\frac{1}{B_R}\right) = \frac{I_S}{A_R} .$$

The elementary transistor model uses three parameters, I_S, B_F, and B_R, while four parameters, I_{ES}, I_{CS}, A_F, and A_R, are used in the Ebers–Moll model. Since the models are equivalent, only three parameters of the Ebers–Moll model can be independent. As is easily verified, the reciprocity relation

$$A_F\, I_{ES} = A_R\, I_{CS} \tag{1.113}$$

holds, which allows the fourth parameter to be calculated as soon as three parameters are known. The Ebers Moll model [26] was developed first, but has been replaced by the equivalent circuit shown in Fig. 1.50, which is closely related to the widely employed Gummel–Poon model. The reason for this is that the parameter I_S, in contrast to I_{ES} and I_{CS}, is not influenced by a bias-dependent current gain – nonideal effects of the current gain are modeled in terms of the eb and bc diodes [54]. Furthermore, the parameters of the Gummel–Poon model are better suited for a description of bipolar transistors in the common-emitter configuration.

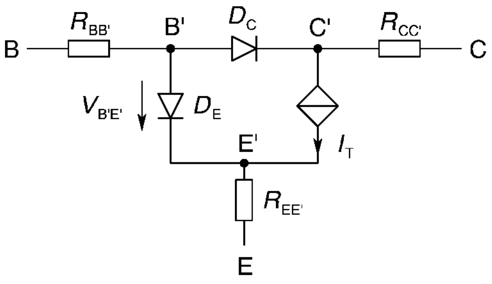


Fig. 1.53. Elementary bipolar transistor model with series resistances

Series Resistances

The elementary transistor model is easily extended by the addition of series resistances to give the equivalent circuit shown in Fig. 1.53. In general, the values of these series resistances are bias dependent.

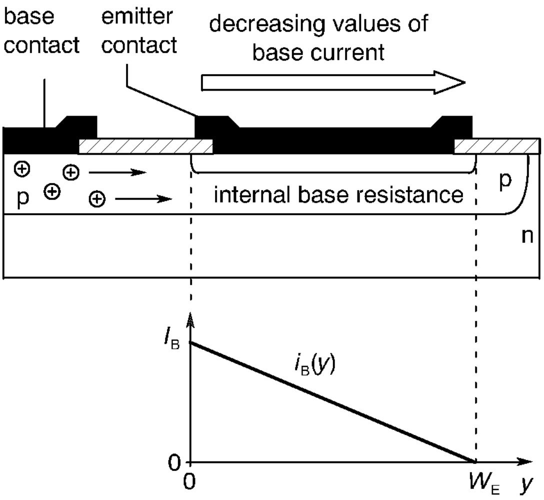


Fig. 1.54. Geometry considered for the estimation of the internal base resistance

For a simple estimate of the internal base resistance, uniform injection of hole current into the emitter will be assumed. The hole current in the base layer will then show a linear decrease with position y (Fig. 1.54), and in a narrow stripe of thickness d_B, length L_E and width dy at y, the power

$$\mathrm{d}P = \frac{\rho}{d_B L_E} i_B^2(y)\,\mathrm{d}y = \frac{\rho I_B^2}{d_B L_E}\left(1 - \frac{y}{W_E}\right)^2 \mathrm{d}y$$

will be dissipated, where ρ is the resistivity of the base layer. From this the total power P dissipated in the base layer is calculated by integration,

$$P = \frac{\rho I_B^2}{d_B L_E}\int_0^{W_E}\left(1 - \frac{y}{W_E}\right)^2 \mathrm{d}y\,.$$

With a base resistance $R_{\mathrm{BB'}}$ defined by the identity[34] $P = R_{\mathrm{BB'}} I_{\mathrm{B}}^2$, the base resistance is estimated to be

$$R_{\mathrm{BB'}} = \frac{\rho}{3} \frac{W_{\mathrm{E}}}{d_{\mathrm{B}} L_{\mathrm{E}}} . \tag{1.114}$$

A transistor with two base contacts corresponds to a parallel connection of two transistors with emitter stripe width $W_{\mathrm{E}}/2$, each driven with base current $I_{\mathrm{B}}/2$. The power consideration then yields

$$R_{\mathrm{BB'}} = \frac{\rho}{12} \frac{W_{\mathrm{E}}}{d_{\mathrm{B}} L_{\mathrm{E}}} . \tag{1.115}$$

1.6.2 Current–Voltage Characteristics

Input and Transfer Characteristics. The input characteristic $I_{\mathrm{B}}(V_{\mathrm{BE}})$ of the bipolar transistor gives the current in the input terminal as a function of the input voltage, while the transfer characteristic $I_{\mathrm{C}}(V_{\mathrm{BE}})$ gives the current in the output terminal as a function of the input voltage. Under forward-bias conditions, the elementary transistor model yields

$$I_{\mathrm{C}} \approx I_{\mathrm{CE}}\,(1+V_{\mathrm{CE}}/V_{\mathrm{AF}}) \qquad \text{and} \qquad I_{\mathrm{B}} = I_{\mathrm{CE}}/B_{\mathrm{F}} ,$$

i.e., both input and transfer currents show an exponential dependence on V_{BE} as $I_{\mathrm{CE}} \sim \exp(V_{\mathrm{BE}}/V_{\mathrm{T}})$. For device characterization, input and transfer characteristics generally are determined for $V_{\mathrm{BC}} = 0$ and plotted semilogarithmically as a so-called Gummel plot. In the elementary model, the forward current gain $B_{\mathrm{N}} = I_{\mathrm{C}}/I_{\mathrm{B}}$ is given by

$$B_{\mathrm{N}} = B_{\mathrm{F}}\,(1+V_{\mathrm{CE}}/V_{\mathrm{AF}}) ; \tag{1.116}$$

its value increases with increasing values of V_{CE}.

Output Characteristics. In a common-emitter configuration the dependence of I_{B} and I_{C} on V_{BE} and V_{CE} is of interest. Neglecting series resistances, one obtains from (1.96), using $V_{\mathrm{BC}} = V_{\mathrm{BE}} - V_{\mathrm{CE}}$, for the transfer current the following

$$I_{\mathrm{T}} = \frac{I_{\mathrm{S}}}{q_{\mathrm{B}}} \exp\left(\frac{V_{\mathrm{BE}}}{V_{\mathrm{T}}}\right) \left[1 - \exp\left(-\frac{V_{\mathrm{CE}}}{V_{\mathrm{T}}}\right)\right] .$$

For $V_{\mathrm{BE}} \gg V_{\mathrm{T}}$ (resp. $I_{\mathrm{CE}} \gg I_{\mathrm{S}}$), the collector current $I_{\mathrm{C}} = I_{\mathrm{T}} - I_{\mathrm{DC}}$ is then

$$I_{\mathrm{C}} = \frac{I_{\mathrm{S}}}{q_{\mathrm{B}}} \exp\left(\frac{V_{\mathrm{BE}}}{V_{\mathrm{T}}}\right) \left[1 - \left(1+\frac{q_{\mathrm{B}}}{B_{\mathrm{R}}}\right) \exp\left(-\frac{V_{\mathrm{CE}}}{V_{\mathrm{T}}}\right)\right] . \tag{1.117}$$

[34] There are alternative definitions, as explained in further detail in Sect. 3.6.

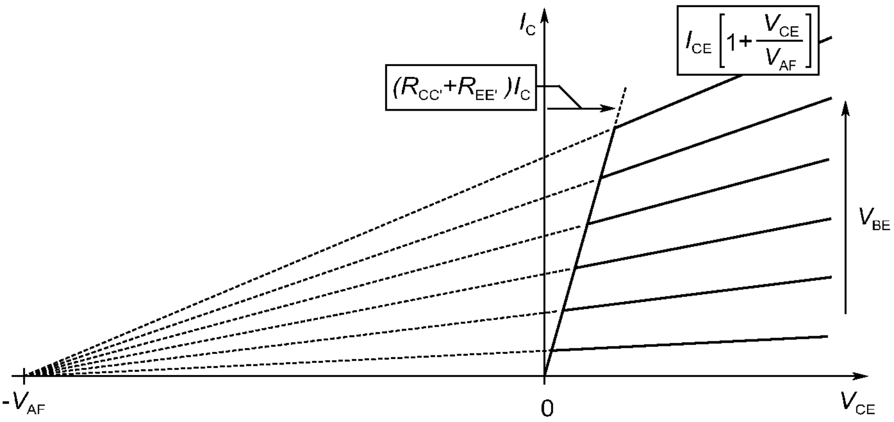


Fig. 1.55. Linear approximation of the output characteristics, Early voltage

Plotting I_C versus V_{CE} for V_{BE} = const. shows a strong increase of I_C within a voltage interval of a few V_T towards:

$$\frac{I_S}{q_B}\exp\left(\frac{V_{BE}}{V_T}\right) \approx I_S\left(1+\frac{V_{CE}}{V_{AF}}\right)\exp\left(\frac{V_{BE}}{V_T}\right) . \tag{1.118}$$

For $V_{CE} > V_{BE}$, (1.118) describes a linear dependence on V_{CE}; for $V_{CE} = -V_{AF}$, I_C would vanish according to (1.118), i.e., extrapolating the output characteristics to $I_C = 0$ should lead to a set of lines that intersect at the same point, $V_{CE} \approx -V_{AF}$ (Fig. 1.55). The base current is described by

$$I_B = I_{DE} + I_{DC} = \frac{I_{CE}}{B_F} + \frac{I_{EC}}{B_R} .$$

If $V_{CE} \gg V_T$, the current I_{DC} carried by the bc diode is small in comparison with the current I_{DE} carried by the eb diode, i.e.,

$$I_C \approx \frac{I_{CE}}{q_B} = \frac{B_F I_{DE}}{q_B} \approx B_F\left(1+\frac{V_{CE}}{V_{AF}}\right) I_B = B_N I_B ,$$

resulting in a current gain that increases with V_{CE}. For $V_{CE} \to 0$, the collector current is

$$I_C = -\frac{I_{CE}}{B_R} = -\frac{B_F}{B_F+B_R} I_B .$$

This corresponds to a parallel connection of eb and bc diode; I_{CE} and I_{EC} show identical values, i.e., there is no transfer current. The base current is composed of the contribution I_{DE} and the contribution I_{DC}, with a ratio determined by the ideal values of forward and reverse current gains B_F and B_R; the collector current is $-I_{DC}$.

If the voltage drop across $R_{\mathrm{EE'}}$ and $R_{\mathrm{CC'}}$ is considered, the value of V_{CE} in (1.117) has to be replaced by

$$V_{\mathrm{C'E'}} = V_{\mathrm{CE}} - R_{\mathrm{EE'}} I_{\mathrm{E}} - R_{\mathrm{CC'}} I_{\mathrm{C}} \approx V_{\mathrm{CE}} - (R_{\mathrm{EE'}} + R_{\mathrm{CC'}}) I_{\mathrm{C}} \,.$$

This corresponds to a voltage shift of the output characteristic to the right by $(R_{\mathrm{EE'}} + R_{\mathrm{CC'}}) I_{\mathrm{C}}$, as is illustrated in Fig. 1.55.

Saturation, Quasi-Saturation. For $V_{\mathrm{CE}} < V_{\mathrm{BE}}$, the bc diode is forward biased and the collector current is composed of the transfer current and the current carried by the bc diode. Under such saturation conditions, a reduced current gain $I_{\mathrm{C}}/I_{\mathrm{B}}$ is observed. If $I_{\mathrm{B}} \gg I_{\mathrm{C}}/B_{\mathrm{F}}$, the value of $V_{\mathrm{C'E'}}$ will be of the order of millivolts. The voltage drop V_{CE} across the transistor will then be

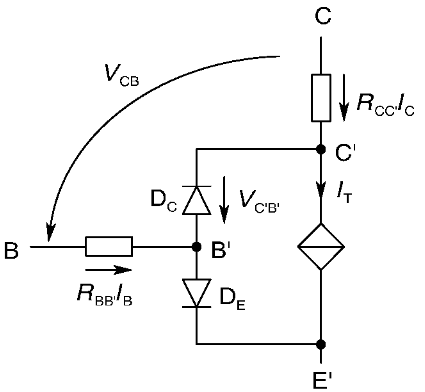


Fig. 1.56. Quasi-saturation. Internally forward-biased bc-diode due to voltage drops across the base and collector series resistances

determined to a large extent by the voltage drop across the series resistances. Due to the voltage drop across the series resistances (Fig 1.56),

$$V_{\mathrm{C'B'}} = V_{\mathrm{CB}} - R_{\mathrm{CC'}} I_{\mathrm{C}} + R_{\mathrm{BB'}} I_{\mathrm{B}}$$

may become negative, even for $V_{\mathrm{CB}} > 0$. This is known as quasi-saturation. In quasi-saturation minority carriers are injected into the collector region, with the consequence of an increased base current and a slow down of the switching operation.

1.6.3 Charge Storage, Charge Control Model

The minority charge stored in the bipolar transistor is divided into the portion $q_{\mathrm{TE}} \sim i_{\mathrm{CE}}$ associated with the eb diode and the portion $q_{\mathrm{TC}} \sim i_{\mathrm{EC}}$ associated with the bc diode. In charge-control theory a quasi-static approach is used for the description of the charges q_{TE} and q_{TC}, i.e., the relationships between stored charge and current derived for dc operation are extended to transient

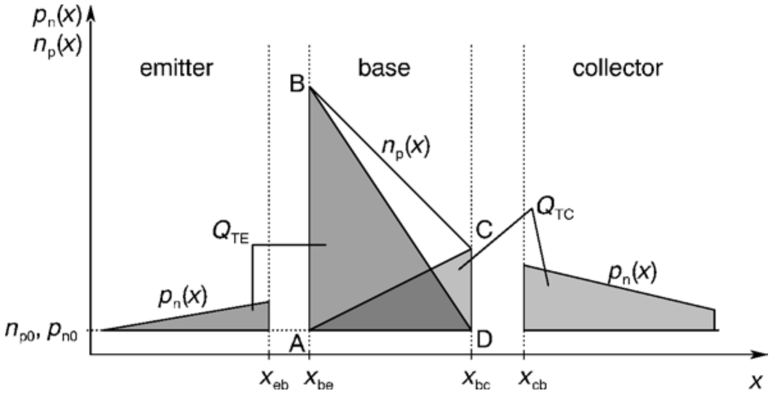


Fig. 1.57. Minority-carrier storage in an npn bipolar transistor (schematic)

operation. Figure 1.57 explains how the minority charge is divided into two contributions. Minority charge in the emitter stems from hole injection into the forward-biased eb diode ($V_{B'E'} > 0$) and is associated with the eb diode. Correspondingly, minority charge in the collector is associated with the bc diode and occurs if holes are injected into the collector region under forward-bias conditions ($V_{B'C'} > 0$). The minority charge stored in the base region is proportional to the trapezoid ABCD, which can be divided[35] into a portion (ABD) that is attributed to the eb diode and is proportional to the stored minority charge in the case $V_{B'E'} > 0$ and $V_{B'C'} = 0$, and a portion (ACD) which is attributed to the bc diode and proportional to the stored minority charge if $V_{B'C'} > 0$ and $V_{B'E'} = 0$.

The charges q_{TE} and q_{TC} are expressed in terms of the forward transit time τ_f and the reverse transit time τ_r as follows:

$$q_{TE} = \tau_f \, i_{CE}/q_B \quad \text{and} \quad q_{TC} = \tau_r \, i_{EC}/q_B \, . \tag{1.119}$$

In general, τ_f and τ_r have to be considered as bias-dependent quantities. In this introductory survey τ_f and τ_r will be set equal to the ideal forward transit time T_F and the ideal reverse transit time T_R.

Figure 1.58 shows the elementary transistor model, extended with series resistances and diffusion and depletion capacitances. This model can only approximately describe the time-dependent behavior of the transistor, since effects due the base and collector series resistances and minority-charge storage are taken into account in a highly simplified manner. Despite this, the model allows to explain the principal behavior of a bipolar transistor in switching operation.

[35] The sum of the areas of triangles ABD and ACD equals the area of the trapezoid, since triangle BCD has the same area as triangle ACD.

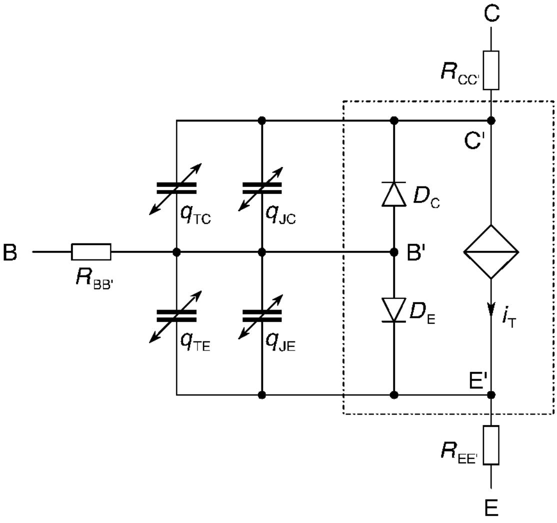


Fig. 1.58. Elementary transistor model (within the box) extended with series resistances and depletion and diffusion capacitances

1.6.4 Switching Operation

A bipolar transistor in a common-emitter configuration, applied in series to a load device, may be operated as a fast switch if a large voltage swing is applied to the input. Here, the saturation voltage that drops across the closed switch, and the delay times associated with turn-on and turn-off of the switch, is of particular interest.

Saturation Voltage

Figure 1.59 shows an equivalent circuit[36] for the calculation of the voltage drop across a bipolar transistor switch. In the ON state, the transistor will be driven into saturation. The saturation voltage V_{CEsat} that remains across the device is determined to a large extent by the emitter and collector series resistance; the corresponding voltage drop across the eb diode is given by V_{BEsat}, where generally $V_{\mathrm{BEsat}} > V_{\mathrm{CEsat}}$. Assuming $q_{\mathrm{B}} = 1$, the equivalent circuit yields

$$I_{\mathrm{B}} = I_{\mathrm{CE}}/B_{\mathrm{F}} + I_{\mathrm{EC}}/B_{\mathrm{R}} \quad \text{and} \quad I_{\mathrm{C}} = I_{\mathrm{CE}} - I_{\mathrm{EC}}\,(1+1/B_{\mathrm{R}}) \ ,$$

i.e., the current that flows into the bc diode must be taken into account. Since both I_{CE} and I_{EC} are large in comparison with the transfer saturation current I_{S}, these equations give

[36]Base-charge modulation and nonideal behavior of the pn junction has been neglected for simplicity.

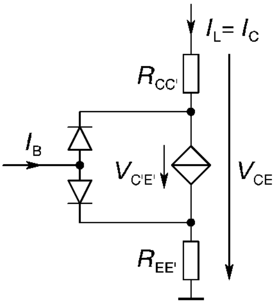


Fig. 1.59. Equivalent circuit for the calculation of the saturation voltage of the bipolar transistor switch

$$V_{C'E'} = V_T \ln\left(\frac{B_F}{B_R}\frac{B+B_R+1}{B_F-B}\right) , \tag{1.120}$$

where $B = I_C/I_B$. The saturation voltage V_{CEsat} is the sum of $V_{C'E'}$ and the voltage drops across the emitter and collector resistances $R_{EE'}$ and $R_{CC'}$:

$$V_{CE} \approx V_{C'E'} + (R_{CC'}+R_{EE'})\, I_C \, .$$

Turn-on and Turn-off Delay Times

Figure 1.60 shows a bipolar transistor switch with an ohmic load R_C and a control current $i_B(t) = [v_1(t) - v_{BE}(t)]/R_B$ delivered by the time-dependent voltage source $v_1(t)$. Together with the relations of the charge-control theory (Fig. 1.60)

$$\begin{aligned} i_B(t) &= \frac{i_{CE}}{B_F} + \frac{i_{EC}}{B_R} + c_{je}\frac{dv_{BE}}{dt} + c_{jc}\frac{dv_{BC}}{dt} \\ &\quad + T_F\frac{di_{CE}}{dt} + T_R\frac{di_{EC}}{dt} \end{aligned} \tag{1.121}$$

$$i_C(t) = i_{CE} - i_{EC}\left(1+\frac{1}{B_R}\right) - c_{jc}\frac{dv_{BC}}{dt} - T_R\frac{di_{EC}}{dt} , \tag{1.122}$$

the turn-on and turn-off switching transients may be calculated [56–58].

Turn-on. Assume a voltage step of v_1 at $t = 0$ from zero to V_1. The value of v_{BE} will first be smaller than the threshold V_{BEon} of the eb diode, resulting in negligible currents i_{CE} and i_{EC}. Since $v_{CE} \approx V_+$, (1.121) yields

$$i_B(t) = (c_{je}+c_{jc})\frac{dv_{BE}}{dt} = \frac{V_1 - v_{BE}}{R_B} .$$

If v_{BE} increases from zero to V_{BEon}, the charges on the eb and bc depletion capacitances change by

$$\Delta Q_{J1} = \int_0^{V_{BEon}} c_{je}(V_{BE})\, dV_{BE} + \int_{V_+}^{V_+ - V_{BEon}} c_{jc}(V_{CB})\, dV_{CB} .$$

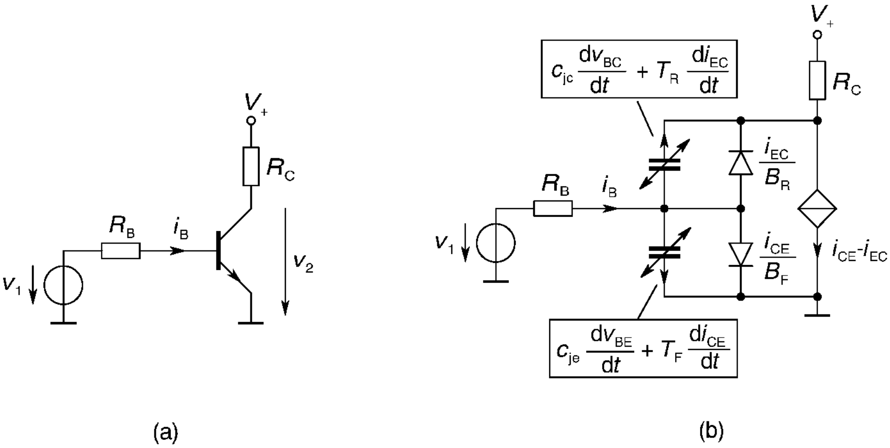


Fig. 1.60. Bipolar transistor operated as a switch. (**a**) Circuit and (**b**) equivalent circuit

Under the condition $V_1 \gg V_{\mathrm{BEon}}$, one may approximate the time-dependent base current by its average value, $\langle i_{\mathrm{B}} \rangle_1 \approx (V_1 - V_{\mathrm{BEon}}/2)/R_{\mathrm{B}}$, to estimate the time t_1 required to reach threshold,

$$t_1 \approx \Delta Q_{\mathrm{J1}}/\langle i_{\mathrm{B}} \rangle_1 \,. \tag{1.123}$$

When v_{BE} has reached V_{BEon}, further charging of the eb depletion capacitance may be neglected. Now the current i_{CE} increases, while i_{EC} may still be neglected as long as $v_{\mathrm{CE}} > V_{\mathrm{BEon}}$. With $\mathrm{d}v_{\mathrm{BE}}/\mathrm{d}t \approx 0$ and $v_{\mathrm{BE}} \approx V_{\mathrm{BEon}} \approx$ const., the base current i_{B} will be approximately constant, and (1.121) can be simplified to

$$i_{\mathrm{B}}(t) = I_{\mathrm{Bon}} = \frac{V_1 - V_{\mathrm{BEon}}}{R_{\mathrm{B}}} = \frac{i_{\mathrm{CE}}}{B_{\mathrm{F}}} + T_{\mathrm{F}}\frac{\mathrm{d}i_{\mathrm{CE}}}{\mathrm{d}t} + c_{\mathrm{jc}}\frac{\mathrm{d}v_{\mathrm{BC}}}{\mathrm{d}t} \,.$$

Integration from t_1 up to the HI-LO delay time t_{PDL}, after which the output voltage v_2 has changed by 50% of the voltage swing (Fig. 1.61), yields

$$I_{\mathrm{Bon}}(t_{\mathrm{PDL}} - t_1) = \Delta Q_{\mathrm{TE}} + \Delta Q_{\mathrm{JC2}} \,, \tag{1.124}$$

if the recombination current $i_{\mathrm{CE}}/B_{\mathrm{F}}$ that flows during this time interval is neglected. The right hand side of (1.124) determines the charge that has to be delivered by I_{B}. The first term ΔQ_{TE} represents the change of the diffusion charge; with $i_{\mathrm{CE}}(t_{\mathrm{PDL}}) \approx V_+/2R_{\mathrm{C}}$ one obtains

$$\Delta Q_{\mathrm{TE}} = T_{\mathrm{F}}\, i_{\mathrm{CE}}(t_{\mathrm{PDL}}) \approx T_{\mathrm{F}} V_+/2R_{\mathrm{C}} \,.$$

The second term on the right hand side of (1.124), ΔQ_{JC2}, describes the change of the charge stored in the bc depletion capacitance,

$$\Delta Q_{\mathrm{JC2}} = \int_{t_1}^{t_{\mathrm{PDL}}} c_{\mathrm{jc}}\frac{\mathrm{d}v_{\mathrm{BC}}}{\mathrm{d}t}\,\mathrm{d}t \approx \int_{V_+ - V_{\mathrm{BEon}}}^{V_+/2 - V_{\mathrm{BEon}}} c_{\mathrm{jc}}(V_{\mathrm{CB}})\,\mathrm{d}V_{\mathrm{CB}}$$

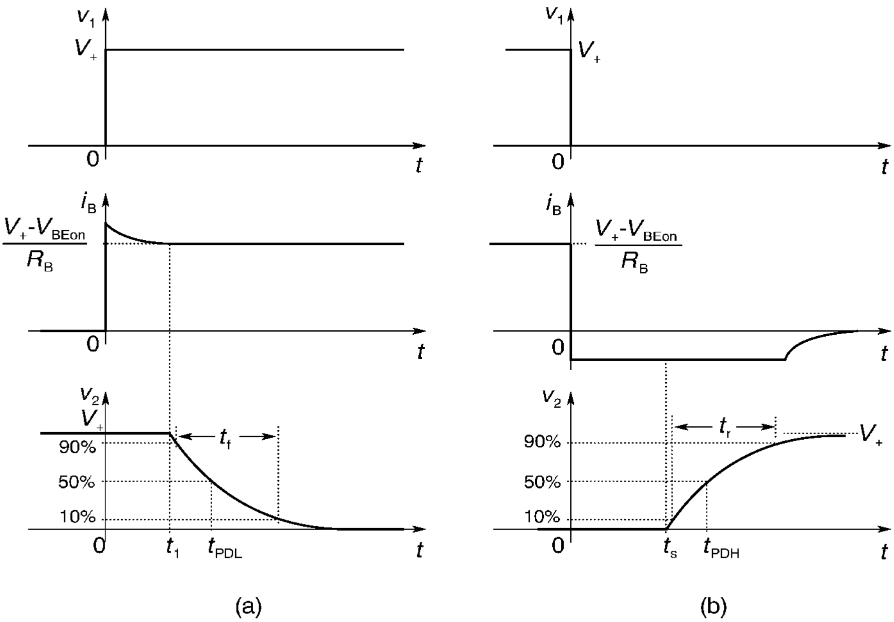


Fig. 1.61. Current and voltage waveforms of a bipolar transistor operated as a switch with ohmic load

if v_{CB} decreases from $V_+ - V_{BEon}$ to $V_+/2 - V_{BEon}$. From (1.124) the HI-LO delay time is obtained:

$$t_{PDL} = t_1 + (\Delta Q_{TE} + \Delta Q_{JC2})/I_{Bon} \,. \tag{1.125}$$

In the same way, the fall time t_f, which is the time required by the output voltage to decrease from 90% of its maximum value to 10%, is obtained:

$$t_f = (\Delta Q_{JC3} + 0.8\, T_F\, V_+/R_C)\,/I_{Bon} \,, \tag{1.126}$$

where

$$\Delta Q_{JC3} = \int_{0.9V_+ - V_{BEon}}^{0.1V_+ - V_{BEon}} c_{jc}(V_{CB})\, dV_{CB} \tag{1.127}$$

is the charge increment of the bc depletion capacitance associated with a decrease of v_{CE} from $0.9\,V_+$ to $0.1\,V_+$. When v_2 becomes smaller than V_{CEon}, i_{EC} begins to increase up to

$$I_{EC} = I_C \frac{B_R(B_F - B)}{B(B_F + B_R + 1)} \,. \tag{1.128}$$

The value of v_{BE} will increase only slightly up to the value V_{BEsat}, while v_{CE} will decrease to V_{CEsat}.

Turn-off. Assume a voltage step of v_1 at $t = 0$ from V_1 to zero. Then, the voltages $v_{\rm BE}$ and $v_{\rm CE}$ will first remain approximately constant, until the charge $T_{\rm R}\, i_{\rm EC}(0)$ associated with the forward-biased bc diode is gone. The time therefore required is called the storage time $t_{\rm s}$. In the time interval from 0 to $t_{\rm s}$, the base current is approximately constant, $i_{\rm B}(t) \approx -V_{\rm BEsat}/R_{\rm B} = -|I_{\rm Boff}|$. From (1.121) one therefore obtains

$$-|I_{\rm Boff}| = \frac{i_{\rm CE}}{B_{\rm F}} + \frac{i_{\rm EC}}{B_{\rm R}} + T_{\rm R}\,\frac{{\rm d}i_{\rm EC}}{{\rm d}t}\,,$$

where $T_{\rm F}\,{\rm d}i_{\rm CE}/{\rm d}t$ has been neglected because $T_{\rm F} \ll T_{\rm R}$. Since, generally, $B_{\rm F} \gg B_{\rm R}$, one obtains[37]

$$i_{\rm EC} + B_{\rm R}T_{\rm R}\,{\rm d}i_{\rm EC}/{\rm d}t = -B_{\rm R}|I_{\rm Boff}|\,.$$

This is a differential equation for $i_{\rm EC}$ of first order, with the general solution

$$i_{\rm EC}(t) = [\,i_{\rm EC}(0){+}B_{\rm R}|I_{\rm Boff}|\,]\exp\!\left(-\frac{t}{T_{\rm R}B_{\rm R}}\right) - B_{\rm R}|I_{\rm Boff}|\,.$$

From $i_{\rm EC}(t_{\rm s}) = 0$, the storage time $t_{\rm s}$ is estimated:

$$t_{\rm s} = T_{\rm R}B_{\rm R}\ln[\,1{+}i_{\rm EC}(0)/(B_{\rm R}|I_{\rm Boff}|)\,]\,, \qquad (1.130)$$

where $i_{\rm EC}(0) = I_{\rm EC}$ according to (1.128). For $t > t_{\rm s}$, i.e., after decharging the bc diffusion charge, $T_{\rm R}\,I_{\rm EC}$, the bc depletion charge and the diffusion charge $T_{\rm F}\,i_{\rm CE}$ have to be cut back. Equation (1.121) yields for the LO-HI delay time $t_{\rm PDH}$ in analogy to (1.125) the estimate

$$t_{\rm PDH} = t_{\rm s} + (\Delta Q_{\rm JC4} + T_{\rm F}V_+/2R_{\rm C})\,/|I_{\rm Boff}|\,, \qquad (1.131)$$

where

$$\Delta Q_{\rm JC4} = \int_{V_+/2-V_{\rm BEon}}^{V_{\rm CEon}-V_{\rm BEon}} c_{\rm jc}(V_{\rm CB})\,{\rm d}V_{\rm CB}$$

is the change of charge in the bc depletion capacitance associated with an increase of $v_{\rm CE}$ from $0.5\,V_+$ to $V_{\rm CEon}$. The rise time $t_{\rm r}$ follows from

$$t_{\rm r} = (\Delta Q_{\rm JC3} + 0.8\,T_{\rm F}V_+/R_{\rm C})\,/|I_{\rm Boff}|\,, \qquad (1.132)$$

with $\Delta Q_{\rm JC3}$ defined according to (1.127).

---

[37] In general, the current $i_{\rm CE}$ may be substituted, with the help of (1.122), which gives

$$\frac{V_+ - V_{\rm CEsat}}{R_{\rm C}} = i_{\rm CE} - i_{\rm EC}\left(1 + \frac{1}{B_{\rm R}}\right) - T_{\rm R}\,\frac{{\rm d}i_{\rm EC}}{{\rm d}t}\,, \qquad (1.129)$$

with this substitution, a second-order differential equation is obtained.

1.7 Elementary Small-Signal Models

1.7.1 Admittance Parameters

The small-signal behavior of a bipolar transistor may be expressed in terms of the admittance parameters $y_{\alpha\beta}$, which determine network elements in the general two-port equivalent circuit depicted in Fig. 1.62b.

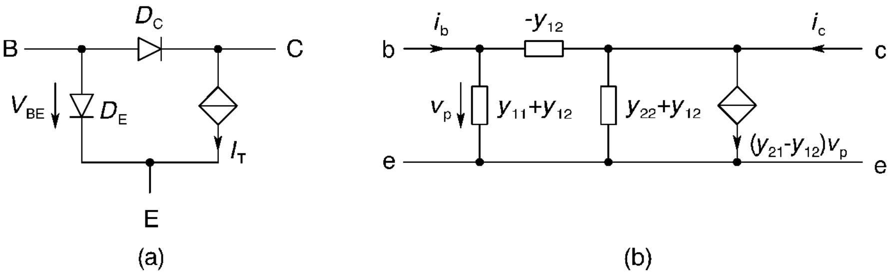


Fig. 1.62. (**a**) Elementary transistor model (series resistances neglected) and (**b**) general two-port equivalent circuit

Under low-frequency conditions, the admittance parameters are determined as partial derivatives of the current–voltage characteristics. Neglecting series resistances, these are given by (Fig. 1.62a)

$$I_C \approx I_S \left(1 + \frac{V_{CE}}{V_{AF}}\right) \exp\left(\frac{V_{BE}}{V_T}\right) \quad \text{and} \quad I_B = \frac{I_S}{B_F} \exp\left(\frac{V_{BE}}{V_T}\right) .$$

In this approximation the y-parameters (NF) are[38]

$$y_{11e} = \left(\frac{\partial I_B}{\partial V_{BE}}\right)_{V_{CE}} = \frac{I_B}{V_T} = g_\pi \tag{1.133}$$

$$y_{12e} = \left(\frac{\partial I_B}{\partial V_{CE}}\right)_{V_{BE}} = 0 \tag{1.134}$$

$$y_{21e} = \left(\frac{\partial I_C}{\partial V_{BE}}\right)_{V_{CE}} = \frac{I_C}{V_T} = g_m \tag{1.135}$$

$$y_{22e} = \left(\frac{\partial I_C}{\partial V_{CE}}\right)_{V_{BE}} = \frac{I_C}{V_{CE} + V_{AF}} = g_o \tag{1.136}$$

in forward operation with $V_{BE} \gg V_T$. The value of y_{12e} equals zero in this approximation; the general small-signal equivalent circuit therefore simplifies to the one depicted in Fig. 1.63, where $g_\pi = I_B/V_T$ is the input conductance,

[38] This are approximate relations for basic calculations, a refined description of small-signal parameters can be found in Sect. 3.9.

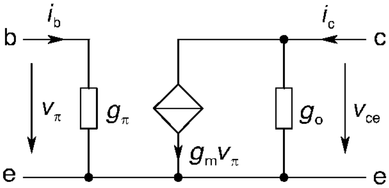


Fig. 1.63. Elementary small-signal equivalent circuit for NF operation

$g_m = I_C/V_T$ is the transconductance and $g_o = I_C/(V_{CE} + V_{AF})$ is the output conductance. Adding small-signal series resistances $r_{bb'}$, $r_{cc'}$ and $r_{ee'}$ leads to the small-signal equivalent circuit depicted in Fig. 1.64.

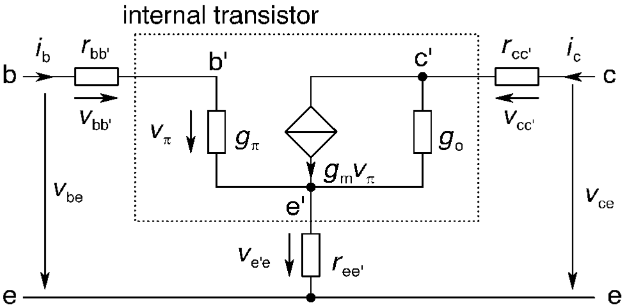


Fig. 1.64. Giacoletto model with series resistances

If the series resistances $R_{BB'}$, $R_{EE'}$ and $R_{CC'}$ of the large-signal equivalent circuit are bias-dependent, the small-signal series resistances $r_{bb'}$, $r_{ee'}$ and $r_{cc'}$ will show different values. As an example, the bias-dependent base resistance $R_{BB'}$ is considered. The current I_B applied at the bias point causes the voltage drop $V_{BB'} = R_{BB'} I_B$ across the base resistance. Superposition of a small-signal contribution i_b causes a change of the base resistance,

$$R_{BB'}(I_B + i_b) \approx R_{BB'} + i_b \left.\frac{dR_{BB'}}{dI_B}\right|_{I_B} ,$$

where $R_{BB'} = R_{BB'}(I_B)$ is the value of the base resistance at the bias point; the voltage drop across the base resistance now is

$$\begin{aligned} V_{BB'} + v_{bb'} &\approx \left(R_{BB'} + i_b \left.\frac{dR_{BB'}}{dI_B}\right|_{I_B} \right) (I_B + i_b) \\ &= R_{BB'} I_B + i_b \left(R_{BB'} + I_B \left.\frac{dR_{BB'}}{dI_B}\right|_{I_B} \right) + i_b^2 \left.\frac{dR_{BB'}}{dI_B}\right|_{I_B} . \end{aligned}$$

Considering only terms of the first order in i_b gives for the small-signal voltage drop across the base resistance

$$v_{\mathrm{bb'}} = \left(R_{\mathrm{BB'}} + I_{\mathrm{B}} \left. \frac{\mathrm{d}R_{\mathrm{BB'}}}{\mathrm{d}I_{\mathrm{B}}} \right|_{I_{\mathrm{B}}} \right) i_{\mathrm{b}} = r_{\mathrm{bb'}}\, i_{\mathrm{b}} \,, \tag{1.137}$$

where $r_{\mathrm{bb'}}$ denotes the small-signal base resistance. SPICE calculates the value of $r_{\mathrm{bb'}}$ during an `.OP`-analysis and displays the results as `RX` in the `.OUT`-file.

1.7.2 Hybrid Parameters

The hybrid parameters determine the response of the small-signal quantities v_{be} and i_{c} to changes of the small-signal quantities i_{b} and v_{ce}. The parameters

$$h_{11\mathrm{e}} = \left(\frac{\partial V_{\mathrm{BE}}}{\partial I_{\mathrm{B}}} \right)_{V_{\mathrm{CE}}} \quad \text{and} \quad h_{21\mathrm{e}} = \left(\frac{\partial I_{\mathrm{C}}}{\partial I_{\mathrm{B}}} \right)_{V_{\mathrm{CE}}}$$

are determined for $V_{\mathrm{CE}} = \text{const.}$, i.e., the small-signal component v_{ce} is zero, corresponding to a short circuit at the output. The parameters

$$h_{12\mathrm{e}} = \left(\frac{\partial V_{\mathrm{BE}}}{\partial V_{\mathrm{CE}}} \right)_{I_{\mathrm{B}}} \quad \text{and} \quad h_{22\mathrm{e}} = \left(\frac{\partial I_{\mathrm{C}}}{\partial V_{\mathrm{CE}}} \right)_{I_{\mathrm{B}}}$$

are determined for $I_{\mathrm{B}} = \text{const.}$ and $i_{\mathrm{b}} = 0$, respectively, corresponding to an input of the small-signal equivalent circuit that is left open. Their values depend on the collector current I_{C} and voltage V_{CE} of the bias point.

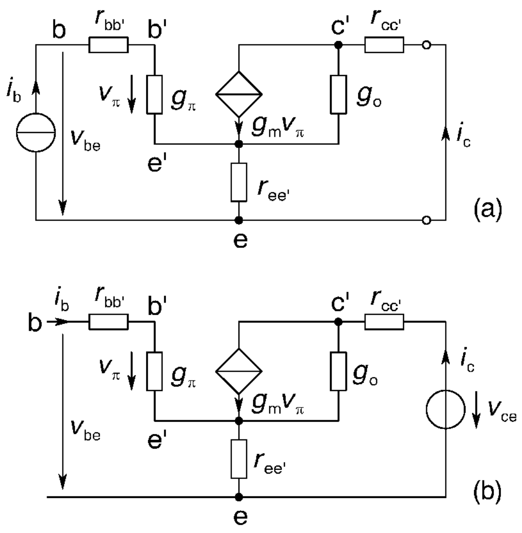


Fig. 1.65. Small-signal equivalent circuits for the determination of (**a**) parameters $h_{11\mathrm{e}}$ and $h_{21\mathrm{e}}$ and (**b**) parameters $h_{12\mathrm{e}}$ and $h_{22\mathrm{e}}$

The hybrid parameters can be calculated with the small-signal equivalent circuits shown in Fig. 1.65. The input resistance results as a series connection of three resistances,

$$h_{11e} = v_{be}/i_b = r_{bb'} + r_\pi + r_{ee'}(h_{21e}+1) \,. \tag{1.138}$$

The emitter series resistance $r_{ee'}$ is weighted with the factor $(h_{21e}+1)$, since both the input current and the output current, which is controlled by the input current, flow through this element. Neglecting series resistances leads to the following:

$$h_{11e} \approx r_\pi \approx V_T/I_C \,.$$

The current gain is

$$h_{21e} = \frac{i_c}{i_b} = \frac{g_m/g_\pi - r_{ee'}g_o}{1 + g_o(r_{ee'}+r_{cc'})} \tag{1.139}$$

and simplifies to

$$h_{21e} = \beta \approx g_m/g_\pi \tag{1.140}$$

if series resistances may be neglected. The small-signal current gain β is related to the large-signal current gain $B_N = I_C/I_B$ as follows:

$$\beta = \left(\frac{\partial I_C}{\partial I_B}\right)_{V_{CE}} = \left(\frac{\partial B_N I_B}{\partial I_B}\right)_{V_{CE}} = B_N + I_B\left(\frac{\partial B_N}{\partial I_B}\right)_{V_{CE}} , \tag{1.141}$$

or, since $(\partial B_N/\partial I_B)_{V_{CE}} = \beta(\partial B_N/\partial I_C)_{V_{CE}}$,

$$\beta = \frac{B_N}{1 - \dfrac{I_C}{B_N}\left(\dfrac{\partial B_N}{\partial I_C}\right)_{V_{CE}}} \,. \tag{1.142}$$

The large-signal current gain B_N and the small-signal current gain β are only equal if B_N is constant. In SPICE, the value of $\beta = g_m/g_\pi$ is calculated during an `.OP`-analysis and printed as `BETAAC` in the `.OUT`-file. The output conductance h_{22e} is determined from the equivalent circuit Fig. 1.65b,

$$h_{22e} = \frac{i_c}{v_{ce}} = \frac{g_o}{1 + g_o(r_{ee'}+r_{cc'})} \,. \tag{1.143}$$

Neglecting the series resistances gives the approximation

$$h_{22e} \approx g_o \,. \tag{1.144}$$

The voltage feedback h_{12e} also follows from Fig. 1.65b,

$$h_{12e} = \frac{v_{be}}{v_{ce}} = \frac{r_{ee'}}{r_{cc'}+r_o+r_{ee'}} \,. \tag{1.145}$$

Since the denominator is dominated by r_o, the following approximation holds:

$$h_{12e} \approx r_{ee'}g_o \ll 1 \,; \tag{1.146}$$

for most applications, one may assume $h_{12e} = 0$.

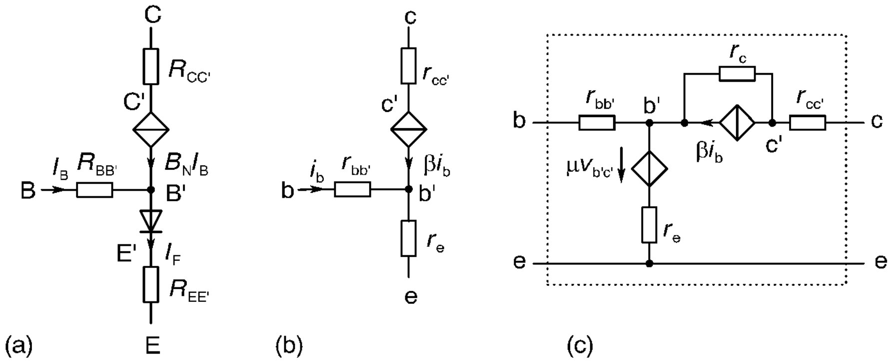


Fig. 1.66. (**a**) Ebers–Moll model for forward operation. (**b**) Elementary T-equivalent circuit. (**c**) T-equivalent circuit that considers voltage feedback and Early effect

1.7.3 T-Equivalent Circuit

Figure 1.66a shows the Ebers–Moll model extended by series resistances for forward operation. Linearization of the network elements yields the T-equivalent circuit depicted in Fig. 1.66b, which, however, does not consider voltage feedback and the Early effect. To consider these quantities the equivalent circuit is extended by the resistance r_c and the controlled source $\mu v_{b'c'}$.

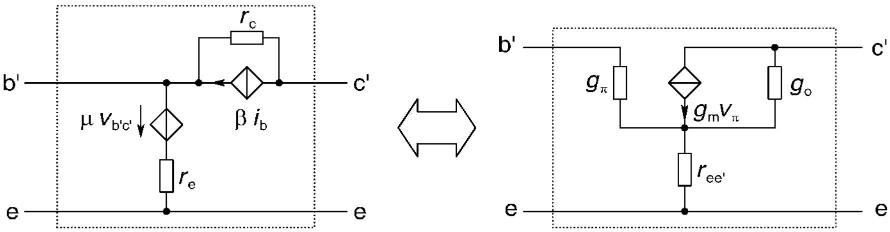


Fig. 1.67. Correspondence between T- and π-equivalent circuits

The extended T-equivalent circuit Fig. 1.66c is equivalent to the π-equivalent circuit for low-frequency operation (capacitances omitted) built from six network elements, where the values of $r_{bb'}$ and $r_{cc'}$ are equal in both models. The quantities β, r_e, r_c and μ of the T-equivalent circuit can be calculated from the values of g_π, g_m, g_o and $r_{ee'}$ of the π-equivalent circuit, since both equivalent circuits must yield the same two-port parameters. The transformation equations are

$$\beta = g_\mathrm{m}/g_\pi \tag{1.147}$$

$$r_\mathrm{c} = r_\pi(1 + r_{\mathrm{ee'}} g_\mathrm{o})/(\beta + 1) + r_{\mathrm{ee'}} \approx r_{\mathrm{ee'}} + 1/g_\mathrm{m} \tag{1.148}$$

$$r_\mathrm{c} = (r_{\mathrm{ee'}} + r_\mathrm{o})/(1 - r_{\mathrm{ee'}} g_\mathrm{o}) \approx r_\mathrm{o} \tag{1.149}$$

$$\mu = r_\mathrm{e}/r_\mathrm{c} + r_{\mathrm{ee'}} g_\mathrm{o} \approx r_\mathrm{e}/r_\mathrm{c} \,. \tag{1.150}$$

1.7.4 Frequency Limits

With each pn junction there is a depletion and a diffusion capacitance. In our small-signal model these have to be modeled as two additional capacitances, c_π (located between base and emitter) and c_μ (located between base and collector), resulting in the small-signal equivalent circuit devised by Giacoletto (Fig. 1.68).

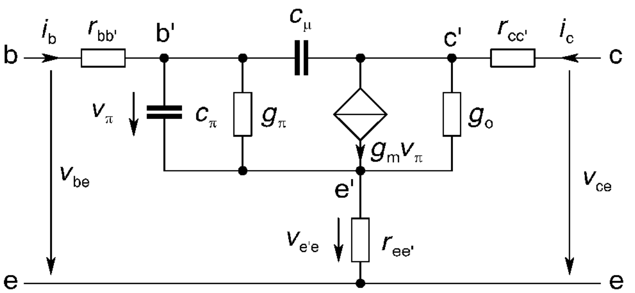


Fig. 1.68. Small-signal π-equivalent circuit devised by to Giacoletto

Under forward operation only the eb diode is forward biased, i.e., there is no diffusion charge associated with the bc diode. If c_{je} denotes the eb depletion capacitance and c_{jc} the bc depletion capacitance one may therefore write

$$c_\pi = c_{\mathrm{je}} + \tau_\mathrm{f} g_\mathrm{m} \quad \text{and} \quad c_\mu = c_{\mathrm{jc}} \,, \tag{1.151}$$

where $\tau_\mathrm{f} g_\mathrm{m}$ denotes the diffusion capacitance of the eb diode. The forward transit time τ_f is set equal to the parameter T_F in the elementary model and is somewhat larger than the base transit time τ_B. The diffusion charge associated with the forward-biased eb junction is $q_{\mathrm{TE}} = \tau_\mathrm{f} I_\mathrm{C}$. Differentiation of this expression with respect to V_{BE} yields the diffusion capacitance $\tau_\mathrm{f} g_\mathrm{m}$.

The capacitances determine the electrical behavior of the transistor at large frequencies. The two-port parameters then become complex, frequency-dependent quantities. Of particular importance is the frequency dependence of the forward common-emitter current gain $h_{21\mathrm{e}}$ and the transconductance $y_{21\mathrm{e}}$, both with the output short circuit.

Figure 1.69 shows a small-signal equivalent circuit for the computation of $h_{21\mathrm{e}}$ and $y_{21\mathrm{e}}$. The series resistances $r_{\mathrm{ee'}}$ and $r_{\mathrm{cc'}}$ were omitted for simplicity;

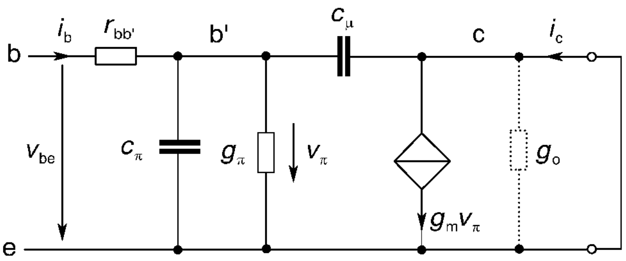


Fig. 1.69. Small-signal π-equivalent circuit for the determination of h_{21e} and y_{21e}

then g_o is short-circuited and has no effect on the two-port parameters. The current $\underline{i}_b$ at the input flows through g_π and the capacitances c_π and c_μ, which short-circuit g_π with increasing frequency. For an input current with constant amplitude, the voltage drop $\underline{v}_\pi$ will then vary in inverse proportion to the frequency and so will the transfer current controlled by $\underline{v}_\pi$. If the contribution of the current through c_μ to the collector current is neglected, one obtains

$$\underline{i}_C = g_m \underline{v}_\pi = g_m r_\pi \frac{g_\pi}{g_\pi + j\,\omega(c_\pi + c_\mu)} \underline{i}_b = \frac{\beta}{1 + j\,f/f_\beta} \underline{i}_b ,$$

where f_β denotes the β-cutoff frequency,

$$f_\beta = \frac{1}{2\pi} \frac{g_\pi}{c_\pi + c_\mu} .$$

For h_{21e} one obtains up to the first order of frequency f

$$h_{21e}(f) \approx \beta/(1 + j\,f/f_\beta) . \tag{1.152}$$

For $f \gg f_\beta$, the magnitude of h_{21e} will therefore decrease in proportion to $1/f$, corresponding to a decrease of 20 dB per decade of frequency in a Bode plot,

$$|h_{21e}(f)| \approx \beta f_\beta / f = f_T / f \quad \text{if } f \gg f_\beta .$$

The quantity $f_T = \beta\, f_\beta$ is known as the cutoff frequency of the transistor. If $f = f_T$, the one-pole approximation of $|h_{21e}|$ becomes unity. Figure 1.70 illustrates this behavior schematically for a bipolar transistor with $\beta = 100$.

For $r_{ee'} \neq 0$, $r_{cc'} \neq 0$and negligible Early effect ($g_o = 0$), the Giacoletto model (Fig. 1.68) yields for the cutoff frequency:

$$\begin{aligned}
\frac{1}{2\pi f_T} &= \frac{c_\pi + c_\mu}{g_m} + (r_{ee'} + r_{cc'}) c_\mu \\
&= \frac{(c_{je} + c_{jc})\, V_T}{I_C} + \tau_f + (r_{ee'} + r_{cc'}) c_{jc} .
\end{aligned} \tag{1.153}$$

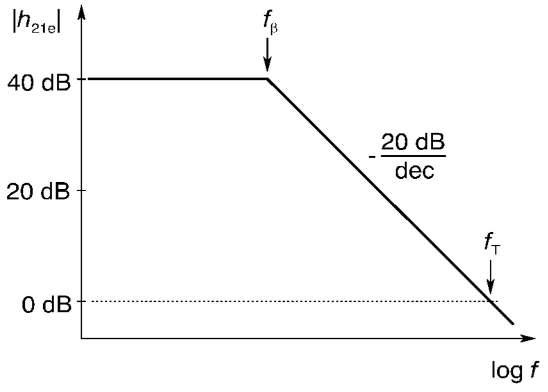


Fig. 1.70. Frequency dependence of $|h_{21e}|$ for a bipolar transistor with $\beta = 100$

As is illustrated in Fig. 1.71, the cutoff frequency is bias-dependent: for small values of the collector current, I_C, its value is dominated by the first term on the right hand side of (1.153),

$$f_T \approx \frac{I_C}{2\pi(c_{je}+c_{jc})V_T} ,$$

which causes f_T to increase with I_C. For large values of I_C the first term in (1.153) may be neglected, resulting in

$$f_T \approx \frac{1}{2\pi\left[\tau_f + (r_{ee'}+r_{cc'})c_{jc}\right]} .$$

In real transistors τ_f increases for large values of the transfer current; the cutoff frequency thus decreases after having passed a maximum. The maximum value of f_T increases with the value of V_{CE}, since c_{jc} and the value of τ_f decrease due to the widening of the bc space-charge layer.

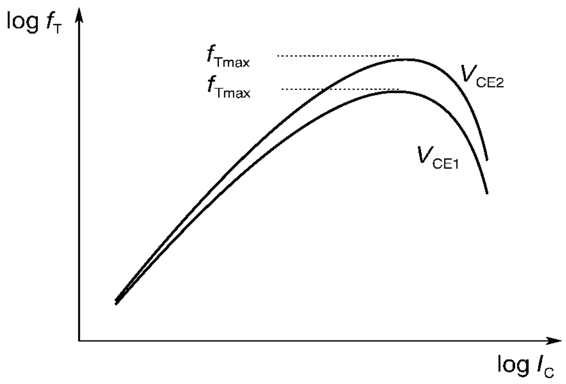


Fig. 1.71. Cutoff frequency f_T versus I_C for two different values of V_{CE} (schematic)

The transconductance y_{21e} is relevant for voltage control of the input. Its frequency dependence is determined by the voltage divider composed of $r_{bb'}$ and the parallel connection of g_π and $(c_\pi + c_\mu)$. With the ratio

$$\frac{\underline{v}_\pi}{\underline{v}_{be}} = \frac{[g_\pi + j\omega(c_\pi + c_\mu)]^{-1}}{r_{bb'} + [g_\pi + j\omega(c_\pi + c_\mu)]^{-1}} = \frac{1}{1 + g_\pi r_{bb'} + j\omega r_{bb'}(c_\pi + c_\mu)} ,$$

one obtains for the transconductance

$$y_{21e} \approx g_m \frac{\underline{v}_\pi}{\underline{v}_{be}} \approx \frac{g_m}{1 + r_{bb'} g_\pi} \frac{1}{1 + jf/f_y} , \tag{1.154}$$

where

$$f_y = \frac{1 + r_{bb'} g_\pi}{2\pi r_{bb'}(c_\pi + c_\mu)} \approx \frac{1}{2\pi r_{bb'}(c_\pi + c_\mu)} \tag{1.155}$$

defines the transconductance-cutoff frequency. Neglecting the term $r_{bb'}\, g_\pi$ in the numerator yields

$$f_y \approx \frac{1}{2\pi r_{bb'}(c_\pi + c_\mu)} = \frac{1}{r_{bb'} g_m} \frac{g_m}{2\pi(c_\pi + c_\mu)} \approx \frac{f_T}{r_{bb'} g_m} = \frac{f_\beta}{r_{bb'} g_\pi} .$$

The transconductance-cutoff frequency f_y is therefore large in comparison with the β-cutoff frequency f_β, as long as $r_{bb'} g_\pi \ll 1$ holds.

For frequencies in excess of the maximum frequency of oscillation, f_{max}, the output power of an amplifier in common-emitter configuration is smaller than the power delivered to the input; the transistor can then no longer be applied as the active element of an oscillator circuit.[39] For an approximate calculation of f_{max}, the small-signal equivalent circuit depicted in Fig. 1.72a is considered. For high frequencies with $\omega c_\pi \gg r_{bb'}$, the input voltage v_1 drops predominantly at $r_{bb'}$; the power delivered to the input is then approximately $P_1 = r_{bb'} I_b^2$, where I_b is the rms value of the small-signal base current i_b. The power P_2 delivered to the load R_L is on the other hand $P_2 = R_L I_c^2$, where I_c is the rms value of the small-signal collector current i_c. The maximum power gain is obtained if the source impedance R_S is matched to the input impedance ($R_S \approx r_{bb'}$ in the frequency range considered), and if the load impedance R_L is matched to the real part of the output impedance $1/h_{22e}$. Under the condition $\omega c_\pi \gg 1/[r_\pi \,\|\, (R_S + r_{bb'})]$, the resistances $r_{bb'}$, R_S and r_π can be neglected in the network Fig. 1.72b, which is used for the calculation of the output impedance $h_{22e} = \underline{i}_c/\underline{v}_2$. The capacitances c_μ and c_π then act as a capacitive voltage divider, i.e., $\underline{v}_\pi/\underline{v}_2 = c_\mu/(c_\mu + c_\pi)$. The collector current therefore is

$$\underline{i}_c = g_m \underline{v}_\pi + j\omega c_\mu(\underline{v}_2 - \underline{v}_\pi) = \frac{c_\mu}{c_\mu + c_\pi}(g_m + j\omega c_\pi)\,\underline{v}_2 ,$$

[39] Practical oscillator circuits operate at frequencies well below f_{max}.

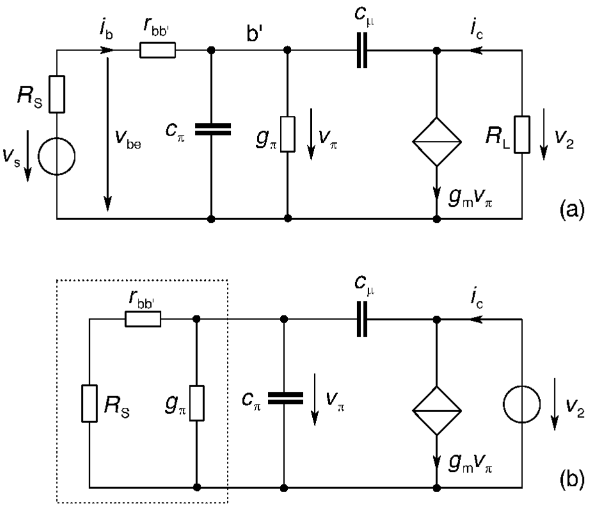


Fig. 1.72. Small-signal equivalent circuit for the computation of f_{max}. (**a**) Circuit used for the computation of the power P_1 delivered to the input and the power P_2 delivered to the load; (**b**) circuit used for the computation of the output impedance

resulting in the output conductance

$$h_{22\mathrm{e}} = \frac{c_\mu}{c_\mu + c_\pi} (g_{\mathrm{m}} + \mathrm{j}\omega c_\pi) \ .$$

Impedance matching at the output therefore requires

$$R_{\mathrm{L}} \approx (c_\mu + c_\pi)/c_\mu g_{\mathrm{m}} \ .$$

Since

$$\frac{\underline{i}_{\mathrm{c}}}{\underline{i}_{\mathrm{b}}} = \frac{\underline{I}_{\mathrm{c}}}{\underline{I}_{\mathrm{b}}} = \frac{h_{21\mathrm{e}}}{1 + h_{22\mathrm{e}} R_{\mathrm{L}}} \approx \frac{h_{21\mathrm{e}}}{2} \ ,$$

one obtains

$$I_{\mathrm{c}}^2 / I_{\mathrm{b}}^2 \approx |h_{21\mathrm{e}}|^2 / 4 \approx f_{\mathrm{T}}^2 / 4 f^2$$

in the frequency range considered. The maximum power gain then is approximately

$$\frac{P_2}{P_1} = \frac{R_{\mathrm{L}}}{r_{\mathrm{bb'}}} \frac{f_{\mathrm{T}}^2}{4 f^2} \approx \frac{1}{4} \frac{c_\mu + c_\pi}{c_\mu g_{\mathrm{m}}} \frac{1}{r_{\mathrm{bb'}}} \frac{f_{\mathrm{T}}^2}{f^2} \approx \frac{f_{\mathrm{T}}}{8\pi r_{\mathrm{bb'}} c_\mu} \frac{1}{f^2} \ ,$$

since $f_{\mathrm{T}} \approx g_{\mathrm{m}} / 2\pi (c_\mu + c_\pi)$. The maximum frequency of oscillation, f_{max}, is computed as the frequency for which $P_2/P_1 \to 1$, resulting in

$$f_{\mathrm{max}} \approx \sqrt{\frac{f_{\mathrm{T}}}{8\pi r_{\mathrm{bb'}} c_\mu}} \ . \qquad (1.156)$$

This result neglects, for example, the effect of the collector–substrate capacitance (see Sect. 3.10).

1.8 Noise Modeling

The term noise, as it is used here, denotes statistical fluctuations of terminal currents or voltages due to the quantization of charge, thermal motion of the carriers and generation or recombination processes.[40] Computation of noise behavior is possible with noise equivalent circuits; for the characterization of the noise produced by a transistor, the noise figure is specified.

1.8.1 Noise and Noise Sources

Figure 1.73 schematically depicts a noisy current signal with a noise current $i_\mathrm{n}(t)$ superposed on the average current I.

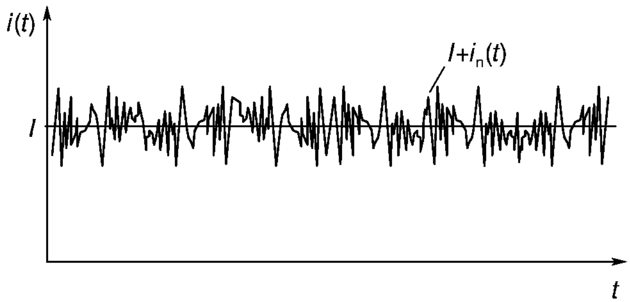


Fig. 1.73. Noisy current signal

The temporal average, $\overline{i_\mathrm{n}}$, of the noise current $i_\mathrm{n}(t)$ is zero by definition

$$\overline{i_\mathrm{n}} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} i_\mathrm{n}(t')\,\mathrm{d}t' = 0 ;$$

therefore, to obtain a measure for the intensity of noise, the rms noise current

$$I_\mathrm{n} = \sqrt{\overline{i_\mathrm{n}^2}} = \sqrt{\lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} i_\mathrm{n}^2(t')\,\mathrm{d}t'}$$

is generally considered. Practical calculations make use of the spectral density[41] $S_i(f)$ of the noise current; $S_i(f)$ determines the mean square noise current within a frequency interval of width 1 Hz and therefore has the dimension A^2/Hz. The rms noise current I_n in the frequency interval $[\,f, f{+}\Delta f\,]$ follows from $S_i(f)$ according to

$$I_\mathrm{n} = \sqrt{\overline{i_\mathrm{n}^2}} = \sqrt{\int_f^{f+\Delta f} S_i(f)\,\mathrm{d}f} \,. \tag{1.157}$$

[40]See Chap. 5.
[41]See Appendix E.

In complete analogy, noise voltages may be described by a spectral density $S_v(f)$ (dimension V^2/Hz), which determines the rms noise voltage V_n in the frequency interval $[\,f, f+\Delta f\,]$

$$V_\mathrm{n} = \sqrt{\overline{v_\mathrm{n}^2}} = \sqrt{\int_f^{f+\Delta f} S_v(f)\,\mathrm{d}f}\;. \tag{1.158}$$

Thermal Noise. The random thermal motion of the carriers within a conducting material causes a noise voltage between two contacts to the material by electrostatic induction. The spectral density of this thermal noise[42] voltage is

$$S_v(f) = 4k_\mathrm{B}TR\,, \tag{1.159}$$

where $k_\mathrm{B} = 1.38066 \times 10^{-23}$ J/K denotes the Boltzmann constant and T the absolute temperature.[43] The noise behavior of an ohmic resistor may therefore

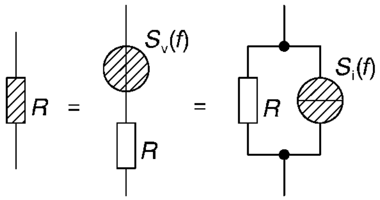


Fig. 1.74. Noise equivalent circuit of an ohmic resistor

be described by a series connection of a noise voltage source with spectral density $S_v(f)$ and a noise-free resistor R, as shown in Fig. 1.74. To distinguish noise-free circuit elements from noisy ones, the latter are illustrated with hatched symbols. An equivalent representation employs a parallel connection of a noise current source with spectral density

$$S_i(f) = S_v(f)/R^2 = 4\,k_\mathrm{B}T/R\,, \tag{1.160}$$

and a noise-free resistor R. Formulas (1.159) and (1.160) describe noise sources with spectral densities which are independent of frequency – so-called white noise sources. The rms noise current and noise voltage generated by a resistor

[42] Thermal noise was first observed by J.B. Johnson in 1927 [59], and it was theoretically described by H. Nyquist in 1928 [60]; it is therefore often termed Johnson noise or Nyquist noise.

[43] Expression (1.159) is an approximation: at large frequencies the spectral density of the noise voltage source has to be multiplied by the Planck factor [61]:

$$p(f) = \frac{hf}{k_\mathrm{B}T}\,\frac{1}{\exp(hf/k_\mathrm{B}T) - 1}\,,$$

where h denotes the Planck constant. For typical operating frequencies in electronic circuits, the deviations of $p(f)$ from one may generally be neglected: at $f = 10$ GHz: $1 - p(f) = 8 \times 10^{-4}$, and at $f = 100$ GHz: $1 - p(f) = 8 \times 10^{-3}$.

therefore depends on the bandwidth B of the circuit: according to formulas (1.157) and (1.158) these quantities will increase in proportion to $\sqrt{B}$.

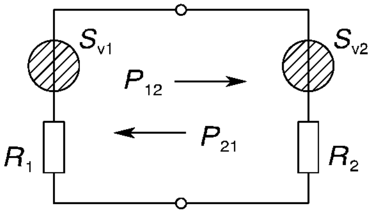


Fig. 1.75. Power exchanged between two noisy resistors connected in parallel

Figure 1.75 shows a noise equivalent circuit of two resistors R_1 (held at temperature T_1) and R_2 (held at temperature T_2). In a frequency interval of width Δf, resistor R_1 produces the rms noise voltage $V_{n1} = \sqrt{4k_B T_1 R_1 \Delta f}$, which drops across the two resistors according to the voltage divider ratio. The power delivered to R_2 by this process is

$$P_{12} = \frac{1}{R_2}\left(\frac{R_2}{R_1+R_2}V_{n1}\right)^2 = \frac{4k_B T_1 R_1 R_2}{(R_1+R_2)^2}\Delta f ;$$

its value is maximum if $R_1 = R_2$ and given by the available noise power

$$P_{NV} = k_B T_1 \Delta f = 4 \times 10^{-21}\,\mathrm{W} \times \left(\frac{T_1}{290\,\mathrm{K}}\right)\frac{\Delta f}{\mathrm{Hz}} . \tag{1.161}$$

Analogously, R_2 delivers the power

$$P_{21} = \frac{1}{R_1}\left(\frac{R_1}{R_1+R_2}\sqrt{4k_B T_2 R_2\,\Delta f}\right)^2 = \frac{4k_B T_2 R_1 R_2}{(R_1+R_2)^2}\Delta f$$

to R_1. If both resistors have the same temperature ($T_1 = T_2$), the system is in thermal equilibrium and $P_{12} = P_{21}$. If R_2 is replaced by an impedance Z, resistor R_1 delivers the power

$$\mathrm{d}P_{12} = 4k_B T R_1 \frac{\mathrm{Re}(Z)}{|R_1+Z|^2}\Delta f$$

to Z within a frequency interval of width Δf. The power delivered by Z to R_1 in this frequency interval is, on the other hand,

$$\mathrm{d}P_{21} = \frac{R_1}{|R_1+Z|^2} S_v^{(Z)}(f)\Delta f ,$$

where $S_v^{(Z)}(f)$ is the spectral density of the noise voltage produced by Z. In thermal equilibrium $\mathrm{d}P_{12} = \mathrm{d}P_{21}$ must be fulfilled, and therefore

$$S_v^{(Z)}(f) = 4k_B T\,\mathrm{Re}(Z) . \tag{1.162}$$

The noise voltage of an impedance is therefore solely determined by $\mathrm{Re}(Z)$.

Shot Noise. The transport of charge across depletion layers in bipolar devices causes shot noise, which over a wide frequency range may be described by a white-noise current source with spectral density (see Chap. 5)

$$S_i(f) \approx 2e(I+2I_\mathrm{S}) \approx 2eI . \tag{1.163}$$

The rms noise current in a frequency interval of width Δf therefore is

$$I_\mathrm{n} = \sqrt{\overline{i_\mathrm{n}^2}} \approx \sqrt{2eI\Delta f} . \tag{1.164}$$

1/f Noise. Thermal noise and shot noise are the predominant causes of noise in electronic devices at large frequencies. At small frequencies additional mechanisms, e.g., due to generation–recombination processes [62, 63], cause extra noise. This noise can generally be described by a noise current source with a frequency-dependent spectral density that obeys the relation

$$S_i(f) = \left(\frac{I}{\mathrm{A}}\right)^{A_\mathrm{F}} \frac{K_\mathrm{F}}{f} , \tag{1.165}$$

where K_F (dimension A^2) and A_F are parameters that are determined from measured data. In pn junctions, where $1/f$ noise appears in addition to shot noise, the noise corner frequency f_c, at which both contributions are equal, is often specified. From $2eI = K_\mathrm{F}(I/\mathrm{A})^{A_\mathrm{F}}/f_\mathrm{c}$, its value is related with the parameter K_F by

$$f_\mathrm{c} = \frac{K_\mathrm{F}}{2eI}\left(\frac{I}{\mathrm{A}}\right)^{A_\mathrm{F}} . \tag{1.166}$$

The noise corner frequency will be bias-dependent if $A_\mathrm{F} \neq 1$.

1.8.2 Noise Circuit Analysis

For the computation of noise signals, noise sources are added to the small-signal equivalent circuit. If two (uncorrelated) noise sources with spectral densities $S_{v1}(f)$ and $S_{v2}(f)$ are connected in series, the resulting noise voltage has spectral density

$$S_v(f) = S_{v1}(f) + S_{v2}(f) .$$

In complete analogy, a parallel connection of two noise current sources with spectral densities $S_{i1}(f)$ and $S_{i2}(f)$ produces a noise current with spectral density

$$S_i(f) = S_{i1}(f) + S_{i2}(f) .$$

These relations must not be applied if the noise sources are correlated. Generally, the superposition of two noise currents, $i_{\mathrm{n}1}(t)$ and $i_{\mathrm{n}2}(t)$, yields a noise current with mean square deviation

$$\overline{(i_{\text{n}1}+i_{\text{n}2})^2} = I_{\text{n}1}^2 + I_{\text{n}2}^2 + 2\overline{i_{\text{n}1}i_{\text{n}2}} \, .$$

In the case of uncorrelated noise currents, $i_{\text{n}1}$ and $i_{\text{n}2}$, the time average $\overline{i_{\text{n}1}i_{\text{n}2}}$ vanishes, and the mean square of the noise current results from a simple addition of the single mean square values. If correlation exists between the noise currents, the correlation coefficient

$$\gamma_{1,2} = \overline{i_{\text{n}1}i_{\text{n}2}}/(I_{\text{n}1}I_{\text{n}2}) \, , \tag{1.167}$$

will be different from zero: uncorrelated noise currents $i_{\text{n}1}$ and $i_{\text{n}2}$ have a correlation coefficient of zero, while completely correlated noise currents $i_{\text{n}1}$ and $i_{\text{n}2}$ have a correlation coefficient of one.

If the investigation is restricted to a small frequency interval, it is useful to introduce complex effective values for the generated noise current and noise voltage. Within a small frequency interval of width Δf at f, one may write

$$\underline{i}_{\text{n}}(t) = \sqrt{2}\,\underline{I}_{\text{n}}\text{e}^{\text{j}\omega t} \, , \quad \text{with} \quad \underline{I}_{\text{n}} = \sqrt{S_i(f)\Delta f} \times \text{e}^{\text{j}\varphi_{\text{n}}} \, ;$$

φ_{n} is the phase of the complex effective value $\underline{I}_{\text{n}}$ and subject to random phase variations between 0 and 2π. The time average is then replaced by an average with respect to the phase angles φ_{n}, a procedure indicated by the brackets $\langle\cdots\rangle$. If two variables are uncorrelated, there is no relation between the two phases and the average of their product will vanish.

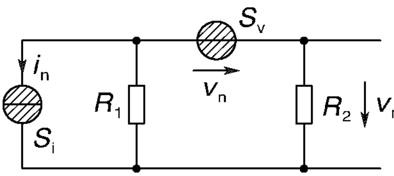


Fig. 1.76. Noise equivalent circuit considered in example 1.8.1

Example 1.8.1 To illustrate the method, the noise voltage across resistor R_2 of the circuit depicted in Fig. 1.76 will be calculated. R_1 and R_2 are assumed to be noiseless; i_{n} and v_{n} are assumed to be correlated with correlation coefficient γ_{iv}. Since the circuit is linear, the complex effective noise voltage across R_2 is easily obtained by superposition, with the result

$$\underline{V}_{\text{n}2} = -\frac{R_2}{R_1+R_2}\,(R_1\underline{I}_{\text{n}} + \underline{V}_{\text{n}}) \, .$$

From this result, the mean-square noise voltage is obtained by multiplication with $\underline{V}_{\text{n}2}^*$ and subsequent averaging with respect to the random phase angles:

$$\langle\underline{V}_{\text{n}2}\underline{V}_{\text{n}2}^*\rangle = \frac{R_2^2}{(R_1+R_2)^2}\left(R_1^2\langle\underline{I}_{\text{n}}\underline{I}_{\text{n}}^*\rangle + \langle\underline{V}_{\text{n}}\underline{V}_{\text{n}}^*\rangle + R_1\langle\underline{I}_{\text{n}}\underline{V}_{\text{n}}^* + \underline{I}_{\text{n}}^*\underline{V}_{\text{n}}\rangle\right) \, .$$

Since $\langle\underline{V}_{\text{n}2}\underline{V}_{\text{n}2}^*\rangle = S_{v2}(f)\Delta f$, $\langle\underline{I}_{\text{n}}\underline{I}_{\text{n}}^*\rangle = S_i(f)\Delta f$, $\langle\underline{V}_{\text{n}}\underline{V}_{\text{n}}^*\rangle = S_v(f)\Delta f$, and

$$\langle \underline{I}_{\mathrm{n}} \underline{V}_{\mathrm{n}}^* + \underline{I}_{\mathrm{n}}^* \underline{V}_{\mathrm{n}} \rangle = 2\gamma_{\mathrm{iv}} \sqrt{S_i(f) S_v(f)}\, \Delta f \ ,$$

the spectral density of the noise voltage across R_2 can be written as follows:

$$S_{v2}(f) = \frac{R_2^2}{(R_1+R_2)^2} \left(R_1^2 S_i(f) + S_v(f) + 2\gamma_{\mathrm{iv}} R_1 \sqrt{S_i(f) S_v(f)} \right) \ .$$

Circuit Analysis with SPICE

Noise analysis in SPICE is possible by inserting the `.NOISE` statement in the input control file after a `.AC` statement, which performs a frequency analysis [61, 64]. The `.NOISE` statement has the principal form

```
.NOISE     V(N)     INPUT SOURCE
```

During a `.NOISE` analysis, the small-signal equivalent circuit is extended by uncorrelated small-signal noise sources. With these, the noise voltage at the specified output node `N` is calculated within a bandwidth 1 Hz as a function of frequency as follows:

$$\texttt{V(ONOISE)} = \sqrt{S_{v,\mathtt{N}}(f) \times 1\,\mathrm{Hz}} \ ;$$

the input-referred noise voltage is

$$\texttt{V(INOISE)} = \texttt{V(ONOISE)}/A_v(f) \ ,$$

where $A_v(f)$ is the voltage gain from the input source to the specified output node `N`. A simple application of the `.NOISE` statement is found in the following example.

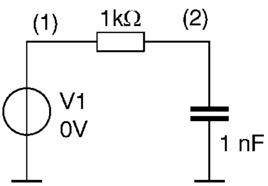


Fig. 1.77. Example circuit used for noise analysis with SPICE

Example 1.8.2 As an example the noise voltage at node (2) of the circuit shown in Fig. 1.77 is calculated with the SPICE `.NOISE` analysis and compared with the result of the analytical computation. The statements

```
.AC    DEC    100    10MEG    21
.NOISE    V(2)    V1
.TEMP    0    50    100
```

initiate (after a mandatory `.AC` analysis) the computation of the noise voltage at node (2) within a frequency band of 1 Hz for three different temperature values (0°C, 50°C and 100°C). In Fig. 1.78 the computed values of `V(ONOISE)` are plotted as a function of frequency.

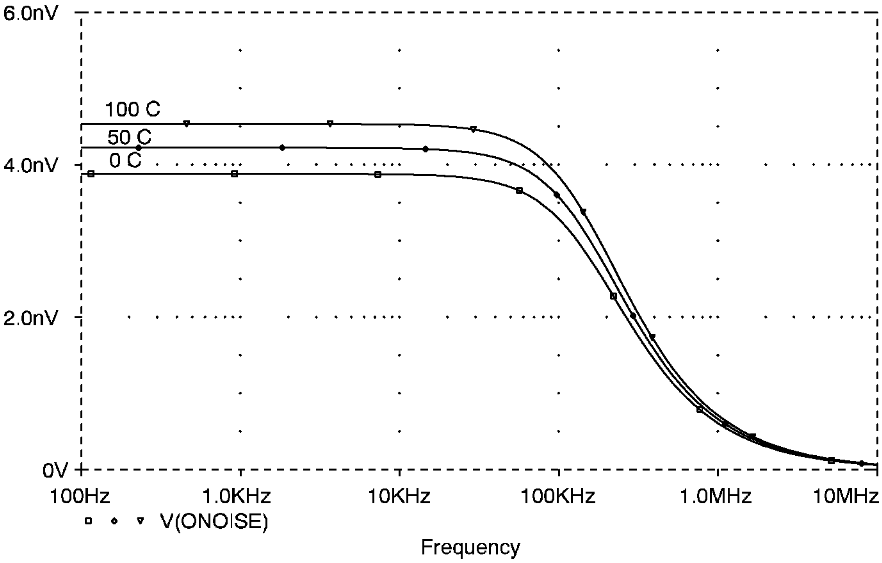


Fig. 1.78. Result of the `.NOISE` analysis

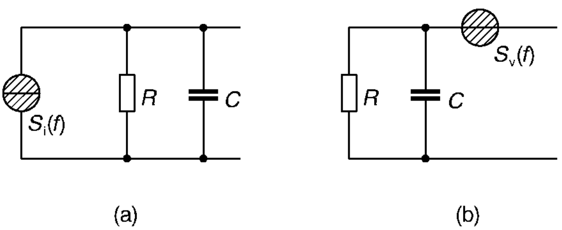


Fig. 1.79. Noise equivalent circuit for a parallel connection of ohmic resistor and capacitor

Figure 1.79a shows the respective noise equivalent circuit; the thermal noise current generated by the ohmic resistor is represented by a noise current source with spectral density $S_i(f) = 4k_{\mathrm{B}}T/R$. An alternative noise equivalent circuit, which represents the generated noise voltage, with a noise voltage generator of spectral density $S_v(f)$ is shown in Fig. 1.79b. The spectral density $S_v(f)$ is related to $S_i(f)$ by the impedance $Z = (\mathrm{j}\omega C + 1/R)^{-1}$ of the parallel connection of R and C as follows:

$$S_v(f) = |Z|^2 S_i(f) = \frac{4k_{\mathrm{B}}TR}{1+(2\pi fRC)^2} .$$

At small frequencies $f \ll (2\pi RC)^{-1}$, the rms value of the noise voltage in a frequency band of width 1 Hz is

$$\sqrt{4k_{\mathrm{B}}TR \times 1\ \mathrm{Hz}} = 2.349 \times 10^{-10}\ \mathrm{V} \times \sqrt{T/\mathrm{K}}\ .$$

Its value increases with temperature; at 0°C (or $T = 273$ K), the analytical computation gives 3.88 nV, while at 100°C (or $T = 373$ K) the value 4.54 nV is obtained in accordance with Fig. 1.78. At large frequencies $f > (2\pi RC)^{-1} \approx 160$ kHz, the noise voltage generated by the resistor is shunted by the capacitor, resulting in a decrease with frequency.

1.8.3 Noisy Linear Two-Ports, Noise Figure

The admittance parameters or any other full set of two-port parameters provide a full description of a linear noise-free two-port. In noisy two-ports, additional noise currents have to be superposed onto the currents of the noise-free two-port. Fig. 1.80a shows such a noise equivalent circuit that has two noise current sources with spectral densities S_{i1} and S_{i2} connected to the input and output port. With the complex effective values $\underline{I}_{\mathrm{n1}}$ and $\underline{I}_{\mathrm{n2}}$ of these noise

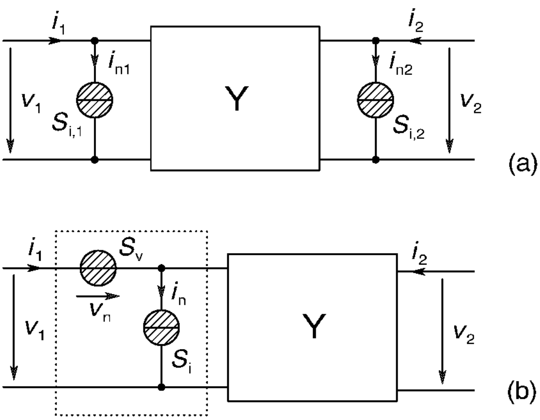


Fig. 1.80. Noisy two-port represented as a (**a**) noise-free two-port with noise current sources parallel to the input and output port, and (**b**) as a noise-free two-port with input-referred noise voltage and noise current sources

sources, the complex effective small-signal terminal currents are determined by

$$\begin{pmatrix} \underline{I}_1 \\ \underline{I}_2 \end{pmatrix} = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix} \begin{pmatrix} \underline{V}_1 \\ \underline{V}_2 \end{pmatrix} + \begin{pmatrix} \underline{I}_{\mathrm{n1}} \\ \underline{I}_{\mathrm{n2}} \end{pmatrix} . \tag{1.168}$$

Often it is helpful to describe the noise of the two-port in terms of noise sources connected to the input of the two-port, as shown in Fig. 1.80b. Here, a noise-free two-port is in cascade to a noise two-port composed of the input-referred current and voltage noise generator with complex effective values

$\underline{I}_\mathrm{n}$ and $\underline{V}_\mathrm{n}$. In terms of the chain parameters (or $ABCD$ parameters) the complex effective small-signal terminal currents then are written as

$$\begin{pmatrix} \underline{V}_1 \\ \underline{I}_1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} \underline{V}_2 \\ -\underline{I}_2 \end{pmatrix} + \begin{pmatrix} \underline{V}_\mathrm{n} \\ \underline{I}_\mathrm{n} \end{pmatrix} . \tag{1.169}$$

If the input-referred noise sources depend on the same physical mechanism to a certain extent, the noise sources will show correlation. Then, it is possible to separate the input-referred noise current i_n into the portion $i_\mathrm{n}^{(2)}$, represented by the complex effective value $\underline{I}_\mathrm{n}^{(2)} = Y_\gamma \underline{V}_\mathrm{n}$, which is completely correlated with v_n, and the portion $i_\mathrm{n}^{(1)}$, which shows no correlation with v_n. The admittance Y_γ is expressed in terms of the complex correlation coefficient $\underline{\gamma}_\mathrm{iv} = \langle \underline{I}_\mathrm{n} \underline{V}_\mathrm{n}^* \rangle / (I_\mathrm{n} V_\mathrm{n})$ of i_n and v_n as follows:

$$Y_\gamma = \underline{\gamma}_\mathrm{iv} \sqrt{S_i/S_v} \, . \tag{1.170}$$

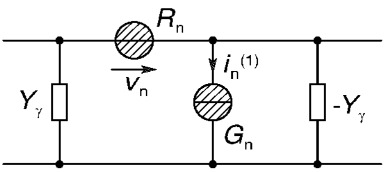


Fig. 1.81. Equivalent noise two-port using uncorrelated noise sources

The noise two-port may be replaced by the two-port depicted in Fig. 1.81, with uncorrelated noise sources v_n and $i_\mathrm{n}^{(1)}$ [65]. The noise sources are often expressed in terms of their equivalent noise conductance

$$G_\mathrm{n} = S_i^{(1)}/4k_\mathrm{B}T \tag{1.171}$$

and their equivalent noise resistance

$$R_\mathrm{n} = S_v/4k_\mathrm{B}T \, . \tag{1.172}$$

Then, four parameters (R_n, G_n and the real and imaginary components of Y_γ) have to be known for a complete specification of the input-referred noise two-port.

Figure 1.82 shows a noisy linear two-port to which a signal generator delivers the signal power ΔP_S1 within the frequency interval between f and $f+\Delta f$. This signal is superposed by a noise signal that delivers the noise power ΔP_N1 to the load in the frequency interval considered. At the input port we therefore observe the signal-to-noise ratio $\Delta P_\mathrm{S1}/\Delta P_\mathrm{N1}$. The power delivered to the load within the frequency interval $[\,f, f+\Delta f\,]$ may also be separated into a signal portion ΔP_S2 and a noise portion ΔP_N2, resulting in the signal-to-noise ratio $\Delta P_\mathrm{S2}/\Delta P_\mathrm{N2}$ at the output port. The noise factor F is defined as the quotient of the signal-to-noise ratio at the input and output port

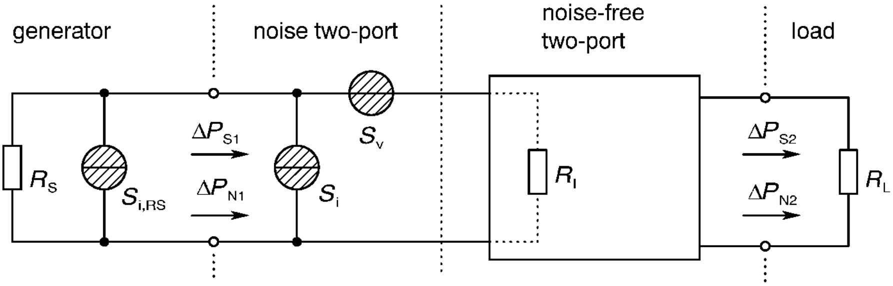


Fig. 1.82. Noisy linear two-port in amplifier operation

$$F = \frac{\Delta P_{S1}/\Delta P_{N1}}{\Delta P_{S2}/\Delta P_{N2}} ; \tag{1.173}$$

it defines the noise figure according to

$$NF = 10\,\mathrm{dB} \times \log_{10}(F) . \tag{1.174}$$

If a signal passes through a noise-free two-port, its signal-to-noise ratio will not change, i.e., it may be attributed the noise factor $F = 1$ corresponding to the noise figure $NF = 0$ dB. The noise figure thus defined is a temperature-dependent quantity and commonly specified for the standard noise temperature $T = 290$ K, specified by the IRE in 1962 [66].

If R_I denotes the input resistance of the circuit shown in Fig. 1.82, the noise power delivered by the source to the two-port within a frequency interval of width Δf is

$$\Delta P_{N1} = \frac{R_I R_S^2}{(R_I + R_S)^2} S_{iRS} \Delta f = \frac{4k_B T R_S R_I}{(R_I + R_S)^2} \Delta f . \tag{1.175}$$

The noise power delivered to the amplifier by the noise two-port with uncorrelated noise current and noise voltage sources of spectral density S_i and S_v is

$$\Delta P'_{N1} = \frac{R_I R_S^2}{(R_I + R_S)^2} S_i \Delta f + \frac{R_I}{(R_I + R_S)^2} S_v \Delta f . \tag{1.176}$$

Both ΔP_{N1} and $\Delta P'_{N1}$ are amplified in the noise-free two-port; if the amplifier has a power gain $G_p = \Delta P_{S2}/\Delta P_{S1}$, the noise power available at the output therefore is $\Delta P_{N2} = G_p(\Delta P_{N1} + \Delta P'_{N1})$. This yields for the noise factor

$$F = \frac{\Delta P_{S1}}{\Delta P_{S2}} \frac{\Delta P_{N2}}{\Delta P_{N1}} = \frac{1}{G_p} \frac{G_p(\Delta P_{N1} + \Delta P'_{N1})}{\Delta P_{N1}} = 1 + \frac{\Delta P'_{N1}}{\Delta P_{N1}}$$

or, considering (1.175) and (1.176),

$$F = 1 + \frac{R_S\, S_i}{4k_B T} + \frac{S_v}{4k_B T R_S} . \tag{1.177}$$

The noise power delivered by the amplifier will change, if the input-referred noise sources are correlated. The noise current $i_\mathrm{n}(t)$ and the noise voltage $v_\mathrm{n}(t)$ cause the noise current

$$i_\mathrm{ni}(t) = \frac{R_\mathrm{S} i_\mathrm{n}(t) + v_\mathrm{n}(t)}{R_\mathrm{S}+R_\mathrm{I}}$$

to flow into the input of the amplifier, i.e., the rms value of the noise power delivered to the input of the amplifier is

$$\begin{aligned}\Delta P'_\mathrm{N1} &= R_\mathrm{I} I_\mathrm{ni}^2 = \frac{R_\mathrm{I}}{(R_\mathrm{S}+R_\mathrm{I})^2}\,\overline{[R_\mathrm{S} i_\mathrm{n}(t)+v_\mathrm{n}(t)]^2} \\ &= \frac{R_\mathrm{I}}{(R_\mathrm{S}+R_\mathrm{I})^2}\left(R_\mathrm{S}^2 I_\mathrm{n}^2 + V_\mathrm{n}^2 + 2R_\mathrm{S}\overline{i_\mathrm{n}(t)v_\mathrm{n}(t)}\right) .\end{aligned}$$

Since $I_\mathrm{n} = \sqrt{S_i \Delta f}$ and $V_\mathrm{n} = \sqrt{S_v \Delta f}$ within a frequency interval of width Δf, (1.176) for $\Delta P'_\mathrm{N1}$ can be modified to read as follows

$$\Delta P'_\mathrm{N1} = \frac{R_\mathrm{I}}{(R_\mathrm{I}+R_\mathrm{S})^2}\left(R_\mathrm{S}^2 S_i + S_v + 2\gamma_\mathrm{iv} R_\mathrm{S}\sqrt{S_i S_v}\right)\Delta f ,$$

if the average $\overline{i_\mathrm{n}(t)v_\mathrm{n}(t)}$ is expressed in terms of the correlation coefficient $\gamma_\mathrm{iv} = \overline{i_\mathrm{n}(t)v_\mathrm{n}(t)}/(I_\mathrm{n}V_\mathrm{n})$. In the case of non-negligible correlation, the noise factor therefore reads

$$F = 1 + \frac{R_\mathrm{S}^2 S_i + S_v + 2\gamma_\mathrm{iv} R_\mathrm{S}\sqrt{S_i S_v}}{4k_\mathrm{B} T R_\mathrm{S}} . \tag{1.178}$$

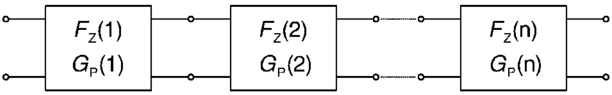


Fig. 1.83. Series connection of noisy linear two-ports

If noisy two-ports are cascaded (Fig. 1.83), the noise factor of the chain is given by the formula of Friis [65]:

$$F = F_1 + \frac{F_2 - 1}{G_p(1)} + \frac{F_3 - 1}{G_p(1)G_p(2)} + \dots , \tag{1.179}$$

where F_k denotes the noise factor of the kth two-port and $G_p(k)$ its power gain. If an amplifier is composed of several amplifying stages, it is therefore important to obtain an input stage with a large power gain, which to a large extent suppresses the noise contributions of the following stages.

1.8.4 Bipolar-Transistor Noise Equivalent Circuit

Noise in bipolar transistors is attributed to shot noise and thermal noise. In addition, at low frequencies $1/f$ noise may be important. A noise equivalent circuit that considers these mechanisms with appropriate noise sources is shown in Fig. 1.84. This model is used automatically in SPICE if a .NOISE

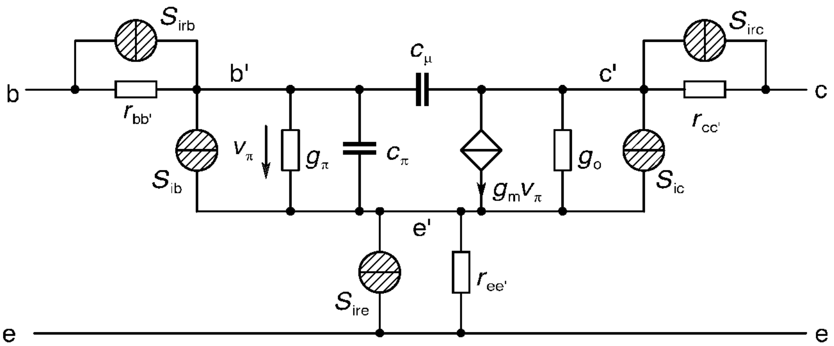


Fig. 1.84. Noise equivalent circuit of the bipolar transistor

analysis is performed. Noise current sources with spectral densities $S_{i\mathrm{rb}} = 4k_\mathrm{B}T/r_{\mathrm{bb}'}$, $S_{i\mathrm{re}} = 4k_\mathrm{B}T/r_{\mathrm{ee}'}$ and $S_{i\mathrm{rc}} = 4k_\mathrm{B}T/r_{\mathrm{cc}'}$ operate parallel to each of the series resistances of the device. Base shot noise and $1/f$ noise are described by a noise current source with spectral density

$$S_{i\mathrm{b}} = 2eI_\mathrm{B} + \left(\frac{I_\mathrm{B}}{\mathrm{A}}\right)^{A_\mathrm{F}} \frac{K_\mathrm{F}}{f} \,. \tag{1.180}$$

The exponent A_F is larger than one, i.e., the $1/f$ noise will show an increase with base current stronger than that of shot noise. Preamplifiers for low-frequency signals, which might suffer from $1/f$ noise, therefore generally are operated with a small collector current. The shot noise of the transfer current component is described in terms of the noise current source with spectral density $S_{i\mathrm{c}} = 2eI_\mathrm{C}$, which depends on the transfer current I_C at the bias point.

1.8.5 Input-Referred Noise Sources

Figure 1.85a shows the (simplified) noise equivalent circuit of a bipolar transistor with a short circuit at the output. This noisy two-port can be separated into a noise-free two-port in series with a noise two-port composed of the input-referred noise current source with spectral density S_i and the input-referred noise voltage source with spectral density S_v (Fig. 1.85b). The input-referred noise sources must give the same spectral density $S_{i\mathrm{a}}$ of the output current noise as the noise equivalent circuit Fig. 1.85a, in any circuit configuration. The correlation between the input-referred noise sources is neglected.

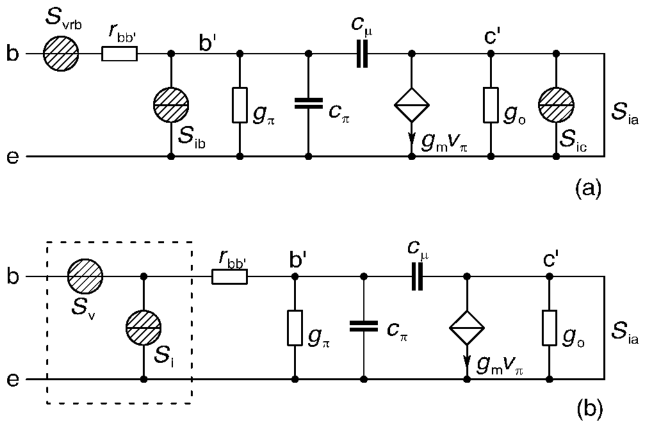


Fig. 1.85. (**a**) Simplified noise equivalent circuit for a bipolar transistor with a short circuit at the output, (**b**) corresponding noise equivalent circuit with input-referred noise sources

Input-Referred Noise Voltage

As a first step, the spectral density S_v of the input-referred noise voltage source is calculated. For that purpose the input is short-circuited and the complex effective value of the noise current at the output is calculated from the elementary noise sources with the help of the superposition theorem. Subcircuits (**a**) (**c**) in the left column of Fig. 1.86 yield the transfer factors $\underline{I}_{\mathrm{a}}/\underline{V}_{\mathrm{nb}} = -y_{21\mathrm{e}}$, $\underline{I}_{\mathrm{a}}/\underline{I}_{\mathrm{nb}} = -r_{\mathrm{bb'}}y_{21\mathrm{e}}$ and $\underline{I}_{\mathrm{a}}/\underline{I}_{\mathrm{nc}} = -1$, where

$$y_{21\mathrm{e}} = \frac{g_{\mathrm{m}} - \mathrm{j}\omega c_\mu}{1 + r_{\mathrm{bb'}}g_\pi + \mathrm{j}\omega r_{\mathrm{bb'}}(c_\pi + c_\mu)} . \tag{1.181}$$

By application of the superposition principle, the spectral density $S_{i\mathrm{a}}$ of the noise current becomes

$$S_{i\mathrm{a}} = |y_{21\mathrm{e}}|^2 S_{v\mathrm{rb}} + r_{\mathrm{bb'}}^2 |y_{21\mathrm{e}}|^2 S_{i\mathrm{b}} + S_{i\mathrm{c}} .$$

This has to be equal to the result $S_{i\mathrm{a}} = |y_{21\mathrm{e}}|^2 S_v$ obtained from subcircuit (d), and therefore

$$S_v = S_{v\mathrm{rb}} + r_{\mathrm{bb'}}^2 S_{i\mathrm{b}} + S_{i\mathrm{c}}/|y_{21\mathrm{e}}|^2 . \tag{1.182}$$

As long as $\omega c_\mu \ll g_{\mathrm{m}}$ and $r_{\mathrm{bb'}}g_\pi \ll 1$ are fulfilled, we may use the approximation $|y_{21\mathrm{e}}|^2 \approx g_{\mathrm{m}}^2/(1+f^2/f_{\mathrm{y}}^2)$, where $f_{\mathrm{y}} = 1/[2\pi r_{\mathrm{bb'}}(c_\pi + c_\mu)]$ is the transconductance cutoff frequency for the simplifying conditions considered ($r_{\mathrm{ee'}} = r_{\mathrm{cc'}} = 0$). Introducing the spectral densities of the individual noise sources gives

$$S_v \approx 4k_{\mathrm{B}}Tr_{\mathrm{bb'}} + r_{\mathrm{bb'}}^2 \left[2eI_{\mathrm{B}} + \frac{K_{\mathrm{F}}}{f}\left(\frac{I_{\mathrm{B}}}{\mathrm{A}}\right)^{A_{\mathrm{F}}} \right] + \frac{2eV_{\mathrm{T}}^2}{I_{\mathrm{C}}}\left(1 + \frac{f^2}{f_{\mathrm{y}}^2}\right) . \tag{1.183}$$

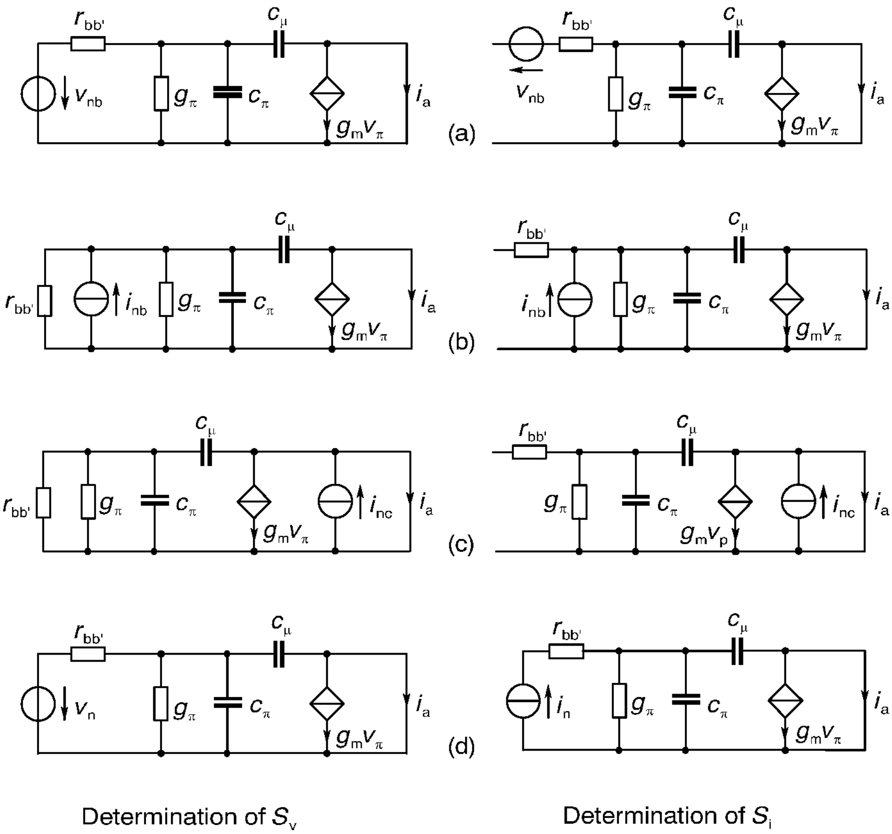


Fig. 1.86. (**a**) - (**d**) Subcircuits used for the computation of the effective two-port noise current and noise voltage sources

For frequencies $f \ll f_{cv1}$, with

$$f_{cv1} = \frac{K_F}{4k_BTr_{bb'} + 2eI_Br_{bb'}^2 + 2eV_T^2/I_C}\left(\frac{I_B}{A}\right)^{A_F}, \tag{1.184}$$

$1/f$ noise will dominate, and S_v will decrease in proportion to $1/f$; then, its value will be approximately constant over several decades of the frequency axis and finally increase in proportion to f^2 if $f \gg f_{cv2}$, where

$$f_{cv2} = f_y\sqrt{1 + 2r_{bb'}I_C/V_T + r_{bb'}^2 I_C I_B/V_T^2}\,, \tag{1.185}$$

as is illustrated in Fig. 1.87.

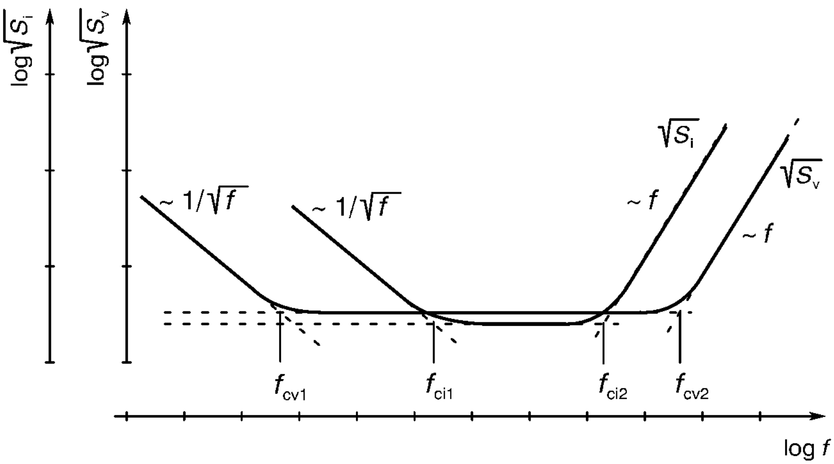


Fig. 1.87. Spectral densities of the noise equivalent sources of a bipolar transistor as a function of frequency

Input-Referred Noise Current

For the determination of S_i, the input will be left open. From subcircuits (**a**) – (**c**) in the right column of Fig. 1.86 we obtain the transfer factors $\underline{I}_a/\underline{V}_{nb} = 0$, $\underline{I}_a/\underline{I}_{nb} = -h_{21e}$ and $\underline{I}_a/\underline{I}_{nc} = -1$, where

$$h_{21e} = \frac{g_m - j\omega c_\mu}{g_\pi + j\omega(c_\pi + c_\mu)} \approx \frac{\beta}{1 + j\, f/f_\beta}\,, \tag{1.186}$$

if $\omega c_\mu \ll g_m$. Together with the result obtained with subcircuit (d),

$$S_{ia} = |h_{21e}|^2 S_{ib} + S_{ic} = |h_{21e}|^2 S_i\,,$$

has to be fulfilled, resulting in

$$S_i = S_{ib} + \frac{g_\pi^2 + \omega^2(c_\pi + c_\mu)^2}{g_m^2} S_{ic} = S_{ib} + \left[\frac{1}{\beta^2} + \left(\frac{f}{f_T}\right)^2\right] S_{ic}\,,$$

where $f_T \approx g_m/2\pi(c_\pi + c_\mu)$ denotes the cutoff frequency (neglecting emitter and collector series resistances). Introducing the spectral densities of the individual sources therefore yields

$$S_i \approx \left(\frac{I_B}{A}\right)^{A_F} \frac{K_F}{f} + 2eI_B + 2eI_C \left(\frac{1}{\beta^2} + \frac{f^2}{f_T^2}\right)\,. \tag{1.187}$$

Since, generally, $2eI_B \gg 2eI_C/\beta^2$ is fulfilled, the second term will dominate in a middle frequency band, with S_i independent of frequency. The frequency dependence of the input-referred spectral density S_i is determined by $1/f$ noise if $f \ll f_{ci1}$, where

$$f_{\mathrm{ci1}} \approx \frac{K_{\mathrm{F}}}{2eI_{\mathrm{B}}}\left(\frac{I_{\mathrm{B}}}{\mathrm{A}}\right)^{A_{\mathrm{F}}} \tag{1.188}$$

if $\beta \ll 1$. For large frequency values, $f \gg f_{\mathrm{ci2}}$, where

$$f_{\mathrm{ci2}} \approx f_{\mathrm{T}}/\sqrt{B_{\mathrm{N}}}\,, \tag{1.189}$$

the value of S_i increases in proportion to f^2, as is illustrated in Fig. 1.87.

1.8.6 Noise Figure

According to (1.178) the noise factor F of a transistor driven with generator series resistance R_{S} is

$$F = 1 + \frac{R_{\mathrm{S}}^2 S_i + S_v}{4k_{\mathrm{B}}TR_{\mathrm{S}}} \tag{1.190}$$

if correlation between the input-referred noise sources is neglected; its value allows the computation of the noise figure NF according to (1.174). Introducing the results (1.183) and (1.187) for S_{v} and S_i yields the following for the noise factor:

$$\begin{aligned} F &= 1 + \frac{1}{R_{\mathrm{S}}}\left[r_{\mathrm{bb'}}\left(1+\frac{r_{\mathrm{bb'}}I_{\mathrm{B}}}{2V_{\mathrm{T}}}\right) + \frac{V_{\mathrm{T}}}{2I_{\mathrm{C}}}\left(1+\frac{f^2}{f_{\mathrm{y}}^2}\right)\right] \\ &\quad + R_{\mathrm{S}}\frac{I_{\mathrm{B}}}{2V_{\mathrm{T}}}\left(1+B_{\mathrm{N}}\frac{f^2}{f_{\mathrm{T}}^2}\right)\,, \end{aligned} \tag{1.191}$$

if $1/f$ noise is neglected and $I_{\mathrm{B}} \gg I_{\mathrm{C}}/\beta^2$ is assumed. With growing values of R_{S}, the second term on the right-hand side (proportional to $1/R_{\mathrm{S}}$) decreases, while the third term (proportional to R_{S}) increases; consequently the noise factor F will show a minimum value. The value of generator series resistance that yields the minimum noise figure is determined from the condition $\mathrm{d}F/\mathrm{d}R_{\mathrm{S}} = 0$. If $f < f_{\mathrm{y}}$ and $f < f_{\mathrm{ci2}} \approx f_{\mathrm{T}}/\sqrt{B_{\mathrm{N}}}$ may be assumed, the optimum source resistance becomes

$$R_{\mathrm{Sopt}} = \sqrt{\frac{2r_{\mathrm{bb'}}V_{\mathrm{T}}}{I_{\mathrm{B}}}\left(1+\frac{r_{\mathrm{bb'}}I_{\mathrm{B}}}{2V_{\mathrm{T}}}\right) + \frac{V_{\mathrm{T}}^2}{B_{\mathrm{N}}I_{\mathrm{B}}^2}}\,; \tag{1.192}$$

note that its value decreases with increasing collector current. In the case of a negligible base resistance ($r_{\mathrm{bb'}} \to 0$), this expression may be further simplified to $R_{\mathrm{Sopt}} \approx \sqrt{B_{\mathrm{N}}}\,V_{\mathrm{T}}/I_{\mathrm{C}}$; this result is suitable for crude estimates of the optimum generator series resistance. An order-of-magnitude estimate of the noise voltages and currents, together with the noise figure of a bipolar transistor, is presented in the following section.

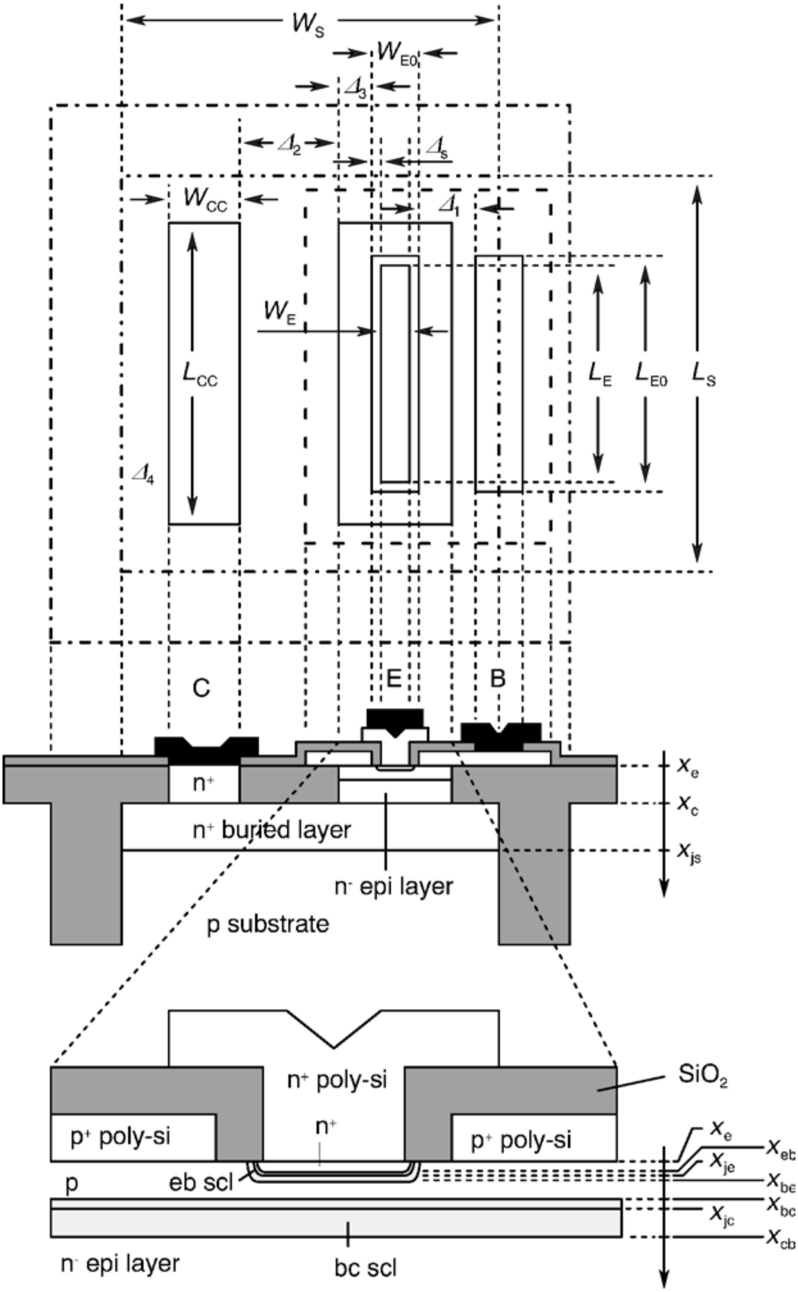


Fig. 1.88. Layout and cross section of an example bipolar transistor

1.9 Orders of Magnitude

As an example, an npn bipolar transistor with a cross section and layout as depicted in Fig. 1.88 is considered. The eb junction depth x_{je} and the bc junction depth x_{jc} are given as 50 nm and 125 nm respectively, whereas the epi-layer thickness is taken to be 800 nm; homogeneous doping of each region is assumed for simplicity. To estimate the transistor parameters at $T = 300\,\mathrm{K}$, we use the following data [67,68]:

Region	Doping (cm^{-3})	Thickness (nm)	μ_{n} ($\mathrm{cm}^2/\mathrm{Vs}$)	μ_{p} ($\mathrm{cm}^2/\mathrm{Vs}$)	ρ ($\Omega\,\mu\mathrm{m}$)
Emitter	5×10^{19}	50	72	150	17.4
Base	2×10^{18}	75	270	117	267
Collector	6×10^{16}	675	845	–	1230
Subcollector	1×10^{19}	1000	108	–	58
Substrate	1×10^{15}	350 µm	–	470	1.3×10^5

The transistor–transistor isolation is assumed to be realized with trenches, resulting in a collector substrate diode of width $W_{\mathrm{S}} = 7\,\mu\mathrm{m}$ and length $L_{\mathrm{S}} = 8.4\mu\mathrm{m}$. Assuming a spacer width $\Delta_{\mathrm{S}} = 0.25\mu\mathrm{m}$, the emitter area is calculated from the opening in the p$^+$-poly layer, with width $W_{\mathrm{E0}} = 1\,\mu\mathrm{m}$ and length $L_{\mathrm{E0}} = 5\,\mu\mathrm{m}$ to $W_{\mathrm{E}} = 0.5\,\mu\mathrm{m}$ and $L_{\mathrm{E}} = 4.5\,\mu\mathrm{m}$. The extensions W_{C} and L_{C} of the bc diode are determined by W_{E0} and L_{E0} together with the overlap $\Delta_3 = 0.7\,\mu\mathrm{m}$ of the p$^+$-poly layer and monocrystalline silicon as follows:

$$W_{\mathrm{C}} = W_{\mathrm{E0}} + 2\Delta_3 = 2.4\,\mu\mathrm{m} \quad \text{and} \quad L_{\mathrm{C}} = L_{\mathrm{E0}} + 2\Delta_3 = 6.4\,\mu\mathrm{m}\,.$$

The built-in voltage and the depletion capacitance of the eb diode are estimated from the doping concentrations N_{DE} and N_{AB} in emitter and base as follows[44]

$$V_{\mathrm{JE}} = V_{\mathrm{T}} \ln\left(\frac{N_{\mathrm{AB}} N_{\mathrm{DE}}}{n_{\mathrm{i}}^2}\right) = 1070\,\mathrm{mV}$$

$$C_{\mathrm{JE}} = L_{\mathrm{E}} W_{\mathrm{E}} \sqrt{\frac{e\epsilon_{\mathrm{Si}} N_{\mathrm{AB}} N_{\mathrm{DE}}}{2(N_{\mathrm{AB}}+N_{\mathrm{DE}})V_{\mathrm{JE}}}} = 8.8\,\mathrm{fF}\,.$$

The built-in voltage and the depletion capacitance of the bc diode result from the doping concentrations N_{AB} and N_{DC} in base and collector

[44] We neglect heavy doping effects (see Sect. 2.5) in this simplified analysis, except for the computation of the current gain.

$$V_{\mathrm{JC}} = V_{\mathrm{T}} \ln\left(\frac{N_{\mathrm{AB}} N_{\mathrm{DC}}}{n_{\mathrm{i}}^2}\right) = 896\,\mathrm{mV}$$

$$C_{\mathrm{JC}} = W_{\mathrm{C}} L_{\mathrm{C}} \sqrt{\frac{e\epsilon_{\mathrm{Si}} N_{\mathrm{AB}} N_{\mathrm{DC}}}{2(N_{\mathrm{AB}}+N_{\mathrm{DC}}) V_{\mathrm{JC}}}} = 11.4\,\mathrm{fF}\,.$$

Only the portion

$$X_{\mathrm{CJC}} = \frac{W_{\mathrm{E}} L_{\mathrm{E}}}{W_{\mathrm{C}} L_{\mathrm{C}}} \approx 0.146$$

of the bc diode is part of the internal transistor.

The built-in voltage and the depletion capacitance of the cs junction are determined by the doping concentrations N_{DSC} and N_{AS} in subcollector and substrate as follows:

$$V_{\mathrm{JS}} = V_{\mathrm{T}} \ln\left(\frac{N_{\mathrm{AS}} N_{\mathrm{DSC}}}{n_{\mathrm{i}}^2}\right) = 832\,\mathrm{mV}$$

$$C_{\mathrm{JS}} = W_{\mathrm{S}} L_{\mathrm{S}} \sqrt{\frac{e\epsilon_{\mathrm{Si}} N_{\mathrm{AS}} N_{\mathrm{DSC}}}{2(N_{\mathrm{AS}}+N_{\mathrm{DSC}}) V_{\mathrm{JS}}}} = 5.9\,\mathrm{fF}\,.$$

Since the pn junctions are assumed to be abrupt, all of the gradation exponents M_{JE}, M_{JC} and M_{JS}, equal 0.5. The extension of the eb space-charge layer into the p-type base region is voltage dependent and given by

$$x_{\mathrm{be}} - x_{\mathrm{je}} = \sqrt{\frac{2\epsilon_{\mathrm{Si}} N_{\mathrm{DE}}(V_{\mathrm{JE}} - V_{\mathrm{BE}})}{e\, N_{\mathrm{AB}}(N_{\mathrm{AB}}+N_{\mathrm{DE}})}} = 26\,\mathrm{nm} \times \sqrt{1 - \frac{V_{\mathrm{BE}}}{V_{\mathrm{JE}}}}\,,$$

while the extension of the bc space-charge layer into the p-type base region is

$$x_{\mathrm{jc}} - x_{\mathrm{bc}} = \sqrt{\frac{2\epsilon_{\mathrm{Si}} N_{\mathrm{DC}}(V_{\mathrm{JC}} - V_{\mathrm{BC}})}{e\, N_{\mathrm{AB}}(N_{\mathrm{AB}}+N_{\mathrm{DC}})}} = 4.1\,\mathrm{nm} \times \sqrt{1 - \frac{V_{\mathrm{BC}}}{V_{\mathrm{JC}}}}\,.$$

The base width at $V_{\mathrm{BE}} = V_{\mathrm{BC}} = 0$ therefore is

$$d_{\mathrm{B0}} = x_{\mathrm{bc}} - x_{\mathrm{be}} = 75\,\mathrm{nm} - 26\,\mathrm{nm} - 4.1\,\mathrm{nm} \approx 45\,\mathrm{nm}\,.$$

This gives the following for the base charge at $V_{\mathrm{BE}} = V_{\mathrm{BC}} = 0$:

$$Q_{\mathrm{B0}} = e N_{\mathrm{AB}} L_{\mathrm{E}} W_{\mathrm{E}} d_{\mathrm{B}} = 3.24 \times 10^{-14}\,\mathrm{As}\,,$$

corresponding to approximately 2×10^5 holes. Using

$$D_{\mathrm{n}} = V_{\mathrm{T}} \mu_{\mathrm{n}} = 25.84\,\mathrm{mV} \times 270\,\mathrm{cm}^2/\mathrm{Vs} \approx 7\,\mathrm{cm}^2/\mathrm{s}$$

gives

$$I_S = \frac{eD_n n_i^2}{N_{AB}} \frac{W_E L_E}{d_{B0}} \approx 2.9 \times 10^{-19}\,\mathrm{A}$$

for the transfer saturation current and

$$\tau_{B0} = \frac{d_B^2}{2D_n} = 1.4\,\mathrm{ps}$$

for the base transit time at $V_{BE} = V_{BC} = 0$. In real transistors a somewhat smaller value of I_S and a larger value of τ_B are to be expected due to deviations from the simple diffusion law. The forward transit time τ_f is composed of the base transit time and the collector transit time (see Sect. 3.5), $\tau_{jc} = d_{jc}/2\,v_{nsat}$. Since $I_C \sim 1/Q_B \approx 1/(Q_{B0} + Q_{JE} + Q_{JC})$, one obtains

$$\left(\frac{\partial I_C}{\partial V_{CE}}\right)_{V_{BE}} \approx -\frac{I_C}{Q_B}\frac{\mathrm{d}Q_{JC}}{\mathrm{d}V_{CB}} \approx I_C \frac{X_{CJC}C_{JC}}{Q_B} = \frac{I_C}{V_{CE}+V_{AF}} ,$$

which leads to the following estimate of the forward Early voltage:

$$V_{AF} \approx Q_{B0}/X_{CJC}C_{JC} \approx 19.5\,\mathrm{V} ;$$

the reverse Early voltage V_{AR} is estimated to be

$$V_{AR} \approx Q_{B0}/C_{JE} \approx 3.7\,\mathrm{V} .$$

The base current is assumed to be due to recombination at the emitter contact only, which is characterized by the effective recombination velocity S_p. The ideal forward current gain is then estimated with (1.105):

$$B_F = \frac{D_n}{D_p}\frac{N_{DE}}{N_{AB}}\left(\frac{n_{iB}}{n_{iE}}\right)^2 \frac{d_E + D_p/S_p}{d_{B0}} .$$

With an apparent bandgap narrowing (see Sect. 2.5) of $\Delta W_g \approx 85\,\mathrm{meV}$, one obtains $(n_{iB}/n_{iE})^2 \approx 0.037$; assuming $S_p \approx 5 \times 10^3$ cm/s yields $B_F \approx 290$ for the ideal forward current gain. In the case of a metal-contacted emitter ($S_p \to \infty$), the forward current gain $B_F \approx 1.8$ would result.

The emitter resistance is estimated from the specific contact resistance $\rho_C \approx 100\,\Omega\,\mu\mathrm{m}^2$ and the emitter area $A_{je} = L_E W_E$:

$$R_{EE'} \approx r_{ee'} \approx \rho_C/A_{je} \approx 44\,\Omega .$$

The collector resistance is composed of the resistance R_{sub} of the subcollector, and the resistance R_{epi} of the epitaxial collector region. R_{sub} is estimated by

$$R_{sub} = \frac{\rho_C + \rho d_{epi}}{L_{CC}W_{CC}} + \frac{\rho}{d_{bl}L_S}\left(\Delta_2 + \frac{W_{CC}}{2} + \Delta_3 + \frac{W_{E0}}{2}\right) \approx 42\,\Omega ,$$

using $L_{CC} = 6.4\mu\mathrm{m}$, $W_{CC} = 1.5\mu\mathrm{m}$, $\Delta_2 = 2\mu\mathrm{m}$, $\Delta_3 = 0.7\mu\mathrm{m}$ and $\rho = 58\Omega\,\mu\mathrm{m}$. R_{epi} is estimated by

$$R_{\rm epi} \; = \; \frac{\rho_{\rm epi}(x_{\rm c}-x_{\rm cb})}{W_{\rm C}L_{\rm C}} \; \approx \; 54\,\Omega\left(1-0.21\sqrt{1-\frac{V_{\rm BC}}{V_{\rm JC}}}\right) \; ;$$

$R_{\rm epi} = 0$ if this formula gives negative results, indicating that the depletion-layer edge reaches through to the buried layer. The base resistance is roughly estimated considering the base contact resistance (area of base contact assumed to equal $L_{\rm E0}W_{\rm E0}$), the series resistance below the spacer oxide and (1.115) for the two-sided base contact:

$$R_{\rm BB'} \; \approx \; r_{\rm bb'} \; \approx \; \frac{\rho_{\rm C}}{W_{\rm E0}L_{\rm E0}} + \frac{\rho\Delta_{\rm S}}{2\,L_{\rm E}d_{\rm B0}} + \frac{\rho W_{\rm E}}{12\,L_{\rm E}d_{\rm B0}} \approx 212\,\Omega \; ,$$

neglecting the resistance of the (silicided) polysilicon sheet, current crowding at the emitter edge and conductivity modulation of the base charge layer.

With these data, the cutoff frequencies can be estimated for a certain point of operation. If, for example, $I_{\rm C} = 1$ mA and $V_{\rm CE} = 3$ V, one obtains $g_{\rm m} = I_{\rm C}/V_{\rm T} \approx 38.6$mS and $V_{\rm BE} \approx V_{\rm T}\ln(I_{\rm C}/I_{\rm S}) \approx 926$mV, resulting in $V_{\rm CB} = 2.07$V and the current gain

$$B_{\rm N} \; \approx \; B_{\rm F}\,(1 + V_{\rm CB}/V_{\rm AF} - V_{\rm EB}/V_{\rm AR}) \; \approx \; 248 \; .$$

The base width at the bias point is

$$d_{\rm B} \; = \; 75\,{\rm nm} - 26\,{\rm nm}\sqrt{1-\frac{0.926}{1.07}} - 4.1\,{\rm nm}\sqrt{1+\frac{2.07}{0.898}} \; \approx \; 58\,{\rm nm} \; ,$$

resulting in a base transit time $\tau_{\rm B} = 2.4$ ps. The bc depletion-layer width at the bias point is

$$d_{\rm jc} \; = \; \sqrt{\frac{2\epsilon_{\rm Si}(N_{\rm DC}+N_{\rm AB})(V_{\rm JC}+V_{\rm CB})}{e\,N_{\rm AB}N_{\rm DC}}} \; = \; 259\,{\rm nm} \; ,$$

corresponding to a bc depletion-layer transit time $\tau_{\rm jc} = d_{\rm jc}/2v_{\rm nsat} = 1.3$ ps. From this, the forward transit time $\tau_{\rm f} = \tau_{\rm B} + \tau_{\rm jc} = 3.7$ ps results. With $g_\pi \approx g_{\rm m}/B_{\rm N} \approx 156\mu$S, $c_{\rm je} \approx 24$fF, $\tau_{\rm f}g_{\rm m} \approx 143$fF, $c_\pi = 167$fF, $c_{\rm jc} = c_\mu \approx 6.2$fF, $r_{\rm ee'} \approx 44\,\Omega$ and $r_{\rm cc'} \approx 75\,\Omega$, one obtains

$$f_{\rm T} \; = \; \frac{1}{2\pi\left[\tau_{\rm f} + \dfrac{c_{\rm je}+c_{\rm jc}}{g_{\rm m}} + (r_{\rm ee'}+r_{\rm cc'})c_{\rm jc}\right]} \; \approx \; 30.5\;{\rm GHz}$$

for the cutoff frequency, corresponding to a β-cutoff frequency $f_\beta \approx f_{\rm T}/B_{\rm N} \approx 123$ MHz. The transconductance cutoff frequency is

$$f_{\rm y} \; = \; \frac{1 + r_{\rm bb'}\,g_\pi}{2\pi r_{\rm bb'}\,(c_\pi + c_\mu)} \; \approx \; 4.5\;{\rm GHz} \; ,$$

considerably larger than the β-cutoff frequency. The maximum frequency of oscillation is

$$f_{\max} \approx \sqrt{\frac{f_{\mathrm{T}}}{8\pi r_{\mathrm{bb'}} c_{\mu}}} \approx 30.4\,\mathrm{GHz}\,.$$

Both f_{y} and $f_{\max}$ are somewhat overestimated, since the effect of the collector–substrate capacitance is not considered here.

In the frequency range between f_{cv1} and f_{cv2}, the spectral density of the input-referred noise current source is dominated by the term $4k_{\mathrm{B}}Tr_{\mathrm{bb'}}$. At $T = 290\,\mathrm{K}$, a noise voltage $V_{\mathrm{n}} \approx 1.9\,\mathrm{nV}$ results in a frequency band of 1 Hz. The spectral density of the input-referred noise current is dominated by $2eI_{\mathrm{B}}$ in the frequency interval between f_{ci1} and f_{ci2}, resulting in a noise current $I_{\mathrm{n}} = 1.1\,\mathrm{pA}$ in a frequency band of 1 Hz. According to (1.192), the minimum noise figure is obtained with the source impedance $R_{\mathrm{Sopt}} = 1.7\ \mathrm{k}\Omega$. With this value the noise factor $F = 1.26$ is obtained from (1.191) for the bias point chosen, resulting in a noise figure $N_{\mathrm{F}} \approx 1$ dB, which is close to what is observed experimentally in comparable transistors.

1.10 References

[1] J.L. Moll. The evolution of the theory for the voltage-current characteristic of pn junctions. *Proc. IRE*, 46:1076–1082, 1958.

[2] J.R. Pierce. The naming of the transistor. *Proc. IEEE*, 86(1):37–47, 1998.

[3] J. Bardeen, W.H. Brattain. The transistor – a semiconductor triode, Phys. Rev., 74:230–231, July 15, 1948; Reprinted in:. *Proc. IEEE*, 86(1):29–30, 1998.

[4] P.R. Morris. *A history of the world semiconductor industry.* P. Peregrinus, London, 1990.

[5] W. Shockley. The theory of pn junctions in semiconductors and pn junction transistors. *Bell Syst. Tech. J.*, 28:435–487, 1949.

[6] W. Shockley. Semiconductor amplifiers, patented Apr. 4, 1950, united states patent office no. 2,502,488; Reprinted in: *Proc. IEEE*, 86(1):34–36, 1998.

[7] W. Shockley, G.K. Teal, M. Sparks. Pn junction transistors. *Phys. Rev.*, 83(1):151–162, 1951.

[8] I.M. Ross. The invention of the transistor. *Proc. IEEE*, 86(1):7–28, 1998.

[9] J.A. Becker, J.N. Shive. The transistor – a new semiconductor amplifier. *The Electrical Engineer 68(3):215–221, March 1949)*, Reprinted in: *Proc. IEEE*, 87(8):1389–1396, 1999.

[10] J.M. Early. PNIP and NPIN junction transistor triodes. *Bell Syst. Tech. J.*, 33(3):517–533, 1954.

[11] H.H. Loor, J.J. Kleimock, H.C. Theurer, H.K. Christensen. New advantages in diffused devices. *Presented at the IRE/AIEC Solid State Device Research Corp., Pittsburgh, PA., June*, 1960.

[12] H. Murrmann. Modern bipolar technology for high-performance ICs. *Siemens Forsch.-u. Entwickl.-Ber.*, 5(6):353–359, 1976.

[13] J. Graul, H. Murrmann, A. Glasl. High-performance transistors with arsenic-implanted polysil emitters. *IEEE J. Solid-State Circuits*, 11(8):491 495, 1976.

[14] D.D. Tang, P.M. Solomon, T.H. Ning, R.D. Isaac, R.E. Burger. 1.25 μm deep-groove-isolated self-aligned bipolar circuits. *IEEE J. Solid-State Circuits*, 17(5):925–931, 1982.

[15] T. Sakai, Y. Kobayashi, H. Yamauchi, M. Sato, T. Makino. High speed bipolar ics using super self-aligned process technology. *Proc. 12th Conf. on Solid State Devices, Tokyo, Jpn. J. Appl. Phys.*, 20(Supplement 20-1):155–159, 1980.

[16] T.H. Ning, D.D. Tang. Bipolar trends. *Proc. IEEE*, 74(12):1669–1677, 1986.

[17] H. Kroemer. Theory of a wide-gap emitter for transistors. *Proc. IRE*, 45(11):1535–1537, 1957.

[18] H. Krömer. Heterostructure bipolar transistors and integrated circuits. *Proc. IEEE*, 70(1):13–25, 1982.

[19] S.S. Iyer, G.L. Patton, J.M.C. Stork, B.S. Meyerson, and D.L. Harame. Heterojunction bipolar transistors using Si-Ge alloys. *IEEE Trans. Electron Devices.*, 36(10):2043–2064, 1989.

[20] J.C. Bean. Silicon-based semiconductor heterostructures: column IV bandgap engineering. *Proc. IEEE*, 80(4):571–587, 1992.

[21] B.S. Meyerson. UHV/CVD growth of Si and Si:Ge alloys: Chemistry, physics and device applications. *Proc. IEEE*, 80(10):1592–1608, 1992.

[22] D.L. Harame, J.H. Comfort, J.D. Cressler, E.F. Crabbe, J.Y.-C. Sun, B.S. Meyerson, and T. Tice. Si/sige epitaxial-base transistors – part I: Materials, physics, and circuits. *IEEE Trans. Electron Devices.*, 42(3):455–468, 1995.

[23] D.L. Harame, J.H. Comfort, J.D. Cressler, E.F. Crabbe, J.Y.-C. Sun, B.S. Meyerson, T. Tice. Si/SiGe epitaxial-base transistors – part II: Process intgration and analog applications. *IEEE Trans. Electron Devices.*, 42(3):469–482, 1995.

[24] J.M. Early. Structure-determined gain-band product of junction triode transistors. *Proc. IRE*, 46:1924–1927, 1958.

[25] H.K. Gummel. A self-consistent iterative scheme for one-dimensional steady state transistor calculations. *IEEE Trans. Electron Devices.*, 11:455–465, 1964.

[26] J.J. Ebers, J.L. Moll . Large-signal behaviour of junction transistors. *Proc. IRE*, 42:1761 1772, 1954.

[27] J.L. Moll. Large-signal transient response of junction transistors. *Proc. IRE*, 42:1773–1784, 1954.

[28] H.K. Gummel, H.C. Poon. An integral charge control model of bipolar transistors. *Bell Syst. Tech. J.*, 49:827–852, 1970.

[29] S.K. Wiedmann. Advancements in bipolar VLSI circuits and technologies. *IEEE J. Solid-State Circuits*, 19(3):282–291, 1984.

[30] J.D. Warnock. Silicon bipolar device structures for digital applications: Technology trends and future directions. *IEEE Trans. Electron Devices.*, 42(3):377–389, 1995.

[31] E. Ohue, et al. A 7.7-ps CML using selective epitaxial SiGe HBTs. *IEEE BCTM Tech. Dig.*, 1998:97–100, 1998.

[32] K. Washio., et al. 82 GHz dynamic frequency divider in 5.5 ps ECL SiGe HBTs. *ISSCC Tech. Dig.*, 2000:12.8, 2000.

[33] C.A. Desoer, E.S. Kuh. *Basic Circuit Theory.* McGraw-Hill, New York, 1969.

[34] L.O. Chua, C.A. Desoer, E.S. Kuh. *Linear and Nonlinear Circuits.* McGraw-Hill, New York, 1991.

[35] M. Abramowith, I.A. Stegun. *Pocketbook of Mathematical Functions.* Harri Deutsch, Frankfurt, 1984.

[36] L.C. Peterson. Equivalent circuits of linear active four-terminal networks. *Bell Syst. Tech. J.*, 27(4):593 622, 1948.

[37] H.C. de Graaff, F.M. Klaassen. *Compact Transistor Modeling for Circuit Design.* Springer, Vienna, 1990.

[38] D.J. Hamilton, A.H. Marshak, F.A. Lindholm. *Principles and Applications of Semiconductor Device Modeling.* Holt, Rinehart and Winston, New York, 1971.

[39] S.O. Kasap. *Principles of Electrical Engineering Materials and Devices.* McGraw-Hill, Boston, 1997.

[40] S.M. Sze. *Semiconductor Devices, Physics and Technology.* J. Wiley, New York, 1985.

[41] D.J. Roulston. *Bipolar Semiconductor Devices.* McGraw-Hill, New York, 1990.

[42] W. Shockley. Transistor electronics: imperfections, unipolar and analog transistors. *Proc. IRE*, 40(11):1289–1313, 1952.

[43] W.E. Beadle, J.C.C. Tsai, R.D. Plummer (Eds.). *Quick Reference Manual for Silicon Integrated Circuit Technology.* J. Wiley, New York, 1985.

[44] M.A. Green. Intrinsic concentration, effective densities of states, and effective mass in silicon. *J. Appl. Phys.*, 67(6):2944–2954, 1990.

[45] A.B. Sproul, M.A. Green, J. Zhao. Improved value for the silicon intrinsic carrier concentration at 300 K. *Appl. Phys. Lett.*, 57(3):255–257, 1990.

[46] S.T. Pantelides. The electronic structure of impurities and other point defects in semiconductors. *Rev. Mod. Phys.*, 50(4):797–858, 1978.

[47] P.A. Tipler. *Physics for Scientists and Engineers.* Worth Publishers, New York, 3rd edn, 1991.

[48] W. van Roosbroeck. Theory of the flow of electrons and holes in germanium and other semiconductors. *Bell Syst. Tech. J.*, 29:560–607, 1950.

[49] H.C. de Graaff, J.W. Slotboom, A. Schmitz. The emitter efficiency of bipolar transistors. *Solid-State Electron.*, 20:515–521, 1977.

[50] S.M. Sze. *Physics of Semiconductor Devices*, 2nd edn. J. Wiley, New York, 1982.

[51] D.L. Heald, J.G. Skalnik, E.N. Nansen, P.F. Ordung. Thermodynamic considerations of pn junction capacitance. *Solid-State Electron.*, 16:1055–1065, 1973.

[52] J.R. Hauser, P.M. Dunbar. Minority carrier reflecting properties of semiconductor high-low junctions. *Solid-State Electron.*, 18:715–716, 1975.

[53] B. Gilbert. Translinear circuits: an historical overview. *Analog Integr. Circ. and Signal Process.*, 9:95–118, 1996.

[54] I.E. Getreu. *Modeling the Bipolar Transistor.* Tektronix, Beaverton, 1976.

[55] J.M. Early. Effects of space-charge layer widening in junction transistors. *Proc. IRE*, 40:1401–1406, 1952.

[56] R.J. Wilfinger. Predicting transistor turn-on delay time in the common emitter configuration. *Sol.St. Design*, 3:34–41, 1962.

[57] H.J. Kuno. Rise and fall time calculations of junction transistors. *IEEE Trans. Electron Devices.*, 11:151–155, 1964.

[58] Y. Hsia, F.-H. Wang. Switching waveform prediction of a simple transistor inverter circuit. *IEEE Trans. Electron Devices.*, 12:626–631, 1965.

[59] J.B. Johnson. Thermal agitation of electricity. *Bell Lab. Rec.*, 3(2):185–187, 1927.

[60] H. Nyquist. Thermal agitation of electric charge in conductors. *Phys. Rev.*, 32:110–113, 1928.

[61] C.D. Motchenbacher, J.A. Connelly. *Low-Noise Electronic System Design.* J. Wiley, New York, 1993.

[62] P. Dutta, P.M. Horn. Low-frequency fluctuations in solids: $1/f$ noise. *Rev. Mod. Phys.*, 53(3):497–516, 1981.

[63] A. van der Ziel. Unified presentation of $1/f$ noise in electronic devices: Fundamental $1/f$ noise sources. *Proc. IEEE*, 76(3):233–258, 1988.

[64] M. Reisch. *Elektronische Bauelemente – Funktion, Grundschaltungen, Modellierung mit SPICE.* Springer, Heidelberg, 1998.

[65] J. Engberg, T. Larsen. *Noise Theory of Linear and Nonlinear Circuits.* J. Wiley, Chichester, 1995.

[66] IRE Standards Committee. IRE standards on electron tubes: Definitions of terms, 1962. *Proc. IEEE*, 50(3):434–435, 1963.

[67] D.B.M. Klaassen. A unified mobility model for device simulation – I. model equations and concentration dependence. *Solid-State Electron.*, 35(7):953–959, 1992.

[68] D.B.M. Klaassen. A unified mobility model for device simulation – I. temperature dependence of carrier mobility and lifetime. *Solid-State Electron.*, 35(7):961–967, 1992.

Part II

PHYSICS AND MODELING OF HIGH-FREQUENCY BIPOLAR TRANSISTORS

- **Semiconductor Physics Required for Bipolar-Transistor Modeling**
- **Physics and Modeling of Bipolar Junction Transistors**
- **Physics and Modeling of Heterojunction Bipolar Transistors**
- **Noise Modeling**

2 Semiconductor Physics Required for Bipolar-Transistor Modeling

The classical description of the electromagnetic field is based on Maxwell's equations [1],

$$\nabla \cdot \boldsymbol{D} = \rho \,, \tag{2.1}$$

$$\nabla \cdot \boldsymbol{B} = 0 \,, \tag{2.2}$$

$$\nabla \times \boldsymbol{E} = -\partial \boldsymbol{B}/\partial t \,, \tag{2.3}$$

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + \partial \boldsymbol{D}/\partial t \,, \tag{2.4}$$

which describe the electric field $\boldsymbol{E}$, the dielectric displacement (or electric flux density) $\boldsymbol{D}$, the magnetic field $\boldsymbol{H}$ and the magnetic induction (or magnetic flux density) $\boldsymbol{B}$; the vector $\boldsymbol{J}$ describes the electric current density due to moving charges, and ρ describes the charge density.

Maxwell's equations allow the determination of the electromagnetic field caused by given charges and currents in a vacuum. In the presence of matter, additional constitutive relations, which describe the response (dielectric and magnetic polarization, current flow, etc.) of the material to applied fields, are necessary. The reason for this is that charges moving in solids interact not only with the macroscopic fields governed by Maxwell's equations, but also with other carriers and the crystal lattice. In the simplified description in the previous chapter, the constitutive relations were written as[1]

$$\boldsymbol{D} = \epsilon \boldsymbol{E} \,, \tag{2.5}$$

where $\epsilon = \epsilon_0 \epsilon_\mathrm{r}$ is the permittivity, ϵ_r the dielectric constant, and

$$\boldsymbol{J} = \boldsymbol{J}_\mathrm{n} + \boldsymbol{J}_\mathrm{p} = e(\mu_\mathrm{n} n + \mu_\mathrm{p} p)\boldsymbol{E} + eD_\mathrm{n}\nabla n - eD_\mathrm{p}\nabla p \tag{2.6}$$

is the electric current density due to moving charges. These equations hold only approximately, and their limitations need to be known if one is to correctly describe the physics of high-frequency bipolar transistors. In this chapter, we therefore look in more detail at the equations used for device modeling in the drift diffusion approximation. In addition to this an extended description, the so-called hydrodynamic model, which also considers energy transport is examined – this approach is particularly useful in the presence of strong electric fields.

[1]The third constitutive relation, $\boldsymbol{B} = \mu_0 \boldsymbol{H}$, which expresses the fact that semiconductors show only weak magnetic polarization, is not of great interest here, since magnetic-field effects have negligible influence on the behavior of high-frequency bipolar transistors.

2.1 Band Structure

Electrons in semiconductors are described in terms of the electronic band structure of the material. For modeling purposes, a quasi-classical approximation is generally employed for the description of crystal electrons.

2.1.1 Bloch Functions

The behavior of microscopic particles such as electrons or ions in solids is governed by the laws of quantum theory, which states that the position and momentum of a particle are complementary quantities, which cannot be measured simultaneously with arbitrary accuracy: according to the quantum theory, the trajectory of an electron, which specifies the position $\boldsymbol{x}$ as a function of time t and therefore also the momentum $\boldsymbol{p}(t) = m_{\mathrm{e}}\,\mathrm{d}\boldsymbol{x}/\mathrm{d}t$, cannot be defined exactly. Electrons are described instead in terms of a complex probability amplitude (or wave function) $\psi(\boldsymbol{x}, t)$, which is calculated by solving Schrödinger's equation.

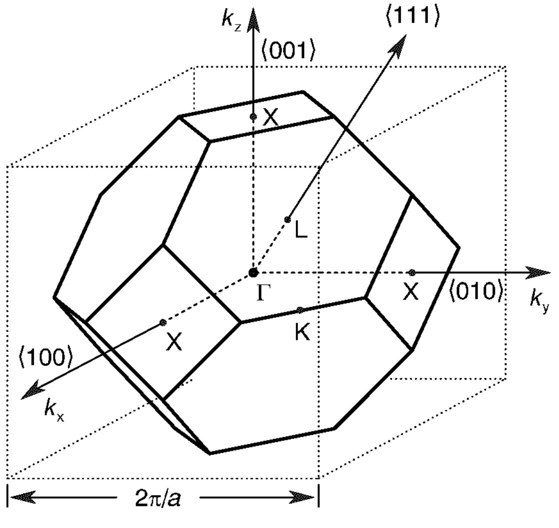


Fig. 2.1. Brillouin zone for a crystal lattice with cubic symmetry, such as the diamond lattice of Si and Ge or the zincblende lattice of GaAs. The zone center is denoted by Γ, the edge along the k_x axis ($\langle 100\rangle$ direction) by X (coordinates $(\pi/a, 0, 0)$) and the edge along the $\langle 111\rangle$ direction by L (coordinates $(\pi/2a, \pi/2a, \pi/2a)$)

The wave functions of electrons in a periodic potential $V(\boldsymbol{x}) = V(\boldsymbol{x}+\boldsymbol{X_m})$ can be written as Bloch functions[2]

$$\psi_{\alpha,\boldsymbol{k}}(\boldsymbol{x}) \;=\; u_{\alpha,\boldsymbol{k}}(\boldsymbol{x})\,\mathrm{e}^{\mathrm{j}\boldsymbol{k}\cdot\boldsymbol{x}} \;; \tag{2.7}$$

the vector $\boldsymbol{k}$ is called the wave vector, and $u_{\alpha,\boldsymbol{k}}(\boldsymbol{x})$ is a function with the lattice symmetry, i.e.

[2]A more detailed account of the fundamentals of solid-state physics can be found, for example, in [2–6].

$$u_{\alpha,\boldsymbol{k}}(\boldsymbol{x}) = u_{\alpha,\boldsymbol{k}}(\boldsymbol{x}+\boldsymbol{X}_{\boldsymbol{m}}) ,$$

where $\boldsymbol{X}_{\boldsymbol{m}}$ denotes a primitive translation of the lattice [6]. Wave functions in Bloch form possess the symmetry of the reciprocal lattice, i.e.

$$\psi_{\alpha,\boldsymbol{k}}(\boldsymbol{x}) = \psi_{\alpha,\boldsymbol{k}+\boldsymbol{K}_{\boldsymbol{m}}}(\boldsymbol{x}) \tag{2.8}$$

for arbitrary reciprocal-lattice vectors $\boldsymbol{K}_{\boldsymbol{m}}$. The wave functions of a given energy band are therefore completely specified if they are known for the first Brillouin zone, which is the unit cell of the reciprocal lattice with center $\boldsymbol{k} = 0$. The energy values $W_\alpha(\boldsymbol{k})$ for different values of α are referred to as the band structure, α is commonly referred to as the band index. In most cases it is sufficient to consider only two bands, the valence band (index V) and the conduction band (index C). The first Brillouin zone is depicted in Fig. 2.1 for the cubic lattice, in which the semiconductors of interest crystallize. Graphical representations of the band structure are usually presented as $W_\alpha(\boldsymbol{k})$ for the paths from Γ to X and L, i.e. for wave vectors directed in the $\langle 100\rangle$ and $\langle 111\rangle$ directions, respectively. Generally the maximum of $W_\mathrm{V}(\boldsymbol{k})$ is found at the Γ point, while the conduction band shows several minima (at the X points in Si and at the L points in Ge and GaAs and also at the Γ point in Ge and GaAs). In the vicinity of these points, $W_\mathrm{C}(\boldsymbol{k})$ may be represented by a parabolic approximation (Fig. 2.2).

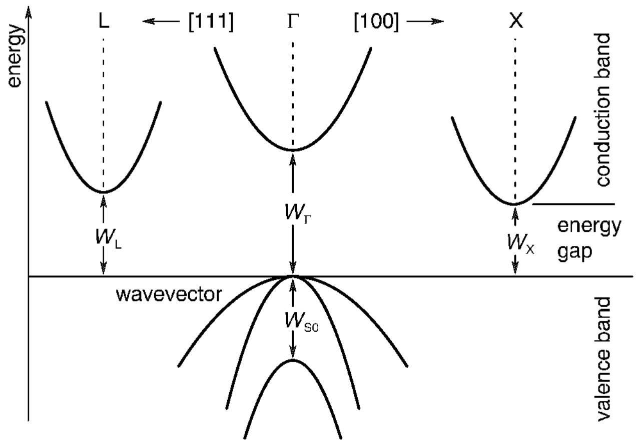


Fig. 2.2. Schematic representation of the band structure of a cubic model semiconductor [7]

The value of the energy gap W_g is determined by the minimum value of W_X, W_L and W_Γ. If $W_\mathrm{g} = W_\Gamma$, as in GaAs, the semiconductor is said to have a direct energy gap, while in the case of Si, with conduction band minima at the X points, and Ge, with conduction band minima at the L points, the energy gap is said to be indirect.

Bloch states are extended states, i.e. a wave function $\psi_{\alpha,\boldsymbol{k}}(\boldsymbol{x})$ that represents such a state extends over the whole crystal volume. In order to describe an electron that has an approximately known position, it is necessary to use a superposition of Bloch functions with different wave vectors $\boldsymbol{k}$ to form a wave packet – in much the same way as a pulse of finite duration is described in terms of its Fourier transform as a superposition of sinusoidal waves of different frequencies. If the wave packet is centered at $\boldsymbol{x}$, with some unavoidable uncertainty $\Delta\boldsymbol{x}$, a superposition of different wavenumbers, assumed to be centered at $\boldsymbol{k}$ with uncertainty $\Delta\boldsymbol{k}$, is required. If the wave packet represents a conduction band electron which is moving in an external electric field $\boldsymbol{E}$, its "position" $\boldsymbol{x}$ and "wavenumber" $\boldsymbol{k}$ obey the semiclassical relations [3]

$$\mathrm{d}\boldsymbol{k}/\mathrm{d}t \;=\; -e\boldsymbol{E}/\hbar\,, \tag{2.9}$$

$$\mathrm{d}\boldsymbol{x}/\mathrm{d}t \;=\; \boldsymbol{u}_\mathrm{n}(\boldsymbol{k}) \;=\; \hbar^{-1}\,\nabla_{\boldsymbol{k}}W_\mathrm{C}(\boldsymbol{k})\,. \tag{2.10}$$

These relations form the basis of the semiclassical theory of transport in semiconductors. Within the effective-mass approximation, the energy dispersion relation for electrons reads

$$W_\mathrm{C}(\boldsymbol{k}) \;\approx\; W_\mathrm{C}(\boldsymbol{k}_0) + \frac{\hbar^2}{2m_\mathrm{n}^*}\,|\boldsymbol{k}|^2 \;=\; W_\mathrm{C} + w_\mathrm{c}(\boldsymbol{k})\,, \tag{2.11}$$

if an isotropic effective mass m_n^* and a conduction band minimum at $\boldsymbol{k}_0 = \boldsymbol{0}$ are assumed; the group velocity of a wave packet centered at $\boldsymbol{k}$ is then

$$\boldsymbol{u}_\mathrm{n}(\boldsymbol{k}) \;=\; \hbar\boldsymbol{k}/m_\mathrm{n}^*\,. \tag{2.12}$$

2.1.2 Temperature Dependence of Bandgap and Intrinsic Carrier Density

The value of the energy gap or bandgap is affected by the lattice constant and the electron–phonon interaction, which vary with temperature. This results in a temperature-dependent bandgap $W_\mathrm{g}(T)$, which is generally represented in the form suggested by Thurmond[3], [10]

$$W_\mathrm{g}(T) \;=\; W_\mathrm{g}(0) - \frac{\alpha T^2}{T+\beta}\,, \tag{2.13}$$

[3] An alternative description of the temperature-dependent bandgap of silicon was given by Bludau and Onton [8], who proposed a quadratic fitting formula

$$W_\mathrm{g}(T) \;=\; 1.1700\ \mathrm{eV} + 1.059\times10^{-5}\,T\times\mathrm{eV/K} - 6.05\times10^{-7}\,T^2\times\mathrm{eV/K^2}$$

for the temperature interval $0 \le T \le 190$ K and

$$W_\mathrm{g}(T) \;=\; 1.1785\ \mathrm{eV} - 9.025\times10^{-5}\,T\times\mathrm{eV/K} - 3.05\times10^{-7}\,T^2\times\mathrm{eV/K^2}$$

for the temperature interval $150\ \mathrm{K} \le T \le 300$ K. A comparison of different formulas used to fit $W_\mathrm{g}(T)$ data can be found in [9].

with the parameters $W_g(0) = 1.170\,\mathrm{eV}$, $\alpha = 4.73 \times 10^{-4}\,\mathrm{eV/K}$ and $\beta = 636\,\mathrm{K}$ for silicon. Figure 2.3 shows both the energy gap $W_g(T)$ and the bandgap voltage

$$V_g(T) = \frac{1}{e}\left(W_g - T\,\frac{\mathrm{d}W_g}{\mathrm{d}T}\right) = \frac{W_g(0)}{e} + \frac{1}{e}\,\frac{\alpha\beta T^2}{(T+\beta)^2} \tag{2.14}$$

for Ge, Si and GaAs as a function of temperature.

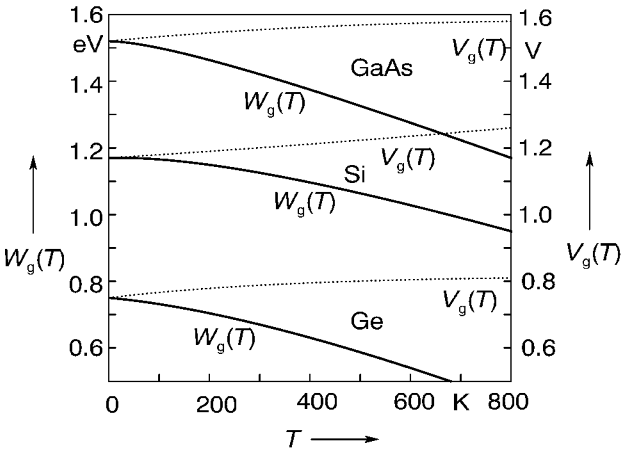


Fig. 2.3. Energy gap W_g and bandgap voltage V_g for the semiconductors Ge, Si and GaAs as a function of temperature

Changes of the density-of-states effective mass with temperature [11,12] cause a temperature dependence of the effective densities of states; for silicon these are described by [13]

$$N_C(T) = 2.86 \times 10^{19} \left(\frac{T}{300\,\mathrm{K}}\right)^{1.58} \mathrm{cm}^{-3} \tag{2.15}$$

and

$$N_V(T) = 3.10 \times 10^{19} \left(\frac{T}{300\,\mathrm{K}}\right)^{1.85} \mathrm{cm}^{-3} \tag{2.16}$$

with an accuracy of 2% over the 200–500 K temperature range. This corresponds to an intrinsic carrier density

$$n_i(T) = 2.98 \times 10^{19} \left(\frac{T}{300\,\mathrm{K}}\right)^{1.71} \exp\left[-\frac{W_g(T)}{2k_B T}\right] \mathrm{cm}^{-3} \tag{2.17}$$

according to [13].

2.2 Thermal Equilibrium

The carrier densities in thermal equilibrium are determined by the density of states and their probability of occupation. If the carrier density is small in comparison with the effective density of states in the band under consideration, Boltzmann statistics apply.

2.2.1 Fermi–Dirac and Boltzmann Statistics

The density $n'(W)\,\mathrm{d}W = Z_C(W)\,f(W)\,\mathrm{d}W$ of electrons with energies in the interval between W and $W + \mathrm{d}W$ is determined by the density of states $Z_C(W)$ in the conduction band together with the probability $f(W)$ that the states considered are indeed occupied.

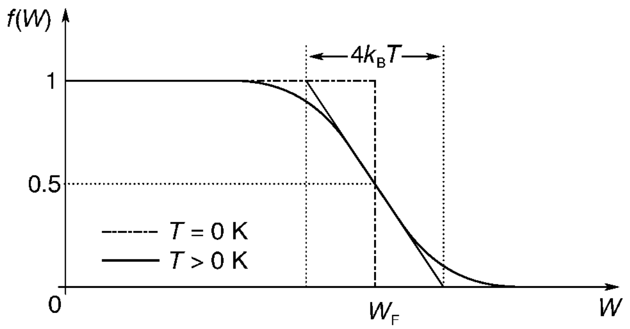


Fig. 2.4. Energy dependence of Fermi distribution $f(W)$

The probability of occupation is given by the Fermi distribution,

$$f(W) = \left[1 + \exp\left(\frac{W - W_F}{k_B T}\right)\right]^{-1}, \tag{2.18}$$

where W_F denotes the Fermi energy. For $T \to 0$ K, the Fermi distribution simplifies to a step function, where $f(W) = 1$ if $W < W_F$ and $f(W) = 0$ if $W > W_F$. This describes the ground state of the solid, in which all electron states below the Fermi energy are occupied, while all states above are unoccupied. For $T > 0$ K, some electrons with energy values $W < W_F$ are excited into states above the Fermi energy – the distribution function is then reduced to values smaller than one at energy values below the Fermi energy, while $f(W) > 0$ above the Fermi energy. At $W = W_F$, the Fermi distribution has a value 1/2 and a slope $-1/(4k_BT)$ (Fig. 2.4). While in an undoped semiconductor the Fermi energy will be approximately in the middle of the energy gap, the introduction of additional electrons by donor atoms will shift the Fermi energy towards the conduction band edge in an n-type semiconductor (Fig. 2.5).

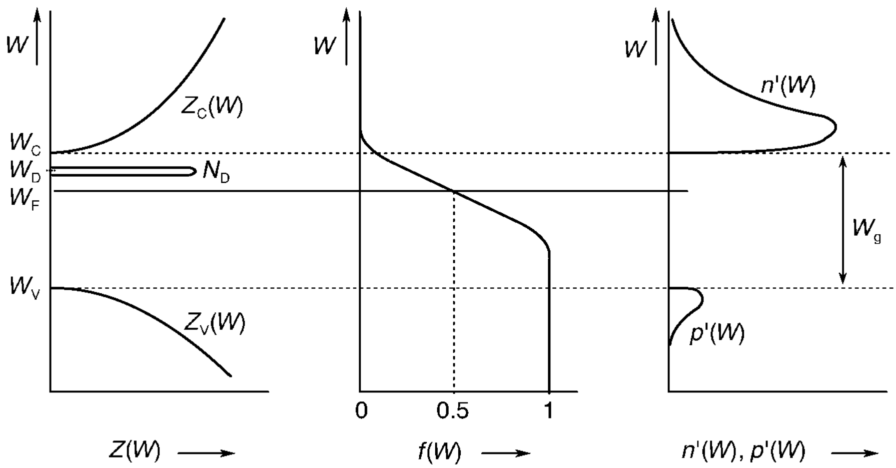


Fig. 2.5. n-type semiconductor. (**a**) Density of states $Z(W)$, (**b**) probability of occupation $f(W)$ and (**c**) energy distribution of electron and hole densities

In contrast, in a p-type semiconductor, the Fermi energy is shifted towards the valence band edge (Fig. 2.6) owing to a decrease of the electron density caused by acceptor atoms.

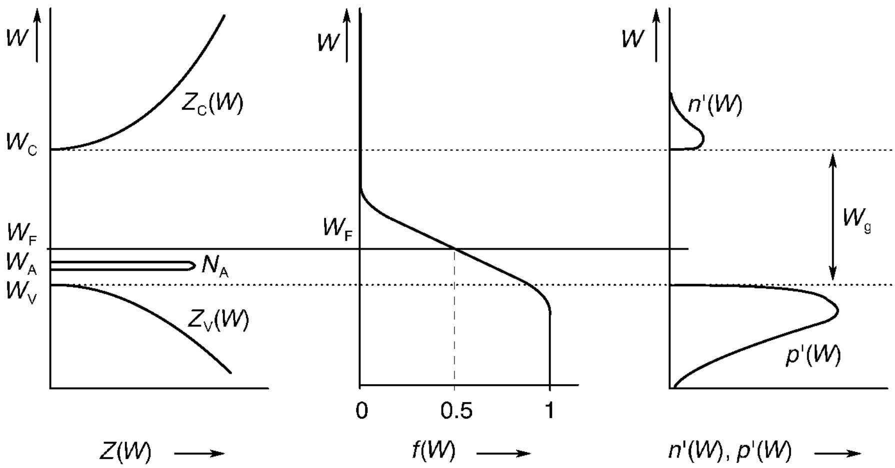


Fig. 2.6. p-type semiconductor. (**a**) Density of states $Z(W)$, (**b**) probability of occupation $f(W)$ and (**c**) energy distribution of electron and hole densities

The density of states $Z_C(W)$ in the conduction band, in the vicinity of the band edge, can be approximated by

$$Z_C(W) \approx \frac{2N_C}{\sqrt{\pi}k_B T}\sqrt{\frac{W-W_C}{k_B T}}\,, \tag{2.19}$$

where N_C denotes the effective density of states in the conduction band. The electron density n_0 in thermal equilibrium comprises all electrons in the conduction band without respect to their energies. Its value is obtained from $n'(W)$ by integration:[4]

$$n_0 = \int_{W_C}^{\infty} Z_C(W) f(W)\,\mathrm{d}W\,. \tag{2.20}$$

In complete analogy, $p'(W)\,\mathrm{d}W = Z_V(W)\,[1-f(W)]\,\mathrm{d}W$ denotes the density of unoccupied states in the valence band, i.e. the density of holes, with energies in the interval between W and $W+\mathrm{d}W$. From this, the hole density p_0 in thermal equilibrium is obtained by integration:

$$p_0 = \int_{-\infty}^{W_V} Z_V(W)\,[1-f(W)]\,\mathrm{d}W\,. \tag{2.21}$$

In the vicinity of the band edge, the density of states $Z_V(W)$ in the valence band may be expressed in terms of the effective density of states in the valence band N_V:

$$Z_V(W) \approx \frac{2N_V}{\sqrt{\pi}k_B T}\sqrt{\frac{W_V-W}{k_B T}}\,. \tag{2.22}$$

In a doped semiconductor with a doping concentration that is small in comparison with the effective density of states, the Fermi energy W_F lies between the band edges W_V and W_C. If the Fermi level lies well below the conduction band edge ($W_C - W_F \gg k_B T$), we may approximate the distribution function in the conduction band by

$$f(W) \approx \exp\left(\frac{W_F - W}{k_B T}\right)\,,$$

whereas, in the valence band, under the assumption $W_F - W_V \gg k_B T$, the approximation

$$1 - f(W) \approx \exp\left(-\frac{W_F - W}{k_B T}\right)$$

holds. Introducing these approximations into (2.20) and (2.21) respectively, yields (1.18) and (1.19).

Degeneracy. In heavily doped semiconductors, the Fermi energy is no longer in the energy gap and Boltzmann statistics may no longer be applied in the computation of majority carrier concentrations; heavily doped semiconductors are therefore sometimes called degenerate. If the density-of-states func-

[4] Owing to the exponential decrease of the Fermi distribution for $W \gg W_F$, the limit of integration has been shifted to infinity.

tion obeys (2.19), the general expression for the equilibrium electron density is given by[5]

$$n_0 = N_C F_{1/2}\left(\frac{W_F - W_C}{k_B T}\right) = \gamma_n N_C e^{x_n} , \tag{2.23}$$

where $x_n = (W_F - W_C)/k_B T$ and

$$F_{1/2}(x) = \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{\xi^{1/2}}{1 + \exp(\xi - x)} \, d\xi \tag{2.24}$$

denotes the so-called Fermi–Dirac integral of order one-half; the degeneracy factor [14] γ_n is defined as $e^{-x_n} F_{1/2}(x_n)$. For larger values of the electron density, when the Fermi energy W_F approaches but still lies below the conduction band energy W_C, γ_n has the series representation

$$\gamma_n \approx 1 - \frac{e^{x_n}}{2^{3/2}} + \frac{e^{2x_n}}{3^{3/2}} - \cdots , \tag{2.25}$$

which converges rather slowly as W_F approaches W_C, corresponding to the situation in which the electron density n_0 approaches the effective density of states in the conduction band N_C.

2.2.2 Ionization

Generally, not all donors or acceptors incorporated into the lattice are ionized. The density of ionized donors is determined from [15]

$$N_D^+ = \frac{N_D}{1 + g \exp\left[(W_F - W_D)/k_B T\right]} ; \tag{2.26}$$

here g denotes the degeneracy of the donor impurity level, which takes account of the fact that each donor state can be unoccupied, or occupied with an electron of either spin (spin degeneracy). If the Fermi energy lies well below the energy of the donor state ($W_D > W_F$) as is the case in moderately doped semiconductors, complete ionization may be assumed at room temperature ($N_D \approx N_D^+$). The density N_A^- of ionized acceptors is determined from the acceptor density N_A according to [15]

$$N_A^- = \frac{N_A}{1 + g \exp\left[(W_A - W_F)/k_B T\right]} , \tag{2.27}$$

where the acceptor state has an energy W_A. The degeneracy g is 4 in Si, Ge and GaAs since each acceptor state can accept an electron of either spin direction from either of the two valence bands, which are degenerate at $\boldsymbol{k} = 0$.

[5] A similar result is obtained for the equilibrium hole density p_0, where the hole degeneracy factor γ_p is introduced.

2.3 The Boltzmann Equation

In thermal equilibrium, the Fermi distribution determines the probability of occupation of a state with wave vector $\boldsymbol{k}$. External disturbances such as an applied electric field or an inhomogeneous temperature distribution, however, will cause deviations from the equilibrium distribution function. If the disturbances vary slowly on the atomic scale, it is possible to introduce a semi-classical approximation,[6], which employs an electron distribution function $f_{\mathrm{n}}(\boldsymbol{x},\boldsymbol{k},t)$, defined such that

$$\frac{1}{4\pi^3}\, f_{\mathrm{n}}(\boldsymbol{x},\boldsymbol{k},t)\,\mathrm{d}^3x\,\mathrm{d}^3k$$

is the number of electrons found at time t in a volume element d^3x at $\boldsymbol{x}$ with wavenumbers within d^3k at $\boldsymbol{k}$. Integration of the distribution function with respect to $\boldsymbol{k}$ yields the electron density

$$n(\boldsymbol{x},t) \;=\; \frac{1}{4\pi^3}\int f_{\mathrm{n}}(\boldsymbol{x},\boldsymbol{k},t)\,\mathrm{d}^3k\;, \tag{2.28}$$

while multiplication of the distribution function by $-e$ times the electron velocity $\boldsymbol{u}_{\mathrm{n}}(\boldsymbol{k}) = \hbar^{-1}\nabla_{\boldsymbol{k}}W_{\mathrm{C}}(\boldsymbol{k})$ and subsequent integration yields the electron current density $\boldsymbol{J}_{\mathrm{n}}(\boldsymbol{x},t)$, defined as

$$\boldsymbol{J}_{\mathrm{n}}(\boldsymbol{x},t) \;=\; -\frac{e}{4\pi^3}\int \boldsymbol{u}_{\mathrm{n}}(\boldsymbol{k})f_{\mathrm{n}}(\boldsymbol{x},\boldsymbol{k},t)\,\mathrm{d}^3k\;. \tag{2.29}$$

If the electrons experience no collisions, the quasi-classical distribution function $f_{\mathrm{n}}(\boldsymbol{x},\boldsymbol{k},t)$ for electrons has to obey

$$\frac{\mathrm{d}}{\mathrm{d}t}f_{\mathrm{n}}(\boldsymbol{x},\boldsymbol{k},t) \;=\; \left(\frac{\partial}{\partial t}+\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}\cdot\nabla_x+\frac{\mathrm{d}\boldsymbol{k}}{\mathrm{d}t}\cdot\nabla_{\mathrm{k}}\right) f_{\mathrm{n}}(\boldsymbol{x},\boldsymbol{k},t) \;=\; 0$$

and

$$\left[\frac{\partial}{\partial t}-\frac{e}{\hbar}\boldsymbol{E}\cdot\nabla_{\boldsymbol{k}}+\boldsymbol{u}_{\mathrm{n}}(\boldsymbol{k})\cdot\nabla\right] f_{\mathrm{n}}(\boldsymbol{x},\boldsymbol{k},t) \;=\; 0\;, \tag{2.30}$$

as a consequence of the relations $\mathrm{d}\boldsymbol{k}/\mathrm{d}t = -e\boldsymbol{E}/\hbar$ and $\mathrm{d}\boldsymbol{x}/\mathrm{d}t = \boldsymbol{u}_{\mathrm{n}}(\boldsymbol{k})$, respectively. Equation (2.30) has to be modified in the presence of collisions, which require the consideration of an additional collision term $C_{\mathrm{n}}[f]$. The collision term is generally expressed in terms of the transition probability $P(\boldsymbol{k},\boldsymbol{k}')$ between the states $\boldsymbol{k}$ and $\boldsymbol{k}'$, which is normalized such that

$$\frac{1}{4\pi^3}\,P(\boldsymbol{k},\boldsymbol{k}')\,\mathrm{d}^3k\,\mathrm{d}^3k'\,\mathrm{d}t$$

is the probability for an electron with a wave vector in the element d^3k at $\boldsymbol{k}$ to be scattered within a time interval $\mathrm{d}t$, into a state with a wave vector lying

[6] For a derivation of the Boltzmann equation from quantum mechanical transport theory see [5, 16, 17], for example.

in the element d^3k' at $\boldsymbol{k'}$ if that state is unoccupied. If this state is occupied with a probability $f_\mathrm{n}(\boldsymbol{x},\boldsymbol{k'},t)$, an additional factor $1 - f_\mathrm{n}(\boldsymbol{x},\boldsymbol{k'},t)$ has to be taken into account since electrons cannot be scattered to occupied states. The changes of $f_\mathrm{n}(\boldsymbol{x},\boldsymbol{k},t)$ due to scattering out of d^3k at $\boldsymbol{k}$ are then given by

$$\left.\frac{\mathrm{d}f_\mathrm{n}}{\mathrm{d}t}\right|_- = -\frac{1}{4\pi^3}\int P(\boldsymbol{k},\boldsymbol{k'})f_\mathrm{n}(\boldsymbol{x},\boldsymbol{k},t)\left[1-f_\mathrm{n}(\boldsymbol{x},\boldsymbol{k'},t)\right]\,\mathrm{d}^3k' ,$$

while those electrons scattered into d^3k at $\boldsymbol{k}$ cause a change

$$\left.\frac{\mathrm{d}f_\mathrm{n}}{\mathrm{d}t}\right|_+ = \frac{1}{4\pi^3}\int P(\boldsymbol{k'},\boldsymbol{k})f(\boldsymbol{x},\boldsymbol{k'},t)\left[1-f_\mathrm{n}(\boldsymbol{x},\boldsymbol{k},t)\right]\,\mathrm{d}^3k' .$$

By summing these equations, the Boltzmann equation

$$\left(\frac{\partial}{\partial t} - \frac{e}{\hbar}\boldsymbol{E}\cdot\nabla_{\boldsymbol{k}} + \boldsymbol{u}_\mathrm{n}(\boldsymbol{k})\cdot\nabla_{\boldsymbol{x}}\right) f_\mathrm{n}(\boldsymbol{x},\boldsymbol{k},t) = C_\mathrm{n}[f] \tag{2.31}$$

is obtained, where

$$C_\mathrm{n}[f] = \left.\frac{\mathrm{d}f_\mathrm{n}}{\mathrm{d}t}\right|_+ + \left.\frac{\mathrm{d}f_\mathrm{n}}{\mathrm{d}t}\right|_- \tag{2.32}$$

is the collision term. A solution of the Boltzmann equation generally requires numerical methods [18].

2.3.1 Collision Term

Lattice vibrations, crystal defects, and collisions with impurities or other carriers cause deviations $\delta V(\boldsymbol{x})$ of the crystal potential from its unperturbed value, which result in frequent transitions between states. The transition rate $P(\boldsymbol{k},\boldsymbol{k'})$ between states with wave vectors $\boldsymbol{k}$ and $\boldsymbol{k'}$ is generally expressed with the help of Fermi's golden rule, which gives

$$P(\boldsymbol{k},\boldsymbol{k'}) = \frac{2\pi}{\hbar}g(\boldsymbol{k'})|M(\boldsymbol{k},\boldsymbol{k'})|^2\delta[\,W_\mathrm{C}(\boldsymbol{k'})-W_\mathrm{C}(\boldsymbol{k})-\mathit{\Delta}W\,] ,$$

where

$$M(\boldsymbol{k},\boldsymbol{k'}) = \frac{1}{\mathcal{V}}\int \delta V(\boldsymbol{x})\,\mathrm{e}^{\mathrm{j}(\boldsymbol{k}-\boldsymbol{k'})\cdot\boldsymbol{x}}\,\mathrm{d}^3x$$

is the matrix element of the perturbation with respect to the unperturbed states $\psi_{\boldsymbol{k}}$ and $\psi_{\boldsymbol{k'}}$, $g(\boldsymbol{k'})$ is the density of final states, and $\mathit{\Delta}W$ is the change in energy caused by an inelastic collision. Elastic collisions are characterized by $\mathit{\Delta}W = 0$, i.e. the energy $W_\mathrm{C}(\boldsymbol{k})$ of the incoming state equals the energy $W_\mathrm{C}(\boldsymbol{k'})$ of the outgoing carrier state.

2.3.2 Thermal Equilibrium

In thermal equilibrium both the time derivative $\partial f_\mathrm{n}/\partial t$ and the collision term $C_\mathrm{n}[f]$ vanish; the equilibrium distribution function $f_\mathrm{n0}(\boldsymbol{x},\boldsymbol{k})$ therefore has to obey

$$\left(-\frac{e}{\hbar}\,\boldsymbol{E}(\boldsymbol{x})\cdot\nabla_{\boldsymbol{k}} + \boldsymbol{u}_\mathrm{n}(\boldsymbol{k})\cdot\nabla_{\boldsymbol{x}}\right) f_\mathrm{n0}(\boldsymbol{x},\boldsymbol{k}) \;=\; 0\,. \tag{2.33}$$

Since $\boldsymbol{E} = -\nabla\psi$ and $\hbar\boldsymbol{u}_\mathrm{n}(\boldsymbol{k}) = \nabla_{\boldsymbol{k}} W_\mathrm{C}(\boldsymbol{k})$, this equation is solved by a function of the form $f_\mathrm{n0}(\boldsymbol{x},\boldsymbol{k}) = g[\,W_\mathrm{C}(\boldsymbol{k}) - e\psi(\boldsymbol{x}) + \mathrm{const.}\,]$, where the constant is chosen as $-W_\mathrm{F}$. In addition to this, it can be shown that the collision term vanishes if $f_\mathrm{n0}(\boldsymbol{x},\boldsymbol{k})$ is of the form

$$f_\mathrm{n0}(\boldsymbol{x},\boldsymbol{k}) \;=\; \frac{1}{1+\exp\left[\dfrac{W_\mathrm{C}(\boldsymbol{k})-e\psi(\boldsymbol{x})-W_\mathrm{F}}{k_\mathrm{B}T}\right]}\,, \tag{2.34}$$

i.e. the Fermi distribution is obtained as an equilibrium solution of the Boltzmann equation.

2.3.3 Limits of Validity

A series of approximations is required for the derivation of the Boltzmann equation from more general quantum transport equations:

1. The Boltzmann equation uses simultaneously the position and momentum of a particle, which, according to the uncertainty principle $\Delta x\Delta p \geq \hbar/2$, can be known accurately at the same time only in the limit $\hbar \to 0$. Predominantly, only electron states within an energy interval of width $\Delta W = k_\mathrm{B}T$ in the vicinity of the conduction band edge are occupied. Therefore, from $\Delta W = \hbar^2\Delta k^2/(2m_\mathrm{n}^*)$, the spread Δk of the electron states that contribute to a wave packet can be estimated as $\Delta k = \sqrt{2m_\mathrm{n}^* k_\mathrm{B}T}\,/\hbar$, resulting in an uncertainty of the position of an electron
$$\Delta x \;\geq\; \frac{\hbar}{2\sqrt{2m_\mathrm{n}^* k_\mathrm{B}T}} \;\approx\; 10\text{--}20\ \mathrm{nm}\,.$$
A classical description of electron dynamics is therefore only justified if the potential varies slowly on the scale of a thermal wavelength, $\theta_\mathrm{B} = h/\sqrt{2m_\mathrm{n}^* k_\mathrm{B}T}$.
2. The finite spread of electron energies, together with the uncertainty relation $\Delta W\Delta t \geq \hbar/2$, limits the frequency range in which the Boltzmann equation may be applied to a maximum frequency
$$f_\mathrm{m} \;\approx\; 1/\Delta t \;\leq\; 2\Delta W/\hbar \;\approx\; 2k_\mathrm{B}T/\hbar\,.$$

The Boltzmann equation can therefore only be applied if the frequency is well below the terahertz range.

3. The interactions of particles are assumed to be small enough to be treated as perturbations resulting in a collision term which is proportional to the corresponding particle densities. Furthermore, collisions are considered to be instantaneous and independent of external fields. If this assumption is dropped, the so-called intracollisional field effect modifies the collision term.
4. Another restriction stems from the fact that the Boltzmann equation provides a single-particle description that provides no information about the correlations between carriers. In order to describe fluctuations and noise the Boltzmann equation therefore has to be extended by stochastic terms.

2.3.4 Relaxation Times

For the formulation of transport theories, it is helpful to derive so-called "relaxation times" from the transition rates $P(\boldsymbol{k}, \boldsymbol{k}')$ in the collision term. The scattering time $\tau_{\mathrm{cn}}(\boldsymbol{k})$, defined by

$$\frac{1}{\tau_{\mathrm{cn}}(\boldsymbol{k})} = \frac{1}{4\pi^3} \int P(\boldsymbol{k}, \boldsymbol{k}')\, \mathrm{d}^3 k' \,, \tag{2.35}$$

determines the rate at which carriers with an incoming state $\boldsymbol{k}$ are scattered to any other state. The probability that a particle in state $\boldsymbol{k}$ at $t = 0$ remains unscattered until time t is given by $\exp[\,-t/\tau_{\mathrm{cn}}(\boldsymbol{k})\,]$. The mean time between two scattering events is therefore

$$\tau_{\mathrm{cn}}(\boldsymbol{k}) = \int_0^{\infty} t \exp\left(-\frac{t}{\tau_{\mathrm{cn}}(\boldsymbol{k})}\right) \mathrm{d}t \,. \tag{2.36}$$

If the scattering mechanism is isotropic, an incoming electron will, on average, have lost all its memory about its original velocity or momentum after one scattering time. In the case of an anisotropic scattering mechanism, the momentum relaxation time, given by

$$\frac{1}{\tau_{\mathrm{vn}}(\boldsymbol{k})} = \frac{1}{4\pi^3} \int P(\boldsymbol{k}, \boldsymbol{k}') \left(1 - \frac{\boldsymbol{k}' \cdot \boldsymbol{k}}{|\boldsymbol{k}|^2}\right) \mathrm{d}^3 k' \,, \tag{2.37}$$

determines the time needed on average to randomize the momentum of an electron. The energy relaxation time, given by

$$\frac{1}{\tau_{\mathrm{wn}}(\boldsymbol{k})} = \frac{1}{4\pi^3} \int P(\boldsymbol{k}, \boldsymbol{k}') \left(1 - \frac{W_{\mathrm{C}}(\boldsymbol{k}')}{W_{\mathrm{C}}(\boldsymbol{k})}\right) \mathrm{d}^3 k' \,, \tag{2.38}$$

determines how quickly carriers may dissipate or absorb energy from their surrounding. In the case of elastic scattering events at fixed impurities, the momentum relaxation time will be infinite.

2.4 The Drift–Diffusion Approximation

The Boltzmann equation is a nonlinear integrodifferential equation and difficult to solve – analytical solutions require simplifying assumptions. In this section only slight deviations of the distribution function from the equilibrium distribution are considered, a situation that is found in the case of low electric field strength. The additional assumption of isotropic scattering then allows one to solve the Boltzmann equation within the relaxation time approximation, which leads to the drift–diffusion approximation of carrier transport.

2.4.1 The Relaxation Time Approximation

In the case of small carrier densities, most of the states are unoccupied; the approximations $1 - f_{\rm n}(\boldsymbol{x}, \boldsymbol{k}, t) \approx 1$ and $1 - f_{\rm n}(\boldsymbol{x}, \boldsymbol{k}', t) \approx 1$ may be applied in the collision term, resulting in

$$C_{\rm n}[f] = -\frac{f_{\rm n}(\boldsymbol{x}, \boldsymbol{k}, t)}{\tau_{\rm cn}(\boldsymbol{k})} + \frac{1}{4\pi^3} \int P(\boldsymbol{k}, \boldsymbol{k}') f_{\rm n}(\boldsymbol{x}, \boldsymbol{k}', t)\, {\rm d}^3 k' \,, \tag{2.39}$$

where the collision time $\tau_{\rm cn}(\boldsymbol{k})$ is defined in (2.36). It is always possible to split up the distribution function $f_{\rm n}(\boldsymbol{x}, \boldsymbol{k}, t)$ into a symmetric contribution $f_{\rm S}(\boldsymbol{x}, \boldsymbol{k}, t)$ and an antisymmetric contribution $f_{\rm A}(\boldsymbol{x}, \boldsymbol{k}, t)$ [19], i.e.

$$f_{\rm n}(\boldsymbol{x}, \boldsymbol{k}, t) = f_{\rm S}(\boldsymbol{x}, \boldsymbol{k}, t) + f_{\rm A}(\boldsymbol{x}, \boldsymbol{k}, t) \,,$$

where

$$f_{\rm S}(\boldsymbol{x}, \boldsymbol{k}, t) = \frac{1}{2} \left[f_{\rm n}(\boldsymbol{x}, \boldsymbol{k}, t) + f_{\rm n}(\boldsymbol{x}, -\boldsymbol{k}, t) \right]$$

and

$$f_{\rm A}(\boldsymbol{x}, \boldsymbol{k}, t) = \frac{1}{2} \left[f_{\rm n}(\boldsymbol{x}, \boldsymbol{k}, t) - f_{\rm n}(\boldsymbol{x}, -\boldsymbol{k}, t) \right] .$$

If $P(\boldsymbol{k}, \boldsymbol{k}')$ remains constant with respect to the sign reversal of one component of $\boldsymbol{k}$, the integral

$$\int P(\boldsymbol{k}, \boldsymbol{k}')\, f_{\rm A}(\boldsymbol{x}, \boldsymbol{k}', t)\, {\rm d}^3 k'$$

vanishes, and the collision integral simplifies to

$$\begin{aligned} C_{\rm n}[f] &= -\frac{f_{\rm A}}{\tau_{\rm cn}(\boldsymbol{k})} - \frac{f_{\rm S}}{\tau_{\rm cn}(\boldsymbol{k})} + \frac{1}{4\pi^3} \int P(\boldsymbol{k}, \boldsymbol{k}') f_{\rm S}(\boldsymbol{x}, \boldsymbol{k}', t)\, {\rm d}^3 k' \\ &= -\frac{f_{\rm A}}{\tau_{\rm cn}(\boldsymbol{k})} + C_{\rm n}[f_{\rm S}] \,. \end{aligned} \tag{2.40}$$

Since $f_{\rm S}$ is an even function of $\boldsymbol{k}$, its gradient $\nabla_{\boldsymbol{k}} f_{\rm S}$ is odd; correspondingly, the gradient $\nabla_{\boldsymbol{k}} f_{\rm A}$ of the odd function $f_{\rm A}$ is even. Since $\boldsymbol{u}_{\rm n}(\boldsymbol{k})$ is odd, and the odd and even contributions to the Boltzmann equation must cancel separately, the Boltzmann equation can be split up into the following system of differential equations:

$$\frac{\partial f_\mathrm{A}}{\partial t} - \frac{e}{\hbar}\boldsymbol{E}\cdot\nabla_{\boldsymbol{k}} f_\mathrm{S} + \boldsymbol{u}_\mathrm{n}(\boldsymbol{k})\cdot\nabla f_\mathrm{S} = -\frac{f_\mathrm{A}}{\tau_\mathrm{cn}(\boldsymbol{k})}\,, \tag{2.41}$$

$$\frac{\partial f_\mathrm{S}}{\partial t} - \frac{e}{\hbar}\boldsymbol{E}\cdot\nabla_{\boldsymbol{k}} f_\mathrm{A} + \boldsymbol{u}_\mathrm{n}(\boldsymbol{k})\cdot\nabla f_\mathrm{A} = C_\mathrm{n}[f_\mathrm{S}]\,. \tag{2.42}$$

This may be used for an approximate computation of distribution functions which are only slightly perturbed from their equilibrium form.

2.4.2 Transport in Low Electric Fields

In the case of low electric fields and small temperature gradients, we may assume only slight deviations of the electron distribution function from its equilibrium form, which is symmetric with respect to $\boldsymbol{k}$. The antisymmetric contribution $f_\mathrm{A}(\boldsymbol{x},\boldsymbol{k},t)$ will then be small in comparison with the symmetric part $f_\mathrm{S}(\boldsymbol{x},\boldsymbol{k},t)$. The latter can be written, in analogy to the equilibrium distribution, as

$$f_\mathrm{S}(\boldsymbol{x},\boldsymbol{k},t) = \frac{1}{1+\mathrm{e}^{\Theta_\mathrm{n}(\boldsymbol{x},\boldsymbol{k},t)}}\,, \tag{2.43}$$

where

$$\Theta_\mathrm{n}(\boldsymbol{x},\boldsymbol{k},t) = \frac{w_\mathrm{c}(\boldsymbol{k})+W_\mathrm{C}(\boldsymbol{x})+e\phi_\mathrm{n}(\boldsymbol{x})}{k_\mathrm{B}T_\mathrm{n}(\boldsymbol{x})} \tag{2.44}$$

allows the possibility of a positional dependence of the conduction band edge $W_\mathrm{C}(\boldsymbol{x})$, caused by an electric field or compositional variations in the semiconductor crystal. The term $w_\mathrm{c}(\boldsymbol{k})$ denotes the kinetic energy of the state (measured from the conduction band edge), while $-e\phi_\mathrm{n}(\boldsymbol{x}) = W_\mathrm{Fn}(\boldsymbol{x})$ can be interpreted as an energy-dependent "Fermi energy" for electrons, defined so as to yield the correct particle density, i.e.

$$n(\boldsymbol{x},t) = \frac{1}{4\pi^3}\int f_\mathrm{S}(\boldsymbol{x},\boldsymbol{k},t)\,\mathrm{d}^3k\,, \tag{2.45}$$

since $\int f_\mathrm{A}(\boldsymbol{x},\boldsymbol{k},t)\,\mathrm{d}^3k = 0$; the quantity ϕ_n is commonly referred to as the quasi-Fermi potential for electrons. The electron temperature $T_\mathrm{n}(\boldsymbol{x})$ is assumed to vary slowly with position and taken to be equal to the lattice temperature. This assumption requires the electrons to be in equilibrium with the lattice and is justified in the presence of a strong electron–phonon interaction and limited electron heating in low electric fields. Since f_S has the equilibrium form, the collision term $C_\mathrm{n}[f_\mathrm{S}]$ vanishes. Under stationary conditions, (2.41) then determines f_A in terms of f_S:

$$f_\mathrm{A} = \tau_\mathrm{cn}(\boldsymbol{k})\,\frac{e}{\hbar}\,\boldsymbol{E}\cdot\nabla_{\boldsymbol{k}} f_\mathrm{S} - \tau_\mathrm{cn}(\boldsymbol{k})\,\boldsymbol{u}_\mathrm{n}(\boldsymbol{k})\cdot\nabla f_\mathrm{S}\ . \tag{2.46}$$

A simple differentiation of (2.43) and (2.44) gives

$$\nabla f_\mathrm{S} = \frac{\mathrm{d}f_\mathrm{S}}{\mathrm{d}\Theta_\mathrm{n}}\,\nabla\Theta_\mathrm{n} \quad \text{and} \quad \nabla_{\boldsymbol{k}} f_\mathrm{S} = \frac{\mathrm{d}f_\mathrm{S}}{\mathrm{d}\Theta_\mathrm{n}}\,\nabla_{\boldsymbol{k}}\Theta_\mathrm{n}\ ,$$

together with

$$\nabla_{\boldsymbol{k}}\Theta_\mathrm{n} = \frac{1}{k_\mathrm{B}T_\mathrm{n}(\boldsymbol{x})}\nabla_{\boldsymbol{k}} w_\mathrm{c}(\boldsymbol{k}) = \frac{\hbar}{k_\mathrm{B}T_\mathrm{n}(\boldsymbol{x})}\,\boldsymbol{u}_\mathrm{n}(\boldsymbol{k})$$

and

$$\nabla\Theta_\mathrm{n} = \frac{e}{k_\mathrm{B}T_\mathrm{n}(\boldsymbol{x})}\,\nabla\phi_\mathrm{n} + \frac{1}{k_\mathrm{B}T_\mathrm{n}(\boldsymbol{x})}\,\nabla W_\mathrm{C} - \frac{\Theta_\mathrm{n}(\boldsymbol{x},\boldsymbol{k},t)}{T_\mathrm{n}(\boldsymbol{x})}\,\nabla T_\mathrm{n}\ .$$

We therefore obtain

$$\begin{aligned} f_\mathrm{A} &= \frac{e\tau_\mathrm{cn}(\boldsymbol{k})}{k_\mathrm{B}T_\mathrm{n}(\boldsymbol{x})}\,\frac{\mathrm{d}f_\mathrm{S}}{\mathrm{d}\Theta_\mathrm{n}}\,\boldsymbol{u}_\mathrm{n}(\boldsymbol{k})\cdot\left(\boldsymbol{E} - \frac{1}{e}\nabla W_\mathrm{C} - \nabla\phi_\mathrm{n}\right) \\ &\qquad + \tau_\mathrm{cn}(\boldsymbol{k})\,\frac{\mathrm{d}f_\mathrm{S}}{\mathrm{d}\Theta_\mathrm{n}}\,\frac{\Theta_\mathrm{n}(\boldsymbol{x},\boldsymbol{k},t)}{T_\mathrm{n}(\boldsymbol{x})}\,\boldsymbol{u}_\mathrm{n}(\boldsymbol{k})\cdot\nabla T_\mathrm{n}\ . \end{aligned} \tag{2.47}$$

Since the symmetric portion of the distribution function does not contribute to the current density, the electron current density under isothermal conditions (where T_n is constant) is given by

$$\begin{aligned} \boldsymbol{J}_\mathrm{n} &= -\frac{e}{4\pi^3}\int \boldsymbol{u}_\mathrm{n}(\boldsymbol{k})\,f_\mathrm{A}(\boldsymbol{x},\boldsymbol{k},t)\,\mathrm{d}^3k \\ &= -en\breve{\mu}_\mathrm{n}\cdot\left(\nabla\phi_\mathrm{n} - \boldsymbol{E} + \frac{1}{e}\,\nabla W_\mathrm{C}\right)\ , \end{aligned} \tag{2.48}$$

where

$$\breve{\mu}_\mathrm{n} = -\frac{1}{V_\mathrm{T}}\int\frac{\mathrm{d}f_\mathrm{S}}{\mathrm{d}\Theta_\mathrm{n}}\,\tau_\mathrm{cn}(\boldsymbol{k})\,\left[\boldsymbol{u}_\mathrm{n}(\boldsymbol{k})\otimes\boldsymbol{u}_\mathrm{n}(\boldsymbol{k})\right]\,\mathrm{d}^3k\bigg/\int f_\mathrm{S}\,\mathrm{d}^3k \tag{2.49}$$

denotes the electron mobility tensor. In the effective-mass approximation (2.12), the mobility tensor is proportional to the unit tensor, i.e. $\breve{\mu}_\mathrm{n} = \mu_\mathrm{n}\,1$, and the electron mobility is given by [20, 21]

$$\mu_\mathrm{n} = -\frac{1}{3V_\mathrm{T}}\left(\frac{\hbar^2}{m_\mathrm{n}^*}\right)^2\int\frac{\mathrm{d}f_\mathrm{S}}{\mathrm{d}\Theta_\mathrm{n}}\,\tau_\mathrm{cn}(\boldsymbol{k})\,|\boldsymbol{k}|^2\,\mathrm{d}^3k\bigg/\int f_\mathrm{S}\,\mathrm{d}^3k\ . \tag{2.50}$$

In a homogeneous semiconductor, the only possible reason for a positional dependence of the conduction band edge is an electric field, i.e. $\nabla W_\mathrm{C} = e\boldsymbol{E}$, and (2.48) simplifies to

$$\boldsymbol{J}_\mathrm{n} = -en\mu_\mathrm{n}\nabla\phi_\mathrm{n}\ . \tag{2.51}$$

According to (2.45), the gradient of the electron density may be expressed as

$$\nabla n \;=\; \frac{1}{4\pi^3}\int \frac{\mathrm{d}f_\mathrm{S}}{\mathrm{d}\Theta_\mathrm{n}}\,\nabla\Theta_\mathrm{n}\,\mathrm{d}^3k\;;$$

in the isothermal case [7] ($|\nabla T_\mathrm{n}| = \mathrm{const.}$), with $\nabla W_\mathrm{C} = e\boldsymbol{E}$, this is equivalent to

$$V_\mathrm{T}\,\nabla n \;=\; \frac{1}{4\pi^3}\int \frac{\mathrm{d}f_\mathrm{S}}{\mathrm{d}\Theta_\mathrm{n}}\,\mathrm{d}^3k\,(\nabla\phi_\mathrm{n} + \boldsymbol{E})$$

and

$$\nabla\phi_\mathrm{n} \;=\; -\boldsymbol{E} - \Lambda_\mathrm{n}\,\frac{V_\mathrm{T}}{n}\,\nabla n\;,$$

respectively, where

$$\Lambda_\mathrm{n} \;=\; -\int f_\mathrm{S}\,\mathrm{d}^3k\Bigg/\int \frac{\mathrm{d}f_\mathrm{S}}{\mathrm{d}\Theta_\mathrm{n}}\,\mathrm{d}^3k\;.$$

Introducing the electron diffusion coefficient

$$D_\mathrm{n} \;=\; \Lambda_\mathrm{n} V_\mathrm{T}\,\mu_\mathrm{n} \tag{2.54}$$

then transforms the electron current equation (2.51) into

$$\boldsymbol{J}_\mathrm{n} \;=\; e\mu_\mathrm{n} n\,\boldsymbol{E} + eD_\mathrm{n}\,\nabla n\;. \tag{2.55}$$

If the electron density n in the conduction band is small in comparison with the effective density of states N_C, the probability of occupation is small in comparison with one, and therefore

$$f_\mathrm{S} \;=\; \frac{1}{1+\mathrm{e}^{\Theta_\mathrm{n}}} \;\approx\; \mathrm{e}^{-\Theta_\mathrm{n}} \;\ll\; 1\;.$$

In this case the approximations $\mathrm{d}f_\mathrm{S}/\mathrm{d}\Theta_\mathrm{n} \approx -f_\mathrm{S}$ and $\Lambda_\mathrm{n} \approx 1$ hold, resulting in the classical Einstein relation $D_\mathrm{n} = V_\mathrm{T}\mu_\mathrm{n}$.

The drift–diffusion approximation describes the linear response of the system to disturbances in the electric field and temperature. Carrier heating in the electric field is not considered, since this is not a phenomenon of linear response. The carrier temperature therefore cannot be calculated selfconsistently within the drift diffusion approximation.

[7] In the presence of a temperature gradient, an additional current component

$$\boldsymbol{J}_\mathrm{n,th} \;=\; enD_\mathrm{nth}\nabla T_\mathrm{n} \tag{2.52}$$

has to be considered, where

$$D_\mathrm{nth} \;=\; -\,\frac{1}{12\pi^3 n T_\mathrm{n}(\boldsymbol{x})}\left(\frac{\hbar^2}{m_\mathrm{n}^*}\right)^2\int \frac{\mathrm{d}f_\mathrm{S}}{\mathrm{d}\Theta_\mathrm{n}}\Theta_\mathrm{n}(\boldsymbol{x},\boldsymbol{k},t)\tau_\mathrm{cn}(\boldsymbol{k})|\boldsymbol{k}|^2\,\mathrm{d}^3k \tag{2.53}$$

is the thermal diffusion coefficient for electrons.

2.5 Hydrodynamic Model

The hydrodynamic model consists of a set of balance equations for the particle density, current density and average kinetic energy and is derived from the Boltzmann equation by integration with respect to $\boldsymbol{k}$. In comparison with the drift–diffusion approximation, the hydrodynamic model provides an improved description of nonlocal effects on carrier transport such as velocity overshoot or hot-carrier processes.[8] Balance equations for (ensemble) averages defined by

$$\langle a\rangle \;=\; \frac{1}{4\pi^3 n(\boldsymbol{x},t)}\int a(\boldsymbol{x},\boldsymbol{k})f_\mathrm{n}(\boldsymbol{x},\boldsymbol{k},t)\,\mathrm{d}^3k \tag{2.56}$$

are derived by multiplication of the Boltzmann equation by the function $a(\boldsymbol{x},\boldsymbol{k})$ and subsequent integration with respect to $\boldsymbol{k}$. Since the distribution function $f_\mathrm{n}(\boldsymbol{x},\boldsymbol{k},t)$ vanishes sufficiently fast as $|\boldsymbol{k}|\to\infty$, one obtains, after partial integration, the balance equation

$$\frac{\partial}{\partial t}n\langle a\rangle + \nabla\cdot(n\langle\boldsymbol{u}_\mathrm{n}a\rangle) + \frac{e}{\hbar}n\boldsymbol{E}\cdot\langle\nabla_{\boldsymbol{k}}a\rangle \;=\; \frac{1}{4\pi^3}\int aC[f]\,\mathrm{d}^3k\,. \tag{2.57}$$

The equations of the hydrodynamic model can be derived as approximations to the balance equations obtained for $a(\boldsymbol{x},\boldsymbol{k})=1,\boldsymbol{u}_\mathrm{n}(\boldsymbol{k})$ and $w_\mathrm{c}(\boldsymbol{k})$, respectively.

2.5.1 Continuity Equation

Integration of the Boltzmann equation for electrons with respect to $\boldsymbol{k}$ yields the continuity equation

$$\frac{\partial n}{\partial t} + \nabla\cdot(n\boldsymbol{v}_\mathrm{n}) \;=\; \left(\frac{\partial n}{\partial t}\right)_\mathrm{c}\,, \tag{2.58}$$

where n is the electron density (2.28), and

$$\left(\frac{\partial n}{\partial t}\right)_\mathrm{c} \;=\; \frac{1}{4\pi^3}\int C_\mathrm{n}[f]\,\mathrm{d}^3k \;=\; -(R-G) \tag{2.59}$$

defines the net recombination rate. The continuity equation requires no simplifying assumptions in its derivation from the Boltzmann equation and holds therefore under general conditions.

[8] In the formulation presented below, lattice heating is neglected. A general approach that takes account of this effect requires the simultaneous solution of the Boltzmann equations for electrons, holes and phonons [22]. The generation rate due to impact ionization is calculated within the hydrodynamic model for both direct and indirect semiconductors in [23]; simple phenomenological expressions are obtained if the "energy formulation" of Chynoweth's formula (see Sect. 2.6) is used [24].

2.5.2 Current Equation

Multiplication of the Boltzmann equation by $\boldsymbol{u}_{\mathrm{n}}(\boldsymbol{k})$ and subsequent integration with respect to $\boldsymbol{k}$ yields the momentum balance equation, which can be written as

$$\begin{aligned}\frac{\partial}{\partial t}(n\boldsymbol{v}_{\mathrm{n}}) + \nabla\cdot[\,n(\boldsymbol{v}_{\mathrm{n}}\otimes\boldsymbol{v}_{\mathrm{n}})\,] + \nabla\cdot(n\langle\boldsymbol{w}_{\mathrm{n}}\otimes\boldsymbol{w}_{\mathrm{n}}\rangle) + \frac{e}{\hbar}n\boldsymbol{E}\cdot\langle\nabla_{\boldsymbol{k}}\otimes\boldsymbol{u}_{\mathrm{n}}\rangle \\ = -\boldsymbol{v}_{\mathrm{n}}(R-G) + \frac{1}{4\pi^3}\int \boldsymbol{w}_{\mathrm{n}}C_{\mathrm{n}}[f]\,\mathrm{d}^3k\,, \qquad (2.60)\end{aligned}$$

where the electron velocity $\boldsymbol{u}_{\mathrm{n}}(\boldsymbol{k})$ has been split up into a contribution due to the streaming of the electrons with the hydrodynamic velocity

$$\boldsymbol{v}_{\mathrm{n}}(\boldsymbol{x},t) = \langle\boldsymbol{u}_{\mathrm{n}}(\boldsymbol{k})\rangle = \frac{1}{4\pi^3 n(\boldsymbol{x},t)}\int \boldsymbol{u}_{\mathrm{n}}(\boldsymbol{k})f_{\mathrm{n}}(\boldsymbol{x},\boldsymbol{k},t)\,\mathrm{d}^3k\,, \qquad (2.61)$$

and a contribution $\boldsymbol{w}_{\mathrm{n}}(\boldsymbol{k}) = \boldsymbol{u}_{\mathrm{n}}(\boldsymbol{k}) - \boldsymbol{v}_{\mathrm{n}}$, which describes the kinetic energy associated with the random thermal motion of the electrons. The term $\boldsymbol{w}_{\mathrm{n}}(\boldsymbol{k})$ obviously averages to zero, i.e. $\langle\boldsymbol{w}_{\mathrm{n}}(\boldsymbol{k})\rangle = 0$, while $\boldsymbol{v}_{\mathrm{n}}$ is related to the electron current density $\boldsymbol{J}_{\mathrm{n}}$ defined in (2.29) by $\boldsymbol{J}_{\mathrm{n}} = -en\boldsymbol{v}_{\mathrm{n}}$. Using the continuity equation, we may write

$$\begin{aligned}\frac{\partial}{\partial t}(n\boldsymbol{v}_{\mathrm{n}}) &= n\frac{\partial\boldsymbol{v}_{\mathrm{n}}}{\partial t} + \boldsymbol{v}_{\mathrm{n}}\frac{\partial n}{\partial t} = n\frac{\partial\boldsymbol{v}_{\mathrm{n}}}{\partial t} - (R-G)\,\boldsymbol{v}_{\mathrm{n}} - \boldsymbol{v}_{\mathrm{n}}\,[\,\nabla\cdot(n\mathbf{v}_{\mathrm{n}})\,] \\ &= n\frac{\partial\boldsymbol{v}_{\mathrm{n}}}{\partial t} - (R-G)\,\boldsymbol{v}_{\mathrm{n}} - \boldsymbol{v}_{\mathrm{n}}(\boldsymbol{v}_{\mathrm{n}}\cdot\nabla n) - n\boldsymbol{v}_{\mathrm{n}}(\nabla\cdot\boldsymbol{v}_{\mathrm{n}})\,,\end{aligned}$$

since $\nabla\cdot[\,n(\boldsymbol{v}_{\mathrm{n}}\otimes\boldsymbol{v}_{\mathrm{n}})] = \boldsymbol{v}_{\mathrm{n}}(\boldsymbol{v}_{\mathrm{n}}\cdot\nabla n) + 2n\boldsymbol{v}_{\mathrm{n}}(\nabla\cdot\boldsymbol{v}_{\mathrm{n}})$. With this equation, (2.60) is transformed into

$$\begin{aligned}n\left(\frac{\partial}{\partial t} + \boldsymbol{v}_{\mathrm{n}}\cdot\nabla\right)\boldsymbol{v}_{\mathrm{n}} + \frac{e}{\hbar}n\boldsymbol{E}\cdot\langle\nabla_{\boldsymbol{k}}\otimes\boldsymbol{u}_{\mathrm{n}}\rangle + \nabla\cdot[\,n\langle\boldsymbol{w}_{\mathrm{n}}\otimes\boldsymbol{w}_{\mathrm{n}}\rangle] \\ = \frac{1}{4\pi^3}\int \boldsymbol{w}_{\mathrm{n}}\,C[f]\,\mathrm{d}^3k\,. \qquad (2.62)\end{aligned}$$

In the effective-mass approximation with an isotropic effective mass m_{n}^* and a scalar electron temperature [9] of the random velocity component $\boldsymbol{w}_{\mathrm{n}}$ the tensor $\langle\boldsymbol{w}_{\mathrm{n}}\otimes\boldsymbol{w}_{\mathrm{n}}\rangle$ is proportional to the unit tensor, i.e.

$$\langle\boldsymbol{w}_{\mathrm{n}}\otimes\boldsymbol{w}_{\mathrm{n}}\rangle_{\alpha\beta} = \frac{1}{3}\langle|\boldsymbol{w}_{\mathrm{n}}(\boldsymbol{k})|^2\rangle\,\delta_{\alpha\beta}\,.$$

[9] This corresponds to the assumption that the expectation values of the squares of all three components are equal: $\langle w_{\mathrm{n}x}^2\rangle = \langle w_{\mathrm{n}y}^2\rangle = \langle w_{\mathrm{n}z}^2\rangle$. Formulations of the hydrodynamic model that also apply to nonparabolic energy dispersion relationships and a nonisotropic "temperature" were published in [25 27]; multivalley semiconductors are considered in [17, 28–30], for example.

Since $m_\mathrm{n}^*\langle|\boldsymbol{w}_\mathrm{n}(\boldsymbol{k})|^2\rangle/2$ is the average kinetic energy (per electron) associated with the random motion of the electrons, we may define a local electron temperature

$$T_\mathrm{n}(\boldsymbol{x}) = \frac{m_\mathrm{n}^*}{3k_\mathrm{B}}\langle|\boldsymbol{w}_\mathrm{n}(\boldsymbol{k})|^2\rangle \tag{2.63}$$

in analogy with the equipartition theorem of classical statistical mechanics.[10] Under these conditions (2.62) simplifies to

$$n\left(\frac{\partial}{\partial t} + \boldsymbol{v}_\mathrm{n}\cdot\nabla\right)\boldsymbol{v}_\mathrm{n} + \frac{en}{m_\mathrm{n}^*}\boldsymbol{E} + \frac{k_\mathrm{B}}{m_\mathrm{n}^*}\nabla\cdot(nT_\mathrm{n}) = n\left(\frac{\partial \boldsymbol{v}_\mathrm{n}}{\partial t}\right)_\mathrm{c}. \tag{2.64}$$

This equation is termed the velocity or momentum balance equation; the term

$$n\left(\frac{\partial \boldsymbol{v}_\mathrm{n}}{\partial t}\right)_\mathrm{c} = \frac{1}{4\pi^3}\int \boldsymbol{w}_\mathrm{n} C_\mathrm{n}[f]\,\mathrm{d}^3k \tag{2.65}$$

gives the rate of change with time of the mean electron velocity due to collisions. This term can be represented as

$$\left(\frac{\partial \boldsymbol{v}_\mathrm{n}}{\partial t}\right)_\mathrm{c} = -\frac{\boldsymbol{v}_\mathrm{n}}{\tau_\mathrm{vn}}, \tag{2.66}$$

where τ_vn is the momentum relaxation time for electrons. Introducing the electron mobility

$$\mu_\mathrm{n} = e\tau_\mathrm{vn}/m_\mathrm{n}^* \tag{2.67}$$

and taking account of the fact that $\boldsymbol{J}_\mathrm{n} = -en\boldsymbol{v}_\mathrm{n}$, the current equation

$$\boldsymbol{J}_\mathrm{n} = e\mu_\mathrm{n} n\boldsymbol{E} + k_\mathrm{B}\mu_\mathrm{n}\nabla(nT_\mathrm{n}) \tag{2.68}$$

is obtained, if the terms $en\tau_\mathrm{vn}\,\partial\boldsymbol{v}_\mathrm{n}/\partial t$ and $en\tau_\mathrm{vn}\,\boldsymbol{v}_\mathrm{n}\cdot\nabla\boldsymbol{v}_\mathrm{n}$ are assumed to be negligible.[11] Under isothermal conditions, the conventional equation for the electron current density in the drift–diffusion approximation results, together with the Einstein relation for nondegenerate semiconductors.

2.5.3 Energy Balance Equation

Using $a(\boldsymbol{k}) = w_\mathrm{c}(\boldsymbol{k})$ in (2.57) gives the energy balance equation

$$\frac{\partial}{\partial t}(nW_\mathrm{n}) + \nabla\cdot(n\langle\boldsymbol{u}_\mathrm{n}w_\mathrm{c}\rangle) - \frac{e}{\hbar}n\boldsymbol{E}\cdot\langle\nabla_{\boldsymbol{k}}w_\mathrm{c}\rangle = \frac{1}{4\pi^3}\int w_\mathrm{c}\,C_\mathrm{n}[f]\,\mathrm{d}^3k,$$

[10] According to this theorem, each degree of freedom contributes on average $k_\mathrm{B}T/2$ to the thermal energy of a system held at a temperature T. When we take account of the fact that each electron has three (translational) degrees of freedom, (2.63) results.

[11] This is reasonable, since the hydrodynamic velocity may be assumed to be approximately constant within a time interval of the order τ_vn and a length interval of the order of $|\tau_\mathrm{vn}\boldsymbol{v}_\mathrm{n}|$, respectively, as the momentum relaxation times lies below 1 ps.

where

$$W_{\rm n}(\boldsymbol{x}) \;=\; \frac{1}{4\pi^3 n(\boldsymbol{x},t)} \int w_{\rm c}(\boldsymbol{k})\, f(\boldsymbol{x},\boldsymbol{k},t)\, {\rm d}^3k \;.$$

In the effective-mass approximation, the average kinetic energy $W_{\rm n}(\boldsymbol{x})$ of the electrons in the conduction band now reads

$$\begin{aligned} W_{\rm n}(\boldsymbol{x}) \;&=\; \frac{1}{2}\, m_{\rm n}^*\, \langle(\boldsymbol{v}_{\rm n}+\boldsymbol{w}_{\rm n})^2\rangle \;=\; \frac{1}{2}\, m_{\rm n}^*\, |\boldsymbol{v}_{\rm n}|^2 + \frac{1}{2}\, m_{\rm n}^*\, \langle|\boldsymbol{w}_{\rm n}(\boldsymbol{k})|^2\rangle \\ &=\; \frac{1}{2}\, m_{\rm n}^*|\boldsymbol{v}_{\rm n}|^2 + \frac{3}{2}\, k_{\rm B}\, T_{\rm n}(\boldsymbol{x}) \;, \end{aligned} \tag{2.69}$$

where the first term on the right-hand side describes the contribution due to the streaming motion, and the second describes the kinetic energy associated with the random thermal motion of the carriers. This allows us to transform the energy balance equation into

$$\begin{aligned} & n\left(\frac{\partial}{\partial t} + \boldsymbol{v}_{\rm n}\cdot\nabla\right) W_{\rm n} + \nabla\cdot\boldsymbol{Q}_{\rm n} + k_{\rm B}\,\nabla\cdot(nT_{\rm n}\boldsymbol{v}_{\rm n}) + en\boldsymbol{v}_{\rm n}\cdot\boldsymbol{E} \\ &\quad =\; n\left(\frac{\partial W_{\rm n}}{\partial t}\right)_{\rm c} \;. \end{aligned} \tag{2.70}$$

In this equation,

$$\boldsymbol{Q}_{\rm n} \;=\; \frac{1}{2} m_{\rm n}^*\, n\, \langle|\boldsymbol{w}_{\rm n}|^2 \boldsymbol{w}_{\rm n}\rangle \tag{2.71}$$

defines the heat flow, which describes the transport of thermal energy due to the random motion of the electrons. The term $(\partial W_{\rm n}/\partial t)_{\rm c}$ gives the rate of change with time of the average electron kinetic energy due to collisions; this term can be represented as

$$\left(\frac{\partial W_{\rm n}}{\partial t}\right)_{\rm c} \;=\; -\frac{W_{\rm n}-W_{\rm n0}}{\tau_{\rm wn}} \;, \tag{2.72}$$

where $\tau_{\rm wn}$ is the energy relaxation time for electrons, and $W_{\rm n0} = 3k_{\rm B}T_{\rm L}/2$ denotes the average thermal energy of electrons that are in equilibrium with the crystal lattice, at a lattice temperature $T_{\rm L}$.

The continuity equation relates the moment of order zero (i.e. the particle density) to the moment of order one (i.e. the mean particle velocity). The velocity balance equation relates the moment of order one to the moment of order two (represented by the temperature tensor). The energy balance equation relates the moment of order two to the moment of order three (represented by the heat flow vector). Introducing higher-moment equations leads to an infinite set of interrelated balance equations. In order to arrive at a tractable problem, the infinite series of balance equations must be truncated by the introduction of a suitable approximation.[12] If the electron temperature is assumed to be a scalar quantity, and the heat flow is approximated by the phenomenological equation[13]

$$\boldsymbol{Q}_{\mathrm{n}} = -\kappa_{\mathrm{n}} \nabla T_{\mathrm{n}} , \tag{2.73}$$

where κ_{n} is the electron thermal conductivity, the energy balance equation transforms into

$$n \left(\frac{\partial}{\partial t} + \boldsymbol{v}_{\mathrm{n}} \cdot \nabla \right) W_{\mathrm{n}} - \kappa_{\mathrm{n}} \nabla^2 T_{\mathrm{n}} + k_{\mathrm{B}} \nabla \cdot (n T_{\mathrm{n}} \boldsymbol{v}_{\mathrm{n}}) + en \boldsymbol{E} \cdot \boldsymbol{v}_{\mathrm{n}}$$
$$= n \left(\frac{\partial W_{\mathrm{n}}}{\partial t} \right)_{\mathrm{c}} . \tag{2.74}$$

The hydrodynamic model has been implemented in various device simulation programs [35–37] and is also available as an option in the widely employed simulation program MEDICI.

Application: Velocity Saturation

As a simple example, the dc current passing through a homogeneously doped n-type resistor with constant cross section and temperature will be considered. The quantities n, W_{n}, T_{n} and $\boldsymbol{Q}_{\mathrm{n}}$ are independent of position, and (2.68) and (2.70) simplify to the following scalar equations governing $J_{\mathrm{n}} = |\boldsymbol{J}_{\mathrm{n}}|$ and $E = |\boldsymbol{E}|$

$$J_{\mathrm{n}} = e \mu_{\mathrm{n}} n E \quad \text{and} \quad J_{\mathrm{n}} E = \frac{W_{\mathrm{n}} - W_{\mathrm{n}0}}{\tau_{\mathrm{wn}}} , \tag{2.75}$$

if one takes account of the fact that $\boldsymbol{J}_{\mathrm{n}}$ and $\boldsymbol{E}$ are parallel. In [38] it was shown that the scattering rate $1/\tau_{\mathrm{vn}}$ increases approximately in proportion to the kinetic energy W_{n} and may therefore be written as

$$\frac{1}{\tau_{\mathrm{vn}}} = \frac{1}{\tau_{\mathrm{v}0}} \left(1 + \eta \frac{W_{\mathrm{n}} - W_{\mathrm{n}0}}{n} \right) , \tag{2.76}$$

where $\tau_{\mathrm{v}0}$ and η are material-specific parameters that vary with temperature, doping concentration, etc. Together with the relation $\mu_{\mathrm{n}} = e\tau_{\mathrm{vn}}/m_{\mathrm{n}}^*$, (2.76) and (2.75) may be combined to give the following equation for $W_{\mathrm{n}} - W_{\mathrm{n}0}$:

[12] Further work addressing the closure of the infinite series of balance equations obtained from the method of moments can be found in [31–33].

[13] This equation is commonly referred to as Fourier's law. In [34], substantial deviations of the actual heat flux from Fourier's law have been reported. There the improved relation

$$\boldsymbol{Q}_{\mathrm{n}} = \frac{5}{2} \left(1 - \frac{\tau_{\mathrm{wn}}}{\tau_{\mathrm{vn}}} \right) \frac{k_{\mathrm{B}} T_{\mathrm{n}}}{e} \boldsymbol{J}_{\mathrm{n}} - \kappa_{\mathrm{n}} \nabla T_{\mathrm{n}} \quad \text{with} \quad \kappa_{\mathrm{n}} = \frac{5}{2} \frac{\tau_{\mathrm{wn}}}{\tau_{\mathrm{vn}}} k_{\mathrm{B}} n D_{\mathrm{n}}$$

with $D_{\mathrm{n}} = \mu_{\mathrm{n}} k_{\mathrm{B}} T_{\mathrm{n}}/e$, is derived.

$$W_{\mathrm{n}} - W_{\mathrm{n0}} = \frac{e\mu_{\mathrm{n0}} n \tau_{\mathrm{wn}} E^2}{1+\eta(W_{\mathrm{n}}-W_{\mathrm{n0}})/n} ,$$

where $\mu_{\mathrm{n0}} = e\tau_{\mathrm{v0}}/m_{\mathrm{n}}^*$. This equation can be solved for $W_{\mathrm{n}} - W_{\mathrm{n0}}$, resulting in

$$\mu_{\mathrm{n}}(E) = \frac{W_{\mathrm{n}} - W_{\mathrm{n0}}}{en\tau_{\mathrm{wn}}E^2} = \frac{2\mu_{\mathrm{n0}}}{1 + \sqrt{1 + 4e\mu_{\mathrm{n0}}\eta\tau_{\mathrm{wn}}E^2}} . \tag{2.77}$$

Owing to the increase of the scattering rate $1/\tau_{\mathrm{vn}}$, the electron mobility μ_{n} decreases with increasing electric-field strength E. For large values of E this results in a saturation of the drift velocity $|v_{\mathrm{n}}| = \mu_{\mathrm{n}}E \to v_{\mathrm{nsat}}$ (see Fig. 2.7), i.e. the mobility shows the asymptotic behavior $\mu_{\mathrm{n}} \approx v_{\mathrm{nsat}}/E$. A comparison with (2.77) results in in $\eta = \mu_{\mathrm{n0}}/(e\tau_{\mathrm{wn}}v_{\mathrm{nsat}}^2)$ and

$$\mu_{\mathrm{n}}(E) = \frac{2\mu_{\mathrm{n0}}}{1 + \sqrt{1 + (2E\mu_{\mathrm{n0}}/v_{\mathrm{nsat}})^2}} . \tag{2.78}$$

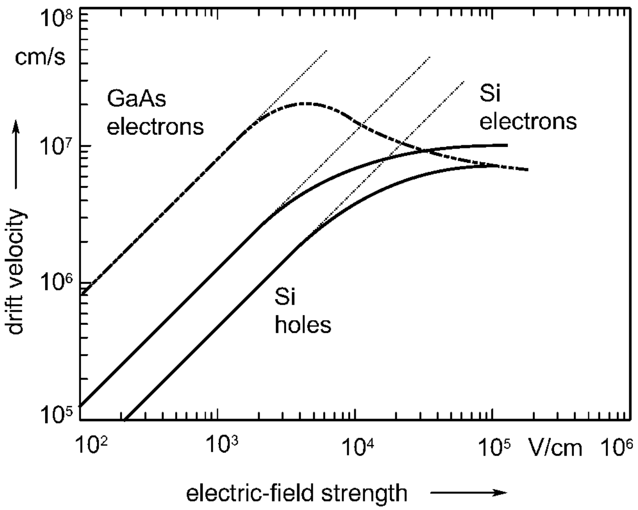


Fig. 2.7. Drift velocity of electrons and holes in Si and drift velocity of electrons in GaAs

Figure 2.7 shows the drift velocities of electrons and holes in silicon which monotonously increase versus the saturation drift velocity v_{nsat}, which is of the order of 10^7 cm/s. The drift velocity of electrons in GaAs, also shown in Fig. 2.7, does not increase monotonously with the electric-field strength E and decreases at large values of E; this is explained by increased scattering of electrons from the conduction band minimum at Γ to the conduction band minima at the L-points, which have a larger effective mass, corresponding to a reduced electron velocity (Gunn effect). To explain such effects an extension of the simple hydrodynamic model presented here is required [28].

2.6 Generation and Recombination

In moderately doped indirect semiconductors such as silicon and germanium, generation–recombination processes via deep traps determine the lifetime of minority carriers to a large extent – in heavily doped regions, Auger processes[14] become relevant. Modeling the reverse currents of pn junctions also requires models for tunneling and impact ionization.

2.6.1 Shockley–Read–Hall Processes

Generation–recombination processes via deep traps were first investigated by Hall [39] and Shockley and Read [40, 41]; such processes are usually referred to as Shockley Read Hall (SRH) processes. Classical SRH theory applies to the situation where only one kind of trapping state is present. As an example, we consider the single-level acceptor-like trap illustrated in Fig. 2.8. There are only two states possible: the trap is either neutral or charged by a trapped electron.

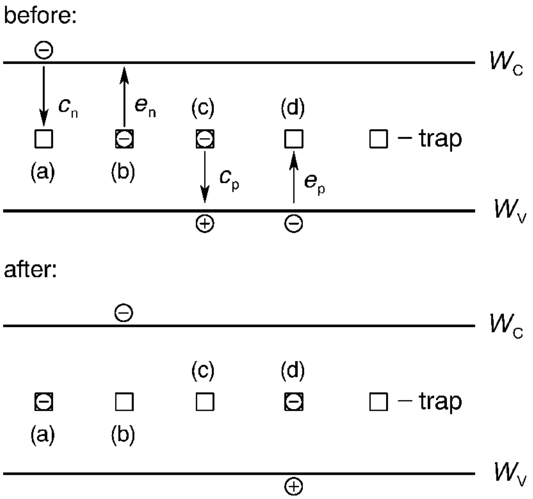


Fig. 2.8. Electron transitions involving a deep acceptor-like state [41]

A conduction band electron can fall into a localized trap state[15] either by cascading down through several excitation levels of the trap, by radiative or Auger processes or by means of multiple phonon emission. We may distinguish four elementary processes: (a) capture of a conduction band electron,

[14]In direct semiconductors, radiative interband transitions also play a substantial role.

[15]The energy level of any trap state that is relevant for SRH processes lies in the energy gap, with a separation from the conduction band edge that significantly exceeds the phonon energy.

(b) electron emission from the deep state to the conduction band, (c) recombination of a trapped electron with a hole in the valence band (this process may equally well be described as the capture of a valence band hole by the charged trap) and (d) capture of a valence band electron leaving a hole in the valence band (this process may equally well be described as emission of a hole from the deep state to the valence band). Following [41,42], we denote the transition rates of the four processes by $c_\mathrm{n}n$, e_n, $c_\mathrm{p}p$ and e_p respectively; the density of deep trap states is denoted by N_T and their probability of occupation by f_t. The change of the electron density due to interaction with the trap is then[16]

$$\left.\frac{\partial n}{\partial t}\right|_\mathrm{SRH} = N_\mathrm{T}\left[e_\mathrm{n}f_\mathrm{t} - c_\mathrm{n}n(1-f_\mathrm{t})\right] = N_\mathrm{T}\,\frac{\tilde{n}_1 f_\mathrm{t} - n(1-f_\mathrm{t})}{\tau_\mathrm{n0}}\,, \tag{2.79}$$

where $\tau_\mathrm{n0} = 1/c_\mathrm{n}$ and $\tilde{n}_1 = e_\mathrm{n}/c_\mathrm{n}$. In this equation, the first term on the right-hand side describes the increase of the electron density n due to electron emission from the trap to the conduction band, while the second term describes the decrease of the electron density due to electrons captured by the trap. The occupation probability f_t of the trap appears in this equation since only occupied traps – with density $N_\mathrm{T}f_\mathrm{t}$ – may emit electrons, while, owing to the Pauli principle, only unoccupied traps – with density $N_\mathrm{T}(1-f_\mathrm{t})$ – are able to capture electrons. In complete analogy, the change of the hole density p due to interaction with the trap is given by

$$\left.\frac{\partial p}{\partial t}\right|_\mathrm{SRH} = N_\mathrm{T}\left[e_\mathrm{p}(1-f_\mathrm{t}) - c_\mathrm{p}pf_\mathrm{t}\right] = N_\mathrm{T}\,\frac{\tilde{p}_1(1-f_\mathrm{t}) - pf_\mathrm{t}}{\tau_\mathrm{p0}}\,, \tag{2.80}$$

where $\tau_\mathrm{p0} = 1/c_\mathrm{p}$ and $\tilde{p}_1 = e_\mathrm{p}/c_\mathrm{p}$. Assuming stationary conditions for the probability of trap occupation ($\partial f_\mathrm{t}/\partial t = 0$), the rate of change of the electron density $\partial n/\partial t|_\mathrm{SRH}$ equals the rate of change of the hole density $\partial p/\partial t|_\mathrm{SRH}$ since electrons and holes are generated and recombine in pairs. From this, the probability of occupation can be derived:

$$f_\mathrm{t} = \frac{c_\mathrm{n}n + e_\mathrm{p}}{c_\mathrm{n}n + e_\mathrm{n} + c_\mathrm{p}p + e_\mathrm{p}}\,.$$

This results in a net recombination rate

$$R - G = -\left.\frac{\partial p}{\partial t}\right|_\mathrm{SRH} = -\left.\frac{\partial n}{\partial t}\right|_\mathrm{SRH} = \frac{N_\mathrm{T}(np - \tilde{n}_1\tilde{p}_1)}{\tau_\mathrm{p0}(n+\tilde{n}_1) + \tau_\mathrm{n0}(p+\tilde{p}_1)} \tag{2.81}$$

---

[16] Equations (2.79) and (2.80) assume noninteracting trap states: The density N_T of traps is assumed to be so small that their average separation $N_\mathrm{T}^{-1/3}$ is large in comparison with the extent of the wave function of the bound state. This assumption implies that the transition rate between adjacent traps is much smaller than the transition rates to the conduction and valence bands.

without any additional assumptions. In particular, no assumptions concerning the ratios $\tilde{n}_1 = e_n/c_n$ and $\tilde{p}_1 = e_p/c_p$ have been introduced. In classical SRH theory $\tilde{n}_1$ and $\tilde{p}_1$ are replaced by their thermal-equilibrium values

$$\tilde{n}_1 \approx n_1 = N_C \exp\frac{W_T - W_C}{k_B T} = n_i \exp\frac{W_T - W_{Fi}}{k_B T} \tag{2.82}$$

and

$$\tilde{p}_1 \approx p_1 = N_V \exp\frac{W_V - W_T}{k_B T} = n_i \exp\frac{W_{Fi} - W_T}{k_B T}\,, \tag{2.83}$$

where W_T is the energy of the trapping state and

$$W_{Fi} = (W_C + W_V)/2 - k_B T \ln\sqrt{N_C/N_V}$$

denotes the intrinsic Fermi level. This approximation yields the widely published expression

$$R - G = N_T \frac{pn - n_i^2}{\tau_{p0}(n + n_1) + \tau_{n0}(p + p_1)}\,, \tag{2.84}$$

where n_1 and p_1 are as defined in (2.82) and (2.83). In the special case where $\tau_{p0} = \tau_{n0} = \tau N_T$, (2.84) simplifies to

$$R - G = \frac{1}{\tau}\frac{pn - n_i^2}{n + p + 2n_i \cosh\left[\,(W_T - W_{Fi})/k_B T\,\right]}\,; \tag{2.85}$$

the recombination rate thus depends strongly on the energy W_T of the recombination center and assumes its maximum value if W_T approaches W_{Fi}. If trap states with different energy levels are present, (2.84) has to be generalized to

$$R - G = \int_{W_V}^{W_C} \frac{(pn - n_i^2)\, N_T'(W_T)\, \mathrm{d}W_T}{\tau_{p0}(W_T)\,[\,n + n_1(W_T)] + \tau_{n0}(W_T)\,[\,p + p_1(W_T)]}\,,$$

where $N_T'(W_T)\,\mathrm{d}W_T$ denotes the density of trapping centers with energy levels in the interval between W and $W + \mathrm{d}W$.

In n-type semiconductors, under low-level-injection conditions, the term $\tau_{p0} n \approx \tau_{p0} N_D^+$ is predominant in the denominator of (2.84). Taking account of the fact that the equilibrium minority-carrier density is $p_{n0} \approx n_i^2/N_D^+$, the net recombination rate therefore reads

$$R - G \approx \frac{p_n n_n - n_i^2}{\tau_{pSRH} N_D^+} = \frac{p_n - p_{n0}}{\tau_{pSRH}}\,, \tag{2.86}$$

where $\tau_{pSRH} = \tau_{p0}/N_T$ is the (minority-carrier) lifetime for excess holes determined by SRH processes. In complete analogy, the term $\tau_{n0} p \approx \tau_{n0} N_A^-$ dominates the denominator of (2.84) in quasi-neutral p-type regions under low-level injection conditions, resulting in

$$R - G \approx \frac{p_\mathrm{p} n_\mathrm{p} - n_\mathrm{i}^2}{\tau_\mathrm{nSRH} N_\mathrm{A}^-} = \frac{n_\mathrm{p} - n_\mathrm{p0}}{\tau_\mathrm{nSRH}} , \tag{2.87}$$

where $\tau_\mathrm{nSRH} = \tau_\mathrm{n0}/N_\mathrm{T}$ is the (minority-carrier) lifetime for excess electrons determined by SRH processes. Equations (2.86) and (2.87) are widely used in the analysis of minority-carrier transport. In combination with the current and continuity equations, the diffusion equation for minority-carrier transport is obtained.

In the derivation of (2.84), the trap occupancy was assumed to be stationary. The formula therefore applies only to operation conditions that vary slowly on the timescale of electron capture and emission processes. The second restriction stems from the assumptions (2.82) and (2.83), which are reasonable if the energy consumed or dissipated during capture and emission processes is exchanged with a heat bath. In the presence of electric fields, electrons may also tunnel into or out of trap states, an effect that will influence the ratios $\tilde{n}_1 = e_\mathrm{n}/c_\mathrm{n}$ and $\tilde{p}_1 = e_\mathrm{p}/c_\mathrm{p}$.

Generation and Recombination at Interfaces, Surface Recombination Velocity

Interfaces between semiconductors and insulators may affect the recombination current of pn junctions in a twofold manner. (1) There will be generation and recombination at the interface due to trapping centers. (2) Charges trapped in the insulator and at the semiconductor–insulator interface cause a band bending in the semiconductor region below the insulator, which has an influence on the SRH recombination in this region.

The formula for SRH processes can be adapted to surface recombination if $N'_\mathrm{T}(W_\mathrm{T})\,\mathrm{d}W_\mathrm{T}$ is replaced by the density (per unit area) of surface states $D_\mathrm{it}(W)\,\mathrm{d}W$ with energy levels in the interval between W and $W+\mathrm{d}W$:

$$(R - G)' = \int_{W_\mathrm{V}}^{W_\mathrm{C}} \frac{(pn - n_\mathrm{i}^2)\, D_\mathrm{it}(W_\mathrm{T})\, \mathrm{d}W_\mathrm{T}}{\tau_\mathrm{p0}(W_\mathrm{T})\,[n+n_1(W_\mathrm{T})] + \tau_\mathrm{n0}(W_\mathrm{T})\,[p+p_1(W_\mathrm{T})]} .$$

This recombination rate forms a boundary condition for the electron and hole current densities at the insulating surface, $e(R-G)' = \boldsymbol{n}\cdot\boldsymbol{J}_\mathrm{p} = -\boldsymbol{n}\cdot\boldsymbol{J}_\mathrm{n}$, where $\boldsymbol{n}$ is the unit vector normal to the insulating surface (Fig. 2.9). Under low-level-injection conditions in n-type material,[17] we may write

$$p_\mathrm{n} n_\mathrm{n} - n_\mathrm{i}^2 \approx N_\mathrm{D}(p_\mathrm{n} - n_\mathrm{i}^2/N_\mathrm{D}) = N_\mathrm{D}(p_\mathrm{n} - p_\mathrm{n0})$$

and modify the general expression for the net surface recombination rate to

$$e(R-G)' = S_\mathrm{p}(p_\mathrm{n} - p_\mathrm{n0}) , \tag{2.88}$$

[17] Similar considerations hold for the surface recombination of electrons in p-type material under low-level-injection conditions.

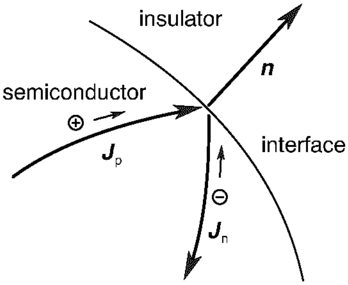


Fig. 2.9. Electron and hole current densities recombining at a semiconductor–insulator interface

where

$$S_{\mathrm{p}} = \int_{W_{\mathrm{V}}}^{W_{\mathrm{C}}} \frac{N_{\mathrm{D}}\, D_{\mathrm{it}}(W_{\mathrm{T}})\, \mathrm{d}W_{\mathrm{T}}}{\tau_{\mathrm{p0}}(W_{\mathrm{T}})\,[n + n_1(W_{\mathrm{T}})] + \tau_{\mathrm{n0}}(W_{\mathrm{T}})\,[p + p_1(W_{\mathrm{T}})]} \tag{2.89}$$

defines the surface recombination velocity at the semiconductor–insulator interface. In heavily doped n-type regions it is reasonable to assume that $n \approx N_{\mathrm{D}} \gg p$, which allows us to simplify the expression for the surface recombination velocity to

$$S_{\mathrm{p}} \approx \int_{W_{\mathrm{V}}}^{W_{\mathrm{C}}} \frac{D_{\mathrm{it}}(W_{\mathrm{T}})\, \mathrm{d}W_{\mathrm{T}}}{\tau_{\mathrm{p0}}(W_{\mathrm{T}})} \,. \tag{2.90}$$

In this approximation, the surface recombination velocity is independent of the hole density at the interface. Generally, however, the surface recombination velocity is not constant; it varies with the injection level and the state of the surface (accumulation or depletion) (see, for example, [43, 44]).

Field-Enhanced Shockley–Read–Hall Recombination

In the presence of electric fields, the rates c_{n}, c_{p}, e_{n} and e_{p} for capture and emission of electrons and holes have to be treated as field-dependent quantities owing to tunneling, thermally assisted tunneling and the Poole–Frenkel effect.[18] The Poole–Frenkel effect describes the enhancement of thermal emission processes out of a potential well induced by an external field E. For a Coulomb well, such a field reduces the potential barrier that has to be overcome by an electron before it can be emitted into the conduction band. The thermal emission rate may be written as

$$e_{\mathrm{n,PF}} = e_{\mathrm{n0}} \exp\left(-\frac{W_{\mathrm{C}} - W_{\mathrm{T}} - e\Delta\psi}{k_{\mathrm{B}}T}\right) , \tag{2.91}$$

where $\Delta\psi = \sqrt{e|E|/\pi\epsilon_{\mathrm{Si}}}$, i.e. the emission rate is enhanced by a factor $\exp(\Delta\psi/V_{\mathrm{T}})$ due to the presence of the electric field (Fig. 2.10). In the presence of a large electric field, electrons may also tunnel from the trap state to

[18] A derivation of the effects of electric field and temperature on the Shockley Read Hall lifetimes in silicon from the microscopic level is given in [45].

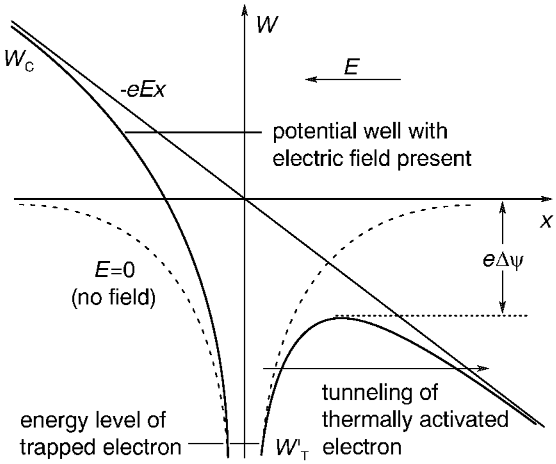


Fig. 2.10. Energy band diagram in the vicinity of a positively charged trapping center of Coulomb type. Electrons may leave the localized state by thermally assisted tunneling or by thermal emission (Poole Frenkel mechanism)

the conduction band. The total emission rate for electrons out of a trap state due to both thermal emission and tunneling was calculated in [46], with the result

$$\frac{e_n}{e_n(0)} = 1 + \int_0^{\Delta W_n/k_B T} \exp\left(z - \frac{4(2m^*)^{1/2} k_B T^{3/2}}{3e\hbar|E|} z^{3/2}\right) dz , \qquad (2.92)$$

where $e_n(0)$ denotes the emission probability in the absence of an electric field, and $\Delta W_n = W_C - W_T$. This formula is used in several approaches to trap-assisted tunneling. In [47], a generalization of the classical SRH recombination rate (2.84), with the lifetimes $\tau_{n0}(E)$ and $\tau_{p0}(E)$ written as field-dependent quantities, has been published. Equation (2.92) has been derived for a Dirac well, which may serve as a model for a neutral trapping state; modified expressions result if the trap center is charged, as was pointed out in [48] for example: if the electron is trapped in a Coulomb well, barrier lowering due to the Poole–Frenkel effect will significantly enhance the emission rate. If this is taken into account, more accurate expressions for the generation rate are obtained.

2.6.2 Auger Recombination

At high carrier or doping concentrations, the predominant recombination mechanism in elemental semiconductors such as silicon is the so-called Auger process. In such a process the energy that is set free if an electron hole pair recombines is transferred[19] to another electron (eeh process) or hole (ehh process). Since such processes require two electrons and one hole or two holes and

[19] At least to a large extent, since more detailed pictures of the Auger effect show that additional interactions with lattice vibrations (phonons) and impurities play an important role in Auger processes.

one electron to collide, it is to be expected that the corresponding recombination probabilities vary in proportion to n^2p and p^2n. Auger recombination is therefore generally modeled by a net recombination rate

$$R - G = (np - n_{\mathrm{i}}^2)(C_{\mathrm{n}}n + C_{\mathrm{p}}p) , \tag{2.93}$$

with constants C_{n} and C_{p} determined from measurements. Table 2.1 shows results determined from the decay of the photoluminescence after excitation of heavily doped silicon samples with a short laser pulse at different temperatures, published in [49].

Table 2.1. Parameters C_{n} and C_{p} for Auger recombination in silicon [49]

Temperature	77 K	300 K	400 K
$C_{\mathrm{n}}/(\mathrm{cm}^6\,\mathrm{s}^{-1})$	2.3×10^{-31}	2.8×10^{-31}	2.8×10^{-31}
$C_{\mathrm{p}}/(\mathrm{cm}^6\,\mathrm{s}^{-1})$	7.8×10^{-32}	9.9×10^{-32}	1.2×10^{-31}

2.6.3 Impact Ionization

The generation rate G_{i} due to impact ionization of electrons and holes varies in proportion to the corresponding current density:

$$G_{\mathrm{i}} = \frac{1}{e}\left(\alpha_{\mathrm{n}}|\boldsymbol{J}_{\mathrm{n}}| + \alpha_{\mathrm{p}}|\boldsymbol{J}_{\mathrm{p}}|\right) . \tag{2.94}$$

The material-specific factors, α_{n} and α_{p}, are termed the ionization coefficients of electrons and holes. A similar formula was employed earlier by Townsend in his theory of impact ionization phenomena due to electrons in gases. Townsends theory was generalized to two-component systems (electrons and holes) by McKay [50], who investigated impact ionization phenomena in reverse-biased pn junctions. His original assumption of identical ionization coefficients for both electrons and holes was soon demonstrated to be incorrect [51]; in particular, $\alpha_{\mathrm{n}} > \alpha_{\mathrm{p}}$ was shown for silicon [52].

Multiplication Factor

Adding the generation rate G_{i} determined from (2.94) to the net recombination rate $(R - G)$ of all other generation–recombination processes yields the following continuity equations for electrons and holes,

$$\frac{\mathrm{d}J_{\mathrm{n}}}{\mathrm{d}x} = e\,(R-G) - \alpha_{\mathrm{n}}J_{\mathrm{n}} - \alpha_{\mathrm{p}}J_{\mathrm{p}}$$

$$\frac{\mathrm{d}J_{\mathrm{p}}}{\mathrm{d}x} = -e\,(R-G) + \alpha_{\mathrm{n}}J_{\mathrm{n}} + \alpha_{\mathrm{p}}J_{\mathrm{p}} ,$$

if one-dimensional current flow with current densities $J_{\mathrm{n}} > 0$, $J_{\mathrm{p}} > 0$ (electron drift from x_{p} to $x_{\mathrm{n}} < x_{\mathrm{p}}$) and stationary conditions are assumed. In this case the current density $J = J_n + J_p$ is independent of position; its value can be calculated from the continuity equations as [53]

$$J = J_n(x_p)M_n + J_p(x_n)M_p - eM_p \int_{x_n}^{x_p} (R-G)\, e^{\theta(x)}\, dx \,, \tag{2.95}$$

where x_n and x_p denote the boundaries of the depletion layer. The function

$$\theta(x) = \int_{x_n}^{x} [\alpha_n(x') - \alpha_p(x')]\; dx'$$

takes account of different ionization coefficients for electrons and holes; if these can be assumed to be equal, $\theta(x) = 0$ and $\exp[\theta(x)] = 1$. The quantity M_n, defined by

$$1 - \frac{1}{M_n} = \int_{x_n}^{x_p} \alpha_n \exp\left[-\int_{x}^{x_p} (\alpha_n - \alpha_p)\, dx'\right]\, dx$$

denotes the multiplication factor for injected electrons: an electron current injected into the space charge layer at x_p is amplified by a factor M_n in traversing the space charge layer. The quantity M_p, defined by

$$1 - \frac{1}{M_p} = \int_{x_n}^{x_p} \alpha_p \exp\left[\int_{0}^{x} (\alpha_n - \alpha_p)\, dx'\right]\, dx$$

gives the corresponding multiplication factor for injected holes: a hole current injected at x_n is amplified by M_p in traversing the space charge layer. M_n and M_p generally have different values; in particular, $M_n > M_p$ in silicon. The third term on the right-hand side of (2.95) describes the contribution due to primary carriers generated in the space charge layer.

Chynoweth Formula, Lucky-Electron Model

Measurements by Chynoweth [54] suggested that the ionization rates for electrons and holes may be expressed as functions of the local-electric-field strength as follows:[20]

$$\alpha_n = a_n \exp(-b_n/E) \quad \text{and} \quad \alpha_n = a_p \exp(-b_p/E) \,,$$

where a_n, b_n, a_p and b_p are material specific constants. Various investigators [53, 56–59] determined multiplication factors in pn junctions and used these data for the computation of ionization coefficients with a local-electric-field dependence according to the above Chynoweth formula. Such a dependence

[20] If there is a substantial electric-field component perpendicular to the current flow, as is the case in the channel of a MOSFET, the Chynoweth expressions should be modified to [55]

$$\alpha_n = a_n \exp\left(-\frac{b_n|\boldsymbol{J}_n|}{|\boldsymbol{E}\cdot\boldsymbol{J}_n|}\right) \quad \text{and} \quad \alpha_p = a_p \exp\left(-\frac{b_p|\boldsymbol{J}_p|}{|\boldsymbol{E}\cdot\boldsymbol{J}_p|}\right) .$$

may be understood in terms of Shockley's simplifying[21] lucky electron model [61–63] of impact ionization, which considers two energy loss mechanisms for electrons crossing the depletion layer of a pn junction:

1. Optical-phonon scattering characterized by a mean free path λ_r. The probability for an electron to traverse a distance d without being scattered by an optical phonon is $P = \exp(-d/\lambda_\mathrm{r})$.
2. Impact ionization. As soon as the kinetic energy of an electron exceeds the ionization energy[22] W_i, it may generate an electron–hole pair. The mean free path of electrons with kinetic energies $W > W_\mathrm{i}$ for impact ionization is denoted by λ_i. To gain an energy W_i in an electric field of strength E, the electron has to travel a distance $W_\mathrm{i}/(eE)$. The probability that the electron is not scattered by an optical phonon before this happens is $P(W_\mathrm{i}) = \exp(-W_\mathrm{i}/e\lambda_\mathrm{r}E)$. As soon as the energy exceeds W_i, electron-hole pairs may be generated. The probability that an electron with kinetic energy $W > W_\mathrm{i}$ indeed produces an electron–hole pair and does not lose its energy owing to optical-phonon scattering is $\lambda_\mathrm{r}/\lambda_\mathrm{i}$, i.e.

$$P_\mathrm{I} = \frac{\lambda_\mathrm{r}}{\lambda_\mathrm{i}} \exp\left(-\frac{W_\mathrm{i}}{e\lambda_\mathrm{r}}\frac{1}{E}\right)$$

determines the probability that an electron which is accelerated in the electric field will generate an electron–hole pair. The ionization coefficient is proportional to this probability.[23]

The generation rate due to impact ionization is represented well by (2.94); the assumption of ionization coefficients $\alpha_\mathrm{n,p}(\boldsymbol{x}) = \alpha_\mathrm{n,p}[\,|\boldsymbol{E}(\boldsymbol{x})|\,]$ that depend only on the magnitude of the local electric field strength[24] however, is only applicable if $|\boldsymbol{E}(\boldsymbol{x})|$ varies slowly with distance. Significant deviations occur in the boundary "dark spaces", in particular [58, 59, 67, 68].

[21] A more general analysis requires numerical methods and yields deviations from the Chynoweth formula (see e.g. the calculations of Baraff [60]).

[22] In indirect semiconductors, W_i is larger than the energy gap W_g, since impact ionization of electrons with kinetic energies $W_\mathrm{kin} \approx W_\mathrm{g}$ is negligibly small owing to momentum conservation [64, 65].

[23] A modification of this elementary discussion which allows for multiple phonon collisions is presented in [62, 63], for example. Here the ionization coefficient is given by

$$\alpha = \frac{eE\exp\left(-\dfrac{W_\mathrm{i}}{eE\lambda_\mathrm{r}}\right)}{W_\mathrm{r}\left\{1+\dfrac{\lambda_\mathrm{i}}{\lambda_\mathrm{r}}\left[1-\exp\left(-\dfrac{W_\mathrm{r}}{eE\lambda}\right)\right]\right\}+\left(W_\mathrm{i}+\dfrac{\lambda_\mathrm{i}}{\lambda_\mathrm{r}}W_\mathrm{r}\right)\exp\left(-\dfrac{W_\mathrm{i}}{eE\lambda_\mathrm{r}}\right)},$$

where $\lambda^{-1} = \lambda_\mathrm{r}^{-1} + \lambda_\mathrm{i}^{-1}$ and W_r denotes the optical-phonon energy.

[24] Experimental investigations presented in [66] indicate, that the ionization coefficient is independent of crystal orientation.

Nonlocal Effects

Expressing the ionization coefficients as a function of local electric field strength is possible in the homogeneous-field case.[25] A comparably simple approach to the consideration of nonlocal effects was published in [73–75], where the ionization coefficient was related to the increase ΔW_n of the kinetic energy of the electrons, rather than to the local electric field strength. Since, in the homogeneous-field case, the excess kinetic energy due to carrier heating in the electric field is independent of position, the power $-ev_\mathrm{n}E$ delivered to an electron due to transport with velocity v_n in the field E equals the average power dissipation $(W_\mathrm{n} - W_\mathrm{n0})/\tau_\mathrm{wn}$ due to energy relaxation. Then $\Delta W_\mathrm{n} = -ev_\mathrm{n}\tau_\mathrm{wn}E$, where $\Delta W_\mathrm{n} = W_\mathrm{n} - W_\mathrm{n0}$, and τ_wn denotes the energy relaxation time (Sect. 2.5). Assuming ionization coefficients in Chynoweth form in the homogeneous-field case allows us to express these as functions of ΔW_n:

$$\alpha_\mathrm{n}(E) \;=\; a_\mathrm{n}\,\exp\left(-\frac{b_\mathrm{n}}{|E|}\right) \;=\; a_\mathrm{n}\,\exp\left(-\frac{eb_\mathrm{n}v_\mathrm{n}\tau_\mathrm{wn}}{\Delta W_\mathrm{n}}\right) . \tag{2.96}$$

Here v_n was taken to be positive, corresponding to a negative value of E. In the inhomogeneous-field case, ΔW_n has to be written as a function of position. If we set $J_\mathrm{n} = -env_\mathrm{n} \approx$ const. in the case of weak carrier multiplication, and $W_\mathrm{n} = 3k_\mathrm{B}T_\mathrm{n}/2$, the one-dimensional energy balance equation (2.70) becomes

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{5}{3}nv_\mathrm{n}W_\mathrm{n}\right) \;=\; -env_\mathrm{n}E - n\,\frac{\Delta W_\mathrm{n}}{\tau_\mathrm{wn}}$$

and transforms to

$$\frac{\mathrm{d}}{\mathrm{d}x}\,\Delta W_\mathrm{n} \;=\; -\frac{3}{5}\,eE - \frac{3}{5}\,\frac{\Delta W_\mathrm{n}}{v_\mathrm{n}\tau_\mathrm{wn}}$$

if W_n0 is assumed to be constant. The solution of this equation is given by

$$\begin{aligned}\Delta W_\mathrm{n}(x) \;&=\; \Delta W_\mathrm{n}(x_\mathrm{p})\exp\left(-\frac{3}{5}\int_{x_\mathrm{p}}^{x}\frac{\mathrm{d}\xi}{v_\mathrm{n}\tau_\mathrm{wn}}\right)\\ &\quad -\frac{3e}{5}\int_{x_\mathrm{p}}^{x}E(x')\exp\left(-\frac{3}{5}\int_{x'}^{x}\frac{\mathrm{d}\xi}{v_\mathrm{n}\tau_\mathrm{wn}}\right)\mathrm{d}x' . \end{aligned}\tag{2.97}$$

With the assumption of a constant energy relaxation length $\lambda_\mathrm{wn} = 5v_\mathrm{n}\tau_\mathrm{wn}/3$ and taking $\Delta W_\mathrm{n}(x_\mathrm{p}) = 0$, one obtains

$$\Delta W_\mathrm{n}(x) \;=\; -\frac{3e}{5}\int_{x_\mathrm{p}}^{x} E(x')\,\exp\left(-\frac{x-x'}{\lambda_\mathrm{wn}}\right)\mathrm{d}x' .$$

[25] In a generalization of the investigation in [69], where the asymptotic solution of the Boltzmann equation was found to determine the fraction of electrons that are able to undergo ionizing collisions, ionization coefficients as a function of a weighted average of the electric field strength were proposed in [70]. An alternative approach, based on history-dependent ionization coefficients and ionization probability densities, is presented in [71,72].

If this is substituted into in (2.96), ionization coefficients in the Chynoweth form are obtained:

$$\alpha_{\rm n} \;=\; a_{\rm n} \exp\left(-b_{\rm n}/|E_{\rm eff}(x)|\right) \;, \tag{2.98}$$

where the local electric field strength is replaced by an effective field strength

$$E_{\rm eff}(x) \;=\; \frac{1}{\lambda_{\rm wn}} \int_{x_{\rm p}}^{x} E(x') \exp\left(\frac{x-x'}{\lambda_{\rm wn}}\right) {\rm d}x' \;. \tag{2.99}$$

In [74, 76], $\lambda_{\rm wn} = 65\,{\rm nm}$ was found to accurately describe the measured data of the electron multiplication factor, while in [75] the value $\lambda_{\rm wn} = 80\,{\rm nm}$ was found to be appropriate. For holes $\lambda_{\rm wn} = 50\,{\rm nm}$ was found in [76].

Impact Ionization at Surfaces

If hot carriers reach an interface between silicon and SiO_2, surface impact ionization may occur. The surface impact ionization rate of electrons was determined in [77], where the ionization rate

$$\alpha_{\rm n}^{\rm (surface)} \;=\; \frac{2.45\cdot 10^6}{\rm cm} \exp\left(-\frac{1.92\,{\rm MV/cm}}{E}\right) \tag{2.100}$$

was found to describe the experimental findings.

2.6.4 Interband Tunneling

Direct tunneling of electrons from the valence to the conduction band is observed in heavily doped pn junctions. If both sides of the junction are degenerate, a forward-bias tunneling current of the kind known from the Esaki diode [78] occurs. In high-frequency bipolar transistors direct generally tunneling is observed in the reversed-biased eb diode, where it is responsible for the low values of the eb breakdown voltage $BV_{\rm EBO}$. Under forward-bias conditions direct tunneling is negligibly small.[26] Interband tunneling, or internal field emission, was first investigated by Zener and Wills [79] in order to explain electrical breakdown of insulators. Internal field emission was later shown to occur in certain pn junctions [80]. Increased interest arose with the invention of the Esaki diode, which triggered a number of theoretical investigations [81–85].

In order to observe a tunneling current, two prerequisites must be fulfilled:

1. The tunneling distance, that is, the extent of the forbidden zone through which the electrons tunnel, must not be more than a few nanometers. In a

[26] Tunneling via impurity states (trap-assisted tunneling) may, however, cause a significant increase of the base current at small forward bias. Such effects are generally considered within extensions of the Shockley Read Hall recombination model, as outlined in Sect. 2.6.1.

semiconductor, the band edges are tilted with a slope eE in the presence of an electric field. From the bandgap W_g, the tunneling distance can be

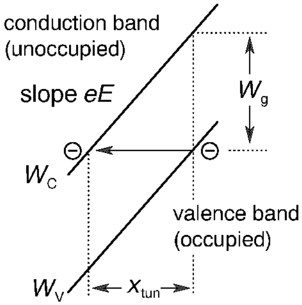


Fig. 2.11. Tunneling distance x_{tun}, defined in terms of the bandgap W_g and the electric field strength E

obtained by simple geometrical reasoning as $x_{\text{tun}} = W_g/eE$ (Fig. 2.11). If we take $x_{\text{tun}} \leq 5$ nm and $W_g \approx 1.1$ eV, the minimum value of the electric field strength necessary to observe tunneling can be estimated as $W_g/(ex_{\text{tun}}) \approx 0.2$ V/nm. Only in heavily doped pn diodes are such large values observed.

2. Since tunneling may occur in both directions, there must be more electrons on one side than on the other in order to observe a net current flow.

A detailed analysis of interband tunneling in silicon pn junctions has to consider electron–phonon interactions[27] and yields rather complicated expressions, even if the investigation is limited to the homogeneous-field case [83,88]. Real pn junctions are even more complicated, since they are heavily doped and therefore show substantial deviations in the density-of-states function; furthermore, the electric field strength varies with position. There exists a fairly simple approach based on Kane's theory [47, 83, 89] of the tunneling current in the homogeneous-field case, which introduces the generation rate

$$G_{\text{tun}}(\boldsymbol{x}) = B\,|\boldsymbol{E}(\boldsymbol{x})|^{\sigma}\,D(\psi, \phi_{\text{n}}, \phi_{\text{p}})\,\mathrm{e}^{-E_0/|\boldsymbol{E}(\boldsymbol{x})|}\,, \tag{2.101}$$

where $\sigma = 2$ for direct transitions and $\sigma = 5/2$ for indirect transitions, including electron phonon interactions.[28] The function $D(\psi, \phi_{\text{n}}, \phi_{\text{p}})$ takes account of the difference in occupation of the valence and conduction band states and may be approximated by [83]

$$D(\psi, \phi_{\text{n}}, \phi_{\text{p}}) = \frac{1}{1 + \exp\left[\,(\phi_{\text{n}} - \psi)/V_{\text{T}}\,\right]} - \frac{1}{1 + \exp\left[\,(\phi_{\text{p}} - \psi)/V_{\text{T}}\,\right]}\,.$$

[27] The importance of electron–phonon interactions is made obvious by measurements of the current–voltage characteristics of Esaki diodes at 4.2 K [86, 87], which show bending points characteristic of the onset of additional phonon-assisted tunneling processes.

[28] For practical purposes, owing to the dominance of the exponential term, the choice of σ is not really significant.

2.7 Heavily Doped Semiconductors

The emitter and base regions of modern high-frequency bipolar transistors use significant doping levels. The density of majority carriers then no longer obeys Boltzmann statistics and has to be described in terms of the Fermi distribution. Furthermore, the band structure is modified owing to the formation of impurity bands, the occurence of band tails in the density-of-states function, and a rigid shift of the band edges caused by self-energy effects [90, 91]. All these effects have to be considered in order to arrive at an accurate description of minority-carrier transport in heavily doped regions.

2.7.1 Modification of the Band Structure

The density-of-states function of an intrinsic semiconductor is zero in the energy gap (Fig. 2.12a). Incorporation of a single donor[29] atom into the lattice of the semiconductor results in a localized state within the energy gap, which may be occupied by an electron. The density-of-states function then shows an additional "peak" in the energy gap (Fig. 2.12b). The "extent" of the localized state can be estimated by considering the Bohr radius $a_0 = 4\pi\epsilon\hbar^2/m_n^* e^2$, which is of the order of 10 nm owing to the large value of the permittivity in a semiconductor. Overlap of the wave functions of adjacent donor atoms may be neglected only if the separation of the impurities is large in comparison with a_0, i.e. if the donor concentration N_D obeys the relation $N_D^{-1/3} \gg a_0$, or $N_D \ll 1/a_0^3$. If $a_0 \approx 10$ nm, this corresponds to $N_D \ll 10^{18}\ \mathrm{cm}^{-3}$. This condition is not fulfilled in the heavily doped emitter and base regions of today's high-frequency bipolar transistors; the picture of localized impurity states is no longer valid in this case.

If the impurity states become closer to one another, their wave functions overlap and electrons may be interchanged between neighboring impurity states, resulting in the formation of an impurity band. Since the electrostatic field of an additional donor impurity is not completely screened, an electron caught in the field of a donor ion will be affected by the field of the surrounding impurities, resulting in a shift of the energy level. Since this effect varies with the average distance between impurities, the separation of the energy level from the conduction band edge is affected by the impurity concentration N according to [90]

$$W_C - W_D(N_D) \approx W_C - W_D(0) - \alpha N_D^{1/3} ,$$

where $\alpha \approx 3.1 \times 10^{-8}$ eV/cm. Owing to the random distribution of the neighboring impurities, a statistical uncertainty arises, which causes a broadening

[29] The effects are described here for n-type material; only one donor state is assumed, for simplicity.

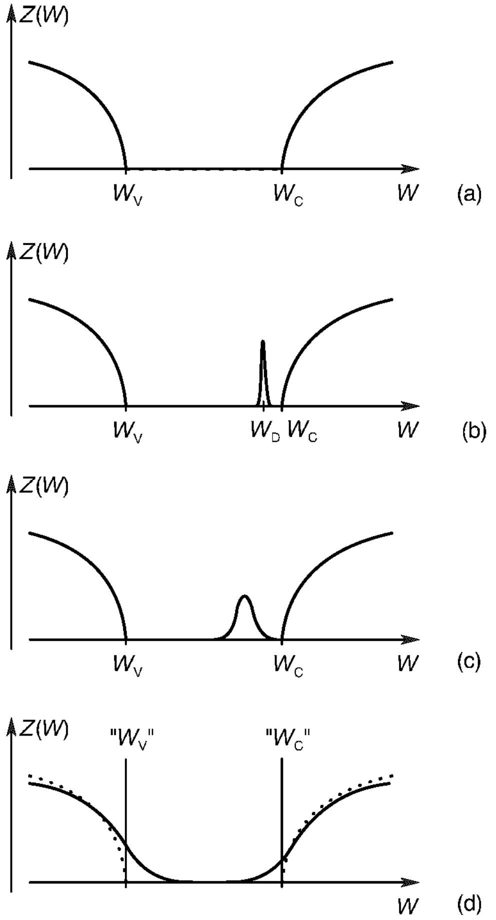


Fig. 2.12. Density of states (schematic representation). (**a**) undoped semiconductor, (**b**) weakly doped semiconductor, (**c**) moderately doped semiconductor, with the beginning of formation of impurity bands, and (**d**) heavily doped semiconductor with band tails

of the energy levels: the density of states in the impurity band is approximately Gaussian [92].

With increasing donor concentration, the impurity band broadens and finally merges with the conduction band;[30] in this case no activation energy is required for the donor electrons to move to the conduction band. In such heavily doped semiconductors it is no longer possible to distinguish between localized impurity states and conduction band states.

[30]Theoretical results for the density of states in phosphorus-doped silicon for doping levels in the range $10^{16}\ \mathrm{cm}^{-3} \leq N_\mathrm{D} \leq 3 \times 10^{20}\ \mathrm{cm}^{-3}$ have been published in [93]

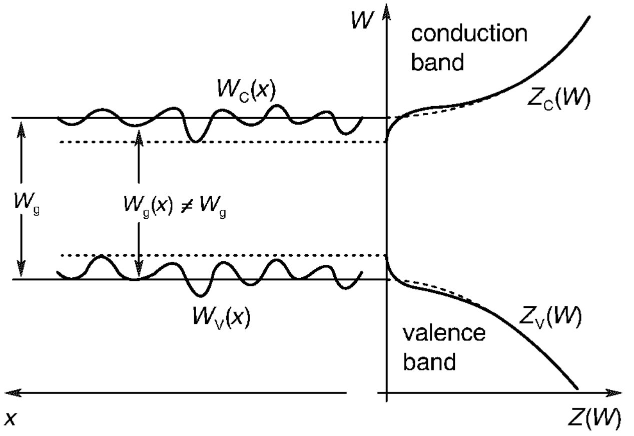


Fig. 2.13. Formation of band tails due to fluctuations of local electrostatic potential (after [94])

The random[31] distribution of dopants results in a spatially random electric field exerted on the electrons in the conduction band, which is particularly attractive in the vicinity of clusters of donor atoms. This leads to the formation of local potential wells in the conduction band (Fig. 2.13). The density-of-states function then shows band tails, as illustrated in Fig. 2.12d. Such band tails appear in both bands and are identical up to a constant factor, as shown in the theoretical analysis of Kane [95].

2.7.2 Bandgap Narrowing in Silicon

The bandgap shrinkage in doped silicon crystals was investigated in [96] and, in an improved form, in [97] using the so-called "random-phase approximation" (see also [94]). Simple expressions for bandgap narrowing in n- and p-type silicon (as well as other semiconductors) were derived in [98], with the results

$$\frac{\Delta W_{\mathrm{g}}}{\mathrm{meV}} = 10.23\left(\frac{N}{10^{18}}\right)^{1/3} + 13.12\left(\frac{N}{10^{18}}\right)^{1/4} + 2.93\left(\frac{N}{10^{18}}\right)^{1/2} \tag{2.102}$$

for n-type silicon and

$$\frac{\Delta W_{\mathrm{g}}}{\mathrm{meV}} = 11.07\left(\frac{N}{10^{18}}\right)^{1/3} + 15.17\left(\frac{N}{10^{18}}\right)^{1/4} + 5.07\left(\frac{N}{10^{18}}\right)^{1/2} \tag{2.103}$$

for p-type silicon. The authors of [98] represent the bandgap narrowing as the superposition of three different effects, which are added. The first term represents the exchange interactions of the majority-carrier densities, which

[31] The dopants are statistically distributed on lattice sites and do not show the regular distribution expected for thermal equilibrium, which minimizes the free energy of the system. The redistribution of dopants at room temperature towards the equilibrium distribution occurs so slowly, however, that it may be neglected for practical purposes.

vary in inverse proportion with the average distance r between carriers and hence in proportion to $N^{1/3}$. The second term represents the effect of electron–hole interaction on bandgap narrowing, which varies in proportion to $1/r^{3/4} \sim N^{1/4}$. The third term takes account of the shift of the majority band due to carrier–impurity interaction, which varies in proportion to $1/r^{3/2} \sim N^{1/2}$ [90].

Electrical Measurements. The early studies of bandgap narrowing in heavily doped silicon were performed because this phenomenon was found to influence the current–voltage characteristics of tunnel diodes (Esaki diodes). In bipolar transistors, bandgap narrowing was considered in 1968 by Kauffman and Bergh [99] to explain the discrepancy between measured and calculated current gains[32]. In 1976, bandgap narrowing of p-type silicon was investigated by Slotboom and de Graaff [100] using npn bipolar transistors with different doping levels in the base region. These authors assumed recombination in the base region to be negligible and determined the product $\mu_n n_{iB}^2$ from the measured transfer saturation current I_S with the help of the Moll–Ross relation (Sect. 3.2). From the resulting data, a fitting formula of the form

$$\Delta W_g^{app}(N) = W_1 \left[\ln(N/N_0) + \sqrt{\ln^2(N/N_0) + 0.5} \right] \tag{2.104}$$

for the reduction of the apparent bandgap,[33] which differs from the actual value $W_g(T)$ owing to Fermi–Dirac statistics, was obtained. Similar investigations were performed by other authors, leading to similar results. All these techniques for the extraction of the intrinsic carrier density rely on a knowledge of the minority-carrier mobility and the majority-carrier concentration, and are therefore subject to uncertainty. Data consistent with the mobility model presented in [102] were presented in [103].

Optical Measurements. Several optical measurement techniques have been employed for the characterization of heavily doped semiconductors. In absorption measurements [104–106], the transmission of light through a sample is recorded as a function of the wavelength λ. The absorption coefficient shows

[32] Taking the base transport factor as unity and under the assumption of ideal behavior of the base current, these authors conclude that the base current I_B should vary in proportion to $n_{iE}^2 \sim \exp(-W_{gE}/k_BT)$, where n_{iE} denotes the intrinsic carrier concentration in the emitter region, while the transfer current $I_T \approx I_C$ should be proportional to n_{iB}^2, where $n_{iB} \sim \exp(-W_{gB}/k_BT)$ is the intrinsic carrier concentration in the base region. The current gain $B_N = I_C/I_B$ should therefore be proportional to $\exp(-\Delta W_g/k_BT)$, where $\Delta W_g = W_{gB} - W_{gE}$ is the reduction of the bandgap W_{gE} in the emitter region relative to the bandgap W_{gB} in the base region.

[33] This term was introduced later, and refers to the fact that $W_g^{app} = k_BT \ln(N_C N_V/n_{ie}^2)$ is the value of the bandgap that yields the correct value of the intrinsic carrier density if Boltzmann statistics are applied. This is, however, not correct in heavily doped semiconductors with a degenerate majority-carrier distribution (see e.g. [101] and references cited therein).

a substantial increase if the incident photons are able to excite electrons from the valence to the conduction band, that is if the energy of the incident photons exceeds the value $W_C - W_V$ in the case of a nondegenerate semiconductor (Fig. 2.14). If an electron is to be excited from the valence to the conduction band, it requires an unoccupied state in the conduction band, which can only be found above the Fermi level. In degenerate semiconductors, with a Fermi level shifted into the conduction band, the absorption edge therefore yields the difference $W_F - W_V$ rather than $W_C - W_V$, and therefore an apparently larger value of the bandgap, i.e. the bandgap narrowing appears to be smaller. Degeneracy thus shifts the optical absorption edge to higher photon energies, an effect known as the Burstein shift [107].

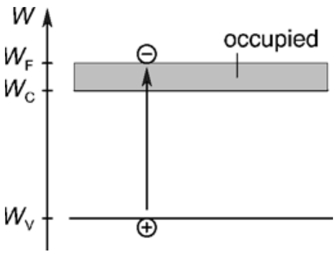


Fig. 2.14. Illustration of the difference between bandgap data determined from electrical and optical measurements

In photoluminescence measurements, [106, 108] the spectral distribution of the photons emitted during a recombination process is recorded. The energy of the emitted photon will be in the interval between $W_C - W_V$ and $W_F - W_V$, if no energy is dissipated or absorbed as a result of phonon emission or absorption. Taking these details into account, optical-aborption and photoluminescence data have been found to be fully consistent with each other [106, 107].

Optical measurements generally result in smaller values of bandgap narrowing, as compared with electrical measurements of the pn product. This is due to degeneracy, which moves the Fermi energy into the conduction band in the case of a heavily doped n-type semiconductor. The electrically and optically measured data for bandgap narrowing are found to be consistent if one takes account of the fact that electrical measurements derived from transport equations in the drift–diffusion approximation always yield the apparent bandgap narrowing

$$\Delta W_{\mathrm{g}}^{\mathrm{app}} = \Delta W_{\mathrm{g}} - k_{\mathrm{B}}T \ln\left[1 + 0.27 \times \exp\left(\frac{W_{\mathrm{F}} - W_{\mathrm{C}}}{k_{\mathrm{B}}T}\right)\right] ,$$

which differs from the real bandgap narrowing ΔW_{g} because effects due to degeneracy are included in the latter parameter [109].

2.8 Silicon Device Modeling in the Drift–Diffusion Approximation

The drift–diffusion approximation is still the most widely employed approach to the modeling and simulation of semiconductor devices.

2.8.1 Basic Equations of the Drift–Diffusion Approximation

Within the drift–diffusion approximation, the continuity equations for both electrons and holes,

$$\partial n/\partial t = G - R + \nabla\cdot\boldsymbol{J}_{\mathrm{n}}/e\,, \tag{2.105}$$

$$\partial p/\partial t = G - R - \nabla\cdot\boldsymbol{J}_{\mathrm{p}}/e\,, \tag{2.106}$$

the current (density) equations,

$$\boldsymbol{J}_{\mathrm{n}} = -e\mu_{\mathrm{n}}n\,\nabla\phi_{\mathrm{n}}\,, \tag{2.107}$$

$$\boldsymbol{J}_{\mathrm{p}} = -e\mu_{\mathrm{p}}p\,\nabla\phi_{\mathrm{p}}\,, \tag{2.108}$$

and the Poisson equation,

$$\nabla^2\psi = -e\,(N_{\mathrm{D}}^{+}-N_{\mathrm{A}}^{-}+p-n)/\epsilon\,, \tag{2.109}$$

together with the definitions

$$\phi_{\mathrm{n}} = \psi - V_{\mathrm{T}}\,\ln(n/n_{\mathrm{ie}})\,, \tag{2.110}$$

$$\phi_{\mathrm{p}} = \psi + V_{\mathrm{T}}\,\ln(p/n_{\mathrm{ie}}) \tag{2.111}$$

are used[34] for the mathematical description of semiconductor devices.

The mobilities μ_{n}, μ_{p} and the net recombination rate $R - G$ have to be modeled as functions of temperature, carrier concentration and electric field strength. In addition to this, the intrinsic carrier density n_{ie} has to be described as a function of both temperature and local doping concentration to take account of rigid bandgap shrinkage (caused by many-body effects), band tails and impurity bands (due to changes of the density-of-states function caused by the presence of impurities), and degenerate statistics [21, 111, 112]. From these equations, the terminal characteristics of a device are obtained by solution of the corresponding boundary value problem.[35]

[34] This set of equations was compiled (in a somewhat different formulation) by van Roosbroeck [110] and is therefore sometimes termed the Roosbroeck equations. The generation and conduction of heat is not considered in these equations, an extended description that takes account of these effects is presented in [113], for example.

[35] The numerical solution of two-dimensional boundary value problems is explained e.g. in [114].

The electron and hole current equations (2.107) and (2.108) relate the current densities to the gradient of the corresponding quasi-Fermi potentials. This representation is suitable if majority-carrier transport is being considered, since the prefactors will be constant and determined by the doping concentration. With the help of (2.110) and (2.111), the electron and hole current equations (2.107) and (2.108) may be expressed in the alternative form

$$\nabla \exp\left(-\frac{\phi_\mathrm{n}}{V_\mathrm{T}}\right) = \frac{1}{k_\mathrm{B} T \mu_\mathrm{n} n_\mathrm{ie}^2}\, p \exp\left(-\frac{\phi_\mathrm{p}}{V_\mathrm{T}}\right) \boldsymbol{J}_\mathrm{n} \tag{2.112}$$

and

$$\nabla \exp\left(\frac{\phi_\mathrm{p}}{V_\mathrm{T}}\right) = -\frac{1}{k_\mathrm{B} T \mu_\mathrm{p} n_\mathrm{ie}^2}\, n \exp\left(\frac{\phi_\mathrm{n}}{V_\mathrm{T}}\right) \boldsymbol{J}_\mathrm{p}\,. \tag{2.113}$$

In these expressions, the prefactor in the electron current density equation is determined by the distribution of holes, while the prefactor in the hole current density equation is determined by the distribution of electrons. The current equations written in this form are therefore particularly suitable for the description of minority-carrier transport under low-level-injection conditions, where the prefactors can be assumed to be constant.

Boundary Conditions. If Maxwell's equations are solved for a given volume, the following boundary conditions at the surface must be considered:

$$\boldsymbol{n} \times (\boldsymbol{E}_2 - \boldsymbol{E}_1) = 0 \quad \text{and} \quad \boldsymbol{n} \cdot (\boldsymbol{D}_2 - \boldsymbol{D}_1) = Q_\mathrm{s}'\,, \tag{2.114}$$

where $\boldsymbol{E}_1$ and $\boldsymbol{D}_1$ are the inner values of the electric field strength and dielectric displacement vector, $\boldsymbol{E}_2$ and $\boldsymbol{D}_2$ are the outer values and $\boldsymbol{n}$ denotes the (outward-oriented) unit vector normal to the interface. The component of the electric field parallel to the boundary thus remains constant, whereas the component of the dielectric displacement orthogonal to the boundary changes by the value of the surface charge density, Q_s', at the boundary. This gives the following boundary condition for the electric field strength at a semiconductor–insulator interface:

$$\epsilon_\mathrm{s}\, \boldsymbol{n} \cdot \boldsymbol{E}_\mathrm{s} = \epsilon_\mathrm{i}\, \boldsymbol{n} \cdot \boldsymbol{E}_\mathrm{i} - Q_\mathrm{s}'\,. \tag{2.115}$$

Here the subscripts s and i are used to denote "semiconductor" and "insulator". If the electric field in the insulator is small, as is usually the case at the boundary of a field oxide region, (2.115) simplifies to $\epsilon_\mathrm{s}\, \boldsymbol{n} \cdot \boldsymbol{E}_\mathrm{s} = -Q_\mathrm{s}'$.

Insulating surfaces are characterized by a small surface recombination velocity, i.e. the normal components of both the majority and the minority current densities vanish to a good approximation: $\boldsymbol{n} \cdot \boldsymbol{J}_\mathrm{n} \approx \boldsymbol{n} \cdot \boldsymbol{J}_\mathrm{p} \approx 0$. Majority carrier flow through an ohmic contact is hindered only by the presence of a possibly current-dependent contact resistance, specified by the specific contact resistance ρ_{c}. This contact resistance causes a difference between the majority-carrier quasi-Fermi potentials on either side of the contact:

$$\begin{aligned}\Delta\phi_{\mathrm{n}} &= \rho_{\mathrm{c}}\,\boldsymbol{n}\cdot\boldsymbol{J}_{\mathrm{n}} \quad \text{for ohmic contacts to n-type material}\,,\\ \Delta\phi_{\mathrm{p}} &= \rho_{\mathrm{c}}\,\boldsymbol{n}\cdot\boldsymbol{J}_{\mathrm{p}} \quad \text{for ohmic contacts to p-type material}\,.\end{aligned}$$

Generally, contacts that are formed to heavily doped semiconductor regions have a very small dielectric relaxation time – this allows one to assume that the semiconductor is neutral in the vicinity of the contact. The assumption of thermal equilibrium at the contact ($np = n_{\mathrm{i}}^2$) is justified only in the case of metal contacts, which act as efficient recombination centers. Together with the assumption of neutrality, the thermal-equilibrium condition yields the majority- and minority-carrier concentrations at metal contacts. The investigation of shallow polysilicon-contacted emitters in modern bipolar transistors is not possible with this assumption. For such contacts, the boundary conditions are formulated in terms of (effective) surface recombination velocities

$$\Delta n = -S_{\mathrm{n}}\,\boldsymbol{n}\cdot\boldsymbol{J}_{\mathrm{n}} \quad \text{for polysilicon contacts to p-type material}\,,$$

and

$$\Delta p = S_{\mathrm{p}}\,\boldsymbol{n}\cdot\boldsymbol{J}_{\mathrm{p}} \quad \text{for polysilicon contact to n-type material}\,,$$

which relate the excess minority-carrier density at the contact to the minority current orthogonal to the surface.

2.8.2 Model Equations for Material Parameters

Reliable simulation of semiconductor device behavior requires models that correctly describe the dependence of material parameters such as the bandgap, mobilities and minority-carrier lifetimes on doping concentration, electric field strength and temperature.

Bandgap Narrowing, Intrinsic Carrier Density

In the absence of bandgap narrowing, the intrinsic carrier density is usually described by

$$n_{\mathrm{i}}(T) = \sqrt{N_{\mathrm{C}}(T)N_{\mathrm{V}}(T)}\,\exp\left[-\frac{W_{\mathrm{g}}(T)}{2k_{\mathrm{B}}T}\right]\,, \tag{2.116}$$

with temperature-dependent values for the effective densities of states $N_{\mathrm{C}}(T)$ and $N_{\mathrm{V}}(T)$ and for the bandgap. Bandgap narrowing due to heavy doping is taken into account by introducing the effective intrinsic carrier density [14]

$$n_{\mathrm{ie}}(T) = n_{\mathrm{i}}(T)\,\exp\left[\frac{\Delta W_{\mathrm{g}}^{\mathrm{app}}(T)}{2k_{\mathrm{B}}T}\right]\,. \tag{2.117}$$

The apparent bandgap narrowing $\varDelta W_{\mathrm{g}}^{\mathrm{app}}(T)$ in this expression takes account of both changes in the density-of-states function and effects due to degeneracy. Measurement results published by several authors can be represented by

$$\varDelta W_{\mathrm{g}}^{\mathrm{app}} = 6.92\,\mathrm{meV} \times \left[\ln(N/N_0) + \sqrt{\ln^2(N/N_0) + 0.5}\right] , \qquad (2.118)$$

where $N_0 = 1.3 \times 10^{17}\ \mathrm{cm}^{-3}$, [102]. In the limit $N \gg N_0$, this results in a logarithmic increase of the apparent bandgap narrowing:

$$\varDelta W_{\mathrm{g}}^{\mathrm{app}} \approx 13.84\,\mathrm{meV} \times \ln\left(\frac{N}{1.3 \times 10^{17}\ \mathrm{cm}^{-3}}\right) . \qquad (2.119)$$

The effective intrinsic carrier density then increases with dopant density according to the power law

$$n_{\mathrm{ie}}(T) \approx n_{\mathrm{i}}(T) \left(\frac{N}{1.3 \times 10^{17}\ \mathrm{cm}^{-3}}\right)^{T_0/T} , \qquad (2.120)$$

where $T_0 \approx 13.84\,\mathrm{meV}/2k_{\mathrm{B}} \approx 80\,\mathrm{K}$.

Mobility

Owing to ionized-impurity scattering, the mobility is a function of dopant density. For majority carriers, according to [115], the experimental mobility data at $T = 300\,\mathrm{K}$ and low values of electric field strength are described well by the relation[36]

$$\mu_{\mathrm{L0}} = \mu_{\mathrm{min}} + \frac{\mu_{\mathrm{max}} - \mu_{\mathrm{min}}}{1 + (N/N_{\mathrm{r1}})^{\alpha_1}} - \frac{\mu_1}{1 + (N_{\mathrm{r2}}/N)^{\alpha_2}} , \qquad (2.121)$$

where N denotes the total impurity concentration, and μ_{min}, μ_{max}, μ_1, N_{r1}, α_1, N_{r2} and α_2 are parameters listed in Table 2.2.

Table 2.2. Parameters for majority-carrier mobility[a] [115]

dopant	μ_{min}	μ_{max}	μ_1	N_{r1}	α_1	N_{r2}	α_2
As	52.2	1417	43.4	9.68×10^{16}	0.68	3.43×10^{20}	2
P	68.5	1414	56.1	9.2×10^{16}	0.711	3.41×10^{20}	1.98
B	44.9	470.5	29.0	2.23×10^{17}	0.719	6.1×10^{20}	2

[a]) mobilities in $\mathrm{cm}^2/(\mathrm{Vs})$, doping concentrations in cm^{-3}.

The mobilities are temperature dependent owing to phonon scattering; the temperature dependence generally fits well to a power-law expression of the form

[36]This formula extends the fitting formula of Caughey and Thomas [116], which does not contain the third term on the right-hand side of (2.121).

$$\mu_{\mathrm{L},0}(T) = \mu_{\mathrm{L},0}\left(\frac{T}{300\ \mathrm{K}}\right)^{-\alpha} = \mu_{\mathrm{L},0}\left(\frac{300\ \mathrm{K}}{T}\right)^{\alpha} . \tag{2.122}$$

A refined mobility model for the low-field mobility in bulk silicon that is claimed to hold form 300 K to 700 K has been presented in [117]. There the low-field bulk mobility is approximated as

$$\begin{aligned}\mu_{\mathrm{L},0} = \mu_0 &+ \frac{\mu_{\mathrm{L}} - \mu_0}{1 + (N_{\mathrm{D}}/C_{\mathrm{r1}})^{\alpha_1} + (N_{\mathrm{A}}/C_{\mathrm{r2}})^{\alpha_2}} \\ &- \frac{\mu_1}{1 + (N_{\mathrm{D}}/C_{\mathrm{s1}} + N_{\mathrm{A}}/C_{\mathrm{s2}})^{-2}}\end{aligned} \tag{2.123}$$

with

$$\mu_0 = \frac{\mu_{0\mathrm{d}} N_{\mathrm{D}} + \mu_{0\mathrm{a}} N_{\mathrm{A}}}{N_{\mathrm{D}} + N_{\mathrm{A}}} , \quad \mu_1 = \frac{\mu_{1\mathrm{d}} N_{\mathrm{D}} + \mu_{1\mathrm{a}} N_{\mathrm{A}}}{N_{\mathrm{D}} + N_{\mathrm{A}}} ,$$

$$\mu_{\mathrm{L}} = \mu_{\max}\left(\frac{T}{300\ \mathrm{K}}\right)^{-\gamma + c(T/300\ \mathrm{K})}$$

and parameters as defined in Table 2.3.

Table 2.3. Parameters for silicon bulk mobility model [117]

Parameter	Phosphorus	Arsenic	Boron
$\mu_{\max}$ ($\mathrm{cm}^2/\mathrm{Vs}$)	1441	1441	470.5
c	0.07	0.07	0.0
γ	2.45	2.45	2.16
$\mu_{0\mathrm{d}}$ ($\mathrm{cm}^2/\mathrm{Vs}$)	$62.2 \times T^{-0.7}$	$55 \times T^{-0.6}$	$90 \times T^{-1.3}$
$\mu_{0\mathrm{a}}$ ($\mathrm{cm}^2/\mathrm{Vs}$)	$132 \times T^{-1.3}$	$132 \times T^{-1.3}$	$44 \times T^{-0.7}$
$\mu_{1\mathrm{d}}$ ($\mathrm{cm}^2/\mathrm{Vs}$)	$48.6 \times T^{-0.7}$	$42.4 \times T^{-0.5}$	$28.2 \times T^{-2.0}$
$\mu_{1\mathrm{a}}$ ($\mathrm{cm}^2/\mathrm{Vs}$)	$73.5 \times T^{-1.25}$	$73.5 \times T^{-1.25}$	$28.2 \times T^{-0.8}$
C_{r1} (cm^{-3})	$8.5 \times 10^{16}\, T^{3.65}$	$8.9 \times 10^{16}\, T^{3.65}$	$1.3 \times 10^{18}\, T^{2.2}$
C_{r2} (cm^{-3})	$1.22 \times 10^{17}\, T^{2.65}$	$1.22 \times 10^{17}\, T^{2.65}$	$2.45 \times 10^{17}\, T^{3.1}$
C_{s1} (cm^{-3})	4×10^{20}	2.9×10^{20}	$1.1 \times 10^{18}\, T^{6.2}$
C_{s2} (cm^{-3})	7×10^{20}	7×10^{20}	6.1×10^{20}
α_1	0.68	0.68	0.77
α_2	0.72	0.72	0.719

In the presence of large electron and hole densities, electron–hole scattering may become relevant. According to Dorkel and Leturq [129], the mobility due to electron–hole scattering is given by

$$\mu_{\mathrm{CC}} = \frac{2 \times 10^7\, T^{3/2}}{\sqrt{pn}} \frac{1}{\ln\left[1 + 8.28 \times 10^8 \dfrac{T^2}{(pn)^{1/3}}\right]} . \tag{2.124}$$

The low-field mobility in silicon in the presence of electron–hole scattering can be obtained from Matthiesen's rule as follows:

$$\frac{1}{\mu_\mathrm{L}} = \frac{1}{\mu_\mathrm{L0}(T)} + \frac{1}{\mu_\mathrm{CC}} \,. \tag{2.125}$$

This assumes that all scattering mechanisms are independent; experimental confirmation of the model is given in [130, 131], for example. Limitations of this approach were noted in [132].

The experiments of Ryder and Shockley [118, 119] demonstrated for the first time the effects of high electric field strength on the electron mobility in homogeneous semiconductors. For large values of electric field strength, a decrease of the mobility with increasing field strength was shown, resulting in a limit for the drift velocity of carriers, which was found to be around 10^7 cm/s. The reason for this nonohmic behavior is an increase of the scattering probability with particle energy due to optical-phonon production [120].[37] The field-dependent mobility can be derived from the low-field mobility μ_L according to the empirical formula of Caughey and Thomas [116]

$$\mu(E) = \frac{\mu_\mathrm{L}}{\left[1 + (\mu_\mathrm{L} E / v_\mathrm{nsat})^\beta\right]^{1/\beta}} \,. \tag{2.126}$$

The saturation drift velocity v_nsat depends on the crystal orientation [122, 123]; the effect is, however, not very pronounced at room temperature. According to [123], the temperature dependence of the saturation drift velocity is described by

$$v_\mathrm{nsat}(T) = \frac{v_\mathrm{n}^*}{1 + C \exp(T/\Theta)} \,, \tag{2.127}$$

where $v_\mathrm{n}^* = 2.4 \times 10^7$ cm/s, $C = 0.8$ and $\Theta = 600$ K.

In heavily doped semiconductors, where mobilities are determined by ionized-impurity scattering to a large extent, minority-carrier mobilities differ from majority-carrier mobilities, since different scattering cross sections are obtained for scattering at attractive and repulsive Coulomb wells. According to a theoretical analysis of Bennett [124], minority-carrier mobilities may exceed majority-carrier mobilities at the same dopant concentration by a factor of three. According to [125, 126], the experimental data for the mobility of holes in n-type silicon are described by

$$\mu_\mathrm{L0} = \mu_\mathrm{min2} + \frac{\mu_2}{1 + (N/N_\mathrm{ref3})^{\alpha_3}} \,, \tag{2.128}$$

where the values $\mu_\mathrm{min2} = 132.54$ cm^2/(Vs), $\mu_2 = 369.73$ cm^2/(Vs), $N_\mathrm{ref3} = 5.96 \times 10^{16}$ cm^{-3} and $\alpha_3 = 0.49$ were determined in [127]; for large impurity concentrations, the minority-carrier mobility exceeds the majority-carrier mobility.

[37] See, for example, [121] for a simple semiempirical discussion.

Table 2.4. Temperature exponent for minority hole mobility in silicon [127]

N_D/cm^{-3}	1.48×10^{18}	3.05×10^{18}	7.13×10^{18}	2.4×10^{19}
m	0.39	0.86	1.19	1.33

The minority hole mobility in silicon was found to be strongly dependent on temperature, especially for the more highly doped devices. In [128], a temperature dependence of the form

$$\mu_p(T) \sim T^{-m} \tag{2.129}$$

was assumed. Fitting experimental data to this formula in the temperature range from 100 K to 295 K gave a monotonic increase of the temperature exponent m with the doping concentration (Table 2.4).

Strain effects due to elastic deformations of the crystal lattice are not considered in most device simulation programs, although substantial mechanical stress is to be expected in the monocrystalline silicon in trench-isolated integrated circuits with high packing density and in heavily doped semiconductor regions. Changes in the energy spectrum due to elastic strain can, in particular, affect the carrier density and scattering rate and modify the effective mass, which in turn affects the mobility [133]. An investigation of strain effects on terminal currents using a modified device simulator was published in [133]; the voltage drop required to yield a certain input or transfer current was found to be reduced by approximately 10 mV, corresponding to an increase of the current of the order of 50% at fixed voltage.

Generation–Recombination Modeling

The net recombination rate is generally modeled as a sum of the net recombination rates due to Shockley–Read–Hall recombination, Auger recombination and direct recombination:

$$R-G = (R-G)_{\mathrm{SRH}} + (R-G)_{\mathrm{Aug}} + (R-G)_{\mathrm{dir}} . \tag{2.130}$$

The Shockley–Read–Hall recombination rate $(R-G)_{\mathrm{SRH}}$ is determined by (2.84):

$$R-G = \frac{pn - n_{\mathrm{ie}}^2}{\tau_{\mathrm{pSRH}}(n+n_1) + \tau_{\mathrm{nSRH}}(n+n_1)} ; \tag{2.131}$$

the lifetimes $\tau_{\mathrm{nSRH}} = \tau_{\mathrm{n0}}/N_T$ and $\tau_{\mathrm{pSRH}} = \tau_{\mathrm{p0}}/N_T$ vary in inverse proportion to the density N_T of trapping centers. Typical values of these parameters are in the region of several microseconds.

The Auger recombination rate for minority carriers varies in proportion with the square of the majority carrier density

$$(R-G)_{\mathrm{Aug}} = (pn-n_{\mathrm{ie}}^2)(C_{\mathrm{n}}n+C_{\mathrm{p}}p) \;, \tag{2.132}$$

where C_{n} and C_{p} are of the order[38] of 10^{-31} cm^6/s .

The rate of direct recombination, associated with photon emission, is small in semiconductors with an indirect bandgap, such as silicon, and can generally be neglected. In direct semiconductors this contribution to the total recombination rate can be written as [114]

$$(R-G)_{\mathrm{dir}} = C_{\mathrm{C}}^{\mathrm{opt}}(pn-n_{\mathrm{ie}}^2) \;. \tag{2.133}$$

In heavily doped pn junctions, where tunneling may be important, an additional generation recombination term $(R–G)_{\mathrm{bbt}}$ has to be added, which takes account of band-to-band tunneling. In addition to this, the lifetimes τ_{n} and τ_{p} in $(R-G)_{\mathrm{SRH}}$ have to be modeled as field-dependent quantities [47].

In heavily doped n-type silicon, the recombination is predominantly determined by Shockley–Read–Hall and Auger processes and can therefore be approximately described by the net recombination rate

$$R-G \approx \left(C_{\mathrm{n}}n + \frac{1}{\tau_{\mathrm{pSRH}}n}\right)(pn-n_{\mathrm{ie}}^2) \;,$$

since $n \gg p$. Introducing the (effective) hole lifetime τ_{p} according to [134], where

$$\tau_{\mathrm{p}} \approx \frac{\tau_{\mathrm{pSRH}}}{1 + C_{\mathrm{n}}\tau_{\mathrm{pSRH}}n^2} \;, \tag{2.134}$$

and taking account of the fact that $p_{\mathrm{n0}} = n_{\mathrm{ie}}^2/n$ under low-level-injection conditions, the net recombination rate is obtained in the form

$$R-G = \frac{p_{\mathrm{n}} - p_{\mathrm{n0}}}{\tau_{\mathrm{p}}} \;.$$

This form is generally applied in the investigation of minority-carrier transport. Owing to Auger processes, the value of τ_{p} will decrease with the dopant density $N_{\mathrm{D}} \approx n$ in proportion to $1/N_{\mathrm{D}}^2$.

2.8.3 Compact Modeling

Numerical solutions of the semiconductor equations in the drift–diffusion approximation based on Gummel's iterative scheme [135–137], were first applied to one-dimensional transistor problems. Two-dimensional simulations followed soon after [114, 138–143]. Device simulation by numerical solution of the semiconductor equations in three dimensions (e.g. [75, 144]) is extremely time-consuming and is therefore only suitable for the study of special problems in device optimization.

[38] In [134] $C_{\mathrm{n}} = C_{\mathrm{p}} = 1.5 \times 10^{-31}$ cm^6/s is suggested.

The numerical solution of the semiconductor equations in the drift–diffusion approximation implies a considerable numerical effort and can therefore only be applied to the study of the behavior of a very small number of devices. Computer-aided design (CAD) of integrated circuits that comprise a large number of devices requires a simplified description in terms of compact models, which provide analytical expressions for the terminal currents.

CAD of extended integrated circuits is possible if closed-form analytical expressions which employ only a comparatively small number of device parameters are used for an approximate description of device behavior. The goal of compact modeling is to derive a representation of device behavior with, preferably, a small number of parameters that is accurate for all modes of operation. For modeling purposes, integral relations (Fig. 2.15) that take account of the boundary conditions but employ only a subset of the basic equations have proven useful. These relations reduce the description of device behavior to a set of ordinary differential equations, since integration removes the space variables.

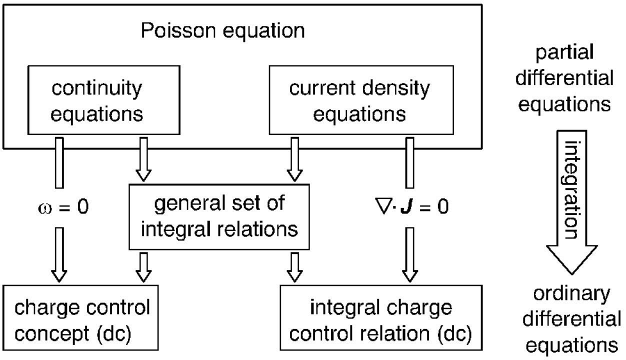


Fig. 2.15. Derivation of integral relations from the basic semiconductor equations. This procedure eliminates the space variables and yields a coupled set of ordinary differential equations

Though rigorously based on the semiconductor equations, these relations do not provide a complete solution of the specific boundary value problem, since they rely on the particular subset of device equations chosen. For npn transistors, two integral relations in particular are of interest: (1) the integral relation derived from the hole continuity equation that underlies the charge control theory (Sects. 3.1 and 3.5), and (2) the integral relation derived from the electron current equation – the integral charge control relation – that underlies the Gummel–Poon model (Sect. 3.2).

Even the most advanced compact models are derived from the semiconductor equations in the drift–diffusion approximation, which have reached their limits of validity in modern high-frequency bipolar transistors. The errors that arise from the drift–diffusion approximation and the approximations required to describe a device in terms of a small number of lumped elements make it difficult to compute the parameters of the equivalent-circuit elements from first principles. The approach that is usually adopted is to define a physics-based model founded on computations performed with the drift–diffusion theory, and to determine the exact parameters of the model from measurements. This procedure ensures a good description of the electrical behavior of a transistor despite the limited validity of the underlying theory.

2.9 References

[1] J.D. Jackson. *Classical Electrodynamics.* Wiley, New York, 2nd edition, 1975.

[2] J.M. Ziman. *Electrons and Phonons – The Theory of Transport Phenomena in Solids.* Clarendon Press, Oxford, 1960.

[3] N.W. Ashcroft, N.D. Mermin. *Solid State Physics.* Holt-Saunders, New York, 1976.

[4] W. Jones, N.H. March. *Theoretical Solid State Physics, Vol. 1: Perfect Lattices in Equilibrium.* Dover, New York, 1973.

[5] W. Jones, N.H. March. *Theoretical Solid State Physics, Vol. 2: Non-equilibrium and Disorder.* Dover, New York, 1973.

[6] O. Madelung. *Introduction to Solid State Theory.* Springer, Berlin, 1978

[7] C. Jacoboni, L. Reggiani. The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials. *Rev. Mod. Phys.*, 55(3):645 705, 1983.

[8] W. Bludau, A. Onton. Temperature dependence of the bandgap of silicon. *J. Appl. Phys.*, 45(4):1846–1848, 1974.

[9] R. Pässler. Comparison of different analytical descriptions of the temperature dependence of the indirect energy gap in silicon. *Solid-State Electron.*, 39(9):1311–1319, 1996.

[10] C.D. Thurmond. The standard thermodynamic function of the formation of electrons and holes in Ge, Si, GaAs and GaP. *J. Electrochem. Soc.*, 122:1133, 1975.

[11] A.B. Sproul, M.A. Green, J. Zhao. Improved value for the silicon intrinsic carrier concentration at 300 K. *Appl. Phys. Lett.*, 57(3):255–257, 1990.

[12] R. Vankemmel, W. Schoenmaker, K. de Meyer. A unified wide temperature range model for the energy gap, the effective carrier mass and intrinsic concentration in silicon. *Solid-State Electron.*, 36(10):1379–1384, 1993.

[13] M.A. Green. Intrinsic concentration, effective densities of states, and effective mass in silicon. *J. Appl. Phys.*, 67(6):2944–2954, 1990.

[14] A.H. Marshak, M.A. Shibib, J.G. Fossum, F.A. Lindholm. Rigid band analysis of heavily doped semiconductor devices. *IEEE Trans. Electron Devices*, 28(3):293–298, 1981.

[15] K. Seeger. *Semiconductor Physics.* 3rd edn,Springer, Berlin, 1985.

[16] W. Kohn, J.M. Luttinger. Quantum theory of electrical transport phenomena. *Phys. Rev.*, 108(3):590–611, 1972.

[17] W. Hänsch. *The Drift–Diffusion Equation and its Applications in MOSFET Modeling.* Springer, New York, 1991.

[18] A. Abramo et al. A comparison of numerical solutions of the Boltzmann transport equation for high-energy electron transport silicon. *IEEE Trans. Electron Devices*, 41(9):1646–1654, 1994.

[19] M. Lundstrom. *Fundamentals of Carrier Transport* Modular Series on Solid State Devices, Vol. X. Addison-Wesley, Reading, Massachusetts, 1990.

[20] R. Stratton. Semiconductor current-flow equations (diffusion and degeneracy). *IEEE Trans. Electron Devices*, 19(12):1288–1292, 1972.

[21] A.H. Marshak, K.M. van Vliet. Electrical current in solids with position-dependent band structure. *Solid-State Electron.*, 21:417–427, 1978.

[22] U. Lindefelt. Heat generation in semiconductor devices. *J. Appl. Phys.*, 75(2):942–957, 1994.

[23] W. Quade, E. Schöll, M. Rudan. Impact ionization within the hydrodynamic approach to semiconductor transport. *Solid-State Electron.*, 36(10):1493–1505, 1993.

[24] K. Souissi, F. Odeh, H.H.K. Tang, A. Gnudi, P.-F. Lu. Investigation of the impact ionization in the hydrodynamic model. *IEEE Trans. Electron Devices*, 40(8):1501–1507, 1993.

[25] R. Thoma, A. Edmunds, B. Meinerzhagen, H.-J. Peifer, W.L. Engl. Hydrodynamic equations for semiconductors with nonparabolic band structure. *IEEE Trans. Electron Devices*, 38(6):1343–1353, 1991.

[26] T.J. Bordelon, X.-L. Wang, C.M. Maziar, A.F. Tasch. Accounting for bandstructure effects in the hydrodynamic model: a first-order approach for silicon device simulation. *Solid-State Electron.*, 35(2):131–139, 1992.

[27] M.C. Vecchi, L.G. Reyna. Generalized energy transport models for semiconductor device simulation. *Solid-State Electron.*, 37(10):1705–1716, 1994.

[28] K. Blotekjaer. Transport equations for electrons in two-valley semiconductors. *IEEE Trans. Electron Devices*, 17(1):38–47, 1970.

[29] R.A. Stewart, L. Ye, J.N. Churchill. Improved relaxation-time formulation of collision terms for two-band hydrodynamic models. *Solid-State Electron.*, 32:497–502, 1989.

[30] J. Jyegal, T.A. Demassa. Correct and rigorous single-electron gas hydrodynamic transport model for multi-valley semiconductors. *Solid-State Electron.*, 37(9):1603–1609, 1994.

[31] D.L. Woolard, H. Tian, M.A. Littlejohn, K.W. Kim, R.J. Trew, M.K. Ieong, T.W. Tang. Construction of higher-moment terms in the hydrodynamic electron-transport model. *J. Appl. Phys.*, 74(10):6197–6207, 1993.

[32] T.-W. Tang, S. Ramaswamy, J. Nam. An improved hydrodynamic transport model for silicon. *IEEE Trans. Electron Devices*, 40(8):1469–1477, 1993.

[33] A.M. Anile, O. Muscato. Improved hydrodynamical model for carrier transport in semiconductors. *Phys. Rev. B*, 51(23):16728 16740, 1995.

[34] M.A. Stettler, M.A. Alam, M.S. Lundstrom. A critical examination of the assumptions underlying macroscopic transport equations for silicon devices. *IEEE Trans. Electron Devices*, 40(4):733–740, 1993.

[35] A. Forghieri, R. Guerreri, P. Ciampolini, A. Gnudi, M. Rudan, G. Baccarani. A new discretization strategy of the semiconductor equations comprising momentum and energy balance. *IEEE Trans. CAD*, 7(2):231–242, 1988.

[36] S.-C. Lee, T.-W. Tang. Transport coefficients for a silicon hydrodynamic model extracted from inhomogeneous Monte-Carlo calculations. *Solid-State Electron.*, 35(4):561–569, 1992.

[37] R.O. Grondin, S.M. El-Ghazaly, S. Goodnick. A review of global modeling of charge transport in semiconductors and full-wave electromagnetics. *IEEE Trans. Microwave Theory Tech.*, 47(6):817–829, 1999.

[38] W. Haensch, M. Miura-Mattausch. The hot-electron problem in small semiconductor devices. *J. Appl. Phys.*, 60(2):650–656, 1986.

[39] R.N. Hall. Electron–hole recombination in germanium. *Phys. Rev.*, 87:387, 1952.

[40] W. Shockley, W.T. Read. Statistics of the recombination of holes and electrons. *Phys. Rev.*, 87:835–842, 1952.

[41] C.-T. Sah, R.N. Noyce, W. Shockley. Carrier generation and recombination in p–n junctions and p–n junction characteristics. *Proc. IRE*, 45:1228–1243, 1957.

[42] C.-T. Sah, F.A. Lindholm. Carrier generation, recombination, trapping, and transport in semiconductors with position-dependent composition. *IEEE Trans. Electron Devices*, 24(4):358–362, 1977.

[43] W.D. Eades, R.W. Swanson. Calculation of surface generation and recombination velocities at the Si–SiO_2 interface. *J. Appl. Phys.*, 58(11):4267–4276, 1985.

[44] D.K. Schroder. Carrier lifetimes in silicon. *IEEE Trans. Electron Devices*, 44(1):160–170, 1997.

[45] A. Schrenk. A model for the field and temperature dependence of Shockley–Read–Hall lifetimes in silicon. *Solid-State Electron.*, 35(11):1585 1596, 1992.

[46] G. Vincent, A. Chantre, D. Bois. Electric field effect on the thermal emission of traps in semiconductor junctions. *J. Appl. Phys.*, 50(8):5484–5487, 1979.

[47] G.A.M. Hurkx, D.B.M. Klaassen, M.P.G. Knuvers. A new recombination model for device simulation including tunneling. *IEEE Trans. Electron Devices*, 39(2):331–338, 1992.

[48] O.K.B. Lui, P. Migliorato. A new generation–recombination model for device simulation including the Poole Frenkel effect and phonon-assisted tunneling. *Solid-State Electron.*, 41(4):575–583, 1997.

[49] J. Dziewior, W. Schmid. Auger coefficients for highly-doped and highly excited silicon. *Appl. Phys. Lett.*, 31(5):346–348, 1977.

[50] K.G. McKay. Avalanche breakdown in silicon. *Phys. Rev.*, 94(4):877–884, 1954.

[51] S.L. Miller. Avalanche breakdown in germanium. *Phys. Rev.*, 99(4):1234–1241, 1955.

[52] S.L. Miller. Ionization rates for holes and electrons in silicon. *Phys. Rev.*, 105(4):1246–1249, 1957.

[53] C.A. Lee, R.A. Logan, R.L. Batdorf, J.J. Kleimack, W. Wiegmann. Ionization rates of holes and electrons in silicon. *Phys. Rev.*, 134(3A):A761–A773, 1964.

[54] A.G. Chynoweth. Ionization rates for electrons and holes in silicon. *Phys. Rev.*, 109(5):1537–1540, 1958.

[55] A. Schütz, S. Selberherr, H.W. Pötzl. A two-dimensional model of the avalanche effect in MOS transistors. *Solid-State Electron.*, 25:177–183, 1982.

[56] J.L. Moll, R. van Overstraeten. Charge multiplication in silicon pn junctions. *Solid-State Electron.*, 6:147 157, 1963.

[57] T. Ogawa. Avalanche breakdown and multiplication in silicon pin junctions. *Jpn. J. Appl. Phys.*, 4(7):473–484, 1965.

[58] R. van Overstraeten, H. de Man. Measurement of the ionization rates in diffused silicon pn junctions. *Solid-State Electron.*, 13:583–608, 1970.

[59] W.N. Grant. Electron and hole ionization rates in epitaxial silicon at high electric fields. *Solid-State Electron.*, 16:1189–1203, 1973.

[60] G.A. Baraff. Distribution functions and ionization rates for hot electrons in semiconductors. *Phys. Rev.*, 128(6):2507–2517, 1962.

[61] W. Shockley. Problems related to p–n junctions in silicon. *Solid-State Electron.*, 2(1):35–67, 1961.

[62] J.J. Moll, N.I. Meyer. Secondary multiplication in silicon. *Solid-State Electron.*, 3:155–158, 1961.

[63] J.L. Moll. *Physics of Semiconductors.* McGraw-Hill, New York, 1964.

[64] E.O. Kane. Electron scattering by pair production in silicon. *Phys. Rev.*, 159(3):624–631, 1967.

[65] C.L. Anderson, C.R. Crowell. Threshold energies for electron–hole pair production by impact ionization in semiconductors. *Phys. Rev. B*, 5(6):2267–2272, 1972.

[66] V.M. Robbins, T. Wang, K.F. Brennan, K. Hess, G.E. Stillman. Electron and hole impact ionization coefficients in (100) and (111) Si. *J. Appl. Phys.*, 58(12):4614–4617, 1985.

[67] Y. Okuto, C.R. Crowell. Ionization coefficients in semiconductors: a nonlocalized property. *Phys. Rev. B*, 10(10):4284–4296, 1974.

[68] Y. Okuto, C.R. Crowell. Threshold energy effect on avalanche breakdown voltage in semiconductor junctions. *Solid-State Electron.*, 18:161–168, 1975.

[69] L.V. Keldysh. Concerning the theory of impact ionization in semiconductors. *Sov. Phys. JETP*, 21(6):1135–1144, 1965.

[70] Z.S. Gribnikov, V.M. Ivastchenko, V.V. Mitin. Nonlocality of carrier multiplication in semiconductor depletion layers. *Phys. Status Solidi B*, 105:451–459, 1981.

[71] R.J. McIntyre. A new look at impact ionization – part i: a theory of gain, noise, breakdown, probability, and frequency response. *IEEE Trans. Electron Devices*, 46(8):1623–1631, 1999.

[72] R.J. McIntyre. A new look at impact ionization – part ii: gain and noise in short avalanche photodiodes. *IEEE Trans. Electron Devices*, 46(8):1632–1639, 1999.

[73] E.F. Crabbe, J.M.C. Stork, G. Baccarani, M.V. Fischetti, S.E. Laux. The impact of non-equilibrium transport on breakdown and transit time in bipolar transistors. *IEDM Tech. Dig.*, pp. 463–466, 1990.

[74] J.W. Slotboom, G. Streutker, M.J.V. Dort, P.H. Woerlee, A. Pruijmboom, D.J. Gravesteijn. Non-local impact ionization in silicon devices. *IEDM Tech. Dig.*, pp. 127–130, 1991.

[75] W. Lee et al. Numerical modeling of advanced semiconductor devices. *IBM J. Res. Develop.*, 36(2):208–230, 1992.

[76] P. Palestri, L. Selmi, G.A.M. Hurkx, J.W. Slotboom, E. Sangiorgi. Energy dependent electron and hole impact ionization in Si bipolar transistors. *IEDM Tech. Dig.*, pp. 885–888, 1998.

[77] J.W. Slotboom, G. Streutker, G.J.T. Davids, P.B. Hartog. Surface impact ionization in silicon devices. *IEDM Tech. Dig.*, pp. 494–497, 1987.

[78] S.M. Sze. *Physics of Semiconductor Devices.* 2nd edn Wiley, New York, 1982.

[79] C. Zener, H.H. Wills. A theory of the electrical breakdown of solid dielectrics. *Proc. R. Soc.*, 145:523–529, 1934.

[80] A.G. Chynoweth, K.G. McKay. Internal field emission in silicon pn junctions. *Phys. Rev.*, 106(3):418–426, 1957.

[81] E.O. Kane. Zener tunneling in semiconductors. *J. Phys. Chem. Solids*, 12:181–188, 1962.

[82] P.J. Price, J.M. Radcliffe. Esaki tunneling. *IBM J.*, (10):364–371, 1959.

[83] E.O. Kane. Theory of tunneling. *J. Appl. Phys.*, 32(1):83–91, 1961.

[84] D.R. Fredkin, G.H. Wannier. Theory of electron tunneling in semiconductor junctions. *Phys. Rev.*, 128(5):2054–2061, 1962.

[85] R.T. Shuey. Theory of tunneling across semiconductor junctions. *Phys. Rev.*, 137(4A):A1268–A1277, 1965.

[86] L. Esaki, Y. Miyahara. A new device using the tunneling process in narrow pn junctions. *Solid-State Electron.*, 1(1):13–21, 1960.

[87] A.G. Chynoweth, D.E. Thomas, R.A. Logan. Phonon-assisted tunneling in silicon and germanium Esaki junctions. *Phys. Rev.*, 125(3):877 881, 1962.

[88] A. Schenk. Rigorous theory and simplified model of the band-to-band tunneling in silicon. *Solid-State Electron.*, 36(1):19–34, 1993.

[89] G.A.M. Hurkx. On the modelling of tunneling currents in reverse-biased p–n junctions. *Solid-State Electron.*, 32(8):665–668, 1989.

[90] R.P. Mertens, R.J. van Overstraeten, H.J. de Man. Heavy doping effects in silicon. *Adv. Electron. Electron Phys.*, 55:77–118, 1981.

[91] R.J. van Overstraeten, R.P. Mertens. Heavy doping effects in silicon. *Solid-State Electron.*, 30(11):1077–1087, 1987.

[92] T.N. Morgan. Broadening of impurity bands in heavily doped semiconductors. *Phys. Rev.*, 139(1A):343–348, 1965.

[93] D.D. Kleppinger, F.A. Lindholm. Impurity concentration dependent density of states and resulting Fermi level for silicon. *Solid-State Electron.*, 14:407–416, 1971.

[94] D.S. Lee, J.G. Fossum. Energy-band distortion in highly doped silicon. *IEEE Trans. Electron Devices*, 30(6):626 634, 1983.

[95] E.O. Kane. Thomas Fermi approach to impure semiconductor band structure. *Phys. Rev.*, 131(1):79–88, 1963.

[96] G.D. Mahan. Energy gap in Si and Ge: impurity dependence. *J. Appl. Phys.*, 51(5):2634–2646, 1980.

[97] K.-F. Berggren, B.E. Sernelius. Band-gap narrowing in heavily doped many-valley semiconductors. *Phys. Rev. B*, 24(4):1971–1986, 1981.

[98] S.C. Jain, D.J. Roulston. A simple expression for band gap narrowing (BGN) in heavily doped Si, Ge, GaAs and Ge_xSi_{1-x} strained layers. *Solid-State Electron.*, 34(5):453–465, 1991.

[99] W.L. Kauffman, A.A. Bergh. The temperature dependence of ideal gain in double diffused silicon transistors. *IEEE Trans. Electron Devices*, 15(10):732–735, 1968.

[100] H.C. de Graaff, J.W. Slotboom. Measurements of bandgap narrowing in Si bipolar transistors. *Solid-State Electron.*, 19:857–862, 1976.

[101] C.M. van Vliet. Bandgap narrowing and emitter efficiency in heavily doped emitter structures revisited. *IEEE Trans. Electron Devices*, 40(6):1140 1147, 1993.

[102] D.B.M. Klaassen. A unified mobility model for device simulation - I. model equations and concentration dependence. *Solid-State Electron.*, 35(7):953–959, 1992.

[103] D.B.M. Klaassen, J.W. Slotboom, H.C. de Graaff. Unified apparent bandgap narrowing in n- and p-type silicon. *Solid-State Electron.*, 35:125–129, 1992.

[104] A.A. Volfson, V.K. Subashiev. Fundamental absorption edge of silicon heavily doped with donor or acceptor impurities. *Sov. Phys. Semicond.*, 1(3):327–332, 1967.

[105] M. Balkanski, E. Amzallag, A. Aziza. Infrared absorption in heavily doped n-type Si. *Phys. Status Solidi*, 31:323–330, 1969.

[106] J. Wagner. Heavily doped silicon studied by luminescence and selective absorption. *Solid-State Electron.*, 28(1):25 30, 1985.

[107] S.T. Pantelides, R. Car, A. Selloni. Energy-gap reduction in heavily doped silicon: causes and consequences. *Solid-State Electron.*, 28(1):17–24, 1985.

[108] J. Wagner. Optical characterization of heavily doped silicon. *Solid-State Electron.*, 30(11):1117 1120, 1987.

[109] J. Wagner, J.A. del Alamo. Band-gap narrowing in heavily doped silicon: a comparison of optical and electrical data. *J. Appl. Phys.*, 63(2):425–429, 1983.

[110] W. van Roosbroeck. Theory of the flow of electrons and holes in germanium and other semiconductors. *Bell Syst. Tech. J.*, 29:560–607, 1950.

[111] R.J. van Overstraeten, H.J. de Man, R.P. Mertens. Transport equations in heavy doped silicon. *IEEE Trans. Electron Devices*, 20(3):290–298, 1973.

[112] A.H. Marshak. Modeling semiconductor devices with position-dependent material parameters *IEEE Trans. Electron Devices*, 36(9):1764–1772, 1989.

[113] G.K. Wachutka. Rigorous thermodynamic treatment of heat generation and conduction in semiconductor device modeling. *IEEE Trans. CAD*, 9(11):1141–1149, 1990.

[114] S. Selberherr. *Analysis and Simulation of Semiconductor Devices.* Springer, Vienna, 1984.

[115] G.D. Masetti, S. Solmi, M. Severi. Modeling of carrier mobility against concentration in arsenic-, phosphorus- and boron-doped silicon. *IEEE Trans. Electron Devices*, 30:764–769, 1983.

[116] D.M. Caughey, R.E. Thomas. Carrier mobilities in silicon empirically related to doping and field. *Proc. IEEE*, 55(12):2192–2193, 1967.

[117] S. Reggiani, M. Valdinoci, L. Colalongo, M Rudan, G. Baccarani, A.D. Stricker, F. Illien, N. Felber, W. Fichtner, L. Zullino. Electron and hole mobility in silicon at large operating temperatures - part I: bulk mobility. *IEEE Trans. Electron Devices*, 49(3):490–499, 2002.

[118] E.J. Ryder, W. Shockley. Mobilities of electrons in high electric fields. *Phys. Rev.*, 81:139–140, 1951.

[119] W. Shockley. Hot electrons in germanium and Ohm's law. *Bell Syst. Tech. J.*, 30:990–1034, 1951.

[120] E.M. Conwell. *High Field Transport in Semiconductors.* Academic Press, New York, 1967.

[121] S.A. Schwarz, S.E. Russek. Semi-empirical equations for electron velocity in silicon: part I – bulk. *IEEE Trans. Electron Devices*, (12):1629–1633, 1983.

[122] C. Canali, C. Jacoboni, F. Nava, G. Ottaviani, A. Alberigi-Quaranta. Electron drift velocity in silicon. *Phys. Rev. B*, 12(4):2265–2284, 1975.

[123] C. Jacoboni, C. Canali, G. Ottaviani, A. Alberigi Quaranta. A review of some charge transport properties of silicon. *Solid-State Electron.*, 20:77 89, 1977.

[124] H.S. Bennett. Hole and electron mobilities in heavily doped silicon: comparison of theory and experiment. *Solid-State Electron.*, 26(12):1157–1166, 1983.

[125] J. del Alamo et al. Simultaneous measurements of hole lifetime, hole mobility, and bandgap narrowing in heavily doped silicon. *IEDM Tech. Dig.*, pp. 290 293, 1985.

[126] J. del Alamo. The temperature dependence of ideal gain in double diffused silicon transistors. *IEDM Tech. Dig.*, pp. 24–27, 1986.

[127] C.H. Wang, K. Misiakos, A. Neugroschel. Minority-carrier transport parameters in n-type silicon. *IEEE Trans. Electron Devices*, 37(5):1314 1322, 1990.

[128] C.H. Wang, K. Misiakos, A. Neugroschel. Temperature dependence of minority hole mobility in heavily doped silicon. *Appl. Phys. Lett.*, 57(2):159–161, 1990.

[129] J.M. Dorkel, P. Leturq. Carrier mobilities in silicon semi-empirically related to temperature, doping and injection level. *Solid-State Electron.*, 24(9):821–825, 1981.

[130] M. Rosling, H. Bleichner, P. Jonsson, E. Nordlander. The ambipolar diffusion coefficient in silicon: dependence on excess-carrier concentration and temperature. *J. Appl. Phys.*, 76(5):2855–2859, 1994.

[131] S. Bellone, G.V. Persiano, A.G.M. Strollo. Electrical measurement of electron and hole mobilities as a function of injection level in silicon. *IEEE Trans. Electron Devices*, 43(9):1459–1465, 1996.

[132] M.E. Levinshtein, T.T. Mnatskanov. On the transport equations in popular commercial device simulators. *IEEE Trans. Electron Devices*, 49(4):702–703, 2002.

[133] J.L. Egley, D. Chidambarrao. Strain effects on device characteristics: implementation in drift–diffusion simulators. *Solid-State Electron.*, 36(12):1653–1664, 1993.

[134] H.C. de Graaff, F.M. Klaassen. *Compact Transistor Modeling for Circuit Design.* Springer, Vienna, 1990.

[135] H.K. Gummel. A self-consistent iterative scheme for one-dimensional steady state transistor calculations. *IEEE Trans. Electron Devices*, 11:455 465, 1964.

[136] B.V. Gokhale. Numerical solutions for a one-dimensional silicon npn transistor. *IEEE Trans. Electron Devices*, 17(8):594–602, 1970.

[137] G.D. Hachtel, R.C. Joy, J.W. Cooley. A new efficient one-dimensional analysis program for junction device modeling. *Proc. IEEE*, 60(1):86–98, 1972.

[138] J.W. Slotboom. Computer-aided two-dimensional numerical analysis of a silicon npn transistor. *IEEE Trans. Electron Devices*, 20(8):669–679, 1973.

[139] O. Manck, W.L. Engl. Two-dimensional computer simulation for switching a bipolar transistor out of saturation. *IEEE Trans. Electron Devices*, 22(6):339–347, 1975.

[140] M. Kurata. *Numerical Analysis for Semiconductor Devices.* D.C. Heath, Lexington, 1982.

[141] W.L. Engl, B. Meinerzhagen, H.K. Dirks. Device modeling. *Proc. IEEE*, 71(1):10 33, 1983.

[142] S.P. Gaur, P.A. Habitz, Y.-J. Park, R.K. Cook, Y.-S. Huang, L.F. Wagner. Two-dimensional device simulation program: 2dp. *IBM J. Res. Develop.*, 29(3):242–251, 1985.

[143] R.W. Knepper, S.P. Gaur, F.-Y. Chang, G.R. Srinivasan. Advanced bipolar tranistor modeling: process and device simulation tools for today's technology. *IBM J. Res. Dev.*, 29(3):218–228, 1985.

[144] T. Toyabe, H. Masude, Y. Aoki, S. Shuluri, T. Hagiwara. Three-dimensional device simulator CADDETH with highly convergent matrix solution algorithms. *IEEE Trans. Electron Devices*, 32(10):2038–2044, 1985.

3 Physics and Modeling of Bipolar Junction Transistors

Physics-based compact models provide insight into the relations between the layout, doping profiles and electrical parameters and therefore form a language for communication between process development and circuit design. This chapter looks at the physics of modern integrated bipolar junction transistors with particular emphasis on the derivation of closed-form expressions for their terminal behavior, which are required for the formulation of compact models. On the basis of these results, the most prominent compact models for large-signal and small-signal operation will be considered, together with a presentation of parameter extraction procedures.

3.1 The Regional Approach

The following analysis of minority-carrier transport in the base region follows [1] and assumes low-level-injection conditions – with a majority-carrier distribution essentially determined by the impurity distribution – and a net space charge distribution that is constant with time. Current flow is assumed to be one-dimensional across the base region, located between coordinates[1] $x_{\mathrm{be}} = 0$ and $x_{\mathrm{bc}} = d_{\mathrm{B}}$.

3.1.1 Drift Transistors – Homogeneous-Field Case

In most transistors, the base region is not homogeneously doped: in comparison with the doping concentration $N_{\mathrm{A}}(0)$ on the emitter side, the doping concentration $N_{\mathrm{A}}(d_{\mathrm{B}})$ on the collector side of the base region is typically reduced by two orders of magnitude. The resulting concentration gradient of holes has to be compensated by an electric field, which accelerates electrons on their way to the collector, resulting in a reduced base transit time. In the following, we consider a doping concentration that varies exponentially with position according to

$$N_{\mathrm{A}}(x) \;=\; N_{\mathrm{A}}(0)\exp\left(-\eta\frac{x}{d_{\mathrm{B}}}\right) ,$$

where $\eta = \ln\left[N_{\mathrm{A}}(0)/N_{\mathrm{A}}(d_{\mathrm{B}})\right]$ (Fig. 3.1). The electric field in the base region is estimated by assuming that the hole drift current due to the electric field compensates the hole diffusion current caused by the hole concentration gradient, i.e. $J_{\mathrm{p}} = e\mu_{\mathrm{p}}pE - eD_{\mathrm{p}}\mathrm{d}p/\mathrm{d}x = 0$. With the help of the Einstein

[1] The coordinate notation is described in Sect. 1.5.

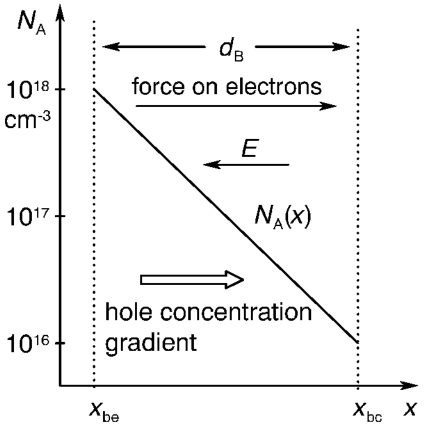


Fig. 3.1. Spatial dependence of doping concentration in the base region for the drift transistor considered here

relation $D_p = V_T \mu_p$ and the approximation $p(x) \approx N_A(x)$, which holds under low-level-injection conditions, the electric field strength can be estimated as

$$E \approx \frac{V_T}{N_A(x)} \frac{dN_A}{dx} = -\eta \frac{V_T}{d_B} . \tag{3.1}$$

If $N_A(0)/N_A(d_B) = 100$ and $d_B = 150\,\mathrm{nm}$, for example, the parameter η is found to be approximately 4.6, corresponding to an electric field strength $|E| \approx 0.79\,\mathrm{V/\mu m}$ at $T = 300\,\mathrm{K}$. Since the space charge in the base is assumed to be constant, this electric field is not a function of time and is entirely determined by the impurity distribution as long as the density of injected minority carriers is small in comparison with the majority-carrier density.

3.1.2 Transfer Current in Frequency and Time Domains

For constant values of the mobility and electric field strength, the current and continuity equations for electrons in the base may be solved after Laplace transformation, as shown in Appendix D. If recombination is neglected, the Laplace transform of the electron current $\underline{i}_{En}(s) = -A_{je}\,\underline{J}_n(0,s)$ that is injected into the base and that of the electron current $\underline{i}_{Cn}(s) = -A_{je}\,\underline{J}_n(d_B,s)$ that reaches the base–collector (bc) depletion layer may be expressed as

$$\underline{i}_{En}(s) = I_S\,\underline{a}_{11}(s)\Lambda_E(s) - I_S\,\underline{a}_{12}(s)\Lambda_C(s) , \tag{3.2}$$

$$\underline{i}_{Cn}(s) = I_S\,\underline{a}_{21}(s)\Lambda_E(s) - I_S\,\underline{a}_{22}(s)\Lambda_C(s) , \tag{3.3}$$

where I_S is the transfer saturation current

$$I_S = \frac{eA_{je}D_n n_{p0}(d_B)}{d_B} \frac{\eta}{2} \frac{\exp(-\eta/2)}{\sinh(\eta/2)} , \tag{3.4}$$

and $\Lambda_E(s)$ and $\Lambda_C(s)$ represent the Laplace transforms

$$\begin{aligned}
\underline{\Lambda}_{\mathrm{E}}(s) &= \mathcal{L}\left[\exp\left(\frac{v_{\mathrm{B'E'}}(t)}{V_{\mathrm{T}}}\right) - 1\right] = \frac{\Delta \underline{n}_{\mathrm{p}}(0,s)}{n_{\mathrm{p0}}(0)} , \\
\underline{\Lambda}_{\mathrm{C}}(s) &= \mathcal{L}\left[\exp\left(\frac{v_{\mathrm{B'C'}}(t)}{V_{\mathrm{T}}}\right) - 1\right] = \frac{\Delta \underline{n}_{\mathrm{p}}(d_{\mathrm{B}},s)}{n_{\mathrm{p0}}(d_{\mathrm{B}})} .
\end{aligned}$$

The coefficients $\underline{a}_{ij}(s)$ approach 1 in the limit $s \to 0$ and are defined as

$$\underline{a}_{21}(s) = \frac{\underline{\vartheta}(s)\sinh[\vartheta(0)]}{\vartheta(0)\sinh[\underline{\vartheta}(s)]} = \underline{a}_{12}(s) , \tag{3.5}$$

$$\underline{a}_{22}(s) = \frac{\underline{\vartheta}(s)\coth[\underline{\vartheta}(s)] - \eta/2}{\vartheta(0)\coth[\vartheta(0)] - \eta/2} , \tag{3.6}$$

$$\underline{a}_{11}(s) = \frac{\underline{\vartheta}(s)\coth[\underline{\vartheta}(s)] + \eta/2}{\vartheta(0)\coth[\vartheta(0)] + \eta/2} , \tag{3.7}$$

where the variable $\underline{\vartheta}(s)$ can be expressed in terms of the (quasi-static) homogeneous base transit time $\tau_{\mathrm{B0}} = d_{\mathrm{B}}^2/2D_{\mathrm{n}}$:

$$\underline{\vartheta}(s) = \sqrt{(\eta/2)^2 + 2s\tau_{\mathrm{B0}}} . \tag{3.8}$$

Since $n_{\mathrm{p0}}(d_{\mathrm{B}}) = n_{\mathrm{ie}}^2/p(d_{\mathrm{B}}) = n_{\mathrm{ie}}^2 \mathrm{e}^{\eta}/p(0)$, (3.4) is equivalent to

$$I_{\mathrm{S}} = \frac{eA_{\mathrm{je}}D_{\mathrm{n}}n_{\mathrm{ie}}^2}{d_{\mathrm{B}}} \frac{\eta}{p(0)\,(1-\mathrm{e}^{-\eta})} .$$

Using the Einstein relation $D_{\mathrm{n}} = k_{\mathrm{B}}T\mu_{\mathrm{n}}/e$ and the identity

$$\int_0^{d_{\mathrm{B}}} p(x)\,\mathrm{d}x = p(0)\int_0^{d_{\mathrm{B}}} \exp\left(-\eta\frac{x}{d_{\mathrm{B}}}\right)\,\mathrm{d}x = p(0)d_{\mathrm{B}}\frac{1-\mathrm{e}^{-\eta}}{\eta} ,$$

which is valid under low-level-injection conditions when $p(x) \approx N_{\mathrm{A}}(x)$, the above equation for I_{S} can therefore be transformed to the Moll–Ross relation [2]

$$(I_{\mathrm{S}} = \frac{eA_{\mathrm{je}}D_{\mathrm{n}}n_{\mathrm{ie}}^2}{\displaystyle\int_0^{d_{\mathrm{B}}} p(x)\,\mathrm{d}x} \approx \frac{eA_{\mathrm{je}}D_{\mathrm{n}}n_{\mathrm{ie}}^2}{\displaystyle\int_0^{d_{\mathrm{B}}} N_{\mathrm{AB}}(x)\,\mathrm{d}x} . \tag{3.9}$$

According (3.9), the transfer current varies in inverse proportion to the base charge. The Moll–Ross relation has been generalized to arbitrary injection levels in Gummel's integral charge control relation (Sect. 3.2).

According to (3.3), the transfer current is the linear response to changes of the minority-carrier densities at the boundaries of the base region: in the frequency domain, these quantities are related by complex frequency-dependent factors. For the purpose of compact modeling, relations between current and voltage, rather than relations between current and minority-carrier densities,

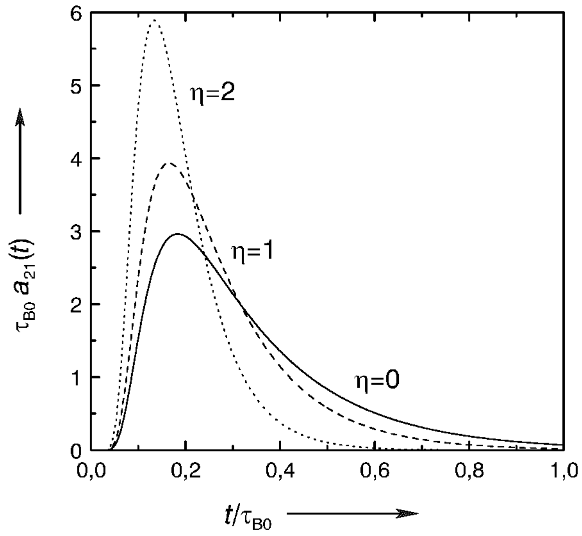


Fig. 3.2. Normalized response function $\tau_{B0}a_{21}(t)$ for drift transistors with different values of η; the case $\eta = 0$ corresponds to a pure diffusion transistor

are required. Unfortunately, no simple relation exists between $\Delta \underline{n}_p(0,s)$ and $\underline{v}_{B'E'}(s)$ in the frequency domain, since

$$\Delta \underline{n}_p(0,s) \;=\; n_{p0}(0)\,\mathcal{L}\left[\exp\!\left(\frac{v_{B'E'}(t)}{V_T}\right) - 1\right]$$

is the Laplace transform of a nonlinear function of $v_{B'E'}(t)$. In the time domain the products relating currents and carrier densities in (3.2) and (3.3) transform to convolution integrals. Using

$$n_p(0,t) \;=\; n_{p0}(0)\left[\exp\left(\frac{v_{B'E'}(t)}{V_T}\right) - 1\right]$$

and

$$n_p(d_B,t) \;=\; n_{p0}(d_B)\left[\exp\left(\frac{v_{B'C'}(t)}{V_T}\right) - 1\right] ,$$

the transfer current $i_T(t) = i_{Cn}(t)$ may therefore be written as the difference

$$i_T(t) \;=\; i_{CE}(t) - i_{EC}(t) \;, \tag{3.10}$$

where

$$i_{CE}(t) \;=\; I_S \int_0^t a_{21}(t-t')\left[\exp\left(\frac{v_{B'E'}(t')}{V_T}\right) - 1\right] \mathrm{d}t' \tag{3.11}$$

is the component of the transfer current that flows from the emitter to the collector, and

$$i_{\mathrm{EC}}(t) \;=\; I_{\mathrm{S}} \int_0^t a_{22}(t-t') \left[\exp\left(\frac{v_{\mathrm{B'C'}}(t')}{V_{\mathrm{T}}} \right) - 1 \right] \mathrm{d}t' \tag{3.12}$$

is the component of the transfer current that flows from the collector to the emitter; this latter component may be neglected under forward-bias conditions, resulting in $i_{\mathrm{T}}(t) \approx i_{\mathrm{CE}}(t)$. Figure 3.2 shows the time dependence of $a_{21}(t)$ for different values of η, as calculated in Appendix D. In particular, at the larger values of η, the function shows a pronounced peak after a characteristic delay $\tau_{\mathrm{a}}(\eta)$, which suggests the approximation

$$a_{21}(t) \;\approx\; \delta[t-\tau_{\mathrm{a}}(\eta)] \,, \tag{3.13}$$

corresponding to a retarded response of $i_{\mathrm{CE}}(t)$ to changes of $v_{\mathrm{B'E'}}(t)$,

$$i_{\mathrm{CE}}(t) \;\approx\; I_{\mathrm{S}} \left[\exp\left(\frac{v_{\mathrm{B'E'}}(t-\tau_{\mathrm{a}})}{V_{\mathrm{T}}} \right) - 1 \right] . \tag{3.14}$$

Such an approximation is used in te Winkel's extended charge control model to be discussed in Sect. 3.1.5.

Since the $\underline{a}_{ij}(s)$ are hyperbolic functions of frequency, no simple expression is obtained for the $a_{ij}(t)$ after transformation to the time domain, as is shown in Appendix D for $a_{21}(t)$, which may be expressed as an infinite series. For the purpose of compact modeling, therefore, approximations are required that provide a trade-off between accuracy and simplicity in order to obtain results with sufficient accuracy in a minimum amount of time.

3.1.3 The Ebers–Moll Model

Introducing the currents $\underline{i}_{\mathrm{f}}(s) = I_{\mathrm{S}}\,\underline{a}_{11}(s)\Lambda_{\mathrm{E}}(s) \approx I_{\mathrm{S}}\Lambda_{\mathrm{E}}(s)$ and $\underline{i}_{\mathrm{r}}(s) = I_{\mathrm{S}}\,\underline{a}_{22}(s)\Lambda_{\mathrm{C}}(s) \approx I_{\mathrm{S}}\Lambda_{\mathrm{C}}(s)$, (3.2) and (3.3) can be written as

$$\begin{aligned} \underline{i}_{\mathrm{En}} &= \underline{i}_{\mathrm{f}}(s) - \underline{\alpha}_{\mathrm{tr}}(s)\,\underline{i}_{\mathrm{r}}(s) \,, \\ \underline{i}_{\mathrm{Cn}} &= \underline{\alpha}_{\mathrm{tf}}(s)\,\underline{i}_{\mathrm{f}}(s) - \underline{i}_{\mathrm{r}}(s) \,, \end{aligned}$$

where $\underline{\alpha}_{\mathrm{tf}}(s) = \underline{a}_{21}(s)/\underline{a}_{11}(s)$ and $\underline{\alpha}_{\mathrm{tr}}(s) = \underline{a}_{12}(s)/\underline{a}_{22}(s)$ are the forward and reverse base transport factors, respectively. The frequency dependence of these quantities is described by a low-pass frequency characteristic (one-pole approximation) in the Ebers–Moll model [3, 4], resulting in

$$\underline{\alpha}_{\mathrm{tf}}(s) \;=\; \frac{\exp(\eta/2)}{\cosh[\underline{\vartheta}(s)] + (\eta/2)\sinh[\underline{\vartheta}(s)]/\underline{\vartheta}(s)} \approx \frac{1}{1+s\tau_{\mathrm{Bf}}} \,, \tag{3.15}$$

$$\underline{\alpha}_{\mathrm{tr}}(s) \;=\; \frac{\exp(-\eta/2)}{\cosh[\underline{\vartheta}(s)] - (\eta/2)\sinh[\underline{\vartheta}(s)]/\underline{\vartheta}(s)} \approx \frac{1}{1+s\tau_{\mathrm{Br}}} \,, \tag{3.16}$$

if recombination in the base region is negligible, as will be assumed in the following. The forward and reverse base transit times τ_{Bf} and τ_{Br} are obtained if the denominators of the exact expressions are developed up to first order in s, with the result

$$\tau_{\mathrm{Bf}} = \frac{2\tau_{\mathrm{B0}}}{\eta^2}\left(\mathrm{e}^{-\eta}+\eta-1\right) \quad \text{and} \quad \tau_{\mathrm{Br}} = \frac{2\tau_{\mathrm{B0}}}{\eta^2}\left(\mathrm{e}^{\eta}-\eta-1\right) . \tag{3.17}$$

The forward and reverse alpha-cutoff frequencies $f_{\alpha\mathrm{f}}$ and $f_{\alpha\mathrm{r}}$ are thus[2]

$$f_{\alpha\mathrm{f}} = 1/2\pi\tau_{\mathrm{Bf}} \quad \text{and} \quad f_{\alpha\mathrm{r}} = 1/2\pi\tau_{\mathrm{Br}} . \tag{3.18}$$

The Ebers–Moll relations[3] are particularly advantageous if the transistor is considered as a device controlled by the emitter current $\underline{i}_{\mathrm{En}}(s)$. This is illustrated by the case of forward operation, where $\underline{i}_{\mathrm{r}}(s) \approx 0$, resulting in a low-pass behavior with a characteristic frequency $f_{\alpha\mathrm{f}} = \omega_{\alpha\mathrm{f}}/2\pi$:

$$\underline{i}_{\mathrm{Cn}}(s) = \frac{1}{1+s/\omega_{\alpha\mathrm{f}}}\,\underline{i}_{\mathrm{En}}(s) . \tag{3.19}$$

The Ebers–Moll model, which understands the bipolar transistor as a current-controlled device, is equivalent to the charge control model considered in the following subsection.

3.1.4 The Charge Control Model

The charge control model [7–9] relates the currents of the transistor to the minority charge stored in the base layer. A relation for the base current is obtained by simple integration of the hole continuity equation

[2]A somewhat different expression for the (forward) alpha-cutoff frequency $f_{c\alpha}$ is obtained if that frequency value is calculated from the condition that the exact expression for $|\underline{\alpha}_{\mathrm{tf}}|$ is reduced by 3 dB; this leads to [5, 6]

$$f_{c\alpha} \approx \frac{1.215}{2\pi\tau_{\mathrm{B0}}}\left[1+\left(\frac{\eta}{2}\right)^{4/3}\right] .$$

The definition employed in this book defines the alpha-cutoff frequency as that frequency value where the one-pole approximation to the base transport factor is reduced by 3 dB with respect to its low-frequency value, and is consistent within the quasi-static approximation.

[3]Improved accuracy is provided by the Narud–Meyer model [1], which uses two-term Taylor series expansions of the coefficients $\underline{a}_{ij}(s)$ to obtain the approximations

$$\underline{a}_{21}(s) = \frac{1}{1+s\tau_{21}} = \underline{a}_{12}(s) ,$$

together with

$$\underline{a}_{22}(s) = 1+s\tau_{22} \quad \text{and} \quad \underline{a}_{11}(s) = 1+s\tau_{11} ,$$

if recombination in the base region is neglected. This gives the following for the frequency dependence of the forward base transport factor:

$$\underline{\alpha}_{\mathrm{tf}}(s) \approx \frac{1}{1+s(\tau_{11}+\tau_{21})+s^2\tau_{11}\tau_{21}} .$$

A detailed discussion of this approach is given in [1].

$$\nabla\cdot\boldsymbol{J}_{\mathrm{p}} = -e\frac{\partial p}{\partial t} - e(R-G) \approx -e\frac{\partial p}{\partial t} - e\,\frac{n_{\mathrm{p}}-n_{\mathrm{p0}}}{\tau_{\mathrm{n}}}$$

over the base region (see Fig. 3.3). Since

$$\int_V \nabla\cdot\boldsymbol{J}_{\mathrm{p}}\,\mathrm{d}^3x = \int_{\partial V}\boldsymbol{J}_{\mathrm{p}}\cdot\mathrm{d}\boldsymbol{\sigma} = i_{\mathrm{BE}}(t)+i_{\mathrm{BC}}(t)-i_{\mathrm{B}}(t)\,,$$

the identity

$$\begin{aligned} i_{\mathrm{B}}(t) &= i_{\mathrm{BE}}(t)+i_{\mathrm{BC}}(t) - \int_V \nabla\cdot\boldsymbol{J}_{\mathrm{p}}(\boldsymbol{x},t)\,\mathrm{d}^3x \\ &= i_{\mathrm{BE}}(t)+i_{\mathrm{BC}}(t)+\frac{\mathrm{d}q_{\mathrm{TB}}}{\mathrm{d}t}+\frac{q_{\mathrm{TB}}}{\tau_{\mathrm{n}}} \end{aligned} \tag{3.20}$$

is obtained, taking account of the fact that

$$\int_V \frac{\partial p}{\partial t}\,\mathrm{d}^3x = \frac{\mathrm{d}}{\mathrm{d}t}\int_V (n_{\mathrm{p}}-n_{\mathrm{p0}})\,\mathrm{d}^3x = \frac{1}{e}\frac{\mathrm{d}q_{\mathrm{TB}}}{\mathrm{d}t}$$

owing to the assumption of quasi-neutrality. Therefore the base current can be expressed in terms of the minority charge[4]

$$q_{\mathrm{TB}}(t) = e\int_V (n_{\mathrm{p}}-n_{\mathrm{p0}})\,\mathrm{d}^3x \tag{3.21}$$

stored in the base region. No approximations are necessary for the derivation of (3.20) from the basic semiconductor equations.

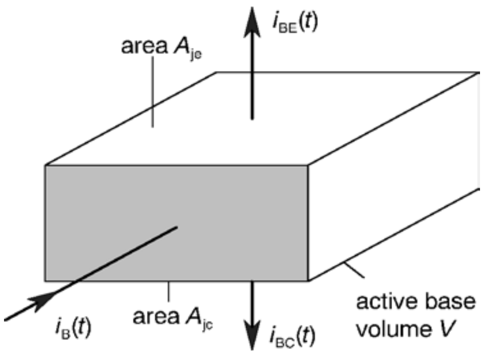


Fig. 3.3. Active base volume considered in the integration of the hole continuity equation

Under dc conditions, the currents I_{CE} and I_{EC} may be expressed in terms of the stored minority charge Q_{TBE} associated with the eb diode and the stored minority charge Q_{TBC} associated with the bc diode. The quantities are related by the forward base transit time τ_{Bf} and the reverse base transit time τ_{Br}, respectively:

[4]Strictly speaking, the injected minority charge is $-q_{\mathrm{TB}}(t)$; the charge q_{TB} corresponds to the excess hole charge in the base region required to neutralize the injected minorities.

$$I_{\mathrm{CE}} = \tau_{\mathrm{Bf}} Q_{\mathrm{TBE}} \quad \text{and} \quad I_{\mathrm{EC}} = \tau_{\mathrm{Br}} Q_{\mathrm{TBC}} \,.$$

If we make the quasi-static assumption, these formulas can be applied to the transient case, resulting in the relations

$$i_{\mathrm{CE}}(t) = \frac{q_{\mathrm{TBE}}(t)}{\tau_{\mathrm{Bf}}} \,, \quad q_{\mathrm{TBE}}(t) = I_{\mathrm{S}} \tau_{\mathrm{Bf}} \left[\exp\left(\frac{v_{\mathrm{B'E'}}(t)}{V_{\mathrm{T}}} \right) - 1 \right] \,, \tag{3.22}$$

and

$$i_{\mathrm{EC}}(t) = \frac{q_{\mathrm{TBC}}(t)}{\tau_{\mathrm{Br}}} \,, \quad q_{\mathrm{TBC}}(t) = I_{\mathrm{S}} \tau_{\mathrm{Br}} \left[\exp\left(\frac{v_{\mathrm{B'C'}}(t)}{V_{\mathrm{T}}} \right) - 1 \right] \,. \tag{3.23}$$

The quasi-static assumption is equivalent to the one-pole approximation to the base transport factor introduced in the Ebers–Moll model as can be seen from the following consideration. The difference $\underline{i}_{\mathrm{En}}(s) - \underline{i}_{\mathrm{Cn}}(s)$ between the electron current $\underline{i}_{\mathrm{En}}(s)$ which is injected across the eb depletion layer into the base region, and the electron current $\underline{i}_{\mathrm{Cn}}(s) = \underline{i}_{\mathrm{CE}}(s)$ which is extracted across the bc depletion layer determines the base current component $\underline{i}_{\mathrm{BB}}(s)$ required to neutralize injected minority carriers and to replace holes lost owing to recombination in the base layer. In terms of the base transport factor $\underline{\alpha}_{\mathrm{tf}}(s) = \underline{i}_{\mathrm{CE}}(s)/\underline{i}_{\mathrm{En}}(s)$ we may write

$$\underline{i}_{\mathrm{BB}}(s) = \left(\frac{1}{\underline{\alpha}_{\mathrm{tf}}(s)} - 1 \right) \underline{i}_{\mathrm{CE}}(s) \,. \tag{3.24}$$

Using the one-pole approximation to the base transport factor (3.19),

$$\underline{\alpha}_{\mathrm{tf}}(s) \approx \frac{1}{1+s/\omega_{\mathrm{c}\alpha}} = \frac{1}{1+s\tau_{\mathrm{Bf}}} \,,$$

we obtain

$$\underline{i}_{\mathrm{BB}}(s) \approx s\, \underline{i}_{\mathrm{CE}}(s)/\omega_{\mathrm{c}\alpha} \,.$$

In a small-signal description where $\underline{i}_{\mathrm{ce}} = g_{\mathrm{m}} \underline{v}_{\pi}$, this current component reads

$$\underline{i}_{\mathrm{bb}} \approx \mathrm{j}(\omega/\omega_{\mathrm{c}\alpha}) \underline{i}_{\mathrm{ce}} = \mathrm{j}\omega c_{\mathrm{tne}} \underline{v}_{\pi} \,,$$

where $c_{\mathrm{tne}} = \tau_{\mathrm{Bf}} g_{\mathrm{m}}$ is the diffusion capacitance of the neutral base layer.[5]

The charge control model utilizes variables which are more closely related to the physics of the bipolar transistor than are those of the Ebers–Moll model. Both approaches, however, describe the currents in terms of first-order ordinary differential equations and are therefore equivalent. As already stated in [7], these relations are only valid at frequencies at which carrier transit times are not significant. So-called "non-quasi-static effects" associated with errors of the quasi-static assumption are considered in the following subsection.

[5] The diffusion capacitance c_{π} of the Giacoletto small-signal model (Sect. 3.9) is larger than this quantity owing to minority-carrier storage in the emitter region.

3.1.5 Non-Quasi-Static Effects

The one-pole approximation underlying the Ebers–Moll and the charge control model is valid only approximately; in particular, deviations are to be expected at high frequencies. In the following, these deviations will be investigated using the analytical results for the drift transistor given in Appendix D. Integration of the electron density $\underline{n}_{\mathrm{p}}(x, s)$ across the base region yields the following for the Laplace transform of the stored minority charge:

$$\underline{q}_{\mathrm{TB}}(s) = \underline{q}_{\mathrm{TBE}}(s) + \underline{q}_{\mathrm{TBC}}(s) , \tag{3.25}$$

where

$$\underline{q}_{\mathrm{TBE}}(s) = \Delta\underline{n}_{\mathrm{p}}(0, s)\frac{eA_{\mathrm{je}}d_{\mathrm{B}}}{\sinh(\underline{\vartheta})}\int_0^1 \mathrm{e}^{\lambda\eta/2}\sinh[\underline{\vartheta}(1-\lambda)]\,\mathrm{d}\lambda \tag{3.26}$$

is that portion of the stored minority charge that is controlled by the eb diode, and

$$\underline{q}_{\mathrm{TBC}}(s) = \Delta\underline{n}_{\mathrm{p}}(d_{\mathrm{B}}, s)\frac{eA_{\mathrm{je}}d_{\mathrm{B}}}{\sinh(\underline{\vartheta})}\mathrm{e}^{-\eta/2}\int_0^1 \mathrm{e}^{\lambda\eta/2}\sinh(\underline{\vartheta}\lambda)\,\mathrm{d}\lambda \tag{3.27}$$

is that portion controlled by the bc diode. For simplicity, we restrict ourselves to forward-bias conditions, where $\underline{q}_{\mathrm{TBC}}(s)$ can be assumed to be negligible. Since

$$\underline{i}_{\mathrm{CE}}(s) = \frac{eA_{\mathrm{je}}D_{\mathrm{n}}}{d_{\mathrm{B}}}\mathrm{e}^{\eta/2}\frac{\underline{\vartheta}}{\sinh(\underline{\vartheta})}\Delta\underline{n}_{\mathrm{p}}(0, s) ,$$

we may write

$$\underline{q}_{\mathrm{TBE}}(s) = \underline{\tau}_{\mathrm{Bf}}(s)\,\underline{i}_{\mathrm{CE}}(s) ,$$

where

$$\begin{aligned}\underline{\tau}_{\mathrm{Bf}}(s) &= 2\tau_{\mathrm{B0}}\frac{\mathrm{e}^{-\eta/2}}{\underline{\vartheta}}\int_0^1 \mathrm{e}^{\lambda\eta/2}\sinh[\underline{\vartheta}(1-\lambda)]\,\mathrm{d}\lambda \\ &= \frac{1}{s}\left[\left(\cosh(\underline{\vartheta}) + \frac{\xi}{\underline{\vartheta}}\sinh(\underline{\vartheta})\right)\mathrm{e}^{-\eta/2} - 1\right]\end{aligned} \tag{3.28}$$

and $\tau_{\mathrm{B0}} = d_{\mathrm{B}}^2/2D_{\mathrm{n}}$. The forward transfer current $i_{\mathrm{CE}}(t)$ and the minority charge $q_{\mathrm{TBE}}(t)$ controlled by the eb junction are therefore related by the transfer factor $\underline{\tau}_{\mathrm{Bf}}(s)$ in the complex frequency domain. The frequency dependence of $\underline{\tau}_{\mathrm{Bf}}(s)$ stems from the factor $1/s$ and from the frequency dependence of $\underline{\vartheta}(s)$. In the presence of a drift field, we may develop the cosh and sinh functions in (3.28) up to second order in s to obtain the approximation

$$\underline{\tau}_{\mathrm{Bf}}(s) \approx \tau_{\mathrm{Bf}}(1+s\tilde{\tau}_1) ,$$

where

$$\tau_{\mathrm{Bf}} = \frac{2\tau_{\mathrm{B0}}}{\eta^2}\,(\eta-1+\mathrm{e}^{-\eta}) \tag{3.29}$$

equals the forward quasi-static base transit time derived using Krömer's relation,[6]

$$\tau_{\mathrm{Bf}} = \frac{1}{D_{\mathrm{n}}}\int_0^{d_{\mathrm{B}}}\frac{1}{p(x)}\int_x^{d_{\mathrm{B}}} p(y)\,\mathrm{d}y\,\mathrm{d}x$$

Here the hole density is $p(x) \approx N_{\mathrm{AB}}(x) = N_{\mathrm{AB}}(0)\,\exp(-\eta x/d_{\mathrm{B}})$. The time constant

$$\tilde{\tau}_1 = \tau_{\mathrm{Bf}}\,\frac{\eta^2-4\eta+6-2(\eta+3)\,\mathrm{e}^{-\eta}}{2\,(\eta-1+\mathrm{e}^{-\eta})^2}$$

describes the phase shift between the stored minority charge and the transfer current; its value increases from $\tau_{\mathrm{Bf}}/6$ in the limit $\eta = 0$ to $\tau_{\mathrm{Bf}}/2$ in the limit of large η.

On the basis of his analysis of the base transport factor of the drift transistor, te Winkel [10] suggested a modification of the charge control model, which describes the transfer current as the delayed response to the stored minority charge in the form

$$i_{\mathrm{CE}}(t) = q_{\mathrm{TBE}}(t-\tau_1)/\tau_{\mathrm{Bf}}\;. \tag{3.30}$$

The additional parameter τ_1 was determined from a comparison with the expansion of the exact solution, with the result

$$\tau_1 \approx \frac{\tau_{\mathrm{Bf}}}{6}\left(1+\frac{4}{15}\,\eta\right)\;. \tag{3.31}$$

This is equivalent to the parameter $\tilde{\tau}_1$, which for small values of η leads to the same approximation. The stored minority charge is described as the delayed response to $\Lambda_{\mathrm{E}}(t) = \exp[v_{\mathrm{BE}}(t)/V_{\mathrm{T}}] - 1$ according to

$$q_{\mathrm{TBE}}(t) = \tau_{\mathrm{Bf}} I_{\mathrm{S}} \Lambda_{\mathrm{E}}(t-\tau_2)\;, \tag{3.32}$$

where

$$\tau_2 \approx \frac{\tau_{\mathrm{Bf}}}{6}\left(1+\frac{2}{5}\,\eta\right) \tag{3.33}$$

has been obtained from an expansion of the exact solution. A comparison with (3.22) shows that this corresponds to the approximation (3.13) suggested by Fig. 3.2, with $\tau_{\mathrm{a}} = \tau_1 + \tau_2$. This model introduces an additional delay and therefore an excess phase shift in the small-signal response. The excess phase shift is further discussed in Sect. 3.9 and Appendix D.

[6] See Sect. 3.5 for further discussion.

3.2 Transfer Current, Early Effect

Within the regional approach described above, the transfer current was derived by analytical solution of the minority-carrier current and continuity equations after the introduction of some idealizations. A different approach, which easily allows one to take account of heavy doping and other effects, is based on Gummel's integral charge control relation [11]. This relation is obtained by transforming the current equation into an integral relation that already takes account of the boundary conditions. In its original form, one-dimensional current flow and a vanishing divergence of the electron current density were assumed, but extensions that take account of a nonvanishing divergence are possible.

3.2.1 The Integral Charge Control Relation

Gummel's integral charge control relation (ICCR) [11,12], which generalizes the Moll–Ross relation to the case of a bias-dependent base charge, underlies the widely known Gummel–Poon model [13] of the bipolar transistor. The only prerequisite inherent in its derivation is the assumption of a constant value of the electron current density along the (one-dimensional) device. To derive the ICCR, the electron current density is written as

$$\begin{aligned} J_{\mathrm{n}} &= -e\mu_{\mathrm{n}} n \frac{\partial \phi_{\mathrm{n}}}{\partial x} = -e\mu_{\mathrm{n}} n_{\mathrm{ie}} \exp\left(\frac{\psi - \phi_{\mathrm{n}}}{V_{\mathrm{T}}}\right) \frac{\partial \phi_{\mathrm{n}}}{\partial x} \\ &= k_{\mathrm{B}} T \mu_{\mathrm{n}} n_{\mathrm{ie}} \exp\left(\frac{\psi}{V_{\mathrm{T}}}\right) \frac{\partial}{\partial x} \exp\left(-\frac{\phi_{\mathrm{n}}}{V_{\mathrm{T}}}\right) , \end{aligned} \tag{3.34}$$

using the fact that $k_{\mathrm{B}}T = eV_{\mathrm{T}}$. The relation $p = n_{\mathrm{ie}} \exp\left[\,(\phi_{\mathrm{p}} - \psi)/V_{\mathrm{T}}\,\right]$ allows us to substitute for $\exp(\psi/V_{\mathrm{T}})$, and (3.34) transforms to

$$\frac{\partial}{\partial x} \exp\left(-\frac{\phi_{\mathrm{n}}}{V_{\mathrm{T}}}\right) = \frac{p}{k_{\mathrm{B}} T \mu_{\mathrm{n}} n_{\mathrm{ie}}^2} \exp\left(-\frac{\phi_{\mathrm{p}}}{V_{\mathrm{T}}}\right) J_{\mathrm{n}} \,.$$

If we introduce the abbreviation

$$\mathit{\Gamma}_{\mathrm{B}}(x,t) = \int_{x_{\mathrm{c}}}^{x} \frac{p}{D_{\mathrm{n}}} \left(\frac{n_{\mathrm{i0}}}{n_{\mathrm{ie}}}\right)^2 \exp\left(\frac{\phi_{\mathrm{pB}} - \phi_{\mathrm{p}}(\xi)}{V_{\mathrm{T}}}\right) \mathrm{d}\xi \,,$$

where ϕ_{pB} denotes the value of the hole quasi-Fermi potential in the base region, this may also be written as

$$\frac{\partial}{\partial x} \exp\left(\frac{\phi_{\mathrm{pB}} - \phi_{\mathrm{n}}}{V_{\mathrm{T}}}\right) = \frac{J_{\mathrm{n}}}{e n_{\mathrm{i0}}^2} \frac{\partial \mathit{\Gamma}_{\mathrm{B}}}{\partial x} \,,$$

making use of the Einstein relation $D_{\mathrm{n}} = V_{\mathrm{T}}\mu_{\mathrm{n}}$. Integrating from the emitter contact x_{e} to the collector contact x_{c} and taking account of the fact that

$$v_{\mathrm{B'E}} = \phi_{\mathrm{pB}} - \phi_{\mathrm{n}}(x_{\mathrm{e}}) \quad \text{and} \quad v_{\mathrm{B'C}} = \phi_{\mathrm{pB}} - \phi_{\mathrm{n}}(x_{\mathrm{c}})$$

gives

$$\exp\left(\frac{v_{\mathrm{B'E}}(t)}{V_{\mathrm{T}}}\right) - \exp\left(\frac{v_{\mathrm{B'C}}(t)}{V_{\mathrm{T}}}\right) = -\frac{1}{en_{\mathrm{i0}}^2}\int_{x_\mathrm{e}}^{x_\mathrm{c}} J_{\mathrm{n}}(x,t)\,\frac{\partial \Gamma_{\mathrm{B}}}{\partial x}\,\mathrm{d}x\ .$$

With the simplifying assumption of a constant electron current density, i.e. $\partial J_{\mathrm{n}}/\partial x = 0$, this yields the following for the transfer current $i_{\mathrm{T}}(t) = -A_{\mathrm{je}}J_{\mathrm{n}}(t)$:

$$i_{\mathrm{T}}(t) = \frac{en_{\mathrm{i0}}^2 A_{\mathrm{je}}}{\Gamma_{\mathrm{B}}(x_\mathrm{c},t)}\left[\exp\left(\frac{v_{\mathrm{B'E}}(t)}{V_{\mathrm{T}}}\right) - \exp\left(\frac{v_{\mathrm{B'C}}(t)}{V_{\mathrm{T}}}\right)\right]\ . \tag{3.35}$$

Except for the assumption $\partial J_{\mathrm{n}}/\partial x = 0$, this equation is an integral representation of the electron current density equation, and therefore equivalent to (3.34), with the advantage, however, that the boundary conditions – in the form of well-defined electron quasi-Fermi potentials at the emitter and collector contacts – have already been taken into account.

If the quasi-Fermi potential for electrons ϕ_{n} can be assumed to be constant across the depletion layers, the limits of integration may be chosen at arbitrary points in the eb and bc space charge regions. This provides a rather good approximation if the corresponding junction is forward biased: in the reverse-biased bc junction, however, the assumption of a constant electron quasi-Fermi potential is not justified, as is discussed in the following subsection. For ease of notation, we use the depletion-layer edges x_{be} and x_{bc} as the limits of integration and introduce the internal voltage drops[7]

$$v_{\mathrm{B'E'}} = \phi_{\mathrm{pB}} - \phi_{\mathrm{n}}(x_{\mathrm{be}}) \quad \text{and} \quad v_{\mathrm{B'C'}} = \phi_{\mathrm{pB}} - \phi_{\mathrm{n}}(x_{\mathrm{bc}})$$

across the eb and bc junctions. If, in addition, the hole quasi-Fermi potential ϕ_{p} is assumed to be constant throughout the internal base region, $\Gamma_{\mathrm{B}}(x,t)$ simplifies to the Gummel number

$$G_{\mathrm{B}} = \int_{x_{\mathrm{be}}}^{x_{\mathrm{bc}}} \frac{p}{D_{\mathrm{n}}}\left(\frac{n_{\mathrm{i0}}}{n_{\mathrm{ie}}}\right)^2 \mathrm{d}x\ ,$$

resulting in the transfer current–voltage characteristic

$$I_{\mathrm{T}} = \frac{en_{\mathrm{i0}}^2 A_{\mathrm{jc}}}{G_{\mathrm{B}}}\left[\exp\left(\frac{V_{\mathrm{B'E'}}}{V_{\mathrm{T}}}\right) - \exp\left(\frac{V_{\mathrm{B'C'}}}{V_{\mathrm{T}}}\right)\right]\ . \tag{3.36}$$

By application of the mean-value theorem, the Gummel number may be rewritten as

$$G_{\mathrm{B}} = \frac{n_{\mathrm{i0}}^2}{\langle D_{\mathrm{n}} n_{\mathrm{ie}}^2\rangle}\int_{x_{\mathrm{be}}}^{x_{\mathrm{bc}}} p(x)\,\mathrm{d}x = \frac{n_{\mathrm{i0}}^2}{\langle D_{\mathrm{n}} n_{\mathrm{ie}}^2\rangle}\,\frac{Q'_{\mathrm{B}}}{e}\ , \tag{3.37}$$

[7]The voltages $v_{\mathrm{B'E'}}$ and $v_{\mathrm{B'C'}}$ defined here may be identified with the applied voltage across the corresponding depletion layer only if the electron quasi-Fermi potential is constant across the depletion layer; otherwise, only the definitions $v_{\mathrm{B'E'}} = \phi_{\mathrm{pB}} - \phi_{\mathrm{n}}(x_{\mathrm{eb}})$ and $v_{\mathrm{B'C'}} = \phi_{\mathrm{pB}} - \phi_{\mathrm{n}}(x_{\mathrm{cb}})$ can be used.

where $Q'_B = Q_B/A_{je}$ is the hole charge per unit area in the active base layer and $\langle D_n n_{ie}^2 \rangle$ denotes a weighted average across the base region. With this notation, the current–voltage characteristic of the transfer current reads

$$I_T = \frac{e^2 A_{je} \langle D_n n_{ie}^2 \rangle}{Q'_B} \left[\exp\left(\frac{V_{B'E'}}{V_T}\right) - \exp\left(\frac{V_{B'C'}}{V_T}\right) \right] . \tag{3.38}$$

Equation (3.38) is commonly referred to as the Gummel transfer current relation, which is a generalization of the Moll–Ross relation developed earlier; (3.38) relates the transfer current to the "base charge" (per unit area) Q'_B and serves as the basis for the derivation of the widely used Gummel–Poon model. The value of Q'_B is closely related with the sheet resistance R_π of the active base layer, which is determined by

$$\frac{1}{R_\pi} = e \int_{x_{be}}^{x_{bc}} \mu_p p \, dx = \langle \mu_p \rangle Q'_B .$$

Replacing Q'_B by $1/\langle \mu_p \rangle R_\pi$ transforms (3.38), for example, to

$$I_T = e^2 A_{je} R_\pi \langle \mu_p \rangle \langle D_n n_{ie}^2 \rangle \left[\exp\left(\frac{V_{B'E'}}{V_T}\right) - \exp\left(\frac{V_{B'C'}}{V_T}\right) \right] .$$

For fixed voltages, the transfer current therefore varies in proportion to the sheet resistance R_π of the active base layer.

3.2.2 Forward Operation, Early Voltage

In forward operation, the bc junction is reverse biased ($V_{BC} < 0$); owing to the saturation of the carrier velocity at the bc junction, the electron density $n_p(x_{bc})$ at the bc depletion-layer edge is given by

$$\begin{aligned} n_p(x_{bc}) &= \frac{I_T}{e A_{je} v_{nsat}} = n_{ie}(x_{bc}) \exp\left(\frac{\psi(x_{bc}) - \phi_n(x_{bc})}{V_T}\right) \\ &= \frac{n_{ie}^2(x_{bc})}{p(x_{bc})} \exp\left(\frac{\phi_{pB} - \phi_n(x_{bc})}{V_T}\right) . \end{aligned}$$

This defines $\phi_{pB} - \phi_n(x_{bc})$ in terms of I_T, and modifies (3.36) to

$$I_T = \frac{e n_{i0}^2 A_{je}}{G_B} \left[\exp\left(\frac{V_{B'E'}}{V_T}\right) - \frac{1}{v_{nsat}} \frac{p(x_{bc})}{n_{ie}^2(x_{bc})} \frac{I_T}{e A_{je}} \right] . \tag{3.39}$$

If (3.39) is solved for I_T, the transfer current–voltage characteristic

$$I_T = e A_{je} n_{i0}^2 \frac{\exp\left(\dfrac{V_{B'E'}}{V_T}\right)}{\displaystyle\int_{x_{be}}^{x_{bc}} \frac{p}{D_n} \frac{n_{i0}^2}{n_{ie}^2} \, dx + \frac{p(x_{bc})}{v_{nsat}} \frac{n_{i0}^2}{n_{ie}^2(x_{bc})}} \tag{3.40}$$

is obtained. In this relation, only x_{bc} varies with the reverse bias $V_{\mathrm{C'B'}}$ at the bc junction. Differentiation with respect to $V_{\mathrm{C'B'}}$ then gives the following for the output conductance $g_{\mathrm{o}} = (\partial I_{\mathrm{T}}/\partial V_{\mathrm{C'B'}})_{V_{\mathrm{B'E'}}}$:

$$g_{\mathrm{o}} = -I_{\mathrm{T}} \frac{\mathrm{d}x_{\mathrm{bc}}}{\mathrm{d}V_{\mathrm{C'B'}}} \frac{\dfrac{p(x_{\mathrm{bc}})}{D_{\mathrm{n}}(x_{\mathrm{bc}})} \dfrac{n_{\mathrm{i0}}^2}{n_{\mathrm{ie}}^2(x_{\mathrm{bc}})} + \dfrac{n_{\mathrm{i0}}^2}{v_{\mathrm{nsat}}} \dfrac{\partial}{\partial x}\left(\dfrac{p}{n_{\mathrm{ie}}^2}\right)\Bigg|_{x_{\mathrm{bc}}}}{G_{\mathrm{B}} + \dfrac{p(x_{\mathrm{bc}})}{v_{\mathrm{nsat}}} \dfrac{n_{\mathrm{i0}}^2}{n_{\mathrm{ie}}^2(x_{\mathrm{bc}})}} .$$

Since $ep(x_{\mathrm{bc}})\,\mathrm{d}x_{\mathrm{bc}}/\mathrm{d}V_{\mathrm{C'B'}} = -c_{\mathrm{jc}}(V_{\mathrm{C'B'}})/A_{\mathrm{je}}$, this result is equivalent to

$$g_{\mathrm{o}} = I_{\mathrm{T}} \frac{n_{\mathrm{i0}}^2}{n_{\mathrm{ie}}^2(x_{\mathrm{bc}})} \frac{\dfrac{1}{D_{\mathrm{n}}(x_{\mathrm{bc}})} + \dfrac{1}{v_{\mathrm{nsat}}} \dfrac{\partial \ln(p/n_{\mathrm{ie}}^2)}{\partial x}\Bigg|_{x_{\mathrm{bc}}}}{G_{\mathrm{B}} + \dfrac{p(x_{\mathrm{bc}})}{v_{\mathrm{nsat}}} \dfrac{n_{\mathrm{i0}}^2}{n_{\mathrm{ie}}^2(x_{\mathrm{bc}})}} \frac{c_{\mathrm{jc}}(V_{\mathrm{C'B'}})}{e\,A_{\mathrm{je}}} .$$

We now use $g_{\mathrm{o}} = I_{\mathrm{T}}/(V_{\mathrm{C'B'}} + V_{\mathrm{AF}})$ to obtain a definition of the Early voltage in terms of the physical field variables, with the result

$$V_{\mathrm{AF}} = \frac{eA_{\mathrm{je}} n_{\mathrm{ie}}^2(x_{\mathrm{bc}})}{n_{\mathrm{i0}}^2 c_{\mathrm{jc}}(V_{\mathrm{C'B'}})} \frac{G_{\mathrm{B}} + \dfrac{p(x_{\mathrm{bc}})}{v_{\mathrm{nsat}}} \dfrac{n_{\mathrm{i0}}^2}{n_{\mathrm{ie}}^2(x_{\mathrm{bc}})}}{\dfrac{1}{D_{\mathrm{n}}(x_{\mathrm{bc}})} + \dfrac{1}{v_{\mathrm{nsat}}} \dfrac{\partial \ln(p/n_{\mathrm{ie}}^2)}{\partial x}\Bigg|_{x_{\mathrm{bc}}}} - V_{\mathrm{C'B'}} . \tag{3.41}$$

Since $V_{\mathrm{AF}} \sim n_{\mathrm{ie}}^2(x_{\mathrm{bc}})$, it is possible to increase the Early voltage and hence the output resistance by increasing $n_{\mathrm{ie}}(x_{\mathrm{bc}})$, as is done in graded-base HBTs (Chapt. 4). In the special case of a homogeneously doped base region, (3.41) simplifies to

$$V_{\mathrm{AF}} = \frac{Q_{\mathrm{B}}}{c_{\mathrm{jc}}(V_{\mathrm{C'B'}})} \left(1 + \frac{D_{\mathrm{n}}}{d_{\mathrm{B}} v_{\mathrm{nsat}}}\right) - V_{\mathrm{C'B'}} , \tag{3.42}$$

where $Q_{\mathrm{B}} = eA_{\mathrm{je}} p d_{\mathrm{B}}$. The Early voltage thus varies in proportion to the base charge Q_{B} and in inverse proportion to the bc depletion-layer capacitance c_{jc}; its value is increased by a factor $(1 + D_{\mathrm{n}}/d_{\mathrm{B}} v_{\mathrm{nsat}})$ owing to velocity saturation in the bc depletion layer, in accordance with [14], if V_{CB} is assumed to be small. This effect is particularly pronounced in narrow-base transistors: for a silicon bipolar transistor with $d_{\mathrm{B}} = 40$ nm and $D_{\mathrm{n}} = 17\ \mathrm{cm^2/Vs}$, this effect yields a 42% increase in comparison with the case $v_{\mathrm{nsat}} \to \infty$. However, this discussion is limited to the low-level-injection case. A discussion of the output conductance under high-level injection has been presented in [15] (see also Sect. 3.7).

3.2.3 Base Charge Partitioning

An extension of Gummel's approach to take account of small-signal ac disturbances superimposed on a dc solution was employed for the numerical solution of the ac small-signal problem in [16]. Equivalent results were derived in [17] starting from a large-signal description, which is summarized here in a slightly modified form. If the spatial variation of the hole quasi-Fermi potential across the neutral base layer is neglected, the electron current density equation yields

$$\exp\left(\frac{v_{\mathrm{B'E'}}(t)}{V_{\mathrm{T}}}\right) - \exp\left(\frac{v_{\mathrm{B'C'}}(t)}{V_{\mathrm{T}}}\right) = -\frac{1}{en_{\mathrm{i0}}^2}\int_{x_{\mathrm{e}}}^{x_{\mathrm{c}}} J_{\mathrm{n}}(x,t)\,\frac{\partial G_{\mathrm{B}}}{\partial x}\,\mathrm{d}x \qquad (3.43)$$

without further approximations, if the function $G_{\mathrm{B}}(x)$ is defined according to

$$G_{\mathrm{B}}(x,t) = \int_{x_{\mathrm{be}}}^{x} \frac{p(x',t)}{D_{\mathrm{n}}(x')}\frac{n_{\mathrm{i0}}^2}{n_{\mathrm{ie}}^2(x')}\mathrm{d}x' \,.$$

The derivation of the Gummel transfer current relation was made possible by the assumption $\partial J_{\mathrm{n}}/\partial x = 0$. If this assumption is dropped and, instead, an integration by parts of the right-hand side of (3.43) is performed, the identity

$$\int_{x_{\mathrm{be}}}^{x_{\mathrm{bc}}} J_{\mathrm{n}}(x,t)\,\frac{\partial G_{\mathrm{B}}}{\partial x}\,\mathrm{d}x = J_{\mathrm{n}}(x_{\mathrm{bc}},t)G_{\mathrm{B}}(x_{\mathrm{bc}},t) - \int_{x_{\mathrm{be}}}^{x_{\mathrm{bc}}} G_{\mathrm{B}}(x,t)\,\frac{\partial J_{\mathrm{n}}}{\partial x}\,\mathrm{d}x$$

is obtained. Replacing $\partial J_{\mathrm{n}}/\partial x$ with the help of the electron continuity equation then yields the following for the transfer current $i_{\mathrm{T}}(t) = -A_{\mathrm{je}}J_{\mathrm{n}}(x_{\mathrm{bc}},t)$, that is, the electron current injected into the bc depletion layer:

$$i_{\mathrm{T}}(t) = i_{\mathrm{TGP}}(t) - eA_{\mathrm{je}}\int_{x_{\mathrm{bc}}}^{x_{\mathrm{bc}}} F_{\mathrm{C}}(x)\left(\frac{\partial n}{\partial t} + R - G\right)\mathrm{d}x \,, \qquad (3.44)$$

where $F_{\mathrm{C}}(x,t) = G_{\mathrm{B}}(x,t)/G_{\mathrm{B}}(x_{\mathrm{c}},t)$ denotes a weighting factor, and $i_{\mathrm{TGP}}(t)$ is the transfer current obtained from Gummel's transfer current relation under the assumption $\partial J_{\mathrm{n}}/\partial x = 0$. Since

$$i_{\mathrm{En}}(t) = i_{\mathrm{T}}(t) + eA_{\mathrm{je}}\int_{x_{\mathrm{be}}}^{x_{\mathrm{bc}}}\left(\frac{\partial n}{\partial t} + R - G\right)\mathrm{d}x \,,$$

the electron current $i_{\mathrm{En}}(t)$ at the emitter-side depletion-layer edge is given by[8]

$$i_{\mathrm{En}}(t) = i_{\mathrm{TGP}}(t) + eA_{\mathrm{je}}\int_{x_{\mathrm{be}}}^{x_{\mathrm{bc}}}[1-F_{\mathrm{C}}(x,t)]\left(\frac{\partial n}{\partial t} + R - G\right)\mathrm{d}x \,. \qquad (3.45)$$

Equations (3.44) and (3.45) are derived from the electron current density equation without any further approximation. If generation and recombination within the base region are neglected, one obtains the following for the electron current at the emitter contact and at the base contact:

[8]The function $1 - F_{\mathrm{C}}(x,t)$ has been introduced as a weighting factor $F_{\mathrm{E}}(x)$ in [17].

$$i_{\mathrm{En}}(t) = i_{\mathrm{TGP}}(t) + \frac{\mathrm{d}q_{\mathrm{BE}}}{\mathrm{d}t} \quad \text{and} \quad i_{\mathrm{T}}(t) = i_{\mathrm{TGP}}(t) - \frac{\mathrm{d}q_{\mathrm{BC}}}{\mathrm{d}t}\,, \tag{3.46}$$

where

$$\frac{\mathrm{d}q_{\mathrm{BE}}}{\mathrm{d}t} = eA_{\mathrm{je}} \int_{\mathrm{x_{be}}}^{x_{\mathrm{bc}}} [1-F_{\mathrm{C}}(x,t)]\,\frac{\partial n}{\partial t}\,\mathrm{d}x\,, \tag{3.47}$$

$$\frac{\mathrm{d}q_{\mathrm{BC}}}{\mathrm{d}t} = eA_{\mathrm{je}} \int_{\mathrm{x_{be}}}^{x_{\mathrm{bc}}} F_{\mathrm{C}}(x,t)\,\frac{\partial n}{\partial t}\,\mathrm{d}x\,. \tag{3.48}$$

Up to this point, generation and recombination are the only processes that have been neglected: the result obtained is therefore as general as the current density equation of the drift–diffusion theory. Under low-level-injection conditions, the weighting factor $F_{\mathrm{C}}(x,t)$ is determined by the dopant distribution and is constant in time; under these conditions, it is possible to define the charges

$$q_{\mathrm{BE}}(t) = eA_{\mathrm{je}} \int_{x_{\mathrm{e}}}^{x_{\mathrm{c}}} [1 - F_{\mathrm{C}}(x)]\, n(x,t)\,\mathrm{d}x\,,$$

$$q_{\mathrm{BC}}(t) = eA_{\mathrm{je}} \int_{x_{\mathrm{e}}}^{x_{\mathrm{c}}} F_{\mathrm{C}}(x) n(x,t)\,\mathrm{d}x$$

as weighted integrals of the minority charge distribution in the base region. With these charges, (3.46) provides a general description of the transfer current under low-level-injection conditions. Although mathematically correct, this approach has a slight disadvantage if applied to analytical device modeling, since the additional terms have lost the explicit dependence on the voltages $v_{\mathrm{B'E'}}(t)$ and $v_{\mathrm{B'C'}}(t)$ which makes the transfer current relation $i_{\mathrm{TGP}}(t)$ so particularly useful. In fact, such a dependence can be reintroduced by the assumption of quasi-static conditions, which assume $q_{\mathrm{BE}}(t) = q_{\mathrm{BE}}[v_{\mathrm{B'E'}}(t), v_{\mathrm{B'C'}}(t)]$ and $q_{\mathrm{BC}}(t) = q_{\mathrm{BC}}[v_{\mathrm{B'E'}}(t), v_{\mathrm{B'C'}}(t)]$ to be functions of $v_{\mathrm{B'E'}}(t)$ and $v_{\mathrm{B'C'}}(t)$ only. The time derivatives then read [16, 17]

$$\frac{\mathrm{d}q_{\mathrm{BE}}}{\mathrm{d}t} = c_{11}\,\frac{\mathrm{d}v_{\mathrm{B'E'}}}{\mathrm{d}t} + c_{12}\,\frac{\mathrm{d}v_{\mathrm{B'C'}}}{\mathrm{d}t}\,, \tag{3.49}$$

$$\frac{\mathrm{d}q_{\mathrm{BC}}}{\mathrm{d}t} = c_{21}\,\frac{\mathrm{d}v_{\mathrm{B'E'}}}{\mathrm{d}t} + c_{22}\,\frac{\mathrm{d}v_{\mathrm{B'C'}}}{\mathrm{d}t}\,, \tag{3.50}$$

where the (bias-dependent) capacitance coefficients are $c_{11} = \mathrm{d}q_{\mathrm{BE}}/\mathrm{d}v_{\mathrm{B'E'}}$, $c_{12} = \mathrm{d}q_{\mathrm{BE}}/\mathrm{d}v_{\mathrm{B'C'}}$, $c_{21} = \mathrm{d}q_{\mathrm{BC}}/\mathrm{d}v_{\mathrm{B'E'}}$ and $c_{22} = \mathrm{d}q_{\mathrm{BC}}/\mathrm{d}v_{\mathrm{B'C'}}$. Making the quasi-static assumption thus results in a charge-partitioning scheme for the minority charge stored in the transistor, and the definition of four capacitances which approximately characterize the high-frequency behavior of the device in forward and reverse operation. Application to the base region of a box-profiled transistor (with a homogeneously doped base region) leads to the decomposition

$$c_{11} = \frac{2}{3}\tau_{\mathrm{Bf}} g_{\mathrm{m}} \quad \text{and} \quad c_{21} = \frac{1}{3}\tau_{\mathrm{Bf}} g_{\mathrm{m}} , \tag{3.51}$$

which is equal to the result obtained by developing the exact solution of the small-signal transport equations up to first order in frequency.

In a small-signal analysis, $F(x,t)$ can be replaced by its dc portion $F_{\mathrm{C0}}(x)$ even under high-level-injection conditions since $\partial n/\partial t = \partial n_1/\partial t$ is a small-signal quantity; the delayed small-signal transfer current then reads

$$i_{\mathrm{t}}(t) = g_{\mathrm{m}} v_\pi - eA_{\mathrm{je}} \frac{\mathrm{d}}{\mathrm{d}t} \int_{x_{\mathrm{be}}}^{x_{\mathrm{bc}}} F_{\mathrm{C0}}(x) n_1(x,t)\, \mathrm{d}x$$

under forward-bias conditions with $v_{\mathrm{c'e'}} = 0$.

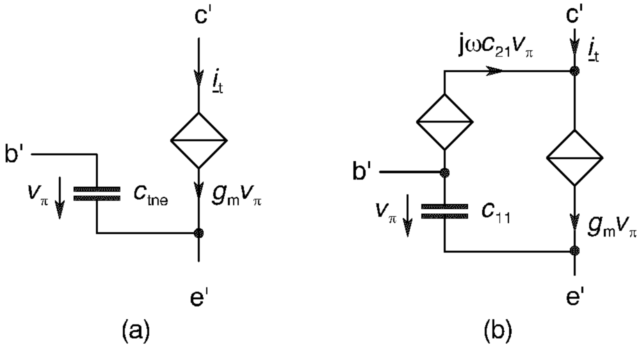


Fig. 3.4. Comparison of the small-signal description of the electron transfer current (**a**) without charge-partitioning and (**b**) with charge-partitioning

If we make the quasi-static assumption again, the phasor of the small-signal transfer current is obtained as

$$\underline{i}_{\mathrm{t}} = g_{\mathrm{m}} \underline{v}_\pi - \mathrm{j}\omega c_{21} \underline{v}_\pi .$$

This description results in an equivalent circuit that differs somewhat from the Giacoletto model, that is, the linearized Gummel–Poon model. This is illustrated in Fig. 3.4, which represents only the network elements associated with the modeling of the small-signal transfer current at $v_{\mathrm{c'e'}} = 0$. Seen from the base node, no difference is associated with the two descriptions, as the diffusion-capacitance c_{tne} due to electrons stored in the base layer in forward operation equals the sum $c_{11} + c_{21}$; however, the transfer current is modeled differently owing to the transcapacitance[9] c_{21}. The difference vanishes to first order in ω if the transconductance g_{m} is replaced by a complex quantity that takes account of the excess phase, i.e. $g_{\mathrm{m}} \to g_{\mathrm{m}}\, \mathrm{e}^{-\mathrm{j}\omega\tau_1}$, where $\tau_1 = c_{21}/g_{\mathrm{m}}$.

[9] This term describes a voltage-controlled current source with a capacitive phase shift.

3.3 Emitter–Base Diode, Current Gain

In dc operation, the base current is caused by generation and recombination within the transistor volume and on its surfaces. Generally, we may distinguish three components I_{BE}, I_{BB} and I_{BC}, associated with minority-carrier injection in the emitter region, neutral base recombination and injection of minority carriers in the collector region, respectively. In forward operation ($V_{\mathrm{BC}} < 0$), the base current of a high-frequency bipolar transistor is dominated[10] by I_{BE}, that is by holes injected into the emitter region. The (static) emitter efficiency of the device, which is defined as

$$\Gamma_{\mathrm{E}} = \frac{I_{\mathrm{T}}}{I_{\mathrm{T}}+I_{\mathrm{BE}}} , \tag{3.52}$$

should be as close to unity as possible, because Γ_{E} determines an upper limit for the current gain,

$$B_{\mathrm{N}} < (1-\Gamma_{\mathrm{E}})^{-1} .$$

Since the thickness of the (monocrystalline) emitter region is small in comparison with the diffusion length, the emitter efficiency is determined by the properties of the emitter contact to a large extent. Polycrystalline emitter contacts have proven advantageous because they show a finite surface recombination velocity and therefore allow the scaling of the monocrystalline emitter region – a process that would result in a diminishing current gain in metal-contacted emitters, as already pointed out in Sect. 1.5. In weakly forward-biased eb junctions, Shockley–Read–Hall recombination in the eb space charge layer may dominate, with the consequence of a bias-dependent current gain. The reverse bias that may be applied to the eb diode is limited, since interband tunneling (the Zener effect) will cause a substantial reverse current even at small reverse bias. This is a consequence of the large electric-field strength in the space charge region that results from the heavily doped emitter and base regions.

3.3.1 Minority-Carrier Transport in Heavily Doped Silicon Emitters

Shallow emitters with contacts of finite surface recombination velocity have already been considered in Sect. 1.4, assuming homogeneous doping throughout the emitter and low-level injection. In general, a mathematical description of minority-carrier transport in heavily doped emitter regions has to take account of effects of heavy doping and position-dependent values of the minority-carrier mobility $\mu_{\mathrm{p}}(x)$ and lifetime $\tau_{\mathrm{p}}(x)$. Owing to the high density of majority carriers, the assumptions of low-level injection and a constant

[10]This is not necessarily true in heterojunction bipolar transistors with wide-bandgap emitters.

value of the electron quasi-Fermi potential $\phi_{\rm n}(x) \approx \phi_{\rm nE}$ throughout the emitter region are fulfilled to a good approximation in the voltage range of practical interest. In the stationary case, minority-carrier transport in the n-type emitter region is then described by the current equation

$$J_{\rm p}(x) = -e\mu_{\rm p}(x)p_{\rm n}(x)\frac{{\rm d}\phi_{\rm p}}{{\rm d}x}$$

and the continuity equation

$$\frac{{\rm d}J_{\rm p}}{{\rm d}x} = -e\,\frac{p_{\rm n}(x)-p_{\rm n0}(x)}{\tau_{\rm p}(x)} = -\frac{en_{\rm ie}^2(x)}{\tau_{\rm p}(x)n(x)}\left[\exp\left(\frac{\phi_{\rm p}(x)-\phi_{\rm nE}}{V_{\rm T}}\right)-1\right] ,$$

rewritten with help of the generalized law of mass action. As is shown in Appendix C, these equations may be combined into an integral equation. In the case of shallow emitters, recombination will occur predominantly at the contact, and the integral equation may be solved by iteration, as is shown in Appendix C, where, for ease of notation, the coordinate system has been chosen such that the emitter depletion-layer boundary has the coordinate $x_{\rm eb} = 0$ and the emitter contact the coordinate $x_{\rm e} = d_{\rm E}$. If recombination in the emitter region is completely negligible, as is the case in so-called transparent emitters ($\tau_{\rm p} \to \infty$), the base saturation current is derived to be

$$I_{\rm BES} = eA_{\rm je}n_{\rm i0}^2\left(G_{\rm E} + \frac{n_{\rm i0}^2}{n_{\rm ie}^2(x_{\rm e})}\frac{n(x_{\rm e})}{S_{\rm nn}}\right)^{-1} , \tag{3.53}$$

where the emitter Gummel number

$$G_{\rm E} = \int_{x_{\rm e}}^{x_{\rm eb}} \frac{n(x)}{D_{\rm p}(x)}\frac{n_{\rm i0}^2}{n_{\rm ie}^2(x)}\,{\rm d}x \tag{3.54}$$

is the analogue of the base Gummel number $G_{\rm B}$ for the emitter region [18,19] and $S_{\rm nn}$ denotes the effective surface recombination velocity at the contact. In the case of a homogeneously doped emitter, $n = N_{\rm D}^+$ and $n_{\rm ie} = n_{\rm iE}$ are constant throughout the emitter region; the integral is then easily evaluated, with the result

$$G_{\rm E} = \frac{N_{\rm D}^+ d_{\rm E}}{D_{\rm p}}\frac{n_{\rm i0}^2}{n_{\rm iE}^2} . \tag{3.55}$$

Using $p_{\rm n0} = n_{\rm iE}^2/N_{\rm D}^+$, this results in

$$I_{\rm BES} = \frac{eA_{\rm je}p_{\rm n0}}{1/S_{\rm nn} + d_{\rm E}/D_{\rm p}} \tag{3.56}$$

in accordance with the findings of Sect. 1.4. From (3.53) and (3.40), one obtains the following for the current gain:

$$B_{\rm N} = \left(G_{\rm E} + \frac{n(x_{\rm e})}{S_{\rm nn}}\frac{n_{\rm i0}^2}{n_{\rm ie}^2(x_{\rm e})}\right) \Big/ \left(G_{\rm B} + \frac{p(x_{\rm bc})}{v_{\rm nsat}}\frac{n_{\rm i0}^2}{n_{\rm ie}^2(x_{\rm bc})}\right) . \tag{3.57}$$

Since $n(x)$ is limited by the solubility of donors in silicon, G_{E} decreases roughly in proportion to the depth of the emitter region d_{E}. Large values of the current gain in down-scaled devices can thus only be maintained if S_{nn} is limited to small values, as is the case if polysilicon emitter contacts are used.

3.3.2 Polycrystalline Emitter Contacts

Polycrystalline silicon was first used by Takagi et al. [20] as a diffusion source, with the intention of forming reliable shallow emitter junctions. The polycrystalline silicon prevented the formation of metal spikes that would otherwise short-circuit shallow junctions. The process was developed further by Graul et al. [21], who deposited an undoped polysilicon layer and employed implantation to precisely control the dopant content of the polycrystalline layer and to form shallow junctions with a reproducible junction depth. The observation of an increase in current gain by a factor of 3 to 7 in comparison with metal-contacted emitters was first reported in [21] for npn transistors. Similar findings were published later for pnp transistors with polysilicon emitter contacts [22].

The observed increase of the current gain is very sensitive to the details of the interface preparation before the deposition of the polysilicon and the details of the emitter drive-in. The reason is the presence of a very thin oxide layer, and dopant segregation at the poly–mono interface. Temperatures in excess of 1000°C during emitter drive-in may cause epitaxial realignment with a substantial decrease in current gain. A few atomic layers at the poly–mono interface are therfore responsible for the increase of the current gain.

In the procedure employed most widely, undoped[11] polycrystalline silicon is deposited on the single-crystal emitter window, followed by n-type doping through implantation of arsenic into the emitter region. Often, a thin amorphous layer with a thickness of the order of 2 nm is formed between the chemical-vapor-deposited polycrystalline silicon and the single-crystal silicon substrate [23]. If this layer is continuous, no epitaxial realignment is observed during emitter drive-in. In the case of a discontinuous amorphous layer, complete epitaxial realignment was found to be possible during emitter drive-in. The speed of realignment depends on the donor concentration and is much greater than in undoped polysilicon. The interface morphology is strongly related to the surface treatment[12] prior to the polysilicon deposition and to the details of the high-temperature annealing employed for emitter drive-in [24]. Wafers which have received an HF dip prior to polysilicon deposition do not

[11]For the fabrication of narrow emitter windows, in-situ doped polysilicon has proven advantageous (see Chap. 7).

[12]For example, an RCA clean of the monocrystalline silicon prior to the deposition of polycrystalline silicon results in a much greater current gain than does an HF etch.

show epitaxial regrowth immediatly after the polysilicon deposition. Subsequent donor implantation and annealing for emitter drive-in, however, lead to partial epitaxial realignment with an areal coverage of several tens of percent [24]. The interfacial oxide layer hinders diffusion of donors and affects the doping profile in the monocrystalline emitter region [25].

The increase in current gain has been attributed to different mechanisms, and a substantial number of papers have been published on the "polysilicon emitter effect". In [21] a reduced bandgap narrowing in heavily doped polycrystalline silicon was assumed, to explain the reduced temperature sensitivity of the current gain in bipolar transistors with polysilicon emitter contacts. The bandgap reductions obtained from temperature measurements of the current gain, however, explained neither the observed increase in current gain nor its strong sensitivity to details of the deposition process. The tunneling model described in [26] is realistic for SiO_2 interfacial layers with a thickness of several nanometers. Since these also hinder majority-carrier transport and result in a substantial emitter resistance, the tunneling model is not very helpful for technologically relevant polysilicon emitters.[13] The segregation model explains the increase of current gain by the presence of a potential barrier at the polysilicon/silicon interface due to dopant segregation (pileup of arsenic at the interface). A reduced value of the hole mobility in the heavily doped n-type polysilicon contact was assumed in [27], to explain the increase in current gain. The oxide tunneling model was found to be inappropriate in [27] as it did not explain the temperature dependence of the base current and a slight dependence of the base current on the thickness of the polysilicon layer. Subsequent investigations [28–30] have shown, however, that the properties of the polysilicon monosilicon interface affect the current gain much more than do the minority-carrier transport properties in the polysilicon layer.

The different mechanisms may be combined to yield unified models [31–33]. Figure 3.5 shows a band diagram of a polysilicon emitter contact with a thin interfacial oxide layer. Holes may cross the interface either by tunneling or by thermal emission, as is illustrated by the current density components J_{pt} and J_{pth}. For small values of the voltage drop across the oxide layer, both components vary in proportion to the difference between the hole densities on either side of the interfacial layer, i.e. $J_{\mathrm{pt}} = eS_{\mathrm{TU}}(p_{\mathrm{R}} - p_{\mathrm{L}})$ and $J_{\mathrm{pth}} = eS_{\mathrm{TH}}(p_{\mathrm{R}} - p_{\mathrm{L}})$, where p_{R} denotes the hole density on the monocrystalline side and p_{L} denotes the hole density on the polycrystalline side [32], and

$$S_{\mathrm{TU}} = e\sqrt{\frac{k_{\mathrm{B}}T}{2\pi m_{\mathrm{p}}^*}}\frac{\mathrm{e}^{-2c_{\mathrm{h}}\chi_{\mathrm{h}}}}{1-c_{\mathrm{h}}k_{\mathrm{B}}T} \quad \text{and} \quad S_{\mathrm{TH}} = A_{\mathrm{h}}^*\frac{T^2}{N_{\mathrm{V}}}\mathrm{e}^{-\chi_{\mathrm{h}}/k_{\mathrm{B}}T} . \qquad (3.58)$$

[13]The kink effect in the base current, described by several authors, is caused not by nonlinearities of the tunneling characteristic but by a substantial collector series resistance, which causes quasi-saturation (see Sect. 3.12).

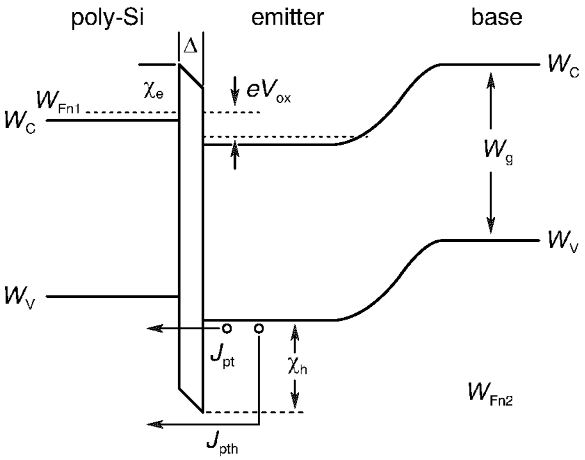


Fig. 3.5. Band diagram of an eb diode with a polycrystalline emitter contact and a thin interfacial oxide layer

Here $c_h = (2\pi\Delta/h)\sqrt{2m_p^*/\chi_h}$, where the effective Richardson constant for holes is A_h^*. The value of p_L is related to the hole current density injected into the polycrystalline silicon by $J_{pt} + J_{pth} \approx S_{poly}(p_L - p_{n0})$, where $S_{poly} = (D_p^{poly}/L_p^{poly}) \coth(d_{poly}/L_p^{poly})$, if minority-carrier transport in the metal-contacted polysilicon is assumed to obey the diffusion equation with a diffusion coefficient D_p^{poly} and diffusion length L_p^{poly}. Taking this together with the recombination velocity S_p, which describes the hole current component $J_{pr} = eS_p(p_R - p_{n0})$ due to recombination of electron hole pairs at the interface, the effective surface recombination velocity at the contact is obtained as

$$S_{nn} = S_p + S_{poly}\frac{S_{TU}+S_{TH}}{S_{poly}+S_{TU}+S_{TH}} = \frac{J_{pt}+J_{pth}+J_{pr}}{e(p_R-p_{n0})}\,. \tag{3.59}$$

This model employs a sufficient number of unknowns, such as the barrier height for holes, the density of recombination centers at the interface and the associated lifetimes, the hole diffusion length in polysilicon, etc., to allow fitting of measured data with good accuracy. Since the parameters employed are not known a priori, however, the method generally adopted is to determine experimentally the effective surface recombination velocity at the poly–mono contact for a specific process technology as a temperature-dependent quantity, written as [34] $S_{nn} = S_{nn,\infty} \exp(-\phi_b/V_T)$, where ϕ_b is a phenomenological thermal "barrier height", responsible for the suppression of hole injection into the polycrystalline emitter contact. Typical data for As-doped emitters with polysilicon contacts can be derived from the data presented in [35]: there $S_{nn} \approx 4 \times 10^4$ cm/s was found at 20°C and $S_{nn} \approx 2 \times 10^5$ cm/s at 200°C, resulting in $S_{nn,\infty} \approx 2.8 \times 10^6$ cm/s and $\phi_b \approx 107$ mV.

3.3.3 Recombination in the Space-Charge Layer

A complete description of the diode current I_{BE} requires the calculation of the current due to recombination in the space charge region,

$$I_{\mathrm{DLE}} \;=\; eA_{\mathrm{je}} \int_{x_{\mathrm{eb}}}^{x_{\mathrm{be}}} (R-G)\,\mathrm{d}x \;.$$

Generation and recombination in the space charge region are mainly caused by SRH processes. If only single-level recombination centers are present, the classical expression for the net recombination rate due to SRH processes reads

$$R-G \;=\; \frac{pn - n_{\mathrm{ie}}^2}{\tau_{\mathrm{p0}}\,(n+n_1) + \tau_{\mathrm{n0}}\,(p+p_1)}\;,$$

where $n_1 = n_{\mathrm{ie}} \exp[(W_{\mathrm{T}} - W_{\mathrm{Fi}})/k_{\mathrm{B}}T]$ and $p_1 = n_{\mathrm{ie}} \exp[(W_{\mathrm{Fi}} - W_{\mathrm{T}})/k_{\mathrm{B}}T]$ represent the electron and hole densities that would be observed if the Fermi level were at W_{T}. If the analysis is restricted to the range of small forward bias, the quasi-Fermi levels ϕ_{n} and ϕ_{p} may be assumed to be constant throughout the space charge region. Using the results

$$n = n_{\mathrm{ie}} \exp\left(\frac{\psi(x)-\phi_{\mathrm{n}}}{V_{\mathrm{T}}}\right)\;, \qquad p = n_{\mathrm{ie}} \exp\left(\frac{\phi_{\mathrm{p}}-\psi(x)}{V_{\mathrm{T}}}\right)$$

and $\phi_{\mathrm{p}} - \phi_{\mathrm{n}} = V_{\mathrm{B'E'}}$, the current due to recombination in the space charge layer may be written as [36]

$$\begin{aligned} I_{\mathrm{DLE}} \;=\; & \frac{eA_{\mathrm{je}}n_{\mathrm{ie}}}{\sqrt{\tau_{\mathrm{p0}}\tau_{\mathrm{n0}}}} \sinh\left(\frac{V_{\mathrm{B'E'}}}{2V_{\mathrm{T}}}\right) \\ & \times \int_{x_{\mathrm{eb}}}^{x_{\mathrm{be}}} \frac{\mathrm{d}x}{b + \cosh\left(\dfrac{2\psi-\phi_{\mathrm{p}}-\phi_{\mathrm{n}}}{2V_{\mathrm{T}}} + \Delta\right)}\;, \end{aligned} \tag{3.60}$$

where $\Delta = \ln\sqrt{\tau_{\mathrm{p0}}/\tau_{\mathrm{n0}}}$ and

$$b \;=\; \exp\!\left(-\frac{V_{\mathrm{B'E'}}}{2V_{\mathrm{T}}}\right) \cosh\left(\frac{W_{\mathrm{T}}-W_{\mathrm{Fi}}}{k_{\mathrm{B}}T} + \Delta\right)\;.$$

Since $\psi(x)$ is a monotonic function of x within the space charge layer, we may substitute for the variable of integration to obtain

$$I_{\mathrm{DLE}} \;=\; eA_{\mathrm{je}} \int_{x_{\mathrm{eb}}}^{x_{\mathrm{be}}} (R-G)\,\mathrm{d}x \;=\; eA_{\mathrm{je}} \int_{\psi(x_{\mathrm{eb}})}^{\psi(x_{\mathrm{be}})} (R-G)\,\frac{\mathrm{d}x}{\mathrm{d}\psi}\,\mathrm{d}\psi\;.$$

In [36], an approximate solution of this integral was presented which assumed an abrupt symmetric pn junction, and that the electric field strength, and therefore $\mathrm{d}x/\mathrm{d}\psi$ were constant throughout the depletion layer. This allows us to approximate $\mathrm{d}\psi/\mathrm{d}x$ as follows:

$$\frac{\mathrm{d}\psi}{\mathrm{d}x} \approx \frac{\psi(x_{\mathrm{be}})-\psi(x_{\mathrm{eb}})}{d_{\mathrm{je}}} = -\frac{V_{\mathrm{JE}}-V_{\mathrm{B'E'}}}{d_{\mathrm{je}}} ,$$

where $d_{\mathrm{je}} = x_{\mathrm{be}} - x_{\mathrm{eb}}$. This yields the approximation[14]

$$I_{\mathrm{DLE}} = \frac{eA_{\mathrm{je}}d_{\mathrm{je}}n_{\mathrm{ie}}}{\sqrt{\tau_{\mathrm{p0}}\tau_{\mathrm{n0}}}} \frac{V_{\mathrm{T}}}{V_{\mathrm{JE}}-V_{\mathrm{B'E'}}} f(b)\, 2\sinh\left(\frac{V_{\mathrm{B'E'}}}{2V_{\mathrm{T}}}\right) , \tag{3.61}$$

where the function $f(b)$ is defined as

$$f(b) = \frac{1}{2}\int_{\psi(x_{\mathrm{bc}})}^{\psi(x_{\mathrm{eb}})} \frac{\mathrm{d}\psi}{b+\cosh\left(\dfrac{2\psi-\phi_{\mathrm{p}}-\phi_{\mathrm{n}}}{2V_{\mathrm{T}}}+\Delta\right)} .$$

The voltage dependence is governed by the term

$$2\sinh\left(\frac{V_{\mathrm{B'E'}}}{2V_{\mathrm{T}}}\right) \approx \exp\left(\frac{V_{\mathrm{B'E'}}}{2V_{\mathrm{T}}}\right) ,$$

which gives an exponential current–voltage characteristic at a forward bias $V_{\mathrm{B'E'}} \gg V_{\mathrm{T}}$. The slope of the logarithmic plot, however, is only approximately half the value expected for an ideal diode. Measured characteristics are generally of the form

$$I_{\mathrm{DLE}} \sim \exp\left(\frac{V_{\mathrm{B'E'}}}{N_{\mathrm{E}}V_{\mathrm{T}}}\right) ,$$

where the value of the emission coefficient N_{E}, which is fitted to the measured data, is close to one if the recombination occurs predominantly in the n- and p-type regions, and close to two if recombination occurs predominantly in the space charge layer.

Figure 3.6 shows (small-signal) emission coefficients N_{E} determined through the relation

$$\frac{1}{N_{\mathrm{E}}} = V_{\mathrm{T}}\frac{\mathrm{d}}{\mathrm{d}V_{\mathrm{BE}}}\ln(I_{\mathrm{B}}) , \tag{3.62}$$

from forward-bias characteristics of eb diodes with different levels of peak base doping density. Deviations from the ideal $\exp(V_{\mathrm{BE}}/V_{\mathrm{T}})$ behavior expected from the Shockley theory are observed for small values of the forward bias, where recombination in the space charge layer predominates, while the classical diffusion current with an emission coefficient close to one is applicable at large forward bias. In the Gummel–Poon model, the base current is therefore modeled as a parallel connection of two diodes with different emission coefficients (Sect. 3.8).

The magnitude of the nonideal current component and the emission coefficient N_{E} increase with the peak base doping concentration N_{A}. This imposes

[14] A generalization of this calculation to asymmetric pn junctions is presented in [37,38].

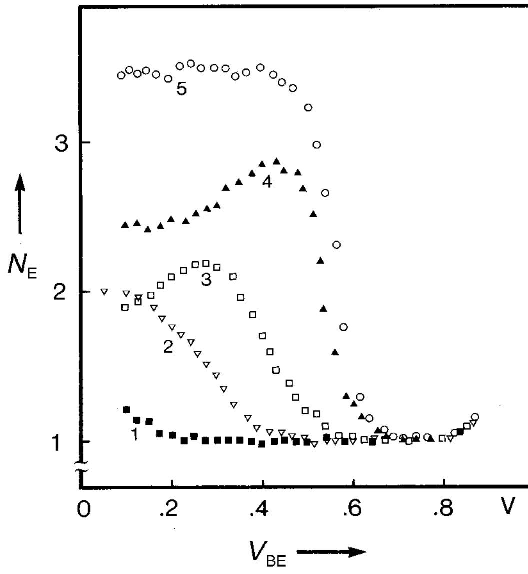


Fig. 3.6. Emission coefficient (small-signal) as a function of forward bias for self-aligned eb diodes with different base doping concentrations. The base implant doses (B, 25 keV) are: (1) $2 \times 10^{13}\,\mathrm{cm}^{-2}$, (2) $4 \times 10^{13}\,\mathrm{cm}^{-2}$, (3) $8 \times 10^{13}\,\mathrm{cm}^{-2}$, (4) $1.6 \times 10^{14}\,\mathrm{cm}^{-2}$, (5) $3.2 \times 10^{14}\,\mathrm{cm}^{-2}$ [39]

a limit on the maximum base dopant concentration in homojunction bipolar transistors [40], which require heavily doped emitter regions in order to obtain good values of emitter efficiency.

For values of N_A in excess of about $3 \times 10^{18}\,\mathrm{cm}^{-3}$, emission coefficients larger than two are observed, an effect that cannot be explained by standard Shockley–Read–Hall theory. Since surface recombination under inverted surface conditions is unlikely, owing to the large doping level, tunneling-assisted SRH processes have to be considered to explain this effect. The effect of field-dependent lifetimes on the emission coefficient becomes evident from the following consideration. The base current under low-level-injection conditions is predominantly due to recombination in the space charge layer; if the bias dependence of $d_\mathrm{je}/(V_\mathrm{JE} - V_\mathrm{BE})$ and series resistances are neglected ($V_\mathrm{BE} = V_\mathrm{B'E'}$), (3.61) may be rewritten as

$$I_\mathrm{B} \sim \frac{1}{\tau(E)} \exp\left(\frac{V_\mathrm{BE}}{2V_\mathrm{T}}\right) ,$$

where $\tau(E) = \sqrt{\tau_\mathrm{n}(E)\tau_\mathrm{p}(E)}$ has to be taken as field-dependent. When this is introduced into the expression (3.62) for the emission coefficient, this yields

$$\frac{1}{N_\mathrm{E}} \approx \frac{1}{2} - V_\mathrm{T}\frac{1}{\tau(E)}\frac{\mathrm{d}\tau}{\mathrm{d}V_\mathrm{BE}} = \frac{1}{2} - V_\mathrm{T}\frac{1}{\tau(E)}\frac{\mathrm{d}\tau(E)}{\mathrm{d}E}\frac{\mathrm{d}E}{\mathrm{d}V_\mathrm{BE}} . \tag{3.63}$$

The second term is negative, since $\mathrm{d}\tau/\mathrm{d}E < 0$ and $\mathrm{d}E/\mathrm{d}V_\mathrm{BE} < 0$, resulting in $1/N_\mathrm{E} < 1/2$ and therefore $N_\mathrm{E} > 2$. Owing to the decrease of the electric field strength in the junction with increasing values of V_BE, the lifetimes increase,

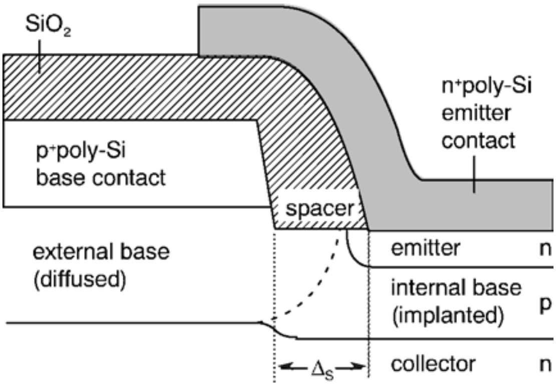


Fig. 3.7. Cross section of self-aligned eb diode fabricated in double-poly technology

with the consequence of a decreasing recombination rate, which causes the base current to increase more slowly with bias, corresponding to a larger value of the emission coefficient.

The nonideal base current component in self-aligned bipolar transistors is often due to an increased defect and dopant density in the "sidewall diode" (Fig. 3.7) and therefore generally varies in proportion to the perimeter of the emitter window rather than to the emitter area. If the emitter area A_{je} is given by $(W_{\mathrm{E}}-2\Delta)(L_{\mathrm{E}}-2\Delta)$ and the emitter perimeter P_{je} is given by $2W_{\mathrm{E}}+2L_{\mathrm{E}}-8\Delta$, where L_{E} and W_{E} denote layout dimensions that deviate by 2Δ from the

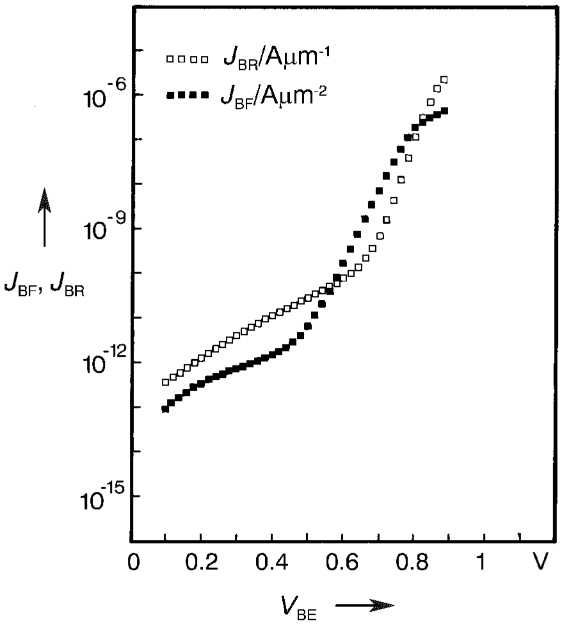


Fig. 3.8. Area and perimeter specific base current contributions [39]

actual dimensions, the base current may be expressed in terms of area and perimeter specific base current densities: $I_B = A_{je} J_{Bf}(V_{BE}) + P_{je} J_{Br}(V_{BE})$. If Δ is estimated from scanning electron microscope pictures, a plot of I_B/A_{je} for given values of V_{BE} versus the ratio P_{je}/A_{je} for transistors with different ratios P_{je}/A_{je} then gives approximately a straight line with axis $J_{Bf}(V_{BE})$ and slope $J_{Br}(V_{BE})$. For small values of V_{BE}, the peripheral component J_{Br} clearly dominates,[15] as is shown by Fig. 3.8. Similar observations in [41] led to the conclusion that the traps causing this nonideal current component are located at the Si/SiO_2 interface of the spacer footing (Fig. 3.7).

The fact that the peripheral base current component is increased because of trap-assisted tunneling is suggested by Fig. 3.9; there $J_{Br}(0.4\,\mathrm{V})$ is plotted versus $1/\sqrt{N_A} \sim 1/E_{max}$, where N_A denotes the value of the implanted boron concentration at the spacer footing as determined by SIMS,[16] and E_{max} is the maximum value of the electric field strength in the sidewall diode.

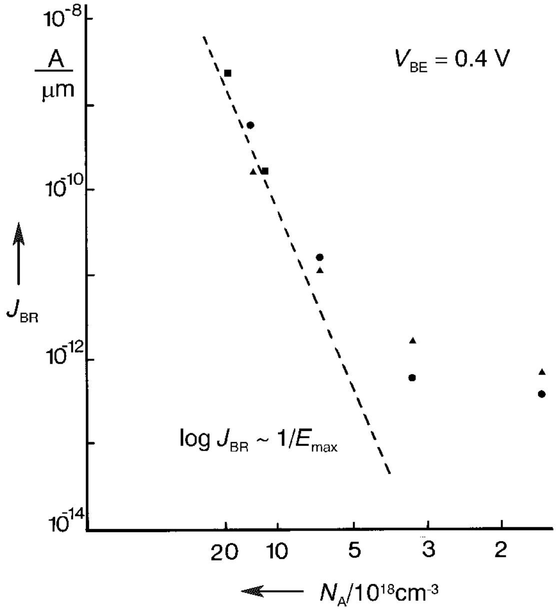


Fig. 3.9. Perimeter specific forward base current component versus peak base doping concentration [39]

For acceptor densities below about $3 \times 10^{18}\,\mathrm{cm}^{-3}$, the observed leakage current density was found to depend weakly on N_A; for peak base doping concentrations larger than about $3 \times 10^{18}\,\mathrm{cm}^{-3}$, however, a strong increase in current

[15] Interestingly, this also occurs at larger bias owing to current crowding at the emitter edge.

[16] The data in Fig. 3.9 were determined from different batches with varying emitter annealing conditions; the variations from batch to batch for devices with comparable base doping concentrations were of the same order of magnitude as the scatter from chip to chip.

density was observed. This increase is far beyond the increase in trapping centers at the boron-doped Si–SiO_2 interface reported in [42], where an increase of the density of surface states roughly in proportion to the boron concentration has been reported. Assuming a similar relation for the density of recombination centers, the four orders of magnitude increase in base current observed in going from $N_A \approx 2 \times 10^{18}\,\mathrm{cm}^{-3}$ to $N_A \approx 2 \times 10^{19}\,\mathrm{cm}^{-3}$ is predominantly due to a decrease of minority-carrier lifetime caused by the tunneling effect. The current is then expected to show an exponential dependence on the value of the maximum electric field strength $E_{\max}$ in the junction, and therefore $\log J_{BR} \sim -1/E \sim -1/\sqrt{N_A}$; a plot of J_{BR} on a logarithmic scale versus $1/\sqrt{N_A}$ should therefore result in a straight line, as is observed in Fig. 3.9 for values of the peak doping concentration exceeding about $5 \times 10^{18}\,\mathrm{cm}^{-3}$.

3.3.4 Reverse-Bias Currents, Breakdown

If a significant reverse bias $-V_{BE} \gg V_T$ is applied to the junction, the net recombination rate simplifies to [36]

$$R - G \approx -\frac{n_{ie}}{2\sqrt{\tau_{n0}\tau_{p0}}}\,\frac{1}{\cosh\left(\dfrac{W_T - W_{Fi}}{k_B T} + \Delta\right)}\,,$$

resulting in a current component

$$I_{DLE} \approx -\frac{eA_{je}d_{je}n_{ie}}{2\sqrt{\tau_{p0}\tau_{n0}}}\,\frac{1}{\cosh\left(\dfrac{W_T - W_{Fi}}{k_B T} + \Delta\right)}\,. \tag{3.64}$$

In this case the generation rate is approximately constant, resulting in a generation current that increases in proportion to the volume $A_{je}d_{je}$ of the space charge region.

Reverse-bias tunneling in heavily doped eb diodes has been investigated, for example, in [39, 43, 44], using polysilicon-contacted bipolar transistors with different base implant doses.[17] In [39], aligned eb configurations were investigated together with self-aligned configurations in order to exclude sidewall diode effects, which are frequently observed in self-aligned bipolar transistors [41]. For a given reverse-bias voltage, the measured current density of the aligned samples scaled with the emitter area and showed a strong increase with the base implant dose, as is expected for interband tunneling. The simplest theoretical models for the description of tunneling currents in pn junctions assume a constant electric field strength E in the junction and predict the tunneling current density to vary as [45]

$$J_{tun} \approx aV_{EB}E\exp(-E_0/E)\,, \tag{3.65}$$

[17] With all other parameters kept fixed.

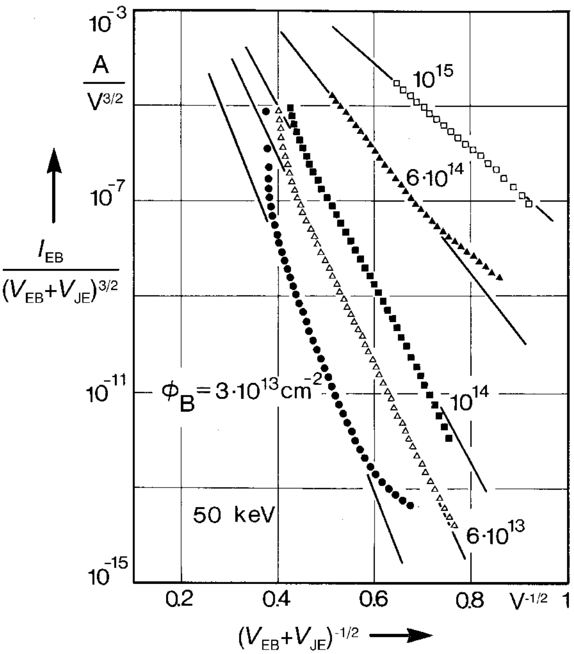


Fig. 3.10. Reverse-bias currents I_{EB} determined for aligned bipolar transistors with different base implant doses ϕ_B (Boron, 50 keV)

where E is generally replaced by the maximum value of the electric field strength[18] in the junction E_{max}, which varies in proportion to $\sqrt{V_{EB}+V_{JE}}$ if the junction is assumed to be abrupt for simplicity. A semilogarithmic plot of $J_{tun}/V_{EB}\sqrt{V_{EB}+V_{JE}}$ versus $1/\sqrt{V_{EB}+V_{JE}}$ should therefore yield a straight line. This is confirmed by Fig. 3.10; the deviations from the straight-line behavior are attributed to Shockley–Read–Hall generation at small values of reverse bias, while the deviations at large voltages are due to carrier multiplication of tunneling-generated carriers. The straight lines were obtained from a least-squares approximation to the linear regime and gave $a = 2.6 \times 10^{-3}\,\mathrm{A/(\mu m\,V^2)}$ and $E_0 = 2.3 \times 10^3\,\mathrm{V/\mu m}$. Although derived from very simplified assumptions, (3.65) together with the experimentally determined parameters a and E_0 provides a valuable means for the quick estimation of reverse-bias currents in heavily doped pn junctions and is also valuable in compact modeling.

[18] In the depletion approximation, the maximum electric field strength is determined by the depletion charge on either side of the junction and is therefore closely related to the charge stored in the depletion capacitance. If E_{m0} denotes the maximum value of the electric field strength in the junction at $V_{EB} = 0$, the maximum electric field strength at a reverse bias V_{EB} is [43]

$$E_{max}(V_{EB}) = E_{m0} + \frac{1}{\epsilon A_{je}} \int_0^{V_{EB}} c_j(V'_{EB})\,dV'_{EB} .$$

3.4 Base–Collector Diode, Breakdown

Advanced bipolar transistors show small values of the open-base breakdown voltage $BV_{\rm CEO}$ of the order of a few volts. This is a consequence of the high current gain and heavily doped collector regions, which are required for large values of the transfer current density. The design of modern ICs is therefore more frequently concerned with the fact that transistors have to be operated with $V_{\rm CE}$ values in the vicinity of or beyond $BV_{\rm CEO}$. Avalanche effects must then be represented in both large- and small-signal models of the bipolar transistor. This requires, in particular, an accurate representation of the multiplication factor $M_{\rm n}(V_{\rm CB}, I_{\rm T})$.

3.4.1 Multiplication Factor

In equivalent-circuit models carrier multiplication is considered with the aid of a controlled current source, which operates in parallel to the bc diode $D_{\rm C}$ as shown in Fig. 3.11. The added current source provides a current

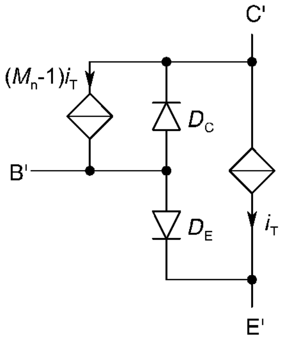


Fig. 3.11. Equivalent circuit for the description of carrier multiplication in the bc diode

$(M_{\rm n} - 1)i_{\rm T}$, where $i_{\rm T}$ is the transfer current of electrons injected into the bc space charge layer and $M_{\rm n}$ is the multiplication factor for injected electrons. In its simplest form, the multiplication factor can be expressed using Miller's empirical formula [46, 47]

$$M_{\rm n} = \frac{1}{1 - (V_{\rm C'B'}/BV_{\rm CBO})^N}\,, \tag{3.66}$$

where $BV_{\rm CBO}$ denotes the breakdown voltage of the bc junction and the exponent N is a parameter determined from a fit to experimental results. The formula (3.66) is not very accurate for silicon junctions; for small values of $M_{\rm n}$, however, it can still be applied in a modified form, if $BV_{\rm CBO}$ is replaced by a smaller voltage $BV_{\rm C}$ that is determined as a fit parameter [48]. This is shown by a double-logarithmic plot of $1 - 1/M_{\rm n}$ versus $V_{\rm CB} \approx V_{\rm C'B'}$, which follows a straight line as long as $1 - 1/M_n$ obeys a power law (see Fig. 3.12 and Sect. 3.12). For larger values of V_{CB}, the increase of $1-1/M_n$ slows down. If the behavior in the vicinity of $M_n = 1$ is described in terms of the tangent to the $1 - 1/M_n$ curve, we arrive at the approximate description

$$1 - \frac{1}{M_n} \approx \begin{cases} (V_{C'B'}/BV_C)^{N_1} & \text{for} \quad 1 - 1/M_n \ll 1 \\ (V_{C'B'}/BV_{CBO})^{N_2} & \text{for} \quad 1 - 1/M_n \approx 1 \end{cases} . \tag{3.67}$$

A weighted average of both limiting formulas often provides an accurate representation of the observed voltage dependence of the multiplication factor M_n (see Appendix D).

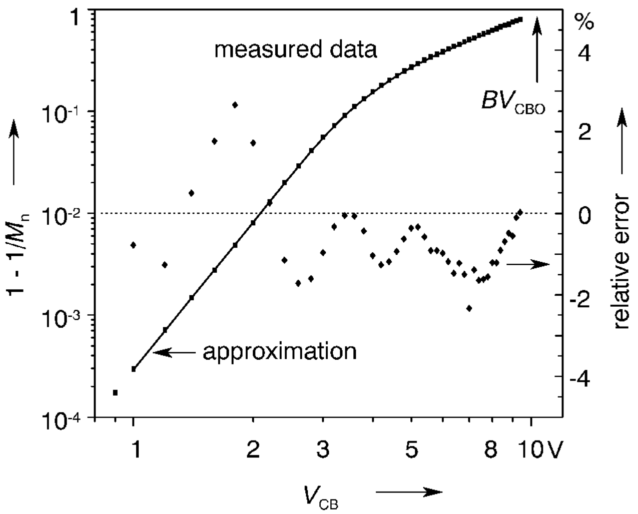


Fig. 3.12. Approximate description of the carrier multiplication factor in self-aligned bipolar transistors according to (3.67) and measured data points (squares). The excellent fit is demonstrated by the relative error (diamonds), which is of the order of 1% over four decades of $1 - 1/M_n$

Figure 3.12 shows a comparison of measured data and the corresponding approximation of $1-1/M_n$ for a high-frequency self-aligned bipolar transistor. Also shown is the relative error,[19] which has a magnitude of the order of 1% over four decades of $1 - 1/M_n$.

An alternative physics-based description of carrier multiplication in the weak avalanche limit has been proposed in [49, 50]. If M_n is close to one, secondary impact ionization may be neglected. If the ionization coefficients are taken to be in the Chynoweth form $\alpha_n = a_n \exp(-b_n/|E|)$, the generation current due to impact ionization I_{AVL} then is [49]

$$I_{AVL} \approx I_T\, a_n \int_{x_{bc}}^{x_{cb}} e^{-b_n/|E(x)|}\, dx = (M_n - 1) I_T .$$

[19]The parameters of the fitting curve were determined using the method outlined in Sect. 3.12.

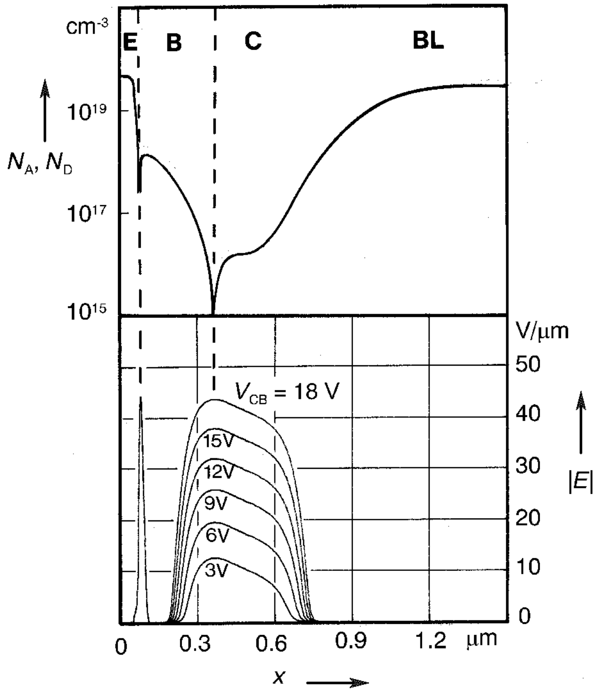


Fig. 3.13. Dopant distribution and electric field strength in a bc diode of a vertical npn bipolar transistor (simulation) [48]

An expansion of the integrand around its maximum now allows us to approximate the ionization integral, leading to

$$M_{\rm n}-1 \;\approx\; \frac{a_{\rm n}}{b_{\rm n}} E_{\max} d_{\rm jc} {\rm e}^{-b_{\rm n}/E_{\max}} \;,$$

where $d_{\rm jc}$ denotes the width of the bc depletion layer[20] In [49, 50], the generation current due to impact ionization in the bc junction is modeled in the form

$$I_{\rm AVL} \;=\; I_{\rm T}\,\frac{a_{\rm n}}{b_{\rm n}}\,\frac{V}{1-M_{\rm JC}}\,\exp\left\{-A_{\rm VL}\left[V\left(1-\theta\frac{I_{\rm T}}{I_{\rm HC}}\right)\right]^{M_{\rm JC}-1}\right\}\;, \qquad (3.68)$$

where $V = V_{\rm C'B'}+V_{\rm JC}$ and $I_{\rm HC} = eA_{\rm je}N_{\rm Depi}v_{\rm nsat}$. The parameter θ takes account of the effect of mobile charge in the space charge layer, which reduces the electric field in the junction and thus the value of the multiplication factor. Neglecting this effect (i.e. setting $\theta = 0$) and introducing the quantities $A_{\rm VC1} = a_{\rm n}/b_{\rm n}(1-M_{\rm JC})$ and $A_{\rm VL} = A_{\rm VC1}$ yields the formula used in the VBIC95 model (Sect. 3.14),

[20] A better approximation replaces $d_{\rm jc}$ by an effective width $d_{\rm eff}$ [52]. For the purpose of compact modeling, the approximation performed here is sufficient, since $d_{\rm jc}$ is fitted to measured data by means of the parameter $A_{\rm VC1}$.

$$M_\mathrm{n} - 1 \;=\; A_\mathrm{VC1} V \exp\Big(-A_\mathrm{VC2} V^{M_\mathrm{JC}-1}\Big)\ . \tag{3.69}$$

This formula uses two parameters to fit experimental data, since V_JC and M_JC are determined from capacitance measurements.

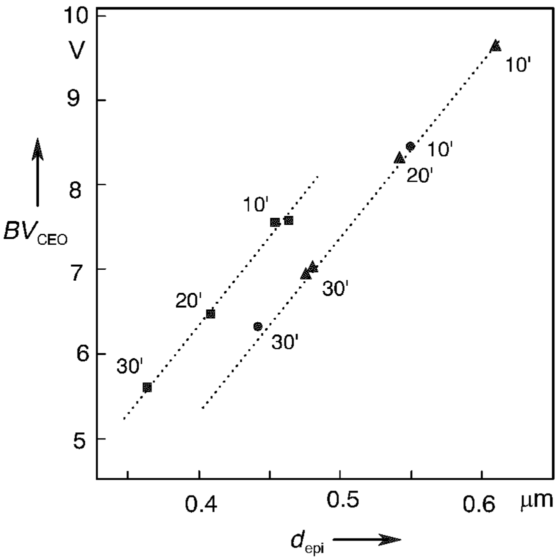


Fig. 3.14. Values of open base breakdown voltage versus epi layer thickness. Identical symbols correspond to transistors with equally doped epitaxial layer and identical forward current gain [53]

3.4.2 Collector–Emitter Breakdown due to Impact Ionization

The open-base breakdown voltage BV_CEO is an important design parameter, since it represents a worst-case situation for collector–emitter breakdown. The condition of a vanishing current at the base terminal ($I_\mathrm{B} = 0$) was used in Sect. 1.5 to derive the condition for open-base breakdown,

$$(M_\mathrm{n} - 1)B'_\mathrm{N} \;=\; 1\ ; \tag{3.70}$$

the value of the open-base breakdown voltage BV_CEO is determined by the current gain[21] B_N and the multiplication factor M_n.

In a bc diode with a completely depleted epilayer, the electric field can be approximated by a trapezoid (Fig. 3.13); if the dopant density is N_Depi and the thickness is d_epi, the maximum value of the electric field strength is approximately

$$E_\mathrm{max} \;\approx\; \frac{V_\mathrm{CB}+V_\mathrm{JC}}{d_\mathrm{epi}} + \frac{eN_\mathrm{Depi}d_\mathrm{epi}}{2\epsilon}\ .$$

[21] $B'_\mathrm{N} \approx I_\mathrm{T}/I_\mathrm{BE} \approx B_\mathrm{N}(V_\mathrm{CB}=0) \times q_\mathrm{B}(V_\mathrm{CB}=0)/q_\mathrm{B}(V_\mathrm{CB})$ is understood as the current gain value that would be observed in the absence of carrier multiplication.

If breakdown occurs when a critical value E_{crit} of the electric field strength in the junction is exceeded, the open-base breakdown voltage [51]

$$BV_{\mathrm{CEO}} \approx V_{\mathrm{CB}}(E_{\mathrm{crit}}) + V_{\mathrm{BEon}} \approx V_{\mathrm{CB}}(E_{\mathrm{crit}}) + V_{\mathrm{JC}}$$

shows the following dependence on the thickness of the epilayer:

$$BV_{\mathrm{CEO}} \approx \left(E_{\mathrm{crit}} - \frac{eN_{\mathrm{Depi}} d_{\mathrm{epi}}}{2\epsilon} \right) d_{\mathrm{epi}} \, .$$

If N_{Depi} and d_{epi} are small, the second term in the brackets may be neglected, and BV_{CEO} varies in proportion to d_{epi}. Such a behavior is demonstrated in Fig. 3.14, where measured values of BV_{CEO} are plotted versus d_{epi}.[22]

3.4.3 Punchthrough

If the base layer becomes very thin, collector–emitter breakdown may occur owing to punchthrough. In this case the bc depletion-layer edge reaches the eb depletion layer and causes a reduction of the barrier height seen by the electrons in the emitter. Thermal emission [54–56] of electrons across the reduced barrier then causes a strong increase of the collector current.

Punchthrough can only be observed if the (metallurgical) base layer is very thin. Except for certain specialized microwave transistors, the base width is chosen to be large enough to prevent punchthrough. As a rough estimate, one may consider the Forward Early voltage V_{AF}: if V_{AF} exceeds the open-base breakdown voltage BV_{CEO}, no punchthrough is to be expected – at least as long as the lateral variations of the base doping are small.

Punchthrough Voltage. In the case of homogeneous doping in both the base and the collector, the extent of the space charge layer into the p-doped region is given by[23]

$$x_{\mathrm{p}}(V_{\mathrm{CB}}) = x_{\mathrm{jc}} - x_{\mathrm{bc}} = \sqrt{\frac{2\epsilon N_{\mathrm{DC}}(V_{\mathrm{JC}} + V_{\mathrm{CB}})}{e(N_{\mathrm{AB}} + N_{\mathrm{DC}})N_{\mathrm{AB}}}} \, ;$$

the base width d_{B} at V_{CB} and $V_{\mathrm{BE}} = 0$ is then given by

$$d_{\mathrm{B}} = d_{\mathrm{B0}} + x_{\mathrm{p}}(0) - x_{\mathrm{p}}(V_{\mathrm{CB}}) \, .$$

Setting $d_{\mathrm{B}} = 0$ and $V_{\mathrm{pt}} \approx V_{\mathrm{CB}} + V_{\mathrm{BE}} \approx V_{\mathrm{CB}} + V_{\mathrm{JC}}$, we therefore obtain

[22] The values of d_{epi} were determined from capacitance measurements and defined as the thickness of the n-type epitaxial layer where $N_{\mathrm{D}} \leq 6 \times 10^{17}\ \mathrm{cm}^{-3}$. The proportionality $BV_{\mathrm{CEO}} \sim d_{\mathrm{epi}}$ is only expected if the epilayer is depleted at voltages considerably below the breakdown voltage; if this condition is not fulfilled, or if breakdown occurs without full depletion of the epilayer, the proportionality $BV_{\mathrm{CEO}} \sim d_{\mathrm{epi}}$ is not fulfilled.

[23] As long as the space charge region does not reach the buried layer on the collector side.

$$V_{\rm pt} \approx \frac{e(N_{\rm DC}+N_{\rm AB})}{2\epsilon}\frac{N_{\rm AB}}{N_{\rm DC}}\,[\,d_{\rm B0}+x_{\rm p}(0)\,]^2 \,.$$

If $N_{\rm DC} \ll N_{\rm AB}$ and $N_{\rm AB}d_{\rm B}$ is expressed in terms of the base sheet resistance $R_\pi = 1/(e\mu_{\rm p}N_{\rm AB}d_{\rm B})$, we may write

$$V_{\rm pt} \approx \frac{\left(1+\mu_{\rm p}R_\pi\sqrt{2e\epsilon N_{\rm DC}V_{\rm JC}}\right)^2}{2e\epsilon\mu_{\rm p}^2}\frac{1}{R_\pi^2}\frac{1}{N_{\rm DC}} \,.$$

The punchthrough voltage therefore increases roughly in inverse proportion to the doping density of the epitaxial collector region $N_{\rm DC}$ and the square of the internal base sheet resistance R_π. In a self-aligned vertical bipolar transistor there may be a reduced doping below the spacer,[24] resulting in a punchthrough at the emitter edge [57].

Punchthrough Current. The current flow under punchthrough conditions may by calculated from a modification of Gummel's approach [54, 55]. Integration of

$$\frac{\partial}{\partial x}\exp\left(-\frac{\phi_{\rm n}}{V_{\rm T}}\right) = \frac{J_{\rm n}}{eD_{\rm n}n_{\rm ie}}\exp\left(-\frac{\psi}{V_{\rm T}}\right)$$

from $x_{\rm eb}$ to $x_{\rm cb}$, with $D_{\rm n}$ taken as constant, yields the following for the electron current density:

$$J_{\rm n} = eD_{\rm n}N_{\rm DE}\frac{\exp\left(-\dfrac{V_{\rm CE}}{V_{\rm T}}\right)-1}{\displaystyle\int_{x_{\rm eb}}^{x_{\rm cb}}\exp\left(\frac{\psi(x_{\rm m})-\psi(x)}{V_{\rm T}}\right)\,{\rm d}x}\exp\left(\frac{\psi(x_{\rm m})-\psi(x_{\rm eb})}{V_{\rm T}}\right) ,$$

since $\phi_{\rm n}(x_{\rm cb}) - \phi_{\rm n}(x_{\rm eb}) = V_{\rm CE}$ and

$$n(x_{\rm eb}) = n_{\rm ie}\exp\left(\frac{\psi(x_{\rm eb})-\phi_{\rm n}(x_{\rm eb})}{V_{\rm T}}\right) \approx N_{\rm DE} \,.$$

In the vicinity of the minimum of the electrostatic potential located at $x = x_{\rm m}$ (see Fig. 3.15), the potential shows a quadratic dependence on x:

$$\psi(x) - \psi(x_{\rm m}) = \frac{eN_{\rm AB}}{2\epsilon}(x-x_{\rm m})^2 \,.$$

Since the integrand decreases quickly if x deviates from $x_{\rm m}$, the limits of the integral can be shifted to infinity for an approximate evaluation:

$$\int_{x_{\rm eb}}^{x_{\rm cb}}\exp\left(\frac{\psi(x_{\rm m})-\psi(x)}{V_{\rm T}}\right)\,{\rm d}x \approx \int_{-\infty}^{\infty}\exp\left(-\frac{(x-x_{\rm m})^2}{2L_{\rm D}^2}\right)\,{\rm d}x = \sqrt{2\pi}\,L_{\rm D} \,,$$

[24] This occurs in particular if the base implant is performed after spacer formation or if the base is diffused from the emitter polysilicon.

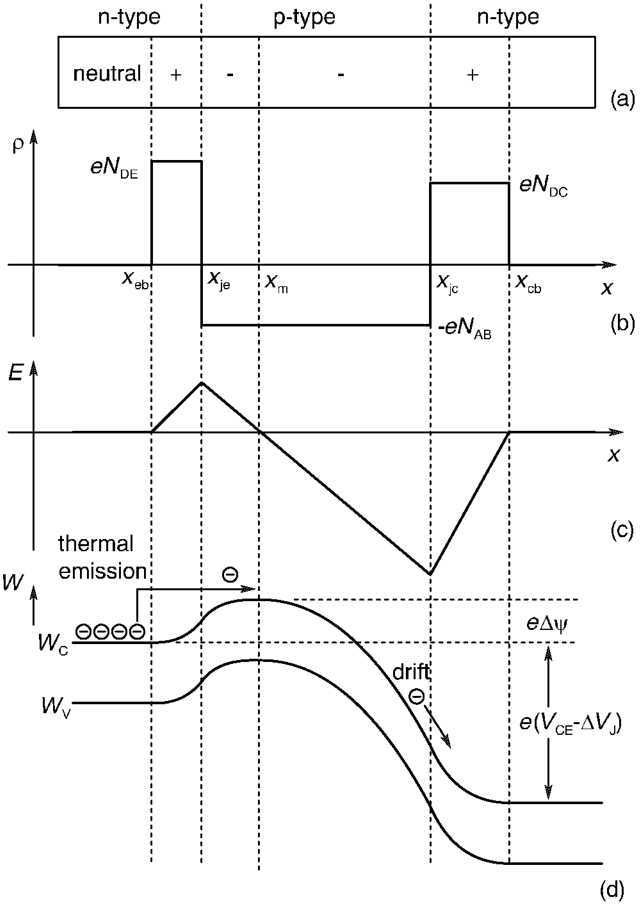


Fig. 3.15. (**a**) npn bipolar transistor under punchthrough conditions, (**b**) distribution of charge, (**c**) electric field and (**d**) one-dimensional band scheme

where $L_D = \sqrt{\epsilon V_T / e N_{AB}}$ is the Debye length in the p-type layer.

In the limit $V_{CE} \gg V_T$, the punchthrough current $I_{pt} = -A_{jc} J_n$ is therefore given by

$$I_{pt} = \frac{e A_{je} D_n N_{DE}}{\sqrt{2\pi} L_D} \exp\left(\frac{-\Delta\psi}{V_T}\right) ; \tag{3.71}$$

its bias dependence stems from the barrier height $\Delta\psi = \psi(x_{eb}) - \psi(x_m)$.

At $V_{CE} < V_{pt}$, the value of $\Delta\psi$ is not affected by V_{CE}; in this case an increase of V_{CE} only increases the voltage drop V_{CB} and hence the extent of the bc depletion layer, which reaches the eb depletion layer at $V_{CE} = V_{pt}$. For voltages $V_{CE} > V_{pt}$, the whole base is depleted and additional positive charge on the collector side has to be neutralized by extra negative charge on the emitter side, i.e. x_{eb} shifts to the right if V_{CE} is further increased. If $x_{cb} - x_{eb} = d_{pt}$

is taken to be approximately constant, an increase of $V_{\rm CE}$ beyond $V_{\rm pt}$ causes the charge increment[25]

$$\Delta Q_{\rm C} \;=\; \frac{\epsilon A_{\rm je}}{d_{\rm pt}}\,(V_{\rm CE}-V_{\rm pt})$$

in the collector region. This charge has to be neutralized by an increment of negative charge in the emitter; the magnitude of this increment can be written as

$$\Delta Q_{\rm E} \;=\; \frac{C_{\rm JE}V_{\rm JE}}{1-M_{\rm JE}}\left[1-\left(\frac{\Delta\psi}{V_{\rm JE}}\right)^{1-M_{\rm JE}}\right]\;,$$

if the bias dependence of the eb depletion capacitance is written in the form[26]

$$c_{\rm je} \;=\; \frac{C_{\rm JE}}{(1-V_{\rm B'E'}/V_{\rm JE})^{M_{\rm JE}}} \;=\; \frac{C_{\rm JE}}{(\Delta\psi/V_{\rm JE})^{M_{\rm JE}}}\;.$$

Solving for $\Delta\psi$ then gives

$$\Delta\psi \;=\; V_{\rm JE}\left(1-\frac{\Delta Q_{\rm E}}{2C_{\rm JE}V_{\rm JE}}\right)^2 \;=\; V_{\rm JE}\left(1-\frac{V_{\rm CE}-V_{\rm pt}}{V_{\rm pts}}\right)^2\;,$$

where

$$V_{\rm pts} \;=\; d_{\rm pt}\sqrt{\frac{2eV_{\rm JE}}{\epsilon}\,\frac{N_{\rm AB}N_{\rm DE}}{N_{\rm AB}+N_{\rm DE}}}\;,$$

if the eb junction is assumed to be abrupt for simplicity ($M_{\rm JE}=1/2$). As long as $V_{\rm CE}-V_{\rm pt}\ll V_{\rm pts}$, the value of $\Delta\psi$ can be approximately represented by

$$\Delta\psi \;\approx\; V_{\rm JE}\left(1-\frac{\Delta Q_{\rm E}}{C_{\rm JE}V_{\rm JE}}\right) \;=\; V_{\rm JE}\left(1-\frac{V_{\rm CE}-V_{\rm pt}}{2V_{\rm pts}}\right)\;.$$

The punchthrough current then shows an exponential increase,

$$I_{\rm pt}\sim\exp\left(\frac{V_{\rm CE}-V_{\rm pt}}{N_{\rm pt}V_{\rm T}}\right)\;,$$

with an emission coefficient $N_{\rm pt}=2V_{\rm pts}/V_{\rm JE}$.

[25] If the increase of $x_{\rm cb}-x_{\rm eb}$ with $V_{\rm CE}$ is taken into account, the charge increment is

$$\Delta Q_{\rm C} \;=\; eN_{\rm DC}\left(\sqrt{1+\frac{2\epsilon(V_{\rm CE}-V_{\rm pt})}{eN_{\rm DC}d_{\rm pt}}}-1\right)\;,$$

where $d_{\rm pt}$ denotes the value of $x_{\rm cb}-x_{\rm eb}$ at $V_{\rm CE}=V_{\rm pt}$.

[26] Taking account of the fact that the height of the potential barrier associated with the eb depletion layer in a junction biased at $V_{\rm B'E'}$ is $\Delta\psi=V_{\rm JE}-V_{\rm B'E'}$.

3.5 Charge Storage, Transit Time

The charge stored in the transistor determines the ac and switching performance, measured by quantities such as the cutoff frequency $f_\mathrm{T} = \beta f_\beta$. The cutoff frequency f_T of a bipolar transistor is shown to be related to the hole charge Q_p stored in the device by

$$\frac{1}{2\pi f_\mathrm{T}} = \left.\frac{\mathrm{d}Q_\mathrm{p}}{\mathrm{d}I_\mathrm{C}}\right|_{V_\mathrm{CE}} .$$

This formula is derived from the hole continuity equation and is valid for arbitrary transistor geometries. This allows to compute cutoff frequencies from static device simulations. For the purpose of compact modeling, the hole charge has to be separated into portions.

3.5.1 Depletion Capacitances

Under reverse bias and small forward bias, the capacitance of an abrupt pn junction is given by [58, 59]

$$c_\mathrm{j}(V') = A_\mathrm{j}\sqrt{\frac{e\epsilon N_\mathrm{D}N_\mathrm{A}}{2(N_\mathrm{D}+N_\mathrm{A})(V_\mathrm{x}-V')}} , \tag{3.72}$$

where V_x is an offset voltage that lies below the built-in voltage V_J by approximately $2V_\mathrm{T}$. A similar result for a linearly graded pn junction motivates the use of the widely employed fitting formula

$$c_\mathrm{j}(V') = C_\mathrm{J0}\,(1-V'/V_\mathrm{J})^{-M} , \tag{3.73}$$

where C_J0, V_J and M are adapted to measured data. An improved description of the depletion capacitance, suggested by Poon and Gummel [60], makes use of the approximation

$$c_\mathrm{j}(V') = \frac{C_0}{\left[(1-V'/V_\mathrm{J})^2+b\right]^{M/2}}\left(1+\frac{M}{1-M}\times\frac{b}{(1-V'/V_\mathrm{J})^2+b}\right) .$$

Here b is an adjustable parameter, which is typically positive and small compared with unity. In the limit $b \to 0$, the conventional representation is obtained. For $b \neq 0$, the parameter

$$C_0 = c_\mathrm{j}(0)\,\frac{1-M}{1+b-M}\,(1+b)^{1+M/2}$$

differs from the zero-bias capacitance. An alternative expression was suggested by de Graaff [61]

$$c_\mathrm{j}(V') = \frac{C_0}{\sqrt{(1-V'/V_\mathrm{J})^2+\delta}}\times\left\{\frac{1}{2}\left[1-\frac{V'}{V_\mathrm{J}}+\sqrt{\left(1-\frac{V'}{V_\mathrm{J}}\right)^2+\delta}\right]\right\}^{1-M} ,$$

which uses parameters C_0, M, V_J and δ that are fitted to measured $c_\mathrm{j}(V')$ curves.

3.5.2 Hole Continuity and Cutoff Frequency

The cutoff frequency f_T is related to the hole charge Q_p stored in the device by

$$\frac{1}{2\pi f_\mathrm{T}} = \left.\frac{\mathrm{d}Q_\mathrm{p}}{\mathrm{d}I_\mathrm{C}}\right|_{V_\mathrm{CE}} , \tag{3.74}$$

as is shown by the following consideration of the continuity equation for holes.[27] We distinguish between two contributions to the base current $i_\mathrm{B}(t)$ of an npn bipolar transistor: the current $i_\mathrm{rec}(t)$ due to recombination in the transistor volume and on its surface, and the rate of change with time of the hole charge $q_\mathrm{p}(t)$ in the transistor volume:[28]

$$i_\mathrm{B}(t) = i_\mathrm{rec}(t) + \mathrm{d}q_\mathrm{p}/\mathrm{d}t . \tag{3.75}$$

The corresponding small-signal relation reads

$$\underline{i}_\mathrm{b}(\omega) = \underline{i}_\mathrm{rec}(\omega) + \mathrm{j}\omega\underline{q}_\mathrm{p}(\omega) , \tag{3.76}$$

where $\underline{i}_\mathrm{b}(\omega)$, $\underline{i}_\mathrm{rec}(\omega)$ and $\underline{q}_\mathrm{p}(\omega)$ denote the phasors of the small-signal deviations of $i_\mathrm{B}(t)$, $i_\mathrm{rec}(t)$ and $q_\mathrm{p}(t)$ from their values I_B, I_rec and Q_p at the bias point. We now divide (3.76) by $\underline{i}_\mathrm{c}(\omega)$ with the result

$$\frac{\underline{i}_\mathrm{b}(\omega)}{\underline{i}_\mathrm{c}(\omega)} = \frac{\underline{i}_\mathrm{rec}(\omega)}{\underline{i}_\mathrm{c}(\omega)} + \mathrm{j}\omega\,\frac{\underline{q}_\mathrm{p}(\omega)}{\underline{i}_\mathrm{c}(\omega)} . \tag{3.77}$$

This expression equals $1/h_{21\mathrm{e}}(\omega)$ if V_CE is kept constant. To obtain ω_β, the expression on the right-hand side of (3.77) has to be expanded correctly up to first order in ω. For the first term, we obtain

$$\frac{\underline{i}_\mathrm{rec}(\omega)}{\underline{i}_\mathrm{c}(\omega)} \approx \frac{1}{\beta}\left(1 + \mathrm{j}\omega\tau_\mathrm{rec}\right) , \tag{3.78}$$

where τ_rec is an as yet undetermined delay, since $i_\mathrm{b}(0) = i_\mathrm{rec}(0)$, as can be seen from (3.76) by setting $\omega = 0$. The second term on the right-hand side of (3.77) is already of first order in ω owing to the factor $\mathrm{j}\omega$. It is therefore

[27] In this quasi-static definition, the transit time is defined as the ratio of stored charge to transfer current. It has to be emphasized that the charge transport through the base region is not strictly a time-of-flight phenomenon, owing to the diffusion mechanism; mathematically, this is a consequence of the fact that the diffusion equation possesses no traveling-wave solutions [62].

[28] This formula is equivalent to the hole continuity equation, integrated with respect to the transistor volume, if the hole current that flows through the emitter contact is treated as a surface recombination current.

sufficient to consider only the zero-order term of $\underline{q}_{\mathrm{p}}(\omega)/\underline{i}_{\mathrm{c}}(\omega)$. The resulting quantity,

$$\frac{q_{\mathrm{p}}(0)}{i_{\mathrm{c}}(0)} = \left.\frac{\mathrm{d}Q_{\mathrm{p}}}{\mathrm{d}I_{\mathrm{C}}}\right|_{V_{\mathrm{CE}}} = \tau_{\mathrm{ec}} \;, \tag{3.79}$$

is the (quasi-static) emitter–collector transit time. To first order in ω, we now have

$$\frac{1}{h_{21\mathrm{e}}^{(1)}(\omega)} = \frac{1 + \mathrm{j}\omega(\beta\tau_{\mathrm{ec}} + \tau_{\mathrm{rec}})}{\beta} = \frac{1 + \mathrm{j}\omega/\omega_{\beta}}{\beta} \;, \tag{3.80}$$

which gives the following for the cutoff frequency $f_{\mathrm{T}} = \beta\omega_{\beta}/2\pi$ defined in Sect. 1.7:

$$\frac{1}{2\pi f_{\mathrm{T}}} = \tau_{\mathrm{ec}} + \frac{\tau_{\mathrm{rec}}}{\beta} \approx \tau_{\mathrm{ec}} \;. \tag{3.81}$$

According to (3.81), a non-quasi-static correction due to the hole recombination current has to be considered. This effect is represented by the time constant τ_{rec}, which, however, is generally smaller than the emitter–collector transit time τ_{ec}. For transistors with a large value of the forward common-emitter current gain, we may therefore neglect τ_{rec} in (3.81) and compute f_{T} solely with the quasi-static emitter–collector transit time τ_{ec}. This approximation becomes inaccurate if the current gain becomes small, as is, for example, the case in the high-current regime. In this case the non-quasi-static correction $\tau_{\mathrm{rec}}/\beta$ to the quasi-static transit time becomes relevant.

Equation (3.81) is directly derived from the hole continuity equation and is therefore completely general. It does not rely on any type of quasi-static approximation, but shows the relation between f_{T} and τ_{ec} to be a consequence of the one-pole approximation used in the definition of the cutoff frequency. The expression for τ_{ec} can be split up into three terms:[29]

$$\frac{\mathrm{d}Q_{\mathrm{p}}}{\mathrm{d}I_{\mathrm{C}}} = \frac{\mathrm{d}Q_{\mathrm{JE}}}{\mathrm{d}I_{\mathrm{C}}} + \frac{\mathrm{d}Q_{\mathrm{JC}}}{\mathrm{d}I_{\mathrm{C}}} + \frac{\mathrm{d}Q_{\mathrm{T}}}{\mathrm{d}I_{\mathrm{C}}} \;, \tag{3.82}$$

where Q_{JE} and Q_{JC} denote the hole charges associated with the emitter and collector depletion layers, and Q_{T} is the hole charge stored in the quasi-neutral regions (e, b, c) together with the neutralized charge of mobile carriers in the space charge layers. The term

$$\mathrm{d}Q_{\mathrm{T}}/\mathrm{d}I_{\mathrm{C}} = \tau_{\mathrm{f}} \tag{3.83}$$

is commonly referred to as the forward transit time. We write the following for the (total) derivatives of the space charges Q_{JE} and Q_{JC} with respect to I_{C}:

[29] Such a decomposition does not necessarily rely on the depletion approximation, as was demonstrated by van den Biesen [63].

$$\frac{dQ_{\rm JE}}{dI_{\rm C}} = \left(\frac{\partial Q_{\rm JE}}{\partial I_{\rm C}}\right)_{V_{\rm B'E'}} + \left(\frac{\partial Q_{\rm JE}}{\partial V_{\rm B'E'}}\right)_{I_{\rm C}} \frac{dV_{\rm B'E'}}{dI_{\rm C}} \,, \tag{3.84}$$

$$\frac{dQ_{\rm JC}}{dI_{\rm C}} = \left(\frac{\partial Q_{\rm JC}}{\partial I_{\rm C}}\right)_{V_{\rm B'C'}} + \left(\frac{\partial Q_{\rm JC}}{\partial V_{\rm B'C'}}\right)_{I_{\rm C}} \frac{dV_{\rm B'C'}}{dI_{\rm C}} \,. \tag{3.85}$$

The quantities

$$\left(\frac{\partial Q_{\rm JE}}{\partial V_{\rm B'E'}}\right)_{I_{\rm C}} = c_{\rm je} \quad \text{and} \quad \left(\frac{\partial Q_{\rm JC}}{\partial V_{\rm B'C'}}\right)_{I_{\rm C}} = c_{\rm jc} \tag{3.86}$$

are the emitter and collector depletion capacitances. The extra terms in (3.84) and (3.85) take account of the effect of the mobile carriers on the space charge layers, which cause the depletion capacitances to be a function of the collector current $I_{\rm C}$. The quantity

$$(\partial Q_{\rm JC}/\partial I_{\rm C})_{V_{\rm B'C'}} = \tau_{\rm jc} \tag{3.87}$$

is called the collector transit time. This term can be estimated with the help of the depletion approximation, which allows (3.87) to be written as

$$\tau_{\rm jc} = (\partial Q_{\rm JC}/\partial I_{\rm C})_{V_{\rm B'C'}} = -A_{\rm je}\rho(x_{\rm bc})\,(\partial x_{\rm bc}/\partial I_{\rm C})_{V_{\rm B'C'}} \,. \tag{3.88}$$

To calculate $\tau_{\rm jc}$, the neutrality condition and the Poisson equation for the bc depletion layer are considered; in the depletion approximation, these equations read

$$\int_{x_{\rm bc}}^{x_{\rm cb}} \rho(x)\,dx = 0 \tag{3.89}$$

and

$$\int_{x_{\rm bc}}^{x_{\rm cb}} x\rho(x)\,dx = \epsilon(V_{\rm C'B'}+V_{\rm JC}) \,. \tag{3.90}$$

Differentiation of (3.89) and (3.90) with respect to $I_{\rm C}$ gives

$$0 = \rho(x_{\rm cb})\,\frac{\partial x_{\rm cb}}{\partial I_{\rm C}} - \rho(x_{\rm bc})\,\frac{\partial x_{\rm bc}}{\partial I_{\rm C}} + \int_{x_{\rm bc}}^{x_{\rm cb}} \frac{\partial \rho}{\partial I_{\rm C}}\,dx \,, \tag{3.91}$$

$$0 = x_{\rm cb}\rho(x_{\rm cb})\frac{\partial x_{\rm cb}}{\partial I_{\rm C}} - x_{\rm bc}\rho(x_{\rm bc})\frac{\partial x_{\rm bc}}{\partial I_{\rm C}} + \int_{x_{\rm bc}}^{x_{\rm cb}} x\,\frac{\partial \rho}{\partial I_{\rm C}}\,dx \,. \tag{3.92}$$

We multiply (3.91) by $x_{\rm cb}$ and subtract the resulting equation from (3.92) to obtain

$$(x_{\rm cb}-x_{\rm bc})\rho(x_{\rm bc})\frac{\partial x_{\rm bc}}{\partial I_{\rm C}} = \int_{x_{\rm bc}}^{x_{\rm cb}} (x_{\rm cb}-x)\,\frac{\partial \rho}{\partial I_{\rm C}}\,dx \,. \tag{3.93}$$

The space charge density in the depletion layer is assumed to be of the form

$$\rho(x) \approx e(N_{\rm D}^{+} - N_{\rm A}^{-}) - \frac{I_{\rm C}}{v_{\rm n}(x)A_{\rm je}} . \tag{3.94}$$

Allowing for a spatial dependence of the electron drift velocity $v_{\rm n}(x)$ within the bc depletion layer gives $A_{\rm je}\partial\rho/\partial i_{\rm C} = -1/v_{\rm n}(x)$. Together with (3.94), this gives the following result for the collector transit time:

$$\tau_{\rm jc} = \int_{x_{\rm bc}}^{x_{\rm cb}} \frac{x_{\rm cb} - x}{d_{\rm jc}} \frac{1}{v_{\rm n}(x)} \, {\rm d}x , \tag{3.95}$$

where $d_{\rm jc} = x_{\rm cb} - x_{\rm bc}$ denotes the bc depletion-layer width. If the electron drift velocity in the bc depletion layer is equal to the saturated drift velocity $v_{\rm nsat}$, (3.95) simplifies to

$$\tau_{\rm jc} = d_{\rm jc}/2v_{\rm nsat} , \tag{3.96}$$

in accordance with the outcome of the ac formulation for the total current passing through the depletion layer [64]. The quantity

$$(\partial Q_{\rm JE}/\partial I_{\rm C})_{V_{\rm B'E'}} = \tau_{\rm je} \tag{3.97}$$

is negligible since $I_{\rm C}$ is predominantly a function of $V_{\rm B'E'}$ in the case of forward operation. Setting ${\rm d}V_{\rm B'E'}/{\rm d}I_{\rm C} = 1/g_{\rm m}$, we obtain the following from $V_{\rm B'C'} = V_{\rm B'E'} + R_{\rm CC'}I_{\rm C} + R_{\rm EE'}I_{\rm E} - V_{\rm CE}$ using the condition $V_{\rm CE} = $ const.:

$$\frac{{\rm d}V_{\rm B'C'}}{{\rm d}I_{\rm C}} = \frac{1}{g_{\rm m}} + r_{\rm cc'} + r_{\rm ee'}\frac{{\rm d}I_{\rm E}}{{\rm d}I_{\rm C}} \approx \frac{1}{g_{\rm m}} + r_{\rm cc'} + r_{\rm ee'} , \tag{3.98}$$

where $r_{\rm cc'} = R_{\rm CC'} + I_{\rm C}({\rm d}R_{\rm CC'}/{\rm d}I_{\rm C})$ and $r_{\rm ee'} = R_{\rm EE'} + I_{\rm E}({\rm d}R_{\rm EE'}/{\rm d}I_{\rm E})$ denote the small-signal collector and emitter series resistances, respectively. We neglect $\tau_{\rm je}$ and the Early effect and combine (3.82), (3.83), (3.84) and (3.85) to obtain the following relation for the emitter-collector transit time:

$$\tau_{\rm ec} = \frac{c_{\rm je} + c_{\rm jc}}{g_{\rm m}} + \tau_{\rm f} + \tau_{\rm jc} + (r_{\rm ee'} + r_{\rm cc'})c_{\rm jc} . \tag{3.99}$$

Values for f_β and $f_{\rm T}$ may alternatively be derived from the Giacoletto small-signal equivalent circuit of the bipolar transistor, which gives, under the assumption $g_{\rm o} = 0$ (Sect. 1.7)

$$\frac{1}{2\pi f_{\rm T}} = \frac{c_\pi + c_\mu}{g_{\rm m}} + (r_{\rm ee'} + r_{\rm cc'})c_\mu . \tag{3.100}$$

A comparison of (3.100) and (3.99) shows that we have to make the identifications

$$c_\pi = c_{\rm je} + (\tau_{\rm f} + \tau_{\rm jc})g_{\rm m} , \tag{3.101}$$

$$c_\mu = c_{\rm jc} , \tag{3.102}$$

i.e., the forward transit time used in the conventional small-signal model (Sects. 1.7 and 3.9) comprises both the forward transit time τ_f, defined as $\mathrm{d}Q_\mathrm{T}/\mathrm{d}I_\mathrm{C}$, and the collector transit time τ_jc.

3.5.3 Forward Transit Time

If the bc junction is reverse biased, minority-carrier storage in the collector region may be neglected. In this case the forward transit time τ_f is the sum of the forward base transit time τ_Bf, the emitter transit time τ_E and the emitter–base transit time τ_BE.

Base Transit Time

The (quasi-static) forward base transit time τ_Bf, due to minority-carrier storage in the base region, may be estimated using the formula of Krömer [65],

$$\tau_\mathrm{Bf} = \int_{x_\mathrm{be}}^{x_\mathrm{bc}} \frac{n_\mathrm{ie}^2(x)}{p(x)} \int_x^{x_\mathrm{bc}} \frac{p(y)}{n_\mathrm{ie}^2(y)} \frac{1}{D_\mathrm{n}(y)}\,\mathrm{d}y\,\mathrm{d}x\ , \tag{3.103}$$

who generalized an earlier result of Moll and Ross [2]. This formula is based on the electron and hole transport equations in the drift–diffusion approximation. In the derivation of (3.103), the electron density $n(x_\mathrm{bc})$ was taken to be zero. Velocity saturation in the bc space charge layer, however, causes this assumption to be somewhat erroneous [66, 67]. In [68], therefore, $J_\mathrm{n} = -en(x_\mathrm{bc})v_\mathrm{n}(x_\mathrm{bc})$, where $v_\mathrm{n}(x_\mathrm{bc}) = v_\mathrm{nsat}$, was used for the electron density at the space-charge-layer boundary, with the result

$$\begin{aligned}\tau_\mathrm{Bf} &= \int_{x_\mathrm{be}}^{x_\mathrm{bc}} \frac{n_\mathrm{ie}^2(x)}{p(x)} \int_x^{x_\mathrm{bc}} \frac{p(y)}{n_\mathrm{ie}^2(y)} \frac{1}{D_\mathrm{n}(y)}\,\mathrm{d}y\,\mathrm{d}x \\ &\quad + \frac{1}{v_\mathrm{nsat}} \frac{p(x_\mathrm{bc})}{n_\mathrm{ie}^2(x_\mathrm{bc})} \int_{x_\mathrm{be}}^{x_\mathrm{bc}} \frac{n_\mathrm{ie}^2(x)}{p(x)}\,\mathrm{d}x\ .\end{aligned} \tag{3.104}$$

Under low-level-injection conditions, $p(x) \approx N_\mathrm{A}^-(x)$; (3.104) then describes the base transit time in terms of local doping and material parameters. In the case of a homogeneously doped base region, (3.104) reduces to [66]

$$\tau_\mathrm{Bf} = \frac{d_\mathrm{B}^2}{2D_\mathrm{n}} + \frac{d_\mathrm{B}}{v_\mathrm{nsat}}\ . \tag{3.105}$$

The increase in τ_Bf caused by finite values of v_nsat is explained in Fig. 3.16, which shows schematically the minority-carrier distribution in a box-profiled base region. An increase of the electron density at x_bc causes both a decrease of the electron current density J_n due to the reduced slope of the minority-carrier distribution and an increase of the stored minority charge ΔQ_nB, and therefore affects the ratio $|\Delta Q_\mathrm{nB}|/|J_\mathrm{n}|$, which determines the base transit

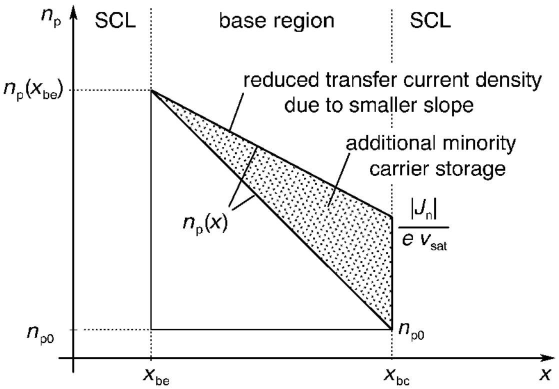


Fig. 3.16. Minority-carrier profile in the base region with and without velocity saturation in the bc space-charge layer (SCL)

time. Equation (3.105) can be derived from Fig. 3.16 by a simple geometrical analysis. Analytical expressions for the base transit time that take account of the field and dopant concentration dependence of the minority-carrier mobility in the base region were derived in [69].

Diffusion in a Short Base. The magnitude of the mean electron velocity due to diffusion is

$$v_n = \frac{D_n}{n} \left| \frac{dn}{dx} \right| ,$$

according to the widely employed drift–diffusion approximation. If D_n is considered to be a constant, the computed value of v_n can assume values in excess of v_{nsat} if $|dn/dx|$ becomes very large. Obviously, this mathematical limit is in contradiction to physical reality: since diffusion is an effect of the thermal motion of carriers, the classical diffusion equation can only hold if v_n is small in comparison with the thermal velocity $v_{th} = \sqrt{k_B T / 2\pi m_n^*}$ of the carriers. Assuming stationary conditions, unidirectional current flow ($\boldsymbol{v}_n = v_n \boldsymbol{e}_x$) and $E = 0$, the momentum balance equation reads

$$n v_n \frac{dv_n}{dx} + \frac{k_B}{m_n^*} T_n \frac{dn}{dx} + \frac{k_B}{m_n^*} n \frac{dT_n}{dx} = -\frac{n v_n}{\tau_{vn}} .$$

Taking account of the relations $J_n = -env_n$ and $\mu_n = e\tau_{vn}/m_n^*$, this equation is easily transformed into

$$J_n = e\mu_n V_T \frac{dn}{dx} + \mu_n n k_B \frac{dT_n}{dx} - \mu_n m_n^* \frac{J_n^2}{e^2 n^2} \frac{dn}{dx} ,$$

since $J_n = \text{const.}$ and

$$n \frac{dv_n}{dx} + v_n \frac{dn}{dx} = 0$$

according to the continuity equation. If the spatial variation of the electron temperature is assumed to be negligible ($\mathrm{d}T_\mathrm{n}/\mathrm{d}x \approx 0$), the quadratic relation

$$J_\mathrm{n}^2 + \frac{e^2}{\mu_\mathrm{n} m_\mathrm{n}^*} \frac{n^2}{\mathrm{d}n/\mathrm{d}x} J_\mathrm{n} - \frac{e^2 k_\mathrm{B} T}{m_\mathrm{n}^*} n^2 = 0$$

is obtained for the electron current density. This equation comprises two limiting situations. At small and moderate values of $\mathrm{d}n/\mathrm{d}x$, the second term is larger than the first term, resulting in the well-known expression for the diffusion current density

$$J_\mathrm{n} = eV_\mathrm{T} \mu_\mathrm{n} \frac{\mathrm{d}n}{\mathrm{d}x} ,$$

whereas, in the case of a large gradient of the electron density, the first term dominates, resulting in

$$J_\mathrm{n} = -en\sqrt{k_\mathrm{B} T_\mathrm{n}/m_\mathrm{n}^*} ,$$

i.e. the average electron velocity is limited to a value of the order of the thermal velocity.

In [70], a numerical solution for narrow-base transistors is reported. While the average electron velocity was found to be somewhat smaller within the base layer as compared with the results of the drift–diffusion theory, a velocity overshoot of the order of 80% was found to occur at the metallurgical junction of the bc diode. The simulation results for the base transit time were compared with results of the drift–diffusion theory, and relative errors of 20% for a 20 nm base width were found.

An investigation of the electron transport through a thin base layer based on an analytical solution of the simplified Boltzmann equation is described in [71, 72]. According to this work, the minority-carrier transport through a thin, homogeneously doped base layer under forward operation conditions still obeys the diffusion equation, i.e.

$$J_\mathrm{n} \sim \frac{n_\mathrm{p}(x_\mathrm{be}) - n_\mathrm{p}(x_\mathrm{bc})}{d_\mathrm{B}} ,$$

if the minority-carrier densities at the depletion-layer boundaries are calculated from

$$n_\mathrm{p}(x_\mathrm{be}) = \frac{2 + 3d_\mathrm{B}/\lambda}{4 + 3d_\mathrm{B}/\lambda} n_\mathrm{BE}^* \quad \text{and} \quad n_\mathrm{p}(x_\mathrm{bc}) = \frac{2}{4+3d_\mathrm{B}/\lambda} n_\mathrm{BE}^* ,$$

where $n_\mathrm{BE}^* = n_\mathrm{p0} \exp(V_\mathrm{B'E'}/V_\mathrm{T})$ denotes the minority-carrier density determined from the Shockley boundary condition and λ is the carrier mean free path. In the limit $d_\mathrm{B}/\lambda \to \infty$, the modified boundary conditions obviously reduce to the classical Shockley boundary conditions under low-level-injection conditions. The forward base transit time τ_Bf is approximated by [72]

$$\tau_{\mathrm{Bf}} = \frac{d_{\mathrm{B}}^2}{2D_{\mathrm{n}}} \left(1 + \frac{4\lambda}{3d_{\mathrm{B}}}\right) ,$$

which, in the limit $\lambda \ll d_{\mathrm{B}}$, reduces to the classical result obtained from elementary diffusion theory.

Emitter Transit Time

A formula similar to (3.103) for the (quasi-static) emitter transit time τ_{E} of metal-contacted emitter regions has been published in [73]:

$$\tau_{\mathrm{E}} = \frac{1}{\beta} \int_{x_{\mathrm{e}}}^{x_{\mathrm{eb}}} \int_{x_{\mathrm{e}}}^{x} \frac{n_{\mathrm{ie}}^2(x)}{n(x)} \frac{n(y)}{n_{\mathrm{ie}}^2(y)} \frac{1}{D_{\mathrm{p}}(y)} \,\mathrm{d}y \,\mathrm{d}x \,. \tag{3.106}$$

This formula neglects recombination in the emitter region and therefore applies only to shallow emitter junctions. In order to describe modern bipolar transistors with nearly transparent emitters and polysilicon contacts, a finite recombination velocity S_{nn} at the polycrystalline emitter contact is assumed.[30] For such devices, the emitter transit time is described by a formula similar to (3.104) if recombination in the emitter region is neglected:

$$\begin{aligned} \tau_{\mathrm{E}} &= \frac{1}{\beta} \int_{x_{\mathrm{e}}}^{x_{\mathrm{eb}}} \frac{n_{\mathrm{ie}}^2(x)}{n(x)} \int_{x_{\mathrm{e}}}^{x} \frac{n(y)}{n_{\mathrm{ie}}^2(y)} \frac{1}{D_{\mathrm{p}}(y)} \,\mathrm{d}y \,\mathrm{d}x \\ &\quad + \frac{1}{\beta S_{\mathrm{nn}}} \frac{n(x_{\mathrm{e}})}{n_{\mathrm{ie}}^2(x_{\mathrm{e}})} \int_{x_{\mathrm{e}}}^{x_{\mathrm{eb}}} \frac{n_{\mathrm{ie}}^2(x)}{n(x)} \,\mathrm{d}x \,; \end{aligned} \tag{3.107}$$

this formula applies if $p(x_{\mathrm{e}}) \gg p_{\mathrm{n0}}$, as is the case under forward operation conditions. Recombination in the emitter region causes a slight modification of the emitter transit time τ_{E}, which, however, is generally small, since most of the recombination takes place at the emitter contact. The effect of the emitter doping gradient on τ_{E} and f_{T} was investigated in [74], where increased values of the cutoff frequency observed for transistors with arsenic emitters were explained by their steeper doping profiles.

Emitter–Base Transit Time

In a forward-biased diode, both electrons and holes are injected into the space charge layer, where they neutralize each other in part and thus provide a con-

[30] In this work, the recombination velocity S_{nn} at the interface is taken to be frequency-independent, i.e. we assume a proportionality between the hole current density orthogonal to the interface and the excess hole density Δp there, $\underline{J}_{\mathrm{p}}(x_{\mathrm{e}}, \omega) = qS_{\mathrm{nn}}\Delta\underline{p}(x_{\mathrm{e}})$, which is valid for all values of the angular frequency ω. This assumption is not well founded in the high-frequency regime, where the time constants associated with the details of the recombination process become relevant. However, since the effect of the recombination delay generally represents a small correction to the cutoff frequency, the assumption $S_{\mathrm{nn}} = \mathrm{const.}$ is justified when one is estimating the effect.

tribution to the diffusion charge. The corresponding capacitance component is sometimes called the neutral capacitance [63, 75]. For a proper distinction

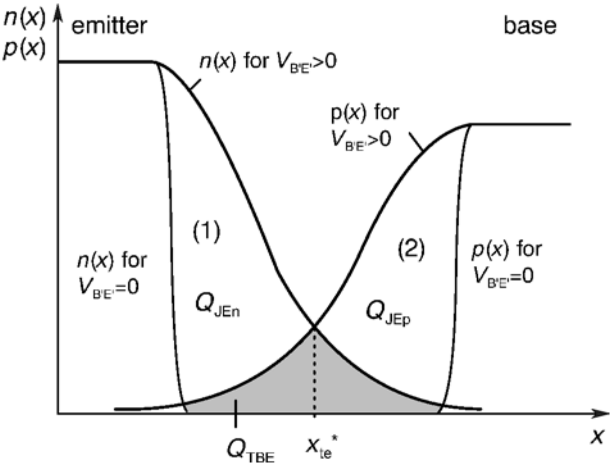


Fig. 3.17. Neutralized base charge $Q_{\rm TBE}$ in the depletion layer defined in terms of the electron and hole current densities (schematic illustration)

between the depletion charge and the diffusion charge (neutralized charge), we can consider Fig. 3.17, which illustrates the changes of the electron and hole densities associated with a forward bias of the eb junction. Additional electrons (1) and holes (2), which are not neutralized, form the electron and hole depletion charges $Q_{\rm JEn}$ and $Q_{\rm JEp}$. The neutralized charge $Q_{\rm TBE}$ is determined by the excess hole density where electrons are in the majority (i.e. for $x < x_{\rm tc}^*$) and by the excess electron density where holes are in the majority (i.e. for $x > x_{\rm te}^*$). At $x = x_{\rm te}^*$ the values of the excess electron and hole densities are equal, and therefore

$$\Delta p(x_{\rm te}^*) \;=\; \Delta n(x_{\rm te}^*) \;=\; n_{\rm ie}\left[\exp\left(\frac{v_{\rm B'E'}}{2V_{\rm T}}\right) - 1\right]$$

if the electron and hole quasi-Fermi potentials are assumed to be constant throughout the depletion layer. Estimating $Q_{\rm TBE}$ from the area of the triangle of height $\Delta p(x_{\rm te}^*)$ and base $2V_{\rm T}/|E(x_{\rm te}^*)|$ formed by the tangents to the $p(x)$ and $n(x)$ curves yields

$$Q_{\rm TBE} \;\approx\; eA_{\rm je}\,\frac{V_{\rm T}\Delta p(x_{\rm te}^*)}{|E_{\rm max}|} \;\sim\; \exp\left(\frac{v_{\rm B'E'}}{2V_{\rm T}}\right) \tag{3.108}$$

if $E(x_{\rm tc}^*)$ is approximated by $E_{\rm max}$. The corresponding transit time $\tau_{\rm BE}$ derived from $\mathrm{d}Q_{\rm TBE}/\mathrm{d}v_{\rm B'E'} = \tau_{\rm BE}g_{\rm m}$ varies in inverse proportion to $\sqrt{I_{\rm C}}$ under low-level-injection conditions, when $I_{\rm C}$ varies in proportion to $\exp(v_{\rm B'E'}/V_{\rm T})$.

3.6 Series Resistances

Series resistances cause deviations from exponential current–voltage characteristics and slow down charging of the transistor capacitances. The base resistance and collector resistance have to be modeled as bias-dependent elements owing to conductivity modulation and current crowding.

3.6.1 Emitter Resistance

The emitter resistance $R_{\mathrm{EE'}}$ can generally be described as ohmic.[31] Owing to the shallow eb junction, the value of $R_{\mathrm{EE'}}$ is determined almost exclusively by the value of the specific contact resistance ρ_{c}:

$$R_{\mathrm{EE'}} \approx \rho_{\mathrm{c}}/A_{\mathrm{je}} .$$

Owing to the small value of the emitter contact area A_{je}, integrated transistors may show values of $R_{\mathrm{EE'}}$ of several tens of ohms. It is important to keep the emitter series resistance low, since this parameter significantly degrades the transconductance of the device.

3.6.2 Base Resistance

The base resistance $R_{\mathrm{BB'}}$ describes the voltage drop caused by the current in the base region and at the base contact. In planar transistors $R_{\mathrm{BB'}}$ may be split up into an external part, which describes the contact resistance and the resistance of the external base region, and an internal part. The external part behaves approximately like an ohmic resistor, with a value determined by the doping and geometry of the external base region and the base contact resistance. The internal part shows a considerable bias dependence and decreases with increasing base current for the following two reasons. (1) An increase in the transfer current causes an increase in the hole diffusion charge in the base region and therefore decreases the sheet resistance of the base layer (conductivity modulation). (2) The voltage drop across the internal base region increases with increasing base current, resulting in emitter current crowding: the base current then flows predominantly at the emitter edge, reducing the distance that has to be traveled by the holes and therefore the internal base resistance.

For planar transistors with a stripe geometry ($W_{\mathrm{E}} \gg L_{\mathrm{E}}$), a one-dimensional description where the base current flows in the y direction can be applied (Fig. 3.18a). Assuming an ideal eb diode with a saturation current $I_{\mathrm{S}}/B_{\mathrm{F}}$ and an

[31] An increase of the emitter resistance in the presence of current crowding was demonstrated in [76] using a partitioned equivalent circuit; the emitter resistance was calculated by considering power dissipation, a method that does not directly yield the effect on the current–voltage characteristics, as is explained for the case of the base resistance.

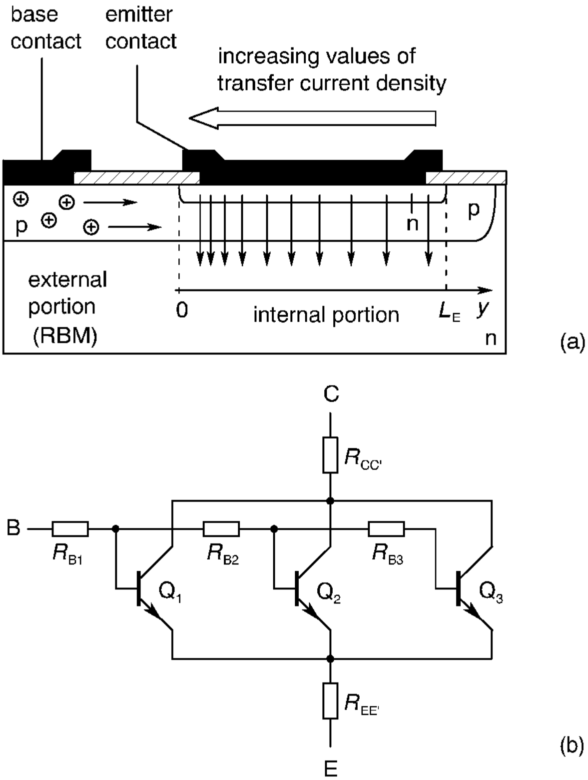


Fig. 3.18. (**a**) Emitter current crowding under dc operating conditions, and (**b**) extended equivalent-circuit model

emission coefficient $N = 1$, the decrease of the base current per unit length due to hole injection into the emitter is

$$\frac{\mathrm{d}I_\mathrm{B}(y)}{\mathrm{d}y} = -\frac{I_\mathrm{S}}{B_\mathrm{F} L_\mathrm{E}} \left[\exp\left(\frac{V_{\mathrm{B'E'}}(y)}{V_\mathrm{T}} \right) - 1 \right] . \tag{3.109}$$

The local voltage across the eb junction $V_{\mathrm{B'E'}}(y)$ is a function of the lateral position y (Fig. 3.18a) owing to the voltage drop across the base layer; if the sheet resistance of the active base layer is R_π, one obtains

$$\frac{\mathrm{d}V_{\mathrm{B'E'}}(y)}{\mathrm{d}y} = -\frac{R_\pi}{W_\mathrm{E}} I_\mathrm{B}(y) .$$

Differentiation of (3.109) allows one to combine the two equations into a second-order differential equation that can be solved analytically [77,78]. The general solution is

$$I_\mathrm{B}(y) = \frac{2V_\mathrm{T}}{3R_\mathrm{B}} z \tan\left[z \left(1 - \frac{y}{L_\mathrm{E}} \right) \right] , \qquad R_\mathrm{B} = \frac{R_\pi}{3} \frac{L_\mathrm{E}}{W_\mathrm{E}} . \tag{3.110}$$

With this result, integration of (3.109) gives

$$V_{\mathrm{B'E'}}(y) \;=\; V_{\mathrm{B'E'}}(0) - 2V_{\mathrm{T}} \ln \frac{\cos\left[\, z(1 - y/L_{\mathrm{E}})\,\right]}{\cos z} , \tag{3.111}$$

where z has to be determined from the boundary condition $I_{\mathrm{B}}(y=0) = I_{\mathrm{B}}$

$$z \tan(z) \;=\; \frac{3I_{\mathrm{B}}R_{\mathrm{B}}}{2V_{\mathrm{T}}} . \tag{3.112}$$

several approaches can now be used to calculate the base resistance [78, 79]. The approach used in SPICE is a power dissipation calculation, that considers the power dissipated in the base region, which leads to the definition

$$P \;=\; R_{\mathrm{BB'}}^{(P)} I_{\mathrm{B}}^2 \;=\; \frac{1}{W_{\mathrm{E}}} \int_0^{L_{\mathrm{E}}} R_\pi I_{\mathrm{B}}(y)^2 \, \mathrm{d}y . \tag{3.113}$$

Since $I_{\mathrm{B}}(y)/I_{\mathrm{B}} = \tan\left[\, z\,(1 - y/L_{\mathrm{E}})\right] / \tan z$, this gives

$$R_{\mathrm{BB'}}^{(P)} \;=\; 3R_{\mathrm{B}} \frac{1}{\tan^2(z)} \int_0^1 \tan^2(\lambda z)\, \mathrm{d}\lambda \;=\; 3R_{\mathrm{B}} \frac{\tan(z) - z}{z \tan^2(z)} , \tag{3.114}$$

where z is determined by (3.112). In the limit of small base current ($I_{\mathrm{B}} \to 0$ and therefore $z \to 0$), the value of the base resistance $R_{\mathrm{BB'}}$ converges to R_{B}, in agreement with the elementary estimate (1.114). SPICE uses $R_{\mathrm{BB'}}^{(P)}$ for the description of the internal base resistance and adds an ohmic term R_{BM} to take account of the external base resistance:

$$R_{\mathrm{BB'}} \;=\; R_{\mathrm{BM}} + R_{\mathrm{B}}\theta(z) , \qquad \text{where} \qquad \theta(z) = 3\, \frac{\tan(z) - z}{z \tan^2(z)} ; \tag{3.115}$$

the value of z can be approximately calculated as a function of the base current I_{B} from [83]

$$z \;=\; \frac{\pi^2}{24} \left(\sqrt{\frac{I_{\mathrm{RB}}}{I_{\mathrm{B}}} + \frac{144}{\pi^2}} - \sqrt{\frac{I_{\mathrm{RB}}}{I_{\mathrm{B}}}} \right) , \tag{3.116}$$

where I_{RB} is the current where the base resistance falls halfway to its minimum value. For small values of base current, $\theta(z) \approx 1$ and the base resistance assumes its maximum value,

$$R_{\mathrm{BB'}} \;\approx\; R_{\mathrm{BM}} + R_{\mathrm{B}} .$$

The power dissipation calculation performed here does not give the dc lumped base resistance, which is defined in terms of the input characteristics of the bipolar transistor by the equation

$$I_{\mathrm{B}} \;=\; \frac{I_{\mathrm{S}}}{B_{\mathrm{F}}} \exp\!\left(\frac{V_{\mathrm{B'E'}}(0) - R_{\mathrm{BB'}} I_{\mathrm{B}}}{V_{\mathrm{T}}} \right) . \tag{3.117}$$

Using

$$I_\mathrm{B} = \frac{I_\mathrm{S}}{B_\mathrm{F}} \exp\left(\frac{V_\mathrm{B'E'}(0)}{V_\mathrm{T}}\right) \frac{1}{L_\mathrm{E}} \int_0^{L_\mathrm{E}} \exp\left(\frac{V_\mathrm{B'E'}(y) - V_\mathrm{B'E'}(0)}{V_\mathrm{T}}\right) \mathrm{d}y \,,$$

we obtain the following definition:

$$R_\mathrm{BB'} = -\frac{V_\mathrm{T}}{I_\mathrm{B}} \ln\left(\frac{1}{L} \int_0^{L_\mathrm{E}} \exp\left(\frac{V_\mathrm{B'E'}(y) - V_\mathrm{B'E'}(0)}{V_\mathrm{T}}\right) \mathrm{d}y\right) .$$

With the help of (3.111), the integral can be evaluated to give [79]

$$R_\mathrm{BB'} = 3R_\mathrm{B} \frac{1}{2z\tan(z)} \ln\left(\frac{z}{\sin(z)\cos(z)}\right) . \tag{3.118}$$

This value of the base resistance could be termed the large-signal dc base resistance. Its value differs from the result of the power dissipation calculation. Yet another expression for the base resistance is obtained if the low-frequency small-signal base resistance $r_\mathrm{bb'}$ is calculated from (3.118) according to

$$r_\mathrm{bb'} = R_\mathrm{BB'} + I_\mathrm{B} \frac{\mathrm{d}R_\mathrm{BB'}}{\mathrm{d}I_\mathrm{B}} .$$

Making use of (3.112), it is possible to express I_B in terms of z and $\mathrm{d}z/\mathrm{d}I_\mathrm{B}$, leading to

$$I_\mathrm{B} \frac{\mathrm{d}R_\mathrm{BB'}}{\mathrm{d}I_\mathrm{B}} = I_\mathrm{B} \frac{\mathrm{d}R_\mathrm{BB'}}{\mathrm{d}z} \frac{\mathrm{d}z}{\mathrm{d}I_\mathrm{B}} = \frac{z\tan(z)}{\tan(z) + z/\cos^2(z)} \frac{\mathrm{d}R_\mathrm{BB'}}{\mathrm{d}z} .$$

Computation of $\mathrm{d}R_\mathrm{BB'}/\mathrm{d}z$ and rearranging terms results in

$$r_\mathrm{bb'} = \frac{3R_\mathrm{B}}{2} \frac{\cot(z) + z\,[1 - \tan(z)]}{z^2} . \tag{3.119}$$

The concept of a lumped base series resistance has to be considered as a rather crude approximation to reality. Furthermore, the following should be noted. (1) The approach considered here does not take conductivity modulation in the base region into account; however, this has a substantial effect on the base series resistance. (2) The analysis performed here assumes that the base current flow is strictly one-dimensional in the y direction, an assumption that is approximately fulfilled in stripe transistors with a large $W_\mathrm{E}/L_\mathrm{E}$ ratio.[32] (3) During switching transients, the distribution of the transfer current density across the base region becomes time-dependent. The problems listed above can be overcome by an increase in the complexity of the model if an equivalent circuit is used in which several transistors are connected in parallel.

[32]The computation of a single lumped base series resistance from layout data is discussed, for example, in [80–82]; the special case of a circular base region is considered also in Appendix D.

3.6.3 Collector Resistance, Quasi-Saturation

The collector resistance generally shows nonohmic behavior, since the bias-dependent extent of the bc space charge region, minority-carrier injection into the epitaxial collector region and velocity saturation effects influence this parameter. This resistance is represented as the sum of an external part R_{CX} due to the contact and the subcollector, and the resistance

$$R_{\mathrm{epi}} \approx \frac{d_{\mathrm{epi}} - x_{\mathrm{n}}}{e\mu_{\mathrm{n}} A_{\mathrm{je}} N_{\mathrm{Depi}}}$$

of the undepleted epilayer (see Fig. 3.19) with a concentration N_{Depi}, assumed to be independent of position for simplicity. The value of R_{epi} vanishes when the edge of the depletion layer reaches the buried layer ($x_{\mathrm{n}} \approx d_{\mathrm{epi}}$).

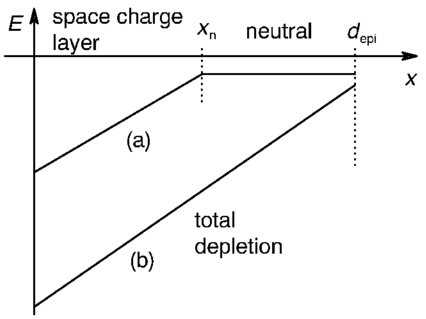


Fig. 3.19. Electric-field distribution in the epitaxial layer for transfer current values smaller than I_1 and (**a**) a small reverse bias V_{CB} resulting in partial depletion of the epilayer, and (**b**) for a large reverse bias V_{CB} resulting in total depletion of the epilayer

The extent x_{n} depends on the internal junction voltage $V_{\mathrm{B'C'}}$ and the space charge density in the depletion layer. The latter is modified in the presence of a transfer current, since the electrons crossing the space charge layer will move on average at their saturated drift velocity v_{nsat}. This results in a reduction of the space charge density of the epitaxial layer to $\rho(x) = eN_{\mathrm{Depi}}\,(1-I_{\mathrm{T}}/I_1)$, where $I_1 = eA_{\mathrm{je}}v_{\mathrm{nsat}}N_{\mathrm{Depi}}$ denotes the Kirk current. Assuming a one-sided junction with a space charge layer that extends almost exclusively into the epilayer, the depletion-layer width $d_{\mathrm{jc}} \approx x_{\mathrm{n}}$ is given by [84]

$$d_{\mathrm{jc}} \approx \sqrt{\frac{2\epsilon V_{\mathrm{JC}}}{eN_{\mathrm{Depi}}}}\sqrt{\frac{1-V_{\mathrm{B'C'}}/V_{\mathrm{JC}}}{1-I_{\mathrm{T}}/I_1}}\,. \tag{3.120}$$

Since $V_{\mathrm{B'C'}} \approx V_{\mathrm{BC}}+(R_{\mathrm{CX}}+R_{\mathrm{epi}})I_{\mathrm{T}}$, the voltage drop across R_{CX} and R_{epi} acts like a forward bias and $V_{\mathrm{B'C'}} > 0$ if $I_{\mathrm{T}} > V_{\mathrm{CB}}/(R_{\mathrm{epi}} + R_{\mathrm{CX}})$. This situation, where the bc diode becomes internally forward biased in spite of an external reverse bias, is known as quasi-saturation and will be discussed further in the following subsection together with the Kirk effect, which occurs if the transfer current increases beyond the critical current I_1.

3.7 High-Level Injection

The current gain and frequency response of a bipolar transistor are a function of the bias current and are rather poor at small current levels. High-speed applications therefore generally require substantial electron current densities. Owing to high-level-injection effects, however, the device speed is not a monotonic function of the current density transported by the device, and decreases with increasing electron current density after passing through a maximum. A thorough understanding of high-level-injection effects is therefore mandatory in order to design optimized high-frequency bipolar transistors and to develop circuits that make the best use of the potentialities of a given technology.

Besides the effects of series resistances considered in the preceding section, current–voltage characteristics and transit times are modified under high-level injection owing to deviations from the Shockley boundary conditions, alteration of the drift field in the base region, deviations from the one-dimensional transistor model, and effects associated with velocity saturation in the bc space-charge region and voltage drops in the epitaxial collector region.

3.7.1 High-Level Injection in the Base Region

Under low-level-injection conditions, the transfer current varies in proportion to $\exp(V_{\mathrm{B'E'}}/V_{\mathrm{T}})$ since the density of minority carriers at the depletion-layer edge at the emitter side of the base region x_{be} shows such a voltage dependence. This follows immediately from the generalized law of mass action

$$p(x_{\mathrm{be}})n(x_{\mathrm{be}}) \;=\; n_{\mathrm{ie}}^2 \exp(V_{\mathrm{B'E'}}/V_{\mathrm{T}}) \tag{3.121}$$

and the condition that $p(x_{\mathrm{bc}}) \approx N_{\mathrm{A}}(x_{\mathrm{bc}}) = \mathrm{const}$. Since an increase in the density of injected electrons causes an increase in the hole density, the latter assumption will only be fulfilled to a good approximation if the density $n(x_{\mathrm{be}})$ of electrons injected into the base layer is small in comparison with the acceptor density N_{A}. Under high-level-injection conditions, when $n(x_{\mathrm{be}}) \ll N_{\mathrm{A}}$ may no longer be assumed, the condition $p(x_{\mathrm{be}}) \approx N_{\mathrm{A}}$ has to be replaced by the neutrality condition

$$N_{\mathrm{A}} + n(x_{\mathrm{be}}) \;=\; p(x_{\mathrm{be}})\;. \tag{3.122}$$

From (3.121) and (3.122), the electron concentration $n(x_{\mathrm{be}})$ at the depletion-layer edge at the emitter side is found to be

$$n(x_{\mathrm{be}}) \;=\; \frac{2n_{\mathrm{ie}}^2 \exp(V_{\mathrm{B'E'}}/V_{\mathrm{T}})}{N_{\mathrm{A}}\left[1+\sqrt{1+(4n_{\mathrm{ie}}^2/N_{\mathrm{A}}^2)\exp(V_{\mathrm{B'E'}}/V_{\mathrm{T}})}\right]}\;. \tag{3.123}$$

Under very high-level-injection conditions the second term in the square root dominates and (3.123) simplifies to

$$n(x_{\mathrm{be}}) \approx n_{\mathrm{ie}} \exp(V_{\mathrm{B'E'}}/2V_{\mathrm{T}}) \,. \tag{3.124}$$

This effect is automatically taken into account in Gummel's transfer current relation derived in Sect. 3.2.

In addition to the effects of the modified boundary condition, the transport of minority carriers is modified owing to changes in the built-in electric field: large values of the transfer current density cause an increase in the density of majority carriers, which neutralize the injected minority carriers. As the density of injected minority carriers decreases from the emitter to the collector junction, the density of the additional majority carriers will be larger at the emitter junction than at the collector junction. The resulting gradient of the majority carrier concentration might be expected to result in a hole current, which cannot flow, however, owing to the reverse-biased bc diode. Consequently an electric field develops, which suppresses hole diffusion and therefore aids electron transport through the base region, resulting in a reduced base transit time, a phenomenon known as the Webster effect [85]. If the neutral base region is assumed to be approximately neutral and $J_{\mathrm{p}} \approx 0$, the electric field strength in the base region has to obey

$$E \approx V_{\mathrm{T}} \frac{1}{n+N_{\mathrm{A}}} \frac{\mathrm{d}}{\mathrm{d}x}(n+N_{\mathrm{A}}) \,,$$

if the Einstein relation $D_{\mathrm{p}}/\mu_{\mathrm{p}} = V_{\mathrm{T}}$ applies. With the aid of this result, the electron current density equation may be written as

$$\begin{aligned} J_{\mathrm{n}} &= e\mu_{\mathrm{n}} n E + eD_{\mathrm{n}} \frac{\mathrm{d}n}{\mathrm{d}x} \\ &\approx eD_{\mathrm{n}} \left(1 + \frac{n}{n+N_{\mathrm{A}}}\right) \frac{\mathrm{d}n}{\mathrm{d}x} + eD_{\mathrm{n}} \frac{n}{n+N_{\mathrm{A}}} \frac{\mathrm{d}N_{\mathrm{A}}}{\mathrm{d}x} \,. \end{aligned} \tag{3.125}$$

In the low-level-injection case, where $n(x) \ll N_{\mathrm{A}}(x)$ throughout the neutral base region, this simplifies to

$$J_{\mathrm{n}} \approx eD_{\mathrm{n}} \frac{\mathrm{d}n}{\mathrm{d}x} + eD_{\mathrm{n}} \frac{n}{N_{\mathrm{A}}} \frac{\mathrm{d}N_{\mathrm{A}}}{\mathrm{d}x} \,,$$

a situation that has already been investigated in Sect. 3.1 for the special situation

$$\frac{1}{N_{\mathrm{A}}} \frac{\mathrm{d}N_{\mathrm{A}}}{\mathrm{d}x} = -\frac{\eta}{d_{\mathrm{B}}} = \mathrm{const.}$$

for very high-level injection, where $n(x) \gg N_{\mathrm{A}}(x)$, the electron current density equation (3.125) can be simplified to

$$J_{\mathrm{n}} \approx 2eD_{\mathrm{n}} \frac{\mathrm{d}n}{\mathrm{d}x} \,.$$

If recombination in the base region is neglected, the electron transfer current can then be written as the difference between a forward transfer current $I_{\rm TF}$ and a reverse transfer current $I_{\rm TR}$,

$$I_{\rm T} = I_{\rm TF} - I_{\rm TR} ,$$

where

$$I_{\rm TF} = 2eA_{\rm je}D_{\rm n}\frac{n(x_{\rm be})}{d_{\rm B}} \quad \text{and} \quad I_{\rm TR} = 2eA_{\rm je}D_{\rm n}\frac{n(x_{\rm bc})}{d_{\rm B}} ,$$

and $d_{\rm B} = x_{\rm bc} - x_{\rm be}$. From the spatial variation of the electron density

$$n(x) \approx n(x_{\rm bc}) + \frac{n(x_{\rm bc})-n(x_{\rm be})}{d_{\rm B}}(x-x_{\rm bc}) ,$$

the minority charge $Q_{\rm TB}$ that is stored in the neutral base region is found to be

$$Q_{\rm TB} = \tau_{\rm Bf}I_{\rm TF} + \tau_{\rm Br}I_{\rm TR} ,$$

where

$$\tau_{\rm Bf} = \tau_{\rm Br} = d_{\rm B}^2/4D_{\rm n} = \tau_{\rm B\infty} \tag{3.126}$$

are the forward and reverse base transit times. Owing to the field that aids transport, these values are half of those for a diffusion transistor under low-level-injection conditions. If a drift field due to a gradient of dopant density in the base region is already present under low-level-injection conditions, the forward base transit time can be smaller than $\tau_{\rm B\infty}$ and will increase towards this value at large current densities. Although a general analytical solution for the base transit times and stored charges is not possible, fitting functions that smoothly connect the low-level result with the high-level result may be used to obtain model equations for arbitrary injection levels [61]. Under forward-bias conditions, $n(x_{\rm bc})$ can be approximated by zero under low-level-injection conditions. For high-level injection, however, a finite value of $n(x_{\rm bc})$ has to be used, owing to the finite drift velocity of electrons in the bc space charge region.

3.7.2 High-Level Injection in the Collector Region

Under high-level-injection conditions, two effects that occur in the collector region have to be distinguished:

1. The voltage drop across the collector resistance, that is, the resistance formed by the undepleted portion of the epilayer, the subcollector and the collector contact, decreases the potential difference across the space charge layer – thus increasing the base width and base transit time –

and may even cause the bc diode to become internally forward biased (quasi-saturation). The critical current associated with the onset of quasi-saturation is [66]

$$I_2 = (V_{\mathrm{CB'}}+V_{\mathrm{JC}})/R_{\mathrm{C0}} \,, \tag{3.127}$$

where R_{C0} denotes the epilayer resistance,

$$R_{\mathrm{C0}} = d_{\mathrm{epi}}/(e\mu_{\mathrm{n0}}N_{\mathrm{Depi}}A_{\mathrm{jc}}) \,. \tag{3.128}$$

If the transfer current I_{T} reaches I_2, the total potential difference between the (internal) base node and the collector node occurs across the epilayer and the depletion layer vanishes. This current represents an upper limit on the amount of current that can flow in a transistor for a given value of V_{CB}, under the assumptions that (1) the current flow is one-dimensional, and (2) the minority-carrier concentration in the collector is negligible [67]. If the transfer current approaches the critical current I_2, the bc junction becomes forward biased and holes will be injected from the collector into the epilayer. In this case a substantial hole charge is stored in parts of the epitaxial collector regions, resulting, for example, in a degradation of the cutoff frequency of the device.

2. Another high-level-injection effect stems from the finite drift velocity $v_{\mathrm{nsat}} \approx 10^7$ cm/s of the electrons in the bc depletion layer, resulting in an electron density $n(x_{\mathrm{bc}}) \approx |J_{\mathrm{n}}|/ev_{\mathrm{nsat}}$, where J_{n} denotes the electron current density. An additional negative space charge is therefore present in the depletion layer, which reduces the electric field that forces holes to stay in the base region. Owing to the mobile charge, the charge density on the p-type side of the depletion layer is given by $\rho(x) = -eN_{\mathrm{A}} - |J_{\mathrm{n}}|/v_{\mathrm{nsat}}$, while on the n-type side the charge density $\rho(x) = eN_{\mathrm{Depi}} - |J_{\mathrm{n}}|/v_{\mathrm{nsat}}$ is obtained. The relations for the extensions x_{p} and x_{n} of the depletion layer into the p-type and n-type regions known from standard pn junction theory therefore have to be modified.

Partial Depletion of the Epilayer

The following calculation applies to the situation where a portion of the epilayer with thickness $d_{\mathrm{epi}} - x_{\mathrm{n}}$ remains undepleted. The electric field strength in the epilayer $E_{\mathrm{epi}} = -|J_{\mathrm{n}}|/e\mu_{\mathrm{n}}N_{\mathrm{Depi}}$ then serves as a boundary condition at the boundary on the n side of the space charge region. In contrast to conventional pn junction theory, the space charge region between x_{bc} and x_{cb} has a net charge density $\epsilon\,[E(x_{\mathrm{cb}}) - E(x_{\mathrm{bc}})] \approx \epsilon E_{\mathrm{epi}}$, as can be found from a solution of the Poisson equation. In the following, the electron transfer current is assumed to be below the value $I_1 = eA_{\mathrm{je}}N_{\mathrm{Depi}}v_{\mathrm{nsat}}$, i.e. the net space charge on the n-type side of the space charge layer is assumed to be positive. The conventional neutrality condition for the pn junction has to be modified to

$$\left(eN_{\mathrm{A}} + \frac{|J_{\mathrm{n}}|}{v_{\mathrm{nsat}}}\right) x_{\mathrm{p}} = \left(eN_{\mathrm{Depi}} - \frac{|J_{\mathrm{n}}|}{v_{\mathrm{nsat}}}\right) x_{\mathrm{n}} - \epsilon E_{\mathrm{epi}} \ ,$$

where $x_{\mathrm{p}} = x_{\mathrm{jc}} - x_{\mathrm{bc}}$ and $x_{\mathrm{n}} = x_{\mathrm{cb}} - x_{\mathrm{jc}}$. In addition to this, the total potential difference $\Delta\psi$ across the space charge region has to obey

$$\begin{aligned} \Delta\psi &= V_{\mathrm{CB'}} + V_{\mathrm{JC}} - \frac{|J_{\mathrm{n}}|(d_{\mathrm{epi}} - x_{\mathrm{n}})}{e\mu_{\mathrm{n}} N_{\mathrm{Depi}}} \\ &= \frac{1}{2\epsilon}\left(eN_{\mathrm{A}} + \frac{|J_{\mathrm{n}}|}{v_{\mathrm{nsat}}}\right) x_{\mathrm{p}}^2 + \frac{1}{2\epsilon}\left(eN_{\mathrm{Depi}} - \frac{|J_{\mathrm{n}}|}{v_{\mathrm{nsat}}}\right) x_{\mathrm{n}}^2 \ . \end{aligned} \tag{3.129}$$

This system of equations can be solved for x_{n} and x_{p}; for simplicity, we consider the situation $N_{\mathrm{A}} \gg N_{\mathrm{Depi}}$, corresponding to $x_{\mathrm{p}} \to 0$. Under these conditions, (3.129) simplifies to a quadratic relation for x_{n}:

$$\frac{eN_{\mathrm{Depi}}}{2\epsilon}\left(1 - \frac{I_{\mathrm{T}}}{I_1}\right) x_{\mathrm{n}}^2 - \frac{R_{\mathrm{C}} I_{\mathrm{T}}}{d_{\mathrm{epi}}} x_{\mathrm{n}} - (V_{\mathrm{CB'}} + V_{\mathrm{JC}} - R_{\mathrm{C}} I_{\mathrm{T}}) = 0 \ ,$$

where $R_{\mathrm{C}} = d_{\mathrm{epi}}/(e\mu_{\mathrm{n}} N_{\mathrm{Depi}} A_{\mathrm{je}}) = R_{\mathrm{C0}}\mu_{\mathrm{n0}}/\mu_{\mathrm{n}}$ and $I_{\mathrm{T}} = A_{\mathrm{jc}}|J_{\mathrm{n}}|$. Solution of this equation gives

$$x_{\mathrm{n}} = \sqrt{\frac{2\epsilon(V_{\mathrm{CB'}} + V_{\mathrm{JC}})}{eN_{\mathrm{Depi}}(1 - I_{\mathrm{T}}/I_1)} - 2x_\delta d_{\mathrm{epi}} + x_\delta^2} \ + \ x_\delta \ , \tag{3.130}$$

where[33]

$$\begin{aligned} x_\delta &= \frac{\epsilon}{eN_{\mathrm{Depi}} d_{\mathrm{epi}}} \frac{R_{\mathrm{C}} I_{\mathrm{T}}}{1 - I_{\mathrm{T}}/I_1} = \frac{\epsilon v_{\mathrm{nsat}}}{e\mu_{\mathrm{n}} N_{\mathrm{Depi}}} \frac{I_{\mathrm{T}}/I_1}{1 - I_{\mathrm{T}}/I_1} \\ &\approx \frac{\epsilon v_{\mathrm{nsat}}}{e\mu_{\mathrm{n0}} N_{\mathrm{Depi}}} \frac{1}{\sqrt{1 - I_{\mathrm{T}}^2/I_1^2}} \frac{I_{\mathrm{T}}/I_1}{1 - I_{\mathrm{T}}/I_1} \end{aligned}$$

is determined by the voltage drop in the undepleted part of the epilayer. If x_δ is small and $\mu_{\mathrm{n}} \approx \mu_{\mathrm{n0}}$, the extension of the bc depletion layer is approximately

$$x_{\mathrm{n}} = \sqrt{\frac{2\epsilon(V_{\mathrm{CB'}} + V_{\mathrm{JC}})}{eN_{\mathrm{Depi}}}} \sqrt{\frac{1 - I_{\mathrm{T}}/I_2}{1 - I_{\mathrm{T}}/I_1}} \ ; \tag{3.131}$$

this expression reduces correctly to the result obtained for a one-sided depletion layer in the limit $I_{\mathrm{T}} \to 0$.

[33] Assuming $I_{\mathrm{T}} = e\mu_{\mathrm{n}} N_{\mathrm{Depi}} A_{\mathrm{je}} E_{\mathrm{epi}}$ and a field-dependent mobility of the form

$$\mu_{\mathrm{n}} = \mu_{\mathrm{n0}} / \sqrt{1 + (\mu_{\mathrm{n0}} E_{\mathrm{epi}}/v_{\mathrm{nsat}})^2} \ ,$$

the identity $\mu_{\mathrm{n}} = \mu_{\mathrm{n0}}\sqrt{1 - I_{\mathrm{T}}^2/I_1^2}$ can be derived, which allows us to express the field-dependent mobility as a function of the transfer current.

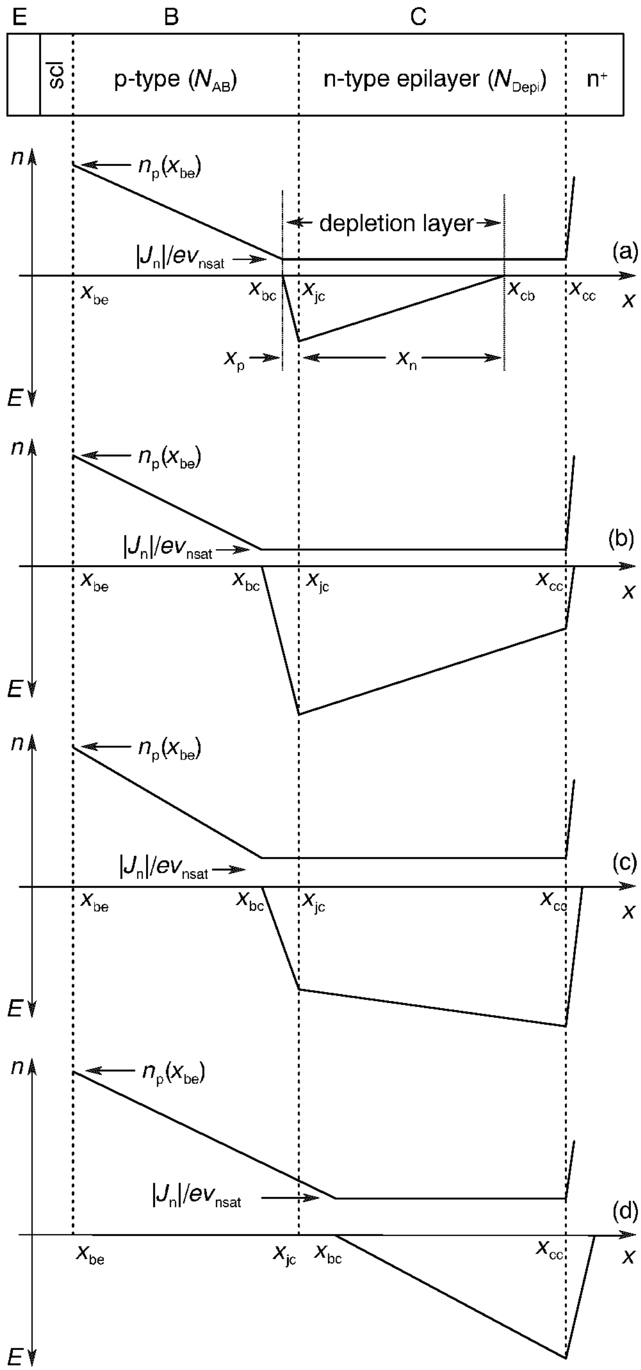


Fig. 3.20. Electron concentration and electric field (schematic) in the bc depletion layer for different bias and injection levels. (**a**) Forward operation with a moderate value of V_{CB} (the depletion layer does not reach the heavily doped subcollector), (**b**) forward operation with large V_{CB} (the epitaxial layer is fully depleted), (**c**) forward operation with a fully depleted epitaxial layer and a transfer current $I_T > I_1$, (**d**) base pushout (after [15])

Equation (3.131) describes two effects of the transfer current on the bc depletion layer associated with the critical current levels I_2 and I_1:

- Owing to the voltage drop across the undepleted part of the epilayer, the voltage drop across the depletion layer decreases with increasing values of I_T, and so does the width of the depletion layer.
- An increase in the width of the bc depletion layer is obtained on the other side as the space charge density in the depleted part of the epilayer decreases with increasing transfer current.

If $I_1 > I_2$, the first effect dominates and the width of the depletion layer will decrease with increasing transfer current, if $I_1 < I_2$ the second effect is more pronounced, resulting in an increase of the depletion layer width with increasing transfer current. Equation (3.130) can be applied if $x_n \le d_{epi}$, a requirement that is equivalent to

$$V_{CB'} \le \frac{eN_{Depi}d_{epi}^2}{2\epsilon}\left(1 - \frac{I_T}{I_1}\right) - V_{JC} .$$

This relation reduces to

$$V_{CB'} = \frac{eN_{Depi}d_{epi}^2}{2\epsilon} - V_{JC}$$

in the limit $|I_T| \to 0$, a result that is in accordance with the elementary depletion-layer analysis of one-sided junctions. If the transfer current $I_T > 0$ increases, the voltage $V_{CB'}$ that is required for the space charge region to meet the buried layer decreases. If

$$I_T < \left(1 - \frac{2\epsilon(V_{CB'}+V_{JC})}{eN_{Depi}d_{epi}^2}\right) I_1 ,$$

the epilayer will be partially depleted, and a change of transfer current will affect the depletion-layer width. This has consequences for the bc depletion capacitance, the output conductance and the forward transit time: the (internal) bc depletion capacitance is determined to a large extent by the extension of the space charge layer into the lightly doped epilayer

$$c_{jc} = \frac{\epsilon X_{CJC}A_{jc}}{d_{jc}} \approx X_{CJC}A_{jc}\sqrt{\frac{eN_{Depi}}{2\epsilon(V_{CB'}+V_{JC})}}\sqrt{\frac{1 - I_T/I_1}{1 - I_T/I_2}} ,$$

resulting in a decrease of c_{jc} in proportion to $\sqrt{1 - I_T/I_1}$. Since c_{jc} affects the value of the Early voltage and thus the output conductance, which varies in proportion to c_{jc} as described in Sect. 3.2, the output conductance $g_o \sim c_{jc}$ of the bipolar junction transistor (BJT) will decrease with increasing transfer current if $I_1 > I_2 \gg I_T$.

Full Depletion of the Epilayer

If V_{CB} is large enough to deplete the epitaxial layer, as illustrated in Fig. 3.20b, the extension of the space charge region and thus the bc depletion capacitance c_{jc} are approximately constant, whereas the epilayer series resistance vanishes. With increasing electron current density, however, the electric-field distribution in the epilayer will change. While the electric field strength decreases in magnitude across the epilayer towards the collector contact if $I_T < I_1$, it increases as soon as $I_T > I_1$, as shown in Fig. 3.20c. If $I_T = I_1$, the epilayer is neutral, since the density of mobile charge equals the donor density. The electric field is then caused by a dipole layer, as illustrated in Fig. 3.21.

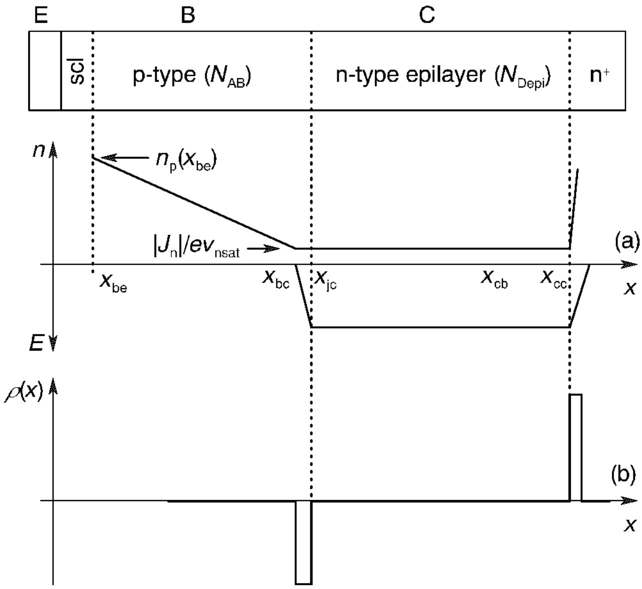


Fig. 3.21. (**a**) Electron concentration and electric-field distribution (schematic) in the bc depletion layer for $I_T = I_1$, and (**b**) corresponding space charge distribution

Further increase of the electron current results in an increase of the negative space charge in the epilayer and causes a shift of the base depletion-layer boundary towards the collector contact, since less negative space charge is required in the p-type region to yield a given potential difference. We may define a critical current I_{KK}, where $x_{bc} = x_{jc}$. Under these conditions the space charge in the p-type region has vanished and the potential difference occurs predominantly across the epilayer. A double integration of the Poisson equation then gives

$$V_{B'C} - V_{JC} = \int_0^{d_{epi}} E(x)\, dx = \frac{1}{\epsilon}\left(eN_{Depi} - \frac{I_{KK}}{A_{je}v_{nsat}}\right)\frac{d_{epi}^2}{2}$$

if $x_{jc} = 0$, corresponding to

$$I_{\mathrm{KK}} = I_1 + \frac{2\epsilon A_{\mathrm{je}} v_{\mathrm{nsat}}(V_{\mathrm{JC}}+V_{\mathrm{CB'}})}{d_{\mathrm{epi}}^2} .$$

Here $I_1 = eN_{\mathrm{Depi}} A_{\mathrm{je}} v_{\mathrm{nsat}}$ if $E(0)$ is taken to be zero. If the transfer current I_{T} exceeds I_{KK}, the base width will increase substantially with the transfer current, an effect commonly referred to as base pushout or the Kirk effect. This effect causes the base transit time to increase and thus the current gain to decrease. The Kirk effect is related to quasi-saturation,[34] which describes the situation where the bc diode becomes internally forward biased owing to a voltage drop at the collector series resistance, despite an external reverse bias ($V_{\mathrm{CE}} > V_{\mathrm{BE}}$).

Current-Induced Base Widening

After the onset of base pushout, the epitaxial collector region can be divided into the induced base region, or injection region [67, 86], between 0 and x_{i}, and the collecting zone [87], or end zone [88], between x_{i} and d_{epi}. In the injection region quasi-neutrality may be assumed, whereas the collecting zone will be either neutral (if $I_{\mathrm{T}} < I_1$, as in Fig. 3.22) or space-charged.

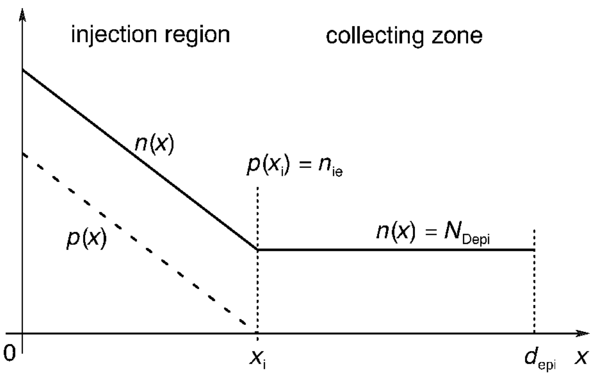


Fig. 3.22. Distribution of electrons and holes in the epitaxial collector region in ohmic quasi-saturation (schematic representation)

In the case of a neutral collecting zone, $I_{\mathrm{T}} < I_1$ and the potential difference is given by

$$V_{\mathrm{CB'}}+V_{\mathrm{JC}} = y_{\mathrm{i}} R_{\mathrm{C}} I_{\mathrm{T}} , \qquad \text{where} \qquad R_{\mathrm{C}} = \frac{R_{\mathrm{C0}}}{\sqrt{1-I_{\mathrm{T}}^2/I_1^2}}$$

denotes the field-dependent epilayer resistance and $y_{\mathrm{i}} = 1-x_{\mathrm{i}}/d_{\mathrm{epi}}$. This relation can be solved for y_{i}, with the result

[34]Sometimes both effects are termed "quasi-saturation". Here the term "quasi-saturation" is applied if holes are injected into the collector region for $I_{\mathrm{T}} < I_1$, while base pushout for $I_{\mathrm{T}} > I_1$ is referred to as the Kirk effect.

$$y_\mathrm{i} = \frac{I_2}{I_\mathrm{T}}\sqrt{1-\frac{I_\mathrm{T}^2}{I_1^2}}\,, \quad \text{or} \quad x_\mathrm{i} = d_\mathrm{epi}\left(1-\frac{I_2}{I_\mathrm{T}}\sqrt{1-\frac{I_\mathrm{T}^2}{I_1^2}}\right)\,,$$

a result that is valid only if y_i is real and $y_i \leq 1$, corresponding to

$$I_1 \geq I_\mathrm{T} \geq I_1 I_2 \Big/ \sqrt{I_1^2 + I_2^2}\,.$$

The collecting zone becomes space-charged if the transfer current I_T is larger than I_1: the electrons then travel at their saturated drift velocity v_nsat and the electron density $n = N_\mathrm{Depi} I_\mathrm{T}/I_1$ exceeds N_Depi. A double integration of the Poisson equation

$$-\frac{\mathrm{d}E}{\mathrm{d}x} = \frac{eN_\mathrm{Depi}}{\epsilon}\left(\frac{I_\mathrm{T}}{I_1}-1\right)$$

with respect to the collecting zone gives

$$V_{\mathrm{CB}'} + V_\mathrm{JC} = -E(x_\mathrm{i})(d_\mathrm{epi}-x_\mathrm{i}) + \frac{eN_\mathrm{Depi}}{2\epsilon}\left(\frac{I_\mathrm{T}}{I_1}-1\right)(d_\mathrm{epi}-x_\mathrm{i})^2\,.$$

In [88], the choice $|E(x_\mathrm{i})| \approx 8 \times 10^3$ V/cm is suggested.

Assuming quasi-neutrality in the injection region, i.e. $n \approx p + N_\mathrm{Depi}$, the electron current density equation reads

$$J_\mathrm{n} = -\frac{I_\mathrm{T}}{A_\mathrm{je}} = e\mu_\mathrm{n0} V_\mathrm{T}\left(2+\frac{N_\mathrm{Depi}}{p}\right)\frac{\mathrm{d}p}{\mathrm{d}x} \tag{3.132}$$

if $N_\mathrm{Depi} =$ const. and if the hole current density is assumed to be negligible. Since the drift field is comparatively small in the injection region, velocity saturation effects can be neglected; integration of (3.132) across the injection region from 0 to x_i gives

$$\frac{x_\mathrm{i}}{d_\mathrm{epi}} R_\mathrm{C0} I_\mathrm{T} = V_\mathrm{T}\left(2\,\frac{p(0)-p(x_\mathrm{i})}{N_\mathrm{Depi}} + \ln\frac{p(0)}{p(x_\mathrm{i})}\right) \tag{3.133}$$

This relation allows to compute x_i if $p(x_\mathrm{i}) = n_\mathrm{ie}$ is used to define the boundary of the injection region. The charge stored in the injection region can be calculated as

$$\begin{aligned} eA_\mathrm{je}\int_0^{x_\mathrm{i}} p\,\mathrm{d}x &= -\frac{e^2 D_\mathrm{n} A_\mathrm{je}^2}{I_\mathrm{T}}\int_{p(0)}^{p(x_\mathrm{i})}(2p + N_\mathrm{Depi})\,\mathrm{d}p \\ &\approx \frac{e^2 D_\mathrm{n} A_\mathrm{je}^2}{I_\mathrm{T}}\,p(0)\left[p(0)+N_\mathrm{Depi}\right] \end{aligned}$$

after substitution of the variable of integration with the help of (3.132). A modified form of these relations is used for the formulation of the MEXTRAM epilayer model (see Sect. 3.16).

3.7.3 The Epilayer Model of Kull et al.

As can be seen from the preceding discussion, the description of the bc junction under forward bias or in quasi-saturation in terms of an ideal diode in series with a collector resistance is poor. In order to improve on the Gummel Poon model in this respect, an extra node C″ was introduced in [89] to describe the resistance and stored charge of the epilayer more accurately. This modification gives a good description of the currents and charges in the epilayer as long as the whole epilayer is quasi-neutral. In [89] the electron mobility is assumed to obey

$$\mu_{\mathrm{n}} = \frac{\mu_{\mathrm{n}0}}{1+\dfrac{\mu_{\mathrm{n}0}}{v_{\mathrm{nsat}}}\left|\dfrac{\partial\phi_{\mathrm{n}}}{\partial x}\right|}\;; \tag{3.134}$$

with this assumption, the electron current density equation reads

$$J_{\mathrm{n}}\left(1+\frac{\mu_{\mathrm{n}0}}{v_{\mathrm{nsat}}}\left|\frac{\partial\phi_{\mathrm{n}}}{\partial x}\right|\right) = -e\mu_{\mathrm{n}0}n\frac{\mathrm{d}\phi_{\mathrm{n}}}{\mathrm{d}x}\;. \tag{3.135}$$

Integration across the lightly doped epitaxial collector region, neglecting recombination ($\mathrm{d}J_{\mathrm{n}}/\mathrm{d}x = 0$), yields[35]

$$\begin{aligned} e\mu_{\mathrm{n}0}\int_0^{d_{\mathrm{epi}}} n\frac{\mathrm{d}\phi_{\mathrm{n}}}{\mathrm{d}x}\,\mathrm{d}x &= e\mu_{\mathrm{n}0}\int_{\phi_{\mathrm{n}}(0)}^{\phi_{\mathrm{n}}(d_{\mathrm{epi}})} n\,\mathrm{d}\phi_{\mathrm{n}} \\ &= -J_{\mathrm{n}}d_{\mathrm{epi}}\left(1+\frac{|\phi_{\mathrm{n}}(0)-\phi_{\mathrm{n}}(d_{\mathrm{epi}})|}{V_0}\right)\;, \end{aligned} \tag{3.136}$$

where

$$V_0 = d_{\mathrm{epi}}v_{\mathrm{nsat}}/\mu_{\mathrm{n}0}\;. \tag{3.137}$$

The parameter V_0 determines a critical value for the voltage drop across the epilayer, above which velocity saturation has to be taken into account

With the assumption $\partial\phi_{\mathrm{p}}/\partial x = 0$, i.e., the assumption of a constant minority-carrier quasi-Fermi level and quasi-neutrality in the collector region ($n \approx p+N_{\mathrm{Depi}}$), the identity

$$\mathrm{d}\phi_{\mathrm{n}} = -\frac{V_{\mathrm{T}}}{p}\,\frac{2p+N_{\mathrm{Depi}}}{p+N_{\mathrm{Depi}}}\,\mathrm{d}p$$

holds. This allows one to evaluate the integral

[35] Since $\partial\phi_{\mathrm{n}}/\partial x$ is monotonic within the region of integration,

$$\int_0^{d_{\mathrm{epi}}}\left|\frac{\partial\phi_{\mathrm{n}}}{\partial x}\right|\mathrm{d}x = \left|\int_0^{d_{\mathrm{epi}}}\frac{\partial\phi_{\mathrm{n}}}{\partial x}\,\mathrm{d}x\right|\;.$$

$$\frac{1}{V_\mathrm{T}}\int_{\phi_\mathrm{n}(0)}^{\phi_\mathrm{n}(d_\mathrm{epi})} n\,\mathrm{d}\phi_\mathrm{n} \;=\; 2\,[\,p(0)-p(d_\mathrm{epi})\,]+N_\mathrm{Depi}\ln\left(\frac{p(0)}{p(d_\mathrm{epi})}\right) \tag{3.138}$$

$$= 2\,[\,p(0)-p(d_\mathrm{epi})\,]-N_\mathrm{Depi}\ln\left(\frac{n(0)}{n(d_\mathrm{epi})}\right)+\frac{N_\mathrm{Depi}}{V_\mathrm{T}}(V_\mathrm{B'C''}-V_\mathrm{B'C'})\ ,$$

where $V_\mathrm{B'C''}=\phi_\mathrm{p}-\phi_\mathrm{n}(0)$ and $V_\mathrm{B'C'}=\phi_\mathrm{p}-\phi_\mathrm{n}(d_\mathrm{epi})$, and

$$n(x)p(x) \;=\; p(x)[\,N_\mathrm{Depi}+p(x)\,] \;=\; n_\mathrm{ie}^2\exp\frac{\phi_\mathrm{p}-\phi_\mathrm{n}(x)}{V_\mathrm{T}}$$

has been assumed. Combining equations yields the following for the current $I_\mathrm{epi}=-J_\mathrm{n}A_\mathrm{je}$ that passes through the epi layer [89]:

$$I_\mathrm{epi} \;=\; \frac{K(v_\mathrm{B'C'})-K(v_\mathrm{B'C''})-\ln\left[\dfrac{1+K(v_\mathrm{B'C'})}{1+K(v_\mathrm{B'C''})}\right]+\dfrac{v_\mathrm{B'C'}-v_\mathrm{B'C''}}{V_\mathrm{T}}}{\dfrac{R_\mathrm{C0}}{V_\mathrm{T}}\left(1+\dfrac{|v_\mathrm{B'C'}-v_\mathrm{B'C''}|}{V_0}\right)}\ ,$$

where

$$K(v) \;=\; \sqrt{1+\gamma\exp\left(\frac{v}{V_\mathrm{T}}\right)} \quad \text{and} \quad \gamma \;=\; \left(\frac{2n_\mathrm{ie}}{N_\mathrm{Depi}}\right)^2\ , \tag{3.139}$$

and R_C0 denotes the epilayer resistance defined in (3.128).

When the device is operated at small current levels, the bc junction is reverse biased internally, i.e. $v_\mathrm{B'C'}<0$ and $v_\mathrm{B'C''}<0$; in this case the functions $K(v)$ are approximately equal to one, and

$$I_\mathrm{epi} \;=\; \frac{v_\mathrm{B'C'}-v_\mathrm{B'C''}}{R_\mathrm{C'C''}}\ , \quad \text{where} \quad R_\mathrm{C'C''} \;=\; R_\mathrm{C0}\left(1+\frac{|v_\mathrm{B'C'}-v_\mathrm{B'C''}|}{V_0}\right)$$

is the epilayer resistance in the presence of velocity saturation effects. This formulation of quasi-saturation effects [89] has been implemented in SPICE (see Sect. 3.8). As has been pointed out, for example, in [90,91], the assumption of quasi-neutrality in the entire epitaxial collector region, which excludes voltage drops due to space charge modulation by hot carriers, gives erroneous results for $I_\mathrm{T}>I_1$. The model was therefore further developed in [90–94] in order to take account of hot-carrier effects and to provide smoother results for the first and higher derivatives of the output characteristic (see Sect. 3.16).

3.8 The Gummel–Poon Model

The Gummel–Poon model [13], which is based on the integral charge-control relation (ICCR, see Sect. 3.2) [11], has served as a standard for more than two decades. On the basis of a rather general integral relation, high-level-injection effects and the Early effect are incorporated a priori. In this section the Gummel–Poon model, together with the extensions introduced in [89], as implemented in SPICE, will be considered.

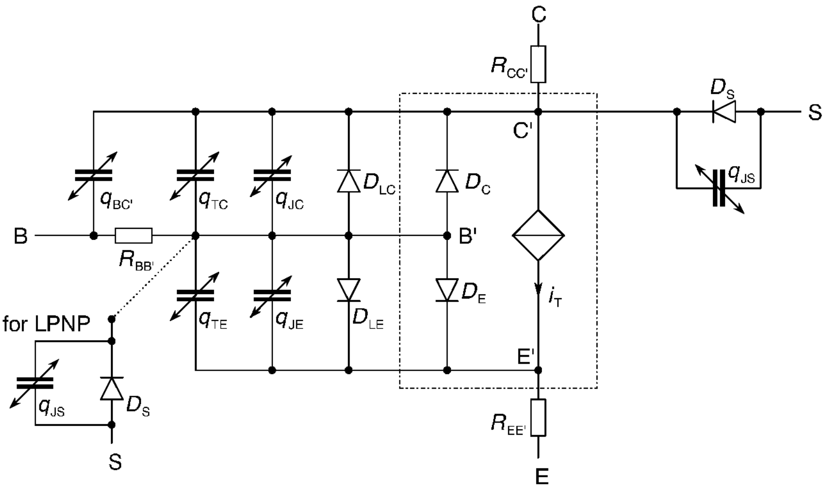


Fig. 3.23. Equivalent circuit of vertical npn bipolar transistor as employed in SPICE and extension for lateral pnp transistor (LPNP)

Figure 3.23 shows the equivalent circuit employed in SPICE for the modeling of the large-signal behavior of bipolar transistors; it extends the elementary transistor model described in Sect. 1.6, which is enclosed in a dashed frame. The internal nodes B′, C′ and E′ are connected to the external nodes B, C and E via the series resistances $R_{BB'}$, $R_{CC'}$ and $R_{EE'}$.

3.8.1 Transfer Current and Current Gain

On the basis of the integral charge control relation, the transfer current is written in the Gummel–Poon model as

$$i_T = \frac{I_S}{q_B}\left[\exp\left(\frac{v_{B'E'}(t)}{N_F V_T}\right) - \exp\left(\frac{v_{B'C'}(t)}{N_R V_T}\right)\right] = \frac{i_{CE} - i_{EC}}{q_B} . \tag{3.140}$$

The additional parameters N_F and N_R were introduced as fitting parameters, but may generally be assumed to be equal to one in homojunction transistors. The normalized base charge q_{B} has to be modeled as a bias-dependent quantity,

$$q_{\mathrm{B}} = 1 + \frac{q_{\mathrm{JE}}}{Q_{\mathrm{B0}}} + \frac{q_{\mathrm{JC}}}{Q_{\mathrm{B0}}} + \frac{q_{\mathrm{TE}}}{Q_{\mathrm{B0}}} + \frac{q_{\mathrm{TC}}}{Q_{\mathrm{B0}}} , \tag{3.141}$$

where $q_{\mathrm{JE}}/Q_{\mathrm{B0}}$ describes the relative change of the base charge due to charging of the eb depletion capacitance and $q_{\mathrm{JC}}/Q_{\mathrm{B0}}$ describes the relative change of the base charge due to charging of the bc depletion capacitance; the terms $q_{\mathrm{TE}}/Q_{\mathrm{B0}}$ and $q_{\mathrm{TC}}/Q_{\mathrm{B0}}$ describe the relative change of the diffusion charges associated with the eb and bc junctions. Under low-level injection the stored minority charge may be neglected, resulting in the approximation

$$q_{\mathrm{B}} \approx 1 + \frac{q_{\mathrm{JE}}}{Q_{\mathrm{B0}}} + \frac{q_{\mathrm{JC}}}{Q_{\mathrm{B0}}} = q_1 , \tag{3.142}$$

Introducing the forward Early voltage V_{AF} and the reverse Early voltage V_{AR}, the value of q_1 is approximated by the relation

$$q_1 = \left(1 - \frac{v_{\mathrm{B'C'}}}{V_{\mathrm{AF}}} - \frac{v_{\mathrm{B'E'}}}{V_{\mathrm{AR}}}\right)^{-1} .$$

Under high-level injection, the terms $q_{\mathrm{TE}}/Q_{\mathrm{B0}}$ and $q_{\mathrm{TC}}/Q_{\mathrm{B0}}$ may no longer be neglected. In addition to q_1, the term

$$\frac{1}{Q_{\mathrm{B0}}}\left(\tau_{\mathrm{f}}\frac{i_{\mathrm{CE}}}{q_{\mathrm{B}}} + \tau_{\mathrm{r}}\frac{i_{\mathrm{EC}}}{q_{\mathrm{B}}}\right) = \frac{q_2}{q_{\mathrm{B}}}$$

must be considered. The normalized base charge q_{B} is then determined from the quadratic relation

$$q_{\mathrm{B}} = q_1 + q_2/q_{\mathrm{B}} . \tag{3.143}$$

Neglecting the bias dependence of τ_{f} allows us to define the forward knee current $I_{\mathrm{KF}} = Q_{\mathrm{B0}}/\tau_{\mathrm{f}}$ and the reverse knee current $I_{\mathrm{KR}} = Q_{\mathrm{B0}}/\tau_{\mathrm{r}}$. Using these quantities, (3.143) may be solved for q_{B} to give

$$q_{\mathrm{B}} = \frac{1}{2}\left(q_1 + \sqrt{q_1^2 + 4q_2}\right) \approx \frac{q_1}{2}\left(1 + (1+4q_2)^{N_{\mathrm{K}}}\right) , \tag{3.144}$$

where

$$q_2 = \frac{i_{\mathrm{CE}}}{I_{\mathrm{KF}}} + \frac{i_{\mathrm{EC}}}{I_{\mathrm{KR}}} . \tag{3.145}$$

The first identity in (3.144) is used in the SPICE Gummel–Poon model to describe the bias dependence of q_{B}. PSPICE employs a modified formula and allows one to modify the “roll-off exponent” N_{K} from its default value of 1/2 for a better fit to measured data.

If the simulation temperature T differs from the nominal temperature T_0, the transfer saturation current $I_{\mathrm{S}}(T)$ can be calculated from the parameters $I_{\mathrm{S}} = I_{\mathrm{S}}(T_0)$ listed in the `.MODEL` statement using

$$I_S(T) = I_S(T_0)\left(\frac{T}{T_0}\right)^{X_{TI}} \exp\left[\frac{E_G}{V_T}\left(\frac{T}{T_0}-1\right)\right] . \tag{3.146}$$

If not specified otherwise, default values of the parameters X_{TI} and E_G corresponding to silicon BJTs are used.

3.8.2 Base Current Components

Leakage diodes (D_{LE} and D_{LC}) are placed parallel to the eb diode D_E and bc diode D_C to model nonideal diode behavior caused by recombination in the space charge layers. The recombination current associated with a forward-biased CS junction is modeled by the diode D_S. All diodes are described by current–voltage characteristics of the form

$$i = I_S\left[\exp\left(\frac{v}{NV_T}\right) - 1\right] .$$

The saturation currents I_S and emission coefficients N of the various diodes are listed in Table 3.1.

Table 3.1. Diode saturation currents and emission coefficients in SPICE

Diode	D_E	D_C	D_{LE}	D_{LC}	D_S
Saturation current	`IS/BF`	`IS/BR`	`ISE`	`ISE`	`ISS`
Emission coefficient	`NF`	`NR`	`NE`	`NC`	`NS`

The parameters `BF` and `BR` denote the ideal forward and reverse current gains already known from the elementary transistor model. These are modeled as temperature-dependent quantities according to

$$B_F(T) = B_F(T_0)\left(\frac{T}{T_0}\right)^{X_{TB}} \quad \text{and} \quad B_R(T) = B_R(T_0)\left(\frac{T}{T_0}\right)^{X_{TB}} \tag{3.147}$$

The saturation current of the eb leakage diode is calculated using

$$I_{SE}(T) = I_{SE}(T_0)\left(\frac{T}{T_0}\right)^{X_{TI}/N_E - X_{TB}} \exp\left[\frac{E_G}{N_E V_T}\left(\frac{T}{T_0}-1\right)\right] , \tag{3.148}$$

with analogous relations for the other diodes.

3.8.3 Current Gain

In forward operation, the Gummel–Poon model gives for the base current

$$I_B = \frac{I_S}{B_F}\exp\frac{V_{B'E'}}{V_T} + I_{SE}\exp\frac{V_{B'E'}}{N_E V_T} ,$$

while the collector current is written as

$$I_C = \frac{I_S}{q_B}\exp\frac{V_{B'E'}}{V_T} .$$

Expressing $\exp(V_{\mathrm{B'E'}}/V_{\mathrm{T}})$ in terms of I_{C} yields the forward common-emitter current gain as follows:

$$B_{\mathrm{N}}(I_{\mathrm{C}}) = \frac{I_{\mathrm{C}}}{I_{\mathrm{B}}} = \frac{1}{\dfrac{q_{\mathrm{B}}}{B_{\mathrm{F}}} + \dfrac{I_{\mathrm{SE}}}{I_{\mathrm{C}}}\left(\dfrac{I_{\mathrm{C}}q_{\mathrm{B}}}{I_{\mathrm{S}}}\right)^{1/N_{\mathrm{E}}}} . \tag{3.149}$$

In integrated bipolar transistors, forward bias leakage currents are frequently small. In this case, the eb leakage diode shows up for small values of I_{C} only. Under these conditions, $q_{\mathrm{B}} \approx q_1$ and

$$B_{\mathrm{N}} \approx \frac{1}{I_{\mathrm{SE}}}\left(\frac{I_{\mathrm{S}}}{q_1}\right)^{1/N_{\mathrm{E}}} I_{\mathrm{C}}^{1-1/N_{\mathrm{E}}} , \tag{3.150}$$

which describes an increase $B_{\mathrm{N}} \sim I_{\mathrm{C}}^{1-1/N_{\mathrm{E}}}$ if the bias dependence of q_1 is neglected for simplicity. In a double-logarithmic plot of B_{N} versus I_{C}, this behavior is represented by a line of slope $1 - 1/N_{\mathrm{E}}$ as illustrated in Fig. 3.24.

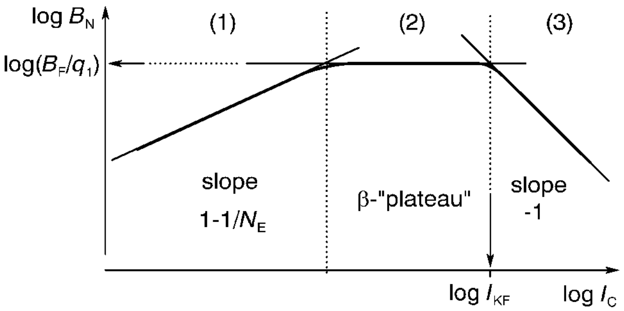


Fig. 3.24. Double-logarithmic plot of current gain versus I_{C}

If the influence of the eb leakage diode may be neglected, the current gain is

$$B_{\mathrm{N}} \approx \frac{B_{\mathrm{F}}}{q_1 + I_{\mathrm{C}}/I_{\mathrm{KF}}} , \tag{3.151}$$

since $q_{\mathrm{B}} = q_1 + I_{\mathrm{C}}/I_{\mathrm{KF}}$. If $I_{\mathrm{C}} \ll I_{\mathrm{KF}}$, the value of $B_{\mathrm{N}} \approx B_{\mathrm{F}}/q_1$ is almost constant, resulting in the β Plateau (Fig. 3.24). Since an increase in I_{C} is associated with an increase in $V_{\mathrm{B'E'}}$ and thus an increase in q_1, a slight decrease in B_{N} with increasing I_{C} is observed in this region. For values of $I_{\mathrm{C}} \gg I_{\mathrm{KF}}$, the term q_1 may be neglected in (3.151) and

$$B_{\mathrm{N}} \approx B_{\mathrm{F}} I_{\mathrm{KF}}/I_{\mathrm{C}} \tag{3.152}$$

is obtained, corresponding to a decrease $B_{\mathrm{N}} \sim 1/I_{\mathrm{C}}$. In a double-logarithmic plot of B_{N} versus I_{C}, this behavior is represented by a line of slope -1 as illustrated in Fig. 3.24.

3.8.4 Charge Storage

The charge stored in the bc depletion capacitance is described by $q_{\rm JC}$ and $q_{\rm BC'}$, the charge stored on the eb depletion capacitance is denoted by $q_{\rm JE}$ and the charge in the cs depletion capacitance is denoted by $q_{\rm JS}$. The bias-dependent depletion capacitances are described by expressions of the form

$$c_{\rm j}(v') = C_{\rm J0}\,(1-v'/V_{\rm J})^{-M} \;,$$

with parameters as listed in Table 3.2.

Table 3.2. Parameters for depletion capacitances in SPICE

Capacitance	$c_{\rm je}$	$c_{\rm jc}$	$c_{\rm bc'}$	$c_{\rm js}$
Zero-bias capacitance $c_{\rm J0}$	`CJE`	`XCJC`×`CJC`	(`1-XCJC`)×`CJC`	`CJS`
Built-in voltage $V_{\rm J}$	`VJE`	`VJC`	`VJC`	`VJS`
Grading coefficient M	`MJE`	`MJC`	`MJC`	`MJS`
Voltage v'	$v_{\rm B'E'}$	$v_{\rm B'C'}$	$v_{\rm BC'}$	$v_{\rm SC'}$

The built-in voltages `VJE`, `VJC` and `VJS` and the parameters `CJE`, `CJC` and `CJS` are defined as temperature-dependent quantities in accordance with (3.222) and (3.221). Linear extrapolation is used for the calculation of the depletion capacitance for a forward bias in excess of $F_{\rm C}V_{\rm J}$ (Fig. 3.67). The parameter $X_{\rm CJC} = A_{\rm je}/A_{\rm jc}$ is used to divide the bc depletion capacitance into an internal part

$$c_{\rm jc} \;=\; X_{\rm CJC}C_{\rm JC}\,(1-\,v_{\rm B'C'}/V_{\rm JC})^{-M_{\rm JC}}$$

and an external part $c_{\rm bc'} \sim (1 - X_{\rm CJC})$. Only the internal portion is charged and discharged via the base resistance $R_{\rm BB'}$.

The minority charge stored in the bipolar transistor is divided into a portion $q_{\rm TE} \sim i_{\rm CE}$ associated with the eb diode and a portion $q_{\rm TC} \sim i_{\rm EC}$ associated with the bc diode. A quasi-static approximation is used for the description of the charges $q_{\rm TE}$ and $q_{\rm TC}$, which are written as

$$q_{\rm TE} \;=\; \tau_{\rm f} i_{\rm CE}/q_{\rm B} \quad \text{and} \quad q_{\rm TC} \;=\; \tau_{\rm r} i_{\rm EC}/q_{\rm B}\;. \tag{3.153}$$

While $\tau_{\rm r} = T_{\rm R}$ is taken to be constant, the value of $\tau_{\rm f}$ is bias-dependent and is described in terms of the parameters `TF`, `XTF`, `ITF` and `VTF`:

$$\tau_{\rm f} \;=\; T_{\rm F}\left[1 + X_{\rm TF}\left(\frac{I_{\rm CE}}{I_{\rm CE}+I_{\rm TF}}\right)^2 \exp\!\left(\frac{v_{\rm B'C'}}{1.44V_{\rm TF}}\right)\right]\;. \tag{3.154}$$

The parameters of this empirical formula are determined from measured $\tau_{\rm f}(I_{\rm C})$ curves.

3.8.5 Series Resistances

The emitter resistance $R_{\mathrm{EE'}}$ is modeled as a temperature-dependent ohmic series resistance

$$R_{\mathrm{EE'}} = R_{\mathrm{E}}\left[1 + T_{\mathrm{RE1}}(T-T_0) + T_{\mathrm{RE2}}(T-T_0)^2\right] , \tag{3.155}$$

where T_{RE1} and T_{RE2} denote the linear and quadratic temperature coefficients of $R_{\mathrm{EE'}}$.

The base resistance is modeled in terms of the parameters $R_{\mathrm{B}}, R_{\mathrm{BM}}$ and I_{RB}. If I_{RB} is not specified (default value ∞), the base series resistance is determined from

$$R_{\mathrm{BB'}} = R_{\mathrm{BM}} + \frac{R_{\mathrm{B}} - R_{\mathrm{BM}}}{q_{\mathrm{B}}} , \tag{3.156}$$

where R_{B} and R_{BM} are temperature-dependent quantities with linear and quadratic temperature coefficients T_{RB1}, T_{RB2}, T_{RM1} and T_{RM2}. Equation (3.156) takes account of the bias dependence of the sheet resistance of the base layer due to changes of the depletion-layer width and conductivity modulation caused by the diffusion charge. If a finite value of the knee current $I_{\mathrm{RB}} > 0$ is defined, the base series resistance can be calculated from

$$R_{\mathrm{BB'}} = R_{\mathrm{BM}} + 3(R_{\mathrm{B}}-R_{\mathrm{BM}})\frac{\tan z - z}{z \tan^2 z} , \tag{3.157}$$

where

$$z = \left(\sqrt{1 + \frac{144}{\pi^2}\frac{i_{\mathrm{B}}}{I_{\mathrm{RB}}}} - 1\right) \Bigg/ \left(\frac{24}{\pi^2}\sqrt{\frac{i_{\mathrm{B}}}{I_{\mathrm{RB}}}}\right) , \tag{3.158}$$

and $i_{\mathrm{B}} = i_{\mathrm{DE}} + i_{\mathrm{DC}} + i_{\mathrm{DLE}} + i_{\mathrm{DLC}}$. This corresponds to the results presented in [77] and [78].

The collector resistance $R_{\mathrm{CC'}}$ was originally represented by a temperature-dependent ohmic resistor with a resistance R_{C} at the nominal temperature T_0 and temperature coefficients T_{RC1} and T_{RC2}. This is, however, a poor description for larger values of the collector current, where velocity saturation and conductivity modulation affect the series resistance.

Specifying R_{C0} in the `.MODEL` statement causes $R_{\mathrm{CC'}}$ to be replaced by the network shown in Fig. 3.25, which in addition is connected to the internal base node B′. The current source i_{epi} is controlled by the voltages $v_{\mathrm{B'C'}}$ and $v_{\mathrm{B'C''}}$, with a voltage dependence as described in Sect. 3.7. The charges stored in the collector region are represented by Q_{w} and Q_0, with a bias dependence modeled by by the relations [89]

$$Q_0 = Q_{\mathrm{C0}}\left[K(v_{\mathrm{B'C'}}) - 1 - \gamma/2\right] , \tag{3.159}$$

$$Q_{\mathrm{w}} = Q_{\mathrm{C0}}\left[K(v_{\mathrm{B'C''}}) - 1 - \gamma/2\right] . \tag{3.160}$$

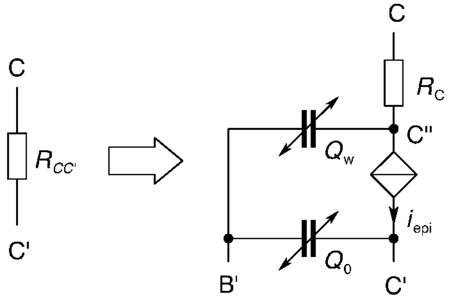


Fig. 3.25. Equivalent circuit to for taking account of quasi-saturation effects

The parameter Q_{C0} is defined as $Q_{C0} = eA_{je}W_{epi}N_{DC}/4$ and is equal (up to a factor of 1/4) to the charge of the ionized donors in the epilayer of the npn bipolar transistor.

3.8.6 Parameters

In the SPICE netlist, bipolar transistors are defined by a statement of the form

```
Q(name)   K_c   K_b   K_e   ⟨ K_s ⟩   Mname   ⟨ AREA ⟩ ,
```

where `K_c` denotes the name of the collector node, `K_b` the name of the base node, `K_e` the name of the emitter node and `K_s` the name of the substrate node.[36] `AREA` is a dimensionless quantity that serves as a scaling factor for area-dependent parameters: currents and depletion capacitances are multiplied by `AREA`, and resistances are divided by `AREA`. If `AREA` is not defined, `AREA = 1` is assumed; the following descriptions assume `AREA = 1`.

`Mname` specifies the model employed, the parameters of which are listed in a separate `.MODEL` statement. The following `.MODEL` statements are available for the description of vertical npn and pnp transistors, as well as lateral pnp transistors:

```
.MODEL   Mname   NPN    (model parameters) ,
.MODEL   Mname   PNP    (model parameters) ,
.MODEL   Mname   LPNP   (model parameters) .
```

[36]The definition of the substrate node may be omitted if `K_s` is zero the substrate node is assumed to be a common ground.

More than 50 parameters are available to characterize the large-signal behavior (cf. Table 3.3). If parameters are not specified in the `.MODEL` statement, the default values are used automatically.

Table 3.3. Parameters of SPICE transistor model

Parameter name	Parameter	Unit	Default
Ideal transistor			
Transfer saturation current	I_S, `IS`	A	10^{-16}
Ideal forward current gain	B_F, `BF`	–	100
Ideal reverse current gain	B_R, `BR`	–	1
Forward emission coefficient	N_F, `NF`	–	1
Reverse emission coefficient	N_R, `NR`	–	1
Early effect			
Forward Early voltage	V_{AF}, `VAF`	V	∞
Reverse Early voltage	V_{AR}, `VAR`	V	∞
Leakage diodes			
Saturation current of eb leakage diode	I_{SE}, `ISE`	A	0
Emission coefficient of eb leakage diode	N_E, `NE`	–	1.5
Saturation current of bc leakage diode	I_{SC}, `ISC`	A	0
Emission coefficient of bc leakage diode	N_C, `NC`	–	2.0
Saturation current of cs diode	I_{SS}, `ISS`	A	0
Emission coefficient of cs diode	N_S, `NS`	–	1
High-level injection			
Forward knee current	I_{KF}, `IKF`	A	∞
Reverse knee current	I_{KR}, `IKR`	A	∞
Roll-off exponent	N_K, `NK`	–	0.5
Series resistances			
Emitter resistance	R_E, `RE`	Ω	0
Base resistance (maximum)	R_B, `RB`	Ω	0
Base resistance (minimum)	R_{BM}, `RBM`	Ω	`RB`
Base resistance knee current	I_{RB}, `IRB`	A	∞
Collector series resistance	R_C, `RC`	Ω	0
Quasi-saturation			
Epilayer charge factor	Q_{CO}, `QCO`	C	0
Epilayer resistance	R_{CO}, `RCO`	Ω	0
Carrier mobility knee voltage	V_o, `VO`	V	10
Epilayer doping factor	γ, `GAMMA`	–	10^{-11}
Transit times			
Ideal forward transit time	T_F, `TF`	sec	0
Parameter for bias dependence	X_{TF}, `XTF`	–	0
Parameter for dependence on V_{BC}	V_{TF}, `VTF`	V	∞
Parameter for dependence on I_C	I_{TF} `ITF`	A	0
"Excess phase" at $f = 1/(2\pi\tau_f)$	P_{TF}, `PTF`	deg	0
Ideal reverse transit time	T_R, `TR`	sec	0

Table 3.3. (continued)

Parameter name	Parameter	Unit	Default
Depletion capacitances			
EB depletion capacitance ($V_{BE} = 0$)	C_{JE}, `CJE`	F	0
EB built-in voltage	V_{JE}, `VJE`	V	0.75
EB grading exponent	M_{JE}, `MJE`	–	0.33
BC depletion capacitance ($V_{BC} = 0$)	C_{JC}, `CJC`	F	0
BC built-in voltage	V_{JC}, `VJC`	V	0.75
BC grading exponent	M_{JC}, `MJC`	–	0.33
Fraction of internal bc capacitance	X_{CJC}, `XCJC`	–	1
CS depletion capacitance ($V_{CS} = 0$)	C_{JS}, `CJS`	F	0
CS built-in voltage	V_{JS}, `VJS`	V	0.75
CS grading exponent	M_{JS}, `MJS`	–	0
Limit for depletion-layer approximation	F_C, `FC`	–	0.5
Temperature dependences			
Bandgap voltage ($\approx W_g/e$)	E_G, `EG`	V	1.11
Temperature exponent of current gain	X_{TB}, `XTB`	–	0
Temperature exponent of I_S	X_{TI}, `XTI`	–	3
Temperature coefficient of R_E (linear)	T_{RE1}, `TRE1`	K^{-1}	0
Temperature coefficient of R_E (quadratic)	T_{RE2}, `TRE2`	K^{-2}	0
Temperature coefficient of R_B (linear)	T_{RB1}, `TRB1`	K^{-1}	0
Temperature coefficient of R_B (quadratic)	T_{RB2}, `TRB2`	K^{-2}	0
Temperature coefficient of R_{BM} (linear)	T_{RM1}, `TRM1`	K^{-1}	0
Temperature coefficient of R_{BM} (quadratic)	T_{RM2}, `TRM2`	K^{-2}	0
Temperature coefficient of R_C (linear)	T_{RC1}, `TRC1`	K^{-1}	0
Temperature coefficient of R_C (quadratic)	T_{RC2}, `TRC2`	K^{-2}	0
Noise			
$1/f$ noise coefficient	K_F, `KF`	A^2	0
$1/f$ noise exponent	A_F, `AF`	–	1

The SPICE Gummel–Poon model in the form summarized here is the standard description of the bipolar transistor. The approximations employed, however, limit the accuracy of the simulation results. The model provides only a crude description of the Early effect, and is not very accurate if the current gain and transit time in the presence of high-level-injection effects have to be modeled. The description of integrated bipolar transistors suffers from the poor description of the parasitic pnp transistor formed by the base layer, collector region and substrate. Furthermore, self-heating effects are not considered in this model. These deficiencies have led to the development of more sophisticated models that improve on the description of the standard Gummel–Poon model – at the cost of increased complexity of the model, however. Sections 3.14–3.16 summarize the improvements achieved by the VBIC, HICUM and MEXTRAM models.

3.9 Small-Signal Description

In the course of a `.AC` analysis, SPICE will first calculate the currents and voltages at the bias point and then, with the help of these data, the elements of the small-signal equivalent circuit. Extensions of this linearized model are required to take account of non-quasi-static effects and for the analysis of nonlinear distortion and intermodulation.

3.9.1 Giacoletto Small-Signal Equivalent Circuit

Figure 3.26 shows the Giacoletto model of an integrated bipolar transistor. Nodes of the small-signal model are denoted by lower-case letters, since only small-signal voltages, such as

$$v_{\mathrm{be}}(t) = v_{\mathrm{BE}}(t) - V_{\mathrm{BE}} \quad \text{or} \quad v_{\mathrm{bc}}(t) = v_{\mathrm{BC}}(t) - V_{\mathrm{BC}}$$

are applied between nodes of the small-signal model. To simplify the notation, the small-signal voltage $v_{\mathrm{b'c'}}$ is abbreviated to v_{π}.

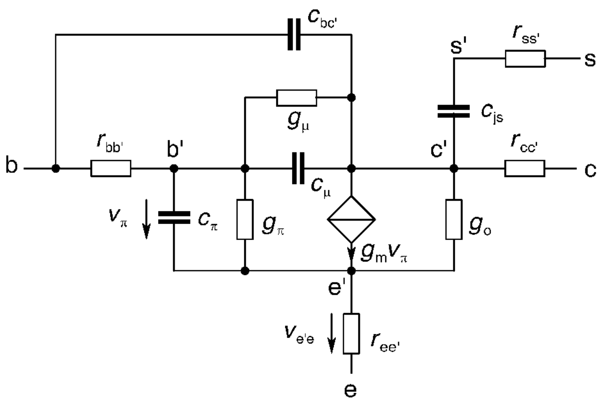


Fig. 3.26. Small-signal model of an integrated bipolar transistor derived from the Gummel–Poon model (small-signal substrate resistance $r_{\mathrm{ss'}}$ added)

The Giacoletto model is derived from the large-signal model by linearization of the network elements; the way from the Gummel–Poon large-signal model to the small-signal model shown in Fig. 3.26 requires the following steps:

- The nonlinear transfer current source is replaced by a parallel connection of a linear voltage-controlled current source $i_{\mathrm{t}} = g_{\mathrm{m}} v_{\pi}$, with an (internal) transconductance[37]

$$g_{\mathrm{m}} = \left(\partial I_{\mathrm{T}} / \partial V_{\mathrm{B'E'}}\right)_{V_{\mathrm{C'E'}}} \approx I_{\mathrm{T}} / V_{\mathrm{T}} , \tag{3.161}$$

[37] This is a very good approximation for hand calculation. In SPICE, a modified expression that also takes account of the reverse Early effect, which yields a correction of the order of $I_{\mathrm{T}}/V_{\mathrm{AR}}$, is used.

and an (internal) output conductance

$$g_{\mathrm{o}} = \left(\frac{\partial I_{\mathrm{T}}}{\partial V_{\mathrm{C'E'}}}\right)_{V_{\mathrm{B'E'}}} \approx -\frac{I_{\mathrm{T}}}{q_{\mathrm{B}}}\left(\frac{\partial q_{\mathrm{B}}}{\partial V_{\mathrm{C'E'}}}\right)_{V_{\mathrm{B'E'}}} , \tag{3.162}$$

which takes account of the change of the collector current with $V_{\mathrm{C'E'}}$ in forward operation, i.e. the slope of the output characteristic; I_{T} denotes the value of the transfer current at the bias point.

- The parallel diodes D_{E} and D_{LE} are replaced by the (internal) input conductance[38]

$$g_{\pi} = \frac{1}{r_{\pi}} = \frac{\mathrm{d}(I_{\mathrm{DE}}+I_{\mathrm{DLE}})}{\mathrm{d}V_{\mathrm{B'E'}}} \approx \frac{I_{\mathrm{DE}}}{V_{\mathrm{T}}} + \frac{I_{\mathrm{DLE}}}{N_{\mathrm{E}}V_{\mathrm{T}}} . \tag{3.163}$$

The large-signal base current I_{B} and transfer current I_{T} are related by the bias-dependent (internal) large-signal current gain $B_{\mathrm{N}}(I_{\mathrm{T}}, V_{\mathrm{C'E'}})$:

$$I_{\mathrm{T}} = B_{\mathrm{N}}(I_{\mathrm{T}}, V_{\mathrm{C'E'}}) I_{\mathrm{B}} .$$

This gives, after differentiation,

$$g_{\mathrm{m}} = \left(\frac{\partial I_{\mathrm{T}}}{\partial V_{\mathrm{B'E'}}}\right)_{V_{\mathrm{C'E'}}} = I_{\mathrm{B}}\left(\frac{\partial B_{\mathrm{N}}}{\partial V_{\mathrm{B'E'}}}\right)_{V_{\mathrm{C'E'}}} + B_{\mathrm{N}}\left(\frac{\partial I_{\mathrm{B}}}{\partial V_{\mathrm{B'E'}}}\right)_{V_{\mathrm{C'E'}}} .$$

Since, under forward-bias conditions, $I_{\mathrm{B}} \approx I_{\mathrm{DE}} + I_{\mathrm{DLE}}$, we may make the identification $g_{\pi} = \partial I_{\mathrm{B}}/\partial V_{\mathrm{B'E'}}$ to obtain

$$g_{\mathrm{m}} = \beta g_{\pi} , \tag{3.164}$$

where

$$\beta = \frac{B_{\mathrm{N}}}{1 - I_{\mathrm{B}}\left(\partial B_{\mathrm{N}}/\partial I_{\mathrm{T}}\right)_{V_{\mathrm{C'E'}}}} = B_{\mathrm{N}} + I_{\mathrm{B}}\left(\frac{\partial B_{\mathrm{N}}}{\partial I_{\mathrm{B}}}\right)_{V_{\mathrm{C'E'}}} \tag{3.165}$$

denotes the (internal) small-signal current gain. In the course of the computation of the bias point, SPICE determines both B_{N}, which is written as `BETADC` in the `.OUT` file, and β, which is written as `BETAAC` in the `.OUT` file.

- Under forward operation, the parallel diodes D_{C} and D_{LC} are reverse biased and generally carry a negligible current, resulting in a negligible conductance g_{μ} in the small-signal model. In the case of a forward-biased bc diode, the conductance $g_{\mu} = \mathrm{d}I_{\mathrm{BC}}/\mathrm{d}V_{\mathrm{B'C'}}$ must be added between nodes b′ and c′.

[38]This approximation omits the terms $I_{\mathrm{S}}/(B_{\mathrm{F}}V_{\mathrm{T}})$ and $I_{\mathrm{SE}}/(N_{\mathrm{E}}V_{\mathrm{T}})$, which are negligible under forward-bias conditions. For the purpose of numerical stability, small extra conductances, expressed in terms of the parameter `GMIN`, are added to g_{m}, g_{o}, g_{π} and g_{μ} in SPICE. The value of `GMIN` can be changed with a `.OPTIONS` statement, its default value is $G_{\mathrm{MIN}} = 10^{-12}/\Omega$. This parameter has to be chosen sufficiently small, in order that these unphysical terms do not substantially affect the simulation result. A detailed listing of the formulas employed in SPICE is given in [83].

A negligible conductance $g_\mu \approx 0$ under forward operation is a consequence of the generally negligible recombination in the base layer. The base current $I_\mathrm{B} = I_\mathrm{BE} + I_\mathrm{BB} + I_\mathrm{BC}$ in forward operation is then almost completely determined by I_BE, resulting in $\partial I_\mathrm{B}/\partial V_\mathrm{C'E'} \approx 0$. In the case of low-frequency operation, the equivalent circuit shown in Fig. 3.27 is obtained for the internal transistor.

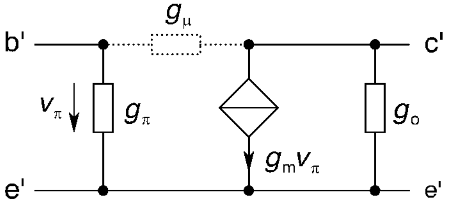


Fig. 3.27. Giacoletto model of internal transistor for low-frequency operation

In real transistors, however, some recombination occurs in the base layer, which causes an additional conductance $g_\mu = -\left(\partial I_\mathrm{BB}/\partial V_\mathrm{C'E'}\right)_{V_\mathrm{B'E'}}$ between nodes b$'$ and c$'$ (see Fig. 3.27). Denoting the forward base transit time by τ_Bf and the electron lifetime in the base layer by τ_n, the recombination current can be estimated from $I_\mathrm{BB} \approx I_\mathrm{T}\tau_\mathrm{Bf}/\tau_\mathrm{n}$, resulting in

$$\left(\frac{\partial I_\mathrm{BB}}{\partial V_\mathrm{C'E'}}\right)_{V_\mathrm{B'E'}} \approx \frac{I_\mathrm{T}}{\tau_\mathrm{n}}\left(\frac{\partial \tau_\mathrm{Bf}}{\partial V_\mathrm{C'E'}}\right)_{V_\mathrm{B'E'}} + \frac{\tau_\mathrm{Bf}}{\tau_\mathrm{n}}\left(\frac{\partial I_\mathrm{T}}{\partial V_\mathrm{C'E'}}\right)_{V_\mathrm{B'E'}} .$$

The derivatives can be estimated for a diffusion transistor, where $\tau_\mathrm{Bf} \approx d_\mathrm{B}^2/2D_\mathrm{n}$ and therefore

$$\left(\frac{\partial \tau_\mathrm{Bf}}{\partial V_\mathrm{C'E'}}\right)_{V_\mathrm{B'E'}} = \frac{d_\mathrm{B}}{D_\mathrm{n}}\left(\frac{\partial d_\mathrm{B}}{\partial V_\mathrm{C'E'}}\right)_{V_\mathrm{B'E'}} = \frac{2\tau_\mathrm{Bf}}{d_\mathrm{B}}\left(\frac{\partial d_\mathrm{B}}{\partial V_\mathrm{C'E'}}\right)_{V_\mathrm{B'E'}} .$$

As the transfer current I_T of a diffusion transistor varies in inverse proportion to d_B under low-level-injection conditions, the relation

$$\left(\frac{\partial I_\mathrm{T}}{\partial V_\mathrm{C'E'}}\right)_{V_\mathrm{B'E'}} = g_\mathrm{o} \approx -\frac{I_\mathrm{T}}{d_\mathrm{B}}\left(\frac{\partial d_\mathrm{B}}{\partial V_\mathrm{C'E'}}\right)_{V_\mathrm{B'E'}}$$

allows us to express $\partial d_\mathrm{B}/\partial V_\mathrm{C'E'}$ in terms of g_o, resulting in

$$g_\mu \approx g_\mathrm{o}\tau_\mathrm{Bf}/\tau_\mathrm{n} . \tag{3.166}$$

If hole injection into the emitter were negligible, virtually all of the holes would recombine in the base layer under forward operation ($I_\mathrm{B} \approx I_\mathrm{BB}$) and $\tau_\mathrm{n}/\tau_\mathrm{Bf}$ would equal the current gain B_N, resulting in $g_\mu = g_\mathrm{o}/B_\mathrm{N}$. Since in modern transistors $I_\mathrm{BE} \gg I_\mathrm{BB}$ and correspondingly $B_\mathrm{N}\tau_\mathrm{Bf} \ll \tau_\mathrm{n}$, the assumption $g_\mathrm{o}/B_\mathrm{N} \gg g_\mu \approx 0$ is justified. If g_μ has to be taken into account, g_π, g_m and g_o have to be calculated from

$$g_\pi = \left(\frac{\partial I_\mathrm{B}}{\partial V_{\mathrm{B'E'}}}\right)_{V_{\mathrm{C'E'}}} + \left(\frac{\partial I_\mathrm{B}}{\partial V_{\mathrm{C'E'}}}\right)_{V_{\mathrm{B'E'}}} ,$$

$$g_\mathrm{m} = \left(\frac{\partial I_\mathrm{C}}{\partial V_{\mathrm{B'E'}}}\right)_{V_{\mathrm{C'E'}}} - \left(\frac{\partial I_\mathrm{B}}{\partial V_{\mathrm{C'E'}}}\right)_{V_{\mathrm{B'E'}}} ,$$

$$g_\mathrm{o} = \left(\frac{\partial I_\mathrm{C}}{\partial V_{\mathrm{C'E'}}}\right)_{V_{\mathrm{C'E'}}} + \left(\frac{\partial I_\mathrm{B}}{\partial V_{\mathrm{C'E'}}}\right)_{V_{\mathrm{B'E'}}}$$

to be in accordance with the general π-equivalent circuit of a two-port in the admittance representation [95].

• The depletion capacitance of the eb diode c_je and the diffusion capacitance associated with the eb diode are calculated at the bias point and combined to give the eb capacitance:

$$c_\pi = c_\mathrm{je}(V_{\mathrm{B'E'}}) + \tau_\mathrm{f} g_\mathrm{m} . \tag{3.167}$$

In the `.OUT` file of a SPICE run, c_π is printed as `CBE`.

• The internal part of the bc depletion capacitance c_jc gives the internal bc capacitance c_μ, which is written to the `.OUT` file as `CBC`. The diffusion capacitance associated with the bc diode can be neglected under forward operation. The external part of the bc capacitance $c_{\mathrm{bc'}}$ is written to the `.OUT` file as `CBX`.

• The series resistances $R_{\mathrm{BB'}}$, $R_{\mathrm{CC'}}$ and $R_{\mathrm{EE'}}$ are substituted by their respective small-signal series resistances $r_{\mathrm{bb'}}$, $r_{\mathrm{cc'}}$ and $r_{\mathrm{ee'}}$ at the bias point. This applies in particular to the base resistance

$$r_{\mathrm{bb'}} = R_{\mathrm{BB'}} + I_\mathrm{B} \frac{\mathrm{d}R_{\mathrm{BB'}}}{\mathrm{d}I_\mathrm{B}} , \tag{3.168}$$

which shows a strong dependence on I_B. The value of $r_{\mathrm{bb'}}$ is written to the `.OUT` file of a SPICE run as `RX`.

3.9.2 Admittance Parameters

The equivalent circuit[39] depicted in Fig. 3.28 is used to compute the admittance parameters. The small-signal base and collector currents of the internal transistor located between nodes b′, c′ and e′ are given by

$$\underline{i}_\mathrm{b} = [g_\pi + \mathrm{j}\omega(c_\pi + c_\mu)]\,\underline{v}_\pi - \mathrm{j}\omega c_\mu \underline{v}_{\mathrm{c'e'}} ,$$

$$\underline{i}_{\mathrm{c'}} = (g_\mathrm{m} - \mathrm{j}\omega c_\mu)\,\underline{v}_\pi + (g_\mathrm{o} + \mathrm{j}\omega c_\mu)\,\underline{v}_{\mathrm{c'e'}} ,$$

as can easily be seen by applying Kirchhoff's current law to nodes b′ and c′, taking account of the fact that $\underline{v}_{\mathrm{b'c'}} = \underline{v}_\pi - \underline{v}_{\mathrm{c'e'}}$.

[39] The dielectric behavior of the substrate may become relevant at high frequencies, owing to the comparatively large resistivity ρ of the substrate. If this is the case, $r_{\mathrm{ss'}}$ is shunted by a capacitance $c_{\mathrm{ss'}} = \rho\epsilon / r_{\mathrm{ss'}}$ (see [96, 97] for further discussion).

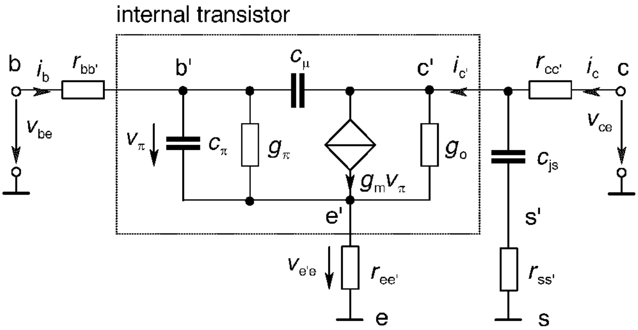


Fig. 3.28. Simplified small-signal model for computation of the admittance parameters in the common-emitter configuration

Owing to the presence of the cs capacitance c_{js}, the small-signal current at the collector terminal of an integrated transistor is given by

$$\underline{i}_c = \underline{i}_{c'} + \frac{j\omega c_{js}}{1+j\omega r_{ss'}c_{js}}\underline{v}_{c'} = \underline{i}_{c'} + y_s\underline{v}_{c'} ;$$

taking the emitter series resistance $r_{ee'}$ into account gives

$$\begin{pmatrix} \underline{i}_b \\ \underline{i}_c \end{pmatrix} = \frac{1}{\Xi}\begin{pmatrix} y'_{11} & y'_{12} \\ y'_{21} & y'_{22} \end{pmatrix}\begin{pmatrix} \underline{v}_{b'} \\ \underline{v}_{c'} \end{pmatrix} , \tag{3.169}$$

where

$$\Xi = 1 + r_{cc'}(g_m+g_\pi+g_o) + j\omega r_{cc'}c_\pi \tag{3.170}$$

and

$$\begin{aligned} y'_{11} &= g_\pi + j\omega c_\pi - y'_{12} , \\ y'_{12} &= -j\omega c_\mu\,\Xi - r_{ee'}g_o\,(g_\pi+j\omega c_\pi) , \\ y'_{21} &= g_m + y'_{12} , \\ y'_{22} &= g_o + \Xi y_s - y'_{12} . \end{aligned}$$

Equation (3.169) relates the small-signal terminal currents to the internal voltages $\underline{v}_{b'}$ and $\underline{v}_{c'}$. These are related to the small-signal terminal voltages by the relations $\underline{v}_b = \underline{v}_{b'} + r_{bb'}\underline{i}_b$ and $\underline{v}_c = \underline{v}_{c'} + r_{cc'}\underline{i}_c$, or, written in matrix form,

$$\begin{pmatrix} \underline{v}_b \\ \underline{v}_c \end{pmatrix} = \frac{1}{\Xi}\begin{pmatrix} \Xi+r_{bb'}y'_{11} & r_{bb'}y'_{12} \\ r_{cc'}y'_{21} & \Xi+r_{cc'}y'_{22} \end{pmatrix}\begin{pmatrix} \underline{v}_{b'} \\ \underline{v}_{c'} \end{pmatrix} .$$

Inversion of this matrix equation and substitution into (3.169) yields

$$\begin{pmatrix} \underline{i}_b \\ \underline{i}_c \end{pmatrix} = \frac{1}{K}\begin{pmatrix} y'_{11}+r_{cc'}\Delta & y'_{12} \\ y'_{21} & y'_{22}+r_{bb'}\Delta \end{pmatrix}\begin{pmatrix} \underline{v}_b \\ \underline{v}_c \end{pmatrix} ,$$

where $\Delta = (y'_{11}y'_{22} - y'_{12}y'_{21})/\Xi$ and

$$K \;=\; \Xi + r_{\mathrm{bb'}}y'_{11} + r_{\mathrm{cc'}}y'_{22} + r_{\mathrm{bb'}}r_{\mathrm{cc'}}\Delta \;.$$

If we neglect terms of order $r_{\mathrm{ee'}}g_{\mathrm{o}}$, $r_{\mathrm{bb'}}g_{\mathrm{o}}$ and $r_{\mathrm{cc'}}g_{\mathrm{o}}$, we may make the approximation

$$\Delta \;\approx\; \mathrm{j}\omega c_\mu(g_{\mathrm{m}}+g_\pi) + y_{\mathrm{s}}\left[g_\pi + \mathrm{j}\omega(c_\pi + \Xi c_\mu)\right] \;,$$

and

$$\begin{aligned} K \;\approx\; & \Xi + r_{\mathrm{bb'}}g_\pi + \mathrm{j}\omega r_{\mathrm{bb'}}c_\pi + \mathrm{j}\omega r_{\mathrm{bb'}}c_\mu\Xi + \mathrm{j}\omega r_{\mathrm{cc'}}c_\mu\Xi + r_{\mathrm{cc'}}\Xi y_{\mathrm{s}} \\ & + \mathrm{j}\omega r_{\mathrm{cc'}}c_\mu r_{\mathrm{bb'}}(g_{\mathrm{m}}+g_\pi) + r_{\mathrm{bb'}}r_{\mathrm{cc'}}y_{\mathrm{s}}\left[g_\pi + \mathrm{j}\omega(c_\pi+\Xi c_\mu)\right] \;. \end{aligned}$$

In the limit $\omega \to 0$ we obtain

$$K \;\to\; 1 + r_{\mathrm{cc'}}(g_{\mathrm{m}}+g_\pi) + r_{\mathrm{bb'}}g_\pi \;=\; 1 + I_{\mathrm{C}}/I_{\mathrm{C}}^* \;.$$

The factor $1/K$ then describes the reduction of the dc small-signal conductances caused by the series resistances. The value of

$$I_{\mathrm{C}}^* = \frac{V_{\mathrm{T}}}{r_{\mathrm{cc'}} + (r_{\mathrm{cc'}}+r_{\mathrm{bb'}})/\beta} \tag{3.171}$$

determines a critical level for the collector current I_{C}: if I_{C} reaches or exceeds I_{C}^*, series resistances will have a severe influence on the dc conductances, but if $I_{\mathrm{C}} \ll I_{\mathrm{C}}^*$, the factor $1/K$ may be replaced by one, under low-frequency conditions. At higher frequencies, we may write

$$K \;\approx\; (1 + I_{\mathrm{C}}/I_{\mathrm{C}}^*)\,(1 + \mathrm{j}f/f_{\mathrm{y}}) \;,$$

if $\omega r_{\mathrm{ss'}}c_{\mathrm{js}} \ll 1$ and only terms of first order in ω are retained; the transconductance cutoff frequency f_{y} is given by

$$\begin{aligned} \frac{1}{2\pi f_{\mathrm{y}}} \;=\; & \frac{(r_{\mathrm{bb'}}+r_{\mathrm{ee'}})c_\pi + \left[1 + r_{\mathrm{ee'}}(g_{\mathrm{m}}+g_\pi) + r_{\mathrm{cc'}}g_{\mathrm{m}}\right]r_{\mathrm{bb'}}c_\mu}{1 + I_{\mathrm{C}}/I_{\mathrm{C}}^*} \\ & + r_{\mathrm{cc'}}(c_\mu + c_{\mathrm{js}}) \;. \end{aligned} \tag{3.172}$$

If the collector series resistance $r_{\mathrm{cc'}}$ is much smaller than the substrate resistance $r_{\mathrm{ss'}}$, as is frequently the case, the cs capacitance c_{js} will short-circuit the branch between nodes c′ and s′, resulting in $y_{\mathrm{s}} \approx 1/r_{\mathrm{ss'}}$. Neglecting terms of order $r_{\mathrm{cc'}}/r_{\mathrm{ss'}}$ and making the approximation $1 + 1/\beta \;\approx\; 1$ then gives

$$\frac{1}{2\pi f_{\mathrm{y}}} \;=\; r_{\mathrm{cc'}}c_\mu + \frac{(r_{\mathrm{bb'}}+r_{\mathrm{ee'}})c_\pi + \left[1 + (r_{\mathrm{ee'}}+r_{\mathrm{cc'}})g_{\mathrm{m}}\right]r_{\mathrm{bb'}}c_\mu}{1 + I_{\mathrm{C}}/I_{\mathrm{C}}^*} \;;$$

in this case the cs capacitance has no effect on the transconductance cutoff frequency. The factor $1/(1+\mathrm{j}f/f_{\mathrm{y}})$ which is common to all of the admittance parameters, describes a low-pass frequency characteristic, which takes account of the fact that the capacitances c_π and c_μ have to be charged via the series resistances. With the approximations described, the input admittance y_{11e} reads

$$y_{11e} = \frac{1}{1+I_C/I_C^*}\frac{1}{1+jf/f_y} g_\pi \left(1+j\frac{f}{f_\beta}\right) , \tag{3.173}$$

where f_β is the β-cutoff frequency, determined by

$$\frac{1}{2\pi f_\beta} = \frac{c_\pi + c_\mu}{g_\pi} + \beta(r_{ee'}+r_{cc'})c_\mu . \tag{3.174}$$

For low frequencies f, the input admittance is real and approximately independent of frequency:

$$y_{11e} \approx \frac{1}{1+I_C/I_C^*}\frac{g_m}{\beta} = \frac{1}{r_\pi + r_{bb'} + r_{ee'}(\beta+1)} = g_{11e} .$$

If the β-cutoff frequency is exceeded, a capacitive behavior dominates, which causes an increase in $|y_{11e}| \sim f$. This behavior is compensated by the additional low-pass characteristic that becomes relevant for $f > f_y$, which causes y_{11e} to approach the value (Fig. 3.29a)

$$y_{11e} \approx \frac{1}{r_{bb'}+r_{ee'}}$$

if $c_\pi \gg c_\mu$. The (complex) transconductance y_{21e} is given by

$$y_{21e} \approx \frac{1}{1+I_C/I_C^*}\frac{1}{1+jf/f_y}\left[g_m - j2\pi f(1+g_m r_{ee'})c_\mu\right] ; \tag{3.175}$$

at low frequencies, the transconductance is approximately given by

$$y_{21e} \approx \frac{g_m}{1+I_C/I_C^*} = g_{21e} , \tag{3.176}$$

while at frequencies in excess of f_y a low-pass frequency characteristic is observed, as illustrated in Fig. 3.29c:

$$y_{21e}(f) \approx \frac{g_{21e}}{1+jf/f_y} . \tag{3.177}$$

The output admittance y_{22e} is given by

$$y_{22e} \approx \frac{1}{1+I_C/I_C^*}\frac{1}{1+jf/f_y}\left[g_o + j2\pi f(1+\delta)c_\mu\right] , \tag{3.178}$$

where $\delta = g_m(r_{bb'} + r_{ee'})$; its value is approximately constant at low frequencies, i.e.

$$y_{22e} \approx \frac{g_o}{1+I_C/I_C^*} = g_{22e} . \tag{3.179}$$

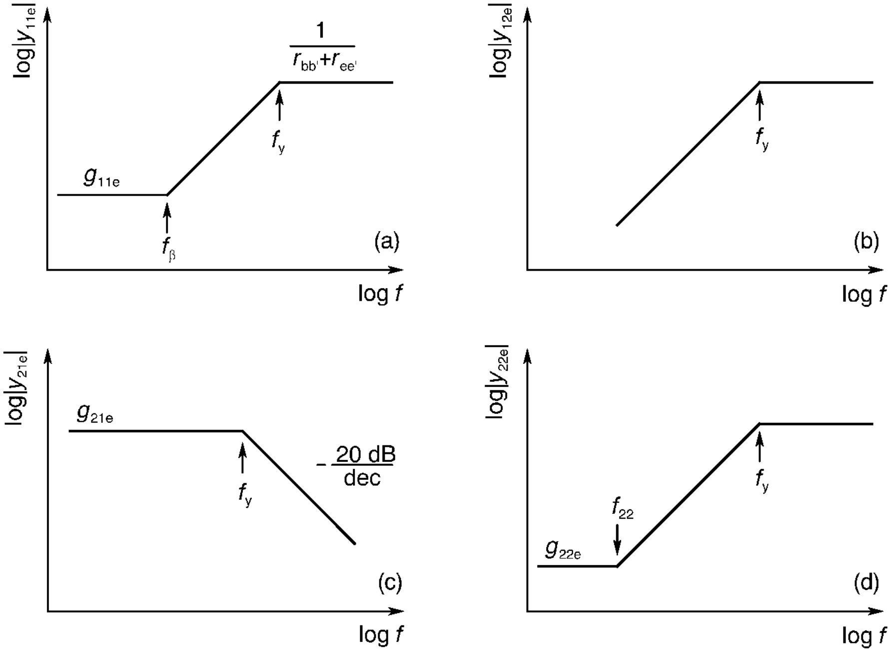


Fig. 3.29. Schematic representation of the frequency-dependent admittance parameters in a Bode diagram. (**a**) Input conductance, (**b**) reverse conductance, (**c**) transconductance and (**d**) output conductance

For frequencies in excess of

$$f_{22} = \frac{g_o}{2\pi\,[\,1 + g_m(r_{bb'} + r_{ee'})\,]\,c_\mu}\,, \tag{3.180}$$

the imaginary part of the numerator in the expression for y_{22e} dominates, resulting in an increase of $|y_{22e}| \sim f$, which is compensated by the low-pass behavior observed for $f > f_y$. The value in y_{22e} then converges to (Fig. 3.29d)

$$y_{22e} \approx g_{22e} f_y / f_{22}\,.$$

Since $|y_{22e}|$ is also determined by terms of order f^2, deviations from this simplified description are observed in practice. The reverse transconductance

$$y_{12e} = \frac{-1}{1 + I_C/I_C^*}\,\frac{1}{1 + \mathrm{j}f/f_y}\,\mathrm{j}2\pi f(1 + g_m r_{ee'})c_\mu \tag{3.181}$$

shows a capacitive behavior, which is compensated by the low-pass characteristic for $f > f_y$. If $c_\pi \gg c_\mu$, the value of y_{12e} then approaches

$$y_{12e} \approx -\frac{(1 + g_m r_{ee'})c_\mu}{(r_{bb'} + r_{ee'})c_\pi}\,,$$

as illustrated in Fig. 3.29b.

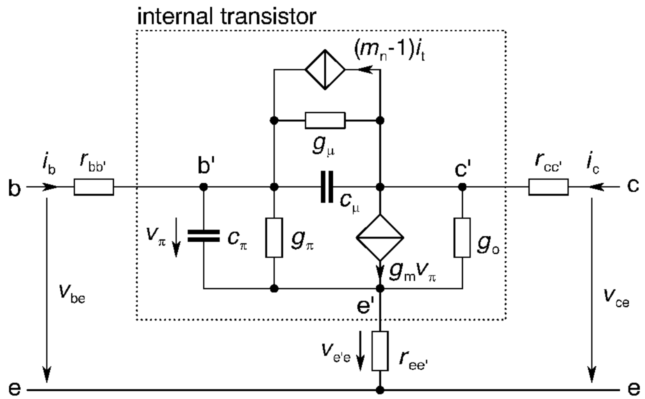


Fig. 3.30. Small-signal equivalent circuit for a bipolar transistor operated in the avalanche regime

3.9.3 Carrier Multiplication Effects

For a small-signal description of a bipolar transistor operated in the avalanche regime, the equivalent circuit depicted in Fig. 3.30 has to be used. The (nonlinear) controlled source $i_\mathrm{m} = (M_\mathrm{n}-1)i_\mathrm{T}$ that describes carrier injection into the base region due to carrier multiplication in the bc space charge region in the large-signal model has to be replaced by a (linear) controlled source and an ohmic conductivity g_μ in parallel. This can be seen if small-signal changes

$$\Delta I_\mathrm{m} = (M_\mathrm{n}-1)\,\Delta I_\mathrm{T} + \Delta M_\mathrm{n}\, I_\mathrm{T}$$

of the injected hole current I_m are considered. Since

$$\Delta M_\mathrm{n} = \left(\frac{\partial M_\mathrm{n}}{\partial V_\mathrm{C'B'}}\right)_{I_\mathrm{T}} \Delta V_\mathrm{C'B'} + \left(\frac{\partial M_\mathrm{n}}{\partial I_\mathrm{T}}\right)_{V_\mathrm{C'B'}} \Delta I_\mathrm{T}\,,$$

we may write

$$\Delta I_\mathrm{m} = (m_\mathrm{n}-1)\,\Delta I_\mathrm{T} + g_\mu\,\Delta V_\mathrm{C'B'}\,, \tag{3.182}$$

where

$$m_\mathrm{n} = M_\mathrm{n} + I_\mathrm{T}\left(\frac{\partial M_\mathrm{n}}{\partial I_\mathrm{T}}\right)_{V_\mathrm{C'B'}} \tag{3.183}$$

denotes the small-signal multiplication factor due to injected electrons, while

$$g_\mu = I_\mathrm{T}\left(\frac{\partial M_\mathrm{n}}{\partial V_\mathrm{C'B'}}\right)_{I_\mathrm{T}} \tag{3.184}$$

is the avalanche conductance [6]. Implementing (3.182) in the form of a small-signal equivalent circuit leads to Fig. 3.30. The value of g_μ in the weak avalanche regime may be estimated from the modified Miller formula (3.67), which gives

$$\frac{\partial M_\mathrm{n}}{\partial V_{\mathrm{C'B'}}} \approx M_\mathrm{n}^2 N_1 \left(\frac{V_{\mathrm{C'B'}}}{BV_\mathrm{c}}\right)^{N_1} \frac{1}{V_{\mathrm{C'B'}}} = \frac{N_1 M_\mathrm{n}(M_\mathrm{n}-1)}{V_{\mathrm{C'B'}}} ,$$

and therefore

$$g_\mu \approx \frac{N_1 M_\mathrm{n}(M_\mathrm{n}-1)I_\mathrm{T}}{V_{\mathrm{C'B'}}} \approx \frac{N_1 M_\mathrm{n}(M_\mathrm{n}-1)V_\mathrm{T}}{V_{\mathrm{C'B'}}} g_\mathrm{m} ; \tag{3.185}$$

obviously, g_μ is small for values of $M_\mathrm{n} - 1 \ll 1$.

Two-Port Parameters

Neglecting series resistances, the admittance parameters are given by

$$y_{11\mathrm{e}} = g_\pi - (m_\mathrm{n}-1)g_\mathrm{m} + g_\mu + \mathrm{j}\omega(c_\pi + c_\mu) , \tag{3.186}$$

$$y_{12\mathrm{e}} = -g_\mu - \mathrm{j}\omega c_\mu - (m_\mathrm{n}-1)g_\mathrm{o} , \tag{3.187}$$

$$y_{21\mathrm{e}} = m_\mathrm{n} g_\mathrm{m} - g_\mu - \mathrm{j}\omega c_\mu , \tag{3.188}$$

$$y_{22\mathrm{e}} = g_\mathrm{o} + g_\mu + (m_\mathrm{n}-1)g_\mathrm{o} + \mathrm{j}\omega c_\mu . \tag{3.189}$$

These are well defined for all values of m_n, i.e. no instability is observed as V_{CE} approaches BV_{CEO}. In the presence of series resistances, however, singular behavior is observed, as can easily be seen if a nonnegligible base series resistance $r_{\mathrm{bb'}}$ is added to the equivalent circuit.

If series resistances may be neglected, the hybrid parameters of the internal transistor (depicted within the dotted frame in Fig. 3.30) are easily derived. As an example, we consider the current gain $h_{21\mathrm{e}}$, determined from

$$h_{21\mathrm{e}} = \frac{m_\mathrm{n}\beta - r_\pi g_\mu - \mathrm{j}\omega r_\pi c_\mu}{1 - \beta(m_\mathrm{n}-1) + r_\pi g_\mu + \mathrm{j}\omega/\omega_\beta} , \tag{3.190}$$

where $\beta = g_\mathrm{m}/g_\pi$ and $\omega_\beta = [r_\pi(c_\pi + c_\mu)]^{-1}$. The term $r_\pi g_\mu$ can be rewritten as

$$r_\pi g_\mu \approx r_\pi g_\mathrm{m} V_\mathrm{T} \left(\frac{\partial M_\mathrm{n}}{\partial V_{\mathrm{C'B'}}}\right)_{I_\mathrm{T}} = \beta V_\mathrm{T} \left(\frac{\partial M_\mathrm{n}}{\partial V_{\mathrm{C'B'}}}\right)_{I_\mathrm{T}} .$$

Introducing the modified small-signal multiplication factor

$$m_\mathrm{n}' = m_\mathrm{n} - V_\mathrm{T} \left(\frac{\partial M_\mathrm{n}}{\partial V_{\mathrm{C'B'}}}\right)_{I_\mathrm{T}}$$

and neglecting the term $\mathrm{j}\omega r_\pi c_\mu$ now allows us to write (3.190) in the form

$$h_{21\mathrm{e}} \approx \frac{\beta'}{1 + \mathrm{j}\omega/\omega_\beta'} , \quad \text{where} \quad \beta' = \frac{m_\mathrm{n}'\beta}{1 - \beta(m_\mathrm{n}'-1)}$$

and $\omega_\beta' = [1 - \beta(m_\mathrm{n}'-1)]\,\omega_\beta$. In the limit $\omega \to 0$, the parameter $h_{21\mathrm{e}}$ – like all other hybrid parameters – shows a singularity as $(m_\mathrm{n}'-1)\beta$ approaches unity. The zero-frequency value β' of h_{21e} increases without limit under these conditions, while the beta cutoff frequency ω'_β converges to zero.

3.9.4 Non-Quasi-Static Effects and Excess Phase

At high frequencies, non-quasi-static effects become relevant and require a modification of the (quasi-static) Giacoletto model. In particular, the measured phase shift of h_{21e} exceeds the value predicted by the quasi-static model. For this purpose, SPICE allows one to modify the internal transconductance of the device by introduction of an additional phase shift factor:

$$g_\mathrm{m} \rightarrow g_\mathrm{m}\, \underline{\phi}(\mathrm{j}\omega) \,.$$

A frequently applied method for modeling the excess phase uses a second-order Bessel polynomial network [98] with a phase shift factor

$$\underline{\phi}(\mathrm{j}\omega) = \frac{1}{1 + \mathrm{j}(\omega/\omega_0) - \omega^2/(3\omega_0^2)} \,, \tag{3.191}$$

where $\omega_0 = 1/(P_\mathrm{TF}\tau_\mathrm{f})$ is expressed in terms of a multiple of the forward transit time τ_f by introducing the additional parameter P_TF. The phase shift

$$\varphi = \arctan \frac{3\omega_0\omega}{3\omega_0^2 - \omega^2}$$

produced by the function $\underline{\phi}(\mathrm{j}\omega)$ can be approximated by $\omega\tau_1$, where $\tau_1 = P_\mathrm{TF}\tau_\mathrm{f}$. If an ac analysis is to be performed, the forward transconductance is simply multiplied by the phase shift factor. For the purposes of transient analysis, the device equations have to be solved in the time domain. Here $\mathrm{j}\omega$ is replaced by s and the phase shift factor corresponds to a second-order differential equation, which determines the delayed transfer current $i_\mathrm{Tx}(t)$ from the undelayed for transfer current $i_\mathrm{Tf}(t)$ according to

$$\frac{1}{3\omega_0^2}\frac{\mathrm{d}i_\mathrm{Tx}}{\mathrm{d}t^2} + \frac{1}{\omega_0}\frac{\mathrm{d}i_\mathrm{Tx}}{\mathrm{d}t} + i_\mathrm{Tx} = i_\mathrm{Tf} \,. \tag{3.192}$$

A systematic analysis of non-quasi-static effects is possible for the drift transistor investigated in Sect. 3.1. A small-signal approximation of the general nonlinear relations (Appendix D) yields the following for the internal transistor in common-base configuration (hole current and depletion capacitances are neglected):

$$y_\mathrm{n11b} = g_\mathrm{m}\, \mathrm{e}^{-\eta/2} \sinh\left(\frac{\eta}{2}\right) \left(\frac{\underline{\vartheta}}{\eta/2}\coth(\underline{\vartheta}) + 1\right) \tag{3.193}$$

and

$$y_\mathrm{n21b} = g_\mathrm{m}\, \frac{\underline{\vartheta}}{\eta/2}\, \frac{\sinh(\eta/2)}{\sinh(\underline{\vartheta})} \,,$$

where

$$\underline{\vartheta} = \sqrt{(\eta/2)^2 + 2\mathrm{j}\omega\tau_{\mathrm{B0}}} \ .$$

Here recombination in the base region is neglected. If, in addition, the delay due to the traversal of the bc depletion layer is taken into account, y_{n21b} has to be multiplied by the factor (see Appendix D)

$$\frac{\sin(\omega\tau_{\mathrm{jc}})}{\omega\tau_{\mathrm{jc}}}\,\mathrm{e}^{-\mathrm{j}\omega\tau_{\mathrm{jc}}} \approx \mathrm{e}^{-\mathrm{j}\omega\tau_{\mathrm{jc}}} \ ,$$

where $\tau_{\mathrm{jc}} = d_{\mathrm{jc}}/2v_{\mathrm{nsat}}$ is the collector transit time. From this result, the input admittance $y_{\mathrm{n11e}} = y_{\mathrm{n11b}} - y_{\mathrm{n21b}}$ in the common-emitter configuration can be derived as

$$\begin{aligned} y_{\mathrm{n11e}} &= g_{\mathrm{m}}\,\frac{\sinh(\eta/2)}{\sinh(\underline{\vartheta})}\,\frac{\underline{\vartheta}}{\eta/2} \\ &\quad \times \left[\mathrm{e}^{-\eta/2}\left(\cosh(\underline{\vartheta}) + \frac{\eta/2}{\underline{\vartheta}}\sinh(\underline{\vartheta})\right) - \frac{\sin(\omega\tau_{\mathrm{jc}})}{\omega\tau_{\mathrm{jc}}}\,\mathrm{e}^{-\mathrm{j}\omega\tau_{\mathrm{jc}}}\right] . \end{aligned}$$

This expression can be developed up to second order in ω resulting in

$$y_{\mathrm{n11e}} = \mathrm{j}\omega c_{\mathrm{tne}}(1-\mathrm{j}\omega\tau_2) \ , \tag{3.194}$$

where $c_{\mathrm{tne}} = \tau_{\mathrm{f}} g_{\mathrm{m}}$; the forward transit time τ_{f} is determined by the base transit time τ_{Bf} defined in (3.29) and the collector transit time τ_{jc}, i.e.

$$\tau_{\mathrm{f}} = \tau_{\mathrm{Bf}} + \tau_{\mathrm{jc}} \ ,$$

and τ_2 is defined in Appendix D. Up to second order of ω this result is equivalent to the modeling approach presented by te Winkel and Seitchik [10, 99, 100], who used the approximation[40]

$$y_{\mathrm{n11e}} \approx \mathrm{j}\omega c_{\mathrm{tne}}\mathrm{e}^{-\mathrm{j}\omega\tau_2} \approx \frac{\mathrm{j}\omega c_{\mathrm{tne}}}{1+\mathrm{j}\omega\tau_2} \ .$$

In complete analogy, the transconductance $y_{\mathrm{n21e}} = y_{\mathrm{n21b}}$ of the internal transistor;

$$y_{\mathrm{n21e}} = g_{\mathrm{m}}\,\frac{\underline{\vartheta}}{\eta/2}\,\frac{\sinh(\eta/2)}{\sinh(\underline{\vartheta})}\,\frac{\sin(\omega\tau_{\mathrm{jc}})}{\omega\tau_{\mathrm{jc}}}\,\mathrm{e}^{-\mathrm{j}\omega\tau_{\mathrm{jc}}} \ ,$$

is approximated in the form

$$y_{\mathrm{n21e}} \approx g_{\mathrm{m}}\mathrm{e}^{-\mathrm{j}\omega(\tau_1+\tau_2)} \approx g_{\mathrm{m}}\,\frac{\mathrm{e}^{-\mathrm{j}\omega\tau_1}}{1+\mathrm{j}\omega\tau_2} \ ,$$

[40] Here a modified formulation is presented that takes account of the collector transit time.

where τ_1 determines the excess phase. A small-signal equivalent circuit that corresponds to the approximations chosen by Seitchik [99],

$$y_{\mathrm{n11e}} \approx \frac{\mathrm{j}\omega c_{\mathrm{tne}}}{1+\mathrm{j}\omega\tau_2} \quad \text{and} \quad y_{\mathrm{n21e}} \approx g_{\mathrm{m}} \frac{\mathrm{e}^{-\mathrm{j}\omega\tau_1}}{1+\mathrm{j}\omega\tau_2} ,$$

is shown in Fig. 3.31 in comparison with its quasi-static counterpart.

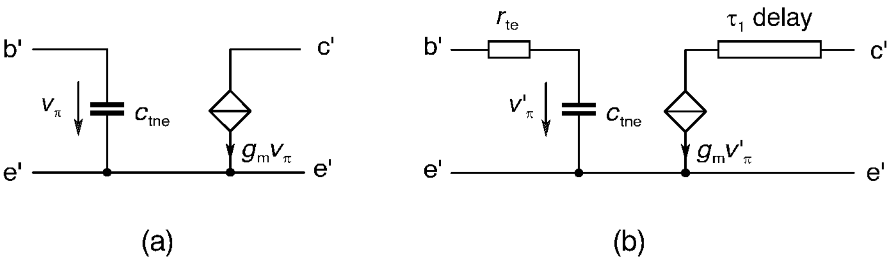


Fig. 3.31. Small-signal equivalent circuit of the internal bipolar transistor, describing electron transport across the base region. (**a**) quasi-static approximation, (**b**) non-quasi-static approximation according to [99]

If the small-signal hole-current component is written as $g_\pi \underline{v}_\pi = g_{\mathrm{m}} \underline{v}_\pi / \beta$, the internal forward common-emitter current gain is obtained as

$$h'_{21\mathrm{e}} = \frac{\beta \mathrm{e}^{-\mathrm{j}\omega\tau_1}}{1+\mathrm{j}\omega\tau_2} .$$

In the small-signal model, the time constant τ_2 can be represented as an RC time constant by introducing an additional series resistance [100]

$$r_{\mathrm{te}} = \frac{\tau_2}{c_{\mathrm{tne}}} = \frac{1}{g_{\mathrm{m}}} \frac{\tau_2}{\tau_{\mathrm{f}}} \tag{3.195}$$

in series with the diffusion capacitance c_{tne}. The time constant τ_1 is represented by a delay line or a voltage-controlled transfer current source with a complex transconductance.

AC Current Crowding. Owing to the distributed nature of the internal base resistance and the depletion and diffusion capacitances, the small-signal response of the transistor input behaves like a transmission line, as depicted in Fig. 3.32a [78, 101], with a short-circuit input impedance

$$h_{11\mathrm{e}} = Z_0 \coth(\underline{\gamma} L_{\mathrm{E}}) .$$

Here

$$Z_0 = \sqrt{\frac{r'}{g'+\mathrm{j}\omega c'}} \quad \text{and} \quad \underline{\gamma} = \sqrt{r'(g'+\mathrm{j}\omega c')}$$

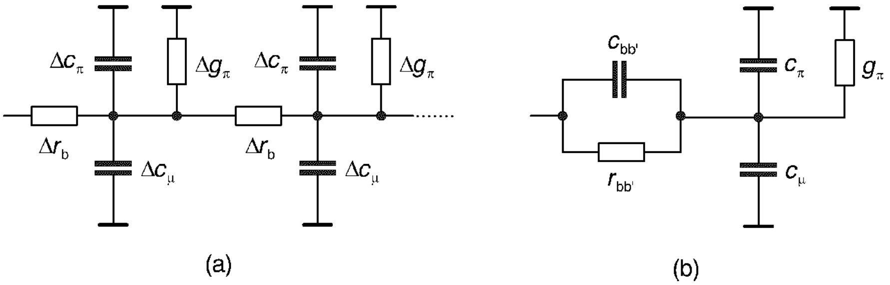


Fig. 3.32. (**a**) Transmission-line equivalent circuit for the small-signal base current in the absence of dc current crowding and conductivity modulation; (**b**) simplified equivalent circuit

denote the characteristic impedance and the propagation constant, respectively. The series impedance r' per unit length of the transmission line is obtained from the sheet resistance of the base layer as $r' = R_\pi/W_E = 3R_B/L_E$, where L_E denotes the extension of the base layer in the direction of the current flow, and W_E the extension orthogonal to the current flow (see Fig. 3.18); the shunt admittance per unit length $g' + j\omega c'$ is determined from $g' = g_\pi/L_E$ and $c' = (c_\pi + c_\mu)/L_E$. This gives

$$h_{11e} = \frac{\gamma L_E \coth(\gamma L_E)}{g_\pi + j\omega(c_\pi + c_\mu)} ,$$

an expression that reduces to

$$h'_{11e} = \frac{1}{g_\pi + j\omega(c_\pi + c_\mu)}$$

in the limit $r' \to 0$. The difference $h_{11e} - h'_{11e}$ can be identified with the base series impedance

$$z_{bb'} = \frac{\gamma L_E \coth(\gamma L_E) - 1}{g_\pi + j\omega(c_\pi + c_\mu)} .$$

Making use of the series expansion

$$x \coth x \approx 1 + \frac{1}{3}x^2 - \frac{1}{45}x^4$$

and the identity $r'L_E/3 = r_{bb'}$, we obtain the following approximation:

$$z_{bb'} \approx r_{bb'}\left(1 - j\omega r_{bb'}\frac{c_\pi + c_\mu}{5}\right) \approx \frac{r_{bb'}}{1 + j\omega r_{bb'}(c_\pi + c_\mu)/5}$$

if $g_\pi r_{bb'} \ll 1$. This can be represented by a parallel connection (Fig. 3.32b) of the dc base series resistance $r_{bb'}$ and a capacitance $c_{bb'} = (c_\pi + c_\mu)/5$ [6]. A similar investigation for a circular geometry [102] suggests a parallel capacitance $(c_\pi + c_\mu)/4$.

3.9.5 Nonlinear Distortion Effects

The design of integrated RF circuits requires the availability of transistor models that provide accurate simulation results for nonlinear effects, such as harmonic distortion and intermodulation distortion, at least up to third order in the input signal. The derivation of a third-order nonlinear small-signal model from a nonlinear transistor model requires model equations that yield accurate derivatives up to third order.

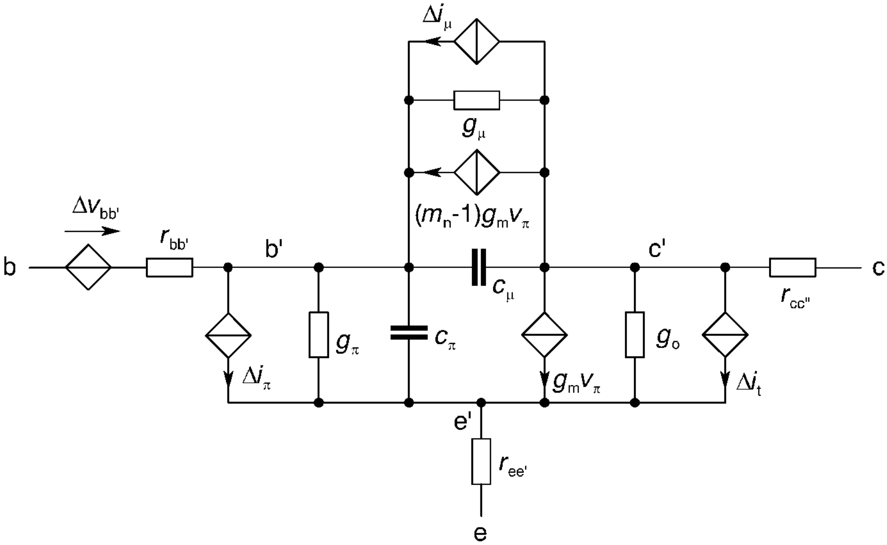


Fig. 3.33. Nonlinear small-signal equivalent circuit of a bipolar transistor used for analysis of distortion in weakly nonlinear circuits

Figure 3.33 shows a small-signal equivalent circuit of a bipolar transistor in forward operation extended by four polynomial controlled sources, which model the second- and third-order distortion terms associated with the nonlinearities of the respective branches of the large-signal equivalent circuit. The voltage source $\Delta v_{bb'}$ is used to model the nonlinear behavior of the base resistance, while Δi_t describes nonlinearities of the transfer current, Δi_π describes those of the eb diode (due to nonlinearities of the recombination current and the eb capacitance), and Δi_μ models nonlinarities of the bc capacitance and carrier multiplication. The effects of a nonlinear collector series resistance are not considered in Fig. 3.33, but could be implemented by adding a nonlinear voltage source $\Delta v_{cc'}$ in series with $r_{cc'}$.

Transfer Current

At low frequencies the transfer current of a bipolar transistor is a nonlinear function of the form $i_{\mathrm{T}}(t) = I_{\mathrm{T}}\left[v_{\mathrm{B'E'}}(t), v_{\mathrm{B'C'}}(t)\right]$, if the variable set of the common-base configuration is employed. If the voltages are written as

$$v_{\mathrm{B'E'}}(t) = V_{\mathrm{B'E'}} + v_\pi(t) \quad \text{and} \quad v_{\mathrm{B'C'}}(t) = V_{\mathrm{B'C'}} + v_\mu(t) ,$$

the transfer current may be written as

$$i_{\mathrm{T}}(t) \approx I_{\mathrm{T}} + g_{\mathrm{m}} v_\pi(t) + g_{\mathrm{o}} v_\mu(t) + \Delta i_{\mathrm{t2}}(t) + \Delta i_{\mathrm{t3}}(t) , \tag{3.196}$$

where I_{T} is the transfer current at the bias point, and

$$g_{\mathrm{m}} = \partial I_{\mathrm{T}} / \partial V_{\mathrm{B'E'}} \quad \text{and} \quad g_{\mathrm{o}} = \partial I_{\mathrm{T}} / \partial V_{\mathrm{C'B'}}$$

are parameters of the small-signal model; the term

$$\Delta i_{\mathrm{t2}} = \frac{1}{2}\frac{\partial^2 I_{\mathrm{T}}}{\partial V_{\mathrm{B'E'}}^2} v_\pi^2 + \frac{\partial^2 I_{\mathrm{T}}}{\partial V_{\mathrm{B'E'}} \partial V_{\mathrm{B'C'}}} v_\pi v_\mu + \frac{1}{2}\frac{\partial^2 I_{\mathrm{T}}}{\partial V_{\mathrm{B'C'}}^2} v_\mu^2 \tag{3.197}$$

describes nonlinear distortion due to terms of second order in the small-signal voltages, while

$$\begin{aligned} \Delta i_{\mathrm{t3}} = {} & \frac{1}{6}\frac{\partial^3 I_{\mathrm{T}}}{\partial V_{\mathrm{B'E'}}^3} v_\pi^3 + \frac{1}{2}\frac{\partial^3 I_{\mathrm{T}}}{\partial V_{\mathrm{B'E'}}^2 \partial V_{\mathrm{B'C'}}} v_\pi^2 v_\mu \\ & + \frac{1}{2}\frac{\partial^3 I_{\mathrm{T}}}{\partial V_{\mathrm{B'E'}} \partial V_{\mathrm{B'C'}}^2} v_\pi v_\mu^2 + \frac{1}{6}\frac{\partial^3 I_{\mathrm{T}}}{\partial V_{\mathrm{B'C'}}^3} v_\mu^3 \end{aligned} \tag{3.198}$$

describes nonlinear distortion due to terms of third order in the small-signal voltages. The current $\Delta i_{\mathrm{t}}(t) = \Delta i_{\mathrm{t2}}(t) + \Delta i_{\mathrm{t3}}(t)$ has to be introduced as a controlled current source parallel to the output conductance in the small-signal model.

Emitter–Base Diode

The large-signal current associated with the eb diode is associated with the recombination current and the current needed to charge and discharge the eb diffusion and depletion capacitances. According to the charge control theory, its value is given by

$$i_{\mathrm{BE}}(t) = \frac{I_{\mathrm{S}}}{B_{\mathrm{F}}}\left[\exp\left(\frac{v_{\mathrm{B'E'}}(t)}{V_{\mathrm{T}}}\right) - 1\right] + \frac{\mathrm{d}q_{\mathrm{TF}}}{\mathrm{d}t} + c_{\mathrm{je}}\frac{\mathrm{d}v_{\mathrm{B'E'}}}{\mathrm{d}t} ,$$

if the recombination current of the eb diode is described by an ideal diode. Using a third-order Taylor series expansion, the first term can be written as

$$I_{\mathrm{BE}} + g_\pi v_\pi(t) + \frac{g_\pi}{2V_{\mathrm{T}}} v_\pi^2(t) + \frac{g_\pi}{6V_{\mathrm{T}}^2} v_\pi^3(t) ,$$

where I_{BE} denotes the recombination current at the bias point.

The bias-dependent eb depletion capacitance may be developed around the bias point $V_{\mathrm{B'E'}}$ as follows:

$$c_{\mathrm{je}}\left[v_{\mathrm{B'E'}}(t)\right] = c_{\mathrm{je}}(V_{\mathrm{B'E'}}) + \sum_{m=1}^{\infty} \frac{1}{m!} \left.\frac{\mathrm{d}^m c_{\mathrm{je}}}{\mathrm{d}V^m}\right|_{V_{\mathrm{B'E'}}} v_\pi^m(t) \,. \tag{3.199}$$

Modeling the bias dependence of the eb depletion capacitance as

$$c_{\mathrm{je}}(V) = \frac{C_{\mathrm{JE}}}{(1-V/V_{\mathrm{JE}})^{M_{\mathrm{JE}}}} \tag{3.200}$$

yields the derivatives

$$\left.\frac{\mathrm{d}c_{\mathrm{je}}}{\mathrm{d}V}\right|_{V_{\mathrm{B'E'}}} = \frac{M_{\mathrm{JE}}}{V_{\mathrm{JE}}-V_{\mathrm{B'E'}}}\, c_{\mathrm{je}}(V_{\mathrm{B'E'}}) \,, \tag{3.201}$$

$$\left.\frac{\mathrm{d}^2 c_{\mathrm{je}}}{\mathrm{d}V^2}\right|_{V_{\mathrm{B'E'}}} = \frac{M_{\mathrm{JE}}(M_{\mathrm{JE}}+1)}{(V_{\mathrm{JE}}-V_{\mathrm{B'E'}})^2}\, c_{\mathrm{je}}(V_{\mathrm{B'E'}}) \,, \tag{3.202}$$

which give the following for the current that flows across the eb depletion capacitance:

$$i_{\mathrm{je}}(t) = c_{\mathrm{je}}(V_{\mathrm{B'E'}})\, \frac{\mathrm{d}v_\pi}{\mathrm{d}t} + \Delta c_{\mathrm{je}}(t)\, \frac{\mathrm{d}v_\pi}{\mathrm{d}t} \,, \tag{3.203}$$

where

$$\Delta c_{\mathrm{je}}(t) = c_{\mathrm{je}}(V_{\mathrm{B'E'}}) \left(\frac{M_{\mathrm{JE}}}{V_{\mathrm{JE}}-V_{\mathrm{B'E'}}}\, v_\pi(t) + \frac{M_{\mathrm{JE}}(M_{\mathrm{JE}}+1)}{2\,(V_{\mathrm{JE}}-V_{\mathrm{B'E'}})^2}\, v_\pi^2(t) \right) .$$

The diffusion charge q_{TF} is a function of the transfer current under forward-bias conditions:[41]

$$q_{\mathrm{TF}} = \int_0^{i_{\mathrm{T}}} \tau_{\mathrm{f}}(i)\, \mathrm{d}i \,;$$

its value can therefore be approximated by a third-order Taylor series

$$q_{\mathrm{TF}}(t) \approx Q_{\mathrm{TF}} + \tau_{\mathrm{f}} i_{\mathrm{t}}(t) + \frac{1}{2}\frac{\mathrm{d}\tau_{\mathrm{f}}}{\mathrm{d}I_{\mathrm{T}}}\, i_{\mathrm{t}}^2(t) + \frac{1}{6}\frac{\mathrm{d}^2\tau_{\mathrm{f}}}{\mathrm{d}I_{\mathrm{T}}^2}\, i_{\mathrm{t}}^3(t) \,,$$

where $i_{\mathrm{t}}(t) = g_{\mathrm{m}} v_\pi + g_{\mathrm{o}} v_\mu + \Delta i_{\mathrm{t2}}(t) + \Delta i_{\mathrm{t3}}(t)$ is the small-signal transfer current. Differentiation with respect to time gives the current

$$\left(\tau_{\mathrm{f}}(I_{\mathrm{T}}) + i_{\mathrm{t}}(t) \left.\frac{\mathrm{d}\tau_{\mathrm{f}}}{\mathrm{d}I_{\mathrm{T}}}\right|_{I_{\mathrm{T}}} + \frac{i_{\mathrm{t}}^2(t)}{2} \left.\frac{\mathrm{d}^2\tau_{\mathrm{f}}}{\mathrm{d}I_{\mathrm{T}}^2}\right|_{I_{\mathrm{T}}} \right) \frac{\mathrm{d}i_{\mathrm{t}}(t)}{\mathrm{d}t} \,.$$

[41] Here the generalized formulation used in the HICUM model is used (see Sect. 3.14).

Taking $\mathrm{d}i_\mathrm{t}/\mathrm{d}t = g_\mathrm{m}\,\mathrm{d}v_\pi/\mathrm{d}t$, the small-signal current $i_\mathrm{be}(t)$ may therefore be approximated as

$$i_\mathrm{be}(t) \;\approx\; g_\pi v_\pi + c_\pi \frac{\mathrm{d}v_\pi}{\mathrm{d}t} + \Delta i_\pi(t)\;, \tag{3.204}$$

where $c_\pi = \tau_\mathrm{f} g_\mathrm{m} + c_\mathrm{je}$ and

$$\Delta i_\pi(t) \;=\; \frac{g_\pi}{2V_\mathrm{T}}\,v_\pi^2(t) + \frac{g_\pi}{6V_\mathrm{T}^2}\,v_\pi^3(t) + \Delta c_\pi(t)\,\frac{\mathrm{d}v_\pi}{\mathrm{d}t}\;. \tag{3.205}$$

Here

$$\begin{aligned}\Delta c_\pi(t) \;&=\; \left(i_\mathrm{t}(t)\,\frac{\mathrm{d}\tau_\mathrm{f}}{\mathrm{d}I_\mathrm{T}}\bigg|_{I_\mathrm{T}} + \frac{i_\mathrm{t}^2(t)}{2}\,\frac{\mathrm{d}^2\tau_\mathrm{f}}{\mathrm{d}I_\mathrm{T}^2}\bigg|_{I_\mathrm{T}}\right) g_\mathrm{m} \\ &\quad + \frac{M_\mathrm{JE} c_\mathrm{je}(V_\mathrm{B'E'})}{V_\mathrm{JE}-V_\mathrm{B'E'}}\left(v_\pi(t) + \frac{M_\mathrm{JE}+1}{2(V_\mathrm{JE}-V_\mathrm{B'E'})}\,v_\pi^2(t)\right)\;.\end{aligned} \tag{3.206}$$

The terms $i_\mathrm{t}(t)$ and $i_\mathrm{t}^2(t)$ in the large parentheses need only be taken into account to second order in the small-signal control voltages.

Base–Collector Diode

Two sources of nonlinear distortion are represented by the polynomial controlled current source: the bias dependence of the bc depletion capacitance and the carrier multiplication factor

$$\Delta i_\mu = \Delta i_\mathrm{jc}(t) + \Delta i_\mathrm{avl}(t)\;.$$

Modeling the bias dependence of the depletion capacitance as

$$c_\mathrm{jc}[\,v_\mathrm{B'C'}(t)] \;=\; \frac{C_\mathrm{JC}}{[\,1-v_\mathrm{B'C'}(t)/V_\mathrm{JC}\,]^{M_\mathrm{JC}}}\;, \tag{3.207}$$

where $v_\mathrm{B'C'}(t) = V_\mathrm{B'C'} + v_\mu(t)$, yields the following for the small-signal current carried by the bc depletion capacitance:

$$c_\mathrm{jc}(V_\mathrm{B'C'})\,\frac{\mathrm{d}v_\mu}{\mathrm{d}t} + \Delta i_\mathrm{jc}(t)\;, \tag{3.208}$$

where

$$\Delta i_\mathrm{jc} \;=\; \frac{M_\mathrm{JC} c_\mathrm{jc}(V_\mathrm{B'C'})}{V_\mathrm{JC} - V_\mathrm{B'C'}}\left(v_\mu(t) + \frac{M_\mathrm{JC}+1}{2(V_\mathrm{JC}-V_\mathrm{B'C'})}\,v_\mu^2(t)\right)\frac{\mathrm{d}v_\mu}{\mathrm{d}t}\;. \tag{3.209}$$

The carrier multiplication factor $M_\mathrm{n}[v_\mathrm{C'B'}(t), i_\mathrm{T}(t)]$ is a nonlinear function of both the voltage $v_\mathrm{C'B'}(t)$ across the bc depletion layer and the transfer current injected into the depletion layer, and therefore is a source of additional distortion. In a fully depleted epitaxial collector region, the maximum value of the electric field strength is [48, 103]

$$E_{\max} \approx \frac{V_{\mathrm{C'B'}}+V_{\mathrm{JC}}}{d_{\mathrm{epi}}} + \frac{e d_{\mathrm{epi}} N_{\mathrm{Depi}}}{2\epsilon}\left(1-\frac{I_{\mathrm{T}}}{I_1}\right) ,$$

where d_{epi} and N_{Depi} denote the thickness and donor concentration of the epilayer and $I_1 = eA_{\mathrm{je}}N_{\mathrm{Depi}}v_{\mathrm{nsat}}$. If the multiplication factor is a function of the maximum electric field strength, its value at $(v_{\mathrm{C'B'}}, i_{\mathrm{T}})$ equals the value of the multiplication factor at $(\tilde{v}_{\mathrm{C'B'}}, i_{\mathrm{T}} = 0)$, where

$$\tilde{v}_{\mathrm{C'B'}} = v_{\mathrm{C'B'}} - R_{\mathrm{AVL}} I_{\mathrm{T}} ,$$

and [48]

$$R_{\mathrm{AVL}} = \frac{eN_{\mathrm{Depi}}d_{\mathrm{epi}}^2}{2\epsilon I_1} = \frac{d_{\mathrm{epi}}^2}{2\epsilon A_{\mathrm{je}} v_{\mathrm{nsat}}} .$$

In this approximation, the multiplication factor can be expanded, resulting in

$$\begin{aligned}\Delta i_{\mathrm{avl}}(t) \approx\ & \frac{I_{\mathrm{T}}}{2}\frac{\partial^2 M_{\mathrm{n}}}{\partial V_{\mathrm{C'B'}}^2}\bigg|_{\tilde{V}_{\mathrm{C'B'}}} \tilde{v}_{\mathrm{c'b'}}^2(t) + \frac{\partial M_{\mathrm{n}}}{\partial V_{\mathrm{C'B'}}}\bigg|_{\tilde{V}_{\mathrm{C'B'}}} \tilde{v}_{\mathrm{c'b'}}(t)\, i_{\mathrm{t}}(t) \\ & + \frac{I_{\mathrm{T}}}{6}\frac{\partial^3 M_{\mathrm{n}}}{\partial V_{\mathrm{C'B'}}^3}\bigg|_{\tilde{V}_{\mathrm{C'B'}}} \tilde{v}_{\mathrm{c'b'}}^3(t) + \frac{1}{2}\frac{\partial^2 M_{\mathrm{n}}}{\partial V_{\mathrm{C'B'}}^2}\bigg|_{\tilde{V}_{\mathrm{C'B'}}} \tilde{v}_{\mathrm{c'b'}}^2(t)\, i_{\mathrm{t}}(t) ,\end{aligned}$$

where $\tilde{v}_{\mathrm{b'c'}} = v_{\mathrm{b'c'}} - R_{\mathrm{AVL}} i_{\mathrm{t}}$, for the nonlinear second- and third-order distortion current caused by carrier multiplication in the bc diode.

Base Resistance

Owing to current crowding and conductivity modulation, the base resistance $R_{\mathrm{BB'}}$ is a function of the base current $i_{\mathrm{B}} = I_{\mathrm{B}} + i_{\mathrm{b}}$; developing this function into a Taylor series yields the following for the small-signal voltage drop across the base resistance:

$$\begin{aligned} v_{\mathrm{bb'}} &= (I_{\mathrm{B}}+i_{\mathrm{b}}) \times R_{\mathrm{BB'}}(I_{\mathrm{B}}+i_{\mathrm{b}}) - I_{\mathrm{B}} R_{\mathrm{BB'}} \\ &= r_{\mathrm{bb'}} i_{\mathrm{b}} + \Delta v_{\mathrm{bb'}} ,\end{aligned}$$

where $r_{\mathrm{bb'}} = R_{\mathrm{BB'}} + I_{\mathrm{B}}(\mathrm{d}R_{\mathrm{BB'}}/\mathrm{d}I_{\mathrm{B}})$ denotes the small-signal base resistance and

$$\Delta v_{\mathrm{bb'}} = \left(\frac{\mathrm{d}R_{\mathrm{BB'}}}{\mathrm{d}I_{\mathrm{B}}} + \frac{I_{\mathrm{B}}}{2}\frac{\mathrm{d}^2 R_{\mathrm{BB'}}}{\mathrm{d}I_{\mathrm{B}}^2}\right) i_{\mathrm{b}}^2 + \left(\frac{1}{2}\frac{\mathrm{d}^2 R_{\mathrm{BB'}}}{\mathrm{d}I_{\mathrm{B}}^2} + \frac{I_{\mathrm{B}}}{6}\frac{\mathrm{d}^3 R_{\mathrm{BB'}}}{\mathrm{d}I_{\mathrm{B}}^3}\right) i_{\mathrm{b}}^3$$

is the third-order nonlinear voltage drop across the base resistance.

3.10 Figures of Merit

The cutoff frequency f_T and the maximum frequency of oscillation f_{max} are widely used as "figures of merit" to characterize high-frequency bipolar technologies. As has already been pointed out in [104], neither f_T nor f_{max} are directly usable to define the switching-speed limitation of a device. Therefore additional figures of merit, such as the transconductance cutoff frequency and various delay times in bipolar transistors with a load device, such as the current-mode logic (CML) gate delay, are of interest.

The parameters of a bipolar transistor are correlated, and a modification that improves one parameter may cause a deterioration of another one. In order to obtain a figure of merit that takes account of how well a trade-off between different parameters is performed, the product of their values is frequently considered. Here the product of the current gain and the Early voltage and the product of the collector–emitter breakdown voltage and the cutoff frequency f_T are of particular interest.

3.10.1 Cutoff Frequency

The β-cutoff frequency, $f_\beta = \omega_\beta/2\pi$, of a bipolar transistor is defined in terms of a one-pole approximation to $h_{21e}(\omega)$,

$$h_{21e}(\omega) \;\approx\; \frac{\beta}{1+\mathrm{j}\omega/\omega_\beta} \;=\; h_{21e}^{(1)}(\omega)\;. \tag{3.210}$$

The magnitude of h_{21e} can therefore be approximated by

$$|h_{21e}(\omega)| \;\approx\; |h_{21e}^{(1)}(\omega)| \;=\; \frac{\beta}{\sqrt{1+\omega^2/\omega_\beta^2}}\;, \tag{3.211}$$

which, in the limit $\omega \ll \omega_\beta$, shows the following asymptotic behavior:

$$|h_{21e}^{(1)}(\omega)| \;\to\; \beta\omega_\beta/\omega \;=\; f_T/f\;, \tag{3.212}$$

where $f = \omega/2\pi$, and f_T is the cutoff frequency of the device. This definition is equivalent to the one proposed by Gummel [105],

$$\frac{1}{2\pi f_T} = -\mathrm{j}\lim_{\omega\to 0}\left[\frac{\mathrm{d}}{\mathrm{d}\omega}\,\mathrm{Im}\left(\frac{1}{h_{21e}(\omega)}\right)\right]\;,$$

and can also be obtained from $f_T = \beta f_\beta$, if f_β is calculated as the Elmore time constant [104, 106] (see Appendix A) of the system function $h_{21e}(\omega)$. According to these definitions, f_T is determined by the asymptotic behavior of $h_{21e}^{(1)}(\omega)$, which is a first-order expansion with respect to frequency. Therefore the quite commonly used definition of f_T as the frequency at which the common-emitter short-circuit current gain becomes unity is somewhat mis-leading and should be avoided – not $|h_{21e}(\omega)|$ but its one-pole approximation $|h_{21e}^{(1)}(\omega)|$ becomes unity when $f \to f_T$.

Values of f_β and f_T can be derived using the Giacoletto small-signal equivalent circuit, which gives, under the assumption $g_o = 0$ (see Sects. 1.7 and 3.9)

$$\frac{1}{2\pi f_T} = \frac{c_\pi + c_\mu}{g_m} + (r_{ee'} + r_{cc'})c_\mu . \tag{3.213}$$

However, this approach provides limited insight into the underlying device physics, and is not suitable for a computation of f_T from data obtained with a numerical device simulator. For technology optimization with the help of numerical device simulation programs, the relation

$$\frac{1}{2\pi f_T} = \left.\frac{dQ_p}{dI_C}\right|_{V_{CE}} \tag{3.214}$$

is used, which directly relates the cutoff frequency f_T to the hole charge Q_p stored in the device (see Sect. 3.5).

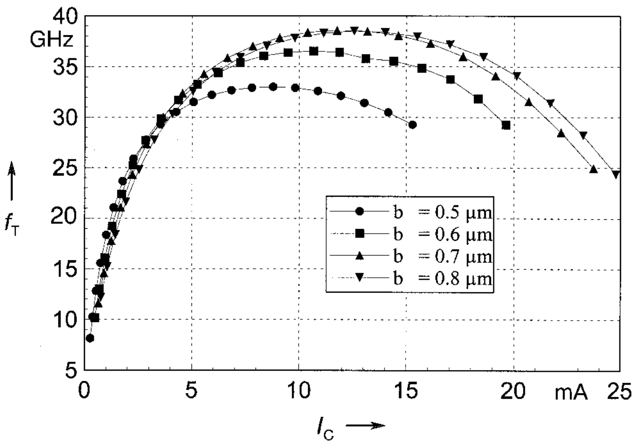


Fig. 3.34. Measured values of cutoff frequency f_T vs. I_C for different values of W_E (after [113])

The maximum cutoff frequency f_{Tmax} is found to decrease with decreasing emitter stripe width, as illustrated in Fig. 3.34. This effect is caused by two-dimensional effects of the current flow and by limitations[42] of the polysilicon emitter contact as a source for donor diffusion. This results in a reduced thickness of the monocrystalline emitter region in a narrow-stripe transistor and

[42] If the boundary between the polysilicon and the adjacent spacer oxide (see Chap. 7) is not vertical, the anisotropic diffusivity in polysilicon can limit the ability of the dopant to diffuse to the corner. This restricts the effectiveness of the polysilicon as a diffusion source in this region [107]. The use of in-situ-doped polysilicon and modified polysilicon spacers (see Chap. 7) allows one to reduce such layout dependence.

therefore an increased base width and transit time τ_f, with the consequence of a reduced cutoff frequency f_T.

The cutoff frequency f_T of a bipolar transistor represents an important figure of merit and gives information about the quality of the vertical doping profile of the transistor. As much abuse of this quantity is seen in the literature, however, we note that the widespread use of f_T as a single figure of merit for technology comparison is not justified, since this quantity does not describe the frequency behavior of the transistor in practical circuit configurations. The reason for this is that f_T is related to the change of stored charge associated with a change of the output current and therefore provides no information about the series resistances seen by the flowing charge. The *RC* delays associated with the series resistances, however, have substantial effect on the high-frequency performance of a transistor in a circuit. Furthermore, the collector–substrate capacitance, which is important for the high-frequency performance of integrated circuits, has no effect on the cutoff frequency. The cutoff frequency therefore does not represent the whole truth.

3.10.2 Maximum Frequency of Oscillation

The maximum frequency of oscillation $f_{\max}$ is defined as the frequency where the unilateral power gain

$$U = \frac{|\, y_{21} - y_{12}\,|^2}{4\,[\mathrm{Re}(y_{11})\mathrm{Re}(y_{22}) - \mathrm{Re}(y_{12})\mathrm{Re}(y_{21})]}$$

becomes unity (see Appendix B). If the admittance parameters are expressed in terms of the approximations derived in Sect. 3.9, we obtain

$$U \approx \frac{\beta g_m}{4 g_o} \frac{1 + \dfrac{f^2}{f_y^2}}{\left(1 + \dfrac{f^2}{f_\beta f_y}\right)\left(1 + \dfrac{f^2}{f_{22} f_y}\right) - \beta\,\dfrac{f^2}{f_y f_{12}} - \dfrac{g_o}{g_\pi}\,\dfrac{f^4}{f_y^2 f_{12}^2}} . \tag{3.215}$$

The low-frequency value of the unilateral gain may be written as

$$U \approx \frac{\beta g_m}{4 g_o} \approx \frac{\beta}{4}\frac{V_{CE} + V_{AF}}{V_T}$$

if the approximations $g_m \approx I_C/V_T$ and $g_o \approx I_C/(V_{CE}+V_{AF})$ are used; its value increases in proportion to the product of current gain and Early voltage. The maximum frequency of oscillation $f_{\max}$ is obtained from (3.215) and the condition $U(f_{\max}) = 1$, which gives a biquadratic equation for $f_{\max}$. The solution gives approximately [111]

$$f_{\max} \approx \sqrt{\frac{f_T}{8\pi(r_{bb'}+r_{ee'}+1/g_m)c_\mu}} , \tag{3.216}$$

which generalizes the result $f_{\max} = \sqrt{f_{\mathrm{T}}/8\pi r_{\mathrm{bb'}} c_\mu}$ obtained in Sect. 1.7. In this discussion, simplified admittance parameters have been used, which neglect the effects of the substrate series resistance $r_{\mathrm{ss'}}$.[43] Measurements on special test structures presented in [112], however, demonstrated an effect of the substrate resistance on $f_{\max}$ that was found to be worst if the condition $r_{\mathrm{ss'}} = 1/(2\pi c_{\mathrm{js}} f_{\max})$ was fulfilled. Larger values of the substrate resistance result in a substantial increase of the maximum frequency of oscillation.

The maximum frequency of oscillation $f_{\max}$ is difficult to determine by measurement from the condition $U(f_{\max}) = 1$, since the required unilateralization is hard to realize in practice. Therefore, the maximum available gain (MAG), and the maximum stable gain (MSG) (see Appendix B) are generally used for the extraction of $f_{\max}$ from measured data (see Fig. 3.35). These quantitities approximately equal the unilateral gain if y_{12} is small. Theoretically, the conditions $\mathrm{MAG}(f_{\max}) = 1$ and $U(f_{\max}) = 1$ yield the same value $f_{\max}$. In measurements, however, $f_{\max}$ is determined by extrapolation of either $\mathrm{MAG}(f)$ or $U(f)$ with a slope of -20 dB/dec resulting in $f_{\max}^{(\mathrm{MAG})}$ or $f_{\max}^{(\mathrm{U})}$, respectively. Since, generally, $f_{\max}^{(\mathrm{MAG})} < f_{\max} < f_{\max}^{(\mathrm{U})}$, the procedure used to determine $f_{\max}$ should be specified in order to obtain comparable results.

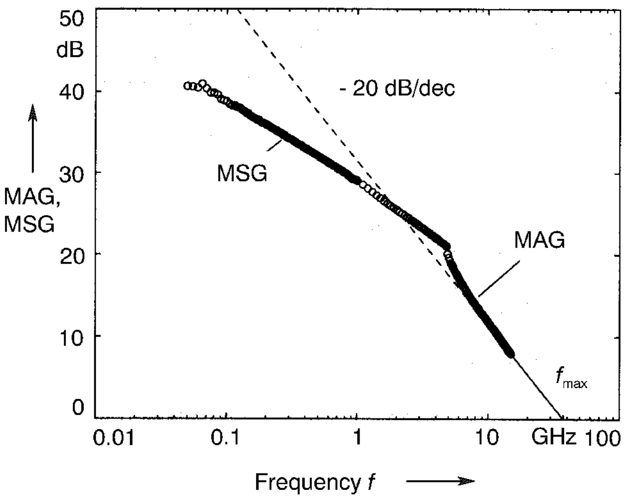


Fig. 3.35. Maximum available gain and maximum stable gain of a self-aligned vertical npn bipolar transistor (after [113])

As was pointed out in [114], the value of $f_{\max}$ is based on neutralizing the feedback between the transistor output and input circuitry by complex net-

[43]The effects of an external base–collector capacitance were investigated in [108, 109], resulting in slightly modified expressions of the form $f_{\max} = \sqrt{f_{\mathrm{T}}/8\pi(RC)_{\mathrm{eff}}}$ with appropriately defined effective values of R and C. In [110] a formula of $f_{\max}$ is derived that takes account of the excess phase.

works, which are quite impractical in integrated circuits. In [114], the frequency f_{PT} at which the power transferred between identical amplifiers is equal to unity is therefore suggested as a practical figure of merit for an integrated bipolar transistor. Unfortunately, this frequency cannot be easily determined using on-wafer s-parameter measurements (unlike f_{T} and $f_{\max}$); it is probably this deficiency that explains why this undoubtedly interesting quantity is not commonly employed. In [115] the bandwidth of a bipolar amplifier in the common–emitter configuration with a load resistance R_{L} and load capacitance C_{L} (see Sect. 6) was calculated as

$$\frac{1}{2\pi f_{\mathrm{g}}} = \tau = \frac{1}{2\pi f_{\mathrm{y}}} + R_{\mathrm{L}}(c_{\mathrm{jc}}+c_{\mathrm{js}}+C_{\mathrm{L}}) + \frac{R_{\mathrm{C}} g_{\mathrm{m}} f_{\mathrm{T}}}{8\pi f_{\max}^2} .$$

The minimum value of the time constant τ was suggested as an alternative figure of merit. This figure of merit takes account of the importance of the transconductance cutoff frequency if the input is driven with a voltage source, as has been assumed in [115]. As the minimum CML gate delay is closely related to the proposed figure of merit, it appears to be more practical to the present author to use this easily measurable quantity as a figure of merit.

3.10.3 CML Gate Delay and Power–Delay Product

The gate delay of a CML inverter circuit[44] and its product with the dissipated power are widely used figures of merit to characterize the potential of a bipolar technology. Closed-form expressions for the gate delay have been published, for example, in [104, 116–122]. The gate delay of a CML gate is approximately given by [116, 118]

$$\tau_{\mathrm{d}} = \tau_{\mathrm{f}} + R_{\mathrm{BB'}} C_{\mathrm{B}} + R_{\mathrm{C}} C_{\mathrm{C}} , \tag{3.217}$$

where the capacitances C_{B} and C_{C} are defined as weighted averages of the transistor and wiring capacitances. As a crude approximation, one may use [118]

$$C_{\mathrm{B}} \approx 2\, c_{\mathrm{jc}}(0) + \frac{\tau_{\mathrm{f}} I_{\mathrm{EE}}}{150\,\mathrm{mV}} , \tag{3.218}$$

$$C_{\mathrm{C}} \approx 2\, c_{\mathrm{jc}}(0) + 0.5\, c_{\mathrm{js}}(0) + C_{\mathrm{W}} , \tag{3.219}$$

where c_{js} denotes the collector–substrate capacitance and C_{W} the capacitance due to the wiring and load capacitances.

Since for a given logic swing V_{S} the value of the collector resistance $R_{\mathrm{C}} = V_{\mathrm{S}}/I_{\mathrm{EE}}$ varies in inverse proportion to the gate current I_{EE}, the third term

[44] We distinguish two types of emitter-coupled logic (see Chap. 6): ECL, which employs emitter followers, and CML, which does not use emitter followers. Published performance data are generally obtained from CML ring oscillator measurements.

on the right-hand side of (3.217) is predominant at small values of I_{EE}. This term varies in inverse proportion to the dissipated power $P = I_{EE}|V_{EE}|$. A double logarithmic plot of the gate delay τ_d versus the dissipated power P then gives a straight line with slope -1.

At large values of P, the delay time is determined by the first two terms on the right-hand side of (3.217). Owing to the increase of the forward transit time τ_f at large values of the transfer current, the value of C_B increases with I_{EE} (and P) and causes the gate delay τ_d to increase after going through a minimum. A plot of the gate delay time versus the power consumed therefore results in a diagram as depicted in Fig. 3.36.

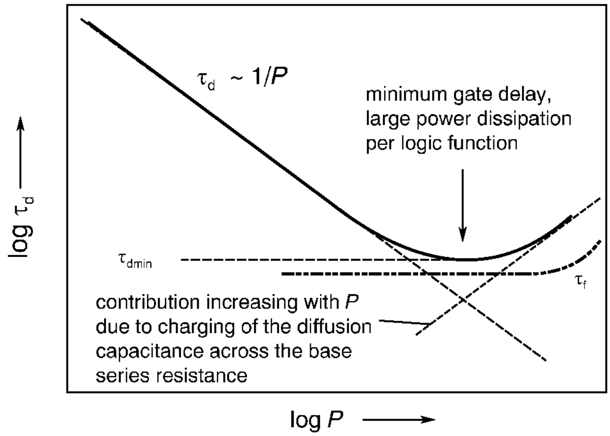


Fig. 3.36. CML gate delay as a function of dissipated power

Besides the minimum value τ_{dmin} of the gate delay, the line that approximates the curve at small values of dissipated power is of particular interest. There the product of the dissipated power and the gate delay is approximately constant; its value is called the power–delay product and is used to characterize bipolar technologies. The power–delay product is given by

$$P\tau_d \geq PR_CC_C = C_CV_S|V_{EE}| .$$

In CML inverter circuits, power–delay products below 10 fJ have already been obtained.

The CML gate delay and the power–delay product are usually determined from ring oscillator measurements (see Sect. 6.10). These circuits generally use differential operation with a small logic swing V_S, whereas practical circuit applications generally employ single-ended ECL circuits (see again Sect. 6.10) with a considerably larger voltage swing and supply voltage, and additional power dissipation due to the emitter follower. Furthermore, there is virtually no wiring capacitance in such ring oscillator test structures. In practical ECL circuits the product of the gate delay and power dissipation will therefore substantially exceed the (ring oscillator) power–delay product.

As was pointed out in [123], an optimum load resistance R_C for the minimum gate delay time may be derived for a given value of the logic swing $V_S = R_C I_{EE}$. The minimum gate delay achievable was estimated in [123] to be

$$\tau_{\mathrm{dmin}} = r_{\mathrm{bb'}}(c_{\mathrm{je}}+3c_{\mathrm{jc}}) + 2\sqrt{2\tau_{\mathrm{f}} r_{\mathrm{bb'}} c_{\mathrm{jc}}} \,.$$

This formula explicitly relates the fundamental device parameters of base resistance, forward transit time and junction capacitances and yields a simple figure of merit for bipolar CML circuits. A somewhat different expression was derived in [120], where

$$\tau_{\mathrm{dmin}} = r_{\mathrm{bb'}}(c_{\mathrm{je}}+3c_{\mathrm{jc}}) + \alpha\tau_{\mathrm{f}} + \sqrt{3}/\omega_{\max}$$

was found. The parameter $\alpha = \theta/\omega = \pi P_{\mathrm{TF}}/180°$ describes the excess phase θ and gives the fraction of the total excess stored charge that is reclaimable at the collector terminal; $\omega_{\max} = \sqrt{\pi f_{\mathrm{T}}/(2 r_{\mathrm{bb'}} c_{\mathrm{jc}})}$ is the maximum angular frequency of oscillation that is approached at high current densities.

3.10.4 Product of Current Gain and Early Voltage

The product of the current gain and the Early voltage determines the quality of current mirror circuits and therefore provides a figure of merit for analog circuit applications. From (3.53) and (3.40) we obtain the following for the current gain:

$$B_{\mathrm{N}} = \frac{\displaystyle\int_{x_{\mathrm{e}}}^{x_{\mathrm{eb}}} \frac{n(x)}{\mu_{\mathrm{p}} n_{\mathrm{ie}}^2}\,\mathrm{d}x + \frac{n(x_{\mathrm{e}})V_{\mathrm{T}}}{n_{\mathrm{ie}}^2(x_{\mathrm{e}})S_{\mathrm{nn}}}}{\displaystyle\int_{x_{\mathrm{be}}}^{x_{\mathrm{bc}}} \frac{p(x)}{\mu_{\mathrm{n}} n_{\mathrm{ie}}^2}\,\mathrm{d}x + \frac{p(x_{\mathrm{bc}})V_{\mathrm{T}}}{n_{\mathrm{ie}}^2(x_{\mathrm{bc}})v_{\mathrm{nsat}}}} \,.$$

According to (3.41), the forward Early voltage is given by

$$V_{\mathrm{AF}} \approx \frac{eA_{\mathrm{je}} n_{\mathrm{ie}}^2(x_{\mathrm{bc}})}{c_{\mathrm{jc}}(V_{\mathrm{CB}})} \frac{\displaystyle\int_{x_{\mathrm{be}}}^{x_{\mathrm{bc}}} \frac{p}{\mu_{\mathrm{n}} n_{\mathrm{ie}}^2}\,\mathrm{d}x + \frac{p(x_{\mathrm{bc}})}{n_{\mathrm{ie}}^2(x_{\mathrm{bc}})}\frac{V_{\mathrm{T}}}{v_{\mathrm{nsat}}}}{\displaystyle \frac{1}{\mu_{\mathrm{n}}(x_{\mathrm{bc}})} + \frac{V_{\mathrm{T}}}{v_{\mathrm{nsat}}} \left.\frac{\partial \ln(p/n_{\mathrm{ic}})}{\partial x}\right|_{x_{\mathrm{bc}}}}$$

as long as $V_{\mathrm{CB}} \ll V_{\mathrm{AF}}$. Combining the two relations yields the following for the product of the current gain and Early voltage:

$$B_{\mathrm{N}} V_{\mathrm{AF}} = \frac{eA_{\mathrm{je}} n_{\mathrm{ie}}^2(x_{\mathrm{bc}})}{c_{\mathrm{jc}}(V_{\mathrm{CB}})} \frac{\displaystyle\int_{x_{\mathrm{e}}}^{x_{\mathrm{eb}}} \frac{n(x)}{\mu_{\mathrm{p}}\, n_{\mathrm{ie}}^2}\,\mathrm{d}x + \frac{n(x_{\mathrm{e}})V_{\mathrm{T}}}{n_{\mathrm{ie}}^2(x_{\mathrm{e}})S_{\mathrm{nn}}}}{\displaystyle \frac{1}{\mu_{\mathrm{n}}(x_{\mathrm{bc}})} + \frac{V_{\mathrm{T}}}{v_{\mathrm{nsat}}} \left.\frac{\partial \ln(p/n_{\mathrm{ie}})}{\partial x}\right|_{x_{\mathrm{bc}}}} \,. \tag{3.220}$$

This product varies in proportion to $n_{\mathrm{ie}}^2(x_{\mathrm{bc}})$ at the collector edge of the base [124]. Heterojunction bipolar transistors with a bandgap grading towards the collector junction (see Chap. 4) therefore show much larger values of the product of current gain and Early voltage than do homojunction transistors with a comparable vertical doping profile. Typical $B_{\mathrm{N}}V_{\mathrm{AF}}$ values for homojunction bipolar transistors are of the order of 1 kV, whereas heterojunction transistors may show values in excess of 10 kV.

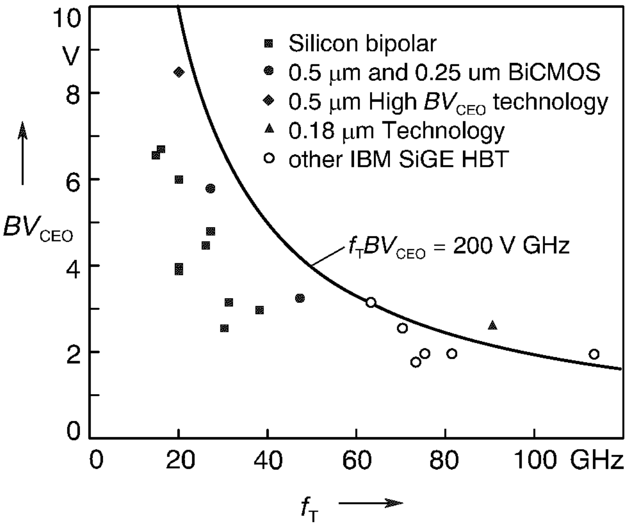


Fig. 3.37. Open-base breakdown voltage BV_{CEO} versus cut-off frequency f_{T} for various bipolar and BiCMOS technologies (after [129])

3.10.5 Johnson Limit

A high cutoff frequency requires a large transfer current density and therefore a high collector doping with the consequence of a strongly reduced breakdown voltage BV_{CEO} (Fig. 3.37). The trade-off between the cutoff frequency f_{T} and the open-base breakdown voltage BV_{CEO} is characterized by the product $f_{\mathrm{T}}BV_{\mathrm{CEO}}$. This product is related to the so-called Johnson limit, which is based on the following elementary considerations [125, 126]. Assume that the current is carried by carriers which move at their saturated drift velocity v_{nsat}. If we neglect all other contributions, the cutoff frequency may be estimated[45] from the time required for the carriers to pass through the drift region of length L, so that $2\pi f_{\mathrm{T}} \approx v_{\mathrm{nsat}}/L$. If the voltage drop V across the drift region is limited by breakdown, which is assumed to occur if a critical field

[45] A more detailed analysis shows that owing to electrostatic induction, part of the current passing between the two electrodes is transported as displacement current, which reduces the transit time to one half of the estimated value.

strength E_{crit} is exceeded, the condition $V < E_{\text{crit}}L$ has to be fulfilled. By forming the product, we obtain the condition

$$f_{\text{T}}V < v_{\text{nsat}}E_{\text{crit}}/2\pi .$$

The right-hand side of this inequality is called the Johnson limit. Assuming $v_{\text{nsat}} \approx 6 \times 10^6$ cm/s and $E_{\text{crit}} = 2 \times 10^5$ cm/s, we can estimate the right-hand side as 200 V GHz. However, the Johnson limit can provide only a crude estimate of the $f_{\text{T}}BV_{\text{CEO}}$ value that can be achieved within a bipolar technology, since neither f_{T} nor BV_{CEO} depends exclusively on the extent of the bc depletion layer, for the following reasons:

- If open-base breakdown occurs because of the avalanche effect, as is generally the case, the breakdown condition $B'_{\text{N}}(M_{\text{n}}-1) = 1$, where $B'_{\text{N}} = I_{\text{T}}/I_{\text{BE}}$, has to be fulfilled. The approximation (see Sect. 3.4) $M_{\text{n}} - 1 \approx 1 - 1/M_{\text{n}} \approx (V_{\text{CB}}/BV_{\text{C}})^{N_1}$ yields

$$BV_{\text{CEO}} \approx (B'_{\text{N}}+1)^{-1/N_1}BV_{\text{C}} + V_{\text{BEon}} .$$

  The collector–emitter breakdown voltage is therefore determined not only by the properties of the bc junction, but also by the current gain and the base–emitter voltage. Transistors with a poor emitter efficiency will therefore show large values of $f_{\text{T}}BV_{\text{CEO}}$, as f_{T} remains unaffected by the current gain.

- A general relation for the cutoff frequency is

$$f_{\text{T}} = \frac{1}{2\pi[(c_{\text{je}}+c_{\text{jc}})V_{\text{T}}/I_{\text{C}} + \tau_{\text{f}}(I_{\text{C}}) + (r_{\text{ee}'}+r_{\text{cc}'})c_{\text{jc}}]} ;$$

  its maximum value f_{Tmax} is predominantly determined by the forward transit time τ_{f}, i.e.

$$f_{\text{Tmax}}BV_{\text{CEO}} \approx \frac{BV_{\text{CEO}}}{2\pi\tau_{\text{f}}(I_{\text{C}})} .$$

  The denominator of this expression comprises the time constants due to storage of minority carriers in the emitter and base regions, and the collector transit time τ_{jc}. Of these, only τ_{jc} depends on the extent of the depletion layer and thus on the collector doping.

These arguments show that $f_{\text{T}}BV_{\text{CEO}}$ may deviate substantially from the Johnson limit. In addition it should be noted that $f_{\text{T}}BV_{\text{CEO}}$ is determined by "dark-space effects" associated with the threshold energy for impact ionization, velocity overshoot effects [127], the displacement current in the bc depletion layer and the fact that the value of the critical electric field strength for breakdown is not a constant. High-frequency bipolar transistors with a SiGe base layer have been realized with $f_{\text{T}}BV_{\text{CEO}} > 200$ V GHz.

3.11 Temperature Dependences, Self-Heating

Most of the parameters used in bipolar-transistor models are temperature-dependent quantities. This holds, in particular, for the transfer saturation current, which has an exponential dependence on temperature. If a device is operated at a temperature T that differs from the reference temperature T_0, at which the parameters were determined, the parameters have to be calculated for the new temperature, using an appropriate temperature model.

The power $p(t)$ dissipated in a biased bipolar transistor elevates its temperature T above the ambient temperature T_A, an effect commonly referred to as self-heating. Self-heating affects the shape of measured current–voltage characteristics, affect two-port parameters and precision analog circuits that depend on close transistor matching [128], and may even lead to bias point instability. Self-heating is not considered in the SPICE Gummel–Poon model, but is taken into account in modern compact models such as VBIC, HICUM and MEXTRAM.

3.11.1 Temperature Dependences

Series resistances vary with temperature owing to the temperature dependence of the majority-carrier mobility. As the mobility usually varies in proportion to a power of the absolute temperature, i.e., $\mu \sim T^{-m}$, temperature-dependent resistances are often written in the form

$$R(T) = R(T_0)\left(\frac{T}{T_0}\right)^m ,$$

where $R(T_0)$ denotes the resistance at the reference temperature T_0. This approach is used in the compact models VBIC, HICUM and MEXTRAM, while SPICE determines temperature-dependent resistance values from a temperature coefficient

$$\alpha_R = \frac{1}{R}\frac{dR}{dT} ,$$

which gives the relative change of the resistance value per Kelvin. A comparison of the two approaches gives $\alpha_R \approx m/T$.

The transfer saturation current varies in proportion to $D_n n_{ie}^2$, i.e. we may write

$$I_S(T) = I_S(T_0)\frac{D_n(T)}{D_n(T_0)}\frac{n_{ie}^2(T)}{n_{ie}^2(T_0)} ,$$

where

$$n_{ie}^2 \sim T^\gamma \exp\left(-\frac{W_g(T)}{k_B T}\right)$$

has a strong temperature dependence; the parameter γ is approximately equal to three. If μ_{n} varies in proportion to T^{-m}, the diffusion coefficient $D_{\mathrm{n}} = V_{\mathrm{T}}\mu_{\mathrm{n}}$ varies in proportion to T^{1-m}. Introducing the parameter $X_{\mathrm{TI}} = \gamma + 1 - m$, we may write the following for the transfer saturation current:

$$I_{\mathrm{S}}(T) = I_{\mathrm{S}}(T_0)\left(\frac{T}{T_0}\right)^{X_{\mathrm{TI}}} \exp\left(\frac{W_{\mathrm{g}}(T_0)}{k_{\mathrm{B}}T_0} - \frac{W_{\mathrm{g}}(T)}{k_{\mathrm{B}}T}\right) .$$

Developing $W_{\mathrm{g}}(T)$ around T_0 up to first order gives

$$W_{\mathrm{g}}(T) \approx W_{\mathrm{g}}(T_0) + (T - T_0)\left.\frac{\mathrm{d}W_{\mathrm{g}}}{\mathrm{d}T}\right|_{T_0} ,$$

and therefore

$$I_{\mathrm{S}}(T) = I_{\mathrm{S}}(T_0)\left(\frac{T}{T_0}\right)^{X_{\mathrm{TI}}} \exp\left(\frac{V_{\mathrm{g}}}{V_{\mathrm{T}}}\frac{T-T_0}{T_0}\right) ,$$

where

$$V_{\mathrm{g}} = W_{\mathrm{g}}(T_0) - T_0\left.\frac{\mathrm{d}W_{\mathrm{g}}}{\mathrm{d}T}\right|_{T_0}$$

denotes the bandgap voltage at the reference temperature. Since the bandgap voltage is substantially less dependent on temperature than is $W_{\mathrm{g}}(T)$, it is a good approximation to assume a constant value of V_{g}.

Depletion capacitances are affected by the temperature dependence of both the dielectric constant and the built-in voltage V_{J}. On the basis of the result

$$V_{\mathrm{J}} = V_{\mathrm{T}} \ln \frac{n_{\mathrm{n0}}p_{\mathrm{p0}}}{n_{\mathrm{ie}}^2} \approx V_{\mathrm{T}} \ln \frac{N_{\mathrm{A}}N_{\mathrm{D}}}{n_{\mathrm{ie}}^2(T)} = V_{\mathrm{T}} \ln \frac{N_{\mathrm{A}}N_{\mathrm{D}}}{n_{\mathrm{ie}}^2(T_0)} - V_{\mathrm{T}} \ln \frac{n_{\mathrm{ie}}^2(T)}{n_{\mathrm{ie}}^2(T_0)} ,$$

the temperature dependences of junction built-in voltages are frequently determined from the relation

$$V_{\mathrm{J}}(T) = \frac{T}{T_0}V_{\mathrm{J}}(T_0) - V_{\mathrm{T}}\left[3\ln\left(\frac{T}{T_0}\right) + \frac{V_{\mathrm{g}}}{V_{\mathrm{T}}}\frac{T-T_0}{T_0}\right] \tag{3.221}$$

that is obtained if $\gamma = 3$ and $W_{\mathrm{g}}(T)$ is assumed to be linear. This results in a decrease of the built-in voltage with increasing temperature, and may even lead to negative values of $V_{\mathrm{J}}(T)$ – a result that is obviously unphysical. The problem arises from the assumptions $n_{\mathrm{n0}} \approx N_{\mathrm{D}}$ and $p_{\mathrm{p0}} \approx N_{\mathrm{A}}$, which are only fulfilled if $N_{\mathrm{A}}, N_{\mathrm{D}} \gg n_{\mathrm{ie}}$; if this is not the case, the relations

$$n_{\mathrm{n0}} = \frac{1}{2}\left(N_{\mathrm{D}} + \sqrt{N_{\mathrm{D}}^2 + 4n_{\mathrm{ie}}^2}\right) \quad \text{and} \quad p_{\mathrm{p0}} = \frac{1}{2}\left(N_{\mathrm{A}} + \sqrt{N_{\mathrm{A}}^2 + 4n_{\mathrm{ie}}^2}\right)$$

have to be used, which result in $V_{\mathrm{J}} \to 0$ as $T \to \infty$. As C_{J0} varies in proportion to $\epsilon_{\mathrm{r}}^{1-M}V_{\mathrm{J}}^{M}$, the temperature coefficient of the zero-bias depletion capacitance is

$$\alpha_{\mathrm{JC0}} = (1-M)\frac{1}{\epsilon_{\mathrm{r}}}\frac{\mathrm{d}\epsilon_{\mathrm{r}}}{\mathrm{d}T} - M\frac{1}{V_{\mathrm{J}}}\frac{\mathrm{d}V_{\mathrm{J}}}{\mathrm{d}T} \,.$$

The temperature coefficient of the dielectric constant of silicon is [130]

$$\frac{1}{\epsilon_{\mathrm{r}}}\frac{\mathrm{d}\epsilon_{\mathrm{r}}}{\mathrm{d}T} \approx 3.5\times 10^{-5}\,\frac{1}{\mathrm{K}}$$

and is small in comparison with the temperature coefficient of the built-in voltage, which is of the order of $-0.2\%/\mathrm{K}$. The temperature dependence of the zero-bias depletion capacitance is therefore dominated by $V_{\mathrm{J}}(T)$, resulting in the widely used model equation

$$C_{\mathrm{J0}}(T) = C_{\mathrm{J0}}(T_0)\left(\frac{V_{\mathrm{J}}(T)}{V_{\mathrm{J}}(T_0)}\right)^M \,. \tag{3.222}$$

3.11.2 Thermal Equivalent Circuit

To relate the device excess temperature $\Delta T(t) = T(t) - T_{\mathrm{A}}$ to the dissipated power $p(t)$, the power balance illustrated in Fig. 3.38 is considered.

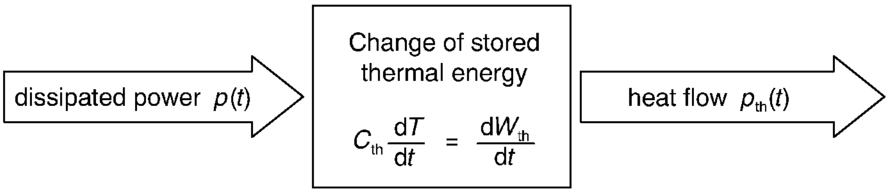


Fig. 3.38. Power balance considered to determine the self-heating effect

For $\Delta T(t) > 0$, heat flows to the ambient: the power

$$p_{\mathrm{th}}(t) = \Delta T(t)/R_{\mathrm{th}} \tag{3.223}$$

is delivered to the ambient, where R_{th} is the thermal resistance between the device and the ambient. The difference between the dissipated power and the extracted thermal power is the rate of change with time of the heat energy of the device:

$$p(t) - p_{\mathrm{th}}(t) = \frac{\mathrm{d}W_{\mathrm{th}}}{\mathrm{d}t} = \frac{\mathrm{d}W_{\mathrm{th}}}{\mathrm{d}T}\frac{\mathrm{d}T}{\mathrm{d}t} \,. \tag{3.224}$$

Since the rate of change of the thermal energy with temperature $\mathrm{d}W_{\mathrm{th}}/\mathrm{d}T$ equals the thermal capacitance C_{th} of the device, the following differential equation for the excess temperature $\Delta T(t)$ is obtained:

$$\frac{\mathrm{d}}{\mathrm{d}t}\Delta T + \frac{\Delta T}{\tau_{\mathrm{th}}} = \frac{p(t)}{C_{\mathrm{th}}} \,. \tag{3.225}$$

The quantity τ_{th} defines the thermal time constant:

$$\tau_{\text{th}} = R_{\text{th}} C_{\text{th}} \,. \tag{3.226}$$

The solution of (3.225) with the initial condition $\Delta T(0) = 0$ is

$$\Delta T(t) = \frac{1}{C_{\text{th}}} \int_0^t p(t') \exp\left(-\frac{t-t'}{\tau_{\text{th}}}\right) \mathrm{d}t' \,. \tag{3.227}$$

This simplifies to

$$\Delta T(t) = P R_{\text{th}} \left[1 - \exp\left(-\frac{t}{\tau_{\text{th}}}\right)\right] \quad \text{for} \quad t > 0 \tag{3.228}$$

if $p(t) = P$ is constant.

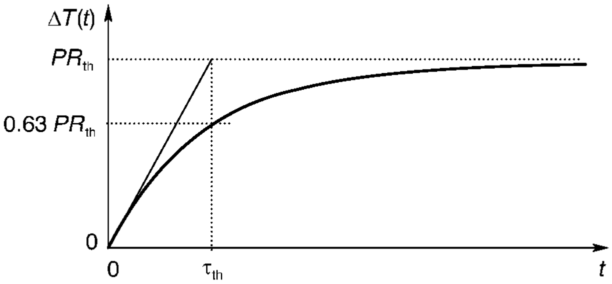


Fig. 3.39. Excess temperature versus time for constant power dissipation

For small times ($t \ll \tau_{\text{th}}$) the excess temperature $\Delta T(t)$ increases approximately linearly with time (see Fig. 3.39):

$$\Delta T(t) \approx P R_{\text{th}} \frac{t}{\tau_{\text{th}}} = \frac{Pt}{C_{\text{th}}} \,;$$

for large times

$$\Delta T(t) \to P R_{\text{th}} \quad \text{for} \quad t \to \infty \,.$$

Equation (3.223) may be considered in analogy to Ohm's law: the heat flow p_{th} due to the temperature difference ΔT is determined by the thermal resistance, while the current i due to a potential difference v is determined by the electrical resistance. This analogy allows one to describe temperature relaxation using thermal equivalent circuits, as depicted in Fig. 3.40.[46] The thermal time constant τ_{th} corresponds to the RC time constant of the equivalent circuit: if p, R_{th} and C_{th} are related to $\tilde{\imath}_{\text{th}}$, $\tilde{R}_{\text{th}}$ and $\tilde{C}_{\text{th}}$ according to

$$\tilde{\imath}_{\text{th}} = 1 \frac{\mathrm{A}}{\mathrm{W}} p \,, \quad \tilde{C}_{\text{th}} = 1 \frac{\mathrm{F\,K}}{\mathrm{J}} C_{\text{th}} \quad \text{and} \quad \tilde{R}_{\text{th}} = 1 \frac{\Omega\,\mathrm{W}}{\mathrm{K}} R_{\text{th}} \,,$$

[46] R_{th} and C_{th} can be replaced by a more accurate thermal impedance $Z_{\text{th}}(s)$ [131] in order to describe self-heating effects beyond the "thermal one-pole approximation".

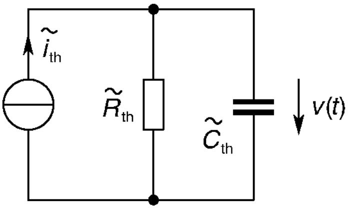


Fig. 3.40. Thermal equivalent circuit

the identity $R_{th}C_{th} = \tilde{R}_{th}\tilde{C}_{th}$ holds. In this case the voltage v across the capacitor varies in direct proportion to the excess temperature:

$$\Delta T = 1\frac{\mathrm{K}}{\mathrm{V}}v\,,$$

i.e. an increase of v by 1 V corresponds to an increase in the excess temperature of 1 K.

The maximum power dissipation P_{max} of a transistor is determined by the maximum allowable junction temperature T_{Jmax}, the thermal resistance between the transistor and the ambient, and the ambient temperature:

$$P_{max} = (T_{max}-T_A)/R_{th}\,.$$

3.11.3 Mitlaufeffekt, Thermal Runaway

If V_{BE} or I_B is held constant, an increase in V_{CE} causes an increase in the power dissipated in the transistor

$$P = I_BV_{BE} + I_CV_{CE} \approx I_CV_{CE}\,.$$

In a slow measurement of the output characteristics, each measurement point will have a specific excess temperature

$$\Delta T \approx R_{th}V_{CE}I_C$$

above the ambient temperature owing to the power dissipated at the bias point; this is the Mitlaufeffekt. In this case the slope of the output characteristic with V_{BE} held constant is

$$\frac{\mathrm{d}I_C}{\mathrm{d}V_{CE}} = \left(\frac{\partial I_C}{\partial V_{CE}}\right)_T + \left(\frac{\partial I_C}{\partial T}\right)_{V_{CE}}\frac{\mathrm{d}T}{\mathrm{d}V_{CE}}\,;$$

its value increases with increasing I_C. Since

$$\frac{\mathrm{d}T}{\mathrm{d}V_{CE}} = \frac{\mathrm{d}T}{\mathrm{d}P}\frac{\mathrm{d}P}{\mathrm{d}V_{CE}} = R_{th}\left(I_C+V_{CE}\frac{\mathrm{d}I_C}{\mathrm{d}V_{CE}}\right)\,, \tag{3.229}$$

one obtains the following after combination of these two equations:

$$\frac{\mathrm{d}I_\mathrm{C}}{\mathrm{d}V_\mathrm{CE}} = \frac{\left(\dfrac{\partial I_\mathrm{C}}{\partial V_\mathrm{CE}}\right)_T + R_\mathrm{th} I_\mathrm{C} \left(\dfrac{\partial I_\mathrm{C}}{\partial T}\right)_{V_\mathrm{CE}}}{1 - R_\mathrm{th} V_\mathrm{CE} \left(\dfrac{\partial I_\mathrm{C}}{\partial T}\right)_{V_\mathrm{CE}}} \; . \tag{3.230}$$

Equation (3.230) diverges as

$$R_\mathrm{th} V_\mathrm{CE} \left(\frac{\partial I_\mathrm{C}}{\partial T}\right)_{V_\mathrm{CE}} \to 1 \; , \tag{3.231}$$

corresponding to an instability: the current increases in an uncontrolled manner this may destroy the transistor under unfavorable conditions. The case of voltage control ($V_\mathrm{BE} = \mathrm{const.}$), where

$$\left(\frac{\partial I_\mathrm{C}}{\partial T}\right)_{V_\mathrm{CE}, V_\mathrm{BE}} \approx \frac{I_\mathrm{C}}{T} \left(\frac{V_\mathrm{g} - V_\mathrm{BE}}{V_\mathrm{T}} + X_\mathrm{TI}\right) \; ,$$

is worse in this respect than the case of current control ($I_\mathrm{B} = \mathrm{const.}$), where

$$\left(\frac{\partial I_\mathrm{C}}{\partial T}\right)_{V_\mathrm{CE}, I_\mathrm{B}} \approx \alpha_\mathrm{B} I_\mathrm{C} \; ,$$

and α_B denotes the temperature coefficient of the current gain.

Two-Port Parameters

Within a small-signal description, it is possible to represent the effects of self-heating by modified admittance parameters $y^\mathrm{T}_{\alpha\beta\mathrm{e}}$, which yield the small-signal terminal currents as a response to the applied voltages according to

$$\begin{pmatrix} \underline{i}_\mathrm{b} \\ \underline{i}_\mathrm{c} \end{pmatrix} = \begin{pmatrix} y^\mathrm{T}_{11\mathrm{e}} & y^\mathrm{T}_{12\mathrm{e}} \\ y^\mathrm{T}_{21\mathrm{e}} & y^\mathrm{T}_{22\mathrm{e}} \end{pmatrix} \begin{pmatrix} \underline{v}_\mathrm{be} \\ \underline{v}_\mathrm{ce} \end{pmatrix} \; . \tag{3.232}$$

If we consider I_B and I_C to be functions of V_BE, V_CE and the temperature T, the small-signal currents can be expressed as total differentials under low-frequency conditions:

$$\begin{aligned} i_\mathrm{b}(t) &= \frac{\partial I_\mathrm{B}}{\partial V_\mathrm{BE}} v_\mathrm{be}(t) + \frac{\partial I_\mathrm{B}}{\partial V_\mathrm{CE}} v_\mathrm{ce}(t) + \frac{\partial I_\mathrm{B}}{\partial T} \vartheta_\mathrm{j}(t) \; , \\ i_\mathrm{c}(t) &= \frac{\partial I_\mathrm{C}}{\partial V_\mathrm{BE}} v_\mathrm{be}(t) + \frac{\partial I_\mathrm{C}}{\partial V_\mathrm{CE}} v_\mathrm{ce}(t) + \frac{\partial I_\mathrm{C}}{\partial T} \vartheta_\mathrm{j}(t) \; , \end{aligned}$$

where $\vartheta_\mathrm{j}(t)$ denotes the small-signal variation of the device temperature. At higher frequencies, these relations have to be generalized to

$$\begin{pmatrix} \underline{i}_\mathrm{b} \\ \underline{i}_\mathrm{c} \end{pmatrix} = \begin{pmatrix} y_{11\mathrm{e}} & y_{12\mathrm{e}} \\ y_{21\mathrm{e}} & y_{22\mathrm{e}} \end{pmatrix} \begin{pmatrix} \underline{v}_\mathrm{be} \\ \underline{v}_\mathrm{ce} \end{pmatrix} + \begin{pmatrix} \partial I_\mathrm{B}/\partial T \\ \partial I_\mathrm{C}/\partial T \end{pmatrix} \underline{\vartheta}_\mathrm{j} \; ,$$

where the phasor $\underline{\vartheta}_{\mathrm{j}}$ of the small-signal device temperature is related to the small-signal portion of the dissipated power by the (small-signal) thermal impedance z_{th}:

$$\underline{\vartheta}_{\mathrm{j}} = z_{\mathrm{th}}\, \underline{p} \,.$$

This thermal impedance generally has to be considered as a function of device temperature, owing to the temperature dependence of the thermal conductivity. Neglecting terms of second order in the small-signal quantities, the phasor of the small-signal portion of the dissipated power reads

$$\underline{p} = V_{\mathrm{BE}}\, \underline{i}_{\mathrm{b}} + I_{\mathrm{B}}\, \underline{v}_{\mathrm{ce}} + V_{\mathrm{CE}}\, \underline{i}_{\mathrm{c}} + I_{\mathrm{C}}\, \underline{v}_{\mathrm{ce}} \,.$$

Since $\underline{i}_{\mathrm{b}} = y_{11\mathrm{e}}^{\mathrm{T}} \underline{v}_{\mathrm{be}} + y_{12\mathrm{e}}^{\mathrm{T}} \underline{v}_{\mathrm{ce}}$ and $\underline{i}_{\mathrm{c}} = y_{21\mathrm{e}}^{\mathrm{T}} \underline{v}_{\mathrm{be}} + y_{22\mathrm{e}}^{\mathrm{T}} \underline{v}_{\mathrm{ce}}$ as a consequence of (3.232), the equations may be combined and solved for the admittance parameters $y_{\alpha\beta\mathrm{e}}^{\mathrm{T}}$, which take account of self-heating effects. The results are

$$\begin{aligned}
y_{11\mathrm{e}}^{\mathrm{T}} &= \frac{1}{\Lambda}\left\{ y_{11\mathrm{e}} + z_{\mathrm{th}}\left[I_{\mathrm{B}}\frac{\partial I_{\mathrm{B}}}{\partial T} - V_{\mathrm{CE}}\left(y_{11\mathrm{e}}\frac{\partial I_{\mathrm{C}}}{\partial T} - y_{21\mathrm{e}}\frac{\partial I_{\mathrm{B}}}{\partial T}\right)\right]\right\} , \\
y_{12\mathrm{e}}^{\mathrm{T}} &= \frac{1}{\Lambda}\left\{ y_{12\mathrm{e}} + z_{\mathrm{th}}\left[I_{\mathrm{C}}\frac{\partial I_{\mathrm{B}}}{\partial T} - V_{\mathrm{CE}}\left(y_{12\mathrm{e}}\frac{\partial I_{\mathrm{C}}}{\partial T} - y_{22\mathrm{e}}\frac{\partial I_{\mathrm{B}}}{\partial T}\right)\right]\right\} , \\
y_{21\mathrm{e}}^{\mathrm{T}} &= \frac{1}{\Lambda}\left\{ y_{21\mathrm{e}} + z_{\mathrm{th}}\left[I_{\mathrm{B}}\frac{\partial I_{\mathrm{C}}}{\partial T} - V_{\mathrm{BE}}\left(y_{21\mathrm{e}}\frac{\partial I_{\mathrm{B}}}{\partial T} - y_{11\mathrm{e}}\frac{\partial I_{\mathrm{C}}}{\partial T}\right)\right]\right\} , \\
y_{22\mathrm{e}}^{\mathrm{T}} &= \frac{1}{\Lambda}\left\{ y_{22\mathrm{e}} + z_{\mathrm{th}}\left[I_{\mathrm{C}}\frac{\partial I_{\mathrm{C}}}{\partial T} - V_{\mathrm{BE}}\left(y_{22\mathrm{e}}\frac{\partial I_{\mathrm{B}}}{\partial T} - y_{12\mathrm{e}}\frac{\partial I_{\mathrm{C}}}{\partial T}\right)\right]\right\}
\end{aligned}$$

where

$$\Lambda = 1 - z_{\mathrm{th}}\left(V_{\mathrm{BE}}\frac{\partial I_{\mathrm{B}}}{\partial T} + V_{\mathrm{CE}}\frac{\partial I_{\mathrm{C}}}{\partial T}\right) .$$

As z_{th} becomes very small in the frequency range where internal transistor capacitances have a substantial effect on the small-signal admittance parameters, we may replace $y_{\alpha\beta\mathrm{e}}$ in the terms multiplied by z_{th} by the corresponding dc values; since

$$\frac{y_{21\mathrm{e}}}{y_{11\mathrm{e}}} = \beta \quad \text{and} \quad \frac{\partial I_{\mathrm{C}}}{\partial T} = \beta\,\frac{\partial I_{\mathrm{B}}}{\partial T} ,$$

the terms $y_{11\mathrm{e}}(\partial I_{\mathrm{C}}/\partial T) - y_{21\mathrm{e}}(\partial I_{\mathrm{B}}/\partial T)$ thus are almost zero and may generally be neglected. In this case, if g_{o} is assumed to be small, the following approximations hold:

$$\begin{pmatrix} \underline{i}_{\mathrm{b}} \\ \underline{i}_{\mathrm{c}} \end{pmatrix} = \frac{1}{\Lambda}\begin{pmatrix} y_{11\mathrm{e}} + z_{\mathrm{th}} I_{\mathrm{B}}\dfrac{\partial I_{\mathrm{B}}}{\partial T} & y_{12\mathrm{e}} + z_{\mathrm{th}} I_{\mathrm{C}}\dfrac{\partial I_{\mathrm{B}}}{\partial T} \\ y_{21\mathrm{e}} + z_{\mathrm{th}} I_{\mathrm{B}}\dfrac{\partial I_{\mathrm{C}}}{\partial T} & y_{22\mathrm{e}} + z_{\mathrm{th}} I_{\mathrm{C}}\dfrac{\partial I_{\mathrm{C}}}{\partial T} \end{pmatrix}\begin{pmatrix} \underline{v}_{\mathrm{be}} \\ \underline{v}_{\mathrm{ce}} \end{pmatrix} .$$

3.12 Parameter Extraction – DC Measurements

Parameters for compact models are usually extracted from measured data. A full data set requires both dc measurements and ac measurements. Measuring the dc characteristics of high-frequency bipolar transistors requires current measurements down to the pA range, as well as the measurement of small voltage differences. This is possible with special instrumentation and an appropriate measurement setup. Owing to the strong temperature dependence of some parameters, accurate temperature control and the consideration of self-heating effects are mandatory.

Ordinary digital multimeters are generally not suited to low-level current measurements or to voltage measurements with high source resistances: current measurements below 1 nA and voltage measurements with source resistances above 1 MΩ require the use of electrometers, which have an extremely high input impedance, typically in excess of 100 TΩ. Such electrometers use operational amplifiers with a field effect transistor input. DC characterization of semiconductor devices is generally performed with special instruments, which combine several current and voltage outputs with several current and voltage measurement inputs in a single case.

To fully exploit the potentialities of electrometers for low-level measurements, high-resistance insulation in the test circuit is mandatory. This implies both careful choice of the insulating material and cabling and periodic tests of test fixtures and cables to ensure their integrity. Because of a specific resistivity that is typically in excess of $10^{17}\ \Omega\,\mathrm{cm}$ and its high resistance to water absorption, Teflon is a satisfactory and widely used insulator for electrometer measurements down to 10 fA. To avoid degradation of insulator performance, the insulator should not be touched with a hand or with any material that might contaminate the surface. Cleaning of a contaminated insulator with alcohol (methanol) can be tried.

3.12.1 Gummel Plot

The parameters I_{S}, N_{F}, B_{F}, I_{SE}, N_{E} and I_{KF} are generally determined in the common-base configuration with $V_{\mathrm{BC}} = 0$, using two ammeters for the measurement of $I_{\mathrm{B}}(V_{\mathrm{BE}})$ and $I_{\mathrm{C}}(V_{\mathrm{BE}})$ as shown in Fig. 3.41,[47] which also shows the input and transfer characteristics $I_{\mathrm{B}}(V_{\mathrm{B'E'}})$ and $I_{\mathrm{C}}(V_{\mathrm{B'E'}})$ expected from the Gummel–Poon model. Neglecting series resistance effects ($V_{\mathrm{BE}} = V_{\mathrm{B'E'}}$), we have

[47] This measurement setup may produce unwanted oscillations under certain biasing conditions; these may generally be suppressed by placing ferrite beads around the supply lines.

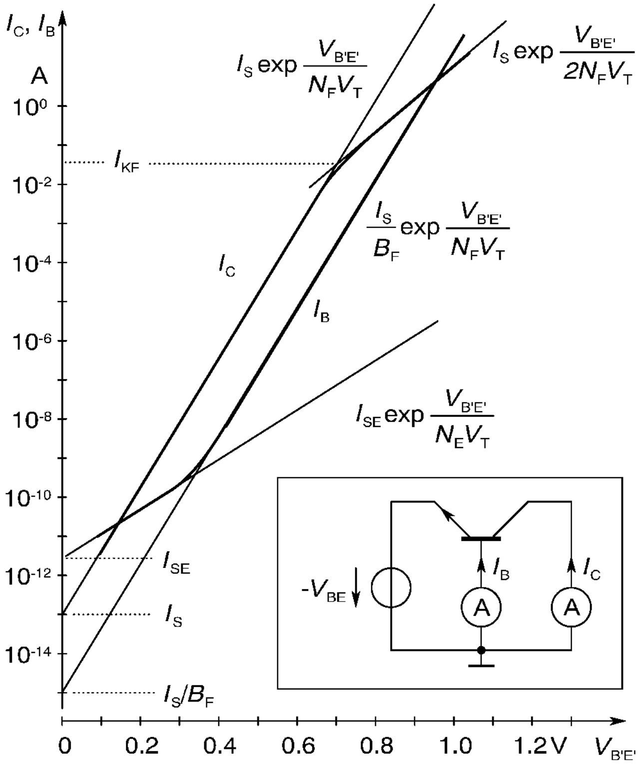


Fig. 3.41. Gummel plot of input and transfer characteristics of a bipolar transistor in the common-base configuration ($V_{BC} = 0$) according to the SPICE Gummel–Poon model, neglecting series resistance effects (after [132])

$$I_C \approx \frac{I_S}{q_B} \exp\left(\frac{V_{BE}}{N_F V_T}\right) \quad \text{and}$$

$$I_B \approx \frac{I_S}{B_F} \exp\left(\frac{V_{BE}}{N_F V_T}\right) + I_{SE} \exp\left(\frac{V_{BE}}{N_E V_T}\right) ,$$

if $V_{CB} = 0$ and $V_{BE} \gg V_T$. A semilogarithmic plot of the base and collector currents versus V_{BE} – usually referred to as a Gummel plot – should therefore result in straight lines. As long as $I_C \ll I_{KF}$, the logarithm of the collector current increases linearly with V_{BE} and allows one to determine the transfer saturation current I_S. The emission coefficients N_R and N_F can be taken to be equal to one, a priori ($N_F = N_R = 1$), as will be assumed in the following. For $V_{BE}/V_{AR} \ll 1$, the approximation $1 - V_{BE}/V_{AR} \approx \exp(-V_{BE}/V_{AR})$ holds, and therefore, under low-level-injection conditions,

$$I_C \approx I_S \exp\left[V_{BE}\left(\frac{1}{V_T} - \frac{1}{V_{AR}}\right)\right] . \tag{3.233}$$

This results in an apparent "emission coefficient"

$$\text{“}N_F\text{”} \approx \frac{V_{AR}}{V_{AR} - V_T} \approx 1 + \frac{V_T}{V_{AR}} ,$$

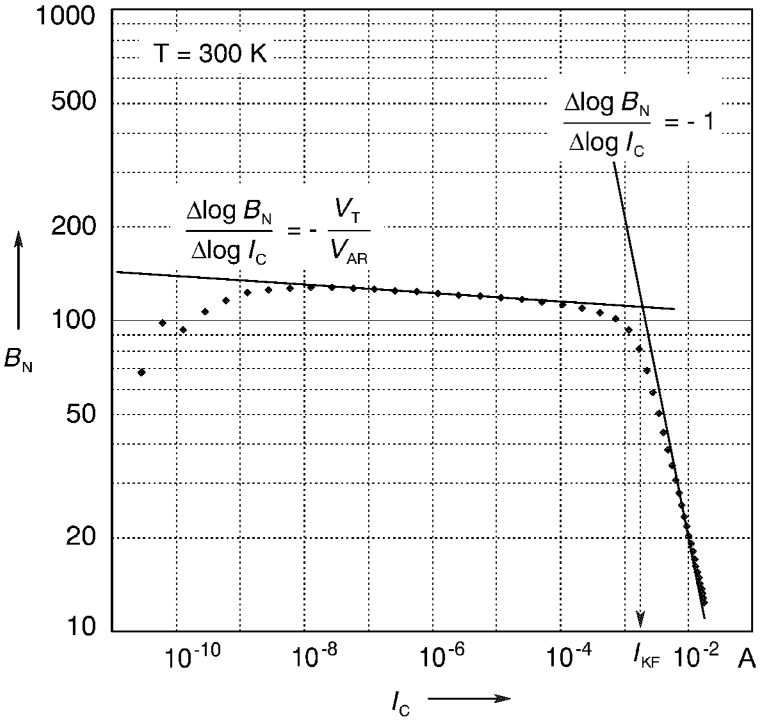


Fig. 3.42. Measured values of large-signal current gain $B_N = I_C/I_B$ and extraction of forward knee current I_{KF}

because of the bias dependence of q_B, described in terms of the reverse Early voltage V_{AR}. For a consistent extraction, both I_S and "N_F" should be fitted to the exponential portion of the transfer current characteristic; this yields the correct value of I_S together with the apparent emission coefficient "N_F", which allows one to determine V_{AR}. For $I_C \gg I_{KF}$, the logarithm of the collector current also increases linearly with $V_{B'E'}$, but with a reduced slope $1/(2V_T)$; taking $I_{EC} \approx 0$ yields

$$I_C \approx \frac{I_{CE}}{q_1/2 + \sqrt{q_1^2/4 + I_{CE}/I_{KF}}} .$$

If $I_{CE} \gg I_{KF}$, the value of $q_1 \approx 1$ may be neglected in this expression, resulting in

$$I_C \approx \sqrt{I_{CE} I_{KF}} \approx \sqrt{I_S I_{KF}} \exp\left(\frac{V_{B'E'}}{2V_T}\right) . \tag{3.234}$$

The intersection of the two approximating lines determines the forward knee current I_{KF}. A plot of I_C versus $V_{B'E'}$ is not possible in this current range, owing to series resistance effects; the value of I_{KF} is therefore determined from a double-logarithmic plot of B_N versus I_C, as shown in Fig. 3.42.

At large values of the collector current I_C, a decrease in $B_N \sim 1/I_C$ is expected from the Gummel Poon model, corresponding to a line of slope -1 in the double-logarithmic plot. At smaller values of I_C, the β Plateau is seen, with a slight bias dependence due to the reverse Early effect. Using $I_B = (I_S/B_F)\exp(V_{B'E'}/V_T)$ and the approximation (3.233), we may estimate

$$B_N \approx B_F \exp\left(-\frac{V_{BE}}{V_{AR}}\right) \approx B_F \left(\frac{I_C}{I_S}\right)^{-V_T/(V_{AR}-V_T)} \sim I_C^{-V_T/V_{AR}} ,$$

corresponding to a line of slope $-V_T/V_{AR}$ in the double-logarithmic representation of B_N versus I_C.

The parameters I_{SE}, N_E and B_F are determined from a semilogarithmic plot of the $I_B(V_{BE})$ characteristic (see Fig. 3.41). The parameters N_R, B_R, I_{SC}, N_C and I_{KR} are determined analogously from the corresponding characteristics in reverse operation ($V_{BE} < 0$, $V_{BC} > 0$).

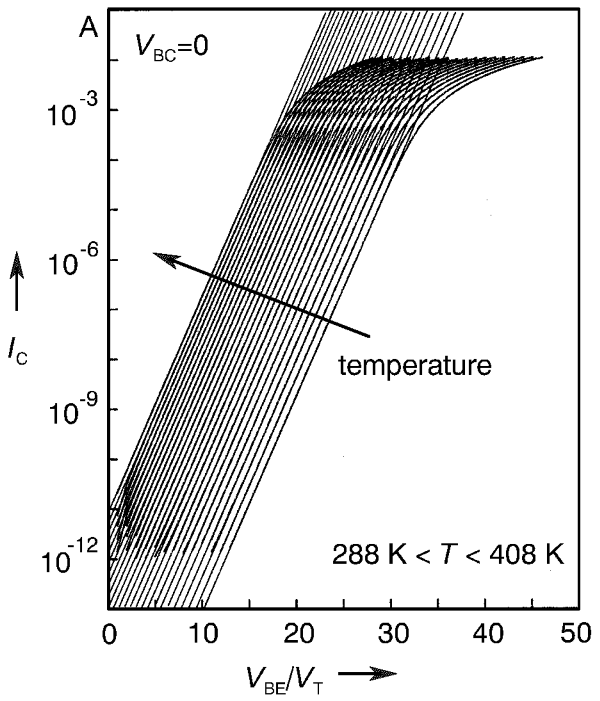


Fig. 3.43. Transfer characteristics measured at different temperatures T between 288 K and 408 K versus V_{BE}/V_T, used for the extraction of $I_S(T)$

The bandgap voltage E_G is determined from a set of transfer-current characteristics measured at different temperatures, such as the family of curves shown in Fig. 3.43, where the transfer current I_C has been plotted on a loga-

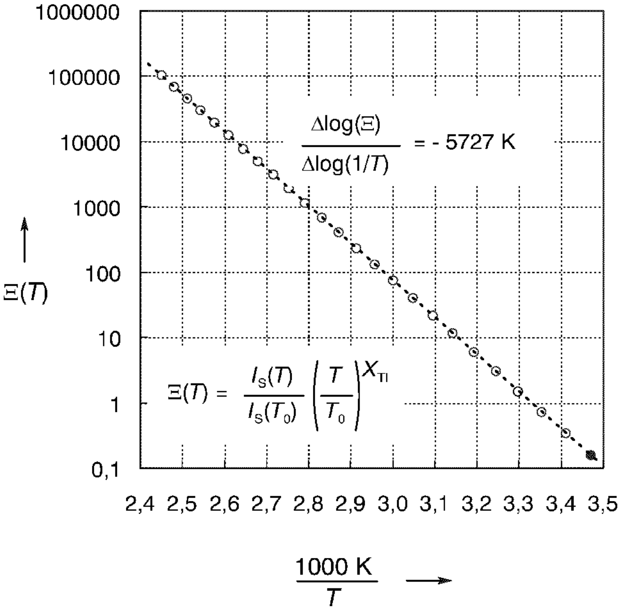


Fig. 3.44. Determination of the parameter E_G from the temperature variation of the saturation current; X_{TI} was chosen to be 3.5, $T_0 = 300\,\mathrm{K}$, $I_S(T_0) = 2.8 \times 10^{-17}$ A

rithmic scale versus V_{BE}/V_T. For each of these curves, the transfer saturation current $I_S(T)$ is extracted by linear extrapolation. For each temperature, the value of

$$\Xi(T) = \frac{I_S(T)}{I_S(T_0)} \left(\frac{T_0}{T}\right)^{X_{TI}}$$

is then calculated, where the parameter X_{TI} is chosen on the basis of theoretical considerations ($X_{TI} = 3.5$ for silicon). Since

$$\log(\Xi) = E_G \frac{e \log(e)}{k_B} \left(\frac{1}{T_0} - \frac{1}{T}\right),$$

a logarithmic plot of $\Xi(T)$ versus $1/T$ should yield a straight line with slope

$$\frac{\Delta \log(\Xi)}{\Delta \log(1/T)} = -\frac{e \log(e)}{k_B} E_G . \tag{3.235}$$

Figure 3.44 shows measured data, which show the expected behavior. From the observed slope, the value $E_G = 1.136$ V can be obtained for the device under test.

The parameter X_{TB} is determined from a double-logarithmic plot of $B_F(T)$ versus T. Since, according to (3.147),

$$\log[B_F(T)] = \log[B_F(T_0)] + X_{TB} \log(T/T_0) , \tag{3.236}$$

the value of X_{TB} is obtained as the slope of the line.

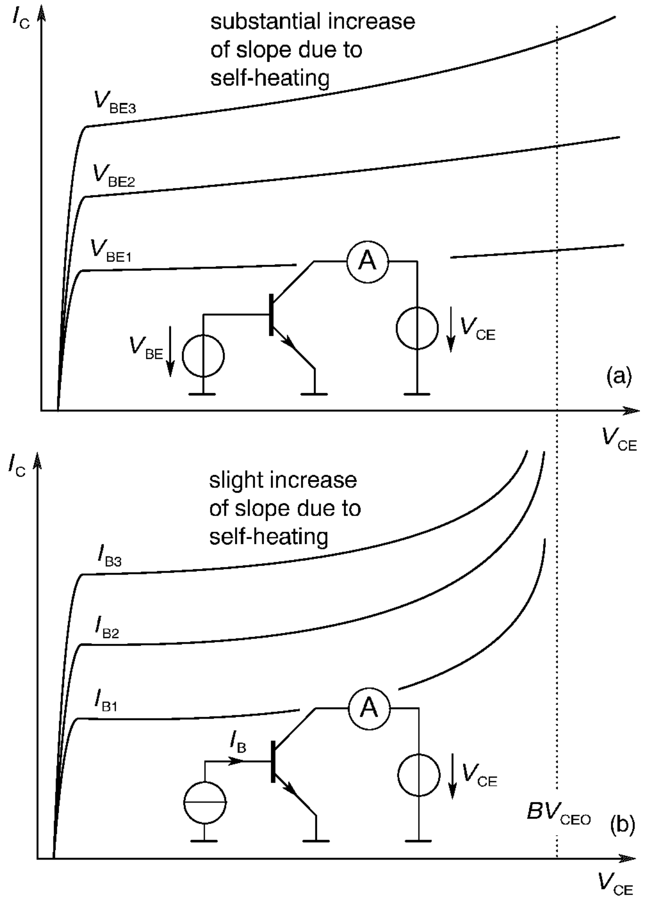


Fig. 3.45. Output characteristics (schematic) measured with (**a**) V_{BE} held constant and (**b**) I_B held constant

3.12.2 Output Characteristics, Early Voltage

The output characteristics $I_C(V_{CE})$ of a bipolar transistor are measured with either V_{BE} or I_B held constant (Fig. 3.45). Since in the Gummel–Poon model the transfer current is described as a voltage-controlled current source, the output characteristics are preferably determined with V_{BE} held constant. In this case carrier multiplication in the bc diode has only a very small effect on the output characteristics, since the additional hole current injected into the base from the bc junction leaves the device via the base contact. The situation is different if the output characteristics are measured with I_B held constant, a method that is often used because in that case self-heating has less effect on the output characteristics (see Sect. 3.11), because

$$I_B \left(\frac{\partial B_N}{\partial T} \right)_{I_B, V_{CE}} \ll \left(\frac{\partial I_C}{\partial T} \right)_{V_{BE}, V_{CE}} .$$

Another feature that makes measurements with I_B = const. appealing is that a set of approximately equally spaced output characteristics is obtained from a series of measurements with a constant increment of the applied base current. For high-frequency bipolar transistors, with their small BV_{CEO} values, however, this method cannot be recommended, owing to the strong effect of carrier multiplication in the bc diode. Figure 3.46 shows a set of output char-

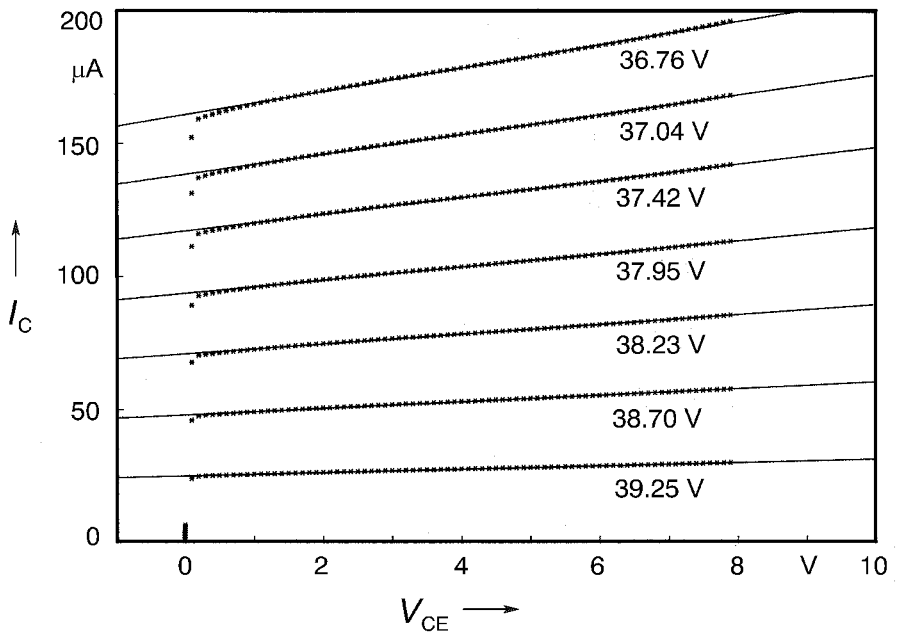


Fig. 3.46. Measured output characteristics and extrapolated Early voltages V_A of a self-aligned vertical npn bipolar transistor with the temperature on the wafer back surface held constant. The curves correspond to values of V_{BE} in the interval $0.65\ \mathrm{V} < V_{BE} < 0.71\ \mathrm{V}$

acteristics measured with V_{BE} used as a parameter. The extrapolated Early voltage V_A is obtained by linear extrapolation of the measured $I_C(V_{CB})$ curves to $I_C = 0$. If this is done by a least-squares fit to the data in the voltage interval $[V_{CB1}, V_{CB2}]$, the parameters I_{CO} and V_A that minimize the integral

$$\int_{V_{CB1}}^{V_{CB2}} \left[I_C(V_{CB}) - I_{CO}\left(1+\frac{V_{CB}}{V_A}\right)\right]^2 \mathrm{d}V_{CB} \,,$$

must be determined; with $\langle V_{CB}\rangle = (V_{CB1} + V_{CB2})/2$ this results in

$$V_A = \frac{(V_{CB2}-V_{CB1})^2 \int_{V_{CB1}}^{V_{CB2}} I_C(V_{CB})\,dV_{CB}}{12\int_{V_{CB1}}^{V_{CB2}} I_C(V_{CB})(V_{CB}-\langle V_{CB}\rangle)\,dV_{CB}} - \langle V_{CB}\rangle\,.$$

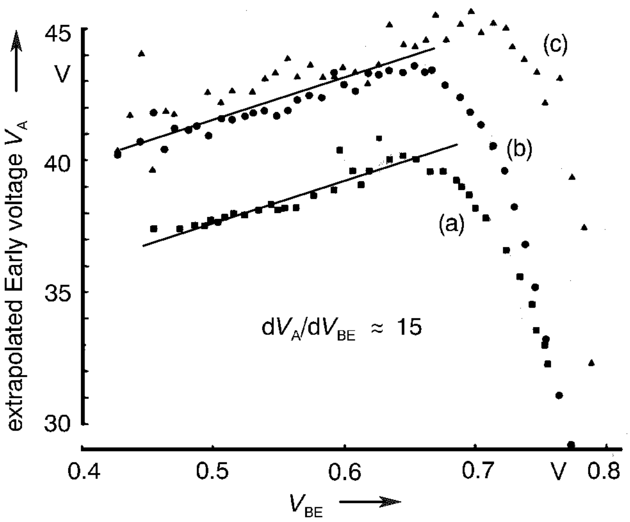


Fig. 3.47. Extrapolated Early voltages for different values of V_{BE} for three transistors with different layouts on the same chip: (**a**) small and (**b**), (**c**) large emitter stripe width

Figure 3.47 shows extrapolated Early voltages determined for different values of the parameter V_{BE}. At small values of V_{BE}, the Early voltage increases with V_{BE}. The reason for this is that the Early voltage determines the relative change of the base charge: Owing to the increase of q_B with V_{BE}, the relative change of q_B with V_{CB} is smaller, resulting in a larger extrapolated Early voltage V_A. This effect is not correctly represented in the conventional SPICE Gummel–Poon model, where [133]

$$q_B \approx q_1 \approx \left(1-\frac{V_{BE}}{V_{AR}}+\frac{V_{CB}}{V_{AF}}\right)^{-1}$$

is used under low-level-injection conditions. Since $V_{CB} = V_{CE} - V_{BE}$, the transfer current in this case obeys the proportionality

$$I_C \sim \frac{1}{q_B} \approx \left(1-\frac{V_{BE}}{V_{AR}^*}\right)\left(1+\frac{V_{CE}}{V_{AF}(1-V_{BE}/V_{AR}^*)}\right),$$

where $V_{AR}^* = V_{AR}V_{AF}/(V_{AR}+V_{AF})$. From this, a bias-dependent value of the extrapolated Early voltage is obtained,

$$V_A \approx V_{AF}(1-V_{BE}/V_{AR}^*)\,,$$

that decreases with increasing V_{BE}, in contrast to what is observed experimentally. Modeling the normalized base charge under low-level-injection conditions in terms of the depletion capacitances, i.e.

$$q_1 = 1 + \frac{1}{Q_{\mathrm{B0}}} \int_0^{V_{\mathrm{BE}}} c_{\mathrm{je}}(V)\,\mathrm{d}V + \frac{1}{Q_{\mathrm{B0}}} \int_0^{V_{\mathrm{BC}}} c_{\mathrm{jc}}(V)\,\mathrm{d}V$$

as in the VBIC, MEXTRAM and HICUM models, is better in this respect.

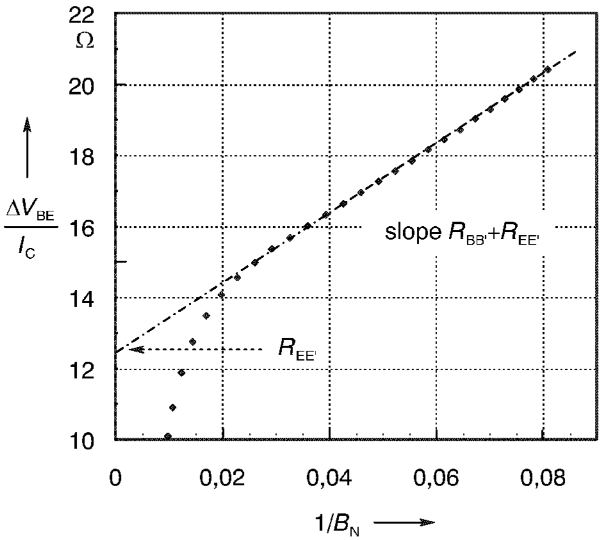


Fig. 3.48. Extraction of base and emitter resistances from measured data using a plot of $\Delta V_{\mathrm{BE}}/I_{\mathrm{C}}$ versus $1/B_{\mathrm{N}}$

3.12.3 Series Resistances

Extraction of Base and Emitter Resistances from the Input and Transfer Characteristics

The method of Ning and Tang [134] for the extraction of the emitter and base resistances assumes negligible high-level-injection effects of the base current and a constant emission coefficient N_{E}, i.e. an input voltage–current characteristic of the form

$$V_{\mathrm{B'E'}} \approx N_{\mathrm{E}} V_{\mathrm{T}} \ln(I_{\mathrm{B}}/I_{\mathrm{SE}}) \,,$$

if $I_{\mathrm{B}} \gg I_{\mathrm{SE}}$. This relation allows one to compute $V_{\mathrm{B'E'}}$ for any given value of I_{B}. The difference $\Delta V_{\mathrm{BE}} = V_{\mathrm{BE}} - V_{\mathrm{B'E'}}$ is equal to the sum of the voltage drops across $R_{\mathrm{BB'}}$ and $R_{\mathrm{EE'}}$:

$$\Delta V_{\mathrm{BE}} = R_{\mathrm{BB'}} I_{\mathrm{B}} + R_{\mathrm{EE'}} I_{\mathrm{E}} = \left(\frac{R_{\mathrm{BB'}} + R_{\mathrm{EE'}}}{B_{\mathrm{N}}} + R_{\mathrm{EE'}} \right) I_{\mathrm{C}} \,,$$

and correspondingly

$$\frac{\Delta V_{\mathrm{BE}}}{I_{\mathrm{C}}} = R_{\mathrm{EE}'} + \frac{R_{\mathrm{BB}'} + R_{\mathrm{EE}'}}{B_{\mathrm{N}}} . \tag{3.237}$$

Since the large-signal current gain B_{N} decreases owing to high-level-injection effects in the base and collector regions, a plot of $\Delta V_{\mathrm{BE}}/I_{\mathrm{C}}$ versus $1/B_{\mathrm{N}}$ should yield a straight line with slope $R_{\mathrm{EE}'}+R_{\mathrm{BB}'}$ and offset $R_{\mathrm{EE}'}$, as shown in Fig. 3.48. This often works fairly well; possible sources of error are high-level-injection effects, nonideal behavior of the base current and self-heating effects.[48]

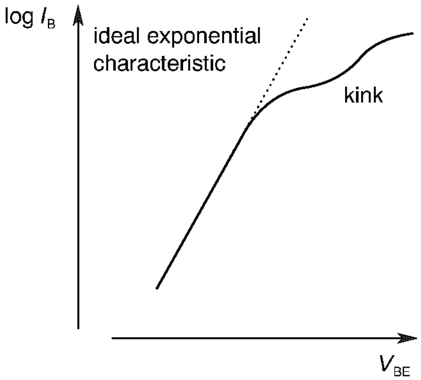


Fig. 3.49. Kink effect in the input characteristic

The semilogarithmic plot of the input characteristic occasionally shows a pronounced kink, as depicted in Fig. 3.49. There are several possible explanations for this effect:

- A substantial value of the collector series resistance may forward bias the bc diode and thus increase the base current owing to the additional component I_{BC}. If this is the case, the kink should vanish if the bc diode is reverse biased.
- An increase in the base charge due to the diffusion charge and base pushout may lead to increased recombination in the base region and therefore to an increase in the base current due to the component I_{BB}.
- The emitter series resistance, together with a bias-dependent value of the current gain due to high-level-injection effects, is another possible reason.

[48]Sometimes $R_{\mathrm{BB}'} = R_{\mathrm{BM}} + R_{\mathrm{Bint}}$ is represented as a sum of an external base resistance R_{BM} (assumed to be ohmic) and a bias-dependent internal base resistance R_{Bint}. If emitter current crowding is neglected, the bias dependence of R_{Bint} is due to an increase of q_{B}. Owing to the proportionality $B_{\mathrm{N}} \sim 1/q_{\mathrm{B}} \sim R_{\mathrm{Bint}}$, the ratio $R_{\mathrm{Bint}}/B_{\mathrm{N}}$ should then be approximately constant. Under this assumption, the offset of the line fitted to the $\Delta V_{\mathrm{BE}}/I_{\mathrm{C}}$ data should equal $R_{\mathrm{EE}'} + R_{\mathrm{Bint}}/B_{\mathrm{N}}$.

If we neglect the effect of the base series resistance, the characteristic is shifted to the right by

$$\Delta V_{\mathrm{BE}} = R_{\mathrm{EE'}}(I_{\mathrm{B}}+I_{\mathrm{C}}) = R_{\mathrm{EE'}}(B_{\mathrm{N}}+1)I_{\mathrm{B}} . \tag{3.238}$$

If the current gain B_{N} decreases with increasing I_{B} owing to high-level-injection effects, the shift to the right of the characteristic from its ideal exponential form becomes less pronounced, resulting in a kink in the input characteristic.

Characterization of Special Test Structures

A direct approach to the determination of the base resistance makes use of a special test structure with two disconnected base contacts [135, 136]. One of

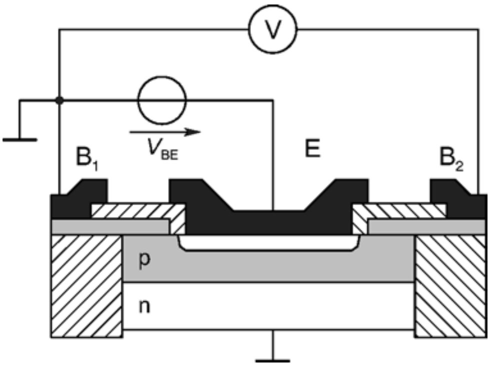


Fig. 3.50. Schematic cross section of a special bipolar-transistor structure with two separate base contacts for measurement of base series resistance [136]

these base contacts (B_1) is used to supply the base current, while the other (B_2) is used to sense the voltage drop $V_{\mathrm{B_1B_2}}$ between the two contacts, as illustrated in Fig. 3.50. The base series resistance approximately follows the relation

$$R_{\mathrm{B_1B_2}} = R_{\mathrm{BM}} + R'_{\mathrm{Bint}}(W_{\mathrm{E}}-2\Delta) ,$$

where Δ describes the deviation of the layout dimension W_{E} of the emitter window from the width of the effective internal base region. The parameters R_{BM}, R'_{Bint} and Δ are determined from a plot of $R_{\mathrm{B_1B_2}}$ values for different values of W_{E}, as illustrated in Fig. 3.51.

The base resistance $R_{\mathrm{B_1B_2}}$ determined by this procedure deviates somewhat from the lumped base resistance of the Gummel–Poon model. It is possible to investigate the relation between these values using the base resistance model of Hauser outlined in Sect. 3.6, which gives the following for the voltage drop across the internal base layer of the test structure [77]:

$$\Delta V_{\mathrm{Bi}} = -2V_{\mathrm{T}}\ln[\cos(z)] = V_{\mathrm{B_1B_2}} - R_{\mathrm{BM}}I_{\mathrm{B}} .$$

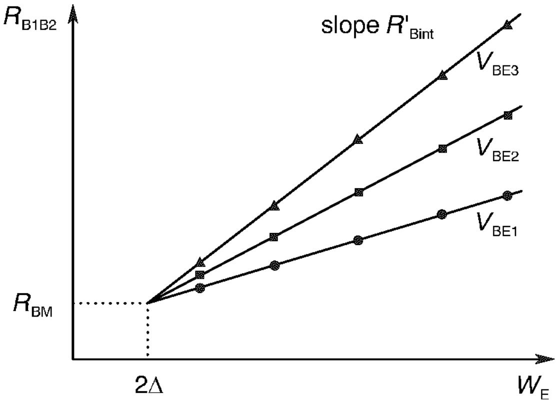


Fig. 3.51. Measurement of base series resistance using a special bipolar-transistor structure with two separate base contacts; typical measurement results, after [136]

If the value of R_{BM} is known, this allows one to determine z as follows [137]:

$$z = \arccos\left[\exp\left(-\frac{V_{B_1B_2} - R_{BM}I_B}{2\,V_T}\right)\right] .$$

This result may be used for the computation of the base resistance from a power dissipation calculation, resulting in

$$R_{BB'} = R_{Bext} + R'_{Bint}\,(W_E - 2\Delta)\,\frac{\tan(z) - z}{z\tan^2(z)} .$$

It should be emphasized, however, that such a transformation neglects conductance modulation of the base layer and can therefore provide only a crude approximation; attempts to improve further on the method have been presented in [138].

The Open-Collector Method

The open-collector method [132, 139, 140] is a dc method for the determination of the emitter resistance $R_{EE'}$. The voltage V_{CE} between the emitter and collector is measured as a function of the base current I_B. Assuming a voltmeter of infinite resistance, $I_C = 0$ and therefore $I_B = I_E$. Since no voltage drop occurs across the collector resistance, because $I_C = 0$, the measured voltage V_{CE} obeys the relation

$$V_{CE} = V_{C'E'} + R_{EE'}I_B . \tag{3.239}$$

If V_{CE} and I_B are given, the value of $V_{C'E'}$ has to be determined from the equivalent circuit, which under dc conditions simplifies to that shown in Fig. 3.52. At node B′, the base current I_B splits up into a portion I_{BC} that flows into the collector region and a portion I_{BE} that flows into the emitter region. Since $I_C = 0$, the relation $I_{BC} = I_{EC}/B_R = (I_{CE} - I_{EC})/q_B$ has to be fulfilled. Use of the current–voltage relations for I_{CE} and I_{EC} yields

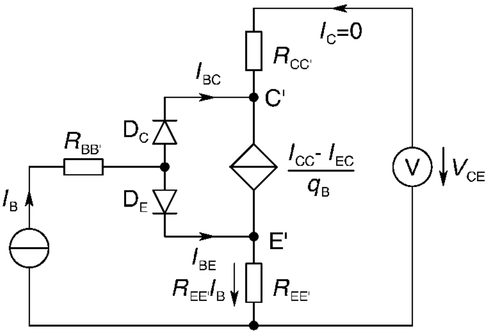


Fig. 3.52. Determination of emitter series resistance by the open-collector method (Gummel–Poon equivalent circuit)

$$V_{C'E'} \approx V_T \ln(1 + q_B/B_R) \ . \tag{3.240}$$

The solution (3.240) is independent of I_B as long as q_B and B_R can be assumed to be constant. In the analysis of measured V_{CE} characteristics, the value of $V_{C'E'}$ can be assumed to be approximately constant if this is the case. Under these conditions the emitter resistance follows from the slope of the $V_{CE}(I_B)$ characteristic: $R_{EE'} \approx dV_{CE}/dI_B$. Application of this method to integrated bipolar transistors will cause errors owing to the influence of the epitaxial collector region and the parasitic pnp transistor formed between the base, collector and substrate; a discussion of the subtleties of the open-collector method in this context has been presented in [141].

3.12.4 Carrier Multiplication and Open-Base Breakdown

The open-base breakdown voltage BV_{CEO}, which determines the maximum voltage that may be applied between the collector and emitter with a high-impedance base drive, is an important parameter that is measured for process monitoring. For the purposes of compact modeling, the bias dependence of the carrier multiplication of injected electrons has to be determined.

Open-Base Breakdown Voltage

Open-base breakdown generally occurs because of carrier multiplication in the bc diode and therefore requires the breakdown condition (3.70) to be fulfilled. Breakdown voltages are measured with an applied test current[49] and a high-impedance voltmeter.

[49]The bias dependence of the current gain causes problems if one intends to characterize emitter–collector breakdown with a voltage source. Since at first only a small current flows, the current gain is low and V_{CE} has to be increased to a value close to the bc breakdown voltage BV_{CBO} to cause breakdown. As soon as the condition for breakdown is fulfilled, an uncontrolled increase in the current and destruction of the transistor result if no precautions are taken to limit the current. For that reason, it is better to apply a current $I_C = I_E$ (open base) and measure the resulting voltage V_{CEO} with a high-impedance voltmeter.

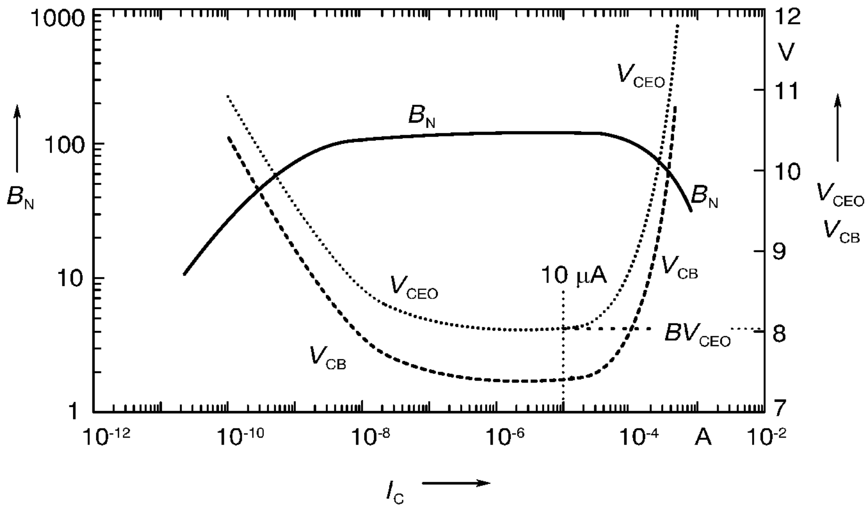


Fig. 3.53. Definition of open-base breakdown voltage

Figure 3.53 shows both the value of the current gain B_N for $V_{CB} = 0$ (i.e. $M_n = 1$) and the value of the open-base collector–emitter voltage V_{CEO} as functions of I_C. An increase in B_N results in a decrease in the value of V_{CEO} – the breakdown voltage BV_{CEO} should therefore be specified as the minimum value of V_{CEO}.

In order to find out whether variations in BV_{CEO} are due to variations in the current gain or due to other reasons, a scatter plot of BV_{CEO}/BV_{CBO} versus $1/[B_N(V_{CB} = 0) + 1]$ on a double-logarithmic scale can be helpful. If Miller's law is approximately fulfilled, the data points should lie on a straight line with slope N_1, since the breakdown condition can be written in the form

$$\left(\frac{BV_{CEO}}{BV_{CBO}}\right)^{N_1} \approx \frac{1}{B_N+1}$$

if the multiplication factor obeys Miller's formula. An example of such a plot is given in Fig. 3.54, which shows the substantial variations in BV_{CEO} observed for lots A and B to be due solely to variations in the emitter efficiency (variations in the surface recombination velocity that characterizes the polysilicon emitter contact), while lot C was found to suffer from small values of BV_{CEO} because of another effect (punchthrough).

Whether collector–emitter breakdown is caused by carrier multiplication in the bc junction or by punchthrough can be determined by comparison of the values of BV_{CEO} and BV_{CES}, which is determined with the emitter and base short-circuited: if $BV_{CES} > BV_{CEO}$, the breakdown is due to carrier multiplication in the bc diode.

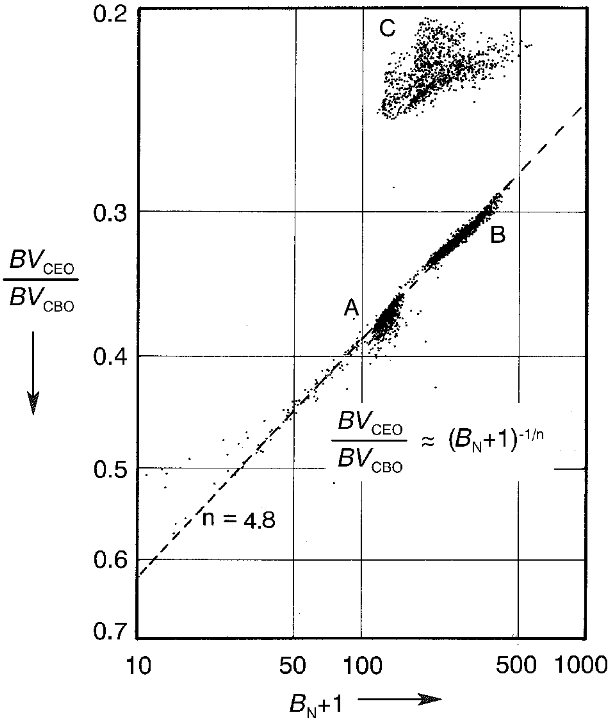


Fig. 3.54. Scatter plot of breakdown voltages

Collector–emitter breakdown at constant terminal current I_B is the worst case: all holes that are generated in the bc depletion layer flow to the emitter and determine the transfer current. Besides the values of BV_{CEO} and BV_{CES}, a breakdown voltage BV_{CER} is sometimes specified; this is determined with an applied current I_C and the base and emitter shunted with a resistor R (see Fig. 3.55a). In this case a portion V_{BE}/R of the generated holes may leave the device. The current I_{BE} that is available to forward-bias the eb junction is therefore

$$I_{BE} = (M_n - 1)I_T - \frac{V_{BE}}{R} ,$$

if the leakage current I_{CBO} of the bc diode is assumed to be small in comparison with the applied current I_C. If we take $I_C = M_n I_T = M_n B'_N I_{BE}$, the breakdown condition

$$M_n = \frac{1 + 1/B'_N}{1 - V_{BE}/(RI_C)}$$

is obtained, which reduces in the limit $R \to \infty$ to the open-base breakdown condition (3.70). Finite values of R cause a reduction in the denominator and require larger values of the multiplication factor in order to meet the breakdown condition.

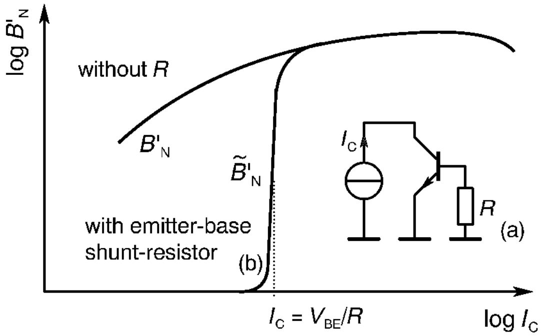


Fig. 3.55. (**a**) Measurement setup for the determination of BV_{CER}, and (**b**) effective current gain used for the discussion of measurement results

The addition of the resistor R corresponds to an effective reduction of the internal current gain B'_N to

$$\tilde{B}'_N = \left(\frac{1}{B'_N} + \frac{V_{BE}}{RI_C}\right)^{-1} ,$$

as illustrated in Fig. 3.55b. At small current values, the effective current gain is very small and $BV_{CER} > BV_{CEO}$ is obtained, whereas if $RI_C \gg V_{BEon}$ the effective current gain is only slightly affected by the shunt resistor, and $BV_{CER} \approx BV_{CEO}$. The measurement of BV_{CER} therefore does not provide any new insight and is therefore not recommended.

Multiplication Factor

For small values of the transfer current – when self-heating effects are negligible – the carrier multiplication factor M_n due to injected electrons can be determined by one of the two methods analyzed in [142]. If the base current at small values of V_{CB} is due to injection into the emitter region only, as is generally the case, its value can be written as

$$I_B(V_{BE}, V_{CB}) = I_B(V_{BE}, 0) - \frac{M_n - 1}{M_n} I_C(V_{BE}, V_{CE}) ,$$

corresponding to

$$1 - \frac{1}{M_n} = \frac{I_B(V_{BE}, 0) - I_B(V_{BE}, V_{CB})}{I_C(V_{BE}, V_{CB})} . \tag{3.241}$$

Since the multiplication of the transfer current I_C/M_n is determined from a change of the base current, which is generally smaller by two orders of magnitude, this method allows one to measure $1 - 1/M_n$ over a range of five decades. A double-logarithmic plot of $1 - 1/M_n$ versus V_{CB} therefore yields a curve such as the one depicted in Fig. 3.56. From this figure, the parameters BV_C, BV_{CBO}, N_1 and N_2 as well as the crossover voltage

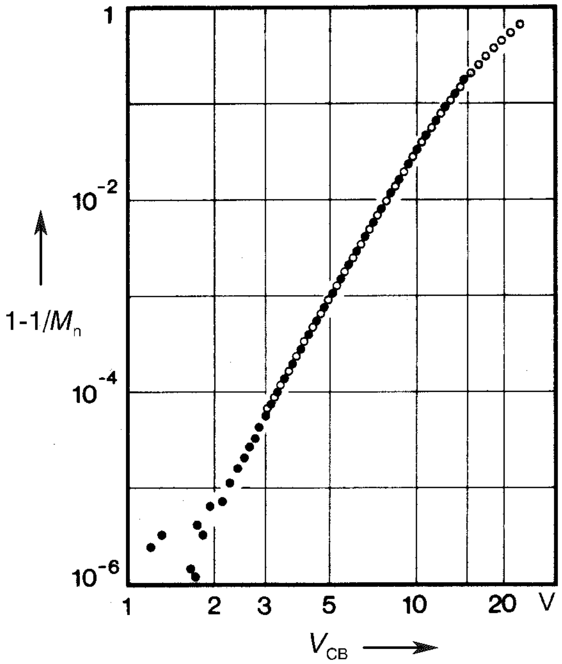


Fig. 3.56. Bias dependence of electron multiplication factor M_n (measured data) plotted in the form of $1 - 1/M_n$ versus V_{CB}

$$V_{cr} = \frac{BV_{C}^{N_1/(N_1-N_2)}}{BV_{CBO}^{N_2/(N_1-N_2)}} \tag{3.242}$$

at which the two asymptotic curves intersect, can be readily obtained, as illustrated in Fig. 3.57. This yields all of the parameters of the avalanche model outlined in Appendix D. If carrier multiplication is described as in the VBIC or MEXTRAM model, the parameters V_{JC} and M_{JC} should be determined first from capacitance measurements; the parameters A_{VC1} and A_{VC2} are then easily determined from a semi-logarithmic plot of

$$\frac{M_n(V_{CB})-1}{V_{CB}+V_{JC}} \quad \text{versus} \quad (V_{CB}+V_{JC})^{M_{JC}-1} \,.$$

Since

$$\log\left(\frac{M_n(V_{CB})-1}{V_{CB}+V_{JC}}\right) = \log(A_{VC1}) - \log(e)A_{VC2}(V_{CB}+V_{JC})^{M_{JC}-1}$$

if (3.69) holds, the parameters A_{VC1} and A_{VC2} can be determined from the offset and slope of the straight line that fits the measurement data.

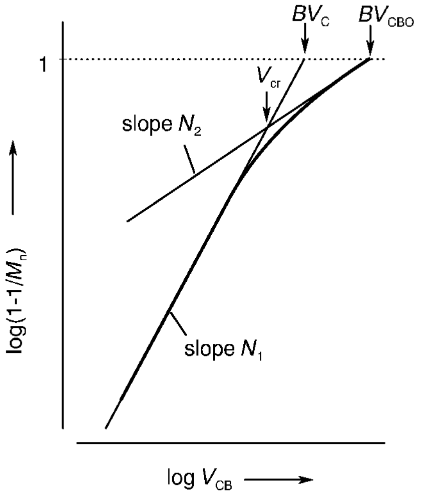


Fig. 3.57. Extraction of parameters related to the carrier multiplication factor M_n

3.12.5 Thermal Resistance, Self-Heating Effects

Current–voltage characteristics determined with standard dc measurement apparatus are generally affected by self-heating effects: in packaged semiconductor devices the temperature at the surface of the package rather than the junction temperature, and in wafer probe stations the temperature of the back surface of the wafer rather than the temperature of the device under test, is held constant. Owing to the finite thermal resistance between the device under test and the heat sink, self-heating effects may therefore influence the results of parameter measurements. As has been shown in [143], it is possible to determine the thermal resistance between the device and the heat sink from the variation of I_B with V_{CB} for values of V_{CB} sufficiently small that carrier multiplication can be neglected. Figure 3.58 shows the relative change

$$\frac{\Delta I_B}{I_B} = \frac{I_B(V_{BE}, V_{CB}) - I_B(V_{BE}, 0)}{I_B(V_{BE}, 0)}$$

in the base current associated with a change in the reverse bias V_{CB} applied to the bc diode for two values of V_{BE}. At $V_{BE} = 0.6\,\mathrm{V}$, the transfer current is small and self-heating effects may be neglected. In this case $\Delta I_B/I_B$ remains approximately zero, up to the onset of an observable hole current component in the bc diode due to impact ionization. At $V_{BE} = 0.8\,\mathrm{V}$, substantial power dissipation occurs, resulting in a device temperature that increases with V_{CB}. If the relative change in the base current $\Delta I_B/I_B$ is plotted versus the increase $\Delta P = P(V_{BE}, V_{CB}) - P(V_{BE}, 0)$ in the dissipated power,[50] an approximately linear increase with slope

[50] $P(V_{BE}, V_{CB}) = I_B V_{BE} + I_C V_{CE}$.

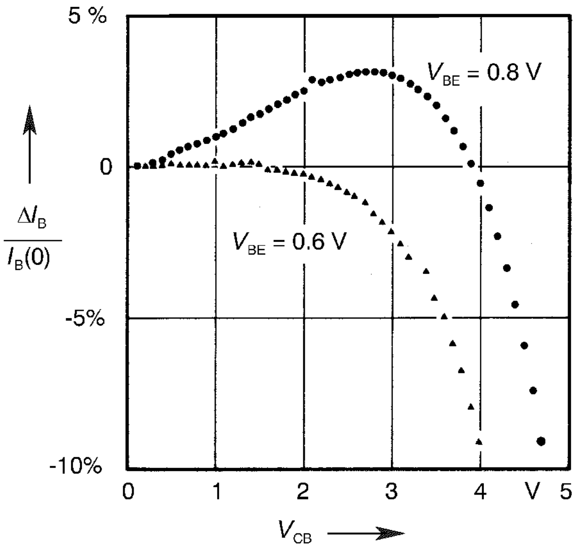


Fig. 3.58. Relative change of I_B with V_{CB} and V_{BE} held constant [143]

$$m_P = \frac{\partial}{\partial P}\left(\frac{\Delta I_B}{I_B}\right) = R_{th}\frac{1}{I_B(V_{BE},0)}\left(\frac{\partial I_B}{\partial T}\right)_{V_{BE}} \tag{3.243}$$

is observed before the onset of carrier multiplication. If series resistances can be neglected, as is the case with transistors which have a rather large emitter area, the temperature dependence

$$I_B \sim T^{X_{TI}}\exp\left(-\frac{W_g - eV_{B'E'}}{k_B T}\right)$$

yields

$$\left(\frac{\partial I_B}{\partial T}\right)_{V_{BE}} = \left(\frac{\partial I_B}{\partial T}\right)_{V_{B'E'}} = \frac{1}{T}\frac{V_g - V_{B'E'} + X_{TI}V_T}{V_T} I_B\,.$$

From this result, the thermal resistance R_{th} can be calculated from the slope m_P as

$$R_{th} \approx T\frac{V_T}{V_g - V_{B'E'} + X_{TI}V_T} m_P\,.$$

In transistors with a small emitter area, a substantial emitter series resistance may occur; in this case a simultaneous extraction of both the thermal resistance and the emitter series resistance is necessary [144]. Alternatively, $(\partial I_B/\partial T)_{V_{BE}}$ can be determined from a series of measurements performed with different temperatures on the back surface of the wafer. In a generalization of the method, the temporal change of $\Delta i_B(t)$ after a v_{CE} voltage step is monitored [145] and used for the extraction of the thermal impedance between the device and the heat sink.

3.13 Parameter Extraction – AC Measurements

AC measurement methods are used for the determination of depletion capacitances and transit times. While depletion capacitances can be determined directly from two-terminal measurements with a vector voltmeter, two-port parameters such as admittance and hybrid parameters are hard to measure directly, as open- and short-circuit conditions are difficult to realize at frequencies in the megahertz and gigahertz regimes. S-parameter measurements [146], which allow a two-port to be terminated with a characteristic impedance (generally 50 Ω), are used instead; from these measurements, admittance or hybrid parameters are calculated (see Appendix B) and used for the extraction of f_T or f_{max}. Using special probes, it is possible to perform s-parameter measurements directly on a wafer. Under these conditions, no errors due to bond inductances or parasitic capacitances of the case appear. De-embedding procedures employing special test structures are necessary for discrimination of parasitics due to pad capacitances, etc.

3.13.1 De-Embedding

Figure 3.59a shows a typical pad configuration of a high-frequency bipolar-transistor test structure. The emitter and substrate pads are shorted and

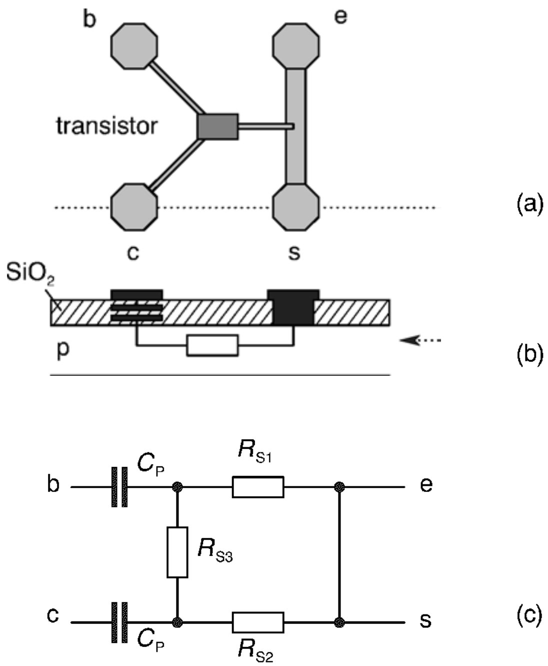


Fig. 3.59. (**a**) Layout, (**b**) cross section and (**c**) simplified equivalent circuit of pad configuration in typical bipolar-transistor test structure [113]

form an ohmic contact to the underlying p-type substrate. The base and collector pads are isolated from the substrate by the LOCOS oxide (see Fig. 3.59b). At high frequencies, however, capacitive coupling to the substrate becomes important. A simple equivalent circuit of the pad configuration and its coupling to the underlying substrate is given in Fig. 3.59c. For a proper determination of the device parameters, the effect of the parasitics has to be separated. For that purpose, additional measurements are performed using open and short test structures, as illustrated in Fig. 3.60.

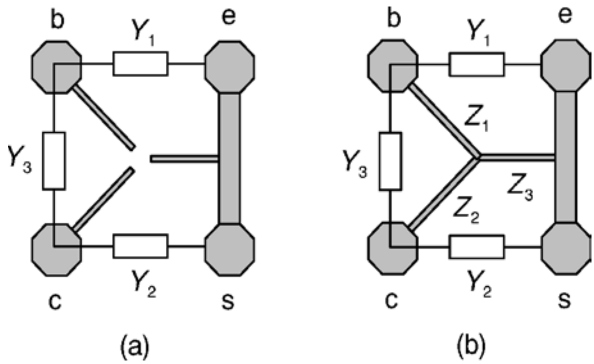


Fig. 3.60. (**a**) Open and (**b**) short test structures used in conventional de-embedding procedures

S-parameter measurements are performed on these test structures and y and z parameters are derived, which are used to subtract the effect of the parasitics [147–150]. These de-embedding procedures rely on simple relations obeyed by the admittance and impedance matrices[51] with parallel admittances and series impedances added as illustrated in Fig. 3.61.

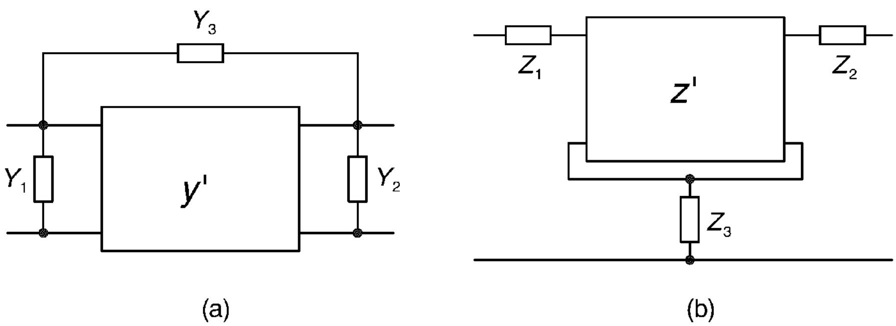


Fig. 3.61. (**a**) Two-port with parallel admittances and (**b**) two-port with series impedances

[51]The computation of admittance and impedance parameters from s-parameters is explained in Appendix B.

According to these relations, the admittance matrix of the two-port shown in Fig. 3.61a is

$$\mathsf{y} = \begin{pmatrix} y'_{11}+Y_1+Y_3 & y'_{12}-Y_3 \\ y'_{21}-Y_3 & y'_{22}+Y_2+Y_3 \end{pmatrix} ,$$

whereas the impedance matrix of the two-port shown in Fig. 3.61b is

$$\mathsf{z} = \begin{pmatrix} z'_{11}+Z_1+Z_3 & z'_{12}+Z_3 \\ z'_{21}+Z_3 & z'_{22}+Z_2+Z_3 \end{pmatrix} .$$

As can be seen from Fig. 3.60a, the admittance matrix of the open configuration is

$$\mathsf{y}^{\mathrm{open}} = \begin{pmatrix} Y_1+Y_3 & -Y_3 \\ -Y_3 & Y_2+Y_3 \end{pmatrix} ;$$

subtraction of the admittance matrix $\mathsf{y}^{\mathrm{open}}$ from the admittance matrix y therefore gives the admittance matrix of the device with series impedances Z_1, Z_2, Z_3. The effect of the series impedances can be removed with the help of the short test structure: transforming the difference between the admittance matrices of the short and open structures to z parameters yields the impedance matrix

$$\mathsf{z}_{\mathrm{ser}} = \begin{pmatrix} Z_1+Z_3 & Z_3 \\ Z_3 & Z_2+Z_3 \end{pmatrix} .$$

Transformation of the admittance matrix $\mathsf{y} - \mathsf{y}^{\mathrm{open}}$ to z parameters and subtraction of $\mathsf{z}_{\mathrm{ser}}$ thus yields the impedance parameters of the device under test.

3.13.2 Transit Time

From the s-parameters of the internal device, it is possible to determine the small-signal current gain $h_{21\mathrm{e}}(f)$ as a function of frequency for different values of the bias current I_{C}. The cutoff frequency $f_{\mathrm{T}} = \beta f_{\beta}$ for each value of the bias current is easily obtained from a Bode plot of $h_{21\mathrm{e}}(f)$.

The parameter τ_{f} is generally determined from a plot of $1/(2\pi f_{\mathrm{T}})$ versus $1/I_{\mathrm{C}}$. As long as the transit time can be assumed to be constant, this should result in a straight line with offset $\tau_{\mathrm{f}} + (r_{\mathrm{ee'}}+r_{\mathrm{cc'}})c_{\mathrm{jc}}$ according to (3.100), which can be rewritten as

$$\frac{1}{2\pi f_{\mathrm{T}}} = \frac{(c_{\mathrm{je}}+c_{\mathrm{jc}})V_{\mathrm{T}}}{I_{\mathrm{C}}} + \tau_{\mathrm{f}} + (r_{\mathrm{cc'}}+r_{\mathrm{cc'}})c_{\mathrm{jc}} . \tag{3.244}$$

The transit time can be assumed to be constant as long as high-level-injection effects are negligible (I_{C} small, and correspondingly $1/I_{\mathrm{C}}$ large): extrapolation

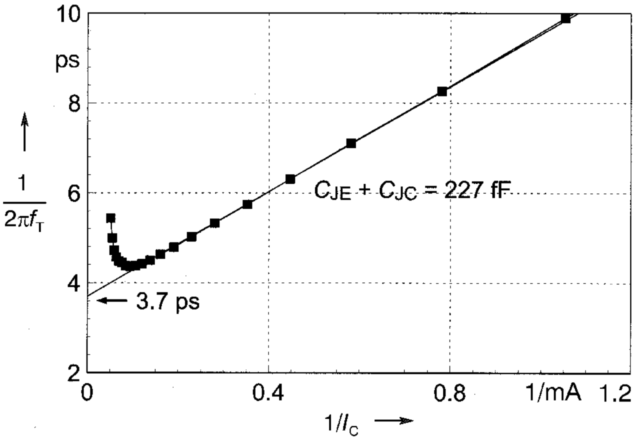


Fig. 3.62. Extraction of forward transit time τ_f (valid under low-level injection) from measurements of the cutoff frequency [113]

of the line to $1/I_C = 0$ and subtraction of $(r_{ee'}+r_{cc'})c_{jc}$ thus yields the forward transit time under low-level-injection conditions. This is demonstrated in Fig. 3.62; the increase of τ_f at large values of I_C due to high-level-injection effects is seen as a deviation from the straight-line behavior.

3.13.3 Capacitances

Depletion capacitances are measured by superposition of a small ac signal on a dc junction voltage; the measurement frequency is generally chosen to be below 1 MHz in order to avoid RC delays. For the characterization of high-frequency bipolar transistors with a very small emitter area, special test structures that employ a parallel connection of several transistors are used to determine a multiple of the original device capacitance. De-embedding of the test structure used has generally to be done.

Three parameters, C_{J0}, M_J and V_J, can be determined to give an optimal fit to the measured data by using a nonlinear least-squares optimization routine. If the program is allowed to choose parameters without any constraints, unphysical values of the diffusion voltage V_J may result. This has the disadvantage that a wrong temperature dependence of the depletion capacitance will be predicted. Therefore V_J is often fixed a priori, in order to give the correct temperature variation, and C_{J0} and M_J are then determined by a fitting procedure.

To estimate the capacitances from layout data, the junction capacitances are generally written as the sum of an intrinsic component that varies in proportion to the junction area $A_j = LW$ and of a peripheral component that varies in proportion to the perimeter length $P_j = 2(W+L)$:

$$c_j = c'_A A_j + c'_p P_j .$$

The parameters c'_{A} and c'_{p} are determined from transistors with different ratios $P_{\mathrm{j}}/A_{\mathrm{j}}$. Plotting measured values of $c_{\mathrm{j}}/A_{\mathrm{j}}$ versus $P_{\mathrm{j}}/A_{\mathrm{j}}$ should then yield a straight line with a y axis intercept given by c'_{A} and a slope determined by c'_{p}.

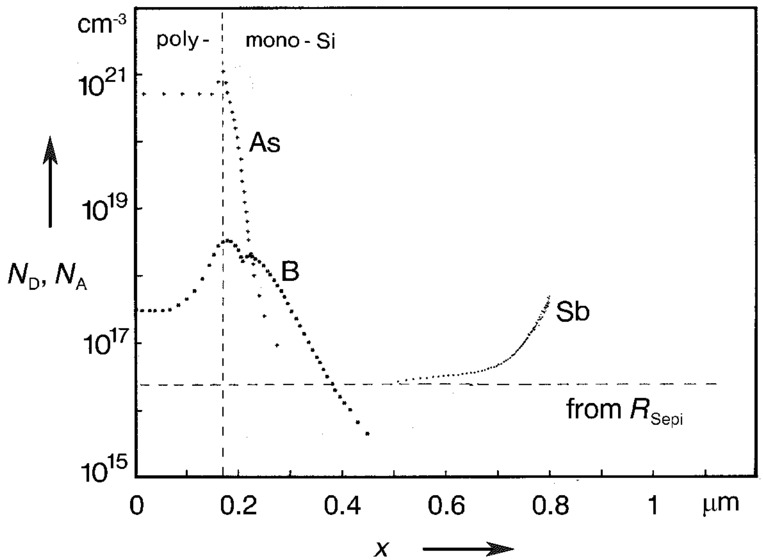


Fig. 3.63. Vertical doping profile of an npn bipolar transistor. The emitter and base dopant distributions were determined by SIMS, and capacitance and resistance measurements were used for the determination of the doping distribution in the epitaxial layer

Extraction of Doping Profiles. If the vertical doping profile of the base region and the area of the bc junction are known,[52] the vertical doping profile in the collector region may be extracted from the bias dependence of the bc capacitance. In the depletion-layer approximation, we may write,[53]

$$\begin{aligned} c_{\mathrm{jc}}(V_{\mathrm{CB}}) &= \frac{\epsilon A_{\mathrm{jc}}}{x_{\mathrm{cb}}(V_{\mathrm{CB}})-x_{\mathrm{bc}}(V_{\mathrm{CB}})} = eA_{\mathrm{jc}}N_{\mathrm{D}}[x_{\mathrm{cb}}(V_{\mathrm{CB}})]\,\frac{\mathrm{d}x_{\mathrm{cb}}}{\mathrm{d}V_{\mathrm{CB}}} \\ &= -eA_{\mathrm{jc}}N_{\mathrm{A}}[x_{\mathrm{bc}}(V_{\mathrm{CB}})]\,\frac{\mathrm{d}x_{\mathrm{bc}}}{\mathrm{d}V_{\mathrm{CB}}}\,, \end{aligned} \tag{3.245}$$

[52] Large-area transistors are to be preferred for this type of measurement; the vertical doping profile in the base region is generally determined by SIMS.

[53] As was pointed out in [151, 152], this yields the density of majority carriers, rather than the density of ionized donors on the collector side. Further errors associated with the depletion-layer approximation have been discussed in [153–155]. For doping densities in excess of 10^{16} cm^{-3}, as are typical for high-frequency bipolar transistors, however, the depletion-layer approximation is fairly good.

where x_{bc} and $x_{\mathrm{cb}} > x_{\mathrm{bc}}$ denote the edges of the depletion layer on the base and collector sides of the bc junction. Using this equation together with

$$\frac{\mathrm{d}c_{\mathrm{jc}}}{\mathrm{d}V_{\mathrm{CB}}} = \frac{c_{\mathrm{jc}}^2(V_{\mathrm{CB}})}{\epsilon A_{\mathrm{jc}}} \left(\frac{\mathrm{d}x_{\mathrm{bc}}}{\mathrm{d}V_{\mathrm{CB}}} - \frac{\mathrm{d}x_{\mathrm{cb}}}{\mathrm{d}V_{\mathrm{CB}}} \right) ,$$

we therefore obtain the following for the doping density in the collector region:

$$\frac{1}{N_{\mathrm{D}}\left[x_{\mathrm{cb}}(V_{\mathrm{CB}})\right]} = -\frac{1}{N_{\mathrm{A}}\left[x_{\mathrm{bc}}(V_{\mathrm{CB}})\right]} - \frac{e\epsilon A_{\mathrm{jc}}^2}{c_{\mathrm{jc}}^3(V_{\mathrm{CB}})} \frac{\mathrm{d}c_{\mathrm{jc}}}{\mathrm{d}V_{\mathrm{CB}}} .$$

An iterative solution, together with the requirement of charge neutrality

$$\int_{x_{\mathrm{bc}}}^{x_{\mathrm{jc}}} N_{\mathrm{A}}(x)\,\mathrm{d}x = \int_{x_{\mathrm{jc}}}^{x_{\mathrm{cb}}} N_{\mathrm{D}}(x)\,\mathrm{d}x ,$$

allows us to extract the vertical doping profile $N_{\mathrm{D}}(x)$. Figure 3.63 shows a result of this kind for a self-aligned vertical bipolar transistor [53].

3.13.4 The Impedance Semicircle Method

A widely used ac method for the characterization of series resistances is the impedance semicircle method [132, 156, 157], which evaluates the frequency-dependent input impedance in the common-emitter configuration. According to the Giacoletto model, the complex input impedance should follow a semicircle if plotted in the complex z plane (Fig. 3.64). The intersections of the semicircle[54] with the real axis should give the values of $r_\pi + r_{\mathrm{bb'}} + r_{\mathrm{ee'}}(\beta+1)$ for $\omega \to 0$, and $r_{\mathrm{bb'}} + r_{\mathrm{ee'}}$ for $\omega \to \infty$. With known values of β and r_π determined from dc characteristics, the values of $r_{\mathrm{bb'}}$ and $r_{\mathrm{ee'}}$ can be computed.

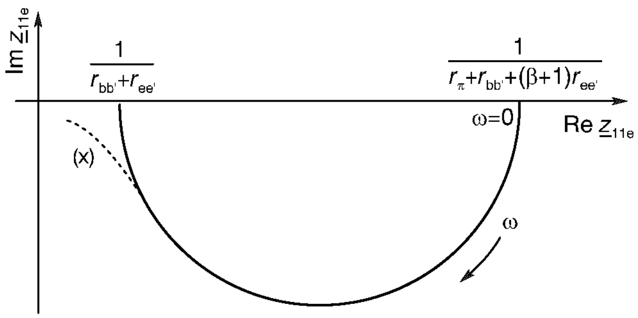


Fig. 3.64. Locus of the common-emitter input impedance $z_{11\mathrm{e}}$ as a function of frequency for the Giacoletto model (schematic representation); measured data (x) will deviate at high frequencies

[54]The same information could be obtained from a Bode plot of the input admittance $y_{11\mathrm{e}}$, which in the approximation (3.173), has the two limit values $g_{11\mathrm{e}} = 1/[r_\pi + r_{\mathrm{bb'}} + r_{\mathrm{ee'}}(\beta+1)]$ for $\omega \to 0$ and $1/(r_{\mathrm{bb'}} + r_{\mathrm{ee'}})$ for $\omega \to \infty$. A modification of the method has been presented in [158].

3.14 The VBIC Model

The VBIC model is a compact model for the modeling of integrated bipolar transistors that improves on the Gummel–Poon model in several details. It was developed by a broadly based committee [159] to serve as an industry standard that would replace the standard Gummel–Poon model and is therefore considered here, although there has not been very much support for the VBIC model recently.[55]

In comparison with the standard Gummel–Poon model, the following modifications and extensions are introduced:

1. The model allows one to consider self-heating of a bipolar transistor by simultaneous simulation of a thermal equivalent circuit, composed of a thermal resistance R_{th}, a thermal capacitance C_{th} and a current source $i_{\mathrm{th}}(t)$ that delivers a current proportional to the power dissipated in the device.
2. The model is extended to take account of the parasitic pnp transistor formed by the base, collector and substrate regions.
3. The Early effect is taken into account in a modified form, resulting in an improved expression for the output conductance of the bipolar transistor.
4. A modified formulation of the transistor quasi-saturation characteristics is implemented that avoids the negative output conductances obtained with the model of Kull et al. [89] under certain biasing conditions.
5. The model provides an improved description of depletion capacitances at large forward bias as an option, and includes additional capacitances to model overlap capacitances.
6. The model improves the description of ac and switching behavior by using a modified description of the charge stored in the epilayer under high-level-injection conditions, by a consistent treatment of the excess phase and by considering current-crowding effects with a distributed input circuit.
7. Weak avalanche effects are modeled with a simple semi-empirical exponential expression.

3.14.1 Vertical NPN Transistor

Figure 3.65 provides a graphical representation of the VBIC model. Except for the additional current source between nodes C′ and B′, which describes the effects of impact ionization in the bc diode, the equivalent circuit used for the description of the internal npn transistor has the same topology as in

[55] The Compact Modeling Council (CMC) recommends (see, for example, www.eigroup.org/cmc/minutes) the use of either MEXTRAM or HICUM; these are discussed in later sections.

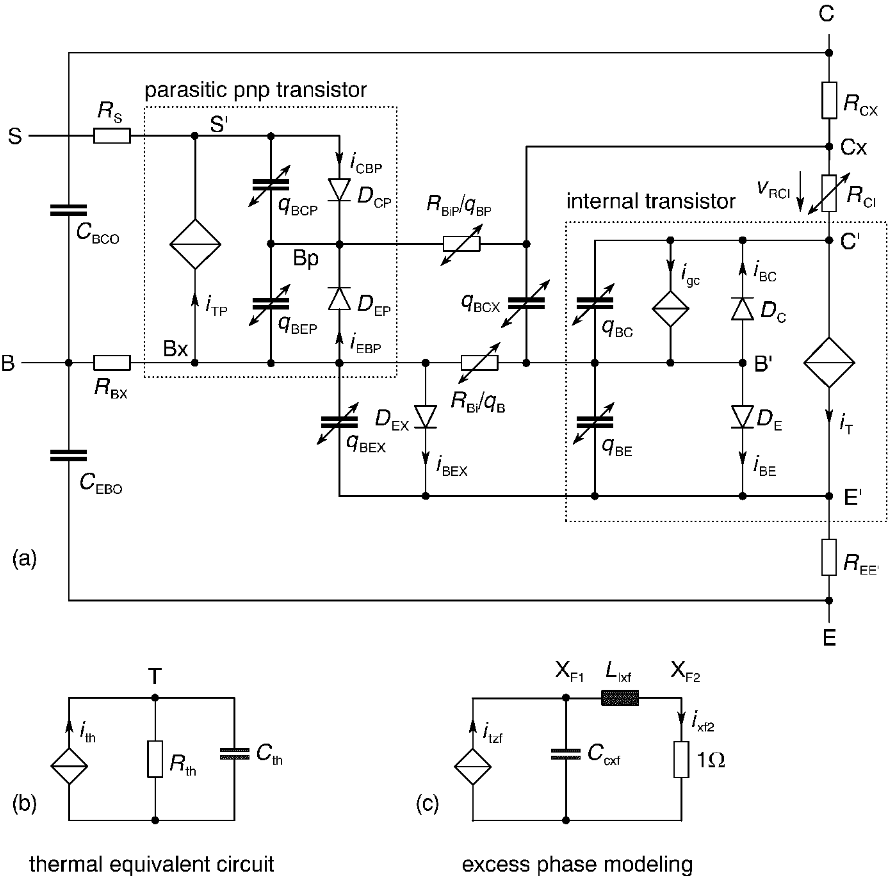


Fig. 3.65. Equivalent circuit of a bipolar transistor according to the VBIC model [159]

the standard Gummel–Poon model. However, somewhat modified modeling equations are employed.

Modeling of Transfer Current and Early Effect

The transfer current is written in a form similar to that of the Gummel–Poon model:

$$i_T = \frac{i_{CE} - i_{EC}}{q_B} \, ,$$

where,[56]

[56]For compatibility with the Gummel–Poon model implemented in SPICE, the model allows separate nonideality factors N_F and N_R, although $N_F = N_R$ is recommended, as otherwise the model can become nonpassive.

$$i_{\mathrm{CE}} = I_{\mathrm{S}}\left[\exp\left(\frac{v_{\mathrm{B'E'}}}{N_{\mathrm{F}}V_{\mathrm{T}}}\right) - 1\right], \qquad i_{\mathrm{EC}} = I_{\mathrm{S}}\left[\exp\left(\frac{v_{\mathrm{B'C'}}}{N_{\mathrm{R}}V_{\mathrm{T}}}\right) - 1\right]$$

and q_{B} is a normalized base charge that is written as

$$q_{\mathrm{B}} = \frac{1}{2}\left[q_1 + \sqrt{q_1^2 + 4\left(\frac{i_{\mathrm{CE}}}{I_{\mathrm{KF}}} + \frac{i_{\mathrm{EC}}}{I_{\mathrm{KR}}}\right)}\right]. \tag{3.246}$$

The normalized base charge under low-level-injection conditions q_1, however, is modeled as

$$q_1 = 1 + q_{\mathrm{je}} + q_{\mathrm{jc}},$$

where the contributions of the eb and bc depletion charges to q_1 are expressed in terms of the respective depletion capacitances, written as

$$q_{\mathrm{je}} = \frac{1}{Q_{\mathrm{B0}}}\int_0^{V_{\mathrm{B'E'}}} c_{\mathrm{je}}(V)\,\mathrm{d}V = \frac{1}{V_{\mathrm{ER}}}\frac{1}{C_{\mathrm{JE}}}\int_0^{V_{\mathrm{B'E'}}} c_{\mathrm{je}}(V)\,\mathrm{d}V,$$

$$q_{\mathrm{jc}} = \frac{1}{Q_{\mathrm{B0}}}\int_0^{V_{\mathrm{B'C'}}} c_{\mathrm{jc}}(V)\,\mathrm{d}V = \frac{1}{V_{\mathrm{EF}}}\frac{1}{C_{\mathrm{JC}}}\int_0^{V_{\mathrm{B'C'}}} c_{\mathrm{jc}}(V)\,\mathrm{d}V.$$

This approach introduces the two parameters V_{EF} and V_{ER}, which are related to the forward and reverse Early voltages V_{AF} and V_{AR} of the Gummel–Poon model.[57] According to the standard Gummel–Poon description[58], one of the parameters C_{JE}, C_{JC}, V_{EF} and V_{ER} is redundant, since the condition

$$Q_{\mathrm{B0}} = V_{\mathrm{ER}}C_{\mathrm{JE}} = V_{\mathrm{EF}}C_{\mathrm{JC}}$$

should be fulfilled. The utilization of four different parameters, however, allows one to fit experimental data better, particularly in modeling heterojunction bipolar transistors.

Excess Phase. The partial network of Fig. 3.65c is used for the modeling of excess phase both in ac and in transient analyses. The controlled current source provides a current i_{tzf} equal to i_{CE}; this current causes a voltage drop v_{xf2} across the ohmic resistance, which is defined as $1\,\Omega$. The transfer factor

[57] If this approach is compared with the standard Gummel–Poon model, where $q_{\mathrm{je}} \approx V_{\mathrm{B'E'}}/V_{\mathrm{AR}}$ and $q_{\mathrm{jc}} \approx V_{\mathrm{B'C'}}/V_{\mathrm{AF}}$ as long as $q_{\mathrm{je}}, q_{\mathrm{jc}} \ll 1$, the estimates

$$V_{\mathrm{ER}} \approx \frac{V_{\mathrm{AR}}}{C_{\mathrm{JE}}V_{\mathrm{B'E'}}}\int_0^{V_{\mathrm{B'E'}}} c_{\mathrm{je}}(V)\,\mathrm{d}V \quad \text{and} \quad V_{\mathrm{EF}} \approx \frac{V_{\mathrm{AF}}}{C_{\mathrm{JC}}V_{\mathrm{B'C'}}}\int_0^{V_{\mathrm{B'C'}}} c_{\mathrm{jc}}(V)\,\mathrm{d}V$$

are obtained for the new parameters V_{ER} and V_{EF}.

[58] As is obvious from the analysis of Sect. 3.2, the transfer current is affected by the bias dependence of the Gummel number, which is a weighted average of the base charge. Its value varies only in proportion to the base charge if the mobilities and intrinsic carrier densities do not show substantial variations with position. This assumption, however, is not valid in graded-base HBTs (see Chap. 4).

$$\frac{\underline{i}_{\mathrm{xf2}}}{\underline{i}_{\mathrm{tzf}}} = \frac{1}{1+\mathrm{j}\omega RC-\omega^2 LC} \tag{3.247}$$

equals the approximation to the phase shift factor $\underline{\phi}(\mathrm{j}\omega)$ considered in Sect. 3.9 if $C = T_{\mathrm{D}}/\Omega$ and $L = (T_{\mathrm{D}}/3)\Omega$, where $T_{\mathrm{D}} = 1/\omega_0$ replaces the SPICE parameter $P_{\mathrm{TF}} = T_{\mathrm{D}}/\tau_{\mathrm{f}}$. Using the current i_{xf2} that flows through R instead of $i_{\mathrm{CE}}/q_{\mathrm{B}} = i_{\mathrm{tzf}}$ for the description of the transfer current introduces an appropriate delay in the voltage response of the transfer current. This formulation works both in ac and in transient analyses and therefore provides a consistent approach for both types of analysis.

Modeling of Internal EB and BC Diodes

The description of the internal eb and bc diodes is similar to that in the Gummel–Poon model, but is more flexible owing to the introduction of additional emission coefficients. The recombination current of the eb diode is written as the sum of an internal component i_{BE} and an external component i_{BEX}:

$$i_{\mathrm{BE}} + i_{\mathrm{BEX}} = I_{\mathrm{BEI}} \left[\exp\left(\frac{v_{\mathrm{B'E'}}}{N_{\mathrm{EI}} V_{\mathrm{T}}} \right) - 1 \right] + I_{\mathrm{BEN}} \left[\exp\left(\frac{v_{\mathrm{B'E'}}}{N_{\mathrm{EN}} V_{\mathrm{T}}} \right) - 1 \right] .$$

This introduces an additional parameter N_{EI}, which can be chosen to be independent of the emission coefficient N_{F} introduced in modeling the transfer current, resulting in more flexibility in modeling the bias- and temperature-dependent current gain: the SPICE counterparts of I_{BEI}, I_{BEN} and N_{EN} are given by $I_{\mathrm{S}}/B_{\mathrm{F}}$, I_{SE} and N_{E}, whereas N_{EI} is taken to be equal to N_{F} in SPICE. The parameter W_{BE} is used to define the fraction of the base current that flows in the internal diode D_{E}, i.e.

$$\frac{i_{\mathrm{BE}}}{W_{\mathrm{BE}}} = I_{\mathrm{BEI}} \left[\exp\left(\frac{v_{\mathrm{B'E'}}}{N_{\mathrm{EI}} V_{\mathrm{T}}} \right) - 1 \right] + I_{\mathrm{BEN}} \left[\exp\left(\frac{v_{\mathrm{B'E'}}}{N_{\mathrm{EN}} V_{\mathrm{T}}} \right) - 1 \right] .$$

The recombination current of the (internal) bc diode is written as

$$i_{\mathrm{BC}} = I_{\mathrm{BCI}} \left[\exp\left(\frac{v_{\mathrm{B'C'}}}{N_{\mathrm{CI}} V_{\mathrm{T}}} \right) - 1 \right] + I_{\mathrm{BCN}} \left[\exp\left(\frac{v_{\mathrm{B'C'}}}{N_{\mathrm{CN}} V_{\mathrm{T}}} \right) - 1 \right] ,$$

which also extends the conventional Gummel–Poon model by the introduction of the extra emission coefficient N_{EI}, which may be chosen to be independent of the emission coefficient N_{R} introduced in the formula for the reverse transfer current. The parameters I_{BCI}, I_{BCN} and N_{CN} correspond to the SPICE parameters $I_{\mathrm{S}}/B_{\mathrm{R}}$, I_{SC} and N_{C}, whereas N_{CI} is taken to be equal to N_{R} in SPICE. The external portion of the bc recombination current is modeled as the base diode of the parasitic pnp transistor.

Impact ionization in the reverse-biased bc diode is taken into account by means of the controlled current source i_{gc}. The current i_{gc} is modeled using the formula of Klosterman and de Graaff [50],

$$i_{\mathrm{gc}} = (i_{\mathrm{T}} - i_{\mathrm{BC}}) A_{\mathrm{VC1}} (P_{\mathrm{C}} - v_{\mathrm{B'C'}}) \exp\Big[-A_{\mathrm{VC2}} (P_{\mathrm{C}} - v_{\mathrm{B'C'}})^{M_{\mathrm{C}}-1}\Big] ,$$

which introduces two additional parameters, A_{VC1} and A_{VC2}; the parameters P_{C} and M_{C} are the built-in voltage and the grading exponent used for modeling of the bc depletion capacitance.

Modeling of Resistances and Quasi-Saturation

The emitter and substrate resistances, as well as the extrinsic portions of the base and collector resistances, are modeled as ohmic resistors. The internal base resistance is modeled as bias-dependent, modulated by the normalized base charge. This approach is justified to first order, as the base partitioning by W_{BE} already takes account of emitter current crowding to a certain extent.

Quasi-saturation effects are described using a modification of the Kull approach [89] outlined in Sect. 3.7 in order to avoid problems with a negative value of the output conductance at high V_{BE}. The modified description makes use of the epi current i_{epi0} given by [89] (see Sect. 3.7) if velocity saturation is ignored:

$$i_{\mathrm{epi0}} = \frac{v_{\mathrm{RCI}} + V_{\mathrm{T}} \left\{ K(v_{\mathrm{B'C'}}) - K(v_{\mathrm{B'Cx}}) - \ln\left[\dfrac{1+K(v_{\mathrm{B'C'}})}{1+K(v_{\mathrm{B'Cx}})}\right]\right\}}{R_{\mathrm{CI}}} ,$$

where R_{CI} corresponds to the parameter R_{C0} of the original model and $K(v)$ is defined as in Sect. 3.7. The effects of velocity saturation are introduced in a different manner, on the basis of an alternative velocity saturation model,

$$\mu_{\mathrm{n}} = \frac{\mu_{\mathrm{n0}}}{\sqrt{1 + (\mu_{\mathrm{n0}} |\nabla \phi_{\mathrm{n}}| / v_{\mathrm{nsat}})^2}} ,$$

which allows one to avoid problems arising from discontinuities in higher-order derivatives. The current i_{epi} across R_{CI} is now written as[59]

$$i_{\mathrm{epi}} = \frac{i_{\mathrm{epi0}}}{\sqrt{1 + \left(\dfrac{R_{\mathrm{CI}} i_{\mathrm{epi0}}}{V_{\mathrm{O}} + 0.5\sqrt{0.01 + V_{\mathrm{rci}}^2} / H_{\mathrm{RCF}}}\right)^2}} , \tag{3.248}$$

where H_{RCF} is a dimensionless fitting parameter that allows one to partially replace V_{O} by V_{RCI}.

[59] The formula reproduced in [159] is obviously erroneous as it does not give the correct dimensions for the current.

3.14.2 The Parasitic PNP Transistor

Integrated npn bipolar transistors are generally realized on a p-type substrate. This results in parasitic vertical pnp transistors formed by the bc and the reverse-biased cs junctions. These transistors are modeled by a simplified Gummel–Poon equivalent circuit. Assuming the cs junction to be reverse biased, the pnp transfer current has only a forward component $i_{\mathrm{TFP}}/q_{\mathrm{Bp}}$; the model equation

$$I_{\mathrm{TFP}} = I_{\mathrm{SP}} \left[W_{\mathrm{SP}} \exp\left(\frac{v_{\mathrm{BxBp}}}{N_{\mathrm{FP}} V_{\mathrm{T}}} \right) + (1 - W_{\mathrm{SP}}) \exp\left(\frac{V_{\mathrm{B'C'}}}{N_{\mathrm{FP}} V_{\mathrm{T}}} \right) - 1 \right]$$

takes account of the fact that the parasitic pnp transistor is distributed between intrinsic (under the emitter) and extrinsic (not under the emitter) components. The Early effect of the parasitic pnp transistor is neglected, i.e. the normalized base charge q_{Bp} takes account of only high-level injection effects:

$$q_{\mathrm{Bp}} = 1 + \frac{q_{2\mathrm{p}}}{q_{\mathrm{Bp}}} \,, \quad \text{where} \quad q_{2\mathrm{p}} = \frac{I_{\mathrm{TFP}}}{I_{\mathrm{KP}}} \,.$$

The base current components of the parasitic pnp transistor are modeled, in complete analogy with the Gummel–Poon model, with saturation currents and emission coefficients I_{BEIP}, I_{BENP}, N_{CI} and N_{CN} for the eb current,

$$i_{\mathrm{EBP}} = I_{\mathrm{BEIP}} \left[\exp\left(\frac{v_{\mathrm{BxBp}}}{N_{\mathrm{CI}} V_{\mathrm{T}}} \right) - 1 \right] + I_{\mathrm{BENP}} \left[\exp\left(\frac{v_{\mathrm{BxBp}}}{N_{\mathrm{CN}} V_{\mathrm{T}}} \right) - 1 \right] \,,$$

and saturation currents and emission coefficients I_{BCIP}, I_{BCNP}, N_{CIP} and N_{CNP} for the collector–base current,

$$i_{\mathrm{CBP}} = I_{\mathrm{BCIP}} \left[\exp\left(\frac{v_{\mathrm{S'Bp}}}{N_{\mathrm{CIP}} V_{\mathrm{T}}} \right) - 1 \right] + I_{\mathrm{BCNP}} \left[\exp\left(\frac{v_{\mathrm{S'Bp}}}{N_{\mathrm{CNP}} V_{\mathrm{T}}} \right) - 1 \right] \,.$$

3.14.3 Stored Charges

In addition to depletion and diffusion capacitances, the model defines two bias-independent overlap capacitances C_{BEO} and C_{BCO}; taking account of these requires the introduction of extra elements in the Gummel–Poon model.

In the default option, the depletion capacitances are modeled in the same manner as in the Gummel–Poon model, although a somewhat different notation is used, as indicated in Table 3.4; a single-piece, smooth model option and a reach-through model option are also provided. If the reach-through option is chosen, the internal and external bc depletion capacitances c_{jc} and c_{jep} approach a constant value if the reverse bias exceeds the critical value V_{RT}.

Table 3.4. VBIC parameters for depletion capacitances

Capacitance	c_{je}	c_{jex}	c_{jc}	c_{jep}	c_{js}
Zero-bias capacitance c_{J0}	WBE×CJE	(1-WBE)×CJE	CJC	CJEP	CJCP
Built-in voltage V_{J}	PE	PE	PC	PC	PS
Grading coefficient M	ME	ME	MC	MC	MS
Voltage v'	$v_{\mathrm{B'E'}}$	$v_{\mathrm{BxE'}}$	$v_{\mathrm{B'C'}}$	v_{BxBp}	$v_{\mathrm{S'Bp}}$

The charge q_{BE} associated with the internal eb diode is described in terms of a depletion charge q_{je} and a diffusion charge

$$q_{\mathrm{BE}} = \int_0^{v_{\mathrm{B'E'}}} c_{\mathrm{je}}(v)\,\mathrm{d}v + \tau_{\mathrm{f}} i_{\mathrm{CE}}/q_{\mathrm{B}} ,$$

where the forward transit time is modeled in analogy to SPICE using the parameters T_{F}, X_{TF}, I_{TF} and V_{TF} already known from the SPICE model. The charge q_{BEx} stored in the eb sidewall capacitance is modeled as a depletion charge $q_{\mathrm{BEx}} = q_{\mathrm{JEx}}$. The charge associated with the internal bc diode is written as

$$q_{\mathrm{BC}} = \int_0^{v_{\mathrm{B'C'}}} c_{\mathrm{jc}}(v)\,\mathrm{d}v + T_{\mathrm{R}} i_{\mathrm{EC}} + Q_{\mathrm{C0}}\sqrt{1+\gamma\exp(v_{\mathrm{B'C'}}/V_{\mathrm{T}})} ,$$

where the reverse transit time T_{R} is already known from SPICE. The third contribution describes charge that is stored in the epilayer according to the Kull model [89], this term is considered only if $R_{\mathrm{CI}} > 0$. Also, the charge q_{BCx} is neglected if $R_{\mathrm{CI}} = 0$; otherwise it is described by

$$q_{\mathrm{BCx}} = Q_{\mathrm{C0}}\sqrt{1+\gamma\exp(v_{\mathrm{B'Cx}}/V_{\mathrm{T}})} .$$

The charge q_{BEp} associated with the eb diode of the parasitic pnp transistor is modeled as

$$q_{\mathrm{BEp}} = \int_0^{v_{\mathrm{BxBp}}} c_{\mathrm{jep}}(v)\,\mathrm{d}v + T_{\mathrm{R}} i_{\mathrm{TFP}} ,$$

whereas the charge q_{BCp} associated with the reverse-biased cs diode is treated as a pure depletion charge,

$$q_{\mathrm{BEp}} = \int_0^{v_{\mathrm{S'Bp}}} c_{\mathrm{js}}(v)\,\mathrm{d}v .$$

3.14.4 Temperature Effects

Saturation currents, current gains, series resistances and depletion capacitances are modeled as temperature-dependent quantities. However, the model equations differ somewhat from those of the standard Gummel–Poon model. The temperature dependence of the series resistances is modeled as follows:

$$R_{\mathrm{X}}(T) = R_{\mathrm{X}}(T_0)\left(\frac{T}{T_0}\right)^{X_{\mathrm{RX}}} , \tag{3.249}$$

with different parameters X_{RX} for the emitter, base, collector and substrate regions. Use of this formulation is motivated by the temperature dependence of the majority-carrier mobilities (see Sect. 2.8). Zero-bias depletion capacitances are modeled in the form

$$C_{\mathrm{X}}(T) = \left(\frac{P_{\mathrm{X}}(T)}{P_{\mathrm{X}}(T_0)}\right)^{M_{\mathrm{X}}} , \tag{3.250}$$

where M_{X} denotes the grading exponent of the corresponding junction. The built-in voltages P_{X} are modeled by a temperature-dependent function derived from

$$P_{\mathrm{X}} = V_{\mathrm{T}} \ln\left(n_{\mathrm{n0}} p_{\mathrm{p0}} / n_{\mathrm{ie}}^2\right) , \tag{3.251}$$

avoiding the approximation $n_{\mathrm{n0}} = N_{\mathrm{D}}$ and $p_{\mathrm{p0}} = N_{\mathrm{A}}$, which may lead to negative values of the built-in voltage at high temperature.

In analogy to the SPICE approach, the general expression

$$I_{\mathrm{X}}(T) = I_{\mathrm{X}}(T_0) \left(\frac{T}{T_0}\right)^{X_{\mathrm{X}}/N_{\mathrm{X}}} \exp\left[-\frac{E_{\mathrm{X}}}{N_{\mathrm{X}} V_{\mathrm{T}}} \left(1 - \frac{T}{T_0}\right)\right] ,$$

with the parameters listed in Table 3.5, is used for the description of temperature-dependent saturation currents; the (nonlinear) temperature dependence of the bandgap is neglected in this approach.

Table 3.5. VBIC parameters for temperature-dependent saturation currents

Circuit element	I_{X}	N_{X}	X_{X}	E_{X}
NPN transfer current source	I_{S}	N_{F}	X_{IS}	E_{A}
EB diode (internal)	I_{BEI}	N_{EI}	X_{II}	E_{AIE}
EB leakage diode (internal)	I_{BEN}	N_{EN}	X_{IN}	E_{ANE}
BC diode (internal)	I_{BCI}	N_{CI}	X_{II}	E_{AIC}
BC leakage diode (internal)	I_{BCN}	N_{CN}	X_{IN}	E_{ANC}
PNP transfer current source	I_{SP}	N_{FP}	X_{II}	E_{AIC}
EB diode (pnp)	I_{BEIP}	N_{CI}	X_{II}	E_{AIC}
EB leakage diode (pnp)	I_{BENP}	N_{CN}	X_{IN}	E_{ANC}
BC diode (pnp)	I_{BCIP}	N_{CIP}	X_{II}	E_{AIS}
BC leakage diode (pnp)	I_{BCNP}	N_{CNP}	X_{IN}	E_{ANS}

No additional parameters are introduced to describe the temperature variation of the current gain, which is determined by the temperature dependences of the saturation currents involved.

The excess temperature due to the power dissipated in the device is calculated from the thermal equivalent circuit Fig. 3.65b. The potential of node T of this equivalent circuit is proportional to the excess temperature ΔT due to self-heating, as outlined in Sect. 3.11.2. By use of the result of this calculation, all temperature-dependent parameters can be updated to the actual device temperature.

3.15 The HICUM Model

The model HICUM,[60] developed at the university of Bochum [160, 161], is a compact physical transistor model that was originally designed to improve on the standard Gummel–Poon model in forward operation or weak saturation (with an internal collector–emitter voltage drop in excess of 250 mV). The model was subsequently extended to take account of the parasitic substrate transistor, saturation and quasi-saturation, avalanche and tunneling currents, and self-heating effects, resulting in the equivalent circuit shown in Fig. 3.66 [162]. In its present form, the model employs as many as 97 transistor-specific model parameters, listed in Table 3.7 – with a substantial number of parameters derived from the separation of the transistor into an internal and an external portion, however. The model has been designed for geometric scalability, and a separate program called TRADICA [163, 164] is available to generate HICUM model parameters from layout data.

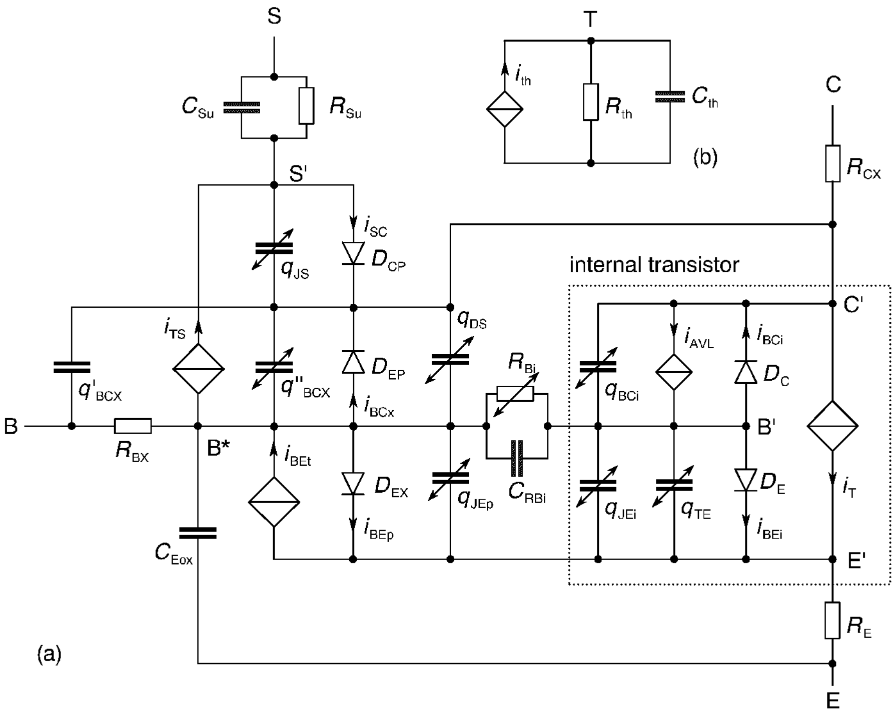


Fig. 3.66. (**a**) HICUM large-signal equivalent circuit, including (**b**) self-heating network, after [165]

[60]The acronym HICUM is an abbreviation of "high current model"; documentation can be found on the website of the Compact Modeling Council at http://www.eigroup.org/cmc

Table 3.6. Parameters of HICUM transistor model (level 2)

Parameter name	Parameter	Unit	Default
Transfer current			
Transfer saturation current	I_{S}, IS	A	10^{-16}
Zero-bias base charge	Q_{P0}, QP0	C	2×10^{-14}
High-current correction	I_{CH}, ICH	A	∞
Emitter diffusion charge weighting factor	H_{FE}, HFE	–	1
Collector diffusion charge weighting factor	H_{FC}, HFC	–	1
EB depletion charge weighting factor	H_{JEI}, HJEI	–	1
BC depletion charge weighting factor	H_{JCI}, HJCI	–	1
BE current			
Saturation current of internal diode	I_{BEIS}, IBEIS	A	10^{-18}
Emission coefficient of internal diode	M_{BEI}, MBEI	–	1
Saturation current of internal leak diode	I_{REIS}, IREIS	A	0
Emission coefficient of internal leak diode	M_{REI}, MREI	–	2
Saturation current of sidewall diode	I_{BEPS}, IBEPS	A	0
Emission coefficient of sidewall diode	M_{BEP}, MBEP	–	1
Saturation current of sidewall leak diode	I_{REPS}, IREPS	A	0
Emission coefficient of sidewall leak diode	M_{REP}, MREP	–	2
BC current			
Saturation current of internal diode	I_{BCIS}, IBCIS	A	10^{-16}
Emission coefficient of internal diode	M_{BCI}, MBCI	–	1
Saturation current of external diode	I_{BCXS}, IBCXS	A	0
Emission coefficient of external diode	M_{BCX}, MBCX	–	1
BE tunneling current			
Tunneling saturation current	I_{BETS}, IBETS	A	0
Tunneling-current exponent	A_{BET}, ABET	–	40
BC avalanche current			
Avalanche current factor	F_{AVL}, FAVL	1/V	0
Avalanche current exponent	Q_{AVL}, QAVL	C	0
Series resistances			
Zero-bias internal base resistance	R_{BI0}, RBI0	Ω	0
External base resistance	R_{BX}, RBX	Ω	0
Emitter current-crowding geometry factor	F_{GEO}, FGEO	–	0.6557
Correction for base sheet resistance	F_{DQR0}, FDQR0	–	0
HF shunt capacitance ratio	F_{RCBI}, FRCBI	–	0
Internal minority-charge ratio	F_{QI}, FQI	–	1
Emitter resistance	R_{E}, RE	Ω	0
External collector resistance	R_{CX}, RCX	Ω	0
Substrate transistor			
pnp transfer saturation current	I_{TSS}, ITSS	A	0
pnp forward transfer emission coefficient	M_{SF}, MSF	–	1
pnp reverse transfer emission coefficient	M_{SR}, MSR	–	1
Saturation current of cs diode	I_{SCS}, ISCS	A	0
Emission coefficient of cs diode	M_{SC}, MSC	–	1
pnp forward transit time	τ_{SF}, TSF	s	0

Table 3.6 (continued)

Parameter name	Parameter	Unit	Default
Substrate coupling			
Substrate series resistance	R_{SU}, RSU	Ω	0
Substrate shunt capacitance	C_{SU}, CSU	F	0
Depletion capacitances			
EB zero-bias depletion capacitance	C_{JEI0}, CJEI0	F	0
EB built-in voltage	V_{DEI}, VDEI	V	0.9
EB grading exponent	Z_{EI}, ZEI	–	0.5
Maximum eb capacitance increase	A_{LJEI}, ALJEI	–	2.5
EB zero-bias sidewall depletion capacitance	C_{JEP0}, CJEP0	F	0
EB sidewall built-in voltage	V_{DEP}, VDEP	V	0.9
EB sidewall grading exponent	Z_{EP}, ZEP	–	0.5
Maximum eb sidewall capacitance increase	A_{LJEP}, ALJEP	–	2.5
CB zero-bias depletion capacitance (internal)	C_{JCI0}, CJCI0	F	0
CB built-in voltage (internal)	V_{DCI}, VDCI	V	0.7
CB grading exponent (internal)	Z_{CI}, ZCI	–	0.4
CB punchthrough voltage (internal)	V_{PTCI}, VPTCI	V	∞
CB zero-bias depletion capacitance (external)	C_{JCX0}, CJCX0	F	0
CB built-in voltage (external)	V_{DCX}, VDCX	V	0.7
CB grading exponent (external)	Z_{CX}, ZCX	–	0.4
CB punchthrough voltage (external)	V_{PTCX}, VPTCX	V	∞
CB partitioning factor	F_{BC}, FBC	–	0
CS zero-bias depletion capacitance	C_{JS0}, CJS0	F	0
CS built-in voltage	V_{DS}, VDS	V	0.6
CS grading exponent	Z_{S}, ZS	–	0.5
CS punchthrough voltage	V_{PTS}, VPTS	V	∞
Diffusion capacitances			
Low-current forward transit time	τ_0, T0	s	0
Parameter for space charge layer effect	$\Delta\tau_{\mathrm{0H}}$, DT0H	s	0
Parameter for velocity saturation effect	τ_{BVL}, TBVL	s	0
Neutral emitter storage time	τ_{EF0}, TEF0	s	0
Parameter for τ_{E} current dependence	G_{TFE}, GTFE	–	1
Saturation time constant	τ_{HCS}, THCS	s	0
Transit time smoothing factor	A_{LHC}, ALHC	–	0.1
Transit time partitioning factor	F_{THC}, FTHC	–	0
Epilayer resistance (low field)	R_{CI0}, RCI0	Ω	150
Velocity saturation voltage	V_{LIM}, VLIM	V	0.5
Internal CE saturation voltage	V_{CES}, VCES	V	0.1
Collector punchthrough voltage	V_{PT}, VPT	V	∞
Reverse storage time	T_{R}, TR	s	0
Isolation capacitances			
EB isolation capacitance	C_{EOX}, CEOX	F	0
CB isolation capacitance	C_{COX}, CCOX	F	0

Table 3.6 (continued)

Parameter name	Parameter	Unit	Default
Non-quasi-static effects			
Minority-charge delay factor	A_{LQF}, ALQF	–	0
Transfer current delay factor	A_{LIT}, ALIT	–	0
Noise			
$1/f$ noise coefficient	K_{F}, KF	–	0
$1/f$ noise exponent	A_{F}, AF	–	2
Internal base resistance factor	K_{RBI}, KRBI	–	1
Collector charge scaling factors			
Emitter width dependence	L_{ATB}, LATB	–	0
Emitter length dependence	L_{ATL}, LATL	–	0
Temperature dependences			
Bandgap voltage extrapolated to 0 K	V_{GB}, VGB	V	1.17
TC[a] of forward current gain	A_{LB}, ALB	1/K	5×10^{-3}
TC (linear) of parameter T0	A_{LT0}, ALT0	1/K	0
TC (quadratic) of parameter T0	K_{T0}, KT0	1/K^2	0
T exponent for RCI0	ζ_{CI}, ZETACI	–	0
TC of saturation drift velocity	A_{LVS}, ALVS	1/K	0
TC of VCES	A_{LCES}, ALCES	1/K	0
T exponent of internal base resistance	ζ_{RBI}, ZETARBI	–	0
T exponent of external base resistance	ζ_{RBX}, ZETARBX	–	0
T exponent of external collector resistance	ζ_{RCX}, ZETARCX	–	0
T exponent of emitter resistance	ζ_{RE}, ZETARE	–	0
TC of FAVL	A_{LFAV}, ALFAV	1/K	0
TC of QAVL	A_{LQAV}, ALQAV	1/K	0
Self-heating			
Thermal resistance	R_{th}, RTH	K/W	0
Thermal capacitance	C_{th}, CTH	J/K	0

[a] Temperature coefficient.

In contrast to the VBIC and MEXTRAM models, only one internal collector node, C′, is used: effects associated with the epitaxial collector region are modeled in an improved description of the normalized base charge q_{B} and transit time τ_{f}. Except for the circuit elements introduced to describe the parasitic effects of the external transistor and the parasitic pnp transistor, which are modeled in a standard way, it is therefore possible to consider HICUM as a Gummel–Poon model with a substantially improved description of high-current effects and with modifications introduced to allow the description of heterojunction bipolar transistors. The underlying ideas are explained in the following subsection.

3.15.1 Modeling Approach

The model is based on Gummel's integral charge control relation, according to which the transfer current in forward operation varies in inverse proportion to the hole charge $Q'_{\mathrm{p}}(V_{\mathrm{B'E'}}, V_{\mathrm{C'E'}})$ stored in the internal transistor volume. In order to accurately predict the bias-dependent value of Q'_{p}, the relation

$$\frac{1}{2\pi f'_{\mathrm{T}}} = \left.\frac{\mathrm{d}Q'_{\mathrm{p}}}{\mathrm{d}I_{\mathrm{C}}}\right|_{V_{\mathrm{CE}}} \tag{3.252}$$

between the cutoff frequency f'_{T} of the internal transistor and Q'_{p} is used [166],[61] leading to

$$Q'_{\mathrm{p}}(I_{\mathrm{C}}, V_{\mathrm{CE}}) = Q'_{\mathrm{p}}(0, V_{\mathrm{CE}}) + \int_0^{I_{\mathrm{C}}} \frac{1}{2\pi f'_{\mathrm{T}}}\,\mathrm{d}I_{\mathrm{C}} .$$

Since $V_{\mathrm{B'C'}} = -V_{\mathrm{CE}}$ if $V_{\mathrm{B'E'}} = 0$, as is the case for $I_{\mathrm{C}} \approx 0$, we may write

$$Q'_{\mathrm{p}}(0, V_{\mathrm{CE}}) = Q_{\mathrm{P0}} + \int_0^{-V_{\mathrm{CE}}} c_{\mathrm{jc}}(v)\,\mathrm{d}v ,$$

where $Q'_{\mathrm{p}}(0, 0)$ has been identified with the zero-bias base charge $Q_{\mathrm{B0}} = Q_{\mathrm{P0}}$ of the Gummel–Poon model. Since, according to Sect. 3.5,[62]

$$\frac{1}{2\pi f'_{\mathrm{T}}} = \tau_{\mathrm{f}} + c_{\mathrm{je}}\frac{\mathrm{d}V_{\mathrm{B'E'}}}{\mathrm{d}I_{\mathrm{C}}} + c_{\mathrm{jc}}\frac{\mathrm{d}V_{\mathrm{B'C'}}}{\mathrm{d}I_{\mathrm{C}}} ,$$

the integral over the inverse cutoff frequency can be split up into a sum in the following way:

$$\int_0^{I_{\mathrm{C}}} \frac{1}{2\pi f'_{\mathrm{T}}}\,\mathrm{d}I_{\mathrm{C}} = \int_0^{I_{\mathrm{C}}} \tau_{\mathrm{f}}\,\mathrm{d}I_{\mathrm{C}} + \int_0^{V_{\mathrm{B'E'}}} c_{\mathrm{je}}(v)\,\mathrm{d}v + \int_{-V_{\mathrm{CE}}}^{V_{\mathrm{B'C'}}} c_{\mathrm{jc}}(v)\,\mathrm{d}v ,$$

where $V_{\mathrm{B'C'}}$ is the internal bc voltage at the bias point. From this, the hole charge stored in the internal transistor volume,

$$\begin{aligned} Q'_{\mathrm{p}}(V_{\mathrm{B'E'}}, V_{\mathrm{B'C'}}) &= Q_{\mathrm{P0}} + \int_0^{I_{\mathrm{C}}} \tau_{\mathrm{f}}\,\mathrm{d}I_{\mathrm{C}} \\ &\quad + \int_0^{V_{\mathrm{B'E'}}} c_{\mathrm{je}}(v)\,\mathrm{d}v + \int_0^{V_{\mathrm{B'C'}}} c_{\mathrm{jc}}(v)\,\mathrm{d}v , \end{aligned}$$

[61] The derivation presented here deviates somewhat from the original formulation but relies on the same idea. As explained in Sect. 3.5, (3.252) provides a good approximation as long as β is large, a requirement that becomes poorly fulfilled for very high-level injection. Assuming that operation conditions are chosen in order that $\beta \gg 1$, (3.252) is fulfilled for all geometries.

[62] In this expression, the forward transit time is understood to comprise the collector transit time τ_{jc}. The depletion capacitances c_{je} and c_{jc} denote depletion capacitances of the internal transistor.

results without any further approximations. As the forward transit time is a bias-dependent small-signal quantity, this formulation consistently relates the Gummel–Poon model to the transit time τ_{f} derived from small-signal measurements.

In order to make the approach adaptable to heterojunction bipolar transistors, further modifications of the Gummel–Poon model were introduced within the so-called generalized integral charge control relation (GICCR). The derivation of the GICCR [167] starts from (3.35), which is written as

$$i_{\mathrm{T}}(t) = eA_{\mathrm{je}}V_{\mathrm{T}}\frac{\exp\left(\dfrac{v_{\mathrm{B'E'}}(t)}{V_{\mathrm{T}}}\right) - \exp\left(\dfrac{v_{\mathrm{B'C'}}(t)}{V_{\mathrm{T}}}\right)}{\displaystyle\int_{x_{\mathrm{e}}}^{x_{\mathrm{c}}} \frac{p}{\mu_{\mathrm{n}} n_{\mathrm{ie}}^2} \exp\left(\frac{v_{\mathrm{B'E'}} - \phi_{\mathrm{p}}}{V_{\mathrm{T}}}\right) \mathrm{d}x}\,, \tag{3.253}$$

where "$v_{\mathrm{B'E'}}$ and $v_{\mathrm{B'C'}}$ are the base–emitter and base–collector voltage, respectively, but with the influence of the voltage drop across the emitter and collector contact resistance, as well as the base resistance being eliminated" [167]. In a further approximation, the integrand in the denominator of (3.253) is simplified by replacing the exponential function by one, not only in the base region, but also in the emitter and collector regions, where holes represent the minority carriers, resulting in

$$i_{\mathrm{T}}(t) = eA_{\mathrm{je}}V_{\mathrm{T}}\frac{\exp\left(\dfrac{v_{\mathrm{B'E'}}(t)}{V_{\mathrm{T}}}\right) - \exp\left(\dfrac{v_{\mathrm{B'C'}}(t)}{V_{\mathrm{T}}}\right)}{\displaystyle\int_{x_{\mathrm{e}}}^{x_{\mathrm{c}}} \frac{p}{\mu_{\mathrm{n}} n_{\mathrm{ie}}^2}\, \mathrm{d}x}\,. \tag{3.254}$$

The integral in the denominator is separated into integrals over the emitter region (E), the base region (B) and the collector region (C) without exactly specifying the limits of integration:

$$\int_{x_{\mathrm{e}}}^{x_{\mathrm{c}}} \frac{p}{\mu_{\mathrm{n}} n_{\mathrm{ie}}^2}\, \mathrm{d}x = \frac{1}{\langle \mu_{\mathrm{n}} n_{\mathrm{ie}}^2 \rangle_{\mathrm{E}}} \int_{\mathrm{E}} p\, \mathrm{d}x + \frac{1}{\langle \mu_{\mathrm{n}} n_{\mathrm{ie}}^2 \rangle_{\mathrm{B}}} \int_{\mathrm{B}} p\, \mathrm{d}x + \frac{1}{\langle \mu_{\mathrm{n}} n_{\mathrm{ie}}^2 \rangle_{\mathrm{C}}} \int_{\mathrm{C}} p\, \mathrm{d}x\,;$$

here appropriately defined averages of $\mu_{\mathrm{n}} n_{\mathrm{ie}}^2$ have been placed in front of each integral with the help of the mean-value theorem. The minority charges q_{TE} and q_{TC} stored in the emitter and collector regions are now defined according to

$$q_{\mathrm{TE}}(t) = eA_{\mathrm{je}} \int_{\mathrm{E}} p(x,t)\, \mathrm{d}x \quad \text{and} \quad q_{\mathrm{TC}}(t) = eA_{\mathrm{je}} \int_{\mathrm{C}} p(x,t)\, \mathrm{d}x\,,$$

while

$$eA_{\mathrm{je}} \int_{\mathrm{B}} p(x,t)\, \mathrm{d}x = Q_{\mathrm{P0}} + q_{\mathrm{JE}}(t) + q_{\mathrm{TB}}(t) + q_{\mathrm{JC}}(t)$$

is represented as the sum of the zero-bias base charge Q_{P0}, the charges q_{JE} and q_{JC} stored in the internal eb depletion capacitances, and the minority charge $q_{\rm TB}$ stored in the neutral base region. With these relations, (3.254) transforms into the "generalized integral charge control relation" [167]

$$i_{\rm T}(t) = \frac{e^2 A_{\rm je}^2 V_{\rm T} \langle \mu_{\rm n} n_{\rm ie}^2 \rangle_{\rm B} \left[\exp\left(\frac{v_{\rm B'E'}(t)}{V_{\rm T}} \right) - \exp\left(\frac{v_{\rm B'C}(t)}{V_{\rm T}} \right) \right]}{Q_{\rm P0} + q_{\rm JE} + q_{\rm JC} + H_{\rm FE} q_{\rm TE} + q_{\rm TB} + H_{\rm FC} q_{\rm TC}} \tag{3.255}$$

where

$$H_{\rm FE} = \frac{\langle \mu_{\rm n} n_{\rm ie}^2 \rangle_{\rm B}}{\langle \mu_{\rm n} n_{\rm ie}^2 \rangle_{\rm E}} \quad \text{and} \quad H_{\rm FC} = \frac{\langle \mu_{\rm n} n_{\rm ie}^2 \rangle_{\rm B}}{\langle \mu_{\rm n} n_{\rm ie}^2 \rangle_{\rm C}} \ .$$

A slight generalization of this transfer current model is implemented in the HICUM model.

3.15.2 Transfer Current

In HICUM the transfer current is written in the form [162]

$$i_{\rm T} = i_{\rm Tf} - i_{\rm Tr} = \frac{I_{\rm S}}{q_{\rm B}^*} \left[\exp\left(\frac{v_{\rm B'E'}(t)}{V_{\rm T}} \right) - \exp\left(\frac{v_{\rm B'C'}(t)}{V_{\rm T}} \right) \right] , \tag{3.256}$$

where the transfer saturation current is

$$I_{\rm S} = \frac{e^2 A_{\rm je}^2 V_{\rm T} \langle \mu_{\rm n} n_{\rm ie}^2 \rangle_{\rm B}}{Q_{\rm p0}} \ ;$$

the normalized "base charge" $q_{\rm B}^*$ is determined from[63]

$$\begin{aligned} q_{\rm B}^* = {} & 1 + \frac{1}{Q_{\rm P0}} \left[H_{\rm JEI} q_{\rm JEI} + H_{\rm JCI} q_{\rm JCI} \right. \\ & \left. + q_{\rm TF} + (H_{\rm FE} - 1) q_{\rm TE} + (H_{\rm FC} - 1) q_{\rm TC} + q_{\rm TR} \right] \ . \end{aligned} \tag{3.257}$$

This relation introduces two parameters, `HJEI` and `HJCI` in addition to `QP0`, `HFE` and `HFC`; these provide additional flexibility in modeling the Early effect, in exactly the same way as the parameters $V_{\rm ER}$ and $V_{\rm EF}$ were introduced to model the Early effect in VBIC and MEXTRAM – this flexibility is required for correct representation of the output characteristics of HBTs (see Chap. 4). $q_{\rm TF}$ describes the total minority charge stored in the neutral region under forward bias conditions in terms of the transit time. The extra terms $(H_{\rm FE} - 1) q_{\rm TE}$ and $(H_{\rm FC} - 1) q_{\rm TC}$ arise from the different weightings of the charges stored in the emitter and collector regions. For an approximate description of saturation effects, the minority charge $q_{\rm TR}$ associated with a forward bias of the bc diode is added.

[63] In order to avoid numerical problems in the case of base layer punchthrough, the charge $Q_{\rm p0} + q_{\rm JEI} + H_{\rm JCI} q_{\rm JCI}$ is limited to a positive value $0.05\, Q_{\rm P0}$ using a "smooth" interpolation formula [162].

Discussion. The author sees the HICUM model as an improved implementation of the Gummel–Poon approach, with a normalized base charge described in terms of the forward transit time, which can be obtained from small-signal measurements. There is only one small objection, concerning the transfer current relation, which can easily be overcome by proper choice of parameters. With the definitions of $v_{\mathrm{B'E'}}$ and $v_{\mathrm{B'C'}}$ used in (3.253) and (3.254), the effects due to voltage drops in the n-type emitter and epitaxial collector regions, which are usually described in terms of series resistances, are included in the Gummel integral. If series resistances due to the monocrystalline n-type regions are to be described as such, a truncation of the Gummel integral is required, as is explained in Appendix D.2. Since the emitter resistance in modern high-frequency bipolar transistors is essentially due to the contact resistance, the variation of the electron quasi-Fermi level across the monocrystalline emitter region can be neglected to a good approximation, thus justifying the step from (3.253) to (3.254). If the electron quasi-Fermi level can be assumed to be constant throughout the monocrystalline emitter region, the limits of integration in the derivation of the transfer current relation may be shifted towards the eb space charge region without affecting the result. Such a step, however, reduces the Gummel number, which now excludes the effects due to holes stored in the emitter region. Thus, making the step from (3.253) to (3.254) and maintaining the effects of the minority charge stored in the emitter seems to the author to be contradictory.

Another argument for excluding the minority charge in the emitter region from Gummel's transfer current relation is that the hole quasi-Fermi potential represents minority carriers there and should therefore not be considered constant in the emitter region. As a simple example, a transparent metal-contacted emitter with thickness d_{E} can be considered, where

$$\int_0^{d_{\mathrm{E}}} p\,\mathrm{d}x \;=\; \frac{d_{\mathrm{E}}p_{\mathrm{n0}}}{2}\exp\left(\frac{\phi_{\mathrm{pB}}-\phi_{\mathrm{nE}}}{V_{\mathrm{T}}}\right)$$

if low-level injection and a constant level of the electron quasi-Fermi potential within the emitter region are assumed. This result obviously deviates from

$$\int_0^{d_{\mathrm{E}}} p\,\exp\left(\frac{\phi_{\mathrm{pB}}-\phi_{\mathrm{p}}}{V_{\mathrm{T}}}\right)\mathrm{d}x \;=\; d_{\mathrm{E}}p_{\mathrm{n0}}\exp\left(\frac{\phi_{\mathrm{pB}}-\phi_{\mathrm{nE}}}{V_{\mathrm{T}}}\right) \;,$$

as is easily seen by the introduction of

$$p(x) \;=\; \frac{n_{\mathrm{ie}}^2}{N_{\mathrm{DE}}}\exp\left(\frac{\phi_{\mathrm{p}}-\phi_{\mathrm{nE}}}{V_{\mathrm{T}}}\right)$$

according to the generalized law of mass action.

Therefore, for a consistent implementation of Gummel's transfer current relation, the general choice $H_{\mathrm{FE}} = 0$ is suggested by the author.

3.15.3 Static Base Current, Parasitic PNP Transistor

The base current is modeled, in analogy with the Gummel Poon model, by a parallel connection of two diodes with different emission coefficients. In contrast to the Gummel–Poon model, the static base–emitter current is divided into a current component

$$i_{\rm BEI} = I_{\rm BEIS}\left[\exp\left(\frac{v_{\rm B'E'}}{M_{\rm BEI}V_{\rm T}}\right)-1\right]+I_{\rm REIS}\left[\exp\left(\frac{v_{\rm B'E'}}{M_{\rm REI}V_{\rm T}}\right)-1\right]$$

across the bottom of the diode, and a sidewall portion

$$i_{\rm BEP} = I_{\rm BEPS}\left[\exp\left(\frac{v_{\rm B^*E'}}{M_{\rm BEP}V_{\rm T}}\right)-1\right]+I_{\rm REPS}\left[\exp\left(\frac{v_{\rm B^*E'}}{M_{\rm REP}V_{\rm T}}\right)-1\right]\,,$$

i.e. a total of eight parameters, `IBEIS`, `MBEI`, `IREIS`, `MREI`, `IBEPS`, `MBEP`, `IREPS` and `MREP`, are available to model the eb recombination current. The recombination current of the forward-biased bc diode is modeled with less effort, in terms of a current

$$i_{\rm BCI} = I_{\rm BCIS}\left[\exp\left(\frac{v_{\rm B'C'}}{M_{\rm BCI}V_{\rm T}}\right)-1\right]$$

across the internal bc diode, and a current

$$i_{\rm BCX} = I_{\rm BCXS}\left[\exp\left(\frac{v_{\rm B^*C'}}{M_{\rm BCX}V_{\rm T}}\right)-1\right]$$

across the external bc diode, i.e. a total of four parameters, `IBCIS`, `MBCI`, `IBCXS` and `MBCX`, are available for modeling of the current in the forward-biased bc diode.

The transfer current $i_{\rm TS}$ of the parasitic pnp transistor is considered to be a second-order effect and is described by an elementary transistor model, which defines

$$i_{\rm TS} = i_{\rm TSf}-i_{\rm TSr} = I_{\rm TSS}\left[\exp\left(\frac{v_{\rm B^*C'}}{M_{\rm SF}V_{\rm T}}\right)-\exp\left(\frac{v_{\rm S'C'}}{M_{\rm SR}V_{\rm T}}\right)\right]$$

and

$$i_{\rm SC} = I_{\rm SCS}\left[\exp\left(\frac{v_{\rm S'C'}}{M_{\rm SC}V_{\rm T}}\right)-1\right]\,.$$

The parasitic pnp transistor may become relevant if the buried layer does not completely separate the epitaxial collector region from the substrate.

3.15.4 Series Resistances

HICUM takes acount of three ohmic series resistances, $R_{\rm BX}$, $R_{\rm CX}$ and $R_{\rm E}$, in series with the base, collector and emitter terminals, which are modeled as temperature-dependent quantities of the form

$$R_{\mathrm{BX}}(T) \;=\; R_{\mathrm{BX}}\left(\frac{T}{T_0}\right)^{\zeta_{\mathrm{RBX}}} \;;$$

analogous relations are used for $R_{\mathrm{CX}}(T)$ and $R_{\mathrm{E}}(T)$, with the model parameters ζ_{RCX} and ζ_{RE}. The model for the internal base resistance takes account of both conductivity modulation and emitter current crowding:

$$R_{\mathrm{BI}} \;=\; R_{\mathrm{BI0}}\,\frac{Q_0}{Q_0+\Delta Q_{\mathrm{p}}}\,\psi(\eta)\;,$$

where R_{BI0} is a model parameter. $Q_0 = (1+F_{\mathrm{DQR0}})Q_{\mathrm{P0}}$ is comparable to the zero-bias base charge Q_{P0} and is determined with the help of the model parameter F_{DQR0}; the modification takes account of the fact, that the hole charge has a different weighting in the Gummel integral and in the computation of the base sheet resistance. ΔQ_{p} denotes the stored hole charge:

$$\Delta Q_{\mathrm{p}} \;=\; q_{\mathrm{JEI}} + q_{\mathrm{JCI}} + q_{\mathrm{TF}} + q_{\mathrm{TR}}\;,$$

and takes account of conductivity modulation. The factor,

$$\psi(\eta) \;=\; \frac{1}{\eta}\ln(1+\eta)\;, \qquad \text{where} \qquad \eta \;=\; F_{\mathrm{GE0}}R_{\mathrm{BI0}}\,\frac{Q_0}{Q_0+\Delta Q_{\mathrm{p}}}\,\frac{i_{\mathrm{BEi}}}{V_{\mathrm{T}}}\;,$$

describes the effect of emitter current crowding [161, 168]. F_{GE0} denotes a model parameter that takes account of the geometry dependence of emitter current crowding.

HICUM takes account of a substrate resistance R_{Su}, introduced between the external substrate node S and the internal substrate node S′; in high-resistivity substrate layers, the displacement current may become relevant at frequencies in excess of that corresponding to the dielectric time constant $\tau_\epsilon = \rho_{\mathrm{sub}}\epsilon$. This effect is modeled by placing a capacitance C_{Su} in parallel with R_{Su}, with C_{Su} chosen in order that $\tau_\epsilon = R_{\mathrm{Su}}C_{\mathrm{Su}}$.

3.15.5 Charge Storage

The depletion capacitance is modeled in analogy with the SPICE Gummel–Poon model under reverse bias and small forward bias; a modified expression is, however, used to describe the behavior at large forward bias. The modified formulation removes the discontinuity in the conventional capacitance–voltage characteristic at $v \to V_{\mathrm{J}}$ and yields a continuously differentiable capacitance–voltage characteristic that is clamped to $A_{\mathrm{LJ}}C_{\mathrm{J}}$ at large values of forward bias. Figure 3.67 illustrates the modeling approach in comparison with other approaches.

The charge stored in the internal eb depletion capacitances is determined from[64]

[64]Equivalent relations hold for the emitter sidewall capacitance with C_{JEI0}, V_{DEI}, Z_{EI} and A_{LJEI} replaced by C_{JEP0}, V_{DEP}, Z_{EP} and A_{LJEP}.

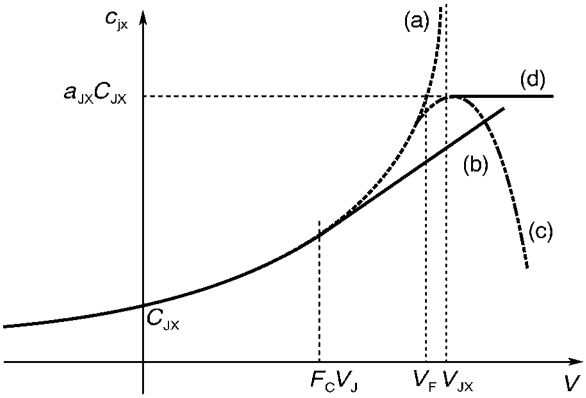


Fig. 3.67. Modeling of depletion capacitances. (**a**) Elementary power law with divergence at $V = V_{JX}$, (**b**) SPICE model with linear approximation for $V > F_C V_{JX}$, (**c**) correct capacitance voltage characteristic, (**d**) HICUM and MEXTRAM approach with a continuously differentiable, clamped capacitance–voltage characteristic

$$q_{JEI} = \int_0^{v_{B'E'}} c_{jei}(V)\, dV \,, \tag{3.258}$$

where the bias-dependent internal eb depletion capacitance is

$$c_{jei}(v_{B'E'}) = \frac{C_{JEI0}}{(1-v_{jE}/V_{DEI})^{Z_{EI}}} \frac{1}{1+\exp[(v_{B'E'}-V_{FE})/V_T]} + \frac{A_{LJEI}\, C_{JEI0}}{1+\exp[(V_{FE}-v_{B'E'})/V_T]} \,. \tag{3.259}$$

Here the HICUM parameters `CJEI0`, `VDEI` and `ZEI` correspond to the SPICE parameters `CJE`, `VJE` and `MJE`. The model parameter `ALJEI` limits the value of the eb depletion capacitance to $A_{LJEI} C_{JEI0}$. The voltage

$$V_{FE} = V_{DEI}\left(1-A_{LJEI}^{-1/Z_{EI}}\right)$$

is the value at which the conventional power-law model would reach the maximum voltage, and $v_{jE} = S(v_{B'E'}, V_{FE}, V_T)$ is the "smoothed junction voltage", defined in terms of the smoothing function

$$S(\xi, a, b) = a - b \ln\left[1+\exp\left(\frac{a-\xi}{b}\right)\right] \,, \tag{3.260}$$

which is a continuously differentiable function that interpolates between the limiting cases $S(\xi,a,b) \approx \xi$ for $a-\xi \gg b$ and $S(\xi,a,b) \approx a$ for $a-\xi \ll -b$.

The charge q_{BCi} is composed of the diffusion charge associated with a forward-biased bc diode, written as $\tau_R i_{Tr}$, and the charge stored in the internal bc depletion capacitance, which is determined from

$$q_{JCI} = \int_0^{v_{B'C'}} c_{jci}(V)\, dV \,; \tag{3.261}$$

the internal bc depletion capacitance c_{jci} is described by

$$
\begin{aligned}
c_{\mathrm{jci}} \;=\; & \frac{C_{\mathrm{JCI0}}}{\left(1-\dfrac{v_{\mathrm{jm}}}{V_{\mathrm{DCI}}}\right)^{Z_{\mathrm{CI}}}} \, \frac{1}{1+\exp\!\left(\dfrac{v_{\mathrm{B'C'}}-V_{\mathrm{FCI}}}{V_{\mathrm{T}}}\right)} \, \frac{1}{1+\exp\!\left(-\dfrac{v_{\mathrm{jr}}+V_{\mathrm{JPCI}}}{V_{\mathrm{R}}}\right)} \\
& + \frac{C_{\mathrm{JCI0}}}{\left(1-\dfrac{v_{\mathrm{jr}}}{V_{\mathrm{DCI}}}\right)^{Z_{\mathrm{CI}}/4}} \left(\frac{V_{\mathrm{DCI}}}{V_{\mathrm{PTCI}}}\right)^{3Z_{\mathrm{CI}}/4} \frac{1}{1+\exp\!\left(\dfrac{v_{\mathrm{jr}}+V_{\mathrm{JPCI}}}{V_{\mathrm{R}}}\right)} \\
& + \frac{A_{\mathrm{JCI}} C_{\mathrm{JCI0}}}{1+\exp\!\left(\dfrac{V_{\mathrm{FCI}}-v_{\mathrm{B'C'}}}{V_{\mathrm{T}}}\right)} \, ,
\end{aligned}
\tag{3.262}
$$

where V_{JPCI} denotes the effective punchthrough voltage, defined in terms of the model parameters V_{PTCI} and V_{DCI} by

$$V_{\mathrm{JPCI}} \;=\; V_{\mathrm{PTCI}} - V_{\mathrm{DCI}} \; ,$$

and $V_{\mathrm{R}} = 4V_{\mathrm{T}} + V_{\mathrm{JPCI}}/10$. The punchthrough voltage V_{PTCI} represents an extra model parameter that may be chosen independently from the punchthrough voltage V_{PT} used for the computation of I_{CK}. Equation (3.262) describes a superposition of a depletion capacitance that follows the conventional power law

$$c_{\mathrm{jci}} \;\approx\; \frac{C_{\mathrm{JCI0}}}{(1-v_{\mathrm{B'C'}}/V_{\mathrm{DCI}})^{Z_{\mathrm{CI}}}} \tag{3.263}$$

for small forward biases $v_{\mathrm{B'C'}} < V_{\mathrm{FCI}} = V_{\mathrm{DCI}}\left(1-A_{\mathrm{JCI}}^{-1/Z_{\mathrm{CI}}}\right)$, and for reverse-bias levels $v_{\mathrm{C'B'}} < V_{\mathrm{JPCI}}$, i.e. before the onset of collector punchthrough. At large forward bias, the capacitance is clamped to $A_{\mathrm{JCI}}C_{\mathrm{JCI0}}$, similarly to the formulation for the eb diode. At large reverse bias, after the onset of collector punchthrough, the bias dependence of the depletion capacitance is approximately

$$c_{\mathrm{jci}} \;\approx\; \frac{C_{\mathrm{JCI0}}}{(1-v_{\mathrm{jr}}/V_{\mathrm{DCI}})^{Z_{\mathrm{CI}}/4}} \left(\frac{V_{\mathrm{DCI}}}{V_{\mathrm{PTCI}}}\right)^{3Z_{\mathrm{CI}}/4} \, ,$$

with a substantially reduced voltage dependence. The factors of the form $1/(1+\mathrm{e}^{x})$ have the sole purpose of switching between the different limiting situations; the "smoothed" junction voltages

$$v_{\mathrm{jr}} \;=\; S(v_{\mathrm{B'C'}}, V_{\mathrm{FCI}}, V_{\mathrm{T}}) \quad \text{and} \quad v_{\mathrm{jm}} \;=\; -S(-v_{\mathrm{jr}}, V_{\mathrm{JPCI}}, V_{\mathrm{R}}) \, ,$$

where $S(x,a,b)$ is defined in (3.260), are introduced to obtain continuously differentiable functions.

The capacitance c_{BCX} associated with the external bc diode is composed of the external bc depletion capacitance c_{jcx} and the capacitance C_{COX}. Its value is split up into a portion $c'_{\mathrm{BCX}} = (1-F_{\mathrm{BC}})c_{\mathrm{BCX}}$ connected to the node

B and a portion $c''_{\mathrm{BCX}} = F_{\mathrm{BC}} c_{\mathrm{BCX}}$ connected to the node B*, using the model parameter FBC. The external bc depletion capacitance c_{jcx} is calculated in complete analogy with c_{jci}, using the parameters CJCIX, VDCX, ZCX and VPTCX.

The minority charge q_{TF} stored in the neutral regions under forward-bias conditions is computed from the bias-dependent forward transit time τ_{f}

$$q_{\mathrm{TF}} = \int_0^{i_{\mathrm{Tf}}} \tau_{\mathrm{f}}(i)\,\mathrm{d}i\,. \tag{3.264}$$

The forward transit time $\tau_{\mathrm{f}}(i)$ is described as a bias-dependent quantity in the form

$$\tau_{\mathrm{f}} = \tau_0 + \Delta\tau_{0\mathrm{h}}\,(c_{\mathrm{c}}-1) + \tau_{\mathrm{BVL}}\left(\frac{1}{c_{\mathrm{c}}}-1\right) + \Delta\tau_{\mathrm{Ef}} + \Delta\tau_{\mathrm{fh}}\,, \tag{3.265}$$

where $c_{\mathrm{c}} = C_{\mathrm{JCI0}}/c_{\mathrm{jci}}(v_{\mathrm{B'C'}})$ and τ_0, $\Delta\tau_{0\mathrm{h}}$ and τ_{BVL} are model parameters; the additional terms $\Delta\tau_{\mathrm{Ef}}$ and $\Delta\tau_{\mathrm{fh}}$ describe effects that become relevant at high current densities. The first term, τ_0, in (3.265) takes account of all voltage-independent contributions to the transit time, determined at small current values and $v_{\mathrm{B'C'}} = 0$; the second term, $\Delta\tau_{0\mathrm{h}}(c_{\mathrm{c}}-1)$, takes account of the effect of the voltage-dependent width of the bc depletion layer on the transit time: increasing values of the reverse bias $v_{\mathrm{C'B'}}$ will cause a reduction of the base width and thus the base transit time, but on the other hand an increase due to the larger value of the collector transit time has to be taken into account. If $\Delta\tau_{0\mathrm{h}} > 0$, the effect of the collector transit time is more pronounced; if the effect on the base transit time predominates, $\Delta\tau_{0\mathrm{h}}$ is negative. The third term, $\tau_{\mathrm{BVL}}\,(1/c_{\mathrm{c}}-1)$, in (3.265) takes account of the fact that the base transit time is increased owing to the finite carrier velocity at the bc depletion layer edge.

The onset of high-current effects is characterized by a critical current

$$I_{\mathrm{CK}} = \frac{v_{\mathrm{Ceff}}}{R_{\mathrm{CI0}}}\,\frac{2 + x + \sqrt{x^2+10^{-3}}}{2\sqrt{1+(v_{\mathrm{Ceff}}/V_{\mathrm{LIM}})^2}}\,, \qquad x = \frac{v_{\mathrm{Ceff}}-V_{\mathrm{LIM}}}{V_{\mathrm{PT}}}\,, \tag{3.266}$$

where

$$R_{\mathrm{CI0}} \approx \frac{d_{\mathrm{epi}}}{e\mu_{\mathrm{n0}} N_{\mathrm{Depi}} A_{\mathrm{je}}} \quad \text{and} \quad V_{\mathrm{PT}} \approx \frac{eN_{\mathrm{Depi}} d_{\mathrm{epi}}^2}{2\epsilon}$$

denote the epilayer resistance and the collector punchthrough voltage; the voltage $V_{\mathrm{LIM}} \approx d_{\mathrm{epi}} v_{\mathrm{nsat}}/\mu_{\mathrm{n0}}$ describes the critical voltage for the onset of velocity saturation effects.[65] R_{CI0}, V_{PT} and V_{LIM} are model parameters; the "effective collector voltage" v_{Ceff} is defined according to

$$v_{\mathrm{Ceff}} = V_{\mathrm{T}}\left\{1 + \ln\left[1 + \exp\left(\frac{v_{\mathrm{C'E'}} - V_{\mathrm{CES}} - V_{\mathrm{T}}}{V_{\mathrm{T}}}\right)\right]\right\}\,,$$

[65] This parameter corresponds to the parameter V_0 of the Kull epilayer model; at this voltage, $I_2 = I_1$ [169].

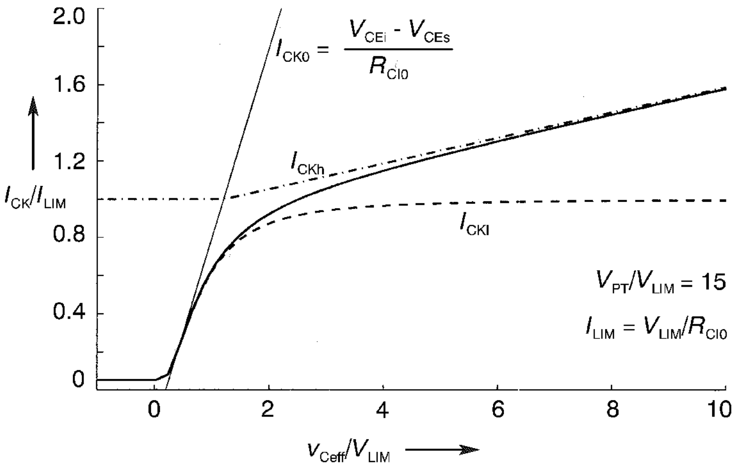


Fig. 3.68. Critical current I_{CK} normalized to I_{LIM}, versus normalized ce voltage (after [162])

where the saturation voltage $V_{\mathrm{CES}} \approx V_{\mathrm{DEI}} - V_{\mathrm{DCI}}$ is a model parameter. The function v_{Ceff} interpolates continuously between the approximation $v_{\mathrm{Ceff}} \approx v_{\mathrm{C'E'}} - V_{\mathrm{CES}}$, valid if $v_{\mathrm{C'E'}} \gg V_{\mathrm{CES}} + V_{\mathrm{T}}$, and $v_{\mathrm{Ceff}} \approx V_{\mathrm{T}}$, valid if $v_{\mathrm{C'E'}} \ll V_{\mathrm{CES}} + V_{\mathrm{T}}$. Equation (3.266) interpolates smoothly (see Fig. 3.68) between the limiting situations of small and large collector–emitter voltage [160],

$$I_{\mathrm{CK}} = \begin{cases} I_{\mathrm{CKl}} = \dfrac{V_{\mathrm{Ceff}}}{R_{\mathrm{CI0}}} \dfrac{1}{\sqrt{1 + (V_{\mathrm{Ceff}}/V_{\mathrm{LIM}})^2}} & \text{if } V_{\mathrm{Ceff}} \ll V_{\mathrm{LIM}} \\[2ex] I_{\mathrm{CHh}} = \dfrac{V_{\mathrm{LIM}}}{R_{\mathrm{CI0}}} \left(1 + \dfrac{V_{\mathrm{Ceff}} - V_{\mathrm{LIM}}}{V_{\mathrm{PT}}}\right) & \text{if } V_{\mathrm{Ceff}} \gg V_{\mathrm{LIM}} \end{cases} .$$

The minority charge q_{TE} stored in the neutral emitter region is expressed in terms of the emitter transit time [170]:

$$\tau_{\mathrm{Ef}} = \tau_{\mathrm{EF0}} \left[1 + \left(\frac{i_{\mathrm{Tf}}}{I_{\mathrm{CK}}} \right)^{G_{\mathrm{TE}}} \right] = \tau_{\mathrm{EF0}} + \Delta\tau_{\mathrm{Ef}} ;$$

the factor $\left[1 + (i_{\mathrm{Tf}}/I_{\mathrm{CK}})^{G_{\mathrm{TE}}} \right]$ takes account of the decrease of the (dc) small-signal current gain β at current levels exceeding I_{CK}, since τ_{E} varies in inverse proportion to β (see Sect. 3.5). The contribution τ_{EF0} to the emitter transit time is included in the parameter τ_0; the result for q_{TE} is

$$q_{\mathrm{TE}} = \int_0^{i_{\mathrm{TF}}} \tau_{\mathrm{EF}}(i)\, \mathrm{d}i = \tau_{\mathrm{EF0}} \left[1 + \frac{1}{1+G_{\mathrm{TE}}} \left(\frac{i_{\mathrm{Tf}}}{I_{\mathrm{CK}}} \right)^{G_{\mathrm{TE}}} \right] i_{\mathrm{Tf}} .$$

High-level-injection effects also affect the minority charge stored in the neutral collector and base regions. The charge stored in the neutral collector region can be calculated in accordance with [86]

$$\Delta Q_{\mathrm{Cf}} = \tau_{\mathrm{pcs}} \left(1-\frac{I_{\mathrm{CK}}}{i_{\mathrm{Tf}}}\right)^2 i_{\mathrm{Tf}} \, ,$$

which can be applied under the assumption $i_{\mathrm{Tf}} > I_{\mathrm{CK}}$; in order to avoid unphysical results at values of i_{Tf} below I_{CK}, the relation is modified to

$$\Delta Q_{\mathrm{Cf}} = \tau_{\mathrm{pcs}}\, i_{\mathrm{Tf}} \left(\frac{i+\sqrt{i^2+A_{\mathrm{LHC}}}}{1+\sqrt{1+A_{\mathrm{LHC}}}}\right)^2 \, ,$$

where $i = 1 - I_{\mathrm{CK}}/i_{\mathrm{Tf}}$ and $A_{\mathrm{LHC}} > 0$ is a model parameter, that is fitted to measured data. The expression for ΔQ_{Cf} is continuously differentiable and results in the transit time [170]

$$\tau_{\mathrm{Cf}} = \tau_{\mathrm{pcs}} \left(\frac{i+\sqrt{i^2+A_{\mathrm{LHC}}}}{1+\sqrt{1+A_{\mathrm{LHC}}}}\right)^2 \left(1+\frac{2I_{\mathrm{CK}}}{i_{\mathrm{Tf}}\sqrt{i^2+A_{\mathrm{LHC}}}}\right) .$$

A corresponding relation is used for the increase of the base transit time:

$$\Delta\tau_{\mathrm{Bf}} = \tau_{\mathrm{Bfvs}} \left(\frac{i+\sqrt{i^2+A_{\mathrm{LHC}}}}{1+\sqrt{1+A_{\mathrm{LHC}}}}\right)^2 \left(1+\frac{2I_{\mathrm{CK}}}{i_{\mathrm{Tf}}\sqrt{i^2+A_{\mathrm{LHC}}}}\right) .$$

This approach is motivated by device simulations [162] and is plausible because of the coupling between the two regions through the carrier density at the bc junction. The sum $\tau_{\mathrm{Cf}} + \Delta\tau_{\mathrm{Bf}}$ yields the transit time component

$$\Delta\tau_{\mathrm{fh}} = \tau_{\mathrm{HCS}} \left(\frac{i+\sqrt{i^2+A_{\mathrm{LHC}}}}{1+\sqrt{1+A_{\mathrm{LHC}}}}\right)^2 \left(1+\frac{2I_{\mathrm{CK}}}{i_{\mathrm{Tf}}\sqrt{i^2+A_{\mathrm{LHC}}}}\right) , \qquad (3.267)$$

where $\tau_{\mathrm{HCS}} = \tau_{\mathrm{pcs}} + \tau_{\mathrm{Bfvs}}$ is a model parameter. The charge stored in the epitaxial region under the external base is described by the diffusion charge

$$q_{\mathrm{DS}} = \tau_{\mathrm{SF}} i_{\mathrm{TSf}} \, ,$$

using the model parameter `TSF`.

HICUM takes account of non-quasi-static effects by a method analogous to the extended charge control model of te Winkel (see Sect. 3.1), which describes the stored minority charge q_{TF} and the transfer current i_{T} as delayed functions of $v_{\mathrm{B'E'}}(t)$. The corresponding delay times for the minority charge and transfer current are calculated as fixed fractions of the forward transit time [171]:

$$\tau_2 = A_{\mathrm{LQF}}\, \tau_{\mathrm{f}} \quad \text{and} \quad \tau_{\mathrm{m}} = A_{\mathrm{LIT}}\, \tau_{\mathrm{f}} \, ,$$

where A_{LQF} and A_{LIT} are model parameters. To avoid numerical problems with small delay times in a time-domain analysis, the delays are implemented using second-order Bessel polynomials (see Sect. 3.9) for both time- and frequency-domain analyses in order to maintain consistency of the corresponding results.

Transient emitter current crowding is taken into account through introduction of a capacitance C_{RBI} parallel to R_{BI}; the value of C_{RBI} is calculated as a fraction of the total capacitance connected to the internal base node B′:

$$C_{\mathrm{RBI}} = F_{\mathrm{CRBI}}\,(c_{\mathrm{jei}}+c_{\mathrm{bci}}+c_{\mathrm{te}}) \ .$$

Here the model parameter F_{CRBI}, which is approximately equal to 0.2 for long stripe transistors (see Sect. 3.9), is used.

3.15.6 BC Avalanche Effect

The avalanche current $i_{\mathrm{AVL}} = (M_{\mathrm{n}} - 1)\, i_{\mathrm{T}}$ is modeled similarly to the VBIC model using the approximation given in [50],

$$\begin{aligned} M_{\mathrm{n}} - 1 &\approx \int_{x_{\mathrm{bc}}}^{x_{\mathrm{cb}}} a_{\mathrm{n}} \exp\left(-\frac{b_{\mathrm{n}}}{|E(x)|}\right) \mathrm{d}x \\ &\approx F_{\mathrm{AVL}}(V_{\mathrm{DCI}}-v_{\mathrm{B'C'}}) \exp\left(-\frac{Q_{\mathrm{AVL}}}{c_{\mathrm{jci}}(V_{\mathrm{DCI}}-v_{\mathrm{B'C'}})}\right) \ , \end{aligned}$$

which is rewritten as[66]

$$M_{\mathrm{n}} - 1 \approx F_{\mathrm{AVL}} V_{\mathrm{DCI}} c_{\mathrm{c}}^{1/Z_{\mathrm{CI}}} \exp\left(-\frac{Q_{\mathrm{AVL}}}{C_{\mathrm{JCI0}} V_{\mathrm{DCI}}}\, c_{\mathrm{c}}^{1/Z_{\mathrm{Ci}}-1}\right) \ ,$$

where $c_{\mathrm{c}} = c_{\mathrm{jci}}(v_{\mathrm{B'C'}})/C_{\mathrm{JCI0}}$ denotes the normalized bc depletion capacitance, in order to avoid a possible numerical instability if $v_{\mathrm{B'C'}} \to V_{\mathrm{DCI}}$. A possible current-dependence of the multiplication factor is not considered in this model.

3.15.7 Emitter–Base Tunneling

Heavily doped eb diodes cause a substantial reverse current due to tunneling at a comparably small reverse bias. The effect is easily taken into account in a compact model by introduction of a voltage-controlled current source between the internal base and emitter nodes of the model. The model adopted by HICUM [162,165] corresponds to (3.65) with a tunneling-current density of

[66] If c_{jci} is modeled by the simple capacitance formula (3.263), the identity

$$V_{\mathrm{DCI}} - v_{\mathrm{B'C'}} = V_{\mathrm{DCI}}\, c_{\mathrm{c}}^{-1/Z_{\mathrm{CI}}}$$

is obtained. Replacing $V_{\mathrm{DCI}} - v_{\mathrm{B'C'}}$ by the term on the right hand side expressed with the improved capacitance model, which limits c_{jci} to a maximum value $A_{\mathrm{JCI}} C_{\mathrm{JCI0}}$, avoids the divergence.

the form $J_T = aV_{EB}E\exp(-E_0/E)$, where E is taken to be equal to the maximum electric field strength E_{max}. Estimating E_{max} in terms of the peripheral eb depletion capacitance (using the parameters C_{JEP}, V_{DEP}, Z_P) yields the following current–voltage characteristic:

$$i_{BET} = I_{BETS}\, c_{ep}^{1-1/Z_P} V_{EB} \exp\left(-A_{BET}\, c_{ep}^{1-1/Z_P}\right) , \tag{3.268}$$

where $c_{ep} = c_{jep}(V_{B'E'})/C_{JEP}$ denotes the normalized eb depletion capacitance, and I_{BETS} and A_{BET} denote two model parameters, with a temperature dependence that is predominantly caused by the temperature dependence of the bandgap [162].

3.15.8 Temperature Effects

The model parameters are specified at the reference temperature T_0 and updated to the actual device temperature T, which is calculated from the ambient temperature and the excess temperature caused by self-heating effects. The temperature dependence of the transfer saturation current is described by

$$I_S(T) = I_S \frac{1}{1+\dfrac{Z_{EI}}{2}\left(1-\dfrac{V_{DEI}(T)}{V_{DEI}}\right)} \left(\frac{T}{T_0}\right)^3 \exp\left[\frac{V_{GB}}{V_T}\left(\frac{T}{T_0}-1\right)\right] ,$$

where V_{GB} denotes an average value of the bandgap voltage in the base region extrapolated to $T = 0$. The first fraction takes acount of the temperature dependence of the zero-bias base charge, which is predominantly due to the temperature dependence of the depletion-layer widths. The remaining factors correspond to the SPICE Gummel–Poon model, with X_{TI} chosen to be equal to three. The temperature dependences of all eb diode saturation currents are modeled in the form

$$I_{BS}(T) = I_{BS}\left(\frac{T}{T_0}\right)^3 \exp\left[\frac{V_{GB}}{MV_T}\left(\frac{T}{T_0}-1\right) - A_{LB}(T-T_0)\right] ,$$

where (I_{BS}, M) is given by (I_{BEIS}, M_{BEI}), (I_{REIS}, M_{REI}), (I_{BEPS}, M_{BEP}) and (I_{REPS}, M_{REP}). This gives an approximately linear temperature dependence (with temperature coefficient A_{LB}) of the forward current gain (measured at constant I_C), under low-level-injection conditions. The saturation currents of the bc and the cs junctions obey corresponding relations, but with A_{LB} set to zero. The temperature dependence of the depletion capacitances and built-in voltages is described analogously to (3.222) and (3.221).

In analogy with the VBIC and MEXTRAM models, self-heating is taken into account by means of a one-pole thermal equivalent circuit, specified by the model parameters R_{th} and C_{th} and a power dissipation calculated from all dissipative elements of the equivalent circuit.

3.16 The MEXTRAM Model

This section describess another physics-based model that has been introduced to improve on the standard Gummel Poon model: the MEXTRAM[67] model [94, 172] developed at Philips in 1985 and updated in 1987 (MEXTRAM 502) [61], 1993 (MEXTRAM 503) and 2000 (MEXTRAM 504).

The following survey is based on the specifications of the Level 504 model released in April 2001 [94], and the specifications of the Level 503 model of December 2000 [172]. The focus will be on the Level 504 model, which is claimed to provide better results for the description of first- and higher-order derivatives, resulting in, for example, more accurate descriptions of the output conductance, cutoff frequency and low-frequency third-order distortion. In the Level 504 model, special attention is paid to SiGe transistors, and self-heating effects are included in the description using a one-pole thermal equivalent circuit similarly to the approach implemented in VBIC. As compared with the Level 503 model, the number of modeling parameters has been increased to provide greater modeling flexibility. In contrast to early MEXTRAM versions, all model equations are now given as explicit functions of internal branch voltages and no internal quantities have to be determined from implicit equations by iteration. The model allows the specification of 66 parameters: 25 are used for forward-current modeling, 6 for reverse-current modeling, 14 for charge modeling, 14 to take account of temperature dependences, 2 for the simulation of self-heating effects, 3 for the noise model, and 2 parameters are included specifically for HBT modeling; 7 additional general parameters determine the level employed, the reference temperature and flags to choose between various modeling approaches.

Table 3.7 lists the 66 transistor parameters specified in the Level 504 model for integrated npn bipolar transistors. These parameters appear in positions 8 to 74 on the parameter list of the model, in the order given in Table 3.7. The parameters in positions 1 to 7 of the list (not shown in Table 3.7) specify the `LEVEL` and the parameter `MULT` (default 1), which can be used to characterize a parallel connection of identical transistors: the parameters `IS`, `IK`, `IBF`, `IBR`, `ISS`, `IKS`, `QB0`, `IHC`, `CJE`, `CJC` and `CJS`, are multiplied by `MULT`, whereas the parameters `RCC`, `SCRCV`, `RCV`, `RBC`, `RBV` and `RE` are divided by `MULT`; the $1/f$ noise coefficients `KF` and `KFN` are scaled as $K_{\mathrm{F}} \times \mathtt{MULT}^{1-A_{\mathrm{F}}}$ and $K_{\mathrm{FN}} \times \mathtt{MULT}^{1-[2(M_{\mathrm{LF}}-1)+A_{\mathrm{F}}(2-M_{\mathrm{LF}})]}$. Further parameters are the reference temperature `TREF` (default 25°C), the excess temperature `DTA` (default 0 K), and three flags `EXMOD`, `EXPHI` and `EXAVL` to allow extended modeling of reverse current gain, distributed high-frequency effects and avalanche effects.

---

[67] The acronym MEXTRAM is an abbreviation of "most exquisite transistor model" – a name that certainly avoids unnecessary understatement. Documentation of the model equations is provided on the web site http://www.semiconductors.philips.com/Philips_Models/

Table 3.7. MEXTRAM model parameters and corresponding parameters in SPICE

Parameter name	Parameter	Unit	SPICE
Internal transistor			
Transfer saturation current	I_S, IS	A	IS
Forward knee current	I_K, IK	A	IKF
Reverse Early voltage	V_{ER}, VER	V	VAR
Forward Early voltage	V_{EF}, VEF	V	VAF
Ideal forward current gain	B_F, BF	–	BF
Saturation current of eb leakage diode	I_{BF}, IBF	A	ISE
Emission coefficient of eb leakage diode	M_{LF}, MLF	–	NE
EB diode partition factor	X_{IBI}, XIBI	–	–
Ideal reverse current gain	B_{RI}, BRI	–	BR
Saturation current of bc leakage diode	I_{BR}, IBR	A	ISC
Crossover voltage of bc leakage diode	V_{LR}, VLR	V	–
BC diode partition factor	X_{EXT}, XEXT	–	–
BC avalanche current			
Avalanche zone thickness	W_{AVL}, WAVL	m	–
Avalanche curvature voltage	V_{AVL}, VAVL	V	–
Avalanche current spreading factor	S_{FH}, SFH	–	–
Series resistances			
Emitter resistance	R_E, RE	Ω	RE
Base resistance (external)	R_{BC}, RBC	Ω	RBM
Base resistance (variable portion)	R_{BV}, RBV	Ω	RB-RBM
Collector resistance (external)	R_{CC}, RCC	Ω	RC
Epilayer resistance	R_{CV}, RCV	Ω	RCO
Epilayer space charge resistance	S_{CRCV}, SCRCV	Ω	–
Critical current for velocity saturation	I_{HC}, IHC	A	–
Quasi-saturation parameter	A_{XI}, AXI	–	–
EB capacitance			
EB depletion capacitance ($V_{BE} = 0$)	C_{JE}, CJE	F	CJE
EB built-in voltage	V_{DE}, VDE	V	VJE
EB grading exponent	P_E, PE	–	MJE
Partition factor for c_{je}	X_{CJE}, XCJE	–	–
EB overlap capacitance	C_{BEO}, CBEO	F	–
BC capacitance			
BC depletion capacitance ($V_{BC} = 0$)	C_{JC}, CJC	F	CJC
BC built-in voltage	V_{DC}, VDC	V	VJC
BC grading exponent	P_C, PC	–	MJC
Constant part of c_{jc}	X_P, XP	–	–
Current-modulation coefficient for c_{jc}	M_C, MC	–	–
Fraction of internal bc capacitance	X_{CJC}, XCJC	–	XCJC
BC overlap capacitance	C_{BCO}, CBCO	F	–
Diffusion charge			
Emitter diffusion charge coefficient	m_τ, MTAU	–	–
Minimum emitter transit time	τ_E, TAUE	s	–
Base transit time	τ_B, TAUB	s	–

Table 3.7. (continued)

Parameter name	Parameter	Unit	SPICE
Epilayer transit time	τ_{EPI}, TEPI	s	–
Reverse base transit time	τ_{R}, TAUR	s	–
HBT modeling options			
Bandgap difference over the base	ΔE_{g}, DEG	eV	–
EB recombination current prefactor	X_{REC}, XREC	–	–
Temperature dependences			
T exponent of zero base charge	A_{QBO}, AQBO	–	–
T exponent of emitter resistivity	A_{E}, AE	–	–
T exponent of base resistivity	A_{B}, AB	–	–
T exponent of epilayer resistivity	A_{EPI}, AEPI	–	–
T exponent of extrinsic base	A_{EX}, AEX	–	–
T exponent of buried-layer resistivity	A_{C}, AC	–	–
BF bandgap voltage difference[a])	ΔV_{GBF}, DVGBF	V	–
BR bandgap voltage difference[b])	ΔV_{GBR}, DVGBR	V	–
Base bandgap voltage	V_{GB}, VGB	V	EG
Collector bandgap voltage	V_{GC}, VGC	V	EG
EB recombination bandgap voltage	V_{GJ}, VGJ	V	EG
TAUE bandgap voltage difference[c])	ΔV_{GTE}, DVGTE	V	–
Noise modeling			
$1/f$ noise coefficient (ideal)	K_{F}, KF	–	KF
$1/f$ noise coefficient (nonideal)	K_{FN}, KFN	–	–
$1/f$ noise exponent	A_{F}, AF	–	AF
CS diode			
PNP transistor saturation current	I_{SS}, ISS	A	–
Substrate knee current	I_{KS}, IKS	A	–
CS depletion capacitance ($V_{\mathrm{CS}} = 0$)	C_{JS}, CJS	F	CJS
CS built-in voltage	V_{DS}, VDS	V	VJS
CS grading exponent	P_{S}, PS	–	0
Substrate bandgap voltage	V_{GS}, VGS	V	EG
T exponent for I_{SS}	A_{S}, AS	–	–
Self-heating			
Thermal resistance	R_{th}, RTH	K/W	–
Thermal capacitance	C_{th}, CTH	J/K	–

[a]) Used to describe the temperature dependence of the ideal forward current gain B_{F}
[b]) Used to describe the temperature dependence of the ideal reverse current gain B_{R}
[c]) Used to describe the temperature dependence of the minimum emitter transit time τ_{E}

The extraction of model parameters is documented in [173]. Figure 3.69 provides a graphical representation of the MEXTRAM model, which (like the VBIC model and the extended Gummel–Poon model [89]) uses an additional internal collector node to model effects of the low-doped collector region (epilayer).

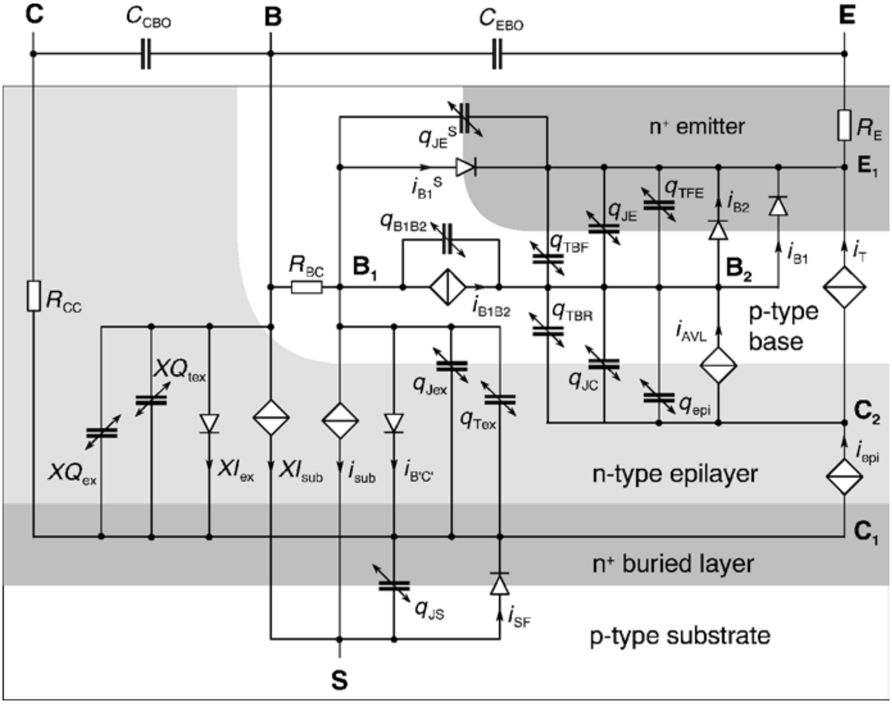


Fig. 3.69. MEXTRAM equivalent circuit of an integrated vertical npn bipolar transistor (after [94]), thermal equivalent circuit for the description of self-heating effects not shown

3.16.1 Transfer Current

In the MEXTRAM model TNS 504 for integrated npn transistors, the transfer current is written in a form that closely resembles that in the Gummel–Poon approach,[68]

$$i_T = \frac{i_{CE} - i_{EC}}{q_B}\,, \tag{3.269}$$

where

$$i_{CE} = I_S\left[\exp\left(\frac{v_{B_2E_1}}{V_T}\right) - 1\right]\,, \quad i_{EC} = I_S\left[\exp\left(\frac{v^*_{B_2C_2}}{V_T}\right) - 1\right]$$

and $q_B = q_1 + q_2$ is the normalized base charge. The voltage $v^*_{B_2C_2}$ may deviate from the internal bc voltage $v_{B_2C_2}$ and is determined so as to be consistent with the epilayer model. The value of q_1 is expressed in terms of integrals over the internal depletion capacitances similarly to the VBIC model:

[68]The notation has been modified in some instances from that used in the documentation of the Level 504 model [94] in order make clear the analogies with other models discussed in this book.

$$q_1 = 1 + \frac{v_{TE}(v_{B_2E_1})}{V_{ER}} + \frac{v_{TC}(v_{B_2C_1}, i_{epi})}{V_{EF}}, \tag{3.270}$$

where the voltage v_{TE} is defined in terms of the internal eb depletion capacitance (3.290) and the voltage v_{TC} is defined in terms of the internal bc depletion capacitance (3.292); two Early voltages are introduced in order to make it possible to adapt the model for the simulation of heterojunction bipolar transistors. The value of q_2, which represents the effects of the stored minority charge, is expressed in the form

$$q_2 = \frac{q_1}{2}\left(\frac{f_1}{1+\sqrt{1+f_1}} + \frac{f_2}{1+\sqrt{1+f_2}}\right), \tag{3.271}$$

where

$$f_1 = \frac{4I_S}{I_K}\exp\left(\frac{v_{B_2E_1}}{V_T}\right) \quad \text{and} \quad f_2 = \frac{4I_S}{I_K}\exp\left(\frac{v^*_{B_2C_2}}{V_T}\right). \tag{3.272}$$

The idea behind this approach is that the diffusion charge in the base region can be represented as the sum of a component q_{TBF} controlled by the eb diode and a component q_{TBR} controlled by the bc diode. The component q_{TBF} is assumed to vary in proportion to (3.123)[69]

$$n(x_{be}) = \frac{2n_{ie}^2\exp\left(\frac{v_{B_2E_1}}{V_T}\right)}{N_A\left[1+\sqrt{1+\frac{4n_{ie}^2}{N_A^2}\exp\left(\frac{v_{B_2E_1}}{V_T}\right)}\right]} = \frac{N_A}{2}\,\frac{f_1}{1+\sqrt{1+f_1}}$$

if we make the identification $n_{ie}^2/N_A^2 = I_S/I_K$. In summary, four parameters are used for the description of the transfer current: the transfer saturation curent `IS`, the forward knee current `IK`, which is used to characterize the onset of high-level-injection effects in the base, and the forward and reverse Early voltages `VEF` and `VER`; these correspond to the parameters `IS`, `IKF`, `VAF` and `VAR` of the SPICE Gummel–Poon model.

3.16.2 Base Current Components, Parasitic PNP Transistor

The ideal forward base current is split up into a bulk component,

$$i_{B1} = (1-X_{IBI})\frac{I_S}{B_F}\left[\exp\left(\frac{v_{B_2E_1}}{V_T}\right) - 1\right], \tag{3.273}$$

and a part that is attributed to the sidewall diode,

$$i_{B1}^S = X_{IBI}\frac{I_S}{B_F}\left[\exp\left(\frac{v_{B_1E_1}}{V_T}\right) - 1\right]; \tag{3.274}$$

[69]The description of q_{TBR} is analogous.

in addition to this, nonideal base current components due to SRH processes in the space charge layer are described by[70]

$$i_{\mathrm{B2}} = I_{\mathrm{BF}} \left[\exp\left(\frac{v_{\mathrm{B_2E_1}}}{M_{\mathrm{LF}} V_{\mathrm{T}}} \right) - 1 \right] + G_{\min} v_{\mathrm{B_2E_1}} \tag{3.275}$$

and[71]

$$i_{\mathrm{B3}} = I_{\mathrm{BR}} \frac{\exp\left(\frac{v_{\mathrm{B_1C_1}}}{V_{\mathrm{T}}} \right) - 1}{\exp\left(\frac{v_{\mathrm{B_1C_1}}}{2V_{\mathrm{T}}} \right) + \exp\left(\frac{V_{\mathrm{LR}}}{2V_{\mathrm{T}}} \right)} + G_{\min} v_{\mathrm{B_1C_1}} \,. \tag{3.276}$$

The parameter `BF` corresponds to the parameter `BF` of the SPICE Gummel–Poon model, while `IBF`, `MLF` and `IBR` correspond to the parameters `ISE`, `NE` and `ISC`. There is no counterpart to the emission coefficient `NC` of the SPICE Gummel–Poon model, which is used to fit the nonideal reverse base current component. The current $i_{\mathrm{B_1C_1}} = i_{\mathrm{B3}} + i_{\mathrm{ex}}$ is composed of i_{B3} and the extrinsic base current i_{ex} due to electrons injected from the collector into the extrinsic base, which is described by

$$i_{\mathrm{ex}} = \frac{I_{\mathrm{S}}}{B_{\mathrm{RI}}} \left[\frac{2\exp\left(\frac{v_{\mathrm{B_1C_1}}}{V_{\mathrm{T}}} \right)}{1 + \sqrt{1 + \frac{4I_{\mathrm{S}}}{I_{\mathrm{K}}} \exp\left(\frac{v_{\mathrm{B_1C_1}}}{V_{\mathrm{T}}} \right)}} - 1 \right]$$

and takes account of high-level-injection effects in the bc diode, a feature not included in the SPICE Gummel–Poon model, VBIC or HICUM. The substrate current is modeled in terms of the pnp transistor saturation current I_{SS} and the substrate knee current (referred to I_{S}) I_{KS}, according to

$$i_{\mathrm{sub}} = \frac{2I_{\mathrm{SS}} \left[\exp\left(\frac{v_{\mathrm{B_1C_1}}}{V_{\mathrm{T}}} \right) - 1 \right]}{1 + \sqrt{1 + 4\frac{I_{\mathrm{S}}}{I_{\mathrm{KS}}} \exp\left(\frac{v_{\mathrm{B_1C_1}}}{V_{\mathrm{T}}} \right)}} \,. \tag{3.277}$$

The currents XI_{ex} and XI_{sub} indicated in Fig. 3.69 are considered only if the option `EXMOD = 1` is chosen,[72] which causes a partitioning of i_{ex} and i_{sub}.

[70] For the purposes of numerical stability, small conductances (not shown in Fig. 3.69) $G_{\min} = 10^{-13}/\Omega$ are connected between nodes B′ and C′ and between B2 and E1.

[71] The formula for i_{B3} is derived from the maximum value of the expression for the net recombination rate, which, under the assumption $\tau_{\mathrm{n0}} = \tau_{\mathrm{p0}} = \tau$, reads [61]

$$R_{\max} = \frac{1}{\tau} \frac{n_{\mathrm{ie}}^2 \left[\exp(V'/V_{\mathrm{T}}) - 1 \right]}{2n_{\mathrm{ie}} \exp(V'/2V_{\mathrm{T}}) + 2n_{\mathrm{ie}} \cosh(\Delta W_{\mathrm{T}}/k_{\mathrm{B}}T)} \,.$$

The formula describes an increase of $i_{\mathrm{B3}} \sim \exp(v_{\mathrm{B_1C_1}}/V_{\mathrm{T}})$ as long as $v_{\mathrm{B_1C_1}} \ll V_{\mathrm{LR}}$; in the limit $v_{\mathrm{B_1C_1}} \gg V_{\mathrm{LR}}$, the proportionality $i_{\mathrm{B3}} \sim \exp(v_{\mathrm{B_1C_1}}/2V_{\mathrm{T}})$ is obtained.

[72] See [94] for the detailed model equations in the case in which this option for extended modeling of the reverse current gain is selected.

The current i_{SF} associated with a forward-biased cs junction is modeled as an ideal diode with saturation current I_{SS} in order to signal a substantial forward bias of this junction to the designer, but without any attempt to accurately describe the current–voltage characteristic.

3.16.3 Epilayer Description

The MEXTRAM epilayer model [91,94,174] includes base widening and quasi-saturation due to an ohmic voltage drop and to the Kirk effect; it represents an improved version of the epilayer model of Kull [89], which is valid only if $I_{\mathrm{T}} < I_1 = eA_{\mathrm{je}}N_{\mathrm{Depi}}v_{\mathrm{nsat}}$ in forward operation. The epilayer model provides the internal bc voltage $v^*_{\mathrm{B_2C_2}}$, which is required for the computation of the transfer current and the bc depletion capacitance; the value of $v^*_{\mathrm{B_2C_2}}$ may deviate from the value of the potential difference $v_{\mathrm{B_2C_2}}$ determined by the circuit simulator in forward operation.

In the first step, the Kull epilayer model is used for the computation of I_{epi}. In reverse operation, the model of Kull is used without modification, and $v^*_{\mathrm{B_2C_2}} = v_{\mathrm{B_2C_2}}$. In the modified formulation for forward operation [90,91], the assumption of quasi-neutrality, which underlies the Kull model, is applied only to the injection region between 0 and x_{i}, where the hole density $p(x)$ is comparable to the electron density. If the adjacent region between x_{i} and d_{epi} is ohmic, the current through this region is[73]

$$I_{\mathrm{epi}} = \frac{1}{R_{\mathrm{CV}}}\frac{d_{\mathrm{epi}}}{d_{\mathrm{epi}}-x_{\mathrm{i}}}\left(V_{\mathrm{C_1B_2}}+V_{\mathrm{B_2C_i}}\right) = I_{\Omega}\ , \tag{3.278}$$

where R_{CV} denotes the ohmic resistance of the total epilayer and $V_{\mathrm{B_2C_i}} = \phi_{\mathrm{pB}} - \phi_{\mathrm{n}}(x_{\mathrm{i}})$. In the Level 504 model, $V_{\mathrm{B_2C_i}}$ is calculated from

$$V_{\mathrm{B_2C_i}} = V_{\mathrm{DC}} + 2V_{\mathrm{T}}\ln\left(\frac{I_{\mathrm{epi}}R_{\mathrm{CV}}}{2V_{\mathrm{T}}}+1\right)\ ,$$

where I_{epi} is calculated from the Kull epilayer model; this relation was found to accurately describe the onset of "hard saturation" [93]. Introducing the new variable[74]

$$V_{\mathrm{qs}} = V_{\mathrm{C_1B_2}} + V_{\mathrm{B_2C_i}}\ ,$$

[73] Since the ohmic region is quasi-neutral, it would in this case be possible to apply the Kull approach to the whole epilayer region, rather than to the region between x_{i} and d_{epi} [91].

[74] The formula implemented in the Level 504 model to calculate V_{qs} is

$$V_{\mathrm{qs}} = \frac{1}{2}\left(V_{\mathrm{qs}}^{\mathrm{th}} + \sqrt{\left(V_{\mathrm{qs}}^{\mathrm{th}}\right)^2 + 0.04\times V_{\mathrm{DC}}^2}\right)\ ,$$

where

$$V_{\mathrm{qs}}^{\mathrm{th}} = V_{\mathrm{C_1B_2}} + V_{\mathrm{DC}} + 2V_{\mathrm{T}}\ln\left(\frac{I_{\mathrm{epi}}R_{\mathrm{CV}}}{2V_{\mathrm{T}}}+1\right)\ ,$$

in order to avoid negative values of $V_{\mathrm{C_1B_2}}+V_{\mathrm{B_2C_i}}$.

(3.278) transforms into

$$I_{\mathrm{epi}} = \frac{V_{\mathrm{qs}}}{R_{\mathrm{CV}} y_{\mathrm{i}}} = I_{\Omega} , \tag{3.279}$$

where $y_{\mathrm{i}} = 1 - x_{\mathrm{i}}/d_{\mathrm{epi}}$. The effects of the field-dependent mobility are not taken into account in I_{Ω}. For $I_{\mathrm{epi}} > I_1 = I_{\mathrm{HC}}$, the region between x_{i} and d_{epi} no longer shows ohmic behavior: the injection region is followed by a space charge region, filled with hot electrons, where the space charge density reads

$$\rho(x) = eN_{\mathrm{Depi}} - \frac{I_{\mathrm{epi}}}{A_{\mathrm{je}} v_{\mathrm{nsat}}} = eN_{\mathrm{Depi}} \left(1 - \frac{I_{\mathrm{epi}}}{I_{\mathrm{HC}}}\right) .$$

A double integration of the Poisson equation from x_{i} to d_{epi} then gives

$$\int_{x_{\mathrm{i}}}^{d_{\mathrm{epi}}} E(x)\,\mathrm{d}x = -V_{\mathrm{qs}} \tag{3.280}$$

$$= -\frac{v_{\mathrm{nsat}}}{\mu_{\mathrm{n0}}}(d_{\mathrm{epi}} - x) + \frac{eN_{\mathrm{Depi}}}{\epsilon}\left(1 - \frac{I_{\mathrm{epi}}}{I_{\mathrm{HC}}}\right)\frac{(d_{\mathrm{epi}} - x_{\mathrm{i}})^2}{2} ,$$

where $E(x_{\mathrm{i}}) = -v_{\mathrm{nsat}}/\mu_{\mathrm{n0}}$ is taken as a boundary condition [91]. Equation (3.280) can be rewritten in the form

$$I_{\mathrm{epi}} = I_{\mathrm{HC}} + \frac{V_{\mathrm{qs}} - I_{\mathrm{HC}} R_{\mathrm{CV}} y_{\mathrm{i}}}{S_{\mathrm{CRCV}} y_{\mathrm{i}}^2} , \tag{3.281}$$

where

$$S_{\mathrm{CRCV}} = \frac{d_{\mathrm{epi}}^2}{2\epsilon v_{\mathrm{nsat}} A_{\mathrm{je}}}$$

denotes the hot-carrier epilayer resistance. To make this result applicable to compact modeling, the thickness of the injection region x_{i} (and, correspondingly, the variable y_{i}) has to be determined and a smooth interpolation between the two limiting cases (3.279) and (3.281) has to be found. A smooth transition between the ohmic and the hot-carrier regime is obtained if I_{HC} in (3.281) is replaced by

$$I_{\mathrm{low}} = \frac{I_{\Omega} I_{\mathrm{HC}}}{I_{\Omega} + I_{\mathrm{HC}}} ,$$

with the result

$$I_{\mathrm{epi}} = \frac{V_{\mathrm{qs}}}{S_{\mathrm{CRCV}} y_{\mathrm{i}}^2} \frac{V_{\mathrm{qs}} + I_{\mathrm{HC}} S_{\mathrm{CRCV}} y_{\mathrm{i}}^2}{V_{\mathrm{qs}} + I_{\mathrm{HC}} R_{\mathrm{CV}} y_{\mathrm{i}}} . \tag{3.282}$$

In the limit $I_{\Omega} \ll I_{\mathrm{HC}}$, this expression correctly reduces to I_{Ω}, whereas in the limit $I_{\Omega} \gg I_{\mathrm{HC}}$ the result (3.281) is obtained. If I_{epi} is determined from the Kull model, it is possible to solve (3.282) for y_{i} in order to obtain the injection width. For numerical reasons, however, I_epi on the left-hand side of (3.282) is replaced by

$$\tilde{I}_\mathrm{epi} = I_\mathrm{qs} \frac{1 + A_\mathrm{XI} \ln \left[1 + \exp\left(\dfrac{I_\mathrm{epi}/I_\mathrm{qs} - 1}{A_\mathrm{XI}}\right)\right]}{1 + A_\mathrm{XI} \ln\left[1 + \exp(-1/A_\mathrm{XI})\right]}, \tag{3.283}$$

where A_XI denotes a model parameter that should be chosen to be around 0.3 for silicon BJTs [93]. I_qs denotes the current at the onset of injection:

$$I_\mathrm{qs} = \frac{V_\mathrm{qs}}{S_\mathrm{CRCV}} \frac{V_\mathrm{qs} + I_\mathrm{HC} S_\mathrm{CRCV}}{V_\mathrm{qs} + I_\mathrm{HC} R_\mathrm{CV}}. \tag{3.284}$$

This is determined from (3.282) by setting $x_\mathrm{i} = 0$ and, correspondingly, $y_\mathrm{i} = 1$. The right-hand side of (3.283) is always larger than I_qs and therefore ensures a nonnegative value of x_i. For the computation of x_i, (3.282) is approximated by the relation

$$\tilde{I}_\mathrm{epi} = \frac{V_\mathrm{qs}}{S_\mathrm{CRCV} y_\mathrm{i}^2} \frac{V_\mathrm{qs} + I_\mathrm{HC} S_\mathrm{CRCV} y_\mathrm{i}}{V_\mathrm{qs} + I_\mathrm{HC} R_\mathrm{CV} y_\mathrm{i}}, \tag{3.285}$$

which is quadratic in y_i and numerically more robust. Using the value of x_i obtained with this procedure, the internal bc voltage $v^*_{\mathrm{B_2C_1}}$ is calculated.

Assuming quasi-neutrality in the injection region, the electron current density equation is given by (3.132), if the hole current density is assumed to be negligible. Since the drift field is comparativey small in the injection region, velocity saturation effects can be neglected; integration of (3.132) across the injection region from 0 to x_i then gives

$$\frac{x_\mathrm{i}}{d_\mathrm{epi}} R_\mathrm{CV} I_\mathrm{epi} = V_\mathrm{T} \left[2 \frac{p(0) - p(x_\mathrm{i})}{N_\mathrm{Depi}} + \ln\left(\frac{p(0)}{p(x_\mathrm{i})}\right)\right].$$

In order to obtain an explicit expression for the normalized hole density $p_0 = p(0)/N_\mathrm{Depi}$, the following approximation is introduced:

$$\frac{x_\mathrm{i}}{d_\mathrm{epi}} R_\mathrm{CV} I_\mathrm{epi} = 2V_\mathrm{T} \left(p_0^* - p_\mathrm{W}\right) \frac{p_0^* + p_\mathrm{W} + 1}{p_0^* + p_\mathrm{W} + 2}, \tag{3.286}$$

where $p_\mathrm{W} = p(d_\mathrm{epi})/N_\mathrm{Depi}$; this approximation was found to produce less than 5% error [93]. With the resulting value of p_0^*, the internal BC voltage $v^*_{\mathrm{B_2C_2}}$ is calculated from

$$p_0^*(p_0^* + 1) = \frac{n_\mathrm{ie}^2}{N_\mathrm{Depi}^2} \exp\left(\frac{v^*_{\mathrm{B_2C_2}}}{V_\mathrm{T}}\right), \tag{3.287}$$

corresponding to

$$v^*_{\mathrm{B_2C_2}} = V_\mathrm{DC} + V_\mathrm{T} \ln\left[p_0^*(p_0^* + 1)\right]$$

with the identification

$$V_{\rm DC} \;=\; V_{\rm T} \ln(N_{\rm Depi}^2/n_{\rm ie}^2) \; . \tag{3.288}$$

It should be emphasized that the definition (3.288) of the bc diffusion voltage differs from the definition of the built-in voltage of the bc diode, which is affected also by the dopant concentration in the base layer. Despite this, $V_{\rm DC}$ is also used to describe the bias dependence of the bc depletion capacitance. This is possible because small deviations of $V_{\rm DC}$ from its optimum value do not severely affect the correctness of the capacitance model. As introduced in the MEXTRAM model, however, this parameter will strongly affect the output conductance in quasi-saturation. Therefore, in MEXTRAM, $V_{\rm DC}$ may not be extracted from capacitance–voltage characteristics as in other models, but is determined from the bias dependence of the current gain [173].

The effects of current spreading have been discussed in [91]; these effects are modeled in the Level 503 model [172]. Since taking account of such effects implies the solution of a third-order equation for $y_{\rm i}$ and since it has been observed, that current-spreading effects are of minor importance [93], current spreading is no longer included in the Level 504 model.

3.16.4 Series Resistances

The external emitter, base and collector series resistances are modeled as (temperature-dependent) ohmic resistors; the bias dependence of the internal base series resistance due to eb current crowding and conductivity modulation is described by a voltage-controlled current source

$$i_{\rm B_1B_2} \;=\; \frac{q_{\rm B}}{3R_{\rm BV}} \left\{ v_{\rm B_1B_2} + 2V_{\rm T} \left[\exp\!\left(\frac{v_{\rm B_1B_2}}{V_{\rm T}} \right) - 1 \right] \right\} , \tag{3.289}$$

where $q_{\rm B}$ is the normalized base charge.

3.16.5 Charge Storage

MEXTRAM Level 504 takes account of constant eb and bc overlap capacitances $C_{\rm BEO}$ and $C_{\rm BCO}$, specified in terms of the parameters `CBEO` and `CBCO`. The eb depletion charge is split up into an internal portion

$$q_{\rm JE} \;=\; (1-X_{\rm CJE}) C_{\rm JE}\, v_{\rm TE}(v_{\rm B_2E_1}) \; ,$$

connected to the nodes $\rm B_2$ and $\rm E_1$, and an external (sidewall) portion

$$q_{\rm JE}^{\rm S} \;=\; X_{\rm CJE} C_{\rm JE}\, v_{\rm TE}(v_{\rm B_1E_1}) \; ,$$

connected to the nodes $\rm B_1$ and $\rm E_1$. The function $v_{\rm TE}(v)$ is defined by

$$v_{\rm TE}(v) \;=\; \frac{V_{\rm DE}}{1-P_{\rm E}} \left[1 - \left(1 - \frac{v_{\rm jE}(v)}{V_{\rm DE}} \right)^{1-P_{\rm E}} \right] + a_{\rm JE}(v - v_{\rm jE}) \; , \tag{3.290}$$

where $v_{\mathrm{JE}}(v) = S(V_{\mathrm{FE}},\, v,\, V_{\mathrm{DE}}/10)$ is the "smoothed" junction voltage defined according to (3.260), and $V_{\mathrm{FE}} = V_{\mathrm{DE}}\left(1-a_{\mathrm{JE}}^{-1/P_{\mathrm{E}}}\right)$; a_{JE} denotes a predefined constant , that is chosen to be equal to 3 for the eb diode. The MEXTRAM parameters CJE, VDE and PE correspond to the SPICE parameters CJE, VJE and MJE; the partition factor XCJE has no counterpart in the conventional Gummel–Poon model. The modified formulation removes the discontinuity in the conventional capacitance–voltage characteristic at $v \to V_{\mathrm{J}}$ and yields a continuously differentiable capacitance–voltage characteristic that is clamped to $a_{\mathrm{JE}}C_{\mathrm{JE}}$ at large values of forward bias. In contrast to earlier versions [61, 172], no attempt is made at a more accurate description of the depletion capacitance at large values of forward bias.

The charge $q_{\mathrm{B_2E_1}}$ associated with the internal eb diode is composed of the charge q_{JE} stored in the internal eb depletion capacitance, the charge stored in the neutral emitter region and the minority charge stored in the base region that is associated with the forward-biased eb diode:

$$\begin{aligned} q_{\mathrm{B_2E_1}}(v_{\mathrm{B_2E_1}}) \;&=\; q_{\mathrm{JE}}(v_{\mathrm{B_2E_1}}) + \tau_{\mathrm{ET}} I_{\mathrm{K}} \left(\frac{I_{\mathrm{S}}}{I_{\mathrm{K}}}\right)^{1/m_\tau} \left[\exp\left(\frac{v_{\mathrm{B_2E_1}}}{m_\tau V_{\mathrm{T}}}\right) - 1\right] \\ &\quad + \frac{2q_1 \tau_{\mathrm{B}} I_{\mathrm{S}} \exp\left(\dfrac{v_{\mathrm{B_2E_1}}}{V_{\mathrm{T}}}\right)}{1+\sqrt{1+\dfrac{4I_{\mathrm{S}}}{I_{\mathrm{K}}}\exp\left(\dfrac{v_{\mathrm{B_2E_1}}}{V_{\mathrm{T}}}\right)}}\,. \end{aligned} \tag{3.291}$$

The parameters TET and MTAU are used here.

In analogy with SPICE, the bc depletion charge is split up into an internal portion q_{JC} and an external portion $q_{\mathrm{JC}}^{\mathrm{ext}}$ by definition of the partition factor X_{CJC}; in MEXTRAM the external portion is further divided into a portion $(1 - X_{\mathrm{EXT}})\, q_{\mathrm{JC}}^{\mathrm{ext}}(v_{\mathrm{B_1C_1}})$ connected to the nodes $\mathrm{B_1}$ and $\mathrm{C_1}$ and a portion $X_{\mathrm{EXT}}\, q_{\mathrm{JC}}^{\mathrm{ext}}(v_{\mathrm{BC_1}})$ connected to the nodes B and $\mathrm{C_1}$ by introduction of the partition factor X_{EXT}. To take account of the finite thickness of the epilayer the additional parameter XP is introduced, which represents the depletion capacitance as a weighted sum of a voltage-dependent and a voltage-independent portion:

$$q_{\mathrm{JC}}(v) \;=\; X_{\mathrm{CJC}} C_{\mathrm{JC}}\, v_{\mathrm{TC}}\,,$$

where

$$v_{\mathrm{TC}} \;=\; \left[(1-X_{\mathrm{P}})\, v_{\mathrm{CV}} + X_{\mathrm{P}} v_{\mathrm{B_2C_1}}\right] \tag{3.292}$$

and

$$v_{\mathrm{CV}} \;=\; \frac{V_{\mathrm{DC}}}{1-P_{\mathrm{C}}}\left[1 - f_{\mathrm{I}}\left(1-\frac{v_{\mathrm{JC}}(v)}{V_{\mathrm{DC}}}\right)^{1-P_{\mathrm{C}}}\right] + f_{\mathrm{I}} b_{\mathrm{JC}}(v-v_{\mathrm{JC}})\,.$$

The parameter b_{JC} is calculated from the predefined constant a_{JC} according to

$$b_{\mathrm{JC}} = \frac{a_{\mathrm{JC}} - X_{\mathrm{P}}}{1 - X_{\mathrm{P}}} \,.$$

The factor

$$f_{\mathrm{I}} = \left(1 - \frac{i_{\mathrm{cap}}}{I_{\mathrm{HC}}}\right)^{M_{\mathrm{C}}}$$

takes account of current modulation of the internal bc depletion capacitance; the current i_{cap} is set equal to I_{epi} in reverse mode ($I_{\mathrm{epi}} \leq 0$), while

$$i_{\mathrm{cap}} = \frac{I_{\mathrm{epi}} I_{\mathrm{HC}}}{I_{\mathrm{epi}} + I_{\mathrm{HC}}} = I_{\mathrm{low}}$$

is used in forward mode ($I_{\mathrm{epi}} > 0$). The current-modulation effect is described in terms of the grading coefficient M_{C}, which may be chosen to be different from the grading coefficient P_{C} used to describe the voltage dependence. In reverse mode ($I_{\mathrm{epi}} \leq 0$), the "smoothed" junction voltage v_{jC} is calculated in analogy with the "smoothed" junction voltage v_{jE} of the eb diode, i.e.

$$v_{\mathrm{jC}} = S(V_{\mathrm{FC}}, v_{\mathrm{B_2C_1}} + v_{\mathrm{C_1C_2}}, V_{\mathrm{DC}}/10) \,,$$

where $V_{\mathrm{FC}} = V_{\mathrm{DC}}\left(1 - b_{\mathrm{JC}}^{-1/P_{\mathrm{C}}}\right)$. In forward mode, v_{jC} is calculated according to

$$v_{\mathrm{jC}} = S(V_{\mathrm{FC}}, v_{\mathrm{B_2C_1}} + V_{\mathrm{XI}}, V_{\mathrm{ch}}) \,, \tag{3.293}$$

where

$$V_{\mathrm{ch}} = \left(0.1 + \frac{2 I_{\mathrm{epi}}}{I_{\mathrm{epi}} + I_{\mathrm{qs}}}\right) V_{\mathrm{DC}}$$

and

$$V_{\mathrm{XI}} = \frac{1}{2} S_{\mathrm{CRCV}} \left(I_{\mathrm{epi}} - I_{\mathrm{HC}} + \sqrt{(I_{\mathrm{epi}} - I_{\mathrm{HC}})^2 + \frac{4 R_{\mathrm{CV}} I_{\mathrm{epi}} I_{\mathrm{HC}}}{S_{\mathrm{CRCV}}}} \right) .$$

The MEXTRAM model parameters `CJC`, `VDC`, `PC` and `XCJC` correspond[75] to the SPICE model parameters `CJC`, `VDC`, `MJC` and `XCJC`; the additional parameters `XEXT`, `XP`, `MC`, `RCV` and `IHC` have been introduced to describe the distributed nature of the external bc depletion capacitance, the effects of an epilayer with finite thickness and the effects of the transfer current on the depletion capacitance, these have no counterpart in the SPICE Gummel–Poon model.

[75] Owing to the importance of the parameter V_{DC} for the correct description of quasi-saturation effects, this parameter is not determined from capacitance–voltage characteristics.

The diffusion charge of the epilayer q_{epi} can be derived by application of the integral charge control relation to the injection region only [93]:

$$q_{\mathrm{epi}} = \tau_{\mathrm{epi}} \frac{2V_{\mathrm{T}}}{R_{\mathrm{CV}}} \frac{x_{\mathrm{i}}}{d_{\mathrm{epi}}} (p_0^* + p_{\mathrm{W}} + 2) ,$$

where the epilayer transit time `TEPI` is a model parameter, and x_{i} and p_0^* are determined from the epilayer model. The value of τ_{epi} can be estimated from

$$\tau_{\mathrm{epi}} \approx \frac{d_{\mathrm{cpi}}^2}{4D_{\mathrm{n}}} \approx I_{\mathrm{S}} Q_{\mathrm{B0}} \left(\frac{R_{\mathrm{CV}}}{2V_{\mathrm{T}}} \right)^2 \exp\left(\frac{V_{\mathrm{DC}}}{V_{\mathrm{T}}} \right)$$

as an initial guess for the purpose of parameter extraction [94].

The charge stored in the cs depletion capacitance is modeled similarly to the eb depletion capacitance, with a predefined constant a_{JS} that is chosen to be equal to 2 for the cs diode. The MEXTRAM parameters `CJS`, `VDS` and `PS` correspond to the SPICE parameters `CJS`, `VJS` and `MJS`.

The excess phase shift is taken approximately into account by base charge partitioning, which redefines the eb and bc diffusion charges according to

$$q_{\mathrm{TE}} \to \frac{2}{3} q_{\mathrm{TE}} \quad \text{and} \quad q_{\mathrm{TC}} \to \frac{1}{3} q_{\mathrm{TE}} + q_{\mathrm{TC}} .$$

A capacitance with stored charge

$$q_{\mathrm{B_1B_2}} = \frac{1}{5} \frac{\partial q_{\mathrm{B_2E_1}}}{\partial v_{\mathrm{B_2E_1}}} v_{\mathrm{B_1B_2}} \tag{3.294}$$

is used to model high-frequency current-crowding effects, in complete analogy with the HICUM model (see also Sect. 3.9). This modeling option requires the flag `EXPHI` to be set to 1; this option is not used if `EXPHI` $= 0$.

3.16.6 Avalanche Effect

The avalanche current $i_{\mathrm{AVL}} \approx I_{\mathrm{cpi}}(M_{\mathrm{n}} - 1)$ is modeled with $M_{\mathrm{n}} - 1$ estimated from the simplified ionization integral:[76]

$$M_{\mathrm{n}} - 1 = \int_{x_{\mathrm{bc}}}^{x_{\mathrm{cb}}} A_{\mathrm{n}} \exp\left(-\frac{B_{\mathrm{n}}}{|E(x)|} \right) \mathrm{d}x .$$

The electric field is approximated around its maximum value according to

[76] A_{n} and B_{n} are predefined constants:

$$A_{\mathrm{n}} = 7.03 \times 10^5 \ \mathrm{cm}^{-1} \quad \text{and} \quad B_{\mathrm{n}} = 1.23 \times 10^6 \ \mathrm{V/cm} .$$

The temperature dependence of B_{n} is modeled according to

$$B_{\mathrm{n}}(T) = B_{\mathrm{n}} \left[1 + 7.2 \times 10^{-4}(T/\mathrm{K} - 300) - 1.6 \times 10^{-6}(T/\mathrm{K} - 300)^2 \right] .$$

$$E(x) \approx E_\mathrm{m}\left(1-\frac{x}{\lambda}\right) \approx \frac{E_\mathrm{M}}{1+x/\lambda},$$

to obtain [175]

$$M_\mathrm{n}-1 = \frac{A_\mathrm{n}}{B_\mathrm{n}} E_\mathrm{m}\lambda \exp\left(-\frac{B_\mathrm{n}}{E_\mathrm{m}}\right)\left[1-\exp\left(-\frac{B_\mathrm{n}W_\mathrm{eff}}{\lambda E_\mathrm{m}}\right)\right]. \tag{3.295}$$

The value of the maximum electric field strength E_m, the effective width of the depletion layer W_eff and the "intersection point" λ, where the linear approximation to $E(x)$ becomes zero, are calculated from the capacitance model of an abrupt junction. Two parameters, WAVL and AVL, are used to calculate these quantities if the flag EXAFL is chosen to be zero (default). If EXAFL is chosen to be one, E_m, λ and W_eff are computed taking account of quasi-saturation and the Kirk effect; in this case the additional parameter SFH is required. If this option is chosen, the output resistance can become negative, resulting in serious convergence problems.

3.16.7 Temperature Effects

The temperature at which the parameters were determined is specified by TREF (in Celsius) in the parameter list, and the actual (ambient) simulation temperature is denoted by TEMP (in Celsius). The device temperature T is increased with respect to this value by the excess temperature due to self-heating. After conversion to kelvin, the device temperature T is calculated from

$$T = (\mathtt{TEMP} + \mathtt{DTA} + \mathtt{VDT} + 273.15)\ \mathrm{K},$$

and the reference temperature is

$$T_0 = (\mathtt{TREF} + 273.15)\ \mathrm{K}.$$

The normalized temperature is defined as $T_\mathrm{N} = T/T_0$. Self-heating is modeled by a one-pole thermal equivalent circuit in complete analogy with the VBIC model, resulting in a voltage $v_\mathrm{dT} = \mathtt{VDT}$ with a value in volts that equals the excess temperature in kelvin. The dissipated power is calculated as the sum of the power dissipated in all elements.

$$\begin{aligned} p_\mathrm{diss} = {} & (v_{\mathrm{B_2E_1}} - v^*_{\mathrm{B_2C_2}})i_\mathrm{T} + (v^*_{\mathrm{B_2C_2}} - v_{\mathrm{B_2C_1}})i_\mathrm{epi} - i_\mathrm{AVL}v^*_{\mathrm{B_2C_2}} \\ & + v^2_{\mathrm{EE_1}}/R_\mathrm{E} + v^2_{\mathrm{CC1}}/R_\mathrm{CC} + v^2_{\mathrm{BB_1}}/R_\mathrm{BC} + i_{\mathrm{B_1B_2}}v_{\mathrm{B_1B_2}} \\ & + (i_\mathrm{B1} + I_\mathrm{B2})v_{\mathrm{B_2E_1}} + I^\mathrm{S}_\mathrm{B1}v_{\mathrm{B_1E_1}} + (i_\mathrm{ex} + i_\mathrm{B3} + i_\mathrm{sub})v_{\mathrm{B_1C_1}} \\ & + (XI_\mathrm{ex} + XI_\mathrm{sub})v_{\mathrm{BC_1}} + (XI_\mathrm{sub} + I_\mathrm{sub} - I_\mathrm{sf})v_{\mathrm{C_1S}}. \end{aligned}$$

The currents XI_{ex} and XI_{sub} define modified expressions for the currents i_{ex} and i_{sub} that are used if the option for extended modeling of the reverse current gain (flag `EXMOD` = 1) is chosen (see [94] for details).

The temperature dependence of the transfer saturation currents is written as[77]

$$I_{\mathrm{S},T} \;=\; I_{\mathrm{S}} T_{\mathrm{N}}^{4-A_{\mathrm{B}}-A_{\mathrm{QB0}}} \exp\left(-\frac{V_{\mathrm{GB}}}{V_{\Delta T}}\right) \;,$$

where $V_{\Delta T} = V_{\mathrm{T}} T_0/(T_0 - T)$. This is equivalent to SPICE if we make the identifications $X_{\mathrm{TI}} = 4 - A_{\mathrm{B}} - A_{\mathrm{QB0}}$ and $V_{\mathrm{GB}} = E_{\mathrm{G}}$. Similar relations, but containing the bandgap voltages V_{GJ} and V_{GC}, are used to describe the saturation currents $I_{\mathrm{BF},T}$ and $I_{\mathrm{BR},T}$. The ideal values of the forward and reverse current gain have the temperature model

$$B_{\mathrm{F},T} \;=\; B_{\mathrm{F}} T_{\mathrm{N}}^{A_{\mathrm{E}}-A_{\mathrm{B}}-A_{\mathrm{QB0}}} \exp\left(-\frac{\Delta V_{\mathrm{GBF}}}{V_{\Delta T}}\right)$$

and

$$B_{\mathrm{R},T} \;=\; B_{\mathrm{R}} \exp\left(-\frac{\Delta V_{\mathrm{GBR}}}{V_{\Delta T}}\right)$$

and introduce the parameters `DVGBF` and `DVGBR` to take account of the effect of different values of the effective energy gap in the emitter, base and collector regions, in order to make the model applicable to HBTs. Temperature-dependent series resistances are modeled by a power law of the form

$$R_{\mathrm{X},T} \;=\; R_{\mathrm{X}} T_{\mathrm{N}}^{A_{\mathrm{X}}} \;,$$

where the exponents A_{E}, $A_{\mathrm{B}} - A_{\mathrm{QB0}}$, A_{EX}, A_{C} and A_{EPI} are used to model the temperature dependence of R_{E}, R_{BV}, R_{BC}, R_{CC} and R_{CV}. The six model parameters `AE`, `AB`, `AQBO`, `AEX`, `AC` and `AEPI` are also used to model the temperature dependence of the transit times, as exemplified by the base transit time

$$\tau_{\mathrm{B},T} \;=\; \tau_{\mathrm{B}} T_{\mathrm{N}}^{A_{\mathrm{QB0}}+A_{\mathrm{B}}-1} \;. \tag{3.296}$$

The eb and cs depletion capacitances are modeled as temperature-dependent quantities according to (3.222); the bc depletion capacitance is divided into a temperature-dependent term and a constant term:

$$C_{\mathrm{JC},T} \;=\; C_{\mathrm{JC}} \left[(1 - X_{\mathrm{P}}) \left(\frac{V_{\mathrm{DC}}}{V_{\mathrm{DC},T}}\right)^{P_{\mathrm{C}}} + X_{\mathrm{P}} \right] \;,$$

[77] Temperature-dependent quantities that have been calculated at the device temperature T are written with an additional subscript T, whereas parameter values at the reference temperature are written without this subscript in accordance with the notation of [94].

where the temperature-dependent built-in voltage is calculated according to (3.221), here the model uses $V_g = V_{GB}$ for the computation of $V_{DE}(T)$, $V_g = V_{GC}$ for the computation of $V_{DC}(T)$ and $V_g = V_{GS}$ for the computation of $V_{DS}(T)$.

3.16.8 Discussion

Presently there are two compact models recommended by the Compact Modeling Council: MEXTRAM and HICUM. Both models have matured for more than a decade and provide an accuracy that should meet most demands. Despite their different appearances (MEXTRAM introduces an extra collector node and employs only 66 parameters in comparison with the 97 parameters used by HICUM), the two models are equivalent to a large extent, as can be seen from the following comparison:

- Both models describe the eb depletion capacitance in essentially the same way, with a division into a bottom and a sidewall portion; however, HICUM uses eight parameters, whereas in MEXTRAM only four parameters are used to describe the voltage dependence of the eb depletion capacitance. This difference stems from the fact that HICUM allows one to define different built-in voltages and grading exponents for the bottom and sidewall portions, whereas in MEXTRAM the separation is performed in terms of the parameter `XCJE`. Furthermore, the maximum capacitance increase a_{JE} is a predefined constant in MEXTRAM, whereas two model parameters (`ALJEI` and `ALJEP`) for the bottom and sidewall portions are introduced to take account of this parameter in HICUM.
- The modeling of the Early effect is equivalent: the parameters `HJEI` and `HJCI` of HICUM do the same thing as the Early voltages `VEF` and `VER` of MEXTRAM. A comparison of the two approaches gives $H_{JEI}C_{JEI0} \equiv Q_{B0}/V_{ER}$ and $H_{JCI}C_{JCI0} \equiv Q_{B0}/V_{EF}$, where $Q_{B0} = \tau_B I_K$ in MEXTRAM.
- Both models take account of ohmic resistances in series with the base, collector and emitter terminals: `RBX`, `RCX` and `RE` in HICUM correspond to `RBX`, `RCX` and `RE` in MEXTRAM. The temperature-exponents `ZETARBX`, `ZETARCX` and `ZETARE` of HICUM correspond to the MEXTRAM parameters `AEX`, `AC` and `AE`.
- Both models describe the parasitic pnp transistor in terms of a simple transistor model; HICUM provides the emission coefficients `MSF`, `MSR` and `MSC` as additional model parameters, and MEXTRAM describes high-level-injection effects in terms of a knee current.
- Both models allow one to define overlap capacitances: the parameters `CEOX` and `CCOX` of HICUM correspond to the parameters `CEBO` and `CCBO` of MEXTRAM.

- AC emitter current crowding is described in both models by a capacitance in parallel with the internal base series resistance.
- The cs capacitance is described in essentially the same way, the parameters `CJS0`, `VDS` and `ZS` of HICUM correspond to the parameters `CJS`, `VDS` and `PS` of MEXTRAM; HICUM allows one to model a punchthrough of the lightly doped substrate region on top of a heavily doped substrate layer with the additional model parameter `VPTS`.
- Both models describe self-heating effects in terms of a one-pole thermal equivalent circuit with identically named parameters `RTH` and `CTH`.
- MEXTRAM uses an additional collector node to model the epilayer in terms of a separate network element, but without making use of the possibilities of partitioning the epilayer charge similarly to Kull's epilayer model. Both MEXTRAM and HICUM assume a constant level of the hole quasi-Fermi potential within the epitaxial collector and are therefore equivalent in this respect.
- The current dependence of the internal bc depletion capacitance is neglected in HICUM for simplicity. This is justified to a large extent because at small current levels, the current dependence of c_{jci} is negligible, whereas at large current levels the effect of c_{jci} on the ac behavior is negligible [162].
- MEXTRAM provides a rather sophisticated model for modeling avalanche effects in the bc junction, an effect that is considered in terms of a rather crude approximation in HICUM.
- Non-quasi-static effects are described by somewhat different approximations: HICUM uses an implementation of te Winkel's delayed charge control model, whereas MEXTRAM uses base charge partitioning.
- HICUM allows one to model reverse-bias tunneling currents in the eb diode, a feature that could be of interest if the eb diode is used as a varactor diode, for example. This explains the two extra parameters `IBETS` and `ABET` in HICUM.
- HICUM provides additional circuit elements, defined by the parameters `RSU` and `CSU`, to treat the substrate contact within the equivalent circuit.

3.17 References

[1] D.K. Lynn, C.S. Meyer, D.J. Hamilton. *Analysis and Design of Integrated Circuits*. McGraw-Hill, New York, 1967.

[2] J.L. Moll, I.M. Ross. The dependence of transistor parameters on the distribution of base layer resistivity. *Proc. IRE*, 44:72–78, 1956.

[3] J.J. Ebers, J.L. Moll. Large-signal behaviour of junction transistors. *Proc. IRE*, 42:1761–1772, 1954.

[4] J.L. Moll. Large-signal transient response of junction transistors. *Proc. IRE*, 42:1773 1784, 1954.

[5] J. Lindmayer, C. Wrigley. Alpha cutoff frequency of junction transistors. *Solid-State Electron.*, 2(5):247–258, 1961.

[6] R.L. Pritchard. *Electrical Characteristics of Transistors.* McGraw-Hill, New York, 1967.

[7] R. Beaufoy, J.J. Sparkes. The junction transistor as a charge-controlled device. *ATE J.*, 13:310–327, 1957.

[8] D.E. Hooper, A.R.T. Turnbull. Applications of the charge-control concept to transistor characterization. *Proc. IRE Australia*, 50(March):132–147, 1962.

[9] J. te Winkel. Past and present of the charge-control concept in the characterization of the bipolar transistor. *Adv. Electron. Electron Phys.*, 39:253–289, 1975.

[10] J. te Winkel. Extended charge-control model for bipolar transistors. *IEEE Trans. Electron Devices*, 20(4):389–394, 1973.

[11] H.K. Gummel. A charge control relation for bipolar transistors. *Bell Syst. Tech. J.*, 49:115–120, 1970.

[12] H.K. Gummel. A self-consistent iterative scheme for one-dimensional steady state transistor calculations. *IEEE Trans. Electron Devices*, 11:455–465, 1964.

[13] H.K. Gummel, H.C. Poon. An integral charge control model of bipolar transistors. *Bell Syst. Tech. J.*, 49:827–852, 1970.

[14] J.J. Liou. Comments on "Early voltage in very-narrow-base bipolar transistors". *IEEE Electron Device Lett.*, 11(5):236, 1990.

[15] S.-G. Lee, R.M. Fox. The effects of carrier-velocity saturation on high-current BJT ouput resistance. *IEEE Trans. Electron Devices*, 39(3):629–633, 1992.

[16] M. Kurata. A small-signal calculation for one-dimensional transistors. *IEEE Trans. Electron Devices*, 18(3):200–210, 1971.

[17] H. Klose, A.W. Wieder. The transient integral charge control relation – a novel formulation of the currents in a bipolar transistor. *IEEE Trans. Electron Devices*, 34(5):1090–1099, 1987.

[18] H.C. de Graaff, J.W. Slotboom, A. Schmitz. The emitter efficiency of bipolar transistors. *Solid-State Electron.*, 20:515–521, 1977.

[19] K. Suzuki. Unified minority-carrier transport equation for polysilicon or heteromaterial emitter contact bipolar transistor. *IEEE Trans. Electron Devices*, 38(8):1868–1877, 1991.

[20] M. Takagi, K. Nakayama, C. Terada, H. Kamioka. Improvement of shallow base transistor technology by using a doped polysilicon diffusion source. *J. Japan Soc. Appl. Phys. (Suppl.)*, 42:101–109, 1973.

[21] H. Murrmann, J. Graul, A. Glasl. High-performance transistors with arsenic-implanted polysil emitters. *IEEE J. Solid-State Circuits*, 11(8):491–495, 1976.

[22] C.M. Maritan, N.G. Tarr. Polysilicon emitter pnp transistors. *IEEE Trans. Electron Devices*, 36(6):1139–1144, 1989.

[23] C.Y. Wong, A.E. Michael, R.D. Isaac, R.H. Kastl, S.R. Mader. The poly–single crystalline silicon interface. *J. Appl. Phys.*, 55(4):1131–1134, 1984.

[24] J.C. Bravman, G.L. Patton, J.D. Plummer. Structure and morphology of polycrystalline silicon–single crystal silicon interfaces. *J. Appl. Phys.*, 57(8):2779–2782, 1985.

[25] K. Sagara, T. Nakamura, Y. Tamaki, T. Shiba. The effect of thin interfacial oxides on the electrical characteristics of silicon bipolar devices. *IEEE Trans. Electron Devices*, 34(11):2286–2290, 1987.

[26] H.C. de Graaff, J.G. de Groot. The SIS tunnel emitter: a theory for emitters with thin interface layers. *IEEE Trans. Electron Devices*, 26(11):1771–1776, 1979.

[27] T.H. Ning, R.D. Isaac. Effect of emitter contact on current gain of silicon bipolar devices. *IEEE Trans. Electron Devices*, 27(11):2051–2055, 1980.

[28] G.L. Patton, J.C. Bravman, J.D. Plummer. Physics, technology, and modeling of polysilicon emitter contacts for VLSI bipolar transistors. *IEEE Trans. Electron Devices*, 33(11):1754–1768, 1986.

[29] D.-L. Chan, D.W. Greeve, A.M. Guzman. Minority-carrier hole diffusion length in heavily doped polysilicon and its influence on polysilicon-emitter transistors. *IEEE Trans. Electron Devices*, 35(7):1045–1054, 1988.

[30] T.F. Meister, K. Ehinger, H. Kabza, C. Fruth, R. Schreiter, M. Biebl. Electrical transport properties in polysilicon emitters investigated by variation of poly-Si thickness. *Proc. IEEE BCTM*, pp.86–89, 1989.

[31] A.A. Eltoukhy, D.J. Roulson. The role of the interfacial layer in polysilicon emitter bipolar transistors. *IEEE Trans. Electron Devices*, 29(12):1862–1869, 1982.

[32] Z. Yu, B. Ricco, R.W. Dutton. A comprehensive analytical and numerical model of polysilicon emitter contacts in bipolar transistors. *IEEE Trans. Electron Devices*, 31(6):773–784, 1984.

[33] I.R.C. Post, P. Ashburn, G.R. Wolstenholme. Polysilicon emitters for bipolar transistors: a review and re-evaluation of theory and experiment. *IEEE Trans. Electron Devices*, 39(7):1717–1731, 1992.

[34] J.D. Cresler, D.D. Tang, K.A. Jenkins, G.-P. Li, E.S. Yang. On the low-temperature static and dynamic properties of high-performance silicon bipolar transistors. *IEEE Trans. Electron Devices*, 36(8):1489–1502, 1989.

[35] M. Pohl, K. Aufinger, J. Böck, T.F. Meister, H. von Philipsborn. DC and AC performance of Si and $Si/Si_{1-x}Ge_x$ bipolar transistors at temperatures up to 300°C. *Proc. ESSDERC*, 28:100–103, 1998.

[36] C.-T. Sah, R.N. Noyce, W. Shockley. Carrier generation and recombination in p–n junctions and p–n junction characteristics. *Proc. IRE*, 45:1228–1243, 1957.

[37] S.C. Choo. Carrier generation–recombination in the space charge region of an asymmetrical pn junction. *Solid-State Electron.*, 11:1069–1077, 1968.

[38] S.C. Choo. On space-charge recombination in pn junctions. *Solid-State Electron.*, 39(2):308–310, 1996.

[39] M. Reisch. Tunneling-induced leakage currents in pn junctions. *Archiv Electronik Übertragungstechnik*, 44(5):368–376, 1992.

[40] J. del Alamo, R.M. Swanson. Forward-bias tunneling limits in scaled bipolar devices. *Extended Abstracts, SSDM, Tokyo*, pp. 283–286, 1986.

[41] G.P. Li, E. Hackbarth, T.-C. Chen. Identification and implication of a perimeter tunneling current component in advanced self-aligned bipolar transistors. *IEEE Trans. Electron Devices*, 35(1):89–95, 1988.

[42] J. Snel. The doped Si/SiO_2 interface. *Solid-State Electron*, 24:135–139, 1981.

[43] J.M.C. Stork, R.D. Isaac. Tunneling in base–emitter junctions. *IEEE Trans. Electron Devices*, 30(11):1527–1534, 1983.

[44] G.A.M. Hurkx. On the modelling of tunneling currents in reverse-biased p–n junctions. *Solid-State Electron.*, 32(8):665–668, 1989.

[45] S.M. Sze. *Physics of Semiconductor Devices*. Wiley, New York, 2nd edition, 1982.

[46] S.L. Miller. Ionization rates for holes and electrons in silicon. *Phys. Rev.*, 105(4):1246–1249, 1957.

[47] R.W. Dutton. Bipolar transistor modeling of avalanche generation for computer circuit simulation. *IEEE Trans. Electron Devices*, 22(6):334–338, 1975.

[48] M. Reisch. On bistable behavior and open-base breakdown of bipolar transistors in the avalanche regime – modeling and applications. *IEEE Trans. Electron Devices*, 39(6):1398–1409, 1992.

[49] H.C. Poon, J.C. Meckwood. Modeling of avalanche effect in integral charge control model. *IEEE Trans. Electron Devices*, 19(1):90–97, 1972.

[50] W.J. Kloosterman, H.C. de Graaff. Avalanche multiplication in a compact bipolar transistor model for circuit simulation. *IEEE Trans. Electron Devices*, 36(7):1376–1380, 1989.

[51] A.S. Grove. *Physics and Technology of Semiconductor Devices.* Wiley, New York, 1st edition, 1967.

[52] J.L. Moll. *Physics of Semiconductors.* McGraw-Hill, New York, 1964.

[53] M. Reisch. *Ladungstrgermultiplikation in selbstjustierten Bipolartransistoren.* Dissertation, TU Wien, 1989.

[54] J. Lohstroh, J.J.M. Koomen, A.T. van Zanten, R.H.W. Salters. Punchthrough-currents in pnp and npn sandwich structures – I. Introduction and basic calculations. *Solid-State Electron.*, 24(9):805–814, 1981.

[55] J. Lohstroh, J.J.M. Koomen, A.T. van Zanten, R.H.W. Salters. Punchthrough-currents in pnp and npn sandwich structures – II. General low-injection theory and measurements. *Solid-State Electron.*, 24(9):815–820, 1981.

[56] S. Esener, S.H. Lee. Punch-through current under diffusion-limited injection: analysis and applications. *J. Appl. Phys.*, 58(3):1380–1387, 1985.

[57] C.T. Chuang, D.D. Tang, G.P. Li, E. Hackbarth. On the punchthrough characteristics of advanced self-aligned bipolar transistors. *IEEE Trans. Electron Devices*, 34(7):1519–1524, 1987.

[58] H.K. Gummel, D.L. Scharfetter. Depletion-layer capacitance of p+n step junctions. *J. Appl. Phys.*, 38(5):2148–2153, 1967.

[59] J.J.H. van den Biesen. P–N junction capacitances, part I: the depletion capacitance. *Philips J. Res.*, 40(2):88–102, 1985.

[60] H.C. Poon, H.K. Gummel. Modeling of emitter capacitance. *Proc. IEEE*, 44(3):200–207, 1990.

[61] H.C. de Graaff, F.M. Klaassen. *Compact Transistor Modeling for Circuit Design.* Springer, Vienna, 1990.

[62] L.J. Varnerin. Stored charge method of transistor base transit analysis. *Proc. IRE*, 47:523–527, 1959.

[63] J.J.H. van den Biesen. A simple regional analysis of transit time in bipolar transistors. *Solid-State Electron.*, 29:529–534, 1986.

[64] J.M. Early. PNIP and NPIN junction transistor triodes. *Bell Syst. Tech. J.*, 33(3):517–533, 1954.

[65] H. Krömer. Two integral relations pertaining to the electron transport through a bipolar transistor with a nonuniform energy gap in the base region. *Solid-State Electron.*, 28(11):1101 1103, 1985.

[66] C.T. Kirk. A theory of transistor cutoff frequency (f_T) falloff at high current densities. *IRE Trans. Electron Devices*, 9(5):164 174, 1962.

[67] R.J. Whittier, D.A. Tremere. Current gain and cutoff frequency falloff at high currents. *IEEE Trans. Electron Devices*, 16(1):39–57, 1969.

[68] K. Suzuki, N. Nakayama. Base transit time of shallow-base bipolar transistors considering velocity saturation at base–collector junction. *IEEE Trans. Electron Devices*, 39(3):623–628, 1992.

[69] N. Rinaldi. Analytical relations for the base transit time and collector transit time in BJTs and HBTs. *Solid-State Electron.*, 41(8):1153–1158, 1997.

[70] G. Baccarani, M.R. Wordeman. An investigation of steady-state velocity overshoot in silicon. *IEEE Trans. Electron Devices*, 28(4):407–416, 1985.

[71] A.A. Grinberg, S. Luryi. Diffusion in a short base. *Solid-State Electron.*, 35(9):1299–1309, 1992.

[72] O. Hansen. Diffusion in a short base. *Solid-State Electron.*, 37(9):1663–1669, 1994.

[73] R. van Overstraeten, H. de Man, R. Mertens. Influence of heavy doping effects on the f_T prediction of transistors. *Electron. Lett.*, 9(8/9):174–176, 1973.

[74] J.A. Kerr, F. Berz. The effect of emitter doping gradient on f_T in microwave bipolar transistors. *IEEE Trans. Electron Devices*, 22(1):15–20, 1975.

[75] J.J.H. van den Biesen. P–N junction capacitances, part i: the neutral capacitance. *Philips J. Res.*, 40(2):103–113, 1985.

[76] J.-S. Park, A. Neugroschel. Current dependence of the emitter resistance of bipolar transistors. *IEEE Trans. Electron Devices*, 37(6):1540–1542, 1990.

[77] J.R. Hauser. The effects of distributed base potential on emitter-current injection density and effective base resistance for stripe transistor geometries. *IEEE Trans. Electron Devices*, 11:238–242, 1964.

[78] H.N. Ghosh. A distributed model of the junction transistor and its application in the prediction of the emitter–base diode characteristic, base impedance, and pulse response of the device. *IEEE Trans. Electron Devices*, 12(10):513–531, 1965.

[79] J.E. Lary, R.L. Anderson. Effective base resistance of bipolar transistors. *IEEE Trans. Electron Devices*, 32(11):2503–2505, 1985.

[80] E.W. Greeneich. Base spreading resistance of polysilicon self-aligned bipolar transistors. *IEEE Trans. Electron Devices*, 36(1):147–149, 1989.

[81] H.-M. Rein, M. Schröter. Base spreading resistance of square-emitter transistors and its dependence on current crowding. IEEE Trans. Electron Devices, 36(4):770–773, 1989.

[82] M. Schröter. Simulation and modeling of the low-frequency base resistance of bipolar transistors and its dependence on current and geometry. *IEEE Trans. Electron Devices*, 38(3):538–544, 1991.

[83] P. Antognetti, G. Massobrio. *Semiconductor Device Modeling with SPICE*. McGraw-Hill, New York, 1989.

[84] H.C. de Graaff, J.A. Pals. On the behaviour of the base–collector junction of a transistor at high collector current densities. *Philips Res. Rep.*, 24:53–69, 1969.

[85] W.M. Webster. On the variation of junction-transistor current-amplification factor with emitter current. *Proc. IRE*, 42:914–920, 1954.

[86] J.R. Beale, J.A. Slatter. The equivalent circuit of a transistor with lightly-doped collector operating in saturation. *Solid-State Electron.*, 11:241–252, 1969.

[87] G. Rey, F. Dupuy, J.P. Bailbe. A unified approach to the base widening mechanisms in bipolar transistors. *Solid-State Electron.*, 18:863–866, 1975.

[88] H.C. de Graaff. High current density effects in the collector of bipolar transistors. In *Process and Device Modeling for Integrated Circuit Design*, F. van de Wiele, W.L. Engl and P.G. Jespers (eds.), Noordhoff International, pp. 419–442, 1977.

[89] S.-W. Lee, P. Lloyd, J. Prendergast, G.M. Kull, L.W. Nagel, H. Dirks. A unified circuit model for bipolar transistors including quasi-saturation effects. *IEEE Trans. Electron Devices*, 32(6):1103–1113, 1985.

[90] H.J. Jeong, J.G. Fossum. Physical modeling of high-current transients for bipolar transistor circuit simulation. *IEEE Trans. Electron Devices*, 34(4):898–905, 1987.

[91] H.C. de Graaff, W.C. Klosterman. Modeling of the collector epilayer of a bipolar transistor in the mextram model. *IEEE Trans. Electron Devices*, 42(2):274–282, 1995.

[92] H.J. Jeong, J.G. Fossum. A charge-based large-signal bipolar transistor model for device and circuit simulation. *IEEE Trans. Electron Devices*, 36(1):124–131, 1989.

[93] J.C.J. Paasschens, W.J. Kloosterman, R.J. Havens, H.C. de Graaff. Improved compact modeling of output conductance and cutoff frequency of bipolar transistors. *IEEE J. Solid-State Circuits*, 36(9):1390–1398, 2001.

[94] J.C.J. Paaschens, W.J. Kloosterman. *The MEXTRAM Bipolar Transistor Model Level 504*. National Laboratory unclassified report NL-UR 2000/811. Internet, http://www.semiconductors.philips.com/Philips_Models/, 2001.

[95] C.A. Desoer, E.S. Kuh. *Basic Circuit Theory*. McGraw-Hill, New York, 1969.

[96] M. Pfost, H.-M. Rein, T. Holzwarth. Modeling substrate effects in the design of high-speed Si-bipolar IC's. *IEEE J. Solid-State Circuits*, 31(10):1493–1591, 1996.

[97] M. Pfost, H.-M. Rein. Modeling and measurement substrate coupling in Si-bipolar IC's up to 40 GHz. *IEEE J. Solid-State Circuits*, 33(4):582–591, 1998.

[98] P.B. Weil, L.P. McNamee. Simulation of excess phase in bipolar transistors. *IEEE Trans. Circuits Syst.*, CAS-25:114–116, 1978.

[99] J. Seitchik, A. Chatterjee, P. Yang. An accurate bipolar model for large signal transient and ac applications. *IEDM Tech. Dig.*, pp. 244–247, 1987.

[100] J.A. Seitchik. Comment on "One-dimensional non-quasi-static models for arbitrarily and heavily doped quasi-neutral layers in bipolar transistors". *IEEE Trans. Electron Devices*, 37(9):2108 2111, 1990.

[101] R.L. Pritchard. Two-dimensional current flow in junction transistors at high frequencies. *Proc. IRE*, 46(June):1152–1160, 1958.

[102] G. Rey. Effets de la defocalisation (c.c. et c.a.) sur le comportement des transistors a jonctions. *Solid-State Electron.*, 12:645–659, 1969.

[103] A.S. Grove. *Physics and Technology of Semiconductor Devices*. Wiley, New York, 1967.

[104] K.G. Ashar. The method of estimating delay in switching circuits and the figure of merit of a switching transistor. *IEEE Trans. Electron Devices*, 11:497–506, 1964.

[105] H.K. Gummel. On the definition of the cutoff frequency f_T. *Proc. IEEE*, 57:2159, 1969.

[106] W.C. Elmore. The transient response of damped linear networks with particular regard to wideband amplifiers. *J. Appl. Phys.*, 19:55–63, 1948.

[107] T.I. Kamins. Effect of polysilicon-emitter shape on dopant diffusion in polysilicon-emitter transistors. *IEEE Electron Device Lett.*, 10(9):401–404, 1989.

[108] K. Kurishima. An analytic expression of $f_{\max}$ for HBT's. *IEEE Trans. Electron Devices*, 43(2):2074 2079, 1996.

[109] M. Vaidyanathan, D.L. Pulfrey. Extrapolated $f_{\max}$ of heterojunction bipolar transistors. *IEEE Trans. Electron Devices*, 46(2):301–309, 1999.

[110] M.B. Das. High-frequency performance limitations of millimeter-wave heterojunction bipolar transistors *IEEE Trans. Electron Devices*, 35(5):604–614, 1988.

[111] M.H. White, M.O. Thurston. Characterization of microwave transistors. *Solid-State Electron.*, 13:523–542, 1970.

[112] D. Gloria, A. Perrotin, J.L. Carbonero, G. Morin. Substrate resistance effect on the $f_{\max}$ parameter of iolated BJT in BICMOS process. *Proc. IEEE Int. Conf. Microelectronic Test Struct.*, 12:24–29, 1999.

[113] K. Aufinger. *Physikalische Modellierung der Rauscheigenschaften von Silizium- und Silizium-Germanium-Höchstfrequenz-Bipolartransistoren.* Dissertation, TU Innsbruck, 2001.

[114] G.W. Taylor, J.G. Simmons. Figure of merit for integrated bipolar transistors. *Solid-State Electron.*, 29(9):941–946, 1986.

[115] G.A.M. Hurkx. The relevance of f_T and $f_{\max}$ for the speed of a bipolar CE amplifier stage. *IEEE Trans. Electron Devices*, 44(5):775–781, 1997.

[116] D.D. Tang, P.M. Solomon. Bipolar transistor design for optimized power-delay logic circuits. *IEEE J. Solid-State Circuits*, 14(4):679–684, 1979.

[117] J.M.C. Stork. Bipolar transistor scaling for minimum switching delay and energy dissipation. *IEDM Tech. Dig.*, pp. 550–553, 1988.

[118] P.K. Tien. Propagation delay in high speed silicon bipolar and GaAs HBT digital circuits. *Int. J. High Speed Electron.*, 1(1):101–124, 1990.

[119] M.Y. Ghannam, R.P. Mertens, R.J. van Overstraeten. An analytical model for the determination of the transient response of CML and ECL gates. *IEEE Trans. Electron Devices*, 37(1):191–201, 1990.

[120] J.M. McGregor, D.J. Roulston, J.S. Hamel, M. Vaidyanathan, S.C. Jain, P. Balk. A simple expression for ECL propagation delay including non-quasi-static effects. *Solid-State Electron.*, 36(3):391–396, 1993.

[121] K.M. Sharaf, M.I. Elmasry. An accurate analytical propagation delay model for high-speed CML bipolar circuits. *IEEE J. Solid-State Circuits*, 29(1):31–45, 1994.

[122] Y. Harada. Delay components of a current mode logic circuit and their current dependency. *IEEE J. Solid-State Circuits*, 30(1):54–60, 1995.

[123] E.W. Greeneich. An appropriate device figure of merit for bipolar CML. *IEEE Electron Device Lett.*, 12(1):18–20, 1991.

[124] E.J. Prinz, J.C. Sturm. Current gain – Early voltage products in heterojunction bipolar transistors with nonuniform base bandgaps. *IEEE Electron Device Lett.*, 12(12):661–663, 1991.

[125] E.O. Johnson. Physical limitations on frequency and power parameters of transistors. *RCA Rev.*, 26:165, 1965.

[126] K.K. Ng, M. Frei, C.A. King. Reevaluation of the $f_T BV_{\mathrm{CEO}}$ limit on Si bipolar transistors. *IEEE Trans. Electron Devices*, 45(8):1854–1855, 1998.

[127] P. Palestri, C. Fiegna, L. Selmi, G.A.M. Hurkx, J.W. Slotboom, E. Sangiorgi. Optimization guidelines for epitaxial collectors of advanced BJT's with improved breakdown voltage and speed *IEDM Tech. Dig.*, pp. 741–744, 1998.

[128] R.M. Fox, S.-G. Lee, D.T. Zweidinger. The effects of BJT self-heating on circuit behavior *IEEE J. Solid-State Circuits*, 28(6):678–685, 1993.

[129] D.L. Harame, D.C. Ahlgren, D.D. Coolbaugh, J.S. Dunn, G.G. Freeman, J.D. Gillis, R.A. Groves, G.N. Hendersen, R.A. Johnson, A.J. Joseph, S. Subbanna, A.M. Victor, K.M. Watson, C.S. Webster, P.J. Zampardi. Current status and future trends of SiGe BICMOS technology. *IEEE Trans. Electron Devices*, 48(11):2575–2594, 2001.

[130] M.H. Norwood, E. Shatz. Voltage variable capacitor tuning: a review. *Proc. IEEE*, 56(5):788–798, 1968.

[131] D.T. Zweidinger, S.G. Lee, R.M. Fox. Compact modeling of BJT self-heating in SPICE. *IEEE Trans. CAD*, 12(9):1368 1375, 1993.

[132] I.E. Getreu. *Modeling the Bipolar Transistor.* Tektronix, Beaverton, 1976.

[133] C.C. McAndrew, L.W. Nagel. Early effect modeling in SPICE. *IEEE J. Solid-State Circuits*, 31(1):136–138, 1996.

[134] D.D. Tang, T.H. Ning. Method for determining the emitter and base series resistances of bipolar tranistors. *IEEE Trans. Electron Devices*, 31(4):409–412, 1984.

[135] R.C. Taft, J.D. Plummer. An eight-terminal Kelvin-tapped bipolar transistor for extracting parasitic series resistances. *IEEE Trans. Electron Devices*, 38(9):2139 2154, 1991.

[136] J. Weng, J. Holz, T.F. Meister. New method to determine the base resistance of bipolar transistors. *IEEE Electron Device Lett.*, 13:158–160, 1992.

[137] E. Dubois, P.H. Bricout, E. Robilliart. Extraction method of the base series resistance in bipolar transistors in presence of current crowding. *IEEE J. Solid-State Circuits*, 31(1):132 135, 1996.

[138] M. Linder, F. Ingvarson, K.O. Jeppson, J.V. Grahn, S.-L. Zhang, M. Östling. A new procedure for extraction of series resistances for bipolar transistors from dc measurements. *Proc. IEEE Int. Conf. Microelectron. Test Struct.*, 12:147–151, 1999.

[139] B. Kulke, S.L. Miller. Accurate measurement of emitter and collector series resistances in transistors. *Proc. IRE*, 45:90, 1957.

[140] L.J. Giacoletto. Measurement of emitter and collector series resistances. *IEEE Trans. Electron Devices*, 19:692–693, 1972.

[141] R. Gabl, M. Reisch. Emitter series resistance from open-collector measurements – influence of the collector region and the parasitic pnp transistor. *IEEE Trans. Electron Devices*, 45(12):2457–2465, 1998.

[142] M. Reisch. Carrier multiplication and avalanche breakdown in self-aligned bipolar transistors. *Solid-State Electron.*, 33(2):189–197, 1990.

[143] M. Reisch. Self-heating in BJT circuit parameter extraction. *Solid-State Electron.*, 35(5):677–679, 1992.

[144] H. Tran, M. Schröter, D.J. Walkey, D. Marchesan, T.J. Smy. Simultaneous extraction and emitter series resistances in bipolar transistors. *Proc. IEEE BCTM*, pp. 170–173, 1997.

[145] D.T. Zweidinger, R.M. Fox, J.S. Brodsky, T. Jong, S.-G. Lee. Thermal impedance extraction for bipolar transistors. *IEEE Trans. Electron Devices*, 43(2):342–346, 1996.

[146] Hewlett Packard. *S-Parameter design.* Application Note 154, Hewlett Packard 1972.

[147] P.J. van Wijnen, H.R. Claessen, E.A. Wolsheimer. A new straightforward calibration and correction procedure for "on wafer" high-frequency s-parameter measurements (45 MHz – 18 GHz). *Proc. IEEE BCTM*, pp. 70–73, 1987.

[148] M.C.A.M. Koolen, J.A.M. Geelen, M.P.J.G. Versleijen. An improved de-embedding technique for on-wafer high-frequency characterization. *Proc. IEEE BCTM*, pp. 188–191, 1991.

[149] D. Costa, W.U. Liu, J.S. Harris. Direct extraction of the AlGaAs/GaAs heterojunction bipolar transistor small-signal equivalent circuit. *IEEE Trans. Electron Devices*, 38(9):2018–2024, 1991.

[150] S. Lee, B.R. Ryum, S.W. Kang. A new parameter extraction technique for small-signal equivalent circuit of polysilicon emitter bipolar transistors. *IEEE Trans. Electron Devices*, 41(2):233–238, 1994.

[151] D.P. Kennedy, P.C. Murley, W. Kleinfelder. On the measurement of impurity atom distributions in silicon by the differential capacitance technique. *IBM J. Res. Dev.*, 12:399–409, 1968.

[152] D.P. Kennedy, R.R. O'Brien. On the measurement of impurity atom distributions in silicon by the differential capacitance technique. *IBM J. Res. Dev.*, 13:212–214, 1969.

[153] W.T. Johnson, P.T. Panousis. The influence of Debye length on the C V measurement of doping profiles. *IEEE Trans. Electron Devices*, 18:965–973, 1971.

[154] W.E. Carter, H.K. Gummel, B.R. Chawla. Interpretation of capacitance vs. voltage measurements of pn junctions. *Solid-State Electron.*, 15:195–201, 1972.

[155] C.P. Wu, E.C. Douglas, C.W. Mueller. Limitations of the CV technique for ion-implanted profiles. *IEEE Trans. Electron Devices*, 22:319–329, 1975.

[156] W.M.C. Sansen, R.G. Meyer. Characterization and measurement of the base and emitter resistances of bipolar transistors. *IEEE J. Solid-State Circuits*, 7(6):492–498, 1972.

[157] T. Fuse, Y. Sasaki. An analysis of small-signal and large-signal base resistances for submicrometer BJTs. *IEEE Trans. Electron Devices*, 42(3):534 539, 1995.

[158] A. Neugroschel. Measurement of the low-current base and emitter resistances of bipolar transistors. *IEEE Trans. Electron Devices*, 34(4):817–822, 1987.

[159] C.C. McAndrew, J.A. Seitchik, D.F. Bowers, M. Dunn, M. Foisy, I. Getreu, M. McSwain, S. Moinian, J. Parker, D.J. Roulston, M. Schröter, P. van Wijnen, L.F. Wagner. VBIC95, the vertical bipolar inter-company model. *IEEE J. Solid-State Circuits*, 31(10):1476–1483, 1996.

[160] H.-M. Rein, H. Stübing. A compact physical large-signal model for high-speed bipolar transistors at high current densities part i: one-dimensional model. *IEEE Trans. Electron Devices*, 34(8):1741–1751, 1987.

[161] H.-M. Rein, M. Schröter. A compact physical large-signal model for high-speed bipolar transistors at high current densities – part ii: two-dimensional model and experimental results. *IEEE Trans. Electron Devices*, 34(8):1752–1761, 1987.

[162] M. Schröter. *HICUM – A Scalable Physics-Based Compact Bipolar Transistor Model, Description of Model Version 2.1.* Internet, http://www.iee.tu.dresden.de/schroter/Models/hicman.pdf, 2001.

[163] M. Schröter, H.-M. Rein, W. Rabe, R. Reimann, H.-J. Wassener, A. Koldehoff. Physics- and process-based bipolar transistor modeling for integrated circuit design. *IEEE J. Solid-State Circuits*, 34(8):1136–1149, 1999.

[164] M. Schröter. *TRADICA – A Program for Sizing and Model Parameter Generation of Integrated Bipolar Transistors*, Manual Version 4.5, June 2000. Internet, http://www.iee.tu.dresden.de/schroter/Models/hicman.pdf, 2000.

[165] M. Schröter, T.Y. Lee , Z. Yan, W. Shi. A compact tunneling current and collector breakdown model. *Proc. IEEE BCTM*, pp. 203–206, 1998.

[166] H.-M. Rein, H. Stübing, M. Schröter. Verification of the integral charge-control relation for high-speed bipolar transistors at high current densities. *IEEE Trans. Electron Devices*, 32(6):1070 1076, 1985.

[167] M. Schröter, M. Friedrich, H.-M. Rein. A generalized integral charge-control relation and its application to compact models for silicon-based HBTs. *IEEE Trans. Electron Devices*, 40(11):2036–2046, 1993.

[168] M. Schröter. Simulation and modeling of the low-frequency base resistance of bipolar transistors and its dependence on current and geometry. *IEEE Trans. Electron Devices*, 38(3):538–544, 1991.

[169] D.L. Bowler, F.A. Lindholm. High current regimes in transistor collector regions. *IEEE Trans. Electron Devices*, 20(3):257–263, 1973.

[170] M. Schröter, T.-Y. Lee. Physics-based minority charge and transit time modeling for bipolar transistors. *IEEE Trans. Electron Devices*, 46(2):288–300, 1999.

[171] M. Schröter, H.-M. Rein. Investigation of very fast and high-current transients in digital bipolar IC's using both a new compact model and a device simulator. *IEEE J. Solid-State Circuits*, 30(5):551–562, 1995.

[172] Philips. *Bipolar NPN Transistors TN/TNS Level 503*. Internet, http://www.semiconductors.philips.com/Philips_Models/, 2000.

[173] J.C.J. Paaschens, W.J. Kloosterman, R.J. Havens. *Parameter Extraction for the Bipolar Transistor Model Mextram Level 504*, National Laboratory unclassified report NL-UR 2001/801. Internet, http://www.semiconductors.philips.com/Philips_Models/, 2001.

[174] J.C.J. Paaschens, W.J. Kloosterman, R.J. Havens, H.C. de Graaff. Improved compact modeling of output conductance and cutoff frequency of bipolar transistors. *IEEE J. Solid-State Circuits*, 36(9):1390–1398, 2001.

[175] W.J. Kloosterman, J.C.J. Paasschens, R.J. Havens, A comprehensive bipolar avalanche multiplication compact model for circuit simulation. *Proc. IEEE BCTM*, pp. 172–175, 2000.

4 Physics and Modeling of Heterojunction Bipolar Transistors

Heterojunctions in bipolar transistors allow improved values of current gain, base resistance, base transit time and Early voltage. If the bandgap in the

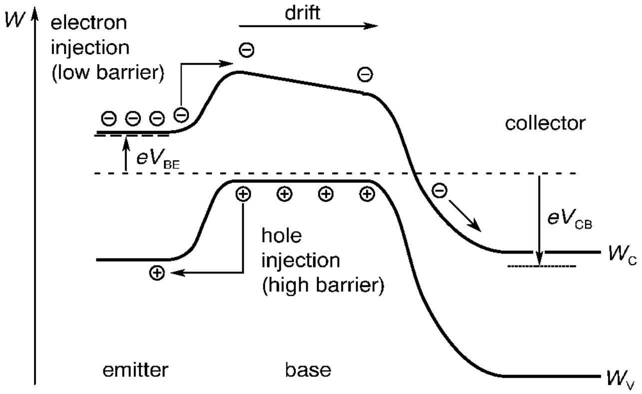


Fig. 4.1. One-dimensional band scheme of heterostructure bipolar transistor with graded base

emitter region exceeds the bandgap in the base region ($W_{gE} > W_{gB}$), fewer carriers will be injected into the emitter at a given transfer current density in comparison with a conventional transistor [1, 2]: the current gain increases according to

$$B_N \sim \exp\left(\frac{W_{gE} - W_{gB}}{k_B T}\right) ,$$

as long as recombination in the base layer is negligible. Its value generally decreases with increasing temperature – in contrast[1] to homojunction BJTs, where bandgap narrowing in the heavily doped emitter causes the bandgap W_{gE} to be smaller than the bandgap W_{gB} in the base region. Wide-gap emitters allow one to dope the base region more heavily than the emitter region, without losing too much current gain. This is of great importance for the realization of high-frequency bipolar transistors, since the extremely thin base widths of these devices require large doping concentrations in order to obtain reasonable values of base resistance and punchthrough voltage.

Gradual changes of alloy composition allow one to realize continuous changes of the bandgap (bandgap grading). If, in particular, the bandgap

[1] A temperature-dependent value of the effective surface recombination velocity used to characterize the polycrystalline emitter contact may also result in a decrease in the current gain of a conventional BJT with increasing temperature [3].

in the base region decreases towards the collector, the associated slope of the conduction band edge (see Fig. 4.1) acts like an additional field that aids in drifting the carriers through the base region. In such devices, the base transit time can therefore be considerably reduced. Furthermore, increased values of the Early voltage can be achieved with a graded base [4]. The doping concentration in the base region is chosen to be large and, preferably, homogeneous, since this allows one to minimize the base transit time without compromising the base series resistance and punchthrough voltage too much.

If the bandgap in the collector region W_{gC} exceeds the bandgap W_{gB} in the base region, the injection of holes into the collector in saturation or quasi-saturation is reduced. This causes holes to pile up at the bc junction [4, 5], resulting in an electric field that slows down the motion of the electrons and therefore increases the transit time. To reduce this effect, the heterojunction is generally shifted into the space charge region of the collector junction [6]. Since impact ionization is less probable in semiconductor materials with a large bandgap, wide-bandgap collectors generally will show an increased value of breakdown voltage.

HBTs with a SiGe base region can be fabricated on a silicon substrate using ultrahigh-vacuum chemical vapor deposition (UHVCVD) or MBE. HBTs made by this technique are compatible with the highly advanced silicon process technology and therefore allow cost-effective circuit integration. Forming a base region in which the number of Ge atoms that substitute for atoms of the silicon lattice increases towards the bc junction yields an HBT with a graded base. In such a transistor, the base current that flows for a given voltage V_{BE} is not decreased in comparison with the base current in a conventional bipolar transistor. Owing to the reduced bandgap in the base region, however, the transfer current increases: bipolar transistors with a SiGe base layer need a reduced bias voltage V_{BE} in order to carry a given transfer current. HBTs with SiGe base regions have been fabricated as laboratory samples with cutoff frequencies exceeding 300 GHz.

Heterostructures can be fabricated without lattice mismatch in III–V semiconductors and allow one to realize comparatively large differences in the bandgap. A typical system is an AlGaAs/GaAs heterostructure on a GaAs substrate material. Since undoped GaAs is semi-insulating (with a specific resistivity of approximately 10^8 Ωcm at room temperature), the collector–substrate capacitance is small. Another example is that of InGaAs/InP HBTs. These benefit from the very large electron mobility and small surface recombination velocity of InGaAs. The bandgap of $In_{0.53}Ga_{0.47}As$, which is typically chosen as the base material, is 0.75 eV, i.e. such transistors require a small forward bias.

4.1 Heterojunctions

Heterojunctions are formed at the interfaces between different semiconductor materials. The different band structures of the adjacent semiconductor materials result in a potential gradient that affects the motion of carriers crossing the heterojunction. In the following, ideal heterojunctions, without surface states at the interface between the two semiconductor materials, are considered. In general, this means that the atomic bonding of the component semiconductors must be maintained without interruption across the heterojunction. This condition, in turn, implies that both the atomic arrangement and the atomic spacing of the two semiconductors must be essentially identical. If this were not the case, misbonded or incompletely bonded interfacial atoms would contribute interfacial electronic states in the same way as unpassivated silicon atoms contribute surface states at the Si–SiO_2 interface [7].

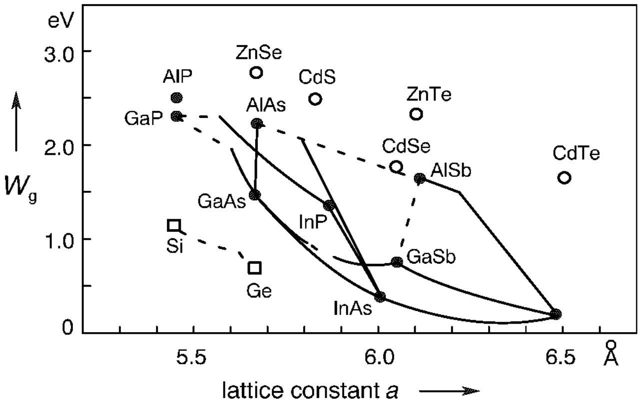


Fig. 4.2. Plot of semiconductor bandgap versus lattice constant a for some common semiconductors. The (solid and broken lines) lines that connect data points for pure semiconductors describe alloys of those semiconductors, and solid lines denote a direct bandgap (after [7])

Figure 4.2 shows the energy gap of some common semiconductor materials plotted versus the lattice constant. An ideal heterojunction is composed of two semiconductor materials with the same lattice constant, such as $Ga_xAl_{1-x}As$ alloys. Heterojunctions could in theory be formed on silicon with GaP, which has approximately the same lattice constant; owing to problems with doping and with the growth of homogeneous layers, however, this possibility is not used in practice [7].

The realization of useful heterostructures on silicon substrates was made possible by the ability to form strained epitaxial layers of Ge_xSi_{1-x} on Si that were thin enough to avoid the formation of misfit dislocations. Before this material system is considered in more detail in Sect. 4.2, a general discussion of heterojunctions between arbitrary semiconducting materials and of heterostructure bipolar transistors will be given.

4.1.1 Thermal Equilibrium

Within the regional approach, the device is subdivided into two quasi-neutral regions – which are taken here to be homogeneously doped – separated by a space-charge layer. In the quasi-neutral regions, the results obtained for a homojunction apply without alteration except that different material parameters have to be used on either side of the junction. The presence of the heterojunction, however, requires a modification of the Shockley boundary conditions. To obtain appropriate boundary conditions, we consider the system depicted in Fig. 4.3 before junction formation.

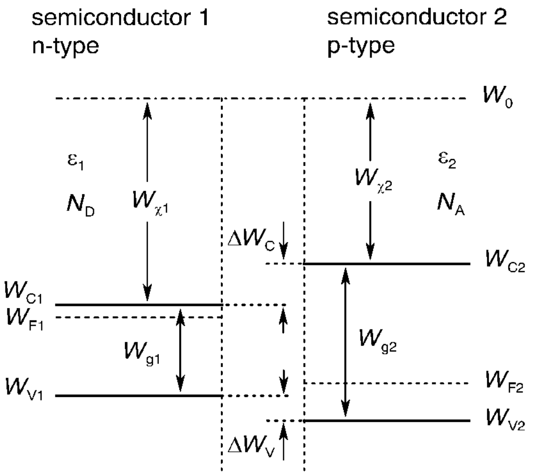


Fig. 4.3. One-dimensional band scheme of heterojunction before junction formation. W_0 denotes the energy of an electron that is able to leave the material

Assuming both sides to be nondegenerate, the equilibrium electron densities on the two sides of the junction are approximately given by

$$n_{n0} \approx N_{C1} \exp\left(\frac{W_{F1}-W_{C1}}{k_B T}\right) \quad \text{and} \quad n_{p0} \approx N_{C2} \exp\left(\frac{W_{F2}-W_{C2}}{k_B T}\right) .$$

Since in thermal equilibrium the Fermi energy is constant across the junction, the condition $W_{F1} = W_{F2}$ must be fulfilled after junction formation (Fig. 4.4). Using

$$W_{C2} - W_{C1} = eV_J + \Delta W_C , \tag{4.1}$$

where the conduction band discontinuity

$$\Delta W_C = W_{\chi1} - W_{\chi2} \tag{4.2}$$

is determined by the values of the electron affinity in the two semiconductors, the relation

$$n_{p0} = n_{n0} \frac{N_{C2}}{N_{C1}} \exp\left(-\frac{V_J}{V_T}\right) \exp\left(-\frac{\Delta W_C}{k_B T}\right) = A_n n_{n0} \tag{4.3}$$

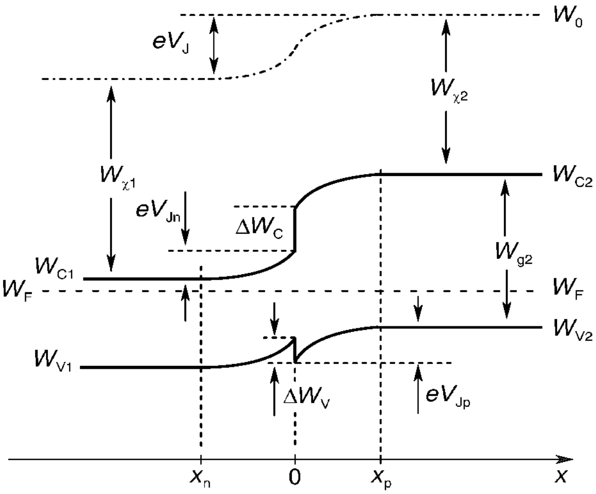


Fig. 4.4. One-dimensional band scheme of heterojunction after junction formation ($\Delta W_C > 0$, $\Delta W_V < 0$)

between the thermal-equilibrium electron densities on either side of the junction is obtained. In complete analogy, using

$$W_{V2} - W_{V1} \;=\; eV_J + W_{V2}(0^+) - W_{V1}(0^-) \;=\; eV_J + \Delta W_V$$

together with

$$p_{n0} \;\approx\; N_{V1} \exp\left(\frac{W_{V1} - W_{F1}}{k_B T}\right) \quad \text{and} \quad p_{p0} \;\approx\; N_{V2} \exp\left(\frac{W_{V2} - W_{F2}}{k_B T}\right) ,$$

the following relation between the thermal-equilibrium hole densities on the two sides is obtained:

$$p_{n0} \;=\; p_{p0} \frac{N_{V1}}{N_{V2}} \exp\left(-\frac{V_J}{V_T}\right) \exp\left(-\frac{\Delta W_V}{k_B T}\right) \;=\; \Lambda_p p_{p0} \,. \tag{4.4}$$

If we take account of the fact that

$$n_{ie}^2(x_n) \;=\; N_{C1} N_{V1} \exp\left(-\frac{W_{g1}}{k_B T}\right) \tag{4.5}$$

and

$$n_{ie}^2(x_p) \;=\; N_{C2} N_{V2} \exp\left(-\frac{W_{g2}}{k_B T}\right) , \tag{4.6}$$

the equations for Λ_n and Λ_p transform into

$$\Lambda_n \;=\; \frac{n_{ie}^2(x_p)}{n_{n0} p_{p0}} \quad \text{and} \quad \Lambda_p \;=\; \frac{n_{ie}^2(x_n)}{n_{n0} p_{p0}} , \tag{4.7}$$

where $n_{n0} \approx N_D$ and $p_{p0} \approx N_A$. The built-in voltage V_J can be determined from Fig. 4.4 to be

$$V_{\rm J} = \frac{W_{\rm g2} - \Delta W_{\rm C}}{e} - \frac{W_{\rm C1} - W_{\rm F}}{e} - \frac{W_{\rm F} - W_{\rm V2}}{e}$$
$$= \frac{W_{\rm g1} - \Delta W_{\rm V}}{e} - \frac{W_{\rm C1} - W_{\rm F}}{e} - \frac{W_{\rm F} - W_{\rm V2}}{e} ,$$

with the values of $W_{\rm C1} - W_{\rm F}$ and $W_{\rm F} - W_{\rm V2}$ determined in analogy to the homojunction case, i.e.

$$W_{\rm C1} - W_{\rm F} = k_{\rm B}T \ln(N_{\rm C1}/n_{\rm n0}) \approx k_{\rm B}T \ln(N_{\rm C1}/N_{\rm D}) , \tag{4.8}$$

$$W_{\rm F} - W_{\rm V2} = k_{\rm B}T \ln(N_{\rm V2}/p_{\rm p0}) \approx k_{\rm B}T \ln(N_{\rm V2}/N_{\rm A}) , \tag{4.9}$$

in the nondegenerate case. The built-in voltage can be separated into a part

$$V_{\rm Jn} = \frac{N_{\rm A}\epsilon_2}{N_{\rm D}\epsilon_1 + N_{\rm A}\epsilon_2} V_{\rm J} = kV_{\rm J} \tag{4.10}$$

on the n-type side, and a part

$$V_{\rm Jp} = \frac{N_{\rm D}\epsilon_1}{N_{\rm D}\epsilon_1 + N_{\rm A}\epsilon_2} V_{\rm J} = (1-k)V_{\rm J} \tag{4.11}$$

on the p-type side of the junction.

4.1.2 Forward-Biased Heterojunction

In an ideal[2] heterojunction, the current under forward bias is caused by minority-carrier injection into the neutral regions and can be computed in analogy to the homojunction case. However, the boundary condition for the minority-carrier density at the depletion layer edge has to be modified if thermal emission across an abrupt heterojunction limits the current of injected minority carriers: at an abrupt heterojunction, the quasi-Fermi potential has a discontinuity,[3] which can be taken into account by applying a modified form of the mass-action law at the depletion layer edge.

Elementary Thermal-Emission Theory

Figure 4.5 shows the discontinuous conduction band edge (spike-and-notch band scheme [2]) of an abrupt heterojunction with an applied forward bias V'. The thermionic emission current for electrons passing from the n-type to the p-type region has to surmount a potential barrier of height

[2] The recombination current due to Shockley–Read–Hall processes in the space charge region of a heterojunction has been investigated in [8]. An increase of the recombination current was found in simulations of graded heterojunctions owing to the increase of intrinsic carrier density associated with the reduction of the energy gap.

[3] In [9] the electron quasi-Fermi-level splitting has been computed for AlGaAs/GaAs HBTs, and differences in the electron quasi-Fermi level of up to $V_{\rm BE}/10$ were found at large forward bias in single-heterojunction HBTs; substantially smaller values were obtained for double-heterojunction HBT's.

$$eV_{\rm n} = ek(V_{\rm J} - V') ,$$

where k is defined in (4.10), resulting in a current density

$$J_{\rm np} = -A_1^* T^2 \exp\left(-\frac{V_{\rm n}}{V_{\rm T}}\right) ,$$

where A_1^* denotes the Richardson constant of semiconductor 1. The thermionic

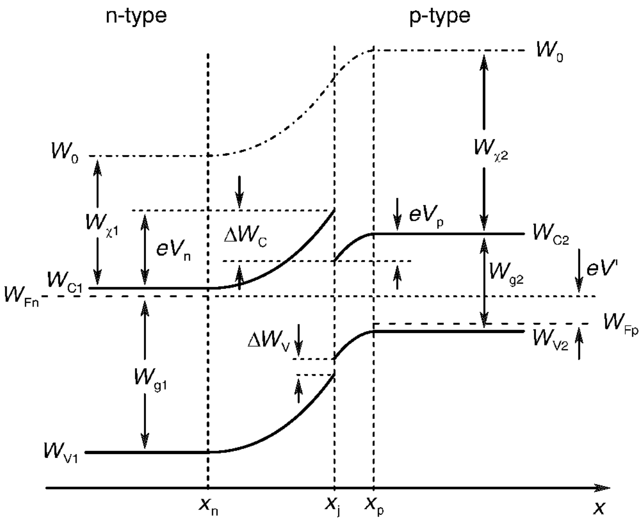


Fig. 4.5. One-dimensional band scheme of an abrupt heterojunction between an n-type wide-gap semiconductor and a p-type semiconductor. This situation is found in an npn HBT with a wide-gap emitter

emission current from the p-type to the n-type side has to surmount a potential barrier of height

$$\Delta W_{\rm C} - eV_{\rm p} = \Delta W_{\rm C} - e(1-k)(V_{\rm J} - V') ,$$

resulting in

$$J_{\rm pn} = -A_2^* T^2 \exp\left(-\frac{\Delta W_{\rm C}/e - V_{\rm p}}{V_{\rm T}}\right) ,$$

where A_2^* denotes the Richardson constant of semiconductor 2. As the two currents cancel in thermal equilibrium ($V' = 0$), the condition

$$\frac{A_2^*}{A_1^*} = \exp\left(\frac{\Delta W_{\rm C}/e - V_{\rm J}}{V_{\rm T}}\right)$$

has to be fulfilled between the Richardson constants on either side of the junction. If the junction is biased, the two currents will no longer cancel, and a net current density $J = J_{\rm np} - J_{\rm pn}$ results, which can be expressed as

$$\begin{aligned} J &= -A_1^* T^2 \left[\exp\left(-\frac{V_\mathrm{n}}{V_\mathrm{T}}\right) - \frac{A_2^*}{A_1^*} \exp\left(-\frac{\Delta W_\mathrm{C}/e - V_\mathrm{p}}{V_\mathrm{T}}\right)\right] \\ &= -A_1^* T^2 \exp\left(-\frac{kV_\mathrm{J}}{V_\mathrm{T}}\right) \left[\exp\left(\frac{kV'}{V_\mathrm{T}}\right) - \exp\left(\frac{(k-1)V'}{V_\mathrm{T}}\right)\right] . \end{aligned} \quad (4.12)$$

The value of k depends on the doping densities on both sides of the junction. If $N_\mathrm{D} \ll N_\mathrm{A}$, the value $k \approx 1$ is obtained, and

$$J_\mathrm{n} \approx -A_1^* T^2 \exp\left(-\frac{V_\mathrm{J}}{V_\mathrm{T}}\right) \left[\exp\left(\frac{V'}{V_\mathrm{T}}\right) - 1\right] . \quad (4.13)$$

Under these conditions there is no notch on the p-type side ($eV_\mathrm{p} \approx 0$), and an exponential current–voltage characteristic with an emission coefficient $N = 1$ results. This situation is typical of the eb junction of an idealized HBT with a wide-bandgap emitter and a heavily doped base[4].

An advantage associated with the conduction band spike is that such a barrier injects the electrons into the base region with a substantial kinetic energy and hence a very high velocity [2, 10, 11]. A thin base region will therefore be crossed at a velocity in excess of the saturated drift velocity (quasi-ballistic transport), an effect that substantially reduces the base transit time.

The thermionic-emission model certainly provides an oversimplified treatment of the current carried by a forward-biased heterojunction: if, for example, the p-type region serves as the base of an HBT, no effects of the base width appear in the electron transfer current (4.13) – in contrast to what is expected. For an adequate description of heterojunctions within the drift–diffusion approximation, boundary conditions for the minority-carrier density at x_n and x_p are required. As will be shown in the following subsection, the quasi-Fermi potentials cannot be assumed to be constant across an abrupt current-carrying heterojunction.

Electron Quasi-Fermi-Level Splitting

Graded-Layer Heterojunctions. In a graded-layer heterojunction, the electron density n and electron quasi-Fermi energy $W_\mathrm{Fn} = -e\phi_\mathrm{n}$, which are related according to[5]

[4] If the doping concentration on the p-type side is of the same order of magnitude as on the n-type side, k is smaller than one and an exponential current–voltage characteristic with an emission coefficient $N = 1/k$ results. In this case a notch appears at the junction x_j, which collects injected electrons and therefore enhances recombination losses. The notch can be removed by incorporating a very thin sheet with a very high acceptor concentration right at the interface [2].

[5] In the case of a degenerate semiconductor N_C has to be replaced by $\gamma_\mathrm{n} N_\mathrm{C}$, where γ_n denotes the degeneracy factor (see Sect. 2.2.1).

$$n(x) = N_{\mathrm{C}}(x)\exp\left(\frac{W_{\mathrm{Fn}}(x) - W_{\mathrm{C}}(x)}{k_{\mathrm{B}}T}\right) , \tag{4.14}$$

vary smoothly throughout the depletion layer; the density

$$J_{\mathrm{n}} = \mu_{\mathrm{n}} n \frac{\mathrm{d}W_{\mathrm{Fn}}}{\mathrm{d}x}$$

of the electron current passing through the depletion layer may therefore be written in the form

$$J_{\mathrm{n}} = k_{\mathrm{B}}T\mu_{\mathrm{n}}(x)N_{\mathrm{C}}(x)\exp\left(-\frac{W_{\mathrm{C}}(x)}{k_{\mathrm{B}}T}\right)\frac{\mathrm{d}}{\mathrm{d}x}\exp\left(\frac{W_{\mathrm{Fn}}(x)}{k_{\mathrm{B}}T}\right) . \tag{4.15}$$

An integration across the depletion layer now gives

$$\exp\left(\frac{W_{\mathrm{Fn}}(x_{\mathrm{p}})}{k_{\mathrm{B}}T}\right) - \exp\left(\frac{W_{\mathrm{Fn}}(x_{\mathrm{n}})}{k_{\mathrm{B}}T}\right) = \frac{J_{\mathrm{n}}}{k_{\mathrm{B}}T}\int_{x_{\mathrm{n}}}^{x_{\mathrm{p}}} \frac{\mathrm{e}^{W_{\mathrm{C}}(x)/k_{\mathrm{B}}T}}{\mu_{\mathrm{n}}(x)N_{\mathrm{C}}(x)}\,\mathrm{d}x ,$$

if J_{n} is assumed to be constant throughout the depletion layer (negligible recombination). Rearranging terms transforms this equation into

$$\exp\left(\frac{\Delta W_{\mathrm{Fn}}}{k_{\mathrm{B}}T}\right) = 1 + \frac{J_{\mathrm{n}}}{k_{\mathrm{B}}T}\int_{x_{\mathrm{n}}}^{x_{\mathrm{p}}} \frac{\mathrm{e}^{W_{\mathrm{C}}(x)/k_{\mathrm{B}}T}}{\mu_{\mathrm{n}}(x)N_{\mathrm{C}}(x)}\,\mathrm{d}x , \tag{4.16}$$

where $\Delta W_{\mathrm{Fn}} = W_{\mathrm{Fn}}(x_{\mathrm{p}}) - W_{\mathrm{Fn}}(x_{\mathrm{n}})$, and $W_{\mathrm{Fn}}(x_{\mathrm{n}}) = 0$ has been assumed for simplicity. The value of J_{n} is related to the excess electron density $\Delta n(x_{\mathrm{p}})$ at the depletion-layer edge on the p-side by a relation of the general form[6]

$$J_{\mathrm{n}} = -eS_{\mathrm{np}}\Delta n(x_{\mathrm{p}}) . \tag{4.17}$$

Since $W_{\mathrm{Fn}} = -e\phi_{\mathrm{n}}$ and

$$\Delta n(x_{\mathrm{p}}) = \frac{n_{\mathrm{ie}}^2(x_{\mathrm{p}})}{p(x_{\mathrm{p}})}\left[\exp\left(\frac{\phi_{\mathrm{p}}(x_{\mathrm{p}}) - \phi_{\mathrm{n}}(x_{\mathrm{p}})}{V_{\mathrm{T}}}\right) - 1\right] ,$$

the electron current density obeys

$$J_{\mathrm{n}} = -eS_{\mathrm{np}}\frac{n_{\mathrm{ie}}^2(x_{\mathrm{p}})}{p(x_{\mathrm{p}})}\left[\exp\left(\frac{\Delta W_{\mathrm{Fn}}}{k_{\mathrm{B}}T}\right)\exp\left(\frac{V'}{V_{\mathrm{T}}}\right) - 1\right] , \tag{4.18}$$

where $V' = \phi_{\mathrm{p}}(x_{\mathrm{p}}) - \phi_{\mathrm{n}}(x_{\mathrm{n}})$ denotes the voltage drop across the junction. Combining (4.16) and (4.18) yields the following for the electron current density:

$$J_{\mathrm{n}} = \frac{-eS_{\mathrm{np}}}{1 + S_{\mathrm{np}}/S_{\mathrm{nj}}}\frac{n_{\mathrm{ie}}^2(x_{\mathrm{p}})}{p(x_{\mathrm{p}})}\left[\exp\left(\frac{V'}{V_{\mathrm{T}}}\right) - 1\right] , \tag{4.19}$$

where

[6] If, for example, the electrons were injected into a base region of thickness d_{B} and transported by diffusion only, the "recombination velocity" S_{np} would equal $D_{\mathrm{n}}/d_{\mathrm{B}}$.

$$\frac{1}{S_{\mathrm{nj}}} = \frac{1}{V_{\mathrm{T}}} \frac{n_{\mathrm{ie}}^2(x_{\mathrm{p}})}{p(x_{\mathrm{p}})} \exp\left(\frac{V'}{V_{\mathrm{T}}}\right) \int_{x_{\mathrm{n}}}^{x_{\mathrm{p}}} \frac{\mathrm{e}^{W_{\mathrm{C}}(x)/k_{\mathrm{B}}T}}{\mu_{\mathrm{n}}(x) N_{\mathrm{C}}(x)} \,\mathrm{d}x \tag{4.20}$$

defines a characteristic velocity S_{nj} for electron transport across the depletion layer. This velocity has to be large in comparison with S_{np} in order that the current is "diffusion-limited" and that the Shockley boundary conditions apply:[7] in the limit $S_{\mathrm{nj}} \gg S_{\mathrm{np}}$, the quasi-Fermi energies may be assumed to be constant across the depletion layer. In this case the generalized mass-action law

$$n(x_{\mathrm{p}})p(x_{\mathrm{p}}) = n_{\mathrm{ie}}^2(x_{\mathrm{p}}) \exp(V'/V_{\mathrm{T}}) \tag{4.21}$$

applies, and (4.19) reduces to the exponential current–voltage characteristic generally employed.

Abrupt Heterojunctions. The situation is somewhat different for abrupt heterojunctions: an increase of the forward bias in an abrupt np heterojunction decreases the barrier for electron flow to the base, but causes the barrier height for the electrons which flow in the opposite direction to increase, owing to a reduction of the notch. Furthermore, the net current crossing the heterojunction may not be small in comparison with the two emission currents crossing the barrier in opposite directions. If this is the case, the electron quasi-Fermi energy W_{Fn} is discontinuous across an abrupt heterojunction interface when a current is flowing, as illustrated in Fig. 4.6 for an npn HBT with an abrupt eb junction. The change of the electron quasi-Fermi energy at the heterojunction interface at $x_{\mathrm{j}} = 0$ can be calculated using an approach similar to that presented in [12]. Owing to the discontinuity, (4.15) may not be applied across the junction but may be applied to each of the intervals $[x_{\mathrm{n}}, 0[$ and $]0, x_{\mathrm{p}}]$ (using the coordinate notation as defined in Fig. 4.5). Integration across these intervals gives the identities

$$\exp\left(\frac{W_{\mathrm{Fn}}(0^-)}{k_{\mathrm{B}}T}\right) - \exp\left(\frac{W_{\mathrm{Fn}}(x_{\mathrm{n}})}{k_{\mathrm{B}}T}\right) = \frac{J_{\mathrm{n}}}{k_{\mathrm{B}}T} \int_{x_{\mathrm{n}}}^{0^-} \frac{\mathrm{e}^{W_{\mathrm{C}}(x)/k_{\mathrm{B}}T}}{\mu_{\mathrm{n}}(x) N_{\mathrm{C}}(x)} \,\mathrm{d}x \,,$$

$$\exp\left(\frac{W_{\mathrm{Fn}}(x_{\mathrm{p}})}{k_{\mathrm{B}}T}\right) - \exp\left(\frac{W_{\mathrm{Fn}}(0^+)}{k_{\mathrm{B}}T}\right) = \frac{J_{\mathrm{n}}}{k_{\mathrm{B}}T} \int_{0^+}^{x_{\mathrm{p}}} \frac{\mathrm{e}^{W_{\mathrm{C}}(x)/k_{\mathrm{B}}T}}{\mu_{\mathrm{n}}(x) N_{\mathrm{C}}(x)} \,\mathrm{d}x$$

and, after addition,

$$\exp\left(\frac{W_{\mathrm{Fn}}(x_{\mathrm{p}})}{k_{\mathrm{B}}T}\right) - \exp\left(\frac{W_{\mathrm{Fn}}(x_{\mathrm{n}})}{k_{\mathrm{B}}T}\right) + \exp\left(\frac{W_{\mathrm{Fn}}(0^-)}{k_{\mathrm{B}}T}\right) - \exp\left(\frac{W_{\mathrm{Fn}}(0^+)}{k_{\mathrm{B}}T}\right) = \frac{J_{\mathrm{n}}}{k_{\mathrm{B}}T} \int_{x_{\mathrm{n}}}^{x_{\mathrm{p}}} \frac{\mathrm{e}^{W_{\mathrm{C}}(x)/k_{\mathrm{B}}T}}{\mu_{\mathrm{n}}(x) N_{\mathrm{C}}(x)} \,\mathrm{d}x \,.$$

[7] See also the discussion in Appendix C.

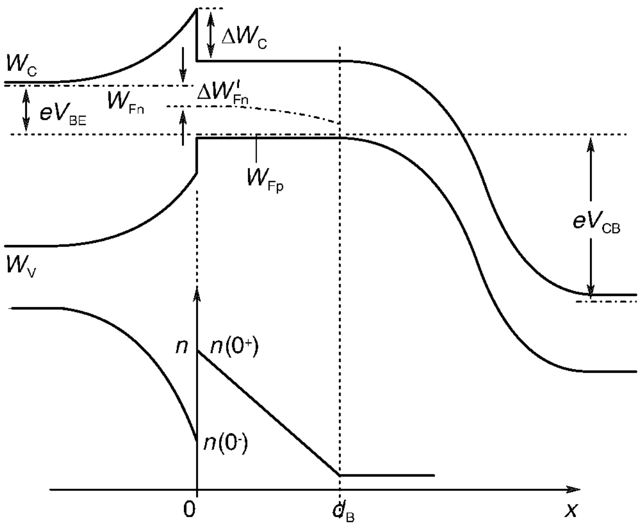


Fig. 4.6. Band scheme of HBT with an abrupt eb diode under forward bias, and schematic representation of electron density (after [13])

If $W_{\mathrm{Fn}}(x_{\mathrm{n}})$ is taken to be zero, this relation is equivalent to

$$\exp\left(\frac{\Delta W_{\mathrm{Fn}}}{k_{\mathrm{B}}T}\right) = 1 + \frac{J_{\mathrm{n}}}{k_{\mathrm{B}}T}\int_{x_{\mathrm{n}}}^{x_{\mathrm{p}}} \frac{\mathrm{e}^{W_{\mathrm{C}}(x)/k_{\mathrm{B}}T}}{\mu_{\mathrm{n}}(x)N_{\mathrm{C}}(x)}\,\mathrm{d}x - \left[1-\exp\left(\frac{\Delta W'_{\mathrm{Fn}}}{k_{\mathrm{B}}T}\right)\right]\exp\left(\frac{W_{\mathrm{Fn}}(0^-)}{k_{\mathrm{B}}T}\right) \quad (4.22)$$

where $\Delta W'_{\mathrm{Fn}} = W_{\mathrm{Fn}}(0^+) - W_{\mathrm{Fn}}(0^-)$ denotes the increment of the electron quasi-Fermi level at the junction. A relation for $\Delta W'_{\mathrm{Fn}}$ can be obtained from the electron current crossing the interface,

$$J_{\mathrm{n}} = -ev_{\mathrm{R}}N_{\mathrm{C}}(0^+)\left[\frac{n(0^-)}{N_{\mathrm{C}}(0^-)} - \frac{n(0^+)}{N_{\mathrm{C}}(0^+)}\exp\left(\frac{\Delta W_{\mathrm{C}}}{k_{\mathrm{B}}T}\right)\right], \quad (4.23)$$

where $\Delta W_{\mathrm{C}} = W_{\mathrm{C}}(0^+) - W_{\mathrm{C}}(0^-)$ is the change of the conduction band energy at the interface, and

$$v_{\mathrm{R}} = \sqrt{k_{\mathrm{B}}T/2\pi m_{\mathrm{n}}^*(0^+)} \quad (4.24)$$

denotes the Richardson velocity[8] [13, 16]. For heterojunctions with different values of the effective mass on either side of the junction, on the effective

[8]This velocity is obtained if the current is due to thermal emission across the barrier only. However, particularly in the vicinity of the maximum, where the potential barrier is thin, electrons may also tunnel through the barrier. Such tunneling requires us to replace v_{R} by an effective Richardson velocity v_{R}^*; in [14], the increase of the effective interface velocity has been expressed in terms of the barrier transparency. Analytical approximations for compact models were presented in [15].

density of states on both sides of the junction $N_\mathrm{C}(0^-)$ and $N_\mathrm{C}(0^+)$ will be different [13, 17–19]; in the case of degeneracy, degeneracy factors have to be introduced [16]. From (4.14), the following relation between the electron densities on the two sides of the junction is obtained:

$$n(0^+) \;=\; n(0^-)\,\frac{N_\mathrm{C}(0^+)}{N_\mathrm{C}(0^-)}\,\exp\!\left(\frac{\Delta W'_\mathrm{Fn}-\Delta W_\mathrm{C}}{k_\mathrm{B}T}\right)\,. \tag{4.25}$$

This allows us to rewrite (4.23) in the form

$$J_\mathrm{n} \;=\; -ev_\mathrm{R}\frac{N_\mathrm{C}(0^+)}{N_\mathrm{C}(0^-)}\left[1-\exp\!\left(\frac{\Delta W'_\mathrm{Fn}}{k_\mathrm{B}T}\right)\right]n(0^-)\,, \tag{4.26}$$

which is found to be equivalent to

$$\left[1-\exp\!\left(\frac{\Delta W'_\mathrm{Fn}}{k_\mathrm{B}T}\right)\right]\exp\!\left(\frac{W_\mathrm{Fn}(0^-)}{k_\mathrm{B}T}\right) \;=\; -\frac{J_\mathrm{n}}{ev_\mathrm{R}}\,\frac{\mathrm{e}^{W_\mathrm{C}(0^-)/k_\mathrm{B}T}}{N_\mathrm{C}(0^+)}$$

after substitution of $n(0^-)$ according to (4.14). Combining this result with (4.22) gives

$$\exp\!\left(\frac{\Delta W_\mathrm{Fn}}{k_\mathrm{B}T}\right) \;=\; 1+\frac{J_\mathrm{n}}{k_\mathrm{B}T}\int_{x_\mathrm{n}}^{x_\mathrm{p}}\frac{\mathrm{e}^{W_\mathrm{C}(x)/k_\mathrm{B}T}}{\mu_\mathrm{n}(x)N_\mathrm{C}(x)}\,\mathrm{d}x+\frac{J_\mathrm{n}}{ev_\mathrm{R}}\,\frac{\mathrm{e}^{W_\mathrm{C}(0^-)/k_\mathrm{B}T}}{N_\mathrm{C}(0^+)}\;;$$

substituting this result into (4.18) finally yields

$$J_\mathrm{n} \;=\; -e\,\frac{S_\mathrm{np}}{1+S_\mathrm{np}/S_\mathrm{n}}\,\frac{n_\mathrm{ie}^2(x_\mathrm{p})}{p(x_\mathrm{p})}\left[\exp\!\left(\frac{V'}{V_\mathrm{T}}\right)-1\right], \tag{4.27}$$

where $1/S_\mathrm{n}=1/S_\mathrm{nj}+1/S_\mathrm{th}$, and

$$S_\mathrm{th} \;=\; v_\mathrm{R}\,\frac{p(x_\mathrm{p})N_\mathrm{C}(0^+)}{n_\mathrm{ie}^2(x_\mathrm{p})}\,\exp\!\left(-\frac{W_\mathrm{C}(0^-)}{k_\mathrm{B}T}\right)\exp\!\left(-\frac{V'}{V_\mathrm{T}}\right) \tag{4.28}$$

defines an effective carrier velocity at the interface. The generalized mass-action law then has to be written as

$$n(x_\mathrm{p})p(x_\mathrm{p}) \;=\; \frac{n_\mathrm{ie}^2(x_\mathrm{p})}{1+S_\mathrm{np}/S_\mathrm{n}}\,\exp\!\left(\frac{V'}{V_\mathrm{T}}\right)\,. \tag{4.29}$$

If $S_\mathrm{np}, S_\mathrm{nj} \gg S_\mathrm{th}$ the electron current across the heterojunction is injection-limited, i.e. the current is predominantly controlled by the barrier at the interface of the abrupt heterojunction; the electron current density can then be approximated by

$$\begin{aligned} J_\mathrm{n} \;&\approx\; -eS_\mathrm{th}\frac{n_\mathrm{ie}^2(x_\mathrm{p})}{p(x_\mathrm{p})}\left[\exp\!\left(\frac{V'}{V_\mathrm{T}}\right)-1\right] \\ &=\; -ev_\mathrm{R}N_\mathrm{C}(0^+)\exp\!\left(-\frac{W_\mathrm{C}(0^-)}{k_\mathrm{B}T}\right)\left[1-\exp\!\left(-\frac{V'}{V_\mathrm{T}}\right)\right]\,, \end{aligned} \tag{4.30}$$

and no longer depends on $S_{\rm np}$. This result is equivalent to that of elementary thermal-emission theory: since

$$W_{\rm C}(0^-) \;=\; W_{\rm C}(x_{\rm n}) + eV_{\rm n} \;, \qquad \text{where} \qquad V_{\rm n} \;=\; k(V_{\rm J} - V') \;,$$

and

$$n(x_{\rm n}) \;=\; N_{\rm C}(x_{\rm n})\, {\rm e}^{-W_{\rm C}(x_{\rm n})/k_{\rm B}T} \;,$$

with the convention $W_{\rm Fn}(x_{\rm n}) = 0$, the electron current density equation can be rewritten as

$$J_{\rm n} \;=\; -ev_{\rm R}\, \frac{N_{\rm C}(0^+)}{N_{\rm C}(x_{\rm n})}\, n(x_{\rm n})\, \exp\!\left(-\frac{kV_{\rm J}}{V_{\rm T}}\right) \times \left[\exp\!\left(\frac{kV'}{V_{\rm T}}\right) - \exp\!\left(\frac{(k-1)V'}{V_{\rm T}}\right)\right] \;,$$

which is equivalent to (4.13) if

$$A_1^* \;=\; ev_{\rm R}\, \frac{N_{\rm C}(0^+)}{N_{\rm C}(x_{\rm n})}\, \frac{n(x_{\rm n})}{T^2} \;.$$

If the injection of minority carriers is controlled by thermal emission, the transfer current of an npn bipolar HBT becomes independent of the base width, resulting in a substantial reduction of the Early effect [13].

Effects of Graded Transition Layers. As was pointed out in [14], grading of the EB heterojunction may greatly enhance the emitter injection efficiency of HBTs if the current injected into the base region in the case of an abrupt heterojunction is limited by a conduction band spike due to a substantial discontinuity of the conduction band edges. If a steeply graded layer in which the material composition changes continuously from that of the n-type semiconductor to that of the p-type semiconductor is introduced, the barrier height for thermal emission of electrons is reduced by $\Delta W_{\rm Bn}$, as illustrated in Fig. 4.7.

In the absence of grading, the conduction band edge on the n-type side obeys the relation

$$W_{\rm C}(x) \;=\; W_{\rm C}(x_{\rm n}) + eV_{\rm n}\left(\frac{x - x_{\rm n}}{x_{\rm j} - x_{\rm n}}\right)^2 \;; \tag{4.31}$$

if a graded layer of thickness $d_{\rm gl}$ is introduced in the interval $[\,x_{\rm j} - d_{\rm gl}, x_{\rm j}\,]$, (4.31) may be applied only to the ungraded region between $x_{\rm n}$ and $x_{\rm j} - d_{\rm gl}$. In the interval between $x_{\rm j} - d_{\rm gl}$ and $x_{\rm j}$, the relation

$$W_{\rm C}(x) \;=\; W_{\rm C}(x_{\rm n}) + eV_{\rm n}\left(\frac{x - x_{\rm n}}{x_{\rm j} - x_{\rm n}}\right)^2 + \Delta W_{\rm C}\, \frac{x - x_{\rm j} + d_{\rm gl}}{d_{\rm gl}}$$

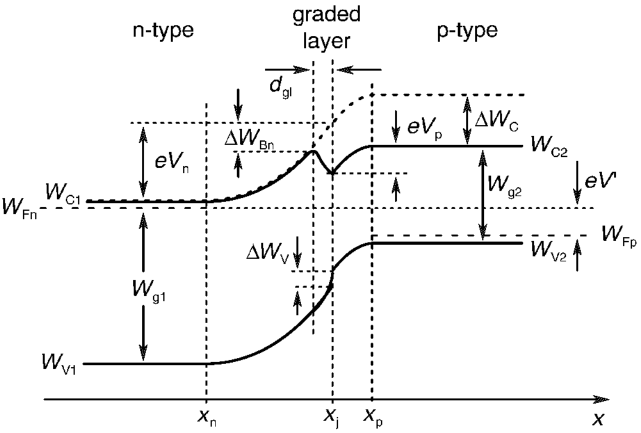


Fig. 4.7. Band scheme of a steeply graded np heterojunction (adapted from [14])

applies if a linearly graded layer is assumed and $\Delta W_C = W_C(x_j^+) - W_C(x_j^-)$ is the conduction band discontinuity in the abrupt junction. The potential difference V_{gl} across the graded layer is

$$V_{gl} = V_n - V_n \left(\frac{x_j - d_{gl} - x_n}{x_j - x_n} \right)^2 = V_n \frac{2d_{gl}(x_j - x_n) - d_{gl}^2}{(x_j - x_n)^2} ,$$

if variation of the dielectric permittivity within the graded layer is neglected. In the case where $eV_{gl} + \Delta W_C < 0$, the conduction band edge becomes lower in the interval between $x_j - d_{gl}$ and x_j. Under this condition, the maximum of the conduction band occurs at $x_j - d_{gl}$ and lies below the value that would have been observed in the case of an abrupt heterojunction by $\Delta W_{Bn} = eV_{gl}$. If the electron current without a graded layer is injection-limited, the decrease of S_n associated with the reduction of the conduction band spike results in enhanced minority-carrier injection – with the consequence of an increased transfer current if a graded layer is used in an HBT.

In a linearly graded layer, the energy position of the conduction band edge increases with x owing to the electric field caused by the space charge and decreases with x owing to the grading, which causes a quasi-electric field $-\Delta W_C/ed_{gl}$. If the the electric field E at $x_j - d_{gl}$ exceeds $-\Delta W_C/e$, the conduction band rises monotonically in the interval between x_n and x_j and merges continuously with the conduction band on the p-type side. This situation is illustrated in Fig. 4.8. In this case thermal emission need not be considered, and the behavior of the heterojunction can be derived from the modified drift–diffusion theory.

4.1.3 Depletion Capacitance

The depletion capacitance of abrupt and graded heterojunctions can be described with the model equation used for homojunctions if modified relations are used for the computation of the zero-bias depletion capacitance and the built-in voltage.

Abrupt Heterojunction. Let $-x_{\mathrm{n}}$ and x_{p} denote the coordinates of the depletion layer edge (see Fig. 4.5) of an abrupt heterojunction located at $x_{\mathrm{j}} = 0$. Under the assumption of homogeneous doping on each side of the junction, the voltage drops V_{n} and V_{p} on either side of the metallurgical junction are given by

$$V_{\mathrm{n}} = \frac{eN_{\mathrm{D}}x_{\mathrm{n}}^2}{2\epsilon_1} \quad \text{and} \quad V_{\mathrm{p}} = \frac{eN_{\mathrm{A}}x_{\mathrm{p}}^2}{2\epsilon_2} .$$

Using the neutrality condition $eN_{\mathrm{D}}x_{\mathrm{n}} = eN_{\mathrm{A}}x_{\mathrm{p}}$ and the fact that the potential drop across the junction is $V_{\mathrm{n}} + V_{\mathrm{p}} = V_{\mathrm{J}} - V'$, the locations of the depletion layer boundaries can be derived as

$$x_{\mathrm{n}} = \sqrt{\frac{2N_{\mathrm{A}}\epsilon_1\epsilon_2(V_{\mathrm{J}}-V')}{eN_{\mathrm{D}}(\epsilon_1 N_{\mathrm{D}}+\epsilon_2 N_{\mathrm{A}})}} \tag{4.32}$$

and

$$x_{\mathrm{p}} = \sqrt{\frac{2N_{\mathrm{D}}\epsilon_1\epsilon_2(V_{\mathrm{J}}-V')}{eN_{\mathrm{A}}(\epsilon_1 N_{\mathrm{D}}+\epsilon_2 N_{\mathrm{A}})}} . \tag{4.33}$$

From these equations, the depletion capacitance[9]

$$c_{\mathrm{j}}(V') = -eN_{\mathrm{D}}A_{\mathrm{j}}\frac{\mathrm{d}x_{\mathrm{n}}}{\mathrm{d}V'} = \frac{C_{\mathrm{J0}}}{\sqrt{1-V'/V_{\mathrm{J}}}}$$

is obtained, where

$$C_{\mathrm{J0}} = A_{\mathrm{j}}\sqrt{\frac{e\epsilon_1\epsilon_2 N_{\mathrm{A}}N_{\mathrm{D}}}{2(\epsilon_1 N_{\mathrm{D}}+\epsilon_2 N_{\mathrm{A}})V_{\mathrm{J}}}} . \tag{4.34}$$

This relation reduces to the result for a homojunction in the special case $\epsilon_1 = \epsilon_2 = \epsilon$; a slight modification is required to take account of graded layers within the space charge region.

[9]This capacitance can be considered as a series connection of two partial capacitances on either side of the metallurgical junction, i.e.

$$\frac{1}{c_{\mathrm{j}}} = \left(\frac{\epsilon_1 A_{\mathrm{j}}}{x_{\mathrm{n}}}\right)^{-1} + \left(\frac{\epsilon_2 A_{\mathrm{j}}}{x_{\mathrm{p}}}\right)^{-1} , \quad \text{equivalent to} \quad c_{\mathrm{j}} = \frac{\epsilon_1\epsilon_2}{\epsilon_1 x_{\mathrm{p}}+\epsilon_2 x_{\mathrm{n}}} A_{\mathrm{j}} .$$

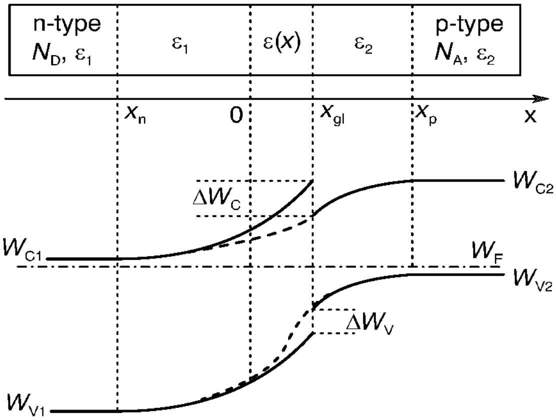


Fig. 4.8. Band scheme of heterojunction without a graded layer (solid curves) and with a graded layer (dashed curves) between 0 and x_{gl}

Graded-Layer Heterojunction. In a graded-layer heterojunction the dielectric permittivity is a function of position, resulting in the relation

$$\frac{d}{dx}\left[\epsilon(x)E(x)\right] = \rho(x) \tag{4.35}$$

between the electric field strength $E(x)$ and the space charge density $\rho(x)$. A double integration of (4.35) yields the following for the potential difference across the space-charge layer:

$$V_J - V' = \int_{x_n}^{x_p} \frac{1}{\epsilon(x)} \int_{x_n}^{x} \rho(x')\, dx'\, dx\ . \tag{4.36}$$

In the following, it will be assumed that the graded layer, with boundaries 0 and x_{gl}, lies completely within the space charge region, with boundaries x_n and x_p (see Fig. 4.8) and that the transition from n-type doping to p-type doping occurs within the graded layer. In this case the space charge region can be separated into three regions, and the neutrality condition can be written as

$$\int_{x_n}^{x_p} \rho(x)\, dx = Q'_n + Q'_{gl} + Q'_p = 0\ ,$$

where

$$Q'_n = \int_{x_n}^{0} \rho(x)\, dx\ , \quad Q'_{gl} = \int_{0}^{x_{gl}} \rho(x)\, dx \quad \text{and} \quad Q'_p = \int_{x_{gl}}^{x_p} \rho(x)\, dx$$

denote the charges per unit area in the three regions. The potential difference across the graded layer is

$$\begin{aligned}\int_0^{x_{gl}} \frac{1}{\epsilon(x)} \int_{x_n}^{x} \rho(x')\, dx' dx &= Q'_n \int_0^{x_{gl}} \frac{1}{\epsilon(x)}\, dx + \int_0^{x_{gl}} \frac{1}{\epsilon(x)} \int_0^{x} \rho(x')\, dx' dx \\ &= \frac{x_{gl} Q'_n}{\epsilon_{gl}} + V_{gl}\ ,\end{aligned}$$

where ϵ_{gl} denotes the reciprocal average of the dielectric permittivity in the graded layer. With this notation, (4.36) reads after an integration by parts

$$V_{\text{J}} - V' = -\int_{x_{\text{n}}}^{0} \frac{x\rho(x)}{\epsilon_1}\,\mathrm{d}x - \int_{x_{\text{gl}}}^{x_{\text{p}}} \frac{x\rho(x)}{\epsilon_2}\,\mathrm{d}x + \frac{x_{\text{gl}}Q'_{\text{n}}}{\epsilon_{\text{gl}}} + \frac{x_{\text{gl}}Q'_{\text{p}}}{\epsilon_2} + V_{\text{gl}}\,.$$

In the case of homogeneously doped n-type and p-type regions, $Q'_{\text{n}} = -eN_{\text{D}}x_{\text{n}}$ and $Q'_{\text{p}} = -eN_{\text{A}}(x_{\text{p}} - x_{\text{gl}})$, and therefore

$$\int_{x_{\text{n}}}^{0} x\rho\,\mathrm{d}x = -\frac{eN_{\text{D}}x_{\text{n}}^2}{2} = -\frac{(Q'_{\text{n}})^2}{2eN_{\text{D}}}$$

and

$$\int_{x_{\text{gl}}}^{x_{\text{p}}} x\rho\,\mathrm{d}x = -\frac{eN_{\text{A}}(x_{\text{p}}^2 - x_{\text{gl}}^2)}{2} = -\frac{(Q'_{\text{p}})^2}{2eN_{\text{A}}} + x_{\text{gl}}Q'_{\text{p}}\,.$$

When the above equations are combined with the neutrality condition, a quadratic relation for Q'_{n} is obtained; the solution of this equation is

$$Q'_{\text{n}} = \sqrt{\frac{2e\epsilon_1\epsilon_2 N_{\text{A}}N_{\text{D}}(V_{\text{J}} - V_{\text{gl}} - V') - \epsilon_1 N_{\text{D}}(Q'_{\text{gl}})^2}{\epsilon_1 N_{\text{D}} + \epsilon_2 N_{\text{A}}} + \Delta_{\text{gl}}^2} - \Delta_{\text{gl}}\,,$$

where

$$\Delta_{\text{gl}} = \frac{\epsilon_1 N_{\text{D}}(e x_{\text{gl}}\epsilon_2 N_{\text{A}} + \epsilon_{\text{gl}}Q'_{\text{gl}})}{\epsilon_1 N_{\text{D}} + \epsilon_2 N_{\text{A}}}\,.$$

Differentiation with respect to V' gives the depletion capacitance of the graded-layer heterojunction in the form

$$c_{\text{j}}(V') = -A_{\text{j}}\frac{\mathrm{d}Q'_{\text{n}}}{\mathrm{d}V'} = \frac{C'_{\text{J0}}}{\sqrt{1 - V'/\tilde{V}_{\text{J}}}}\,,$$

where the "depletion capacitance offset voltage" is given by

$$\tilde{V}_{\text{J}} = V_{\text{J}} - V_{\text{gl}} - \frac{(Q'_{\text{gl}})^2}{2e\epsilon_2 N_{\text{A}}} + \frac{(\epsilon_1 N_{\text{A}} + \epsilon_2 N_{\text{D}})}{2e\epsilon_1\epsilon_2 N_{\text{A}}N_{\text{D}}}\Delta_{\text{gl}}^2\,, \tag{4.37}$$

and C'_{J0} is calculated according to (4.34) with V_{J} replaced by $V_{\text{J}'}$. The presence of the graded layer therefore results in a shift of the depletion capacitance offset voltage with respect to the built-in voltage V_{J} of the junction. Since in the limit $x_{\text{gl}} \to 0$ all of Q'_{gl}, V_{gl} and Δ_{gl} become zero, (4.37) reduces to the abrupt-heterojunction solution in the limit of vanishing thickness of the graded layer.

4.2 Heterojunction Bipolar Transistors

Heterojunction bipolar transistors employ one or two heterojunctions, which may be graded or abrupt, in order to improve device performance in comparison with homojunction BJTs. Single-heterojunction bipolar transistors (SHBTs) benefit from a base layer with a reduced bandgap, which, however, equals the bandgap within the collector region. Double-heterojunction bipolar transistors (DHBTs) have an additional heterojunction between the base and the wide-gap collector region. This offers the advantage of an increased breakdown voltage and reduces the offset voltage in switching applications.

4.2.1 Transfer Current

In gently graded HBTs, the material parameters vary continuously and the transfer current can be described in terms of the integral charge control relation derived in Sect. 3.2. In the case of HBTs with steeply graded or abrupt junctions, the discontinuity of the electron quasi-Fermi potential across the heterojunctions has to be considered, if the current is limited by thermal emission. In this case the integral charge control relation may be applied only to the base region between the two abrupt heterojunctions, resulting in

$$\frac{I_\mathrm{T}}{eA_\mathrm{je}} \int_{x_\mathrm{be}}^{x_\mathrm{bc}} \frac{p}{D_\mathrm{n} n_\mathrm{ie}^2}\,\mathrm{d}x = \exp\left(\frac{\phi_\mathrm{pB} - \phi_\mathrm{n}(x_\mathrm{be})}{V_\mathrm{T}}\right) - \exp\left(\frac{\phi_\mathrm{pB} - \phi_\mathrm{n}(x_\mathrm{bc})}{V_\mathrm{T}}\right) .$$

Owing to the possible discontinuity of the quasi-Fermi potential across the heterojunction, however, it is not possible to identify $\phi_\mathrm{pB} - \phi_\mathrm{n}(x_\mathrm{bc})$ and $\phi_\mathrm{pB} - \phi_\mathrm{n}(x_\mathrm{bc})$ with the potential drops

$$V_{\mathrm{B'E'}} = \phi_\mathrm{pB} - \phi_\mathrm{n}(x_\mathrm{be}) - \frac{W_\mathrm{Fn}(x_\mathrm{be}) - W_\mathrm{Fn}(x_\mathrm{eb})}{e} \tag{4.38}$$

and

$$V_{\mathrm{B'C'}} = \phi_\mathrm{pB} - \phi_\mathrm{n}(x_\mathrm{bc}) - \frac{W_\mathrm{Fn}(x_\mathrm{bc}) - W_\mathrm{Fn}(x_\mathrm{cb})}{e} \tag{4.39}$$

across the junction. In analogy with (4.26), we may write

$$\exp\left(\frac{W_\mathrm{Fn}(x_\mathrm{be}) - W_\mathrm{Fn}(x_\mathrm{eb})}{k_\mathrm{B}T}\right) = 1 - \frac{I_\mathrm{T}}{eA_\mathrm{je}} \frac{p(x_\mathrm{be})}{n_\mathrm{ie}^2(x_\mathrm{bc})} \frac{1}{S_\mathrm{ne}} \mathrm{e}^{-V_{\mathrm{B'E'}}/V_\mathrm{T}} \tag{4.40}$$

and

$$\exp\left(\frac{W_\mathrm{Fn}(x_\mathrm{bc}) - W_\mathrm{Fn}(x_\mathrm{cb})}{k_\mathrm{B}T}\right) = 1 - \frac{I_\mathrm{T}}{eA_\mathrm{je}} \frac{p(x_\mathrm{bc})}{n_\mathrm{ie}^2(x_\mathrm{bc})} \frac{1}{S_\mathrm{nc}} \mathrm{e}^{-V_{\mathrm{B'C'}}/V_\mathrm{T}} , \tag{4.41}$$

with the effective interface velocities S_ne and S_nc at the heterojunctions defined in analogy with (4.27).[10] Combining the above equations yields the modified integral charge control relation

[10] The effective Richardson velocities v_R^* at the two junctions may be different if tunneling across differently shaped barriers at the eb and bc heterojunctions is taken into account.

$$I_{\mathrm{T}} = eA_{\mathrm{je}} \frac{\exp\left(\frac{V_{\mathrm{B'E'}}}{V_{\mathrm{T}}}\right) - \exp\left(\frac{V_{\mathrm{B'C'}}}{V_{\mathrm{T}}}\right)}{\int_{x_{\mathrm{be}}}^{x_{\mathrm{bc}}} \frac{p(x)}{D_{\mathrm{n}}(x) n_{\mathrm{ie}}^2(x)} \mathrm{d}x + \frac{1}{S_{\mathrm{ne}}} \frac{p(x_{\mathrm{be}})}{n_{\mathrm{ie}}^2(x_{\mathrm{be}})} + \frac{1}{S_{\mathrm{nc}}} \frac{p(x_{\mathrm{bc}})}{n_{\mathrm{ie}}^2(x_{\mathrm{bc}})}} . \quad (4.42)$$

Equation (4.42) generalizes the formulation given in [20] to HBTs with graded base layers. Base layers with a gradual reduction of the bandgap towards the collector junction lead to a reduced base transit time τ_{B} [21]. Alternatively, for a given value of τ_{B}, the width of the base layer may be increased and hence the internal base resistance decreased by grading the base layer. This is exploited in most HBTs with a SiGe base layer (see Sect. 4.3); an investigation of the effect of base grading on the gain and high-frequency performance of AlGaAs/GaAs HBTs has been presented in [22].

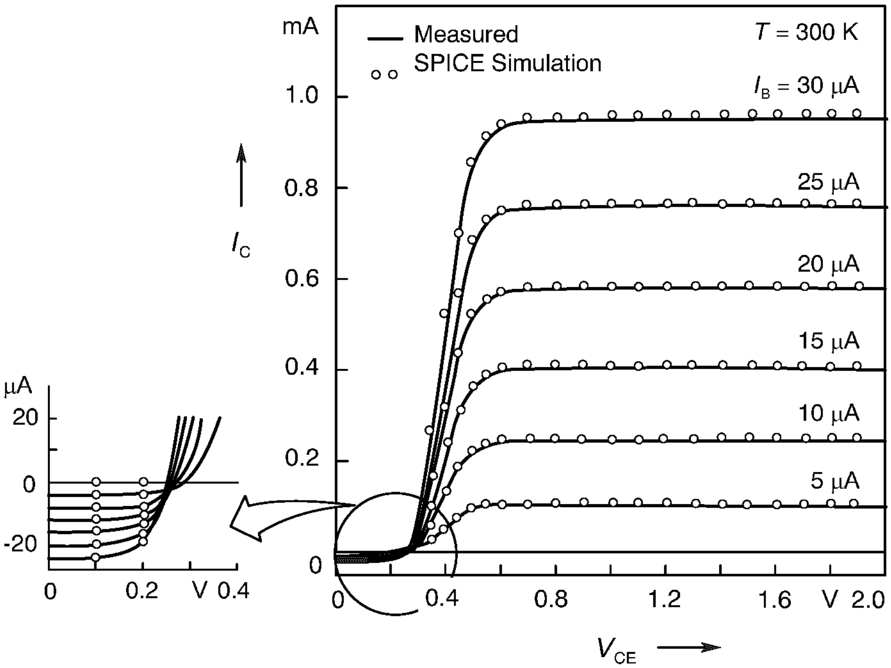


Fig. 4.9. Output characteristics of GaAs/AlGaAs HBT (after [23])

4.2.2 Offset Voltage

Different values of the turn-on voltage of the eb and bc diode result in an offset voltage $V_{\mathrm{CE}}(I_{\mathrm{C}} = 0)$ that may reach several hundred millivolts (see Fig. 4.9). The exact value of the offset voltage is determined by the positional dependence of the bandgap, the ratio of eb and bc junction areas, high-level injection effects, and series resistances. The offset voltage can be determined from the relations

$$I_{\rm C} = I_{\rm T} - I_{\rm BC} \quad \text{and} \quad I_{\rm B} = I_{\rm BE} + I_{\rm BC}$$

if the current–voltage relations are known. In HBTs with wide-gap emitter the transfer saturation current $I_{\rm S}$ is generally small in comparison with the saturation current $I_{\rm SC}$ of the bc diode. If this is the case, the simplified equations

$$I_{\rm C} \approx I_{\rm S} \exp\left(\frac{V_{\rm B'E'}}{N_{\rm F} V_{\rm T}}\right) - I_{\rm SC} \exp\left(\frac{V_{\rm B'C'}}{N_{\rm C} V_{\rm T}}\right) = 0 \tag{4.43}$$

$$I_{\rm B} \approx I_{\rm SC} \exp\left(\frac{V_{\rm B'C'}}{N_{\rm C} V_{\rm T}}\right) \tag{4.44}$$

can be applied. This yields for the offset voltage

$$V_{\rm CE}(I_{\rm C} = 0) = R_{\rm EE'} I_{\rm B} + V_{\rm T} \ln\left[\left(\frac{I_{\rm B}}{I_{\rm S}}\right)^{N_{\rm F}} \left(\frac{I_{\rm SC}}{I_{\rm B}}\right)^{N_{\rm F}}\right] \tag{4.45}$$

This result considers the effect of different areas of the eb and bc junctions as $I_{\rm S} \sim A_{\rm je}$ and $I_{\rm SC} \sim A_{\rm jc}$ as well as effects of the positional dependence of the bandgap and the emitter series resistance. Not considered are high-level injection effects and voltage drops associated with the epilayer of the transistor.

4.2.3 Nonequilibrium Carrier Transport

In an abrupt-junction HBT, the transfer current is limited by thermal emission across the conduction band spike at the eb junction if the condition $v_{\rm th} \exp(-\Delta W_{\rm C}/k_{\rm B}T) \ll D/d_{\rm B}$ is fulfilled, where $v_{\rm th} \approx 10^7$ cm/s denotes the average thermal velocity in the emitter in the x direction [24]. In such a transistor, the transfer current will be independent of the base thickness $d_{\rm B}$ and the current gain will vary in inverse proportion to $d_{\rm B}$ owing to the reduction of the base current with the thickness of the base layer, an effect which can be expected if the base current is essentially due to recombination in the base layer. This is illustrated in Fig. 4.10, where the current gain $B_{\rm N}$ is plotted versus the base width $d_{\rm B}$ for AlInAs/InGaAs HBTs with an abrupt eb heterojunction. For $d_{\rm B} < 100$ nm, a dependence $B_{\rm N} \sim 1/d_{\rm B}$ is observed, as is expected if the base current varies in proportion to the width of the base region, i.e. $I_{\rm B} \sim d_{\rm B}$, and if the transfer current is injection-limited and therefore independent of $d_{\rm B}$. At larger values of $d_{\rm B}$, the current gain is found to vary in proportion to $1/d_{\rm B}^2$, as is expected for a transfer current that varies in proportion to $1/d_{\rm B}$; nevertheless, this result cannot be explained by simple diffusive transport of the electrons across the base layer, as was demonstrated by determination of the base transit time $\tau_{\rm B}$, which was well below the value

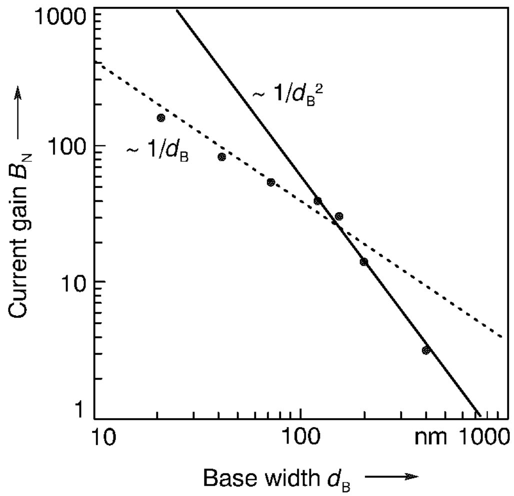


Fig. 4.10. Common-emitter current gain B_N measured at room temperature versus base thickness d_B for $Al_{0.48}In_{0.52}As/In_{0.53}Ga_{0.47}As$ HBTs (after [24])

expected for pure diffusive transport. This reduction of the base transit time can be attributed to nonequilibrium transport (velocity overshoot), as the electrons are injected with a large kinetic energy into the base region owing to the high conduction band offset $\Delta W_C \approx 0.47$ eV [24]. For distances comparable to the mean free path, the mean carrier velocity is larger than the saturated drift velocity and close to the group velocity $|\boldsymbol{v}_n| = \hbar^{-1}|\nabla_{\boldsymbol{k}} W_C|$ (Fig. 4.11). A survey of nonequilibrium electron transport in the base and collector regions of III–V semiconductor HBTs is given in [25, 26].

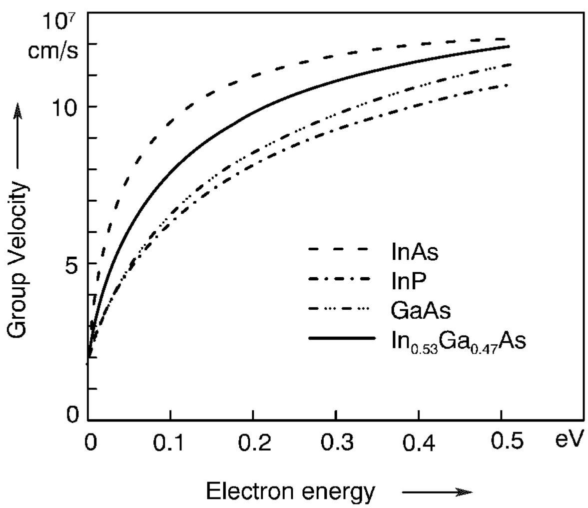


Fig. 4.11. Group velocity for electrons in the Γ-valley moving in the ΓX direction as a function of energy for different semiconductor materials (after [25])

4.3 Silicon-Based Semiconductor Heterostructures

Material systems and growth techniques for the realization of heterostructures that are compatible with silicon technology have been developed during the past two decades. This section gives an overview of the widely investigated narrow-gap SiGe alloys, which have now found their way into production processes.[11]

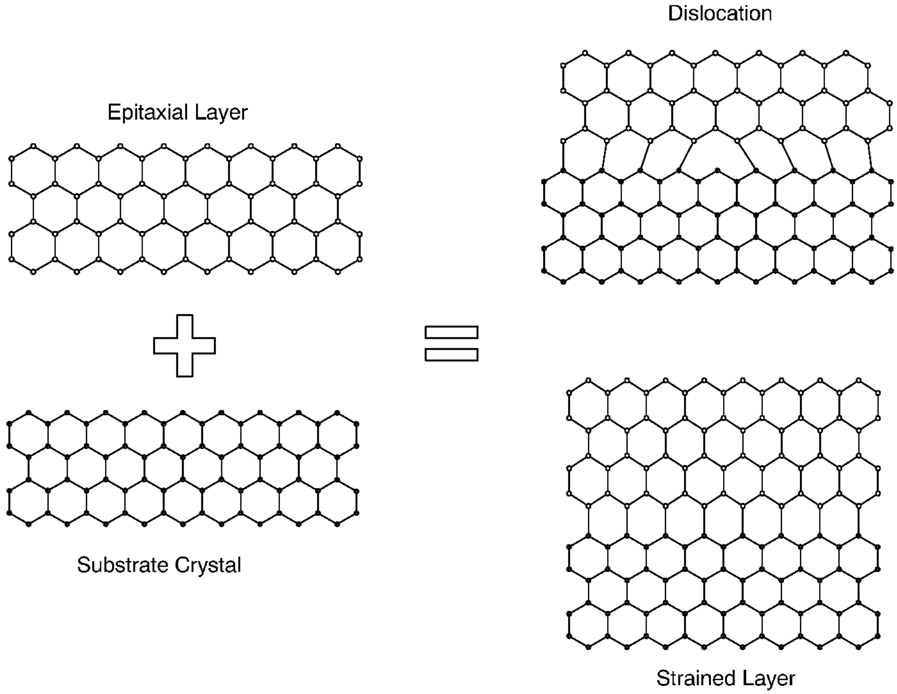


Fig. 4.12. Schematic representation of alternative growth modes for lattice-mismatched, tetrahedrally bonded semiconductors (after [7])

Silicon and germanium both crystallize in the diamond lattice and are completely miscible. The atomic spacing in a Ge crystal is, however, 4.17% larger than that for Si; in an $Si_{1-x}Ge_x$ crystal with a small Ge concentration x, a diamond lattice with lattice constant

[11] Besides SiGe alloys, various wide-gap materials, such as oxygenated and nitrogenated Si, microcrystalline hydrogenated Si and β-Si, have been investigated, in an attempt to produce wide-gap emitters.

$$a_{\mathrm{Si}_{(1-x)}\mathrm{Ge}_x} = a_{\mathrm{Si}} + (a_{\mathrm{Ge}} - a_{\mathrm{Si}})x \tag{4.46}$$

is obtained, according to Vegard's rule [27]. The formation of arbitrary Si-$Si_{1-x}Ge_x$ heterostructures is therefore not possible, owing to lattice mismatch, which results in misfit dislocations (see Fig. 4.12). The relaxation that results from the formation of such dislocations reduces the energy associated with the elastic deformation of the deposited layer. For very thin deposited layers,

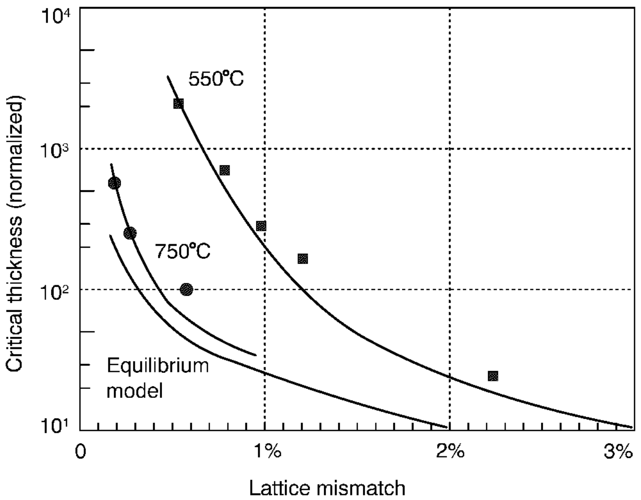


Fig. 4.13. Critical thickness of SiGe layers on Si substrate as a function of lattice mismatch for different values of growth temperature (in units of the magnitude $b \approx 0.4$ nm of the Burger vector, after [5,28])

however, the energy due to elastic deformation is small – smaller than the energy associated with the formation of misfit dislocations. In this case a $Si_{1-x}Ge_x$ layer deposited on a Si substrate will adopt the lattice constant of the underlying substrate and show a compressive in-plane strain, with extension in the growth direction (see Fig. 4.12). In such pseudomorphic or strained-layer heterostructures, the constraint of having a lattice-matched substrate is relaxed, provided that a critical thickness is not exceeded.

Figure 4.13 shows the critical thickness of $Si_{1-x}Ge_x$ layers versus the (relative) lattice mismatch $\Delta a/a$ in thermal equilibrium. The figure also shows that SiGe films grown at higher temperatures tend to relax at thickness values that are considerably smaller than those observed for layers grown at lower temperatures. Exceeding the critical thickness for high temperatures by using low-temperature epitaxy can nevertheless be dangerous, as thicker films are thermodynamically metastable at reduced temperatures: if the original growth temperature is exceeded in a subsequent process step, such as rapid thermal annealing, the metastable layer will spontaneously relax.

4.3.1 Growth of SiGe/Si Heterostructures

SiGe/Si heterostructures are predominantly[12] realized using the UHVCVD growth techniques [31] developed during the last two decades. This technique offers excellent film uniformity across large wafer areas, conformality, and apparatus costs that are small in comparison with those of physical vapor deposition techniques such as MBE. For epitaxial growth, the grown layer has to replicate the crystal lattice of the underlying substrate. This requires a highly perfect, chemically pure initial growth interface. In classical silicon epitaxy, such an interface is obtained by the use of high temperatures to volatilize or dissolve contaminating species such as oxygen or carbon. Epitaxial layers can be grown with a rather precisely controlled dopant concentration using classical CVD systems; however, attempts to create steep vertical doping profiles suffer from the effect of autodoping, which is due to solid-state diffusion of dopant species from the substrate into the epilayer and to evaporation of dopant from the substrate with subsequent reincorporation in the grown layer. In order to reduce this effect, both the substrate temperature and the gas pressure in the CVD system have to be lowered.

Deposition temperatures around 550°C have been found to be favorable for the growth of pseudomorphic SiGe layers with moderate Ge concentrations (up to 15%) [27]. Epitaxial growth under such conditions requires an ultra-pure growth environment throughout the deposition process and a UHV base pressure (typical operating pressures are in the range of 0.001 mbar). The growth of the deposited layers occurs as a result of pyrolysis of the reactive gases used (e.g. silane, germane and diborane); doping and compositional transitions are accomplished via changes in the inlet gas composition. The growth rate depends on the substrate temperature – which is approximately equal to the wall temperature of the CVD reactor[13] in a UHVCVD system – and the reactant pressure; the typical values of the growth rate of the order of 1 nm/min allow precise layer control.

[12] Molecular-beam epitaxy has been used for the production of laboratory samples (see e.g. [27] and references therein); the first SiGe/Si HBTs were realized with this technique [29, 30]. This technique allows one to deposit the entire device structure without exposure to high-temperature steps; the growth rate and film composition achieved with this physical deposition process are essentially determined by the impinging species and are independent of substrate orientation and temperature. However, the high cost of the apparatus, difficulties with selective growth and problems in achieving high doping levels, required for high current gain and small base resistance, do not make this technique appealing for mass production.

[13] In a modified process scheme, called limited-reaction processing, the wall temperature of the CVD reactor is well below the substrate temperature. The substrate is radiatively heated in this case and can change its temperature rapidly, allowing one to control the growth rate [27].

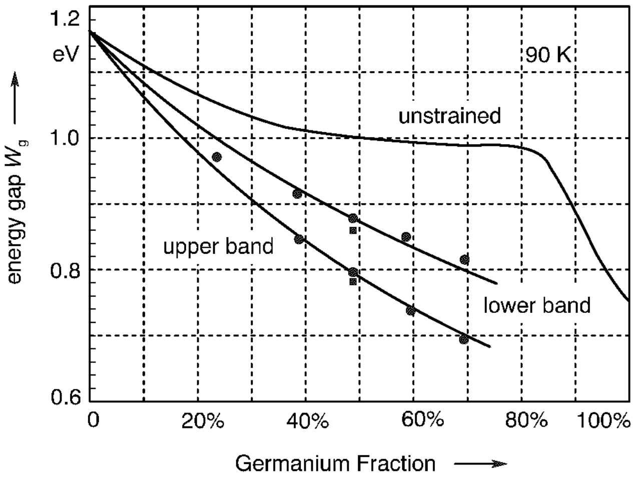


Fig. 4.14. Bandgaps for unstrained bulk $Si_{1-x}Ge_x$ alloys and strained $Si_{1-x}Ge_x$ layers on a Si substrate for light- and heavy-hole bands, together with experimental data points obtained from optical absorption measurements (after [5, 28])

4.3.2 SiGe Material Parameters

Bandgap Reduction and Band Offsets

In unstrained SiGe alloys, the bandgap is reduced in comparison with silicon owing to the presence of the germanium atoms (see Fig. 4.14). Silicon is an indirect semiconductor with a sixfold degeneracy of the conduction band minimum: in $\boldsymbol{k}$ space there are six minima of the conduction band, at a distance of about $0.85\,(2\pi/a)$ from the Γ point in the X direction (see Fig. 2.1) [32]. A graphical representation of the constant-energy surfaces in silicon (in $\boldsymbol{k}$ space) at an energy slightly above the conduction band minimum is shown in Fig. 4.15a. The ellipsoidal shape results from the anisotropic effective mass, which has a value m_l^* along the symmetry axis and a value m_t^* transverse to the symmetry axis. The kinetic energy $w_\mathrm{C}(\boldsymbol{k}) = W_\mathrm{C}(\boldsymbol{k}) - W_\mathrm{C}$ of an electron slightly above the conduction band minimum W_C at $k_{x0} = 0.85\,(2\pi/a)$ is given by

$$w_\mathrm{C}(\boldsymbol{k}) = \frac{\hbar^2 (k_x - k_{x0})^2}{2m_\mathrm{l}^*} + \frac{\hbar^2 k_y^2}{2m_\mathrm{t}^*} + \frac{\hbar^2 k_z^2}{2m_\mathrm{t}^*} .$$

The constant-energy surface $w_\mathrm{C}(\boldsymbol{k}) = \mathrm{const.}$ thus describes an ellipsoidal surface in $\boldsymbol{k}$ space. The values of $m_\mathrm{l}^* \approx 0.98\,m_\mathrm{e}$ and $m_\mathrm{t}^* \approx 0.19\,m_\mathrm{e}$ differ by more than a factor of five. For unstrained SiGe alloys with a Ge content below 85%, the conduction band is silicon-like, with six equivalent conduction band minima in the X directions; for larger values of the Ge content, a germanium-like conduction band, with eight equivalent conduction band minima in the L directions (see Fig. 2.1) is observed.

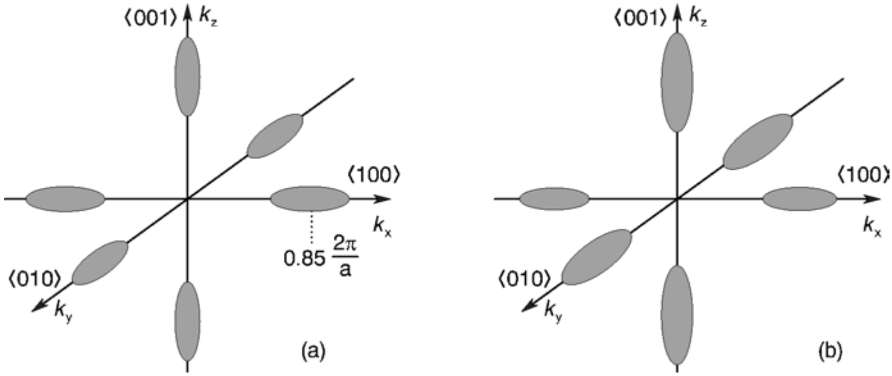


Fig. 4.15. Constant-energy surfaces (for energies slightly above the the conduction band minimum) for (**a**) silicon, showing the sixfold degeneracy of the conduction band states, and (**b**) a strained SiGe layer grown on a ⟨100⟩ interface. In the SiGe layer, the degeneracy of the conduction band minima is removed owing to the stress, which causes the valleys in the ⟨100⟩ direction to shift upwards in energy and the valleys in the ⟨001⟩ and ⟨010⟩ directions to shift downwards in energy

The bandgap reduction of strained SiGe layers grown on a Si substrate was investigated both experimentally and theoretically in [28, 33]. As compared with the unstrained SiGe alloy, a substantial reduction of the indirect bandgap results from the strain: there is a uniform shift of the bandgap due to dilation, and in addition, uniaxial splittings of the degenerate conduction and valence band edges occur. The difference between the lowest-lying conduction band edge and the highest valence band edge in a $\mathrm{Ge}_x\mathrm{Si}_{1-x}$ layer on Si is a function of the Ge content x and can be written in the form [7]

$$\frac{W_{\mathrm{g}}(x,T)}{\mathrm{eV}} = \frac{W_{\mathrm{g}}(0,T)}{\mathrm{eV}} - 0.96\,x + 0.43\,x^2 - 0.17\,x^3\ .$$

In $\mathrm{Ge}_x\mathrm{Si}_{1-x}$/Si heterojunctions, a Ge content of 20% yields a bandgap reduction of 176 meV in the strained layer, corresponding to $7k_{\mathrm{B}}T$ at room temperature.[14]

The valence band offset of a $\mathrm{Ge}_x\mathrm{Si}_{1-x}$/Si heterojunction is a function of the Ge content and increases by approximately 7 meV for each percent of Ge, i.e. [35, 36]

$$\Delta W_{\mathrm{V}}(x,T) \approx x \times 700\,\mathrm{meV}\ .$$

This result has been derived theoretically and has been verified by various experimental investigations, as shown in Fig. 4.16. The conduction band offset

[14] Also, in strained $\mathrm{Ge}_x\mathrm{Si}_{1-x}$ layers grown on a Si substrate, the hole effective mass is found to be reduced in comparison with silicon. In [34] it is shown theoretically, that this change of hole effective mass modifies the effective density of states in the valence band and causes results for the bandgap reduction obtained from of electrical measurements of the transfer current to be different from optical measurements.

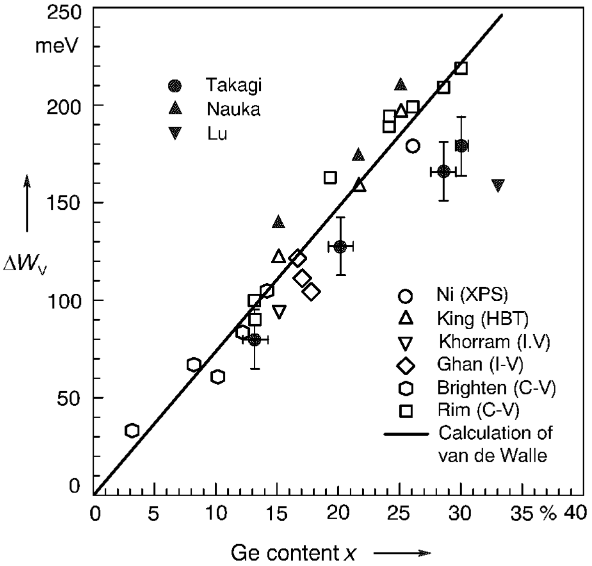


Fig. 4.16. Dependence of valence band discontinuity ΔW_V on Ge content x in strained Ge_xSi_{1-x} layers on a Si substrate as obtained from various measurements (after [36]; references to data points as cited therein)

in strained layers of Ge_xSi_{1-x} on Si has been calculated in [28, 33], where a small conduction band offset of about 20 meV has been found.

Electron and Hole Mobilities

The electron mobility of pure Ge exceeds that of pure Si by about a factor of four at low temperatures [5]. As already mentioned, in unstrained $Si_{1-x}Ge_x$ alloys, six equivalent conduction band minima occur along the $\langle 100\rangle$ directions up to a Ge-content $x = 0.85$, and for larger values of x the conduction band is "Ge-like" with eight equivalent conduction band minima along the $\langle 111\rangle$ directions. The electron mobility in unstrained $Si_{1-x}Ge_x$ alloys therefore decreases with the Ge content x up to $x = 0.85$ owing to alloy scattering; for larger values of x, an increase of μ_n towards the value for pure Ge at $x = 1$ is observed.

The sixfold degeneracy of the conduction band minimum and the anisotropy of the effective mass requires that for unstrained silicon, the electron mobility in the $\langle 100\rangle$ direction has to be calculated as a weighted average of the mobilities $\mu_{n\|}$ and $\mu_{n\perp}$, i.e.

$$\mu_n^{Si} = \frac{\mu_{n\|} + 2\mu_{n\perp}}{3},$$

where $\mu_{n\|}$ is the mobility of electrons in valleys oriented along the $\langle 100\rangle$ direction, and $\mu_{n\perp}$ is the mobility of electrons in valleys along the $\langle 010\rangle$ and $\langle 001\rangle$ directions (see Fig. 4.15). If we write $\mu_{n\perp}/\mu_{n\|} = m_l/m_t = B$, we can

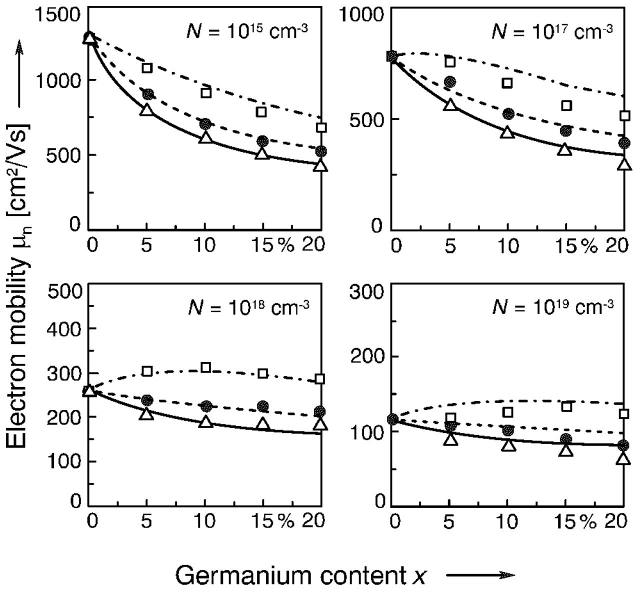


Fig. 4.17. Electron mobility of $Si_{1-x}Ge_x$ layers on Si substrate as a function of Ge content x for four different doping levels (after [37]). Curves describe the results of the expressions derived in [37], symbols denote the outcome of Monte Carlo simulations. The filled circles and dashed curves describe unstrained SiGe layers, the triangles and solid curves show the mobility in strained layers parallel to the growth interface, and the squares and dot–dashed curves show the mobility in strained layers orthogonal to the growth interface

express $\mu_{n||}$ and $\mu_{n\perp}$ in terms of μ_n as follows, as is done in Arora's mobility model [37]:

$$\mu_{n||} = \frac{3\mu_n}{2B+1} \quad \text{and} \quad \mu_{n\perp} = \frac{3B\mu_n}{2B+1} .$$

In an unstrained SiGe material these mobilities are reduced owing to alloy scattering, which can be taken into account using Matthiesen's rule to obtain the mobilities $\mu_{||}$ and $\mu_{\perp}$.

The strain inherent in a pseudomorphic SiGe layer affects the electron mobility. Owing to the strain, the energy of the two conduction band minima with wave vectors normal to the surface is increased, whereas the energy of the minima with wave vectors parallel to the interface is decreased in comparison with the unstrained situation. In a strained layer, the populations of the valleys in the $\langle 010 \rangle$ and $\langle 001 \rangle$ directions is different from the population of the valleys in the $\langle 100 \rangle$ direction, if the latter is the growth direction. This modifies the weighting factors used for the determination of the combined mobility of the six valleys, resulting in a mobility parallel to the growth plane given by

$$\mu_{n\|}^{\mathrm{SiGe}} = \frac{\left(\mu_\perp + \mu_\|\right)\exp\left(-\dfrac{\Delta W_x}{k_\mathrm{B}T}\right) + \mu_\perp \exp\left(-\dfrac{\Delta W_z}{k_\mathrm{B}T}\right)}{2\exp\left(-\dfrac{\Delta W_x}{k_\mathrm{B}T}\right) + \exp\left(-\dfrac{\Delta W_z}{k_\mathrm{B}T}\right)} \tag{4.47}$$

and a mobility perpendicular to the growth plane given by

$$\mu_{n\perp}^{\mathrm{SiGe}} = \frac{2\mu_\perp \exp\left(-\dfrac{\Delta W_x}{k_\mathrm{B}T}\right) + \mu_\| \exp\left(-\dfrac{\Delta W_z}{k_\mathrm{B}T}\right)}{2\exp\left(-\dfrac{\Delta W_x}{k_\mathrm{B}T}\right) + \exp\left(-\dfrac{\Delta W_z}{k_\mathrm{B}T}\right)} . \tag{4.48}$$

Here $\Delta W_x = -\Xi_\mathrm{u}\epsilon_\mathrm{T}/3$ and $\Delta W_z = 2\Xi_\mathrm{u}\epsilon_\mathrm{T}/3$, where $\Xi_\mathrm{u} \approx 9.2$ eV is the deformation potential constant for silicon, and the strain in the material is

$$\epsilon_\mathrm{T} = \frac{1+\nu}{1-\nu}\frac{a_\mathrm{SiGe}(x) - a_\mathrm{Si}}{a_\mathrm{SiGe}(x)} ,$$

where ν is the Poisson's ratio given by $\nu = 0.280 - 0.007x$ for $Si_{1-x}Ge_x$. These data have been used for the computation of the electron mobility in strained layers of SiGe (see Fig. 4.17), and good agreement with the outcome of Monte Carlo simulations was obtained [37].

The hole mobility in strained $Si_{1-x}Ge_x$ alloys has been investigated in [37]. The hole mobility parallel to the interface in a p-type strained $Si_{1-x}Ge_x$ layer with an acceptor concentration $N_\mathrm{A} = 10^{15}\ \mathrm{cm}^{-3}$, grown on $\langle 100 \rangle$ silicon, was found to increase by 50% for $x = 0.1$, by 80% for $x = 0.2$ and by 140% for $x = 0.3$. This increase was attributed to a decrease of the hole effective mass. At larger dopant densities, the mobility decreases owing to ionized-impurity scattering. The increased hole mobility decreases the internal base series resistance.

4.4 SiGe HBTs

Heterojunction bipolar transistors with a SiGe base layer on a silicon substrate became possible with the advent of SiGe growth techniques. As mentioned earlier, such HBTs generally use graded layers, as in Fig. 4.18, and can be described with a slightly modified Gummel–Poon model. The incorporation of low concentrations of carbon ($< 10^{20}\ \mathrm{cm}^{-3}$ into the SiGe region can substantially suppress boron outdiffusion during subsequent processing steps [38, 39]. This allows one to use high boron concentrations in very thin base layers with the consequence of improved high-frequency performance.

Silicon-based HBTs with a SiGe base layer are particularly suited for the realization of npn transistors since most of the bandgap offset occurs in the valence band, while the conduction band spike is of the order of a few meV. As compared with homojunction BJTs of comparable size, such HBTs offer several performance advantages:

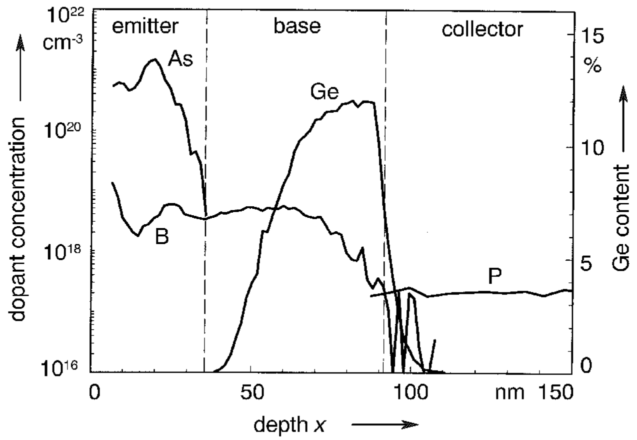


Fig. 4.18. Typical vertical doping profile and Ge content in the active region of a graded-base SiGe HBT(after [40])

- The transit time can be considerably reduced, owing to the drift field in the base region associated with the grading of the Ge content in the base layer; this results in an increased cutoff frequency f_{T}.
- The same value of the current gain can be achieved with a substantially increased number of acceptors in the base layer; this yields a considerable reduction of the internal base resistance, resulting in increased values of the maximum frequency of oscillation f_{max} and of the transconductance cutoff frequency f_{y}, as well as a reduced noise figure at high frequencies.
- With a heterojunction, an adequate current gain can be achieved without an oxide layer at the polysilicon emitter contact; this yields $1/f$-noise corner frequencies below 500 Hz [41].
- The Early voltage can be considerably increased; this improves the potential of the transistor as a current source.
- Si BJTs and SiGe HBTs show comparable base current levels at the same V_{BE} value, but different values of the transfer current, as is illustrated in Fig. 4.19: in a SiGe HBT the same transfer current is achieved with a reduced voltage, allowing a slight reduction of the supply voltage.

In this section, the physics and the electrical behavior of SiGe HBTs are considered in more detail.

4.4.1 Transfer Current

If velocity saturation effects in the bc diode are neglected ($v_{\mathrm{nsat}} \rightarrow \infty$), the transfer current–voltage characteristic for forward operation (3.40) reads

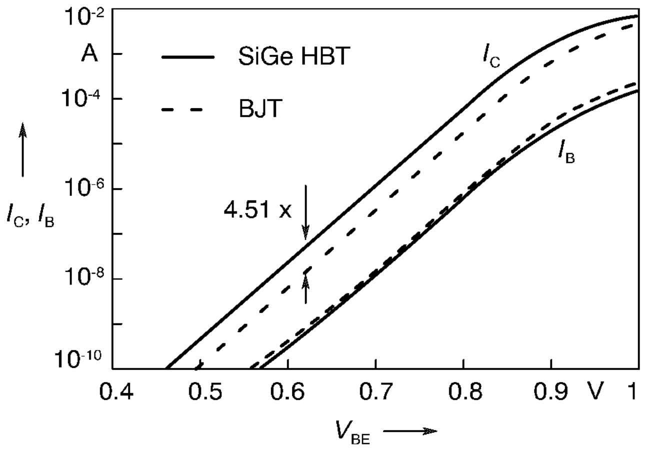


Fig. 4.19. Gummel Plot of graded-base SiGe HBT and Si BJT (after [4])

$$I_T = \frac{eA_{je}n_{i0}^2}{\int_{x_{be}}^{x_{bc}} \frac{p(x)}{D_n(x)} \frac{n_{i0}^2}{n_{ie}^2(x)} \, dx} \exp\left(\frac{V_{B'E'}}{V_T}\right) .$$

In the case of a homogeneously doped base, the approximation $p(x) \approx N_A$ may be used under low-level-injection conditions. If $\Delta W_g^{(1)}$ denotes the apparent bandgap narrowing associated with the dopant concentration N_A, and $\Delta W_g^{Ge}(x)$ denotes the bandgap reduction due to the Ge content as a function of position x, we may write

$$n_{ie}^2(x) = \gamma n_{i0}^2 \exp\left(\frac{\Delta W_g^{(1)}}{k_B T}\right) \exp\left(\frac{\Delta W_g^{Ge}(x)}{k_B T}\right) ,$$

where $\gamma = (N_C N_V)_{SiGe}/(N_C N_V)_{Si}$. The transfer current now reads

$$I_T = \frac{eA_{je}\langle \gamma D_n \rangle n_{i0}^2}{N_A \int_{x_{be}}^{x_{bc}} e^{-\Delta W_g^{Ge}(x)/k_B T} \, dx} \exp\left(\frac{\Delta W_g^{(1)}}{k_B T}\right) \exp\left(\frac{V_{B'E'}}{V_T}\right) , \quad (4.49)$$

where $\langle \gamma D_n \rangle$ denotes the mean value of γD_n. In the case of a trapezoidal Ge distribution, as illustrated in Fig. 4.20, the bandgap narrowing associated with the germanium is[15]

$$\Delta W_g^{Ge}(x) = \begin{cases} \Delta W_g^{Ge}(x_{bc}) + \Delta W_{g,\Delta}^{Ge} \dfrac{x - x_{be}}{x_T - x_{be}} , & x_{bc} \leq x \leq x_T \\ \Delta W_g^{Ge}(x_{be}) + \Delta W_{g,\Delta}^{Ge} , & x_T \leq x \leq x_{bc} \end{cases} ,(4.50)$$

[15] This is an approximation, as theoretical investigations indicate that the bandgap narrowing due to heavy doping is more pronounced in strained layers than in unstrained layers [44].

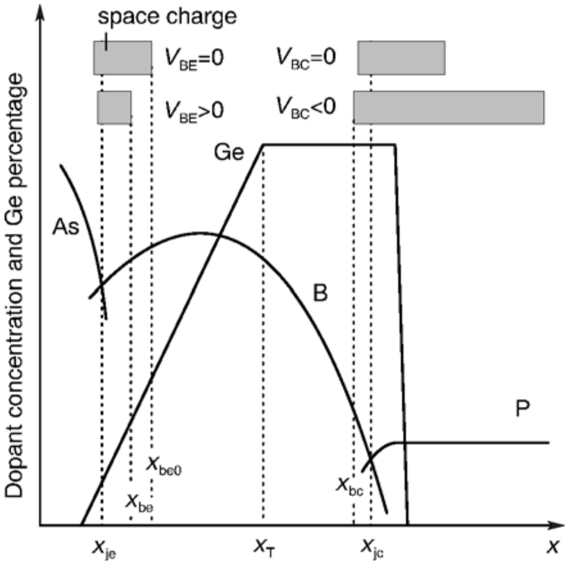


Fig. 4.20. Dopant concentration and Ge distribution in a graded-base SiGe HBT (schematic, after [42, 43])

where $\Delta W_{g,\Delta}^{Ge} = \Delta W_g^{Ge}(x_{bc}) - \Delta W_g^{Ge}(x_{be})$ denotes the bandgap reduction across the base layer due to the graded Ge profile. With this result, the integral in the denominator of (4.49) can be evaluated resulting in

$$I_T = \frac{eA_{je}\langle\gamma D_n\rangle n_{i0}^2}{N_A d_B} \exp\left(\frac{\Delta W_g^{(1)} + \Delta W_{g,eff}^{Ge}}{k_B T}\right) \exp\left(\frac{V_{B'E'}}{V_T}\right) ,$$

where $d_B = x_{bc} - x_{be}$ denotes the base width and

$$\exp\left(-\frac{\Delta W_{g,eff}^{Ge}}{k_B T}\right) = \frac{(1-\lambda)k_B T}{\Delta W_{g,\Delta}^{Ge}} \exp\left(-\frac{\Delta W_g^{Ge}(x_{be})}{k_B T}\right) + \left(\lambda - \frac{(1-\lambda)k_B T}{\Delta W_{g,\Delta}^{Ge}}\right) \exp\left(-\frac{\Delta W_g^{Ge}(x_{bc})}{k_B T}\right) ,$$

where $\lambda = (x_{bc} - x_T)/d_B$. The essential difference from the result for a homojunction is an increase of the transfer current by a factor $\exp\left(\Delta W_{g,eff}^{Ge}/k_B T\right)$, which also affects the temperature dependence of the device. This factor allows one to increase the denominator $N_A d_B$, resulting in a substantially reduced internal base resistance, as illustrated in Fig. 4.21. Since $\Delta W_{g,eff}^{Ge}$ represents an average over the base region, its value is bias-dependent and consequently affects the Early voltage of the device.

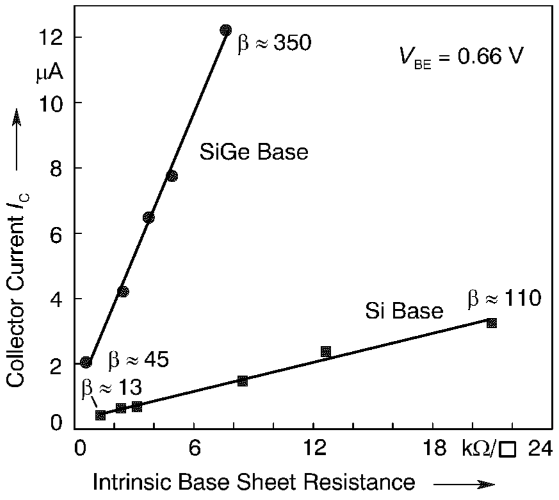


Fig. 4.21. Collector current density versus base sheet resistance of internal base layer for fixed values of V_{BE} (after [27,45])

Early Voltage

For a discussion of the Early effect in SiGe HBTs, we may use the general result

$$V_{AF} = \frac{eA_{je}n_{ie}^2(x_{bc})}{n_{i0}^2 c_{jc}(V_{C'B'})} \frac{G_B + \dfrac{p(x_{bc})}{v_{nsat}} \dfrac{n_{i0}^2}{n_{ie}^2(x_{bc})}}{\dfrac{1}{D_n(x_{bc})} + \dfrac{1}{v_{nsat}} \left.\dfrac{d\ln(p/n_{ie}^2)}{dx}\right|_{x_{bc}}} - V_{C'B'}\,, \qquad (4.51)$$

which has been derived from the transfer current relation given in Sect. 3.2. Since $V_{AF} \sim n_{ie}^2(x_{bc})$, it is possible to increase the Early voltage and hence the output resistance by increasing $n_{ie}(x_{bc})$, as is done in graded-base HBTs. Bandgap grading allows a substantial increase of the Early voltage, which then shows a substantial temperature dependence, as illustrated in Fig. 4.22.

If $p(x) \approx N_A$ and the diffusion coefficient D_n are assumed to be constant throughout the base region and if a trapezoidal Ge profile with $x_T < x_{bc}$ is assumed, this expression simplifies to

$$V_{AF} = \frac{eA_{je}N_A}{c_{jc}(V_{C'B'})} \left(\int_0^{d_B} \frac{n_{ie}^2(d_B)}{n_{ie}^2(x)}\, dx + \frac{D_n}{v_{nsat}} \right) - V_{C'B'}\,,$$

if the coordinate system is chosen such that $x_{be} = 0$ and $x_{bc} = d_B$. Assuming a Ge-induced bandgap narrowing according to (4.50) and $\gamma \approx 1$, the integral can be evaluated, resulting in

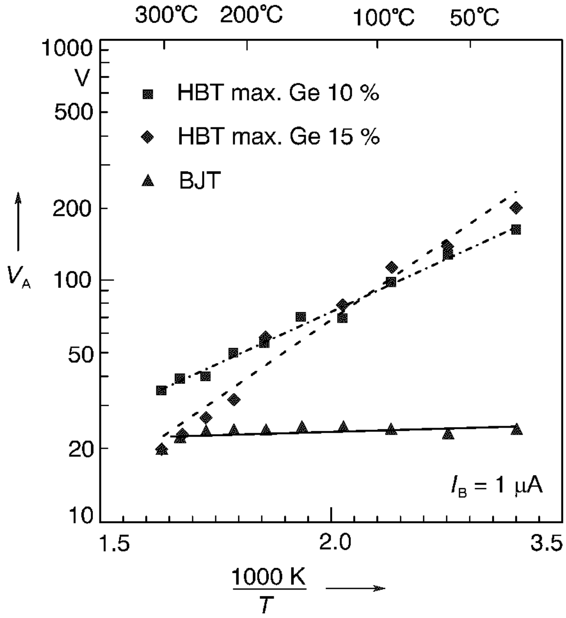


Fig. 4.22. Temperature dependence of extrapolated Early voltage V_A of BJTs and HBTs (after [3])

$$V_{AF} = \frac{eA_{je}N_A}{c_{jc}(V_{C'B'})}\left[\frac{D_n}{v_{nsat}} + d_B - x_T + \frac{k_BT}{\Delta W_{g,\Delta}^{Ge}} x_T \left(e^{\Delta W_{g,\Delta}^{Ge}/k_BT} - 1\right)\right]$$
$$- V_{C'B'} \,. \qquad (4.52)$$

The situation is different if $d_B < x_T$, i.e. if the bc depletion layer extends into the graded region. In this case

$$V_{AF} = \frac{eA_{je}N_A}{c_{jc}(V_{C'B'})} \frac{\displaystyle\int_0^{d_B} \frac{n_{ie}^2(d_B)}{n_{ie}^2(x)}\,dx + \frac{D_n}{v_{nsat}}}{1 + \displaystyle\frac{D_n n_{ie}^2(d_B)}{v_{nsat}} \frac{d}{dx}\frac{1}{n_{ie}^2}\bigg|_{d_B}} - V_{C'B'} \,.$$

Using the results

$$n_{ie}^2(d_B)\frac{d}{dx}\frac{1}{n_{ie}^2}\bigg|_{d_B} = -\frac{1}{x_T}\frac{\Delta W_{g,\Delta}^{Ge}}{k_BT}$$

and

$$\int_0^{d_B} \frac{n_{ie}^2(d_B)}{n_{ie}^2(x)}\,dx = x_T \frac{k_BT}{\Delta W_{g,\Delta}^{Ge}}\left[\exp\left(\frac{\Delta W_{g,\Delta}^{Ge}}{k_BT}\right) - 1\right] ,$$

the following result for the Early voltage can be derived:

$$V_{\mathrm{AF}} = \frac{eA_{\mathrm{je}}N_{\mathrm{A}}}{c_{\mathrm{jc}}(V_{\mathrm{C'B'}})} \times \frac{x_{\mathrm{T}} \dfrac{k_{\mathrm{B}}T}{\Delta W_{\mathrm{g},\Delta}^{\mathrm{Ge}}} \left[\exp\left(\dfrac{\Delta W_{\mathrm{g},\Delta}^{\mathrm{Ge}}}{k_{\mathrm{B}}T} \right) - 1 \right] + \dfrac{D_{\mathrm{n}}}{v_{\mathrm{nsat}}}}{1 - \dfrac{\Delta W_{\mathrm{g},\Delta}^{\mathrm{Ge}}}{k_{\mathrm{B}}T} \dfrac{D_{\mathrm{n}}}{v_{\mathrm{nsat}} x_{\mathrm{T}}}} - V_{\mathrm{C'B'}} \,.$$

From this result, even negative values of V_{AF} (corresponding to a negative

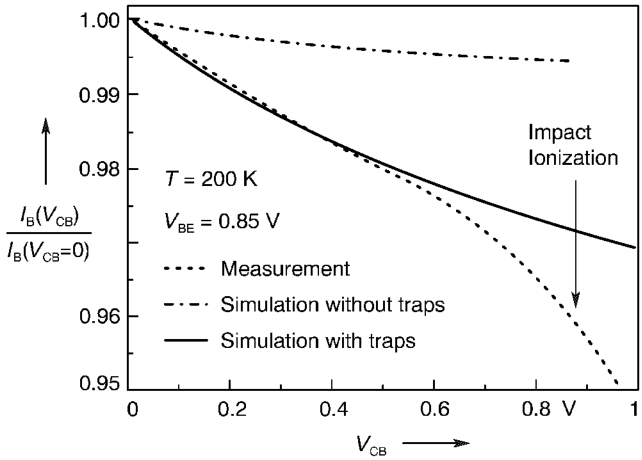


Fig. 4.23. Measured values of normalized base current for a fixed value of V_{BE} as a function of V_{CB} and simulated behavior with and without traps in the bc junction (after [46])

slope of the output characteristic under isothermal conditions) are expected for small values of x_{T}. The calculations above are derived from the modified drift–diffusion theory and apply to graded HBTs; in HBTs with abrupt heterojunctions, thermionic emission and neutral-base recombination will affect the output conductance of the device and thus the extrapolated Early voltage V_{A} [47].

At low temperatures, enhanced recombination in the base region or in the bc junction may affect the output conductance of SiGe HBTs, as was demonstrated in [46]. Here the measurements showed a decrease of the base current with V_{CB} at fixed V_{BE} before the onset of impact ionization (see Fig. 4.23); this would not be expected if recombination occurred only in the emitter region, as can generally be assumed in the case of high-frequency homojunction BJTs.

Temperature Dependence

The temperature dependence of the transfer current in graded-base SiGe HBTs depends on the Ge grading; its value is important in many analog circuits, such as bandgap references (see Chap. 6). Neglecting series resistances, the $V_{\mathrm{BE}}(T)$ characteristic can be obtained from the transfer current characteristic

$$I_{\mathrm{T}} = I_{\mathrm{T}}(\Delta W_{\mathrm{g}}=0) \exp\left(\frac{\Delta W_{\mathrm{g}}^{(1)}}{k_{\mathrm{B}}T}\right) \exp\left(\frac{\Delta W_{\mathrm{g,eff}}^{\mathrm{Ge}}}{k_{\mathrm{B}}T}\right) ,$$

which gives

$$V_{\mathrm{BE}}(T) = V_{\mathrm{BE}}[I_{\mathrm{T}}(\Delta W_{\mathrm{g}}=0)] - \frac{1}{e}\left(\Delta W_{\mathrm{g}}^{(1)} + \Delta W_{\mathrm{g,eff}}^{\mathrm{Ge}}\right)$$

i.e. the voltage drop across the eb diode is reduced by the effective bandgap narrowing in the base region due to the heavy doping and the germanium content. As $\Delta W_{\mathrm{g,eff}}^{\mathrm{Ge}}$ is determined by the location of the emitter-side depletion layer edge x_{be}, a change of the forward bias V_{BE} will result in a different offset voltage in the temperature characteristic of V_{BE}. This has negative effects on a special class of bandgap references [48].

4.4.2 Base Transit Time

Within the drift–diffusion approximation, the base transit time can be derived from the modified Krömer formula (3.106),

$$\tau_{\mathrm{Bf}} = \int_0^{d_{\mathrm{B}}} \frac{n_{\mathrm{ic}}^2(x)}{p(x)} \int_x^{d_{\mathrm{B}}} \frac{p(y)}{D_{\mathrm{n}}(y) n_{\mathrm{ie}}^2(y)} \,\mathrm{d}y\,\mathrm{d}x + \frac{1}{v_{\mathrm{nsat}}} \frac{p(d_{\mathrm{B}})}{n_{\mathrm{ie}}^2(d_{\mathrm{B}})} \int_0^{d_{\mathrm{B}}} \frac{n_{\mathrm{ie}}^2(x)}{p(x)} \,\mathrm{d}x ,$$

where the coordinates are chosen such that $x_{\mathrm{be}} = 0$ and $x_{\mathrm{bc}} = d_{\mathrm{B}}$. The above expression can be computed analytically for a trapezoidal Ge profile if $\gamma = 1$ and if $p(x) \approx N_{\mathrm{A}}$ and D_{n} are assumed to be constant throughout the base region.[16] In this case the integral that determines the forward base transit time can be split into a sum of four integrals

$$\tau_{\mathrm{Bf}} = \frac{1}{D_{\mathrm{n}}} \int_0^{x_{\mathrm{T}}} \int_x^{x_{\mathrm{T}}} \frac{n_{\mathrm{ie}}^2(x)}{n_{\mathrm{ie}}^2(y)} \,\mathrm{d}y\,\mathrm{d}x + \frac{1}{D_{\mathrm{n}}} \int_0^{x_{\mathrm{T}}} \int_{x_{\mathrm{T}}}^{d_{\mathrm{B}}} \frac{n_{\mathrm{ie}}^2(x)}{n_{\mathrm{ie}}^2(y)} \,\mathrm{d}y\,\mathrm{d}x + \frac{1}{D_{\mathrm{n}}} \int_{x_{\mathrm{T}}}^{d_{\mathrm{B}}} \int_x^{d_{\mathrm{B}}} \frac{n_{\mathrm{ie}}^2(x)}{n_{\mathrm{ic}}^2(y)} \,\mathrm{d}y\,\mathrm{d}x + \frac{1}{v_{\mathrm{nsat}}} \int_0^{d_{\mathrm{B}}} \frac{n_{\mathrm{ie}}^2(x)}{n_{\mathrm{ic}}^2(d_{\mathrm{B}})} \,\mathrm{d}x ,$$

which can be evaluated assuming a Ge-induced bandgap narrowing according to (4.50), to give

[16] This situation corresponds to a constant value of the quasi-electric field due to the bandgap grading throughout the base region. Corresponding results for the base transit time which take account of a position-dependent mobility have also been derived, for example in [49].

$$\tau_{\mathrm{Bf}} = \frac{1}{D_{\mathrm{n}}}\left(\frac{k_{\mathrm{B}}T}{\Delta W_{\mathrm{g},\Delta}^{\mathrm{Ge}}}(1-\theta)x_{\mathrm{T}}^2 + (d_{\mathrm{B}}x_{\mathrm{T}}-x_{\mathrm{T}}^2)\theta + \frac{(d_{\mathrm{B}}-x_{\mathrm{T}})^2}{2}\right)$$
$$+\frac{d_{\mathrm{B}}+(\theta-1)x_{\mathrm{T}}}{v_{\mathrm{nsat}}}\,, \tag{4.53}$$

where

$$\theta = \frac{k_{\mathrm{B}}T}{\Delta W_{\mathrm{g},\Delta}^{\mathrm{Ge}}}\left[1-\exp\left(-\frac{\Delta W_{\mathrm{g},\Delta}^{\mathrm{Ge}}}{k_{\mathrm{B}}T}\right)\right].$$

The formula (4.53) reduces to the result published in [50] in the limit $v_{\mathrm{nsat}} \to \infty$. In the limit $x_{\mathrm{T}} \to d_{\mathrm{B}}$, the forward base transit time is

$$\tau_{\mathrm{Bf}} = \frac{d_{\mathrm{B}}^2}{D_{\mathrm{n}}}\frac{k_{\mathrm{B}}T}{\Delta W_{\mathrm{g},\Delta}^{\mathrm{Ge}}}(1-\theta) + \frac{\theta d_{\mathrm{B}}}{v_{\mathrm{nsat}}}\,;$$

in the limit $v_{\mathrm{nsat}} \to \infty$, the result for a linearly graded base published in [27] is obtained. The optimum Ge profile in the base region for a given maximum value of the Ge content can be obtained by setting $\mathrm{d}\tau_{\mathrm{Bf}}/\mathrm{d}x_{\mathrm{T}}$ equal to zero; this results in the optimum value of x_{T}, given by

$$\frac{x_{\mathrm{T}}}{d_{\mathrm{B}}} = \frac{1-\theta}{1-2\theta+(2k_{\mathrm{B}}T/\Delta W_{\mathrm{g},\Delta}^{\mathrm{Ge}})(1-\theta)}\left(1+\frac{D_{\mathrm{n}}}{d_{\mathrm{B}}v_{\mathrm{nsat}}}\right).$$

In this optimization, the base width and thus the base sheet resistance have been taken to be constant; as the Gummel number changes with the value of x_{T}, different values of the forward current gain appear if x_{T} is changed. Numerical investigations [51] suggest that a graded Ge profile has only a small influence on the maximum cutoff frequency f_{Tmax} as long as the base Gummel number

$$G_{\mathrm{B}} = \int_{x_{\mathrm{be}}}^{x_{\mathrm{bc}}} \frac{p(x)}{D_{\mathrm{n}}(x)}\frac{n_{\mathrm{i0}}^2}{n_{\mathrm{ie}}^2(x)}\,\mathrm{d}x$$

remains constant.

4.4.3 High-Level-Injection Effects

At high injection levels, the electron flow into the collector may be reduced owing to the formation of a potential barrier at the bc heterojunction: the Kirk effect is more pronounced in SiGe HBTs, resulting in a nearly abrupt degradation of transistor performance above a critical current level. The reason for this is the valence band offset at the SiGe/Si heterointerface in the bc junction. This offset has a negligible effect on the device characteristics under low-level injection, when $J < J_{\mathrm{K}}$. After the onset of the Kirk effect, however, the space charge region in the junction shifts towards the subcollector. In the case of a conventional BJT, holes are injected into the collector region (see Sect. 3.7). However, as the heterojunction blocks hole injection into the collector, holes pile up at the heterointerface and induce a conduction band barrier that opposes the electron transfer current and therefore degrades the transconductance. Furthermore, a strong increase of the stored minority charge in the base region causes both a decrease of f_{T} and f_{max} and an increase of the base recombination current, with the consequence of a substantial reduction of the current gain.

This effect cannot be eliminated completely, but a careful optimization of the Ge and dopant distribution in the bc junction allows one to delay the onset of the effect to current densities above those required for normal circuit operation [52]. Several constraints have to be taken into account when this optimization is done:

- Moving the heterojunction more deeply into the bc depletion layer or using a graded heterojunction at the bc junction (retrograding) allows one to shift the onset of high-level-injection effects to a larger current density. As these methods increase the thickness of the strained layer, they cause a possible film stability problem. Furthermore, a decrease of the breakdown voltage is expected if the heterojunction is shifted too far into the space charge layer.
- The possibility of increasing the collector doping level in order to shift the Kirk current density to a larger value is limited by the negative impact on the open-base breakdown voltage BV_{CEO}.

A model equation that considers the barrier effect on the forward transit time τ_{f} has been derived in [53].[17] The density of minority carriers in the pushed-out base is assumed to be large in comparison with the density of donors N_{D}, i.e. $n \approx p \gg N_{\mathrm{D}}$. The generalized law of mass action then gives

$$np \approx n^2 \approx n_{\mathrm{i0}}^2 \exp\left(\frac{\Delta W_{\mathrm{g}}}{k_{\mathrm{B}}T}\right) \exp\left(\frac{\Delta W_{\mathrm{F}}}{k_{\mathrm{B}}T}\right) \tag{4.54}$$

where $\Delta W_{\mathrm{F}} = W_{\mathrm{Fn}} - W_{\mathrm{Fp}}$ is the difference of the electron and hole quasi-Fermi levels. If the hole current density is assumed to be small in the epitaxial layer, the electron current density equation gives

$$\frac{\mathrm{d}}{\mathrm{d}x}\Delta W_{\mathrm{F}} = \frac{J_{\mathrm{n}}}{\mu_{\mathrm{n}} n} .$$

Taking the derivative of the root of (4.54) the two equations can be combined with the result:

$$\frac{1}{2k_{\mathrm{B}}T}\frac{J_{\mathrm{n}}}{\mu_{\mathrm{n}}} = \frac{\mathrm{d}n}{\mathrm{d}x} - \frac{n}{2k_{\mathrm{B}}T}\frac{\mathrm{d}\Delta W_{\mathrm{g}}}{\mathrm{d}x} . \tag{4.55}$$

[17] A model for HBTs with an abrupt heterojunction at the pn transition has been presented in [54].

An increase of the electron current density to $J_n + \Delta J_n$ causes an increase of the electron density to $n(x) + \Delta n(x)$; Δn and ΔJ_n are related by

$$\frac{1}{2k_BT}\frac{J_n+\Delta J_n}{\mu_n} = \frac{d}{dx}(n + \Delta n) - \frac{n+\Delta n}{2k_BT}\frac{d\Delta W_g}{dx} . \tag{4.56}$$

The difference of (4.56) and (4.55) yields

$$\frac{1}{2k_BT\mu_n} = \frac{d}{dx}\left(\frac{\Delta n}{\Delta J_n}\right) - \frac{1}{2k_BT}\left(\frac{\Delta n}{\Delta J_n}\right)\frac{d\Delta W_g}{dx} .$$

Since $dW_g/dx = 0$ for $x > x_T$, this differential equation has the solution

$$\frac{\Delta n}{\Delta J_n} = -\frac{1}{ev_{nsat}} + \frac{x - x_b}{2k_BT\mu_n}$$

in the range $x_{bar} < x < x_b$ if the boundary condition $J_n = -env_{sat}$ is applied at $x = x_b$. At the barrier the electron density is increased by the Boltzmann factor, i.e.

$$\left.\frac{\Delta n}{\Delta J_n}\right|_{x_{bar}^-} = \left[-\frac{1}{ev_{nsat}} + \frac{x_{bar} - x_b}{2k_BT\mu_n}\right]\exp\left(\frac{W_{bar}}{k_BT}\right) ,$$

where W_{bar} denotes the height of the potential barrier. In the case of a trapezoidal Ge profile, the Ge content is constant in the region between x_T and x_{bar}. Therefore

$$\frac{\Delta n}{\Delta J_n} = \left.\frac{\Delta n}{\Delta J_n}\right|_{x_{bar}^-} + \frac{x - x_{bar}}{2k_BT\mu_n}$$

in this region. If x_T is assumed to be close to the metallurgical junction $x_{jc} = 0$, the increase of the transit time can be estimated from

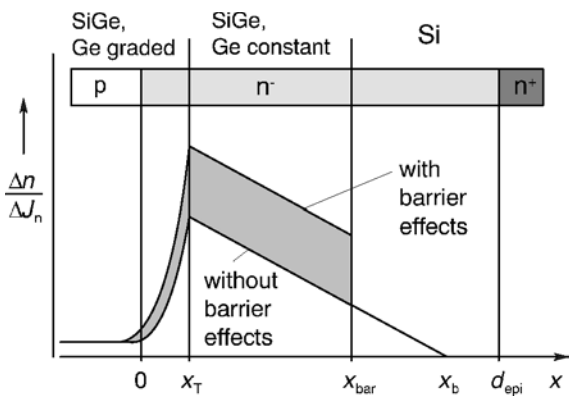


Fig. 4.24. Illustration of $\Delta n/\Delta J_n$ as a function of position for a SiGe HBT both with and without barrier effect present. x_T is where the Ge profile becomes constant, x_{bar} is the location of the heterojunction, and x_b is the edge of the extended base (after [53])

$$
\begin{aligned}
\Delta\tau_{\mathrm{B}} &\approx -e\int_0^{x_{\mathrm{bar}}}\left(\left.\frac{\Delta n}{\Delta J_{\mathrm{n}}}\right|_{W_{\mathrm{bar}}} - \left.\frac{\Delta n}{\Delta J_{\mathrm{n}}}\right|_{W_{\mathrm{bar}}=0}\right)\mathrm{d}x \\
&= x_{\mathrm{bar}}\left[\frac{1}{v_{\mathrm{nsat}}} + \frac{x_{\mathrm{b}} - x_{\mathrm{bar}}}{2D_{\mathrm{n}}}\right]\exp\left(\frac{W_{\mathrm{bar}}}{k_{\mathrm{B}}T}\right)
\end{aligned}
$$

if $x_{\mathrm{b}} > x_{\mathrm{bar}}$ and $D_{\mathrm{n}} = V_{\mathrm{T}}\mu_{\mathrm{n}}$. The value of x_{b} is determined by the collector current density and can be written as

$$
x_{\mathrm{b}}(I_{\mathrm{C}}) = \begin{cases} 0, & J_{\mathrm{C}} \leq I_{\mathrm{HC}} \\ d_{\mathrm{epi}}(1 - I_{\mathrm{C}}/I_{\mathrm{HC}})\ , & I_{\mathrm{C}} > I_{\mathrm{HC}} \end{cases} ,
$$

where I_{HC} is the Kirk current and d_{epi} denotes the thickness of the epilayer. With the increase of the transit time can be written as

$$
\Delta\tau_{\mathrm{b}} = \begin{cases} 0\ , & I_{\mathrm{C}} \leq I_{\mathrm{bar}} \\ \tau_{\mathrm{bb}}(1 - I_{\mathrm{bar}}/I_{\mathrm{C}})\ , & I_{\mathrm{C}} > I_{\mathrm{bar}} \end{cases} \tag{4.57}
$$

where

$$
\tau_{\mathrm{bb}} = x_{\mathrm{bar}}\left(\frac{d_{\mathrm{epi}} - x_{\mathrm{bar}}}{2D_{\mathrm{n}}} + \frac{1}{v_{\mathrm{nsat}}}\right)\left[\exp\left(\frac{W_{\mathrm{bar}}}{k_{\mathrm{B}}T}\right) - 1\right] \tag{4.58}
$$

$$
I_{\mathrm{bar}} = \frac{I_{\mathrm{HC}}}{1 - \dfrac{x_{\mathrm{bar}}}{d_{\mathrm{epi}}} + \dfrac{2D_{\mathrm{n}}}{v_{\mathrm{nsat}}d_{\mathrm{epi}}}} \tag{4.59}
$$

The model introduces two additional parameters τ_{bb} and I_{bar} to fit experimental data and was found to improve the description of transistors with an SiGe base region; Fig. 4.25 demonstrates the improvements on the standard HICUM model.[18]

4.4.4 Compact Models for SiGe HBTs

A compact model for SiGe HBTs based on the integral charge control relation has been presented in [55,56], where effects of thermionic emission in abrupt heterojunction transistors are neglected because there is only a small spike in the conduction band. Compact models for HBTs must differ from those for conventional bipolar transistors in several respects:

- The temperature dependence is different, owing to the different values of the bandgap in the base and emitter regions.
- The modeling of the Early effect has to be different, owing to the effect of the intrinsic carrier density at the bc depletion layer edge on the output conductance.

[18] The comparison considers SiGe HBTs with a reduced collector doping ("high-breakdown HBTs"); the effect is less pronounced if the doping concentration in the collector region is increased.

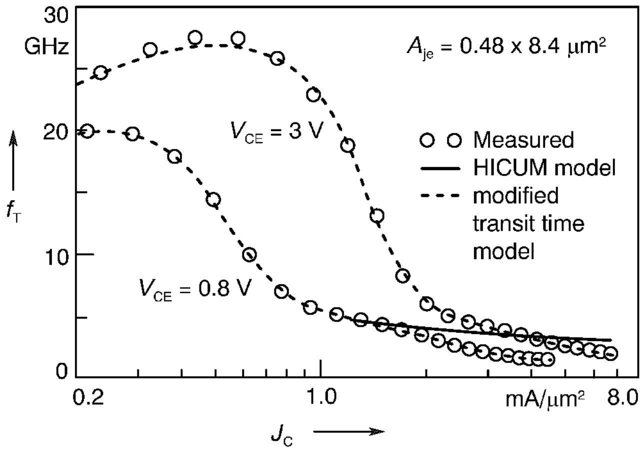


Fig. 4.25. Cutoff frequency f_T versus collector current density J_C for high-breakdown SiGe HBT (after [53])

- High-level-injection effects have to be taken into account in a different manner, owing to the base–collector heterojunction.

MEXTRAM provides two features to model SiGe HBTs [57]:

- The effects of a graded base layer are described in terms of the parameter ΔE_g, which modifies the description of the normalized base charge under low-level-injection conditions:

$$q_1 = \frac{\exp\left[\left(\frac{v_{Te}}{V_{er}}+1\right)\frac{\Delta E_g}{V_T}\right] - \exp\left(-\frac{v_{Tc}}{V_{ef}}\frac{\Delta E_g}{V_T}\right)}{\exp\left(\frac{\Delta E_g}{V_T}\right) - 1} \tag{4.60}$$

  with voltages v_{Te} and v_{Tc} as defined in Sect. 3.16. Formula (4.60) is derived from the Gummel number

$$G_B = \frac{N_A}{D_n}\int_{x_{bc}}^{x_{bc}} \frac{n_{i0}^2}{n_{ic}^2(x)}\,dx$$

  by assuming homogeneous doping and a linear grading of the bandgap with the bandgap difference ΔE_g across the zero-bias base width d_B. Formula (4.60) reduces to $q_1 = v_{Te}/V_{er} + V_{Tc}/V_{ef}$ in the limit $\Delta E_g \to 0$.

- In some SiGe process the current component I_{BB} owing to neutral-base recombination may be nonnegligible. This may be caused by the high doping density in the base layer with the consequence of increased Auger recombination or by barrier effects at the bc junction. If I_B is affected by I_{BB} a decrease of I_B with increasing values of v_{Tc} is observed. Neutral-base recombination under low-level-injection conditions is described in terms of the parameter X_{REC}, as follows:

$$I_{\mathrm{B}} = \frac{I_{\mathrm{S}}}{B_{\mathrm{F}}} \left[\exp\left(\frac{v_{\mathrm{B2E1}}}{V_{\mathrm{T}}} \right) - 1 \right] \left(1 + X_{\mathrm{REC}} \frac{v_{\mathrm{Tc}}}{V_{\mathrm{EF}}} \right) .$$

For high-level injection conditions with a nonnegligible electron concentration at the bc junction, this formula is generalized to

$$\begin{aligned} I_{\mathrm{B}} = \frac{I_{\mathrm{S}}}{B_{\mathrm{F}}} \Bigg\{ & (1 - X_{\mathrm{REC}}) \left[\exp\left(\frac{v_{\mathrm{B2E1}}}{V_{\mathrm{T}}} \right) - 1 \right] \\ & + X_{\mathrm{REC}} \left[\exp\left(\frac{v_{\mathrm{B2E1}}}{V_{\mathrm{T}}} \right) + \exp\left(\frac{v^*_{\mathrm{B2C2}}}{V_{\mathrm{T}}} \right) - 2 \right] \left(1 + X_{\mathrm{REC}} \frac{v_{\mathrm{Tc}}}{V_{\mathrm{EF}}} \right) \Bigg\} . \end{aligned}$$

4.5 Compound Semiconductor HBTs

Silicon-based HBTs are of particular interest for large-scale integration at a moderate cost level. Alternative heterostructures based on compound semiconductors may offer specific advantages because of a larger electron mobility or saturated drift velocity or because of a different value of the bandgap. As the focus of this book is on silicon-based devices, this subject will only be considered briefly.[19]

Heterojunction bipolar transistors using wide-bandgap III–V semiconductor materials such as $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{N/GaN}$ are under investigation for high-power, high-frequency applications[20] [64–67]. The large bandgaps of these semiconductor materials, in excess of 3 eV, allows one to operate such devices at high temperatures and voltages, as both the intrinsic carrier density and the ionization coefficient are small. There are, however, some technological problems, such as the incomplete ionization of the Mg dopant used in the GaN base layer, to be overcome.

4.5.1 GaAlAs/GaAs HBTs

The GaAlAs/GaAs material system has some excellent features:

- The GaAlAs/GaAs material system provides an excellent lattice match: the difference between the lattice constants of AlAs and GaAs is only about 0.14% at room temperature, and thus coherently matched heterostructures are possible for any Al content.
- The freedom in the alloy composition due to the nearly perfect lattice matching allows one to realize substantial bandgap differences. The bandgap of $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$ is given by

$$\frac{W_{\mathrm{g}}(x)}{\mathrm{eV}} = \begin{cases} 1.424 + 1.247\,x & \text{for} \quad x < 0.45 \\ 1.424 + 1.27\,x + 1.147(x - 0.45)^2 & \text{for} \quad x > 0.45 \end{cases} .$$

[19] See, for example, [25, 58–62].

[20] A comparison of various candidates for the fabrication of power HBTs is given in [63].

Below $x = 0.45$ the semiconductor alloy has a direct bandgap, for larger values of x the bandgap is indirect, with conduction band minima at the X points of the Brillouin zone.

- The small effective mass associated with the direct conduction band minimum results in a large electron mobility, close to 8000 cm^2/Vs in pure GaAs.
- The direct conduction band minimum allows one to integrate light-emitting devices such as LEDs or laser diodes with circuits constructed with AlGaAs/GaAs HBTs.
- Owing to the wide bandgap, GaAs has a small intrinsic carrier density and thus a substantial specific resistance. "Semi-insulating" GaAs substrates obtained from intrinsic GaAs result in extremely small collector–substrate capacitances and reduce the coupling between devices and their interconnects to the substrate. This is of particular importance in microwave monolithic integrated circuits (MMICs), which frequently employ wide interconnects in order to control the impedance and to limit attenuation.

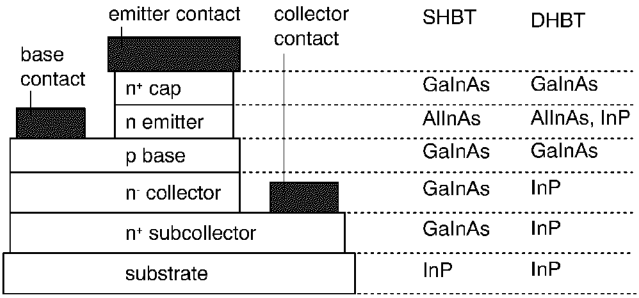


Fig. 4.26. Layers in single- (SHBT) and double-hetero-junction (DHBT) bipolar transistors on an InP substrate (after [68])

4.5.2 Indium Phosphide

Indium phosphide provides excellent electron mobility with a saturated drift velocity that exceeds the value for silicon by nearly a factor of three, a base–emitter voltage that lies some 200 mV below the corresponding value for a Si BJT, and a larger value of the bc breakdown voltage. These benefits are somewhat reduced by the smaller value of the thermal conductivity, which limits the maximum allowable power dissipation [68]. Figure 4.26 shows a schematic illustration of the layers typically employed in single- and double-heterojunction bipolar transistors (SHBTs and DHBTs, respectively) on an InP substrate. Using InAs ($W_g = 1.45\,\text{eV}$) or InP ($W_g = 1.35\,\text{eV}$) as emitter material on a base layer formed from GaInAs (e.g. with $W_g = 0.75\,\text{eV}$) yields a wide-gap emitter and a small turn-on voltage owing to the small bandgap of the base layer (Fig. 4.27). DHBTs formed on top of a wide-gap collector layer (e.g. InP) benefit additionally from the high breakdown field of InP, which allows one to increase the doping and reduce the thickness of the n^- collector layer, resulting in an increased cutoff frequency for a given value of the breakdown voltage: InP/InGaAs HBTs show larger values of[21] $f_T BV_{CEO}$ than do SiGe HBTs.

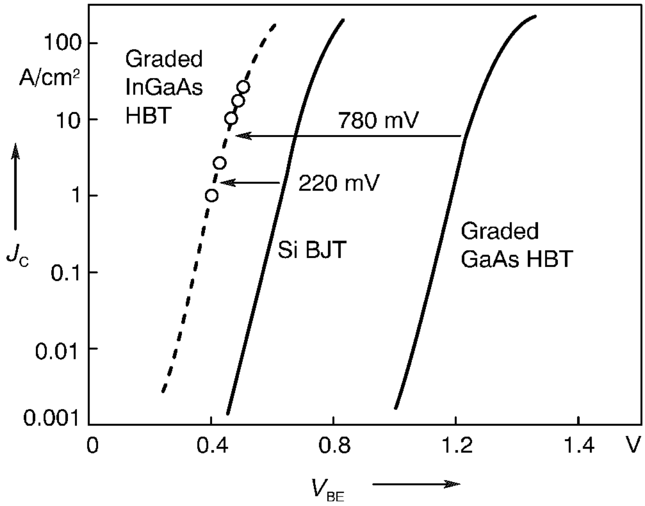


Fig. 4.27. Collector-current density J_C versus base–emitter voltage V_{BE} for different npn bipolar transistor technologies

Laboratory samples of InP/InGaAs HBTs grown by MOCVD showing a cutoff frequency $f_T = 161$ GHz and a maximum frequency of oscillation $f_{max} = 167$ GHz with a 2 μm $\times$ 10 μm emitter were demonstrated in [69]. An IC process for an HBT with an InGaP emitter, developed for RF and microwave instrumentation with minimum lateral dimensions of 1 μm, showing a cutoff frequency $f_T = 65$ GHz, a maximum frequency of oscillation $f_{max} = 75$ GHz, a current gain $\beta = 132$ and an open-base breakdown voltage $BV_{CEO} = 8.3$ V, is described in [70]. Values of f_T and f_{max} in excess of 500 GHz seem be achievable with 0.2 μm lithography [68].

4.5.3 Microwave Power Transistors

Microwave power transistors are intended to deliver large amounts of power at high frequencies. Such devices are generally realized as multiemitter transistors (Fig. 4.28), which cover a substantial area of the chip surface.[22] This may lead to an inhomogeneous or even unstable current distribution among the various emitter fingers (Fig. 4.29), a problem that can be overcome by

[21] The quantity $f_T BV_{CEO}$ serves as a figure of merit to characterize the trade-off between breakdown and speed (see Sect. 3.10).

[22] Design rules for AlGaAs/GaAs HBTs for power applications are discussed in [71].

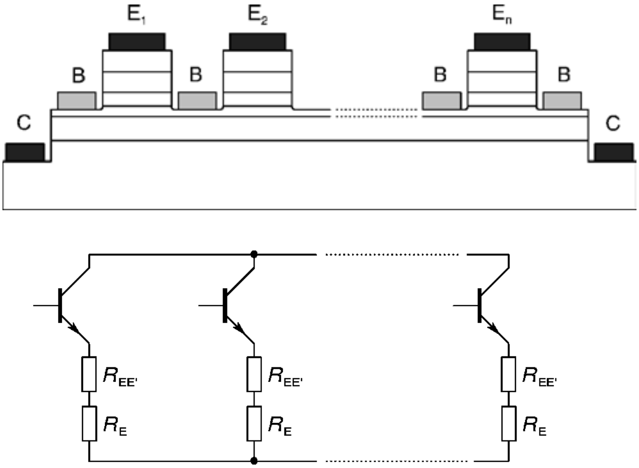


Fig. 4.28. Cross section (schematic) and simplified equivalent circuit of a power HBT with n emitter fingers (after [72])

adding an emitter ballasting resistor [73, 74] in series with each emitter finger (Fig. 4.28). The transistor may then be considered as a parallel connection of N transistors with emitter series resistors. If $R = R_{\mathrm{E}} + R_{\mathrm{EE'}}$ denotes the combination of the (internal) emitter resistance $R_{\mathrm{EE'}}$ and the emitter ballasting resistance R_{E}, a change of the junction temperature T_{j} or the emitter current I_{E} causes a change

$$\mathrm{d}V_{\mathrm{BE}} = \left(\frac{\partial V_{\mathrm{BE'}}}{\partial I_{\mathrm{E}}}\right)_T \mathrm{d}I_{\mathrm{E}} + \left(\frac{\partial V_{\mathrm{BE'}}}{\partial T_{\mathrm{j}}}\right)_{I_{\mathrm{E}}} \mathrm{d}T_{\mathrm{j}} + R\,\mathrm{d}I_{\mathrm{E}} + I_{\mathrm{E}}\,\frac{\mathrm{d}R}{\mathrm{d}T_{\mathrm{j}}}\,\mathrm{d}T_{\mathrm{j}}$$

of the voltage drop $V_{\mathrm{BE}} = V_{\mathrm{BE'}} + RI_{\mathrm{E}}$. To achieve stability with respect to temperature and/or current fluctuations, the value of $\mathrm{d}V_{\mathrm{BE}}$ must be positive [72]; the minimum allowable value of the emitter ballasting resistor is obtained from the condition $\mathrm{d}V_{\mathrm{BE}} = 0$, corresponding to

$$\left(1 + \alpha_{\mathrm{R}} I_{\mathrm{E}}\,\frac{\mathrm{d}T_{\mathrm{j}}}{\mathrm{d}I_{\mathrm{E}}}\right) R = -\left(\frac{\partial V_{\mathrm{BE'}}}{\partial I_{\mathrm{E}}}\right)_T - \left(\frac{\partial V_{\mathrm{BE'}}}{\partial T_{\mathrm{j}}}\right)_{I_{\mathrm{E}}} \frac{\mathrm{d}T_{\mathrm{j}}}{\mathrm{d}I_{\mathrm{E}}}\,,$$

where $\mathrm{d}R/\mathrm{d}T_{\mathrm{j}} = \alpha_{\mathrm{R}} R$ has been expressed in terms of the temperature coefficient α_{R}. Taking account of the relation

$$T_{\mathrm{j}} - T_{\mathrm{A}} \approx R_{\mathrm{th}} V_{\mathrm{CE}} I_{\mathrm{E}}\,,$$

one obtains

$$\frac{\mathrm{d}T_{\mathrm{j}}}{\mathrm{d}I_{\mathrm{E}}} = \frac{\mathrm{d}R_{\mathrm{th}}}{\mathrm{d}T_{\mathrm{j}}}\frac{\mathrm{d}T_{\mathrm{j}}}{\mathrm{d}I_{\mathrm{E}}} V_{\mathrm{CE}} I_{\mathrm{E}} + R_{\mathrm{th}} V_{\mathrm{CE}} = \frac{R_{\mathrm{th}} V_{\mathrm{CE}}}{1 - \alpha_{\mathrm{Rth}} R_{\mathrm{th}} V_{\mathrm{CE}} I_{\mathrm{E}}}$$

if V_{CE} is constant; here $\mathrm{d}R_{\mathrm{th}}/\mathrm{d}T_{\mathrm{j}} = \alpha_{\mathrm{Rth}} R_{\mathrm{th}}$ has been written in terms of the temperature coefficient α_{Rth} of the thermal resistance. Combining equations yields the following for the minimum emitter ballasting resistance:

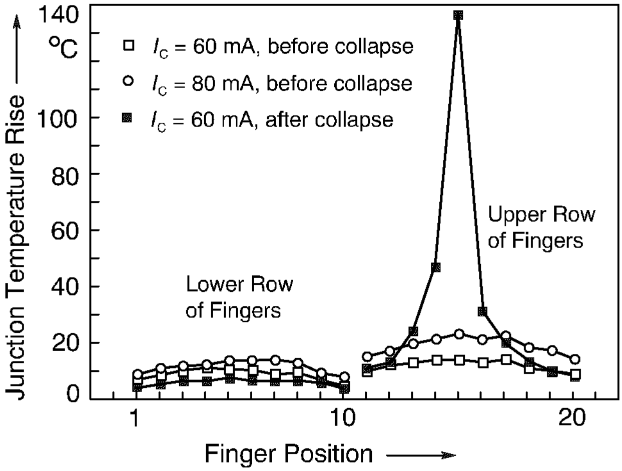


Fig. 4.29. Distribution of excess temperature in a 20-finger power HBT (2 × 10 fingers with a central base feedthrough) measured at V_{CE} = 6 V with a precision infrared scope (after [75])

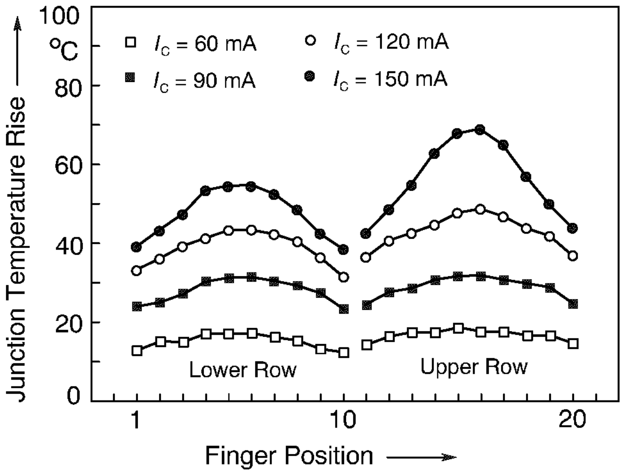


Fig. 4.30. Distribution of excess temperature in a 20-finger power HBT with emitter ballasting resistors, measured at V_{CE} = 6 V with a precision infrared scope (after [75])

$$R_{\mathrm{Emin}} = -\frac{R_{\mathrm{th}}V_{\mathrm{CE}}\left(\dfrac{\partial V_{\mathrm{BE'}}}{\partial T_{\mathrm{j}}}\right)_{I_{\mathrm{E}}} + (1-\alpha_{\mathrm{Rth}}R_{\mathrm{th}}V_{\mathrm{CE}}I_{\mathrm{C}})\left(\dfrac{\partial V_{\mathrm{BE'}}}{\partial I_{\mathrm{E}}}\right)_T}{1+(\alpha_{\mathrm{R}}-\alpha_{\mathrm{Rth}})R_{\mathrm{th}}V_{\mathrm{CE}}I_{\mathrm{E}}} - R_{\mathrm{EE'}}\,.$$

This result takes account of the variation with temperature of both the emitter ballasting resistance and the thermal resistance; thermal coupling between neighboring emitter fingers is not considered.

In Sect. 3.11, the following condition for thermal instability was derived[23]

[23] If the thermal resistance is assumed to be temperature-dependent, the condition should be modified to [76]

$$R_{\mathrm{th}}V_{\mathrm{CE}}\left[\left(\frac{\partial I_{\mathrm{C}}}{\partial T}\right)_{V_{\mathrm{CE}}} + \frac{I_{\mathrm{C}}}{R_{\mathrm{th}}}\frac{\mathrm{d}R_{\mathrm{th}}}{\mathrm{d}T}\right] \to 1\,.$$

$$R_{\rm th} V_{\rm CE} \left(\frac{\partial I_{\rm C}}{\partial T}\right)_{V_{\rm CE}} \to 1 \, .$$

Since

$$\left(\frac{\partial I_{\rm C}}{\partial T}\right)_{V_{\rm CE}} = \frac{1}{1 + R g_{\rm m}} \frac{I_{\rm C}}{T} \left(\frac{V_{\rm g} - V_{\rm BE}}{V_{\rm T}} + X_{\rm TI}\right) ,$$

where $R = R_{\rm EE'} + R_{\rm E}$ denotes the effective emitter series resistance, applies when $V_{\rm BE}$ is constant, the instability criterion results in the following condition for the maximum allowable current per emitter finger:

$$R_{\rm th} V_{\rm CE} \frac{1}{1 + R I_{\rm C}/V_{\rm T}} \frac{I_{\rm C}}{T} \left(\frac{V_{\rm g} - V_{\rm BE}}{V_{\rm T}} + X_{\rm TI}\right) = 1 \, . \tag{4.61}$$

The temperature T and the temperature voltage $V_{\rm T}$ at the operating temperature differ from the values $T_{\rm A}$ and $V_{\rm TA}$ determined by the ambient temperature owing to self-heating

$$T = T_{\rm A} + R_{\rm th} V_{\rm CE} I_{\rm C} \quad \text{and} \quad V_{\rm T} = V_{\rm TA} + \frac{k_{\rm B}}{e} R_{\rm th} V_{\rm CE} I_{\rm C} \, .$$

If this is introduced into (4.61) a quadratic equation for the maximum allowable current per emitter finger results. A simplified solution gives the following estimate for the maximum allowable current for a power transistor with n emitter fingers:

$$I_{\rm Cmax} = n \frac{V_{\rm TA}}{R_{\rm th} V_{\rm CE} \, \Theta - R} , \tag{4.62}$$

where[24]

$$\Theta = \frac{k_{\rm B}}{e} \left(\frac{V_{\rm g} - V_{\rm BE}}{V_{\rm T}} + X_{\rm TI}\right) - n \frac{k_{\rm B}}{e} \, .$$

Experimental investigations of the instability criterion and its dependence on the substrate temperature have been published in [76].

The uniformity of the current distribution in multiemitter transistors can be improved substantially by using emitter ballasting resistors, as is shown by comparison of Figs. 4.29 and 4.30. If instabilities of the current distribution are avoided, the current-carrying capability is limited by high-level-injection effects in the collector region. A positive value of the temperature coefficient of the ballasting resistor and a negative value of the temperature coefficient of the thermal resistance reduce the minimum value of the ballasting resistance required.

---

[24] If $R_{\rm th}$ is temperature-dependent, the relation for Θ should be modified to [76]

$$\Theta = \frac{k_{\rm B}}{e} \left(\frac{V_{\rm g} - V_{\rm BE}}{V_{\rm T}} + X_{\rm TI}\right) \left(1 - \frac{T - T_{\rm A}}{R_{\rm th}} \frac{{\rm d}R_{\rm th}}{{\rm d}T}\right)^{-1} - n \frac{k_{\rm B}}{e} \, .$$

Carrier multiplication in the bc diode may have a substantial effect on the current distribution. At a given value of V_{CB}, the base current that flows into the device is given by $[1-B'_{\mathrm{N}}(M_{\mathrm{n}}-1)]\,I_{\mathrm{BE}}$, where I_{BE} denotes the base current injected into the emitter and $B'_{\mathrm{N}} = I_{\mathrm{T}}/I_{\mathrm{BE}}$, and the base current component due to recombination in the base region I_{BB} has been neglected. If B'_{N} and M_{n} decrease with temperature, as is the case for AlGaAs/GaAs HBTs, "cold" fingers may have a negative base current, i.e. produce holes, which acts as an additional base current component for "hot" fingers, which as a consequence carry a larger current and heat up even further [77]. This mechanism has to be taken into account if current gain collapse occurs at V_{CE} values near the open-base breakdown voltage BV_{CEO}.

4.6 References

[1] H. Krömer. Theory of a wide-gap emitter for transistors. *Proc. IRE*, 45(11):1535–1537, 1957.

[2] H. Krömer. Heterostructure bipolar transistors and integrated circuits. *Proc. IEEE*, 70(1):13–25, 1982.

[3] M. Pohl, K. Aufinger, J. Böck, T.F. Meister, H. von Philipsborn. DC and AC performance of Si and Si/$\mathrm{Si}_{1-x}\mathrm{Ge}_x$ bipolar transistors at temperatures up to 300°C. *Proc. ESSDERC*, 28:100–103, 1998.

[4] D.L. Harame, J.H. Comfort, J.D. Cressler, E.F. Crabbe, J.Y.-C. Sun, B.S. Meyerson, T. Tice. Si/SiGe epitaxial-base transistors – part i: materials, physics, and circuits. *IEEE Trans. Electron Devices*, 42(3):455–468, 1995.

[5] H. Morkoc, B. Sverdlov, G.-B. Gao. Strained layer heterostructures, and their applications to MODFET's, HBT's and lasers. *Proc. IEEE*, 81(4):493–556, 1993.

[6] J.W. Slotboom, G. Streutker, A. Pruijmboom, D.J. Gravesteijn. Parasitic energy barriers in SiGe HBTs. *IEEE Electron Device Lett.*, 12(9):486–488, 1991.

[7] J.C. Bean. Silicon-based semiconductor heterostructures: column IV bandgap engineering. *Proc. IEEE*, 80(4):571–587, 1992.

[8] C.D. Parikh, F.A. Lindholm. Space-charge region recombination in heterojunction bipolar transistors. *IEEE Trans. Electron Devices*, 39(10):2197–2205, 1992.

[9] D.L. Pulfrey, S. Searles. Electron quasi-fermi level splitting at the base–emitter junction of AlGaAs/GaAs HBT's. *IEEE Trans. Electron Devices*, 40(6):1183–1185, 1993.

[10] H. Fukano, H. Nakajima, T. Ishibashi, Y. Takanashi, M. Fujimoto. Effect of hot-electron injection on high-frequency characteristics of abrupt $\mathrm{In}_{0.52}(\mathrm{Ga}_{1-x}\mathrm{Al}_x)_{0.48}\mathrm{As}$/InGaAs HBTs. *IEEE Trans. Electron Devices*, 39(3):500–506, 1992.

[11] J. Laskar, R.N. Nottenburg, J.A. Baquedano, A.F.J. Levi, J. Kolodzey. Forward transit delay in $\mathrm{In}_{0.53}\mathrm{Ga}_{0.47}\mathrm{As}$ heterojunction bipolar transistors with nonequilibrium electron transport. *IEEE Trans. Electron Devices*, 40(11):1942–1949, 1993.

[12] M.S. Lundstrom. Boundary conditions for pn heterojunctions. *Solid-State Electron.*, 27(5):491–496, 1984.

[13] A.A. Grinberg, S. Luryi. On the thermionic-diffusion theory of minority transport in heterojunction bipolar transistors. *IEEE Trans. Electron Devices*, 40(5):859–866, 1993.

[14] A.A. Grinberg, M.S. Shur, R.J. Fischer, H. Morkoc. An investigation of the effect of graded layers and tunneling on the performance of AlGaAs/GaAs heterojunction bipolar transistors. *IEEE Trans. Electron Devices*, 31(12):1758 1765, 1984.

[15] S. Searles, D.L. Pulfrey, T.C. Kleckner. Analytical expressions for the tunnel current at abrupt semiconductor-semiconductor heterojunctions. *IEEE Trans. Electron Devices*, 44(11):1851–1856, 1997.

[16] K.-M. Chang, J.-Y. Tsai, C.-Y. Chang. New physical formulation of the thermionic emission current at the heterojunction interface. *IEEE Electron Device Lett.*, 14(7):338–341, 1993.

[17] S.J. Fonash. Band structure and photocurrent collection in crystalline and polycrystalline pn heterojunction solar cells. *Solid-State Electron.*, 22:907–910, 1979.

[18] S.J. Fonash. General formulation of the current-voltage characteristics of a pn heterojunction solar cell. *J. Appl. Phys.*, 51(4):2115–2118, 1980.

[19] A.A. Grinberg. Thermionic emission in heterosystems with different effective electronic masses. *Phys. Rev. B*, 33(10):7256–7258, 1986.

[20] C.D. Parikh, F.A. Lindholm. A new charge-control model for single and double-heterojunction bipolar transistors. *IEEE Trans. Electron Devices*, 39(6):1303–1311, 1992.

[21] H. Kroemer. Heterostructure bipolar transistors: what should we build? *J. Vac. Sci. Technol.*, B1:126–130, 1983.

[22] S.C.M. Ho, D.L. Pulfrey. The effect of base grading on the gain and high-frequency performance of AlGaAs/GaAs heterojunction bipolar transistors. *IEEE Trans. Electron Devices*, 36(10):2173–2182, 1989.

[23] M.E. Hafizi, C.R. Crowell, M.E. Grupen. The DC characteristics of GaAs/AlGaAs heterojunction bipolar transistors with application to device modeling. *IEEE Trans. Electron Devices*, 37(10):2121 2129, 1990.

[24] A.F.J. Levi, B. Jalali, R.N. Nottenburg, A.Y. Cho. Vertical scaling in hterojunction bipolar transistors with nonequilibrium base transport. *Appl. Phys. Lett.*, 60(4):460–462, 1992.

[25] B. Jalali, S.J. Pearton (eds.). *InP HBTs: Growth, Processing and Applications.* Artech House, Boston, 1995.

[26] T. Ishibashi. Nonequilibrium electron transport in HBTs. *IEEE Trans. Electron Devices*, 48(11):2595–2605, 2001.

[27] S.S. Iyer, G.L. Paton, J.M.C. Stork, B.S. Meyerson, D.L. Harame. Heterojunction bipolar transistors using Si–Ge alloys. *IEEE Trans. Electron Devices*, 36(10):2043–2064, 1989.

[28] R. People. Physics and applications of Ge_xSi_{1-x}/Si strained-layer heterostructures. *IEEE J. Quantum Electron.*, 22(9):1696–1710, 1986.

[29] S.S. Iyer, G.L. Patton, S.L. Delage, S. Tiwari, J.M.C. Stork. Silicon-germanium base heterojunction bipolar transistors by molecular beam epitaxy. *IEDM Tech. Dig.*, 1987:874–876, 1987.

[30] T. Tatsumi, H. Hirayama, N. Aizaki. Si/$Ge_{0.3}Si_{0.7}$/Si heterojunction bipolar transistor made with Si molecular beam epitaxy. *Appl. Phys. Lett.*, 52:895–897, 1988.

[31] B.S. Meyerson. UHV/CVD growth of Si and Si:Ge alloys: chemistry, physics and device applications. *Proc. IEEE*, 80(10):1592–1608, 1992.

[32] O. Madelung (Ed.). *Semiconductors – Basic Data.* Springer, Berlin, 2nd edition, 1996.

[33] R. People. Indirect bandgap of coherently strained Ge_xSi_{1-x} bulk alloys on ⟨001⟩ silicon substrates. *Phys. Rev. B*, 32(2):1405–1408, 1985.

[34] S.C. Jain, J. Poortmans, S.S. Iyer, J.J. Loferski, J. Nijs, R. Mertens, R. van Overstraeten. Electrical and optical bandgaps of Ge_xSi_{1-x} strained layers. *IEEE Trans. Electron Devices*, 40(12):2338 2343, 1993.

[35] C.H. Gan, J.A. del Alamo, B.R. Bennett, B.S. Meyerson, E.F. Crabbe, C.G. Sodini, L.R. Reif. $Si_{1-x}Ge_x/Si$ valence band discontinuity measurements using a semiconductor-insulator-semiconductor (SIS) heterostructure. *IEEE Trans. Electron Devices*, 41(12):2430–2439, 1994.

[36] S.-I. Takagi, J.L. Hoyt, K. Rim, J.J. Welser, J.F. Gibbons. Evaluation of the valence band discontinuity of $Si/Si_{1-x}Ge_x/Si$ heterostructures by application of admittance spectroscopy to MOS capacitors. *IEEE Trans. Electron Devices*, 45(2):494–501, 1998.

[37] T. Manku, A. Nathan. Electron drift mobility for devices based on unstrained and coherently strained $Si_{1-x}Ge_x$ grown on (001) silicon substrate. *IEEE Trans. Electron Devices*, 39(9):2082–2089, 1992.

[38] H.J. Osten, D. Knoll, B. Heinemann, H. Rücker, B. Tillack. Carbon doped SiGe heterojunction bipolar transistors for high frequency applications. *Proc. IEEE BCTM*, pp. 109 116, 1999.

[39] H.J. Osten, D. Knoll, B. Heinemann, P. Schley. Increasing process margin in SiGe heterojunction bipolar technology by adding carbon. *IEEE Trans. Electron Devices*, 46:1910–1912, 1999.

[40] T.F. Meister, H. Schäfer, M. Franosch, W. Molzer, K. Aufinger, U. Scheler, C. Walz, M. Stolz, S. Boguth, J. Böck. SiGe base bipolar technology with 74 GHz $f_{\max}$ and 11 ps gate delay. *IEDM Tech. Dig.*, pp. 739–742, 1995.

[41] J.D. Cressler. SiGe HBT technology: A new contender for Si-based RF and microwave circuit applications. *IEEE Trans. on Microwave Theory and Techniques*, 46(5):572–589, 1998.

[42] J.D. Cressler, J.H. Comfort, E.F. Crabbe, G.L. Patton, J.M.C. Stork, J. Y.-C. Sun, B.S. Meyerson. On the profile designa and optimization of epitaxial Si- and SiGe-base bipolar technology for 77 K applications: part i: transistor dc design considerations. *IEEE Trans. Electron Devices*, 40(3):525–541, 1993.

[43] J.D. Cresler, E.F. Crabbe, J.H. Comfort, J.M.C. Stork, J. Y.-C. Sun. On the profile designa and optimization of epitaxial Si- and SiGe-base bipolar technology for 77 K applications: part ii: circuit performance issues. *IEEE Trans. Electron Devices*, 40(3):525–541, 1993.

[44] S.C. Jain, T.J. Gosling, D.H.J. Totterdell, J. Pootmans, R.P. Mertens, R. van Overstraeten. The combined effects of strain and heavy doping on the indirect band gap of Si and Ge_xSi_{1-x} alloys. *Solid-State Electron.*, 34(5):445–451, 1991.

[45] J.M.C. Stork, G.L. Patton, E.F. Crabbe, D.L. Harame, B.S. Meyerson, S.S. Iyer, E. Ganin. Design issues for SiGe heterojunction bipolar transistors. *IEDM Tech. Dig.*, pp. 57–64, 1989.

[46] G. Niu, J.D. Cressler, A.J. Joseph. Quantifying neutral base recombination and the effects of collector-base junction traps in UHV/CVD SiGe HBT's. *IEEE Trans. Electron Devices*, 45(12):2499–2504, 1998.

[47] C.M. Krowne, K. Ikossi-Anastasiou, E. Kouginos. Early voltage in heterojunction bipolar transistors: quantum tunneling and base recombination effects. *Solid-State Electron.*, 39:1979–1991, 1995.

[48] S.L. Salmon, J.D. Cresler, R.C. Jaeger, D.L. Harame. The influence of Ge grading on the bias and temperature characteristics of SiGe HBTs for precision analog circuits. *IEEE Trans. Electron Devices*, 47(2):292–298, 2000.

[49] N. Rinaldi. Analytical relations for the base transit time and collector transit time in bjts and hbts. *Solid-State Electron.*, 41(8):1153–1158, 1997.

[50] J. Song, J.S. Yuan. Optimum Ge profile for base transit time minimization of SiGe HBT. *Solid-State Electron.*, 41(12):1957 1959, 1997.

[51] R.J.E. Hueting, J.W. Slotboom, A. Pruijmboom, W.B. de Boer, C.E. Timmering, N.E.B. Cowern. On the optimization of SiGe-base bipolar transistors. *IEEE Trans. Electron Devices*, 43(9):1518–1524, 1996.

[52] A.J. Joseph, J.D. Cresler, D.M. Richey, G. Niu. Optimization of SiGe HBTs for operation at high current densities. *IEEE Trans. Electron Devices*, 46(7):1347–1354, 1999.

[53] Q. Liang, J.D. Cressler, G. Niu, R.M. Malladi, K. Newton, D.L. Harame. A physics-based high-injection transit-time model applied to barrier effects in SiGe HBTs. *IEEE Trans. Electron Devices*, 49(10):1807–1813, 2002.

[54] S. Wilms, H.-M. Rein. Analytical high-current model for the transit time of SiGe HBTs. *Proc. IEEE BCTM*, pp. 199–202, 1998.

[55] M. Friedrich, H.-M. Rein. Analytical current–voltage relations for compact SiGe HBT models - part i: the 'idealized' HBT. *IEEE Trans. Electron Devices*, 46(7):1384–1393, 1999.

[56] M. Friedrich, H.-M. Rein. Analytical current–voltage relations for compact SiGe HBT models - part ii: applicationto practical HBT's and parameter extraction. *IEEE Trans. Electron Devices*, 46(7):1394–1401, 1999.

[57] J.C.J. Paasschens, W.J. Kloosterman, R.J. Havens. Modelling two SiGe HBT specific features for circuit simulation. *Proc. IEEE BCTM*, pp. 38–41, 2001.

[58] P.M. Asbeck, F. M.-C. Chang, K.-C. Wang, G.J. Sullivan, D.T. Cheung. GaAs-based heterojunction bipolar transistors for very high performance electronic circuits. *Proc. IEEE*, 81(12):1709–1726, 1993.

[59] C.T.M. Chang, H.-T. Yuan. GaAs HBTs for high-speed digital integrated circuit applications. *Proc. IEEE*, 81(12):1727–1743, 1993.

[60] B.K. Oyama, B.P. Wong. GaAs HBTs for analog circuits. *Proc. IEEE*, 81(12):1744–1761, 1993.

[61] B. Bayraktaroglu. GaAs HBTs for microwave integrated circuits. *Proc. IEEE*, 81(12):1762–1785, 1993.

[62] P. Asbeck. III-V HBTs for microwave applications: technology status and modeling challenges. *Proc. IEEE BCTM*, pp. 52–57, 2000.

[63] G.-B. Gao, H. Morkoc. Material-based comparison for power heterojunction bipolar transistors. *IEEE Trans. Electron Devices*, 38(11):2410–2416, 1991.

[64] C. Monier, F. Ren, J. Han, P.-C. Chang, R.J. Shul, K.-P. Lee, A. Zhang, A.G. Baca, S. Pearton. Simulation of npn and pnp AlGaN/GaN heterojunction bipolar transistor performances: limiting factors and optimum design. *IEEE Trans. Electron Devices*, 48(3):427 432, 2001.

[65] B.S. Shelton, D.J.H. Lambert, J.J. Huang, M.M. Wong, U. Chowdhury, T.G. Zhu, H.K. Kwon, Z. Liliental-Weber, M Benarama, M. Feng, R.D. Dupuis. Selective area growth and characterization of AlGaN/GaN heterojunction bipolar transistors by metallorganic chemical vapor deposition. *IEEE Trans. Electron Devices*, 48(3):490–494, 2001.

[66] L.S. McCarthy, I.P. Smorchkova, H. xing, P. Kozodoy, P. Fini, J. Limb, D.L. Pulfrey, J.S. Speck, M.J.W. Rodwell, S.P. DenBaars, U.K. Mishra. GaN HBT: Toward an RF device. *IEEE Trans. Electron Devices*, 48(3):543–551, 2001.

[67] D.L. Pulfrey, S. Fathpour. Performance predictions for npn $Al_xGa_{1-x}N$/GaN HBTs. *IEEE Trans. Electron Devices*, 48(3):597–602, 2001.

[68] G. Raghavan, M. Sokolich, W. Stanchina. Indium phosphide ICs unleash the high-frequency spectrum. *IEEE Spectrum*, 37(10):47–52, 2000.

[69] H. Shigematsu, T. Iwai, Y. Matsumiya, H. Ohnisihi, O. Ueda, T. Fuji. Ultrahigh f_T and $f_{\max}$ new self-alignment InP/InGaAs HBTs with a highly Be-doped base layer grown by ALE/MOCVD. *IEEE Electron. Device Lett.*, 16(2):55–57, 1995.

[70] T. Low, T. Shirley, C. Hutchinson, G. Essilfie, W. Whiteley, B. Yeats, D. D'Avanzo. InGaP HBT technology for RF and microwave instrumentation. *Solid-State Electron*, 43:1437–1444, 1999.

[71] G.-B. Gao, H. Morkoc, M.-C.F. Chang. Heterojunction bipolar transistor design for power applications. *IEEE Trans. Electron Devices*, 39(9):1987–1997, 1992.

[72] G.-B. Gao, M. Selim Ünlü, H. Morkoc, D.L. Blackburn. Emitter ballasting resistor design for, and current handling capability of AlGaAs/GaAs power heterojunction bipolar transistors. *IEEE Trans. Electron Devices*, 38(2):185–196, 1991.

[73] R.P. Arnold, D.S. Zoroglu. A quantitative study of emitter ballasting. *IEEE Trans. Electron Devices*, 21(7):385–391, 1974.

[74] P.L. Hower, P.K. Govil. Comparison of one- and two-dimensional models of transistor thermal instability. *IEEE Trans. Electron Devices*, 21(10):617–623, 1974.

[75] W. Liu, S. Nelson, D.G. Hill, A. Khatibzadeh. Current gain collapse in microwave multifinger heterojunction bipolar transistors operated at very high power densities. *IEEE Trans. Electron Devices*, 40(11):1917–1927, 1993.

[76] W. Liu, A. Khatibzadeh. The collapse of current gain in multi-finger heterojunction bipolar transistors: its substrate temperature dependence, instability criteria, and modeling. *IEEE Trans. Electron Devices*, 41(10):1698–1707, 1994.

[77] W. Liu. The interdependence between the collapse phenomenon and the avalanche breakdown in AlGaAs/GaAs power heterojunction bipolar transistors. *IEEE Trans. Electron Devices*, 42(4):591–597, 1995.

5 Noise Modeling

The quantization of electric charge, in conjunction with fluctuations of the density and velocity of charge carriers, is responsible for the noise in electronic devices. Generally we may distinguish between shot noise, due to fluctuations of the current associated with directed carrier motion, for example, caused by an applied electric field, and thermal noise (also termed diffusion noise or velocity fluctuation noise), due to the thermal motion of the carriers.

In semiconductors generation–recombination noise, which is due to fluctuations of the carrier density caused by interband transitions,[1] may be important in addition. In order to take account of these effects within the framework of transport theory, the continuity and current equations have to be extended by stochastic terms, which represent local noise sources. The calculation of the noise properties of a device then requires one to relate these local, microscopically produced fluctuations in the interior of the device to the (measurable) fluctuations of the terminal currents and voltages.

5.1 Noise in Semiconductors

This section gives an introduction to thermal noise, shot noise and generation–recombination noise, and also examines $1/f$ noise and extra noise due to hot-carrier effects and carrier multiplication.

5.1.1 Shot Noise and Thermal Noise

Consider the situation depicted in Fig. 5.1, where a current flows between two electrodes E1 and E2 held at constant potential. An electron between electrodes E1 and E2 carries a current $i_k(t) = -ev_{x,k}(t)/L$, where $v_{x,k}(t)$ denotes the x component of its velocity at time t, and L denotes the distance between the electrodes. Adding the current carried by all of the electrons involved, we obtain the total current at time t as

$$i(t) = -\frac{e}{L}\sum_{k=1}^{N(t)} v_{x,k}(t) , \tag{5.1}$$

where $N(t)$ is the number of electrons between the electrodes at time t. If we denote the drift velocity by $\langle v_x \rangle = v_{\rm n}$, the ensemble average[2] of the current is therefore

[1] Or transitions between band states and localized trapping centers.

[2] See Appendix E for a brief summary of the statistical foundations of noise theory.

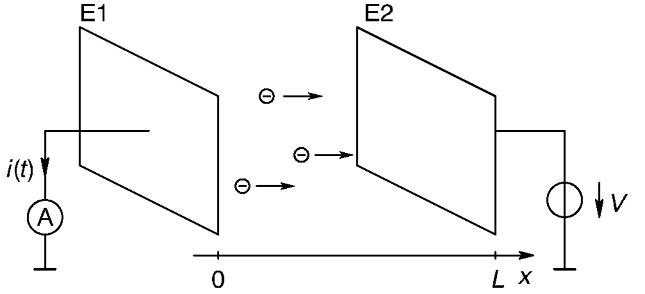


Fig. 5.1. Electrons carrying a current between electrodes E1 and E2

$$I = -\frac{e}{L}\langle N\rangle\, v_{\mathrm{n}} , \tag{5.2}$$

if no correlation exists between the fluctuations of the electron number and the velocity fluctuations. Fluctuations of the particle current may be caused by either fluctuations of the number $N(t)$ of electrons between the electrodes or fluctuations of the x component of the velocity of the carriers. If $\Delta N(t)$ denotes the deviation of the electron number $N(t)$ from its mean value $\langle N\rangle$, and $\Delta v_{x,k}(t)$ the deviation of the x component of the velocity of electron k at time t from the mean value $\langle v_x\rangle = v_{\mathrm{n}}$, the current fluctuation is

$$\Delta i(t) \approx -\frac{e}{L}v_{\mathrm{n}}\Delta N(t) - \frac{e}{L}\sum_{k=1}^{\langle N\rangle}\Delta v_{x,k}(t) . \tag{5.3}$$

Since no correlation exists between the velocity fluctuations $\Delta v_{x,k}(t)$ and $\Delta v_{x,l}(t)$ if $k \neq l$, the variance of the current is

$$\langle \Delta I^2\rangle = \frac{e^2}{L^2}\left(v_{\mathrm{n}}^2\langle\Delta N^2\rangle + \langle N\rangle\langle\Delta v_x^2\rangle\right) . \tag{5.4}$$

The two terms on the right-hand side of this equation are of different nature. While the first term, referred to in the following as the shot noise, is related to the directed motion of electrons, the second term, referred to in the following as the thermal noise, is associated with the random thermal motion of the electrons. Shot noise is therefore associated with an average current flow and is a typical nonequilibrium phenomenon, while thermal noise will always be observed, even in thermal equilibrium. Since the average carrier velocity v_{n} used to describe the directed motion of carriers does not fluctuate, shot noise will be observed only if the number of carriers fluctuates. Depending on whether the term determined by the velocity fluctuations $\langle N\rangle\langle\Delta v_x^2\rangle$ or the term determined by the fluctuations of the particle number $v_{\mathrm{n}}^2\langle\Delta N^2\rangle$ dominates, the device will show predominantly thermal noise or predominantly shot noise, as illustrated by the following examples.

Shot Noise of a Vacuum Diode

Figure 5.2 shows the current voltage characteristics of a vacuum diode for different values of cathode temperature T_C. In the saturation region, the device acts as a current source that delivers a current limited by the rate at which electrons are emitted from the cathode, since virtually all emitted electrons are collected by the anode. The number $N(t)$ of electrons in the region

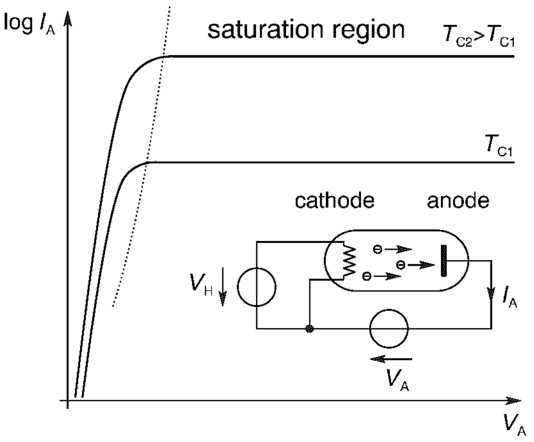


Fig. 5.2. Current–voltage characteristics of a vacuum diode for different values of cathode temperature T_C

between the electrodes is controlled by a stochastic emission process. Owing to the exponential decrease of the electron distribution function in the cathode metal at large energy values, the kinetic energy of the electrons at the cathode surface is small: most of the electrons emitted have kinetic energies well below $3k_BT_C$. Since operation in the saturation regime requires several volts to be applied to the anode, the energy acquired by the electrons in the electric field is much larger than their initial (random) kinetic energy. For a vacuum diode in the saturation region we may therefore assume $\langle \Delta v_x^2 \rangle \ll v_n^2$. In this limit, the autocorrelation function of the noise current $R_i(\tau)$ is essentially determined by the autocorrelation function $R_{\Delta N}(\tau)$ of the number of particles in the drift volume:

$$R_i(\tau) \approx \frac{e^2}{L^2} v_n^2 R_{\Delta N}(\tau) . \tag{5.5}$$

Since the presence of one electron in the drift volume is independent of the presence of other electrons there, the autocorrelation function of the particle number is $\langle N \rangle$ times the autocorrelation function of the function $f_1(t)$ that describes the presence of one electron in the drift region between the two electrodes, defined by

$$f_1(t) = \begin{cases} 1 & \text{if the electron is inside the drift volume} \\ 0 & \text{if not} \end{cases} .$$

If it takes a flight time t_f for an electron to cross the drift region, $f_1(t)$ is of rectangular form, as shown in Fig. 5.3a; this results in an autocorrelation function $R_1(\tau)$ of triangular shape, as shown in Fig. 5.3b.

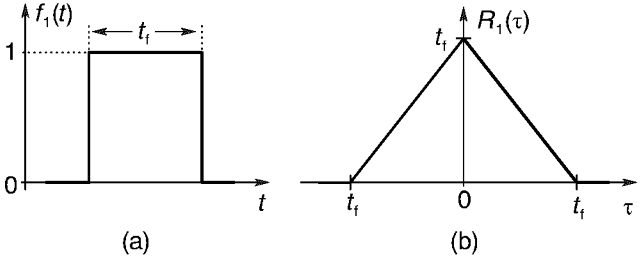


Fig. 5.3. (**a**) Function $f_1(t)$ describing the presence of an electron in the drift volume, and (**b**) the corresponding autocorrelation function

The spectral density $S_1(\omega)$ of $R_1(\tau)$ can be determined with the help of the Wiener–Khintchin relation

$$S_1(\omega) = 4\int_0^{t_f} \frac{t_f - \tau}{t_f} \cos(\omega\tau)\,\mathrm{d}\tau = 4t_f \frac{1-\cos(\omega t_f)}{(\omega t_f)^2} . \tag{5.6}$$

For frequencies that obey $\omega t_f \ll 1$, it is possible to approximate $S_1(\omega)$ by its low-frequency limit $S_1(\omega \to 0) = 2t_f$ to obtain

$$S_{\Delta N}(\omega) = \langle N \rangle S_1(\omega) \approx 2\langle N \rangle t_f . \tag{5.7}$$

The spectral density of the noise current then yields, if we write $t_f = L/v_n$,

$$S_i(\omega) = \frac{e^2}{L^2} v_n^2 S_{\Delta N}(\omega) = 2e\frac{e}{L}\langle N \rangle v_n \frac{t_f v_n}{L} = 2eI_A , \tag{5.8}$$

which is Schottky's classical result [1] for shot noise.[3] The same result can be obtained by application of Carson's theorem, as shown in Appendix E[4].

[3]This result is only correct if the carrier flow is unidirectional. For small values of the anode voltage V_A, not all electrons emitted by the cathode are transported to the anode – some are captured again by the cathode, resulting in a noise current with a spectral density exceeding $2eI_A$.

[4]Since the electrons are accelerated in the electric field between the cathode and anode, the pulse shape of the individual current pulses is triangular, rather than rectangular as has been assumed here for simplicity. Application of Carson's theorem to triangular current pulses yields

$$S_i = 2eI_A \frac{4(\omega t_f)^2 + 8\,[\,1-\cos(\omega t_f) - (\omega t_f)\sin(\omega t_f)\,]}{(\omega t_f)^4} ,$$

which is identical to the result obtained for rectangular pulses in the limit $\omega t_f \ll 1$.

Thermal Noise of an Ohmic Resistor

In a metal resistor,[5] the electron number shows negligibly small fluctuations since deviations from charge neutrality settle down within the dielectric relaxation time τ_ϵ. In this case the number of electrons between the electrodes is not controlled by an emission process, and insufficient or superfluous electrons are easily exchanged with the contact electrodes, resulting in negligibly small fluctuations of the electron number. Another difference from the vac-

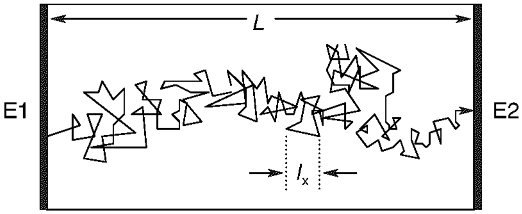


Fig. 5.4. Path of an electron (schematic representation) between electrodes E1 and E2 in a resistive material

uum diode results from the scattering of electrons: in a solid, the mean free path is much less than the distance between the electrodes, and each electron is involved in many scattering events on its way from E1 to E2 (Fig. 5.4). The average drift velocity is therefore much smaller than the random thermal velocity, i.e. $\langle \Delta v_x^2 \rangle \langle N \rangle \gg v_n^2 \langle \Delta N^2 \rangle$. In this limit the autocorrelation function of the noise current is predominantly determined by the autocorrelation function $R_{\Delta v}(\tau)$ of the x component of the velocity of a carrier:

$$R_i(\tau) \approx \frac{e^2}{L^2} \langle N \rangle R_{\Delta v}(\tau) . \tag{5.9}$$

Owing to the randomizing effect of the scattering process, the autocorrelation function $R_{\Delta v}(\tau)$ only differs significantly from zero within a time interval of the order of the collision time τ_c. For frequencies which obey $\omega\tau_c \ll 1$, we obtain, with the help of the Wiener–Khintchin relation, the relation between the diffusion coefficient and the velocity autocorrelation function derived in Appendix E, and the Einstein relation $D_n = \mu_n V_T$,

$$S_{\Delta v}(\omega) \approx S_{\Delta v}(\omega = 0) = 4 \int_0^\infty R_{\Delta v}(\tau) \, d\tau = 4D_n = 4\mu_n V_T . \tag{5.10}$$

If we write $\langle N \rangle = nAL$, the spectral density of the noise current reads

[5]In a resistor composed of a semiconductor material, current may be carried by electrons and holes. Fluctuations of the electron and hole densities due to generation and recombination processes do not result in a change of charge density and therefore do not attract electrons and holes across the contacts: therefore the noise observed in thermal equilibrium will not increase because of generation or recombination processes. The situation is different if the resistor carries a current. In this case a resistor composed of a semiconductor material shows an additional generation recombination noise due to fluctuations of the carrier densities, which result in fluctuations of the specific resistivity (see Sect. 5.1.2).

$$S_i(\omega) = \frac{e^2}{L^2}\langle N\rangle S_{\Delta v}(\omega) = 4eV_{\mathrm{T}}\frac{e\mu_{\mathrm{n}}nA}{L} = 4k_{\mathrm{B}}TG\,, \tag{5.11}$$

which is the Nyquist theorem for the thermal noise of a resistor. This theorem relates the spectral density S_i of the noise current of a resistor to its conductance G, i.e. to the quantity that describes the response of the current to a change of applied voltage such that $\Delta I = G\Delta V$. That the thermal noise of an ohmic resistor is related to its resistance can be understood if one considers the fact that both thermal noise and ohmic resistance have the same physical origin, that is, the random scattering of mobile carriers by impurities and lattice vibrations. The Nyquist theorem for the thermal noise of a resistor is a special case of the fluctuation–dissipation theorem [2], which relates the fluctuations of a linear system from its equilibrium state to the response of the system to an external perturbation.

The thermal noise of a resistor as described by the Nyquist theorem is observed in thermal equilibrium, i.e. without a bias applied to the resistor terminals, and does not vary significantly in biased resistors – at least as long as the field in the resistor volume is small enough to avoid significant carrier heating. The shot noise contribution associated with the directed motion of carriers is negligibly small in the resistor, since the electron number in the drift volume may be assumed to be constant. This is the result of huge currents crossing the contact electrodes in both directions, which almost cancel except for a small difference that is observed as the current flowing in the external circuit.

Noise of Hot Electrons. Consider the noise of an n-type semiconductor sample to which a voltage V is applied. In a low electric field, the average kinetic energy of an electron in the conduction band will be close to $3k_{\mathrm{B}}T_{\mathrm{L}}/2$, where T_{L} is the temperature of the crystal lattice, and the noise will be related to the impedance by the Nyquist theorem, i.e. noise measurements do not provide information that is not otherwise available from mobility data. However, if the electric field increases, the mean kinetic energy $\langle W\rangle$ of the electrons may considerably exceed $3k_{\mathrm{B}}T_{\mathrm{L}}/2$; the corresponding electron temperature[6] $T_{\mathrm{n}} = 2\langle W\rangle/3k_{\mathrm{B}} \gg T_{\mathrm{L}}$ will then exceed the lattice temperature T_{L}, resulting in increased noise.[7] A measurement of the noise produced by a biased semiconductor sample in the nonohmic regime therefore provides information about the mean energy of the electrons [4, 5].

[6] The energy relaxation time τ_{wn} is considerably larger than the momentum relaxation time τ_{vn}, i.e. an electron generally loses only a small fraction of its kinetic energy in a collision process, although that process almost completely randomizes its direction of flight. This is a prerequisite for the "thermalization" of the electron distribution and the introduction of an electron temperature.

[7] See e.g. [3] for a survey of fluctuations and noise of hot carriers in semiconductors.

5.1.2 Generation–Recombination Noise

Generation and recombination processes change the number of electrons and holes in a stochastic manner and contribute to device noise.

Fluctuations of Particle Density

As explained in Appendix E, generation and recombination processes cause fluctuations of the number N of electrons or holes in a semiconductor volume with a variance

$$\langle \Delta N^2 \rangle \;=\; \frac{R+G}{2} V\tau \,, \tag{5.12}$$

and a relaxation time τ that can be expressed in terms of the derivatives of the generation and recombination rates with respect to the average electron density $n = \langle N \rangle / V$. In the special case of a p-type semiconductor under low-level-injection conditions,[8]

$$G \;=\; \frac{N_{\mathrm{T}} n_{\mathrm{ie}}^2}{(n+n_1)\tau_{\mathrm{p0}} + (p+p_1)\tau_{\mathrm{n0}}} + C_{\mathrm{p}} p n_{\mathrm{ie}}^2 \;\approx\; \frac{n_{\mathrm{p0}}}{\tau_{\mathrm{n}}} \,,$$

$$R \;=\; \frac{N_{\mathrm{T}} p n}{(n+n_1)\tau_{\mathrm{p0}} + (p+p_1)\tau_{\mathrm{n0}}} + C_{\mathrm{p}} p^2 n \;\approx\; \frac{n}{\tau_{\mathrm{n}}}$$

and

$$\tau \;=\; \left(\frac{\mathrm{d}R}{\mathrm{d}n} - \frac{\mathrm{d}G}{\mathrm{d}n} \right)^{-1} \;\approx\; \tau_{\mathrm{n}} \,.$$

As fluctuations Δn of the electron density from the stationary value n relax with the time constant τ, the autocorrelation function of the electron density obeys [6]

$$\langle \Delta n(0) \Delta n(t) \rangle \;=\; \langle \Delta n^2 \rangle \, \mathrm{e}^{-t/\tau} \,,$$

where $\langle \Delta n^2 \rangle = \langle \Delta N^2 \rangle / V^2$. Application of the Wiener–Khintchin relation now yields the following for the spectral density of the particle density fluctuations:

$$S_{\Delta n}(\omega) \;=\; 4 \int_0^{\infty} \langle \Delta n(0) \Delta n(t) \rangle \cos(\omega t) \, \mathrm{d}t \;=\; \frac{4 \langle \Delta n^2 \rangle \tau}{1+\omega^2 \tau^2} \,, \tag{5.13}$$

or, with the help of (5.12),

$$S_{\Delta n}(\omega) \;=\; \frac{2(G+R)\tau^2}{1+\omega^2\tau^2} \frac{1}{V} \,. \tag{5.14}$$

[8] The minority-carrier lifetime τ_{n} takes account of both SRH and Auger processes (see Sect. 2.6):

$$\frac{1}{\tau_{\mathrm{n}}} \;=\; C_{\mathrm{p}} p^2 + \frac{N_{\mathrm{T}}}{\tau_{\mathrm{n0}}} \,.$$

A resistor composed of a semiconducting material with average resistance $R \sim 1/n$ will show resistance fluctuations with a spectral density $S_R(\omega) = R^2 S_{\Delta n}(\omega)/n^2$ if these are solely determined by fluctuations of the carrier density. If a constant voltage V is applied to the resistor, the resistance fluctuations imply a noise current with a spectral density

$$S_i(\omega) = \frac{I^2}{R^2} S_R(\omega) = I^2 \frac{S_{\Delta n}(\omega)}{n^2} ,$$

which varies in proportion to the square of the average current I carried by the resistor owing to the external voltage V; this noise is not observed in thermal equilibrium.

Noise Due to Carrier Multiplication Effects

Carrier multiplication due to impact ionization is a random process and therefore increases device noise. Investigations of this effect were originally motivated by the need to characterize the current amplification process in avalanche photodiodes, which determines the detectivity of such sensors. In modern high-frequency bipolar transistors with their small values of open-base breakdown voltage, carrier multiplication in the bc diode may be non-negligible and may affect the nonlinearity and noise of the device. From the theory of the avalanche photodiode, it is known that the noise of a reverse-biased pn junction, into which a primary current I_T is injected can be described by a noise current source with spectral density [7, 8]

$$S_i = 2em_\mathrm{n}^2 \phi(m_\mathrm{n}) I_\mathrm{T} , \tag{5.15}$$

where $m_\mathrm{n} = \langle m_\mathrm{n} \rangle$ denotes the average multiplication factor and $\phi(m_\mathrm{n}) = \langle m_\mathrm{n}^2 \rangle / \langle m_\mathrm{n} \rangle^2$ represents the increase of noise due to the stochastic carrier multiplication process. For pure electron injection, the value of $\phi(m_\mathrm{n})$ is given by [8]

$$\phi(m_\mathrm{n}) = 2 - \frac{1}{m_\mathrm{n}} + k \left(m_\mathrm{n} - 2 + \frac{1}{m_\mathrm{n}} \right) , \tag{5.16}$$

where $k = \alpha_\mathrm{p}/\alpha_\mathrm{n}$ denotes the ratio of the hole and electron ionization coefficients, which was assumed to be constant. Since $k \ll 1$ for silicon, we may make the approximation

$$\phi(m_\mathrm{n}) \approx 2 - 1/m_\mathrm{n} \tag{5.17}$$

in the weak avalanche regime.[9] Noise due to carrier multiplication in the bc diode of a high-frequency bipolar transistor becomes relevant when the transistor is operated near the open-base breakdown voltage, as is shown by the calculations in Sect. 5.6.3.

[9] See e.g. [9, 10] and references therein for further discussion.

5.1.3 Low-Frequency Noise (1/f Noise)

At low frequencies, $1/f$ noise, with a spectral density proportional to $f^{-\gamma}$, where $\gamma = 1.0 \pm 0.1$, is observed [11] in all electronic devices over a wide frequency range and is usually measured from 1 Hz to 10 kHz. At higher frequencies the $1/f$ noise disappears into (white) thermal noise or shot noise. While $1/f$ noise of electronic devices was first considered to be a surface effect that could be avoided by careful treatment of surfaces and interfaces, an investigation of measured $1/f$-noise data for metal and semiconductor resistors by Hooge [11] has shown that the spectral density of the current or resistance fluctuations follows an empirical relation of the form

$$\frac{S_i(f)}{I^2} = \frac{S_R(f)}{R^2} = \frac{\alpha_{\mathrm{H}}}{Nf}, \tag{5.18}$$

where N denotes the number of carriers in the sample. Equation (5.18) is today commonly referred to as the Hooge equation; the Hooge parameter α_{H} typically has values between 10^{-4} and 2×10^{-3}, dependent on crystal quality [11]. Several physical mechanisms are responsible for the $1/f$ noise in electronic devices.

Mobility fluctuations are one of the physical mechanisms leading to $1/f$ noise in resistors. According to a theory by Handel [12], such $1/f$ noise occurs because of interactions of scattered particles with photons emitted during the scattering process and is unavoidable, just like thermal noise, for example. The theoretical predictions of this model, however, result in Hooge parameters that are substantially below experimental findings; for further discussion of this theory see, for example, [11,13–16]. The mechanism described by Handel therefore determines a lower limit to the $1/f$ noise. According to [11], $1/f$ noise due to mobility fluctuations is associated with a Hooge parameter $\alpha_{\mathrm{H}} \approx 10^{-4}$. The resistance of a sample is determined by the product of the carrier density and the mobility of the carriers, leading to

$$\frac{\Delta R}{R} = -\frac{\Delta\mu}{\mu} - \frac{\Delta n}{n} \quad \text{and} \quad \frac{S_R(f)}{R^2} = \frac{S_{\Delta\mu}(f)}{\mu^2} + \frac{S_{\Delta n}(f)}{n^2},$$

where both terms on the right-hand side of each expression may show $1/f$ noise. If the first term predominates, the noise is called mobility fluctuation $1/f$ noise; if the second term predominates, the noise is called number fluctuation $1/f$ noise [15]. Number fluctuation $1/f$ noise due to particle density variations caused by generation and recombination processes at centers in the semiconductor bulk would, however, require an unrealistic energy spectrum of the trap states, as pointed out in [15]. The $1/f$ noise of resistors is therefore generally attributed to mobility fluctuation $1/f$ noise.

The specific resistivity of polysilicon resistors is determined to a large extent by the potential barriers at the boundaries between the grains of crystalline silicon (see Chap. 7). In order for electrons to be able to travel from one grain to another, they have either to surmount the barrier or to tunnel through it. The height of the barrier is determined by the number of charge carriers trapped at the grain boundary, which is subject to random fluctuations. A theoretical analysis that considers fluctuations of trap occupancy at the grain boundaries similarly to the McWorther model (see Appendix E) and relates these trap occupancy fluctuations to fluctuations of the current was presented in [17]. In experimental investigations presented in [17], the low-frequency noise current of medium and large polysilicon resistors was found to have a spectral density that obeys

$$S_i(f) = \frac{I^2}{WL}\frac{\alpha}{f}.$$

In this expression, W and L denote the layout dimensions of the polysilicon sheet resistor, while α is a process-specific and temperature-dependent constant. In [17] a substantial dependence on the doping concentration and on the type of doping[10] was reported, whereas details of the deposition conditions and the grain size did not have significant effects on the observed low-frequency noise.

Low-frequency $1/f$ noise may also result from fluctuations of the surface recombination velocity at an insulating boundary of a semiconducting region. This situation was investigated in [18], where the surface recombination velocity S_p for minoritiy carriers is considered to be a function of the surface potential ψ_S, which in turn is affected by the density of surface-trapped carriers n_it. These dependences result in a relation between the spectral density of the diffusion current and the spectral density of the number of interface-trapped carriers S_nit as follows:

$$S_i = \left(\frac{\partial I}{\partial S_\mathrm{p}}\right)^2 \left(\frac{\partial S_\mathrm{p}}{\partial \psi_\mathrm{S}}\right)^2 \left(\frac{\partial \psi_\mathrm{S}}{\partial N_\mathrm{IT}}\right)^2 S_\mathrm{nit}\,. \tag{5.19}$$

In [18], S_nit was computed using the McWorther model, which assumes interface traps with a wide range of relaxation time constants. Fluctuations of the effective surface recombination velocity cause an increase of $1/f$ noise in BJTs with a polysilicon emitter contact; this mechanism is an important source of $1/f$ noise in modern high-frequency bipolar transistors, beyond the $1/f$ noise produced by the series resistances (see Sect. 5.6).

Multistep tunneling processes [19] in heavily doped junctions are another mechanism leading to $1/f$ noise. Such currents are increased in the course of device degradation and cause a reduction of forward common-emitter current gain (see Chap. 7).

[10]Phosphorus-doped resistors showed larger $1/f$ noise than boron-doped resistors. This was explained by different barrier properties and the larger hole effective mass, which leads to a smaller influence of traps on the current at the grain boundaries.

5.2 Transport Theory of Noise

The transport equations in the drift–diffusion approximation considered up to now do not consider fluctuations. Fluctuations due to generation and recombination processes (i.e. fluctuations of the number of particles) and due to thermal motion of the carriers (thermal noise or diffusion noise) can be taken into account by adding stochastic noise sources to the continuity and current equations resulting in the Langevin equations of transport theory.[11] The stochastic source terms cause fluctuations that have to obey the following system of partial differential equations:

$$\frac{\partial n}{\partial t} = -(R-G) + \frac{1}{e}\nabla\cdot\boldsymbol{J}_\mathrm{n} + \gamma(\boldsymbol{x},t)\ , \tag{5.20}$$

$$\frac{\partial p}{\partial t} = -(R-G) - \frac{1}{e}\nabla\cdot\boldsymbol{J}_\mathrm{p} + \gamma(\boldsymbol{x},t)\ , \tag{5.21}$$

$$\boldsymbol{J}_\mathrm{n} = e\mu_\mathrm{n} n\boldsymbol{E} + eD_\mathrm{n}\nabla n + e\boldsymbol{\eta}_\mathrm{n}(\boldsymbol{x},t)\ , \tag{5.22}$$

$$\boldsymbol{J}_\mathrm{p} = e\mu_\mathrm{p} p\boldsymbol{E} - eD_\mathrm{p}\nabla p - e\boldsymbol{\eta}_\mathrm{p}(\boldsymbol{x},t)\ . \tag{5.23}$$

The stochastic noise sources $\gamma(\boldsymbol{x},t)$, $\boldsymbol{\eta}_\mathrm{n}(\boldsymbol{x},t)$ and $\boldsymbol{\eta}_\mathrm{p}(\boldsymbol{x},t)$ have a white spectrum[12] [22,23]. The spectral density functions of the generation–recombination noise sources are determined by the local generation and recombination rates, i.e. by the rates at which electrons or holes appear or disappear in the conduction or valence band,[13] respectively

$$S_\gamma(\boldsymbol{x},\boldsymbol{x}',\omega) = 2\,[G(\boldsymbol{x})+R(\boldsymbol{x})]\,\delta(\boldsymbol{x}-\boldsymbol{x}')\ . \tag{5.24}$$

---

[11] Here, the discussion is restricted to extensions of the drift-diffusion approximation. Langevin equations that extend the hydrodynamic model are considered in [20, 21].

[12] Modeling of low-frequency $1/f$ noise (e.g. noise due to mobility fluctuations) requires additional local $1/f$ noise sources, as explained in [14, 15]. In order to allow the analysis of time-dependent SRH processes, the density n_t of filled SRH trapping centers has to be considered in a separate continuity equation [22].

[13] This spectral density function is consistent with (5.14), as can be seen from the small-signal continuity equation for the electron density in a p-type semiconductor in the absence of current flow under low-level-injection conditions,

$$\mathrm{j}\omega\Delta\underline{n} = -\frac{\Delta\underline{n}}{\tau_\mathrm{n}} + \underline{\gamma}\ ,$$

which gives $\Delta\underline{n}(\omega) = \tau_\mathrm{n}\underline{\gamma}/(1+\mathrm{j}\omega\tau_\mathrm{n})$. The spectral density of the fluctuations of the electron density is obtained after a double integration with respect to the semiconductor volume,

$$S_{\Delta n}(\omega) = \iint\langle\Delta\underline{n}(\omega)\Delta\underline{n}^*(\omega)\rangle\,\mathrm{d}^3x\,\mathrm{d}^3x' = \frac{2(G+R)\tau_\mathrm{n}^2}{1+\omega^2\tau_\mathrm{n}^2}\frac{1}{V^2}\iint\delta(\boldsymbol{x}-\boldsymbol{x}')\,\mathrm{d}^3x\,\mathrm{d}^3x'\ ,$$

resulting in (5.14), since

$$\iint\delta(\boldsymbol{x}-\boldsymbol{x}')\,\mathrm{d}^3x\,\mathrm{d}^3x' = V\ .$$

The spectral density functions $S^{kl}_{\eta n}(\boldsymbol{x},\boldsymbol{x}')$ and $S^{kl}_{\eta p}(\boldsymbol{x},\boldsymbol{x}')$ of the electron and hole diffusion noise sources, which describe the correlation of the component k of $\boldsymbol{\eta}_\mathrm{n}(\boldsymbol{x},t)$ or $\boldsymbol{\eta}_\mathrm{p}(\boldsymbol{x},t)$ with the component l of $\boldsymbol{\eta}_\mathrm{n}(\boldsymbol{x},t)$ or $\boldsymbol{\eta}_\mathrm{p}(\boldsymbol{x},t)$, are given by

$$S^{kl}_{\eta n}(\boldsymbol{x},\boldsymbol{x}',\omega) = 4D_\mathrm{n} n(\boldsymbol{x})\delta(\boldsymbol{x}-\boldsymbol{x}')\delta_{kl} \,, \tag{5.25}$$

$$S^{kl}_{\eta p}(\boldsymbol{x},\boldsymbol{x}',\omega) = 4D_\mathrm{p} p(\boldsymbol{x})\delta(\boldsymbol{x}-\boldsymbol{x}')\delta_{kl} \,, \tag{5.26}$$

where $\delta(\boldsymbol{x})$ denotes tha Dirac δ-function and δ_{kl} the Kronecker delta. No correlation is assumed between different components of the diffusion noise sources or between the electron and hole diffusion noise sources.

For the calculation of noise, a small-signal analysis is generally performed. The resulting system of linear stochastic differential equations may then be solved using Green's function methods [24]; a method of this kind has also been employed for the numerical computation of semiconductor device noise [25].

5.2.1 Langevin Approach to the Noise of Ohmic Resistors

An n-type resistor of length L and constant cross section A, in which no generation or recombination takes place, serves here as a simple example to illustrate the Langevin method. If we neglect the hole current and assume one-dimensional conditions, the Langevin transport equations and the Poisson equation yield the following system of small-signal equations:

$$\frac{\partial}{\partial t}\Delta n = \frac{1}{e}\frac{\partial}{\partial x}\Delta J_\mathrm{n} \,, \tag{5.27}$$

$$\Delta J_\mathrm{n} = e\mu_\mathrm{n} E\Delta n + e\mu_\mathrm{n} n\Delta E + eD_\mathrm{n}\frac{\partial}{\partial x}\Delta n + e\eta_\mathrm{n} \,, \tag{5.28}$$

$$\frac{\partial}{\partial x}\Delta E = -\frac{e}{\epsilon}\Delta n \,, \tag{5.29}$$

which may be combined to give the following for the total noise current density:

$$\begin{aligned}\Delta J &= \Delta J_\mathrm{n} + \epsilon\frac{\partial}{\partial t}\Delta E \\ &= e\mu_\mathrm{n} E\Delta n + e\mu_\mathrm{n} n\Delta E + eD_\mathrm{n}\frac{\partial}{\partial x}\Delta n + e\eta_\mathrm{n} + \epsilon\frac{\partial}{\partial t}\Delta E \,.\end{aligned}$$

Since the value of $\Delta i(t) = A\Delta J(t)$ does not depend on the position x, this is equivalent to

$$\Delta i(t) = \frac{A}{L}\int_0^L \Delta J(t)\,\mathrm{d}x \,.$$

If a constant voltage is applied to the resistor, the condition

$$\Delta V(t) \;=\; -\int_0^L \Delta E\,\mathrm{d}x = 0$$

holds, and therefore

$$\Delta i(t) \;=\; \frac{eA\mu_\mathrm{n}E}{L}\int_0^L \Delta n\,\mathrm{d}x + \frac{eA}{L}\int_0^L \eta_\mathrm{n}\,\mathrm{d}x\;, \tag{5.30}$$

if we consider only the drift component of the electron current. Under these conditions, it is possible to combine (5.27), (5.28) and (5.29) to obtain the following partial differential equation for the determination of Δn:

$$\begin{aligned}\frac{\partial}{\partial t}\Delta n \;&=\; \mu_\mathrm{n}E\frac{\partial}{\partial x}\Delta n + \mu_\mathrm{n}n\frac{\partial}{\partial x}\Delta E + \frac{\partial}{\partial x}\eta_\mathrm{n}\\ &=\; \mu_\mathrm{n}E\frac{\partial}{\partial x}\Delta n - \frac{e\mu_\mathrm{n}n}{\epsilon}\Delta n + \frac{\partial}{\partial x}\eta_\mathrm{n}\;.\end{aligned}$$

This corresponds, if we write $E = -|E|$, to

$$\Delta n + \tau_\epsilon\left(\frac{\partial}{\partial t} + \mu_\mathrm{n}|E|\frac{\partial}{\partial x}\right)\Delta n \;=\; \tau_\epsilon\frac{\partial\eta_\mathrm{n}}{\partial x}\;, \tag{5.31}$$

where $\tau_\epsilon = \epsilon/(e\mu_\mathrm{n}n) = \epsilon\rho$ denotes the dielectric relaxation time. If the carrier density is large, small values of τ_ϵ result. Fluctuations of the electron density may then be neglected and the fluctuation of the total noise current is determined solely by the second term on the right-hand side of (5.30), since (5.31) gives $\Delta n \to 0$ in the limit $\tau_\epsilon \to 0$. The autocorrelation function of the current fluctuations then is

$$R_i(\tau) \;=\; \frac{e^2A^2}{L^2}\int_0^L\int_0^L \langle\eta_\mathrm{n}(x,t)\eta_\mathrm{n}(x',t+\tau)\rangle\,\mathrm{d}x'\,\mathrm{d}x\;.$$

Using the result[14]

$$S_\eta(x,x',\omega) \;=\; \frac{4n(x)D_\mathrm{n}(x)}{A}\,\delta(x-x')\;,$$

a transformation of $R_i(\tau)$ to the frequency domain yields

$$S_i(\omega) \;=\; \frac{4e^2AnD_\mathrm{n}}{L^2}\int_0^L\int_0^L \delta(x-x')\,\mathrm{d}x\mathrm{d}x' \;=\; \frac{4e^2AnD_\mathrm{n}}{L}\;.$$

Making use of the Einstein relation $D_\mathrm{n} = V_\mathrm{T}\mu_\mathrm{n}$ and introducing the conductance $G = e\mu_\mathrm{n}nA/L = 1/R$, we therefore obtain

$$S_i(\omega) \;=\; 4k_\mathrm{B}TG \;=\; 4k_\mathrm{B}T/R\;. \tag{5.32}$$

[14] The factor $1/A$ results from the reduction of the three-dimensional δ-function $\delta(\boldsymbol{x}-\boldsymbol{x}')$ to a one-dimensional δ-function $\delta(x-x')$.

As can be shown by analysis of (5.31), this approximation is valid as long as the condition $\mu_{\mathrm{n}}|E|\tau_\epsilon \ll L\,|1+\mathrm{j}\omega\tau_\epsilon|$ is fulfilled. Under this condition, the length L of the noisy conductor is large in comparison with the distance $\mu_{\mathrm{n}}|E|\tau_\epsilon$ traveled by a carrier with a drift velocity of magnitude $\mu_{\mathrm{n}}|E|$ within a dielectric relaxation time.

5.3 Noise of pn Junctions

The noise of a pn junction is frequently explained using the corpuscular approach, which starts from the diode equation

$$I = I_{\mathrm{S}}\left[\exp\left(\frac{V}{V_{\mathrm{T}}}\right) - 1\right] , \tag{5.33}$$

interpreted as the superposition of a forward current $I_{\mathrm{S}}\exp(V/V_{\mathrm{T}})$ and a reverse current $-I_{\mathrm{S}}$. These current components are assumed to fluctuate independently from each other and to show full shot noise, resulting in a noise current with spectral density

$$S_i = 2e(I + 2I_{\mathrm{S}}) . \tag{5.34}$$

Although this yields a result that is considered to be correct for an ideal diode at low frequencies, the partition performed here is rather ambiguous since much larger currents the diffusion current and the drift current cross the depletion layer both in the forward and in the reverse direction, as pointed out in [26, 27]. In [28], a somewhat different interpretation of (5.34) was given. There the currents carried by minority-carrier diffusion in each of the bulk regions are split up into a "forward diffusion current" proportional to the minority-carrier density at the depletion layer edge (i.e. $p_{\mathrm{n0}}\exp(V/V_{\mathrm{T}})$ for the electrons diffusing into the n-type region) and a "reverse diffusion current" proportional to the minority-carrier density deep in the bulk region (i.e. p_{n0} in a long-base diode for the electrons diffusing out of the p-type region). These currents are assumed to show uncorrelated shot noise, resulting in (5.34). If, however, this argument is applied to the hole current injected into a transparent emitter of thickness d with a contact of finite surface recombination velocity S_{p}, the current $I = I_{\mathrm{f}} - I_{\mathrm{r}}$ is obtained as the difference between a forward diffusion current I_{f} and a reverse diffusion current I_{r}, determined by the excess hole densities at the space charge layer boundary ($x = 0$) and the contact ($x = d$):

$$I_{\mathrm{f}} = \frac{eA_{\mathrm{j}}D_{\mathrm{p}}}{d}\Delta p(0) \quad \text{and} \quad I_{\mathrm{r}} = \frac{eA_{\mathrm{j}}D_{\mathrm{p}}}{d}\Delta p(d) .$$

If these currents were assumed to fluctuate independently from each other, the shot noise would be considerably enhanced, in contrast to experimental observations for short-base diodes with a polysilicon contact. Therefore a certain amount of correlation between the forward and reverse currents has to be taken into account in order to obtain a proper description of such short-base diodes. As this correlation is expected to depend on the extension d of the bulk region and on frequency, a systematic procedure, such as the Langevin approach described in the preceding section, is needed for the computation of the noise of pn junctions and its frequency dependence. Before such an investigation is performed in Sect. 5.3.2, the mechanisms leading to noise in biased pn junctions will be discussed in some more detail.

5.3.1 Noise Mechanism of Biased pn Junctions

In thermal equilibrium, the noise produced by a pn junction can be obtained from the Nyquist theorem, which states that in thermal equilibrium (at the bias point $V = 0$) a pn diode shows thermal noise with a noise current determined by its small-signal conductance:

$$S_i = 4k_{\mathrm{B}}Tg_{\mathrm{d}} \,. \tag{5.35}$$

If we consider the current–voltage characteristic (5.33), we obtain the small-signal conductance $g_{\mathrm{d}} = \mathrm{d}I/\mathrm{d}V|_{V=0} = I_{\mathrm{S}}/V_{\mathrm{T}}$, resulting in

$$S_i = 4k_{\mathrm{B}}TI_{\mathrm{S}}/V_{\mathrm{T}} = 4eI_{\mathrm{S}}$$

for the spectral density of the noise current in thermal equilibrium in agreement with the shot noise formula (5.34).

If the pn junction is reverse biased and generation of electron–hole pairs in the space charge layer is neglected, a reverse current $-I_{\mathrm{S}}$ flows because of injection of minority carriers across the space charge layer boundaries, as explained in Sect. 1.4.3. The situation is analogous to the vacuum diode, which has a current flow that is determined solely by the random injection of carriers into the drift region. From this consideration, the spectral density of the noise current density is expected to be

$$S_i = 2eI_{\mathrm{S}} \,, \tag{5.36}$$

in accordance with (5.34). Under reverse-bias conditions, therefore, the diode shows full shot noise.[15]

In a forward-biased pn junction diode, current flows because of minority-carrier injection. If, for example, an electron succeeds in escaping the n bulk region and enters the p bulk region it will be screened by the holes present

[15] In contrast to the vacuum diode, however, the electrons acquire substantial amounts of random thermal energy owing to scattering processes: the electron gas heats up, resulting in an additional thermal-noise term in the spectral density function of the noise current. The shot noise formula, which takes account of the effect of random injection of electrons into the drift region, assumes this noise term to be negligible.

there. All its movements in the p-bulk region will be followed by the screening cloud, which neutralizes the electron and therefore suppresses the effect of its movements on the terminal current. If the electron finally recombines with a hole of its screening cloud, the charge density remains unchanged and therefore no effect on the terminal current results. The elementary process that causes carriers to flow in the device terminals is therefore an electron[16] crossing the space charge layer – on approaching the p bulk region, it is already attracting holes because of electrostatic induction, holes that have to be supplied via the anode contact. Thus the noise-generating process is similar to the situation investigated in Sect. 5.1.1, and a superposition of a shot noise current due to the directed motion of the carriers and a thermal-noise current due to the thermal motion of the carriers is expected to be seen. However, in contrast to the current carried between two metal electrodes, which have well-defined surfaces, the "electrodes" are not clearly defined in the case of a forward-biased pn junction, where the region of large electron density has to be considered as one electrode and the region of large hole density as the other. Furthermore, electrons are accelerated by the electric field to the anode in the example of Sect. 5.1.1, i.e. virtually all electrons injected into the drift region will reach the anode. In a diode, however, only a small fraction of carriers will be able to surmount the potential barrier associated with the decelerating field of the space charge layer. Therefore a "corpuscular approach" that tries to calculate the noise current by evaluation of the currents that are induced by all carriers in the space charge region is cumbersome at the very least.[17]

Instead of calculating the current due to electrostatic induction caused by the carriers in the space charge layer, we can calculate the noise current in a way similar to the procedure used for the calculation of the average diode current in drift–diffusion theory: in the absence of generation and recombination in the space charge layer, the noise current of the pn diode may be calculated as the noise current of the minority carriers injected into the bulk regions as will be shown in the following section. Owing to the large currents crossing the space charge layer in both directions, we may assume that small fluctuations of minority-carrier density at the space charge layer boundaries are compensated almost immediately, resulting in negligibly small fluctuations of the minority-carrier density at the space charge layer boundaries.[18] Despite this, fluctuations of the minority-carrier densities in the interior of the bulk regions due to velocity fluctuations and generation recombination processes will cause fluctuations of the minority currents injected into the bulk regions.

[16]This discussion considers the electron current, it is taken as understood that the hole current behaves analogously.

[17]See also [26, 29–31].

[18]The adiabatic assumption (see the following subsection) is based on the same approximation as the Shockley boundary conditions; in situations where the Shockley boundary conditions cannot be applied (see Appendix C), the adiabatic boundary condition is expected to fail also.

The adiabatic assumption, which assumes that the carrier densities in the space charge layer are not subject to fluctuations under forward operation conditions, together with the coupling of fluctuations in the interior of the bulk regions to the injected minority currents, explains the somewhat astonishing phenomenon that the noise current of a diode may be calculated from fluctuations in the bulk regions. In the following subsection, the spectral density of the noise current of a diode will be calculated using the Langevin approach as described by van Vliet [22–24, 32, 33]. In order to not obscure the physical insight by mathematical details, we restrict ourselves to one-dimensional conditions and solve the differential equations by elementary methods.

5.3.2 Langevin Approach to the Noise of pn Junction Diodes

In this section, the set of stochastic transport equations is solved under small-signal conditions to obtain the total noise current in the space charge layer,

$$\Delta i(t) = A_{\mathrm{j}} \left(\Delta J_{\mathrm{n}}(x,t) + \Delta J_{\mathrm{p}}(x,t) + \epsilon \frac{\partial \Delta E(x,t)}{\partial t} \right) . \tag{5.37}$$

Since the total noise current $\Delta i(t)$ is independent of position, it equals the noise current carried by the device terminals. The fluctuations $\Delta J_{\mathrm{n}}(x,t)$, $\Delta J_{\mathrm{p}}(x,t)$ and $\Delta E(x,t)$ have to be determined from the linearized stochastic transport equations, which are difficult to solve in the general case. If the calculation is restricted to the space charge layer, however, it is possible to simplify the set of small-signal stochastic transport equations by using the adiabatic approximation, according to which the electron and hole densities in the space charge region are assumed to be in equilibrium with the n- and p-type regions respectively. According to this assumption, fluctuations $\Delta n(x,t)$ and $\Delta p(x,t)$ of the mobile-carrier densities are restored almost immediately and may therefore be neglected. Assuming $\Delta n(x,t) = \Delta p(x,t) = 0$ within the space charge region, the linearized stochastic transport equations read

$$\Delta J_{\mathrm{n}} = e\mu_{\mathrm{n}} n \Delta E + e\eta_{\mathrm{n}}(x,t) , \tag{5.38}$$

$$\Delta J_{\mathrm{p}} = e\mu_{\mathrm{p}} p \Delta E - e\eta_{\mathrm{p}}(x,t) , \tag{5.39}$$

$$\frac{1}{e}\frac{\partial \Delta J_{\mathrm{n}}}{\partial x} = R - G - \gamma(x,t) , \tag{5.40}$$

$$\frac{1}{e}\frac{\partial \Delta J_{\mathrm{p}}}{\partial x} = -(R - G) + \gamma(x,t) . \tag{5.41}$$

Integration of (5.40) and (5.41) yields the current densities in the space charge layer:

$$\Delta J_\mathrm{n}(x,t) = e\int_{x_\mathrm{p}}^{x} [R-G-\gamma(x',t)]\,\mathrm{d}x' + \Delta J_\mathrm{n}(x_\mathrm{p},t)\,,$$

$$\Delta J_\mathrm{p}(x,t) = -e\int_{x_\mathrm{n}}^{x} [R-G-\gamma(x',t)]\,\mathrm{d}x' + \Delta J_\mathrm{p}(x_\mathrm{n},t)\,,$$

where x_n and x_p denote the space charge layer boundaries. If the potential difference across the junction is assumed to be constant, the condition

$$\frac{\partial}{\partial t}\int_{x_\mathrm{n}}^{x_\mathrm{p}} \Delta E(x,t)\,\mathrm{d}x = 0$$

must be fulfilled, i.e. averaging the total noise current (5.37) across the depletion layer yields

$$\Delta i(t) = \frac{1}{d_\mathrm{j}}\int_{x_\mathrm{n}}^{x_\mathrm{p}} \Delta i(t)\,\mathrm{d}x = \Delta i_1(t) + \Delta i_2(t)\,, \tag{5.42}$$

where $d_\mathrm{j} = x_\mathrm{p} - x_\mathrm{n}$ denotes the depletion layer width,

$$\Delta i_1(t) = A_\mathrm{j}\,[\,\Delta J_\mathrm{n}(x_\mathrm{p},t) + \Delta J_\mathrm{p}(x_\mathrm{n},t)\,] \tag{5.43}$$

is the noise current associated with the injection of minority carriers in the bulk regions, and

$$\begin{aligned}\Delta i_2(t) &= \frac{eA_\mathrm{j}}{d_\mathrm{j}}\int_{x_\mathrm{n}}^{x_\mathrm{p}}\int_{x_\mathrm{p}}^{x} [R-G-\gamma(x',t)]\,\mathrm{d}x'\,\mathrm{d}x \\ &\quad -\frac{eA_\mathrm{j}}{d_\mathrm{j}}\int_{x_\mathrm{n}}^{x_\mathrm{p}}\int_{x_\mathrm{n}}^{x} [R-G-\gamma(x',t)]\,\mathrm{d}x'\,\mathrm{d}x \\ &= -eA_\mathrm{j}\int_{x_\mathrm{n}}^{x_\mathrm{p}} [R-G-\gamma(x,t)]\,\mathrm{d}x\end{aligned} \tag{5.44}$$

is the noise current due to generation and recombination in the space charge layer. Since $\Delta i_1(t)$ and $\Delta i_2(t)$ are uncorrelated, the spectral density S_i of the noise current is the sum of the spectral densities of the two noise current contributions. Owing to the high quality of state-of-the-art semiconductor material, recombination occurs predominantly within the bulk regions; recombination in the space charge layer is observed only at small current levels.[19] In the following, therefore, only noise due to minority-carrier injection into the bulk regions will be investigated; we shall consider hole injection into

[19] A theory for the current noise associated with carrier generation and recombination in the space charge region was presented in [34], the problem was investigated within the framework of the Langevin approach in [22], and a somewhat different approach was used in [31]. All investigations led basically to the same result: in a forward-biased junction, the noise due to recombination at centers located within the space charge layer ranges from three-quarters of the shot noise to the full shot noise, depending on the injection level; and in a reverse-biased pn junction, the noise due to generation at centers within the space charge layer ranges from two-thirds of the shot noise to the full shot noise, depending on frequency.

an n-type region. The small-signal admittance and the noise current associated with hole injection into the n-type region have to be calculated from the following system of small-signal Langevin equations:

$$\begin{aligned} \mathrm{j}\omega\Delta\underline{n} &= -\frac{\Delta\underline{n}}{\tau_{\mathrm{n}}} + \frac{1}{e}\frac{\mathrm{d}}{\mathrm{d}x}\Delta\underline{J}_{\mathrm{n}} + \underline{\gamma}\,, && (5.45)\\ \mathrm{j}\omega\Delta\underline{p} &= -\frac{\Delta\underline{n}}{\tau_{\mathrm{n}}} - \frac{1}{e}\frac{\mathrm{d}}{\mathrm{d}x}\Delta\underline{J}_{\mathrm{p}} + \underline{\gamma}\,, && (5.46)\\ \Delta\underline{J}_{\mathrm{n}} &= e\mu_{\mathrm{n}}n\Delta\underline{E} + eD_{\mathrm{n}}\frac{\mathrm{d}}{\mathrm{d}x}\Delta\underline{n} + e\,\underline{\eta}_{\mathrm{n}}\,, && (5.47)\\ \Delta\underline{J}_{\mathrm{p}} &= e\mu_{\mathrm{p}}p\Delta\underline{E} - eD_{\mathrm{p}}\frac{\mathrm{d}}{\mathrm{d}x}\Delta\underline{p} - e\,\underline{\eta}_{\mathrm{p}}\,, && (5.48)\\ \frac{\mathrm{d}}{\mathrm{d}x}\Delta\underline{E} &= \frac{e}{\epsilon}\left(\Delta\underline{p} - \Delta\underline{n}\right)\,, && (5.49) \end{aligned}$$

It is assumed here that the electron and hole recombination rates are equal, i.e. trapping and detrapping effects at the recombination centers are not considered. In the following, complete screening of injected minority carriers will be assumed, i.e. $\Delta\underline{n} = \Delta\underline{p}$. This is allowed if the timescale is large in comparison with the dielectric relaxation time, as can be seen from the following considerations. Combining the continuity and current equations for holes with the small-signal Poisson equation results in

$$\mathrm{j}\omega\Delta\underline{p} = -\frac{\Delta\underline{n}}{\tau_{\mathrm{n}}} - \mu_{\mathrm{p}}\frac{\mathrm{d}p}{\mathrm{d}x}\Delta\underline{E} - \frac{e\mu_{\mathrm{p}}p}{\epsilon}(\Delta\underline{p} - \Delta\underline{n}) + D_{\mathrm{p}}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\Delta\underline{p} - \underline{\zeta}_{\mathrm{p}}\,,$$

where

$$\underline{\zeta}_{\mathrm{p}}(x) = \underline{\gamma}(x) + \mathrm{d}\underline{\eta}_{\mathrm{p}}/\mathrm{d}x = \underline{\gamma}(x) + \underline{\eta}'_{\mathrm{p}}(x)\,, \qquad (5.50)$$

and the small-signal electric field strength is calculated from the linearized Poisson equation (5.49). If we write the dielectric relaxation time in the p region as $\tau_\epsilon = \epsilon/(e\mu_{\mathrm{p}}p)$, this equation transforms to

$$\Delta\underline{n} - \Delta\underline{p} = \mathrm{j}\omega\tau_\epsilon\Delta\underline{n} + \frac{\tau_\epsilon}{\tau_{\mathrm{n}}}\Delta\underline{n} + \tau_\epsilon\mu_{\mathrm{p}}\frac{\mathrm{d}p}{\mathrm{d}x}\Delta\underline{E} - \tau_\epsilon D_{\mathrm{p}}\frac{\mathrm{d}^2}{\mathrm{d}x^2}\Delta\underline{p} + \tau_\epsilon\underline{\zeta}_{\mathrm{p}}\,.$$

Since $\tau_\epsilon \ll \tau_{\mathrm{n}} = \tau_{\mathrm{p}}$ is a very small quantity, the excess electron density will be almost completely screened by holes, as can be seen by taking the limit $\tau_\epsilon \to 0$. Therefore $\Delta\underline{p} = \Delta\underline{n}$ may be assumed if the system is considered on a timescale large compared with τ_ϵ and on a length scale large compared with the Debye length $L_{\mathrm{D}} = \sqrt{\epsilon V_{\mathrm{T}}/ep}$. In this case $\Delta\underline{E} = 0$, and it suffices to analyze the following stochastic diffusion equation to obtain the small-signal hole density $\Delta\underline{p}$:

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\Delta\underline{p} - \frac{\Delta\underline{p}}{\underline{L}_\mathrm{p}^2} = \frac{1}{D_\mathrm{p}}\underline{\zeta}_\mathrm{p}(x)\ . \tag{5.51}$$

Equation (5.51) allows us to calculate the phasor $\Delta\underline{I}_\mathrm{p}(0)$ of the small-signal noise current due to holes injected into the n-type region. Equation (5.51) is an inhomogeneous linear differential equation of second order, and two boundary conditions must be specified in order to obtain the required solution. Owing to the large forward and reverse currents crossing the space charge layer, the hole density at the edge of the space charge layer of the n-type region (i.e. at $x = 0$) may be assumed to be in equilibrium with the holes on the p-type side, i.e. the fluctuations of the hole density at the space charge layer edge may be neglected, resulting in the boundary condition $\Delta\underline{p}(0) = 0$ (the adiabatic assumption). The solution of (5.51) is then

$$\Delta\underline{p} = \underline{A}\sinh\left(\frac{x}{\underline{L}_\mathrm{p}}\right) + \frac{\underline{L}_\mathrm{p}}{D_\mathrm{p}}\int_0^x \underline{\zeta}_\mathrm{p}(x')\sinh\left(\frac{x-x'}{\underline{L}_\mathrm{p}}\right)\mathrm{d}x'\ ; \tag{5.52}$$

the parameter $\underline{A}$ has to be chosen in order to satisfy the second boundary condition, which depends on the extent of the n-type region and on the type of contact. From these data, the noise portion $\Delta\underline{I}_\mathrm{p}(0)$ of the hole current injected into the n-type layer can be calculated, and from that, $\langle\Delta\underline{I}_\mathrm{p}(0)\Delta\underline{I}_\mathrm{p}^*(0)\rangle$. Together with the spectral functions

$$S_{\gamma\mathrm{p}}(x,x',\omega) = \frac{2\left[p_\mathrm{n}(x)+p_\mathrm{n0}\right]}{A_\mathrm{j}\tau_\mathrm{p}}\,\delta(x-x') \tag{5.53}$$

and

$$S_{\eta\mathrm{p}}(x,x',\omega) = \frac{4}{A_\mathrm{j}}\,D_\mathrm{p}p_\mathrm{n}(x)\delta(x-x') \tag{5.54}$$

of the local noise sources, $\langle\Delta\underline{I}_\mathrm{p}(0)\Delta\underline{I}_\mathrm{p}^*(0)\rangle$ leads to the spectral density of the "bulk portion" of the noise current.

Long-Base Diode

In the case of a one-sided long-base diode with an n-type region large in comparison with the hole diffusion length L_p, we require[20] $\Delta\underline{p} \to 0$ as $x \to \infty$; this requires

$$\underline{A} = \frac{\underline{L}_\mathrm{p}}{D_\mathrm{p}}\int_0^\infty \underline{\zeta}_\mathrm{p}(x)\mathrm{e}^{-x/\underline{L}_\mathrm{p}}\,\mathrm{d}x\ .$$

An integration by parts then gives the following for hole noise current $\underline{I}_\mathrm{p}(0) = eA_\mathrm{j}D_\mathrm{p}\underline{A}/\underline{L}_\mathrm{p}$ injected into the n-type region:

[20]This requirement has to be considered as an approximation, since $p_\mathrm{n} \to p_\mathrm{n0}$ as $x \to \infty$, and p_n0 is subject to fluctuations.

$$\underline{I}_{\mathrm{p}}(0) \;=\; eA_{\mathrm{j}}\int_0^\infty \underline{\gamma}(x)\mathrm{e}^{-x/\underline{L}_{\mathrm{p}}}\,\mathrm{d}x + \frac{eA_{\mathrm{j}}}{\underline{L}_{\mathrm{p}}}\int_0^\infty \underline{\kappa}_{\mathrm{p}}(x)\mathrm{e}^{-x/\underline{L}_{\mathrm{p}}}\,\mathrm{d}x\,.$$

From this result, the spectral density $S_{i\mathrm{p}}(\omega) = \langle \underline{I}_{\mathrm{p}}(0)\underline{I}_{\mathrm{p}}^*(0)\rangle$ of the noise current due to injected holes can be calculated as

$$\begin{aligned} S_{i\mathrm{p}}(\omega) &= e^2A_{\mathrm{j}}^2\int_0^\infty\int_0^\infty S_{\gamma\mathrm{p}}(x,x',\omega)\mathrm{e}^{-x/\underline{L}_{\mathrm{p}}}\,\mathrm{e}^{-x'/\underline{L}_{\mathrm{p}}^*}\,\mathrm{d}x\,\mathrm{d}x' \\ &+ \frac{e^2A_{\mathrm{j}}^2}{|\underline{L}_{\mathrm{p}}|^2}\int_0^\infty\int_0^\infty S_{\eta\mathrm{p}}(x,x',\omega)\mathrm{e}^{-x/\underline{L}_{\mathrm{p}}}\,\mathrm{e}^{-x'/\underline{L}_{\mathrm{p}}^*}\,\mathrm{d}x\,\mathrm{d}x'\,. \end{aligned}$$

Taking account of the fact that

$$p_{\mathrm{n}}(x) \;=\; [\,p_{\mathrm{n}}(0)-p_{\mathrm{n}0}\,]\,\mathrm{e}^{-x/L_{\mathrm{p}}} + p_{\mathrm{n}0} \;=\; \frac{1}{eA_{\mathrm{j}}}\frac{L_{\mathrm{p}}}{D_{\mathrm{p}}}\left(I_{\mathrm{p}}\mathrm{e}^{-x/L_{\mathrm{p}}} + I_{\mathrm{Sp}}\right)\,,$$

where I_{p} is the (dc) hole current injected at the bias point, $I_{\mathrm{Sp}} = eA_{\mathrm{j}}D_{\mathrm{p}}p_{\mathrm{n}0}/L_{\mathrm{p}}$ and $L_{\mathrm{p}} = \sqrt{D_{\mathrm{p}}\tau_{\mathrm{p}}}$, the spectral densities (5.53) and (5.54) of the local noise sources can be rewritten as

$$S_{\gamma\mathrm{p}}(x,x',\omega) \;=\; \frac{2}{eA_{\mathrm{j}}^2L_{\mathrm{p}}}\left(I_{\mathrm{p}}\mathrm{e}^{-x/L_{\mathrm{p}}} + 2I_{\mathrm{Sp}}\right)\delta(x-x') \tag{5.55}$$

and

$$S_{\eta\mathrm{p}}(x,x',\omega) \;=\; \frac{4L_{\mathrm{p}}}{eA_{\mathrm{j}}^2}\left(I_{\mathrm{p}}\mathrm{e}^{-x/L_{\mathrm{p}}} + I_{\mathrm{Sp}}\right)\delta(x-x')\,. \tag{5.56}$$

Evaluating the integrals with respect to x' then yields the following for the spectral density of the noise current:

$$\begin{aligned} S_{i\mathrm{p}}(\omega) &= \frac{2eI_{\mathrm{p}}}{L_{\mathrm{p}}}\int_0^\infty \mathrm{e}^{-(1+2a)x/L_{\mathrm{p}}}\,\mathrm{d}x + \frac{4eI_{\mathrm{Sp}}}{L_{\mathrm{p}}}\int_0^\infty \mathrm{e}^{-2ax/L_{\mathrm{p}}}\,\mathrm{d}x \\ &+ \frac{4eI_{\mathrm{p}}\sqrt{1+\omega^2\tau_{\mathrm{p}}^2}}{L_{\mathrm{p}}}\int_0^\infty \mathrm{e}^{-(1+2a)x/L_{\mathrm{p}}}\,\mathrm{d}x \\ &+ \frac{4eI_{\mathrm{Sp}}\sqrt{1+\omega^2\tau_{\mathrm{p}}^2}}{L_{\mathrm{p}}}\int_0^\infty \mathrm{e}^{-2ax/L_{\mathrm{p}}}\,\mathrm{d}x\,, \end{aligned}$$

where $a = \mathrm{Re}\left(\sqrt{1+\mathrm{j}\omega\tau_{\mathrm{p}}}\right)$. Evaluating the integrals with respect to x now yields

$$\begin{aligned} S_{i\mathrm{p}}(\omega) &= 2eI_{\mathrm{p}}\left(\frac{1}{1+2a} + \frac{2\sqrt{1+\omega^2\tau_{\mathrm{p}}^2}}{1+2a}\right) + 2eI_{\mathrm{Sp}}\left(\frac{1}{a} + \frac{\sqrt{1+\omega^2\tau_{\mathrm{p}}^2}}{a}\right) \\ &= 4e(I_{\mathrm{p}}+I_{\mathrm{Sp}})a - 2eI_{\mathrm{p}} \;=\; 4k_{\mathrm{B}}Tg_{\mathrm{dp}}a - 2eI_{\mathrm{p}}\,, \end{aligned} \tag{5.57}$$

where $g_{\mathrm{dp}} = (I_{\mathrm{p}} + I_{\mathrm{Sp}})/V_{\mathrm{T}}$ is the low-frequency small-signal conductance of the diode due to injected holes. Since $g_{\mathrm{dp}}a = \mathrm{Re}(y_{\mathrm{p}})$, where y_{p} denotes the small-signal admittance due to injected holes, this formula is equivalent to[21]

$$S_{i\mathrm{p}}(\omega) \; = \; 4k_{\mathrm{B}}T\,\mathrm{Re}(y_{\mathrm{p}}) - 2eI_{\mathrm{p}} \; . \tag{5.58}$$

In the limit $\omega\tau_{\mathrm{p}} \ll 1$, this expression simplifies to the standard shot noise formula (5.34). At low frequencies we may expand a to obtain the approximation

$$S_i(\omega) \; \approx \; 2e(I_{\mathrm{p}}+2I_{\mathrm{Sp}}) + \frac{1}{2}e(I+I_{\mathrm{Sp}})(\omega\tau_{\mathrm{p}})^2 \; .$$

If $I \gg I_{\mathrm{Sp}}$, as is the case in a forward-biased diode, the relative error $\Delta S_i(\omega)/S_{\mathrm{p}}(0) \approx (\omega\tau_{\mathrm{p}})^2/4$ of the classical shot noise formula increases in proportion to ω^2.

Short-Base Diode with Metal Contact

In the case of a metal-contacted short-base diode, we require $\Delta \underline{p} = 0$ at the contact, assuming the metal contact to be able to restore thermal-equilibrium conditions without delay. This results in a noise current with a spectral density that can be written as

$$S_{i\mathrm{p}}(\omega) \; = \; 2eI_{\mathrm{p}}f_{\mathrm{p1}}(\omega) + 4eI_{\mathrm{Sp}}f_{\mathrm{p2}}(\omega) \; , \tag{5.59}$$

where the frequency-dependent functions are calculated in Appendix E. In the limit $d/L_{\mathrm{p}} \ll 1$, these functions can be expanded into power series of (d/L_{p}) up to second order, resulting in the approximations $f_{\mathrm{p1}}(\omega) \approx 1$ and $f_{\mathrm{p2}}(\omega) \approx 1$. Expanding $f_{\mathrm{p1}}(\omega)$ up to second order in ω yields

$$f_{\mathrm{p1}}(\omega) \; \approx \; 1 + (\omega\tau_{\mathrm{fp1}})^2 \; ,$$

[21] This result is equivalent to that obtained by van der Ziel using a "transmission line analogy" [35] and has also been obtained by Guggenbuehl and Strutt [36] using an interesting approach based on two assumptions. The Nyquist formula $S_i(f) = 4k_{\mathrm{B}}T\,\mathrm{Re}(y_{\mathrm{d}})$ cannot be applied to a biased diode, as such a device is not in thermal equilibrium, and an additional fluctuation current is observed due to the current I passing through the diode. In [36] it has been assumed that this additional noise current can be balanced out by supplying an additional noise current $i_{\mathrm{n}I}$ to the diode, and quasi-thermal equilibrium conditions can be restored. This additional noise current source was assumed to show no correlation with the Nyquist noise, to deliver power to the diode under forward operation, and to sink power if the diode was reverse-operated. This assumption was found to be reasonable, as the majority of the carriers constituting the direct current I have to surmount the internal potential hill of the diode, thus consuming energy, if I is a forward current; the additional current source that is supposed to restore a thermal-equilibrium energy distribution should therefore deliver power to the diode in forward operation. The noise current associated with the directed motion of carriers has been described as shot noise with spectral density $2eI$, as long as the motion of carriers across the depletion layer may be assumed to be uncorrelated (low-level injection), resulting in (5.58).

where

$$\tau_{\mathrm{fp1}}^2 = \frac{\tau_{\mathrm{p}}^2}{2} \frac{\sinh^2(d/L_{\mathrm{p}}) + (d/L_{\mathrm{p}}) \tanh(d/L_{\mathrm{p}}) - 2(d^2/L_{\mathrm{p}}^2)}{\cosh(2d/L_{\mathrm{p}}) - 1} .$$

For small values of (d/L_{p}), the approximation

$$\tau_{\mathrm{fp1}}^2 \approx \frac{2\tau_{\mathrm{p}}^2}{45} \left(\frac{d}{L_{\mathrm{p}}}\right)^4$$

may be used. Corrections to the elementary shot noise formula are therefore of order $(d/L_{\mathrm{p}})^4$; in heavily doped short-base diodes with a minority-carrier lifetime of the order of nanoseconds and $d/L_{\mathrm{p}} \ll 1$, these corrections can be assumed to be negligible.

Short-Base Diode with Polysilicon Contact

In a forward-biased short-base diode with a poly-silicon contact, the excess hole density at the contact will be different from zero. If the contact is characterized by a surface recombination velocity S_{p}, the following relation between the small-signal hole current density and the small-signal excess hole density at the contact applies:

$$\Delta \underline{J}_{\mathrm{p}}(d) = -eD_{\mathrm{p}} \frac{\mathrm{d}}{\mathrm{d}x} \Delta \underline{p} \bigg|_d = eS_{\mathrm{p}}[\Delta \underline{p}(d) + \underline{\kappa}_{\mathrm{r}}] . \tag{5.60}$$

The term $\underline{\kappa}_{\mathrm{r}}$ denotes a stochastic term that takes account of fluctuations of the recombination at the surface. The spectral density of this noise source is determined by

$$S_{\kappa}(\omega) = \frac{2}{A_{\mathrm{j}}} S_{\mathrm{p}}[p_{\mathrm{n}}(d) + p_{\mathrm{n0}}] .$$

Equation (5.60) serves as a boundary condition that allows one to determine $\underline{A}$ and the spectral density of the noise current at $x = 0$. These calculations are outlined in Appendix E and lead to the result

$$S_{i\mathrm{p}}(\omega) = 2eI_{\mathrm{p}} f_{\mathrm{p1}}(\omega) + 4eI_{\mathrm{Sp}} f_{\mathrm{p2}}(\omega) \tag{5.61}$$

where

$$f_{\mathrm{p1}}(\omega) = \frac{\theta_1(d)}{\theta_2(d)\,|\underline{\theta}_1(d)|^2} \Bigg[a \left(\nu^2 \sqrt{1+\omega^2\tau_{\mathrm{p}}^2} + 1\right) \sinh\left(\frac{2ad}{L_{\mathrm{p}}}\right)$$
$$+\nu \left(\sqrt{1+\omega^2\tau_{\mathrm{p}}^2} + 1\right) \cosh\left(\frac{2ad}{L_{\mathrm{p}}}\right) + \nu \left(\sqrt{1+\omega^2\tau_{\mathrm{p}}^2} - 1\right) \cos\left(\frac{2bd}{L_{\mathrm{p}}}\right)$$
$$-b \left(\nu^2 \sqrt{1+\omega^2\tau_{\mathrm{p}}^2} - 1\right) \sin\left(\frac{2bd}{L_{\mathrm{p}}}\right) \Bigg] - 1 , \tag{5.62}$$

and $\underline{\theta}_1(x) = \sinh(x/\underline{L}_\mathrm{p}) + \underline{\nu}\cosh(x/\underline{L}_\mathrm{p})$, $\underline{\theta}_2(x) = \cosh(x/\underline{L}_\mathrm{p}) + \underline{\nu}\sinh(x/\underline{L}_\mathrm{p})$ and $\underline{\nu} = D_\mathrm{p}/(S_\mathrm{p}\underline{L}_\mathrm{p})$, if fluctuations of the surface recombination velocity are neglected. From this solution, the result obtained for a metal-contacted emitter can be derived in the limit $\nu \to 0$. In the limit $\omega \to 0$ we obtain $f_{\mathrm{p}1} \to 1$, and (5.61) reduces to the classical shot noise formula, i.e. $f_{\mathrm{p}1} \to 1$ as $\omega \to 0$. As shown in Appendix E, this result is equivalent to

$$S_{i\mathrm{p}} = 4k_\mathrm{B}T\,\mathrm{Re}(y_\mathrm{p}) - 2eI_\mathrm{p}\,, \tag{5.63}$$

where

$$y_\mathrm{p} = \frac{I_\mathrm{p} + I_\mathrm{Sp}}{V_\mathrm{T}}\,\frac{\theta_1(d)}{\theta_2(d)}\,\frac{\underline{\theta}_2(d)}{\underline{\theta}_1(d)}\sqrt{1 + \mathrm{j}\omega\tau_\mathrm{p}} \tag{5.64}$$

denotes the ac small-signal diode admittance due to hole injection into the n-type emitter region.

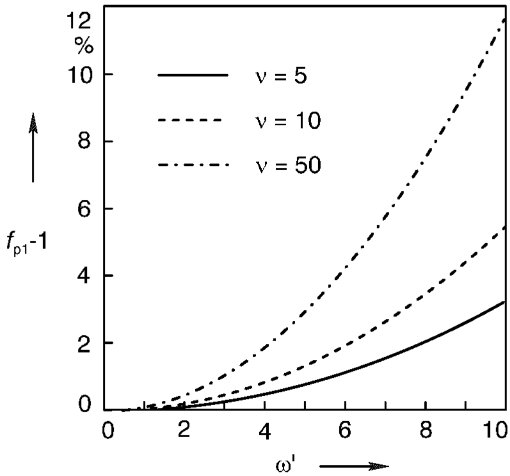


Fig. 5.5. Plot of $f_{\mathrm{p}1} - 1$ vs. $\omega\tau_\mathrm{p}$ for short-base diodes with $d/L_\mathrm{p} = 1/20$ for different values of the effective surface recombination velocity at the contact, characterized in terms of the parameter $\nu = D_\mathrm{p}/S_\mathrm{p}L_\mathrm{p}$

Figure 5.5 shows $f_{\mathrm{p}1}(\omega) - 1$ as a function of $\omega\tau_\mathrm{p}$. Short-base polysilicon diodes thus show an increase of noise current with frequency, which becomes more pronounced as the effective surface recombination velocity S_p becomes smaller (i.e. as ν increases). This behavior is to be expected if the current carried by a short-base diode is described as the difference of two strongly correlated currents I_f and I_r, where I_r increases with ν and the correlation between the two currents decreases with frequency.

5.4 Noise Generated by the Transfer Current

Reliable noise models of the bipolar transistor are needed to accurately predict the noise performance of high-frequency circuits over the whole frequency range of interest. In Sect. 1.8 the small-signal noise equivalent circuit of the bipolar transistor used in SPICE was described. If series resistances are neglected, the simplified equivalent circuit of Fig. 5.6 is obtained.

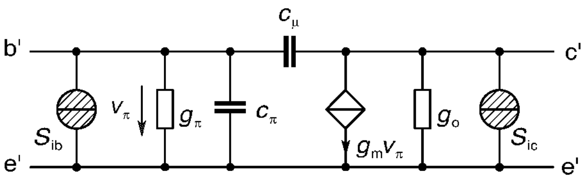


Fig. 5.6. Noise equivalent circuit of the internal bipolar transistor as used in SPICE, neglecting series resistance effects

At high frequencies $1/f$ noise is negligible and the noise current generators have spectral densities

$$S_{ib} = 2eI_B \quad \text{and} \quad S_{ic} = 2eI_T , \tag{5.65}$$

where we can assume that $I_B \approx I_{BE}$ and $I_T \approx I_C$ for forward operation. Correlation between the two noise current sources is neglected. This model is a good approximation over a wide frequency range; at very high frequencies, however, the spectral densities become frequency-dependent and show correlation.

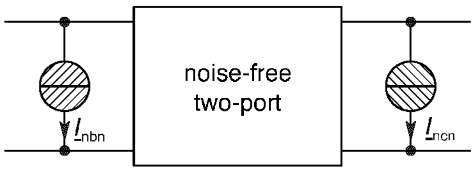


Fig. 5.7. General noise equivalent circuit for the description of the noise due to electron transport through the base region with input and output short-circuited

A general description can be obtained from a computation of the noise current phasors $\underline{I}_{nen}$ and $\underline{I}_{ncn}$, which describe the noise current of the electrons crossing the emitter- and collector-side depletion layer boundaries of the base region. In the common-emitter configuration, this noise can be described by a noise current source $\underline{I}_{nbn}$ connected to the input port and a noise current source $\underline{I}_{ncn}$ connected to the output port, as shown in Fig. 5.7. Since

$$\underline{I}_{nbn} = \underline{I}_{nen} - \underline{I}_{ncn} ,$$

the spectral densities of the input and output noise current sources $\langle \underline{I}_{nbn} \underline{I}^*_{nbn} \rangle$ and $\langle \underline{I}_{ncn} \underline{I}^*_{ncn} \rangle$, and the complex correlation coefficient $\langle \underline{I}_{ncn} \underline{I}^*_{nbn} \rangle$, can be calculated. In order to obtain analytical solutions, a drift transistor with a constant electric field strength E throughout the base region is considered in the following. For ease of notation, we choose the coordinate system such that $x_{\rm be} = 0$ and $x_{\rm bc} = d_{\rm B}$. Assuming $\Delta\underline{E} = 0$, i.e. instantaneous screening of injected minority carriers, the small-signal Langevin equations for the electron density read

$$\mathrm{j}\omega\Delta\underline{n} = -\frac{\Delta\underline{n}}{\tau_{\rm n}} + \frac{1}{e}\frac{\mathrm{d}}{\mathrm{d}x}\Delta\underline{J}_{\rm n} + \underline{\gamma}\,, \tag{5.66}$$

$$\Delta\underline{J}_{\rm n} = e\mu_{\rm n}E\Delta\underline{n} + eD_{\rm n}\frac{\mathrm{d}}{\mathrm{d}x}\Delta\underline{n} + e\,\underline{\eta}_{\rm n}\,. \tag{5.67}$$

Combining these equations, we obtain the second-order stochastic differential equation

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\Delta\underline{n} + \frac{E}{V_{\rm T}}\frac{\mathrm{d}}{\mathrm{d}x}\Delta\underline{n} - \frac{\Delta\underline{n}}{\underline{L}_{\rm n}^2} = -\frac{1}{D_{\rm n}}\underline{\zeta}_{\rm n}(x)\,, \tag{5.68}$$

where $\underline{\zeta}_{\rm n} = \underline{\gamma} + \mathrm{d}\underline{\kappa}_{\rm n}/\mathrm{d}x$. This equation can be solved (see Appendix E) using the boundary conditions $\Delta\underline{n}(0)$ and $\Delta\underline{n}(d_{\rm B}) = 0$. The first condition is in accordance with the adiabatic assumption for the forward-biased eb diode, while the second neglects velocity saturation in the bc depletion layer. For a finite value of the electron drift velocity in the bc depletion layer, the electron density and its fluctuations will not be zero; under low-level-injection conditions, the electron density will be small, however, justifying our approximation. From the solutions obtained, the electron noise currents $\underline{I}_{\rm nen}$ and $\underline{I}_{\rm ncn}$ at the emitter- and collector-side depletion layer edges can be derived, with the result

$$\underline{I}_{\rm nen} = \frac{eA_{\rm jc}}{\sinh(\underline{\vartheta})}\int_0^{d_{\rm B}}\underline{\zeta}_{\rm n}(x)\exp\left(-\frac{\eta x}{2d_{\rm B}}\right)\sinh\left(\frac{\underline{\vartheta}(d_{\rm B}-x)}{d_{\rm B}}\right)\mathrm{d}x \tag{5.69}$$

and

$$\underline{I}_{\rm ncn} = \frac{eA_{\rm je}}{\sinh(\underline{\vartheta})}\int_0^{d_{\rm B}}\underline{\zeta}_{\rm n}(x)\exp\left(\frac{\eta(d_{\rm B}-x)}{2d_{\rm B}}\right)\sinh\left(-\frac{\underline{\vartheta}\,x}{d_{\rm B}}\right)\mathrm{d}x\,, \tag{5.70}$$

where

$$\underline{\vartheta} = \sqrt{(\eta/2)^2 + 2\mathrm{j}\omega\tau_{\rm B0}} = \frac{\eta}{2}(\alpha + \mathrm{j}\beta)\,. \tag{5.71}$$

In the above,

$$\alpha = \sqrt{\frac{\sqrt{1+(\omega\tilde{\tau})^2}+1}{2}}\,,\qquad \beta = \sqrt{\frac{\sqrt{1+(\omega\tilde{\tau})^2}-1}{2}} \tag{5.72}$$

and

$$\tilde{\tau} = \frac{4\tau_{\rm Bf}}{\eta-1+\exp(-\eta)}\,,\quad \text{where}\quad \tau_{\rm Bf} = \frac{2\tau_{\rm B0}}{\eta^2}\left(\eta-1+\mathrm{e}^{-\eta}\right)\,. \tag{5.73}$$

The above results apply in the limit $\tau_\mathrm{n} \to \infty$, i.e. if recombination in the base is neglected. From these results the spectral densities of the noise current sources connected to the input and output, as well as their correlation, can be calculated.

Transfer Current Noise

If we take $n_\mathrm{p0} \approx 0$, we obtain from these equations (see Appendix E)

$$\langle \underline{I}_\mathrm{ncn} \underline{I}^*_\mathrm{ncn} \rangle = 2eI_\mathrm{T}\Delta f \qquad (5.74)$$

without any further approximation. The output current noise associated with the transfer current thus follows the classical shot noise formula. This result is also expected from comparison with the vacuum tube: the bc depletion layer of a bipolar transistor closely resembles a vacuum tube, where electrons are injected into a drift region in a random manner.

Base Current Noise

In the limit $\tau_\mathrm{n} \to \infty$, the spectral density of the electron noise current injected into the base region is

$$\begin{aligned}\frac{\langle \underline{I}_\mathrm{nen} \underline{I}^*_\mathrm{nen} \rangle}{\Delta f} &= \frac{2eI_\mathrm{T}}{\cosh(\eta\alpha) - \cos(\eta\beta)} \left\{ \mathrm{e}^{-\eta} \left[\cos(\eta\beta) - \cosh(\eta\alpha)\right] \right. \\ &\qquad \left. + \left(1 - \mathrm{e}^{-\eta}\right) \left[\alpha \sinh(\eta\alpha) + \beta \sin(\eta\beta)\right] \right\} . \qquad (5.75)\end{aligned}$$

This result may be expressed in terms of the internal common-base input admittance associated with electron injection into the base layer,

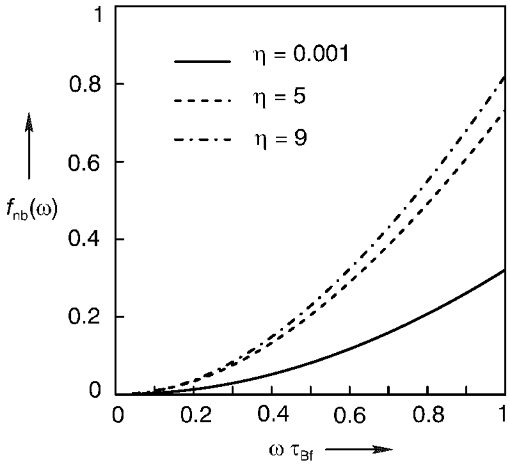


Fig. 5.8. Factor $f_\mathrm{nb}(\omega)$ vs. $\omega\tau_\mathrm{Bf}$ for different values of η

$$y_{\mathrm{n11b}} \;=\; \frac{I_{\mathrm{T}}}{V_{\mathrm{T}}}\,\mathrm{e}^{-\eta/2}\sinh\!\left(\frac{\eta}{2}\right)\left(\frac{\underline{\vartheta}}{\eta/2}\coth(\underline{\vartheta})+1\right)$$

as

$$\langle \underline{I}_{\mathrm{nen}}\underline{I}^*_{\mathrm{nen}}\rangle \;=\; 4k_{\mathrm{B}}T\mathrm{Re}(y_{\mathrm{n11b}})\Delta f - 2eI_{\mathrm{T}}\,\Delta f\;, \tag{5.76}$$

a result that was derived earlier by van der Ziel [35] for the case of a diffusion transistor. The noise currents represented by $\underline{I}_{\mathrm{nen}}$ and $\underline{I}_{\mathrm{ncn}}$ show correlation, i.e.

$$\langle \underline{I}_{\mathrm{ncn}}\underline{I}^*_{\mathrm{nen}}\rangle \;=\; 2eI_{\mathrm{T}}\,\frac{\underline{\vartheta}}{\eta/2}\,\frac{\sinh(\eta/2)}{\sinh\underline{\vartheta}} \tag{5.77}$$

differs from zero. Since

$$y_{\mathrm{n21b}} \;=\; g_{\mathrm{m}}\,\frac{\underline{\vartheta}}{\eta/2}\,\frac{\sinh(\eta/2)}{\sinh\underline{\vartheta}}\;, \tag{5.78}$$

where $g_{\mathrm{m}} = I_{\mathrm{T}}/V_{\mathrm{T}}$, represents the transfer admittance associated with the electron transport across the base layer, (5.77) may be written as

$$\langle \underline{I}_{\mathrm{ncn}}\underline{I}^*_{\mathrm{nen}}\rangle \;=\; 2k_{\mathrm{B}}Ty_{\mathrm{n21b}}\Delta f\;, \tag{5.79}$$

in analogy with van der Ziel's result for the diffusion transistor. The spectral density $\langle \underline{I}_{\mathrm{nbn}}\underline{I}^*_{\mathrm{nbn}}\rangle$ is obtained from

$$\begin{aligned}\langle \underline{I}_{\mathrm{nbn}}\underline{I}^*_{\mathrm{nbn}}\rangle \;&=\; \langle(\underline{I}_{\mathrm{nen}}-\underline{I}_{\mathrm{ncn}})(\underline{I}^*_{\mathrm{nen}}-\underline{I}^*_{\mathrm{ncn}})\rangle\\ &=\; 4k_{\mathrm{B}}T\mathrm{Re}(y_{\mathrm{n11b}}-y_{\mathrm{n21b}})\Delta f \;=\; 4k_{\mathrm{B}}T\,\mathrm{Re}(y_{\mathrm{n11e}})\Delta f\;.\end{aligned}$$

Computation of $\mathrm{Re}(y_{\mathrm{n11e}})$ yields the spectral density of the input noise current source,

$$\langle \underline{I}_{\mathrm{nbn}}\underline{I}^*_{\mathrm{nbn}}\rangle \;=\; 2eI_{\mathrm{T}}f_{\mathrm{nb}}(\omega)\Delta f\;, \tag{5.80}$$

where

$$f_{\mathrm{nb}}(\omega) \;=\; (1-\mathrm{e}^{-\eta})\left(1+\frac{\alpha\sinh(\eta\alpha)+\beta\sin(\eta\beta)}{\cosh(\eta\alpha)-\cos(\eta\beta)}\right) - 2\,[1+\psi_1(\omega)] \tag{5.81}$$

(Fig. 5.8). Here

$$\begin{aligned}\psi_1(\omega) \;&=\; \frac{2\sinh(\eta/2)}{\cosh(\eta\alpha)-\cos(\eta\beta)}\\ &\quad\times\left[\alpha\sinh\!\left(\frac{\eta\alpha}{2}\right)\cos\!\left(\frac{\eta\beta}{2}\right)+\beta\cosh\!\left(\frac{\eta\alpha}{2}\right)\sin\!\left(\frac{\eta\beta}{2}\right)\right]-1\end{aligned}$$

(Fig. 5.9); this quantity vanishes in proportion to β^2 in the low-frequency limit, where $\beta \to 0$.

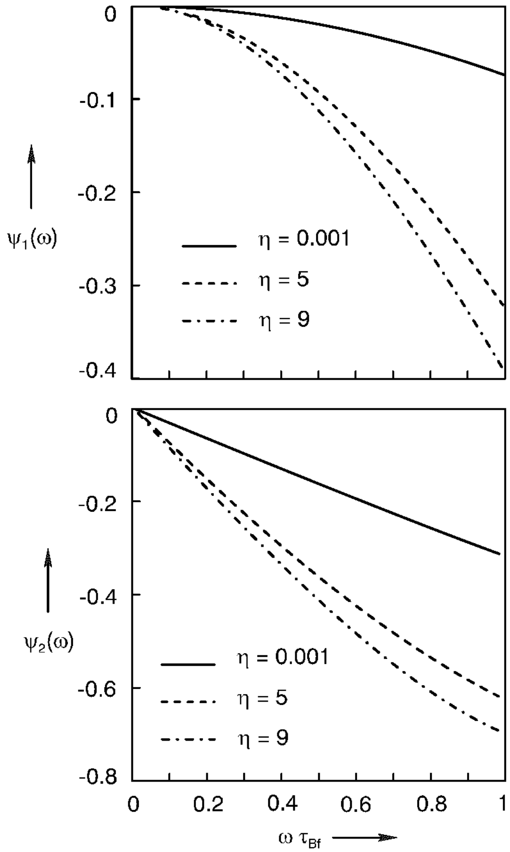


Fig. 5.9. Real part $\psi_1(\omega)$ and imaginary part $\psi_2(\omega)$ of the complex correlation coefficient (in units of $2eI_\mathrm{T}$) vs. $\omega\tau_\mathrm{Bf}$ for different values of η

The correlation between the transfer current and base current is described by

$$\begin{aligned}\langle \underline{I}_\mathrm{ncn}\underline{I}^*_\mathrm{nbn}\rangle &= \langle \underline{I}_\mathrm{ncn}\underline{I}^*_\mathrm{nen}\rangle - \langle \underline{I}_\mathrm{ncn}\underline{I}^*_\mathrm{ncn}\rangle = 2k_\mathrm{B}T(y_\mathrm{n21b}-g_\mathrm{m})\,\Delta f \\ &= 2eI_\mathrm{T}\left[\,\psi_1(\omega)+\mathrm{j}\psi_2(\omega)\right] , \qquad (5.82)\end{aligned}$$

where

$$\begin{aligned}\psi_2(\omega) &= \frac{2\sinh(\eta/2)}{\cosh(\eta\alpha)-\cos(\eta\beta)} \\ &\times \left[\beta\sinh\!\left(\frac{\eta\alpha}{2}\right)\cos\!\left(\frac{\eta\beta}{2}\right) - \alpha\cosh\!\left(\frac{\eta\alpha}{2}\right)\sin\!\left(\frac{\eta\beta}{2}\right)\right] .\end{aligned}$$

These results will be used for the computation of the noise factor in Sect. 5.6.

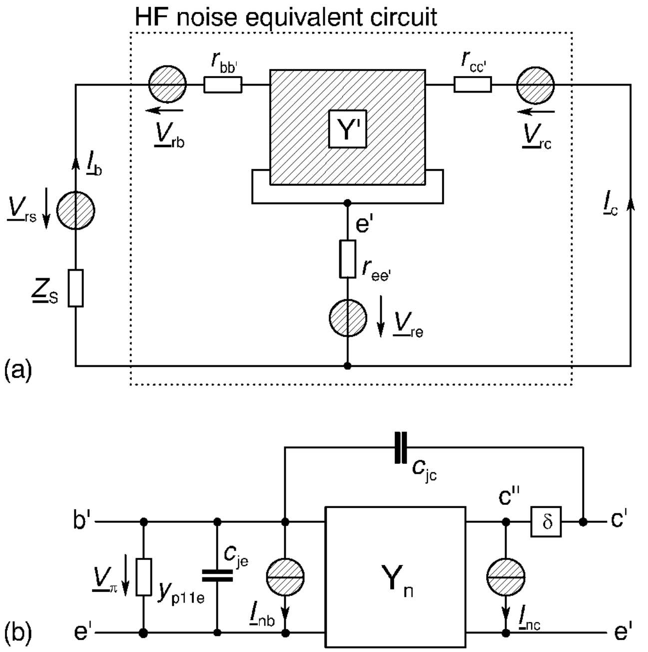


Fig. 5.10. (**a**) High-frequency noise equivalent circuit for the computation of the noise figure of a one-dimensional transistor; (**b**) equivalent circuit for the description of the high-frequency behavior of the internal transistor

5.5 High-Frequency Noise Equivalent Circuit

The results calculated up to now have not considered the noise current contributions of the hole current injected into the emitter or of the depletion capacitances and series resistances. For computation of the noise figure, the noise equivalent circuit shown in Fig. 5.10 has to be considered. The noise current source $\underline{I}_{nb} = \underline{I}_{nbn} + \underline{I}_{nbp}$ takes account of the base noise current associated with electrons and holes crossing the eb depletion layer. The two-port, described in terms of the admittance parameters $y'_{\alpha\beta}$, describes electron transport across the base layer, the delay due to the base-collector transit time, the hole current injected into the emitter region and charging and decharging of the eb and bc depletion capacitances. The admittance parameters can be expressed as

$$
\begin{aligned}
y'_{11} &= y_{n11e} + y_p + j\omega(c_{je} + c_{jc}) \;, \\
y'_{12} &= y_{n12e} - j\omega c_{jc} \;, \\
y'_{21} &= y_{n21e}\underline{\phi}(\omega) - j\omega c_{jc} \;, \\
y'_{22} &= y_{n22e}\underline{\phi}(\omega) + j\omega c_{jc} \;,
\end{aligned}
$$

where y_p denotes the small-signal admittance due to hole current injection in the emitter region (see Sect. 5.3), the $y_{\mathrm{n}\alpha\beta\mathrm{e}}$ denote the small-signal admittances associated with the electron transfer current, and

$$\underline{\phi}(\omega) = \frac{\sin(\omega\tau_\mathrm{jc})}{\omega\tau_\mathrm{jc}}\,\mathrm{e}^{-\mathrm{j}\omega\tau_\mathrm{jc}} \approx \mathrm{e}^{-\mathrm{j}\omega\tau_\mathrm{jc}}$$

describes the signal propagation across the bc depletion layer. In the equivalent circuit, this term is represented by a delay line with delay $\tau_\mathrm{jc} = d_\mathrm{jc}/2v_\mathrm{nsat}$ where d_jc denotes the width of the bc depletion layer and v_nsat is the saturation drift velocity for electrons. The noise sources in the noise equivalent circuit Fig. 5.10 are represented by phasors in the standard way, for example

$$\underline{V}_\mathrm{rb} = \sqrt{4k_\mathrm{B}Tr_{\mathrm{bb}'}\Delta f}\,\mathrm{e}^{\mathrm{j}\varphi_\mathrm{vrb}} \quad\text{and}\quad \underline{I}_\mathrm{nc} = \sqrt{2eI_\mathrm{T}\Delta f}\,\mathrm{e}^{\mathrm{j}\varphi_\mathrm{inc}} \, .$$

Except for $\underline{I}_\mathrm{nb}$ and $\underline{I}_\mathrm{nc}$, all these noise sources stem from independent physical processes and are therefore independent; ensemble averages of products of the stochastic phase factors therefore yield[22]

$$\langle \mathrm{e}^{\mathrm{j}\varphi_\alpha}\mathrm{e}^{-\mathrm{j}\varphi_\beta}\rangle = \delta_{\alpha\beta} \, .$$

The nonvanishing ensemble averages are listed in the following:

$$\langle \underline{V}_\mathrm{rb}\underline{V}_\mathrm{rb}^*\rangle = 4k_\mathrm{B}Tr_{\mathrm{bb}'}\Delta f \, , \tag{5.83}$$

$$\langle \underline{V}_\mathrm{re}\underline{V}_\mathrm{re}^*\rangle = 4k_\mathrm{B}Tr_{\mathrm{ee}'}\Delta f \, , \tag{5.84}$$

$$\langle \underline{V}_\mathrm{rc}\underline{V}_\mathrm{rc}^*\rangle = 4k_\mathrm{B}Tr_{\mathrm{cc}'}\Delta f \, , \tag{5.85}$$

$$\langle \underline{I}_\mathrm{nb}\underline{I}_\mathrm{nb}^*\rangle = \left[\mathrm{Re}(y_\mathrm{n11e}) + 2eI_\mathrm{T}f_\mathrm{pb}(\omega)/B_\mathrm{N}\right]\Delta f \tag{5.86}$$

$$\langle \underline{I}_\mathrm{nc}\underline{I}_\mathrm{nc}^*\rangle = 2eI_\mathrm{T}\Delta f \, , \tag{5.87}$$

$$\langle \underline{I}_\mathrm{nc}\underline{I}_\mathrm{nb}^*\rangle = 2k_\mathrm{B}T(y_\mathrm{n11e} - g_\mathrm{m})\Delta f \, . \tag{5.88}$$

Effects of Carrier Multiplication

Carrier multiplication in the base–collector junction is a stochastic phenomenon and therefore contributes to the noise of the transistor. The conventional noise equivalent circuit of the bipolar transistor has to be extended to take this into account.[23] Carrier multiplication amplifies transfer current noise, which is modeled according to (5.15). For $m_\mathrm{n} \to 1$, we have $\phi(m_\mathrm{n}) \to 1$ and $S_i \to 2eI_\mathrm{T} \approx 2eI_\mathrm{C}$. This noise is represented by a noise source between the internal collector node c$'$ and the internal emitter node e$'$ of the

[22] The symbol $\delta_{\alpha\beta}$ denotes the Kronecker delta, i.e. $\delta_{\alpha\beta} = 1$ if $\alpha = \beta$ and $\delta_{\alpha\beta} = 0$ if $\alpha \neq \beta$.

[23] For ease of notation, the internal transistor is described in terms of the quasi-static approximation (Giacoletto model).

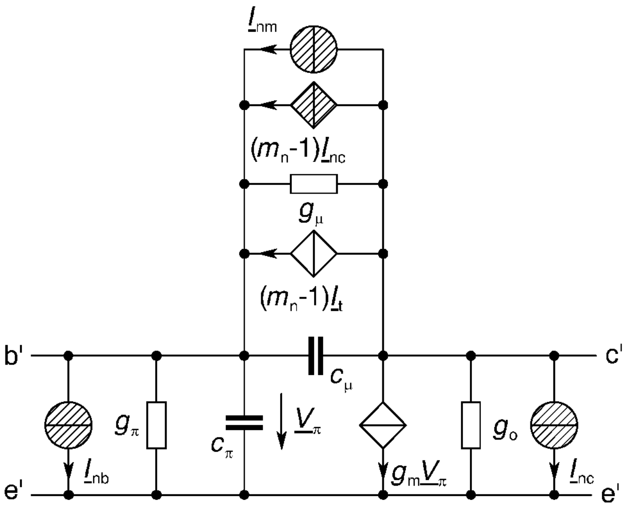


Fig. 5.11. Noise equivalent circuit of bipolar transistors (series resistances neglected) operated in the avalanche regime

small-signal equivalent circuit. The increment for $m_n > 1$ represents the noise of the generated current, which is separated in the bc space charge region and causes hole injection into the base region. Therefore an additional noise source between the internal collector node c′ and the internal base node b′ of the small-signal equivalent circuit has to be introduced, as shown in the noise equivalent circuit Fig.5.11. The noise due to carrier multiplication in the bc space charge layer is a superposition of two contributions. The first contribution describes the average multiplication of the statistically fluctuating transfer current; this part is represented by a controlled current source (diamond-shaped symbol) $(m_n - 1)\underline{I}_{nc}$ and is fully correlated with $\underline{I}_{nc}$. The second term describes stochastic fluctuations of the multiplication process and is therefore independent of all the other noise sources. This part is represented by an independent noise current source $\underline{I}_{nm}$ described by the phasor

$$\underline{I}_{nm} = \sqrt{2eI_T m_n^2[\phi(m_n)-1]\Delta f}\,e^{j\varphi_{nm}} . \tag{5.89}$$

The combined effect of the two noise sources yields a spectral density

$$S_i = \frac{1}{\Delta f}\langle|\underline{I}_{nm} + m_n\underline{I}_{nc}|^2\rangle = 2em_n^2\phi(m_n)I_T , \tag{5.90}$$

in accordance with (5.15). A formula for the noise figure of a bipolar transistor operated in the avalanche regime is derived in the following section on the basis of this model.

5.6 Noise Figure

An expression for the noise figure[24] of a noisy two-port connected to a real source impedance has already been considered in Sect. 1.8. If R_S is replaced by a complex source impedance Z_S, the expression (1.178) for the noise factor has to be modified to

$$F = 1 + \frac{|Z_\mathrm{S}|^2 S_i + S_v + 2\,\mathrm{Re}(Z_\mathrm{S}\underline{\gamma}_{iv})\sqrt{S_i S_v}}{4k_\mathrm{B}T\,\mathrm{Re}(Z_\mathrm{S})} \tag{5.91}$$

5.6.1 Noise Caused by the Transfer Current

In this section, the noise factor is calculated using the results of Sect. 5.4 and compared to the outcome of the conventional SPICE model; for simplicity only the noise caused by the transfer current is considered. Neglecting series resistance effects, the phasors of the equivalent noise sources $\underline{V}_\mathrm{n}$ and $\underline{I}_\mathrm{n}$ are

$$\underline{V}_\mathrm{n} = -\underline{I}_\mathrm{ncn}/y_\mathrm{n21e}\,, \tag{5.92}$$

$$\underline{I}_\mathrm{n} = \underline{I}_\mathrm{nbn} - y_\mathrm{n11e}\,\underline{I}_\mathrm{ncn}/y_\mathrm{n21e}\,, \tag{5.93}$$

where y_n11e and y_n21e denote the internal common-emitter input and transfer admittances associated with electron transport through the base region. These are related to the admittances in common-base configuration by

$$y_\mathrm{n21e} = y_\mathrm{n21b} = g_\mathrm{m}\,\frac{\underline{\vartheta}}{\eta/2}\,\frac{\sinh(\eta/2)}{\sinh\underline{\vartheta}}\,, \tag{5.94}$$

$$y_\mathrm{n11e} = y_\mathrm{n11b} - y_\mathrm{n21b}\,; \tag{5.95}$$

where $g_\mathrm{m} = I_\mathrm{T}/V_\mathrm{T}$. Using the results

$$\begin{aligned}
\langle \underline{I}_\mathrm{ncn}\underline{I}^*_\mathrm{ncn}\rangle &= 2eI_\mathrm{T}\,\Delta f = 2k_\mathrm{B}Tg_\mathrm{m}\Delta f\,,\\
\langle \underline{I}_\mathrm{nbn}\underline{I}^*_\mathrm{nbn}\rangle &= 4k_\mathrm{B}T\,\mathrm{Re}(y_\mathrm{n11e})\Delta f\,,\\
\langle \underline{I}_\mathrm{ncn}\underline{I}^*_\mathrm{nbn}\rangle &= 2k_\mathrm{B}T(y_\mathrm{n21e}-g_\mathrm{m})\Delta f\,,
\end{aligned}$$

we obtain

$$S_v = \frac{\langle \underline{V}_\mathrm{n}\underline{V}^*_\mathrm{n}\rangle}{\Delta f} = \frac{2eI_\mathrm{T}}{|y_\mathrm{n21e}|^2} = \frac{2k_\mathrm{B}Tg_\mathrm{m}}{|y_\mathrm{n21e}|^2} \tag{5.96}$$

for the spectral density of the equivalent input noise voltage generator, and

$$S_i = \frac{\langle \underline{I}_\mathrm{n}\underline{I}^*_\mathrm{n}\rangle}{\Delta f} = 2k_\mathrm{B}Tg_\mathrm{m}\left[\frac{|y_\mathrm{n11e}|^2}{|y_\mathrm{n21e}|^2} + 2\,\mathrm{Re}\left(\frac{y_\mathrm{n11e}}{y_\mathrm{n21e}}\right)\right] \tag{5.97}$$

for the spectral density of the input noise current generator. The correlation

[24] Measurement of noise figures requires careful de-embedding to separate the noise characteristics of the device from the effects of pads and interconnect lines [37].

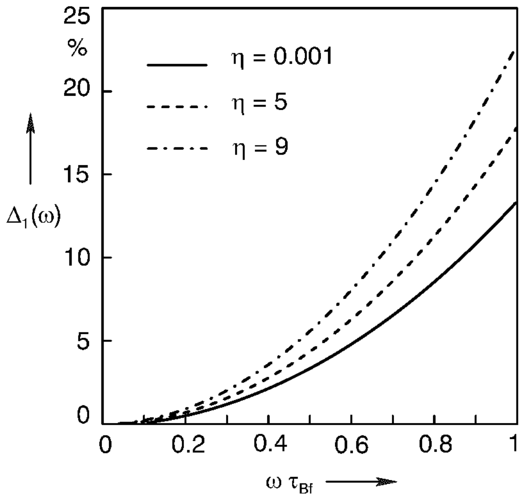


Fig. 5.12. Correction term $\Delta_1(\omega)$ vs. $\omega\tau_{\mathrm{Bf}}$ for different values of η

between the equivalent input noise current sources is

$$\underline{\gamma}_{iv}\sqrt{S_i S_v} = \frac{\langle \underline{I}_{\mathrm{n}} \underline{V}_{\mathrm{n}}^* \rangle}{\Delta f} = 2k_{\mathrm{B}}T\,\frac{g_{\mathrm{m}}\,(y_{\mathrm{n21e}} + y_{\mathrm{n11e}}) - |y_{\mathrm{n21e}}|^2}{|y_{\mathrm{n21e}}|^2}\,. \tag{5.98}$$

If these expressions are introduced into the formula for the noise factor, we obtain

$$\begin{aligned} F &= 1 + \frac{g_{\mathrm{m}}}{2\,\mathrm{Re}(Z_{\mathrm{S}})}\left\{|Z_{\mathrm{S}}|^2\left[\frac{|y_{\mathrm{n11e}}|^2}{|y_{\mathrm{n21e}}|^2} + 2\,\mathrm{Re}\left(\frac{y_{\mathrm{n11e}}}{y_{\mathrm{n21e}}}\right)\right]\right. \\ &\quad \left. + \frac{1}{|y_{\mathrm{n21e}}|^2} + 2\,\mathrm{Re}\left(Z_{\mathrm{S}}\,\frac{(y_{\mathrm{n21e}}+y_{\mathrm{n11e}}) - |y_{\mathrm{n21e}}|^2/g_{\mathrm{m}}}{|y_{\mathrm{n21e}}|^2}\right)\right\}. \end{aligned}$$

In the case of a real source impedance, this result simplifies to

$$F = 1 + \Delta_1(\omega) + \frac{g_{\mathrm{m}}R_{\mathrm{S}}}{2} f_{\mathrm{n1}}(\omega) + \frac{g_{\mathrm{m}}}{2R_{\mathrm{S}}|y_{\mathrm{n21e}}|^2}\,,$$

where

$$f_{\mathrm{n1}}(\omega) = \frac{|y_{\mathrm{n11e}}|^2}{|y_{\mathrm{n21e}}|^2} + 2\,\mathrm{Re}\left(\frac{y_{\mathrm{n11e}}}{y_{\mathrm{n21e}}}\right)$$

and

$$\Delta_1(\omega) = \mathrm{Re}\left(\frac{g_{\mathrm{m}}(y_{\mathrm{n21e}}+y_{\mathrm{n11e}})}{|y_{\mathrm{n21e}}|^2}\right) - 1$$

(Fig. 5.12). In the case of the drift transistor, y_{n21e} is given by (5.94), and

$$\frac{y_{\mathrm{n11e}}}{y_{\mathrm{n21e}}} = \mathrm{e}^{-\eta/2}\left(\cosh(\underline{\vartheta}) + \frac{\eta/2}{\underline{\vartheta}}\sinh(\underline{\vartheta})\right) - 1\,, \tag{5.99}$$

resulting in

$$f_{\mathrm{n1}}(\omega) \;=\; \frac{\alpha\,(\mathrm{e}^{\eta\alpha}-1)+\beta\,[\sin(\eta\beta)+\cos(\eta\beta)-1]}{\sqrt{1+(\omega\tilde{\tau})^2}}\,\mathrm{e}^{-\eta}+\mathrm{e}^{-\eta}-1$$

and

$$\Delta_1(\omega) \;=\; \frac{\alpha\sinh(\eta\alpha)+\beta\sin(\eta\beta)+\cosh(\eta\alpha)-\cos(\eta\beta)}{2\,\sinh(\eta/2)\,\sqrt{1+(\omega\tilde{\tau})^2}}\,\mathrm{e}^{-\eta/2}-1\,.$$

This result has to be compared with the outcome of the quasi-static approach used in the conventional PSPICE noise model. There, small-signal transport of the electron current across the base layer is described in terms of an (electron) diffusion capacitance c_{tne} and a linearized transfer current source $g_{\mathrm{m}}\underline{v}_\pi$, as shown in Fig. 5.13.

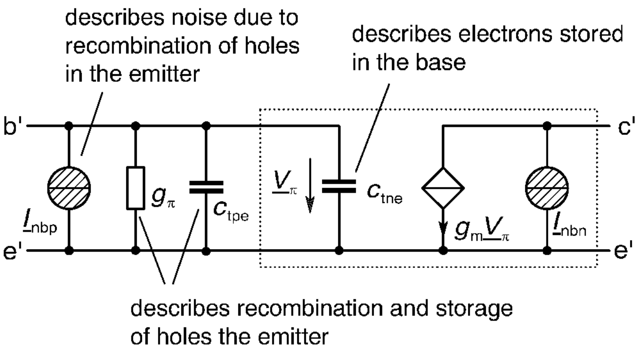


Fig. 5.13. Quasi-static small-signal equivalent circuit for the description of the electron current injected into and traversing the base layer

In this approximation, $y_{\mathrm{n11e}} = \mathrm{j}\omega c_{\mathrm{tne}}$ and $y_{\mathrm{n21e}} = g_{\mathrm{m}}$, while $\underline{\gamma}_{iv}$ is purely imaginary; this results in the approximations

$$f_{\mathrm{n1}} \;\approx\; \frac{\omega^2 c_{\mathrm{tne}}^2}{g_{\mathrm{m}}^2} \;\approx\; \omega^2\tau_{\mathrm{Bf}}^2\,, \quad \Delta_1 \;\approx\; 0 \quad \text{and} \quad f_{\mathrm{pb}} \;\approx\; 1\,,$$

corresponding to

$$F \;=\; 1+\frac{1}{2R_{\mathrm{S}}g_{\mathrm{m}}}+\frac{R_{\mathrm{S}}g_{\mathrm{m}}}{2}\,\frac{\omega^2 c_{\mathrm{tne}}^2}{g_{\mathrm{m}}^2}$$

for the noise factor. The deviation of the correct noise factor from the result obtained from the quasi-static approximation may then be represented as

$$\Delta F \;=\; \Delta_1+g_{\mathrm{m}}R_{\mathrm{S}}\Delta_2+\frac{1}{g_{\mathrm{m}}R_{\mathrm{S}}}\Delta_3\,,$$

where

$$\Delta_2 = \frac{\alpha\,(\mathrm{e}^{\eta\alpha}-1) + \beta\,[\,\sin(\eta\beta)+\cos(\eta\beta)-1\,]}{2\sqrt{1+(\omega\tilde{\tau})^2}}\,\mathrm{e}^{-\eta} + \frac{\mathrm{e}^{-\eta}-1}{2} - \frac{\omega^2\tau_{\mathrm{Bf}}^2}{2}\,,$$

$$\Delta_3 = \frac{\cosh(\eta\alpha)-\cos(\eta\beta)}{4\,\sinh^2(\eta/2)\sqrt{1+(\omega\tilde{\tau})^2}} - \frac{1}{2}\,.$$

These functions have been computed numerically as a function of frequency for different values of η; graphical representations of the results as a function of $\omega\tau_{\mathrm{Bf}}$ are given in Figs. 5.12 and 5.14.

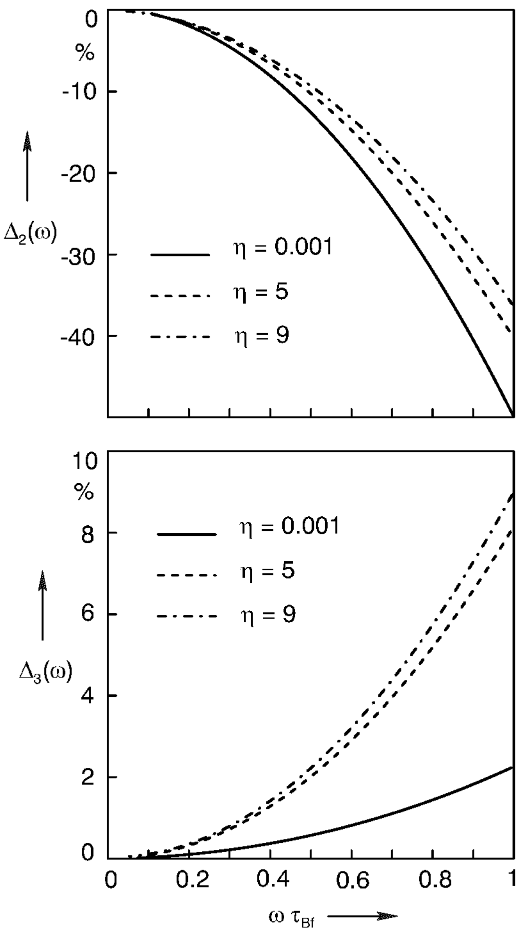


Fig. 5.14. Correction terms $\Delta_2(\omega)$ and $\Delta_3(\omega)$ vs. $\omega\tau_{\mathrm{Bf}}$ for different values of η

5.6.2 Noise Figure

A general expression for the noise figure in the absence of carrier multiplication effects can be derived from the noise equivalent circuit shown in Fig. 5.10. The transfer factors $\underline{H}_{\mathrm{rs}} = \underline{I}_{\mathrm{c}}/\underline{V}_{\mathrm{rs}}$, $\underline{H}_{\mathrm{ib}} = \underline{I}_{\mathrm{c}}/\underline{I}_{\mathrm{nb}}$, $\underline{H}_{\mathrm{ic}} = \underline{I}_{\mathrm{c}}/\underline{I}_{\mathrm{nc}}$, $\underline{H}_{\mathrm{re}} = \underline{I}_{\mathrm{c}}/\underline{V}_{\mathrm{re}}$ and $\underline{H}_{\mathrm{rc}} = \underline{I}_{\mathrm{c}}/\underline{V}_{\mathrm{rc}}$ allow us to compute the spectral density of the output noise current in the absence of transistor noise sources,

$$\langle \underline{I}_{\mathrm{c}}\underline{I}_{\mathrm{c}}^{*}\rangle = |\underline{H}_{\mathrm{rs}}|^2 \langle \underline{V}_{\mathrm{rs}}\underline{V}_{\mathrm{rs}}^{*}\rangle ,$$

where

$$\langle \underline{V}_{\mathrm{rs}}\underline{V}_{\mathrm{rs}}^{*}\rangle = 4k_{\mathrm{B}}TR_{\mathrm{S}}\Delta f .$$

We can also compute the spectral density of the output noise current with all noise sources active

$$\begin{aligned}\langle \underline{I}_{\mathrm{c}}\underline{I}_{\mathrm{c}}^{*}\rangle = {} & |\underline{H}_{\mathrm{rs}}|^2 \langle \underline{V}_{\mathrm{rs}}\underline{V}_{\mathrm{rs}}^{*}\rangle + |\underline{H}_{\mathrm{rs}}|^2 \langle \underline{V}_{\mathrm{rb}}\underline{V}_{\mathrm{rb}}^{*}\rangle + |\underline{H}_{\mathrm{re}}|^2 \langle \underline{V}_{\mathrm{re}}\underline{V}_{\mathrm{re}}^{*}\rangle \\ & + |\underline{H}_{\mathrm{rc}}|^2 \langle \underline{V}_{\mathrm{rc}}\underline{V}_{\mathrm{rc}}^{*}\rangle + |\underline{H}_{\mathrm{ib}}|^2 \langle \underline{I}_{\mathrm{nb}}\underline{I}_{\mathrm{nb}}^{*}\rangle + |\underline{H}_{\mathrm{ic}}|^2 \langle \underline{I}_{\mathrm{nc}}\underline{I}_{\mathrm{nc}}^{*}\rangle \\ & + 2\,\mathrm{Re}\left(\underline{H}_{\mathrm{ic}}\underline{H}_{\mathrm{ib}}^{*}\langle \underline{I}_{\mathrm{nc}}\underline{I}_{\mathrm{nb}}^{*}\rangle\right) ,\end{aligned}$$

using the spectral densities given in (5.83) – (5.88). This yields the following for the noise factor:

$$\begin{aligned}F = {} & 1 + \frac{r_{\mathrm{bb'}}}{R_{\mathrm{S}}} + \frac{r_{\mathrm{ee'}}}{R_{\mathrm{S}}}\frac{|\underline{H}_{\mathrm{re}}|^2}{|\underline{H}_{\mathrm{rs}}|^2} + \frac{r_{\mathrm{cc'}}}{R_{\mathrm{S}}}\frac{|\underline{H}_{\mathrm{rc}}|^2}{|\underline{H}_{\mathrm{rs}}|^2} + \frac{g_{\mathrm{m}}}{2R_{\mathrm{S}}}\frac{|\underline{H}_{\mathrm{ic}}|^2}{|\underline{H}_{\mathrm{rs}}|^2} \\ & + \left(\frac{\mathrm{Re}(y_{\mathrm{n11c}})}{R_{\mathrm{S}}} + \frac{g_{\mathrm{m}}}{2R_{\mathrm{S}}}\frac{f_{\mathrm{pb}}(\omega)}{B_{\mathrm{N}}}\right)\frac{|\underline{H}_{\mathrm{ib}}|^2}{|\underline{H}_{\mathrm{rs}}|^2} + 2\,\mathrm{Re}\left(\frac{\underline{H}_{\mathrm{ic}}\underline{H}_{\mathrm{ib}}^{*}}{|\underline{H}_{\mathrm{rs}}|^2}\frac{y_{\mathrm{n21c}} - g_{\mathrm{m}}}{2R_{\mathrm{S}}}\right) ,\end{aligned}$$

where $R_{\mathrm{S}} = \mathrm{Re}\,Z_{\mathrm{S}}$. The transfer factors are

$$\frac{\underline{H}_{\mathrm{rc}}}{\underline{H}_{\mathrm{rs}}} = \frac{y'_{22} + (Z_{\mathrm{S}} + r_{\mathrm{bb'}} + r_{\mathrm{ee'}})\Delta'_y}{y'_{21} - r_{\mathrm{cc'}}\Delta'_y} ,$$

$$\frac{\underline{H}_{\mathrm{re}}}{\underline{H}_{\mathrm{rs}}} = -\frac{y'_{21} + y'_{22} + (Z_{\mathrm{S}} + r_{\mathrm{bb'}})\Delta'_y}{y'_{21} - r_{\mathrm{ee'}}\Delta'_y} ,$$

$$\frac{\underline{H}_{\mathrm{ib}}}{\underline{H}_{\mathrm{rs}}} = -\frac{(Z_{\mathrm{S}} + r_{\mathrm{bb'}} + r_{\mathrm{ee'}})\,y'_{21} + r_{\mathrm{ee'}}y'_{22}}{y'_{21} - r_{\mathrm{ee'}}\Delta'_y} ,$$

$$\frac{\underline{H}_{\mathrm{ic}}}{\underline{H}_{\mathrm{rs}}} = \frac{1 + (Z_{\mathrm{S}} + r_{\mathrm{bb'}} + r_{\mathrm{cc'}})y'_{11} + r_{\mathrm{cc'}}y'_{12}}{y'_{21} - r_{\mathrm{ee'}}\Delta'_y} ,$$

where $\Delta'_y = y'_{11}y'_{22} - y'_{12}y'_{21}$. This formula allows to compute the optimum source resistance Z_{Sopt} that minimizes the noise factor. The general result for the minimum noise figure can then be computed using Z_{Sopt}. Various expressions for the minimum noise figure of a bipolar transistor have been presented in the literature [38–43]; these generally can be considered as approximations to the result of the general approach presented here [44].

Series resistances and cutoff frequencies are layout-dependent, and so is the minimum noise figure that may be achieved with a given bipolar technology. It is therefore of interest to consider the effects of scaling on the minimum noise figure [43,45]. Heterojunction bipolar transistors with a wide-gap emitter allow one to reduce the base series resistance without compromising the current gain. As the base resistance has a substantial effect on the high-frequency noise of a bipolar transistor, HBTs have smaller noise figures than comparable BJTs.[25]

5.6.3 Effects of Carrier Multiplication on Noise Figure

To calculate the noise figure of a bipolar transistor, it is common practice to include the noise sources in a separate noise two-port connected to the input of an otherwise noise-free two-port that represents the small-signal behavior of the transistor (see Sect. 1.8.3). However, in the avalanche regime, where $V_{\mathrm{CE}} \geq BV_{\mathrm{CEO}}$, this procedure does not work, as $h_{21\mathrm{e}}$ becomes singular as $V_{\mathrm{CE}} \to BV_{\mathrm{CEO}}$. To determine the noise figure of a bipolar transistor under such circumstances, the circuit depicted in Fig. 5.15 may be considered, where non-quasi-static effects, the collector series resistance $r_{\mathrm{cc'}}$, the output conductance g_{o} and the conductance g_μ have been neglected for simplicity. For this circuit, we determine the spectral density of the output noise current

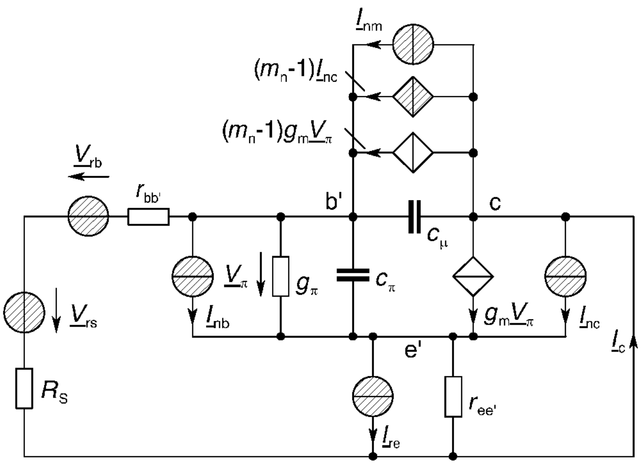


Fig. 5.15. Noise equivalent circuit for calculation of the noise figure of a bipolar transistor operated in the avalanche regime

in the case when only the noise source $\underline{V}_{\mathrm{rS}}$ is active and in the case when

[25] A computational study of base-profile optimization to obtain the minimum noise figure in a SiGe HBT based on the noise model of Hawkins [40], has been presented in [46].

all noise sources are active, leading to the values $S_{i\mathrm{a1}}(f)$ and $S_{i\mathrm{a2}}(f)$, respectively. The ratio $S_{i\mathrm{a2}}/S_{i\mathrm{a1}}$ defines the noise factor, from which the noise figure is determined. We obtain

$$S_{i\mathrm{a1}} = 4k_{\mathrm{B}}TR_{\mathrm{S}}|\underline{H}_{\mathrm{vrs}}|^2 \,, \tag{5.100}$$

where $\underline{H}_{\mathrm{vrs}} = \underline{I}_{\mathrm{c}}/\underline{V}_{\mathrm{rs}}$, and

$$\begin{aligned} S_{i\mathrm{a2}} &= \frac{1}{\Delta f}\Big\langle|\underline{H}_{\mathrm{vrs}}\,\underline{V}_{\mathrm{rs}} + \underline{H}_{\mathrm{vrb}}\,\underline{V}_{\mathrm{rb}} + \underline{H}_{\mathrm{ib}}\,\underline{I}_{\mathrm{nb}} \\ &\quad +\underline{H}_{\mathrm{im}}\,[\underline{I}_{\mathrm{nm}} + (m_{\mathrm{n}}-1)\underline{I}_{\mathrm{nc}}\,] + \underline{H}_{\mathrm{ic}}\,\underline{I}_{\mathrm{nc}} + \underline{H}_{\mathrm{ire}}\,\underline{I}_{\mathrm{re}}|^2\Big\rangle \,, \end{aligned} \tag{5.101}$$

where the transfer factors are $\underline{H}_{\mathrm{vrb}} = \underline{I}_{\mathrm{c}}/\underline{V}_{\mathrm{rb}}$, $\underline{H}_{\mathrm{ib}} = \underline{I}_{\mathrm{c}}/\underline{I}_{\mathrm{nb}}$, $\underline{H}_{\mathrm{im}} = \underline{I}_{\mathrm{c}}/\underline{I}_{\mathrm{nm}}$, $\underline{H}_{\mathrm{ic}} = \underline{I}_{\mathrm{c}}/\underline{I}_{\mathrm{nc}}$ and $\underline{H}_{\mathrm{ire}} = \underline{I}_{\mathrm{c}}/\underline{I}_{\mathrm{re}}$. In calculating $S_{i\mathrm{a2}}$, correlations between different noise current sources may be neglected, resulting in

$$\begin{aligned} S_{i\mathrm{a2}}\Delta f &= |\underline{H}_{\mathrm{vrs}}|^2\langle\underline{V}_{\mathrm{rs}}\underline{V}_{\mathrm{rs}}^*\rangle + |\underline{H}_{\mathrm{vrb}}|^2\langle\underline{V}_{\mathrm{rb}}\underline{V}_{\mathrm{rb}}^*\rangle + |\underline{H}_{\mathrm{ib}}|^2\langle\underline{I}_{\mathrm{nb}}\underline{I}_{\mathrm{nb}}^*\rangle \\ &\quad +|\underline{H}_{\mathrm{im}}|^2\langle\underline{I}_{\mathrm{nm}}\underline{I}_{\mathrm{nm}}^*\rangle + |\underline{H}_{\mathrm{ic}}+(m_{\mathrm{n}}-1)\underline{H}_{\mathrm{im}}|^2\langle\underline{I}_{\mathrm{nc}}\underline{I}_{\mathrm{nc}}^*\rangle + |\underline{H}_{\mathrm{ire}}|^2\langle\underline{I}_{\mathrm{re}}\underline{I}_{\mathrm{re}}^*\rangle \,. \end{aligned}$$

This expression is increased by

$$\begin{aligned} &\Delta S_{i\mathrm{a2}}\Delta f = \\ &\quad |\underline{H}_{\mathrm{im}}|^2\langle\underline{I}_{\mathrm{nm}}\underline{I}_{\mathrm{nm}}^*\rangle + \Big[|\underline{H}_{\mathrm{ic}} + (m_{\mathrm{n}}-1)\underline{H}_{\mathrm{im}}|^2 - |\underline{H}_{\mathrm{ic}}|^2\Big]\langle\underline{I}_{\mathrm{nc}}\underline{I}_{\mathrm{nc}}^*\rangle \end{aligned}$$

with respect to the result that would be obtained if $m_{\mathrm{n}} = 1$. Making use of

$$\langle\underline{I}_{\mathrm{nc}}\underline{I}_{\mathrm{nc}}^*\rangle = 2eI_{\mathrm{T}}\Delta f \quad \text{and} \quad \langle\underline{I}_{\mathrm{nm}}\underline{I}_{\mathrm{nm}}^*\rangle = 2eI_{\mathrm{T}}m_{\mathrm{n}}^2\,[\phi(m_{\mathrm{n}})-1]\,\Delta f$$

yields the following for the increase of the noise figure due to carrier multiplication effects:

$$\Delta F = \frac{g_{\mathrm{m}}}{2R_{\mathrm{S}}}\left(2(m_{\mathrm{n}}-1)\frac{\mathrm{Re}(\underline{H}_{\mathrm{ic}}\underline{H}_{\mathrm{im}}^*)}{|\underline{H}_{\mathrm{vrs}}|^2} + \Big[m_{\mathrm{n}}^2\phi(m_{\mathrm{n}}) - 2m_{\mathrm{n}}+1\Big]\frac{|\underline{H}_{\mathrm{im}}|^2}{|\underline{H}_{\mathrm{vrs}}|^2}\right) \,;$$

this term reduces to zero as $m_{\mathrm{n}} \to 1$. To illustrate this effect, we consider the special situation $V_{\mathrm{CE}} = BV_{\mathrm{CEO}}$, where $(m_{\mathrm{n}}-1)\beta = 1$. Making use of the approximation (5.17), $\phi(m_{\mathrm{n}}) \approx 2 - 1/m_{\mathrm{n}}$, we may write

$$m_{\mathrm{n}}^2\phi(m_{\mathrm{n}}) \approx 1 + \frac{3}{\beta} \,,$$

and therefore

$$m_{\mathrm{n}}^2\phi(m_{\mathrm{n}}) - 2m_{\mathrm{n}} + 1 \approx 1/\beta$$

if terms of order $1/\beta^2$ are neglected. Since $g_{\mathrm{m}}/\beta = g_\pi$, the excess noise due to carrier multiplication increases the noise factor by

$$\Delta F(BV_{\mathrm{CEO}}) \approx \frac{g_\pi}{2R_{\mathrm{S}}} \frac{2\,\mathrm{Re}(\underline{H}_{\mathrm{ic}}\underline{H}_{\mathrm{im}}^*) + |\underline{H}_{\mathrm{im}}|^2}{|\underline{H}_{\mathrm{vrs}}|^2}$$

at $V_{\mathrm{CE}} = BV_{\mathrm{CEO}}$. As long as device capacitances and delay times can be neglected, the transfer factors are real, and

$$\frac{H_{\mathrm{im}}^2}{H_{\mathrm{vrs}}^2} = r_{\mathrm{b}}^2 \left(1+\frac{\Xi}{\beta m_{\mathrm{n}}}\right)^2 ,$$

$$\frac{H_{\mathrm{im}}H_{\mathrm{ic}}}{H_{\mathrm{vrs}}^2} = r_{\mathrm{b}}^2 \left(1+\frac{\Xi}{\beta m_{\mathrm{n}}}\right)\left(\frac{\Xi}{\beta m_{\mathrm{n}}} - \frac{r_{\mathrm{ee'}}}{r_{\mathrm{b}}}\right) ,$$

where $r_{\mathrm{b}} = R_{\mathrm{S}} + r_{\mathrm{bb'}}$ and

$$\Xi = 1 - \beta(m_{\mathrm{n}} - 1) + \frac{r_\pi + r_{\mathrm{ee'}}(\beta+1)}{r_{\mathrm{b}}} .$$

At $V_{\mathrm{CE}} = BV_{\mathrm{CEO}}$, this expression simplifies to

$$\Xi = \frac{r_\pi + r_{\mathrm{ee'}}(\beta+1)}{r_{\mathrm{b}}} .$$

From these expressions, we obtain

$$\begin{aligned}\Delta F(BV_{\mathrm{CEO}}) &\approx \frac{g_\pi r_{\mathrm{b}}^2}{2R_{\mathrm{S}}}\left(1+\frac{\Xi}{\beta+1}\right)\left(1+\frac{3\Xi}{\beta+1} - \frac{2r_{\mathrm{ee'}}}{r_{\mathrm{b}}}\right)\\ &= \frac{g_\pi r_{\mathrm{b}}^2}{2R_{\mathrm{S}}}\left(1+\frac{r_{\mathrm{ee'}}}{r_{\mathrm{b}}} + \frac{r_\pi}{(\beta+1)r_{\mathrm{b}}}\right)\left(1+\frac{3r_\pi}{(\beta+1)r_{\mathrm{b}}} + \frac{r_{\mathrm{ee'}}}{r_{\mathrm{b}}}\right)\end{aligned}$$

at $V_{\mathrm{CE}} = BV_{\mathrm{CEO}}$, using $\beta m_{\mathrm{n}} = \beta+1$. The increase of the noise factor depends

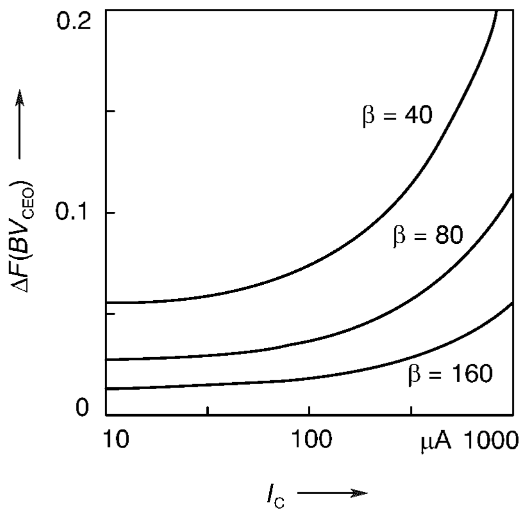


Fig. 5.16. Increase of noise factor due to carrier multiplication at $V_{\mathrm{CE}} = BV_{\mathrm{CEO}}$ as a function of the transfer current at the bias point

on both the source resistance and the bias point; Fig. 5.16 shows the dependence of ΔF on I_C if R_S is chosen as the optimum source resistance in the absence of carrier multiplication effects, i.e.

$$R_S = \frac{\sqrt{\beta + 2\beta g_m(r_{bb'}+r_{ee'}) + g_m^2(r_{bb'}+r_{ee'})^2}}{g_m} \tag{5.102}$$

for three different values of the current gain, assuming $B_N = \beta$ for simplicity.[26] The excess noise due to carrier multiplication effects at $V_{CE} = BV_{CEO}$ decreases with current gain, as can easily be understood from the breakdown condition $\beta(m_n - 1) = 1$, which implies that at larger current gain, less carrier multiplication (and hence less noise) is required for open-base breakdown. A plot of the excess noise associated with carrier multiplication versus the source impedance R_S shows minima at R_S values that lie considerably below the value given by (5.102).

5.7 Low-Frequency Noise

Low-frequency noise in bipolar transistors is generally characterized by a spectral density that varies in proportion to $1/f$ ($1/f$ noise). Low-frequency noise in conjunction with device nonlinearities may cause phase noise in oscillators and is therefore also of great importance in high-frequency applications: Oscillators constructed from transistors with low $1/f$ noise may yield a sufficiently stable oscillator frequency without relying on an external dielectric resonator, and thus allow a substantial decrease in the size of microwave and RF circuitry.

For the characterization of low-frequency noise, the corner frequencies of the input-referred noise sources are usually specified. Since these are bias-dependent quantities, the values are only meaningful if the measurement conditions are known. Modeling $1/f$ noise in terms of a base current noise generator with spectral density $S_i = (I_B/\mathrm{A})^{A_F} K_F/f$ in addition to the base current shot noise with spectral density $S_i = 2eI_B$ yields the corner frequency of the input-referred noise current (see Sect. 1.8)

$$f_{ci} = \frac{K_F}{2eI_B}\left(\frac{I_B}{\mathrm{A}}\right)^{A_F} . \tag{5.103}$$

Determination of f_{ci} for different values of I_B allows one to determine A_F and K_F from a double-logarithmic plot of f_{ci}/Hz versus I_B/A; this should give a straight line [47] with slope $A_F - 1$ since

[26] The fact that ΔF is small in comparison with one does not necessarily lead to the conclusion that the effect is negligible, as an increase of the noise factor by 0.1 results in an increase of the noise figure by approximately 0.4 dB.

$$\log_{10}\left(\frac{f_{\mathrm{ci}}}{\mathrm{Hz}}\right) = \log_{10}\left(\frac{K_{\mathrm{F}}/2e}{\mathrm{A\,Hz}}\right) + (A_{\mathrm{F}}-1)\log_{10}\left(\frac{I_{\mathrm{B}}}{\mathrm{A}}\right) ,$$

according to (5.103). In polysilicon-emitter BJTs, $1 < A_{\mathrm{F}} < 2$ is observed – the corner frequency of the input-referred noise current source thus increases with the bias current I_{B}. For SiGe HBTs with a base current of several microamperes, values of the noise current corner frequency of the order of 1 kHz and values of the noise voltage corner frequency well below 1 kHz have been reported by several authors [48–51].

The $1/f$ noise in high-frequency BJTs and HBTs may be due to series resistances (contacts are usually formed using polycrystalline silicon, which is known to cause $1/f$ noise) and surface effects (a shallow emitter with a contact of finite surface recombination velocity may also cause an increased level of $1/f$ noise). For the optimization of BJTs and HBTs with respect to their $1/f$-noise properties, the mechanisms that lead to $1/f$ noise in bipolar transistors must be understood; therefore measurement techniques that allow one to determine the location of $1/f$ noise sources are required.

For the measurement of $1/f$ noise, three resistors R_{E}, R_{B} and R_{C} are usually placed in series with the device terminals, as shown in Fig. 5.17. These resistors are wire-wound or metal film resistors, in order not to introduce additional $1/f$-noise sources; biasing is performed with batteries [52]. Measuring

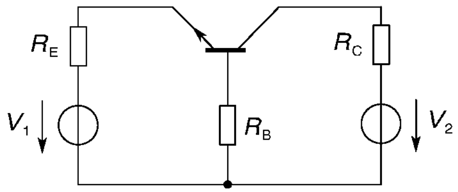


Fig. 5.17. General circuit for low-frequency noise measurements (after [52])

the noise voltages across the series resistors for different resistance values and currents allows the location of the $1/f$-noise sources with the help of the small-signal noise equivalent circuit shown in Fig. 5.18 [52]. If the noise current sources are assumed to be uncorrelated, the spectral densities of the noise voltages across the external resistors R_{E}, R_{C} and R_{B} are related to the spectral densities of the transistor noise sources by the relations [52]

$$\begin{aligned}\frac{Z^2}{R_{\mathrm{E}}^2}S_{v\mathrm{e}} &= \left[r_\pi - \beta(r_{\mathrm{bb'}}+R_{\mathrm{B}})\right]^2 S_{i\mathrm{b}} + \left(r_\pi+r_{\mathrm{bb'}}+R_{\mathrm{B}}\right)^2 S_{i\mathrm{c}} \\ &\quad +4k_{\mathrm{B}}T\Big[(\beta+1)^2(R_{\mathrm{B}}+R_{\mathrm{E}}) - 2(\beta+1)Z + Z^2/R_{\mathrm{E}}\Big] \\ &\quad +(\beta+1)^2 S_{v\mathrm{r}}\end{aligned}$$

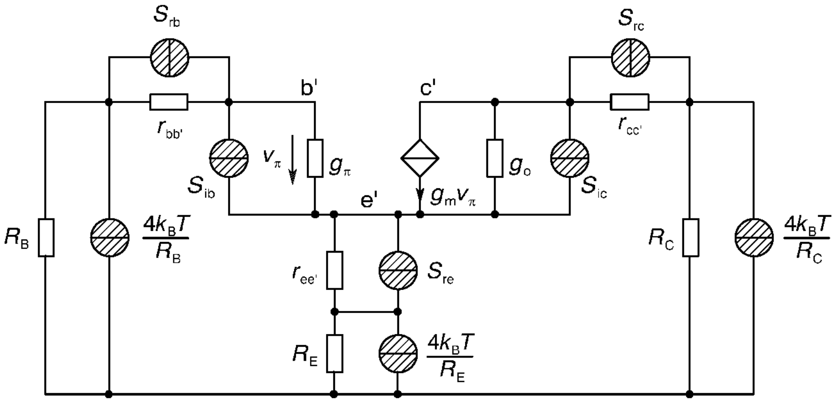


Fig. 5.18. Small-signal noise equivalent circuit of a BJT, with metal film terminal series resistances R_{B}, R_{E} and R_{C} with negligible $1/f$ noise

$$
\begin{aligned}
\frac{Z^2}{R_{\mathrm{C}}^2} S_{v\mathrm{c}} \;=\; & \beta^2 \left(r_{\mathrm{bb'}}+R_{\mathrm{B}}+r_{\mathrm{ee'}}+R_{\mathrm{E}}\right)^2 S_{i\mathrm{b}} \\
& + \left(r_\pi+r_{\mathrm{bb'}}+R_{\mathrm{B}}+r_{\mathrm{ee'}}+R_{\mathrm{E}}\right)^2 S_{i\mathrm{c}} \\
& +4k_{\mathrm{B}}T\Big[\beta^2(R_{\mathrm{B}}+R_{\mathrm{E}}) + Z^2/R_{\mathrm{C}}\Big] + \beta^2 S_{v\mathrm{r}}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{Z^2}{R_{\mathrm{B}}^2} S_{\mathrm{vb}} \;=\; & \left[r_\pi + \beta(r_{\mathrm{ee'}}+R_{\mathrm{E}})\right]^2 S_{i\mathrm{b}} + (r_{\mathrm{ee'}}+R_{\mathrm{E}})^2 S_{i\mathrm{c}} \\
& +4k_{\mathrm{B}}T\Big(R_{\mathrm{B}}+R_{\mathrm{E}}) - 2Z + Z^2/R_{\mathrm{B}}\Big] + S_{v\mathrm{r}} \;,
\end{aligned}
$$

where

$$
S_{v\mathrm{r}} = I_{\mathrm{B}}^2 S_{\mathrm{rb}} + I_{\mathrm{E}}^2 S_{\mathrm{re}} + 4k_{\mathrm{B}}T(r_{\mathrm{bb'}}+r_{\mathrm{ee'}}) \tag{5.104}
$$

and

$$
Z = r_\pi + r_{\mathrm{bb'}} + R_{\mathrm{B}} + (\beta+1)(r_{\mathrm{ee'}}+R_{\mathrm{E}}) \;. \tag{5.105}
$$

Low-frequency $1/f$ noise in high-frequency Si BJTs and SiGe HBTs has been investigated in [49, 51, 54–59]. The dominant noise source of the $1/f$-noise component was found to be situated in the emitter junction at the emitter side. Both fluctuations of the emitter resistance and fluctuations of the surface recombination velocity at the contact are mechanisms that result in such noise. In devices with a polysilicon emitter, the interface treatment was found to have a substantial effect on the observed $1/f$ noise: increasing the barrier at the interface and thus decreasing the effective surface recombination velocity results in an increase in the $1/f$ noise, as the interface affects both S_{re} and $S_{i\mathrm{b}}$. A model of $1/f$ noise that considers the effect on the noise sources S_{re} and $S_{i\mathrm{b}}$ of barrier height fluctuations of the oxide layer and the resulting

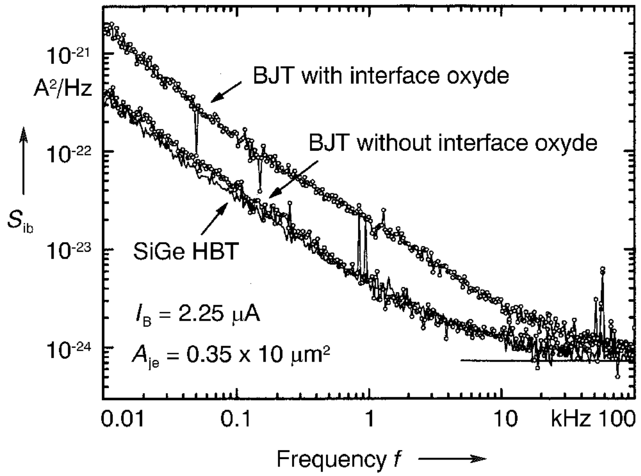


Fig. 5.19. Spectral density S_{ib} of base current noise for BJTs with and without an interfacial oxide layer in comparison with a SiGe graded-base HBT (after [53])

transparency fluctuations at the contact, for both minority and majority carriers, is presented in [54]. Oxide breakup due to a rapid thermal annealing (RTA) sequence was found to result in a decrease of low-frequency noise; this is also demonstrated by the measurement results shown in Fig. 5.19. A reduction of the $1/f$ noise associated with the polysilicon contact by implantation of F atoms was reported in [60].

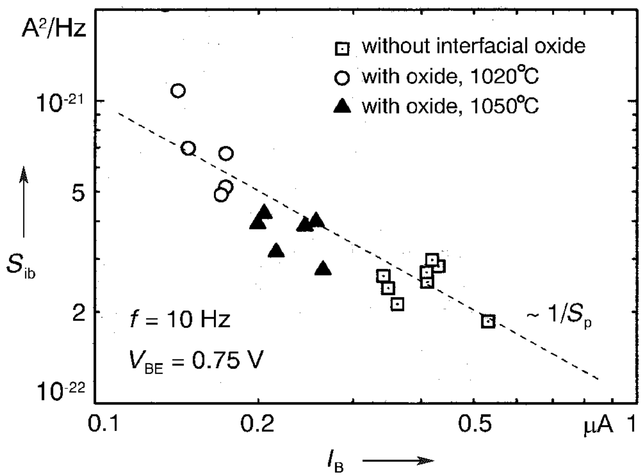


Fig. 5.20. Spectral density S_{ib} of base current noise, measured at a fixed value of V_{BE}, versus I_B for Si BJTs with and without interfacial oxide and with different temperatures used for the emitter drive-in (after [53])

The effect of the interface on $1/f$ noise in high-frequency BJTs with a polysilicon emitter contact can also be seen from Fig. 5.20, where the spectral density of the $1/f$-noise current is plotted versus the base current for a fixed value of V_{BE} and of frequency. If V_{BE} is kept fixed, the base current will vary approximately in proportion to the surface recombination velocity ($I_B \sim S_p$); larger values of the base current thus correspond to a less pronounced barrier at the interface, with a reduced effect on the $1/f$ noise produced there.

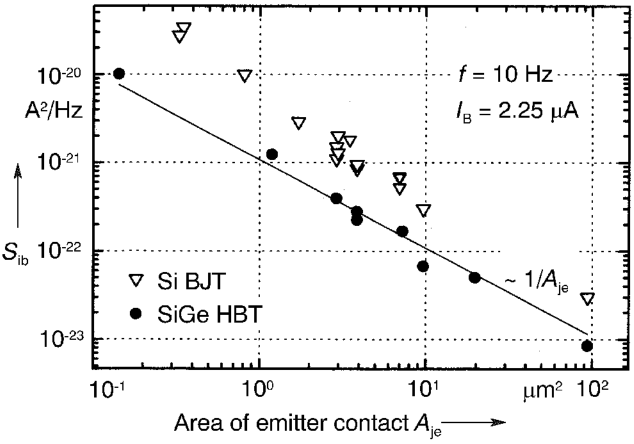


Fig. 5.21. Spectral density S_{ib} of base current noise for Si BJTs and SiGe HBTs, at a fixed frequency, versus the area of the emitter contact (after [53])

The spectral density $S_{1/f}$ of the $1/f$-noise current was found to vary in proportion to I_B^2 (i.e. $A_F \approx 2$) and in inverse proportion to the emitter area $1/A_{je}$ (i.e. $K_F \sim 1/A_{je}$), as demonstrated by the measured data in Fig. 5.21. This behavior led to the conclusion that the $1/f$-noise sources are homogeneously distributed over the entire emitter area [55]. Therefore, for the purpose of comparison of various technologies, the quantity

$$B_{1/f} = K_F A_{je} = \frac{A_{je} f S_{1/f}}{(I_B/A)^2}$$

has been suggested as a figure of merit to characterize $1/f$ noise [48].

The effect of mobility fluctuations in crystalline silicon on the $1/f$ noise in pn junction diodes has been investigated in [61,62]. Low-frequency $1/f$ noise associated with bulk defect centers in n^+p silicon diodes was investigated in [63] using a gated diode. The noise was attributed to defect centers associated with precipitated oxygen/dislocation complexes in the bulk section of the space charge region at 0.43 eV below the conduction band.

Low-frequency noise in npn GaAs/AlGaAs HBTs has been investigated in [64], where $1/f$ noise of both the collector current, with $S_{ic}^{1/f} \sim I_C^{1.4}$, and the base current, with $S_{ib}^{1/f} \sim I_B^{1.5}$, has been observed, where S_{ic} is larger than S_{ib} at fixed emitter–base voltage. At higher forward currents the parasitic series resistances were found to play an important role, because of feedback effects as well as $1/f$ resistance fluctuations. Results on the low-frequency noise of InP/InGaAs HBTs have been reported in [65].

5.8 References

[1] W. Schottky. Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern. *Ann. Phys.*, 57:541–567, 1918.

[2] H.B. Callen, T.A. Welton. Irreversibility and generalized noise. *Phys. Rev.*, 83(1):34–40, 1951.

[3] J.-P. Nougier. Fluctuations and noise of hot carriers in semiconductor materials and devices. *IEEE Trans. Electron Devices*, 41(11):2034–2049, 1994.

[4] P.J. Price. *Fluctuations of hot electrons. in Fluctuation Phenomena in Solids*, R.E. Burgess (ed.) Academic Press, New York, 1965.

[5] A. Gisolf, R.J. Zijlstra. Scattering noise of hot holes in space-charge-limited current flow in p-type Si. *J. Appl. Phys.*, 47(6):2727–2734, 1976.

[6] K.M. van Vliet. Noise in semiconductors and photodetectors. *Proc. IRE*, 46:1004–1018, 1958.

[7] A.S. Tager. Current fluctuations in a semiconductor (dielectric) under the conditions of impact ionization and avalanche breakdown. *Sov. Phys. Solid State*, 6(8):1919–1925, 1965.

[8] R.J. McIntyre. Multiplication noise in uniform avalanche diodes. *IEEE Trans. Electron Devices*, 13(1):164–168, 1966.

[9] R.J. McIntyre. A new look at impact ionization – part I: a theory of gain, noise, breakdown, probability, and frequency response. *IEEE Trans. Electron Devices*, 46(8):1623–1631, 1999.

[10] R.J. McIntyre. A new look at impact ionization part II: gain and noise in short avalanche photodiodes. *IEEE Trans. Electron Devices*, 46(8):1632–1639, 1999.

[11] F.N. Hooge. 1/f noise sources. *IEEE Trans. Electron Devices*, 41(11):1926–1935, 1994.

[12] P.H. Handel. Fundamental quantum $1/f$ noise in semiconductor devices. *IEEE Trans. Electron Devices*, 41(11):2023–2033, 1994.

[13] P. Dutta, P.M. Horn. Low-frequency fluctuations in solids: $1/f$ noise. *Rev. Mod. Phys.*, 53(3):497–516, 1981.

[14] B. Pellegrini. On mobility-fluctuation origin of $1/f$ noise. *Solid-State Electron.*, 29(12):1279–1287, 1986.

[15] A. van der Ziel. Unified presentation of $1/f$ noise in electronic devices: fundamental $1/f$ noise sources. *Proc. IEEE*, 76(3):233–258, 1988.

[16] C.M. van Vliet. A survey of results and future prospects on quantum $1/f$ noise and $1/f$ noise in general. *Solid-State Electron.*, 34(1):1–21, 1991.

[17] R. Brederlow, W. Weber, C. Dahl, D. Schmitt-Landsiedel, R. Thewes. Low-frequency noise of integrated poly-silicon resistors. *IEEE Trans. Electron Devices*, 48(6):1180–1187, 2001.

[18] R.A. Schiebel. A model for $1/f$ noise in diffusion current based on surface recombination velocity fluctuations and insulator trapping. *IEEE Trans. Electron Devices*, 41(5):768–778, 1994.

[19] S.T. Hsu. Flicker noise in metal semiconductor Schottky barrier diodes due to multistep tunneling processes. *IEEE Trans. Electron Devices*, 18(10):882–887, 1971.

[20] C. Jungemann, B. Neinhüs, B. Meinerzhagen. Hierarchical 2-D DD and HD noise simulations of Si And SiGe devices - part I: Theory. *IEEE Trans. Electron Devices*, 49(7):1250–1257, 2002.

[21] C. Jungemann, B. Neinhüs, S. Decker, B. Meinerzhagen. Hierarchical 2-D DD and HD noise simulations of Si And SiGe devices - part I: Results. *IEEE Trans. Electron Devices*, 49(7):1258–1264, 2002.

[22] K.M. van Vliet. Noise and admittance of the generation–recombination current involving SRH centers in the space-charge region of junction devices. *IEEE Trans. Electron Devices*, 23(11):1236–1246, 1976.

[23] K.M. van Vliet. Noise sources in transport equations associated with ambipolar diffusion and Shockley–Read recombination. *Solid-State Electronics*, 13:649–657, 1970.

[24] K.M. van Vliet. General transport theory of noise in pn junction-like devices – I. Three-dimensional Green's function formulation. *Solid-State Electron.*, 15:1033–1053, 1972.

[25] F. Bonani, M.R. Pinto, R.K. Smith, G. Ghione. An efficient approach to noise analysis through multidimensional physics-based models. *IEEE Trans. Electron Devices*, 45(1):261–269, 1998.

[26] M.J. Buckingham, E.A. Faulkner. The theory of inherent noise in pn junction diodes and bipolar transistors. *Radio Electron. Engineer*, 44(3):125–140, 1974.

[27] M.J. Buckingham. *Noise in Electronic Devices and Systems*. Wiley, 1983.

[28] J.L. Wyatt Jr., G.J. Coram. Nonlinear device noise models: satisfying the thermodynamic requirements. *IEEE Trans. Electron Devices*, 46(1):184–193, 1999.

[29] A. van der Ziel, J.B. Anderson, A.N. Birbas, W.C. Chen, P. Fang, V.M. Hietala, C.S. Park, P.R. Pukite, M.F. Toups, X. Wu, J. Xu, A.C. Young. Shot noise in solid state diodes. *Solid-State Electron.*, 29(10):1069–1071, 1986.

[30] B. Pellegrini. Semiconductor noise. *Phys. Rev. B*, 38(12):8269–8278, 1988.

[31] B. Pellegrini. Noise of junction devices. *Phys. Rev. B*, 38(12):8279–8292, 1988.

[32] K.M. van Vliet, M.L. Tarng. General transport theory of noise in pn junction-like devices – I. Carrier correlations and fluctuations for high injection. *Solid-State Electron.*, 15:1055–1069, 1972.

[33] C.M. van Vliet. Macroscopic and microscopic methods for noise in devices. *IEEE Trans. Electron Devices*, 41(11):1902–1915, 1994.

[34] P.O. Lauritzen. Noise due to generation and recombination of carriers in pn junction transition regions. *IEEE Trans. Electron Devices*, 15(10):770–776, 1968.

[35] A. van der Ziel. Theory of shot noise in junction diodes and junction transistors. *Proc. IRE*, 43:1639–1646, 1955.

[36] W. Guggenbuehl, M.J.O. Strutt. Theory and experiments on shot noise in semiconductor junction diodes and transistors. *Proc. IRE*, 45:839–854, 1957.

[37] K. Aufinger, J. Böck. A straightforward noise de-embedding method and its application to high-speed silicon bipolar transistors. *Proc. ESSDERC*, pp. 957–960, 1996.

[38] H. Fukui. The noise performance of microwave transistors. *IEEE Trans. Electron Devices*, 13(3):329–341, 1966.

[39] S.D. Malaviya, A. van der Ziel. A simplified approach to noise in microwave transistors. *Solid-State Electron.*, 13:1511 1518, 1970.

[40] R.J. Hawkins. Limitations of Nielsen's and related noise equations applied to microwave bipolar transistors, and a new expression for the frequency and current dependent noise figure. *Solid-State Electron.*, 20:191–196, 1977.

[41] A. van der Ziel, G. Bosman. Accurate expression for the noise temperature of common emitter microwave transistors. *IEEE Trans. Electron Devices*, 31(9):1280 1283, 1984.

[42] L. Escotte, J.-P. Roux, R. Plana, J. Graffeuil, A. Gruhle. Noise modeling of microwave heterojunction bipolar transistors. *IEEE Trans. Electron Devices*, 42(5):883–889, 1995.

[43] S.P. Voinigescu, M.C. Maliepaard, J.L. Showell, G.E. Babcock, D. Marchesan, M. Schroter, P. Schvan, D.L. Harame. A scalable high-frequency noise model for bipolar transistors with application to optimal transistor sizing for low-noise amplifier design. *IEEE J. Solid-State Circuits*, 32(9):1430–1439, 1997.

[44] K. Aufinger, M. Reisch. RF noise models for bipolar transistors - a critical comparison. *Proc. IEEE BCTM*, pp. 110–113, 2001.

[45] L.C.N. de Vreede, H.C. de Graaff, G.A.M. Hurkx, J.L. Tauritz, R.G.F. Baets. A figure of merit for the high-frequency noise behavior of bipolar transistors. *IEEE J. Solid-State Circuits*, 29(10):1220–1226, 1994.

[46] W.E. Ansley, J.D. Cressler, D.M. Richey. Base-profile optimization for minimum noise figure in advances UHV/CVD SiGe HBT's. *IEEE Trans. Microwave Theory Tech.*, 46(5):653–660, 1998.

[47] J.C. Costa N.D. Ngo, R. Jackson, N. Camillieri, J. Jaffee. Extracting $1/f$ noise coefficients for BJT's. *IEEE Trans. Electron Devices*, 41(11):1992–1999, 1994.

[48] J.D. Cressler, L. Vempati, J.A. Babcock, R.C. Jaeger, D.L. Harame. Low-frequency noise characteristics of UHV/CVD epitaxial Si- and SiGe-base bipolar transistors. *IEEE Electron Device Lett.*, 17(1):13 15, 1996.

[49] R. Gabl, K. Aufinger, J. Böck, T.F. Meister. Low-frequency noise characteristics of advanced Si and SiGe bipolar transistors. *Proc. ESSDERC*, pp. 536–539, 1997.

[50] B. van Haaren, M. Regis, O. Llopis, L. Escotte, A. Gruhle, C. Mhner, R. Plana, J. Graffeuil. Low-frequency noise properties of SiGe HBT's and application to ultra-low phase-noise oscillators. *IEEE Trans. Microwave Theory Tech.*, 46(5):647–652, 1998.

[51] K. Aufinger, H. Knapp, R. Gabl, T.F. Meister, J. Böck, H. Schäfer, M. Pohl, L. Treitinger. Noise characteristics of 0.5 µm/50 GHz Si and 0.5 µm/70 GHz SiGe bipolar technologies. *European microwave conference*, pp. 647-652, 1999.

[52] T.G.M. Kleinpenning. Low-frequency noise in modern bipolar transistors: impact of intrinsic transistor and parasitic series resistances. *IEEE Trans. Electron Devices*, 41(11):1981–1991, 1994.

[53] R. Gabl. *Beiträge zur physikalischen Modellierung integrierter Silizium-Bipolartransistoren.* Dissertation, Universität Innsbruck, 1998.

[54] H.A.W. Markus, T.G.M. Kleinpenning. Low-frequency noise in polysilicon emitter bipolar transistors. *IEEE Trans. Electron Devices*, 42(4):720–727, 1995.

[55] L.S. Vempati, J.D. Cressler, J.A. Babcock, R.C. Jaeger, D.L. Harame. Low-frequency noise in UHV/CVD epitaxial Si and SiGe bipolar transistors. *IEEE J. Solid-State Circuits*, 31(10):1458–1467, 1996.

[56] E. Simoen, S. Decoutere, A. Cuthbertson, C.L. Claeys, L. Deferm. Impact of polysilicon emitter interfacial layer engineering on the $1/f$-noise of bipolar transistors. *IEEE Trans. Electron Devices*, 43(12):2261–2268, 1996.

[57] H.A.W. Markus, Ph. Roche, T.G.M. Kleinpenning. On the 1/f noise in polysilicon emitter bipolar transistors: coherence between base current noise and emitter series ressitance noise. *Solid-State Electron.*, 41(3):441–445, 1997.

[58] S.P.O. Bruce, L.K.J. Vandamme, A. Rydberg. Measurement of low-frequency base and collector current noise and coherence in SiGe heterojunction bipolar transistors using transimpedance amplifiers. *IEEE Trans. Electron Devices*, 46(5):993 1000, 1999.

[59] J.M. Routoure, J. Lepaisant, D. Bloyet, S. Bardy, C. Biard, L. Gambus, J. Lebailly. Low frequency excess noise measurements in high frequency polysilicon emitter bipolar transistors. *Solid-State Electron.*, 43:729–740, 1999.

[60] N. Lukyanchikova, N Garbar, M. Petrichuk, J.F.W. Schiz, P. Ashburn. The influence of BF_2 and F implants on the $1/f$ noise in SiGe HBTs with a self-aligned link base. *IEEE Trans. Electron Devices*, 48(12):2808–2815, 2001.

[61] T.G.M. Kleinpenning. $1/f$ noise in pn junction diodes. *J. Vac. Sci. Technol. A*, 3(1):176–182, 1985.

[62] L.K.J. Vandamme, Gy Trefan. Review of low-frequency noise in bipolar transistors over the last decade. *Proc. IEEE BCTM*, pp. 68 73, 2001.

[63] F.-C. Hou, G. Bosman, E. Simoen, J. Vanhellemont, C. Claeys. Bulk defect induced low-frequency noise in n^+p silicon diodes. *IEEE Trans. Electron Devices*, 45(12):2528–2536, 1998.

[64] T.G.M. Kleinpenning, A.J. Holden. $1/f$ noise in npn GaAs/AlGaAs heterojunction bipolar transistors: impact of intrinsic transistor and parasitic series resistances. *IEEE Trans. Electron Devices*, 40(6):1148–1153, 1993.

[65] Y. Takanashi, H. Fukano. Low-frequency noise of InP/InGaAs heterojunction bipolar transistors. *IEEE Trans. Electron Devices*, 45(12):2400–2406, 2001.

Part III

CIRCUITS AND TECHNOLOGY

- **Basic Circuit Configurations**
- **Process Integration**
- **Applications**

6 Basic Circuit Configurations

Since the requirements of circuit design trigger process development, as well as the effort undertaken in device modeling, a knowledge of the basic circuit configurations is helpful for everyone working in the bipolar field. This chapter briefly[1] explains the most important building blocks of analog bipolar circuits. The last section explains the application of bipolar transistors in digital circuits.

6.1 Common-Emitter Configuration

The common-emitter configuration is the most widely employed amplifier configuration as it provides the best power gain.

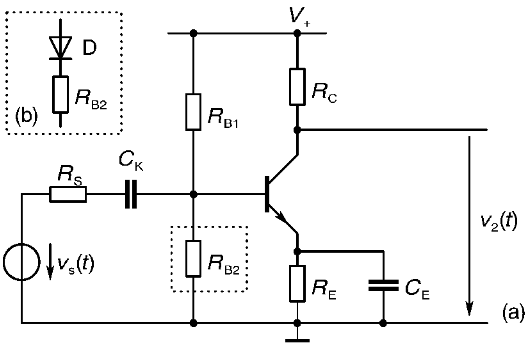


Fig. 6.1. (**a**) Basic amplifier in common-emitter configuration; (**b**) modification of voltage divider for temperature compensation of bias point

6.1.1 Biasing

Figure 6.1 shows the basic common-emitter amplifier, with a biasing network formed by the voltage divider R_{B1} and R_{B2} and the emitter series resistance R_E. At high frequencies, R_E is shunted by the capacitor C_E. The capacitor C_K couples the small-signal input signal delivered by the voltage source $v_S(t)$ to the input (base) of the amplifier. In the limit of large current gain ($B_N \to \infty$), the input bias point V_B and the emitter resistance R_E, required to obtain the bias current I_C are given by

$$V_B = \frac{R_{B2}}{R_{B1}+R_{B2}} V_+ \quad \text{and} \quad R_E = \frac{V_B - V_{BE}}{I_C} .$$

[1]For a more detailed presentation of bipolar-circuit techniques, the excellent textbook of Gray and Meyer [1] is recommended.

From

$$V_2 = V_+ - R_C I_C \approx V_+ - R_C \frac{V_B - V_{BE}}{R_E} ,$$

we obtain the change of the bias point with temperature as

$$\frac{dV_2}{dT} \approx -\frac{R_C}{R_E}\left(\frac{dV_B}{dT} - \frac{dV_{BE}}{dT}\right) ,$$

if the ratio R_C/R_E is assumed to be independent of temperature. Therefore, if $dV_B/dT = dV_{BE}/dT$, the bias point is to a good approximation independent of temperature. Under the condition $V_+ \gg V_{BE}$, this requirement may be fulfilled by adding a diode in series with R_{B2}, as is depicted in Fig. 6.1b.

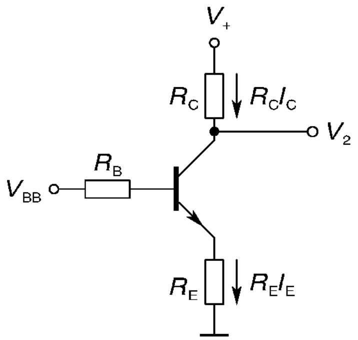


Fig. 6.2. Circuit used to investigate temperature sensitivity

Effect of Series Feedback on Bias Point Stability

Series feedback significantly reduces the sensitivity of the output bias point to variations of temperature or device parameters. To demonstrate this, we investigate the temperature variation of V_2 in the circuit shown in Fig. 6.2, neglecting the temperature dependence of the resistances for simplicity. Differentiation of the relation

$$V_2 = V_+ - R_C I_C = V_+ - R_C B_N \frac{V_{BB} - V_{BE} - R_E I_E}{R_B}$$

with respect to temperature yields

$$\frac{dV_2}{dT} = -\frac{1}{B_N}\frac{dB_N}{dT} R_C I_C + \frac{R_C}{R_B} B_N \frac{dV_{BE}}{dT} + R_C \frac{R_E B_N}{R_B}\frac{dI_E}{dT} .$$

Using the approximations

$$\left(\frac{\partial V_{BE}}{\partial I_C}\right)_T \approx \frac{1}{g_m} \approx \frac{V_T}{I_C} \quad \text{and} \quad \frac{dV_2}{dT} = -R_C \frac{dI_C}{dT}$$

gives, in addition,

$$\frac{dV_{BE}}{dT} = \left(\frac{\partial V_{BE}}{\partial T}\right)_{I_C} + \left(\frac{\partial V_{BE}}{\partial I_C}\right)_T \frac{dI_C}{dT} = \left(\frac{\partial V_{BE}}{\partial T}\right)_{I_C} - \frac{V_T}{R_C I_C}\frac{dV_2}{dT} .$$

Combining these equations and taking account of the fact that

$$-R_C \frac{dI_E}{dT} \approx -R_C \frac{dI_C}{dT} = \frac{dV_2}{dT}$$

results in

$$\frac{dV_2}{dT} = -\frac{R_B R_C I_C}{R_B + B_N(R_E + V_T/I_C)} \frac{1}{B_N}\frac{dB_N}{dT} + \frac{R_C B_N}{R_B + B_N(R_E + V_T/I_C)}\left(\frac{\partial V_{BE}}{\partial T}\right)_{I_C} . \tag{6.1}$$

In the limit $R_E \to 0$ and $R_B I_C/V_T \gg B_N$, we obtain

$$\frac{dV_2}{dT} \approx -\frac{1}{B_N}\frac{dB_N}{dT} R_C I_C \approx -6\,\frac{\mathrm{mV}}{\mathrm{K}}\frac{R_C I_C}{\mathrm{V}} - 1.5\,\frac{\mathrm{mV}}{\mathrm{K}}\frac{R_C}{R_B} B_N ,$$

where we have estimated the temperature coefficient of the current gain as 0.6%/K and $\partial V_{BE}/\partial T \approx -1.5$ mV/K. In this situation, V_{BB} in conjunction with the large base resistance R_B essentially acts like a current source. Owing to the temperature coefficient of the current gain, a considerable temperature drift of the bias point results (typically -30 mV/K for $R_C I_C = 5$ V and $R_B \to \infty$). The bias point is therefore always defined using emitter series feedback. In the limit $R_B \to 0$ we obtain the approximation

$$\frac{dV_2}{dT} \approx \frac{R_C}{R_E + V_T/I_C}\left(\frac{\partial V_{BE}}{\partial T}\right)_{I_C} \approx -1.5\,\frac{\mathrm{mV}}{\mathrm{K}}\frac{R_C}{R_E} ,$$

resulting in a much smaller temperature drift for resistance ratios R_C/R_E of the order of one.

Stability of Bias Point in the Avalanche Regime

The open-base breakdown voltage BV_{CEO} is relevant if the base is driven with a current source of infinite output resistance. Larger values of V_{CE} may be applied if the base is driven with a source of finite resistance. The stability of an amplifier circuit, which employs a bipolar transistor with $V_{CE} > BV_{CEO}$, can be determined from a small-signal analysis. For that purpose, we transform the amplifier circuit depicted in Fig. 6.1 to the small-signal equivalent circuit shown in Fig. 6.3, where g_o, g_μ and $r_{cc'}$ have been omitted for simplicity. The resistances R_{B1} and R_{B2} have been combined into $R_B = R_{B1} \,\|\, R_{B2}$, and the combined effect of the emitter series resistance $r_{cc'}$, the series feedback resistor R_E and the bypass capacitor C_E is represented by the impedance Z_E.

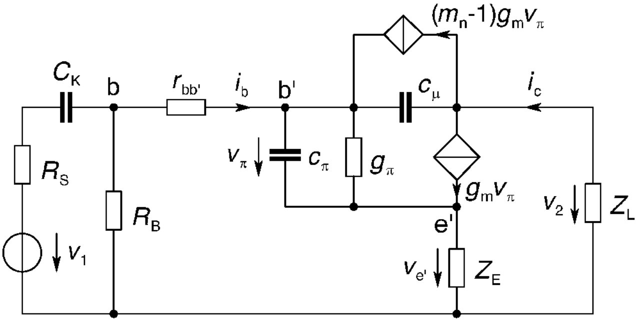


Fig. 6.3. Small-signal equivalent circuit corresponding to Fig. 6.1

For typical operation frequencies, C_K and C_E can be considered as short-circuited, while c_μ and c_π can be omitted. In this case $Z_E = r_{ee'}$, and the transfer function is obtained as

$$\frac{\underline{v}_2}{\underline{v}_1} = -\frac{K m_n R_C g_m}{1 + g_\pi \left\{ r_b[1 - \beta(m_n - 1)] + (\beta + 1) r_{ee'} \right\}} , \tag{6.2}$$

where $r_b = r_{bb'} + R_S \,||\, R_B$ and $K = R_B/(R_S + R_B)$. This expression diverges for

$$1 + g_\pi r_b[1 - \beta(m_n - 1)] + r_{ee'} g_\pi (\beta + 1) \rightarrow 0 . \tag{6.3}$$

If we make the approximation $g_\pi(\beta + 1) \approx \beta g_\pi = g_m$, this condition is equivalent to

$$m_n - 1 = \frac{1}{\beta} + \frac{1/g_m + r_{ee'}}{r_{bb'} + R_S \,||\, R_B} . \tag{6.4}$$

With the approximation

$$m_n - 1 \approx 1 - \frac{1}{M_n} \approx \left(\frac{V_{C'B'}}{BV_C} \right)^{N_1} , \tag{6.5}$$

the value of $V_{C'B'}$ that leads to breakdown can be estimated as

$$V_{CBR} \approx BV_C \left(\frac{1}{\beta} + \frac{1/g_m + r_{ee'}}{r_{bb'} + R_S \,||\, R_B} \right)^{1/N_1} . \tag{6.6}$$

Since (6.5) tends to overestimate the value of $m_n - 1$, the value of V_{CBR} determined from (6.6) provides a conservative estimate of the maximum allowed V_{CE}-value, $V_{CEmax} = V_{BE} + V_{CBR}$. In the limit $r_b \rightarrow \infty$, corresponding to the case of a current source applied to the input, the result derived here reduces to $V_{CEmax} \rightarrow V_{BE} + BV_C \, \beta^{-1/N_1} \approx BV_{CEO}$, as expected. In order to obtain a stable dc bias point, the value of R_B has to be chosen to be sufficiently small. For $\omega \rightarrow 0$, the impedances represented by C_K and C_E can be assumed to be infinite. In this case $Z_E = r_{ee'} + R_E$, and $R_S \,||\, R_B$ reduces to R_B.

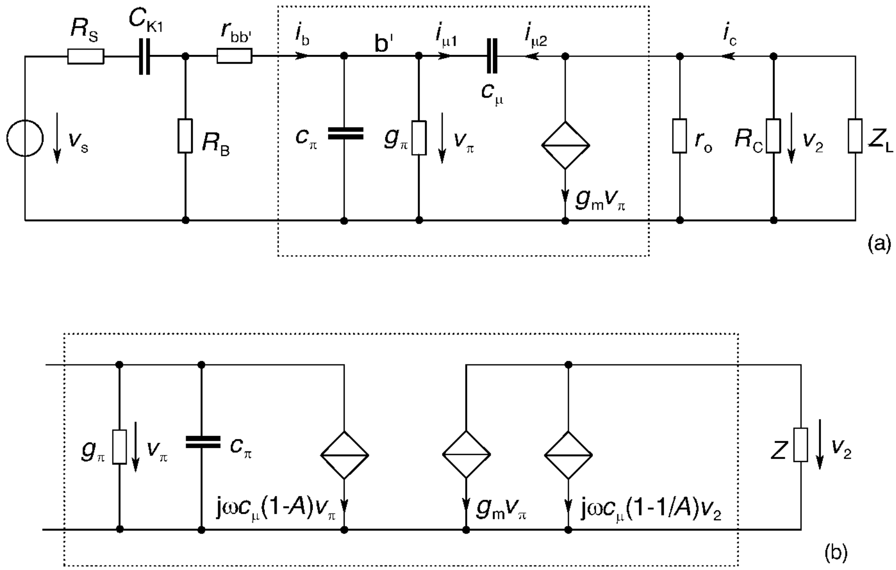


Fig. 6.4. small-signal equivalent circuit of a bipolar transistor amplifier stage in the common-emitter configuration, (**a**) before and (**b**) after network transformation

6.1.2 AC Characteristics

The frequency-dependent voltage transfer ratio of an amplifier in the common-emitter configuration can be determined from the small-signal equivalent circuit shown in Fig. 6.4, which describes the load connected to the output in terms of a load impedance Z_L. The input and output ports are linked by a voltage-controlled current source that delivers a small-signal transfer current $g_m \underline{v}_\pi$, and by a capacitance c_μ. As seen from the input, a current

$$\underline{i}_{\mu 1} \;=\; \mathrm{j}\omega c_\mu(\underline{v}_\pi - \underline{v}_2) \;=\; \mathrm{j}\omega c_\mu(1-\underline{A})\,\underline{v}_\pi \;, \quad \text{where} \quad \underline{A} = \underline{v}_2/\underline{v}_\pi \;,$$

flows into c_μ, while, as seen from the output, a current

$$\underline{i}_{\mu 2} \;=\; \mathrm{j}\omega c_\mu(\underline{v}_2 - \underline{v}_\pi) \;=\; \mathrm{j}\omega c_\mu(1-1/\underline{A})\,\underline{v}_2$$

flows. If these currents are represented by voltage-controlled current sources, the equivalent circuit depicted in Fig. 6.4b results, where Z denotes an impedance composed of R_C, r_o and the load impedance Z_L. From Kirchhoff's current law, which gives

$$\underline{v}_2/Z + g_m \underline{v}_\pi + \mathrm{j}\omega c_\mu(\underline{v}_2 - \underline{v}_\pi) \;=\; 0 \;,$$

the voltage transfer factor

$$\underline{A} \;=\; -\frac{g_m - \mathrm{j}\omega c_\mu}{1/Z + \mathrm{j}\omega c_\mu} \;\approx\; -\frac{g_m}{1/Z + \mathrm{j}\omega c_\mu} \tag{6.7}$$

can be obtained. If the load is composed of a load resistance R_L parallel to a load capacitance C_L, we have

$$\frac{1}{\underline{Z}} + j\omega c_\mu = \frac{1}{R_C} + \frac{1}{r_o} + \frac{1}{R_L} + j\omega(c_\mu + C_L) = \frac{1 + j\omega R(c_\mu + C_L)}{R},$$

where $R = R_C \parallel r_o \parallel R_L$. If $\omega c_\mu \ll g_m$, the voltage transfer ratio $\underline{A}$ shows a low-pass characteristic of first order

$$\underline{A} \approx -\frac{g_m R}{1 + jf/f_{go}},$$

where

$$f_{go} = \frac{1}{2\pi R(c_\mu + C_L)}$$

is the cutoff frequency due to the low-pass characteristic of the output circuit. For frequencies $f \ll f_{go}$, the transfer factor is approximately constant, i.e. $\underline{A} \approx -g_m R$.

As seen from the input, the Miller capacitance $c_m = (1 - \underline{A})c_\mu$ has to be considered as being parallel to c_π. In an inverting amplifier where $\underline{A} = -A$, the Miller capacitance is just $(A+1)$ times the bc depletion capacitance c_μ.[2] That the contribution of c_μ is weighted by the voltage gain becomes obvious if we consider a small change v_π of the base potential: since the amplifier is inverting, the collector potential changes by $v_2 = -Av_\pi$, i.e. the voltage swing across c_μ is $(A+1)v_\pi$. As seen from the input, the input capacitance $c_i = c_\pi + (1 - \underline{A})c_\mu$ has to be considered. The input impedance z_i of the amplifier is therefore (Fig. 6.5b)

$$z_i = \frac{1}{g_\pi + j\omega[c_\pi + c_\mu(1 - \underline{A})]}. \tag{6.8}$$

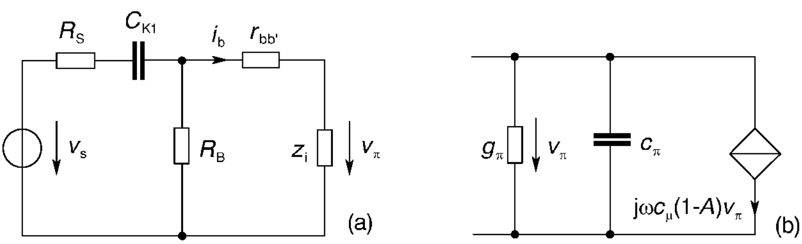


Fig. 6.5. (**a**) Equivalent circuit of the network connected to the input of a terminated transistor two-port (represented by the input impedance z_i; (**b**) network elements that contribute to z_i

[2]At high frequencies, the Miller capacitance is a complex quantity, since in this case $\underline{A}$ describes a phase shift that differs from 180°.

The voltage drop $\underline{v}_\pi$ across z_i controls the transfer current source. We neglect the effect of R_B, which is assumed to be large in comparison with the other impedances, and combine the generator series resistance R_S and the base series resistance $r_{\mathrm{bb'}}$ to obtain $r_\mathrm{b} = R_\mathrm{S} + r_{\mathrm{bb'}}$. Analysis of the resulting voltage divider yields

$$\frac{\underline{v}_\pi}{\underline{v}_\mathrm{s}} = \frac{z_\mathrm{i}}{z_\mathrm{i} + r_\mathrm{b} + (\mathrm{j}\omega C_{\mathrm{K1}})^{-1}} .$$

At low frequencies, $z_\mathrm{i} \approx r_\pi$, resulting in a high-pass characteristic

$$\frac{\underline{v}_\pi}{\underline{v}_\mathrm{s}} \approx \frac{r_\pi}{r_\pi + r_\mathrm{b}} \frac{1}{1 - \mathrm{j} f_\mathrm{K}/f} , \quad \text{where} \quad f_\mathrm{K} = \frac{1}{2\pi(r_\pi + r_\mathrm{b}) C_{\mathrm{K1}}} .$$

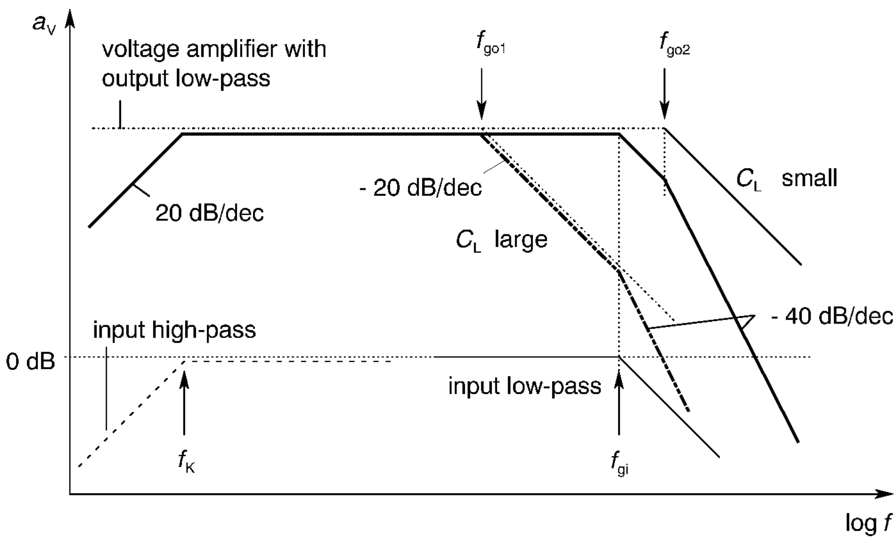


Fig. 6.6. Frequency characteristic of a bipolar-transistor amplifier in common-emitter configuration

Generally, the frequency of operation is chosen to be large in comparison with f_K, and C_{K1} may be replaced by a short circuit in this range. For $f \gg f_\mathrm{K}$,

$$\frac{\underline{v}_\pi}{\underline{v}_\mathrm{s}} = \frac{1}{1 + r_\mathrm{b}/z_\mathrm{i}} = \frac{1}{1 + r_\mathrm{b} g_\pi + \mathrm{j}\omega r_\mathrm{b} c_\mathrm{i}} . \tag{6.9}$$

Since generally the condition $r_\mathrm{b} g_\pi \ll 1$ is fulfilled, this corresponds to a low-pass behavior with a cutoff frequency

$$f_\mathrm{gi} = \frac{1}{2\pi r_\mathrm{b} c_\mathrm{i}} . \tag{6.10}$$

If we write $g_\mathrm{m} r_\pi = \beta$, the voltage transfer factor $\underline{v}_2/\underline{v}_\mathrm{s}$ of the amplifier

$$\frac{\underline{v}_2}{\underline{v}_\mathrm{s}} = \frac{\underline{v}_\pi}{\underline{v}_\mathrm{s}}\frac{\underline{v}_2}{\underline{v}_\pi} \approx -\frac{\beta R}{r_\pi+r_\mathrm{b}}\frac{1}{1-\mathrm{j}f_\mathrm{K}/f}\frac{1}{1+\mathrm{j}f/f_\mathrm{gi}}\frac{1}{1+\mathrm{j}f/f_\mathrm{go}}, \tag{6.11}$$

can be approximated by the product of the maximum voltage transfer ratio $-\beta R/(r_\mathrm{b}+r_\pi)$ with a high-pass transfer function (cutoff frequency f_K) and two low-pass transfer functions (cutoff frequencies f_gi for the low-pass function associated with the input circuit, and f_go for the low-pass function associated with the output circuit). In a Bode diagram, $a_v = 20\,\mathrm{dB}\log(|\underline{H}_v|)$ is plotted versus the logarithm of the frequency. Since the logarithm of a product equals the sum of the logarithms of the individual factors, we can write

$$\begin{aligned} a_v &= 20\,\mathrm{dB}\,\log\left(\frac{\beta R}{r_\pi+r_\mathrm{b}}\right) + 20\,\mathrm{dB}\,\log\left|\frac{1}{1-\mathrm{j}f_\mathrm{K}/f}\right| \\ &+20\,\mathrm{dB}\,\log\left|\frac{1}{1+\mathrm{j}f/f_\mathrm{gi}}\right| + 20\,\mathrm{dB}\,\log\left|\frac{1}{1+\mathrm{j}f/f_\mathrm{go}}\right|, \end{aligned}$$

and a_v may be determined by geometrical construction, as illustrated in Fig. 6.6 for different values of C_L.

6.1.3 Nonlinear Distortion

There are several sources of nonlinear distortion[3]: (1) distortion caused by the exponential current–voltage characteristic; (2) distortion due to the nonlinear bc depletion capacitance, a contribution that becomes increasingly important as the operating frequency increases, (3) distortion due to the avalanche effect in the bc diode; and (4) effects due to the nonlinear bias dependence of the base and collector series resistances.

In [7], it was found that bipolar-transistor amplifiers operating under class A conditions achieve the highest linearity in a bias range where current spreading in the epilayer is basically absent. Both the quasi-saturation region and the avalanche region are unfavorable if optimum IP_3 performance (see Appendix B) is required. The influence of the nonlinear base resistance was found to be of minor importance.

The bias dependence of the bc depletion capacitance is an important mechanism of third-order intermodulation distortion, which is determined to a large extent by the vertical doping profile in the collector region. The effect of modifications of the collector doping profile[4] on nonlinear distortion has been investigated in [8, 9]; there the incorporation of an epitaxially grown arsenic doping spike into an otherwise lightly doped collector was found to reduce the bias dependence of c_jc and and thus third-order intermodulation at high

[3]See Appendix A for basic definitions. A detailed exposition of the computational methods employed in small-signal distortion analysis is given in [2, 3]; see also [4–6].

[4]Such modifications have to take account of high-level-injection effects and breakdown requirements.

frequencies. Negative feedback results in a reduction of nonlineary. In [10] it was found that common-emitter transconductance stages with inductive series feedback are more linear than resistively or capacitively degenerated stages. The reduction of third-order intermodulation distortion in a balanced common-emitter configuration is described in [11].

6.2 Common-Collector Configuration

A bipolar-transistor amplifier in the common-collector configuration has near-unity voltage gain and provides only current gain. It serves as a level shifter and impedance transformer in integrated bipolar-transistor circuits.

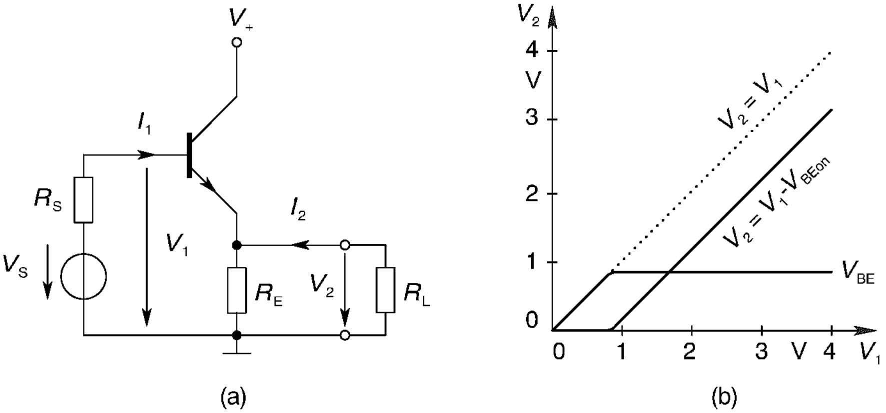


Fig. 6.7. (**a**) Bipolar-transistor amplifier in common-collector configuration (emitter follower), and (**b**) dc voltage transfer characteristic

6.2.1 Basic Principles

Figure 6.7 shows a bipolar-transistor amplifier in the common-collector configuration, also known as the emitter follower. If V_1 exceeds the threshold voltage of the eb diode, an emitter current I_E flows, which produces a voltage drop V_2 across R_E, i.e. $I_E = V_2/R_E + I_L$, if $I_L = -I_2$ is the current that flows into the load. Since in the limit $B_N \gg 1$ the voltage drop across the eb diode is $V_{BE} = V_T \ln(I_E/I_S)$, the following relation between the voltages at the input and at the output is obtained:

$$V_1 = V_2 + V_T \ln\left(\frac{V_2}{R_E I_S} + \frac{I_L}{I_S}\right) . \tag{6.12}$$

Since V_{BE} remains approximately constant for emitter current values in the milliampere range, the output voltage V_2 is related to V_1 by a shift of the voltage level by V_{BEon} (Fig. 6.7b). The output (i.e. the emitter) therefore follows the input (base) of the transistor; for this reason this circuit is frequently called the emitter follower. Since generally $B_{\mathrm{N}} \ll 1$, this circuit has current gain, but does not provide voltage gain ($\mathrm{d}V_2/\mathrm{d}V_1 \approx 1$).

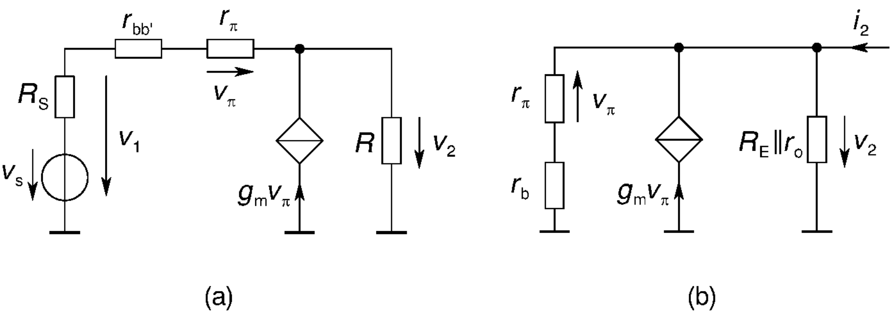


Fig. 6.8. Small-signal equivalent circuits used for the computation of (**a**) the input resistance and voltage gain of the emitter follower, and (**b**) the output resistance of the emitter follower

The input resistance and voltage gain of the emitter follower circuit can be calculated with the help of the small-signal equivalent circuit shown in Fig. 6.8a, where $R = R_{\mathrm{E}} \,\|\, R_{\mathrm{L}} \,\|\, r_{\mathrm{o}}$, and $r_{\mathrm{b}} = R_{\mathrm{S}} + r_{\mathrm{bb'}}$ is dominated by R_{S} in most cases. From Kirchhoff's current law, we obtain $(g_{\mathrm{m}}+g_\pi)v_\pi = v_2/R$, or

$$\frac{v_\pi}{v_2} = \frac{1}{R(g_{\mathrm{m}}+g_\pi)} \approx \frac{1}{Rg_{\mathrm{m}}} \,.$$

If we use this result in $v_{\mathrm{s}} = v_2 + v_\pi + r_{\mathrm{b}} g_\pi v_\pi$, this yields the following for the small-signal voltage gain:

$$\frac{v_{\mathrm{s}}}{v_2} = 1 + (1+r_{\mathrm{b}}g_\pi)\frac{v_\pi}{v_2} \approx 1 + \frac{1 + r_{\mathrm{b}}g_\pi}{Rg_{\mathrm{m}}} \,,$$

or, if we take account of the relation $\beta = g_{\mathrm{m}}/g_\pi$,

$$\frac{v_2}{v_{\mathrm{s}}} = \frac{1}{1 + r_{\mathrm{b}}/\beta R + 1/g_{\mathrm{m}}R} \,. \tag{6.13}$$

Using the result

$$v_1 = \left(1+r_{\mathrm{bb'}}g_\pi+\frac{v_2}{v_\pi}\right) v_\pi \approx (1+r_{\mathrm{bb'}}g_\pi+Rg_{\mathrm{m}})v_\pi$$

we obtain the input resistance of the emitter follower:

$$r_1 = \frac{v_1}{i_1} = r_\pi(1+r_{\mathrm{bb'}}g_\pi+Rg_{\mathrm{m}}) \approx \beta R \,. \tag{6.14}$$

The output resistance $r_2 = v_2/i_2$ of the emitter follower can be calculated from the small-signal equivalent circuit shown in Fig. 6.8b. Combining the relations

$$i_2 + (g_{\mathrm{m}}+g_\pi)v_\pi - \frac{v_2}{R_{\mathrm{E}} \,\|\, r_{\mathrm{o}}} = 0 \quad \text{and} \quad v_\pi = -\frac{r_\pi}{r_\pi + r_{\mathrm{b}}} v_2$$

yields

$$r_2 = \left(\frac{\beta+1}{r_\pi + r_{\mathrm{b}}} + \frac{1}{R_{\mathrm{E}} \,\|\, r_{\mathrm{o}}}\right)^{-1} = R_{\mathrm{E}} \,\|\, r_{\mathrm{o}} \,\|\, \frac{r_\pi + r_{\mathrm{b}}}{\beta+1} \,. \tag{6.15}$$

Typically, the output resistance is much smaller than the input resistance, which is why this circuit is used for impedance transformation.

6.2.2 AC Characteristics

At high frequencies, the transistor and load capacitances have to be considered. The voltage transfer ratio of a capacitively loaded emitter follower is derived in the following small-signal analysis based on the equivalent circuit shown in Fig. 6.9. The generator series resistance R_{S} and the base series re-

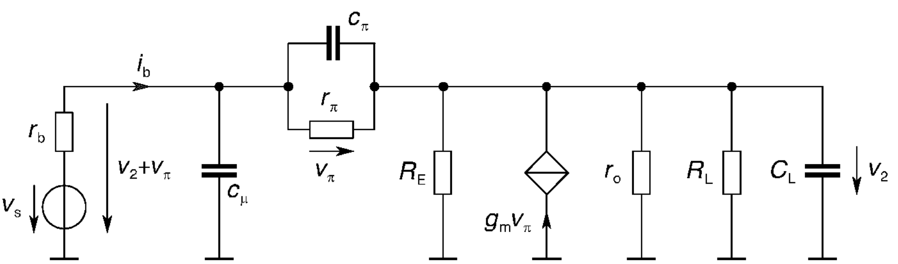


Fig. 6.9. Small-signal equivalent circuit used for the discussion of the ac behavior of a capacitively loaded emitter follower

sistance $r_{\mathrm{bb'}}$ have been combined into r_{b}; R denotes the parallel connection of R_{E}, r_{o} and R_{L}. Using

$$\underline{i}_{\mathrm{b}} = \mathrm{j}\omega c_\mu(\underline{v}_2 + \underline{v}_\pi) + (g_\pi + \mathrm{j}\omega c_\pi)\underline{v}_\pi$$

in $\underline{v}_{\mathrm{s}} = \underline{v}_2 + \underline{v}_\pi + r_{\mathrm{b}}\underline{i}_{\mathrm{b}}$ and taking account of the relation $\omega_\beta = g_\pi/(c_\pi + c_\mu)$ yields

$$\underline{v}_{\mathrm{s}} = (1 + \mathrm{j}\omega r_{\mathrm{b}} c_\mu)\,\underline{v}_2 + [1 + r_{\mathrm{b}} g_\pi\,(1 + \mathrm{j}\omega/\omega_\beta)]\,\underline{v}_\pi \,,$$

or, for the inverse of the voltage transfer ratio

$$\frac{\underline{v}_{\mathrm{s}}}{\underline{v}_2} = 1 + \mathrm{j}\omega r_{\mathrm{b}} c_\mu + \left[1 + r_{\mathrm{b}} g_\pi \left(1 + \mathrm{j}\frac{\omega}{\omega_\beta}\right)\right] \frac{\underline{v}_\pi}{\underline{v}_2} \,.$$

The voltage transfer ratio $\underline{v}_\pi/\underline{v}_2$ is obtained from Kirchhoff's current law, which gives

$$(g_\mathrm{m}+g_\pi+\mathrm{j}\omega c_\pi)\underline{v}_\pi \;=\; \frac{\underline{v}_2}{R}+\mathrm{j}\omega C_\mathrm{L}\underline{v}_2$$

where $R = R_\mathrm{E}\,\|\,R_\mathrm{L}\,\|\,r_\mathrm{o}$, with the result

$$\frac{\underline{v}_\pi}{\underline{v}_2} \;=\; \frac{1+\mathrm{j}\omega RC_\mathrm{L}}{R(g_\mathrm{m}+g_\pi+\mathrm{j}\omega c_\pi)} \;\approx\; \frac{1+\mathrm{j}\omega RC_\mathrm{L}}{Rg_\mathrm{m}}\,.$$

The approximation is valid if the conditions $g_\mathrm{m}/g_\pi = \beta \gg 1$ and $\omega c_\pi/g_\mathrm{m} \ll 1$ are fulfilled, which is generally the case as long as $f \ll f_\mathrm{T}$. Multiplication of the voltage transfer ratios yields using $\omega_\mathrm{T} = \beta\omega_\beta$,

$$\frac{\underline{v}_\mathrm{s}}{\underline{v}_2} \;\approx\; 1+\mathrm{j}\omega\left[r_\mathrm{b}c_\mu+\left(\frac{1}{g_\mathrm{m}}+\frac{r_\mathrm{b}}{\beta}\right)C_\mathrm{L}+\frac{r_\mathrm{b}}{\beta R}\,\frac{1}{\omega_\beta}\right]-\omega^2\,\frac{r_\mathrm{b}C_\mathrm{L}}{\omega_\mathrm{T}}\,,$$

where the constant terms $1/Rg_\mathrm{m}$ and $r_\mathrm{b}/\beta R$ have been neglected. The voltage transfer factor thus has a frequency dependence that is analogous to the frequency dependence of an LRC series resonant circuit with resonant frequency

$$\omega_0 \;=\; \sqrt{\frac{\omega_\mathrm{T}}{r_\mathrm{b}C_\mathrm{L}}} \tag{6.16}$$

and damping constant

$$\delta \;=\; \left(\frac{c_\mu}{2C_\mathrm{L}}+\frac{1}{2r_\mathrm{b}g_\mathrm{m}}+\frac{1}{2\beta}\right)\omega_\mathrm{T}+\frac{1}{2RC_\mathrm{L}}\,.$$

Owing to the inductive nature of the emitter follower, voltage noise and ringing may occur during switching transients if the quality factor $Q = \omega_0/2\delta$ exceeds $1/\sqrt{2}$.

6.3 Common-Base Configuration

Figure 6.10 shows a bipolar-transistor amplifier in the common-base configuration. In this circuit, the input current $I_1 = -I_\mathrm{E}$ approximately equals the output current I_C, i.e. this amplifier configuration does not provide current gain. The common-base configuration therefore provides only voltage gain and has a small input resistance. The voltage transfer ratio and the input and output resistance can be calculated with help of the small-signal equivalent circuit shown in Fig. 6.11. There, the base potential is assumed to be constant ($\underline{v}_\mathrm{b} = 0$), a reasonable assumption if the resistances R_B1 and R_B2 are small enough. If we write $\underline{v}_1 = -\underline{v}_\pi$ and use the abbreviation $R = R_\mathrm{L}\,\|\,R_\mathrm{C}\,\|\,r_\mathrm{o}$, Kirchhoff's current law, applied to the output node, yields

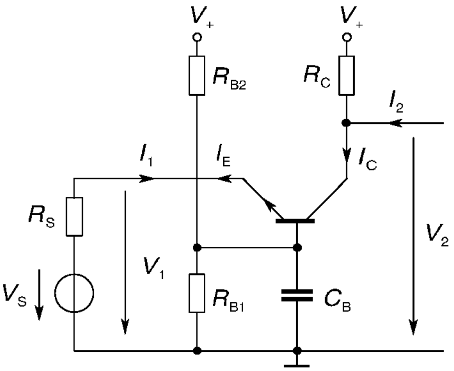


Fig. 6.10. Bipolar-transistor amplifier in common-base configuration

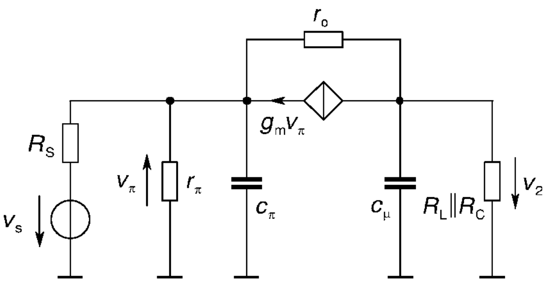


Fig. 6.11. Small-signal equivalent circuit of bipolar-transistor amplifier in common-base configuration

$$\left(\frac{1}{R} + \mathrm{j}\omega c_\mu\right) \underline{v}_2 \;=\; -\left(g_\mathrm{m} + \frac{1}{r_\mathrm{o}}\right) \underline{v}_\pi \;\approx\; g_\mathrm{m} \underline{v}_1 \,.$$

This leads to the following result for the voltage transfer ratio:

$$\frac{\underline{v}_2}{\underline{v}_1} \approx \frac{g_\mathrm{m} R}{1 + \mathrm{j}\omega R c_\mu} \approx g_\mathrm{m} R \,;$$

the voltage gain is approximately equal to the result for an amplifier in the common-emitter configuration. If applied to the input node, Kirchhoff's current law yields

$$\underline{i}_1 + (g_\mathrm{m} + g_\pi)\underline{v}_\pi + \mathrm{j}\omega c_\pi \underline{v}_\pi + g_\mathrm{o}(\underline{v}_2 + \underline{v}_\pi) \;=\; 0 \,.$$

If we write $\underline{v}_1 = -\underline{v}_\pi$ and $\underline{v}_2 = A_v \underline{v}_1$, this gives the following result for the input impedance:

$$z_\mathrm{i} \;=\; \frac{1}{g_\mathrm{m} + g_\pi + g_\mathrm{o}(1 - A_v) + \mathrm{j}\omega c_\pi} \approx \frac{1}{g_\mathrm{m}} \frac{1}{1 + \mathrm{j} f / f_\mathrm{T}} \,,$$

where the conditions $\beta \gg 1$, $g_\mathrm{o} A_v \ll g_\mathrm{m}$ and $2\pi f_\mathrm{T} \approx g_\mathrm{m}/c_\pi$ has been assumed. Generally, the input impedance is small in comparison with the generator series resistance ($|z_\mathrm{i}| \ll R_\mathrm{S}$), and the input of the amplifier is driven essentially by the current $\underline{i}_1 \approx \underline{v}_\mathrm{s}/R_\mathrm{S}$. In this case $\underline{v}_1 \approx z_\mathrm{i} \underline{v}_\mathrm{s}/R_\mathrm{S}$, and

$$\underline{v}_2 \approx \frac{g_\mathrm{m} R}{1 + \mathrm{j}\omega R c_\mu} \frac{z_\mathrm{i}}{R_\mathrm{S}} \underline{v}_\mathrm{s} \approx \frac{R}{R_\mathrm{S}} \frac{1}{1 + \mathrm{j}\omega R c_\mu} \frac{1}{1 + \mathrm{j}f/f_\mathrm{T}} \underline{v}_\mathrm{s} ,$$

i.e. the bandwidth of this amplifier configuration is very large, owing to the large cutoff frequency f_T and because no Miller effect is present.

The output resistance can be obtained from the small-signal equivalent circuit, neglecting the capacitances and assuming $R_\mathrm{S} \,\|\, r_\pi \,\|\, r_\mathrm{o} \approx R_\mathrm{S} \,\|\, r_\pi$, with the result

$$r_\mathrm{a} = R_\mathrm{C} \parallel \left(1 + \frac{\beta R_\mathrm{S}}{R_\mathrm{S} + r_\pi} \right) r_\mathrm{o} \ ,$$

using $g_\mathrm{m} r_\pi = \beta$. For small values of the source impedance R_S, the output resistance is approximately equal to the output resistance for an amplifier in the common-emitter configuration.

6.4 The Diode-Connected Bipolar Transistor

The pn diodes used in bipolar integrated circuits are generally derived from bipolar transistors. In principle, there are six ways to operate a three-terminal bipolar transistor as a diode (Fig. 6.12).

6.4.1 Realizations

Among the possibilities depicted in Fig. 6.12, circuit (a) is generally employed if the low breakdown voltage BV_EBO of the eb diode is acceptable. In this circuit, saturation of the bc diode is avoided. This diode shows a fast reverse recovery time, which is determined by the minority-carrier lifetime in the base region and the depletion capacitance of the eb diode.

If a larger breakdown voltage is necessary, circuit (b) should be employed. This diode is equivalent to the bc diode of a transistor; the emitter diffusion is omitted. Forward operation of this diode, however, causes minority-carrier injection into the substrate and requires sufficient substrate contacts nearby in order to avoid problems. The reverse recovery time is determined by the minority-carrier lifetime in the collector region and is therefore generally considerably larger than the value observed for diode (a).

If the designer has to employ readily drawn transistors, the diode-connected transistor (c) will do approximately the same job; the emitter could be left open, without too much loss of performance. The latter possibility, which is not shown, is, however, not recommended, as it can cause operation of the bipolar transistor in saturation and lead to unwanted minority-carrier injection into the substrate.

Circuit (d) uses a floating collector, which becomes forward biased owing to electrons injected into the collector region by the transfer current. This

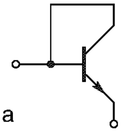
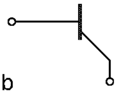
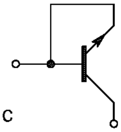
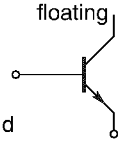
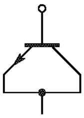
diode realization	saturation current	breakdown voltage	substrate injection	reverse recovery
 a	$I_S \dfrac{B_F+1}{B_F}$	BV_{EBO}	no	fast
 b	I_S/B_R	BV_{CBO}	yes, under forward operation	slow
 c	$I_S \dfrac{B_R+1}{B_R}$	BV_{CBO}	yes, under forward operation	slow
 d	$I_S \dfrac{B_F+B_R+1}{B_R+1}$	BV_{EBO}	yes, under forward operation	slow
 e	$I_S \dfrac{B_F+B_R}{B_F B_R}$	BV_{EBO}	yes, under forward operation	slow

Fig. 6.12. Realizations of a diode in an integrated bipolar technology

results in minority-carrier injection into the substrate and a considerably increased reverse recovery time – two reasons why this circuit should not be used. In circuit (e), the eb and bc diodes are connected in parallel. Since the breakdown voltage of this configuration is determined by the smaller breakdown voltage BV_{EBO}, circuit (a) should be used instead, since circuit (e) shows both minority-carrier injection into the substrate and an increased reverse recovery time due to the parallel-connected bc diode.

6.4.2 Current–Voltage Characteristic

The voltage drop $V = V_{CE} = V_{BE}$ across the diode-connected BJT corresponding to circuit a equals the base–emitter voltage; the current carried by the diode is therefore

$$I = I_C + I_B = \frac{I_S}{q_B}\left(1+\frac{1}{B_N}\right)\exp\left(\frac{V}{V_T}\right) ,$$

if series resistance effects may be neglected. Neglecting the Early effect ($q_B \approx 1$) and nonideal base current characteristics ($B_N \approx B_F$) gives

$$I \approx I_S\left(1+\frac{1}{B_F}\right)\exp\left(\frac{V}{V_T}\right) ,$$

i.e., in the elementary model, the saturation current of the diode-connected BJT is $I_S(1+1/B_F)$, and is approximately equal to the transfer saturation current if $B_F \ll 1$. Owing to the bias dependence of q_B and B_N, a slight deviation of the emission coefficient from one will be observed:

$$N = \frac{1}{V_T}\left(\frac{\mathrm{d}\ln I}{\mathrm{d}V}\right)^{-1} = \left[1 - \frac{V_T}{q_B}\frac{\mathrm{d}q_B}{\mathrm{d}V} - \frac{V_T}{B_N^2+B_N}\frac{\mathrm{d}B_N}{\mathrm{d}V}\right]^{-1} .$$

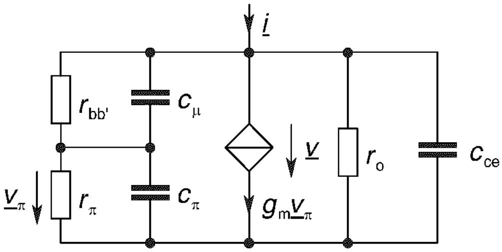


Fig. 6.13. Small-signal equivalent circuit of a diode-connected BJT

6.4.3 High-Frequency Behavior

The high-frequency behavior of a diode-connected bipolar transistor can be derived from the small-signal equivalent circuit[5] shown in Fig. 6.13. Since

$$\underline{i} = (g_m + g_\pi + \mathrm{j}\omega c_\pi)\underline{v}_\pi + (g_o + \mathrm{j}\omega c_{ce})\underline{v}$$

and

$$\underline{v}_\pi = \frac{1 + \mathrm{j}\omega r_{bb'} c_\mu}{1 + r_{bb'} g_\pi + \mathrm{j}\omega r_{bb'}(c_\pi + c_\mu)}\,\underline{v} ,$$

the small-signal admittance of the diode is obtained as

$$\underline{y} = \frac{\underline{i}}{\underline{v}} = g_o + \mathrm{j}\omega c_{ce} + \frac{g_m + g_\pi}{1 + r_{bb'} g_\pi}\,\frac{1 + \mathrm{j}f/f_1 - f^2/f_3^2}{1 + \mathrm{j}f/f_2} ,$$

where

[5]In a diode-connected integrated bipolar transistor, the coupling between the collector and emitter will be due to the cs capacitance (in series with the substrate resistance) to a certain extent; in this case c_{ce} has to be replaced by a more general coupling admittance.

$$f_1 = \frac{g_m + g_\pi}{2\pi[\,c_\pi + r_{bb'}(g_m + g_\pi)\,c_\mu\,]}, \quad f_2 = \frac{1 + r_{bb'}g_\pi}{2\pi r_{bb'}(c_\pi + c_\mu)}$$

and

$$f_3 = \frac{1}{2\pi}\sqrt{\frac{g_m + g_\pi}{r_{bb'}c_\pi c_\mu}}\,.$$

As the transistor model is accurate only to first order of the frequency $f = \omega/2\pi$, a consistent description of effects associated with terms of order ω^2 is not possible with the equivalent circuit of Fig. 6.13. However, even though the frequency f_3 does not provide a complete description of second-order frequency effects, the above result shows that the base series resistance in conjunction with the device capacitances causes resonant behavior, which can be seen from an increase of the frequency-dependent small-signal diode impedance [12].

6.5 Current Sources and Active Loads

A bipolar transistor with an output conductance $\underline{y}_{22e}$ equal to zero works as a current source if V_{BE} is held constant: if $\underline{y}_{22e} = 0$, the collector current is independent of V_{CE}, i.e. independent of the voltage drop that appears across a load connected in series with the collector. In real transistors, $\underline{y}_{22e} \approx 1/r_o$ is finite, and the corresponding current source is nonideal. Another problem with constant-voltage biasing is the strong temperature dependence of the collector current. Therefore some sort of feedback, such as the series feedback considered in the following section or the compensation of temperature effects provided by a current mirror, is required in order to arrive at a good current source.

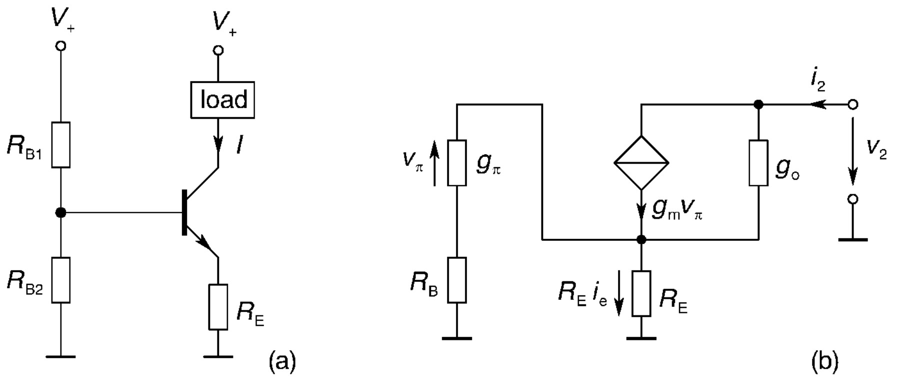


Fig. 6.14. Bipolar current source with series feedback: (**a**) circuit and (**b**) small-signal equivalent circuit

6.5.1 Current Source with Series Feedback

The output resistance of the current source shown in Fig. 6.14a can be obtained from the small-signal equivalent circuit shown in Fig. 6.14b. The voltage divider composed of $R_{\mathrm{B}} = R_{\mathrm{B1}} \| R_{\mathrm{B2}}$ and r_π gives

$$v_\pi = -\frac{r_\pi}{r_\pi + R_{\mathrm{B}}} R_{\mathrm{E}} i_{\mathrm{e}} ,$$

and therefore

$$i_2 = g_{\mathrm{m}} v_\pi + g_{\mathrm{o}}(v_2 - R_{\mathrm{E}}\, i_{\mathrm{e}}) = -\left(\frac{r_\pi g_{\mathrm{m}}}{r_\pi + R_{\mathrm{B}}} + g_{\mathrm{o}}\right) R_{\mathrm{E}}\, i_{\mathrm{e}} + g_{\mathrm{o}} v_2 .$$

Combinining this result with

$$i_{\mathrm{e}} = i_2 + g_\pi v_\pi = i_2 - \frac{g_\pi r_\pi}{r_\pi + R_{\mathrm{B}}} R_{\mathrm{E}}\, i_{\mathrm{e}}$$

and using the relation $\beta = g_{\mathrm{m}} r_\pi$ gives the following for the output resistance:

$$r_2 = \frac{v_2}{i_2} = r_{\mathrm{o}} + R_{\mathrm{E}} \frac{\beta r_{\mathrm{o}} + r_\pi + R_{\mathrm{B}}}{r_\pi + R_{\mathrm{B}} + R_{\mathrm{E}}} . \tag{6.17}$$

The output resistance of the current source equals the output resistance r_{o} of the bipolar transistor if there is no series feedback present ($R_{\mathrm{E}} = 0$). In the case of a weak series feedback ($R_{\mathrm{E}} \ll r_\pi$), we obtain

$$r_2 \approx r_{\mathrm{o}} + R_{\mathrm{E}}(1 + \beta r_{\mathrm{o}}/r_\pi) = R_{\mathrm{E}} + r_{\mathrm{o}}(1 + g_{\mathrm{m}} R_{\mathrm{E}}) ,$$

whereas in the case of strong series feedback ($R_{\mathrm{E}} \gg r_\pi$),

$$r_2 \approx r_{\mathrm{o}} + \beta r_{\mathrm{o}} = (\beta + 1) r_{\mathrm{o}}$$

is obtained, if $R_{\mathrm{B}} \ll r_\pi$ may be assumed. Owing to the series feedback, the output resistance increases at first in proportion to R_{E}, but its value saturates at $(\beta + 1) r_{\mathrm{o}}$ in the case of strong series feedback. This limit is proportional to the product of the current gain and the Early voltage $B_{\mathrm{N}} V_{\mathrm{AF}}$ (see Sect. 3.10), if the approximations $\beta + 1 \approx B_{\mathrm{N}}$ and $r_{\mathrm{o}} \approx V_{\mathrm{AF}}/I_{\mathrm{C}}$ are used. The bipolar-transistor current source differs from the corresponding field-effect-transistor current source in that the latter does not show an upper limit for the output resistance as the series feedback increases: as no current flows into the input of the field-effect transistor, the limit $\beta \to \infty$ has to be used in (6.17) to describe the output resistance of a field-effect-transistor current source.

6.5.2 Current Mirror

A simple realization of a current source used in integrated circuits is the current mirror, shown in Fig. 6.15a. The eb diodes are connected in parallel and therefore show the same voltage V_{BE}. Transistor T_1 acts as a diode and carries collector current

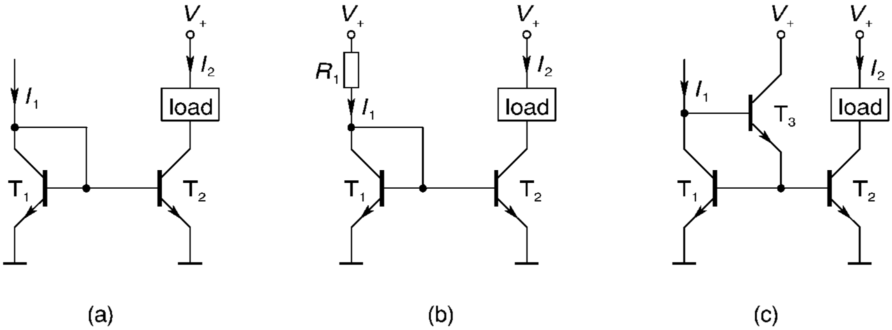


Fig. 6.15. Current mirror using bipolar transistors. (**a**) Principle; (**b**) elementary current source; (**c**) improved current mirror, which reduces the effect of base currents

$$I_{C1} = \frac{I_{S1}}{q_{B1}} \exp\left(\frac{V_{BE}}{V_T}\right) \qquad \text{if} \qquad V_{BE} \gg V_T \,.$$

The current I_1 that flows in the left branch of the current mirror is composed of I_{C1} and the base currents of both transistors:

$$I_1 = I_{C1} + I_{B1} + I_{B2} = \left(1+\frac{1}{B_{N1}}\right) I_{C1} + \frac{I_{C2}}{B_{N2}} \,.$$

As long as T_2 does not saturate, its collector current I_{C2} and therefore the current I_2 in the right branch of the current mirror is

$$I_2 = I_{C2} = \frac{I_{S2}}{q_{B2}} \exp\left(\frac{V_{BE}}{V_T}\right) \,.$$

The current ratio is therefore

$$\frac{I_1}{I_2} = \frac{I_{S1}}{I_{S2}} \frac{q_{B2}}{q_{B1}} \left(1+\frac{1}{B_{N1}}\right) + \frac{1}{B_{N2}} \approx \frac{I_{S1}}{I_{S2}} \,, \tag{6.18}$$

i.e., if the current gains B_{N1} and B_{N2} are large and if the Early effect is negligible, the current ratio is given by the ratio of the two transfer saturation currents. In integrated-circuit design, this ratio can be easily modified by the choice of the emitter area.[6] Current mirrors are only applicable if both transistors have good thermal coupling [13, 14], and are therefore only useful if both transistors are integrated on the same chip. In that case self-heating of transistor T_2 also heats T_1, resulting in a decreasing value of V_{BE} (since $I_1 \approx$ const.), which compensates for the effects of self-heating.

[6] If a current ratio different from one is required, parallel connections of identical transistors are often employed, since then all transistors are subject to the same process variations.

Figure 6.15b shows an elementary constant-current source. The current $I_1 = (V_+ - V_{\mathrm{BE}})/R_1$ through R_1 is "mirrored" to the right branch. Deviations of the mirror ratio from the ideal value $I_{\mathrm{S1}}/I_{\mathrm{S2}}$ are determined by the factor

$$\frac{q_{\mathrm{B2}}}{q_{\mathrm{B1}}}\left(1+\frac{1}{B_{\mathrm{N1}}}\right)+\frac{I_{\mathrm{S2}}}{I_{\mathrm{S1}}}\frac{1}{B_{\mathrm{N2}}}$$

In the elementary transistor model, the first contribution, which is caused by the Early effect, can be written as

$$\frac{q_{\mathrm{B2}}}{q_{\mathrm{B1}}} \approx \frac{V_{\mathrm{AF}}+V_{\mathrm{CE1}}}{V_{\mathrm{AF}}+V_{\mathrm{CE2}}} \approx 1-\frac{V_{\mathrm{CE2}}-V_{\mathrm{CE1}}}{V_{\mathrm{AF}}+V_{\mathrm{CE2}}} .$$

This value is small if transistors with a large Early voltage are used. The second term is caused by the base currents of the two transistors, and may be significantly reduced if the current source is modified as in Fig. 6.15c. In this circuit, the base current of transistors T_1 and T_2 is supplied by the emitter current of T_3, which requires a base current

$$I_{\mathrm{B3}} = \frac{I_{\mathrm{B1}}+I_{\mathrm{B2}}}{B_{\mathrm{N3}}+1} .$$

The only current that is "stolen" from I_1 and thus not mirrored to the right branch is this current. The effect of the base current on the current ratio is now of order $1/B_{\mathrm{N}}^2$ and is negligible for most practical purposes.

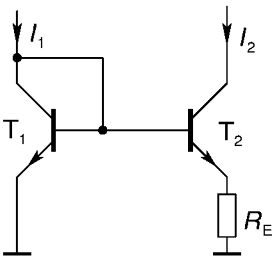


Fig. 6.16. Current mirror for reduced output current

The output current of the current mirror can be reduced by insertion of a resistance R_{E} in series with the emitter of transistor T_2 (Fig. 6.16). This resistor reduces the voltage drop across the eb diode of transistor T_2 by $R_{\mathrm{E}}I_2$. If the Early effect is neglected, this yields a modified mirror ratio

$$\frac{I_1}{I_2} = \exp\left(-\frac{I_2 R_{\mathrm{E}}}{V_{\mathrm{T}}}\right) . \tag{6.19}$$

Owing to the temperature dependence of V_{T} in the argument of the exponential function, the mirror ratio is now temperature-dependent.

Current Mirrors with Increased Output Resistance

Current sources realized with elementary current mirrors show an output resistance equal to the output resistance of the transistor. The Wilson current mirror, shown in Fig. 6.17a, yields a substantially enhanced output resistance. Here the output current $I_2 = I_{C3}$ is fed back by the current mirror composed of T_2 and T_1; an increase of I_{C3} therefore causes an increase of I_R and therefore a decrease of $V_{BE1} + V_{BE2}$, which compensates the increase of I_{C3} to a certain extent.

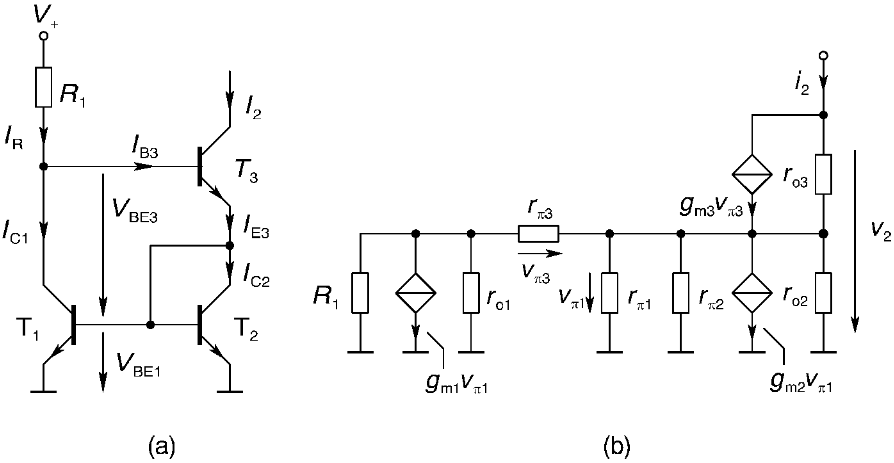


Fig. 6.17. Wilson current mirror. (**a**) Circuit, and (**b**) small-signal equivalent circuit

For an analysis of the operational principles, we shall assume that all transistors have an identical current gain B_N. The current I_R through R_1 is then composed of the collector current I_{C1} of T_1 and the base current I_{C3}/B_N of T_3:

$$I_R = I_{C1} + I_{C3}/B_N \, . \tag{6.20}$$

I_{C1} is mirrored from I_{C2}, i.e.

$$I_{C1} = (1+\delta)I_{C2} \, ,$$

where δ denotes the relative error of the mirror ratio. From Kirchhoff's current law applied to the emitter node of transistor T_3, we obtain, in addition,

$$I_{C3} = \frac{B_N}{B_N+1}(I_{C2}+I_{B1}+I_{B2}) \approx \frac{B_N}{B_N+1}\left(1+\frac{2}{B_N}\right)I_{C2} \, ,$$

which, together with (6.20) results in

$$I_R = I_{C3}\left(\frac{1}{B_N} + (1+\delta)\frac{B_N+1}{B_N+2}\right) \approx I_{C3}\left(1+\delta+\frac{2}{B_N^2}\right) .$$

The output current $I_2 = I_{C3}$ therefore deviates from the input current I_R only by a term of the order of $2/B_N^2$, in addition to the error δ caused by the incorrect mirror ratio.

The output resistance $r_2 = v_2/i_2$ of the Wilson current mirror can be calculated with the help of the small-signal equivalent circuit shown in Fig. 6.17b; this gives the approximate result

$$r_2 \approx r_o\beta/2 ,$$

where $\beta \gg 1$ and identical small-signal quantities for all transistors have been assumed. This assumption is reasonable, since all transistors carry approximately the same current.

6.5.3 Active Load

The voltage gain of an amplifier in the common-emitter configuration with a collector resistance R_C is limited by ($g_o = 0$, no series resistances)

$$|\,dV_2/dV_1\,| \approx g_m R_C \approx R_C I_C/V_T ,$$

and is therefore determined by the voltage drop across R_C at the bias point. Large values of voltage gain therefore require a large supply voltage V_+ together with a large value of R_C.

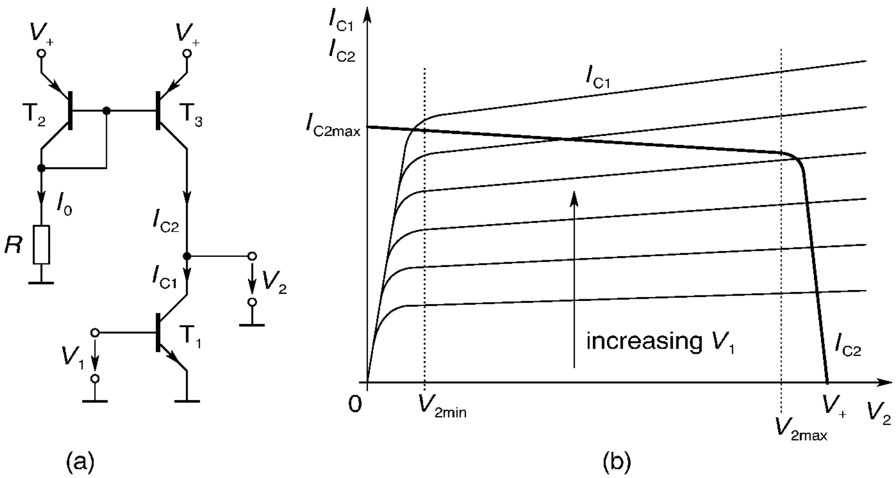


Fig. 6.18. Common-emitter amplifier stage with active load. (**a**) Circuit, and (**b**) graphical determination of bias point and transfer characteristic

Figure 6.18 shows an amplifier in the common-emitter configuration where the ohmic load element has been replaced by a pnp current mirror. This allows one to substantially increase the voltage gain achievable with a given supply voltage V_+. In the voltage range $V_{2\min} < v_2 < V_{2\max}$, the output characteristic of transistor T_2 is approximately linear, with a slope determined by r_o. To realize such a characteristic with a collector resistance $R_\mathrm{C} = r_\mathrm{o}$, a supply voltage $r_\mathrm{o} I_{\mathrm{C2max}} \gg V_+$ would be required. The voltage transfer factor of the unloaded amplifier is

$$\frac{\mathrm{d}V_2}{\mathrm{d}V_1} = -\frac{g_{\mathrm{m1}} r_{\mathrm{o2}}}{1 + r_{\mathrm{o2}}/r_{\mathrm{o1}}} .$$

Since

$$g_{\mathrm{m1}} \approx \frac{I_{\mathrm{C1}}}{V_\mathrm{T}} , \qquad r_{\mathrm{o1}} \approx \frac{V_{\mathrm{AF1}}}{I_{\mathrm{C1}}} , \qquad r_{\mathrm{o2}} \approx \frac{V_{\mathrm{AF2}}}{I_{\mathrm{C2}}} \qquad \text{and} \qquad I_{\mathrm{C1}} \approx I_{\mathrm{C2}} ,$$

this gives

$$\frac{\mathrm{d}V_2}{\mathrm{d}V_1} = -\frac{1/V_\mathrm{T}}{1/V_{\mathrm{AF1}} + 1/V_{\mathrm{AF2}}} . \tag{6.21}$$

With Early voltages V_{AF1} and V_{AF2} in the range of several tens of volts, voltage gains in the range of several thousand are possible.

6.6 Differential Amplifiers

Differential amplifiers are widely used in bipolar analog circuits and also form the basis of digital ECL circuits. Figure 6.19 shows an emitter-coupled differential amplifier.

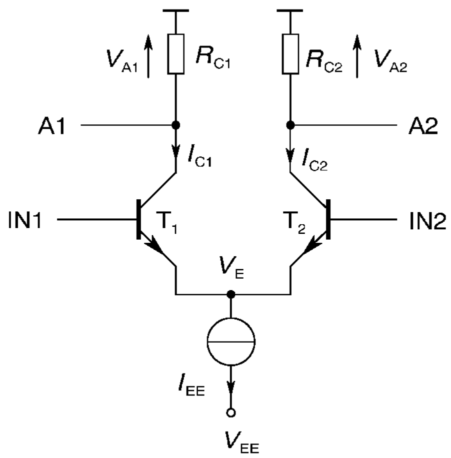


Fig. 6.19. Emitter-coupled differential amplifier

6.6.1 DC Transfer Characteristic

The output voltages V_{A1} and V_{A2} are determined by the voltage drops across the collector resistances $R_{C1} = R_{C2} = R_C$:

$$V_{A1} = -R_C I_{C1} \quad \text{and} \quad V_{A2} = -R_C I_{C2} \, . \tag{6.22}$$

If none of the transistors is saturated, I_{C1} and I_{C2} are predominantly determined by the voltages $V_{BE1} = V_{IN1} - V_E$ and $V_{BE2} = V_{IN2} - V_E$. Assuming identical transistors and neglecting the Early effect ($q_B = 1$) yields the following for the ratio of the two collector currents:

$$\frac{I_{C2}}{I_{C1}} = \exp\left(-\frac{V_{ID}}{V_T}\right) , \tag{6.23}$$

where $V_{ID} = V_{IN1} - V_{IN2}$ denotes the difference input voltage. If the current gain of the transistors is large, the collector current of each transistor approximately equals the corresponding emitter current. The emitter currents add up to the current I_{EE} delivered by the current source:

$$I_{EE} = I_{E1} + I_{E2} = \frac{I_{C1}}{A_N} + \frac{I_{C2}}{A_N} \approx I_{C1} + I_{C2} \, , \tag{6.24}$$

where the common-base current gains $A_N = B_N/(B_N + 1)$ of the two transistors have been taken to be equal and approximately unity. Solving the two equations for I_{C1} and I_{C2} gives

$$I_{C1} = \frac{A_N I_{EE}}{1 + \exp(-V_{ID}/V_T)} \quad \text{and} \quad I_{C2} = \frac{A_N I_{EE}}{1 + \exp(V_{ID}/V_T)} \, . \tag{6.25}$$

The output voltages V_{A1} and V_{A2} can be obtained from (6.22) and (6.25); this gives the difference output voltage $V_{AD} = V_{A2} - V_{A1}$ as follows:

$$V_{AD} = V_S \tanh\left(\frac{V_{ID}}{2V_T}\right) ,$$

where $V_S = A_N R_C I_{EE} \approx R_C I_{EE}$ denotes the maximum voltage swing of the amplifier.

Offset Voltage

The (input) offset voltage V_O of a differential amplifier is the difference input voltage that is required to obtain zero difference output voltage. A nonzero offset voltage may be caused by variations of the characteristics of the load as well as by variations of the transistor parameters of the differential pair. The difference output voltage is zero if the condition

$$R_{C1} I_{S1} \exp\left(\frac{V_{BE1}}{V_T}\right) = R_{C2} I_{S2} \exp\left(\frac{V_{BE2}}{V_T}\right)$$

is fulfilled. From this condition, the input offset voltage $V_O = V_{BE1} - V_{BE2}$ is obtained as

$$V_O = V_T \ln\left(\frac{R_{C2} I_{S2}}{R_{C1} I_{S1}}\right) \approx V_T \left(\frac{\Delta R_C}{R_C} + \frac{\Delta I_S}{I_S}\right) .$$

The approximation in the above equation is obtained by expanding the logarithm up to first order, and is valid under the assumption of small relative deviations, that is, if

$$\frac{R_{C2}}{R_{C1}} - 1 = \frac{\Delta R_C}{R_C} \ll 1 \quad \text{and} \quad \frac{I_{S2}}{I_{S1}} - 1 = \frac{\Delta I_S}{I_S} \ll 1 .$$

6.6.2 Differential-Mode and Common-Mode Voltage Gain

The differential-mode voltage gain, A_D, describes the response of the output voltages to a change of the difference input voltage:

$$A_D = -\frac{dV_{A1}}{dV_{ID}} = \frac{dV_{A2}}{dV_{ID}} = \frac{1}{2}\frac{dV_{AD}}{dV_{ID}} = \frac{V_S}{4V_T} .$$

If I_{EE} is determined by an ideal current source, the common-mode voltage gain is zero, i.e. the output voltages V_{A1} and V_{A2} remain unaffected if V_{IN1} and V_{IN2} change by the same amount. In practice, however, I_{EE} depends somewhat on V_E and therefore on the common-mode input voltage $V_{CM} = (V_{IN1} + V_{IN2})/2$. In common mode, $V_{IN1} = V_{IN2} = V_{CM}$; if V_{CM} is applied to both inputs of the differential amplifier, we have $V_{BE1} = V_{BE2} = V_{BE}$, and therefore $V_{CM} = V_{BE} + V_E$. Since the output resistance of the current source $r_{ee} = dV_E/dI_{EE}$ is large in comparison with $1/g_m$, we obtain $dV_E/dV_{CM} \approx 1$, i.e. the voltage drop across the diodes is approximately constant. Since $V_{A1} = V_{A2} \approx -R_C I_{EE}/2$, we obtain the following for the common-mode voltage gain:

$$A_{CM} = -\frac{dV_{A1}}{dV_{CM}} = -\frac{dV_{A1}}{dI_{EE}}\frac{dI_{EE}}{dV_E}\frac{dV_E}{dV_{GL}} \approx \frac{R_C}{2r_{ee}} . \tag{6.26}$$

The common-mode rejection ratio CMRR is therefore given by

$$\text{CMRR} = \frac{A_D}{A_{CM}} \approx \frac{R_C I_{EE}}{4V_T}\frac{2r_{ee}}{R_C} = \frac{r_{ee} I_{EE}}{2V_T} ; \tag{6.27}$$

its value increases in proportion to the output resistance of the current source.

Small-Signal Analysis

Figure 6.20 shows a small-signal equivalent circuit of the differential amplifier. There, the output resistance r_{ee} of the current source is split into two parallel resistances of value $2r_{ee}$ to illustrate the symmetry of the circuit. Application

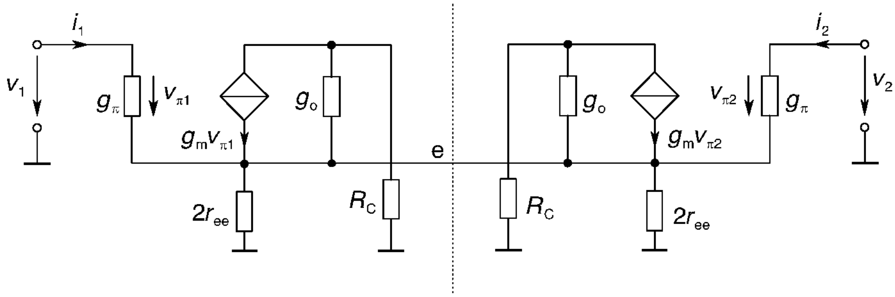


Fig. 6.20. Low-frequency small-signal equivalent circuit of a differential amplifier

of the superposition theorem now allows us to investigate differential- and common-mode operation separately.

In differential-mode operation, the potential of node e remains unchanged under these conditions, i.e., in a small-signal analysis, node e is at ground potential. For the determination of the differential-mode voltage gain, it is therefore sufficient to investigate the subcircuit shown in Fig. 6.21a. From

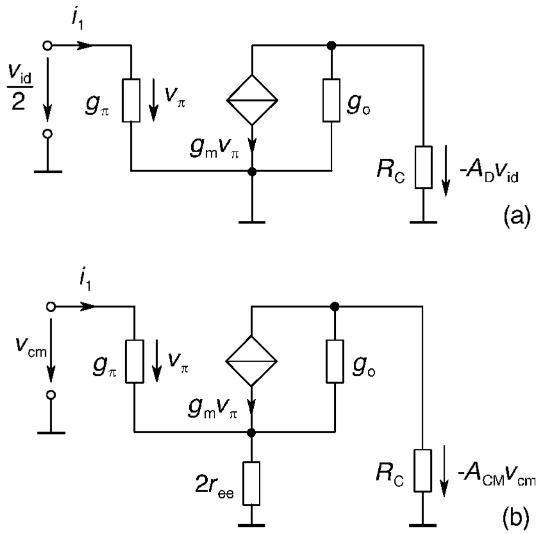


Fig. 6.21. Small-signal equivalent circuit of an differential amplifier for (**a**) differential mode and (**b**) common mode

this subcircuit, the differential-mode voltage gain can be obtained as follows:

$$A_\mathrm{D} = \frac{g_\mathrm{m}}{2}\left(R_\mathrm{C} \,\|\, r_\mathrm{o}\right) \quad \text{where} \quad g_\mathrm{m} \approx \frac{1}{V_\mathrm{T}}\frac{I_\mathrm{EE}}{2} .$$

Apart from the factor 1/2, the value of this gain equals the value obtained for an amplifier in the common-emitter configuration without feedback.

In common-mode operation, no current crosses the symmetry axis shown in Fig. 6.20 – it is therefore sufficient to investigate the half-circuit shown in Fig. 6.21b. This half-circuit corresponds to an amplifier in the common-emitter configuration with strong series feedback. The common-mode voltage gain is therefore

$$A_{\mathrm{CM}} = \frac{R_{\mathrm{C}}}{2r_{\mathrm{ee}}} .$$

This symmetry consideration can also be applied under conditions of high-frequency operation, when parasitic transistor capacitances have to be considered.

6.7 Analog Multipliers

An analog multiplier provides an output signal that is proportional to the product of its input signals. A simple example can be derived from the basic

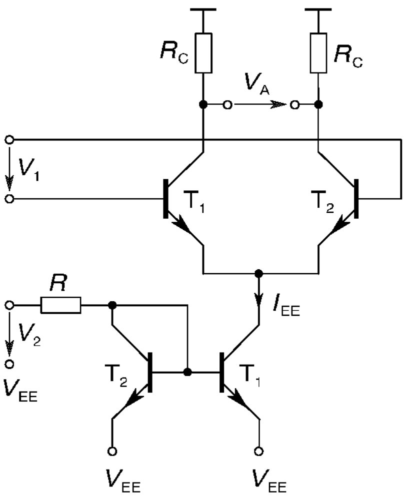


Fig. 6.22. Analog multiplier (two-quadrant) derived from the differential amplifier

differential amplifier. Under small-signal conditions ($V_1 \ll V_{\mathrm{T}}$), the difference output voltage of the amplifier was found to be

$$v_{\mathrm{ad}} = \frac{R_{\mathrm{C}} I_{\mathrm{EE}}}{2V_{\mathrm{T}}} v_{\mathrm{id}} . \tag{6.28}$$

If I_{EE} is to be chosen proportional to another signal V_2, as in the circuit depicted in Fig. 6.22, where $I_{\mathrm{EE}} = \Gamma(V_2 - V_{\mathrm{BEon}})/R$ if Γ denotes the current mirror ratio, (6.28) becomes

$$v_{\mathrm{ad}} = \frac{\Gamma R_{\mathrm{S}}}{2RV_{\mathrm{T}}}(V_2 - V_{\mathrm{BEon}})v_1 \ . \tag{6.29}$$

Obviously, v_{ad} is proportional to the product of v_1 and V_2 if $V_2 > V_{\mathrm{BEon}}$. Since the circuit does not work for negative values of V_2, it is termed a two-quadrant multiplier (it functions only in two quadrants of the (V_1, V_2) plane). This circuit may also be considered as a variable-gain small-signal voltage amplifier, with a voltage gain that is controlled by V_2.

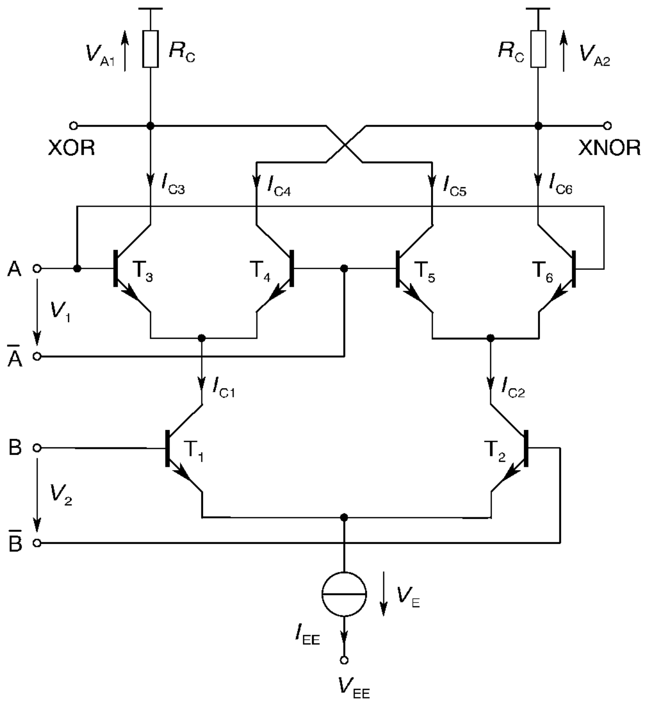


Fig. 6.23. Four-quadrant analog multiplier (Gilbert cell). This circuit can also be employed to realize the XOR and XNOR functions in differential ECL with series gating

A circuit that allows four-quadrant multiplication is the Gilbert cell, shown in Fig. 6.23. It consists of three differential stages with total currents I_{EE}, I_{C1} and I_{C2}, respectively. In the limit $A_{\mathrm{N}} \to 1$, we obtain from (6.25)

$$I_{\mathrm{C1}} = \frac{I_{\mathrm{EE}}}{1+\exp(-V_2/V_{\mathrm{T}})} \ , \qquad I_{\mathrm{C2}} = \frac{I_{\mathrm{EE}}}{1+\exp(V_2/V_{\mathrm{T}})} \ , \tag{6.30}$$

and also

$$I_{\mathrm{C3}} = \frac{I_{\mathrm{C1}}}{1+\exp(-V_1/V_{\mathrm{T}})} \ , \qquad I_{\mathrm{C4}} = \frac{I_{\mathrm{C1}}}{1+\exp(V_1/V_{\mathrm{T}})} \tag{6.31}$$

and

$$I_{\mathrm{C5}} = \frac{I_{\mathrm{C2}}}{1+\exp(V_1/V_{\mathrm{T}})} \ , \qquad I_{\mathrm{C6}} = \frac{I_{\mathrm{C2}}}{1+\exp(-V_1/V_{\mathrm{T}})} \ . \tag{6.32}$$

Combining these equations, we obtain the output voltage difference

$$V_{\mathrm{A}} = V_{\mathrm{A2}} - V_{\mathrm{A1}} = R_{\mathrm{C}} \left(I_{\mathrm{C3}} + I_{\mathrm{C5}} - I_{\mathrm{C4}} - I_{\mathrm{C6}} \right) \tag{6.33}$$

which results in

$$V_{\mathrm{A}} = R_{\mathrm{C}} I_{\mathrm{EE}} \tanh\left(\frac{V_1}{2V_{\mathrm{T}}}\right) \tanh\left(\frac{V_2}{2V_{\mathrm{T}}}\right) . \tag{6.34}$$

The output voltage is therefore proportional to the product of the hyperbolic tangents of the input functions. Under small-signal conditions, (6.34) simplifies to

$$V_{\mathrm{A}} = R_{\mathrm{C}} I_{\mathrm{EE}} \frac{V_1}{2V_{\mathrm{T}}} \frac{V_2}{2V_{\mathrm{T}}} . \tag{6.35}$$

6.8 Two-Transistor Amplifier Stages

This section considers two-transistor amplifier stages: the Darlington configuration, which is widely used to improve the current gain and input resistance, and the cascode configuration, which is used to improve the output resistance and frequency response.

6.8.1 The Darlington Configuration

The Darlington configuration (Fig. 6.24a) is useful when a small current at the input is required to drive a large current at the output. The Darlington

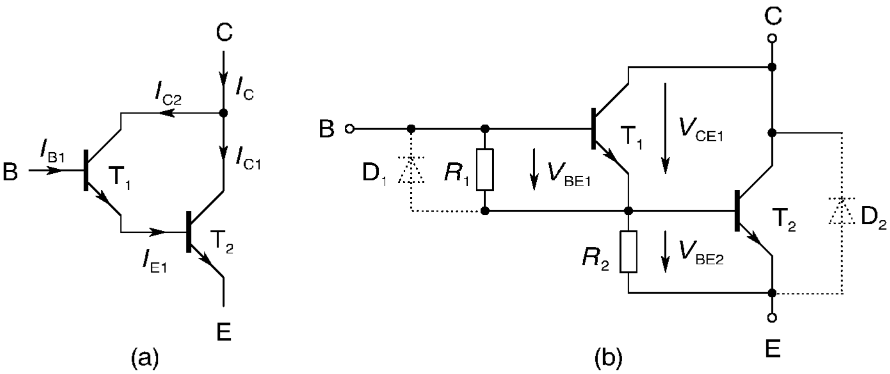


Fig. 6.24. (**a**) Elementary Darlington transistor pair, (**b**) modified Darlington configuration

pair of transistors acts like a single transistor with base current $I_{\mathrm{B}} = I_{\mathrm{B1}}$ and collector current $I_{\mathrm{C}} = I_{\mathrm{C1}} + I_{\mathrm{C2}}$. Since the base current I_{B2} of T_2 equals the emitter current $I_{\mathrm{E1}} = (B_{\mathrm{N1}}+1) I_{\mathrm{B1}}$ of T_1, the collector current is given by

$$I_C = B_{N1} I_{B1} + B_{N2}(B_{N1}+1) I_{B1} \approx B_{N1} B_{N2} I_{B1} \,, \tag{6.36}$$

i.e. the current gain of the Darlington configuration approximately equals the product of the current gains of the individual transistors. The voltage required at the input of the Darlington pair is the sum of V_{BE1} and V_{BE2}, and therefore approximately twice the value for an individual transistor. Another difference has to be considered if the Darlington pair is applied in switching circuits: since only transistor T_1 may go into saturation and since $V_{CE1} > 0$, the condition $V_{CB2} > 0$ results, and the saturation voltage across the Darlington pair is larger than V_{BE2}.

Application of the Darlington configuration as a switch is problematic – particularly during turn off – since T_1 cannot sink charges stored in T_2. Therefore additional circuit elements are added to the Darlington pair, as shown in Fig. 6.24b. There, the diode connected antiparallel to the eb diode allows a current to flow out of the base terminal of T_2 if the eb voltage of the Darlington becomes negative. The ce leakage current of the Darlington pair may become unacceptable especially at elevated temperatures as the leakage current of T_1 is amplified by the current gain B_{N2} of T_2. This problem may be circumvented by placing resistors R_1 and R_2 parallel to the eb diodes of the transistors. The resistance values are chosen such that the leakage currents of the individual transistors cause a voltage drop that lies well below the diode voltage. In this case, the current generated in the bc diode is sunk by the resistor and does not cause a substantial forward bias of the eb diode, thus keeping the leakage current small.

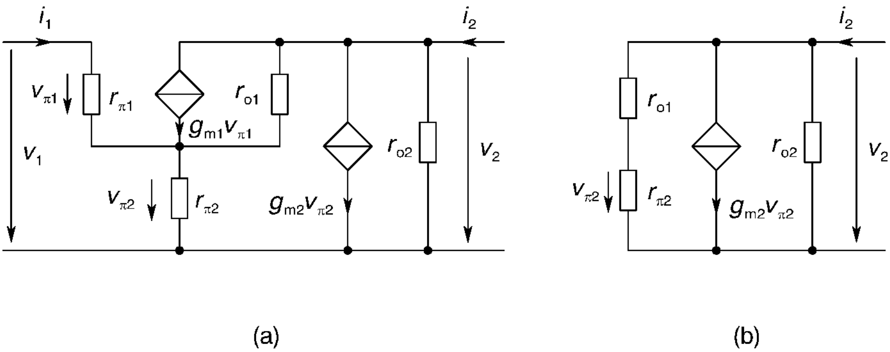


Fig. 6.25. (**a**) Small-signal equivalent circuit of Darlington transistor pair, (**b**) simplified equivalent circuit used for the computation of the output resistance

The input resistance $r_1 = v_1/i_1$ of the Darlington pair can be obtained from the small-signal equivalent circuit shown in Fig. 6.25a. Neglecting r_{o1} for simplicity, we obtain

$$v_1 = v_{\pi1} + v_{\pi2} = r_{\pi1} i_1 + r_{\pi2}(\beta_1+1) i_1 \,,$$

and therefore, if $\beta_1 \gg 1$,

$$r_1 \approx r_{\pi 1} + \beta_1 r_{\pi 2} \approx r_{\pi 1}(1 + \beta_1/\beta_2) \approx 2r_{\pi 1} , \tag{6.37}$$

since $r_{\pi 2} \approx r_{\pi 1}/\beta_2$. This is because the input resistances of the individual transistors vary in inverse proportion to the respective collector current values.

The output resistance $r_2 = v_2/i_2$ can be determined from the small-signal equivalent circuit shown in Fig. 6.25b. Owing to the open input, the condition $v_{\pi 1} = 0$ applies and allows us to remove $r_{\pi 1}$ and the transfer-current source of transistor $\mathrm{T_1}$, which is controlled by $v_{\pi 1}$ from the equivalent circuit. The output current is given by

$$v_{\pi 2} = \frac{r_{\pi 2}}{r_{\mathrm{o}1} + r_{\pi 2}} v_2$$

and, using the fact that $r_{\pi 2} g_{\mathrm{m}2} = \beta_2$,

$$i_2 = \frac{v_2}{r_{\mathrm{o}2}} + \frac{\beta_2}{r_{\mathrm{o}1} + r_{\pi 2}} v_2 + \frac{v_2}{r_{\mathrm{o}1} + r_{\pi 2}} .$$

The output resistance r_2 of the Darlington pair is therefore

$$r_2 = \frac{v_2}{i_2} = \left(\frac{1}{r_{\mathrm{o}2}} + \frac{\beta_2 + 1}{r_{\mathrm{o}1} + r_{\pi 2}} \right)^{-1} \approx 2r_{\mathrm{o}2} , \tag{6.38}$$

assuming $r_{\mathrm{o}1} \approx \beta_2 r_{\mathrm{o}2}$. Since the output resistances of the individual transistors vary in inverse proportion to the collector currents, which differ by the value of the current gain $B_{\mathrm{N}2} \approx \beta_2$, the latter assumption is approximately valid, if equal values of the Early voltage are assumed.

6.8.2 The Cascode Configuration

The cascode configuration combines an amplifier stage in common-emitter configuration with an amplifier stage in common-base configuration (Fig. 6.26). This circuit configuration acts like a single transistor with emitter $\mathrm{E_1}$,

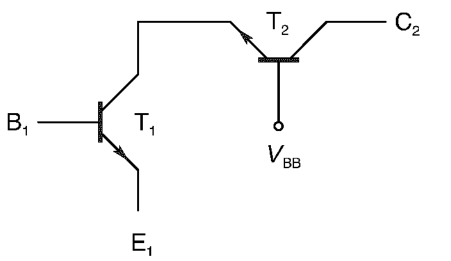


Fig. 6.26. Cascode configuration

base $\mathrm{B_1}$ and collector $\mathrm{C_2}$; it has a very high output resistance and shows virtually no high-frequency feedback from the output back to the input through the bc capacitance (no Miller effect). Combination of the cascode configuration with its high output resistance with an active pnp load allows to realize single amplifying stages with a very high voltage gain. The ac behavior of

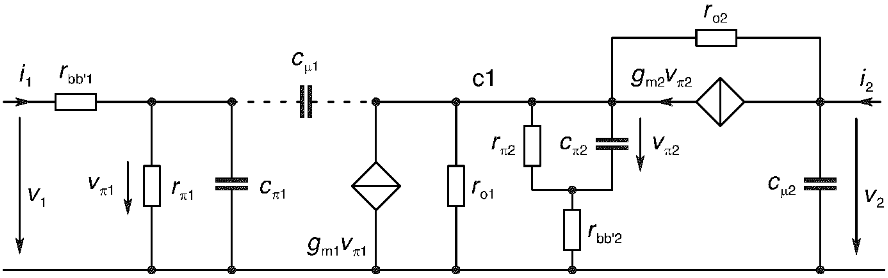


Fig. 6.27. Small-signal equivalent circuit of the cascode configuration

the cascode configuration is represented by the small-signal equivalent circuit depicted in Fig. 6.27. Owing to the very low input impedance of the common-base configuration, the voltage swing of node c1 will be of the same order of magnitude as the input voltage swing. If $c_{\pi 1} \gg c_{\mu 1}$ it is possible to neglect $c_{\mu 1}$ in the analysis of the frequency response; $v_{\pi 1}/v_1$ will then show a low-pass characteristic with a cutoff frequency approximately equal to the transconductance cutoff frequency f_{y} independent of the load connected to the output of the cascode stage.

6.9 Bandgap References

A stable reference voltage is required in many electronic circuits. For that purpose, temperature-compensated Zener diodes can be used, in principle. However, Zener diodes are not suited for integrated circuits, since Zener diodes require voltage drops in excess of typical supply voltages and impose additional technological constraints. Therefore a special class of circuits, called bandgap references, which derive a reference voltage from the temperature dependence of the diode voltage, are frequently employed in IC design for the generation of reference voltages. In such bandgap references, a correction voltage $V_{\mathrm{corr}}(T)$ is added to $V_{\mathrm{BE}}(T)$ in order to compensate for at least the first-order temperature dependence of $V_{\mathrm{BE}}(T)$. As $V_{\mathrm{BE}}(T)$ shows an approximately linear decrease with increasing temperature, $V_{\mathrm{corr}}(T)$ should increase in proportion to the absolute temperature T. Such a correction voltage can be obtained from the difference $\Delta V_{\mathrm{BE}} = V_{\mathrm{T}} \ln(J_{\mathrm{C1}}/J_{\mathrm{C2}}) \sim T$ between the base–emitter voltages of two transistors operated at different collector current densities.

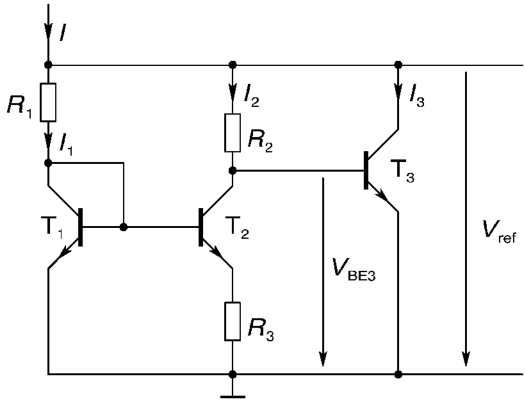


Fig. 6.28. Bandgap voltage reference due to Widlar [15]

Figure 6.28 shows the bandgap voltage reference suggested by Widlar [15]. The output voltage V_{ref} of this circuit is determined by the voltage drop V_{BE3} across the eb diode of transistor T_3 and the voltage drop across R_2:

$$V_{ref} = V_{BE3} + R_2 I_2 .$$

For an illustration of the idea behind this circuit, we shall assume the base currents to be negligible. Since T_1 and T_2 form a current mirror,

$$R_3 I_2 = V_T \ln\left(\frac{I_1}{I_2}\right) , \tag{6.39}$$

and thus

$$V_{ref} = V_{BE3} + V_T \frac{R_2}{R_3} \ln\left(\frac{I_1}{I_2}\right) , \tag{6.40}$$

i.e. V_{ref} is composed of $V_{BE3}(T)$ and a voltage that varies in proportion to the absolute temperature T, if the ratios R_2/R_3 and I_1/I_2 are independent of temperature. The reference voltage becomes temperature-independent (to first order) if the condition

$$\frac{k_B}{e}\frac{R_2}{R_3} \ln\left(\frac{I_1}{I_2}\right) = -\frac{dV_{BE3}}{dT} \approx -\frac{V_{BE3} - V_g - X_{TI} V_T}{T} \tag{6.41}$$

is fulfilled (I_{C3} is assumed to be constant). Combinination of (6.39) and (6.41) yields the reference voltage

$$V_{ref} = V_{BE3} - T\frac{dV_{BE3}}{dT} \approx V_g + X_{TI} V_T ;$$

its value is close to the bandgap voltage V_g, and therefore, in the case of silicon $V_{ref} \approx 1.3$ V.

Figure 6.29 shows another bandgap reference, which determines the reference voltage V_{ref} as the sum of the base–emitter voltage $V_{BE1}(T)$ of transistor

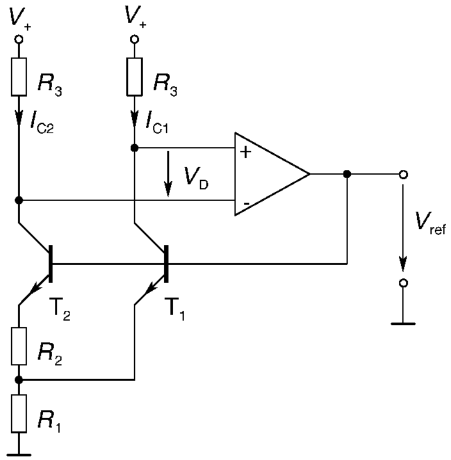


Fig. 6.29. Alternative realization of bandgap voltage reference [16]

T_1 and the voltage drop across resistor R_1. Assuming an ideal differential amplifier, a bias point where $V_D = 0$ and thus $I_{C1} = I_{C2} = I_C$ is obtained, since T_1 and T_2 have the same collector resistance. If the saturation currents of the two transistors are different, the relation $V_{BE1} - V_{BE2} = V_T \ln(K) = R_2 I_C$ applies, where $K = I_{S2}/I_{S1}$. Since K is approximately temperature-independent, the voltage difference $V_{BE1} - V_{BE2}$ varies in proportion to T, and so does the voltage drop across R_1:

$$V_1 = 2R_1 I_C = 2\frac{R_1}{R_2} V_T \ln(K) \sim T .$$

The increase of V_1 with temperature is used to compensate for the decrease of V_{BE1} with temperature.

The concepts considered so far provide only a first-order correction of the temperature-dependent base–emitter voltage; if second-order effects are taken into the account, a curvature-corrected bandgap reference results. The non-linear temperature dependence of $V_{BE}(T)$ when $I_C = \text{const.}$ can be obtained from the relations $V_{BE}(T) = V_T \ln[I_C/I_S(T)]$ and

$$I_S(T) = I_S(T_0)\left(\frac{T}{T_0}\right)^{X_{TI}} \exp\left(\frac{W_g(T_0)}{k_B T_0} - \frac{W_g(T)}{k_B T}\right)$$

(see section 3.11 and [17]), resulting in

$$V_{BE}(T) = \frac{T}{T_0} V_{BE}(T_0) - X_{TI} V_T \ln\left(\frac{T}{T_0}\right) + \frac{1}{e}\left(W_g(T) - W_g(T_0)\frac{T}{T_0}\right)$$

if I_C is constant. If we chose $T_0 = 300\,\text{K}$, we obtain, up to second order of the temperature difference $\Delta T = T - T_0$

$$V_{BE}(T) = V_{BE}(300\,\text{K}) + \Delta T + b\,(\Delta T)^2 , \tag{6.42}$$

where

$$a = \left(\frac{\partial V_{\mathrm{BE}}}{\partial T}\right)_{I_{\mathrm{C}}} = \frac{V_{\mathrm{BE}}(T_0) - V_{\mathrm{g}}(T_0) - X_{\mathrm{TI}} V_{\mathrm{T0}}}{T_0}$$

and

$$b = -\frac{1}{2}\left(\frac{X_{\mathrm{TI}} V_{\mathrm{T0}}}{T_0^2} - \frac{1}{e}\left.\frac{\mathrm{d}^2 W_{\mathrm{g}}}{\mathrm{d}T^2}\right|_{T_0}\right)$$

$$= (2.333 + 1.435\, X_{\mathrm{TI}}) \times 10^{-4}\, \frac{\mathrm{mV}}{\mathrm{K}^2} \, .$$

There are many ways to compensate for the second-order nonlinearity of the $V_{\mathrm{BE}}(T)$ characteristic, such as the generation of a correction voltage that varies in proportion to the square of the absolute temperature, or exploitation of a temperature-dependent current gain. A widely used method exploits the difference between the curvature of the base–emitter voltage $V_{\mathrm{BE1}}(T)$ of a transistor T_1 biased with a constant current and the base emitter voltage $V_{\mathrm{BE2}}(T)$ of a transistor T_2 biased with a current that varies in proportion to the absolute temperature. To second order in ΔT, this gives

$$V_{\mathrm{BE2}}(T) = V_{\mathrm{BE2}}(T_0) + \left[\left(\frac{\partial V_{\mathrm{BE2}}}{\partial T}\right)_{I_{\mathrm{C2}}} + \frac{V_{\mathrm{T0}}}{T_0}\right] \Delta T + b'(\Delta T)^2 \, ,$$

where

$$b' = [2.333 + 1.435(X_{\mathrm{TI}} - 1)] \times 10^{-4}\, \mathrm{mV/K}^2 \, .$$

The difference $V_{\mathrm{BE1}}(T) - (b/b')V_{\mathrm{BE2}}(T)$ has no second-order temperature dependence. A low-voltage bandgap reference that compensates for the second-order nonlinearities in V_{BE}, as well as those introduced by the temperature drift of the resistors, has been presented in [18]. For further examples of curvature-corrected bandgap references, see [18–20]. Bandgap references can also be realized with SiGe HBTs [21], although substantially reduced voltage stability may result.

6.10 Digital Circuits

Bipolar transistors may be operated as switches and are therefore suited for the realization of digital circuits. We shall briefly review the most important approaches. However, these have mostly become obsolete owing to advances in integrated digital CMOS circuits: only ECL circuit techniques are still of interest for high-speed logic circuits. This appplication is described in more detail in Chap. 8.

6.10.1 Characteristics of Digital Circuits

This subsection explains important electrical characteristics of digital circuit technologies, such as noise margin, gate delay and power–delay product.

Voltage Transfer Characteristic, Noise Margin

In the following, we shall assume that the binary variables are represented by different voltages V_H and V_L, and attribute the HI state (logic 1) to the more positive voltage (this is called positive logic). The voltage transfer characteristic of an inverter with input voltage V_1 and output voltage V_2 in this case has the form depicted in Fig. 6.30. The voltage difference $V_S = V_H - V_L$ is called the logic swing. The unity gain points (UGPs) are defined as those points

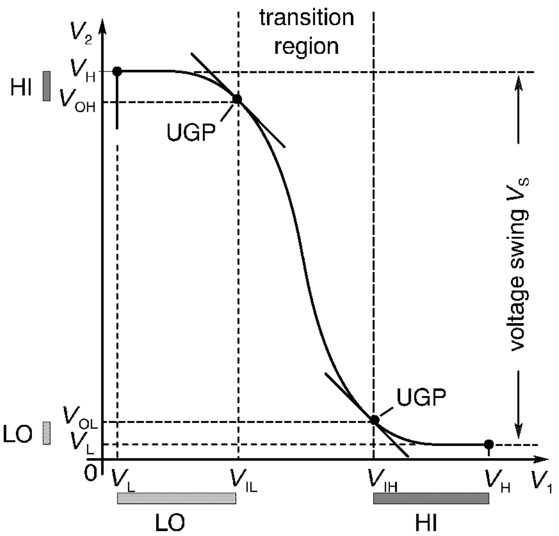


Fig. 6.30. Voltage transfer characteristic of an inverter circuit

on the voltage transfer characteristic where the (small-signal) voltage gain is unity. The transition region, which lies between these points, is characterized by a voltage gain larger than one. A nonlinear voltage transfer characteristic, with a voltage gain that exceeds one over a certain range of input voltages, is required for the regeneration of digital signal levels. Such regeneration is necessary to guarantee that errors do not accumulate if a signal propagates through a series of gates. The effect of a nonlinear voltage transfer characteristic is illustrated in Fig. 6.30: wide spans of input voltages V_1 within the intervals $[V_L, V_{IL}]$ (LO) and $[V_{IH}, V_H]$ (HI) cause output voltages V_2 within the comparably narrow intervals $[V_{OH}, V_H]$ (HI) and $[V_L, V_{OL}]$ (LO).

The immunity of a digital circuit technology with respect to noise voltages is characterized in terms of its noise margin. For the definition of the noise margin, a flip-flop is generally considered [22–26]; this is formed by feedback of the output of an inverter (INV1) to its input via a second inverter (INV2), as illustrated in Fig. 6.31. This flip-flop has two stable points of operation, which are determined by the intercept points of the voltage transfer characteristic and its inverse, obtained by mirroring the voltage transfer characteristic in the bisector of the (V_1-V_2) plane.

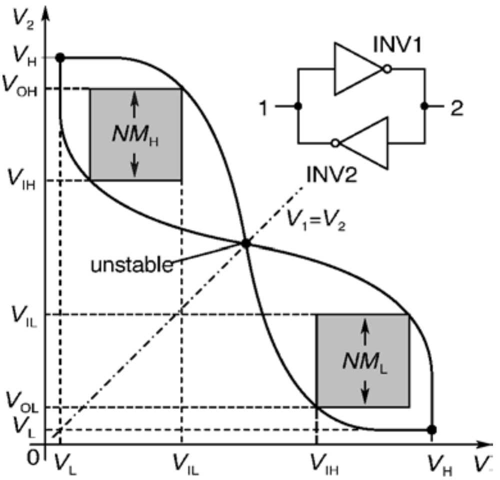


Fig. 6.31. Voltage transfer characteristic of an inverter circuit and definition of static noise margin

A noise voltage ΔV applied to the input of the inverters will shift the third intercept point, which represents an unstable point of operation, to V_L (if $\Delta V < 0$) or V_H (if $\Delta V > 0$). If the noise voltage exceeds a certain value, only one intercept point will remain. In this case the flip-flop no longer operates as a bistable element. The noise voltage required for this to occur is called the static noise margin,[7] and may be determined as the side length of a square within the voltage transfer characteristic and its inverse, resulting in the LO-level and HI-level noise margins

$$NM_L = V_{IL} - V_{OL} \qquad \text{and} \qquad NM_H = V_{OH} - V_{IH} . \tag{6.43}$$

If the two values are different, the smaller one determines the noise immunity of the circuit technology since the sign of the noise voltage is generally not known a priori. The amplitude of a noise signal may exceed the static noise margin if the noise is applied in pulsed form. The maximum amplitude of a

[7] The situation considered here is an idealization, since the voltage transfer characteristics of practical circuit technologies show statistical deviations. In order to characterize noise immunity, therefore, worst-case characteristics have to be considered.

noise pulse with a long duration must not, however, deviate very much from the static noise margin, owing to the quasistatic conditions. The dynamic noise margin converges to the static value in the limit of long pulse width; its value increases for shorter pulse widths.

Transition and Delay Times

The dynamic performance of logic circuits is characterized in terms of specific transition and delay times (Fig. 6.32). The rise time t_r and fall time t_f are

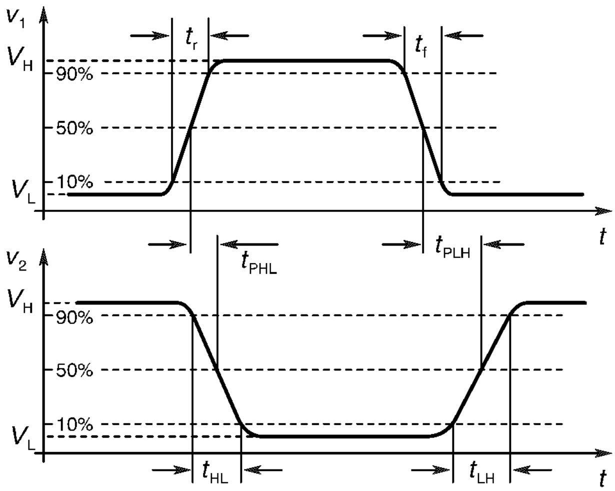


Fig. 6.32. Transition and delay time definitions

defined between the 10% and 90% points of the input signal v_1, and the HI–LO transition time t_HL and LO–HI transition time t_LH are defined between the 10% and 90% points of the output signal v_2. The propagation delay times t_PHL and t_PLH from input to output are defined between the 50%-points of the input and output signals. The gate delay τ_d is generally defined as the average propagation delay $\tau_\mathrm{d} = (t_\mathrm{PHL} + t_\mathrm{PLH})/2$. These quantities depend substantially on the signal source and on the load impedance and should therefore be considered with caution if used for a comparison of different technologies.

Power Consumption and Power–Delay Product

The average power consumption P is composed of a static portion P_0, and a dynamic portion P_1 associated with the switching transient. The static power dissipation P_0 is the average of the power dissipated in both logical states; the dynamic portion P_1 increases in proportion to the switching frequency f. The average power consumption P of bipolar digital circuits is dominated by P_0, in contrast to CMOS circuits, where the dynamic contribution P_1 is dominant.

The switching speed, which is generally characterized in terms of the gate delay time τ_d, increases with the power consumed by the gate. A figure of merit that represents this trade-off is the power–delay product. The power–delay product is frequently determined from ring oscillator measurements, since gate delays in the picosecond range are difficult to determine directly. A ring oscillator[8] is formed from a series of $(2n+1)$ inverters, with the output of the last one fed back to the input of the first, as shown in Fig. 6.33. Such circuits perform self-sustained oscillations with a time period $2(2n+1)\tau_d$; if $n \gg 1$, the signal frequency is easily measured.

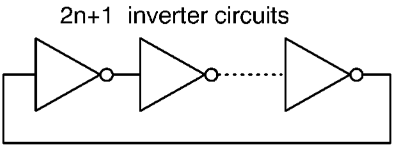


Fig. 6.33. Ring oscillator

The power–delay product depends on the switching frequency and thus on the number of inverter stages employed in the ring oscillator, if the power consumption shows a considerable frequency-dependent component. In this case, an appropriate figure of merit can be obtained if the gate delay τ_d is multiplied by the average power required by the gate to perform the switching transient; the power–delay product then corresponds to the energy per logic operation [27], required by that type of digital circuit.

6.10.2 Bipolar-Digital-Circuit Techniques

Various bipolar-digital-circuit techniques have been developed;[9] however, with the exception of the ECL technique, which still yields the fastest silicon digital integrated circuits, all these approaches have been rendered outdated by CMOS.

Direct-Coupled Transistor Logic and Resistor–Transistor Logic

Direct-coupled transistor logic (DCTL) is derived from the elementary bipolar-transistor inverter in the common-emitter configuration. By introducing additional transistors, one can easily extend this circuit to a NOR circuit.

[8]Ring oscillator circuits represent an idealized situation, since each inverter output is fed directly to the input of the following inverter, resulting in unrealistically small loading capacitances.

[9]See [28, 29] for a survey of earlier achievements in integrated bipolar circuits for digital applications.

In DCTL circuits, the output of one gate is directly applied to the bases of the

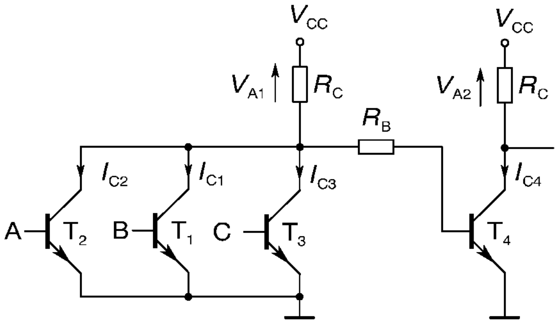


Fig. 6.34. OR gate realized in resistor–transistor logic (RTL) as a three-input NOR gate followed by an inverter

input transistors of subsequent gates. If more than one gate is connected to the output, process variations will cause an unequal distribution of the output current $(V_{CC} - V_{BEon})/R_C$ between the various inputs. This effect, which is known as current hogging, severely limits the fan-out capability of DCTL. In resistor–transistor logic (RTL), additional base resistances R_B (Fig. 6.34) are therefore introduced to reduce this effect. This circuit technique is very simple but suffers from several drawbacks: a logic gate with its output LO dissipates a static power $P \approx V_{CC}R_C$; furthermore, the switching speed is slow, owing to the unavoidable saturation of the switching transistors. In addition to this, the voltage transfer characteristic is far from ideal, with the consequence of a low noise margin [30].

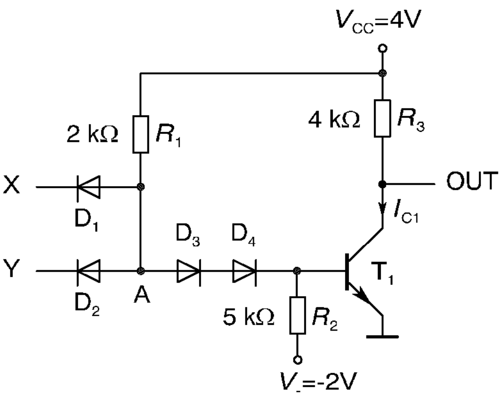


Fig. 6.35. NAND gate with a fan-in of two in diode–transistor logic (after [30])

Diode–Transistor Logic and Transistor–Transistor Logic

Figure 6.35 shows a two-input diode–transistor logic (DTL) NAND circuit, which has output voltage levels $V_H = V_{CC}$ and $V_L = V_{CEsat}$. If both X and Y are HI, no current will cross diodes D1 and D2 and the base potential of transistor T_1 will be V_{BEon}, resulting in a LO at the output. If, however, one of the input signals X or Y is LO, the potential at node A will be one diode voltage, and the base potential of T_1 will lie below zero, i.e. T_1 is switched off. Diode–transistor logic provides an output HI level that is independent of the number of load gates connected to it, and better noise margins than RTL. The extra supply voltage V_-, however, is a disadvantage, which was overcome with the introduction of transistor–transistor logic (TTL).

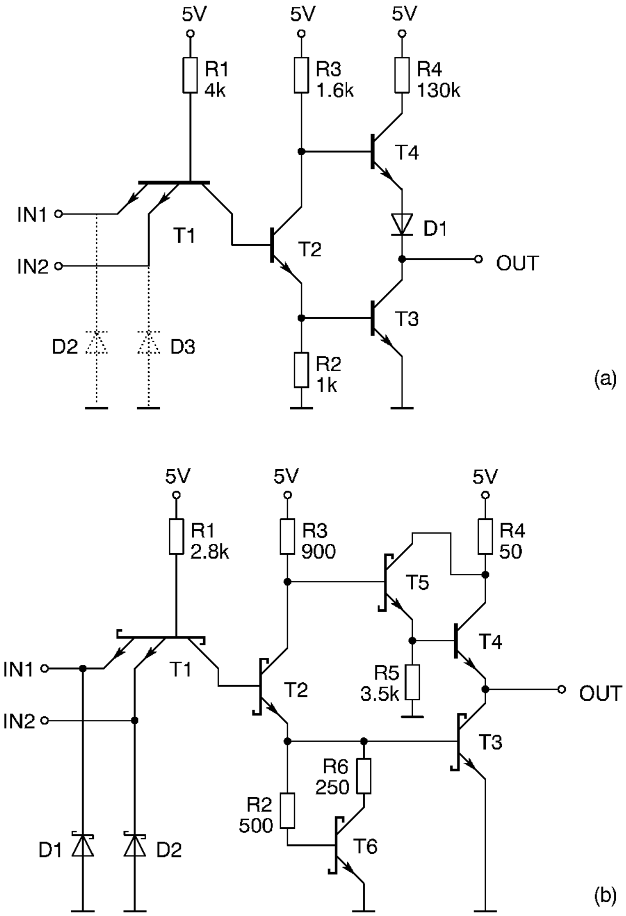


Fig. 6.36. TTL circuits: (**a**) standard TTL NAND, (**b**) Schottky TTL NAND gate (after [30])

Figure 6.36a shows a standard TTL NAND gate with a fan-in of two. The two inputs are connected to the emitters of a multi emitter transistor. If at least one of the inputs is LO, T_1 receives a base current via R_1 and goes into forward active mode. As the collector of T_1 is connected with the base of T_2, the collector current of T_1 is limited to a small value, and T_1 is driven into saturation and turns off T_2. Now no voltage occurs at R_3, i.e. the output is HI if at least one of the inputs is LO. If both inputs are HI, T_1 operates in reverse mode and delivers a base current to T_2: the output is driven to LO, as expected for a NAND gate. Transistors T_3 and T_4 serve as the output stage and improve the fan-out of the gate. Diodes D_2 and D_3, shown by broken lines, limit noise voltage spikes, for example due to inductances, during the switching transient.

Combination of TTL circuits is comparably simple, since interconnects are not terminated – however, this is at the cost of significant reflections. TTL circuits are generally operated with a 5 V bias and have a voltage swing of almost the same size, and are therefore rather insensitive to device tolerances and external noise signals. Some disadvantages of the large voltage swing are rather large switching times and noise generation.[10] If a LO appears at the output, the switching transistors are driven into saturation; the minority charge stored in the bc diode guarantees a very good dynamic noise margin, but implies a substantial delay during switch-off. Saturation of the switching transistors cannot be avoided in TTL circuits, resulting in gate delays of typically 50 ns.

To improve on this situation, Schottky TTL gates were developed, which employ a Schottky diode parallel to the bc diode. Figure 6.36b shows a schematic of a Schottky TTL NAND gate with a fan-in of two. In comparison with the circuit shown in Fig. 6.36a, all transistors that might be driven into saturation are realized as Schottky transistors, i.e. as vertical npn bipolar transistors with a Schottky diode parallel to the bc diode. A further modification concerns diode D_1 in Fig. 6.36a, which has been replaced by a Darlington pair; this yields a larger current at the output during a LO–HI transition and improves the rise time. An improved voltage transfer characteristic and therefore a better value of the static noise margin is achieved by the addition of transistor T_6 [30].

Integrated Injection Logic

Integrated injection logic (I2L) [31] (Fig. 6.37), sometimes also termed merged transistor logic (MTL) [32], improves on the bipolar digital-circuit techniques discussed above in terms of packing density and power consumption. In this technique, all emitters are held at ground potential, and an n^+

[10] Crosstalk due to capacitive coupling becomes large as dv/dt increases.

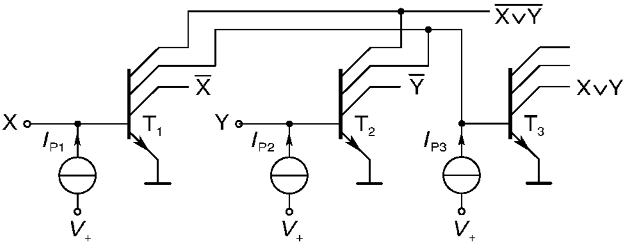


Fig. 6.37. Realization of OR/NOR function with I2L logic gates

substrate (or buried layer) may be used as the common emitter (Fig. 6.38). Since, in addition, no resistors are needed as load devices, I2L logic gates show excellent packing density. Owing to the reverse operation of the npn bipolar transistors, switching speed is poor, however. A study of I2L logic gates realized with SiGe HBTs has been presented in [33].

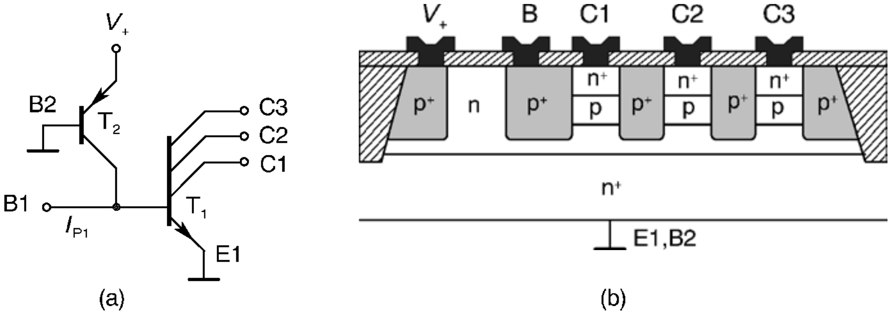


Fig. 6.38. Realization of I2L logic gate. (**a**) Circuit diagram, (**b**) schematic cross section (after [34]

Emitter-Coupled Logic

Emitter-coupled logic (ECL) is derived from the differential amplifier discussed in Sect. 6.7. If the differential input voltage swing is large in comparison with $2V_{\mathrm{T}}$, the differential amplifier operates as an inverter. The voltage drop across the collector series resistances therefore changes by $V_{\mathrm{S}} \approx R_{\mathrm{C}} I_{\mathrm{EE}}$ (i.e. $|\Delta V_{\mathrm{A}}| = 2V_{\mathrm{S}}$). This is exploited for the realization of fast digital integrated circuits: gate delays below 10 ps have already been realized (see Chap. 8).

Figure 6.39 shows how logic operations can be performed with differential amplifiers in nonlinear operation.[11] The first differential stage acts as an OR gate: the noninverting output is fed into the following gate, which acts as an

[11] In ECL circuits voltages are generally referenced to V_{CC}, which is taken as ground, i.e. V_{EE} is the (negative) supply voltage.

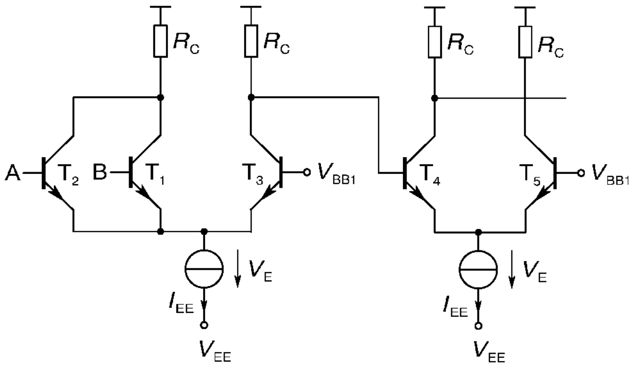


Fig. 6.39. Emitter-coupled logic circuit

inverter.[12]. Whether the current in the first differential stage is carried by T_3 or by the parallel connection of T_1 and T_2 is determined by the potentials v_A and v_B at the inputs A and B: if either v_A or v_B exceeds the fixed reference voltage V_{BB1} by several times V_T, T_3 carries no current, and $V_{A1} \approx 0$.

In Fig. 6.39, the potential at one input of the differential amplifier is kept at a fixed reference voltage, a technique commonly referred to as single-ended. An alternative approach is to use two complementary signals to feed both inputs of the differential amplifier (differential mode). The voltage swing V_S can be chosen to be as small as approximately 200 mV in differential mode. In single-ended mode, the voltage swing has to be considerably larger ($\geq$ 500 mV). If the output of the first gate is fed directly to the input of the second gate, as in Fig. 6.39, the bc diode of a current carrying transistor becomes forward-biased with the voltage swing, resulting in a substantial slowdown of switching speed. This can be avoided by shifting the level of the output signal with the help of an emitter follower.[13]. Saturation of the transistors can then be avoided, resulting in extremely fast digital circuits.

Since ECL circuits are generally operated with properly terminated lines, very small gate delays are achievable. Some disadvantages of ECL circuit technology are the comparably small packing density of ECL circuits, the increased effort required in circuit design, and a large static power dissipation, which limits the integration of ECL gates into a silicon chip.

[12]This part of the circuit is redundant, as the same result can be achieved by using the collector node of T_1 as the output: both the signal and its complement are available in ECL circuits.

[13]Sometimes a distinction is made between current-mode logic (CML), which uses no emitter follower and ECL, which uses emitter followers [35]. Owing to the large voltage swing required in single-ended circuits, CML should be applied only in differential mode.

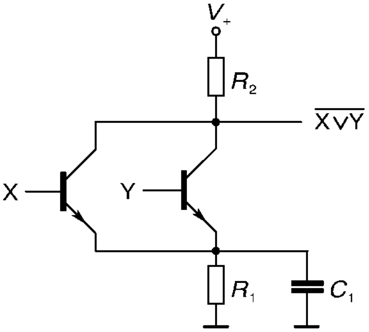


Fig. 6.40. Two-input NOR gate in nonthreshold logic

Nonthreshold Logic (NTL)

Figure 6.40 shows a basic NOR gate realized in nonthreshold logic (NTL). The two resistors have a ratio $R_2/R_1 \approx 1.5$, and R_1 is shunted with a speedup capacitance [36]. The circuit uses a temperature-dependent supply voltage in order to obtain a temperature-independent logic swing $V_S \approx R_2(V_+ - V_{BEon})/R_1$, which otherwise would increase with temperature owing to the decrease of V_{BEon}. In order to avoid strong saturation, V_S is chosen to be several hundred millivolts, resulting in a supply voltage $V_+ \approx 1$ V at $T = 300$ K. The gate delays and power–delay product of circuits that use this technique may be superior to those achieved with ECL, but at the cost of a smaller noise margin.

Complementary Bipolar Logic

Several attempts were undertaken to realize complementary bipolar logic circuits, characterized by a small standby current and significant available currents, in order to obtain fast switching transients. The basic configuration closely resembles a CMOS inverter with a pnp transistor as the pull-up device and an npn transistor as the pull-down device. Such circuit concepts require fast vertical pnp transistors [37, 38]. Since this implies a remarkable process complexity, and since it is difficult to obtain well-defined logic levels without operating the circuits rail-to-rail[14] as in CMOS, complementary bipolar logic was found not to be competitive with CMOS.

6.11 References

[1] P.R. Gray, R.G. Meyer. *Analysis and Design of Analog Integrated Circuits.* Wiley, New York, 3rd edition, 1993.

[2] D.D. Weiner, J.F. Spina. *Sinusoidal Analysis and Modeling of Weakly Nonlinear Circuits.* van Nostrand, New York, 1980.

[14]This would result in saturation of the transistors employed and is therefore unwanted.

[3] P. Wambacq, W. Sansen. *Distortion Analysis of Analog Integrated Circuits*. Kluwer, Dordrecht, 1998.

[4] R.G. Meyer, M.J. Shensa, R. Eschenbach. Cross modulation and intermodulation in amplifiers at high frequencies. *IEEE J. Solid-State Circuits*, 7(1):16–23, 1972.

[5] S. Narayanan, H.C. Poon. An analysis of distortion in bipolar transistors using integral charge control model and Volterra series. *IEEE Trans. Circuit Theory*, 20(4):341–351, 1973.

[6] H.E. Abraham, R.G. Meyer. Transistor design for low distortion at high frequencies. *IEEE Trans. Electron Devices*, 23(12):1290–1297, 1976.

[7] L.C.N. de Vreede, H.C. de Graaff, J.A. Willemen, W. van Noort, R. Jos, L.E. Larson, J.W. Slotboom, J.L. Tauritz. Bipolar transistor epilayer design using the MAIDS mixed-level simulator. *IEEE J. Solid-State Circuits*, 34(9):1331–1338, 1999.

[8] W.D. van Noort, L.C.N. de Vreede, H.F.F. Jos, L.K. Nanver,J.W. Slotboom. Reduction of UHF power transistor distortion with a nonuniform collector doping profile. *IEEE J. Solid-State Circuits*, 36(9):1399–1406, 2001.

[9] W.D. van Noort, L.K. Nanver, J.W. Slotboom. Arsenic-spike epilayer technology applied to bipolar transistors. *IEEE Trans. Electron Devices*, 48(11):2500–2505, 2001.

[10] K.L. Fong, R.G. Meyer. High-frequency nonlinearity analysis of common-emitter and differential-pair transconductance stages. *IEEE J. Solid-State Circuits*, 33(4):548–555, 1998.

[11] M.P. van der Heijden, H.C. de Graaff, L.C.N. de Vreede. A novel frequency-independent third-order intermodulation distortion cancellation technique for BJT amplifiers. *IEEE J. Solid-State Circuits*, 37(9):1176–1183, 2002.

[12] L.M. Rucker. Monolithic bipolar diodes and their models. *IEEE Circuits Devices Mag.*, (March):257–263, 1991.

[13] P.C. Munro, F.-Q. Ye. Simulating the current mirror with a self-heating BJT model. *IEEE J. Solid-State Circuits*, 26(9):1321–1324, 1991.

[14] R.M. Fox, S.-G. Lee, D.T. Zweidinger. The effects of BJT self-heating on circuit behavior. *IEEE J. Solid-State Circuits*, 28(6):678–685, 1993.

[15] R.J. Widlar. New developments in IC voltage regulators. *IEEE J. Solid-State Circuits*, 6(1):2–7, 1971.

[16] A.P. Brokaw. A simple three-terminal IC bandgap reference. *IEEE J. Solid-State Circuits*, 9(6):388–393, 1974.

[17] Y.P. Tsividis. Accurate analysis of temperature effects in $I_{\rm C} - V_{\rm BE}$ characteristics with application to bandgap reference sources. *IEEE J. Solid-State Circuits*, 15(6):1076–1084, 1980.

[18] R.J. Widlar. Low voltage techniques. *IEEE J. Solid-State Circuits*, 13(6):838–846, 1978.

[19] G.C.M. Meijer, P.C. Schmale, K. van Zalinge. A new curvature-corrected bandgap reference. *IEEE J. Solid-State Circuits*, 17(6):1139–1143, 1982.

[20] M. Gunawan, G.C.M. Meijer, J. Fonderie, J.H. Huising. A curvature-corrected low-voltage bandgap reference. *IEEE J. Solid-State Circuits*, 28(6):667–670, 1993.

[21] S.L. Salmon, J.D. Cresler, R.C. Jaeger, D.L. Harame. The influence of Ge grading on the bias and temperature characteristics of SiGe HBTs for precision analog circuits. *IEEE Trans. Electron Devices*, 47(2):292–298, 2000.

[22] C.F. Hill. Noise margin and noise immunity in logic circuits. *Microelectronics*, 1:16–21, 1968.

[23] J. Lohstroh. Static and dynamic noise margins of logic circuits. *IEEE J.Solid-State Circuits*, 14(3):591 598, 1979.

[24] J. Lohstroh, E. Seevinck, J. de Groot. Worst-case static noise margin criteria for logic circuits and their mathematical equivalence. *IEEE J. Solid-State Circuits*, 18(6):803–807, 1983.

[25] E. Seevinck, F.J. List, J. Lohstroh. Static-noise margin analysis of MOS SRAM cells. *IEEE J. Solid-State Circuits*, 22(5):748 754, 1987.

[26] J.B. Foley, J.A.R. Bannister. Analyzing ECL's noise margin. *IEEE Circuits Devices Mag.*, (May):32–37, 1994.

[27] R. Müller, H.-J. Pfleiderer, K.-U. Stein. Energy per logic operation in integrated circuits: Definition and determination. *IEEE J. Solid-State Circuits*, 11(5):657–661, 1976.

[28] J. Lohstroh. Devices and circuits for bipolar (V)LSI. *Proc. IEEE*, 69(7):812–826, 1981.

[29] T.H. Ning, D.D. Tang. Bipolar trends. *Proc. IEEE*, 74(12):1669–1677, 1986.

[30] D.A. Hodges, H.G. Jackson. *Analysis and Design of Digital Integrated Circuits.* McGraw Hill, New York, 2nd edition, 1988.

[31] C.M. Hart, A. Slob. Integrated injection logic – a new approach to LSI. *IEEE J. Solid-State Circuits*, 7:346 351, 1972.

[32] H.H. Berger, S.K. Wiedmann. Merged transistor logic – a low cost bipolar logic concept. *IEEE J. Solid-State Circuits*, 7:340–346, 1972.

[33] S.P. Wainwright, S. Hall, P. Ashburn, A.C. Lamb. Analysis of Si:Ge heterojunction integrated injection logic (I2L) strucures using a stored charge model. *IEEE Trans. Electron Devices*, 45(12):2437–2447, 1998.

[34] S.K. Wiedmann. Advancements in bipolar VLSI circuits and technologies. *IEEE J. Solid-State Circuits*, 19(3):282–291, 1984.

[35] R.L. Treadway. DC analysis of current mode logic. *IEEE Circuits Devices Mag.*, (March):21–35, 1989.

[36] J. Lohstroh, J.J.M. Koomen, A.T. van Zanten, R.H.W. Salters. Punchthrough-currents in pnp and npn sandwich structures – I. Introduction and basic calculations. *Solid-State Electron.*, 24(9):805–814, 1981.

[37] S.K. Wiedmann. Charge buffered logic (CBL) – a new complementary bipolar circuit concept. *Symp. VLSI Tech., Dig. Tech. Papers, Kobe*, pp. 38–39, 1985.

[38] S.K. Wiedmann, D.F. Wendel. Speed enhancemant and key design aspects of charge buffered logic. *Symp. VLSI Tech., Dig. Tech. Papers, San Diego*, pp. 43–44, 1986.

7 Process Integration

Today, a wide variety of transistor circuits is available in integrated form, discrete transistors are employed only if an appropriate integrated circuit is not available or if the transistor properties required are not compatible with integrated circuit technology. This chapter briefly explains the fabrication of integrated npn transistors, passive components and pnp transistors. The last section considers reliability issues associated with forward- and reverse-bias stress and electrostatic discharges.

7.1 Fabrication of Integrated npn Transistors

This section outlines the fabrication of modern high-frequency bipolar transistors, focusing on the basic process steps for collector isolation and emitter and base formation, which are specific to high-frequency bipolar transistors. For a general introduction to the technology of integrated circuits see [1], for example.

7.1.1 Collector Isolation

In integrated bipolar transistor circuits, the collectors of different transistors have to be isolated from each other. This is usually done by embedding the n-doped collector regions in a p-type substrate, which is connected to the most negative potential, resulting in a reverse-biased collector–substrate (cs) junction. A large bc breakdown voltage and a small bc depletion capacitance require a low-doped collector region. For that purpose, an epitaxial collector region on top of a heavily doped buried layer, which is introduced to reduce the collector series resistance, is generally employed. Advances in collector isolation techniques have allowed manufacturers to improve the packing density and to reduce the cs capacitance. Figure 7.1 shows three widely employed approaches to isolate the collectors of different transistors from each other.

PN-Junction Isolation. The pn-junction isolation technique shown in Fig. 7.1a employs a p-type ring, which surrounds the n-type epitaxial and buried-layers region of each transistor. This approach consumes a lot of chip area and yields substantial device capacitances, resulting in large power consumption; additional drawbacks are a substantial cs capacitance and the thermal stress during the drive-in of the p-type isolation rings, which causes significant dopant diffusion out of the buried layer, with the consequence of poor controllability of the doping profile in the epitaxial collector region. This approach is therefore no longer employed in modern integrated circuits.

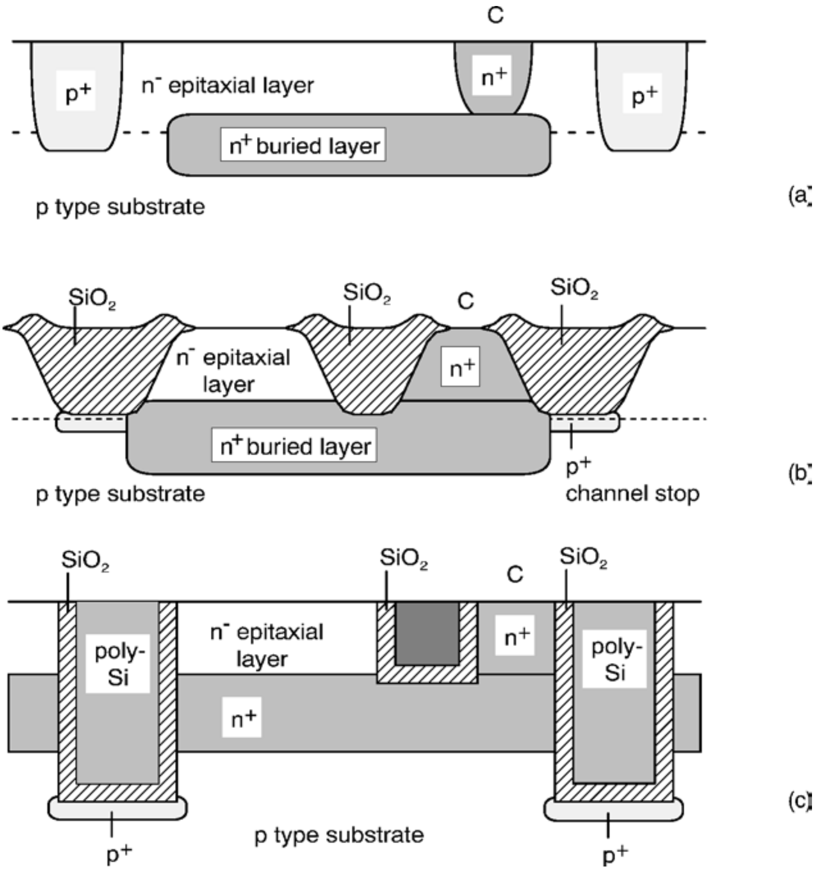


Fig. 7.1. Collector isolation in integrated bipolar circuits: (**a**) p-type diffusion, (**b**) LOCOS isolation, (**c**) trench isolation (U-groove isolation)

LOCOS Isolation. A substantial increase in packing density and a reduction in parasitic transistor capacitances was achieved with the introduction of LOCOS[1] isolation, shown in Fig. 7.1b. In this approach, a partial oxidation of the chip surface is performed, by shielding active device areas with silicon nitride during the oxidation process. This results in a transistor area that is surrounded by LOCOS oxide (field oxide), which also allows one to reduce the capacitance of device interconnects to the underlying substrate. As compared with pn-junction isolation, a reduction of the transistor area to approximately one third of the area used before can be achieved. In order to avoid short circuits between neighboring devices, owing to unwanted formation of inversion layers below the LOCOS oxide, an additional p-type implant, called

[1]The acronym LOCOS is an abbreviation of local oxidation of silicon.

the "channel stop" (Fig. 7.1b), increases the doping concentration below the LOCOS oxide and thus the threshold voltage for inversion-layer formation.

U-Groove Isolation. Narrow isolation regions are difficult to realize with the LOCOS technique, as the width of the isolation layer has to be larger than the thickness by a factor of approximately two. Further reduction of the lateral extension is, however, possible with plasma etching. This process allows one to realize almost vertically etched grooves, which surround the transistor region, and which are filled with insulating material (trench isolation or U-groove isolation). This technique[2] has the additional advantage that the buried layer may be deposited as an unstructured layer covering the whole chip. The reduced thermal budget allows one to realize steep vertical doping profiles in the collector region. The trench isolation capacitance depends on the depth and width of the trench, the state of the bottom (open or closed) and the filling of the trench [3, 4].

Selectively Grown Collector Regions. Selective epitaxial growth was first reported in 1962 and is usually performed under low pressure (< 100 Torr) and at low temperature ($< 900°C$) with a substantial HCl flow. This allows to overcome problems resulting from the fact that the growth rate of selective epitaxial silicon is a function of the nucleation site seed area and the ratio of the area of the SiO_2 mask to the silicon area exposed. The selectivity of the growth process increases with (1) decrease in temperature, (2) decrease in pressure, (3) decrease in growth rate and (4) increase in HCl concentration. Selective epitaxial growth of the collector region (Fig. 7.2) is a means to reduce isolation capacitances. In contrast to the LOCOS technique, the thickness of the field oxide can be chosen independently of the lateral isolation dimensions.

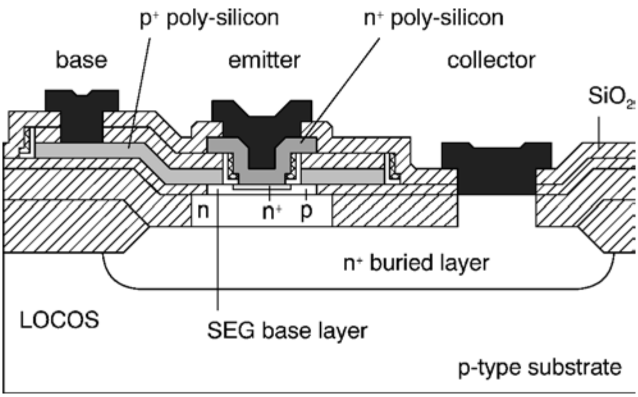


Fig. 7.2. Cross section of self-aligned bipolar transistor with selectively grown epitaxial base and collector layers (after [5])

[2] See [2] for an early example of this technique.

In [6, 7], selective epitaxial growth and chemical–mechanical polishing were used for the realization of devices with a cross section of the kind depicted in Fig. 7.3a. In these devices, the cs capacitance is reduced by two means: insertion of an intrinsic layer between the p-type substrate and n^+ buried subcollector to reduce the cs capacitance per unit area, and epitaxial overgrowth to reduce the area of the cs junction. Such a device cross section increases the thermal resistance between the device and substrate and therefore results in increased self-heating. The realization of collector–substrate windows with a larger width W_{CS} is therefore advantageous for devices with substantial power dissipation, such as those used in high-speed digital applications.

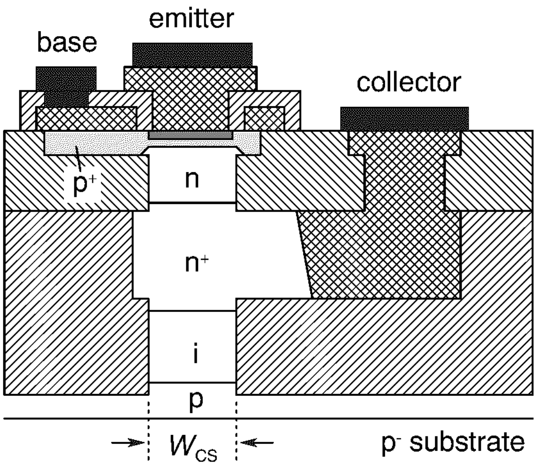


Fig. 7.3. Cross section of SPIRIT (sequentially planarized interlevel isolation technology) isolation scheme (white and gray areas, mono-crystalline silicon; hatched areas, oxide and cross-hatched areas, polycrystalline silicon, after [6])

7.1.2 Emitter and Base Formation

The formation of the emitter and base is crucial for the performance achieved by a bipolar process, as these steps determine most of the essential device parasitics, such as the base resistance, base transit time and depletion capacitances. Lateral and vertical scaling of bipolar transistors requires the use of self-alignment, emitter contacts with finite surface recombination velocity or HBTs with a wide-gap emitter. Several techniques are available for the realization of base layers with a thickness below 100 nm: (1) low-energy implantation through a thin screen oxide to reduce channeling effects, (2) diffusion of the active base layer from polysilicon or some other solid diffusion source, and (3) epitaxial growth of thin p-doped layers (selective and nonselective).

After the formation of mutually isolated collector regions, the base and emitter regions are formed. In conventional planar technology, first the base region and then the emitter region are diffused through openings in appropri-

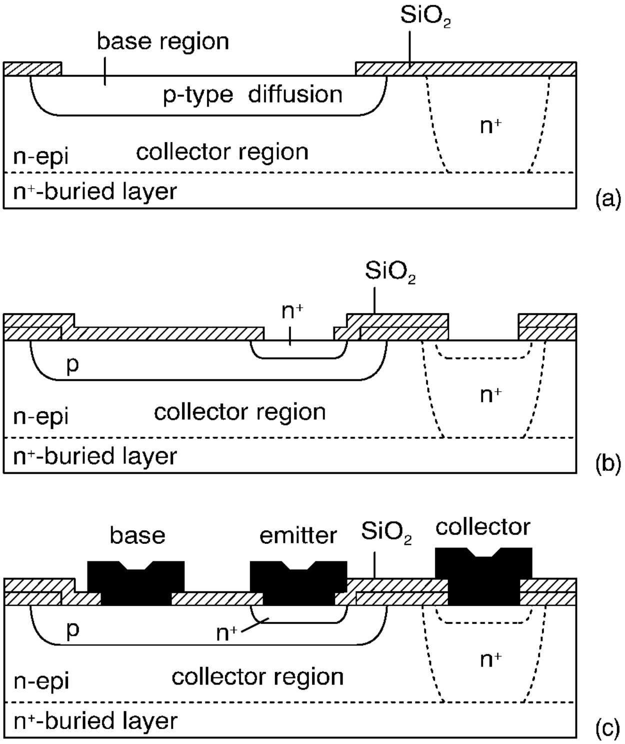


Fig. 7.4. Formation of emitter and base layer in classical planar technology. (**a**) diffusion of base layer, (**b**) emitter diffusion, and (**c**) formation of metal contacts

ately structured layers of silicon dioxide (Fig. 7.4). Since the base contact is formed above the bc diode, this technique has a very unfavorable ratio X_{CJC} of the areas of the eb and bc depletion capacitances. The limitations with respect to vertical scaling of the doping profile imposed by the metal contact were overcome with self-aligned polysilicon-contacted emitters.

Self-Aligned Emitter–Base Diodes

A significant improvement of the device performance came with the introduction of a self-aligned emitter–base isolation scheme [8–14]. The double-polysilicon technique, illustrated in Fig. 7.5, where both the base contact region and the emitter region are diffused from heavily doped polysilicon, is the most prominent approach. In the first step, a double layer of heavily doped p-type silicon and silicon dioxide is deposited. In the subsequent plasma etching process, this double layer is removed except in those regions where the external base is located (Fig. 7.5a). Since there is no "etch stop" at the interface between the poly- and monocrystalline silicon (the etch rates in poly- and monocrystalline silicon are approximately equal), the etch time and etch rate have to be controlled carefully in order to minimize the overetch into the

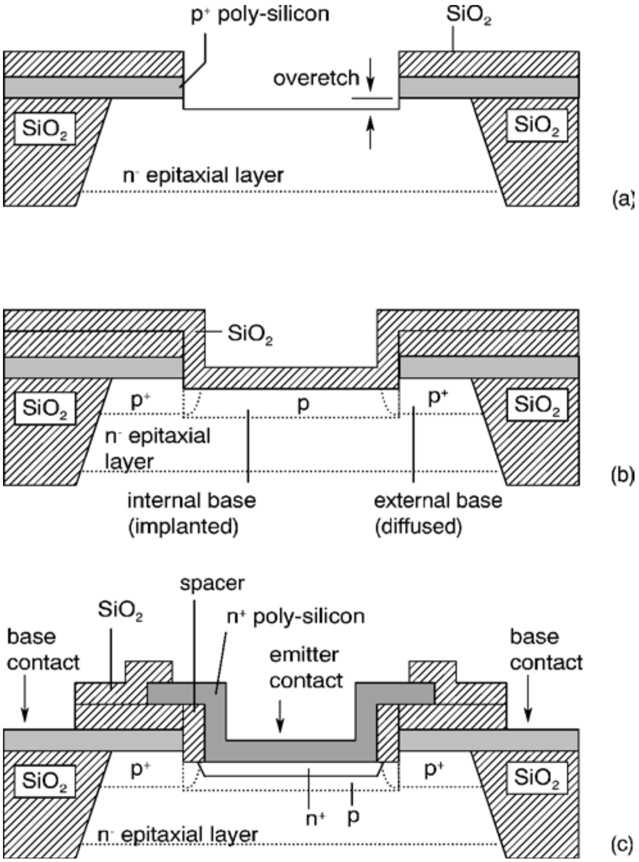


Fig. 7.5. Formation of a self-aligned eb diode. (**a**) Anisotropic etch of emitter window; (**b**) diffusion of external base, base implantation and deposition of oxide layer with good step coverage; and (**c**) deposition and etching of n^+ polysilicon, oxide deposition, emitter diffusion and etching of contact holes

monocrystalline silicon. The lateral extent of the hole etched into the double layer determines the emitter area, which, however, is somewhat smaller owing to subsequent processing steps. After diffusion of the external base region from the remaining ring of p^+ polysilicon and implantation[3] of the internal base region, followed by an annealing step, a layer of SiO_2 is deposited (Fig. 7.5b) and subsequently removed by plasma etching. Only the "spacers" (Fig. 7.5c) remain. In the next step, polysilicon is deposited, which is heavily doped either during deposition (in-situ-doped polysilicon) or by a subsequent implantation

[3]If the active base layer is formed by implantation before spacer formation, a good contact between the external and internal base regions is guaranteed. With this technique, however, it may be difficult to control the eb breakdown voltage, which is determined by tunneling in the sidewall diode and thus by the doping density below the spacer region. The latter is determined by the overlap of the internal and external base regions and cannot be controlled easily. If the active base layer is formed by implantation after spacer formation or by diffusion from the (emitter) polysilicon layer, a poor contact between the external base region and the active base region may result. Special process schemes have been developed to overcome this problem [15].

of donors. This layer serves as the emitter contact and as a source of dopants for the diffusion of the shallow eb junction. Special techniques, such as rapid thermal processing[4] (RTP), allow the reproducible formation of eb junctions situated less than 50 nm below the polysilicon emitter contact. Etching the n^+ poly, oxide deposition and etching of contact holes follow. Since the contact holes are placed above thick insulating regions, the bc capacitance and bs capacitances are low. Furthermore, the internal base, which lies below the emitter region, is well contacted owing to the low-resistance, ring-shaped base contact region, which is separated from the internal base by only a thin region below the spacer.

Another approach to reducing the base resistance and bc capacitance is the SICOS (sidewall base contact structure) configuration depicted in Fig. 7.6. This process aims to minimize the bc depletion capacitance by forming the

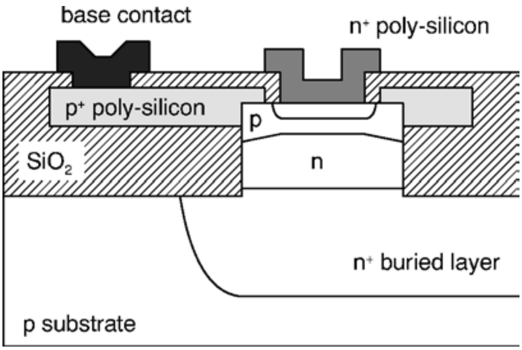


Fig. 7.6. Cross-section of sidewall-contacted base transistor (SICOS configuration)

external base on top of a thick oxide layer [16]. However, this process never reached the maturity of the double-polysilicon process outlined above.

Scaling of the Double-Polysilicon Bipolar Transistor. Improved lithography and the polysilicon emitter contact allowed the "scaling" of the double-polysilicon bipolar transistor, i.e. the coordinated reduction of the lateral and vertical device dimensions. In "scaled" devices, increased dopant concentrations in the base and collector regions are required to reduce the depletion layer thickness.

Improved device performance in scaled devices depends to a large extent on larger transfer current densities. Therefore an increased dopant density in the collector region is required to increase the critical current density for the onset of the Kirk effect. A selective collector implantation before spacer

[4] In this technique, the diffusion source is heated rapidly by radiation from strong lamps. This allows one to combine a very high diffusion temperature (required for good electrical activation of the donors) with a small diffusion time (required for shallow junctions).

formation [17] allows one to increase the dopant density in the collector region of the active transistor, whereas the collector doping density in the external bc diode remains unaffected. Such a pedestal collector or selectively implanted collector (SEC) is widely used. The decrease of the open-base breakdown voltage BV_{CEO} with increased collector doping requires careful optimization of the collector doping profile. In [18], numerical simulations were used to demonstrate that a thin highly doped layer adjacent to the base–collector junction can shift the onset of high-level-injection effects in the collector region to larger current densities and therefore improve the device cutoff frequency without causing significant deterioration of BV_{CEO} and f_{max}, provided the voltage drop across the heavily doped layer is lower than an effective threshold of approximately 1.2 V.

Originally, the polysilicon used to form the emitter contact was deposited undoped, and dopants were introduced before emitter drive-in by implantation. This allows rather precise control of the dopant content in the diffusion source. A problem with this technique is the emitter plug effect [19], which is relevant for narrow emitter stripes, as the polysilicon layer at the edge of the emitter window is much thicker in the direction of implantation than the bottom portion is. After implantation, the sidewall is therefore undoped. If redistribution of dopants within the polysilicon layer is slow on the timescale of the rapid thermal emitter drive-in, the sidewall portion will not act as an efficient diffusion source for the monocrystalline emitter region. A reduction of this problem is possible if the thick oxide spacer is replaced by a thin L-shaped spacer [20]. The problem can be avoided by using in-situ-doped polysilicon [21] to form the emitter contact. An additional advantage of in-situ-doped polysilicon diffusion sources is a reduction of the emitter series resistance by up to a factor of about two.

Reducing the energy of the base implant, the use of BF_2 instead of B and minimization of the thermal budget[5] have allowed substantial improvements of device performance (see, for example, [20,22–24]). In [25,26], a low-energy base implant and subsequent RTP, together with a salicide[6] base layer to reduce the external base resistance, have been employed to realize manufacturable self-aligned double-poly BJTs with a cutoff frequency in excess of 50 GHz and a minimum ECL gate delay of 12 ps. Double-polysilicon devices optimized with respect to the power–delay product were described in [28]; there, ECL delays of 50 ps and a CML power–delay product of 4.5 fJ at 1.1

[5] Sometimes phosphorus is used instead of arsenic as an emitter dopant, owing to the smaller thermal budget required for electrical activation of phosphorus.

[6] This term is an abbreviation for "self-aligned silicide"; in this technique the isolating material, which covers the polycrystalline base contact, is uncovered by anisotropic etching, using the emitter polysilicon as a mask. Silicidation of the uncovered polysilicon layer results in a reduced value of the external base resistance. Silicidation of the sidewall of the polysilicon base contact has also been used for the reduction of the layout dimensions [27].

V supply voltage[7] were achieved within a 0.35 µm BiCMOS technology. Self-aligned silicon bipolar transistors with $f_T = 50\,\text{GHz}$ and $f_{max} = 81\,\text{GHz}$ were demonstrated in [29]. Double-polysilicon transistors with a cutoff frequency $f_T = 74\,\text{GHz}$ on an SOI substrate using in-situ phosphorus-doped polysilicon as the emitter contact were described in [30].

Epitaxially Grown Base

Epitaxially grown base layers are of particular interest for the fabrication of HBTs with a SiGe base layer. The UHVCVD technology employed in early developments [31–33] has some advantages, such as precise controllability of impurity profiles and low growth temperatures; however, the memory effect of the dopant in the chamber and the low throughput are problems in mass production [34]. Higher throughput can be achieved with low-pressure chemical vapor deposition (LPCVD), which has a higher growth rate and avoids the memory effect, owing to the self-cleaning of the chamber wall by HCl etching. HCl gas has been assumed to be necessary to obtain selectivity. Its use, however, decreases the growth rate and sometimes makes it difficult to optimize the growth conditions. Therefore HCl-free LPCVD growth techniques were developed [34], and were demonstrated to allow the production of SiGe HBTs with $f_{max} = 160\,\text{GHz}$ and high yields.

Various self-alignment schemes have been investigated for forming contacts to epitaxially grown base layers [35, 36]. As an example, the process flow depicted in Fig. 7.7 will be described [37–39]. First, a thin SiO_2 film, an undoped polysilicon buffer layer (thickness 40 nm) and a thin Si_3N_4 film are deposited on a silicon substrate. The subsequently deposited layers of in-situ boron-doped polysilicon (IBDP), SiO_2 (thickness 20 nm) and undoped polysilicon (thickness 150 nm), and a thick layer of SiO_2 are patterned in order to define the emitter window (Fig. 7.7a). Then, SiO_2 and Si_3N_4 sidewalls are formed around the emitter area (Fig. 7.7b). Wet etching is used to remove the Si_3N_4 sidewall, and also the Si_3N_4 layer, the polysilicon buffer layer and the SiO_2 layer in the active device area (Fig. 7.7c). Selective growth of doped SiGe layers and a Si cap-layer (thickness 20 nm) in a UHVCVD system follows; a 20 nm thick boron-doped layer with $N_A = 10^{19}\ \text{cm}^{-3}$ forms the intrinsic base. During the formation of the epitaxial base, a polycrystalline SiGe layer grows on the buffer layer and the IBDP surfaces (Fig. 7.7d). This provides a connection between the intrinsic base and the IBDP layer, which serves as the extrinsic base. After deposition of a 20 nm thick screen oxide on the SiGe and Si cap layers, a selectively implanted collector is formed

[7] This value benefits from the fact, that in BiCMOS technology current sources of ECL or CML stages can be realized with n-channel MOSFETs, which yield approximately constant current at smaller voltage drop as BJT current sources with their unavoidable emitter series feedback.

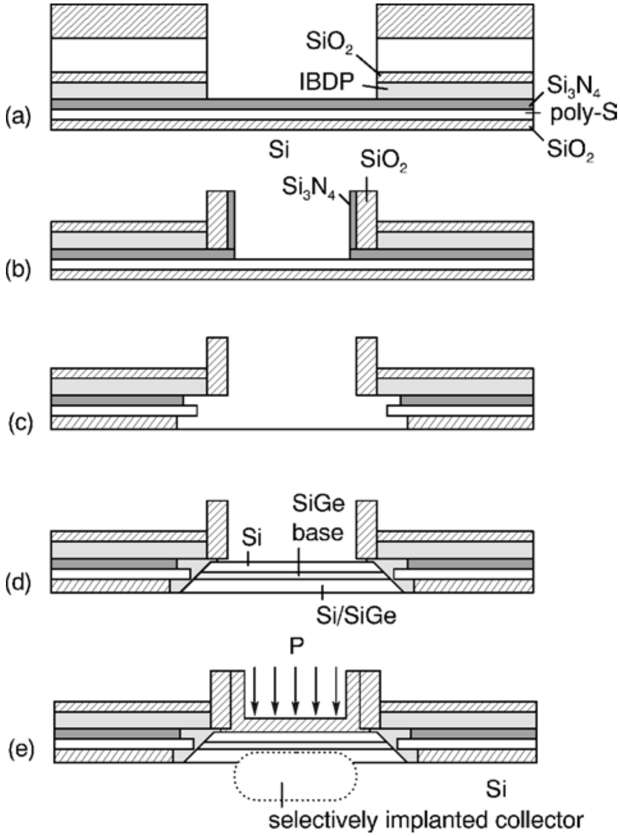


Fig. 7.7. Process flow for fabrication of self-aligned bipolar transistor with selectively grown epitaxial base layer (after [37])

by phosphorus implantation (Fig. 7.7e). After the screen oxide is removed, in-situ phosphorus-doped polysilicon is used to form the emitter with rapid thermal annealing. This step also activates the donors implanted into the collector region. The external base resistance was finally reduced by depositing a selective-CVD tungsten[8] layer on the IBDP. Self-aligned HBTs with a selectively grown SiGe base layer, trench isolation and Ti salicide electrodes produced in a BiCMOS process including varactors ($Q = 45$ at 10 GHz) and spiral inductors ($Q = 19$ at 10 GHz) were demonstrated to have a cutoff frequency of 76 GHz and an ECL gate delay of 6.7 ps in [41].

[8]The formation of a $TiSi_2$ layer was found to affect the diffusion of impurities: $TiSi_2$ formation results in shallower emitters with a substantial dependence on the width W_E of the emitter window, with the consequence of a layout-dependent current gain B_N [40]. This dependence can be reduced if emitter diffusion occurs before $TiSi_2$ formation.

7.2 Passive Components

The realization of RF integrated circuits requires passive components such as resistors, capacitors, inductors and transformers on the chip.

7.2.1 Resistors

There are several possible ways to realize resistors in an integrated bipolar technology. Classical diffused resistors (Fig. 7.8a) are problematic in high-frequency circuits owing to the capacitance of the depletion layer formed between the resistive sheet and the substrate. Therefore, polysilicon resistors (Fig. 7.8b) are generally employed; these are deposited on a thick oxide and have a smaller capacitance with the underlying substrate. As compared with As-doped n-type polysilicon resistors, B-doped p-type polysilicon resistors were found to show substantially better performance in terms of temperature coefficient and linearity [23]. Polysilicon resistors were found to increase their resistance after thermal or electrical stress. This is attributed to an increase of trapping states at the grain boundaries, which causes larger potential barriers and thus larger resistivity. Measures to reduce this problem include incorporation of fluorine to reduce the number of dangling bonds, and compensation doping [42].

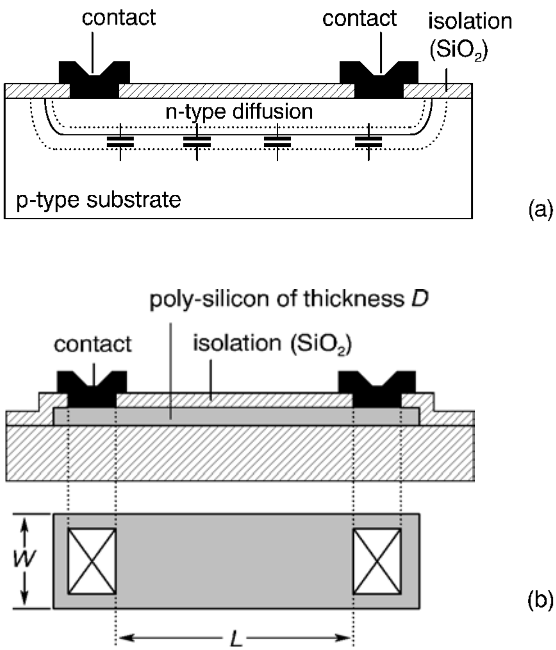


Fig. 7.8. Two widely employed methods to form resistors in integrated bipolar technology. (**a**) Diffused resistor and (**b**) polysilicon sheet resistor

Since the thickness and dopant concentration of the polysilicon layer generally cannot be influenced by the designer, the sheet resistance, $R_F = \rho/D$, is used as a design parameter. A resistor of width W and length $L \gg W$ has a resistance

$$R = R_F \frac{L}{W} + 2R_{Ce} \,, \tag{7.1}$$

where R_{Cc} denotes the contact end resistance. The contact end resistance is determined by the specific contact resistance ρ_C and the area A_C of the contact, as well as by the sheet resistance of the material below the contact area (Fig. 7.9a). For a long ohmic contact, the contact end resistance can be

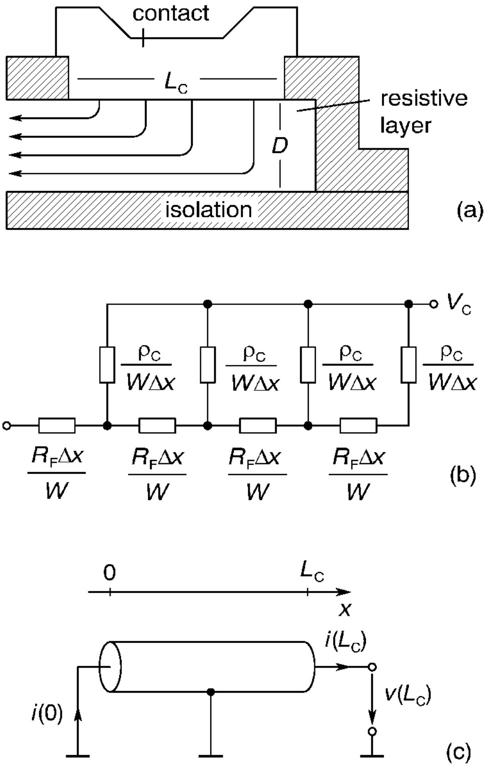


Fig. 7.9. Calculation of the contact end resistance. (**a**) Cross section and current flow, (**b**) equivalent circuit and (**c**) transmission line model

calculated from the transmission line model [43,44]. The equivalent circuit of the contact, shown in Fig. 7.9b, may be applied if the width of the contact area equals the width W of the resistor sheet. In this equivalent circuit, the contact is divided into small segments of length Δx with contact resistance $\rho_C/(W\Delta x)$ to the underlying sheet. The latter is represented as a series connection of resistances $R_F\Delta x/W$. In the limit $\Delta x \to 0$ the contact can be seen as analogous to a transmission line of the RG type.[9] Using $G' = W/\rho_C$ and $R' = R_F/W$, we obtain the propagation constant

$$\gamma = \sqrt{R'G'} = \sqrt{R_F/\rho_C}$$

and the characteristic impedance

$$Z_0 = \sqrt{R'/G'} = \sqrt{\rho_C R_F}/W .$$

With these parameters, the contact end resistance R_{Ce} can be obtained from the RG line depicted in Fig. 7.9c. Here the contact is assumed to be at ground potential ($V_C = 0$); if the requirement $\underline{i}(0) = 0$ is imposed, the contact end resistance is given by $R_{Ce} = -\underline{v}(L_C)/\underline{i}(L_C)$:

$$R_{Ce} = Z_0 \coth(\gamma L_C) = \frac{\sqrt{\rho_C R_F}}{W} \coth\left(L_C \sqrt{R_F/\rho_C}\right) . \tag{7.2}$$

For values of the contact length L_C large in comparison with the transfer length $\lambda_T = \sqrt{\rho_C/R_F}$, the contact end resistance approaches $\sqrt{\rho_C R_F}/W$.

The contact end resistance of a contact with a more complicated geometry can be calculated with the help of a quasi-three-dimensional approximation [45], if the vertical dimensions are small in comparison with the lateral ones. In this approach, the current and continuity equations

$$\boldsymbol{J} = \sigma \boldsymbol{E} = -\sigma \nabla \psi \quad \text{and} \quad \nabla \cdot \boldsymbol{J} = 0 \tag{7.3}$$

in the resistive sheet are integrated with respect to the vertical coordinate (here denoted by z). If the resistor is isolated on the backside ($J_z(0) = 0$), this results in

$$\boldsymbol{J}_{\|} = -\sigma_s \nabla_{\|} \psi(x,y) \quad \text{and} \quad \nabla_{\|} \cdot \boldsymbol{J}_{\|} = \frac{J_z(x,y,D)}{D} , \tag{7.4}$$

where

$$\boldsymbol{J}_{\|} = \boldsymbol{e}_x \frac{1}{D} \int_0^D J_x \, \mathrm{d}z + \boldsymbol{e}_y \frac{1}{D} \int_0^D J_y \, \mathrm{d}z , \qquad \sigma_s = \frac{1}{D} \int_0^D \sigma \, \mathrm{d}z ,$$

and $J_z(x,y,D) = 0$ outside the contact area. If the contact is held at a potential V_C and ρ_C denotes the specific contact resistance, the boundary condition $V_C = \psi(x,y) + \rho_C J_z(x,y,D)$ has to be fulfilled, resulting in the second-order differential equation for $\psi(x,y)$

$$-\nabla_{\|} \cdot [\, \sigma_s(x,y) \, \nabla_{\|} \psi(x,y) \,] = \frac{V_C - \psi(x,y)}{\rho_C D} , \tag{7.5}$$

which simplifies to

[9] A transmission line with negligible capacitance and inductance.

$$\left(\lambda_T^2 \nabla_\parallel^2 - 1\right)\psi(x,y) = -V_C \,, \tag{7.6}$$

where the transfer length is $\lambda_T = \sqrt{\rho_C \sigma_s D} = \sqrt{\rho_C / R_F}$, if σ_s shows negligible lateral variation. The solution of the boundary value problem generally requires numerical methods, analytical solutions may, however, be found for some special geometries.

7.2.2 Capacitors

Linear capacitors are required in various electronic-circuit configurations and can be realized as MOS capacitors in accumulation or as parallel-plate capacitors between two layers of metal and/or polysilicon. Metal–insulator–metal (MIM) capacitors between two layers of metal offer the largest Q-value, but a comparatively small capacitance per unit area. In [46], for example, a 3.5 pF MIM capacitor with an area of $240 \times 415 \mu\mathrm{m}^2$, a Q-value of 36 at $f = 1\mathrm{GHz}$ and a self-resonance frequency of 6.6 GHz has been compared with a 2 pF MOS capacitor with an area of $35 \times 31.8 \mu\mathrm{m}^2$, a Q-value of 7 at $f = 1\mathrm{GHz}$ and a self-resonance frequency of 8GHz. These results demonstrate that MOS capacitors (if available) have a large capacitance per unit area at moderate Q-values; for the maximum Q-value, however, the MIM capacitor is the preferable device. MIM capacitors generally yield good linearity and quality factors, but suffer from a low capacitance density, which arises mainly from the relatively large thickness and small dielectric permittivity of the insulating layer. The thicknesses of the insulating layer which separate different metal layers cannot be reduced without constraint, as such a step will generally increase crosstalk capacitances between metal interconnect lines in different layers. Parallel-plate capacitors therefore often consume a substantial amount of chip area. The capacitance density can be improved by exploiting both vertical and horizontal electric field components [47].

Varactors

Varactor diodes employ the bias-dependent small-signal capacitance of the depletion layer of a pn junction. They are widely used in parametric amplification, harmonic generation, mixing, detection and voltage-variable tuning [48]. The realization of integrated voltage-controlled oscillators (VCOs) requires voltage-dependent capacitors. If the underlying process does not feature thin gate oxides for the purpose of forming MIS varactors, varactors must be realized with either the eb, the bc or the cs junction.

The most important parameters of a varactor are its Q-value, its area consumption and its control voltage range. Only a poor quality can be achieved with the cs junction, which is in series with a comparatively large substrate resistance. The following considerations therefore focus on the bc and the eb junctions.

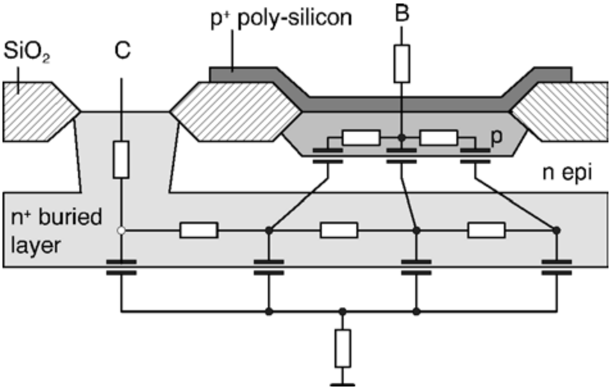


Fig. 7.10. Cross section through and equivalent circuit of a bc varactor diode. To minimize the series resistance r_S it may be advantageous to use a collector contact that surrounds the diode

Figure 7.10 shows a cross section through a bc diode, which shows a very small leakage current. In circuit applications the collector should be kept at constant potential, since otherwise the cs junction, with its poor quality factor, acts in parallel and reduces the Q-value.

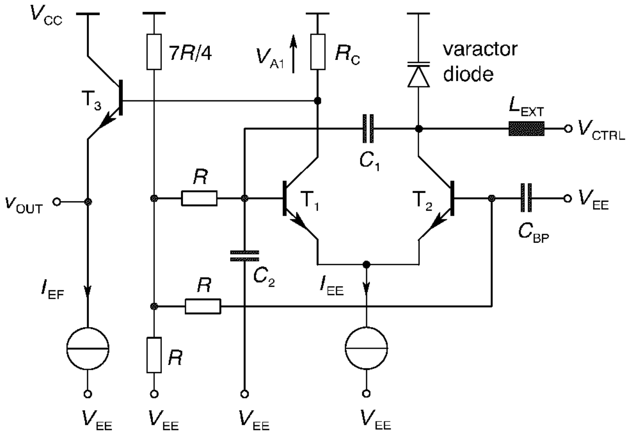


Fig. 7.11. Schematic of a varactor-tuned voltage-controlled oscillator (simplified, after [49])

An application of a varactor diode in a voltage-controlled oscillator is illustrated in Fig. 7.11. The oscillator uses an emitter-coupled pair, which provides a total harmonic distortion smaller than that of a single-transistor oscillator, since the tanh nonlinearity of a differential pair has only odd harmonics [49]. The effect of the varactor substrate junction has been minimized by connecting the cathode of the varactor diode to V_{CC}.

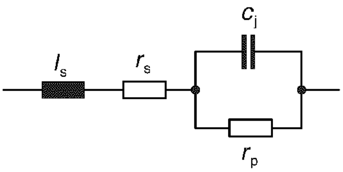


Fig. 7.12. Simplified equivalent circuit of a varactor diode

For a discussion of the Q-value of a varactor, we consider the simplified small-signal equivalent circuit of the varactor diode shown in Fig. 7.12. At low frequencies f, the Q-value is governed by the parallel resistance r_p caused by the diode leakage current ($Q \approx \omega r_\mathrm{p} c_\mathrm{j}$). In this range the Q-value increases in proportion to the frequency. At high frequencies, however, the series resistance causes a decrease ($Q \approx 1/(\omega r_\mathrm{s} c_\mathrm{j})$) of the Q-value. In bc varactor diodes, r_p is large and the Q-value is limited by r_s. The value of r_s is layout-dependent and is essentially determined by the buried-layer resistance in large-area varactor diodes [46]. Maximum Q-values between 5 and 7 for such varactor diodes were reported in [46].

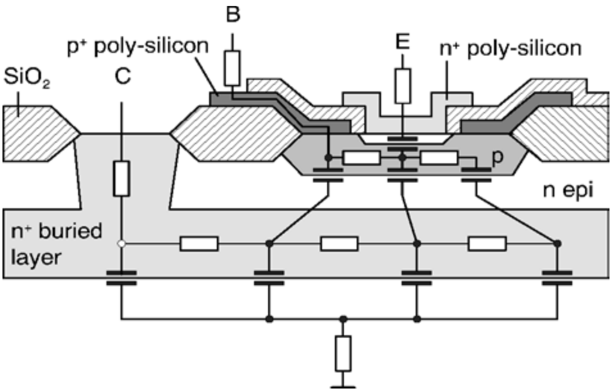


Fig. 7.13. Cross section through and equivalent circuit of an eb varactor diode

Figure 7.13 shows a cross section through the eb diode of a self-aligned double-polysilicon bipolar transistor. A reverse bias applied to such a diode implies a nonnegligible current due to tunneling, which causes a parallel conductance in the small-signal equivalent circuit and decreases the Q-value. If the base potential is held constant, the capacitance will equal the eb depletion capacitance, if the emitter potential is held constant, the eb and bc depletion capacitances will act in parallel. This diode provides a larger capacitance per unit area, the cost of a generally reduced Q-value owing to the typical increased series resistance and increased leakage current of the reverse-biased eb diode.

7.2.3 Inductors

On-chip inductors are helpful for impedance matching of low-noise amplifiers, for high-frequency filtering and for the realization of selective amplifiers. Therefore integrated inductors are highly desirable for monolithic silicon RF circuits, and many publications on this theme here appeared since the first demonstration of spiral inductors on silicon with a useful value of Q [50]. Two planar inductors interwound to promote magnetic coupling allow one to form planar transformers and baluns [51].

The fact that the integration of inductors in silicon technology has come into focus only recently is explained by the comparably coarse lithography and low frequencies of operation typical of earlier technologies, which resulted in large area consumption and poor Q-values. With frequencies in the gigahertz range and lithography techniques that allow a metal pitch in the micron range, more windings can be realized in a given chip area, while smaller inductance values are required. Using the possibilities of a modern bipolar process, inductance values in the range of 0.5–100 nH with Q-values up to 40 have been shown to be feasible [52]. The area consumed by an integrated planar inductor (Fig. 7.14) is generally a few hundred micrometers by a few hundred micrometers. An optimum inductor design yields the maximum value of Q (at a fixed maximum area consumption) and a self-resonance frequency that lies well above the frequency of operation. Constraints on the Q-value achievable arise from the specific resistivity of the metallization layers and the low-resistivity and therefore lossy substrate. Silicon substrates are not particularly suited for the realization of high-quality inductors, as the electromagnetic field produced by the inductor windings causes eddy currents in the substrate, which result in inductor losses and a decrease of Q. A semi-insulating substrate as provided by undoped GaAs with its small intrinsic carrier density, would be better. The cost advantages of silicon technology, however, are leading designers to optimize inductors produced with conventional process technologies. If several metal layers are available, the inductor winding can be formed in more than one layer; these are connected in parallel, with the effect of a reduced series resistance of the inductor [46]. If only two metallization layers are present, the inductor winding should be formed in layer 2, which has a larger distance from the silicon substrate and therefore causes smaller substrate losses.

Electrical Characteristics of Integrated Inductors

The principal features of the behavior of integrated inductors may be understood by consideration of the equivalent circuit[10] shown in Fig. 7.15. The spi-

[10] The accuracy of this simplified model [50] can be improved by the addition of parallel capacitances to the substrate resistance to take account of the displacement current in the substrate and by the introduction of a frequency-dependent series resistance r_s, which takes account of the skin effect [53].

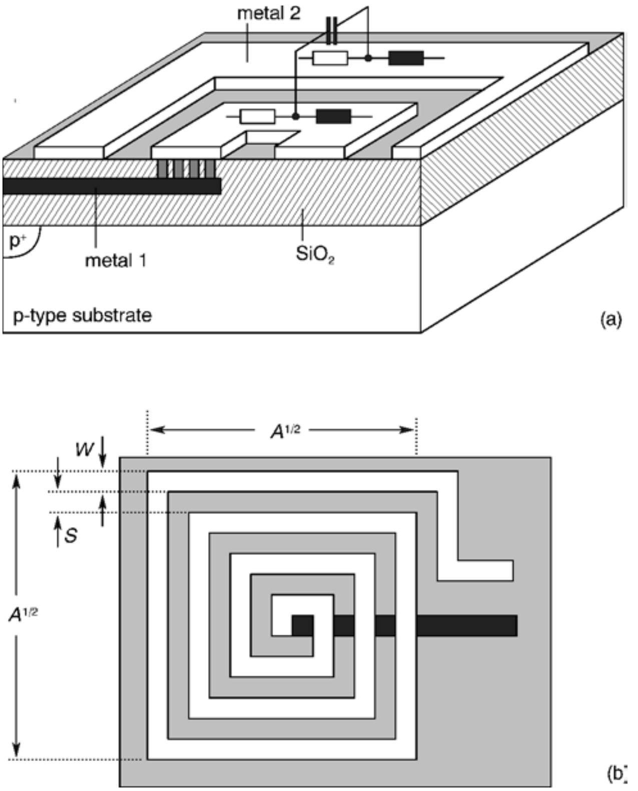


Fig. 7.14. (**a**) Cross section of planar inductor realized with two metal layers, (**b**) layout of planar inductor

ral coil is modeled by its inductance l, the series resistance r_{s} and a capacitor c_{p}, which represents the capacitive coupling between neighboring windings. The capacitances c_{s} and resistances r_{b} take account of the coupling to the substrate. The admittance between the two terminals is easily calculated to be

$$y = \mathrm{j}\omega c_{\mathrm{p}} + \frac{1}{r_{\mathrm{s}} + \mathrm{j}\omega l} + \frac{\mathrm{j}\omega c_{\mathrm{s}}/2}{1 + \mathrm{j}\omega r_{\mathrm{b}} c_{\mathrm{s}}} \,. \tag{7.7}$$

After separation of real and imaginary contributions, we obtain the Q-value of the inductor as follows up to terms in ω^3:

$$Q = \frac{\mathrm{Im}(\underline{z})}{\mathrm{Re}(\underline{z})} = -\frac{\mathrm{Im}(\underline{y})}{\mathrm{Re}(\underline{y})} \approx \omega\left(\frac{l}{r_{\mathrm{s}}} - r_{\mathrm{s}} c_{||}\right) \frac{1 + \omega^2 \tau_1^2}{1 + \omega^2 \tau_2^2} \,, \tag{7.8}$$

where $c_{||} = c_{\mathrm{p}} + c_{\mathrm{s}}/2$,

$$\tau_1^2 = \frac{r_{\mathrm{b}}^2 c_{\mathrm{s}}^2 (1 - r_{\mathrm{s}}^2 c_{\mathrm{p}}/l) - c_{||} l}{1 - r_{\mathrm{s}}^2 c_{||}/l} \approx r_{\mathrm{b}}^2 c_{\mathrm{s}}^2 - c_{||} l \tag{7.9}$$

and

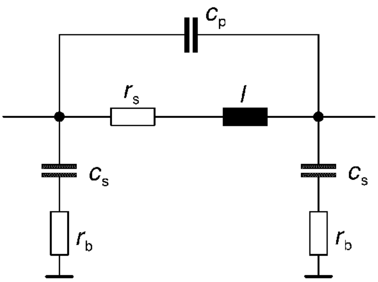


Fig. 7.15. Equivalent circuit of integrated planar inductor

$$\tau_2^2 = r_b(r_b+r_s/2)c_s^2 \,. \tag{7.10}$$

At low frequencies ($\omega \ll \omega_1, \omega_2$), the quality factor Q roughly increases in proportion to frequency. The equivalent circuit considered here illustrates the principles, but is too simple for the optimization of integrated inductors; models and design tools which take into account capacitive coupling between inductor windings and substrate coupling, as well as current constriction and proximity effects, are required for that purpose [54–57].

Several approaches allow one to improve the performance of integrated inductors. Examples are the use of higher-conductivity metal layers or multimetal layers in order to reduce the loss resistance of the inductor, the connection of multilayer spirals in series to reduce the area of the inductor, low-loss substrates to reduce losses caused by eddy currents in the substrate at high frequencies, and thick layers of field oxide to reduce the capacitive coupling to the substrate [55]. Q-factors up to 12 at 7.5 GHz for a 3.2 nH micromachined inductor realized with a six-level copper interconnect and a low-K dielectric were demonstrated in [58].

Substrate Coupling, Crosstalk. The large area of planar inductors and the associated large coupling capacitance to the substrate are responsible for the collection and generation of unwanted substrate noise.[11] In amplifying stages, unwanted parasitic feedback loops, which may lower the gain or even cause unwanted oscillations, may occur. The substrate noise produced by a planar inductor can be reduced by a guard ring, which surrounds the substrate region below the inductor. A broken guard ring was found to help reduce the eddy current induced in the substrate by the inductor [60] and thus reduce substrate noise, while maintaining the inductance value. Patterned ground shields were found to increase the peak Q by up to 33% [61] or 25% [62], while reducing the substrate coupling between two adjacent inductors.

[11] See, for example, [55, 59] for an analysis of substrate coupling effects.

7.3 PNP Transistors

In integrated bipolar transistor circuits, pnp transistors are of interest as current sources, active loads, etc. The implementation of such devices in a high-performance npn bipolar process requires a substantial increase in process complexity if high-performance vertical pnp transistors have to be fabricated in addition to npn transistors.[12] Therefore lateral pnp transistors are often employed if performance requirements are low.

7.3.1 Vertical pnp Transistors with Polysilicon Emitter

The realization of shallow eb junctions with an emitter diffused out of polysilicon, which has been previously doped by implantation is problematic owing to the lower grain boundary diffusivity of the boron dopant and its reduced segregation to the grain boundaries [63]. Owing to the reduced boron concentration in the polysilicon layer at the edge of the emitter window, the perimeter of the emitter–base junction may be positioned inside the polysilicon (Fig. 7.16), resulting in nonideal input characteristics and increased emitter–base leakage current.

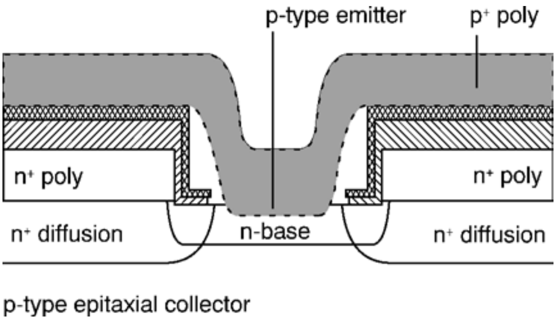


Fig. 7.16. Schematic cross section of the emitter–base diode of a double-poly self-aligned vertical pnp transistor. The gray area illustrates the region doped with boron. The edge of the eb junction is positioned inside the polysilicon layer [63]

P-type polysilicon emitters formed with a polysilicon layer doped by implantation of BF_2 were investigated in [64]. The experimental findings were explained using a model for the bonding of fluorine to oxygen and silicon at the polysilicon-monocrystalline-silicon interface. While the fluorine was found to give rise to increased current gain owing to the passivation of traps at the polysilicon-monocrystalline-silicon interface in devices given a low thermal

[12] Vertical pnp transistors that use the substrate as the collector can easily be integrated into a vertical npn bipolar process if the p-type external base region is used as the emitter, and the n-type epitaxial collector layer is used as the base (no buried layer) above the p-type substrate which serves as the collector. However, such a transistor can be used only if the collector is at ground potential, since no collector–collector isolation can be realized with this concept. A possible application of such transistors is in pnp emitter followers if the performance requirements are low.

budget during emitter drive-in, a decrease of the current gain was found after emitter drive-in with a higher thermal budget, an effect that has been explained by the accelerated breakup of the interfacial layer.

The high-frequency performance of vertical pnp transistors suffers somewhat from the smaller hole mobility in the base layer, which results in a base transit time that is increased by a factor of about two in comparison with a corresponding npn transistor. On the other hand, increased values of the open-base breakdown voltage BV_{ECO} are obtained, since this quantity is determined by the multiplication factor for injected holes, which is considerably smaller than the multiplication factor for injected electrons.

The realization of self-aligned pnp transistors with a polysilicon emitter contact was reported in [65]. A complementary bipolar processes with aligned eb diodes and n^+ and p^+ buried layers (isolated from the p-type substrate by an n-well) was described in [66]. High-speed, low-power complementary bipolar technologies with self-aligned npn and pnp transistors with cutoff frequencies exceeding 20 GHz have been described in [67, 68]. A summary of the silicon complementary process technology with vertically integrated pnp transistors reported in recent years has been presented in [69].

7.3.2 Lateral pnp Transistors

In many applications the performance requirements placed on the pnp transistors are much less severe than the demands on the npn transistors. In this case lateral pnp transistors, which are compatible with vertical npn transistor technology, are of interest. The term "lateral bipolar transistor" originates from the fact that the transfer current flows parallel to the semiconductor surface, as illustrated in Fig. 7.17.

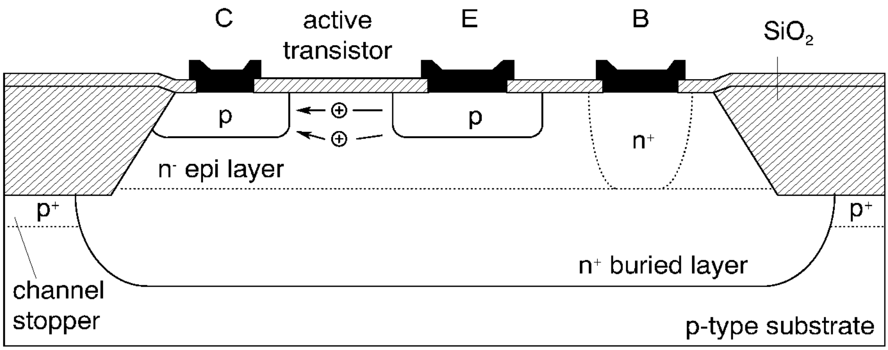


Fig. 7.17. Cross section (schematic) of lateral pnp transistor realized in vertical npn bipolar transistor technology

Such lateral pnp transistors cannot reach the performance obtained with npn transistors for several reasons:

1. The base width of the lateral pnp transistor is defined by lithography and is larger than the base width of the vertical npn transistor, which is defined by diffusion or implantation. Lateral pnp transistors therefore show a significantly reduced cutoff frequency.
2. The doping concentration of the n-type epilayer is chosen in order to obtain vertical npn transistors with optimum performance. The base doping of the lateral pnp transistor is therefore too small, with the consequence of a limited current-carrying capability, as the current gain degrades owing to high-level injection at comparably low values of transfer current.
3. Since the base doping is much smaller than the collector doping in a lateral pnp transistor, the bc depletion layer extends predominantly into the base layer, with the consequence of a strong dependence of the base width on V_{CB} and thus a small value of the Early voltage.
4. Both the emitter and the collector form parasitic vertical pnp transistors, at least one of which is in forward active mode under the usual operation conditions. This implies a reduction of the achievable current gain, and a current flow into the substrate.
5. The area consumed by a lateral pnp transistor designed for a certain value of collector current is much larger than the area consumed by a vertical npn transistor designed for the same current. The reasons for this are the small cross section of the current-carrying epilayer and the low base doping, which limits the transfer current density to values below those achievable in npn transistors.

Despite these drawbacks, lateral pnp transistors are widely used in analog integrated-circuit designs whenever the cutoff frequency is not critical.

Realization and Electrical Characteristics of Lateral PNP Transistors

Figure 7.18 shows a possible realization of a lateral pnp transistor within a double-polysilicon npn bipolar-transistor technology.[13] Each lateral pnp transistor forms two parasitic pnp transistors with the substrate (see Fig. 7.18). While the transistor formed by the collector, base and substrate is in the

[13] In addition to this technique, which forms lateral pnp transistors in the silicon substrate, approaches using silicon-on-insulator (SOI) have yielded interesting results. A lateral bipolar transistor with an $f_{\max}$ value of 67 GHz and small capacitances ($c_{\mathrm{je}} = 1.5\,\mathrm{pF}$, $c_{\mathrm{jc}} = 1.4\,\mathrm{pF}$, $c_{\mathrm{js}} = 2.5\,\mathrm{pF}$) on thin-film SOI for RF analog applications has been described in [70]. Lateral pnp bipolar transistors formed with a comparatively simple process and with peak cutoff frequencies in excess of 10 GHz were demonstrated in [71].

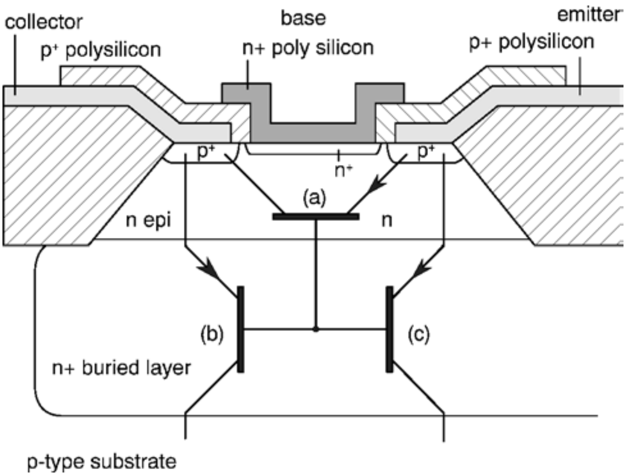


Fig. 7.18. Cross section through a lateral pnp transistor in self-aligned double-polysilicon technology. Associated with each integrated lateral pnp transistor (**a**) are two parasitic pnp transistors (**b**) and (**c**)

cutoff mode and is therefore negligible as long as the lateral pnp transistor does not operate in saturation, the parasitic transistor formed by the emitter, base and substrate will always be in its forward active mode if the lateral pnp transistor is in forward operation. In order to minimize the hole current injected into the substrate, such a transistor is usually laid out as a ring transistor with the emitter located in the center. A separate contact to the buried layer is generally not necessary: the area consumption associated with such contacts does not result in substantial improvement of the base series resistance.

Base Current and Current Gain. As in the case of the npn transistor, the base current of the lateral pnp transistor can be split up into three components:

$$I_{\mathrm{B}} = I_{\mathrm{EB}} + I_{\mathrm{BB}} + I_{\mathrm{CB}} \, , \tag{7.11}$$

where I_{EB} denotes the current due to recombination of electrons injected into the emitter region, I_{CB} denotes the current due to recombination of electrons injected into the collector region,[14] and I_{BB} denotes the current due to recombination in the base volume and at its surfaces. While I_{BB} is generally negligibly small for state-of-the-art vertical npn transistors, this component may dominate the base current in lateral pnp transistors.[15] The base current then increases with base width.

In lateral pnp transistors, high-level-injection effects occur in the base region at a comparatively small transfer current (Fig. 7.19). The reduction of the

[14]This component is generally negligible in forward active mode.

[15]This holds in particular for devices as depicted in Fig 7.18, where substantial recombination occurs in the n^+ regions, which contact the base layer.

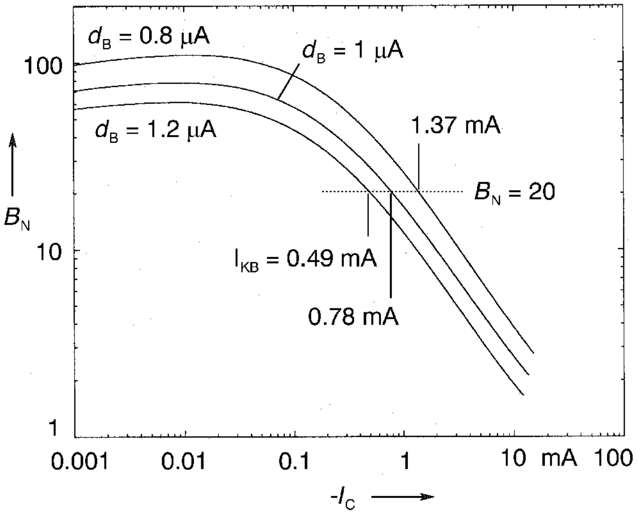


Fig. 7.19. Current gain B_N of lateral pnp transistors with different values of base width d_B vs. collector current I_C

current gain due to high-level-injection effects limits the transfer current that can be controlled to values smaller than a critical value I_{KB}: for practical purposes the current gain must not fall below a certain value B_{Nmin} (for example $B_{Nmin} = 20$) if one is to obtain a device with an output current an order of magnitude larger than the control current required. To see how I_{KB} is affected by the base width, the forward knee current per unit length I_{KF}/W can be roughly estimated as follows:

$$\frac{I_{KF}}{W} = \frac{1}{W}\frac{Q_{B0}}{\tau_B} = \frac{eN_{Depi}d_B d_{epi}}{d_B^2/2D_p} = 2eD_p N_{Depi} d_{epi}\,\frac{1}{d_B}\,.$$

Its value varies in inverse proportion to the base width d_B according to this estimate. The current gain observed at small current levels shows a similar dependence, since $I_T \sim 1/d_B$ the current gain[16] varies as

$$B_F \sim 1/d_B\,. \tag{7.12}$$

In the vicinity of I_{KB}, the current gain varies approximately as $1/I_C$. The maximum allowable collector current I_{Cmax} that results from the requirement $B_N > B_{Nmin}$ can therefore be estimated from

$$B_{Nmin} \approx B_F I_{KF}/I_{KB}\,. \tag{7.13}$$

The maximum allowable transfer current therefore varies as

$$I_{KB} \sim \frac{1}{B_{Nmin}}\frac{1}{d_B^2}\,. \tag{7.14}$$

[16]This discussion neglects the dependence of the base current on the base width; this is only valid if I_B is not dominated by the component I_{BB}, which increases with d_B.

Figure 7.19 shows measured values of B_N versus I_C for different values of d_B. A plot of $1/\sqrt{I_{KB}}$ versus the layout dimension yields a straight line, which confirms the dependence of the current-carrying capability $I_{KB} \sim 1/d_B^2$. The parameter I_{KB} therefore shows a substantial increase if the base width d_B is decreased – a step that causes a decrease of the Early voltage, however.

Early Effect. If the lateral pnp transistor is operated as an active load or employed in a current mirror, the forward Early voltage V_{AF} is important. This parameter is generally low in lateral pnp transistors, since the base doping is small in comparison with the collector doping. The bc depletion layer therefore extends predominantly into the base region and causes a strong decrease of the base width with increasing V_{BC}. Figure 7.20 shows measured output characteristics of a lateral pnp transistor with a nominal base width of 1.2 µm. The output resistance determined at $V_{EC} = 2$ V and $I_C = -0.06$ mA is 238 kΩ, corresponding to an Early voltage $V_{AF} = -r_o I_C - V_{EC} \approx 12.3$ V.

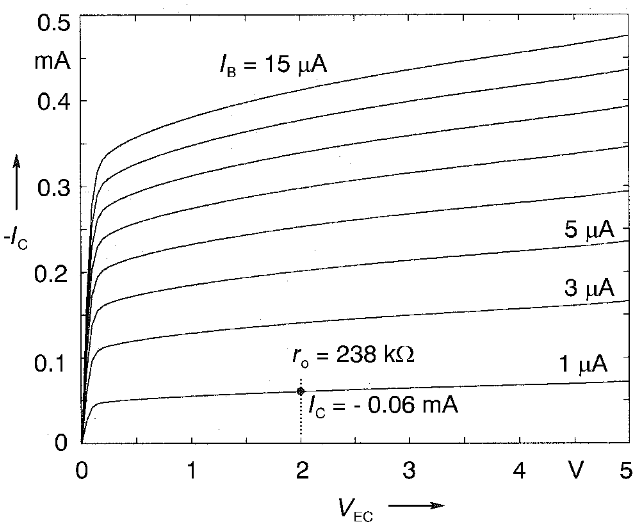


Fig. 7.20. Output characteristics of lateral pnp transistor with nominal base width of 1.2 µm

High-Frequency Behavior, Transit Time. Lateral pnp transistors generally show poor performance at high frequencies, owing to their small current-carrying capability, large capacitances and a forward transit time that is large in comparison with values achieved for vertical npn transistors. The large forward transit time is a consequence of the comparatively large base thickness and the minority charge stored in the epitaxial region below the emitter contact.

7.4 Reliability

This section surveys two important reliability issues: device degradation and failure caused by electrostatic discharges.

7.4.1 Device Degradation

Under reverse stress, hot carriers which have been accelerated in the electric field of the eb space-charge region may cause an increase in the number of recombination centers, with the consequence of an increase in the forward-bias base current and a degradation of the current gain. In addition, junction degradation has also been observed after very high-level forward injection.

Reverse-Bias Stress

In [72], a constant-voltage reverse-bias stress was applied to the eb diodes of some self-aligned bipolar transistors. The change ΔI_B of the forward-bias base current at a fixed value of V_{EB} was measured periodically and used to monitor the effects of the stress. Investigations of aligned silicon bipolar transistors with polysilicon contacts presented in [73] show an increase of the base current in proportion to $I_{EB}\sqrt{t_{stress}}$. This behavior is confirmed by the measured data reproduced in Fig. 7.21. These data also demonstrate the relaxation

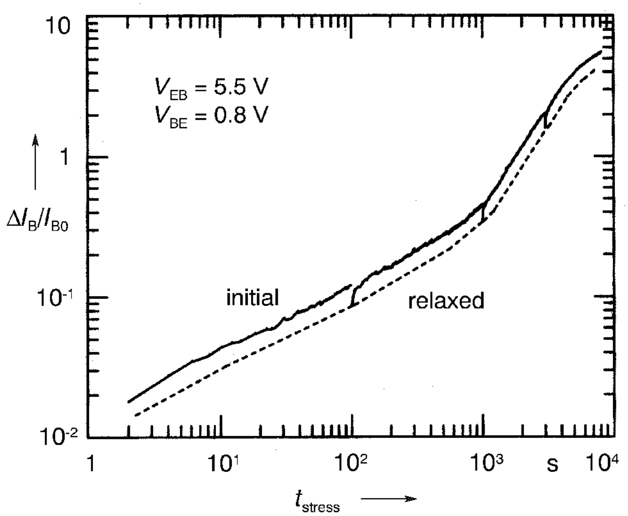


Fig. 7.21. Relative change of stress-induced excess base current versus stress time. The base current was allowed to relax for 10 min after 100 s, 1000 s and 3000 s (after [74])

of the device degradation: the stress-induced excess base current decreases immediately after the removal of the stress voltage, as was observed earlier in [72]. In [74], a quantitative model of this relaxation transient is presented.

There the transient is attributed to the reduction of stress-generated positive charge trapped in the oxide layer near the eb junction, caused by holes tunneling from oxide hole traps to silicon band states or SiO_2–Si interface traps. A spatial distribution of the oxide traps, from the SiO_2–Si interface into the oxide, was used to explain the experimentally observed logarithmic time dependence of the base-current relaxation. The relaxation of the stress-induced excess base current was found to be strongly reduced for stress voltages below 3 V; in [74] this threshold energy is interpreted as the energy required for hot holes to tunnel to oxide hole traps, which lie approximately 1.4 eV above the valence band edge of the oxide layer and thus about 2.9 eV below the silicon valence band edge.

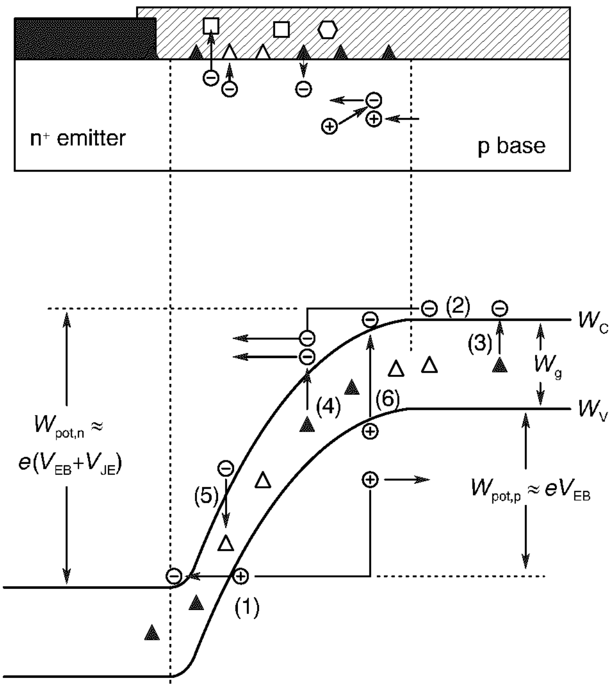


Fig. 7.22. Cross section and band scheme of reverse-biased eb diode, illustrating the creation of interface traps (open triangles) at weak bonds (filled triangles) and charging of oxide electron traps (open squares) and hole hole traps (open hexagons) by hot electrons and hot holes, in an oxide-covered eb junction under reverse-bias voltage stress (after [75]). (1) Interband tunneling, (2) electron injection, (3) thermal generation of electrons at weak bonds, (4) generation of electrons at weak bonds by impact ionization, (5) trapping of electrons and (6) generation of electron–hole pairs by impact ionization

It has been suggested in [75] that the kinetic energy of the charge carriers, rather than the stress voltage, is the fundamental parameter for device degradation. As has been pointed out in [75], the degradation of the current gain during reverse-bias stress at small values of V_{EB} is caused by primary hot holes, which are generated by interband tunneling in the vicinity of the metallurgical n^+p junction (Fig. 7.22, path (1)). Hot electrons might be injected into the reverse-biased junction either by injection of thermally generated electrons from the p-type region into the space charge region (Fig.

7.22, path (2)) or by impact ionization of hot holes (Fig. 7.22, path (6)). The possibility, that thermally generated carriers play a substantial role can be excluded by consideration of the temperature dependence of the results of constant-voltage stress experiments: owing to the exponential temperature dependence of the thermal generation rate, a strong increase of the stress damage with temperature would be expected if thermally generated carriers were important, in contradiction to what is observed experimentally. Owing to the small ionization coefficient of holes and the small potential energy of hot holes $W_{\mathrm{pot,p}} \approx eV_{\mathrm{EB}}$ (see Fig. 7.22), the generation of secondary electrons by impact ionization is also rather unlikely.

The temperature dependence of reverse-bias stress effects has been investigated in [76, 77]. For constant values of the reverse bias and stressing time, an increase of degradation with increasing temperature is generally observed, owing to the increase of reverse current, which is due to tunneling in the eb diode and increases with temperature due to the reduction of the bandgap. As the number of hot electrons varies in proportion to the reverse current, increased stress effects are expected if the temperature increases. However, this tendency is opposed by the fact that the electron mean free path decreases at higher temperatures. At a given reverse bias – and therefore at an approximately constant value of the electric field strength – the fraction of electrons which acquire sufficient energy to be able to cause damage therefore decreases. If the latter effect dominates, a decrease of reverse-bias stress effects with increasing temperature may result [76]. Defects created at the Si–SiO_2 interface at the periphery of the eb junction after aging were observed to substantially affect the base current noise of self-aligned polysilicon BJTs [78].

Reverse-bias stress experiments presented in [79] showed no effect of the Ge profile on the eb stress damage, suggesting that reliability of SiGe HBTs is comparable to that of silicon bipolar transistors.[17]

If the reverse stress current is due to tunneling in the reverse-biased eb sidewall diode, long stress times are required for a given value of stress charge. The reverse current at a given value of V_{EB} can be increased by injection of minority carriers across the forward-biased bc junction [75, 81]. However, it should be noted that the results of measurements under these conditions cannot be compared directly with measurements of the kind described above, since different degradation mechanisms apply: it is predominantly hot holes that may cause damage if the collector is left open, whereas electron injection

[17] The degradation of HBTs with a SiGe base layer after exposure to ionizing radiation was investigated in [80], where only a small degradation of the current gain was observed after irradiation to 2.0 Mrad. This led to the conclusion that the addition of a SiGe strained layer in the epitaxial base does not create any additional reliability risk. This conclusion was modified to a certain extent by the observation that some SiGe HBTs showed enhanced low-frequency noise after exposure to ionizing irradiation [80].

into the reverse-biased eb diode from the collector yields predominantly hot electrons. Furthermore, the spatial distribution of the stress current density will differ from that in stress measurements with the collector left open.

A reduction of the nonideal base current may possibly occur because of passivation of the interface traps with hydrogen atoms. These form Si–H complexes at the dangling bonds and make them electrically inactive. Furthermore, annealing is possible if one uses elevated temperatures and a significant forward bias of the junction [82]. During hot-carrier stress, bombardment of the Si–H complexes may free hydrogen ions (protons), which may diffuse through the lattice, form B–H complexes and thus add extra positive charges to the B^- ions incorporated into the lattice. This has the same consequence as a reduction of the acceptor concentration, and causes an increase of the base series resistance and of BV_{EBO}, and a decrease of the depletion layer capacitance [83].

Forward-Bias Stress

Transistor operation at high forward current densities in excess of $1\,\mathrm{mA/cm^2}$ has been shown to cause degradation of polysilicon-contacted bipolar transistors.[18] At small values of V_{BE}, a decrease of the current gain B_{N} has been observed, which is attributed to an increase in the number of SRH recombination centers at the Si SiO_2 interface, which terminates the eb space charge layer. At large values of V_{BE}, a decrease of the base current and hence an increase of the current gain has been observed. This is attributed to a decreased recombination velocity at the polysilicon emitter contact. Furthermore, a decrease of the emitter series resistance has been observed during high-current stress. A physical model that explains these effects by hydrogenation and dehydrogenation of electronic interface traps (which are also hydrogen traps) at the Si–SiO_2 interface and at the polysilicon-monocrystalline silicon interface is presented in [84].

7.4.2 Failure of Bipolar Devices due to Electrostatic Discharges

Device failure induced by electrostatic discharge (ESD) causes severe reliability problems for MOS integrated circuits, owing to the sensitivity of the gate dielectric. Bipolar integrated circuits used to be less delicate in this respect, but with the drastically reduced device dimensions typical of modern self-aligned bipolar technologies, a considerable increase in ESD sensitivity has been observed. Suitable protection techniques [85,86] are therefore necessary.

The determination of failure threshold levels of (aligned) semiconductor diodes and transistors subjected to pulsed voltages has been the subject of several publications that appeared during the last 40 years. Wunsch and

[18] See, for example, [84] and references therein.

Bell [87] applied rectangular pulses of varying height and pulse width directly to bipolar devices. The resulting failure levels were compared with the predictions of a simple thermal failure model based on an analytical solution of the time-dependent heat equation for a one-dimensional heat source of negligible extent. With the assumption of a minimum temperature difference ΔT_{f} necessary to produce device failure, Wunsch and Bell derived the following failure condition for rectangular voltage pulses:

$$P/A_{\mathrm{j}} = \sqrt{\pi \lambda \rho_{\mathrm{m}} c_{\mathrm{p}}}\, \Delta T_{\mathrm{f}} / \sqrt{t_{\mathrm{p}}}\,. \tag{7.15}$$

This relates the power P dissipated over the area A_{j} of the junction to the pulse width t_{p}, the value of ΔT_{f} and the material parameters λ (thermal conductivity), ρ_{m} (mass density) and c_{p} (specific heat).

Besides the approach of applying rectangular voltage pulses directly to the device, a different stress condition, represented by the (simplified) equivalent circuit of a charged person and his/her body contact resistance (Fig. 7.23) is widely used for the investigation of ESD reliability. This is the human-

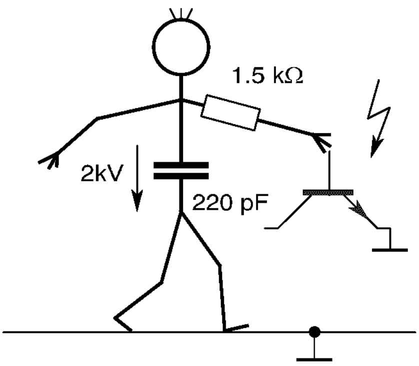


Fig. 7.23. Human-body model for the investigation of the sensitivity of devices with respect to electrostatic discharges

body model,which was introduced as a military standard. The test apparatus discharges a capacitor C (typically 100 pF $< C <$ 220 pF) precharged to a voltage $V_{\mathrm{C}}(0)$, via a resistor R, which is typically chosen to be of the order of 1 kΩ, and the device under test.

The most sensitive part of a self-aligned bipolar transistor with respect to electrostatic discharges is generally the eb sidewall diode, i.e. the region where the external base, defined by outdiffusion of the p^+ polysilicon, is linked to the internal base, usually defined by implantation. If a large reverse bias is applied to the diode, breakdown due to tunneling and carrier multiplication occurs. The threshold for this process is given by the eb breakdown voltage BV_{EBO}, which is determined by the doping concentrations in the sidewall diode. This breakdown leads to a considerable current that flows through the sidewall diode, and results in a voltage drop in the resistor R of the test apparatus, which limits the voltage drop across the device to values below the threshold voltages of alternative current paths, such as that provided by oxide breakdown.

If a series of discharges, starting with a small initial value of $V_C(0)$ is applied to a self-aligned eb diode, only small changes of the input characteristics are observed initially; these can be attributed to device degradation caused by reverse-bias stress (Fig. 7.24). If $V_C(0)$ exceeds a certain value V_F, however, the

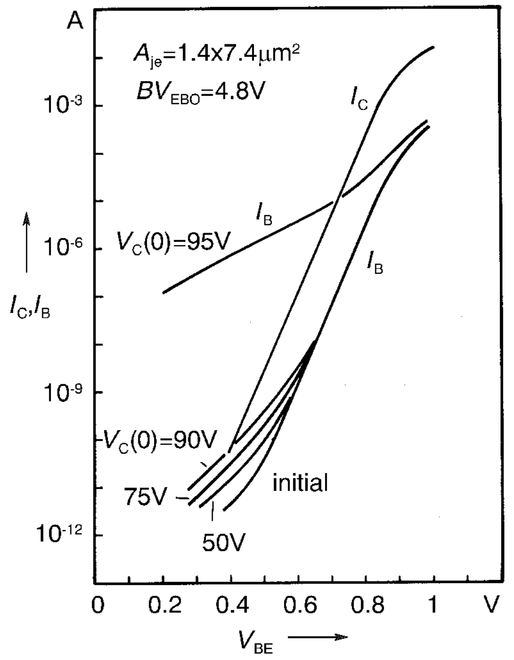


Fig. 7.24. Damage to the eb diode of a self-aligned double-polysilicon bipolar transistor caused by discharges applied in accordance with the human-body model. All input characteristics were determined successively on the same device

input characteristics change drastically and the eb diode loses its rectifying property. Device failure is therefore only observed if the failure threshold voltage V_F is exceeded. For an accurate determination of the failure threshold, the following procedure can be applied. For each value of the voltage $V_C(0)$, the base current is measured at $V_{BE} = 0.5\,\mathrm{V}$ for an ensemble of N transistors before and after the discharge, giving the values $I_{B0,i}$ and $I_{B1,i}$, respectively. From these data, the quantity

$$\Gamma = \frac{1}{N}\sum_{i=1}^{N} \log_{10}(I_{B0,i}/I_{B1,i})$$

is computed. Since the occurrence of ESD damage leads to a substantial increase in the base current by approximately four orders of magnitude for the value of V_{BE} chosen, a plot of Γ versus $V_C(0)$ should show a pronounced

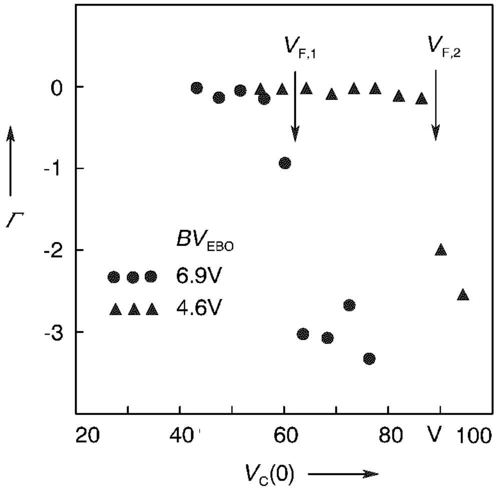


Fig. 7.25. Definition of failure threshold for two transistors with identical layout but different values of BV_{EBO}, in terms of the parameter Γ defined in the text

kink if there exists a well-defined failure threshold. This is indeed observed (Fig. 7.25), and the kink observed in this representation is used to determine the failure threshold. Such investigations have been performed for transistors fabricated with different technologies and with different layouts. However, the failure threshold did not vary in proportion to the emitter area, as predicted by the model of Wunsch and Bell, but in proportion to the length L^* of the emitter side adjacent to the base contact (Fig. 7.26).

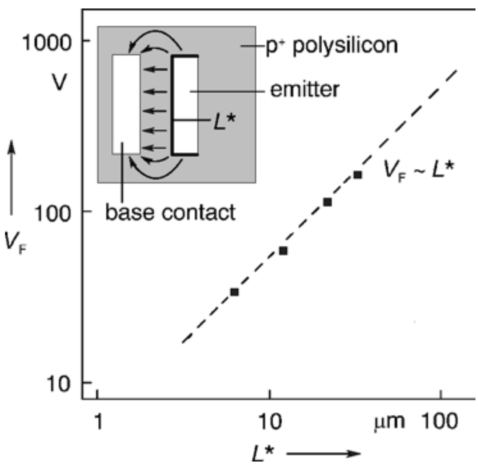


Fig. 7.26. Failure threshold V_F versus length L^* of that part of the perimeter of the emitter area adjacent to the base contact (length L plus twice the width of the narrow emitter stripe)

This behavior can be explained by a localized overheating of the device in the course of the discharge: as the discharge occurs predominantly in the sidewall adjacent to the base contact, the energy deposited per unit volume will vary in inverse proportion to L^*. The maximum temperature increase during an electrostatic discharge will be investigated in the following.

Temperature Stress During an Electrostatic Discharge

To study the impact of the stress conditions (the voltage $V_C(0)$, capacitance C and series resistance R of the test apparatus) and the effect of device-specific parameters (the eb breakdown voltage BV_{EBO}, space charge layer width d_j, heat capacity c_p, mass density ρ_m and thermal conductivity λ) on the value of the maximum temperature that occurs in the course of the discharge, the time-dependent heat equation for an extended time-dependent heat source has to be solved. Details of the calculation are presented in Appendix F; it relies on the following assumptions:

1. The voltage drop that occurs at the eb diode in the course of the discharge is assumed to equal the eb breakdown voltage BV_{EBO}. The power dissipated in the device may then be estimated as $p(t) = BV_{EBO}\, i(t)$, resulting in
$$W \approx BV_{EBO} \int_0^\infty i(t)\,dt \approx BV_{EBO} C [V_C(0) - BV_{EBO}] \tag{7.16}$$
for the energy deposited in the device during the discharge. This assumption is justified by the observation that the failure threshold V_F varies approximately in proportion to BV_{EBO} (Fig. 7.25).
2. Simplified boundary conditions are used. The heat source is completely embedded in the silicon crystal, which is considered to be of infinite extent.
3. The influence of a heat sink (such as the back surface of the wafer, held at constant temperature) is neglected. This assumption restricts the validity of the results to a short time interval of length ℓ^2/α, where ℓ denotes the distance to the heat sink and $\alpha = \lambda/(\rho_m c_p)$ the thermal diffusivity (see Appendix F). For a typical wafer thickness ($\ell = 400\,\mu m$) and $\alpha \approx 1\,cm^2/s$, this condition means that the results are valid within a time interval of the order of milliseconds, which is much larger than the time constant of the discharge.
4. The thermal conductivity λ is assumed to be constant, independent of temperature, in contrast to the well-known decrease of λ from 1.56 W/(cm K) at room temperature (300 K) to 0.31 W/(cm K) at 1000 K. Using this assumption, the heat equation may be treated analytically [88]. For the estimates given in this section, we have used the value of the thermal conductivity, at 500 K, i.e. $\lambda = 0.4$ W/(cm K). This approach is justified because the greatest changes in thermal conductivity occur in the low-temperature range from 300 K to 500 K.

5. The rise of the pulse is not considered, since the capacitance of the eb diode is smaller than the capacitance of the test apparatus by several orders of magnitude. However, it should be noted that in practice, series inductances may cause a substantial deviation of the real waveform from the assumed exponential decay, with the consequence of modified failure threshold levels.
6. Quantitative results are obtained by application of the general results to the special case of a homogeneous cuboid heat source of length L, width B and thickness $D = d_{\mathrm{j}}$. This model assumption represents the special case where all the power dissipated in the device during the discharge produces heat in the space charge layer. Stronger localization of the discharge may however, be studied by a reduction of the extent of the discharge region considered.

Despite these simplifying assumptions, the model is instructive in that it clarifies several aspects of the problem, as the heat equation can be treated analytically. The calculations show, in particular, that the maximum temperature that occurs during the discharge is essentially determined by two different time constants: the time constant of the electric discharge, $\tau = RC$, and the time constant associated with the temperature relaxation, $\tau^* = \rho_{\mathrm{m}} c_{\mathrm{p}} D^2/\lambda$, where D denotes the minimum side length of the cuboid considered. The external time constant τ describes the influence of the test apparatus, while the internal time constant τ^* is determined by the properties of the device under test. The maximum overtemperature ΔT_{m} that occurs during the discharge can be shown to be of the form

$$\Delta T_{\mathrm{m}} = \Delta T_0 \, F(\eta) \, , \tag{7.17}$$

where

$$\Delta T_0 = \frac{BV_{\mathrm{EBO}}[\,V_{\mathrm{C}}(0) - BV_{\mathrm{EBO}}\,]}{c_{\mathrm{p}} \rho_{\mathrm{m}} LBD} \tag{7.18}$$

describes the adiabatic temperature increase that would be observed in the case of negligible thermal conduction. The function $F(\eta)$ is a geometry-specific function, which for given values of B/D and L/D depends only on the ratio $\eta = \tau^*/\tau$ of the two time constants. Figure 7.27 shows $F_{B/D,L/D}(\eta)$ for different values of the parameters B/D and L/D. The model calculations show:

1. The series resistance R of the test apparatus has (for typical values of R in the kΩ range) a negligible influence on the energy deposited in the device during the discharge. However, owing to its impact on the time constant $\tau = RC$ and therefore on the parameter η, the value of R determines ΔT_{m} to a great extent.

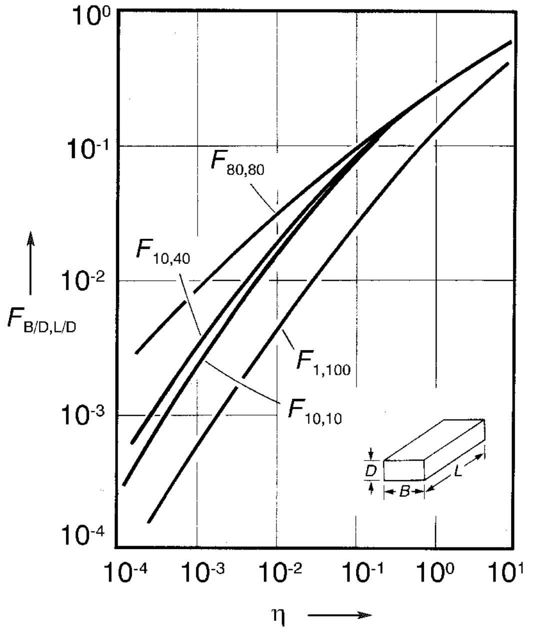


Fig. 7.27. Correction factors $F_{B/D,L/D}$, which take account of the temperature relaxation during an electrostatic discharge in a reverse-biased junction

2. A decrease of the device dimensions causes a decrease of the size of the discharge region, which again causes an increase of the energy deposited per unit volume. This effect is compensated to a large extent by better coupling of the smaller device to the heat sink. A crude estimate of the combined effect may be obtained from the following considerations. In the region of small η, the function $F(\eta)$ is approximately proportional to a power of η (see Fig. 7.27):

$$F(\eta) \sim \eta^{\theta} \sim D^{2\theta} \quad \text{where} \quad 0.5 \leq \theta \leq 0.8 \,.$$

If D is considered as a scaling factor, i.e. if D is decreased while the ratios B/D and L/D are fixed, the adiabatic temperature increase varies according to

$$\Delta T_0 \sim \frac{BV_{\mathrm{EBO,new}}}{BV_{\mathrm{EBO,old}}} \frac{1}{D^3} \,,$$

i.e. the maximum temperature that occurs in the course of the discharge varies in proportion to[19]

$$\Delta T_{\mathrm{m}} \sim \frac{BV_{\mathrm{EBO,new}}}{BV_{\mathrm{EBO,old}}} D^{2\theta-3} \,. \tag{7.19}$$

[19]This behavior compares (astonishingly) well with the behavior predicted by the model of Wunsch and Bell ($\Delta T_{\mathrm{m}} \sim A \sim 1/D^2$), which is based on a one-dimensional calculation for a δ-shaped heat source.

7.5 References

[1] S.M. Sze (Ed.). *VLSI Technology*. McGraw-Hill, New York, 1985.

[2] K. Ueno, H. Goto, E. Sugiyama, H. Tsunoi. A sub-40 ps ECL circuit at a switching current of 1.28 mA. *IEDM Tech. Dig.*, pp. 371–374, 1987.

[3] C.T. Chuang, P.F. Lu. On the scaling property of trench isolation capacitance for advanced high-performance ECL circuits. *IEDM Tech. Dig.*, pp. 799–802, 1989.

[4] P.F. Lu, C.T. Chuang. On the scaling of trench isolation capacitance for advanced high-performance ECL circuits. *IEEE Trans. Electron Devices*, 37(10):2270–2274, 1990.

[5] T.F. Meister, H. Schäfer, M. Franosch, W. Molzer, K. Aufinger, U. Scheler, C. Walz, M. Stolz, S. Boguth, J. Böck. SiGe base bipolar technology with 74 GHz $f_{\max}$ and 11 ps gate delay. *IEDM Tech. Dig.*, pp. 739–742, 1995.

[6] J.N. Burghartz, R.C. McIntosh, C.L. Stanis. A low-capacitance bipolar/BiCMOS isolation technology, part I – concept, fabrication process, and characterization. *IEEE Trans. Electron Devices*, 41(8):1379–1387, 1994.

[7] J.N. Burghartz, A.O. Cifuentes, J.D. Warnock. A low-capacitance bipolar/BiCMOS isolation technology, part II - circuit performance and device self-heating. *IEEE Trans. Electron Devices*, 41(8):1388–1395, 1994.

[8] J. Graul, H. Murrmann, A. Glasl. High-performance transistors with arsenic-implanted polysil emitters. *IEEE J. Solid-State Circuits*, 11(8):491–495, 1976.

[9] T. Sakai, Y. Kobayashi, H. Yamauchi, M. Sato, T. Makino. High speed bipolar ICs using super self-aligned process technology. *Proc. 12th Conf. Solid-State Devices, Tokyo, Japan. J. Appl. Phys.*, 20 Supplement 20-1:155–159, 1980.

[10] T.H. Ning, R.D. Isaac. Effect of emitter contact on current gain of silicon bipolar devices. *IEEE Trans. Electron Devices*, 27(11):2051–2055, 1980.

[11] H. Nakashiba, I. Ishida, K. Aomura, T. Nakamura. An advanced PSA technology for high-speed bipolar LSI. *IEEE J. Solid-State Circuits*, 15(4):455–459, 1980.

[12] D.D. Tang, P.M. Solomon, T.H. Ning, R.D. Isaac, R.E. Burger. 1.25 µm deep-groove-isolated self-aligned bipolar circuits. *IEEE J. Solid-State Circuits*, 17(5):925–931, 1982

[13] T.H. Ning, D.D. Tang. Bipolar trends. *Proc. IEEE*, 74(12):1669–1677, 1986.

[14] T.H. Ning. History and future perspectives of the modern silicon bipolar transistor. *IEEE Trans. Electron Devices*, 48(11):2485–2491, 2001.

[15] J.D. Hayden, J.D. Burnett, J.R. Pfiester, M.P. Woo. A new technique for forming a shallow link base in a double polysilicon bipolar transistor. *IEEE Trans. Electron Devices*, 41(1):63–68, 1994.

[16] K. Nakazato, T. Nakamura, T. Okabe, M. Nagata. Characteristics and scaling properties of npn transistors with a sidewall base contact structure. *IEEE J. Solid-State Circuits*, 20(2):248–252, 1985.

[17] S. Konaka, E. Yamamoto, K. Sakuma, Y. Amemiya, T. Sakai. A 20-ps Si bipolar IC using advanced super self-aligned process technology with collector ion implantation. *IEEE Trans. Electron Devices*, 36:1370–1375, 1989.

[18] P. Palestri, C. Fiegna, L. Selmi, M.S. Peter, G.A.M. Hurkx, J.W. Slotboom, E. Sangiorgi. A better insight into the performance of silicon BJT's featuring highly nonuniform collector doping profiles. *IEEE Trans. Electron Devices*, 47(5):1044 1051, 2000.

[19] J.N. Burghartz, J.Y.C. Sun, S.R. Mader, C.L. Stanis, B.J. Ginsberg. Perimeter and plug effects in deep sub-micron polysilicon emitter bipolar transistors. *Symp. VLSI Tech. Dig. Tech. Papers*, pp. 55–56, 1990.

[20] K. Ehinger, E. Bertagnolli, J. Weng, R. Mahnkopf, R. Köpl, H. Klose. Narrow BF_2 implantated bases for 35 GHz/24 ps high-speed Si bipolar technology. *IEDM Tech. Dig.*, pp. 459–462, 1991.

[21] T. Uchino, T. Shiba, T. Kikuchi, Y. Tamaki, A. Watanabe, Y. Kiyota. Very-high-speed silicon bipolar tranistors with in-situ doped polysilicon emitter and rapid vapor-phase doping base. *IEEE Trans. Electron Devices*, 42(3):406–412, 1995.

[22] J. Warnock, J.D. Cresler, K.A. Jenkins, T.-C. Chen, J.Y.-C. Sun, D.D. Tang. 50-GHz self-aligned bipolar transistors with ion-implanted base profiles. *IEEE Electron Device Lett.*, 11(10):475–477, 1990.

[23] T. Yamaguchi, S. Uppili, J.S. Lee, G.H. Kawamoto, T. Dosluoglu, S. Simpkins. Process and device characterization for a 30-GHz f_T submicrometer double poly-Si bipolar technology using BF_2-implanted base with rapid thermal process. *IEEE Trans. Electron Devices*, 40(8):1484 1495, 1993.

[24] S. Konaka, M. Ugajin, T. Matsuda. Deep submicrometer super self-aligned Si bipolar technology with 25.4 ps ECL. *IEEE Trans. Electron Devices*, 41(1):44–49, 1994.

[25] J. Böck, T.F. Meister, H. Knapp, K. Aufinger, M. Wurzer, R. Gabl, M. Pohl, S. Boguth, M. Franosch, L. Treitinger. 0.5 µm/60 GHz $f_{\max}$ implanted base Si bipolar technology. *IEEE BCTM Tech. Dig.*, pp. 160 163, 1998.

[26] J. Böck, H. Knapp, K. Aufinger, T.F. Meister, M. Wurzer, S. Boguth, L. Treitinger. High-performance implanted base silicon bipolar technology for RF applications. *IEEE Trans. Electron Devices*, 48(11):2514–2519, 2001.

[27] T. Shiba, Y. Tamaki, T. Onai, Y. Kiyota, T. Kure, T. Nakamura. A very small bipolar transistor technology with sidewall polycide base electrode for ECL-CMOS LSI's. *IEEE Trans. Electron Devices*, 43(9):1357–1363, 1996.

[28] R.C. Taft, C.S. Lage, J.D. Hayden, H.C. Kirsch, J.-H. Lin, D.J. Denning, F.B. Shapiro, D.E. Bockelman, N. Camillieri. The SCC BJT: a high-performance bipolar transistor compatible with high-density deep-submicrometer BiCMOS SRAM technologies. *IEEE Trans. Electron Devices*, 42(7):1277–1286, 1995.

[29] T. Onai, E. Ohue, M. Tanabe, K. Washio. 12-ps ECL using low-base-resistance Si bipolar transistor by self-aligned metal/IDP technology. *IEEE Trans. Electron Devices*, 44(12):2207–2212, 1997.

[30] T. Shiba, T. Uchino, K. Ohnishi, Y. Tamaki. In situ phosphorous-doped polysilicon emitter technology for very high-speed small emitter bipolar transistors. *IEEE Trans. Electron Devices*, 43(6):889–897, 1996.

[31] D.L. Harame, J.H. Comfort, J.D. Cressler, E.F. Crabbe, J.Y.-C. Sun, B.S. Meyerson, T. Tice. Si/SiGe epitaxial-base transistors – part I: materials, physics, and circuits. *IEEE Trans. Electron Devices*, 42(3):455–468, 1995.

[32] D.L. Harame, J.H. Comfort, J.D. Cressler, E.F. Crabbe, J.Y.-C. Sun, B.S. Meyerson, T. Tice. Si/SiGe epitaxial-base transistors – part II: process integration and analog applications. *IEEE Trans. Electron Devices*, 42(3):469–482, 1995.

[33] D.L. Harame, B.S. Meyerson. The early history of IBM's SiGe mixed signal technology. *IEEE Trans. Electron Devices*, 48(11):2555–2567, 2001.

[34] Y. Kiyota, T. Udo, T. Hashimoto, A. Kodama, H. Shimamoto, R. Hayami, E. Ohue, K. Washio. HCl-free selective epitaxial Si–Ge growth by LPCVD for high-frequency HBTs. *IEEE Trans. Electron Devices*, 49(5):739–745, 2002.

[35] J.N. Burghartz, S.R. Mader, B.J. Ginsberg, B.S. Meyerson, J.M.C. Stork, C.L. Stanis, J.Y.-C. Sun, M.R. Polcari. Self-aligned bipolar epitaxial base n–p–n transistors by selective epitaxy emitter window (SEEW) technology. *IEEE Trans. Electron Devices*, 38(2):378 385, 1991.

[36] F. Sato, T. Tatsumi, T. Hashimoto, T. Tashiro. A super self-aligned selectively grown SiGe base (SSSB) bipolar transistor fabricated by cold-wall type UHV/CVD technology. *IEEE Trans. Electron Devices*, 41(8):1373–1378, 1994.

[37] E. Ohue, K. Oda, R. Hayami, K. Washio. A 7.7-ps CML using selective-epitaxial SiGe HBTs. *Proc. IEEE BCTM*, 1998:97–101, 1998.

[38] M. Kondo, K. Oda, E. Ohue, H. Shimamoto, M. Tanabe, T. Onai, K. Washio. Ultra-low-power and high-speed SiGe base bipolar transistors for wireless communication systems. *IEEE Trans. Electron Devices*, 45(6):1287 1294, 1998.

[39] K. Washio, M. Kondo, E. Ohue, K. Oda, R. Hayami, M. Tanabe, H. Shimamoto, T. Harada. A 0.2 µm self-aligned selective-epitaxial-growth SiGe HBT featuring 107-GHz $f_{\max}$ and 6.7-ps ECL. *IEEE Trans. Electron Devices*, 48(9):1989–1994, 2001.

[40] M. Kondo, H. Shimamoto, K. Washio. Variation in emitter diffusion depth of $TiSi_2$ formation on polysilicon emitters of Si bipolar transistors. *IEEE Trans. Electron Devices*, 48(9):2108–2117, 2001.

[41] K. Washio, E. Ohue, H. Shimamoto, K. Oda, R. Hayami, Y. Kiyota, M. Tanabe, M. Kondo, T. Hashimoto, T. Harada. A 0.2 µm 180-GHz-$f_{\max}$ 6.7 ps-ECL SOI/HRS self-aligned SEG SiGe HBT/CMOS technology for microwave and high-speed digital applications. *IEEE Trans. Electron Devices*, 49(2):271–278, 2002.

[42] M. Rydberg, U. Smith. Long-term stability and electrical properties of compensation doped poly-Si IC-resistors. *IEEE Trans. Electron Devices*, 47(2):417–426, 2000.

[43] H. Murrmann, D. Widmann. Current crowding on metal contacts to planar devices. *IEEE Trans. Electron Devices*, 16:1022–1024, 1969.

[44] H.H. Berger. Models for contacts to planar devices. *Solid-State Electron.*, 15:145–158, 1972.

[45] W.M. Loh, S.E. Swirhun, T.A. Schreyer, R.M. Swanson, K.C. Saraswat. Modeling and measurement of contact resistances. *IEEE Trans. Electron Devices*, 34(3):512–524, 1987.

[46] J.N. Burghartz, M. Soyuer, K.A. Jenkins. Integrated RF and microwave components in BiCMOS technology. *IEEE Trans. Electron Devices*, 43(9):1559 1570, 1996.

[47] R. Aparicio, A. Hajimiri. Capacity limits and matching properties of integrated capacitors. *IEEE J. Solid-State Circuits*, 37(3):384–393, 2002.

[48] M.H. Norwood, E. Shatz. Voltage variable capacitor tuning: A review. *Proc. IEEE*, 56(5):788–798, 1968.

[49] M. Soyuer, R.G. Meyer. High-frequency phase-locked loops in monolithic bipolar technology. *IEEE J. Solid-State Circuits*, 24(3):787–795, 1989.

[50] N.M. Nguyen, R.G. Meyer. Si IC-compatible inductors and LC passive filters. *IEEE J. Solid-State Circuits*, 25(4):1028–1031, 1990.

[51] D.T.S. Cheung, J.R. Long, R.A. Hadaway, D.L. Harame. Monolithic transformers for silicon RF IC design. *IEEE BCTM Tech. Dig.*, pp. 105–108, 1998.

[52] J.N. Burghartz, D.C. Edelstein, M. Soyuer, K.A. Jenkins. RF circuit design aspects of spiral inductors on silicon. *IEEE J. Solid-State Circuits*, 33(12):2028–2034, 1998.

[53] K.B. Ashby, I.A. Koullias, W.C. Finley, J.J. Bastek, S. Moinian. High Q inductors for wireless applications in a complementary silicon bipolar process. *IEEE J. Solid-State Circuits*, 31(1):4–9, 1996.

[54] J.R. Long, M.A. Copeland. The modeling, characterization, and design of monolithic inductors for silicon RF IC's. *IEEE J. Solid-State Circuits*, 32(3):357–368, 1997.

[55] A.M. Niknejad, R.G. Meyer. Analysis, design, and optimzation of spiral inductors and transformers for Si RF IC's. *IEEE J. Solid-State Circuits*, 33(10):1470–1481, 1998.

[56] C.P. Yue, S.S. Wong. Physical modeling of spiral inductors on silicon. *IEEE Trans. Electron Devices*, 47(3):560–568, 2000.

[57] S. Jenei, B.K.J.C. Nauwelaers, S. Decoutere. Physics-based closed-form inductance expression for compact modeling of integrated spiral inductors. *IEEE J. Solid-State Circuits*, 37(1):77–80, 2002.

[58] H. Lakdawala, X. Zhu, H. Luo, S. Santhanam, L.R. Carley, G.K. Fedder. Micromachined high-Q inductors in a 0.18 µm copper interconnect low-k dielectric CMOS process. *IEEE J. Solid-State Circuits*, 37(3):394–403, 2002.

[59] E. Chiprout. Interconnect and substrate modeling and analysis: an overview. *IEEE BCTM Tech. Dig.*, pp. 158–165, 1997.

[60] A.L.L. Pun, T. Yeung, J. Lau, F.J.R. Clement, D.K. Su. Substrate noise coupling through planar spiral inductor. *IEEE J. Solid-State Circuits*, 33(6):877–884, 1998.

[61] C.P. Yue, S.S. Wong. On-chip spiral inductors with patterned ground shields for Si-based RF IC's. *IEEE J. Solid-State Circuits*, 33(5):743–752, 1998.

[62] S.-M. Yim, T. Chen, K.K. O. The effects of a ground shield on the characteristics and performance of spiral inductors. *IEEE J. Solid-State Circuits*, 37(2):237–244, 2002.

[63] P.-F. Lu, J.D. Warnock. On the perimeter base leakage of double-poly self-aligned pnp transistors. *IEEE Trans. Electron Devices*, 39(12):2823–2826, 1992.

[64] N.E. Moiseiwitsch, P. Ashburn. The benefits of fluorine in pnp polysilicon emitter bipolar transistors. *IEEE Trans. Electron Devices*, 41(7):1249–1256, 1994.

[65] P.-F. Lu, J.D. Warnock, J.D. Cressler, K.A. Jenkins, K.-Y. Toh. The design and optimization of high-performance double-poly self-aligned pnp technology. *IEEE Trans. Electron Devices*, 38(6):1410–1418, 1991.

[66] T. Yamaguchi, T.M. Archer, R.E. Johnston, J.S. Lee. Process and device optimization of an analog complementary bipolar IC technology with 5.5-GHz f_T pnp transistors. *IEEE Trans. Electron Devices*, 41(6):1019–1026, 1994.

[67] T. Onai, E. Ohue, Y. Idei, M. Tanabe, H. Shimamoto, K. Washio, T. Nakamura. Self-aligned complementary bipolar technology for low-power dissipation and ultra-high-speed LSI's. *IEEE Trans. Electron Devices*, 42(3):413–418, 1995.

[68] M.C. Wilson, P.H. Osborne, S. Nigrin, S.B. Goody, J. Green, S.J. Harrington, T. Cook, S. Thomas, A.J. Manson, A. Madni. Process HJ: a 30 GHz npn and 20 GHz pnp complementary bipolar process for high linearity RF circuits. *IEEE BCTM Tech. Dig.*, pp. 164–167, 1998.

[69] R. Bashir, F. Hebert, J. DeSantis, J.M. McGregor, W. Yindeepol, K. Brown, F. Moraveji, T.B. Mills, A. Sadovnikov, J. MacGinty, P. Hopper, R. Sabsowitz, M. Khidr, T. Krakowski, L. Smith, R. Razouk. A complementary bipolar technology family with a vertically integrated pnp for high-frequency analog applications. *IEEE Trans. Electron Devices*, 48(11):2525–2534, 2001.

[70] H. Nii, T. Yamada, K. Inoh, T. Shino, S. Kawanaka, M. Yoshimi, Y. Katsumata. A novel lateral bipolar transistor with 67 GHz $f_{\max}$ on thin-film SOI for RF analog applications. *IEEE Trans. Electron Devices*, 47(7):1536–1541, 2000.

[71] T. Shino, S. Yoshitomi, H. Nii, S. Kawanaka, K. Inoh, T. Yamada, T. Fuse, Y. Katsumata, M. Yoshimi, S. Watanabe, J.-I. Matsunaga. Analysis on high-frequency characteristics of SOI lateral BJTs with self-aligned external base for 2-GHz RF applications. *IEEE Trans. Electron Devices*, 49(3):414–421, 2002.

[72] D.D. Tang, E. Hackbarth. Junction degradation in bipolar transistors and the reliability imposed constraints to scaling and design. *IEEE Trans. Electron Devices*, 35(12):2101–2107, 1988.

[73] J.D. Burnett, C. Hu. Modeling hot-carrier effects in polysilicon emitter bipolar transistors. *IEEE Trans. Electron Devices*, 35(12):2238–2244, 1988.

[74] A. Neugroschel, C.-T. Sah, M.S. Carroll, K.G. Pfaff. Base current relaxation transient in reverse emitter–base bias stressed silicon bipolar junction transistors. *IEEE Trans. Electron Devices*, 44(5):792–800, 1997.

[75] A. Neugroschel, C.-T. Sah, M.S. Carroll. Degradation of bipolar transistor current gain by hot holes during reverse emitter–base bias stress. *IEEE Trans. Electron Devices*, 43(8):1286 1290, 1996.

[76] C.-J. Huang, T.A. Grotjohn, J. Sun, D.K. Reinhard, C.-C.W. Yu. Temperature dependence of hot-electron degradation in bipolar transistors. *IEEE Trans. Electron Devices*, 40(9):1669–1674, 1993.

[77] H.S. Momose, H. Iwai. Analysis of the temperature dependence of hot-carrier induced degradation in bipolar transistors for BiCMOS. *IEEE Trans. Electron Devices*, 41(6):978–987, 1994.

[78] A. Mounib, F. Balestra, N. Mathieu, J. Brini, G. Ghibaudo, A. Chovet, A. Chantre, A. Nouailhat. Low-frequency noise sources in polysilicon emitter BJTs: influence of hot-electron-induced degradation and post-stress recovery. *IEEE Trans. Electron Devices*, 42(9):1647–1652, 1995.

[79] U. Gogineni, G. Niu, S.J. Mathew, J.D. Cressler, D.C. Ahlgren. Comparison of current gain and low-frequency noise degradation by hot electrons and hot holes under reverse eb stress in UHV/CVD SiGe HBT's. *Proc. IEEE BCTM*, pp. 172–175, 1998.

[80] J.A. Babcock, J.D. Cressler, L.S. Vempati, S.D. Clark, R.C. Jaeger, D.L. Harame. Ionizing radiation tolerance and low-frequency noise degradation in UHV/CVD SiGe HBT's. *IEEE Electron Device Lett.*, 16(8):351–353, 1995.

[81] F. Ingvarsson, L.-A. Ragnarsson, P. Lundgren, K.O. Jeppson. Stress and recovery transients in bipolar transistors and MOS structures. *Proc. IEEE Int. Conf. Microelectron. Test Struct.*, 12:173–178, 1999.

[82] H. Wurzer, R. Mahnkopf, H. Klose. Annealing of degraded npn-transistors – mechanisms and modeling. *IEEE Trans. Electron Devices*, 41(4):533–538, 1994.

[83] D. Quan, P.K. Gopi, G.J. Sonek, G.P. Li. Hot carrier induced bipolar transistor degradation due to base dopant compensation by hydrogen: theory and experiment. *IEEE Trans. Electron Devices*, 41(10):1824–1830, 1994.

[84] M.S. Carroll, A. Neugroschel, C.-T. Sah. Degradation of silicon bipolar junction transistors at high forward current densities. *IEEE Trans. Electron Devices*, 44(1):110–117, 1997.

[85] S.H. Voldmann. The state of the art of electrostatic discharge proection: physics, technology, circuits, design, simulation and scaling. *IEEE J. Solid-State Circuits*, 34(9):1272–1282, 1999.

[86] J.E. Vinson, J.J. Liou. Electrostatic discharge in semiconductor devices: protection techniques. *Proc. IEEE*, 88(12):1878–1900, 2000.

[87] D.C. Wunsch, R.R. Bell. Determination of threshold failure levels of semiconductor diodes and transistors due to pulse voltages. *IEEE Trans. Nucl. Sci.*, NS-15:244–259, 1968.

[88] H.S. Carslaw, J.C. Jäger. *Conduction of Heat in Solids*. Clarendon, Oxford, 1959.

8 Applications

In this chapter, four important applications of high-frequency bipolar transistors – ECL digital circuits, high-speed optical data transmission systems, RF circuits and BiCMOS circuits – are briefly considered.

8.1 Emitter-Coupled Logic

Emitter-coupled logic (ECL) is a bipolar-digital-circuit technique that is important in high-speed logic and has gate delays of the order of 10 ps. The static power dissipation associated with this technique prevents ECL from being competitive with CMOS for low-frequency operation. The situation changes at high frequencies, however, as the power dissipated in a CMOS gate increases in proportion to the switching frequency: as a consequence, ECL circuits operated in the GHz regime may even show a smaller power consumption than their CMOS counterparts. Another advantage of the ECL circuit concept is that lines are easily terminated. Furthermore, less noise is produced by ECL logic owing to the balanced current steering and the small voltage swing.

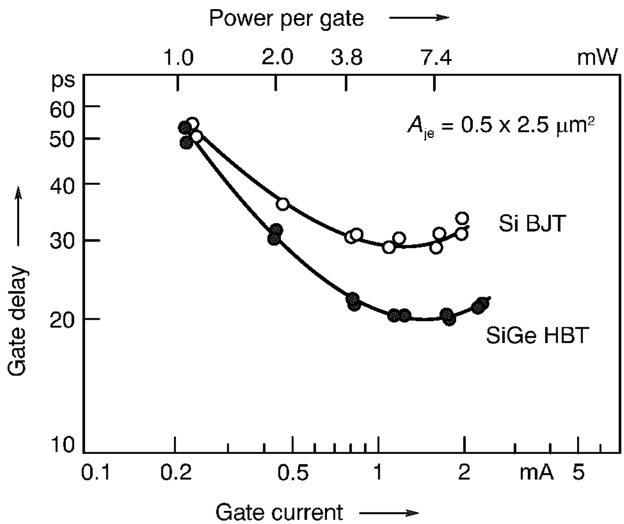


Fig. 8.1. Comparison of the gate delays of unloaded ECL ring oscillators (voltage swing V_S = 500 mV) as a function of bias current for a SiGe HBT and a Si BJT of comparable layout and doping profile (after [1])

Figure 8.1 shows power–delay data for ECL inverters constructed from comparable Si BJTs and SiGe HBTs. For small values of the gate current, the gate delay is determined to a large extent by the time required to charge and discharge the depletion capacitances (see Sect. 3.10). As the introduction of a SiGe base layer does not affect these capacitances, both technologies show approximately the same gate delay in this situation. At larger values of the gate current, the gate delay is determined essentially by the charge storage in the diffusion capacitance, and hence by the forward transit time τ_f. This parameter is, however, substantially reduced in SiGe HBTs, which thus show a minimum gate delay that lies well below the value achievable with Si BJTs.

8.1.1 Single-Ended, Differential and Feedback ECL

ECL digital circuits (Fig. 8.2) are derived from the bipolar differential stage considered in Sects. 6.7 and 6.10. If operated in the nonlinear regime, this circuit can switch the gate current I_{EE} from one of the two branches of the differential stage to the other. In the classical ECL10K series, the gate current I_{EE} is defined by a resistor R_E of approximately 1 kΩ. This simple approach results, however, in an unfavorable voltage transfer characteristic [2], which is in addition sensitive to variations of temperature and of the supply voltage. Improved approaches therefore employ transistor current sources driven with a special reference voltage source, the "bias driver" [3], to fix the value of the gate current. The bias voltage generated by the driver shows a predefined temperature dependence, designed to provide a voltage transfer characteristic that is independent of temperature.

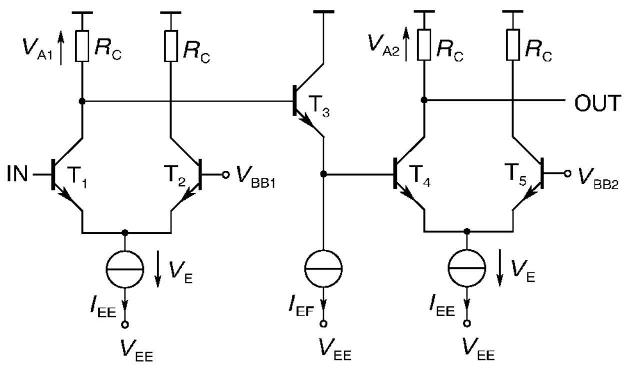


Fig. 8.2. ECL digital circuits

Switching of an ECL gate occurs if the input voltage difference changes its sign. This can be achieved in several ways. One approach is to connect one input of the differential amplifier to a fixed reference voltage. This is the most widely employed technique and is commonly referred to as single-ended ECL. An alternative approach is to use two complementary signals to feed both inputs of the differential amplifier (differential ECL). In feedback ECL, the output of one branch is used to define the input voltage of the second branch of the differential stage.

Single-Ended ECL

In single-ended ECL, the reference input is held at a fixed potential V_{BB1}. The voltage swing V_{S} may not be chosen smaller than 500 mV, because of the need to provide acceptable noise immunity in the presence of temperature variations, supply voltage variations and process tolerances.[1]

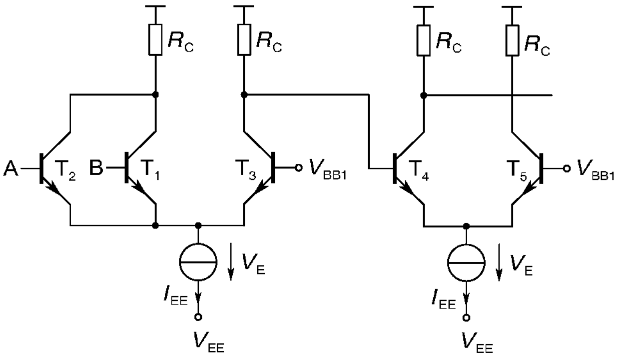


Fig. 8.3. CML digital circuits (emitter-coupled logic without emitter followers)

CML (current-mode logic) circuits feed the output of one differential stage directly into the input of the following stage, as illustrated in Fig. 8.3.[2] The problem with this technique is that the bc diode of one of the transistors of the differential stage is forward biased by the voltage swing V_{S} if the corresponding output node is HI. This degrades the switching time and is therefore unwanted. High-speed digital systems therefore generally employ ECL circuits, where emitter followers are added to the outputs (Fig. 8.2). These emitter followers shift the output voltage level by one diode voltage V_{BEon}. As the reference voltage $V_{\mathrm{BB1}} \approx -V_{\mathrm{S}}/2 - V_{\mathrm{BEon}}$ is shifted by the same amount, a logic swing of approximately 700 mV does not result in saturation of the transistors of the differential pair of the following stage. Furthermore, the emitter followers isolate the collector nodes of the gate from the load capacitance and provide the current gain necessary to drive the latter: with their small output impedance of the order of several ohms, emitter followers are well suited for driving loads with a characteristic impedance of 50 Ω.

For an emitter follower that drives a wiring capacitance C_{W}, an additional time constant [6],

$$\tau_{\mathrm{EF}} \approx 0.69 \left(\frac{1}{2\pi f_{\mathrm{T}}} + \frac{C_{\mathrm{W}} V_{\mathrm{T}}}{I_{\mathrm{EF}}} \right) ,$$

[1] A typical value of V_{S} in practical ECL circuits is 700 mV.

[2] This example combines an OR function with an inverter, and is redundant since two complementary signals (OR function and NOR function) are available from the first gate.

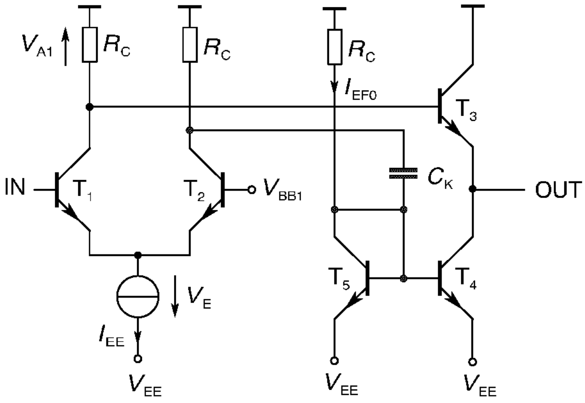


Fig. 8.4. An active-pull-down ECL inverter circuit (after [5])

has to be considered in the expression for the gate delay (see Sect. 3.10). The optimization of the power partition, that is of the ratio I_{EF}/I_{EE}, between the differential stage and the emitter followers using an automated approach has been investigated in [4]. A problem with emitter followers is that the current that is available to discharge the output node is limited by I_{EF} (Fig. 8.2). To overcome this problem, active pull-down circuits have been investigated. These are intended to increase the current available during a HI–LO switching transient at the output by capacitive coupling of the inverted node of the differential stage to the current source of the emitter follower; an example is illustrated in Fig. 8.4. Such circuits result in reduced HI–LO transition times, but at the cost of increased area consumption and therefore reduced packing density.

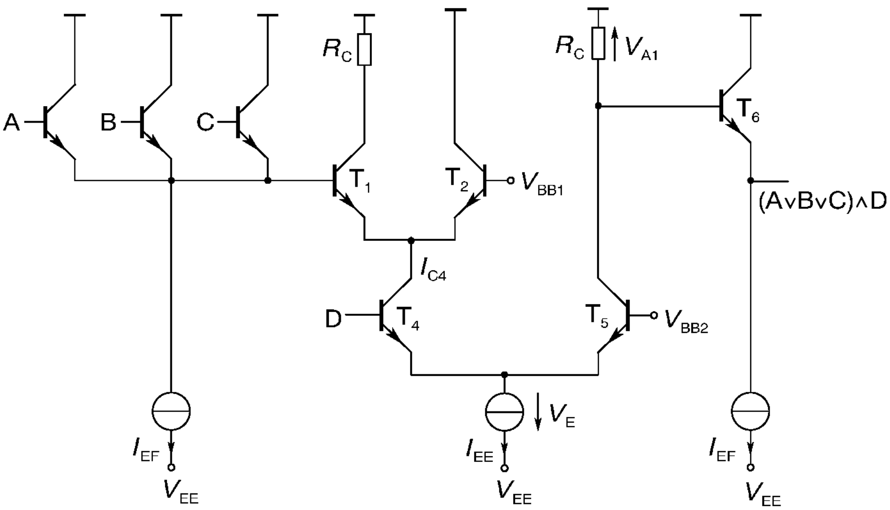


Fig. 8.5. Wired-OR connection and series gating in ECL circuits

Figure 8.5 shows two approaches for the reduction of power consumption in ECL circuits. A parallel connection of emitter followers at the input of the differential stage corresponds to an OR connection of the signals (this is called a wired-OR connection): if only one of the inputs A, B or C is HI, the base potential of T_1 will be HI.[3] Introducing an additional differential stage (T_4 and T_5) into one branch allows one to realize an additional AND/NAND function without increasing the gate current. This technique is commonly referred to as series gating.

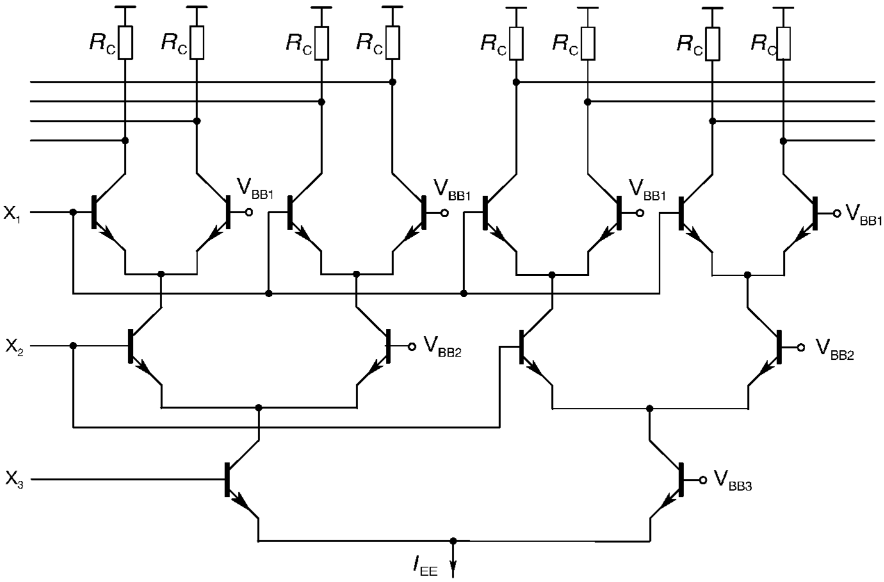


Fig. 8.6. 8-bit-wide decoder constructed using a three-level series gating technique

Figure 8.6 shows an example of how complex logic functions can be merged into one gate by use of the series gating technique. The circuit operates as an 8-bit-wide decoder with three input and eight output signals. For each three-bit word $X_3X_2X_1$ at the input, one of the output signals is HI, while all others are LO.

[3] In a wired-OR connection with n input transistors, the current I_{EF} is distributed between an unknown number of transistors ranging from 1 to n. This implies an uncertainty in the input HI and LO levels owing to the finite transconductance and its degradation by the emitter series resistance. The input voltage difference between the two limiting cases "1 transistor on" and "n transistors on" is

$$\Delta V_{IN} = V_T \ln(n) + (n-1)R_{EE'}I_{EF}/n .$$

This uncertainty has to be compensated by an increase of the voltage swing.

Differential ECL

Differential ECL exploits the fact that each ECL gate provides two complementary signals, which can be used as inputs to a subsequent differential stage. Owing to the conjugate input signals, the differential input voltage changes by $2V_S$. Differential ECL shows excellent noise immunity and minimum gate delay. In differential ECL, the voltage swing may be as low as 200 mV – this allows one to avoid emitter followers in some signal paths. A severe disadvantage is, however, that two complementary signals have to be routed for each logical signal, with the consequences of a complicated layout and a poor packing density. Furthermore, logic functions have to be realized by use of series gating techniques, which reduces power dissipation [7] but imposes limits on how far the supply voltage can be reduced.[4] For these reasons, complex ECL systems are generally realized as single-ended gates.

An example where differential ECL in combination with the series gating technique is beneficial is the XOR and XNOR gate that is obtained if a Gilbert cell (Fig. 6.23) is operated in the nonlinear region. With this approach the whole logic function is realized within one gate, i.e. only one current source is required. If a smaller supply voltage is required, series gating has to be avoided, and an XNOR function can be realized in single-ended ECL using the circuit depicted in Fig. 8.7. This circuit requires two current sources and therefore has a larger power consumption than the solution with series gating has.

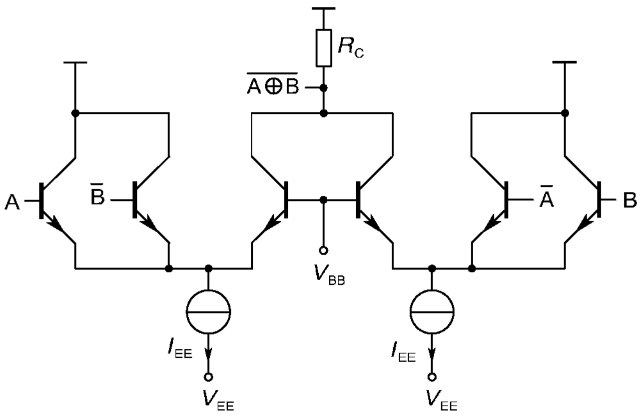


Fig. 8.7. Realization of XNOR function in single-ended ECL so as to allow a reduced supply voltage

[4] Each level requires a voltage drop of at least V_{BEon}. A minimum voltage drop $V_E \approx$ 500 mV across the current sources requires a minimum supply voltage of about 2.2 V for two-level series gating with CML input (see also [8]).

Feedback ECL

In feedback ECL circuits, such as the basic inverter shown in Fig. 8.8a, the input of transistor T_2 is connected to the output and is therefore complementary to the inverter input voltage V_{IN}. Such circuits are suitable, for example,

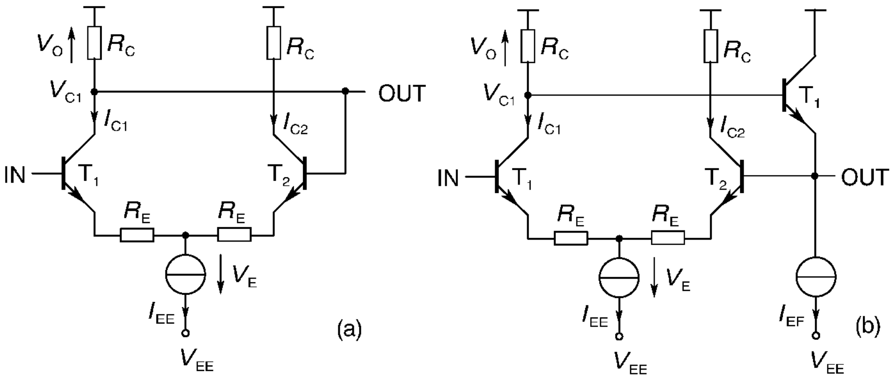


Fig. 8.8. (**a**) Basic feedback ECL (FECL) inverter circuit [9,10] and (**b**) modified feedback ECL inverter [11]

for the generation of differential signals from a single-rail input signal. Bandwidth limitations arise in the circuit of Fig. 8.8a because of saturation of the transistors of the differential pair and because of the input capacitance of transistor T_2. Increased switching speed is achieved by adding an emitter follower, which acts as a level shifter and low-impedance base drive, as shown in Fig. 8.8b. To illustrate the principle,we neglect the Early effect here and assume identical transistors. Kirchhoff's voltage law then gives

$$\begin{aligned} V_{IN} &= V_{C1} - V_{EF} - R_{BB'}I_{B2} - V_{BE2} - R_E I_{E2} \\ &\qquad + R_E I_{E1} + V_{BE1} + R_{BB'}I_{B1} \, . \end{aligned} \tag{8.1}$$

Making use of the identities

$$I_{C1} = I_S \exp\left(\frac{V_{BE1}}{V_T}\right) = -\frac{V_{C1}}{R_C}$$

and

$$I_{C2} = I_S \exp\left(\frac{V_{BE2}}{V_T}\right) = \frac{V_S + V_{C1}}{R_C} \, ,$$

where $V_S = R_C(I_{C1} + I_{C2})$ denotes the logic swing, we obtain

$$V_{BE1} - V_{BE2} = V_T \ln\left(\frac{-V_{C1}}{V_S + V_{C1}}\right) \; ;$$

taking account of the fact that $I_{C1} = -V_{C1}/R_C$ and $I_{C2} = -(V_S + V_{C1})/R_C$ therefore allows us to rewrite (8.1) in the form

$$V_{IN} = (1-2\gamma)V_{OUT} - 2\gamma V_{EF} - \gamma V_S + V_T \ln\left(\frac{-V_{OUT}-V_{EF}}{V_S+V_{OUT}+V_{EF}}\right) , \quad (8.2)$$

where $\gamma = [R_E+(R_E+R_{BB'})/B_N]/R_C$. Equation (8.2) may associate more than one value of V_{OUT} with a given value of V_{IN}, i.e. the voltage-transfer characteristic may show a hysteresis effect, as shown in Fig. 8.9 for two different values of γ. If the input voltage is increased from $(V_{IN}+V_{EF})/V_S = -1$, the

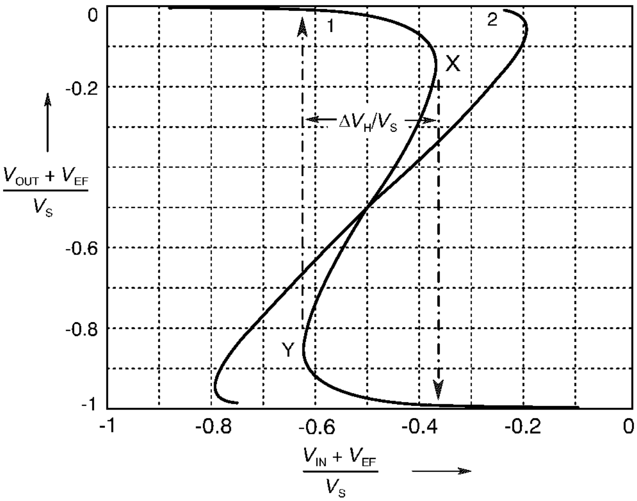


Fig. 8.9. Voltage transfer characteristic of a FECL gate for two different values of γ, indicating the hysteresis effect described in the text (after [11])

output voltage will follow the upper branch until point X is reached, where the output voltage will jump to the lower branch, which is followed for further increase of V_{IN}. If V_{IN} is then decreased again, the output voltage will follow the lower branch until point Y is reached, when it jumps back to the upper branch. The switching points can be obtained, after differentiation of (8.2), from the condition $dV_{IN}/dV_{OUT} = 0$. This leads to a quadratic equation for V_{OUT}, which gives

$$V_{OUT} + V_{EF} = \frac{V_S}{2}\left(-1 \pm \sqrt{1+X}\right) ,$$

where $X = 4V_T/[(2\gamma - 1)V_S]$. Subtraction of the two corresponding input voltages calculated from (8.2) yields the width of the hysteresis loop:

$$\frac{\Delta V_H}{V_S} = (1-\gamma)\sqrt{1+X} + \frac{2V_T}{V_S}\ln\left(\frac{1-\sqrt{1+X}}{1+\sqrt{1+X}}\right) .$$

A 6 Gb/s decision circuit based on a modified FECL gate has been demonstrated in [11].

8.1.2 Noise Margin

In a single-ended inverter, one input is tied to a fixed reference voltage, denoted here by V_{BB}, while the other is used as the signal input. In the limit $B_N \to \infty$, the following relation exists between the output voltage V_2 and the input voltage V_1 (see Sect. 6.7):

$$V_2 = -\frac{R_C I_{EE}}{1+\exp\left(\dfrac{V_1 - V_{BB}}{V_T}\right)} .$$

The maximum voltage gain is obtained for $V_1 = V_{BB}$ and given by

$$|dV_2/dV_1|_{max} = V_S/4V_T = A_D/2 ,$$

where $V_S = R_C I_{EE}$ denotes the logic swing. The unity gain points are obtained from the condition $|dV_2/dV_1| = 1$; the corresponding equation,

$$\left[1+\exp\left(\frac{V_1 - V_{BB}}{V_T}\right)\right]^2 = 2A_D \exp\left(\frac{V_1 - V_{BB}}{V_T}\right) ,$$

can be solved for V_1, with the result

$$V_{IL} = V_{BB} + V_T \ln\left[A_D - 1 + \sqrt{(A_D-1)^2 - 1}\right] ,$$

$$V_{IH} = V_{BB} + V_T \ln\left[A_D - 1 - \sqrt{(A_D-1)^2 - 1}\right] .$$

The corresponding inverter output signals are

$$V_{OH} = -\frac{V_S}{A_D + \sqrt{(A_D-1)^2 - 1}} ,$$

$$V_{OL} = -\frac{V_S}{A_D - \sqrt{(A_D-1)^2 - 1}} .$$

Using these relations, the results $NM_L = V_{IL} - V_{OL}$ and $NM_H = V_{OH} - V_{IH}$ are easily calculated. Owing to the symmetry of the voltage transfer characteristic, $NM_H = NM_L$ are equal to the noise margin $NM = (NM_H + NM_L)/2$:

$$NM = \frac{\sqrt{(A_D-1)^2-1}}{A_D} V_S + \frac{V_T}{2} \ln\left(\frac{A_D - 1 + \sqrt{(A_D-1)^2-1}}{A_D - 1 - \sqrt{(A_D-1)^2-1}}\right)$$

$$= \frac{V_S}{2}\left[\sqrt{1-4x} + x\ln\left(\frac{1-2x+\sqrt{1-4x}}{1-2x-\sqrt{1-4x}}\right)\right] ,$$

where $x = V_T/V_S$. In order for an output voltage error of less than 5% to be achieved, the input voltage difference has to exceed $3V_T$. In order to meet this constraint even at an elevated temperature of 150°C, a minimum voltage swing of $3V_T \approx 109\,\mathrm{meV}$ is required for differential operation, whereas a logic swing of $6V_T \approx 218$ mV is required for single-ended operation. In practical ECL circuits, additional errors have to be considered [6,12]:

- The differential stages have offset voltages caused by statistical variations of the saturation currents and load resistances.
- Voltage drops in the supply lines cause a shift of the output voltage. This problem can be compensated by adding additional series resistances if the current distribution on the chip is known. Another technique to reduce voltage drops in the supply lines is the use of full metallization on the front and back surfaces to supply power [13].
- The voltage swing is subject to errors caused by deviations of the current supplied by the current source (errors in the bias driver and errors in the individual current mirrors) from its nominal value.
- The output current delivered by an output that is HI causes a voltage drop in the load resistance and causes the output voltage to be approximately $nR_C I_{EF}/B_N$ below V_{CC} if n emitter followers are connected to the output.

In single-ended ECL circuits, signal levels must meet absolute specifications of their values since the difference between a signal voltage and a fixed reference voltage determines the state of the output. The voltage swing V_S has to be chosen large enough to ensure that the input voltage difference exceeds $3V_T$ even in the worst case: single-ended ECL circuits require voltage swings in excess of 500 mV for reliable operation to be obtained. Some of the errors noted above do not affect differential ECL circuits, for which a voltage swing as low as 200 mV is sufficient for reliable operation.

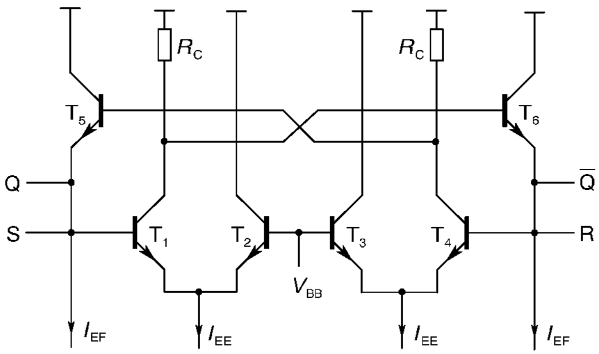


Fig. 8.10. RS flip-flop realized with single-ended ECL NOR gates

8.1.3 Flip-Flops

The RS flip-flop can be realized by the cross-coupling of two ECL NOR gates. Such a realization using two single-ended ECL gates is shown in Fig. 8.10. The two reference transistors can be omitted. Merging of the two gates then results in the circuit depicted in Fig. 8.11. In this example, multiemitter transistors are used for the emitter followers in order to decouple the intrinsic feedback loop from the load-dependent switching of the output nodes Q and $\overline{\mathrm{Q}}$. Note that the set and reset nodes (1) and (2) form a wired-OR circuit with the emitter follower of the preceding circuit. Therefore, only three current sources are required by the flip-flop; the current sources that deliver the bias currents I_{EF2} are those of the preceding emitter followers.

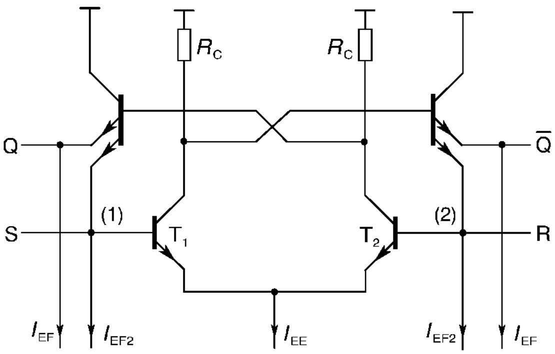


Fig. 8.11. Alternative realization of an RS flip-flop, omitting the reference transistors

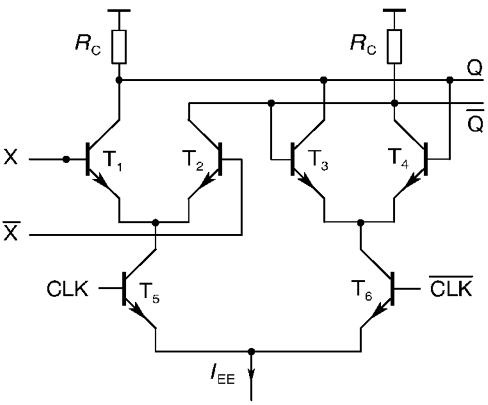


Fig. 8.12. D flip-flop realized with a two-level series gating technique

Figure 8.12 shows a (transparent) D flip-flop. This circuit changes the output variables Q and $\overline{\mathrm{Q}}$ to $\overline{\mathrm{X}}$ and X if CLK is HI. If CLK is LO, the circuit is nontransparent and changes of the input variables do not affect the output variables.

8.1.4 Frequency Dividers

Frequency dividers, or prescalers, are frequently employed in frequency synthesizers that operate at high frequencies to extend the frequency range of phase-locked loop (PLL) circuits. A static frequency divider is usually realized as a master–slave D flip-flop (Fig. 8.13). Examples of static frequency

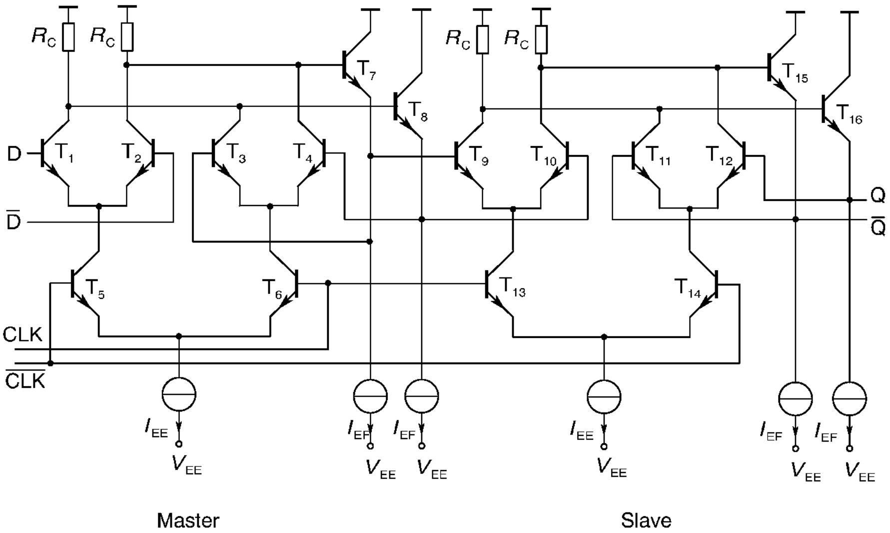


Fig. 8.13. Circuit diagram of a 2:1 static frequency divider (master–slave flip-flop)

dividers are a 53 GHz static 2:1 frequency divider constructed in Si/SiGe bipolar technology that operates with a current of 122 mA at a 6.3 V supply voltage [14], a 15 GHz static 8:1 frequency divider that consumes only 22 mA from a 3.6 V supply voltage [15], and a dual-modulus prescaler with a maximum operating frequency of 20 GHz with 27 mW power consumption at a supply voltage of 2.3 V [16]. A static frequency divider with a maximum operating frequency of 72.2 GHz has been described in [27].

An analog multiplier can be used in the realization of a dynamic frequency divider; an example has been presented in [17], featuring a maximum frequency of operation of 79 GHz, achieved with a self-aligned SiGe bipolar technology with an effective emitter width of 0.25 µm. Another example of a dynamic frequency divider with a maximum operating frequency of 92.4 GHz has been described in [27]. Figure 8.14 shows a circuit diagram of a dynamic frequency divider. The input signal is applied to one of the multiplier ports, and the multiplier output is connected to the second multiplier input via three cascaded emitter followers. These emitter followers provide the necessary level shifting and help to optimize the frequency response of the feedback

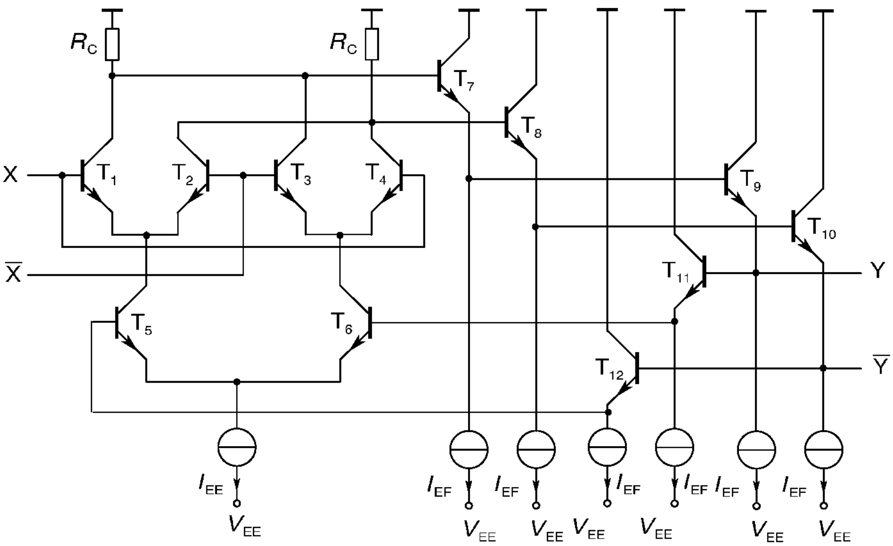


Fig. 8.14. Circuit diagram of a 2:1 dynamic frequency divider with input X, $\overline{X}$ and output Y, $\overline{Y}$ (after [17])

loop. Such dynamic frequency dividers have a limited bandwidth – a ratio of approximately three between the maximum and the minimum frequency of operation can be achieved [17]. This is, however, no problem in applications which require operation of the divider in only a limited frequency range.

8.2 High-Speed Optical Transmission Systems

Figure 8.15 shows a block diagram of a fiber-optic link of the kind used in synchronous optical networks (SONET). Data from different input channels

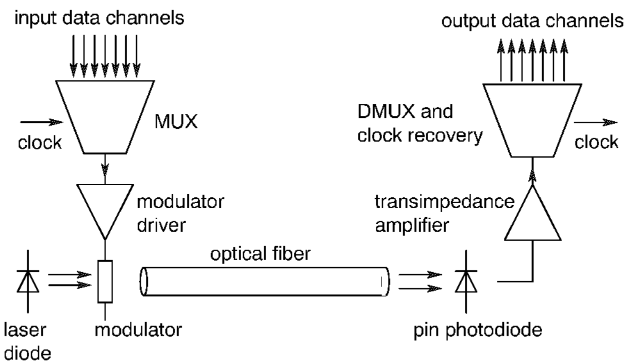


Fig. 8.15. Block diagram of fiber-optic link

are collected into a single high-speed data-stream by means of a time-division multiplexer (MUX). This signal is used to modulate the light emitted by a laser diode with the help of an electro-optic light modulator. The light signal traverses the optical fiber and generates a photocurrent in a pin diode at the other end. This current is amplified and transformed to a voltage signal by means of a wide-dynamic-range automatic-gain transimpedance amplifier, which is followed by a clock and data recovery circuit with a demultiplexer (DMUX) for parallelization of the data stream.

The transimpedance amplifier amplifies the small input current delivered by the pin photodiode and transforms it to an output voltage. In light-wave receiver systems, automatic gain control is required to adapt the system to a broad range of input signal levels. Figure 8.16 shows a schematic of a gain-controlled transimpedance amplifier cell that can be used in an automatic-gain-control amplifier. This circuit combines an analog multiplier (Gilbert cell) with a transimpedance stage, and provides a maximum voltage gain [18]

$$A_{\mathrm{V}} = \frac{R_{\mathrm{F}}}{2V_{\mathrm{T}}/I_{\mathrm{EE1}}+R_{\mathrm{E}}+r_{\mathrm{ee'}}+r_{\mathrm{bb'}}},$$

which can be reduced by changing the voltage V_{ctrl}. An automatic-gain amplifier with a wide bandwidth of 32.7 GHz and wide dynamic range has been described in [19].

A 40 Gb/s integrated clock and data recovery circuit realized in self-aligned bipolar technology, with a cutoff frequency of 50 GHz, has been presented in [20]; another example of a clock and data recovery IC for a 40 Gb/s fiber-optic receiver, fabricated with InP HBTs, has been presented in [21]. An example of a SiGe receiver IC for 10 Gb/s data rate has been presented in [22]. An example of a fully integrated 40 Gb/s clock and data recovery IC realized in SiGe technology for use in high-speed optical transmission systems has been presented in [23]. An IC chipset with a static frequency divider, a MUX/DMUX, a preamplifier and an automatic-gain-control amplifier, fabricated in SiGe bipolar technology, for optical-fiber-link communications, operating at a data rate of 40 Gb/s has been reported in [24]. Another chipset for a 40 Gb/s system, using SiGe HBTs, has been presented in [25]; a solution that employs InP HBTs is described in [26].

8.3 RF Microelectronics

Considerable interest in high-frequency bipolar transistors has arisen from the rapid development of personal communication systems (cellular phones, pagers, wireless LANs, etc.) over the last decade. Such communication systems generally operate in the frequency band between 800 MHz and 2.5 GHz,

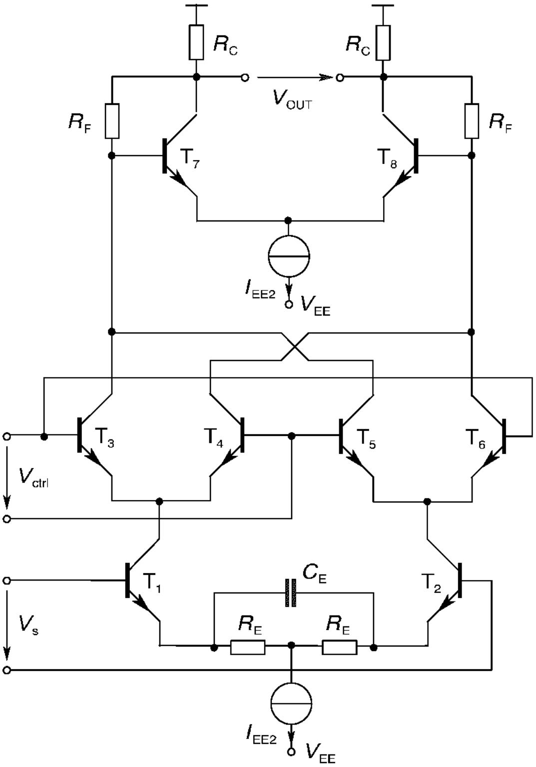


Fig. 8.16. Gain-controlled amplifier cell (after [18])

since above this value signal attenuation in the atmosphere and within buildings rises rapidly with frequency.[5]

RF integrated-circuit design is among the most demanding design tasks: the designer has to optimize the circuit with respect to requirements for low noise, high linearity, large bandwidth, small power consumption, small supply voltage, etc., while taking account of the trade-offs that exist between these issues.[6] Third-order intermodulation is a severe problem in mobile communication systems, as can be seen from the following example. Consider a narrow-band amplifier with a bandwidth of 1 MHz and a center frequency

[5]Wireless systems operating at higher frequencies are of interest for application in wireless local area networks. For example, there are unlicensed frequency bands from 5.15 to 5.35 GHz and from 5.725 to 5.825 GHz in North America, and from 5.47 to 5.725 GHz in Europe, which allow to realize portable multimedia applications with data rates between 20 and 150 Mb/s. An example of a receiver for this frequency range, realized with 0.5 µm silicon bipolar technology (f_T = 25 GHz), has been presented in [28]; and [29] presents a downconverter, realized with SiGe technology, that operates in this frequency range.

[6]A detailed discussion is beyond the scope of this book. The fundamentals of RF microelectronics are presented, in [30], for example. A compilation of selected reprints on integrated circuits for wireless communications is found in [31]. Important design considerations for very-high-speed Si bipolar ICs are addressed in [32].

of 900 MHz. Obviously, the signal frequency $f = 900$ MHz is well within the passband, while the frequencies $f_1 = 902$ MHz and $f_2 = 904$ MHz are outside it and should therefore not affect the output of the amplifier. The frequency $2f_1 - f_2$, which is a product of mixing, however, will lie within the passband of the amplifier. An example of an RF integrated circuit is an integrated silicon bipolar receiver for digital cordless telephones in the 900 MHz ISM band that operates with supply voltages between 2.7 V and 6.5 V, consumes 26 mA and has 118 dB dynamic range; this is described in [33].

Low-noise amplifiers (LNAs) are required for the amplification of the incoming RF signal prior to the frequency conversion process. This is necessary in order to reduce the effect of the noise produced by the mixer; the amplifier provides a substantially increased signal level at the input. Figure 8.17

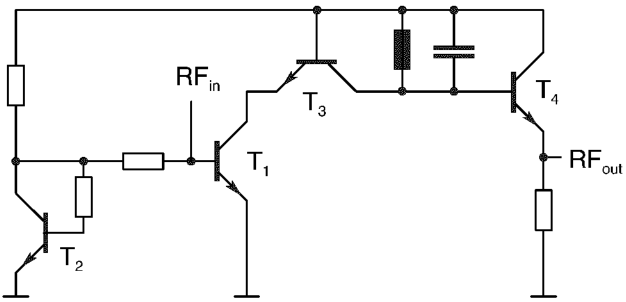


Fig. 8.17. Simplified circuit diagram of low-noise amplifier (after [34])

shows a simplified circuit diagram of a low-noise amplifier. The noise of this circuit is determined by the input transistor T_1, which has a bias point defined by a modified current mirror. The layout of T_1 is chosen to achieve the optimum base resistance and collector current density for the minimum noise figure. Together with T_3, the input transistor forms a cascode stage, in order to minimize the Miller effect. A parallel resonant circuit tuned to the frequency of operation is used to provide a high gain at a small supply voltage. The emitter follower T_4 provides a low-impedance output signal. The circuit works for a supply voltage as low as 2.4 V. Using SiGe technology and a 3.6 V supply voltage, a gain of 26.3 dB and a noise figure of 2.0 dB were achieved at $f = 10.5$ GHz and a power dissipation of 26.6 mW [34]; with silicon bipolar technology, a power gain of 26 dB and noise figure of 1.8 dB have been achieved at $f = 5.6$ GHz and 31.3 mW power dissipation [35]. A low-noise amplifier that operates with a current of 4.8 mA at a 2.7 V supply voltage with a noise figure of 1.3 dB at 1.8 GHz has been constructed using a 30 GHz SiGe bipolar process [36]. At high frequencies, the LNA noise figure increases with frequency (Fig. 8.18); for high-frequency low-noise amplifiers, the forward transit time should be as small as possible [37].

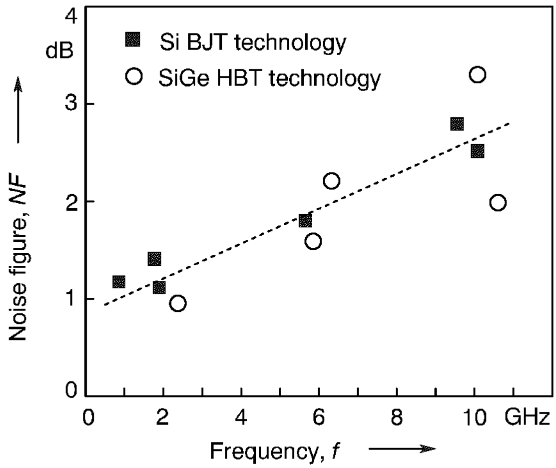


Fig. 8.18. Noise figures achieved in practice versus frequency for Si and SiGe technologies (after [35])

Voltage-controlled oscillators (VCOs) must exhibit low phase noise, small dc current consumption and low-voltage operation. VCOs are generally realized with integrated varactor diodes,[7] as in the circuit shown in Fig. 7.11. A recent example of a monolithic VCO with phase noise -106 dBc/Hz at a 100 kHz offset from the 800 kHz carrier which consumes 1.6 mA from a 2.7 V supply, has been presented in [40]. A 108 GHz VCO, realized with InP HBTs, with a tuning bandwidth of 2.73 GHz, and phase noise of -88 dBc/Hz and -109 dBc/Hz at offsets of 1 MHz and 10 MHz, respectively, has been presented in [41].

In a mixer, the RF signal, i.e. a modulated narrowband signal

$$v_{\mathrm{RF}}(t) = \hat{v}_1(t)\cos(\omega_{\mathrm{RF}}t) + \hat{v}_2(t)\sin(\omega_{\mathrm{RF}}t) ,$$

is multiplied by the local-oscillator signal, with angular frequency ω_{LO}, to obtain the IF (intermediate frequency) signal:

$$\begin{aligned} v_{\mathrm{IF}}(t) &= \frac{\hat{v}_1(t)}{2}\{\sin[(\omega_{\mathrm{RF}}+\omega_{\mathrm{LO}})t] + \sin[(\omega_{\mathrm{RF}}-\omega_{\mathrm{LO}})t]\} \\ &\quad + \frac{\hat{v}_2(t)}{2}\{\cos[(\omega_{\mathrm{RF}}+\omega_{\mathrm{LO}})t] + \cos[(\omega_{\mathrm{RF}}-\omega_{\mathrm{LO}})t]\} . \end{aligned} \tag{8.3}$$

This is composed of two mixing products, with angular frequencies $\omega_{\mathrm{RF}}+\omega_{\mathrm{LO}}$ and $\omega_{\mathrm{RF}}-\omega_{\mathrm{LO}}$. Of these, only the difference frequency is used in receiver systems, and the frequency component $\omega_{\mathrm{RF}}+\omega_{\mathrm{LO}}$ is rejected by filters. Mixers are derived from the analog multiplier considered in Sect. 6.7. Figure 8.19

[7] A VCO that uses a variable impedance converter to simulate the varactor function, operates from a single 2V supply voltage and consumes 15 mA, with a maximum oscillation frequency of 2 GHz and -86 dBc/Hz at 100 kHz offset in the 2 GHz range, has been described in [38]; a corresponding 2 V, 1.8 GHz BJT phase-locked loop has been described in [39].

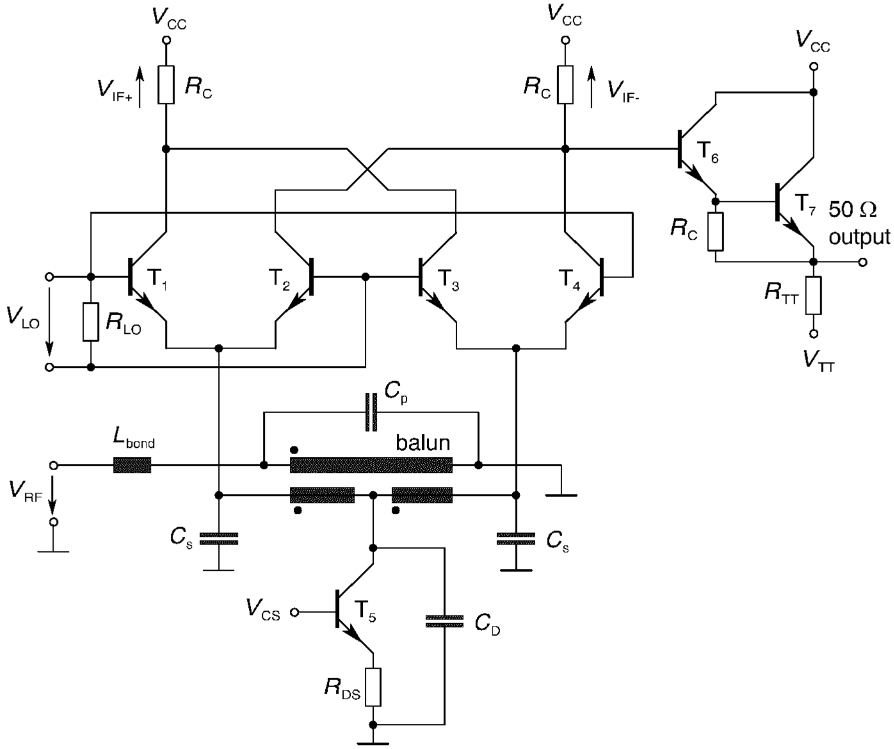


Fig. 8.19. Schematic of a doubly balanced mixer (after [42]; the use of a transformer to match the output to a 50 Ω load has also been described in [42], but is not included in this figure)

shows an example of a doubly balanced mixer that uses a balun at the input to split the RF input signal into anti-phase and in-phase components. These components are fed to the cross-coupled differential stages formed by transistors T_1 to T_4. The bias current is fed from the current source T_5 through the center tap in the secondary winding of the balun. The output Darlington pair matches IF output signal to a 50 Ω load.

Mobile phones require low-cost power amplifiers which operate at a small supply voltage that is fed directly from the battery and is therefore subject to variations. In digital cellphones, such amplifiers are required to deliver an output power of more then 2 W directly to the antenna. The power-added efficiency of power amplifier stages is an important quantity that substantially affects the battery standby time. Practical power amplifiers can achieve efficiencies of about 65%, i.e. almost two-thirds of the battery power is turned into useful RF power at the antenna and only one-third is dissipated in the power amplifier circuit. The impedance matching required is typically performed with integrated transformers. An example of an integrated 5 W power amplifier, realized with a 25 GHz f_T silicon bipolar technology, with a power-added efficiency of 59% has been presented in [43]; a monolithic 2.5 V, 1 W power amplifier with 55% power-added efficiency at 1.9 GHz has been presented in [44]; the circuit consists of an on-chip transformer, which acts as an input balun and input matching network, a driver stage, two transformers acting as an interstage matching network, and a power output stage. An example of an AlGaAs/GaAs HBT MMIC power amplifier for dual-band operation that operates with a 3.2 V supply voltage is described in [45].

8.4 BiCMOS

The acronym "BiCMOS" describes an integrated-circuit technology that combines CMOS technology with bipolar transistors on one chip. The technique can be employed to speed up digital circuits and to produce circuits for RF and mixed-signal applications. It should be noted, however, that the rather complex process flows that are required to integrate high-performance bipolar transistors and MOSFETs make this approach inappropriate for the design of products with a low production volume or a low margin.

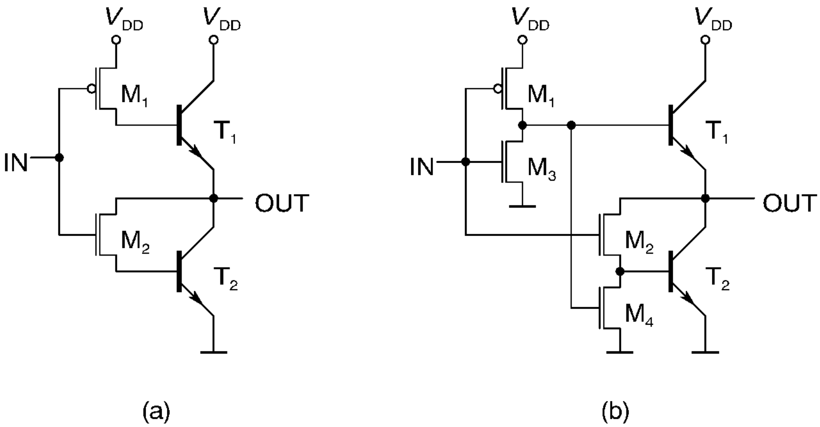


Fig. 8.20. Tow different realizations of a BiCMOS inverter

Figure 8.20 shows two different realizations of an inverter in BiCMOS technology. Both of these can be used for driving large capacitive loads. These circuits combine the high-impedance input of a MOSFET with a low-impedance bipolar output stage and are faster in charging or discharging a capacitive load than their pure CMOS counterparts are if equal area consumption is assumed.[8] The circuit shown in Fig. 8.20a requires fewer devices but suffers

[8]See, for example, [46–50] for analytical investigations of the transient behavior of BiCMOS buffers.

from considerable power consumption in the case of a slowly varying input voltage. The reason for this is that, when $V_{\mathrm{BEon}} + V_{\mathrm{THn}} < V_{\mathrm{IN}} < V_{\mathrm{DD}} + V_{\mathrm{THp}}$, both M_1 and M_2 deliver a base current, with the result that T_1 and T_2 form a low-impedance path from V_{DD} to ground. The additional MOSFETs M_3 and M_4 in the circuit shown in Fig. 8.20b allow one to circumvent this dilemma and are dimensioned such that the range of input voltages in which both T_1 and T_2 may carry a large current is minimum.

Besides the realization of fast output buffers, BiCMOS allows one to combine different circuit techniques on one chip. One such approach is the combination of CMOS and ECL circuit techniques. Realizing time-critical paths with an ECL circuit technique may allow a reduced cycle time, without the huge power consumption implied by a pure ECL solution. High-speed SRAM circuits with CMOS memory cells and ECL peripheral circuits have been developed for use as cache and control memories in mainframe computers. An example of a 1 Mb ECL–CMOS SRAM with a 550 ps clock access time and a power dissipation of 43 W constructed with a 0.2 μm BiCMOS technology is given in [51].

Probably the most important application of BiCMOS technologies is the construction of mixed analog–digital circuits, which combine RF and analog building blocks. These building blocks are realized using bipolar transistors together with CMOS digital-circuit techniques. In RF and mixed-signal applications, bipolar transistors are desirable owing to their larger transconductance and better device-to-device matching, which allows smaller offset voltages. In addition, bipolar amplifiers provide larger bandwidth, lower noise, higher gain for a given layout size and power consumption level, and easier matching to off-chip RF passive components.

A typical example is an RF PLL, which usually consists of a bipolar prescaler and a programmable CMOS divider, phase detector and charge pump. In low-cost mobile communication systems, these are preferrably integrated on one chip, using a BiCMOS technology. A 10 GHz SiGe BiCMOS frequency synthesizer with a maximum operational frequency of 10 GHz and a power consumption of 17 mW has been presented in [52]. A BiCMOS process for RF telecommunications with a Si epitaxial base and high-voltage transistors with $BV_{\mathrm{CEO}} = 10$ V has been presented in [53]. Integrated systems that combine analog or RF circuits with digital circuit sections impose additional design difficulties: owing to the small signal levels, the analog and RF circuits are sensitive to noise in the power supply and on the substrate generated by the digital CMOS section.

Earlier BiCMOS technologies tried to minimize the process overhead associated with the integration of bipolar transistors in a CMOS process. The modular approach of modern BiCMOS processes [54, 55] is instead based on a core CMOS process and various add-on-modules that can be added as required by the design project.

Table 8.1. SiGe HBT parameters for different BiCMOS technology generations [54]

Parameter	0.5 µm	0.25 µm	0.18 µm
$A_{\mathrm{je}}/\mu\mathrm{m}^2$	0.5×2.5	0.44×3	0.18×0.82
B_{N}	100	100	200
$V_{\mathrm{A}}/\mathrm{V}$	65	75	120
$BV_{\mathrm{CEO}}/\mathrm{V}$	3.35	3.35	2.5
$BV_{\mathrm{CBO}}/\mathrm{V}$	10.5	10.5	7.5
$f_{\mathrm{T}}/\mathrm{GHz}$	47	47	90
$f_{\mathrm{max}}/\mathrm{GHz}$	65	65	90
$NF_{\mathrm{min}}/\mathrm{dB}$	0.8	0.8	0.4

The performance of BiCMOS systems on chip has to be competitive with multiple-chip solutions in pure CMOS and SiGe bipolar technologies. Thus high-performance bipolar transistors have to be integrated. Table 8.1 shows some typical parameters achieved with different BiCMOS technologies. The small breakdown voltages of the high-performance transistors, which result from the selective collector implant, can be increased by masking this implant in the case of certain "high-breakdown" transistors. Additional components that are generally required on chip in mixed-signal systems are highly linear resistors, capacitors, varactors and high-Q inductors. The cost increase associated with the additional process modules should be well below 50%, and the BiCMOS process should require only a small additional time relative to the underlying CMOS process.

8.5 References

[1] J.D. Cressler. SiGe HBT technology: a new contender for Si-based RF and microwave circuit applications. *IEEE Trans. Microwave Theory Tech.*, 46(5):572–589, 1998.

[2] D.K. Lynn, C.S. Meyer, D.J. Hamilton. *Analysis and Design of Integrated Circuits*. McGraw-Hill, New York, 1967.

[3] H.H. Muller, W.K. Owens, P.W.J. Verhofstadt. Fully compensated emitter-coupled logic: eliminating the drawbacks of conventional ECL. *IEEE J. Solid-State Circuits*, 8(5):362–367, 1973.

[4] H.Y. Hsieh, K. Chin, C.-T. Chuang. Power partition and emitter size optimization for bipolar ECL circuits. *IEEE J. Solid-State Circuits*, 28(5):548–552, 1993.

[5] J.M. McGregor, D.J. Roulston, J.S. Hamel, M. Vaidyanathan, S.C. Jain, P. Balk. A simple expression for ECL propagation delay including non-quasi-static effects. *Solid-State Electron.*, 36(3):391–396, 1993.

[6] G.R. Wilson. Advances in bipolar VLSI. *Proc. IEEE*, 78(11):1707–1719, 1990.

[7] S. Colaco, R. Davies, D. Healey, O. Choy. Multilevel differential logic – the bipolar alternative. *J. Semicustom ICs*, 3(4):21–27, 1986.

[8] B. Razavi, Y. Ota, R.G. Swartz. Design techniques for low-voltage high-speed digital bipolar circuits. *IEEE J. Solid-State Circuits*, 29(3):332–339, 1994.

[9] H.M. Rein, R. Ranfft. Improved feedback ECL gate with low delay–power product for the subnanosecond region. *IEEE Trans. Electron Devices*, 12(1):80–82, 1977.

[10] J.B. Hughes. Comments on "Improved Feedback ECL with low delay power product for the subnanosecond region". *IEEE J. Solid-State Circuits*, 13(2):276–278, 1978.

[11] V. Ramakrishnan, J.N. Albers, R.N. Nottenburg. Modified feedback ECL gate for Gb/s applications. *IEEE J. Solid-State Circuits*, 34(2):205–211, 1999.

[12] R.L. Treadway. DC analysis of current mode logic. *IEEE Circuits Devices Mag.*, (March):21 35, 1989.

[13] M. Kokado, M. Hyoshida, N. Miyoshi, K. Suzuki, M. Takaoka, N. Tsuzuki, H. Harada. A 54000-gate ECL array with substrate power supply. *IEEE J. Solid-State Circuits*, 24(5):1271–1274, 1989.

[14] M. Wurzer, T.F. Meister, H. Knapp, K. Aufinger, R. Schreiter, S. Boguth, L. Treitinger. 53 GHz/static frequency divider in a Si/SiGe bipolar technology. *IEEE ISSCC Tech. Dig.*, pp. 206–207, 2000.

[15] H. Knapp, W. Wilhelm, M. Wurzer. A low-power 15-GHz frequency divider in a 0.8µm silicon bipolar technology. *IEEE Trans. Microwave Theory Tech.*, 48(2):205–208, 2000.

[16] H. Knapp, T.F. Meister, M. Wurzer, K. Aufinger, S. Boguth, L. Treitinger. A low-power 20-GHz SiGe dual-modulus prescaler. *IEEE MTT-S Dig.*, pp. 731–734, 2000.

[17] H. Knapp, T.F. Meister, M. Wurzer, D. Zöschg, K. Aufinger, L. Treitinger. A 79-GHz dynamic frequency divider in SiGe bipolar technology. *IEEE ISSCC Tech. Dig.*, pp. 208–209, 2000.

[18] M. Möller, H.-M. Rein, H. Wernz. 13 Gb/s Si-bipolar AGC amplifier IC with high gain and wide dynamic range for optical-fiber receivers. *IEEE J. Solid-State Circuits*, 29(7):815 822, 1994.

[19] K. Ohhata, T. Masuda, E. Ohue, K. Washio. Design of a 32.7-GHz bandwidth AGC amplifier IC with wide dynamic range implemented in SiGe HBT. *IEEE J. Solid-State Circuits*, 34(9):1290–1297, 1999.

[20] M. Wurzer, J. Böck, H. Knapp, W. Zirwas, F. Schumann, A. Felder. A 40-Gb/s integrated clock and data recovery circuit in a 50-GHz f_T silicon bipolar technology. *IEEE J. Solid-State Circuits*, 34(9):1320–1324, 1999.

[21] G. Georgiou, Y. Bayens, Y.-K. Chen, A.H. Gnauck, C. Gröpper, P. Pachke, R. Pullela, M. Reinhold, C. Dorschky, J.-P. Mattia, T.W. von Mohrenfels, C. Schulien. Clock and data recovery IC for 40-Gb/s fiber-optic receiver. *IEEE J. Solid-State Circuits*, 37(9):1120–1125, 2002.

[22] Y.M. Greshishchev, P. Schvan, J.L. Showell, M.-L. Xu, J.J. Ojha, J.E. Rogers. A fully integrated SiGe receiver IC for 10-Gb/s data rate. *IEEE J. Solid-State Circuits*, 35(12):1949–1957, 2000.

[23] M. Reinhold, C. Dorschky, E. Rose, R. Pullela, P. Mayer, F. Kunz, Y. Baeyens, T. Link, J.-P. Mattia. A fully integrated 40-Gb/s clock and data recovery IC with 1:4 DEMUX in SiGe technology. *IEEE J. Solid-State Circuits*, 36(12):1937–1945, 2001.

[24] K. Washio. SiGe HBTs and ICs for optical-fiber communication systems. *Solid-State Electron.*, 43:1619–1625, 1999.

[25] G. Freeman, M. Meghelli, Y. Kwark, S. Zier, A. Rylyakov, M.A. Sorna, T. Tanji, O.M. Schreiber, K. Walter, J.-S. Rieh, B. Jagannathan, A. Joseph, S. Subbanna. 40-Gb/s circuits built from a 120-GHz f_T SiGe technology. *IEEE J. Solid-State Circuits*, 37(9):1106 1114, 2002.

[26] Y. Baeyens, G. Georgiou, J.S. Weiner, A. Leven, V. Houtsma, P. Paschke, Q. Lee, R.F. Kopf, Y. Yang, L. Chua, C. Chen, C.T. Liu, Y.-K. Chen. InP D-HBT ICs for 40 Gb/s and higher bitrate lightwave transceivers. *IEEE J. Solid-State Circuits*, 37(9):1152–1159, 2002.

[27] K. Washio, E. Ohue, K. Oda, R. Hayami, M. Tanabe, H. Shimamoto. Optimization and characteristics related to the emitter–base junction in self-aligned SEG SiGe HBTs and their application in 72-GHz-static/92-GHz-dynamic frequency dividers. *IEEE Trans. Electron Devices*, 49(10):1755–1760, 2002.

[28] J.P. Maligeorgos, J.R. Long. A low-voltage 5.1–5.8-GHz image-reject receiver with wide dynamic range. *IEEE J. Solid-State Circuits*, 35(12):1917–1926, 2000.

[29] J.R. Long. A low-voltage 5.1-5.8-GHz image-reject downconverter RF-IC. *IEEE J. Solid-State Circuits*, 35(9):1320–1328, 2000.

[30] B. Razavi. *RF Microelectronics*. Prentice Hall, Upper Saddle River, 1997.

[31] A.A. Abidi, P.R. Gray, R.G. Meyer. *Integrated Circuits for Wireless Communications*. IEEE Press, New York, 1999.

[32] H.-M. Rein, M. Möller. Design considerations for very-high-speed Si-bipolar IC's operating up to 50 Gb/s. *IEEE J. Solid-State Circuits*, 31(8):1076–1090, 1996.

[33] J. Durec. An integrated silicon bipolar receiver subsystem for 900-MHz ISM band applications. *IEEE J. Solid-State Circuits*, 33(9):1352–1372, 1998.

[34] D. Zöschg, W. Wilhelm, T.F. Meister, H. Knapp, H.-D. Wohlmuth, K. Aufinger, M. Wurzer, J. Böck, H. Schäfer, A.L. Scholtz. 2dB noise figure, 10.5 GHz LNA using SiGe bipolar technology. *Electron. Lett.*, 35:2195–2196, 1999.

[35] D. Zöschg, W. Wilhelm, J. Böck, H. Knapp, M. Wurzer, K. Aufinger, H.-D. Wohlmuth, A.L. Scholtz. Monolithic LNAs up to 10 GHz in a production-near 65 GHz $f_{\max}$ silicon bipolar technology. *Proc. IEEE Radio Frequency IC Symp.*, pp. 135–138, 2000.

[36] O. Shana'a, I. Linscott, L. Tyler. Frequency-scalable SiGe bipolar RF front-end design. *IEEE J. Solid-State Circuits*, 36(6):888–895, 2001.

[37] D. Zöschg, W. Wilhelm, T.F. Meister, H. Knapp, M. Wurzer, K. Aufinger, J. Böck, H.-D. Wohlmuth, A.L. Scholtz. Low noise amplifiers in SiGe bipolar technology. *Microwave Eng. Europe*, June, pp. 47–49, 2000.

[38] W.-Z. Chen, J.-T. Wu. A 2-V 2-GHz BJT variable frequency oscillator. *IEEE J. Solid-State Circuits*, 33(9):1406–1410, 1998.

[39] W.-Z. Chen, J.-T. Wu. A 2-V 1.8-GHz BJT phase-locked loop. *IEEE J. Solid-State Circuits*, 34(6):784–789, 1999.

[40] M.A. Margarit, J. L. Tham, R.G. Meyer, M.J. Deen. A low-noise, low power VCO with automatic amplitude control for wireless applications. *IEEE J. Solid-State Circuits*, 34(6):761–771, 1999.

[41] K.W. Kobayashi, J.C. Cowles, L.T. Tran, A. Gutierrez-Aitken, M. Nishimoto, J.H. Elliott, T.R. Block, A.K. Oki, D.C. Streit. A 44-GHz-high IP3 InP HBT MMIC amplifier for low DC power millimeter-wave receiver applications. *IEEE J. Solid-State Circuits*, 34(9):1188–1195, 1999.

[42] J.R. Long, M.A. Copeland, P. Schvan, R.A. Hadaway. A low-voltage silicon bipolar RF front-end for PCN receiver applications. *Proc. IEEE International Solid-State Circuits Conf.*, pp. 140–141, 1995.

[43] W. Simbürger, H.-D. Wohlmuth, P. Weger, A. Heinz. A monolithic transformer coupled 5-W silicon power amplifier with 59% PAE at 0.9 GHz. *IEEE J. Solid-State Circuits*, 34(12):1881–1892, 1999.

[44] W. Simbürger, A. Heinz, H.-D. Wohlmuth, J. Böck, K. Aufinger, M. Rest. A monolithic 2.5 V, 1 W silicon bipolar power amplifier with 55% PAE at 1.9 GHz. *IEEE MTT-S Dig.*, 2000:853–856, 2000.

[45] K. Yamamoto, S. Suzuki, K. Mori, T. Asada, T. Okuda, A. Inoue, T. Miura, K. Chomei, R. Hattori, M. Yamanouchi, T. Shimura. A 3.2-V operation single-chip dual-band AlGaAs/GaAs HBT MMIC power amplifier with active feedback circuit technique. *IEEE J. Solid-State Circuits*, 35(8):1109–1120, 2000.

[46] H.J. de los Santos, B. Hoefflinger. Optimization and scaling of CMOS bipolar drivers for VLSI interconnects. *IEEE Trans. Electron Devices*, 33(11):1722–1730, 1986.

[47] E.W. Greeneich, K.L. McLaughlin. Analysis and characterization of BiCMOS for high-speed digital logic. *IEEE J. Solid-State Circuits*, 23(2):558–565, 1988.

[48] G.P. Rosseel, R.W. Dutton. Influence of device parameters on the switching speed of BiCMOS buffers. *IEEE J. Solid-State Circuits*, 24(1):90 99, 1989.

[49] S. Zhang, T.S. Kalkur, S. Lee, D. Chen. Analysis of the switching speed of BiCMOS buffer under high current. *IEEE J. Solid-State Circuits*, 29(7):787–796, 1994.

[50] S. Zhang, T.S. Kalkur. Analysis of BiCMOS buffer for input voltages with finite rise time. *IEEE J. Solid-State Circuits*, 29(7):797–806, 1994.

[51] H. Nambu, K. Kanetani, K. Yamasaki, K. Higeta, M. Usami, M. Nishiyama, K. Ohhata, F. Arakawa, T. Kusunoki, K. Yamaguchi, A. Hotta, N. Homma. A 550-ps access 900-MHz 1-Mb ECL-CMOS SRAM. *IEEE J. Solid-State Circuits*, 35(8):1159–1168, 2000.

[52] B.-U. H. Klepser, M. Scholz, E. Götz. A 10-GHz SiGe BiCMOS phase-locked-loop frequency synthesizer. *IEEE J. Solid-State Circuits*, 37(3):328–335, 2002.

[53] H. Nii, C. Yoshino, S. Yoshitomi, K. Inoh, H. Furuya, H. Nakajima, H. Sugaya, H. Naruse, Y. Katsumata, H. Iwai. An 0.3 µm Si epitaxial base BiCMOS technology with 37-GHz $f_{\max}$ and 10-V BV_{CEO} for RF telecommunications. *IEEE Trans. Electron Devices*, 46(4):712–721, 1999.

[54] D.L. Harame, D.C. Ahlgren, D.D. Coolbaugh, J.S. Dunn, G.G. Freeman, J.D. Gillis, R.A. Groves, G.N. Hendersen, R.A. Johnson, A.J. Joseph, S. Subbanna, A.M. Victor, K.M. Watson, C.S. Webster, P.J. Zampardi. Current status and future trends of SiGe BiCMOS technology. *IEEE Trans. Electron Devices*, 48(11):2575–2594, 2001.

[55] D.A. Rich, M.S. Carroll, M.R. Frei, T.G. Ivanov, M. Mastrapasqua, S. Moinian, A.S. Chen, C.A. King, E. Harris, J. de Blauwe, H.-H. Vuong, V. Archer, K. Ng. BiCMOS – technology for mixed-digital, analog and RF applications. *IEEE Microwave Mag.*, 3(2):44–55, 2002.

APPENDIX

- Linear and Nonlinear Response
- Linear Twoports, S-Parameters
- PN Junctions: Details
- Bipolar Transistors: Details
- Noise: Details
- Overtemperature Developed During Electrostatic Discharges

A Linear and Nonlinear Response

Electronic devices and circuits are examples of systems, i.e. they have an[1] output variable $y(t)$ determined by an input variable $x(t)$ (Fig. A.1). In a

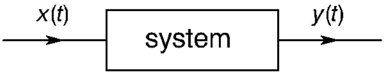


Fig. A.1. System with input variable $x(t)$ and output variable $y(t)$

system without memory, the output variable $y(t)$ is a function of the value of the input variable at the same instant, i.e. $y(t) = y[x(t)]$, whereas in a system with memory the output variable $y(t)$ is determined also by past values of the input variable.

A.1 Linear Response

A system is called linear if a superposition[2] $x(t) = \alpha_1 x_1(t) + \alpha_2 x_2(t)$ of arbitrary input signals $x_1(t)$ and $x_2(t)$ that have individual output signals $y_1(t)$ and $y_2(t)$ causes the output signal $y(t) = \alpha_1 y_1(t) + \alpha_2 y_2(t)$. If the input variable $x(t)$ of a linear system is harmonic with angular frequency ω, which is written in complex notation as

$$x(t) = \mathrm{Re}\left(\underline{\hat{x}}\,\mathrm{e}^{\mathrm{j}\omega t}\right) = \mathrm{Re}(\underline{x}) \,,$$

the output variable $y(t)$ shows the same frequency dependence

$$y(t) = \mathrm{Re}\left(\underline{\hat{y}}\,\mathrm{e}^{\mathrm{j}\omega t}\right) = \mathrm{Re}(\underline{y}) \,.$$

The ratio $\underline{A}(\mathrm{j}\omega) = \underline{y}/\underline{x}$ of the phasors $\underline{y}$ and $\underline{x}$ is the transfer factor of the system. If, in particular, $x(t)$ is a superposition of sinusoidal signals with angular frequencies $\omega_1, \ldots, \omega_\mathrm{n}$, i.e.,

$$x(t) = \sum_{k=1}^{N} \mathrm{Re}\left(\underline{\hat{x}}_k\,\mathrm{e}^{\mathrm{j}\omega_k t}\right) = \frac{1}{2} \sum_{k=-N}^{N} \underline{\hat{x}}_k\,\mathrm{e}^{\mathrm{j}\omega_k t} \,, \tag{A.1}$$

the output signal is

$$y(t) = \sum_{k=1}^{N} \mathrm{Re}\left[\underline{A}(\mathrm{j}\omega_k)\,\underline{\hat{x}}_k\,\mathrm{e}^{\mathrm{j}\omega_k t}\right] = \frac{1}{2} \sum_{k=-N}^{N} \underline{A}(\mathrm{j}\omega_k)\,\underline{\hat{x}}_k\,\mathrm{e}^{\mathrm{j}\omega_k t} \,, \tag{A.2}$$

[1] For simplicity, we consider only one input and one output variable; the definition is easily generalized to more input and output variables.

[2] The variables α_1, α_2 denote arbitrary real numbers.

where $\underline{\hat{x}}_k = \hat{x}_k \, \mathrm{e}^{\mathrm{j}\varphi_k} = \underline{\hat{x}}^*_{-\mathrm{k}}$, $\omega_{-\mathrm{k}} = -\omega_k$ and $\underline{A}(-\mathrm{j}\omega_k) = \underline{A}^*(\mathrm{j}\omega_k)$. The Fourier coefficients $\underline{\hat{y}}_k$ of the output signal are therefore derived from the corresponding Fourier coefficients $\underline{\hat{x}}_k$ of the input signal by multiplication by $\underline{A}(\mathrm{j}\omega_k)$. An analogous relation exists between the Laplace transform $\underline{y}(s)$ of the output variable and the Laplace transform $\underline{x}(s)$ of the input variable,

$$\underline{y}(s) \;=\; \underline{A}(s)\,\underline{x}(s) \; . \tag{A.3}$$

This multiplicative connection in the (complex) frequency domain corresponds to a convolution integral in the time domain,

$$y(t) \;=\; y(0) + \int_0^t A(t-t')\,x(t')\,\mathrm{d}t' \; . \tag{A.4}$$

The response of a linear system to an arbitrary input signal $x(t)$ is therefore completely described in terms of either the frequency-dependent transfer factor $\underline{A}(s)$ or the response function $A(t)$, which may be considered as the system's response to a pulse of negligible duration applied to the input. This is a direct consequence of the superposition principle characteristic of linear systems.

A.1.1 Step Response, Elmore Delay

Consider a time-dependent output signal of the form $y(t) = y_0 + y_\mathrm{s} f(t)$, where y_s is the output swing and $f(t)$ may take only values between zero and one. All information concerning the temporal response is contained in the normalized output variable $f(t)$. A possible form of $f(t)$, obtained as the response to a step function $x(t) = x_\mathrm{s}\Theta(t)$ applied to the input, is sketched in Fig. A.2. Two time constants are generally used to characterize the temporal response to a step function applied to the input. The definition of the delay time due to Elmore [1],

$$t_\mathrm{d} \;=\; \int_0^\infty t\dot{f}(t)\,\mathrm{d}t \; , \tag{A.5}$$

is derived from the transfer factor of the system according to

$$t_\mathrm{d} \;=\; -\,\frac{1}{\underline{A}(0)}\,\frac{\mathrm{d}\underline{A}}{\mathrm{d}s}\bigg|_{s=0} \;=\; -\,\frac{\mathrm{d}}{\mathrm{d}s}\,\ln\left[\underline{A}(s)\right]\bigg|_{s=0} \; , \tag{A.6}$$

while the rise time, according to Elmore's definition,

$$t_\mathrm{r} \;=\; \sqrt{2\pi\Big(\int_0^\infty t^2\dot{f}(t)\,\mathrm{d}t - t_\mathrm{d}^2\Big)} \; , \tag{A.7}$$

is related to the transfer factor by

$$t_\mathrm{r} \;=\; \sqrt{2\pi\left(\frac{1}{\underline{A}(0)}\,\frac{\mathrm{d}^2\underline{A}}{\mathrm{d}s^2}\bigg|_{s=0} - t_\mathrm{d}^2\right)} \; . \tag{A.8}$$

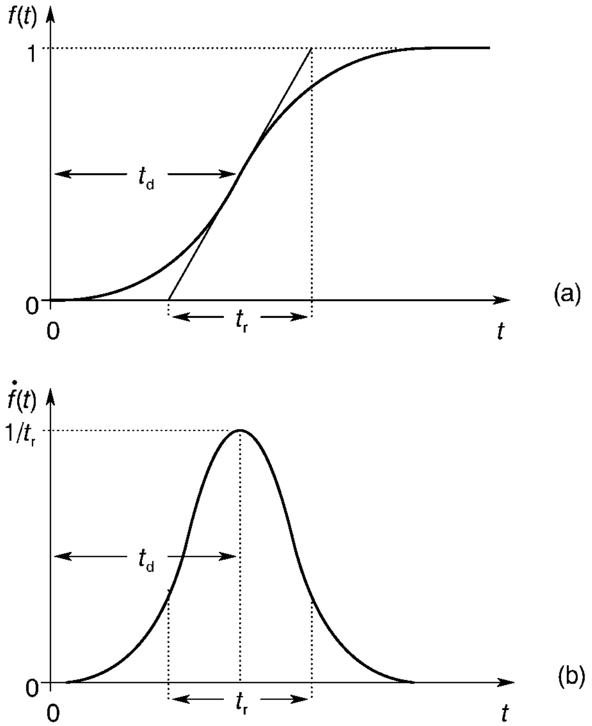


Fig. A.2. Temporal variation of (**a**) $f(t)$ and (**b**) $\dot{f}(t)$, and definition of delay and rise times

For lumped-element networks, the transfer factor is a rational function

$$\underline{A}(s) \;=\; \underline{A}(0)\,\frac{1 + a_1 s + a_2 s^2 + \cdots + a_n s^n}{1 + b_1 s + b_2 s^2 + \cdots + b_m s^m}\,,$$

leading to

$$t_\mathrm{d} \;=\; b_1 - a_1$$

and

$$t_\mathrm{r} \;=\; \sqrt{2\pi\Big[\,b_1^2 - a_1^2 + 2(a_2 - b_2)\Big]}\;.$$

Developing the transfer factor into a power series in s yields

$$\underline{A}(s) \;=\; \underline{A}(0)\,\left[1 - s t_\mathrm{d} + \frac{1}{2} s^2 \left(t_\mathrm{d}^2 + \frac{t_\mathrm{r}^2}{2\pi}\right) - \ldots\right]\;. \qquad \text{(A.9)}$$

A.2 Nonlinear Systems Without Memory

The following section summarizes the most important facts about weakly nonlinear systems [2, 3] without memory. An example of a memoryless nonlinear system is a nonlinear resistor with a terminal current that is a only function of the terminal voltages at time t, i.e. $i(t) = i[v(t)]$, where $i[v] = I(V)$ is the current–voltage characteristic of the resistor (see Sect. 1.2).

If a nonlinear system has no memory, its output may be represented as a Taylor series in the input signal, which is truncated after the first few terms. In the case of nonlinear systems with memory, which are considered in Sect. A.3, a generalization of the Taylor series, the Volterra series, is used to characterize the response. If a nonlinear system with an Nth-order polynomial response

$$y(t) = \sum_{p=1}^{N} \alpha_p \, x^p(t) \tag{A.10}$$

is excited by a sinusoidal signal of frequency f (one-tone excitation), the spectrum of the output signal is a superposition of harmonics, i.e. sinusoidal signals with frequencies mf, where $1 \leq m \leq N$ – a phenomenon that is known as harmonic distortion.

If the input signal is a superposition of two sinusoidal signals with frequencies f_1 and f_2 (two-tone excitation) the output signal is a superposition of mixing products, i.e. sinusoidal signals with frequencies $mf_1 + nf_2$, where m and n are integers (i.e. $m, n \in \{-N, \ldots, -2, -1, 0, 1, 2, \ldots, N\}$), restricted by $1 \leq |m| + |n| \leq N$. The harmonics of a one-tone excitation correspond to mixing products with either m or n equal to zero. The sum $|m| + |n|$ is called the order of the mixing product. If the polynomial representation (A.10) of $y(t)$ in terms of $x(t)$ ends with a term of order N, mixing products of order up to N will appear in the output signal.

A.2.1 Harmonic Distortion, Gain Compression

The output $y(t)$ of a nonlinear system, which is fed by a sinusoidal signal $x(t) = \hat{x} \cos(\omega t)$ will always be a superposition of harmonics

$$y(t) = y_0 + \sum_{k=1}^{\infty} \hat{y}_k \cos(k\omega t + \varphi_k) \, .$$

The ratio $HD_m = \hat{y}_m/\hat{y}_1$ is called the mth-order harmonic distortion; the quantity

$$THD = \sqrt{\sum_{k=2}^{\infty} HD_k^2}$$

is termed the total harmonic distortion or the distortion factor. Consider a circuit with a third-order polynomial response, fed by an input voltage $v_{\mathrm{i}}(t) = \hat{v}_{\mathrm{i}} \cos(\omega t)$. Using the identities

$$\cos^2(\omega t) = \frac{1+\cos(2\omega t)}{2} \quad \text{and} \quad \cos^3(\omega t) = \frac{3\cos(\omega t)+\cos(3\omega t)}{4},$$

one obtains the following for the output signal:

$$\begin{aligned} v_{\mathrm{o}}(t) = \sum_{p=1}^{3} \alpha_p v_{\mathrm{i}}^p(t) &= \frac{\alpha_2}{2}\hat{v}_{\mathrm{i}}^2 + \alpha_1(1+3\kappa_3)\,\hat{v}_{\mathrm{i}}\cos(\omega t) \\ &\quad + \alpha_1\kappa_2\cos(2\omega t) + \alpha_1\kappa_3\cos(3\omega t) \,. \end{aligned} \tag{A.11}$$

The coefficients

$$\kappa_2 = \frac{\alpha_2}{2\alpha_1}\hat{v}_{\mathrm{i}} \approx HD_2 = \frac{\text{amplitude of 2nd harmonic}}{\text{fundamental amplitude}}, \tag{A.12}$$

$$\kappa_3 = \frac{\alpha_3}{4\alpha_1}\hat{v}_{\mathrm{i}}^2 \approx HD_3 = \frac{\text{amplitude of 3rd harmonic}}{\text{fundamental amplitude}} \tag{A.13}$$

approximate the second- and third-order harmonic distortions. Figure A.3

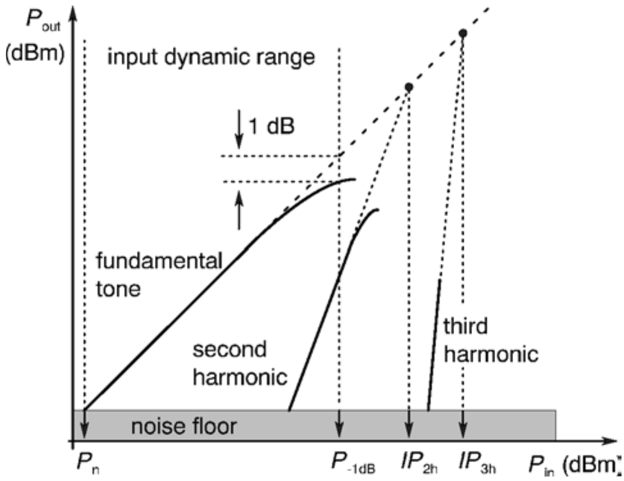


Fig. A.3. Output power associated with the first three harmonics as a function of the input power; definition of (1 dB) compression point, input dynamic range and second- and third-order harmonic intercept points

shows the output power[3] associated with the three different harmonics as a function of the input power level. Figure A.3 and (A.11) demonstrate three effects of the nonlinearity:

[3]It is common practice to express signal levels as the power dissipated in a 50 Ω resistor in dBm. A signal amplitude $\hat{v}_{\mathrm{i}} = 100$ mV then corresponds to a power dissipation $P_{\mathrm{i}} = \hat{v}_{\mathrm{i}}^2/2R = 0.1$ mW or -10 dBm.

1. An additional dc component appears, which is caused by the second-order nonlinearity.
2. At small values of the input power, the output spectrum shows only the response at the fundamental tone above the noise floor. With increasing input levels, second- and third-order harmonics contribute to the output signal. As $\kappa_2 \sim \hat{v}_\mathrm{i}$ and $\kappa_3 \sim \hat{v}_\mathrm{i}^2$, the output amplitudes of the second- and third-order harmonics rise faster than the output amplitude at the fundamental tone. Linear extrapolation of the corresponding curves allows to determine the second- and third-order harmonic intercept points $IP_{2\mathrm{h}}$ and $IP_{3\mathrm{h}}$, defined as the input power levels where the second- and third-order harmonic distortion coefficients HD_2 and HD_3, respectively, equal one.
3. The gain at the fundamental frequency is affected by the third-order response. The output amplitude of the fundamental tone is given by $\alpha_1\hat{v}_\mathrm{i}(1+3\kappa_3)$, resulting in a small-signal gain $|\alpha_1(1+3\kappa_3)|$. In amplifiers, κ_3 is generally negative, i.e. the signal gain decreases as a function of the input signal amplitude $\hat{v}_\mathrm{i}$, a phenomenon commonly referred to as gain compression. This effect is frequently quantified in terms of the (1 dB) compression point, which specifies the input signal level that causes a drop of the gain by 1 dB with respect to its small-signal value $|\alpha_1|$. The amplitude $\hat{v}_\mathrm{i,cp}$ of the input signal at the compression point can be derived from the condition

$$-1\,\mathrm{dB} \;=\; 20\,\mathrm{dB}\,\log[\,1+3\kappa_3(\hat{v}_\mathrm{i,cp})] \;\approx\; 20\,\mathrm{dB}\times 3\kappa_3$$

as $\hat{v}_\mathrm{i,cp} \approx \sqrt{|\alpha_1/15\,\alpha_3|}$. This corresponds to a power level $P_{-1\,\mathrm{dB}} = \hat{v}_\mathrm{i,cp}^2/R_\mathrm{i}$ delivered to the input of the amplifier. The 1 dB compression point is generally used together with the input-referred noise power P_n for the definition of the input dynamic range, calculated as the ratio $P_{-1\,\mathrm{dB}}/P_\mathrm{n}$ and expressed in dB.

A.2.2 Intermodulation Distortion

In the case of a two-tone excitation, the input signal $x(t)$ is the sum of two sinusoidal signals with angular frequencies ω and ω':

$$x(t) \;=\; \hat{x}\cos(\omega t) + \hat{x}'\cos(\omega' t)\;. \tag{A.14}$$

The terms $\sim x^2(t)$ and $\sim x^3(t)$ cause terms of the form $\cos(\omega t)\cos(\omega' t)$ and $\cos^2(\omega t)\cos(\omega' t)$. With the help of the identities

$$\cos(\omega t)\cos(\omega' t) \;=\; \frac{\cos[(\omega-\omega')t\,] + \cos[(\omega+\omega')t\,]}{2}$$

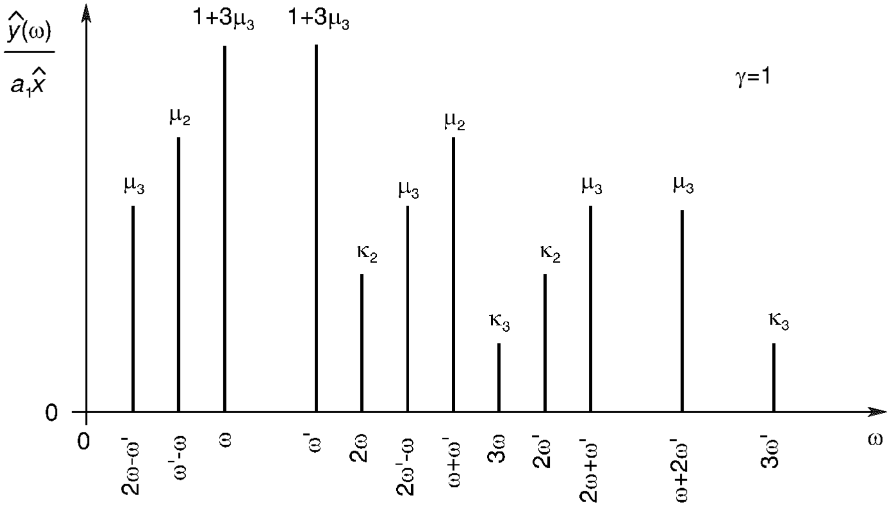


Fig. A.4. Normalized output spectrum caused by two sinusoidal inputs of equal amplitude $\hat{x}$ and angular frequencies ω, ω' including third order intermodulation product terms

and

$$\cos^2(\omega t)\,\cos(\omega' t) \;=\; \frac{\cos[(2\omega+\omega')t] + \cos[(2\omega-\omega')t] + 2\cos(\omega' t)}{4} \,,$$

these terms can be seen to produce contributions with angular frequencies $(\omega+\omega')$, $(\omega-\omega')$, $(2\omega+\omega')$, $(2\omega-\omega')$, $(2\omega'+\omega)$ and $(2\omega'-\omega)$ in the spectrum of the output signal. This effect is called intermodulation distortion and is predominantly determined by the coefficients

$$\mu_2 \;=\; \alpha_2\hat{x}/\alpha_1 \quad \text{and} \quad \mu_3 \;=\; 3\alpha_3\hat{x}^2/4\alpha_1 \,, \tag{A.15}$$

which are related to the coefficients used for the specification of harmonic distortion by $\mu_2 = 2\kappa_2$ and $\mu_3 = 3\kappa_3$. If $\gamma = \hat{x}'/\hat{x}$ is the ratio of the two input signal amplitudes, the output signal is, to third-order,

$$\begin{aligned} y(t) \;=\; & a_1\hat{x}\cdot\Big\{\kappa_2(1+\gamma^2) + \Big[1 + (1+2\gamma^2)m_3\Big]\cos(\omega t) \\ & + \Big[1 + (\gamma^2+2)m_3\Big]\,\cos(\omega' t) \\ & + \kappa_2\cos(2\omega t) + \gamma^2\kappa_2\cos(2\omega' t) + \kappa_3\cos(3\omega t) + \gamma^3\kappa_3\cos(3\omega' t) \\ & + \gamma\mu_2\cos\big[(\omega+\omega')t\big] + \gamma\mu_2\cos\big[(\omega-\omega')t\big] \\ & + \gamma\mu_3\cos[(2\omega+\omega')t] + \gamma\mu_3\cos[(2\omega-\omega')t] \\ & + \gamma^2\mu_3\cos[(2\omega'+\omega)t] + \gamma^2\mu_3\cos[(2\omega'-\omega)t]\Big\} \;. \end{aligned} \tag{A.16}$$

In the special case of two interfering signals of equal amplitude ($\gamma = 1$), the output spectrum illustrated in Fig. A.4 results. If two strong interfering signals with angular frequencies ω' and ω'' pass through the nonlinear system, intermodulation products with angular frequencies $2\omega'' - \omega'$ and $2\omega' - \omega''$ will occur, which may be close to the angular frequency ω of a weak signal that has to be amplified. In linear amplifiers the coefficient μ_3 must be small to avoid desensitizing and intermodulation distortion.

Desensitizing. Assume the circuit under consideration to be an amplifier that has to amplify a weak signal $\hat{x}\cos(\omega t)$. If now $\hat{x}'\cos(\omega' t)$ is a strong interfering signal ($\gamma \gg 1$), the amplified signal, i.e. the component of the output spectrum with angular frequency ω, has an amplitude

$$\left(1 + (1+2\gamma^2)\mu_3\right) a_1\hat{x} \;=\; \left(1 + \frac{3a_3(\hat{x}^2+2\hat{x}'^2)}{4a_1}\right) a_1\hat{x}\;. \tag{A.17}$$

If $a_3 < 0$, as is generally the case, the output signal will decrease with increasing amplitude of the interferer, an effect commonly referred to as desensitizing or blocking [4].

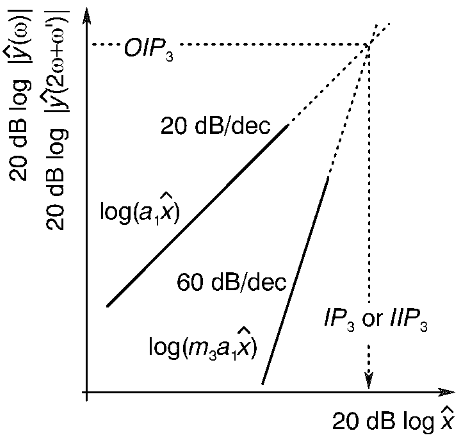


Fig. A.5. Definition of third-order intercept point

Intercept Point. In order to characterize intermodulation effects, the third-order intercept point IP_3 (Fig. A.5) is specified for interfering signals of equal amplitude ($\gamma = 1$). The amplitude of the third-order intermodulation products is then

$$m_3\alpha_1\hat{x} \;=\; 3\alpha_3\hat{x}^3/4\;, \tag{A.18}$$

i.e. its magnitude increases in proportion to the input signal amplitude $\hat{x}$, while the amplitude of the fundamental harmonic increases only in proportion to $\hat{x}$.

A.3 Nonlinear Systems with Memory

In linear systems without memory, the transfer function is independent of frequency, i.e. the input and output variables are related by an expression of the form

$$\underline{y}(s) = A\,\underline{x}(s) \tag{A.19}$$

in the frequency domain. In the time domain, the system function is represented as a δ-function: $A(t) = A\,\delta(t)$. Such behavior is found in resistive elements, where the current is a function of the voltage at that particular time only; in contrast, the current depends on the history of the applied voltage in reactive elements with capacitive or inductive behavior. In this case the transfer function is frequency-dependent as described in (A.3), resulting in the convolution integral (A.4) in the time domain.

A.3.1 Volterra Series

In nonlinear systems, higher-order response terms must be considered in addition, resulting in a Volterra series [3, 5, 6]:

$$\begin{aligned} y(t) &= \sum_{n=1}^{\infty} y_n(t) \\ &= \int_0^t A_1(t-t_1)\,x(t_1)\,\mathrm{d}t_1 \\ &\quad + \int_0^t\int_0^t A_2(t-t_1,t-t_2)\,x(t_1)x(t_2)\,\mathrm{d}t_1\,\mathrm{d}t_2 \\ &\quad + \int_0^t\int_0^t\int_0^t A_3(t-t_1,t-t_2,t-t_3)\,x(t_1)x(t_2)x(t_3)\,\mathrm{d}t_1\,\mathrm{d}t_2\,\mathrm{d}t_3 \\ &\quad + \cdots \end{aligned}$$

Generally, the nth-order term is expressed in terms of the Volterra kernel $A_n(t_1,\ldots,t_n)$:

$$y_n(t) = \int_0^t\cdots\int_0^t A_n(t-t_1,\ldots,t-t_n)\,x(t_1)\cdots x(t_n)\,\mathrm{d}t_1\cdots\mathrm{d}t_n\ .$$

Assume that $x(t)$ may be represented as the sum of K sinusoidal signals in the form

$$x(t) = x_0 + \sum_{k=1}^{K} \hat{x}_k \cos\,(\omega_k t+\varphi_k) = \frac{1}{2}\sum_{k=-K}^{K} \underline{\hat{x}}_k\,\mathrm{e}^{\mathrm{j}\omega_k t}\ , \tag{A.20}$$

where $\underline{\hat{x}}_k = \hat{x}_k\,\mathrm{e}^{\mathrm{j}\varphi_k} = \underline{\hat{x}}^*_{-k}$ and $\omega_{-k} = -\omega_k$. The nth-order response may then be written in the form [3]

$$y_n(t) = \frac{1}{2^n} \sum_{k_1=-K}^{K} \cdots \sum_{k_n=-K}^{K} \underline{A}_n(\omega_{k1}, \ldots, \omega_{kn}) \times$$
$$\times\, \underline{\hat{x}}_{k1} \cdots \underline{\hat{x}}_{kn} \exp\left[\mathrm{j}(\omega_{k1} + \cdots + \omega_{kn})t\right] \tag{A.21}$$

the nonlinear transfer factors $\underline{A}_n(\omega_1, \ldots, \omega_n)$ are independent of the amplitude of the excitation and completely characterize the nth-order response of a weakly nonlinear system (Fig. A.6).

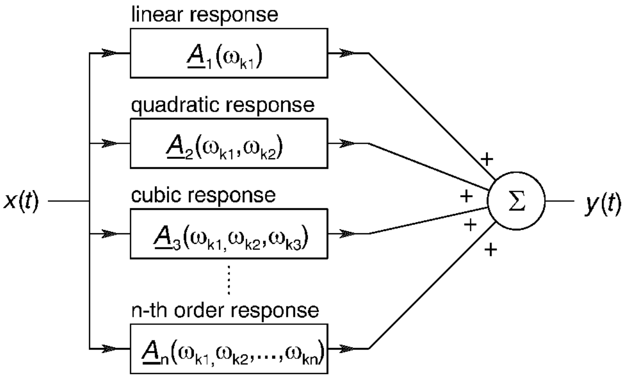


Fig. A.6. Graphical illustration of the procedure employed for the calculation of the output signal of a weakly nonlinear system in terms of nonlinear transfer functions

Evaluation of the sum (A.21) requires the addition of $(2K)^2$ summands, of which many, however, are equal: nonlinear transfer functions that differ only by a permutation of their arguments are equal. As an example, we consider, the third-order transfer function $\underline{A}_3(\omega_1, \omega_2, \omega_3)$, which satisfies the symmetry relations

$$\begin{aligned} \underline{A}_3(\omega_1, \omega_2, \omega_3) &= \underline{A}_3(\omega_1, \omega_3, \omega_2) = \underline{A}_3(\omega_3, \omega_1, \omega_2) \\ &= \underline{A}_3(\omega_3, \omega_2, \omega_1) = \underline{A}_3(\omega_2, \omega_1, \omega_3) = \underline{A}_3(\omega_2, \omega_3, \omega_1)\,. \end{aligned}$$

A.4 References

[1] W.C. Elmore. The transient response of damped linear networks with particular regard to wideband amplifiers. *J. Appl. Phys.*, 19:55–63, 1948.

[2] K.A. Simons. The decibel relationships between amplifier distortion products. *Proc. IEEE*, 58(7):1073–1086, 1970.

[3] D.D. Weiner, J.F. Spina. *Sinusoidal Analysis and Modeling of Weakly Nonlinear Circuits.* Van Nostrand, New York, 1980.

[4] R.G. Meyer, A.K. Wong. Blocking and desensitization in RF amplifiers. *IEEE J. Solid-State Circuits*, 30(8):944–946, 1995.

[5] S.A. Maas. *Nonlinear Microwave Circuits.* Artech House, Norwood, 1988.

[6] M. Schetzen. Nonlinear system modeling based on the Wiener theory. *Proc. IEEE*, 69(12):1557–1573, 1981.

B Linear Two-Ports, s-Parameters

This appendix compiles important properties of linear two-ports and s-parameters for ready reference by the reader.

B.1 Indefinite Admittance Matrix

A transistor has three terminals, with potentials $v_\alpha(t)$ with respect to ground and terminal currents $i_\beta(t)$, as illustrated in Fig. B.1.

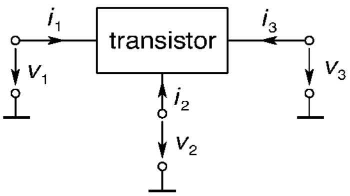


Fig. B.1. Terminal currents and terminal voltages with respect to ground for a three-terminal transistor

Under small-signal conditions, the terminal potentials

$$v_\alpha(t) \;=\; V_\alpha + \mathrm{Re}\left(\hat{\underline{v}}_\alpha \mathrm{e}^{\mathrm{j}\omega t}\right) \;=\; V_\alpha + \mathrm{Re}(\underline{v}_\alpha) \;, \qquad \alpha = 1,2,3$$

will cause terminal currents

$$i_\beta(t) \;=\; I_\beta + \mathrm{Re}\left(\hat{\underline{i}}_\beta \mathrm{e}^{\mathrm{j}\omega t}\right) \;=\; I_\beta + \mathrm{Re}(\underline{i}_\beta) \;, \qquad \beta = 1,2,3$$

with phasors $\underline{v}_\alpha$ and $\underline{i}_\beta$ which are related by the coefficients $(\tilde{y}_{ij})$ of the indefinite admittance matrix

$$\underline{i}_\alpha \;=\; \sum_{\beta=1}^{3} \tilde{y}_{\alpha\beta}\, \underline{v}_\beta \;, \qquad \alpha = 1,2,3 \;. \tag{B.1}$$

The indefinite admittance matrix has 3×3 coefficients, of which, however, only four are independent: Kirchhoff's current law $(\underline{i}_1 + \underline{i}_2 + \underline{i}_3 = 0)$ requires the coefficients $(\tilde{y}_{ij})$ in each row to add up to zero [1]. Furthermore, since the ground potential can be chosen at will, the result does not change if a constant potential is added to each potential value. This requires the coefficients $(\tilde{y}_{ij})$ in each column to add up to zero. The conditions

$$\forall\, \beta \;:\; \sum_{\alpha=1}^{3} \tilde{y}_{\alpha\beta} \;=\; 0 \quad \text{and} \quad \forall\, \alpha \;:\; \sum_{\beta=1}^{3} \tilde{y}_{\alpha\beta} \;=\; 0 \tag{B.2}$$

allow one to transform two-port parameters in common-emitter, common-base and common-collector configurations into one another.

Example. The small-signal description of a bipolar transistor in terms of the indefinite admittance matrix reads

$$\begin{pmatrix} \underline{i}_\mathrm{b} \\ \underline{i}_\mathrm{c} \\ -\underline{i}_\mathrm{e} \end{pmatrix} = \begin{pmatrix} \tilde{y}_{11} & \tilde{y}_{12} & \tilde{y}_{13} \\ \tilde{y}_{21} & \tilde{y}_{22} & \tilde{y}_{23} \\ \tilde{y}_{31} & \tilde{y}_{32} & \tilde{y}_{33} \end{pmatrix} \begin{pmatrix} \underline{v}_\mathrm{b} \\ \underline{v}_\mathrm{c} \\ \underline{v}_\mathrm{e} \end{pmatrix} .$$

Choosing $\underline{v}_\mathrm{e} = 0$ and thus $\underline{v}_\mathrm{b} = \underline{v}_\mathrm{be}$ and $\underline{v}_\mathrm{c} = \underline{v}_\mathrm{ce}$, one obtains

$$\begin{pmatrix} \underline{i}_\mathrm{b} \\ \underline{i}_\mathrm{c} \end{pmatrix} = \begin{pmatrix} y_\mathrm{11e} & y_\mathrm{12e} \\ y_\mathrm{21e} & y_\mathrm{22e} \end{pmatrix} \begin{pmatrix} \underline{v}_\mathrm{be} \\ \underline{v}_\mathrm{ce} \end{pmatrix} = \begin{pmatrix} \tilde{y}_{11} & \tilde{y}_{12} \\ \tilde{y}_{21} & \tilde{y}_{22} \end{pmatrix} \begin{pmatrix} \underline{v}_\mathrm{b} \\ \underline{v}_\mathrm{c} \end{pmatrix} ,$$

where the parameters $y_{\alpha\beta\mathrm{e}}$ are the admittance parameters of the common-emitter configuration. Choosing $\underline{v}_\mathrm{b} = 0$ and thus $\underline{v}_\mathrm{c} = \underline{v}_\mathrm{cb}$ and $\underline{v}_\mathrm{e} = \underline{v}_\mathrm{eb}$, one obtains

$$\begin{pmatrix} -\underline{i}_\mathrm{e} \\ \underline{i}_\mathrm{c} \end{pmatrix} = \begin{pmatrix} y_\mathrm{11b} & y_\mathrm{12b} \\ y_\mathrm{21b} & y_\mathrm{22b} \end{pmatrix} \begin{pmatrix} \underline{v}_\mathrm{eb} \\ \underline{v}_\mathrm{cb} \end{pmatrix} = \begin{pmatrix} \tilde{y}_{33} & \tilde{y}_{32} \\ \tilde{y}_{23} & \tilde{y}_{22} \end{pmatrix} \begin{pmatrix} \underline{v}_\mathrm{c} \\ \underline{v}_\mathrm{e} \end{pmatrix} .$$

The coefficients $y_{\alpha\beta\mathrm{b}}$ are termed the admittance parameters of the common-base configuration and may be calculated from the admittance parameters in the common-emitter configuration with the help of the identities (B.2), which give

$$\begin{aligned} y_\mathrm{11b} &= y_\mathrm{11e} + y_\mathrm{12e} + y_\mathrm{21e} + y_\mathrm{22e} \ , \\ y_\mathrm{12b} &= -(y_\mathrm{12e} + y_\mathrm{22e}) \ , \\ y_\mathrm{21b} &= -(y_\mathrm{21e} + y_\mathrm{22e}) \ , \\ y_\mathrm{22b} &= y_\mathrm{22e} \ . \end{aligned}$$

B.2 Terminated Two-Ports

In the case of small-signal operation, transistors may be represented as linear two-ports. Since transistors are active devices, power gain may be achieved, depending on the termination of the two-port. This may lead to unwanted oscillatory behavior, and conditions for stability are required. Another problem arises in the measurement of the two-port parameters over a wide range of frequencies; in the case of a characterization in terms of admittance parameters, this requires the realization of a short-circuit condition[1] for the ac small signal superimposed on the dc bias voltages. Since this is difficult to achieve at high frequencies, s-parameters are generally employed for the characterization of the small-signal behavior of bipolar transistors at high frequencies.

[1] The open-circuit condition required for the measurement of impedance parameters is even more difficult to realize at high frequencies.

B.2.1 Input and Output Impedance

In calculating the properties of two-port networks it is often helpful to replace these by equivalent one-ports. If the two-port is seen from the input, its

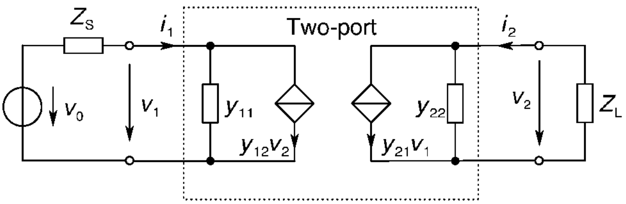


Fig. B.2. Terminated two-port circuit

input impedance Z_i, together with the source impedance Z_S, determines the power delivered to the two-port by the signal source; if it is seen from the output, the output impedance Z_o (determined for $v_0 = 0$), together with the load impedance Z_L, determines the power delivered to the load (Fig. B.2). The value of the input impedance depends on the load connected to the output port, while the value of the output impedance depends on the circuitry connected to the input port. If a voltage source with source impedance Z_S is applied to the input of a two-port whose output is connected to a load impedance Z_L, the following equations must be fulfilled:

$$\underline{V}_1 = \underline{V}_0 - Z_\mathrm{S}\,\underline{I}_1 \quad \text{and} \quad \underline{V}_2 = -Z_\mathrm{L}\,\underline{I}_2 \,. \tag{B.3}$$

Combination of (B.3) with the two-port equations in admittance and hybrid form yields the input impedance

$$Z_\mathrm{i} = \frac{\underline{V}_1}{\underline{I}_1} = \frac{1 + y_{22}Z_\mathrm{L}}{y_{11} + \Delta_y Z_\mathrm{L}} = \frac{h_{11} + \Delta_h Z_\mathrm{L}}{1 + h_{22}Z_\mathrm{L}} \,, \tag{B.4}$$

and the output impedance ($\underline{V}_0 = 0$)

$$Z_\mathrm{o} = \frac{\underline{V}_2}{\underline{I}_2} = \frac{1 + y_{11}Z_\mathrm{S}}{y_{22} + \Delta_y Z_\mathrm{S}} = \frac{h_{11} + Z_\mathrm{S}}{\Delta_h + h_{22}Z_\mathrm{S}} \,. \tag{B.5}$$

B.2.2 Voltage and Current Gain

The voltage transfer factor of the terminated two-port is given by

$$\underline{A}_v = \frac{\underline{V}_2}{\underline{V}_1} = \frac{-y_{21}Z_\mathrm{L}}{1 + y_{22}Z_\mathrm{L}} = \frac{-h_{21}Z_\mathrm{L}}{h_{11} + \Delta_h Z_\mathrm{L}} \,; \tag{B.6}$$

its magnitude $A_v(\omega) = |\underline{A}_v(\mathrm{j}\omega)|$ is called the voltage gain. In complete analogy, the current transfer factor of the terminated two-port is defined as the ratio of the phasors of the input and output currents,

$$\underline{A}_i = \frac{\underline{I}_2}{\underline{I}_1} = \frac{y_{21}}{y_{11} + \Delta_y Z_\mathrm{L}} = \frac{h_{21}}{1 + h_{22} Z_\mathrm{L}} , \tag{B.7}$$

and its magnitude $A_i(\omega) = |\underline{A}_i(\mathrm{j}\omega)|$ is called the current gain. The voltage and current gains may be expressed in dB in the forms

$$a_v = 20\,\mathrm{dB} \times \log(A_v) \tag{B.8}$$

and

$$a_i = 20\,\mathrm{dB} \times \log(A_i) , \tag{B.9}$$

respectively.

B.2.3 Power Gain

Several definitions of power gain are used to characterize the power gain of the terminated two-port circuit shown in Fig. B.2 [2].

Operating Power Gain. The operating power gain[2] G_p is defined as the ratio of the effective power $P_2 = -\mathrm{Re}\,(\underline{V}_2\,\underline{I}_2^*)$ delivered to the load to the effective power $P_1 = \mathrm{Re}\,(\underline{V}_1\,\underline{I}_1^*)$ absorbed by the input port. If we set $\underline{V}_1 = Z_\mathrm{i}\,\underline{I}_1 = \underline{I}_1/Y_\mathrm{i}$ and $\underline{V}_2 = -Z_\mathrm{L}\,\underline{I}_2 = -\underline{I}_2/Y_\mathrm{L}$, the power gain is obtained as

$$G_p(f) = \frac{P_2}{P_1} = A_i^2(f)\frac{\mathrm{Re}(Z_\mathrm{L})}{\mathrm{Re}(Z_\mathrm{i})} = A_v^2(f)\frac{\mathrm{Re}(Y_\mathrm{L})}{\mathrm{Re}(Y_\mathrm{i})} . \tag{B.10}$$

The value of G_p depends on the load impedance $Z_\mathrm{L} = 1/Y_\mathrm{L}$, but is not affected by the source impedance $Z_\mathrm{S} = 1/Y_\mathrm{S}$. In terms of the admittance parameters, the power gain is given by

$$G_p = \frac{|y_{21}|^2 \mathrm{Re}(Y_\mathrm{L})}{|y_{22} + Y_\mathrm{L}|^2} \left(y_{11} - \frac{y_{12}\,y_{21}}{y_{22} + Y_\mathrm{L}} \right) . \tag{B.11}$$

The power gain may also be expressed[3], in dB according to

$$a_p = 10\,\mathrm{dB} \times \log(G_p) ; \tag{B.12}$$

since power varies in proportion to V^2 or I^2, the different "prefactors" (20 dB and 10 dB) used for the specification of voltage or current gain and power gain, respectively, guarantee at least a close relation between the two quantities.

[2] Also termed "average power gain" or simply "power gain".

[3] Only dimensionless quantities may be expressed in dB, i.e. this "quantity" may only be used for the specification of ratios, namely relative signal levels. The specification of absolute signal levels is possible if a reference level is given. Power specifications in dBm, which specify the rms value of the power P relative to 1 mW, i.e.

$$P \text{ in dBm} \equiv 10\,\mathrm{dB} \times \log\,(P/1\,\mathrm{mW})$$

are frequently employed.

Example. As an example, consider a linear system with input resistance R_1 and a load resistance R_2 connected to the output, as depicted in Fig. B.3. The (average) power delivered to the input is $P_1 = V_1^2/R_1$, where V_1 is the

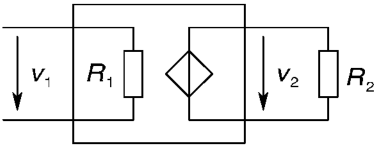


Fig. B.3. The relation between power and voltage or current gain (see text)

rms value of the input voltage; in complete analogy, the power delivered to the load can be derived from the rms value of the voltage V_2 according to $P_2 = V_2^2/R_2$. Therefore

$$\begin{aligned} 10\,\mathrm{dB} \times \log\left(\frac{P_2}{P_1}\right) &= 10\,\mathrm{dB} \times \log\left(\frac{V_2^2}{V_1^2}\frac{R_1}{R_2}\right) \\ &= 20\,\mathrm{dB} \times \log\left(\frac{V_2}{V_1}\right) + 10\,\mathrm{dB} \times \log\left(\frac{R_1}{R_2}\right) \ , \end{aligned}$$

i.e. a_p equals a_v if R_1 and R_2 have the same value.

Transducer Power Gain. The average power P_1 delivered from the signal source (rms voltage V_0) to the input of the two-port ia a maximum[4] in the case of conjugate matching, i.e. if $Z_\mathrm{i} = Z_\mathrm{S}^*$; in this case the power

$$P_1 = \frac{\mathrm{Re}(Z_\mathrm{i})}{|Z_\mathrm{i} + Z_\mathrm{S}|^2} V_0^2 = \frac{V_0^2}{4\,\mathrm{Re}(Z_\mathrm{i})} = P_1^+$$

is absorbed by the input of the two-port. This is just half of the power P_0 delivered by the signal source (the other half is dissipated in the source impedance Z_S). The transducer power gain G_T is defined as P_2/P_1^+, where P_2 is the power delivered to the load. In terms of the admittance parameters, the transducer power gain is given by

$$G_\mathrm{T} = \frac{4\,|y_{21}|^2\,\mathrm{Re}(Y_\mathrm{S})\mathrm{Re}(Y_\mathrm{L})}{|(y_{11}+Y_\mathrm{S})(y_{22}+Y_\mathrm{L}) - y_{12}y_{21}|^2} \ . \tag{B.13}$$

The value of G_T is equal to value of G_p if the input of the two-port is matched properly to the source. The transducer power gain relates the power that may be obtained from the source under optimum conditions to the power delivered to the load.

[4] A proof of this statement is given in [2], for example.

Available Power Gain. The available gain is defined for the case of conjugate matching at the input and output of the two-port: in this case the power P_2 delivered to the load achieves its maximum value (see Fig. B.2)

$$P_2 = \frac{|y_{21}|^2}{4\,\mathrm{Re}(y_{22})}\,V_1^2 = P_2^-$$

if the output impedance of the two-port is conjugately matched to the load impedance ($Z_o = Z_L^*$). The available gain G_A is now defined as the ratio of the average power available at the output to the power delivered by the signal source: $G_A = P_2^-/P_1^+$. In terms of the admittance parameters, the available power gain is given by

$$G_A = \frac{|y_{21}|^2 \mathrm{Re}(Y_S)}{|y_{11}+Y_S|^2}\left(y_{22} - \frac{y_{12}y_{21}}{y_{11}+Y_S}\right) . \tag{B.14}$$

The value of G_A is equal to the value of G_p and G_T if both the input and the output are matched properly to the source and load.

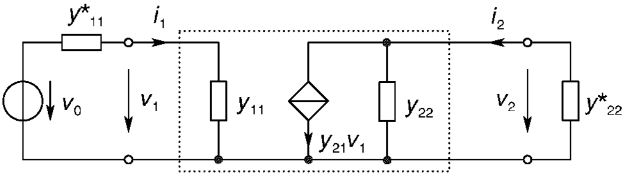


Fig. B.4. Circuit used for the computation of unilateral power gain

Unilateral Power Gain. A two-port is termed unilateral if $y_{12} = 0$; this situation is depicted in Fig. B.4, where conjugate matching at the input and output ports is assumed. Since

$$\underline{I}_2 = y_{21}\,\underline{V}_1 + y_{22}\,\underline{V}_2 = -y_{22}^*\,\underline{V}_2 ,$$

the complex rms value of the output voltage is

$$\underline{V}_2 = -\frac{y_{21}}{2\,\mathrm{Re}(y_{22})}\,\underline{V}_1 .$$

This yields the following for the power absorbed by the load:

$$P_2^- = -\frac{1}{2}\left(\underline{I}_2\underline{V}_2^* + \underline{I}_2^*\underline{V}_2\right) = \frac{|y_{21}|^2}{4\,\mathrm{Re}(y_{22})}\,V_1^2 .$$

Since the power delivered to the input is given by $P_1^+ = \mathrm{Re}(y_{11})V_1^2$ if the input is conjugately matched to the load, the unilateral power gain U is obtained as

$$U = \frac{P_2^-}{P_1^+} = \frac{|y_{21}|^2}{4\,\mathrm{Re}(y_{11})\mathrm{Re}(y_{22})} . \tag{B.15}$$

Transistor two-ports are not strictly unilateral, since y_{12} is generally different from zero. However, by unilateralization, i.e. by adding external feedback, it is

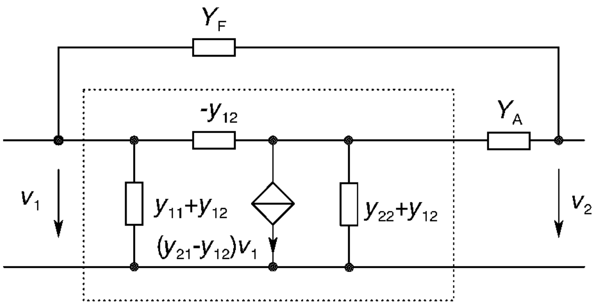


Fig. B.5. Unilateralization by adding two lossless admittances Y_F and Y_A

possible to realize unilateral conditions. This is illustrated in Fig. B.5, where two lossless admittances Y_A and Y_F are added to the transistor two-port. The admittance matrix $Y_{\alpha\beta}$ of the feedback circuit is given by

$$\mathsf{Y} = \frac{Y_A}{Y_A + y_{22}} \begin{pmatrix} Y_{11} + \Delta_y / Y_A & y_{12} \\ y_{21} & y_{22} \end{pmatrix} + \begin{pmatrix} Y_F & -Y_F \\ -Y_F & Y_F \end{pmatrix} . \tag{B.16}$$

The added lossless admittances are purely imaginary; unilateral conditions are achieved if

$$Y_F = \mathrm{j}\left(\mathrm{Im}(y_{12}) - \frac{\mathrm{Re}(y_{12})}{\mathrm{Re}(y_{22})}\mathrm{Im}(y_{22})\right) \quad \text{and} \quad Y_A = \frac{\mathrm{Re}(y_{22})}{\mathrm{Re}(y_{12})} Y_F$$

resulting in the unilateral gain

$$\begin{aligned} U &= \frac{|Y_{21}|^2}{4\,\mathrm{Re}(Y_{11})\,\mathrm{Re}(Y_{22})} \\ &= \frac{|y_{21} - y_{12}|^2}{4\left[\mathrm{Re}(y_{11})\mathrm{Re}(y_{22}) - \mathrm{Re}(y_{12})\mathrm{Re}(y_{21})\right]} . \end{aligned} \tag{B.17}$$

In contrast to other definitions of power gain, the unilateral power gain is invariant under lossless reciprocal embedding [5] and therefore provides a unique measure for the specification of the power amplification capability of a device. If the transistor can be considered as a three-terminal device,[5] the unilateral gain is the same in the common-emitter, common-base and common-collector configurations. Oscillations are possible if $U \geq 1$. The maximum frequency of oscillation f_{max}, defined by the condition $U(f_{max}) = 1$, is the maximum frequency for which power gain can be obtained from the transistor.

[5]In integrated bipolar transistors, the grounded substrate contact has to be considered in addition. In this case a change of the transistor configuration changes the equivalent circuit, resulting in different values of the unilateral gain for different configurations.

B.2.4 Stability

A linear circuit is unstable if its output increases without limit.[6]. The stability of a two-port network depends on the terminations. A two-port is potentially unstable at the angular frequency ω if there are passive two-port terminations that produce self-sustained oscillations with angular frequency ω. If no such terminations can be found, the two-port is absolutely stable. For absolutely stable operation, the real part of the input admittance

$$Y_{\mathrm{i}} = \frac{y_{11}Y_{\mathrm{L}} + \Delta_y}{y_{22} + Y_{\mathrm{L}}}$$

has to be positive for all passive load admittances Y_{L}. Using the representations $y_{\alpha\beta} = g_{\alpha\beta} + \mathrm{j}b_{\alpha\beta}$ and $Y_{\mathrm{L}} = G_{\mathrm{L}} + \mathrm{j}B_{\mathrm{L}}$, the real part of Y_{i} is [2]

$$\mathrm{Re}(Y_{\mathrm{i}}) = g_{11}\,\frac{(G_{\mathrm{L}}+g_{22}-G_1)^2 + (B_{\mathrm{L}}+b_{22}-B_1)^2 - G_1^2 - B_1^2}{(g_{22}+G_{\mathrm{L}})^2 + (b_{22}+B_{\mathrm{L}})^2}\,,$$

where $G_1 = \mathrm{Re}(y_{21}y_{12})/(2g_{11})$ and $B_1 = \mathrm{Im}(y_{21}y_{12})/(2g_{11})$. Since the denominator is positive, $\mathrm{Re}(Y_{\mathrm{i}})$ is positive if $g_{11} > 0$ and [7]

$$(G_{\mathrm{L}}+g_{22}-G_1)^2 + (B_{\mathrm{L}}+b_{22}-B_1)^2 - G_1^2 - B_1^2 > 0\,.$$

Since the second term on the left-hand side is zero in the worst case, this requires that

$$\Xi = (G_{\mathrm{L}}+g_{22}-G_1)^2 - G_1^2 - B_1^2 > 0$$

has to be fulfilled. This is the case if no positive value of G_{L} can be found for which $\Xi(G_{\mathrm{L}}) = 0$. Since the solution of this quadratic equation is given by

$$G_{\mathrm{L}} = G_1 - g_{22} + \sqrt{G_1^2+B_1^2}\,,$$

the circuit is absolutely stable if

$$\sqrt{G_1^2+B_1^2} = \frac{|y_{12}y_{21}|}{2g_{11}} < g_{22} - G_1 = g_{22} - \frac{\mathrm{Re}(y_{12}y_{21})}{2g_{11}}$$

is fulfilled. If we introduce the stability factor

$$k = \frac{2g_{11}g_{22} - \mathrm{Re}(y_{12}y_{21})}{|y_{12}y_{21}|}\,, \tag{B.18}$$

this is equivalent to the condition $k > 1$. The value of the stability factor is between -1 and ∞; if k lies between -1 and $+1$, the circuit is potentially unstable. The stability factor is calculated solely from the two-port parameters,

[6] In practical circuits, nonlinear effects limit the signal amplitude.

[7] If $g_{11} < 0$, oscillation is possible if the output is short-circuited: letting $G_{\mathrm{L}} \to \infty$ results in $\mathrm{Re}(Y_{\mathrm{i}}) = g_{11}$.

and serves therefore as a device-specific parameter that is independent of the termination chosen. If $k > 1$ the optimum terminations $Y_\mathrm{S}^{(\mathrm{opt})}$ and $Y_\mathrm{L}^{(\mathrm{opt})}$ are obtained from the matching conditions, which provide two equations for two unknowns, resulting in

$$Y_\mathrm{S}^{(\mathrm{opt})} = \frac{|y_{12}\,y_{21}|\sqrt{k^2-1}}{2\,\mathrm{Re}(y_{22})} + \mathrm{j}\left(\frac{\mathrm{Im}(y_{12}y_{21})}{2\,\mathrm{Re}(y_{22})} - \mathrm{Im}(y_{11})\right) ,$$

$$Y_\mathrm{L}^{(\mathrm{opt})} = \frac{|y_{12}\,y_{21}|\sqrt{k^2-1}}{2\,\mathrm{Re}(y_{11})} + \mathrm{j}\left(\frac{\mathrm{Im}(y_{12}y_{21})}{2\,\mathrm{Re}(y_{11})} - \mathrm{Im}(y_{22})\right) .$$

The maximum available gain (MAG)

$$\mathrm{MAG} = \frac{|y_{21}|}{|y_{12}|}\left(k - \sqrt{k^2-1}\right) \tag{B.19}$$

equals the transducer power gain if both sides are conjugately matched. If the MAG cannot be calculated, the maximum stable gain (MSG) determined as the MAG obtained with the minimum series resistance added to achieve stable operation, is of interest. The maximum stable gain is given by

$$\mathrm{MSG} = |y_{21}|/|y_{12}| \; ; \tag{B.20}$$

obviously MSG = MAG if $k = 1$.

B.2.5 Incident and Reflected Power

If the input is conjugately matched to the source, the input voltage (complex rms value) is given by

$$\underline{V}_1 = Z_\mathrm{S}^*\underline{V}_0/2R_\mathrm{S} = \underline{V}_1^+ , \quad \text{where} \quad \mathrm{Re}(Z_\mathrm{S}) = R_\mathrm{S} ,$$

while the input current is

$$\underline{I}_1 = \underline{V}_0/2R_\mathrm{S} = \underline{I}_1^+ .$$

Following [3], $\underline{V}_1^+$ and $\underline{I}_1^+$ are called the incident voltage and the incident current (at the input port). The product of these quantities determines the power $P_1^+ = \mathrm{Re}\,(\underline{V}_1^+\underline{I}_1^{+*})$ available from the generator. If the matching condition is not fulfilled, the input voltage $\underline{V}_1$ may be represented as the sum of the incident voltage $\underline{V}_1^+$ and the reflected voltage $\underline{V}_1^-$:

$$\underline{V}_1 = \frac{Z_\mathrm{i}}{Z_\mathrm{S}+Z_\mathrm{i}}\,\underline{V}_0 = \underline{V}_1^+ + \underline{V}_1^- = (1+\Gamma_\mathrm{iv})\,\underline{V}_1^+ .$$

The voltage reflection coefficient

$$\Gamma_\mathrm{iv} = \frac{\underline{V}_1}{\underline{V}_1^+} - 1 = \frac{Z_\mathrm{i}}{Z_\mathrm{i}+Z_\mathrm{S}}\,\frac{Z_\mathrm{S}+Z_\mathrm{S}^*}{Z_\mathrm{S}^*} - 1 = \frac{Z_\mathrm{i}-Z_\mathrm{S}^*}{Z_\mathrm{i}+Z_\mathrm{S}}\,\frac{Z_\mathrm{S}}{Z_\mathrm{S}^*} \tag{B.21}$$

vanishes if the input is conjugately matched to the source ($Z_\mathrm{i} = Z_\mathrm{S}^*$). In the same way, the input current may be represented as the difference of an incident current $\underline{I}_1^+$ and a reflected current $\underline{I}_1^-$:

$$\underline{I}_1 = \frac{1}{Z_\mathrm{S}+Z_\mathrm{i}}\underline{V}_0 = \underline{I}_1^+ - \underline{I}_1^- = (1-\Gamma_\mathrm{ii})\underline{I}_1^+ ,$$

where Γ_ii denotes the current reflection coefficient

$$\Gamma_\mathrm{ii} = 1 - \frac{\underline{I}_1}{\underline{I}_1^+} = 1 - \frac{Z_\mathrm{S}+Z_\mathrm{S}^*}{Z_\mathrm{i}+Z_\mathrm{S}} = \frac{Z_\mathrm{i}-Z_\mathrm{S}^*}{Z_\mathrm{i}+Z_\mathrm{S}} .$$

The voltage and current reflection coefficients are related by

$$\Gamma_\mathrm{iv}\, Z_\mathrm{S}^* = \Gamma_\mathrm{ii}\, Z_\mathrm{S} , \tag{B.22}$$

i.e. the two quantities are equal if the source impedance is real. The power

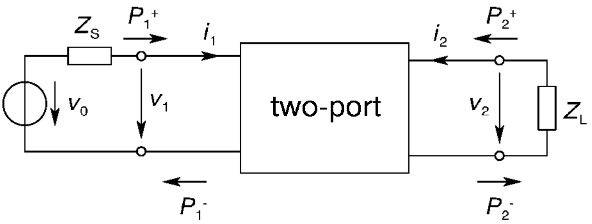


Fig. B.6. Two-port with terminated output and excitation of the input

absorbed by the input of the two-port is given by

$$P_1 = \frac{1}{2}\left(\underline{V}_1\underline{I}_1^* + \underline{V}_1^*\underline{I}_1\right) = \frac{R_\mathrm{i}V_0^2}{|Z_\mathrm{S}+Z_\mathrm{i}|^2} ,$$

where $R_\mathrm{i} = \mathrm{Re}(Z_\mathrm{i})$. It may be represented as the power (Fig. B.6)

$$P_1^+ = \frac{V_0^2}{4R_\mathrm{S}} = \mathrm{Re}(\underline{V}^+\underline{I}^{+*}) = \frac{R_\mathrm{S}}{|Z_\mathrm{S}|^2}\,|\underline{V}_1^+|^2 = |\underline{a}_1|^2$$

available from the generator, diminished by the power

$$P_1^- = \frac{V_0^2}{4R_\mathrm{S}} - \frac{R_\mathrm{i}V_0^2}{|Z_\mathrm{S}+Z_\mathrm{i}|^2} = \frac{1}{4R_\mathrm{S}}\,\frac{|Z_\mathrm{S}-Z_\mathrm{i}^*|^2}{|Z_\mathrm{S}+Z_\mathrm{i}|^2}\,V_0^2$$

$$= \frac{R_\mathrm{S}}{|Z_\mathrm{S}|^2}\,|\underline{V}_1^-|^2 = |\underline{b}_1|^2$$

reflected at the input of the two-port, where we have introduced the incident- and reflected-power parameters

$$\underline{a}_1 = \frac{\sqrt{R_\mathrm{S}}}{Z_\mathrm{S}^*}\,\underline{V}_1^+ \quad \text{and} \quad \underline{b}_1 = \frac{\sqrt{R_\mathrm{S}}}{Z_\mathrm{S}}\,\underline{V}_1^- . \tag{B.23}$$

These are related to the input voltage $\underline{V}_1$ and input current $\underline{I}_1$ by

$$\underline{V}_1 = \underline{V}_1^+ + \underline{V}_1^- = \frac{Z_S^*}{\sqrt{R_S}}\underline{a}_1 + \frac{Z_S}{\sqrt{R_S}}\underline{b}_1 , \tag{B.24}$$

$$\underline{I}_1 = \underline{I}_1^+ - \underline{I}_1^- = \frac{1}{\sqrt{R_S}}\underline{a}_1 - \frac{1}{\sqrt{R_S}}\underline{b}_1 , \tag{B.25}$$

or, vice versa,

$$\underline{a}_1 = \frac{\underline{V}_1 + Z_S\underline{I}_1}{2\sqrt{R_S}} \quad \text{and} \quad \underline{b}_1 = \frac{\underline{V}_1 - Z_S^*\underline{I}_1}{2\sqrt{R_S}} . \tag{B.26}$$

Together with the analogously defined parameters $\underline{a}_2$ and $\underline{b}_2$ for the output port, this analysis introduces a new set of variables ($\underline{a}_1$, $\underline{b}_1$, $\underline{a}_2$ and $\underline{b}_2$) for the description of the behavior of a two-port (Fig. B.7). These variables allow us to define the transmission properties of the two-port in terms of scattering parameters, or s-parameters.

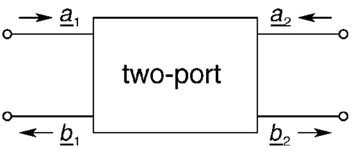


Fig. B.7. Incident and reflected waves in a two-port network

B.3 S-Parameters

The high-frequency behavior of linear two-ports is usually characterized in terms of s-parameters, since the open- and short-circuit conditions required for the measurement of y- and h-parameters are hard to realize at high frequencies, owing to lead inductances and fringing capacitances. Furthermore, in the case of active two-ports, the required open-circuit or short-circuit conditions often result in oscillation. These problems are overcome in s-parameter measurements: s-parameters are directly related to the incident and reflected power and do not vary in amplitude along a lossless transmission line. This allows one to place the device under test at some distance from the measurement apparatus (Sect. 3.10). In addition, problems caused by unstable oscillation may be avoided by impedance matching. Since s-parameters may be determined using commercially available measurement equipment far into the gigahertz range [4], the behavior of two-ports in the megahertz and gigahertz range is generally characterized with s-parameters.

For linear two-ports, a linear relation exists between the power parameters at the input port and those at the output port:

$$\begin{pmatrix} \underline{b}_1 \\ \underline{b}_2 \end{pmatrix} = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} \begin{pmatrix} \underline{a}_1 \\ \underline{a}_2 \end{pmatrix} . \tag{B.27}$$

The frequency-dependent parameters $s_{\alpha\beta}$ are referred to as scattering parameters, the matrix composed of these parameters is the scattering matrix. The scattering parameters are defined as ratios of ingoing and outgoing wave amplitudes and are therefore dimensionless. The measurement conditions $\underline{a}_1 = 0$ and $\underline{a}_2 = 0$ required for the determination of the s-parameters correspond to the proper termination of the input and output port, respectively. The input reflection coefficient

$$s_{11} = \left. \frac{\underline{b}_1}{\underline{a}_1} \right|_{\underline{a}_2=0} = \frac{\underline{V}_1/\underline{I}_1 - Z_{01}}{\underline{V}_1/\underline{I}_1 + Z_{01}} = \frac{Z_{\mathrm{i}1} - Z_{01}}{Z_{\mathrm{i}1} + Z_{01}} = \Gamma_{\mathrm{i}} \tag{B.28}$$

equals the reflection coefficient at the input of the two-port if the output port is properly terminated. Its squared magnitude $|s_{11}|^2$ is equal to the ratio of the power reflected at the input port to the power available at the input port. If $|s_{11}| > 1$ the reflected power is larger than the incident power, i.e. the input port acts as a power source, and oscillations may occur. The forward transmission coefficient

$$s_{21} = \left. \frac{\underline{b}_2}{\underline{a}_1} \right|_{\underline{a}_2=0} = \sqrt{\frac{Z_{01}}{Z_{02}}} \frac{\underline{V}_2 - Z_{02}\underline{I}_2}{\underline{V}_1 + Z_{01}\underline{I}_1} = \sqrt{\frac{Z_{01}}{Z_{02}}} \frac{\underline{V}_2}{\underline{V}_{01}/2} \tag{B.29}$$

determines the reaction of the output to the input signal: the squared magnitude $|s_{21}|^2$ represents the ratio of the power delivered by the output terminated with Z_{02}, to the power available at the input port. Analogously, the reverse transmission coefficient

$$s_{12} = \left. \frac{\underline{b}_1}{\underline{a}_2} \right|_{\underline{a}_1=0} \tag{B.30}$$

determines the reaction of the input to the output signal. The output reflection coefficient

$$s_{22} = \left. \frac{\underline{b}_2}{\underline{a}_2} \right|_{\underline{a}_1=0} = \frac{Z_{\mathrm{i}2} - Z_{02}}{Z_{\mathrm{i}2} + Z_{02}} \tag{B.31}$$

equals the reflection coefficient at the output port if the input port is properly terminated. Its squared magnitude $|s_{22}|^2$ equals the ratio of the power reflected at the output port to the power available at the output port if the input port is terminated with Z_{01}.

B.3.1 Relations between s-Parameters and Two-Port Parameters

A full set of s-parameters completely describes a linear two-port and must therefore contain the same information as a full set of admittance or hybrid parameters. The admittance parameters may be calculated from the scattering parameters as follows:

$$
\mathsf{y} = \frac{1}{Z_0}\begin{pmatrix} \dfrac{1-s_{11}+s_{22}-\Delta_s}{1+s_{11}+s_{22}+\Delta_s} & \dfrac{-2s_{12}}{1+s_{11}+s_{22}+\Delta_s} \\ \dfrac{-2s_{21}}{1+s_{11}+s_{22}+\Delta_s} & \dfrac{1+s_{11}-s_{22}-\Delta_s}{1+s_{11}+s_{22}+\Delta_s} \end{pmatrix} .
$$

Vice versa, writing $y'_{\alpha\beta} = Z_0\, y_{\alpha\beta}$,

$$
\mathsf{s} = \begin{pmatrix} \dfrac{1-y'_{11}+y'_{22}-\Delta'_y}{1+y'_{11}+y'_{22}+\Delta'_y} & \dfrac{-2y'_{12}}{1+y'_{11}+y'_{22}+\Delta'_y} \\ \dfrac{-2y'_{21}}{1+y'_{11}+y'_{22}+\Delta'_y} & \dfrac{1+y'_{11}-y'_{22}-\Delta'_y}{1+y'_{11}+y'_{22}+\Delta'_y} \end{pmatrix} .
$$

In the same way, the hybrid parameters can be obtained from given scattering parameters according to

$$
\mathsf{h} = \begin{pmatrix} Z_0\,\dfrac{1+s_{11}+s_{22}+\Delta_s}{1-s_{11}+s_{22}-\Delta_s} & \dfrac{2s_{12}}{1-s_{11}+s_{22}-\Delta_s} \\ \dfrac{-2s_{21}}{1-s_{11}+s_{22}-\Delta_s} & \dfrac{1}{Z_0}\,\dfrac{1-s_{11}-s_{22}+\Delta_s}{1-s_{11}+s_{22}-\Delta_s} \end{pmatrix}
$$

and, vice versa,

$$
\mathsf{s} = \begin{pmatrix} \dfrac{-1+h_{11}/Z_0-h_{22}Z_0+\Delta_h}{1+h_{11}/Z_0+h_{22}Z_0+\Delta_h} & \dfrac{2h_{12}}{1+h_{11}+h_{22}+\Delta_h} \\ \dfrac{-2h_{21}}{1+h_{11}/Z_0+h_{22}Z_0+\Delta_h} & \dfrac{1+h_{11}/Z_0-h_{22}Z_0-\Delta_h}{1+h_{11}/Z_0+h_{22}Z_0+\Delta_h} \end{pmatrix} .
$$

B.3.2 Matching and Power Gain

Consider a generator that emits a power wave $\underline{b}_S$ towards a terminated two-port with input reflection coefficient Γ_i. If $\Gamma_i \neq 0$, reflections at the input

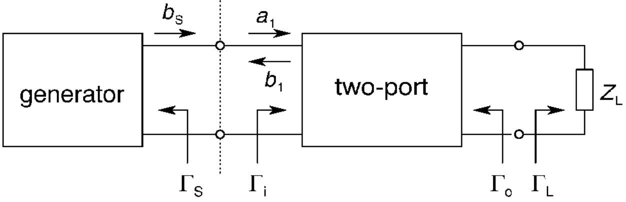


Fig. B.8. Reflection at the input of a terminated two-port

occur, resulting in a reflected power wave $\Gamma_i b_S$ incident on the generator output (Fig. B.8). If the reflection coefficient Γ_S of the generator as seen from the two-port differs from zero, a further reflected power wave $\Gamma_S \Gamma_i b_S$ incident on the input of the two-port results, which again causes reflections, etc. Summing these reflections, we obtain a geometric series for the power wave reflected at the input:

$$b_1 = \Gamma_i b_S + (\Gamma_i \Gamma_S)\Gamma_i b_S + (\Gamma_i \Gamma_S)^2 \Gamma_i b_S + \cdots = \frac{\Gamma_i b_S}{1-\Gamma_i \Gamma_S} . \tag{B.32}$$

Since $b_1 = \Gamma_i a_1$, this results in

$$a_1 = \frac{b_S}{1-\Gamma_i \Gamma_S} = b_S - \Gamma_S b_1 . \tag{B.33}$$

The transducer power gain G_T of the two-port is given by

$$G_T = |s_{21}|^2 \frac{(1-|\Gamma_L|^2)(1-|\Gamma_G|^2)}{|(1-\Gamma_L s_{22})(1-\Gamma_G s_{11}) - \Gamma_L \Gamma_G s_{12} s_{21}|^2} , \tag{B.34}$$

where Γ_o is the reflection of the two-port output seen from the load. In a 50 Ω system, $\Gamma_S = \Gamma_L = 0$, and G_T is simply $|s_{21}|^2$. Maximum available gain MAG, maximum stable gain MSG and the unilateral gain U can be expressed in s-parameters as follows:

$$\mathrm{MAG} = \frac{|s_{21}|}{|s_{12}|}\left(k - \sqrt{k^2-1}\right) \tag{B.35}$$

$$\mathrm{MSG} = \frac{|s_{21}|}{|s_{12}|} \tag{B.36}$$

$$U = \frac{|s_{21}/s_{12} - 1|^2}{2k|s_{21}/s_{12}| - \mathrm{Re}(s_{21}/s_{12})} , \tag{B.37}$$

where

$$k = \frac{1 - |s_{11}|^2 - |s_{22}|^2 + |\Delta_s|^2}{2\,|s_{21}||s_{12}|} \tag{B.38}$$

These relations are used for the determination of the maximum frequency of oscillation from s-parameter measurements.

B.4 References

[1] L.O. Chua; C.A. Desoer; E.S. Kuh. *Linear and Nonlinear Circuits.* McGraw Hill, New York, 1991.

[2] R.S. Carson. *High-Frequency Amplifiers.* Wiley, New York, 1982.

[3] U.L Rohde G.D. Vendelin, A.M. Pavio. *Microwave Circuit Design Using Linear and Nonlinear Techniques.* Wiley, New York, 1990.

[4] Hewlett Packard. S-parameter design. *Application Note 154*, 1972.

[5] J. Choma Jr., W.-K. Chen. *Linear two-port networks.* In: The Circuits and Filters Handbook (W.-K. Chen, ed.). CRC Press, Boca Raton, 1995.

C PN Junctions: Details

In this appendix the boundary conditions at pn junctions, epitaxial diodes, minority-carrier transport in heavily doped emitter regions and the small-signal admittance of diodes at high frequencies are considered in more detail.

C.1 Boundary Conditions at pn Junctions

Under high-level-injection conditions, the approximations $n(x_\mathrm{n}) \approx n_{\mathrm{n}0} \approx N_\mathrm{D}(x_\mathrm{n})$ and $p(x_\mathrm{p}) \approx p_{\mathrm{p}0} \approx N_\mathrm{A}(x_\mathrm{p})$ are no longer fulfilled. In this case, the assumption of Boltzmann quasiequilibrium across the depletion layer

$$\frac{n_\mathrm{p}(x_\mathrm{p})}{n_\mathrm{n}(x_\mathrm{n})} = \frac{p_\mathrm{n}(x_\mathrm{n})}{p_\mathrm{p}(x_\mathrm{p})} = \exp\left(\frac{V' - V_\mathrm{J}}{V_\mathrm{T}}\right) ,$$

together with the neutrality conditions

$$n_\mathrm{p}(x_\mathrm{p}) + N_\mathrm{A}(x_\mathrm{p}) = p_\mathrm{p}(x_\mathrm{p}) \quad \text{and} \quad p_\mathrm{n}(x_\mathrm{n}) + N_\mathrm{D}(x_\mathrm{n}) = n_\mathrm{n}(x_\mathrm{n}) ,$$

provides a set of four equations that determines the four unknowns $n_\mathrm{p}(x_\mathrm{p})$, $n_\mathrm{n}(x_\mathrm{n})$, $p_\mathrm{p}(x_\mathrm{p})$ and $p_\mathrm{n}(x_\mathrm{n})$. The resulting formulas for the minority-carrier densities at the depletion layer edges are commonly referred to as the Fletcher boundary conditions:

$$n_\mathrm{p}(x_\mathrm{p}) = \frac{N_\mathrm{D}(x_\mathrm{n}) + N_\mathrm{A}(x_\mathrm{p}) \exp\left(\frac{V' - V_\mathrm{J}}{V_\mathrm{T}}\right)}{1 - \exp\left(\frac{2(V' - V_\mathrm{J})}{V_\mathrm{T}}\right)} \exp\left(\frac{V' - V_\mathrm{J}}{V_\mathrm{T}}\right) \tag{C.1}$$

$$p_\mathrm{n}(x_\mathrm{n}) = \frac{N_\mathrm{A}(x_\mathrm{p}) + N_\mathrm{D}(x_\mathrm{n}) \exp\left(\frac{V' - V_\mathrm{J}}{V_\mathrm{T}}\right)}{1 - \exp\left(\frac{2(V' - V_\mathrm{J})}{V_\mathrm{T}}\right)} \exp\left(\frac{V' - V_\mathrm{J}}{V_\mathrm{T}}\right) . \tag{C.2}$$

According to the generalized mass-action law, the pn product at the boundary of the space charge layer on the p side is

$$n_\mathrm{p}(x_\mathrm{p})\, p_\mathrm{p}(x_\mathrm{p}) = n_\mathrm{ie}^2(x_\mathrm{p}) \exp\left(\frac{\phi_\mathrm{p}(x_\mathrm{p}) - \phi_\mathrm{n}(x_\mathrm{p})}{V_\mathrm{T}}\right) . \tag{C.3}$$

Since the external voltage drop across the junction is $V' = \phi_\mathrm{p}(x_\mathrm{p}) - \phi_\mathrm{n}(x_\mathrm{n})$ and

$$\phi_\mathrm{n}(x_\mathrm{p}) = \phi_\mathrm{n}(x_\mathrm{n}) + \int_{x_\mathrm{n}}^{x_\mathrm{p}} \frac{\mathrm{d}\phi_\mathrm{n}}{\mathrm{d}x}\, \mathrm{d}x = \phi_\mathrm{n}(x_\mathrm{n}) - \int_{x_\mathrm{n}}^{x_\mathrm{p}} \frac{J_\mathrm{n}}{e\mu_\mathrm{n} n}\, \mathrm{d}x ,$$

(C.3) may be rewritten as

$$n_{\mathrm{p}}(x_{\mathrm{p}})\,p_{\mathrm{p}}(x_{\mathrm{p}}) \;=\; n_{\mathrm{ie}}^2(x_{\mathrm{p}})\,\exp\!\left(\frac{V'}{V_{\mathrm{T}}}\right)\exp\!\left(\frac{1}{V_{\mathrm{T}}}\int_{x_{\mathrm{n}}}^{x_{\mathrm{p}}}\frac{J_{\mathrm{n}}}{e\mu_{\mathrm{n}}n}\,\mathrm{d}x\right)\,. \tag{C.4}$$

This result reduces to the conventional boundary condition in the limit $J_{\mathrm{n}} \to 0$. Since under forward-bias conditions $J_{\mathrm{n}} < 0$ if the coordinate system is chosen such that $x_{\mathrm{n}} < x_{\mathrm{p}}$, the hole–electron product will always be smaller than $n_{\mathrm{ie}}^2 \exp(V'/V_{\mathrm{T}})$, as stated in [1]. Since under stationary conditions $\mathrm{d}J_{\mathrm{n}}/\mathrm{d}x = e(R-G)$, the electron current density $J_{\mathrm{n}}(x)$ is given by

$$J_{\mathrm{n}}(x) \;=\; J_{\mathrm{n}}(x_{\mathrm{p}}) + e\int_{x_{\mathrm{p}}}^{x}(R-G)\,\mathrm{d}x\,, \tag{C.5}$$

and the integral in (C.4) thus transforms to

$$\int_{x_{\mathrm{n}}}^{x_{\mathrm{p}}}\frac{J_{\mathrm{n}}}{e\mu_{\mathrm{n}}n}\,\mathrm{d}x \;=\; J_n(x_{\mathrm{p}})\int_{\mathrm{x_n}}^{x_{\mathrm{p}}}\frac{1}{e\mu_{\mathrm{n}}n}\,\mathrm{d}x + \int_{x_{\mathrm{n}}}^{x_{\mathrm{p}}}\frac{1}{\mu_{\mathrm{n}}n}\int_{x_{\mathrm{p}}}^{x}(R-G)\,\mathrm{d}x'\,\mathrm{d}x\,.$$

Since $1/(e\mu_{\mathrm{n}}n)$ is the conductivity of the conduction band electrons, the first term on the right-hand side can be considered as the change of the quasi-Fermi potential ϕ_{n} caused by the electron current that traverses the space charge layer. The second term on the right-hand side describes the change of the quasi-Fermi potential ϕ_{n} associated with the electron current that recombines in the space charge layer; this term may be neglected in comparison with the first term if most of the recombination takes place in the quasi-neutral p-type region. This will be assumed in the following, i.e. only the first term on the right-hand side will be given further consideration. To estimate the error associated with this term, an exponential variation of the electron density across the depletion layer can be assumed for simplicity:

$$n(x) \;=\; n(x_{\mathrm{n}})\exp\!\left(-\gamma\,\frac{x-x_{\mathrm{n}}}{d_{\mathrm{j}}}\right)\,,$$

where $\gamma = \ln[n(x_{\mathrm{n}})/n(x_{\mathrm{p}})]$ and $d_{\mathrm{j}} = x_{\mathrm{p}} - x_{\mathrm{n}}$. If μ_{n} is taken to be constant, the integral may be evaluated, with the result

$$\Delta\phi_{\mathrm{n}} \;=\; J_{\mathrm{n}}\int_{x_{\mathrm{n}}}^{x_{\mathrm{p}}}\frac{1}{e\mu_{\mathrm{n}}n}\,\mathrm{d}x \;\approx\; \frac{J_{\mathrm{n}}}{\gamma}\frac{d_{\mathrm{j}}}{e\mu_{\mathrm{n}}n(x_{\mathrm{p}})}$$

if $n(x_{\mathrm{p}}) \ll n(x_{\mathrm{n}})$. In a long-base diode $J_{\mathrm{n}} \approx -eD_{\mathrm{n}}n(x_{\mathrm{p}})/L_{\mathrm{n}}$, and therefore

$$\Delta\phi_{\mathrm{n}}/V_{\mathrm{T}} \;\approx\; -d_{\mathrm{j}}/(\gamma L_{\mathrm{n}})$$

as $D_{\mathrm{n}} = V_{\mathrm{T}}\mu_{\mathrm{n}}$. Since $\gamma > 1$ in general, Shockley's boundary condition will be fulfilled if the width of the depletion layer is small in comparison with the minority-carrier diffusion length [2, 3], a requirement that is generally fulfilled. As was pointed out in [4], however, in (metal-contacted) short-base diodes $J_{\mathrm{n}} \approx -eD_{\mathrm{n}}n(x_{\mathrm{p}})/d_{\mathrm{p}}$, where d_{p} is the thickness of the p-type region, and therefore

$$\Delta\phi_{\mathrm{n}}/V_{\mathrm{T}} \approx -d_{\mathrm{j}}/(\gamma d_{\mathrm{p}}) \; .$$

In this case d_{j} could be of the same order of magnitude as d_{p} and $\Delta\phi_{\mathrm{n}}$ would come into the range of V_{T}, resulting in substantial deviations from Shockley's boundary conditions.

C.2 Epitaxial Diode

A description of short epilayers that also applies under high-level-injection conditions is derived in the following. We neglect injection of electrons into the p-type region and recombination in the epitaxial layer, i.e. we take $J_{\mathrm{n}} \approx 0$ within the epilayer. Together with the assumption of neutrality, we therefore require

$$J_{\mathrm{n}} = e\mu_{\mathrm{n1}}(p_{\mathrm{n}}+N_{\mathrm{D1}})E + eD_{\mathrm{n}}\,\frac{\mathrm{d}p_{\mathrm{n}}}{\mathrm{d}x} = 0 \; ,$$

where μ_{n1} denotes the electron mobility and N_{D1} denotes the donor density in the epilayer. With the help of the Einstein relation $D_{\mathrm{n}} = V_{\mathrm{T}}\mu_{\mathrm{n}}$, this condition may be solved for the electric field strength in the epitaxial layer, resulting in

$$E = -\frac{V_{\mathrm{T}}}{p_{\mathrm{n}}+N_{\mathrm{D1}}}\,\frac{\mathrm{d}p_{\mathrm{n}}}{\mathrm{d}x} \; . \tag{C.6}$$

With this result, the expression for the hole density transforms to

$$J_{\mathrm{p}} = e\mu_{\mathrm{p1}}p_{\mathrm{n}}E - eD_{\mathrm{p1}}\frac{\mathrm{d}p_{\mathrm{n}}}{\mathrm{d}x} \;=\; -eD_{\mathrm{p1}}\frac{2p_{\mathrm{n}}+N_{\mathrm{D1}}}{p_{\mathrm{n}}+N_{\mathrm{D1}}}\frac{\mathrm{d}p_{\mathrm{n}}}{\mathrm{d}x} \; . \tag{C.7}$$

Since recombination is assumed to be negligible, $J_{\mathrm{p}} \approx$ const. throughout the epitaxial layer. After a separation of variables in (C.7), an integration from 0 to d therefore gives

$$\begin{aligned} J_{\mathrm{p}}d &= \int_0^d J_{\mathrm{p}}\,\mathrm{d}x = -eD_{\mathrm{p1}}\int_{p_{\mathrm{n}}(0)}^{p_{\mathrm{n}}(d)}\frac{2p_{\mathrm{n}}+N_{\mathrm{D1}}}{p_{\mathrm{n}}+N_{\mathrm{D1}}}\,\mathrm{d}p_{\mathrm{n}} \\ &= 2eD_{\mathrm{p1}}\left[p_{\mathrm{n}}(0) - p_{\mathrm{n}}(d) - \frac{N_{\mathrm{D1}}}{2}\ln\left(\frac{p_{\mathrm{n}}(0)+N_{\mathrm{D1}}}{p_{\mathrm{n}}(d)+N_{\mathrm{D1}}}\right)\right] \; . \end{aligned} \tag{C.8}$$

The value of $p_{\mathrm{n}}(0)$ is determined from the voltage drop V' across the p$^+$n junction with the help of the generalized mass-action law,

$$p_{\mathrm{n}}(0)\,[\,N_{\mathrm{D1}}+p_{\mathrm{n}}(0)\,] \;=\; n_{\mathrm{ie1}}^2\exp(V'/V_{\mathrm{T}}) \; ,$$

while the value of $p_{\mathrm{n}}(d)$ is related to the current density J_{p} by the surface recombination velocity $S_{\mathrm{nn^+}}$ at the nn$^+$ junction in accordance with (1.85). Solving these equations for $p_{\mathrm{n}}(0)$ and $p_{\mathrm{n}}(d)$ yields

$$p_{\mathrm{n}}(0) = \frac{N_{\mathrm{D1}}}{2}\left[\sqrt{1+\frac{4n_{\mathrm{ie1}}^2}{N_{\mathrm{D1}}^2}\exp\left(\frac{V'}{V_{\mathrm{T}}}\right)} \;-\; 1\right] \tag{C.9}$$

and

$$p_n(d) = \frac{N_{D1}}{2}\left[\sqrt{1+\frac{4}{N_{D1}}\left(\frac{J_p}{eS_{nn^+}}+p_{n0}\right)}\;-\;1\right]. \tag{C.10}$$

Combining (C.8), (C.9) and (C.10) results in the following implicit relation for the hole current density J_p:

$$J_p = \frac{eN_{D1}D_{p1}}{d_{epi}}\left[\sqrt{1+\frac{4n_{ie1}^2}{N_{D1}^2}\exp\left(\frac{V'}{V_T}\right)}-\sqrt{1+\frac{4}{N_{D1}}\left(\frac{J_p}{eS_{nn}}+p_{n01}\right)}\right]$$
$$-\frac{eN_{D1}D_{p1}}{d_{epi}}\ln\left(\frac{1+\sqrt{1+\frac{4n_{ie1}^2}{N_{D1}^2}\exp\left(\frac{V'}{V_T}\right)}}{1+\sqrt{1+\frac{4}{N_{D1}}\left(\frac{J_p}{eS_{nn^+}}+p_{n01}\right)}}\right). \tag{C.11}$$

An approximate solution of this equation is possible under low-level-injection conditions ($p_n(0) \ll N_D$), where

$$\sqrt{1+\frac{4n_{ie1}^2}{N_{D1}^2}\exp\left(\frac{V'}{V_T}\right)} \;\approx\; 1+\frac{2n_{ie1}^2}{N_{D1}^2}\exp\left(\frac{V'}{V_T}\right), \tag{C.12}$$

$$\sqrt{1+\frac{4}{N_{D1}}\left(\frac{J_p}{eS_{nn}}+p_{n01}\right)} \;\approx\; 1+\frac{2}{N_{D1}}\left(\frac{J_p}{eS_{nn}}+p_{n01}\right). \tag{C.13}$$

Using this result together with $p_{n01} = n_{ie1}^2/N_{D1}$ and the approximation $\ln(1+x) \approx x$, which is valid for $|x| \ll 1$, one therefore obtains

$$J_p \approx \frac{eD_{p1}p_{n01}}{d+D_{p1}/S_{nn^+}}\left[\exp\left(\frac{V'}{V_T}\right)-1\right] \tag{C.14}$$

in accordance with Sect. 1.4.6. Under high-level-injection conditions, the term determined by

$$p_n(0) \;\approx\; n_{ie1}\exp(V'/2V_T) \;\gg\; N_D \;\gg\; p_{n01}$$

will dominate in (C.8); if we take account of only this term, we obtain

$$I_p \;\approx\; \frac{2eA_jD_{p1}n_{ie}}{d}\exp\left(\frac{V'}{2V_T}\right).$$

In these calculations the hole diffusion coefficient D_{p1} has been assumed to be a constant, independent of the injection level. This assumption is not strictly valid, however, since carrier–carrier scattering will decrease the mobility under high-level-injection conditions.

C.3 Minority-Carrier Transport in Heavily Doped Emitter Regions

In the following, an approximate expression for the hole current I_{BE} injected into the emitter region that takes account of heavy-doping effects and position-dependent material parameters will be developed.[1] Owing to the heavy doping of the emitter region, the assumption $\phi_{\mathrm{n}} \approx$ const. yields a good approximation, and for ease of notation we set $\phi_{\mathrm{n}}(x) = 0$ throughout the emitter region. The coordinate system is chosen such that minority carriers are injected at $x = 0$ and reach the contact at $x = d$; the hole current injected into the emitter region is then $I_{\mathrm{BE}} = A_{\mathrm{je}}\, J_{\mathrm{p}}(0)$. To calculate $J_{\mathrm{p}}(0)$, we combine $J_{\mathrm{p}} = -e\mu_{\mathrm{p}} p(\mathrm{d}\phi_{\mathrm{p}}/\mathrm{d}x)$ with the law of mass action $pn = n_{\mathrm{ie}}^2 \exp(\phi_{\mathrm{p}}/V_{\mathrm{T}})$ to obtain

$$J_{\mathrm{p}}(x) \;=\; -k_{\mathrm{B}}T\mu_{\mathrm{p}}(x)\frac{n_{\mathrm{ie}}^2(x)}{n(x)}\frac{\mathrm{d}}{\mathrm{d}x}\exp\left(\frac{\phi_{\mathrm{p}}}{V_{\mathrm{T}}}\right) \,, \tag{C.15}$$

or equivalently

$$\frac{\mathrm{d}\Lambda_{\mathrm{p}}}{\mathrm{d}x} \;=\; -J_{\mathrm{p}}(x)\,\frac{\mathrm{d}G_{\mathrm{E}}}{\mathrm{d}x} \,, \tag{C.16}$$

introducing the abbreviations

$$\Lambda_{\mathrm{p}}(x) = \exp\left(\frac{\phi_{\mathrm{p}}(x)}{V_{\mathrm{T}}}\right) - 1 \quad \text{and} \quad G_{\mathrm{E}}(x) = \frac{1}{k_{\mathrm{B}}T}\int_0^x \frac{n(x')}{\mu_{\mathrm{p}}(x')\,n_{\mathrm{ie}}^2(x')}\,\mathrm{d}x' \,.$$

An integration by parts of (C.16) gives

$$\Lambda_{\mathrm{p}}(x) \;=\; \Lambda_{\mathrm{p}}(0) - J_{\mathrm{p}}(x)G_{\mathrm{E}}(x) + \int_0^x \frac{\mathrm{d}J_{\mathrm{p}}}{\mathrm{d}x'}\,G_{\mathrm{E}}(x')\,\mathrm{d}x' \,.$$

Together with the stationary continuity equation, which may be written as

$$\frac{\mathrm{d}J_{\mathrm{p}}}{\mathrm{d}x} = -\frac{e\,[p_{\mathrm{n}}(x) - p_{\mathrm{n0}}(x)]}{\tau_{\mathrm{p}}(x)} = -\frac{e n_{\mathrm{ie}}^2(x)}{n(x)\tau_{\mathrm{p}}(x)}\,\Lambda_{\mathrm{p}}(x) \,,$$

this is equivalent to

$$\Lambda_{\mathrm{p}}(x) = \Lambda_{\mathrm{p}}(0) - J_{\mathrm{p}}(x)G_{\mathrm{E}}(x) - \int_0^x \frac{e n_{\mathrm{ie}}^2(x')\Lambda_{\mathrm{p}}(x')G_{\mathrm{E}}(x')}{n(x')\tau_{\mathrm{p}}(x')}\,\mathrm{d}x' \,. \tag{C.17}$$

Since

$$J_{\mathrm{p}}(x) = J_{\mathrm{p}}(0) + \int_0^x \frac{\mathrm{d}J_{\mathrm{p}}}{\mathrm{d}x'}\,\mathrm{d}x' = J_{\mathrm{p}}(0) - \int_0^x \frac{e n_{\mathrm{ie}}^2(x')\Lambda_{\mathrm{p}}(x')}{n(x')\tau_{\mathrm{p}}(x')}\,\mathrm{d}x' \,, \tag{C.18}$$

(C.17) may be rewritten in the form

[1] This approach is related to the derivation given in [5]; alternative formulations that yield basically the same result may be found in [6], for example.

$$\Lambda_{\mathrm{p}}(x) = \Lambda_{\mathrm{p}}(0) - J_{\mathrm{p}}(0)G_{\mathrm{E}}(x) + \int_0^x \Gamma(x,x')\Lambda_{\mathrm{p}}(x')\,\mathrm{d}x' \,, \tag{C.19}$$

where

$$\Gamma(x,x') \;=\; [\,G_{\mathrm{E}}(x) - G_{\mathrm{E}}(x')\,]\frac{e n_{\mathrm{ie}}^2(x')}{n(x')\tau_{\mathrm{p}}(x')} \,. \tag{C.20}$$

Equation (C.19) is an integral equation for $\Lambda_{\mathrm{p}}(x)$, which may be solved iteratively according to the scheme

$$\Lambda_{\mathrm{p}}^{(\alpha+1)}(x) = \Lambda_{\mathrm{p}}(0) - J_{\mathrm{p}}(0)G_{\mathrm{E}}(x) + \int_0^x \Gamma(x,x')\Lambda_{\mathrm{p}}^{(\alpha)}(x')\,\mathrm{d}x' \,.$$

In the case of negligible recombination ($\tau_{\mathrm{p}} \to \infty$) one obtains, in particular

$$\Lambda_{\mathrm{p}}^{(0)}(x) \;=\; \Lambda_{\mathrm{p}}(0) - J_{\mathrm{p}}(0)G_{\mathrm{E}}(x) \,. \tag{C.21}$$

With this taken as a first guess, the next step of the iteration yields

$$\begin{aligned}\Lambda_{\mathrm{p}}^{(1)}(x) \;&=\; \Lambda_{\mathrm{p}}(0)\left(1 + \int_0^x \Gamma(x,x')\,\mathrm{d}x'\right)\\ &\quad - J_{\mathrm{p}}(0)\left(G_{\mathrm{E}}(x) + \int_0^x \Gamma(x,x')G_{\mathrm{E}}(x')\,\mathrm{d}x'\right)\end{aligned}$$

and, by induction,

$$\Lambda_{\mathrm{p}}(x) \;=\; \Lambda_{\mathrm{p}}(0)\sum_{n=0}^{\infty}\zeta_n(x) - J_{\mathrm{p}}(0)\sum_{m=0}^{\infty}\theta_m(x) \,,$$

where $\zeta_n(x)$ and $\theta_m(x)$ denote recursively defined functions:

$$\begin{aligned}\zeta_{n+1}(x) \;&=\; \int_0^x \Gamma(x,x')\,\zeta_n(x')\,\mathrm{d}x' \\ \theta_{m+1}(x) \;&=\; \int_0^x \Gamma(x,x')\,\theta_m(x')\,\mathrm{d}x' \,.\end{aligned} \tag{C.22}$$

and $\zeta_0(x) = 1$ and $\theta_0(x) = G_{\mathrm{E}}(x)$. Introducing (C.20) into (C.22) and performing an integration by parts yields

$$\zeta_{n+1}(x) = \frac{1}{V_{\mathrm{T}}}\int_0^x\int_0^{x'} \frac{n_{\mathrm{ie}}^2(x'')}{n_{\mathrm{ic}}^2(x')}\frac{n(x')}{n(x'')}\frac{\zeta_n(x'')}{\mu_{\mathrm{p}}(x')\tau_{\mathrm{p}}(x'')}\,\mathrm{d}x''\,\mathrm{d}x' \,,$$

together with an analogous expression for $\theta_{m+1}(x)$. At the emitter contact one obtains, in particular,

$$\Lambda_{\mathrm{p}}(d) = \Lambda_{\mathrm{p}}(0)\sum_{n=0}^{\infty}\zeta_n(d) - J_{\mathrm{p}}(0)\sum_{m=0}^{\infty}\theta_m(d) \,. \tag{C.23}$$

Using this together with (C.18), one obtains the following for the hole current density at the emitter contact ($x = d$):

$$\begin{aligned} J_\mathrm{p}(d) &= J_\mathrm{p}(0) + J_\mathrm{p}(0) \sum_{m=0}^{\infty} \int_0^d \frac{e n_\mathrm{ie}^2(x)}{n(x)\tau_\mathrm{p}(x)}\, \theta_m(x)\,\mathrm{d}x \\ &\qquad - \Lambda_\mathrm{p}(0) \sum_{n=0}^{\infty} \int_0^d \frac{e n_\mathrm{ie}^2(x)}{n(x)\tau_\mathrm{p}(x)}\, \zeta_n(x)\,\mathrm{d}x \,. \end{aligned} \tag{C.24}$$

If the excess hole density at the emitter contact is determined by S_p, the boundary condition

$$\begin{aligned} J_\mathrm{p}(d) &= e S_\mathrm{p}\,[\,p_\mathrm{n}(d) - p_\mathrm{n0}(d)\,] = \frac{e S_\mathrm{p} n_\mathrm{ie}^2(d)}{n(d)} \left[\exp\left(\frac{\phi_\mathrm{p}(x_\mathrm{e})}{V_\mathrm{T}} \right) - 1 \right] \\ &= \frac{e S_\mathrm{p} n_\mathrm{ie}^2(d)}{n(d)}\, \Lambda_\mathrm{p}(d) \end{aligned} \tag{C.25}$$

must be fulfilled.

Combining (C.24), (C.25) and (C.23) and using $\Lambda_\mathrm{p}(0) = \exp(V_\mathrm{BE}/V_\mathrm{T}) - 1$ yields the hole current $I_\mathrm{BE} = A_\mathrm{je} J_\mathrm{p}(0) = I_\mathrm{S}\,[\,\exp(V_\mathrm{BE}/V_\mathrm{T}) - 1\,]$ injected into the emitter region; the saturation current may be expressed as

$$I_\mathrm{S} = A_\mathrm{je} \frac{1 + \chi_1}{G_\mathrm{E}(d) + \dfrac{n(d)}{e n_\mathrm{ie}^2(d)\, S_\mathrm{p}} + \chi_2} \,, \tag{C.26}$$

where

$$\chi_1 = \sum_{n=1}^{\infty} \zeta_n(d) + \frac{1}{S_\mathrm{p}} \frac{n(d)}{n_\mathrm{ie}^2(d)} \sum_{n=0}^{\infty} \int_0^d \frac{n_\mathrm{ie}^2(x) \zeta_n(x)}{n(x)\tau_\mathrm{p}(x)}\,\mathrm{d}x \,, \tag{C.27}$$

$$\chi_2 = \sum_{m=1}^{\infty} \theta_m(d) + \frac{1}{S_\mathrm{p}} \frac{n(d)}{n_\mathrm{ie}^2(d)} \sum_{m=0}^{\infty} \int_0^d \frac{n_\mathrm{ie}^2(x) \theta_m(x)}{n(x)\tau_\mathrm{p}(x)}\,\mathrm{d}x \tag{C.28}$$

are correction terms which vanish in the limit $\tau_\mathrm{p} \to \infty$. Considering only terms which are of first order in $1/\tau_\mathrm{p}$, the correction terms read

$$\begin{aligned} \chi_1 &= \int_0^d \int_0^x \frac{n_\mathrm{ie}^2(x')}{n_\mathrm{ie}^2(x)} \frac{n(x)}{n(x')} \frac{1}{D_\mathrm{p}(x)\tau_\mathrm{p}(x')}\,\mathrm{d}x'\,\mathrm{d}x \\ &\quad + \frac{1}{S_\mathrm{p}} \frac{n(d)}{n_\mathrm{ie}^2(d)} \int_0^d \frac{n_\mathrm{ie}^2(x)}{n(x)} \frac{1}{\tau_\mathrm{p}(x)}\,\mathrm{d}x \end{aligned}$$

and

$$\begin{aligned} \chi_2 &= \int_0^d \int_0^x \frac{n_\mathrm{ie}^2(x')}{n_\mathrm{ie}^2(x)} \frac{n(x)}{n(x')} \frac{G_\mathrm{E}(x')}{D_\mathrm{p}(x)\tau_\mathrm{p}(x')}\,\mathrm{d}x'\,\mathrm{d}x \\ &\quad + \frac{1}{e S_\mathrm{p}} \frac{n(d)}{n_\mathrm{ie}^2(d)} \int_0^d \int_0^x \frac{n_\mathrm{ie}^2(x)}{n_\mathrm{ie}^2(x')} \frac{n(x')}{n(x)} \frac{1}{D_\mathrm{p}(x)\tau_\mathrm{p}(x')}\,\mathrm{d}x'\,\mathrm{d}x \,. \end{aligned}$$

C.4 High-Frequency Diode Admittance

In order to obtain accurate results for the nonstationary behavior of a forward-biased pn junction, the time-dependent diffusion equations must be solved. For that purpose we consider the density $p_{\mathrm{n}}(x,t)$ of holes injected into an n-type region of thickness d, under small-signal conditions; in this case the hole density in the n-type region may be expressed as

$$p_{\mathrm{n}}(x,t) \;=\; p_{\mathrm{n}}(x) + \mathrm{Re}\left[\underline{\hat{p}}_{\mathrm{n1}}(x)\,\mathrm{e}^{\mathrm{j}\omega t}\right] \;.$$

In this expression $p_{\mathrm{n}}(x)$ denotes the hole density in the n-type region in dc operation with a bias voltage V. The small-signal portion has to obey

$$\frac{\mathrm{d}^2\underline{\hat{p}}_{\mathrm{n1}}}{\mathrm{d}x^2} - \frac{\underline{\hat{p}}_{\mathrm{n1}}(x)}{\underline{L}_{\mathrm{p}}^2} \;=\; 0\;, \qquad \text{where} \quad \underline{L}_{\mathrm{p}} \;=\; \frac{L_{\mathrm{p}}}{\sqrt{1+\mathrm{j}\omega\tau_{\mathrm{p}}}}\;, \tag{C.29}$$

which is analogous to the stationary diffusion equation, except for the complex diffusion length. The boundary condition for $\underline{\hat{p}}_{\mathrm{n1}}(x)$ at the space charge layer boundary ($x = 0$) under low-level-injection conditions is obtained by developing the Shockley boundary conditions

$$p_{\mathrm{n}}(0,t) \;=\; p_{\mathrm{n0}}\,\exp\!\left(\frac{V+\mathrm{Re}(\underline{v})}{V_{\mathrm{T}}}\right) \;\approx\; p_{\mathrm{n0}}\,\exp\!\left(\frac{V}{V_{\mathrm{T}}}\right)\left(1+\frac{\mathrm{Re}(\underline{v})}{V_{\mathrm{T}}}\right)\;,$$

resulting in

$$\underline{\hat{p}}_{\mathrm{n1}}(0) \;=\; p_{\mathrm{n}}(0)\,\underline{\hat{v}}/V_{\mathrm{T}}\;, \tag{C.30}$$

where $p_{\mathrm{n}}(0)$ denotes the hole density at the depletion layer boundary at the bias point. As the second boundary condition, we require

$$-eD_{\mathrm{p}}\left.\frac{\mathrm{d}\underline{\hat{p}}_{\mathrm{n1}}}{\mathrm{d}x}\right|_{d} \;=\; eS_{\mathrm{p}}\underline{\hat{p}}_{\mathrm{n1}}(d)\;, \tag{C.31}$$

where the surface recombination velocity S_{p} at the contact has been assumed to be a frequency-independent constant. The solution of (C.29) that obeys the boundary conditions (C.30) and (C.31) allows us to compute the small-signal hole current that is injected into the n-type region

$$\underline{\hat{i}} \;=\; -eA_{\mathrm{j}}D_{\mathrm{p}}\left.\frac{\mathrm{d}\underline{\hat{p}}_{\mathrm{n1}}}{\mathrm{d}x}\right|_{0} \;=\; \underline{y}(\omega)\,\underline{\hat{v}}\;,$$

where the admittance is

$$\underline{y}(\omega) \;=\; \frac{eA_{\mathrm{j}}D_{\mathrm{p}}p_{\mathrm{n}}(0)}{V_{\mathrm{T}}}\,\frac{1}{\underline{L}_{\mathrm{p}}}\,\frac{\cosh(d/\underline{L}_{\mathrm{p}}) + \underline{\nu}_{\mathrm{p}}\sinh(d/\underline{L}_{\mathrm{p}})}{\sinh(d/\underline{L}_{\mathrm{p}}) + \underline{\nu}_{\mathrm{p}}\cosh(d/\underline{L}_{\mathrm{p}})}\;, \tag{C.32}$$

and $\underline{\nu}_{\mathrm{p}} = D_{\mathrm{p}}/(S_{\mathrm{p}}\underline{L}_{\mathrm{p}})$. Equation (C.32) does not take account of the depletion layer capacitance and series resistance effects, and therefore corresponds to $g_\mathrm{d}(1+\mathrm{j}\omega\tau_\mathrm{t})$ in the quasi-static model. If (C.32) is developed up to first order in ω, one obtains an improved description of the small-signal diffusion capacitance. The dc portion,

$$\begin{aligned} y(0) &= \frac{eA_\mathrm{j}D_\mathrm{p}p_\mathrm{n}(0)}{V_\mathrm{T}}\frac{1}{L_\mathrm{p}}\frac{\cosh(d/L_\mathrm{p})+\nu_\mathrm{p}\sinh(d/L_\mathrm{p})}{\sinh(d/L_\mathrm{p})+\nu_\mathrm{p}\cosh(d/L_\mathrm{p})} \\ &= g_\mathrm{d} = (I+I_\mathrm{S})/V_\mathrm{T}\,, \end{aligned} \tag{C.33}$$

equals the small-signal conductance of the diode, which can also be obtained from a small-signal analysis of the current–voltage characteristic. The term of first order in ω is given by

$$\omega\left.\frac{\mathrm{d}y}{\mathrm{d}\omega}\right|_0 = \mathrm{j}\omega\frac{g_\mathrm{d}\tau_\mathrm{p}}{2}\left[1+F\left(\frac{d}{L_\mathrm{p}}\right)\right], \tag{C.34}$$

where

$$\begin{aligned} F(\xi) &= \frac{(\nu_\mathrm{p}^2-1)\,\xi-\nu_\mathrm{p}}{[\sinh(\xi)+\nu_\mathrm{p}\cosh(\xi)]\,[\cosh(\xi)+\nu_\mathrm{p}\sinh(\xi)]} \\ &= \frac{2(\nu_\mathrm{p}^2-1)\xi-\nu_\mathrm{p}}{2\nu_\mathrm{p}\cosh(2\xi)+(\nu_\mathrm{p}^2+1)\sinh(2\xi)}\,. \end{aligned} \tag{C.35}$$

Since $\mathrm{d}y/\mathrm{d}\omega|_0$ is purely imaginary, we may define the non-quasi-static diffusion capacitance

$$\tilde{c}_\mathrm{t} = -\mathrm{j}\left.\frac{\mathrm{d}y}{\mathrm{d}\omega}\right|_0 = \frac{g_\mathrm{d}\tau_\mathrm{p}}{2}\left[1+F\left(\frac{d}{L_\mathrm{p}}\right)\right],$$

which results in a non-quasi-static definition of the transit time,

$$\tau_\mathrm{t} = \frac{\tau_\mathrm{p}}{2}\left[1+F\left(\frac{d}{L_\mathrm{p}}\right)\right].$$

In a long-base diode $d/L_\mathrm{p}\to\infty$ and $F(d/L_\mathrm{p})\to 0$, so that in this case

$$\tilde{c}_\mathrm{t} = g_\mathrm{d}\tau_\mathrm{p}/2 \quad \text{and} \quad \tau_\mathrm{t} = \tau_\mathrm{p}/2\,.$$

The small-signal solution of the diffusion equation therefore yields a diffusion capacitance that differs from the value obtained with the quasi-static assumption by a factor 1/2 in the case of a long-base diode (see also [7]). In a diode with a metal contact ($S_\mathrm{p}\to\infty$ and therefore $\nu_\mathrm{p}\to 0$), the function $F(d/L_\mathrm{p})$ simplifies to

$$F\left(\frac{d}{L_\mathrm{p}}\right) = -\frac{d}{L_\mathrm{p}}\frac{1}{\sinh(d/L_\mathrm{p})\cosh(d/L_\mathrm{p})} = -\frac{2d}{L_\mathrm{p}}\frac{1}{\sinh(2d/L_\mathrm{p})}\,.$$

If, in addition, $d \ll L_\mathrm{p}$, the substitution $\sinh(x)\approx x+x^3/6$ and $\cosh(x)\approx 1-x^2/2$ results in

$$1+F\left(\frac{d}{L_{\mathrm{p}}}\right) \approx \frac{2}{3}\frac{d^2}{L_{\mathrm{p}}^2} = \frac{2}{3}\frac{d^2}{D_{\mathrm{p}}\tau_{\mathrm{p}}}$$

and therefore

$$\tilde{c}_{\mathrm{t}} \approx \frac{2}{3}g_{\mathrm{d}}\frac{d^2}{2D_{\mathrm{p}}} \quad \text{and} \quad \tau_{\mathrm{t}} = \frac{2}{3}\frac{d^2}{2D_{\mathrm{p}}}\,,$$

which is two-thirds of the quasi-static value. For a short-base diode with a contact of finite surface recombination velocity, the function $1-F(d/L_{\mathrm{p}})$ may be developed into a power series in d/L_{p} to obtain the following for the transit time $\tau_{\mathrm{t}} = \tilde{c}_{\mathrm{t}}/g_{\mathrm{d}}$:

$$\tau_{\mathrm{t}} \approx \tau_{\mathrm{p}}\frac{d}{L_{\mathrm{p}}}\frac{\nu_{\mathrm{p}}(\nu_{\mathrm{p}}+d/L_{\mathrm{p}})}{\nu_{\mathrm{p}}+(1+\nu_{\mathrm{p}}^2)d/L_{\mathrm{p}}}\,, \tag{C.36}$$

to second order in d/L_{p}. If only first-order terms are retained, the transit time reads

$$\tau_{\mathrm{t}} \approx \tau_{\mathrm{p}}\nu_{\mathrm{p}}d/L_{\mathrm{p}} = d/S_{\mathrm{p}}\,. \tag{C.37}$$

C.5 References

[1] H.K. Gummel. Hole–electron product of pn junctions. *Solid-State Electron.*, 10:209–212, 1967.

[2] J.L. Moll. *Physics of Semiconductors.* McGraw-Hill, New York, 1964.

[3] J.R. Hauser. Boundary conditions at pn junctions. *Solid-State Electron.*, 14:133–139, 1970.

[4] F. van de Wiele. Modified Boltzmann boundary conditions in junction theory. *Solid-State Electron.*, 41:1699–1706, 1997.

[5] E. de Castro, M. Rudan. Integral-equation solution of minority-carrier transport problems in heavily doped semiconductors. *IEEE Trans. Electron Devices*, 31(6):785–792, 1984.

[6] A. Neugroschel, J.-S. Park, F.A. Lindholm. Systematic analytical solutions for minority-carrier transport in semiconductors with position-dependent composition, with application to heavily doped silicon. *IEEE Trans. Electron Devices*, 33(2):240–249, 1986.

[7] A. Arendt, M. Illi. Die exakte Berechnung der Diffusionskapazität von Halbleiterdioden aus der stationären Ladungsverteilung – Auflösung eines Widerspruchs. *Archiv Elektronik Übertragungstechnik*, 22(12):669–674, 1967.

D Bipolar Transistor: Details

This appendix presents details concerning minority-carrier transport through the base region of a drift transistor, the computation of the base resistance, the generation of model parameter from layout data, Gummel's integral charge control relation and carrier multiplication.

D.1 Drift Transistor

D.1.1 Electron Transport Through the Base Region

For constant mobility and electric field strength in the base region, the current and continuity equations for electrons in the base region read

$$J_{\mathrm{n}} = e\mu_{\mathrm{n}} n_{\mathrm{p}} E + eD_{\mathrm{n}} \frac{\partial n_{\mathrm{p}}}{\partial x} \quad \text{and} \quad \frac{\partial n_{\mathrm{p}}}{\partial t} = \frac{1}{e}\frac{\partial J_{\mathrm{n}}}{\partial x} - \frac{n_{\mathrm{p}} - n_{\mathrm{p0}}}{\tau_{\mathrm{n}}} .$$

With the help of the Einstein relation $D_{\mathrm{n}} = V_{\mathrm{T}}\mu_{\mathrm{n}}$, these equations may be combined into a differential equation for the excess electron density $\Delta n_{\mathrm{p}} = n_{\mathrm{p}} - n_{\mathrm{p0}}$,

$$\frac{\partial}{\partial t}\Delta n_{\mathrm{p}} = -\frac{\Delta n_{\mathrm{p}}}{\tau_{\mathrm{n}}} + \mu_{\mathrm{n}} E \frac{\partial}{\partial x}\Delta n_{\mathrm{p}} + \mu_{\mathrm{n}} V_{\mathrm{T}} \frac{\partial^2}{\partial x^2}\Delta n_{\mathrm{p}} , \tag{D.1}$$

if n_{p0} is assumed to be constant. In this equation an electric field strength $E > 0$ describes a retarding field, while $E < 0$ corresponds to an accelerating field. Laplace transformation reduces the partial differential equation (D.1) to an ordinary differential equation with respect to the space coordinate,

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \frac{E}{V_{\mathrm{T}}}\frac{\mathrm{d}}{\mathrm{d}x} - \frac{1+s\tau_{\mathrm{n}}}{L_{\mathrm{n}}^2}\right)\Delta \underline{n}_{\mathrm{p}}(x,s) = -\frac{\Delta n_{\mathrm{p}}(x,t{=}0)}{\mu_{\mathrm{n}} V_{\mathrm{T}}} . \tag{D.2}$$

Assuming $\Delta n_{\mathrm{p}}(x, t = 0) = 0$, the general solution of (D.2) is

$$\Delta \underline{n}_{\mathrm{p}}(x,s) = \Delta n_1(s)\exp\left(\frac{\alpha_1 x}{d_{\mathrm{B}}}\right) + \Delta n_2(s)\exp\left(\frac{\alpha_2 x}{d_{\mathrm{B}}}\right) , \tag{D.3}$$

where the coefficients α_1 and α_2 are determined by the corresponding characteristic equation. The solution of the characteristic equation gives

$$\alpha_1 = \frac{\eta}{2} + \underline{\vartheta}(s) \quad \text{and} \quad \alpha_2 = \frac{\eta}{2} - \underline{\vartheta}(s) ,$$

where $\underline{\vartheta}(s)$ is a dimensionless parameter given by

$$\underline{\vartheta}(s) = \sqrt{(\eta/2)^2 + (1+s\tau_{\mathrm{n}})d_{\mathrm{B}}^2/L_{\mathrm{n}}^2} .$$

Rewritten in terms of the homogeneous base transit time $\tau_{B0} = d_B^2/2D_n$ and the parameter $\delta = d_B/L_n$, the result for $\underline{\vartheta}$ reads

$$\underline{\vartheta} = \sqrt{(\eta/2)^2 + 2s\tau_{B0} + \delta^2} \,. \tag{D.4}$$

If the coefficients $\Delta n_1(s)$ and $\Delta n_2(s)$ are expressed in terms of the excess electron densities $\Delta\underline{n}_p(0,s)$ and $\Delta\underline{n}_p(d_B,s)$ at the depletion layer boundaries, (D.3) reads

$$\begin{aligned}\Delta\underline{n}_p(x,s) &= \exp\left(\frac{\eta\lambda}{2}\right)\frac{\sinh[(1-\lambda)\underline{\vartheta}]}{\sinh(\underline{\vartheta})}\,\Delta\underline{n}_p(0,s)\\ &+ \exp\left(\frac{\eta(\lambda-1)}{2}\right)\frac{\sinh(\lambda\underline{\vartheta})}{\sinh(\underline{\vartheta})}\,\Delta\underline{n}_p(d_B,s)\,,\end{aligned}$$

where $\lambda = x/d_B$. This allows us to calculate the electron current density

$$\underline{J}_n(x,s) = e\mu_n E\Delta\underline{n}_p(x,s) + eD_n\frac{\mathrm{d}}{\mathrm{d}x}\Delta\underline{n}_p(x,s)$$

and, in particular, the electron current $\underline{i}_{En}(s) = -A_{je}\underline{J}_n(0,s)$ that is injected into the base region, and the electron current $\underline{i}_{Cn}(s) = -A_{je}\underline{J}_n(d_B,s)$ that reaches the bc depletion layer. These currents may be expressed as

$$\underline{i}_{En}(s) = A_{11}(s)\frac{\Delta\underline{n}_p(0,s)}{n_{p0}(0)} - A_{12}(s)\frac{\Delta\underline{n}_p(d_B,s)}{n_{p0}(d_B)}\,, \tag{D.5}$$

$$\underline{i}_{Cn}(s) = A_{21}(s)\frac{\Delta\underline{n}_p(0,s)}{n_{p0}(0)} - A_{22}(s)\frac{\Delta\underline{n}_p(d_B,s)}{n_{p0}(d_B)}\,, \tag{D.6}$$

where the coefficients $A_{ij}(s)$ are given by

$$A_{11}(s) = \frac{eD_n A_{je} n_{p0}(0)}{d_B}\left[\underline{\vartheta}\coth(\underline{\vartheta}) + \eta/2\right]\,, \tag{D.7}$$

$$A_{12}(s) = \frac{eD_n A_{je} n_{p0}(d_B)}{d_B}\,\mathrm{e}^{-\eta/2}\frac{\underline{\vartheta}}{\sinh(\underline{\vartheta})}\,, \tag{D.8}$$

$$A_{21}(s) = \frac{eD_n A_{je} n_{p0}(0)}{d_B}\,\mathrm{e}^{\eta/2}\frac{\underline{\vartheta}}{\sinh(\underline{\vartheta})}\,, \tag{D.9}$$

$$A_{22}(s) = \frac{eD_n A_{je} n_{p0}(d_B)}{d_B}\left[\underline{\vartheta}\coth(\underline{\vartheta}) - \eta/2\right]\,. \tag{D.10}$$

Since $n_{p0}(0)N_{AB}(0) = n_{p0}(d_B)N_{AB}(d_B)$, or $n_{p0}(0)\mathrm{e}^{\eta/2} = n_{p0}(d_B)\mathrm{e}^{-\eta/2}$ owing to the law of mass action, the reciprocity relation

$$A_{21}(s) = A_{12}(s) \tag{D.11}$$

is valid.

D.1.2 Computation of $a_{21}(t)$

Transformation of $\underline{a}_{21}(s)$ to the time domain is possible if $\sinh(\underline{\vartheta})$ is represented as a geometric series according to

$$\frac{1}{\sinh(\underline{\vartheta})} = \frac{2\,\exp(-\underline{\vartheta})}{1-\exp(-2\underline{\vartheta})} = 2\sum_{n=0}^{\infty}\exp[-(2n+1)\underline{\vartheta}]\ .$$

This results in

$$\begin{aligned}\underline{a}_{21}(s) &= 2\,\frac{\sinh[\vartheta(0)]}{\vartheta(0)}\sum_{n=0}^{\infty}\underline{\vartheta}(s)\exp[-2(n+1)\underline{\vartheta}(s)]\\ &= \frac{2\sqrt{\tau_{\mathrm{B0}}}\,\sinh[\vartheta(0)]}{\vartheta(0)}\sum_{n=0}^{\infty}\sqrt{s+s_\vartheta}\,\exp(-\alpha_n\sqrt{s+s_\vartheta})\ ,\end{aligned}$$

where $s_\vartheta = \vartheta^2(0)/2\tau_{\mathrm{B0}}$ and $\alpha_n = (2n+1)\sqrt{2\tau_{\mathrm{B0}}}$. Since

$$\begin{aligned}&\mathcal{L}^{-1}[\sqrt{s+s_\vartheta}\,\exp(-\alpha_n\sqrt{s+s_\vartheta})]\\ &\quad= \exp(-s_\vartheta t)\,\mathcal{L}^{-1}[\sqrt{s}\,\exp(-\alpha_n\sqrt{s})]\\ &\quad= -\exp(-s_\vartheta t)\,\frac{\mathrm{d}}{\mathrm{d}\alpha_n}\,\mathcal{L}^{-1}[\exp(-\alpha_n\sqrt{s})]\\ &\quad= \frac{1}{2\sqrt{\pi}\,t^{3/2}}\left(\frac{\alpha_n^2}{2t}-1\right)\exp\left(-\frac{\alpha_n^2}{4t}-s_\vartheta t\right)\ ,\end{aligned}$$

the response function $a_{21}(t)$ may be represented by an infinite series:

$$\begin{aligned}a_{21}(t) &= \frac{1}{\tau_{\mathrm{B0}}}\sqrt{\frac{2}{\pi}}\,\frac{\sinh[\vartheta(0)]}{\vartheta(0)}\,\tilde{t}^{-3/2}\exp\left(-\frac{\vartheta^2(0)\tilde{t}}{2}\right)\\ &\qquad\times\sum_{n=0}^{\infty}\left(\frac{(2n+1)^2}{\tilde{t}}-1\right)\exp\left(-\frac{(2n+1)^2}{2\,\tilde{t}}\right)\ ,\end{aligned} \tag{D.12}$$

where $\vartheta^2(0) = (\eta/2)^2+\delta^2$ and $\tilde{t} = t/\tau_{\mathrm{B0}}$.

D.1.3 Excess Phase

In Sect. 3.1, the transfer current i_{CE} was shown to be delayed with respect to the stored minority charge q_{TBE}. That there is an excess phase shift between the transfer current $\underline{i}_{\mathrm{CE}}(s)$ and the base current $\underline{i}_{\mathrm{BB}}(s)$ that is required to neutralize the changing minority charge in the base layer can also be seen from an approximate analysis of the base transport factor. With the approximation

$$\begin{aligned}\cosh(\underline{\vartheta}) + \frac{\eta \sinh(\underline{\vartheta})}{2\underline{\vartheta}} &= \frac{\mathrm{e}^{\underline{\vartheta}}}{2}\left[1+\frac{\eta}{2\underline{\vartheta}}+\left(1-\frac{\eta}{2\underline{\vartheta}}\right)\mathrm{e}^{-\underline{\vartheta}}\right] \\ &\approx \frac{\mathrm{e}^{\underline{\vartheta}}}{2}\left(1+\frac{\eta}{2\underline{\vartheta}}\right),\end{aligned}$$

the frequency-dependent base transport factor (3.15) can be approximated by

$$\underline{\alpha}_{\mathrm{t}}(s) \approx 2\,\frac{\exp[-\underline{\vartheta}(s)+\eta/2]}{1+\eta/2\underline{\vartheta}(s)} \tag{D.13}$$

in the limit $|\vartheta| \gg 1$, corresponding to $\eta \gg 1$, with an accuracy for both amplitude and phase better than 1% if $\eta > 3$, as pointed out by te Winkel [1,2]. If $\underline{\vartheta}$ is expressed in terms of the quasi-static forward transit time τ_{Bf}, we may write

$$\underline{\vartheta} = \sqrt{\left(\frac{\eta}{2}\right)^2 + 2s\tau_{\mathrm{B0}}} = \frac{\eta}{2}\sqrt{1+\frac{4s\tau_{\mathrm{Bf}}}{\eta-1+\exp(-\eta)}}\,.$$

In the case of a substantial drift field in the base region ($\eta \gg 1$), the term $\exp(-\eta)$ in the denominator can be neglected, and we obtain

$$\underline{\vartheta}(s) \approx \frac{\eta}{2}\left(1+\frac{2s\tau_{\mathrm{Bf}}}{\eta-1}\right) \quad \text{and} \quad \frac{1}{\underline{\vartheta}(s)} \approx \frac{2}{\eta}\left(1-\frac{2s\tau_{\mathrm{Bf}}}{\eta-1}\right)$$

up to first order in s; with this result, (D.13) transforms to

$$\underline{\alpha}_{\mathrm{t}}(s) \approx \frac{\exp[-\eta s\tau_{\mathrm{Bf}}/(\eta-1)]}{1-s\tau_{\mathrm{Bf}}/(\eta-1)}\,, \tag{D.14}$$

i.e. an expression with an excess phase shift factor of the form proposed earlier by Thomas and Moll [3]. When this relation is substituted into (3.24), the following approximation is obtained:

$$\begin{aligned}\underline{i}_{\mathrm{BB}}(s) &\approx \left[1-\frac{s\tau_{\mathrm{Bf}}}{\eta-1}-\exp\left(-\frac{\eta s\tau_{\mathrm{Bf}}}{\eta-1}\right)\right]\exp\left(\frac{\eta s\tau_{\mathrm{Bf}}}{\eta-1}\right)\underline{i}_{\mathrm{CE}}(s) \\ &\approx s\tau_{\mathrm{Bf}}\exp\left(\frac{\eta s\tau_{\mathrm{Bf}}}{\eta-1}\right)\underline{i}_{\mathrm{CE}}(s)\,,\end{aligned}$$

in the limit $|s\tau_{\mathrm{Bf}}| \ll 1$.

D.1.4 Collector Transit Time

The collector transit time describes the delay associated with electron transport through the bc depletion layer. Owing to the finite electron drift velocity v_{nsat}, the time required by an electron to travel through a depletion layer of thickness d_{jc} is $2\tau_{\mathrm{jc}} = d_{\mathrm{jc}}/v_{\mathrm{nsat}}$. This value, however, is not the collector transit time, since the electrons traversing the depletion layer push electrons of the epilayer towards the collector contact by means of electrostatic induction. Assume a sinusoidal current described by the phasor $\underline{i}_{\mathrm{n}}(x_{\mathrm{bc}}) = \sqrt{2}\,\underline{I}_{\mathrm{n}}\mathrm{e}^{\mathrm{j}\omega t}$ to be injected at the base-side edge of the bc depletion layer. Since each electron carries a current $-ev_{\mathrm{nsat}}/d_{\mathrm{jc}}$ during a time interval of length $2\tau_{\mathrm{jc}}$, the current at the collector-side edge of the bc depletion layer can be obtained from [4]

$$\underline{i}_{\mathrm{n}}(x_{\mathrm{cb}}) = \sqrt{2}\,\underline{I}_{\mathrm{n}} \int_0^t f(t-t')\mathrm{e}^{\mathrm{j}\omega t'}\,\mathrm{d}t'$$

where

$$f(t) = \begin{cases} v_{\mathrm{nsat}}/d_{\mathrm{jc}} & \text{for} \quad 0 \le t \le 2\tau_{\mathrm{jc}} \\ 0 & \text{for} \quad t < 0 \ \text{and} \ t > 2\tau_{\mathrm{jc}} \end{cases} .$$

Carrying out the integration yields the following for the current-transfer factor:

$$\frac{\underline{i}_n(x_{\mathrm{cb}})}{\underline{i}_{\mathrm{n}}(x_{\mathrm{bc}})} = \frac{1-\mathrm{e}^{-\mathrm{j}2\omega\tau_{\mathrm{jc}}}}{\mathrm{j}\omega 2\tau_{\mathrm{jc}}} = \frac{\sin(\omega\tau_{\mathrm{jc}})}{\omega\tau_{\mathrm{jc}}}\mathrm{e}^{-\mathrm{j}\omega\tau_{\mathrm{jc}}} \approx \mathrm{e}^{-\mathrm{j}\omega\tau_{\mathrm{jc}}} .$$

Up to very high frequencies, the passage of electrons through the bc depletion is thus described by a phase shift $\mathrm{e}^{-\mathrm{j}\omega\tau_{\mathrm{jc}}}$ corresponding to the collector transit time τ_{jc}.

D.1.5 Small-Signal Analysis

For an approximate characterization of non-quasi-static effects, we expand the admittance parameters

$$\begin{aligned} y_{\mathrm{n11e}} &= g_{\mathrm{m}} \frac{\sinh(\eta/2)}{\sinh(\underline{\vartheta})} \frac{\underline{\vartheta}}{\eta/2} \\ &\quad \times \left[\mathrm{e}^{-\eta/2}\left(\cosh(\underline{\vartheta}) + \frac{\eta/2}{\underline{\vartheta}}\sinh(\underline{\vartheta})\right) - \frac{\sin(\omega\tau_{\mathrm{jc}})}{\omega\tau_{\mathrm{jc}}}\mathrm{e}^{-\mathrm{j}\omega\tau_{\mathrm{jc}}} \right] \end{aligned}$$

and

$$y_{\mathrm{n21e}} = g_{\mathrm{m}} \frac{\underline{\vartheta}}{\eta/2} \frac{\sinh(\eta/2)}{\sinh\underline{\vartheta}} \frac{\sin(\omega\tau_{\mathrm{jc}})}{\omega\tau_{\mathrm{jc}}}\mathrm{e}^{-\mathrm{j}\omega\tau_{\mathrm{jc}}} ,$$

which describe electron transport through the base and the bc depletion layer, up to second order in ω. With the approximations

$$\begin{aligned} \underline{\vartheta}\coth(\underline{\vartheta}) &\approx \frac{\eta}{2}\coth\left(\frac{\eta}{2}\right) + \mathrm{j}\omega\tau_{\mathrm{B0}}\left[\frac{2}{\eta}\coth\left(\frac{\eta}{2}\right) - \frac{1}{\sinh^2(\eta/2)}\right] \\ &\quad +\omega^2\frac{2\tau_{\mathrm{B0}}^2}{\eta^2}\left[\frac{2}{\eta}\coth\left(\frac{\eta}{2}\right) + \frac{1}{\sinh^2(\eta/2)} - \frac{\eta\cosh(\eta/2)}{\sinh^3(\eta/2)}\right] , \end{aligned}$$

$$\frac{\underline{\vartheta}}{\sinh(\underline{\vartheta})} \approx \frac{\eta}{2}\,\frac{1}{\sinh(\eta/2)} + \mathrm{j}\omega\tau_{\mathrm{B0}}\,\frac{1}{\sinh(\eta/2)}\left[\frac{2}{\eta} - \coth\left(\frac{\eta}{2}\right)\right]$$
$$+ \omega^2\frac{2\tau_{\mathrm{B0}}^2}{\eta^2}\,\frac{1}{\sinh(\eta/2)}\left[\coth\left(\frac{\eta}{2}\right) - \eta\coth^2\left(\frac{\eta}{2}\right) + \frac{2}{\eta} + \frac{\eta}{2}\right]$$

and

$$\frac{\sin(\omega\tau_{\mathrm{jc}})}{\omega\tau_{\mathrm{jc}}}\,\mathrm{e}^{-\mathrm{j}\omega\tau_{\mathrm{jc}}} \approx 1 - \mathrm{j}\omega\tau_{\mathrm{jc}} - \frac{2}{3}\,\omega^2\tau_{\mathrm{jc}}^2\,,$$

which are valid to second order of ω, we obtain

$$y_{\mathrm{n11e}} \approx \mathrm{j}\omega g_{\mathrm{m}}(\tau_{\mathrm{Bf}}+\tau_{\mathrm{jc}}) + \omega^2 g_{\mathrm{m}}(\tau_{\mathrm{Bf}}+\tau_{\mathrm{jc}})\tau_2\,, \tag{D.15}$$

where

$$\tau_{\mathrm{Bf}} = \frac{2\tau_{\mathrm{B0}}}{\eta^2}(\mathrm{e}^{-\eta} + \eta - 1)$$

and

$$\tau_2 = \frac{\tau_{\mathrm{Bf}}}{1+\gamma}\left(\frac{1+\mathrm{e}^{-\eta}+\eta+\eta^2/2-\eta\coth(\eta/2)}{(\mathrm{e}^{-\eta}+\eta-1)^2}\right.$$
$$\left. + \gamma\,\frac{\eta\coth(\eta/2)-2}{\mathrm{e}^{-\eta}+\eta-1} + \frac{2}{3}\gamma^2\right)\,,$$

where $\gamma = \tau_{\mathrm{jc}}/\tau_{\mathrm{Bf}}$. The transfer admittance can be approximated by

$$y_{\mathrm{n21e}} \approx g_{\mathrm{m}}\left[1 - \mathrm{j}\omega(\tau_1+\tau_2)\right]\,,$$

where

$$\tau_1 + \tau_2 = \tau_{\mathrm{Bf}}\left(\frac{\eta\coth(\eta/2)-2}{\mathrm{e}^{-\eta}+\eta-1} + \gamma\right)\,.$$

The time constant τ_2 can be expressed as

$$\tau_2 = \frac{\tau_{\mathrm{f}}}{(1+\gamma)^2}\left[f_1(\eta) + \gamma f_2(\eta) + \frac{2}{3}\gamma^2\right]\,, \tag{D.16}$$

where $\gamma = \tau_{\mathrm{jc}}/\tau_{\mathrm{Bf}}$ denotes the ratio of the collector transit time to the base transit time, and $f_1(\eta)$ and $f_2(\eta)$ are functions defined as

$$f_1(\eta) = \frac{\eta^2/2 - \eta^2\coth(\eta/2) + \eta + 1 + \mathrm{e}^{-\eta}}{(\mathrm{e}^{-\eta}+\eta-1)^2}$$

and

$$f_2(\eta) = \frac{\eta\coth(\eta/2) - 2}{\mathrm{e}^{-\eta}+\eta-1}\,.$$

D.2 Quasi-Three-Dimensional Computations of the Base Resistance

The quasi-two-dimensional analysis of Hauser [5] for a stripe transistor of infinite extent is easily extended to the quasi-three-dimensional analysis of planar base regions. The resulting system of equations allows the analysis of more complicated geometries and the consideration of base charge modulation effects. However, these equations require numerical solution in the general case. In the following, the active base layer is assumed to be parallel to the xy

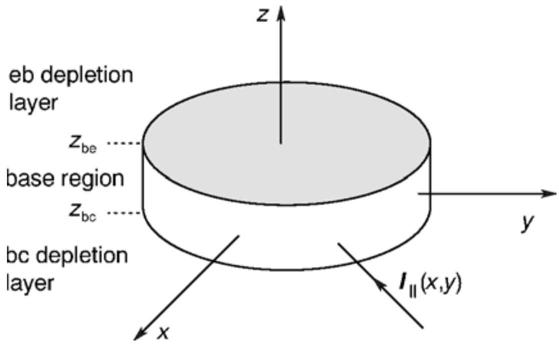


Fig. D.1. Coordinate system used for the quasi-three-dimensional analysis of the active base layer

plane, with boundary planes $z = z_{\mathrm{be}}$ with the eb depletion layer and $z = z_{\mathrm{bc}}$ with the bc depletion layer (see Fig. D.1). Integrating the hole current density equation $\boldsymbol{J}_{\mathrm{p}} = -e\mu_{\mathrm{p}} p \nabla \phi_{\mathrm{p}}$ with respect to z then yields

$$\boldsymbol{I}_{||} = \int_{z_{\mathrm{be}}}^{z_{\mathrm{bc}}} \boldsymbol{J}_{\mathrm{p}}\,\mathrm{d}z = -e\int_{z_{\mathrm{be}}}^{z_{\mathrm{bc}}} \mu_{\mathrm{p}} p\,\nabla_{||}\phi_{\mathrm{p}}\,\mathrm{d}z = -\langle\mu_{\mathrm{p}}\rangle Q'_{\mathrm{p}}\,\nabla_{||}\phi_{\mathrm{p}}\ ,$$

where $\nabla_{||} = \boldsymbol{e}_x\,\partial/\partial x + \boldsymbol{e}_y\,\partial/\partial y$ and the hole quasi-Fermi potential ϕ_{p} has been assumed to be independent of z within the base layer. The quantity

$$Q'_{\mathrm{p}} = e\int_{z_{\mathrm{be}}}^{z_{\mathrm{bc}}} p\,\mathrm{d}z$$

denotes the hole charge per unit area in the base layer, and

$$\langle\mu_{\mathrm{p}}\rangle = \frac{e}{Q'_{\mathrm{p}}}\int_{z_{\mathrm{be}}}^{z_{\mathrm{bc}}} \mu_{\mathrm{p}} p\,\mathrm{d}z$$

is the average hole mobility. Integration of the continuity equation with respect to z yields

$$\nabla_{||}\cdot\boldsymbol{I}_{||} = J_{\mathrm{p},z}(z_{\mathrm{bc}}) - J_{\mathrm{p},z}(z_{\mathrm{be}}) - \partial Q'_{\mathrm{p}}/\partial t\ , \tag{D.17}$$

if recombination in the base layer may be neglected ($\nabla\cdot\boldsymbol{J}_{\mathrm{p}} = -e\,\partial p/\partial t$). The current density of holes injected into the emitter region is

$$J_{\mathrm{p,z}}(z_{\mathrm{be}}) \approx \frac{I_{\mathrm{S}}}{A_{\mathrm{je}}B_{\mathrm{F}}} \exp\left(\frac{\phi_{\mathrm{p}}(x,y)}{V_{\mathrm{T}}}\right)$$

if the electron quasi-Fermi potential in the emitter region is taken to be zero and an ideal forward-biased pn junction characteristic with emission coefficient $N = 1$ is assumed. If stationary conditions and $J_{\mathrm{p},z}(z_{\mathrm{bc}}) = 0$ are assumed, the following system of differential equations is obtained:

$$\boldsymbol{I}_{\|} = -\langle\mu_{\mathrm{p}}\rangle Q'_{\mathrm{p}} \nabla_{\|}\phi_{\mathrm{p}} \quad \text{and} \quad \nabla_{\|}\cdot\boldsymbol{I}_{\|} = -\frac{I_{\mathrm{S}}}{A_{\mathrm{je}}\, B_{\mathrm{F}}} \exp\left(\frac{\phi_{\mathrm{p}}}{V_{\mathrm{T}}}\right) . \tag{D.18}$$

This has to be solved by numerical means in the general case (see [6] for numerical examples). For specialized geometries, however, symmetry considerations may allow one to reduce the partial differential equations to ordinary differential equations, as is shown in the following example.

Analytical Solution for Circular Geometry

If the base region is of circular geometry, it is possible to reduce the partial differential equations (D.18) to a set of ordinary differential equations with the radial coordinate as the independent variable. If $\boldsymbol{I}_{\|}$ is expressed in terms of the portion $I_{\mathrm{B}}(r)$ of the base current that flows into a circular base region of radius r, i.e.

$$\boldsymbol{I}_{\|} = -\frac{1}{2\pi r}\, I_{\mathrm{B}}(r)\boldsymbol{e}_r \ , \tag{D.19}$$

the system

$$\frac{\mathrm{d}I_{\mathrm{B}}}{\mathrm{d}r} \approx \frac{2\pi I_{\mathrm{S}}}{A_{\mathrm{je}}B_{\mathrm{F}}}\, r \exp\left(\frac{\phi_{\mathrm{p}}(r)}{V_{\mathrm{T}}}\right) , \tag{D.20}$$

$$\frac{\mathrm{d}\phi_{\mathrm{p}}}{\mathrm{d}r} = \frac{1}{2\pi\langle\mu_{\mathrm{p}}\rangle} \frac{1}{Q'_{\mathrm{p}}} \frac{I_{\mathrm{B}}(r)}{r} \tag{D.21}$$

has to be solved. Neglecting base charge modulation, these equations may be combined to give [7,8]

$$\frac{\mathrm{d}^2 I_{\mathrm{B}}}{\mathrm{d}r^2} = \frac{1}{r}\frac{\mathrm{d}I_{\mathrm{B}}}{\mathrm{d}r} + 2\lambda_{\mathrm{p}} \frac{I_{\mathrm{B}}(r)}{r}\frac{\mathrm{d}I_{\mathrm{B}}}{\mathrm{d}r} = \frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left(r\frac{\mathrm{d}I_{\mathrm{B}}}{\mathrm{d}r}\right) - \frac{1}{r}\frac{\mathrm{d}I_{\mathrm{B}}}{\mathrm{d}r} ,$$

where $\lambda_{\mathrm{p}} = (4\pi V_{\mathrm{T}}\langle\mu_{\mathrm{p}}\rangle Q'_{\mathrm{p}})^{-1}$; rearranging terms results in

$$\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left(2I_{\mathrm{B}}(r) + \lambda_{\mathrm{p}} I_{\mathrm{B}}^2(r) - r\frac{\mathrm{d}I_{\mathrm{B}}}{\mathrm{d}r}\right) = 0 . \tag{D.22}$$

For $r \neq 0$, this equation can only be fulfilled if the term in the large parantheses is constant. To fulfill the boundary conditions

$$\lim_{r\to 0} I_{\mathrm{B}}(r) = 0 \quad \text{and} \quad \lim_{r\to 0}\left(\frac{\mathrm{d}I_{\mathrm{B}}}{\mathrm{d}r}\right) = 0 ,$$

this constant must equal zero. Therefore, writing $I_B(r) = r^2 f(r)$, we obtain the following differential equation:

$$\mathrm{d}f/\mathrm{d}r = r\lambda_p f^2(r) ;$$

this can be solved by separation of variables, with the general solution

$$f(r) = -\frac{2}{\lambda_p r^2 + \gamma} .$$

The value of γ has to be chosen to fulfill the condition $I_B(r_B) = r_B^2 f(r_B) = I_B$ at the boundary of the base region, which is at a radius r_B. This gives $\gamma = -r_B^2(\lambda_p + 2/I_B)$. The value of I_B as a function of $\phi_p(r_B) = V_{BE}$ is obtained from (D.20), which gives

$$\left.\frac{\mathrm{d}I_B}{\mathrm{d}r}\right|_{r_B} = \frac{2\pi I_S}{A_{je}B_F} r_B \exp\left(\frac{V_{BE}}{V_T}\right) = \frac{\lambda_p I_B^2 + 2I_B}{r_B} , \tag{D.23}$$

by using the fact that

$$\frac{\mathrm{d}I_B}{\mathrm{d}r} = \frac{\mathrm{d}}{\mathrm{d}r}[r^2 f(r)] = 2rf(r) + r^2\frac{\mathrm{d}f}{\mathrm{d}r} = \frac{2I_B + \lambda_p I_B^2}{r} .$$

Solution of the quadratic equation (D.23) for I_B gives

$$I_B = \frac{1}{\lambda_p}\left[\sqrt{1 + 2\frac{\lambda_p I_S}{B_F}\exp\left(\frac{V_{BE}}{V_T}\right)} - 1\right] .$$

For small values of V_{BE}, a Taylor series expansion of the square root yields

$$I_B \approx \frac{I_S}{B_F}\exp\left(\frac{V_{BE}}{V_T}\right) \quad \text{or} \quad V_{B'E'} = V_T \ln\left(\frac{B_F I_B}{I_S}\right) ,$$

since base resistance effects are negligible in this case; for large values of V_{BE}, the approximate relation

$$I_B \approx \sqrt{\frac{2I_S}{B_F\lambda_p}} \exp\left(\frac{V_{BE}}{2V_T}\right) ,$$

or

$$V_{BE} = 2V_T \ln\left(I_B\sqrt{\frac{B_F\lambda_p}{2I_S}}\right) = V_T\ln\left(\frac{B_F I_B}{I_S}\right) + V_T\ln\left(\frac{I_B\lambda_p}{2}\right) ,$$

is obtained. In this result, the first term describes the voltage drop $V_{B'E'}$ in the absence of series resistance effects, whereas the second term is the voltage drop across the internal base resistance. Since the latter depends logarithmically on the current I_B, analogously to an ideal diode, the internal base resistance is sometimes represented by an ideal diode in equivalent circuits (e.g. in [7]).

D.3 Generation of Model Parameters from Layout Data

Demand for careful optimization of transistor layout stems from the fact, that many applications such as mobile or fiber-optic communication systems require device operation near the performance limits of production technologies in order to meet the specified speed, noise and power consumption demands. With stripe transistors as the standard layout, circuit designers focus on the dimension and number of emitter stripes and the number of base contacts: High-speed circuits usually require transistors with minimum emitter stripe width, low-power applications demand for minimum capacitances, while driver circuits and power amplifiers generally show the best performance if the emitter stripe width is chosen in excess of the minimum achievable value.

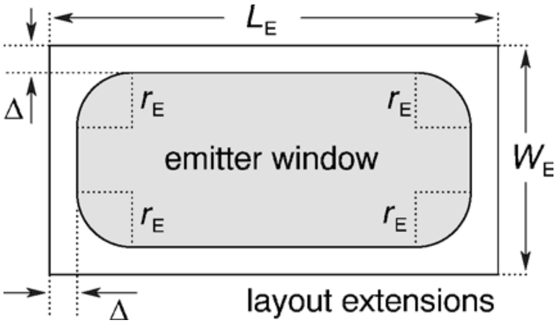


Fig. D.2. Layout extensions and extensions of emitter window with corner rounding

With model parameters X described in terms of a portion that varies in proportion with the area and a portion that varies with the perimeter

$$X = A_{\mathrm{je}} X'_{\mathrm{A}} + P_{\mathrm{je}} X'_{\mathrm{P}} \tag{D.24}$$

it is possible to compute the parameters of the internal transistor from layout data. If corner rounding with radius r_{E} is taken into account, area A_{je} and perimeter P_{je} of the emitter window are calculated as

$$A_{\mathrm{je}} = (W_{\mathrm{E}} - 2\Delta)(L_{\mathrm{E}} - 2\Delta) - (4-\pi) r_{\mathrm{E}}^2$$

$$P_{\mathrm{je}} = 2\,[\,W_{\mathrm{E}} + L_{\mathrm{E}} - 4\Delta - (4-\pi) r_{\mathrm{E}}\,]$$

as is seen from Fig. D.2. X'_{A} and X'_{P} can be determined from measurements of X for transistors with different ratios $\Gamma = P_{\mathrm{je}}/A_{\mathrm{je}}$: a plot of X/A_{je} then gives a straight line with slope X'_{P} and offset X'_{A} if (D.24) is fulfilled. Together with similar calculations for the external transistor, it is possible to generate the model parameters for different transistor geometries from layout data. Such an approach has been presented e.g. in [9] for the HICUM model (see Sect. 3.15).

D.4 Generalization of the Gummel Transfer Current Relation to Arbitrary Geometries

Gummel's integral charge control relation can be generalized to transistors of arbitrary geometry [10]. In the general case, the electron current density has to be described by a vector field $\boldsymbol{J}_{\mathrm{n}}(\boldsymbol{x})$. The continuity equation then reads

$$\frac{\partial J_{\mathrm{n}x}}{\partial x} = e\frac{\partial n}{\partial t} - e(R-G) - \frac{\partial J_{\mathrm{n}y}}{\partial y} - \frac{\partial J_{\mathrm{n}z}}{\partial z}\,, \tag{D.25}$$

i.e. the assumption $\partial J_{\mathrm{n}x}/\partial x$ is not fulfilled even in the absence of recombination and under stationary conditions owing to the terms $\partial J_{\mathrm{n}y}/\partial y$ and $\partial J_{\mathrm{n}z}/\partial z$, which appear in the case of current spreading, i.e. if the current flow is not one-dimensional. In order to treat spreading effects, generalized

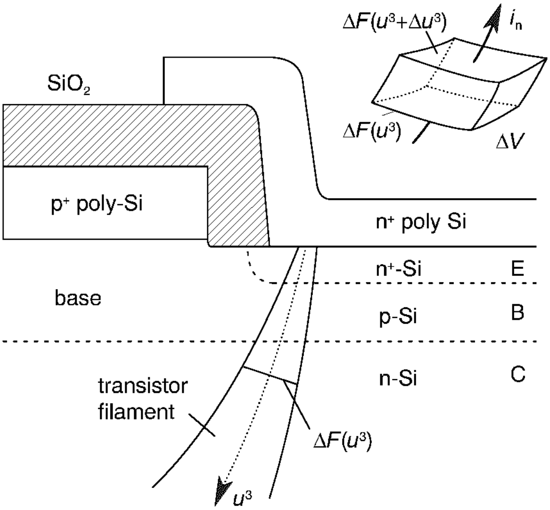


Fig. D.3. Transistor filament defined in terms of electron streamlines, and generalized coordinates used for description in terms of transistor filaments

coordinates are used. This allows one to treat the active transistor volume as a parallel configuration of transistor filaments following the electron streamlines.[1] Figure D.3 shows such a transistor filament – since the boundaries are defined by electron streamlines, the electron current traverses only the infinitesimal contact areas at the ends of the filament on the emitter and collector sides. We use the coordinate u^3 as a length coordinate along the center of the transistor filament and denote the intersections of the center line with the emitter and collector contact surfaces by $u^{3\mathrm{e}}$ and $u^{3\mathrm{c}}$, respec-

[1] The electron streamlines are solutions of the differential equation

$$\mathrm{d}\boldsymbol{x}(\lambda)/\mathrm{d}\lambda = \boldsymbol{J}_{\mathrm{n}}[\,\boldsymbol{x}(\lambda)\,]/|\boldsymbol{J}_{\mathrm{n}}[\,\boldsymbol{x}(\lambda)\,]|\,,$$

where λ denotes a length coordinate.

tively. Mathematically, the description of the active transistor volume as a parallel configuration of transistor filaments corresponds to a transformation of coordinates $\boldsymbol{x} \mapsto \boldsymbol{u}$. We restrict ourselves to orthogonal curvilinear coordinates and use a standard notation [11] that employs the metric coefficients $h_\alpha = |\partial \boldsymbol{x}/\partial u^\alpha|$. The coordinates u^α are assumed to be dimensionless, i.e. the h_α have the dimensions of length. The cross section of the filament at u^3 is $\Delta F(u^3) = h_1 h_2 \, \Delta u_1 \Delta u_2$. Since by definition the electron flow is parallel to the electron streamlines, we obtain the following from the electron continuity equation and the Gauss theorem:

$$I_{\Delta F}(u^3 + \Delta u^3) - I_{\Delta F}(u^3) \; = \; e \, (\partial n/\partial t + R - G) \, \Delta V \; , \tag{D.26}$$

where $\Delta V = h_1 h_2 h_3 \, \Delta u^1 \Delta u^2 \Delta u^3$. Division by $\Delta u^1 \Delta u^2 \Delta u^3$ yields, in the limit $\Delta u^3 \to 0$, the continuity equation

$$\partial i_{\mathrm{n}}/\partial u^3 \; = \; e \, h_1 h_2 h_3 (\partial n/\partial t + R - G) \; , \tag{D.27}$$

in a form similar to the one-dimensional situation considered in Sect. 3.2; the new variable $i_{\mathrm{n}}(u^3)$ is defined by $I_{\Delta F}(u^3) = i_{\mathrm{n}}(u^3) \Delta u^1 \Delta u^2$, and correspondingly $i_{\mathrm{n}}(u^3) = h_1 h_2 |\boldsymbol{J}_{\mathrm{n}}(u^3)|$. Written in u-coordinates, the electron current density equation $\boldsymbol{J}_{\mathrm{n}} = -e\mu_{\mathrm{n}} n \nabla \phi_{\mathrm{n}}$ gives

$$\frac{\partial \phi_{\mathrm{n}}}{\partial u^3} \; = \; -\frac{1}{e} \frac{1}{\mu_{\mathrm{n}} n} \, h_3 \, |\boldsymbol{J}_{\mathrm{n}}(u^3)| \; = \; -\frac{1}{e \mu_{\mathrm{n}} n} \, \frac{h_3}{h_1 h_2} \, i_{\mathrm{n}} \; . \tag{D.28}$$

Taking account of the fact that $n = n_{\mathrm{ic}} \exp[\,(\psi - \phi_{\mathrm{n}})/V_{\mathrm{T}}\,]$ and $D_{\mathrm{n}} = V_{\mathrm{T}} \mu_{\mathrm{n}}$, this equation can be rewritten as

$$e n_{\mathrm{i0}}^2 h_1(u^{3\mathrm{e}}) h_2(u^{3\mathrm{e}}) \frac{\partial}{\partial u^3} \exp\!\left(\frac{\phi_{\mathrm{pB}} - \phi_{\mathrm{n}}}{V_{\mathrm{T}}} \right) \; = \; i_{\mathrm{n}} \, \frac{\partial \gamma_{\mathrm{B}}}{\partial u^3} \; , \tag{D.29}$$

where

$$\gamma_{\mathrm{B}}(\boldsymbol{u}) \; = \; \int_{u^{3\mathrm{e}}}^{u^3} \frac{h_1(u^{3\mathrm{e}}) h_2(u^{3\mathrm{e}})}{h_1(u^3) h_2(u^3)} \, \frac{p}{D_{\mathrm{n}}} \, \frac{n_{\mathrm{i0}}^2}{n_{\mathrm{ic}}^2} \, \exp\!\left(\frac{\phi_{\mathrm{pB}} - \phi_{\mathrm{p}}(u^1, u^2)}{V_{\mathrm{T}}} \right) h_3 \, \mathrm{d}u^3 \; .$$

Assuming $\partial n/\partial t$ and $R - G$ to be negligible, an integration of (D.29) yields the following for the transfer current carried by each filament:

$$i_{\mathrm{n}}(u^{3c}) \; = \; \frac{e n_{\mathrm{i0}}^2 h_1(u^{3e}) h_2(u^{3e})}{\gamma_{\mathrm{B}}(u^1, u^2, u^{3c})} \left[\exp\!\left(\frac{V_{\mathrm{BE}}}{V_{\mathrm{T}}} \right) - \exp\!\left(\frac{V_{\mathrm{BC}}}{V_{\mathrm{T}}} \right) \right] \; ,$$

where $V_{\mathrm{BE}} = \phi_{\mathrm{pB}} - \phi_{\mathrm{n}}(u^{3\mathrm{e}})$ and $V_{\mathrm{BC}} = \phi_{\mathrm{pB}} - \phi_{\mathrm{n}}(u^{3\mathrm{c}})$. As the ratio $h_1(u^{3\mathrm{e}}) h_2(u^{3\mathrm{e}})/h_1(u^3) h_2(u^3) = \Delta F(u^{3\mathrm{e}})/\Delta F(u^3)$ describes changes of the cross section of the transistor filament, the value of $\gamma_{\mathrm{B}}(u^{3\mathrm{c}})$ is reduced (and hence the transfer current is increased) if $\Delta F(u^{3\mathrm{e}})/\Delta F(u^3) < 1$ (i.e. if the filament widens). In contrast, the value of $\gamma_{\mathrm{B}}(u^{3\mathrm{c}})$ is reduced (and hence the transfer current is increased) if the transistor filament narrows, i.e.

$\Delta F(u^{3e})/\Delta F(u^3) > 1$; in the case of one-dimensional carrier motion, where $\Delta F(u^{3e})/\Delta F(u^3) = 1$, the result reduces to the conventional formulation. Integration over the area of the emitter contact yields the transfer current relation

$$I_{\rm T} = \frac{en_{\rm i0}^2 A_{\rm je}}{G_{\rm B}} \left[\exp\left(\frac{V_{\rm BE}}{V_{\rm T}}\right) - \exp\left(\frac{V_{\rm BC}}{V_{\rm T}}\right)\right] \Omega_{\Delta\phi_{\rm p}} ,$$

with a Gummel number defined by

$$\frac{1}{G_{\rm B}} = \frac{1}{A_{\rm je}} \int_{A_{\rm je}} \frac{h_1(u^{3e})h_2(u^{3e})}{g_{\rm B}(u^1,u^2)}\, {\rm d}u^1\, {\rm d}u^2 ,$$

where

$$g_{\rm B}(u^1,u^2) = \int_{u^{3e}}^{u^{3c}} \frac{h_1(u^{3e})h_2(u^{3e})}{h_1(u^3)h_2(u^3)} \frac{p}{D_{\rm n}} \frac{n_{\rm i0}^2}{n_{\rm ie}^2}\, h_3\, {\rm d}u^3 .$$

The term

$$\Omega_{\Delta\phi_{\rm p}} = \frac{G_{\rm B}}{A_{\rm je}} \int_{A_{\rm je}} \frac{h_1(u^{3e})h_2(u^{3e})}{g_{\rm B}(u^1,u^2)} \exp\left(\frac{\phi_{\rm p}(u^1,u^2) - \phi_{\rm pB}}{V_{\rm T}}\right) {\rm d}u^1\, {\rm d}u^2$$

denotes a factor that stems from the spatial variation of the hole quasi-Fermi potential and can be expressed in terms of the internal base resistance $R_{\rm BB'}$ by making the identification $\Omega_{\Delta\phi_{\rm p}} = \exp(-R_{\rm BB'}I_{\rm B}/V_{\rm T})$. Note that if the current flow is not one-dimensional, the Gummel number is not proportional to the "base charge"

$$Q_{\rm B} = e \int p(\boldsymbol{u})\, h_1 h_2 h_3\, {\rm d}u^1 {\rm d}u^2 {\rm d}u^3 ,$$

even if $D_{\rm n}$ and $n_{\rm ie}$ are assumed to be constant.

D.5 Definition of Series Resistances Within the Integral Charge Control Relation

In the application of Gummel's integral charge control relation, which transforms the electron current density equation into its integral form, the question arises as to where the limits of integration should be chosen: if the integration is performed from the emitter to the collector contact, emitter and collector series resistance effects will modify the Gummel integral (see Sect. 3.2). The aim of the following discussion is to show how a consistent separation of the two effects can be obtained. As was pointed out in Sect. 2.8, the electron current density equation can be written in two equivalent forms, namely

$$\frac{\partial \phi_{\rm n}}{\partial x} = -\frac{1}{e\mu_{\rm n} n} J_{\rm n} \tag{D.30}$$

and

$$\frac{\partial}{\partial x}\exp\left(-\frac{\phi_\mathrm{n}}{V_\mathrm{T}}\right) = \frac{p}{eD_\mathrm{n}n_\mathrm{ie}^2}\exp\left(-\frac{\phi_\mathrm{p}}{V_\mathrm{T}}\right)J_\mathrm{n}\ , \tag{D.31}$$

where the first form is particularly suited for the description of majority-carrier flow (resistive phenomena), whereas the second is particularly suited for the description of minority-carrier flow (diffusive phenomena). In order to take account of series resistances within Gummel's integral charge control relation, it is therefore reasonable to use (D.30) in regions where electrons represent the majority of carriers and (D.31), where holes represent the majority. If we introduce appropriate weight functions[2] Γ_n and Γ_p, where $\Gamma_\mathrm{n}+\Gamma_\mathrm{p}=1$, together with the function

$$\Delta\phi_\mathrm{n} = -\frac{1}{e}\int_{x_0}^{x}\frac{J_\mathrm{n}}{\mu_\mathrm{n}n}\Gamma_\mathrm{n}\mathrm{d}x\ ,$$

(D.30) can be written as

$$\frac{\partial\phi_\mathrm{n}}{\partial x} = -\frac{\Gamma_\mathrm{p}+\Gamma_\mathrm{n}}{e\mu_\mathrm{n}n}J_\mathrm{n} = -\frac{\Gamma_\mathrm{p}}{e\mu_\mathrm{n}n}J_\mathrm{n}+\frac{\partial}{\partial x}\Delta\phi_\mathrm{n}\ , \tag{D.32}$$

or in the equivalent form (where $D_\mathrm{n}=V_\mathrm{T}\mu_\mathrm{n}$)

$$\frac{\partial}{\partial x}\exp\left(-\frac{\phi_\mathrm{n}-\Delta\phi_\mathrm{n}}{V_\mathrm{T}}\right) = \frac{1}{eD_\mathrm{n}}\frac{\Gamma_\mathrm{p}}{n_\mathrm{ie}^2}p\exp\left(-\frac{\phi_\mathrm{p}}{V_\mathrm{T}}\right)\exp\left(\frac{\Delta\phi_\mathrm{n}}{V_\mathrm{T}}\right)J_\mathrm{n}\ .$$

If J_n is taken to be constant for simplicity, an integration from the emitter to the collector contact yields for the transfer current $I_\mathrm{T}=-A_\mathrm{je}J_\mathrm{n}$:

$$I_\mathrm{T} = eA_\mathrm{je}\frac{\exp\left(-\dfrac{\phi_\mathrm{n}(x_\mathrm{e})-\Delta\phi_\mathrm{n}(x_\mathrm{e})}{V_\mathrm{T}}\right)-\exp\left(-\dfrac{\phi_\mathrm{n}(x_\mathrm{c})-\Delta\phi_\mathrm{n}(x_\mathrm{c})}{V_\mathrm{T}}\right)}{\displaystyle\int_{x_\mathrm{e}}^{x_\mathrm{c}}\frac{p}{D_\mathrm{n}n_\mathrm{ie}^2}\Gamma_\mathrm{p}\exp\left(-\frac{\phi_\mathrm{p}}{V_\mathrm{T}}\right)\exp\left(\frac{\Delta\phi_\mathrm{n}}{V_\mathrm{T}}\right)\mathrm{d}x}\ .$$

In the integral in the denominator, the spatial dependence of $\phi_\mathrm{p}=\phi_\mathrm{pB}$ may be neglected, as the integrand becomes very small outside the region dominated by holes. Application of the mean-value theorem allows us then to write

$$\begin{aligned}&\int_{x_\mathrm{e}}^{x_\mathrm{c}}\frac{p}{D_\mathrm{n}n_\mathrm{ie}^2}\Gamma_\mathrm{p}\exp\left(-\frac{\phi_\mathrm{p}}{V_\mathrm{T}}\right)\exp\left(\frac{\Delta\phi_\mathrm{n}}{V_\mathrm{T}}\right)\mathrm{d}x\\ &\qquad = \exp\left(-\frac{\phi_\mathrm{pB}}{V_\mathrm{T}}\right)\exp\left(\frac{\Delta\phi_\mathrm{n}(x_\mathrm{b})}{V_\mathrm{T}}\right)\int_{x_\mathrm{e}}^{x_\mathrm{c}}\frac{p}{D_\mathrm{n}n_\mathrm{ie}^2}\Gamma_\mathrm{p}\,\mathrm{d}x\ ,\end{aligned}$$

where x_b denotes a point within the base region; the exact location within the base region is not of practical interest, since $\Gamma_\mathrm{n}\approx 0$ within the base layer, and

[2] Possible choices are $\Gamma_\mathrm{p}=p/(n+p)$ and $\Gamma_n=n/(n+p)$ or, even simpler, $\Gamma_\mathrm{p}=1$ if $p>n$ and $\Gamma_\mathrm{p}=0$ elsewhere. The latter choice corresponds to the situation where the limits of integration are chosen near the base depletion-layer boundaries.

therefore $\Delta\phi_{\mathrm{n}}(x) \approx$ const. within this region. If we write $V_{\mathrm{B'E}} = \phi_{\mathrm{pB}} - \phi_{\mathrm{n}}(x_{\mathrm{e}})$ and $V_{\mathrm{B'C}} = \phi_{\mathrm{pB}} - \phi_{\mathrm{n}}(x_{\mathrm{c}})$ and use the definitions

$$R_{\mathrm{EE'}} I_{\mathrm{E}} = \Delta\phi_{\mathrm{n}}(x_{\mathrm{e}}) - \Delta\phi_{\mathrm{n}}(x_{\mathrm{b}}) = \frac{1}{e}\int_{x_{\mathrm{e}}}^{x_{\mathrm{b}}} \frac{J_{\mathrm{n}}}{\mu_{\mathrm{n}} n}\, \Gamma_{\mathrm{n}}\, \mathrm{d}x \,, \tag{D.33}$$

$$R_{\mathrm{CC'}} I_{\mathrm{C}} = \Delta\phi_{\mathrm{n}}(x_{\mathrm{b}}) - \Delta\phi_{\mathrm{n}}(x_{\mathrm{c}}) = \frac{1}{e}\int_{x_{\mathrm{b}}}^{x_{\mathrm{c}}} \frac{J_{\mathrm{n}}}{\mu_{\mathrm{n}} n}\, \Gamma_{\mathrm{n}}\, \mathrm{d}x \,, \tag{D.34}$$

the integral charge control relation transforms into

$$I_{\mathrm{T}} = eA_{\mathrm{jc}} \frac{\exp\left(\dfrac{V_{\mathrm{B'E}} - R_{\mathrm{EE'}} I_{\mathrm{E}}}{V_{\mathrm{T}}}\right) - \exp\left(\dfrac{V_{\mathrm{B'C}} + R_{\mathrm{CC'}} I_{\mathrm{C}}}{V_{\mathrm{T}}}\right)}{\displaystyle\int_{x_{\mathrm{e}}}^{x_{\mathrm{c}}} \frac{p}{D_{\mathrm{n}} n_{\mathrm{ie}}^2}\, \Gamma_{\mathrm{p}}\, \mathrm{d}x} \,.$$

This discussion clearly demonstrates that a consistent definition of emitter and collector series resistances within the integral charge control relation requires a truncation of the Gummel integral. If the truncation is performed close to the collector-side depletion layer edge of the base region (forward operation is assumed), the definition of the collector series resistance in (D.34) becomes problematic, as its computation requires an integration across the depletion layer, where carrier transport is nonohmic. In this case it is better to use the approach outlined in Sect. 3.2.

D.6 Multiplication Factor

The multiplication factor M_{n} due to injected electrons often shows the following asymptotic behavior:

$$1 - \frac{1}{M_{\mathrm{n}}} \approx \begin{cases} (V_{\mathrm{C'B'}}/BV_{\mathrm{C}})^{N_1} & \text{for} \quad 1 - 1/M_{\mathrm{n}} \ll 1 \\ (V_{\mathrm{C'B'}}/BV_{\mathrm{CBO}})^{N_2} & \text{for} \quad 1 - 1/M_{\mathrm{n}} \approx 1 \end{cases} \,. \tag{D.35}$$

By adding the reciprocals of the two expressions, we obtain a formula that contains both limiting cases. Since this approach of reciprocal addition results in a value of $1 - 1/M_{\mathrm{n}}$ that is somewhat too small, weighting factors

$$f_k(V - V_{\mathrm{cr}}) = \frac{1}{1 + \alpha \exp[\,\beta_k(V - V_{\mathrm{cr}})\,]} \tag{D.36}$$

are used to "continuously switch" between the two asymptotic relations around the crossover voltage V_{cr} defined in Fig. 3.57, at which both asymptotic curves show the same limiting behavior. The value of α is determined so as to obtain the correct value of $1 - 1/M_{\mathrm{n}}$ at $V_{\mathrm{CB}} = V_{\mathrm{cr}}$, i.e.

$$\alpha = 2\left(1 - \frac{1}{M_{\mathrm{n}}}\right)\left(\frac{BV_{\mathrm{C}}}{V_{\mathrm{cr}}}\right)^{N_1} - 1 \,. \tag{D.37}$$

The parameters β_k are typically chosen as approximately 2/V, which yields good results. An improved description is obtained if the parameters $V_{\rm cr}$, β_1 and β_2 are determined by a least-squares fit. The multiplication factor $M_{\rm n}$ is then approximated by

$$1 - \frac{1}{M_{\rm n}} = \frac{1}{f_1(V - V_{\rm cr}) \left(\frac{BV_1}{|V_{\rm CB}| + \delta}\right)^{N_1} + f_2(V_{\rm cr} - V) \left(\frac{BV_2}{|V_{\rm CB}| + \delta}\right)^{N_2}} ,$$

where δ is a small positive constant (e.g. $\delta = 10^{-8}$), introduced to avoid singularities in the case $V_{\rm BC} = 0$. This formula has been implemented in a SPICE model, which showed good convergence of the results for voltages $V_{\rm CB}$ up to $BV_{\rm CBO}$. The advantage of this approach is that it usually provides a good representation of experimental results, and that it employs parameters that are easily obtained from measured data; furthermore, simple formulas are obtained for hand calculation in the region of weak carrier multiplication.

D.7 References

[1] J. te Winkel. Drift transistor, simplified electrical characterization. *Radio Electron. Eng.*, 36:280 288, 1957.

[2] J. te Winkel. Transmission line analogue of a drift transistor. *Philips Res. Repts.*, 14:52–64, 1959.

[3] D.E. Thomas, J.L. Moll. Junction transistor short-circuit current gain and phase determination. *Proc. IRE*, 46:1177–1184, 1958.

[4] R.L. Pritchard. *Electrical Characteristics of Transistors.* McGraw-Hill, New York, 1967.

[5] J.R. Hauser. The effects of distributed base potential on emitter-current injection density and effective base resistance for stripe transistor geometries. *IEEE Trans. Electron Devices*, 11:238–242, 1964.

[6] R.E. Lowther, J. Johnston. Fast, quasi-3d modeling of base resistance for circuit simulation. *IEEE Trans. Electron Devices*, 38(3):518 526, 1991.

[7] G. Rey. Effets de la defocalisation (c.c. et c.a.) sur le comportement des transistors a jonctions. *Solid-State Electron.*, 12:645–659, 1969.

[8] E.W. Greeneich. Base spreading resistance of polysilicon self-aligned bipolar transistors. *IEEE Trans. Electron Devices*, 36(1):147–149, 1989.

[9] M. Schröter, H.-M. Rein, W. Rabe, R. Reimann, H.-J. Wassener, A. Koldehoff. Physics- and process-based bipolar tranistor modeling for integrated circuit design. *IEEE J. Solid-State Circuits*, 34(8):1136–1149, 1999.

[10] M. Reisch. Integral relations for bipolar transistor modeling. *IEDM Tech. Dig.*, pp. 126–129, 1988.

[11] P.M. Morse, H. Feshbach. *Methods of Theoretical Physics.* McGraw-Hill, New York, 1953.

E Noise: Details

This section includes a brief summary of some statistical concepts relevant to noise calculations. In addition, details of noise calculations for pn diodes and junction transistors are provided.

E.1 Some Statistics

E.1.1 Stochastic Variables, Correlation

The voltages and currents of noisy devices are examples of stochastic, or random, variables. The temporal average of the stochastic variable $x(t)$ is defined as

$$X = \overline{x} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)\,\mathrm{d}t\ ;$$

the variance $\sigma_x^2 = \overline{[\,x(t) - X\,]^2}$ determines the mean square deviation of the random variable $x(t)$ from its mean value X:

$$\sigma_x^2 = \mathrm{var}(x) = \lim_{T\to\infty} \frac{1}{T} \int_{-\mathrm{T}/2}^{T/2} [\,x(t) - X\,]^2\,\mathrm{d}t\ . \tag{E.1}$$

The correlation function $R_{xy}(\tau)$ of two random variables $x(t)$ and $y(t)$ is defined as

$$R_{xy}(\tau) = \overline{\Delta x(t)\Delta y(t+\tau)} = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} \Delta x(t)\Delta y(t+\tau)\,\mathrm{d}t\ , \tag{E.2}$$

where $\Delta x(t) = x(t) - X$ and $\Delta y(t) = y(t) - Y$. The correlation coefficient c_{xy} of the two random variables is related to the correlation function by

$$R_{xy}(0) = c_{xy}\sigma_x\sigma_y\ . \tag{E.3}$$

The correlation between two noisy signals is of importance if the variance of superposed noisy signals is to be calculated:

$$\overline{(\Delta x + \Delta y)^2} = \sigma_x^2 + \sigma_y^2 + 2c_{xy}\sigma_x\sigma_y\ . \tag{E.4}$$

The autocorrelation function of the random variable $x(t)$ is given by

$$R_x(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} \Delta x(t)\,\Delta x(t+\tau)\,\mathrm{d}t \tag{E.5}$$

and determines how fast a given fluctuation decays within a time interval of length τ. The autocorrelation function is symmetric, i.e.

$$R_x(\tau) \;=\; R_x(-\tau) \tag{E.6}$$

and has its maximum value at $\tau = 0$,

$$R_x(0) \;=\; \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} \Delta x^2(t)\,\mathrm{d}t \;=\; \sigma_x^2\,; \tag{E.7}$$

it decays to zero with increasing τ. This corresponds to the fact that for large time differences there is no correlation between the values of the random variable.

E.1.2 Ensemble Average, Distribution Function

Besides the temporal average, which is useful for the characterization of noise in stationary systems, the ensemble average of the stochastic variable of interest is employed, this ensemble average can also be applied to nonstationary systems. The ensemble average of $x(t)$ is defined in terms of an ensemble of identical systems, each characterized by a stochastic variable $x_\alpha(t)$:

$$\langle x(t)\rangle \;=\; \lim_{N\to\infty} \frac{1}{N}\sum_{\alpha=1}^{N} x_\alpha(t)\;.$$

Ensemble averages are usually calculated with the help of the probability density $f(x,t)$, defined such that $f(x,t)\,\Delta x$ gives the probability of finding the random variable x at time t within the interval $[\,x, x+\Delta x\,]$. Since the probability of finding the system in any one of its allowed states must equal one, the probability density is normalized to one:

$$\int f(x,t)\,\mathrm{d}t \;=\; 1 \qquad \text{for all values of } t\;. \tag{E.8}$$

The ensemble average of a function $g(x)$ is expressed in terms of the probability density according to

$$\langle g(t)\rangle \;=\; \int g(x)\,f(x,t)\,\mathrm{d}x\;. \tag{E.9}$$

A frequently employed example of a probability density function is the Gaussian distribution,

$$f_{\mathrm{G}}(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}}\,\exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)\,, \tag{E.10}$$

which describes a distribution of values of x around their average value x_0 with a variance σ^2. Most fluctuating quantities generated in electronic devices can be described by a probability density function of this form. The reason for this is the central limit theorem, which states that a random variable which is determined by a large number of independent processes tends to have a Gaussian probability density function.

A stationary system is called ergodic if its temporal and ensemble averages are equal. This will be assumed in the following. The variance and the correlation function may then each be expressed in two alternative ways:

$$\sigma_x^2 \;=\; \overline{[\,x(t) - X\,]^2} \;=\; \langle\,[\,x(t) - \langle x\rangle]^2\,\rangle \tag{E.11}$$

and

$$R_{xy}(\tau) \;=\; \overline{\Delta x(t)\Delta y(t+\tau)} \;=\; \langle \Delta x(0)\Delta y(\tau)\rangle\;. \tag{E.12}$$

E.1.3 Spectral Density

For the analysis of electronic circuits, the spectral density of the noisy variable under consideration is required. The spectral density $S_x(f)$ of a random variable $x(t)$ can be calculated from its autocorrelation function by use of the the Wiener-Khintchin theorem, which gives

$$S_x(f) \;=\; 4\int_0^\infty R_x(\tau)\cos(2\pi f\tau)\,\mathrm{d}\tau \tag{E.13}$$

and, vice versa,

$$R_x(\tau) \;=\; \int_0^\infty S_x(f)\cos(2\pi f\tau)\,\mathrm{d}f\;; \tag{E.14}$$

in particular,

$$R_x(0) \;=\; \int_0^\infty S_x(f)\,\mathrm{d}f \;=\; \sigma_x^2\;. \tag{E.15}$$

If the fluctuations of the measured quantity y are caused by fluctuations of a parameter x, then the spectral density S_y depends on the spectral density S_x according to

$$S_y \;=\; \left(\frac{\partial y}{\partial x}\right)^2 S_x\;. \tag{E.16}$$

E.1.4 Carson Theorem, Shot Noise

If a stationary random variable $x(t)$ is the sum of a large number of identical events $\phi(t-t_i)$,

$$x(t) \;=\; \sum_i \phi(t-t_i)\;, \tag{E.17}$$

where the t_i are assumed to be statistically distributed, with $\lambda\,\Delta t$ events on average in a time interval of length Δt, the temporal average is given by

$$X \;=\; \lambda\int_{-\infty}^{\infty}\phi(t)\,\mathrm{d}t\;, \tag{E.18}$$

and the variance of $x(t)$ can be shown to be given by

$$\sigma_x^2 = \lambda \int_{-\infty}^{\infty} \phi^2(t)\, \mathrm{d}t , \tag{E.19}$$

a relation commonly referred to as the Campbell theorem. The fluctuations $\Delta x(t) = x(t) - X$ are described by the spectral density

$$S_x(\omega) = 2\lambda\, |\phi(\omega)|^2 , \tag{E.20}$$

a relation that is known as the Carson theorem. A typical application of the Carson theorem is the shot noise model, which describes the current carried by individual carriers as a random succession of individual charge transfer events, adding up to an average current $I = \lambda e$, which results in

$$S_i(\omega) = 2I|\phi(\omega)|^2/e , \tag{E.21}$$

where $\phi(\omega)$ is the Fourier transform of an individual current pulse. Assuming the current pulses to be rectangular with length t_f, we obtain

$$\phi(\omega) = \frac{e}{t_\mathrm{f}} \int_0^{t_\mathrm{f}} \mathrm{e}^{-\mathrm{j}\omega t}\, \mathrm{d}t , \tag{E.22}$$

and therefore

$$|\phi(\omega)|^2 = e^2 \frac{\sin^2(\omega t_\mathrm{f}/2)}{(\omega t_\mathrm{f}/2)^2} = 2e^2 \frac{1-\cos(\omega t_\mathrm{f})}{(\omega t_\mathrm{f})^2} , \tag{E.23}$$

in accordance with (5.7).

E.2 Velocity Fluctuations and Diffusion

The velocity fluctuation $\Delta v(t) = v(t) - v_\mathrm{n}$ denotes the deviation of the velocity $v(t)$ of a carrier at time t from the (average) drift velocity $v_\mathrm{n} = \langle v(t) \rangle$. Velocity fluctuations cause variations of the distance traveled by carriers during a time interval $[\,0, t\,]$ equal to

$$\Delta x(t) = \int_0^t \Delta v(t')\, \mathrm{d}t' , \tag{E.24}$$

where $\langle \Delta x(t) \rangle = 0$. The mean square deviation therefore is

$$\langle \Delta x(t)^2 \rangle = \int_0^t \int_0^t \langle \Delta v(t') \Delta v(t'') \rangle\, \mathrm{d}t'\, \mathrm{d}t'' . \tag{E.25}$$

In the case of stationary processes, the velocity autocorrelation function

$$\langle \Delta v(t') \Delta v(t'') \rangle = \langle \Delta v(0) \Delta v(\tau) \rangle = R_{\Delta v}(\tau)$$

depends only on the difference $t'' - t' = \tau$; a corresponding substitution transforms (E.25) into

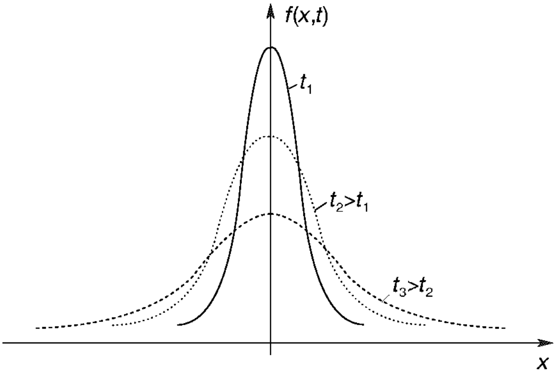


Fig. E.1. Temporal evolution of the probability density function $f(x,t)$ for a diffusing particle initially located at $x = 0$

$$\langle \Delta x(t)^2 \rangle \;=\; \int_0^t \int_{-\infty}^{\infty} R_{\Delta v}(\tau)\,\mathrm{d}\tau\,\mathrm{d}t \;=\; t\int_{-\infty}^{\infty} R_{\Delta v}(\tau)\,\mathrm{d}\tau \;, \tag{E.26}$$

where the limits of integration for τ have been replaced by infinity assuming that $t \gg \tau_{\mathrm{c}}$, where τ_{c} is the collision time; this is justified because the autocorrelation function $R_{\Delta v}(\tau)$ differs significantly from zero only in a time interval of the order of τ_{c}, because after a few scattering events practically no correlation exists between velocity values. A nonvanishing mean square deviation corresponds to carrier diffusion; the diffusion coefficient D is given by

$$\langle \Delta x(t)^2 \rangle \;=\; 2Dt \;. \tag{E.27}$$

The probability that a carrier at time t is in an interval between x and $x+\Delta x$ is $f(x,t)\Delta x$; the probability density $f(x,t)$ has to obey the diffusion equation

$$\frac{\partial}{\partial t} f(x,t) \;=\; eD\,\frac{\partial^2}{\partial x^2} f(x,t) \;. \tag{E.28}$$

The solution for an electron initially located at $x = 0$ (i.e. $f(x,0) = \delta(x)$) is

$$f(x,t) = \frac{1}{\sqrt{4\pi Dt}}\,\exp\left(-\frac{x^2}{4Dt}\right) \;, \tag{E.29}$$

and is sketched in Fig. E.1; at time t, the mean square deviation from $x = 0$ is given by

$$\langle x(t)^2 \rangle \;=\; 2Dt \;. \tag{E.30}$$

A comparison with (E.27) therefore yields a definition of the diffusion coefficient in terms of the velocity autocorrelation function:

$$D \;=\; \frac{1}{2}\int_{-\infty}^{\infty} R_{\Delta v}(\tau)\,\mathrm{d}\tau \;. \tag{E.31}$$

E.3 Thermodynamics and Noise

The Nyquist theorem, which describes thermal noise produced by ohmic resistors, may be generalized to linear two-terminal networks characterized by a frequency-dependent admittance $Y(f)$, which produce a noise current with a spectral density

$$S_i(f) = 4k_\mathrm{B}T\,\mathrm{Re}\,[\,Y(f)]\ . \tag{E.32}$$

The Nyquist theorem predicts a white noise spectrum and therefore implies an infinitely large deliverable noise power if integrated over all frequencies, an obvious error that may be traced back to the application of classical thermodynamics (e.g. in the form of the equipartition theorem). A calculation based on the rules of quantum theory shows that the Nyquist theorem may be applied only for frequencies that obey $hf \ll k_\mathrm{B}T$; its generalization reads [1]

$$S_i(f) = 2hf\,\mathrm{Re}\,[\,Y(f)]\coth\left(\frac{hf}{2k_\mathrm{B}T}\right)\ . \tag{E.33}$$

For frequencies $f \ll k_\mathrm{B}T/h$, i.e. for the range of frequencies considered in this book, the approximation $\xi\coth\xi \approx 1$, which is valid for $|\xi| \ll 1$ yields (E.32).

The Nyquist theorem represents a special form of the fluctuation dissipation theorem [2], which was formulated for fluctuations from a state of thermodynamic equilibrium, a situation that does not exist if a voltage or a current is applied to an electronic device. In such a driven system, i.e. a system operated in a nonequilibrium steady state, there is not necessarily a connection between the noise and the impedance, as was pointed out by Richardson [3]. However we may often assume that the system differs only slightly from its equilibrium form and expect a somewhat modified Nyquist relation to be applicable.

A generalization of the fluctuation–dissipation theorem for purely resistive nonlinear devices was presented in [4, 5]. According to this generalization, the spectral density of the noise voltage generated in a nonlinear device that carries an average current I_0 is given by

$$S_v(f) = 4k_\mathrm{B}TP_1/I_1^2\ ,$$

where I_1 is the rms value of a small-signal test current that flows through the device in addition to I_0, and P_1 is the excess power dissipation caused by this additional noise current. In terms of the current–voltage characteristic of the nonlinear resistor, this results in a noise voltage with spectral density

$$S_v(f) = 4k_\mathrm{B}T\left(\left.\frac{\mathrm{d}V}{\mathrm{d}I}\right|_{I_0} + \frac{I_0}{2}\left.\frac{\mathrm{d}^2V}{\mathrm{d}I^2}\right|_{I_0}\right)\ . \tag{E.34}$$

Under the assumption that the noise generated by a device is determined solely by the device and not by the circuit connected to it, it is possible to transform the spectral density of the generated noise voltage to the spectral density of an equivalent noise current source

$$S_i(f) \;=\; 4k_{\mathrm{B}}T(g-\beta I_0) \ , \tag{E.35}$$

where $g = \mathrm{d}I/\mathrm{d}V|_{I_0}$ is the small-signal conductance at the bias point, and

$$\beta \;=\; \frac{1}{2}\;\frac{\mathrm{d}^2I/\mathrm{d}V^2}{\mathrm{d}I/\mathrm{d}V}\bigg|_{I_0}$$

serves as a measure of the nonlinearity of the device.[1] The formulas (E.34) and (E.35) were derived on the basis of several assumptions listed in [5]. Here we wish to emphasize only the assumption that the system is purely resistive, i.e. it may not store energy except in the form of heat.

Example. As an example, we apply the theorem to an ideal diode with the current-voltage characteristic

$$I(V) \;=\; I_{\mathrm{S}}\left[\exp\!\left(\frac{V}{V_{\mathrm{T}}}\right) - 1\right] \ ,$$

where

$$\frac{\mathrm{d}I}{\mathrm{d}V}\bigg|_{I_0} \;=\; \frac{I_0+I_{\mathrm{S}}}{V_{\mathrm{T}}} \qquad \text{and} \qquad \frac{\mathrm{d}^2I}{\mathrm{d}V^2}\bigg|_{I_0} \;=\; \frac{I_0+I_{\mathrm{S}}}{V_{\mathrm{T}}^2} \ .$$

Since $V_{\mathrm{T}} = k_{\mathrm{B}}T/e$, the spectral density of the diode noise current is obtained from (E.35) as

$$S_i(f) \;=\; 2e(I_0+2I_{\mathrm{S}}) \ ,$$

in accordance with the classical shot noise formula. Application of (E.35) is possible only to purely resistive devices. While this assumption seems acceptable for a Schottky diode, it is problematic in the case of pn junctions, which show minority-carrier storage in the n and p bulk regions under forward-bias conditions. Each minority carrier stores an energy of about W_{g} in a form that is not associated with thermal motion, in contradiction to the assumption of purely resistive behavior. The theorem may therefore be applied only for frequencies that obey $\omega\tau \ll 1$, where τ denotes the minority-carrier lifetime, since in this limit minority-carrier storage may be neglected.

[1] A similar result was obtained earlier by Gunn [6]. Gupta's approach [4,5] was criticized by Stratonovich [7], who stated that Gupta's formulas are not generally applicable. In [8] it was shown that the assumption of a Gaussian white-noise current model for a nonlinear resistor is in contradiction with thermodynamic principles. Unfortunately, the author could not find a formula that allows one to estimate the errors associated with Gupta's formula. For further results on the thermodynamic approach to noise in nonlinear networks, see [9–11] and references therein.

E.4 Generation–Recombination Noise

This section gives a justification for the spectral density function $S_{\Delta n}(\omega)$ introduced in (5.24) to characterize fluctuations of the particle density. Let $P(N,t)$ denote the probability that N electrons are in the conduction band at time t, $g(N)\Delta t$ the probability that an additional electron is generated within a time interval of length Δt, and $r(N)\Delta t$ the probability of the recombination of an electron–hole pair within Δt. The value of $P(N,t)$ then obeys a so-called master equation

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}P(N,t) &= -g(N)P(N,t) - r(N)P(N,t) \\ &+ \ r(N{+}1)P(N{+}1,t) + g(N{-}1)P(N{-}1,t)\ . \qquad \text{(E.36)}\end{aligned}$$

Under stationary conditions $\mathrm{d}P/\mathrm{d}t = 0$ has to be fulfilled for all values of N, which gives the following recursive relation[2] for the stationary-state probabilites $P_0(N)$:

$$\begin{aligned}&r(N{+}1)P_0(N{+}1) + g(N{-}1)P_0(N{-}1) \\ &\qquad - [g(N){+}r(N)]P_0(N) = 0\ . \qquad \text{(E.37)}\end{aligned}$$

If N is taken as a continuous variable and if $P_0(N)$ is maximum at N_0, $\ln P_0(N)$ can be expanded around its maximum into a Taylor series of the form

$$\ln[P_0(N)] = \ln[P_0(N_0)] - \frac{1}{2\sigma^2}(N{-}N_0)^2\ , \qquad \text{(E.38)}$$

where $\sigma^2 = \langle\Delta N\rangle^2$. The probability $P_0(N)$ of finding N electrons in the conduction band thus obeys a Gaussian distribution

$$P_0(N) \approx P_0(N_0)\exp\left(-\frac{(N{-}N_0)^2}{2\sigma^2}\right)\ , \qquad \text{(E.39)}$$

as long as N is not too far from the expectation value $\langle N\rangle = N_0$. The particle density thus fluctuates around N_0 with a variance determined by

$$\frac{1}{\langle\Delta N^2\rangle} = -\frac{\mathrm{d}^2}{\mathrm{d}N^2}\ln P_0\bigg|_{N_0} = -\frac{1}{P_0(N_0)}\frac{\mathrm{d}^2P_0}{\mathrm{d}N^2}\bigg|_{N_0} = -\frac{P_0''(N_0)}{P_0(N_0)}\ , \qquad \text{(E.40)}$$

[2]For $N = 0$, we obtain, taking $r(0) = 0$ and $g(-1) = 0$,

$$r(1)P_0(1) - g(0)P_0(0) = 0 \quad \text{and therefore} \quad P_0(1) = P_0(0)\,\frac{g(0)}{r(1)}\ .$$

From this result, together with (E.37), we obtain by induction

$$P_0(N) = P_0(0)\,\frac{g(0)g(1)\cdots g(N-1)}{r(0)r(1)\cdots r(N)}\ .$$

owing to the vanishing derivative $\mathrm{d}P_0/\mathrm{d}N|_{N_0} = P_0'(N_0) = 0$ at the maximum. The second derivative of P_0 at the maximum can be obtained from (E.37) if the terms $r(N+1)P_0(N+1)$ and $g(N-1)P_0(N-1)$ are expanded up to second order, which allows us to transform (E.37) into the relation

$$0 = \left[r'(N) - g'(N) + \frac{1}{2}r''(N) + \frac{1}{2}g''(N)\right]P_0(N)$$
$$+ \left[r(N) - g(N) + r'(N) + g'(N)\right]P_0'(N_0) + \frac{r(N)+g(N)}{2}P_0''(N) ,$$

which is valid for all values of N. If N is chosen equal to N_0, we obtain

$$\frac{P_0''(N_0)}{P_0(N_0)} = -\frac{2r'(N_0) - 2g'(N_0) + r''(N_0) + g''(N_0)}{r(N_0) + g(N_0)} ,$$

since $P_0'(N_0) = 0$. Taking only terms of first order in the numerator in (E.40) yields

$$\langle \Delta N^2 \rangle = \frac{r(N_0) + g(N_0)}{2\left[r'(N_0) - g'(N_0)\right]} . \tag{E.41}$$

Deviations $\Delta P(N,t) = P(N,t) - P_0(N)$ from the stationary state obey the master equation (E.36), with $P(N,t)$ replaced by $\Delta P(N,t)$. For large values of N, we may assume

$$\Delta P(N+1,t) \approx \Delta P(N,t) \approx \Delta P(N-1,t) ;$$

if we take this relation together with

$$r(N+1) \approx r(N) + \left.\frac{\mathrm{d}r}{\mathrm{d}N}\right|_N \quad \text{and} \quad g(N-1) \approx g(N) - \left.\frac{\mathrm{d}g}{\mathrm{d}N}\right|_N ,$$

the master equation simplifies to

$$\frac{\mathrm{d}}{\mathrm{d}t}\Delta P(N,t) = \left(\left.\frac{\mathrm{d}r}{\mathrm{d}N}\right|_N - \left.\frac{\mathrm{d}g}{\mathrm{d}N}\right|_N\right)\Delta P(N,t) \approx \frac{1}{\tau}\Delta P(N,t) .$$

The relaxation time τ can be expressed in terms of the derivatives of the generation and recombination rates $G = g/V$ and $R = r/V$ with respect to the average electron density $n = N_0/V$:

$$\left.\frac{\mathrm{d}r}{\mathrm{d}N}\right|_N - \left.\frac{\mathrm{d}g}{\mathrm{d}N}\right|_N \approx \left.\frac{\mathrm{d}r}{\mathrm{d}N}\right|_{N_0} - \left.\frac{\mathrm{d}g}{\mathrm{d}N}\right|_{N_0} = \left.\frac{\mathrm{d}R}{\mathrm{d}n}\right|_{n_0} - \left.\frac{\mathrm{d}G}{\mathrm{d}n}\right|_{n_0} = \frac{1}{\tau} .$$

With this approximation, (E.41) gives

$$\langle \Delta N^2 \rangle = \frac{R+G}{2}V\tau , \tag{E.42}$$

a relation that shows explicitely that both generation and recombination contribute to the noise.

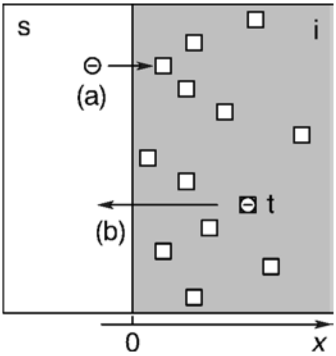


Fig. E.2. Semiconductor–insulator interface in the McWorther model. (**a**) Trapping: an electron tunnels from the semiconductor s to a trapping center t in the insulator i. (**b**) Detrapping: an electron tunnels from a trapping center in the insulator to the semiconductor

E.5 McWorther Model of 1/f Noise

The McWorther model assumes traps at various depths within the insulating layer, as illustrated in Fig. E.2. Trapping and detrapping at a center located at x within the insulating layer is slowed down because carriers have to tunnel from the semiconductor to the trap site and vice versa. If τ_0 denotes the relaxation time constant for a trap located at x_0, the following dependence may be assumed:

$$\tau(x) = \tau_0 \exp\left(\frac{x}{x_0} - 1\right) .$$

The spectral density of the fluctuations of the particle density per unit volume at x is given by

$$\frac{S_{\Delta n}}{V} = 4\langle \Delta n^2 \rangle \frac{\tau(x)}{1 + \omega^2 \tau^2(x)} .$$

From this relation, the spectral density of the fluctuations of particle density per unit area can be obtained after an integration with respect to x

$$\frac{S_{\Delta n}}{A} = 4\,\langle \Delta n^2 \rangle \int_0^\infty \frac{\tau(x)}{1 + \omega^2 \tau^2(x)}\, \mathrm{d}x .$$

If we write $\mathrm{d}\tau/\mathrm{d}x = \tau/x_0$,

$$\frac{S_{\Delta n}}{A} = 4x_0 \langle \Delta n^2 \rangle \int_{\tau_0}^\infty \frac{\mathrm{d}\tau}{1 + \omega^2 \tau^2} = \frac{4x_0 \langle \Delta n^2 \rangle}{\omega} \left(\frac{\pi}{2} - \arctan(\omega\tau_0)\right) ,$$

where the insulator is assumed to be thick in comparison with x_0, which is of the order of a few nanometers. For frequencies such that $\omega\tau_0 \ll 1$, the second term in the large paranthesis may be neglected ($\arctan(\omega\tau_0) \approx 0$), resulting in a $1/f$ dependence of the spectral density,

$$\frac{S_{\Delta n}}{A} = \frac{x_0 \langle \Delta n^2 \rangle}{f} .$$

E.6 Short-Base Diode with Metal Contact

In the case of a metal-contacted short-base diode, we require $\Delta \underline{p} = 0$ at the contact, assuming the metal contact to be able to restore thermal-equilibrium conditions without delay. This results in

$$\underline{A} = \frac{1}{\sinh(d/\underline{L}_\mathrm{p})} \frac{\underline{L}_\mathrm{p}}{D_\mathrm{p}} \int_0^d \underline{\zeta}_\mathrm{p} \sinh\left(\frac{d-x}{\underline{L}_\mathrm{p}}\right) \mathrm{d}x \ ;$$

the current $\Delta \underline{I}_\mathrm{p} \approx \Delta \underline{I}$ injected into the n-type region is therefore

$$\Delta \underline{I} = eA_\mathrm{j} \frac{1}{\sinh(d/\underline{L}_\mathrm{p})} \int_0^d \underline{\zeta}_\mathrm{p} \sinh\left(\frac{d-x}{\underline{L}_\mathrm{p}}\right) \mathrm{d}x$$

or, after integration by parts,

$$\begin{aligned}\Delta \underline{I} &= eA_\mathrm{j} \frac{1}{\sinh(d/\underline{L}_\mathrm{p})} \int_0^d \underline{\gamma} \sinh\left(\frac{d-x}{\underline{L}_\mathrm{p}}\right) \mathrm{d}x \\ &+ \frac{eA_\mathrm{j}}{\underline{L}_\mathrm{p}} \frac{1}{\sinh(d/\underline{L}_\mathrm{p})} \int_0^d \underline{\eta}_\mathrm{p} \cosh\left(\frac{d-x}{\underline{L}_\mathrm{p}}\right) \mathrm{d}x \ .\end{aligned}$$

From this relation, the spectral density $S_i(\omega) = \langle \Delta \underline{I} \, \Delta \underline{I}^* \rangle$ of the noise current can be calculated as

$$\begin{aligned}S_i(\omega) &= \frac{2e^2 A_\mathrm{j}}{\tau_\mathrm{p} |\sinh(d/\underline{L}_\mathrm{p})|^2} \int_0^d [\, p_\mathrm{n}(x) + p_\mathrm{n0}\,] \left| \sinh\left(\frac{d-x}{\underline{L}_\mathrm{p}}\right)\right|^2 \mathrm{d}x \\ &+ \frac{4e^2 A_\mathrm{j} D_\mathrm{p}}{|\underline{L}_\mathrm{p}|^2 |\sinh(d/\underline{L}_\mathrm{p})|^2} \int_0^d p_\mathrm{n}(x) \left| \cosh\left(\frac{d-x}{\underline{L}_\mathrm{p}}\right)\right|^2 \mathrm{d}x \ ,\end{aligned}$$

using the spectral densities of the noise sources. Expressing the hole density

$$p_\mathrm{n}(x) = p_\mathrm{n0} \left[\exp\left(\frac{V}{V_\mathrm{T}}\right) - 1 \right] \frac{\sinh[(d-x)/L_\mathrm{p}]}{\sinh(d/L_\mathrm{p})} + p_\mathrm{n0}$$

in terms of the injected hole current I_p yields

$$p_\mathrm{n}(x) = \frac{I_\mathrm{p}}{eA_\mathrm{j}} \frac{L_\mathrm{p}}{D_\mathrm{p}} \frac{\sinh[(d-x)/L_\mathrm{p}]}{\cosh(d/L_\mathrm{p})} + p_\mathrm{n0} \ ,$$

and p_n0 may be expressed in terms of the saturation current I_Sp as

$$p_\mathrm{n0} = \frac{I_\mathrm{Sp}}{eA_\mathrm{j}} \frac{L_\mathrm{p}}{D_\mathrm{p}} \frac{\sinh(d/L_\mathrm{p})}{\cosh(d/L_\mathrm{p})} \ .$$

With this result, the expression for the spectral density of the noise current associated with holes injected into the metal-contacted emitter region can be written as

$$S_{i\mathrm{p}}(\omega) = 2eI_{\mathrm{p}}f_{\mathrm{p}1}(\omega) + 4eI_{\mathrm{Sp}}f_{\mathrm{p}2}(\omega) \ ,$$

where

$$\begin{aligned} f_{\mathrm{p}1}(\omega) &= \frac{1}{\cosh(d/L_{\mathrm{p}})|\sinh(d/\underline{L}_{\mathrm{p}})|^2} \\ &\times \left[\frac{1}{L_{\mathrm{p}}} \int_0^d \left| \sinh\left(\frac{d-x}{\underline{L}_{\mathrm{p}}}\right) \right|^2 \sinh\left(\frac{d-x}{L_{\mathrm{p}}}\right) \mathrm{d}x \right. \\ &\quad \left. + 2\sqrt{1+\omega^2\tau_{\mathrm{p}}^2}\, \frac{1}{L_{\mathrm{p}}} \int_0^d \left| \cosh\left(\frac{d-x}{\underline{L}_{\mathrm{p}}}\right) \right|^2 \sinh\left(\frac{d-x}{L_{\mathrm{p}}}\right) \mathrm{d}x \right] \end{aligned}$$

and

$$\begin{aligned} f_{\mathrm{p}2}(\omega) &= \frac{\tanh(d/L_{\mathrm{p}})}{\left|\sinh(d/\underline{L}_{\mathrm{p}})\right|^2} \left\{ \frac{1}{L_{\mathrm{p}}} \int_0^d \left| \sinh\left(\frac{d-x}{\underline{L}_{\mathrm{p}}}\right) \right|^2 \mathrm{d}x \right. \\ &\quad \left. + \sqrt{1+\omega^2\tau_{\mathrm{p}}^2}\, \frac{1}{L_{\mathrm{p}}} \int_0^d \left| \cosh\left(\frac{d-x}{\underline{L}_{\mathrm{p}}}\right) \right|^2 \mathrm{d}x \right\} . \end{aligned}$$

By making use of the decomposition

$$\begin{aligned} \left| \sinh\left(\frac{d-x}{\underline{L}_{\mathrm{p}}}\right) \right|^2 &= \frac{1}{2} \left[\cosh\left(\frac{2a(d-x)}{L_{\mathrm{p}}}\right) - \cos\left(\frac{2b(d-x)}{L_{\mathrm{p}}}\right) \right] , \\ \left| \cosh\left(\frac{d-x}{\underline{L}_{\mathrm{p}}}\right) \right|^2 &= \frac{1}{2} \left[\cosh\left(\frac{2a(d-x)}{L_{\mathrm{p}}}\right) + \cos\left(\frac{2b(d-x)}{L_{\mathrm{p}}}\right) \right] , \end{aligned}$$

where $a = \mathrm{Re}\left(\sqrt{1+\mathrm{j}\omega\tau_{\mathrm{p}}}\right)$ and $b = \mathrm{Im}\left(\sqrt{1+\mathrm{j}\omega\tau_{\mathrm{p}}}\right)$, the integrals can be evaluated to give

$$f_{\mathrm{p}1}(\omega) = 2\tanh\left(\frac{d}{L_{\mathrm{p}}}\right) \frac{a\sinh\left(\frac{2ad}{L_{\mathrm{p}}}\right) + b\sin\left(\frac{2bd}{L_{\mathrm{p}}}\right)}{\cosh\left(\frac{2ad}{L_{\mathrm{p}}}\right) - \cos\left(\frac{2bd}{L_{\mathrm{p}}}\right)} - 1 \tag{E.43}$$

and

$$f_{\mathrm{p}2}(\omega) = \frac{a\sinh\left(\frac{d}{L_{\mathrm{p}}}\right)\sinh\left(\frac{2ad}{L_{\mathrm{p}}}\right) + b\sinh\left(\frac{d}{L_{\mathrm{p}}}\right)\sin\left(\frac{2bd}{L_{\mathrm{p}}}\right)}{2\cosh\left(\frac{d}{L_{\mathrm{p}}}\right)\left[\cosh\left(\frac{2ad}{L_{\mathrm{p}}}\right) - \cos\left(\frac{2bd}{L_{\mathrm{p}}}\right)\right]} . \tag{E.44}$$

E.7 Short-Base Diode with Polysilicon Contact

The boundary condition (5.58) results in

$$\underline{A} = -\frac{1}{\underline{\theta}_1(d)} \left(\frac{\underline{L}_\mathrm{p}}{D_\mathrm{p}} \int_0^d \underline{\zeta}_\mathrm{p}(x)\underline{\theta}_1(d-x)\,\mathrm{d}x + \frac{1}{S_\mathrm{p}}\,\underline{\kappa}_\mathrm{S} \right) , \tag{E.45}$$

where $\underline{\theta}_1(x) = \sinh(x/\underline{L}_\mathrm{p}) + \underline{\nu}\cosh(x/\underline{L}_\mathrm{p})$ and $\underline{\nu} = D_\mathrm{p}/(S_\mathrm{p}\underline{L}_\mathrm{p})$. The hole current $\underline{I}_\mathrm{p}(0) = A_\mathrm{j}\Delta\underline{J}_\mathrm{p}(0) = -eA_\mathrm{j}D_\mathrm{p}\underline{A}/\underline{L}_\mathrm{p}$ injected into the n-type region is therefore, after an integration by parts, given by the following:

$$\underline{I}_\mathrm{p}(0) = \frac{eA_\mathrm{j}}{\underline{\theta}_1(d)} \left[\int_0^d \left(\underline{\gamma}(x)\underline{\theta}_1(d-x) + \frac{\underline{\eta}_\mathrm{p}}{\underline{L}_\mathrm{p}}\,\underline{\theta}_2(d-x) \right) \mathrm{d}x + \underline{\nu}_\mathrm{p}\underline{\kappa}_\mathrm{S} \right] ,$$

where $\underline{\theta}_2(x) = \cosh(x/\underline{L}_\mathrm{p}) + \underline{\nu}\sinh(x/\underline{L}_\mathrm{p})$. Using the spectral functions

$$S_\gamma(x,x',\omega) = \frac{2\,[\,p_\mathrm{n}(x) + p_\mathrm{n0}\,]}{A_\mathrm{j}\tau_\mathrm{p}}\,\delta(x-x') ,$$

$$S_{\eta\mathrm{p}}(x,x',\omega) = \frac{4}{A_\mathrm{j}} D_\mathrm{p} p_\mathrm{n}(x)\,\delta(x-x')$$

and

$$S_\kappa(\omega) = \frac{2}{A_\mathrm{j}} S_\mathrm{p}\,[\,p_\mathrm{n}(d) + p_\mathrm{n0}\,] ,$$

the spectral density $S_{i\mathrm{p}}(\omega) = \langle \underline{I}_\mathrm{p}(0)\underline{I}_\mathrm{p}^*(0)\rangle/\Delta f$ of the noise current is obtained as

$$\begin{aligned} S_{i\mathrm{p}}(\omega) &= \frac{2e^2A_\mathrm{j}}{\tau_\mathrm{p}|\,\underline{\theta}_1(d)|^2} \int_0^d [\,p_\mathrm{n}(x) + p_\mathrm{n0}\,]\,|\,\underline{\theta}_1(d-x)|^2\,\mathrm{d}x \\ &+ \frac{4e^2A_\mathrm{j}D_\mathrm{p}}{|\underline{L}_\mathrm{p}|^2|\,\underline{\theta}_1(d)|^2} \int_0^d p_\mathrm{n}(x)\,|\,\underline{\theta}_2(d-x)|^2\,\mathrm{d}x \\ &+ \frac{2e^2|\underline{\nu}|^2A_\mathrm{j}S_\mathrm{p}}{|\,\underline{\theta}_1(d)|^2}\,[\,p_\mathrm{n}(d) + p_\mathrm{n0}\,] . \end{aligned} \tag{E.46}$$

The hole density $p_\mathrm{n}(x)$ increases in proportion to the hole current I_p injected into the n-type region and can be expressed as

$$p_\mathrm{n}(x) = \frac{L_\mathrm{p}}{D_\mathrm{p}}\,\frac{\theta_1(d-x)}{\theta_2(d)}\,\frac{I_\mathrm{p}}{eA_\mathrm{j}} + \frac{L_\mathrm{p}}{D_\mathrm{p}}\,\frac{\theta_1(d)}{\theta_2(d)}\,\frac{I_\mathrm{Sp}}{eA_\mathrm{j}} , \tag{E.47}$$

where I_Sp denotes the hole saturation current, and

$$\theta_1(x) = \sinh(x/L_\mathrm{p}) + \nu\cosh(x/L_\mathrm{p}) ,$$

$$\theta_2(x) = \cosh(x/L_\mathrm{p}) + \nu\sinh(x/L_\mathrm{p}) .$$

Use of the identities $L_\mathrm{p} = \sqrt{D_\mathrm{p}\tau_\mathrm{p}}$ and $|\underline{L}_\mathrm{p}|^2 = L_\mathrm{p}^2/\sqrt{1+\omega^2\tau_\mathrm{p}^2}$, together with (E.47) and the substitution $\xi = d-x$, allows us to transform (E.46) into

$$S_{i\mathrm{p}}(\omega) \;=\; 2eI_\mathrm{p}f_\mathrm{p1}(\omega) + 4eI_\mathrm{Sp}f_\mathrm{p2}(\omega)\;, \tag{E.48}$$

where

$$f_\mathrm{p1} \;=\; \int_0^d \frac{\theta_1(\xi)\Big[|\,\underline{\theta}_1(\xi)|^2+2\sqrt{1+\omega^2\tau_\mathrm{p}^2}\,|\,\underline{\theta}_2(\xi)|^2\Big]}{L_\mathrm{p}\,|\,\underline{\theta}_1(d)|^2\theta_2(d)}\,\mathrm{d}\xi + \frac{|\underline{\nu}|^2\theta_1(0)}{\nu|\,\underline{\theta}_1(d)|^2\theta_2(d)}$$

$$f_\mathrm{p2} \;=\; \int_0^d \frac{\theta_1(d)\Big[|\,\underline{\theta}_1(\xi)|^2+\sqrt{1+\omega^2\tau_\mathrm{p}^2}\,|\,\underline{\theta}_2(\xi)|^2\Big]}{L_\mathrm{p}\,|\,\underline{\theta}_1(d)|^2\theta_2(d)}\,\mathrm{d}\xi + \frac{|\underline{\nu}|^2\theta_1(d)}{\nu|\,\underline{\theta}_1(d)|^2\theta_2(d)}$$

Making use of

$$\underline{\nu} \;=\; \frac{D_\mathrm{p}}{S_\mathrm{p}\underline{L}_\mathrm{p}} \;=\; \nu(a+\mathrm{j}b) \qquad \text{and} \qquad \frac{\xi}{\underline{L}_\mathrm{p}} \;=\; \frac{\xi}{L_\mathrm{p}}(a+\mathrm{j}b)\;,$$

where $a = \mathrm{Re}\left(\sqrt{1+\mathrm{j}\omega\tau_\mathrm{p}}\right)$ and $b = \mathrm{Im}\left(\sqrt{1+\mathrm{j}\omega\tau_\mathrm{p}}\right)$, gives

$$\begin{aligned}|\,\underline{\theta}_{1,2}(\xi)|^2 \;&=\; \frac{\nu^2\sqrt{1+\omega^2\tau_\mathrm{p}^2}+1}{2}\cosh\left(\frac{2a\xi}{L_\mathrm{p}}\right) + \nu a\sinh\left(\frac{2a\xi}{L_\mathrm{p}}\right)\\ &\pm \frac{\nu^2\sqrt{1+\omega^2\tau_\mathrm{p}^2}-1}{2}\cos\left(\frac{2b\xi}{L_\mathrm{p}}\right) \pm \nu b\sin\left(\frac{2b\xi}{L_\mathrm{p}}\right)\;,\end{aligned}$$

where the "+" sign applies to the subscript 1 and the "−" sign to the subscript 2. If this result is substituted in the integrand that determines $f_\mathrm{p1}(\omega)$, eight integrals such as

$$\begin{aligned}&\int_0^d \sinh\left(\frac{\xi}{L_\mathrm{p}}\right)\cosh\left(\frac{2a\xi}{L_\mathrm{p}}\right)\,\mathrm{d}\xi\\ &= \frac{L_\mathrm{p}\left[2a\sinh\left(\frac{d}{L_\mathrm{p}}\right)\sinh\left(\frac{2ad}{L_\mathrm{p}}\right) - \cosh\left(\frac{d}{L_\mathrm{p}}\right)\cosh\left(\frac{2ad}{L_\mathrm{p}}\right)+1\right]}{2\sqrt{1+\omega^2\tau_\mathrm{p}^2}+1}\end{aligned}$$

have to be computed; summing the corresponding terms then gives (5.59), which is equivalent to

$$S_{i\mathrm{p}} \;=\; 4k_\mathrm{B}T\,\mathrm{Re}(y_\mathrm{p}) - 2eI_\mathrm{p}\;, \tag{E.49}$$

where y_p denotes the ac small-signal diode admittance due to hole injection into the n-type emitter region. From this solution, the result (E.43) for a metal-contacted emitter can be derived in the limit $\nu \to 0$. In the limit $\omega \to 0$ one obtains $f_\mathrm{p1} \to 1$ and the formula (E.48) reduces to the classical shot noise formula, i.e. $f_\mathrm{p1} \to 1$ as $\omega \to 0$.

E.8 Equivalent-Circuit Representation of Transfer Current Noise

The fundamental solutions of the stochastic differential equation (5.63) are

$$f_1(x) = \exp\left(\frac{(\eta/2+\underline{\vartheta})x}{d_\mathrm{B}}\right) \quad \text{and} \quad f_2(x) = \exp\left(\frac{(\eta/2-\underline{\vartheta})x}{d_\mathrm{B}}\right) ,$$

where with $\eta = -Ed_\mathrm{B}/V_\mathrm{T}$ and $\underline{\vartheta} = \sqrt{\xi^2 + d_\mathrm{B}^2/\underline{L}_n^2}$. By use of the Wronskian

$$\begin{vmatrix} f_1 & f_2 \\ f_1' & f_2' \end{vmatrix} = -\frac{2\vartheta}{d_\mathrm{B}} \exp\left(\frac{\eta x}{d_\mathrm{B}}\right) ,$$

the solution of the stochastic diffusion equation is found (e.g. using variation of constants) to be

$$\begin{aligned} \Delta\underline{n} &= \underline{A}\exp\left(\frac{\eta x}{2d_\mathrm{B}}\right)\sinh\left(\frac{\underline{\vartheta}x}{d_\mathrm{B}}\right) \\ &\quad -\frac{d_\mathrm{B}}{D_\mathrm{n}}\frac{1}{\underline{\vartheta}}\int_0^x \underline{\zeta}_\mathrm{n}(x')\exp\left(\frac{\eta(x-x')}{2d_\mathrm{B}}\right)\sinh\left(\frac{\underline{\vartheta}(x-x')}{d_\mathrm{B}}\right)\mathrm{d}x' , \end{aligned}$$

where $\Delta\underline{n}(0)$ has been assumed to be zero, in accordance with the adiabatic assumption already applied for the investigation of pn junction noise. As a second boundary condition, we take $\Delta\underline{n}(d_\mathrm{B}) = 0$, neglecting velocity saturation in the bc depletion layer,[3] to obtain

$$\underline{A} = \frac{1}{\underline{\vartheta}\sinh(\underline{\vartheta})}\frac{d_\mathrm{B}}{D_\mathrm{n}}\int_0^{d_\mathrm{B}} \underline{\zeta}_\mathrm{n}(x)\exp\left(-\frac{\eta x}{2d_\mathrm{B}}\right)\sinh\left(\frac{\underline{\vartheta}(d_\mathrm{B}-x)}{d_\mathrm{B}}\right)\mathrm{d}x .$$

When we apply the boundary conditions $\Delta\underline{n}(0) = \Delta\underline{n}(d_\mathrm{B}) = 0$, this yields (5.66) and (5.67) for the electron noise currents $\underline{I}_\mathrm{nen}$ and $\underline{I}_\mathrm{ncn}$ at the emitter- and collector-side depletion layer edges. After integration by parts, these expressions transform to

$$\begin{aligned} \underline{I}_\mathrm{nen} &= \frac{eA_\mathrm{je}}{\sinh(\underline{\vartheta})}\int_0^{d_\mathrm{B}}\left(\underline{\gamma}+\frac{\eta}{2d_\mathrm{B}}\underline{\kappa}_\mathrm{n}\right)\exp\left(-\frac{\eta x}{2d_\mathrm{B}}\right)\sinh\left(\frac{\underline{\vartheta}(d_\mathrm{B}-x)}{d_\mathrm{B}}\right)\mathrm{d}x \\ &\quad + \frac{eA_\mathrm{je}}{\sinh(\underline{\vartheta})}\frac{\underline{\vartheta}}{d_\mathrm{B}}\int_0^{d_\mathrm{B}}\underline{\kappa}_\mathrm{n}\exp\left(-\frac{\eta x}{2d_\mathrm{B}}\right)\cosh\left(\frac{\underline{\vartheta}(d_\mathrm{B}-x)}{d_\mathrm{B}}\right)\mathrm{d}x \end{aligned}$$

and

[3] For a finite value of the electron drift velocity in the bc depletion layer, the electron density and its fluctuations will not be zero. Under low-level-injection conditions, the electron density will be small, however, justifying our approximation. Under high-level-injection conditions, velocity saturation effects are expected to affect the spectral density of transfer current noise.

$$
\begin{aligned}
\underline{I}_{\mathrm{ncn}} &= \frac{eA_{\mathrm{je}}}{\sinh(\underline{\vartheta})} \int_0^{d_{\mathrm{B}}} \left(\underline{\gamma} + \frac{\eta}{2d_{\mathrm{B}}}\, \underline{\kappa}_{\mathrm{n}}\right) \exp\left(\frac{\eta(d_{\mathrm{B}}-x)}{2d_{\mathrm{B}}}\right) \sinh\left(-\frac{\underline{\vartheta} x}{d_{\mathrm{B}}}\right)\, \mathrm{d}x \\
&\quad + \frac{eA_{\mathrm{je}}}{\sinh(\underline{\vartheta})} \frac{\underline{\vartheta}}{d_{\mathrm{B}}} \int_0^{d_{\mathrm{B}}} \underline{\kappa}_{\mathrm{n}} \exp\left(\frac{\eta(d_{\mathrm{B}}-x)}{2d_{\mathrm{B}}}\right) \cosh\left(-\frac{\underline{\vartheta} x}{d_{\mathrm{B}}}\right) \mathrm{d}x\, .
\end{aligned}
$$

Expressing the sinh and cosh functions in terms of exponential functions, we may rewrite the above equations as

$$
\begin{aligned}
\underline{I}_{\mathrm{nen}} &= \frac{eA_{\mathrm{je}}}{2\sinh(\underline{\vartheta})} \int_0^{d_{\mathrm{B}}} \underline{\gamma} \left(\mathrm{e}^{\underline{\vartheta}}\, \underline{\theta}_+ - \mathrm{e}^{-\underline{\vartheta}}\, \underline{\theta}_-\right) \mathrm{d}x \\
&\quad + \frac{eA_{\mathrm{je}}}{2\sinh(\underline{\vartheta})} \frac{1}{d_{\mathrm{B}}} \int_0^{d_{\mathrm{B}}} \underline{\kappa}_{\mathrm{n}} \left[\left(\frac{\eta}{2} + \underline{\vartheta}\right) \mathrm{e}^{\underline{\vartheta}}\, \underline{\theta}_+ - \left(\frac{\eta}{2} - \underline{\vartheta}\right) \mathrm{e}^{-\underline{\vartheta}}\, \underline{\theta}_-\right] \mathrm{d}x
\end{aligned}
$$

and

$$
\begin{aligned}
\underline{I}_{\mathrm{ncn}} &= \frac{eA_{\mathrm{je}}}{2\sinh(\underline{\vartheta})}\, \mathrm{e}^{\eta/2} \int_0^{d_{\mathrm{B}}} \underline{\gamma}\, (\underline{\theta}_+ - \underline{\theta}_-)\, \mathrm{d}x \\
&\quad + \frac{eA_{\mathrm{je}}}{2\sinh(\underline{\vartheta})} \frac{\mathrm{e}^{\eta/2}}{d_{\mathrm{B}}} \int_0^{d_{\mathrm{B}}} \underline{\kappa}_{\mathrm{n}} \left[\left(\frac{\eta}{2} + \underline{\vartheta}\right) \underline{\theta}_+ - \left(\frac{\eta}{2} - \underline{\vartheta}\right) \underline{\theta}_-\right] \mathrm{d}x\, ,
\end{aligned}
\tag{E.50}
$$

where $\theta_\pm(x) = \exp[-(\eta/2 \pm \underline{\vartheta})x/d_{\mathrm{B}}]$. Taking the difference yields

$$
\begin{aligned}
\underline{I}_{\mathrm{nbn}} &= \frac{eA_{\mathrm{je}}}{2\sinh(\underline{\vartheta})} \int_0^{d_{\mathrm{B}}} \underline{\gamma}\, (a_+ \underline{\theta}_+ - a_- \underline{\theta}_-)\, \mathrm{d}x \\
&\quad + \frac{eA_{\mathrm{je}}}{2\sinh(\underline{\vartheta})} \frac{1}{d_{\mathrm{B}}} \int_0^{d_{\mathrm{B}}} \underline{\kappa}_{\mathrm{n}} \left[\left(\frac{\eta}{2} + \underline{\vartheta}\right) a_+ \underline{\theta}_+ - \left(\frac{\eta}{2} - \underline{\vartheta}\right) a_- \underline{\theta}_-\right] \mathrm{d}x\, ,
\end{aligned}
\tag{E.51}
$$

where $a_\pm = \exp(\pm\underline{\vartheta}) - \exp(\eta/2)$. The formulas (E.50) and (E.51) allow one to compute the spectral density of the transfer and base current noise associated with electron transport through the base region. From (E.50), the spectral density $\langle \underline{I}_{\mathrm{ncn}} \underline{I}_{\mathrm{ncn}}^* \rangle$ of the transfer noise current can be calculated, with the result

$$
\begin{aligned}
&\frac{\langle \underline{I}_{\mathrm{ncn}} \underline{I}_{\mathrm{ncn}}^* \rangle}{\Delta f} \\
&\quad = \frac{e^2 A_{\mathrm{je}} \mathrm{e}^{\eta}}{2\tau_{\mathrm{n}} |\sinh(\underline{\vartheta})|^2} \int_0^{d_{\mathrm{B}}} [n_{\mathrm{p}}(x) + n_{\mathrm{p0}}] \left[|\underline{\theta}_+|^2 + |\underline{\theta}_-|^2 - 2\mathrm{Re}\,(\underline{\theta}_+ \underline{\theta}_-^*)\right] \mathrm{d}x \\
&\quad + \frac{e^2 A_{\mathrm{je}} D_{\mathrm{n}} \mathrm{e}^{\eta}}{d_{\mathrm{B}}^2\, |\sinh(\underline{\vartheta})|^2} \int_0^{d_{\mathrm{B}}} n_{\mathrm{p}}(x) \Bigg[\left|\frac{\eta}{2} + \underline{\vartheta}\right|^2 |\underline{\theta}_+|^2 + \left|\frac{\eta}{2} - \underline{\vartheta}\right|^2 |\underline{\theta}_-|^2 \\
&\qquad - 2\left(\frac{\eta^2}{4} - |\underline{\vartheta}|^2\right) \mathrm{Re}(\underline{\theta}_+ \underline{\theta}_-^*) + 2\eta \mathrm{Im}(\underline{\vartheta})\, \mathrm{Im}(\underline{\theta}_+ \underline{\theta}_-^*) \Bigg] \mathrm{d}x\, .
\end{aligned}
$$

In the limit $\tau_{\mathrm{n}} \to \infty$, we obtain

$$
\begin{aligned}
&\frac{\langle \underline{I}_{\mathrm{ncn}} \underline{I}_{\mathrm{ncn}}^{*} \rangle}{\Delta f} \\
&\quad = \left| \frac{\eta}{2} + \underline{\vartheta} \right|^2 \Xi_1 + \left| \frac{\eta}{2} - \underline{\vartheta} \right|^2 \Xi_2 - 2 \left(\frac{\eta^2}{4} - |\underline{\vartheta}|^2 \right) \Xi_3 + 2\eta \, \mathrm{Im}(\underline{\vartheta}) \, \Xi_4 \, ,
\end{aligned}
$$

where

$$
\Xi_1 = \frac{e^2 A_{\mathrm{je}} D_{\mathrm{n}} \mathrm{e}^{\eta}}{d_{\mathrm{B}}^2 \, |\sinh(\underline{\vartheta})|^2} \int_0^{d_{\mathrm{B}}} n_{\mathrm{p}}(x) \, |\underline{\theta}_{+}(x)|^2 \, \mathrm{d}x \, , \tag{E.52}
$$

$$
\Xi_2 = \frac{e^2 A_{\mathrm{je}} D_{\mathrm{n}} \mathrm{e}^{\eta}}{d_{\mathrm{B}}^2 \, |\sinh(\underline{\vartheta})|^2} \int_0^{d_{\mathrm{B}}} n_{\mathrm{p}}(x) \, |\underline{\theta}_{-}(x)|^2 \, \mathrm{d}x \, , \tag{E.53}
$$

$$
\Xi_3 = \frac{e^2 A_{\mathrm{je}} D_{\mathrm{n}} \mathrm{e}^{\eta}}{d_{\mathrm{B}}^2 \, |\sinh(\underline{\vartheta})|^2} \int_0^{d_{\mathrm{B}}} n_{\mathrm{p}}(x) \, \mathrm{Re}[\underline{\theta}_{+}(x) \underline{\theta}_{-}^{*}(x)] \, \mathrm{d}x \, , \tag{E.54}
$$

$$
\Xi_4 = \frac{e^2 A_{\mathrm{je}} D_{\mathrm{n}} \mathrm{e}^{\eta}}{d_{\mathrm{B}}^2 \, |\sinh(\underline{\vartheta})|^2} \int_0^{d_{\mathrm{B}}} n_{\mathrm{p}}(x) \, \mathrm{Im}[\underline{\theta}_{+}(x) \underline{\theta}_{-}^{*}(x)] \, \mathrm{d}x \, . \tag{E.55}
$$

The electron density $n_{\mathrm{p}}(x)$ in the base region is related to the electron current injected into the bc depletion layer by

$$
n_{\mathrm{p}} = \frac{I_{\mathrm{T}}}{eA_{\mathrm{jc}}} \frac{d_{\mathrm{B}}}{D_{\mathrm{n}}} \frac{1}{\vartheta} \exp\left(-\frac{\eta(d_{\mathrm{B}} - x)}{2 d_{\mathrm{B}}} \right) \sinh\left(\frac{\vartheta(d_{\mathrm{B}} - x)}{d_{\mathrm{B}}} \right) + n_{\mathrm{p0}} \, ,
$$

as can be seen from the results derived in appendix D; if recombination in the base layer is neglected, we may make the identification $\vartheta = \underline{\vartheta}(0) = \eta/2$ to obtain

$$
n_{\mathrm{p}} = \frac{I_{\mathrm{T}}}{eA_{\mathrm{je}}} \frac{d_{\mathrm{B}}}{D_{\mathrm{n}}} \frac{1}{\eta} \left[1 - \exp\left(\frac{\eta(x - d_{\mathrm{B}})}{d_{\mathrm{B}}} \right) \right] + n_{\mathrm{p0}}
$$

If we use $x = \lambda d_{\mathrm{B}}$ and $\underline{\vartheta}$, α, β and $\tilde{\tau}$ as defined in Sect. 5.4, we may write

$$
|\theta_{\pm}(x)|^2 = \mathrm{e}^{-\eta(1 \pm \alpha)\lambda}
$$

and

$$
\underline{\theta}_{+}(x) \underline{\theta}_{-}^{*}(x) = \mathrm{e}^{-\eta\lambda} \left[\cos(\eta\beta\lambda) - \mathrm{j} \sin(\eta\beta\lambda) \right] \, .
$$

Neglecting the contributions that stem from the thermal-equilibrium minority-carrier density n_{p0}, the integrals Ξ_{α} can then be evaluated, with the result

$$
\Xi_1 \approx \frac{eI_{\mathrm{T}}}{\eta^2 \, |\sinh(\underline{\vartheta})|^2} \frac{1}{1+\alpha} \left(\mathrm{e}^{\eta} - 1 - \frac{1}{\alpha} \left(1 - \mathrm{e}^{-\eta\alpha} \right) \right) \, ,
$$

$$
\Xi_2 \approx \frac{eI_{\mathrm{T}}}{\eta^2 \, |\sinh(\underline{\vartheta})|^2} \frac{1}{1-\alpha} \left(\mathrm{e}^{\eta} - 1 + \frac{1}{\alpha} \left(1 - \mathrm{e}^{\eta\alpha} \right) \right) \, ,
$$

$$\Xi_3 \approx \frac{eI_{\mathrm{T}}}{\eta^2\,|\sinh(\underline{\vartheta})|^2}\,\frac{1}{1+\beta^2}\left(\mathrm{e}^{\eta}-\cos(\eta\beta)-\frac{1}{\beta}\,\sin(\eta\beta)\right)$$

$$\Xi_4 \approx \frac{eI_{\mathrm{T}}}{\eta^2\,|\sinh(\underline{\vartheta})|^2}\,\frac{1}{1+\beta^2}\left(\sin(\eta\beta)-\beta(\mathrm{e}^{\eta}-1)-\frac{1}{\beta}\,[\cos(\eta\beta)-1]\right)\,.$$

Summing, we arrive at (5.74).

In the limit $\tau_{\mathrm{n}} \to \infty$, the spectral density of the electron noise current injected into the base region across the eb depletion layer is

$$\begin{aligned}\frac{\langle \underline{I}_{\mathrm{nen}}\underline{I}_{\mathrm{nen}}^{*}\rangle}{\Delta f} &= \left|\frac{\eta}{2}+\underline{\vartheta}\right|^2 \mathrm{e}^{\eta\alpha}\mathrm{e}^{-\eta}\,\Xi_1 + \left|\frac{\eta}{2}-\underline{\vartheta}\right|^2 \mathrm{e}^{-\eta\alpha}\mathrm{e}^{-\eta}\,\Xi_2 \\ &\quad -2\left[\left(\frac{\eta^2}{4}-|\underline{\vartheta}|^2\right)\cos(\eta\beta)-\eta\,\mathrm{Im}(\underline{\vartheta})\sin(\eta\beta)\right]\mathrm{e}^{-\eta}\,\Xi_3 \\ &\quad +2\left[\left(\frac{\eta^2}{4}-|\underline{\vartheta}|^2\right)\sin(\eta\beta)+\eta\,\mathrm{Im}(\underline{\vartheta})\cos(\eta\beta)\right]\mathrm{e}^{-\eta}\,\Xi_4\end{aligned}$$

if n_{p0} is taken to be zero for simplicity as above. Computation of the integrals results in (5.75).

E.9 References

[1] A. van der Ziel. *Noise – Sources, Characterization, Measurement.* Prentice-Hall, Englewood Cliffs, New Jersey, 1970.

[2] T.A. Welton, H.B. Callen. Irreversibility and generalized noise. *Phys. Rev.*, 83(1):34–40, 1951.

[3] J.M Richardson. Noise in driven systems. *IRE Trans. Inf. Theory*, (March):62–65, 1955.

[4] M.S. Gupta. Thermal fluctuations in driven nonlinear resistive systems. *Phys. Rev. A*, 18(6):2725–2731, 1978.

[5] M.S. Gupta. Thermal noise in nonlinear resistive devices and its circuit applications. *Proc. IEEE*, 70(8):788–804, 1982.

[6] J.B. Gunn. Thermodynamics of nonlinearity and noise in diodes. *J. Appl. Phys.*, 39(12):5357–5361, 1968.

[7] R.L. Stratonovich. Fluctuation–dissipation models of a nonlinear resistor and Gupta's formulas. *Radiofizika*, translation by Plenum, 1988, 31(2):220–230, 1986.

[8] J.L. Wyatt Jr., G.J. Coram. Nonlinear device noise models: satisfying the thermodynamic requirements. *IEEE Trans. Eectron Devices*, 46(1):184–193, 1999.

[9] N.G. van Kampen. *Stochastic Processes in Physics and Chemistry.* North-Holland, Amsterdam, 1992.

[10] Y.V. Mamontov, M. Willander. Accounting thermal noise in mathematical models of quasi-homogeneous regions in silicon diodes. *IEEE Trans. CAD*, 14(7):815–823, 1995.

[11] L. Weiss, W. Mathis. A thermodynamical approach to noise in non-linear networks. *Int. J. Circuit Theory Appl.*, 26:147–165, 1998.

F Overtemperature Developed During Electrostatic Discharges

A temperature gradient ∇T will induce a heat flux $\boldsymbol{q} = -\lambda \nabla T$ proportional to ∇T and to the thermal conductivity λ. A heat flux is a transport of thermal energy and may, in conjunction with the internal power dissipation per unit volume $p'(\boldsymbol{x}, t)$, cause a change of the local temperature $T(\boldsymbol{x}, t)$. This is described by by the heat equation

$$\rho_m c_\mathrm{p} \frac{\partial T}{\partial t} - \nabla \cdot (\lambda \nabla T) \;=\; p'(\boldsymbol{x}, t) \; . \tag{F.1}$$

F.1 Thermal Conductivity

The thermal conductivity λ is a function of temperature and is determined by various physical processes. For relatively pure semiconductors and temperatures well below the melting point, heat is conducted almost exclusively by lattice vibrations (phonons). Only at very high temperatures do mobile carriers and photons provide a substantial contribution to the heat transport in the crystal. The thermal conductivity of a highly doped semiconductor might be expected to be higher than that of the pure material because of the large number of electrons available for transporting heat. Measurements indicate, however, that the thermal conductivity is lowered at high doping levels by as much as 30% [1].

The thermal conductivity of silicon is temperature-dependent (Fig. F.1). Experimental observations indicate that the thermal conductivity of monocrystalline silicon decreases with temperature for $T > 20$ K: the thermal conductivity first falls somewhat faster than $1/T$ until a temperature of about 200 K is reached, and then varies approximately as $1/T$ up to the melting point. In the range above 100 K it is possible to approximate the temperature dependent thermal conductivity by

$$\lambda(T) \;\approx\; 3130 \, \frac{\mathrm{W}}{\mathrm{cm\,K}} \, \frac{1}{(T/\mathrm{K})^{4/3}}$$

or by the relation[1]

$$\lambda(T) \;\approx\; 350 \, \frac{\mathrm{W}}{\mathrm{cm\,K}} \, \frac{\mathrm{K}}{T - 75\,\mathrm{K}} \; ,$$

which is helpful in analytical investigations of the heat flow equation [4].

[1] Alternative approximations are given in [3].

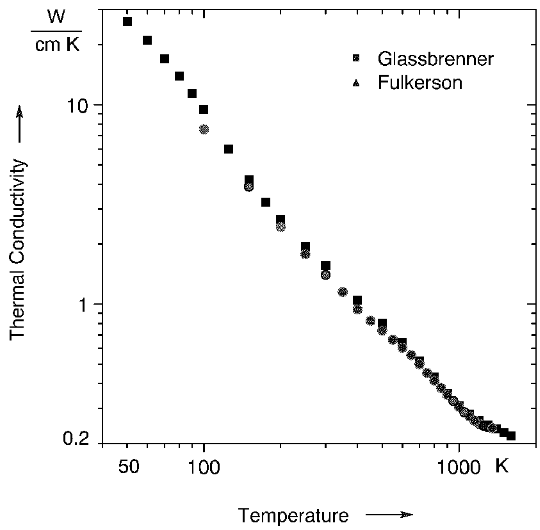


Fig. F.1. Thermal-conductivity data measured by Glassbrenner and Slack [2] and Fulkerson et al. [3]

The thermal conductivity of deposited polysilicon films has been determined in [5] using an inhomogeneously doped polysilicon microbridge. The values of λ obtained for heavily doped polysilicon range from 0.30 to 0.35 W/(cm K), with no temperature dependence found in the range between 30°C and 150°C. The reported value is about 20% to 25% of the value for monocrystalline silicon.

F.2 Transient Overtemperature During a Short Pulse

For an approximate computation of the overtemperature developed during an electrostatic discharge, we solve the time-dependent heat equation (F.1), with the assumption of a constant thermal conductivity λ. Assuming $\lambda = \text{const.}$, (F.1) may be transformed to

$$\frac{1}{\alpha}\frac{\partial T}{\partial t} - \nabla^2 T = \frac{1}{\lambda}p' , \quad \text{where} \quad \alpha = \frac{\lambda}{\rho_\mathrm{m} c_\mathrm{p}} .$$

We need to determine $p'(\boldsymbol{x}, t)$ for stress conditions corresponding to the human-body model. For this purpose, we assume that the total power dissipated in the device generates heat, and write

$$p'(\boldsymbol{x}, t) \approx BV_\mathrm{EBO}\, i(t) f(\boldsymbol{x})/\mathcal{V} ,$$

where $\mathcal{V}$ denotes the volume of the discharge and $f(\boldsymbol{x})$ is assumed to be normalized to one[2] ($\int f(\boldsymbol{x})\,\mathrm{d}^3x = 1$). The current $i(t)$ that flows through the device during a discharge described by the human-body model can be obtained from $i(t) = C\,\mathrm{d}v_{\mathrm{C}}/\mathrm{d}t$ and gives

$$p'(t) \;\approx\; \frac{BV_{\mathrm{EBO}}(V_{\mathrm{C}}(0)-BV_{\mathrm{EBO}})}{R\mathcal{V}}\,\mathrm{e}^{-t/\tau}\;,$$

where $\tau = RC$ denotes the time constant of the discharge. In addition to this external time constant, an internal time constant $\tau^* = \zeta^2/\alpha$, where ζ denotes a typical measure of the dimensions of the heat source (the spatial extent of the discharge),[3] determines the temperature response of the system. By normalizing the temperature with respect to the initial temperature T_0, the time with respect to τ^* and the spatial dimensions with respect to ζ, we transform the heat equation into the dimensionless form

$$\frac{\partial\tilde{T}}{\partial\sigma} - \nabla_{\xi}^2\tilde{T} \;=\; \tilde{Q}\;, \tag{F.2}$$

where $\tilde{T} = T/T_0$, $\boldsymbol{\xi} = \boldsymbol{x}/\zeta$, $\sigma = t/\tau^*$ and $\tilde{Q} = (p'\zeta^2)/(\lambda T_0)$. Using the Green's function of the heat equation

$$G(\boldsymbol{\xi}-\boldsymbol{\xi}',\sigma-\sigma') \;=\; \Theta(\sigma-\sigma')\,\frac{1}{8\pi^{3/2}(\sigma-\sigma')^{3/2}}\,\exp\left(-\frac{|\boldsymbol{\xi}-\boldsymbol{\xi}'|^2}{4(\sigma-\sigma')}\right)\;,$$

we find the solution to (F.2) to be

$$\tilde{T}(\boldsymbol{\xi},\sigma) \;=\; 1 + \int_0^{\sigma}\int_{\mathcal{V}}\tilde{Q}(\boldsymbol{\xi}',\sigma')\,G(\boldsymbol{\xi}-\boldsymbol{\xi}',\sigma-\sigma')\,\mathrm{d}^3\xi'\,\mathrm{d}\sigma'\;.$$

If the coordinate system is chosen such that the maximum temperature occurs at the origin ($\boldsymbol{\xi} = \boldsymbol{0}$), the normalized overtemperature at this point is

$$\Delta\tilde{T}(\sigma) \;=\; \tilde{Q}(0)\mathrm{e}^{-\eta\sigma}\int_0^{\sigma}(\sigma')^{-3/2}\Lambda(\sigma')\mathrm{e}^{\eta\sigma'}\,\mathrm{d}\sigma'\;, \tag{F.3}$$

where

$$\Lambda(\sigma) \;=\; \frac{1}{8\pi^{3/2}}\int_{\mathcal{V}} f(\boldsymbol{\xi}')\exp\left(-\frac{|\boldsymbol{\xi}|^2}{4\sigma}\right)\,\mathrm{d}^3\xi' \tag{F.4}$$

and $\eta = \tau^*/\tau$ denotes the ratio of the two time constants characterizing the system. If the maximum temperature occurs at $\sigma = \sigma_{\mathrm{m}}$, the condition

$$\left.\frac{\partial}{\partial\sigma}\Delta\tilde{T}(\sigma)\right|_{\sigma_{\mathrm{m}}} \;=\; -\eta\,\Delta\tilde{T}(\sigma_{\mathrm{m}}) + \tilde{Q}(0)\sigma_{\mathrm{m}}^{-3/2}\Lambda(\sigma_{\mathrm{m}}) \;=\; 0$$

[2]This separation is an idealization, since a possible time-dependent redistribution of the heat generation pattern is excluded.

[3]Since temperature relaxation is determined essentially by the minimum spatial dimension D of the heat source, the choice $\zeta = D$ is generally adequate.

must be fulfilled. This yields the following for the maximum overtemperature:

$$\Delta T(\sigma_{\mathrm{m}}) = \Delta T_0 \sigma_{\mathrm{m}}^{-3/2} A(\sigma_{\mathrm{m}}) \,, \tag{F.5}$$

where $\Delta T_0 = T_0 \tilde{Q}(0)/\eta$ and σ_{m} is determined by the equation

$$\int_0^{\sigma_{\mathrm{m}}} \sigma^{-3/2} A(\sigma) \mathrm{e}^{\eta\sigma} \,\mathrm{d}\sigma = \frac{1}{\eta \sigma_{\mathrm{m}}^{3/2}} A(\sigma_{\mathrm{m}}) \,\mathrm{e}^{\eta\sigma_{\mathrm{m}}} \,. \tag{F.6}$$

For a given geometry, the value of σ_{m} is a function only of the ratio η. In the following, we illustrate the dependence on the geometry by investigation of the special case of a cuboid heat source.

Homogeneous Cuboid Heat Source

As an example, we consider a homogeneous heat source of length L, width B and thickness D as a model for a discharge homogeneously distributed over the space charge layer. If $D \ll B, L$, the thickness D should be chosen as the characteristic length, resulting in

$$A(\sigma) = \sigma^{3/2} \operatorname{erf}\left(\frac{1}{4\sqrt{\sigma}}\right) \operatorname{erf}\left(\frac{B/D}{4\sqrt{\sigma}}\right) \operatorname{erf}\left(\frac{L/D}{4\sqrt{\sigma}}\right) \,.$$

This relation can be used for the determination of σ_{m}. For a given value of B/D and L/D, the value of σ_{m} is a function of η only and the correction factor $F_{B/D,\,L/D} = \sigma_{\mathrm{m}}^{-3/2} A(\sigma_{\mathrm{m}})$ is obtained as

$$F_{B/D,\,L/D} = \operatorname{erf}\left(\frac{1}{4\sqrt{\sigma_{\mathrm{m}}}}\right) \operatorname{erf}\left(\frac{B/D}{4\sqrt{\sigma_{\mathrm{m}}}}\right) \operatorname{erf}\left(\frac{L/D}{4\sqrt{\sigma_{\mathrm{m}}}}\right) \,.$$

The dependence of $F_{B/D,\,L/D}$ on η is shown in Fig. 7.27 for different values of B/D and L/D.

F.3 References

[1] P.D. Maycock. Thermal conductivity of silicon, germanium, III–V compounds and III–V alloys. *Solid-State Electron.*, 10:161–168, 1967.

[2] C.J. Glassbrenner, G.A. Slack. Thermal conductivity of silicon and germanium from 3 K to the melting point. *Phys. Rev.*, 134(4A):A1058–A1069, 1964.

[3] W. Fulkerson, R.K. Williams, R.S. Graves, D.L. McElroy, J.P. Moore. Thermal conductivity, electrical resistivity and Seebeck coefficient of silicon from 100 K to 1300 K. *Phys. Rev.*, 167(3):765–782, 1968.

[4] E.L. Heasell. The heat-flow problem in silicon: an approach to an analytical solution with application to the calculation of thermal instability in bipolar devices. *IEEE Trans. Electron Devices*, 25(12):1382–1388, 1978.

[5] Y.C. Tai, C.H. Mastrangelo, R.S. Muller. Thermal conductivity of heavily doped LPCVD polysilicon. *IEDM Tech. Dig.*, pp. 278 281, 1987.

Index