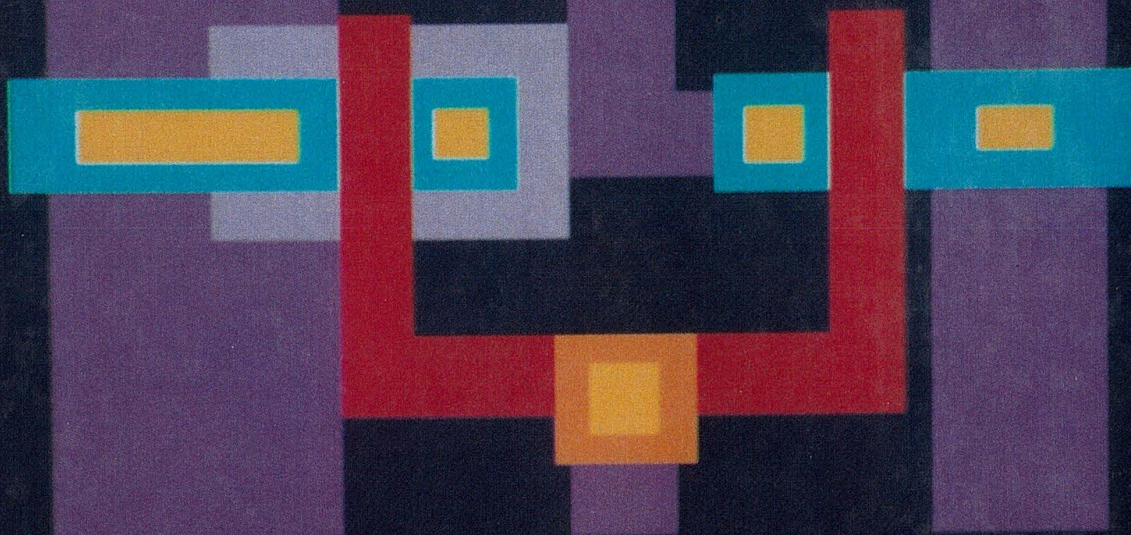


Silicon Processing

for the VLSI Era
Volume 2 - Process Integration



S. Wolf

THE COMPANION VOLUMES TO THIS BOOK

Silicon Processing

for the VLSI Era

Volume 1 - Process Technology

By **STANLEY WOLF** and **RICHARD N. TAUBER**

TABLE of CONTENTS - **Ch.1** Silicon: Single-Crystal Growth and Wafering; **Ch. 2** Crystalline Defects, Thermal Processing, and Gettering; **Ch. 3** Vacuum Technology for VLSI Applications; **Ch. 4** Basics of Thin Films; **Ch. 5** Silicon Epitaxial Film Growth; **Ch. 6** Chemical Vapor Deposition of Amorphous and Polycrystalline Films; **Ch. 7** Thermal Oxidation of Silicon; **Ch. 8** Diffusion in Silicon; **Ch. 9** Ion Implantation for VLSI; **Ch. 10** Aluminum Thin Films and Physical Vapor Deposition in VLSI; **Ch. 11** Refractory Metals and Their Silicides in VLSI; **Ch. 12** Lithography I: Optical Photoresists - Material Properties and Process Technology; **Ch. 13** Lithography II: Optical Aligners and Photomasks; **Ch. 14** Advanced Lithography: E-Beam and X-Ray; **Ch. 15** Wet Processing: Cleaning & Etching; **Ch. 16** Dry Etching for VLSI; **Ch. 17** Material Characterization Techniques for VLSI Fabrication; **Ch. 18** Structured Approach to Design of Experiments for Process Optimization.

1986 LATTICE PRESS 660 pp. ISBN 0-9616721-3-7

Volume 3 - The Submicron MOSFET

By **STANLEY WOLF**

TABLE OF CONTENTS - **Ch. 1** Role of Process & Device Models in Microelectronics Technology; **Ch. 2** Numerical Methods for Solving the Partial Differential Equations which Model Submicron Devices and Processes; **Ch. 3** Basic MOS Physics & MOS Capacitors; **Ch. 4** Long-Channel MOSFETs; **Ch. 5** The Submicron MOSFET; **Ch. 6** Isolation Structures in CMOS; **Ch. 7** Thin Gate Oxides: Growth & Reliability; **Ch. 8** Well Formation in CMOS; **Ch. 9** Hot-Carrier Resistant Processing & Device Structures.

1995 LATTICE PRESS 722 pp. ISBN 0-9616

Order by using form on last page of book

ISBN 0-9616721-4-5



Silicon Processing
for the VLSI Era

Volume 2
Process
Integration

S. Wolf

TK7874
.W831
V.2

LATTICE
PRESS

SILICON PROCESSING
FOR
THE VLSI ERA
VOLUME 2:
PROCESS INTEGRATION

**SILICON PROCESSING
FOR
THE VLSI ERA**

**VOLUME 2:
PROCESS INTEGRATION**

STANLEY WOLF Ph.D.
Professor, Department of Electrical Engineering
California State University, Long Beach
Long Beach, California

LATTICE PRESS

Sunset Beach, California

DISCLAIMER

This publication is based on sources and information believed to be reliable, but the authors and Lattice Press disclaim any warranty or liability based on or relating to the contents of this publication.

Published by:

Lattice Press,
Post Office Box 340
Sunset Beach, California 90742, U.S.A.

Cover design by Roy Montibon, Visionary Art Resources, Inc., Santa Ana, CA.

Copyright © 1990 by Lattice Press.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system without written permission from the publisher, except for the inclusion of brief quotations in a review.

Library of Congress Cataloging in Publication Data
Wolf, Stanley

Silicon Processing for the VLSI Era
Volume 1 : Process Integration

Includes Index

1. Integrated circuits-Very large scale
integration. 2. Silicon. I. Title

86-081923

ISBN 0-961672-4-5

9 8 7 6 5

PRINTED IN THE UNITED STATES OF AMERICA

5.6 HOT-CARRIER EFFECTS IN MOSFETs

If device dimensions are reduced and the supply voltage remains constant (e.g., 5 V), the lateral electric field generated in MOS devices increases. If the electric field becomes strong enough, it can give rise to so-called *hot-carrier* effects in MOS devices. This has indeed become a significant problem in NMOS devices with channel lengths smaller than $1.5\ \mu\text{m}$ (and in PMOS devices with submicron channel lengths).³⁰ Hot-electron effects are more severe than hot-hole effects because of the higher electron mobility. Therefore, we begin our discussion by considering hot-electron effects in *n*-channel devices. At the end of this section we will also discuss the impact of hot-carrier effects on *p*-channel devices.

The maximum electric field, E_M , in a MOSFET occurs near the drain during saturated operation. A rigorous calculation of the field near the drain is a complex procedure, requiring a computer-aided solution of the two-dimensional Poisson equation, with the results of one such analysis being shown in Fig. 5-26. Nevertheless, the value of E_M can be estimated from³²

$$E_M = (V_{DS} - V_{Dsat}) / m \quad (5-19)$$

where

$$m = 0.22\ t_{ox}^{1/3} / x_j'^{1/2} \quad (5-19a)$$

and t_{ox} is the gate oxide thickness and x_j' approximately corresponds to the source/drain junction depth. Although V_{Dsat} depends on L_{eff} , the dependence is weaker than a linear

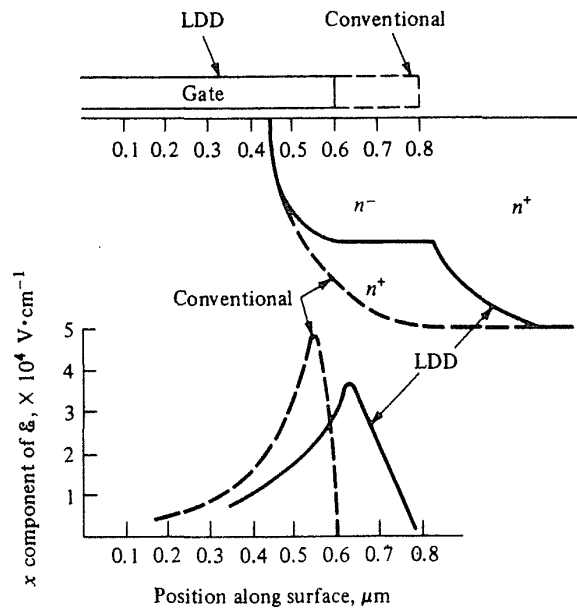


Fig. 5-26 Magnitude of the electric field at the Si-SiO₂ interface as a function of distance: $L = 1.2\ \mu\text{m}$, $V_{DS} = 8.5\ \text{V}$, $V_{GS} = V_T$.³⁸ (© 1980 IEEE).

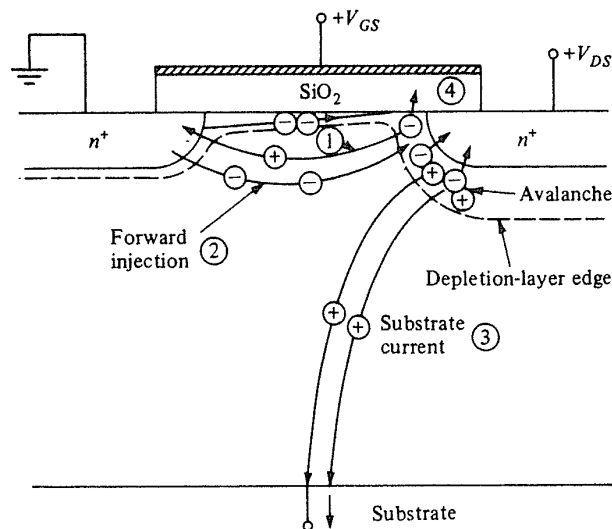


Fig. 5-27 Hot-carrier generation and current components. ①- Holes reaching the source. ②- Electron injection from the source. ③- Substrate hole current. ④- Electron injection into the oxide.

relationship (especially if $L_{eff} > 1 \mu m$). Thus, we can see that E_M increases as device dimensions shrink, but this is due to thinner gate oxides and shallower junctions, as well as to the reduction in L_{eff} .

Regardless of the factors that increase their magnitude, such high electric fields cause the electrons in the channel to gain kinetic energy and become "hot" (i.e., their energy distribution is shifted to a much higher value than that of electrons which are in thermal equilibrium with the lattice). Such hot electrons (which become hot near the drain edge of the channel, since that is where E_M exists) can cause several effects in the device. First, those electrons that acquire ≥ 1.5 eV of energy can lose it via *impact ionization*, which generates electron-hole pairs. The total number of electron-hole pairs generated by impact ionization is *exponentially dependent on the reciprocal of the electric field*, $\sim 1/E_M$. In the extreme, this electron-hole pair generation can lead to a form of *avalanche breakdown* (Process 1, shown in Fig. 5-27). Second, the hot holes and electrons can overcome the potential energy barrier between the silicon and the SiO_2 (~ 3.1 eV), thereby causing *hot carriers to become injected into the gate oxide*. Each of these events brings about its own set of repercussions.

5.6.1 Substrate Currents Due to Hot Carriers

When electron-hole pairs are created by impact ionization, the electrons are normally attracted to the drain, and they add to the drain current. The *holes*, on the other hand,

enter the substrate and constitute a part of the parasitic substrate current (*Process 3* in Fig. 5-27). This substrate current, I_{sub} , can itself produce several problems:

- If a substrate bias-generator circuit is included on-chip, the output of the bias generator will be less negative with increasing substrate current.
- If some of these holes are collected by the source (instead of by the body contact), *and* this collected hole current causes a voltage drop in the substrate material on the order of 0.6 V, the substrate-source *pn* junction will conduct significantly. Electrons will then be injected from the source to the substrate, just like electrons injected from emitter to base of an *npn* transistor (the *forward injection* shown in Fig. 5-27). These electrons can, in turn, gain sufficient energy as they travel toward the drain to cause additional impact ionization and create new electron-hole pairs. A positive-feedback mechanism thus exists, one that can sustain itself if the drain voltage exceeds a certain value. This is observed externally as a form of breakdown, referred to as a *snapback breakdown*. A particularly clear explanation of this effect, including the reason for the negative-resistance, or "snapback," portion of the curves, is given in reference 3.
- As some of the holes are accelerated through the drain-substrate depletion region, they may acquire enough energy to cause secondary impact ionization. This will create electrons far from the drain region. Some these electrons, instead of being collected by the drain, may escape the drain field and instead travel (sometimes hundreds of microns) to other nodes on the chip. This may lead to a reduction of the storage time of dynamic circuit nodes in DRAMs (i.e., manifested as a degradation of the refresh time).³³ This excess electron current is reported to be around 10^{-4} times smaller than the substrate ionization impact current itself.
- I_{sub} may induce latchup in CMOS circuits.

The magnitude of the substrate current depends exponentially on the value of E_M . An example of how I_{sub} depends on decreasing channel length is seen in Fig. 5-28.³⁴ This shows the maximum I_{sub} generated at a voltage of 5 V versus the effective channel length for MOS transistors processed with the same technology. The magnitude of I_{sub} would increase even more rapidly with shrinking L_{eff} if the oxide thickness and junction depth were scaled together with the channel length.

5.6.2 Hot-Carrier Injection into the Gate Oxide

The hot holes and electrons that are injected into the gate oxide cause another set of problems. First, some of these carriers pass to the gate electrode (mostly hot holes) and thus constitute a gate current, I_G , typically in the fA (10^{-15} A) range. For higher, but still nondestructive, biases, the gate current can grow rapidly to become several pA (10^{-12} A).

However, some fraction of the hot carriers injected into the gate oxide do not reach the gate electrode. This is because the gate oxide contains empty electron states (also known as *traps*), which can be filled by the injected hot carriers. Such occupied traps generally become electron traps, even if they are initially filled with holes. As a result, there is a negative charge density caused by the trapping of the hot carriers in the oxide. This *trapped charge* acts like a contribution to the fixed oxide charge term in the expression for the device threshold voltage. Furthermore, the trapped charge accumulates with time. Due to the polarity of the trapped charge, the resulting shift in the *n*-channel device threshold voltage is *positive* (i.e., V_T increases in NMOS devices). If I_G becomes of the order of pA, the trapping of a fraction of the electrons injected into the oxide can become a significant effect (the fraction that gets trapped is ~ 1 in 10^6 injected electrons). As a result, this increase in *negative* stored charge can lead to a permanent change in V_T in the MOSFET. (The means by which charge in the oxide impacts V_T was described earlier.)

EXAMPLE 5-6: V_T Shift due to Hot-Electron Effects. Given a short-channel NMOS device with $t_{ox} = 20$ nm, a gate *width* of $5\ \mu\text{m}$, and a gate current of 1 nA that is momentarily caused by hot carrier injection. The current flows through a region of the gate oxide near the drain end which is $5\ \mu\text{m}$ wide, and $0.2\ \mu\text{m}$ long. Assume that 1 in 10^6 electrons of the gate current becomes trapped at an average distance of $0.1\ t_{ox}$ from the Si-SiO₂ interface. How long does the gate current have to flow to change V_T by 0.1 V in the region where the injection is occurring?

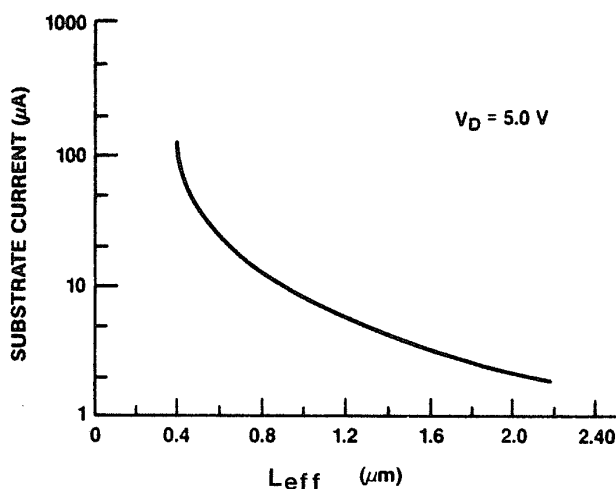


Fig. 5-28 The maximum substrate current due to impact ionization produced at a drain voltage of 5 V vs effective channel length for 250Å gate oxide transistors.³⁴ (© 1986 IEEE).

where x_C is the thickness of the epi layer from the metallurgical C-B junction to the edge of the epi/buried-layer boundary.

The increase in base width leads to an increase in GN_B , and hence to a decreased current gain at high collector currents. In addition, the larger base width reduces f_T and the transistor speed. The transistor doping profiles must thus be optimized with this effect in mind. Once a transistor structure has been designed, operation at current levels above J_{CK} for the device should be avoided.

From the equation used to calculate J_{CK} , it is seen that this value can be raised by increasing the doping in the collector. Unfortunately, this also produces a decrease in the BV_{CBO} and BV_{CEO} of the device, as well as an increase in the collector-base junction capacitance, C_{CB} . Since transistors cannot be fabricated with smaller breakdown voltages than those required by the system power-supply voltage levels, the breakdown voltage constraints will set the maximum value of N_{epi} .

On the other hand a tradeoff between J_{CK} and C_{CB} can be made, based on the operating point of the transistor. That is, if the transistor is to be operated at low current densities, the speed will be limited by C_{CB} . Thus, C_{CB} should be minimized (by reducing N_{epi}), since the Kirk effect will not impact the speed performance. However, for high-current applications, the speed will be limited by high-level injection effects, and C_{CB} can thus be allowed to rise when N_{epi} is increased in order to minimize speed degradation due to the Kirk effect.

Finally, it should be noted that in digital ICs, where BV_{CEO} is smaller than in analog ICs, a larger value of collector doping can be tolerated. This also implies that higher collector currents can be sustained in such devices without the onset of the Kirk effect.

7.3.4 Operating-Voltage Limits in Bipolar Transistors

The mechanism of avalanche breakdown in a pn junction limits the maximum reverse-bias voltages that can be applied to a junction and is also responsible for some of the maximum operating voltage values in bipolar transistors. The breakdown voltage of the C-B junction with $I_E = 0$ (BV_{CBO}) is due to this mechanism. The doping on the lightly doped side of the collector typically determines BV_{CBO} (except in cases of *reachthrough*, as will be discussed in section 7.2.4.1).

In C-B junctions in planar bipolar transistors, however, the pn junction is formed by diffusion into the silicon at openings in the surface SiO_2 layer. Thus, impurities will diffuse downward and sideways under the edges of the SiO_2 layer. Hence, the junction has a plane region with nearly cylindrical edges (as can be seen in Fig. 7-4c). Such junction curvature enhances the electric field in the curved part of the depletion region (see chap. 5, Fig. 5-10a), which reduces the breakdown voltage below that predicted by one-dimensional junction theory. Figure 5-10a also shows that as the junction depths get shallower (making the radius of curvature, r_j , of the spherical and cylindrical structures smaller), the breakdown voltage is significantly reduced, especially for small substrate-impurity concentrations. Thus, the measured value of BV_{CBO} is less than the value of the breakdown voltage that would be predicted if the junction were only one-

dimensional (as would be the case if the transistor had a structure like that shown in Fig. 7-1a).

The breakdown voltage of the E-B junction with $I_C = 0$ (BV_{EBO}) is also limited by avalanche breakdown. Since the emitter is so heavily doped, it is the base region that is the lightly doped side of this junction. Since the base is more heavily doped than the collector, BV_{EBO} is much smaller than BV_{CBO} and is typically 6 – 8 V. Since the base is formed by diffusion from the surface, the doping in the base region is highest there so the E-B junction will undergo breakdown first in this area. One of the reasons for performing a drive-in step following predeposition when a diffusion step is used to form the base is to reduce the surface concentration of boron, which thus increases BV_{EBO} to a sufficiently high value.

While high doping in the base region adjacent to the emitter sidewalls is advantageous for reducing the base series resistance, R_B (see section 7.5.2.6), if the doping is too high, BV_{EBO} becomes too small. Tradeoffs must therefore be made in the selection of the doping value in this region. The curvature of the junction enhances the electric field and reduces BV_{EBO} in the same manner as described for BV_{CBO} .

For a bipolar transistor with a voltage applied between the emitter and collector and in which $I_B = 0$, breakdown occurs at BV_{CEO} . In this situation, the breakdown may be due either to avalanche breakdown multiplied by the effect of transistor action, or to punchthrough.

In the first instance, the E-B junction will be very slightly forward-biased as a result of the base region's acquiring a potential that is intermediate between the potentials of the collector and emitter. Thus, the collector current will also be greater than I_{CBO} (the leakage current of the C-B junction when $I_E = 0$) because most of the carriers injected from the slightly forward-biased E-B junction will be collected by the collector. This increased current will precipitate avalanche breakdown at a smaller applied C-B voltage, which implies that BV_{CEO} must be smaller than BV_{CBO} . An empirical expression has been derived that relates BV_{CEO} and BV_{CBO} (plane):¹¹

$$BV_{CEO} = \frac{BV_{CBO}(\text{plane})}{(\beta)^{\frac{1}{n}}} \quad (7 - 13)$$

where BV_{CBO} (plane) is the collector-base breakdown voltage along the planar (flat) portion of the base region. The empirical parameter n has been found to have values of between 3 and 6.

7.3.4.1 Reachthrough Breakdown. Since most bipolar transistors used in ICs are fabricated with collector regions that consist of a lightly doped epitaxial layer over a heavily doped buried layer, BV_{CBO} may in some cases be limited by *reachthrough breakdown*. If the epitaxial-layer thickness, t_{epi} , is too small, under increasing reverse-bias voltage the depletion region of the C-B junction will entirely penetrate (or "reach through") the epitaxial layer before the predicted BV_{CBO} voltage has been reached. If the voltage is increased beyond this point, the depletion region must enter the heavily doped buried layer, resulting in BV_{CBO} values smaller than those

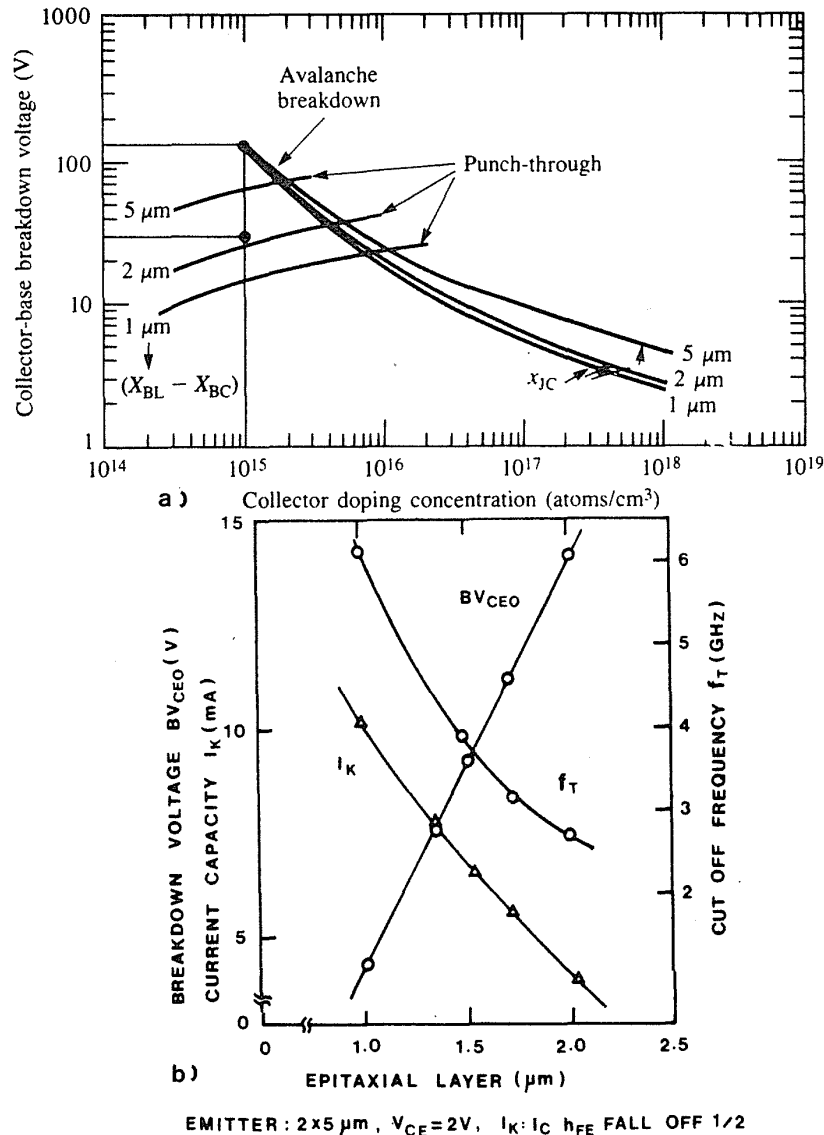


Fig. 7-10 (a) Collector-base-junction breakdown voltage as a function of collector-doping concentration with collector-base junction depth and punchthrough limits as parameters.¹²⁵ Reprinted by permission of the publisher, The Electrochemical Society, Inc. (b) Epitaxial layer thickness dependence of cutoff frequency, f_T , breakdown voltage BV_{CEO} , and knee current, I_K .¹² (© 1987 IEEE).

predicted merely by the doping in the epi layer. Figure 7-10a gives the collector-base junction-breakdown voltage as a function of the collector-doping concentration, with the collector-base junction depth and the thickness of the lightly doped region of the epi layer ($x_{BL} - x_{BC}$) as parameters.¹²⁵ Figure 7-10b shows how a decrease in t_{epi} causes BV_{CEO} to decrease in a bipolar transistor designed for use in 5-V digital ICs.¹² Section

7.12.1.2 gives another example of a BV_{CEO} calculation in which this effect is considered.

7.3.4.2 Punchthrough Breakdown. In some cases the value of BV_{CEO} can be limited by *punchthrough* rather than by avalanche breakdown. If the base width is too narrow and/or the doping level in the base is too small, the depletion region of the C-B junction will penetrate the base region and join the E-B junction depletion region before the avalanche-limited value of BV_{CEO} has been reached. (Section 7.2.2 describes such base-width modulation.) If the collector voltage is increased beyond this value, the potential-energy barrier at the E-B junction will be lowered by ΔV , as shown in Fig. 7-11. As a result, a large emitter current (and correspondingly, a large collector current) can flow in the device, even with no avalanche breakdown. (Note that a large current flow across the C-B junction when both the E-B and C-B junctions are reverse biased is a manifestation of breakdown.)

Only a slight increase in the collector voltage beyond that needed to merge the two depletion regions is required for a large increase in I_C . Thus, punchthrough appears as a form of device breakdown. To prevent punchthrough, appropriate combinations of base doping and base width for the circuit operating voltages must be used. Most bipolar transistors are designed so that avalanche breakdown of the C-B junction will occur before punchthrough.

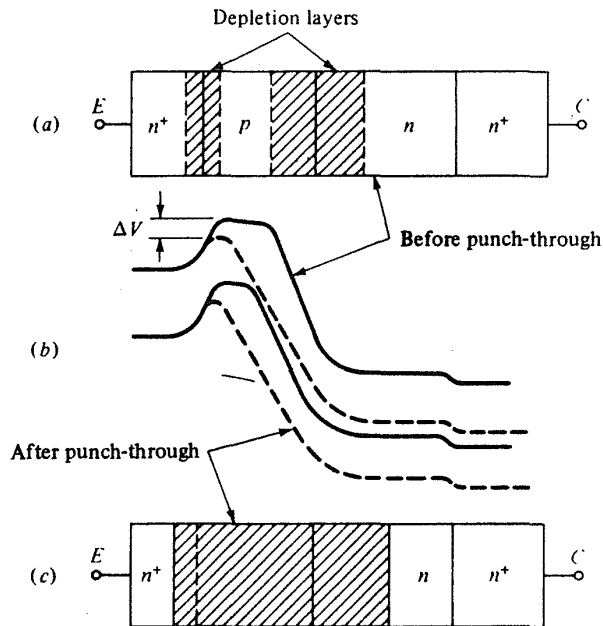


Fig. 7-11 Punchthrough in an *nnp* transistor: (a) space-charge regions before punchthrough, (b) energy-band diagram (and space-charge regions) after punchthrough. Copyright, 1988, McGraw-Hill Book Company. Reprinted with permission from ref. [2].

7.3.4.3 Breakdown Voltage and High-Level Injection Limits in Advanced Bipolar Transistors. To maintain the sufficiently high current levels needed to achieve minimum gate delays (see section 7.3.8), the current density is increased as the emitter area is scaled down. Higher collector doping levels must be used to avoid the Kirk effect. Furthermore, higher doping concentrations in the base are needed to avoid punchthrough of the narrow base regions. Thus, as of 1990, the base and collector doping concentrations in advanced bipolar devices are typically on the order of $1 \times 10^{18} \text{ cm}^{-3}$ and $1 \times 10^{17} \text{ cm}^{-3}$, respectively. The resulting peak electric field in the C-B junction is high enough to produce intolerably large avalanche-multiplication currents at low C-B reverse bias voltages (e.g., less than 3 V). Lu and Chen describe this problem in detail in reference 13.

7.3.5 Parasitic Series Resistances in Bipolar Transistors

Due to the planar structure of IC transistors, the current that flows from the emitter to the collector must travel parallel to the surface of the wafer after it has been transported vertically through the base. It must then flow upward to a contact at the wafer surface. Because the current-flow paths within the transistor have significant resistivity, parasitic series resistances exist in each of the three bulk regions of the device: a collector series resistance, R_C ; a base series resistance, R_B ; and an emitter series resistance, R_E . Each must be as small as possible.

7.3.5.1 Collector Series Resistance, R_C . A low value of R_C is important in both high-frequency ac circuits and high-speed digital circuits. In the former, a large R_C value would decrease the cutoff frequency and introduce a voltage drop that would increase V_{CEsat} . In the latter, a large R_C value would reduce both f_T (see section 7.3.7) and the maximum current-drive capability. Figure 7-12 shows the components of R_C .

Several approaches are used to minimize R_C . First, the lightly doped epitaxial layer used as the collector is doped as much as can be tolerated with regard to breakdown-voltage and junction-capacitance considerations. This means that in low-voltage digital devices (in which a low value of V_{CEsat} is critical but lower operating voltages are

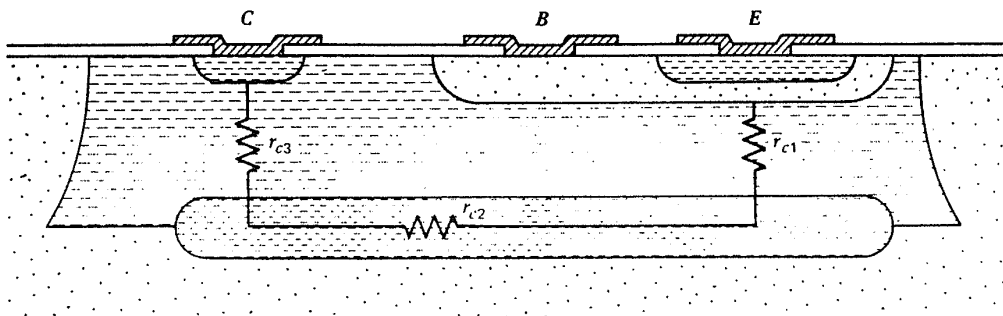


Fig. 7-12 Components of collector resistance, R_C . Copyright, 1984, John Wiley and Sons, Inc. Reprinted with permission from ref. [10].