Wiley Classics Library

Grebene

Bipolar and MOS Analog Integrated Circuit Design

EX-1014 0001
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Copyright © 2003 by John Wiley & Sons, Inc. All rights reserved.

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representation or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services please contact our Customer Care Department within the U.S. at 877-762-2974, outside the U.S. at 317-572-3993 or fax 317-572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print, however, may not be available in electronic format.

Library of Congress Cataloging-in-Publication Data is available.

ISBN 0-471-43078-1

Printed in the United States of America.

10 9 8 7 6 5 4 3 2

CONTENTS

CHAPTER 1. INTEGRATED-CIRCUIT FABRICATION 1

1.1 The Planar Process 1
1.2 Electrical Resistivity of Silicon 4
1.3 Solid-State Diffusion 5
1.4 Epitaxial Deposition 12
1.5 Oxidation of Silicon 14
1.6 Photomasking 17
1.7 Ion Implantation 20
1.8 Thin-Film Processes 22
1.9 Bipolar Integrated-Circuit Fabrication Steps 26
1.10 Modifications of Basic Process 31
1.11 Assembly and Packaging 38
1.12 Integrated-Circuit Packages 41
1.13 Testing of Integrated Circuits 46
1.14 Reliability Considerations 47

CHAPTER 2. ACTIVE DEVICES IN INTEGRATED CIRCUITS 53

2.1 *npn* Transistors 54
2.2 *npn* Transistors for Special Applications 75
2.3 *pnp* Transistors 83
2.4 Junction Field-Effect Transistors 95
2.5 MOS Field-Effect Transistors 106

CHAPTER 3. PASSIVE COMPONENTS: DIODES, RESISTORS, AND CAPACITORS 121

Part I Integrated Diodes 122

3.1 Junction Diodes 122

EX-1014 0003
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

3.2 Schottky Diodes 126
3.3 Zener Diodes 130

Part II Integrated Resistors **135**

3.4 Diffused Resistors 136
3.5 Pinched Resistors 144
3.6 Epitaxial Resistors 147
3.7 Ion-Implanted Resistors 150
3.8 Thin-Film Resistors 154
3.9 Trimming of Resistors 155

Part III Integrated Capacitors **160**

3.10 Junction Capacitors 160
3.11 MOS Capacitors 164

CHAPTER 4. BIAS CIRCUITS **169**

4.1 Constant-Current Stages 170
4.2 *pnp* Current Sources 183
4.3 Voltage-Controlled Current Sources 187
4.4 Supply-Independent Biasing 189
4.5 Voltage Sources 193
4.6 DC Level-Shift Stages 197
4.7 Temperature-Independent Biasing 204
4.8 Stabilization of Chip Temperature 210

CHAPTER 5. BASIC GAIN STAGES **215**

5.1 Differential Gain Stages 215
5.2 Gain Stages with Active Loads 233
5.3 Output Stages 246

CHAPTER 6. ANALOG DESIGN WITH MOS TECHNOLOGY **263**

6.1 Basic Characteristics of MOS Transistors 264
6.2 Building Blocks of NMOS Analog Design 271
6.3 Analog Design with Depletion-Mode Load Devices 284
6.4 Analog Design with CMOS Technology 290
6.5 MOS Voltage References 299
6.6 MOS Transistor as an Analog Switch 303

CHAPTER 7. OPERATIONAL AMPLIFIERS 309

7.1 Fundamentals of Operational Amplifiers 310
7.2 Circuit Configurations for Monolithic Operational Amplifiers 320
7.3 Frequency Compensation 325
7.4 Large-Signal Operation 333
7.5 Input Stage Design 339
7.6 Practical Op Amp Circuits 350
7.7 MOS Operational Amplifiers 368
7.8 Special-Purpose Op Amp 375
7.9 Other Operational Amplifier-Based Circuits: Buffers and Comparators 383

CHAPTER 8. WIDEBAND AMPLIFIERS 397

8.1 General Design Considerations 398
8.2 High-Frequency Transistors 399
8.3 High-Frequency Device Models 401
8.4 Frequency Response of Single-Transistor Gain Stages 403
8.5 Compound Devices 410
8.6 Neutralization of Collector-Base Capacitance 415
8.7 Amplifier Circuits Using Local Feedback 417
8.8 Amplifier Circuits Using Overall Feedback 423
8.9 Dual-Loop Feedback Amplifiers 429
8.10 Root-Locus Techniques 433
8.11 Current Amplifiers: The Gilbert Gain Cell 437
8.12 Electronic Gain Control 443

CHAPTER 9. ANALOG MULTIPLIERS AND MODULATORS 451

9.1 A Classification of Modulators and Multipliers 451
9.2 Properties of an Analog Multiplier 452
9.3 Applications of an Analog Multiplier 454
9.4 Variable-Transconductance Multiplier 456
9.5 Four-Quadrant Multipliers with Wide Dynamic Range 459
9.6 Practical Analog Multiplier Circuits 462
9.7 Balanced Modulators 469
9.8 Applications of Balanced Modulators 472

CHAPTER 10. VOLTAGE REGULATORS 481

Part I Series Regulators 482

10.1 Fundamentals of Series Regulators 482

EX-1014 0005
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

10.2 Protection Circuits 489
10.3 Practical Series Regulator Circuits 497
10.4 Layout Considerations for Power Circuits 509
10.5 Failure Mechanisms in Power Devices 512

Part II Switching Regulators 514

10.6 Fundamentals of Switching Regulators 514
10.7 Modes of Operation with Inductive Output Circuits 521
10.8 Efficiency Considerations 527
10.9 Practical Switching Regulator Circuits 528

CHAPTER 11. INTEGRATED-CIRCUIT OSCILLATORS AND TIMERS 541

Part I Integrated-Circuit Oscillators 541

11.1 An Overview of Oscillator Types 541
11.2 Tuned Oscillator Circuits 543
11.3 Relaxation Oscillators 556
11.4 Emitter-Coupled Multivibrators 571
11.5 CMOS Relaxation Oscillators 581
11.6 Limitations of Relaxation Oscillators 586
11.7 Monolithic Wave-Shaping Techniques 591

Part II Integrated-Circuit Timers 599

11.8 Fundamentals of Integrated-Circuit Timers 599
11.9 One-Shot Timers 600
11.10 Timer/Counter Circuits 609

Part III Frequency-to-Voltage and Voltage-to-Frequency Converters 615

11.11 Voltage-to-Frequency Converters 615
11.12 Frequency-to-Voltage Converters 622

CHAPTER 12. PHASE-LOCKED-LOOP CIRCUITS 627

Part I Fundamentals of Phase-Locked Loops 628

12.1 Principle of Operation of a PLL System 628
12.2 PLL in Locked Condition 635
12.3 Effects of Loop Filter and Loop Gain on PLL Performance 637
12.4 Applications of Phase-Locked Loops 647

Part II Building Blocks of Monolithic Phase-Locked-Loop Circuits 657

12.5 Phase Detectors 657

12.6 Voltage-Controlled Oscillators 668
12.7 Monolithic PLL Design Example 673

CHAPTER 13. INTEGRATED-CIRCUIT FILTERS 679

Part I A Review of Filter Characteristics 681

13.1 Basic Filter Specifications 681
13.2 A Review of Basic Filter Types 684
13.3 Biquadratic Filter Function 687
13.4 Sensitivity Considerations 696
13.5 Analog Sampled-Data Filters 698

Part II Switched-Capacitor Filters 703

13.6 Fundamentals of Switched-Capacitor Circuits 703
13.7 Characteristics of MOS Circuit Elements 712
13.8 Effects of Parasitic Capacitances 719
13.9 Practical Design Constraints 725
13.10 Second-Order Filter Configurations 727
13.11 Higher Order Filters 739
13.12 Applications and Limitations of Switched-Capacitor Filters 750

CHAPTER 14. DATA CONVERSION CIRCUITS: DIGITAL-TO-ANALOG CONVERTERS 753

14.1 Principles of D/A Conversion 754
14.2 Basic D/A Converter Circuits 757
14.3 Definitions of D/A Converter Terms 764
14.4 D/A Converter Architecture 770
14.5 Current Switches 780
14.6 Resistor and Capacitor Networks 785
14.7 Voltage References 790
14.8 Biasing of Current Sources 791
14.9 Effects of Device Mismatches 795
14.10 Accuracy Considerations 799
14.11 Monolithic Design Examples 802
14.12 Ultraprecision D/A Converter Circuits 817

CHAPTER 15. DATA CONVERSION CIRCUITS: ANALOG-TO-DIGITAL CONVERTERS 825

15.1 Fundamentals of A/D Conversion 827
15.2 Integrating-Type A/D Converters 835

EX-1014 0007
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

15.3 Digital-Ramp-Type A/D Converters 846
15.4 Successive-Approximation A/D Converters 847
15.5 Successive-Approximation Converters Using MOS Technology 852
15.6 Parallel A/D Converters 865
15.7 Other High-Speed A/D Conversion Techniques 871
15.8 Nonlinear A/D Converters for Telecommunications 873
15.9 An Overview of A/D Converter Techniques 876

INDEX 881

EX-1014 0008
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

CHAPTER ONE

INTEGRATED-CIRCUIT FABRICATION

The aim of this chapter is to familiarize the reader with the fabrication processes for analog integrated circuits. As a designer, one does not need to know the specific details of each and every process step, since these fabrication processes often become, by themselves, areas of specialization and rapid technological development. However, these manufacturing processes determine the capabilities and limitations of monolithic integrated-circuit (IC) products. Therefore, it is imperative at all times that a successful design engineer be familiar with the fundamental properties and the constraints of each of the major process steps in the fabrication of integrated circuits.

1.1. THE PLANAR PROCESS

The fabrication of a monolithic integrated circuit involves a complex sequence of processing steps. Even though the specific nature of these processes is well diversified, the bulk of the manufacturing steps associated with the present-day IC technology can be grouped under the term *planar process*.[1] Prior to the invention of the planar process in 1959, the solid-state electronics field was dominated by germanium devices. Introduction of the planar process has revolutionized the field of microelectronics almost overnight; and silicon, rather than germanium, emerged as the predominant semiconductor material.

When exposed to air, silicon forms an insulating oxide layer, called silicon dioxide (SiO_2). The formation of this oxide layer can be enhanced by exposing the single-crystal silicon wafer to steam or dry oxygen at high temperatures. The electrical and chemical properties of the SiO_2 layer are key to the planar process for the following reasons. First, the SiO_2 layer forms an inert cover over the

EX-1014 0009
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

silicon surface, and protects it from external contamination. Second, it serves as a barrier to the diffusion of impurities into silicon; thus, by etching well-defined patterns or *windows* in the oxide layer, one can diffuse desired impurities into selected areas of the silicon wafer. Finally, the SiO_2 layer provides an insulating surface over which the metal interconnections can be formed.

The planar process technology is comprised of five independent processes: epitaxy, oxidation, photolithography, diffusion, and thin-film deposition. Sometimes a sixth process called *ion implantation* is also used as a supplement or substitute for the diffusion process. A schematic illustration of these basic process steps is given in Figure 1.1 in terms of the cross-sectional view of a silicon wafer. For illustrative purposes, the vertical and horizontal dimensions of the wafer cross section are not drawn to scale.

Epitaxy is a deposition technique during which additional silicon atoms can be deposited on a single-crystal silicon substrate, without changing the crystalline structure of the silicon wafer. In other words, during the epitaxial deposition step, the single-crystal silicon substrate can be extended by the vapor

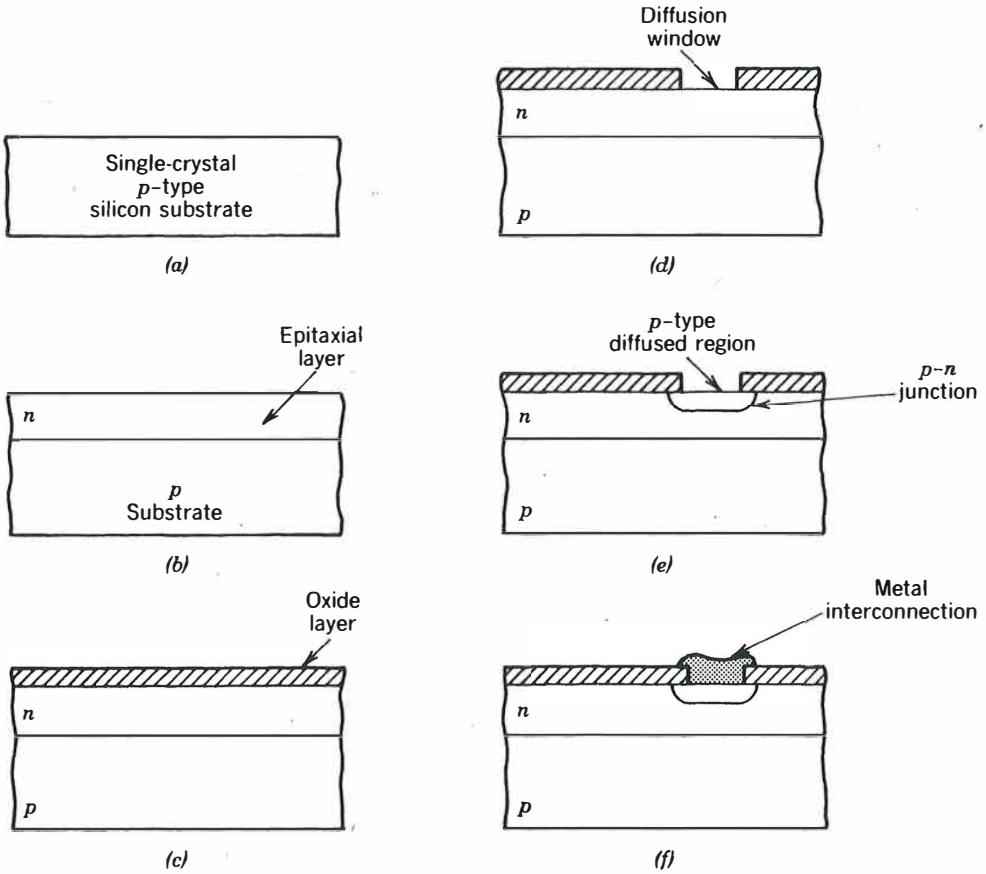


FIGURE 1.1. Basic steps in planar technology: epitaxy, oxidation, photolithography, diffusion, and metallization.

EX-1014 0010
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

phase deposition of additional atomic layers of silicon. By controlling the deposition rates, and introducing selected types and amounts of impurities into the carrier gases during the epitaxial deposition process, the thickness and the resistivity type of the epitaxial layer can be accurately controlled. In Figure 1.1*b*, this is illustrated for the case of an *n*-type epitaxial layer, deposited on a single-crystal *p*-type silicon substrate.

Oxidation, or surface passivation, is achieved by exposing the silicon wafer surface to an oxidizing atmosphere at high temperatures. As described earlier, this results in the formation of an oxide layer, SiO_2, which protects the silicon surface from contamination by undesired impurities. This is illustrated in Figure 1.1*c*.

Photolithography, or a masking technique, is then used to etch selective openings into the oxide layer. These openings serve as diffusion windows from which controlled amounts of impurity doping can be introduced into localized areas of the silicon wafer, as shown in Figure 1.1*d*. By using photographic techniques, the sizes of the diffusion windows can be greatly reduced without sacrificing the accuracy of alignment and the resolution of the line widths.

Controlled amounts of dopant impurities can be introduced into the preselected areas of the silicon surface through the diffusion windows in the oxide layer. The solid-state diffusion of these dopants into silicon at high temperatures (usually in excess of 1000°C) results in the formation of a *p*-*n* junction, as shown in Figure 1.1*e*. Since the diffusion of the impurities from the diffusion window proceeds sideways as well as downwards, the resulting junction edge on the silicon surface is located *under* the oxide layer and is not exposed to air on the surface. This protects the junction from possible contamination. During the diffusion process, the diffusion window can also be closed by exposing the wafer to oxidizing atmosphere, which would result in reoxidizing the exposed silicon surface.

Electrical contact to the semiconductor regions can be formed by opening new windows on the oxide layer and depositing a thin metal film of high electrical conductivity, such as aluminum (Al), over these windows. This conductive film can then be etched into a desired pattern, interconnecting the devices on the silicon wafer, thus completing the monolithic circuit structure (Fig. 1.1*f*).

As compared with earlier solid-state device fabrication methods, the planar process offers two unique advantages which make it ideally suited for integrated circuits:

1. The semiconductor junctions are protected on the surface by an inert oxide layer and are not exposed to air. This results in very low leakage currents and high reliability.
2. Photolithographic reduction, masking, and etching techniques are used to determine device geometries. This makes it possible to reduce device dimensions greatly and facilitates the simultaneous fabrication, or *batch processing*, of a large number of devices and circuits on the same silicon surface.

EX-1014 0011
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

The monolithic IC technology owes its rapid growth and acceptance to these two unique features of the planar technology.

1.2. ELECTRICAL RESISTIVITY OF SILICON

The addition of small concentrations of selected impurities, called *dopants,* into otherwise pure single-crystal silicon has a very significant effect on its electrical characteristics. These dopants are normally Group III or Group V elements from the Periodic Table, as shown in Table 1.1. When introduced into the silicon lattice, under proper process conditions, these dopant atoms substitutionally take the place of silicon atoms within the crystal structure. Thus, for the small amount of impurities added, the crystalline structure of silicon remains unchanged, but its electrical characteristics and its resistivity are profoundly affected.

When a silicon wafer is doped with Group III impurities such as boron (B) it is said to be p type, where the majority of the carriers available for current conduction are the holes. Similarly, when a Group V dopant such as phosphorus (P) or antimony (Sb) is used, the resulting silicon wafer will be n type, and the majority of carriers available for conduction will be the electrons.

For IC fabrication, boron and phosphorus are the most commonly used dopants, with arsenic (As) and antimony (Sb) also finding application in special cases.

In extrinsicly doped semiconductors, it is practical to assume that the majority carrier concentration is approximately equal to the density of the donor or acceptor dopant atom within the crystal. Thus, for example, for p- and n-type doped material, respectively,

$$p_p \approx N_A \quad \text{and} \quad n_n \approx N_D \tag{1.1}$$

where p_p and n_n are the equilibrium concentrations of the holes and the electrons (expressed in cm^{-3}) and N_A and N_D are the concentrations of the acceptor (Group III) and the donor (Group V) dopants.

The electrical resistivity $\rho(\Omega\text{-cm})$ for a p-type doped silicon can be approximated as

$$\rho \approx \frac{1}{q\mu_p N_A} \tag{1.2}$$

TABLE 1.1. p- and n-Type Dopants for Silicon Device Fabrication

p-Type Dopants		n-Type Dopants	
Boron	(B)	Phosphorus	(P)
Aluminum	(A1)	Arsenic	(As)
Gallium	(Ga)	Antimony	(Sb)
Indium	(In)		

EX-1014 0012
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

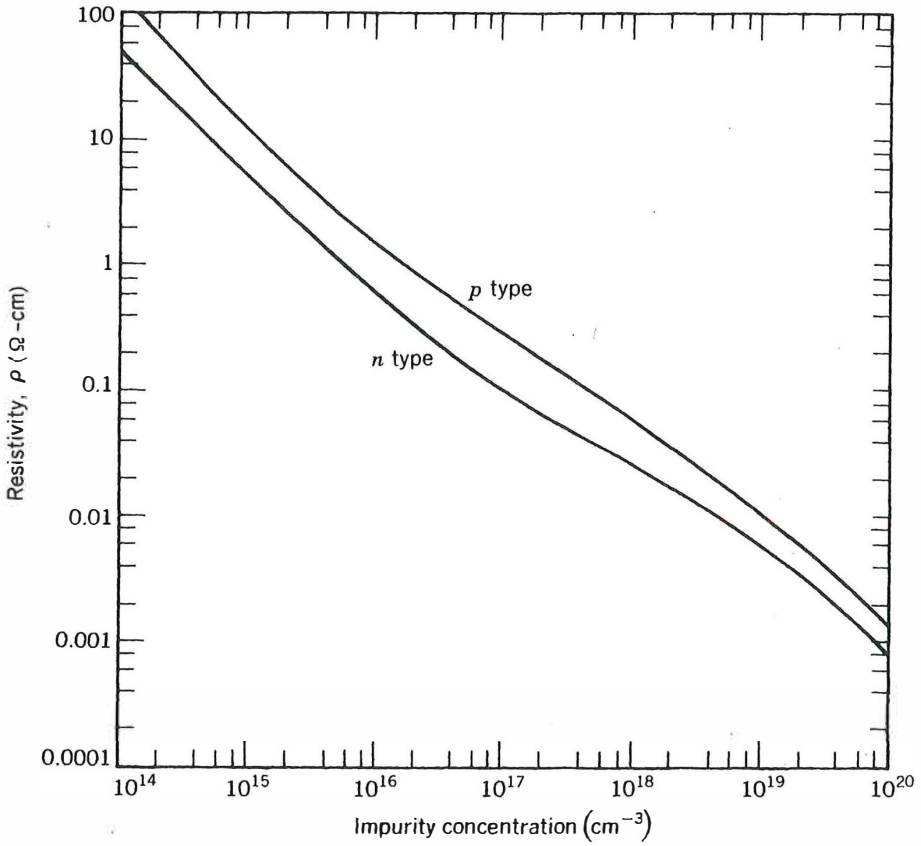


FIGURE 1.2. Resistivity of p- and n-type silicon as a function of impurity concentration.

where q is the electronic charge and μ_p (cm^2/V-sec) is the hole mobility. Conversely, the resistivity of a n-type doped silicon can be written as

$$\rho \approx \frac{1}{q\mu_n N_D} \tag{1.3}$$

where μ_n is the electron mobility.

Figure 1.2 gives a plot of the resistivity of uniformly doped p- and n-type silicon as a funciton of impurity concentration.

In certain calculations, conductivity, rather than resistivity, is used as a parameter. Conductivity σ $(\Omega\text{-cm})^{-1}$ is the reciprocal of resistivity (i.e., $\sigma = 1/\rho$).

1.3. SOLID-STATE DIFFUSION

The diffusion process is by far the most widely used method of introducing controlled amounts of impurities into the silicon substrate. It is a relatively well understood and highly reproducible process step which readily lends itself to the batch processing advantages of the planar technology, since a large number of silicon wafers can be processed simultaneously.

EX-1014 0013
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Diffusion, as a general process, is the mechanism by which different sets of particles confined to the same volume tend to spread out and redistribute themselves evenly throughout the confining volume. In the case of solid-state devices and integrated circuits, the relevant diffusion process is the one dealing with the movement and distribution of impurity atoms in a crystalline lattice structure. In crystalline solids, the diffusion process is significant only at elevated temperatures where the thermal energy of the individual lattice atoms has a statistical chance of overcoming the interatomic forces which hold the lattice together.

In the single-crystal silicon lattice, the impurities can move through the lattice by any one or a combination of the two dominant diffusion mechanisms. These physical mechanisms can be briefly outlined as follows.

Substitutional Diffusion. The impurity atoms propagate through the lattice by replacing a silicon atom at a given lattice site.

Interstitial Diffusion. The impurity atoms do not replace the silicon atoms at the regular lattice sites, but instead move into the interstitial voids in the three-dimensional lattice structure.

In the fabrication of analog integrated circuits, substitutional diffusion is by far the most important mechanism, since all the impurities intentionally introduced into the silicon substrate to form the junction and the device structures diffuse substitutionally. This is not necessarily the case for digital integrated circuits where controlled amounts of interstitial impurities, such as gold, copper, or nickel, can be introduced into the silicon lattice to reduce the minority carrier lifetime.

Diffusion Theory

Although the diffusion process proceeds in all three dimensions simultaneously, for the analysis of the fundamental properties of that process it is sufficient to consider only one dimension. As will be described in Chapters 2 and 3, the geometry and dimensions of most semiconductor devices fabricated by the planar process justify this one-dimensional assumption.

The fundamental physical property of the diffusion process is that the particles tend to diffuse from a region of high concentration to that of lower concentration at a rate proportional to the concentration gradient between the two regions. This is known as Fick's first law, and can be mathematically expressed as

$$F = -D\frac{\partial N}{\partial x} \tag{1.4}$$

where F is the net particle flux density, that is, the net number of particles flowing through a unit surface area normal to the direction of flow per unit time, N is the number of particles per unit volume, and x is the distance measured parallel to the direction of flow. D is the diffusion coefficient and has units of

EX-1014 0014
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

(length)^2/time. The magnitude of D gives a measure of the relative ease or difficulty with which the diffusing particle can move about in its environment. The negative sign appears in Eq. (1.4) to indicate that the particle flow is directed from a region of high concentration toward one of lower concentration.

In all diffusion problems, one is interested in the variations of the impurity concentration with time as well as with distance. The fundamental law of diffusion, which relates the time rate of change of concentration to the spatial coordinates of the region in question, can be derived from Fick's law by applying the continuity principle to Eq. (1.4) as follows. Consider a region or a volume in the material enclosed by an area A normal to the flow and a width dx parallel to the flow. The net flow of particles into this volume can be written as

$$FA - \left(F + \frac{\partial F}{\partial x}dx\right)A = -\frac{\partial F}{\partial x}dx\,A \tag{1.5}$$

This net flow into the volume can be related to the concentration change within the volume as

$$\frac{\partial N}{\partial x}dx\,\mathrm{A} = -\frac{\partial F}{\partial x}dx\,\mathrm{A} \tag{1.6}$$

The expression for $\partial F/\partial x$ can be obtained by differentiating Eq. (1.4). Substituting the result into Eq. (1.6), one obtains

$$\frac{\partial N}{\partial t} = D\frac{\partial^2 N}{\partial_x{}^2} \tag{1.7}$$

This equation is known as Fick's second law, and it defines the impurity distribution at any given point within the semiconductor crystal as a function of time.

The diffusion coefficient D is a property of the particular impurity, and is an exponential function of temperature. Figure 1.3 shows the diffusion coefficients of various dopants in silicon as a function of temperature. Note that the diffusion coefficient increases by approximately one order of magnitude for every 100°C change in temperature. Thus, the diffusion process is effective only at temperatures in excess of 1000°C.

Since the basic equation [Eq. (1.7)], which describes the distribution of the impurities in silicon as a function of both time and distance, is a second-order partial differential equation, its particular solution applicable to a diffusion problem depends on the boundary conditions associated with the diffusion process. Most of the diffusion processes encountered in the fabrication of integrated circuits fall into one or the other of the following two classes of boundary conditions: constant-source (complementary error function) diffusion or limited-source (Gaussian) diffusion. The properties of each of these diffusion profiles are described below.

Constant-Source Diffusion. In this type of diffusion, the wafer is exposed to the impurity source during the entire duration of the diffusion. Thus, there is

EX-1014 0015
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

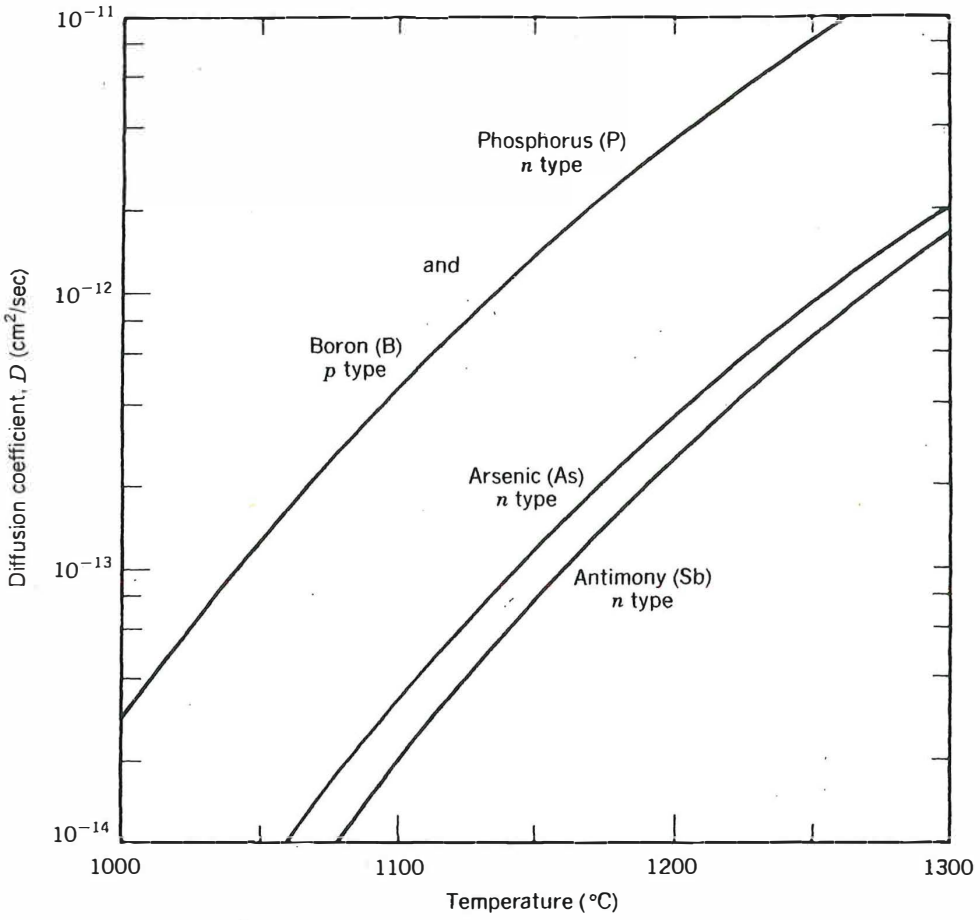


FIGURE 1.3. Diffusion coefficients of various dopants in silicon.[3]

an undiminished supply of impurities at the wafer surface; and the impurity concentration at the wafer surface, N_o is constant, as set by the solid solubility of the particular dopant in silicon. If we also assume that there are no impurities in silicon at time $t = 0$, then the solution describing the distribution of impurities in silicon becomes the *complementary error function,* erfc, where

$$N(x, \text{t}) = N_o \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right) \tag{1.8}$$

In the above expression, N is the density of impurity atoms (atoms/cm^3) at a distance x from the diffusion surface at time t after the start of diffusion. Figure 1.4 shows a logarithmic plot of the impurity concentration as a function of time t, subsequent to the start of the diffusion cycle.

It should be noted that if the diffusion is performed into a wafer of opposite conductive type, then a junction is formed at a depth x_j, where the concentration of the diffused impurity is equal to that of the background concentration N_B.

EX-1014 0016
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

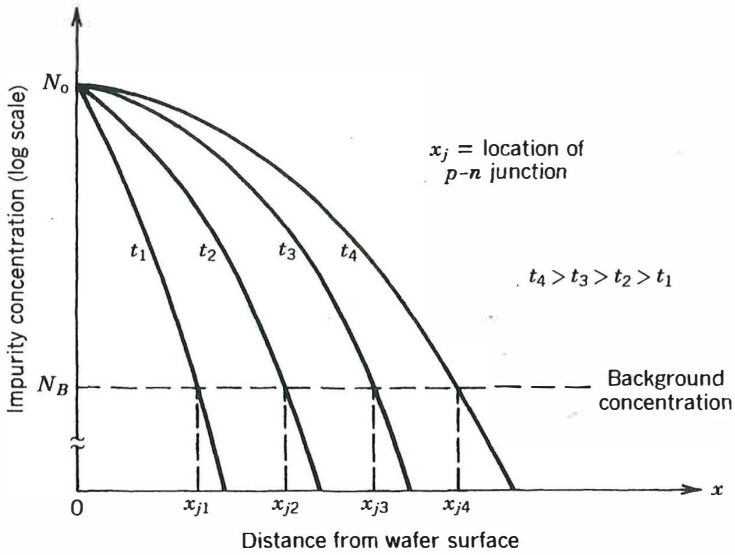


FIGURE 1.4. Constant-source diffusion profiles as a function of time.

Limited-Source (Gaussian) Diffusion. In this type of diffusion, the wafer is exposed briefly to impurities during a so-called predeposition step, where a thin layer of dopant atoms is placed on the silicon surface. After that, the impurity source is turned off, and the total amount of impurities deposited on the surface serves as the impurity source for the rest of the diffusion cycle. In this case, the resulting impurity distribution is approximated by the Gaussian distribution given as

$$N(x, t) = \frac{Q}{\sqrt{\pi Dt}} \exp\left(\frac{-x^2}{4Dt}\right) \tag{1.9}$$

where Q (atoms/cm^2), is the initial concentration of impurity atoms deposited on the surface during the predeposition step, which precedes the diffusion cycle. Figure 1.5 shows a logarithmic plot of the impurity concentration into the wafer as a function of increasing time t. Note that since the total amount of impurities, Q, available for diffusion is constant, the surface concentration N_o decreases with increasing time.

Both the complementary error function and the Gaussian distributions are well-defined functions. Figure 1.6 gives a plot of each of these functions for various values of their arguments. Note that, as shown in Figures 1.4 and 1.5, the junction depth x_j is monotonically increasing function of time.

In monolithic IC fabrication, the complementary error-function diffusions are used for either very deep or very shallow diffusions, such as the p-type isolation or the n-type emitter diffusion of *npn* transistors. The Gaussian diffusion is usually used for medium-depth diffusions, such as the p-type base diffusion of *npn* bipolar transistors.

EX-1014 0017
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

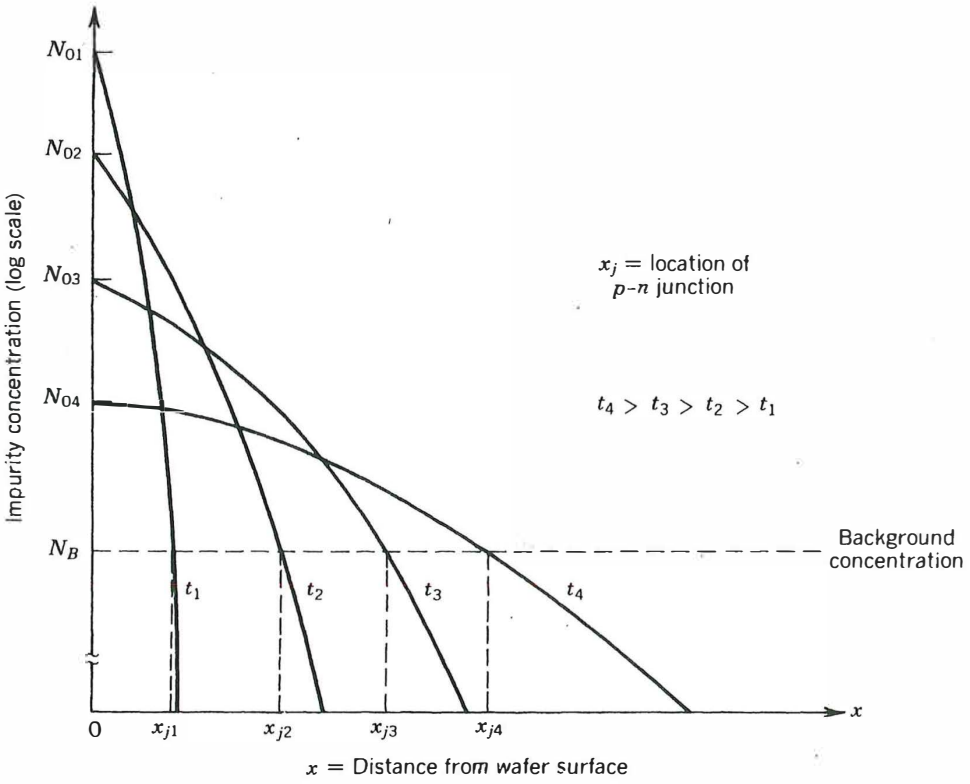


FIGURE 1.5. Limited-source (Gaussian) diffusion profiles as a function of time.

Basic Properties of the Diffusion Process

In the design and fabrication of monolithic integrated circuits, the following three fundamental properties of the diffusion process must be considered:

1. All diffusions proceed simultaneously. The impurities introduced in an earlier diffusion step continue to diffuse during subsequent diffusion cycles. Therefore, when calculating the total effective diffusion time for a given impurity profile, one must often consider the effects of subsequent diffusion cycles. The effects of the subsequent diffusion on a given impurity profile can be estimated by defining an effective Dt product for the particular impurity profile as

$$(Dt)_{\text{eff}} \approx D_1 t_1 + D_2 t_2 + D_3 t_3 + \cdots \tag{1.10}$$

where $t_1, t_2, t_3, \ldots$ are the different diffusion times, and $D_1, D_2, D_3, \ldots$ are the corresponding diffusion coefficients as determined by the respective temperatures of the diffusion cycles. Thus, for example, in the planar device fabrication, the emitter region of a bipolar transistor is formed by a diffusion process which succeeds the base diffusion step. Therefore, the effective Dt product of the base region contains a finite contribution from the emitter diffusion step.

EX-1014 0018
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

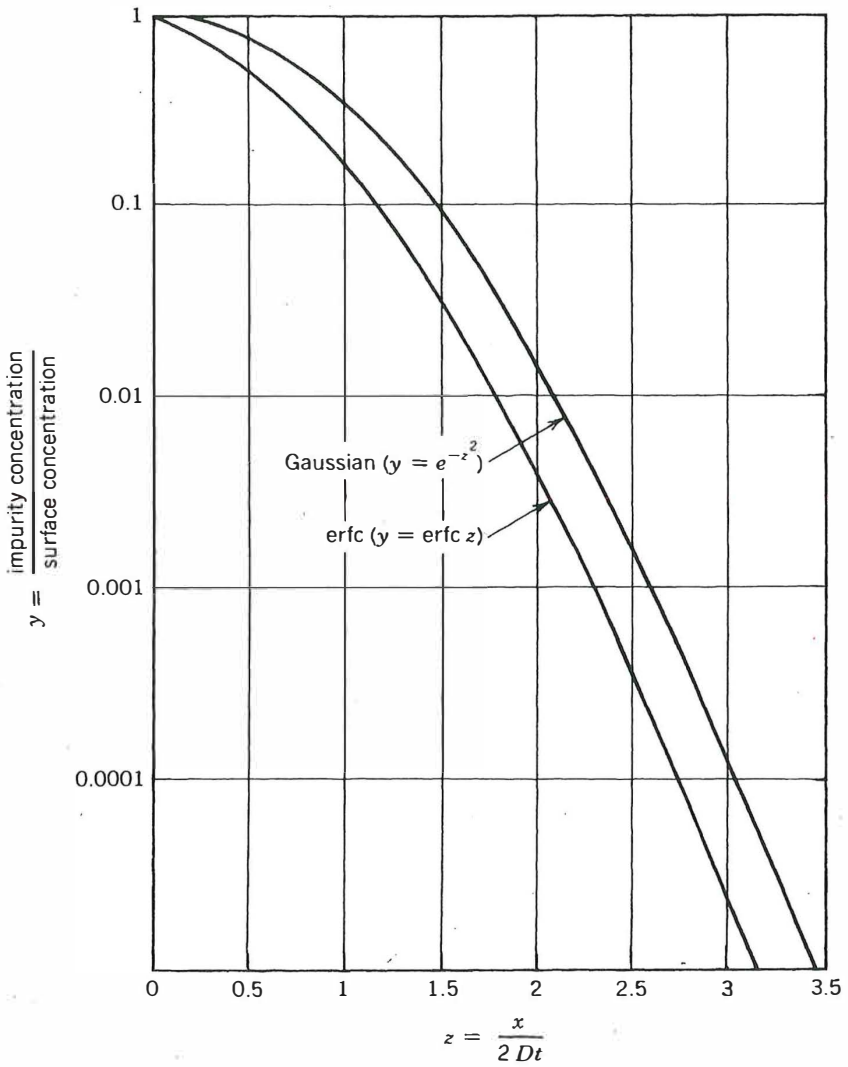


FIGURE 1.6. Values of the complementary error function (erfc) and the Gaussian distribution as functions of their arguments.

2. The diffusion profiles of Eqs. (1.8) and (1.9) are both functions of $(x/\sqrt{Dt})$. Therefore, for a given surface and background concentration, the junction depths x_1 and x_2 associated with the two separate diffusions having different times and temperatures, can be related as

$$\frac{x_1}{x_2} = \sqrt{\frac{D_1 t_1}{D_2 t_2}} \tag{1.11}$$

3. The diffusions proceed sideways from a diffusion window as well as downward. In considering the lateral dimensions of the planar devices, particularly in the case of lateral *pnp* transistors and MOS transistors, these lateral

diffusion effects need to be considered. Typically, the lateral diffusion distance is about 75–80% of the vertical penetration.

1.4. EPITAXIAL DEPOSITION

Epitaxy is a deposition technique where the single-crystal structure of a silicon substrate can be extended by vapor phase deposition of additional atomic layers of silicon. Epitaxial growth, or deposition, is carried out in a special furnace called a *reactor*, where silicon wafers having clean and chemically polished surfaces are heated up to temperatures comparable to those encountered in the diffusion step (i.e., 1000°–1200°C). During the epitaxial growth, vapors containing silicon are passed over the heated substrate. Normally hydrogen is used as the carrier gas, with either silicon tetrachloride ($SiCl_4$) or silane (SiH_4) as the source of silicon. Normally the $SiCl_4$ process requires somewhat higher temperatures than SiH_4 decomposition, and also has a slower growth rate. During the expitaxial growth process, the source compound is chemically reduced, resulting in free silicon atoms, some of which are deposited on the single-crystal substrate. Under proper deposition conditions, the interatomic forces of the single-crystal silicon lattice constrain the deposited silicon atoms to follow the original crystal structure. Thus, structurally, the deposited epitaxial layer forms a continuation of the original crystal structure.

During the process of epitaxial growth, controlled amounts of p- or n-type impurities are also introduced into the carrier gas to control the type and resistivity of the deposited layer. Unlike the diffusion process, epitaxial growth proceeds by uniform addition of atomic layers onto the substrate. Thus, the dopant impurities are uniformly distributed through the epitaxial layer, and do not show a concentration gradient. Furthermore, epitaxial layers can be grown over diffused regions or over other epitaxial layers.

Redistribution of Impurities during Epitaxy

Since epitaxial growth is a high-temperature process, the impurities at the interface of the epitaxial layer (epi) and the substrate tend to redistribute themselves via the diffusion process. For example, in the case of an n-type epitaxial layer grown on a p-type substrate, the epi–substrate interface no longer represents a step junction, but becomes graded due to the diffusion of impurities from the epitaxial layer into the substrate, and vice versa. Consequently, the impurity distribution at the epi–substrate interface may look as illustrated in Figure 1.7. The dashed line shows the ideal p-type impurity distribution at the interface, and the solid line corresponds to the actual distribution.

For relatively rapid rates of epitaxial growth (i.e., > 0.2 μm/min), the impurity distribution $N(x, t)$ across the interface can be written as

EX-1014 0020
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

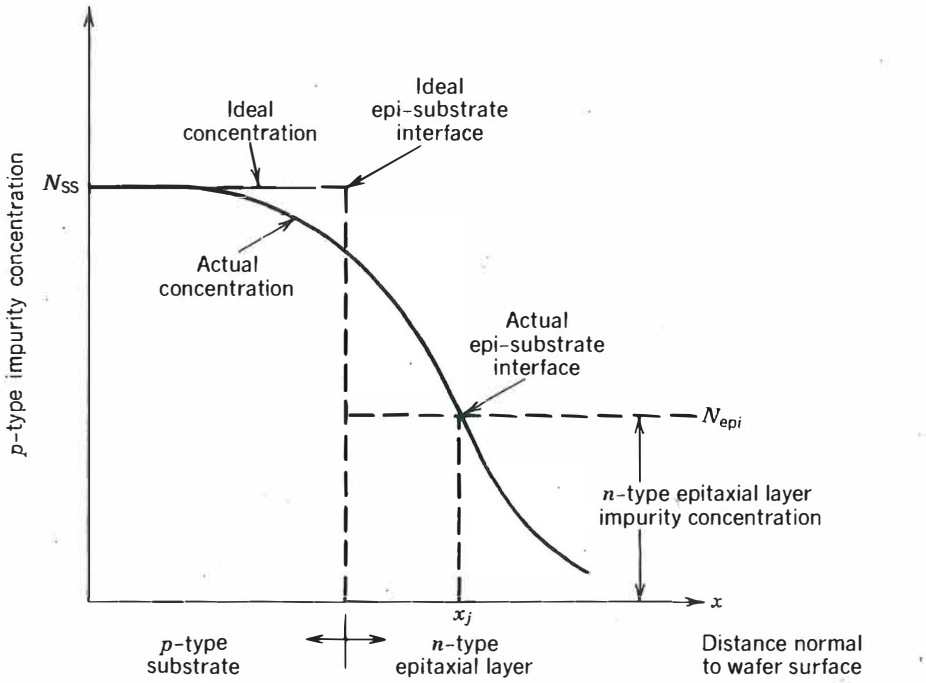


FIGURE 1.7. Impurity redistribution at epi–substrate interface during subsequent epitaxial growth and diffusion cycles.

$$\frac{N(x, t)}{N_{SS}} \approx \frac{1}{2}\,\text{erfc}\left(\frac{x}{2\sqrt{Dt}}\right) \tag{1.12}$$

where N_{SS} is the impurity concentration within the bulk of the substrate. The total amount of interdiffusion across the epi–substrate interface during the entire device fabrication cycle can be estimated by defining an effective Dt product for the out-diffusion process, in accordance with Eq. (1.10).

In practical *npn* bipolar transistor structures (see Figs. 1.11 and 1.13), a heavily doped n-type layer, called the *buried layer*, is diffused into selected regions of the p-type substrate prior to the growth of the n-type epitaxial layer. This buried layer diffusion serves as a low-resistivity conduction path for the collector current of the *npn* transistor. However, during subsequent epitaxial growth or additional diffusion steps, this buried layer tends to out-diffuse into the epitaxial layer and, thus, increase the n-type impurity concentration within the region of the epitaxial layer close to the subepitaxial n-type buried layer diffusion. In such cases, Eq. (1.12) can also be used to estimate the amount of anticipated out-diffusion from the buried layer into the epitaxial region by replacing N_{SS} by N_o where N_o (atoms/cm^3) is the surface concentration of the n-type donor atoms on the buried layer surface, prior to the epitaxial growth step.

In order to minimize the out-diffusion of the buried layer during the epitaxial growth or the subsequent diffusion steps, donor impurities having low diffusion coefficients, such as arsenic or antimony, are used for buried layer dopants.

EX-1014 0021
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Crystal Defects in Epitaxial Layers

During the epitaxial deposition process, a number of crystal defects may occur. Depending on their nature and density, these defects and imperfections may affect the overall electrical characteristics of the epitaxial layer and of the device junctions formed in it. When present in large numbers, these crystal defects reduce the minority carrier lifetime within the epitaxial layer, increase the junction leakages, and cause localized voltage breakdowns.

The most commonly encountered imperfections in the epitaxial structures are *dislocations* and *stacking faults*. The dislocation defects are caused by imperfect arrays of atoms within the localized regions of the substrate lattice and by mechanical stress. The dislocations tend to appear in clusters and be oriented along lines known as *slip planes*. They can be reduced by taking additional care during the substrate preparation, prior to epitaxy.

Another prevalent crystal defect in the epitaxial regions is the stacking fault caused by improper stacking of the crystal planes over a localized region. Such a fault may start at the epi–substrate interface and propagate through the deposited layer thickness; or it may originate entirely within the epitaxial region. This latter occurrence is often associated with excessively high growth rates.

Additional types of crystal defects in epitaxial layers are pits or pyramids on the epitaxial layer surface, which are, in general, due to either excessive or parasitic impurities present in the epitaxy system.

Polycrystalline Silicon Growth

If silicon is epitaxially grown on a noncrystalline substrate, such as an SiO_2 layer, the deposited layers of silicon tend to be oriented in random directions, thus leading to a polycrystalline structure. Polycrystalline silicon does not have the electrical properties associated with single-crystal semiconductors, but can be used for a variety of special applications in IC fabrication. One particular application of polycrystalline silicon growth will be described later in this chapter, in connection with dielectrically isolated device structures (see Fig. 1.14). Another commonly encountered application of polycrystalline silicon is in forming the gate electrode of MOS devices (see Fig. 2.50), where it has the effect of reducing the gate threshold voltage.

1.5. OXIDATION OF SILICON

When exposed to an oxidizing atmosphere, silicon forms an insulating silicon dioxide (SiO_2) layer on its surface. This oxide layer is an inert dielectric; in sufficient thickness, it is impervious to impurities or most forms of contamination. Thus, it forms a natural *passivation* layer over the active silicon area within the wafer. In the fabrication of monolithic integrated circuits, this oxide layer performs three fundamental functions:

EX-1014 0022
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

1. It serves as a diffusion barrier and allows selective diffusions into silicon through the windows etched into the oxide.
2. It protects the junctions from exposure to moisture and other contaminants in the atmosphere.
3. It serves as an insulator on the device surface on which the metal interconnections can be formed.

In addition to these three functions, the SiO_2 layer is often utilized as the dielectric region of monolithic capacitances, or as the gate dielectric of MOS devices. The SiO_2 layers can be formed on the silicon surface by any one of several methods. Some commonly used techniques consist of thermal oxidation, pyrolytic deposition, and anodic or gas plasma oxidation. Among these, thermal oxidation is by far the most commonly used growth technique.

Thermal Oxidation of Silicon

During the thermal oxidation step, an oxide layer is formed on the silicon surface through the basic chemical reaction

$$Si + O_2 \rightarrow SiO_2 \tag{1.13}$$

In the presence of some water vapor, the oxidation process is significantly accelerated and proceeds in accordance with the chemical reaction

$$Si + 2H_2O \rightarrow SiO_2 + 2H_2 \tag{1.14}$$

Except for the initial oxidation step (see Fig. 1.1*c*), the thermal oxide growth is not generally performed as a separate fabrication step. Instead, it is often incorporated into the diffusion process by providing an oxidizing atmosphere during part of the diffusion cycle. This in turn provides sufficient oxide on the previous diffusion windows for the next masking step.

Thermal oxidation normally proceeds at a temperature range of 900–1200°C. During oxidation, a carrier gas containing the oxidizing agent (normally oxygen gas or water vapor) is passed over the heated wafer substrate. The kinetics of thermal oxide growth are well understood and covered in the literature.[4,5] Oxidation proceeds by an inward motion of the oxidizing species toward the silicon–SiO_2 interface. Therefore, as the oxidation process proceeds, it is necessary for the oxygen molecules to diffuse through a thicker layer of SiO_2 to get to the silicon surface where the chemical reactions (1.13) or (1.14) can take place. Consequently, the time rate of oxide growth decreases rapidly with increasing oxide thickness. It can be shown[5] that for very thin layers of SiO_2, the growth rate is linear with time. However, as the oxide thickness x_o increases, the growth rate becomes proportional to $\sqrt{t}$.

The practical thicknesses of thermally grown SiO_2 layers used in monolithic IC fabrication are in the range of 500–20,000 Å (10,000 Å = 1 μm). The lower limit of thickness is often dictated by electrical breakdown or random defect densities (i.e., pin holes) in the oxide layer. The upper limit is set by required

EX-1014 0023
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

oxidation times and the difficulty of etching the oxide layer during the photomasking step.

Masking Properties of SiO_2

The diffusion coefficients of most dopants in SiO_2 are about two to four orders of magnitude smaller than those in silicon. Therefore, for these impurities, which include all those listed in Table 1.1 with the exception of gallium, an SiO_2 layer of proper thickness can serve as a diffusion barrier. The minimum oxide thickness necessary to mask against a given diffusion step depends to a large extent on the specifics of the diffusion process, such as type of dopant used, surface concentration, predeposition temperature, and time. Figure 1.8 gives some typical curves showing the minimum oxide thickness needed to mask against the two most commonly used dopants, boron and phosphorous.[6]

Several positively charged ionic species, such as sodium (Na^+) or hydrogen (H^+) ions, can diffuse through the SiO_2 layer with relative ease at temperatures as low as 150°C. Therefore, oxide passivation is prone to ionic contamination. These ions tend to generate a positive space charge within the silicon–SiO_2 interface, which in turn leads to an increased free-electron concentration of the silicon side of the interface. As a consequence, the surface layer of silicon directly against the SiO_2 layer tends to appear less p type or more n type than would be expected from the dopant impurity concentration. This effect, when coupled with the depletion of the p- type boron concentration during the oxide growth cycle, may result in the formation of a parasitic n-type inversion layer at the Si–SiO_2 interface. This parasitic inversion layer, known as *channeling*, is a dominant failure mechanism for integrated devices containing lightly doped p-type regions. It can be eliminated by maintaining a relatively high surface concentration of boron within the p-type regions (typically $\geq 10^{17}$ atoms/cm^3) and by avoiding ionic contamination of the SiO_2 layer.

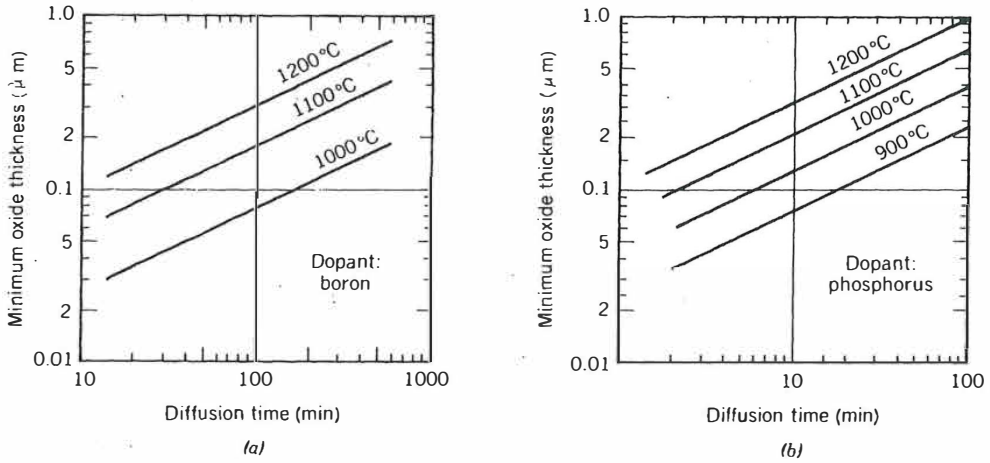


FIGURE 1.8. Minimum oxide thickness required to mask against diffusion: (*a*) Boron and (*b*) phosphorus.[6]

EX-1014 0024
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Chemically Deposited Oxide Layers

Sometimes it is advantageous to form an inert dielectric coating over the surface of the integrated circuit by pyrolytically depositing an oxide layer. Such a deposition process is often referred to as chemical vapor deposition (CVD) step. During the CVD step, the silicon wafer is maintained at a relatively low temperature (typically 400°C). Thus, such a step is particularly useful as a final passivation layer over the IC surface, subsequent to the completion of the metal interconnection or the thin-film deposition step, and protects the device surface from mechanical damage or scratches.

Silicon Nitride Passivation

Silicon nitride (Si_3N_4) is far more resistant to ionic contamination than SiO_2. Therefore, it is frequently utilized as a passivating layer for IC structures whose performance can be easily degraded by surface contamination. This is particularly true for analog integrated circuits involving MOS devices or operating at low current levels. An additional advantage of Si_3N_4 over the thermally grown oxide is its superior masking properties against the dopant impurities. Even such dopants as gallium, which readily diffuse through SiO_2, can be effectively masked by Si_3N_4.

The Si_3Ni_4 passivating layer is most conveniently formed by a pyrolytic deposition process at a temperature range of 800–1000°C.

The deposition is obtained by the decomposition of SiH_4 and ammonia (NH_3) in the presence of hydrogen gas, in accordance with the reaction

$$3\ SiH_4 + 4\ NH_3 \rightarrow Si_3N_4 + 12\ H_2 \qquad (1.15)$$

Silicon nitride is often used to complement the SiO_2 passivation process. In such an application, a layer of Si_3N_4 (typically 1000 Å thick) is sandwiched between two SiO_2 layers on the wafer surface to provide an added degree of surface passivation. The second layer of SiO_2 over the nitride layer is normally formed by pyrolytic deposition. This second oxide layer also serves as a mask during the photomasking and etching of the contact windows through the nitride layer.

1.6. PHOTOMASKING

The initial layout of an integrated circuit is normally done at a scale several hundred times larger than the final dimensions of the finished monolithic chip. This initial layout is then decomposed into individual mask layers, each corresponding to a masking step during the fabrication process. The individual mask layers are then reduced photographically to the final dimensions of the integrated unit. The reduced form of each of these patterns is then contact-printed on a transparent glass slide to form a photographic *mask* of the patterns to be etched

EX-1014 0025
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

on to the SiO_2 surface. To facilitate batch processing, a large number of such masks are contact-printed on the same glass slide, forming a *masking plate*. The plate is sufficiently large to cover the entire surface of the silicon wafer to be masked. Thus, in a single masking operation, an array of a large number of identical masks can be applied simultaneously over the wafer surface.

During the masking operation, the mask pattern is transferred from the masking plate to the wafer surface by photolithographic techniques. The wafer surface to be masked is initially coated with a photosensitive coating known as *photoresist* or *resist*. The resist-coated wafer surface is then brought into intimate contact with the masking plate and exposed under an ultraviolet light. The portions of the photosensitive resist not covered by opaque portions of the mask polymerize and harden as a result of this exposure. Then the unexposed parts of the resist can be washed away, leaving a photoresist mask on the wafer surface. As a consequence of the masking step, the pattern to be etched through the oxide is transferred to the wafer surface in the form of a hardened etch-resistant photoresist pattern.

The photomasking step is followed by an etching step during which the parts of the SiO_2 layer not protected by the exposed resist mask are etched away, forming the diffusion or the contact windows on the oxide. In this process, a buffered hydroflouric acid (HF) solution is used as the etchant. Following the etching step, the photoresist is washed away by a special cleaning solution, and the silicon wafer is ready for the next diffusion step. A similar photomasking step is also used in forming the metal interconnection patterns. The typical sequence of steps in the photomasking process is illustrated in Figure 1.9.

Dimensional Tolerances

In most monolithic circuit structures, the lateral dimensions of the integrated components are determined by the limitations of the photolithographic reduction, masking, and etching processes. The two fundamental limitations on the photolithography process are the alignment and the resolution of the mask patterns.

Since the monolithic IC fabrication steps require the successive application of a number of masks, it is necessary that each new mask applied to the silicon surface align with the previous set of masks over the entire surface of the wafer. This requires a good degree of dimensional accuracy associated with the initial layout of the circuit. To ensure this dimensional accuracy, the initial layout is carried out at the largest possible magnification within the capabilities of the photoreduction system. Typically, a 500× size is preferred for the initial layout for circuits having final reduced dimensions of up to approximately 70 sq. mils. For larger overall chip dimensions, a smaller initial layout scale, such as 400×, may be preferred to avoid optical distortion during the reduction process. Note that the drafting inaccuracies associated with the initial layout are also reduced at the same scale as the original layout. Thus, for example, a 0.01-in. dimen-

EX-1014 0026
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

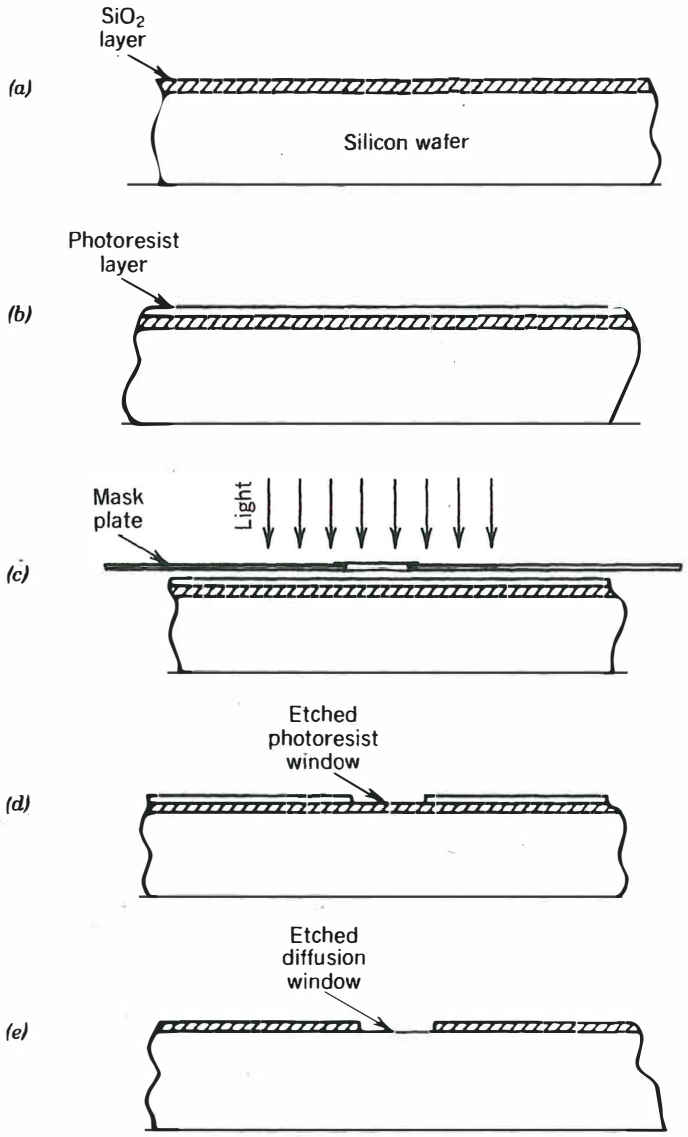


FIGURE 1.9. Typical sequence of steps in the photomasking process which result in a diffusion window pattern in the SiO_2 layer that duplicates the pattern on the mask plate. (*a*) Growing SiO_2 on wafer surface; (*b*) applying a thin coat of photoresist on oxidized wafer surface; (*c*) exposing photoresist through a mask plate; (*d*) developing and etching photoresist; (*e*) etching the exposed SiO_2 layer and stripping the photoresist to end up with a diffusion window in the SiO_2 layer.

sional inaccuracy in the initial layout leads to a $\pm$ 0.5-1 μm error in the final dimension at a 500$\times$ reduction.

A possible source of error in the masking step is the tolerance associated with the "step-and-repeat" process in contact-printing the mask array on the masking plate. The source of error in this case is the mechanical advance mechanism involved. An additional factor limiting the alignment tolerances of a mask set is

EX-1014 0027
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

the accuracy of positioning the mask on the wafer surface. This is done with the aid of a mechanical alignment jig under a high-powered microscope. However, it is still subject to some operator error. To minimize the alignment errors at the stage of the masking operation, it is customary to use concentric alignment patterns on successive mask layers. The alignment accuracy for a typical mask set, under production conditions, is approximately $\pm$ 1 μm for concentric patterns.

The ability of the mask to define or reproduce fine details on the wafer surface is determined by the resolution of the photomasking step. A good measure of the resolution is the minimum line width needed to resolve, reproducibly, two parallel lines spaced one line width apart. The main limitations to the resolving power of the photomasking techniques are the statistical fluctuations in the molecular structure of photographic emulsions and the diffraction of the light at the mask edges. At present, the minimum line width that can be resolved under production conditions is approximately 2 μm.

The etching step also introduces random irregularities, or errors, which tend to reduce the overall mask resolution. The grainy structure of the exposed and polymerized photoresist does not define a true edge during the etching step, but can cause random irregularities of the order of ± 0.5 μm along straight edges. This effect, along with the nonuniform etching properties of the oxide layers, also tends to round the sharp corners on the masks to a typical radius of 2–3 μm. At present, the minimum dimensions of a diffusion window that can be formed routinely under production environment is approximately 4 μm $\times$ 4 μm. However, these dimensional tolerances can be improved significantly by using more advanced and complex pattern-forming techniques, such as electron-beam photolithography[7].

1.7. ION IMPLANTATION

In the ion-implantation process, the impurities are introduced into silicon by bombarding the wafer surface with high-energy ions of the desired impurity type.[8,9] The implantation operation takes place in a vacuum. Impurity ions are accelerated from an ion source, and a mass spectrometer is used to separate the undesired impurities from the beam. The ion beam is then focused to a small area (typically smaller than $\frac{1}{4}$ in.^{2}) and is scanned across the semiconductor wafer which serves as the target. During the implantation process, the depth of penetration of the impurity ions into the silicon lattice is controlled by their energy, which is set by the accelerating field; and the density of the implanted ions is controlled by the beam current. Typical energy levels used in the ion-implantation process are in the range of 30–200 kilo-electron-volts (keV). When ions penetrate the silicon wafer, they produce lattice defects or dislocations. These are removed by annealing the wafer at temperatures as the order of 500–600°C, subsequent to the implantation step.

The impurity profile resulting from ion implantation has a Gaussian distribu-

EX-1014 0028
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

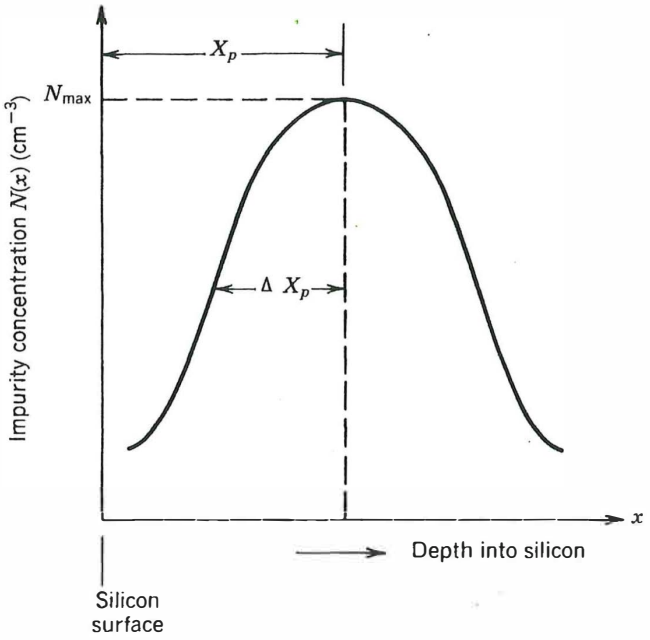


FIGURE 1.10. Typical distribution of implanted impurity atoms in silicon.

tion, with the peak of the distribution appearing *below* the surface of the silcon wafer, as shown in Figure 1.10. The peak of the distribution occurs at a depth X_p, called the *mean range*, which increases with increasing mass and energy of the incident ions. The relative spread of the distribution is measured in terms of its standard deviation, ΔX_p. The relative spread of the distribution $\Delta X_p/X_p$ depends on the ratio between the incident ion mass and the silicon atom mass. Heavier ions produce narrower profiles. Table 1.2 gives some typical values of X_p and ΔX_p associated with various dopants at different energy levels.[8,9]

TABLE 1.2. Mean Range X_p and Spread ΔX_p of Implanted Impurity Concentration for Various Dopants at Different Implant Energy Levels.[8,9]

Dopant Impurity	Implant Energy Level (keV)				
	30	50	70	100	200
Boron					
X_p (μm)	0.095	1.52	2.10	2.80	4.85
ΔX_p (μm)	0.0375	0.05	0.059	0.07	0.092
Phosphorus					
X_p (μm)	0.036	0.062	0.086	0.122	0.25
ΔX_p (μm)	0.0165	0.022	0.034	0.047	0.078
Arsenic					
X_p (μm)	—	0.031	0.042	0.057	0.11
ΔX_p (μm)	—	0.011	0.015	0.02	0.037

EX-1014 0029
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

The semiconductor surface can be easily masked against the implanted ions by using a surface oxide (SiO_2) layer as a mask. Thus, the ion-implanted regions can be readily patterned on the silicon surface, in the same manner as the diffused regions, using photomasking techniques. Since ion implantation is a low-temperature step, an unremoved photoresist layer can also be used to mask against implanted ions, instead of an SiO_2 layer.

Ion implantation provides an alternate method to diffusion for the introduction of dopant impurities into silicon. Thus, it gives the device designer an added degree of flexibility in the fabrication of monolithic devices. One key advantage of ion implantation is that it is a low-temperature process. Thus, it can be added to the manufacturing process without affecting the diffusion processes which precede it. It is particularly useful for forming very shallow junctions in silicon, or achieving well controlled impurity concentrations on the silicon surface.

The most frequent application for ion implantation is in controlling MOS device thresholds and fabricating junction-gate field-effect transistors (JFET) or precision resistor-ladder networks. The shallow device junctions formed by ion implantation are also very useful in the design of high-freqency transistors.

1.8. THIN-FILM PROCESSES

After completion of the epitaxy, diffusion, and ion-implantation steps, conductive, resistive, or dielectric thin-film layers can be deposited on the silicon wafer surface. These thin-film layers are then etched and patterned by the conventional photomasking techniques to perform a multiplicity of circuit functions. The conductive thin films, such as aluminum, are used for interconnection of the circuit components. Deposited resistive or dielectric layers can be used in forming thin-film resistors or capacitor structures. The term *thin film* is used to imply deposited film thicknesses of several micrometers or less, as compared with the larger geometry and thicker films associated with hybrid integrated circuits.

The deposition of dielectric films has been described earlier in connection with the surface passivation technology (see Section 1.5). In this section, particular attention will be given to the deposition of conductive and resistive films. Thin-film resistors formed by the deposition and patterning of resistive thin-film layers on the wafer surface have some distinct advantages over the conventional diffused resistors. Thin-film resistors, in general, exhibit lower temperature coefficients and offer a wide range of sheet resistivity values, which can be chosen independently of active device design requirements. Furthermore, in some cases they can be trimmed to a final value by postdeposition heat treatment or anodization techniques. In forming the resistor patterns, resistive thin films, such as tantalum (Ta), nickel–chromium (NiCr) alloys, or tin oxide (SnO_2), are the most commonly used materials. Electrical properties of thin-film resistors are discussed further in Chapter 3.

EX-1014 0030
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

For interconnection purposes, aluminum is the most commonly used thin-film material because of its high electrical conductivity and good adherence to the SiO_2 surface.

Deposition Techniques

Resistive or conductive thin films can be deposited on the silicon wafer surface by a variety of techniques. Some of these are outlined below.

Vacuum Evaporation. The passivated silicon substrate, together with the source of the material to be evaporated, is placed in a bell jar under high-vacuum conditions (10^{-5}–10^{-6} torr). The material to be evaporated is heated electrically by a tungsten or tantalum filament or by an electron gun until it vaporizes. Under the high-vacuum conditions used, the mean free path of the vaporized molecules is comparable to the dimensions of the bell jar. Therefore, the vaporized material radiates in all directions within the bell jar. Some of the vaporized material then deposits on the substrate, which is placed some distance from the source to ensure uniformity of deposition. The substrate is also maintained at an elevated temperature to provide a good adhesion of the deposited film.

Both conductive and resistive films can be deposited by vacuum evaporation. Aluminum, gold, and silver are among the conductive films formed in this manner. Nickel–chromium resistors can also be deposited by vacuum evaporation techniques, except in this case, due to high power densities required to vaporize the source, electron-beam bombardment, rather than thermal heating of the source material, is used. The films deposited by vacuum evaporation exhibit a fine-grained structure. The grain structure of the film becomes finer as the evaporation rate is increased and the angle of incidence of the radiating vapor on the wafer surface is made steeper. For uniformity and repeatability of film properties, a fine-grained structure is desirable.

Cathode Sputtering. The sputtering process takes place in a low-pressure gas atmosphere. A glow discharge is formed by applying a high voltage (typically 5000 V) between the cathode and the anode sections of the sputtering apparatus. The cathode is coated with the material to be evaporated, and the substrate is attached to the anode or placed within the glow-discharge region. Normally an inert gas such as argon (A) is used as the sputtering medium. The A^+ ions generated by the glow discharge accelerate toward the cathode due to the negative cathode potential. When these high-energy ions impinge on the cathode, they cause the atoms or the molecules of the cathode to break away, or sputter, from the surface. Then some of these cathode particles which float away are intercepted by the substrate and deposit in the form of a thin layer. Under the low-vacuum conditions used in sputtering, the mean free path of the source atoms is much shorter than the source-to-substrate spacing. Therefore, the depo-

EX-1014 0031
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

sition rates in the sputtering process are much slower than in vacuum evaporation.

By adding small amounts of reactive gases, such as oxygen or nitrogen, to the inert argon atmosphere, the chemical composition of the deposited layer can be modified. This is known as *reactive sputtering* and is particularly useful for tantalum deposition.

Vapor Phase Deposition. In vapor or gas phase deposition, halide compounds of the material to be deposited are chemically reduced, and the resulting metal atoms are deposited on the substrate. This basic deposition process very closely resembles the epitaxial growth step of the planar process. Vapor deposition is particularly useful for obtaining thick layers of deposited films (up to 20 μm). It is commonly utilized for forming aluminum oxide ($Al_2O_3 \cdot SiO_2$) dielectric layers or SnO_2 resistive films. The sheet resistance of SnO_2 films can be controlled by introducing Group III or Group V ions [such as indium (In) or antimony (Sb)] to increase or reduce the sheet resistance. In this manner, sheet resistances in the range of 100–5000 ohms per square ($\Omega/\square$) can be obtained.

Patterning and Etching of Thin Films

With minor modifications, the basic photomasking and etching techniques described in Section 1.6 can be utilized in patterning the thin-film components. One significant exception is the case of multiple thin-film layers (such as aluminum interconnections over thin-film resistors) where additional care should be taken in the choice of the etchant to ensure that the bottom film is not damaged by the patterning of the top layer.

In the case of Al, SnO_2, Ta, or $Al_2O_3 \cdot SiO_2$ layers, patterning and etching can be achieved by direct photoresist techniques. In the case of very thin (300–500-Å) nickel-chromium films, an *inverse metal-masking* technique can be used. In this process, a thin layer of metal film (typically copper) is deposited and etched into an inverse, or negative, of the desired final metal pattern. Then the desired thin-film layer is deposited on this inverse metal pattern. In the final etching step, the inverse metal pattern of the initial metal film is etched away, taking with it the layer of desired metal deposited on it, and only the portions of the desired metal layer that adhere directly to the substrate are left behind.

Interconnections and Ohmic Contacts

The basic prerequisite for the conductive films used for interconnections is that they should make good ohmic contact with the diffused components or other metallic films deposited on the device surface. A good ohmic contact is defined as one that exhibits a linear current–voltage (I–V) relationship which passes through the origin of the I–V characteristic. In a great majority of IC applications, aluminum is used as the interconnection layer. This is because of its ease

EX-1014 0032
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

of deposition and patterning, as well as its high degree of conductivity and its ability to form ohmic contacts with silicon.

The exposure of contact areas to the ambient atmosphere often results in the formation of parasitic oxide layers over the chip areas to be interconnected. Therefore, to provide good ohmic contact, the interconnecting metal must be chemically active so that it can be "alloyed" through these parasitic oxide layers. The most commonly used interconnection metal is aluminum. It can be readily alloyed into the silicon substrate to form ohmic contacts. Since aluminum is a p-type dopant (see Table 1.1), to avoid the formation of a nonohmic rectifying contact, the contact areas on the lightly doped n-type semiconductor regions are n^+ doped prior to metallization (see, for example, the collector contacts of the npn bipolar transistors of Figs. 1.13 and 1.14). The heavy n^+ doping causes a high degree of damage to the silicon lattice at the surface. Therefore, the parasitic pn junction formed by the alloying of a p-type aluminum interconnection into n-type silicon is very leaky and nearly ohmic in its conduction properties.

In conventional monolithic IC fabrication, the alloying of the aluminum interconnections into silicon is the last step of the planar process. It is normally accomplished by a short heat treatment in an inert atmosphere, typically about 10 min at 500°C.

A troublesome metal interconnection problem can occur in devices which employ two active dissimilar metals in their interconnection scheme. At the interface of two dissimilar metals, parasitic intermetallic compounds and oxides can form. A typical example of this is the intermetallic gold–aluminum compounds forming between the aluminum bonding pads and the gold wires that may be used to connect the chip to the package terminals. These compounds, which are brittle and nonconductive, are commonly referred to as the "purple plague" because of their dark color. Under certain circumstances (see Section 1.14) this may be a detriment to the reliability of IC interconnections using gold-wire bonds and aluminum bonding pads.

Although aluminum is a good conductor, it still introduces a finite amount of series resistance into the device interconnections. Typical resistivity of aluminum is of the order of 2.8×10^{-6} ohm-cm. In the case of a typical aluminum interconnection trace of 1-μm thickness, this corresponds to a sheet resistance of approximately 0.03 $\Omega/\square$.

Under very high current densities, conductive thin films such as aluminum exhibit a failure mode due to the so-called *electromigration* effect. It causes the metal atoms to gradually migrate away from the high-current-density points within the conductor. This is a progressive failure mode, under continuous operation, and comes about from the momentum exchange between the conducting electrons and the stationary metal atoms. Electromigration is a slow process which speeds up as the current density or the temperature is increased. It starts as a formation of localized *voids*, or gaps, in the conducting metal strip and eventually leads to a complete open circuit at the point of highest current density within the conductor strip.

EX-1014 0033
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

To avoid electromigration effects, current densities in the IC interconnection paths are normally kept at less than the 10^5-A/cm^2 level. In the case of a typical aluminum interconnection path of 1-μm thickness, this corresponds to a maximum allowable continuous current of approximately 50 mA per mil width of the interconnection path.

1.9. BIPOLAR INTEGRATED-CIRCUIT FABRICATION STEPS

The fabrication of a bipolar integrated circuit involves a sequence of five to eight masking and diffusion steps. The sequence of some of these basic steps is illustrated in Figure 1.11 for the case of an *npn* bipolar transistor and a *p*-type diffused resistor.

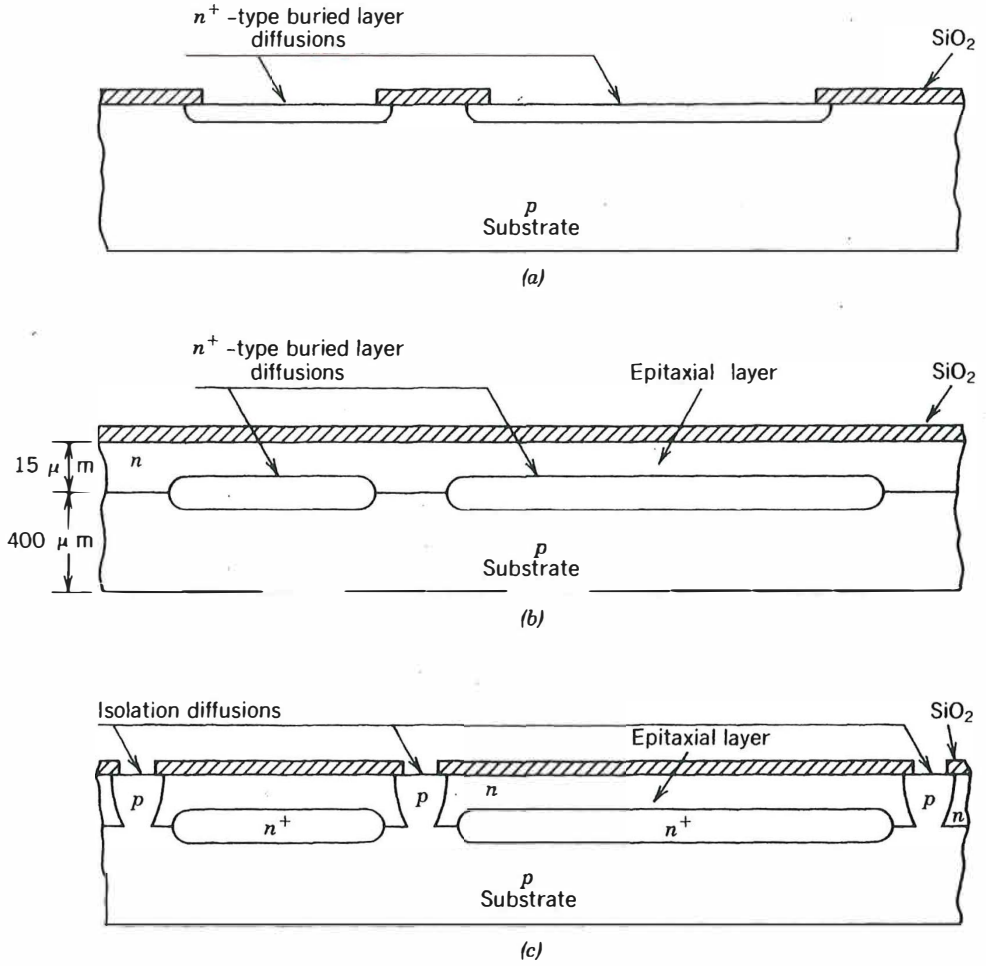
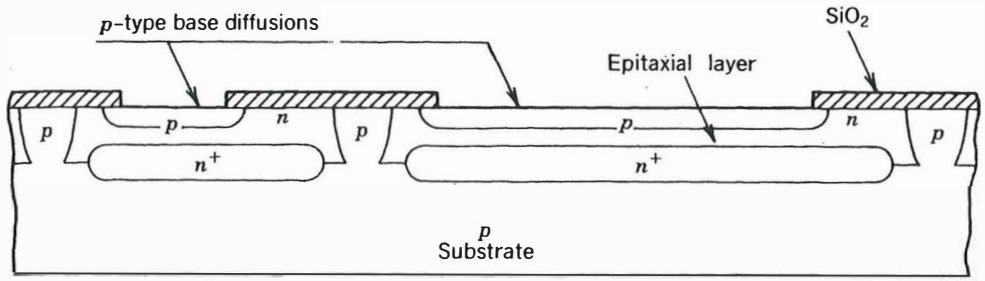
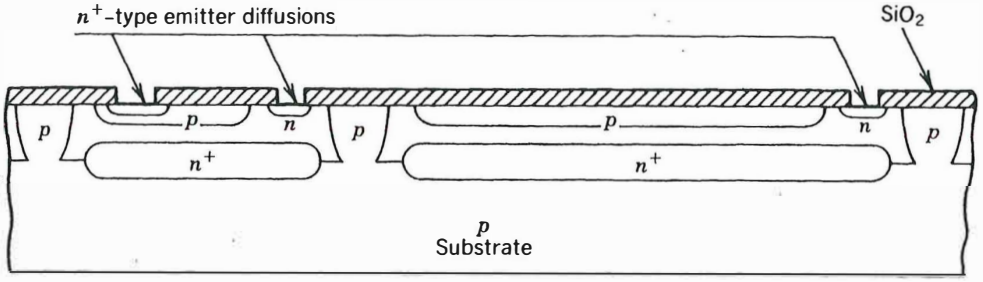


FIGURE 1.11. Basic sequence of steps in the fabrication of a bipolar monolithic circuit: (*a*) Initial oxidation and buried layer diffusion; (*b*) expitaxial layer growth and second oxidation; (*c*) isolation diffusion; (*d*) base diffusion; (*e*) n^+ emitter diffusion; (*f*) contact windows and interconnections.

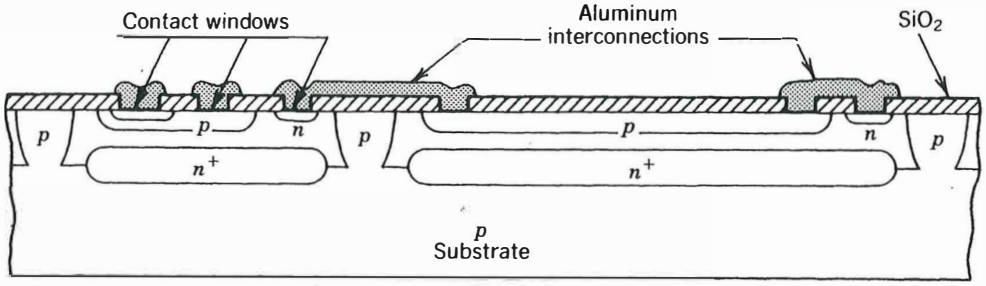
EX-1014 0034
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.



(d)



(e)



(f)

The starting material is a wafer of p-type silicon, typically 4 inches in diameter and approximately 400 μm thick, with a typical impurity concentration of 10^{16} atoms/cm^3. The first masking and diffusion step, illustrated in Figure 1.11*a* forms a low-resistivity n^+-type layer* which will eventually form a low-resistance current path within the collector region of the resulting *npn* transistor. Since this heavily doped n-type layer will be covered by the epitaxial layer, it is called a buried layer, and the corresponding diffusion is called the buried layer diffusion. The sheet resistance of the buried layer diffusion is in the range of 10–30 Ω/□, and the impurity used is either arsenic or antimony. These impurities

*In this context, the + sign is used to imply heavy impurity concentration, not electrical charge.

EX-1014 0035
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

are chosen because they diffuse slowly and, thus, do not redistribute significantly during the subsequent diffusion steps.

After the buried layer diffusion step, the oxide covering on the wafer is stripped and an n-type epitaxial layer is grown over the entire wafer surface, as shown in Figure 1.11*b*. The thickness and the impurity concentration of this epitaxial layer determine the breakdown voltage of the resulting transistor structure. Assuming a minimum transistor breakdown voltage LV_{CEO} of 30 V is required, the epitaxial layer will be chosen to be approximately 15 μm thick, with an impurity concentration of 2×10^{15} atoms/cm^3, which corresponds to a resistivity of approximately 2.5 Ω/cm (see Fig. 1.2). Note that, as shown in Figure 1.11*b*, the n^+-type buried layer also out-diffuses somewhat into the epitaxial layer during this process.

Following the epitaxial growth, an oxide layer is formed on the wafer surface. Then, after a masking step, a p-type (boron) diffusion is made, as shown in Figure 1.11*c*. The function of this diffusion is to form the deep p-type *isolation walls,* which reach through the n-type epitaxial layer into the p-type substrate. Because of the depth to which this diffusion must penetrate, it requires several hours of diffusion time at temperatures in excess of 1200°C. The sheet resistance of the p-type isolation diffusion is in the range of 20–40 $\Omega/\square$. Note that the n^+-type buried layer diffusion is omitted in the regions directly under the isolation diffusion, and that the n^+-type buried layer does *not* touch the p-type isolation wall. This is done to avoid forming a low-breakdown p-n junction between the n-type *tub* and the p-type isolation; and to ensure that the p-type isolation wall can reach down to the p-type substrate, thus forming a continuous wall surrounding the n-type tub.

The next masking and diffusion step (Fig. 1.11*d*) forms the p-type base region of the *npn* transistor. It results in a sheet resistance in the range of 100–200 $\Omega/\square$ and a junction depth of 1–3 μm. Since this diffusion also forms many of the resistors in the circuit, its sheet resistance is closely controlled to be within $\pm 20\%$ of the target value.

Following the base diffusion, the n-type emitter regions of the *npn* transistor are formed by the emitter mask and the subsequent emitter diffusion step, as shown in Figure 1.11*e*. Normally, phosphorus is used as a dopant for the emitter diffusion. The resistance of the emitter diffusion is of the order of 2–10 $\Omega/\square$, and the resulting junction depth is in the range of 0.5–2.5 μm. Since the difference in the junction depth of the base and the emitter diffusions determines the base width of the *npn* transistor, the depth of the emitter diffusion is controlled to be approximately 0.5–1 μm, *less* than that of the base diffusion. The heavily doped (n^+-type) emitter diffusion also serves as a low-resistance contact to the n-type epitaxial layer to form the ohmic collector contact for the transistor. This is necessary because a direct ohmic contact between the aluminum interconnection and the lightly doped n-type epitaxial region is difficult to form (see Section 1.8).

After the emitter diffusion, the wafer undergoes a masking step called the *contact mask*, which opens all the contact windows over all the passive and

EX-1014 0036
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

active devices on the chip. Then the entire wafer is coated with a thin layer of aluminum (0.5–1 μm), which will form a conductive interconnection path between the devices. Then, in the next masking step, called the *metal mask*, the aluminum is etched away, leaving behind the desired interconnection pattern between the components on the chip. The resulting device structure is shown in Figure 1.11*f*.

Figure 1.12 shows a plane view of the completed device structure of Figure 1.11*f*. For illustrative purposes, the collector of the *npn* transistor is assumed to be connected to the resistor to form a simple common-emitter amplifier stage.

Figure 1.13 shows the typical impurity concentration within the vertical cross section of the resulting transistor structure. Assuming a base depth of 3 μm and an emitter depth of 2.5 μm, a base width of $\approx$0.5 μm is achieved. Also, note that the n^+-type buried layer has out-diffused approximately 6 μm from the nominal epi–substrate interface.

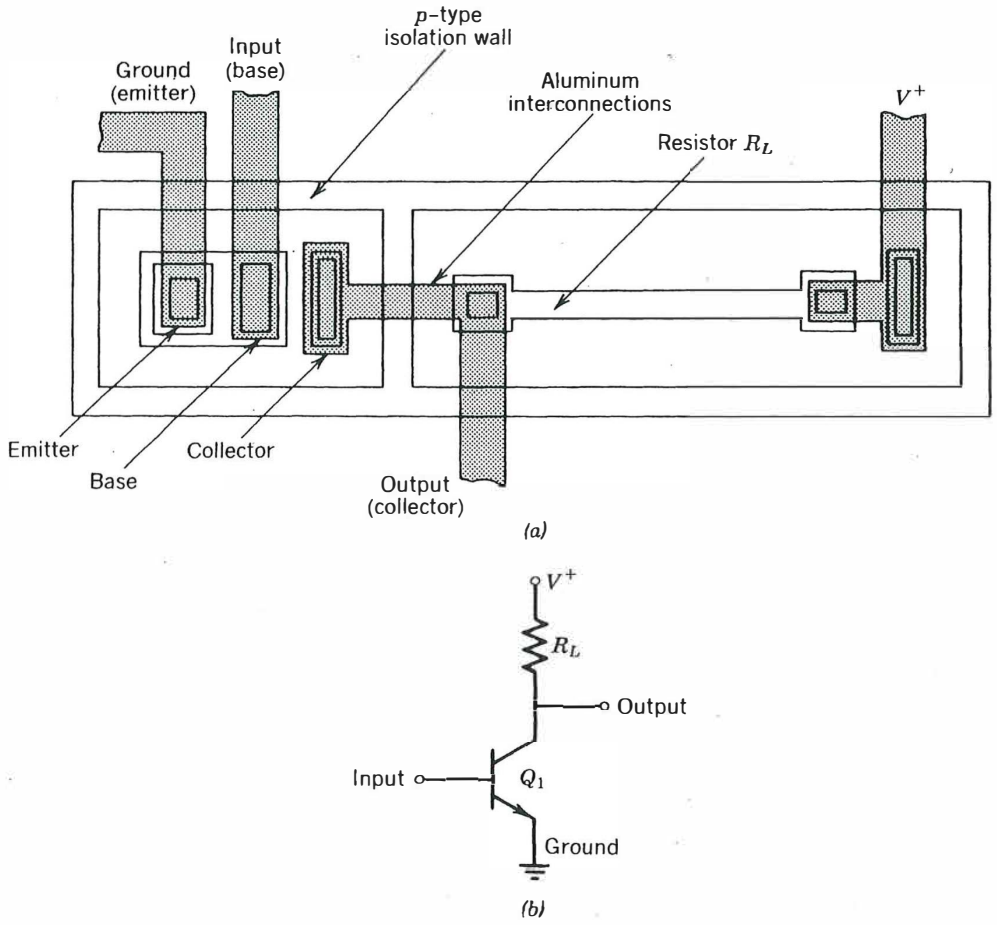


FIGURE 1.12. Plane view of *npn* transistor and resistor combination shown in Figure 1.11*f*, illustrating its connection as a common-emitter gain stage: (*a*) Plane view of circuit layout; (*b*) equivalent circuit.

EX-1014 0037
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

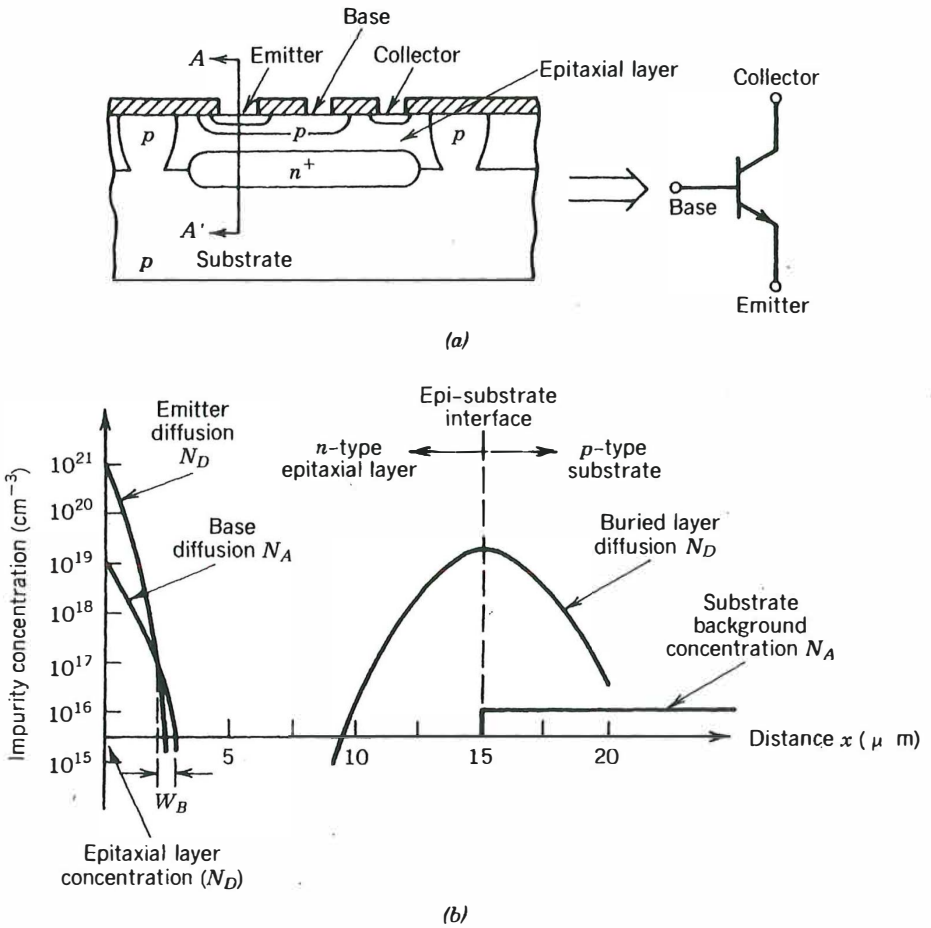


FIGURE 1.13. Device cross section and typical impurity profile for a monolithic *npn* transistor: (*a*) device cross section; (*b*) impurity profile along A–A^+.

In the normal fabrication process, subsequent to the metal mask step, the silicon wafer undergoes an *alloy* or *sinter* step at low temperatures (typically 5–10 min at 450°–500°C) to assure proper ohmic contact between the aluminum interconnections and the contact windows on the wafer surface. Finally, a dielectric oxide layer is deposited on the wafer surface using a low-temperature CVD process (see Section 1.5) to form a passivation layer over the entire chip surface, which also covers the aluminum interconnections. Finally, a passivation mask is applied to the silicon wafer to etch large windows on top of the bonding pads on the IC chip, to which external connections will be made. The purpose of this passivation layer is to protect the chip surface from dust, mechanical damage, and scratches.

After completion of the sequence of fabrication steps described above, the finished silicon wafer is ready for evaluation and the functional electrical testing of the individual circuit chips.

Normally, the wafers proceed through the fabrication process steps in groups

EX-1014 0038
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

called *lots*, with typically 15–25 wafers to a lot. Thus, many wafers, each containing hundreds of complex circuit chips, are manufactured simultaneously. This batch processing capability of the planar process, which enables one to fabricate thousands of circuits in one process sequence, is the key to the economic advantages of integrated circuits.

1.10 MODIFICATIONS OF BASIC PROCESS

The vast majority of the monolithic bipolar analog circuits presently in production are fabricated using the basic sequence of steps illustrated in Figure 1.11. However, when certain performance characteristics are required, additional steps can be incorporated into the basic manufacturing sequence to enhance such characteristics. Some of these modifications will be briefly discussed in this section. Although these modifications improve specific device characteristics, they often require additional process steps or tighter process controls and, thus, add to the manufacturing cost of the monolithic circuit.

Dielectric Isolation

In certain circuit applications, the parasitic capacitances, leakage currents, and breakdown characteristics associated with the conventional junction isolated device structures may not be acceptable. Typical examples of such applications are circuits that must operate at high frequencies or very high voltages (i.e., in excess of 100 V) or circuits that must withstand high amounts of ionizing radiation (such as gamma rays) without suffering permanent damage. In such cases, superior electrical isolation can be obtained by surrounding each n-type pocket or tub not with a reverse-bias p-n junction, but with a dielectric layer. Normally, thermally grown SiO_2 is used as the dielectric material.

In forming the dielectrically isolated pockets on the wafer surface, a number of alternate fabrication techniques can be utilized.[9] Figure 1.14 shows a typical sequence of fabrication steps in forming the dielectrically isolated single-crystal silicon pockets or islands. Starting with an n-type substrate, a nonselective n^+-type layer is diffused into the wafer surface. For reasons which will be explained shortly, a $< 100 >$ oriented crystal is utilized as the starting material, as opposed to the $< 111 >$ oriented crystal normally used for junction isolation.*

Following the initial n^+ diffusion, the wafer surface is oxidized, and a mirror image mask of the desired isolation grid pattern is applied to the wafer to remove the oxide along the isolation grid. The exposed silicon surface is then etched by

*During the initial silicon crystal growth, the <111> oriented crystal is easier to grow defect-free than the <100> orientation and is, therefore, somewhat less expensive. In addition, the <111> crystal is less affected by surface contamination than the <100> crystal. These are the main reasons for the choice of the <111> oriented substrate for most bipolar device structures.

EX-1014 0039
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

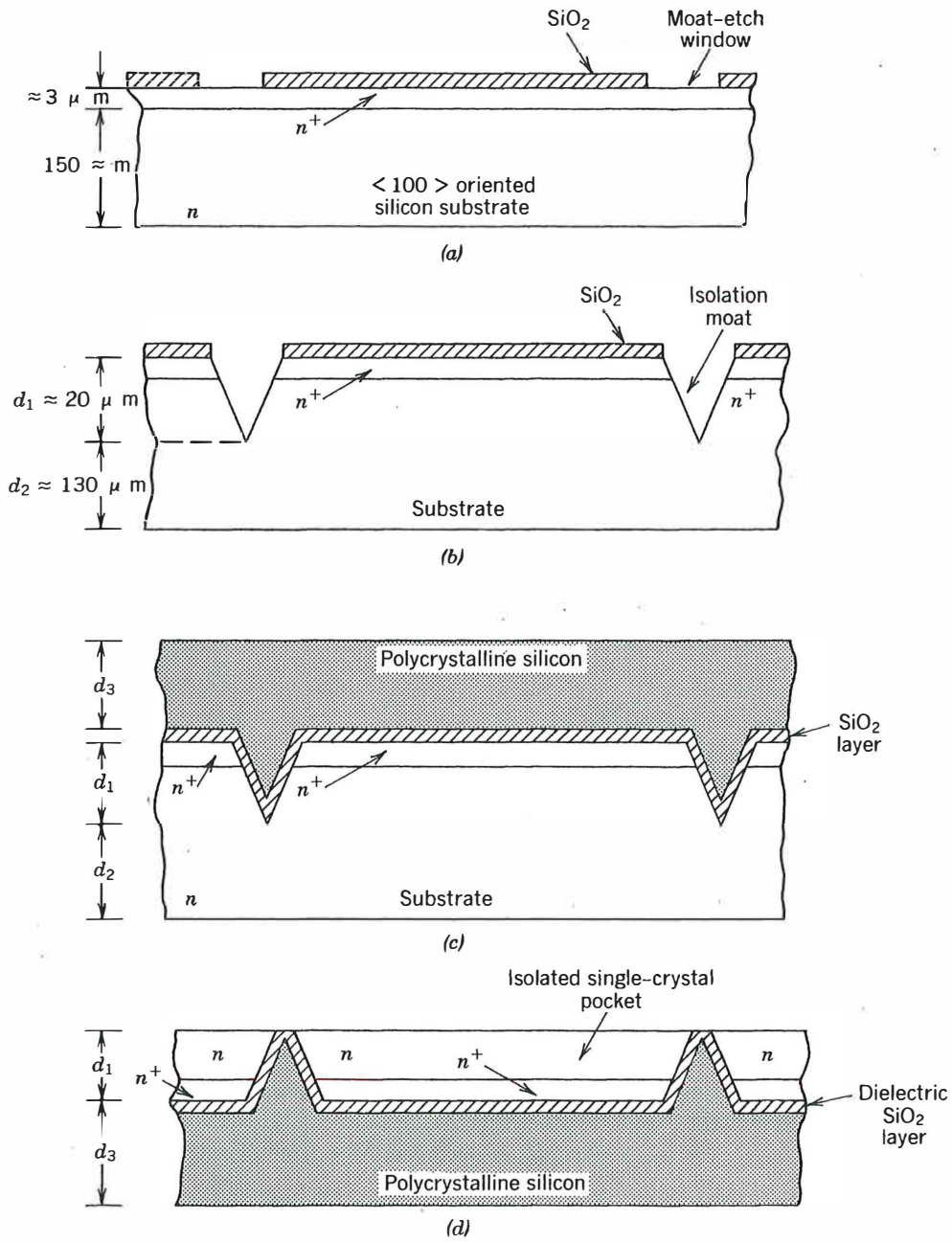


FIGURE 1.14. Sequence of steps in dielectric-isolation process: (*a*) n^+ layer diffusion and opening of moat-etch windows; (*b*) etching of isolation moats; (*c*) forming SiO_2 dielectric layer and growing polycrystalline silicon layer; (*d*) forming isolated *n*-type pocket by grinding away substrate until peak of V groove appears.

EX-1014 0040
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

a potassium hydroxide (KOH) based etch. The etchant used in this step etches away the exposed silicon anisotropically, that is, the etch rate is much faster along the [111] planes than along the [100] crystal planes. This preferential etching results in formation of a V-shaped isolation *groove* or *moat* on the wafer surface, as shown in Figure 1.14*b*.

Referring to Figure 1.14*b*, it should be noted that the vertical dimensions of the drawings are not to scale, and the depth d_1 of the isolation moat is a small fraction of the total wafer thickness. After the preferential etching step, the exposed silicon is reoxidized, and a thick layer of polycrystalline silicon is deposited over the oxide layer, as shown in Figure 1.14*c*. The thickness and the electrical characteristics of this polycrystalline silicon layer are of no consequence since its main function is to serve as a mechanical support for the wafer. Then the original wafer is flipped around, and the bottom surface of Figure 1.14*c* is now corresponding to the top of the device structure. Then the single-crystal n-type layer of thickness d_2 is backlapped until the isolation grid appears on the wafer surface, resulting in the isolated n-type single-crystal pocket shown in Figure 1.14*d*. After the isolated pockets are formed, the fabrication of integrated devices within the pockets is completed by a sequence of conventional masking and diffusion steps, resulting in the isolated device structure of Figure 1.15.

In some cases, it may be desirable to reverse the order of the initial n^+-type layer which covers the sides as well as the bottom of the isolated pocket. Such a structure offers lower ohmic resistance directly below the base. However, in this case a larger clearance is required between the p-type base and the dielectric sidewall to prevent the n^+-type layer from touching the base-diffused regions.

The basic sequence of processing steps has been known for quite some time. However, the process has become practical only after the development of anisotropic etch techniques,[10] which allow very accurate control of the etch depths by the width of the initial oxide window.

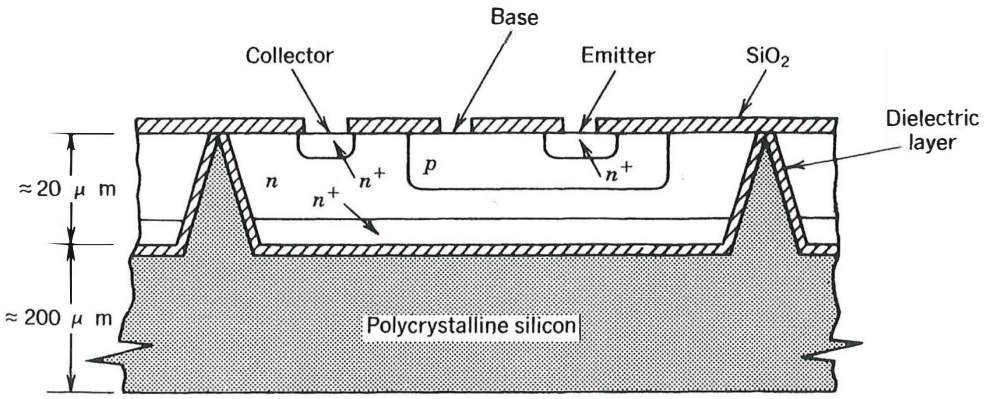


FIGURE 1.15. Cross section of a dielectrically isolated n-type pocket containing a *npn* transistor.

EX-1014 0041
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Deep n^+ Diffusion

As will be described in Chapter 2, the basic bipolar transistor structure of Figure 1.11 has a relatively large parasitic resistance in series with its active collector area. Most of this resistance is due to the resistivity of the n-type epitaxial region. This parasitic series resistance can be significantly reduced by extending the heavily doped (i.e., low-resistivity) n^+-type collector contact region all the way down through the epitaxial layer into the n^+-type buried layer. In practice, this can be achieved by means of a deep n^+ diffusion which is performed subsequent to the p-type isolation diffusion and prior to the p-type base diffusion (i.e., between steps *c* and *d* of Fig. 1.11). The resulting deep diffusion is called an n^+ sinker, plug, or sidewall diffusion.

Figure 1.16 shows the cross section of a *npn* transistor with such a deep n^+ diffusion. Although this step adds an extra masking and diffusion cycle to the basic fabrication sequence, its effect on the overall manufacturing cost is minimal.

Up-Down-Diffused Isolation

The p-type isolation diffusion is the longest diffusion cycle during the bipolar IC fabrication process. This is because the depth of the isolation diffusion must be sufficient to penetrate the entire thickness of the n-type epitaxial layer, which can be anywhere from 5 to 25 μm, depending on the particular fabrication process used, which in turn depends on the transistor breakdown voltage requirements. Since the diffusion process proceeds sideways as well as downward into the silicon, the side diffusion of the isolation wall in turn determines the base-to-isolation spacing of a *npn* transistor. As shown in Figure 2.4, the base-to-isolation spacing, in turn, is the major contributor to the chip area taken up by a small-geometry *npn* transistor.

The minimum base-to-isolation spacing can be significantly reduced by using the so-called *up-down isolation diffusion* technique illustrated in Figure 1.17. In this method, the isolation wall is diffused simultaneously from the epi–substrate interface as well as from the wafer surface. This is achieved by forming a p^+-type subepitaxial layer directly under the isolation wall, prior to the epitaxial

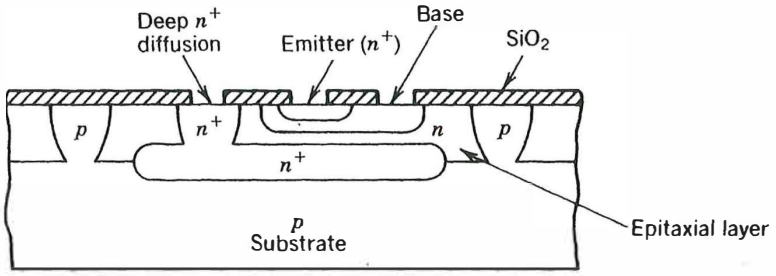


FIGURE 1.16. Cross section of *npn* transistor structure using deep n^+ diffusion.

EX-1014 0042
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

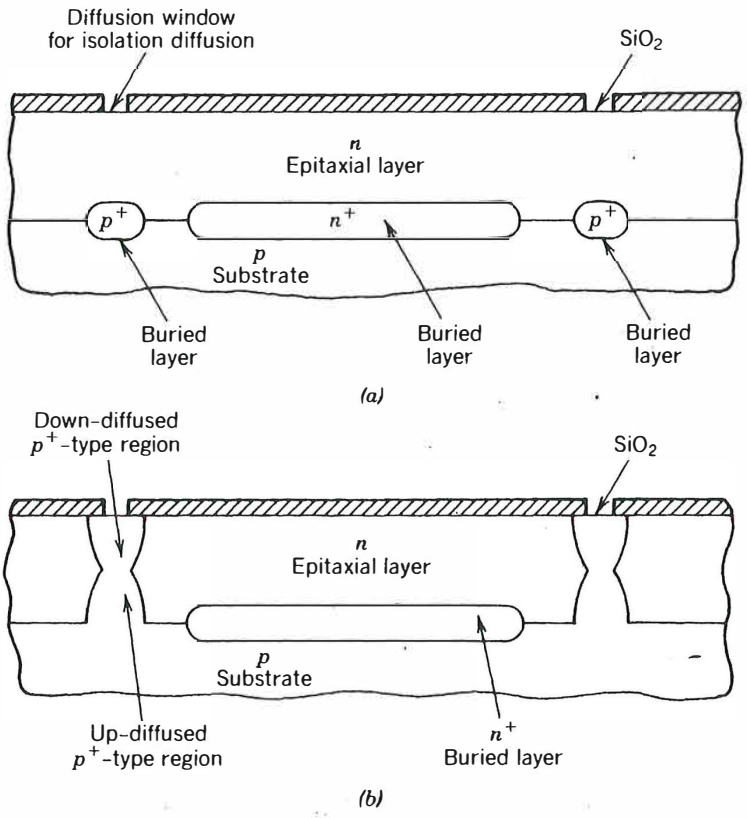


FIGURE 1.17. Forming an isolation wall with simultaneous upward and downward diffusion of p-type impurities. Device cross section (*a*) before isolation diffusion and (*b*) after isolation diffusion.

growth. In other words, both the p^+-type and the n^+-type buried layer regions are formed before the epitaxy process. This results in the device cross section shown in Figure 1.17*a*, prior to the start of the conventional isolation diffusion step. Then, during the isolation diffusion, the p^+-type subepitaxial layer out-diffuses into the epitaxial layer and meets the down-diffusing isolation wall at approximately halfway in the epitaxial layer to complete the isolation wall, as shown in Figure 1.17*b*. In this manner, the total depth of the top-diffused isolation wall is only 50–70% of the epitaxial layer thickness, which results in a very significant reduction of isolation side diffusion. As a result, the minimum allowable base-to-isolation spacing can be reduced significantly, allowing a more compact device layout.

An added benefit of the up-down-diffused isolation is that the total isolation diffusion time is reduced, resulting in the reduction of the undesired out-diffusion of the n^+-type buried layer into the epitaxial region. This out-diffusion is undesirable since it reduces the epitaxial layer resistivity and lowers the transistor collector–base breakdown voltage.

The up-down isolation, similar to the case of deep n^+ diffusion, adds one extra masking and diffusion step to the conventional bipolar fabrication process

EX-1014 0043
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

shown in Figure 1.11. However, it is a relatively noncritical additional step and can result in significant chip area reduction (typically of the order of 15–20% of the active chip area) in high-voltage integrated circuits with 20-μm or thicker epitaxial layers.

Two-Step Emitter Diffusion

The common-emitter current gain β_F of bipolar transistors is a strong function of the transistor base width and increases very rapidly as the base width is reduced. Unfortunately, the transistor breakdown voltage BV_{CEO} is also reduced greatly as the base width is reduced (see Section 2.1). In certain applications, such as in designing the input stages of monolithic operational amplifiers, it is necessary to have very high-gain (i.e., *superbeta*) transistors on the same chip as the conventional medium-gain but high-breakdown *npn* transistors. Superbeta transistors exhibit a current gain β_F on the order of 2000–5000, with a breakdown voltage of 2–5 V, whereas the conventional *npn* transistors would have $\beta_F \approx 200$–500, with a breakdown voltage of 30–50 V.

The superbeta transistors can be fabricated simultaneously with conventional *npn* transistors by using a two-step emitter diffusion process: the emitters of superbeta transistors are diffused first, and the emitters of conventional *npn* transistors are then masked and diffused as the next step. Thus, the emitter regions of the supergain transistors undergo a *longer* diffusion cycle, and since both transistors have the same base diffusion, the superbeta transistor ends up with a *deeper* emitter diffusion, resulting in a *narrower* base width. The basic two-step emitter diffusion process is also used for fabricating double-diffused JFETs on the same chip as *npn* bipolar transistors (see Fig. 2.36).

Figure 1.18 shows the structure of a superbeta transistor side by side with a conventional *npn* transistor. The two-step emitter diffusion process requires one additional masking and diffusion step, as well as tighter process controls, than the basic bipolar process.

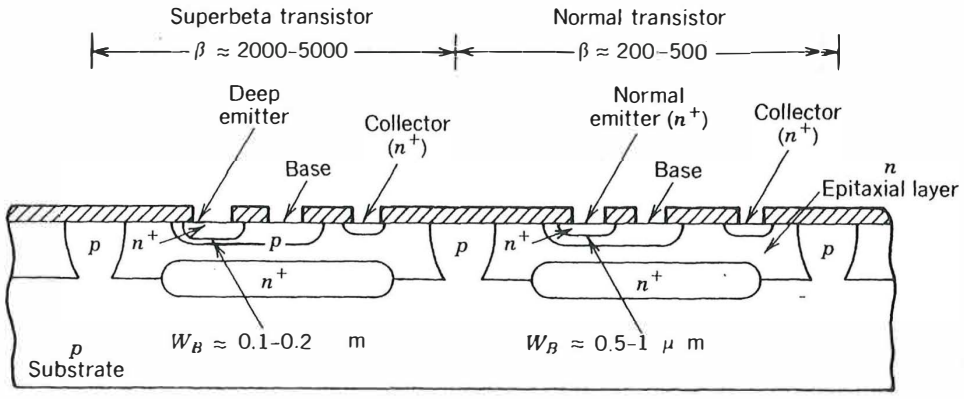


FIGURE 1.18. Superbeta transistor using two-step emitter diffusion.

EX-1014 0044
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

In certain applications, it is also possible to replace the two-step emitter process with a two-step base process, where the base can be diffused to a *shallower* depth. Thus, the same emitter diffusion applied to both devices shown in Figure 1.18 would then result in a narrower base width for the superbeta transistor.

The applications and the electrical characteristics of superbeta transistors are described further in Chapter 2.

Ion Implantation

As described in Section 1.7, ion implantation provides another alternate method to diffusion for introducing impurities into silicon. Since it is a low-temperature process, it is normally implemented *after* all the diffusion steps have been completed (i.e., between steps *e* and *f* of Fig. 1.11). It is normally used to form high-value *p*-type resistors, bipolar compatible JFET (BIFET) or MOS structures, or high-frequency integrated injection logic (I^2L) gates. Ion implantation normally requires one masking and one implantation step, followed by a low-temperature ($\approx$ 500–600°C) heat treatment. In certain cases, such as bipolar compatible JFETs, two ion-implantation steps are used.

Double-Layer Metallization

In certain complex circuits, it may be necessary to use two separate interconnection layers on the chip. This is done by first forming the initial layer of interconnections (see Fig. 1.11*f*) and then depositing a layer of dielectric. Then interconnecting windows, or *via holes* are etched through this dielectric. Next, a second layer of metal is evaporated and etched to provide a second layer of interconnections, which connect to the first layer through the via holes cut in the deposited dielectric layer. Double-layer metallization is a difficult process requiring careful process control.

Thin-Film Resistors

Thin-film resistors are deposited and etched on the oxidized silicon surface prior to the aluminum evaporation and etching. Formation of thin-film resistors requires a well-controlled deposition step (by either vacuum evaporation or sputtering) followed by a separate masking and etching step.

High-Performance Complementary Transistors

Monolithic bipolar fabrication technology is designed around the *npn* bipolar transistor. This basic process also produces a simple low-frequency *pnp* device, called the *lateral pnp transistor*, which is covered in Section 2.3. However, in certain applications, it is required to have high-performance *pnp* devices with the gain and frequency response characteristics comparable to those of normal *npn*

EX-1014 0045
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

transistors. This can be accomplished by adding a number of additional diffusion and masking steps to the basic fabrication sequence shown in Figure 1.11.[11, 12] However, in practice this is at best a compromise solution which has a significant impact on manufacturing costs and yields because of the following factors:

1. The basic device structure of Figure 1.11 requires only *one* critical diffusion step (i.e., the n^+ emitter diffusion) to control the *npn* transistor characteristics. If additional diffusion steps were provided to form a *pnp* transistor, that transistor would in turn require at least one critical diffusion step to set its base width. Since all diffusions proceed simultaneously, it is very difficult to perform two independent critical diffusion steps under a continuous high-volume production environment.
2. Adding additional masking and diffusion steps increases the amount of handling of the silicon wafer during the fabrication process, which in turn increases the chances of creating circuit defects, parameter variation, wafer surface contamination, wafer breakage, and operator error.

Economic Considerations

Many variations and modifications can be added to the basic bipolar IC fabrication sequence of Figure 1.11 to obtain specialized or improved device characteristics, or to obtain complementary or dissimilar device combinations. However, in practice, particularly in a high-volume production environment, the economic considerations (i.e., the manufacturing costs and the resulting yield of good circuits at the end of the wafer fabrication process) are of paramount importance. Thus, the performance benefit of any process modification has to be carefully judged against the resulting increase in manufacturing cost and complexity.

Table 1.3 gives the approximate relative manufacturing costs of various modifications to the basic process described in this section. For comparison purposes, the basic four-diffusion, seven-mask manufacturing process illustrated in Figure 1.11 is taken as a reference (i.e., as unity). The relative complexity figures indicated give a rough estimate of the relative manufacturing cost of one completely finished wafer of silicon under high-volume manufacturing conditions. It should be noted, however, that the complexity figures given in Table 1.3 are somewhat subjective (i.e., they reflect the view of the author) and are subject to change as the IC fabrication technology continues to evolve.

1.11. ASSEMBLY AND PACKAGING

Fabrication of the monolithic circuit on the surface of a silicon wafer represents only one part of the total manufacturing process. Additional asembly and packaging steps are required to make the circuit electrically functional.

EX-1014 0046
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

TABLE 1.3. Estimate of Relative Manufacturing Costs for Various Bipolar IC Fabrication Processes

Process Description	Estimated Relative Cost Factor[a]
Basic bipolar process (Fig. 1.11)	1.0
Dielectric isolation (Fig. 1.14)	2 to 4
Deep n^+ diffusion (Fig. 1.16)	1.1
Up-down isolation diffusion (Fig. 1.17)	1.15
Two-step emitter diffusion (Fig. 1.18)	1.3
Ion implantation (single implant)	1.2
Bipolar compatible JFET (using two ion-implant steps)	1.3
Thin-film resistors (untrimmed)	1.8
Thin-film resistors (laser trimmed)	2.5
Double-layer metallization	1.5–2

[a]Relative cost factors do not add linearly. If combinations of these modifications are used in a given process, the resulting relative cost factor is usually higher than the sum of the corresponding cost factors.

The wafer fabrication steps described up to this point represent the most efficient part of the IC fabrication process since they allow thousands of IC chips to be fabricated simultaneously through batch processing methods. However, once the wafer processing is complete, each chip on the wafer must be individually tested. Then the wafer is separated into *chips* or *dice,* and each circuit must be handled and packaged individually. Figure 1.19 gives a flowchart of the manufacturing steps that follow the wafer fabrication operation.

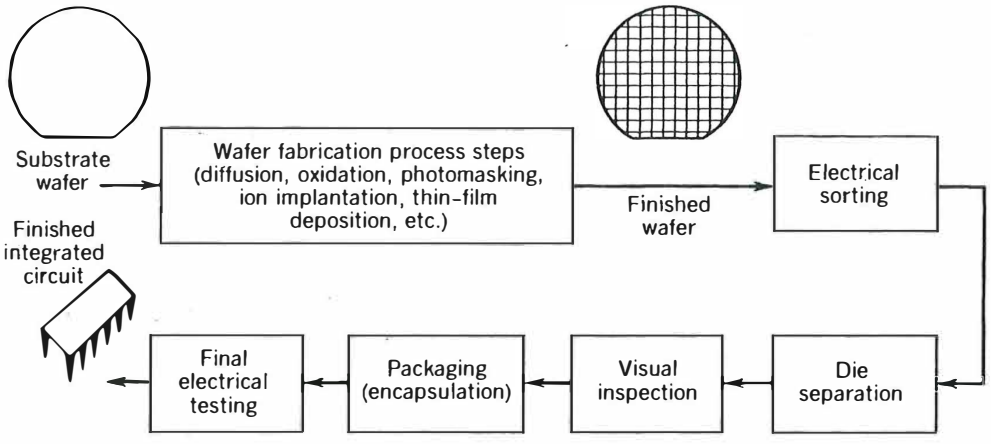


FIGURE 1.19. Flowchart of IC manufacturing steps.

EX-1014 0047
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Electrical Sorting

After the wafer fabrication steps and prior to encapsulation, each of the individual circuit chips on the wafer has to be electrically tested to ensure that it meets the desired electrical performance requirements. Since the assembly process is a rather costly step in the overall IC manufacturing process, it is necessary to screen the monolithic chips for defects and electrical failures while the units are still on the wafer. For this purpose, an automatic probing station is utilized. During the electrical probing operation, the bonding pads on each circuit chip are contacted by needle-like metallic probes, and various current and voltage levels at the terminals of the circuit are measured. Normally, these electrical tests are performed by programmed automatic testers which can perform up to 100 tests per second on a single chip. If the electrical characteristics of the circuit are not acceptable, an automatic marking pen at the probing station is activated to mark the unit as a reject. When the test sequence for a given circuit is completed, the wafer is automatically indexed, and the probes advance to the next chip. In this manner, the entire wafer is automatically sorted, and the rejected devices are marked with ink dots.

Die Separation

Subsequent to electrical sorting, the wafer is cut along the rectangular grids separating the individual chips or dice. A number of methods have been developed for separating the individual chips. The most commonly used ones are scribing the wafer with a diamond-tipped cutting tool, called *scriber*, or using a circular diamond-tipped saw to cut a series of deep grooves into the wafer along the rectangular grid separating the chips. Subsequent to the scribing or sawing operation, the wafer is fractured along the scribed channels, or grooves, into physically separated chips. Recently the use of a laser beam, rather than a diamond saw or scriber, has also gained acceptance in the IC industry. Laser scribing is preferred over conventional scribing techniques since it is a much faster process and eliminates accidental breakage, chipping, or cracking of the individual dice during the die separation step.

Visual Inspection

After die separation, the electrically good chips (i.e., those with no ink marks) are visually separated from the rest of the chips on the wafer. These good chips are then placed in plastic carriers, with proper orientation such that all the chips on the carrier face in the same direction. At this point, each chip is examined under magnification to ensure that it has no visually obvious defects, such as scratches or cracks, or was not damaged by the probe contacts during the electrical sorting operation. The purpose of this visual inspection is to separate out and discard any mechanically damaged or defective chips prior to the assembly step. To avoid potential reliability problems, electrically good chips

EX-1014 0048
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

which may exhibit severe over- or underetching of various mask patterns, and particularly the aluminum interconnection lines, are also discarded. At this point, the remaining chips are ready to be mounted into final circuit packages.

Encapsulation

The packaging or encapsulation of the visually inspected IC chip proceeds in three steps. First, during the *die attach* step, the IC chip is attached or bonded to the gold-plated header, or the Kovar lead frame. Then, in the second step, called the *wire bonding* step, gold- or aluminum-wire leads are used to connect the bonding pads on the circuit chip to the package leads or posts. During this step, any one of several bonding techniques, such as thermocompression bonding (for gold wires) or ultrasonic bonding (for aluminum wires), can be used.

The last step of the encapsulating process is sealing the monolithic circuit package by injection molding (in the case of plastic packages), by soldering or welding a cap (in the case of side-brazed or metal can type packages), or by forming a hermetic glass seal between the package and its cap (in the case of ceramic packages). In most assembly operations, it is also customary to inspect each chip in the package after the wire-bonding step and prior to sealing. In this manner, obvious assembly defects can be detected quickly, and the rejects are discarded.

1.12. INTEGRATED-CIRCUIT PACKAGES

An IC package is expected to satisfy a large number of partly conflicting requirements: low cost, mechanical strength, high packing density, hermeticity, low parasitic reactances, low thermal resistance, and ease of handling and testing. No single circuit package exists which ideally fulfills these characteristics. For a majority of monolithic analog circuits, the package choice is narrowed down to the three most commonly used package types: *dual-in-line packages, metal cans,* and *flat packs*.

The dual-in-line (DIP) package, the dimensional diagrams of which are shown in Figure 1.20, is by far the most commonly used package type because of its relatively low cost and ease of handling. The in-line bent structure of the leads makes it convenient for automatic handling during electrical testing or board-insertion steps. The DIP packages are available in 8-, 14-, 16-, 18-, 20-, and 22-pin versions, with the narrow package dimensions shown in Figure 1.20*a*. Higher pin count versions of the DIP packages, from 24 to 40 pins, have the wider bodied structure shown in Figure 1.20*b*.

The DIP packages are available in three types: plastic packages which are nonhermetic and made of injection-molded epoxy compounds, black ceramic (CERDIP) packages, and the combined metal–ceramic or side-brazed packages. The plastic DIP package, which is rated for operation up to 85°C, is the lowest cost package type when hermeticity is not required. Both CERDIP and side-

EX-1014 0049
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

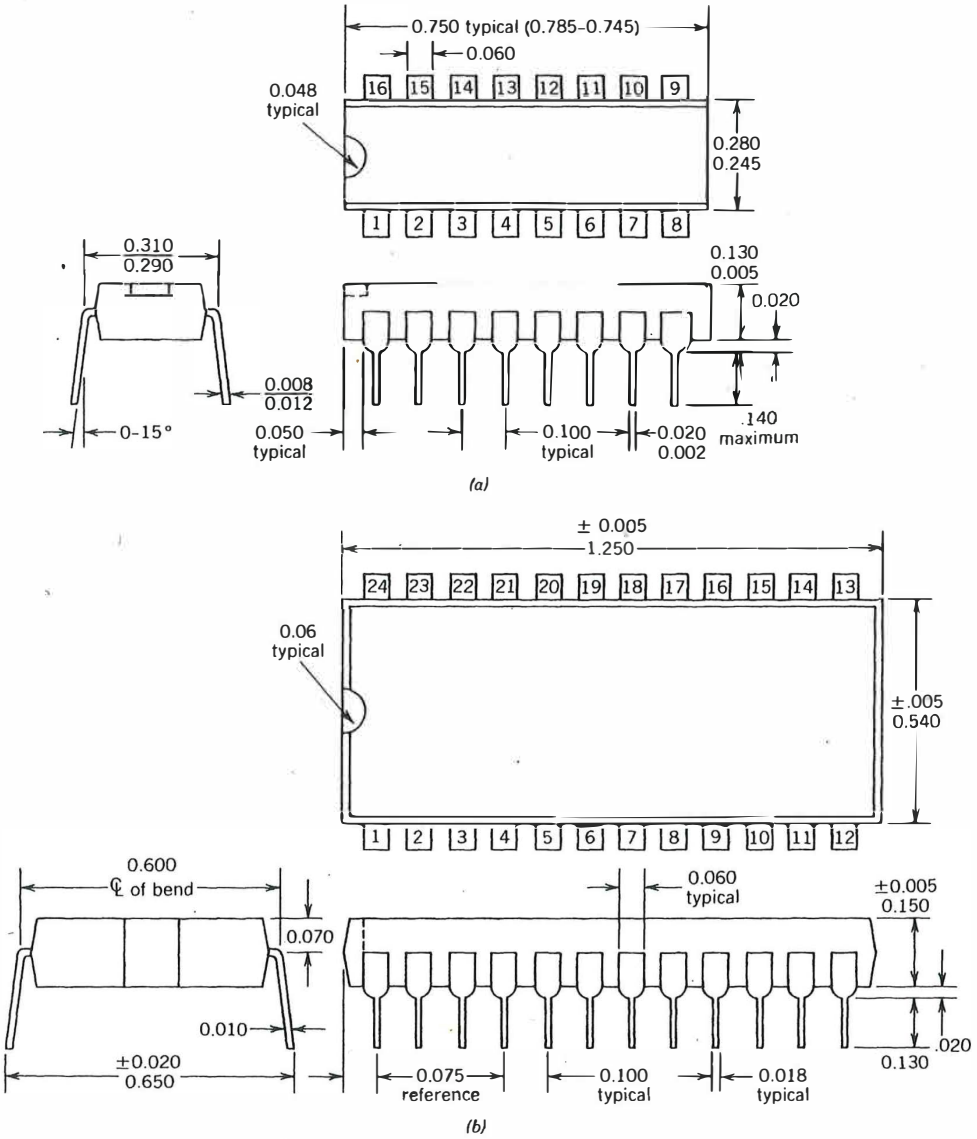


FIGURE 1.20. Dual-in-line (DIP) packages most commonly used in IC packaging (dimensions in inches): (*a*) Narrow-body DIP for 8–22 pins, (*b*) wide-body DIP for 24–40 pins.

brazed DIP packages are hermetic and can operate over a temperature range of −55 to +125°C.

The two most commonly used metal can type packages are the TO-99 and the TO-3, as shown in Figure 1.21. The metal can packages have a welded cap which hermetically seals the chip. Their key advantages are good thermal characteristics, high mechanical strength, and a very high reliability rating. Their disadvantages are (1) the available number of leads is limited (12 pins maximum), (2) leads bend easily and are difficult to insert into sockets, and (3)

EX-1014 0050
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

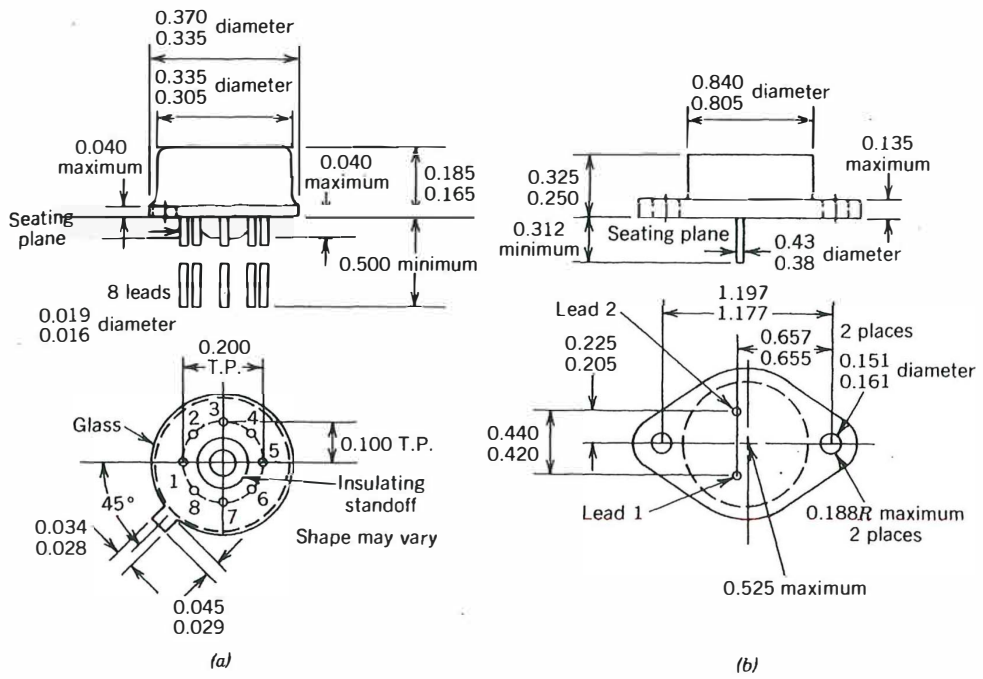


FIGURE 1.21. Metal can type IC packages (dimensions in inches): (*a*) TO-99 type package; (*b*) TO-3 type power package.

higher cost, particularly due to gold plating needed for certain parts of the package.

The flat pack type, small-geometry packages (see Fig. 1.22) were developed to improve on the volume, weight, and pin count limitations of metal can packages. Flat packs with up to 22 leads are commercially available. They have approximately one-fifth the volume and weight of conventional DIP packages and can be produced in both round and rectangular shapes. Their key advantages are lightweight and small volume; their major disadvantages are high cost and difficulty in handling.

Packaging and assembly are the major considerations in the low-cost high-volume manufacture of monolithic circuits. As such, an extensive amount of engineering and development effort has gone into these areas within recent years. The cost element, in particular, has brought about a multiplicity of new, fully automated packaging and assembly techniques, such as *beam-lead*, *flip-chip*, and *spider-bond* methods. A detailed analysis and comparison of these techniques is a highly specialized subject, which is well covered in the literature.[13,14]

Thermal Considerations

One of the basic limitations of integrated circuits is the dissipation of heat produced during operation of the circuit. This heat must be transfered to some

EX-1014 0051
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

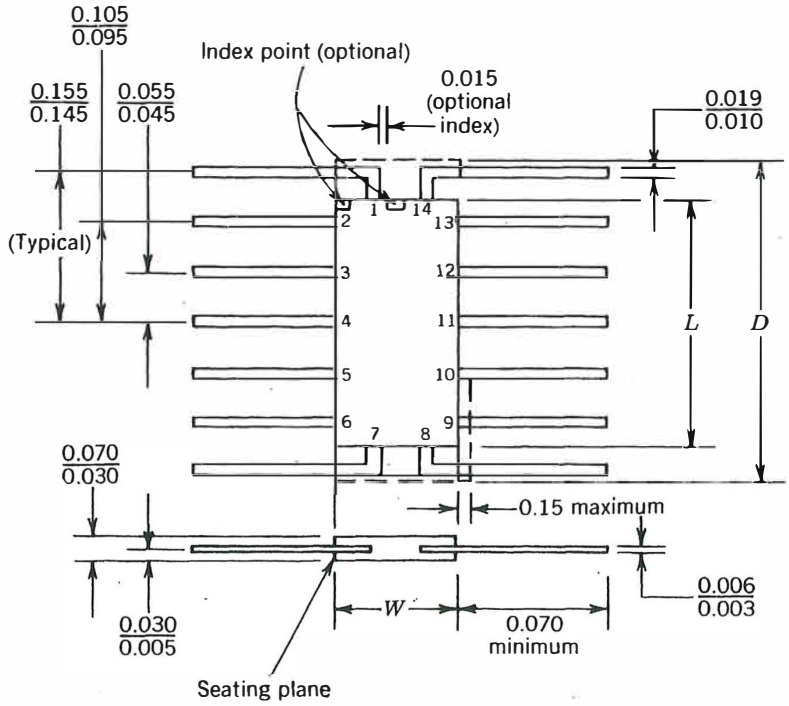


FIGURE 1.22. Physical dimensions of a 14-pin flat pack (dimensions in inches).

sink without causing excessive temperature rise in the circuit elements or interconnections. For silicon devices, reliability considerations dictate that the junction temperature on the chip T_j be kept below about 150°C. This basic limitation, along with the thermal conduction properties of the IC package, determines the maximum allowable power dissipation or ambient temperature range of operation.

The steady-state thermal behavior of the circuit chip and the IC package can be estimated by using its electrical analog shown in Figure 1.23.[15] In this model, the current is analogous to the flow of heat, and the voltage is equivalent to temperature. The current source P_d represents the power dissipation of the IC chip. The resistors R_{jc} and R_{ca} represent the thermal resistance from junction to

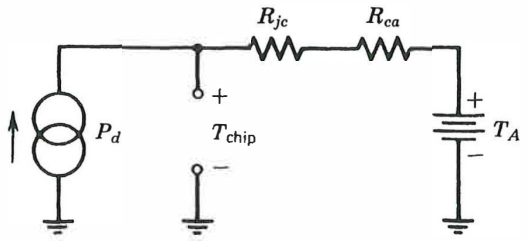


FIGURE 1.23. Electrical analog of package power dissipation.

EX-1014 0052
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

case and from case to ambient for the package. The sum R_T of the two resistances is the total thermal resistance of the system. From the approximate electrical model of Figure 1.23, the chip temperature can be written as

$$T_{\text{chip}} = T_A + (R_{\text{jc}} + R_{\text{ca}})P_d \qquad (1.16)$$

where T_{chip} and T_A are the temperatures of the chip and the ambient air surrounding the package.

The junction-to-case thermal resistance depends mainly on the particular package type used. However, it is also affected by the chip size and the manner in which the chip is attached to the package. The case-to-ambient thermal resistance is a function of the package surface, external heat sinking, and whether or not forced air cooling is used. Table 1.4 lists the approximate values for R_{jc} and R_{ca} for various package types, assuming no heat sinks and still air ambient.

Many manufacturers also like to state the package power handling capability in terms of the power dissipation capability at $T_A = 25°C$, with a derating factor expressed in mW/°C, for operation at temperatures above 25°C. These typical ratings and derating factors are also listed in Table 1.4 for a maximum allowable chip temperature of 150°C. Since different IC manufacturers may assume different maximum chip temperatures for their products, the 25°C power ratings and the derating factors may vary between different manufacturers.

TABLE 1.4. Approximate Thermal Resistance and Allowable Power Dissipation for Various IC Packages.[a]

Package Type	Thermal Resistance (°C/W)			Allowable Power Dissipation	Derating Factor for
	Junction to Case	Case to Ambient	Total	at $T_A = 25°C$ (W)	$T_A > 25°C$ (mW/°C)
Dual in line (DIP), 14/16-pin, plastic (Fig. 1.20*a*)	70	100	170	0.750	6
Dual in line (DIP), 14/16-pin, ceramic (CERDIP)	40	100	140	0.90	7
Metal can (Fig. 1.21)					
TO-99	50	100	150	0.8	6
TO-3	10	30	40	3.0	25
Flat pack (Fig. 1.22)	100	150	250	0.5	4

[a] $T_{\text{chip}} \leq 150°C$, no heat sink, still air ambient.

EX-1014 0053
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

1.13. TESTING OF INTEGRATED CIRCUITS

Although integrated circuits are batch processed during wafer fabrication, each and every circuit chip must be individually tested at least twice during the manufacturing cycle: once at the electrical sort step prior to scribing and die separation, and a second time after completion of the assembly operation. This second test cycle, called the final test step, includes a detailed testing of all circuit parameters according to the minimum and maximum specifications given in the product data sheet. Unless specified otherwise, both the electrical sort and the final test operations are done at room temperature.

A large number of tests are often necessary to characterize an analog integrated circuit. Consequently, testing is one of the most important, expensive, and time-consuming parts of the overall IC manufacturing process. The types of testing that can be performed on integrated circuits can be divided into three categories:

1. *DC Testing*. This measures the static parameters of the circuit, such as operating voltage and current levels, input bias currents, offsets, and so on. It is performed by forcing preprogrammed current or voltage levels to various circuit terminals and then sensing the resulting voltage or current levels.
2. *AC Testing*. This evaluates the circuit performance under operating bias, with sinusoidal ac signals applied to it.
3. *Dynamic Testing*. This includes testing the circuit in an environment or operating condition which simulates its actual application and includes pulse, amplitude, and time measurements as well as complex waveforms.

In normal production testing of analog integrated circuits, extensive dc testing with some ac testing is utilized. Dynamic testing is usually quite complex and time consuming. Therefore, it is used mainly for very specialized circuits which combine analog and digital functions on the same chip, and whose end applications are very well defined.

Because of the large number and the complexity of the tests required, computer-controlled and fully automated test systems have become a major element in IC testing. Figure 1.24 shows a generalized block diagram of such a computer-controlled test system.[16] The test program can be loaded into the computer by punched cards, paper tape, or magnetic tape. Instructions from the computer are then sent to an interface or control unit which controls the various elements of the system. Stimulus instructions for the integrated circuits are buffered, converted into analog voltages, and delivered to the pins of multiplexed test stations or wafer probes, which are time shared under computer control. Analog-to-digital (A/D) converters convert the output functions of the integrated circuits into digital form.

This information is then buffered and returned to the computer for processing. The computer makes a "go" or "no go" decision based on the test results and

EX-1014 0054
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

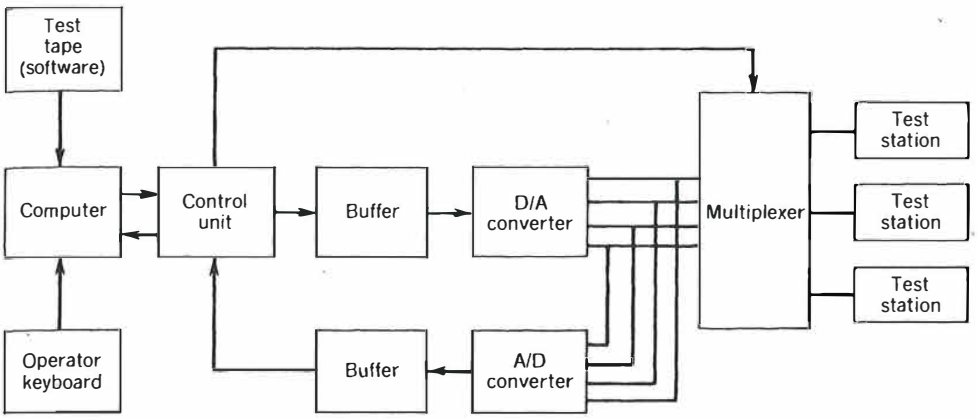


FIGURE 1.24. Functional block diagram of a computer-controlled test system.

automatically sorts the device into the appropriate container. Normally, such a tester would also have a data-logging capability such that preliminary modes of failure can be recorded for failure analysis and yield improvement purposes.

1.14. RELIABILITY CONSIDERATIONS

One of the most significant attributes of monolithic integrated circuits is high reliability. Integrated circuits are far more reliable than their discrete component counterparts, and their reliability is improving rapidly with increased knowledge of the processing techniques and an understanding of the possible failure mechanisms.

Failure Modes and Mechanisms

The most commonly encountered failure mechanisms in integrated circuits can be attributed to one of the following four sources: (1) bulk failures, (2) surface-related failures, (3) failures of metallization or interconnections, and (4) package-related failures.

Bulk Failure. Bulk failure is a relatively unimportant failure mode. Good starting material is essential in the fabrication of reliable integrated circuits. Crystallographic defects, such as dislocations, stacking faults, and growth strains, enhance long-term degradation mechanisms and, therefore, contribute to the unreliability of integrated circuits. Failure modes associated with bulk silicon include die breakage, short-circuits due to secondary breakdown, uncontrolled *pnpn* switching, and degradation of electrical characteristics. Bulk failure mechanisms are accelerated at high operating current densities due to localized heating effects.

EX-1014 0055
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Surface-Related Failures. These failures are statistically second only to failures in the interconnection system. Typically, 35% of all IC failures result from surface effects. Surface effects significantly influence *p-n* junction characteristics and tend to control transistor gains, junction breakdown voltages, and leakage currents. Charge migration along the silicon surface, especially in the vicinity of a *p-n* junction, is a major mechanism of surface instability. This instability is often caused by ionic contaminants in the oxide on the surface or near the silicon–oxide interface. The charge buildup due to ionic contamination may be high enough to cause the formation of an inversion layer along the surface, where the resistivity type of the underlying silicon may be reversed. Failures due to surface effects are accelerated by increasing the temperature and reverse biasing the *p-n* junctions. Both conditions tend to increase the ionic charge mobility and enable ionic contaminants to induce inversion layers near the junctions.

Metallization and Interconnection Failures. The most common failure modes in integrated circuits are open or short circuits in the circuit metallization and in bonding. These conditions contribute to more failures than all other failure types combined (typically between 50 and 60%).

Under high current densities ($\geq 5 \times 10^5$ A/cm^2), electromigration effects become a dominant failure mode. Electromigration is a mass transport effect which causes the atoms of the interconnection metal to migrate gradually toward the more positive end of the conductor. This mass transport phenomenon takes place along the grain boundaries of the metal interconnections and results in the formation of voids in the interconnection pattern which may eventually lead to an open circuit. Electromigration effects are enhanced at elevated temperatures. An additional failure mechanism associated with the metal interconnections is the formation of micro-cracks in the aluminum interconnections as a metal trace passes over a step on the SiO_2 layer.

Package-Related Failures. One of the serious reliability problems in integrated circuits is associated with the bonding of the wire leads between the package and the chip. A serious failure mechanism associated with gold-wire bonds on aluminum is known as purple plague, and is due to the formation of gold-rich intermetallic compounds such as Au_2Al, Au_4Al, and Au_5Al_2. These compounds create porous regions in the bond which are mechanically weak and electrically nonconductive. As these intermetallic compounds are formed, the differences in their structure and thermal expansion can stress the porous interface layer to the point of rupture. However, the "plague" formations are a serious problem only at elevated temperatures (typically $\approx$ 200°C) and are more likely to occur in step stressing than in actual use. Therefore, gold-wire–aluminum metallization and bonding systems are still considered to be the most reliable under normal operating conditions.

EX-1014 0056
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Testing for Package Reliability

Most packages use a number of different materials, such as metal, glass, ceramics, or plastics, to isolate the IC chip from its environment. Special testing procedures have been developed to test the actual sealing ability and the hermeticity of these packages. One of the tests commonly used for this purpose is the helium-leak test, where the package is immersed into a helium atmosphere under pressure for extended periods of time (usually 1 h). The package is then transferred to a mass spectrometer chamber and tested for helium leaks. Radioactive tracer methods can also be used to detect or trace leaks in the package seals.

Thermal stresses introduced between leak tests can point out losses in the package integrity and the cracking of seals. Thermal shock tests typically consist of cycling the package 10–20 times between the temperature extremes of -55 and $+125$°C. Other structural tests include lead fatigue tests, where the leads are bent back and forth for a given number of times; soldering tests, where the device must withstand typical soldering temperatures applied to the leads; and acceleration and shock tests, where the integrity of the package and the leads is examined under centrifugal or inertial shock conditions. In specialized applications, other parameters of the circuit and package may have to be measured. An example of such a parameter is the radiation resistance test, which tests the package integrity and circuit operation during and after irradiation by neutrons, X rays, and gamma rays.

Reliability Measurements

Providing a quantitative measure of reliability is a difficult task. In general, reliability is measured or compared to standards in terms of a *mean time between failures* (MTBF). The difficulty of demonstrating a given failure rate, or MTBF, becomes apparent from the requirement that the testing time be at least as long as the MTBF. In general, the MTBF for integrated circuits is greater than 10^7 h. To demonstrate this reliability with 90% confidence, approximately 2.3×10^7 h of operational life testing with no failures is required. To reduce this testing time, accelerated life tests may be used. This can be done by aging the integrated circuit at accelerated stress conditions, such as elevated temperatures. As shown in Figure 1.25, the reliability of the circuit decreases rapidly with increasing junction temperature.[17] This effect can be utilized to accelerate the life testing process.

Since a quantitative measure of reliability is very difficult to obtain, a number of relatively simple test or screening procedures have evolved, which provide at least a qualitative measure of reliability. These are briefly described below.

Burn-In. In this test, the units are operated at elevated temperatures (typically 125°C) under static bias or dynamic operating conditions for a relatively short

EX-1014 0057
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

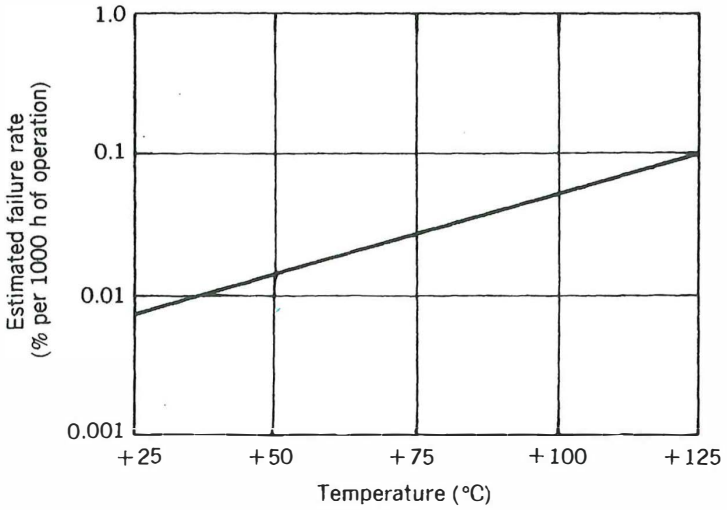


FIGURE 1.25. Estimated failure rate as a function of junction temperature.[17]

period (typically 168 h). It is primarily intended to detect and weed out the early failures.

Storage Life Tests. Storage life tests are the basic reliability tests for IC manufacturing. Circuits under test are stored at elevated ambient temperatures for 1000 h and up. Periodically, their characteristics are measured after cooling to room temperature. Usually, storage life test temperatures are between 125°C and 350°C.

Operating Life Tests. The operating life test is another basic reliability test, which is carried out at 25°C and/or 125°C under power, with the circuit operating under bias conditions similar to those encountered in its actual use. Normally, supply voltage and bias conditions are chosen to provide maximum stress conditions which may be encountered in practice. The key device parameters, such as offset currents and voltages, are then read and recorded at periodic intervals. These tests simulate actual use conditions more closely.

REFERENCES

1. J. A. Hoerni, U.S. Patent No. 3,025,589, Assigned to Fairchild Camera and Instrument Corp., New York, 1960.
2. J. C. Irvin, "Resistivity of Bulk Silicon and of Diffused Layers in Silicon," *Bell System Tech. J.* **41,** 387–410 (March 1962).
3. C. S. Fuller and J. A. Ditzenberger, "Diffusion of Donor and Acceptor Elements in Silicon," *J. Appl. Phys.* **27,** 544–553 (1956).
4. B. E. Deal, "The Oxidation of Silicon in Dry Oxygen, Wet Oxygen and Steam," *J. Electrochem Soc.*, No. 110, 527 (1963).

EX-1014 0058
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

5. A. S. Grove, *Physics and Technology of Semiconductor Devices*, Wiley, New York, 1967, Chap. 2.
6. R. M. Burger and R. P. Donavan, *Fundamentals of Silicon Integrated Device Technology*, Vol. 1, Prentice-Hall, Englewood Cliffs, N.J., 1967, p. 254.
7. J. Lyman, "Lithography Chases the Incredible Shrinking Line," *Electronics*, 105–116 (April 1979).
8. J. W. Mayer, L. Eriksson, and J. A. Davis, *Ion Implantation in Semiconductors*, Academic Press, New York, 1970.
9. G. Carter and W. A. Grant, *Ion Implantation of Semiconductors*, Wiley, New York, 1976.
10. H. A. Waggener, R. C. Kragness, and A. L. Tyler, "Anisotropic Etching for Forming Isolation Slots in Silicon Beam-Leaded Integrated Circuits," IEEE Int. Electron Dev. Conf., Washington, DC, 1967.
11. B. Polata, "Compatible High-Performance and Complementary Bipolar Transistors for Integrated Circuits," IEEE Int. Electron Dev. Conf., Washington, DC, 1969.
12. P. C. Davis, S. F. Moyer, and V. R. Saari, "High Slew Rate Monolithic Operational Amplifier Using Compatible Complementary pnp's," *IEEE J. Solid-State Circuits*, **SC-9,** 340–346 (December 1974).
13. A. B. Glaser and G. E. Subak-Sharpe, *Integrated Circuit Engineering*, Addison-Wesley, Reading, MA, 1977, Chap. 10.
14. D. J. Rose, "Packaging and Assembly: The 1980's Semiconductor Technology Forecast," Semiconductor International, 41–50 (January 1980).
15. P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, Wiley, New York, 1977, Chap. 2, pp. 122–126.
16. F. VanVeen, "An Introduction to IC Testing," *IEEE Spectrum*, 28–37 (December 1971).
17. Research Triangle Institute, "Integrated Silicon Device Technology—Reliability," Vol. 15, Rep. ASD-TDR 63-316, May 1967.

EX-1014 0059
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

CHAPTER TWO

ACTIVE DEVICES IN INTEGRATED CIRCUITS

In designing monolithic circuits, designers do not have the luxury of choosing their devices at will. Instead, they are constrained to use those device types that can be fabricated easily and economically with standard manufacturing processes. The limitations imposed by the fabrication techniques, and the strong interdependence of device parameters and parasitics, often require significant design and layout compromises. Therefore, it is necessary for an analog IC designer to be familiar with the characteristics and limitations of monolithic devices. The purpose of this chapter is to provide a comprehensive overview of the active components available in monolithic circuits. It will be assumed that the reader is already familiar with the fundamentals of semiconductor device theory. In surveying the integrated device and component structures, particular attention will be given to relating the device characteristics and the associated parasitic effects to the physical properties of integrated structures.

An important distinction between IC design and the conventional circuit design with discrete components is that IC designers also have the capability to determine, or specify, the geometry and the layout of the devices they use. This gives IC designers an added degree of freedom in optimizing their circuit performance. Thus, IC design involves a certain amount of device design as well. For example, the need often exists for a transistor with a high current-carrying capability to be used in the output stage of an amplifier. Such a device can be made by using a device geometry other than the standard one, and effectively consists of many standard devices connected in parallel. However, the larger device in turn will exhibit a somewhat different set of parasitics than the small-signal devices due to added stray capacitance effects. Thus it is the responsibility of IC designers to be familiar with the characteristics, the parasitics, and the design trade-offs associated with the active devices at their disposal so that they can make optimum use of the tools available to them.

EX-1014 0060
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

2.1. *npn* TRANSISTORS

The *npn* bipolar transistor is by far the most significant active component in analog integrated circuits. The basic fabrication steps described in Chapter 1 were initially developed around the *npn* bipolar transistor, and were later extended to other active devices. Consequently, in going from a discrete to an integrated device, the basic *npn* bipolar transistor structure involves the least amount of design compromise. In most analog IC designs, the characteristics of the available *npn* transistors serve as the starting point for the rest of the design; and the remainder of the circuit components are then chosen or designed to be compatible with the fabrication steps required for the *npn* transistor.

Device Structure

The monolithic *npn* transistor differs from its discrete counterpart in one important aspect, as illustrated in Figure 2.1. In the case of the discrete transistor, one can make a direct electrical contact to the collector portion of the device through the backside of the chip. In the case of the integrated *npn* transistor, the device is surrounded by a reverse-biased *p-n* junction (i.e., the isolation wall and

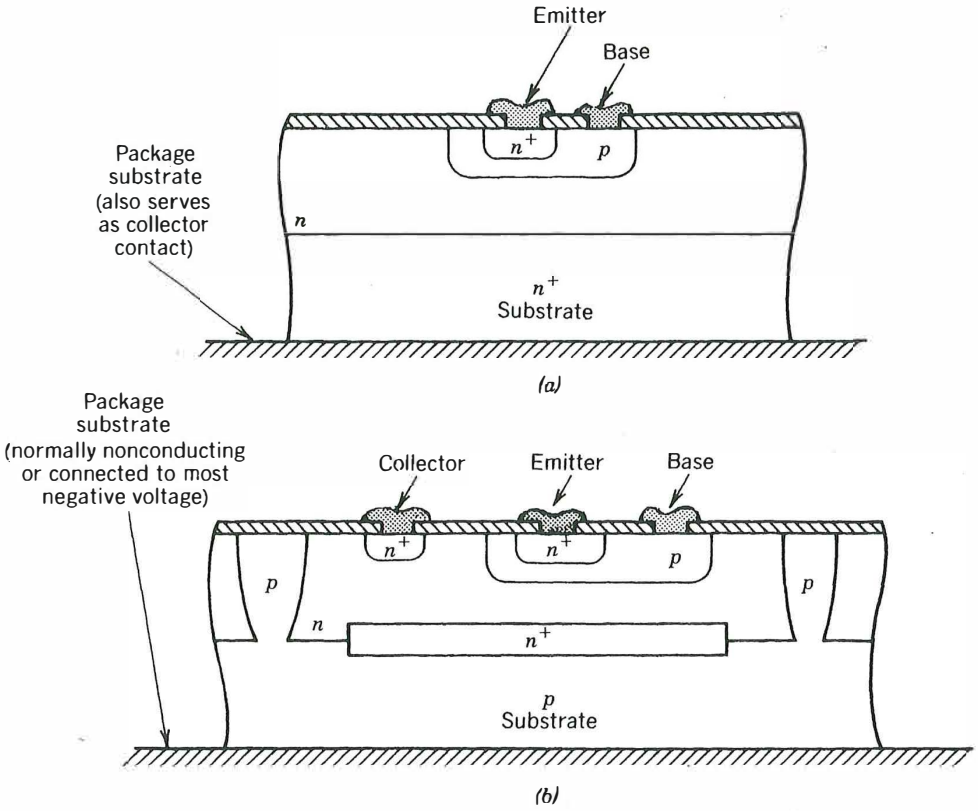


FIGURE 2.1. Comparison of discrete and integrated *npn* bipolar transistor structures: (*a*) Discrete transistor; (*b*) integrated transistor.

EX-1014 0061
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

the p-type substrate), which electrically isolates it from the other devices on the same chip. As a result, the collector portion of the device is accessible only through an electrical contact at the top surface of the device. This introduces an additional parasitic series resistance r_{cs}, into the collector terminal of the integrated *npn* transistor.

The presence of a reverse-biased junction isolation pocket surrounding the collector region of the integrated device introduces a potentially parasitic *pnp* transistor within the same device structure, as illustrated in Figure 2.2*a*. This *pnp* transistor is formed by the p-type base of the *npn* transistor along with the n-type collector and the p-type isolation regions. In normal operation, the substrate is *always* biased at a more negative voltage than the n-type collector of the *npn* transistor which in turn causes the emitter–base junction of the parasitic *pnp* transistor to be permanently reversed biased, and thus maintains it in the off or nonconducting state. Under this condition, the isolation pocket can be considered as a reverse-biased diode D_{CS}, with its associated parasitic capacitance C_{CS} as shown in Figure 2.2*b*.

The subepitaxial n^+-type layer in the integrated *npn* transistor (Fig. 2.1*b*) serves a dual purpose: it provides a low-resistivity current path from the active collector region to the physical collector contact, and it minimizes the possibility of parasitic *pnp* action by reducing the current gain of the parasitic *pnp* transistor of Figure 2.2*a*.

Figure 2.3 shows the two possible conditions which may cause the parasitic *pnp* transistor to be active. One such case, shown in Figure 2.3*a*, is when the collector of the *npn* transistor is pulled to a potential below the substrate voltage. This condition may occur if several IC chips operating with different supply voltages and different substrate bias levels have to be interfaced. The second case illustrated in Figure 2.3*b*, comes about when the *npn* transistor is driven into saturation (i.e., when its collector–base junction becomes forward biased). This condition may also come about when the *npn* transistor is operated in an inverted

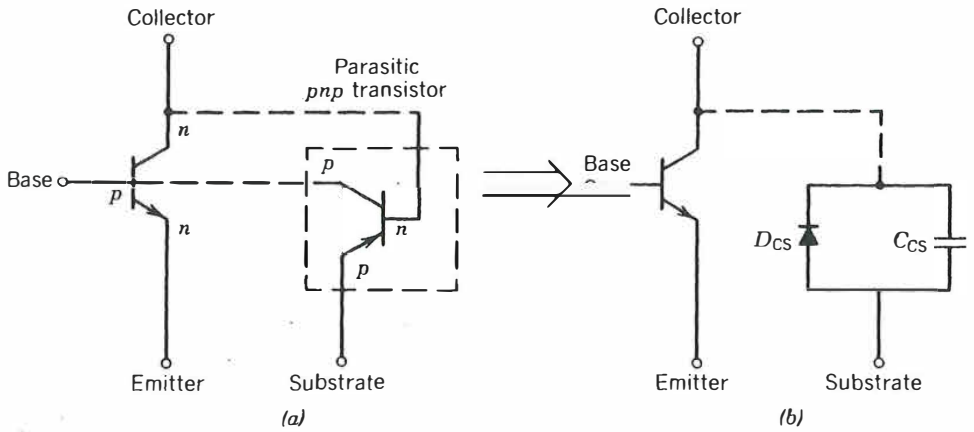


FIGURE 2.2. Parasitic devices associated with a junction-isolated *npn* transistor: (*a*) Parasitic *pnp* transistor to substrate; (*b*) its equivalent circuit under proper reverse-bias condition.

EX-1014 0062
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

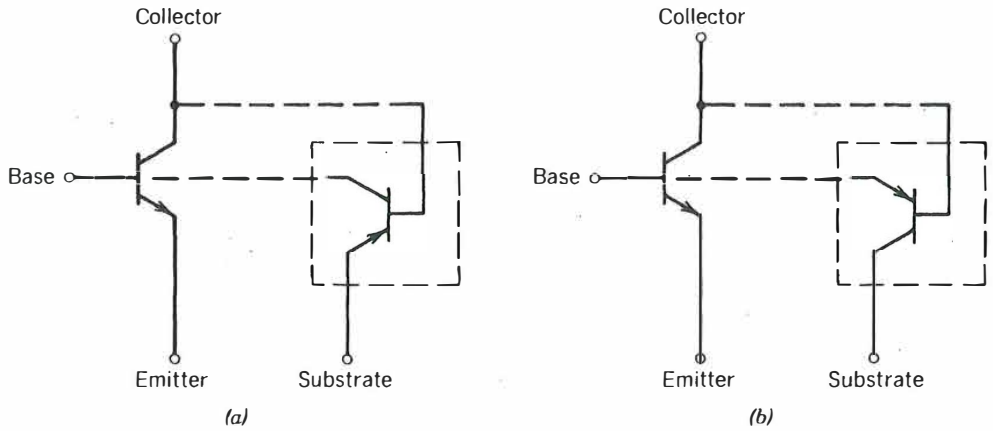


FIGURE 2.3. Two possible conditions where parasitic *pnp* action may result: (*a*) When the collector of the *npn* transistor is pulled to a voltage below substrate potential; (*b*) when the *npn* transistor is saturated so that its collector–base junction is forward biased, or when it is operated in the "inverted" mode with its emitter and collector reversed.

mode, with the roles of its emitter and collector reversed. The parasitic *pnp* action of Figure 2.3*b* is harmful only if the emitter of the *npn* transistor is *not* connected to the same potential as the substrate, since under this condition the parasitic *pnp* transistor would shunt some of the collector current of the *npn* transistor into the substrate. In normal design practices, both of these possible parasitic *pnp* conditions are avoided by proper biasing or design analysis, and the substrate junction remains as a reverse-biased diode, as shown in Figure2.2*b*.

Figure 2.4 shows the lateral geometry of a typical small-signal *npn* transistor. To give an idea of the lateral dimensions and tolerances, the scale of the drawing, in micrometers is shown alongside the figure. The composite mask layers which form the device are superimposed on the figure, with appropriate coding to identify their functions. The structural cross-section of the same device, sectioned normal to the wafer surface, is also shown. For illustrative purposes, a medium-voltage fabrication process with 15-μm epitaxial layer thickness and 30-V breakdown voltage is assumed (see Fig. 1.13).

The minimum size for the lateral dimensions of the device is limited by two significant factors: (1) masking and mask alignment tolerances and (2) side-diffusion effects. Normally, a clearance of approximately 5 μm is left around an oxide contact cut and the edge of the corresponding diffusion, as is the case for the base and emitter contacts. This tolerance is left to account for any possible misalignment of the mask patterns during the masking operation, or the overetching of the oxide window during the subsequent etching step.

The transistor action takes place directly below the emitter region. Therefore, to be able to supply the collector current with a minimum amount of series voltage drop, it is preferable to locate the collector contact as close to the emitter as possible. The distance between the edge of the base region and the collector contact is chosen to be significantly more than the respective side diffusions of the p-type base and the n^+-type collector contact areas. If this precaution is not

EX-1014 0063
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

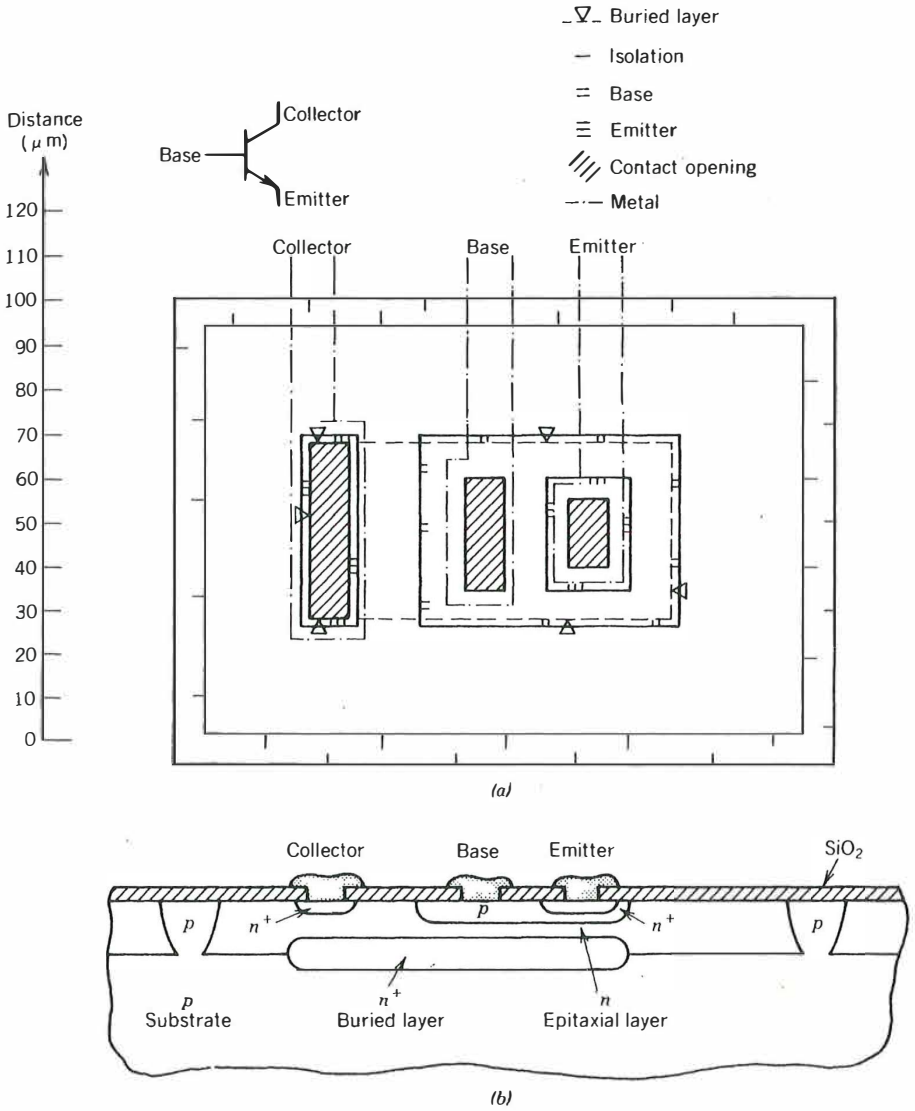


FIGURE 2.4. Lateral dimensions and cross section of a typical small-geometry *npn* transistor. (Mask layers coded for identification as shown.)

taken, the n^+-type collector contact region may touch the p-type base and result in a low collector–base breakdown voltage. In the case of the typical device geometry shown in Figure 2.4, this dimension is approximately 10 μm. The subepitaxial n^+-type layer is located directly below the base region and extends to the area directly below the collector contact.

The distance between the p-type isolation wall and the inner transistor structure is set by the side-diffusion effects, as well as by the thickness of the *depletion layer* associated with the collector–base and the collector–isolation junctions. Since the isolation diffusion is a deep one, it also tends to side-diffuse significantly more than the base. Therefore, it is customary to leave a typical

EX-1014 0064
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

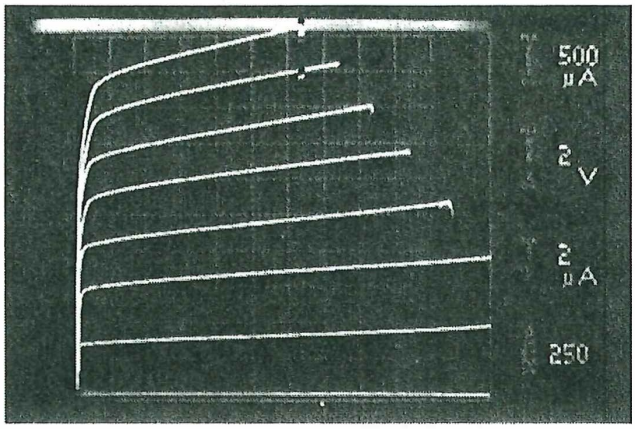


FIGURE 2.5. Typical curve-tracer picture of current–voltage characteristics for the small-geometry *npn* transistor. (*Photo:* Exar Integrated Systems, Inc.)

clearance of $1.5–1.8X_e$ between the inner edge of the isolation wall and the edge of the base diffusion, where X_e is the thickness of the epitaxial layer. In the case of the 15 μm thickness of the epitaxial layer shown in Figure 2.4, this corresponds to a spacing of approximately 25 μm. The base-to-isolation spacing is the dominant factor in determining the minimum size of a small-signal *npn* transistor. Since this spacing is directly proportional to the epitaxial layer thickness, high-breakdown transistors which require thicker epitaxial layers do not provide as good a component packing density on the chip as those devices fabricated with a low-voltage process and a thinner epitaxial layer.

Figure 2.5 shows the typical current–voltage characteristics for the small-geometry *npn* transistor of Figure 2.4.

Electrical Characteristics

When biased in its active region, the bipolar transistor functions as a nonideal current-controlled current amplifier. A given amount of base current I_B injected into the base terminal causes a much larger collector current I_C to flow. Figure 2.6 shows a simplified equivalent circuit for an *npn* bipolar transistor which

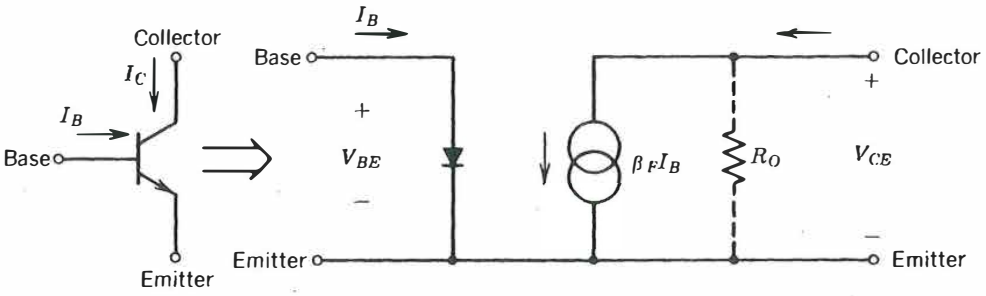


FIGURE 2.6. Simplified equivalent circuit of *npn* transistor for bias calculations. [*Note* $R_O = (V_A + V_{CE})/I_C \approx V_A/I_C$ signifies the finite output resistance due to the Early effect.]

EX-1014 0065
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

approximates its current–voltage characteristics. When the transistor is in its active region, the base–emitter junction is forward biased and the base–collector junction is under reverse bias.

The collector current I_C is related to the base current I_B as

$$I_C = \beta_F I_B \tag{2.1}$$

where β_F is the forward-current gain. Similarly, the base current I_B is an exponential function of the applied base–emitter voltage V_{BE},

$$I_B = I_{BO} \exp\left(\frac{q V_{BE}}{k T}\right) \tag{2.2}$$

where I_{BO} = reverse saturation current of the base–emitter diode
q = electronic charge
k = Boltzmann's constant
T = temperature (in °K)

The term kT/q, which has the dimensions of voltage and is proportional to temperature, is often referred to as the thermal voltage V_T,

$$V_T = \frac{kT}{q} \approx 26 \text{ mV at } 25°\text{C} \tag{2.3}$$

The collector current I_C is related to the base current through the forward gain factor β_F as

$$I_C = \beta_F I_B = I_{CO} \exp\left(\frac{V_{BE}}{V_T}\right) \tag{2.4}$$

Similarly, the base–emitter voltage to sustain a given collector current can be expressed as

$$V_{BE} = V_T \ln\left(\frac{I_C}{I_{CO}}\right) \tag{2.5}$$

The parameter I_{CO}, which is called the collector reverse saturation current, is given as

$$I_{CO} = \frac{q D_n n_i^2}{Q_B} A = I_S A \tag{2.6}$$

where D_n is the diffusion constant of electrons in the base, n_i is the intrinsic carrier concentration in silicon, Q_B is the total number of dopant atoms in the base region per unit area of the emitter, and A is the area of the base–emitter junction. The main significance of Eq. (2.6) is that the collector reverse saturation current is *directly proportional* to the emitter area. The first term in Eq. (2.6) is a parameter which depends on the particulars of the device impurity profile and the semiconductor material properties. For the IC devices fabricated simul-

EX-1014 0066
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

taneously on the same chip, this term will be the same. Thus, as implied by Eq. (2.6), the emitter area A of the *npn* bipolar transistor can be used as a scaling factor such that if the same base–emitter voltage is applied to two transistors on the same chip, their collector currents I_{C1} and I_{C2} will be related to their respective emitter areas, A_1 and A_2 as

$$\frac{I_{C1}}{I_{C2}} = \frac{A_1}{A_2} \tag{2.7}$$

The scaling of transistor currents by scaling their emitter areas is one of the most often used circuit design techniques in analog IC design and will be discussed in more depth in Chapter 4.

The common-base current gain factor α_F is the ratio of the collector and emitter currents, and is given as

$$\alpha_F = \frac{I_C}{I_E} = \frac{I_C}{I_B + I_C} = \frac{\beta_F}{1 + \beta_F} \tag{2.8}$$

In terms of the physical current conduction mechanism within the transistor, α_F indicates the fraction of the carriers injected from the emitter which reach the collector, and is a number very close to unity. It can be expressed as a product of a number of device parameters,

$$\alpha_F = \gamma\beta^* M \tag{2.9}$$

where γ = emitter efficiency
β^* = base transport factor
M = collector avalanche multiplication factor.

The emitter efficiency γ is defined as the ratio of the electron current (for the case of an *npn* transistor), injected into the base from the emitter, to the total hole and electron current crossing the emitter–base junction. It can be closely approximated by an expression of the form[1]

$$\gamma = \left(1 + \frac{\rho_E}{\rho_B}\frac{W_B}{L_E}\right)^{-1} \tag{2.10}$$

where L_E is the diffusion length of the minority carriers into the emitter, ρ_E and ρ_B are the average resistivities of the emitter and base regions within a diffusion length of the junction, and W_B is the width of the base region. As implied by Eq. (2.10), γ approaches unity as the emitter is more heavily doped with respect to the base, and as the base width W_B is narrowed. For an integrated device structure similar to that shown in Figure 2.4, with the impurity profile of Figure 1.13, the emitter efficiency is in the range of 0.992–0.998 for I_E in the low milliampere range.

The base transport factor β^* is the fraction of minority carriers injected from the emitter that reach the collector, and it can be approximated as

$$\beta^* \approx 1 - \frac{W_B^2}{2D\tau} \tag{2.11}$$

EX-1014 0067
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

where D is the minority carrier diffusion coefficient, and τ is the excess minority carrier lifetime within the base region. For a typical integrated *npn* transistor, β^* is of the order of 0.995.

The collector avalanche multiplication factor M is the ratio of the carriers entering the collector region to the number of minority carriers arriving at the base side of the collector–base junction depletion layer. Because of the secondary ionization effects within the collector–base depletion region, M is usually slightly larger than unity. It can be related to the base–collector reverse bias V_{CB} by the empirical relationship

$$M = \left[1 - \left(\frac{V_{CB}}{BV_{CB}}\right)^m\right]^{-1} \tag{2.12}$$

where BV_{CB} is the breakdown voltage of the collector–base junction. The exponent m has the approximate values of 4 and 2, respectively, for the *npn* and *pnp* transistors.

Figure 2.5 shows the typical current–voltage characteristics of a monolithic *npn* transistor. The finite slope of the collector current–voltage characteristics is due to the modulation of the effective base width by the widening of the collector–base depletion layer, as the collector–base voltage is increased. This base width modulation, known as the *Early effect*, becomes more significant as the base width W_B is reduced, or as the resistivity of the base region is increased with respect to that of the collector region. For dc modeling of the transistor, the finite slope of the output characteristics can be extrapolated to a common point on the collector–emitter voltage V_{CE} axis as shown in Figure 2.7, and this extrapolated point can be used to define an effective Early voltage V_A. This allows the base width modulation effects to be incorporated into Eq. (2.4) by

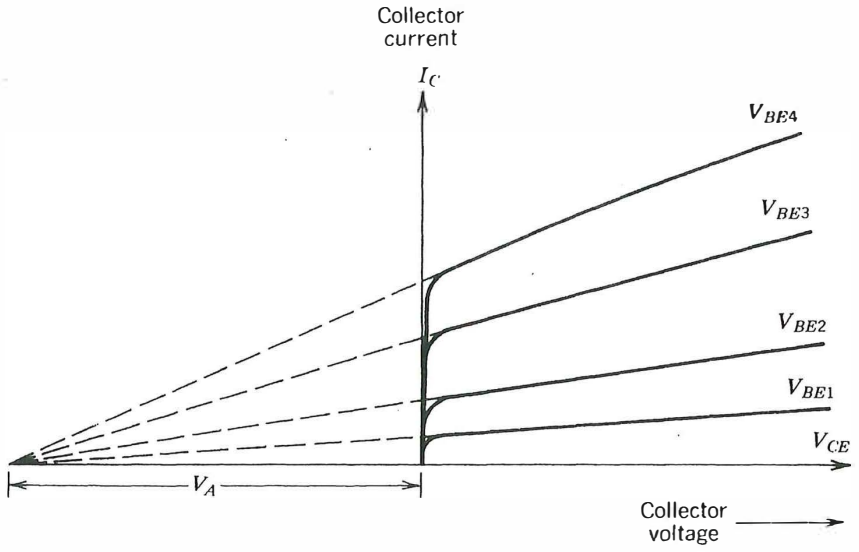


FIGURE 2.7. Bipolar transistor output characteristics showing the Early effect voltage V_A.

EX-1014 0068
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

rewriting it as

$$I_C = I_{CO}\left(1 + \frac{V_{CE}}{V_A}\right)\exp\left(\frac{V_{BE}}{V_T}\right) \tag{2.13}$$

For typical *npn* transistors, the Early voltage V_A is in the range of 50–100 V and decreases as the forward-current gain β_F is increased due to the reduced base width.

The simplified equivalent circuit of Figure 2.6 is often sufficient for the first-order analysis of the dc bias conditions for an *npn* transistor stage. Depending on the required accuracy of the calculations, the effect of base width modulation can be incorporated into the model by the addition of the resistor R_O, shown by dashed lines in Figure 2.6.

The forward-current gain β_F varies with both the temperature and the collector current. It exhibits a strong positive temperature coefficient of approximately +5000 to +7000 ppm/°C (where ppm stands for parts per million). This temperature dependence of β_F is primarily due to the increase in emitter efficiency γ [see Eq. (2.10)] with increasing temperature.[2]

Figure 2.8 gives some typical β_F versus I_C curves for the small-signal *npn* transistor of Figure 2.4 at three different temperatures. As shown, the dependence of β_F on I_C can be divided into three regions. In the low-current region, the parasitic surface recombination and the recombination of carriers in the base–emitter depletion region are primarily responsible for the β_F falloff. The low-current β_F can be improved by minimizing surface recombination effects through additional surface passivation steps, such as silicon nitride deposition, or by reducing the emitter periphery and the emitter–base junction area. In the

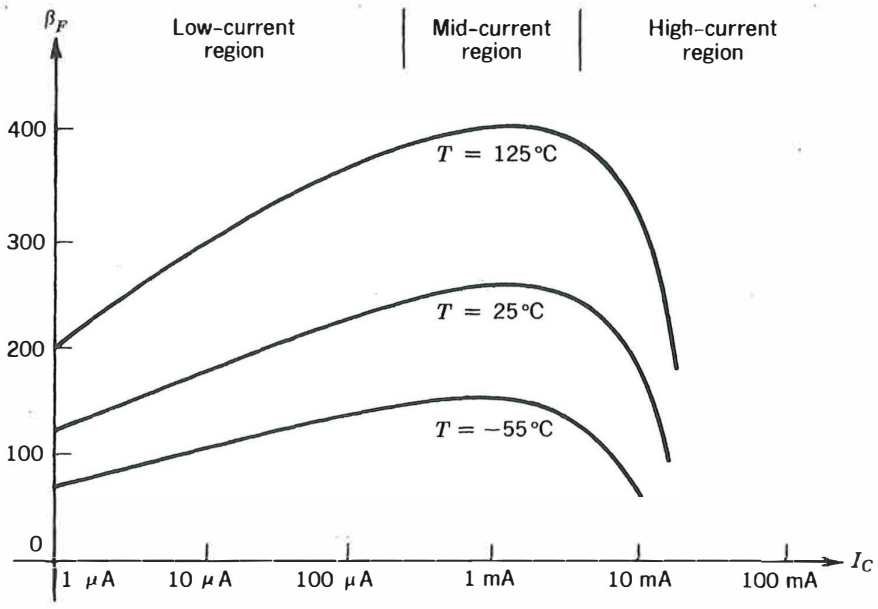


FIGURE 2.8. Typical β_F versus I_C characteristics for a small-geometry *npn* transistor.

EX-1014 0069
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

mid-current range, which stretches from approximately 100 μA to 3 mA, β_F is relatively constant.

The decrease of β_F at high currents is due to two dominant factors: decrease of emitter efficiency and emitter-crowding effects. The decrease of emitter efficiency at high currents results from the presence of a large number of excess minority carriers in the base, reducing the effective base resistivity ρ_B near the base–emitter junction. The emitter-crowding effect is caused by the ohmic drop within the active base region due to the flow of base current. As a consequence of this, a voltage gradient is created within the active base region and the edge of the emitter becomes more forward biased than the bottom of the emitter area. Thus, the emitter region injects carriers preferentially along its periphery, and only the edge of the emitter is electrically active. To reduce the β_F falloff at high current levels, it is necessary to maximize the emitter periphery-to-area ratio, and to minimize the base-spreading resistance. This in turn leads to an interdigitized transistor structure for high-current applications (see Fig. 2.16). Another factor which reduces β_F at high current levels is the onset of the so-called *Kirk effect*[3] which occurs when the minority carrier concentration in the collector becomes comparable to the donor atom density. This causes the effective base with W_B of the transistor to appear to be larger by stretching into the collector region, and this in turn reduces β_F by decreasing the base transport factor β^* of Eq. (2.11).

Voltage Breakdown

As the reverse bias across a *p-n* junction is increased beyond a critical value, the current through the junction increases rapidly. This critical voltage is known as the junction breakdown voltage BV. In silicon, two separate breakdown mechanisms exist. These are the avalanche and the Zener breakdowns.

If the impurity concentration on either side of the junction is less than $\approx 10^{18}$ atoms/cm^3, the breakdown voltage is determined by the onset of avalanche multiplication. It occurs when the electric field within the depletion layer provides sufficient energy for the free carriers to knock off additional valence electrons from the lattice atoms. These secondary electrons, in turn, generate additional free carriers, leading to an avalanche multiplication of the free carriers within the depletion layer. This phenomenon is similar to the ionization breakdown in gases. The avalanche breakdown voltage is normally determined by the impurity concentration on the lighter doped side of the junction. For example, in the case of a monolithic *npn* transistor, the base impurity concentration would have the dominant influence on the emitter–base breakdown voltage; and the collector doping will determine the collector–base breakdown voltage. Figure 2.9 gives the avalanche breakdown voltage versus concentration for a *p-n* junction where the lighter doped side is assumed to have a uniform impurity distribution. This provides a good approximation to the collector–base junction of an *npn* transistor.[1] The avalanche breakdown voltage shows a strong temperature dependence with a typical temperature coefficient on the order of +300 ppm/°C, the dependence being stronger for higher breakdown values.

EX-1014 0070
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

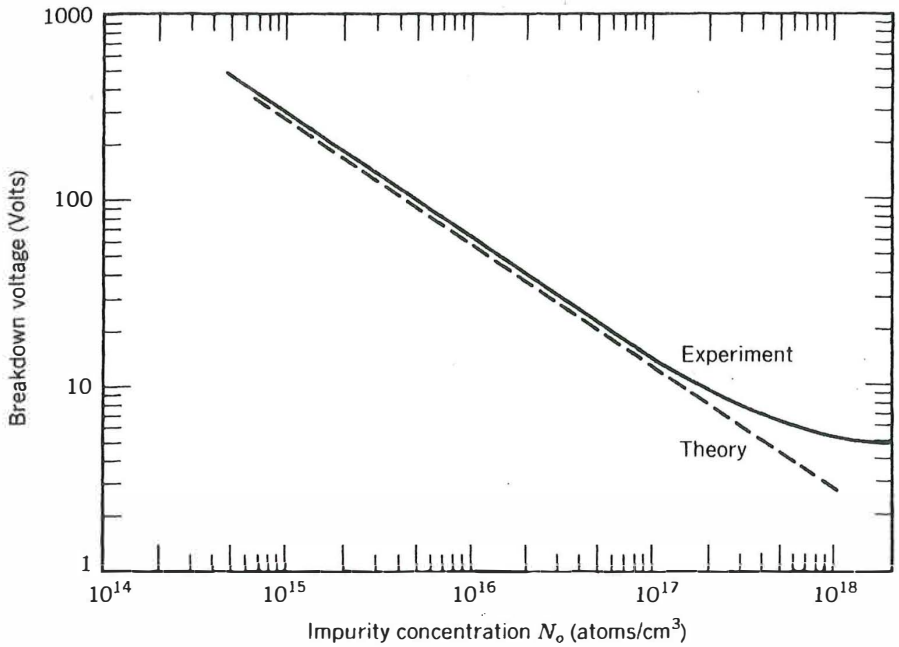


FIGURE 2.9. Avalanche breakdown voltage versus concentration for a step junction.[1]

The basic breakdown voltages associated with the integrated transistor structures are the base–collector BV_{CBO} and the base–emitter BV_{EBO} breakdowns. In designating the transistor breakdown voltage, a triple subscript designation is used where the last subscript O designates that the remaining terminal is open-circuited. Both the BV_{EBO} and the BV_{CBO} values can be closely approximated from the data of Figure 2.9 once the base and the collector impurity levels are known. The typical values of BV_{EBO} are in the range of 6–8 V for integrated transistors, which makes the base–emitter breakdown diode a convenient device for dc voltage reference and level shifting within the circuit.

Due to carrier multiplication effects within the base region of the transistor, the collector–emitter breakdown voltage BV_{CEO}, is smaller than BV_{CBO}. It can be approximated as[4]

$$BV_{CEO} \approx \frac{BV_{CBO}}{(\beta_F)^{\frac{1}{m}}} \tag{2.14}$$

where m is the exponent associated with the avalanche multiplication factor described in Eq. (2.12). Since β_F can vary significantly over the operating current range of a transistor, the lowest value of BV_{CEO} within the current–voltage characteristics of the transistor is called the *lowest sustaining voltage* LV_{CEO}. In normal operation of the transistor, its maximum operating voltage is determined by its lowest sustaining voltage rating. Thus, for example, when one talks about a 25-V or a 40-V transistor process, one is normally referring to the LV_{CEO} value, which can be reliably and repeatedly obtained from that process.

The second type of breakdown mechanism in semiconductors is the so-called

EX-1014 0071
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Zener breakdown, which happens if both sides of the junction are very heavily doped. Zener breakdown is a quantum-mechanical phenomenon where the electrons from the valence band on one side of the junction can "tunnel" to an available energy state in the conduction band on the other side. For junctions which breakdown at 5 V or less, the Zener breakdown is the main conduction mechanism. In analog IC structures, impurity concentration levels that can give way to true Zener breakdown are not normally encountered. However, in the IC jargon, the base–emitter reverse breakdown voltage BV_{EBO}, which normally lies in the range of 6–8 V is mistakenly referred to as the Zener diode. As will be described in Chapter 4, this diode is often used in analog integrated circuits to generate reference voltages within the circuit.

If the current through the junction is limited to a safe value, the breakdown is not destructive and the device junctions recover to normal when the voltage level is reduced. However, one exception to this is the case of the base–emitter junction where the repeated avalanche breakdown of the junction will cause β_F to be degraded.[5] If the device is being used as a voltage reference, this is of no consequence; however, if the transistor is being used as a gain stage, particularly at the input of an amplifier, the β_F degradation may seriously affect the performance.

An additional voltage breakdown mechanism in a transistor is the *punch-through* breakdown. In a narrow and lightly doped base structure, the collector depletion layer can extend through the entire base region into the emitter, thus forming a current path between the collector and the emitter. Punch-through is the dominant breakdown mechanism in the superbeta transistors, which will be covered later in this section.

Matching of Device Characteristics

The absolute-value tolerances in integrated devices are, in general, poorer than their discrete counterparts. However, the monolithic components are far superior to discrete devices in the matching and tracking of device parameters. Table 2.1 shows the typical range of values of β_F, V_{BE}, and the base-emitter reverse breakdown V_{EBO} for integrated *npn* transistors. Typical values of the absolute-value and matching tolerances, as well as the temperature coefficients, are also listed in the table. In all cases, the tolerances and the temperature coefficients

TABLE 2.1. Typical Parameter Values and Tolerances for Monolithic *npn* Transistors

Device Parameter	Typical Range of Values	Absolute Value Tolerance	Matching Tolerance	Temperature Coefficient	Thermal Tracking
β_F	50–300	±20%	± 5%	+5000–+7000 ppm/°C	±500 ppm/°C
V_{BE}	0.6–0.7 V	±20 mV	± 1 mV	−2 mV/°C	±10 μV/°C
V_{EBO}	6–9 V	±200 mV	±25 mV	+2–+6 mV/°C	±200 μV/°C

EX-1014 0072
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

shown refer to a parameter design value falling halfway between the range limits, that is, $\beta_F \approx 150$, $V_{BE} \approx 0.65$ V, and $V_{EBO} \approx 7$ V.

Small-Signal Model

For small-signal applications, the frequency response of the bipolar transistor can be closely approximated by the *hybrid-π* model shown in Figure 2.10.

In the equivalent circuit, the parameters r_b, C_c, r_o and r_{cs} are the parasitics inherent to the basic bipolar transistor structure. The base-spreading resistance r_b is the resistance of the current path which the base current must traverse from the physical base contact to the active base region below the emitter. C_c is the collector–base junction capacitance which depends on the area A_B of the collector–base junction as well as on the reverse bias V_{CB} across it. Assuming a uniformly doped collector region, it can be approximated by an expression of the form

$$C_c \approx A_B \sqrt{\frac{q\epsilon N_C}{V_{CB}}} \tag{2.15}$$

where ϵ = dielectric constant of silicon
N_C = collector doping concentration
A_B = collector–base junction area

The device transconducture g_m can be expressed as

$$g_m = \frac{\partial I_C}{\partial V_{BE}} \approx \frac{\partial I_E}{\partial V_{BE}} = \frac{qI_E}{kT} = \frac{I_E}{V_T} \tag{2.16}$$

where V_T is the thermal voltage defined in Eq. (2.3)

The small-signal forward gain β_o is defined as the *incremental* change of the collector current with the change of base current, that is,

$$\beta_o = \frac{\Delta I_C}{\Delta I_B} \tag{2.17}$$

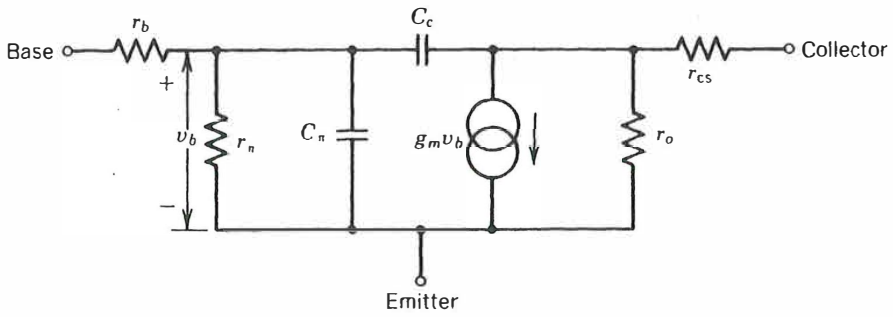


FIGURE 2.10. Small-signal hybrid-π equivalent circuit for a bipolar transistor.

EX-1014 0073
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

If the dc current gain β_F is relatively constant for small changes of I_C or for I_B, as is the case in virtually all applications, then

$$\beta_o \approx \beta_F \tag{2.18}$$

If Eq. (2.18) is valid, that is, if the *small-signal* current gain is approximately equal to the *large-signal* current gain, then one often does not differentiate between the two, but uses a single value of β for both ac and dc calculations. In order to simplify the subscript notation, the symbol β will be used to signify β_o and β_F interchangeably.

The dynamic resistance of the base–emitter junction r_π is also related to the direct bias current I_E as

$$r_\pi = \beta_o \frac{kT}{qI_E} = \frac{\beta_o}{g_m} \tag{2.19}$$

The output resistance r_o is due to the Early effect, and can be related to the Early voltage V_A of Figure 2.7 as

$$r_o = \frac{V_A}{V_T g_m} \tag{2.20}$$

For typical IC transistors, r_o is in the range of 50–100 kΩ.

The collector series resistance r_{cs}, is the bulk resistance of the semiconductor region through which the collector current must flow from the collector contact to the active collector–base junction. In discrete transistors, r_{cs} is negligibly small (in the 5–30Ω range) due to direct contact to the backside of the chip (see Fig. 2.1*a*). However, in monolithic IC transistors, r_{cs} is much higher (typically in the range of 100–500 Ω for small-geometry devices), depending on the device layout and the epitaxial layer thickness and resistivity. This will be discussed further in the next section, in connection with other device parasitics which are unique to integrated transistor structure.

C_π represents the total effective capacitance associated with the base–emitter junction and can be expressed as

$$C_\pi = C_D + C_{JE} \tag{2.21}$$

where C_D is the emitter diffusion capacitance corresponding to the charge in transit through the base region, and C_{JE} is the physical junction capacitance associated with the emitter–base junction. The latter is analogous to C_c of the collector–base junction, and assuming a graded junction, it can be expressed as

$$C_{JE} \approx A_E \left(\frac{q\epsilon N_B}{V_J} \right)^{1/3} \tag{2.22}$$

where A_E = area of the emitter–base junction
N_B = base impurity concentration at the emitter–base junction
V_J = net voltage across the junction

EX-1014 0074
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

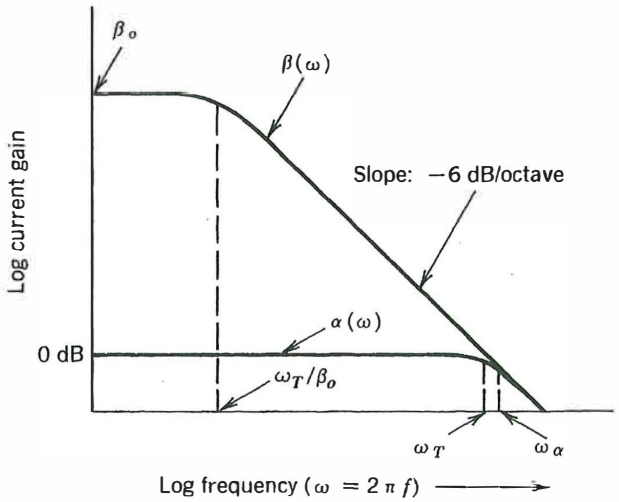


FIGURE 2.11. Frequency dependence of transistor current gain.

The frequency dependence of the common-base current gain α can be closely approximated as a one-pole gain rolloff function,

$$\alpha(\omega) = \frac{1}{1 + j\omega/\omega_\alpha} \tag{2.23}$$

where $\omega_\alpha = 2\pi f_\alpha$ is known as the *alpha cutoff* frequency. Since the common-emitter current gain is related to α by Eq. (2.8), the frequency dependence of β can be, to a first order, approximated as*

$$\beta(\omega) = \frac{1}{1 - \alpha} \frac{\beta_o}{1 + j\omega/\beta_o\omega_\alpha} \tag{2.24}$$

However, in the case of graded base transistor structures, it has been experimentally observed that an additional phase delay is introduced into the common-emitter current gain expression above and beyond that predicted by Eq. (2.24). This *excess phase* effect can be incorporated into $\beta(\omega)$ by means of an empirical phase-delay factor by rewriting Eq. (2.24) as

$$\beta(\omega) = \frac{\beta_o \exp(-jm\omega/\omega_\alpha)}{1 + j\omega/\beta_o\omega_\alpha} \tag{2.25}$$

where the excess phase factor m is approximately 0.4 for integrated transistors.

Figure 2.11 shows a plot of the magnitude of $\beta(\omega)$ as a function of frequency in a logarithmic scale. The frequency at which the common-emitter current gain reaches unity is known as the *unity current gain–bandwidth product* ω_T for the

*Note that in this analysis, it is assumed that small-signal low-frequency incremental gain is approximately the same as dc gain, that is $\beta_F \approx \beta_o$ similar to the assumption implied by Eq. (2.18).

EX-1014 0075
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

transistor. It can be shown that ω_T is related to ω_α as

$$\omega_T = \frac{\omega_\alpha}{1 + m\alpha} \approx \frac{\omega_\alpha}{1 + m} \tag{2.26}$$

Due to −6-dB/octave rolloff of $\beta(\omega)$ at high frequencies, the beat cutoff frequency ω_β can be related to ω_T as

$$\omega_\beta = \frac{\omega_T}{\beta_o} \tag{2.27}$$

Normally ω_T is measured with a current source input into the base terminal, and with the collector ac grounded. Then, referring to the equivalent circuit of Figure 2.10 and neglecting r_{cs}, ω_T can be expressed as

$$\omega_T \approx \frac{\beta_o}{(C_\pi + C_c)r_\pi} = \frac{g_m}{C_\pi + C_c} \tag{2.28}$$

Thus, once the ω_T of the transistor is known at a given bias setting, the base–emitter capacitance C_π of Figure 2.10 can be estimated as

$$C_\pi \approx \frac{g_m}{\omega_T} - C_c \tag{2.29}$$

Figure 2.12 shows the typical f_T versus current characteristics for a small-signal integrated *npn* transistor. The reduction of f_T at low current levels is mostly due to the emitter–base junction capacitance C_{JE}. The drop at high currents is a result of current crowding and the conductivity modulation effects in the base region. For a typical small-signal transistor of the geometry shown in Figure 2.4, the maximum value of f_T is in the range of 600–800 MHz and peaks at a current range of approximately 1–5 mA.

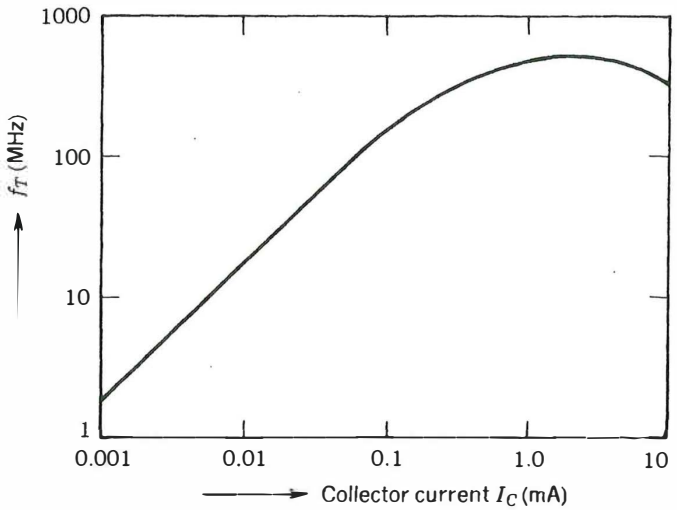


FIGURE 2.12. Typical f_T versus current characteristics of a small-geometry *npn* transistor.

EX-1014 0076
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Parasitic Elements in IC Transistors

The simplified hybrid-π model of Figure 2.10 shows the *intrinsic* equivalent circuit associated with *any* bipolar transistor, whether it is discrete or integrated. However, the basic fabrication processes associated with the integrated *npn* transistor also add a number of unique parasitics to the device structure in the form of additional series resistances and shunt capacitances. Figure 2.13 shows the location of these additional parasitics. For illustrative purposes, the figure is not drawn to scale.

The three parasitic resistances associated with the device structure are the base-spreading resistance r_b the collector series resistance r_{cs}, and the emitter series resistance r_{ex}. Their relative values and effects are briefly described below.

Base-Spreading Resistance r_b. This resistance comes about because of the electrical contact to the base region being physically removed from the active base region, directly below the emitter surface, as illustrated in Figure 2.14. It is made up of two sections in series, r_{b1} and r_{b2}, as shown. r_{b1} is the bulk resistance from the base contact to the edge of the active emitter area and can be easily calculated from the knowledge of the sheet resistivity of the base diffusion process. r_{b2} is a nonlinear resistance since it is spread two-dimensionally through the active base region, and the current flow out of it is distributed over the active base area. At low currents, its value is approximately a factor of 5–8 *higher* than r_{b1}; however, it drops to approximately 50% of its low-current value at higher current levels, where the carrier injection from the

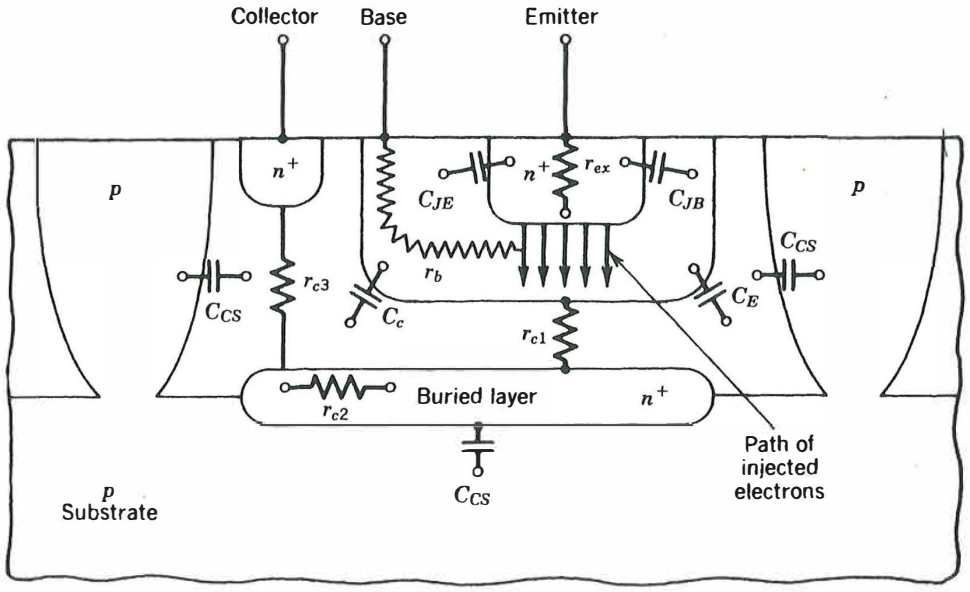


FIGURE 2.13. An illustration of inherent parasitics associated with junction-isolated *npn* transistor. (After Ref. 4).

EX-1014 0077
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

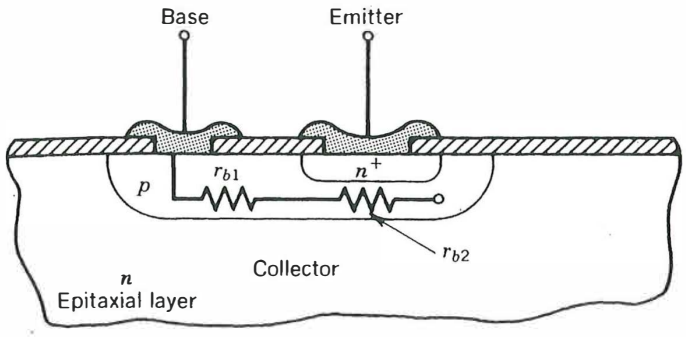


FIGURE 2.14. Components of base resistance r_b.

emitter into the base region takes place primarily along the emitter periphery nearest to the base contact.

In the typical small-geometry *npn* transistor of Figure 2.4, with a typical base sheet resistance of 150 Ω/□, the value of r_{b1} is approximately 60 Ω, and r_{b2} varies from 400 Ω at low currents to approximately 200 Ω at collector currents in excess of 1 mA.

The base-spreading resistance, r_b is particularly detrimental for low-noise or high-frequency applications. In low-noise transistors, it is often the dominant source of thermal noise. It degrades the transistor frequency response, particularly if the transistor is driven from a low-impedance signal source, since it serves as a voltage divider and reduces the effective signal voltage v_b in Figure 2.10, which reaches the active region of the transistor. Thus, to minimize the above problems, low-noise and high-frequency transistors are designed with multiple base contacts and long, stripe-shaped emitters which maximize the emitter periphery facing the base contact.

Collector Series Resistance r_{cs}. This is the series resistance of the total current path which must be traversed from the physical collector contact to the active collector area, just below the emitter–base junction. As shown in Figure 2.15, it is made up of three components, r_{c1}, r_{c2}, and r_{c3}. In general, r_{c2}, which is set by the sheet resistance of the subepitaxial n^+-type buried layer, is relatively small compared to r_{c1} and r_{c3}, which depend on the epitaxial layer thickness and resistivity. For the small-geometry *npn* transistor of Figure 2.4 with a 20-Ω/□ buried layer and a 15-μm 2.5-Ω/cm epitaxial layer, the respective values of these resistances are as follows:

$$r_{c1} \approx r_{c2} = 150\ \Omega \qquad r_{c2} \approx 30\ \Omega \qquad r_{cs} = r_{c1} + r_{c2} + r_{c3} = 330\ \Omega$$

The collector series resistance is particularly important for the saturated operation of the transistor, since at high current levels it is the main contributor to the saturation voltage V_{CEsat} across the transistor. In such applications, r_{cs} is often referred to as the *saturation resistance* R_{sat}. In high-current transistors, its effects are minimized by surrounding the active transistor with a large collector-

EX-1014 0078
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

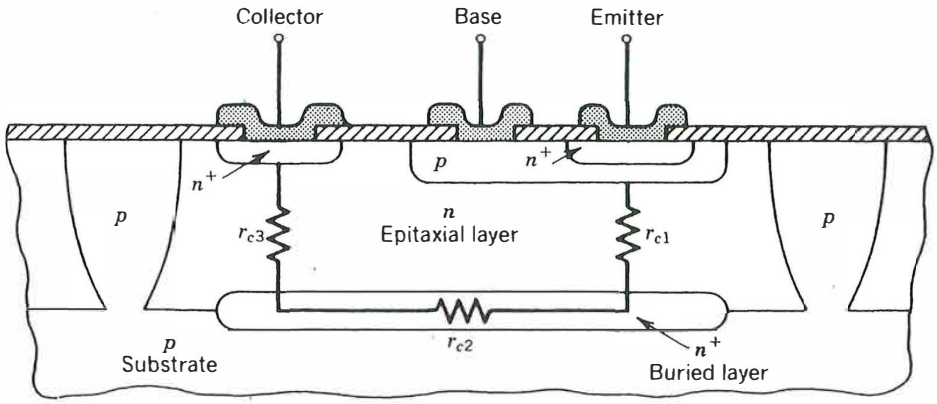


FIGURE 2.15. Components of collector series resistance r_{cs}.

contact area, and by using deep n^+ diffusion (see Fig. 1.16) to eliminate the contribution of r_{c3}.

Emitter Series Resistance r_{ex}. This is the bulk resistance of the emitter region between the emitter contact and the emitter–base junction. Since the emitter area is very heavily doped, its value is of the order of several ohms. In general, its effect in the ac equivalent circuit is negligible, except in the case of high-current transistors, where it may affect the current distribution within the emitter area. This subject is discussed further, in connection with power control and regulator circuits, in Chapter 10.

Parasitic Capacitances. As indicated in Figure 2.13, there are several parasitic junction capacitances associated with the integrated *npn* transistor structure. These are the emitter–base junction capacitance C_{JE}, the base–collector junction capacitance C_{CJ}, and the collector–substrate junction capacitance C_{CS}. The values of these capacitors depends on the total area of the junction, its impurity profile, and the *net* applied voltage across the junction that is, the applied voltage, V minus the built-in junction potential ψ_o. As will be described in Chapter 3, the general expression for any arbitrary junction capacitance, C_X can be written as

$$C_X = \frac{C_{XO}\, A_X}{(1 - V/\psi_o)^n} \tag{2.30}$$

where V = applied voltage across the junction (positive for forward bias)
C_{XO} = capacitance per unit area of junction under zero bias (i.e., $V = 0$)
A_X = total junction area.

The exponent n is equal to $\frac{1}{2}$ for a step junction and equal to $\frac{1}{3}$ for a linearly graded junction. For most semiconductor junctions, either the step or the linear-grading approximation holds very closely.

EX-1014 0079
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

ψ_o is the built-in junction potential which depends on the impurity concentration on either side of the junction. Assuming a step, or a very steeply graded junction, it can be expressed as

$$\psi_o = V_T \ln\left(\frac{N_A N_D}{n_i^2}\right) \tag{2.31}$$

where N_A and N_D are the acceptor and donor concentrations on the respective sides of the junction, and n_i is the intrinsic carrier concentration in pure silicon ($n_i \approx 1.5 \times 10^{10}\,\text{cm}^{-3}$ at 300°K). For IC transistors, typical values of ψ_o are 0.7 V for base–emitter junctions and 0.5–0.6 V for base–collector and collector–substrate junctions.

Table 2.2 gives typical values of the unit-area zero-bias capacitances C_{JEO}, C_{CO} and C_{CSO}, the corresponding values of the exponent n and the built-in junction potential ψ_o for the three commonly used analog IC processes. Once these values are known and the corresponding junction area is measured from the device layout, the values of each of these three parasitic capacitances can be directly calculated from Eq. (2.30).

Summary of *npn* Device Parameters

Many of the *npn* device parameters, including the device parasitics discussed in the previous sections, depend somewhat on the specifics of the IC fabrication processes and the impurity profiles, even when the lateral dimensions (i.e., the layout) of the device are fixed.

Basic bipolar analog IC fabrication processes available from many IC suppliers differ in their specifics and details. However, the majority of the bipolar IC processes tend to fall into one of the following general categories:

1. *Low-Voltage Process.* This process uses an epitaxial layer thickness in the range of 8–10 μm with an epitaxial layer resistivity of 0.8–1 Ω/cm. It results in an *npn* bipolar transistor with a collector–base breakdown voltage LV_{CEO} of 20 V.
2. *Medium-Voltage Process.* This process uses 2–2.5 Ω/cm epitaxial layer resistivity, with an epitaxial layer thickness of approximately 15 μm, and gives a breakdown voltage of approximately 30 V. The impurity profile of Figure 1.13 and the device geometry of Figure 2.4 correspond to this process.
3. *High-Voltage Process.* This process is used for circuits such as operational amplifiers, which are designed to operate with supply voltages as high as 40 V. It uses 4–5-Ω/cm resistivity, 17–19 μm thick epitaxial layers, and gives a device breakdown voltage in excess of 45 V.

The typical small-geometry device layout is applicable to all three processes. However, in the case of the low-voltage process, a smaller base-to-isolation

EX-1014 0080
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

TABLE 2.2. Summary of Typical Values for Parasitic Junction Capacitances for Monolithic *npn* Transistors

Junction Characteristics	Low-Voltage Process: 1 Ω-cm, 9-μm Epitaxial Layer, 20-V Device	Medium-Voltage Process: 2.5 Ω-cm, 14-μm Epitaxial Layer, 30-V Device	High-Voltage Process: 5 Ω-cm, 18-μm Epitaxial Layer, 40-V Device
Emitter–base junction			
Zero-bias capacitance C_{JEO}	10^{-3} pF/μm^2	10^{-3} pF/μm^2	10^{-3} pF/μm^2
Built-in potential ψ_{JEO}	0.7 V	0.7 V	0.7 V
Exponent n	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
Base–collector junction			
Zero-bias capacitance C_{CO}	2.3×10^{-4} pF/μm^2	1.4×10^{-4} pF/μm^2	10^{-4} pF/μm^2
Built-in potential ψ_{CO}	0.6 V	0.58 V	0.55 V
Exponent n	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
Collector-isolation junction			
Zero-bias capacitance C_{CSO}	10^{-4} pF/μm^2	10^{-4} pF/μm^2	10^{-4} pF/μm^2
Built-in potential ψ_{CSO}	0.52 V	0.55 V	0.58 V
Exponent n	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

EX-1014 0081
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

spacing can be used ($\approx$ 18 μm versus 25 μm as shown in the figure) to result in a smaller overall isolation pocket size.

Table 2.3 gives a summary of all the relevant device parameters, including the parasitic elements, associated with a small-geometry transistor structure similar to that of Figure 2.4, fabricated with each of these three popular IC fabrication processes.

2.2. *npn* TRANSISTORS FOR SPECIAL APPLICATIONS

Unlike the discrete circuit designer, monolithic IC designers can exercise control over the geometry and the layout of the devices at their disposal. Thus, by properly modifying and scaling the dimensions of their devices, they can optimize their performance for special applications, such as high-current or high-voltage operation. In this section some of these special *npn* transistor structures will be reviewed.

TABLE 2.3. Summary of Device Parameters for Small-Geometry *npn* Transistor

Device Parameter	Low-Voltage Process: 1Ω-cm, 9-μm Epitaxial Layer, 20-V Device	Medium-Voltage Process: 2.5Ω-cm, 14-μm Epitaxial Layer, 30-V Device	High-Voltage Process: 5Ω-cm, 18-μm Epitaxial Layer, 40-V Device
Forward-current gain β_F	150	200	200
Reverse-current gain β_R	2	2	2
Early Voltage V_A	50 V	80 V	120 V
Collector–base leakage I_{CO}	10^{-7} mA	10^{-7} mA	10^{-7} mA
Emitter–base breakdown BV_{EBO}	6.5 V	6.8 V	7.0 V
Collector–base breakdown BV_{CBO}	40 V	60 V	90 V
Collector–emitter breakdown BV_{CEO}	25 V	40 V	50 V
Lowest sustaining voltage LV_{CEO}	20 V	30 V	40 V
Base resistance r_b			
135-Ω/□ process	200 Ω	200 Ω	200 Ω
200-Ω/□ process	400 Ω	400 Ω	400 Ω
Collector series resistance r_{cs}	70 Ω	120 Ω	200 Ω
Emitter series resistance r_{ex}	2 Ω	2 Ω	2 Ω

EX-1014 0082
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

High-Current Transistors

There are two parameters which degrade the transistor characteristics at high current levels. These are the β falloff at high currents and the large voltage drop across the collector series resistor r_{cs}. As described earlier in connection with Figure 2.8, β falloff at high currents is due to the decrease of emitter efficiency and the crowding of the emitter current to the periphery of the emitter nearest to the base contact. These effects can be minimized by using an interdigitated transistor structure, as shown in Figure 2.16 which is made up of alternating stripes of emitter and base contact "fingers." Such a structure maximizes the emitter periphery-to-area ratio and reduces the base-spreading resistance r_b.

The collector series resistance r_{cs}, which is also referred to as the saturation resistance R_{sat}, is detrimental to high-current operation because of excessive voltage drop and/or power dissipation generated across it. This resistance can be minimized by using a "wrap around" collector contact structure, as shown in Figure 2.16, as well as using a deep n^+ diffusion from the collector contact to the n^+-type buried layer (See Fig. 1.16).

High-Voltage Transistors

In certain applications, such as interfacing with high-voltage display systems, analog integrated circuits are required to operate with supply voltages in excess

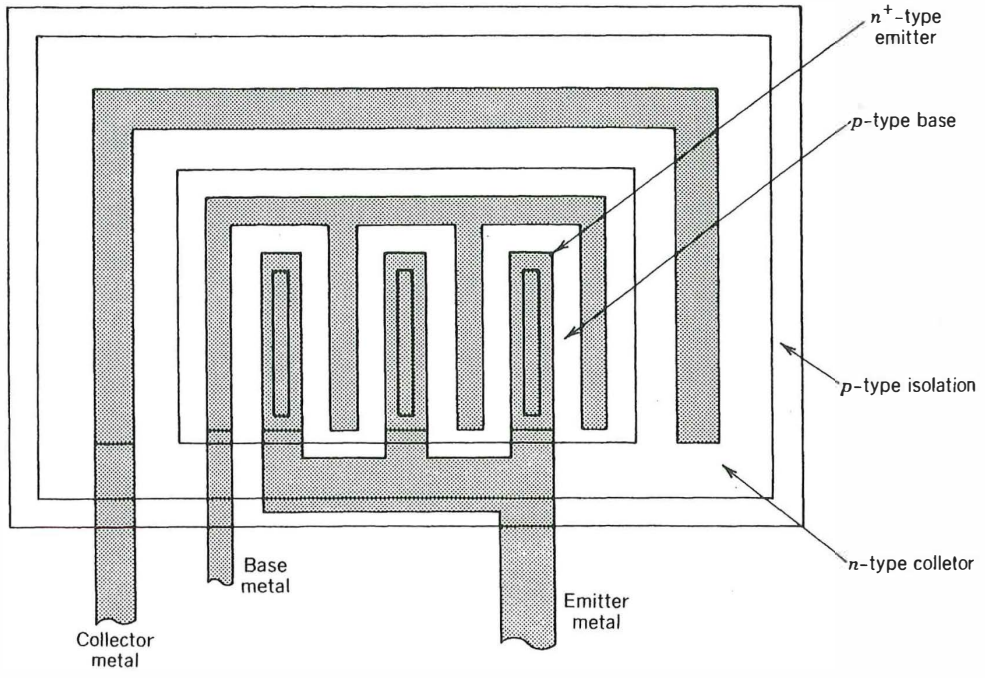


FIGURE 2.16. Typical lateral geometry of a high-current transistor.

EX-1014 0083
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

of 50 V. By proper device design and layout, it is possible to extend the breakdown voltage characteristics of monolithic *npn* transistors up to 100 V. There are two main limitations to high-voltage operation: (1) the breakdown voltages of semiconductor junctions and (2) the surface inversion problem where a parasitic *p*-channel MOS transistor may be formed on the epitaxial layer surface.

Breakdown Voltage. The junction breakdown voltage imposes a fundamental limit on the high-voltage capability of monolithic devices, as shown in Figure 2.9. However, the data of Figure 2.9 only refer to the avalanche breakdown within the "bulk" silicon and neglect the surface breakdown effects. The discontinuity of the silicon lattice at the device surface, as well as the presence of electrostatic charge trapped at the Si–SiO_2 interface, can modify the electrostatic field distribution within the junction at or near the device surface. This can lead to a narrowing down of the junction depletion layer at the surface, or at the sharp corners of the device, and results in a localized surface breakdown at a lower voltage than that predicted in Figure 2.9. This effect is illustrated in Figure 2.17*a* in terms of a simple *p*-*n* junction. This surface breakdown can be greatly reduced or eliminated by using an electrostatic shield, or *field plate*, over the junction surface, as shown in Figure 2.17*b*. When connected to a more negative potential than the *n* side of the *p*-*n* junction, this field plate tends to spread out the depletion layer near the surface and, thus, avoids surface breakdown.[6] In the case of a high-voltage transistor, this field plate can be formed by simply overlapping the base and the emitter metal interconnection traces over the base–collector junction.

In fabricating a high-voltage transistor with a breakdown voltage well in excess of 50 V, one would use a high-resistivity epitaxial layer (typically 5–8 Ω/cm) with extended thickness of the order of 25 μm. The base and emitter junctions are made relatively deep, typically on the order of 3.5 and 2.5 μm, respectively, to avoid surface breakdown effects, and sufficient space is pro-

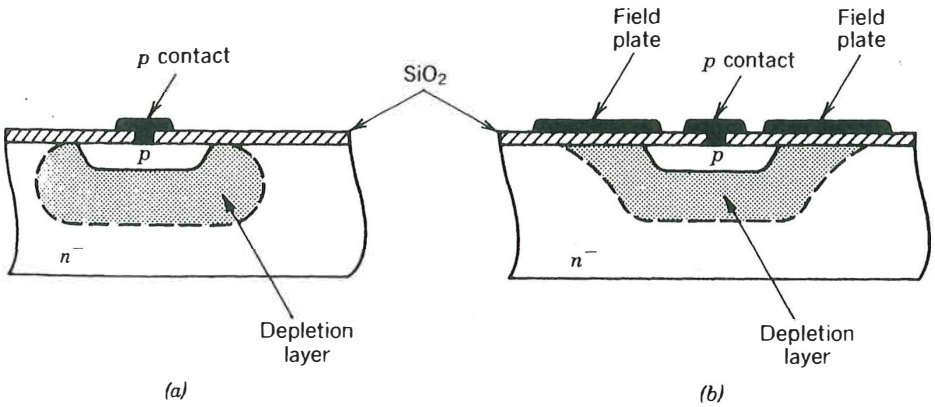


FIGURE 2.17. Effect of a surface field plate on the *p*-*n* junction depletion layer: (*a*) Without field plate and (*b*) with field plate.

EX-1014 0084
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

vided within the collector pocket for the extension of the collector–base and the collector–isolation depletion layers, without touching each other or the n^+-type buried layer.

Figure 2.18 shows the typical layout of such a high-voltage transistor. Note that the base and the emitter metal traces are made to overlap the base–collector junction to serve as the field plate to minimize the surface breakdown effects. Also, the base-to-isolation spacing is increased significantly beyond what is shown in Figure 2.4 to account for increased isolation side diffusion and collector–base and collector–isolation depletion layers.

Surface Inversion. One of the problems associated with high-voltage integrated circuits is the formation of a parasitic p-channel MOS transistor on the surface of the lightly doped n-type epitaxial layer surface. The presence of a conducting metal layer above the oxide layer, biased at a *negative* voltage relative to the epitaxial layer, creates an electric field within the oxide layer

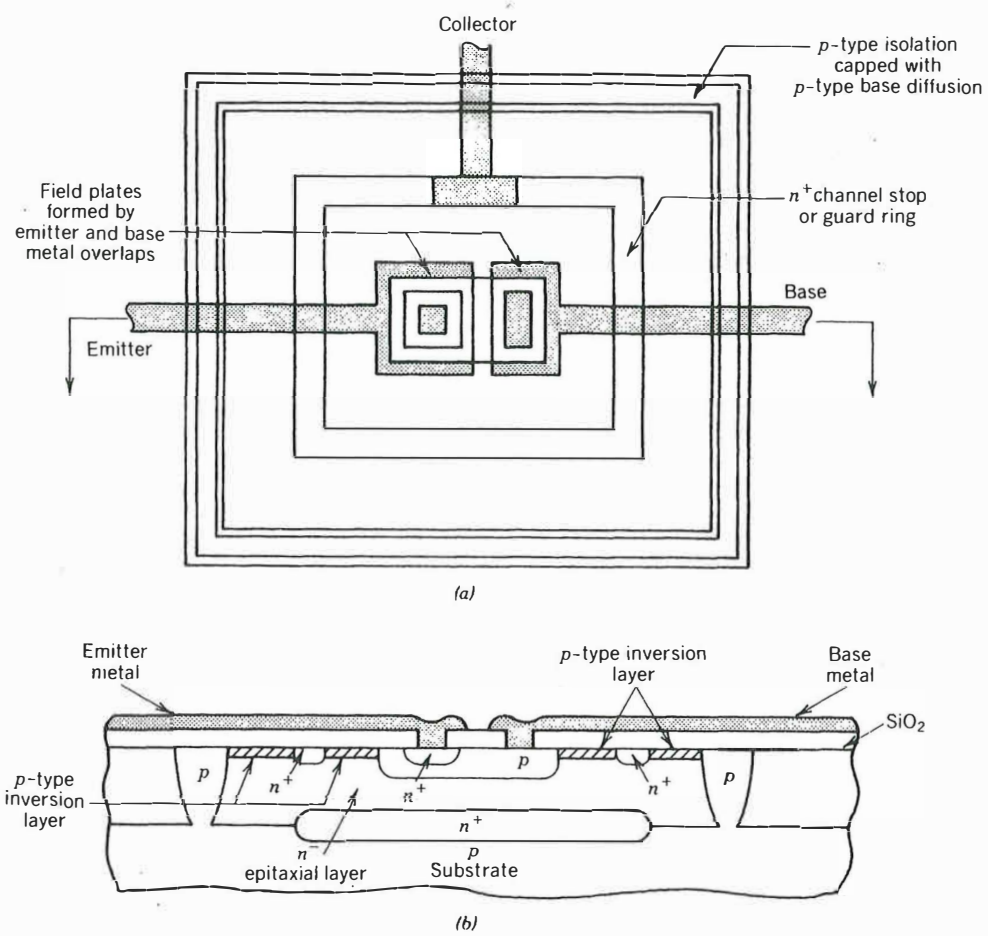


FIGURE 2.18. Device layout and cross section of a high-voltage *npn* transistor using field plate and n^+ channel stops.

EX-1014 0085
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

which pushes the free electrons *away* from the epi–oxide interface, and creates a p-type inversion layer along the epitaxial layer surface directly under the metal trace. Typically, with lightly doped epitaxial layers, with resistivities $\geq 5\ \Omega$/cm, this inversion layer sets in at voltage differentials of 35–40 V across the SiO_2 layer. If this parasitic p-type "channel" is not blocked by a heavily doped n^+ guard ring, or *channel stop*, completely surrounding the base region, it would result in base–isolation short circuit. Since the heavily doped n^+ ring, which also serves as the collector contact, does not invert, it serves as a channel stop to inhibit this parasitic p-channel MOS transistor action. Since this parasitic channel appears only under the metal traces, in many cases it is not necessary to encircle the entire base area with a n^+ channel stop, but instead, the channel stops can be placed in discrete segments under the metal traces. The layout example of Figure 2.18*a* shows a device structure with a continuous guard-ring-type n^+ channel stop.

It should also be noted that in high-voltage integrated circuits it is customary to deposit an additional p-type layer on top of the isolation during the subsequent p-type base diffusion. This second p-type layer on the isolation surface, which is often called a *p-cap*, also serves as a channel stop to avoid parasitic n-type channel formation over the isolation region. Such a *parasitic* channel may be formed when a metal trace with a positive potential, such as the V^+ line, passes over the p-type isolation wall.

Superbeta Transistors

In certain analog circuits, such as the input stages of operational amplifiers, it is necessary to have a very high input impedance and low input bias currents. For such an application, the β of a typical integrated *npn* transistor is not high enough, since the actual device design requires a compromise between the current gain and the voltage breakdown requirements [see Eq. (2.14)].

It is possible to increase the value of β by improving the base transport efficiency β^* [see Eq. (2.11)]. This can be done by using a very narrow base structure ($W_B \approx 2500$ Å). However, the collector–emitter voltage breakdown of such a device structure is limited to the 3–4 V range because of the collector–base depletion layer punching through the active base region into the emitter. Figure 2.19 shows the typical current–voltage characteristics of such an ultrahigh-gain, or superbeta, transistor. Such transistors exhibit typical β values of 2000–5000 at collector current levels of 10–20 μA. Due to excessive base width modulation effects, the Early voltage V_A and the output resistance r_o of a superbeta transistor are typically an order of magnitude lower than those of the conventional *npn* transistor. Typical values of V_A are in the 4–10 V range.

The superbeta transistors can be fabricated simultaneously with conventional *npn* bipolar transistors, with the addition of one extra masking and emitter diffusion step. This is done using the *two-step emitter diffusion* process described in Section 1.10 (see Figure 1.17).

In circuit design, superbeta transistors are often used together with a conventional IC transistor which can provide the necessary voltage protection.

EX-1014 0086
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

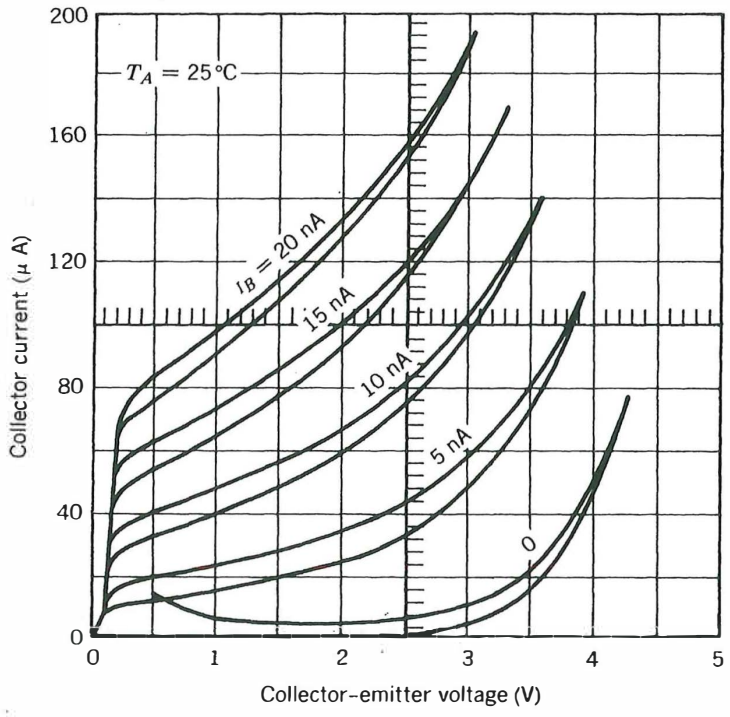


FIGURE 2.19. Curve-tracer display of typical current–voltage characteristics for a superbeta transistor.

Figure 2.20 shows a composite connection of a superbeta transistor with a lateral *pnp* transistor where the base–emitter diode of the *pnp* transistor clamps the voltage swing across the superbeta transistor to a level below its breakdown voltage. Thus, the resulting transistor structure has the effective β of the punch-through *npn* transistor and the breakdown voltage of the lateral *pnp* transistor. The current source I_X shown in the figure is included to ensure proper direction

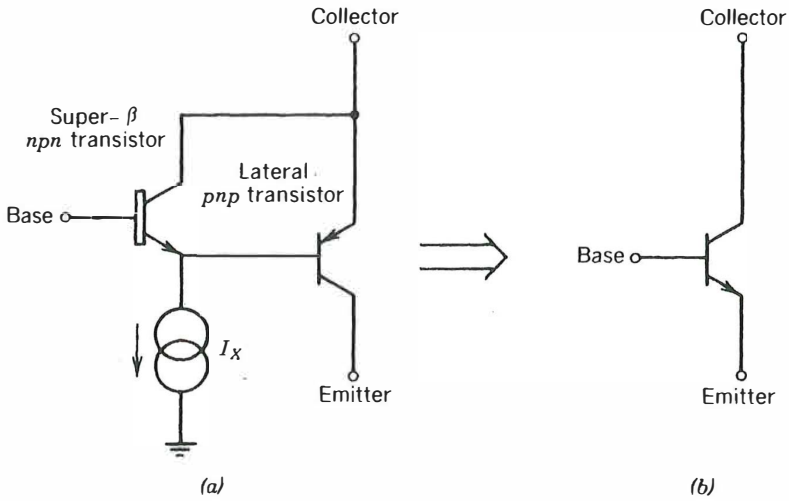


FIGURE 2.20. Composite connection of a superbeta *npn* transistor with a lateral *pnp* transistor to improve breakdown characteristics: (*a*) Actual connection; (*b*) electrical equivalent circuit.

EX-1014 0087
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

of current flow out of the base of the *pnp* transistor; it is normally a small fraction of the collector current of the composite device.

Multiple-Emitter Transistor

One can easily enlarge the base area of the basic *npn* transistor structure and insert a number of separate emitter regions. This results in a multiple-emitter *npn* structure, as shown in Figure 2.21. Since the emitter region of an integrated *npn* transistor takes up the smallest amount of device area, the resulting device size and the required chip area are not increased significantly. Such a multiple-emitter *npn* transistor, which was traditionally used in TTL-type digital circuits, also finds a range of applications in analog IC design. Some of these applications, such as providing multiple-output emitter-follower circuits, buffered biasing, and temperature-compensated avalanche diodes, are discussed in more detail in Chapter 4.

One word of caution is in order regarding multiple-emitter transistors. If one of the emitters is biased in the off condition relative to the base potential while the other emitter is conducting, a parasitic lateral *npn* action may take place between the two emitters. In that case, a parasitic *npn* transistor will be formed, where the off emitter will serve as the collector for the electrons injected by the on emitter.

Inverted Operation of *npn* Transistor

In certain analog circuit applications, it may be advantageous to reverse the emitter and the collector terminals of an *npn* transistor and operate it in an inverted or reverse mode. This inverted operation is also the key concept behind the integrated injection logic (I^2L) circuits.

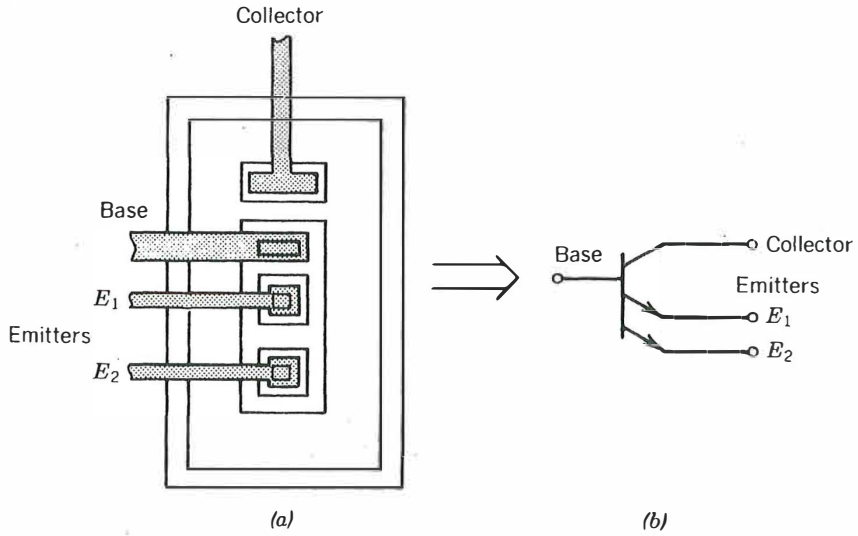


FIGURE 2.21. Multiple-emitter *npn* transistor: (*a*) Device layout; (*b*) equivalent device symbol.

EX-1014 0088
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

One advantage of the inverted operation of the *npn* transistor is that the normal emitter now becomes the collector of the inverted transistor, and that one can form a multiple-collector transistor in a common isolation tub by using the device structure shown in Figure 2.22.

In inverted operation, the transistor characteristics are very poor. The current gain, which is now given by the inverse beta β_R, is only a small percentage (typically 1–5%) of β_F. This is due to the following reasons. The emitter efficiency γ [see Eq. (2.10)] is degraded due to the emitter being doped lighter than the base, the base transport factor β^* is reduced due to the retarding field gradient in the base region; and a number of carriers injected into the base region have to diffuse a long distance to reach the collector region, thus making the effective base width appear to be much larger than W_B. In addition, a parasitic *pnp* transistor is also formed in the isolation tub between the base of both the *npn* transistor and the *p*-type isolation wall and the substrate, as shown in Figure 2.3*b*. However, the effects of this parasitic *pnp* transistor can be greatly reduced, or nearly eliminated, by surrounding the entire *p*-type base region in an n^+-type pocket by means of a deep n^+ diffusion, as shown in Figure 2.22. The collector–base breakdown of the inverted transistor is also limited to the 6–7 V range by the BV_{EBO} of the normal base–emitter junction.

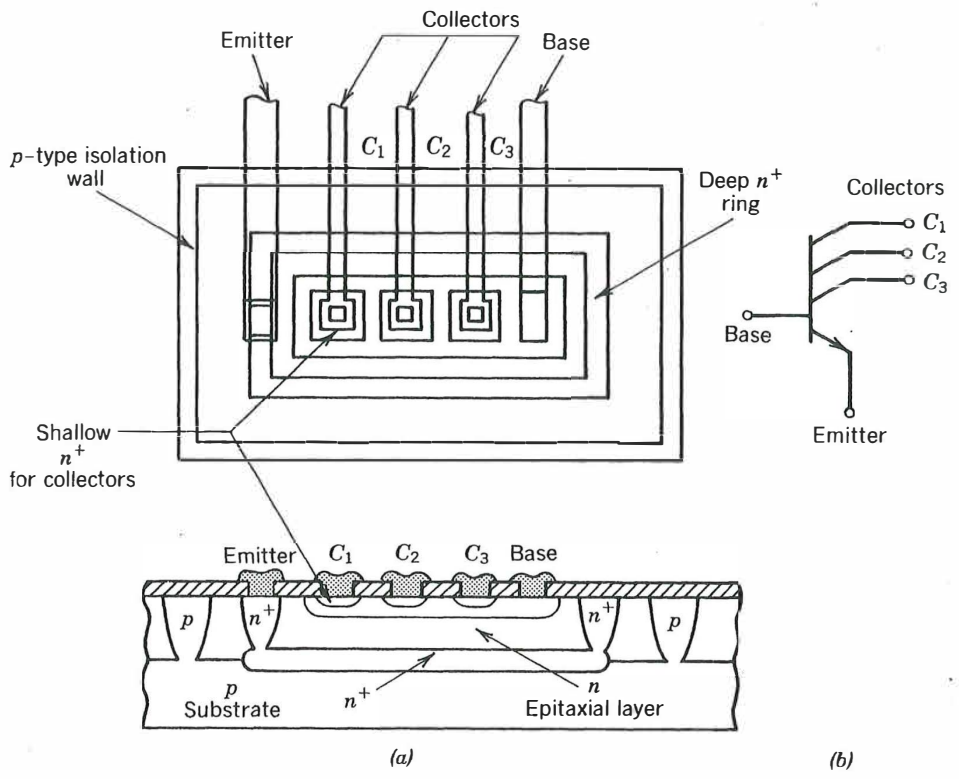


FIGURE 2.22. Device structure of a multiple-collector inverted *npn* transistor: (*a*) Device layout and cross section; (*b*) equivalent device symbol.

EX-1014 0089
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

In spite of all the drawbacks outlined above, the inverted transistor still offers some unique advantages in selected analog circuit applications. Some of these are listed below:

1. Under saturated conditions, its offset voltage is lower than that of the normal *npn* transistor. Thus, it makes a good grounding switch for shunting low-level currents (i.e., ≤ 1 mA) to ground.
2. Its collector–base capacitance is lower, since it is equal to C_{JE} under reverse-bias conditions. Thus, it can be used as a low-capacitance current source for biasing differential amplifier stages.
3. It allows a large number of independent *npn* current sources to be formed within a common isolation tub, using the *current mirror* configuration shown in Figure 4.2.
4. In the multiple-collector device structure of Figure 2.22, the fraction of the total emitter current reaching each of the collectors is proportional to the area of the collector–base junction. Thus, by varying the geometry of the collectors, the current distribution between them can be scaled.

2.3. *pnp* TRANSISTORS

Some analog circuit functions may require the use of complimentary bipolar transistors. In such cases, it is necessary to fabricate functional *pnp* transistors on the same substrate with the *npn* devices. For this purpose, a number of monolithic *pnp* transistor structures have been developed which are totally compatible with the standard *npn* bipolar process technology, that is, they can be fabricated simultaneously with the *npn* transistors without requiring additional diffusion or masking steps. Two such *pnp* transistors commonly used in analog IC design are the *lateral pnp* and the *substrate pnp* transistors. Both of these devices use the *lightly doped n*-type epitaxial region of the *npn* transistor as the base region. They are, in general, inferior to the basic *npn* transistor in their current-handling or high-frequency characteristics. However, they are extremely useful in biasing, dc level shifting and serving as active load devices in gain stages.

Lateral *pnp* Transistor

Figure 2.23 shows the plan view and the structural diagram of a typical lateral *pnp* transistor. The base region of the device is formed by the *n*-type epitaxial layer which serves as the collector of the *npn* transistors. The *p*-type base diffusion of the *npn* transistor is used to form the emitter and the collector regions of the lateral *pnp* transistor, the n^+-type emitter diffusion of the *npn* transistor is used to form the n^+-type contact region for the *pnp* base. In such a device structure, the transistor action takes place in the *lateral direction*, that is,

EX-1014 0090
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

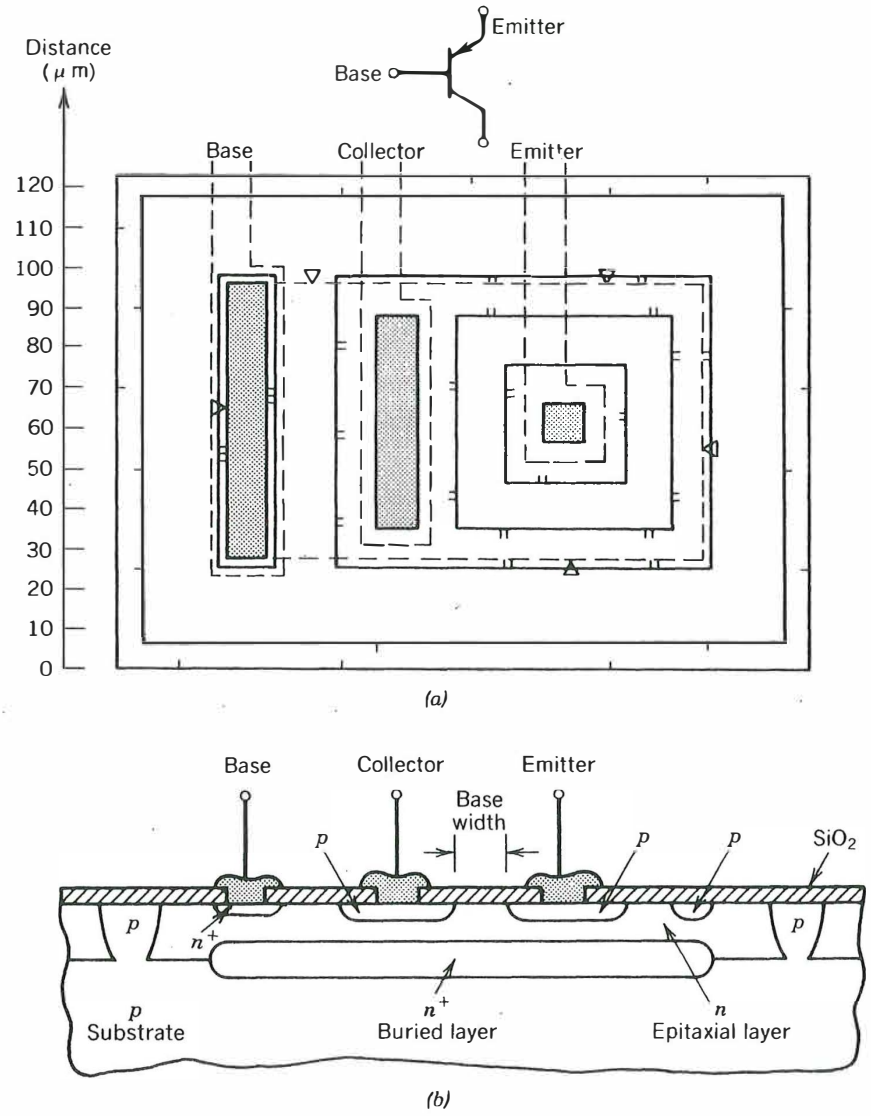


FIGURE 2.23. Typical device layout and cross section of a lateral *pnp* transistor.

parallel to the device surface. The minority carriers injected into the base diffuse laterally toward the collector region. In order to collect the great majority of carriers injected from the emitter, and to minimize parasitic current flow into the substrate, the collector region of the lateral *pnp* transistor is normally made to surround the emitter region along the lateral surface of the device as shown in Figure 2.23*a*. Although the device layout of the figure shows a square emitter surrounded by a square or rectangular collector region, it is also very common to use a circular emitter region surrounded with a circular collector ring.

Figure 2.24 illustrates the direction of current flow in terms of the minority carriers (holes) injected into the base region of the lateral *pnp* transistor. The

EX-1014 0091
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

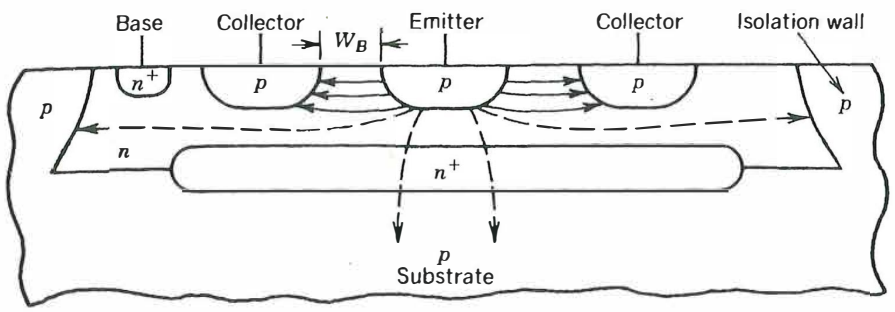


FIGURE 2.24. Minority carrier flow in lateral *pnp* transistor. Solid lines indicate normal carrier flow paths; dashed lines indicate parasitic carrier flow to substrate.

carrier transport across the base region is most efficient at or near the surface of the device, where the separation between the collector and the emitter is minimal. This minimum spacing is the effective base with W_B for the device. Due to masking tolerances, side-diffusion effects, and voltage breakdown requirements, W_B is constrained to be of the order of 6–12 μm. This value for the effective base width is much larger than that of a vertical *npn* transistor ($W_B \approx 0.7$ μm). Thus, the current gain and the frequency performance characteristics of the lateral *pnp* transistors are inferior to those of the *npn* devices.

The current gain β of the lateral *pnp* transistor is low (typically in the range of 5–50) due to its poor emitter efficiency and the wide base width. The emitter efficiency of the lateral *pnp* transistor is degraded due to two factors: (1) low impurity concentration in the *p*-type emitter region, and (2) small effective area of the emitter. This latter effect is due to the fact that only the lateral emitter edge facing the collector is active; the holes injected from the rest of the emitter have a much lower probability of reaching the collector. The wide base width W_B of the lateral *pnp* transistor also affects β_F adversely since it lowers the base transport factor β^* [see Eq. (2.11)].

Another effect which adversely influences the lateral *pnp* transistor current gain is the reduction of β^* due to the parasitic minority carrier flow from the emitter directly into the substrate, as shown by the dashed lines in Figure 2.24. Thus, the substrate serves as the collector of a parasitic *pnp* transistor (see Fig. 2.25). The n^+-type buried layer, which is still retained under the emitter and collector regions of the lateral *pnp* transistor, sets up a retarding field to reduce this parasitic *pnp* action. In normal operation, with a wrap around collector geometry and an n^+-type subepitaxial layer, as shown in Figure 2.23, this parasitic substrate current is in the range of 3–5% of the normal collector current.

The current-handling capability of the lateral *pnp* transistor is severely limited by the lightly doped *n*-type epitaxial layer which serves as the base region. Thus, as the collector current is increased, the high-level injection effects set in at the emitter base region, and the emitter efficiency and β_F fall off rapidly. For a small-geometry lateral *pnp* transistor with approximately 100-μm emitter periphery (see Fig. 2.23*a*), β_F starts falling off at approximately 20–50 μA of collector current. Figure 2.25 shows typical current gain characteristics of a

EX-1014 0092
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

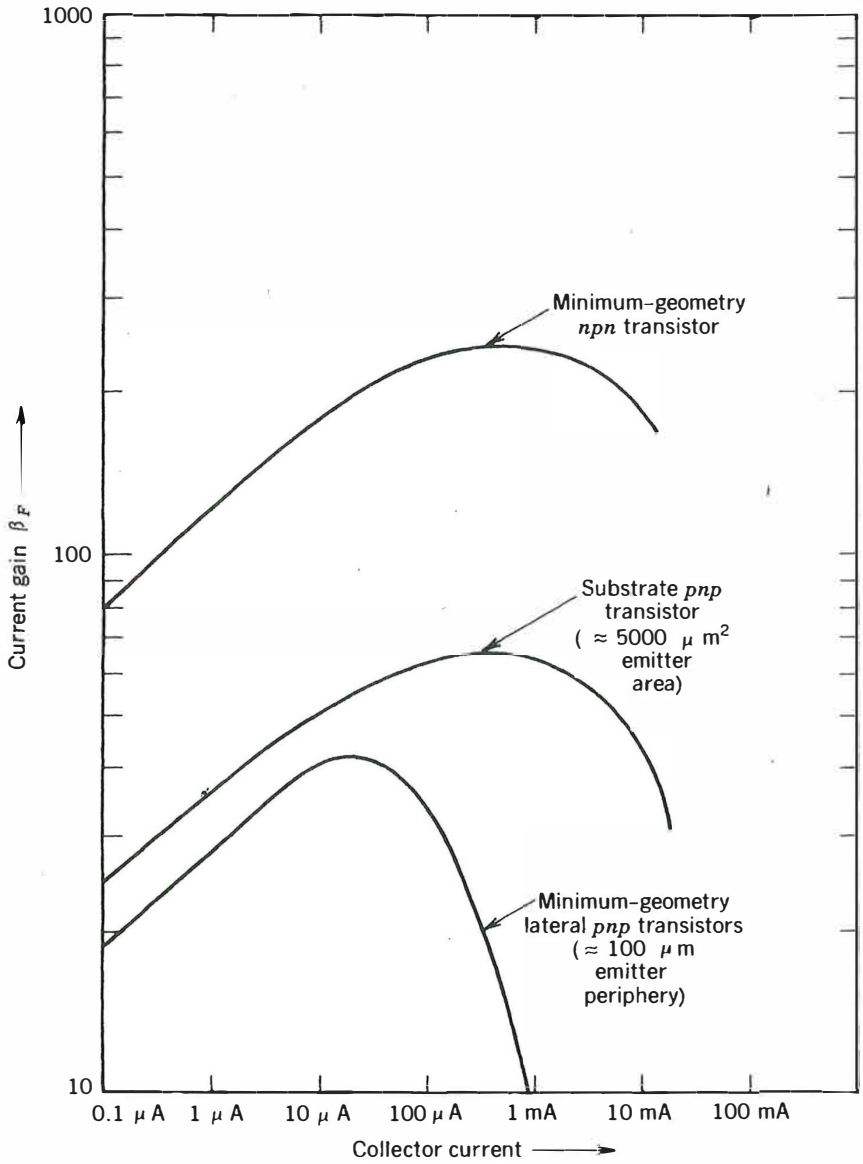


FIGURE 2.25. Comparison of current gain characteristics of the small-geometry *npn*, lateral *pnp*, and substrate *pnp* transistors.

small-area lateral *pnp* transistor as a function of collector current. Note that the current-handling capability of the lateral *pnp* transistor is approximately 50–100 times lower that that of a small-geometry *npn* transistor and its β_F is typically a factor of 10–50 lower.

The frequency response of the lateral *pnp* transistor is limited by the transit time of the carriers through the wide base region. It can be shown that this transit time is proportional to $(W_B)^2$. Similarly, the f_T of the lateral *pnp* transistor is inversely proportional to $(W_B)^2$. Since the base width of the later *pnp* transistor is approximately 10 times larger than that of the normal *npn* transistor (i.e.,

EX-1014 0093
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

0.8 μm vs. 8 μm), the f_T of the later *pnp* transistor is approximately 100 times lower than that of an *npn* transistor. Typical values of f_T for the lateral *pnp* transistors are in the range of 3–5 MHz.

Parasitics Associated with Lateral *pnp* Transistor. The basic lateral *p-n-p* transistor structure of Figure 2.23 has two inherent parasitic *pnp* transistors associated with it, as shown in Figure 2.26. The first Q_A, is the parasitic *pnp* transistor between the emitter of the *pnp* transistor and the *p*-type substrate and isolation walls. Since the emitter-base junction of the lateral *pnp* transistor Q_1 is normally forward biased, the parasitic transistor Q_A is active at all times. However, its effective current gain is greatly reduced by the use of an n^+-type buried layer and that wraparound collector structure of Figure 2.23. Thus, its net effect is shunting a small amount of the normal emitter current of Q_1 to the substrate. This current, shown as I_X in Figure 2.26, is typically of the order of 3–5% of I_E.

The second parasitic *pnp* transistor, Q_B, formed between the collector of Q_1 and the *p*-type substrate, is normally off since the collector–base junction of Q_1 is reverse biased. However, if the *pnp* transistor Q_1 were operated in its inverse mode (i.e., its emitter and collector functions are reversed), this parasitic transistor would turn on and shunt a substantial amount of current I_Y to the substrate. The parasitic *pnp* transistor Q_B also comes into conduction when Q_1 is driven into saturation. Both of these conditions *must be* avoided by proper circuit design. In certain specialized applications, where the later *pnp* transistor may be driven into saturation, or must operate in an inverted mode, the effects of the parasitic *pnp* transistor Q_B can be reduced by placing a deep n^+ ring around the whole *pnp* transistor and the *p*-type isolation pocket. Such a deep n^+ ring would reduce the value of I_Y to approximately 5–10% of the collector current; however, it would cause a significant increase in the chip area required by the device.

Multiple-Collector *pnp* Transistor. The collector region of the lateral *pnp* transistor can be split into several segments around a center emitter area, as shown in Figure 2.27. This results in a multiple-collector transistor structure

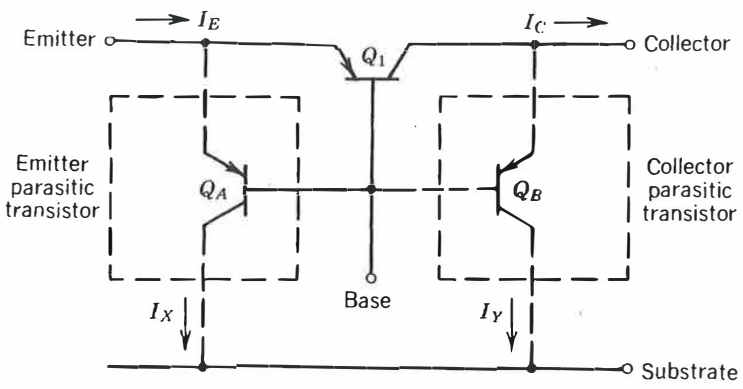


FIGURE 2.26. Parasitic *pnp* transistors associated with the lateral *pnp* transistor.

EX-1014 0094
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

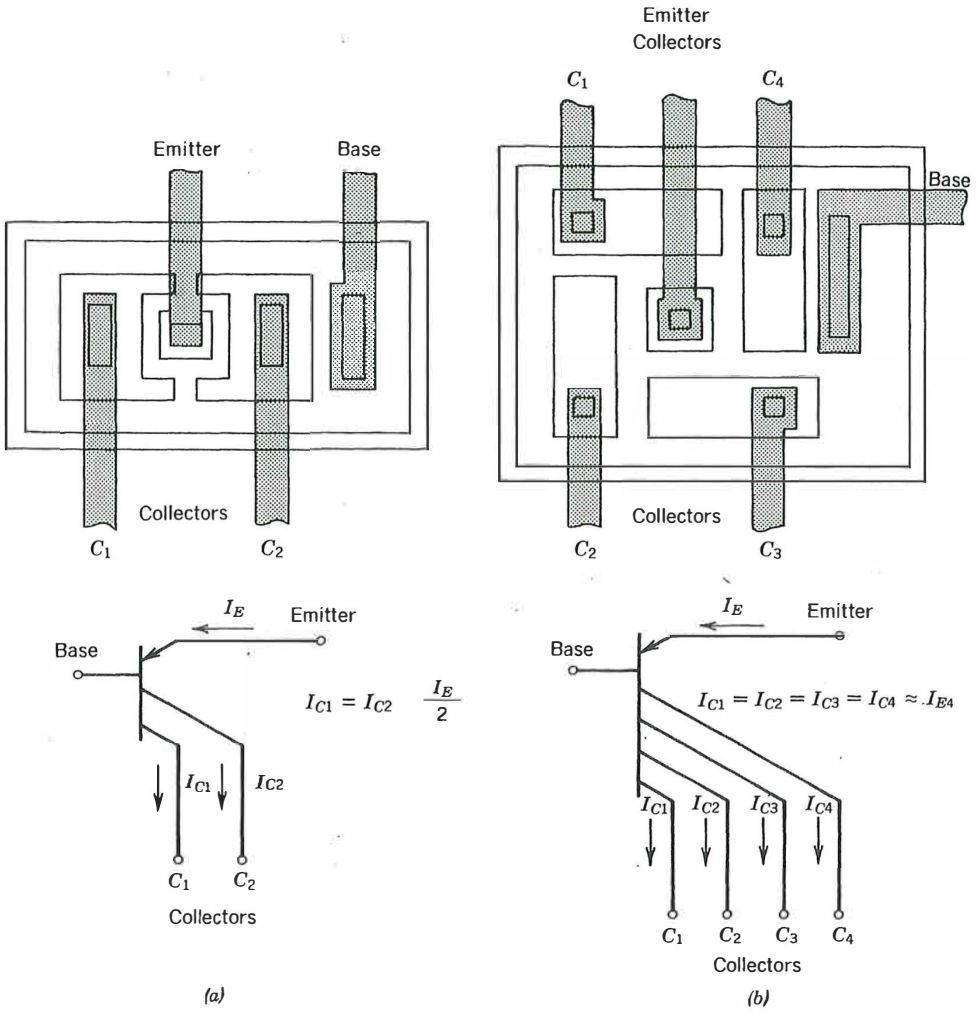


FIGURE 2.27. Typical multiple-collector lateral *pnp* transistor structures.

which is very useful in a number of design applications. Since the collector current associated with each collector segment is directly proportional to the collector periphery facing the emitter, the partitioning of the emitter current between the different collectors can be controlled and scaled by varying the corresponding collector periphery. As will be described in Chapter 4, this scheme is often used to create accurate and stable ratios of currents which are scaled by the device layout.

When multiple-collector *pnp* transistors are used as ratioed current sources or active loads, care must be taken to ensure that none of the collectors are driven into saturation. Otherwise a parasitic *pnp* action would result between the collector *segments*, as well as between the collectors and the substrate.

Composite Connection of Lateral *pnp* and Vertical *npn* Transistors. The low β of the lateral *pnp* transistor can be improved by combining it with an integrated

EX-1014 0095
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

npn transistor to form a composite transistor, as shown in Figure 2.28. The polarity and the electrical characteristics of such a composite device are equivalent to that of a single *pnp* transistor having a higher current gain β_T, given as

$$\beta_T = \beta_n \beta_p \tag{2.32}$$

where β_p and β_n are the respective current gains of the *pnp* and *npn* transistors. The transconductance of the composite device is the same as that of the *npn* transistor. Although the emitter of the *npn* transistor acts as the collector of the composite device, the resulting output impedance is still quite high since the base of the *npn* transistor is driven by a current source. Since the lateral *pnp* transistor only carries the base current of the *npn* transistor, the current-handling capability of the composite device is primarily the same as that of the *npn* transistor. However, the frequency response of the composite device is determined by the lateral *pnp* transistor.

For purposes of biasing and network analysis, it is convenient to consider the composite device of Figure 2.28 as a high-gain *pnp* transistor. However, a word of caution is in order. The composite transistor is essentially a cascade of two devices. Therefore, it combines the inherent parasitics and the charge storage effects of both devices. Thus, its frequency response characteristics exhibit a multiple-pole response; and any feedback loop built around it may tend to instability.

A Summary of Device Characteristics. To some degree, the electrical characteristics of the lateral *pnp* transistor also depends on the particular fabrication process used. With reference to the three basic bipolar processes (i.e., the low-, medium-, and high-voltage processes) discussed in Section 2.1, the two key parameters which affect the lateral *pnp* transistor are the epitaxial resistivity and the collector–emitter spacing (i.e., the base width). For low-voltage devices requiring 20-V breakdown and 1Ω-cm epitaxial layer resistivity, the base width W_B can be as low as 6 μm. For devices with 30- or 40-V breakdown, W_B is required to be of the order of 8 μm. Since such devices also require higher epitaxial layer resistivity (i.e., a lighter doped *pnp* base) the β falloff with current comes about at lower current levels.

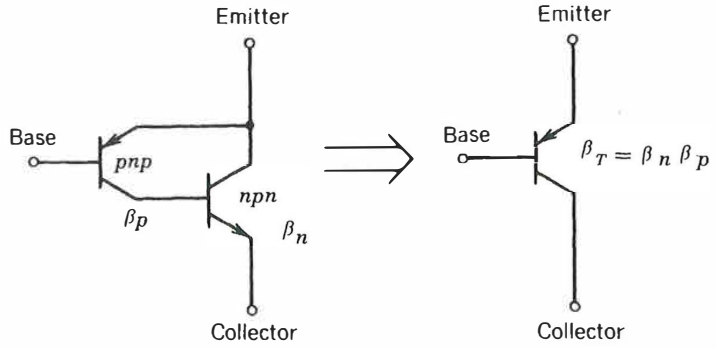


FIGURE 2.28. Composite connection of a lateral *pnp* transistor and a normal *npn* transistor.

EX-1014 0096
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Table 2.4 gives a summary of the electrical characteristics and parameters of a small-geometry lateral *pnp* transistor, similar to that shown in Figure 2.23, for each of the three basic fabrication processes discussed earlier. Note that in the table, I_O refers to the current level where β_F starts to fall off, and $I_{1/2}$ refers to the current level where β_F has fallen to 50% of its peak value.

The parasitic capacitance values associated with the lateral *pnp* transistor are not shown separately since they are the same as those associated with the collector–base and the collector–isolation junctions of the *npn* transistor given in Table 2.2.

Substrate *pnp* Transistor

A functional *pnp* transistor can also be obtained by using the base region of the *npn* transistor as the emitter, the *n*-type epitaxial layer as the base, and the *p*-type substrate as the collector. Such a device is known as the *substrate pnp* transistor. The substrate *pnp* transistor is inherently present as a parasitic, associated with the basic *npn* or the lateral *pnp* structures (see Fig. 2.3*a* or Fig. 2.26).

TABLE 2.4. Typical Device Parameters for Small-Geometry Lateral *pnp* Transistor of Figure 2.23[a]

Device Parameter	Low-Voltage Process: *npn* LV_{CEO}, 20 V, 1-Ω-cm Epitaxial Layer, $W_B = 6\ \mu m$	Medium-Voltage Process: *npn* LV_{CEO}, 30 V, 2.5-Ω-cm Epitaxial Layer, $W_B = 8\ \mu m$	High-Voltage Process: *npn* LV_{CEO}, 40 V, 5-Ω-cm Epitaxial Layer, $W_B = 10\ \mu m$
Forward-current gain β_F	20	30	40
Peak-gain current I_0	50 μA	30 μA	20 μA
Half-gain current $I_{1/2}$	500 μA	300 μA	200 μA
Early voltage V_A	60 V	50 V	50 V
Collector–base leakage I_{CO}	10^{-7} mA	10^{-7} mA	10^{-7} mA
Breakdown voltages $BV_{EBO} = BV_{CBO}$	50 V	70 V	90 V
Lowest sustaining voltage LV_{CEO}	35 V	50 V	60 V
f_T (at $I_C = 100\ \mu A$)	6 MHz	4 MHz	3 MHz
Base resistance r_b	150 Ω	220 Ω	300 Ω
Collector series resistance r_{cs}	150 Ω	150 Ω	150 Ω
Emitter series resistance r_{ex}	10 Ω	10 Ω	10 Ω

[a]Emitter periphery ≈ 100 μm.

EX-1014 0097
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

However now it is put to good use as a functional device. The substrate *pnp* transistor can be fabricated simultaneously with the *npn* bipolar transistors, without requiring additional masking or diffusion steps. However, since the epitaxial layer thickness is now directly related to the effective base width W_B of the *pnp* transistor, a somewhat tighter control of the epitaxial layer thickness is necessary to ensure repeatability of it chararacteristics.

Figure 2.29 shows a typical device layout and the cross section for a substrate *pnp* transistor. There are many variations possible to the basic device layout illustrated. Figure 2.30 shows the basic carrier flow path within the transistor for the minority carriers (holes) injected into the n-type epitaxial layer, which serves as the base region of the transistor. Note that the subepitaxial n^+-type layer is omitted between the epi–substrate interface in order to enhance the minority carrier transport through the base region. In the substrate *pnp* transistor, the isolation walls also serve as the collector, provided they are fairly close (i.e., within 20–30μm) to the p-type emitter area.

There is one key limitation to the substrate *pnp* transistor: the collector of the substrate *pnp* transistor is formed by the p-type substrate, which is common to the rest of the circuit, and is at all times ac grounded. Therefore, the substrate *pnp* transistor is only available in the grounded-collector configuration and cannot be used for level shifting or voltage amplification. However, it still provides current amplification and can be used as a low-impedance output device in class-B complementary stages.

Compared to the lateral *pnp* transistor, the substrate *pnp* transistor can handle a higher amount of current for a given device area. This is because in the case of the substrate *pnp* transistor, the entire emitter–base area, not just the emitter periphery, is active in injecting minority carriers into the base region (see Fig. 2.30). However, since the n-type epitaxial layer forming the base region is lightly doped, the current gain drops off rapidly at high currents. For a given emitter area, the current-handling capability of a substrate *pnp* transistor is approximately one-third to one-fifth of that of an equivalent *npn* transistor. Figure 2.25 shows the typical β_F versus current characteristics for the substrate *pnp* transistor of Figure 2.29, which has an effective emitter area of approximately 5000 μm^2.

A few comments on the basic layout of the substrate *pnp* transistor shown in Figure 2.29 are in order. First, note that the base contact is made by means of an n^+ diffusion, made into a "hole" within the p-type emitter region. This minimizes the base resistance r_b for the device. Although the presence of this n^+ diffusion, in direct contact with the p-type emitter, would cause a low reverse breakdown for the base–emitter junction of the device (typically $BV_{EBO} \approx 7$ V), it would not be a detriment in normal operation of the device since the substrate *pnp* transistor is only used in the grounded-collector (i.e., emitter follower) configuration. Second, an electrical contact to the substrate, or the isolation wall, is placed in the immediate vicinity of the device so that the collector current, which is flowing into the substrate common to the entire circuit, can be extracted with a minimum amount of voltage drop generated within the sub-

EX-1014 0098
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

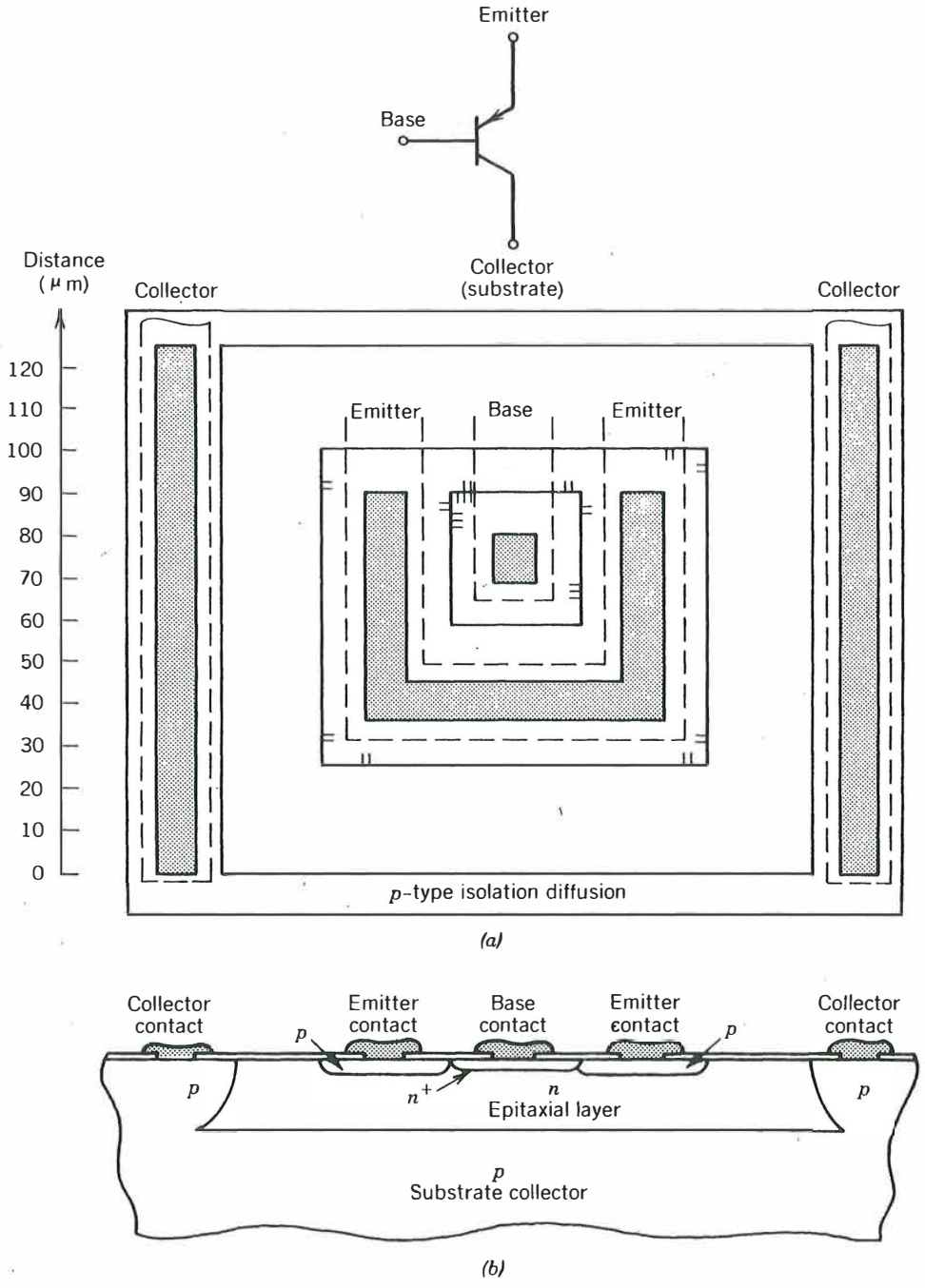


FIGURE 2.29. Typical device layout and cross section of a substrate *pnp* transistor.

EX-1014 0099
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

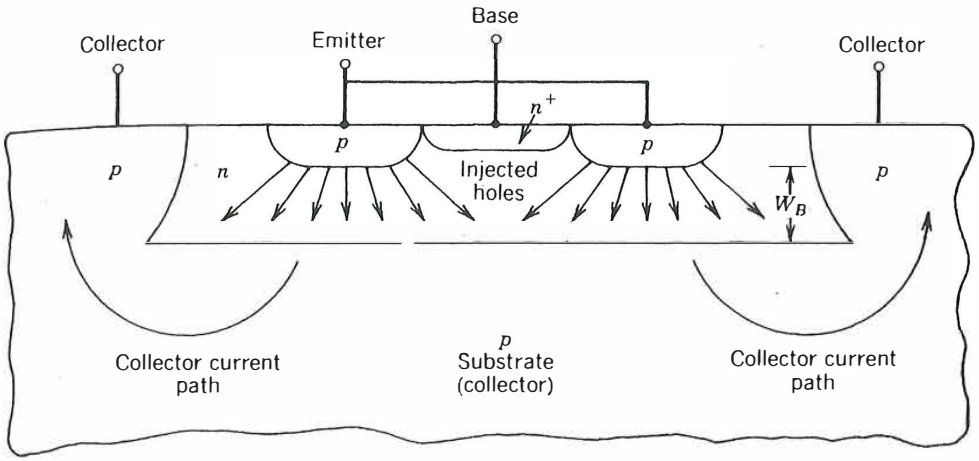


FIGURE 2.30. Direction of minority carrier flow, and the collector current path in a substrate *pnp* transistor.

strate. If this were not done, the voltage gradients generated within the substrate due to collector current of the substrate *pnp* transistor may cause some of the isolation junctions within the chip to be forward biased.

The high-frequency capability of the substrate *pnp* transistor is dominated by the carrier transit time through the wide base region W_B, which is determined by the epitaxial layer thickness. Typical values of the device cutoff frequency are in the range of 8–30 MHz.

Table 2.5 gives a summary of the substrate *pnp* device parameters for a device structure similar to that shown in Figure 2.29 for various basic bipolar manufacturing processes.

High-Performance *pnp* Transistors

In the design of high-performance analog circuits, particularly for operation within a radiation environment, the performance characteristics of the lateral and the substrate *pnp* transistors are not acceptable. A number of high-performance *pnp* transistor structures have been developed for these applications. However, each of these device structures requires additional processing steps above and beyond what is required for the basic *npn* transistor, and they are, therefore, limited to special design applications where the added fabrication cost or complexity can be justified.

The use of dielectric isolation techniques provides an added degree of freedom in fabricating high-performance complementary devices by providing access to the backside of the device structure. Therefore, with dielectric isolation methods, it is possible to fabricate high-performance *pnp* transistors simultaneously with *npn* bipolars, using a device structure as shown in Figure 2.31. The process steps for such a device structure follow the basic sequence of steps shown in Figure 1.14, except that prior to an n^+ deposition on the wafer, a

EX-1014 0100
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

TABLE 2.5. Typical Device Parameters for the Substrate *pnp* Transistor of Figure 2.29

Device Parameter	Low-Voltage Process: 1 Ω-cm, 9-μm Epitaxial Layer	Medium-Voltage Process: 2.5 Ω-cm, 14-μm Epitaxial Layer	High-Voltage Process: 5 Ω-cm, 18-μm Epitaxial Layer
Forward-current gain β_F	30	40	50
Peak-gain current I_0	5 mA	3 mA	2 mA
Half-gain current $I_{1/2}$	30 mA	20 mA	10 mA
Early voltage V_A	40 V	50 V	50 V
Collector–base leakage I_{CO}	4×10^{-7} mA	4×10^{-7} mA	4×10^{-7} mA
Base–emitter breakdown BV_{EBO}	6.5 V	6.8 V	7.0 V
Collector–base breakdown BV_{CBO}	50 V	70 V	90 V
Lowest sustaining voltage LV_{CEO}	35 V	50 V	70 V
f_T (at I_C = 1 mA)	30 MHz	20 MHz	10 MHz
Base resistance r_b	50 Ω	100 Ω	150 Ω
Emitter series resistance r_{ex}	5 Ω	5 Ω	5 Ω
Collector series resistance r_{cs}	50 Ω	50 Ω	50 Ω

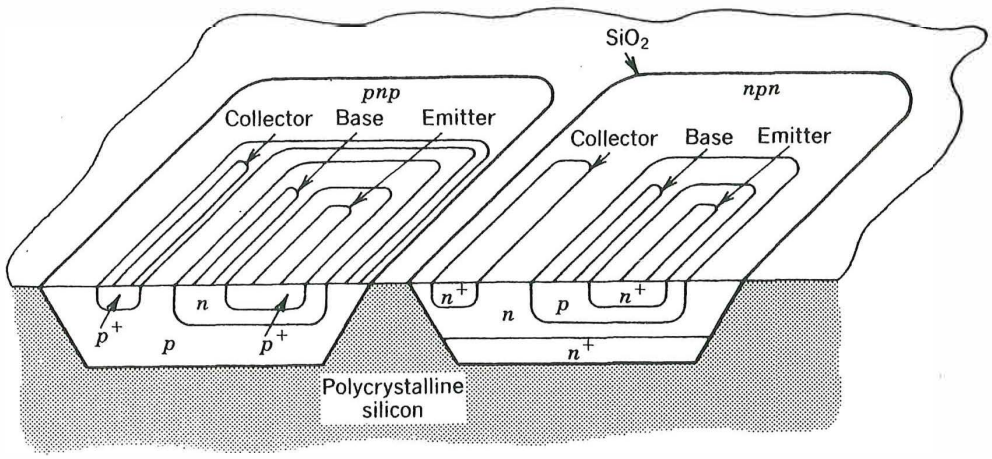


FIGURE 2.31. Dielectrically isolated vertical *npn* and *pnp* transistors.

EX-1014 0101
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

selective p diffusion is made from the backside. This is a deep ($\approx$20 μm) noncritical diffusion step, forming the p-type collector region for the *pnp* transistor. This step is then followed by the selective n^+ predeposition for the *npn* collectors. After moat etching, polycrystalline silicon growth, and backlapping operations, one ends up with dielectrically isolated p- and n-type pockets on the silicon surface. The p-type pockets do not have a buried layer similar to the n^+-type layer below the *npn* collector. However, due to the backside diffusion step forming the p-type islands, the *pnp* collector region has a reverse impurity gradient, being more heavily doped near the bottom of the pocket. This reverse impurity profile results in a low collector series resistance for the *pnp* transistor and eliminates the need for a separate p^+-type buried layer. Once the dielectrically isolated p- and n-type islands are formed, the device structure of Figure 2.31 is completed by the following sequence of diffusion steps: (1) *pnp* base, (2) *npn* base, (3) *pnp* emitter, and (4) *npn* emitter.

In this process, the *pnp* diffusion steps are interleaved with the standard *npn* diffusion cycles without changing the *npn* impurity profile. Therefore, the electrical characteristics of the resulting *npn* devices remain unchanged. However, the *pnp* devices show a significant improvement over the lateral and the substrate *pnp* transistors, with the following typical performance characteristics:

$$\beta_F \approx 50\text{–}100 \quad (\text{at } I_C = 1\ mA)$$
$$LV_{CEO} \approx 60\ \text{V}$$
$$BV_{EBO} \approx 9\ \text{V}$$
$$BV_{CBO} \approx 80\ \text{V}$$
$$f_T = 150\ \text{MHz}$$

Typical absolute-value tolerances, matching tolerances, and thermal tracking properties of the high-performance *pnp* transistors are comparable to those listed in Table 2.1 for the *npn* devices.

Again, the reader is reminded that the economics and the yield of monolithic IC fabrication are extremely sensitive to additional manufacturing steps. Thus, complex device structures such as those shown in Figure 2.31, which require several additional masking and diffusion steps, are difficult to manufacture, economically, under a high-volume production environment (see Table 1.2 for relative cost factors). They should be used with discretion and limited to those special applications where the added fabrication costs can be justified.

2.4. JUNCTION FIELD-EFFECT TRANSISTORS

The principle of operation of the junction field-effect transistor (JFET) differs significantly from that of the bipolar type. In the case of JFETs, the current conduction mechanism through the device relies solely on the majority carriers, whereas in the bipolar transistor *both* the majority and the minority carriers participate actively in the current transport process. A detailed coverage of JFET

EX-1014 0102
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

device theory is readily available in the literature[7,8] and will not be repeated here. Instead, this section is aimed at covering the salient features of JFET structures which are readily compatible with the basic *npn* bipolar process technology.

The JFET is a voltage-controlled device, where the current conduction through a channel connecting the *source* and the *drain* regions is controlled or modulated by means of a control voltage applied to the *gate* terminal. The JFET can be either a p-channel or an n-channel device, depending on the conductivity type of the source and drain regions and the channel that connects the two. The gate regions of opposite conductivity type, and a reverse-biased p-n junction exists at all times between the channel and the gate regions.

Figure 2.32 shows the basic symbol and the voltage polarities associated with the p- and n-channel JFET. In monolithic IC design, p-channel JFET are more widely used than n-channel devices because their structure is more readily compatible with the basic *npn* bipolar fabrication process.

Device Characteristics

Figure 2.33 shows an idealized p-channel JFET structure which is helpful in deriving the basic device characteristics and understanding its principle of operation. In the figure, a uniformly doped p-type channel is assumed, with symmetrical, heavily doped n-type gate regions on either side. With zero applied bias to device terminals, a finite depletion region exists across the p-n junction surrounding the channel, due to the built-in junction potential ψ_0 [see Eq. (2.31)]. Assume, for the time being, that the gate region is short-circuited to the source. If a negative voltage V_{DS} were applied to the drain, a drain current I_D would flow from the source to the drain through the channel. This current will produce a voltage gradient within the channel and cause the channel–gate junction to be more heavily reverse biased at the drain side of the channel. This

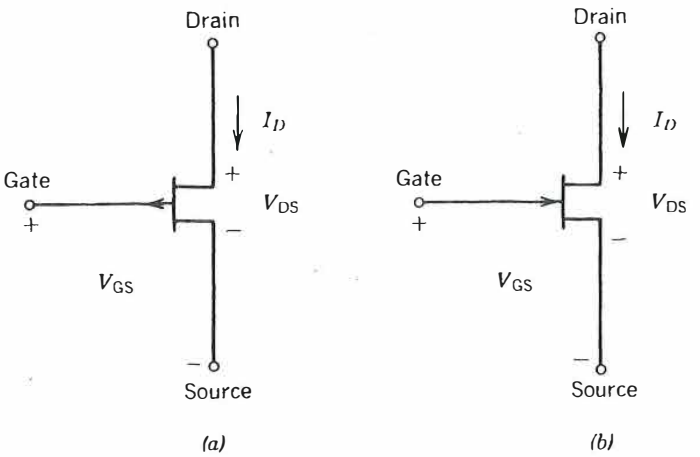


FIGURE 2.32. Circuit symbols and sign conventions for p- and n-channel JFET: (*a*) p-channel; (*b*) n-channel.

EX-1014 0103
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

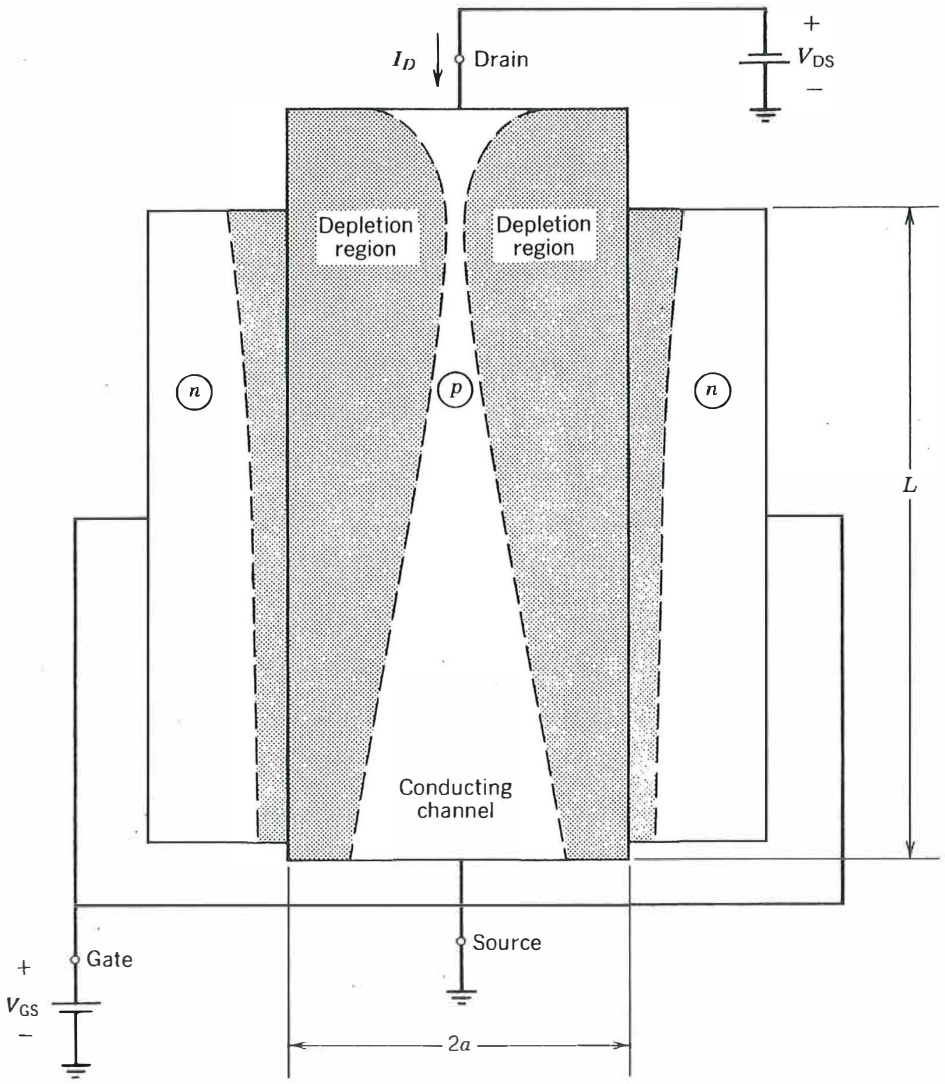


FIGURE 2.33. Idealized *p*-channel JFET model for deriving basic device characteristics.

reverse bias, in turn, causes the depletion layer to stretch into the channel and constrict the conductive region of the channel further, with the narrowest region appearing at the drain end of the channel.

If the drain–source voltage V_{DS} is increased further, the channel width at the drain end approaches zero. The voltage level at which this happens, that is, when the gate–channel depletion layer reaches completly across the entire width of the channel, is called the pinch-off voltage V_P. If V_{DS} is increased further, beyond V_P, the drain current I_D becomes insensitive to V_{DS} and reaches a saturation value I_{DSS}. Device operation in this region, where the drain current is relatively constant, is called *pinched* operation or operation in the *pinched-off* region.

If an external bias V_{GS} is applied to reverse-bias the gate region with respect to the source, as shown in Figure 2.33, this would create an additional initial

EX-1014 0104
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

depletion layer within the source end of the channel. This additional depletion layer would then cause the drain current I_{DS} to saturate sooner and remain constant at a lower value than the zero gate bias saturation current I_{DSS}. Finally, if the applied gate–source bias V_{GS} is equal to or greater than the pinch-off voltage V_P, the entire channel region becomes depleted of mobile carriers, the drain current becomes zero, and the source becomes electrically disconnected from the drain.

The pinch-off voltage is an important parameter of device design and can be given as

$$V_P = \frac{a^2 q N_A (1 + N_A/N_D)}{2\epsilon} - \psi_o \tag{2.33}$$

where a is the half-width of the channel region transverse to the direction of current flow, as shown in Figure 2.33, ϵ is the dielectric coefficient of silicon, and N_A and N_D are the doping levels of the p-type channel and the n-type gate regions, respectively. Note that V_P is a very strong function of the channel width and the channel impurity concentration.

Using the idealized device model of Figure 2.33, one can show that I_D can be related to the gate and drain voltages V_{GS} and V_{DS} as[9]

$$I_D = G_o\left[V_{DS} + \frac{3}{2}\frac{(\psi_o + V_{GS} - V_{DS})^{3/2} - (\psi_o + V_{GS})^{3/2}}{(\psi_o + V_P)^{1/2}}\right] \tag{2.34}$$

as long as the device is not operated in its pinched mode, that is, as long as the total reverse bias across the junction at the drain end of the channel does not exceed V_P.

In Eq. (2.34), the constant G_o is determined by the geometry and the conduc-

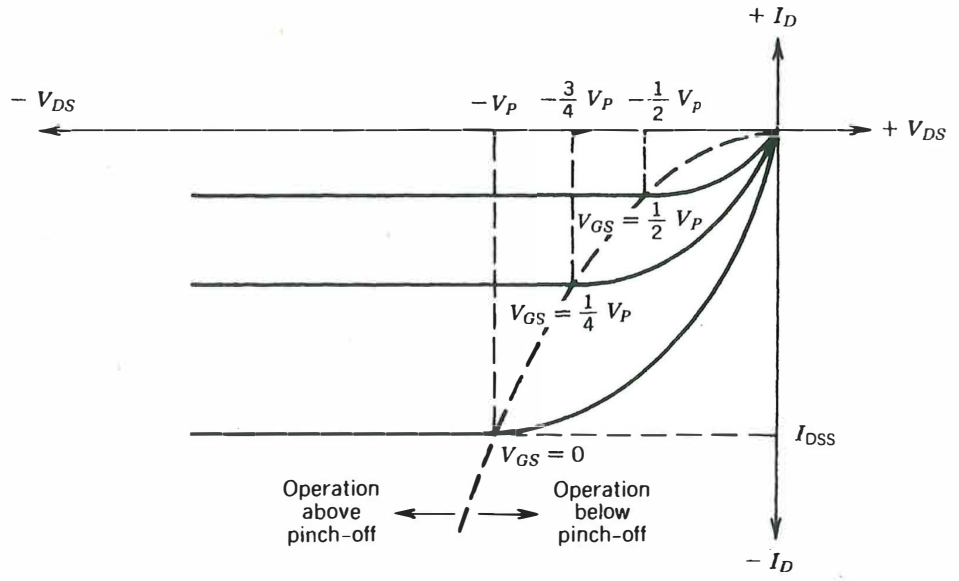


FIGURE 2.34. Idealized current–voltage characteristics for a p-channel JFET.

EX-1014 0105
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

tivity of the channel as

$$G_o = \frac{2az}{L}\sigma_c \tag{2.35}$$

where L is the length of the active channel region in the direction of current flow, Z is the depth of the channel (i.e., the dimension going *into* the paper in Fig. 2.33), and σ_c is the conductivity of the channel.

Figure 2.34 shows the idealized p-channel JFET characteristics implied by Eq. (2.34). This equation predicts the current–voltage characteristics to the right of the dashed line in Figure 2.34, corresponding to operation below pinch-off, and assumes that I_D is constant and independent of V_{DS} for higher voltages.

The maximum value of drain current I_{DSS}, which corresponds to $V_{GS} = 0$, is given as

$$I_{DSS} = G_o\left[-V_P + \frac{2}{3}\frac{(\psi_o + V_P)^{3/2} - \psi_o^{3/2}}{(\psi_o + V_P)^{1/2}}\right] \tag{2.36}$$

The saturation value I_{DS} of the drain current in the pinched-off region of the device (i.e., the flat portion of the current–voltage characteristics) is related to I_{DSS} as

$$I_{DS} = I_{DSS}\left(1 - \frac{V_{GS}}{V_P}\right)^2 \tag{2.37}$$

Actual JFET characteristics show a finite slope to the current–voltage characteristics for pinched operation of the device, as illustrated in Figure 2.35. This finite slope comes about from the modulation of the effective channel length L by the drain-to-gate depletion layer, and is analogous to the Early effect in

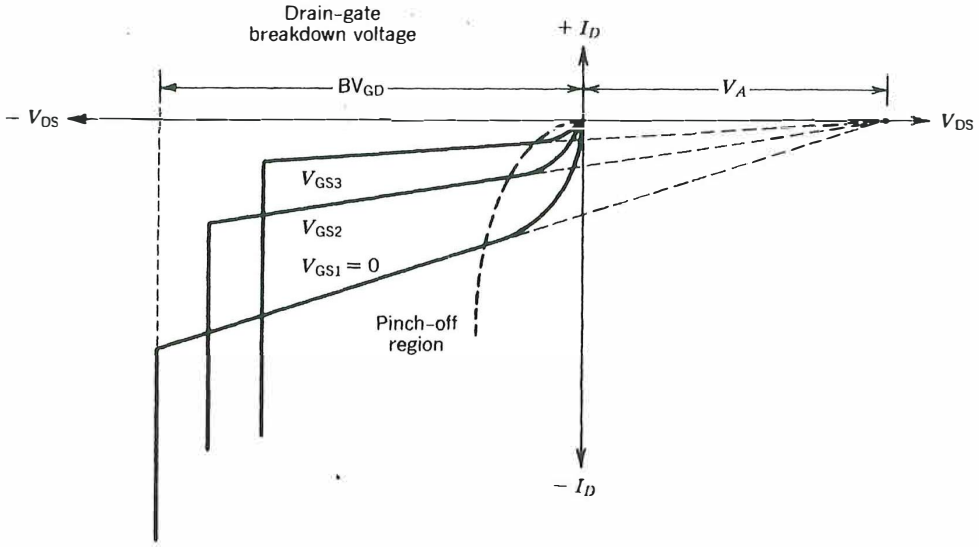


FIGURE 2.35. Actual p-channel JFET current–voltage characteristics showing effects of channel length modulation and gate–channel junction breakdown.

EX-1014 0106
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

bipolar transistors. Thus, one can similarly extrapolate the device characteristics to an intercept voltage V_A on the V_{DS} axis, as shown in the figure. This allows the idealized current characteristics of Eq. (2.37) to be modeled more accurately as

$$I_{DS} = I_{DSS}\left(1 - \frac{V_{GS}}{V_P}\right)^2\left(1 - \frac{V_{DS}}{V_A}\right) \tag{2.38}$$

For practical JFET structures, V_A is in the range of 50–100 V and increases as the channel length-to-width ratio $L/2a$ is increased.

As shown in Figure 2.35, the maximum value of V_{DS} is limited by the avalanche breakdown of the gate–channel junction. This breakdown occurs at the drain end of the channel, where the total reverse bias across the junction, which is equal to $V_{GS} + V_{DS}$, reaches the junction avalanche breakdown voltage, BV_{GD}.

Integrated JFET Structures

The key requirement for integrated JFET devices is that they should be process compatible with the basic *npn* bipolar technology, such that they can be manufactured with only a minimum number of additional process steps. An additional requirement is that the resulting devices should have a repeatable and predictable pinch-off voltage in the range of 1–5 V so that they can operate with signal and bias levels common to bipolar transistors. These requirements are best fulfilled by *p*-channel devices, made by either the diffusion or the ion-implantation techniques.

Double-Diffused JFET. Figure 2.36 shows the typical cross section of a *p*-channel JFET structure which can be fabricated simultaneously with an *npn*

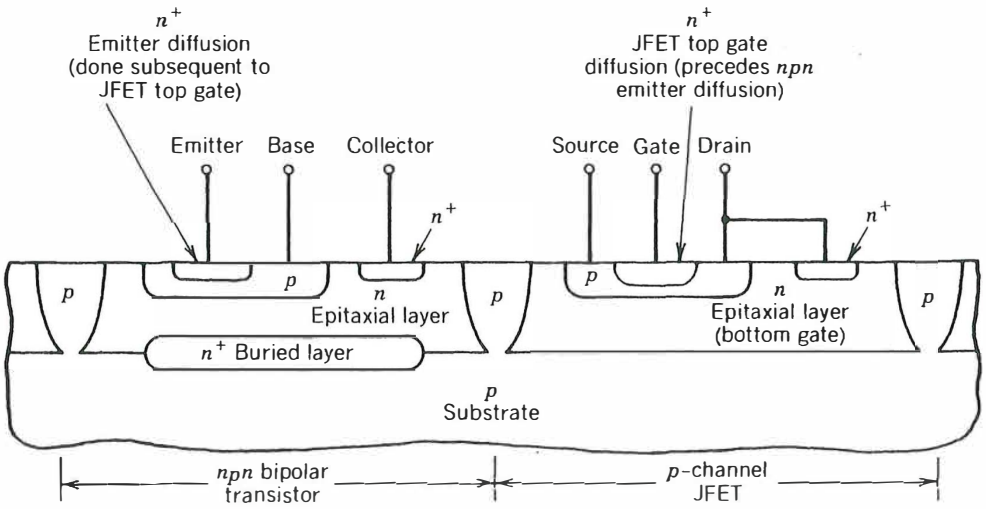


FIGURE 2.36. Bipolar compatible double-diffused *p*-channel JFET structure.

EX-1014 0107
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

bipolar transistor. This device uses the p-type base region of the *npn* transistor as the channel, where the n^+ emitter diffusion and the n-type epitaxial region serve as the top and bottom gates. However, if only the normal n^+ emitter diffusion is used to form the top gate, the resulting pinch-off voltage is too high (i.e., in the 8–15-V range). This problem can be avoided by using a separate n^+ gate diffusion step for the JFET which is diffused slightly deeper than the normal n^+ emitter diffusion, thus resulting in a narrower p-type channel. This process is very similar to that described in Section 1.10 in connection with the superbeta transistors (see Fig. 1.17) and results in a V_P in the range of 2–5 V.

Although double-diffused JFET devices have been used in a number of applications such as operational amplifier input stages, they have several drawbacks. The most important limitation is the difficulty of controlling the channel width by means of the critical n^+ diffusion step. This results in poor control of the absolute value of V_P as well as the matching and tracking of V_P between different devices on the same chip. A second limitation is the low breakdown characteristic: the gate–drain breakdown is essentially limited to that of the base–emitter junction of a bipolar transistor, that is, approximately 7 V. Still a third disadvantage is that since the p-type channel is more heavily doped than the n-type epitaxial region, the bottom gate does not contribute significantly to the device transconductance.

Ion-Implanted JFET. The drawbacks and limitations of double-diffused JFETs can be avoided by using an ion-implanted device structure, as shown in Figure 2.37. This device, which is also called a *BIFET*, has gained wide acceptance with the advent of ion-implantation technology, it has replaced the double-diffused JFET in most of the circuit applications.

The ion-implanted JFET structure uses the p-type base diffusion of the *npn* transistor to form the source and drain contact regions. The channel is then formed by a p-type implant step. This is followed by an n-type implant step to form the gate region. The channel implant is the critical step in the device fabrication. Since the ion-implantation process is capable of very uniform and precise placement of impurities into silicon, it allows the channel resistivity and thickness to be accurately controlled. This in turn allows significant improvement in the control of the absolute values and the matching of the pinch-off voltage and the saturation current across the wafer. The n-type gate implant is a heavy-dose shallow implant. Since it does not extend significantly into the channel, it does not affect the device characteristics significantly.

Figure 2.38 shows the typical impurity profile for the implanted channel and gate regions. Note that the total channel width is $\approx 0.5\ \mu$m. The impurity concentration in the channel is kept low (typically in the range of 2×10^{16} atoms/cm^3) in order to maintain a high gate–channel breakdown voltage (typically $\geq$40 V). Typical values of pinch-off voltage in ion-implanted JFETs are in the range of 1–3 V, with a typical matching of $\approx \pm 10$ mV between adjacent devices on the chip.

EX-1014 0108
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

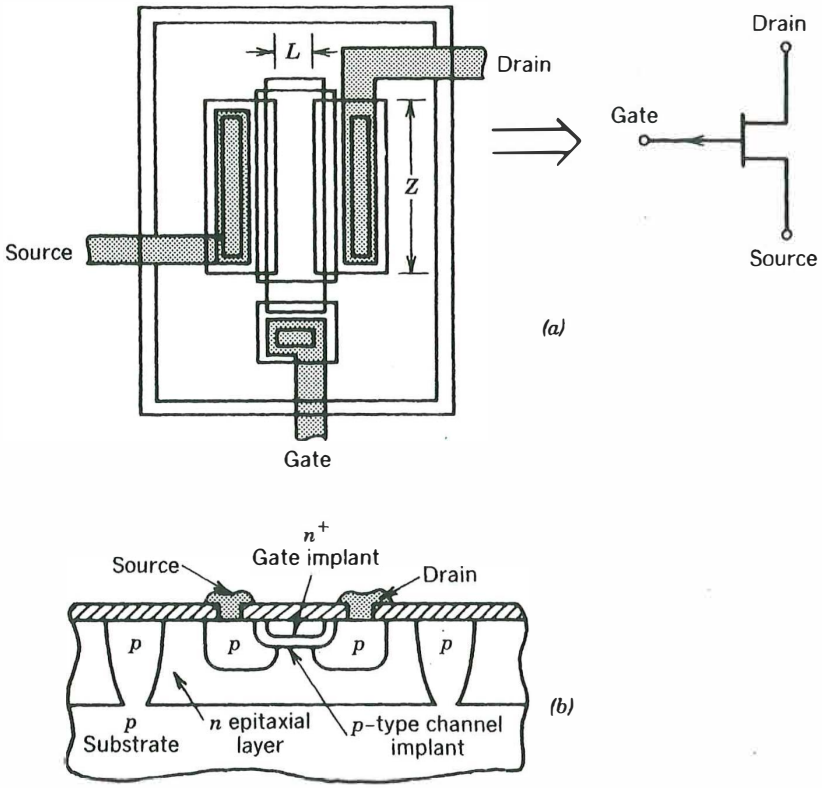


FIGURE 2.37. Top view and cross section of an ion-implanted p-channel JFET.

Small-Signal Model

In most analog circuit applications, the JFET is operated in its pinch-off region, where its current–voltage characteristics approximate that of a voltage-controlled current source. Figure 2.39 shows that the basic small-signal equivalent circuit for a JFET for operation in its pinch-off region. The parasitic capacitances associated with the device structure are also shown.

The small-signal transconductance g_m can be derived from Eq. (2.37) as

$$g_m = \frac{\partial I_{DS}}{\partial V_{GS}} = -\frac{2I_{DSS}}{V_P}\left(1 - \frac{V_{GS}}{V_P}\right) \tag{2.39}$$

which can be rewritten as

$$g_m = g_{mo}\left(1 - \frac{V_{GS}}{V_P}\right) \tag{2.40}$$

where g_{mo} is the zero bias transconductance corresponding to $V_{GS} = 0$, given as

$$g_{mo} = -\frac{2I_{DSS}}{V_P} \tag{2.41}$$

EX-1014 0109
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

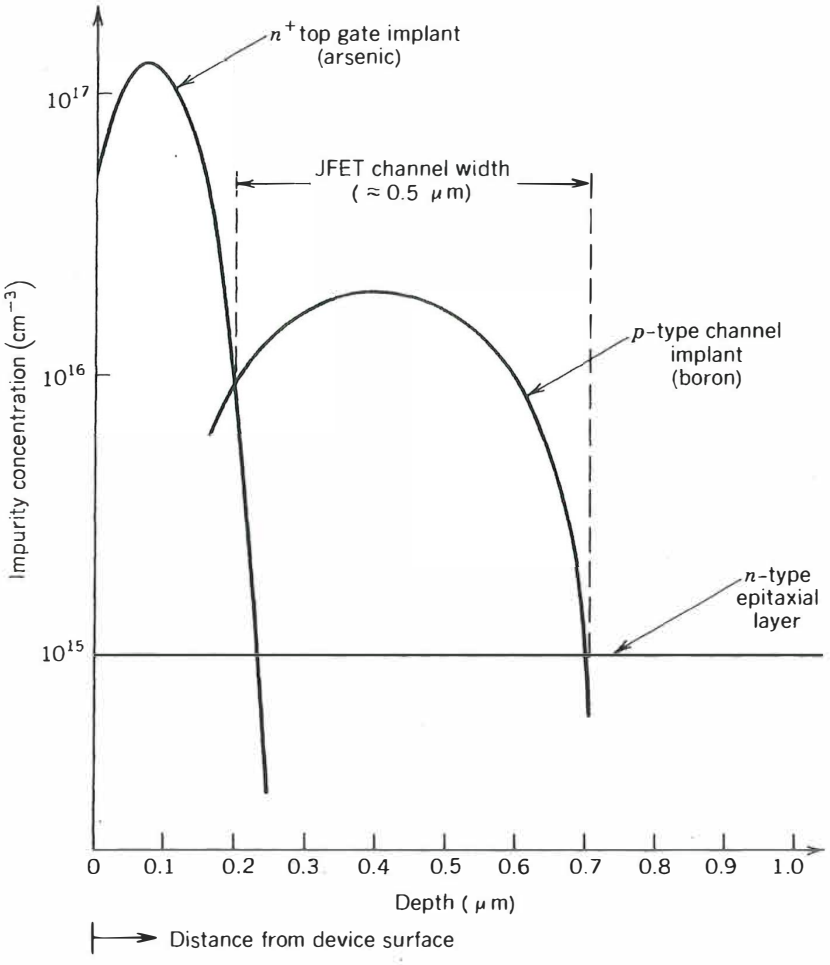


FIGURE 2.38. Typical impurity profile of an ion-implanted *p*-channel JFET.

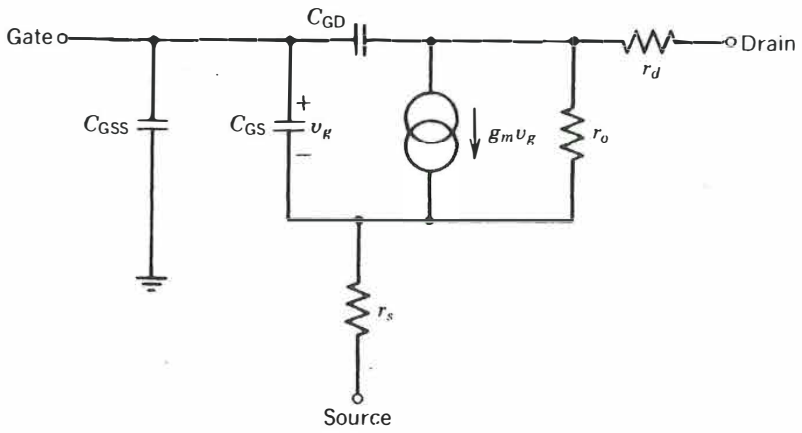


FIGURE 2.39. Small-signal equivalent circuit for a JFET.

EX-1014 0110
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

It should be noted that for a p-channel device I_{DSS} is negative and V_P is positive, whereas the opposite is true for an n-channel JFET. Thus, g_{mo} is always a positive number. For a pinch-off voltage $V_P \approx 1.0$ V and $I_{DSS} = 250\ \mu$A, g_{mo} is around 0.5 mA/V.

The small-signal output resistance r_o is due to the channel length modulation effect. It can be derived by differentiating Eq. (2.38),

$$\frac{1}{r_o} = \frac{\partial I_D}{\partial V_{DS}} = \frac{I_{DSS}}{V_A}\left(1 - \frac{V_{GS}}{V_P}\right)^2 \tag{2.42}$$

where V_A is the extrapolated voltage intercept shown in Figure 2.35. Typical values of V_A are in the range of 50–100 V. Assuming that $V_{DS}/V_A \ll 1$, Eq. (2.42) simplifies to

$$r_o \approx \frac{V_A}{I_D} \tag{2.42a}$$

The parasitic drain and source resistances r_d and r_s, shown in the equivalent circuit of Figure 2.39, are due to the parasitic bulk resistance between the respective drain and source contacts and the active channel region directly below the gate. Typical values of these resistances are of the order of 30–100 Ω, depending on the device geometry and the impurity profile. In most applications, their presence can be neglected.

Compared to the bipolar transistor, one key advantage of the JFET is the high input impedance of the gate terminal. Since the gate is in series with a reverse-biased p-n junction, it is a virtual open circuit at dc. This makes the JFET virtually an ideal device for circuits, such as operational amplifiers which require high input impedance. However, compared to bipolar devices, the JFETs have one serious limitation: in bipolar transistors, the transconductance g_m is determined by the bias current [see Eq. (2.16)] and is approximately 40 millimhos per milliampere of bias current. In JFETs g_m is determined by the channel geometry (i.e., the channel Z/L ratio, as shown in Fig. 2.37) and the channel width and resistivity, and is of the order of 0.5–1 millimho at $I_{DS} \approx 1$ mA. Since the overall voltage gain which can be obtained from the active device is proportional to g_m, a bipolar stage can provide much higher voltage gain than a corresponding JFET stage at similar bias levels.

Parasitic Capacitances. The parasitic capacitances shown in Figure 2.39 are inherent to the monolithic JFET structures. The gate–substrate capacitance C_{GSS} is the capacitance of the reverse-biased epi–substrate junction of Figures 2.36 and 2.37. This is essentially the same as the collector–substrate capacitance C_{DS} of a bipolar transistor. The capacitances, C_{GS} and C_{GD} are the junction capacitances associated with the gate–source and the gate–drain junctions.

In the bipolar compatible p-channel JFET structures of Figures 2.36 and 2.37, the gate–source capacitance C_{GS} is made up of two sections

$$C_{GS} = C_{GS1} + C_{GS2} \tag{2.43}$$

EX-1014 0111
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

where C_{GS1} is the capacitance due to the heavily doped top gate and source junction, and C_{GS2} is the capacitance due to the lightly doped back gate (i.e., the n-type epitaxial layer) and source junction. The capacitance of the gate–channel junction is normally associated with the source terminal.

The drain–gate capacitance C_{GD} is primarily due to the p-type drain and the n-type back gate (i.e., the epitaxial layer) junction. The values of these capacitances depend on the device geometries (i.e., the junction areas) and the applied reverse bias across the junction. Similar to the case of bipolar transistors, they can be evaluated using Eqs. (2.30) and (2.31). The zero-bias junction capacitances, the built-in junction potentials, and the exponent n associated with each junction are given in Table 2.6 for typical double-diffused and ion-implanted JFET structures of Figures 2.36 and 2.37, assuming a 5-Ω/cm n-type epitaxial layer. In both cases, a stripe-type device layout similar to that shown in Figure 2.37*a* is assumed. Assuming a Z/L of 20, a gate substrate bias of 10 V, and V_{DG} of 5 V, typical values of capacitances calculated from the data of Table 2.6 are

$$C_{GSS} = 2 \text{ pF} \qquad C_{GS1} = 1 \text{ pF}$$
$$C_{GD} = 0.5 \text{ pF} \qquad C_{GS2} = 0.6 \text{ pF}$$

It should be noted that although C_{GD} has the lowest value, it is the most detrimental parasitic to the voltage gain of the device at high frequencies, due to the *Miller effect*, similar to the collector–base capacitance C_c of a bipolar transistor.

TABLE 2.6. Typical Values of Parasitic Capacitances Associated with Double-Diffused and Ion-Implanted JFETS

Parasitic Capacitances	Double-Diffused JFET	Ion-Implanted JFET
Gate–substrate capacitance C_{GSS}		
Zero-bias capacitance	10^{-4} pF/μm^2	10^{-4} pF/μm^2
Built-in potential ψ_o	0.52 V	0.52 V
Exponent n	$\frac{1}{2}$	$\frac{1}{2}$
Gate–source capacitance C_{GS}		
1. Top gate capacitance C_{GS1}		
Zero-bias capacitance	8×10^{-4} pF/μm^2	5×10^{-4} pF/μm^2
Built-in potential ψ_o	0.7 V	0.5 V
Exponent n	$\frac{1}{3}$	$\frac{1}{3}$
2. Bottom gate capacitance C_{GS2}		
Zero-bias capacitance	10^{-4} pF/μm^2	10^{-4} pF/μm^2
Built-in potential ψ_o	0.55 V	0.55 V
Exponent n	$\frac{1}{2}$	$\frac{1}{2}$
Gate–drain capacitance C_{GD}		
Zero-bias capacitance	10^{-4} pF/μm^2	10^{-4} pF/μm^2
Built-in potential ψ_o	0.55 V	0.55 V
Exponent n	$\frac{1}{2}$	$\frac{1}{2}$

EX-1014 0112
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Summary of Device Characteristics. Table 2.7 gives a comparison of the electrical characteristics of the double-diffused and the ion-implanted p-channel JFET structures. For comparison purposes, both devices are assumed to have the stripe geometry of Figure 2.37*a* with a Z/L ratio of 20, where the channel length $L = 12\ \mu$m and the width of the device $Z = 240\ \mu$m. The important point to note from the table is that the ion-implanted JFET structure gives a much tighter control of the key device parameters, such as I_{DSS} and V_P and provides a much higher gate–drain breakdown voltage.

2.5. MOS FIELD-EFFECT TRANSISTORS

In the metal–oxide semiconductor field-effect transistor (MOSFET), a thin dielectric barrier is used to isolate the gate and the channel. The control voltage applied to the gate terminal induces an electric field across the dielectric barrier and modulates the free-carrier concentration in the channel region. In the literature, these devices are also often referred to as insulated-gate field-effect transistors (IGFET) or simply as MOS transistors.

MOS transistors are classified as p-channel or n-channel devices, depending on the conductivity type of the channel region. In addition, these devices can also be classified according to their mode of operation as *enhancement* or *depletion* type devices.

In a depletion-mode MOSFET, a conducting channel exists under the gate with no applied gate voltage. The applied gate voltage controls the current flow

TABLE 2.7. Typical Parameter Values and Tolerances for Integrated JFET Devices[a]

Device Characteristic	Double-Diffused JFET	Ion-Implanted JFET
Saturation current I_{DSS}		
Typical value	$-400\ \mu$A	$-250\ \mu$A
Absolute-value tolerance[b]	-200 to $-800\ \mu$A	-150 to $-400\ \mu$A
Matching tolerance[c]	$\pm 10\%$	$\pm 5\%$
Pinch-off voltage V_P		
Typical value	2 V	1 V
Absolute-value tolerance	± 500 mV	± 150 mV
Matching tolerance	$\pm$ 15 mV	$\pm$ 8 mV
Zero-bias transconductance g_{mo}	800 μmho	400 μmho
Gate–drain breakdown BV_{DGO}	7 V	50 V
Early voltage V_A	100 V	100 V

[a]For comparison purposes, identical device geometries are assumed with $L = 12\ \mu$m, $Z = 240\ \mu$m, $Z/L = 20$.
[b]Sigma limit for absolute-value distribution under production environment.
[c]Sigma limit for matching characteristics of identical device geometries, adjacent to each other.

EX-1014 0113
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

between the source and the drain by depleting a part of this channel. This is very similar to the operation of the JFET described previously.

In the case of the enhancement-mode MOS transistor, no conductive channel exists between the source and the drain at zero applied gate voltage. As a gate bias of proper polarity is applied and increased beyond a threshold value V_{TH}, a localized inversion layer is formed directly below the gate. This serves as a conducting channel between the source and the drain electrodes. If the gate bias is increased further, the resistivity of the induced channel is reduced, and the current conduction from the source to the drain is enhanced. Figure 2.40 shows a cross-section diagram and the circuit symbol of an n-channel enhancement-mode MOSFET. The particular polarities of the gate and drain bias for the proper operation of the device are also identified. Normally, an SiO_2 layer is used as the gate dielectric. The thickness of this oxide layer is normally much less than that of the oxide layers commonly used for masking or surface passivation.

The enhancement-mode MOSFET is preferred over its depletion-mode counterpart because it is a *self-isolating* device, does not require tight control of

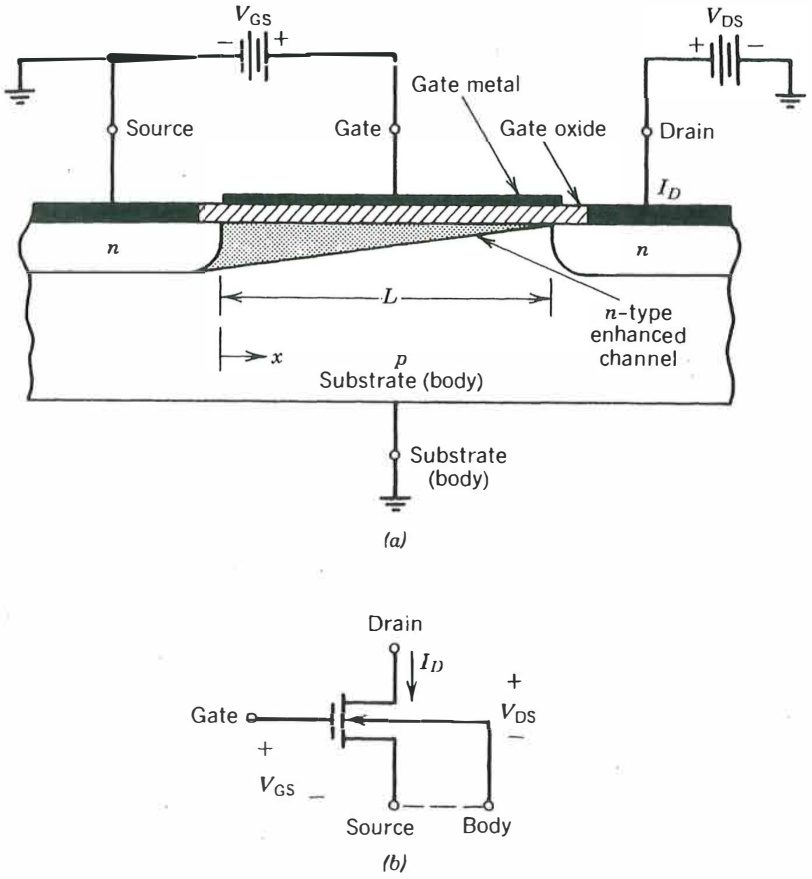


FIGURE 2.40. Cross section and device symbol for a n-channel enhancement-mode MOSFET.

EX-1014 0114
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

diffusion cycles, and can be fabricated by a single diffusion step forming the source and the drain pockets. Since all the active regions of the MOSFET are reverse biased with respect to the substrate, adjacent devices fabricated on the same substrate are electrically isolated without requiring a separate isolation diffusion. Because of this self-isolation advantage, MOSFET devices offer a much higher packing density per unit area of silicon surface than the bipolar transistors. The depletion-mode MOSFET transistors are often used in conjunction with enhancement-mode MOS transistors to serve as active loads, to improve the gain or the switching speed of the enhancement devices.

The bulk of the semiconductor region, shown as substrate in Figure 2.40, is normally inactive since the current flow is confined to a thin surface channel directly below the gate. This part of the MOS transistor is called the *body* and is normally tied to the same potential as the source. However, in certain circuit applications, particularly in designing analog circuits with MOS transistors, the body may not be connected to the source. This may have a significant effect in the device characteristics, as will be described later.

Device Characteristics

In the device structure of Figure 2.40, with the body short-circuited to the source and with both $V_{GS} = 0$ and $V_{DS} = 0$, there will be no conducting channel between the n-type *beds* which form the source and drain regions. If the gate voltage V_{GS} is raised gradually, the electric field within the gate dielectric layer will increase correspondingly and cause an accumulation of n-type free carriers (electrons) in the region of the device directly below the gate dielectric. If the gate voltage is raised beyond a certain threshold voltage, the free-carrier concentration below the gate will be sufficient to form a conductive n-type channel, which will then connect the n-type source and drain region. In other words, if V_{GS} is raised beyond a threshold voltage V_{TH}, the p-type region directly below the gate would invert its conductivity type and form an n-type channel. It can be shown that this uniform conducting channel has a sheet conductance g_c given as[10]

$$g_c = \mu_s C_{ox}(V_{GS} - V_{TH}) \tag{2.44}$$

where μ_s = surface mobility of the majority carriers in the induced channel
C_{ox} = capacitance per unit area of the gate electrode

The value of μ_s is in the range of 150–250 cm^2/V-sec for holes and 300–600 cm^2/V-sec for electrons, depending on the crystal orientation of slicon and the impurity profile at, or near, the device surface.

The gate capacitance C_{ox} is related to the thickness T_{ox} and the dielectric coefficient ϵ_{ox}, of the gate dielectric, and can be expressed as

$$C_{ox} = \frac{\epsilon_{ox}}{T_{ox}} \tag{2.45}$$

EX-1014 0115
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

T_{ox} is normally chosen to be $\geq$ 800 Å for breakdown purposes. For practical MOS structures, the design value of T_{ox} is in the range of 1000–2000 Å.

If the drain voltage V_{DS} is increased in the polarity shown in Figure 2.40, a finite drain current I_D flows through the induced channel. This current also causes an ohmic drop along the channel which subtracts from the net gate voltage $V_{GS}-V_{TH}$. Since at any point along the channel, the apparent sheet conductance is proportional to this net gate voltage, the voltage gradient due to I_D causes the channel to deplete along its length, and the sheet conductance becomes a function of the distance x along the channel,

$$g_c(x) = \mu_s C_{ox}\Big(V_{GS} - V(x) - V_{TH}\Big) \tag{2.46}$$

where $V(x)$ is the channel potential at a distance x from the source. Thus, an increase of V_D causes the drain current to increase, which in turn causes the channel to deplete or pinch off near the drain. For values of the drain voltage in excess of the net gate bias, that is, for

$$V_{DS} \geq V_{GS} - V_{TH} \tag{2.47}$$

a space-charge layer is formed at the drain end of the channel, since the net gate bias (i.e., the applied gate voltage minus the voltage drop along the channel) at this point is no longer sufficient to maintain an induced channel. This leads to saturation of the drain current, similar to the case of JFETs, and results in a set of idealized drain current versus voltage characteristics shown in Figure 2.41. The voltage values shown on the curve are for illustrative purposes only.

For values of V_{DS} less than $V_{GS} - V_{TH}$, the drain current I_D can be related to V_{DS} as

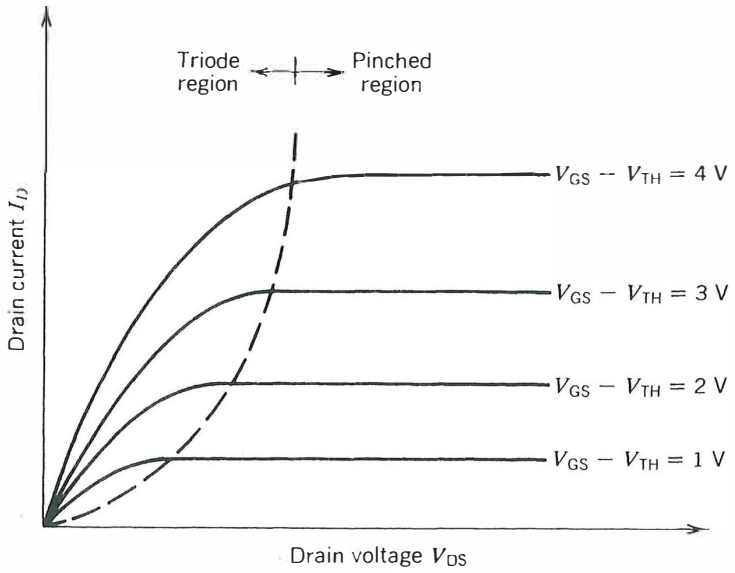


FIGURE 2.41. Idealized current–voltage characteristics of a *n*-channel enhancement-mode MOSFET.

EX-1014 0116
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

$$I_D = \frac{\mu_s C_{ox} Z}{L}\left(V_{GS} - V_{TH} - \frac{V_{DS}}{2}\right) V_{DS} \tag{2.48}$$

where $V_{GS} > V_{TH}$, and Z is the depth of the channel measured normal to the cross section shown in Figure 2.40 (see Fig. 2.47).

The current–voltage characteristics given by Eq. (2.48) correspond to the part of the device characteristics to the left of the dashed line in Figure 2.41. This region of the device characteristics, where I_D increases with increasing V_{DS}, is called the *triode* region of device characteristics. If V_{DS} is increased further, such that $V_{DS} \geq V_{GS} - V_{TH}$, then the channel will be nearly pinched off at the drain end, and the output current level will saturate in a manner similar to the pinched operation of a JFET (see Section 2.4). This region of device characteristics is again called the pinched-off region or the pinched operation.

The saturation value of the drain current I_{DO} can be obtained from Eq. (2.48) by setting $V_{DS} = V_{GS} - V_{TH}$, that is,

$$I_{DO} = \frac{\mu_s C_{ox} Z}{2L}(V_{GS} - V_{TH})^2 \tag{2.49}$$

Comparing Eq. (2.49) with Eq. (2.37) for the case of the JFET, one sees that the MOSFET also exhibits a *square-law* characteristic, where the controlled current I_{DS} is proportional to the square of the net control voltage $V_{GS} - V_{TH}$. Similar to the case of the JFET, the MOS transistor current–voltage characteristics are linearly dependent on the channel width-to-length ratio Z/L. Thus, for a given channel length L, the current levels within the device can be scaled by making the channel width Z larger or smaller.

In actual MOSFET devices, the drain current I_D does not saturate at its ideal value of I_{DO}, but continues to increase slightly with increasing V_{DS}, as shown in Figure 2.42. This finite conductance of the output characteristics is due to the channel length modulation effects, where the channel drain depletion layer

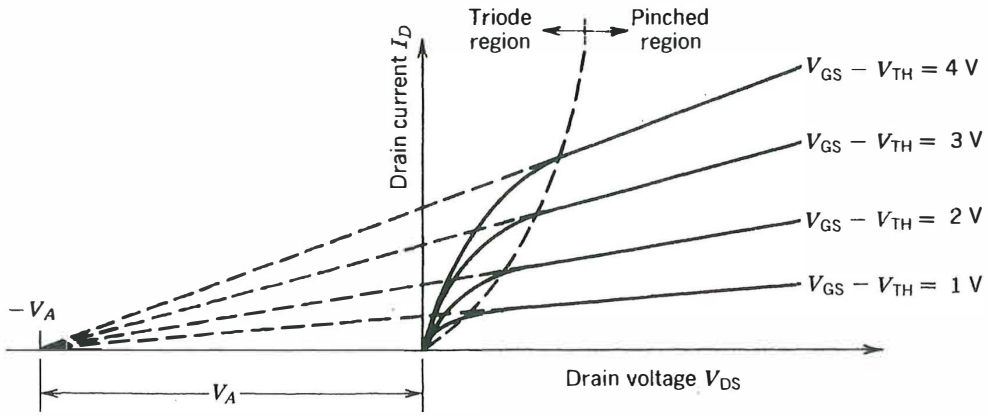


FIGURE 2.42. Actual current–voltage characteristics of a *n*-channel enhancement-mode MOSFET showing the effect of channel length modulation.

EX-1014 0117
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

extends back into the channel and causes the effective channel length L to appear lower than its actual value. This effect can be included in the device characteristics by defining a "channel-length modulation voltage" intercept V_A on the voltage axis of the current–voltage characteristics, as shown in Figure 2.42. Then Eq. (2.49) can be written as

$$I_{DO} = \frac{\mu_s C_{ox} Z}{2L} (V_{GS} - V_{TH})^2 \left(1 - \frac{V_A}{V_{DS}}\right) \tag{2.50}$$

to include the effects of channel length modulation. Note that for an n-channel device, V_A is negative and V_{DS} is positive; for a p-channel device, V_A is positive and V_{DS} is negative. Typical values of V_A are in the range of 30–50 V for MOS transistors with 8-μm channel length.

Depletion-Mode Devices

In the case of depletion-mode MOS transistors, a conducting channel exists under the gate region, with $V_{GS} = 0$. Such devices would exhibit a threshold voltage V_{TH} of opposite polarity. In other words, to bring the channel to the verge of conduction, one has to apply a gate bias of opposite polarity to "deplete" the existing channel from the free carriers.

Figure 2.43 shows the typical current–voltage characteristics of an n-channel depletion-mode MOSFET. For illustrative purposes, a threshold voltage V_{TH} of -2 V is assumed. Note that the device characteristics are virtually identical to those of Figure 2.42, except that the threshold voltage now has an opposite polarity.

All the basic equations derived for the enhancement-mode MOSFET are directly applicable to the depletion-mode devices, with the appropriate change of the polarity of V_{TH}.

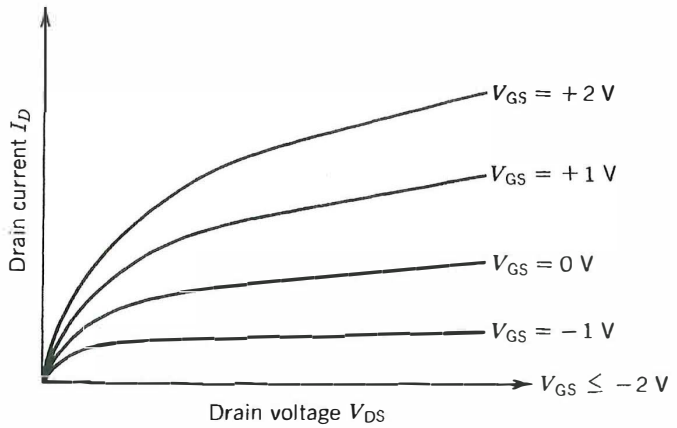


FIGURE 2.43. Current–voltage characteristics of a n-channel depletion-mode MOSFET with $V_{TH} = -2$ V.

EX-1014 0118
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Circuit Symbols

Several circuit symbols and sign conventions have been developed for MOS transistors. A set of the most commonly used circuit symbols for various MOSFET types are shown in Figure 2.44. In the device symbols, the channel conductivity type (i.e., either *n* or *p* type) is designated by the direction of the arrow on the body or the substrate terminal. This arrow designates the polarity of the *p–n* junction formed between the channel and the substrate, and is in the same direction as a forward-biased diode, that is, it points from the *p* side to the *n* side of the junction. The depletion-mode MOSFETs are shown with a double line across the source–drain regions to imply the existence of a finite conducting channel under the gate, with zero applied bias.

In many circuit drawings where the body or substrate connection is not explicitly shown, it is easier to describe or identify the MOS devices by a slightly different set of device symbols, as indicated in Figure 2.45. In this case, the source terminal is identified with an arrow, similar to that of the emitter of a bipolar transistor. The polarity of the arrow is in the direction of the *p*-*n* junction formed by the source-body junction, pointing from the *p* region to the *n* region.

In the rest of this book, unless specified otherwise, the circuit symbols of Figure 2.45 will be used since they are very similar to the basic bipolar *npn* and *pnp* transistor symbols. This makes it easier for an analog circuit designer trained in bipolar IC design to adapt to the use of MOS devices, and to derive many of the basic MOS building blocks, or subcircuits, by direct analogy to their bipolar counterparts. This subject will be covered in depth in Chapter 6.

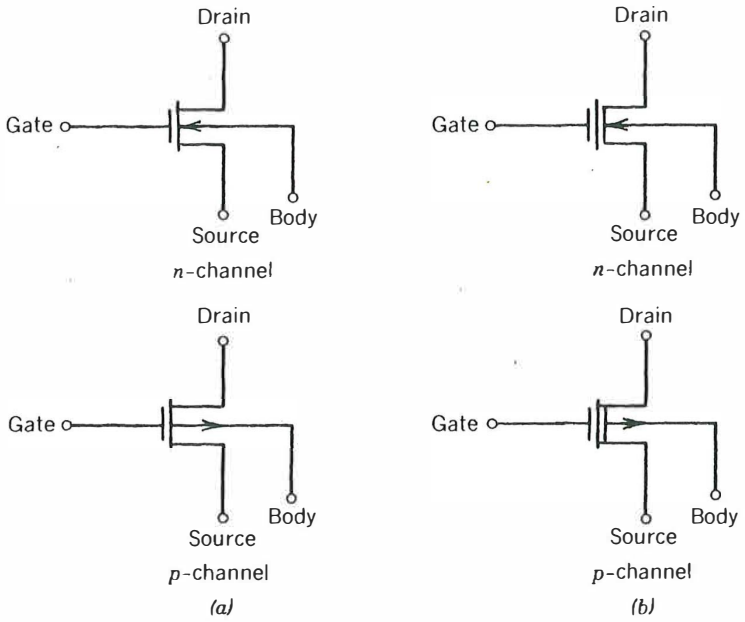


FIGURE 2.44. Set of commonly used circuit symbols for *p*- and *n*-channel (a) Enhancement mode; (b) depletion mode. MOS transistors.

EX-1014 0119
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

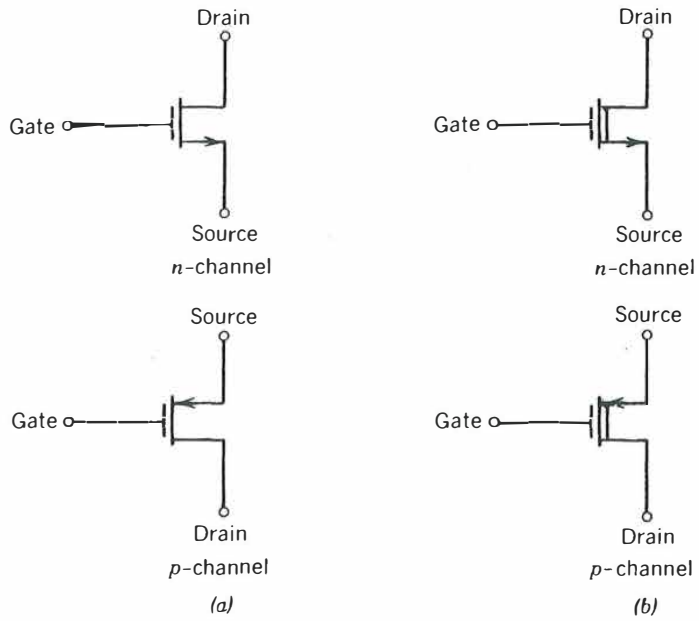


FIGURE 2.45. Alternate set of device symbols for p- and n-channel MOSFETS. These symbols are often preferred in MOS analog design because of their similarity to bipolar transistor notation. (a) Enhancement mode; (b) depletion mode.

Small-Signal Mode

For ac small-signal operation, the frequency and gain characteristics of a MOSFET can be closely approximated by the equivalent circuit of Figure 2.46 for $V_{DS} > V_{GS} - V_{TH}$, that is, for operation in the pinched region of device characteristics. This ac equivalent circuit is very similar to that for a JFET (see Fig. 2.39).

The device transconductance g_m in the pinched region can be derived by differentiating Eq. (2.49) as

$$g_m = \frac{\partial I_{DO}}{\partial V_{GS}} = \frac{\mu_s C_{ox} Z}{L}(V_{GS} - V_{TH}) \tag{2.51}$$

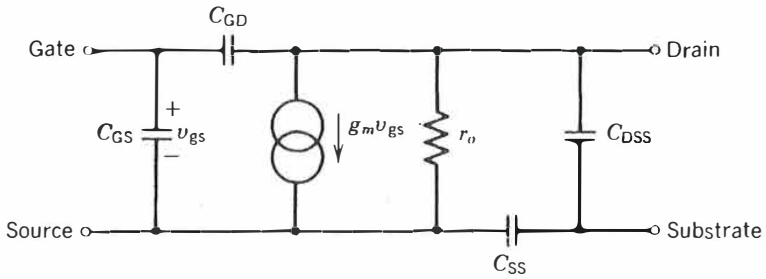


FIGURE 2.46. Ac small-signal equivalent circuit for a MOSFET.

EX-1014 0120
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

As was the case for JFET, the transconductance is directly proportional to the Z/L ratio of the device layout. Similar to the JFET, g_m also increases linearly with the gate voltage. A device exhibiting this device characteristic is known as a square law device and is particularly useful for multiplication or modulation of ac signals (i.e., mixer applications) or automatic gain control (AGC) functions.

The finite output resistance r_o is due to the channel length modulation effect and depends on the actual length L of the channel. By defining an intercept voltage V_A, as shown in Figure 2.42, the value of r_o can be approximated as

$$r_o \approx \frac{|V_A|}{I_{DO}} \tag{2.52}$$

for the cases where $V_A >> V_{GS}-V_{TH}$. For $V_A = 30$ V and $I_{DO} = 1$ mA, a typical value for r_o would be approximately 30 kΩ.

The parasitic capacitances C_{SS} and C_{DSS} are the capacitance between the source and drain bed diffusion and the body, or substrate, region. In general, the MOSFET is operated in a grounded-source configuration, with the source and the body regions connected together. In that case, the parasitic capacitance C_{SS} is short-circuited and has no effect on the circuit performance.

C_{GS} is the capacitance from the gate to the source and the unpinched region of the channel. It is an MOS capacitance. Thus, its value is equal to the gate capacitance per unit area C_{ox} of Eq. (2.45) times the area of overlap between the gate oxide and the source region. C_{GD} is the capacitance from the gate electrode to the drain. Its value is equal to C_{ox} times the area of overlap between the gate oxide and the drain region.

The value of C_{ox} is on the order of 4×10^{-4} pF per square micron of gate oxide area. Typical values of the zero-bias capacitances for C_{SS} and C_{DSS} are approximately 10^{-4} pF/μm^2, with a built-in potential ψ_0 of approximately 0.6 V. Thus, for a given device layout and bias condition, their values can be approximated using the generalized junction capacitance expression given in Eq. (2.30) with the exponent n equal to $\frac{1}{2}$.

A basic figure of merit for the high-frequency performance of a MOSFET is its transconductance frequency f_c given as

$$f_c = \frac{g_m}{2\pi C_{in}} \tag{2.53}$$

where C_{in} is the total input capacitance $C_{GS} + C_{GD}$ at the gate terminal, with the drain ac grounded. Since C_{in} is proportional to C_{ox} times the gate area, that is,

$$C_{in} = C_{ox}LZ \tag{2.54}$$

and g_m is given by Eq. (2.51), one can express f_c as

$$f_c = \frac{\mu_s(V_{GS} - V_{TH})}{2\pi L^2} \tag{2.55}$$

Equation (2.55) indicates that for optimizing high-frequency performance, a

EX-1014 0121
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

short channel length is required. As will be described in the next section, such short-channel device structures can be obtained using the so called self-aligned gate MOSFET structures.

Effect of Substrate Bias

In the derivation of the basic device equations and the ac equivalent circuit of Figure 2.46, it was assumed that the substrate, or body, of the MOS transistor is connected to the source. Although this is the case in most digital applications, a number of important exceptions to this case occur in analog IC design using MOSFETs. A typical example of such a case is when two MOS transistors of the same conductivity type are stacked such that one serves as an active load for the other.

The dependence of the MOSFET characteristics on the source-to-body bias V_{SB} is called the *body effect* and may be significant for a number of analog amplifier applications of monolithic MOSFET devices. The primary effect of an added reverse bias, V_{SB} between the body and the source is the increase of the gate threshold voltage V_{TH} above its nominal value. Another undesired result of the body effect is the reduction of the device transconductance and the output impedance when operated in a *cascode* configuration. These effects will be analyzed and discussed in more detail in Chapter 6.

Integrated MOSFET Devices

MOS transistors can be fabricated by the standard IC processing techniques described in Chapter 1. Figure 2.47 shows the cross sections and typical device layouts for p- and n-channel MOSFETs. For a p-channel device, a n-type substrate is used. First the n-type source and drain diffusions are performed. Next, the oxide is removed over the channel region and a thin oxide is grown. Although it is desirable to have a very thin layer of oxide to reduce the threshold voltage and to increase C_{ox}, in production, gate oxide thicknesses less than 1000 Å are generally not used because of reliability and yield problems. Following the thin oxide regrowth, windows are opened in the thick oxide over source and drain regions, and aluminum is deposited for contact and for gate metal. For an n-channel device, a similar procedure is used. The typical resistivity of the substrate region is on the order of 3–5 Ω/cm. The resistivity and the diffusion depth of the source and device drain regions are not critical, except that the side diffusion of these regions under the gate oxide must be taken into account in calculating the effective channel length L. Typically, these regions are diffused to a depth of 2–3 μm, and have a sheet resistivity range of 50–100 Ω/□.

The lateral geometry of the devices is also shown in Figure 2.47 for the so-called stripe geometry. Normally, the channel length L, is fixed by the process and masking tolerances, but the channel width Z is a design parameter which lets the designer optimize the device characteristics by setting the Z/L ratio.

EX-1014 0122
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

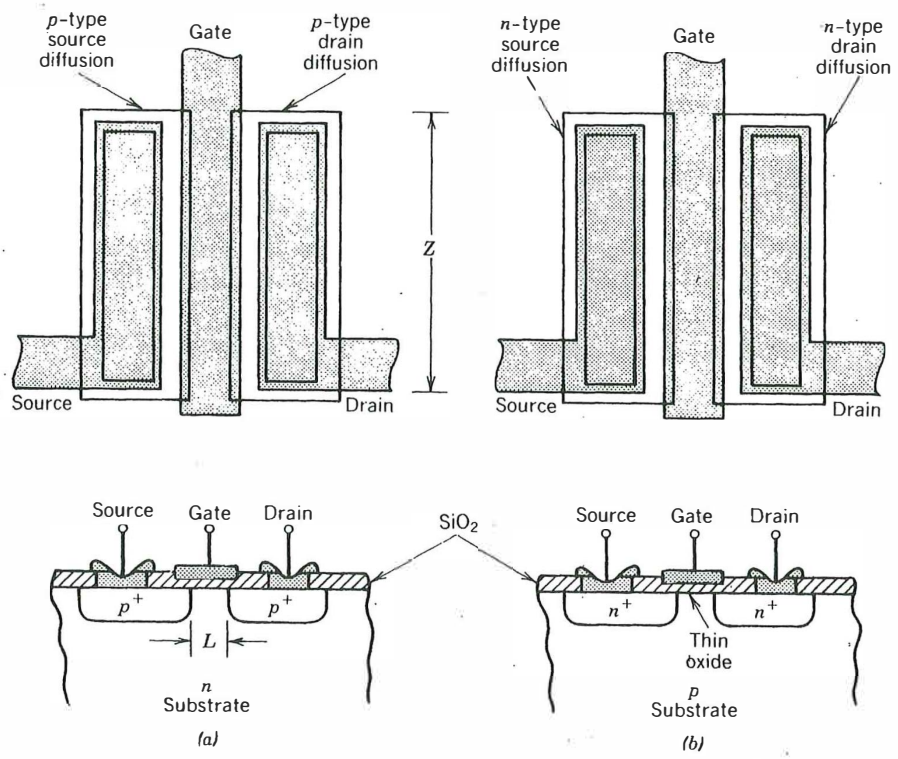


FIGURE 2.47. Typical device layouts and cross sections of p- and n-channel MOS transistors: (*a*) p-channel MOSFET; (*b*) n-channel MOSFEt.

Bipolar Compatible MOSFET Structures. The metal gate p-channel MOSFET transistor structure of Figure 2.47*a* can be formed with the standard bipolar IC processing steps, with the addition of one extra masking step to form the thin oxide layer over the gate region. The resulting device cross section is shown in Figure 2.48. The source and drain regions are formed by the p-type base diffusion of the *npn* transistor, and the n-type epitaxial layer forms the body of a MOSFET. The p-type inversion layer is then formed under the thin oxide

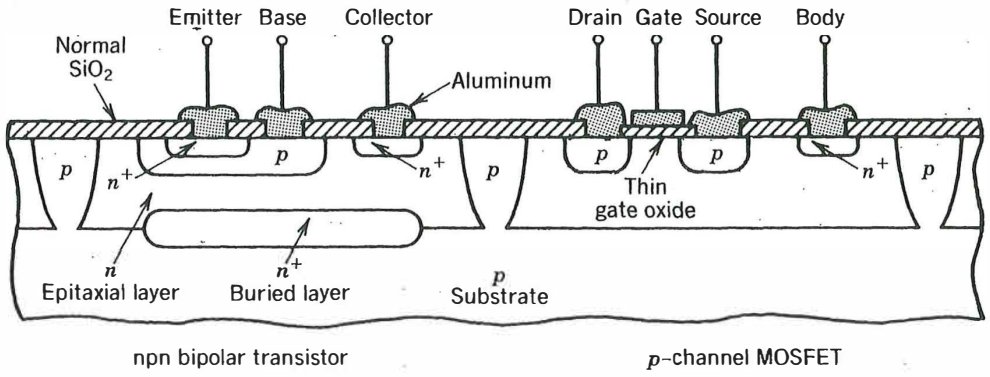


FIGURE 2.48. Bipolar compatible p-channel MOSFET structure.

EX-1014 0123
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

layer, when the gate voltage exceeds the threshold voltage. For such a device structure, an epitaxial layer resistivity of 3–5 Ω/cm is most readily compatible with either the medium- or the high-voltage process described in Section 2.1.

Although the process complexity of the bipolar compatible p-channel MOSFET structure of Figure 2.48 is not great, the dependence of the MOS transistor characteristics on the surface properties requires a more careful control of the surface conditions on the wafer than is typically required for standard bipolar IC processing. This is particularly true for controlling the absolute value and the long-term stability of the MOS threshold voltage V_{TH}. Thus, although the inclusion of p-channel MOSFETs along with bipolar transistors is feasible, it is more difficult than it may at first appear.

Self-Aligning Gate Structures. In the basic MOSFET device structures of Figure 2.47, it is necessary that the thin oxide and the metal gate overlap both the source and the drain regions in order to account for registration alignment errors. This means that there will be gate–source and gate–drain overlap capacitances. The latter is the more objectionable in applications where the device is used in the common-source connection, since this may lead to the Miller-effect multiplication of the gate–drain capacitance.

To overcome the capacitance problem associated with the gate-drain overlap, and to minimize the mask alignment problems in defining the gate region, a number of processing techniques have been developed. These techniques make the gate-channel registration automatic and eliminates the mask alignment problems. Figure 2.49 illustrates one such method, using ion implantation. The source and drain regions are diffused with standard processing techniques; however, they are spread somewhat further apart than the final channel length L. Then a thin oxide is regrown between the two p diffusions, and the gate metal is deposited and masked. However, an *undersize* gate metal mask is used such that the gate metal falls in between the two p diffusions but *does not* overlap them. Then the entire wafer surface is implanted with p-type (boron) ions. The gate metal now serves as a mask and blocks the boron ions from penetrating the region below it. Thus, the implantation step causes the p-type source and drain regions to extend to the edge of the gate metal, but not under it, as shown in Figure 2.49*b*. In this manner, both the gate overlap capacitance and the mask alignment tolerances are greatly reduced.

Silicon Gate Structures. Both the overlap capacitance and the threshold voltage V_{TH} can be reduced by using a gate electrode made up of doped silicon, rather than metal.[12] In this process, after the thin oxide growth, a layer of polycrystalline silicon is deposited over the entire wafer by CVD techniques. This silicon is then etched away from all areas except where the gate is to be located. Next, the windows are opened in the oxide layer to diffuse the p-type source and drain regions. In this process, the polycrystalline silicon gate region becomes p-type doped, and also serves as a self-aligning mask to separate the source and drain diffusions. Next, an oxide layer is deposited on the wafer

EX-1014 0124
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

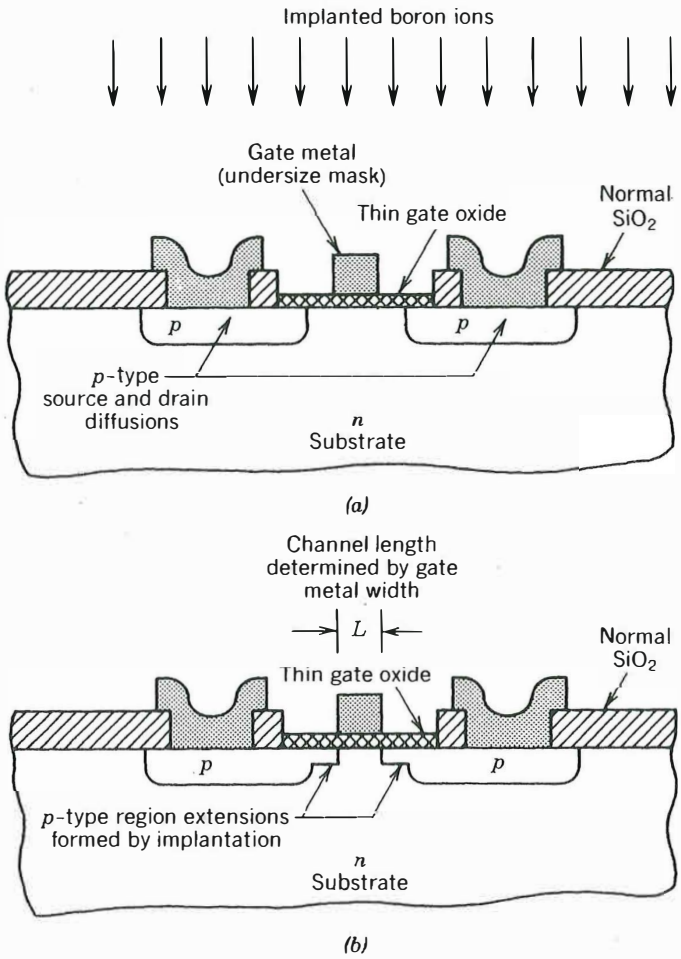


FIGURE 2.49. Self-aligning gate MOSFET structure using ion implantation and undersize gate metal mask to determine channel length. (Vertical scale exaggerated for clarity) (*a*) Before ion implantation; (*b*) after ion implantation.

surface and the contact windows are opened. This is followed by the conventional metal deposition to form the interconnections. The sequence of the process steps and the resulting device structure are shown in Figure 2.50. Note that the polycrystalline silicon gate electrode is sandwiched within the SiO_2 layer over the gate region.

Often a thin silicon nitride (Si_3N_4) layer is used as a gate dielectric, in silicon gate devices, in conjunction with the thin SiO_2 layer. This additional gate dielectric layer increases the gate capacitance C_{ox} since Si_3N_4 has a higher dielectric constant than SiO_2 and also helps stabilize the threshold voltage against long-term drift.

Complementary MOS (CMOS) Structures. In a number of circuit applications, it is desirable to have both the *n*-channel and the *p*-channel enhancement-

EX-1014 0125
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

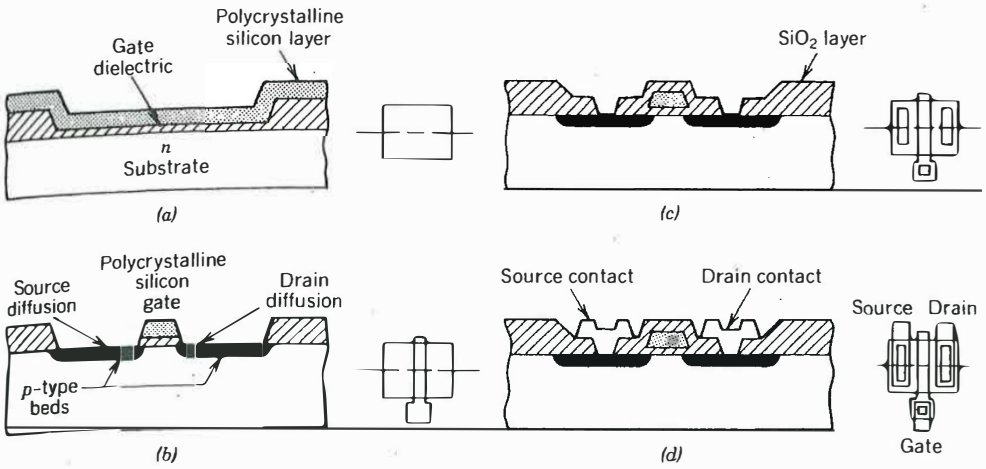


FIGURE 2.50. Typical sequence of steps in the fabrication of a p-channel silicon gate MOSFET: (*a*) Gate oxide and polycrystalline silicon are grown; (*b*) source and drain beds are diffused; (*c*) source and drain contacts are defined; (*d*) aluminum interconnections are deposited and etched.

mode devices on the same chip. This is particularly true for micropower circuits where low standby power is essential. In the case of analog design, the availability of complementary devices also simplifies the circuit design and biasing significantly.

It is possible to fabricate MOS transistors simultaneously by using the device structure shown in Figure 2.51. The basic process sequence proceeds as follows. First, a deep p-type pocket is diffused into the n-type substrate. The diffusion of this p-type pocket, called a *p well,* is critical since it requires a deep diffusion but low impurity concentration. Its doping concentration affects both the threshold voltage and the breakdown characteristics of the n-channel MOSFET. Next, after a masking step, a second p-type diffusion, called the p^+ *diffusion,* is made to form the source and drain regions of the p-channel MOSFET. At the same time, a ring of this p^+ diffusion is also applied around the periphery of the deep p well to avoid possible parasitic surface inversion along the lightly doped

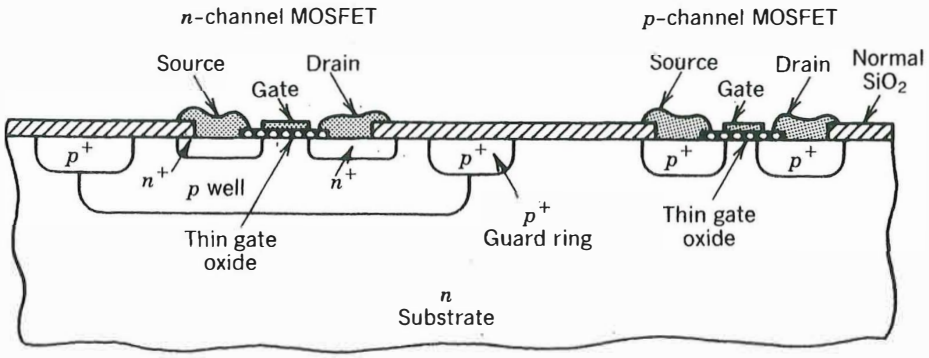


FIGURE 2.51. Complementary MOS transistor (CMOS) structure.

EX-1014 0126
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

surface of the p well. This p^+ ring around the p well is called a *channel stop*. Next, after a masking step, n^+ diffusion is performed to form the source and drain of the n-channel MOSFET within the p well. Finally, the thin gate oxide is masked and grown over the gate regions, and the metal interconnection pattern is formed.

CMOS has one added advantage over single-polarity MOSFET, it is immune to the substrate bias effect (body effect) in most circuit applications. This makes it particularly suitable for analog design, as will be described in more detail in Chapter 6.

In terms of the number of process steps, the extra masking steps make the CMOS structure somewhat more complex and more expensive to manufacture than the simple p- or n-channel MOSFETs. The side-diffusion effects of the deep p-well diffusion also consume a good fraction of the chip area, similar to the isolation diffusion of the bipolar process. Thus, the packing density of the devices on a CMOS chip is somewhat lower than that of comparable circuits using only p- or n-channel devices.

REFERENCES

1. S. K. Ghandhi, *The Theory and Practice of Microelectronics*, Wiley, New York, 1968, Chaps. 13 and 14.
2. H. DeMan, "The Influence of Heavy Doping on the Emitter Efficiency of a Bipolar Transistor,"*IEEE Trans. Electron Dev.* ED-18, 833–835 (October 1971).
3. R. J. Whittier and D. A. Tremere,"Current Gain and Cutoff Frequency Falloff at High Currents,"*IEEE Trans. on Electron Dev.* ED-16, 39–57 (January 1969).
4. P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, Wiley, New York, 1977, Chap. 1.
5. B. A. McDonald,"Avalanche Degradation of h_{FE}," *IEEE Trans. on Electron Dev.* **ED-17,** 871–878 (October, 1970).
6. A. B. Grebene, *Analog Integrated Circuit Design*, VanNostrand Reinhold, New York, 1972, Chap. 11.
7. A. B. Grebene and S. K. Ghandi,"General Theory for Pinched Operation of the Junction-Gate FET,"*Solid State Electron.* 12, 573–589 (1969).
8. R. Cobbold, *Theory and Applications of Field-Effect Transistor*, Wiley, New York, 1970.
9. P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, Wiley, New York, 1977, pp. 46–53.
10. A. S. Grove, *Physics and Technology of Semiconductor Devices*, Wiley, New York, 1967, Chaps. 8, 9.
11. W. M. Penney, Ed. *MOS Integrated Circuits*, VanNostrand Reinhold, New York, 1972, Chap. 20.
12. F. Faggin and T. Klein, "Silicon-Gate Technology,"*Solid State Electron.* **13,** 1125 (1970).

EX-1014 0127
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

CHAPTER FOUR

BIAS CIRCUITS

The starting point of a monolithic IC design is the design of the bias circuitry internal to the chip. This design step is very critical since it determines the internal voltage and current levels over all operating conditions of the integrated circuit as well as over all manufacturing process variations.

For a linear circuit designer trained in the area of discrete circuits, the basic constraints and limitations of monolithic circuit technology often pose a difficult challenge. This is particularly true with regard to the biasing circuitry. Many of the conventional biasing techniques cannot be directly applied to monolithic designs because of the following limitations of IC components:

1. Poor absolute-value tolerances.
2. Poor temperature coefficients.
3. Limitations on component values.
4. Lack of coupling capacitors.
5. Limited choice of compatible active devices.

On the other hand, IC fabrication methods offer a number of unique and powerful advantages to the circuit designer:

1. Availability of a large number of active devices.
2. Good matching and tracking of component values.
3. Close thermal coupling.
4. Control of device layout and geometry.

Fortunately, over the years a number of basic circuit configurations, or subcircuits, have been developed, which make efficient use of the advantages of monolithic technology, while avoiding most of its shortcomings. The purpose of

EX-1014 0128
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

this chapter is to outline some of these biasing techniques, along with the design philosophy and the guidelines associated with them. The basic subcircuits discussed in this chapter déal with the dc design of a monolithic circuit, with regard to the current and voltage bias levels within the circuit, and their drift with temperature or power supply changes. These basic subcircuits form a useful set of building blocks which, along with the basic gain stages described in the next chapter, serve as the starting point of more complex analog functions and subsystems.

4.1. CONSTANT-CURRENT STAGES

In a constant-current stage, the reference current in one branch of the circuit is accurately reproduced or reflected in a second branch, relatively independent of the absolute values of the device parameters. Because of this property, these subcircuits are also known as *current mirror* circuits. Such circuit configurations are particularly useful building blocks for analog circuit design, since they provide a means of establishing the dc bias levels within the circuit, within the accuracy of the matching or tracking properties of the monolithic components.

Basic Current Mirror

The basic current mirror subcircuit is shown in Figure 4.1. It is made up of two matched transistors, where one transistor, Q_1, is connected as a diode and sets the base–emitter voltage V_{BE} of Q_2. The operating principle of the current mirror is derived from the basic properties V_{BE} of the transistor and the collector current I_C as given by Eq. (2.5),

$$V_{BE} = V_T \ln\left[\frac{I_C}{I_{CO}}\right] \tag{4.1}$$

where V_T is the thermal voltage ($=kT/q$), and I_{CO} is the reverse saturation current. The reverse saturation current is related to the area A of the emitter–base junction as

$$I_{CO} = I_S A \tag{4.2}$$

where the constant of proportionality I_S depends on the intrinsic semiconductor parameters, such as the minority carrier diffusion lengths and the concentrations on both sides of the junction. For the IC devices on the same chip, which are fabricated simultaneously, this term will be the same. Therefore, if the two *npn* transistors are operated with the same base–emitter voltage, their collector currents will be related to each other as the ratio of their emitter areas, that is,

$$\frac{I_{C1}}{I_{C2}} = \frac{A_1}{A_2} \tag{4.3}$$

EX-1014 0129
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

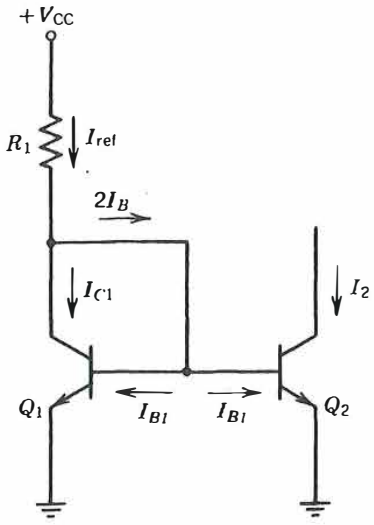


FIGURE 4.1. Basic current mirror.

In the current mirror circuit of Figure 4.1, the base–emitter voltages of the two transistors are forced to be equal. This in turn forces the collector currents of Q_1 and Q_2 to be equal, that is,

$$I_2 = I_{C1} = I_{\text{ref}} - I_{B1} - I_{B2} \tag{4.4}$$

where I_{B1} and I_{B2} are the base currents of Q_1 and Q_2. Since the base currents are also proportional to the respective emitter areas, I_2 can be related to I_{ref} from Eq. (4.3) as

$$I_2 = I_{\text{ref}} \frac{A_2}{A_1} \left(\frac{\beta}{\beta + 1 + A_1/A_2} \right) \tag{4.5}$$

which, for the case of identical devices (i.e., $A_2 = A_1$) reduces to

$$I_2 = I_{\text{ref}} \left(\frac{\beta}{\beta + 2} \right) \approx I_{\text{ref}} \left(1 - \frac{2}{\beta} \right) \tag{4.6}$$

If $\beta >> 1$,* both equations converge to

$$I_2 \approx I_{\text{ref}} \tag{4.7}$$

Another potential source of error in the current mismatch is the mismatch of the V_{BE} of Q_1 and Q_2. For small values of mismatch ΔV_{BE} this effect can be incorporated into Eq. (4.6) as an additive term,

$$I_2 = I_{\text{ref}} \left(1 - \frac{2}{\beta} + \frac{\Delta V_{\text{BE}}}{V_T} \right) \tag{4.8}$$

*In order to minimize the subscript notation, the symbol β will be used to mean either β_F or β_o, interchangeably.

EX-1014 0130
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

where $\Delta V_{BE} = V_{BE2} - V_{BE1}$. Note that ΔV_{BE} can be positive or negative, depending on the polarity of the mismatch. The above expression indicates that *each millivolt of mismatch in the V_{BE} of the two transistors results in approximately 4% of current mismatch.*

The basic current–voltage equations (4.1) through (4.8) hold over a broad temperature range (typically −60 to +150°C) and over six orders of magnitude of current. Thus, the basic current mirror circuit of Figure 4.1 provides a means of obtaining a current reference level, independent of implicit device parameters, which can be "scaled" by proper choice of the emitter areas of the two transistors. Assuming typical V_{BE} mismatch of $\leq \pm$ 1 mV and β in the range of 100–200, the typical output current I_2 would be within ±5% of I_{ref} over a wide range of temperature and current levels.

Since I_{ref} is set by resistor R_1, I_2 can also be expressed in terms of the setting resistor R_1 as

$$I_2 \approx \frac{V_{CC} - V_{BE}}{R_1} \frac{A_2}{A_1} \approx \frac{V_{CC}A_2}{R_1A_1} \tag{4.9}$$

assuming that $V_{CC} \gg V_{BE}$.

The output impedance of the basic current mirror circuit is primarily due to the Early effect of Q_2 (see Fig. 2.7). This effect also tends to contribute to the current mismatch as the collector voltage V_2 of Q_2 is increased. It can be incorporated into the basic current expression by means of a multiplication term as

$$I_2 = I_{ref}\left(1 - \frac{2}{\beta}\right)\frac{1 + V_2/V_A}{1 + V_{BE}/V_A} \tag{4.10}$$

where V_A is the Early voltage associated with the *npn* transistor (see Table 2.2). It should be noted that for typical values of $V_A \approx 100$ V, this second term can contribute as much as 25% of current mismatch at a collector voltage V_2 of 25 V.

The output resistance R_{out} of the current mirror is equal to the collector

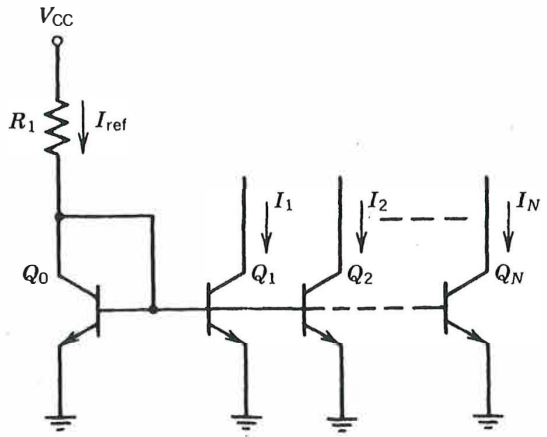


FIGURE 4.2. Basic current mirror with multiple outputs.

EX-1014 0131
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

resistance r_o of the hybrid-π model, given by Eq. (2.20), which can be written as

$$R_{\text{out}} \approx \frac{V_A}{I_2} \tag{4.11}$$

For $V_A \approx 100$ V and $I_2 = 1$ mA, this results in an output impedance of approximately 100 kΩ. For ac calculations, it should be remembered that R_{out} is also shunted to the substrate by means of the parasitic collector–substrate capacitance C_{CS} of Q_2.

In many applications, it is advantageous to derive not one but several currents from a reference current I_{ref}. This can be done by using a multiple-output current mirror, as shown in Figure 4.2. In this case, assuming N identical transistors and neglecting the Early effect of the output transistors, the resulting currents are related to I_{ref} as

$$I_1 = I_2 = \cdots = I_N = I_{\text{ref}}\left(1 - \frac{N+1}{\beta}\right) \tag{4.12}$$

When using the multiple-output current mirror of Figure 4.2, a word of caution is in order. If any one of the output transistors, Q_1 through Q_N, is driven into saturation, its collector–base junction becomes forward biased and causes a part of I_{ref} to be shunted to ground, through it. This effect, which is known as *current hogging*, then causes saturated transistor to use up more than its share of the available base current. If only a limited amount of base current is available, this causes the currents of the remaining outputs to *decrease* below the nominal value given by Eq. (4.12). As will be discussed later, this phenomenon is sometimes useful as a technique for detecting the onset of saturation.

The basic current mirror configuration is also very useful as a controlled-gain transistor, as illustrated in Figure 4.3. The resulting three-terminal device is

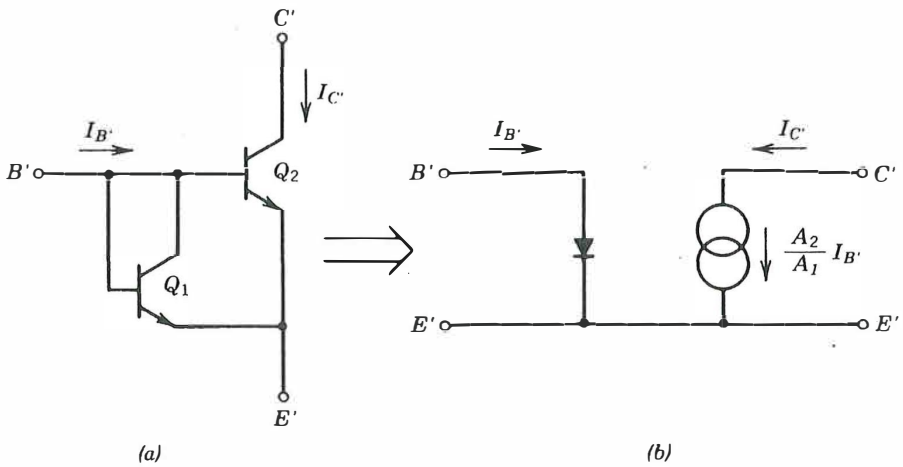


FIGURE 4.3. Current mirror as a controlled-gain transistor: (*a*) Circuit connection; (*b*) simplified equivalent circuit for large-signal operation. (*Note:* A_1 and A_2 are the respective emitter areas of Q_1 and Q_2.)

EX-1014 0132
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

effectively equivalent to an *npn* transistor where the effective current gain β_{eff} is equal to the area ratio of Q_1 and Q_2, that is,

$$\beta_{\text{eff}} = \frac{I_C'}{I_B'} = \frac{A_2}{A_1}\left(1 - \frac{1 + A_2/A_1}{\beta}\right) \tag{4.13}$$

or

$$\beta_{\text{eff}} = \frac{A_2}{A_1} \qquad \text{for} \quad \beta \gg 1 \tag{4.14}$$

As will be described later, this controlled-gain transistor is particularly useful in designing high-current *pnp* current mirrors (see Fig. 4.19) or output stages with fixed current drive capability.

Base Current Compensated Current Mirror

The main error source in the basic current mirror circuits of Figures 4.1 and 4.2 are the finite values of the base currents which directly subtract from I_{ref}. This source of error can be greatly reduced by adding an extra transistor Q_3 into the circuit, as shown in Figure 4.4. The base currents of Q_1 and Q_3 are then supplied from Q_3, and only the base current of Q_3 is subtracted from I_{ref}. Thus, assuming that Q_1 and Q_2 are well matched, the output current differs from I_{ref} by I_{B3},

$$I_2 = I_{\text{ref}} - I_{B3} = I_{\text{ref}} - \frac{I_{B1} + I_{B2}}{\beta_3} \tag{4.15}$$

and it can be related to I_{ref} as

$$I_2 = I_{\text{ref}}\left(1 - \frac{1}{2 + \beta\beta_3}\right) \tag{4.16}$$

where β is the current gain of Q_1 or Q_2, and β_3 is the gain of Q_3. In the above expression, β_3 is identified separately since Q_3 operates at a much lower current

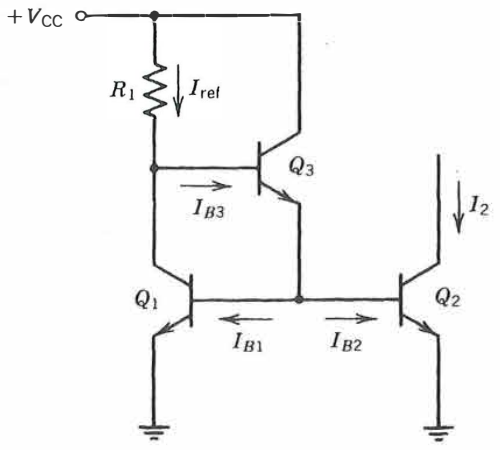


FIGURE 4.4. Base current compensated current mirror.

EX-1014 0133
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

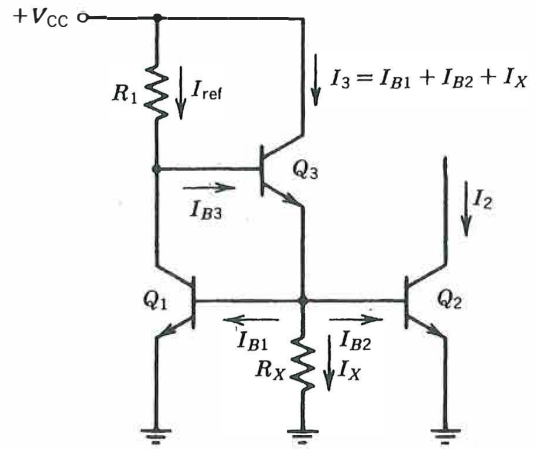


FIGURE 4.5. Base current compensated current mirror with current shunting resistor R_X. (*Note:* Normally I_X is set 5–10 times higher than I_{B1} or I_{B2}.)

level than Q_1 and Q_2 and, thus, may exhibit a lower value of β. In some designs, to avoid this problem, an additional resistance, R_X, is connected from the bases of Q_1 and Q_2 to ground, shown in Figure 4.5, to increase the current in Q_3 by an amount I_X, where

$$I_X = \frac{V_{BE}}{R_X} \tag{4.17}$$

I_X is the current in R_X. Typically R_X is chosen to set I_X to within 5–10 times I_{B1} or I_{B2}.

Since the accuracy of the base current compensated current mirror is largely insensitive to the base current errors, such a circuit is often used to set up a multiple number of constant-current stages from a given reference current, as shown in Figure 4.6. In a generalized case with N identical outputs, assuming

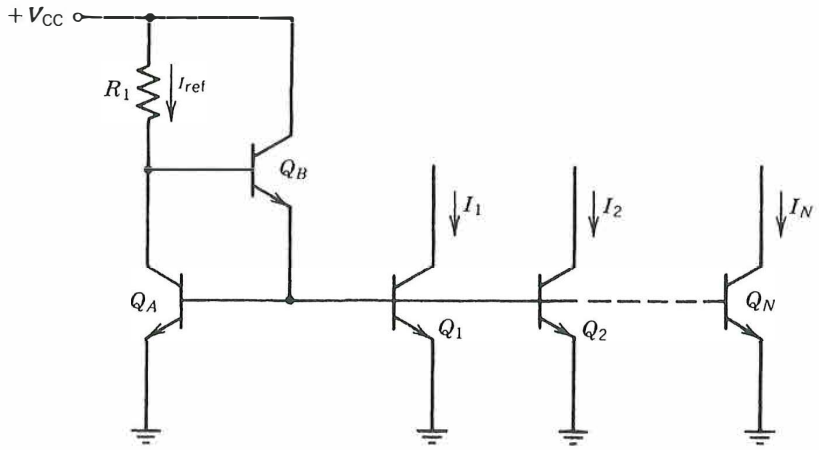


FIGURE 4.6. Base current compensated current mirror with multiple outputs.

EX-1014 0134
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

all transistors to be matched, the output currents are related to I_{ref} as

$$I_1 = I_2 = \cdots = I_N = I_{ref}\left(1 - \frac{N+1}{N+1+\beta^2}\right) \tag{4.18}$$

which simplifies to

$$I_1 = I_2 = \cdots I_N \approx I_{ref}\left(1 - \frac{N+1}{\beta^2}\right) \tag{4.19}$$

The base current compensated current mirror is also very useful for designing constant current stages that have a much higher output current than I_{ref}. This can be done by making the emitter are of the output transistor Q_2 to be much larger than that of Q_1. Assuming that the two emitter areas are related by the ratio N, such that $N = A_2/A_2$, the output current I_2 can be related to I_{ref}

$$I_2 \approx NI_{ref}\left(1 - \frac{N+1}{\beta^2}\right) \approx NI_{ref} \tag{4.20}$$

Note that Eq. (4.20) is very similar to the case of the multiple-output current source [Eq. (4.19)], except that N does not have to be an integer.

To summarize, the key advantage of base current compensation is the greatly reduced dependence of the output current on the transistor β. This makes it particularly suited to the design of multiple-output current mirror circuits, or to the scaling of output currents. However, the output impedance R_{out} is not affected by the base current compensation and is still the same as that given by Eq. (4.11). The methods of increasing this output impedance will be discussed in later sections of this chapter.

Resistor-Ratioed Current Mirror

The basic current mirror subcircuit can be modified to use resistor ratios, rather than the transistor emitter areas, to scale the current. Such a modified version of the circuit, often called the *resistor-ratioed current mirror*, is shown in Figure 4.7.

Neglecting the base currents, the current levels through each of the two transistors are related as

$$I_{ref}R_1 + V_{BE1} = I_2R_2 + V_{BE2} = V_A \tag{4.21}$$

The base–emitter voltage drop difference between two identical transistors operating at the respective collector currents I_{ref} and I_2 can be written as

$$\Delta V_{BE} = V_{BE2} - V_{BE1} = V_T \ln\left(\frac{I_2}{I_{ref}}\right) \tag{4.22}$$

EX-1014 0135
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

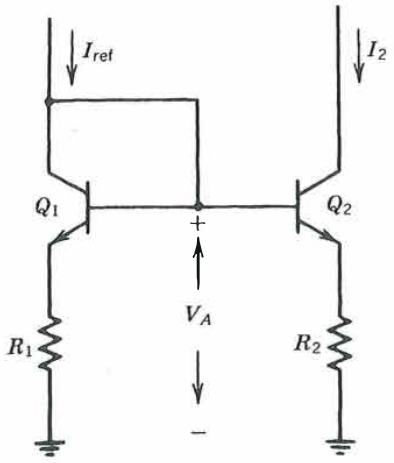


FIGURE 4.7. Resistor-ratioed current mirror.

Thus, from Eqs. (4.21) and (4.22) the ratio of the two currents can be expressed as

$$\frac{I_2}{I_{ref}} = \frac{R_1}{R_2}\left[1 - \frac{V_T \ln(I_2/I_{ref})}{R_1 I_{ref}}\right] \tag{4.23}$$

If the voltage drop across R_1 is made to be significantly larger than V_T and is comparable to V_{BE}, then the second term within the brackets becomes negligibly small and the two currents are closely related by the resistor ratios, that is,

$$\frac{I_2}{I_{ref}} \approx \frac{R_1}{R_2} \tag{4.24}$$

For $I_{ref}R_1 \geq V_{BE}$, the two currents follow the resistor ratio of Eq. (4.24) with a maximum error of less than $\pm 10\%$ over two orders of magnitude in current, that is,

$$\frac{1}{10} < \frac{I_2}{I_{ref}} < 10 \tag{4.25}$$

independent of temperature.

The resistor-biased constant-current stage is preferred over the simple diode-biased current mirror of Figure 4.1 for the cases where the ratio of I_2/I_{ref} is significantly different than unity, since the resistor ratios can be varied over a broader range of values than the emitter areas.

In the resistor-ratioed current mirror circuit, the resistor R_2 in series with the emitter of Q_2, serves as a series feedback resistor and increases the output resistance of the circuit. Using the simple hybrid-π model of Figure 2.10, the output resistance of the resistor-ratioed current mirror can be approximated as

$$R_{out} \approx r_o \frac{R_1 + \beta R_2}{R_1 + R_2} \tag{4.26}$$

EX-1014 0136
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

In practical cases, where $\beta R_2 >> R_1$, this expression simplifies to

$$R_{\text{out}} \approx \beta r_o \frac{R_2}{R_1 + R_2} \tag{4.27}$$

where r_o is the collector–emitter resistance of the hybrid-π model given in Eq. (2.20). Comparing Eq. (4.11) with Eq. (4.27), one sees that the output resistance of the resistor-ratioed current mirror is approximately β times higher than that of the ordinary current mirror, provided that R_2 is sufficiently large.

It should be remembered that the basic results given in Eqs. (4.24) and (4.27) are valid only if the voltage drop across R_1 and R_2 is comparable to V_{BE}. In certain design applications requiring very low current levels, in the low microampere range, this requirement may result in excessively large values of R_1 or R_2, and will make the use of resistor ratioing impractical.

Figure 4.8 shows a special case of the resistor-ratioed current mirror, with R_1 set equal to zero. Such a configuration is particularly useful for obtaining very low values of output current with relatively large values of the reference current, and without requiring high-value resistors. Assuming matched device geometries, Q_1 operates at a higher current density than the base–emitter diode of Q_2; and the voltage drop V_2 across R_2 is constrained to be

$$V_2 = I_2 R_2 = V_{\text{BE1}} - V_{\text{BE2}} = V_{\text{T}} \ln \left(\frac{I_{\text{ref}}}{I_2} \right) \tag{4.28}$$

Solving for R_2, we can write

$$R_2 = \frac{V_T}{I_2} \ln \left(\frac{I_{\text{ref}}}{I_2} \right) \tag{4.29}$$

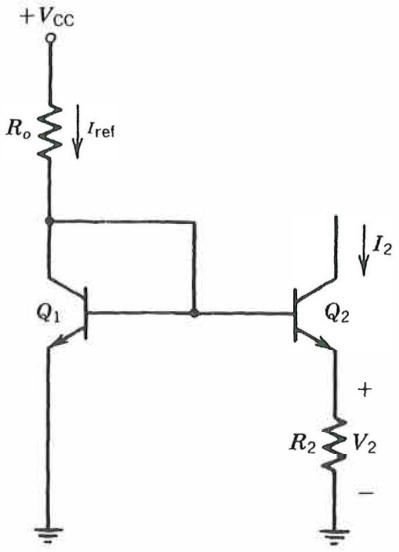


FIGURE 4.8. Constant-current stage for low current levels.

EX-1014 0137
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Since R_2 is proportional to the logarithm of the current ratio, a high degree of current mismatch can be obtained without requiring extremely high resistor values. For example, letting $R_2 = 20\ \text{k}\Omega$, one obtains a 100 : 1 ratio between the two currents, with $I_{\text{ref}} = 1$ mA and $I_2 \approx 10\ \mu\text{A}$.

An added advantage of the constant-current circuit of Figure 4.8 is that the output current I_2 is relatively independent of the supply voltage, that is,

$$I_2 = \frac{V_T}{R_2} \ln\left(\frac{V_{\text{CC}} - V_{\text{BE}}}{R_o I_2}\right) \tag{4.30}$$

Thus, for $I_{\text{ref}} \gg I_2$, the output current varies as the *logarithm* of the supply voltage. Such a low-value constant-current stage is particularly useful for biasing the input stage of an operational amplifier where it can set quiescent current levels relatively independently of supply voltage changes. However, it should be noted that in Eq. (4.30), V_T is directly proportional to temperature. Thus, I_2 would also exhibit a strong temperature dependence, depending on the temperature coefficient of R_2 and V_T.

The output resistance of the low-current constant-current stage can be easily calculated from the simplified hybrid-π model of Figure 2.10. Assuming that the transconductance g_{m2} of Q_2 is low enough such that $g_{m2}R_2 \ll \beta$, one can show that[1]

$$R_{\text{out}} \approx r_o\left(1 + \frac{I_2 R_2}{V_T}\right) \tag{4.31}$$

Thus, the output resistance R_{out} depends strongly on the value of the voltage drop across R_2. In the current source configuration of Figure 4.8, I_2R_2 is limited to several hundred millivolts for practical current ratios, which limits the maximum value of R_{out} to approximately $10r_o$.

In many applications, a large number of low-value constant-current stages can be driven from a single current reference, as shown in Figure 4.9. Since each

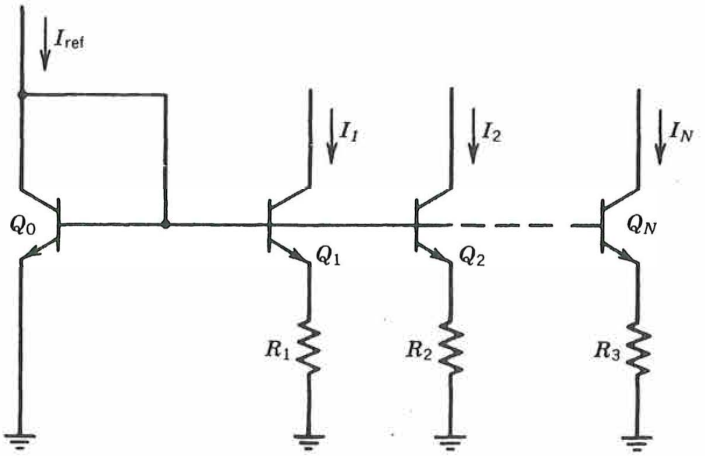


FIGURE 4.9. Driving a multiple number of low-value current sources from a single reference.

EX-1014 0138
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

of the current outputs operates at much lower current levels than I_{ref}, the base currents of Q_1 throught Q_N have a negligible effect in calculating the output currents. Each of the output currents can be set to its nominal value by calculating the value of the emitter resistor corresponding to it.

Wilson Current Mirror[2]

The β dependence and the low output impedance of the basic current mirror can be greatly improved by using the so-called *Wilson current mirror* configuration shown in Figure 4.10. In this configuration, the base current of Q_2, which is extracted from the reference current I_{ref}, is resupplied back to the base of the reference transistor Q_1, thus keeping the current levels in Q_1 and Q_2 unaffected by the base current changes. The V_{BE} drop across Q_3 sets the bias level for Q_1, which in turn sets the current level in Q_3. It can be shown by straightforward nodal analysis that the output current I_2 is related to the reference current I_{ref} as

$$I_2 = I_{ref} + (I_{B1} + I_{B3} - 2I_{B2}) \tag{4.32}$$

Note that the additional terms within the parentheses represent the third-order errors due to the base current mismatches, and reduce to zero if the transistor β are matched. If the transistor β are matched to better than $\pm 20\%$, the difference between I_2 and I_{ref} is less than 0.5% for a typical β value of 200.

Further insight can be gained into the operation of the improved constant-current stage of Figure 4.10 by treating it as a special case of the shunt-series feedback, with a unity feedback connection in the shunt branch. This corresponds to the unity current feedback condition from the emitter of Q_2 back to the

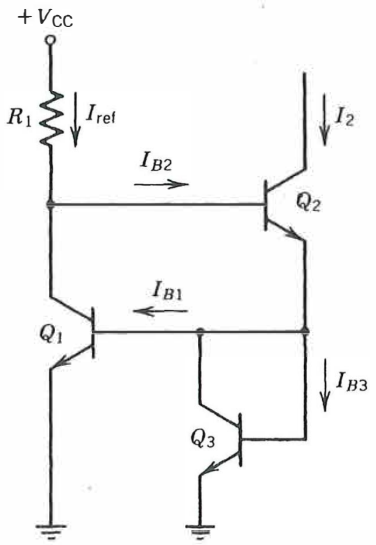


FIGURE 4.10. Wilson current mirror.

EX-1014 0139
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

base of Q_1. As a consequence of this feedback arrangement, the output impedance of the current mirror stage is greatly increased. By using the hybrid-π model, the output resistance of the Wilson current mirror can be calculated to be[3]

$$R_{\text{out}} \approx \frac{\beta r_o}{2} \tag{4.33}$$

One basic drawback of the Wilson current mirror is that it cannot be easily extended to a multiple number of outputs from a single I_{ref} without upsetting the base current cancellation feature given in Eq. (4.32). This problem can be overcome by the modified circuit shown in Figure 4.11, by splitting the output transistor Q_2 into two matched transistors Q_{2A} and Q_{2B}. Then the base current cancellation feature is retained, but two matched output currents are provided, where

$$I_{2A} = I_{2B} = \frac{I_{\text{ref}}}{2} \tag{4.34}$$

Cascode-Connected Current Mirrors

In many applications where a high output impedance is very desirable, it may be advantageous to cascode or "stack" two current mirrors together, as shown

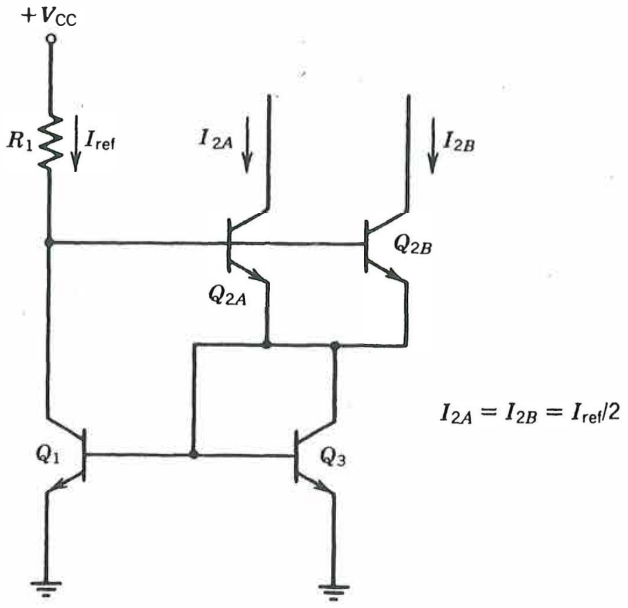


FIGURE 4.11. Wilson current mirror with multiple outputs.

EX-1014 0140
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

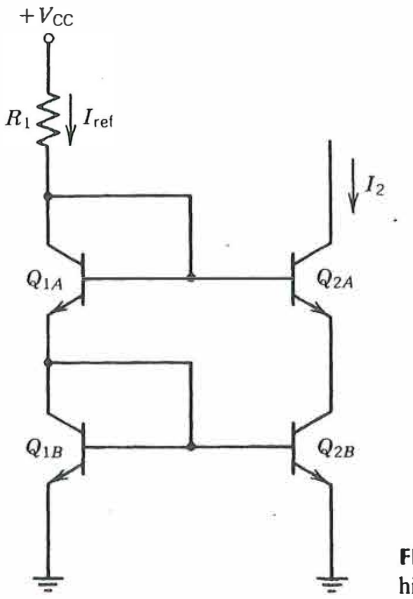


FIGURE 4.12. Cascode-connected current mirror with high output impedance.

in Figure 4.12. In such a configuration, assuming that all the transistors are well matched, the output current I_2 is related to the reference current I_{ref} as

$$I_2 = I_{ref}\left(1 - \frac{4}{\beta}\right) \tag{4.35}$$

In this configuration, transistor Q_{2B} serves as a constant-current source in series with the emitter of Q_{2A} and forces the output resistance R_{out} to be

$$R_{out} \approx \beta r_o \tag{4.36}$$

In a typical small-signal *npn* transistor, where $r_o \approx 100\ \text{k}\Omega$ at $I_E = 1$ mA, with $\beta = 200$, this results in $R_{out} \approx 20\ \text{M}\Omega$.

In certain applications where not one but a multiple number of high-impedance outputs are required, the multiple-output cascode current mirror circuit of Figure 4.13 can be used. In this configuration, Q_0 sets the current levels in the lower sets of current source transistors, with Q_B providing the base current compensation. Q_A provides the base current compensation for the upper set of current source transistors. Since Q_A and Q_B are a set of Darlington, or *common collector*, transistors, they can share the same isolation pocket.

The output currents I_1 through I_N are related to the reference current by a modified version of Eq. (4.19) as

$$I_1 = I_2 = \cdots = I_N = I_{ref}\left(1 - \frac{2(N+1)}{\beta^2}\right) \tag{4.37}$$

The output resistance of each of the outputs will still be that of the basic cascode current mirror, that is, $R_{out} = \beta r_o$.

EX-1014 0141
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

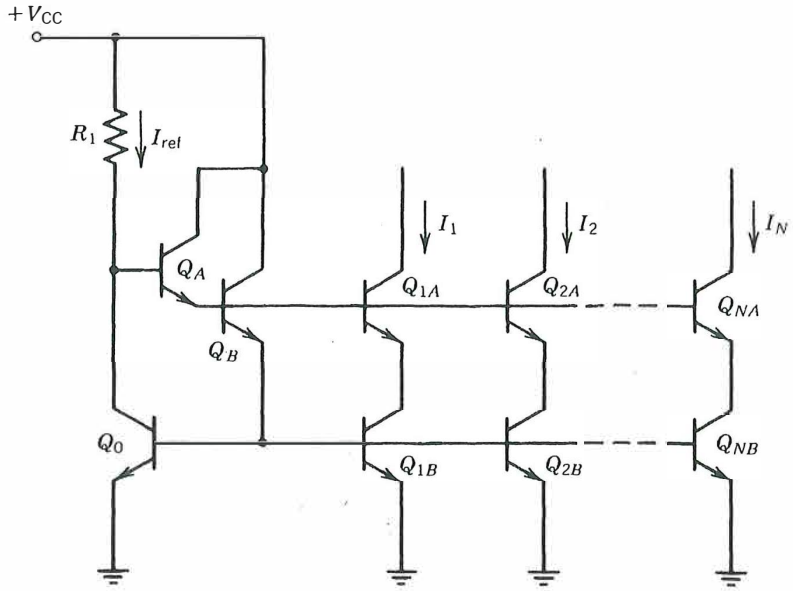


FIGURE 4.13. Cascode-connected current mirror with multiple outputs and base current compensation.

4.2. *pnp* CURRENT SOURCES

In principle, all of the *npn* current mirror or constant-current stage configurations discussed in the previous section apply to the *pnp* transistors, with the appropriate polarity reversal of the currents and the bias voltages. However, in practice, these stages suffer somewhat from the three basic shortcomings of the lateral *pnp* transistor, which are the following:

1. Low value of β. (Typically, β is in the range of 10–50 for lateral *pnp* devices.)
2. Low output resistance. (Since the base region of the lateral *pnp* transistor is lighter doped than its collector, the collector–base depletion region extends mostly into the base region, resulting in excessive base width modulation and a relatively low Early voltage, $V_A \approx 50$ V.)
3. Limited current-handling capability. (For practical lateral *pnp* transistor geometries, β falls off rapidly for I_C in excess of 50–100 μA.)

Thus, in choosing the appropriate *pnp* current mirror configuration, these three basic limitations associated with the lateral *pnp* transistors must be kept in mind. On the other hand, the availability of the lateral *pnp* transistor in a multiple-collector configuration (see Fig. 2.27) provides some unique advantages, particularly in terms of device interconnection and layout. Figure 4.14*a* shows the basic *pnp* current mirror. Since Q_1 and Q_2 both share a common emitter and a

EX-1014 0142
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

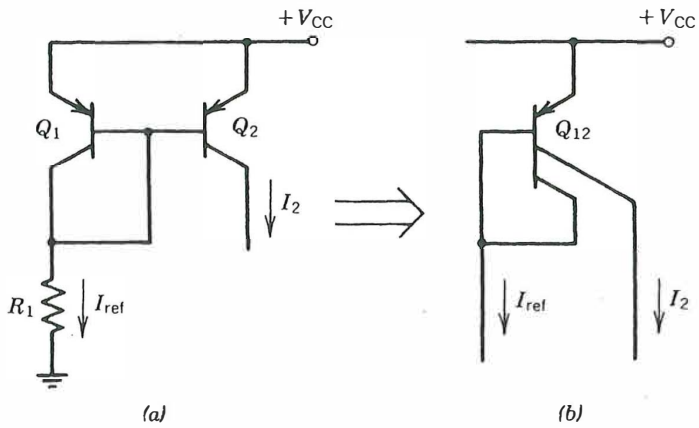


FIGURE 4.14. Basic *pnp* current mirror; (*a*) Circuit diagram; (*b*) actual implementation using split-collector lateral *pnp* transistor.

common base region, their combination can be replaced by a split-collector *pnp* transistor Q_{12}, as shown in Figure 4.14*b*, which has the device layout of Figure 2.27*a*.

As discussed earlier in Chapter 2, the collector currents in multiple-collector lateral *pnp* transistor split in proportion to the collector *periphery* facing the emitter. This allows the designer to control the ratio of the reference current I_{ref} to the output current I_2 by controlling the periphery ratio of the two collectors,

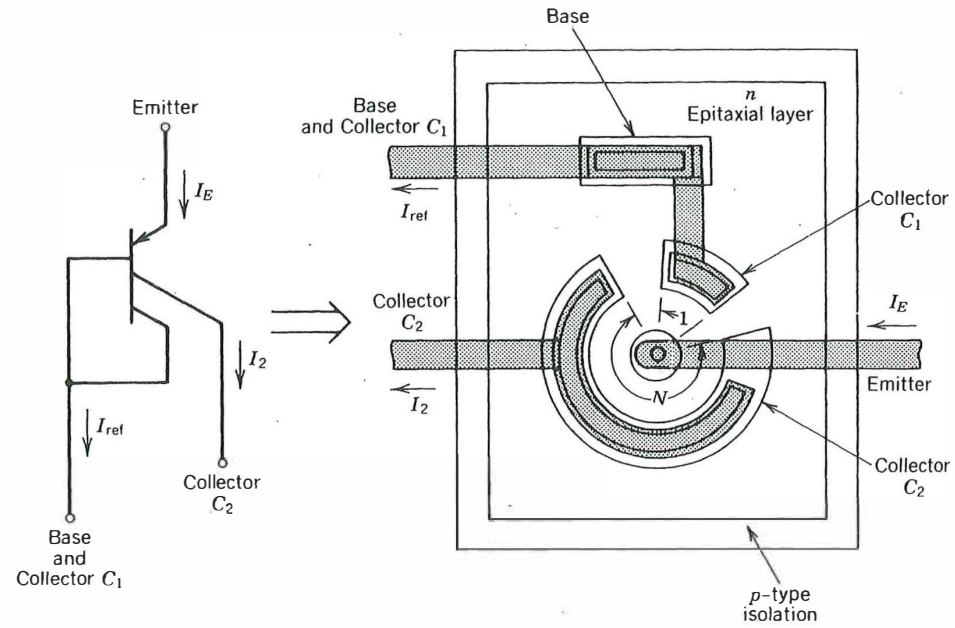


FIGURE 4.15. Current scaling by controlling the collector periphery ratio of a split-collector lateral *pnp* transistor.

EX-1014 0143
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

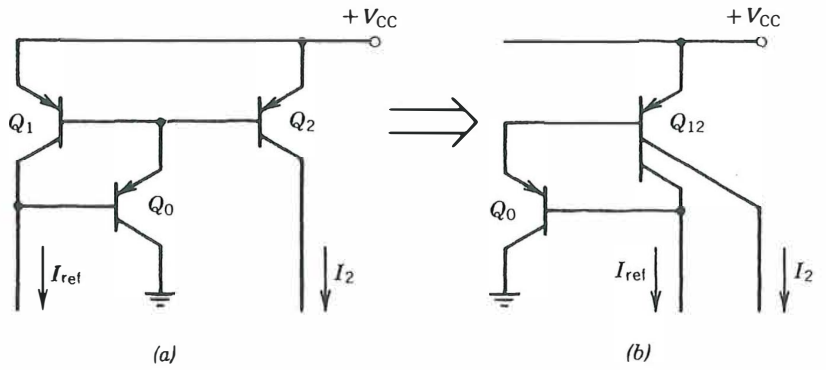


FIGURE 4.16. Base current compensated *pnp* current mirror: (*a*) Circuit diagram; (*b*) actual implementation.

as shown in Figure 4.15. Assuming that the two collectors C_1 and C_2 have a periphery ratio of $1:N$ between them, then by following the basic current mirror equation [Eq. (4.6)], one can relate the output current to the reference current as

$$I_2 = NI_{\text{ref}}\left(1 - \frac{1+N}{\beta}\right) \tag{4.38}$$

It should be noted that since the *pnp* transistor β is relatively low, the error term in parentheses in Eq. (4.38) can be significant for large values of N. For example, for $N = 5$ and $\beta \approx 20$, the output current I_2 may be as much as 50% below its ideal value of I_{ref}.

The current mismatches due to low β can be reduced significantly by using the *pnp* equivalent of the base current compensated current mirror shown in Figure 4.16. In this case, the compensating transistor Q_0 is normally designed as a substrate *pnp* transistor (see Fig. 2.29) with Q_1 and Q_2 made up of the split-collector *pnp* transistor Q_{12}, as shown in Figure 4.16*b*.

Figure 4.17 shows the other *pnp* current mirror configurations which are direct derivatives of the *npn* current mirrors discussed in the previous section. Another convenient *pnp* current mirror configuration is the multiple-output Wilson current source configuration shown in Figure 4.18, which can be designed with only two split-collector *pnp* transistors.

One of the major limitations of *pnp* current sources is the limited current-handling capability of the lateral *pnp* transistor. This can be solved by connecting an *npn* transistor to the output port to form a composite *pnp* transistor, similar to that shown in Figure 2.28. Such a transistor basically has the current-handling capability of the *npn* transistor and the polarity of the *pnp* transistor. If a controlled-β *npn* transistor (see Fig. 4.3) is used in place of the normal *npn* transistor, the output current level can be accurately controlled. Figure 4.19 demonstrates a circuit configuration which utilizes these principles. In the circuit, Q_1 and Q_2 from the basic current mirror. The collector current I_2 of

EX-1014 0144
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

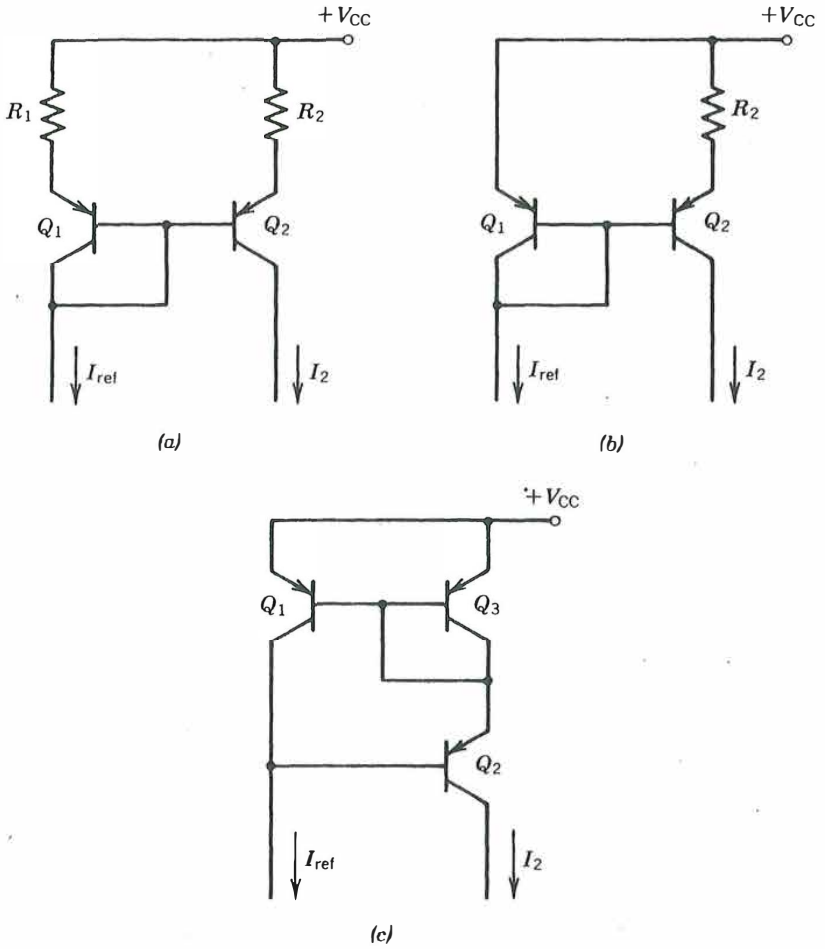


FIGURE 4.17. Other *pnp* current mirror configurations: (*a*) Resistor-biased current mirror; (*b*) low-current current mirror; (*c*) Wilson current mirror.

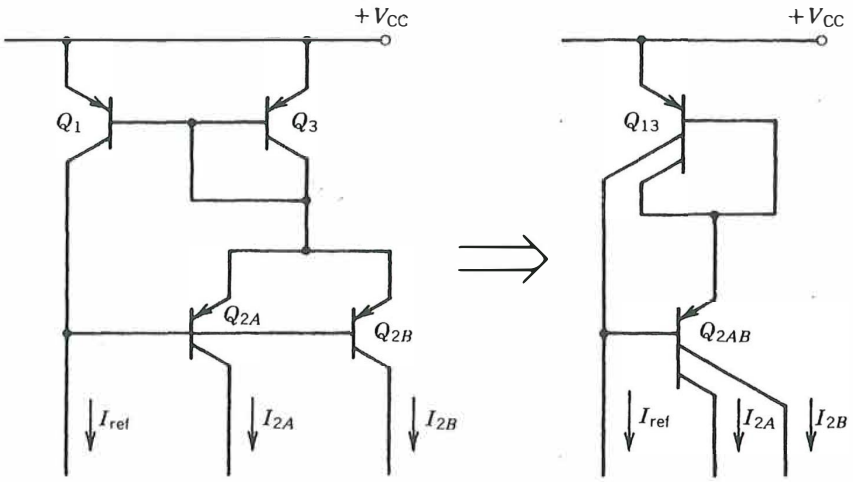


FIGURE 4.18. Multiple-output Wilson current source using split-collector *pnp* transistors. (*Note:* For symmetrical split collectors, $I_{2A} = I_{2B} = I_{ref}/2$.)

EX-1014 0145
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

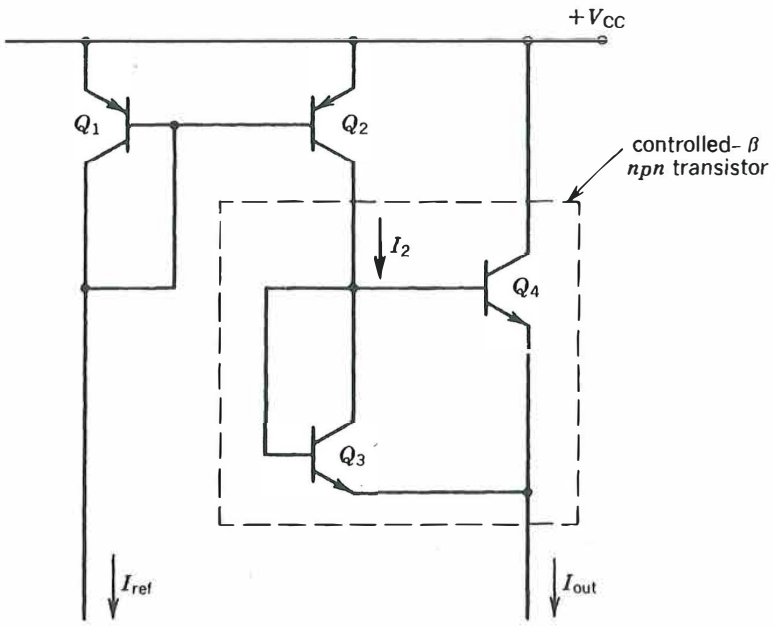


FIGURE 4.19. High-current *pnp* current source with composite controlled-β *npn* transistor.

transistor Q_2 is further amplified by the controlled-β *npn* transistor made up of Q_3 and Q_4.

Assuming that Q_3 and Q_4, have respective emitter areas A_3 and A_4, the output current I_{out} is related to I_2 as

$$I_{out} = I_2\left(1 + \frac{A_4}{A_3}\right) \tag{4.39}$$

Since I_2 is related to I_{ref} by the basic current mirror equation [Eq. (4.6)] for the matched pair of *pnp* transistors Q_1 and Q_2, the overall current transfer equation becomes

$$\frac{I_{out}}{I_{ref}} = \left(1 - \frac{2}{\beta_P}\right)\left(1 + \frac{A_4}{A_3}\right) \approx 1 + \frac{A_4}{A_3} \tag{4.40}$$

where β_P is the current gain of the *pnp* transistors Q_1 or Q_2. In this manner, accurate current ratios can be maintained while providing a high output drive current.

4.3. VOLTAGE-CONTROLLED CURRENT SOURCES

All of the *npn* or *pnp* current mirror circuits discussed so far can be used in practice as voltage-controlled current sources simply by varying the voltage applied across the resistor R_1 which determines the reference current I_{ref}. For example, in Figure 4.1, if the supply voltage V_{CC} was replaced by the control

EX-1014 0146
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

voltage V_C, then the output current I_2 can be related to V_C as

$$I_2 \approx I_{\text{ref}} = \frac{V_C - V_{\text{BE}}}{R_1} \tag{4.41}$$

where V_{BE} is the base–emitter drop of the diode connected transistor Q_1 of Figure 4.1.

In practice, the control characteristics given by Eq. (4.41) are not very useful because of the dependence of the output current on V_{BE} as well as V_C. This V_{BE} dependence can introduce a very strong temperature drift to the output current, particularly at low values of V_C, approaching V_{BE}. This effect is even more pronounced in the base current compensated current source of Figure 4.5 or in the Wilson current mirror of Figure 4.10, where the collector of the current-setting transistor Q_1 is clamped at not one but two V_{BE} above ground.

Figure 4.20*a* shows a simple two-transistor voltage-controlled constant-current stage which overcomes this problem. In the circuit, the V_{BE} dependence of the output current is virtually eliminated by effectively canceling the V_{BE} drops of Q_1 and Q_2. Assuming that the V_{BE1} and V_{BE2} associated with Q_1 and Q_2 are approximately equal, the voltage level V_A at the emitter of Q_2 can be written as

$$V_A = V_C - V_{\text{BE1}} + V_{\text{BE2}} \approx V_C \tag{4.42}$$

where V_{BE1} and V_{BE2} are the base–emitter drops of the corresponding transistors. For convenience, in Figure 4.20 V_A and V_C are measured with reference to the positive supply, rather than ground. Assuming that the base current of Q_1 is

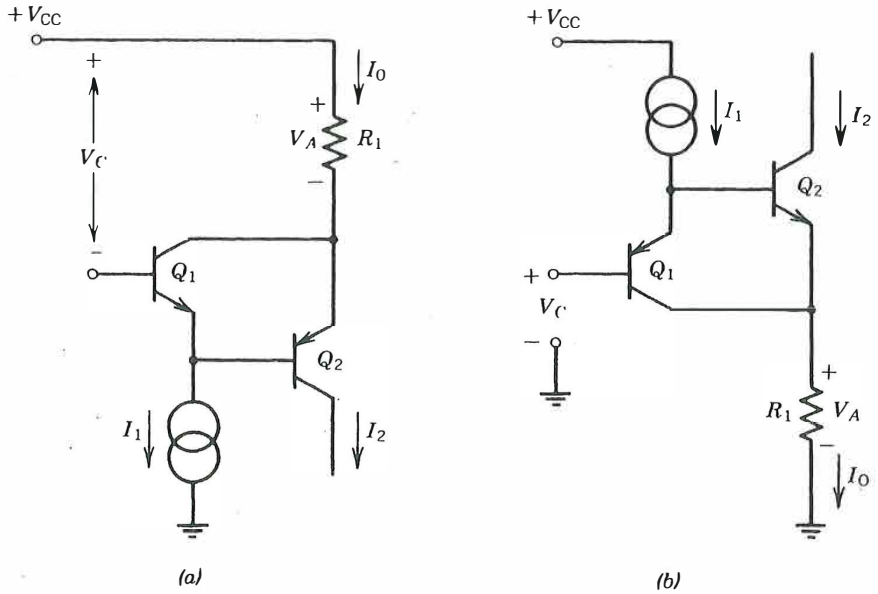


FIGURE 4.20. Voltage-controlled current source stages.

EX-1014 0147
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

negligible, the output current I_2 is related to the other two branch currents as

$$I_2 = I_0 - I_1 \tag{4.43}$$

However, since I_0 is set by the voltage drop across R_1 as

$$I_0 = \frac{V_A}{R_1} \approx \frac{V_C}{R_1} \tag{4.44}$$

the output current then becomes proportional to the control voltage as

$$I_2 = \frac{V_C}{R_1} - I_1 \tag{4.45}$$

where I_1 is a constant bias current.

Figure 4.20*b* shows the complementary equivalent of the same voltage-controlled constant-current stage. However, in this case, the input stage is a *pnp* transistor, and care must be taken in neglecting its base current. The voltage-controlled current source circuits of Figure 4.20 are useful building blocks in the design of voltage-controlled oscillators or automatic gain control circuits.

4.4. SUPPLY-INDEPENDENT BIASING

The constant-current stages described in Sections 4.1 and 4.2 require the generation of a reference current I_{ref}, which is then reproduced at the output. As shown in the case of the simple current mirror of Figure 4.1, I_{ref} is normally derived by connecting a current-setting resistor R_1 directly to the supply voltage. This results in a supply dependence of the output current I_2 given as

$$I_2 = I_{ref} = \frac{V_{CC} - V_{BE}}{R_1} \approx \frac{V_{CC}}{R_1} \tag{4.46}$$

In many applications, this supply dependence is not desirable and alternate bias solutions must be found. In this section, some of these alternate approaches shall be examined.

Using V_{BE} as Reference

The supply dependence of output current can be greatly reduced by using the transistor base–emitter voltage V_{BE} as a reference to generate the output current. Figure 4.21 shows two such biasing schemes. In the circuit of Figure 4.21*a*, the diodes D_1 and D_2, are formed by diode-connected transistors. Assuming that the voltage drops across D_1 and D_2 are each equal to V_{BE}, the base voltage of Q_2 is fixed at $2V_{BE}$, and the voltage drop across R_2 is equal to V_{BE}. Neglecting the base current of Q_2, this results in an output current

$$I_{out} \approx \frac{V_{BE}}{R_2} \tag{4.47}$$

EX-1014 0148
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

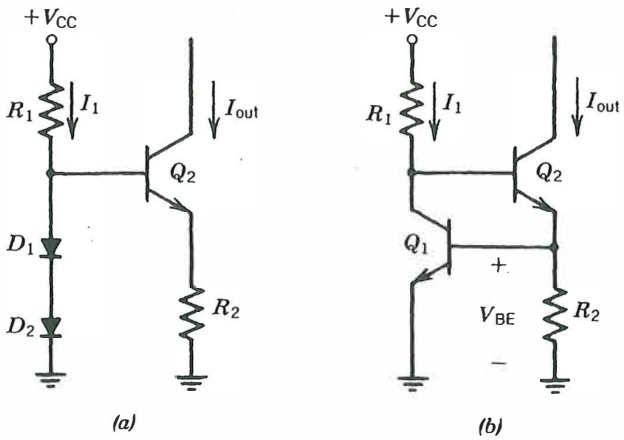


FIGURE 4.21. Supply-independent current bias sources using V_{BE} as reference.

The circuit of Figure 4.21*b* also works in a similar manner. The grounded-emitter transistor Q_1 forces a net voltage drop of V_{BE1} to appear across R_2. Neglecting the base current of Q_2, this again results in an output current expression similar to that of Eq. (4.47).

The bias circuits of Figure 4.21 are not completely independent of the supply voltage, since the current I_1, which sets the V_{BE} drops is dependent on the supply voltage. Using Eq. (2.5) of Chapter 2, which relates the V_{BE} drop to the current through the diode, one can write for the case of Figure 4.21, or Eq. (4.47),

$$V_{BE} = V_T \ln\left(\frac{I_1}{I_S}\right) \tag{4.48}$$

where I_S is the reverse saturation current of diodes D_1 and D_2 of Figure 4.21*a* or the transistor Q_1 of Figure 4.21*b*. The current I_1 is in turn set by the supply voltage as

$$I_1 = \frac{V_{CC} - 2V_{BE}}{R_1} \approx \frac{V_{CC}}{R_1} \tag{4.49}$$

Thus, taking into account the current dependence of V_{BE}, Eq. (4.47) can be written as

$$I_{out} \approx \frac{V_T}{R_2} \ln\left(\frac{V_{CC}}{R_1 I_S}\right) \tag{4.50}$$

which shows a logarithmic dependence on the supply voltage.

A common problem associated with the V_{BE}-referenced current sources is the strong temperature dependence of V_{BE}, which is of the order of -3000 ppm/°C. If the resistor R_2 of Figure 4.21 is a base-diffused resistor, it would introduce an additional temperature coefficient of $\approx +2000$ ppm/°C into the denominator of Eq. (4.47) which would result in a total temperature coefficient of I_{out} on the order of -5000 ppm/°C. This would result in approximately 50% reduction in the value of I_{out} over a 100°C temperature change.

EX-1014 0149
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

An alternate approach to supply insensitive biasing is the low-current current source of Figure 4.8, discussed earlier. As indicated by Eq. (4.30), this circuit also exhibits a logarithmic dependence on the supply voltage, where

$$I_2 \approx \frac{V_T}{R_2} \ln\left(\frac{V_{CC}}{R_1 I_2}\right) \tag{4.51}$$

However, since $V_T (= kT/q)$ exhibits a positive temperature coefficient of approximately +3000 ppm/°C at room temperature, its temperature drift can be compensated, in part, by the temperature coefficient of R_2. Assuming that R_2 is a diffused resistor with a temperature coefficient of approximately +2000 ppm/°C, the resulting temperature coefficient of I_2 in Eq. (4.51) is on the order of +1000 ppm/°C. This is a significant improvement over the simple V_{BE}-referenced bias circuits of Figure 4.21.

Self-Biasing References

The power supply dependence can be greatly improved by using so-called *self-biased* or *boot-strapped* current references. Instead of generating the reference current by connecting a resistor to the supply voltage, the reference current can be made to depend on the output current of the current source. This is done by sensing the output current by means of a current mirror and forcing the reference current to be equal to it. Figure 4.22 shows the application of this technique to the V_{BE}-referenced current source shown in Fig. 4.21*b*.

Assuming that the effects of finite base currents and the finite output conductances are negligible, the operation of the circuit can be analyzed as follows: Q_1,

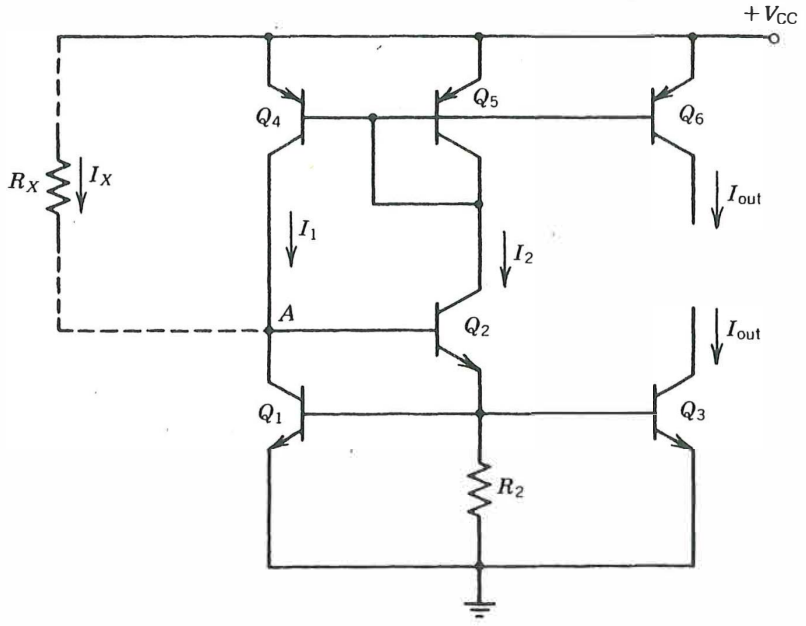


FIGURE 4.22. Self-biasing V_{BE} reference.[1]

EX-1014 0150
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Q_2, and R_2 function as described in Figure 4.21*b*, and set the collector current I_2 of Q_2. This current is then sensed by the current mirror transistors Q_5 and Q_4, and the reference current I_1 is forced to equal I_2. The operating point of the circuit is determined by two constraints: (1) I_2 must show a logarithmic dependence on I_1 as given by Eqs. (4.47) and (4.48). (2) I_2 must equal I_1 as dictated by the current mirror transistors Q_4 and Q_5. The operating point of the circuit which satisfies these conditions can be graphically determined as the intersection of the two curves shown in Figure 4.23. The intersection of the two curves, point D, determines the desired operation of the circuit. Once the current level I_2 is set, it can be reflected to the outputs of the circuit as the output current I_{out} through the *pnp* or the *npn* current mirrors, as shown in Figure 4.22.

In most of the self-biased circuits of the type described, there are not one but *two* stable operating points. The second operating point, which is not desirable, normally occurs at zero current state. In other words, the circuit will be perfectly stable if all the currents are zero. This can be seen from the graphic solution of Figure 4.23, as well as by examining the circuit of Figure 4.22. For example, if Q_1 was nonconducting, its V_{BE} would be zero, thus the voltage across R_2 would be zero, resulting in $I_2 = 0$, which would cause I_1 to be zero. This problem, which is common to self-biased circuits using internal feedback, is often referred to as the *startup problem*. This problem can be avoided by eliminating the undesired zero-current state. In its simplest form, this can be done by supplying a very small amount of startup current I_X to node A in Figure 4.22 by means of a high-value resistor R_X, connected to the supply voltage, as shown by dashed lines. If the startup current is chosen such that $I_X \ll I_1$, it has a negligible effect on circuit operation.

The circuit of Figure 4.22 is shown in its simplest form to illustrate the principle of self-biasing circuits. If desired, the output of the circuit can be greatly improved by using Wilson current mirrors or cascoded current mirrors. Many varieties of self-biased references, or their startup circuits, have been developed. These are well covered in the literature.[4]

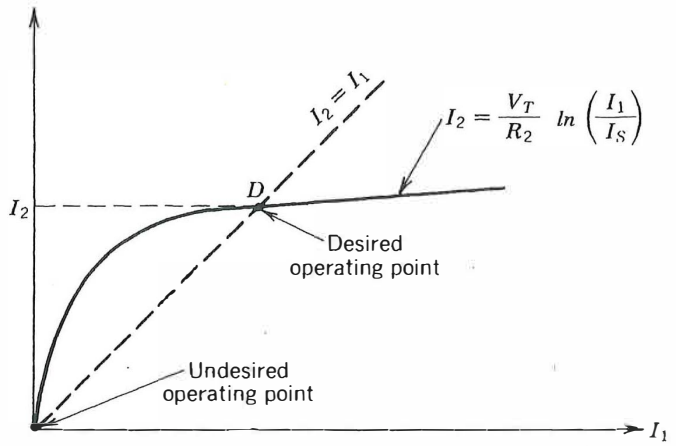


FIGURE 4.23. Graphic solution for operating point of circuit of Figure 4.22.

EX-1014 0151
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

It must be emphasized once more that when designing self-biased circuits it is *extremely* important to examine the possible operating points of the circuit to avoid startup instability problems.

4.5. VOLTAGE SOURCES

In a variety of circuit applications, it is necessary to establish a low-impedance point within the circuit which can serve as an internal voltage supply. Ideally, such a voltage reference point is required to have both a very low ac impedance and a very stable dc voltage level which is insensitive to power supply and temperature variations. In most applications, however, only one of these two requirements, that is, either the low impedance or the dc voltage stability, is of prime importance. The circuits which primarily fulfill the low-impedance requirements are known as voltage sources, whereas those specifically designed to provide a constant voltage, independent of the supply or the temperature changes, are called voltage references.

The voltage source stages are normally used to provide independent bias levels within the circuit. In such an application, the low ac impedance of the voltage source is necessary to buffer or decouple adjacent gain stages. An example of such an application is the use of voltage source stages to form a common bias point for the inputs of a differential gain stage. Figure 4.24 shows some of the practical voltage source configurations often used in IC design. In the circuit of Figure 4.24*a*, the low output impedance of an emitter-follower stage is used to simulate a low-impedance voltage source with an output voltage

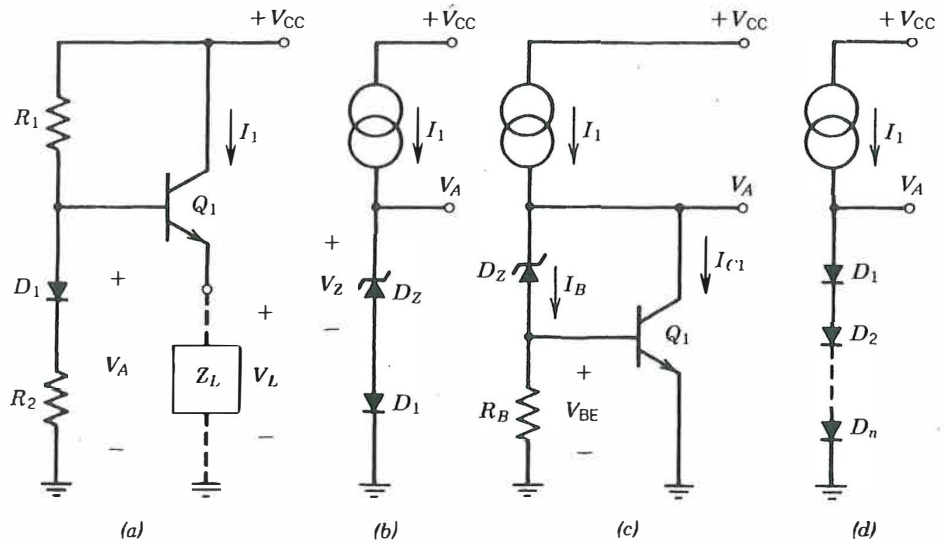


FIGURE 4.24. Practical voltage source configurations: (*a*) Common-collector stage; (*b*) temperature-compensated Zener diode; (*c*) temperature-compensated Zener diode with buffered output; (*d*) diode string.

EX-1014 0152
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

level V_L given as

$$V_L = (V_{CC} - V_{BE})\frac{R_2}{R_1 + R_2} \approx V_{CC}\frac{R_2}{R_1 + R_2} \tag{4.52}$$

Note that in this configuration the diode D_1 in the bias string is used to partially offset the dc value and the temperature dependence of the V_{BE} drop across Q_1. The load Z_L represents the rest of the circuitry biased by the current through Q_1. Using the hybrd-π model for the transistor (see Fig. 2.10), the resistance level R_O looking into the emitter of Q_1 can be expressed as

$$R_O \approx \frac{V_T}{I_1} + \frac{R_1 R_2}{\beta(R_1 + R_2)} \tag{4.53}$$

Due to the resistive bias string in Figure 4.24*a*, the values of the bias voltage V_A and the output voltage V_L are both dependent on the supply voltage V_{CC}. This dependence can be avoided using the bias circuits of Figure 4.24*b*, *c*, or *d*. In each of these circuits, the impedance level looking into the bias terminal is low enough for most applications to eliminate the need for an additional emitter-follower stage. In many applications, the current source I_1 may also be replaced by a resistor connected between $+V_{CC}$ and V_A.

The circuits of Figure 4.24*b* and *c* both provide an output voltage level

$$V_A = V_Z + V_{BE} \tag{4.54}$$

where V_Z is the breakdown voltage of the Zener diode. The forward-biased diode D_1 of Figure 4.24*b* provides partial compensation for the positive temperature coefficient of V_Z. In a monolithic structure, the combination of D_Z and D_1 can be conveniently designed as a single transistor structure with two separate emitters, as shown in Figure 3.10. The impedance level measured looking into the output terminal is

$$R_{out} = R_Z + \frac{V_T}{I_1} \tag{4.55}$$

where R_Z is the dynamic resistance of the base–emitter breakdown diode. For typical integrated device structures, R_Z is in the range of 40–100 Ω.

The circuit of Figure 4.24*c* also provides an output voltage level equal to $V_Z + V_{BE}$ by impressing a voltage drop equal to V_{BE} across the bias resistor R_B. In this case, Q_1 serves as an active gain stage and automatically adjusts its current I_{C1} to maintain the output voltage level constant. The output impedance of the circuit is approximately the same as that given by Eq. (4.55). However, since the current I_B through the Zener diode is maintained constant, independent of supply voltage changes, the sensitivity of V_A to supply voltage variations is greatly reduced.

Since the base–emitter breakdown voltage is set by the IC fabrication process, the values of V_A available from the circuits of Figure 4.24*b* and *c* are restricted to be within the 6.5–8-V range.

EX-1014 0153
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Figure 4.24*d* shows how a number of diodes, or diode-connected transistors, can be cascaded in series to simulate a low-impedance output voltage V_A, where

$$V_A = nV_{BE} \tag{4.56}$$

and an output resistance level of

$$R_{out} = \frac{nV_T}{I_1} \tag{4.57}$$

where n is the number of diodes in the string. Such a voltage source has a strong negative temperature coefficient, given as

$$\frac{\partial V_A}{\partial T} = n\frac{\partial V_{BE}}{\partial T} \approx -2n \text{ mV/°C} \tag{4.58}$$

Since each diode in the string requires a separate collector pocket, a string involving large numbers of diodes may occupy a significant portion of the chip area. In many applications, an alternate solution to the diode string is the so-called V_{BE}-*multiplier* circuit shown in Figure 4.25. Such a circuit can produce an output voltage that is an arbitrary multiple of the transistor base–emitter drop. In the circuit of Figure 4.25*a*, the voltage drop across R_2 is constrained to be equal to the transistor V_{BE}. Assuming that the base current of Q_1 is negligible, the current through R_2 is the same as that through R_1. Therefore, the output dc level can readily be related to the transistor V_{BE} as

$$V_A = I_2(R_1 + R_2) = V_{BE}\left(1 + \frac{R_1}{R_2}\right) \tag{4.59}$$

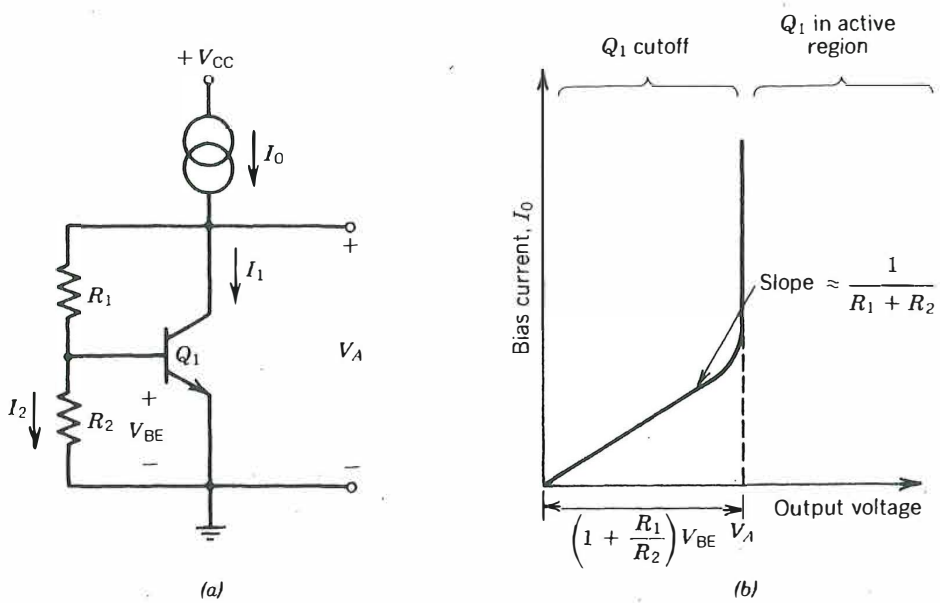


FIGURE 4.25. The V_{BE}-multiplier circuit: (*a*) Circuit connection; (*b*) output level as a function of bias current I_0.

EX-1014 0154
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Note that, as long as $I_0R_2 > V_{BE}$, the transistor is in its active region, and due to the shunt feedback provided by R_1, the transistor current I_1 automatically adjusts itself to maintain I_2 and V_A relatively independent of the supply voltage or current. The corresponding current–voltage characteristics of the V_{BE}-multiplier circuit are shown in Figure 4.25*b*.

In most applications, the circuit of Figure 4.25*a* provides a convenient substitute for a diode string, particularly when a large number of diodes or a noninteger multiple of V_{BE} are required. Using the hybrid-π model, the output resistance of the V_{BE} multiplier can be readily calculated as

$$R_O = \frac{R_1}{\beta} + \frac{R_1 + R_2}{g_m R_2} \tag{4.60}$$

where $g_m = (= I_1/V_T)$ is the transconductance of Q_1. For most applications, the value of R_O is in the range of 50–200 Ω.

Figure 4.26 shows a modified version of the basic shunt feedback circuit which is useful for obtaining high-value voltage sources without requiring high-voltage Zener diodes. Assuming that D_Z is the reverse-biased base–emitter diode with a break-down voltage V_Z, the voltage level at the output terminal can be expressed as

$$V_A = (V_Z + V_{BE})\left(1 + \frac{R_1}{R_2}\right) \tag{4.61}$$

provided that $I_0R_2 > V_Z + V_{BE}$ so that Q_1 is in its active region. The circuit of Figure 4.26 is particularly useful for high-voltage integrated circuits, where it can be used as a high-value voltage source, or it can be substituted for a high-voltage avalanche diode, for overvoltage protection in circuits prone to high-voltage transients.

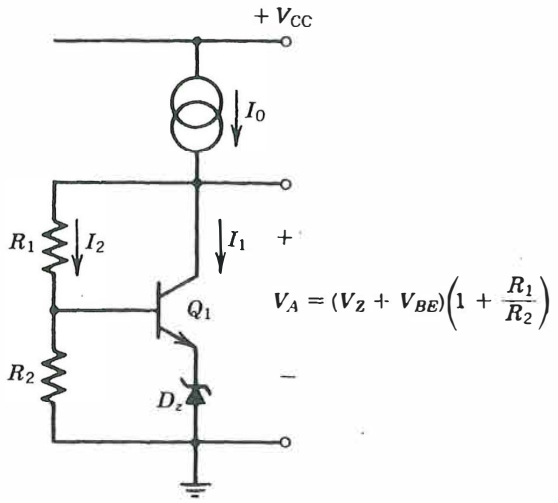


FIGURE 4.26. Circuit for simulating high-voltage breakdown diode.

EX-1014 0155
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

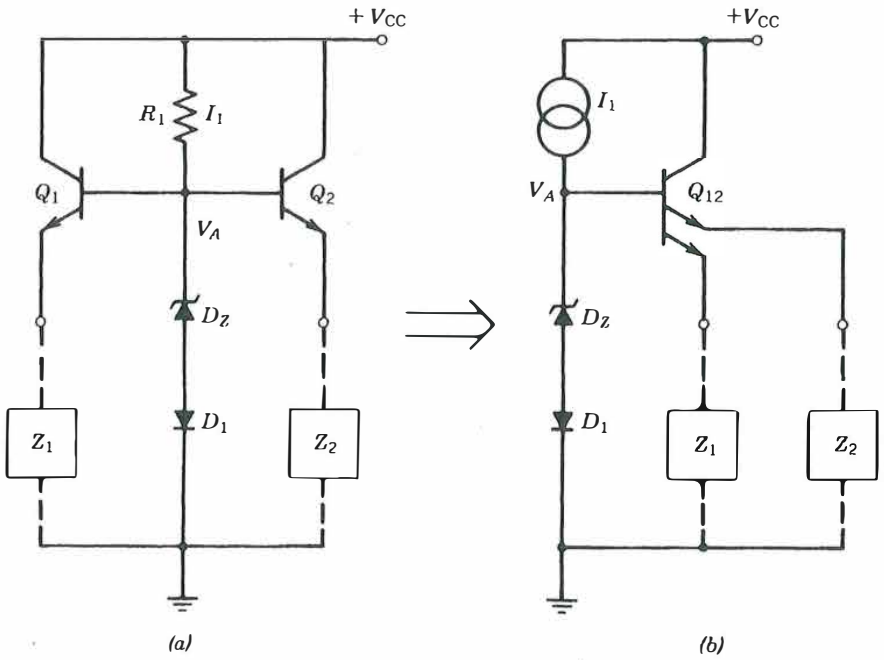


FIGURE 4.27. Multiple voltage sources biased from same reference voltage: (*a*) Discrete design; (*b*) its monolithic counterpart using multiple-emitter *npn* transistors.

In providing low-impedance bias points within the circuit, the emitter-follower stage is the most commonly used buffer circuit. Since many emitter-follower stages can share the same isolation pocket, they require a minimum amount of chip area. In some analog circuit applications, it is necessary to provide multiple voltage sources which are biased from the same reference voltage, but are buffered from each other such that the ac signals in one source would be relatively isolated from those in the other. In conventional circuit design using discrete devices, this can be done using two separate emitter-follower stages, as shown in Figure 4.27*a*. However, using the design and layout advantages of monolithic circuits, the same circuit can be designed using a multiple-emitter transistor (see Fig. 2.21), as shown in Figure 4.27*b*, with a minimum increase in chip area.

4.6. DC LEVEL-SHIFT STAGES

Since large-value coupling capacitors are not available in monolithic circuits, all broadband gain stages need to be dc coupled. This means that the output dc level of a gain stage should be compatible with the dc level at the input of the next stage. In an *npn* common-emitter gain stage, the output dc level is always higher than the dc level of the input. Therefore, if a number of such gain stages are cascaded, the output dc level rapidly builds up toward the positive supply

EX-1014 0156
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

voltage. This in turn limits the amplitude and the linearity of the available output swing. Ideally, such a dc level buildup can be avoided by using complementary *pnp-npn* gain stages. However, the *pnp* transistors available in monolithic form have relatively poor frequency response and current gain characteristics.

If an analog integrated circuit is comprised of a cascade of *npn* gain stages, this positive dc level buildup can be overcome by using a level-shift stage between each gain stage to shift the output dc level toward the negative supply with minimum attenuation of the ac signal. In general, such a stage also serves as a unilateral buffer between successive gain stages. Therefore, it is required to have a high input impedance and a relatively low output impedance to prevent interstage loading. Figure 4.28 shows some practical dc level-shift stages for monolithic circuit applications. In each case, a common-collector stage is used at the input to avoid loading the output of the gain stage connected to V_1.

The resistive level-shift stage of Figure 4.28*a* provides a simple means of shifting the input level V_1 to a more negative dc level V_2, where

$$V_2 = (V_1 - V_{BE})\frac{R_2}{R_2 + R_1} \tag{4.62}$$

The main drawback of the resistive level-shift stage is attenuation of the ac signal along with the dc level shift. The ac voltage gain A_v for the resistive level-shift

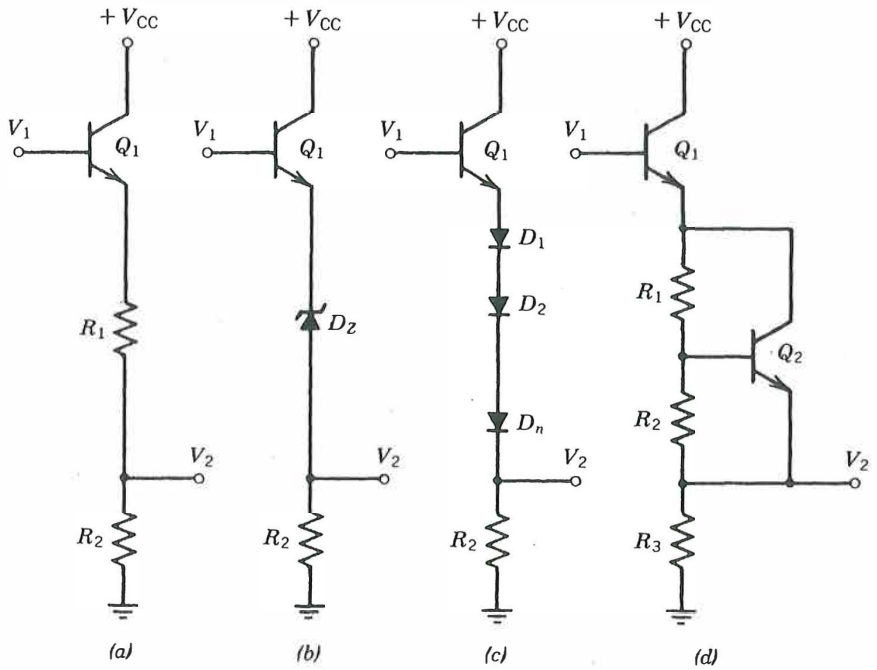


FIGURE 4.28. Some practical dc level-shift stages: (*a*) Resistive; (*b*) Zener diode; (*c*) diode string; (*d*) V_{BE} multiplier.

EX-1014 0157
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

stage is less than unity, and is given as

$$A_V = \frac{R_2}{R_1 + R_2} < 1 \tag{4.63}$$

Thus, as the value of R_2 is decreased to improved the net dc level shift from V_1 to V_2, the ac gain of the stage deteriorates rapidly. The output impedance of such a stage is also relatively high, being equal to the shunt combination of R_1 and R_2.

The Zener diode level-shift stage of Figure 4.28*b* provides an alternate means of shifting the dc level by the amount

$$V_1 - V_2 = V_{BE} + V_Z \tag{4.64}$$

where V_Z is the breakdown voltage of the Zener diode D_Z. The reverse breakdown characteristic of the base–emitter junction is used to form the Zener diode (i.e., $V_Z \approx 6$–9 V). If the bulk resistance of D_Z is negligible compared to R_2, the voltage gain for the stage is approximately unity. The two main disadvantages of the Zener diode level-shift stage are the limitations on the values of V_2 available in integrated circuits and the excess noise generated by the breakdown diode D_Z. Therefore, such a level-shift scheme is not suitable for low-level ac signals or for circuits operating with relatively low supply voltages, such as $V_{CC} < 10$ V.

The diode string level-shift stage of Figure 4.28*c* provides a net dc level shift $V_1 - V_2$ given as

$$V_1 - V_2 = (n + 1)V_{BE} \tag{4.65}$$

where n is the number of diodes in the string. Normally, the diodes D_1 through D_n would be formed by diode-connected transistors. Assuming that the dynamic impedance of the diodes is negligible compared to R_2, the voltage gain of the stage is approximately unity. The output impedance R_O for the circuit is quite low, given as

$$R_O \approx (n + 1)\frac{V_T}{I_1} \tag{4.66}$$

where V_T is the thermal voltage. The diode string level-shift stage has two disadvantages: (1) Since each diode requires a separate isolation pocket, such a stage may take up a sizable chip area and have appreciable shunt capacitance to the substrate. (2) The output dc level shows a strong temperature dependence due to the change of the diode voltage V_{BE} with temperature.

The circuit of Figure 4.28*d* is often used as a substitute for the diode string level-shift stage. In this circuit, the diode chain is replaced by the V_{BE}-multiplier stage described in Figure 4.25. The net dc level change across the stage is

$$V_1 - V_2 = V_{BE}\left(2 + \frac{R_1}{R_2}\right) \tag{4.67}$$

EX-1014 0158
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Assuming that $\beta >> 1$, the net voltage gain across the stage can be expressed as

$$A_V \approx \frac{R_3 R_2 g_m}{1 + R_3 R_2 g_m} \approx 1.0 \tag{4.68}$$

where g_m is the transconductance of Q_2. Similarly, the output resistance R_O for the stage can be calculated to be

$$R_O = R_3 \left\| \frac{R_1}{R_2 g_m} \right. \tag{4.69}$$

The main disadvantages of the level-shift circuits of Figure 4.28*c* and *d* are the temperature dependence of V_2 due to the V_{BE} change with temperature.

Another commonly used level-shift stage is the resistor and current source combination shown in Figure 4.29. Since the dynamic output resistance of the current source is much higher than R_1, the voltage gain of the stage is very close to unity. However, the output dc level is shifted toward negative supply (or ground) by the amount

$$V_1 - V_2 = 2V_{BE} + I_0 R_1 \tag{4.70}$$

This type of level-shift stage is susceptible to capacitive loading at node A in Figure 4.29, which would limit its frequency response. This capacitive loading is mainly due to the collector–substrate capacitance of the *npn* transistor which forms the current source I_0. Therefore, in the layout of the circuit, care must be taken to minimize the collector area of this transistor.

pnp Level-Shift Stages

In spite of its relatively poor gain and frequency response, lateral *pnp* transistors still find a number of applications as gain or level-shift stages. In general, their

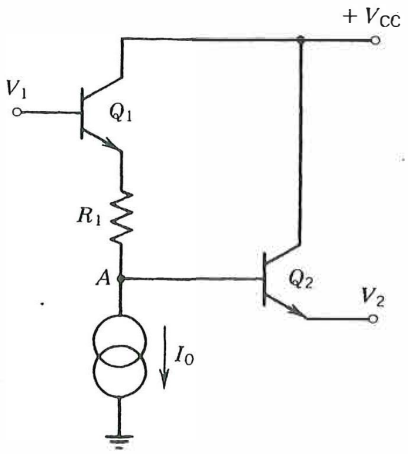


FIGURE 4.29. Level-shift stage using resistor and current source combination.

EX-1014 0159
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

applications are limited to low-frequency circuits, where the low f_T of the lateral *pnp* transistor ($\approx$5 MHz) does not present a problem.

The low-gain characteristics of the lateral *pnp* transistor can be improved by combining with an *npn* transistor to form a so-called *composite pnp* transistor, as described in Chapter 2 (see Fig. 2.28). Figure 4.30 shows the use of such a composite *pnp* transistor as both a level-shift and a gain stage.

Since the composite *pnp* transistor is a feedback circuit made up of two devices, it exhibits a two-pole rolloff in its frequency response and is prone to oscillations. In practical applications, this problem can be avoided by connecting a resistor R_X between the base and the emitter of Q_2 to reduce the overall current gain of the combined Q_1 and Q_2 and to have Q_1 operate at a slightly higher current level than just the base current of Q_2. As a rule of thumb, the current I_X through R_X is chosen to be $\approx$ 10% of I_1. This is done by choosing R_X such that

$$R_X = \frac{10V_{BE}}{I_1} \tag{4.71}$$

Normally, R_X is designed as a noncritical pinch resistor.

The output dc voltage of the composite *pnp* level-shift stage of Figure 4.30 is given as

$$V_2 = \frac{R_2}{R_1}(V_{CC} - V_{BE} - V_1) \tag{4.72}$$

Similarly, the small-signal ac voltage gain A_V of the circuit can be expressed as

$$A_V \approx -\frac{R_2}{R_1} \tag{4.73}$$

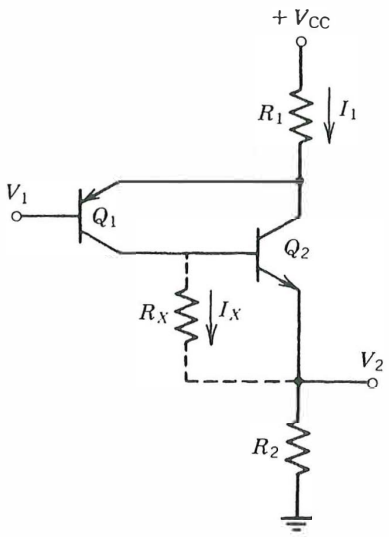


FIGURE 4.30. Level-shift stage using composite *pnp* configuration.

EX-1014 0160
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Figure 4.31 shows other commonly used level-shift stages comprised of *pnp-npn* transistors. The circuit of Figure 4.31*a* is derived from the voltage-controlled current source circuit described earlier (see Fig. 4.20). It provides an output voltage level V_2 given as

$$V_2 = I_2R_2 = \frac{R_2}{R_1}(V_{CC} - V_1 - I_1R_1) \tag{4.74}$$

where I_1 is a constant-current source. The ac gain can be evaluated by differentiating Eq. (4.74)

$$A_V \approx -\frac{R_2}{R_1} \tag{4.75}$$

In the circuit of Figure 4.31*b*, the lateral *pnp* transistor is operated in its common–base configuration in order to enhance its frequency capability. The output voltage level V_2 is equal to

$$V_2 = \alpha_p I_1 R_2 = \frac{\alpha_p R_2}{R_1}(V_1 - V_{bias} - 2V_{BE}) \tag{4.76}$$

where α_p is the common-base current gain of Q_2. The ac voltage gain of the circuit is

$$A_V = \alpha_p \frac{R_2}{R_1 + 2/g_m} \approx \alpha_p \frac{R_2}{R_1} \tag{4 . 77}$$

where $g_m (= I_1/V_T)$ is the transconductance of Q_1 or Q_2.

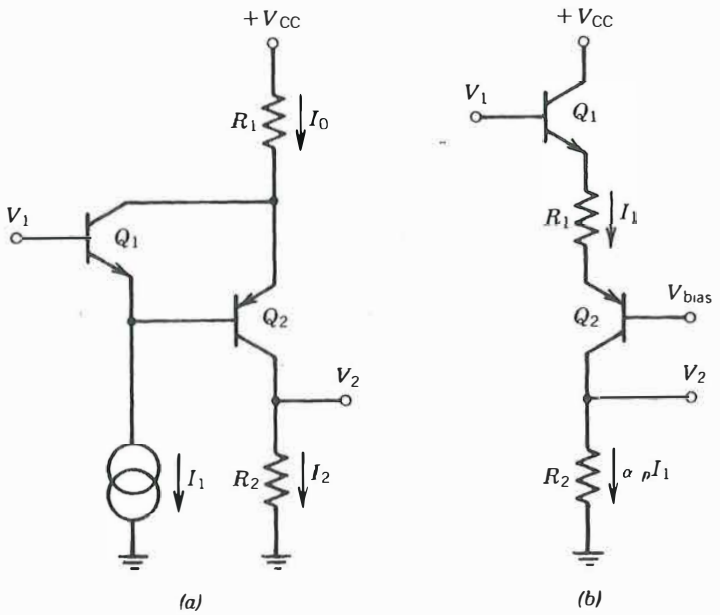


FIGURE 4.31. Other commonly used *pnp–npn* level-shift stages.

EX-1014 0161
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

The circuit of Figure 4.31*b* offers the best frequency performance among the *pnp–npn* level-shift stages discussed so far. However, it has two disadvantages: it requires a low-impedance bias source V_{bias}, and its gain and level-shift characteristics depend on α_p, which may vary between 0.8 and 0.98 in practical lateral *pnp* devices.

pnp current mirrors are also suitable for dc level shifting, provided that the frequency requirements are relatively low (i.e., $\leq$ 1 MHz). They are particularly suitable for level shifting in differential gain stages. Figure 4.32 shows their application in a different gain stage. The circuit of Figure 4.32*a* is a basic differential gain stage with a voltage gain of R_L/R_E, and with an output common-mode voltage $(V_{out})_{CM}$, close to the power supply, where

$$(V_{out})_{CM} = V_{CC} - \frac{I_0 R_L}{2} \tag{4.78}$$

In the level-shifted version of the same circuit (see Fig. 4.32*b*), the *pnp* current mirrors, made up of D_3, Q_3 and D_4, Q_4, reflect the output current toward ground (or negative supply) so that the voltage gain remains unchanged, but the output common-mode voltage now becomes

$$(V_{out})_{CM} \approx \frac{I_0 R_L}{2} \tag{4.79}$$

In the actual implementation of the design, D_3, Q_3 and D_4, Q_4 would each be designed as split-collector *pnp* transistors to conserve chip area.

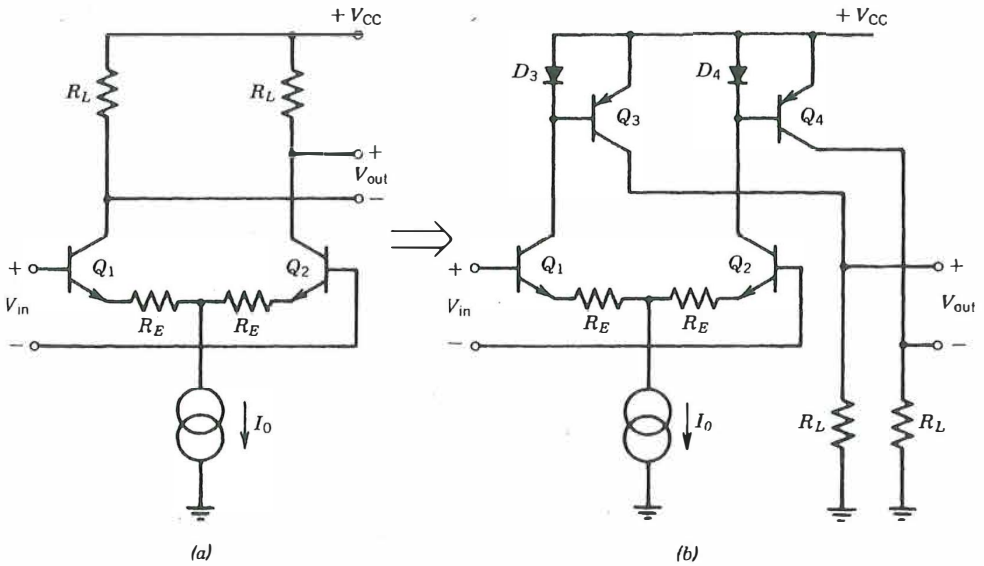


FIGURE 4.32. Differential level shifting using *pnp* current mirrors: (*a*) Basic differential stage; (*b*) its modified version with current mirror level shifting.

EX-1014 0162
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

4.7. TEMPERATURE-INDEPENDENT BIASING

In the design of various analog circuits, such as D/A converters, voltage regulators, or low-drift amplifiers, it is necessary to establish a temperature-independent bias reference within the circuit. This stable bias reference can be either a current or a voltage. In most applications, voltage rather than current references are preferred since they are easier to interface with the rest of the circuitry.

In the case of a voltage reference, unlike the case of voltage sources, the main emphasis is not on low output impedance but on the temperature stability of the voltage level. Temperature stability requirements on a voltage reference are typically $\leq$ 100 ppm/°C. In many cases, by careful design and compensation even this figure can be significantly improved to be within the range of 20–40 ppm/°C.

Typical temperature coefficients of monolithic components are in the range of a few thousand ppm/°C or more. However, by making use of the matching and tracking characteristics of monolithic components, and their close thermal coupling on the chip, it is possible to compensate the thermal drifts to a few parts per million. The basic principle of temperature compensation used in monolithic IC design is very simple. One starts with a *predictable* temperature drift, then finds another predictable temperature source of opposite polarity which can be scaled by a temperature-independent scale factor. Then, by proper circuit design, the effects of the two opposite-polarity drifts are made to cancel, resulting in a nominally zero temperature coefficient voltage level.

Among the monolithic components, there are three basic temperature drift sources which are reasonably predictable and repeatable:

1. The temperature dependence of the base–emitter drop V_{BE} which shows a strong negative temperature coefficient, typically on the order of -2 mV/°C.
2. The temperature dependence of the V_{BE} difference ΔV_{BE}, which is proportional to the absolute temperature, through the thermal voltage V_T [see Eq (4.28)] and, thus, exhibits a positive temperature coefficient.
3. The temperature drift of the base–emitter Zener diode V_Z, which is inherently low and positive in polarity (typically on the order of +200 to +500 ppm/°C.

In designing temperature-compensated bias references, one achieves the required compensation by scaling one or more of these drift sources and subtracting them from each other.

The temperature coefficient of the monolithic resistors (with the exception of thin-film resistors) is too high and too nonlinear to be used for any predictable temperature compensation. However, the resistor ratios show excellent tracking over temperature, with the temperature coefficient of the resistor ratio being on the order of ± 5–± 20 ppm/°C for well-matched resistors. Thus, monolithic

EX-1014 0163
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

resistors are suitable for generating temperature-insensitive scale factors, when used in a ratio rather than in an absolute-value form.

Figure 4.33*a* shows a simple voltage reference circuit which makes use of the opposite-polarity drift between the Zener voltage V_Z and the forward diode voltage V_{BE}.[5] The base–emitter avalanche diode D_Z is supplied by a constant current I_1 and provides a bias voltage V_Z with a positive temperature coefficient (typically $\approx +3$ mV/°C). The temperature dependence of the V_{BE} drop across Q_1 and D_1 results in a temperature coefficient of about +7 mV at the cathode of D_1. Similarly, the thermal variation of the voltage drop across D_2 creates a temperature coefficient of ≈ -2 mV/°C at the anode of D_2. Thus, by tapping the resistor string R_1 and R_2, connecting these two points of opposite temperature drift, a voltage reference V_{ref} can be made to have a nominally zero temperature coefficient. The voltage level of V_{ref} is given as

$$V_{ref} = \frac{R_2 V_Z + V_{BE}(R_1 - 2R_2)}{R_1 + R_2} \tag{4.80}$$

The temperature coefficient of V_{ref} can be nominally set to zero by setting the resistor ratio as

$$\frac{R_1 - 2R_2}{R_2} = -\frac{\partial V_Z / \partial T}{\partial V_{BE} / \partial T} \tag{4.81}$$

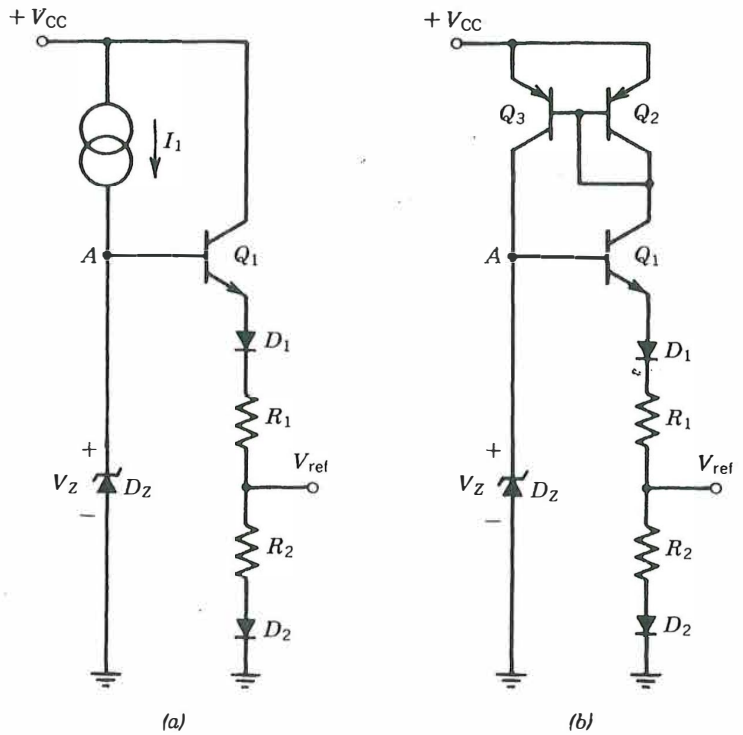


FIGURE 4.33. Simple voltage reference circuit: (*a*) Basic circuit; (*b*) its self-biased version.

EX-1014 0164
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

For typical values of $\partial V_Z/\partial T$ and $\partial V_{BE}/\partial T$ associated with IC components, the nominal value of V_{ref} for a zero temperature coefficient is in the range of 1.7–2.5 V. Figure 4.33*b* shows a self-biased version of the same circuit, which can generate the supply-independent constant current I_1 internally by means of the *pnp* current mirrors Q_2 and Q_3. A word of caution is in order, however, regarding the circuit of Figure 4.33*b*. It is a self-biased circuit; therefore, it may require startup circuitry which will inject an initial startup current to node A in the circuit to avoid the zero-current stable stage (see Figs. 4.22 and 4.23).

Stable voltage references with temperature coefficients on the order of ± 30 to ± 50 ppm/°C have been reported using the basic circuit technique shown in Figure 4.33.[5]

The basic disadvantages of the Zener-referenced bias circuits of the type shown in Figure 4.33 are that they require a relatively high value of power supply (typically $\geq$ 10 V) and introduce substantial noise into the circuit, due to the avalanche breakdown within the diode, as well as exhibiting some long-term drift of the Zener voltage V_Z. However, in the more recent designs, the long-term drift problems have been largely eliminated by using buried-Zener structures (see Fig. 3.11), which confine the breakdown to the subsurface region of silicon.

Band-Gap Reference Circuits

The high supply voltage and the Zener noise problems associated with the Zener-biased reference circuit can be avoided by using the so-called *band-gap* reference circuit. Such a reference circuit operates on the principle of compensating the negative temperature drift of V_{BE} with the positive temperature coefficient of the thermal voltage V_T. Its principle of operation is illustrated, symbolically, in Figure 4.34. First, we generate a known negative temperature drift due to V_{BE}; next, we produce a positive temperature drift due to the thermal voltage $V_T = kT/q$; and finally, we scale the latter with a constant (i.e., temperature-independent scale factor K) and subtract it from the former to obtain nominally zero temperature dependence. In the simplified model of Figure 4.34, the output voltage V_{out} is given as

$$V_{out} = V_{BE} + KV_T \tag{4.82}$$

Since each of the two terms in the above equation exhibits opposite-polarity temperature drifts, it should be possible, at least in theory, to make V_{out} nominally independent of temperature. This, in summary, is the principle of a band-gap reference. The temperature-stabilized output dc level, where $\partial V_{out}/\partial T$ is nominally equal to zero, comes about at an output voltage level on the order of +1.25 V. One can show mathematically that this voltage level is very nearly equal to the band-gap voltage of silicon.[6] The name *band-gap reference* is derived from this relationship.

Figure 4.35 shows a simple implementation of the band-gap reference concept. In this circuit, Q_1 and Q_2 operate as a low-current bias stage, as shown in

EX-1014 0165
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

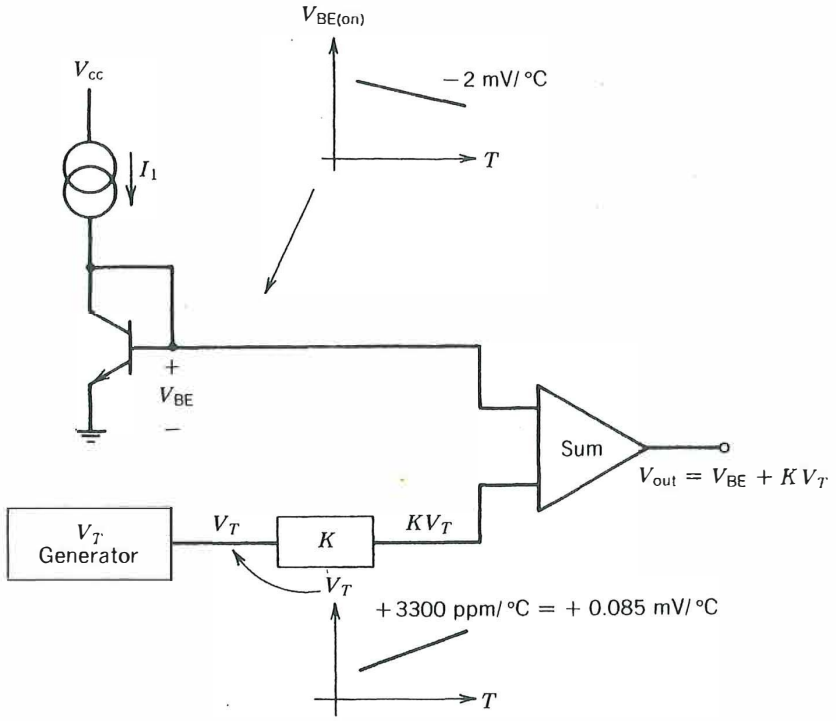


FIGURE 4.34. Symbolic model for illustrating the principle of operation of a band-gap reference.[1]

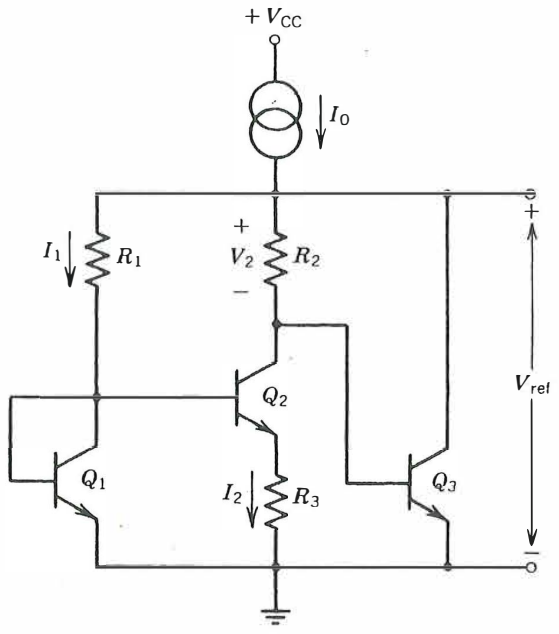


FIGURE 4.35. Simple band-gap reference circuit.

EX-1014 0166
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Figure 4.8. The voltage across R_3 is equal to the V_{BE} difference between Q_1 and Q_2, that is,

$$I_2R_3 = V_{BE1} - V_{BE2} = V_T \ln\left(\frac{I_1}{I_2}\right) = \Delta V_{BE} \tag{4.83}$$

Assuming $\beta >> 1$, the net voltage drop V_2 across R_2 can be written as

$$V_2 = \frac{R_2}{R_3}\Delta V_{BE} = V_T\frac{R_2}{R_3}\ln\left(\frac{I_1}{I_2}\right) \tag{4.84}$$

The output voltage V_{ref} is then equal to the base–emitter drop of Q_3 plus the voltage drop V_2, that is,

$$V_{ref} = V_{BE} + V_T\frac{R_2}{R_3}\ln\left(\frac{I_1}{I_2}\right) \tag{4.85}$$

Assuming that the ratio I_1/I_2 can be kept relatively insensitive to temperature, Eq. (4.85) is the same form as Eq. (4.82), and the first-order temperature dependence of V_{ref} can be reduced to zero. Practical voltage reference circuits with temperature coefficients in the range of 30–60 ppm/°C can be obtained in this manner.

The main drawback of the basic band-gap reference circuit of Figure 4.35 is the difficulty of maintaining the current ratio I_1/I_2 independent of temperature. Figure 4.36 shows two alternate circuit approaches which largely overcome this problem at the expense of added circuit complexity, that is, by using a high-gain operational amplifier in a feedback loop.

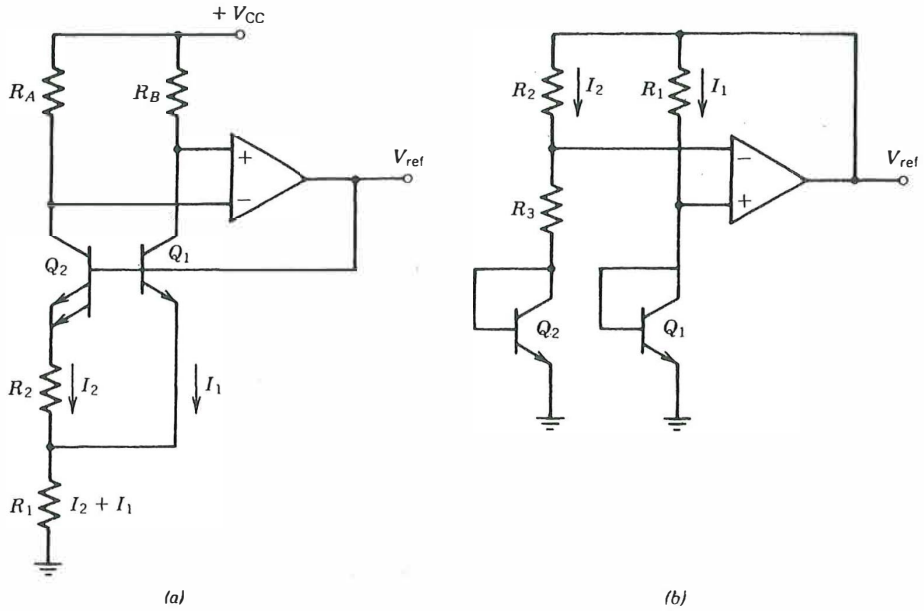


FIGURE 4.36. Alternate configurations for band-gap voltage reference circuits.

EX-1014 0167
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

In the circuit of Figure 4.36*a*, the emitter area of Q_2 is made to be n times that of Q_1. The output voltage can be expressed as

$$V_{\text{ref}} = V_{\text{BE1}} + (I_1 + I_2)R_1 \tag{4.86}$$

Assuming that the collector resistors R_A and R_B are identical, the collector currents of Q_1 and Q_2 are *forced* to be equal in order to set the differential voltage at the input of the operational amplifier equal to zero. This in turn forces I_1 to be equal to I_2, and the voltage drop across R_2 is equal to the ΔV_{BE} between Q_1 and Q_2. Thus, Eq. (4.86) can be rewritten as

$$V_{\text{ref}} = V_{\text{BE1}} + \frac{2R_1}{R_2} V_T \ln(n) \tag{4.87}$$

which is of the form given by Eq. (4.82). In practical applications, the emitter area ratio n is usually taken as 2, that is, Q_2 is made to have *twice* the emitter area of Q_1.

The circuit of Figure 4.36*b* also operates on a similar principle.[1] Since the differential voltage at the input of the operational amplifier has to be zero, the currents I_1 and I_2 are forced to have the ratio

$$\frac{I_1}{I_2} = \frac{R_2}{R_1} \tag{4.88}$$

Assuming that Q_1 and Q_2 are well matched and neglecting the effect of base currents, the difference in their base-emitter voltages can be expressed as

$$V_{\text{BE1}} - V_{\text{BE2}} = \Delta V_{\text{BE}} = V_T \ln\left(\frac{I_1}{I_2}\right) = V_T \ln\left(\frac{R_2}{R_1}\right) \tag{4.89}$$

Note that this differential voltage ΔV_{BE} appears directly across resistor R_3, that is,

$$V_{\text{BE}} = I_2 R_3 = \frac{I_1 R_2 R_3}{R_1} \tag{4.90}$$

The output voltage V_{ref} is equal to V_{BE1} plus the voltage drop across R_1,

$$V_{\text{ref}} = V_{\text{BE1}} + I_1 R_1 \tag{4.91}$$

Substituting the results of Eqs. (4.89) and (4.90) into Eq. (4.91), one obtains

$$V_{\text{ref}} = V_{\text{BE1}} + V_T \frac{R_2}{R_3} \ln\left(\frac{R_2}{R_1}\right) \tag{4.92}$$

which is again of the same form as the basic band-gap reference equation [Eq. (4.82)].

The advantage of the circuits of Figure 4.36 over that of Figure 4.35 is the added degree of stability, since the internal current levels and their ratios are well controlled. Their disadvantage is the need for additional resistors, which need to be perfectly matched, and the need for a stable high-gain feedback amplifier.

EX-1014 0168
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

4.8. STABILIZATION OF CHIP TEMPERATURE

The close thermal coupling between the integrated components on a monolithic chip allows the circuit designer an added degree of freedom in controlling the thermal environment of the IC chip. Because of the small thermal capacity of the chip, the entire chip can be maintained at a constant elevated temperature within the IC package. This can be done by incorporating a heating element, along with a temperature sensor–controller circuit, on the same IC chip, alongside the circuit to be stabilized. This temperature sensor–controller unit is basically a temperature regulator whose function is to maintain the IC chip at a constant elevated temperature with minimum dependence on ambient variations. In other words, the IC package can serve as a miniature *temperature chamber*, where the IC chip itself is both the heating and the temperature-regulating element. Since the thermal capacity of the IC chip is relatively small, such a stabilization scheme can often be achieved without requiring excessive amounts of power dissipation.[8]

Figure 4.37 shows a functional block diagram of a temperature sensor–controller circuit for stabilizing the substrate temperature. All the necessary circuit elements to form such a temperature regulating system are readily available in the form of integrated components. The predictable temperature dependence of the transistor V_{BE} can be utilized as the temperature-sensing element, and a power transistor can be used as the heating element. In order to minimize the thermal gradients through the chip, the circuit to be stabilized is laid out symmetrically with respect to the heating and sensing elements, and the most critical components, such as the input stage of a high-gain amplifier, are placed equidistant from the heating element and located nearest to the sensor. The threshold level of the heater unit is set such that the heater is operative over

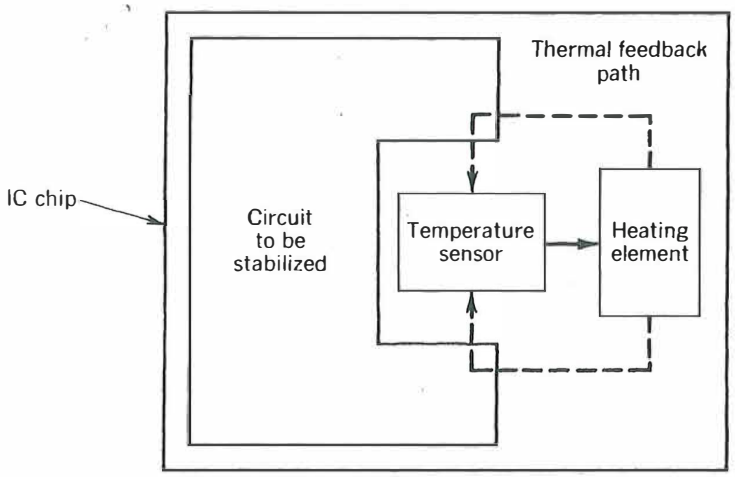


FIGURE 4.37. Block diagram for substrate temperature stabilization of an IC chip.

EX-1014 0169
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

the entire temperature range of interest and keeps the chip temperature at a relatively constant level, above the highest ambient temperature to be encountered.

The most important application of temperature-controlled substrate circuits is in the design of precision voltage reference circuits or in very-low-drift differential amplifiers. For example, if the temperature of the chip can be stabilized to ±5°C over an ambient temperature change of from −55°C to +125°C, an inherently stable voltage reference, which exhibits ±40-ppm drift per degree celsius change of chip temperature, can appear to be stable to ±2-ppm/°C change of ambient temperature.

Using this technique, voltage reference circuits with temperature drifts of ≤ 1 ppm/°C have been developed as successful commercial products.[9,10]

Figure 4.38 shows a practical design example to illustrate the principle of operation of a temperature-stabilized integrated circuit. For illustrative purposes, the circuit to be stabilized is assumed to be a voltage reference. In the circuit, the diodes D_1 and D_2, which have a temperature-dependent voltage drop $V_{BE}(T)$, serve as the temperature-sensing elements and generate a temperature-sensitive control voltage V_0, where

$$V_0 = V_1 - V_2 = V_{BE}(T) - V_{ref}\frac{R_2}{R_1 + R_2} \tag{4.93}$$

This voltage drives a differential amplifier which activates the heater transistor Q_0.

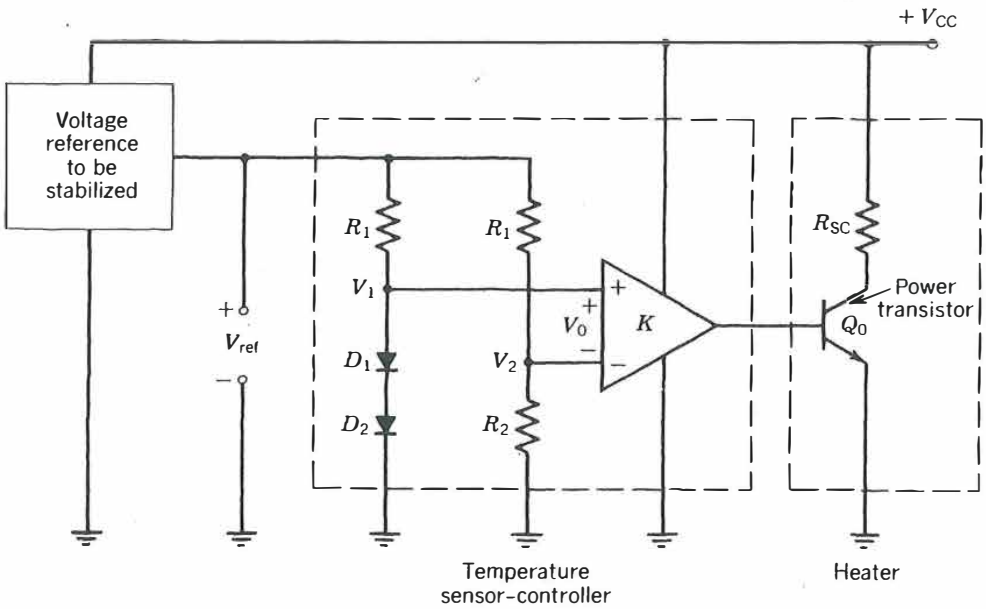


FIGURE 4.38. Simplified example of chip temperature stabilization for precision voltage reference.

EX-1014 0170
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

The diode voltage drop makes an excellent temperature sensor, since it varies very linearly with temperature T, and is given as

$$V_{BE}(T) = V_{BE}(T_0) + \gamma_D(T - T_0) \tag{4.94}$$

where $V_{BE}(T_0)$ is the nominal diode drop at the reference temperature, which is usually taken as +25°C. For base–emitter diodes, the temperature coefficient γ_D is ≈ -2 mV/°C.

In most applications, the chip temperature T_C is normally chosen to be 25–50°C higher than the highest operating ambient temperature. Once T_C is chosen, one would then choose the resistor divider ratio in Eq. (4.93) to have $V_0 = 0$ at $T = T_C$. Thus, the heater will be turned on until the chip temperature reaches T_C, and then would stabilize itself at a power dissipation level sufficient to maintain the chip temperature at a steady-state value very close to T_C. The current-limiting resistor R_{SC}, at the collector of the heater transistor, is used to minimize the surge current transients when the heater transistor is first turned on. The accuracy of the temperature sensor–controller circuit to maintain the chip temperature at or very near T_C over wide changes of ambient temperature is determined by the amplifier gain K. The total power dissipation required, as well as the thermal time constants of the system, are determined by the thermal properties of the IC package. To avoid excessive power dissipation, a thermally insulated package structure is normally used.

At this point, a note of caution is also in order. The temperature sensor–controller circuit of Figure 4.38 is a closed-loop feedback system and, thus, may have stability problems, particularly if the amplifier gain K is set too high.

Although circuits with stabilized substrate temperatures are useful in special circuit applications requiring an unusually high degree of precision, their general acceptance has been somewhat limited due to the following reasons:

1. Relatively high power dissipation (typically on the order of 500 mW or more).
2. Need for special packaging with good thermal insulation, in order to minimize power dissipation.
3. Relatively large surge currents and long time constants (typically on the order of 5–10 secs) needed for the initial warm-up time when the power is first applied.

As mentioned in the beginning of this chapter, the excellent matching of integrated components, as well as the close thermal coupling between them, provides the IC designer with a diverse set of design techniques. With imaginative application of these unique circuit techniques, some of which are outlined in this chapter, it is possible to design monolithic analog circuits which offer superior performance characteristics over their discrete counterparts.

EX-1014 0171
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

REFERENCES

1. P. R. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, Wiley, New York, 1977, Chap. 4.
2. G. R. Wilson, "A Monolithic Junction FET-NPN Operational Amplifier," *IEEE J. Solid-State Circuits*, **SC-3,** 341–348 (December 1968).
3. J. Davidse, *Integration of Analogue Electronic Circuits*, Academic Press, London, 1979, Chap. 4.
4. A. B. Grebene, Ed., *Analog Integrated Circuits*, IEEE Press, New York, 1978.
5. R. J. Widlar, "A Versatile Monolithic Voltage Regulator," National Semiconductor Appl. Note AN-1, 1967.
6. R. J. Widlar, "New Developments in IC Voltage Regulators," *IEEE J. Solid-State Circuits*, **SC-6,** 2–7 (February 1971).
7. A. P. Brokaw, "A Three-Terminal IC Bandgap Reference," *IEEE J. Solid-State Circuits*, **SC-9,** 388–393 (December 1974).
8. T. F. Prosser, "An Integrated Temperature Sensor-Controller," *IEEE J. Solid-State Circuits*, **SC-1,** 8–13 (1966).
9. R. Dobkin, "IC Reference Has 1 ppm/°C Drift," National Semiconductor Appl. Note 181, 1977.
10. D. P. Laude and J. D. Beason, "5 Volt Temperature Regulated Voltage Reference," *IEEE J. Solid-State Circuits*, **SC-15,** 1070–1076 (December 1980).

EX-1014 0172
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

CHAPTER FIVE

BASIC GAIN STAGES

The choice of input and output gain stages is among the most critical steps in analog IC design. This chapter will examine some of the basic gain stages very commonly utilized in analog integrated circuits. These gain stages serve as building blocks or subcircuits in almost all of the different classes of analog integrated circuits discussed in the later chapters.

The differential amplifier is one of the most widely used classes of gain stages in analog IC design. The first part of this chapter covers the subject of differential gain stages, with particular emphasis on the use of active devices as "active loads." The current mirror subcircuit, which was covered as a biasing element in Chapter 4, is now utilized as a dynamic load to obtain very high voltage gains from a single-stage amplifier. The second part of the chapter deals with output stages. These are the gain blocks or subcircuits designed to provide large signal swings into an output load, with minimum signal distortion or standby power requirements.

The analyses and design examples presented in this chapter are limited to bipolar technology. However, most of the design criteria and circuit properties are readily applicable to MOS devices, which will be covered in the next chapter.

5.1. DIFFERENTIAL GAIN STAGES

Differential amplifiers represent a broad class of circuits whose basic function is to amplify the *difference* between two input signals. For this reason, they are also referred to as *difference amplifiers*. The bias levels and the gain characteristics of a differential stage, by and large, depend on the symmetry between the two branches of the circuit.[1,2] This balanced nature of the differential amplifier makes it ideal as a gain block for integrated circuits, since close matching is

EX-1014 0173
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.

Wiley Classics Library

The **Wiley Classics Library** consists of selected books that have become recognized classics in their respective fields. With these new unabridged and inexpensive editions, Wiley hopes to extend the life of these important works by making them available to future generations of mathematicians and scientists.

ISBN 978-0-471-43078-0



EX-1014 0174
Kangxi Communication Technologies (Shanghai) Co., Ltd. v. Skyworks Solutions, Inc.