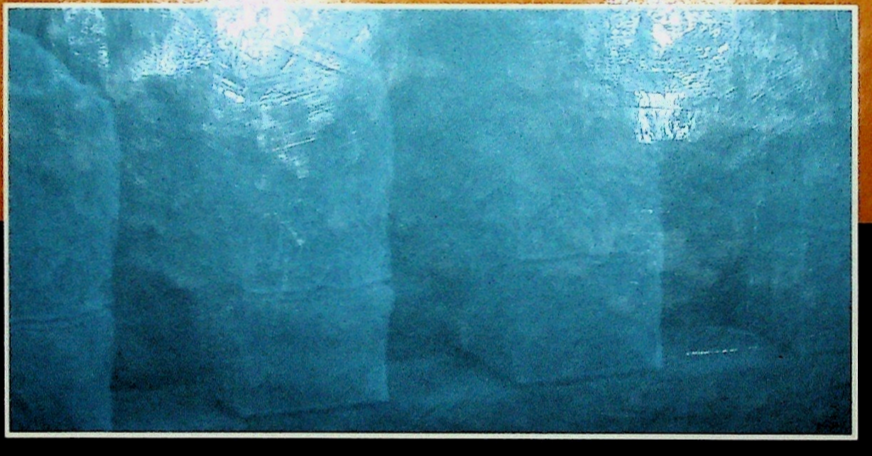
WILEY

PHYSICS OF SEMICONDUCTOR DEVICES



THIRD EDITION

IPR2025-00845
Apple-Sony EX1021 Page 1

Physics of Semiconductor Devices

IPR2025-00845
Apple-Sony EX1021 Page 2

Physics of Semiconductor Devices

Third Edition

S. M. Sze
National Chiao Tung University
Hsinchu, Taiwan
and
Stanford University
Stanford, California

Kwok K. Ng
Semiconductor Research Corporation
Durham, North Carolina

WILEY-INTERSCIENCE
A JOHN WILEY & SONS, INC., PUBLICATION

IPR2025-00845
Apple-Sony EX1021 Page 3

Description of cover photograph:

A scanning electron micrograph of an array of the floating-gate nonvolatile semiconductor memory (NVSM) magnified 100,000 times. NVSM was invented at Bell Telephone Laboratories in 1967. There are more NVSM cells produced annually in the world than any other semiconductor device and, for that matter, any other human-made item. For a discussion of this device, see Chapter 6. Photo courtesy of Macronix International Company, Hsinchu, Taiwan, ROC.

Copyright © 2007 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 Section 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, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. 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, or online at http://www.wiley.com/go/permission.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations 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 or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States 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 may not be available in electronic format. For information about Wiley products, visit our web site at www.wiley.com.

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

ISBN-13: 978-0-471-14323-9
ISBN-10: 0-471-14323-5

Printed in the United States of America.

10 9 8 7 6 5 4 3

Introduction

The book is organized into five parts:

Part I: Semiconductor Physics
Part II: Device Building Blocks
Part III: Transistors
Part IV: Negative-Resistance and Power Devices
Part V: Photonic Devices and Sensors

Part I, Chapter 1, is a summary of semiconductor properties that are used throughout the book as a basis for understanding and calculating device characteristics. Energy band, carrier concentration, and transport properties are briefly surveyed, with emphasis on the two most-important semiconductors: silicon (Si) and gallium arsenide (GaAs). A compilation of the recommended or most-accurate values for these semiconductors is given in the illustrations of Chapter 1 and in the Appendixes for convenient reference.

Part II, Chapters 2 through 4, treats the basic device building blocks from which all semiconductor devices can be constructed. Chapter 2 considers the *p-n* junction characteristics. Because the *p-n* junction is the building block of most semiconductor devices, *p-n* junction theory serves as the foundation of the physics of semiconductor devices. Chapter 2 also considers the heterojunction, that is a junction formed between two dissimilar semiconductors. For example, we can use gallium arsenide (GaAs) and aluminum arsenide (AlAs) to form a heterojunction. The heterojunction is a key building block for high-speed and photonic devices. Chapter 3 treats the metal-semiconductor contact, which is an intimate contact between a metal and a semiconductor. The contact can be rectifying similar to a *p-n* junction if the semiconductor is moderately doped and becomes ohmic if the semiconductor is very heavily doped. An ohmic contact can pass current in either direction with a negligible voltage drop and can provide the necessary connections between devices and the outside world. Chapter 4 considers the metal-insulator-semiconductor (MIS) capacitor of which the Si-based metal-oxide-semiconductor (MOS) structure is the dominant member. Knowledge of the surface physics associated with the MOS capacitor is important, not only for understanding MOS-related devices such as the MOSFET and the floating-gate nonvolatile memory but also because of its relevance to the stability and reliability of all other semiconductor devices in their surface and isolation areas.

IPR2025-00845
Apple-Sony EX1021 Page 5

1.5.8 Space-Charge Effect

The space charge in a semiconductor is determined by both the doping concentrations and the free-carrier concentrations,

$$\rho = (p - n + N_D - N_A)q \quad . \tag{115}$$

In the neutral region of a semiconductor, $n = N_D$ and $p = N_A$, so that the space-charge density is zero. In the vicinity of a junction formed by different materials, dopant types, or doping concentrations, n and p could be smaller or larger than N_D and N_A, respectively. In the depletion approximation, n and p are assumed zero so that the space charge is equal to the majority-carrier doping level. Under bias, the carrier concentrations n and p can be increased beyond their values in equilibrium. When the injected n or p is larger than its equilibrium value as well as the doping concentration, the *space-charge effect* is said to occur. The injected carriers thus control the space charge and the electric-field profile. This results in a feedback mechanism where the field drives the current, which in turn sets up the field. The space-charge effect is more common in lightly doped materials, and it can occur outside the depletion region.

In the presence of a space-charge effect, if the current is dominated by the drift component of the injected carriers, it is called the *space-charge-limited current*. Since it is a drift current, it is given by, in the case of electron injection,

$$J = qnv. \tag{116}$$

The space charge again is determined by the injected carriers giving rise to the Poisson equation of the form

$$\frac{d^2\psi_i}{dx^2} = \frac{qn}{\varepsilon_s} \quad . \tag{117}$$

The carrier velocity v is related to the electric field by different functions, depending on the field strength. In the low-field mobility regime,

$$v = \mu\mathscr{E} \quad . \tag{118}$$

In the velocity-saturation regime, velocity v_s is independent of the field. In the limit of ultra-short sample or time scale, we have the ballistic regime where there is no scattering, and

$$v = \sqrt{\frac{2qV}{m^*}} \quad . \tag{119}$$

From Eqs. 116–119, the space-charge-limited current in the mobility regime (the Mott-Gurney law) can be solved to be (see Vol. 4 of Ref. 4)

$$J = \frac{9\varepsilon_s \mu V^2}{8L^3}, \tag{120}$$

in the velocity-saturation regime

$$J = \frac{2\varepsilon_s v_s V}{L^2}, \tag{121}$$

IPR2025-00845
Apple-Sony EX1021 Page 6

$$\alpha = \frac{4\pi k_e}{\lambda}. \tag{126}$$

In semiconductors, the absorption coefficient is a strong function of the wavelength or photon energy. Near the absorption edge, the absorption coefficient can be expressed as[5]

$$\alpha \propto (h\nu - E_g)^{\gamma} \tag{127}$$

where $h\nu$ is the photon energy and γ is a constant. There exist two types of band-to-band transitions: allowed and forbidden. (Forbidden transitions take into account the small but finite momentum of photons and are much less probable.) For direct-bandgap materials, transitions mostly occur between two bands of the same k value, as transitions (a) and (b) in Fig. 29. While allowed direct transitions can occur in all k values, forbidden direct transitions can only occur at $k \neq 0$. In the one-electron approximation, γ equals 1/2 and 3/2 for allowed and forbidden direct transitions, respectively. Note that for $k = 0$ at which the bandgap is defined, only allowed transition ($\gamma = 1/2$) occurs and thus it is used in determining the bandgap experimentally. For indirect transitions [transition (c) in Fig. 29], phonons are involved in order to conserve momentum. In these transitions, phonons (with energy E_p) are either absorbed or emitted, and the absorption coefficient is modified to

$$\alpha \propto (h\nu - E_g \pm E_p)^{\gamma}. \tag{128}$$

Here the constant γ equals 2 and 3 for allowed and forbidden indirect transitions, respectively.

In addition, increased absorption peaks and steps can be due to formation of excitons, which are bound electron-hole pairs with energy levels within the bands that move through the crystal lattice as a unit. Near the absorption edge, where the values of $(E_g - h\nu)$ become comparable with the binding energy of an exciton, the Coulomb interaction between the free electron and hole must be taken into account. The photon energy required for absorption is lowered by this binding energy. For $h\nu \gtrsim E_g$ the

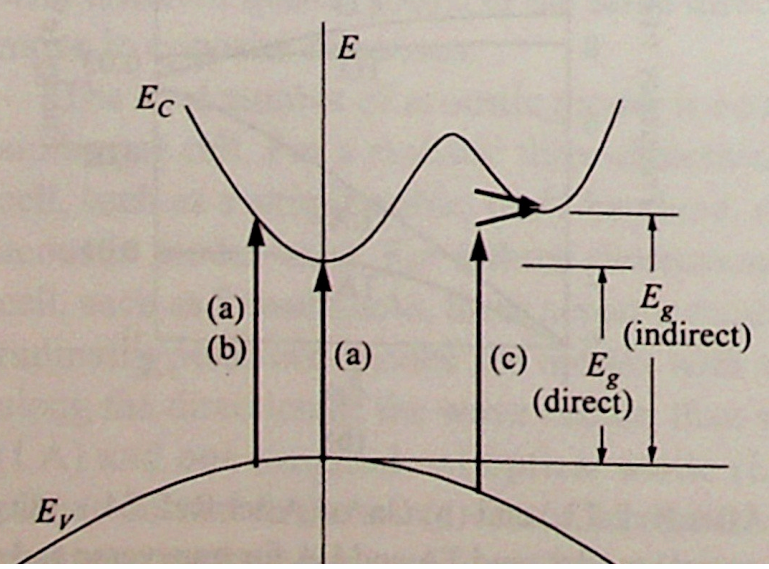


Fig. 29 Optical transitions: (a) allowed and (b) forbidden direct transitions; (c) indirect transition involving phonon emis-

IPR2025-00845
Apple-Sony EX1021 Page 7

absorption merges continuously into the fundamental absorption. When $h\nu \gg E_g$, higher energy bands participate in the transition processes, and complicated band structures are reflected in the absorption coefficient.

Figure 30 plots the experimental absorption coefficients α near and above the fundamental absorption edge (band-to-band transition) for Si and GaAs. The shift of the curves toward higher photon energies at lower temperature is associated with the temperature dependence of the bandgap (Fig. 6). An α of 10^4 cm^{-1} means that 63% of light will be absorbed in one micron of semiconductor.

When light passes through a semiconductor, absorption of light and generation of electron-hole pairs (G_e) occur, and the light intensity P_{op} diminishes with distance according to

$$\frac{dP_{op}(x)}{dx} = -\alpha P_{op}(x) = G_e h\nu. \tag{129}$$

Solution of the above gives an exponential decay of intensity

$$P_{op}(x) = P_0(1-R)\exp(-\alpha x) \tag{130}$$

where P_0 is the external incident light intensity and R is the reflection of the ambient-semiconductor interface at normal incidence,

$$R = \frac{(1-n_r)^2 + k_e^2}{(1+n_r)^2 + k_e^2}. \tag{131}$$

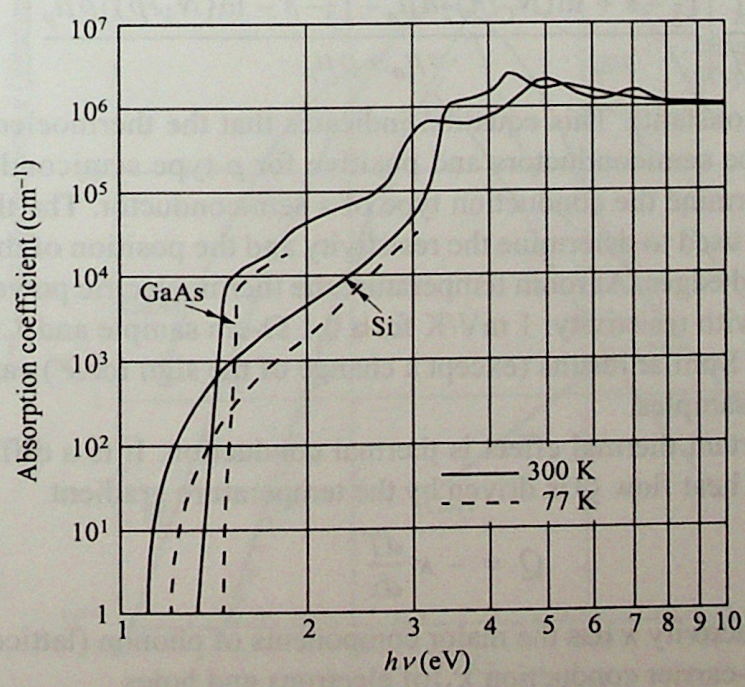


Fig. 30 Measured absorption coefficients near and above the fundamental absorption edge for Si and GaAs. (After Refs. 75–78.)

IPR2025-00845
Apple-Sony EX1021 Page 8

2.2 DEPLETION REGION

2.2.1 Abrupt Junction

Built-in Potential and Depletion-Layer Width. When the impurity concentration in a semiconductor changes abruptly from acceptor impurities N_A to donor impurities N_D, as shown in Fig. 1a, one obtains an abrupt junction. In particular, if $N_A \gg N_D$ (or vice versa), one obtains a one-sided abrupt p^+-n (or n^+-p) junction.

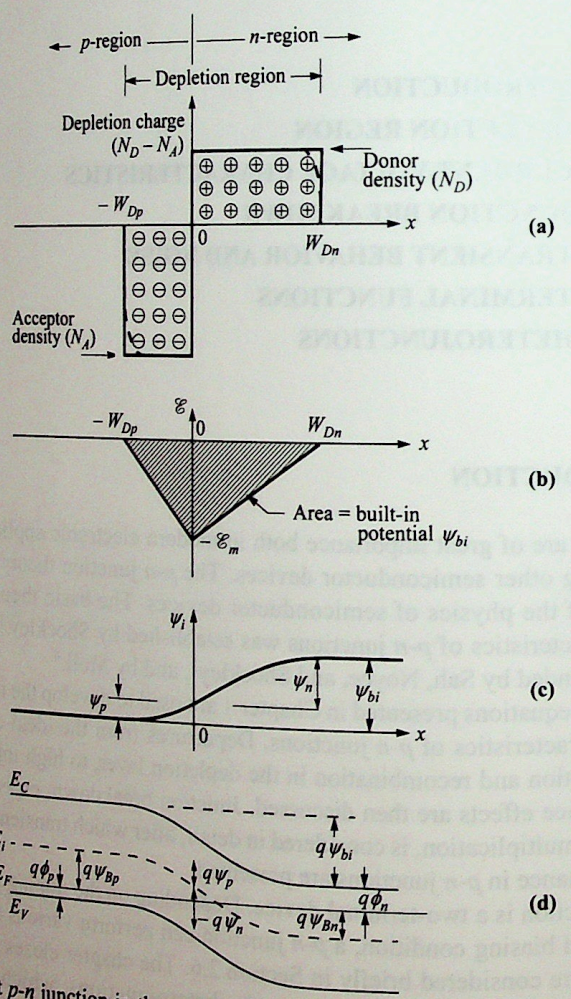


Fig. 1 Abrupt *p-n* junction in thermal equilibrium. (a) Space-charge distribution. Dashed lines indicate corrections to depletion approximation. (b) Electric-field distribution. (c) Potential distribution where ψ_{bi} is the built-in potential. (c) Energy-band diagram.

IPR2025-00845
Apple-Sony EX1021 Page 9

$$|\mathscr{E}_m| = \frac{qaW_D^2}{8\varepsilon_s}. \tag{32}$$

Integrating Eq. 30 once again gives the potential distribution shown in Fig. 5c

$$\psi_i(x) = \frac{qa}{6\varepsilon_s}\left[2\left(\frac{W_D}{2}\right)^3 + 3\left(\frac{W_D}{2}\right)^2 x - x^3\right] \qquad -\frac{W_D}{2} \le x \le \frac{W_D}{2} \tag{33}$$

from which the built-in potential can be related to the depletion width

$$\psi_{bi} = \frac{qaW_D^3}{12\varepsilon_s} \tag{34}$$

or

$$W_D = \left(\frac{12\varepsilon_s\psi_{bi}}{qa}\right)^{1/3}. \tag{35}$$

Since the values of the impurity concentrations at the edges of the depletion region ($-W_D/2$ and $W_D/2$) are the same and equal to $aW_D/2$, the built-in potential for a linearly graded junction can be approximated by an expression similar to Eq. 5:

$$\psi_{bi} \approx \frac{kT}{q}\ln\left[\frac{(aW_D/2)(aW_D/2)}{n_i^2}\right]$$
$$\approx \frac{2kT}{q}\ln\left(\frac{aW_D}{2n_i}\right). \tag{36}$$

Equations 35 and 36 can thus be used to solve for W_D and ψ_{bi}.

Based on an accurate numerical technique,[8] the built-in potential can be calculated explicitly by an expression as a *gradient voltage* V_g:

$$V_g = \frac{2kT}{3q}\ln\left(\frac{a^2\varepsilon_s kT}{8n_i^3q^2}\right). \tag{37}$$

The gradient voltages for Si and GaAs as a function of impurity gradient are shown in Fig. 6. These voltages are smaller than the ψ_{bi} calculated from Eq. 36, using the depletion approximation, by more than 100 mV. The depletion-layer width and the corresponding capacitance for silicon using this V_g as the built-in potential are plotted in Fig. 7 as a function of net potential $(V_g - V)$.

The depletion-layer capacitance for a linearly graded junction is given by

$$C_D = \frac{\varepsilon_s}{W_D} = \left[\frac{qa\varepsilon_s^2}{12(\psi_{bi} - V)}\right]^{1/3} \tag{38}$$

where V is positive/negative for forward/reverse bias.

2.2.3 Arbitrary Doping Profile

In this section we consider the doping near the junction to be of any arbitrary shape. Limiting the discussion to the *n*-side of a p^+-*n* junction, the net potential change at the junction is given by integrating the total field across the depletion region;

IPR2025-00845
Apple-Sony EX1021 Page 10

bound by layers of opposite types. This double-heterojunction design yields the highest efficiency and is the preferred approach.

Furthermore, if the central active layer is reduced to the range of 10 nm or smaller, a quantum well is formed. In this case, the 2-dimensional carrier densities become relevant and have to be calculated based on quantum mechanics. The effective 3-dimensional carrier densities (per unit volume), however, is given by the 2-D values divided by the quantum-layer width. This phenomenon pushes the carrier densities to higher levels and can result in higher efficiency. Another advantage of a thin active layer comes about in epitaxial growth, since a thin strained layer can accommodate higher level of lattice mismatch (see Section 1.7). Another feature of a quantum well is that quantization levels can theoretically extend the radiation energy (or to shorter wavelength) beyond the energy gap, but this feature is rarely used.

12.3.2 Materials of Choice

The most-important semiconductors for LED applications are listed in Figure 10, where the relative luminosity for the human eye is also added for reference. The spectrum covers most of the visible and extends into the infrared region. For display applications, since the human eye is only sensitive to light of energy $h\nu \geq 1.8$ eV ($\lambda \leq 0.7$ μm), semiconductors of interest must have energy bandgaps larger than this value. In general, all of these semiconductors are direct-bandgap materials except for some of the alloy composition in the GaAsP system which will be discussed in more details later. Direct-bandgap semiconductors are particularly important for electroluminescent devices, because the radiative recombination is a first-order transition process (no phonon involved) and the quantum efficiency is expected to be much higher than that for an indirect-bandgap semiconductor, where a phonon is involved.

AlGaAs. The $Al_xGa_{1-x}As$ systems covers a wide range of wavelengths from red to infrared. In the form of GaAs, it is the earliest material used for high-efficiency LEDs

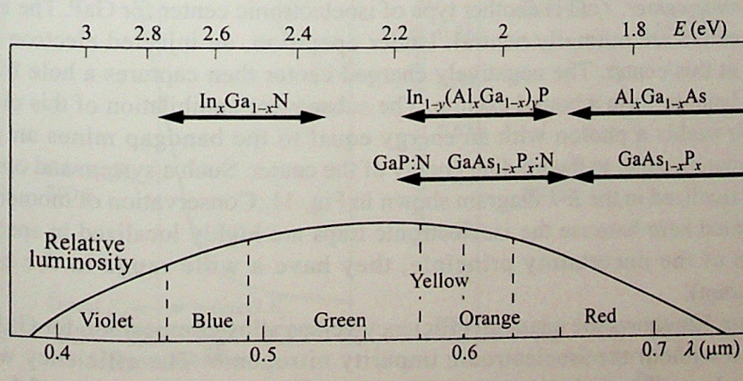


Fig. 10 Semiconductors of interest for LEDs, including the relative luminosity function of the human eye.

IPR2025-00845
Apple-Sony EX1021 Page 11

and reemits light of different wavelengths. The wavelength converters can be phosphors,[23] dyes, and other semiconductors. They usually convert to lower energy (longer wavelength) and have a wider spectrum compared to the original light. Their efficiencies are usually quite high. These are the reasons that they are quite common in white-light LEDs. Not as common is conversion to higher energy. Blue LEDs can be obtained from infrared-to-visible up-converters,[24] where the infrared emission from a GaAs LED is absorbed by phosphor doped with rare-earth ions, such as ytterbium (Yb^{3+}) and erbium (Er^{3+}). The operation depends on the successive absorption of two photons in the infrared region followed by the emission of a single photon in the visible region.

12.3.3 Definitions of Efficiencies

The main function of an LED is to convert electrical energy into light in the visible part of the spectrum for display and illumination purposes. In this section, the different terms of LED efficiencies are discussed. Knowing the origins of these efficiencies, optimization can be made accordingly.

Internal Quantum Efficiency. For a given input power, the radiative recombination processes are in direct competition with the nonradiative ones. Each of the band-to-band transitions and transitions via traps can be either radiative or nonradiative. Examples of nonradiative band-to-band recombination are those in indirect-bandgap semiconductors. Conversely, examples of radiative recombination via traps are those via isoelectronic levels.

The internal quantum efficiency η_{in} is the efficiency of converting carrier current to photons, defined as

$$\eta_{in} = \frac{\text{number of photons emitted internally}}{\text{number of carriers passing junction}}. \tag{8}$$

It can be related to the fraction of the injected carriers that combine radiatively to the total recombination rate, and may also be written in terms of their lifetimes as

$$\eta_{in} = \frac{R_r}{R_r + R_{nr}} = \frac{\tau_{nr}}{\tau_{nr} + \tau_r} \tag{9}$$

where R_r and R_{nr} are the radiative and nonradiative recombination rates, and τ_r and τ_{nr} are their associated radiative and nonradiative lifetimes, respectively. For low-level injection, the radiative recombination rate in the p-side of the junction is given by

$$\begin{aligned} R_r &= R_{ec} np \\ &\approx R_{ec} \Delta n N_A, \end{aligned} \tag{10}$$

where R_{ec} is the recombination coefficient and Δn is the excess carrier density which is much larger than the minority carrier density in equilibrium $\Delta n \gg n_{po}$. R_{ec} is a function of the band structure and temperature. Its value would be very small for indirect-bandgap semiconductors. ($R_{ec} \approx 10^{-10}$ cm^3/s for direct-bandgap materials, and $\approx 10^{-15}$ cm^3/s for indirect-bandgap materials.)

IPR2025-00845
Apple-Sony EX1021 Page 12

out that the wavelength ranges indicated are from interband transitions across their bandgaps. For intersubband transitions within the conduction band as in a quantum cascade laser (Section 12.6.3), the wavelength can be much extended for a material system.

Since heterojunction lasers are most common, the bandgap-lattice constant relationship is critical in choosing the appropriate material combination. Such relationships for some common material systems are shown in Fig. 32 of Chapter 1. To achieve heterojunctions with negligible interface traps, the lattices between the two semiconductors must be closely matched. Meanwhile, large bandgap difference is desirable for carrier confinement, and large refractive-index difference is beneficial in waveguiding. Also, as shown in the figure, some of the materials span from direct bandgap to indirect bandgap where lasing is prohibited. Therefore, this composition has to be avoided.

One of the most-important applications for lasers is optical-fiber communication. Figure 23 shows the loss characteristics achieved in experimental fibers. Three wavelengths of particular interest are also indicated on the figure. Around 0.9-μm wavelength, the GaAs/$Al_xGa_{1-x}As$ heterostructure lasers serve as the optical source, and the Si photodiode can be used as the less-expensive photodetector. Around 1.3-μm wavelength the fiber has low loss (0.6 dB/km) and low dispersion; and around 1.55-μm wavelength, the loss reaches a minimum of 0.2 dB/km. For these two wavelengths, III-V quaternary compound lasers, such as $In_xGa_{1-x}As_yP_{1-y}$/InP lasers, are candidates for optical sources, and photodiodes in ternary or quaternary compounds as well as Ge avalanche photodiodes are candidates for optical detectors.[47]

Device Structures. The basic structure of a laser is a *p-n* junction surrounded by optically designed surfaces, as shown in Fig. 24. A pair of parallel planes are cleaved or polished perpendicular to the plane of the junction. The two remaining sides of the

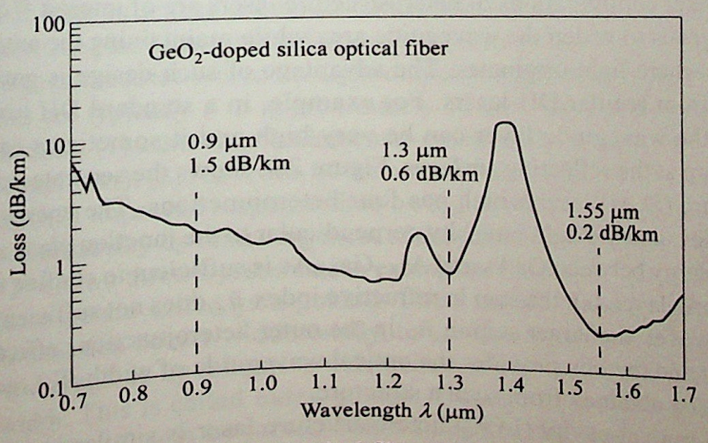


Fig. 23 Loss characteristics of silica optical fiber. The three wavelengths of interest are also shown. (After Ref. 46.)

IPR2025-00845
Apple-Sony EX1021 Page 13

noise. This is due to random effects associated with surface traps and generally has $1/f$ characteristics that are more pronounced at low frequencies. The generation-recombination noise comes from the fluctuations of these generation and recombination events. Generation noise can originate from both optical and thermal processes.

Since all the noises are independent events, they can be added together as the total noise. A related figure-of-merit[2] is the noise-equivalent power (NEP) that corresponds to the incident rms optical power required to produce a signal-to-noise ratio of one in a 1-Hz bandwidth. To the first order, this is the minimum detectable light power. Finally, the detectivity D^* is defined as

$$D^* = \frac{\sqrt{AB}}{\text{NEP}} \qquad \text{cm-Hz}^{1/2}/W, \tag{4}$$

where A is the area and B is the bandwidth. This is also the signal-to-noise ratio when one watt of light power is incident on a detector of area 1 cm^2, and the noise is measured over a 1-Hz bandwidth. The parameter is normalized to the area since the device noise is generally proportional to the square root of area. The detectivity depends on the detector's sensitivity, spectral response, and noise. It is a function of wavelength, modulation frequency, and bandwidth, and is recommended to be expressed as $D^*(\lambda,f,B)$.

The last section of this chapter deals with solar cells which to some extent share some similarity with photodetectors in that they both convert light to electricity. The purpose of the solar cells, however, is for power generation from the sunlight, as opposed to detection of faint light. So one difference between them is the intensity of light involved. The second difference being that solar cells are power generators and as such no external bias is required, whereas photodetectors usually require some bias and the change of current is detected as signal.

13.2 PHOTOCONDUCTOR

A photoconductor consists simply of a slab of semiconductor, in bulk or thin-film form, with ohmic contacts affixed to the opposite ends (Fig. 2). When incident light falls on the surface of the photoconductor, carriers are generated either by band-to-band transitions (intrinsic) or by transitions involving forbidden-gap energy levels (extrinsic), resulting in an increase in conductivity. The processes of intrinsic and extrinsic photoexcitations of carriers are shown in Fig. 3.

For the intrinsic photoconductor, the conductivity is given by $\sigma = q(\mu_n n + \mu_p p)$, and the increase of conductivity under illumination is mainly due to the increase in the number of carriers. The wavelength cutoff is given by Eq. 1, where ΔE is the semiconductor bandgap E_g in this case. For shorter wavelengths, the incident radiation is absorbed by the semiconductor, and electron-hole pairs are generated. For the extrinsic photoconductor, photoexcitation occurs between a band edge and an impurity energy level in the energy gap.

IPR2025-00845
Apple-Sony EX1021 Page 14

tors are cooled to lower temperatures (such as 77 K and 4.2 K). The lower temperatures reduce thermal effects which cause thermal ionization and deplete the energy levels, and increase the gain and detection efficiency. Near 0.5 μm, a CdS photoconductor has high sensitivity, whereas at 10 μm a HgCdTe photoconductor is preferred.[6] In the wavelength range from 100 to 400 μm, a GaAs extrinsic photoconductor is a better choice because of its high detectivity.[7] This photoconductor has high dynamic range and can give comparable performance for high-level (strong light intensity) detection. For low-level detection at microwave frequencies, however, a photodiode will provide considerably more speed and higher signal-to-noise ratio. Thus photoconductors have limited use in high-frequency optical demodulators, such as in optical mixing. They have been, however, extensively used for infrared detection especially beyond a few microns of wavelength.

13.3 PHOTODIODES

13.3.1 General Consideration

A photodiode has a depleted semiconductor region with a high electric field that serves to separate photogenerated electron-hole pairs. For high-speed operation, the depletion region must be kept thin to reduce the transit time. On the other hand, to increase the quantum efficiency (the number of electron-hole pairs generated per incident photon), the depletion layer must be sufficiently thick to allow a large fraction of the incident light to be absorbed. Thus, there is a trade-off between the speed of response and quantum efficiency.

For the visible and near-infrared wavelength range, photodiodes are usually reverse-biased with moderate biasing voltages, because this reduces the carrier transit time and lowers the diode capacitance. The reverse voltage is, however, not large enough to cause avalanche multiplication or breakdown. This biasing condition is in contrast to avalanche photodiodes, where an internal current gain is obtained as a result of the impact ionization under avalanche breakdown conditions. All photodiodes, with the exception of the avalanche photodiode which is not included in this section, thus have a maximum gain of one (Table 1). The photodiode family includes the *p-i-n* photodiode, *p-n* photodiode, heterojunction photodiode, and metal-semiconductor (Schottky barrier) photodiode.

We shall now briefly consider the general characteristics of a photodiode: its quantum efficiency, response speed, and device noise.

Quantum Efficiency. As mentioned previously, quantum efficiency is the number of electron-hole pairs generated per incident photon (Eq. 2). A related figure of merit is the responsivity, which is the ratio of the photocurrent to the optical power (Eq. 3). Therefore, for a given quantum efficiency, the responsivity increases linearly with wavelength. For an ideal photodiode ($\eta = 1$), $\mathscr{R} = \lambda/1.24$ (A/W) where λ is expressed in microns.

Since the optical absorption coefficient α is a strong function of the wavelength, for a given semiconductor the wavelength range in which appreciable photocurrent

IPR2025-00845
Apple-Sony EX1021 Page 15

$$P_{\text{opt}}|_{\text{min}} = \frac{2h\nu}{\eta}\sqrt{\frac{(S/N)I_{eq}B}{q}} \tag{22}$$

where

$$I_{eq} = I_B + I_D + \frac{2kT}{qR_{eq}}. \tag{23}$$

The noise-equivalent power (NEP) is given by ($S/N = 1$, $B = 1$ Hz)

$$\begin{aligned}\text{NEP} &= \text{rms optical power } P_{\text{opt}}|_{\text{min}} \\ &= \left(\frac{h\nu}{\eta}\right)\sqrt{\frac{2I_{eq}}{q}} \quad \text{W/cm}^2\text{-Hz}^{1/2}.\end{aligned} \tag{24}$$

To improve the sensitivity of a photodiode, both η and R_{eq} should be increased while I_B and I_D should be minimized. The NEP decreases with R_{eq} until it saturates to a constant value limited by dark-current or background-current shot noise.

13.3.2 *p-i-n* and *p-n* Photodiodes

The *p-i-n* photodiode is a special case of the *p-n* junction photodiodes, and is one of the most-common photodetectors, because the depletion-region thickness (the intrinsic layer) can be tailored to optimize the quantum efficiency and frequency response. Figure 6 shows schematic representation of a *p-i-n* diode and its energy-band diagram under reverse-bias conditions, together with optical absorption characteristics. We shall discuss the operation of *p-i-n* photodiode in some detail with the help of Fig. 6. This discussion applies also to *p-n* junction photodiodes. Light absorption in the semiconductor produces electron-hole pairs. Pairs produced in the depletion region or within a diffusion length of it will eventually be separated by the electric field, leading to current flow in the external circuit as carriers drift across the depletion layer.

Quantum Efficiency. Under steady-state conditions the total photocurrent density through the reverse-biased depletion layer is given by[10]

$$J_{\text{tot}} = J_{dr} + J_{\text{diff}} \tag{25}$$

where J_{dr} is the drift current due to carriers generated within the depletion region and J_{diff} is the diffusion current due to carriers generated outside the depletion layer in the bulk of the semiconductor and diffusing into the reverse-biased junction. We shall now derive the total current under the assumptions that the thermal generation current can be neglected and that the surface *p*-layer is much thinner than $1/\alpha$. Referring to Fig. 6c, the electron-hole generation rate is given by

$$G_e(x) = \Phi_0 \alpha \exp(-\alpha x) \tag{26}$$

where Φ_0 is the incident photon flux per unit area given by $P_{\text{opt}}(1-R)/Ah\nu$, R is the reflection coefficient, and A is the device area. The drift current J_{dr} is thus given by

IPR2025-00845
Apple-Sony EX1021 Page 16

toward the semiconductor contribute to the photocurrent. The internal photoemission process is energy dependent, and the quantum efficiency is given by

$$\eta = C_F \frac{(h\nu - q\phi_B)^2}{h\nu}, \tag{40}$$

where C_F is the Fowler emission coefficient. This phenomenon is often applied to measure the barrier height. When a Schottky-barrier diode is scanned with light of variable wavelength, Fig. 11c shows that the quantum efficiency has a threshold of $q\phi_B$, and it increases with the photon energy. When the photon energy reaches the energy-gap value, the quantum efficiency jumps to a much higher value. In practical applications, the internal photoemission has typical quantum efficiencies of only less than 1%.

A typical configuration is illustrated in Fig. 9c. To avoid large reflection and absorption losses when the diode is illuminated through the metal contact, the metal film must be very thin, ≈ 10 nm, and an antireflection coating must be used. By using a low-doping *i*-layer, a metal-*i*-*n* photodiode similar to a *p*-*i*-*n* diode can be made. This structure is advantageous mainly for band-to-band excitation. A special metal-semiconductor diode is the point-contact photodiode shown in Fig. 9e.[16] The active volume is very small, and as a result both the drift time and capacitance are small. It is thus suitable for very-high modulation frequencies.

For detectors with internal photoemission, it is more efficient to direct the incoming light through the substrate. Since the barrier height is always smaller than the energy gap, light with $q\phi_B < h\nu < E_g$ is not absorbed in the semiconductor, and intensity is not reduced at the metal/semiconductor interface. The metal layer, in this case, can be thicker for easier thickness control and to minimize series resistance. For Si devices, options are available using silicides in place of the metal. A silicide usually has a more reproducible interface since it is formed by reacting metal with Si so that the new interface is never exposed. Common silicides used for this purpose are PtSi, Pd_2Si, and IrSi. Another advantage of a Schottky-barrier diode is that high-temperature processing for diffusion or implantation anneal is not required.

Metal-semiconductor photodiodes are particularly useful in the visible and ultraviolet regions. In these regions the absorption coefficients α in most of the common semiconductors are very high, of the order of 10^5 cm^{-1} or higher, corresponding to an effective absorption length of $1/\alpha = 0.1$ μm or less. It is possible to choose a proper metal and antireflection coating so that a large fraction of the incident radiation will be absorbed near the surface of the semiconductor.

The dark current in a Schottky diode is known to be due to thermionic emission of majority carriers and it does not suffer from charge storage of minority diffusion current, which limits the speed capability in a *p*-*n* photodiode. Ultrafast Schottky-barrier photodiodes operating beyond 100 GHz have been reported. The main advantages of the Schottky-barrier photodiode are high speed and long-wavelength detection capability without having to use a semiconductor with a small energy gap.

IPR2025-00845
Apple-Sony EX1021 Page 17

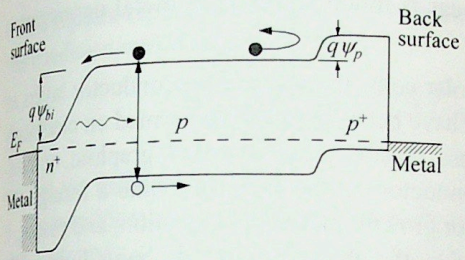


Fig. 48 Energy-band diagram for an n^+-p-p^+ back-surface field junction solar cell. (After Ref. 85.)

consider in the following other features that have contributed to the high performance of the solar cell.

The idea of a *back surface field* (BSF) has improved the output voltage over that in conventional cells. A schematic band diagram is shown in Fig. 48. The front surface is made in the normal way, but the back of the cell has a very heavily doped region adjacent to the contact. The potential energy barrier $q\psi_p$ tends to confine minority carriers (electrons) in the more lightly doped region and helps to drive them toward the front. The BSF cell is equivalent to a normal cell having a very small recombination velocity at the back ($S_n < 100$ cm/s). The low S_n will enhance the spectral response at low photon energies. Therefore, the short-circuit current density will increase. The open-circuit voltage is also increased due to increased short-circuit current, decreased diode recombination current at the back contact, and the added potential energy $q\psi_p$.

To reduce light reflection, a textured surface, either on the front or on the back, has been used to trap the light. The *textured* cell, as an example, has pyramidal surfaces produced by anisotropic etching of ⟨100⟩-oriented Si surface as shown in Fig. 49. Light incident on the side of a pyramid will be reflected onto another pyramid instead of being reflected backward. The reflectivity of bare Si is reduced from about 35% for flat surfaces to around 20% for the textured surface. The addition of an antireflection coating reduces the overall reflection to a few percent.

Another area of cost saving is the thick metallization process since solar cells are power devices and they conduct much higher current than regular integrated circuits. A process known as screen printing is commonly used to deposit thick metal layers in

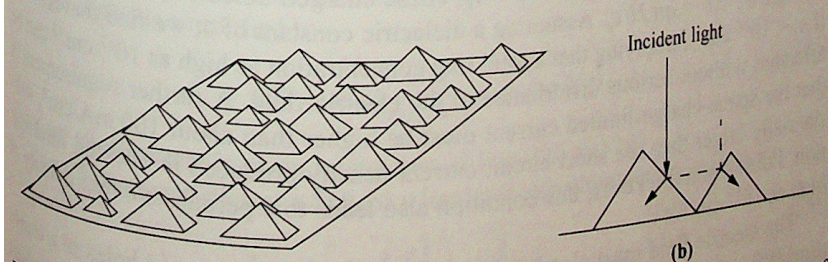


Fig. 49 (a) Textured cell with pyramidal surfaces. (b) Optical path showing the trapping of light to reduce reflection. (After Ref. 86.)

IPR2025-00845
Apple-Sony EX1021 Page 18