

Photoelectric Materials and Devices

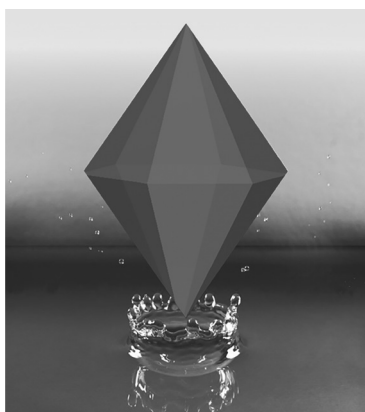


Tao Han | Haibo Ruan | Shixiu Cao | Youwei Guan
Dianyong Tang | Xin Yang

Photoelectric Materials and Devices

This page intentionally left blank

Photoelectric Materials and Devices



Tao Han
Haibo Ruan
Shixiu Cao
Youwei Guan
Dianyong Tang
Xin Yang

Chongqing University of Arts and Sciences, China

Published by

World Scientific Publishing Co. Pte. Ltd.

5 Toh Tuck Link, Singapore 596224

USA office: 27 Warren Street, Suite 401-402, Hackensack, NJ 07601

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

光电材料与器件

Originally published in Chinese by China Science Publishing & Media Ltd.

Copyright © China Science Publishing & Media Ltd., 2017

PHOTOELECTRIC MATERIALS AND DEVICES

Copyright © 2021 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 978-981-123-060-8 (hardcover)

ISBN 978-981-123-061-5 (ebook for institutions)

ISBN 978-981-123-062-2 (ebook for individuals)

For any available supplementary material, please visit

<https://www.worldscientific.com/worldscibooks/10.1142/12109#t=suppl>

Desk Editor: Nur Syarfeena Binte Mohd Fauzi

Typeset by Stallion Press

Email: enquiries@stallionpress.com

Printed in Singapore

Profile

This book mainly introduces the basic theory and physical characteristics of photoelectric materials, the preparation technology of photoelectric components, the working principle, the latest application, the latest progress of photoelectric materials and devices technology and the correlation with other technologies. The content mainly involves the theoretical basis of photoelectric materials, micro-nano photoelectric materials and devices, semiconductor luminescent materials and devices, inorganic photoluminescence materials, LED packaging technology, transparent conductive materials, touch screen, display screen, solar cell materials and the basic principles and development trend of their applications. In particular, the book gives a systematic theoretical analysis of new photoelectric materials and devices, such as optoelectronic materials and devices, optoelectronic materials and devices, and transparent conductive materials, and provides application examples.

This book can be used as a reference for undergraduate and graduate students, majoring in engineering such as optical information science and technology, information display and photoelectric technology, photoelectric information engineering, and photoelectric materials and devices, as well as for researchers and engineers.

This page intentionally left blank

Preface

With the rapid development of photoelectric technologies such as information, display, semiconductor lighting and photoelectric sensing, the photoelectric industry will be the most attractive sunrise industry in the 21st century and has a huge development potential. Photoelectric materials are the foundation and the precursor of the whole photoelectric industry. The production of new materials and devices promotes the significant progress of photoelectric technology. Without these basic materials and devices, it is difficult to assemble high-performance photoelectric equipment and build high-performance photoelectric system. Photoelectric materials and devices involve many disciplines such as optics, electronics and materials. However, there are seldom reference books.

According to the “13th Five-Year Plan” of our university and the idea of integrated construction of disciplines and specialties, key disciplines and specialties of materials will be built to form a discipline and specialty system that can meet the development needs of local pillar industries and strategic emerging industries. At present, the school has built materials science and engineering, microelectronics science and engineering and other majors, planning to build photoelectric information science and engineering, optical engineering and other majors. As the main course of the above majors, the teaching material construction of photoelectric materials and devices is of great significance.

This book mainly deals with the theoretical basis of photoelectric materials, micro-nano photoelectric materials and devices, semiconductor luminescent materials and devices, LED packaging

technology, inorganic photoluminescence materials, transparent conductive materials, touch screen, display screen, solar cell materials and the basic principles and development trends of applications.

This book is compiled by the teachers of new material technology research institute of Chongqing University of Arts and Sciences. It is divided into ten chapters. Tao Han organized the compilation of the book and wrote Chapters 1 and 6, Dianyong Tang wrote Chapters 2 and 3, Shixiu Cao wrote Chapters 4 and 5, Haibo Ruan wrote Chapter 7, Youwei Guan wrote Chapter 8, Xin Yang and Haibo Ruan and Tao Han wrote Chapter 9, Xin Yang wrote Chapter 10.

The writing features of this book are as follows:

- (1) starting from the current situation of technology, pay attention to the close combination with the development environment of modern photoelectric technology, and adapt to the idea of higher education teaching reform;
- (2) pay attention to the scientific, rigorous, advanced, practical and targeted contents in the compilation;
- (3) relevant content in simple terms, step by step, strengthen the application, the details slightly appropriate.

Thanks to the textbook project of Chongqing University of Arts and Sciences for the publication of this book.

Limited to the level of the editor, there are inevitably omissions and deficiencies in the book, the majority of readers are invited to criticize and correct, so as to correct.

January 2017

Tao Han
Chongqing University
of Arts and Sciences

Contents

<i>Profile</i>	v
<i>Preface</i>	vii
Chapter 1. Introduction	1
1.1 Brief Introduction of Photoelectronics Technology	1
1.1.1 Development course of optoelectronics technology	1
1.1.2 Related concepts of optoelectronic technology	3
1.2 The Concept, Position and Function of Photoelectric Materials and Devices	6
1.2.1 Basic concepts of photoelectric materials and devices	6
1.2.2 The position and function of photoelectric materials and devices in photoelectric technology	7
1.3 The Theoretical Basis of This Book	8
1.3.1 Photoelectric conversion materials and devices	8
1.3.2 Electro-optical conversion materials and devices	11
1.3.3 Transparent conductive material	11
References	12

Chapter 2. Theoretical Basis of Photoelectric

Materials	13
2.1 Energy Band Theory	13
2.1.1 Schrödinger equation of crystal and its approximate solution	14
2.1.2 Bloch's theorem	19
2.1.3 Periodic boundary conditions	21
2.1.4 Energy bands and their general properties	23
2.1.5 Brillouin zone	26
2.1.6 Metals, semiconductors and insulators	28
2.1.7 Electrons, holes and carriers	31
2.2 The Process of Absorption in a Material	33
2.2.1 The basic absorption	35
2.2.2 Direct transition of permission and prohibition	36
2.2.3 Indirect transition	39
2.2.4 Exciton	44
2.3 The Photoelectric Effect	46
2.3.1 External photoelectric effect	47
2.3.2 Photoconductive effect	53
2.3.3 Photovoltage	59
2.3.4 Thermoelectric effect	67
2.3.5 Pyroelectric effect	68
2.3.6 Photon traction effect	70
References	71

Chapter 3. Micro/Nano Photoelectric

Materials and Devices	73
3.1 Nanometer Photoelectric Materials and Devices . . .	74
3.1.1 Nano photoelectric materials	74
3.1.2 Nanometer photoelectric devices	76
3.2 Photonic Crystals and Photonic Crystal Devices	80
3.2.1 Structure of photonic crystals	80

3.2.2	Basic properties of photonic crystals	81
3.2.3	Photonic crystal devices	83
3.3	Metamaterials and Related Devices	89
3.3.1	Metamaterials	89
3.3.2	Negative refractive index materials and devices	89
3.3.3	Stealth cloak	95
3.4	Surface Plasmon Polaritons and Devices	98
3.4.1	Basic principles and properties	98
3.4.2	Surface plasmon optical waveguide	100
3.4.3	Surface plasmon resonance sensor	103
	References	106

Chapter 4. Semiconductor Luminescent Materials and Devices

109

4.1	Introduction to Semiconductor Luminescent Materials Crystals	110
4.1.1	Crystal structure	110
4.1.2	Defects and their effects on luminescence	116
4.1.3	Energy band structure	118
4.1.4	Conditions of semiconductor luminescent materials	124
4.2	Absorption of Light by Semiconductors	128
4.2.1	Classification of light absorbing mechanisms in semiconductors	128
4.2.2	Semiconductor optical absorption theory	132
4.3	Excitation and Luminescence of Semiconductors	134
4.3.1	PN junction and its characteristics	134
4.3.2	Recombination of injected carriers	144
4.4	Light-Emitting Diode Lighting Technology	151
4.4.1	Basic characteristics of LED	152
4.4.2	LED luminescence principle	155
4.4.3	Characteristic parameters of LED	158

4.4.4	LED driving technology	167
4.4.5	LED application	170
4.5	Photoluminescence and Electroluminescence of Organic Materials	172
4.5.1	Photoluminescence principle of organic materials	172
4.5.2	Relation between molecular structure and luminescence properties	176
4.5.3	Luminescence quenching phenomenon	177
4.5.4	Application of organic photoluminescent materials	179
4.5.5	Structure and luminescence principle of organic electroluminescent materials	180
4.5.6	Advantages of organic electroluminescent materials	181
4.5.7	Main problems of organic electroluminescent materials	182
4.5.8	Future development trends of organic electroluminescent materials	182
	References	183

Chapter 5. Inorganic Photoluminescent Materials 185

5.1	Inorganic Photoluminescence	185
5.1.1	Photoluminescence process	185
5.1.2	Light return to ground state: luminescence	187
5.1.3	Non-radiative return to ground state	189
5.2	Luminescence Principle of Phosphor	190
5.2.1	Basic concepts	192
5.2.2	Characteristics of phosphors	194
5.3	Development History and Present Situation of Phosphors	197
5.3.1	Development history of fluorescent powders for fluorescent lamps	197
5.3.2	Rare earth tribasic phosphor	199
5.3.3	White LED phosphor	207

5.4	Main Preparation Methods of Phosphors	209
5.4.1	High-temperature solid-phase method	210
5.4.2	Combustion synthesis method	213
5.4.3	Solvent (hydrothermal) method	215
5.4.4	Sol-gel method	217
5.4.5	Precipitation method	220
5.4.6	Spray pyrolysis	221
5.4.7	Microemulsion method	224
5.4.8	Polymer network gel method	225
5.4.9	Microwave method	226
	References	228

Chapter 6. Light-Emitting Diode Packaging Technology

231

6.1	LED Packaging Mode	231
6.1.1	Pin package	232
6.1.2	Surface mount packaging	233
6.1.3	Power package	234
6.1.4	Integrated multichip device packaging	235
6.1.5	Other packaging methods	236
6.2	LED Packaging Technology	238
6.3	LED Packaging Materials and Equipment	244
6.3.1	LED packaging material	244
6.3.2	LED packaging equipment	252
6.4	Fluorescent Powder Coating Technology	259
6.4.1	Mix phosphor powder	260
6.4.2	Phosphor coating	262
6.5	LED Heat Dissipation Technology	264
6.5.1	Source of heat	265
6.5.2	Effect of heat on LED	266
6.5.3	Heat dissipation mechanism and solution of LED	266
6.6	Optical Structure of LED	274
6.6.1	LED light conversion structure	277
6.6.2	LED light distribution structure	278

6.6.3	Simulation and design of LED packaging	279
6.7	Key Technology of Power LED Packaging	286
6.7.1	Ways to improve luminous efficiency	287
6.7.2	Improving the optical characteristics of LED	291
6.7.3	Increase the single light flux and input power of LED	293
6.7.4	Reduce the cost of LED	294
6.7.5	Improving the reliability of LED	294
	References	294
Chapter 7. Transparent Conductive Materials		297
7.1	Brief Introduction of Transparent Conductive Films	297
7.2	Typical TCO Materials	298
7.3	Conductivity of TCO	300
7.3.1	Conductivity principle of TCO	300
7.3.2	Energy band, orbital domain and mobility	302
7.3.3	N-type and P-type TCO	304
7.3.4	Carrier generation	305
7.3.5	The relation between the conductivity of TCO and temperature and carrier concentration	307
7.3.6	Relation between carrier scattering and resistance in TCO	309
7.4	Optical Properties of TCO	312
7.4.1	Transparency principle of TCO	312
7.4.2	Plasma vibration and plasma frequency	312
7.4.3	Burstein–Moss effect	314
7.4.4	Carrier concentration and transparency	316
7.5	Transparent Conductive Material Technology	316
7.5.1	Indium tin oxide	318
7.5.2	Other compromises between conductivity and transparency	319

7.5.3	Soft ITO films	324
7.5.4	Silver nanowires	328
	References	332

Chapter 8. Touch Panel 335

8.1	Development of Touch Technology	335
8.1.1	The generation of touch control technology	335
8.1.2	Definition of touch panel	336
8.1.3	Classification of touch panel	338
8.2	Resistance Touch Panel	339
8.3	Capacitance Touch Panel	341
8.3.1	Surface capacitance touch panel	342
8.3.2	Projection capacitive touch panel	344
8.4	Other Touch Technology	347
8.4.1	Infrared touch panel technology	347
8.4.2	Surface acoustic wave touch panel	349
8.5	The Frontier of Touch Control Technology	351
8.5.1	Embedded touch panel	351
8.5.2	Force touch technology	353
8.5.3	Flexible touch technology	353
8.6	Introduction of Touch Panel Production Technology	354
8.6.1	Production technology of glass touch panel	354
8.6.2	Production technology of film touch panel	362
8.6.3	Yellow light process for film touch panel	366
8.7	Capacitive Touch Panel Production Equipment and Materials	368
8.7.1	Major production equipment of capacitive touch panel	368
8.7.2	Capacitive touch panel manufacturing auxiliary materials	370
	References	376

Chapter 9. Display Screen 379

9.1	Introduction of Display Technology	379
9.2	Working Principle and Characteristic of Display . . .	383
9.2.1	Cathode ray tube (CRT) display	383
9.2.2	Liquid crystal display	384
9.2.3	Plasma display	385
9.2.4	Organic electroluminescent display	386
9.2.5	LED display	387
9.2.6	Organic Light Emitting Diode (OLED) display	388
9.3	Structure and Working Principle of TFT-LCD Device	389
9.3.1	Liquid crystal materials and their properties	389
9.3.2	Structure and working principle of panel plate	393
9.3.3	Structure and principle of backlight	397
9.4	Material Technology and Technology of TFT-LCD	399
9.4.1	TFT-LCD material technology	399
9.4.2	TFT-LCD technology	414
9.5	OLED Display Screen	422
9.5.1	Principle and application of OLED display	422
9.5.2	Fabrication of OLED devices	422
9.5.3	Advantages and disadvantages of OLED . . .	425
9.5.4	Driving mode of OLED	426
9.5.5	Challenges for OLED industry	429
	References	433

Chapter 10. Materials and Applications of Solar Cells 435

10.1	Photovoltaic Technology and Solar Cells	435
10.1.1	Development of solar cells	435
10.1.2	Principle of solar cell	437
10.1.3	Classification of solar cells	444

10.2	Crystalline Silicon Solar Cells	444
10.2.1	Polycrystalline silicon thin-film solar cells	444
10.2.2	Amorphous silicon thin-film solar cells	447
10.3	Compound Semiconductor Thin-Film Batteries . . .	448
10.3.1	CdTe solar cells	448
10.3.2	CIGS solar cells	453
10.3.3	GaAs solar cells	460
10.4	New Solar Cells	464
10.4.1	Dye-sensitized solar cells	464
10.4.2	Organic thin-film solar cells	468
10.4.3	Perovskite solar cells	471
	References	475

This page intentionally left blank

Chapter 1

Introduction

1.1. Brief Introduction of Photoelectronics Technology

According to the latest market analysis report of Global Industrial Analysis, the global output value of photoelectric industry reached 932 billion US dollars in 2015. Photoelectric industry is the first leading industry in the 21st century and the commanding point of economic development. Its strategic position is self-evident. Photoelectronics technology is the pillar and foundation of the photovoltaic industry, involving the frontier theories and technologies of optoelectronics, optics, materials, microelectronics, computer technology and so on. It is a high-tech science formed by the infiltration and intersection of many disciplines. Its application involves solar photovoltaic, light-emitting diode (OLED), flat panel display screen, laser, computer and communication. Information storage, modern testing instruments, smart glass and many other fields.

1.1.1. *Development course of optoelectronics technology*

The earliest optoelectronic devices are photodetectors, and the basis of photodetectors is the discovery and research of photoelectric effects. In 1873, Smith discovered the photoconductivity of selenium. In 1888, German Hertz observed that when ultraviolet light irradiated the metal, it could make the metal emit charged particles. In 1890, Lerner determined the charge-mass ratio of charged particles and proved them to be electrons, thus clarifying the essence of photoelectric effect. In 1900, German physicist Planck introduced

energy quantum into the study of blackbody radiation, and proposed the famous Planck formula to describe the phenomenon of blackbody radiation, which laid the foundation for quantum theory. In 1929, Kohler made a silver–oxygen–cesium photocathode and a photocell appeared. In 1939, Zvorakin of the Soviet Union made a practical photomultiplier tube. In the late 1930s, lead sulfide (PbS) infrared detectors were invented, which can detect radiation up to 3 microns. Thermoelectric infrared detectors and radiocalorimeters made of semiconductor materials appeared in the 1940s. In the mid-1950s, cadmium sulfide (CdS), cadmium selenide (CdSe), photoresistors and short-wave infrared lead sulfide photodetectors were put into use. In 1954, the first silicon-based solar cell was born at Bell Laboratory in the United States. In 1958, HgCdTe infrared detector was invented by Lawson *et al.*

In 1960, American Mayman developed ruby laser, the first laser in the world. In 1961, the first laser rangefinder appeared. Since then, various laser guided weapons, laser blinding weapons and laser destructive weapons have been successfully developed.

In 1964, RCA discovered the photoelectric effect, guest-host effect, dynamic scattering effect and phase-shift storage effect of liquid crystal, which laid a technical foundation for the development of liquid crystal display, liquid crystal light valve and other devices. Since then, flat panel display technology has developed fastest with liquid crystal display. Other flat panel displays, including plasma display and organic electroluminescent display, have come out one after another.

In 1966, optical fiber technology began to develop. In the same year, British-Chinese scientist Gao Rong and others proposed the possibility of realizing low-loss optical fibers, which opened the way for optical fiber communication. In the 1970s, the landmark achievements in the field of optoelectronics were the realization of low-loss optical fibers, the maturity of semiconductor lasers and the advent of charge coupled device (CCD). In 1970, the United States developed a quartz optical fiber with a loss of 20 dB/km and a semiconductor laser operating continuously at room temperature, making optical fiber communication possible. This year is recognized

as the first year of optical fiber communication. Since then, optical fiber communication has developed rapidly. At the same time of technological development, applications are also developing. In the early 1970s, the US laser-guided bomb was put into use. In 1972, Philips of the Netherlands demonstrated an analog laser video disk. In the mid- and late-1970s, Japan, the United States and the United Kingdom began to build backbone optical fiber communication networks.

In the 1990s, optoelectronic technology has achieved success in the field of storage. CD-ROM has become an important means of computer storage. CD and VCD have penetrated into millions of households. DVD also entered into people's lives in the mid-1990s. In addition, optoelectronic technology has made great progress in lighting and display. In 1983, Chinese American professor Deng Qingyun discovered organic light-emitting diode (OLED) in his laboratory. Its low voltage and high quantum efficiency make it a new generation of flat panel display technology. In 1993, Nishia Nakamura invented a blue light-emitting diode (LED) with commercial application value based on wide bandgap semiconductor materials GaN and InGaN, which innovated the light source technology.

1.1.2. Related concepts of optoelectronic technology

Optoelectronic technology is a very broad concept. It encompasses the generation, transmission, processing and reception of optical signals. It covers a series of fields from basic to application, such as new materials (new light-emitting photosensitive materials, nonlinear optical materials, substrates, transmission materials and micro-structures of artificial materials), micro-fabrication and micro-electromechanical, device and system integration. The science of optoelectronics technology is the pillar and foundation of optoelectronics information industry. It involves the frontier disciplines such as optoelectronics, optics, materials science, microelectronics, computer technology and so on. It is a high-tech discipline formed by the infiltration and intersection of many disciplines. At present, the scope of optoelectronic technology is shown in Table 1.1.

Table 1.1. Overview of optoelectronic technology.

Category	Industrial technical content		
Photoelectronic materials and components	Photoelectric material	Optical glass, optical plastics, optical crystals, infrared materials, laser materials, optical fibers, nonlinear optical materials, semiconductor optoelectronic materials and epitaxial materials, etc.	
	Photoelectric components	Light-emitting devices, light detecting/receiving components, light conduits, phototransistors, solar cells, etc.	
Photoelectric display	LED display, liquid crystal display, plasma display, organic light-emitting diode, organic light-emitting display, field-emitting display, laser holographic projection, etc.		
Optical sensor	Magneto-optic effect sensor, ambient light sensor, infrared light sensor, solar light sensor		
Optical input/output	Scanner, form reader, character recognition machine, digital camera, laser printer, laser copier, optical fax machine, etc.		
Optical storage	CD, VCD, DVD, optical quantum data storage, three-dimensional volume storage and CD-ROM, recorder		
Optical communication	Components and parts	Single-mode/multi-mode optical fibers, optical fiber connectors, etc.	
		Active devices	Optical amplifier, light source-fiber coupler, light source-detector coupler, light source Transistor Outline (TO) packaging device, optical pump device, etc.
		Passive devices	Optical switch, optical coupler, optical attenuator, optical isolator, spectrometer, wavelength division multiplexer, etc.
	Systems and equipment	Optical fiber transmission equipment, optical fiber area network equipment, optical fiber communication detection and monitoring equipment, optical fiber radio and television (CATV) system, etc.	
Laser and other applications	Laser	Semiconductor lasers, solid-state lasers, gas lasers, dye lasers, excimer lasers, etc.	
	Laser applications	Laser industrial processing, laser medical treatment, laser weapons, laser scientific research and other applications	

Photoelectronic technology is closely related to electronic technology. Electronics is a science and technology with application as its main purpose. It mainly studies the characteristics and behavior of electronics and the physics of electronic devices. Electronics involves many scientific categories, including physics, chemistry, mathematics, materials science and so on. Photoelectronics refers to the electronics of light wave bands, namely infrared, visible, ultraviolet and soft X-ray (wavelength range 1–10 nm). It is a science that studies photons as information carriers and energy carriers. It mainly studies how photons produce, move and transform. Electronic technology is a technology that studies the motion law of electrons and is applied to electronic devices, circuits and equipment. Photoelectronics technology is a technology that studies the interaction between light and electricity, that is, photons or light waves and electrons. It includes photoelectronics energy technology and photoelectronics information technology. Photoelectronics technology is inseparable from electronic technology. Photoelectronics technology needs energy–light conversion, light source needs driving and control circuit to emit light, and photoelectric signal processing is needed in the information field. At present, a large number of information processing is still based on electronic technology, and the conversion and amplification between electrical information and optical information are still needed in the process of transmitting and receiving. The core of future information technology is optoelectronic technology. The advantages of optoelectronic technology in the field of information technology are embodied in the following three aspects.

- (1) The frequency of light wave is higher and the information capacity that can be transmitted is larger.
- (2) Light wave propagation in medium has certain advantages.
- (3) Great progress has been made in optoelectronic information technology, especially after the emergence of lasers (optical communication, optical storage).

1.2. The Concept, Position and Function of Photoelectric Materials and Devices

1.2.1. *Basic concepts of photoelectric materials and devices*

The optoelectronic technology includes materials, devices, modules, equipment and systems. The devices are made of materials with certain properties (chemical, electrical, optical and mechanical properties). The modules are composed of devices suitable for specified parameters (current, voltage, response speed or frequency). The devices are operated by multiple functional modules in a certain software environment. The equipment is constructed under a certain standard agreement. Therefore, the bottom and foundation of photoelectric technology is photoelectric materials and devices.

Photoelectric materials are the foundation and pioneer of the whole photoelectric industry. Photoelectric materials refer to materials that can generate, convert, transmit, process and store optical signals. They mainly include semiconductor photoelectric materials (III–V), organic semiconductor photoelectric materials, inorganic crystals and quartz glass. Compound semiconductor materials such as AlGaAs and InGaAsN can be formed by any combination of III–V elements. Their lattice constants, bandgap width and absorption/emission wavelength are the three most important parameters determining the photoelectric properties of compound semiconductor materials. Photoelectric devices refer to devices that can realize the conversion function between light radiation energy and signal or the transmission, processing and storage of photoelectric signals. At present, most commercial semiconductor optoelectronic devices are made of GaAs-based, InP-based and GaN-based compound semiconductor material systems. They are widely used in optical communication network, photoelectric display, photoelectric storage, photoelectric conversion and photoelectric detection.

1.2.2. The position and function of photoelectric materials and devices in photoelectric technology

The emergence of new devices and materials promotes significant advances in optoelectronic technology, including LED, OLED, solar cells, touch screens, lasers, optical fibers, all-optical network devices (semiconductor lasers, Erbium Doped Fiber Amplifier (EDFA), Optical Cross Connector (OXC), Optical Add-Drop Multiplexing (OADM)), as well as laser printers, digital cameras, VCD, DVD and so on. Without these devices or without knowledge of them, it is difficult to assemble high-performance optoelectronic devices and build high-performance optoelectronic systems.

Photoelectric devices and products made of photoelectric materials are gradually applied in every important link of the information industry. From information acquisition, processing, transmission to information storage and display, the requirements of the information industry for high-speed, large-capacity, high-definition, ultra-thin and ultra-light information-related products are constantly increasing, which promotes the sustained and high-speed development of the photoelectric industry. New photoelectric materials and new products and new technologies are emerging. Among them, optoelectronic sensor system and its related devices, optoelectronic display, optical communication and optical storage are the most important application fields in the optoelectronic industry. In the field of optoelectronic display, the flat panel display products with liquid crystal display (LCD) as the mainstream have replaced the traditional cathode ray tube (CRT) market and occupied half of the display market. In photoelectric flat panel display devices and products, LCD has penetrated into every field of display devices; OLED, known as Dream Display, has also begun to be applied in small-scale display fields such as mobile phones, digital cameras, personal digital assistants (PDA). The successful development of GaN-based blue light-emitting diodes

(LED) and the appearance of commercial devices make it possible for full-color display and white light illumination of LED products, and set off a blue-light heat worldwide. White light has become the fourth generation of lighting sources.

With the successful development of semiconductor lasers, digital optical disks represented by CD ROM, VCD and DVD have become one of the indispensable storage technologies in today's multimedia information age. They have been widely used in computer storage, digital household appliances, radio and television, vehicle navigation and electronic publishing. Optical storage is developing along the direction of CD–DVD–3D holographic storage.

In recent years, the research and application of organic and organic/inorganic composite optoelectronic materials and devices have made great progress and development, which has attracted great attention from the International Optoelectronic academia and industry. Organic materials have become a new generation of optoelectronic information materials with the advantages of high speed, high density and low cost. The development and industrialization of optoelectronic devices based on organic materials, such as OLED, plastic optical fibers, organic thin-film lasers, polymer-based holographic optical memory, organic waveguide devices, organic transistors and field effect transistors, organic optical switches, etc., will push the optoelectronic industry to a new height.

1.3. The Theoretical Basis of This Book

1.3.1. *Photoelectric conversion materials and devices*

Photoelectric conversion is based on the principle of photoelectric effect, which is the phenomenon that matter emits electrons under the action of light radiation. The photoelectric phenomenon was first discovered by Hertz, but Einstein was the first to successfully explain the photoelectric effect. Einstein believed that the energy of a photon is transmitted to a single electron in a metal. When the electron absorbs a photon, it uses part of the energy to break away from the metal's bondage. The remaining part becomes the kinetic energy of

the electron after it leaves the metal surface. Einstein's photoelectric effect equation is as follows:

$$\frac{1}{2}mv^2 = h\nu - W, \quad (1.1)$$

where, h is Planck constant, ν is light frequency, $\frac{1}{2}mv^2$ is photoelectron kinetic energy, and W is the minimum energy required for photoelectron to escape from metal surface, which is called metal work. The emitted electrons are called photoelectrons.

The photoelectric effect can be divided into external photoelectric effect and internal photoelectric effect. External photoelectric effect refers to the phenomenon that the electrons in the object escape from the surface of the object and emit outward under the action of light, which is also called photoelectric emission effect. Internal photoelectric effect refers to the phenomenon that illumination changes the conductivity of an object or produces photoelectromotive force. It is also divided into photoelectric conductivity effect and photovoltaic effect (i.e. photovoltaic effect). Photoconductivity effect is a phenomenon that the electron absorbs photon energy from bonding state to free state under the action of light, which results in the change of material conductivity. When the light irradiates on the photoconductor, if the photoconductor is an intrinsic semiconductor material and the radiation energy is strong enough, the electrons in the valence band of the photoconductor will be stimulated to the conductivity of the photoconductor, which will increase the conductivity of the photoconductor. Photovoltaic effect refers to the phenomenon that light causes potential difference between different parts of inhomogeneous semiconductor or semiconductor bonding with metal.

Photoelectric effect and photovoltaic effect are quite different. By definition, photovoltaic effect is the precondition of photovoltaic effect. Photovoltaic effect is a special place where photovoltaic effect acts on semiconductor, which results in potential difference. In materials, the materials that produce photovoltaic effect can only be semiconductors, and photovoltaic effect materials can also be metals. Photovoltaic effect is a minority carrier process. It is the potential

difference between the two ends of PN junction when a few carriers absorb photons in semiconductor. The photovoltaic effect is that semiconductor or metal emits free electrons under the excitation of photons, and the surface emits electrons after overcoming the surface barrier. In photovoltaic effect, carriers cannot leave the material, while the latter can leave the material. The former has a certain absorption spectrum for the spectrum and is related to the intensity of light, while the latter has a cut-off wavelength of electron escape velocity independent of the intensity of light, only related to the frequency.

Solar energy is converted into electrical materials by photovoltaic effect, which is mainly used to make solar cells. The working principle of photoelectric conversion material is as follows: the same material or two different semiconductor materials are made into PN junction battery structure. When the solar light shines on the surface of PN junction battery structure material, the solar energy is converted into electricity through PN junction. The requirement of solar cells for photoelectric conversion materials is high conversion efficiency and large area devices, so as to better absorb sunlight. The main photoelectric conversion materials used are monocrystalline silicon, polycrystalline silicon and amorphous silicon. At present, the maximum energy conversion efficiency of inorganic silicon photovoltaic cells (photovoltaic cells) has reached 24%, and the energy conversion efficiency of GaAs semiconductor photovoltaic cells has even reached 32%. However, they are demanding on the purity and fabrication process of materials, and will produce some highly toxic substances in the manufacturing process. In addition, the non-flexibility and difficult processing of inorganic photovoltaic cells also limit their wide application process. The photovoltaic cells based on conjugated polymers, or polymer solar cells, not only have the advantages of flexible molecular design, simple fabrication method, low cost and large area flexible devices, but also inherit the advantages of high mobility and chemical stability of inorganic nanocrystalline carriers. It has attracted great attention.

1.3.2. Electro-optical conversion materials and devices

Solid light-emitting materials emit light under electric field, which is called electroluminescence. It is a process of converting electric energy directly into light energy. Devices made from this phenomenon are called electroluminescent devices, such as LED, liquid crystal display, semiconductor lasers, etc. LED uses solid semiconductor chip as luminescent material, which emits excess energy through carrier recombination in semiconductor, and directly emits red, yellow, blue and green light. On this basis, using the principle of three primary colors and adding phosphor, any color of light can be emitted, such as red, yellow, blue, green, blue, orange, purple, white, etc. It has the characteristics of small size, low power consumption, long service life, high brightness, low heat, environmental protection and durability. LED lamp is a kind of illumination device which uses LED as light source. It has been widely used in illumination, display, indication and other fields. OLED, also known as organic electro-excitation light, uses organic polymer materials as semiconductor materials in light-emitting diodes. OLED technology has the characteristics of self-luminescence. It uses very thin organic coatings and glass substrates. When current passes through, these organic materials will emit light, and OLED display screen has a large visual angle, and can save electricity. OLED display technology is widely used in mobile phones, digital cameras, DVD players, PDA, notebook computers, car audio and television, etc.

1.3.3. Transparent conductive material

Transparent conductive materials are a kind of special materials with high transmittance to visible light and high conductivity, such as indium-tin oxide (ITO), nanowire, graphene and so on. Transparent conductive materials are widely used because of their unique photoelectric properties. In many modern electronic information technology, photoelectric technology, new energy technology and national defense technology, the design, manufacture and use of

transparent conductive materials is an indispensable key technology. For example, transparent conductive materials are indispensable electrode materials in touch screen. In photovoltaic solar cells, high transmittance electrode materials can ensure that photovoltaic cells can fully absorb sunlight. Coating a layer of high transparency and high conductivity film on the surface of building glass can effectively prevent the infrared radiation from entering or escaping, and greatly improve the energy-saving effect of the building. Transparent conductive films can also effectively shield electromagnetic radiation, so they have important applications in aviation, communications and other national defense industries.

References

- An Y *et al.*, 2011. *Photoelectronic Technique*, Beijing: Electronics Industry Press.
- Hou H, 2012. *Optoelectronic Materials and Devices*, Beijing: National Defence Industry Press.
- Jiang W, Shi J, 2009. *Photoelectronic Technique*, Beijing: Science Press.
- Shi S, Liu J, 2010. *Optoelectronic Technology and Its Application*, Beijing: Science Press.
- Wang X, Ye C, 2013. *Organic Photoelectric Materials and Devices*, Beijing: Chemical Industry Press.
- Zhu J, 2009. *Fundamentals of Optoelectronic Technology*, Beijing: Science Press.

Chapter 2

Theoretical Basis of Photoelectric Materials

2.1. Energy Band Theory

The energy spectra of electrons in a complete crystal are bands of dense energy levels, which are called bands. The energy band is separated from the energy band by the energy forbidden zone. Among them, the lowest energy band completely empty at 0K is called conduction band, and the highest energy band completely occupied by electrons is called valence band. The energy gap between them is called forbidden band. Energy band theory, also known as solid energy band theory, is a quantum mechanics theory about the motion state of electrons in crystals. It predicts that the energy of electrons in crystals will always fall in some limited range or “energy band”. The electrical, optical and magnetic properties of crystals are all related to the motion of electrons. The energy band theory is used to study these problems. The band theory successfully explains the difference between metals, semiconductors and insulators and the Hall effect phenomenon. Semiconductor physics is based on energy band theory.

With the development of experimental technology, the electronic band structures of many crystals have been successfully determined by means of cyclotron resonance, electro-optic, magneto-optic and spectroscopy. Especially in recent years, due to the wide application of computer technology, the band structure of electrons can be calculated more accurately in theory. Nevertheless, since band theory is a simplified approximation theory after all, it can only be applied to

ordered crystals, and even for ordered crystals, when their structures are complex, band theory is often difficult to deal with.

2.1.1. *Schrödinger equation of crystal and its approximate solution*

2.1.1.1. *Schrödinger equation*

Crystals consist of a large number of atoms arranged periodically, and atoms consist of nuclei and electrons. Since the inner electrons do not participate in the physical process of the crystal, it can be presumed that the crystal is composed of the outermost electrons of the atom and ions that lose electrons. If $r_1, r_2, r_3, \dots, r_j \dots$ denotes the potential vectors of electrons, $R_1, R_2, R_3, \dots, R_j \dots$ denoting the potential vectors of an ion that has lost its electron, then the steady-state Schrödinger equation of the crystal is

$$\hat{H}\psi = E\psi, \quad (2.1)$$

where ψ is a wave function, E is an eigenvalue of energy, $\hat{H}$ is a Hamiltonian operator, and

$$\hat{H} = T_e + T_Z + \mathbf{u}_e + \mathbf{u}_Z + \mathbf{u}_{eZ} + \mathbf{V} \quad (2.2)$$

where $T_e = \sum_i (-\frac{h^2}{2m}\Delta_i^2)$ is the kinetic energy operator of all electrons, m is the electronic mass, and $\Delta_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}$ is the Laplacian operator of the i th electron,

$$T_z = \sum_{\alpha} T_{\alpha} = \sum_{\alpha} \left(-\frac{h^2}{2M_{\alpha}} \Delta_{\alpha}^2 \right)$$

is the kinetic energy operator for all ions, M_{α} is the mass of ions, and Δ_{α}^2 is the Laplace operator for the α th ion,

$$\hat{\mathbf{u}}_e = \frac{1}{2} \sum_{i,j \neq i} \frac{e^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_j|} = \frac{1}{2} \sum_{i,j \neq i} \hat{\mathbf{u}}_{ij}$$

represents the interaction energy between electrons,

$$\hat{\mathbf{u}}_Z = \frac{1}{2} \sum_{\alpha, \beta \neq \alpha} \frac{z_\alpha z_\beta e^2}{4\pi\epsilon_0 |\mathbf{R}_\alpha - \mathbf{R}_\beta|} = \frac{1}{2} \sum_{\alpha, \beta \neq \alpha} \hat{\mathbf{u}}_{\alpha\beta}$$

denotes the interaction energy between ions, and $z_\alpha e$ and $z_\beta e$ are the charge quantities of α and β ions, respectively, and

$$\hat{\mathbf{u}}_e Z = - \sum_{i, \alpha} \frac{z_\alpha e^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{R}_\alpha|} = \sum_{i, \alpha} \hat{\mathbf{u}}_{i\alpha}$$

represents the interaction energy between electrons and ions, and

$$\hat{\mathbf{V}} = V(\mathbf{r}_1 \cdot \mathbf{r}_2, \dots, \mathbf{r}_n, \mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N)$$

is the potential energy of all electrons and ions in the field.

The density of atoms in crystals is about $5 \times 10^{22} \text{ cm}^{-3}$, so the above equation cannot be solved strictly. Generally, the single-electron approximation method is used.

2.1.1.2. *Adiabatic approximation and atomic valence approximation*

- (1) *Adiabatic approximation*: Generally, the average kinetic energy of heavy particles (e.g. nuclei) is the same order of magnitude as that of light particles (e.g. extranuclear electrons). Because $M_\alpha \gg m$, the electron velocity is much larger than that of the nucleus (about two orders of magnitude), so it is possible to consider the motion of the electron in the crystal separately from that of the nucleus. This simplification assumes that the overall motion of atoms has a weak influence on the motion of electrons, just as there is no exchange of energy between the overall motion of atoms and the motion of electrons. This simplification is usually called adiabatic approximation.

Furthermore, if we assume that the nucleus is fixed, then the nuclear coordinates are no longer variables, but $R_{10}, R_{20}, \dots, R_{\alpha 0}, \dots$. The form of R_{N0} appears to represent the coordinates of lattice points. In this case, the nuclear kinetic energy is zero,

and the interaction energy $\mathbf{u}_Z$ is constant and can be chosen as zero. In addition, if there is no outfield, there is $\mathbf{V} = 0$.

At this time, the Schrödinger equation of crystal can be simplified to describe the electron motion equation in a fixed nuclear field.

$$\begin{aligned}\hat{\mathbf{H}}\psi_e &= (\hat{\mathbf{T}}_e + \hat{\mathbf{u}}_e + \hat{\mathbf{u}}_e z)\psi_e = \left[\sum_i \left(-\frac{\hbar^2}{2m} \nabla_i^2 \right) \right. \\ &\quad \left. + \frac{1}{2} \sum_{i,j \neq i} \frac{e^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,a} \frac{z_a e^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{R}_a|} \right] \psi_e \\ &= E_e \psi_e.\end{aligned}\tag{2.3}$$

- (2) *Atomic valence approximation*: To further simplify the above equation, the so-called atomic valence approximation is adopted, that is, all electrons except valence electrons form fixed ionic substances with their nuclei.

2.1.1.3. *Single-electron approximation — Hartree–Fock method*

Crystals contain a large number of electrons, which belong to the multi-electron system. Every electron in the system is subject to Coulomb action of other electrons. So even if we only study the motion of electrons, it is still very complicated. At present, the most effective way to deal with multi-electron problems is the so-called single-electron approximation method, which considers the motion of each electron separately. The single-electron approximation method is also known as the Hartree–Fock method. In this method, in order to approximate the separation of the motion of each electron, an appropriate simplification is adopted: when studying the motion of an electron, the Coulomb effect of other electrons on the electron everywhere in the crystal is considered equally according to their probability distribution. This average consideration is achieved by introducing a self-consistent electron field. For example, for the i th electron, supposing that with the aid of an external potential field, a potential field can be applied to the position of the electron at any time with the same effect as other electrons, which is recorded as Ω_i ,

then Ω_i is only related to the potential vector r_i of the i th electron, and can be recorded as $\Omega_i = \Omega_i(r_i)$, which is called self-consistent electron field. If all other electrons are treated in the same way, there are

$$\sum_i \Omega_i(\mathbf{r}_i) = \frac{1}{2} \sum_{i,j \neq i} \hat{\mathbf{u}}_{ij} = \frac{1}{2} \sum \frac{e^2}{4\pi\epsilon_0 |\mathbf{r}_i - \mathbf{r}_j|} \quad (2.4)$$

Assuming that $\Omega_i(\mathbf{r}_i)$ is known, the Hamiltonian operator of the system can be written as

$$\begin{aligned} \hat{\mathbf{H}}_e &= \sum_i -\frac{\hbar^2}{2m} \nabla_i^2 + \frac{1}{2} \sum_{i,j \neq i} \hat{\mathbf{u}}_{ij} + \sum_{i,a} \hat{\mathbf{u}}_{i,a} \\ &= \sum_i -\frac{\hbar^2}{2m} \nabla_i^2 + \sum_i \Omega_i(\mathbf{r}_i) + \sum_i \left(\sum_a \hat{\mathbf{u}}_{i,a} \right) \\ &= \sum_i \hat{\mathbf{H}}_i. \end{aligned} \quad (2.5)$$

So for the i th electron, the Hamiltonian operator is

$$\hat{\mathbf{H}}_i = \frac{\hbar^2}{2m} \nabla_i^2 + \Omega_i(\mathbf{r}_i) + \sum_a \hat{\mathbf{u}}_{ia} = -\frac{\hbar^2}{2m} \nabla_i^2 + \Omega_i(\mathbf{r}_i) + \hat{\mathbf{u}}_i(\mathbf{r}_i), \quad (2.6)$$

where $\mathbf{u}_i(\mathbf{r}_i)$ is the potential energy of the i th electron in all ion fields, and $\Omega_i(\mathbf{r}_i)$ is the potential energy of the i th electron in all other electron fields. Thus the eigenfunction of the system can be expressed as the product of each electron wave function and the total energy is the sum of the energy of each electron:

$$\psi, (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_a) = \prod_i \psi_i(\mathbf{r}_i), \quad (2.7a)$$

$$E_e = \sum_i E_i \quad (2.7b)$$

where, $\psi_i(r_i)$ and E_i satisfy the Schrödinger equation of a single electron:

$$\hat{\mathbf{H}}_i\psi_i(\mathbf{r}_i) = E_i\psi_i(\mathbf{r}_i). \quad (2.8)$$

By introducing the concept of self-consistent electron field, the multi-electron problem can be transformed into a single-electron problem. Since the i th electron can be any electron, the above single-electron equation can be generally expressed as

$$\hat{\mathbf{H}}\psi(\mathbf{r}) = E\psi(\mathbf{r}). \quad (2.9)$$

Here $\hat{\mathbf{H}} = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})$, $V(\mathbf{r}) = \Omega(\mathbf{r}) + \hat{\mathbf{u}}(\mathbf{r})$, where $-\frac{\hbar^2}{2m}\nabla^2$ is the kinetic energy operator of a single electron; $V(\mathbf{r})$ is its potential energy operator, which contains the average Coulomb interaction energy of all other electrons to it and the Coulomb interaction energy of all ions to it.

For specific crystals, as long as the potential function $V(\mathbf{r})$ is written, a series of energy spectrum values E and corresponding wave function $\psi(r)$ can be found by solving Schrödinger equation in principle.

2.1.1.4. Atomic orbits and lattice orbits

There are two kinds of single-electron wave functions for electrons in crystals, one is atomic orbital and the other is lattice orbital. In atomic orbits, electrons do not get rid of the bondage of atoms, and basically move around atoms. Their wave functions are larger only near individual atoms. Atomic orbits are suitable for internal electrons in crystals. In lattice orbits, electrons move not only around each atom, but also between atoms. They communalize in the whole crystal, and their wave functions extend throughout the crystal. Lattice orbits are suitable for external electrons.

The external electrons in crystals are usually concerned, and lattice orbits are generally chosen. In addition, it is considered that all atoms are stationary in their equilibrium positions. Therefore, the potential energy $V(\mathbf{r})$ of the external electrons should have the symmetry of the lattice, especially the periodicity.

2.1.1.5. State distribution of electrons

After finding all possible energy spectra and motion states of a single electron, if we know the distribution of a large number of electrons in these states, the problem of electron motion in a crystal will be solved.

The distribution of electrons in states is a quantum statistical problem. In the case of thermal equilibrium, the distribution of electrons in the state is approximately determined by the Fermi–Dirac distribution. New distribution functions can also be found in unbalanced cases.

2.1.2. Bloch's theorem

The wave equation of a single electron in a crystal is

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}) \right] \psi(\mathbf{r}) = E\psi(\mathbf{r}),$$

where the potential function $V(\mathbf{r})$ has the microcosmic symmetry of the lattice, especially the periodicity of the lattice. For example, the form of electron potential function in one-dimensional periodic potential field is shown in Fig. 2.1.

Bloch's theorem: If $V(\mathbf{r})$ has lattice periodicity, i.e., $V(\mathbf{r} + \mathbf{R}_m) = V(\mathbf{r})$, the solution of Schrödinger equation of crystal can be written

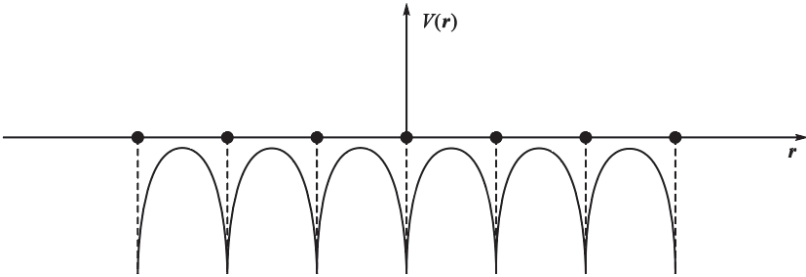


Fig. 2.1. Electrons moving in one-dimensional periodic potential field.

in the form of Bloch's function as follows.

$$\psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u(\mathbf{r}), \quad (2.10)$$

where k is called wave vector, which is a real number. Here $u(\mathbf{r})$ is a function with lattice periodicity, i.e.

$$u(\mathbf{r} + \mathbf{R}_m) = u(\mathbf{r}),$$

where $\mathbf{R}_m$ is the lattice vector.

Another common form of Bloch's theorem is

$$\psi(\mathbf{r} + \mathbf{R}_m) = e^{i\mathbf{k}\cdot\mathbf{R}_m}\psi(\mathbf{r}). \quad (2.11)$$

The formula shows that the wave function $\psi(\mathbf{r})$ in the periodic potential field is translated by any lattice vector $\mathbf{R}_m$, and the wave function $\psi(\mathbf{r} + \mathbf{R}_m)$ is obtained. There is only one constant factor of modulus 1 between the two wave functions.

In a word, the electron wave function in periodic field can be generally expressed as the product of a plane wave and a periodic factor. The wave vector of plane wave is the real vector k , which can be used to indicate the motion state of electrons. Different k represents different states. Therefore, k also acts as a quantum number. For clarity, an index k is added to the wave function and energy spectrum value (eigenvalue).

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}}(\mathbf{r}), \quad (2.12a)$$

$$E = E(k). \quad (2.12b)$$

It can be seen from the formula that the wave vector k can only be taken as a real value in order to make the electron wave propagate in the whole crystal undamped. We can give a rough explanation of the wave function: the plane wave factor $e^{i\mathbf{k}\cdot\mathbf{r}}$ is the same as the free electron wave function, which describes the movement of electrons among the primitive cells; the periodic factor $u_{\mathbf{k}}(\mathbf{r})$ describes the movement of electrons in a single cell, because it only repeats periodically among the primitive cells. Because

$$|\psi_{\mathbf{k}}(\mathbf{r} + \mathbf{R}_m)|^2 = |e^{i\mathbf{k}\cdot\mathbf{R}_m}\psi_{\mathbf{k}}(\mathbf{r})|^2 = |\psi_{\mathbf{k}}(\mathbf{r})|^2.$$

The results show that the probability of electrons appearing at the corresponding points in each cell is equal.

Because the momentum operator $-i\hbar\nabla = \frac{\hbar}{i}\nabla$ and H of electrons in crystals are not interchangeable, their wave functions are not only plane waves, but also have a periodic factor. The product of wave vector k and $\hbar$ has the dimension of momentum. For electrons moving in periodic fields, $\hbar k$ is usually referred to as the “quasi-momentum” of electrons, which is expressed by p : $p = \hbar k$. Quasi-momentum is also called lattice momentum.

2.1.3. Periodic boundary conditions

According to Bloch’s theorem, the wave function of an electron in a periodic field can be expressed as the product of a plane wave and a periodic factor. After considering the boundary conditions, k is limited and only intermittent values can be taken. The size of the actual crystal is always limited. The environment of the electron in the cell near the surface is different from that in the corresponding position of the inner cell, so it is periodically destroyed, which brings some inconvenience to the theoretical analysis. In order to overcome this difficulty, Born–Carmen periodic boundary conditions are usually used: if an infinite crystal is generated by periodic repetition of finite crystals, and requiring the motion of electrons to repeat periodically in space with finite crystals as periodic periods.

It is assumed that the finite crystal considered is a parallel hexahedron with N_1 primordia along the a_1 direction, N_2 primordia in the a_2 direction and N_3 primordia in the a_3 direction. The total number of primordia is $N = N_1 N_2 N_3$. According to the periodic boundary condition, the wave function along the $\mathbf{a}_1$ direction is required to take $N_1 \mathbf{a}_1$ as the period, that is to say

$$\begin{aligned}\psi(\mathbf{r} + N_1 \mathbf{a}_1) &= \psi(\mathbf{r}) = e^{i\mathbf{k} \cdot N_1 \mathbf{a}_1} \psi(\mathbf{r}) \implies e^{i\mathbf{k} \cdot N_1 \mathbf{a}_1} \\ &= 1 \implies \mathbf{k} \cdot N_1 \mathbf{a}_1 = 2\pi.\end{aligned}$$

Let $k = \beta_1 b_1 + \beta_2 b_2 + \beta_3 b_3$, because $\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\delta_{ij}$, there is

$$\mathbf{k} \cdot N_1 \mathbf{a}_1 = 2\pi\beta_1 N_1 = 2\pi l_1 \implies \beta_1 = \frac{l_1}{N_1} \quad (l_1 \text{ is an arbitrary integer}).$$

Similarly, there are

$$\beta_2 = l_2/N_2, \beta_3 = l_3/N_3 \quad (l_2, l_3 \text{ are arbitrary integers}).$$

So, there are

$$\mathbf{k} = \frac{l_1}{N_1}\mathbf{b}_1 + \frac{l_2}{N_2}\mathbf{b}_2 + \frac{l_3}{N_3}\mathbf{b}_3. \quad (2.13)$$

Under periodic boundary conditions, $\mathbf{k}$ can only take discontinuous values, so the energy $E(k)$ corresponding to these wave vectors can only take discontinuous values. The wave vectors k determined by formula (2.13) are located at the top angles of some parallel hexahedrons with three sides of $b_1/N_1, b_2/N_2$ and b_3/N_3 in the inverted space. In inverted space, the volume of the representative point of each wave vector k is

$$\frac{\mathbf{b}_1}{N_1} \cdot \left(\frac{\mathbf{b}_2}{N_2} \times \frac{\mathbf{b}_3}{N_3} \right) = \frac{\Omega^x}{N_1 N_2 N_3} = \frac{(2\pi)^3/\Omega}{N} = \frac{(2\pi)^3}{N\Omega} = \frac{(2\pi)^3}{V}, \quad (2.14)$$

where V is the volume of the whole finite crystal, so the wave vector in the unit inverted space is $V/(2\pi)^3$, which is the distribution density of the representative point of k in the inverted space. So the number of representations of k in each inverted cell is

$$\frac{\Omega^* V}{(2\pi)^3} = \frac{(2\pi)^3 N \Omega}{(2\pi)^3 \Omega} = N. \quad (2.15)$$

In each inverted cell, the number of representative points of k is equal to the total number of primitive cells of crystal N . This is an important conclusion derived from periodic boundary conditions. Each wave vector k represents a space motion state (quantum state) of electrons in a crystal, so that the number of states of electrons in the range of $dk = dk_x dk_y dk_z$ is the number of states of electrons.

$$\frac{V}{(2\pi)^3} d\mathbf{k} = \frac{V}{(2\pi)^3} dk_x dk_y dk_z. \quad (2.16)$$

2.1.4. Energy bands and their general properties

2.1.4.1. Energy band

The wave function of electron motion in crystal is Bloch function.

$$\psi_k(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_k(\mathbf{r}). \quad (2.17)$$

Given a k , the plane wave part is determined. In order to determine $u_k(\mathbf{r})$, the wave equation needs to be solved.

$$\begin{aligned} \hat{H}\psi_k(\mathbf{r}) &= \left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) \right] \psi_k(\mathbf{r}) = E(\mathbf{k})\psi_k(\mathbf{r}) \\ \Rightarrow \left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) \right] e^{i\mathbf{k}\cdot\mathbf{r}} u_k(\mathbf{r}) &= E(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{r}} u_k(\mathbf{r}) \\ \Rightarrow \left[-\frac{\hbar^2}{2m}(\nabla^2 + i2\mathbf{k} \cdot \nabla - k^2) + V(\mathbf{r}) \right] u_k(\mathbf{r}) &= E(\mathbf{k})u_k(\mathbf{r}) \\ \Rightarrow \left[-\frac{\hbar^2}{2m} \left(\frac{1}{i}\nabla + \mathbf{k} \right)^2 + V(\mathbf{r}) \right] u_k(\mathbf{r}) &= E(\mathbf{k})u_k(\mathbf{r}). \end{aligned} \quad (2.18)$$

The above formula is the differential equation satisfied by $\mathbf{u}_k(\mathbf{r})$, and $\mathbf{u}_k(\mathbf{r} + \mathbf{R}_m) = \mathbf{u}_k(\mathbf{r})$. For a given problem, $V(\mathbf{r})$ is certain. When k is given, the form of differential equation is determined. For the eigen equation of this property, there can be many separated energy spectrum values:

$$E_1(k), E_2(k), \dots, E_n(k). \quad (2.19)$$

By substituting these energy spectra into differential equations, the corresponding functions $u_k(r)$ can be obtained.

$$u_{1,\mathbf{k}}(\mathbf{r}), u_{2,\mathbf{k}}(\mathbf{r}), \dots, u_{n,\mathbf{k}}(\mathbf{r}). \quad (2.20)$$

These functions are multiplied by the plane wave factor $e^{i\mathbf{k}\cdot\mathbf{r}}$ to obtain the corresponding wave functions:

$$\psi_{1,k}(r), \psi_{2,k}(r), \dots, \psi_{n,k}(r). \quad (2.21)$$

The above relationship can be abbreviated as

$$E_n(k), \psi_{n,k}(r) = e^{ikr} u_{n,k}(\mathbf{r}) \quad (n = 1, 2, 3, \dots). \quad (2.22)$$

The electronic energy spectrum value $E_n(k)$ in the crystal has the following properties.

- (i) $E_n(-k) = E_n(k)$, that is, $E_n(k)$ has inversion symmetry. In particular, for one-dimensional cases, $E_n(k)$ is a dual function.
- (ii) $E_n(k + K_1) = E_n(k)$, where K_1 is a reciprocal lattice vector and $K_1 = l_1 b_1 + l_2 b_2 + l_3 b_3$. This is because the physical meaning of k and $k + K_1$ is equivalent.

Therefore, two quantum numbers n and k are needed to mark the state of electron motion and the corresponding energy spectrum in the crystal.

Because $\mathbf{k} = \frac{l_1}{N_1} \mathbf{b}_1 + \frac{l_2}{N_2} \mathbf{b}_2 + \frac{l_3}{N_3} \mathbf{b}_3$ takes the discrete value, $E_n(k)$ is a quasi-continuous energy band, that is, the relationship between $E_n(k)$ and k is quasi-continuous; index n is the label of energy band, different n corresponds to different energy band $E_n(k)$; k is the label of different states and levels in each energy band, and each k is represented by a point in inverted space, which is the starting point of vector k at the origin point. The idea is at the end. For each energy band, a point in the inverse space can represent a single-electron state and level, and the number of k points is N . Figure 2.2 shows the band structure of quasi-free electrons in one-dimensional case:

$$E_n(\mathbf{k}) = \frac{\hbar^2 k^2}{2m}.$$

2.1.4.2. General characteristics of energy bands

- (1) $V(\mathbf{r})$ has the periodicity of the lattice: $V(\mathbf{r} + \mathbf{R}_m) = V(\mathbf{r})$.
- (2) $E_n(k)$ has the periodicity of reciprocal lattice: $E_n(k + K_1) = E_n(k)$, and K_1 is any reciprocal lattice vector.
- (3) The wave function also has the periodicity of inverted lattice: $\psi_{n,k+K_1}(\mathbf{r}) = \psi_{n,k}(\mathbf{r})$.
- (4) $E_n(k)$ has inversion symmetry: $E_n(-k) = E_n(k)$.

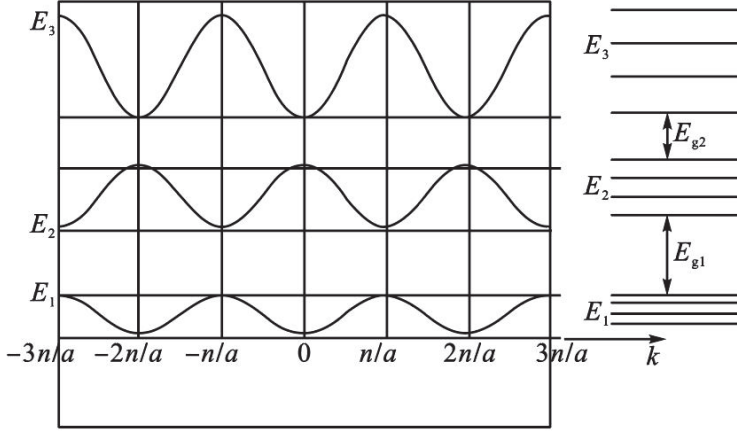


Fig. 2.2. Schematic diagram of band structure of one-dimensional quasi-free electrons.

- (5) $E_n(k)$ has the macroscopic point group symmetry of crystals: $E_n(\alpha k) = E_n(k)$, and α is any macroscopic point group symmetry operation of crystals. Here αk represents another wave vector obtained by rotation or inversion of k , which is equal to their corresponding energy spectrum. It should be noted that although the energy spectra corresponding to αk and k are equal here, the wave functions are generally independent. This means that the symmetry of energy bands can cause degeneracy of energy levels, but only the states corresponding to k and αk , which differ from each other by one reciprocal vector K_1 , are consistent.

The wave vectors k and k' ($k' = k + K_1$) of a reciprocal lattice vector differ from each other, indicating the same electronic states and calling them equivalent, while the wave vectors k and k' ($k' = \alpha k$), which are connected by the point symmetry operation, have the same energy spectrum value and are called symmetrical.

- (6) In each energy band, the number of space wave functions $\psi_{n,k}(\mathbf{r})$ of electrons is N , and N is the total cell number of crystals. Considering the two possible orientations of electron spin, the

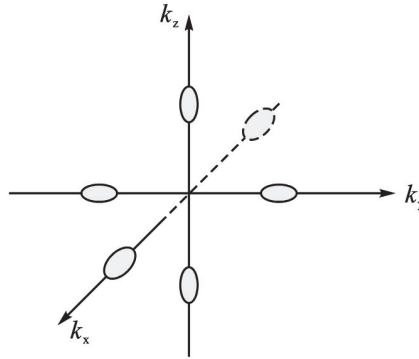


Fig. 2.3. Ellipsoidal isoelectric surface near the conduction band minimum of silicon.

number of states of each band is 2 times the total number of crystal primitives and $2N$.

- (7) It can be inferred from the symmetry of the energy band that the extreme value of the energy band is symmetrically distributed in the inverted space, and the wave vectors are connected by the symmetrical operation. In inverted space, the surface composed of representative points with equal energy is called isosurface, and the location where the minimum energy appears is called valley. It can be seen from $E_n(\alpha k) = E_n(k)$ that the isoenergetic surfaces of inverted space coincide with each other under the macroscopic point symmetry operation of the crystal. For example, the isoenergetic surface near the conduction band minimum of silicon is a revolving ellipsoid. As shown in Fig. 2.3, it has the point group symmetry of the cube and therefore has six valleys symmetrical to each other.

2.1.5. Brillouin zone

The change of energy band $E_n(k)$ in inverted space is symmetrical. For those wave vectors associated with symmetrical operation of macroscopic point groups of crystals, the corresponding energy spectrum values are equal. The inverted cell can be used to analyze

the periodicity of the band, but it is not suitable to discuss the symmetry of the band. In inverted space, the representative points of K linked by symmetric operations are generally not in the same inverted cell. Therefore, it is necessary to divide the inverted space into some periodic and symmetrical repetitive elements, Brillouin region.

2.1.5.1. Brillouin Zoning method

In inverted space, the mid-vertical plane connecting the origin with all inverted lattice points are shown in Fig. 2.4. These planes divide the inverted space into several regions. The region closest to the origin is called the first Brillouin region, and the regions closest to the origin constitute the second Brillouin region. By analogy, we can get the third and fourth Brillouin zone. It can also be said that the area surrounded by the interface near the origin is the first Brillouin zone, the area entering through an interface from the origin is the second Brillouin zone, and the area entering after passing through the interface $(n-1)$ is the n Brillouin zone. The representative points of $\mathbf{k}$ on the boundary of Brillouin zone are located on the mid-vertical plane of the inverted lattice vector K_n and satisfy the plane equation.

$$k \cdot \left(\frac{\mathbf{K}_n}{K_n} \right) = \frac{1}{2} K_n \text{ or } \mathbf{k} \cdot \mathbf{K}_n = \frac{1}{2} K_n^2. \quad (2.23)$$

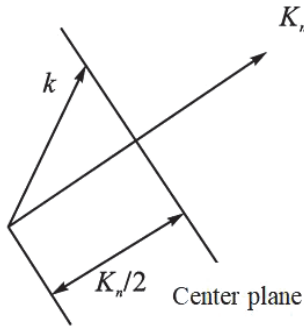


Fig. 2.4. Mid-vertical plane.

2.1.5.2. *Characteristics of Brillouin region*

- (1) The volume of each Brillouin region is equal and equal to that of an inverted cell.
- (2) After translating the proper inverse lattice vector $\mathbf{K}_n$, the parts of each Brillouin region can coincide with another Brillouin region.
- (3) Each Brillouin region is symmetrically distributed around the origin, and has the point group symmetry of positive and inverse lattices.

2.1.5.3. *Simple Brillouin zone*

In order to find the independent states in each band, it is only necessary to limit k to a Brillouin region. The first Brillouin zone is the most convenient to use. It is usually called the simple Brillouin zone.

2.1.6. *Metals, semiconductors and insulators*

An important aspect of the band theory's success lies in its ability to explain why some elements combine into crystals to form good conductors, while others form semiconductors or insulators. The physical properties of conductors and insulators are very different. For example, at 1 K, the resistivity of good conductors (excluding superconductors) can be as low as $10^{-10} \Omega \cdot \text{cm}$, while the resistivity of good insulators can be as high as $10^{22} \Omega \cdot \text{cm}$.

Metals are generally conductors, and the conductivity decreases with the increase of temperature. Semiconductors have poor conductivity, and the conductivity increases rapidly with the increase of temperature. Insulators have the worst conductivity and are basically non-conductive. These differences can be well explained by using energy band theory.

2.1.6.1. *Full-filled band and partially-filled band*

The contribution of an electron to the current density in the crystal is $j_k = \frac{-e}{V} \mathbf{v}$, and the total current density is $j = -e \frac{2}{(2\pi)^3} \int f \mathbf{v} d\mathbf{k}$.

Because $E(-k) = E(k)$, for a pair of electrons in the $\pm k_0$ state, their velocities are equal and opposite. According to the formula of average electron velocity, there are

$$\begin{aligned}\mathbf{v}(-\mathbf{k}_0) &= \frac{1}{\hbar} [\nabla_{\mathbf{k}} E(\mathbf{k})]_{k=-k_0} = \frac{1}{\hbar} [\nabla_{-k'} E(-\mathbf{k}')]_{k'=k_0} \\ &= -\frac{1}{\hbar} [\nabla_{k'} E(\mathbf{k}')]_{k'=k_0} = -\mathbf{v}(k_0).\end{aligned}$$

- (1) In the absence of an external field, the electrons are in thermal equilibrium, and the distribution function f is only a function of energy E . The wave vector is in the state of $\pm k$, and the corresponding energy is equal, so the probability of occupation by electrons is equal. In this pair of states, the electron velocities are equal in magnitude and opposite in direction, so the contributions to the current cancel each other out. There is no current flow in the crystal, as shown in Fig. 2.5.
- (2) When there is an electric field ε ,

$$\frac{d\mathbf{k}}{dt} = \frac{1}{\hbar} \mathbf{F} = \frac{-e}{\hbar} \varepsilon.$$

For full-filled band case, the change of electrons in Brillouin region under the action of electric field is shown in Fig. 2.6,

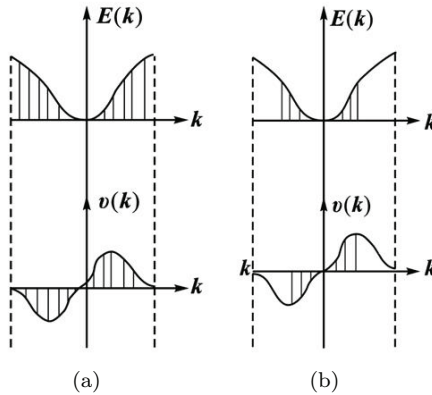


Fig. 2.5. $E(\mathbf{k})$ and $v(\mathbf{k})$ diagrams without outfield: (a) full-filled band condition and (b) partial-filled band condition.

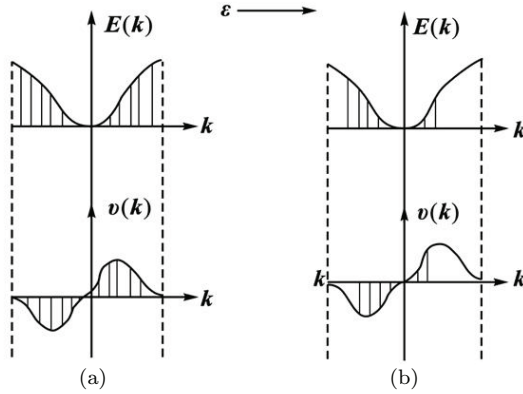


Fig. 2.6. $E(k)$ and $v(k)$ diagrams with outfield: (a) Full-filled band condition and (b) partial-filled band condition.

$\pm k$ states occupy electrons at the same time, so the contribution to current is zero.

For partial-filled cases, there are more electrons in the state of wave vector k opposite to the electric field, less electrons in the same state as the electric field, or more electrons in the opposite direction of v and ε , and fewer electrons in the same direction of ε . Electrons are negatively charged, resulting in a net current along the electric field in the crystal.

2.1.6.2. Metals, semiconductors and insulators

The difference between semiconductor, insulator and metal depends on whether there is a partially filled unsatisfactory band at 0 K. The two basic conditions for determining whether a crystal is a semiconductor or an insulator are as follows.

- (1) Electrons are sufficient to fill integer bands. If there are N primitive cells in a crystal, considering the two orientations of electrons, each band can hold $2N$ electrons, and the total number of electrons in a crystal is multiplied by the number of primitive cells, the condition can be expressed as

$$\text{Number of electrons in each cell} * N/2N = \text{integer},$$

that is, the number of electrons in each cell should be even.

- (2) There is an energy gap between the highest energy of occupied band by electrons and the higher energy band, which does not overlap. If this condition is not satisfied, electrons can be filled into overlapping bands so that they cannot be full-filled.

One of the above two items is not satisfied, that is, it may be metal.

The difference between semiconductor and insulator lies in the narrow bandgap of the former, which is generally less than 3 eV.

The above discussion is applicable to most crystals, but not to some transition metal oxides. For example, cobalt oxide (CoO) is a semiconductor material rather than a metal. Although the number of electrons in each cell of cobalt oxide is odd, the single-electron approximation and the shared motion model cannot be simply applied to the case where the d-electron motion is tightly bound by atoms in such materials. This shows that the band theory has limitations.

2.1.7. *Electrons, holes and carriers*

For semiconductors and insulators, when the temperature rises from 0 K, there will always be a small number of electrons, which jump from the highest occupied band to the adjacent empty band due to thermal excitation. There are also a few electrons in the previously empty adjacent energy bands that can conduct electricity. Usually, the lowest empty band on the highest full band is called conduction band; the former full maximum band is now in empty state, and electrons have room to move and can conduct electricity. The highest occupied bands are usually called valence bands because they are the bands occupied by valence electrons forming chemical bonds. The void state in the valence band is called a void. Electrons in conduction band and holes in valence band can conduct current, so they are collectively called carriers.

- (i) *Holes*: Thermal excitation causes a part of the valence band electrons to jump to the conduction band and form an empty state.

In order to analyze the problem conveniently, a hypothetical particle called a hole is introduced correspond to the empty state in the valence band.

(ii) *Hole current*: The current density caused by an empty state in a full band, that is, the current density of a hole, is equal in magnitude and opposite in direction to that caused by an electron in the same corresponding state. If the hole current density is j_h , then there is

$$j_h + j_e = 0. \quad (2.24)$$

In the formula, $\mathbf{j}_e = \frac{-e}{V} \mathbf{v}(\mathbf{k})$ is the current density of electrons. Thus, the hole current density is

$$\mathbf{j}_h = -j_e = \frac{e}{V} \mathbf{v}(\mathbf{k}). \quad (2.25)$$

This corresponds to the charge carried by the hole (+e) and moves at the electron speed corresponding to the state $\mathbf{k}$. If the average velocity of a hole is $v_h(\mathbf{k})$, then there is

$$\mathbf{v}_h(\mathbf{k}) = \frac{1}{\hbar} \nabla_k E(\mathbf{k}). \quad (2.26)$$

(iii) *Cavitation acceleration*: When there is an external electric field, the velocity of change of electrons in inverted space is

$$\frac{d\mathbf{k}}{dt} = \frac{1}{\hbar} \mathbf{F} = \frac{-e}{\hbar} (\boldsymbol{\varepsilon} + \mathbf{v} \times \mathbf{B}).$$

The acceleration is

$$a = \frac{d\mathbf{v}}{dt} = \frac{1}{\hbar} \nabla_k \nabla_k E(\mathbf{k}) \cdot \frac{d\mathbf{k}}{dt} = \frac{1}{\hbar^2} \nabla_k \nabla_k E(\mathbf{k}) \cdot (-e)(\boldsymbol{\varepsilon} + \mathbf{v} \times \mathbf{B}).$$

Because electrons tend to occupy low energy states, the empty states are near the top of the full band. Consider the case where the isoenergetic surface near the top of the valence band is spherical. At this time

$$E(\mathbf{k}) = E_v + \frac{\hbar^2 k^2}{2m_e}, \quad (2.27)$$

where E_v is the valence band top energy, and m_e is the effective mass of electrons. Since E_v is the maximum energy value of the valence band, the second derivative at this point is less than zero, so there is m_e less than zero. For $m_h = -m_e$, the effective mass of the hole is

$$E(\mathbf{k}) = E_v - \frac{\hbar^2 k^2}{2m_h}. \quad (2.28)$$

$$\Rightarrow \frac{d\mathbf{v}_h}{dt} = \frac{-e}{-m_h}(\boldsymbol{\varepsilon} + \mathbf{v} \times \mathbf{B}) = \frac{e}{m_h}(\boldsymbol{\varepsilon} + \mathbf{v} \times \mathbf{B}). \quad (2.29)$$

(iv) *Hole energy*:

$$E_h(k) = -E_e(k)$$

2.2. The Process of Absorption in a Material

In general, semiconductor materials possess all optical properties of dielectrics and metals to varying degrees. When a semiconductor material absorbs energy from the outside in some form (such as light, electricity, etc.), its electrons will be excited from the ground state to the excited state, that is, light absorption. Electrons in the excited state will spontaneously or stimulate and then transit from the excited state to the ground state, and radiate the absorbed energy in the form of light (radiation recombination), i.e., luminescence; of course, the absorbed energy can also be released in the form of non-radiation (such as heating). Figure 2.7 is a schematic diagram of possible absorption spectra in materials. Different absorption spectra correspond to different physical processes. The optical absorption region of the material is mainly divided into six regions.

- (1) Basic absorption region: the spectrum range is in the ultraviolet-visible-near infrared band. The transition of electrons from valence band to conduction band leads to strong absorption of light. The absorption coefficient is very high. Photoconductivity often occurs with the electron and hole that can be transferred.

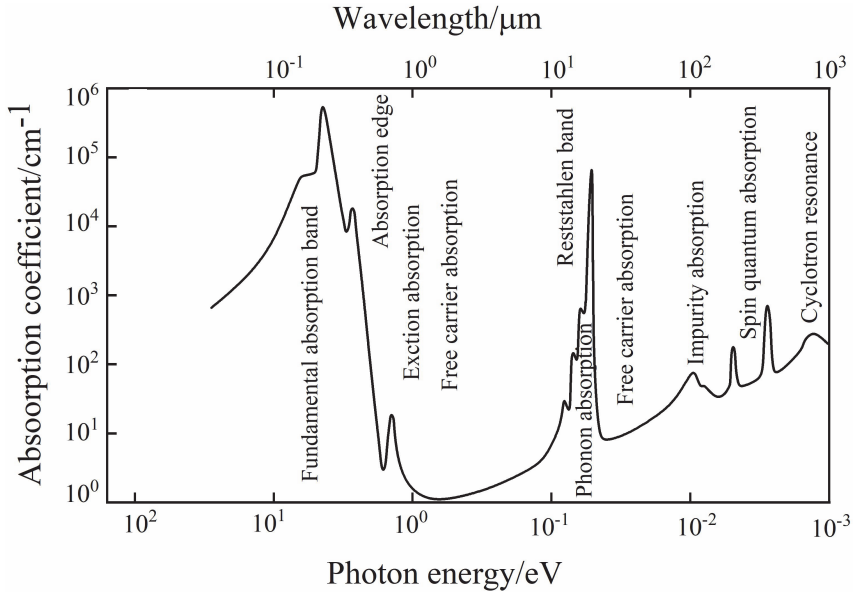


Fig. 2.7. Absorption spectra of a hypothetical semiconductor material.

- (2) Absorption edge: The smallest energy gap across which the electron transition crosses. For non-metallic materials, fine spectral lines are often generated by the absorption of excitons (stimulated electrons and holes bound together to form a new system, excitons).
- (3) Free carrier absorption: It is caused by the absorption of photon energy by electrons in conduction band or holes in valence band in the same band. It can be extended to the whole infrared band or even to microwave band. Obviously, the absorption coefficient is a function of electron (hole) concentration. The carrier concentration of metal materials is high, so the absorption line intensity in this band is very strong, even covering up the spectrum of other absorption areas.
- (4) Absorption caused by crystal vibration: It is caused by the interaction between incident photons and lattice vibration (phonon), and the wavelength is 20–50 micron.

- (5) Impurity absorption: Impurities introduce shallow levels into the intrinsic band structure, and the ionization energy is about 0.01 eV, which can be easily observed only at low temperature.
- (6) Spin-wave quantum or cyclotron resonance absorption: spin-wave quantum, cyclotron resonance and incident light have a lower energy, longer wavelength, up to millimeter magnitude.

2.2.1. The basic absorption

Figure 2.8 shows the absorption spectra of GaAs in the near infrared region. It can be seen that the absorption curve changes dramatically near 1.4 eV, forming the so-called absorption edge. It is found that for most semiconductors and insulators, similar absorption edges exist in visible or near infrared regions.

A careful study of the structure of the absorption edge reveals some regularities.

- (1) Strong absorption region: absorption coefficient $\alpha(\omega)$ is 10^4 – 10^6 cm⁻¹. The change of $\alpha(\omega)$ with photon energy $\hbar\omega$ is a power exponential rule, and its exponents may be 1/2, 3/2, 2, etc.

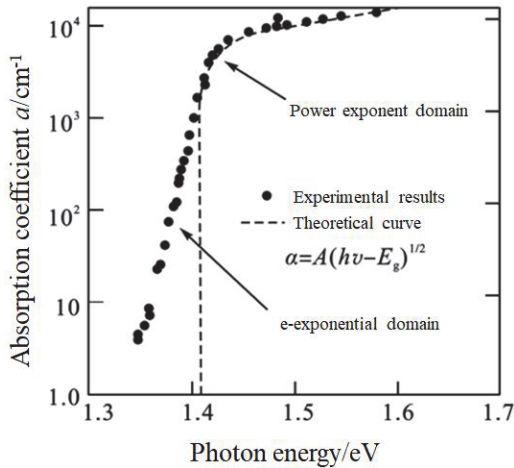


Fig. 2.8. Absorption spectra of semiconductor GaAs.

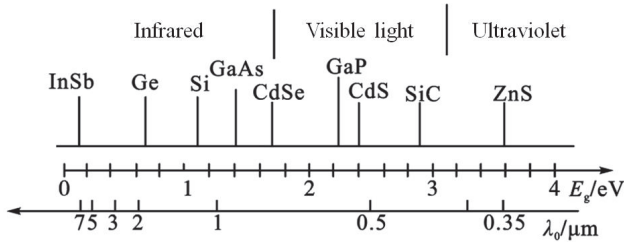


Fig. 2.9. Corresponding relations between E_g and λ_0 of several important semiconductor materials.

- (2) E -index absorption region: the absorption coefficient is about 10^2 cm^{-1} , and the change of $\alpha(\omega)$ with ω is e -index.
- (3) Weak absorption region: absorption coefficient $\alpha(\omega)$ is generally below 10^2 cm^{-1} .

Therefore, an absorbing edge contains abundant information. The optical process, rules and mechanism of power exponential absorption region are studied by spectroscopy.

The electron absorbs photons and then transits from valence band to conduction band. Obviously, only when the photon energy $h\nu$ is larger than the bandgap width E_g , that is $h\nu \geq E_g$, the basic absorption phenomenon can occur. So there is a long wave limit, that is, $\lambda = ch/E_g$. If the wavelength is greater than this value, it cannot cause basic absorption. In addition to energy requirements, the transition from valence band to conduction band of electrons also meets a certain momentum selection rule — The law of conservation of momentum.

Figure 2.9 shows the relationship between E_g and λ_0 of several important semiconductor materials.

2.2.2. Direct transition of permission and prohibition

Basic absorption can be divided into two categories: direct transition and indirect transition.

Assuming that the semiconductor is a pure semiconductor material, the valence band is full and the conduction band is empty at 0 K. Electrons absorb photon energy to produce a transition, keeping

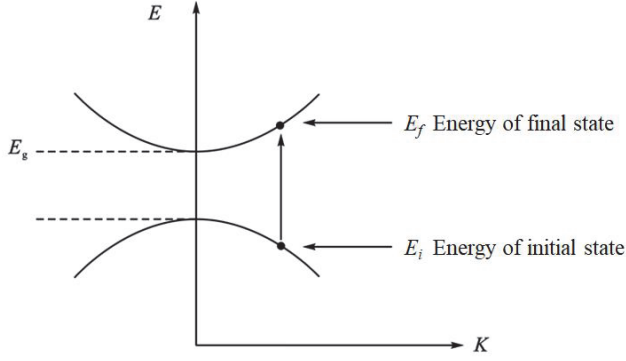


Fig. 2.10. Direct transition of photon energy absorbed by electrons from valence band to conduction band.

the wave number (quasi-momentum) unchanged, which is called direct absorption. This process does not require phonon assistance, as shown in Fig. 2.10. The common semiconductor GaAs belongs to this kind of direct bandgap semiconductor.

2.2.2.1. Permissible direct transitions

If all transitions are permissible, the transition probability P_{if} is a constant. In this case, the absorption coefficient α can be approximately expressed as

$$\alpha = AP_{if} \sum_{i,f} N_i(E_i)N_f(E_f) = AP_{if}N, \quad (2.30)$$

where $N_i(E_i)$, $N_f(E_f)$ and N represent the initial state density, the final state density and the joint state density of the electron transition, respectively.

For a simple parabolic band structure, the top of the valence band is assumed to be the coordinate origin (and the direct bandgap), with

$$\begin{cases} E_i = -\frac{\hbar^2 k^2}{2m_h}, \\ E_f = E_g + \frac{\hbar^2 k'^2}{2m_e} = E_g + \frac{\hbar^2 k^2}{2m_e}, \end{cases} \quad (2.31)$$

where m_e and m_h are the effective masses of conduction band electrons and valence band holes, respectively. According to the conservation of energy, there are

$$h\nu = Ef(k') - E_i(k) = E_g + \frac{\hbar^2 k^2}{2} \left(\frac{1}{m_e} + \frac{1}{m_h} \right) = E_g + \frac{\hbar^2 k^2}{2m_r}, \quad (2.32)$$

where $1/m_r = 1/m_e + 1/m_h$, and m_r is the reduced effective mass. Combined with (2.33) in the unit energy interval, the number of states (or density of states) in k space from k to $k + dk$ is

$$N(h\nu)d(h\nu) = \frac{8\pi k^2 dk}{(2\pi)^3} = \frac{(2m_r)^{3/2}}{2\pi^2 \hbar^3} (h\nu - E_g)^{1/2} d(h\nu). \quad (2.33)$$

Absorption coefficient $\alpha(h\nu)$ is, of course, proportional to $N(h\nu)$:

$$\alpha(h\nu) = AP_{\text{if}}N(h\nu) = B(h\nu - E_g)^{1/2}. \quad (2.34a)$$

It can be found in theory,

$$B \approx \left[e^2 \left(2 \frac{m_e m_h}{m_e + m_h} \right)^{3/2} \right] / nch^2 m_e, \quad (2.34b)$$

where B is independent of ν , and n is the refractive index of a pure semiconductor material.

The above discussion is based on the assumption that the direct transition of electrons is permissible for any k -value transition. It is assumed that the transition is permissible under the selection rule.

$$P_{\text{if}}(k = 0) \neq 0.$$

2.2.2.2. The direct transition of prohibition

In some materials, due to the difference of symmetry, in some cases, even in materials with direct bandgap, the direct transition of electrons is prohibited at $k = 0$ due to the restriction of the selection rule of quantum mechanics, while the transition of $k \neq 0$ is allowed.

In this case, there are

$$\begin{cases} P_{\text{if}}(k = 0) = 0, \\ P_{\text{if}}(k \neq 0) \neq 0. \end{cases}$$

Such a transition is called a forbidden transition of $k = 0$. The reason for this prohibition is similar with the selection rule of electron-level transitions in atomic physics. The electronic orbital composition of conduction band and valence band is different, there will be $k = 0$ forbidden transition and $k \neq 0$ transition. At this point, the transition probability P_{if} is no longer a constant. It is proportional to k^2 , that is, to $h\nu - E_g$. Currently, there is

$$\alpha(h\nu) = A' P_{\text{if}} N(h\nu) = B' (h\nu - E_g)^{3/2}, \quad (2.35a)$$

$$B' \approx \frac{2}{3} B \left(\frac{m_r}{m_h} \right) \frac{1}{h\nu}, \quad (2.35b)$$

where B' is related to ν .

Therefore, there are

$$\alpha(h\nu) = C (h\nu - E_g)^{3/2} / h\nu.$$

It can be seen that not all absorption can be described by the law of $1/2$ power, and the law of approximate $3/2$ power is often found in experiments.

2.2.3. Indirect transition

Square absorption edges, i.e. $[\alpha(h\nu)]^{1/2} \propto h\nu$, are often found in pure semiconductors such as germanium, silicon and heavily doped semiconductors. This absorption comes from indirect transitions. There are two situations that can lead to this absorption: one is the transition with phonon participation. Electrons not only absorb photons, but also exchange certain vibration energy with the lattice, i.e. emit or absorb a phonon. This absorption is different from that of direct transition light, and its absorption coefficient is closely related to temperature. The reason is that the lattice vibration is different at different temperatures. The number density of phonons has a distribution with temperature, and the optical absorption

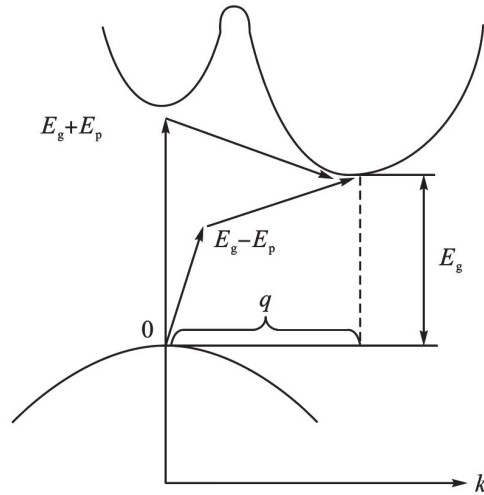


Fig. 2.11. Indirect transition of photon energy absorbed by electrons from valence band to conduction band.

coefficient ($1\text{--}10^3\text{ cm}^{-1}$) is much smaller than that of direct transition ($10^4\text{--}10^6\text{ cm}^{-1}$). The other is absorption involving impurity scattering.

For some semiconductor materials, the transition from valence band to conduction band is accomplished by phonons, as shown in Fig. 2.11, because the bottom k value of conduction band and the top k' value of valence band are different (indirect bandgap materials). E_p represents the energy of phonons. When the photon energy is in $E_g - E_p$, the electron absorbs a phonon to transit to the conduction band. If the photon energy is in $E_g + E_p$, the electron emits a phonon to transit to the conduction band.

When the law of conservation of energy is satisfied, momentum must also be conserved. Photon momentum is too small to change the momentum of electrons, so phonons must be involved.

$$\hbar k' - \hbar k \pm \hbar q = \text{photon momentum}, \quad (2.36)$$

where q is the phonon wave vector, $\mp$ means that the electron emits (−) or absorbs (+) a phonon during the transition. Formula (2.36)

can be simplified to

$$\hbar k' - \hbar k = \mp \hbar q. \quad (2.37)$$

If phonons have energy E_p , the law of conservation of energy is expressed as

$$E_f - E_i \pm E_p = h\nu. \quad (2.38)$$

For materials with simple parabolic band structure, the initial state density of energy in E_i is

$$N(E_i) = \frac{1}{2\pi^2 \hbar^3} (2m_h)^{3/2} |E_i|^{1/2} \quad (2.39)$$

where m_h is the effective mass of valence band holes. The final state density of energy in E_f is

$$N(E_f) = \frac{1}{2\pi^2 \hbar^3} (2m_e)^{3/2} (E_f - E_g)^{1/2}. \quad (2.40)$$

If the formula (2.38) is substituted for the formula (2.40), then there are

$$N(E_f) = \frac{1}{2\pi^2 \hbar^3} (2m_e)^{3/2} (h\nu - E_g \mp E_p + E_i)^{1/2}. \quad (2.41)$$

Referring to Section 2.2.1, it is obvious that the absorption coefficient is proportional to the product of the initial and final state densities, and the possible combinations of all states separated by $h\nu \pm E_p$ are integrated, while the convolution of the state density is converted into the integral of the initial E_i (valence band). Considering that the absorption coefficient is proportional to the probability of electron-phonon interaction $f(N_p)$, N_p denotes the number density of phonons whose energy is E_p , the absorption coefficient is

$$\alpha(h\nu) = Af(N_p) \int_0^{-E_i^m} |E_i|^{1/2} (h\nu - E_g \mp E_p + E_i)^{1/2} dE_i, \quad (2.42)$$

where the integral upper limit $E_i^m = h\nu - E_g \mp E_p$, $-E_i^m$ denotes the lowest initial energy value of indirect transition for a photon

whose frequency is ν . It is noted that the phonon distribution obeys the Bose distribution, and the probability $f(N_p)$ of the interaction between electron and phonon is proportional to the phonon number density N_p , i.e.

$$f(N_p) \propto N_p = \frac{1}{\exp(E_p/k_B T) - 1}. \quad (2.43)$$

The integral of formula (2.43) is obtained by considering the following two absorption modes, it is obtained that

- (1) For $h\nu > E_g - E_p$, the absorption process of phonons is accompanied (because only the energy of phonons can be absorbed, the absorption coefficient of $h\nu + E_p > E_g$ can be guaranteed). The absorption coefficient is

$$\alpha_a(h\nu) = \frac{A_a(h\nu - E_g + E_p)^2}{\exp(E_p/k_B T) - 1}. \quad (2.44a)$$

- (2) For $h\nu > E_g + E_p$, it can be accompanied by both phonon emission and phonon absorption (when the energy of photons is large enough to ensure $h\nu - E_p > E_g$). The absorption coefficient of the accompanying phonon emission is

$$\alpha_e(h\nu) = \frac{A_e(h\nu - E_g - E_p)^2}{1 - \exp(-E_p/k_B T)}. \quad (2.44b)$$

So, if the photon energy $h\nu > E_g + E_p$, there are two kinds of absorption, the total absorption coefficient is

$$\alpha(h\nu) = \alpha_a(h\nu) + \alpha_e(h\nu) \quad (2.45)$$

According to the above formula, the actual measurement data can be analyzed. Make $\sqrt{\alpha}-h\nu$ chart. If the above absorption mechanism is satisfied, the linear relationship between them can be seen, as shown in Fig. 2.12. The following important information can be obtained by analyzing formula (2.44a) and (2.44b).

- (1) When $E_g + E_p > h\nu > E_g - E_p$, α_a was the dominant factor. When $h\nu = E_g - E_p$, $\alpha_a = 0$. The slope of absorption

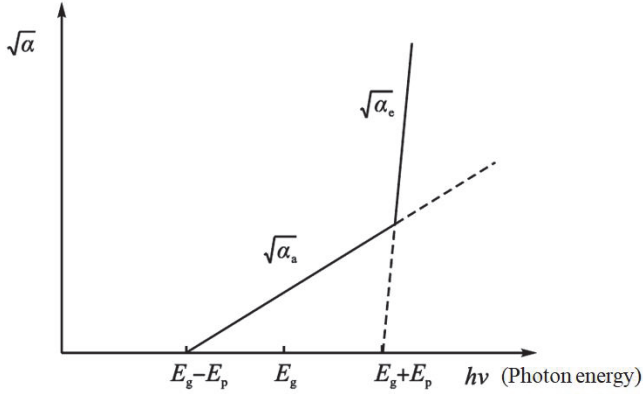


Fig. 2.12. Relation of indirect absorption coefficient of electrons to temperature and photon energy.

edge is $\left\{ \frac{A_a}{\exp(E_p/R_B T) - 1} \right\}^{1/2}$ obtained from the line segment of $[\alpha(h\nu)]^{1/2} - h\nu$.

In this case, the phonon absorption process is accompanied by a line segment with a lower absorption coefficient. The line segment is extended to intersect with $h\nu$ to obtain $h\nu = E_g - E_p$. As the temperature decreases, the slope of the line decreases.

- (2) When $h\nu > E_g + E_p$, $[\alpha(h\nu)]^{1/2} - h\nu$ corresponds to the line segment with higher absorption coefficient, which includes both phonon emission and phonon absorption process. When $h\nu = E_g + E_p$, $\alpha_e = 0$. However, the comparison formulas (2.44a) and (2.44b) show that the probability of launching a phonon at low temperature is much greater than that of absorbing a phonon. Therefore, the slope of this line is basically determined by the probability of launching a phonon, i.e. $\left\{ \frac{A_e}{1 - \exp(-E_p/k_B T)} \right\}$.

As the temperature decreases, the slope of the straight line decreases, extending the line to intersect with the energy axis to obtain $h\nu = E_g + E_p$.

- (3) From the above two points, two important parameters E_p and E_g can be obtained by measuring the $h\nu$ relationship of $[\alpha(h\nu)]^{1/2}$.
- (4) At different temperatures, E_g may be different. With the decrease of temperature, E_g generally increases. In this case,

it will be found that with the decrease of measurement temperature, the absorption edge “blue shift” will occur. It may also be that with the decrease of temperature, the absorption edge “red shift” will occur, whether it is “red shift” or “blue shift” depends on the specific situation.

(Note: When $h\nu > E_g + E_p$, α_e is dominant. If $h\nu \leq E_g - E_p$, $\alpha_a = 0$; If $h\nu \leq E_g + E_p$, $\alpha_e = 0$.)

2.2.4. Exciton

The term exciton comes from excitation, which means the quantum of the elementary or excited states in solids. It can also be simply understood as bound electron–hole pairs. Electrons excited from the valence band to the conduction band are usually free. Holes moving freely in the valence band and electrons moving freely in the conduction band may be re-bound to form bound electron–hole pairs, that is, excitons. So why are free electrons and holes bound together to become excitons? This is mainly the result of Coulomb interaction. Because of the binding, the energy of exciton is lower than that of free electron. The absorption and emission spectra of excitons are different from those of band-to-band transitions and have characteristic structures. Exciton binding energy (or exciton binding energy) in semiconductors is generally very low, about several or more millivolts. Therefore, at room temperature, the exciton absorption is generally not observed.

It is found that there are a series of discrete absorption peaks in the low energy of the absorption edge of the interband transition, and the distribution of the peaks is regular. Figure 2.13 shows the absorption spectrum near the band edge of high purity GaAs.

Compared with the GaAs band transition absorption edge represented by the dotted line in the lower right corner of the figure, the main feature is that there are a series of absorption peaks in the low energy direction of the absorption edge, and the absorption intensity is much higher than that of the band transition absorption. The absorption spectrum labeled $n = 1, 2, 3, \dots, \infty$ in the figure is attributed to free exciton absorption, while the absorption

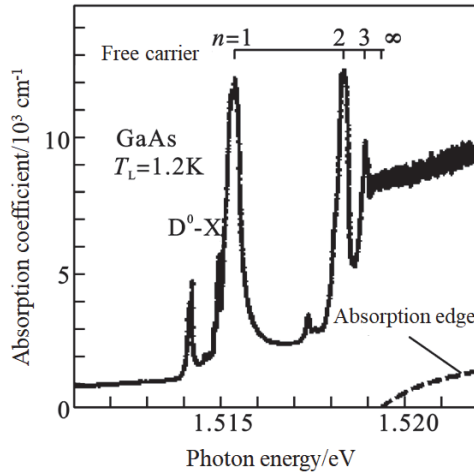


Fig. 2.13. Near-band-edge absorption spectra of high purity GaAs at low temperature.

peak labeled $D^0 - X$ is bound exciton absorption on neutral donor impurities. The experiment was carried out at low temperature.

Figure 2.14 shows the band-edge absorption spectra of Cu_2O at 1.8 K low temperature. Comparing with Fig. 2.13, they have a series of absorption peaks in the low-energy direction of the absorption edge. The difference is that the label of the absorption peak in Cu_2O does not start from $n = 1$, but from $n = 1, 2, 3, 4, \dots$, which is determined by the selection rule. In addition, the absorption peaks are asymmetric due to the influence of band-edge background absorption.

Unlike the absorption spectra of interband transitions, the above discrete absorption peaks are not accompanied by photoconductivity, which indicates that these discrete absorption peaks are not caused by the transition from valence band electrons to conduction bands, but are probably caused by the exciton transition when valence band electrons are excited to some discrete energy levels below the bottom of conduction bands. Why is the motion of excitons not accompanied by changes in photoconductivity? The reason is that the exciton itself is electrically neutral and the motion of the hole-electron pair is in one direction.

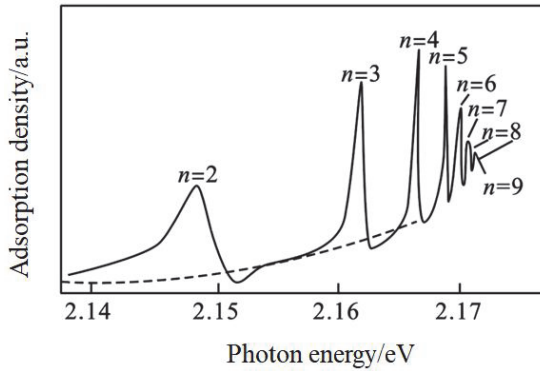


Fig. 2.14. Cu_2O “yellow exciton” absorption spectrum.

Since the energy of exciton is lower than that of free electron, it can be understood that the energy absorbed by exciton is lower than that of bandgap E_g , and that the exciton absorption peaks are distributed in the low energy direction along the absorption edge of the band-to-band transition.

2.3. The Photoelectric Effect

Light irradiation on an object causes the object to emit electrons, or changes its conductivity, or produces photoelectromotive force. This change in electrical properties of the object caused by light is collectively called photoelectric effect. Photoelectric effects can be classified into two categories.

- (1) The phenomenon that matter emits electrons outward after being illuminated is called external photoelectric effect. This effect occurs mostly in metals and metal oxides.
- (2) The phenomenon that the photoelectrons generated by the illumination of a substance only move inside the substance without escaping outside the substance is called the internal photoelectric effect. This effect includes photoconductive effect and photovoltaic effect.

2.3.1. External photoelectric effect

If the excited electrons escape from the surface of photosensitive material and form photocurrent under the action of external electric field, this is photoemission effect or external photoelectric effect. Some special optoelectronic devices, such as phototransistors and photomultipliers, are based on the external photoelectric effect. The main laws and properties of the photoelectron emission effect are as follows.

2.3.1.1. Stoletov's law

Stoletov's law is also known as the first law of photoelectric emission. When the frequency component of the incident light is constant (monochrome light of the same wavelength or light of the same frequency component), the saturated photoemission current I_k of the photocathode is proportional to the flux Φ_k absorbed by the cathode, that is to say, the saturated photoemission current I_k of the photocathode is proportional to the flux K absorbed by the cathode, i.e.

$$I_k = S_k \Phi_k, \quad (2.46)$$

where S_k is the coefficient to characterize the photoelectric emission sensitivity. This relationship seems very simple, but it is very important because it is an important basis for photometric measurement and photoelectric conversion with photoelectric detectors.

2.3.1.2. Einstein's law

Einstein's law is also called the second law of photoelectric emission. The maximum kinetic energy of emitted photoelectrons increases linearly with the increase of incident light frequency, but has nothing to do with the intensity of incident light. That is to say, the energy relationship of emitted photoelectrons conforms to Einstein's formula:

$$h\nu = \left(\frac{1}{2} m_e v^2 \right)_{\max} + \Phi_0, \quad (2.47)$$

where h is Planck constant, ν is the frequency of incident light, m_e is the mass of photoelectrons, v is the velocity of photoelectrons emitted, and Φ_0 is the work of the photocathode. Electron escape work is a physical quantity describing the binding strength of the material surface to the electrons, which is equal to the minimum energy required for the electrons to escape from the surface quantitatively. It can also be called the energy threshold of photoemission.

According to the quantum theory of light put forward by Einstein in 1905, it is easy to explain the sum of formula (2.46) and formula (2.47). In fact, under the action of light, the electrons in the photosensitive object absorb the energy of photons, and have enough kinetic energy to overcome the effect of the boundary barrier of the photosensitive object and escape from the surface. According to Einstein's hypothesis, the escape of each electron is the result of absorption of a photon quantum, and all the energy of a photon has the energy of radiation into photoelectrons. So the stronger the light is, the more quantum numbers acting on the cathode surface, the more electrons will escape from the cathode surface. At the same time, the higher the frequency of incident light, that is to say, the greater the energy of each photon, the greater the kinetic energy of the electrons at the highest level in the cathode material when they get this energy and overcome the barrier effect and escape from the interface.

2.3.1.3. *Red limit of photoelectric emission*

In the frequency range of incident light, there exists a critical wavelength of photocathode. When the wavelength of light wave equals the critical wavelength, the photoelectron just escapes from the cathode. This wavelength is usually called the "red limit" of photoemission, or the threshold wavelength of photoemission (photocathode wavelength threshold λ_0). Obviously, the initial velocity (kinetic energy) of the photoelectron at the red limit should be zero. Therefore, $h\nu_0 = \Phi_0$, critical frequency $\nu_0 = \Phi_0/h$, so the critical wavelength is

$$\lambda_0 = \frac{C}{\nu_0} = \frac{ch}{\Phi_0} = \frac{1.24}{\Phi_0}, \quad (2.48)$$

where, the unit of λ_0 is micron. The shortest wavelength of visible light (380 nm) generates photoemission in cathode materials with surface work (also called work function) less than 3.2 eV, while the longest wavelength of visible light (780 nm) generates photoemission only in cathode materials with work function less than 1.6 eV.

2.3.1.4. *Instantaneous photoelectric emission*

The instantaneity of photoelectric emission is an important characteristic of photoelectric emission. Experiments show that the delay time of photoelectric emission does not exceed the order of 3×10^{-13} s. Therefore, in fact, it can be considered that the photoelectric emission is inertial, which determines that the external photoelectric effect device has a high-frequency response. The instantaneity of photoemission is due to the fact that it does not involve the physical process in which electrons migrate to metastable levels in atoms.

The above conclusion is strictly correct only when the temperature is 0 K. As the temperature increases, the energy of electrons in the cathode material will also increase, and it is possible to escape from the surface below the original red limit. In practice, however, the number of such energetic electrons is very small due to the increase of temperature. In practical measurement of photoelectric emission at high temperature, Einstein's law and red limit are still correct for most metals due to the limitation of instrument sensitivity.

At the earliest time, it was believed that the photoemission effect occurred only on the surface of the cathode material, i.e. the monoatomic layer on the cathode surface or within a distance of tens of nanometers from the surface. However, after the discovery of cathode materials with high sensitivity, it is considered that photoemission occurs not only on the surface of the object, but also deep into the cathode material, which is usually called the volume effect of photoemission, while the former is called the surface effect of photoemission.

The process of optical emission includes three basic stages.

- (1) The electron absorbs photons and generates excitation, i.e. obtain energy.

- (2) Electrons (stimulated electrons) which obtain photon energy move from the emitter to the vacuum interface (electron transmission).
- (3) The stimulated electrons escape to vacuum through the surface barrier.

The state of electron excitation depends on the optical properties of the material. All light-emitting materials should have optical absorption capacity. Optical absorption coefficient should be as large as possible, so that stimulated electrons occur near the surface, that is to say, the excitation depth is shallow. In solids, when stimulated electrons move to the surface, some energy will be lost due to various interactions. The transport capacity of stimulated electrons can be expressed by the effective escape depth. It refers to the average distance through which the stimulated electrons reach the vacuum interface. The greater the ratio of escape depth to stimulated depth, the higher the efficiency of the emitter. In order to complete the photoemission, that is, the electron eventually escapes into vacuum, the energy of the electron reaching the surface should be greater than the work of the material. The smaller the work of escape, the greater the probability of electron emission from an object to a vacuum. After the electrons are emitted from the object, they are compensated by the electronic current of the external power supply, so as to meet the requirements of the conductivity of photocathode materials.

2.3.1.5. *Photoelectric emission of metals*

Metal reflects most of the incident visible light (the reflection coefficient is over 90%) and its absorption efficiency is very low. Photoelectrons collide with a large number of free electrons in metals and lose a lot of energy in motion. Only the photoelectrons close to the surface can reach the surface and overcome the potential barrier escape, that is, the depth of photoelectrons escape from metal is very shallow, only a few nanometers, and most of the work of metal escape is more than 3 eV. For visible light with energy less than 3 eV ($\lambda > 410 \text{ nm}$), it is difficult to generate photoemission. Only cesium

(2 eV work of escape) is the most sensitive to visible light, so it can be used as photocathode. However, the quantum efficiency of pure metal cesium is very low, less than 0.1% due to the excessive energy loss in the first two stages of photoemission. Metals have a large number of free electrons and no bandgap. The below Fermi level is basically filled by electrons. The above Fermi level is basically empty. The surface energy band is little affected by the internal and external electric fields. E_F depends on the material. Therefore, the electron escape work of metals is defined as the difference between the vacuum level and E_F at $T = 0$ K. It is a parameter of materials and can be used as the energy threshold of photoemission.

2.3.1.6. Photoelectric emission of semiconductors

Semiconductor photoelectric escape parameters have electron affinity. The electron affinity reflects the minimum energy required for the electron at the bottom of the conduction band to escape into vacuum, which is equal to the difference between the vacuum energy level (the static electron energy in vacuum) and the conduction band energy level E_c . It has surface electron affinity potential energy χ_a and *in vivo* electron affinity potential energy χ_{ac} . χ_a is a parameter of material, independent of doping and bending of surface energy band, while χ_{ac} is not a parameter of material and can change with bending of surface energy band.

Semiconductor has fewer free electrons and a forbidden band. Fermi level E_F is generally in the forbidden band and varies with doping and internal and external electric fields, so the difference between vacuum level and Fermi level is not a material parameter. Semiconductor electron escape work is defined as the difference between the vacuum level and the energy level of the electron emission center at $T = 0$ K. The energy levels of the electron emission center are either at the top of valence band, impurity level or at the bottom of conduction band. The situation is complex. Therefore, the concept of electron escape work is seldom used in semiconductors. Since the work of electron escape contains affinity (the difference between vacuum level and conduction band bottom level), it is more practical to use the concept of electron affinity than the concept

of work of escape in order to express the difficulty of photoemission. Therefore, the concept of electron affinity is generally used instead of the concept of escape work for semiconductors. In order to represent the energy threshold of photoemission, many data are calculated according to the difference between vacuum energy level and valence bandtop (electron affinity plus bandgap width).

Surface energy band is bent. When semiconductor is unbounded, the band structure is flat; when semiconductor is bounded, the periodicity of lattice arrangement (potential field) is destroyed at the surface, and the surface is easy to oxidize and contaminated by impurities, so the additional energy level (surface energy level) is introduced into the forbidden band. Because of the existence of surface energy levels, the energy band bends at the surface. Surface band bending has an effect on photoelectron emission *in vivo*. Because surface electron affinity χ_a is a parameter of material, it does not change with the bending of surface energy band, while *in vivo* electron affinity χ_{ac} increases and decreases with the bending of surface energy band.

For N-type semiconductors, when electrons at donor level transit to surface level, a negative space charge region will be generated on the surface of semiconductors, while a positive volume charge is distributed in the body a little farther away from the surface, so the surface energy band will bend upward. The degree of upward bending can be expressed by surface barrier eU_s . e is the electron charge, and U_s is the surface potential, which is equal to the potential difference between the body and the surface. For N-type semiconductor, the electron affinity E_{ea} *in vivo* increases a barrier height eU_s when the surface energy band bends upward, which makes the photoelectron emission *in vivo* more difficult than when the surface energy band does not bend.

For P-type semiconductors, the opposite is true. Some of the electrons with higher energy than the acceptor level in the surface level have to transit to the acceptor level, so a positive space charge region is generated on the surface of the semiconductor, and an equal amount of negative charge is distributed in the body a little farther away from the surface, so the surface energy band is bent downward.

Especially when the surface of P-type semiconductor adsorbs positive atoms (such as cesium atoms) or N-type materials, the positive charge of the coupling layer on the surface is outside, and the band bending is more severe. The degree of band bending is also expressed by surface barrier eU_s . The surface energy band bends downward, which reduces the height of the barrier eU_s *in vivo* when the energy band does not bend. Thus, the bending of the surface energy band is very beneficial to the photoelectron emission *in vivo*.

Therefore, almost all practical photocathodes are made of P-type semiconductor material as the substrate, and then coated with positive metal or N-type material on its surface. In this way, the downward curved surface energy band can be obtained, and the escape work can be reduced. If the bending of the band is small enough (the width of the band is z) to be much smaller than the reciprocal of the material absorption coefficient ($z < 1/\alpha$), the main part of the photoelectron emission can come from the body. At this time, the quantum efficiency is much larger than the simple band bending. In addition, the Fermi level of strongly implanted P-type semiconductors is very close to the valence band, which makes the emission of hot electrons (dark current) smaller.

2.3.2. Photoconductive effect

Here we will discuss the phenomenon of photoconductivity. The former is called photoconductive effect, and the latter is called photovoltaic effect. The phenomenon that the conductivity of semiconductor materials increases under the action of light is the photoconductivity effect. The main sources of photoconductivity are interband carrier transition and impurity excitation, so there are intrinsic photoconductivity and impurity photoconductivity.

According to the introduction of Chapter 1, photoconductivity can be expressed as

$$\sigma = e(n_0 + \Delta n)\mu_e + e(p_0 + \Delta p)\mu_h \quad (2.49)$$

where n_0 and p_0 are thermally balanced carrier concentrations, respectively. The conductivity at thermal equilibrium is

$$\sigma_0 = n_0 e \mu_e + p_0 e \mu_h. \quad (2.50)$$

Let $\Delta\sigma = \sigma - \sigma_0$ be called additional photoconductivity. When $\Delta\sigma - \sigma_0$ is defined as the photoconductive sensitivity, there is

$$\frac{\Delta\sigma}{\sigma_0} \equiv \frac{\sigma - \sigma_0}{\sigma_0} = \frac{\Delta n \mu_e + \Delta p \mu_h}{n_0 \mu_e + p_0 \mu_h} = \frac{(1 + b) \Delta n}{b n_0 + p_0}, \quad (2.51)$$

where let $b = \mu_e / \mu_h$. In the eigenstate, there is $\Delta n = \Delta p$. Because the intrinsic carrier concentration n_0 and p_0 increase exponentially with temperature. Therefore, the lower the temperature, the smaller the N_0 and p_0 , and the greater the $\Delta\sigma / \sigma_0$, that is to say, the lower the temperature, the higher the sensitivity. In experiments, it is often observed that the photoconductivity is closely related to the impurities in the material as follows.

2.3.2.1. *The relationship between photoconductivity and light intensity under constant illumination*

There are mainly linear and nonlinear photoconductivity.

(i) *Linear photoconductivity*: When light intensity is low, the relationship between photoconductivity and light intensity is linear. Electron-hole pair yield: I is defined as the number of photons passing through a unit area in a unit time, α as the absorption coefficient, and β as the quantum yield of electron-hole pairs produced by each photon. Therefore, the electron-hole pair generation rate is $I\alpha\beta$, and the electron-hole pair recombines continuously, the recombination rate is $\Delta n / \tau$, and τ is the electron lifetime. Under steady-state conditions, the yield and the recombination rate are balanced.

$$I\alpha\beta = \Delta n / \tau, \quad (2.52)$$

$$\Delta n = \tau I\alpha\beta. \quad (2.53)$$

Typical representative systems are silicon and cuprous oxide.

(ii) *Parabolic photoconductivity*: The photoconductivity is proportional to the square root of light intensity.

$$(\Delta n)^2 = \tau I\alpha\beta / \gamma, \quad (2.54)$$

where γ is a proportional constant. The representative systems that accord with this relationship are thallium sulfide (Tl_2S_3).

2.3.2.2. Relaxation time of photoconductivity

The relaxation time of photoconductivity reflects the speed of semiconductor reaction to light. When the illumination is not constant, the discussion begins with illumination or with the cancellation of illumination.

(i) For linear photoconductivity, when $t = 0$, light begins to shine, then

$$d(\Delta n)/\Delta t = I\alpha\beta - \Delta n/\tau. \quad (2.55)$$

It represents the number of net remaining carriers per unit time. If $t = 0$ and $n = 0$, then

$$\Delta n = \tau(I\alpha\beta)[1 - \exp(-t/\tau)]. \quad (2.56)$$

Formula (2.56) represents the ascending curve.

When the light is canceled,

$$D(\Delta n)/\Delta t = -\Delta n/\tau, \quad (2.57)$$

$$\Delta n = \tau(I\alpha\beta) \exp(-t/\tau). \quad (2.58)$$

Formula (2.58) represents the descent curve. The ascending and descending curves are shown in Fig. 2.15. The relaxation time is defined as $t \equiv \tau \ln 2$. Its physical meaning is that during the period of time, Δn rises or falls to $1/2$ of the stationary value $\tau(I\alpha\beta)$.

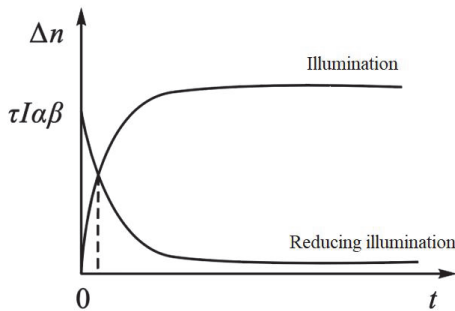


Fig. 2.15. Linear photoconductivity rising and falling curves.

(ii) For parabolic linear photoconductivity, when illumination begins,

$$d(\Delta n)/\Delta t = I\alpha\beta - \gamma(\Delta n)^2. \quad (2.59)$$

The solution yields

$$\frac{1}{(I\alpha\beta)^{1/2}} \tanh^{-1} \left(\frac{\Delta n}{\sqrt{I\alpha\beta/\gamma}} \right) = t. \quad (2.60)$$

Formula (2.60) represents the ascending curve.

When the light stops,

$$d(\Delta n)/\Delta t = -\gamma(\Delta n)^2. \quad (2.61)$$

The solution yields

$$\Delta n = \left(\frac{I\alpha\beta}{\gamma} \right)^{1/2} \frac{1}{1 + (I\alpha\beta\gamma)^{1/2}t}. \quad (2.62)$$

Formula (2.62) represents the descent curve.

Relaxation time $t = (I\alpha\beta\gamma)^{-1/2}$. Within this time value, the photoconductivity increases to 0.76 in the steady state, and decreases to 1/2 after the light stops.

2.3.2.3. Photoresistors

Photoresistor is a kind of non-polar photoelectric element made of photoconductive material, also known as photoconductive tube. It works based on semiconductor photoconductive effect. Because the photoresistor has no polarity, DC bias or AC voltage can be added when working. When there is no light, the resistance (dark resistance) of the photoresistor is very large and the current in the circuit is very small. When it is illuminated by light in a certain wavelength range, the resistance (bright resistance) decreases sharply and the current in the circuit increases rapidly. The current can be measured by an ammeter, as shown in Fig. 2.16. According to the change of current value, the intensity of illumination can be calculated.

- (1) Dark resistance and bright resistance: The resistance value of photosensitive resistor when it is not illuminated is called dark

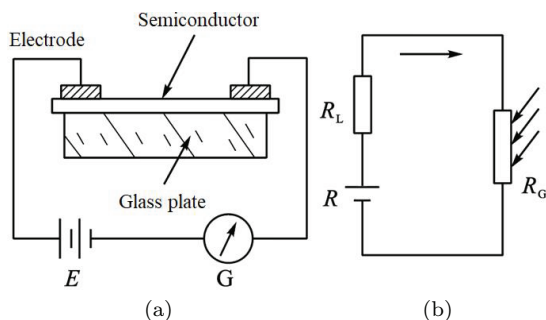


Fig. 2.16. Photoresistor structure diagram.

resistance, and the resistance value when it is illuminated by strong light is called bright resistance. The greater the dark resistance and the smaller the bright resistance, the higher the sensitivity.

- (2) Volt-ampere characteristics of photoresistors: As shown in Fig. 2.17, the higher the voltage applied under a certain illumination, the greater the current; under a certain voltage, the stronger the illumination of the incident light, the greater the current, but not necessarily a linear relationship.
- (3) Spectral characteristics of photoresistors: As shown in Fig. 2.18, the sensitivity of photoresistors is different for different wavelengths of light. This characteristic must be fully considered when optoelectronic devices are selected.
- (4) Response time and (modulation) frequency characteristics of photoresistors: The response time of photoelectric devices reflects its dynamic characteristics. The shorter the response time, the better the dynamic characteristics. For optoelectronic devices using modulated light, the upper limit of modulation frequency is limited by response time. The response time of the photoresistor is generally 10^{-3} – 10^{-1} s, and that of the photodiode is about 2×10^{-5} s, as shown in Fig. 2.19.
- (5) Temperature characteristic of photoresistor: With the increase of temperature, dark resistance and sensitivity of photoresistor will decrease, and the change of temperature will also affect the spectral characteristic curve. With the increase of temperature,

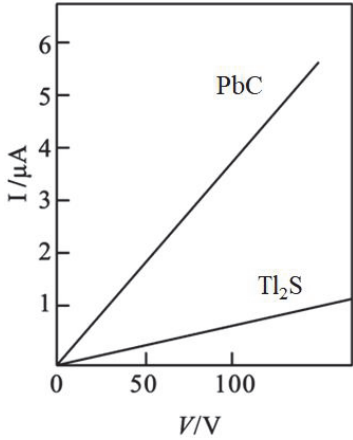


Fig. 2.17. Volt-ampere characteristic of photoresistor.

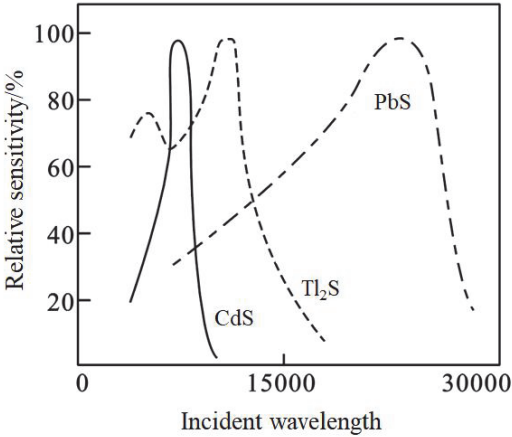


Fig. 2.18. Spectral characteristic of photoresistor.

the peak value of spectral response of photoelectric devices such as lead sulfide photoresistor will move to short wave direction, so infrared detectors often take refrigeration measures, as shown in Fig. 2.20.

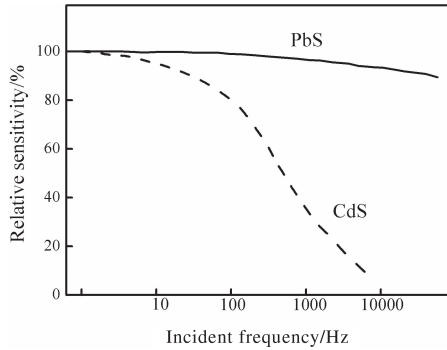


Fig. 2.19. Frequency characteristic of photoresistors.

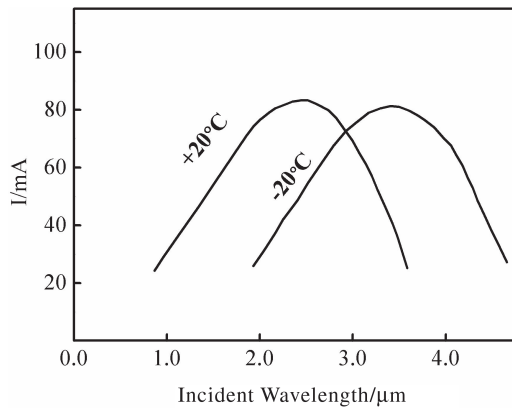


Fig. 2.20. Spectral temperature characteristics of lead sulfide photoresistors.

2.3.3. Photovoltage

For semiconductor PN junctions, the phenomenon of voltage difference after illumination, which converts solar energy into electricity, is the working principle of solar cells. The photovoltaic principle of PN junction is described below.

Semiconductor photodiodes have a lot in common with ordinary diodes. They all have a PN junction, which belongs to the nonlinear

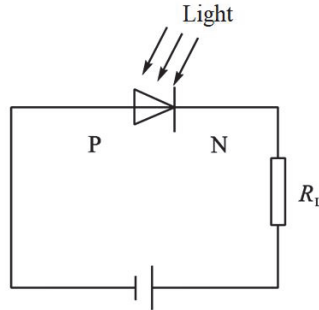


Fig. 2.21. Photodiode wiring method.

components with unidirectional conductivity. Photodiodes are usually used under negative bias. Their illumination characteristics are linear, so they are suitable for detection and other applications, as shown in Fig. 2.21. When photodiodes are not illuminated by light, the reverse resistance is very large and the reverse current (dark current) is very small (in the cut-off state).

When the energy of incident light is larger than the gap of semiconductor, i.e. $h\nu > E_g$, it irradiates on PN junction and generates electron-hole pairs in the junction region due to intrinsic absorption. Under light excitation, the concentration of most carriers generally changes little, while the concentration of a few carriers varies greatly. Therefore, we mainly discuss the motion of photogenerated minority carriers.

Because of the built-in electric field in PN junction, the photogenerated minority carriers on both sides of the junction move in opposite directions under the action of the built-in electric field. So the electrons in P region pass through PN junction into N region, and the holes in N region enter into P region, thus there is charge accumulation in P and N region, as shown in Fig. 2.22(a); the potential at P end increases and the potential at N end decreases, thus forming a photovoltaic potential at both ends of PN-junction, as shown in Fig. 2.22(b), which is the photovoltaic effect of PN-junction. If the PN-junction is loaded at both ends, the current will flow over the load under appropriate illumination. The relationship between

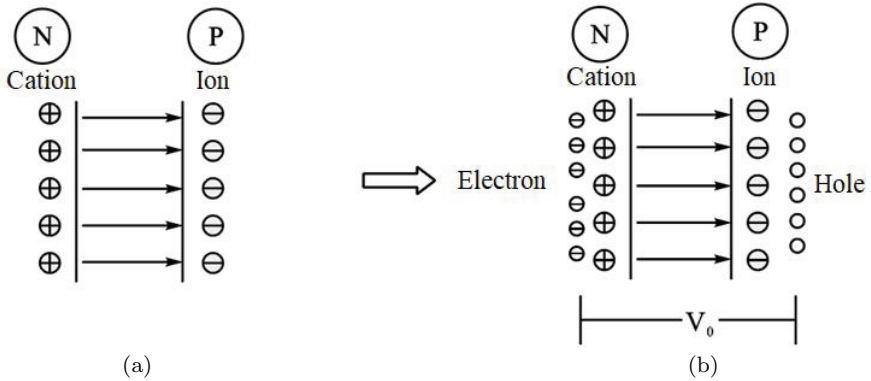


Fig. 2.22. Photovoltaic principle of semiconductor diode PN junction: (a) Start lighting (b) Equilibrium state.

the photoelectric current I and illumination of the photodiode is linear. However, in the absence of light, although PN junctions have built-in electric fields, there are no migrtable carriers and no current can be formed.

2.3.3.1. Open-circuit voltage and short-circuit current

Two important concepts that must be clearly understood about photodiodes are the open-circuit voltage and short-circuit current of photovoltaic cells. Next, we discuss the behavior of open-circuit voltage and short-circuit current.

- (i) Open-circuit voltage: There is light, but the external circuit is disconnected, and the potential difference V_0 formed at both ends of PN-junction, that is, the open-circuit voltage of photovoltaic cell (V_0); the expression of the open-circuit voltage will be given later.
- (ii) Short-circuit current: When illuminated, the external circuit is short-circuit, and the photovoltaic voltage cannot be formed at both ends of PN-junction, but the current flowing through the external circuit is the largest, which is the short-circuit current I_0 of the photovoltaic cell. Open-circuit voltage and short-circuit

current are two important parameters of photovoltaic cells. The main sources of photocurrent are discussed below.

2.3.3.2. Composition of photocurrent I_l

As the carriers generated by light move in opposite directions, the photogenerated current I_l (i.e. the hole in N region migrates to P region and the electron in P region migrates to N region) is formed in PN junction. Therefore, the photocurrent I_l consists of two components, I_{ge} and I_{gh} , where I_{ge} represents the electron part generated by P region and I_{gh} represents the hole part generated by N region which can diffuse to the barrier region, i.e. the change of carrier (electric quantity) per unit time.

$$I_l = I_{ge} + I_{gh} = eA(L_n + L_p)G, \quad (2.63)$$

where, G is the carrier generation rate per unit volume, A is the junction area of PN junction, L_n and L_p are the diffusion lengths of electron and hole carriers respectively, and e is the charge of electron. It is closely related to the characteristic parameters and illumination of PN junction. It can be seen that the larger the junction area A is, the greater I_l will be.

2.3.3.3. Photovoltaic equivalent circuit with load

Because light generates photoelectromotive force at both ends of PN junction, which is equivalent to adding forward voltage V at both ends of PN junction (attention is equal to open voltage V_0 when opening circuit), positive current I_f flows through PN junction after forming circuit with load. It must be pointed out that I_f exists only when there is a load. Otherwise, the circuit is equivalent to open circuit. Although photoelectromotive force is generated by illumination, there is no carrier flow. When there is a load, because the output voltage of the photovoltage is V , the current I flowing over the load is photoelectric current I_l minus the forward current I_f in the PN junction, i.e.

$$I = I_l - I_f. \quad (2.64)$$

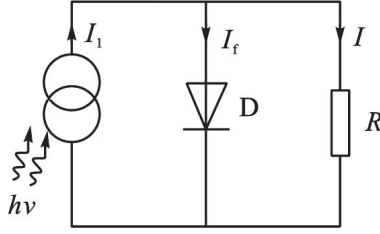


Fig. 2.23. Equivalent circuit diagram of semiconductor photodiode D with load R.

Figure 2.23 is a photovoltaic equivalent circuit with load.

From the above discussion, we can see that there are three currents in the photovoltaic cell: photogenerated current I_1 , positive current I_f of PN junction under photovoltaic voltage V , current I flowing through the external circuit. Both I_1 and I_f flow through PN junctions, but in the opposite direction.

From semiconductor physics (Liu Enke *et al.*, 2003), we can see that the forward current I_f of PN junction with V forward voltage bias is in the following form:

$$I_f = I_s[\exp(eV/k_B T) - 1], \quad (2.65)$$

where I_s is the reverse saturated current of PN junction (under certain temperature conditions, the minority carrier concentration determined by intrinsic excitation is certain, so the drift current formed by minority carrier is constant, basically independent of the magnitude of the applied reverse voltage, which is also called reverse saturated current I_s), k_B is the Boltzmann constant, and T is the temperature of PN junction. The size of V can be deduced from formula (2.64) and formula (2.65).

$$V = \frac{k_B T}{e} \ln \left[\frac{I_1 - I}{I_s} + 1 \right] \quad (2.66)$$

In this way, the magnitude of the open circuit voltage V_0 of PN junction and the short circuit current I_0 of PN junction in short circuit can be obtained. In fact, the relationship between the open-circuit voltage and short-circuit current of PN junction and the

characteristic parameters and illumination of PN junction itself can be obtained. When opening the circuit, let $I = 0$, V_0 can be obtained by formula (2.66).

$$V_0 = \frac{k_B T}{e} \ln \left(\frac{I_1}{I_s} + 1 \right). \quad (2.67)$$

In short circuit, let $V = 0$, $I_f = 0$ from formula (2.53), then I_0 from formula (2.52).

$$I_0 = I_1. \quad (2.68)$$

It is obvious that it is completely photocurrent. I_1 is related to the characteristic parameters of PN junction itself and to light. Generally, I_1 is proportional to light intensity (as can be seen from formula (2.63)). From formula (2.66), we can see that the open circuit voltage V_0 is logarithmic to light intensity. Fig. 2.24(a) shows the I - V relationship of PN junction photodiodes formed by a typical semiconductor GaAs material. Figure 2.24(b) shows the relationship between open-circuit voltage and short-circuit current and light intensity.

It must be pointed out that V_0 does not increase infinitely with the intensity of light. When the photovoltaic voltage V_0 increases to the disappearance of PN junction barrier, the maximum photovoltaic

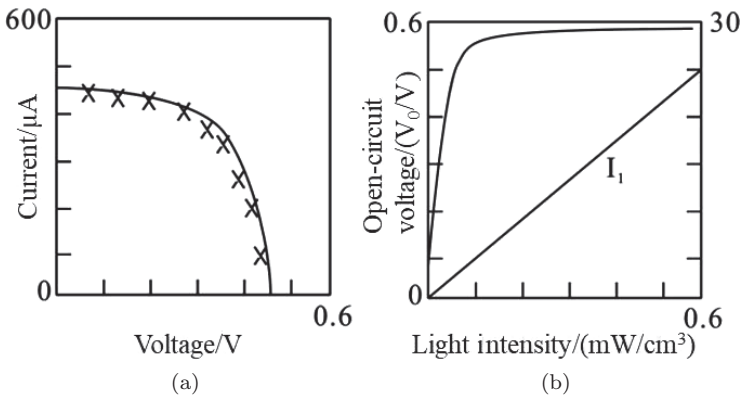


Fig. 2.24. (a) I - V relationship of GaAs photodiodes; (b) I - V relationship of photodiodes and light intensity.

voltage $V_{\max}$ is obtained, which is related to the doping degree of the material. In practice, $V_{\max}$ is equivalent to the bandgap width E_g .

How to make the load obtain the highest possible voltage or current? In order to obtain the highest voltage and large current, on the one hand, several photodiodes are used in series and parallel, on the other hand, the photoelectric conversion efficiency of solar cells is improved. A solar cell efficiency parameter η is introduced, which is defined as

$$\eta = \frac{\text{Power Consumption in Load}}{\text{Light power incident on the junction area}}. \quad (2.69)$$

The theoretical calculation shows that the bandgap of semiconductor materials is 1.1–1.5 eV, and the utilization efficiency of sunlight is the highest. The bandgap of single crystal silicon is 1.1–1.5 eV. People try to make photovoltaic cells from single crystal silicon by doping, but its price is expensive. At the same time, single crystal silicon is an indirect bandgap semiconductor material with low photoelectric conversion efficiency. It has been found that substituting doped amorphous silicon for monocrystalline silicon not only reduces the cost, but also improves the photoelectric conversion efficiency.

One of the most important applications of photovoltaic effect is the direct conversion of solar energy into electrical energy. Solar cells are typical photovoltaic cells, generally consisting of a large area of silicon PN-junction. In fact, there are many kinds of photovoltaic cells. In the early stage, cuprous oxide photovoltaic cells were seldom used because of their low conversion efficiency. At present, selenium photovoltaic cells and silicon photovoltaic cells are widely used. Selenium photovoltaic cells are mostly used in analysis and measurement instruments such as exposure meters and Illuminometers because of their similar spectral characteristics and wide spectrum. Compared with other semiconductor photocells, silicon photocells not only have stable performance, but also have the highest conversion efficiency (up to 17%) which is almost close to the theoretical limit. In addition, there are thin-film photovoltaic cells, purple photovoltaic cells, heterojunction photovoltaic cells and so on.

Thin-film photovoltaic cells are made of cadmium sulfide and other materials into thin-film structures to reduce weight, simplify array structure, improve radiation resistance and reduce costs. Purple photovoltaic cells reduce the PN junction of silicon photovoltaic cells to $0.2\text{--}0.3\text{ }\mu\text{m}$ junction depth and shift the peak spectral response to about 600 nm to improve the short-wave response, so as to adapt to the use of outer space.

Heterojunction photovoltaic cells use semiconductor materials with different bandgaps to make heterojunction PN. The incident light almost passes through one side of the wide bandgap material, and is absorbed in the narrow bandgap material in the junction region, resulting in electron–hole pairs. This “window” effect can be used to improve the collection efficiency of incident light in order to obtain a higher conversion efficiency than that of homogeneous junction silicon photovoltaic cells. The maximum conversion efficiency can be up to 30% theoretically. However, the technology is not yet mature and the conversion efficiency is still lower than that of silicon photovoltaic cells.

The core part of the photovoltaic cell is a PN junction, which is generally made into a large sheet to receive more incident light. Figure 2.25 is a schematic diagram of the structure of selenium photovoltaic cells. The manufacturing process is as follows: firstly, a layer of P-type selenium is coated on the aluminum sheet, then a layer of cadmium is evaporated, then N-type cadmium selenide is formed after heating, forming a large area PN-junction with the

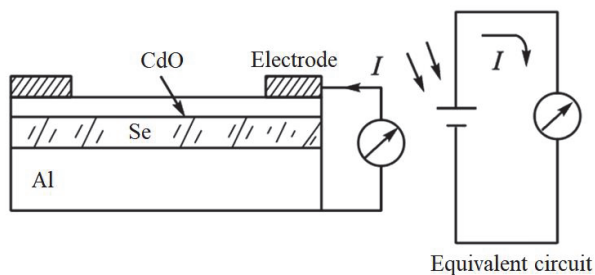


Fig. 2.25. Structural diagram of selenium photovoltaic cell.

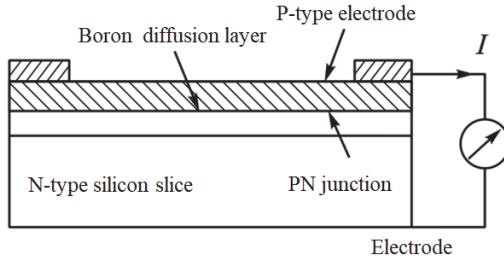


Fig. 2.26. Structure diagram of silicon photovoltaic cell.

original P-type selenium, and finally coated with a semi-transparent protective layer, welded on the electrode, aluminum sheet as positive electrode, cadmium selenide as negative electrode.

Silicon photovoltaic cells are made of monocrystalline silicon (there are also amorphous silicon products at present), as shown in Fig. 2.26. A diffusion PN (P + N) junction is formed by diffusing P-type impurities (e.g. boron) on an N-type silicon wafer or by diffusing N-type impurities (e.g. phosphorus) on a P-type silicon wafer to form a PN junction of N + P and then welding two electrodes. P-terminal is the positive electrode of photovoltaic cell, and N-terminal is the negative electrode. Generally, P + N-type photodetectors are used on the ground, such as domestic 2CR-type photodetectors. N + P-type silicon photovoltaic cells have strong radiation resistance and are suitable for space applications. They can be used as solar cells in space, such as domestic 2DR type.

Semiconductor photovoltaic effect is also widely used in radiation detectors, including light radiation and other radiation. Its advantage is that it does not need external power supply, but generates unbalanced carriers by radiation or high-energy particle excitation, and detects radiation or particle current intensity by measuring photovoltaic voltage.

2.3.4. Thermoelectric effect

When two different pairing materials (metal or semiconductor) are welded in parallel, if the temperature of the two joints is different,

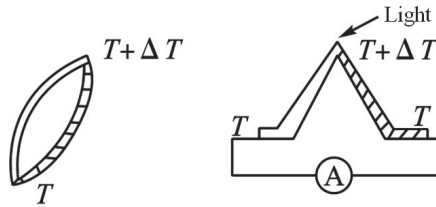


Fig. 2.27. Thermoelectric effect.

an electromotive force (EMF) is generated in the parallel circuit, which is called thermoelectric EMF. If there is an electromotive force, there will be a current flow in the circuit, as shown in Fig. 2.27. If the cold end is separated and connected with an ammeter, then when the light fusion end (called the coupler joint) absorbs light energy, the temperature of the coupler joint rises, and the ammeter has the corresponding current reading. The current value indirectly reflects the size of the light energy. This is the principle of using thermocouples to detect light energy. In practice, in order to improve the measurement sensitivity, several thermocouples are often used in series, called thermopile, which is used in laser energy meter.

2.3.5. Pyroelectric effect

Pyroelectric effect refers to the release of partial charges adsorbed on the surface of some crystals with the change of polarization intensity with temperature, which is achieved by so-called pyroelectric materials. Pyroelectric material is a kind of piezoelectric crystal with poor crystalline symmetry, so it has self-generating polarization (i.e. inherent electric dipole moment) under normal conditions. According to the electromagnetic theory, surface bound charge appears on the surface of material perpendicular to the polarization vector P_s , and the surface charge density $\sigma_s = |P_s|$. Because the polarization vectors of spontaneous generation in the crystal are disorderly arranged, the total P_s are not large. Additionally, the neutralization of external free charges near the surface of the material usually does not detect the existence of surface charges. If the DC electric field is applied to the thermoelectric body, the spontaneous polarization vectors will

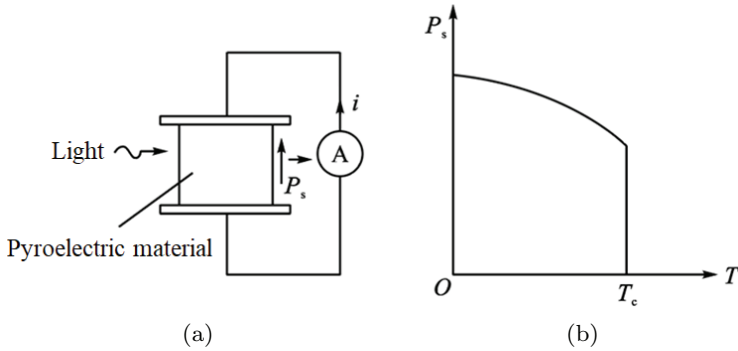


Fig. 2.28. Pyroelectric effect.

tend to align uniformly (forming single domain polarization), and the total P_s will increase. When the electric field is removed, the thermoelectric body is sometimes called thermoelectric-ferroelectric if the total P_s can still be maintained. It is an ideal material for realizing thermoelectric phenomena.

Thermoelectric $|P_s|$ determines the surface charge density σ_s , and when $|P_s|$ emission changes, σ_s also changes. The single-domain thermoelectric maintains a larger $|P_s|$. $|P_s|$ is a function of temperature, as shown in Fig. 2.28(b). Temperature increases and $|P_s|$ decreases. When the temperature rises to T_c , the natural polarization suddenly disappears. T_c is called Curie temperature. Pyroelectric phenomena occur only below T_c . When the intensity of light irradiates the thermoelectric body, the temperature and size of the thermoelectric body change, and the surface charge changes accordingly. It is very important that the free charge near the surface of the thermoelectric body neutralizes the surface charge slowly. Good thermoelectric, the process is slow. Before neutralization, the surface charge on the side of the thermoelectric body changes with respect to temperature, which is called pyroelectric phenomenon. If the thermoelectric body is placed between the capacitor plates and a ammeter is connected with both ends of the capacitor, there will be a current flowing through the ammeter. This current is called short-circuit pyroelectric current, as shown in Fig. 2.28. If the

plate area is A , the current is

$$i = A \frac{dP_s}{dt} = A \frac{dP_s}{dT} \frac{dT}{dt} = A\beta \frac{dT}{dt}, \quad (2.70)$$

where $\beta = dP_s/dT$ is called pyroelectric coefficient. Obviously, if the illuminated light is constant, then T is constant, P_s is constant, and the current is zero. So pyroelectric detector is an AC or instantaneous response device.

2.3.6. Photon traction effect

When the energy of a photon beam is not enough to cause the electron-hole pair to irradiate a sample, a potential difference can be established at both ends of the sample in the direction of the beam, and the magnitude of the potential difference is proportional to the light power. This is called the photon pull effect. Its research began in the 1960s. It was 1970 that laser was used as light source.

Photons have momentum. When photons collide with carriers in materials, they can transfer momentum to carriers. As a result, the directional motion of carriers is induced in the direction of the beam, and charge accumulation is formed at the end surface, resulting in additional electric field. When the electric field force of the additional electric field on the carrier is balanced with the impulse produced by the photon, a stable potential difference is established.

Because the photon drag potential difference symbols produced by electrons and holes are opposite, this effect is very weak for intrinsic semiconductors. To observe the photon traction effect, N-type semiconductor or P-type semiconductor must be selected.

In the photoelectric effects related to carrier generation mentioned above, photon energy is transmitted to electrons, while photon traction effect ignores carrier generation and transfers photon momentum to electrons. This transfer occurs very quickly, and the detector made of photon traction effect has a fast response (response time is about 10^{-9} s).

At present, available photon traction detectors can be found in the band of 1–10000 μm , and their sensitivity can reach 0.1–40 $\mu\text{V/W}$.

References

- Chaikin P M, Lubensky T C, 2001. *Principles of Condensed Matter Physics*. Cambridge: Cambridge University Press.
- Cheng T, 2013. *The Foundations of Modern Quantum Mechanics*, Beijing: Peking University Press.
- Fan M, Yao J, 2013. *Optical Functional Materials Science*, Beijing: Science Press.
- Huang K, Xie X, 2012. *Physics of Semiconductor*, Beijing: Science Press.
- Liu E, Zhu B, Luo J, 2003. *Semiconductor Physics*, Shang Hai: Shanghai Scientific and Technical Publishers.
- Xie X, Lu D, 2007. *Band Theory of Solids*, Shang Hai: Fudan University Press.
- Yang Y, He G, Ma J, 2009. *Photoelectric Information Technology*, Shang Hai: Donghua university press.
- Zeng J, 2003. *Course in Quantum Mechanics*, Beijing: Science Press.

This page intentionally left blank

Chapter 3

Micro/Nano Photoelectric Materials and Devices

The function and performance of optoelectronic devices are greatly limited by materials. This limitation comes from the physical and chemical properties of materials themselves. Many specific functions and properties expected by human beings cannot be realized. However, the physical and chemical properties of materials will change dramatically by introducing micro- and nano-structures or reducing the size of materials to the order of μm and nanometer, so they can be used for the development and preparation of new materials and devices. In recent years, the rapid development of micro/nano materials has benefited from this, especially the rapid development of micro/nano optoelectronic materials. For example, nano-optoelectronic materials including nanoparticles, nanowires and nanofilms, photonic crystals with special optical properties through the periodic distribution of dielectric constants, supermaterials with negative refraction or electromagnetic stealth properties, etc., are extremely novel and have great application prospects. Due to the rapid development of micro/nano optoelectronic materials, the research of micro/nano optoelectronic devices is also very active. On the one hand, traditional optoelectronic devices are fabricated with new photoelectric materials in order to achieve better performance. On the other hand, new optoelectronic devices are developed based on new materials to achieve functions that traditional optoelectronic materials cannot achieve. In this chapter, the development of materials and devices in the fields of nano photoelectric materials,

photonic crystals, metamaterials and plasma polaritons will be briefly summarized. Because any one of these fields has a wide range of connotations and applications, and with the rapid change of people's in-depth research, limited to the length of the textbook and the ability of the editor, the description here will be more concise. Interested readers can refer to the latest literature monographs at home and abroad for further in-depth understanding.

3.1. Nanometer Photoelectric Materials and Devices

3.1.1. *Nano photoelectric materials*

With the deepening of people's research on the basic structure and law of matter, human beings have been able to manipulate matter at atomic and molecular scales. Depending on this ability, a new discipline, nanoscience and technology, has been developed in recent years. It studies the basic structure, properties and engineering applications of substances at nanoscale. Materials with at least one dimension of 0.1–100 nm in three-dimensional space are called nanomaterials. When the scale of materials is reduced to nanometer scale, they will show many strange characteristics, which is the driving force of nanomaterials research. According to the dimensions, nanomaterials can be divided into nanoparticles, nanowires and nanofilms.

Nano-scale photoelectric materials, namely nanomaterials with special photoelectric properties, are called nano photoelectric materials. The optical and electrical properties of nano-optoelectronic materials are different from those of macro-optoelectronic materials. This mainly comes from the following three aspects.

- (1) *Small size effect.* Because the size of nanoparticles is close to the physical characteristics such as light wavelength and electron de Broglie wavelength, the acoustic, optical, electrical, magnetic and thermodynamic properties of materials are changed.
- (2) *Surface effect.* The size of nanoparticles decreases, the surface area increases, and the surface density of states increases, which not only changes the atomic transport and configuration of

the material surface, but also changes the surface electron spin distribution and energy spectrum.

- (3) *Quantum size effect*. Because electrons are limited in the three-dimensional direction, the energy levels of electrons show a discrete-level structure similar to that of atoms, which makes the absorption and light-emission properties of materials different from those of macroscopic materials.

Nano-optoelectronic materials have wide application value. The research on the basic properties of nano photoelectric materials, the preparation of nano photoelectric materials, and the development, design and fabrication of devices based on nano photoelectric materials are the current international research hotspots. There are many kinds of nano photoelectric materials, such as nano-luminescent materials, nano photoelectric conversion materials and nano-photocatalytic materials.

3.1.1.1. *Nano-luminescent materials*

In 1994, Bharagava *et al.*, reported for the first time that the luminescence lifetime of transition metal ions doped nano-semiconductor particle ZnS:Mn was reduced by five orders of magnitude, and the external quantum efficiency was still up to 18%. Although the experimental results are controversial, they have aroused great interest in nano-luminescent materials.

Nano-luminescent materials mainly use nanoparticles with the size of 1–100 nm as the luminescent matrix, including pure and doped nano-semiconductor luminescent materials, rare earth ions and transition metal ions doped nano-oxides, sulfides, composite oxides and various nano-inorganic salts luminescent materials. Nano-luminescent materials are mainly used in the design and fabrication of luminescent devices, which can achieve the luminescent properties that macro-bulk materials do not have.

3.1.1.2. *Nano photoelectric conversion materials*

Nano photoelectric conversion material is a kind of nanomaterial which can directly convert light energy into electric energy.

Nano photoelectric conversion materials are mainly used for solar cells, which can improve the conversion efficiency of solar cells and realize the efficient use of green energy solar energy.

3.1.1.3. *Nano-photocatalytic materials*

Nano-photocatalytic material is a kind of nanomaterial which converts light energy into chemical energy. Nanoparticles have high specific surface area and high surface defect density, so they can be used as efficient catalytic materials because of their relatively higher reactivity than macromaterials. As a highly active and selective catalyst, nanocatalytic materials have attracted widespread attention. Among them, nanophotocatalytic materials are also the focus of research, which use nanomaterials to accelerate the chemical reaction under light conditions. Nano-TiO₂ is a very typical nano-photocatalyst. Because it can effectively treat many toxic compounds, it has important applications in air purification and water treatment.

People attach great importance to the research of nano photoelectric materials. At present, many methods have been used to prepare nano photoelectric materials, such as chemical precipitation, sol-gel, hydrothermal synthesis, laser-induced vapor deposition, etc. With the deepening of research, people also study nano-optoelectronic devices based on nano-optoelectronic materials, such as quantum dot light-emitting diodes, quantum dot lasers, quantum dot single photon detectors, nanowire light-emitting diodes, nanowire optical waveguides, nanowire lasers, nanowire photoelectric sensors and so on. In fact, quantum well semiconductor lasers, which have been commercialized successfully, also belong to the category of nano-optoelectronic devices.

3.1.2. *Nanometer photoelectric devices*

There are many kinds of nano-optoelectronic devices based on nano-optoelectronic materials, including nano-light-emitting devices, such as quantum dot lasers, quantum dot light-emitting diodes, nanowire lasers, nanowire light-emitting diodes, quantum well lasers, quantum well light-emitting diodes, quantum cascade lasers, and

nano-detection and sensor devices, such as quantum dot light detectors, quantum dot single-chip lasers, etc. Photon detectors, nanowire sensors, nano-optical memory devices, such as quantum dot optical memory, nano-optical transmission devices, such as nano-optical fibers, nano photoelectric converter devices, such as quantum dot solar cells, in addition to nano-optical switches, light modulators and so on. These devices cover all aspects of optoelectronic devices and will undoubtedly greatly change the face of optoelectronics with the progress of research. This book cannot cover all aspects of the research results of nano-optoelectronic devices, only a brief introduction of some typical nano-optoelectronic devices.

3.1.2.1. *Quantum dot optoelectronic devices*

Quantum dots are nanomaterials consisting of a small number of atoms that constrain the movement of electrons in three-dimensional directions. Quantum dots have many characteristics similar with isolated atoms, so they are also called “artificial atoms”. Quantum dots can be used to design and fabricate many photoelectric devices with excellent performance, such as quantum dot solar cells, quantum dot light-emitting diodes, quantum dot lasers and so on.

(1) Quantum dot solar cells

Solar cell is a kind of photoelectric converter which directly converts light energy into electric energy. It has important application value for human society. However, because the fact that only light energy larger than the gap of the semiconductor can be absorbed in the semiconductor and that only one photon can excite one electron–hole pair, the conversion efficiency of ordinary solar cells is low. Quantum dot solar cells can theoretically achieve a maximum conversion efficiency of 65%, which is twice the maximum conversion efficiency of ordinary solar cells. There are three main types of quantum dot solar cells: PIN structure quantum dot solar cells, quantum dot sensitized solar cells and quantum dot solar cells based on multi-exciton effect. At present, quantum dot solar cells are still in the research stage. Once the high conversion efficiency quantum dot solar

cells are successfully developed and put into use, it will inevitably have a far-reaching impact on the way people use energy.

(2) Quantum dot light-emitting diodes

The structure of quantum dot light-emitting diodes is similar with OLED. The light-emitting layer consisting of quantum dot materials is sandwiched between the electron transport layer and the hole transport layer consisting of organic materials. Under the action of an applied electric field, electrons and holes are injected into the quantum dot layer, which is captured by quantum dots and recombined to produce light radiation. Due to the limitation of semiconductor band structure, ordinary light-emitting diodes usually emit only a certain wavelength of monochromatic light. By changing the size and composition of quantum dots, the light emitted by quantum dot light-emitting diodes can cover near infrared and visible light. Quantum dot light-emitting diodes have important application prospects in display technology. In addition, it can also be used as a fluorescent probe for biomolecular and cellular imaging.

(3) Quantum dot laser

Quantum dot lasers have no difference in structure from conventional semiconductor lasers, but quantum dot materials are used in the active region. Because quantum dots have electronic energy-level structure similar to isolated atoms, the laser characteristics of quantum dot lasers are similar to gas lasers, and some shortcomings of ordinary semiconductor lasers and quantum well lasers can be overcome. Compared with conventional semiconductor lasers and quantum well lasers, quantum dot lasers have lower threshold current density, higher luminous efficiency and differential gain, narrower spectral linewidth and better temperature stability.

3.1.2.2. *Nanowire optoelectronic devices*

Nanowires are quasi-one-dimensional nanomaterials. They can also be used to design and fabricate various nano-optoelectronic devices, such as nanofibers, nanowire light-emitting diodes, nanowire lasers, nanowire photovoltaic cells, nanowire photodiodes, etc. Compared

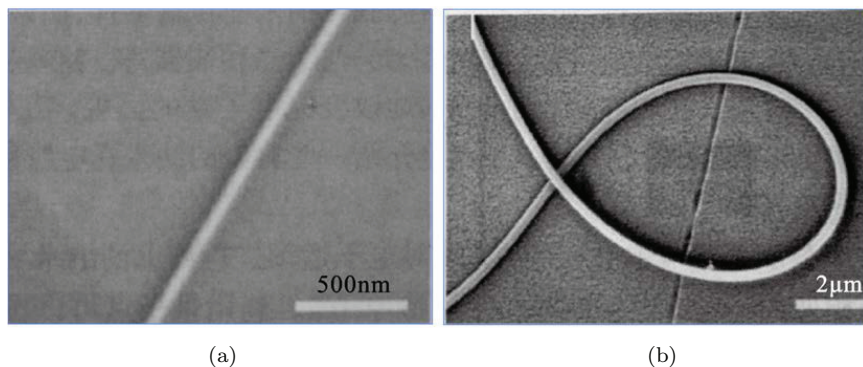


Fig. 3.1. Nanofibers. (a) Silicon dioxide nanofibers with a diameter of about 50 nm. (b) Nanofibers with a bending diameter of 280 nm and a bending radius of about $2.7 \mu\text{m}$.

with quantum dot optoelectronic devices, nanowire optoelectronic devices have their own unique characteristics and unique applications. Here is a brief introduction to nano-optical fibers. Nanofibers are optical fibers with diameters ranging from tens to hundreds of nanometers, as shown in Fig. 3.1. Because the diameter of the optical fiber is smaller than the wavelength, it is also called sub-wavelength optical fiber.

At present, there are many different methods to prepare nano-optical fibers. The most typical preparation method is to heat and elongate ordinary optical fibers on the flame. Tong Limin of Zhejiang University successfully used this method to fabricate nanofibers with a diameter of 50 nm in 2003. This kind of optical fiber has the advantages of simple structure, high uniformity, low transmission loss and high mechanical strength, and can be easily coupled and integrated with existing optical fiber systems.

Nano-optical fibers have special optical properties, and have special research and application value. For example, because the electromagnetic field of light wave in nanofibers is limited to a very small range and has a very high power density, supercontinuum spectrum can be generated by nanofibers; nanofibers have high mechanical strength and can be bent to the μm level, so nanofibers can be used to fabricate very small ring resonators for various

microfilters and lasers. Because of the strong evanescent wave field around the fiber, nanofibers can be used to fabricate chemical and biological sensors.

3.2. Photonic Crystals and Photonic Crystal Devices

Photonic crystal is a material structure with periodic refractive index distribution in space. It is usually composed of two or more materials with different refractive index. This periodic structure will produce Bragg scattering for electromagnetic wave propagating in a specific direction. If the difference of material and low refractive index of photonic crystal is large enough, part of the frequency band of electromagnetic wave cannot propagate in any direction, and photonic crystal is equivalent to the “insulator” of light, photons will be formed. *Photonic bandgap* (PBG): Outside the photonic bandgap, electromagnetic waves propagate in photonic crystals in the form of Bloch waves. Although the microscopic structure of photonic crystals has a strong scattering effect on electromagnetic waves, this Bloch wave has a definite directivity.

Photonic crystals can control the radiation and propagation of light flexibly and effectively, so they have wide and important application value. Since Yablonovitch and John independently proposed the concept of photonic crystals in 1987, photonic crystals have attracted much attention from researchers all over the world. They have become a major research hotspot in the fields of electromagnetics and photonics. They have made important breakthroughs in theory, experiment and application. Some research results have been successfully commercialized.

3.2.1. Structure of photonic crystals

According to the dimension, photonic crystals can be divided into one-dimensional, two-dimensional and three-dimensional, as shown in Fig. 3.2. One-dimensional photonic crystals exhibit periodicity only in one direction, while they exhibit translation invariance in the other two orthogonal directions. For example, a Bragg mirror can be considered as a one-dimensional photonic crystal. The study

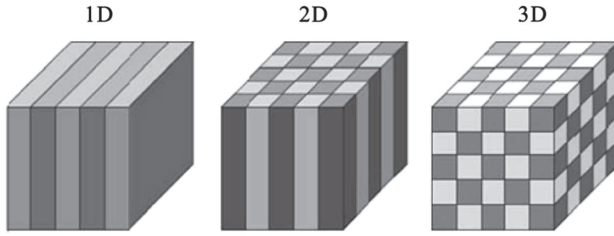


Fig. 3.2. Structural sketches of one-dimensional, two-dimensional and three-dimensional photonic crystals.

of one-dimensional photonic crystals can be traced back to 1887, when Rayleigh firstly studied the electromagnetic properties of this periodic layered dielectric structure. The refractive index of two-dimensional photonic crystals exhibits periodicity in the two-dimensional direction, while it satisfies translation invariance in the vertical direction of the two-dimensional. Krauss successfully fabricated two-dimensional near infrared photonic crystals for the first time in 1996 by using semiconductor material processing technology. Three-dimensional photonic crystals exhibit periodicity in three directions. This kind of photonic crystal can be fabricated by self-assembly, laser holography and electron beam lithography. However, it is still difficult to fabricate complex three-dimensional photonic crystal structures and apply them in practice. In fact, photonic crystals, such as natural proteins (opal), butterflies, beetles and other organisms, were invented and utilized earlier than human beings, and they usually show brilliant colors, which are related to their internal photonic crystal structure.

3.2.2. Basic properties of photonic crystals

3.2.2.1. Photonic crystal forbidden band

In photonic crystals, the refractive index (or dielectric constant) is periodically distributed, as shown in Fig. 3.3. When the periodicity is comparable to the optical wavelength and the refractive index changes sufficiently, the bandgap appears in the photonic crystal. The light corresponding to the forbidden band frequency does not

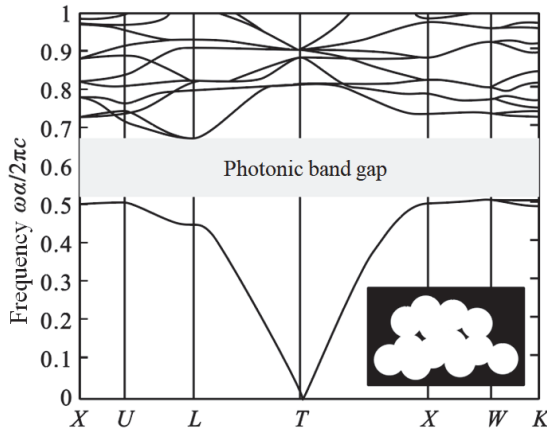


Fig. 3.3. Photonic crystal bandgap (three-dimensional diamond structure).

have any mode of electromagnetic wave propagation, so the light is strictly prohibited. For example, if a beam of light whose frequency is within the photonic crystal forbidden band is incident on the surface of the photonic crystal, it will be completely reflected because it cannot propagate in the photonic crystal. By breaking the periodicity of local region in photonic crystals and introducing point defect or line defect into photonic crystals, light can be confined to the point defect or line defect, thus forming photonic crystal microcavities or photonic crystal waveguide devices. The photonic crystal waveguide with photonic crystal forbidden band has a very good restraint effect on the optical frequency electromagnetic field, and can realize the right angle or even greater angle bend transmission of light wave in very small scale, so it is very helpful to reduce the size of integrated optoelectronic devices.

3.2.2.2. Abnormal dispersion

Because of the periodic distribution of refractive index, when light propagates in photonic crystals (forbidden band frequency), the special dispersion relationship between frequency ω and propagation vector κ , i.e. $\omega = \omega(\kappa)$, can be obtained by calculating the band structure of photonic crystals. Due to the special dispersion characteristics of photonic crystals, photonic crystals show some new

effects and phenomena, such as hyperprism, negative refraction, self-collimation, etc. According to the band structure of photonic crystals, the group velocity $\nu_g = d\omega/d\kappa$ can be further obtained. The group velocity is strongly influenced by photonic crystals, which can be far lower than that of vacuum light or even equal to zero. Since the group velocity represents the speed of light energy propagation, photonic crystals can be used to control the speed of light propagation and achieve slow light.

3.2.2.3. *Inhibition and enhancement of spontaneous radiation*

Photonic crystals can control spontaneous emission. The spontaneous emission probability of an atom is

$$W = \frac{2\pi}{h} |V|^2 \rho(E). \quad (3.1)$$

In the formula, $|V|$ is a matrix element with zero points, and $\rho(E)$ is the density of states of the light field. Obviously, the probability of spontaneous emission is proportional to the density of states in the field. The spontaneous emission can be controlled by controlling the density of states of the light field. In photonic crystals, the density of states of light field is modulated by photonic crystals. In the bandgap frequency range of photonic crystals, the density of states of the optical field is zero. If the atom is in the photonic crystal, if its spontaneous emission frequency falls in the forbidden band, the spontaneous emission probability is zero and the spontaneous emission is completely suppressed. If a defect exists in the photonic crystal, the defect state of QF is introduced into the forbidden band of the photonic crystal. The defect state has a high density of states in the optical field, and the spontaneous emission of the corresponding frequency is enhanced.

3.2.3. *Photonic crystal devices*

With the special properties of photonic crystals, many kinds of photonic crystal devices can be fabricated. For example, by introducing point defects into photonic crystals, a photonic crystal microresonator is formed and used in photonic crystal lasers.

Photonic crystal fibers and photonic crystal waveguides are formed by introducing line defects into photonic crystals. High efficiency photonic crystal light-emitting diodes can also be fabricated by using the photonic crystal bandgap effect. Using the unique dispersion characteristics of photonic crystals in the non-bandgap region, open resonators, polarization beam splitters and compact wavelength division multiplexers can be designed and fabricated. Because photonic crystals can be used as a platform to fabricate and integrate various optoelectronic devices, photonic crystals have opened broad application prospects for optical integration, photonic chips and optical computing.

3.2.3.1. *Photonic crystal fibers*

Photonic crystal fiber (PCF), also known as microstructural fiber or holey fiber, is another two-dimensional photonic crystal waveguide. Unlike planar two-dimensional photonic crystal waveguide, the line defect used to restrain and transmit light wave in photonic crystal fiber is in the same direction as the dielectric column (or air hole), and the light is perpendicular to the medium column (or air hole). Directional propagation is in the periodic plane of photonic crystals. Since Knight of Southampton University and his collaborators produced the world's first photonic crystal fiber (Fig. 3.4) in 1996, due to its unique performance and flexible design, the research of photonic crystal fibers has attracted wide attention. Various types and special functions of photonic crystal fibers have been proposed and fabricated one after another, and some of the research results have been commercially applied. Nowadays, photonic crystal fibers have become a mature field in the research of photonic crystals.

According to the different guided wave mechanism, photonic crystal fibers can be divided into two categories: index guided PCF and PBG guided PCF. The former utilizes the total internal reflection effect, because air micro-holes are introduced into the cladding, the average refractive index of the cladding decreases. If the core is solid, the average refractive index of the cladding is lower than the refractive index of the core, so the light wave is confined to the core

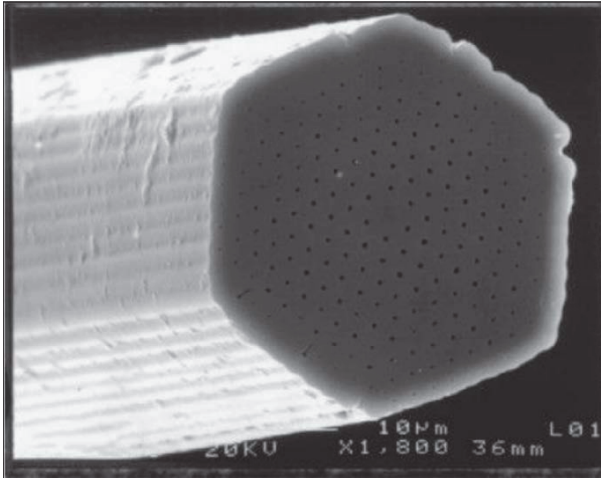


Fig. 3.4. The first photonic crystal fiber in the world. It has the characteristics of single mode without cut-off.

in a way similar with that of traditional optical fibers. The latter utilizes the photonic bandgap effect, and two-dimensional photonic crystals can produce photonic bandgap, which is the bandgap barrier. If a defect is introduced in a two-dimensional photonic crystal along the direction perpendicular to the photonic crystal plane, the light can be confined to propagate along the defect in the two-dimensional photonic crystal plane. Because of the confinement effect of photonic bandgap, the refractive index of the core can be lower than the average refractive index of the cladding, and even air can be used to form hollow light in crystal fibers.

Compared with traditional optical fibers, photonic crystal fibers have improved significantly in many properties, and have some advantages that traditional optical fibers do not possess, such as no cut-off single mode, large mode area, cheek birefringence, etc. In addition, the dispersion of photonic crystal fibers can be controlled to a certain extent by the structure design, so various dispersion compensation fibers and dispersive flat-port photonic crystal fibers can be designed. Photonic crystal fibers can also enhance optical nonlinearity and improve the efficiency of various nonlinear optical

effects. Among them, the generation of supercontinuum spectrum by photonic crystal fibers is one of the highlights of current research.

In short, photonic crystal fibers are flexible in design and can possess various excellent performances that traditional optical fibers do not possess. They are widely used in communication, sensing, laser and nonlinear optics, and their research has exerted great influence on these fields. It can be said that photonic crystal fiber technology is an important progress in the field of optical fibers in recent years.

3.2.3.2. *Photonic crystal light-emitting diodes*

Although the luminescent diodes used by people have very high internal quantum efficiency, which can reach more than 90%, the external quantum efficiency is usually only 3–20%. Most of their energy is limited to the material internal reabsorption or conversion of heat energy, and most of the energy is wasted. Photonic crystal light-emitting diodes (PCLD) fabricated by photonic crystals can greatly improve the luminous efficiency and control the direction of light emission.

Photonic crystal light-emitting diodes are two-dimensional photonic crystal microcavity structures fabricated on thin-film light-emitting diodes (SLD). Because the luminous region is surrounded by two-dimensional photonic crystals, the light emitted from the luminous region cannot propagate in the thin-film due to the photonic crystal forbidden band, and can only be coupled to the vertical direction and emitted to the outer space, thus greatly improving the light yield. In 2009, Philips Lumileds scientists designed and fabricated InGaN photonic crystal light-emitting diodes to achieve up to 73% luminous output. Matsushita is also actively studying photonic crystal light-emitting diodes. They use sapphire crystal as a substrate to produce photonic crystal light-emitting diodes. The light-emitting efficiency of the photonic crystal light-emitting diodes is 50% higher than that of ordinary light-emitting diodes, and the theoretical value should be three times that of ordinary light-emitting diodes.

Through the improvement of process and device structure, the characteristics of photonic crystal light-emitting diodes still have a lot of room to improve, and will greatly improve the basic characteristics of light-emitting diodes (including light efficiency).

3.2.3.3. *Photonic crystal lasers*

Photonic crystals used in laser technology can effectively break through the technical bottleneck faced by traditional laser technology and achieve laser performance that traditional methods cannot achieve. Using photonic crystals, a micro-resonator with Q value can be formed to reduce the laser threshold and achieve a low threshold laser. Using photonic crystals to design micro-resonators can significantly improve the noise characteristics of lasers because the number of cavity modes is greatly reduced. In addition, photonic crystal lasers without geometric boundaries can be designed and realized by using the tailoring effect of photonic crystal band structure on the density of states and the low group velocity characteristics of photonic crystal band at some special positions. By adjusting the structure and refractive index parameters of photonic crystals, fine tuning of laser wavelength can also be achieved. In terms of structure, photonic crystal lasers are mainly photonic crystal defect microcavity lasers and photonic crystal ring waveguide lasers. This book introduces the former.

In 1999, researchers from California Institute of Technology and University of Southern California successfully fabricated a point defect two-dimensional photonic crystal laser for the first time, which realized room temperature-pulsed laser radiation. Photonic crystal microcavity structure was fabricated on InGaAsP plate by complex etching technology, in which an air hole was removed from the central position of photonic crystal. The photonic crystal microcavity achieves three-dimensional constraints of light through two mechanisms: total internal reflection in vertical direction and photonic crystal bandgap in planar plane. The active layer of the laser is an InGaAsP/InP strained quantum well structure with a designed output wavelength is 1.55 μm . The lattice period of photonic crystal

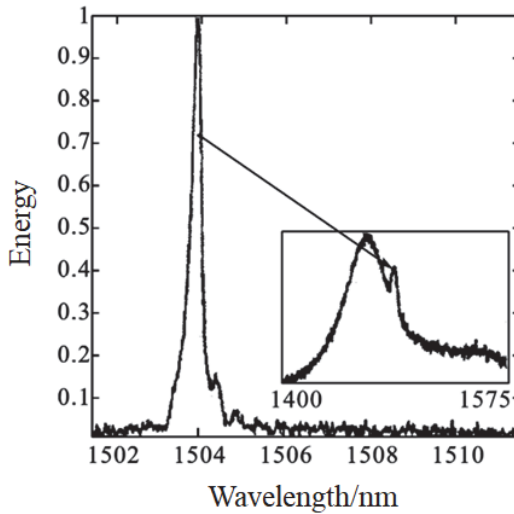


Fig. 3.5. Pulse laser output (central wavelength 1504 nm) (illustrated spontaneous emission spectrum below threshold).

is 515 nm, the radius of air hole is 180 nm, and the radius of two large air holes around the defect is 240 nm. The simulation results show that the quality factor of the resonator can reach 250. Pulsed laser output with a central wavelength of 1504 nm at room temperature is realized by using a 830 nm semiconductor laser, as shown in Fig. 3.5.

In the near infrared band, photonic crystal lasers are mainly studied in the active region of InGaAsP and InGaAs quantum wells. In 2003, Monat *et al.* successfully fabricated InP-based photonic crystal microcavity laser on silicon wafer. The width of microcavity is 5 μm , the quality factor is 1000, the emission wavelength is 1465 nm, and the laser threshold is 10 mW. Thereafter, photonic crystal microcavity lasers with different processes have been developed. In the short wavelength band, the blue-violet laser is emitted by the wide bandgap GaN and ZnO materials. Noda's team used GaN/InGaN quantum wells to implement 406.5 nm surface-emitting photonic crystal lasers. Scharrer *et al.* fabricated three-dimensional photonic crystals of ZnO and realized ultraviolet laser with wavelength of 330–383 nm. In addition, there are silicon-based hybrid photonic crystal microcavity lasers.

3.3. Metamaterials and Related Devices

3.3.1. Metamaterials

Metamaterial, or special material, is a kind of artificial electromagnetic material, which possesses super electromagnetic (optical) properties that natural materials do not possess. These properties are derived from the special microstructures of artificial designs. At present, the research on metamaterials mainly focuses on the negative index materials and invisibility cloak, which are the research hotspots in recent years.

3.3.2. Negative refractive index materials and devices

Negative refractive index material is a medium in which the permittivity and permeability are both negative. Veselago, a Soviet scientist, firstly put forward the concept of negative refractive index in 1967, but human beings have not been able to find this material in nature, so the concept of negative refractive index has not attracted people's attention. From 1996 to 1999, Pendry *et al.* proposed that using metal wires and split ring resonator (SRR) can realize the simultaneous negative dielectric constant and permeability in microwave band. In 2001, Shelby *et al.* first confirmed the negative refractive properties of negative refractive index materials in experiments. So far, the negative refractive index materials have regained their importance.

3.3.2.1. Physical properties of negative refractive index materials

Electromagnetic wave propagates in a homogeneous, isotropic and lossless medium, which satisfies Maxwell's equations:

$$\begin{aligned}
 \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t}, \\
 \nabla \times \mathbf{H} &= -\frac{\partial \mathbf{D}}{\partial t}, \\
 \nabla \cdot \mathbf{B} &= 0, \\
 \nabla \cdot \mathbf{D} &= 9.
 \end{aligned}
 \tag{3.2}$$

In the formula, $\mathbf{D} = \varepsilon \mathbf{E} = \varepsilon_0 \varepsilon_r \mathbf{E}$, $\mathbf{B} = \mu \mathbf{H} = \mu_0 \mu_r \mathbf{H}$. The propagation properties of electromagnetic waves are determined by the dielectric constant ε and permeability μ of the dielectrics.

It is assumed that the electromagnetic wave is a homogeneous plane wave with a satisfying relation of electric field vector $\mathbf{E}$, magnetic field vector $\mathbf{H}$ and wave vector $\boldsymbol{\kappa}$, i.e.

$$\begin{aligned}\boldsymbol{\kappa} \times \mathbf{E} &= \omega \mu_r \mu_0 \mathbf{H}, \\ \boldsymbol{\kappa} \times \mathbf{H} &= -\omega \varepsilon_r \varepsilon_0 \mathbf{E}.\end{aligned}\tag{3.3}$$

For ordinary dielectric materials with positive dielectric constant ε and permeability μ , the electric field vector, magnetic field vector and wave vector satisfy the right-hand relationship. For dielectric materials whose dielectric constant ε and permeability μ are negative at the same time, there are

$$\begin{cases} \boldsymbol{\kappa} \times \mathbf{E} = -\omega |\mu_r| \mu_0 \mathbf{H}, \\ \boldsymbol{\kappa} \times \mathbf{H} = \omega |\varepsilon_r| \varepsilon_0 \mathbf{E}. \end{cases}\tag{3.4}$$

The electric field vector $\mathbf{E}$, magnetic field vector $\mathbf{H}$ and wave vector $\boldsymbol{\kappa}$ satisfy the left-hand relationship. Therefore, negative refractive index materials are also called left-handed materials. Poynting vector $\mathbf{s} = \mathbf{E} \times \mathbf{H}$, wave vector $\boldsymbol{\kappa}$ represents the direction of electromagnetic wave propagation (direction of velocity), and Poynting vector $\mathbf{s}$ represents the direction of energy flow. Obviously, in negative refractive index materials, the propagation direction of electromagnetic wave is opposite to that of energy flow. $\boldsymbol{\kappa} = -\sqrt{\varepsilon_r \mu_r} \omega / c$ can be obtained, so there is a refractive index $n = -\sqrt{\varepsilon_r \mu_r}$, the refractive index is negative, which is also the source of the name of negative refractive index material.

The propagation of light in negative refractive index materials will show many strange phenomena. The typical phenomena are as follows.

- (1) *Negative refraction.* When light is incident from a normal positive refractive index material into a negative refractive index

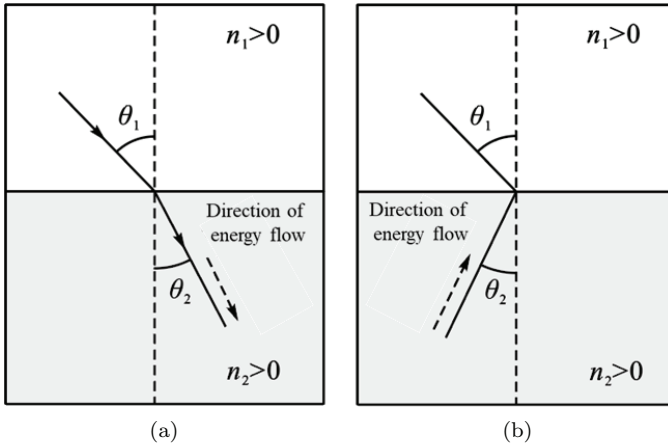


Fig. 3.6. Refractive index in materials: (a) refraction in ordinary materials; (b) refraction in negative refractive index materials.

material, or vice versa, the light is located on the same side of the interface normal, as shown in Fig. 3.6.

- (2) *Inverse Doppler effect.* If there is a relative motion of a light source and detector in the negative refractive index material, when the detector is relatively close to the light source, the measured light frequency decreases and the red shift occurs. When the detector is relatively far away from the light source, the measured light frequency increases and the blue shift occurs. This is contrary to the usual Doppler effect.
- (3) *Abnormal Cherenkov radiation.* It is found that when charged particles move faster than the speed of light in the medium, they generate electromagnetic radiation, which is Cherenkov radiation. It is impossible for superluminal phenomena to occur in vacuum. However, in the medium, because the speed of light is $v = c/n$, the velocity of charged particles can exceed the speed of light. The equal phase plane of Cherenkov radiation is a cone. The angle θ between the radiated energy flow and the direction of motion of charged particles satisfies $\cos \theta = c/nv$. For positive refractive materials, the angle θ is acute, and the radiation of light wave is forward positive radiation. In negative refractive index materials, the angle θ is passive, and the energy flow of

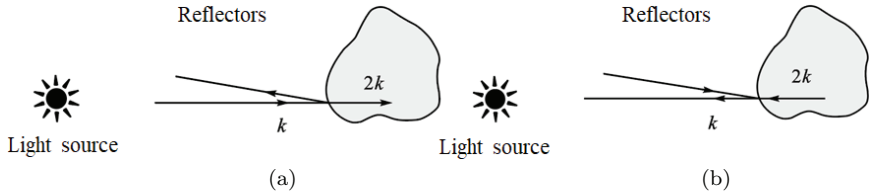


Fig. 3.7. Optical pressure in left-handed and right-handed Materials: (a) right-handed medium; (b) left-handed medium.

radiated light wave is backward to the direction of motion of charged particles.

- (4) *Abnormal optical pressure.* Any light wave can be regarded as a photon flow composed of a group of photons propagating in the form of discrete energy. The photon carrying momentum with a wave vector of κ is $p = \hbar\kappa$. As shown in Fig. 3.7, the impulse transmitted to object $2\hbar\kappa$ is reflected when a beam of light is incident vertically on the surface of an object in a positive refractive index material. It is easy to deduce that the light intensity I is incident vertically to the surface of the object, which produces a repulsive force of pressure $2I/c$ on the object. If, in negative refractive index materials, because the wave vector κ is opposite to the direction of energy flow, photons will transmit $-2\hbar\kappa$ from incident to reflection, and the optical pressure will change from repulsion to attraction.

3.3.2.2. Typical structure of negative refractive index materials

Negative refractive index materials have not been found in nature. All negative refractive index materials are manually prepared in the laboratory. The periodically arranged metal wires can simulate the response characteristics of plasma to electromagnetic waves, and produce negative dielectric constant in a certain frequency band. The periodic array composed of metal open-loop resonators can produce negative permeability. Negative refractive index materials can be prepared by combining the structures of two artificial materials, as shown in Fig. 3.8. This material is mainly used in microwave and terahertz bands.

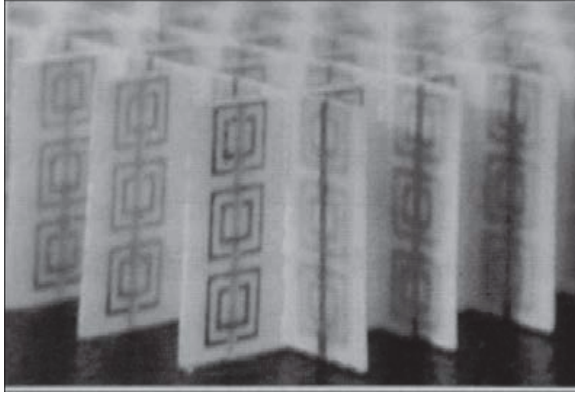


Fig. 3.8. Typical structures of negative refractive index materials (consisting of periodic arrangement of metal wires and metal gap rings).

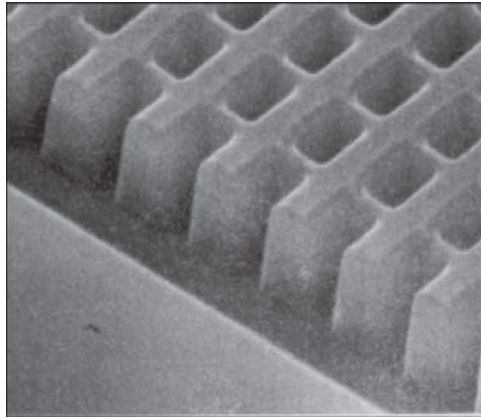


Fig. 3.9. Negative refractive index material of fishing net structure.

It is difficult to realize the negative refractive index in optical frequency band. In 2008, Valetine *et al.* alternately deposited silver and magnesium fluoride layers on a substrate and etched nanoscale fishing net structure. The material has negative refractive index in near infrared band. See Fig. 3.9. In addition, although photonic crystals have no negative dielectric constant and permeability, they can also exhibit negative refractive index material properties in a certain frequency band, resulting in negative refractive phenomena.

3.3.2.3. Negative refractive index device — superlens

Due to the limitation of diffraction limit, the resolution of lenses and conventional optical imaging devices is limited to the wavelength order of light. Negative refractive index materials could amplify evanescent waves, which makes it possible for negative refractive index materials to break the diffraction limit and achieve ultra-high resolution of sub-wavelength magnitude. The device made of negative refractive index material can break the diffraction limit. It is called superlens or perfect lens.

Negative refraction of light or electromagnetic wave occurs on the surface of ordinary positive and negative refractive index materials, i.e. the incident light and refractive light are located on the same side of the normal line of the interface. This makes it possible to focus light by using negative refractive index materials parallel plates without using curved surfaces. As shown in Fig. 3.10, the parallel plates of negative refractive index materials can play the role of ordinary lenses. In 2000, Pendry proved theoretically that the parallel plate structure composed of negative refractive index could break through the diffraction limit and is used for sub-wavelength imaging. In 2004, Grbic and Eleftheriades firstly achieved a superlens with a diffraction limit of $1/3$ in the microwave band. Due to the immaturity of negative refractive index materials in optical band, it is still

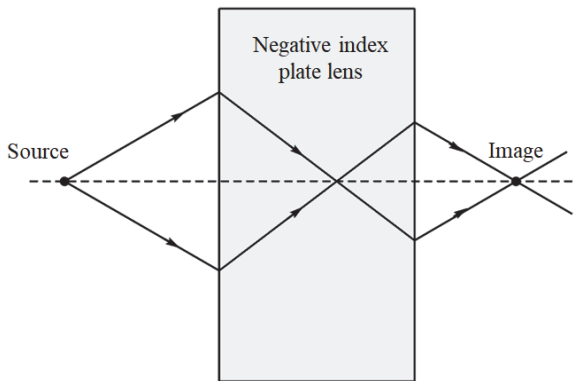


Fig. 3.10. Principle of negative refractive index plate lens.

difficult to fabricate optical band superlenses based on negative refractive index materials. However, with the maturity of material processing technology, it is believed that there will be breakthroughs in this field.

3.3.3. *Stealth cloak*

Stealth and invisibility often appear in myths, legends, science fiction and movies, which is a long-standing dream of mankind. The so-called stealth, from the optical point of view, means that light does not reflect, refract and absorb on the object, thus does not change the state of light transmission, making the object invisible. Stealth cloak, with the aid of the principle of transformation optics, can reasonably design the optical or electromagnetic parameters of metamaterial, which can make light or electromagnetic wave pass around the object like fluid without changing the propagation state, thus achieving the purpose of “stealth” of the object.

In 2006, Schurig *et al.* successfully fabricated a two-dimensional electromagnetic stealth cloak at microwave frequency band for the first time. Its structure is shown in Fig. 3.11. It uses 10-layer periodic SRR structure. By carefully designing the structural dimensions of SRR units on each layer, the effective dielectric constant and permeability of the material can meet the requirements of

$$\varepsilon_r = \left(\frac{b}{b-a} \right)^2, \quad \mu_r = \left(\frac{r-a}{r} \right)^2, \quad \mu_0 = 1. \quad (3.5)$$

where a and b are inner and outer diameters.

Under the geometrical optical limit conditions, light travels along the same path as the electromagnetic parameters given by the principle of transformation optics. Figures 3.11(b) and 3.11(c), respectively, give the transient electric field distribution measured in theory and experiment by incident 12GHz electromagnetic wave. It shows that the structure has successfully maintained the propagation state of electromagnetic wave, thus it can hide the function of objects inside the structure and realize electromagnetic stealth. Although the reflectivity of the structure is not zero, it successfully verifies

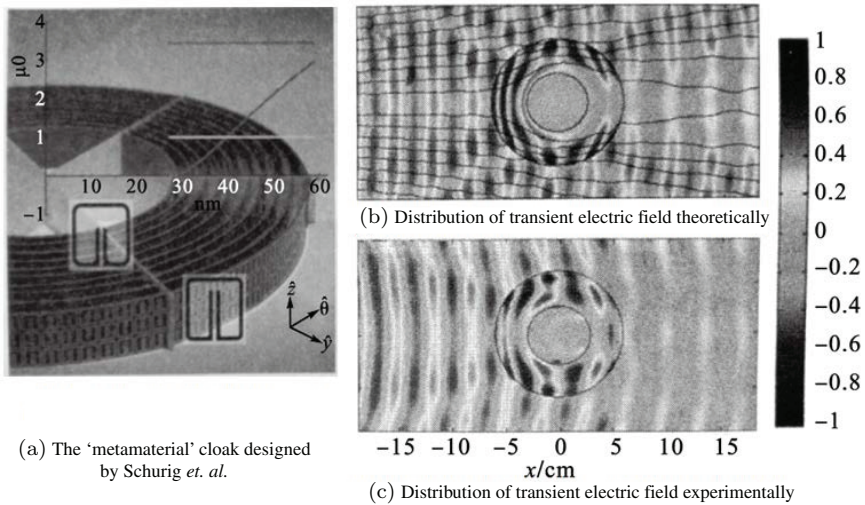


Fig. 3.11. Structure and electric field distribution of stealth cloak.

the theory of transform optics and the feasibility of electromagnetic stealth.

Due to the limitation of materials, it is relatively difficult to realize the invisibility cloak in visible band. However, in June 2011, researchers at the University of California, Berkeley, successfully prepared the first visible band stealth cloak, making the 300-nm-tall, 6- μm -wide object under the cloak "disappear". The structure is shown in Fig. 3.12. It is realized by fabricating silicon nitride waveguides on nanoporous SiO_2 . Among them, the thickness of silicon nitride layer is 300 nm. There is a small bump on the surface of SiO_2 . Its shape satisfies the function $y = h \cos^2(\pi x/\omega)$, where $h = 300 \text{ nm}$ and $\omega = 6 \mu\text{m}$. According to the design method of QCM (quasi-conformal mapping), the refractive index of the convex part changes according to the specific law, the refractive index of the center is the lowest and the refractive index of the bottom is the highest. Periodic air hole arrays with a period of 130 nm in hexagonal lattice were fabricated on silicon nitride. The required refractive index distribution was achieved by adjusting the size of the air hole.

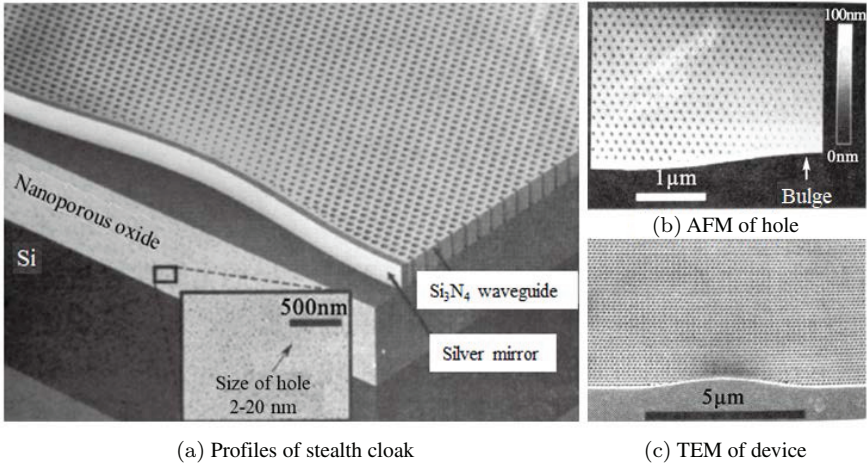


Fig. 3.12. Visible band invisibility cloak.

Figures 3.13(a) and 3.13(b), respectively, give the simulation results of the instantaneous distribution of the optical frequency electromagnetic field in the presence and absence of the invisibility cloak. Figure 3.13(a) shows that the surface bulge has a strong scattering effect on the incident light. Figure 3.13(b) shows that under the presence of the invisible cloak, the incident light wave propagates according to the mode of light propagation in the absence of the bulge, and thus achieves invisibility. The purpose of the body. Figure 3.13(c) gives the light intensity distribution under 480 nm, 530 nm and 700 nm illumination plane structure, bulge structure and bulge structure with invisibility cloak. The convex structure significantly changes the light intensity distribution, and the optical cavity distribution with the invisibility cloak is the same or very close to that without the convex. Therefore, the experimental results show the effectiveness of the invisibility cloak.

Although the research of stealth cloak is still in the experimental stage, the practical application of stealth cloak is still far away from people. However, there is no interruption in the study of invisibility cloak. With the efforts of researchers, or one day, the invisibility cloak in human fantasy can enter people's lives.

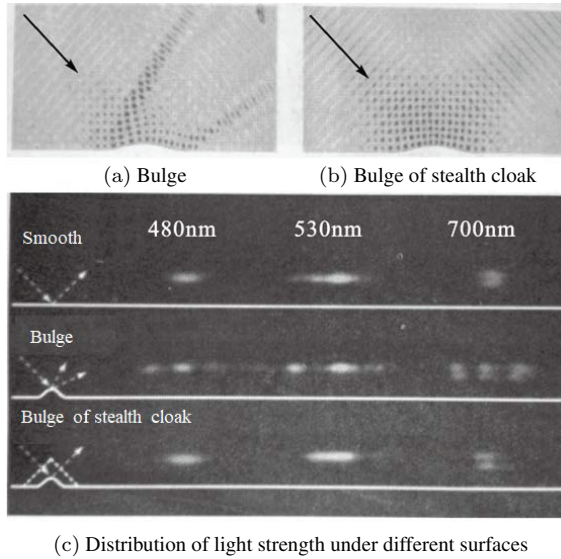


Fig. 3.13. Theoretical simulation results of instantaneous distribution and intensity distribution of optical-frequency electromagnetic field of electromagnetic stealth cloak.

3.4. Surface Plasmon Polaritons and Devices

3.4.1. Basic principles and properties

Surface plasmon polariton (SPP) is a mode of electromagnetic wave propagation caused by the interaction of free electrons near the surface of light and metal, and a mixed excitation state formed by the interaction of photons and electrons localized on the metal surface. Surface plasmon polaritons can propagate along the interface of metals and dielectrics, and decay exponentially rapidly in the direction perpendicular to the interface, which makes light confined in a spatial dimension far smaller than the free space wavelength of light waves (Fig. 3.14). Therefore, surface plasmon polaritons can be used to control the behavior of light at sub-wavelength scale, thus meeting some specific application needs.

For polarization wave, according to the boundary conditions, the electric field must be continuous at the interface of metal and

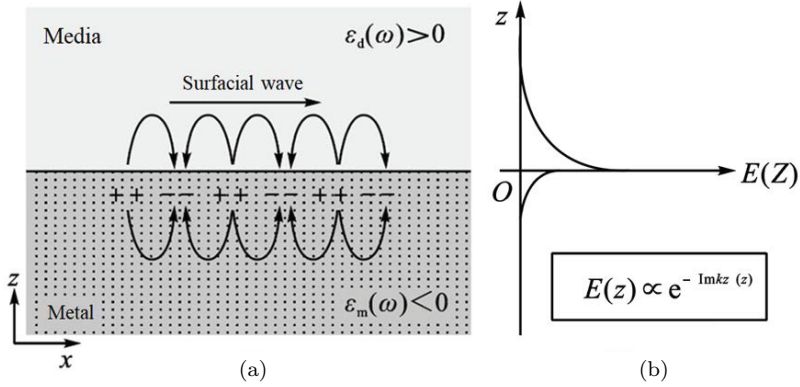


Fig. 3.14. Surface plasmon structure and electric field distribution. (a) Coupling of surface plasmon electromagnetic field with electrons. (b) Electric field amplitude distribution perpendicular to the interface direction.

medium, so the induced charge cannot be formed near the surface, and the surface plasmon cannot be formed without the induced charge. Surface plasmon polaritons are for TM (Transverse Magnetic) polarization waves.

The electromagnetic wave propagating in the XZ -plane can be expressed as

$$E = E_0 e^{i(k_r x + k_z z - \omega t)}. \quad (3.6)$$

The dielectric constant of the medium is set as d , and the dielectric constant of the metal is set as m when the loss is ignored. According to Drude's model, there are $\varepsilon_m = 1 - \frac{\omega_p^2}{\omega^2}$. According to the boundary conditions of electromagnetic field, the Max equation is solved. For plasma polaritons propagating on metal and dielectric interfaces, there are

$$\frac{k_{dz}}{\varepsilon_d} + \frac{k_{mz}}{\varepsilon_m} = 0, \quad (3.7)$$

$$k_r^2 + k_{dz}^2 = \varepsilon_d \left(\frac{\omega}{c} \right)^2, \quad (3.8)$$

$$k_r^2 + k_{mz}^2 = \varepsilon_m \left(\frac{\omega}{c} \right)^2. \quad (3.9)$$

The expressions of wave vector components of surface plasmon electromagnetic field in dielectrics and metals can be obtained, i.e.

$$k_x = \frac{\omega}{c} \left(\frac{\varepsilon_m \varepsilon_d}{\varepsilon_m + \varepsilon_d} \right)^{1/2} = \frac{\omega}{c} \left[\frac{\left(1 - \frac{\omega_p^2}{\omega^2}\right) \varepsilon_d}{1 + \varepsilon_d - \frac{\omega_p^2}{\omega^2}} \right]^{1/2}, \quad (3.10)$$

$$k_{dz} = \frac{\omega}{c} \left[\frac{\varepsilon_d^2}{1 + \varepsilon_d - \frac{\omega_p^2}{\omega^2}} \right]^{1/2}, \quad (3.11)$$

$$k_{mz} = \frac{\omega}{c} \left[\frac{\left(1 - \frac{\omega_p^2}{\omega^2}\right)^2}{1 + \varepsilon_d - \frac{\omega_p^2}{\omega^2}} \right]^{1/2}, \quad (3.12)$$

On the premise that k_x is a real number, it can be seen from expressions (3.11) and (3.12) that surface plasmon polaritons can be divided into two types: radiative and non-radiative. When the excitation frequency of surface plasmon is less than $\omega_p/\sqrt{2}$ (if the medium is air, then the frequency is $\omega_p/\sqrt{1+\omega_d}$), k_{dz} and k_{mz} are imaginary numbers. The electromagnetic field of surface plasmon generated by this method propagates along the interface and decays exponentially in the direction perpendicular to the interface. Currently, the plasma polaron is non-radiative. When the frequency of electromagnetic wave of surface plasmon is greater than ω_p , k_{dz} and k_{mz} are also real numbers. At this time, electromagnetic wave will radiate to the space outside the dielectric interface, so it is radiative. This book mainly considers non-radiative surface plasmon polaritons. The dispersion curve of surface plasmon is shown in Fig. 3.15.

Surface plasmon polaritons have important application prospects in many fields, such as optical integration, optical storage, optical sensing, super-resolution imaging and so on.

3.4.2. Surface plasmon optical waveguide

Because the electromagnetic field of surface plasmon is confined near the interface of metal and dielectric and propagates along

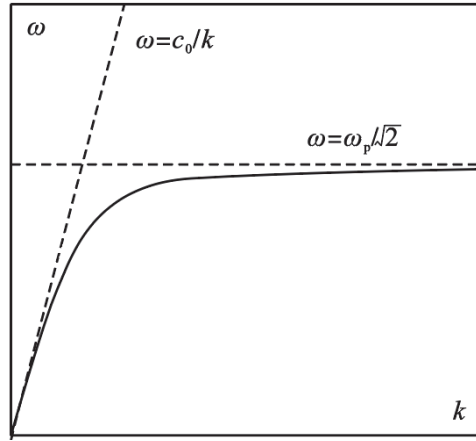


Fig. 3.15. Dispersion relation curve of surface plasmon.

the interface, this metal dielectric interface structure can be used as optical waveguide, that is, surface plasmon optical waveguide. Surface plasmon optical waveguides are only confined in a two-dimensional plane and can propagate freely in any direction at the interface between metal and dielectrics, so SPWs are only equivalent to two-dimensional waveguide structures. In order to further control the propagation behavior of surface plasmon, it is necessary to attach certain restrictions to make the optical-frequency electromagnetic field of surface plasmon propagate in a specific direction.

In order to restrict the propagation of electromagnetic wave in a direction in the plane, another one-dimensional restriction is needed. For example, a dielectric loaded surface plasmon optical waveguide has a structure as shown in Fig. 3.16(a). A dielectric strip is introduced into the metal air plane structure, so that the surface plasmon is confined to transport in the medium because the refractive index of the dielectric strip is higher than that of the air. The grooves of certain shape can be etched on the metal surface, and the propagation of surface plasmon can be restrained. The interface structures of plasma optical waveguides on the metal rectangular groove surface and on the metal V-groove surface are given in Figs. 3.16(b) and 3.16(c), respectively.

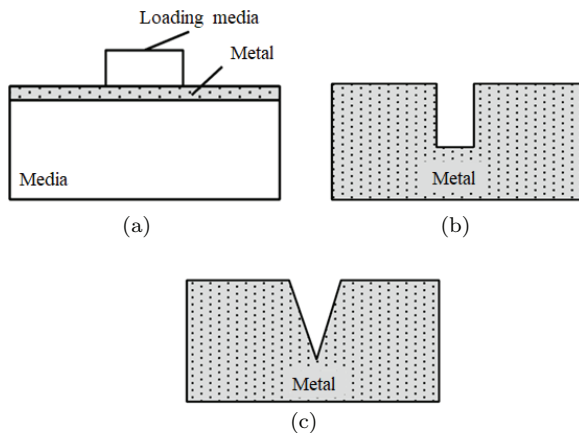


Fig. 3.16. Different types of surface plasmon optical waveguides: (a) dielectric-loaded; (b) metal rectangular groove optical waveguide; and (c) metal V-groove optical waveguide.

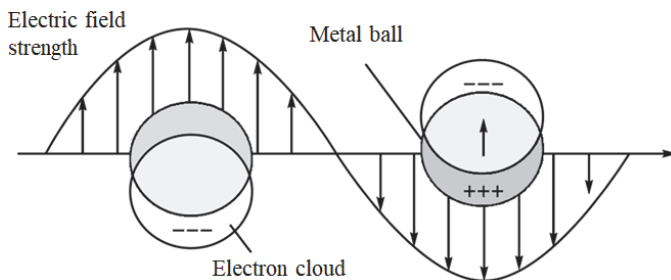


Fig. 3.17. Propagation of surface plasmon in nanospherical chains.

Surface plasmon optical waveguides can also be formed by aligning metal nanospheres into linear arrays (chains), as shown in Fig. 3.17. Under the irradiation of external light, the electrons in metal nanospheres move in groups, the electron density is rearranged, the electrons in metal nanospheres interact with the electrons in the optical frequency electromagnetic field to form surface plasmons, and the surface plasmons between adjacent nanospheres will be coupled with each other, so the propagation of optical frequency electromagnetic wave along the metal nanosphere chain can be realized. In 1998,

Quinten *et al.* reported for the first time the coupling of surface plasmon polaritons between two nanospheres. In 2003, Maier *et al.* observed the transport of surface plasmons along silver nanosphere chains for the first time. These results indicate the feasibility of using metal nanosphere chains as surface plasmon optical waveguides to transmit optical frequency electromagnetic fields.

Surface plasmon optical waveguide is the most basic passive device for many surface plasmon optical integrated devices.

3.4.3. Surface plasmon resonance sensor

Surface plasmon polaritons (SPP) can be used for high sensitivity sensing because of their local electric field enhancement effect, especially in biological and chemical sensing. One of the most typical is surface plasmon resonance (SPR) sensing technology. SPR sensors have been developed and applied in many scientific fields. Since 1982, Nylander *et al.* first used SPR technology in the field of immunosensor, surface plasmon biosensors have been deeply studied and widely used, and have become the main means of studying the interaction of biological molecules.

Evanescent waves generated by total internal reflection of light on the inner surface of dielectric materials such as glass are used to excite plasma polaritons on the surface of metal films. Electromagnetic waves are localized on the surface of metal films and propagate along the surface of metal films. When the incident angle and wavelength satisfy certain conditions, the surface plasmon polaritons have the same frequency and wave number as evanescent waves, which results in resonance. The incident light is strongly absorbed, resulting in a sharp decrease in the energy of the reflected light. A resonance peak appears in the reflection spectrum, i.e. the lowest intensity of the reflected light, as shown in Fig. 3.18(a). The position of the resonance peak is a function of the incident wavelength and the incident angle. The angle corresponding to the resonance peak is called the resonance angle, and the wavelength corresponding to the resonance peak is called the resonance wavelength. The position of the resonance peak is also affected by the refractive index of the surface medium of the

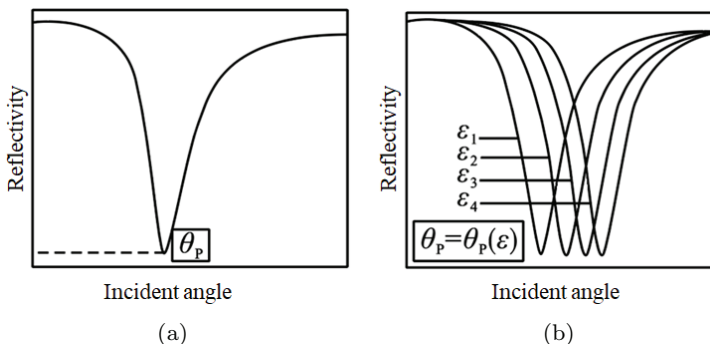


Fig. 3.18. Surface plasmon resonance sensor. (a) The relationship between incident angle and reflectivity (the lowest reflectivity at resonance position). (b) The relationship between incident angle and reflectivity at different dielectric constants of samples.

metal film and is extremely sensitive to the change of the refractive index near the surface, as shown in Fig. 3.18(b). Therefore, this effect can be used to realize the detection of biological and chemical sensors and to construct an exceptionally sensitive surface plasmon resonance sensor.

SPP is generated by excitation, and SPR is the key of surface plasmon sensing technology. In the late 1960s, Kretschman and Otto used prism coupling method to realize SPR of light wave, which laid a foundation for the wide application of SPR technology. Their method is simple and ingenious, and is still the most widely used method in SPR sensing. The basic structures of Kretschman prism coupling method and Otto prism coupling method are given in Figs. 3.19(a) and 3.19(b), respectively. For the former, the metal film is evaporated at the bottom of the prism, illuminated on the prism, evanescent waves generated by total internal reflection penetrate the metal film at the bottom of the prism, and surface plasmon polaritons are formed on the outer layer of the metal film. For the latter, the metal film is located very close to the bottom of the prism. The evanescent wave generated by total internal reflection interacts with the plasma on the surface of the metal film and forms surface plasmons on the surface of the metal film.

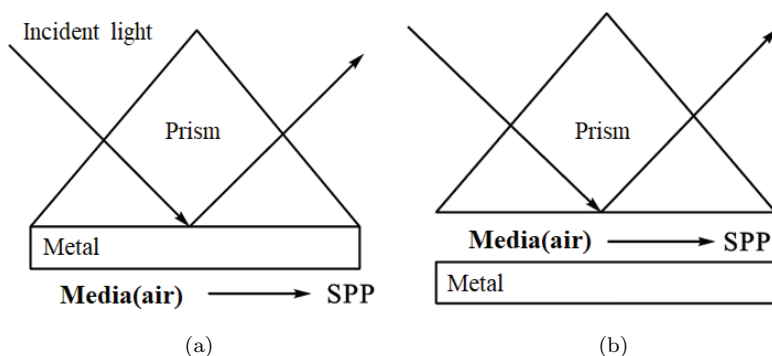


Fig. 3.19. Principle of resonance by prism coupling: (a) Kretschman prism coupling method; (b) Otto prism coupling method.

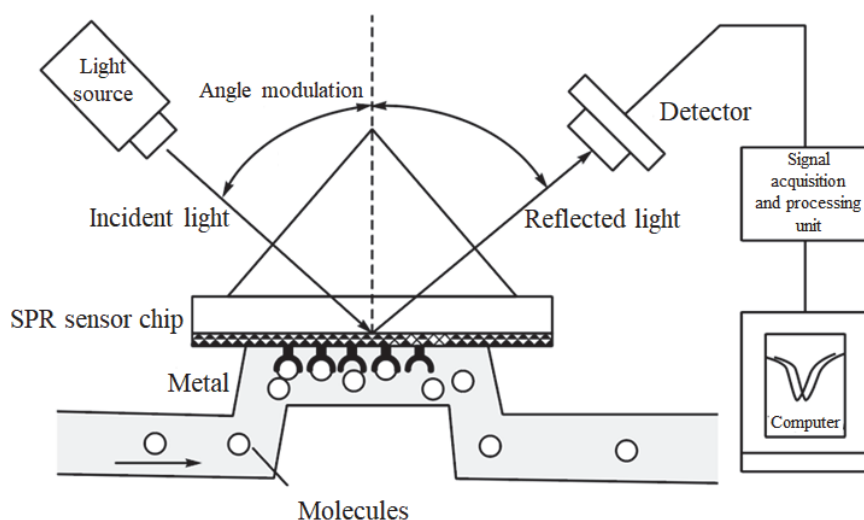


Fig. 3.20. Principle of typical surface plasmon resonance sensor.

Figure 3.20 gives the structure of a typical surface plasmon resonance sensor. A set of surface plasmon resonance sensor generally includes four parts: optical system, sensing element, data acquisition and processing unit.

The optical system includes light source, optical coupler, angle adjusting component and photodetector, which is used to generate

surface plasmon resonance (SPR) and detect the change of SPR spectrum. The light source can choose either the multi-color light source or the monochrome light source, such as He-Ne laser, according to the actual needs.

Sensitive elements are metal thin-films and surface-modified sensitive substances, which are used to convert the chemical and biological information of the subject to be measured into changes in refractive index. SPR sensors measure the specific properties of samples mainly depending on the change of SPR spectrum caused by the change of refractive index of samples. Without surface modification, the SPR sensor can only perform some simple measurements. Through specific surface modification, SPR sensor can obtain selective recognition of the object under test, so as to meet specific needs.

The data acquisition and processing unit is used to collect and post-process the information obtained by the photodetector in the optical system.

References

- Guo P., Liang L., 2005. *Fundamentals of Optoelectronic Technology*, Beijing: Beijing University of Aeronautics and Astronautics Press.
- Gharghi M *et al.*, 2011. A carpet cloak for visible light, *Nano Lett.* 11(7):2825–2828.
- Grbic A, Eleftheriades G V, 2004. Overcoming the diffraction limit with a planar left-handed transmission-line lens, *Phys. Rev. Lett.* 92(11):117403.
- Joannopoulos D, Meade R D, Winn J N, 1995. *Photonic Crystals*, Princeton: Princeton University Press.
- John S, 1987. Strong localization of photons in certain disordered dielectricsuperlattices, *Phys. Rev. Lett.* 58(23):2486–2489.
- Knight J *et al.*, 1996. All-silica single-mode optical fiber with photonic crystal cladding, *Opt. Lett.* 21(19):1547–1549.
- Kosaka H *et al.*, 1999. Self-collimating phenomena in photonic crystals. *Appl. Phys. Lett.* 74:1212–1214.
- Kosaka H *et al.*, 1999. Superprism phenomena in photonic crystals: Toward microscale lightwave circuits, *J. Lightwave Technol.* 17(11):2032–2038.
- Krauss T, De La Rue R, Brand S, 1996. Two-dimensional photonic-bandgap structures operating at near-infrared wavelengths, *Nature* 383(6602):699–702.
- Luo C *et al.*, 2002. All-angle negative refraction without negative effective index, *Phys. Rev. B* 65(20):201104.

- Maier A *et al.*, 2003. Local detection of electromagnetic energy transport below the diffraction limit in metal nano particle plasmon waveguides, *Nat. Mater.* 2(4):229–232.
- Matsubara H *et al.*, 2008. GaN photonic-crystal surface-emitting laser at blue-violet wavelengths, *Science* 319(5862):445–447.
- Monat C *et al.*, 2003. InP based photonic crystal microlasers on silicon wafer, *Physica E* 17:475–476.
- Notomi M, 2000. Theory of light propagation in strongly modulated photonic crystals: Refractionlike behavior in the vicinity of the photonic bandgap, *Phys. Rev. B* 62(621):10696–10705.
- Painter O *et al.*, 1999. Two-dimensional photonic bandgap defect mode laser. *Science* 284(5421):1819–1821.
- Pendry J B, 2000. Negative refraction makes a perfect lens, *Phys. Rev. Lett.* 85(18):3966–3969.
- Quinten K *et al.*, 1998. Electromagnetic energy transport via linear chains of silver nanoparticles, *Opt. Lett.* 23(17):1331–1333.
- Scharrer M *et al.*, 2006. Ultraviolet lasing in high-order bands of three-dimensional ZnO photonic crystals, *Appl. Phys. Lett.* 88(20):201103–201105.
- Schurig D *et al.*, 2006. Metamaterial electromagnetic cloak at microwave frequencies, *Science* 314(5801):977–980.
- Shelby R A, Smith D R, Shultz S, 2001. Experimental verification of a negative index of refraction, *Science* 292(5514):77–79.
- Tong L M *et al.*, 2003. Subwavelength-diameter silica wires for low loss optical wave guiding, *Nature* 426(6968):816–819.
- Valentine J *et al.*, 2008. Three-dimensional optical metamaterial with a negative refractive index, *Nature* 455(7211):376–379.
- Veselago G, 1968. The electrodynamics of substances with simultaneously negative values of ϵ and μ , *Sov. Phys. Usp.* 10(4):509–514.
- Wierer J J, David A, Megens M M, 2009. III-nitride photonic-crystal light-emitting diodes with high extraction efficiency, *Nat. Photonics* 3(3):163–169.
- Yablonovitch E, 1987. Inhibited spontaneous emission in solid-state physics and electronics, *Phys. Rev. Lett.* 58:2059–2062.

This page intentionally left blank

Chapter 4

Semiconductor Luminescent Materials and Devices

Luminescence is a process in which energy is absorbed by an object in some way and then converted into light radiation. Light radiation can be divided into equilibrium radiation and non-equilibrium radiation according to its energy conversion process. Luminescence refers to the non-equilibrium radiation part of light radiation. Non-equilibrium radiation is a kind of external energy excitation, the object deviates from the original equilibrium state, if the object in the process of returning to the equilibrium state, its excess energy is emitted in the way of light radiation, it is called luminescence. Luminescent materials such as gallium arsenide, gallium phosphide and gallium phosphorus arsenide are the basis of light-emitting devices, and the improvement of device performance depends largely on the progress of materials. Semiconductor luminescent materials mainly consist of III–V compound semiconductors and ternary and quaternary solid solutions composed of them, such as GaAs, GaP, GaN, or ternary crystals $\text{GaAs}_{1-x}\text{P}_x$, $\text{Ga}_{1-x}\text{Al}_x\text{As}$ and quaternary crystals $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$, $\text{Al}_x\text{Ga}_{1-x}\text{As}_y\text{Sb}_{1-y}$, etc. Inorganic luminescent materials are mainly composed of crystalline materials as matrix and rare earth or transition metal elements doped in them. In order to better explain the luminescent phenomena of inorganic luminescent materials, it is very important to understand the crystal structure, energy band theory and related knowledge of inorganic materials.

4.1. Introduction to Semiconductor Luminescent Materials Crystals

4.1.1. *Crystal structure*

The forms of substances in nature can be roughly divided into three categories: solid, liquid and gas. Solids can also be divided into crystals and amorphous crystals. Crystals are solid materials consisting of atoms (ions or molecules) arranged periodically and repeatedly in space. This periodic structure is often called a lattice. The periodically repetitive elements are called structural elements. The selection of primitives is arbitrary. There are two practical methods: one is to select the smallest repetitive unit, even if the atoms in the cell are the least, which is called the primitive cell; the other is to select the smallest repetitive unit which can reflect the symmetry of the lattice to the greatest extent. Crystallographic cell is called unit cell. The actual length of each side of crystallographic cell is called lattice constant. Crystals can also be divided into single and polycrystals. Monocrystals are crystals with regular arrangement inside them. Polycrystals are not. Locally, their origins, ions or molecules are arranged regularly, but in general, they are irregular. Therefore, it can be said that polycrystals are composed of single crystals. Single crystals have relatively symmetrical shapes, such as Yila GaAs single crystals as the base of GaAsP epitaxy sheet, three symmetrical edges with 120° apart in $\langle 111 \rangle$ direction, and four symmetrical edges with 90° apart in $\langle 100 \rangle$ direction. This is also the regular arrangement of atomic structure in single crystals. Reflections of columns.

4.1.1.1. *Spatial lattice*

After a long period of research, Bravi's theory of spatial lattice was put forward in the 19th century. It is believed that the internal structure of crystals can be summarized as the periodic and infinite distribution of some identical points in space. These points represent the center of gravity of atoms, ions, molecules or groups. These points are collectively called lattices, and their structures are called spatial

lattices. The positions of these points are called nodes, so each node in the lattice corresponds to a certain element of a structure.

Spatial lattice is the periodic arrangement of repetitive elements (structural elements) of crystals in three-dimensional space. The repetitive elements of these periodic arrangements can be abstracted into some identical points arranged in space with the same periodicity. The array of these points is called lattice, or spatial lattice. The whole crystal can be represented by a three-dimensional lattice. In a crystal space lattice, each lattice has exactly the same surroundings; under the symmetric operation of translation (connecting vectors of any two points in the lattice, translating according to this vectors), all points can be restored; each point represents the same position in the structure; its corresponding specific contents include the kinds and quantities of atoms or molecules and their arrangement in space in a certain way. Structures are called structural units of crystals, or basis for short. Primitives refer to the specific content in the repetition period; lattice points are abstract points representing the spatial repetition arrangement of structural primitives. If the structural elements are placed in the same way at the positions of each lattice in the crystal lattice, the structure of the whole crystal can be obtained.

Through the nodes in the lattice, many parallel linear families and parallel crystal plane families can be made, so that lattices form lattices. Because of the periodicity of the lattice, a parallel hexahedron with the node as the vertex and the edge length equal to the period in the direction can be taken as the repetitive unit, which is called the cell, to characterize the characteristics of the lattice. The three prisms of the parallel hexahedron are called crystal axes, which are expressed by x -, y - and z -axes. Let the unit cell vectors in these axes be expressed by $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$, and the angles between the three axes are expressed by α , β and γ , respectively. a , b and c are the lengths of unit cell vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$, which are called lattice constants. a , b , c , α , β and γ are the six parameters of the prototype, as shown in Fig. 4.1.

The crystal analysis shows that the crystal can be divided into seven crystal systems with six parameters of cell. The seven

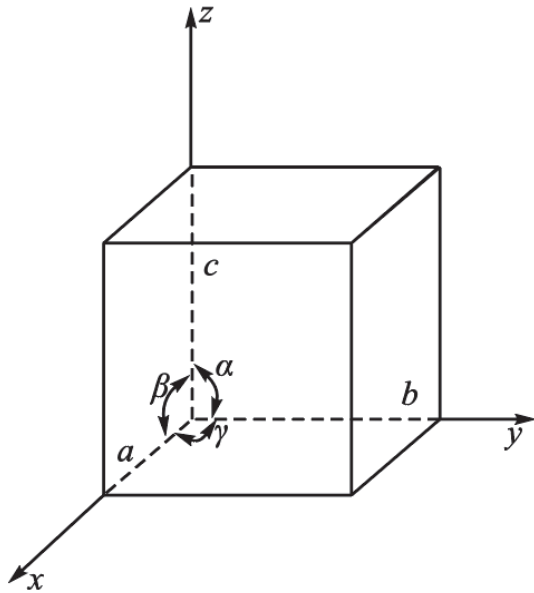


Fig. 4.1. Cell.

Table 4.1. Seven crystal system lattices.

Crystal system	Side length	Included angle	Examples of crystals
Cubic system	$a = b = c$	$\alpha = \beta = \gamma = 90^\circ$	NaCl
Trigonal system	$a = b = c$	$\alpha = \beta = \gamma \neq 90^\circ$	Al_2O_3
Tetragonal system	$a = b \neq c$	$\alpha = \beta = \gamma = 90^\circ$	SnO_2
Hexagonal system	$a = b \neq c$	$\alpha = \beta = 90^\circ, \gamma = 120^\circ$	AgI
Orthogonal crystal system	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90^\circ$	HgCl_2
Monoclinic system	$a \neq b \neq c$	$\alpha = \beta = 90^\circ, \gamma \neq 90^\circ$	KClO_3
Triclinic system	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma \neq 90^\circ$	$\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$

crystal systems include triclinic, monoclinic, orthogonal, tetragonal, triangular, hexagonal and cubic systems, as shown in Table 4.1. The three cell types of cubic system are shown in Fig. 4.2. The lattice structures of semiconductor materials Ge, Si, GaAs, GaP and GaAsP belong to the face-centered cubic structure of the cubic system.

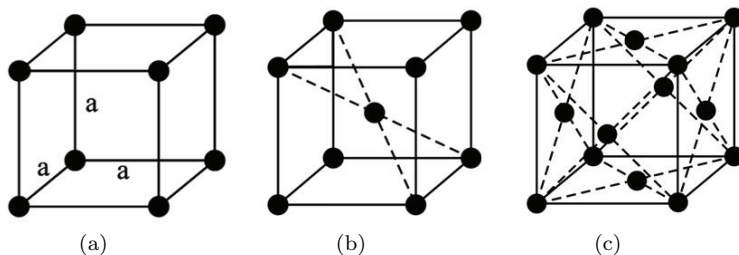


Fig. 4.2. Three crystal cells of cubic system: (a) simple cube; (b) body centric cube; and (c) face centric cube.

4.1.1.2. *Crystal plane and orientation*

A plane composed of several nodes in a lattice is called a lattice plane, and in crystals it is called a crystal plane. A crystal plane can usually be expressed by a crystal plane index, also known as Miller index. The determination method is as follows.

- (1) The crystal axis X , Y and Z are made for the cell, and the cell edge length is taken as the unit length of the crystal axis.
- (2) Calculate the intercept of the undetermined crystal plane on the three crystal axes (if the crystal plane is parallel to a certain axis, the intercept is ∞), such as 1, 1, ∞ , 1, 1, 1, 1, $1/2$, etc.
- (3) Take the reciprocal of these intercepts, such as 110, 111, 112, etc.
- (4) The reciprocal is reduced to the smallest simple integer and parentheses are added to denote the index of the crystal plane, which is generally denoted as (hkl) , such as (110), (111), (112), etc.

The crystal plane $a_1b_1c_1$ marked in Fig. 4.3 has corresponding intercepts of $1/2$, $1/3$ and $2/3$. Its reciprocal numbers are 2, 3 and $3/2$, which are reduced to simple integers of 4, 6 and 3. Therefore, the crystal plane index of $a_1b_1c_1$ is (463). Figure 4.4 shows the surface indices of some crystal planes in the crystal.

The crystal plane index should be explained as follows: h , k and l correspond to X -, Y - and Z -axes, respectively, and the order cannot be changed at will. If a certain number is 0, it means that the crystal plane is parallel to the coordinate axis corresponding to the number.

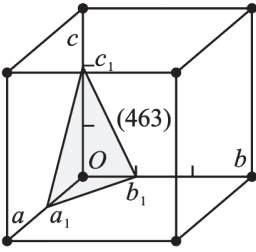


Fig. 4.3. Representation of crystal plane index.

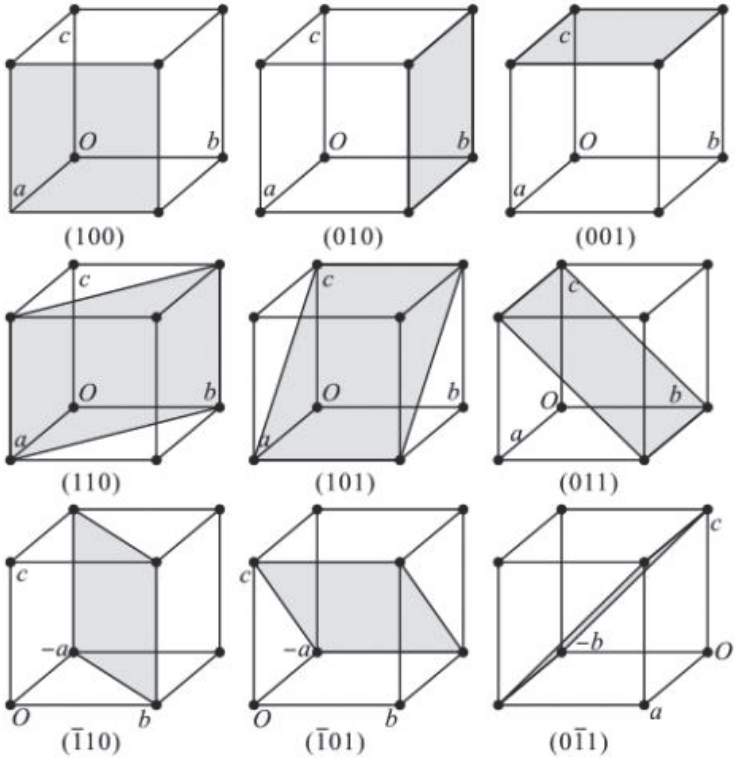


Fig. 4.4. Crystal plane index of several crystal planes.

For example, $(h0l)$ indicates that the crystal plane is parallel to the Y -axis. If the interception of an axis is negative, the corresponding index is labeled with “-”, such as $(h\bar{k}l)$, $(\bar{h}kl)$. In crystals, any crystal plane always recurs in a certain period, and its number can be infinite and parallel to each other, so it can be expressed by the same crystal plane index (hkl) . So (hkl) is not just a crystal plane, but a group of parallel planes. H , K and l represent the number of the crystal planes in the unit length range along the three coordinate axes, i.e. the linear density of the crystal planes.

Any straight line direction passing through many nodes in a crystal lattice is called crystal orientation. Its index is determined by the following methods.

- (1) Take a lattice of the cell as the origin, three basic vectors as the coordinate axis, and the length of the lattice basic vectors as the unit length of the three coordinates.
- (2) Make a straight line OP through the origin and make it parallel to the crystal orientation AB to be calibrated (Fig. 4.5). This straight line must pass through some lattices.
- (3) Choose the nearest point P from the origin O on the straight line OP , and determine the coordinate value of the point P .
- (4) The value is multiplied by the smallest common multiple to the smallest integers u , v and w , with square brackets, $[uvw]$ is the orientation index of AB orientation. If a number in u , v and

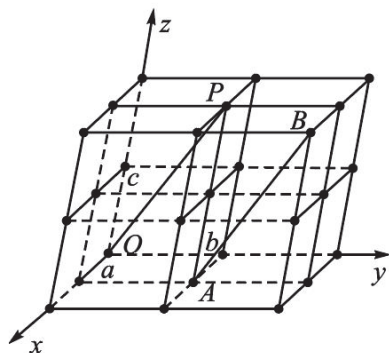


Fig. 4.5. Determination of crystal direction index.

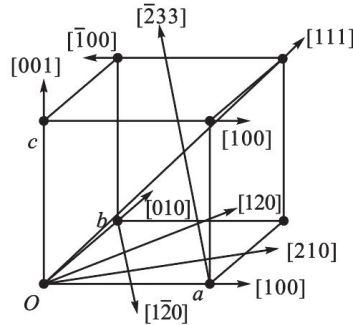


Fig. 4.6. Crystal direction index of several crystal directions in orthogonal lattice.

w is negative, the negative sign is marked above the number. Figure 4.6 gives the orientation indices of some orientations in orthogonal lattices.

Obviously, the orientation index represents a group of parallel and consistent orientations. If two straight lines in a crystal are parallel to each other but in opposite directions, the number of their orientation indices is the same and the sign is the opposite.

4.1.2. Defects and their effects on luminescence

The ideal crystal has a strict periodic structure, while the actual crystal is always incomplete. This incompleteness is called defect. The existence of defects can affect the solubility, diffusion rate and distribution of impurities in crystals. Therefore, defects are closely related to the luminescent properties of materials.

According to the geometric structure of the defect, it can be divided into the following four types.

- (1) Point defect. The positions deviating from the ideal lattice structure are confined to one atom or several atoms, such as lattice vacancies, impurity atoms, interstitial atoms, etc. Point defect can be divided into intrinsic point defect and non-intrinsic point defect. Non-intrinsic defects refer to defects caused by the introduction of impurity atoms. Eigenpoint defect is a point defect formed by the misalignment of the matrix atoms that

make up the crystal when there are no foreign impurities. For example, vacancies and interstitial atoms caused by thermal vibration fluctuations of lattice atoms caused by temperature rise are typical intrinsic point defects, and their number depends on temperature, also known as thermal defects. The classifications of intrinsic point defects are as follows: (1) Frank defects, vacancies and interstitial atoms appear in pairs, and interstitial atoms may return to vacancies and compound; (2) Schottky defects, only vacancies without interstitial atoms (the most probabilistic defect, the lowest energy to form vacancies), interstitial atoms are arranged in the lattice of normal atoms on the surface. The third kind of defect has only gap but no vacancy, and the atoms at the normal lattice point on the surface run into the lattice gap. Only two of the three defects are independent, so the third one is unnamed.

- (2) Line defects, such as dislocations.
- (3) Surface defect. Two-dimensional defects, such as grain boundaries, twin boundaries, phase boundaries, stacking faults at solid-solid interfaces, and the external surface of crystals.
- (4) Body defects, such as voids, second-phase inclusions, etc.

The localized energy level formed by the defect can become the radiation recombination center, that is, the luminescence center. Structural defect luminescence centers are formed by structural defects of the lattice itself, such as vacancies, interstitial atoms, etc. For example, the self-activated luminescence center of self-activated blue emission band in ZnS is the Zn^{2+} vacancy. The green luminescence of cathode-ray luminescence powder can be obtained without activator, and its luminescence center is excess Zn. Another kind of luminescence center is called impurity defect luminescence center, which is the association defect composed of activator ion or activator ion and other defects. For example, in the ZnS: Mn luminescent body, the luminescent center is the substitution ion Mn^{2+} , which produces orange emission light. The luminescence centers formed by rare earth ions in luminescent bodies also belong to this category.

Structural defects can also form non-radiative recombination centers. For example, when electrons are captured by non-radiative recombination centers formed by gallium vacancies in P-GaAs, the energy is not emitted as photons, but converted into heat, and the intensity of photoluminescence decreases with the increase of its concentration. The Ga vacancy, a non-radiative recombination center in N-GaP, also contributes to the green luminescence efficiency.

4.1.3. Energy band structure

Chapter 2 of this book has introduced the energy band theory in detail. This paper focuses on the application of the theory in semiconductor luminescent materials. All kinds of solid-state luminescence are the results of electron transitions in different energy states of solids. Therefore, the study of the energy states of electrons in solids is one of the bases for understanding the phenomenon of solid-state luminescence. Energy-level theory is a theory to explain the orbit of the electron outside the nucleus. It holds that electrons can only move in specific, discrete orbits, and that electrons in each orbit have discrete energy, which is energy levels. Electrons can transit between different orbits. The energy absorbed by electrons can transit from low level to high level or from high level to low level to radiate photons. The energy levels of hydrogen atoms are shown in Fig. 4.7.

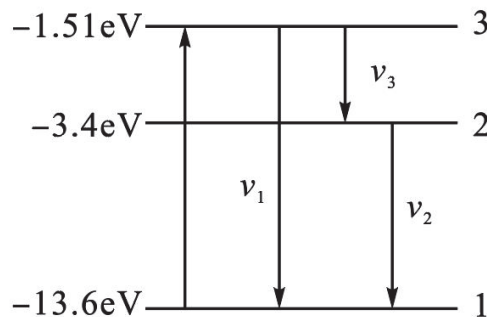


Fig. 4.7. Energy-level diagram of hydrogen atom.

When many atoms in the crystal approach, the electron orbits of the inner and outer layers overlap in different degrees. Of course, the outermost electron orbits of two adjacent atoms overlap most in the crystal. At this time, electrons are no longer confined to certain atoms, but can be transferred to adjacent atoms, so that electrons can move in the whole crystal, which is the commonalization of electrons in the crystal. This process can be qualitatively represented by Fig. 4.8. The dotted line represents the motion of electronic commonalization. Its expression can still be expressed by $E(\kappa)$ relation similar to that of free electrons. The relationship $E(\kappa)$ between the wave vector κ describing the electron energy and the electron commonalization motion in semiconductor crystals is called energy band structure. Of course, the complex interaction in crystals makes the $E(\kappa)$ relationship extremely complex. So the adiabatic approximation, static approximation and single-electron approximation are simplified to single-electron problem. Then the Schrodinger equation is solved by using modern calculation method and high-speed computer with appropriate potential function. The essence of determining the eigenwave function and the eigenenergy value $E(\kappa)$ is to consider only the motion of an electron in a fixed nuclear potential field and the average potential field of other electrons, and then the energy state of the electron becomes an energy band. Therefore, the theory of using single-electron approximation to

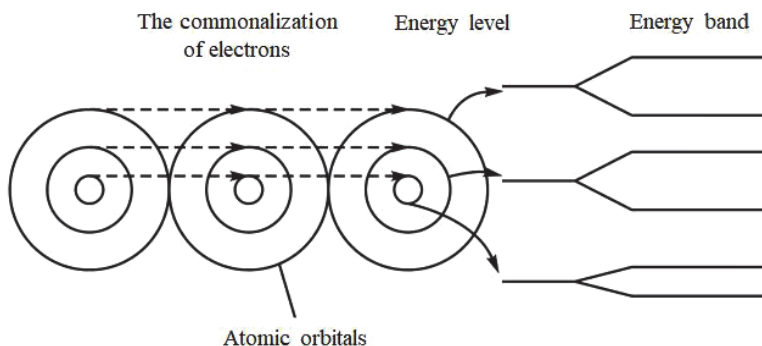


Fig. 4.8. The commonalization of electrons in crystals transforms electrical energy levels into energy bands.

study the energy state of electrons in crystals is also called energy band theory.

Energy band theory is an important approximation theory for discussing the states and motions of electrons in crystals (including crystals of metals, insulators and semiconductors). Among them, the absolutely empty minimum energy band is called conduction band, and the absolutely electron-occupied maximum energy band is called valence band. The energy gap between them is called forbidden band, as shown in Fig. 4.8. The energy band theory regards the motion of each electron in a crystal as an independent motion in an equivalent potential field, i.e. the theory of single-electron approximation. For valence electrons in a crystal, the equivalent potential field includes the potential field of atomic reality, the average potential field of other valence electrons and the exchange effect caused by the antisymmetry of electron wave function, which is a periodic potential field of a crystal. The electrical, optical and magnetic properties of crystals are all related to the motion of electrons. The energy band theory is used to study these problems. The band theory successfully explains the difference between metals, semiconductors and insulators and the Hall effect phenomenon. Semiconductor physics is based on energy band theory.

Energy band of semiconductor: after continuous growth of N-type and P-type layers of semiconductor on crystalline substrates. That is to say, PN junction is formed. When P-type and N-type layers are combined, energy band diagrams and Fermi levels without voltage and plus forward voltage are formed, as shown in Fig. 4.9. In Fig. 4.9,

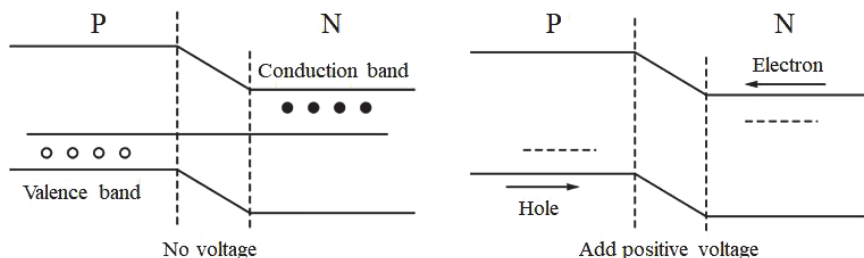


Fig. 4.9. Energy band diagram and Fermi level diagram of diode.

the vertical direction represents energy. For electrons, the higher the energy is, for holes, the higher the energy is. Therefore, in order for electrons to enter P-type field from N-type field (or holes to enter N-type field from P-type field), obstacles must be overcome first. The “height” of the barrier is equivalent to the starting voltage V_b (built in voltage) — the forward voltage drop of the diode.

The band structures of GaAs and GaP are shown in Fig. 4.10. The parabolic curves with peak values are the bands describing the outer valence electrons in crystals. They are called full bands because they are usually filled by electrons. The parabolic curves with peak values downward describe the energy states of the electrons involved in conduction after excitation, which are called conduction bands. Between the top of the valence band and the bottom of the conduction band is the forbidden band, whose width is called bandgap width or forbidden bandwidth, expressed in E_g and expressed in eV. Simplified band structures are often used in practical applications, as shown in Fig. 4.11.

Whether the bottom of the conduction band and the top of the valence band are at the same position in the Brillouin region has a great influence on the optical properties of the crystal. Direct bandgap semiconductors with the same K value are called direct bandgap semiconductors, and indirect bandgap semiconductors with

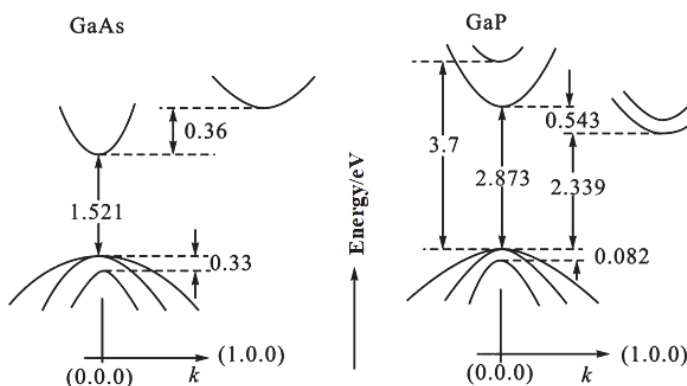


Fig. 4.10. Energy band structures of GaAs and GaP.

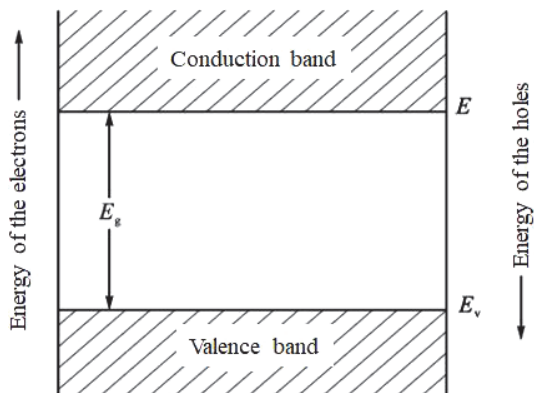


Fig. 4.11. Semiconductor simplified energy band structure.

different K values are called indirect bandgap semiconductors. The former is also called direct transition semiconductor, and the latter is also called indirect transition semiconductor. The latter must have lattice to participate in momentum exchange, that is, phonon to participate in the transition composite luminescence, which is a second-order process with low probability and low luminescence efficiency. GaAs band structure is a direct transition type, with the same K value at the bottom of conduction band and the top of valence band, both at the Γ point. The conduction band bottom and valence band top of GaP have different k values, X and Γ points, respectively. Table 4.2 gives the experimental and calculated values of band structure and bandgap width of semiconductor crystals.

Semiconductor crystals as luminescent materials are not impurity-free and defect-free intrinsic semiconductors. Some impurities are deliberately mixed in, and some are brought in by contamination. Impurities generate impurity levels in the bandgap. Usually, the impurity levels are shallow donor level and shallow acceptor level. When the impurity atoms replace the atoms in semiconductor crystals, the redundant valence electrons will be emitted as donor impurity or N -type impurity. The impurity atoms with fewer valence electrons than those without atoms will receive the electron as

Table 4.2. Experimental and calculated values of band structure and bandgap width of semiconductor crystals.

Crystal	Bandgap type	Experimental value of bandgap width (eV)		Calculated bandgap width (eV)	Melting point (°C)
		0 K	300 K		
C (diamond)	Indirect	5.48	5.47	5.48	1420
Si	Indirect	1.166	1.12	1.04	1420
Ge	Indirect	0.744	0.663	0.61	958
a-Sn	Direct	0.082	*	0.13	-150
Sic (6H)	Indirect	3.033	2.996	4.64	
BN	Indirect	—	57	9.57	
BP	Indirect	—	2	1.31	2000–5000
BA ₉	Indirect	—	—	0.85	
AlN	Indirect	—	5.9	8.85	>2400
AlP	Indirect	2.52	2.45	2.63	>1500
AlAs	Indirect	2.288	2.16	1.87	~1700
AlSb	Indirect	1.6	1.5	2.15	>1050
GaN	Direct	2.52	2.45	2.63	>1500
GaP	Indirect	2.333	2.261	2.75	1465
GaAa	Direct	1.521	1.435	1.53	1237
GaSb	Direct	0.813	0.72	1	712
InN	Direct	—	2	42	33
InP	Direct	1.421	1.351	1.45	1062
InAs	Direct	0.42	0.35	0.84	942
InSb	Direct	0.228	0.18	0.39	625

*Room temperature instability.

acceptor impurity or P-type impurity. In the semiconductor of group III–V compound, the acceptor is formed when group II atoms such as Zn replace group IV atoms, when group VI atoms such as S replace group V atoms, and when group IV atoms such as Si are amphoteric impurities. It forms the donor when replacing group III atoms and when replacing group V atoms, it forms the acceptor. In fact, the donor or acceptor of group IV atoms is determined by the conditions of crystal growth and heat treatment. The deep level near the center of the bandgap, the shallow donor level near the bottom of the conduction band and the shallow acceptor level near the top of the valence band.

Impurities with the same arrangement of outer electrons as atoms in a crystal, such as the impurity atom N in a GaP crystal, cannot form impurity levels by its charge. However, the potential field around the impurity atom is quite different from that of the crystal itself, and the energy levels of trapping electrons or holes are often formed by the action of short-range potential field. This energy level is called isoelectronic trap. The light atoms in the same column on the periodic table of elements have high electronegativity. When they replace heavy atoms, they can form electron traps. If heavy atoms replace light atoms, they can form hole traps. When the isoelectronic trap captures the carriers, it can further capture the carriers with opposite charges, which forms the bound excitons bound by the isoelectronic trap. The electrons and holes in the bound excitons can be directly transited and recombined to produce high efficiency luminescence. This is very important for indirect transition semiconductor luminescent materials. For example, GaP has a very low luminescent efficiency. After adding isoelectronic trap impurities N and Zn–O pairs, a large number of green and red luminescent devices are made.

4.1.4. *Conditions of semiconductor luminescent materials*

Not all semiconductor materials can emit light. Semiconductor materials are divided into direct bandgap materials and indirect bandgap materials. Only direct bandgap materials can emit light. Direct bandgap materials, i.e. electrons can transit vertically from the bottom of the conduction band to the top of the valence band. They have the same momentum in the conduction band and the valence band and have high luminescence efficiency. Indirect bandgap materials refer to the fact that electrons cannot transit vertically to the top of the valence band at the bottom of the conduction band. Their momentum is not equal in the conduction band and valence band. Therefore, another particle must be involved to make the momentum equal. The energy of this particle is E_p and the momentum is K_p . This indirect bandgap material is difficult to emit light, so its luminescence efficiency is very low, such as semiconductor

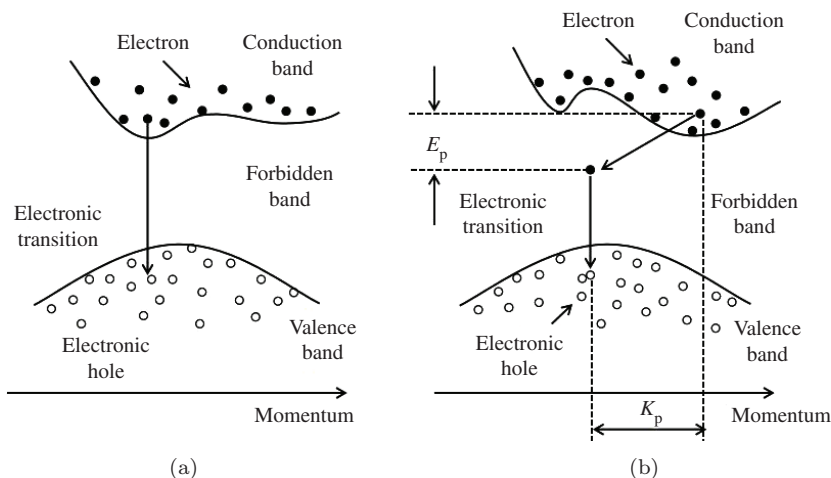


Fig. 4.12. Semiconductor (a) direct and (b) indirect bandgap material diagrams.

materials silicon (Si) and germanium (Ge). Figure 4.12 is a diagram of semiconductor direct and indirect bandgap materials. It illustrates the way in which semiconductor resistivity is generated, with various proprietary names attached. Among them, the carrier located in the conduction band is called the electron, the carrier located in the valence band is called the hole, and the part between the conduction band and the valence band is called the bandgap width (E_g). The resistivity of doped semiconductors is related to doping concentration and temperature, which indicates that electrons and holes are functions of impurity concentration and temperature. Electrons and holes are two important carriers related to electrical conduction. Electrons have negative charges and holes have positive charges. Figure 4.13 is a schematic diagram of the current flow direction and the electron and hole walking direction of the electrons and holes in semiconductor after adding voltage. As can be seen from the graph, the current flow direction is the direction of hole flow, while the electrons move in the opposite direction. In fact, electrons and holes exist in semiconductors at the same time, and their densities will increase sharply as the temperature rises. As long as the temperature is not changed, the densities will remain roughly the same. The more electrons, the fewer holes; and the more holes, the

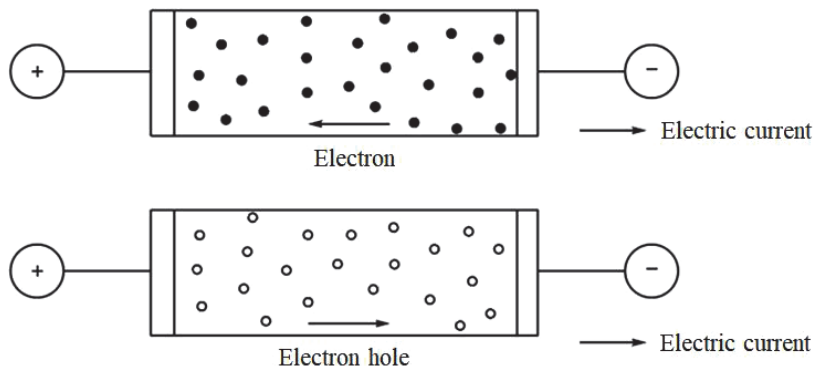


Fig. 4.13. Hole and electronic conductivity diagram.

fewer electrons. Therefore, semiconductors with more electrons are called N-type semiconductors, and those with more holes are P-type semiconductors. The more electrons in N-type semiconductors (or holes in P-type semiconductors), the lower the resistance.

Therefore, the conditions for the formation of semiconductor luminescent materials are as follows.

- (1) The bandgap width is suitable. The photon energy released by the recombination of minority carriers and majority carriers injected into PN junctions is less than the bandgap width. Therefore, the bandgap width of the crystal must be larger than the photon energy of the required light-emitting wavelength. Because the long-wave limit of visible light is about 700 nm, E_g must be greater than 1.78 eV for visible light-emitting diodes. Because the peak value of visual sensitivity is at 550 nm, the crystal $E_g \geq 2.3$ eV should be used to obtain light-emitting diodes with high visible light efficiency. If the blue light-emitting diode with short wavelength is to be obtained, the condition of $E_g \geq 2.7$ eV must be satisfied. The combination of light-emitting diodes and photodetectors can be used to transmit light signals, and silicon photodetectors are suitable for this purpose. The sensitivity of silicon photodetectors peaks at 900 nm with the distribution of wavelength. The long wavelength limit is

determined by $E_g = 1.12$ eV of silicon. In short wavelength, the sensitivity decreases slowly. Therefore, GaAs with a bandgap width slightly larger than that of silicon is the most suitable crystal for infrared light-emitting diodes.

- (2) P-type and N-type crystals with high conductivity can be obtained. In order to fabricate excellent PN junctions, two kinds of crystals, P-type and N-type, are needed, and the conductivity of these two crystals should be very high. Although the bandgap width of II–VI compound crystal is appropriate, it only shows N-type (or P-type) conductivity, so it is not suitable as the crystal of light-emitting diode.
- (3) High-quality crystals with good integrity can be obtained. The incompleteness of the crystal has a great influence on the luminescence phenomenon. The incompleteness here refers to the impurities and lattice defects that can shorten the lifetime of a few carriers and reduce the luminescence efficiency. Therefore, it is necessary to obtain high-quality crystals with good integrity to fabricate high-efficiency light-emitting diodes. The properties of crystals and the growth methods of crystals are related to the integrity of crystals. SiC can satisfy the conditions of (1) and (2), but the crystal growth temperature is very high and the crystal with good integrity cannot be obtained, which becomes the obstacle to the development of SiC light-emitting diodes. It is difficult for GaN to produce blue light-emitting devices in large quantities, mainly because of the quality of crystal.
- (4) The probability of luminescence recombination is high. The high probability of luminous recombination is necessary to improve the luminous efficiency, which is the reason why direct transition crystals are used to fabricate light-emitting diodes. However, when using direct transition crystals, there is a great loss of light emission to the outside. In addition, even in indirect transition crystals, as long as high quality crystals can be used and appropriate impurities can be doped to form high concentration luminescence centers with high probability of luminescence recombination, high efficiency luminescence can be obtained, and the loss of light emission to the outside is also small.

If we look for the crystals of visible or near infrared light-emitting diodes according to the above conditions, the direct transition crystals are GaAs, GaN, InP and so on. The indirect transition crystals are AlP, AlAs, AlSb, GaP, etc. But AlP, AlAs and AlSb are unstable in the air. GaAs and GaP are the only crystals that can emit visible light. In order to extend the direct transition bandgap of GaAs to visible band, GaAs–GaP and GaAs–AlAs mixtures can be used. These two mixtures are complete solid solutions in the range of all components.

4.2. Absorption of Light by Semiconductors

4.2.1. *Classification of light absorbing mechanisms in semiconductors*

As far as the general law of light absorption of matter is concerned, light waves incident on the surface of matter. Without considering thermal excitation and impurities, electrons in semiconductors are basically in valence bands, and electrons in conduction bands are very few. When light is incident on the semiconductor surface, the valence electrons in the outer layer of the atom absorb enough photon energy, cross the forbidden band and enter the conduction band, and become free electrons that can move freely. At the same time, a free hole is left in the valence band to generate an electron–hole pair. Semiconductor light absorption mechanism can be roughly divided into the following five types.

- (1) Intrinsic absorption. When photons with sufficient energy act on semiconductors, the valence band electron absorption energy is stimulated to the conduction band and forms an electron–hole pair, which is called intrinsic absorption. As shown in Fig. 4.14, the valence band electron absorbs photon energy and transits into the conduction band. The phenomenon of electron–hole pairs is called intrinsic absorption. Obviously, the condition for intrinsic absorption is that the photon energy must be larger than the bandgap of the semiconductor E_g , so that the electron absorption on the valence band E_v can jump into the bottom

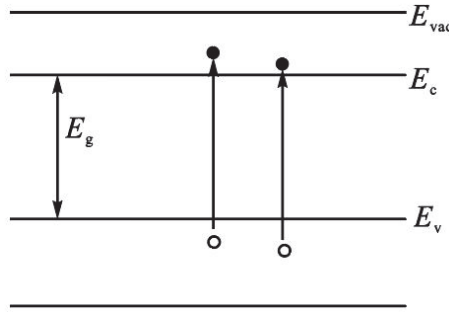


Fig. 4.14. Intrinsic absorption.

level E_c of the conduction band, i.e.

$$h\nu \geq E_g. \quad (4.1)$$

Thus, the wavelength limit of the intrinsic absorption can be obtained as

$$\lambda \leq hc/E_g = 1.24/E_g \quad (4.2)$$

That is to say, only when the wavelength is shorter than the incident radiation can the device produce intrinsic absorption and change the conductivity of intrinsic semiconductor. The intrinsic absorption has the following characteristics: it can make the semiconductor have a high absorption coefficient; the absorption coefficient of direct bandgap semiconductor is higher than that of indirect bandgap semiconductor material; the penetration depth of absorption coefficient is related to wavelength λ .

- (2) Exciton absorption. Because of the Coulomb effect, the electrons actually excited in the conduction band and the holes left in the valence band are in a binding state, which is called exciton. The physical essence of exciton absorption is the absorption of light caused by the transition of valence band electrons to exciton levels below the direct conduction band. When the incident photon energy $h\nu$ on the intrinsic semiconductor is less than E_g , or the incident photon energy $h\nu$ on the impurity semiconductor is less than the impurity ionization energy (ΔE_D or ΔE_A),

the electron does not produce the transition between bands to become a free carrier and is still in the stimulated state under the constraint of the original bound charge. The electrons in the stimulated state are called excitons. The phenomenon that excitons are produced by absorbing photon energy is called exciton absorption. Obviously, exciton absorption does not change the conductivity of semiconductors.

- (3) Lattice vibration absorption. The absorption of photon energy by lattice atoms in the far infrared spectral region is directly transformed into the increase of lattice vibration kinetic energy, which macroscopically shows that the temperature of the object rises and causes the thermal sensitivity of the material.
- (4) Impurity absorption. Non-ionized impurity atoms (donor atoms) in N-type semiconductors absorb photon energy $h\nu$. If $h\nu$ is greater than or equal to donor ionization energy ΔE_D , the outer electrons of impurity atoms will jump from the impurity level (donor level) to the conduction band and become free electrons. Similarly, in P-type semiconductors, the valence electrons jump into the acceptor level and leave holes in the valence band after the electrons in the valence band absorb photons with energy $h\nu$ greater than ΔE_A (acceptor ionization energy). It is equivalent to the hole absorption photon energy jumping into the valence band at the acceptor level. These two impurity semiconductors absorb photons of sufficient energy, and the process of ionization is called impurity absorption. Obviously, the long wave limit of impurity absorption is

$$\lambda_L \leq 1.24/\Delta E_D, \quad (4.3)$$

$$\lambda_L \leq 1.24/\Delta E_A. \quad (4.4)$$

Because $E_g > \Delta E_D$ or ΔE_A , the long wavelength of impurity absorption is always longer than that of intrinsic absorption. Impurity absorption can change the conductivity of semiconductor and also cause photoelectric effect.

- (5) Free carrier absorption. Free carriers, i.e. carriers that can move freely in the band, are electrons in the conduction band and vacancies in the valence band in semiconductors. For general

semiconductor materials, when the frequency of incident photons is not high enough to cause the transition between bands or exciton formation, absorption still exists, and its intensity increases with the increase of wavelength. This is caused by the transition of free carriers between energy levels in the same energy band, which is called free carrier absorption. Free carrier absorption does not change the conductivity of semiconductors.

Generally speaking, only intrinsic absorption and impurity absorption can directly produce unbalanced carriers and cause photoelectric effect. Other absorption transforms radiation energy into heat energy in varying degrees, which increases the device temperature and accelerates the motion of thermally excited carriers without changing the conductivity of semiconductors.

The types of electrons involved in the optical absorption transition of semiconductor materials are as follows.

- (1) Valence electrons. An electron that interacts with other atoms to form a chemical bond in an electron outside the nucleus. The valence electrons of the main group elements are the outermost electrons of the atoms of the main group elements; the valence electrons of the transition elements are not only the outermost electrons, but also the penultimate third-layer electrons of the sub-outer electrons and some elements can become valence electrons. The number of valence electrons is the number of outermost electrons in the main group elements. The valence electrons of atoms of transition group elements include not only the outermost electrons, but also the sub-outer electrons.
- (2) Inner shell electrons. The number of electrons in the inner layer is equal to the total number of electrons minus the number of electrons in the outer layer.
- (3) Free electrons. Off-domain electrons refer to electrons that are not confined to the interior of an atom. Such electrons can move in matter or vacuum when they are subjected to external electric or magnetic fields. In general, some valence electrons can move freely from atoms in crystal lattices. Such electrons are called free electrons.

- (4) For example, after substituting Si with P, there are a positive center of P and a isolated electron. The electron will be bound by the P positive center.

4.2.2. *Semiconductor optical absorption theory*

4.2.2.1. *Light absorption coefficient*

The mechanisms of light absorption in semiconductors include inter-band transition absorption (intrinsic absorption), carrier absorption and lattice vibration absorption. The absorption intensity of light is usually expressed by the parametric absorption coefficient describing the decay rate of light in semiconductors in the unit of cm^{-1} . If the incident light intensity is I and the distance of light entering the semiconductor is x , then the definition is given.

$$\alpha = -1/I(dI/dx). \quad (4.5)$$

4.2.2.2. *Characteristics of interband absorption spectrum curve*

The relationship between the absorption coefficient α and the photon energy $h\nu$ of Si and GaAs in the case of interband transition is measured as shown in Fig. 4.15.

- (i) The absorption coefficient increases with the photon energy.
- (ii) All kinds of semiconductors have a lower limit of photon energy absorption (or the long wavelength limit of photon absorption — cut-off wavelength), and the lower limit of photon energy decreases with the increase of temperature (i.e., the cut-off wavelength increases).
- (iii) The absorption spectrum curve of GaAs is steeper than that of Si.

4.2.2.3. *A simple description of the interband absorption spectrum curve*

Because the inter-band optical absorption of semiconductor is caused by the transition of valence band electrons to conduction band, the optical absorption coefficient is related to the energy density

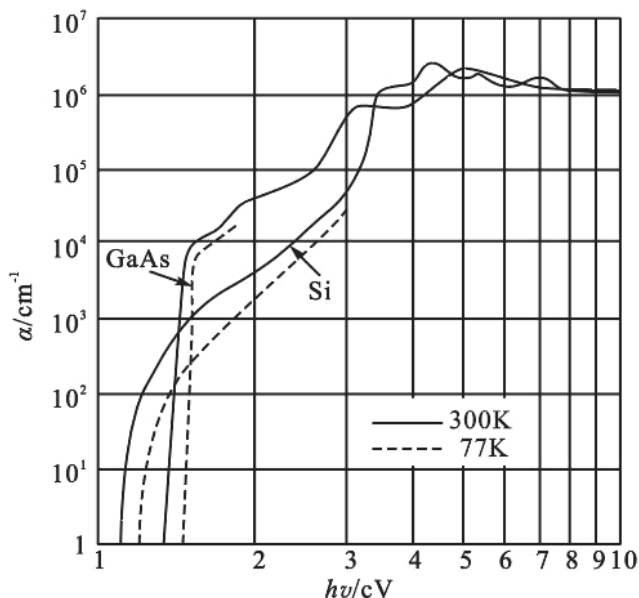


Fig. 4.15. Si and GaAs absorption spectra.

of valence band and conduction band. The distribution of energy density in valence band and conduction band is complex. Under the approximation of free electron and spherical isosurface, the relationship between energy density and energy is subparabolic. The energy density near the top of valence band and the bottom of conduction band is generally very small. Therefore, the absorption coefficient of transition between the top of valence band and the bottom of conduction band is very small. With the increase of energy, the energy density increases, so the absorption coefficient is very small. The absorption spectrum curve increases with the increase of photon energy. However, due to the complexity of the distribution function of the density of states in the actual semiconductor energy band and the fact that the transition of the electron absorption light must conform to the transition rule of quantum mechanics, k -selection rule, the absorption spectrum curve of the semiconductor becomes very complex, and steps and multiple peaks or valleys may appear as shown in Fig. 4.15.

For valence electrons to be able to transition from valence band to conduction band, they should absorb at least the energy of the bandgap width E_g , so as to conform to the law of energy conservation, there exists a minimum light absorption energy, the lower limit of photon energy, which corresponds to the long wavelength limit of light absorption, the cut-off wavelength:

$$\lambda_g = 1.24/E_g. \quad (4.6)$$

Because the bandgap of semiconductor will decrease with the increase of temperature, so the λ_g will also increase with the increase of temperature.

4.3. Excitation and Luminescence of Semiconductors

There are many ways of energy conversion, such as incandescent lamp is tungsten wire through current, high-temperature tungsten wire radiates light; television picture tube is the electron gun emits electrons, hit the screen, phosphor is stimulated to emit light, the energy of electronic movement into light.

In semiconductor light-emitting diodes, electrical energy is directly transformed into light energy. Electric energy produces more electrons and holes than heat balance does. At the same time, it reduces electrons and holes due to recombination, resulting in a new heat balance. In the process of recombination, energy is emitted in the form of light.

4.3.1. *PN junction and its characteristics*

Doping some trace elements into intrinsic semiconductors as impurities can significantly change the conductivity of semiconductors. The impurities are mainly trivalent or pentavalent elements. The intrinsic semiconductor doped with impurities is called impurity semiconductor. Impurity semiconductors are usually doped in intrinsic semiconductors in the order of one-millionth of magnitude, also known as doped semiconductors. Impurities in semiconductors have a great influence on conductivity. Generally, they can be divided into N-type semiconductors and P-type semiconductors. N-type

semiconductors incorporate a small amount of impurity phosphorus (or antimony) into silicon crystals (or germanium crystals). Since semiconductor atoms (such as silicon atoms) are replaced by impurity atoms, four of the five outer electrons of phosphorus atoms form covalent bonds with the surrounding semiconductor atoms, and one of the extra electrons is hardly bound, making it easier to become free electrons. As a result, N-type semiconductors become semiconductors with high electron concentration, whose conductivity is mainly due to free electron conduction. P-type semiconductors incorporate a small amount of impurity boron (or indium) into silicon crystals (or germanium crystals). Since semiconductor atoms (such as silicon atoms) are replaced by impurity atoms, when the three outer electrons of boron atoms form covalent bonds with the surrounding semiconductor atoms, a “hole” will be created, which may attract bound electrons to “fill” the boron atoms. Become a negatively charged ion. In this way, these semiconductors become electrically conductive due to their high concentration of “holes” (equivalent to “positive charges”). In P-type semiconductors, there are more holes, fewer free electrons, more holes and fewer free electrons; in N-type semiconductors, on the contrary, more free electrons and fewer holes.

The essential structure of LED is semiconductor PN junction. When the forward voltage is applied to PN junction, a few carriers are injected, and the light-emitting recombination of a few carriers is the working mechanism of the light-emitting diode. PN junction refers to the structure of adjacent P and N regions in a single crystal. It is usually made on one type of conductive crystal by alloying, diffusion, ion implantation or growth to produce another type of conductive thin layer. Other methods, such as using light and avalanche processes, are not considered because of their low efficiency. The working mechanism of light-emitting diodes is shown in Fig. 4.16.

PN junction not only can efficiently create excess electrons and holes, but also the matching with semiconductor integrated circuits and other semiconductor transistors is extremely advantageous in order to create the necessary voltage for excess electrons and holes (usually DC voltage below 2V). In this paper, a simple mutation

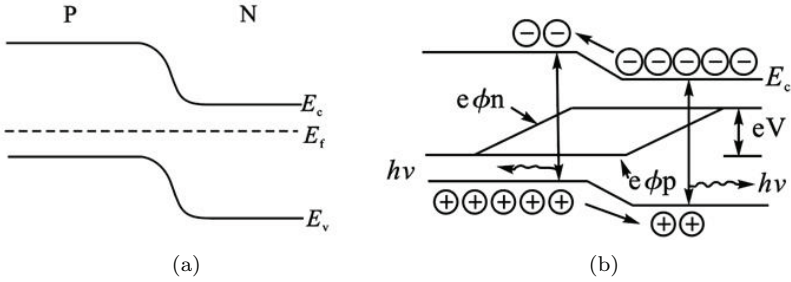


Fig. 4.16. Principle of PN junction luminescence of light-emitting diodes: (a) zero bias (b) forward bias V .

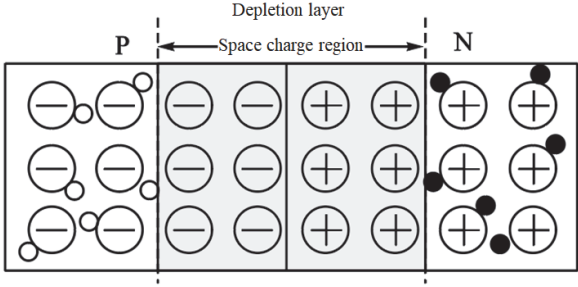


Fig. 4.17. Mutation PN junction model diagram.

junction (as the interface of PN junction, i.e. the junction of N-type and P-type regions with a certain concentration of impurities) is first placed under ideal conditions. Holes in the P region move to the N region due to diffusion; electrons in the N region diffuse to the P region, forming a space charge region near the interface between the P region and the N region, i.e. the depletion layer, as shown in Fig. 4.17.

The depletion layer of space charge is sandwiched between P and N regions, resulting in contact potential difference, thus inhibiting the continuous diffusion of holes and electrons. When the difference between diffusion and contact potential equalizes, it is the state of thermal equilibrium. The charge in the depletion layer is negative in the P region and positive in the N region. Because it was electrically neutral, the holes and electrons were electrically neutral when they moved to each other. In addition, the charge in the depletion layer is

fixed on the lattice and cannot move, so the total negative charge in the depletion layer in P region is equal to the total positive charge in the depletion layer in N region.

In addition, the interface between P and N regions is called PN junction. It can be clearly seen from Fig. 4.18 that the power line density of transverse PN junction is the highest in the depletion layer, and the electric field of PN junction is the largest. For the potential energy of electrons, it can be expected that in the depletion layer, the N region rises continuously. Thermodynamics shows that when the Fermi levels in P and N regions are equal, the process of forming the depletion layer stops. Figure 4.18 is a model of the band diagram near PN junction in thermal equilibrium state.

From the formation principle of PN junction, it can be seen that in order to make PN junction turn on and form current, the resistance of electric field in space charge region must be eliminated. Apparently, adding a larger electric field in the opposite direction, i.e. the positive pole of the external power supply in the P region and the negative pole in the N region, can offset the self-built electric field in the P region, so that the carriers can continue to move, thus forming a linear positive current. The external reverse voltage is equivalent to the resistance of the built-in electric field, and the PN junction cannot be turned on, only a very weak reverse current

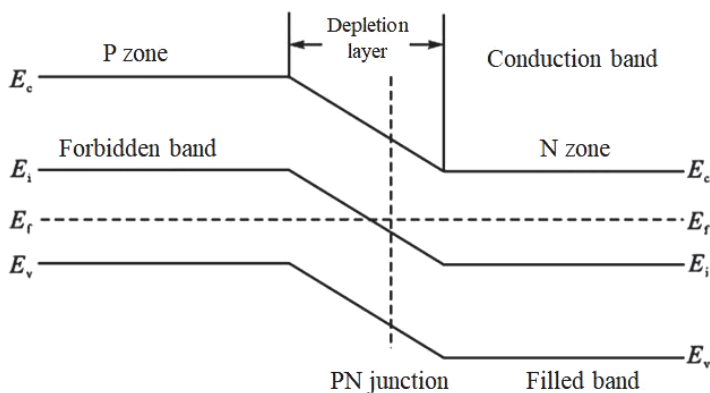


Fig. 4.18. Energy band diagram near PN junction in thermal equilibrium state.

(formed by the drift motion of a few carriers, due to the limited number of minority carriers, current saturation). When the reverse voltage increases to a certain value, because the number and energy of a few carriers increase, the covalent bonds will be destroyed by collision, and the bound electrons and holes will be released, and the current will increase continuously. Finally, the PN junction will be damaged by breakdown (into conductors), and the reverse current will increase sharply.

4.3.1.1. *Reverse breakdown*

When the reverse voltage is applied to PN junction, the space charge region becomes wider and the electric field in the region increases. When the reverse voltage increases to a certain extent, the reverse current will suddenly increase. If the external circuit can't limit the current, the current will burn down PN. When the reverse current suddenly increases, the voltage is called breakdown voltage. There are two basic breakdown mechanisms: tunnel breakdown (also known as Zener breakdown) and avalanche breakdown. The former has a breakdown voltage less than 6 V and a negative temperature coefficient, while the latter has a breakdown voltage greater than 6 V and a positive temperature coefficient.

The avalanche breakdown is that when the carrier drift speed in the barrier layer accelerates to a certain extent with the increase of the internal electric field, the kinetic energy of avalanche breakdown is enough to collide the valence electrons bound in the covalent bond to produce free electron-hole pairs. The newly generated carriers collide with other neutral atoms under the action of strong electric field, and then produce new ones. Free electron-hole pairs, such a chain reaction, dramatically increase the number of carriers in the barrier layer, like an avalanche. Avalanche breakdown occurs in PN junctions with lower doping concentration. The barrier layer is wide, the chance of collision ionization is more, and the breakdown voltage of avalanche breakdown is higher.

Ziner breakdown occurs in PN junctions with high doping concentration. Because of the high doping concentration, the

PN junction is very narrow, so the electric field in the junction layer is very strong (up to 2.5×10^5 V/m) even if a small reverse voltage (below 5 V) is applied. Under the action of strong electric field, the valence electrons of atoms in PN junction will be forced to pull out from the covalent bond to form “electron–hole pairs”, thus producing a large number of carriers. Under the action of reverse voltage, they form a large reverse current and breakdown occurs. Obviously, the physical essence of Zener breakdown is field ionization. The avalanche breakdown voltage of silicon PN junction can be controlled at 8–1000 V by appropriate doping process. Zener breakdown voltage is less than 5 V. Two kinds of breakdown may occur simultaneously at 5–8 V.

Thermoelectric breakdown is that when the reverse voltage is applied to the PN junction, the reverse current flowing through the PN junction will cause heat loss. When the reverse voltage increases gradually, the power loss for a certain reverse current will also increase, which will generate a lot of heat. If there is no good heat dissipation condition to transfer these heat energy in time, the junction temperature will rise. The breakdown caused by thermal instability is called thermal breakdown. The temperature characteristic of breakdown voltage is that when the temperature rises, the lattice vibration intensifies, which shortens the average free path of carrier motion and reduces the kinetic energy before collision. The avalanche breakdown can only occur if the reverse voltage is increased, which has a positive temperature coefficient. However, when the temperature rises, the valence electron energy state in the covalent bond is high, thus Zener strike occurs. The piercing voltage decreases with the increase of temperature and has a negative temperature coefficient.

4.3.1.2. *Unidirectional electricity*

PN junction is turned on when positive voltage is applied. If the positive pole of the power supply is connected to the P region and the negative pole to the N region, a part of the applied positive voltage falls to the PN junction, and the PN junction is positively biased,

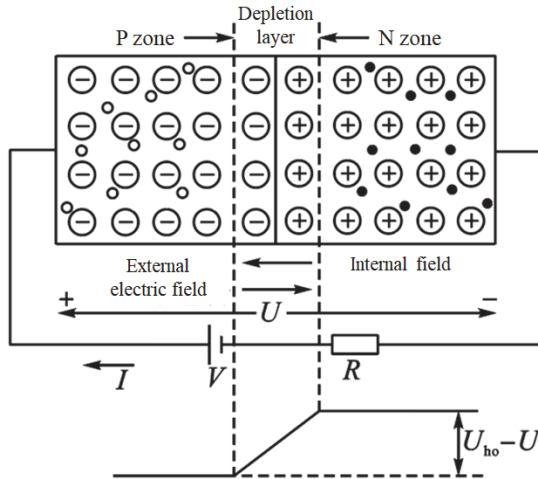


Fig. 4.19. Forward voltage, PN junction conduction.

as shown in Fig. 4.19. Current flows from P-type to N-type. Holes and electrons move toward the interface, narrowing the space charge region, and the current can pass smoothly. The direction is opposite to the direction of electric field in PN junction, which weakens the internal electric field. As a result, the impediment of the internal electric field to the multi-carrier diffusion motion is weakened and the diffusion current is increased. Diffusion current is much larger than drift current, so the influence of drift current can be neglected. PN junction shows low resistance.

PN junction is cut off when reverse voltage is applied. If the positive pole is connected to N area and the negative pole is connected to P area, some of the applied reverse voltage falls to PN junction, and the PN junction is in reverse bias, as shown in Fig. 4.20. Both holes and electrons move far away from the interface, which widens the space charge region and prevents the current from flowing through. The direction is the same as that of the electric field in PN junction, which strengthens the internal electric field. The impediment of the internal electric field to the multi-carrier diffusion motion increases and the diffusion current decreases. At this time, the drift current of minority carriers in PN junction is larger than that of diffusion current

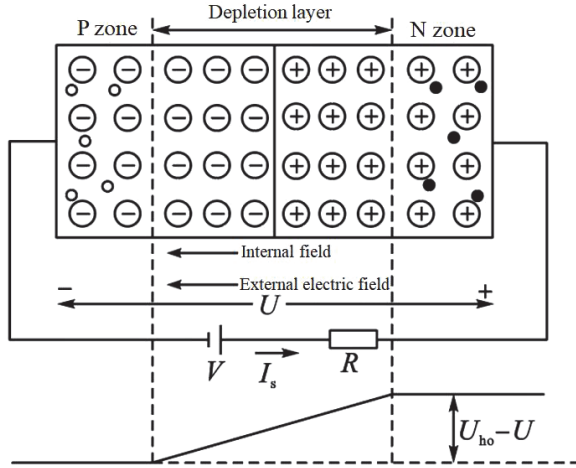


Fig. 4.20. Reverse voltage PN junction cut-off.

under the action of internal electric field, and the diffusion current can be neglected. The PN junction shows high resistance. Under certain temperature conditions, the minority carrier concentration determined by intrinsic excitation is certain, so the drift current formed by minority carrier is constant, basically independent of the magnitude of the added reverse voltage, which is also called reverse saturated current.

When PN junction is applied with forward voltage, it exhibits low resistance and large forward diffusion current; when PN junction is applied with reverse voltage, it exhibits high resistance and small reverse drift current. It can be concluded that PN junctions have unidirectional conductivity.

4.3.1.3. Volt-ampere characteristics

The volt-ampere characteristic (external characteristic) curve of PN junction is shown in Fig. 4.21, which visually shows the unidirectional conductivity of PN junction.

The expression of the volt-ampere characteristic is

$$i_D = I_S \left(e^{\frac{V_D}{V_T}} - 1 \right), \quad (4.7)$$

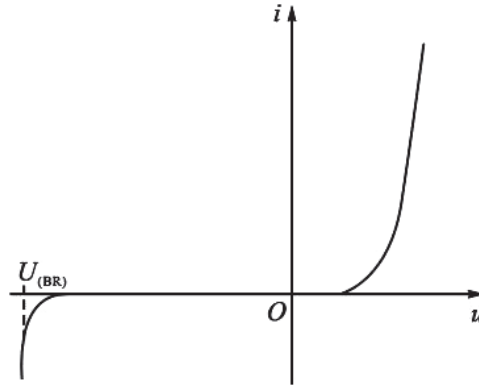


Fig. 4.21. PN junction volt-ampere characteristic curve.

where, i_D is the current passing through PN junction, V_D is the applied voltage at both ends of PN junction and V_T is the voltage equivalent of temperature.

$$V_T = kT/q = T/11600 = 0.026 \text{ V}. \quad (4.8)$$

Among them, k is Boltzmann constant ($1.38 \times 10^{-23} \text{ J/K}$), T is thermodynamic temperature (300 K) and q is electronic charge ($1.6 \times 10^{-19} \text{ C}$). At room temperature, $V_T \approx 26 \text{ mV}$. I_s is a reverse saturated current. For discrete devices, the typical values are 10^{-14} – 10^{-8} A . Diode PN junctions in integrated circuits have smaller I_s values.

- When $V_D \gg 0$ and $V_D > V_T$, $i_D = I_s e^{\frac{V_D}{V_T}}$.
- When $V_D < 0$ and $|V_D| \geq V_T$, $i_D \approx -I_s$.

4.3.1.4. Capacitance characteristics

When PN junction is applied with reverse voltage, the positive and negative charges in the space charge region constitute a capacitive device. Its capacitance varies with applied voltage, mainly including barrier capacitance (C_B) and diffusion capacitance (C_D). Barrier capacitance and diffusion capacitance are both non-linear capacitors.

Barrier capacitance is formed by a thin layer of ions in the space charge region. When the applied voltage changes the voltage drop

on the PN junction, the thickness of the ion thin layer changes accordingly, which corresponds to the amount of charge stored in the PN junction. Barrier region is similar to flat-plate capacitor. Ion charges with equal value and opposite polarity are stored on both sides of the junction. The charge varies with the applied voltage. It is called barrier capacitance. It is expressed by C_B and its value is

$$C_B = dQ/dT. \quad (4.9)$$

The junction resistance is very large when the PN-junction is reverse biased, so the role of C_B cannot be ignored, especially in high frequency, it has a great impact on the circuit. C_B is not constant, but varies with V . Varistor diodes can be fabricated by using this characteristic.

PN junction has abrupt junction and slow junction. Considering abrupt junction, PN junction is equivalent to flat-plate capacitor. Although the applied electric field will widen or narrow the barrier area, the change is relatively small and can be neglected. Then

$$C_B = \varepsilon_s/L. \quad (4.10)$$

When the width of the barrier layer L_0 is known in dynamic equilibrium, the substitution formula (4.10) yields

$$C_T = \varepsilon_s/L_0. \quad (4.11)$$

Diffusion capacitance refers to the accumulation and concentration distribution of polygons on the boundary of PN junction when they diffuse into the opposite region when conducting electricity in the positive direction. The accumulated charge varies with the change of applied voltage. When the forward voltage of PN junction increases, the forward current increases, which requires more carriers to accumulate to meet the requirement of increasing current. When the forward voltage decreases, the forward current decreases, and the holes accumulated in the electrons or N region decrease relatively. Thus, when the applied voltage changes, the carriers are present. “Fill” and “Release” to PN junction. The diffusion capacitance C_D of PN junction describes the capacitance effect of electrons accumulated

in P region or holes accumulated in N region varying with applied voltage.

When the PN junction is biased, the electrons diffuse from the N region to the P region and recombine with the holes provided by the external power supply to form a forward current. The newly diffused electrons accumulate in the P region near the PN junction, forming a certain multi-carrier concentration gradient distribution curve. Conversely, similar concentration gradient distribution curves are formed in the N region when the voids diffuse from the P region to the N region.

C_D is a nonlinear capacitor. When PN junction is biased positively, C_D is larger and the number of carriers is small when PN-junction is biased inversely. Therefore, the value of diffusion capacitance is very small when PN junction is biased inversely, which can be generally ignored.

PN junction capacitance refers to the total capacitance C_j of PN junction, which is the sum of C_T and C_D . $C_j = C_T + C_D$. When the applied forward voltage C_D is very large, C_j is mainly diffusion capacitance (tens to thousands of pictures); when the applied reverse voltage C_D tends to zero, C_j is mainly barrier capacitance (tens to tens of pictures).

4.3.2. *Recombination of injected carriers*

Composition can be divided into two categories: one is the radiation compounding with light (radiation compounding); the other is the non-radiation compounding without light radiation (non-radiation compounding). The former is because the recombination of holes and electrons radiates energy in the form of light energy, which is an important recombination for solid luminescence, while the recombination without light radiation is harmful for solid luminescence. In this regard, many methods to solve the luminescence efficiency are proposed, which can be divided into the following two categories.

Radiative recombination. Electrons and holes recombine by collision, which can be divided into two types: one is not through phonon recombination (direct transition type) and the other is

through phonon medium recombination (indirect transition type); the other is through impurity level recombination, through adjacent level recombination and exciton recombination.

Non-radiative recombination. It can be divided into multi-phonon recombination, Auger recombination and device surface recombination. Multiphonon recombination can be divided into two categories: those related to defect energy levels such as lattice defects and those unrelated to them; those related to defect energy levels and those unrelated to Auger recombination.

4.3.2.1. Radiative compound

Direct and indirect transitions. When the electron at the bottom of the conduction band falls into the full band and recombines with the hole, the energy difference between the initial state and the final state radiates in the form of light. There are two cases as shown in Fig. 4.22, namely direct transition and indirect transition. The following two cases are explained separately.

In the vicinity of room temperature, it is considered that the electrons in the conduction band converge near the lowest energy point, i.e. the cases of Figs. 4.22(a) and 4.22(b). Electrons have

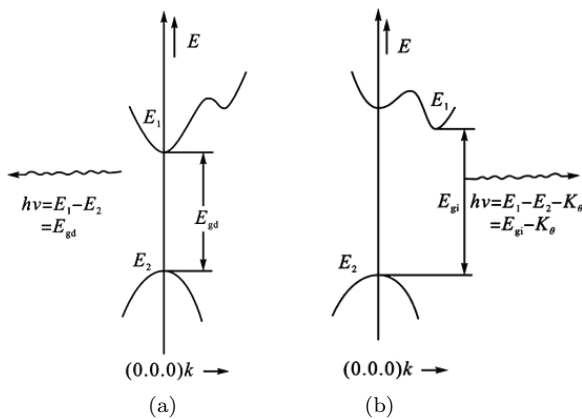


Fig. 4.22. Two energy band structures and radioluminescence: (a) direct transition type and (b) indirect transition type.

energy E_1 . In addition, when the electrons mentioned above fall into full-band holes, the energy becomes E_2 , and light is radiated. Electrons should have energy and momentum during the transition from E_1 to E_2 .

Because the momentum of light is negligible in comparison with the energy of electricity, for light radiation, the same energy state between energy and momentum as shown in Fig. 4.22(a) is beneficial for vertical transition.

If the collision model is taken into account, when electrons and holes collide, the radiation frequency ν is

$$h\nu = E_1 - E_2, \quad (4.12)$$

where, h is Planck constant.

GaAs and GaN are typical examples of direct transition type. $\text{CaAs}_{1-x}\text{P}_x$ and $\text{Ga}_{1-x}\text{Al}_x\text{As}$ ternary compound semiconductors are also direct transition type when x is small. In Fig. 4.22(b), the photon radiation is obtained because the momentum difference between the initial and composite states cannot cause the electron transition. At this time, the phonon retains momentum due to the lattice vibration. If it is used as a collision model, the collisions of electrons, holes and phonons will occur less frequently than in the case of both. In addition, the frequency of radiated light is

$$h\nu = Q_1 - Q_2 - K_\theta, \quad (4.13)$$

where, K_θ is the energy of phonons, that is, the heat energy of lattice vibration.

The energy K_θ is consumed by phonon formation, which is why the radiated light energy becomes smaller. GaP is a typical example of indirect transition. Table 4.3 lists the theoretical recombination probabilities of direct and indirect bandgap semiconductors. In order to improve the luminescence efficiency, it is necessary to find new dopants to form excitons.

- (1) Through the recombination of impurity levels. Consider that most of the impurities in semiconductors containing impurities are ionized near room temperature, as shown in Fig. 4.23(a).

Table 4.3. Theoretical compound probability of direct and indirect bandgap semiconductors (300 K).

Materials	Bandgap type	Compound probability (cm ³ /s)
GaAs	Direct	7.21×10^{-10}
GaSb	Direct	2.39×10^{-10}
InP	Direct	1.23×10^{-9}
InAs	Direct	3.5×10^{-10}
InSb	Direct	4.6×10^{-11}
Si	Indirect	1.79×10^{-15}
Ge	Indirect	5.25×10^{-14}
GaP	Indirect	5.37×10^{-11}

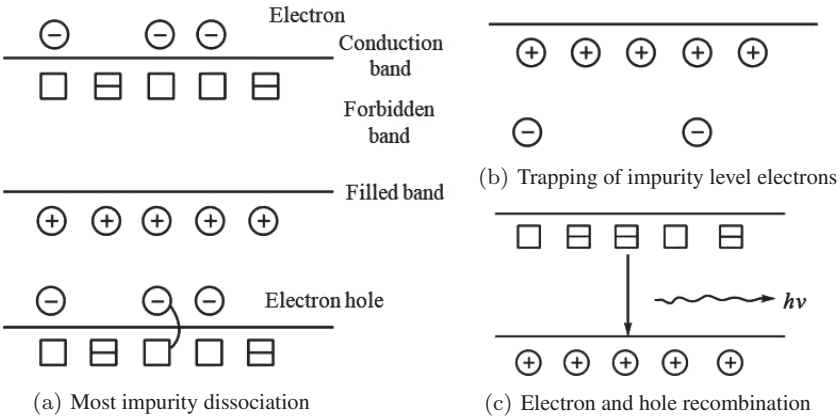


Fig. 4.23. Composition between impurity levels.

Then, as mentioned earlier, the conduction band electrons are captured at the empty impurity level, as shown in Fig. 4.23(b). At this point, the excess energy is consumed by heat. Secondly, the electron captured by the impurity level, which falls into the hole in the full band and recombines, must reabsorb the heat energy of the electron captured by the impurity level and recombine with the hole before returning to the conduction band, as shown in Fig. 4.23(c). Otherwise, the captured electrons cannot move freely.

- (2) Composition of adjacent levels. Active impurities and other impurities, due to the large number of defects, produce mutual influence, that is, conductive impurities, and there is Coulomb attraction between defect groups. In other words, in wave mechanics, the fluctuation factor of electrons spans two levels, causing the transition of electrons. The energy difference between these two levels radiates in the form of light, and in the end, the recombination of electrons and holes continuously causes light radiation, as shown in Fig. 4.24. The infrared luminescence of silicon doped GaAs is such a luminescent mechanism. Since the energy is smaller than the bandgap, i.e. the radiation wavelength is longer, the crystal absorbs less by itself. Therefore, light radiation is obtained outside the crystal.
- (3) Exciton recombination. In semiconductor crystals, besides electrons fixed on lattice atoms (full-band electrons) and electrons (conduction band electrons) which can move freely in crystals, there are electrons fixed on lattice points at their intermediate energies. If explained by wave mechanics, it can be seen from the wave function that there are not only one atom, but also electrons up to several atoms.

In the case of gas atoms, under normal conditions, the external electrons are firmly fixed on the nucleus (full of electrons). If the extranuclear electrons get enough energy, they become free electrons (conduction band electrons) completely free from the

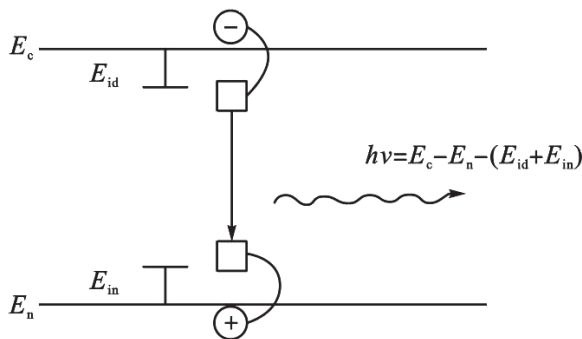


Fig. 4.24. Composition between adjacent levels.

nuclear restriction. However, when they get a little less energy than this, the extranuclear electrons transit to another lateral orbit (excited state electrons), then they have only weak binding force with the nucleus.

This is the electron in the excited state, the atom with the excited state. Then a hole–electron pair is generated with the hole, and the exciton can be transferred to another atom by diffusion. Electron–hole pairs are generally electrically neutral, and the energetic electron–hole pairs radiate outward in the form of light due to the release of energy from the recombination.

It is well known that light emitted by the recombination of excitons, such as the red and green light of GaP. In the case of red light, the oxygen atom replacing P atom and the Zn atom replacing Ga atom in GaP crystal form the above-mentioned electron–hole pairs when they are adjacent to each other. Because of the strong electron affinity of oxygen, the electrons injected into P region are captured by oxygen atoms first and become excited electrons. This electron is recombined by a hole trapped in zinc due to the Coulomb force. Although GaP is an indirect transition crystal, the external quantum efficiency of this mechanism is still very high. The efficiency of GaP:Zn–O can reach 15% and that of GaP:N can reach 0.7%.

Due to the introduction of isoelectronic traps, bound excitons (which cannot transport energy, but can transfer energy through recombination, and emit photons and phonons through recombination) are formed. Both are radiation recombination and non-radiation recombination. Exciton recombination bound to isoelectronic traps is an efficient radiation recombination.

4.3.2.2. *Non-radiative composition*

- (1) The phonon recombination is emitted in stages. As an important semiconductor for solid-state luminescence, the bandgap must be above 1 eV in terms of luminescence wavelength. In addition, because the energy of phonons is about 0.06 eV, when the conduction band electrons fall into full band, if all the energy of electrons generate phonons, then more than 20 phonons will be generated.

The probability of simultaneous generation of so many phonons equals zero. Considering the collision model, the probability of simultaneous collision of so many phonons is minimal.

In practical crystals, it is impossible to generate most phonons at the same time. Because of harmful metal and crystal defects, electrons fall to many energy levels in the forbidden band, and phonons are produced periodically.

- (2) Auger process. The important process of transforming electronic energy into thermal energy is called Auger process, which is a non-radiative composite process. Auger process occurs in the case of free carriers and lattice defects. The effect of the latter is great because of the lower luminous efficiency of the actual devices. Firstly, it is illustrated by Fig. 4.25(a). When an electron in the conduction band recombines with a hole in the full band, the energy is transferred to the other electrons in the conduction band. The electrons that get the energy rise to the high energy level in the conduction band, gradually emit phonons and fall to the lower end of the conduction band. In this way, the energy emitted by the combination of electrons and holes does not radiate in the form of light and is converted into heat

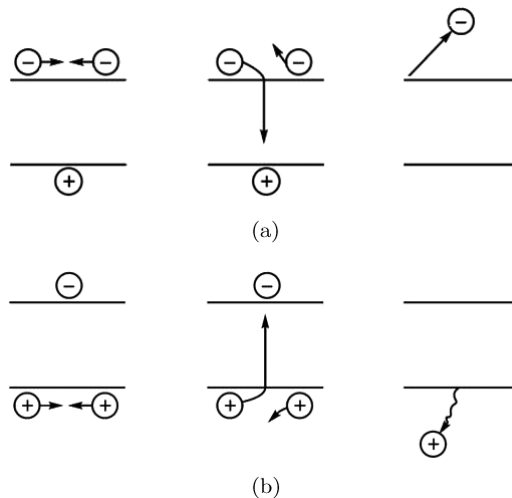


Fig. 4.25. Auger process of free carriers: (a) electron–electron collisions and (b) cavity–cavity collision.

energy. Figure 4.25(b) shows that another hole obtains the energy emitted by the recombination.

In addition to the above, due to the existence of crystal defects, in the process of electron transition, the excess energy is obtained by other electrons and holes, resulting in the production of phonons and depletion. In fact, the main reason for the low luminescence efficiency of light-emitting diodes is the Auger process through the defect level. How to reduce such energy levels has become an important issue in manufacturing technology.

- (3) Surface composite. In the intuitive crystal surface, it can be imagined that there are more defects than the interior. Therefore, the probability of non-radiative recombination on the surface is higher than that inside the crystal. In addition, the carrier migration on the surface of the crystal is different from that on the inside due to periodic damage.

Like other semiconductor tubes, the characteristics of solid-state luminescent tubes, which mainly depend on luminescent efficiency, and the lifetime and reliability of the tubes, are closely related to the surface. When GaP is used, the formation of oxide film on the core surface and the covering of silicon nitride film with $\text{GaAs}_{1-x}\text{P}_x$ can improve the surface condition.

4.4. Light-Emitting Diode Lighting Technology

Diodes using semiconductor PN junctions as light sources were invented in the early 1960s. Red LED first appeared in 1964, and then yellow LED. It was not until 1994 that blue and green LED were successfully developed. In 1996, Nichia (Japan) successfully developed white LED. LED is widely used in the fields of indicator lamp, signal lamp, display screen, landscape lighting and so on, because of its characteristics of power saving, long life, vibration resistance and fast response. It can also be seen everywhere in our daily life. In recent years, with the deepening of the research on semiconductor luminescent materials, the continuous progress of LED manufacturing technology and the development and application of new materials, various colors of ultra-high brightness LED have

made breakthroughs, and high brightness LED will become the fourth generation of green lighting source.

4.4.1. Basic characteristics of LED

4.4.1.1. Basic structure of LED

LED (light-emitting diode), a kind of solid-state semiconductor device, can directly convert electricity into light. An electroluminescent semiconductor chip, encapsulated in epoxy resin, acts as a positive and negative electrode through a pin support, as shown in Fig. 4.26.

The heart of LED is a semiconductor chip. One end of the chip is attached to a bracket, one end is negative, the other end is connected to the positive pole of the power supply, so that the whole chip is encapsulated by epoxy resin. Semiconductor chips consist of two parts, one is P-type semiconductor, in which holes dominate, and the other is N-type semiconductor, in which electrons dominate. But when the two semiconductors are connected, a PN junction is formed between them. When the current acts on the wafer through a wire, the electrons are pushed to the P region, where the electrons recombine with holes and then emit energy in the form of photons, which is the principle of LED luminescence. The wavelength of light,

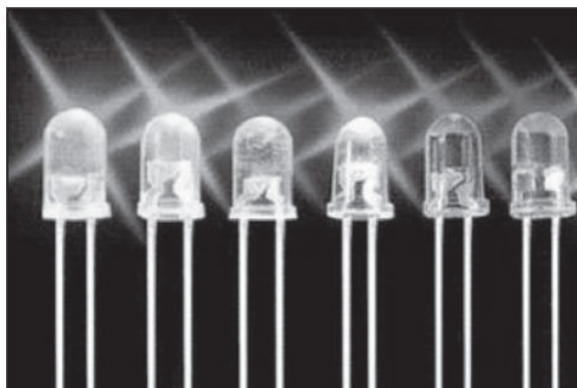


Fig. 4.26. LED beads.

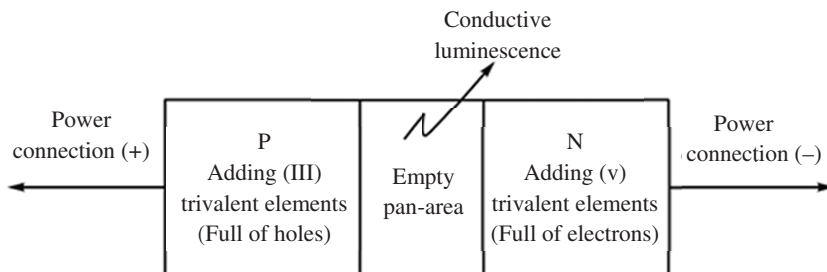


Fig. 4.27. LED structural diagram.

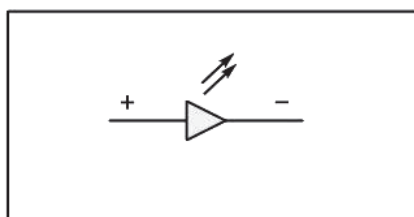


Fig. 4.28. Circuit symbol diagram of LED.

the color of light, is determined by the material that forms PN junctions.

Actually, as a PN junction, as shown in Fig. 4.27, it is a spontaneous emission device that can emit ultraviolet, visible and infrared light. Its luminescence principle is electro-excitation light. At the bulk contact surface of P-type and N-type semiconductors, i.e., the free valence electrons and holes recombine after adding forward (forward) current to the PN junction, thus converting electrical energy into visible light radiation energy.

4.4.1.2. Symbols of LED in electronic circuits

The circuit symbols of light-emitting diodes in electronic circuits are shown in Fig. 4.28. In DC power supply, they are all connected to the line forward, that is, P pole connected to the positive pole of the power supply and N pole connected to the negative pole of the power supply. In AC power supply, because of the low reverse breakdown

voltage of the LED, it is necessary to connect a current limiting resistance with a larger resistance value or to connect a silicon diode in series.

Light-emitting diodes (LED) are made of III-V compounds, such as GaAs (GaAs), GaP (GaP), GaAsP (GaAs), GaN (nitride grafting) and other semiconductors. The core of LED is PN junction. Therefore, it has general V-I characteristics of PN junctions, i.e. forward conduction, reverse cut-off and breakdown characteristics.

When the forward voltage is applied to the light-emitting diode (P plus positive voltage, N plus negative voltage), the forward voltage of ordinary silicon diode is greater than 0.6 V, and the forward voltage of germanium diode is greater than 0.3 V, the forward voltage V_F of the light-emitting diode is greater than 1.5–3.8 V (generally speaking, the working voltage of red and yellow light-emitting diodes is about 2 V, and that of other color light-emitting diodes is about 2 V. When the voltage is about 3 V, the holes in N region and the electrons in P region are injected from P region of the light-emitting diode. The holes in N region and P region are combined with the electrons in N region and P region respectively in the μm region near the PN junction to produce spontaneous emission fluorescence, as shown in Fig. 4.29 (E_g is the forbidden bandwidth). The energy states of electrons and holes in different semiconductor materials are different. When electrons and holes recombine, the photon energy released is different. The larger the photon energy released, the shorter the light wavelength emitted. Commonly used are red, green, blue and yellow light diodes. If the reverse voltage is applied at both ends of the diode (P pole plus negative voltage, N pole plus positive voltage), its current is very small and almost zero, which is called reverse leakage current. However, when the applied reverse voltage exceeds the withstand voltage, there will be a large reverse current. This voltage is called the reverse breakdown voltage of the diode, which will damage the diode when the infinite current measures are taken. Generally, the reverse breakdown voltage of luminescent tubes is about 5 V. The reverse breakdown voltage of red, yellow, yellow and green wafers is about 20–40 V, and that of blue, pure green and purple wafers is

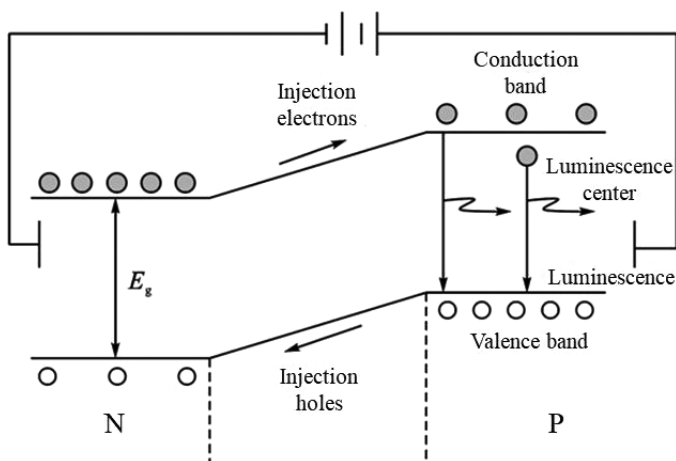


Fig. 4.29. Semiconductor band and compound luminescence.

only about 5 V, depending on the manufacturer and the process of various chips.

4.4.2. LED luminescence principle

Light-emitting diode (LED) is an injection-type electroluminescent device, which injects a large number of unbalanced carriers (injecting holes and electrons into P and N regions respectively) under the applied positive voltage. Under the action of applied electric field, electrons in N region drift to P region, holes in P region drift to N region, recombine and emit light across PN junction barrier.

4.4.2.1. Luminescence conditions for manufacturing light-emitting diodes

- (i) The bandgap of compound semiconductor crystal should be used to obtain the desired wavelength.
- (ii) Luminescent materials can be easily fabricated into PN diodes, and the lattice constants a of each layer should be matched. Make the so-called DH layer, but both sides a must be consistent.

- (iii) The material with bandgap is sandwiched on both sides of the luminous region of the active layer. The bandgap of the active layer is smaller than that of the cladding layer. The refractive index of the active layer is larger than that of the cladding layer, and the luminescence is easily emitted from the inside.
- (iv) It has stable physical and chemical structure. Crystallization has high ionicity, wide bandgap E_g and high melting point. Compound semiconductor crystal materials can work at higher temperatures, such as GaP, AlP, GaN and other compounds.
- (v) Crystals with direct or indirect migration bands. Most of the luminescent regions are direct-migrated bandgap materials, which have higher luminescent efficiency and higher mobility of electrons (holes) than indirect-migrated bandgap materials.

4.4.2.2. *White light mode of light-emitting diodes*

The technical scheme of white LED will be discussed in detail in Chapter 6. Here is a brief introduction.

- (i) Multi-chip. Red LED, green LED and blue LED chips are combined into one pixel to achieve white light.
- (ii) Single LED chip + phosphor. (1) Blue LED + yellow phosphor is the most widely used and mature phosphor at present, but its color rendering is insufficient (70 ~ 80). (2) Blue LED + red powder and green powder have good color rendering performance, but the luminous efficiency of red powder developed at present is not high enough. (3) Near-ultraviolet LED + red, green and blue phosphor has high color rendering index, high light efficiency and adjustable color temperature. From the current feasibility, practicability and commercialization, the third way (near) ultraviolet chips to excite tricolor phosphors is the main direction of white LED development in the future.

4.4.2.3. *Basic characteristics of light-emitting diodes*

- (1) High luminous efficiency. The luminous efficiency of incandescent lamp and halogen tungsten lamp is 12–24 lm/W, that of fluorescent lamp is 50–70 lm/W and that of sodium lamp is

90–140 lm/W. After improvement, the light efficiency of LED will reach 50–200 lm/W, and its spectral monochrome is good, its spectrum is narrow, and it can directly emit colored visible light without filtering. With the research of LED light efficiency in various countries, its luminous efficiency will be improved greatly in the near future.

- (2) Low power consumption. Low power LED single transistor power is $0.03 \sim 0.06$ W, driven by DC, single transistor driving voltage is 1.5–3.5 V, current is $15 \sim 20$ mA, fast reaction speed, high frequency operation. Under the same lighting effect, the power consumption is 1/8 of incandescent lamp and 1/2 of fluorescent lamp. Take 45 W fluorescent lamp as an example. Under the same lighting effect, only 8 W white LED tube is needed.
- (3) Long service life. LED is small, light weight, epoxy resin packaging, can withstand high strength mechanical impact and vibration, not easy to break. The theoretical life is 100,000 hours, and the actual life is 5–10 years, which can greatly reduce the maintenance cost of lamps and lanterns.
- (4) Strong safety and reliability. Low calorific value, no thermal radiation, cold light source, can accurately control the light intensity and luminous angle, light color is soft, no glare, no mercury, lead and other substances that may endanger health.
- (5) Green environmental protection. LED is an all-solid luminous body, shock-resistant, impact-resistant, not easy to break; LED lighting does not contain mercury, xenon, lead and other harmful elements, waste can be recycled, no pollution. LED light does not contain ultraviolet and infrared rays, and has no radiation pollution. The working principle of common energy-saving lamp is that mercury vapor is stimulated by current to emit ultraviolet radiation, while phosphor is irradiated to produce white light and there is ultraviolet leakage.
- (6) The luminescence response speed is fast ($10^{-9} - 10^{-7}$ s), the high frequency characteristic is good, and the pulse information can be displayed. The response time of incandescent lamp is millisecond.
- (7) The light source is small in size and can be assembled freely. It is easy to develop into light, thin and short small lighting products for installation and maintenance.

4.4.3. Characteristic parameters of LED

LED is an optoelectronic device which uses compound material to make PN junction. It has the electrical characteristics (I - V characteristics, C - V characteristics), optical characteristics (spectral response characteristics, light intensity pointing characteristics, time characteristics) and thermal characteristics of PN junction devices.

4.4.3.1. Electrical characteristics of LED

- (i) I - V characteristics. The main parameters for characterizing the fabrication performance of PN junctions on LED chips are shown in Fig. 4.30. The I - V characteristic of LED is nonlinear and rectifier: single conductivity, i.e. the external positive bias voltage is low contact resistance, and vice versa, high contact resistance.

Forward dead zone: (Fig. 4.30, OA or OA') Point A is the open voltage for V_0 . When $V < V_a$, the applied electric field can overcome many barrier electric fields due to carrier diffusion, and R is very large at this time. For different LED, the open voltage is different, GaAs is 1 V, red GaAsP is 1.2 V, GaP is 1.8 V, GaN is 2.5 V.

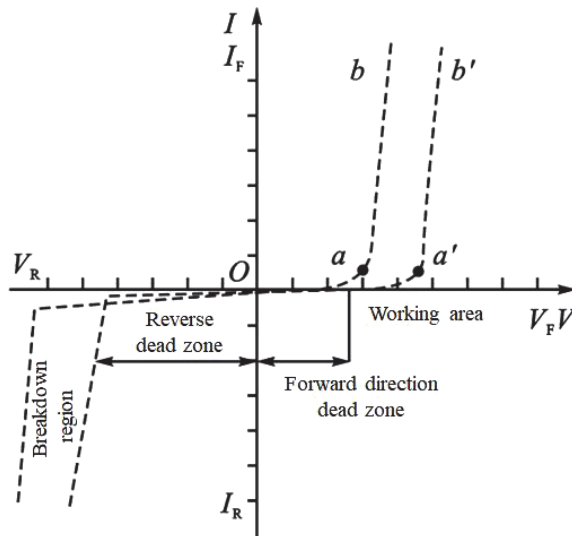


Fig. 4.30. I - V characteristic curve.

Workspace: Current I_F is exponentially related to applied voltage.

$I_F = I_S(e^{qV_F/KT} - 1)$, I_S is reverse saturated current. When $V > 0$, the forward workspace I_F of $V > V_F$ increases with V_F index, $I_F = I_S e^{qV_F/KT}$.

Reverse dead zone: PN junction with reverse bias when $V < 0$, GaP is 0 V and GaN is 10 μ A when $V = -V_R$ and reverse leakage current IR ($V = -5$ V).

Breakdown zone: When $V < -V_R$, V_R is called reverse breakdown voltage; V_R voltage corresponds to I_R is reverse leakage current. When the reverse bias keeps increasing and $V < -V_R$, the I_R increases suddenly, that is, the breakdown phenomenon occurs. Due to the different kinds of compounds used, the reverse breakdown voltage V_R of various LED is also different.

- (ii) C - V characteristic. Because the chip of LED has 9 mil $\times$ 9 mil ($250 \mu\text{m} \times 250 \mu\text{m}$), 10 mil $\times$ 10 mil ($250 \mu\text{m} \times 250 \mu\text{m}$), 11 mil $\times$ 11 mil ($280 \mu\text{m} \times 280 \mu\text{m}$), 12 mil $\times$ 12 mil ($300 \mu\text{m} \times 300 \mu\text{m}$), the size of PN junction is different, which makes its junction capacitance (zero bias) $C \approx n + pf$. The C - V characteristic is quadratic function. As shown in Fig. 4.31, it is measured by a C - V characteristic tester for 1 MHz AC signal.

Maximum permissible power consumption P_{Fm} , when the current flowing through the LED is I_F and the tube voltage drop is U_F , the

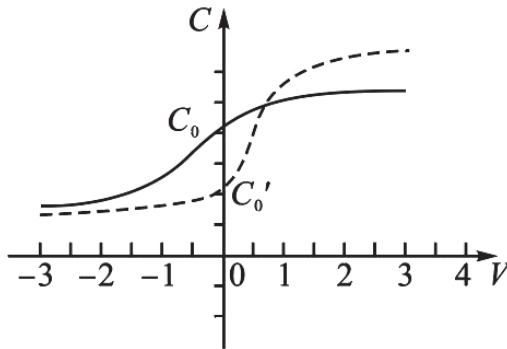


Fig. 4.31. LED C - V characteristic curve.

power consumption is $P = U_F I_F$. When the LED works, the external bias voltage and bias current will make the carriers compound to emit light, and some of them become hot, which will increase the junction temperature. If the junction temperature is T_j and the external ambient temperature is T_a , when $T_j > T_a$, the internal heat transfers outward by means of the tube seat, and the dissipated heat (power) can be expressed as $P = KT (T_j - T_a)$.

Response time, which represents how fast a display tracks changes in external information. The response time of existing LCD (liquid crystal display) is 10^{-5} – 10^{-3} s, and CRT, PDP and LED all reach 10^{-7} – 10^{-6} s (μ s level).

From the point of view of use, the response time is the delay time of LED lighting and extinguishing, that is, t_r and t_f in Fig. 4.32. The value of t_0 in the figure is very small and negligible. Response time depends mainly on carrier lifetime, junction capacitance and circuit impedance. Lighting time of LED — rising time t_r refers to the time it takes to turn on the power supply to make the luminous brightness reach 10% of the normal value until the luminous brightness reaches 90% of the normal value. LED extinction time — descent time t_f

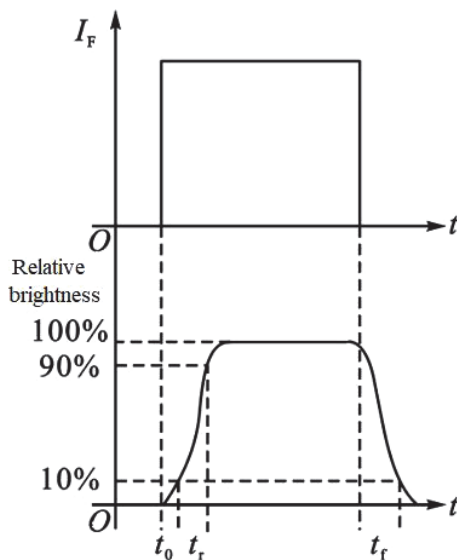


Fig. 4.32. LED response time.

refers to the time it takes for normal luminescence to decrease to 10% of its original level. The response time of different materials is different, such as GaAs, GaAsP, GaAlAs, GaP is less than 10^{-9} s and GaP is 10^{-7} s. Therefore, they can be used in high frequency systems ranging from 10 MHz to 100 MHz.

4.4.3.2. Optical properties of LED

There are two series of light-emitting diodes, infrared (non-visible) and visible light. The former can be used to measure the radiation, and the latter can be used to measure its optical properties.

- (1) Normal intensity and angular distribution I_θ . The luminous intensity (normal intensity) characterizes the important performance of luminous devices. A large number of applications of LED require cylindrical and spherical packaging. Because of the role of convex lenses, they all have strong directivity: the maximum light intensity is located in the normal direction, and the angle between them and the horizontal plane is 90° . When deviating from the normal direction and different angle of θ , the light intensity also changes. Angular distribution I_θ of luminous intensity describes the distribution of luminous intensity of LED in all directions of space. It mainly depends on the packaging process (including bracket, mold head, epoxy resin with or without scattering agent). In order to obtain high directivity angular distribution (Fig. 4.33), the position of the LED core is far from the die head; the conical (bullet) die head is used; and no scattering agent is added to the encapsulated epoxy resin. Semi-intensity angle $\theta_{1/2}$ refers to the angle between the direction where the luminous intensity value is half of the axial intensity value and the luminous axis (normal direction). The above measures can make the scattering angle of LED ($2\theta_{1/2}$) = about 6° , and greatly improve the directivity. At present, the scattering angles of circular LEDs commonly used in packaging are 5° , 10° , 30° and 45° .
- (2) Peak wavelength and its spectral distribution. The luminous intensity or power output of LED varies with the wavelength, and a distribution curve is drawn, which is the spectrum

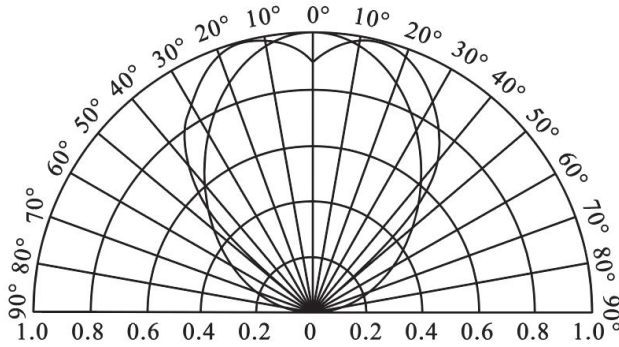


Fig. 4.33. Angular distribution of LED luminescence intensity.

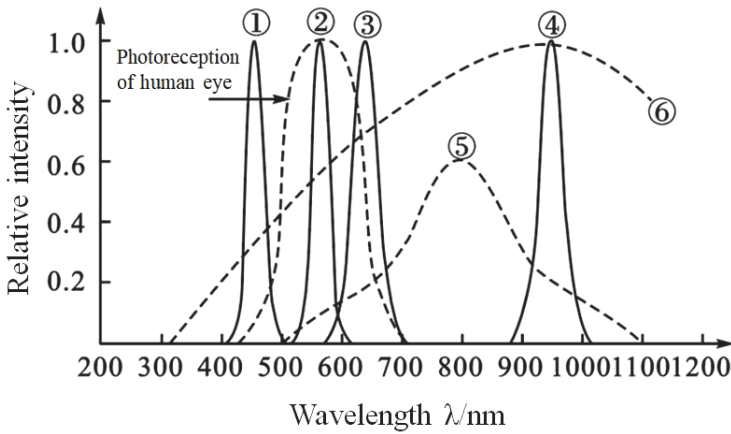


Fig. 4.34. Spectral response curve of LED.

distribution curve. After the determination of this curve, the relevant chromaticity parameters such as the main wavelength and purity of the device will also be determined. The spectrum distribution of LED is related to the type and properties of compound semiconductor used in preparation and PN junction structure (epitaxial layer thickness, doping impurities), but not to the geometry and packaging mode of device.

As shown in Fig. 4.34, the spectral response curves of LED fabricated from different compound semiconductors and dopants are

plotted. Among them, (1) is a blue InGaN/GaN light-emitting diode with a peak of $\lambda_p = 460\text{--}465\text{ nm}$. (2) is a green GaP:N LED with a peak of $\lambda_p = 550\text{ nm}$. (3) is red GaP:Zn-O LED with a peak of $\lambda_p = 680\text{--}700\text{ nm}$. (4) is an infrared LED using GaAs material, with a peak of $\lambda_p = 910\text{ nm}$. (5) is a Si photodiode, usually used for photoelectric reception. (6) is standard tungsten filament lamp. As can be seen from the figure, no matter what material the LED is made of, it has the strongest relative light intensity (the largest light output), which should have a relative wavelength, which is called peak wavelength, expressed in λ_p . Only monochromatic light has a wavelength of λ_p .

- (3) Spectral line width. There are two points whose light intensity equals half of the peak value (maximum light intensity) at $\pm\Delta\lambda$ on both sides of the peak value of the LED spectrum line. The width between the two points corresponds to the width of $\lambda_p - \Delta\lambda$ and $\lambda_p + \Delta\lambda$ is called the width of the spectrum line, also called the half power width or half-high width. The half-height width reflects the narrow spectral line width, that is, the parameters of LED monochrome, and the half-height width of LED is less than 40 nm.
- (4) Main wavelength. Some LED luminescence is not only monochrome, that is, not only has a peak wavelength, but also has several peaks, not monochrome light. For this purpose, the main wavelength is introduced to describe the chroma characteristics of LED. The main wavelength is the wavelength of the main monochrome light emitted by the LED, which can be observed by the human eye. The better the monochromaticity, the better the wavelength of λ_p is. For example, GaP materials can emit multiple peak wavelengths, while the main wavelength is only one. With the long-term operation of the LED, the junction temperature increases and the main wavelength shifts to the long wave.
- (5) Light flux. Luminous flux F is the radiation energy of the total light output of LED, which indicates the performance of the device. F is the sum of the energy emitted by the LED in all directions. It is directly related to the working current. As the current increases, F increases. The unit of luminous flux of visible LED is lumen (lm).

The power and luminous flux of LED radiation are related to chip material, packaging technology level and the size of external constant current source. At present, the maximum luminous flux of monochrome LED is about 1 lm, and the F of white LED is 1.5–1.8 lm (microchip). For power level chip of $1\text{ mm} \times 1\text{ mm}$, white LED is made with $F = 18\text{ lm}$.

- (6) Luminescence efficiency and visual sensitivity. The quantum efficiency of LED can be divided into internal efficiency (the efficiency of converting electric energy into light energy near PN junction) and external efficiency (radiation to external efficiency). The former is only used to analyze and evaluate the characteristics of chips. The most important photoelectric characteristic of LED is the ratio of radiated light energy (luminous quantity) to input electric energy, that is, luminous efficiency. Visual sensitivity is the use of some parameters in illumination and photometry. The human visual sensitivity has a maximum value of 680 lm/W at $\lambda = 555\text{ nm}$.

If the visual sensitivity is recorded as K_λ , the relationship between luminous energy P and visible light flux F is $P = \int P\lambda D\lambda$ and $F = \int K\lambda P\lambda D\lambda$.

Luminescence efficiency — quantum efficiency η = number of photons emitted/number of carriers in PN junction = $(e/hcI) \int \lambda P\lambda D\lambda$. If the input energy is $W = UI$, the luminous energy efficiency $\eta_P = P/W$; if the photon energy $hc = ev$, then $\eta = \eta_P$, the total optical flux $F = (F/P)P = K\eta_P W$, $K = F/P$.

Luminance efficiency: The luminous flux F /power $W = K\eta_P$ of LED is used to evaluate the characteristics of external packaged LED. Luminance efficiency of LED is higher than that of visible light under the same applied current, so it is called visible light luminescence efficiency (Table 4.4).

- (7) Luminous brightness. Brightness is another important parameter of LED luminous performance, which has strong directivity. The brightness $B_0 = I_0/A$ in the positive normal direction, specifying that the surface brightness of the luminous body in a certain direction is equal to the luminous flux emitted by the unit projection area on the surface of the luminous body in the unit solid angle, in the unit of cd/m^2 or Nit.

Table 4.4. Several common LED lumen efficiencies (visible light years).

LED luminous color	λ_p (nm)	Materials	Visible light-emission efficiency (lm/W)	External quantum efficiency (%)	
				Maximum	Average value
Red light	700	GaP:Zn-O	2.4	12	1 ~ 3
	660	GaPALAs	0.27	0.5	0.3
	650	GaAsP	0.38	0.5	0.2
Yellow light	590	GaP:N-N	0.45	0.1	
Green light	555	GaP:N	4.2	0.7	0.015–1.15
Blue light	465	GaN		10	
White light	Band	GaN-YAG	Microchip 1.6, Large Chip 18		

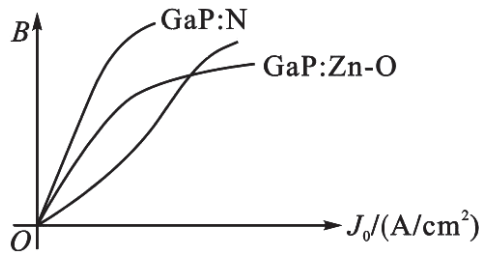


Fig. 4.35. Brightness — current density curve of LED.

If the surface of the light source is an ideal diffuse reflector, the brightness B_0 is constant, independent of direction. The surface brightness of clear blue sky and fluorescent lamp is about 7000 Nit, and that of the sun is about 14×10^8 Nit from the ground. The brightness of LED is related to the applied current density. Generally, the brightness of LED, J_0 (current density) and B_0 increase approximately. In addition, the brightness is also related to the ambient temperature. As the ambient temperature increases, η_C (composite efficiency) decreases and B_0 decreases. When the ambient temperature remains unchanged, the current increases enough to cause the PN junction temperature to rise, and the brightness will be saturated (Fig. 4.35).

- (8) Life span. The phenomenon of light intensity or brightness attenuation of LED with long working time is called aging. The

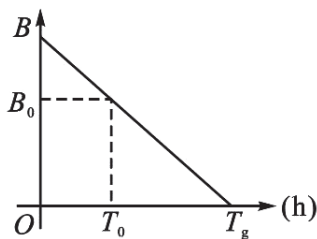


Fig. 4.36. Brightness life curve of LED.

aging degree of the device is related to the size of the external constant current source. It can be described as $B_t = B_0 e^{-t/\tau}$, B_t is the brightness after t time and B_0 is the initial brightness. Typically, the time t experienced by reducing the brightness to $B_t = 1/2 B_0$ is called the lifetime of the diode. It takes a long time to determine t . Usually the lifetime is estimated (Fig. 4.36).

Measuring method: After burning the LED with a constant current source for 10^3 – 10^4 hours, we measured B_0 , $B_t = 1000$ – 10000 , substituted $B_t = B_0 e^{-t/\tau}$ to get τ , and then substituted $B_t = 1/2 B_0$ to get the lifetime t . For a long time, it has been considered that the lifetime of LED is 10^6 hours, which refers to a single LED. With the development and application of power-type LED under $I_F = 20$ mA, foreign scholars believe that the percentage of light attenuation of LED is the basis of lifetime. For example, the light attenuation of the LED is 35% and its lifetime is more than 6000 h.

4.4.3.3. Thermal characteristics

The optical parameters of LED are closely related to PN junction temperature. Generally, it works at low current $I_F < 10$ mA, or 10–20 mA for a long time to light the LED continuously, and the temperature rise is not obvious. If the ambient temperature is high, the main wavelength or λ_p of the LED will drift to the long wavelength and B_0 will decrease. Especially, the temperature rise of the lattice and large display screen has a great influence on the reliability and stability of the LED. The scattering ventilation device should be specially designed.

The relationship between the main wavelength of LED and temperature can be expressed as

$$\lambda_P(T') = \lambda_0(T_0) + T_\varepsilon \times 0.1 \text{ nm}/^\circ\text{C}. \quad (4.14)$$

Formula (4.14) shows that when the junction temperature rises by 10°C , the wavelength drifts to long wave by 1 nm, and the uniformity and consistency of luminescence become worse. As a light source for lighting, miniaturization and dense arrangement are required to improve the intensity and brightness of light per unit area. In particular, attention should be paid to the use of heat-dissipating lamp shell or special general equipment to ensure the long-term work of LED.

4.4.4. LED driving technology

LED driver requirements: no matter how the input voltage changes, no matter how the forward voltage (VF) of the LED changes, no matter how the ambient temperature changes, to ensure that the current flowing through the LED is a constant value. There are two driving modes of LED: one is classified according to the characteristics of driving power supply, the other is classified according to the connection mode of LED load.

4.4.4.1. Commonly used LED driving mode

- (1) Step-down LED driver — the output voltage must be lower than the input voltage, mainly for general lighting.
- (2) Boost-type LED driver — the output voltage must be higher than the input voltage, which is mainly used in the backlight system of portable liquid crystal display.
- (3) Boost/buck LED driver — can work in Buck mode or boost mode according to the relationship between input voltage and output voltage. It is suitable for all kinds of portable lamps powered by batteries.
- (4) SEPIC LED driver — the output voltage is not only higher than the input voltage, but also lower than the input voltage. The polarity of the output voltage and the input voltage is the same,

and the isolation of the input and output circuits can be realized. Applicable to battery-powered automotive lighting driving power supply.

4.4.4.2. Load connection mode

- (1) Series LED driving mode. As shown in Fig. 4.37, the main features of the device are uniform brightness, high efficiency and simple wiring (only two leads are needed between the driver and the LED). The disadvantage is that as long as one LED is damaged, all other LEDs will be extinguished; the output voltage of the power supply must be high enough, and the output capacitor has a large capacity.
- (2) Parallel LED drive mode. As shown in Fig. 4.38, the main advantage is that it is suitable for low voltage and small current LED, and can drive common anode or common cathode LED

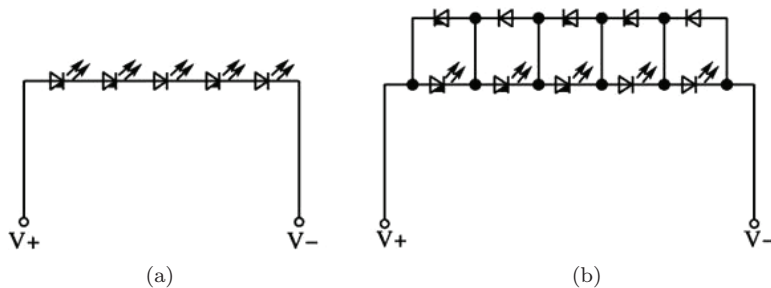


Fig. 4.37. LED series connection.

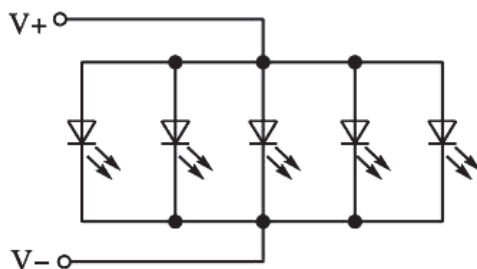


Fig. 4.38. LED parallel connection.

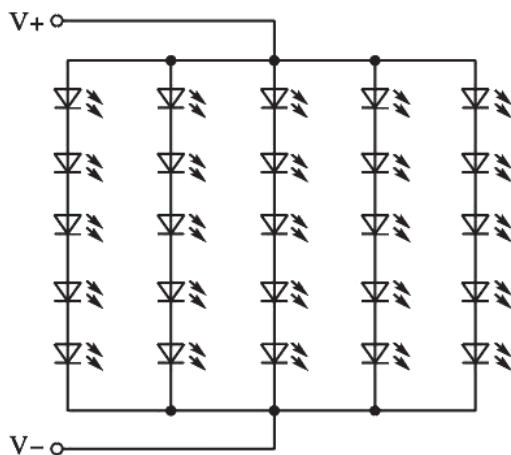


Fig. 4.39. LED hybrid connection.

module. The disadvantage is that the current of each branch must be stable in order to ensure uniform brightness.

- (3) Driving mode of hybrid LED. As shown in Fig. 4.39, the advantages are flexible design and the ability to drive common anode or common cathode LED modules.

4.4.4.3. Characteristics of LED driving technology

Because of this characteristic, single-guide electrical devices need to supply power to LED with DC current or one-way pulse current.

The barrier potential determines the threshold voltage, and the LED will be fully turned on when the voltage applied on the LED exceeds the threshold voltage. The threshold voltage is generally above 2.5 V, and the tube voltage drop is 3–4 V in normal operation.

The temperature coefficient of PN junction is negative, and the barrier potential of LED decreases with the increase of temperature. Therefore, the LED cannot be directly powered by a voltage source. Current limiting measures must be taken. Otherwise, with the increase of the working temperature of the LED, the current will become larger and larger, resulting in damage.

The ratio of the current flowing through the LED to the light flux of the LED is also nonlinear. The luminous flux of LED increases

with the increase of current flowing through the LED, but it is not proportional. There is an extreme efficiency point with the change of current, so the LED should work at a current value with high luminous efficiency.

Due to the differences in production technology and material characteristics, the barrier potential and internal resistance of the same type of LED are not exactly the same, which leads to the inconsistency of tube voltage drop when the LED works, and the negative temperature coefficient of the barrier potential of the LED. Therefore, the LED cannot be directly used in parallel.

The development trend of LED driving power supply can be summarized as high reliability, long life, high efficiency, high power constant, intelligent, adjustable light and short, light and thin.

4.4.5. LED application

Because of its high luminous efficiency, low power consumption, long service life, strong safety and reliability, and environmental protection, LED light source has been widely used in the field of lighting in recent years. The application of LED light source in lighting field is the product of the high-speed development of semiconductor light-emitting materials technology and the gradual penetration of the concept of “green lighting”. “Green lighting” is a new concept put forward in the field of lighting abroad in the late 1980s. The implementation of “Green Lighting Project” in China began in 1996. Important steps to achieve this plan are to develop and promote efficient and energy-saving lighting devices, save lighting power, reduce environmental and light pollution, and establish a high-quality, efficient, economical, comfortable, safe and reliable lighting system beneficial to the environment. The application fields of LED light source are mainly shown in the following six aspects.

4.4.5.1. Building exterior lighting

The projection of an area of a building is nothing more than the use of projection lamps which control the shape of the round head and square head of the beam angle, which is completely consistent

with the traditional concept of projection lamps. However, due to the small and thin source of LED, the research and development of linear projection lamps has undoubtedly become a major highlight of LED projection lamps, because many buildings do not have traditional projection lights in place at all. It is easy to install, can be installed horizontally or vertically, and is better integrated with the surface of the building. It brings new lighting ideas for lighting designers and expands the creative space. It will also have an impact on the lighting methods of modern and historical buildings.

4.4.5.2. *Landscape lighting*

Because unlike the traditional light source, LED is mostly glass bulb shell, it can be well integrated with urban street furniture, and can be used in urban leisure space such as path, staircase, deck, waterfront, gardening for lighting. For flowers or low shrubs, LED can be used as a light source for lighting. LED concealed projection lamps will be particularly popular. The fixed end can be designed as plug-and-pull type, which can be adjusted conveniently according to the height of plant growth.

4.4.5.3. *Marking and indicative lighting*

Places requiring space restriction and guidance, such as road and pavement partition display, staircase step local lighting, emergency exit indicator lighting, can use LED self-luminous buried lights with appropriate surface brightness or lamps embedded in vertical walls, such as floor guiding lights or seat side guiding lights in theatre auditorium, and guiding lights in shopping mall inner floors. In addition, compared with neon lights, because of its low voltage, no fragile glass, it will not increase the cost because of the bending in the production, so it is worth popularizing in the logo design.

4.4.5.4. *Interior space display lighting*

As far as lighting quality is concerned, because LED light source does not have heat, ultraviolet and infrared radiation, it will not damage exhibits or commodities. Compared with traditional light

source, lamps do not need additional filters. The lighting system is simple, cheap and easy to install. Its precise light distribution can be used as a substitute for optical fiber lighting in museums. Commercial lighting mostly uses color LED, interior decorative white LED combined with interior decoration to provide auxiliary lighting, hidden light band can use LED, especially for low space.

4.4.5.5. *Entertainment places and stage lighting*

Because of the dynamic and digital control of color, brightness and dimming of LED, active saturated color can create static and dynamic lighting effect. From white light to any color in the whole spectrum, the use of LED opens a new way of thinking in the lighting of this kind of space. Compared with the 50–250 h life of par lamp and halogen lamp, the maintenance cost and the frequency of replacing light source are reduced. In addition, LED overcomes the phenomenon of color deviation of metal halide lamp after a period of time. Compared with par lamp, LED has no thermal radiation, which can make the space more comfortable. At present, the application of LED color decorative wall in catering buildings has become popular.

4.4.5.6. *Vehicle indicator lighting*

LED navigation information display for vehicle road traffic. In the field of urban transportation and expressway, LED is widely used as a variable indicator and lighting function to replace similar products abroad. Among them, electric power dispatching, vehicle dynamic tracking, vehicle dispatching management and so on, are also gradually using high-density LED display screen to play the role of indicator lighting.

4.5. Photoluminescence and Electroluminescence of Organic Materials

4.5.1. *Photoluminescence principle of organic materials*

As long as any object has a certain temperature, it must have radiation (red light, infrared radiation) which is in thermal equilibrium

with this temperature. There are two kinds of light radiation: equilibrium radiation and non-equilibrium radiation, namely thermal radiation and luminescence. Non-equilibrium radiation refers to the system deviates from the original equilibrium state under the excitation of external energy. If the object returns to the equilibrium state, its excess energy is released in the form of light radiation, it is called luminescence. Luminescence is that matter emits excess energy in the form of light outside of thermal radiation. Photoluminescence refers to the phenomenon that objects rely on external light sources for irradiation, thus obtaining energy, generating excitation, leading to luminescence. It also refers to the process that substances absorb photons (or electromagnetic waves) and re-emit photons (or electromagnetic waves). Therefore, when the external light source (such as ultraviolet, visible light or even laser) irradiates the photoluminescent materials, the luminescent materials will emit such as visible light, ultraviolet light and so on. The luminescent process is generally as follows: the matrix lattice or activator (or luminescent center) absorbs excitation energy; the matrix lattice transfers the excitation energy absorbed to the activation. Activated activator emits fluorescence and returns to the ground state, accompanied by some non-luminescent transitions, and energy is emitted in the form of heat.

One of the most important components of organic matter is hydrocarbons. Organic compounds with double or triple bonds between carbon atoms, known as unsaturated hydrocarbons, usually have strong photoluminescence (PL). These organic molecules all have π bonds, and their excited states are closely related to their luminescence. Molecules with double bonds, such as aromatic hydrocarbons (i.e., benzene series compounds, including dyes), polyenes, nucleic acid, amino acid, and some polymers, play an important role in the luminescence of their π bonds. When atoms form molecules, s electrons form σ bonds with each other, while p electrons form π bonds. In the ground state, electrons bond. Whether it is a σ bond or a π bond, there are two bonding electrons with opposite spins, whose total spin is zero ($S = 0$). So it becomes singlet, usually S_0 (Fig. 4.40). When an electron is excited, if its spin remains unchanged, there is still total spin $S = 0$, and the excited state is singlet, with $S_1, S_2, S_3, \dots$ represents different singletons. If the spin reverses and the

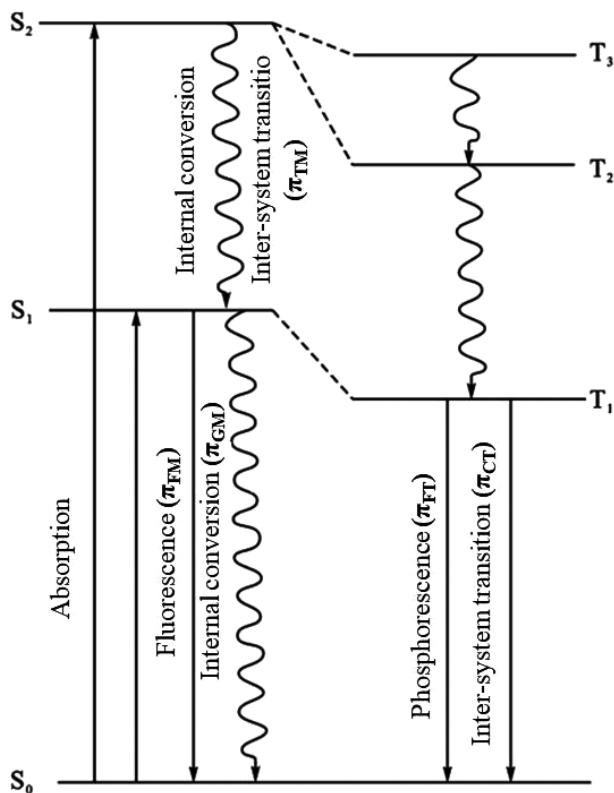


Fig. 4.40. Schematic diagram of excitation and transition of π electrons.

spins of the two electrons are parallel, then the total spin is $S = 1$, which becomes a triplet state: $T_1, T_2, T_3, \dots$. Some people also call the T state a trilinear state. In fact, the T state is degenerate, that is, the energy of the three states is equal, so it is generally shown as a line. Only under certain conditions can the T state split and two or three lines appear in the spectrum. So it's better to call it triplet. According to the spin selection rule, the transition between singlet and triplet states is forbidden. Generally, the triplet energy level is lower than the corresponding singlet state, that is, S_1 is higher than T_1 , S_2 is higher than $T_2, \dots$. Of course, this is not strict. Sometimes S_1 is higher than T_1 and T_2 .

Figure 4.40 gives a schematic diagram of excitation and transition of π electrons. Most of the luminescence is from S_1 to ground state. The luminescence of S_2 level is rare, because the probability of its energy transferring to S_1 through non-radiative multi-phonon transition is very high, which is about $10^{12} S^{-1}$ order of magnitude. This non-radiative process chemist is usually called internal conversion. The life of S_1 is about 10^{-9} s, which is quite long compared with that of S_2 . Fluorescence characterization is the transition of $S_1 \rightarrow S_0$. As with atomic spectroscopy, the selection rule of $S \rightarrow T$ is not strict. Spin-orbit coupling results in mixing of different spin states. Therefore, the transition of $S_1 \rightarrow T_1$ may occur, and the transition of $T_1 \rightarrow S_0$ is naturally possible, but the corresponding decay time of luminescence is much longer than that of fluorescence, some of which can be up to the order of second, which is phosphorescence. The process of transferring from S_1 to T_1 without radiation is called inter-system crossing. As can be seen from the figure, the energy of phosphorescent photons is less than that of fluorescence, so the wavelength is longer. When the gap between T_1 and S_1 is in the order of kT , the molecule can also transfer from T_1 to S_1 by means of the energy of thermal vibration, and then emit light from S_1 . This fluorescence is called delayed fluorescence, also known as E-type delayed fluorescence. Because of its long duration, it can be compared with phosphorescence, and the fluorescence of electrons returning to S_0 immediately from S_1 is also called prompt fluorescence. When the two T_1 states with high density or close to each other can interact, their energy addition can excite electrons to higher states than the S_1 state (e.g. S_2), resulting in delayed fluorescence. Lumb calls this fluorescence P-type delayed fluorescence, which is different from the E-type delayed fluorescence. In the latter case, the total number of $T_1 \rightarrow S_0$ transitions decreases, and the phosphorescent intensity decreases accordingly. So this process is also called triplet annihilation. The energy of the lowest excited state (S_1) of the π electron is relatively small, generally equivalent to the wavelength of near-ultraviolet to visible light. This is why it is particularly important in luminescence. Let me show you here that there should also be a molecular vibration level in Fig. 4.40, which is omitted here.

Among organic molecules, only aromatic molecules and their derivatives have the highest luminescence efficiency. Benzene is the basic unit of aromatics. Its emission spectrum covers the wavelength range of 260–330 nm, but its quantum efficiency is very low, less than 10%. Hanging a group on the benzene ring can greatly improve the efficiency, and the wavelength shifts slightly to the short wave direction. But if we hang straight-chain phenyl, such as terphenyl, tetrabiphenyl, etc., or hang several phenyls at the same time on several median of benzene ring, such as m-terphenyl, 1,3,5-triphenylbenzene, it will have a great impact, making the emission wavelength move to long-wave direction a lot, such as polycyclic naphthalene and pyrene.

For organic materials, the luminescence of organic compounds is a radiation transition phenomenon from excited state to ground state. Most organic compounds have even electrons, which exist in pairs in various molecular orbitals in the ground state. When an electron in a molecule absorbs the energy of light being excited, fluorescence is generated from the excited state (singlet S), the vibrational level relaxation to the lowest excited singlet state (S_1), and finally from S_1 to the ground state S_0 .

4.5.2. *Relation between molecular structure and luminescence properties*

The structure of organic molecules and their luminescent properties are obviously related, but the general rules have not yet been found. Here are some of the conclusions Krasovitskii and Baltin have drawn from their published work.

Conjugate chain length, red shift of spectrum and efficiency increase. This is for simple aromatic compounds. But for more complex compounds, chain length is not the only determinant.

Different series have different “critical” lengths. The efficiency of the chain will decrease if it is longer. This is due to the increase of rotational and vibrational modes of molecules with complex structures and the increase of the probability of non-radiative transition. For compounds with long conjugated bonds, the rigid structure of molecules is an important condition for high efficiency, because it can

minimize the radiation-free loss. In some cases, if hydrogen bonds are formed within the molecule, the rigidity of the molecule can be increased.

In addition, if the molecule forms a new ring, it will also make the original non-luminous molecule into luminous molecule. This is also due to the formation of a heterocycle, which increases the molecular rigidity, i.e. the structure of heterocycles, the properties and number of heteroatoms can affect the characteristics of luminescence. Another factor is the three-dimensional structure of molecules. The substituents of aromatic and heterocyclic rings are also important factors affecting the luminescence. Some substituents affect the charge transfer, and others enhance the polarization of molecules. Nitro is generally a luminescent toxicant, but some nitro-containing molecules have strong luminescence, and most of them are yellow and green. Halogens sometimes have quenching effects.

4.5.3. *Luminescence quenching phenomenon*

Luminescence quenching refers to the weakening or disappearance of the intensity of a given luminescence band (or line). There can be two cases: the intensity of the given band decreases, and there is no change in other bands, which means that the lost excitation energy is converted to heat energy; the given band decreases, and other bands are generated or enhanced, which means that the excitation energy of this part is partially converted to another energy level system or another luminous center.

When the concentration of luminescent substances in solution or solid is relatively high, the phenomenon that the luminescent intensity tends to saturate or even decrease often occurs, which is called concentration quenching. It was Förster and Kasper who first discovered this phenomenon. When they observed the luminescent properties of pyrene $C_{16}H_{10}$ (molecular structure shown in Fig. 4.41), they found that the blue luminescence of pyrene decreased with the increase of concentration, while a wide unstructured band appeared in the long wave direction. However, no change in absorption was observed when the absorption spectra were measured. Therefore,

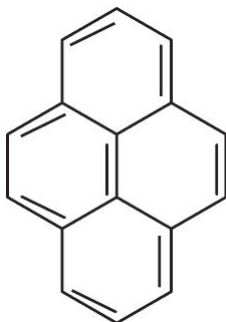


Fig. 4.41. Molecules of pyrene.

they believe that the luminescence centers involved are formed in the excitation state, that is to say, a new compound is formed. Förster and Kasper called this compound transient dimer, and then formally named excimer, or the simplified expression of excited dimer, to distinguish it from the normal combination of two molecules, dimer. Excited molecules are also “dimers” composed of two molecules, but one of them is in the excited state, whether it is singlet or triplet. Once the relaxation returns to the ground state, it breaks down into two original molecules. Therefore, only when the concentration of luminescent substances is high, the number of excited molecules can be observed. At this time, only the emission spectrum changes, but the absorption spectrum does not. The formation of excitation molecules is not uncommon, nor unique to pyrene molecules. On the contrary, for aromatic hydrocarbons and their derivatives, this is almost a rule and often happens.

Of course, concentration quenching is not necessarily due to the formation of excitation molecules. It is also possible that as the concentration increases, the distance between molecules decreases, and the energy transfer between them becomes easier. This means that the lifetime of the excited state is prolonged. Since there is a certain probability of non-radiative transition in each excited molecule, the longer the residence time in the excited state, the greater the possibility of non-radiative transition. In addition, there may be some quenching impurities or defects near some molecules.

The long-term transfer of exciting molecules in crystals also increases the chance of contact with these quenchers. These are the factors that cause concentration quenching. For the latter two cases, no new bands are generated.

When it comes to quenching agents, oxygen can be said to be the most common quenching agent, especially in solution. The role of oxygen is quite complex. It can promote (or catalyze) intra-system transition or inter-system transition. It may also combine with organic molecules to form so-called excitation complexes, which are now called excitation complexes by many people in the third edition of English–Chinese Vocabulary of Chemistry and Chemistry. The complex may emit light as well as exciting molecules, but its wavelength changes. Therefore, the quenching of the existing luminescent bands is also shown. Excited state complexes can also occur in solids, especially near the interface between the two materials. Another common quencher is halogen. The luminescence of organic molecules always decreases when a halogen substituent is attached to them. Luminescence quenches even if it is not in contact with a substituent but only with halides. The reason for quenching is that the halogen enlarges the internal conversion.

4.5.4. Application of organic photoluminescent materials

Fluorescent whitening agents, which absorb ultraviolet light, can emit purple or blue fluorescence, and can complement each other with the yellow formed over time on the articles. Adding in various fibers, plastics and paper can produce whitening effect.

Fluorescent dyes and fluorescent pigments, after absorbing visible light of short wavelength, can not only reflect visible light of long wavelength, but also emit visible fluorescence, so they have higher brightness and bright color. Fluorescent dyes and organic fluorescent compounds can be used as laser medium in laser technology. Using monochrome laser source or scintillator as optical pump, they can excite fluorescent compounds molecules and emit fluorescence for fluorescence analysis.

Fluorescence analysis is a widely used analytical technique with high sensitivity, good selectivity, fast analysis speed and good reproducibility.

4.5.5. *Structure and luminescence principle of organic electroluminescent materials*

The research of organic electroluminescent materials began in 1960s. In 1987, Tang of Kodak company in USA selected 8-hydroxyquinoline aluminium as luminescent material with strong electron transmission ability. The organic electroluminescent devices with low working voltage and high luminescent brightness were fabricated by using ultra-thin-film technology and new device structure. The research of electrolytic materials has entered a new stage of research and breakthrough. In 1990, the Journal *Nature* reported the electroluminescence of *ppv*, which opened up a new field of light-emitting devices. The research of polymer thin-film electroluminescent devices made organic electroluminescent dyes develop from small organic molecules to polymers, and became a hot research project. Subsequently, flexible polymer light-emitting devices based on plastic substrates emerged, which greatly promoted the development of light-emitting devices and made the research of organic electroluminescence widely spread around the world.

Organic light-emitting devices have the simplest three-layer structure, and the organic light-emitting layer is sandwiched between the upper and lower electrodes — cathode and anode. With the deepening of technical research and the development of manufacturing technology, in order to improve the ability of electrodes to inject holes and electrons in order to improve the luminescence efficiency, the device is usually made into a multi-layer structure, and then a hole transmission layer and an electronic transmission layer are added on both sides of the luminous layer. The screen projecting light uses a substrate glass and a driving circuit. The main function of the anode is to produce voids, indicating that the material of the electrode is tin indium. The main function of the cathode is to generate electrons. When the device is applied with positive

voltage, under the action of electric field, holes and electrons emit in the organic light-emitting layer, and then emit through transparent anode.

Because the number of holes produced by the anode and the cathode is usually different, it means that when a part of the holes pass through the whole OLED structure layer, they will not encounter electrons coming from the opposite direction, so the energy consumption is very large and the efficiency is very low. In this way, holes and electron transport layer are introduced: when the hole transport layer of the anode transports holes, the electron transport layer on the cathode side transports electrons, which correspondingly blocks each other's electrons and holes. Thus, the efficiency is obviously improved, and the material of the luminous layer needs to be doped with a certain amount of fluorescent dopant. Usually used to increase light efficiency and luminous color. Substrate glass and drive circuit play the role of light transmission and support fixation. In order to make the image bright, many OLEDs fake color filters in the internal test. Driving circuit is an electrical circuit used to control the operation of cathode and anode.

4.5.6. Advantages of organic electroluminescent materials

- (1) Low energy consumption, organic electroluminescent materials do not need backlighting.
- (2) The response speed is very fast (from several microseconds to tens of microseconds), which is very important in displaying active images.
- (3) It has strong environmental adaptability, good temperature characteristics and can be displayed in low-temperature environment.
- (4) It can realize wide angle of view, high resolution display and high contrast.
- (5) Organic electroluminescent materials have simple structure and low cost. They do not need background light sources and filters. They can produce ultra-thin, light weight and portable products.

4.5.7. *Main problems of organic electroluminescent materials*

- (i) Life problem. The main factors affecting life are chemical aging of organic matter, dissolution of organic film by heating during driving, insulation damage caused by micro-defects, aging of interface between electrode/organic film or organic film/organic film, and aging caused by instability of amorphous organic film. If low-temperature polycrystalline silicon is introduced as its driving circuit, the lifetime of the material will be greatly prolonged.
- (ii) Chromaticity. Most of the luminescent materials have the problem of insufficient color purity, which makes it difficult to display bright colors. Especially red color performance is particularly poor.
- (iii) Large size problem. When the size of the device becomes larger, there will be many problems, such as the problem of driving mode, the problem of material life under scanning mode, and the problem of uniformity of luminescence of display screen, etc.

4.5.8. *Future development trends of organic electroluminescent materials*

At present, the development of OLED is quite popular in the world. It is believed that OLED will be a powerful competitor of TFT-LCD. Active matrix-driven organic electroluminescent display technology, which combines low-temperature polycrystalline silicon technology with OLED, is the future development direction. Japanese and Taiwan manufacturers have transformed the amorphous silicon production line into a polycrystalline silicon production line with OLED, and have realized the mass production of small-size display panels. In order to occupy a place in the industrialization process of OLED, while developing new OLED materials, actively developing low-temperature polycrystalline silicon technology is the most important.

References

- Fang Z, 1992. *Semiconductor Luminescent Materials and Devices*, Shanghai: Fudan University Press.
- Pan J, Tian M, Tong J, 2011. *Fundamentals of Material Science*, Beijing: Tsinghua University Press.
- Xiao Z, 2008. *Semiconductor Lighting Luminous Materials and Applications*, Beijing: Chemical Industry Press.
- Zhang Z, Jiang X, Xu S, 1996. Stability of organic films electroluminescence, *J. Luminescence* 17(2):178–182.
- Zhang Z *et al.*, 2000. Multi color organic light emitting diodes and its stability, *J. Luminescence* 21(4):308–313.

This page intentionally left blank

Chapter 5

Inorganic Photoluminescent Materials

Human beings have long noticed the existence of luminescent materials in nature, and since the 17th century, luminescent phenomena have gradually become the research object of experimental science. Luminescent materials widely exist in people's lives. They are functional materials that can convert various forms of energy absorbed from the outside world into unbalanced light radiation. There are many kinds of luminescent materials, such as photoluminescence, cathode ray luminescence, electroluminescence, thermoluminescence, photoluminescence, radiation luminescence and so on. The inorganic photoluminescent materials are described in detail below.

5.1. Inorganic Photoluminescence

5.1.1. *Photoluminescence process*

Photoluminescence (PL) refers to the phenomenon of light-emitting materials stimulated by ultraviolet, visible or infrared light. It roughly undergoes three main stages: absorption, energy transfer and light emission. The absorption and emission of light occur in the transitions between energy levels, both passing through the excited state, and the energy transfer is due to the motion of the excited state. The energy of exciting light radiation can be absorbed directly by the luminescent center (activator or impurity) or by the matrix of the luminescent material. In the first case, the absorption energy of the luminescence center transits to a higher level, and then back to a lower level or ground-state level to emit light. The study of the properties of these

excitation states involves the interaction between the impurity center and the lattice, which can be analyzed by the crystal field theory. With the enhancement of the effect of crystal field, the absorption and emission spectra become narrower and narrower, and the temperature effect becomes stronger and weaker, which makes part of the excitation energy become lattice vibration. In the second case, the matrix absorbs light energy and forms electron-hole pairs in the matrix. They may move in the crystal and be bound to various luminescence centers. The luminescence is caused by the combination of electrons and holes. When the luminescent center ions are in the energy band of the matrix, a local energy level will be formed, which lies between the conduction band and valence band of the matrix, i.e. in the forbidden band of the matrix. In different matrix structures, the localized energy levels formed by the luminescent center ions in the forbidden band are different, which results in different transitions and different luminescent colors under the excitation of light.

In fact, the luminescence process of photoluminescent materials is more complex, generally consisting of three processes.

- (1) The excitation energy is absorbed by the matrix lattice or activator (or luminescence center).
- (2) The excitation energy absorbed by the matrix lattice is transferred to the activator.
- (3) The activated activator emits fluorescence and returns to the ground state, accompanied by some non-luminescent transition energy emitted in the form of heat.

The whole luminescence process is illustrated in Figs. 5.1 and 5.2. Sometimes, in addition to doping activator, another ion, called sensitizer, is doped in the matrix. As shown in Fig. 5.2, this ion can strongly absorb the excitation energy, and then transfer the energy to the activator. The sensitized rare earth ion emits fluorescence and returns to the ground state, accompanied by a non-luminescent transition, and the energy emits in the form of heat.

The process of light absorption is that when light shines on an object, one part is reflected and scattered, the other part enters the object, and the light is absorbed except for the part through which

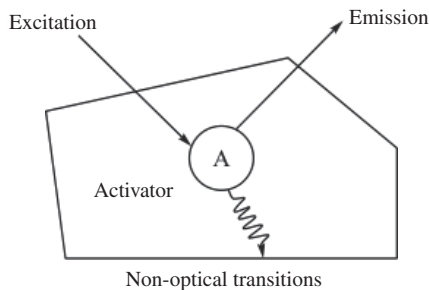


Fig. 5.1. Photoluminescence process of phosphors.

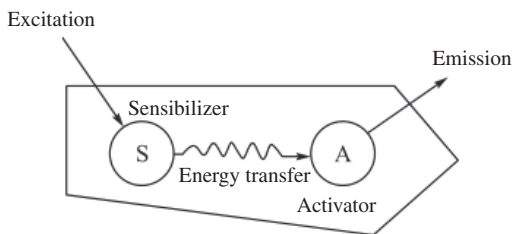


Fig. 5.2. Luminescence process of energy transfer from sensitizer to activator.

it passes. Luminescent materials can only emit light after absorbing excitation energy, so absorption spectrum is an important index to characterize the luminescent properties of luminescent materials. But the light absorbed by the material does not always cause luminescence, so there is the concept of excitation spectrum, which indicates the change of luminous intensity with the wavelength of excitation light, which indicates what light can excite a certain emission light. Absorption spectra and excitation spectra are interrelated.

5.1.2. *Light return to ground state: luminescence*

Luminescence is that substances emit excess energy in the form of light in addition to thermal radiation, and the emission process of this excess energy has a certain duration. It refers to a special phenomenon of light emission, which is fundamentally different from thermal radiation. It is a kind of light emission superimposed

on thermal radiation. It is a non-equilibrium radiation, reflecting the characteristics of luminescent substances. Luminescence has a relatively long duration, that is, after the excitation that the external action stops, the luminescence does not disappear immediately, but gradually weakens. This process is also known as afterglow. Luminescence includes three processes: excitation, energy transfer and luminescence.

5.1.2.1. *Excitation process*

Excitation process: After the excitable system (luminescent center, matrix, exciton, etc.) in the luminescent body absorbs energy, the process of transition from ground state to higher energy state is called excitation process.

5.1.2.2. *Energy transfer process*

Energy transfer process includes energy transfer and energy transport.

Energy transfer refers to the transfer of all or part of the excitation energy from one excitation center to another. Energy transport refers to the process of transferring excitation energy from one crystal to another by means of the motion of electrons, holes, excitons, etc. There are roughly four mechanisms of energy transfer and transport: reabsorption, resonance transfer, energy transport by carriers and energy transfer by excitons.

The decay rate of luminous intensity of characteristic luminescent materials is related to the intensity and temperature of excitation light. Luminescent materials based on silicate, phosphate, arsenate and germinate generally decay exponentially, but sometimes at the beginning of decay they decay exponentially, and then hyperbolic. The attenuation rate is related to temperature when the hyperbolic law is used to attenuate.

In the case of composite luminescent materials, the growth and decay processes follow other rules. In this case, the number of electrons compounded by ionization centers increases with the increase of the number of ionization centers. When the excitation

light is removed, the emission attenuation of the composite luminescent materials is very complex. The excited electrons leave the luminescent center and may recombine with an ionization center, or may be trapped. The “afterglow” is due to the possible heat release of electricity from the trap and recombination with the ionization center until all traps are exhausted.

The emission spectra of different luminescent centers are also significantly different due to their different energy level composition and transition properties. For example, rare earth ions, which are the main components of luminescent materials, also have various forms of luminescent properties. From the width of their emission spectra, they can be divided into linear emission and broadband emission. The former includes Eu^{3+} , Gd^{3+} , Tb^{3+} , Sm^{3+} , Dy^{3+} , Pr^{3+} , etc. The latter includes Ce^{3+} , Pr^{3+} , Nd^{3+} , Eu^{2+} , Sm^{2+} , Yb^{2+} , etc.

For linear emission rare earth ions, the common characteristic is that the emission transition comes from the electrons inside 4fN. Because 4f electrons are well shielded from the surrounding environment, the emission transition produces sharp lines in the spectrum. In the configuration coordinate diagram, these levels are parabolic, and the lifetime of the excited state is longer because the parity of the transition is unchanged.

5.1.2.3. *Luminescence process*

The process in which the stimulated system jumps from the excited state to the ground state and emits part of the energy absorbed during the excitation in the form of light radiation is called luminescence process.

5.1.3. *Non-radiative return to ground state*

When the absorbed light is stimulated, the system will return from the excited state to the ground state. However, the transition from the excited state to the ground state is not the only process. Another possibility is non-radiative return, i.e. non-luminous non-radiative return. Non-radiation process always competes with radiation process. Because the important requirement of luminescent materials

is higher light output, it is required that the radiation process in this material has higher probability than that in non-radiation process. The part of energy absorbed by materials that does not emit through radiation (luminescence) will dissipate in the lattice (non-radiation process), so the non-radiation process competing with the luminescence process must be suppressed. However, some non-radiative processes also promote high light output, that is, to ensure more effective excitation of luminescent activators and/or to promote the occupancy of luminescent levels.

The radiative return transition from the excited state to the ground state has been discussed previously, but in many materials this return process is non-radiative. This is because many luminescent centers do not emit light at all in some substrates. Non-radiation processes are ubiquitous in luminescent materials. It is difficult to find luminescent materials with energy efficiency close to 100%. For the same material, its luminescence phenomenon is not invariable. For example, some materials emit light at low temperature, but they do not emit light at room temperature. Some materials emit light at room temperature, but they do not emit light when they rise to a certain temperature, that is, the thermal quenching of light occurs.

5.2. Luminescence Principle of Phosphor

Phosphorescent powder is an inorganic powder material that can emit light under certain excitation conditions. It is a kind of luminescent material that can convert external energy into visible light. It is an important supporting material in the field of lighting and display. It is an extremely important material in today's life. The matrix absorption band absorbs energy and transfers it to the luminescent ion (luminescent center) to reach the excitation state, or the luminescent ion absorbs energy directly to the excitation state, and the ions in the excitation state emit photons in the process of returning to the ground state, i.e. absorb high-energy radiation, and then emit light. The energy of the emitted photons is lower than that of the excitation radiation. The high-energy radiation of phosphors

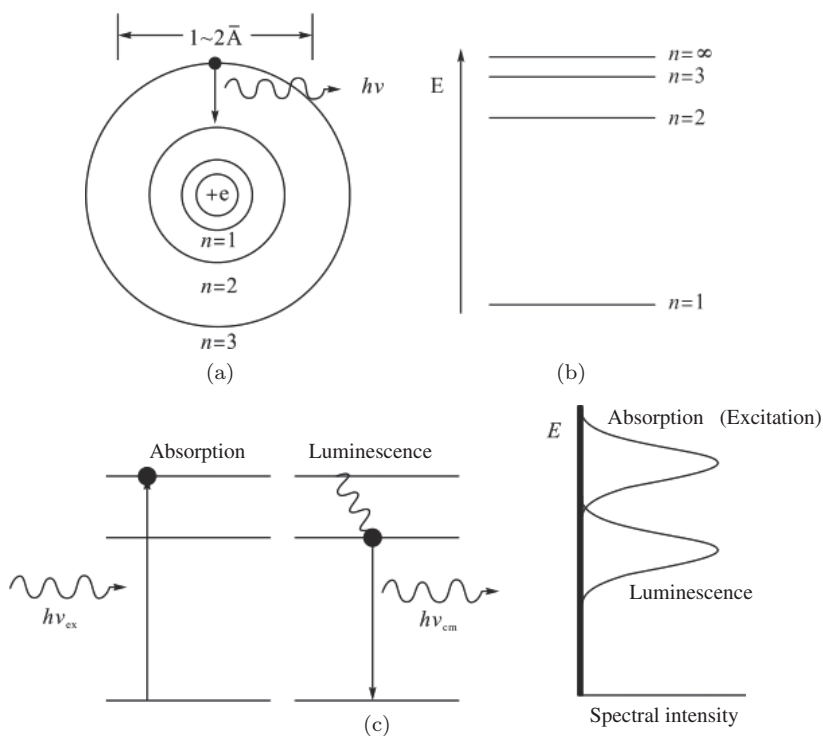


Fig. 5.3. Atom structure and light conversion: (a) schematic diagram of hydrogen atom; (b) energy level diagram of hydrogen atom; and (c) absorption (excitation), luminescence process and corresponding spectra.

can be electrons or high-speed ions, or photons from gamma rays to visible light.

According to quantum theory, there are many levels in an isolated single atom or ion. As shown in Fig. 5.3(a), when bound electrons in an atom or ion transit from a high level to a low level, they will form their own intrinsic luminescence. The following is an example of the simplest hydrogen atom. The hydrogen atom contains one electron and is called 1s, 2s, 3s, from the outside of the nucleus. The electron orbitals correspond to different energy levels. The electron of hydrogen atom is usually located in the innermost 1s orbital. The state of the electron is called ground state. If the electron is

stimulated by external energy such as electron collision or light, it will absorb the excitation energy and migrate to its outer orbit such as 2s orbit. The energy of the 2s orbital is higher than that of the 1s orbital. As shown in Fig. 5.3(b), this state of the electron is called the excited state. Atomic luminescence is produced when electrons return from the excited state to the ground state, as shown in Fig. 5.3(c).

5.2.1. *Basic concepts*

5.2.1.1. *Luminescence center*

Inorganic solid-state luminescent materials consist of two parts: one is the main component of materials, namely matrix; the other is a small amount of components intentionally incorporated, called activators. Luminescence centers, i.e. the specific centers of photons emitted by excited electrons in the luminescent body, transit back to the ground state (or recombine with holes). They can be ions, ionic groups or impurities incorporated into the matrix. Rare earth ions, transition metal ions (Ti, Cr, Mn, etc.) and some other heavy metal ions (Sb, Tl, Bi, etc.) and ion groups such as WO_4^{-2} can all be luminescent centers. If the excited electrons do not leave the center and return to the ground state, these centers are called discrete luminescence centers. If the electrons are ionized after excitation, and the holes emit light through specific centers, such centers are called composite luminescence centers.

5.2.1.2. *Emission spectrum*

The distribution of luminous energy according to wavelength or frequency is called emission (or luminescence) spectrum. There are usually two kinds of emission spectra of luminescent materials. One is the spectral band, i.e. the distribution of emission energy varies continuously in a certain wavelength range (hundreds or even thousands of angstrom); the other is the line spectrum, i.e. the spectrum consists of many lines of different intensity.

5.2.1.3. *Excitation spectra*

The excitation spectrum reflects the luminescence of substances under different wavelength light excitation. The higher the ordinate value, the stronger the luminescence, the higher the energy. The transverse axis represents the wavelength of the excitation light, and the longitudinal axis represents the efficiency or intensity of the excitation light.

5.2.1.4. *Absorption spectra*

Absorption spectrum is the variation of absorption coefficient $a(\nu)$ with frequency ν or wavelength λ . $a(\nu)$ is the main parameter in the following formula: $I(\nu) = I_0(\nu) \exp[-a(\nu)l]$, where $I_0(\nu)$ and $I(\nu)$ are the incident and transmitted light intensities respectively, and l is the thickness of the sample. Some of the incident light is reflected off the surface of the sample, and it is reflected again before it is emitted from another surface of the sample. The shape of absorption spectra of luminescent materials is similar to that of emission spectra, some are band spectra and some are line spectra.

5.2.1.5. *Luminescence attenuation*

Luminescence decay is one of the most important characteristics of luminescent phenomena. It is a key sign to distinguish between luminescent phenomena and other luminescent phenomena. The persistence of luminescence reflects the retention of matter in the excited state. The duration corresponds to the lifetime of the excited state.

5.2.1.6. *Luminescence efficiency*

Luminescent efficiency is another important parameter of luminescent materials and devices. Efficiency is expressed in terms of power efficiency or energy efficiency, quantum efficiency, photometric or lumen efficiency. Efficiency shows how much of the energy of the

excited light is still emitted into light, and how much heat is consumed by other means. So, I hope the higher it is, the better.

5.2.2. Characteristics of phosphors

5.2.2.1. Primary properties of phosphors (test performance)

Absorption spectra indicate the relationship between the absorption energy of phosphors and the wavelength of irradiated light. The absorption spectra of phosphors mainly depend on the matrix materials, and activators also play a role. The absorption peaks of most phosphors are in the ultraviolet region. Absorption spectra can only indicate the absorption characteristics of materials, but absorption does not mean that certain luminescence.

Excitation spectra indicate the change of luminous intensity with the excitation wavelength at a specific wavelength, and reflect the excitation effect of light at different wavelengths on the luminous materials. The wavelength range of the excitation light contributing to the emission can be determined by the excitation spectrum of the luminescent material.

Emission spectra indicate the relationship between luminous energy and wavelength of luminescent materials.

Quantum efficiency, the ratio of the number of photons emitted by phosphors to the number of exciting photons absorbed.

Luminescence efficiency, the ratio of luminous flux of phosphor to excitation energy.

Afterglow: Luminescence of phosphor after excitation stops.

Particle size: The particle size of phosphors must take into account both the process and the requirements of obtaining excellent luminescent properties. If the particle size is too large, the coating will be uneven, which will affect the luminous life of the lamp. If the particle size is too small, the reflection of ultraviolet radiation will increase, and the absorption of ultraviolet radiation will decrease, resulting in the decrease of the luminous efficiency of the lamp.

Temperature characteristics: The relationship between the emission characteristics of phosphors and temperature. Usually refers to the amount of change when the powder is heated to 120°C and kept at constant temperature for 10 minutes, including luminance, excitation wavelength, emission peak and color coordinates.

Color coordinates: In RGB three primary color system, the brightness of three primary colors is different. The ratio of brightness is R:G:B=1:4.5907:0.0601. In a trichrome system, the color stimulus of any color can be matched by an appropriate number of three primary color stimuli. The ratio of the stimulus of each primary color to the total stimulus of the three primary colors is called the chromaticity coordinate of the color, or the color coordinate for short.

5.2.2.2. *Quadratic properties of phosphors (performance)*

Dispersion, phosphor must have good dispersion, in order to get a uniform coating.

Stability, including thermal stability, chemical stability and ultraviolet radiation stability.

The light decay characteristic refers to the property that the light output of phosphor decreases with the ignition time.

5.2.2.3. *Factors determining the conversion efficiency of phosphors*

- (1) Chemical composition, i.e. formula. It should be clearly distinguished between the matrix and activator, as well as the main activator and the auxiliary activator. The matrix does not emit light, only the main activator and the auxiliary activator emit light.
- (2) Crystal structure and morphology. The morphologies of amorphous, flat, granular, square, spherical and degree of polymerization crystals are closely related to light conversion.
- (3) Synthesis process. The traditional methods include solid-phase synthesis, liquid-phase synthesis, solid-liquid-phases synthesis and spray method.

- (4) Particle size and distribution. Different synthesis processes have different particle sizes. The traditional solid-phase crushing process has different particle size distribution, wide particle distribution, poor luminescence characteristics, good consistency of particles in liquid-phase method and good luminescence characteristics.
- (5) Crystal emission spectra. It is a single line spectrum, or several line spectra, or a narrow or broad spectrum or broadband spectrum.
- (6) Luminescence intensity of crystal. Phosphors have high absorbance and luminescence intensity, and their absorbance depends on the crystal structure and morphology. Crystal structure is closely related to chemical composition and molar number. The luminescence intensity is closely related to the crystal morphology.
- (7) Light attenuation index. If the crystal morphology is irregular and the particle size is not uniform, the light scattering and diffuse reflection are serious, the light output efficiency is low, the heat accumulation is large, and the heat dissipation effect is not good, which will affect its life. When the particle size is reached by mechanical breaking method, even if the initial light flux is maintained at a higher rate, the light decay is large.
- (8) Optical conversion efficiency. In the process of YAG + chip, it belongs to the typical downward emission. In this device, the quantum efficiency is less than 1. How to make the quantum efficiency close to or equal to 1? The synthesis process plays a decisive role. The unit that synthesizes the light conversion efficiency of the light-emitting device is lm/W.

It can be seen from the above that the chemical composition determines the spectral wavelength and crystal structure, and the morphology determines the light intensity, light conversion efficiency and lifetime. The synthesis process affects the crystal morphology and structure, so these eight factors are interrelated and equally important.

5.3. Development History and Present Situation of Phosphors

5.3.1. *Development history of fluorescent powders for fluorescent lamps*

5.3.1.1. *First generation phosphors (1938–1948)*

The first fluorescent powders used for fluorescent lamps were CaWO_4 blue powder, zinc silicate (Zn_2SiO_4 : Mn) green powder activated by manganese ions and cadmium borate (CdB_2O_5 : Mn) red powder activated by manganese ions. At that time, the light efficiency of 40/W fluorescent lamp was 40lm/W.

Soon, beryllium zinc silicate ($(\text{Zn}, \text{Be})_2\text{SiO}_4$: Mn) phosphors were successfully developed and replaced zinc silicate and cadmium borate phosphors. The phosphor is also activated by divalent manganese ions. The luminescent color can vary between green and orange according to the different proportion of zinc and beryllium. In addition, calcium tungstate phosphor is also replaced by magnesium tungstate. The luminous flux of 40 W fluorescent lamp has risen to 2300 lm in 1948. As beryllium is a toxic substance, it is gradually replaced by calcium halophosphate phosphors.

5.3.1.2. *Second generation phosphors (1949–)*

In 1942, Mckeag *et al.* invented $3\text{Ca}_3(\text{PO}_4)_2 \cdot \text{Ca}(\text{F}, \text{Cl})_2$: Sb, Mn, abbreviated as halogen powder. It was popularized in 1948. Since this material is a single matrix, high luminous efficiency, adjustable light color, rich raw materials, low price, it has been the main phosphor for straight-tube fluorescent lamps since its practicality. There are two defects in the application of halogen powder in fluorescent lamps.

- (i) The lack of blue light below 450 nm and red light above 600 nm in the luminescence spectrum makes the CRI R_a value of the lamp low. The CRI R_a value can be increased by adding a certain proportion of blue and red powder, but the light efficiency of the lamp decreases obviously.

- (ii) Under the action of ultraviolet 185 nm, the color center is formed, which makes the light of the lamp fade greatly.

With the appearance of straight-tube fluorescent lamp and compact fluorescent lamp, the application of halogen powder in thin-tube fluorescent lamp has been limited, and the requirement of high-quality lighting source has not been met, so new fluorescent powder has been developed and studied.

5.3.1.3. *Rare earth phosphors*

The 1970s is the golden age for the development and research of rare earth phosphors. Many kinds of phosphors have been successfully developed and applied. The outer electronic structure of rare earth elements is $4f^{0-14}5d^{0-16}s^2$. The energy of 4f shell electrons is lower than that of 5d shell electrons and higher than that of 6s shell electrons. Rare earth ions in compounds usually lose two 6s electrons and one 4f electrons in a trivalent state. Because rare earth ions contain special 4f electron configurations, when they are excited, 4f electrons can generate excitation transitions between different energy levels. When they are de-excited, the excited electrons migrate to different energy levels and return to the original 4f electron configurations, resulting in the luminescence spectra, i.e. the mutual transitions between 4f–4f and 4f–5d. The unique electronic structure of rare earth elements determines that they have special luminescent properties. Therefore, rare earth phosphors have the following advantages.

- (1) Compared with the general elements, the 4f electron layer configuration of rare earth elements makes their compounds have a variety of fluorescence properties. Except for $3Sc^{3+}$, Y^{3+} without 4f sublayer, the 4f sublayer of La^{3+} and Lu^{3+} is hollow or full. The 4f electrons of other rare earth elements can be arbitrarily distributed among seven 4f orbitals, thus generating abundant electronic energy levels, which can absorb or emit electromagnetic radiation of various wavelengths from ultraviolet, visible

to near infrared region, thus making the rare earth luminescent materials present rich and varied fluorescence characteristics.

- (2) Because the 4f electron of rare earth element is in the inner orbit and shielded effectively by the outer S and P orbits, it is difficult to be disturbed by the external environment. The 4f energy level difference is very small. The f-f transition shows a sharp linear spectrum, and the color purity of luminescence is high.
- (3) Fluorescence lifetime spans six orders of magnitude from nanosecond to millisecond. Long-life excited states are one of its important characteristics. The average lifetime of excited states of atoms or ions is 10^{-10} – 10^{-8} s, while some of the rare earth elements have an average lifetime of 10^{-6} – 10^{-2} s, which is mainly caused by the low spontaneous transition probability between 4f electron levels.
- (4) It has strong ability to absorb excitation energy and high conversion efficiency.
- (5) The physical and chemical properties are stable and can withstand the effects of high-power electron beam, high-energy radiation and strong ultraviolet radiation.

5.3.2. *Rare earth tribasic phosphor*

Rare earth trichrome phosphors consist of red, green and blue phosphors activated by rare earth ions. They are mainly aluminate, phosphate and borate. Among them, aluminate technology is quite mature, most of which are aluminate Series in China, with the largest production and sales. In 1974, the Dutch scientist Verstegen first invented the rare earth tricolor fluorescent lamp by using the blue material $\text{BaMg}_2\text{Al}_{16}\text{O}_{27}:\text{Eu}$, $\lambda_{\text{max}} = 450\text{ nm}$, green material $\text{MgAl}_{11}\text{O}_9:\text{Ce}$, Tb , $\lambda_{\text{max}} = 545\text{ nm}$, and the existing red material $\text{Y}_2\text{O}_3:\text{Eu}$, $\lambda_{\text{max}} = 611\text{ nm}$. To replace the incandescent lamp with the problems that cannot be solved by halogen powder, reduce energy consumption by 3/4, and realize the revolutionary progress of lighting source. The compact phosphor made of halogen powder has very serious light decay due to its small size, thin diameter and much higher ultraviolet irradiation energy than the ordinary straight tube

fluorescent lamp. The rare earth tricolor phosphor has the advantages of good ultraviolet radiation stability and high thermal quenching temperature, which makes the compact fluorescent lamp possible. Yes.

- (1) The characteristics of rare earth tricolor phosphors are as follows: narrow luminous band, more concentrated luminous energy, high stability under short-wave ultraviolet excitation, good high-temperature characteristics, and more suitable for high-load tubular fluorescent lamps and various single-ended compact fluorescent lamps.
- (2) The shortcomings of rare earth tricolor phosphors are as follows: Rare earth raw materials are expensive, resulting in high cost of tricolor lamps, which limits the development of tricolor lamps. Reducing the diameter of the tube or adopting new coating technology to reduce the amount of tricolor and replacing one or two rare earth tricolor with other cheap color powders can also produce fluorescent lamps with high luminous efficiency and high color rendering, but the light fading may be a little larger.

At present, among the rare earth tricolor phosphors, the red powder passes through basically. The main problem is the high cost. Efforts are still needed in terms of consistency and particle size distribution. The initial luminous flux of green powder is very high, but after the lamp is ignited for 2000 hours, the luminance decreases. The most prominent problem of blue powder is its high light decay, which reaches 5% after lamp making, leading to color drift after lamp making. In addition, green powder has the highest light efficiency, followed by red powder and blue powder, which is about 1/5 of green powder. Red powder is mainly used to reduce color temperature, green powder is used to increase light efficiency and brightness, blue powder is used to improve color rendering index.

At present, $\text{Y}_2\text{O}_3:\text{Eu}^{3+}$ red powder is the only red powder that has reached the practical level. Its performance is still unmatched so far. It is a perfect luminescent material for red lamps. Its density is 5.1 g/cm^3 , its chemical properties are stable, and it is insoluble in water, weak acid and weak base. Under the excitation of 254 nm

ultraviolet radiation, the emission peak is 611 nm, and the color coordinates are $x = 0.650$ and $y = 0.345$. $\text{Y}_2\text{O}_3:\text{Eu}^{3+}$ pink is expensive, and the cost of raw materials exceeds the sum of blue and green. The key to reducing the price of three basic pinks is to reduce the cost of pink. Researchers tried to replace yttrium oxide (Y_2O_3) with Y_2O_3 , $\text{aSiO}_2:\text{Eu}^{3+}$ and Y_2O_3 , $\text{bAl}_2\text{O}_3:\text{Eu}^{3+}$. The brightness, spectral characteristics and chroma coordinates were similar to $\text{Y}_2\text{O}_3:\text{Eu}^{3+}$ and the cost of raw materials could be reduced by about 15%. Figures 5.4 and 5.5 are $\text{Y}_2\text{O}_3:\text{Eu}^{3+}$ excitation spectra, diffuse reflectance spectra and chroma maps, respectively.

The effect of blue powder is mainly to improve the light efficiency and color rendering. The emission wavelength and spectral power distribution of blue powder have a great influence on the light efficiency, color temperature, light decay and color rendering of fluorescent lamps. The three basic blue luminescent materials developed now are $\text{Sr}(\text{PO}_4)_6\text{C}_{12}:\text{Eu}^{2+}$; $\text{Sr}_4\text{Al}_{14}\text{O}_{25}:\text{Eu}^{2+}$; $(\text{CaSrBa})_{10}(\text{PO}_4)_6\text{C}_{12} \cdot n\text{B}$; $\text{BAM}:\text{BaMgAl}_{10}\text{O}_{17}:\text{Eu}^{2+}$ (single peak), $\text{BaMgAl}_{10}\text{O}_{17}:\text{Eu}$, Mn (double peak); $\text{SCA}:(\text{SrBaMgCa})_5(\text{PO}_4)_3\text{Cl}:\text{Eu}^{2+}$.

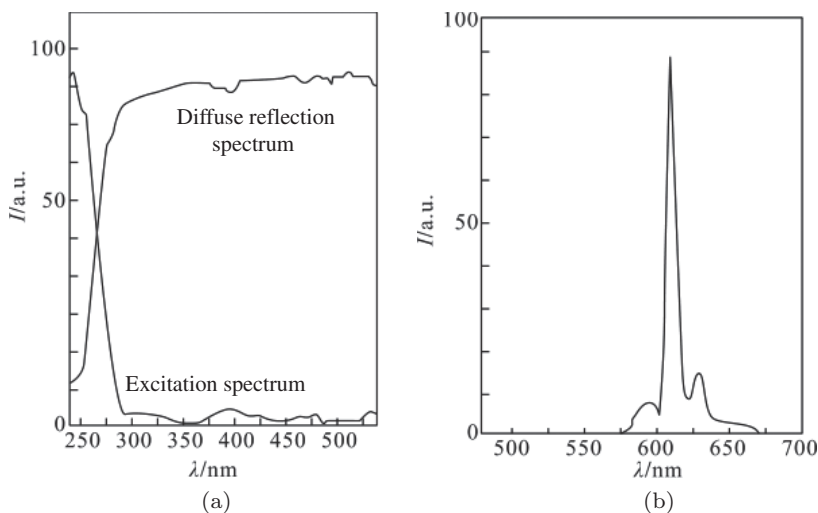
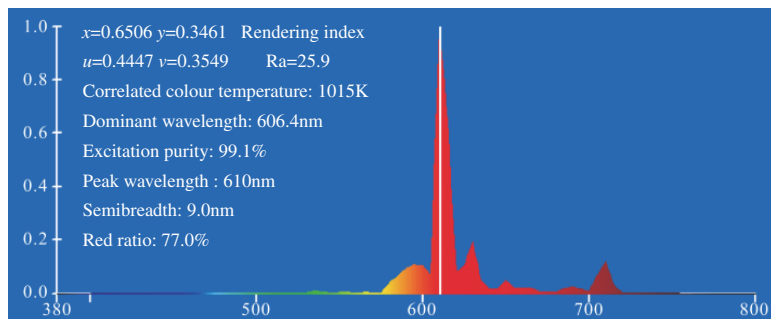
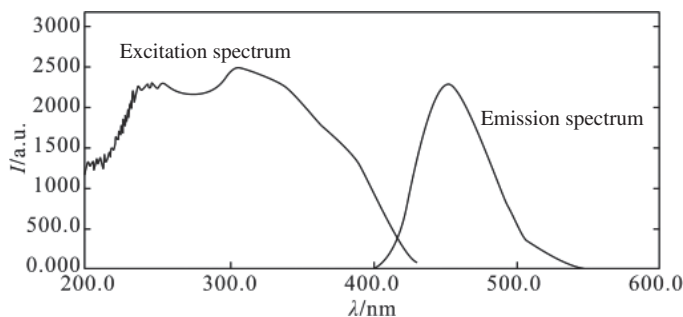


Fig. 5.4. $\text{Y}_2\text{O}_3:\text{Eu}^{3+}$ excitation spectrum and diffuse reflectance spectrum. (a) Excitation spectra of $\text{Y}_2\text{O}_3:\text{Eu}^{3+}$ (b) $\text{Y}_2\text{O}_3:\text{Eu}^{3+}$ diffuse reflectance spectra.

Fig. 5.5. $\text{Y}_2\text{O}_3:\text{Eu}^{3+}$ chroma diagram.Fig. 5.6. Excitation and emission spectra of $\text{BaMgAl}_{10}\text{O}_{17}:\text{Eu}^{2+}$.

Among them, BAM belongs to hexagonal crystal system, white crystal, the most mature application, the most extensive field. Eu^{2+} replaces Ba in the mirror layer and becomes a luminescent center. It absorbs 254 nm ultraviolet radiation and emits 450 nm blue light. It is a broad-band luminescent material with a half-height and a half-width of 50 nm. Quantum efficiency is about 95%, chemical properties are stable and temperature quenching characteristics are good. Figures 5.6 and 5.7 are excitation spectra, diffuse reflectance spectra and chroma maps of $\text{BaMgAl}_{10}\text{O}_{17}:\text{Eu}^{2+}$.

The bimodal BAM emits 450 nm blue light and 515 nm blue-green light under 254 nm ultraviolet excitation. BAM was used as the matrix and Eu^{2+} as the luminescent center and sensitizer. Mn^{2+} is also the luminescent center, replacing Mg^{2+} , and its emission peak

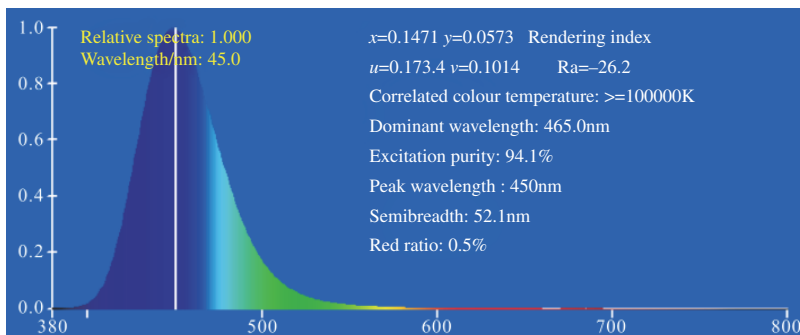


Fig. 5.7. Chromaticity diagram of BaMgAl₁₀O₁₇:Eu²⁺.

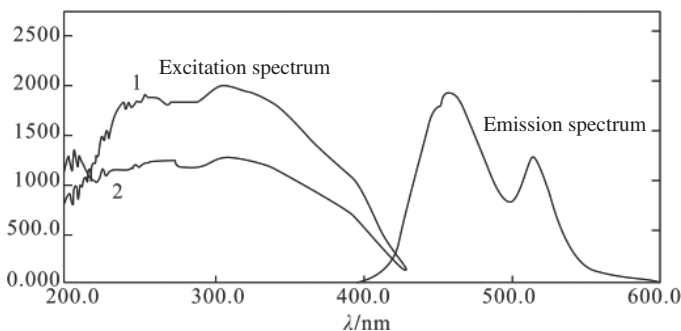


Fig. 5.8. Excitation and emission spectra of BaMgAl₁₀O₁₇:Eu²⁺, Mn²⁺.

is 515 nm. Most of the Eu²⁺ transitions produce blue light, while a few Eu²⁺ transitions transmit to Mn²⁺, and then the Mn²⁺ transitions emit blue-green light. When the concentration of Mn²⁺ increases, the green light increases and the blue light of Eu²⁺ decreases. Bimodal blue powder can improve the color rendering index at the expense of certain brightness. Figures 5.8 and 5.9 are the excitation spectra, diffuse reflectance spectra and chroma maps of BaMgAl₁₀O₁₇:Eu²⁺, Mn²⁺.

Among the three primary phosphors, green powder contributes most to the luminous flux of the lamp. Tb³⁺ is used as activator and Ce³⁺ as sensitizer in most of the three-color green powders. The maximum emission peak of Tb³⁺ is located at 545 nm, which belongs

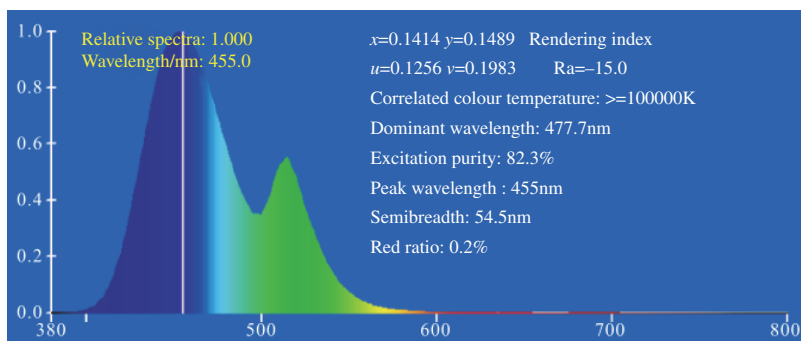


Fig. 5.9. Chromaticity diagram of $\text{BaMgAl}_{10}\text{O}_{17}:\text{Eu}^{2+}, \text{Mn}^{2+}$.

to the $^5\text{D}_4-^7\text{F}_5$ transition of Tb^{3+} . Ce^{3+} has strong absorption near 254 nm and strong emission in the long-wave ultraviolet region of 330–400 nm. Ce^{3+} can transfer energy to Tb^{3+} effectively through non-radiation energy transfer. At present, the commonly used three-color green powders are $\text{MgAl}_{11}\text{O}_{19}:\text{Ce}^{3+}, \text{Tb}^{3+}$ (abbreviated as CAT), emission peak 543 nm, color coordinate $x = 0.335, y = 0.595$; $\text{LaPO}_4:\text{Ce}^{3+}, \text{Tb}^{3+}$ (abbreviated as LAP), emission peak 543 nm, color coordinate $x = 0.360, y = 0.574$; $\text{Zn}_2\text{SiO}_4:\text{Mn}$ (abbreviated as ZSM), $\text{GdMgB}_5\text{O}_{10}:\text{Ce}^{3+}, \text{Tb}^{3+}$, emission peak 525 nm, color coordinate $x = 0.251, y = 0.698$. Figures 5.10 and 5.11 are excitation spectra, diffuse reflectance spectra, emission spectra and chroma maps of $\text{MgAl}_{11}\text{O}_{19}:\text{Ce}^{3+}, \text{Tb}^{3+}$.

LAP is a kind of green luminescent material with high efficiency, which is developed by Japan and widely used in Japan, the United States and the Soviet Union. LAP belongs to monoclinic system with finer crystal particles than CAT. The luminous color of LAP is yellowish, and the x value in color coordinates is large. It is beneficial to save expensive red powder when forming three-base color powder; the quantum efficiency is 3% higher than that of CAT; the synthesis temperature of LAP is low; the relative density and particle size of LAP and red powder and blue powder can be matched reasonably, so the comprehensive performance of LAP after lamp making is better than that of CAT, which tends to replace CAT. Figure 5.12 shows the excitation and diffuse reflectance spectra of LAP.

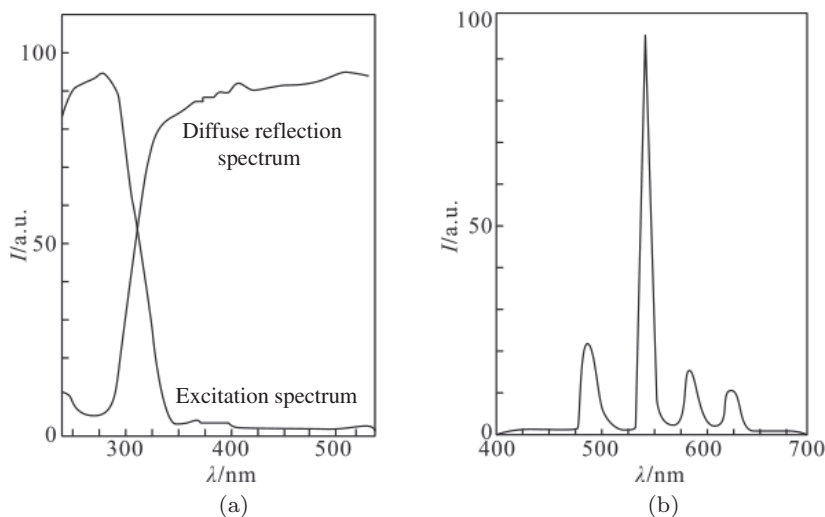


Fig. 5.10. Excitation spectra, diffuse reflectance spectra (a) and emission spectra (b) of $\text{MgAl}_{11}\text{O}_{19}:\text{Ce}^{3+}, \text{Tb}^{3+}$.

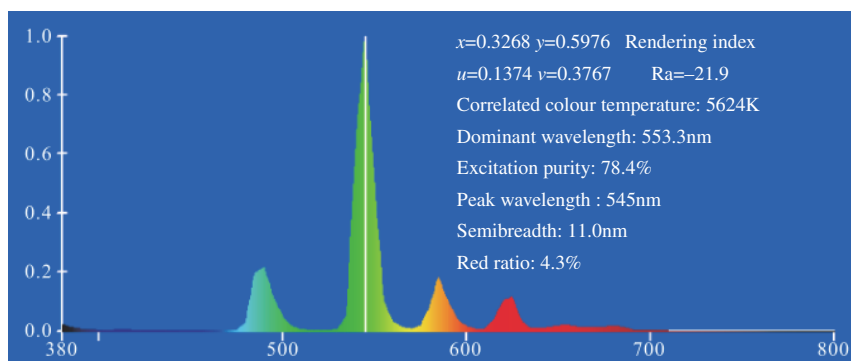


Fig. 5.11. Chromaticity diagrams of $\text{MgAl}_{11}\text{O}_{19}:\text{Ce}^{3+}, \text{Tb}^{3+}$.

The biggest obstacle in the application of LAP is that the temperature quenching is very serious, and the brightness at 200°C is only $1/2$ of that at 20°C . As powder for lamp, because of its small diameter and high wall temperature, the flue-cured tube temperature in lamp making process is as high as 550°C , so serious temperature

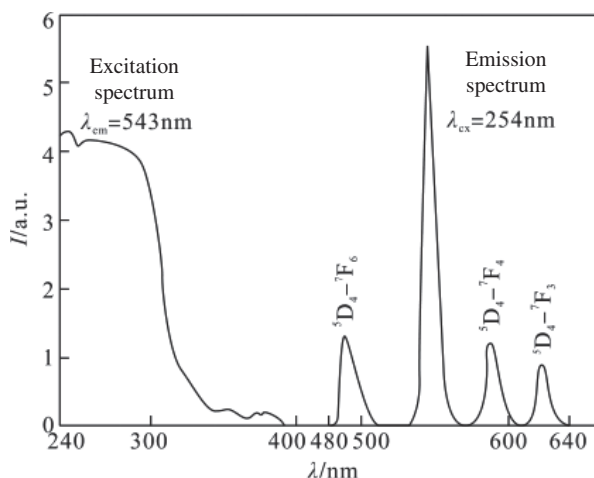


Fig. 5.12. Excitation and diffuse reflectance spectra of $\text{LaPO}_4:\text{Ce}^{3+}, \text{Tb}^{3+}$.

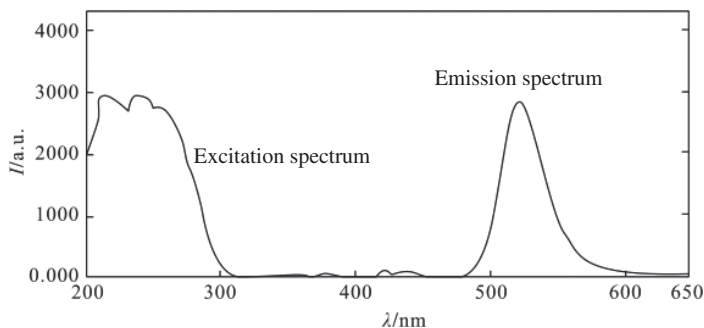


Fig. 5.13. Excitation and emission spectra of $\text{Zn}_2\text{SiO}_4:\text{Mn}$.

quenching effect must be overcome. However, due to the limitation of process and production cost, the amount of LAP is limited in China.

ZSM is an excellent luminescent material. Mn occupies the position of Zn and emits green light. It has been widely used in photoluminescence and cathode ray luminescence. High luminance, good color purity, no near infrared emission, good chemical stability, strong environmental adaptability, good moisture resistance, easy preparation of matrix, low price.

5.3.3. White LED phosphor

White LED is recognized as a new light source in the 21st century. It is the fourth generation light source after incandescent lamp, fluorescent lamp and high intensity gas discharge lamp. LED has many advantages, such as energy saving, environmental protection (shock resistance, non-breaking, recyclable waste), small size, all-solid state, high luminous efficiency, low calorific value, low energy consumption (only 1/8 of incandescent lamp), low voltage and low current start-up, long life, fast reaction speed, planar packaging, easy development into light and small products, etc. It has been widely used in urban landscape lighting, LCD display. Display backlight, indoor and outdoor general lighting and other lighting fields, is a new generation of green lighting source to replace incandescent lamp and fluorescent lamp.

As early as 1976, Blasse, a Dutch scientist, studied that cerium-activated yttrium aluminium garnet (YAG) could emit yellow light. Although the luminance was weak at that time, it was widely used in flying-point scanners. This is the embryonic form of YAG: Ce^{3+} . Twenty years later, in 1996, Japanese scientist Nishida Nakamura invented GaN blue wafer, and found that YAG yellow phosphor on the blue chip could emit false white light. So Niya applied for GaN invention patent and LED packaging patent, patent number US5998925. Due to the blockade of patents and the lack of authorization from Japan and Asia, the technology has been confined for eight years, and there are many complaints from all over the world. Scientists in Europe, America and Japan have taken a new path to study new products into Baiguang in order to circumvent this patent. Osram has developed $(\text{Tb}_{1-x-y}\text{Re}_x\text{Ce}_y)_3(\text{Al}, \text{Ga})_5\text{O}_{12}$, or TAG for short, which is still the crystal structure of yttrium aluminium garnet. It only adds Tb to obtain a new yellow phosphor different from YAG, and has applied for a patent for white LED, the patent number is US6669866.

As a light conversion material, phosphor plays an important role. It directly affects the luminous efficiency, service life, color rendering index, color temperature and other major indicators of

white LED products. With the breakthrough of LED chip technology, the luminous efficiency of LED will gradually approach its theoretical luminous efficiency. The performance of phosphor will directly determine the product performance of LED light source. At present, there are not many phosphors that can match blue light, near ultraviolet or other chips. It is necessary to develop phosphors with high luminescence efficiency, long service life, high color rendering index, more stable physical and chemical properties and simpler preparation process.

5.3.3.1. *Problems of LED phosphors*

(1) Problems of phosphors matched with blue light excitation:

Taking YAG:Ce^{3+} as an example, YAG:Ce^{3+} has high physical and chemical stability and high quantum efficiency under blue light excitation, which has been commercialized. But there are mainly the following problems.

The emission spectrum of InGaN emitted by YAG:Ce^{3+} blue light is low because YAG:Ce^{3+} does not emit red light. Red phosphor should be added in order to improve its color rendering index.

YAG:Ce^{3+} has poor temperature characteristics, and the emission quantum efficiency decreases with the increase of temperature.

YAG:Ce^{3+} is synthesized by high-temperature solid-state method with high temperature (1500–1600°C). Other soft chemical methods are not yet mature. The most efficient red phosphors are nitrogen compounds and sulfides activated by Ce^{3+} or Eu^{2+} ions. Nitrogen-containing compounds activated by Ce^{3+} or Eu^{2+} have high photoemission quantum efficiency, but the synthesis temperature is high and the synthesis conditions are harsh. A simple synthesis method of Ce^{3+} or Eu^{2+} activated sulfides usually uses toxic H_2S as raw material, and its instability affects the lifetime of LED.

(2) Problems of phosphors matched with ultraviolet and near-ultraviolet excitation:

The blue and red phosphors in the tricolor phosphors used for lamp matching with ultraviolet excitation have high quantum efficiency and stability under short-wave ultraviolet excitation, but the

quantum efficiency under near-ultraviolet excitation (370–405 nm) is very low.

Ultraviolet and near-ultraviolet excitation matches less green and yellow phosphors, and when assembled into white LED, there is a problem of blue light reabsorption.

5.3.3.2. *Development trend and application requirements of white LED phosphors*

The application of white LED phosphors requires the widest excitation band from ultraviolet to blue, high quantum efficiency (>90%) and strong absorption in the excitation band, and smaller peak width in the appropriate emission band. Therefore, we focus on the following four aspects to improve the existing phosphors for white LED.

- (1) The synthesis of Phosphors by soft chemical method and microwave method was studied. Using soft chemical method and microwave method to prepare phosphors can obtain phosphors with good secondary characteristics, thus improving the light efficiency of phosphors.
- (2) On the basis of the original phosphor for lamp, the crystal properties of the phosphor are changed by doping rare earth and alkaline earth metals, so as to change the excitation spectrum of the phosphor, achieve the best match with the current LED, and improve the light efficiency of white LED.
- (3) Find red, green and blue phosphors that match the emission spectrum of LED well, and the emission spectrum is narrow band spectrum, in order to improve the color rendering index.
- (4) The stability of phosphors can be improved by subsequent treatment of the powder, such as coating.

5.4. Main Preparation Methods of Phosphors

The properties of the materials are mainly determined by the chemical composition and microstructure of the materials. Therefore,

the chemical composition and preparation technology of the powders become the important factors to determine the luminescence efficiency of the phosphors. At present, the preparation methods of phosphors are high-temperature solid-phase method, combustion synthesis method, sol-gel method, solvothermal method, chemical coprecipitation method, spray pyrolysis method, etc. The various methods of synthesizing inorganic powder luminescent materials will be introduced one by one below.

5.4.1. *High-temperature solid-phase method*

High-temperature solid-state reaction (also known as high-temperature solid-state reaction) is a traditional synthesis method of luminescent materials. Solid-state reaction usually depends on the crystal structure and defect structure of materials, not only the inherent reactivity of components. Each mass transfer phenomenon and reaction process occurring in solid materials are related to various defects in the lattice. Generally, the more defects in solid phase, the stronger the corresponding mass transfer ability, and the higher the solid-phase reaction rate related to mass transfer ability. The necessary and sufficient condition for solid-state reaction is that the reactants must contact each other, that is, the reaction is carried out through the particle interface. The finer the reactant particles are, the larger the specific surface area is, and the larger the contact area between reactant particles is, which is conducive to the solid-phase reaction. Therefore, grinding and fully mixing the reactants can increase the contact area between the reactants and make the diffusion and transport of atoms or ions easier, so as to increase the reaction rate. In addition, some external factors, such as temperature, pressure, additives, radiation and so on, are also important factors affecting the solid-phase reaction. Solid-state reactions usually include solid-state interface diffusion, atomic-scale chemical reactions, nucleation of new phases, transport of solid phases and growth of new phases.

Nucleation and diffusion rate are two important factors determining solid state reaction. If there is structural similarity between

products and reactants, nucleation is easy. In high-temperature solid-state reaction, it is often necessary to control the reaction atmosphere. Some reactants will produce different products in different reaction atmosphere. Especially for the reaction containing variable valence ions, if we want to obtain a desired product, we must control the reaction atmosphere.

The preparation of luminescent materials by solid-state reaction mainly involves two processes: batching and calcination. The main role of calcination process is to make the chemical reaction between the components of raw materials, form a matrix with a certain lattice structure, and make the activator enter the matrix, stay in the gap of the matrix lattice or displace the lattice atoms. Obviously, calcination is the key step to form the luminescence center. Calcination conditions (temperature, atmosphere, time, etc.) directly affect the luminescent properties.

The calcination temperature mainly depends on the properties of matrix, melting point, diffusion rate and crystallization ability of components. The lower the diffusion rate and crystallization ability between components, the higher the temperature required. Generally, 2/3 of the highest melting point in the matrix component is suitable, but the choice of flux also has an effect. The optimum temperature should be determined by experiments. The calcination temperature of luminescent materials is generally 800–1400°C.

Flux plays an important role in the calcination process. It is a kind of material that plays a role of flux in the process of calcination of luminescent materials. It makes the activator easy to enter the matrix and promotes the formation of micro-crystals in the matrix. Commonly used flux materials are halides, alkali metals and alkaline earth metals salts, boron oxides and salts, the amount of which is 5–25% of the matrix. The type, content and purity of flux have a direct impact on the luminescent properties.

The ambient atmosphere around the burden has a great influence on the luminescent properties during calcination. Usually it is necessary to prevent metal vapor from poisoning the luminescent body and oxygen from oxidizing the material in the air. The ambient atmosphere has a direct impact on the brightness and color of

the phosphor. For example, Cadmium halophosphate lamp powder activated by Sb and Mn can be calcined in oxidizing atmosphere to obtain green material, while in nitrogen atmosphere to obtain orange material.

The phosphor prepared by solid-state reaction usually needs post-treatment, including grinding, powder selection, powder washing, coating and screening. These links often directly affect the secondary properties of phosphors, such as coating properties and anti-aging properties. Powder washing methods include water washing, acid washing and alkali washing. The purpose is to remove flux, excessive activator and other impurities. Because the flux is mostly alkali metal or alkaline earth metal salts, these metal ions have poor ability to resist ion and electron bombardment and ultraviolet radiation, and remain in the luminescent body to make the phosphor black and deteriorate, and shorten its life. For example, the removal of excess VO_4^{3-} in $YVO_4:Eu^{3+}$ by NaOH can improve brightness.

The main advantages of synthesizing luminescent materials by solid-state reaction are as follows: good crystal quality, less surface defects, high luminous brightness and long afterglow time, which are conducive to industrial production. At present, all kinds of luminescent materials and lighting display devices based on luminescent materials are flooding the market, and these luminescent materials still use the traditional high-temperature solid-state reaction method in production. However, the disadvantage of solid-state reaction method is that the calcination temperature is high, the holding time is longer (more than 2 hours), and the equipment requirements are higher; the particle size distribution is not uniform, it is difficult to obtain spherical particles, which are easy to agglomerate and need to be smashed to reduce the particle size, so that the crystal shape of the luminescent body is destroyed, the crystallinity of the phosphor is reduced, and the luminescent properties are reduced, at the same time, the particle morphology is incomplete and as a result, the coating is not uniform and the compactness is poor, which is not conducive to the acquisition of high quality phosphors or fluorescent display products.

By adding flux, such as B_2O_3 , H_3BO_3 , or B and other metal halides such as Sr, Ba, Ca, Al, etc., the calcination temperature has been significantly reduced, and some of them can be reduced to $1200^\circ C$. But with the addition of flux, fluorescent powder is easy to be calcined into hard lumps. Only by crushing and sieving, can the available fine powder be obtained. In this process, the performance of fluorescent powder decreases, and it must be calcined twice at high temperature, so that the performance can be improved. In addition, the solid-state reaction method has the advantages of high labor intensity, long production cycle and high cost of expensive rare earth materials. Therefore, it has become a hot topic to find new synthetic pathways for phosphors. Through these years' efforts, many new methods have been developed, such as sol-gel method, precipitation method, polymer network gel method, hydrothermal method, microwave method, combustion synthesis method, LHPG (laser heating substrate growth method) and so on. These new synthetic methods have their own characteristics. Compared with the traditional high-temperature solid-state method, the calcination temperature is relatively lower and the finished product particles are finer. With its mild reaction conditions and flexible and diverse operation methods, they have shown great potential in the preparation of multi-functional optical materials. The following will introduce various new preparation methods.

5.4.2. *Combustion synthesis method*

Combustion synthesis (also known as olefin burning) is a method of synthesizing materials by combustion of precursors. It was first developed by Soviet experts and named as self-propagating high-temperature synthesis. It is a method of preparing inorganic compounds with high-temperature resistance. The process is as follows: when the reactant reaches the ignition temperature of exothermic reaction, it is ignited in some way, and the heat released from the combustion of raw materials keeps the system at high temperature. The synthesis process continues. The combustion product is the

material prepared. Chemical reactions occurring during combustion include combustion of solutions and decomposition of materials.

The products made by traditional methods have great influence on the secondary characteristics of phosphors after lamp making. Combustion method is another new method invented on the basis of some shortcomings. The phosphor prepared by this method can effectively absorb blue and violet light, and the product has obvious advantages. For example, $4\text{SrO} \cdot 7\text{Al}_2\text{O}_3 \cdot \text{Eu}^{2+}$ luminescent materials have been successfully synthesized by combustion method. The main preparation process is that the reactants are stoichiometrically mixed with water and appropriate amount of urea. The samples are heated to be dissolved and then burned in an electric furnace. The reaction is completed immediately. In addition, $\text{MgCeAl}_{11}\text{O}_{18}$ luminescent materials were successfully synthesized.

The process of combustion synthesis generally has two forms: self-propagating and ignition-detonation. The former starts at one end of the reactant and propagates automatically to the other end at a speed of 0.1–25 cm/s until the reactant is exhausted. Burning and detonation are heated uniformly under control, and then the whole reactant reacts in an instant, releasing high temperature, forming an explosion. The combustion synthesis of phosphors belongs to the ignition–explosion type. The furnace charge used in combustion synthesis is generally composed of oxidant and reductant. When synthesizing phosphors, the oxidants used are usually cationic nitrates, perchlorates and other highly purified compounds, which constitute the chemical components of the products. They require high water solubility to adapt to the solution ingredients. Reducing agents are mostly organic compounds, requiring simple structure, composition and low carbon content, so as not to pollute the product by residual carbon after burning, and the reaction is mild at high temperature, the released gas does not cause public hazards, and is easily soluble in water. They have strong complexing power to metal cations in aqueous solution, so as to avoid the precipitation of some component crystals in the middle of combustion volatilization and destroy the overall uniformity.

Compared with the traditional high-temperature solid-state method, combustion method has the advantages of simple preparation process, rapid heating, small product particles, uniform particle size distribution, high purity, light-emitting brightness is not easy to be destroyed, energy saving and cost saving. However, the reaction process is intense and difficult to control, and it is not easy for large-scale industrial production.

The combustion method has many advantages, such as relying entirely on exothermic reaction and spontaneous high temperature to complete the reaction in an instant; the combustion gas can protect Ce^{3+} and Eu^{2+} rare earth ions from oxidation, thus eliminating the need for reductive protective atmosphere; the furnace temperature can be greatly reduced (about 500°C), the synthesis time is short (about 10 min), the product quality is uniform and the process is simple; the products prepared are single phase, high purity, and small size. So it is no need for additional ball milling. And it is time-saving, labor-saving, energy-saving, pollution-free, less material loss, and low cost. However, the density of the primary products synthesized by combustion method is small and the specific surface area is large, so the luminescence intensity is affected to a certain extent. It needs a short time of post-treatment at a certain temperature to achieve the desired level. The reason may be that the combustion time is too short. If new fuels can be developed, the combustion time can be prolonged at the current temperature level, so that both the higher reaction temperature and sufficient reaction process can be achieved, or the unique problems of combustion synthesis of luminescent materials can be solved, which will be the direction of future efforts.

5.4.3. *Solvent (hydrothermal) method*

Solvent (hydrothermal) synthesis (also known as solvent (hydrothermal method) refers to the synthesis by chemical reaction of substances in water or solvent at a certain temperature ($100\text{--}1000^{\circ}\text{C}$) and pressure ($1\text{--}100\text{ MPa}$).

Hydrothermal method is also a new synthetic method invented in recent years to study inorganic luminescent materials. This method

is mainly a kind of inorganic preparation method which makes the substance react in solution under certain temperature and pressure. In recent years, the typical composite fluoride is a kind of inorganic functional material, which has good optical, electrical, magnetic and thermal properties, and can realize multivalent ion doping. These characteristics provide favorable conditions for exploring new materials. Hydrothermal method plays a very important role in the synthesis of composite fluoride. Composite fluorides synthesized by hydrothermal method include KMgF_3 , BaBeF_4 , BaY_2F_8 , KYF_4 and other products. In addition, precursors of $\text{BaMgAl}_{10}\text{O}_{17}:\text{Eu}^{2+}$ were synthesized. The hydrothermal method is mainly to dissolve the weighted reaction mixture, heat it to $60\text{--}70^\circ\text{C}$, add ammonia water to form colloidal precipitation, wash acid radical ions with distilled water, heat and concentrate the suspended matter with precipitation, then transfer it into the autoclave, in the constant temperature box at 240°C for several hours, and steam the sample into the evaporating dish. The precursor is then converted into a crucible and calcined at a certain temperature to obtain the required fluorescent materials. The advantages of hydrothermal method lie in the following four points.

- (1) Low and medium temperature liquid-phase control, low-energy consumption, wide applicability.
- (2) The raw material is relatively cheap and easy to obtain. The reaction is carried out in the rapid convection of liquid phase with high yield, homogeneous phase and high purity.
- (3) The process is simple and the powder with good crystallinity and narrow particle size distribution can be obtained directly without high-temperature calcination treatment, and the product has good dispersion. The reaction temperature, pressure, treatment time and pH, the type and concentration of precursors used in hydrothermal process have great influence on the reaction rate, crystal shape, particle size and morphology of the product. The cutting of product properties can be achieved by controlling the above experimental parameters.

- (4) The synthesis reaction is always carried out in closed conditions, which can control the atmosphere to form appropriate redox reaction conditions, and realize the formation and crystallization of some phases which are difficult to obtain by other means.

But there are obvious shortcomings: it cannot be used in the reaction of water-sensitive compounds, high production costs, organic solvents are not easy to remove, and it pollutes the environment.

5.4.4. *Sol-gel method*

Sol-gel method is a new wet chemical synthesis method. The preparation of rare earth luminescent materials by this method has made great progress in the past decade. The luminescent materials synthesized by sol-gel method can obtain smaller particle size without grinding, and the synthesis temperature is lower than that of traditional synthesis methods. Therefore, this method has considerable potential in the synthesis of luminescent materials, and is one of the important methods for the synthesis of nano-luminescent materials.

In 1984, Morlotti prepared $\text{Mg}_2\text{SiO}_4:\text{Sm}^{3+}$ with $\text{Mg}(\text{NO}_3)_2$, $\text{Sm}(\text{NO}_3)_2$ and $\text{Si}(\text{OC}_2\text{H}_5)_4$ as raw materials by sol-gel method, but only discussed the solubility of Sm^{3+} in the matrix and did not discuss its luminescent properties. The sol-gel method was used to prepare luminescent materials in 1987. At the same time, Rabinovich and other rare earth nitrate $\text{Y}(\text{NO}_3)_3$ and $\text{Si}(\text{OC}_2\text{H}_5)_4$ were used as raw materials to produce $\text{Y}_2\text{SiO}_5:\text{Tb}^{3+}$ cathodoluminescence thin-films on the quartz glass substrate. In 1990s, sol-gel began to go to its prosperity stage, which attracted the attention of the scientific community. It has been widely applied. It also shows its striking superiority in the field of luminescent materials synthesis. New or improved luminescent materials prepared by this method have been successfully applied to optical devices. Sol-gel method has many advantages, such as high purity, uniform chemical composition, low synthesis temperature, controllable particle size and uniform particle size. Therefore, it is of great theoretical significance and application prospect to develop new rare earth luminescent materials by sol-gel method.

5.4.4.1. *Basic principles of sol-gel method*

Sol is a system in which tiny solid particles are suspended and dispersed in liquid phase and Brownian motion is carried out continuously. According to the strength of the interaction between particles and solvents, sols are usually classified into two types: hydrophilic and hydrophobic. Because the Gibbs free energy of interfacial atoms is higher than that of internal atoms, the sol is a thermodynamically unstable system. Without other restrictions, the micelles tend to coagulate spontaneously and reach a low specific surface state. If the above process is reversible, it is called flocculation; if it is irreversible, it is called gelation. Gel refers to colloidal particles or polymer molecules cross-linked, forming a space network structure, in the network structure pores filled with liquid (dispersed in dry gel medium can also be gas) dispersion system, not all sol can be changed into gel, gel formation is the key to the role of colloidal particles. Whether the force is strong enough to overcome the interaction between colloidal particles and solvents. There are four ways to prepare gelatin from sol.

- (1) volatilizing or cooling the sol, such as water, alcohol and other disperse medium, making it become supersaturated liquid and forming gel.
- (2) add non-solvent, such as adding proper amount of ethanol in pectin solution, then form gel.
- (3) gelatin can be formed by adding appropriate amount of electrolyte into colloidal particles with strong hydrophobicity (especially asymmetric shape), and gel can be formed if $\text{Fe}(\text{OH})_3$ can be formed under appropriate electrolytes.
- (4) using the chemical reaction to produce insoluble substances and controlling the reaction conditions, the gel can be obtained.

5.4.4.2. *Basic process of sol-gel method*

The inorganic salts and metal alkoxides or other organic salts are dissolved in water or organic solvents to form homogeneous solutions. Solutes and solvents produce hydrolysis, alcoholysis or chelating

reactions. The reaction results in the formation of ions aggregated into 1 nm and the formation of sols. The latter is changed into gel by evaporation drying, and the gel is transformed into the desired product after drying and heat treatment. According to the different raw materials used, sol-gel method can be divided into two categories: one is aqueous sol-gel method, the other is alkoxide sol-gel method.

5.4.4.3. *Characteristics of sol-gel method*

The test and analysis of products obtained by sol-gel method showed that sol-gel method has the following four advantages compared with the traditional high-temperature solid-state reaction method.

- (1) The homogeneity of the product is good, especially for multi-component products. The homogeneity of the product can reach the level of molecule or atom, so that the activated ions can be evenly distributed in the matrix lattice, which is conducive to finding the lowest concentration of the activated ions when the luminescence of the luminescent body is strongest.
- (2) Calcination temperature is lower than that of solid-state reaction at high temperature. Therefore, it can save energy and avoid impurities being introduced from the reactor due to high calcination temperature. At the same time, gelatin is formed partly before calcination, which has large surface area and is beneficial to product formation.
- (3) The purity of the product is high, because the reaction can use high-purity raw materials, and the solvent can be easily removed in the process of treatment. The reaction process and the microstructure of the gel are easy to control, which greatly reduces the support reaction.
- (4) The narrowing of the banded emission peak can improve the relative luminescence intensity and quantum efficiency of the luminescent body.
- (5) Films, fibers or bulk functional materials can be prepared at different stages of the reaction according to the need.

The disadvantage of sol-gel method is that the production process is too long and the cost is high. The gel washing of the precursor is difficult, and it is easy to form two particles during drying. During heat treatment, it will cause hard agglomeration of powder particles, resulting in poor dispersion of the final powder, and the alcohol salt has greater toxicity, which is harmful to the human body and the environment.

5.4.5. *Precipitation method*

The method of preparing inorganic and organic powders by precipitation of solute from homogeneous solution is called precipitation method. At the same time, many kinds of precipitation are precipitated to prepare many kinds of mixed powders, which is called coprecipitation method. The precipitation process is closely related to the concentration, pH and temperature of the solute in the solvent. The state of the precipitate can be controlled by adjusting the parameters such as pH and temperature. After the precipitate is heated and decomposed, ceramic powders or precursors such as oxides, sulfides, carbonates, oxalates and phosphates can be obtained. The theoretical basis of precipitation reaction is multiphase ion equilibrium of insoluble electrolyte. The precipitation reaction includes the formation, dissolution and transformation of precipitation. The formation and dissolution of new precipitation can be judged by the solubility product rule, and the transformation of precipitation can be judged by the solubility product constant of insoluble electrolyte. The precipitates after precipitation are generally crystalline with higher solubility, thinner solution, lower relative supersaturation and higher reaction temperature, while the precipitates directly precipitated are amorphous with smaller solubility, stronger solution, higher relative supersaturation and lower reaction temperature. Crystalline precipitation particles are large, high purity and easy to filter and wash. Amorphous precipitation particles are small and have many adsorbed impurities. Adsorbates are difficult to filter and wash. Precipitates and impurities can be separated by dilute electrolyte solution washing and aging.

Precipitation is also a common method in the preparation of luminescent materials. It has unique advantages in the preparation of metal oxides and nanomaterials. The product prepared by precipitation method has the advantages of low reaction temperature, high sample purity, uniform particle size, small particle size and good dispersion. The following points should be paid attention to in precipitation–coprecipitation method.

For the multi-ion coprecipitation process, the control of pH and the selection of precipitant should be paid special attention to. pH is very important for the precipitation process. Low pH may lead to incomplete precipitation, and high pH may lead to the re-dissolution of some precipitates.

For the multi-ion coprecipitation process, it is difficult to obtain a homogeneous coprecipitation mixture because of the different alkalinity required for the precipitation of each metal ion, i.e. the different pH. At this time, the homogeneous mixture of raw materials can be poured into NH_4OH to obtain a homogeneous precipitation mixture.

The particle size of the powder synthesized by precipitation method is closely related to the concentration of metal ions in the solution, and the particle size obtained by dilute solution is smaller. Moreover, the particle size of the powders synthesized by precipitation method is closely related to the temperature.

Precipitation method overcomes the shortcomings of hard mixing and uniformity of raw materials in solid-phase method, realizes mixing at molecular level of raw materials. Powders with controllable particle size, high dispersion, good chemical uniformity and high purity are prepared directly at low temperature, but the morphology of particles is difficult to control.

5.4.6. *Spray pyrolysis*

Luminescent materials usually include various metal sulfides, metal oxides, composite oxides and inorganic salts doped with rare earth ions and transition metal ions. These luminescent materials have been applied in many fields, especially in luminescent display such as

cathode ray tube (CRT), vacuum fluorescence display (VFD), plasma flat panel display (PDP), field emission display (FED), electroluminescent display (EL), fluorescent lamp lighting and some light-emitting diodes (LED). It has broad application prospects. In order to obtain luminescent materials with good performance for future display technologies such as high definition projection television and flat panel display, the shape and size of luminescent materials is a crucial aspect. For the luminescent body, the ideal particle shape is spherical. Spherical luminescent particles are necessary for high brightness and high resolution display. At the same time, spherical luminescent materials can also obtain higher stacking density, thus reducing the light scattering of luminescent bodies. Recent studies have shown that spherical luminescent particles can minimize the irregular shape of the luminescent layer, thereby prolonging the lifetime of the screen. In addition, the luminescent materials must have smaller size and narrower size distribution, and be non-agglomerated, so that they can have good luminescent properties, i.e. high resolution and high luminescent efficiency. Smaller particles can also increase their service life by forming compact phosphor layers. The results show that the optimum luminescent size should be 1–2 μm . Therefore, the preparation of non-agglomerated spherical luminescent materials with particle size of 1–2 μm has become the goal of current luminescent workers. In order to prepare spherical luminescent particles, many methods have been tried, such as spray pyrolysis, polymer microgel, complex precipitation, etc. In these methods, spray pyrolysis is the most effective and universal method to prepare spherical luminous powder. This method originated in the early 1960s and has been widely used in the preparation of inorganic materials, catalysts and ceramic materials in recent years. The development of spray pyrolysis is mainly divided into four stages: atomization droplet drying and drying particle thermal decomposition respectively; the two-stage method: atomization drying and thermal decomposition at high-temperature reaction zone simultaneously; continuous spray drying and thermal decomposition method; atomization droplet direct joining in gas-phase chemical reaction synthesis method.

The process of preparing materials by spray pyrolysis is as follows: first, water, ethanol or other solvents are used to mix the reaction raw material into a solution, then atomizing the reaction solution through a spray device and introducing it into a reactor, where the precursor solution is dried by spray drying, and the reactants are subjected to thermal decomposition or combustion and other chemical reactions, thus obtaining the initial reaction. Ultrafine particles with completely different chemical composition. The spray pyrolysis process is generally divided into two stages: the first stage is evaporation from the surface of the droplet, similar to direct heating evaporation. With the evaporation of solvent, the solute appears supersaturated state, which precipitates fine solid phase at the bottom of the droplet, then gradually expands around the droplet, and finally covers the entire surface of the droplet, forming a solid shell layer. The second stage of droplet drying is more complex, including the formation of stomata, fracture, expansion, shrinkage and grain size. Hair grows. In fact, the structure and properties of the initial precipitates on the liquid surface not only determine the properties of the solid particles to be formed, but also determine the conditions for the continuous precipitation of the solid phase. Every step of particle drying has a great influence on the next step. The equipment used in spray pyrolysis includes atomizer, pressure nozzle, quartz tube and heating furnace.

The aerosol process of liquid-phase precursor can be precipitated by spray pyrolysis. The solute can be precipitated in a short time. It has many advantages of traditional liquid-phase method and gas-phase method, such as the same composition of particles, spherical particles, controllable size, continuous process and high industrial potential.

- (1) As the micro-powder is dried by droplets suspended in the air, the prepared particles are generally spherical in shape and uniform in size and composition, which is difficult to achieve for other preparation methods such as precipitation, thermal decomposition and alcoholic hydrolysis, because the inner shape of a droplet is the main reason. The microreactor was formed

and the drying time was short. The whole process was completed quickly.

- (2) The composition of the product is controllable. Since the starting material is homogeneously mixed in the solution state, the final composition of the synthesized compound or functional material can be precisely controlled.
- (3) The morphology and properties of the product can be controlled. By controlling different operating conditions, such as reasonably selecting solvents, reaction temperature, spray velocity and carrier gas velocity, various fine powders with different shapes and properties can be prepared. Because the method itself utilizes the thermal decomposition of materials, the reaction temperature during the preparation of materials is low, and it is especially suitable for the preparation of ultrafine crystalline composite oxide powders. Compared with other materials prepared by other methods, the product has smaller apparent density, larger specific surface area and better calcination performance.
- (4) The preparation process is a continuous process, which does not require subsequent filtration, washing, drying and crushing processes in various liquid-phase methods. It is easy to operate and is therefore conducive to industrial scale-up.
- (5) There is no need to grind during the whole process, which can avoid introducing impurities and destroying crystal structure, thus ensuring high purity and high activity of the product.

5.4.7. *Microemulsion method*

Microemulsion method is adopted by the preparation of nanoparticles in recent years, a novel method, it is insoluble in water nonpolar substances as dispersion medium, the aqueous solution with different reactants as the dispersed phase, using the appropriate surfactant as emulsifier, forming type water-in-oil (W/O) microemulsion, make the formation of the particles space limit in the interior of the microemulsion droplets, and narrow particle size distribution, shape of nanoparticles. For example, using NP25/N29 as emulsifier,

Lee *et al.*, prepared $\text{Y}_2\text{O}_3\text{:Eu}$ nanocrystalline particles with the size of 20–30 nm by microemulsion method, and compared them with the samples prepared by precipitation method. It was found that the nanocrystalline particles obtained by microemulsion method have narrow size distribution, small particle size, high degree of crystallization and cathode ray luminescence efficiency. Kao *et al.*, prepared ZnS nano phosphors in $\text{H}_2\text{O}/$ supercritical CO_2 microemulsions. X-ray diffraction analysis results showed that the average size of ZnS nanocrystals was 2–3 nm, and emission spectrum test results showed that the luminescence intensity was higher than that of ZnS with micron size.

5.4.8. Polymer network gel method

With the progress of science and technology, luminescent materials are changed from single-component compounds to multi-component composites, which are more and more complex. Therefore, a new synthesis method — polymer network gel method is produced.

The structure of the product prepared by polymer network gel method was analyzed by X-ray diffraction, and the morphology of the particle was analyzed by transmission electron microscope. At the same time, the polymer network gel method has relatively simple requirements for raw materials, such as the use of inorganic salt solution. The product size is small, the uniformity of multiple components can reach the molecular level, and the synthesis temperature is greatly reduced. However, this method needs to select the appropriate network agent and initiator.

Take YAG:Ce^{3+} powder synthesized by polymer network gel method as an example. Firstly, network agent and initiator were added into the raw material solution, and the gel was polymerized at 80°C . Then, the gel was heated up to 700°C at $2^\circ\text{C}/\text{min}$ and kept constant for 2 hours. The analysis shows that YAG directly changes from amorphous form to YAG phase with the increase of temperature in the preparation of YAG by polymer network gel method, which is the difference of this method from other methods.

5.4.9. *Microwave method*

The improvement of luminescent material performance indexes needs to overcome the inherent defects of traditional synthesis methods, such as high-temperature solid-phase reaction temperature, large powder particles and easy to agglomerate. In order to carry out post-sequence work, products must be ground, and the grinding process seriously reduces the luminescent properties of materials and other indicators. Microwave method is a new interdisciplinary subject which has developed rapidly in recent ten years, and it is one of the most characteristic methods in the synthesis of new materials. This method overcomes the disadvantages of high-temperature solid-phase method, has the characteristics of fast, high efficiency and even heating, can significantly improve the luminescent material's many performance indexes, and has made a significant contribution to the preparation of inorganic powder luminescent materials. Fluorescence bodies, such as CaWO_4 and $(\text{YGd})\text{BO}_3\text{:Eu}^{3+}$, have been synthesized by microwave method. In this method, the reactants are weighed according to a certain chemical ratio, fully mixed and placed in a crucible to be heated for a certain time, and then taken out and cooled.

Microwave refers to electromagnetic wave with a frequency of 300 MHz–300 GHz, and its corresponding wavelength is 0.1–100 cm. Unlike visible light, microwaves are continuous and polarizable. There are obvious differences between microwave heating and traditional heating methods. Microwave heating is the bulk heating of materials caused by the loss of medium in the electromagnetic field. The microwave enters the interior of the material, and the microwave field interacts with the material to transform the energy of the electromagnetic field into the thermal energy of the material. However, traditional heating means that a heat source transfers heat to the surface of the heated material by means of heat radiation, conduction and convection, raising its surface temperature. Then, heat is transferred from the outside to the inside by virtue of conduction. The temperature gradient is high outside and low inside. The remarkable feature of microwave heating is that the material

is always in the microwave field. In addition to following the thermodynamic law, the internal particle movement is also affected by electromagnetic field. The higher the temperature is, the more active the particle is and the more strongly affected by electromagnetic field.

The obvious advantages of microwave synthesis method is fast, time saving, less energy consumption, easy operation, in a microwave oven products, product by the analysis of all sorts of luminous performance and indicators are not less than the conventional method, product loose and small particle size, uniform distribution, pure color, high luminous efficiency, has good application value, microwave synthesis has the following advantages.

- (1) By the uniqueness of selective heating mode and microwave rapid heating, microwave heating and is closely related to the dielectric constant of the medium, the dielectric constant of the medium is easy to use microwave heating, heating medium is not easy, small dielectric constant throughout the microwave devices only sample under high temperature and the rest is still in normal temperature, so both economic and convenient, and the whole device has the advantages of simple structure, low cost. Microwave heating can be heated at different depths at the same time, this "body heating effect" makes the heating fast and uniform, side reactions are reduced, the product is relatively simple. In addition, the efficiency of microwave energy conversion to heat energy is up to 80–90%, so microwave calcining can effectively save energy.
- (2) Microwave synthesis can improve the structure and properties of synthetic materials. Due to the fast microwave heating rate, the abnormal growth of grains in the process of material synthesis is avoided, and the materials with high purity, relatively complete crystal shape, fine particle size and uniform distribution can be synthesized in a short time and at low temperature. Generally, they can be directly applied without grinding. In addition, the sample is heated from the inside, so the temperature gradient of the treated material is opposite to that of the traditional heating method. Therefore, microwave heating can heat both large and

small workpiece, and reduce the thermal stress that causes cracks in the process of treatment.

- (3) Microwave heating has little thermal inertia. As long as the microwave tube and filament are added for 15 h, high pressure can be applied to heat the heated object instantly. After the power is turned off, the sample can be cooled down rapidly in the low-temperature environment.
- (4) Can improve the working environment and working conditions. Microwave heating starts by heating the product itself, rather than indirectly by conducting heat back to other media (such as air), so the equipment itself basically radiates no energy and does not have high ambient temperatures, which can improve the working environment and working conditions.

References

- Zhang T, Zhang J, 2001. *Inorganic Photoluminescence Materials and Applications*, Beijing: Chemical Industry Press.
- Jilin Institute of Physics, Chinese Academy of Sciences and Solid Luminescence Compilation Group, University of Science and Technology of China, 1976. *Luminescence of Solids*, Beijing: Science Press.
- Huang X *et al.*, 1997. Preparation of nanosized ZnS particles in emulsions, *Chinese J. Appl. Chem.* 14(1):117–118.
- Kao T W *et al.*, 2004. Supercritical microemulsions as nanoreactors for manufacturing ZnS nanophosphors, *Chem. Lett.* 33(7):802–803.
- Cao S, Han T, Tu M, 2011. The effect of Eu^{2+} doping concentration on luminescence properties of $\text{Ca}_2\text{MgSi}_2\text{O}_7:x\text{Eu}^{2+}$ green phosphor, *J. Phys.* 60(12):569–574.
- Han T *et al.*, 2006. Effect of K^+ ions on the photoluminescence properties of $\text{Ca}_{4.75}(\text{PO}_4)_3\text{Cl}:0.05\text{Eu}^{2+}, 0.2\text{Mn}^{2+}$ phosphors, *Mater. Lett.* 167:50–53.
- Han T *et al.*, 2012. Effects of sintering route and flux on the luminescence and morphology of YAG:Ce phosphors for white emitting-light diodes (LEDs). *Appl. Mech. Mat.* 236–237:9–15.
- Han T *et al.*, 2012. Chemical substitution effects of elements on photoluminescence properties of YAG:Ce phosphors using orthogonal experimental design, *Optical Mater.* 34(34):1618–1621.
- Han T *et al.*, 2013. Effects of annealing temperature on YAG:Ce synthesized by spray drying method, *Optik* 124(18):3539–3541.
- Han T *et al.*, 2014. Preparation and formation mechanism of a core-shell structured $\text{Al}_2\text{O}_3/\text{YAG:Ce}$ phosphor by spray drying method, *Raremetal Mater. Eng.* 43(10):2311–2315.

- Han T *et al.*, 2015. Large micro-sized $\text{K}_2\text{TiF}_6\text{:Mn}^{4+}$ red phosphors synthesised by a simple reduction reaction for high colour-rendering white light-emitting diodes, *RSC Adv.* 5:100054–100059.
- Wang J *et al.*, 2015. Morphology and photoluminescence of tunable green orange cerium doped terbium lutetium aluminum garnet, *Int. J. Electrochem.* 10:2554–2563.

This page intentionally left blank

Chapter 6

Light-Emitting Diode Packaging Technology

LED packaging plays a connecting role in the LED industry chain. LED packaging technology is developed and evolved on the basis of semiconductor discrete device packaging. The function of packaging is to provide adequate protection for the chip, prevent the chip from long-term exposure in the air or mechanical damage and failure, so as to improve the stability of the chip. In the packaging process, packaging materials and packaging methods are the main factors. Packaging materials and processes account for 30% to 60% of the total cost of LED lamps. Different packaging structures and materials can improve the light extraction efficiency and heat dissipation performance of LED, reduce light decay and improve its service life. In short, the key technology of LED packaging is to extract as much light as possible from the chip within a limited cost range, while reducing the thermal resistance of packaging and improving reliability.

6.1. LED Packaging Mode

LED packaging has undergone four stages: (1) pin-type LED (lamp-LED), suitable for LED with diameter of 3–5 mm, current less than 30 mA, power less than 0.1 W, mainly used for signal indication. (2) Surface mount device (SMD-LED), a packaging technology that can directly attach or weld packaged devices to the designated position of printed circuit board (PCB) surface, mainly metal brackets. The two structures of stand-type LED and PCB chip-type LED have the advantages of high reliability, good high frequency characteristics,

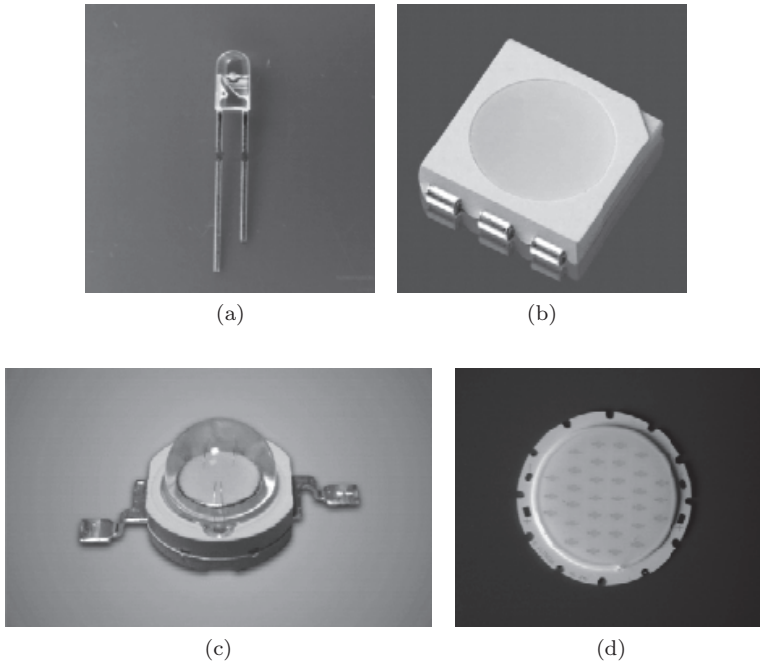


Fig. 6.1. Typical LED packaging: (a) Pin-type LED; (b) SMD-LED; (c) Power LED; and (d) COB LED.

easy automation and integration, etc. (3) High-power LED can reach 1 W, 3 W or even 5 W due to the increase of heat sink, which enlarges the application field of LED and has great significance for the application of white-light LED lighting. (4) Integrated LED (Chip On-Power LED) Board, COB type LED, encapsulates several LED chips directly on metal or ceramic substrates, and reduces thermal resistance by directly cooling the substrates. Typical LED packaging is shown in Fig. 6.1.

6.1.1. *Pin package*

Pin-type packaging technology is the first LED product successfully developed and put into the market. It has mature technology and a wide variety of products. Pin-type LED mainly consists of lead frame, radiation cup installed on chip, LED chip placed in reflector cup and

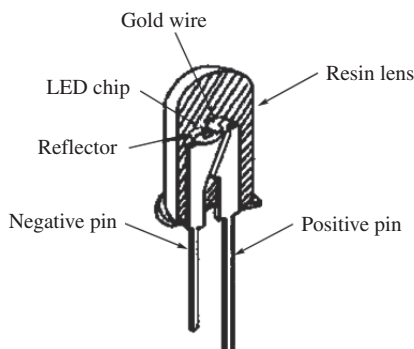


Fig. 6.2. Structural schematic diagram of pin-type encapsulated LED beads.

epoxy encapsulation lens. The electrical interconnection between chip and lead frame is established by gold wire binding. The light beam angle can be controlled by designing special lens. Figure 6.2 shows the structure of pin-encapsulated LED beads.

Pin-type LED packaging is mainly used in low-power chip packaging below 0.1 W. The LED emits 90% of the heat, which is transmitted from the negative pin rack to the circuit board and then to the external environment. Therefore, the biggest disadvantage of this packaging structure is that the thermal resistance is very large, reaching more than 250 K/W, and it cannot be used for high-power LED packaging. Moreover, the epoxy resin used for packaging is easy to aging under ultraviolet light, which will further accelerate the photodegradation.

6.1.2. Surface mount packaging

Surface mount technology (SMT) is an electronic assembly technology that directly solders chip-based, miniaturized, non-lead and short-lead surface mounted devices (SMD) onto the surface of PCB and other wiring substrates by automatic assembly equipment. It is an important technology link that integrates dispersed components into components and components. The LED packaging structure wraps the metal lead frame in the high-temperature resistant nylon Polyphthalamide (PPA) PPA plastic by injection molding process,

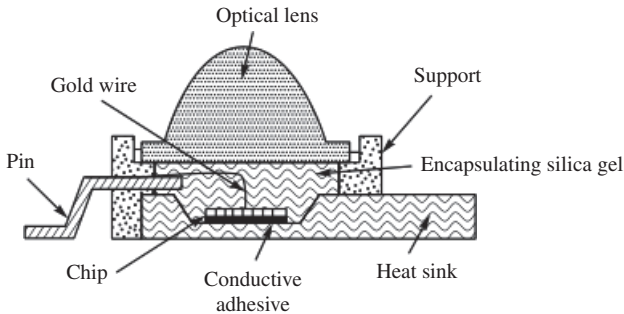


Fig. 6.3. Structural schematic diagram of surface mounted packaging LED chip.

and forms a reflecting cup with a specific shape. The metal lead frame extends from the bottom of the reflecting cup to the side of the device, forming the device pin by flattening outward or bending inward. One or several small and medium power LED chips are solidified on the metal lead frame in the reflecting cup through conductive silver paste and form electrical connection through gold wire binding. The chips radiate heat outward through the lead frame. As shown in Fig. 6.3, they are mainly used in applications with power less than 0.3 W. Common models are 3528, 5050, etc. Compared with pin-type packaging technology, SMD-LED uses lighter packaging circuit board and reflective layer material, which reduces size and weight, and is especially suitable for indoor and semi-outdoor full-color display applications. The biggest disadvantage is that the bonding force between PPA plastics and copper lead frame is poor. There are cracks at the interface between the two materials. Water vapor or corrosive gases are easy to penetrate into the device along the bonding gap, resulting in device failure.

6.1.3. *Power package*

In recent years, LED chips and packaging technology have been developing towards high power. Luxeon LED launched by Lumi LEDs in 1998 is the first commercialized W-class LED high power device. The luminous flux of $\phi 5$ mm pin-type packaging LED under high current is only 5–10% of that of power-type packaging LED.

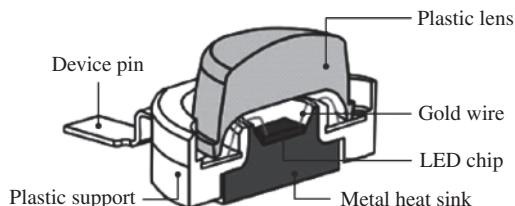


Fig. 6.4. High power LED packaging architecture.

As the power becomes larger, the heat generated will inevitably increase, so it is necessary for power LED to design heat and select heat dissipation materials to solve the problem of light decay and life caused by temperature rise. This structure adopts the idea of thermoelectric separation design. The LED chip is placed on a specially designed heat sink of high thermal conductivity material (usually copper heat sink). The chip electrodes are bound to the left and right pins by gold wire. The chip heat is transmitted downward through the heat sink and the current is injected through the pins. As long as the heat sink is in good contact with the external heat sink, the junction temperature of the LED chip can be obtained. Effective control, as shown in Fig. 6.4.

6.1.4. *Integrated multichip device packaging*

With the popularity of LED applications, engineers found that the LED light source based on LED devices has the following shortcomings: (1) Because the LED device is a point light source, it cannot provide the same uniform luminous effect as fluorescent lamp and incandescent lamp. In addition to causing glare, the illumination of point light source will produce double shadows when working under the light source, which seriously affects the illumination effect. (2) The assembly route of LED light source discrete device Metal Core printed circuit board (MCPCB) LED light source module LED lamps and lanterns is usually adopted, which not only consumes time and adds extra material consumption, but also introduces multi-stage thermal interface layer into many assemblies, and the thermal resistance of the module is large. In practical application, we can

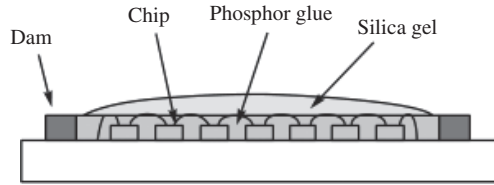


Fig. 6.5. Schematic diagram of COB structure.

merge the “LED light source discrete device metal core circuit board LED light source module” and package the LED chip directly into the light source module. The assembly route of “LED light source module LED lamps” is adopted. This light source module is called integrated multi-chip packaging or COB packaging.

COB package directly attaches the chip to the surface of the substrate, and realizes the electrical connection with the substrate by wire bonding. The encapsulation structure is manufactured by injecting reflecting cup on copper plate and fixing lead frame electrode at the same time. The main structure is shown in Fig. 6.5. After fixing the chip and welding the gold wire, fluorescent glue is poured into the substrate (some of which are dotted with silica gel) to protect the chip from the external environment and improve the thermal conductivity and heat dissipation ability.

COB packaging structure encapsulates multiple light-emitting diodes (LEDs) in a small area. It has the advantages of small size, low cost, good heat dissipation and high light output, making the assembled lamp shell lighter and simpler, easy to achieve secondary light distribution and achieve specific optical distribution. The silica gel on the surface of common COB packaging structure is flat or slightly convex, but under this structure, there is a serious problem of total reflection of light at the interface between colloid and air.

6.1.5. Other packaging methods

6.1.5.1. Flip-chip LED package

LED chips usually adopt sapphire substrates as the front-mounted structure. However, due to the poor thermal conductivity of sapphire,

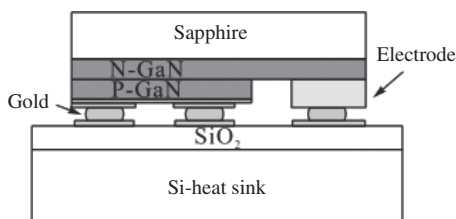


Fig. 6.6. Flip-chip packaging for LED chips.

the heat generated by the chips can hardly be transferred to the heat sink, which limits the application of power-type LED. Flip chip packaging is one of the current development directions. Compared with the formal structure, heat does not need to pass through the sapphire substrate of the chip, but directly transfers to the silicon or ceramic substrate with higher thermal conductivity, and then emits to the external environment through the metal base.

Flip-chip packaging requires the preparation of large size LED chips suitable for eutectic welding, as shown in Fig. 6.6. At the same time, it is necessary to fabricate the corresponding size of the silicon substrate, to fabricate the gold conductive layer of eutectic welding electrodes on the silicon substrate, and to extract the conductive layer (the joints of ultrasonic gold wire ball welding). Next, the silicon substrate and the LED chip are welded together by the eutectic welding equipment. Through this method, heat is not transmitted to sapphire substrate of LED chip, but directly to silicon or ceramic substrate. This method reduces the thermal resistance of internal heat sink. Therefore, the heat dissipation effect of flip-chip structure has been greatly improved. Theoretically, the thermal resistance can be as low as 1.34°C/W , and in practical measurement it can reach $6\text{--}8^{\circ}\text{C/W}$. However, the general GaN-based flip-chip LED is still the most transverse structure, and the phenomenon of current congestion is still widespread.

6.1.5.2. LED filament

In 2008, Niuwei Light Source launched the bulb lamp “LED filament bulb” with incandescent lamp as the prototype. The process of LED

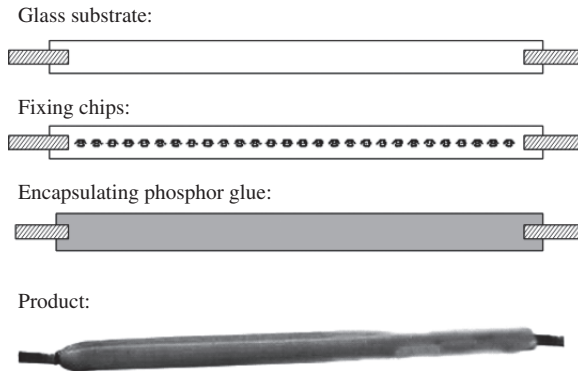


Fig. 6.7. Packaging technology of LED filament.

filament is usually to weld 28 piece of 1016 LED chips (0.02 W) in series with gold wires on a glass substrate with a length of 38 mm and a diameter of 1.5 mm, and then package the fluorescent glue to achieve as shown in Fig. 6.7. The LED filament is driven by 10 mA current, with a voltage of 84 V, a power of 0.84 W and a luminous flux of 100 lm. The luminous efficiency can reach 120 lm/W. If matched with red chip, the color rendering index can reach more than 95. In the past, LED light sources, such as pin-type LED, patch LED, COB, and other LED beads, can only be planar light sources without optical devices such as lenses, but the LED filament breaks through this point and achieves 360° all-angle light-emitting three-dimensional light source, avoiding the problem of affecting the light effect and causing light loss due to adding lenses. In addition, the characteristics of low current and high voltage of LED filament effectively reduce the cost of LED heating and driver, and have outstanding advantages. LED filament has been used in crystal pendant lamp, candle lamp, bulb lamp, wall lamp and so on, and has a very broad prospect.

6.2. LED Packaging Technology

LED packaging process is the implementation part of packaging design, which has an important impact on the performance of the

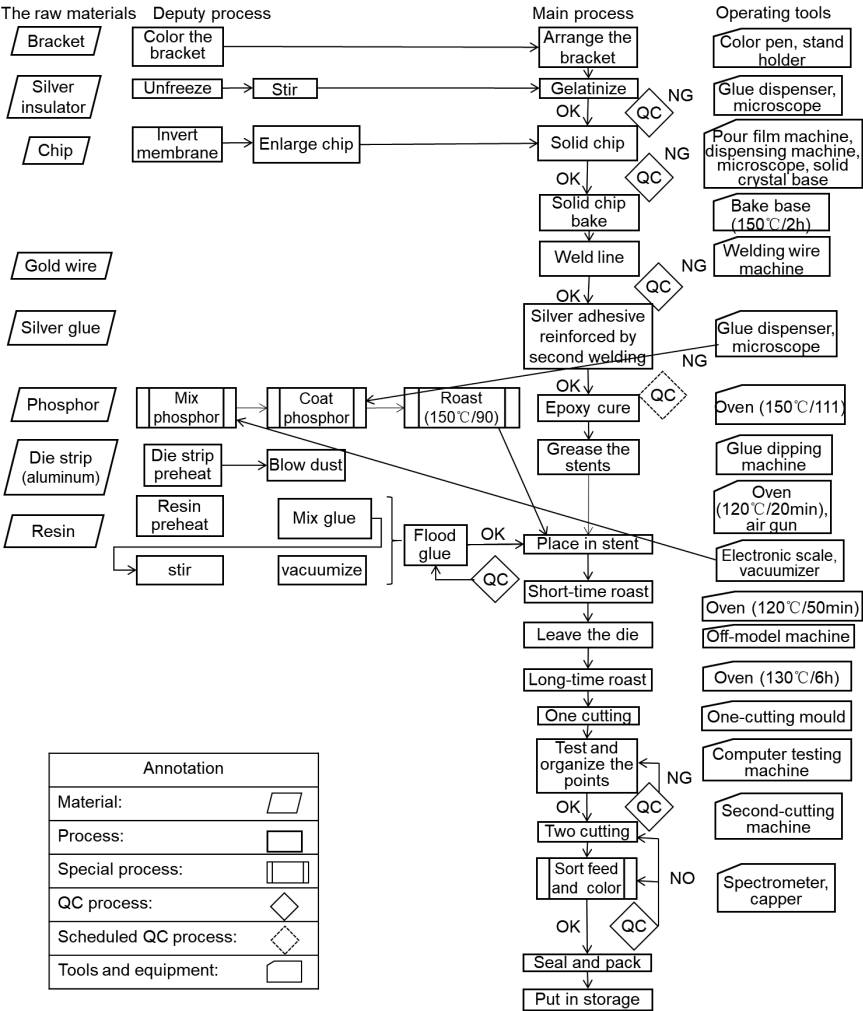


Fig. 6.8. Process flow chart of pin white LED.

product. This section mainly introduces the LED packaging process with pin-type LED packaging process as an example. Figure 6.8 shows the flow chart of pin-type white-light LED packaging process. The chip-type LED packaging process should omit everything and dichotomy. For high-power LED packaging, lens placement and other links need to be added. If the white light is not needed, the phosphor

links corresponding to the chip can be omitted in the colloid coating links.

- (1) Chip inspection: Whether there is mechanical damage, pitting or pitting on the surface of the mirror material, whether the chip size and size meet the technological requirements of LED packaging, and whether the pattern of the electrode is complete or not.
- (2) Expansion: The arrangement of LED chip products after scratching is still relatively close, and the distance between them is only about 0.1 mm, which is not conducive to the subsequent process operation. Therefore, it is necessary to use a expander, also known as a crystal expander, to expand the film of the chip, and the chip is pulled to a distance of more than 0.5 mm.
- (3) Glue dispensing: Silver glue (conductive) or insulating glue should be placed on the corresponding position of the LED encapsulated bracket or reflector cup. For conductive substrates such as GaAs and SiC, silver glue is used to fix the LED chip. For blue and green LED chips, because they are sapphire insulating substrates, they are fixed with insulating glue. Controlling dispensing quantity is a difficult point in dispensing technology. At the same time, there are specific technical standards and requirements for dispensing position and height of colloid formation.
- (4) Glue preparation: Contrary to dispensing, glue preparation is to apply silver glue on the back electrodes of LED products first, then fix the LED products with silver glue on the back and install them on the support of LED products. So the efficiency of preparing glue is generally much higher than dispensing glue, but not all products are suitable for the process of preparing glue.
- (5) Crystallization: It is a particularly important process, the purpose of which is to stick the chip to the cup on the bracket through silver glue or insulating glue. After the dispensing process is completed, the wafer is placed, the position of the chip is adjusted, the two poles of the chip are seen along the

vertical direction through the interface, and the cross fork on the interface is located at the edge through the adjustment of the handle and keyboard, and then the front and rear hook claws are adjusted respectively to make it just catch the bracket. The chip is located in the center of the cup by adjusting the knobs on both sides, and the silver glue or insulating glue is uniform by adjusting the built-in knobs. There are three key points in crystallization: adjusting the parameters of thimble, suction nozzle and suction according to wafer. adjusting the parameters of silver glue according to wafer and base. adjusting the position of grain, offset angle and thrust process. Figure 6.9 is the schematic diagram of the crystallization process.

- (6) Manual stitching: The expanded LED product chip (with or without glue) is fixed on the fixture of the stitching table. At the same time, the LED product support is placed under the fixture, and the LED product chip is stabbed to the corresponding position one by one with a needle under the microscope. Comparing with automatic mounting, manual spinning has the advantage that different chips can be replaced at any time. It is suitable for products that need different kinds of chips.
- (7) Mounting: In fact, mounting combines two steps of glue dispensing and crystal fixing. Firstly, silver glue (insulating glue) is applied on the support of LED products. Then, the chip of LED products is sucked up and moved by vacuum suction nozzle, and finally placed on the corresponding position of the support.

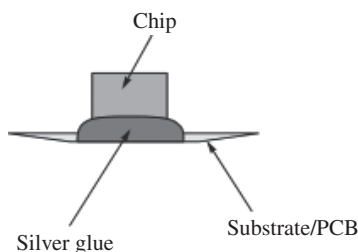


Fig. 6.9. Solidification process.

- (8) Sintering: The purpose of sintering is to solidify silver glue. Sintering requires strict temperature control in order to prevent bad batch performance. The sintering temperature of silver glue is generally controlled at about 150°C and the sintering time is about 2h. According to the actual situation, the temperature and time can be adjusted to 170°C , 1h. The time interval of silver glue sintering oven must be 2 hours (or 1 hour) to open and replace the sintered products, and it is not allowed to open freely in the sintering process. The sintering oven shall not be used for other purposes after being used to prevent pollution.
- (9) Welding wire: The purpose of the welding wire is to connect the LED chip with the bracket, and to connect the pin with the circuit board by soldering tin when assembling, thus realizing the electrical connection of the internal LED chip. The welding process of LED chips can be divided into ball welding and pressure welding. Gold wire is usually used for ball welding. Firstly, the gold wire magnet nozzle is melted into a golden ball, then pressed on the electrode of the LED chip, and then the gold wire is arced over the corresponding bracket. At the second point, the gold wire is melted into a golden ball, and the gold wire is broken after pressing. Aluminum wire pressure welding does not require melting balls, other similar (Fig. 6.10).
- (10) Dispensing packaging: dispensing packaging of LED chips mainly includes dispensing, filling and molding. The difficulty of process control is the existence of bubbles, black spots and irregularities. In the design process, epoxy resin and bracket

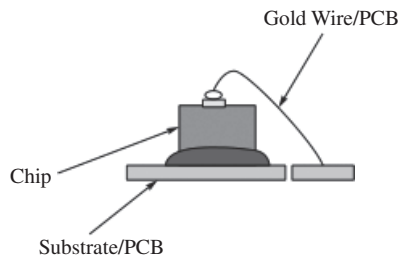


Fig. 6.10. Welding wire technology.

with good bonding should be selected for material selection. In addition, the main difficulty of manual dispensing packaging lies in the control of dispensing volume, which requires a high level of operation (especially white LED products), because epoxy resin will become thicker in use. It should be pointed out that the dispensing of white LED products still has the problem of color difference caused by phosphor precipitation.

- (11) Filling packaging: The pin-type LED packaging is in the form of filling. The filling process is to inject liquid epoxy resin into the cavity of the LED forming die, then insert the pressure welded LED bracket into the oven to solidify the epoxy. After the epoxy is solidified, the LED is removed from the cavity and formed.
- (12) Molding and packaging: Put the welded LED bracket into the mold, close the upper and lower dies with hydraulic press and vacuum them, put the solid epoxy resin into the inlet of the injection duct to heat and press the hydraulic ejector into the mold duct, and the epoxy resin enters each LED forming slot along the duct and solidifies.
- (13) Curing and post-curing: Curing is the baking and curing of the encapsulated epoxy resin. The curing temperature and time of the general epoxy resin are 135°C and 1h. The general temperature and time of die-pressing packaging are 150°C and 4 min. In order to make the curing of epoxy resin more fully, post-curing is necessary. At the same time, thermal aging of LED products is also required. The general temperature condition is 120°C. Post-curing is very important to improve the bond strength between epoxy resin and scaffolds.
- (14) Reinforcement and slicing: Lamp-LED products generally use the method of cutting the reinforcement of the LED product bracket because the LED products are connected (rather than single) in the actual production. SMD-LED products are on a PCB and need a slicing machine to complete the separation work.
- (15) Testing: Testing the photoelectric parameters and shape dimensions of LED products after scratching, and sorting the LED products according to different requirements of customers.

- (16) Packaging: counting and packaging the finished products. It is noteworthy that ultra-bright LED products need anti-static packaging.

6.3. LED Packaging Materials and Equipment

6.3.1. *LED packaging material*

Packaging materials play an important role in the function of LED chips. Improper selection of packaging materials will result in poor heat dissipation or low light output, which will lead to functional failure of the chip. LED packaging materials mainly include bracket (substrate), chip, gold wire, conductive adhesive, phosphor, packaging adhesive, filling adhesive and optical lens, etc.

6.3.1.1. *Bracket (substrate)*

The bracket is the key path in the heat dissipation channel of high-power LED, and the bracket has both the function of electrical connection and mechanical support. How to transfer excess heat to heat sink effectively, the selection of material and structure of the bracket is very important. At present, the most widely used scaffolds are mainly metal and ceramic substrates. Packaging bracket requires high thermal conductivity, high insulation, high heat resistance, thermal expansion coefficient matching with chip and high strength.

Metal Core printed circuit boards (MCPCB) are a kind of printed boards, which are mainly prepared by lamination process. It was produced in the United States in the 1960s. MCPCB has the advantages of good heat dissipation, small thermal expansion coefficient, high-dimensional stability and shielding. It is widely used in high frequency equipment, high power module power supply and electronic products. The MCPCB substrate consists of three layers, i.e. circuit copper layer, insulation layer and metal base layer, from top to bottom, as shown in Fig. 6.11. The copper layer of the circuit forms an electrical connection, and the metal base usually adopts copper or aluminum with high thermal conductivity. Insulation layer usually has high requirements, including high insulation, high

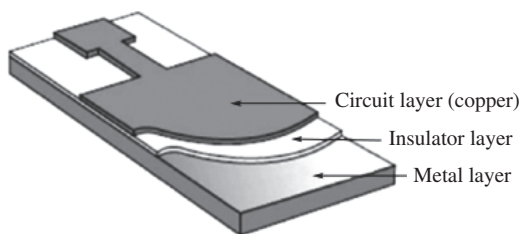


Fig. 6.11. Metal-based printed circuit board.

stability, high thermal conductivity, thermal expansion coefficient matching with chip and good surface quality.

Ceramic materials are widely used to prepare heat dissipation substrates because of their good material properties. Ceramic substrates are mainly divided into co-fired ceramic substrates, ceramic copper clad laminates Direct Bonded Copper (DBC), Direct Aluminum Bonded (DAB) and Direct Plate Copper (DPC). Co-fired ceramic multilayer substrates are divided into High-Temperature Co-fired Ceramic (HTCC) and Low-Temperature Co-fired Ceramic (LTCC). The basic technology is to mix alumina powder and organic binder into mud slurry evenly and scrape the slurry into flakes with a scraper. After drying, thin ceramic green blanks are formed. According to the interconnection requirements of each layer, through holes are designed and processed. After filling with metal, through holes are used as vertical transmission channels of signals of each layer. Each layer of electrical interconnection is printed by screen printing technology, and the electrodes are made of copper, silver, gold and other metals. After laminating the raw porcelain sheets which are made of multi-layer circuit and filled with holes, they are sintered in a co-fired furnace and finally split into packaging substrates.

Compared with metal packaging substrate, ceramic substrate eliminates the complex fabrication process of insulating layer. Multilayer Ceramic Metal Packaging (MLCMP) technology has been greatly improved in heat treatment compared with traditional packaging methods. A new type of AlN ceramics has the characteristics of high thermal conductivity, low dielectric constant and low dielectric loss. It is an ideal material for new generation semiconductor

Table 6.1. Performance comparison of common LED packaging bracket substrates.

Performance indicators	PCB	MCPCB	LTCC	TFC*	DBC*	DPC*
Thermal conductivity [W/(m · K)]	0.3–0.4	1.0–3.0	2.0–3.0	20–25	20–25	20–25
CTE ($\times 10^{-6}/^{\circ}\text{C}$)	13–17	17–23	5.0–8.0	5.0–8.0	5.0–8.0	5.0–8.0
Process temperature $^{\circ}\text{C}$	200	200	850–900	700–800	1065	200–300
Thickness of metal layer(μm)	—	<100	—	5–20	100–600	10–100
Current carrying capacity	Poor	Poor	Excellent	Good	Excellent	Good
Adhesion	Low	Low	Strong	Strong	Strong	Strong
Heat shock resistance	Very poor	Poor	Excellent	Excellent	Good	Good
Minimum line width (μm)	>200	>200	>300	100	150	30–50
Maximum operating temperature ($^{\circ}\text{C}$)	115	150	500	300~500	500~800	500~600
Main applications	Low power	Medium power is widely used	Medium power	Medium power COB packaging	High power COB packaging	High power COB packaging
Cost	Very low	Low	High	Low	High	High

Note: *Uses alumina ceramic substrates.

packaging. DBC is also a kind of ceramic substrate with excellent thermal conductivity. The ultra-thin composite substrate has excellent electrical insulation and high thermal conductivity. Its thermal conductivity can reach $24\text{--}28\text{ W}/(\text{m}\cdot\text{K})$.

6.3.1.2. *Packaging glue*

The packaging adhesive must have the characteristics of high thermal conductivity, high transmittance, good heat resistance and good UV (ultraviolet) shielding resistance. Because of its special requirements for transparency, packaging adhesives are mainly used in the market at present, such as epoxy resin, silicone, polycarbonate, glass, polymethacrylate and other high transparency materials. The traditional packaging material of LED is epoxy resin, which is widely used in monochrome LED and low power white LED devices. However, epoxy resin has poor ultraviolet resistance and thermal aging performance. When used in high power white LED packaging, the transmittance of materials will be significantly reduced with the extension of the device's use time, and the device's life will be greatly shortened. Moreover, the hardness of epoxy material is relatively high and its processing is not convenient, so it is basically used for outer lens materials. Silicone material is a kind of material with high ultraviolet resistance, high aging resistance and low stress. It has become an ideal choice for LED packaging materials.

According to the law of refraction, when light is incident from a dense medium to a sparse medium, total reflection occurs when the incident angle reaches a certain value, i.e. greater than or equal to the critical angle. In the case of GaN blue chip, the refractive index of GaN material is 2.3. When light radiates from the inside of the crystal to air, according to the law of refraction:

$$\theta_0 = \arcsin(n_1/n_2), \quad (6.1)$$

where, n_2 is equal to 1, that is, the refractive index of air. n_1 is the refractive index of GaN, from which the critical angle of θ_0 is calculated to be about 25.8 degrees. The light emitted by the active layer is extracted only a small part of the light produced by the active

layer, and most of the light is easily absorbed by multiple reflections in the interior, resulting in excessive light loss. In order to improve the light extraction efficiency of LED packaging, it is necessary to increase the value of n_2 , that is, to improve the refractive index of the packaging material, so as to improve the critical angle of the product, so as to improve the packaging luminescence efficiency of the product. At the same time, packaging materials are required to absorb less light. The transmittance of silicone resin is proportional to the luminous intensity and efficiency of LED devices. The higher the transmittance, the more conducive to increasing the luminous intensity and efficiency of LED devices.

Because GaN chips have high refractive index (about 2.2), the refractive index of ordinary silicone materials is only 1.4. In order to improve the refractive index of silicone materials, the first method is to introduce phenyl groups into the molecular structure. Since the content of phenyl cannot be increased indefinitely, the refractive index of silicone materials can be increased to about 1.55 by introducing phenyl into the molecular structure. At present, the refractive index of silicone materials for LED packaging with the highest refractive index can be seen in the market, which is about 1.54. The second is that the refractive index can be increased to more than 1.7 by introducing inorganic nanoparticles such as nanotitanium dioxide and zinc oxide. However, there are still many technical problems in the modification of nanoparticles.

For white LED sealant, the traditional two-component epoxy resin is selected. There are two problems in this kind of sealant: the optical grade resin is easy to be heated and yellowed. the short wavelength radiation can also cause the aging of the epoxy resin. This is because the luminous spectrum of white LED contains short wavelength light, while epoxy resin is easily degraded by short wavelength light in white LED. Low-power white-light LEDs have already caused the degradation of epoxy resin, and high-power white-light LEDs contain more short-wavelength light, which will cause accelerated aging. White LED is encapsulated in silica gel. Silica gel can disperse blue light and near ultraviolet rays, besides having better heat resistance and less ageing for short wavelength. Therefore,

compared with epoxy resin, silicone resin can inhibit the deterioration of materials caused by short wavelength light. In addition, the transmittance and refractive index of silica gel are ideal. Silicone resin sealant is a kind of stable flexible cementitious material. It will not produce internal stress due to fusion of temperature at 40–120°C. It will disconnect the gold wire from the lead frame and prevent the “lens” formed by the epoxy resin encapsulated outside.

6.3.1.3. *Conductive adhesive*

Conductive adhesives for LED are used to heat and fix chips. Silver gum is mainly composed of resin, silver powder and hardener, in which the content of silver powder is 65–80%. It is expected to become a new direction of heat dissipation. Conductive silver adhesives are widely used in packaging and conductive circuits in microelectronics, LED, LCD, electronic components, integrated circuits, automotive electronics, radio frequency identification, electronic tags and other fields because of their environmental friendliness and high conductivity. Table 6.2 shows the technical parameters of partially conductive silver glue.

6.3.1.4. *Golden thread*

Gold wire is mainly used to connect wafer and bracket in LED packaging, which is 99.99% pure gold. The commonly used gold wires are 0.9 mil (1 mil = 25.4 μm , diameter unit), 1.0 mil, 1.2 mil, etc.

6.3.1.5. *Phosphors*

The generation of white LED is mainly based on the principle of light recombination. Various types of white light can be obtained by changing the proportion of different color light. There are three main technologies in white light packaging technology: single chip, dual chip, multi-chip and so on.

Single chip packaging: A single chip with different phosphors can be generated in a variety of ways. Blu-ray chip + YAG yellow phosphor is the most widely used and simplest packaging method at

Table 6.2. Technical parameters of partially conductive silver glue.

Parameter	Model					
		Ablestik 826-1DS	T3007-20	T-11	DAD-87	TK129-L
Volume resistivity ($\Omega \cdot \text{cm}$)		5×10^{-4}	4.9×10^{-4}	2×10^{-4}	5×10^{-4}	2.3×10^{-4}
Shear strength (MPa)		9.3	30.53	3.5	4	20
Heat conductivity [$\text{W}/(\text{m K})$]		2	1.2		14.58	2.3
$\eta(\text{Pa} \cdot \text{s})$		19	15–25	120	15–25	8.6
Working life (h)		24	12			
Curing conditions	T($^{\circ}\text{C}$)	150	150	150	150	150
	t(min)	30	30	60	240	60~90
Ion number (mg/kg)	Cr ⁺	175		16	10	
	Na ⁺	200		16	5	
	K ⁺	2		7		
Curing loss	loss rate (%)	2.2	0.1	0.019	1	0.65
	T/ $^{\circ}\text{C}$	300	250	250	341	300
Coefficient of thermal expansion [$\mu\text{m}/(\text{m}^{\circ}\text{C})$]	<T _g	50	90		22	40
	>T _g	1700	500		74	100
Storage life	t/a	1	0.5	1	1	0.5
	T/ $^{\circ}\text{C}$	−40	−50 to −40	0	−5–0	−15
Price (Ten thousand yuan/kg)		1.1	1.5	0.5	0.35	0.9
Place of origin		USA	Japan	Britain	Shanghai	Shenzhen

present. It was first proposed by Japan Niya Company. In the specific process, a 400–470 nm blue light chip is coated with YAG (yttrium aluminium garnet) phosphor. The blue light emitted by the chip is absorbed partly by the phosphor, and the other part is combined with the yellow light emitted by the phosphor to form a white light. White LEDs with different color temperatures can be prepared by different bands of chips and different amounts of phosphors. Due to the lack of red part in phosphor emission spectrum, the color rendering is poor, and blue LED will shift due to the influence of temperature and current, resulting in the color change of white light source. Blue chip + red phosphor + green phosphor, this method can achieve high color rendering, but because the efficiency of red phosphor and green phosphor is not high, the efficiency of white LED encapsulated by this method is low. Ultraviolet chip + red, green and blue phosphors. The principle of this method is that LED chip generates ultraviolet light to excite triple-color phosphors, and then compounds them into white light. This method has good color rendering. However, the low efficiency of red and green phosphors also leads to low luminescence efficiency. In addition, the temperature stability of phosphors also needs to be considered. Figure 6.12 shows a schematic diagram of the implementation of white LED.

The phosphors used in white LED should also meet other characteristics, such as strong absorption, wide band excitation and emission, high quantum efficiency, low thermal quenching, stable performance, suitable particle size and morphology, etc. to meet the

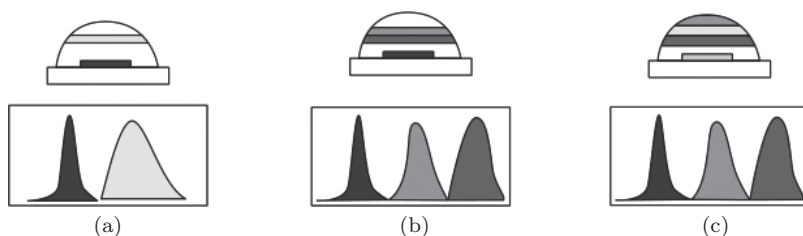


Fig. 6.12. Realization of white LED: (a) blue LED + yellow phosphor; (b) blue LED + green and red phosphors; and (c) ultraviolet LED + yellow, green and red phosphors.

Table 6.3. Fluorescent powder for white LED.

Phosphor	Chemical composition	Luminescence intensity	Peak width	Durability	Thermal quenching or thermal stability
Green phosphor	$Y_3(Al, Ga)_5O_{12}:Ce$	Δ	wide	o	Δ
	$SrGa_2S_4:Eu$	o	secondary	$\times$	$\times$
	$(Ba, Sr)_2SiO_4:Eu$	o	secondary	Δ	Δ
	$Ca_3Sc_2Si_3O_{12}:Ce$	o	wide	o	o
	$CaSc_2O_4:Ce$	o	wide	o	o
	B-sialon:Eu	o	secondary	o	o
	$(Sr, Ba)Si_2O_2N_2:Eu$	o	secondary	o	o
	$Ba_3Si_6O_{12}N_2:Eu$	o	secondary	o	o
Yellow phosphor	$(Y, Gd)_3Al_5O_{12}:Ce$	o	wide	o	Δ
	$Tb_3Al_5O_{12}:Ce$	Δ	wide	o	Δ
	$CaGa_2S_4:Eu$	o	secondary	$\times$	$\times$
	$(Sr, Ca, Ba)_2SiO_4:Eu$	o	wide	o	Δ
	Ca-R-sialon:Eu	o	secondary	o	o
Red phosphor	$(Sr, Ca)S:Eu$	o	wide	$\times$	$\times$
	$(Ca, Sr)_2Si_5N_8:Eu$	o	wide	Δ	Δ
	$CaAlSiN_3:Eu$	o	wide	o	o
	$(Sr, Ba)_3SiO_5:Eu$	o	wide	$\times$	o
	$K_2SiF_6:Mn$	o	narrow	o	o

Note: o means good, Δ means medium, $\times$ means bad.

requirements of white LED packaging. In recent years, a large number of phosphors for lighting have been studied, but only a few phosphors of garnet, silicate, sulfide and nitride can be effectively stimulated by blue-light InGaN chips, as shown in Table 6.3.

6.3.2. LED packaging equipment

6.3.2.1. Metallographic microscope

It is used to detect problems in the process of LED packaging, such as chip surface damage, as shown in Fig. 6.13.

6.3.2.2. Wafer expander

Using the heating plasticity of the LED film and the up and down control of double cylinders, the single LED wafer is evenly diffused to all sides, and the wafer gap is satisfied, and the film is formed



Fig. 6.13. Metallurgical microscope.

automatically, so that it can be better implanted into the welding workpiece, as shown in Fig. 6.14.

6.3.2.3. *Dispenser*

LED (light-emitting diode) with silver glue, epoxy resin, etc. The dispensing machine, like the crystal fixing machine, requires high precision, so that the glue quantity can be effectively controlled. If the amount of glue is too large, the chip will easily extrude the excess glue after pasting, block and absorb the light around the chip, and absorb the light emitted from the reflecting cup wall, which will affect the brightness. If the amount of glue is too small, especially when entering the welding process, the chip will easily fall off from the bottom of the cup, causing dead lights, leakage and so on; secondary products. The dispenser is shown in Fig. 6.15.

6.3.2.4. *Backing glue machine*

It is used for dipping silver glue after crystallization of LED (light-emitting diode). The backing glue machine is shown in Fig. 6.16.



Fig. 6.14. Crystal expander.



Fig. 6.15. Dispenser.

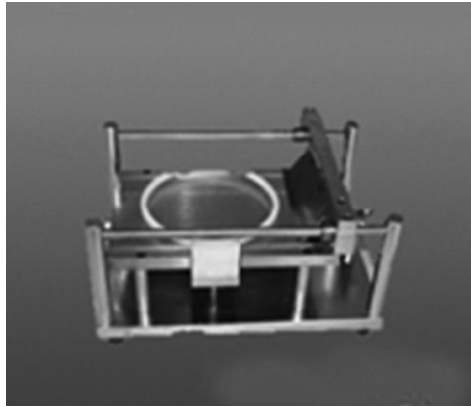


Fig. 6.16. Backing glue machine.

6.3.2.5. *Crystallizer*

When the LED chip is placed on the bracket, the precise placement of the LED grains in the encapsulation position will affect the luminous efficiency of the whole package device. If the position of the grains in the reflecting cup is deviated, the light cannot be emitted completely, which will affect the brightness of the finished product. Therefore, it is necessary to choose a high-precision crystal fixer, preferably with advanced image recognition system. The crystal fixer is shown in Fig. 6.17.

6.3.2.6. *Wire welder*

Weld the gold wire to the chip electrode. Before the wire welder is used, the power, temperature and pressure of first and second welding, and the temperature and power of ultrasonic wave should be adjusted so that these parameters can make the wire withstand 5 g pull force, so that the subsequent baking process will not be broken or de-welded because of the different expansion coefficient of the material. The wire welder is shown in Fig. 6.18.



Fig. 6.17. Crystallizer.



Fig. 6.18. Wire welder.

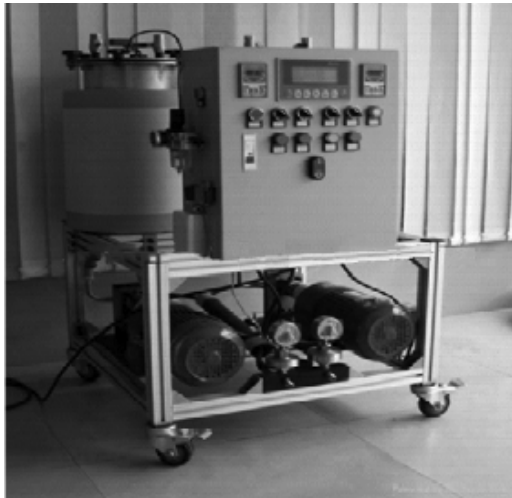


Fig. 6.19. Glue-filling machine.

6.3.2.7. *Glue-filling machine*

The LED is encapsulated with epoxy resin. The needles of the glue-filling machine must be kept at the same level, and the leaking passage must not contain dregs, and it is well sealed, and the needles must be cleaned up at intervals. Because the encapsulated lens is a layer of optical “lens” formed by epoxy resin, if the lens is mixed with impurities, it will make the light output efficiency poor, and there will be black spots in the spot. The glue filling machine is shown in Fig. 6.19.

6.3.2.8. *Oven*

Cured epoxy resin. The oven must be circulating air, and the tray on the baffle must be level. In the production of white LED, the fluorescent powder must be dried in the oven, but if it is not the tray of circulating air and interlayer, the distribution of fluorescent powder is uneven, resulting in uneven spot, and may cause the overflow of fluorescent powder. The oven is shown in Fig. 6.20.



Fig. 6.20. Oven.

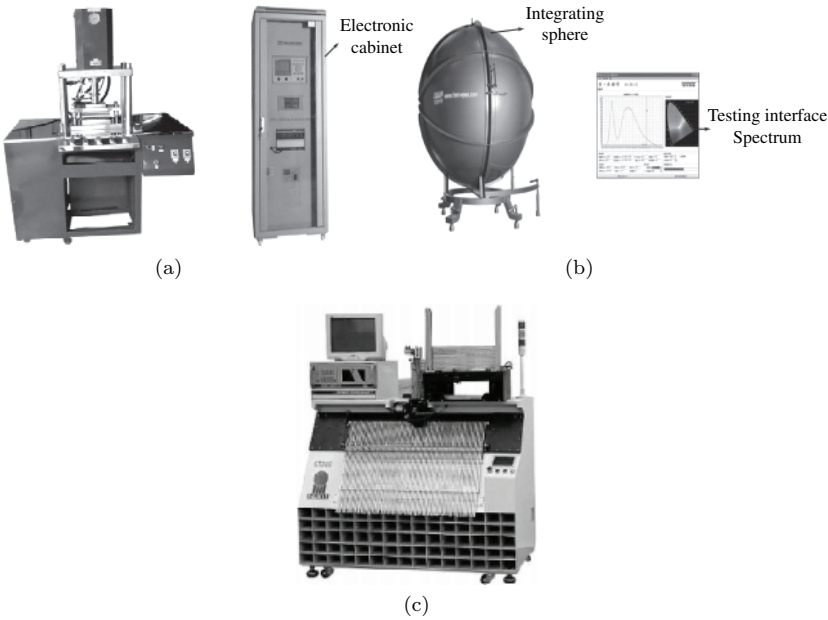


Fig. 6.21. Other equipment: (a) hydraulic press; (b) testing machine; and (c) spectrophotometer.

6.3.2.9. Other equipment

Other LED packaging devices include hydraulic press, foot cutter, testing machine and spectrophotometer, as shown in Fig. 6.21.

6.4. Fluorescent Powder Coating Technology

In the design of white LED, the most important step is to point phosphor, which is the key to the formation of white light. The wavelength of the chip is 460–470 nm, and the selected phosphors are also in this band. The dot phosphors are divided into two very important operations — the fluorescent powder and the coated phosphor.

In the process of phosphor-to-white LED, an important link is the phosphor coating process. The fluorescent silica gel fluids flow out of the spot coating machine and spread on the chip surface under the action of surface tension. After the curing process, the phosphor silica gel layer is finally formed. The coating process of phosphor glue is essentially a two-phase flow process. The final geometrical morphology of phosphor layer is determined by this process, and the final optical performance of LED is determined by the morphology of phosphor layer. The function of LED packaging phosphor is light-color recombination. If the phosphor with white light is not uniform in thickness or shape, it will lead to partial yellowing or bluing of the emitted light, resulting in non-uniform spot, thus affecting the performance of white LED. Therefore, the concentration, thickness and coating shape of phosphors must be strictly controlled.

At present, the traditional phosphor coating process, i.e. the point phosphor mixed adhesive process, is mainly used in our country. The phosphor is directly coated on the chip surface. The phosphor powder and resin (e.g. epoxy resin, organosilicon resin, etc.) are mixed in a certain proportion. After “expanding”, “spinning” and “fixing”, the phosphor resin mixture prepared in advance is coated on the surface of the LED chip between “lead-to-stand bonding” and “packaging” processes. Other processes are basically the same as those of monochrome LED packaging. The specific process of LED phosphor layer is to make YAG yellow phosphor and transparent resin into yellow phosphor slurry according to a certain proportion, stirring evenly, under the microscope, the phosphor slurry is coated on the chip surface with a fine needle or the phosphor is coated on the chip surface with an automatic dispensing machine.

6.4.1. *Mix phosphor powder*

The common phosphors are powdery, so they cannot be directly covered on the chip. Only when the phosphors are dispersed in a kind of glue, and then the mixed phosphor glue is baked, can they be covered on the blue chip.

6.4.1.1. *Glue selection*

The selected solvent must not destroy the structure of the phosphor itself, and the solvent cannot react with the phosphor chemically. Commonly used glues are epoxy resin, acrylic resin and silicone resin, which form suspension with phosphor.

6.4.1.2. *Preparation of suspension*

The materials selected here are yellow phosphors and epoxy resins in corresponding bands. According to the principle of white light emission, too much phosphor will cause the white light to be yellow, and too little phosphor will make the white light to be blue. Therefore, the fluorescent powder should be prepared reasonably according to its luminous efficiency. But the finished facula encapsulated with phosphor and epoxy resin are blue, white and yellow. The reason for this spot formation is that the phosphor is not uniformly stimulated by blue light, that is to say, the fine particles of the phosphor are not completely stimulated by blue light. In order to solve the problem of excitation, it is necessary to introduce a substance such as diffuser, which can enhance the efficiency of blue light excitation phosphors, thereby enhancing the luminescence efficiency of phosphors. Experiments show that the diffuser improves the spot, making the spot no longer one by one, but new problems arise. Although the spot presents a color as a whole, there is a layer of yellow on the outer ring. In order to improve the yellow circle, we must know the reason. By dissecting the finished LED, we can see the precipitation of phosphor, as shown in Fig. 6.22.

Through theoretical analysis, it is known that this phenomenon is caused by the large proportion of yellow light. Firstly, it is necessary



Fig. 6.22. Profile of fluorescent powder dried in solution.

to change the ratio of phosphor glue to find a suitable ratio to improve the yellow circle, and then, the problem of phosphor precipitation. From Fig. 6.22, we can see that the thickness of phosphor covering the gap between the chip and the support cup is much thicker than the thickness of the chip surface. This is because epoxy volatilizes part of the resin during baking. Epoxy resin is two-component: one is resin and the other is curing agent, which belongs to anhydride. The curing agent acts as a crosslinking agent to solidify small molecules. The reaction between curing agent and resin is exothermic reaction, while the heat conductivity of epoxy resin is very poor and its viscosity is very high. Therefore, the heat generated is not easy to dissipate, the volume becomes larger and the density decreases, which makes the phosphor precipitate easily. In addition, the size of the chip is different from that of the bottom of the bracket cup, which can easily lead to high phosphor concentration around the chip. The uneven distribution of phosphor concentration will result in uneven distribution of color temperature of white LED, which makes the brightness and spot of white LED unable to achieve the desired effect. How to improve the uneven distribution of phosphors caused by precipitation is a problem for further study. In theory, two aspects can be improved: (1) The production process, that is, in the production process, in a very short interval of time uniform stirring, and the speed of spot phosphor is accelerated, and the connection time with the next link is also tightened. Spotted phosphor semi-finished products quickly enter the baking process. (2) Adding a new substance makes the phosphor easy to maintain a good uniform mixing state at high temperature. Surfactant is introduced into phosphor solution, which can adsorb organic matter partly and inorganic surfactant partly. After repeated experiments,

the optimal mass ratio of phosphor, surfactant, diffuser and epoxy resin is 10:5:3:100.

6.4.2. *Phosphor coating*

The outer package of pin-type white LED has a molding die. The epoxy resin sealed at the top is made into a certain shape, which has the following functions: first, to protect the core from external erosion. Second, to use different shapes and material properties (with or without dispersant), to function as a lens or diffuse lens to control the divergence angle of light. However, the “lens” formed by epoxy resin cannot be adjusted. In order to achieve better light efficiency, it is necessary to design a “lens” formed by the coated phosphor. The phosphor can form three kinds of lens on the cup surface of the support: concave lens, plane lens and convex lens. According to the light radiation pattern of the two-layer lens, the convex lens is selected. The angle of the convex lens is the same as that of the lens formed by the outer packaging glue. In this way, the light emitted by the chip can be emitted vertically, and the light emittance can be improved. However, the coating method of phosphor is still not perfect, and the light intensity distribution of four sides around the chip is also different. Although some adjustments have been made to the composition and proportion of the phosphor glue, the precipitation of the phosphor can only be improved, but cannot be completely solved. This coating method affects the color temperature and color coordinates of white LED. This problem can be solved if the phosphor can be completely thin covered on the chip. But for pin white LED packaging process is difficult to achieve, and to suit the production and sales of factories, this coating technology is inappropriate. This idea can be achieved for high-power packaging. In high-power white-light LED, the requirement of light-emitting efficiency is high, so large-scale LED chips with 10 times larger area than small chips (about 1mm^2) are used. At present, the packaging of high-power white LED mainly adopts flip-chip. The flip-chip is to flip-chip the GaN LED grains on the radiator board, and make a

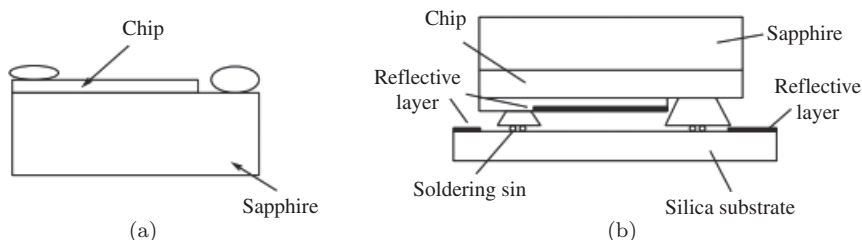


Fig. 6.23. Structural diagrams of two kinds of power chips: (a) forward-mounted high-power chips and (b) flip-mounted high-power chips.

reflective layer with high reflectivity above the P-electrode. That is to say, the light originally emitted from the top of the component is derived from other light-emitting angles of the component, while the light is taken along the end of the sapphire substrate. Figure 6.23 is a schematic diagram of two kinds of power chips: forward and flip chip. This packaging method reduces the light loss on the side of the electrode, and can obtain about two times the light output of the forward mode. Because there is no hindrance of gold wire pad, it is helpful to improve the brightness.

Compared with the advantages and disadvantages of the two chips, flip chip is used in the packaging of high-power white-light LED to replace the traditional high-power chip. The phosphor coating technology of high power white LED is to apply the phosphor evenly on the surface instead of around the chip. In this way, the phosphor mixture solution is directly applied to the chip, so the glue used is no longer epoxy resin, because the fluidity of epoxy resin is strong. If the phosphor is mixed with traditional epoxy resin, the phosphor solution will spill over from the chip surface. Auto-forming UV glue can be selected, and the UV glue and common phosphor can be blended evenly according to a certain weight ratio. The prepared raw materials are added to the dispensing machine to dispense the coating on the high power light-emitting diode chip, so that the coating thickness can be controlled from 0.5 to 0.6 mm. The coated chip was cured by ultraviolet lamp, and the curing process was completed.

6.5. LED Heat Dissipation Technology

Because of its low power, the heat generated by the early LED can be emitted into the air with the packaging lead frame or packaging material, without special consideration of its heat dissipation. With the continuous development of LED technology, especially in the wide application of high-power LED, LED has been upgraded from micro-watt level to watt level or even 100-watt level. The heat generated by LED has become one of the most important factors seriously affecting the performance and life of LED. The thermal control of LED has also become a key issue that needs priority consideration in packaging and application of LED.

The problem of thermal control is first determined by the semiconductor characteristics of the LED itself. Considering the stability of materials, compared with traditional light sources, the heat resistance of LED chips and their packaging materials is poor. Taking incandescent lamp as an example, under normal working conditions, its filament can withstand the high temperature of over 2000°C, while the temperature of common LED nodes cannot exceed 120°C. Even among the latest devices introduced by Lumileds, Nichia, CREE, the highest node temperature cannot exceed 150°C. In addition, the size of LED is relatively small, and it has high heat flux under the same power. For example, CREE DA1000 high power chip, the chip area is only 1 mm × 1 mm, the maximum driving current of the chip (1000 mA), the voltage drop is about 3.5 V, and the thermal power is 2.5 W except the optical power. The corresponding heat flux density is as high as 250 W/cm².

Generally, the luminous wavelength of LED varies with temperature, and the spectral width increases, which affects the brightness of the color. In addition, when the forward current flows through the PN junction, the heating loss causes temperature rise in the junction area. Near room temperature, the luminous intensity of LED decreases by about 1% for each temperature rise. It is very important to maintain the color purity and luminous intensity when encapsulating heat dissipation. The driving current is limited to about 20 mA. However, the light output of LED will increase with the increase of current. The driving current of many power

LED can reach 70 mA, 100 mA or even 1 A level. Therefore, it is necessary to improve the packaging structure, new design concept of LED packaging and low thermal resistance packaging structure and technology to improve the thermal characteristics. For example, using large area chip flip-chip structure, selecting silver glue with good thermal conductivity, increasing the surface area of metal support, and directly mounting the silicon carrier of solder bumps on the heat sink. In addition, the thermal design and thermal conductivity of PCB are also very important in application design.

6.5.1. *Source of heat*

For general lighting use, a large number of LED components need to be integrated into a module to achieve the required illumination. But the conversion efficiency of LED is not high, only about 15–20% of electric energy is converted to optical output, and the rest is converted to thermal energy. Heat is one of the greatest threats to LED, which affects the electrical performance of LED and eventually leads to the failure of LED. How to keep the LED reliable for a long time is the key technology of high-power LED device packaging and system packaging.

For light-emitting diodes consisting of PN junctions, when the forward current flows through the PN junction, the PN junction has heat loss, which radiates into the air through bonding, encapsulating materials, heat sink, etc. In this process, each part of the material has a thermal impedance to prevent heat flow, that is, thermal resistance, which is a fixed value determined by the size, structure and material of the device. Let the thermal resistance of the light-emitting diode be $R_{th} (^{\circ}\text{C} / \text{W})$ and the thermal dissipation power be $P_D (\text{W})$. At this time, the temperature of PN junction caused by the thermal loss of the current rises to

$$\Delta T = R_{th} \times P_D. \quad (6.2)$$

PN junction temperature is

$$T_j = T_a + R_{th} \times P_D, \quad (6.3)$$

where, T_a is the ambient temperature.

6.5.2. *Effect of heat on LED*

The heat generated in the LED luminescence process will cause the LED module temperature rise, when the temperature rises, will cause the following effects: At the same time, due to the increase of temperature rise caused by heat loss, the brightness of LED will no longer continue to increase proportionally with the current, that is, it shows the phenomenon of thermal saturation. The light wavelength shift, as the junction temperature rise, light-emitting peak wavelength will drift towards long wave, 0.2–0.3 nm/°C, for by blue chip coated YAG phosphors mixture of white LED, blue light wavelength drift, will cause the mismatch of and phosphor excitation wavelength, thus reducing the whole white LED luminous efficiency, and leads to a change in the white light color temperature, severely reduce the service life of leds, accelerate the LED light failure.

6.5.3. *Heat dissipation mechanism and solution of LED*

6.5.3.1. *Heat dissipation mechanism*

There are three basic ways of heat dissipation: heat conduction, convection and radiation. Compared with other solid semiconductor devices, LED devices are more sensitive to temperature. Due to the limitation of the working temperature of the chip, the chip can only work below 120°C, so the thermal radiation effect of the device can be neglected. Conduction and convection are important to heat dissipation of LED. From the thermal energy analysis, it is assumed that $Q = \text{divergent power } (Pd) = Vf * If$, and the relative variations of Vf and If are relatively small. Therefore, in the heat dissipation design, the heat transfer should be considered first, and the heat is transmitted from the LED module to the radiator in advance.

(1) Heat conduction

Figure 6.24 is a schematic diagram of heat conduction.

Firstly, considering that the heat source is uniformly loaded on the whole surface of the thermal conductive material, it is

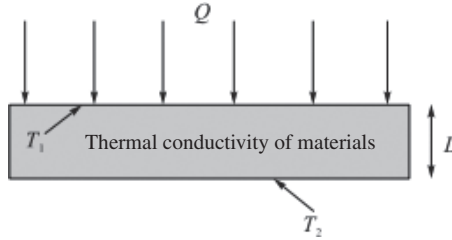


Fig. 6.24. Schematic diagram of heat conduction.

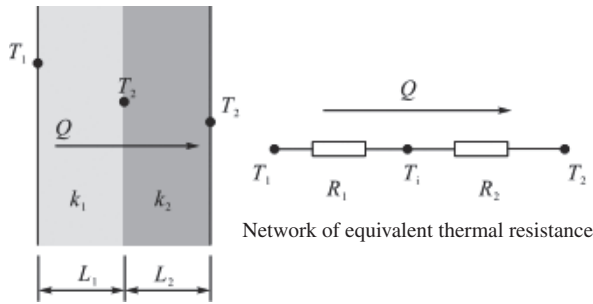


Fig. 6.25. Heat conduction of two-layer materials.

known from the Fourier heat conduction law that the heat flux is proportional to the temperature gradient:

$$q = Q/A = k(T_1 - T_2)/L = k(\Delta T/\Delta x). \quad (6.4)$$

where, K is the thermal conductivity, A is the area, x is the thickness of the thermal conductive material, and q is the heat flux density, representing the dissipated power per unit area. For multilayer composites, the total thermal resistance can be reduced to

$$\frac{\Delta T}{Q} \sim \sum_i \frac{\Delta x_i}{k_j}. \quad (6.5)$$

For example, the heat conduction of two-layer materials in Fig. 6.25 is illustrated:

$$R = R_1 + R_2 = L_1/(k_1 A) + L_2/(k_2 A). \quad (6.6)$$

From Eqs. (6.5) and (6.6), the basic method of heat dissipation can be roughly obtained: reducing the thickness of materials and selecting materials with high thermal conductivity. However, high thermal conductivity materials, such as copper, will also introduce relatively large residual stresses, and when bonding, a thicker bonding layer is needed. This will inevitably increase the excess contact thermal resistance. The thermal resistance formulas of packaging are mostly described by Eq. (6.7):

$$R_{ja} = (T_j - T_a)/Q, \quad (6.7)$$

where, T_j is the junction temperature of the chip, T_a is the ambient temperature, and Q is the heating power of the chip. At the same heating power and ambient temperature, the greater the thermal resistance, that is to say, the more heat cannot be released from the package, but accumulated inside the chip, the faster the junction temperature T of the chip rises, the worse the reliability.

(2) Convection

Figure 6.26 is a thermal convection analysis.

The heat exchange occurs at the interface between solid and fluid. The heat transfer on the surface is carried away by the flow of fluid. The convective heat exchange formula is derived from Newton's cooling law:

$$q = h(T_1 - T_2)A^*/A, \quad (6.8)$$

where, A is the cross-sectional area of the material.

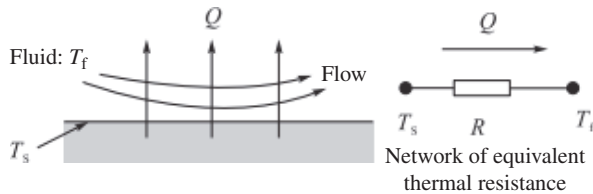


Fig. 6.26. Thermal convection analysis.

The heat resistance formula of convective heat exchange is obtained from Eq. (6.8):

$$R = \Delta T/Q = \Delta T/qA = 1/hA^*, \quad (6.9)$$

where, h is the coefficient of heat conduction and A^* is the area of heat convection. Because the heat exchanged by convection is proportional to the surface area of convection, it is necessary to optimize the structure of microfluidic channels to increase the heat dissipation area A^* , and to increase the heat transfer coefficient h .

By combining Eqs. (6.5) and (6.9), the heat transfer mechanism of heat conduction and convection can be obtained. The total thermal resistance formula is

$$\frac{\Delta T}{q} = \frac{1}{f} \cdot \frac{1}{h} + \sum \frac{\Delta x_i}{k_i}, \quad (6.10)$$

where, f is A^*/A ; A^* is the area involved in heat convection and A is the cross-sectional area of the material, as a surface enhancement factor, which is related to the internal structure of the microfluidic channel.

Equation (6.10) can be a preliminary estimate encapsulation device of the largest temperature difference ΔT , namely $(T_1 - T_2)$, or in a cooling conditions, packaging can hold the biggest density of heat convection.

6.5.3.2. Solution to heat dissipation problem

For power LED, the driving current is generally more than several hundred milliamperes. The current density of PN junction is very high, so the temperature rise of PN junction is very obvious. For packaging and application, how to reduce the thermal resistance of the product and make the heat generated by PN junction emit as soon as possible can not only improve the saturated current of the product, improve the luminous efficiency of the product, but also improve the reliability and life of the product. In order to reduce the thermal resistance of products, the choice of packaging materials is particularly important, including brackets, substrates and filling materials. The thermal resistance of each material should be low,

which requires good thermal conductivity. Secondly, the structure design should be reasonable, the thermal conductivity of each material should be matched continuously, and the thermal conductivity between materials should be good, so as to avoid conduction. Heat dissipation bottlenecks are created in the heat channel to ensure that heat is emitted from the inner to the outer layers.

(1) Selection of connecting materials from chip to substrate. Silver glue is commonly used to connect chips and substrates. However, the thermal resistance of silver glue is very high, and the internal structure of silver glue cured is epoxy resin skeleton and silver powder filled thermal conductive structure. This kind of structure has very high thermal resistance, which is very harmful to the stability of heat dissipation and physical properties of devices. Therefore, it is reasonable to choose tin paste as adhesive material.

(2) Material selection of support (base plate). Table 6.4 is the thermal conductivity of common substrates and supports. It is shown that the thermal conductivity of silver, pure copper and gold is higher than others, but the price of silver, pure copper and gold is higher. In order to achieve a good cost performance ratio, the substrates are made of copper or aluminium.

(3) Selection of the external cooling device of the substrate and selection of the connecting material between the substrate and the external cooling equipment. Most of the losses of high-power LED devices become heat when they work. If no heat dissipation measures are taken, the temperature of the chip can reach or exceed the allowable temperature saving, and the device will be

Table 6.4. Thermal conductivity of common substrate materials.

Materials	Platinum	Silver	Tin	Zinc	Copper	Gold	Aluminum	Aluminum alloy (60Cu-40Ni)	Aluminum alloy (87Al-13Si)
Thermal conductivity [W/(m · K)]	71.4	427	67	121	398	315	236	22.2	162

damaged. Therefore, heat dissipation devices must be added. The most common way is to install the power device on the radiator, using the radiator to dissipate heat to the surrounding space, its main direction of heat flow is from the chip to the bottom of the device, through the radiator to dissipate heat to the surrounding space. The radiator is made of aluminium alloy sheet by stamping process and surface treatment. The surface treatment includes electrophoretic coating or black oxidation treatment. The purpose is to improve the heat dissipation efficiency and insulation performance. The smaller the heat resistance of the radiator itself is, the faster the convection speed is. In terms of interfacial thermal resistance, interstitial air is the biggest obstacle. Although the gap between the substrate and the radiator can be observed with the naked eye is very small, there are actually tiny gaps due to the uneven surface of the material. Because the interface thermal resistance of air is very large, which is not conducive to diffusion, it greatly increases the overall interface thermal resistance. The method of reducing the thermal resistance of the interface is to increase the smoothness of the material surface, reduce the capacity of air and apply contact pressure. Therefore, on the filler material of the base plate and the external radiator, the thermal conductive silicone resin is chosen.

6.5.3.3. *Refrigeration devices*

The traditional refrigeration methods include air refrigeration, water cooling, heat pipe refrigeration, and Partel effect element refrigeration (semiconductor refrigeration). Now some new methods have been put forward, such as ultrasonic refrigeration, superconducting refrigeration, and the effective integration of various refrigeration methods into one device. Following is a brief introduction of several refrigeration methods.

(1) Air cooling

The coefficient of heat conductivity of heat sink can be changed by several methods. The most popular method is to speed up the flow through heat sink. But when the air velocity is increased to 10 m/s, noise will be introduced. Another method is to change the shape of

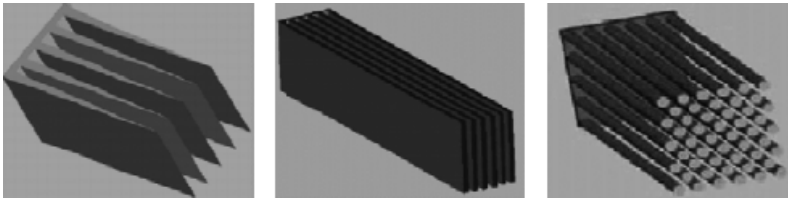


Fig. 6.27. Several heat sink shapes.

the heat sink to expand the effective heat dissipation area, as shown in Fig. 6.27. Radiator shapes can be designed into a variety of array shapes, such as cylindrical arrays, strip arrays or pyramid shapes.

Usually a combination of radiator and fan is used. The radiator transmits heat from the chip to the radiator through close contact with the chip surface. The radiator is usually a good conductor with many blades. Its fully expanded surface makes the heat convection greatly increased. At the same time, the circulating air can also take away more heat energy. Fans are designed to make cooling more efficient and less noise.

(2) Water cooling

The water cooling system is composed of pumps, heat sinks, water pipes and other components. The pump is responsible for driving the water circulation. The heat on the chip is transferred to the water. The liquid flow is used to take the heat away. The water pipes transfer the hot water to the heat sink. Heat sink and chip are not in one, can effectively improve the heat dissipation capacity, heat sink plays a role in heat dissipation. As shown in Fig. 6.28, a hollow metal disc connects with the chip. The liquid flows through the groove inside the chip. The chip transmits heat to the chassis and then transfers heat to the liquid. Then the liquid flows through the heat sink, where it releases heat into the air. When cooled, the liquid enters the chassis again. In addition, the micro-structure of the micro-channel can increase the contact area between the liquid and the heat sink, thus greatly increasing the temperature drop and prolonging the service life of the device.

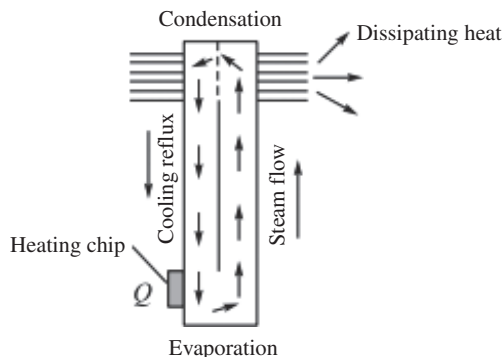


Fig. 6.28. Schematic diagram of heat pipe cooling.

In some cooling devices, heat pipes are used to dissipate heat, and heat pipes are used to remove heat from the surface of central processing unit (CPU) or electronic chips. Coolants in heat pipes become gas after heating, rise in heat pipes, and when they reach the upper part, they are cooled by flowing air, heat is carried away by air, and coolants become liquid again, and flow downward. It's going round and round.

(3) Thermoelectric refrigeration

Thermoelectric refrigeration, also known as thermoelectric refrigeration, or semiconductor refrigeration, is a refrigeration method utilizing thermoelectric effect (that is, Peltier effect). Semiconductor refrigerators have the advantages of high refrigeration density, compatibility with IC process, no moving parts, no wear and tear, and compact structure, which can improve the integration.

As shown in Fig. 6.29, when an N-type semiconductor element and an P-type semiconductor element are connected to form a thermocouple and connected to a DC power supply, the temperature difference and heat transfer will occur at the junction. At one junction above, the current direction is $N \rightarrow P$, the temperature drops and the heat is absorbed, which is the cold end. At the other junction below, the current direction is $P \rightarrow N$, the temperature rises and the heat is released, so it is the hot end.

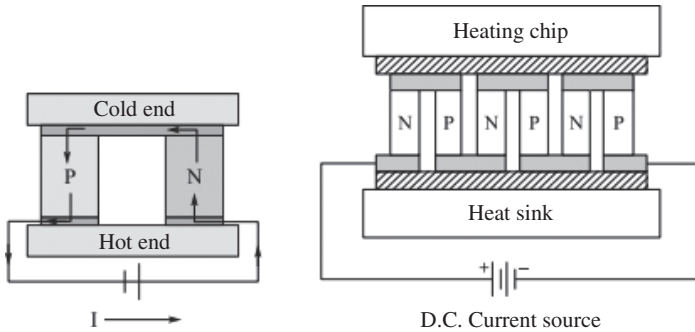


Fig. 6.29. Schematic diagram of thermoelectric refrigeration.

The Pelter effect of metal thermocouple can be qualitatively explained by the phenomenon of contact potential difference. Because of the contact potential difference, the electrons at the junction undergo a sudden change in potential. When the contact potential difference is in the same direction as the external electric field, the electric field force does work to increase the electronic energy. At the same time, the collision between the electron and the crystal lattice transforms this energy into an increment of the energy in the crystal. As a result, the temperature of the bonding position is increased and the heat is released. When the contact potential difference is opposite to the external electric field, the electrons work against the electric field force, and their energy comes from the crystal lattice at the junction. As a result, the temperature at the junction decreases and heat is absorbed from the surrounding environment.

In order to further improve the efficiency of thermoelectric refrigeration, multi-stage thermoelectric refrigeration (Fig. 6.30) is proposed to increase the heat exchange between the integrated heat sink and the external environment.

6.6. Optical Structure of LED

LED optical structure (optical functional structure) mainly includes light extraction structure, light conversion structure and light distribution structure. It involves the whole process of industrial chain

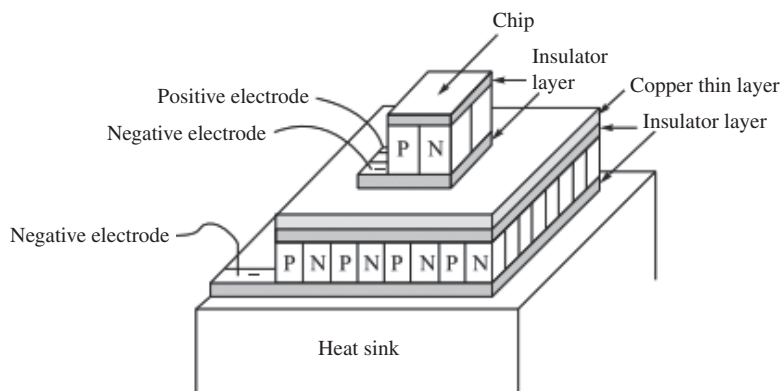


Fig. 6.30. Two-stage thermoelectric cooler and heat sink are integrated as a whole.

from epitaxy to terminal application. Its main function is to improve the luminous efficiency of light source, the color performance of light source, and the utilization efficiency of light energy on the target irradiation surface. In theory, the efficiency of LED devices can reach 85% from current injection to photon generation. In practice, the efficiency of electro-optic conversion of LED devices is only 30–55%. A large amount of light is confined to the inner part of the device package, and finally converted into heat energy. The issue of optimizing the LED chip and packaging structure to improve the probability of light escaping from the package is collectively referred to as the problem of light enhancement.

The main material of LED chip is GaN or AlGaInP (refractive index $n = 2.5\text{--}3.5$). Compared with the packaging agent (epoxy resin or silica gel refractive index $n = 1.4\text{--}1.6$), the difference of refractive index is large. When the normal angle between the emitted light from the chip and the interface is larger than the critical angle (24–40 degrees), the total internal reflection of the light will occur. It is found that the probability of total light reflection in the chip can be effectively reduced and the external quantum efficiency of the chip can be improved by means of patterned sapphire substrate (PSS), optimizing the chip shape, surface modification of the chip, and fabrication of photonic crystals on the chip surface.

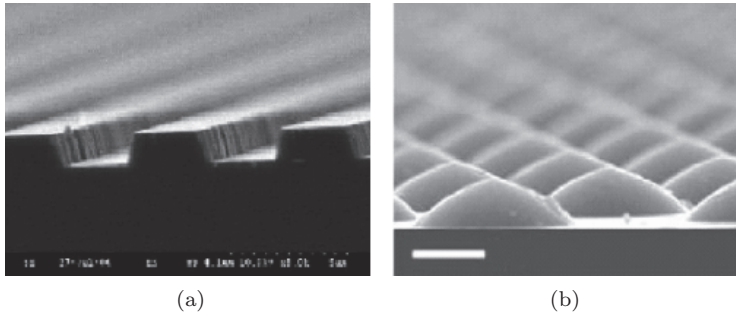


Fig. 6.31. InGaN/GaN multiple quantum well LED devices: (a) parallel groove pattern structure and (b) microlens array pattern structure.

As shown in Fig. 6.31, InGaN/GaN MQW LED devices are fabricated. Compared with the traditional planar substrate devices, the patterned substrate structure can reduce the dislocation density in the epitaxial layer, improve the crystal growth quality, and release the light guide phenomenon inside the chip. It can improve the reverse bias characteristics of the device without changing the forward voltage drop, and improve the light output efficiency.

The influence of geometric parameters of the patterned sapphire substrate on the output light intensity of GaN-based LED is studied in detail. It is found that the inclined surface of the cone platform can effectively improve the light intensity of LED chip at 25° – 60° , and the light intensity reaches its maximum value at around 33° . At the same time, the light output enhancement becomes significant with the increase of the density of the cone platform array.

A special shape chip is designed, and the side of the chip is processed into a positive or inverted trapezoid, or a v -shaped reinforced structure is processed on the surface of the chip. The experimental results show that the brightness of the chip after special treatment is significantly improved compared with the traditional shape LED chip, and the highest brightness is more than 200%.

Surface modification is also an important means to reduce the total internal reflection of light in the chip. The growth process of Mg-doped p-GaN was reduced to 800° , and the surface with random roughness effect was obtained. The test results showed that the

brightness of LED was increased by 80%. By wet etching, randomly distributed pits or protrusions are processed on the surface of P-type GaN. Compared with the unroughened chip, this process can improve the light output by 29.4–40%. Another effective method to improve the light extraction efficiency of LED chips is to fabricate photonic crystal structures with photonic bandgaps periodically arranged by media with different refractive index on the chip surface. Photonic crystal's contribution to the LED chip light extraction efficiency not only from the role of the photonic bandgap, photonic crystal is equivalent to a “diluted” semiconductor, with encapsulated colloid is closer to the effective refractive index, and enlarged the escape cone on the surface of the chip, and photonic crystals in the lateral spread with strong colors and light scattering effect.

As for LED packaging devices, the colloid (refractive index $n = 14\text{--}16$) and air (refractive index $n = 1$) also have the problem of difficult light emission. The phenomenon of “whisperinggallery” in the high-power LED package is analyzed. It is pointed out that when the surface of the package holder reflector cup is smooth, some light will be reflected along the circumferential direction of the reflector cup. This mode of light propagation limits light energy to the inside of the reflecting cup, and some of the reflection will be absorbed by the surface of the reflecting cup. After sandblasting the wall surface of the LED reflection cup into a surface with diffuse reflection characteristics, a part of the light is scattered and escapes each time it reflects on the wall surface.

6.6.1. LED light conversion structure

Light conversion structure, i.e. phosphor coating structure, is mainly oriented to white light illumination technology of LED. The purpose of this structure is to convert the short wavelength light emitted by the LED chip into the long wavelength light complementary to it (color complementary to form white light). There are two ways to realize it: blue LED excitation phosphor (such as YAG yellow-green phosphor, orange nitride phosphor, etc.) or near-ultraviolet LED excitation full-color phosphor.

Phosphor coating technology and formulation are important factors affecting the light efficiency of LED. The backscattering characteristics of phosphors can make 50–60% of the incident light backscatter. The traditional coating process usually covers the phosphor coating material (powder and silica gel mixture) directly on the LED chip, which will cause a large amount of light absorbed into the LED after powder scattering, resulting in the loss of light energy. If the phosphor is directly covered on the chip, the temperature of the phosphor will rise, which will reduce the quantum efficiency of the phosphor and seriously affect the conversion efficiency of the package. Using the technology far away from phosphor coating can reduce the probability that the backward heat radiation is absorbed by the chip, and the luminous efficiency of the LED can be increased by 7–16%. The temperature of phosphor coating can be reduced by about 16–8°C by far away from phosphor coating, and the conversion efficiency of phosphor can be significantly improved. On the basis of studying the optimization of phosphor coatings, a multi-layer phosphor structure is proposed. The red phosphor layer is separated from the yellow phosphor layer, and the yellow phosphor is placed on the red phosphor. The experimental results show that this phosphor coating structure can reduce the mutual absorption between phosphor coatings and the lumen efficiency of packaging products. It can be increased by 18%.

6.6.2. LED light distribution structure

Light distribution structure. Although the application of LED does not require the light source to be clearly imaged on the target irradiation surface as traditional geometric optical design, the users of the light source hope the higher utilization rate of light energy is the better. This kind of optical design theory, which does not aim at imaging but starts from capability allocation and utilization, is called “non-imaging optical theory”. This theory was developed in the mid-1960s to detect electrons in lambda particle decay in high-energy physics experiments. In order to avoid using multiple large photocells in the experiment, the radiation during attenuation

needs to be convergent to the smallest area possible. With help from mathematical engineers, they designed the first non-imaging optical concentrator — a CPC (parabolic concentrator) reflector. The successful application of CPC reflector marks the birth of non-imaging optics.

In the application of LED, it is often necessary to adjust the LED light path through the theory of non-imaging optical design. For example, the projector light source system needs to converge and evenly illuminate the light source to LCD, LCOS or DMD chips. In the lighting system of street lamps, it is necessary to consider to spread the light source to form rectangular light spots, so as to effectively cover the dark area between lamp posts. In the design of automobile headlamp, it is not only necessary to meet the illumination requirement in the specified area, but also need to take into account the safety of the vehicle.

6.6.3. *Simulation and design of LED packaging*

In the optical simulation of LED package, chip, bracket cup, lens (or filler colloid) are three elements. In package design, chip characteristics remain unchanged, and cup structure, lens structure and properties can be changed through simulation design. Take lehman high-power LED as the sample below, and try to find the best light-emitting efficiency by changing the structure of cup and the shape of lens, and design the packaging product with light-emitting angle of 60° .

The shape of the inner wall of the bracket and the shape of the bowl mold is shown in Fig. 6.32. In order to more intuitively simulate the effect, the bottom side of the cup is provided with a 100% reflection effect instead of being filled as a gel. Figure 6.33 simulates the light intensity distribution. According to the light intensity distribution, the half-angle is 65° , the luminous angle is 130° , and the efficiency is 87.95%.

Change the side of the conical bowl into an arc as shown in Fig. 6.34, with the bottom and side set to 100% reflective effect. Figure 6.35 simulates the distribution of light intensity. According to

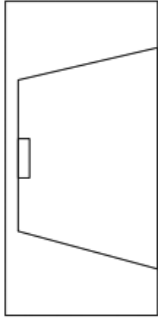


Fig. 6.32. Structural diagram.

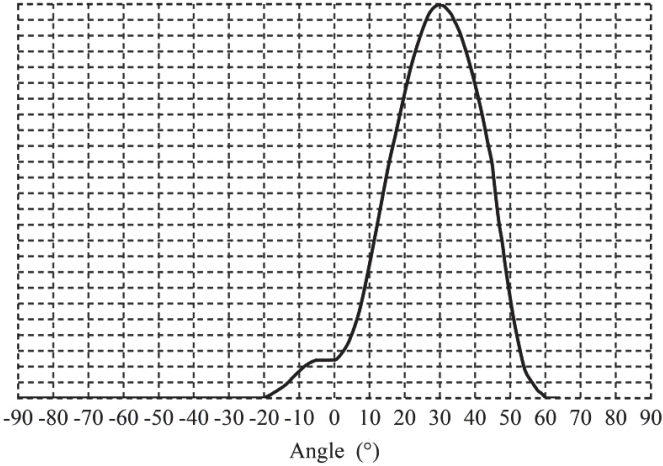


Fig. 6.33. Distribution of luminescence intensity.

the distribution of light intensity, the half-angle is 20° , the luminous angle is 40° , and the efficiency is 100%.

Change the side of the conical bowl to an outward-facing structure as shown in Fig. 6.36, with the bottom and side set to a 100% reflective effect. Figure 6.37 simulates the light intensity distribution. According to the light intensity distribution, the half-angle is 65° , the luminous angle is 130° , and the efficiency is 88.35%.

The change of the colloid volume makes its surface is flush with the cup mouth (Fig. 6.38); the light angle and the light efficiency is

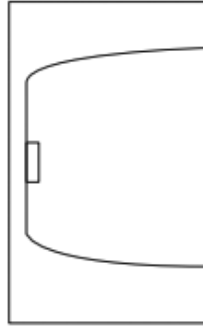


Fig. 6.34. Structural diagram.

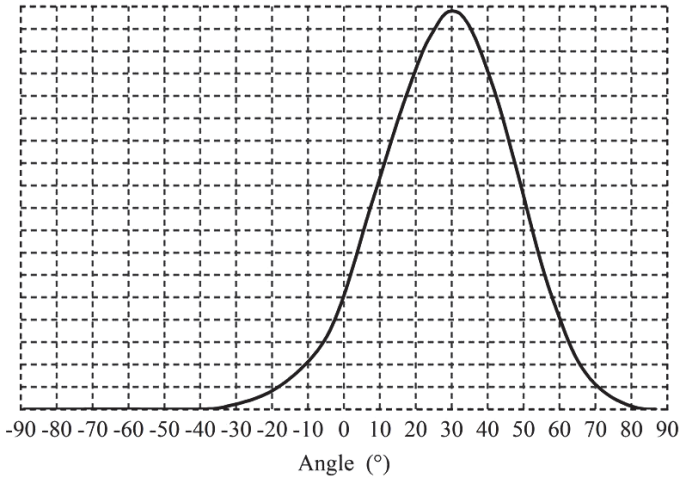


Fig. 6.35. Distribution of luminescence intensity.

simulated by 116° , and 36.15%, respectively (Fig. 6.39); the concave depth of segment is set as 0.15 mm (Fig. 6.40); the light angle and the light efficiency is simulated by 110° , and 36.28%, respectively (Fig. 6.41); the bump height of segment is set as 0.1 mm (Fig. 6.42); the light angle and the light efficiency is simulated by 122° , and 39.53%, respectively (Fig. 6.43); the half-value angle is 30° according the luminescence intensity distribution map (Fig. 6.44), and the light angle is 60° and the light efficiency is 381.51% (Fig. 6.45).

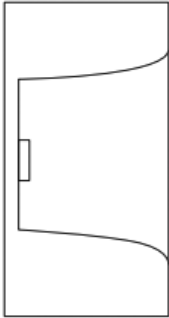


Fig. 6.36. Structural diagram.

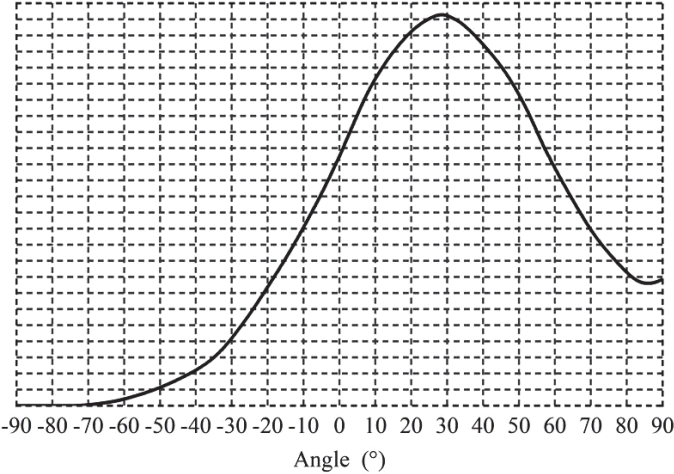


Fig. 6.37. Distribution of luminescence intensity.

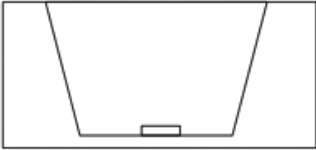


Fig. 6.38. Colloid and cup mouth level.

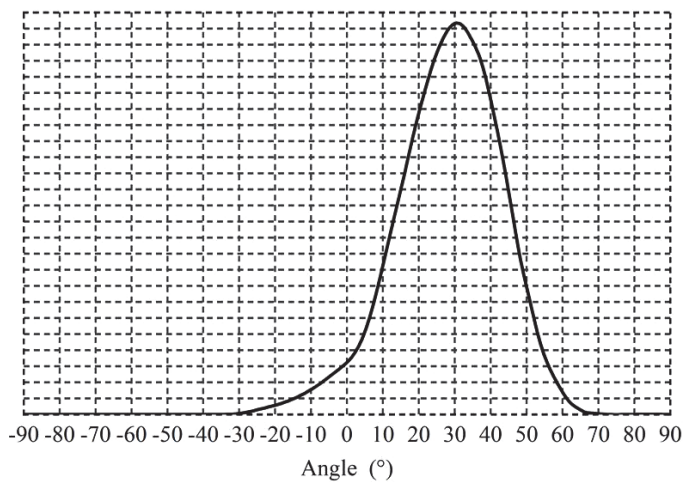


Fig. 6.39. Distribution of luminescence intensity.

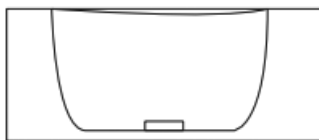


Fig. 6.40. Colloidal indentation map.

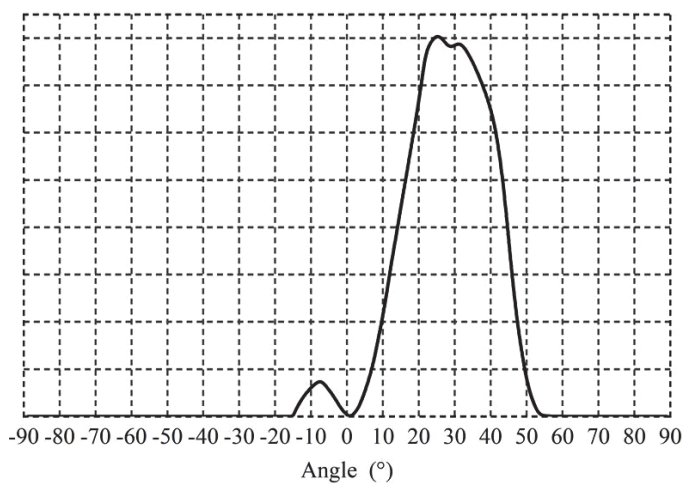


Fig. 6.41. Distribution of luminescence intensity.

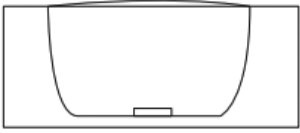


Fig. 6.42. Colloidal bulge diagram.

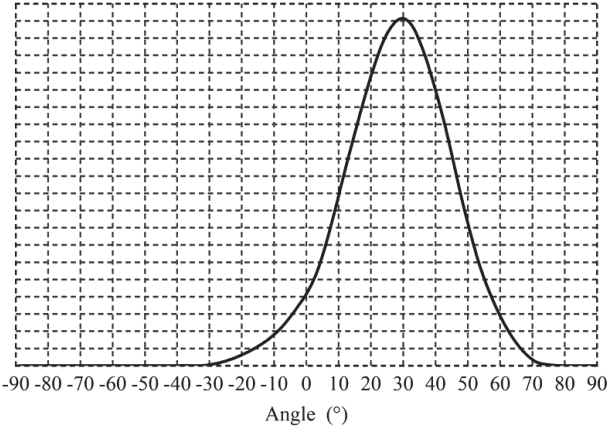


Fig. 6.43. Distribution of luminescence intensity.

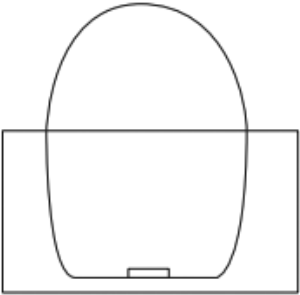


Fig. 6.44. Final sketch.

The influence of silicon refractive index (n_{silicon}), phosphor particle size (D_{phosphor}), reflective cup surface type and reflective rate (R_{specular} , R_{diffuser}) on the luminous flux of white LED was systematically studied by using Lighttools, a professional optical simulation design software.

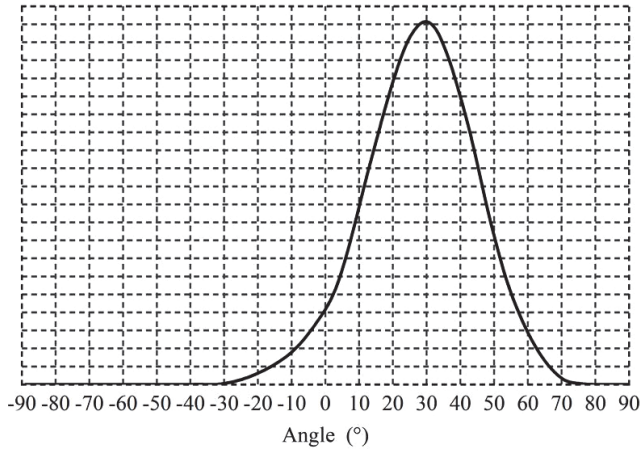


Fig. 6.45. Distribution of luminescence intensity.

Lighttools is an optical simulation software developed by ORA company based on Monte Carlo non-sequence light tracing, which is widely used in the design and simulation of lighting systems. There are mainly three stages when it is applied to the optical simulation of white light LED. In the first stage, inside the LED chip, blue light photons generate optical phenomena such as reflection, transmission and absorption due to different optical properties such as refractive index, transmittance and so on inside the material of each layer and at the interface. The second phase, blue light photons through the chip-phosphor powder adhesive interface after entering phosphor glue inside, when the blue light photon transfers to the surface of phosphor, scattering and absorption happens, which absorb blue light photons according to quantum conversion efficiency of phosphor powder into yellow light photons exit, yellow light photons at the interface between phosphor-silica gel according to the theory of total reflection, less than the critical angle of the light will exist within the phosphor particles and greater than the critical angle of the light will be transmitted into the phosphor glue inside. The third stage, as a result of the phosphor grade for micro/nano size particles, phosphor glue can be treated as individual scattering material, internal blue light photons and yellow light photon transfer due to the effect of

scattering of the particles, when the light to fluorescent powder glue–air interface, also there is a full reflection angle, less than the angle of the light return to phosphor glue inside, is greater than the angle of the light exit to the outside world, finally through the statistics and emergent photons can be concluded that the LED luminous flux.

6.7. Key Technology of Power LED Packaging

At present, 1 W power LED has been produced on a large scale, and 3 W, 5 W and even 10 W high-power LED chips have also been introduced and entered the market. With the continuous improvement of input power of LED chip, large dissipated power brings large heating and high light output efficiency, which puts forward new and higher requirements for LED packaging technology, packaging equipment and packaging materials. China has taken packaging technology as a key project of research and development.

The earliest power LED by HP company launched in 1993 “piranha” packaging structure (Superflux LED), improved to “SnapLED” in 1994. Osram later launched the PowerTOPLED, a PLCC container with a metal frame. Lumileds introduced Luxeon-type LED in 1998. The package structure adopted the separation scheme of thermal channel and electrical channel at first. The flip chip was directly welded on the heat sink with silicon carrier, and new structures and materials such as reflective cup, optical lens and flexible transparent adhesive were adopted to achieve high light efficiency and small thermal resistance. In 2001, American UOE company introduced Norlux series LED with hexagonal aluminum plate as the substrate, which can be assembled with 40 chips in the central luminous area, with a light efficiency of 201 lm/W and a light flux of 100 lm. Osram introduced single chip “GoldenDragon” series LED in 2003, this structure heat sink directly contact with metal circuit board, has good heat dissipation performance, the input power is 1 W. Lanina ceramics has used a metal-substrate low-temperature ceramics (LTCC) multilayer printed circuit board manufacturing technology to produce power LED arrays. Because LTCC technology

connects LED chips directly to a sealed array configuration package, the operating temperature can be up to 250°C.

The flux of conventional LED is a long way from the usual sources of light, such as incandescent and fluorescent bulbs. LED to enter the field of lighting, the first task is to improve its luminous efficiency and luminous flux to the level of existing lighting source, power type white LED is the key to solve the above problems. Due to the continuous improvement of LED chip input power, higher requirements are put forward for the packaging technology of power LED. According to the requirement of light source in lighting field, the packaging of power LED for lighting faces the following challenges: higher luminous efficiency; higher single light flux; better optical properties (optical directivity, color temperature, color rendering, etc.); greater input power; higher reliability (lower failure rate, longer life, etc.); lower luminous flux cost. The requirements of these challenges are fully reflected in the semiconductor lighting blueprint of the United States, as shown in Table 6.5. In fact, it can be achieved step by step by improving the key technology of LED packaging.

6.7.1. *Ways to improve luminous efficiency*

LED luminescence efficiency is determined by the luminescence efficiency of chip and the luminescence efficiency of package structure. The main ways to improve LED luminous efficiency are to improve the luminous efficiency of chip. Effectively extract the light from the chip. The extracted light is efficiently exported out of LED tube. Improve the excitation efficiency of phosphors (for white light). Reduce LED thermal resistance.

(1) Chip selection

LED's luminous efficiency mainly depends on the luminous efficiency of chip. With the continuous progress of chip manufacturing technology, the luminous efficiency of chip is increasing rapidly. Currently, the chips with high luminous efficiency mainly include TS chips of HP company, XB chips of CREE company, WB (wafer bonding) chips, ITO chips, surface roughening chips, reverse bonding chips and so on. According to different application requirements and LED packaging

Table 6.5. US semiconductor lighting development blueprint.

Technical indicators	Lighting LED 2002	Lighting LED 2007	Lighting LED 2012	Lighting LED 2020	Incandescent lamp	Fluorescent lamp
Luminous efficiency/(lm/W)	25	75	150	200	16	85
Life (kh)	20	>20	>100	>100	1	10
Luminous flux/(lm)	25	200	1000	1500	1200	3400
Input power/W	1	2.7	6.7	7.5	75	40
Cost per 1000 lumen/(\$)	200	20	<5	<2	0.4	1.5
Single lamp cost/(\$)	5	4	<5	<3	0.5	5
Chromogenic index (CRI)	75	80	>80	>80	95	75
Permeable lighting market	Low luminous flux requirements	Incandescent lamp market	Fluorescent lamp market	Permeable lighting market		

structure characteristics, suitable chip with high luminous efficiency can be selected for packaging.

(2) Design and material selection of light outlet channel

After the selection of the chip, it is very important to extract and export the light emitted by the chip efficiently to improve the LED's luminous efficiency.

Is the extraction of light as chip light-emitting layer higher refractive index (GaN $n = 2.4$, GaP $n = 3.3$), if the optical channel and surface bonding material refractive index and the chip is large (such as epoxy resin for $n = 1.5$), will result in less total reflection critical angle on the surface of the chip, chip only part of the light can escape through the interface be effectively used, a considerable part of the trapped inside the chip by total reflection of light, as shown in Fig. 6.46, cause the light extraction efficiency is low, directly affect the LED luminous efficiency. In order to improve the extraction efficiency, it is necessary to consider the refractive index matching with the material on the chip surface. Using flexible silica gel with high refractive index as the material for bonding with the chip surface can not only improve the extraction efficiency, but also make the chip and bonding lead get good stress protection. The refractive index changes of the LED encapsulated by GaN flip chip are as follows: active layer ($n = 2.4$) $\rightarrow$ sapphire ($n = 1.8$) $\rightarrow$ epoxy resin ($n = 1.5$) $\rightarrow$ air ($n = 1$). The refractive index of the light outlet channel of the LED encapsulated by GaN-type formal chip changes as follows: active layer ($n = 2.4$) $\rightarrow$ epoxy resin ($n = 1.5$) $\rightarrow$ air ($n = 1$). The refractive

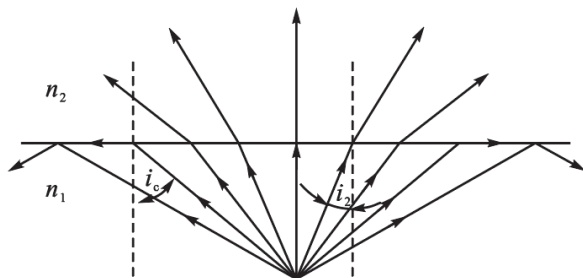


Fig. 6.46. Refraction and reflection of light at interfaces of different media.

index of the light outlet channel of the LED encapsulated by flip chip is better than that of the formal chip, and the light output efficiency is higher. Optical channel material selection is as follows: (1) high transmittance; (2) well-matched refractive index; (3) anti-UV, anti-yellowing characteristics; (4) high-temperature tolerance and good stress characteristics.

A well-designed light outlet channel enables light to be efficiently exported to the outside of the LED tube, including: (1) the design of the reflecting cavity; (2) the design of the lens; (3) the design of the interface between different materials in the light outlet channel and the matching of refractive index; (4) the unnecessary light absorption and leakage in the light outlet channel can be reduced as much as possible.

(3) Process selection of phosphor coating

As far as white LED is concerned, whether the use of phosphor is reasonable or not has a great influence on its luminous efficiency. Firstly, the phosphor with high stimulated conversion efficiency matching the wavelength of the chip should be selected; secondly, the appropriate carrier glue should be selected to prepare the phosphor and make it cover the chip uniformly and effectively in a good coating way. The traditional method of filling the reflecting cup with the fluorescent glue is shown in Figure 6.47.

The shortcoming of traditional coating method is as follows: not only coating uniformity protected, coating fluorescent powder in the chip of an arch, center powder, with a thick layer of four edges and corners in powder thin layer, the white light is not uniform, but also in the reflective cavity is formed in the uneven distribution of

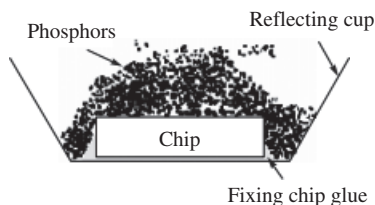


Fig. 6.47. Traditional coating method of phosphor.

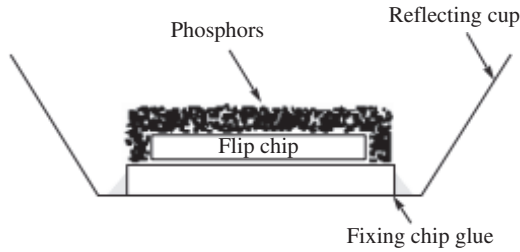


Fig. 6.48. Coating method of phosphor film.

phosphor, unnecessary light leakage loss, and both affect the quality of the light color, and reduce the LED lights.

The above problems can be solved by using planar phosphor film technology, as shown in Fig. 6.48. The process realizes the controllability of the concentration, thickness and shape of the phosphor coating, makes the phosphor layer on the chip surface uniform, and achieves uniform white light.

(4) Heat dissipation design

The junction temperature of the chip increases due to the heating of the LED itself, which leads to the decrease of the luminous efficiency of the chip. Power LED must have a good heat dissipation structure, so that the heat inside the LED can be exported and dissipated as soon as possible, in order to reduce the junction temperature of the chip and improve its luminous efficiency.

The heat resistance is a sign of the heat dissipation structure of LED. Using excellent heat dissipation technology to reduce the thermal resistance of packaging structure will effectively guarantee the improvement of LED luminous efficiency.

6.7.2. Improving the optical characteristics of LED

(1) Regulating the spatial distribution of light intensity

Compared with the traditional light source, the light emitted by LED has stronger directivity. If properly controlled, it can improve the overall lighting efficiency and make the lighting effect better.

In LED lighting engineering, it is usually realized through the following steps: to understand the distribution characteristics of chip luminescence. According to the characteristics of chip luminescence distribution and the requirements of LED final light intensity distribution, the optical channel and reflection cavity are designed. Lens design: consideration of refraction and diffusion of light in the exit channel; design and match the geometric dimensions of each part of the light outlet channel; select the right material and processing technology of light outlet channel.

(2) Improve the uniformity of light color

At present, the most commonly used technology route of LED white light generation is blue chip + yellow phosphor (YAG/TAG). The process method is to mix phosphor and carrier adhesive and then spread onto the chip. In the actual operation, the viscosity of the carrier adhesive is a dynamic parameter. It is difficult to control the coating amount and uniformity of phosphors due to the influence of factors such as the density of phosphors is higher than the carrier gum and easy to produce precipitation. The methods to improve the uniformity of light color are as follows: reasonable selection of phosphor particle size; grasp the viscosity characteristics of carrier adhesive. The preparation process of fluorescent adhesive was improved to prevent the fluorescent powder from settling in the carrier adhesive. High precision phosphor coating equipment is used, and the method and form of fluorescent adhesive coating are improved.

(3) Improve color temperature and color rendering

The color temperature of white LED is mainly controlled by the wavelength of blue chip and stimulated wavelength of phosphor, and the amount and uniformity of phosphor coating. Based on the mechanism of blue chip + yellow phosphor (YAG/TAG) LED white light generation technology route and the characteristics of phosphor, the early traditional white LED is in the high color temperature region (>5500 K), color temperature control is easier to achieve, and color rendering is better ($R_a > 80$). However, in the low color temperature area (2700–5500 K) which is usually required for lighting

applications, the color temperature control of traditional white LED is difficult, and the color rendering is not good ($R_a < 80$), which has a certain gap with the requirements of lighting source. Even if white light with low color temperature can be generated, its color coordinates are far away from the blackbody radiation trajectory (usually above the trajectory), which makes its light color incorrect and color rendering poor. The key to solve this problem is the improvement of phosphor, and the color coordinates of white light emitted by LED can be as close as possible to the blackbody radiation trajectory by adding red phosphor, so as to improve its light color and color rendering. At present, there are four main methods to improve the color rendering of white LED in low color temperature region: choose blue chip with short wavelength as far as possible ($\lambda_D < 460\text{ nm}$); analyze the defects of white LED luminescence spectrum, select appropriate phosphors containing these defects; improve the coating technology of phosphors, to ensure that phosphors get full and uniform excitation; white light generation technology with color dominance.

6.7.3. Increase the single light flux and input power of LED

At present, the pearlescent flux of single LED lamp is too small to be used independently in lighting, and its input power is too small, so it needs more peripheral application circuits to cooperate. In order to enter the lighting field, the single light flux and input power of LED must be increased. The ways to improve the single light flux and input power of LED are as follows: under the premise of a certain input power, improving the luminous efficiency of LED is the most direct way to obtain a larger single light flux; using large area chip to encapsulate the LED and increasing the working current can obtain a higher single light flux and input power; using multi-chip high density integrated package power LED is the most common method to obtain high single light flux and high input power. Among the above three ways, heat dissipation technology is the key. Improving the heat dissipation ability and reducing the thermal resistance of

LED is the fundamental guarantee to improve the luminous flux and input power of LED.

6.7.4. Reduce the cost of LED

High price is the ultimate bottleneck for semiconductor LED to enter the lighting field. As far as packaging technology is concerned, in order to reduce the cost of LED, the following five problems must be solved: mature and feasible technology route; simple and reliable process method; general product design; high product performance and reliability; and high yield.

6.7.5. Improving the reliability of LED

In practical applications, people generally pay attention to the reliability of LED, such as dead light, light decay, color shift, scintillation and life. The input power of power-type LED is high, and the application environment is bad, which puts forward higher requirements for reliability. LED life is the core index of reliability. Accelerated aging life test is needed for LED products. At the same time, in view of the slow degradation of LED performance, there are relevant international standards, most typical of which is the North American and International Lighting Commission system.

References

- Chen M X *et al.*, 2006. Advances in packaging design and research on high-power white LED, *Semiconductor Optoelectron.* 27(6):653–658.
- Cho H K *et al.*, 2006. Light extraction enhancement from nano-imprinted photonic crystal GaN-based blue light-emitting diodes, *Opt. Express* 14 (19):8654–8660.
- Fang L *et al.*, 2011. Study on the fabrication and performance of aluminum substrate for heat dissipation of high power LED, *Mater. Rep. B: Res. Papers* 25(2):130–134.
- Huang S H *et al.*, 2006. Improved light extraction of nitride-based flip-chip light-emitting diodes via sapphire shaping and texturing, *IEEE Photonics Technol. Lett.* 18(24):2623–2625.
- Huang X H *et al.*, 2011. Effect of patterned sapphire substrate shape on light output power of GaN-based LEDs, *Photonics Technol. Lett.* 23(14):944–946.

- Ioselevich A S, Kornyshev A A, 2002. Approximate symmetry laws for percolation in complex systems: Percolation in polydisperse composites, *Phys. Rev. E*, 65(2):95–129.
- Kissinger S *et al.*, 2010. Enhancement in emission angle of the blue LED chip fabricated on lens patterned sapphire (0001), *Solid-State Electron.* 54(5): 509–515.
- Laubsch A *et al.*, 2010. High-power and high-efficiency InGaN-based light emitters, *IEEE Trans. Electron Devices* 57(1):79–87.
- Lee Y *et al.*, 2008. GaN light-emitting diode with deep-angled mesa sidewalls for enhanced light emission in the surface-normal direction, *IEEE Trans. Electron Devices* 55(2):523–526.
- Lee Y *et al.*, 2005. Improvement in light-output efficiency of near-ultraviolet InGaN-GaN LEDs fabricated on stripe patterned sapphire substrates. *Mater. Sci. Eng. B* 122(3):184–187.
- Li B C *et al.*, 2009. Development of packaging design of high-power white LED, *Laser Optoelectron. Progress* 46(9):35–39.
- Li H P *et al.*, 2007. Package and MCPCB for high-power LEDs, *Semiconductor Optoelectron.* 28(1):47–50.
- Lin C C, Liu R S, 2011. Advances in phosphors for light-emitting diodes, *J. Phys. Chem. Lett.* 2:1268–1277.
- Liu L *et al.*, 2007. Promoting of the phosphor-based white-LED and optical properties, *Chinese J. Luminescence* 28(6):890–893.
- Liu Y B, Ding J, 2008. Summarization on packaging technique of power-LED, *Chinese J. Liquid Crystals Displays*, 23(4):508–513.
- Liu Y B *et al.*, 2008. Research of heat release technology of based on power-LED, *China Illuminating Eng. J.* 19(1):269–272.
- Medalia A I, 1986. Electrical conduction in carbon black composites. *Rubber Chem. Techno.* 59(3):432–454.
- Miyauchi S, Togashi E, 1985. The conduction mechanism of polymer-filler particles, *J. Appl. Polymer Sci.* 30(7):2743–2751.
- Quan X, 1987. Investigation of the short-range coherence length in polymer composites below the conductive percolation threshold. *J. Polym. Sci., Polymer Phys. Ed.* 25:1557–1561.
- Sim J K *et al.*, 2012. Characteristic enhancement of white LED lamp using low temperature co-fired ceramic-chip on board package, *Current Appl. Phys.* 12(2):494–498.
- Simmons J G, 1963. Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film, *J. Appl. Phys.* 34(6):1793–1803.
- Tang K *et al.*, 2014. Research status and development trends of LED packaging, *China Illuminating Eng. J.* 25(1):26–30.
- Tu D W *et al.*, 2008. Effect of optical structure on output light intensity distribution in LED package, *Optics Precision Eng.* 16(5):832–838.
- van Beek L K H, van Pul B I C F, 1962. Internal field emission in carbon black-loaded natural rubber vulcanizates, *J. Appl. Polymer Sci.* 6(24): 651–655.

- Wan Z P, Zhao X L, Tang Y, 2012. Situation development of LED chips on light extraction structure, *China Surface Eng.* 25(3):6–12.
- Yan J, Yu Y, 2004. LED's optical encapsulation structure design based on monte carlo simulation method, *Chinese J. Luminescence* 25(1):90–94.
- Zhang Q Y, Huang X Y, 2010. Recent progress in quantum cutting phosphors. *Prog. Mater. Sci.* 55:353–427.

Chapter 7

Transparent Conductive Materials

7.1. Brief Introduction of Transparent Conductive Films

If a thin-film material has more than 80% transmittance in the visible light range (wavelength 380–760 nm), high conductivity and resistivity less than $1 \times 10^{-3} \Omega \text{cm}$, it can be called transparent conductive thin-film. Au, Ag, Pt, Cu, Rh, Pd, Al, Cr and other metals have a certain degree of visible light transmittance when forming thin-films with thickness of 3–15 nm, so they have been used as transparent electrodes in history. However, metal thin-films that absorb too much light have low hardness and poor stability, so people begin to study the formation methods and physical properties of transparent conductive thin-films such as oxides, nitrides, fluorides and so on. Among them, transparent conducting oxide (TCO), which is composed of metal oxides, has become the leading role of transparent conducting film, and its application field and demand have been expanding in recent years. With the vigorous development of 3C industry, the output of flat panel display (FPD) led by LCD has been increasing year by year. At present, it has occupied an important position in the global display market. Among them, indium oxide ($\text{In}_2\text{O}_3:\text{Sn}$), which means indium oxide doped with tin, ITO, is a transparent electrode material for FPD. In addition, SnO_2 and other low-E windows, which can reflect infrared rays on buildings, have become the largest application area of transparent conductive films. In the future, with the increase of functional requirements and the global trend of energy saving, electrochromic

(EC) window (a kind of transparent glass that can change with applied voltage) with both dimming and energy saving effects is also expected to become very important materials, which are widely used in buildings, automobiles and daily life. And the demand for transparent conductive films will be increasing.

As shown in Table 7.1, TCO accounts for the majority of transparent conductive films. This is because TCO has ionicity and appropriate energy gap, and is chemically stable, so it becomes an important material for transparent conductive films.

7.2. Typical TCO Materials

Typical TCO materials are In_2O_3 , SnO_2 , ZnO , CdO , CdIn_2O_4 , Cd_2SnO_4 , Zn_2SnO_4 , $\text{In}_2\text{O}_3\text{--ZnO}$, etc. The energy gaps of these oxide semiconductors are all above 3 eV, so the energy of visible light (1.6–3.3 eV) is not enough to excite valence band electrons to conduction band. Only light with wavelengths of 350–400 nm (ultraviolet) can excite valence band electrons. Therefore, the absorption of light generated by electron migration between bands will not occur in the visible range, and TCO is transparent to visible light.

The resistivity of these materials ranges from 10^{-3} to $10^{-1} \Omega \text{ cm}$. If Sn is further added to In_2O_3 to become ITO, adding Sb and F to SnO_2 , adding In, Ga or Al to ZnO can increase carrier concentration to $10^{20}\text{--}10^{21} \text{ cm}^{-3}$ and reduce resistivity to $10^{-4}\text{--}10^{-3} \Omega \text{ cm}$. For example, in ITO, Sn with four valence replaces In with three valence, and Ga or Al with three valence replaces Zn with two valence in GZO or AZO, so a dopant atom can provide a carrier. However, in reality, not all dopants are displaced solid solutions. They may exist as neutral atoms between lattices, become scattering centers, or segregate at grain boundaries or surfaces. It is very important for the fabrication of transparent conductive films with low resistance to form displacement solid solution and improve the doping efficiency.

In_2O_3 , SnO_2 and ZnO are three most noticeable TCO materials at present. $\text{In}_2\text{O}_3\text{:Sn}$ (ITO) has become a very important TCO material with the popularity of FPD in recent years because it is a transparent electrode material on FPD. The transparent electrode

Table 7.1. Some commonly used transparent conductive films.

Material	Purpose	Nature requirements
SnO ₂ :F	Low-E glass for cold zone buildings	Plasma wavelength 2 micron (increase the penetration of sunlight infrared region)
Ag, TiN	Low radiation glass for tropical buildings	Plasma wavelength $\leq 1 \mu\text{m}$ (reflecting sunlight infrared region)
SnO ₂ :F	Outer surface of solar cell	Thermal stability and low cost
SnO ₂ :F	EC windows	Chemical stability, high transmittance and low cost
ITO	Electrodes for flat panel displays	Easily etchable, low film forming temperature, low resistance
ITO, Ag, Ag-Cu alloy	Defogging glass (refrigerator, airplane, car)	Low cost, durability, low resistance
SnO ₂	Oven glass	High-temperature stability, chemical and mechanical durability, low cost
SnO ₂	De-static glass	Chemical and mechanical durability
SnO ₂	Touch screen	Low cost and durability
Ag, ITO	Electromagnetic shielding (computer, communication equipment)	Low resistance

material on FPD uses ITO because it has the following excellent properties:

- (1) The resistivity is low, about $1.5 \times 10^{-4} \Omega \text{ cm}$.
- (2) It has strong adhesion to glass substrates, close to TiO₂ or chrome films.

- (3) The transparency is high and the transmittance in the central region of visible light is better than that of SnO_2 .
- (4) Proper resistance to strong acid and alkali.
- (5) Good electrochemical and chemical stability.

SnO_2 film has hardly been used since 1975 because of its poor conductivity compared with ITO. However, due to its excellent chemical stability, SnO_2 film began to be a transparent conductive substrate for amorphous silicon solar cells around 1990. Amorphous silicon solar cells are formed by plasma chemical vapor deposition (CVD), and plasma is formed by SiH_4 gas and hydrogen, which becomes a strong reducing atmosphere. This will reduce the transmittance of ITO from 85% to 20%, while SnO_2 will remain at 70%. Therefore, instead of ITO film, SnO_2 film is used in amorphous silicon solar cells.

In recent years, ZnO (ZnO: Al , AZO) has attracted much attention as a TCO material. Among them, zinc oxide doped with aluminum (AZO) has the most potential as a substitute for ITO. Due to the improvement of the process, the physical properties of ZnO films prepared in laboratory are close to ITO, and zinc is superior to indium in terms of production cost and toxicity, especially the low price of zinc, which is an important point for the popularization of materials. The properties of In_2O_3 , SnO_2 and ZnO are shown in Table 7.2. The principles of conductivity and transmittance of TCO and some properties in Table 7.2 are described in detail later.

7.3. Conductivity of TCO

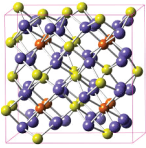
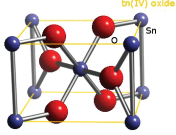
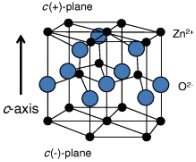
7.3.1. Conductivity principle of TCO

If the material is to be conductive, there must be carriers carrying charges and paths for carriers to move at high speed inside the material. The conductivity of the material can be expressed by

$$\sigma = ne\mu, \quad (7.1)$$

where, n is the carrier concentration, e is the electron charge, and μ is the carrier mobility. When the electron orbital overlap (interaction)

Table 7.2. Properties of In_2O_3 , SnO_2 and ZnO .

Name of material	In_2O_3	SnO_2	ZnO
Crystal structure name	Bixbyite	Rutile	Wurtzite
Crystal structure diagram			
Guideway track area	$\text{In}^{+3}5s$	$\text{Sn}^{+4}5s$	$\text{Zn}4s\text{--}2p$
Valence belt track region	$\text{O}^{-2}2p$	$\text{O}^{-2}2p$	anti-bonding $\text{Zn}4s\text{--}2p$ bonding (The upper part is $\text{O}2p$ and the bottom part is $\text{Zn}4s$)
Energy gap (eV)	3.5–4.0	3.8–4.0	3.3–3.6
Donor level sources	Oxygen holes or Sn dopants	Oxygen holes or interlattice solid solution Sn	Oxygen holes or interlattice solid solution Zn
Dopant	$\text{Sn}^{(4+)}$	$\text{Sb}^{(5+)}$	$\text{Al}^{(3+)}$
Donor level location	When $E_d = E_{d0} - an_d^{1/3}$ (eV), $E_{d0} = 0.093$ eV, $a = 8.15 \times 10^{-8}$ eV cm, $n_d > 1.49 \times 10^{18}$ cm^{-3} , the donor level enters the conduction band and becomes a degenerate semiconductor	Under conduction band 15–150 meV Under conduction band 10–30 meV (Sb-doped)	Under conduction band 200 meV

(Continued)

Table 7.2. (*Continued*)

Name of material	In ₂ O ₃	SnO ₂	ZnO
Mobility [cm ² /(V s)]	103	18–31	28–120
Carrier concentration (cm ⁻³)	1.4×10^{21}	2.7×10^{20} – 1.2×10^{21}	1.1×10^{20} – 1.5×10^{21}
Resistivity (Ω cm)	4.3×10^{-5}	7.5×10^{-5} – 7.5×10^{-4}	1.9×10^{-4} – 5.1×10^{-4}

between adjacent atoms constituting a solid is large, that is, when the orbital expansion in space is large, the carrier is easy to move from one atomic position to another, that is, the mobility is larger.

To explain the origin of TCO conductivity, it can be simply described as follows: when metal atoms are bonded to oxygen atoms, they tend to lose electrons and become cations, while in metal oxides, metal cations with $(n - 1)d10ns0$ ($n > 4$, n is the main quantum number) electronic configuration have an isotropic expansion of their s-orbital domains. If there is a lock-like structure in the crystal, these cations can be very close to each other and their s-orbital domains overlap, then conduction paths can be formed. With the addition of movable carriers (the material itself or from dopants), it is conductive.

7.3.2. Energy band, orbital domain and mobility

If a simple formula is added, the description in Section 7.3.1 can be further explained as follows: the mobility is

$$\mu = e\tau/m^*, \quad (7.2)$$

where, τ is relaxation time (the time from one scattering to the next scattering when the carrier moves), which is related to the crystalline structure, and m^* is the effective mass of the carrier. The smaller the effective mass is, the faster the carrier moves in the electric field, so

μ depends mainly on the effective mass. The definition of effective mass m^* is as follows:

$$m^* = \frac{\hbar^2 \partial^2 \mathbf{E}(k)}{\partial^2 \mathbf{E}(k)}. \quad (7.3)$$

Among them, $\mathbf{E}(k)$ is the energy of the energy band and k is the size of the wave vector. It can be seen that the bigger the bending degree of E curve is, the smaller the m^* is. Near the origin (point) of k space, $\mathbf{E}$ can be expressed as

$$\mathbf{E} = H_{nn} + 2H_{mn} \cos(ka) \approx H_{nn} + 2H_{mn} - 2H_{mn}(ka)^2, \quad (7.4)$$

where, $H_{mn} = \int \phi^*(x_m) H \phi(x_n) dx$ is the interaction between m -orbital and n -orbital, and a is the atomic interval. It can be seen that the larger the electron-orbit interaction between adjacent atoms is, the smaller the m^* is.

For most wide gap oxides, the bottom of the guide band is mainly composed of the cationic air orbit, and the valence band is composed of the occupied oxygen 2p orbit. In N-type transparent conductive materials, the ionic airspace is the electron's moving path; therefore, the enlargement of this airspace is very important for the formation of high-speed moving path. Generally speaking, metal cations with $(n-1)d^{10}ns^0$ ($n \geq 4$, n is the dominant quantum number) electronic configuration have an isotropic expansion of the s-orbital domain. In this crystal structure where cations are close to each other, the overlap between orbital domains is large and a wide conduction band is formed.

Therefore, if we want to get high mobility, we should choose cations with large space expansion and shorten the distance between cations. This principle applies not only to crystals with orderly ion arrangement, but also to amorphous substances.

Cations in oxides interact with oxygen ions to form oxygen-ion polyhedrons, so the distance between cations is related to the stereo configuration of oxygen-ion polyhedrons. As far as conductivity is concerned, in order to form the carrier moving path in the crystal, the polyhedrons must be arranged in a continuous line. The continuous

arrangement of polyhedrons includes vertex-sharing, edge-sharing and surface-sharing, while the distance between ions decreases according to the order of vertex-sharing, edge-sharing and surface-sharing. Therefore, it is easy to understand that the overlap between cationic orbital domains increases according to the order of vertex-sharing, edge-sharing and surface-sharing. In fact, because of the repulsion of Coulomb force between cations, crystals with continuous faces and polyhedrons almost do not exist. Therefore, materials with high mobility are concentrated in the crystal structure of polyhedral locks with common edges. In N-type crystalline conductive oxides, except for ZnO, all crystal structures have a rutile lock structure with oxygen octahedral edges.

Amorphous oxides cannot directly form oxygen-ion octahedral prism-shared crystalline structure, but the cations are also surrounded by oxygen ions. Although the orbital overlap cannot be obtained to the same extent as that of crystals, if the cationic airspace can be fully expanded, the conductive orbital overlap can be obtained. In cations with $(n - 1)d^{10}ns^0$ ($n \geq 4$) electronic configuration, Cd^{2+} and In^{3+} have wide orbital domains. If carriers can be introduced, they will exhibit conductivity.

7.3.3. *N-type and P-type TCO*

The principle of conduction described in Sections 7.3.1 and 7.3.2 is mainly aimed at N-type TCO. In the wide gap oxides without transition metals, there are much fewer P-type conductive materials than N-type ones. The path of P-type TCO holes is at the upper part of valence band, which is mainly composed of the occupied oxygen 2p-orbital domain.

In the typical metal oxide (MO) orbital domain, the bottom of the guide band is mainly empty *ns*-orbital domain of metal cations, while the upper part of the valence band is mainly non-bonding oxygen 2p-orbital domain. The so-called non-bonding means that there is almost no interaction with other elements, when the band expansion is very small, even if there are holes, it will localize. Therefore, there are fewer P-type conductive oxides with wide energy gap.

In order to overcome this problem, a substance containing Cu^+ can be selected as a candidate for the substance with large expansion in the upper valence band. Cu^+ is a closed-shell electronic configuration of d^{10} . There is no optical absorption of transition metal ions due to d-orbital migration, and its d-orbital energy is approximately the same as that of oxygen in 2p-orbital region. Between the two orbital domains with very close energy levels, it is easy to form general bonding orbital domains or mixed orbital domains, so that the energy bands formed by orbital domains can be extended to form the moving path of holes. Most metal cations other than Cu^+ do not have orbital domains similar to the energy of oxygen 2p-orbital domains. Oxygen 2p-orbital domains above valence bands are non-bonding. Therefore, common P-type substances are conductive due to the expansion of valence bands due to the presence of Cu^+ .

7.3.4. *Carrier generation*

Another factor affecting conductivity is the generation and concentration of carriers. TCO is the same as semiconductor. When the material itself is a pure ideal crystal, the carrier does not exist and becomes an insulator. The thermal energy at room temperature is about 30 meV, and the energy gap of TCO is generally above 3 eV, so the carriers will not be thermally excited at room temperature.

Carriers are generated by generalized defects, which include ion deficiency or dopant incorporation. The speed of carrier movement and the size of energy gap (transparency) are related to the material itself, but the generation of carriers is different.

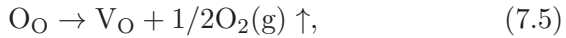
In order to generate carriers in TCO, there must be defects, but the converse may not be true, that is to say, appropriate defects can effectively generate carriers. In ionic crystals such as oxides or halides, the carrier generation reaction of defects is mainly based on the expression proposed by Krögen–Vink. This representation is centered on the valence of each location in the electroneutral and ideal crystalline states, and the electrons or holes are represented by “/” and “.” respectively. For example, in CaO , Mg^{2+} at the original

Ca^{2+} position is represented by MgCa , Al^{3+} by Al'_{Ca} , and K^+ by K_{Ca} . Every case must satisfy the condition of electrical neutrality. Like typical semiconductors, the generation of effective carriers depends on the energy difference between the defect level and the moving path band.

(1) ITO (N-type):

$\text{Sn}'_{\text{In}} \rightarrow \text{Sn}_{\text{In}} + e'$ (due to thermal energy, the reaction produces electron carriers to the right)

Oxygen atom O is replaced by void V:



So Sn'_{In} and VO are both donors.

(2) Cu_2O (P-type):

$\text{V}_{\text{Cu}} \rightarrow \text{V}'_{\text{Cu}} + h'$ (cavity carrier caused by Cu deficiency)

So V_{Cu} is acceptor.

The energy E_d and orbital radius a_d of donor electrons are

$$E_d = -\frac{e^4 m^*}{2(4\pi\epsilon\hbar^2)} = -\frac{m^*}{m} \left(\frac{\epsilon_0}{\epsilon}\right)^2 E_H, \quad E_H = 13.6 \text{ eV}, \quad (7.7)$$

$$a_d = \frac{4\pi\epsilon\hbar^2}{e^2 m^*} = \frac{m}{m^*} \frac{\epsilon_0}{\epsilon} a_H, \quad a_H = 0.529 \text{ \AA} \text{ (Bohr radius)}, \quad (7.8)$$

where, ϵ is the permittivity of the parent material, and m^* is the effective mass of the electron. The larger the ϵ or the smaller the m^* is, the smaller the free energy of the donor is.

In Si, $\epsilon = 12\epsilon_0$, $m^* = 0.3m$, so $E_d = 28 \text{ meV}$, $a_H = 21 \text{ \AA}$, which is close to the experimental V donor level (40–50 meV). The room temperature thermal energy (kT) is about 30 meV, so it can be judged that the donor atoms are mostly free to generate carriers at room temperature. If only point defects are considered, some oxides can be estimated by using the hydrogen atom model.

- (1) The m^* and ε of SnO_2 (rutile structure) perpendicular to and parallel to the C -axis are, respectively, $m_{\perp}^* = 0.229m_0$, $m_{\parallel}^* = 0.234m_0$, $\varepsilon_{\perp} = 14.0\varepsilon_0$, $\varepsilon_{\parallel} = 9.9\varepsilon_0$; the result is close to that of Si.
- (2) TiO_2 is $m^* = 10\text{--}30m_0$, $\varepsilon_{\perp} = 89\varepsilon_0$, $\varepsilon_{\parallel} = 173\varepsilon_0$; $E_d = 5\text{--}52\text{ meV}$ can be obtained.
- (3) MgO or Al_2O_3 is $\varepsilon = 10\varepsilon_0$, assuming $m^* = 10m_0$, $E_d = 1.36\text{ eV} \gg 30\text{ meV}$ can be obtained.

The results of these estimates indicate qualitatively that SnO_2 or TiO_2 with impurities are conductive at room temperature, but MgO or Al_2O_3 are not.

As carriers and their energy levels come from dopants, the solution limit and free energy of dopants have become a concern. Because there is no theory of solution limit, it can only be obtained by trial and error. Dopants are also the cause of scattering. The maximum conductivity may not be obtained with the maximum amount of dopants, so the solution limit may not be so important.

7.3.5. *The relation between the conductivity of TCO and temperature and carrier concentration*

Even the ITO with the best conductivity in TCO has an order of magnitude less conductivity than the typical metal because of its low carrier concentration. In order to form donor level, the carrier concentration of TCO must reach $10^{18}\text{--}10^{19}\text{ cm}^{-3}$, while the typical metal carrier concentration is generally above 10^{22} cm^{-3} (the carrier is in a plasma state, and the interaction between the carrier and light is very strong, and the metallic luster comes from the reflection of the carrier to light).

Transparent conductive oxides such as ITO explain their electrical properties roughly the same as silicon semiconductors. In the following, N-type TCO of In_2O_3 and SnO_2 are used as representatives to observe the relationship between their conductivity and temperature and carrier concentration.

The relationship between conductivity and temperature of In_2O_3 single crystal is similar to that of Si, as shown in Fig. 7.1. At low temperature (L region), the carrier (electron) is supplied by

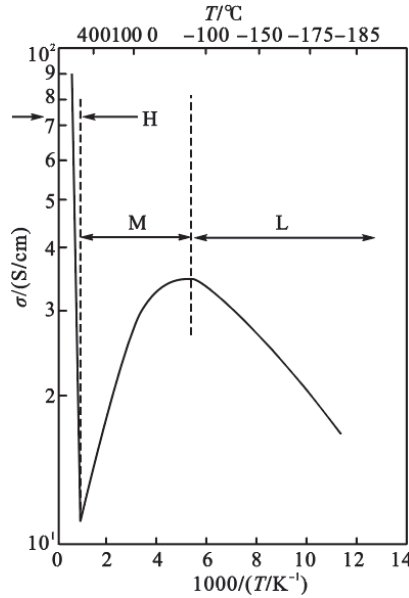


Fig. 7.1. Temperature dependence on electrical conductivity of In_2O_3 single crystal.

the doped donor impurity Sn, and when the temperature rises, the electron excitation from donor level to conduction band increases, so the conductivity also increases. When the temperature rises to a certain extent, all donor level electrons are stimulated to the conduction band. At this time, the number of carriers does not increase, but the thermal vibration of the crystal increases with the temperature, which makes the carrier scattering intensified. Therefore, the conductivity decreases with the temperature rising (M region). Finally, when the temperature rises to a certain extent, the electron in the valence band is also heated to the conduction band, so the conductivity increases again with the temperature (H region).

Defects must be introduced to control conductivity. The conductivity of SnO_2 thin-films with different carrier concentration varies with temperature as shown in Fig. 7.2. With the increase of carrier concentration, the conductivity increases, while the activation energy decreases. When the carrier concentration is high, the conductivity

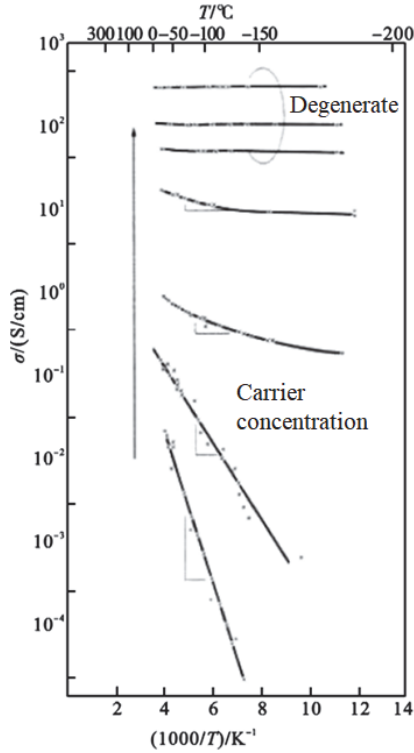


Fig. 7.2. Temperature dependence of conductivity of SnO_2 thin-films.

becomes almost independent of temperature, or exhibits the same temperature tendency as that of metal. When the carrier concentration is very high, the Fermi level enters the conduction band and becomes a so-called degenerate semiconductor. At this time, TCO will show the properties of metal.

7.3.6. *Relation between carrier scattering and resistance in TCO*

The mobility of transparent conductors is $160 \text{ cm}^2/(\text{V s})$, $260 \text{ cm}^2/(\text{V s})$ and $180 \text{ cm}^2/(\text{V s})$ in the samples of In_2O_3 , SnO_2 and ZnO without dopants, respectively. Generally, the crystallinity of thin-film samples is worse than that of single crystal, so these values are

the upper limit of mobility. The actual mobility is also determined by the carrier scattering in the crystal.

Generally speaking, there are five kinds of carrier scattering mechanisms: free dopant scattering, neutral dopant scattering, lattice vibration scattering, differential row scattering and grain boundary scattering. These scattering mechanisms have different temperature dependence. Using Hall effect measurements at low temperatures (from room temperature to liquid nitrogen temperature), we can infer which mechanism actually contributes the most. For ITO with resistivity of about $10^{-4} \Omega \text{cm}$, Hall effect measurements show that the mobility at low temperature is almost independent of temperature, so the scattering mechanism of lattice vibration scattering and differential scattering is not important.

In TCO with good quality (high carrier concentration and large grain size), the average free radius of carriers is less than one order of magnitude smaller than the grain size, so it is not important to judge the grain boundary scattering. In addition, it is known that with the increase of dopant concentration in ZnO, the scattering mechanism will change from grain boundary scattering to free dopant scattering. That is to say, with the addition of dopants, the free Dopant Center in the crystal is the main reason for carrier scattering in highly doped transparent conductors.

Free dopants refer to dopants with different valences from ions originally present in crystals. For example, Sn^{4+} in solid solution replacing the position of In^{3+} in In_2O_3 crystals. The Coulomb force between free dopants and free electrons is the cause of scattering. Using Born approximation and Thomas Fermi-shaped shielding potential, and assuming that all dopants in $\text{In}_2\text{O}_3\text{:Sn}$, ZnO:Al and $\text{SnO}_2\text{:Sb}$ can effectively generate carriers, the curve of resistivity ρ versus n in the presence of only free dopants can be obtained, as shown by the solid line on the left side of Fig. 7.3.

The point above the real line is the resistance measurements of In_2O_3 , ZnO and SnO_2 reported in the literature. The numerical values of each material have the same trend as the theoretical curves, which indicates that the main factor determining the resistivity is the

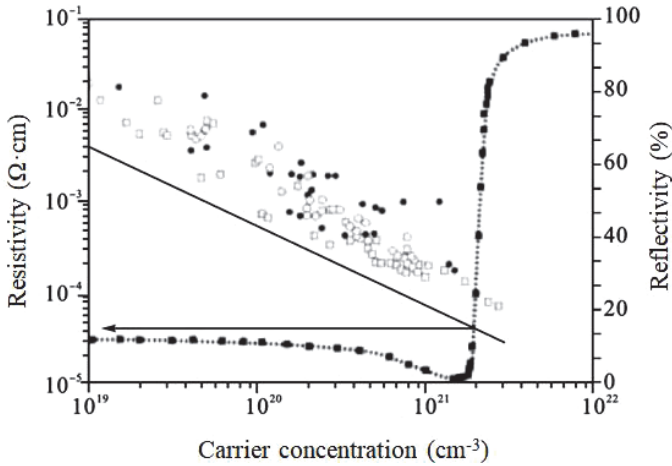


Fig. 7.3. The relation between resistivity, reflectivity and carrier concentration in the presence of free dopants.

scattering of free dopants. The theoretical value shows the lower limit of resistivity. From the tendency of the curve, it can be seen that the limit of resistivity is inversely proportional to carrier concentration, and the upper limit of mobility is $90 \text{ cm}^2/(\text{V} \cdot \text{s})$. Because the upper limit of carrier concentration determined by the requirement of transparency is $2.5 \times 10^{21} \text{ cm}^{-3}$, the theoretical lower limit of TCO resistivity is about $4 \times 10^{-5} \Omega \text{ cm}$, based on the intersection point of resistivity and reflectance curve. In fact, even in ITO with the lowest resistivity, its resistivity is about $1 \times 10^{-4} \Omega \text{ cm}$, which is larger than the theoretical value mentioned above. Therefore, it is considered that the scattering of neutral dopants in ITO (i.e. Sn atoms or SnO_2 complexes located between In_2O_3 lattices) is also a very important scattering mechanism, which can make the measured values deviate from the solid line of Fig. 7.3; without the scattering of neutral dopants, the resistivity should be reduced. To the above theoretical value. However, it is also believed that the concentration and scattering cross-section of neutral dopants are much smaller than that of free dopants, and their scattering effects can be neglected.

7.4. Optical Properties of TCO

7.4.1. *Transparency principle of TCO*

The representative spectra of light penetration, reflection and absorption of TCO are shown in Fig. 7.4. When the incident light energy is larger than the energy gap, the valence band electrons will be excited to the conduction band, so the transmittance range in the short wavelength is determined by the energy gap. In addition, the limit of the transmittance range at the long wavelength side is determined by the plasma frequency. Typical metal or TCO carriers are in a plasma state and interact strongly with light. When the wavelength of the incident light is longer than a certain wavelength, the incident light will be reflected; the wavelength determined by the plasma frequency is in the ultraviolet region for metals and in the infrared region for TCO. So metals are opaque in general, and TCO can just make visible light penetrate and transparent.

7.4.2. *Plasma vibration and plasma frequency*

In the infrared region, besides penetration and reflection, there is absorption, which is related to the resonance of plasma. In general, the plasma vibration which is stable in the plasma is longitudinal wave, and the quantized plasma vibration is called plasmon. Therefore, electromagnetic wave (shear wave) will not interfere with plasma vibration, but if the electric field vector is taken into account, the plasma vibration generated on the surface of the plasma can interfere. The absorption seen in Fig. 7.4 may be the resonance absorption of surface plasmon vibration.

The vibration of plasma can be measured by Electron Energy Loss Spectroscopy (EELS). EELS is an electronic gun that injects electrons with specific energy into the sample. At the same time, the energy of reflecting or penetrating electrons is analyzed by analyzer, and then processed by multiplier tube and counter. Most electrons do not interact with the sample and retain their original energy, but some electrons can cause lattice vibration, carrier plasma vibration

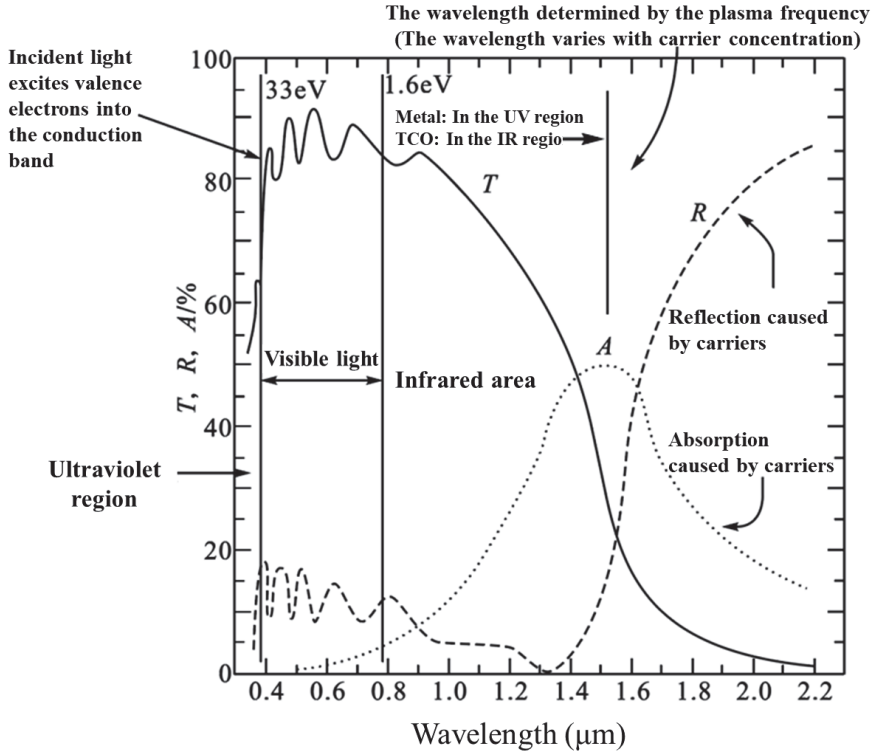


Fig. 7.4. Representative charts of light penetration, reflection and absorption spectra of TCO.

or electron inter-orbital migration, which reduces the energy of these electrons.

The plasma frequency is determined by the carrier vibration and is a function of carrier concentration. Figure 7.5 shows the EELS of ITO. It can be seen that there are plasma absorption peaks near 0.4–0.6 eV, and the plasma frequency increases with the increase of carrier concentration. With the increase of Sn doping, the carrier concentration increases, and the peak of energy loss (absorption) moves towards high energy. In Fig. 7.5, the carrier concentration of the sample with the highest Sn doping content is about $5 \times 10^{20} \text{ cm}^{-3}$. If the concentration is as high as $1.5 \times 10^{21} \text{ cm}^{-3}$, the peak will move towards higher energy.

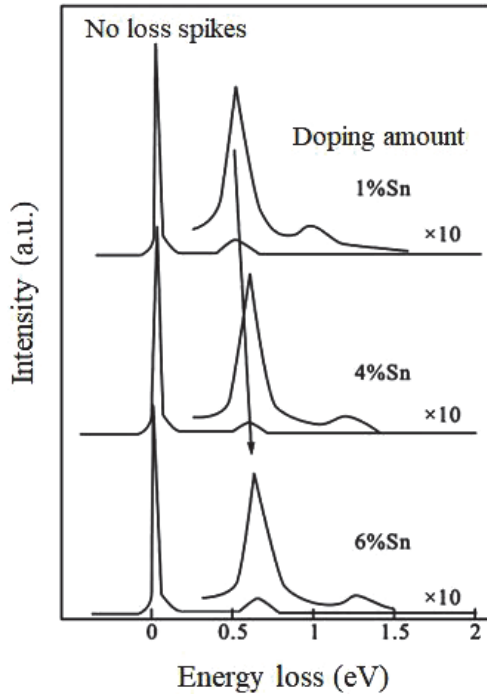


Fig. 7.5. Electron energy loss spectroscopy of ITO.

7.4.3. Burstein–Moss effect

The generated carriers will occupy the bottom of the conduction band, so that the light with the original energy gap cannot transfer the electron excitation of valence band to the bottom of the conduction band; to migrate to the vacancy of the conduction band, higher energy is required. This phenomenon of energy moving towards high energy at the absorption end is called Burstein–Moss (BM) shift, which can be explained simply by Fig. 7.6. When the bottom of the conduction band is occupied, the light with the original gap energy E_{g0} cannot excite the valence band electrons to the bottom of the conduction band. It is necessary to have higher energy E_g to excite the electrons to the conduction band.

As can be seen from the figure, when the carrier concentration is fixed (the area of the oblique part is fixed), the larger the curvature

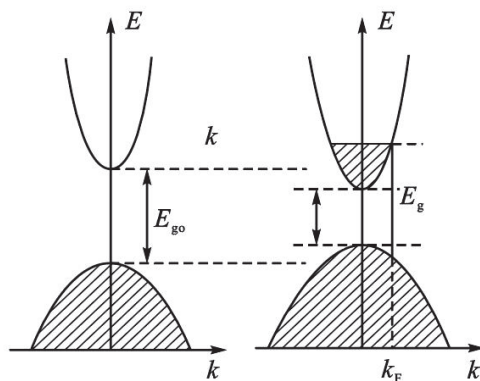


Fig. 7.6. The diagram of BM shift.

of the conduction band, the higher the energy of the filling/vacancy boundary will be, and the more obvious the dependence of this movement on the carrier concentration will be. Therefore, the larger the conduction band curvature, that is, the greater the mobility of the material, the more significant the BM shift.

Moss initially observed the shift of the absorption end in InSb [μ is several $\text{cm}^2/(\text{V s})$] with electron mobility higher than ITO by two orders of magnitude. Strictly speaking, as Burstein's analysis shows, in order to generate carriers, the original energy gap itself will also have some changes. Since only the shift of the absorption end can be observed from the transmittance measurements, the effect of gap change cannot be simply distinguished from the effect of carrier occupying the conduction band, so the two effects are combined as BM shift. However, it is generally believed that the Moss effect should be relatively large in essence.

BM shift will move the absorber to the high-energy side and make the "transparent window" larger. In fact, in ITO, the absorption end of In_2O_3 is only visible light. After adding Sn to generate carriers, the absorption end moves to the ultraviolet region. Another example is the color change of CdO at the absorption end in the visible range. The energy gap is 2.2 eV and the CdO is dark brown. The energy gap is expanded to 3.1 eV after adding $6.5 \times 10^{20} \text{ cm}^{-3}$ carriers, which turns yellow-green.

7.4.4. Carrier concentration and transparency

In order to reduce resistivity and increase carrier concentration, Sn, Al and Sb dopants must be added into transparent conductors such as In_2O_3 , SnO_2 and ZnO , respectively. However, the transparency is affected by the increase of carrier concentration, because free electrons absorb light with a lower vibration frequency than their plasma frequency. The plasma frequency ω is

$$\omega = \left(\frac{ne^2}{\varepsilon_0\varepsilon_\infty m_\infty} \right)^{1/2}, \quad (7.9)$$

where, n is the carrier density, e is the electron charge, ε_0 is the dielectric coefficient in vacuum, ε_∞ is the high-frequency dielectric coefficient, and m_∞ is the conductivity effective mass. With the increase of n , ω becomes larger and the range of light absorption is extended from near infrared to visible.

By substituting the typical dielectric coefficients of TCO, we can calculate the reflectivity of the long wavelength side boundary (800 nm) of visible light and the carrier concentration function, as shown in the reflectivity curve of Fig. 7.3. In order to maintain the transparency of visible light range, n must be suppressed below $2 \times 10^{21} \text{ cm}^{-3}$.

7.5. Transparent Conductive Material Technology

The principle of TCO material is briefly introduced. It has been more than 50 years since TCO was first used. However, people pay more attention to the practical use of oxide electronic materials, so the basic physical properties of these materials are not much. In recent years, due to the progress of science and technology, they have a deeper understanding of them. With these understandings, scholars and experts have begun to try to design new TCO based on basic concepts. It can be expected that new TCO materials will emerge in the near future.

For a long time, the development and market demand of transparent and conductive materials in optoelectronic related industries have been growing continuously. However, the indium mineral deposits

required by mainstream ITO have not been found yet. Research and development of new transparent conductive materials to replace ITO has always been the main research direction of materials researchers, such as metal oxides, nanosilver wires, metal mesh, carbon nanotubes, graphene and conductive polymers. Nowadays, the market of optoelectronic industry and products is changing rapidly. Therefore, this section will analyze the trend of optoelectronic industry, and introduce the progress of research and development of transparent conductive materials, so that relevant researchers can grasp the right direction in the development of conductive materials and their application products.

With the increasing demand for transparent conductive materials in the photoelectric industry, it is difficult to predict the potential size of transparent conductive materials in the future because of the numerous related industries. Until February 2012, Electronics. ca Publications, a well-known market research institution in the electronics industry, presented its views on the future output value of transparent conductive materials (Fig. 7.7). At present, there is a large market demand for transparent conductive materials in three major industries: solar energy, flat panel display and touch panel. The future value of transparent electrodes is expected to increase significantly from \$190 million in 2012 to \$1.8 billion in 2019. Because in addition to the solar energy industry, another energy-saving industry — intelligent temperature control glass — also has strong demand, and both of them are large-scale products. As long as energy-saving issues continue to fever, the use of transparent conductive materials in related products will increase year by year. In contrast, 2012 may be just the beginning of the market scale of transparent conductive materials and energy-saving industries. It is believed that the future photoelectric products may have a more diversified look. At present, the requirements for transparent conductive materials are not clear (large-scale products for energy-saving applications may have the most potential), but it is not too late for large factories in various countries to invest in the development of new transparent conductive materials.

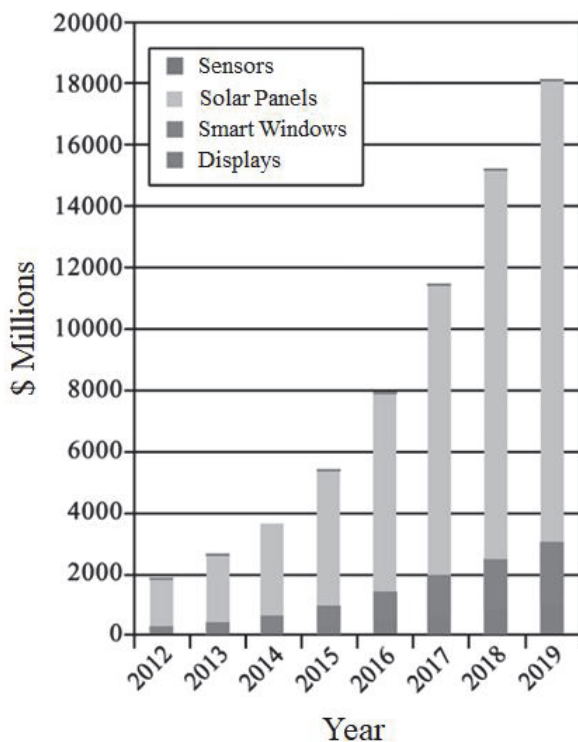


Fig. 7.7. Estimation of transparent conductive materials' output value from 2012 to 2019.

7.5.1. Indium tin oxide

At present, the mainstream of transparent conductive materials market is still ITO (Indium Tin Oxide). ITO is an inorganic metal oxide composed of Indium and Tin. It is also a small number of materials with good transmittance (more than 90%) and low sheet resistance (200–500 Ω/sq). As early as the 1980s, low temperature sputtering method was developed, which belongs to the mature process technology. However, indium ore belongs to rare earth deposit, and its source is mainly in a few countries in China. Even though there are different opinions about the service life of indium deposits, it must be believed that the application of energy saving and display industries may increase dramatically. If high-alternative

materials fail to appear later, the most pessimistic view is that indium deposits will be exhausted by about 2050. In order to avoid monopoly or scarcity of single material, the market always expects materials with various characteristics matching ITO as alternatives. If we want to import other metal oxides as substitutes into the existing ITO electrode process under the minimum change, we can start with the conductivity (σ) of metal oxides and analyze the related factors. As the conductivity is proportional to the electron mobility (μ) and carrier concentration (n), it is found that the electron mobility of metal oxides with high-electron mobility is 100–1000 cm²/(V s), which is comparable to that of gold, silver and copper, but the ITO with the best carrier concentration is only in the order of $10^{20} - 10^{21}$ cm⁻³: only 1/100 or less of the metal. Among them, GZO (low-cost ZnO-doped Ga element) is the most representative. Although the conductivity close to ITO may be obtained by doping, it is still very difficult to make a large area metal chloride transparent electrode with uniform, humidity and temperature resistance. Other ideas for jumping out of existing ITO processes are analyzed below.

7.5.2. *Other compromises between conductivity and transparency*

Electrode materials must provide good conductivity in optoelectronic components. Metal materials are still the best choice, while transparent materials (such as glass and plastics) are often poor conductors of electricity. Therefore, in order to develop other transparent conductive materials in the market, it is necessary to find a compromise between conductive and transparent properties. Major alternatives on the market are material manufacturers, including metal mesh (3M, Fujifilm, PolyIC, etc.), nanosilver wire (CAM-BRIOS, Carestream, BlueNano, ITRI), carbon nanotubes (Eikos, Canatu, SWeNT, Unidym, Mitsui, Chiel, LG Chem, TopNanosys, ITRI), graphene (Samsung), conductive polymers ((AGFA Orgacon, the predecessor of Bayer Baytron), Heraeus, Kodak,) and other metal oxides (Diren of Japan, etc.). The concept of metal mesh is

to combine metal and glass (plastic) to make surface electrodes. If the ratio of metal to glass is 1:99, the theoretical transmittance is 99%, and the resistance value of the sheet should be about $1\ \Omega/\text{sq}$. The UTC (unpatterned transparent conductor) 88xx series products produced by American Merchant 3M are used for isolating EMI. The transmittance is 86–90% and the sheet resistance is $10\text{--}35\ \Omega/\text{sq}$; the patterned transparent metal mesh electrode film (Fig. 7.8) introduced by Fujifilm has a diagonal line of $500\ \mu\text{m}$ and a transmittance of about 80%. The sheet resistance is very low, about $0.5\ \Omega/\text{sq}$, which can isolate EMI efficiency of 300 MHz up to 40 dB; PolyIC provides customized transparent conductive film. (The sheet resistance is $15\text{--}100\ \Omega/\text{sq}$, the transmittance is more than

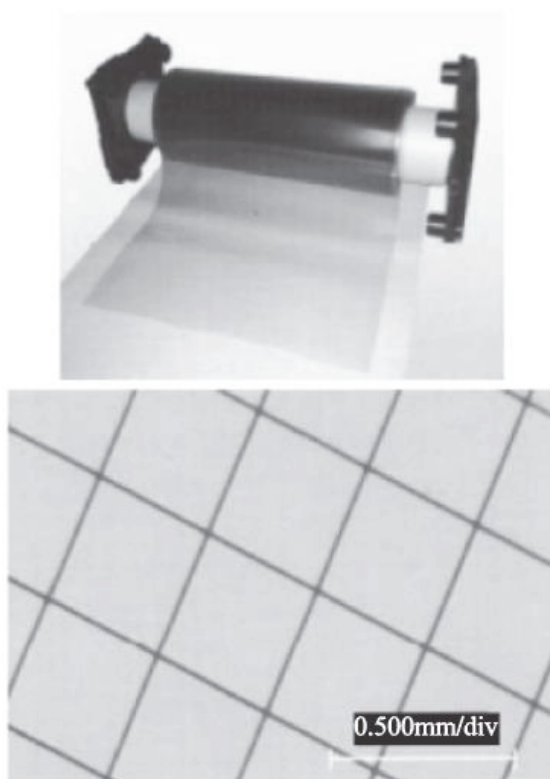


Fig. 7.8. Patternable metal mesh transparent electrode.

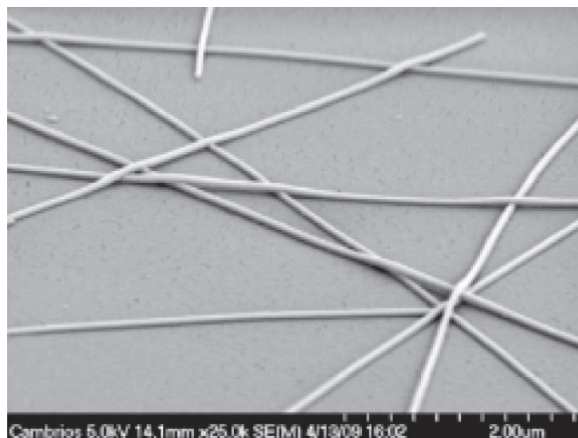


Fig. 7.9. Transparent conductive Film of silver Nanowires.

85%, and the metal lattice can be reduced to $10\ \mu\text{m}$). However, the disadvantage of metal mesh is that the close-range mesh structure still belongs to the nanometer level visible to the naked eye, which is not suitable for small-size display products. CAMBRIOS, a leading manufacturer of nanosilver wires, synthesizes nanosilver wires with aspect ratio greater than 300 nm and radius less than 100 nm by chemical method. The nanosilver wires are coated on transparent substrates and randomly form grids about $1\ \mu\text{m}$ in size (Fig. 7.9).

Its ClearOhm series products have a high transmittance of nearly 99% in the case of monolayer, with a resistance of $50\text{--}300\ \Omega/\square$; and the transparent electrode film produced by Carestream contains a transparent substrate with a transmittance of 87–89% and a resistance of $100\text{--}300\ \Omega/\square$. Photoelectric properties are comparable to ITO, but the transmittance of silver nanowires in long wavelength range has advantages, while haze (%) exists at a disadvantage (Fig. 7.10). Graphene (Fig. 7.11) materials are locked into good transparent electrode substitutes besides their excellent photoelectric properties (single-layer graphene transmittance can reach 97.7%, sheet resistance value is about $100\ \Omega/\square$), and their tough mechanical properties make them widely used. Therefore, many researchers have invested in the research and development.

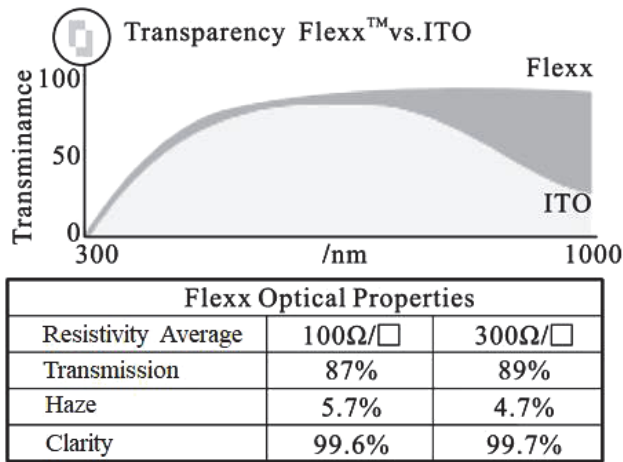


Fig. 7.10. Light transmittance and haze of Silver nanowires electrodes.

At present, the highest quality single-layer graphene material is grown on copper foil by CVD and then transferred to other transparent substrates. The cost of production is slightly higher. Carbon nanotubes (CNT, Fig. 7.12) are the most possible transparent electrode materials, including Inkjet, Shadow Mask, screen printing, gravure printing and so on. The graphene as another widely used carbon materials, the transmittance of widely used carbon materials. The transmittance of single layer is 80–89%, the resistance of sheet is 200–700 Ω/□, but the color is gray and black. In general, PEDOT:PSS with the best conductivity is chosen as transparent electrode material, because organic polymer has many differences from other transparent conductive materials, PEDOT:PSS has the advantages of coating and flexibility in essence. However, the conjugated molecule structure makes it absorb specific wavelength and show obvious blue color. The stability of the conjugated molecule is also worse than that of other inorganic materials and carbon materials. The resistance value of the conjugated molecule is slightly higher than that of other inorganic materials, which is more than 300 Ω/□. But last year Kodak launched a new product with almost

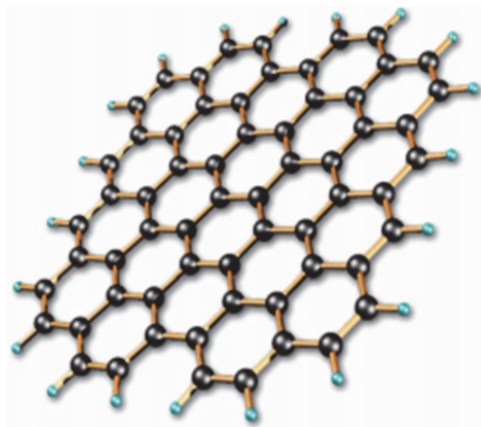


Fig. 7.11. The structure diagram of Graphene.

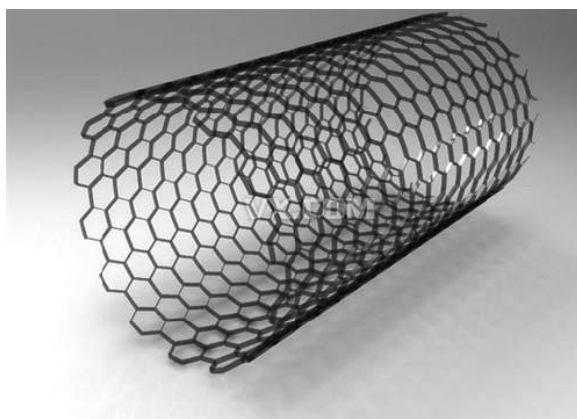


Fig. 7.12. The structure diagram of single-layer CNTs.

no color bias ($a^*:-0.24$, $b^*:0.48$). The transmittance of visible light is 88.3%, and the film resistance is about $230 \Omega/\square$. Because the requirement specification of capacitive touch panel is about 90% of the single-layer transmittance and the resistance value of the panel is less than $300 \Omega/\square$, the progress of this product keeps the competitiveness of conductive polymer and other materials.

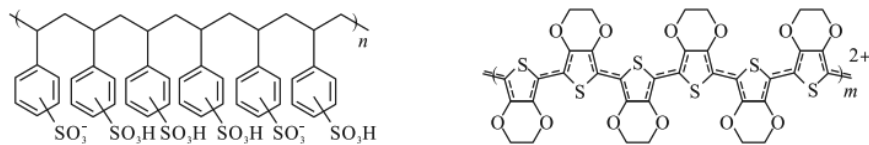


Fig. 7.13. PEDOT: Structural Diagram of PSS.

7.5.3. Soft ITO films

Despite the rapid development of new transparent conductive materials, the progress of ITO materials has not been alarmed (the process of ITO on glass is not discussed here). Looking to the future, the demand for flexibility of photoelectric 3C products is bound to increase. Diren locked in the application industry of flexible touch control and display panel, and developed ITO film based on soft substrate materials.

Soft transparent conductive films are manufactured by Roll-to-Roll (roll-to-roll) method. It is necessary to master three technologies: film, wet coating and vacuum evaporation. For example, the ITO film is sandwiched between the display panel and cover lenses, and the polarizer can produce high reflection because of the retardation. Therefore, the circular polarization system (Fig. 7.14) which combines the phase difference plate and the polarization plate can reduce the reflectivity from more than 70% to less than 20%. In addition, the process temperature that the general soft base plate can withstand is not high, and the material of PET often begins to deform below 100°C, resulting in many restrictions on the product and its process. In order to improve the temperature resistance, Diren displayed ITO/PC transparent conductive film which can withstand up to 150°C. Among them, flameproof polycarbonate (Table 7.3) with special structure design is used as the base material. The glass conversion temperature can reach 264°C and the heat shrinkage rate of 250°C is less than 1%. Although these are the data of experimental stage, the possibility of partitioning the market of glass-making products with plastic substrates has been given.

In the design of Hard Coating, the Diren used wet coating technology to suppress the interference between two ITO of touch

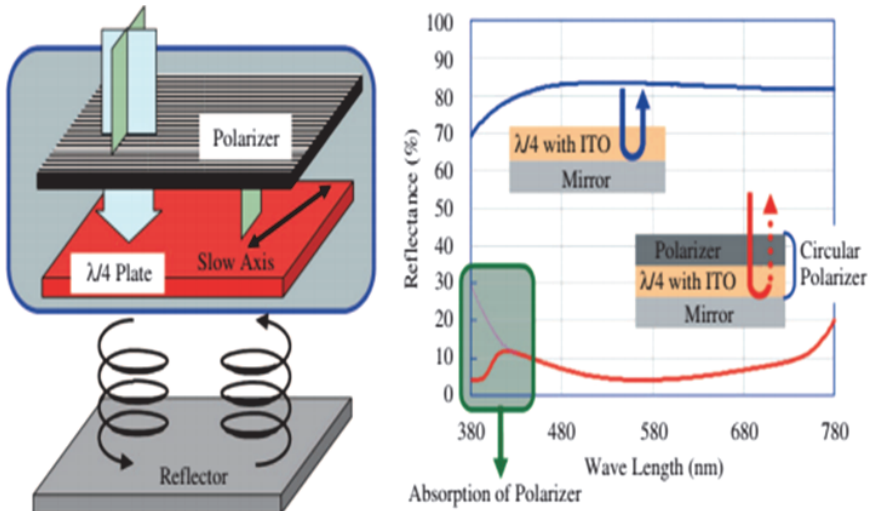


Fig. 7.14. Circular polarizer thin-film technology for reducing reflection.

Source: Diren Cheng was published in Fine Tech Japan in 2012.

elements. Hard Coating usually adds a few particles to the ITO film on the surface to reduce the interference due to the scattering of light in the two layers. However, these particles are often destroyed in component durability testing, so the Diren's approach is to form a slightly scatterable surface on the bottom of ITO film by wet coating. Compared with Hard Coating by adding particles, it can maintain stability more than 30 million times. The key point of vacuum evaporation technology is to improve the optical properties and stability of ITO. Comparing crystallized ITO and amorphous ITO deposited at 25 μm PET with the same thickness, Diren found that crystallized ITO had better optical properties. Its transmittance could be increased from 89.1–90.3%, b^* could be reduced from 2.8 to 1.9, and the resistance of the sheet increased slightly from 262 $\Omega/\square$ to 290 $\Omega/\square$. In high temperature and humidity environment, the ratio of sheet resistance change after 500 hours test shows that the change rate of sheet resistance of crystalline ITO is 40–50% compared with that of non-crystalline ITO, only about 10% of the sheet resistance change, and the stability is remarkably excellent. There will be

Table 7.3. Transparent conductive film ITO/flameproof PC with temperature resistance up to 264℃.

Project	Condition		Unit	Flame retardant polycarbonate (experimental value)	Conventional polycarbonate D-92
Optical properties	Film thickness		μm	100	92
	density		kg/m ³	1.93	1.2
	Tortuosity (nD)		—	1.622	1.585
	Density (haxe)		%	1	0.2
	Total light transmittance		%	89	90
	L^* C/2		—	95.2	95.1
	a^*		—	−0.1	0
	b^*		—	0.8	0.5
	Re		nm	<10	<10
	Rth		nm	—	10
Physical/thermal properties	Photoelastic coefficient		10 ^{−12} m ² /N	47	80
	Glass transfer temperature	20℃C/min(DSC)	□	264	148
	Linear expansion coefficient	50~100℃(TMA)	10 ^{−6} /□	56	80
	Heat shrinkage	150 □, 10 min	%	<0.01	—
		200 □,10 min	%	<0.01	—
		250 □,10 min	%	<0.01	—
	Flammability	UL94	—	V-0	Not-V

Table 7.3. (Continued)

Project	Condition		Unit	Flame retardant polycarbonate (experimental value)	Conventional polycarbonate D-92
Chemical property	Water absorption	24 h in Water	%	0.2	0.3
	Water vapor permeability		g/m ² day	52	50
	Oxygen permeability		ml/m ² day	1.11	900
	Drug resistance	Acid resistance	—	○	○
		Tolerance test	—	×	×
		Resistance to organic solvents	—	×	×
Mechanical properties	Elastic modulus		GPa	2.6	2.1 (TD)
	Tensile strength		MPa	91	56 (TD)
	Tensile percent		%	5	5 (TD)

no peeling in 10,000 write cycles. No residual acid will cause discoloration and deterioration of optical clear adhesive (OCA) after bonding. At this time of increasing demand for capacitive touch panel, the development of larger conductive film must be aimed at reducing the resistance of the chip, because low resistance components have the following advantages: larger operating range; flexible design; faster response time; and meticulous lines. At present, the use of ITO materials to increase thickness to meet the demand for chip resistance values of various sizes. However, when the transmittance decreases with the thickness and the cost increases, the Diren uses Anneal technology to reduce the resistance from $218 \Omega/\square$ to $147 \Omega/\square$, and the transmittance increases from 87–89%. This further makes the properties of ITO conductive films an indicator. In addition, the ITO material is not bending-resistant, and its flexibility radius can reach 2–2.5 mm after improvement.

7.5.4. *Silver nanowires*

At present, there are two main methods to prepare nanosilver wires: template method and solution chemical method. Although the products with precise line length and diameter can be obtained by template method, they must undergo complex processes such as filling reduction and electrolytic deposition of silver, and this technology cannot be manufactured in large quantities. At present, there are many methods of synthesizing nanosilver wires by solution chemistry, most of which are prepared by ethylene glycol. Polyvinyl pyrrolidone (PVP, polyvinyl pyrrolidone) was used as a stabilizer to reduce silver nitrate in the presence of high temperature and seeds (above 160°C). Seed is a decahedral structure. By using PVP as a protective agent, the two ends can extend outward into a linear shape. The cross-section of this line is pentagonal, because the two end lattices are (111) pentahedral and each side is (100) lattice. PVP molecule adsorbed on (100) surface by force, so that silver can only grow on (111) surface in the process of reduction, so it can grow into silver line (Fig. 7.15). The amount of PVP is also different from the protection effect, which indirectly affects the

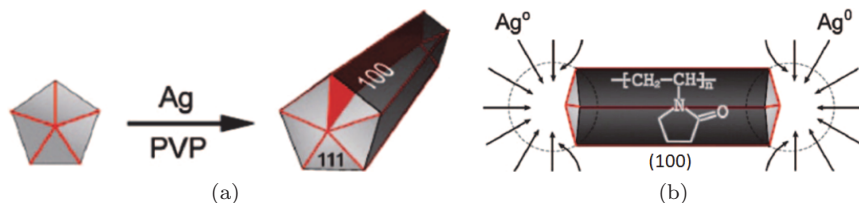


Fig. 7.15. Schematic diagram of growth mechanism of silver nanowires.

diameter of nanowires. Therefore, the synthesis method is affected by many factors, including seed selection, stabilizer selection, solvent/reductant selection, buffer ion addition, silver-ion concentration control, reaction time/temperature and so on. CAMBRIOS, the leading manufacturer of this material, has been investing in the research and development of nanosilver conductors for quite a long time. The company believes that the nanosilver wires are in an advantageous position in the comparison of the existing transparent conductive (alternative) materials (Table 7.4). The analysis is as follows.

Excellent properties: In the future, transparent conductive films will not only require less than $300 \Omega/\square$, but also have high conductivity of silver wires. With the increase of coating concentration, the resistance can be greatly reduced, and the visible light transmittance can be maintained at more than 80%. Only the photoelectric properties of evaporated graphene are comparable to that of graphene.

High-process performance: Thin-films can be prepared by low temperature/wet coating, and patterned electrodes can be made by etching, transfer and screen printing. Only conductive polymers and carbon nanotubes are the same, but their photoelectric properties have a big bottleneck at present.

Cost advantage: In terms of production cost, Roll-to-Roll can be used as the production mode, and the coverage area after film forming is only less than 1%. Although silver is a precious metal material, it is widely distributed, frequently used and easily recovered.

It is generally believed that the two most difficult problems in this material technology are the atomicity and adhesion after

Table 7.4. Comparisons among transparent conductive (alternative) materials.

	ITO	Metal lattice	Silver nanowires	Conductive polymer	Carbon nanotubes	Graphite field
Single-layer transmittance/%	89–90	86–90	90–99	87–89	80–89	97
Chip resistance ($\Omega/\square$)	150–300	15–100	10–300	200–500	200–700	100
Process	Sputter plating	Printing	R2R	R2R	R2R	CVD/transfer
Flexibility	Bad	Good	Good	Good	Good	Good
Color appearance	Yellowish	Dew, near fragrance with micron grid	Dew degree	Slant blue	Partial dark gray	Good

Source: Diren Cheng was published in Fane Tech Japan in 2012.

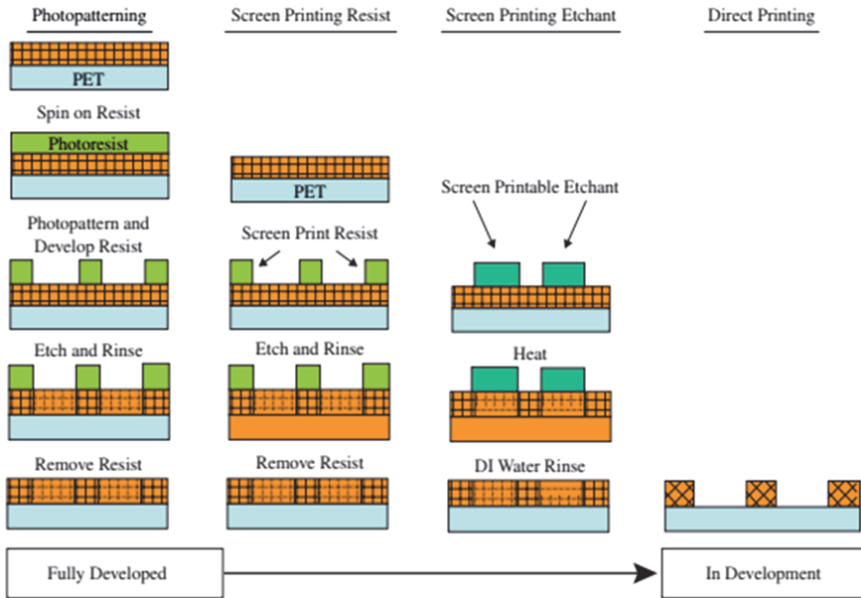


Fig. 7.16. Evolution of patterning of silver nanowires electrode.

Source: CAMBRIOS.

film formation. Silver nanowires developed by CAMBRIOS can effectively reduce fog by controlling the ratio of diameter to diameter. As for the surface of silver wire and base material, because the conductive continuous phase is formed by dispersive deposition, it is necessary to add some adhesives in the solution of silver wire coating to form an overcoat on the silver wire network. The effect on the overall conductivity of the adhesives is not significant because only the appropriate proportion must be controlled. CAMBRIOS has proposed the latest patterning method (Fig. 7.16) in order to further improve the possibility of silver nanowires as transparent conductive layer materials. The previous concept was to fabricate etchable patterns on the top of silver nanowires after they were fabricated into thin-films, and then to leave the desired patterns in the etching and deinking process. If the lithography step is simplified, photoresist etching and ink removal are applied to screen printing. At present, CAMBRIOS shows that etchant can be printed directly

over the silver film by screen printing. When etchant penetrates the protective layer, it can directly cause silver line breakage and lose conductivity. The circuit can be fabricated without subsequent deinking. In addition, CAMBRIOS also works with Hitachi Chemical to insert a layer of Photo Sensitive Layer between Cover Film and Silver Film to demonstrate a 20 μm high-resolution line printing capability. However, direct printing of conductive patterns is still the ultimate goal. There are opportunities for ink jet, screen printing or gravure printing, but the most difficult part of the direct coating technology is to adjust the proper viscosity and adhesion, which is still in the development stage.

The objectives of developing new transparent conductive materials are as follows: stable source and price; easy production (non-vacuum film forming); bending resistance; excellent optical and electrical properties; patternability; environmental change resistance; large area fabrication; high uniformity of properties (the first three are the worries of current use of ITO). According to the recent development of various transparent conductive materials, the trend of progress is still optimistic, and each has its own advantages and disadvantages. ITO materials generally give people the impression that bending resistance is also gradually improving. Since the industry has invested heavily in ITO process related equipment, it is unlikely that ITO materials will be replaced before they are used up. Even if some new material technology can eliminate the limitation of vacuum film forming in the process, if there is no significant difference in photoelectric characteristics, it will only be set for the record first. However, if a breakthrough growth in material technology or industry demand (such as green energy industry) occurs in the future, it is possible to accelerate the process of introducing new material technology into existing products. Therefore, the industry or R&D units should pay attention to the development of all transparent conductive materials technology.

References

- Chen P, 2009. Development of new transparent conductive materials for touch panel. *OPTOLINK* (80):21–24.

- Committee 166 on Transparent Conductive Materials, The Japan Society for the Promotion of Science, 1999. Transparent Conductive Film Technology.
- Coutts T J, Young D L, Li X, 2000. Characterization of transparent conducting oxides. *MRS Bull.* 25(8):58–65.
- Freeman A J *et al.*, 2000. Chemical and thin-film strategies for new transparent conducting oxides. *MRS Bull.* 25(8):45–51.
- Gordon R G, 2000. Criteria for choosing transparent conductors. *MRS Bull.*, 25(8):52–57.
- Huang C-J, Huang S-J, 2012. Development and application trends of graphene materials. *J. Indust. Mater.*, (304):62–74.
- Kawazoe H *et al.*, 2000. Transparent p-type conducting oxides: Design and fabrication of *pn*-heterojunctions. *MRS Bull.* 25(8):28–36.
- Lewis B G, Paine D C, 2000. Applications and processing of transparent conducting oxides. *MRS Bull.* 25(8):22–27.
- Zhang D H, Ma H L, 1996. Scattering mechanisms of charge carriers in transparent conducting oxide films. *Appl. Phys. A* 62(5):487–492.

This page intentionally left blank

Chapter 8

Touch Panel

8.1. Development of Touch Technology

8.1.1. *The generation of touch control technology*

If the invention of mouse in 1964 brought computer operation into a new era, then the emergence of touch panel makes the graphical human–computer interaction interface more intuitive and easy to use. In 1971, Sam Hurst, an American, invented the world’s first touch sensor. Although this instrument is different from the touch panel people see today, it is regarded as the beginning of the development of touch panel technology. Figure 8.1 is the original touch panel and touch panel mobile phone.

In 1971, Sam Hurst, a teacher at the University of Kennedy, was bothered by the huge amount of graphics data he had to process every day. He began to figure out how to improve his work efficiency and how to work out these graphics in the simplest way.

He transformed his three basements into workshops, one for processing wood, one for manufacturing electronic components, one for assembling these parts, and ultimately produced the earliest touch panels. The earliest touch panel was named “AccuTouch” because it was manually assembled and produced only a few devices a day. In 1973, the technology was selected by *American Journal of Industrial Research* as one of the 100 most important new technology products of the year. Soon after, Sam Hurst set up his own company and worked with Siemens to improve the technology. During this period, touch panel technology was mainly adopted by the US military. Until 1982, Sam Hurst’s company exhibited 33 TV sets with touch panels

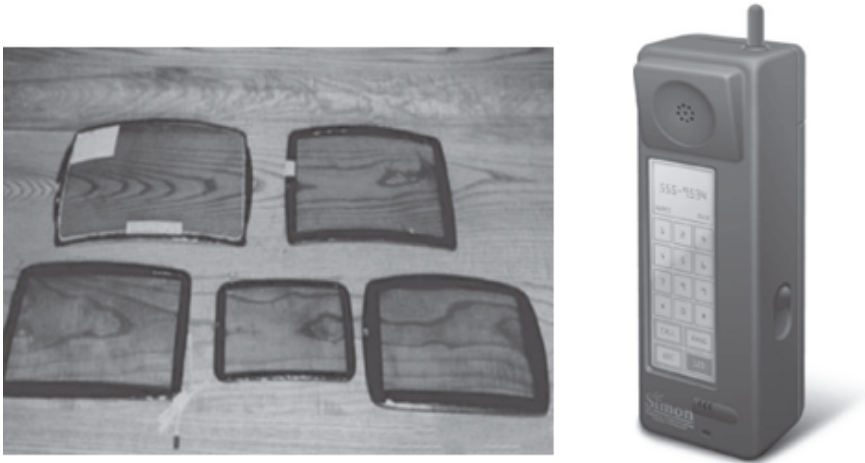


Fig. 8.1. Initial touch panel and touch panel mobile phone.

at a science and technology exhibition in the United States. It was the first time that ordinary people “touched” the magic touch panel by hand. The earliest touch phone was Simo Personal Communicator launched by IBM in 1994, which used pressure sensing black-and-white screen instead of physical buttons. The phone is equipped with a touch pen to record text and even send faxes.

8.1.2. *Definition of touch panel*

Touch panel is a device for continuous coordinate positioning of touch input. In principle, the standard touch panel needs to meet three requirements.

8.1.2.1. *Transparency*

Transparency directly affects the visual effect of touch panel. Touch panel as the input device on the screen, the first requirement is to try not to affect the display effect, which puts forward high requirements for the optical performance of touch panel. Transparency has the problem of transparency. Infrared technology touch panel and surface acoustic touch panel are separated by only one layer of pure glass. Transparency is the best. Other touch panels need to make a

thorough study on this point. “Transparency” is a very broad concept in the touch panel industry. Many touch panels are multi-layer composite films. It is not enough to generalize their visual effects with transparency only. It should include at least four characteristics: transparency, color distortion, reflectivity and clarity. It can be further divided, for example, the degree of reflection includes the degree of mirror reflection and diffraction reflection, but the degree of diffraction reflection on the surface of the touch panel has not reached the level of CD disk, for users, these four measures are enough.

Because of the existence of light transmittance and wavelength curve, the image seen through the touch panel inevitably produces color distortion with the original image. The static image feeling is only color distortion, and the dynamic multimedia image feeling is not very comfortable.

Color distortion refers to the maximum color distortion in the picture, the smaller the better. The usual transparency can only be the average transparency in the graph, of course, the higher the better.

Reflectance mainly refers to the overlapping light and shadow on the image caused by specular reflection, such as figures, windows, lights and so on. Reflection is a negative effect brought by the touch screen, the smaller the better. It affects the user’s browsing speed, even in severe cases, it is impossible to recognize the image characters. The use environment of the reflective touch screen will be limited, and the lighting layout of the scene will also be forced to adjust.

Clarity refers to the degree of clarity of each detail and its boundary on an image. In the touch panel industry, we measure it by the haze. The haze refers to the percentage of transmitted light intensity above 2.5 degree deviation from the incident light. The higher the haze, the lower the clarity.

8.1.2.2. *Touch panel is absolute coordinate system*

The essential difference between absolute coordinate system and relative coordinate system, such as mouse, is the intuitive in place at one time. The characteristic of absolute coordinate system is

that each location coordinate has nothing to do with the last location coordinate. Touch panel is a set of independent coordinate positioning system physically. The data of each touch is converted to the coordinates of the screen through the calibration data. In this way, it is required that the output data of the same point of the touch panel is stable under any circumstances. If it is not stable, then the touch panel cannot guarantee absolute coordinate positioning. Point inaccuracy, this is the most feared problem of touch panel, that is, drift. In technical principle, if the touch panel cannot guarantee that the same touch point sampled the same data each time, there will inevitably be a drift problem. At present, mutual capacitive touch panel does not exist drift phenomenon.

8.1.2.3. *Detecting touch and locating*

All kinds of touch panel technology rely on their own sensors to work, and even some touch panel itself is a set of sensors. The response speed, reliability, stability and life of touch panel are determined by their respective positioning principles and sensors.

8.1.3. *Classification of touch panel*

From the technical principle to distinguish the touch panel, it can be divided into five categories: vector pressure sensing technology touch panel, resistance technology touch panel, capacitance technology touch panel, infrared technology touch panel and surface acoustic technology touch panel. Infrared touch panel technology is cheap, but its frame is fragile, easy to produce light interference, distortion in the case of curved surface. Capacitance technology touch panel design is reasonable, but its image distortion problem is difficult to be fundamentally solved. Resistance touch panel has accurate positioning, but its price is quite high, and it is afraid of scratching and fragility. The surface acoustic wave touch panel solves various defects of the previous touch panel. It is clear and not easy to damage. It is suitable for various occasions. The disadvantage is that if there are water droplets and dust on the screen surface, the touch panel will become dull or even not work.

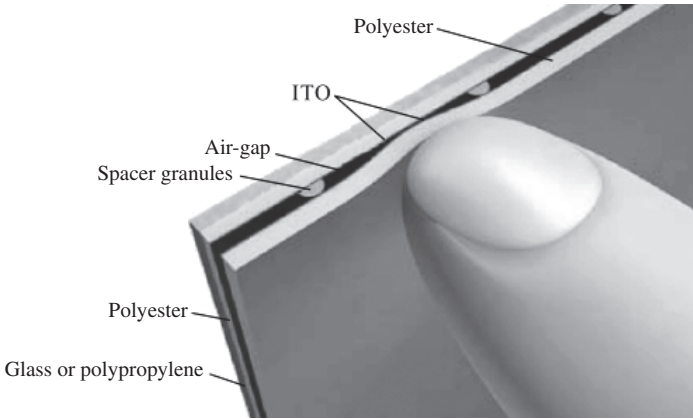


Fig. 8.2. Structure of resistance touch panel.

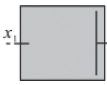
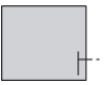
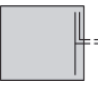
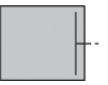
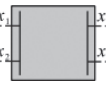
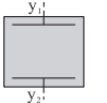
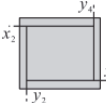
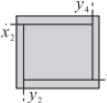
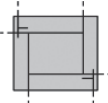
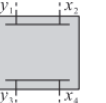
8.2. Resistance Touch Panel

Resistance touch panel is a kind of touch sensor which can convert the touch position into X- and Y-axis voltage and calculate the specific touch position. Its general structure is shown in Fig. 8.2. The lower layer is a rigid transparent base material, such as glass. The surface is coated with ITO, and the upper layer is coated with a flexible film hardened by surface treatment. The inner layer is coated with ITO. The two layers of ITO are separated by transparent particles. Touch control function can be realized by pressing the upper layer film to make the upper and lower ITOs turn on. Position coordinates can be obtained by the algorithm. Resistance touch panel has the advantage of low cost and is not affected by dust, water and dirt. Its disadvantage is that it is easy to be fragile and fragile. After a certain time of using surface ITO, there will be small cracks, even deformation, astringent sliding operation and poor user experience.

According to the different electrode structures, resistance touch panels can be divided into five categories (Table 8.1). The principle of four-wire resistor screen (Fig. 8.3) is analyzed as an example.

The two layers of ITO coatings of resistance touch panel are separated by transparent particles, and the two layers of ITO do not touch each other in the non-pressing state. The X + connected VDD

Table 8.1. Electrode structure of resistance touch panel.

	Four-wire type	Five-line type	Six-line type	Seven-line type	Eight-line type
Upper electrode					
Lower electrode					
Patent right	Wiring is the simplest, not subject to patent restrictions	ELO, 3M Patents	Universal Photo-electric Patent	Japan's Rich Patents	3M Patent
Charac- teristic	The product is not scratch-proof and the price is the lowest	Modified four-wire type, not scratch-resistant	Increase scratch resistance, anti-electro-magnetic wave and anti-noise function compared with four lines	Scraping resistance and high accuracy	Temperature resistance and ambient temperature variation, resolution twice as high as four lines

(reference voltage) of the lower coating, X – grounding. Similarly, Y + connected to VDD, Y – grounded. When the fingers press the resistance screen, the flexible film deforms. When the upper ITO coating contacts the lower ITO coating, it can be considered as the conduction of the upper circuit to the lower layer. The circuit is shown in Fig. 8.3. The R_{touch} is the contact resistance. The coordinates of X and Y can be calculated by the following methods. When Y+ pole plus VDD voltage, Y– pole grounding, X+ pole and

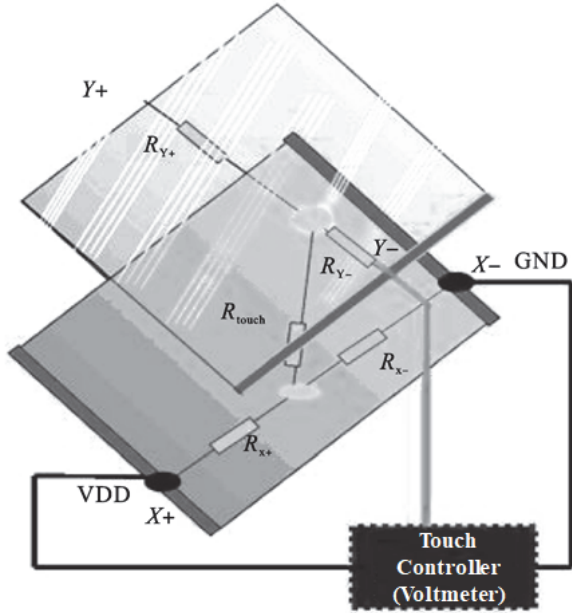


Fig. 8.3. Principle of four-wire resistance touch panel.

X- disconnected voltage, the contact point voltage is

$$V_x = V_{DD} [R_{X-} / (R_{X+} + R_{X-})]. \quad (8.1)$$

Because the magnitude of R_{X+} and R_{X-} value is proportional to the distance from the contact point to the Y+ pole and Y- pole, $R_{X-} = L_{X-} \cdot \rho$ and $R_{X+} = L_{X+} \cdot \rho$ (ρ is unit length resistance), we can know that $V_x = V_{DD} [L_{X-} / (L_{X+} + L_{X-})]$, that is, $V_x / V_{DD} = [L_{X-} / (L_{X+} + L_{X-})]$. Measuring the contact point voltage V_x can determine the L_{X-} , that is, the X- axis coordinates of the contact point. Similarly, the Y- axis coordinates of the contact point can be obtained. Resistance screens of other structures have similar principles.

8.3. Capacitance Touch Panel

The basic principle of capacitive touch panel is shown in Fig. 8.4. The capacitance is formed between the capacitive screen and the contacts,

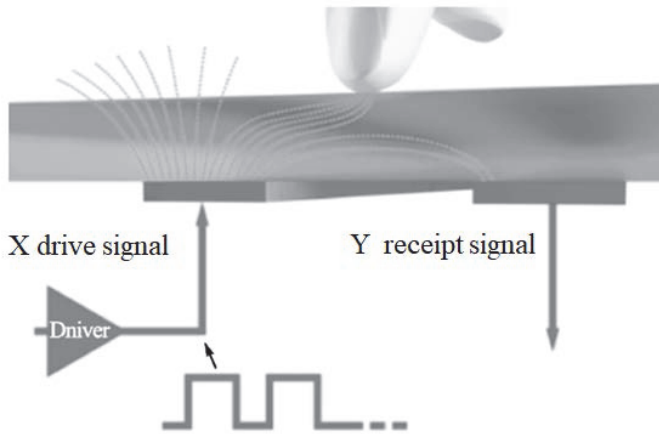


Fig. 8.4. Basic principle of capacitance touch panel.

and the contacts are required to be conductors, such as fingers. When the finger touches the capacitive screen, a small part of the current is transferred to the human body through the capacitance between the finger and the capacitive screen. The touch panel detection circuit detects the change of the current, so as to determine the coordinates of the contact points, that is, to realize the touch control function.

There are two kinds of capacitive touch panel, surface capacitive touch panel and projected capacitive touch panel.

8.3.1. *Surface capacitance touch panel*

Surface capacitive touch panel only uses single layer of uniform ITO on the whole surface, and is coated with long and narrow electrodes on all sides of the touch panel, forming a low voltage AC electric field in the conducting body. When the finger touches the surface of the screen, a coupling capacitance will be formed between the finger and the conductor layer, and a certain amount of charge will be transferred to the human body. In order to recover these charge losses, the charge is added from the four corners of the screen. The amount of charge added in each direction is proportional to the

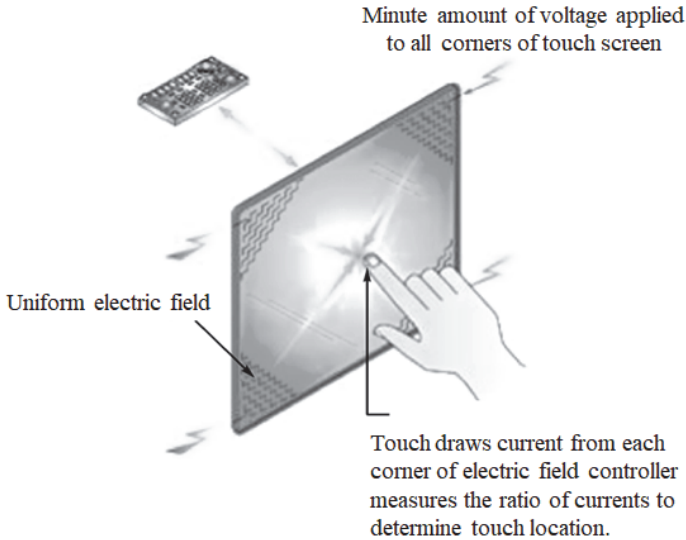


Fig. 8.5. Structural principle of surface capacitance touch panel.

distance of the touch point, from which the position of the touch point can be calculated. This kind of surface capacitive touch panel has simple principle and production process, but it can only support single touch. The structure principle of the surface capacitive touch panel is shown in Fig. 8.5.

Uniform deposition of ITO results in pillow distortion, which is usually corrected by low impedance edge patterns, as shown in Fig. 8.6. Surface capacitive ITO coatings usually require linearized metal electrodes around the screen to reduce the effect of corner/edge effects on the electric field. When a large area of the palm or handheld conductor is close to the capacitive screen instead of touching, it can cause the misoperation of the capacitive screen, especially in wet weather. The misoperation of the capacitive screen can be caused by holding the display, holding the palm close to the display within 7 cm or the body close to the display within 15 cm. Gloves or handheld non-conductive objects do not respond when touched because of the addition of more insulating media. When the environmental

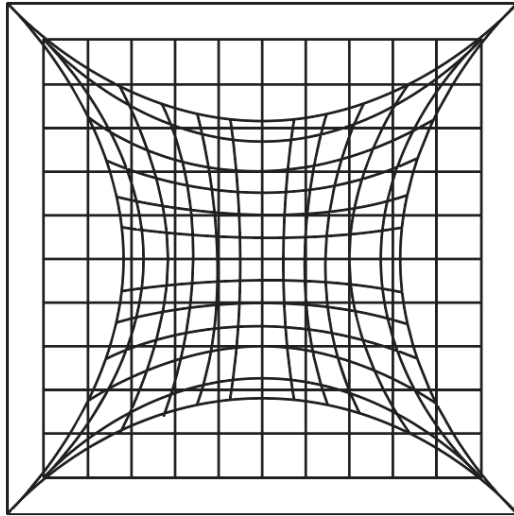


Fig. 8.6. Pillow distortion schematic.

temperature, humidity and electric field change, it will cause the drift of the capacitance screen and cause inaccuracy.

8.3.2. *Projection capacitive touch panel*

The projective capacitive touch panel can be divided into two types: self-capacitive touch panel and mutual capacitive touch panel. Such capacitive screens require one or more carefully designed, etched ITO layers. The ITO layer is etched to form multiple horizontal and vertical electrodes, all of which are driven by a capacitive induction chip. The chip cannot only transmit data to a main processor, but also handle the XY axis position of the contacts by itself. Usually, both horizontal and vertical electrodes are driven by single-ended induction, that is to say, there is no difference between a row of driving circuits and a row of driving circuits, called “single-ended” induction (self-capacitance), as shown in Fig. 8.7. However, in some methods, an axis is driven by a set of AC signals, and the response through the touch panel is sensed by the electrodes on other axes, called “transverse” induction. In this case, the electric field is transmitted from one group of electrodes (e.g. rows) to another group

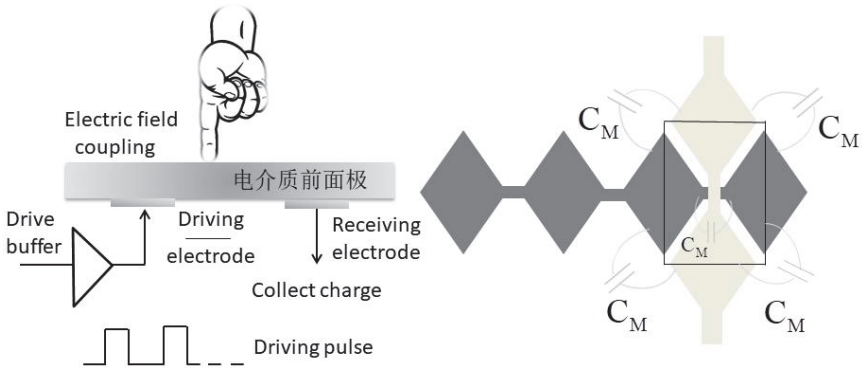


Fig. 8.7. Structural principle of self-capacitive touch panel.

of electrodes (e.g. columns) through the dielectric layer of the upper layer plate in a transverse manner.

8.3.2.1. Self-capacitive touch panel

The general structure of self-capacitive touch panel is as follows: ITO (a transparent conductive material) is used to fabricate transverse and longitudinal electrode arrays on the glass surface. These transverse and longitudinal electrodes form capacitors with the ground respectively. This capacitance is commonly called self-capacitance, that is, the capacitance from the electrode to the ground. When the finger touches the capacitive screen, the capacitance of the finger will be superimposed on the capacitance of the screen, which will increase the capacitance of the screen. In touch detection, the transverse and longitudinal electrode arrays are detected by the self-capacitive screen in turn. According to the changes of capacitance before and after touch, the transverse coordinates and the longitudinal coordinates are determined respectively, and then combined into the plane touch coordinates. The scanning mode of self-capacitance is equivalent to projecting the touch points on the touch panel to the X -axis and Y -axis directions respectively, then calculating the coordinates in the X -axis and Y -axis directions respectively, and finally combining them into the coordinates of the touch points.

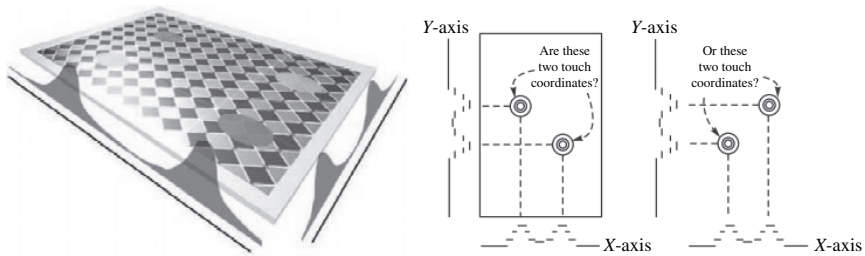


Fig. 8.8. Multi-touch and ghost-point phenomena of self-capacitive touch panel.

If it is a single touch, the projection in the X -axis and Y -axis directions is unique, and the coordinates combined are also unique; if there are two touches on the touch panel and the two points are not in the same X -direction or the same Y -direction, then there are two projections in the X -direction and Y -direction respectively, then four coordinates are combined. Obviously, only two coordinates are real, and the other two are commonly known as ghost points, as shown in Fig. 8.8. Therefore, the self-capacitive screen cannot achieve real multi-touch.

The characteristics of the self-capacitive touch panel are as follows: (1) It needs to be calibrated for the first time in use or when the environment changes greatly. (2) It has ghost point effect and cannot realize real multi-touch. (3) It is directly affected by temperature, humidity, finger wetness, body weight, and dryness of the ground. It is also very disturbed by large areas of external objects, and easy to produce “drift”.

8.3.2.2. *Mutual capacitance touch panel*

The general structure of mutual capacitive touch panel is as follows: ITO is used to make transverse and longitudinal electrodes. The difference between it and self-capacitive screen is that capacitance will be formed where two groups of electrodes intersect, that is, the two groups of electrodes constitute the two poles of capacitance respectively. When the finger touches the capacitive screen, it affects the coupling between the two electrodes near the touch point, thus changing the capacitance between the two electrodes. When the

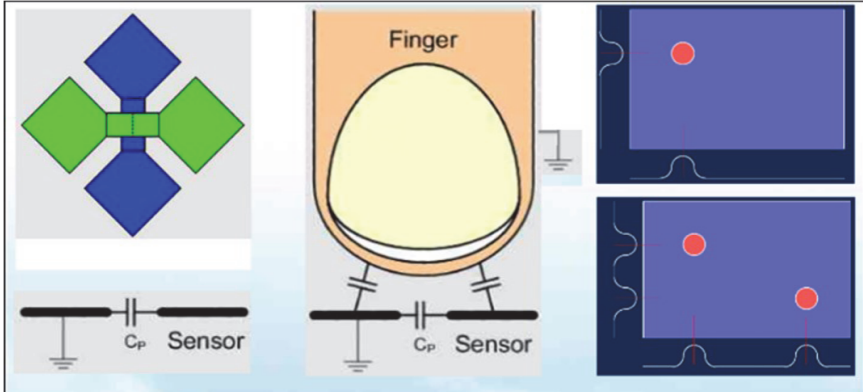


Fig. 8.9. Structural principle of mutual capacitance touch panel.

mutual capacitance is measured, the transverse electrodes send out the excitation signals in turn, and the longitudinal electrodes receive the signals at the same time. Thus, the capacitance values of the intersection points of the transverse and longitudinal electrodes can be obtained, that is, the capacitance values of the two-dimensional plane of the whole touch panel. According to the two-dimensional capacitance variation data of touch panel, the coordinates of each touch point can be calculated, as shown in Fig. 8.9. Therefore, even if there are multiple touch points on the screen, the real coordinates of each touch point can be calculated.

The characteristics of mutual capacitive touch panel are as follows: (1) No calibration is needed. (2) Avoiding the ghost point effect can realize real multi-touch. (3) It is not affected by temperature, humidity, finger wetness, body weight and ground dryness, and will not produce “drift” phenomenon.

8.4. Other Touch Technology

8.4.1. Infrared touch panel technology

Infrared touch panel technology consists of infrared transmitting and receiving sensors installed on the outer frame of the touch panel. On the surface of the screen, an infrared detection network is formed.

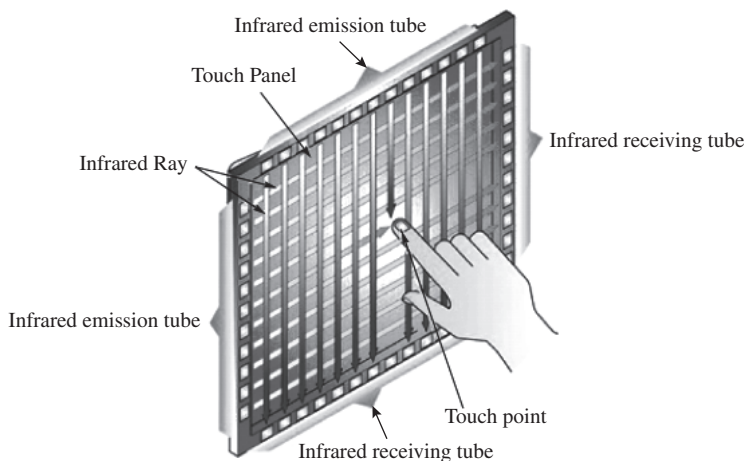


Fig. 8.10. Structure principle of infrared touch panel.

Any touching object can change the infrared line at the contact point to realize the touch panel operation. The principle of infrared touch panel is similar to that of surface acoustic wave touch panel. It uses infrared emitting and receiving sensing elements. These elements form an infrared detection network on the surface of the screen. Touch-operated objects (such as fingers) can change the conducting infrared ray, and then convert it into the coordinate position of the touch-operated to achieve the response of the operation. On the infrared touch panel, the circuit board devices arranged on the four sides of the screen have infrared emitters and infrared receivers, which correspond to the formation of the infrared matrix crossing vertically and horizontally, as shown in Fig. 8.10.

Infrared touch panel is not susceptible to changes in the external environment, such as electrostatic discharge, high current, high voltage, etc. It can be used in harsh environment.

Infrared touch panel has the advantages of low price, convenient installation and fast response; high transparency, no intermediate medium, long service life, high durability, no fear of scratches, long touch life; good use characteristics, no need for strength, no special requirements for touch body. The disadvantage is that the infrared transmitter and receiver are easy to be damaged and have low

resolution; they will be subject to strong infrared interference, such as remote control, high-temperature objects, sunlight or incandescent lamp irradiating the infrared receiver; they will be subject to strong electromagnetic interference, such as transformers.

8.4.2. *Surface acoustic wave touch panel*

Surface acoustic wave (SAW) is a mechanical wave propagating along the surface of a medium. Surface acoustic wave touch panel is composed of touch panel, acoustic generator, reflector and acoustic receiver. The acoustic generator can send a high-frequency sound wave across the screen surface. When the finger touches the screen, the sound wave at the contact point is blocked, thus determining the coordinate position. The touch panel part of the SAW touch panel can be a flat, spherical or cylindrical glass plate mounted in front of the plasma display screen. This glass plate is only a pure reinforced glass, unlike other touch panel technologies, there is no film or cover. The upper left corner and the lower right corner of the glass screen are fixed with vertical and horizontal ultrasonic transmitting transducers, while the upper right corner is fixed with two corresponding ultrasonic receiving transducers. Taking the X -axis as an example, the control circuit generates the emission signal (electrical signal), which is converted into ultrasonic wave in thickness direction by the X -axis emission transducer on the glass screen, and the component propagating along the glass surface is produced by the refraction of the ultrasonic wave through the wedge-shaped seat under the transducer. When the ultrasonic wave encounters a 45° oblique reflection in the course of its advance, it produces a component parallel to the incident wave in 90° and Y -axis, which is transmitted to the other side of the glass screen in the X -direction and 45° oblique reflection. After reflection, the component is transmitted to the X -axis receiving transducer in the opposite direction to the emission direction. The X -axis receiving transducer converts the recovered sound waves into electrical signals. The control circuit processes the electric signal to obtain the waveform which characterizes the sound energy distribution of the glass screen. When touching, the

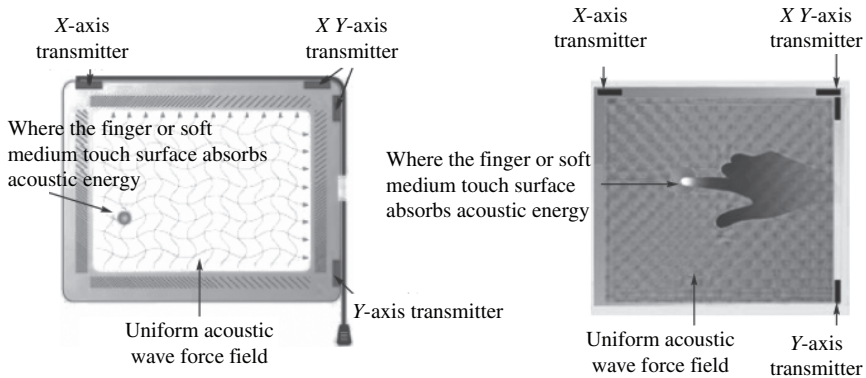


Fig. 8.11. Structural principle of surface acoustic wave touch panel.

finger will absorb part of the acoustic energy, and the recovered signal will attenuate. Program analysis of attenuation can determine the coordinates of the touch points in the X -direction. Similarly, the coordinates in the direction of Y -axis can be judged. Once the coordinates in the direction of X and Y are determined, the touch point will be determined uniquely and naturally. Figure 8.11 is the structure principle of SAW touch panel.

The subtlety of sound screen design is that only one pair of transducers in each direction can detect the whole touch surface; unlike infrared touch panel, at least dozens of transmitter/receiver pairs are needed in each direction. With the increase of screen size or resolution, the number of transmitter/receiver pairs needs to be increased proportionally. The key to its realization lies in the reflection array, whose precise and accurate spacing distribution ensures the consistency of the recovered signals.

Surface acoustic wave touch screen is not affected by temperature, humidity and other environmental factors. It has high resolution, excellent scratch-proof, long life (50 million trouble-free); can maintain clear and transparent image quality; no drift, only need to be corrected once when installed; has the third axis (pressure axis) response, and is most suitable for use in public places. The main disadvantage of SAW touch panel is that it is affected by dust, water droplets, oil pollution, etc. It needs to clean the reflection array

regularly, and the touch panel is relatively large, so it is not suitable for smartphones and other devices.

8.5. The Frontier of Touch Control Technology

8.5.1. *Embedded touch panel*

In 2007, Apple, led by Jobs, released its first smartphone, the iPhone 1, with a capacitive touch panel and scratch-resistant corning glass cover. With its smooth touch experience and fashionable appearance, this mobile phone has led the trend of smart touch terminal devices. Since then, smart phones have become popular all over the world, and its capacitive touch panel has become the standard of smart phones. Nowadays, smartphones are developing in the direction of larger and thinner, in which the thickness of touch panel is more and more stringent, so embedded touch panel emerges at the historic moment. Essentially, the embedded touch panel is to insert the touch electrode of the touch panel into other components of the mobile phone. Unlike the plug-in touch panel, it does not have a separate touch panel component. Embedded touch panel is divided into four structures: OGS (one glass solution), on cell, in cell and hybrid in cell (Fig. 8.12).

8.5.1.1. *OGS technology*

The OGS touch screen directly forms ITO conductive layer and sensor on the protective glass. A piece of glass acts as both cover glass and touch sensor. Its advantages are that it saves the cost of one layer of glass and reduces the cost of one fit, can be processed together at the same time, reduces the weight, and increases the transmittance. Its outstanding advantage is that this structure can reduce the thickness of mobile phones, more fashionable appearance, better user experience, some high-end mobile phones and laptops touch panel using this technology.

8.5.1.2. *On cell technology*

It is one of the development directions of high-end smart phones in recent years to integrate touch sensors into LCD, so that LCD

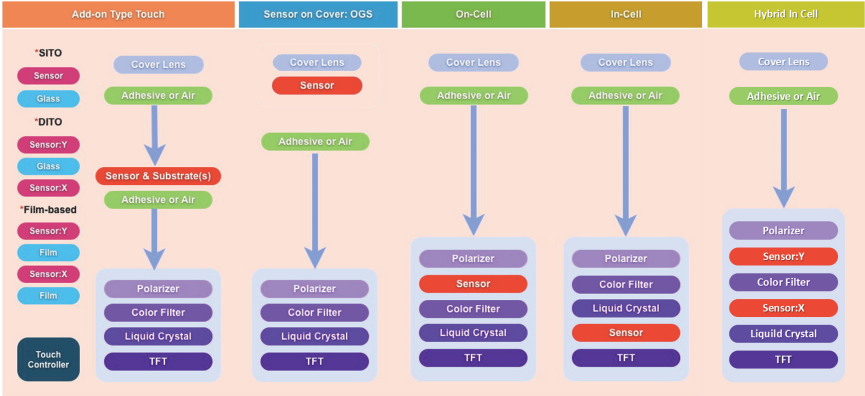


Fig. 8.12. Structure classification of capacitive touch panel.

has dual functions of display and touch. Among them, some smart phones, led by Samsung GALAXY mobile phones, adopt the method of integrating touch sensors into the surface of the substrate on the display panel, which is called on cell touch technology in the industry. In this way, touch electrodes need to be made on the display panel, which requires high process conditions and avoids high-temperature process, so as to avoid the impact on the display panel.

8.5.1.3. *In cell technology*

The high-end smartphone manufacturer, led by Apple Mobile, integrates touch sensor electrodes with TFT drive electrodes of LCD, simplifies the components of mobile phones, and achieves a lighter, more fashionable appearance and a smoother user experience. But this kind of in-cell technology requires higher production process and lower qualified rate, and the interference between touch electrode and display drive electrode needs more high-end processors and algorithm support.

8.5.1.4. *Hybrid in cell technology*

Some mobile phone manufacturers, led by Huawei Mobile, combine the above two technologies and adopt a hybrid touch technology

of on-cell and in-cell. Specifically, the driving electrode and the sensing electrode of the touch electrode are respectively made on the upper and lower surfaces of the substrate on the display panel. This technology has the advantages of the other two technologies. The process is relatively simple and the qualified rate is high.

8.5.2. *Force touch technology*

Force touch technology, also known as pressure touch and force touch, has no unified name at present. It is a touch sensing technology used by Apple for its products Apple Watch, new MacBook and new 13in MacBook Pro. Through force touch, the device can perceive the force difference between light and heavy pressure, and call out different corresponding functions. There are many tiny electrodes around the display screen. Force touch senses the difference between light and press, and then triggers a series of related reactions according to the actual situation. It can sense the force that the user clicks on the touch panel, and give feedback to different functions or controls. Touch control is not only a point control function, so that short points and long press will have different reactions. Apple Watch not only recognizes every touch, but also perceives strength, thus opening up a new space for user control. Figure 8.13 is a force touch technical sketch.

8.5.3. *Flexible touch technology*

The development of wearable technology has led to the development of non-rigid flexible display technology, and the supporting flexible touch technology has also made rapid progress. Traditional screens have been lightweight to a certain extent, but they are also very fragile and fragile. Flexible touch panels are far thinner than traditional screens and feel like a piece of paper. At the same time, because it is a flexible material, flexible touch panel can be applied to various surface shapes of electronic products, not limited to flat-panel materials, and relatively more anti-fall. In addition, the flexible technology of flexible touch panel will also drive the large-scale mobile terminal, making it more convenient to carry. Figure 8.14 is a flexible touch technology product.

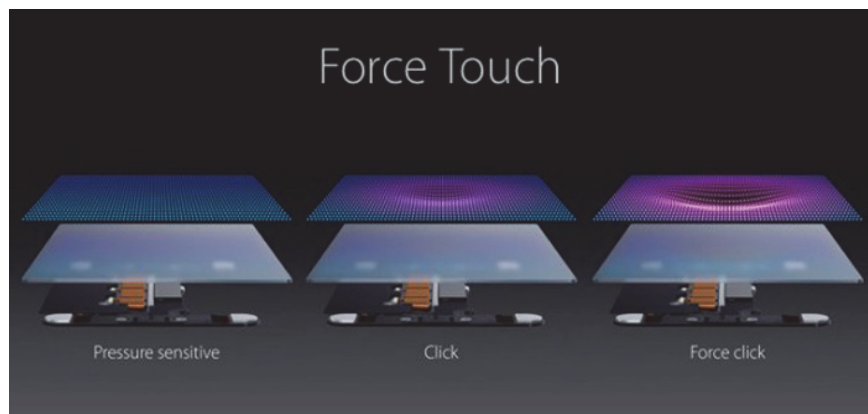


Fig. 8.13. Force touch technology.



Fig. 8.14. Flexible touch technology.

8.6. Introduction of Touch Panel Production Technology

8.6.1. *Production technology of glass touch panel*

The overall process flow of glass touch panel is shown in Fig. 8.15.

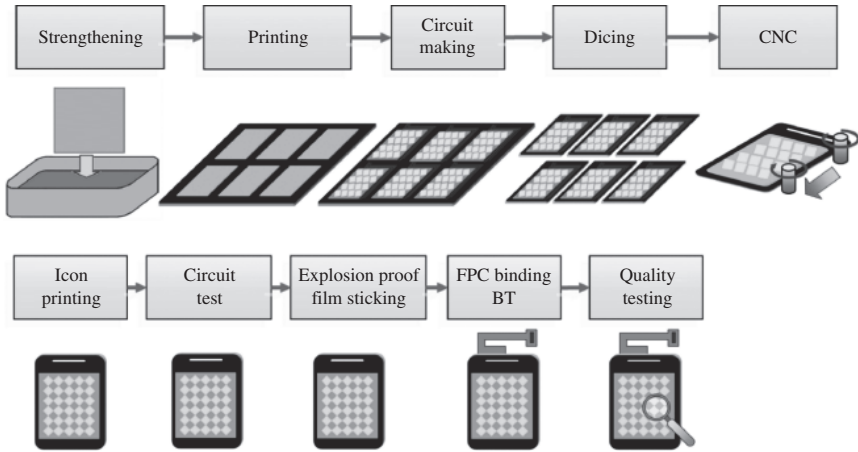


Fig. 8.15. Overall process flow of glass touch panel.

8.6.1.1. *Process of shading layer*

The shading layer refers to the rim ink of the glass touch panel, which is used to beautify the non-window area of the touch panel and to shield the internal structure. Compared with traditional screen printing process, the exposure process has the characteristics of high-dimensional accuracy and low film thickness, which is conducive to the transition between the line on the shading layer and the line in the window area in the later stage. Figure 8.16 is the process diagram of the shading layer.

8.6.1.2. *ITO process*

ITO process refers to the process of making the electrode circuit of the touch sensor in the window area. Firstly, ITO thin-films with a certain thickness are sputtered on the glass which completes the shading process by magnetron sputtering equipment. Then, photoresist materials are coated and exposed under the corresponding film shield. After the development solution is washed, the pattern on the film is transferred to the photoresist. Next, the developed glass is etched in acid etching solution, and the ITO film which is not protected by photoresist is etched. Finally, the photoresist is washed out with the

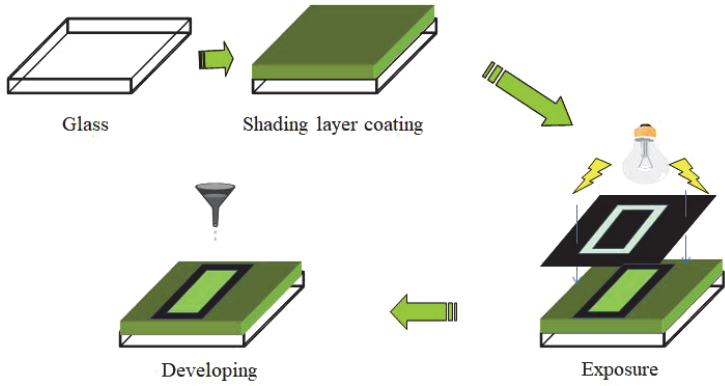


Fig. 8.16. Schematic diagram of shading layer process.

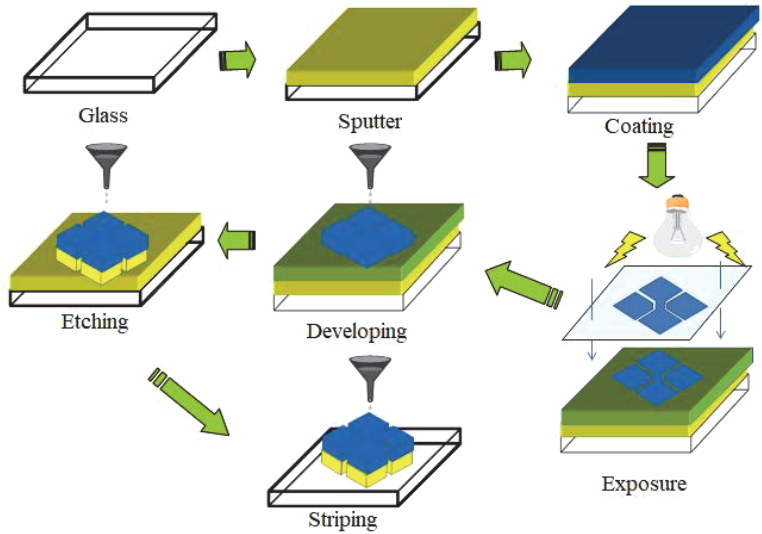


Fig. 8.17. ITO process diagram.

film, and the touch electrode circuit is completed. Figure 8.17 is an ITO process diagram.

8.6.1.3. *OG process*

Because the single ITO cannot form the XY electrode circuit intersecting the capacitive touch panel, it requires OG process and

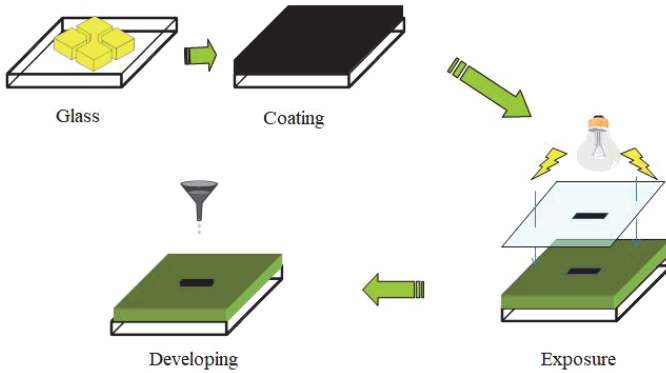


Fig. 8.18. OG process diagram.

metal process to complete the bridging work. Among them, OG process refers to the production of insulating layer, which uses exposure technology. In addition to isolating electrical connections, the insulating layer also plays a role of shielding the metal layer. Figure 8.18 is the OG process diagram.

8.6.1.4. *Metal processing*

Metal fabrication refers to the process of connecting two non-adjacent ITO films together through a metal layer. Firstly, an insulating layer is fabricated by exposure at the place where the bridge needs to be built. Then, a metal conductive layer is evaporated and a photoresist film is coated. The required photoresist patterns are obtained by similar exposure and development processes, and the unprotected metal layer is removed by etching. Then two adjacent ITO films are connected by metal bridges. OG process and metal process require high precision of exposure development, which is the key step of the whole circuit manufacturing process. In addition, the lines on the shading layer are made of metal layer in this process. Figure 8.19 is a schematic diagram of the metal process.

8.6.1.5. *Processing of SiO_2 protective layer*

SiO_2 protective layer means that a layer of SiO_2 is coated on the product which completes the process of touching electrode circuit,

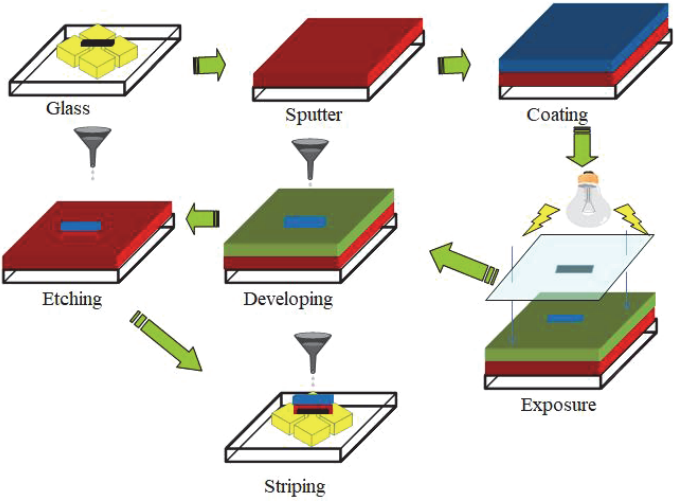


Fig. 8.19. Schematic diagram of metal manufacturing process.

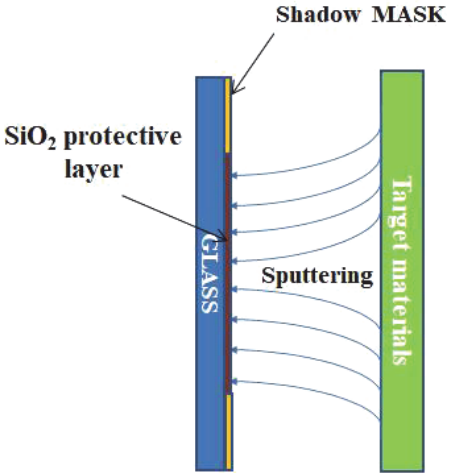


Fig. 8.20. Process diagram of SiO_2 protective layer.

which plays a protective role on ITO circuit. Protective shield is needed in parts that do not need to be coated during the coating process. Figure 8.20 is the process diagram of SiO_2 protective layer.

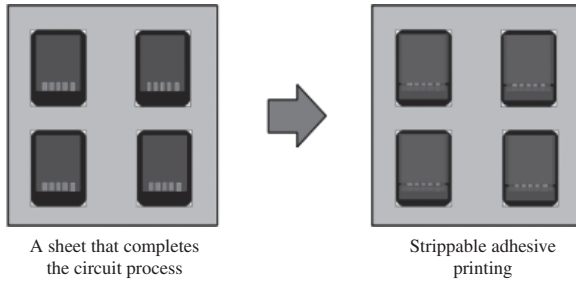


Fig. 8.21. Sketch of peelable rubber processing.

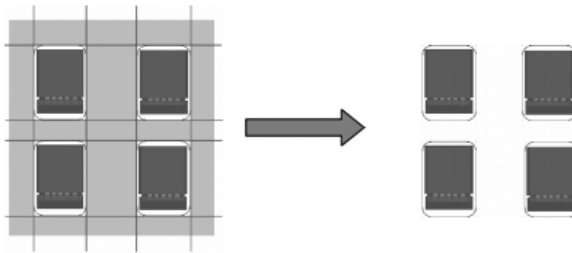


Fig. 8.22. Schematic diagram of cutting process.

8.6.1.6. *Peelable rubber processing*

Peelable glue process refers to the printing of peelable glue on a large piece of glass which has completed the production of circuit. The purpose is to protect the electrode circuit and glass surface during subsequent cutting and CNC process. After the follow-up process is completed, the peelable glue can be directly torn off, which is very convenient. Figure 8.21 is a sketch of strippable rubber manufacturing process.

8.6.1.7. *Cutting process*

Circuit fabrication and other processes can be completed on a large piece of glass, each piece of glass can be made on multi-mode products, which greatly improves production efficiency. Then we need to use special cutting equipment to complete the work of slicing. Figure 8.22 shows the cutting process.

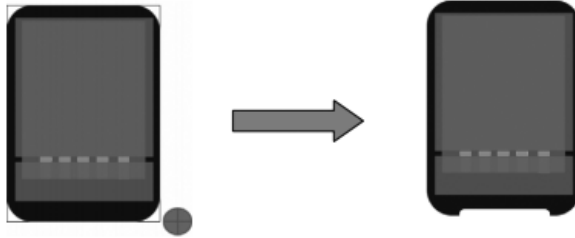


Fig. 8.23. CNC process diagram.

8.6.1.8. *CNC process*

CNC process refers to the precise grinding process for the edges of the fragmented products. It uses a precise CNC carving machine to grind the edges according to the pre-designed graphic size. The grinding process includes the processing of plane shape, such as chamfering, grooving and punching, and the grinding and shaping of section, such as the grinding of 2.5D cover glass. Figure 8.23 is a CNC process diagram.

8.6.1.9. *Secondary intensification process*

Secondary hardening process generally refers to the product that completes CNC process, under the condition of double-sided protection with strippable glue, the cutting cracks of glass segments are micro-etched in certain concentration HF solution to improve its mechanical pressure resistance. This process requires higher control of etching parameters. Excessive etching will result in water ripple and BM shedding at the edge, while insufficient etching cannot eliminate or alleviate the cracks, and cannot achieve the purpose of increasing mechanical pressure resistance. Different kinds of glass also have different requirements for the composition of strengthening liquid. Figure 8.24 is a schematic diagram of the secondary intensification process.

8.6.1.10. *FPC process*

Flexible Circuit Board (FPC) process refers to the electrical connection process between the touch sensor electrode and the Flexible

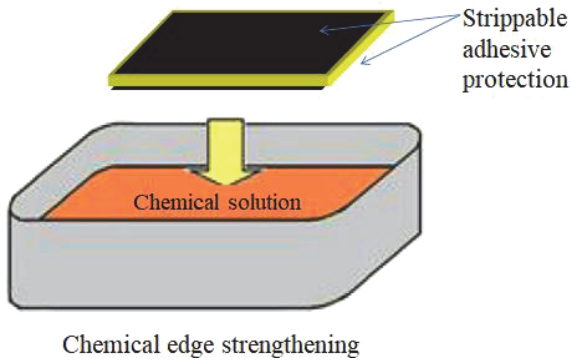


Fig. 8.24. Schematic diagram of secondary intensification process.

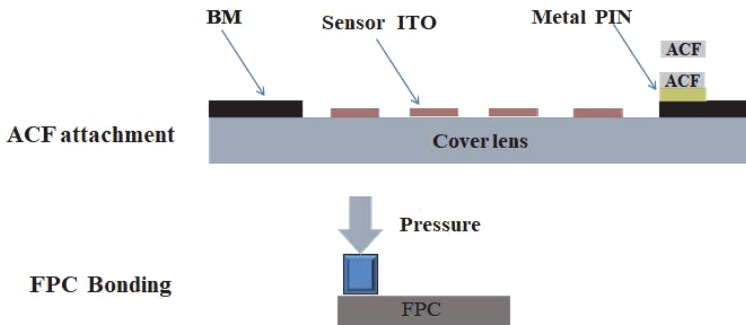


Fig. 8.25. FPC process diagram.

Circuit Board. In general, anisotropic conductive adhesives (ACA) are used as bonding materials in FPC process. Plastic hollow microspheres coated with metal were dispersed into a thermosetting semi-dry adhesive, and anisotropic conductive adhesive was obtained by coating the film. When used, a layer of ACA is pre-pasted on the gold finger of FPC. Then align with the PIN foot of the touch sensor, press it under a certain pressure and temperature, and the FPC and the touch sensor can be turned on. Figure 8.25 is the FPC process diagram.

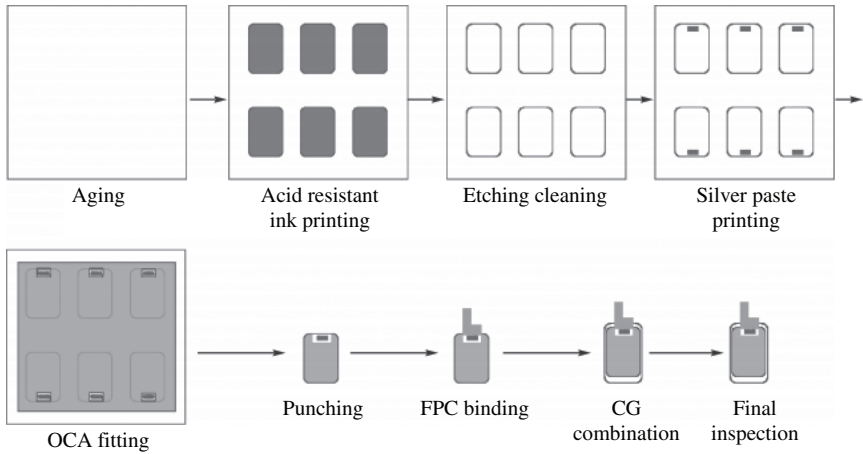


Fig. 8.26. Sketch of production process for thin-film touch panel.

8.6.2. Production technology of film touch panel

The production process of thin-film touch panel is mainly divided into the following processes, including: aging, acid-resistant ink printing, etching cleaning, silver paste printing, OCA fitting, punching, FPC binding, CG combination, etc. Figure 8.26 is the flow chart of the production process of the thin-film touch panel.

8.6.2.1. Aging

ITO films are usually produced by coil-to-coil method. In order to prevent damage to ITO caused by curling and transportation, such as cracking and rising resistance, ITO films are not fully crystallized in order to improve its bending resistance. But correspondingly, the sheet resistance is too high to meet the practical requirements. Therefore, it is necessary to ageing at high temperature before use to improve its crystallinity and reduce the resistance value of the sheet. The specific technology is to bake ITO film at 130–170°C for about 30 min, either in oven or in tunnel infrared oven. Figure 8.27 is a schematic diagram of aging process.

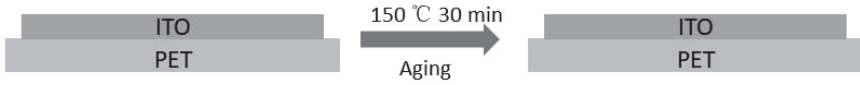


Fig. 8.27. Sketch of aging process.

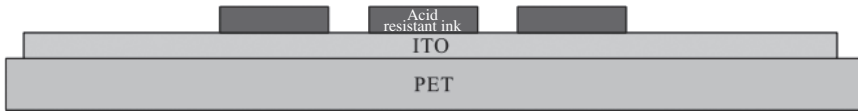


Fig. 8.28. Drawing of acid-resistant ink printing process.

8.6.2.2. *Acid-resistant ink printing*

Similar to ITO process of 8.6.1 glass touch panel, acid-resistant ink is printed on ITO film by screen printing to protect this part of ITO from being etched during etching process, and the corresponding pattern of electrode circuit is transferred to ITO film. Acid-resistant ink is usually cured by ultraviolet irradiation, and the color is more bright, good printability, can print fine electrode circuit. Figure 8.28 is a schematic diagram of acid-resistant ink printing process.

8.6.2.3. *Etching and cleaning*

Etching and cleaning process is to etch ITO in acidic condition and then wash acid-resistant ink in alkaline condition to obtain ITO film with pre-designed electrode circuit. In this process, the concentration of acid and alkali is controlled strictly, and the control of etching time is also key. Figure 8.29 is a schematic diagram of etching cleaning process.

8.6.2.4. *Silver paste printing*

Silver paste printing refers to the printing of silver paste lines on ITO etched films. As the edge line of touch sensor, ITO has higher resistance than conductive silver paste. So screen printing is usually

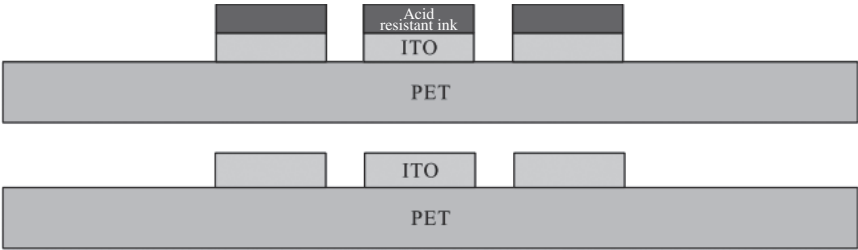


Fig. 8.29. Schematic diagram of etching cleaning process.

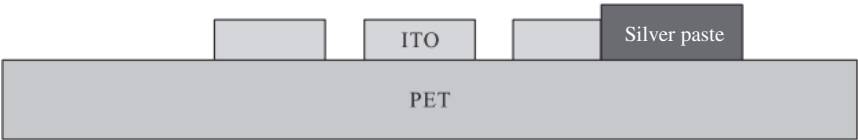


Fig. 8.30. Silver paste printing process diagram.

used to print the edge line of touch sensor on ITO film. This process requires high accuracy, and requires a high precision screen printer with CCD alignment. In recent years, the borders have become narrower and narrower. The technology of using laser silver paste and photosensitive silver lines to obtain narrower borders has also been developed. Figure 8.30 is a schematic diagram of silver paste printing process.

8.6.2.5. *OCA fit*

Film touch panel generally consists of two layers of ITO film, and another layer of cover glass as a protective layer, namely GFF structure. These layered structures are usually bonded with optical transparent adhesives. In the large-scale process, some specific openings are needed in OCA to facilitate later FPC binding. In order to reduce the bonding bubbles, the rolling pressing of COTS is usually used for OCA bonding. Figure 8.31 is a schematic diagram of OCA bonding process.

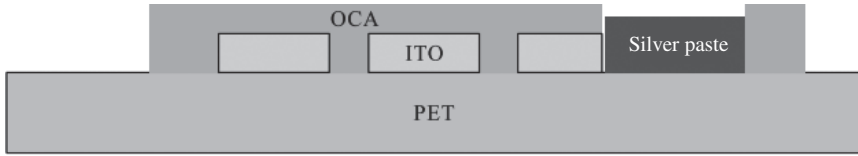


Fig. 8.31. OCA fitting process diagram.

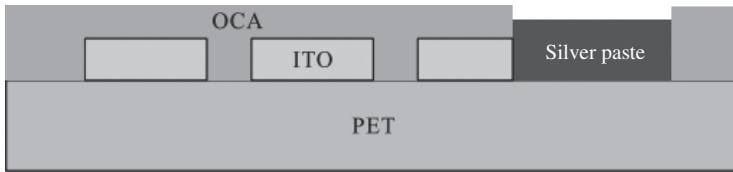


Fig. 8.32. Drawing of punching process.

8.6.2.6. *Punching*

Punching technology refers to the use of die-cutting equipment to complete the large-scale process of product shape cutting to remove redundant part. This process can also be accomplished by laser cutting equipment, but its edge is burned by laser at high temperature, which is not conducive to the subsequent tearing operation. Figure 8.32 is a drawing of punching process.

8.6.2.7. *FPC binding*

For the binding process of thin-film touch panel, except for the lower curing temperature of ACA, other operations are the same as glass touch panel, and their description will not repeated again. Figure 8.33 is a schematic diagram of FPC binding process.

8.6.2.8. *CG Combination*

CG (cover glass) combination refers to the process of bonding the touch sensor and cover glass together, usually by rolling cots.

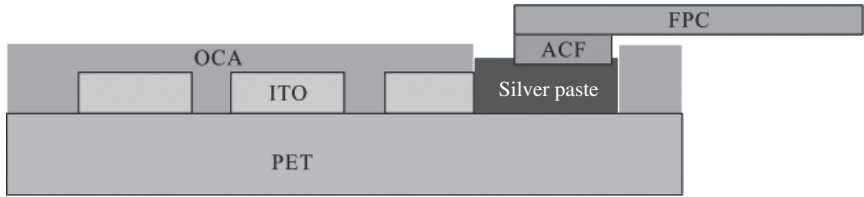


Fig. 8.33. FPC binding process diagram.

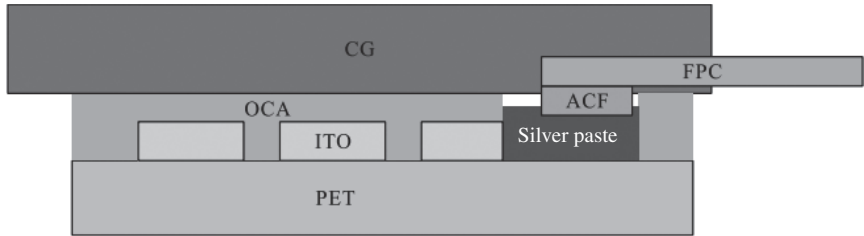


Fig. 8.34. CG combined process diagram.

Because there is a certain height difference between the ink layer and the window area in the glass border area of the cover plate, it is easy to produce the bonding bubbles, which is the difficulty of this process. High pressure heating is usually used to remove bubbles. Figure 8.34 is a CG combined process diagram.

8.6.3. Yellow light process for film touch panel

Yellow light technology refers to the use of exposure development method to fabricate thin-film touch panel touch sensor lines and border wiring, in order to obtain a smaller line width, line spacing, improve the resolution of touch sensor, reduce the border width. The process is generally divided into aging, photoresist bonding, exposure development, etching and peeling, OCA bonding, punching, FPC bonding, CG combination, etc. Figure 8.35 is a diagrammatic diagram of yellow light process for thin-film touch panel.

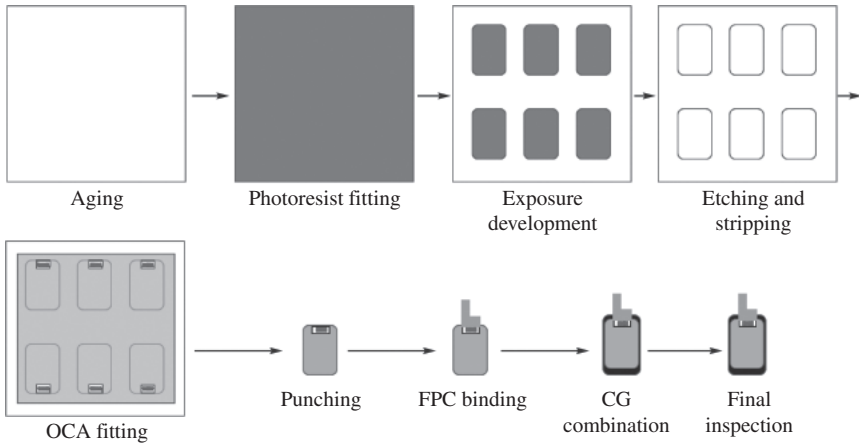


Fig. 8.35. Process diagram of yellow light technology for thin-film touch panel.

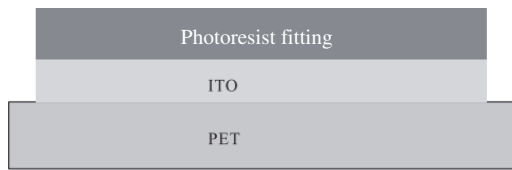


Fig. 8.36. Schematic diagram of photoresist process.

8.6.3.1. Photoresist bonding

Photoresist bonding refers to the bonding of photoresist dry films to ITO films. Generally, the cots are pressed together. Figure 8.36 is a schematic diagram of the photoresist process.

8.6.3.2. Exposure development

Similar to the glass touch panel, the pattern of the touch sensor and the pattern of the frame photosensitive silver are all made by exposure development, and their linewidth and distance are much lower than those of the conventional printing process. This is the core of yellow light technology. The linearity of the light source and the

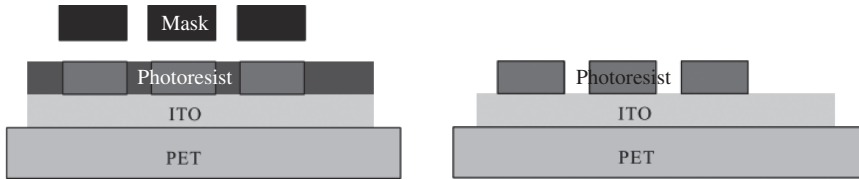


Fig. 8.37. Exposure development process diagram.

resolution of the photoresist determine the precision of the line and the cost. Figure 8.37 is a schematic diagram of exposure development process.

The follow-up process is the same as the conventional thin-film touch panel, and will not be repeated.

8.7. Capacitive Touch Panel Production Equipment and Materials

8.7.1. *Major production equipment of capacitive touch panel*

Large-scale precision screen printer plays an important role in touch panel production. Silver pulp line printing, acid-resistant protection ink printing, insulation printing and border ink printing all need to be used. Large-scale precision screen printer has the characteristics of high printing accuracy, good printing quality and high automation. Figure 8.38 is a large precision screen printer.

UV Curing Machine: Used in acid-resistant ink and other processes requiring UV curing. Figure 8.39 is a UV curing machine.

Infrared Tunnel Furnace: Used for film aging and ink thermal curing process. Figure 8.40 is an infrared tunnel furnace.

Laser etcher: It is used for the processing of ITO, silver paste and other circuits. It is also used for the processing of nano silver wire, graphene and other new materials. Figure 8.41 is a laser etcher.

Laser cutting machine: It is used for cutting the shape of touch function sheet and cutting and grooving of various thin-films. Figure 8.42 is a laser cutting machine.



Fig. 8.38. Large precision screen printer.



Fig. 8.39. UV curing machine.

Soft-to-hard bonding machine: for the bonding of optical transparent glue. Figure 8.43 is a soft-to-hard laminator.

ACA bonding machine: used for ACA (anisotropic conductive adhesive) bonding. Figure 8.44 is ACA laminator.

FOG Hot Press: Used for binding FPC (Flexible Circuit Board). Figure 8.45 is a FOG hot press.

Etching Cleaning Line: Etching Cleaning Line for ITO. Figure 8.46 is the etching cleaning line.

Other touch panel production equipment also includes dispensing machine, electronic magnifier and so on.

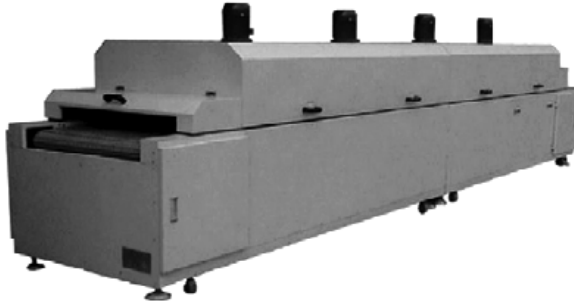


Fig. 8.40. Infrared tunnel furnace.



Fig. 8.41. Laser etcher.

8.7.2. Capacitive touch panel manufacturing auxiliary materials

The production of auxiliary materials for capacitive touch panel has an important impact on the function of touch panel. Failure to meet the requirements of optical performance and weatherability will affect the service life of touch panel. The auxiliary materials mainly include silver paste, OCA, optical transparent adhesive (OTA), anisotropic conductive adhesive (ACA) and so on.



Fig. 8.42. Laser cutting machine.



Fig. 8.43. Soft-to-hard laminator.

8.7.2.1. *Silver paste*

Silver slurry is a viscous slurry consisting of high purity (99.9%) silver particles, adhesives, solvents and auxiliaries. It is an indispensable material for electronic circuit of touch sensor. Conductive silver paste



Fig. 8.44. ACA laminator.



Fig. 8.45. FOG hot press.

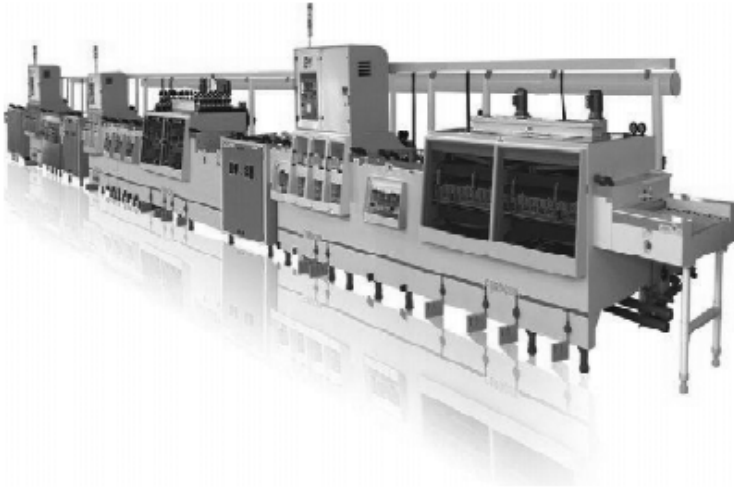


Fig. 8.46. Etching cleaning line.

has strict requirements on its composition. Its quality, content, shape and size are closely related to the properties of silver paste.

Silver particles are the main components of conductive silver paste, and the conductive characteristics of thin-film switches are mainly reflected by it. The content of silver in the slurry is directly related to the conductivity. In a sense, the high content of silver is beneficial to improve its conductivity, but when its content exceeds the critical volume concentration, its conductivity cannot be improved. When the silver content is 80–90% (weight ratio), the conductivity reaches the highest value. When the content continues to increase, the conductivity no longer increases, and the resistance value shows an upward trend. When the content is less than 60%, the change of resistance is unstable. In practical applications, the content of silver particles in silver paste should consider not only stable resistance, but also curing characteristics, bonding strength, economy and other factors. For example, the content of silver particles is too high, the probability of being coated by bonding resin is low, the bonding force of silver conductor decreases after solidification and the risk of silver particles falling off. Therefore, the silver content in silver paste is generally 60–70% suitable.

The size of silver particles is related to the conductivity of silver paste. Under the same volume, the contact probability between the coarse particles is relatively low, and there is a larger space, which is occupied by non-conducting resin, thus forming a barrier to the conducting particles, and the conductivity of the conducting particles decreases. On the contrary, the contact probability of fine particles is increased and the conductivity is improved. The effect of particle size on conductivity is only a relative relationship from the above situation. Due to the influence of processing conditions and screen printing methods, it is very good to control the particle size at 3–5 μm , which is only 1/10–1/5 of the diameter of 250 mesh screen. It can make the conductive particles pass through the mesh smoothly and densely. The collector is deposited on the substrate to form a full conductive pattern.

The shape of silver particles is closely related to their electrical conductivity. From the general impression, particles are only understood as spherical or nearly spherical particles. The conductive particles used for making conductive printing materials are flaky, flat and needle-like, especially flaky particles. The circular particles contact with each other point, while the sheet particles can form the contact between the surface and the surface. After printing, the sheet particles overlap with each other in a certain thickness, thus showing better electrical conductivity. Under the same ratio and volume, the resistance of spherical particles is $10^{-2} \Omega$, while that of flake particles is $10^{-4} \Omega$.

8.7.2.2. *OCA*

OCA is one of the important raw materials of touch panel. It is a kind of double-sided adhesive tape without matrix material, which is made of optical acrylic adhesive without substrate and then bonded with a layer of separating film on the upper and lower layers respectively. It is the best adhesive for touch panel.

OCA has the advantages of high clarity, high transmittance (all-light penetration >99%), high adhesion, high weather resistance, high water resistance, high-temperature resistance, ultraviolet

resistance, thickness control, uniform spacing, long-term use will not produce yellowing (yellowing), peeling and deterioration problems.

OCA can be applied in different fields according to its thickness. It is mainly used for bonding electronic paper, transparent devices, projection screen assembly, aerospace or military optical devices assembly, display assembly, lens assembly, resistance touch panel GFF structure, FF structure, capacitive touch panel, panel, ICON, glass and polycarbonate and other plastic materials. Special adhesives for bonding transparent optical elements such as lenses. It is required to be colorless and transparent, with a light transmittance of more than 90%, good cementing strength, curable at room or medium temperature, and with low shrinkage. Silicone rubber, acrylic resin, unsaturated polyester, polyurethane, epoxy resin and other adhesives can bind optical components. In order to improve the optical properties or reduce the curing shrinkage, some additives are usually added in the preparation. Suitable for fixed mobile machine display peripheral film, screen (acrylic acid, glass screen, touch panel, etc.).

8.7.2.3. ACF/ACA

Anisotropic Conductive Film/Adhesive (ACF/ACA) is relative to isotropic conductive adhesive (ICA).

The fillers in ACF are conductive microspheres with narrow distribution, and the size accuracy is within ($\pm 5\%$). The conduction mechanism is simply that the conductive microspheres will be pinched by the upper and lower conductive contacts when the proper pressure is applied to the glue, while the conductive particles that have been pressed into the gap between the two parallel PIN feet will not contact each other because of the lack of direct pressure compression, so they are relatively insulated.

At present, it is used in COG, TCP/COF, COB and FPC. Other devices are used to drive IC-related components, touch panel wiring, high precision wiring, instead of expensive probe testing. Most notable is ACF conductive film, which is suitable for fixed volume and large output of wiring bonding, requiring high-priced

tape coiling equipment and hot press. ACA (anisotropic conductive glue) is more suitable for the application of a small amount of varied volume. Sony, Hitachi and other anisotropic conductive glues need to be refrigerated at 5°C storage environment and used up within three months, otherwise variation will occur. For manufacturers, high-priced ACA is easy to use, but it costs a lot. In recent years, only Weifeng and Guanpin have successfully developed ACA. Nowadays, Nanowell has developed an ACP that can be stored at room temperature. It has good application prospects in IC, radio frequency identification (RFID), touch panel and other packaging industries. With the continuous development of high and new technology, the demand for anisotropic conductive adhesives has become more and more vigorous in major semiconductor manufacturers and packaging enterprises such as liquid crystal modules.

References

- Chen S S, 2011. *Study on Capacitive Touch-Screen*, Suzhou: Soochow University.
- ELO Touch Systems Inc, 2001. *Projective Capacitive Touch Screen*, US, 6297811.
- IBM, 2007. *Capacitive Two Dimensional Tablet with Single Conductive Layer*, US, 4087625.
- Li W T, Guo Q, 2005. *Latest LCD Display Applications*, Beijing: Publishing House of Electronics Industry.
- Liu X B, 2001. Touch-a new feel for the user interface, *Comput. World* 28:B1–B3.
- Li W W, Liu B, 2011. *Design and Implementation of a Multiple Touch System Based on Infrared Technology Awareness Science and Technology*. Dalian: IEEE, pp. 233–236.
- Ma Y L, Xu Z S, 2009. Strength analyzing and proof test for chemical strengthened optical window glass, *Optics Optoelectron. Technol.* 7(6):56–60.
- Microcrop Technologies, 2012. *Capacitive Touch System Using Both Self-Capacitance and Mutual Capacitance*, CN, 201180012691.1.
- Qiu X X, 2012. *Research on Moire Removal Algorithm Based on CCD Static Image*, Nanjing: Nanjing University of Science and Technology.
- ROHM Co., Ltd., 2010. *Position Input Device*, CN, 201010136588.5.
- Shen Y Z, Zhao L J, Wang C Y, 1996. The method and principle of determining the X- and Y-coordinates of the touch point in front of the screen, *Electron. Sci. Technol.* 3:45–48.
- Scott M, 1985. *Touch Sensitive Control Device*, US, 4550221.
- Weng X P, 2010. *Touch Sensing Technology and Its Application*, Beijing: Beijing University of Aeronautics and Astronautics Press.

- Xing L J, Yang S Z, 2006. Performance and application of touch screen, *Electron. Products* 7:71–73.
- Yuan B Z *et al.*, 2003. A conceptual model and features of new generation (fourth generation) human-computer interactive systems, *Acta Electronica Sinica* 31(12A):1945–1954.
- Zhang X F, 2004. Talking about touch screen technology, *Modern Phys.* 16(3): 43–45.
- Zheng S Y, 2009. *Research on Capacitive Touch Screen*, Shantou: Shantou University.

This page intentionally left blank

Chapter 9

Display Screen

9.1. Introduction of Display Technology

Display technique is an electronic system consisting of one or more display devices to provide visual information. It uses electronic technology to provide flexible visual information. Visual organs (eyes) receive the most information in human sensory organs. In production and life, people need to make more and more use of rich visual information. The task of display technology is to change the intensity of light, the wavelength of light (i.e. color) and other characteristics of light according to the psychological and physiological characteristics of human beings, and to form different forms of visual information. Visual information is generally expressed in characters, graphics and images.

In 1897, Braun, Germany, invented the cathode ray tube (CRT) for measuring instruments to display rapidly changing electrical signals. During World War II, it was also used to display radar signals. After the war, the development of television technology has become an important basis for the development of display technology. In the early 1950s, electron beam tubes began to be used for computer output display. In the early 1950s, electroluminescent display devices were manufactured to explore AC, DC powder type and AC, DC film display technology, and gradually improve brightness and luminescent efficiency. LCD devices were made in the 1960s. During this period, plasma display and light-emitting diode display appeared, and electrochromic display and electrophoretic display were also studied. Since the appearance of laser, the application

of laser in display has been paid more attention and holographic display has been produced. In order to meet the needs of military command center, a variety of large screen display devices have been developed. In the early 1970s, the emergence of microcomputers and the development of large-scale integrated circuit technology have greatly improved the processing components of display equipment. Display software has also been developed accordingly. Therefore, the application of graphics, image and color display equipment based on electron beam tube has entered a new period of development.

Flat panel display industry is one of the “core pillar industries” in the field of electronic information. It is a strategic emerging industry with an annual output value of over 100 billion US dollars. It integrates many disciplines such as optoelectronics, microelectronics, chemistry, manufacturing equipment, semiconductor engineering and materials. It has the characteristics of long industrial chain and cross-cutting in many fields. The pulling effect on upstream and downstream industries is obvious. The output value of products related to display technology in China accounts for about 30% of the total output value of the information industry, and the manufacturing capacity of color TV has accounted for 47% of the total processing capacity of the global color TV industry. China has become a big manufacturing country of new display-related products, and has the technological and industrial basis for leapfrog development.

Flat panel display technology refers to the display technology represented by TFT LCD (thin-film transistor liquid crystal display). It also includes PDP (plasma display), OLED (organic light-emitting display), FED (field emission display), E-PAPER (electronic paper display) and other new display technologies. The overall trend of flat panel display technology will be toward the direction of high quality, high presence, interactive multi-functional integration, energy saving, consumption reduction and health and environmental protection.

In recent years, relevant enterprises, universities and scientific research institutions have made active and effective efforts for the development of flat panel display industry in China, and have made some remarkable achievements under the support of the national

science and technology plan. During the 11th Five-Year Plan period, the Ministry of Science and Technology attaches great importance to the development of TFT-LCD technology and industry. It has developed the whole industrial chain of upstream raw materials, equipment, screen, module and whole machine. At present, it has mastered the production technology and product design technology of TFT-LCD completely and independently, through independent innovation and cross-authorization. Backbone enterprises have mastered core patent technologies such as AFFS and new technologies such as GOA, ODF, 4Mask, 120 Hz driver, dynamic backlight, etc. They have certain patent and technical risk prevention capabilities. From the point of view of product development, they have covered all fields such as mobile phones, laptops, computer monitors, TV and special applications.

In the field of plasma display panel manufacturing, surface discharge barrier plasma display technology is mainly adopted in China. Based on the research of plasma display technology acquired by Korea Orion Company and the plasma display technology acquired by domestic enterprises in recent ten years, aiming at mass production, the cost is reduced by localized mass production and the display performance is continuously improved. At present, we have mastered the mass production technology and the development technology of module products in an all-round way. The qualified rate of mass production has exceeded 90%. The high definition series modules of 42in and 50in have successfully achieved mass production. The full HD module will be mass produced soon after completion of trial production. At present, some progress has been made in the research of materials and devices in the upstream of plasma display industry in China. Most of them have entered the stage of trial production and verification, and large-scale industrial applications have not yet been realized.

Our country has accumulated abundant experience in OLED mechanism research, material development, device structure design and so on, and has done a lot of research work, especially in material and process technology development, and has made great progress and valuable research results. Our country has carried

out fruitful research work on blue-light complex materials, active organic light-emitting display driving technology, small molecular light-emitting materials, interface materials and so on. A series of red fluorescent materials with independent intellectual property rights have been developed. Their lifetime exceeds 15,000 hours (initial brightness is 1000 cd/m^2). It has a promising industrial application prospect. The single-layer structure devices developed in China break through the traditional double-layer and multi-layer structure, simplify the preparation process of materials, equipment and technology, and greatly reduce production costs. In terms of cathode structure, it breaks through the technical barrier of LiF/Al cathode structure, one of Kodak's core patents, and develops new cathode structures of nitrides and amines of alkali and alkaline earth metals. Domestic polymer luminescent materials mainly rely on their own development, although the overall performance of domestic materials is still lower than foreign products, but the rapid development of luminescent materials has greatly improved the efficiency, but the stability still needs to be greatly improved. Great progress has also been made in the research and development of silicon-based and metal oxide-based TFT substrates for active organic light-emitting display and driving IC development in China.

Since 1985, the field emission cold cathode has been studied in our country. After more than 20 years of unremitting efforts, we have carried out research on new field emission materials, field emission theory and the design and fabrication of field emission display devices, which have good foundation and research characteristics. In particular, great progress has been made in the field emission display of printed LWF with low escape power, field emission display of new nanowire cold cathode and field emission display of diamond-like materials. The principled prototype of color field emission display and backlight has been developed on the two technical routes of low escape power printing and micro-nano materials. Our country has acquired important intellectual property rights in field emission display cathode materials, gate structure and device preparation, which lays a good foundation for the research of field emission display in the 12th Five-Year Plan.

In the 1990s, electronic book products with LCD panel appeared in China. Because of the poor display characteristics of LCD panel paper, it is not accepted by people. With the development of new liquid crystal materials, it will be possible to promote the production of a new generation of liquid crystal electronic paper.

Electrophoretic display has become the mainstream technology of commercialized electronic paper display products because of its good performance. The research of electrophoretic display panel technology in China started late, but it made rapid progress. The domestic thin-film transistor liquid crystal display manufacturers also actively invested in the research and development of TFT active electrophoretic display electronic paper panel technology. E-book products with electrophoretic display panel appeared in China in 2006, and the market scale expanded rapidly. In addition, with the development of new liquid crystal materials, it will be possible to promote the production of a new generation of liquid crystal electronic paper.

9.2. Working Principle and Characteristic of Display

9.2.1. Cathode ray tube (CRT) display

CRT is a cathode ray tube display. It consists of five parts: electron gun, deflection coils, shadow mask, phosphor and glass shell. It used to be the most widely used display, but now it has been replaced by the second generation display such as LCD. CRT flat panel display has the advantages of large visual angle, no damage, high color reductivity, uniform color, short response time and so on.

The working principle of CRT is as follows: CRT picture tube emits high-speed electrons with an electron gun, controls the deflection angle of high-speed electrons through vertical and horizontal deflection coils, and finally bombards the fluorescent material on the screen to make it glow. By adjusting the power of the electron beam by accelerating the voltage, different light spots will be generated on the screen, forming various patterns and words.

Each pixel on the screen of a color picture tube is composed of three dyes: red, green and blue. The phosphorescent coatings of these three colors are activated by three electron beams respectively. The required color can be obtained by adjusting the brightness of the three colors with different energy electron beams, similar to the color matching process in painting. If the electron beam is not pointed accurately enough, it may hit the adjacent fluorescent coating, which will result in incorrect color or slight duplication. Therefore, more precise control of the electron beam is necessary.

9.2.2. *Liquid crystal display*

The LCD consists of two glass panels about 1 mm thick, which are evenly separated by a 5-micron liquid crystal material. Because the liquid crystal material itself does not emit light, there is a backlight plate on the back of the liquid crystal display. Its main function is to provide a uniform background light source.

The light emitted by the backlight plate passes through the first polarization filter layer and enters the liquid crystal layer containing tens of thousands of liquid crystal droplets. Droplets in the liquid crystal layer are contained in a small cell structure, one or more cells constitute a pixel on the screen. There are transparent electrodes between the glass plate and the liquid crystal material. The electrodes are divided into rows and rows. At the intersection of rows and rows, the rotatory state of the liquid crystal is changed by changing the voltage. The function of the liquid crystal material is similar to that of a small light passing switch. Around the liquid crystal material are control circuit and drive circuit. When the electrodes in the LCD generate electric fields, the liquid crystal molecules will be distorted, and the light passing through them will be refracted regularly, and then filtered through the second filter layer and displayed on the screen.

LCD technology also has weaknesses and technical bottlenecks. Compared with CRT display, brightness, visual angle and reaction time are weaker than CRT display. The reaction time and visual angle depend on the quality of LCD panel. LCD can be divided into

passive matrix LCD and active matrix LCD according to different control modes.

Passive matrix LCD is limited in brightness and visual angle, and its reaction speed is slow. Due to the problem of picture quality, this display device is not conducive to the development of desktop display, but because of its low cost, passive matrix LCD still occupies part of the market. Passive matrix LCD can be divided into twisted nematic LCD, super TN LCD and double layer STN LCD.

Active matrix LCD, also known as thin-film transistor (TFT) LCD. TFT LCD is a transistor built in every pixel of the screen, which can make the brightness brighter, the color richer and the visual area wider. Compared with CRT display, LCD display has fewer assembly parts, smaller space and lower energy consumption.

9.2.3. Plasma display

The principle of plasma display panel (PDP) is similar to that of fluorescent lamp. PDP is a display technology using gas discharge. Each plasma tube on the screen corresponds to a pixel. The screen uses glass as the base plate. The base plate is spaced at a certain distance and sealed by air tightness to form a discharge space. The discharge space is filled with rare neon, xenon and other mixed inert gases as working medium. The inner side of two glass substrates is coated with metal oxide conductive films as exciting electrodes. When voltage is added to the electrode, plasma discharge occurs in the mixed gas in the discharge space. Ultraviolet radiation is produced by gas plasma discharge. Ultraviolet excitation coats the fluorescent screen with three primary colors (also known as tribasic colors), which emits three primary colors of red, green and blue. When each primary color unit achieves 256 gray levels and then mixes the colors, the color display is realized.

Because PDP does not need backlight, there is no uniformity of angle and brightness of LCD display, and high brightness and contrast can be achieved. Compared with LCD display technology, the larger the screen of PDP, the higher the depth of field and fidelity of image. In addition to the advantages of brightness, contrast and

visual angle, PDP display does not have response time problem and can achieve high dynamic non-smear display. Compared with CRT, PDP display has no scanning line and is completely point-to-point imaging, so the image is clear, stable and without flicker, which will not lead to eye fatigue; PDP also has no X-ray radiation. However, PDP shows the inherent shortage of life, which determines its short vitality and market share is difficult to improve, and ultimately determines that PDP is only a transitional technology.

9.2.4. Organic electroluminescent display

The structure of organic light-emitting display (OLED) is shown in Fig. 9.1.

When driven by external voltage, the electrons injected by cathode and the holes injected by anode migrate from the electron transport layer and the hole transport layer to the organic light-emitting layer, compounding in the organic layer, releasing energy. The organic light-emitting substance molecules are excited, transiting to the excited state, and the stimulated molecules are

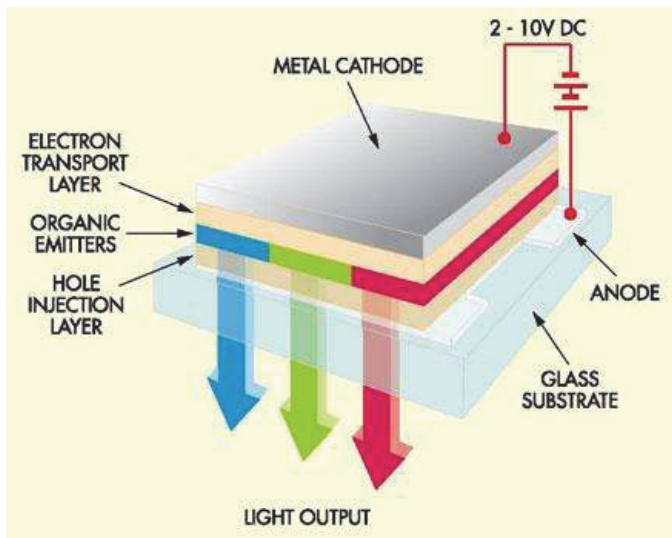


Fig. 9.1. The schematic diagram of OLED device.

generated when they return to the ground state from the excited state. Luminescence phenomenon.

Compared with LCD, OLED can actively luminous, high brightness, high contrast, ultra-thin, low cost, low power consumption, fast response, wide angle of view, flexible display; all solid state, wide operating temperature range, not afraid of vibration, suitable for harsh environment.

The main application fields of OLED include PDA, digital camera, on-board display device, MP3, mobile phone, head-wear display, maritime handheld satellite phone, GPS, digital photo frame, medical instrument, television and other display fields and lighting, military industry, special display (flexible display, micro-display) and many other fields.

9.2.5. LED display

LED panel: It is a kind of display screen that displays text, graphics, images, animation, market, video, video signals and other information by controlling the display mode of semiconductor light-emitting diodes. LED display refers to the display which takes LED (light-emitting diode) as the pixel light-emitting element directly. The light-emitting diode composed of the array emits red, green and blue light directly, and then forms a color picture. However, due to the large diameter of the LED itself and the large distance between the same color pixels, the LED display is usually only suitable for large screen display which can be viewed from a long distance.

LED display screen is divided into graphic display screen and video display screen, which are composed of LED matrix blocks. Picture and text display screen can display Chinese characters, English text and graphics synchronously with computer; Video display screen is controlled by micro-computer, with pictures, text and images simultaneously. It can broadcast all kinds of information in real-time, synchronous and clear way. It can also display two-dimensional, three-dimensional animation, video, television, VCD program and live scene. LED display screen has bright color, strong stereoscopic sense, static as oil painting, movies, and is widely

used in stations, wharfs, airports, shopping malls, hospitals, hotels, banks, securities markets, construction markets, auction houses, industrial enterprise management and other public places. It has the advantages of high brightness, low working voltage, low power consumption, miniaturization, easy to match with integrated circuits, simple driving, long life, impact resistance and stable performance.

LED backlight display is only the backlight of liquid crystal display, which transits from traditional CCFL (cold cathode fluorescent lamp, similar to fluorescent lamp) to LED (light-emitting diode). The imaging principle of liquid crystal can be simply understood as that the external voltage deflects the liquid crystal molecule to change the light transmittance of the backlight, and then projects the light into the color filters of different colors to form an image. The transition of backlight module from cold cathode fluorescent lamp to LED can bring many benefits, which can make the brightness of display screen more uniform, lower power consumption, lighter and fashionable.

9.2.6. Organic Light Emitting Diode (OLED) display

OLED display is the abbreviation of organic light-emitting diode, which can also be called organic light-emitting display. As early as 1936, Pope in the United States discovered that organic matter can emit light at high voltage. But it was not until 1987 that Deng Qingyun, a Chinese researcher at Kodak, prepared sandwich fluorescent devices based on 8-hydroxyquinoline aluminum (Alq_3), that OLED, as a commercially available and excellent flat panel display technology, attracted the interest of academia and industry. At present, hundreds of research units and enterprises have invested in the research and development and production of OLED, including display giants in the current market, such as Samsung, LG, Philips, Sony and other companies. On the whole, the industrialization of OLED has begun. Among them, monochrome, polychromatic and color devices have reached the level of mass production. Large-scale full-color displays have been put on the market in batches and really entered people's daily life.

OLED belongs to carrier double-injection luminescent device. Its luminescent mechanism is as follows: driven by external voltage, electrons injected by cathode and holes injected by anode are combined in organic materials to release energy, and transfer energy to molecules of organic luminescent substances, which are stimulated to transit from ground state to excited state when they are subjected to. When the excitator returns to the ground state, the radiation transition produces luminescence. OLED emits monochrome light such as red, green and blue by driving the organic film itself by electric current, which can also achieve full-color effect.

Various display technologies are described above, such as those shown in Table 9.1.

9.3. Structure and Working Principle of TFT-LCD Device

The main component of the finished Thin-Film Transistor (TFT) LCD is LCD module. LCD module consists of panel and back light. Panel is the core part of the whole LCD, and its manufacturing process is also the most complex. What people usually call bright spot also happens in the manufacturing process of panel board. The quality of backlight directly affects the display effect, and it is usually the key to the life of LCD.

9.3.1. *Liquid crystal materials and their properties*

Liquid crystal is a special form of matter. It is different from solid crystals with fixed shape and anisotropy in optical properties, and from liquid with no fixed shape and isotropy in optical properties. It is a viscous liquid with anisotropy in optical properties. Figure 9.2 shows the molecular arrangement of solids, liquid crystals and liquids.

It has been recognized that liquid crystals with special properties have existed for more than 100 years. It has been found that some solid crystals can be transformed into liquid crystals after being heated to a certain temperature. This kind of liquid crystals is called thermotropic liquid crystals. This kind of liquid crystals is applied in display technology and photoelectric technology. It has also been

Table 9.1. Contrast of various display technologies.

Display technique	Advantage	Inferiority	Application
Cathode ray tube (CRT)	High definition, low cost and mature technology	It is difficult to realize large screen and thin screen; high energy consumption	Color picture tubes and computer monitors (obsolete)
Liquid crystal display (LCD)	Low voltage, low energy consumption, light weight, long life, no electromagnetic radiation	Non-active luminescence, small visual angle, contrast and brightness affected by environmental impact, narrow working temperature range, slow response speed and complex process	The mainstream display used to be widely used
Plasma display panel (PDP)	Simple structure, small volume and thin thickness	Low luminous efficiency, difficult to achieve simultaneous improvement of brightness and contrast, thin display glass, high energy consumption, high driving voltage, high cost	Public information display, wall-mounted large screen television and automatic monitoring system (outdated)
Light emitting diode (LED) display	Small size, light weight and long life	Cannot achieve high-density, high-definition display, high-energy consumption, poor consistency	Instrument display, lighting, advertising screen

Table 9.1. (*Continued*)

Display technique	Advantage	Inferiority	Application
Fluorescent tube display (VFD)	Long life, high brightness and bright color	Cannot display animation and quality is not fine enough	Instrument Display
Field emission display (FED)	Field of view, energy consumption, response time and working temperature are better than LCD.	Technological difficulties and complex processes	Avionics
Organic light-emitting diode (OLED) display	Active luminescence, low driving voltage, high brightness, light and thin, large area, flexible display, high luminous efficiency, wide viewing angle, fast response, low energy consumption, wide operating temperature range	Continuous improvement is needed in luminance, stability and durability	Wide application prospects, a kind of display technology in the future
Quantum dot light-emitting diode (QLED) display	Low cost, high luminous efficiency, light weight, high color reduction, high color purity, large screen display, flexible display	Short life, late start, immature industrial chain	Wide application prospects, a kind of display technology in the future

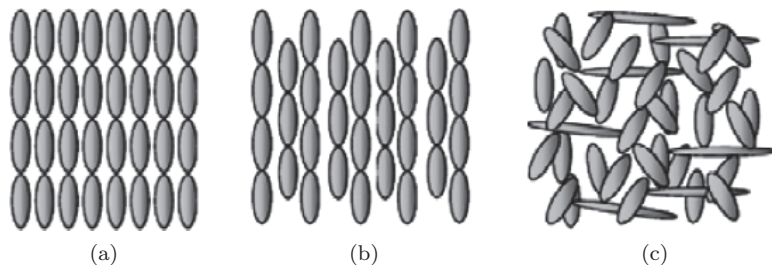


Fig. 9.2. Molecular arrangement of (a) solids, (b) liquid crystals and (c) liquids.

found that some solid substances in plants and animals have liquid crystalline properties after dissolution. These substances often play an important role in the metabolism of organisms or in the process of life. These liquid crystals are called lyotropic liquid crystals. From the molecular structure, liquid crystals can be divided into three types: nematic, near-crystalline and cholesteric.

In near crystal liquid crystals, rod-like molecules are arranged in layers, forming parallel arrangement of molecules, which is nearly perpendicular to the layers. This kind of interlayer bonding is weak and the interlayer is easy to slide with each other. Therefore, the near crystal liquid crystal shows the properties of two-dimensional liquid. But its viscosity is much higher than that of ordinary liquids. In nematic liquid crystals, rod molecules are arranged in parallel in the same way. Each molecule can move freely along the long axis without layered structure. Therefore, it is fluidity-rich and has low viscosity. Cholesteric liquid crystals and near-crystalline liquid crystals also form layered structures, and the long axes of molecules are parallel to nematic liquid crystals in the layers. However, the orientation of the long axis of the molecule is somewhat different between the adjacent layers, and the whole liquid crystal forms a helical structure. The optical properties of cholesteric liquid crystals, such as optical rotation, selective light scattering and circular polarization dichroism, are based on this helical structure.

Although the existence of liquid crystal has been found for a long time, the study of liquid crystal remains in the laboratory stage for a long time after its discovery, and no practical application has been

found. However, since the 1930s, many important achievements have been made in the research and application of liquid crystal materials through the persistent exploration of scientists. The application of plane display is one of the important achievements. The liquid crystal materials used in planar display need to have high birefringence, high dielectric anisotropy and low liquid viscosity. From the chemical structure of liquid crystal materials put into practice at present, there are mainly cholesterol esters, biphenyl aromatic hydrocarbons, diphenylacetylenes, polyacetylenes and so on.

9.3.2. *Structure and working principle of panel plate*

9.3.2.1. *Twisted nematic LCD*

Twisted nematic (TN) liquid crystals are nematic liquid crystals with 90-degree twist. Twisted nematic liquid crystal display (TNLCD) appeared in the 1970s. It not only has the basic characteristics of liquid crystal display, but also has the characteristics of high contrast, simple fabrication technology and low cost. At present, this type of LCD is widely used in portable calculators, clocks and instruments. At present, LCD manufacturers in China produce more of these products.

Twisted nematic liquid crystal display (TNLCD) is formed by two Indium Tin Oxide (ITO) glass panels clamped with twisted nematic liquid crystal material. The thickness of liquid crystal is usually $5\mu\text{m}$. Its specific thickness is related to the birefringence of liquid crystal materials. A layer of orientation layer is coated on the top and bottom ITO glass substrates. By using the interaction force between the liquid crystal molecules and the surface of orientation layer, the liquid crystal molecules are arranged parallel to the orientation direction of the surface friction and have an inclination angle of 2–3 degrees, as shown in Fig. 9.3. The frictional orientation of the upper and lower substrates is 90 degrees, which makes the twist of the liquid crystal molecules 90 degrees. At the same time, a small amount of chiral materials are added into the liquid crystal to determine the twist direction of the liquid crystal molecules. A polarizer is attached to the outer side of the upper and lower

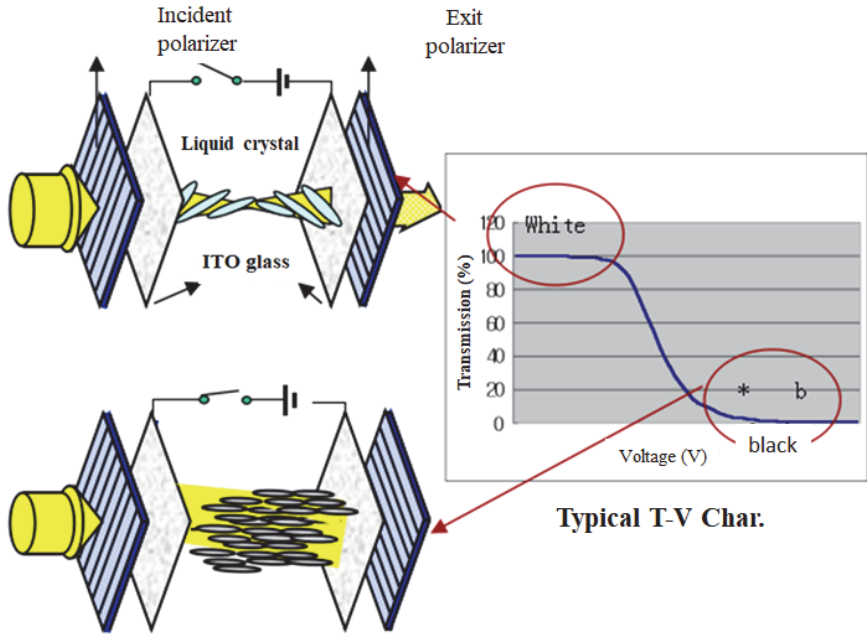


Fig. 9.3. Structure of twisted nematic liquid crystal display (TNLCD).

glass substrates. The optical axis of the polarizer is consistent with the frictional direction of the glass substrates, so that it can be displayed in normal white on the LCD screen. When the polarization plane of the incident light rotates 90 degrees with the liquid crystal molecule, the polarized light passes through the polarizer, and the bright state is obtained. When applied voltage, the positive liquid crystal molecules are arranged in the direction of electric field, and the polarization plane of linearly polarized light is unchanged. The polarized light cannot get dark state through the polarizer on the side of the emitting light, so the liquid crystal display is a controlled light valve. However, under the condition of optimizing the parameters, the maximum number of scanning rows can only reach 32, the information capacity is very small, and the twisted nematic liquid crystal display can only be made into black-white, monochrome and low contrast liquid crystal display (20:1). The angle of view is only 30 degrees, which is relatively narrow, and the maximum size of

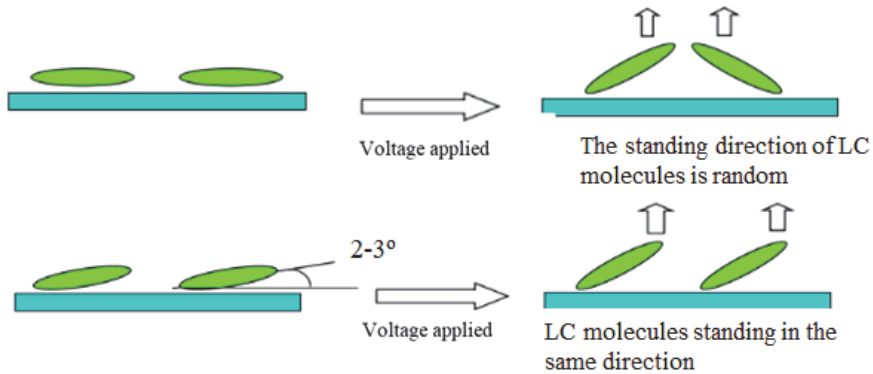


Fig. 9.4. Comparison of rotation directions of LC molecules.

the panel is only 3 inches, so its application scope is limited to a great extent. At present, it can only be used on electronic meters, calculators and simple handheld game machines. If there is no tilt angle, LCD has a high probability of random voltage movement, so it stands at any angle in one direction. Figure 9.4 is a comparison of the rotational directions of liquid crystal molecules.

9.3.2.2. *Thin-film transistor liquid crystal display*

Thin-film transistor (TFT) liquid crystal display (LCD) is an active matrix display formed by introducing thin-film transistor switches into twisted nematic liquid crystal display (LCD), which overcomes the shortcomings of passive matrix display such as cross interference, less information and slow writing speed. It greatly improves the display quality and can be applied to computer high resolution panchromatic display and other fields. At present, thin-film transistors are based on the structure of amorphous silicon thin-film transistors (α -Si TFTAM-LCD). Figure 9.5 is the structure of TFT-LCD panel.

TFT arrays are built on the lower glass substrates. The ITO electrodes of each pixel are connected with the TFT drain electrodes, the gate is connected with the scanning bus, and the source power is connected with the signal bus. When the scanning signal voltage is applied, the signal voltage is applied to the storage capacitor

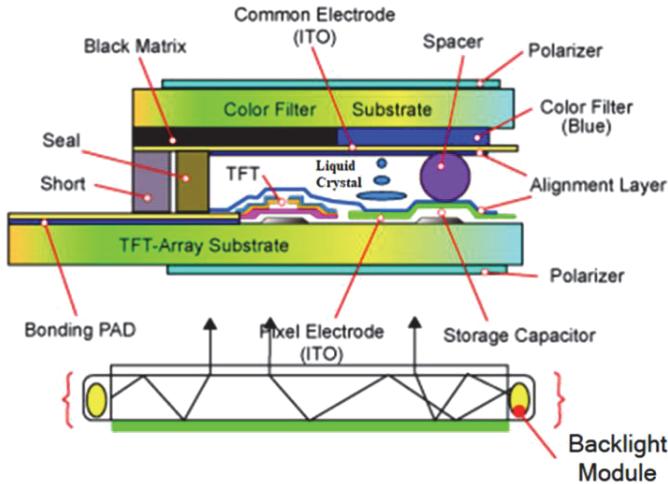


Fig. 9.5. Structure of TFT-LCD panel.

by the conduction of the original source electrode and charged. The signal voltage of the storage capacitor is applied to the liquid crystal pixel in the frame frequency to make it in the on-state. When addressing again, the signal voltage is used to charge or discharge. In this way, each pixel is separated by thin-film transistor switching elements, which not only prevents cross-interference, but also ensures that the LCD response speed is satisfied with the frame rate. At the same time, the gray level can be obtained by storing information. At present, the gray level has reached 256 levels, and 16.7 million colors can be obtained, and almost full-color display can be obtained. Figure 9.6 is a TFT cross-section. Since the formation of the industry in the 1990s, the production line of thin-film transistor liquid crystal display has developed from the first generation to the sixth generation. The area of glass substrates has increased greatly every time, and the output has been continuously increasing and the cost has been decreasing. For example, the glass substrate size of the seventh generation thin-film transistor LCD production line will reach $1870\text{ mm} \times 2200\text{ mm}$, the maximum size of the LCD TV screen can be made up to 94 cm (37 in), the maximum size of the notebook computer screen is 38.1 cm (15 in), and the maximum size

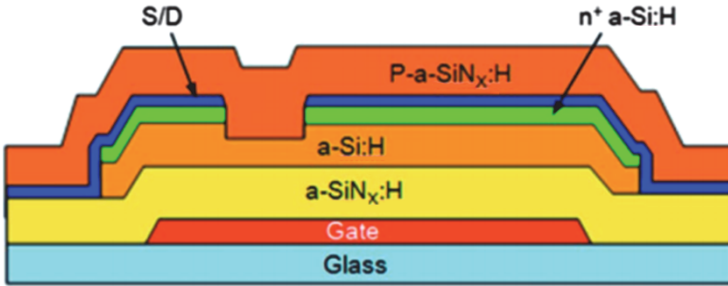


Fig. 9.6. Cross-section view of TFT.

of the monitor screen is 63.5 cm(25 in). Another development trend of thin-film transistor liquid crystal display is thinning, lightweight and low power consumption. Based on the development of new materials, the innovation of manufacturing technology, the improvement of equipment accuracy and automation, and the progress of software technology, the updating speed of thin-film transistor liquid crystal display products is very fast.

9.3.3. Structure and principle of backlight

9.3.3.1. Classification of backlight sources and construction of lamp tubes

Backlight (hereinafter referred to as B/L) is divided into straight down and side light according to lamp arrangement. Side light type needs a guide plate which can direct the light emitted from the side lamp tube to the front of B/L, but straight down type needs no guide plate because the light emitted from the side lamp tube directly emits to the front of B/L. Figure 9.7 and Table 9.2 are two kinds of backlight sources and their performance comparisons respectively.

TFT-LCD B/L light source uses a cathode fluorescent lamp (CFL), which supplies a certain voltage from the outside, emits electrons on the cathode, scans the fluorescent body and emits visible light. The structure of CFL is composed of glass plate, electrode, sealed gas (Hg, Ar, Ne) and fluorescent body. CFL scans the ultraviolet radiation emitted by sealed mercury (Hg) on the

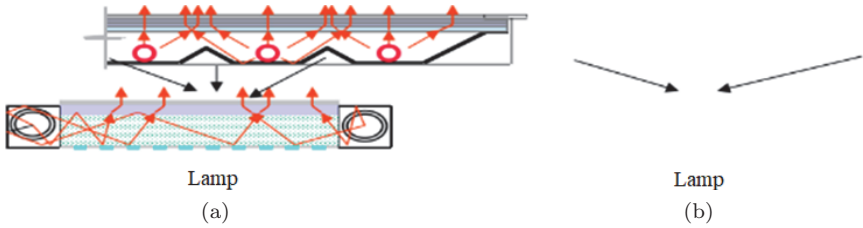


Fig. 9.7. Direct light type (a) and side light type (b).

Table 9.2. Performance comparison between straight down and side lighting.

Category	Straight down	Side light
Weight	Light	Heavy
Brightness	High-brightness	Lower
Consumption of electricity	Large (above 10CCFL cold cathode fluorescent lamp)	Small (above 6CCFL cold cathode fluorescent lamp)
Picture size	Medium and Big Screen (MNT, TV)	Medium and small pictures (full model can be used)
Thickness	Thick	Thin
Is there a light guide?	No (diffusion plate used)	Yes
Price	High-grade	Low-grade

fluorescent body coated on the inner wall of the glass tube to produce visible light. In order to make a small amount of mercury easy to start, and to inhibit the evaporation of cathode material, argon is sealed in glass tube. The types of CFL are CCFL (cold cathode fluorescent lamp) and HCFL (hot cathode fluorescent lamp). Figure 9.8 is the structure of CCFL.

9.3.3.2. Construction of backlight

Lamp is a light source which receives high voltage and produces visible light, mainly using CCFL and HCFL. The lamp box reflects the light source from lamp and incident on the light guide plate, using brass, aluminium and brass composite Ag, etc. The light guide panel of materials is mainly made of PMMA by injection molding or casting

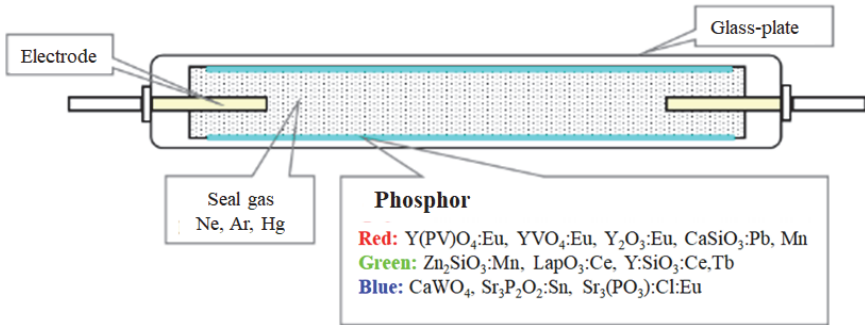


Fig. 9.8. The construction of CCFL.

method. It is used as light source for guiding light incident and has the function of uniform distribution of light source. Reflector is mainly used in PET equipment to reduce the loss of light source when the light guide plate is incident, and it has reflective function. Diffusion sheet (diffuser down) is mainly a spherical light source formed by propylene resin on PET equipment, which uniformly diffuses the light emitted by the self-guide plate and collects the light at the same time. Bottom prism is mainly acrylic resin on polyether (PET) equipment; it can regularly form prism shape and collect light, and the brightness increase rate is 1.55 times of user surface. Top prism has the same function as bottom prism, which increases the brightness by 1.33 times of the surface of bottom prism. Prism is arranged cross-wise to collect X -axis and Y -axis light sources. Diffuser up (protector film) has the same structure as diffuser down to protect prism, also known as protective film. In order to use the diffuser with transmittance, the top prism collector will lose more or less the light source, but it will be used to reduce the bad characteristics of the top prism. Figure 9.9 is the structure of backlight.

9.4. Material Technology and Technology of TFT-LCD

9.4.1. TFT-LCD material technology

TFT-LCD material technology includes material design, selection and use. The display principle of TFT-LCD is closely related to

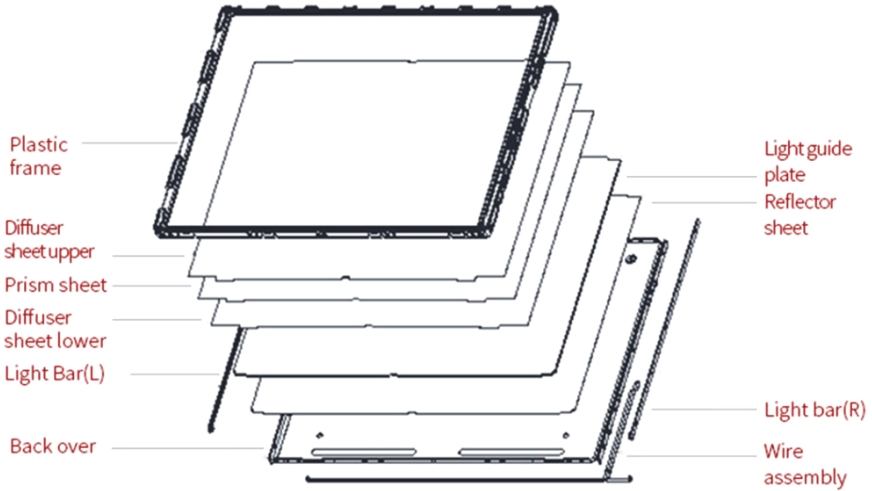


Fig. 9.9. Structure of Backlight module.

the materials used. The materials needed for making TFT-LCD include glass substrates, metal targets, chemical materials, liquid crystals, CF substrates, polarizers, circuit elements, backlights, etc. In addition to various chemical gases used in TFT substrates, chemical materials also include alignment dies, sealing clothes and clearances.

9.4.1.1. *Glass substrate*

Glass substrate is a basic component of LCD. This is a very flat surface float method to produce thin glass sheets. The glass substrates currently used commercially have a thickness of 0.7 mm and 0.5 mm, and are about to enter a thinner (e.g. 0.4 mm) thickness process. Basically, a TFT-LCD panel needs two glass substrates, which are used as the substrates of the bottom glass substrates and color filters respectively. General glass substrate manufacturers supply about 1:1.1 to 1:1.3 for LCD panel assembly plants and color filter processing plants. The glass substrates used in LCD can be divided into alkali glass and alkali-free glass. Alkali glass includes sodium glass and neutral boron silicate glass, which are mostly used

in TN and STN LCD. The main manufacturers are Nitrogen Nitrate (NHT), Asahi Nitrate (ASAHI) and Central Glass, etc. Floating process is the main production method. Alkali-free glass is mainly alumino silicate glass, with SiO_2 , Al_2O_3 , B_2O_3 and BaO as its main components. The total content of alkali metal is less than 1%, which is mainly used in TFT LCD. The leading manufacturer is Corning Company, USA, which mainly produces by overflow melting process.

Only Corning, Asahi, Japan and other four manufacturers can provide large-scale LCD screen glass substrates. Among them, Corning of the United States accounts for 51% of the market, Asahi of Japan accounts for 28% of the market, and only these two companies can provide a complete set of production lines for more than five generations. Although glass substrates only account for 6–7% of the cost of TFT-LCD products, the technological oligopoly makes glass substrates become the main parts of TFT-LCD upstream materials. The TFT-LCD glass project developed by Rainbow and Dongxu in China should receive special support and encouragement.

The properties of ultra-thin flat glass substrate mainly depend on the composition of glass, while the composition of glass affects the thermal expansion coefficient, viscosity (strain, annealing, transformation, softening and working point), chemical resistance, optical penetration and absorption, and electrical properties at various frequencies and temperatures. The quality of products depends not only on the composition of materials, but also on the production process. In the application of TN/STN and TFT-LCD, glass substrates require surface properties, heat resistance, drug resistance and alkali metal content. The main physical characteristics affecting glass substrates for TFT-LCD are described below.

- (1) Strain point: As an index of glass densification, it must be resistant to high temperature in the production process of photoelectric liquid crystal display.
- (2) Density: For TFT LCD, notebook computers are the largest market at present. Therefore, the smaller the density of glass substrates, the better, so that they can be easily transported and carried.

- (3) Thermal expansion coefficient: This coefficient will determine the proportion of expansion or contraction of glass material due to temperature change. The lower the coefficient, the better, so as to minimize the thermal expansion and cooling of large screen.

Other physical properties such as melting point, softening point, chemical resistance, mechanical strength, optical properties and electrical properties can be regulated according to the specific needs of users.

In the whole process of glass substrate, the main technology includes three parts: feeding, sheet forming and post-processing. The feeding technology is mainly controlled by the formula. First, glass raw materials are melted into low viscosity and uniform glass melt in a high-temperature furnace. Not only the physical and chemical properties of glass should be considered, but also the optimum formula of raw materials should be selected without changing the chemical composition, so as to effectively reduce the melting temperature of glass and clarify the glass, at the same time achieve the specific properties of glass and meet the requirements of practical application. Thin plate forming technology is related to dimensional accuracy, surface properties and whether further processing and grinding are needed to achieve special physical and chemical characteristics. Later processing includes glass separation, grinding, cleaning and heat treatment.

At present, there are three main process technologies for producing glass substrates for flat panel displays: float technology, slot down drawing and overflow fusion technology. Because of the horizontal extension, the surface of the floating method will produce scars and bumps, which need to be grinded and processed. Therefore, the investment amount is higher, but it has the advantages of producing a wider glass product (the width can reach 2.5 m) and a larger production capacity (about 100,000 m²/month). The overflow melting process has many advantages, such as better control of surface characteristics, no grinding and simpler process. It is especially suitable for producing ultra-thin flat glass with thickness less than 2 mm. However, the width of the glass produced is limited to

less than 1.5 m, and the productivity is low. Floating method can be used to produce glass substrates for various flat panel displays, while overflow melting method is currently only used to produce TFT-LCD glass substrates. The following three processes are described separately.

(1) Floating method: the most famous flat glass manufacturing technology at present, this method transfers the melted glass paste in the furnace to the liquid tin bed. Because of the low viscosity, the thickness of glass can be controlled by baffle or pull rod. With the increase of the distance flowing through tin bed, the glass paste gradually solidifies into flat glass. The solidified flat glass is drawn out by guide wheel and then processed by annealing and cutting.

In the production of ultra-thin flat glass by float method, a lower amount of glass paste should be controlled. The ribbon which enters the tin bed should be cooled to about 700°C. At this time, the viscosity of the ribbon is about 108 g/(cm s). The edge roller is used to pull the glass paste floating on the liquid tin and stretch it outward. After that, the glass belt is heated to 850°C and pulled by external force with the conveyor belt roller.

Floating technology uses horizontal extraction method, so it is easier to use the production line with elongated horizontal direction to meet the requirements of annealing. The main reason why float technology has not been widely used in the production of ultra-thin flat glass with thickness less than 2 mm is that it cannot meet the required economic scale. For example, the daily output of floating technology can almost meet the current monthly consumption of Taiwan market; if the ultra-thin flat glass is produced by floating technology, it is usually produced by discontinuous tank, so the optimization of tank kiln design is very important.

(2) Flow hole down-drawing method: As far as the special ultra-thin flat glass for flat panel display is concerned, many manufacturers use the flow hole down-drawing method to produce it. In this method, homogeneous glass paste with low viscosity is introduced into slot bushing trough made of platinum alloy, and the thickness of glass is controlled by gravity, pull-down force and the size of die opening. Temperature and the size of the orifice jointly determine the glass

output, while the size of the orifice and the downward drawing speed jointly determine the glass thickness, and the temperature distribution determines the warpage of the glass. The ultra-thin flat glass is drawn by the downward drawing method of the orifice. Ultra-thin flat glass with thickness of 0.03–1.1 mm and thickness of 5–20 t per day can be produced by the process of down-hole induction. Because platinum metal cannot withstand high mechanical stress, most of the platinum alloy molds are made of platinum alloy. However, the flow hole often deforms when it bears external force, which has the shortcomings of uneven thickness and surface flatness that cannot meet the requirements of specifications. The downward flow hole method must be annealed vertically. If it is turned horizontally, it may increase the contact between the glass surface and the roller and the warpage caused by the horizontal conveying, resulting in a large increase in the unqualified rate. Such concerns make it necessary to adopt the design of elevation in the construction of melting furnace. At the same time, the height required for annealing must be taken into account accurately, which greatly increases the difficulty of the project and also reflects the cost of the plant.

(3) Overflow melting method: The melted glass paste is transported to the center of the melt pump by a long strip melt pump, and then two outflow glass pastes are synthesized under the pump by overflow method. The use of this forming technology also requires the aid of dies, so the melt pump die also faces the problems of mechanical stress deformation, maintaining the level of the melt pump and how to stably inject the melt glass paste into the melt pump. Because the thickness and surface quality of super-flat glass fabricated by overflow melting method depend on the amount of glass paste, stability, level, surface properties of the pump and the amount of glass introduced into the melt pump. Overflow melting technology can produce ultra-thin glass substrates with double original glass surfaces. Compared with float method and downward drainage method, it can avoid post-processing processes such as grinding or polishing. At the same time, it is not necessary to pay attention to the difference of glass surface properties caused by the contact between original and liquid tin, or grinding media, which has become the mainstream of ultra-thin

flat glass forming. Because alkali-free glass has special ingredient formula, and has strict standard specifications in thermal stability, mechanical, electrical, optical, chemical characteristics, appearance size, surface smoothness and so on, its production line adjustment and learning time are longer, and the technical threshold for new manufacturers to join the industry is higher.

9.4.1.2. ITO thin-films

ITO is a kind of transparent conductive oxide (TCO). Because of its best combination of conductivity and transparency, ITO has become the most important transparent conductive material. It is mainly used in liquid crystal display, touch screen, solar thin-film battery, organic light-emitting elements for lighting and other fields. The structure of the display screen is shown in Fig. 9.10. Indium oxide only absorbs ultraviolet rays, does not absorb visible light, doped with tin, although it will lose the transmittance, but it can improve

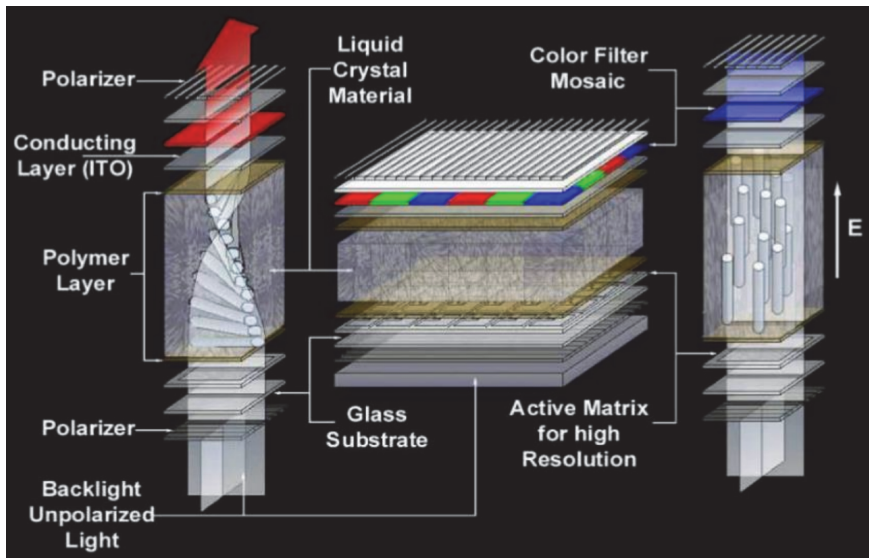


Fig. 9.10. Construction of display screen.

the conductivity. Therefore, transmittance and conductivity are two interdependent indicators. ITO raw material is ITO powder, known as ITO target. ITO conductive film is formed when ITO target is deposited on PET substrate, and ITO conductive glass is formed when ITO target is deposited on glass substrate. Now the most mature and widely used ITO deposition technology is magnetron sputtering, that is, bombarding the target with high-energy particles, making atoms sputter out of the target and forming thin-films on the surface of the substrate, which is better than vacuum evaporation, pyrolysis spraying, chemical vapor deposition, sol-gel method and so on.

The conductivity of ITO thin-films is related not only to the composition of ITO thin-films (including tin content and oxygen content), but also to the process conditions (including substrate temperature, sputtering voltage, etc.) when ITO thin-films are prepared. The factors affecting the conductivity of ITO thin-films include sheet resistance (γ), film thickness (d) and resistivity (ρ). The relationship among these three factors is $\gamma = \rho/d$. In order to obtain ITO thin-films with different sheet resistance, it is actually necessary to obtain different film thickness and resistivity. From the above formulas, it can be seen that the conductivity of ITO thin-films is better (that is, the resistance value of ITO thin-films is lower), and the thickness of ITO thin-films increases with a certain resistivity, which will lead to the decrease of the transmittance of ITO thin-films, and vice versa. Therefore, the resistivity of ITO thin-films is the key to the preparation of ITO thin-films. In order to obtain good transmittance and conductivity, the resistivity (ρ) of ITO thin-films should be as small as possible.

The factors affecting resistivity (ρ) mainly include carrier concentration, carrier mobility and so on. The higher carrier concentration or carrier mobility, the smaller the resistivity of thin-films. In terms of control conditions, carrier concentration can be achieved by adjusting tin and oxygen content of ITO deposited materials, while carrier mobility is related to the crystalline state and crystal structure of ITO films, and can be achieved by adjusting deposition temperature, sputtering voltage and film forming conditions.

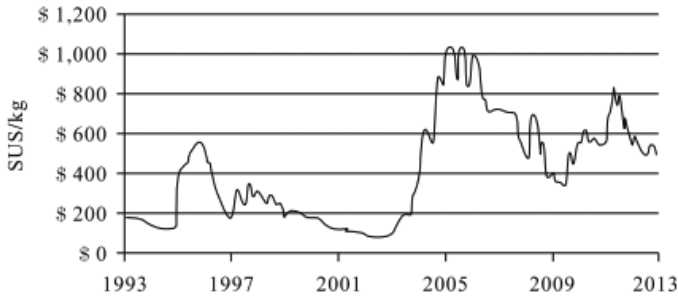


Fig. 9.11. Price trend chart of indium in the past 20 years.

At present, the shortcomings of ITO are as follows: (1) Indium, the main component, is expensive. Indium is a rare metal. It distributes in the earth's crust relatively small and scattered. It mainly exists in cassiterite and sphalerite in trace amounts. With the popularity of liquid crystal display and touch screen products, the price of indium has risen several times. Figure 9.11 shows the price trend of indium in the past 20 years. Seventy percent of copper consumption is in the production of ITO conductive materials. (2) Deposition technology needs expensive vacuum deposition equipment and high maintenance cost in vacuum environment; less than 30% of ITO targets are sputtered on the substrate during deposition, and the rest are sputtered on the wall of the chamber, resulting in a great waste of raw materials. (3) ITO has relatively high resistivity. With the increase of screen size, the resistance will continue to increase, affecting screen brightness and sensor responsiveness. With the increase of the number of electrodes, the area of the border wiring will also increase. On the larger touch screen, although the wiring electrodes can be taken out at both ends of the long sensor electrodes and other methods to deal with, they greatly increase the process difficulty and cost. (4) ITO is brittle and the processing difficulty will increase with the increase of ITO size. Moreover, due to the lack of flexibility and flexibility, it is not suitable for flexible touch screen.

The industry has been developing materials that can replace ITO, such as other transparent conductive oxides, conductive polymer materials, nanomaterials, metal nets, etc. In the field of solar

thin-film cells, cost plays an important role in product competition. The other two transparent conductive oxides, FTO and ZnO, have lower cost although their conductivity is not as good as ITO. In addition, solar thin-film cells do not require uniform and smooth film surface, but have a certain degree of concave and convex, to improve the scattering ability of transmitted light (i.e. fog). FTO and ZnO have better light scattering ability than ITO, and laser etching is easier than ITO, so they gradually replace ITO in the field of solar thin-film cells. However, in the field of LCD and touch screen, there is no commercial alternative material that can be produced on a large scale. Especially in the field of touch screen, the touch screen industry led by smart phones, tablets and so on is developing rapidly. The manufacturing cost of large-scale touch screen and the application of flexible touch screen are all restricted by ITO characteristics.

9.4.1.3. *Directional membranes*

The alignment layer is the key material to control the quality of LCD display. In order to achieve good rotation effect, alignment layer must be coated on the inner side of the upper and lower electrode substrates of LCD, and then rubbing process is carried out. The surface of alignment film will form grooves arranged in a certain direction because of friction. The liquid crystal materials on the alignment film will achieve align effect because of the force between molecules. Thus, the alignment of liquid crystal molecules in a specific direction and a predetermined tilt angle can be controlled, which will be beneficial to the action of the display. Polyimide (PI) resin is the most widely used alignment material because of its high transparency, uniform liquid crystal alignment, high charge retention, high heat resistance and chemical stability.

9.4.1.4. *Liquid crystal materials*

Liquid crystal is a material state between liquid and crystalline state. In addition to some properties of liquids and crystals (such as fluidity, anisotropy, etc.), it also has its unique properties. The study of liquid crystals has developed into a noticeable subject.

Liquid crystal materials are mainly aliphatic, aromatic, stearic acid and other organic compounds. Liquid crystals also exist in biological structures, and soap aqueous solutions of appropriate daily concentration are also liquid crystals. There are thousands of liquid crystal materials synthesized by organic compounds. Due to the different environmental conditions, liquid crystals can be divided into two categories: thermotropic liquid crystals which only exist in a certain temperature range; liquid crystalline phases which appear after some compounds dissolve in water or organic solvents are called lyotropic liquid crystals. Lyotropic liquid crystals are related to biological tissues. Studying the relationship between liquid crystals and living cells is one of the contents of biophysical research nowadays.

Thermotropic liquid crystals (TLC) are liquid crystals formed by homogeneous mixtures of a single compound or a few compounds. The liquid crystal phase usually appears within a certain temperature range. Typical long rod thermotropic liquid crystals have molecular weights ranging from 200 to 500 g/mol, and the aspect ratio of molecules ranges from 4 to 8. Thermotropic crystals are classified into three types according to the arrangement of rod molecules: nematic liquid crystal, smectic liquid crystal and cholesteric liquid crystal.

Nematic liquid crystals: Its molecules are rod-shaped and tend to align in the same direction locally. Molecular short-range interaction is weak, and its arrangement and movement are relatively free. Molecular arrangement makes its viscosity small and fluidity strong. Nematic liquid crystal is the main material of liquid crystal display devices because of its optical properties of uniaxial crystals and its sensitivity to external effects.

Near crystal liquid crystal: The molecule of near crystal liquid crystal is rod-like, and the molecule is arranged in layers. The direction of long axis of each layer molecule is the same, but the long axis of molecule is at a certain angle with the layers. The thickness of the layer is approximately equal to the length of the molecule, and the distance between the layers can vary. Because of the strong molecular binding force in the molecular layer and the weak interlayer binding force, this liquid crystal has fluidity, but its viscosity is larger than

that of nematic liquid crystal. Because of the positive birefringence of smectic liquid crystal, the properties of smectic liquid crystal display devices are superior to those of nematic liquid crystal display devices.

Cholesteric liquid crystal: Its molecules are arranged in flat layers, the long axes of the molecules are parallel to the plane of the layers, the long axes of the molecules in the layers are parallel to each other (corresponding directions), the long axes of the molecules in the two adjacent layers have a small twist angle, and the orientation vectors of the molecules in each layer rotate continuously and uniformly along the normal direction of the layers, which makes the whole structure of the liquid crystal form a spiral structure. The distance between the two planes of helical torsion 360 degrees is called pitch, which is expressed by L , usually in the order of 10^2 nm. This special helical structure makes the crystal have obvious optical properties such as optical rotation, circularly polarized light dichroism and selective light scattering. Therefore, cholesteric liquid crystals are often used as additives to control the alignment of liquid crystal molecules or as chromotropic liquid crystal films directly.

Lyotropic liquid crystal (LLC) is a liquid crystal formed by two or more compounds including solvent compounds, which occurs when the concentration of solute molecules in solution is within a certain range. Its solvent is mainly water or other polar molecular liquid agent. The aspect ratio of long rod solute molecules in lyotropic liquid crystals is about 15. The main reason for the long-range ordering of molecules in this liquid crystal is the interaction between solute and solvent molecules, while the interaction between solute molecules is secondary. The biofilm has the characteristics of lyotropic liquid crystal.

The liquid crystal molecules are disk-shaped, bowl-shaped and other shapes, but most of them are slender rods. According to the arrangement of molecules, liquid crystals can be divided into three phases: near-crystal phase, nematic phase and cholesteric phase, among which nematic phase and cholesteric phase are most frequently used.

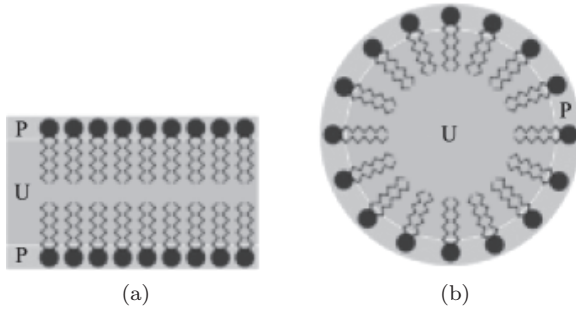


Fig. 9.12. Lyotropic liquid crystal structure: (a) double deck and (b) micelles.

Figure 9.12 is a lyotropic liquid crystal structure. The surfactant molecules on the head contact with water while the tail is impregnated in oil.

9.4.1.5. *CF substrate*

Color filter (CF) is an optical filter that represents color. It can accurately select small bands of light to pass through and reflect other undesirable bands. Color filters are usually installed in front of the light source, so that the eye can receive a saturated color light, including infrared filters, green filters, blue filters and so on. Compared with UV filter and VD filter, CF filter is the general term of colored filter, such as contrast filter, color separation filter, LB filter, etc. Color filter is the key component of liquid crystal display. LCD is a non-active light-emitting component. Its color display must provide light source through the internal backlight module (penetrating LCD) or the external environment incident light (reflecting or semi-penetrating LCD). Then it is combined with driving IC and liquid crystal to control the formation of gray scale. Then through R, G and B color layers of color filters, chromacity is provided to form a color display screen. The color layer formed by the pigment dispersion method is similar to the yellow light microlithography process of semiconductor. Firstly, the dispersive color photoresist is coated on the glass substrate which has

formed a black matrix. After soft bake, aligned exposure, developed photoresist stripping and post bake, the three-color pattern of R, G and B is formed three times by repeating the process.

9.4.1.6. *Polarizer*

The imaging of liquid crystal display must depend on polarized light. All liquid crystals have two polarizers, front and back, which are closely attached to the liquid crystal glass to form a liquid crystal with a total thickness of about 1 mm. If there is no polarizer, the LCD cannot display the image. The basic structure of polarizer includes the middle polyvinyl alcohol (PVA), two layers of cellulose triacetate (TCA), pressure sensitive film, release film and protective film. In order to protect the physical properties of the polarizing film, a layer of high light transmittance, good water resistance and mechanical strength (TCA) film was compounded on both sides of the PVA to protect the polarizing film, which formed the polarizing plate. In the production of ordinary TN-type LCD polarizer, according to different requirements, it is necessary to coat PSA with a certain thickness on one side of the original plate of the polarizer, and to compound the isolation film for PSA protection; on the other side, according to the product type, the composite protective film, reflective film and semi-transparent and semi-reflective film are respectively used to form the finished polarizer. For STN-type LCD polarizer products, the phase difference film and protective film with positioning phase difference compensation value should be compounded on the side of PSA layer according to the different needs of customers, thus forming STN-type LCD polarizer products, which is the basic structure and function principle of LCD polarizer.

The pressure sensitive adhesives used are high-temperature resistant, moisture proof and pressure sensitive adhesives, and special dipping treatment of PVA (dye series products). The polarizer produced is a wide temperature type polarizer. The anti-ultraviolet polarizer can be made by adding a component to the pressure sensitive adhesive used to prevent ultraviolet rays from passing through. A polarizer for STN can be made by compounding a birefringent optical

compensation film on the transmission film. A wide-angle polarizer or a narrow-angle polarizer can be made by compounding a light diversion film on the transmission film. The color polarizer is used to color the pressure sensitive adhesive, PVA film or TCA film. In fact, with the development of new LCD products, more and more types of polarizers are available.

9.4.1.7. *Backlight*

Backlight is a kind of light source behind liquid crystal display (LCD). Its luminous effect will directly affect the visual effect of liquid crystal display module (LCM). The LCD itself does not emit light. It displays graphics or results of light modulation. Backlight is mainly composed of light source, light guide plate, optical module and structural parts.

Light source: There are three types of backlight: EL, CCFL and LED.

Light guide plate: It is divided into printing, chemical etching, V cut, stamper and internal diffusion.

Optical Molds: Light-enhanced film/plate, diffusion film/plate, reflector, black/white glue.

Structural parts: There are backboard (iron backboard, aluminium backboard, plastic backboard), rubber frame, lamp holder, aluminium profile and aluminium base bar in the structural parts, among which backboard and rubber frame are necessary parts, while other structural parts are not fully used.

With the rapid development of international information technology industry, the related LCD industry is constantly pushing forward the old and bringing forth the new. The size of LCD products is developing towards diversification and portability. Backlight, as one of the core components of LCD products, is bound to cooperate with this trend and devote itself to the diversification and portability of products.

The core technology of backlight module is optical technology of light guide plate. There are mainly two types of light guide plate,

printing and injection molding. Others, such as injection molding plus printing, laser dotting and corrosion, account for a small proportion, which are not suitable for the principle of mass production. Printing has become the mainstream technology for a long time because of its low cost, but the low qualification rate has always been its main shortcoming. LCD products require more precise guide plate structure, and injection-shaped guide plate will inevitably become the mainstream of backlight development. However, the corresponding technical difficulties of die and mold can only be overcome by a few large Japanese factories.

LCD always requires high brightness of backlight, but high brightness also makes the power consumption of LCD remain high. Backlight, as the most power-consuming part of LCD module, cannot meet the requirements of portable information products. Therefore, it is not suitable for portable products to increase the brightness of backlight without increasing power consumption, and then increase the brightness of LCD without increasing power consumption. Therefore, it is also one of the main development trends to improve the brightness of backlight and then increase the brightness of LCD without increasing power consumption. The current goal is to achieve a brightness of 500–700 cd/m² and a lifetime of 100,000 hours for flat-panel fluorescent lamps.

9.4.2. TFT-LCD technology

9.4.2.1. Manufacturing process of thin-film transistor (TFT)

The manufacturing process of TFT is divided into six major processes: hard film, cleaning, photo, etching, film removal and detection, as shown in Fig. 9.13.

(1) Hardening process

Hard film process refers to the process of forming film on glass (grid) by physical or chemical methods on gate electrodes, data (source/drain) electrodes, pixel electrodes, insulating film, protective film and semiconductor film.

The gate electrode, data (source/drain) electrode and pixel electrode are metal materials (aluminium, chromium, ITO,

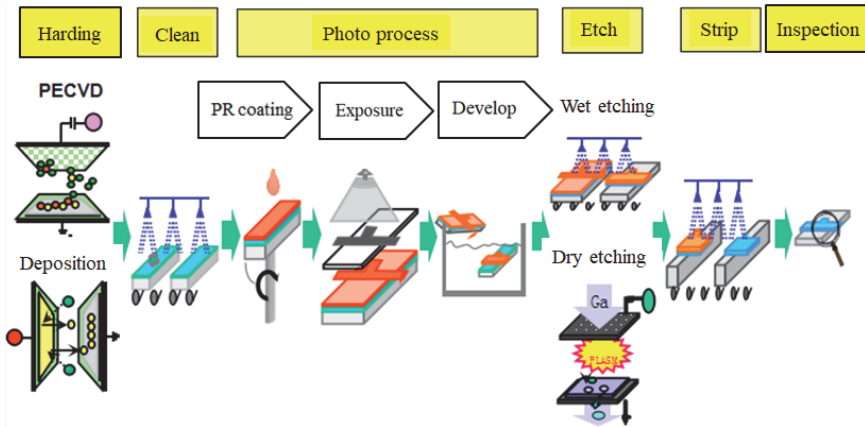


Fig. 9.13. TFT manufacturing process diagram.

molybdenum). The target material is plated on glass by sputtering physical method in the ionic region between target (mainly metal) and glass. Ion region (plasma) is the formation of ionization by applying high voltage to the inert gas injected between two electrodes. The ionized inert gas impacts the target material, and then the removed target material moves to the glass to form a film. Figure 9.14 is the hardening process.

Insulation film, protective film and semiconductor film are chemical PECVD (plasma enhanced chemical vapor deposition) processes, that is, plasma formed by injecting inert gas between two electrodes and applying high frequency power supply to form film on glass.

(2) Cleaning process

Cleaning process refers to removing the contamination of glass or membrane surface in the initial stage or process in advance, so as to avoid bad process, which is helpful to ensure the adhesion between the membrane and the membrane. The main devices are UV cleaning device, brushing cleaning device, mega sonic unit ultrasound, CJ (cavitation jet) unit cavitation jet device.

(3) Photo process

Photo process refers to the process of forming a mask to be made on the film to pass through light, and its shape moves from

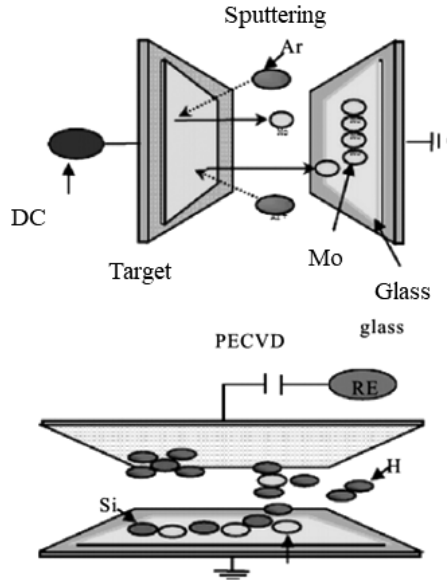


Fig. 9.14. Hardening process.

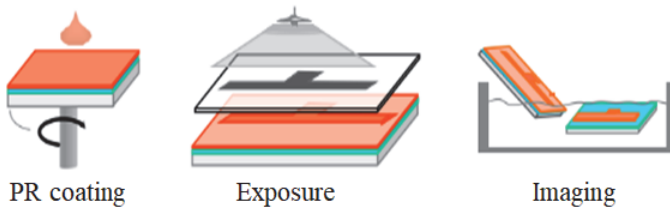


Fig. 9.15. Photo process.

the mask to the photosensitizer (PR), including the process of photosensitizer coating, exposure and imaging. Figure 9.15 is a photo process.

(4) Etching process

The etching process refers to the process of selective removal of the film of the photosensitizer (PR) by physical and chemical methods. There are two ways of etching: (1) wet etching is a chemical solution etching of metal materials (aluminium, chromium, ITO, molybdenum); (2) dry etching is a gas plasma etching of SiN_x , a-Si. Figure 9.16 is the etching process.

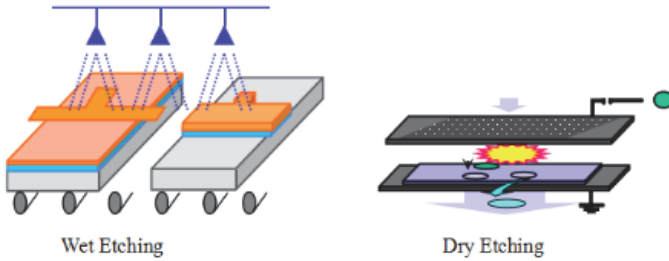


Fig. 9.16. Etching process.

(5) De-film process

Stripping process refers to the process of removing the photosensitizer (PR) left behind for patterning after etching process. The necessary condition of the membrane removal process is to remove PR completely, and the lower membrane should not be damaged, but also to maintain a uniform surface state for the next process.

(6) Inspection procedure

Inspection process refers to the investigation/evaluation process, semi-finished products, product quality determination of good or bad process.

9.4.2.2. Cell/screen making process

The assembly process of TFT-LCD panel (screen). Firstly, the washed color film substrate and the TFT array substrate are coated with a directional film coating solution, and the frictional orientation is achieved. Then, the TFT array substrate is coated with encapsulation glue, and the spacer of 5–10 μm size is scattered on it as a support point. Then, the array substrate and the color film substrate are combined to form an empty cell by encapsulating the encapsulation glue. Next, liquid crystal is injected in two ways. One way is to cut off the empty base plate and take the size needed for the final display product. After inspection, liquid crystal material is injected into vacuum and sealed. The other way is to inject liquid crystal first, then cut off the sheet and then seal. The production time of these two methods is different, which will affect the total qualified

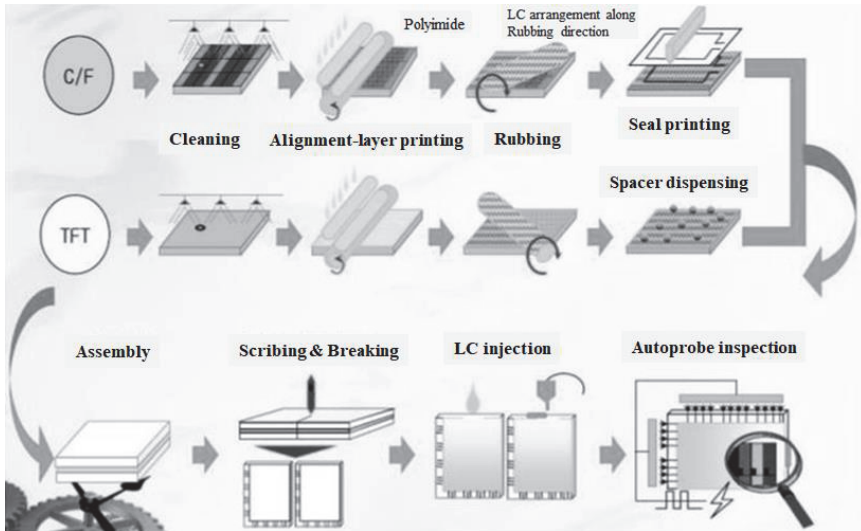


Fig. 9.17. Cell manufacturing process.

rate and also result in different production capacity. Figures 9.17 and 9.18 are the manufacturing processes of boxes.

The process flow of the module: Finally, the LCM module is completed by connecting the TFT-LCD panel with drive IC and printed circuit board, and installing backlight and fixed frame. Figure 9.19 is a component of the module process. The process includes polarizer bonding, TAB bonding, PCB bonding, B/L assembly, aging testing and packaging, as shown in Fig. 9.20. In addition, each LCM should be checked before and after aging.

Polarizer bonding is mainly divided into two processes: cleaning and bonding. Cleaning includes knife washing, brushing, washing and drying: knife washing is to remove glass debris by rotating blades; brushing is to remove dust and fingerprints by brushing; washing is to remove residual impurities by pure water; drying is to remove micro-moisture by hot air at high and low temperatures. Polarizer bonding refers to the use of information from the panel and the polarizer to attach the upper and lower polarizer plates to the panel respectively. Figure 9.21 is the cleaning diagram, and Fig. 9.22 is the polarizer fitting diagram.

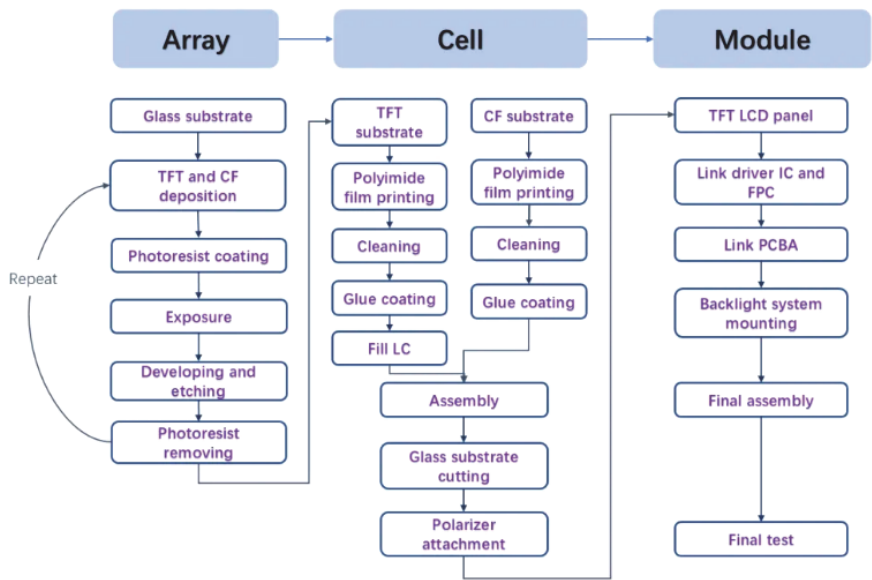


Fig. 9.18. TFT LCD Manufacturing flow chart.

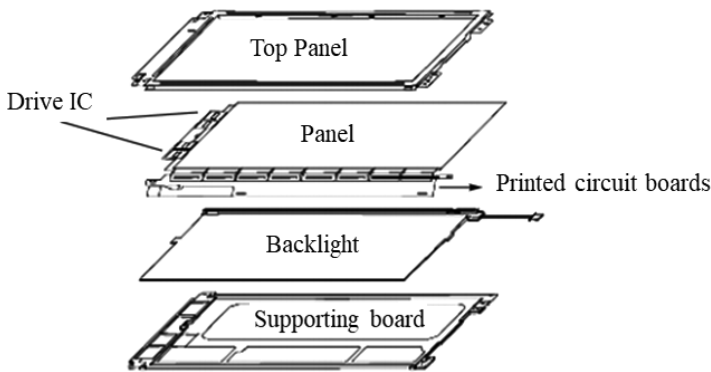


Fig. 9.19. Components of modular process.

TAB bonding refers to the use of ACF (anisotropic conductive film) to connect TAB (tape automated bonding) to the panel. The process includes ACF attachment, TAB positioning and this pressing: ACF attachment refers to attaching ACF to the panel; TAB

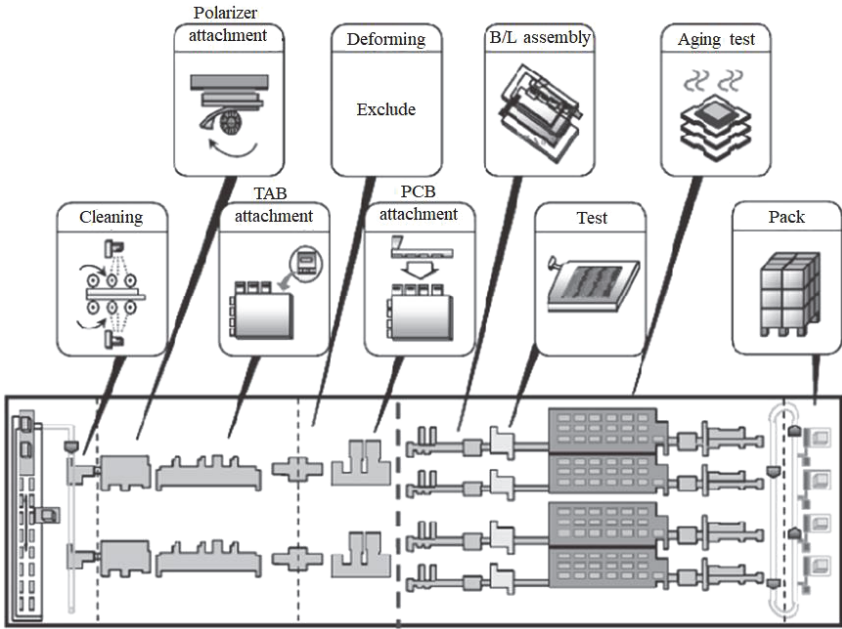


Fig. 9.20. Detailed description of module manufacturing process.

positioning refers to pre-pressing TAB on the panel using positioning information on the panel; this pressing is to completely press TAB on the panel under high temperature and pressure and make the ACF of the connecting part open. Figure 9.23 is a TAB map.

PCB bonding and TAB bonding are all connected by ACF, but the difference is that the connection here is TAB and PCB, so because of different materials, ACF used is different, and the engineering conditions are different. Specific processes are divided into resin coating, ACF bonding and PCB formal pressing: resin coating is to prevent moisture and other foreign bodies from entering the panel; ACF bonding is to attach ACF to the PCB panel; PCB formal pressing is to press PCB and TAB under high temperature and pressure, and make the connecting part of ACF conductive. Figure 9.24 is a PCB map.

B/L assembly is the process of connecting the top board, the back board connecting TAB and PCB, the backlight board and the supporting motherboard manually.

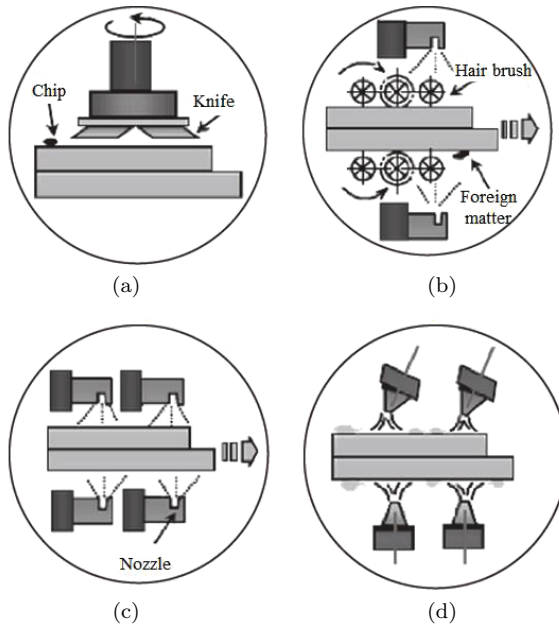


Fig. 9.21. Cleaning chart: (a) knife washing; (b) brushing; (c) flushing, and (d) drying.

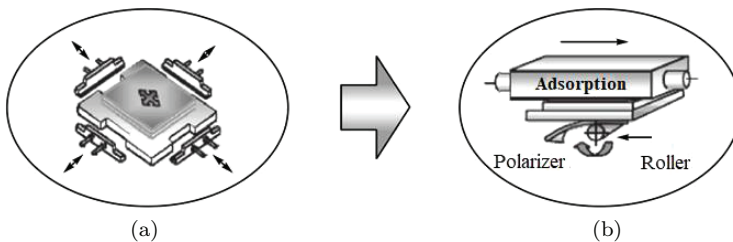


Fig. 9.22. Polarizer matching map: (a) location determination and (b) sticking the upper/lower plates.

Aging test is to place the assembled LCD module in the aging room at 50°C to test the connection status of the module. According to the customer's requirements, the aging time of the module has its own length.

Packaging is also a manual way to test qualified LCD module packaging out of the factory.

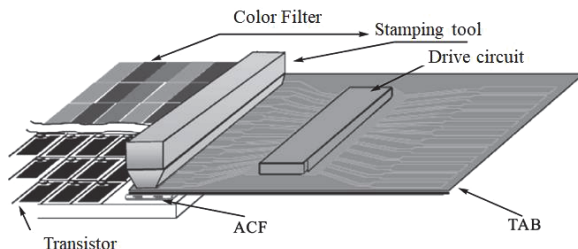


Fig. 9.23. TAB match map.

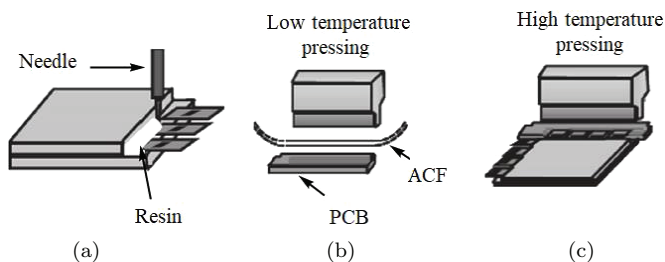


Fig. 9.24. PCB match map: (a) resin-coated screen, (b) PCB to ACF, and (c) PCB formal pressing.

9.5. OLED Display Screen

9.5.1. Principle and application of OLED display

At present, OLED's two main application markets are display and lighting. Among them, display application develops rapidly. At present, there are many consumer electronic products, such as mobile phones, MP3/MP4, digital cameras, automotive displays and so on, which are imported into PMOLED panel or AMOLED panel. Especially, under the impetus of brand manufacturers such as Samsung, Nokia and HTC, the flagship mobile phone application using AMOLED panel display gradually pushed to the mid-end positioning market. Figure 9.25 is the application of OLED display technology.

9.5.2. Fabrication of OLED devices

Because OLED is sensitive to water and oxygen, the device fabrication process must be carried out in vacuum or inert gas protection



Fig. 9.25. Application of OLED display technology.

environment. After the organic film is formed, the sample must be packaged or passivated immediately. Therefore, for OLED display, circuit fabrication using traditional semiconductor technology (such as photolithography, etching and deposition) should precede thin-film technology.

Typically, traditional OLED devices are fabricated on transparent ITO conductive glass anodes, and the light emitted by the organic layer is transmitted through transparent ITO and glass substrates. Thin-film transistors (TFT) are needed for AM drive technology. In OLED display, at least two TFT are required for a pixel due to the diode characteristics of the device. In order to provide stable electron-hole pairs, the carrier mobility in thin-film transistors should be kept at a relatively high level, with longer lifetime and open voltage stability. For small molecule OLED, organic materials are sublimated by thermal evaporation under high vacuum. Accurate deposition rate (0.01 nm/s) can be achieved by controlling temperature. For polymer materials, because of their high molecular weight, it is necessary to prepare them by solution process, such as spin coating and inkjet printing. When the device is encapsulated under inert gas protection, a device with longer life can be obtained. The passivation layer can also be deposited directly on OLED, which is more conducive to reducing the thickness of the device and can be applied to flexible substrates.

9.5.2.1. *Fabrication of OLED thin-films*

Small molecule OLED materials are stable and easy to sublime in vacuum. Usually, thin-films are prepared by vacuum evaporation. Organic layers and organic materials are deposited on the graphical ITO substrates in turn. Organic layers and metal cathodes are controlled by masks. Deposition rate of organic matter is regulated by controlling evaporation temperature. The higher the temperature, the faster the deposition rate. It will also affect the efficiency and life of the device. When the deposition temperature is higher than the glass transition temperature of the material, the organic material will decompose and lead to failure.

For full color OLED display, red, green and blue pixels need to be made. Full color display can be achieved by using three color OLED deposition directly. However, because the traditional lithography process is not suitable for the formation of OLED films, full color display can be achieved indirectly through the following

three schemes: using transverse parallel light-emitting units; using white OLED with color filter film; and using blue OLED with color conversion materials.

9.5.2.2. *Packaging and passivation of OLED devices*

Because most organic materials are very sensitive to moisture and oxygen in the air, and the active interface materials and metal cathodes used in OLED are very active, the water–oxygen isolation treatment of OLED plays an important role in prolonging the life of the device. Ultraviolet curable epoxy resin is usually used as binder to seal OLED. The cover can be made of glass or metal. Moisture and oxygen in the environment may permeate into the device along the interface of the binder. Therefore, the gap between the substrate and the cover should be as small as possible, and the width of epoxy resin should be as long as possible to reduce the water and oxygen permeability. At the same time, CaO and other drying aids are added to the device during packaging, which can absorb residual or infiltrated water.

Oxygen–water isolation can also be achieved by fabricating a passivation layer with a certain thickness on OLED, and the thickness of the device can be greatly reduced. However, because organic materials cannot withstand high temperature, the process temperature of passivation layer must be below 100°C. At the same time, there should be no cracks or pinholes in passivation layer, and its thermal stress should be as small as possible to prevent damage of bottom emission cathode or top emission anode caused by thermal shrinkage. At present, the passivation layer materials studied are polymer film deposition to prepare passivation layer, such as PET, PI and other polymers.

9.5.3. *Advantages and disadvantages of OLED*

9.5.3.1. *Advantages of OLED*

- (1) Ultrathin thickness, only 1/3 of the LCD screen, and lighter weight.

- (2) Solid state shows that it has good seismic performance and is not afraid of falling.
- (3) Large visual angle and good fidelity.
- (4) The response time is too fast, and there is absolutely no shadow in displaying the moving picture.
- (5) It has good low-temperature characteristics and can still display normally at -40°C .
- (6) The manufacturing process is simple and the cost is lower.
- (7) Higher luminous efficiency and lower energy consumption.
- (8) It can be manufactured on different material substrates, and can be made into flexible displays that can bend.

9.5.3.2. *Disadvantages of OLED*

Large-scale mass production technology has not yet been solved, the qualified rate is low, and the cost is high.

9.5.4. *Driving mode of OLED*

9.5.4.1. *Passive driver*

Passive drive (PM-OLED) is divided into static drive circuit and dynamic drive circuit.

(1) Static driving mode

In static-driven organic light-emitting display devices, cathodes of each organic light-emitting pixel are connected together, and the anodes of each pixel are separately drawn, that is, common-negative connection. Static driving circuit is generally used for driving segment display screen.

(2) Dynamic driving mode

In the dynamically driven organic light-emitting display devices, two electrodes of the pixel are made into a matrix structure, that is, the electrodes of the same nature of the horizontal group of display pixels are common, and the electrodes of the same nature of the vertical group of display pixels are common. If the pixels can be divided into N rows and M columns, there are N row electrodes and M column electrodes. Rows and columns

correspond to the two electrodes of the luminous pixel, namely the cathode and the anode. In the actual circuit-driven process, to light up the pixels row by row, the method of row scanning is usually adopted. When scanning row by row, the row electrode is the data electrode. The method is as follows: Pulse is applied to each row of electrodes cyclically, and the driving current pulse of each row of pixels is given by all column electrodes, so as to realize the display of all the pixels in one row. In order to avoid "crossover effect", the pixels in the same row or column are not displayed by reverse voltage. This kind of scan is carried out in a row-by-row sequence, and the time required to scan all rows is called frame period.

The selection time of each line in a frame is equal. Assuming that the number of scanning rows of a frame is N and the time of scanning a frame is 1, the selection time of a row is $1/N$ of the time of a frame, which is called duty cycle coefficient. At the same current, the increase of scanning lines will decrease the duty cycle, resulting in the decrease of the effectiveness of current injection on organic electroluminescent pixels in a frame and the decrease of display quality. Therefore, with the increase of display pixels, in order to ensure the display quality, it is necessary to increase the driving current or use double screen electrode mechanism to improve duty cycle coefficient.

In addition to the cross-effect caused by the common use of electrodes, the mechanism of the composite formation of positive and negative charge carriers in the organic electroluminescent display screen is that any two light-emitting pixels may be crosstalk between the two light-emitting pixels as long as any functional film constituting their structure is directly connected together. That is, one pixel emits light, and another pixel may emit weak light. This phenomenon is mainly caused by the poor thickness uniformity and transverse insulation of organic functional films. From the driving point of view, in order to alleviate this unfavorable crosstalk, the reverse cut-off method is also an effective method.

Gray level of display refers to the brightness level of black-and-white image from black to white. The more gray levels, the richer the image from black to white, and the clearer the details. Gray level is a

very important index for image display and colorization. Generally, the screen with gray level display is mostly dot matrix display, and its driver is mostly dynamic. The methods to realize gray level control are control method, space gray level modulation and time gray level modulation.

9.5.4.2. *Active driver*

Each pixel of active drive (AM OLED) is equipped with low temperature poly Si thin-film transistor (LTP Si TFT) with switching function, and each pixel is equipped with a charge storage capacitor. The whole system of peripheral drive circuit and display array is integrated on the same glass substrate. Active driving is a static driving mode with storage effect and can be driven by 100% load. This kind of driving is not limited by the number of scanning electrodes. It can adjust each pixel independently and selectively. Active driving has no duty cycle problem. Active driving is not limited by the number of scanning electrodes. It is easy to achieve high brightness and high resolution. Active driver is more conducive to OLED colorization because it can independently adjust the gray level of red and blue pixels of brightness. Active matrix drive circuit in the

Table 9.3. Comparison of active and passive displays.

Passive form	Subject-verb form
Transient high density luminescence (Dynamically driven/selective)	Continuous luminescence
Extra-panel IC chips	TFT driver circuit/embedded thin-film driver IC
Line stepwise Scanning	Line stepwise erasing data
Stage control is easy	Formation of organic light-emitting pixels on TFT substrate
Low cost and high driving voltage	Low voltage drive, low power consumption, high cost
Easy to design and easy to manufacture	Long life and complex process of light-emitting module
Matrix driven/OLED	TFT driver/OLED

display screen, easier to achieve integration and miniaturization. In addition, it solves the connection problem between peripheral driving circuit and screen, and improves the yield and reliability.

9.5.5. Challenges for OLED industry

The development of OLED industry faces the following three challenges.

- (1) In technology, the qualified rate of OLED panel is too low. It is known that the qualified rate of large-size OLED panel is less than 70%, while the yield of liquid crystal panel can reach more than 99%.
- (2) OLED industry needs huge investment, which is difficult to accomplish by one or two enterprises alone. OLED is a technology-intensive and capital-intensive industry, as well as a frontier technology in the information industry. It requires relevant departments of the state to plan the OLED industry as a whole from a strategic height, concentrate resources, strengthen independent innovation, guide and encourage enterprises to join the OLED industry chain and provide financial support to form the ability of sustainable development.
- (3) Cost problem, because OLED is current-mode equipment, which needs high current drive, so LTPS TFT substrate is needed. LTPS is bound to keep the cost of OLED high. At present, OLED display screen is mainly concentrated in small and medium size, and OLED pixel distribution is not high, so OLED is more suitable for a slightly larger size, but the cost advantage is worse if the size is too large.

Following is the production line construction of OLED manufacturers in and outside China from 2010 to 2011 based on various information for reference.

Table 9.4 shows the production line construction of OLED manufacturers in and out of China from 2010 to 2011 based on various information for reference.

Table 9.4. Production line construction of OLED manufacturers at home and abroad.

Company	Line construction	Product development	Follow-up plan
LGD	One 730×920 baseboard line, 10 K/M; two 730×460 evaporation line, 20 K/M; construction of an 8th generation line has been started and production capacity is planned for 2011	Mass production of AM has been achieved in 3–15 in, and 19/20.7 in samples have been developed	An 8th generation line has the capacity of mass production
Samsung SMD	One 730×920 baseboard line, 25 K/M; five 730×460 evaporation lines, 50 K/M; one 1300×1500 baseboard line, 20 K/M; nine 750×650 evaporation lines, 70 K/M; and one 5.5 generation AM production line were put into operation in May 2011	Mass production of 2.2–3.8 in smartphone screens; 14, 19, 21, 30, 40 in AM display panel samples have been developed and completed	Since January 2011, we have invested US\$240 million in the construction of OLED project in Tianjin Development Zone in two years. This project will produce OLED screen for mobile phones, and the new plant will be put into use in November 2011. In February 2011, in alliance with Taiwan and Xin, a 5.5th-generation AMOLED panel factory was set up. Hexin's investment will exceed NT\$10 billion. It plans to

(Continued)

Table 9.4. (*Continued*)

Company	Line construction	Product development	Follow-up plan
			produce 50,000 pieces per month in the first phase and 80,000 pieces in the second phase in the fourth quarter of 2011. It plans to start construction of an 8th-generation AMOLED production line in the third quarter of 2011
SONY	Two 600×720 baseboard lines, 40 K/M; one 550×600 baseboard line, 25 K/M	In February 2011, 24.5 in OLED TV and 16.5 inch OLED TV for business were released at present, four OLED TVs have been released. Suspension of 11-minute OLED TV production, but continued to invest in R&D	Unknown
Auo	One 620×750 baseboard line, two 620×375 evaporation lines, 7 K/M	AM products under 7.6 inches have been mass produced and 25 in AM samples have been developed.	Plans to start mass production of small and medium-sized OLED panels for mobile phones and handheld devices

(*Continued*)

Table 9.4. (Continued)

Company	Line construction	Product development	Follow-up plan
Tianma	One 730×920 TFT substrate line and one 365×460 evaporation line. In January 2010, the 4.5 generation AMOLED pilot project was launched	A 3.2 in sample has been developed.	Achieve trial production of 4.5th-generation AMOLED pilot line; plan to build a 5.5th-generation AMOLED production line in Xiamen
Visionox	A 370×470 PMOLED production line, a 200×200 AM panel test line and a 400×500 TFT substrate line	Develop and complete 2.4 in AMOLED samples in cooperation with Youda	Planning and Construction of a 5.5th-Generation AMOLED production line
Rainbow	A 335×550 PMOLED production line; a 4.5th-generation AMOLED production line was constructed in Shunde in October 2010, with a total investment of 4.96 billion yuan	PMOLED or AMOLED samples not yet developed	Construction of Article 2 4.5th-Generation AMOLED production line
Beijing Oriental	Construction of a 5.5th-generation a-Si & LTPS R&D line	Introduce 40 in AM sample	Planning and construction of a 4.5th-generation AMOLED production line will transfer OLED project to Ordos

References

- Li J, Liu N, Wu S, 2013. *Introduction to Flat Panel Display*, Beijing: Electronic Industry Press.
- Li W, 2002. *Organic Light-Emitting Materials, Devices and their Flat-Panel Display — A New Optoelectronic Technology*, Beijing: Science Press.
- Yu J, Tian C, 2015. *Fundamentals and Industrialization of OLED Display*, Chengdu: University of Electronic Science and Technology Press.

This page intentionally left blank

Chapter 10

Materials and Applications of Solar Cells

10.1. Photovoltaic Technology and Solar Cells

10.1.1. *Development of solar cells*

10.1.1.1. *Development of international solar cell technology*

In 1839, Becquerel, France, reported that photovoltaic effects, i.e. photoelectrochemical systems, occur in systems where light electrodes are inserted into electrolytes.

In 1876, Adams found that crystal selenium can produce current under light, that is, solid photovoltaic phenomenon.

In 1884, Charles Fritts made the first 1% selenium battery in the United States.

In 1954, Pearson and Charpin of Bell Laboratory developed the first silicon solar cell with 6% practical value. The New York Times calls this breakthrough “the beginning of a new era that ultimately leads to unlimited sunlight serving human civilization”, the pioneer of modern solar cells.

In 1958, silicon solar cells were first used in space.

In the early 1960s, the design of space batteries tended to be stable. In the 1970s, a large number of applications began in space, and ground applications began. At the end of the 1970s, the production of ground solar cells has greatly exceeded that of space batteries.

When the power, coal, oil and other non-renewable energy sources become more and more urgent, and the energy problem has increasingly become a bottleneck restricting the economic development

of the international community, more and more countries begin to implement the “Sunshine Plan” to develop solar energy resources and seek new impetus for economic development. Some high-level nuclear research institutions in Europe are also turning to renewable energy. Promoted by the huge potential of the International Photovoltaic market, solar cell manufacturers in various countries are scrambling to invest huge amounts of money to expand production in order to gain a place.

The global solar cell industry grew 17-fold from 1994 to 2004, and solar cell production was mainly distributed in Japan, Europe and the United States. In 2006, the global solar cell installation scale has reached 1744 MW, an increase of 19% over 2005, and the total market output value has officially exceeded the \$10 billion mark. Global solar cell production reached 3436 MW in 2007, an increase of 56% over 2006.

10.1.1.2. *Domestic solar cell technology development*

In 1959, the first solar cell with practical value was born; in March 1971, the solar cell was first applied to the second satellite of our country, “Practice No. 1”; in 1973, the solar cell was first applied to buoy lamps; in 1979, the production of single crystal silicon cells was started by using waste single crystal and semiconductor device technology of semiconductor industry. In the mid and late 1980s, the importation of foreign key equipment or complete production lines led to the initial formation of China’s solar cell manufacturing industry.

At present, China has become a major solar cell producer in the world. In 2007, the national solar cell production reached 1188 MW, an increase of 293% year on year. China has successfully surpassed Europe and Japan to become the world’s largest producer of solar cells. In terms of industrial layout, China’s solar cell industry has formed a certain agglomeration trend. In the Yangtze River Delta, the Bohai Sea Rim, the Pearl River Delta and the central and Western regions, solar energy industry clusters with different characteristics have been formed.

Solar photovoltaic power generation will occupy an important position in the world energy consumption in the near future. It will not only replace some conventional energy sources, but also become the main body of world energy supply. It is expected that by 2030, solar renewable energy will account for more than 30% of the total green energy structure, and solar photovoltaic power will account for more than 10% of the world's total power supply; by 2040, renewable energy will account for more than 50% of the total energy consumption, and solar photovoltaic power will account for more than 20% of the total power. By the end of the 21st century, renewable energy will account for more than 80% of the energy structure and solar power will account for more than 60%. These figures are enough to show the development prospects of solar photovoltaic industry and its important strategic position in the field of energy. From this, we can see that the solar cell market has broad prospects.

In the effective use of solar energy, solar photovoltaic utilization is the fastest growing and most dynamic research field in recent years, and is one of the most attractive projects.

10.1.2. *Principle of solar cell*

10.1.2.1. *Electron and hole transport*

Conductivity of semiconductor materials is realized by the directional movement of two carriers, electrons and holes. At low temperatures, valence electrons are completely bound around the nucleus and cannot move in crystals. At this time, the valence band in the band diagram is full and the conduction band is empty. With the increase of temperature and other reasons, some electrons break away from the bond of the nucleus, resulting in valence electron commonalization and become free electrons, which can move in the whole crystal lattice, leaving a vacancy in the original position of electrons, known as holes. When valence electrons become free electrons, they can move irregularly as negative charges (e) in crystals. At this time, from the energy band point of view, the electron absorbs energy and transits from valence band to conduction

band. Under the action of external electric field, besides thermal motion, the electron drifts in the opposite direction to the electric field and generates electric current, which has the same direction as the electric field. This conductive mechanism of free electron carrying current is called electronic conduction, and electrons are called carriers.

Before electrons become free electrons, atoms are clearly neutral. After electrons become free electrons moving in the whole crystal, the original position of electrons lacks a negative charge, showing a positive charge, called a hole. At this time, from the energy band point of view, because of the transition of electrons to the conduction band, holes are left in the valence band. If the adjacent electrons enter this position, the position of the electrons will be empty, showing positive charge. If the process continues, it will be as if the holes have moved. Under the action of external electric field, besides thermal motion, the hole drifts along the direction of electric field to generate current, which is opposite to the direction of electric field. This kind of conductive mechanism of hole carrying current is called hole conduction.

At a certain temperature, due to the absorption of thermal vibration energy, the electron-hole pairs in semiconductor materials are constantly generated. At the same time, when the electrons and holes meet, the recombination occurs, that is, the electrons in the conduction band transit to the valence band and recombine with the holes in the valence band, resulting in the disappearance of the electron-hole pairs. Obviously, for pure semiconductors, the concentration of electron-hole pairs depends mainly on temperature in the thermal equilibrium state if impurities are not consciously doped. The higher the temperature, the higher the concentration of electron-hole pairs. Such semiconductor materials become intrinsic semiconductor materials with electron and hole concentration is

$$n = p = n_i(T), \quad (10.1)$$

where, T is the thermodynamic temperature and n_i is the intrinsic carrier concentration. At room temperature of 300 K, the intrinsic carrier concentration of silicon material is $1.5 \times 10^{10} \text{ cm}^{-3}$.

Under the action of external electric field, electrons and holes move. Due to the periodic repetitive arrangement of atoms in crystals, their motion state is different from that in free space. Therefore, effective mass is used instead of mass to characterize the difference. Let the effective mass of electrons and holes be m_n and m_p respectively, and their accelerations in the external electric field are

$$a_n = -qE/m_n, \quad (10.2)$$

$$a_p = -qE/m_p, \quad (10.3)$$

where, q is the elementary charge quantity.

10.1.2.2. *Doping characteristics of semiconductors*

Taking silicon and germanium semiconductors as examples, if impurities are acceptor impurities, such as boron or indium, one of the third group elements in the periodic table, their valence bands have only three electrons, and the minimum energy level of their conduction bands is lower than that of the fourth group elements. Therefore, electrons can easily transit from the valence band of germanium or silicon to the conduction band of boron or indium. In this process, a positive ion is generated by the loss of electrons, which is a “vacancy” for other electrons, so it is usually called a “hole”, and this material is called a “P”-type semiconductor. Conduction in such materials is mainly caused by positively charged holes, so electrons are “minority carriers” in this case. Figure 10.1 is P-type-doped Si.

If the impurities are donor impurities such as arsenic or antimony, the valence bands of these elements have five electrons. However, the maximum energy level of valence electrons of impurity elements is larger than that of germanium (or silicon), so it is easy for electrons to enter the conduction band of group IV elements from this energy level. These materials become semiconductors. Because conductivity is caused by excess negative ions, it is called “N”-type. The conductivity of some materials is due to the excess positive ions in the materials, but mainly due to the large number of electrons, so electrons (in N-type materials) are called “majority carriers”. Figure 10.2 is N-type-doped Si.

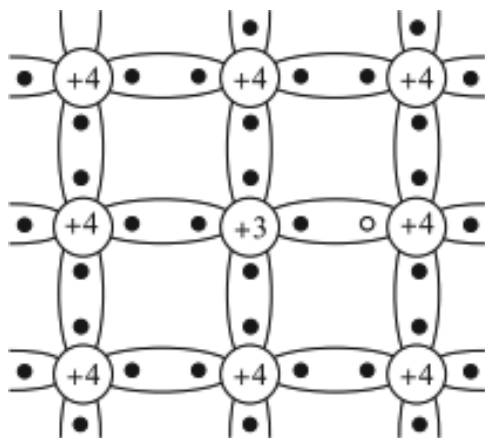


Fig. 10.1. P-type-doped Si.

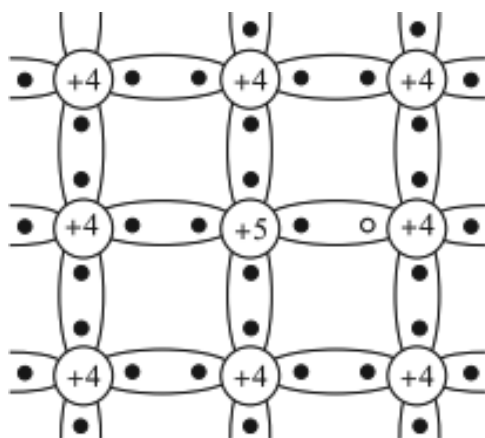


Fig. 10.2. N-type doped Si.

For general conductive materials, the conductivity σ can be expressed by the following formula:

$$\sigma = ne\mu, \quad (10.4)$$

where, n is the carrier concentration, cm^{-3} ; e is the charge of electrons; μ is the mobility of carriers (the velocity of the downloaded carriers per unit electric field intensity), $\text{cm}^2/(\text{V s})$. Carriers are

electrons here. For semiconductor materials, because both electrons and holes conduct electricity at the same time, there are two carriers, so the formula can be changed to

$$\sigma = ne\mu_e + pe\mu_p, \quad (10.5)$$

where, n is the electron concentration, p is the hole concentration, e is the electron charge, and μ_e and μ_p are the electron and hole mobility, respectively. If the electron concentration is much higher than the hole concentration, the conductivity of the material is $\sigma \approx ne\mu_e$, and if the hole concentration is much higher than the electron concentration, the conductivity of the material is $\sigma \approx pe\mu_p$.

10.1.2.3. Working principle and structure of solar cells

When P-type and N-type semiconductors are combined, a special thin layer will be formed in the interface area of the two semiconductors. The P-type side of the interface is negatively charged and the N-type side is positively charged. This is due to the concentration difference between P-type semiconductor and N-type semiconductor due to their multi-hole and multi-free electrons. Electrons in the N region will diffuse to the P region, and holes in the P region will diffuse to the N region. Once diffused, an “internal electric field” from N to P will be formed, thus preventing the diffusion. When the balance is reached, a special thin layer is formed, which forms a potential difference, thus forming a PN junction, as shown in Fig. 10.3.

When the wafer is illuminated, the holes in the N-type semiconductor move to the P-type region, while the electrons in the P-type region move to the N-type region, thus forming the current from the N-type region to the P-type region. Then a potential difference is formed in the PN junction, which forms the power supply. Figures 10.4 and 10.5 are schematic diagrams of PN junction solar cells and working principles of photovoltaic cells, respectively.

Since semiconductors are not good conductors of electricity, if electrons flow through PN junctions in semiconductors, the resistance is very large and the loss is very large. But if all the metal is coated on the upper layer, the sunlight cannot pass and the current cannot be

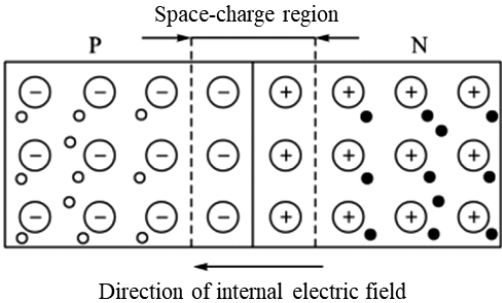


Fig. 10.3. Diagram of built-in electric field in PN junction.

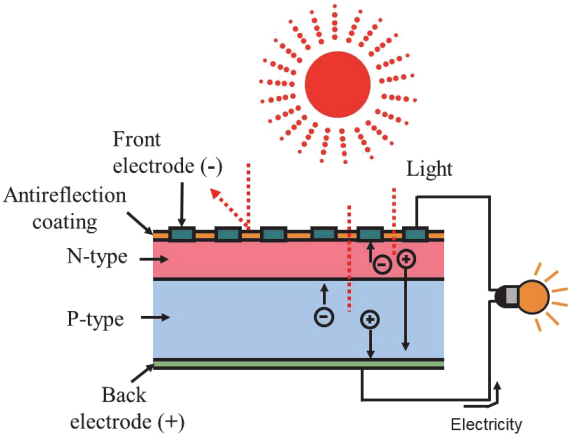


Fig. 10.4. PN junction solar cell schematic diagram.

generated. Therefore, the PN junction is usually covered with metal mesh to increase the area of incident light.

In addition, the surface of silicon is very bright, which reflects a lot of sunlight and cannot be used by batteries. For this reason, scientists have coated it with a protective film with very small reflection coefficient. In actual industrial production, a layer of silicon nitride film is deposited by chemical vapor deposition. Its thickness is about 1000 Å, which reduces the reflection loss to 5%, or even less.

This electric field is equivalent to a diode, allowing (or even pushing) electrons to flow from the P side to the N side, rather than

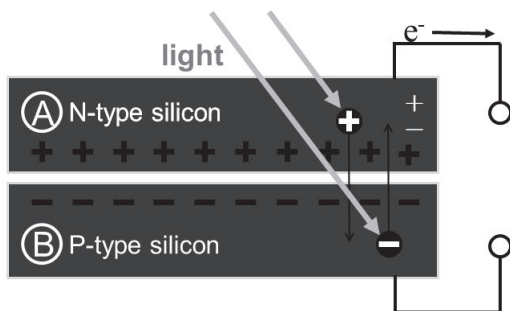


Fig. 10.5. Operating principle of photovoltaic cells.

vice versa. It's like a mountain — electrons can easily slide down the hill (to the N side), but can't climb up (to the P side). In this way, an electric field equivalent to a diode is obtained, in which the electrons can only move in one direction. When light strikes a solar cell in the form of photons, its energy releases electron-hole pairs. Each photon carrying enough energy usually releases just one electron, creating a free hole. If this happens close enough to the electric field, or if the free electrons and free holes are within its influence, the electric field will send electrons to the N side and holes to the P side. This will lead to further damage to electrical neutrality. If an external current path is provided, electrons will pass through it and flow to their original side (P side), where they merge with the holes sent by the electric field and work during the flow. Electron flow provides current and the electric field of the battery generates voltage. With current and voltage, power is the product of both.

Monocrystalline silicon is not the only material used in photovoltaic cells. Polycrystalline silicon is also used in the battery material. Although the efficiency of the battery produced by this method is not as high as that of single crystal silicon battery, it can reduce the cost. In addition, amorphous silicon without crystal structure is also used to reduce costs. Other materials used include gallium arsenide, indium copper selenide and cadmium telluride. Because different materials have different bandgaps, they seem to “tune” photons with different wavelengths or different energies. One way to improve efficiency is to use two or more layers of different

materials with different bandgaps. The material with higher bandgap is placed on the surface to absorb higher energy photons, while the material with lower bandgap is placed below to absorb lower energy photons. This technology can greatly improve efficiency. Such batteries are called multifaceted batteries, which can have multiple electric fields.

10.1.3. *Classification of solar cells*

1. According to the maturity of technology

- (i) Crystalline silicon batteries: monocrystalline silicon and polycrystalline silicon.
- (ii) Thin-film batteries: a-Si, CIGS, CdTe, spherical batteries, polycrystalline silicon films, organic batteries.
- (iii) New concept batteries (third generation batteries): quantum dots, quantum well batteries, laminated (bandgap transition) batteries, intermediate band batteries, impurity band batteries, up-down converter batteries, a-Si/C-Si heterojunctions (increasing infrared absorption), dipole antenna batteries, hot carrier batteries.

2. According to material

- (i) Silicon-based batteries: monocrystalline silicon, polycrystalline silicon, microcrystalline (nanocrystalline), amorphous silicon.
- (ii) Compound semiconductor batteries: CdTe, CIGS, GaAs, InP, organic batteries, photochemical batteries.
- (iii) Band range: It is divided into solar photovoltaic cells and thermal photovoltaic cells.

10.2. Crystalline Silicon Solar Cells

10.2.1. *Polycrystalline silicon thin-film solar cells*

Polycrystalline silicon has a gray metallic luster with a density of $2.32\text{--}2.34\text{ g/cm}^3$. The melting point is 1410°C and the boiling point is 2355°C . It is soluble in the mixed acid of hydrofluoric acid and nitric acid, insoluble in water, nitric acid and hydrochloric acid.

Its hardness is between germanium and quartz. It is brittle at room temperature and fragile at cutting time. It has ductility when heated above 800°C and obvious deformation at 1300°C . It is not active at room temperature and reacts with oxygen, nitrogen and sulphur at high-temperature. It has great chemical activity and can interact with almost any material in high temperature melting state. It has semiconductor properties and is a very important excellent semiconductor material, but a small amount of impurities can greatly affect its conductivity. Polycrystalline silicon is widely used in the electronic industry to manufacture semiconductor radio, recorder, refrigerator, color TV, video recorder, computer, etc. It is obtained by chlorination of dry silicon powder and dry hydrogen chloride gas under certain conditions, then condensation, distillation and reduction.

According to the different raw materials used in the deposition reaction of silicon, the methods of mass production of polysilicon in the world are divided into the new thermal decomposition of silane as raw material and the improved Siemens method using trichlorosilane as raw material. The former can produce both granular polysilicon and rod polysilicon, and the latter can produce rod polysilicon. The thermal decomposition of new silane for producing rod polysilicon and granular polysilicon differs greatly in the preparation of silane, reaction equipment and process, which can be regarded as two different methods. Modified Siemens process is the mainstream of polysilicon production, accounting for more than 75% of the total output.

The process of producing battery sheet is very complicated. It usually goes through the process as shown in Fig. 10.6.

- (1) Slice: The silicon rod is cut into square silicon wafers by multi-wire cutting.
- (2) Silicon wafer cleaning: The surface of silicon wafer is cleaned by conventional cleaning method, and then the cutting damage layer is removed by acid (or alkali) solution for $30\text{--}50\text{ }\mu\text{m}$.
- (3) Suede preparation: Anisotropic etching of silicon wafer with alkali solution, and suede preparation on the surface of silicon wafer.

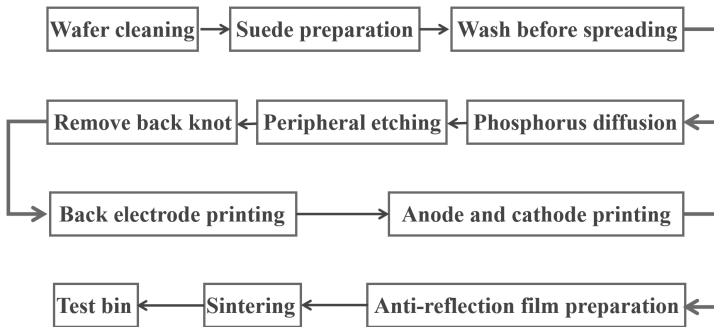


Fig. 10.6. Process flow chart of crystalline silicon solar cells.

- (4) Phosphorus diffusion: PN+ junctions are formed by diffusion from coating source (liquid source or Solid Nitriding scale source), and the junction depth is generally $0.3\text{--}0.5\ \mu\text{m}$.
- (5) Peripheral etching: The diffusion layer formed on the surface of the silicon wafer during diffusion will shorten the upper and lower electrodes of the battery and remove the diffusion layer by shielding wet etching or plasma dry etching.
- (6) Removal of back PN+ junction: Wet etching or abrasive disk method is commonly used to remove back PN+ junction.
- (7) Preparation of upper and lower electrodes: Vacuum evaporation, electroless nickel plating or printing and sintering of aluminium slurry. First, the lower electrode is made, and then the upper electrode is made. Aluminum slurry printing is a widely used process.
- (8) Antireflective coating preparation: In order to reduce the input reflection loss, a layer of antireflective coating is coated on the surface of silicon wafer. There are MgF_2 , SiO_2 , Al_2O_3 , Si_3N_4 , TiO_2 , Ta_2O_5 and so on. Vacuum plating, ion plating, sputtering, printing, PECVD or spraying can be used in the process.
- (9) Sintering: Sintering battery chips on nickel or copper substrates.
- (10) Test classification: According to the specified parameters and specifications, test classification.

10.2.2. Amorphous silicon thin-film solar cells

Two key issues in developing solar cells are to improve conversion efficiency and reduce costs. As amorphous silicon thin-film solar cells have low cost and are easy to produce on a large scale, they are widely valued and developed rapidly. In fact, in early 1970s, Carlson and others have begun the research and development of amorphous silicon batteries. In recent years, the research and development of amorphous silicon batteries has developed rapidly. At present, many companies in the world have produced this kind of battery products.

Although amorphous silicon is a good solar cell material, its optical bandgap is 1.7 eV, and the material itself is insensitive to the long-wave region of solar radiation spectrum, which limits the conversion efficiency of amorphous silicon solar cells. In addition, the photoelectric efficiency decreases with the duration of illumination time, i.e. the so-called photoinduced fading (S W) effect, which makes the performance of the battery unstable. The way to solve these problems is to prepare laminated solar cells. The laminated solar cells are made by depositing one or more p-i-n sub-cells on the p, i, n-layer single-junction solar cells. The key problems for stacked solar cells to improve conversion efficiency and solve the instability of single junction cells are as follows: (1) It combines materials with different bandgaps to improve the spectral response range. (2) The I layer of the top battery is thin, and the electric field intensity produced by the illumination does not change much, which ensures the extraction of the photogenerated carriers in the I layer. (3) The carrier produced by the bottom cell is about 1/2 of that of the single cell, and the photoinduced degradation effect decreases. (4) The sub-batteries of the stacked solar cells are connected in series.

Amorphous silicon thin-film solar cells have many preparation methods, including reactive sputtering, PECVD, LPCVD (low pressure chemical vapor deposition) and so on. The raw material gas is H_2 diluted SiH_4 , and the substrates are mainly glass and stainless steel. The amorphous silicon thin-films can be made into single-junction and multi-layer solar cells through different cell processes. At present, two major advances have been made in the research

of amorphous silicon solar cells: first, the conversion efficiency of three-layer amorphous silicon solar cells reaches 13%, which sets a new record; second, the annual production capacity of three-layer amorphous silicon solar cells reaches 5 MW. The maximum conversion efficiency of single-junction solar cells manufactured by United Solar Energy Corporation (VSSC) is 9.3%, and that of three-bandgap three-layer solar cells is 13%.

The maximum conversion efficiency is achieved on small area (0.25 cm^2) batteries. It has been reported that the conversion efficiency of single crystalline amorphous silicon solar cells exceeds 12.5%. A series of new measures have been adopted by the Japanese Academy of Central Research. The conversion efficiency of the amorphous silicon solar cells is 13.2%. There is little research on amorphous silicon thin-film solar cells, especially the laminated solar cells in China. Geng Xinhua of Nankai University and others used industrial materials to fabricate a-Si/a-Si laminated solar cells with an area of $20\text{ cm} \times 20\text{ cm}$ and a conversion efficiency of 8.28% using aluminium back electrode.

Amorphous silicon solar cells have great potential due to their high conversion efficiency, low cost and light weight. But at the same time, because of its low stability, it directly affects its practical application. If stability and conversion rate can be further solved, amorphous silicon solar cells are undoubtedly one of the main development products of solar cells.

10.3. Compound Semiconductor Thin-Film Batteries

10.3.1. *CdTe solar cells*

CdTe is a class II–VI compound, $E_g = 1.5\text{ eV}$, theoretical efficiency 28%, stable performance, and has been valued by photovoltaic industry. Technologies include near space sublimation (CSS), electrodeposition, sputtering, vacuum evaporation and screen printing; laboratory battery efficiency is 16.4%; commercial battery efficiency is 8–10%; CdTe battery has achieved large-scale production in the early 1990s, with a market share of 0.3% in 2002.

The first CdTe solar cell was made by plating in alloy on CdTe single crystal in RCA laboratory. Its photoelectric conversion efficiency was 2.1%. In 1963, the first heterojunction CdTe thin-film cell was born. The research on CdTe solar cell in China began in the early 1980s. Inner Mongolia University adopts evaporation technology and Beijing Solar Energy Research Institute adopts electrodeposition technology. The efficiency reached 5.8% in 1983. In the late 1990s, Sichuan University adopted near-space sublimation technology. During the 10th Five-Year Plan period, it was included in the national “863” key project and required to establish a 0.5 MW/year pilot production line.

10.3.1.1. *CdTe solar cell structure*

Figure 10.7 is the structure of CdTe solar cells and Fig. 10.8 is the energy band diagram of CdTe thin-film solar cells.

The cladding structure is composed of transparent oxide (TCO), CdS and CdTe thin-films on glass substrates in turn, and sunlight is irradiated from below the glass substrates, first through the TCO layer and then into the CdS/CdTe junction. For the underlying structure, CdTe thin-films are first grown on the appropriate substrates, then CdS and TCO thin-films. Among them, superstrategy is the most efficient.

Substrate: Mainly supports batteries, prevents pollution and incident sunlight.

TCO layer: transparent conductive oxide layer. Its main function is to transmit light and conduct electricity.

CdS layer: N-type semiconductor, PN junction with P-type CdTe. The absorption edge of CdS is about 521 nm, and almost all visible light can be transmitted.

CdTe layer: The main absorber layer of the battery. The PN junction formed by it and the N-type CdS window layer is the core part of the battery. Polycrystalline CdTe thin-films have ideal bandgap ($E_g = 1.45 \text{ eV}$) and high optical absorption (about 10^4 cm^{-1}) for solar cells. The spectral response of CdTe is almost the same as that of the sun.

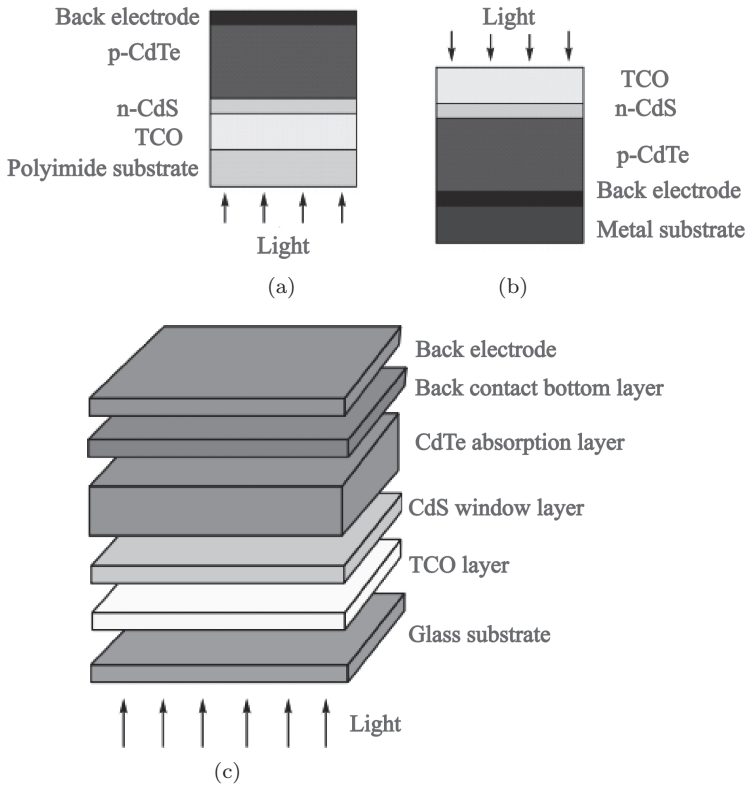


Fig. 10.7. CdTe solar cell structure: (a) Cladding structure; (b) bottom structure; (c) CdTe solar cell structure.

Back contact layer and back electrode: Reduce the contact barrier between CdTe and metal electrodes, induce current, and make metal electrodes form ohmic contact with CdTe.

10.3.1.2. Fabrication technology of CdTe solar cells

CdTe thin-film preparation: The thin-films prepared by low-temperature deposition technology such as electrodeposition and chemical bath deposition are compact and have fine grains. After treatment, the grains grow. High-temperature deposition technologies such as screen printing, near-space sublimation and elemental vapor-phase combination require post-treatment in chlorinated compounds

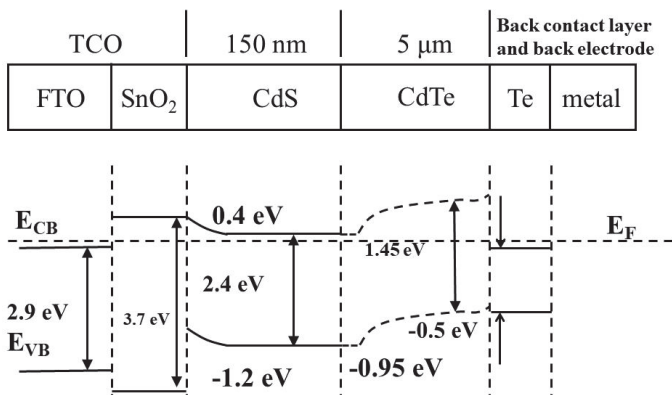
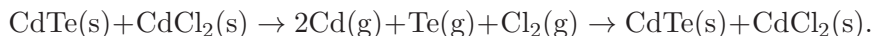


Fig. 10.8. Energy band diagram of CdTe thin-film solar cells.

and oxygen atmosphere to prepare high conversion efficiency batteries with grain size of 2–3 μm or more. The possible reason is that chlorine not only promotes grain growth, but also passivates grain boundary defects as acceptor impurity in CdTe. Various preparation techniques of cadmium telluride films are shown in Fig. 10.9.

CdCl₂ treatment of CdTe absorption layer: Almost all CdTe films obtained by deposition technology must be treated by CdCl₂ again. CdCl₂ treatment can further improve the conversion efficiency of CdTe/CdS heterojunction solar cells. The reasons are as follows: (1) It can form an interface layer between CdTe and CdS to reduce the concentration of interface defect states. (2) It can lead to the recrystallization of CdTe films and the growth of grains, and reduce the defect of crystal boundary. (3) Heat treatment can passivate defects and improve carrier lifetime of absorption layer. When the CdTe film is placed in a CdCl₂ environment at about 400°C, the following reactions will occur:



Therefore, the presence of CdCl₂ promotes the recrystallization of CdTe by means of regional gas phase transport. Not only did the smaller grains disappear, but the interface structure between CdTe and CdS was also ordered.

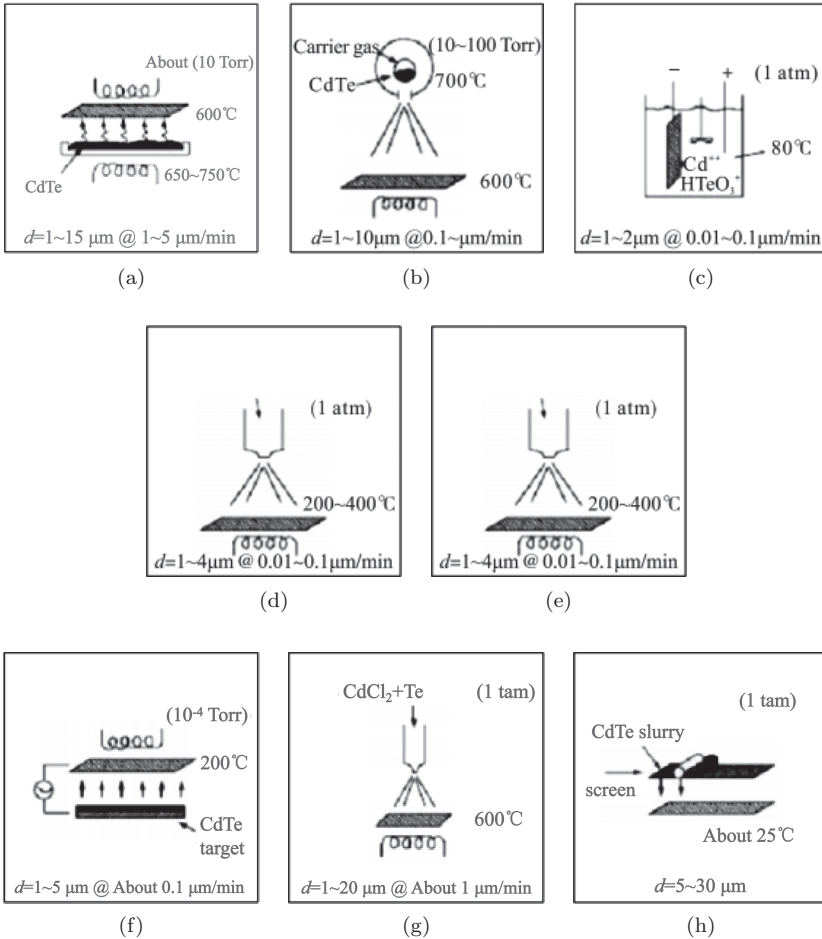


Fig. 10.9. Preparation technology of cadmium telluride films: (a) near-space sublimation; (b) vapor transport deposition; (c) electrodeposition; (d) metal-organic chemical vapor deposition; (e) physical vapor deposition; (f) sputtering method; (g) spray deposition method; and (h) screen printing.

Back contact layer preparation: CdTe has a high work function (about 5.5 eV), and it is difficult to form ohmic contact with most metals. A feasible method is to chemically etch the surface of CdTe thin-films and deposit highly doped back contact materials, which can significantly improve the quantum conversion efficiency. Table 10.1 shows the CdTe contact layer and its conversion efficiency.

Table 10.1. CdTe contact layer and its conversion efficiency.

Structure	Back contact layer thickness	V_{∞} (mV)	I_{∞} (mA/cm ²)	FF (%)	η (%)
CdS/CdTe	—	706	20.8	56.4	8.3
CdS/CdTe/ ZnTe:Cu	50	768	22.1	61.4	10.4
CdS/CdTe/ ZnTe/ZnTe:Cu	50/50	784	22.8	64.7	11.6

10.3.1.3. *Advantages and disadvantages of CdTe solar cells*

Low cost. The cost of CdTe thin-film solar cells is far superior to that of crystalline silicon and other materials in industrial scale. The cost of production is only US\$0.87/W.

High absorbance. It is most consistent with the spectrum of the sun and absorbs more than 95% of the sunlight.

The process is relatively simple, standard process, low energy consumption after the end of the life cycle, recyclable, strong and weak light can generate electricity, the higher the temperature, the better the performance.

Tellurium is scarce. There is no guarantee of the increasing demand for CdTe solar cells.

Cadmium is toxic as a heavy metal. Once the CdTe solar cells are discharged and polluted in the process of production and use, it will affect the environment.

10.3.2. *CIGS solar cells*

CIGS thin-film solar cells are the best proportion of chalcopyrite crystalline thin-film solar cells composed of four elements: Cu, In, Ga and Se. They are the key technologies to make up the solar panels. CIGS solar cell is the future star of solar cell products, which can compete with traditional crystalline silicon solar cell because of its strong light absorption capacity, good power generation stability, high conversion efficiency, long daytime power generation time, high



Fig. 10.10. CIGS thin-film solar panels.

power generation capacity, low production cost and short energy recovery cycle. CIGS thin-film solar panels are shown in Fig. 10.10.

The laboratory technology of CIGS thin-film solar cell materials and devices is becoming mature in developed countries. The development of large area battery components and mass production is the general trend of CIGS battery development. Flexible and cadmium-free batteries are the research hotspots in recent years. The National Renewable Energy Labs (NREL) fabricated 19.9% cells on glass substrates by a three-step co-evaporation process. This flexible substrate CIGS solar cell has a very promising military application prospect. Recently, the efficiency of CIGS small area batteries has set a new record, reaching 20.1%, which is almost the same as that of polycrystalline silicon batteries. The efficiency of CIGS batteries prepared on stainless steel substrates by NREL and Panasonic Electric Co. of Japan is over 17.5%. After years of efforts, a team led by Tiwari, a scientist at the Swiss Federal Laboratory of Materials Science and Technology (EMPA), has improved the previously developed flexible stainless steel substrates solar cells and achieved 18.7% efficiency. The efficiency of Cd-free CIGS batteries developed by the Department of Energy National Photovoltaic Center and Japan's New Energy and Industrial Technology Development Agency

(NEDO) reached 18.6%. This shows that CIGS solar cells with high conversion efficiency can be prepared even without CdS.

Research on CIGS thin-film solar cells in China began in the 1980s. CIS materials and batteries were studied in Inner Mongolia University, Yunnan Normal University and Nankai University. The absorption layer CIS thin-films, N-type CdS thin-films and window low resistivity CdS:In thin-films were prepared by evaporation method in Nankai University. In 1999, CIS cell efficiency was 8.83% and CIGS cell efficiency was 9.13%. In 1999, supported by the “211” project of the Ministry of Education, CIGS thin-films were prepared by selenization of metal pre-layers, CdS thin-films were prepared by chemical water bath (CBD), and intrinsic ZnO, ZnO: Al thin-films were sputtered. In 2002, the research platform of CIGS thin-film solar cells 10 cm × 10 cm area module was established, which laid a foundation for the development of CIGS thin-film solar cells and compound electronic thin-films and devices in China. In early June 2011, the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, in cooperation with the Chinese University of Hong Kong, successfully developed CIGS thin-film solar cells with photoelectric conversion efficiency of 17%, which is the highest conversion efficiency of CIGS solar cells reported in China.

With the increasing shortage of raw materials and rising prices of crystalline silicon solar cells, many companies have invested heavily, and the CIGS industry is showing a vigorous development trend. In January 2012, Manz’s acquisition of the whole innovative production line of WU RTH Solar CIGS accelerated its technological development, creating photovoltaic modules with 14% actual photovoltaic module efficiency (15.1% area efficiency) and setting a world record in this field. In 2016, Manz and the German Research Center for Solar and Hydrogen Energy (ZSW) achieved 22.6% CIGS thin-film solar cell efficiency. A new record of thin-film photovoltaic cells has been created. Among all thin-film technologies, CIGS is the most potential technology to further improve efficiency and reduce costs. Because of its excellent performance, CIGS is called the next generation of low-cost solar cells in the world. It has broad market prospects in both ground solar power generation and space micro-satellite

power supply applications. China's CIGS industry lags far behind countries and regions such as Europe, America and Japan. Nankai University, relying on the National "10th Five-Year Plan" 863, has built a 0.3 MW pilot line, and has prepared a sample of integrated components with a efficiency of $30\text{ cm} \times 30\text{ cm}$ and a efficiency of 7%. In February 2008, Shandong Furi Photovoltaic Technology Co., Ltd. cooperated with Johanna of Germany to introduce China's first CIGSSe (copper, indium, gallium, sulfur and selenium compounds) commercial production line. On September 15, 2011, the largest domestic copper, indium, gallium and selenium (3 MW CIGS) thin-film solar photovoltaic roof power station in Gaomi City, Shandong Province, was launched. The first phase of the 300 kW project has been successfully connected with the State Grid system. The photovoltaic roof power plant is the largest demonstration area in China using CIGS thin-film solar cell modules. Following the solar module heat and polycrystalline silicon heat, thin-film battery has become a new investment hot spot in the domestic photovoltaic field. Domestic investment enthusiasm for thin-film solar cells continues to rise, and thin-film solar cell projects are blooming everywhere. In the next few years, the sales of CIGS thin-film solar cells will accelerate. It is expected that the global sales of CIGS thin-film solar cells will grow to \$2.1 billion by 2020. Revenue from CIGS thin-film photovoltaic modules will reach \$4.4 billion in 2017.

10.3.2.1. *Structure and material properties of CIGS thin-film solar cells*

The structure of CIGS thin-film solar cells is shown in Fig. 10.11.

The chemical composition of CIGS thin-film material is $\text{CuIn}_x\text{Ga}_{1-x}\text{Se}_2$. Its crystal structure is chalcopyrite. It is a mixed crystal semiconductor of CuInSe_2 and CuGaSe_2 . The crystal structure of CIGS is shown in Fig. 10.12.

CIGS is characterized by direct bandgap semiconductor with optical absorption coefficient of 10^5 cm^{-1} , which is conducive to film formation; bandgap of 1.04–1.68 eV; radiation resistance and long life; high-quality crystals can be deposited on soda-lime glass,

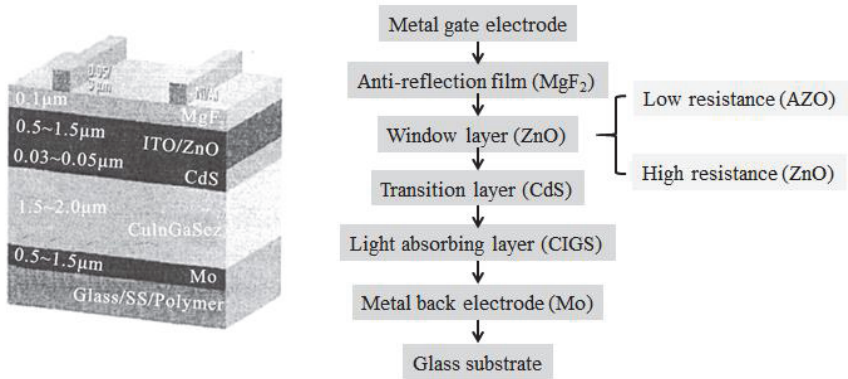


Fig. 10.11. CIGS thin-film solar cell structure.

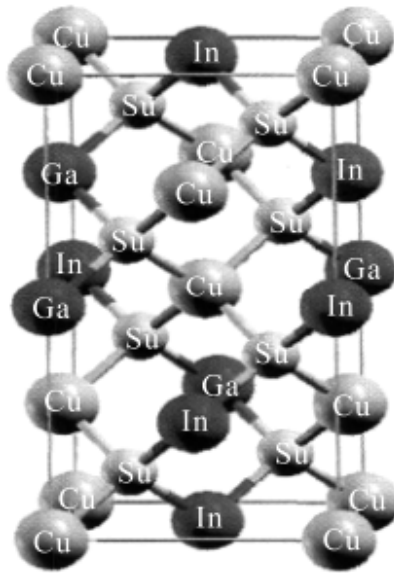


Fig. 10.12. Crystal structure of CIGS.

and flexible deposits such as stainless steel foil and polyimide. The preparation method is as follows.

(i) Sputtering–Selenization

Figure 10.13 is a flowchart of sputtering-selenium process.

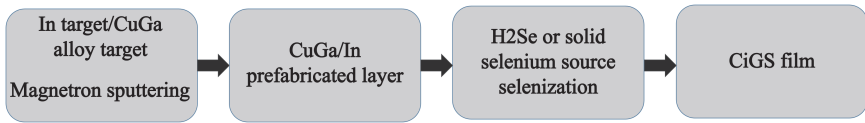


Fig. 10.13. Flowchart of sputtering selenium process.

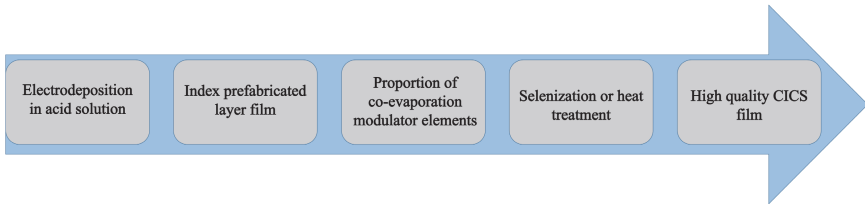


Fig. 10.14. Flowchart of electrodeposition process.

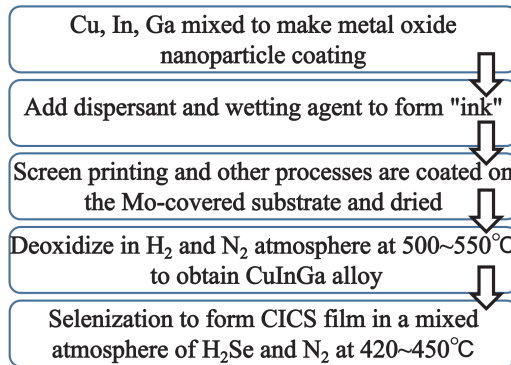


Fig. 10.15. Flowchart of nanoparticle coating process.

(ii) Electrodeposition

Figure 10.14 is a flowchart of electrodeposition process.

(iii) Nanoparticle coating method

Figure 10.15 is the flow chart of nanoparticle coating process.

10.3.2.2. *Advantages of CIGS Technology*

- (i) Strong light absorption capacity: CIGS solar cells consist of chalcopyrite crystals with the best proportion of copper, In,

Ga and Se as absorption layer. The absorption spectrum has a wide range of wavelength. Besides the visible spectrum range of absorbable light of crystalline and amorphous silicon solar cells, CIGS solar cells can also cover the infrared region with wavelengths ranging from 700 to 1200 nm. That is, the time of absorbable photovoltaic power generation is the longest in a day. Compared with crystalline silicon solar cells of the same wattage level, CIGS thin-film solar cells can generate more than 20% of the total electricity per day.

- (ii) High power generation stability: Because of the photoinduced attenuation of crystalline silicon batteries, their power generation efficiency will gradually decline after a long period of sunshine exposure, while CIGS solar cells have no photoinduced attenuation characteristics and have high power generation stability. After a long period of power generation, crystalline silicon solar cells have more or less hot spot phenomenon, which leads to less power generation and higher maintenance costs. CIGS solar cells can adopt internal connection structure to avoid this phenomenon, which is lower than crystalline silicon solar cells.
- (iii) Conversion efficiency: According to the National Renewable Energy Laboratory of the United States, the highest conversion efficiency of solar cells can reach 20.2%, while the industry record can reach 17%, and the general standard is 12%.
- (iv) Low production cost: The main cost of CIGS solar cells is raw materials consisting of glass substrates and four elements: Cu, In, Ga and Se. Among them, only sodium glass used in general building materials is used for glass, and no special ultra-white glass or thin-film conductive glass for solar energy is needed. The four metal elements are not precious metals, and the thickness of CIGS absorbing layer for each panel is not more than 3 microns. The demand for raw materials is not high, and the cost of each panel is very competitive.
- (v) Short cycle of energy recovery: According to the research of the U.S. Department of Energy, the recovery period of crystalline silicon solar cells is 2–4 years, and that of thin-film solar cells

is 1–2 years. In other words, every solar power generation system can enjoy a truly pollution-free use cycle of 26–29 years, and the use of CIGS solar energy is undoubtedly the best choice.

10.3.3. *GaAs solar cells*

GaAs is an important semiconductor material. Because its electron mobility is 5–6 times larger than that of silicon, it has been widely used in the fabrication of microwave devices and high-speed digital circuits. Semiconductor devices made of GaAs have the advantages of high frequency, high temperature, good low-temperature performance, low noise and strong radiation resistance. Gallium arsenide is one of the semiconductor materials, which has many advantages. However, the crystal triode fabricated with gallium arsenide has a small amplification factor and poor thermal conductivity, so it is not suitable for high-power devices. Although gallium arsenide has superior properties, because it decomposes at high temperature, it is technically demanding to grow high purity single crystal materials with ideal chemical ratio.

GaAs has a wider bandgap than silicon, which makes its spectral responsiveness and space solar spectral matching ability better than silicon. At present, the theoretical efficiency of silicon battery is about 23%, the theoretical efficiency of single-junction gallium arsenide battery is 27%, and that of multi-junction gallium arsenide battery is more than 50%.

Structural design and epitaxial material growth of multi-junction solar cells are very important links in battery preparation. Domestic research institutes in this field mainly include Xi'an Institute of Optical Precision Machinery, Chinese Academy of Sciences, Sichuan University, Shanghai Jiao Tong University and so on. At present, Wuhan National Photoelectric Laboratory is also conducting related research. MOCVD (metal organic vapor deposition) technology is mainly used for the growth of epitaxy materials. GaInP₂/GaAs/Ge cascade is mostly used for multi-junction battery structure.

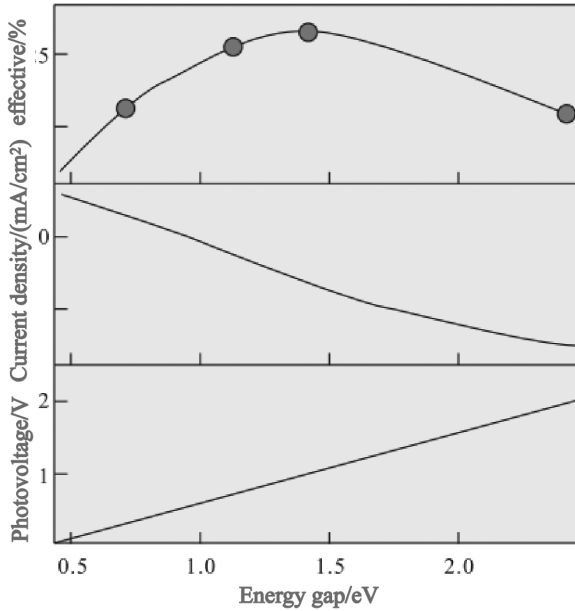


Fig. 10.16. Performance parameters of GaAs solar cells.

10.3.3.1. Structure of GaAs solar cells

(i) Single junction design

If the current generated is to be maximized, solar cells should be able to capture photons in the solar spectrum as much as possible. Therefore, the smaller the bandgap, the better the material can achieve this goal. However, the materials with small bandgap will lead to smaller photovoltaic, and some photons with higher energy (i.e. shorter wavelength) will not convert the energy higher than the bandgap into electric energy, but will be wasted in the form of heat. If the material with large bandgap is chosen, the smaller photoelectric current will be caused. The theoretical efficiency of these single junction solar cell materials is below 30%. Figure 10.16 is the performance parameters of GaAs solar cells.

In order to overcome the shortcomings of poor mechanical strength and fragility of GaAs/GaAs solar cells, Ge was

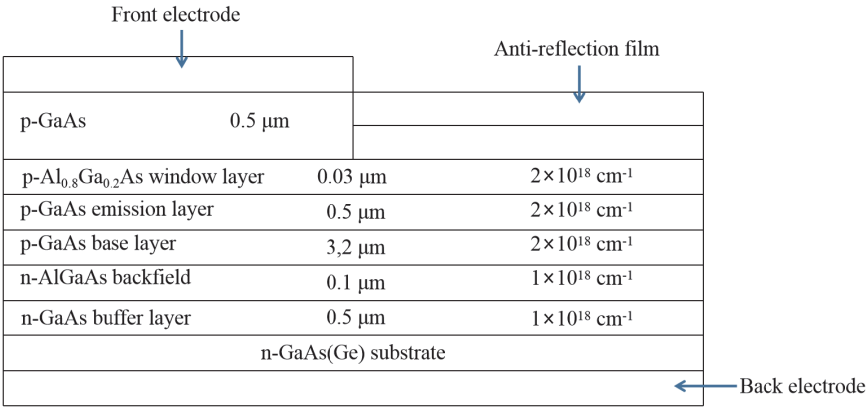


Fig. 10.17. Single-junction GaAs solar cell structure.

gradually used to replace GaAs to prepare substrates since 1983 (Fig. 10.17).

(ii) Multi-junction design

Because single-junction solar cells can only absorb and convert sunlight in a specific spectral range, the efficiency of energy conversion is not high. Multi-junction solar cells, superimposed from top to bottom according to the bandgap width, selectively absorb and convert different energy of solar spectrum, can greatly improve the conversion efficiency of solar cells. Figure 10.18 is the principle of multi-junction cascade solar energy spectral absorption.

Multi-junction solar cells can selectively absorb and convert different energy of solar spectrum, and greatly improve the conversion efficiency of solar cells. For multi-junction solar cells, the higher the bandgap of the upper layer, the larger the bandgap.

10.3.3.2. Fabrication technology of GaAs thin-films

Liquid-phase epitaxy (LPE) is an epitaxy growth technique proposed by NELSON in 1963. The principle is to use low melting point metals (such as Ga, In, etc.) as solvents, growth materials (such as

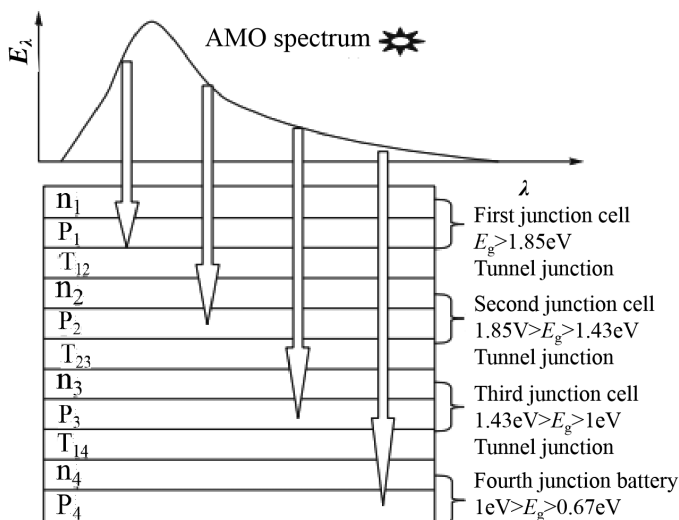


Fig. 10.18. Spectral absorption principle of multi-junction cascade solar energy.

GaAs, Al, etc.) and dopants (such as Zn, Te, Sn, etc.) as solutes to make the solutes saturated or supersaturated in solvents. Solvents are precipitated from solvents by cooling, crystallized on substrates, and realize crystal epitaxial growth.

LPE equipment has low cost and simple technology. It can be used in batch production of single junction GaAs/GaAs solar cells. Heterogeneous interface growth is impossible, multi-layer complex structure growth is difficult to achieve, and epitaxial layer parameters are difficult to accurately control, which limits the further improvement of GaAs solar cell performance.

In the early 1990s, the technology was no longer developed in foreign countries, but the LPE equipment was still retained in Europe, Japan and other countries for the development of small satellite power supply.

MOCVD is a method proposed by MANASEVIT in 1968 to prepare compound semiconductor thin-layer single crystals. The principle is that the metal organic compounds $\text{Ga}(\text{CH}_3)_3$, $\text{Al}(\text{CH}_3)_3$, $\text{Zn}(\text{C}_2\text{H}_5)_2$ and the hydrides of V and VI elements (PH_3 , AsH_3 ,

H₂Se) are used as raw materials for crystal growth. Vapor deposition (vapor-phase epitaxy) was carried out on the substrate by thermal decomposition to grow III–V, II–VI compound semiconductors and their ternary and quaternary compound semiconductors thin-films.

10.4. New Solar Cells

10.4.1. *Dye-sensitized solar cells*

Among many new types of solar cells, dye-sensitized solar cells (DSC) have developed rapidly in recent years. Its research history can be traced back to the 1960s. Tributsch found that dyes adsorbed on semiconductors can generate current under certain conditions, which laid an important foundation for photoelectrochemistry. In fact, before 1991, the photoelectric conversion efficiency of most dye-sensitized solar cells was relatively low (<1%). In 1991, a team led by Michael Graetzel of the Federal Institute of Higher Technology in Lausanne, Switzerland, introduced nanocrystalline porous films into dye-sensitized solar cells, which greatly improved their photoelectric conversion efficiency. Compared with silicon-based solar cells, dye-sensitized solar cells have the advantages of low cost, simple process and high photoelectric conversion efficiency.

10.4.1.1. *Structure of dye-sensitized solar cells*

Typical structures of dye-sensitized solar cells include porous nano-TiO₂ semiconductor films, transparent conductive glass, dye photosensitizers, hole transport media and pair electrodes, as shown in Fig. 10.19. Porous nano-TiO₂ film is the photoanode of solar cells, and its performance is directly related to the efficiency of solar cells. This kind of film is usually coated with nanocrystalline TiO₂ particles on the surface of conductive glass and sintered at high temperature to form porous electrodes.

Transparent conductive glass is usually ITO glass or TCO glass, which plays the role of transmitting and collecting electrons. Dye photosensitizers are adsorbed on the surface of porous electrodes,

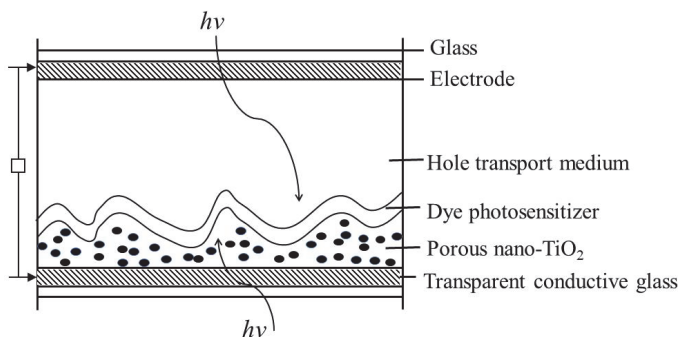


Fig. 10.19. Structure of dye-sensitized solar cells.

requiring wide visible absorption and long-term stability. Hole transport media mainly play the role of redox and electron transport. The main difference of various dye-sensitized batteries is also the difference of the hole transmission medium. Platinum electrodes or platinum electrodes with a single-electron layer are generally used for collecting electrons.

10.4.1.2. Principle of dye-sensitized solar cells

The basic working principle of dye-sensitized solar cells is as follows: when the energy is lower than the forbidden bandwidth of porous nano-TiO₂ film, but equal to the incident light of the characteristic absorption wavelength of dye molecules irradiated on the porous electrode, the electrons adsorbed in the dye molecules on the surface of the porous electrode are stimulated to transfer to the excited state, and then injected into the conduction band of TiO₂. The dye molecule itself becomes oxidized. Electrons injected into TiO₂ are enriched into conductive glass substrates by diffusion and then enter the external circuit. The oxidized dye molecules get electrons from the electrolyte solution and are reduced to the ground state. The oxidized electrons in the electrolyte diffuse to the opposite electrodes, which completes a photoelectrochemical reaction. In dye-sensitized solar cells, light energy is directly converted into electrical energy, and there is no net chemical change inside the cells.

DSC batteries work in a similar way to natural photosynthesis, which is different from traditional silicon batteries. The absorption of light is mainly achieved by dyes, while the separation and transmission of charges are controlled by the kinetic reaction rate. Charge transport in TiO_2 is accomplished by most carriers, so the requirements for material purity and preparation process are not very strict, which makes the cost of production greatly reduced.

10.4.1.3. *Advantages of dye-sensitized solar cells*

Low price, simple process: traditional solar cell light absorption and carrier transmission are accomplished by the same substance. In order to prevent the recombination of electrons and holes, the materials used must be of high purity and have no structural defects. Therefore, the process requirements of semiconductor are very high, which makes it difficult to reduce the cost. Dye-sensitized photoelectrochemical cells only produce carriers on one band, i.e. after photosensitization of the anode, electrons are injected into nano- TiO_2 conduction bands, while holes remain on the dyes on the surface. Therefore, the recombination of charges is limited, so polycrystalline and low purity materials can be used. The process is relatively simple and the cost is greatly reduced. At present, the price of dye-sensitized solar cells is 1/10–1/5 of that of silicon solar cells.

The theoretical photoelectric conversion efficiency is high: at present, the dye-sensitized solar cells are mainly liquid electrolyte, and the theoretical photoelectric conversion efficiency has been stable at more than 10%, which is no inferior to that of polycrystalline silicon solar cells. All-solid-state batteries using solid organic hole transport materials as electrolytes can even reach 33% under monochromatic light.

Other advantages: Dye-sensitized solar cells have high transparency and can be made into transparent products. When prepared on flexible substrates, batteries can be made into various forms, which greatly enlarges their applications. It can be used under various

illumination conditions. It is insensitive to the incident angle of light and can make full use of refracted light and reflected light. The working temperature is wide and the upper limit can be as high as 70°C.

10.4.1.4. *Problems and development prospects of dye-sensitized solar cells*

(i) Main problems of dye-sensitized solar cells at present

At present, the photoelectric conversion efficiency of dye-sensitized solar cells (area $< 0.5 \text{ cm}^2$) has reached 11.04%. However, for large-area dye-sensitized solar cells with practical significance, the photoelectric conversion efficiency has been around 5% (the highest 5.9%) and the area of batteries larger than 100 cm^2 has not been reported. Compared with traditional silicon solar cells, the conversion efficiency of dye-sensitized solar cells still needs to be improved.

At present, liquid electrolyte dye-sensitized solar cells, which are widely used, mainly use liquid organic small molecule compound solvents, have low boiling point, easy volatilization and high fluidity, which will cause a series of problems, such as electrode corrosion, electrolyte leakage, short life and so on, and bring difficulties to battery sealing and long-term use.

(ii) Development prospects of dye-sensitized solar cells

Because there are a series of problems in liquid electrolyte dye-sensitized solar cells, it is an important research direction to find suitable solid-state hole transport materials to replace liquid electrolyte to prepare all-solid-state dye-sensitized solar cells. In addition to all-solid-state sensitized solar cells, the future development direction of dye-sensitized solar cells includes the following aspects: low-temperature preparation and flexibility of high-efficiency electrodes (photoanodes and counterelectrodes). Design and development of cheap and stable full spectrum dyes. The encapsulation of liquid electrolyte, the preparation of high efficient solid electrolyte and the solution of related problems.

10.4.2. *Organic thin-film solar cells*

Although the power supply efficiency of organic thin-film solar cells is not as high as that of traditional solar cells, they are inexpensive in cost and have various uses, so their prospects are bright, and they have the following advantages.

Organic materials have the advantages of low cost, easy modulation, flexibility and film forming. Organic solar cells have relatively simple processing process, low-temperature operation and low cost of device fabrication. In addition, the potential advantages of organic solar cells include large-scale manufacturing, flexible substrates, environmentally friendly, portable and so on. Therefore, it is expected to play a power supply role in watches, portable calculators, semi-transparent chargers, toys, flexible curling systems and other systems.

10.4.2.1. *Structure of organic solar cells*

(i) Schottky structure

The basic physical process of the first organic solar cell device structure is as follows: the electrons in organic semiconductors are stimulated from HOMO level to LUMO level under solar irradiation to generate electron-hole pairs. Electrons are extracted by electrodes with low work function, and holes are filled by electrons from electrodes with high work function, thus forming photocurrent. Figure 10.20 is the structure and charge transfer diagram of organic solar cells.

The excitons formed by light excitation can be separated only in the diffusion layer of the Schottky junction by the electric field in the nodal region, while the excitons formed at other locations must first move into the diffusion layer to form the contribution of light current. However, the exciton migration distance in organic molecular materials is quite limited, usually less than 10 nm. So most excitons recombine before they are separated into electrons and holes, resulting in low photoelectric conversion efficiency.

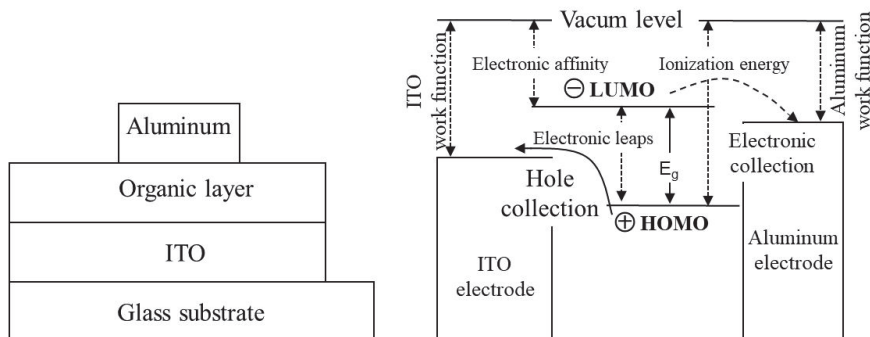


Fig. 10.20. Structure and charge transfer diagram of organic solar cells.

(ii) Double-layer organic solar cells

The basic physical processes of double-layer organic solar cells are as follows: light irradiation on the organic semiconductor materials as donors produces excitons, which dissociate at the interface between donors and acceptors, and then electron injection into the organic semiconductor materials as acceptors, the holes and electrons are separated. In this system, the electron donor is P-type and the electron acceptor is N-type, so the holes and electrons are transferred to two electrodes to form photocurrent.

Dual-layer organic solar cells were first developed by Kodak's Deng Qingyun. Compared with the Schottky battery mentioned above, this kind of structure is characterized by introducing the mechanism of charge separation, which makes it easier for excitons generated in organic materials to dissociate at the interface of the two materials to achieve charge separation, and greatly improves the efficiency of exciton dissociation. Thus, the efficiency of battery devices can be increased.

(iii) Bulk heterojunction structure

Based on the photoinduced electron transfer theory of conjugated polymer C60 (fullerene) system, the conjugated polymer MEH PPV and the derivative PCBM of fullerene were doped into bulk heterojunction structure in a certain proportion. Because the two materials are doped with each other and the

doping size is between several to tens of nanometers, the excitons formed at any point in the doped layer can be separated and formed charge at the interface within their diffusion length, thus achieving very high exciton separation efficiency.

10.4.2.2. *Common materials for organic solar cells*

(i) Photosensitive materials

Photosensitive materials are functional layers of photovoltaic effect, including donor materials and acceptor materials. Buffer layer, or barrier layer, transport layer and modification layer, mainly includes electron transport layer (ETL) and hole transport layer (HTL). The former blocks holes and transmits electrons, while the latter blocks electrons and transmits holes, thus playing the role of charge collection. The window electrode is TCO. Back electrodes are metal electrodes, such as Al, Ag, etc. Photosensitive materials generally include small molecular materials and polymer materials.

Small molecule materials are some dye molecules with conjugated system. They can absorb visible light very well, thus exhibiting good photoelectric conversion characteristics. They have the advantages of designable compound structure, light material, low production cost, good processing performance and easy preparation of large area solar cells. However, due to the poor solubility of organic small molecule materials, evaporation is usually used to prepare small molecule thin-films in organic solar cells.

The molecular structure of conductive polymer materials is characterized by large π -electron conjugated system, and the molecular weight of polymer materials affects the degree of conjugated system. The condensed state (amorphous and crystalline), crystallinity, orientation of crystal plane and crystalline morphology of materials have great influence on device performance. The main polymer materials are poly (p-styrene) (PPV), polyaniline (PANI), polythiophene (PTH) and their derivatives, such as the most common donor material P3HT and host material PCBM.

(ii) Electrode materials

In order to improve the output efficiency of electrons and holes in solar cell devices, materials with the lowest work function should be chosen as cathodes and materials with the highest work function as anodes. The selection of electrode material has a great influence on the determination of ohmic contact or rectification contact between electrode and organic material.

The material requirements of organic thin-film solar cells are as follows: the absorption spectrum range is wider and the absorption coefficient is higher, so more solar energy can be obtained. Appropriate HOMO and LUMO levels cannot only achieve effective charge separation, but also ensure enough V_{oc} . Molecules have better planarity, which is conducive to the accumulation of molecules and higher hole mobility, which is conducive to the transport of photogenerated carriers. The active layer can form ideal morphology and excellent charge transfer network by solution treatment, heat treatment or additives. It has good solubility, is conducive to solution processing, and has excellent chemical and optical stability, so it has practical value.

10.4.3. *Perovskite solar cells*

Perovskite-based solar cells have revolutionized the field of photovoltaics, aiming at high efficiency and low cost devices. Yang Yang of UCLA even calls it a new generation of solar cells. Therefore, the rapid development of perovskite in the past year can be predicted. With the continuous efforts of the relevant research groups, it is entirely reasonable to believe that it is possible to produce large-scale conversion efficiency poles through low-cost fabrication process by utilizing structural engineering, material engineering, interface engineering, energy band engineering and incident light management engineering. High green and efficient perovskite-based solar energy has really become the mainstream product of the new generation of low-cost and green energy industry.

In 2009, Akihiro Kojima first prepared quantum dots (9–10 nm) from $\text{CH}_3\text{NH}_3\text{PbI}_3$ and $\text{CH}_3\text{NH}_3\text{PbBr}_3$ and applied them to dye-sensitized solar cells (DSC). The performance of this kind of material

sensitized TiO_2 solar cells in visible light range was studied, and 3.8% photoelectric conversion efficiency was obtained.

In 2011, the experimental scheme was improved and optimized. The $\text{CH}_3\text{NH}_3\text{PbI}_3$ quantum dots were fabricated up to 2–3 mm, and the battery efficiency increased by 1 fold to 6.54%.

Spiro-OMeTAD is used as an organic hole transport material in perovskite batteries. The stability and process repeatability of perovskite batteries are greatly improved.

In 2013, with the continuous optimization of technology, the photoelectric conversion efficiency of perovskite solar cells soared to 15% in only about half a year.

In 2014, the highest photoelectric conversion efficiency of perovskite solar cells was close to 20%.

The biggest advantage of this new perovskite absorbent material is that its absorption coefficient is very large and its absorption capacity is more than 10 times higher than that of traditional dyes. However, the micro-mechanism of the new perovskite absorbent material is still uncertain.

10.4.3.1. *Perovskite material structure*

Perovskite material is a kind of peculiar structure with ABX_3 crystalline form, showing rich and colorful physical properties, including insulator, ferroelectricity, antiferromagnetism, giant magnetism/giant magnetism effect, famous for its superconductivity. The ABX_3 perovskite structure is composed of metal Pb atom as octahedral core, halogen Br atom as octahedral apex, and organic methylamines at the apex of face-centered cubic lattice. The characteristics of this organic halide perovskite structure are as follows.

Halogen octahedrons are connected by common vertices to form a three-dimensional network. According to Pauli's coordination polyhedron connection rules, this structure is more stable than common edge and coplanar connection.

The common top connection makes the gap between octahedral networks larger than that between common edges and coplanar connections. It allows larger size ions to fill in. Even if a large number

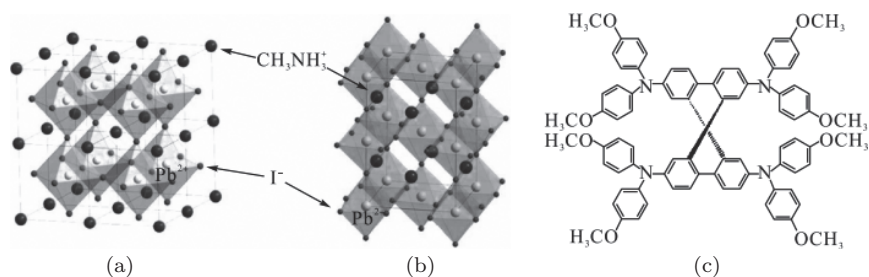


Fig. 10.21. Molecular structure of commonly used perovskite materials: (a) cubic $\text{CH}_3\text{NH}_3\text{PbI}_3$; (b) tetragonal $\text{CH}_3\text{NH}_3\text{PbI}_3$; and (c) spiro-OMeTAD.

of crystal defects are produced, or the size of the ions differs greatly from the geometrical requirements, the structure is stable and the defects are propitious to diffusion and migration. Figure 10.21 is the molecular structure of perovskite materials commonly used.

10.4.3.2. Typical battery structure of perovskite batteries

The typical structure and energy level diagram of perovskite batteries are shown in Fig. 10.22.

10.4.3.3. Principle of perovskite battery group

In dye-sensitized batteries, the transport of electrons in the electrolyte is very poor, and the stability of $\text{CH}_3\text{NH}_3\text{PbX}_3$ in the electrolyte is very poor. Therefore, solid transport is used in the structure of perovskite batteries to overcome these disadvantages.

In organic batteries, the exciton binding energy is very large (about 400 meV), which requires a strong built-in electric field to separate excitons, while in perovskite batteries, the exciton binding energy is only 50–76 meV. The exciton in the material is Wannier–Mott type, which means that at room temperature, photogenerated electron–hole pairs can be separated in the material and the open-circuit voltage of the battery can be increased.

The principle of perovskite batteries is to generate excitons in the photosensitive layer under light. Because of the small binding energy of excitons, separation can take place inside the material, which is

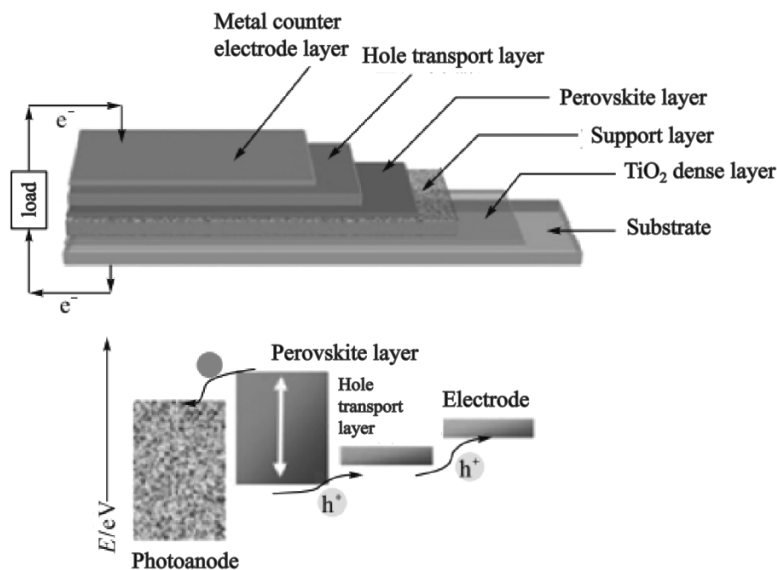


Fig. 10.22. Typical structure and energy level diagram of perovskite batteries.

transported through the electron-hole layer and finally collected by the electrodes.

The photophysical processes in the photosensitive layer are as follows: photoabsorption produces electron-hole pairs, which then evolve into highly detached Wannier excitons. A small part of them will spontaneously form free carriers. Excitons and free carriers coexist, and their dynamic number varies according to their lifetime. The double group of excitons is very weak. Defect-assisted reorganization can also be inhibited in some $\text{CH}_3\text{NH}_3\text{PbX}_3$ perovskites. The recombination of electrons and holes produced by exciton quenching is also very weak. Auger recombination is dominant here. Under the condition of high pump-in incentive, self-amplification recombination will compete with Auger recombination. At low light intensity, Auger and self-amplification recombination are suppressed.

Figure 10.23 is a schematic diagram of perovskite solar cell energy and electron transfer.

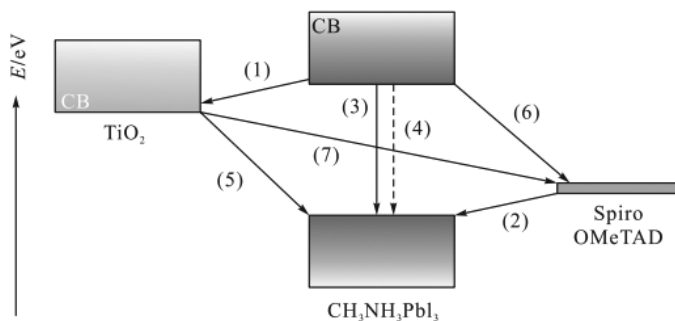


Fig. 10.23. Schematic chart of energy level and electron transfer of perovskite solar cells.

References

- An Q, Cao G-C, Li G-X, 1984. *Solar Cell Theory and Technology*. Shanghai: Shanghai Science and Technology Press.
- Luo Y, 2011. *Photovoltaic Cell Theory and Technology*. Beijing: Central Radio and Television University Press.
- Weng M, 2013. *Solar Cells — Materials, Manufacturing and Testing Technology*. Beijing: Science Press.
- Xiao X, Yang C, 2014. *Thin-Film Solar Cells*. Beijing: Science Press.
- Xiong S, Zhu M, 2009. *Fundamentals and Applications of Solar Cells*. Beijing: Science Press.
- Yang D, 2011. *Solar Cell Materials*. Beijing: Chemical Industry Press.