Brought to You by

Like the book? Buy it!

Semiconductor Materials and Devices Series

Semiconductor
Nanostructures
for Optoelectronic
Applications

Todd Steiner, editor

Semiconductor Nanostructures for Optoelectronic Applications

For a listing of recent titles in the *Artech House*
Semiconductor Materials and Devices Library, turn to the back of this book.

Semiconductor Nanostructures for Optoelectronic Applications

Todd Steiner

Editor

Artech House, Inc.
Boston • London
www.artechhouse.com

Library of Congress Cataloging-in-Publication Data

A catalog record of this book is available from the U.S. Library of Congress.

British Library Cataloguing in Publication Data

Semiconductor nanostructures for optoelectronic applications
—(Artech House semiconductor materials and devices library)
1. Semiconductors 2. Nanostructured materials 3. Optoelectronic devices
I. Steiner, Todd
621.3’8152

ISBN 1-58053-751-0

Cover design by Gary Ragaglia

© 2004 ARTECH HOUSE, INC.
685 Canton Street
Norwood, MA 02062

All rights reserved. Printed and bound in the United States of America. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Artech House cannot attest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

International Standard Book Number: 1-58053-751-0

10 9 8 7 6 5 4 3 2 1

Contents

CHAPTER 1

Introduction 1

1.1 Synopsis 1
1.2 Growth 1
1.3 Optoelectronic Devices Based on Semiconductor Nanostructures 2
1.4 Materials for Semiconductor Nanostructures 2
1.5 Summary 3

CHAPTER 2

Review of Crystal, Thin-Film, and Nanostructure Growth Technologies 5

2.1 Introduction 5
2.2 Review of Thermodynamics 6
2.2.1 Chemical Reactions 7
2.2.2 Phase Diagrams 7
2.3 Bulk Crystal Growth Techniques 8
2.3.1 Czochralski Method 8
2.3.2 Bridgman Method 11
2.3.3 Float-Zone Method 13
2.3.4 Lely Growth Methods 14
2.4 Epitaxial Growth Techniques 16
2.4.1 Liquid Phase Epitaxy 16
2.4.2 Vapor Phase Epitaxy 17
2.4.3 Molecular Beam Epitaxy 20
2.4.4 Metalorganic Chemical Vapor Deposition 24
2.4.5 Atomic Layer Epitaxy 29
2.5 Thin-Film Deposition Techniques 29
2.5.1 Plasma-Enhanced Chemical Vapor Deposition 29
2.5.2 Vacuum Evaporation 31
2.5.3 Sputtering 33
2.6 Growth of Nanostructures 34
2.6.1 Properties and Requirements of Quantum Dot Devices 35
2.6.2 Growth Techniques 36
References 41

CHAPTER 3

Quantum Dot Infrared Photodetectors 45

3.1 Introduction 45
3.2 QD and QDIP Structure Growth and Characterization 49
3.2.1 GaAs Capped Large and Small InAs QDs 50
3.2.2 AlGaAs Capped Large InAs MQD QDIP Structures 57
3.2.3 $In_xGa_{1-x}As$ Capped Small and Large InAs MQD-Based QDIP Structures 64
3.3 QDIP Device Characteristics 76
3.3.1 Device Structures 76
3.3.2 Unintentionally Doped Large (PIG) InAs/GaAs MQD-Based Detectors 77
3.3.3 QDIPs with AlGaAs Blocking Layers 87
3.3.4 InAs/InGaAs/GaAs QDIPs 92
3.3.5 Dual-Color QDIPs 102
3.4 Prognosis 107
Acknowledgments 109
References 109

CHAPTER 4

Quantum Dot Lasers: Theoretical Overview 113

4.1 Introduction: Dimensionality and Laser Performance 113
4.2 Advantages of an Idealized QD Laser 115
4.3 Progress in Fabricating QD Lasers 115
4.4 State-of-the-Art Complications 116
4.4.1 Nonuniformity of QDs 117
4.4.2 Parasitic Recombination Outside QDs 126
4.4.3 Violation of Local Neutrality in QDs 129
4.4.4 Excited States 131
4.4.5 Spatial Discreteness of Active Elements: Hole Burning 132
4.4.6 Intrinsic Nonlinearity of the Light-Current Characteristic 134
4.4.7 Critical Sensitivity to Structure Parameters 139
4.4.8 Dependence of the Maximum Gain on the QD Shape 142
4.4.9 Internal Optical Loss 143
4.5 Novel Designs of QD Lasers with Improved Threshold and Power Characteristics 148
4.5.1 Temperature-Insensitive Threshold 148
4.5.2 Enhanced Power Performance 150
4.6 Other Perspectives 151
References 153

CHAPTER 5

High-Speed Quantum Dot Lasers 159

5.1 Introduction 159
5.2 MBE Growth of Self-Organized QDs and Their Electronic Properties 160
5.2.1 Self-Organized Growth of In(Ga)As QDs 160
5.2.2 Electronic Spectra of In(Ga)As/GaAs QDs 161
5.3 Separate Confinement Heterostructure QD Lasers and Their Limitations 163
5.3.1 Carrier Relaxation and Phonon Bottleneck in Self-Organized QDs 164
5.3.2 Hot Carrier Effects in SCH QD Lasers 167
5.4 Tunnel Injection of Carriers in QDs 168
5.4.1 Tunneling-Injection Laser Heterostructure Design and MBE Growth 169
5.4.2 Measurement of Phonon-Assisted Tunneling Times 170
5.5 Characteristics of High-Speed Tunneling-Injection QD Lasers 172
5.5.1 Room Temperature DC Characteristics 172
5.5.2 Temperature-Dependent DC Characteristics 172
5.5.3 High-Speed Modulation Characteristics 174
5.6 Conclusion 183
Acknowledgments 183
References 183

CHAPTER 6

Zinc Oxide-Based Nanostructures 187

6.1 Introduction 187
6.1.1 General Properties of ZnO 187
6.1.2 ZnO One-Dimensional Nanostructures 189
6.2 Growth Techniques 191
6.2.1 Growth Mechanisms 191
6.2.2 Growth Techniques 194
6.2.3 Summary 210
6.3 Characterizations 211
6.3.1 Structural Characterizations 211
6.3.2 Optical Characterizations 215
6.4 Device Applications 219
6.4.1 Optical Devices 219
6.4.2 Electronic Devices 221
References 224

CHAPTER 7

Antimony-Based Materials for Electro-Optics 229
7.1 Introduction 229
7.1.1 Antimony 229
7.1.2 Sb-Based III-V Semiconductor Alloys 230
7.1.3 Bulk Single-Crystal Growth 232
7.1.4 Applications 232
7.2 III-Sb Binary Compounds: GaSb, AlSb, and InSb 235
7.2.1 GaSb 235
7.2.2 AlSb 239
7.2.3 InSb 242
7.3 InAsSb 250
7.3.1 Physical Properties 250
7.3.2 Growth of InAsSb 253
7.3.3 Characterizations 253
7.3.4 Device Measurement 256
7.4 InTlSb 259
7.4.1 MOCVD Growth of InTlSb 259
7.4.2 InTlSb Photodetectors 262
7.5 InBiSb 262
7.5.1 MOCVD Growth of InSbBi 262
7.5.2 InSbBi Photodetectors 265
7.6 InTlAsSb 266
7.7 InAsSb/InAsSbP for IR Lasers 267
7.7.1 Growth and Characterization of InAsSb and InAsSbP 268
7.7.2 Strained-Layer Superlattices 269
7.7.3 Device Results 271
7.8 GaSb/InAs Type II Superlattice for IR Photodetectors 273
7.8.1 Introduction 273
7.8.2 Experimental Results for Type II Photodetectors 275
Acknowledgments 284
References 285

CHAPTER 8

Growth, Structures, and Optical Properties of III-Nitride Quantum Dots 289
8.1 Introduction 289
8.2 Growth of III-Nitride QDs 291
8.2.1 MBE Growth of III-Nitride QDs 292
8.2.2 Other Techniques 314
8.3 Optical Properties of III-Nitride QDs 317
8.3.1 Effects of Quantum Confinement, Strain, and Polarization 318

8.3.2 GaN QDs 323
8.3.3 InGaN QDs 337
8.4 Summary 343
References 344

CHAPTER 9

Self-Assembled Germanium Nano-Islands on Silicon and Potential Applications 349

9.1 Introduction 349
9.2 Heteroepitaxy Mechanisms 349
9.3 Uniform Ge Islands 350
9.4 Registration and Regimentation of Ge Islands 355
9.5 Novel Device Applications 362
9.5.1 Optoelectronics 362
9.5.2 Thermoelectricity 365
9.5.3 Electronics Applications 366
9.5.4 Quantum Information Applications 366
9.6 Conclusion 367
References 367

CHAPTER 10

Carbon Nanotube Engineering and Physics 371

10.1 Introduction 371
10.2 Controlled Fabrication of Uniform Nanotubes in a Highly Ordered Array 373
10.3 Interfacing with Biomolecules and Cells 379
10.4 Intrinsic Quantum Electromechanical Couplings 382
10.5 Extrinsic Coupling to Radiation Fields 391
10.6 Heterojunction Nanotubes 392
10.7 Prospects for Future Advances 396
Acknowledgments 398
References 398

Acronyms 403

About the Editor 407

Index 409

CHAPTER 1

Introduction

Todd Steiner, Air Force Office of Scientific Research

1.1 Synopsis

As we begin the twenty-first century, nanoscience and technology are advancing at a rapid pace and making revolutionary contributions in many fields including electronics, materials science, chemistry, biology, structures and mechanics, and optoelectronics. Although nanoscience and technology are progressing along many fronts, the most impressive progress has been made in the area of semiconductor technology. This book reviews recent progress in semiconductor nanostructure growth and materials development and also reviews progress in semiconductor devices using nanostructures, with a particular emphasis on 3D nanostructures that have emerged during the last 10 years.

1.2 Growth

Semiconductor nanostructures have been enabled by the advancements in epitaxial growth techniques, which are now capable of growing epilayers as thin as one atomic layer and with interface roughnesses that are a mere fraction of a monolayer. The development of advanced crystal and thin-film growth technologies capable of realizing high crystalline quality and purity of materials is an enabling step in bringing semiconductor devices to reality. These growth techniques are reviewed in Chapter 2. Chapter 2 starts with an overview of the bulk crystal growth techniques that are required for obtaining high-quality substrates, then looks at the primary means for producing high-quality epilayers, including liquid phase epitaxy, vapor phase epitaxy, molecular beam epitaxy, *metalorganic chemical vapor deposition* (MOCVD), and *atomic layer epitaxy* (ALE), as well as techniques for thin-film deposition including plasma-enhanced chemical vapor deposition, electron cyclotron resonance, vacuum evaporation, and sputtering. Chapter 2 then discusses the different growth modes of low-dimensional structures such as quantum wires and quantum dots.

1.3 Optoelectronic Devices Based on Semiconductor Nanostructures

Since the successful development of quantum well lasers in the 1970s, one of the richest areas of application of semiconductor nanostructures has been in the area of optoelectronic devices, with the two most important areas being semiconductor lasers and detectors. Early efforts focused on band-to-band transitions and have progressed more recently to intersubband devices. In addition, the early devices utilized 2D nanostructures, either superlattices or quantum wells. In recent years, the growth of quantum dots and their integration into working devices has revolutionized semiconductor devices. This book highlights results in semiconductor devices based on *quantum dots* (QDs).

In Chapter 3, we review progress on *quantum dot infrared detectors* (QDIPs) by providing a comprehensive discussion of the growth, structural and optical characterization, and device figures of merit. We discuss the QD and the QDIP structure growth, QD size distribution, and the tailoring of the QD electronic energy levels and wave functions via manipulation of the QD confinement potential. We also show how to take advantage of stress manipulation to realize multiple-color QDIPs. One section focuses on the QDIP device characteristics (dark current, responsivity, noise, photoconductive gain, detectivity) for each of three classes of QDIPs discussed: InAs/GaAs/AlGaAs, InAs/InGaAs/GaAs, and dual-color InAs/InGaAs/GaAs QDIPs.

In Chapter 4, we provide a theoretical overview of QD lasers, including the advantages of QD lasers over quantum well lasers, the recent progress in fabricating QD lasers, and a theoretical treatment of many issues of practical importance in developing QD lasers, such as the nonuniformity of QDs, parasitic recombination outside of QDs, threshold and power characteristics, and nonlinear properties. The chapter also includes novel designs for QD lasers with improved threshold and power characteristics.

In Chapter 5, we provide an overview of InGaAs tunnel injection QD lasers, which have demonstrated the lowest thresholds for QD lasers and the highest modulation bandwidths. This chapter describes the growth of these QD lasers, the unique carrier dynamics observed in self-organized QDs, their effect on high-frequency performance of QD lasers, and the novel injection technique whereby electrons are injected into the QD ground state by tunneling. The enhanced performance of these tunnel injection QD lasers is also described and discussed.

1.4 Materials for Semiconductor Nanostructures

Progress in semiconductor nanostructures is advancing to a wide variety of material systems. In this book we highlight the progress in five important material systems of technological importance. Each of these material systems has demonstrated 2D and 3D nanostructures and has had varying degrees of success in the fabrication of optoelectronic devices.

In Chapter 6 we review progress in zinc oxide-based nanostructures, including the Zno/ZnMgO system. Zinc oxide is emerging as an important material for

ultraviolet and visible optoelectronic applications, due to the ease with which light emission can be obtained. In Chapter 7 we review progress in antimony-based nanostructures, including the binary compounds GaSb, InSb, and AlSb; the tertiary compounds InAsSb, InAsP, InTlSb, and InSbBi; and the quaternary compounds InTlAsSb and InAsSbP. Devices based on these materials are also discussed. In Chapter 8 we review recent advances in the growth of III-nitride quantum dots and their unique properties. The growth techniques and the structural and optical properties associated with quantum confinement, strain, and polarization in GaN and $In_xGa_{1-x}N$ quantum dots are discussed in detail.

In Chapter 9 we review the progress of nanostructures in the silicon/germanium material system, which has the potential for bringing optoelectronics and photonics to silicon. Specifically, we review issues of Ge island formation on Si. We show uniform Ge island formation on planar Si and ordered island formation on prepatterned mesa structures. We discuss the effect of growth conditions such as growth temperature, deposition rate, deposition coverage, and substrate patterning on the formation of the islands. We discuss the potential applications of Ge islands in the fields of optoelectronics, thermoelectricity, electronics, and quantum information. In Chapter 10, we present a review of carbon nanotubes, especially for optoelectronics applications. The field of carbon nanotubes has advanced quickly and widely on many fronts during the past decade. Controlled fabrication of carbon nanotubes of uniform diameter, length, and spacing is now feasible. Real and perceived potential applications in electronics, sensing, molecular biology, actuation, composite material, and energy storage have been demonstrated. We introduce some of these advances and some of the fundamental properties of the carbon nanotubes, discuss the underlying physics of new effects and phenomena observed or anticipated, and describe the controllable fabrication processes of new forms of nanotubes, as well as some interesting and relatively new and unconventional directions of potential applications.

1.5 Summary

As we enter the twenty-first century, semiconductor nanostructures are revolutionizing many areas of electronics, optoelectronics, and photonics. We present in this volume some of the more interesting results that are leading the revolution in the area of optoelectronics. It is in this area that the real benefits of 3D structures are being realized for practical devices. These achievements will serve to enhance the contributions of semiconductor nanostructures in other areas, helping to maintain the leading position of semiconductor nanotechnology in the more general world of nanoscience and technology.

CHAPTER 2

Review of Crystal, Thin-Film, and Nanostructure Growth Technologies

Alireza Yasan and Manijeh Razeghi, Northwestern University

2.1 Introduction

During the latter half of the twentieth century, in an effort to increase integration, enhance functionality, and reduce energy consumption, the major focus of the development of semiconductor devices was on miniaturization. As a result, semiconductor devices have evolved from millimeter-sized devices capable of manipulating electricity (e.g., transistors) into micrometer-sized devices that can handle both electricity and light (e.g., light-emitting diodes). As we enter the twenty-first century, we envision nanometer-sized semiconductor devices that can directly interact with individual atoms and molecules at the nanometer level (e.g., quantum sensors). In this regard, the development of advanced crystal and thin-film synthesis technologies capable of realizing high crystalline quality and purity of materials is an enabling step toward making such semiconductor devices a reality.

We begin this chapter by giving an overview of *thermodynamics*. Chemical reactions and phase diagrams are the subject of this first section, after which we move on to a discussion of crystal growth techniques.

The earliest crystal growth techniques consisted of growing semiconductor crystals in bulk form using one of the *bulk crystal growth techniques*: Czochralski, Bridgman, or float zone. These methods are appropriate for the synthesis of large-volume semiconductor crystals under thermodynamic equilibrium conditions, but offer nearly no flexibility in terms of alloy composition or the heterostructures needed for advanced semiconductor devices. Nevertheless, these are excellent techniques for manufacturing high-purity, near perfect, single-crystal wafers to be used as substrates for epitaxial growth.

Epitaxial growth techniques have been specifically developed to enable the growth of high-quality semiconductor alloys under controlled conditions. Using these techniques, single-crystal semiconductor thin films are synthesized on a substrate. As the need for even more complex semiconductor devices increased, several techniques have been successively developed and refined to satisfy these ever-evolving needs. Liquid phase epitaxy is the oldest epitaxial growth technique. Although still used in some instances, this technique is losing momentum because of

its poor thickness uniformity and control and poor interface. The second technique, vapor phase epitaxy, has enjoyed broader success, but the material generally suffers from surface defects. It is nevertheless gaining interest in the case of GaN-based semiconductors. Two other techniques, molecular beam epitaxy and metalorganic chemical vapor deposition, are the most widely used techniques and have demonstrated unsurpassed capabilities in the epitaxial growth of numerous semiconductor structures, in terms of material quality, process control, and reliability.

Other thin-film deposition techniques exist that are primarily used for the deposition of dielectric films, but can also be used for the deposition of semiconductors in a polycrystalline form. These techniques include plasma-enhanced chemical vapor deposition, electron cyclotron resonance, vacuum evaporation, and sputtering. They are much simpler and cheaper than the epitaxial growth techniques, but are not as flexible and do not yield material that is as high in quality. Nevertheless, they are well suited for the deposition of the dielectric films commonly employed in the manufacturing process used for semiconductor devices.

Finally, we conclude by discussing low-dimensional structures such as quantum wires and quantum dots. Different growth modes and possible growth techniques are presented. The requirements for room-temperature operation of devices based on nanostructures are briefly discussed.

2.2 Review of Thermodynamics

In this section we briefly review the thermodynamics of materials. Thermodynamics tells us whether or not a reaction is possible. It can also determine, to some extent, the feasibility of a chemical reaction. To get such information the free-energy function, G, is often used:

$$G = H - TS \tag{2.1}$$

where H is the enthalpy, S is the entropy, and T is the absolute temperature. Let's assume that the initial state of the system (i) changes to a final state (f) due to a chemical reaction while the temperature is kept constant. The free-energy change can be written as

$$\Delta G = G_f - G_i = \Delta H - T\Delta S \tag{2.2}$$

The second law of thermodynamics states that "in all energy exchanges, if no energy enters or leaves the system, the potential energy of the state will always be less than that of the initial state ($\Delta G < 0$)." This implies that systems tend to minimize the free energy to a lower value than the initial value. After the system has achieved the equilibrium, $\Delta G = 0$. For a process that cannot occur, $\Delta G > 0$. Therefore, the possibility of occurrence of a particular reaction can be determined through the sign of ΔG.

2.2.1 Chemical Reactions

For a typical chemical reaction involving materials X, Y, and Z in equilibrium with x, y, and z as the stoichiometric coefficients,

$$xX + yY \rightarrow zZ \tag{2.3}$$

The free-energy change of the reaction is given by

$$\Delta G = zG_Z - xG_X - yG_Y \tag{2.4}$$

Free energy of individual reactants is often written as

$$G_i = G_i^0 + RT \ln a_i \tag{2.5}$$

where G_i^0 is the free energy of the species in its standard state and a_i is a term called *activity* that reflects the change in free energy when the material is not in its standard state. The standard state is typically 1 atmosphere partial pressure for a gas at 25°C. A pure liquid or solid is the standard state of the relevant substance. Table 2.1 lists the standard values of enthalpy and entropy for various substances [1]. Substitution of (2.5) into (2.4) and letting $\Delta G = 0$ yields

$$-\Delta G^0 = RT \ln K \tag{2.6}$$

where

$$K = \frac{a_{Z(eq)}^z}{a_{X(eq)}^x a_{Y(eq)}^y} \tag{2.7}$$

2.2.2 Phase Diagrams

Phase diagrams are the primary visualization tools in materials science because they allow one to predict and interpret changes in the composition of a material from

Table 2.1 Standard Values of Enthalpy and Entropy for Various Species

Species	*State*	*ΔH_f (kJ/mol)*	*S (J/mol·K)*
CO_2	Gas	–393.51 ± 0.13	213.785 ± 0.010
Cl_2	Gas	0	223.081 ± 0.010
H	Gas	217.998 ± 0.006	114.717 ± 0.002
H^+	Aqueous	0	0
H_2O	Liquid	–285.830 ± 0.040	69.95 ± 0.03
H_2O	Gas	–241.826 ± 0.040	188.835 ± 0.010
N	Gas	472.68 ± 0.40	153.301 ± 0.003
NH_3	Gas	–45.94 ± 0.35	192.77 ± 0.05
O_2	Gas	0	205.152 ± 0.005

phase to phase. As a result, phase diagrams have been proven to provide an immense understanding of how a material forms microstructures within itself, leading to an understanding of its chemical and physical properties. However, in some instances materials have failed to perform to their proposed potential. One can deduce, by referring to a material's phase diagram, what may have happened to the material when it was made to cause failure. In these instances, one can use thermodynamic relations to go into the phase diagrams and extrapolate the data.

A few simple rules are associated with phase diagrams with the most important of them being the *Gibb's phase rule*. The phase rule describes the possible number of degrees of freedom in a (closed) system at equilibrium, in terms of the number of separate phases and the number of chemical constituents in the system. It can be simply written as follows:

$$f = C - P + 2 \tag{2.8}$$

where C is the number of components, P is the number of phases, and f is the number of degrees of freedom in the system. The number of degrees of freedom (f) is the number of independent intensive variables (i.e., those that are independent of the quantity of material present) that need to be specified in value to fully determine the state of the system. Typical such variables might be temperature, pressure, or concentration. This rule states that for a two-component, one-phase system, there are two degrees of freedom. For example, on a P-T diagram, pressure and temperature can be chosen independently. On the other hand, for a two-phase system, there is only one degree of freedom and there is only one pressure possible for each temperature. Finally, for a three-phase system, there exists only one point with fixed pressure and temperature (Figure 2.1). As a real-world example, the P-T-x phase diagram of the Ga-N system at a fixed pressure of 1 atm is shown in Figure 2.2 [2].

2.3 Bulk Crystal Growth Techniques

The historical starting point for virtually all semiconductor devices has been in the synthesis of single crystals. Today, three major methods have been developed to realize large-volume semiconductor crystals under thermodynamic equilibrium conditions: the Czochralski, Bridgman, and float-zone methods, which are discussed in the following subsections.

2.3.1 Czochralski Method

The *Czochralski* (CZ) crystal growth method was developed in 1916 by accident. Jan Czochralski, an engineer at the AEG Company in Berlin at that time, accidentally dipped his pen into a crucible containing molten tin and withdrew it quickly. He observed a thin wire of solidified metal hanging at the tip. This small observation later led to development of the Czochralski method for obtaining single crystals [3]. The Czochralski method is by far the most popular crystal growth method,

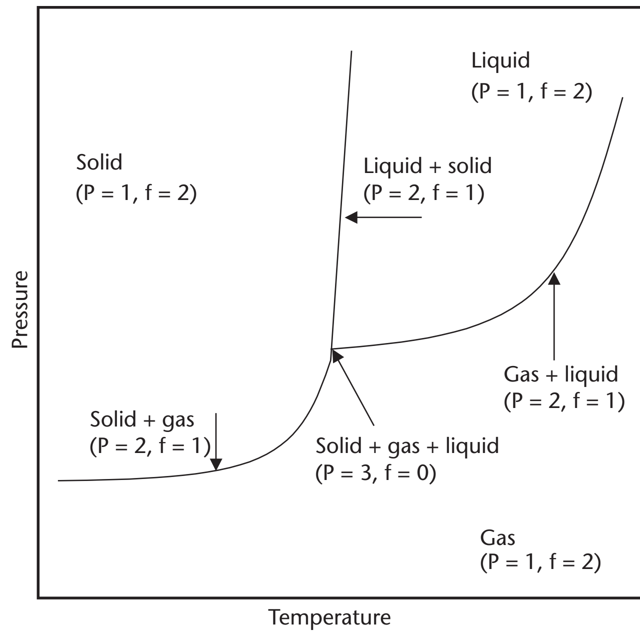


Figure 2.1 P-T diagram of a one-component system showing degrees of freedom for a different number of phases.

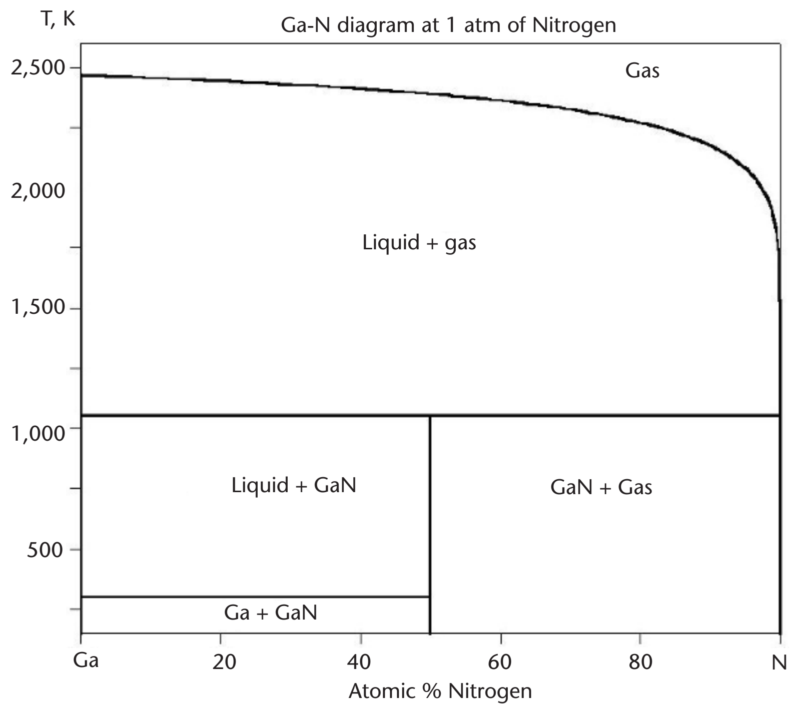


Figure 2.2 Calculated P-T-x phase diagram for Ga-N at atmospheric pressure.

accounting for between 80% and 90% of all silicon crystals grown for the semiconductor industry.

The Czochralski method uses a high-purity quartz (SiO_2) crucible filled with pieces of polycrystalline material, called *charge,* which are heated above their melting point (e.g., 1,415°C for silicon). The crucible, shown in Figure 2.3, is heated either by induction using radio-frequency (RF) energy or by thermal resistance methods. A "seed" crystal, about 0.5 cm in diameter and 10 cm long, with the desired orientation is lowered into molten crystal, termed *melt,* and then drawn up at a carefully controlled rate. When the procedure is properly done, the material in the melt will make a transition into a solid phase crystal at the solid/liquid interface, so the newly created material accurately replicates the crystal structure of the seed crystal. The resulting single crystal is called the *boule.* Modern boules of silicon can reach diameters of more 300 mm and be up to 2m long.

During the entire growth period, the crucible rotates in one direction at 12 to 14 *rotations per minute* (rpm), while the seed holder rotates in the opposite direction at 6 to 8 rpm while being pulled up slowly. This constant stirring prevents the formation of local hot or cold regions. The crystal diameter is monitored by an optical pyrometer that is focused at the interface between the edge of the crystal and the melt. An automatic diameter control system maintains the correct crystal diameter through a feedback loop control. Argon is often used as the ambient gas during this

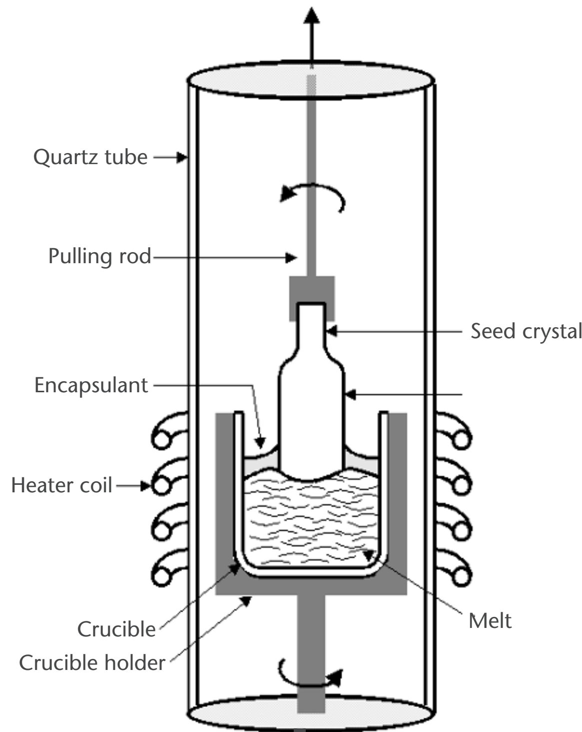


Figure 2.3 Cross section of a furnace used for the growth of single-crystal semiconductor boules by the Czochralski process.

crystal-pulling process. By carefully controlling the pull rate, the temperature of the crucible, and the rotation speed of both the crucible and the rod holding the seed, precise control over the diameter of the crystal is obtained.

During the Czochralski growth process, several impurities are incorporated into the crystal. Because the crucibles are made from fused silica (SiO_2) and the growth process takes place at temperatures around 1,500°C, small amounts of oxygen will be incorporated into the boule. For extremely low concentrations of oxygen impurities, the boule can be grown under magnetic confinement. In this situation, a large magnetic field is directed perpendicularly to the pull direction and used to create a Lorentz force. This force will change the motion of the ionized impurities in the melt in such a manner as to keep them away from the solid/liquid interface and therefore decrease the impurity concentration. Using this arrangement, the oxygen impurity concentration can be reduced from about 20 *parts per million* (ppm) to as low as 2 ppm.

It is also common to introduce dopant atoms into the melt in order to tailor the electrical properties of the final crystal: the carrier type and concentration. Simply weighing the melt and introducing a proportional amount of impurity atoms is all that is theoretically required to control the carrier concentration. However, impurities tend to segregate at the solid/liquid interface, rather than be uniformly distributed inside the melt. This will in turn affect the amount of dopant incorporated into the growing solid.

The growth of GaAs with the Czochralski method is far more difficult than for silicon because of the vast differences in vapor pressure of the constituents at the growth temperature of ~1,250°C: 0.0001 atm for gallium and 10,000 atm for arsenic. The liquid encapsulated Czochralski (LEC) method utilizes a tightly fitting disk and sealant around the melt chamber (see the encapsulant in Figure 2.3) to prevent the out-diffusion of arsenic from the melt. The most commonly used sealant is boric oxide (B_2O_3). Additionally, *pyrolytic boron nitride* (pBN) crucibles are used instead of quartz (silicon oxide) in order to avoid silicon doping of the GaAs boule. Once the charge is molten, the seed crystal can be lowered through the boric oxide until it contacts the charge, at which point it may be pulled.

Because the thermal conductivity of GaAs is about one-third that of silicon, the GaAs boule is not able to dissipate the latent heat of fusion as readily as silicon. Furthermore, the shear stress required to generate a dislocation in GaAs at the melting point is about one-fourth that in silicon. Consequently, the poorer thermal and mechanical properties allow GaAs boules to be only about 8 inches in diameter [4] and they contain many orders of magnitude larger defect densities than are realized in silicon.

2.3.2 Bridgman Method

The Bridgman crystal growth method is similar to the Czochralski method except that all of the semiconductor material (melt, seed, crystal) is kept completely inside the crucible during the entire heating and cooling processes, as shown in Figure 2.4.

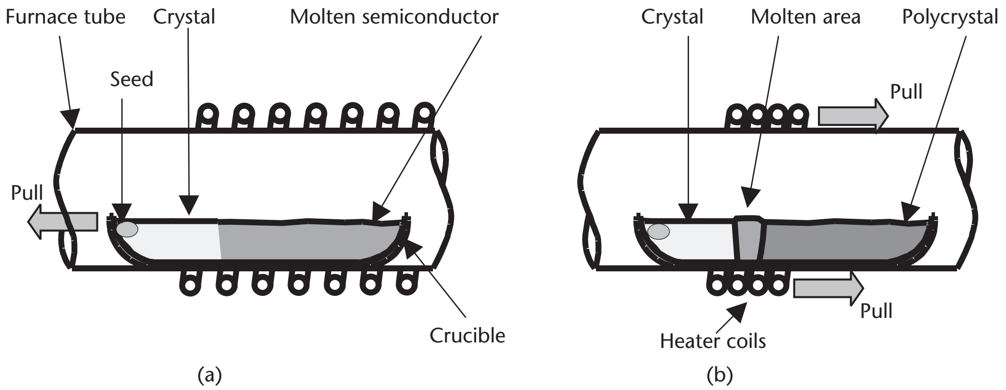


Figure 2.4 The Bridgman growth method in a crucible: (a) solidification from one end of the melt and (b) melting and solidification in a moving heated zone.

A quartz crucible filled with polycrystalline material is pulled horizontally through a furnace tube. As the crucible is drawn slowly from the heated region into a colder region, the seed crystal induces single-crystal growth. The shape of the resulting crystal is determined by the shape of the crucible. As a variation to this procedure, the heater may move instead of the crucible. As an alternative, the heater may move instead of the crucible.

A couple of disadvantages are associated with the Bridgman growth method. They result from the fact that the material is constantly in contact with the crucible. First, the crucible wall introduces stresses in the solidifying semiconductor. These stresses will result in deviations from the perfect crystal structure. Also, at the high temperatures required for bulk crystal growth, silicon tends to adhere to the crucible.

In the case of compound semiconductors, the process is slightly different from that for silicon. The solid gallium and arsenic components are loaded onto a fused silica ampule, which is then sealed. The arsenic in the chamber provides the overpressure necessary to maintain stoichiometry. A tube furnace is then slowly pulled past the charge. The temperature of the furnace is set to melt the charge when it is completely inside. As the furnace is pulled past the ampule, the molten GaAs charger in the bottom of the ampule recrystallizes. A seed crystal may be mounted so as to contact the melt.

Typical compound semiconductor boules grown by the Bridgman method have diameters of 2 inches. The growth of larger crystals requires very accurate control of the stoichiometry and the radial and axial temperature gradients. Dislocation densities of lower than 10^3 cm^{-2}, compared to 10^4 cm^{-2} for boules grown by the CZ method, are routinely achieved with the Bridgman method. This method produces the best results for compound semiconductor growth such as GaAs, and approximately 75% of the compound semiconductor boules are grown by the Bridgman growth method.

2.3.3 Float-Zone Method

Unlike the previous two methods, the *float-zone* (FZ) technique proceeds directly from a rod of polycrystalline material obtained from the purification process as shown in Figure 2.5. Moreover, this method does not make use of a crucible. For this reason, extremely high-purity silicon boules, with carrier concentrations lower than 10^{11} cm^{-3}, have been grown by the FZ method. But in general, this method is not used for compound semiconductor growth.

The principle of the FZ method is as follows. A rod of an appropriate diameter is held at the top of the growth furnace and placed in the crystal-growing chamber. A single-crystal seed is clamped in contact at the other end of the rod. The rod and the seed are enclosed in a vacuum chamber or inert atmosphere, and an inductive-heating coil is placed around the rod outside the chamber. The coil melts a small length of the rod, starting with part of the single seed crystal. A "float zone" of melt is formed between the seed crystal and the polysilicon rod. The molten zone is slowly moved up along the length of the rotating rod by moving the coil upward. High-purity crystals can be obtained with FZ method.

The molten region that solidifies first remains in contact with the seed crystal and assumes the same crystal structure as the seed. As the molten region is moved along the length of the rod, the polycrystalline rod melts and then solidifies along its

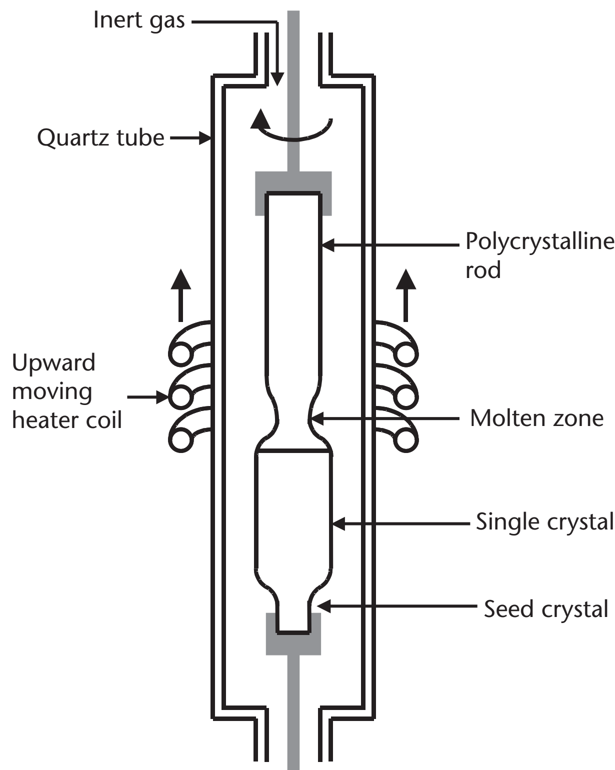


Figure 2.5 Cross section of the FZ crystal growth furnace.

entire length, becoming a single-crystal rod of silicon in the process. The motion of the heating coil controls the diameter of the crystal. Because of the difficulties in preventing collapse of the molten region, this method has been limited to small-diameter crystals (less than 76 mm). However, because no crucible is involved in the FZ method, oxygen contamination that might arise from the quartz (SiO_2) crucible is eliminated. Wafers manufactured by this method are used in applications requiring low-oxygen-content, high-resistivity starting materials for devices such as power diodes and power transistors.

One disadvantage of FZ crystal growth is in the difficulty of introducing a uniform concentration of dopants. Currently, four techniques are used: core doping, gas doping, pill doping, and neutron doping.

Core doping uses a doped polysilicon boule as the starting material. Undoped material can be deposited on top of the doped boules until the desired overall doping concentration is obtained. This process can be repeated several times to increase the uniformity or the dopant distribution and, neglecting the first few melt lengths, the dopant distribution is very good. Gas doping simply uses the injection of gases, such as $AsCl_3$, PH_3, or BCl_3, into the polycrystalline rod as it is being deposited or into the molten ring during FZ refining. Pill doping is accomplished by inserting a small pill of dopant into a hole that is bored at the top of the rod. If the dopant has a relatively low segregation coefficient, most of it will diffuse into the rod as the melt passes over the rod. Gallium and indium are commonly used as pill dopants. Finally, light n-type doping of silicon can be achieved with neutron bombardment. This is possible because approximately 3.1% of silicon mass is the mass 30 isotope.

2.3.4 Lely Growth Methods

Although they account for nearly all bulk semiconductor boules grown commercially, the previously described techniques all make use of the crystallization process from a melt. This is not possible for a growing number of semiconductor materials, such as silicon carbide (SiC) and gallium nitride (GaN) based materials, because they do not have a liquid phase under reasonable thermodynamic conditions. For example, SiC melt can exist only under pressures higher than 10^5 atm and temperatures higher than 3,200°C. Furthermore, under these conditions, the stoichiometry and stability of the melt could no longer be ensured. At this time, two techniques deserve to be mentioned as successful for the growth of bulk SiC semiconductor boules: the Lely method and the modified Lely method.

The Lely growth method [5] has yielded the highest quality crystals to date and is carried out in a cylindrical crucible, as depicted schematically in Figure 2.6. The growth process is basically driven by a temperature gradient. This temperature gradient is maintained between the outer and the inner areas of the crucible, with a lower temperature at the center. At the same time, the system is kept near chemical equilibrium, with lower partial pressures of SiC precursors in the inner and colder region. The two areas are separated by a porous graphite, which also provides nucleation centers.

The chemical gradient results in a mass transport occurring from the outer toward the inner region. Because the inner region is colder than the outer, SiC will

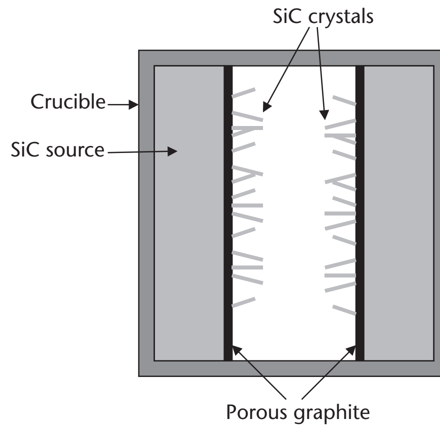


Figure 2.6 Cross section of a cylindrical crucible used for the Lely growth of SiC.

nucleate on the graphite and crystals will start to grow under their most energetically stable form. Although of the highest quality in terms of low defect density, the resulting crystals are limited in size and their dimensions are random (typically smaller than 1 cm^2). These are nevertheless used as seed crystals for all other bulk SiC growth methods, including the modified Lely method.

The modified Lely method [6] is the historical name for the seeded sublimation growth or physical vapor transport technique. Its principle is similar to the Lely method except that a SiC seed crystal is used to achieve a controlled nucleation. This method is currently used for the growth of all commercial SiC single-crystal boules. A modern crucible for the modified Lely technique is depicted schematically in Figure 2.7. The cooler seed is placed at the top to prevent falling contaminants. The polycrystalline SiC source is heated (up to 2,600°C) at the bottom of the crucible, and it sublimes at low pressure. Mass transport occurs naturally and SiC naturally recrystallizes through supersaturation at the seed.

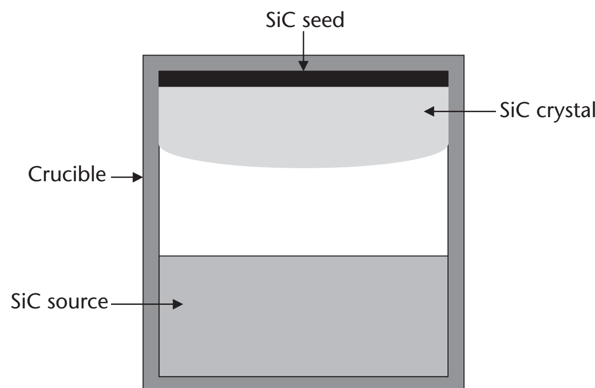


Figure 2.7 Cross section of the most commonly used modified Lely growth configuration.

Although the modified Lely method is more than 20 years old and has been able to advance the growth of bulk SiC semiconductor crystals, major issues remain. Indeed, the polytype formation and the growth shape are poorly controlled. The doping is not uniform and a high density of defects, such as micropipes and dislocations, remains [7].

2.4 Epitaxial Growth Techniques

Even the simplest semiconductor device requires the deposition of a series of crystal films on top of finely polished wafer substrates obtained through the bulk crystal growth techniques previously described. This process of extending the crystal structure of the underlying substrate material into the grown layer is called *epitaxy*. The term *epitaxy* is a combination of two Greek words, *epi* (meaning "placed" or "resting on") and *taxis* (meaning "arrangement") and refers to the formation of single-crystal films on top of a crystalline substrate.

The term *epitaxy* can be further qualified depending on the relationship between the film and substrate: *Homoepitaxy* is employed when the film and the substrate are the same material, and *heteroepitaxy* when the film and the substrate are different materials. Homoepitaxy results in a film that is totally lattice matched to the substrate, whereas heteroepitaxy can result in a strained or a relaxed film depending on the difference in lattice parameters and thermal expansion coefficients of the film and the substrate. An example of homoepitaxy can be growth of Si on Si substrate and an example of heteroepitaxy can be growth of InP on GaAs substrate or GaN on sapphire substrate.

The discovery of quantum wells and superlattices has revolutionized the area of semiconductor devices. These devices require ever more precise control, more uniform thickness over larger areas, excellent homogeneity, high purity, very sharp interfaces between the substrate and epitaxial layers, and low misfit dislocations in the epilayers. Historically, epitaxial techniques have been developed in order to progressively satisfy these requirements, from liquid phase epitaxy to vapor phase epitaxy, molecular beam epitaxy, and metalorganic chemical vapor deposition, which are reviewed in the following subsections.

2.4.1 Liquid Phase Epitaxy

The *liquid phase epitaxy* (LPE) growth technique [8] involves the precipitation of material from a supercooled solution onto an underlying substrate. The LPE reactor includes a horizontal furnace system and a sliding graphite boat as shown in Figure 2.8. The apparatus is quite simple and excellent quality layers and high purity levels can be achieved.

LPE is a thermodynamic equilibrium growth process. The composition of the layers that are grown on the substrate depends mainly on the equilibrium phase diagram and to a lesser extent on the orientation of the substrate. This melt is placed in a graphite boat and is slid inside a hot furnace that is at a suitable atmosphere. A

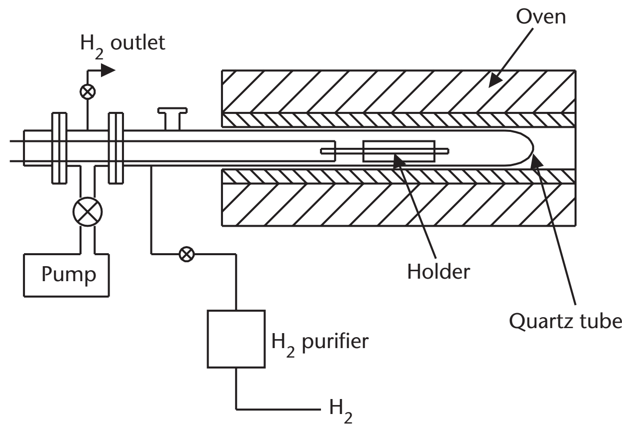


Figure 2.8 Cross section of an LPE system.

subsequent cooling causes the solute to come out and deposit on an underlying substrate, thus forming an epitaxially grown layer. The three major parameters that can affect the growth are the melt composition, growth temperature, and growth duration.

The advantages of LPE are the simplicity of the equipment used, higher deposition rates, low defect concentration, excellent control of stoichiometry, and the high purity that can be obtained. Background elemental impurities are eliminated by using high-purity metals and the inherent purification process that occurs during the liquid-to-solid phase transition.

The disadvantages of the LPE include poor thickness uniformity, high surface roughness, a meltback effect, and the high growth rates that prevent the growth of multilayer structures with abrupt interfaces. Growing films as thin as a few atomic layers is therefore out of the question using LPE, which has led to the development of more advanced and complex techniques. Furthermore, only small wafers can be used with LPE, which makes it a small-scale process.

2.4.2 Vapor Phase Epitaxy

Like LPE, *vapor phase epitaxy* (VPE) is also a thermodynamic equilibrium growth process. However, unlike LPE, the VPE growth technique involves reactive compounds in their gaseous form. A VPE reactor typically consists of a hot wall quartz chamber composed of several zones serving different purposes and maintained at different temperatures using a multielement furnace, as illustrated in Figure 2.9. In the VPE growth process, the source materials are generally hotter than the substrate.

The gaseous species for the group III source materials are synthesized by reacting hydrogen chloride gas (HCl) with a melted pure metal, for example, gallium (Ga) and indium (In), contained in a small vessel. This occurs in the first zone, called the group III species synthesis zone, which is maintained at a temperature T_s

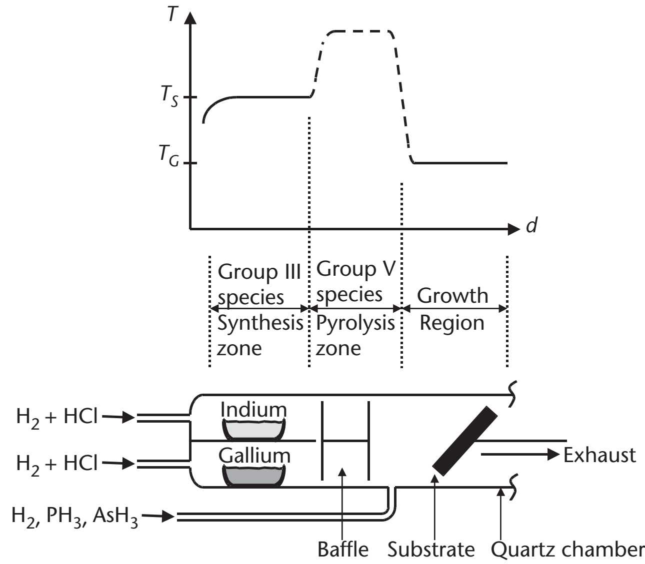
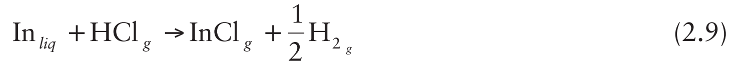
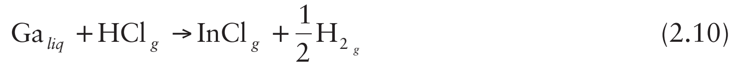


Figure 2.9 Cross section of a typical VPE reactor.

(~750–850°C for GaAs or InP growth). The reaction between the metal and HCl occurs in the following manner to form group III-chloride vapor compounds, which can be transported to the growth region:

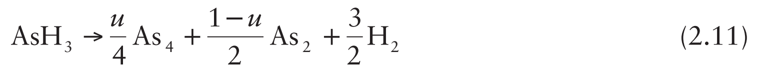


$$\mathrm{In}_{liq} + \mathrm{HCl}_g \rightarrow \mathrm{InCl}_g + \frac{1}{2}\mathrm{H}_{2_g} \tag{2.9}$$



$$\mathrm{Ga}_{liq} + \mathrm{HCl}_g \rightarrow \mathrm{InCl}_g + \frac{1}{2}\mathrm{H}_{2_g} \tag{2.10}$$

The group V source materials are provided in the form of hydrided gases, for example, arsine (AsH_3) and phosphine (PH_3). They are pyrolized in the second zone, called the group V species pyrolysis zone, which is maintained at a temperature $T > T_S$. These compounds are decomposed into elemental group V compounds, for example:



$$\mathrm{AsH}_3 \rightarrow \frac{u}{4}\mathrm{As}_4 + \frac{1-u}{2}\mathrm{As}_2 + \frac{3}{2}\mathrm{H}_2 \tag{2.11}$$

$$\mathrm{PH}_3 \rightarrow \frac{v}{4}\mathrm{P}_4 + \frac{1-v}{2}\mathrm{P}_2 + \frac{3}{2}\mathrm{H}_2 \tag{2.12}$$

where u and v represent the mole fraction of AsH_3 or PH_3, which is decomposed into As_4 or P_4, respectively.

The gas flow is then cooled by a temperature gradient that is imposed between the sources and the substrate. This cooling of reactants results in growth of the

semiconductor, such as GaAs or InP, onto the substrates in the growth region, which is maintained at a temperature T_G (~680–750°C for GaAs or InP growth).

The advantages of VPE include its low cost, high deposition rates, and the possibility of multiwafer growth. VPE also offers a high degree of flexibility in introducing dopant into the material as well as the control of the composition gradients by accurate control of the gas flows. Localized epitaxy can also be achieved using VPE.

One of VPE's main disadvantages is the difficulty of achieving multiple quantum wells or superlattices, which are periodic structures with numerous very thin layers (on the order a few tens of angstroms). Other disadvantages include the potential formation of hillocks and haze and interfacial decomposition during the preheat stage. These disadvantages have discounted this technique in the development of advanced semiconductor devices.

Recently, however, VPE has received renewed interest thanks to the development of wide bandgap gallium nitride (GaN) based semiconductors. Indeed, because of their extreme thermodynamic properties, GaN substrates cannot be achieved using the bulk crystal growth techniques discussed previously. Although other high-temperature and high-pressure techniques have been demonstrated, they are not controllable and have only yielded crystals with dimensions smaller than 1 cm. The development of GaN-based semiconductors has therefore been developed through heteroepitaxy on sapphire substrates.

Due to its high deposition rates, vapor phase epitaxy has proved successful in growing thick (>100 μm) GaN films on sapphire substrates in a reasonable time and under reasonable growth conditions. After lifting these thick films from the substrate, quasi-GaN substrates can be obtained. This is currently the most promising technology to achieve GaN (and AlGaN in the future) substrates with areas potentially comparable with those realized through bulk crystal growth techniques.

In the VPE growth of GaN, hydrogen chloride gas passes over a crucible containing metallic Ga at a high temperature (~850°C) and forms gaseous GaCl. HCl and ammonia (NH_3) are injected into the hydride pyrolysis zone using N_2 as a carrier gas. Through a showerhead, GaCl is injected into the growth zone, which is maintained between 950°C and 1,050°C, where it reacts with NH_3 on the substrate surface to produce GaN through the following reaction:

$$GaCl + NH_3 \rightarrow GaN + HCl + H_2 \quad (2.13)$$

The NH_3:HCl ratio is typically around 30:1 with a growth rate of ~0.3 μm/min.

A few disadvantages are still associated with the VPE growth of GaN. For example, NH_3 has the potential to dissociate and react with HCl to produce NCl_3, which is highly explosive. In addition, if not enough care is taken, HCl could potentially cause leaks in the reactor. $GaCl_3$ can be used instead of HCl to avoid such a problem [9]. Furthermore, undesired products of the VPE growth process such as NH_3Cl and $GaCl_3$ are produced, which can clog the exhaust line unless heated to high temperature (>150°C) or evaporated at reduced pressure. Finally, due to exchange reactions with the hot quartz walls of the reactor, AlGaN growth or *p*-type doping is difficult to realize using VPE.

2.4.3 Molecular Beam Epitaxy

Molecular beam epitaxy (MBE) is an advanced technique for the growth of thin epitaxial layers of semiconductors, metals, or insulators [10, 11]. In this method, the epitaxial growth takes place through reactions between the atomic and molecular beams of the source materials and the substrate surface, which is heated to a certain temperature in an ultrahigh vacuum environment. Depending on the nature of the precursor sources used, different variants of MBE exist. If all source materials are in solid form, the MBE process is called *solid source MBE* (SSMBE). *Gas source MBE* (GSMBE) utilizes sources in the form of gas, and, finally, *metalorganic MBE* (MOMBE) uses metalorganic material sources.

Solid precursor sources are generally solids heated above their melting points in effusion cells, also known as Knudsen cells, until atoms of the source material are able to escape the cell into the vacuum chamber by thermionic emission. The beam flux of the source materials is a function of its vapor pressure and can thus be controlled by its temperature. Gases can also be used as potential precursor sources, generally for group V elements in the synthesis of III-V compounds, and are connected through an injector and cracker. Its molecular beam flux can be controlled using a mass flow controller. Finally, metalorganic precursor sources are either liquids or fine solids with a properly controlled vapor pressure. By flowing a controlled amount of inert carrier gas through the liquid/solid, the vapor of the metalorganic compound is collected and a controlled molecular beam flux ensues.

The thickness, composition, and other properties of the epitaxial layers and heterostructures are directly controlled by the interruption of the unwanted atomic beams with specially designed shutters. A computer remotely operates the shutter controls. The typical rate of growth with MBE is around a single monolayer per second. Although slow, this allows for abrupt changes in material composition. Under appropriate conditions, the beam of atoms and molecules will attach to the substrate material and an epitaxial layer will begin to form.

The epitaxial layers crystallize through a reaction between the atomic beams of the source materials and the heated substrate surface. The thickness, composition, and doping level of the epilayer can be very precisely controlled via an accurate control of the atomic beam fluxes. The substrate is mounted on a block and rotated continuously to promote uniform crystal growth on its surface.

A schematic diagram of an MBE reactor is shown in Figure 2.10. Generally, an MBE reactor consists of three vacuum sections. Preparation and storage of the wafers is done in the buffer section. Loading and unloading of the samples into and out of the growth chamber is done in the load lock. Samples are loaded onto a rotational magnetic holder using a process known as *continual azimuthal rotation* (CAR). Cryopanels are used in conjunction with the vacuum system to keep the partial pressure of undesirable gases such as CO_2 and H_2O around 10^{-11} Torr.

The major difference between MBE and other epitaxial growth techniques stems from the fact that the growth is carried in an ultrahigh vacuum environment. Therefore, the growth is far from thermodynamic equilibrium conditions and is mainly governed by the kinetics of the surface processes. This is in contrast to the other growth techniques, such as LPE and VPE, in which the growth condition is near the

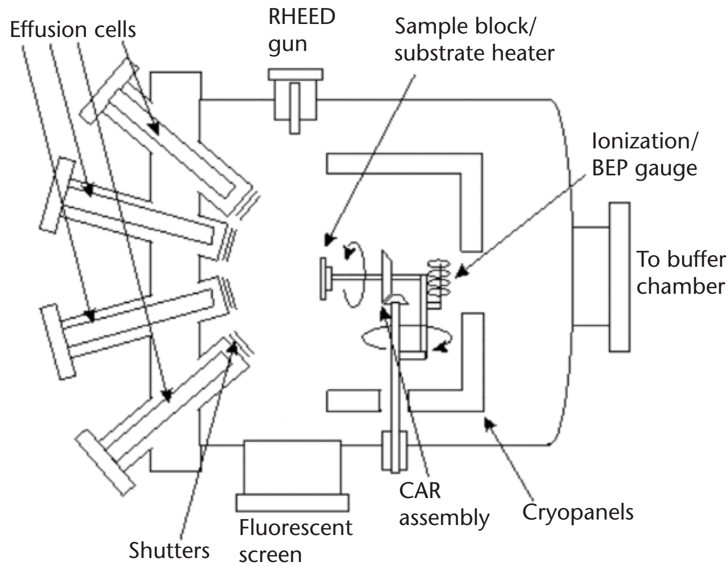


Figure 2.10 Schematic diagram of an MBE system.

thermodynamic equilibrium and is controlled primarily by diffusion processes near the surface of the substrate. The most important processes in MBE growth occur at the atomic level in the crystallization zone and can be summarized into four fundamental steps, as illustrated in Figure 2.11: (a) the adsorption of the constituent atoms or molecules impinging on the substrate surface; (b) the surface migration and dissociation of the absorbed species; (c) the incorporation of the constituent atoms into the crystal lattice of the substrate or the epilayer, at a site where sufficiently strong bonding exists; that site is usually at the edge of a spreading atomic layer, the growing epitaxial crystal; and (d) the thermal desorption of the species not incorporated into the crystal lattice.

The atoms impinging on the substrate surface must be allowed sufficient time to reach their proper position at the step edge before an entire new layer comes down and buries them. Otherwise, we would get a very rough surface with mountain-like and valley-like features on it. Worse yet, the crystal could actually end up with defects, such as missing atoms at sites in the crystal structure that would result in undesirable electrical properties.

Within the ultrahigh vacuum, the atoms in the chamber have a long mean free path, and collisions with other atoms are infrequent before reaching the substrate. Atoms from the sources are thus able to travel in a straight line until they collide with the substrate material. The mean free path L of an atom is related to the concentration n of this species and its atomic or molecular diameter d through the relation:

$$L = \frac{1}{\sqrt{2}\pi n d^2} \tag{2.14}$$

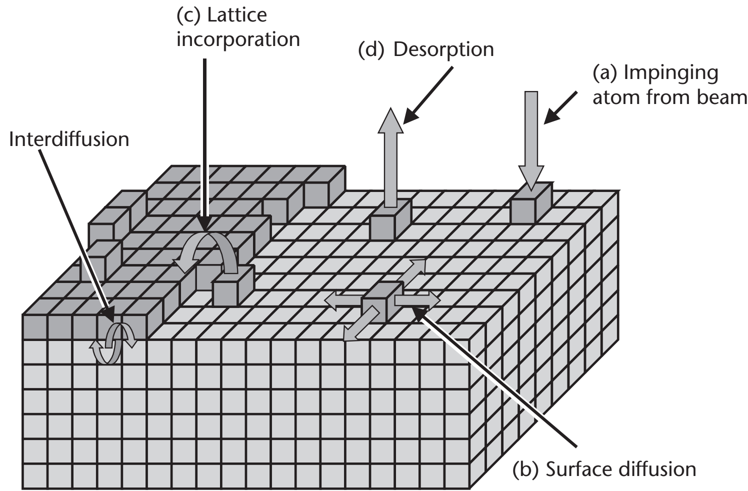


Figure 2.11 Schematic illustration of the surface processes during MBE epitaxial growth.

The concentration n is determined by the pressure P and temperature T in the MBE:

$$n = \frac{P}{k_b T} \tag{2.15}$$

where k_b is the Boltzmann constant. Under ultrahigh vacuum conditions (10^{-10}–10^{-11} Torr), the mean free path of gases is found to be several orders of magnitude higher than the distance between source and substrate (~10–20 cm).

Thanks to its ultrahigh vacuum growth environment, one of the primary advantages of MBE systems is the ability to use advanced *in situ* characterization tools such as *reflection high-energy electron diffraction* (RHEED), *auger electron spectroscopy* (AES), X-ray photoelectron spectroscopy, low-energy electron diffraction, secondary-ion mass spectroscopy, and ellipsometry, in order to monitor the film growth process.

In a RHEED system, a beam of electrons with energies in the range of 5 to 50 keV is directed on the substrate at a grazing angle θ as shown in Figure 2.12. The electrons are then diffracted by the epitaxial wafer surface, which leads to the appearance of intensity-modulated streaks on a fluorescent screen. What is observed is called a RHEED pattern.

There are two types of RHEED characterization: static and dynamic. In the first type, the atomic construction of the surface can be determined from the RHEED diffraction pattern. Such information is of particular interest since the atomic surface construction is a function of the flux of the incoming electron beam, the substrate temperature, and the strain of the epilayer. Dynamic RHEED is based on the change of the intensity of the main (central) diffraction streak as the wafer surface roughness changes over time. Indeed, during the epitaxial growth process, starting from an atomically flat surface, the roughness of the epitaxial layer increases as a new

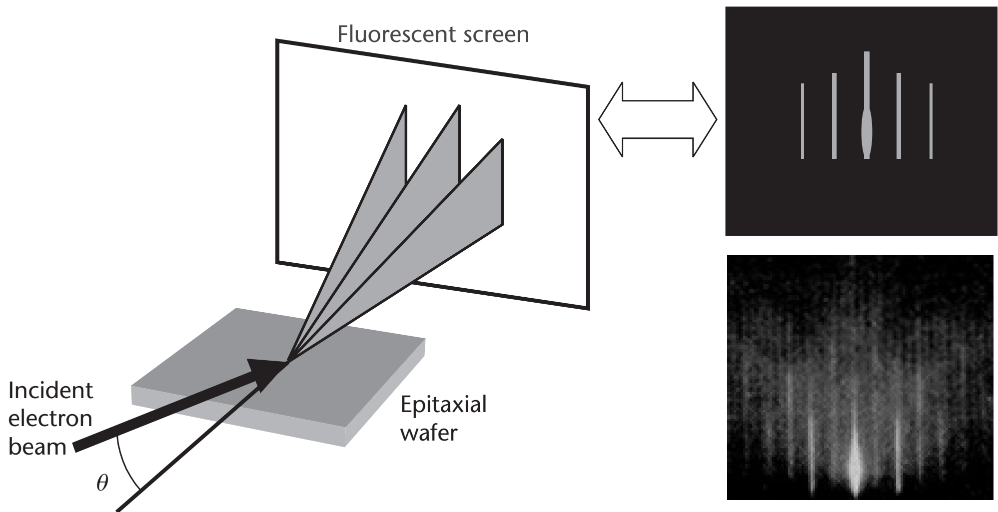


Figure 2.12 Schematic diagram of the geometrical configuration of RHEED measurements. Note that is usually less than 1°.

atomic layer nucleates. Once the surface coverage reaches 50%, the roughness is maximal and will start to decrease as the growing layer is filled. Once the new layer is completed, the roughness is minimal and will start to increase again. The intensity of the main RHEED streak thus follows this periodic oscillating pattern during growth, with the maximal intensity corresponding to the minimal roughness. The time separation between two adjacent peaks yields the time required for the growth of a single layer of the crystal. This is a powerful method that provides an accurate thickness calibration technique that is sensitive to within one single atomic layer.

Auger electron spectroscopy is another *in situ* surface monitoring technique that can be used during the MBE epitaxial growth process to study surface stoichiometry. This technique utilizes the Auger effect to measure the elemental composition of surfaces: a beam of energetic electrons, 3 to 25 keV, is used to excite surface atoms by knocking a core-level electron into higher orbitals. When relaxing into equilibrium, the atoms release their extra energy by emitting Auger electrons with characteristic energies. This energy is measured and the quantity of Auger electrons is proportional to the concentration of the atoms on the surface. In addition to measuring the two-dimensional distribution of elements on a surface, AES can also realize elemental depth profiles when it is accompanied by ion sputtering.

Other advantages of MBE over other epitaxial growth techniques include excellent thickness control and low growth temperatures. The latter reduces the diffusive rearrangement of dopants and semiconductor constituent atoms, and thus reduces the blurring of doping and composition profiles across interfaces.

In spite of its technological advantages, MBE suffers from the high costs associated with maintaining the ultrahigh vacuum environment. In addition, technological challenges remain, such as increasing the growth rate, which remains rather

slow; alleviating the difficulty associated with growing phosphorous-bearing alloys such as InP and InGaAsP; and alloy composition control.

2.4.4 Metalorganic Chemical Vapor Deposition

Metalorganic chemical vapor deposition (MOCVD) has become one of the most widely used techniques for the epitaxial growth of advanced semiconductors thin films and devices at the commercial scale. The technology has now established its ability to produce high-quality epitaxial layers and sharp interfaces, and to grow multilayer structures with thicknesses as thin as a few atomic layers, especially for III-V compound semiconductors [12–14].

The MOCVD growth process is based on the pyrolysis of alkyls or metalorganics (of group III elements typically) in an atmosphere of hydrides (of group V elements typically). The controlled amounts of volatile compounds of alkyls and hydride gases are introduced into a reaction chamber in which a semiconductor substrate is placed on a heated susceptor. The latter has a catalytic effect on the decomposition of the gaseous products, such that the semiconductor crystal growth takes place in this hot region. By contrast to VPE, the substrate is hotter than the precursor sources in MOCVD. A schematic diagram of an MOCVD reactor is shown in Figure 2.13, which depicts the gas handling system and the reactor chamber [15].

The gas handling system includes the alkyl and hydride sources and the valves, pumps, and other instruments necessary to control the gas flows and mixtures. Hydrogen (H_2), nitrogen (N_2), argon (Ar), and helium (He) are the most common inert carrier gases used in the MOCVD growth process.

The alkyl sources are metalorganic (or organometallic) compounds that are liquids or finely crushed solids usually contained in a stainless steel cylinder called a *bubbler*. The partial pressure of the source is regulated by precise control of the temperature and total pressure inside the bubbler. Electronic mass flow controllers are used to accurately and reliably measure and/or control the mass flow rate of hydride and carrier gases through the gas handling system. Thus, by sending a controlled flow of carrier gas through the bubbler, a controlled mass flow in the form of dilute vapors of the metalorganic compounds can be achieved. The purity of the sources is one of the most important issues in modern semiconductor technology. Constant effort is devoted to purifying every source material used in order to avoid any kind of contamination. Gas purifiers are often used to further purify hydride sources and carrier gases.

The mixing of volatile alkyl and hydride compounds in the gas handling system is achieved within a manifold that first stabilizes the flows, then mixes them and selectively directs them either to the reaction chamber or into the vent (waste). The manifold is designed to uniformly mix metalorganic and hydride sources prior to reaching the growth zone.

The reactor chamber is usually made of quartz or stainless steel and contains the susceptor on which the substrate wafer is resting. The susceptor can be heated using one of the following three methods: RF induction heating, radiative (lamp) heating, or resistance heating. The shape of the reactor chamber is carefully designed and

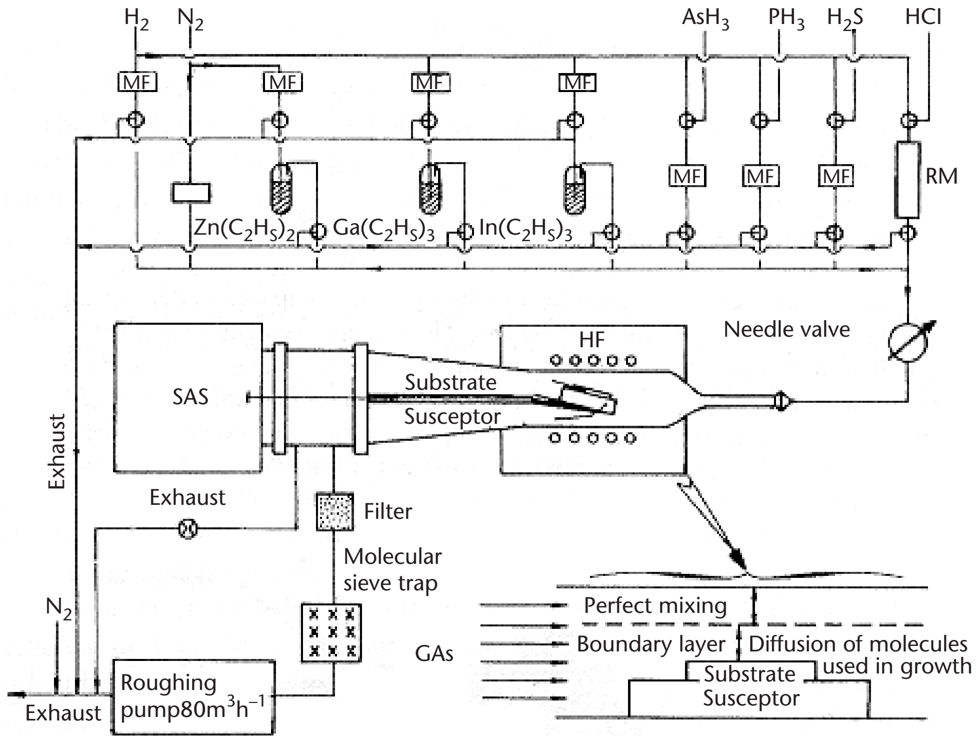


Figure 2.13 Schematic diagram of typical low-pressure MOCVD reactor.

engineered to eliminate the development of vortices and dead volumes. The growth parameters (e.g., pressure, temperature, and total gas flow) are chosen such that a laminar flow free of convection is realized. This is generally easier to do by operating at low pressure. By doing so, one ensures that a stable, reproducible, and uniform growth process is achieved.

Two types of fundamental processes occur during crystal growth on the substrate surface: thermodynamic and kinetic processes. Thermodynamic processes determine the driving force for the overall growth process, whereas kinetic processes define the rates at which the various processes occur. Hydrodynamics and mass transport, which take into account the gas velocities and temperature gradients in the vicinity of the hot susceptor, control the rate of transport of material to the growing solid/vapor interface. The rates of the chemical reactions occurring during growth, either homogeneously in the gas phase, or heterogeneously at the substrate surface, also play a role. Each of these factors will dominate some aspect of the overall growth process. A study of the dependence of a macroscopic quantity, such as growth rate, on external parameters, such as substrate temperature and precursor source flow rates, gives insight into the overall growth mechanism.

Thermodynamic calculations are useful in obtaining information about the solid composition of a multicomponent system when vapor phase compositions are known. They are also useful in obtaining the phase diagram of a multicomponent

system by calculating the compositions of the crystal for different temperatures and pressures. However, the MOCVD process is by definition not an equilibrium process. Thermodynamics can thus only define certain limits for the MOCVD growth process and is unable to provide any information about the time required to attain equilibrium, the actual steps involved in the pursuit of the lowest energy state, or the rates of the various processes occurring during the transition from the initial input gases to the final semiconductor solid. These problems can only be approached in terms of kinetics.

A much simplified description of the MOCVD growth process for III-V compounds, such as the growth of GaAs using trimethylgallium ($Ga(CH_3)_3$) and AsH_3, occurring near and at the substrate surface is illustrated in Figure 2.14. Several processes are involved, including (1) the transport of reactants through the boundary layer to the substrate, (2) the adsorption of reactants at the substrate, (3) the atomic and/or molecular surface diffusion and chemical reactions, (4) the incorporation of GaAs into the lattice, and (5) the transport of by-products away from the substrate. The typical reactants and growth temperatures for a few III-V compound semiconductors are summarized in Table 2.2.

Although the MOCVD growth process cannot usually accommodate as many *in situ* characterization techniques as can the MBE process, recent advances in the design and manufacture of MOCVD growth equipment have led to a few viable techniques. One of the pioneering works in this area was done in the late 1980s and consisted of conducting reflectance difference spectroscopy measurements during epitaxial growth. This technique consisted of conducting near-normal incidence ellipsometry measurements on the growing epitaxial layer, which yielded the relative difference of the two reflection coefficients for polarizations along the <011> and <01–1> crystal axes. This method was found to be very sensitive to any transient fluxes at the substrate surface and could be used to optimize the gas switching procedures and any heterojunction growth processes.

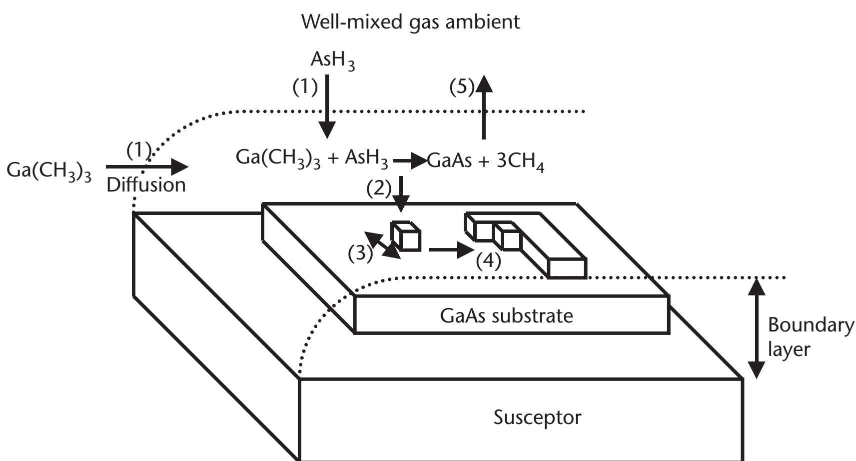


Figure 2.14 Illustration of the various processes involved in the MOCVD growth of GaAs from AsH_3 and $Ga(CH_3)_3$.

Table 2.2 Reactants and Growth Temperatures for Some of the III-V Semiconductors Grown by the MOCVD Process

Compound	*Reactants*	*Growth Temperature (°C)*
GaAs	TMGa + AsH_3	650–700
GaP	TMGa + PH_3	750
GaN	TMGa + NH_3	1,000–1,050
GaSb	TEGa + TMSb	500–550
AlAs	TMAl + AsH_3	700
AlN	TMAl + NH_3	1,100–1,200
InAs	TEIn + AsH_3	650–700
InP	TEIn + PH_3	725

Most current techniques now make use of some sort of optical interferometry, using monochromatic light with a photon energy lower than the bandgap energy of the growing semiconductor, in order to assess the surface condition, that is, the roughness of the growing epitaxial layer. By doing so, one can instantaneously estimate the atomic layer growth rate in a manner similar to how it is done in the dynamic RHEED method in MBE.

The MOCVD growth technique has proved advantageous in terms of multiple-heterostructure growth control, high versatility, high uniformity of composition, material morphology, sharp interfaces, and the ability to control solid composition while maintaining good lattice matching. Even for lattice-mismatched material systems such as III-nitrides, this method has been proven to produce high-quality layers. For example, growth of AlN on sapphire substrate has been successfully conducted despite an ~13% lattice mismatch between the AlN epitaxial layer and sapphire (Figure 2.15). The ability to grow high-quality epilayers has in turn led to the realization of an increasingly large number of high-performance devices, both in electronics and optoelectronics applications. MOCVD is also one of the major techniques used in industry, because modern equipment is capable of yielding the required high industrial throughput. The large lattice mismatch between sapphire and AlN induces a dislocated interface with the thickness on the order of 1 *monolayer* (ML), after which the AlN epilayer assumes its own lattice parameter.

However, MOCVD still suffers from the highly toxic, flammable, pyrophoric, and corrosive nature of the reactants (such as arsine and phosphine) and by-products. Like any other *chemical vapor deposition* (CVD) process, the high temperatures needed to decompose a molecule in MOCVD sometimes lead to substantial diffusive rearrangement of both dopants and the semiconductor species in heterostructures. This rearrangement produces a blurring of the intended composition profiles or even the out-diffusion of species from the back of the wafer into the vapor and back into the growing epilayer (autodoping). Furthermore, in MOCVD, one controls what enters the reactor but not what arrives at the growing

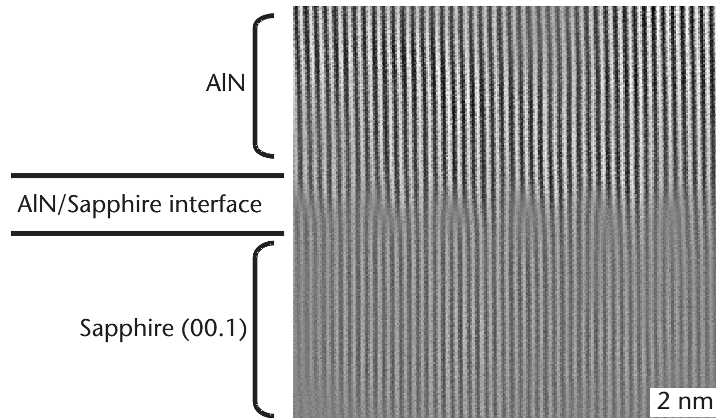


Figure 2.15 High-resolution TEM image of AlN/sapphire interface grown by MOCVD. The large lattice mismatch between sapphire and AlN induces a dislocated interface with the thickness on the order of 1 mL, after which the AlN epilayer assumes its own lattice parameter.

semiconductor surface. A change in the incoming gas chemistry does not produce a corresponding change at the surface. A delay and a time averaging will result.

Thermodynamics of CVD

The most important information we can get out of thermodynamics is the feasibility of a chemical reaction. Thermodynamics can also provide some information on the partial pressures of the involved precursors. To use thermodynamics, we should assume that the system under consideration is at its equilibrium state, which is not the case for a reactor through which gases flow in and out continuously. Therefore, we should expect thermodynamics to merely give us some guidelines when it comes to CVD. The assumption of having an equilibrium state can also be translated into the approximation of the free-energy change ΔG by the *standard* free-energy change ΔG^0.

Feasibility of Reactions

Let us see how thermodynamics can help us find out about feasibility of a chemical reaction. Table 2.3 includes several reactions with different free-energy changes. This table shows that oxidation and nitridation of silane are favorable reactions and cannot be reversed, because ΔG is a strongly negative value. Decomposition of silane, however, can be reversible as the reaction has a small value for free-energy change and, in fact, by adding small amounts of chlorine the reaction will go the other way. Deposition of TiN is not thermodynamically favorable at room temperature, however the reaction can take place at slightly higher temperatures (ΔG is a small positive value). As for deposition of Ti metal, the value of free-energy change is much higher; therefore, much higher temperatures (in excess of 1,000°C) are required for deposition of Ti.

Table 2.3 Free-Energy Change and Classification of Some Reactions

Reactants	*Products*	*ΔG (kJ/mol)*	*Classification*
$SiH_4 + 2O_2$	$SiO_2 + 2H_2O$	–1,307	Highly favorable, highly irreversible
$2SiH_4 + 4NH_3$	$Si_3N_4 + 12H_2$	–742	Favorable, irreversible
SiH_4	$Si + 2H_2$	–57	Moderately favorable, can be reversible
$TiCl_4 + 2NH_3$	$TiN + 4HCl + H_2$	92	Not favorable, possible at elevated temperatures
$TiCl_4 + 2H_2$	$Ti + 4HCl$	287	Not favorable, possible only at very high temperatures

2.4.5 Atomic Layer Epitaxy

Atomic layer epitaxy (ALE) is a growth technique that is implemented primarily during MOCVD process. It is an attractive method because it allows the digital control of the growth rate at a monolayer scale. Indeed, during ALE, the precursor source materials are alternatively injected onto the substrate in the reactor. As a result, gas phase mixing and homogeneous chemical reactions of source materials commonly found in MOCVD are suppressed and the growth reaction occurs only on the substrate surface. Therefore, the film thickness can be controlled with a single atomic layer accuracy. Furthermore, the ALE process exhibits self-limitation, that is, the layer thickness per cycle is independent of subtle variations of growth parameters. The growth rate is only dependent on the number of growth cycles and the lattice constant of the deposited material.

ALE is in fact a particular case of a self-limiting process that takes place in the gas phase. Other types of self-limiting growth processes exist, but they use ionic species reactants in solution, in which case the methods are known as *successive ionic layer adsorption and reaction* (SILAR) or *electrochemical ALE* (ECALE).

2.5 Thin-Film Deposition Techniques

2.5.1 Plasma-Enhanced Chemical Vapor Deposition

Plasma-enhanced chemical vapor deposition (PECVD) is a thin-film deposition technique that uses plasma to promote the decomposition of precursor sources into reactive species. By doing so, the deposition temperature can be lowered, which thus reduces thermal damage to the semiconductor wafer or device. Today, this technique is most often used for the deposition of dielectric films, even though other types of materials can also be deposited, including metals and semiconductors.

Figure 2.16 shows a schematic diagram of a commonly used PECVD reactor chamber. It is a cold wall parallel-plate reactor inside a cylindrical aluminum chamber. The chamber is maintained at low pressure through the use of vacuum pumps, except when loading and unloading wafers. The top plate or electrode is a showerhead through which gases are injected. The wafers are resting on the bottom electrode plate. The high side of the RF power supply is connected to the top electrode

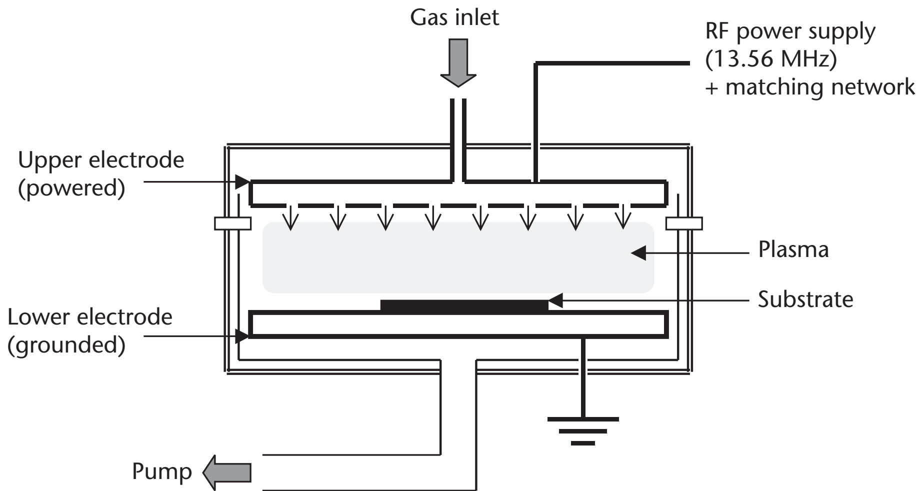


Figure 2.16 Schematic diagram of a parallel-plate PECVD reactor.

and is operated at a typical frequency of 13.56 MHz. The bottom electrode is grounded [16].

Most of the energy needed for the decomposition reaction is supplied by high-energy electrons created inside a glow discharge (plasma). The plasma is created by applying a strong oscillating electric field across the gas mixture between the parallel electrodes, which dissociates and ionizes the gas molecules, thus generating the plasma. The plasma contains electrons, gas molecules, ions, and free radicals of high energy. During each cycle, electrons and ions gain high-kinetic energies; however, this energy is much higher for electrons due to their light mass. Therefore, the effective electron temperature can reach high values (10^4–10^5 K) while ions remain at much lower temperature (100–400°C). In this way the source gas molecules are excited to higher energy states, primarily by inelastic collisions with the energetic electrons, and dissociate into a variety of radicals, ions, atoms, and more electrons. An *avalanche effect* results that continues until a steady-state plasma is established [17].

Radicals and atoms generated in the plasma travel to the substrate surface through a gas phase diffusion process. Many of these radicals undergo secondary reactions during their diffusion toward the substrate, mainly with parent molecules. Upon arrival, they are adsorbed onto the surface where some then diffuse on the growing film surface and make chemical bonds at favorable sites to form an amorphous network. Others are desorbed depending on their respective sticking coefficients.

The properties of the thin films deposited by this technique depend on various parameters such as electrode configuration, power, frequency, gas composition, pressure, flow rate, and substrate temperature.

The PECVD deposition process can be greatly enhanced through the use of microwave plasmas in which microwave energy is coupled to the natural resonant

frequency of the electrons in the presence of a magnetic field. This is then called an *electron cyclotron resonance* (ECR) process.

RF plasmas usually have a charge density of the order of 10^{10} cm^{-3} at a pressure of about 10^{-2} to 1 Torr. By contrast, the ECR discharge is easily generated at much lower pressures (10^{-5}–10^{-3} Torr) and has a degree of ionization that is 1,000 times as high as in an RF plasma. In addition, the controllability of the ion energy, high deposition rates, and the absence of source contamination have made ECR plasmas popular not only for film deposition but also for the etching process.

2.5.2 Vacuum Evaporation

Vacuum evaporation is a simple thin-film deposition technique that is generally used for metal or dielectric thin films on a semiconductor wafer or device. The source material consists of a heated metal filament/wire or fine crushed solids that are heated above their melting points in a high vacuum chamber. The evaporated atoms travel the distance between the source and the substrate and deposit on the wafer surface [18].

Figure 2.17 shows a typical vacuum deposition system. It consists of a vacuum chamber, maintained under high vacuum (10^{-3}–10^{-7} Torr) by a pumping system. The shape of the chamber is generally a bell jar that is made of quartz or stainless steel, inside of which many components are located, including the metal or dielectric material sources, a wafer holder, a shutter, a thickness rate monitor, heaters, and an ion gauge to monitor the chamber pressure. The precursor sources are located at the bottom of the chamber, whereas the wafer holder is placed on the top with the

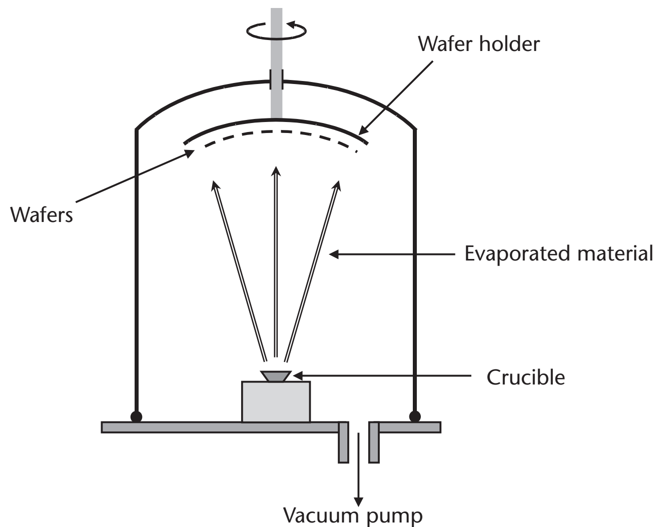


Figure 2.17 Schematic cross section of a typical vacuum evaporation system, which includes a glass bell jar under vacuum, a sample holder, a metal filament (evaporation source), and a thickness monitor.

wafers facing down. The wafer holder may be rotated to achieve a uniform deposition.

It is important to operate at a reduced pressure for a number of reasons. First, reducing the pressure ensures that the concentration of residual oxygen molecules is small enough to minimize the oxidation reaction of metals (e.g., aluminum). Secondly, the coating uniformity is enhanced at a higher vacuum. Indeed, under these conditions, the mean free path of the evaporated atoms is increased. When the mean free path exceeds the dimensions of the chamber, this ensures that the metal atoms will strike the wafers before hitting another atom, which could cause nonuniform depositions.

Three different types of evaporation techniques are used, depending on the method used to physically evaporate the metal from its solid state: filament evaporation, electron beam evaporation, and flash hot plate.

The *filament evaporation* method is the simplest of these methods and is applicable only for metals. The metal can be in the form of a wire wrapped around a coiled tungsten that can sustain high temperatures and current. The metal can also be stored in tungsten boats if large quantities of material are required. An electrical current is passed through the tungsten boat, thus heating and melting the metal into a liquid which then evaporates into the chamber at low pressure. Filament evaporation is not very controllable due to temperature variations along the filament. Another drawback when using filaments is that the source material can be easily contaminated and the contaminants can subsequently be evaporated onto the wafers. Moreover, mixtures of metal alloys containing, for example, titanium, platinum, nickel, and gold are difficult to achieve using the filament evaporation method because each metal has a different evaporation rate at a given temperature.

To avoid such problems, the *electron beam evaporation* technique was developed. Figure 2.18 is an illustration of the principle of an electron beam source, which consists of a copper holder or crucible with a center cavity that contains the metal material. A beam of electrons is generated and bent by a magnet flux so that it strikes the center of the charge cavity, as shown in Figure 2.18. In addition, the solid metal within the crucible is heated to its melting point such that it presents a smooth and uniform surface where the electron beam hits, thus ensuring that the deposition on the wafer is uniform. The crucible is cooled with water to maintain the edges of the metal in a solid state. Electron beam evaporation is relatively controlled for a variety of metals such as aluminum and gold. This technique can also be used to deposit dielectric materials. This method had its own limitations: It can only evaporate one alloy at a time. But, over the years electron beam systems have incorporated multiple guns so that each material will have its own electron beam.

The *flash hot plate* method uses a fine wire as the source material. This fine wire that contains an alloy material is fed automatically onto a hot plate surface. On contact, the tip of the wire melts and the material "flashes" into a vapor and coats the wafers in the chamber. Because all of the elements are flashed simultaneously, the composition of the metal film deposited on the wafer is close to the alloy composition of the wire.

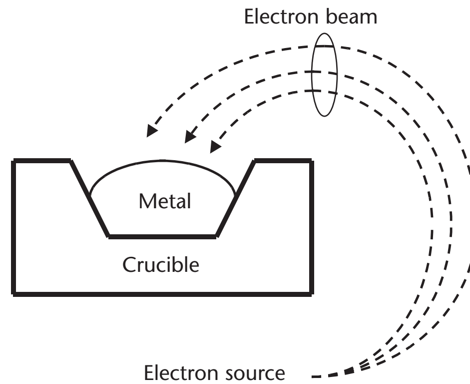


Figure 2.18 Schematic diagram of an electron beam evaporation source in which electrons are generated by a high-temperature filament and accelerated into an electron beam.

2.5.3 Sputtering

Sputtering is another vacuum deposition technique that is primarily used to deposit dielectric and metal thin films. A typical sputtering deposition system is shown in Figure 2.19. It contains a slab or target of the desired metal material, which is electrically grounded and serves as the cathode. Electrons are accelerated between two metal plates holding wafers (anode) and the cathode target. Under vacuum conditions, argon gas is introduced into the chamber and is ionized into positively charged ions through collisions with the electrons to generate secondary electrons that acquire enough energy to ionize new argon atoms.

This process thus generates a plasma inside the chamber. Argon is usually used for sputtering because of its inert nature and high purity. The ions are then

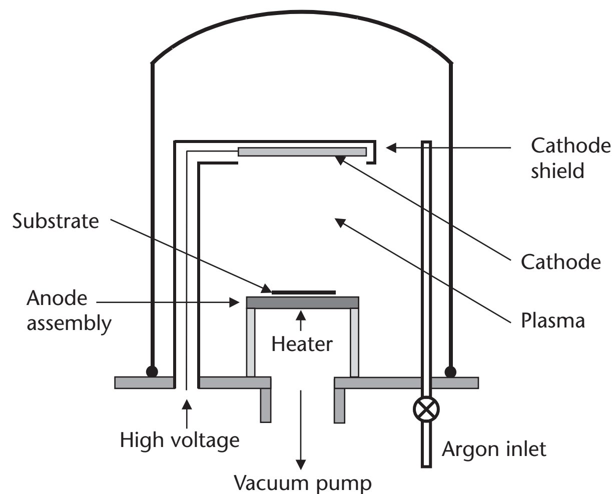


Figure 2.19 Schematic cross section of a typical sputtering system.

accelerated toward the cathode target. By impacting the cathode target, enough target atoms are sputtered such that they deposit onto the wafer surface that faces the target. The main feature of the sputtering method is that the target material is deposited on the wafer without chemical or compositional change [19]. The vacuum inside the chamber should be chosen so that minimal collisions happen when the sputtered material moves toward the wafers, while allowing enough collisions between electrons and argon atoms to maintain the plasma. Although the deposition rate does not depend on wafer temperature, it should be heated in order to have a stable film deposited.

The process just described is called *dc sputtering*. For insulating targets, made of dielectric materials for example, an RF source is used to keep the target neutral [20]. Other types of sputtering methods include ion beam sputtering, magnetron sputtering, reactive sputter deposition, bias sputter deposition, and collimated sputter deposition.

The efficiency of the sputtering process is measured by a parameter called *sputter yield*, which represents the number of atoms sputtered from the target per incident ion. The sputter yield is proportional to the following factors: (1) the number of atoms displaced toward the substrate per primary collision, (2) the number of atomic layers that contain these atoms and contribute to sputtering, (3) the number of target atoms per unit area, and (4) the cross-section $\sigma_0 = \pi a^2$ where

$$a = \frac{a_b}{\sqrt{Z_1^{2/3} + Z_2^{2/3}}}$$

where a_b is the Bohr radius of the incident ion, and Z_1, Z_2 are the atomic numbers of the incident ion and target atom, respectively. The sputter yield usually ranges from 0.01 to 4 and increases with the mass and energy of the sputtering ion.

Sputtering has several advantages over other traditional evaporation techniques. For example, the composition of the deposited film is precisely determined by the composition of the target material, step coverage is improved, and sputtered films have a higher adhesion. As with the evaporation technique, a high-quality film can only be obtained with a clean environment (e.g., clean chamber, pure source material, and clean wafer surface).

2.6 Growth of Nanostructures

The introduction of quantum wells in the early 1970s was a turning point in the direction of research on electronic structures [21]. A quantum well is a very thin layer of a semiconductor sandwiched between two layers of another semiconductor with wider energy gaps. The motion of electrons in a quantum-well structure is bound in two directions if the thickness of the quantum-well layer is of the order of the de Broglie wavelength:

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{3kTm_{eff}}} \tag{2.16}$$

In the 1980s the interest of researchers shifted toward structures with further reduced dimensionality: one-dimensional confinement (quantum wires) [22] and zero-dimensional confinement (quantum dots). Localization of carriers in all three dimensions breaks down the classical band structure of a continuous dispersion of energy as a function of momentum. Unlike quantum wells and quantum wires, the energy level structure of quantum dots is discrete (Figure 2.20). The study of nanostructures opens a new chapter in fundamental physics. In addition, quantum dots have many applications in devices such as quantum dot lasers and photodetectors.

2.6.1 Properties and Requirements of Quantum Dot Devices

In the last few years, nanostructures have been successfully realized the use of self-organized effects during growth of strained heterostructures. The most celebrated material system in this regard is InGaAs/AlGaAs. Other material systems such as InP/InGaP, (In,Ga)Sb/GaAs, and (In,Ga,Al)N systems have also seen rapid progress in the field of nanotechnology.

QDs are expected to possess many interesting and useful properties for optoelectronic devices. It is proposed that semiconductor lasers with a QD active region will have very low, temperature-independent threshold currents [23–25] and high-frequency modulation with negligible chirping effect [26].

To demonstrate useful QD-based devices several requirements need to be fulfilled [27]:

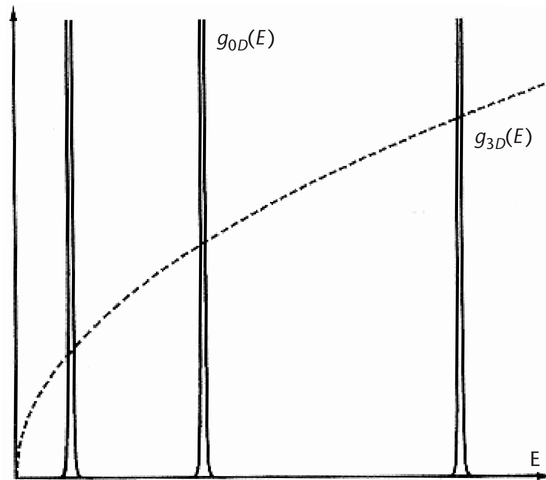


Figure 2.20 Density of states in the conduction band for a quantum dot (0D) and a bulk crystal (3D).

- Small QD size and sufficiently deep localizing potential for observation of zero-dimensional confinement,
- High density of QDs and a high filling factor,
- Low dislocation density material.

The QD size should not exceed a lower and an upper size limit dictated by the population of energy levels. The minimum QD size is the one that ensures existence of at least one energy level of an electron or a hole or both. For a spherical QD, assuming a type I heterostructure, the minimum diameter (D_{min}) is given by

$$D_{min} = \frac{\pi\hbar}{\sqrt{2m_e^* \Delta E_c}} \tag{2.17}$$

where m_e^* is the electron effective mass and ΔE_c is the conduction band offset between the two different materials used. For instance, for the InAs/AlGaAs system, the critical diameter of the QDs is ~3 to 5 nm, whereas this value is about 1 nm for the GaN/AlN system.

The maximum size of a QD is based on the thermal population of the higher lying energy levels, which is not desirable for some applications such as laser diodes. Limiting the thermal population of higher lying energy levels to 5% will result in the following condition:

$$kT \leq \frac{1}{3}\left(E_2^{QD} - E_1^{QD}\right) \tag{2.18}$$

where E_1^{QD} and E_2^{QD} are the first and second energy levels in the QD, respectively. For room temperature operation, this condition imposes an upper limit of 20 nm for InAs/AlGaAs material system if electrons are considered.

The uniformity issue arises when a device performance relies on the integrated gain in a narrow energy range, such as QD lasers. This imposes a constraint on the fluctuation in the QD size and shape. Large size and shape dispersion values result in variations in the position of energy levels, resulting in inhomogeneous broadening, which typically follows a Gaussian shape.

In addition to uniformity issues, the QD density and the fill factor are of importance. To obtain a device that provides the required performance, a certain number of QDs packed closely into a given volume are required.

Finally, as a general requirement, the density of defects in a QD material should be as low as possible in order to minimize the nonradiative processes.

2.6.2 Growth Techniques

Several methods have been used for fabrication of QD structures. Postgrowth patterning of quantum dots and transferring of the pattern to the semiconductor layer by etching was one of the earliest implemented methods [28]. This method usually suffers from low lateral resolution and etching-induced damages. Another method is selective growth of a compound semiconductor with a narrower bandgap on the

surface of another semiconductor with a wider bandgap. For the formation of quantum dots, the surface of the underlying semiconductor is patterned with a mask (SiO_2). Nucleation takes place on the area that is not covered by the mask, resulting in tetrahedral pyramids (GaAs/AlGaAs system; Figure 2.21) [29] or hexagonal pyramids (InGaN/GaN system; Figure 2.22) [30].

Probably the most useful method of formation of nanostructures is the self-organization phenomena on crystal surfaces. Strain relaxation may lead to formation of ordered arrays of quantum wires and quantum dots [31].

MBE and MOCVD are commonly used techniques for the growth of uniform arrays of 3D QDs [32]. Both lattice-matched and lattice-mismatched heteroepitaxial systems can be used for the formation of nanostructures. In lattice-matched systems, the growth mode is determined by the relation between the energies of two surfaces and the interface energy [33]. If the sum of the surface energy (γ_f) of the epitaxial layer and the energy of the interface (γ_i) is lower than the substrate surface energy (γ_s), that is, $\gamma_f + \gamma_i < \gamma_s$, upon deposition the top material wets the substrate leading to the Frank-van der Merwe growth mode [Figure 2.23(a)]. Changing the value of $\gamma_f + \gamma_i$ may result in a transition from this growth mode to the Volmer-Weber growth mode where 3D islands are formed [Figure 2.23(b)].

In a lattice-mismatched material system, such as a GaAs/InAs heterostructure with 7% lattice mismatch, only the first few deposited monolayers form epitaxial strained layers with the lattice constant equal to that of the substrate. When the critical thickness is exceeded, the significant strain occurring in the top layers leads to the spontaneous formation of randomly distributed islands. The phase transition

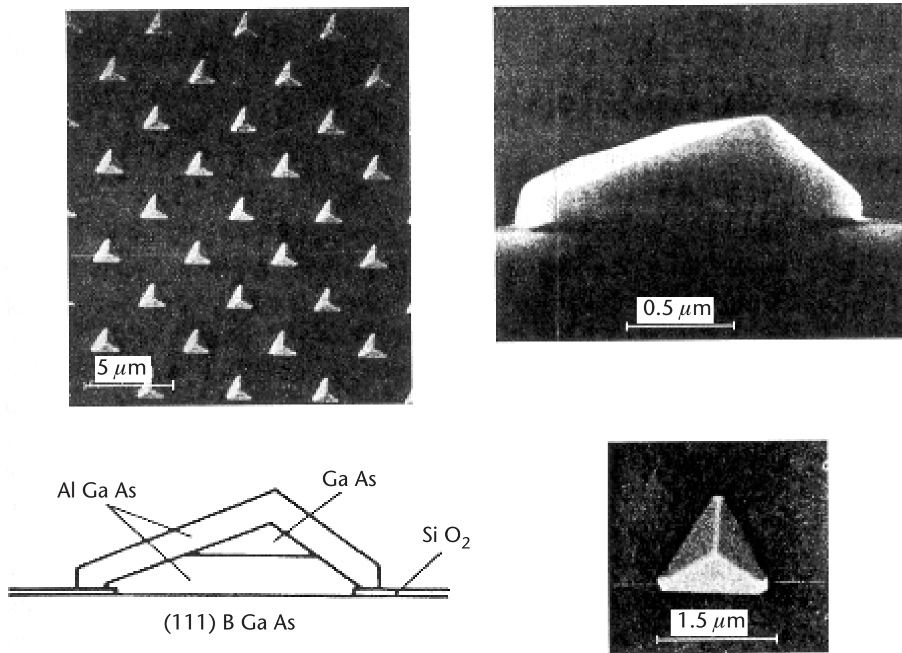


Figure 2.21 GaAs quantum dots grown on the surface of GaAs by selective MOCVD growth.

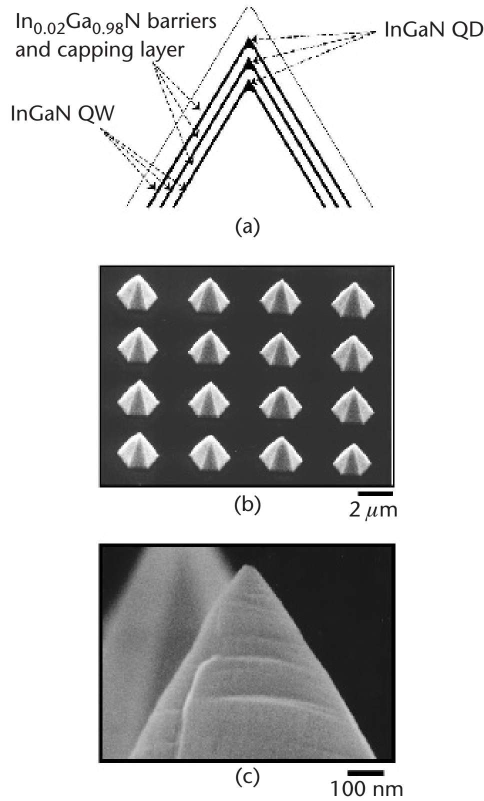


Figure 2.22 (a) Schematic diagram of InGaN quantum dots formed on top of hexagonal GaN pyramids, (b) bird's-eye view, and (c) cross section.

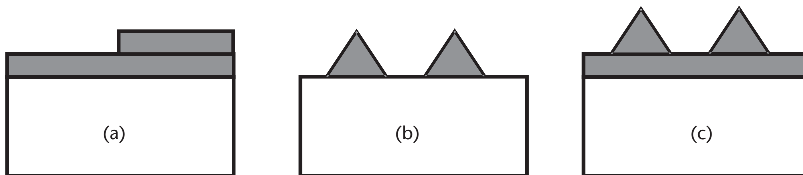


Figure 2.23 Schematic presentation of various growth modes: (a) Frank-van der Merwe, (b) Volmer-Weber, and (c) Stranski-Krastanov.

from the epitaxial structure to random arrangement of islands is called the Stranski-Krastanov transition [Figure 2.23(c)] [34].

The shape of the quantum dots can be significantly modified during regrowth or postgrowth annealing or by applying complex growth sequences. The quantum dots may assume the shape of pyramids [35] or flat circular lenses [36].

InAs/GaAs QD Growth

We consider growth of InAs QDs on GaAs substrate as a typical example of growth of nanostructures. Quantum dots form due to the strain mismatch between the MBE-grown InAs and the GaAs substrate. If a very thin InAS layer is grown on the GaAs substrate, the growth will be in the form of a two-dimensional wetting layer. The growth mode will switch to three-dimensional island growth by increasing the thickness of the InAs layer, and the density of dots will increase up to the point that the neighboring islands coalesce to form larger islands. Substrate temperature is another important parameter in the MBE growth of InAs islands. The diameter and density of QDs depend heavily on the growth temperature. Figure 2.24 shows the base diameter and density of InAs islands grown with a 2.1-ML supply of InAs as an *Atomic Force Microscope* (AFM) function of growth temperature [37]. The AFM images are also shown on the right. This temperature dependence is due to the surface migration length of InAs, which increases with increasing temperature.

Multiple-layer stacking of islands is often required when implementing QDs in the active region of optical devices to obtain enough interaction between the confined electrons and the electromagnetic field. In this case, a spacer layer determines the size and density of QDs. Below a certain thickness of the spacer (GaAs in this example), the islands align themselves according to the bottom layer and their size increases as we move toward the upper layers. Increasing the thickness of the spacer may result in independent growth of islands with nearly constant size and density in different layers. By keeping the width of the spacer below the critical thickness, one can grow closely stacked multiple layers with perpendicularly aligned islands suitable for electrical coupling in the vertical direction. In this case, the thickness of the

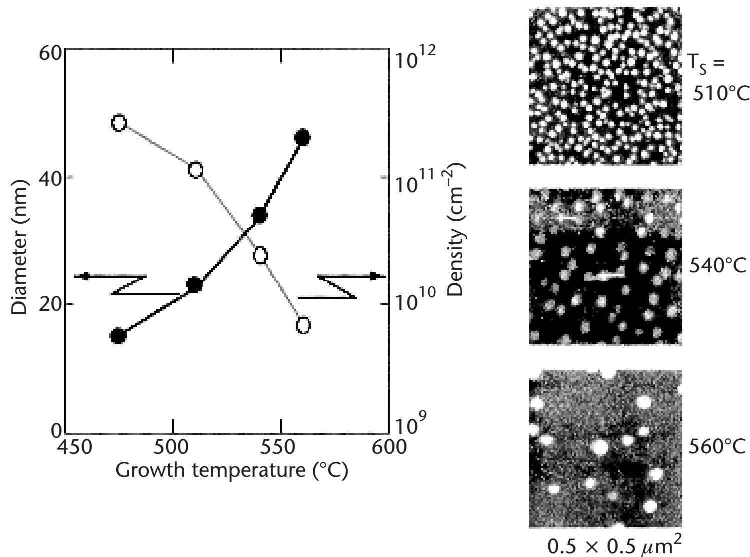


Figure 2.24 Dependence of diameter and density of QDs for a 2.1-ML InAs supply on the growth temperature and the *Atomic Force Microscope* (AFM) images of islands at each temperature.

spacer, which is chosen small enough so that perpendicular alignment can take place, determines the height of the islands. A schematic representation of this process is shown in Figure 2.25. As discussed earlier, growth of quantum dots on a given substrate in the Stranski-Krastanov growth mode begins with the two-dimensional growth of a *wetting layer*, followed by the 3D growth of islands. The shrinkage of islands (InAs) takes place during the annealing step after overgrowth of the spacer (GaAs). During this period, part of the island that resides above the spacer leaves the island to partially form the subsequent wetting layer. After this, the spacer replaces the top portion of the island to reduce the total energy, which leads to shrinkage of the island [38, 39].

MOCVD is another popular method for growth of low-dimensional structures. Various groups have reported growth of self-organized InGaAs QDs on GaAs using MOCVD [40–42]. Similarly, other material systems have also been studied for growth of ordered arrays of islands: InAlAs/AlGaAs on GaAs [43], InAs on InP [44], and InGaAs/InGaP [45].

III-Nitride QD Growth

Relatively new to the field are the III-nitride compounds consisting of (Al,Ga,In)N alloys. Self-assembled QDs in the III-nitride material system are generally grown either by using an “antisurfactant” or by taking advantage of the lattice mismatch between InGaN/GaN or GaN/AlGaN heterostructures.

In the GaN/AlGaN system, adding a third element (antisurfactant) such as TESi to the substrate material (AlGaN) can decrease the surface free energy and reverse the inequality $\gamma_f + \gamma_i > \gamma_s$ to force the 3D growth in this system [46]. In this case, the shape and density of QDs can be controlled by the Si dose, the growth temperature, the growth duration, and the Al content in the underlying AlGaN layer [46–48].

The Stranski-Krastanov growth mode has also been studied in the GaN/AlN material system. This is driven by the fact that an ~2.5% lattice mismatch exists between GaN and AlN. Using this method, a high density of dots (~10^{10} cm^{-2}) has been achieved by MOCVD [49]. The density of QDs depends on the number of GaN monolayers grown on the AlN substrate. Figure 2.26 shows this relationship. The minimum number of MLs for formation of 3D islands appears to be ~4. Similarly, it has been shown that 3D islands form above a critical thickness of 3 MLs by plasma-assisted MBE [50]. In general, the critical thickness is reached when the gain of elastic energy becomes greater than the surface free-energy term [51]. For the GaN/AlN

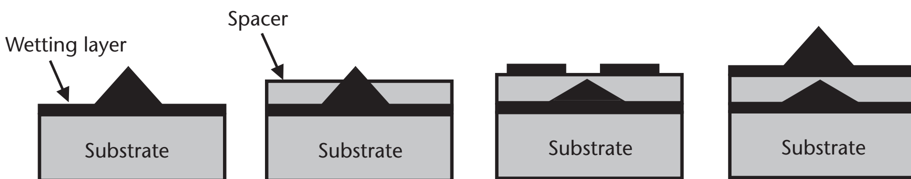


Figure 2.25 Schematic representation of the growth of closely stacked islands.

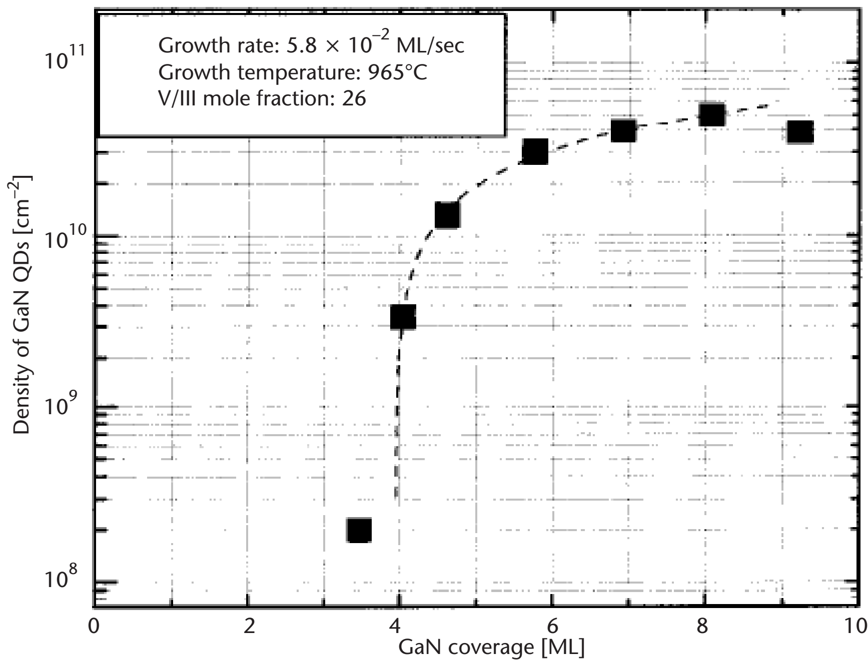


Figure 2.26 Density of self-assembled GaN QDs (cm^{-2}) on AlN as a function of GaN coverage in monolayers.

material system, usually the constraint is to grow more than 3 MLs for the island formation and less than 8 MLs to avoid the plastic relaxation of the stress [52].

The ratio of Ga/N is one of the main factors in determination of the growth mode. It has been found that growth of GaN on AlN in an N-rich environment results in the formation of GaN QDs [50]. On the other hand, the 2D/3D transition can be completely inhibited by performing the growth in a Ga-rich environment. This behavior is attributed to formation of a stable Ga film, about 2 ML thick, which acts as a *self-surfactant* [53].

References

[1] Cox, J. D., D. D. Wagman, and V. A. Medvedev, *CODATA Key Values for Thermodynamics,* New York: Hemisphere Publishing, 1989.

[2] Davydov, A. V., et al., "Thermodynamic Assessment of the Gallium-Nitrogen System," *Physica Status Solidi (a)*, Vol. 188, 2001, pp. 407–410.

[3] Czochralski, J., "Ein neues Verfahren zur Messung der Kristallisationsgeschwindigkeit der Metalle," *Z. Physik. Chem.,* Vol. 92, 1918, p. 219.

[4] Inada, T., et al., *The 2nd Asia Conference on Crystal Growth and Crystal Technology (CDCT-2)*, Seoul, Korea, 2002.

[5] Lely, J.A., *Ber. Deut. Keram.,* Vol. 32, 1955, p. 229.

[6] Tairov, Y. M., and V. F. Tsvetkov, *J. Cryst. Growth,* Vol. 43, 1978, p. 209.

[7] Han, R., et al., "Development of Bulk SiC Single Crystal Grown by Physical Vapor Transport Method," *Opt. Mat.,* Vol. 23, 2003, p. 415.

[8] Nelson, H., "Epitaxial Growth from the Liquid State and Its Applications to the Fabrication of Tunnel and Laser Diodes," *RCA Rev.,* Vol. 24, 1963, p. 603.

[9] Kuramoto, M., *IEICE Trans. Electron.*, Vol. E83-C, 2000, p. 552.

[10] Cho, A. Y., *The Technology and Physics of Molecular Beam Epitaxy,* New York: Plenum Press, 1985.

[11] Foxton, C. T., and J. J. Harris, *J. Cryst. Growth,* Vol. 81, 1984, pp. 1–4.

[12] Razeghi, M., *The MOCVD Challenge, Volume 1: A Survey of GaInAsP-InP for Photonic and Electronic Applications,* Bristol, England: Adam Hilger, 1989.

[13] Razeghi, M., *The MOCVD Challenge, Volume 2: A Survey of GaInAsP-GaAs for Photonic and Electronic Device Applications,* Bristol, England: Institute of Physics, 1995.

[14] Stringfellow, G. B., *Organometallic Vapor-Phase Epitaxy: Theory and Practice,* Boston, MA: Academic Press, 1989.

[15] Razeghi, M., *Fundamentals of Solid State Engineering,* Boston, MA: Kluwer Academic Publishers, 2002.

[16] Rsoler, R. S., W. C. Benzing, and J. Balod, "A Production Reactor for Low Temperature Plasma-Enhanced Silicon Nitride Deposition," *Solid-State Technol.*, Vol. 19, 1976, p. 45.

[17] El-Kareh, B., *Fundamentals of Semiconductor Processing Technology,* Boston, MA: Kluwer Academic Publishers, 1995.

[18] O'Hanlon, J. F., *A User's Guide to Vacuum Technology,* New York: John Wiley & Sons, 1981.

[19] Chapman, B. N., and S. Mangano, "Introduction to Sputtering," in *Handbook of Thin-Film Deposition Processes and Techniques,* K. K. Schugeraf, (ed.), Park Ridge, NJ: Noyes Publications, 1988.

[20] Chapman, B. N., *Glow Discharge Processes: Sputtering and Plasma Etching,* New York: John Wiley & Sons, 1980.

[21] Dingle, R., W. Wiegmann, and C. H. Henry, "Quantum States of Confined Carriers in Very Thin $Al_xGa_{1-x}As$/GaAs/$Al_xGa_{1-x}As$ Heterostructures," *Phys. Rev. Lett.,* Vol. 33, 1974, p. 827.

[22] Petroff, P. M., et al., "Toward Quantum Well Wires: Fabrication and Optical Properties," *Appl. Phys. Lett.*, Vol. 41, 1982, p. 635.

[23] Arakawa, Y., and H. Sakaki, "Multidimensional Quantum Well Laser and Temperature Dependence of Its Threshold Current," *Appl. Phys. Lett.,* Vol. 40, 1982, p. 939.

[24] Asada, M., Y. Miyamoto, and Y. Suematsu, "Gain and Threshold of Three-Dimensional Quantum-Box Lasers," *IEEE J. Quantum Electron.*, Vol. 22, 1986, p. 1915.

[25] Yariv, A., "Scaling Laws and Minimum Threshold Currents for Quantum-Confined Semiconductor Lasers," *Appl. Phys. Lett.,* Vol. 53, 1988, p. 1033.

[26] Arakawa, Y., K. Vahala, and A. Yariv, "Quantum Noise and Dynamics in Quantum Well and Quantum Wire Lasers," *Appl. Phys. Lett.,* Vol. 45, 1984, p. 950.

[27] Bimberg, D., M. Grundmann, and N. N. Ledentsov, *Quantum Dot Heterostructures,* New York: John Wiley & Sons, 1999.

[28] Reed, M. A., et al., "Spatial Quantization in GaAs/AlGaAs Multiple Quantum Dots," *J. Vacuum Sci. Technol. B,* Vol. 4, 1986, p. 358.

[29] Fukui, T., et al., "GaAs Tetrahedral Quantum Dot Structures Fabricated Using Selective Area Metalorganic Chemical Vapor Deposition," *Appl. Phys. Lett.*, Vol. 58, 1991, p. 2018.

[30] Tachibana, K., et al., "Selective Growth of InGaN Quantum Dot Structures and Their Microphotoluminescence at Room Temperature," *Appl. Phys. Lett.*, Vol. 76, 2000, p. 3212.

[31] Shchukin, V. A., et al., *Phys. Rev. Lett.,* Vol. 75, 1995, p. 2968.

[32] Ledentsov, N. N., et al., "Luminescence and Structural Properties of In,GaAs/GaAs Quantum Dots," *Proc. 22nd Int. Conf. Physics of Semiconductors,* D. J. Lockwood, (ed.), Vancouver, Canada, Vol. 3, 1994, p. 1855.

[33] Grundmann, M., *Nano-Optoelectronics: Concepts, Physics and Devices,* Berlin, Germany: Springer-Verlag, 2002.

[34] Starnski, I. N., and L. von Krastanov, *Akad. Wiss. Lett. Mainz Math. Natur K1. Iib.,* Vol. 146, 1939, p. 797.

[35] Grundmann, M., O. Stier, and D. Bimberg, "InAs/GaAs Pyramidal Quantum Dots: Strain Distribution, Optical Phonons and Electronic Structure," *Phys. Rev. B,* Vol. 52, 1995, p. 11969.

[36] Fafard, S., et al., "Selective Excitation of the Photoluminescence from N-Dot Ensembles and the Linewidth of Ultrasmall $Al_yIn_{1-y}As/Al_xGa_{1-x}As$ Quantum Dots," *Phys. Rev. B,* Vol. 50, 1994, p. 8086.

[37] Sugawara, M., (ed.), *Self-Assembled InGaAs/GaAs Quantum Dots, Volume 60, Semiconductors and Semimetals,* San Diego, CA: Academic Press, 1999, p. 125.

[38] Xie, Q., et al., *Appl. Phys. Lett.,* Vol. 75, 1995, p. 2542.

[39] Ledentsov, N. N., et al., *Appl. Phys. Lett.,* Vol. 69, 1996, p. 1095.

[40] Mukai, K., et al., *Jpn. J. Appl. Phys.,* Vol. 33, 1994, p. L1710.

[41] Oshinowo, J., et al., *Appl. Phys. Lett.,* Vol. 65, 1994, p. 1421.

[42] Heinrichsdorff, A., et al.,, *Appl. Phys. Lett.,* Vol. 68, 1996, p. 3284.

[43] Leon, R., et al., *Appl. Phys. Lett.,* Vol. 67, 1995, p. 521.

[44] Ponchet, A., et al., *Appl. Phys. Lett.,* Vol. 67, 1995, p. 1850.

[45] Kim, S., et al., "Growth and Characterization of InGaAs/InGaP Quantum Dots for Midinfrared Photoconductive Detector," *Appl. Phys. Lett.,* Vol. 73, 1998, p. 963.

[46] Tanaka, S., S. Iwai, and Y. Aoyagi, "Self-Assembling GaN Quantum Dots on $Al_xGa_{1-x}N$ Surfaces Using a Surfactant," *Appl. Phys. Lett.,* Vol. 69, 1996, p. 4096.

[47] Shen, X. Q., et al., *Appl. Phys. Lett.,* Vol. 72, 1998, p. 344.

[48] Ramvall, P., et al., *J. Appl. Phys.*, Vol. 87, 2000, p. 3883.

[49] Miyamura, M., K. Tachibana, and Y. Arakawa, "High-Density and Size-Controlled GaN Self-Assembled Quantum Dots Grown by Metalorganic Chemical Vapor Deposition," *Appl. Phys. Lett.,* Vol. 80, 2002, p. 3937.

[50] Widmann, F., et al., "Blue-Light Emission from GaN Self-Assembled Quantum Dots Due to Giant Piezoelectric Effect," *Phys. Rev. B,* Vol. 58, 1998, p. R15989.

[51] Snyder, C. W., et al., *Phys. Rev. Lett.,* Vol. 66, 1991, p. 3032.

[52] Gogneau, N., et al., "Structure of GaN Quantum Dots Under Modified Stranski-Krastanow Conditions on AlN," *J. Appl. Phys.,* Vol. 94, 2003, p. 2254.

[53] Mula, G., et al., *Phys. Rev. B*, Vol. 64, 2001, p. 5406.

CHAPTER 3

Quantum Dot Infrared Photodetectors

Anupam Madhukar, Eui-Tae Kim, and Zhonghui Chen, University of Southern California, and Joe Campbell and Zhengmao Ye, University of Texas at Austin

3.1 Introduction

A solid, when confined in all three spatial dimensions to length scales comparable to or smaller than the electron de Broglie wavelength, gives rise to discrete electron energy levels. As a result of the attendant discrete, ideally δ-function-like density of energy states, such structures acquire new and unique electronic and optical properties not found in solid structures confined in only one direction (i.e., quantum wells) or two directions (i.e., quantum wires). These structures are called *quantum dots* and are a particular class of semiconductors.

QDs are the defect-free 3D islands that can form spontaneously for a deposition amount beyond a critical value during growth of a compressively strained semiconductor overlayer on an appropriate substrate [1]. The critical value depends on the lattice mismatch and material combination. A variety of III-V, IV-IV, and II-VI semiconductor combinations exceeding ~2% lattice mismatch exhibit the phenomenon. For the InAs on GaAs(001) combination involving ~7% mismatch, the critical deposition amount is ~1.6 ML of In. For Ge on Si(001) involving a ~4% mismatch, it is ~3 ML. For InAs/GaAs(001), the typical size of the coherent islands is ~20 nm for the base length and 3 to 8 nm in height. An atomic force microscope image of the InAs island size distribution is shown in Figure 3.1. The exciton diameter in bulk InAs is ~70 nm. Strong electronic quantum confinement effects are thus expected and observed. Dubbed *self-assembled QDs* (SAQDs), the SAQDs have, in the past decade, represented the most extensively employed vehicle for studying the physics of zero-dimensional structures as well as for examining the potential of QDs for electronic and optoelectronic devices [1]. In this chapter we provide a brief report on the potential of SAQDs as *infrared* (IR) photodetectors in the near-IR to long-IR wavelength (1- to 20-μm) regime.

The IR detection wavelength regimes originate in the range of the interband and intraband optical transition energies afforded by the tetrahedrally bonded semiconductor SAQDs, for example, those in the (InGaAl)/As family. The schematic of Figure 3.2(a) is an idealized representation of the 3D coherent (i.e., defect-free) pyramidal islands formed during InAs deposition on GaAs(001) substrates due to the ~7% larger bulk lattice constant of the InAs overlayer. Figure 3.2(b) represents a

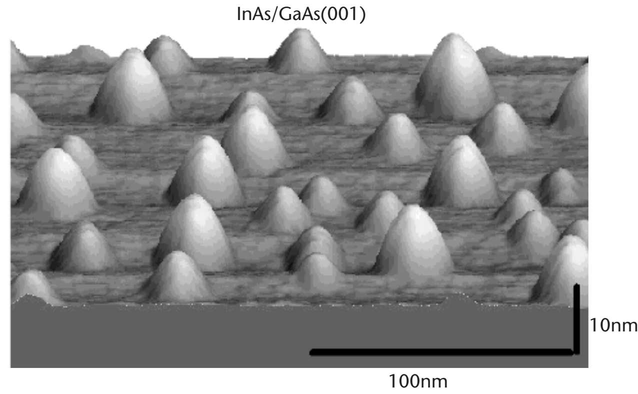


Figure 3.1 An atomic force microscope image of InAs islands on GaAs(001). Note the fluctuation in size and random spatial location.

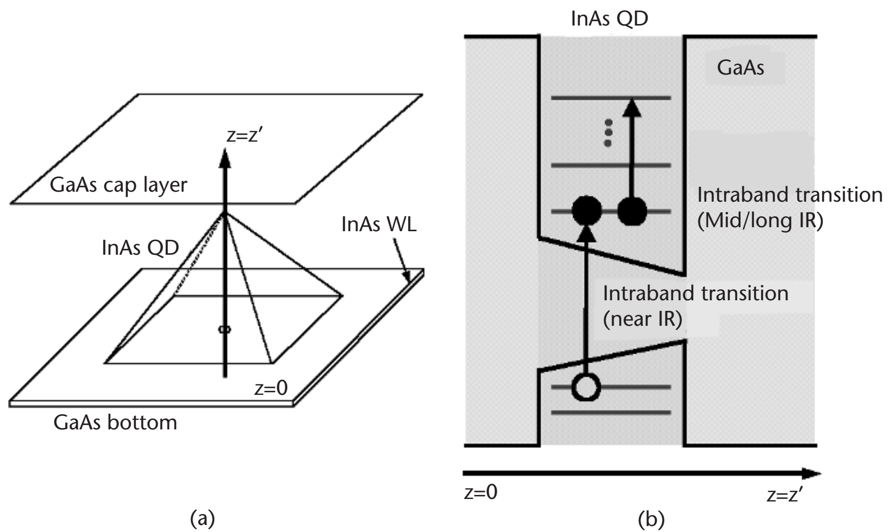


Figure 3.2 Schematic of (a) a buried pyramidal island and (b) the conduction and valence band edge lineup that defines the confinement potential along the line passing through the island base. The tapering toward the apex symbolizes lower compressive strain toward the apex. The 3D quantum confined states, arising from the full 3D confinement potential, are indicated along with the interband and intraband transitions to be tailored for detectors.

one-dimensional potential defined by the conduction and valence band edges of the GaAs/InAs/GaAs system along a vertical line passing through the pyramidal island base center and apex. In such a spatially varying potential, the electrons and holes of InAs in certain regions of the otherwise bulk band structure acquire spatially confined wave functions along with discrete energies, as symbolically indicated in Figure 3.2(b). Of course, in reality the QDs are 3D with mostly imprecisely known shapes and chemical compositions in all directions along with accompanying 3D

spatial distributions of stress (strain) fields. The actual energies and wave functions are determined by the full 3D spatially varying potential even for an idealized conceptualization of a QD, such as that depicted in Figure 3.2(b).

A variety of QD electronic structure calculations based on a range of theoretical approaches abound in the literature [2]. These provide much needed help and guidance in gaining some understanding of the experimentally inferred electronic structure, largely based on optical studies, sometimes supplemented with optoelectronic or electrical studies. For our main purpose here—SAQD-based IR photodetectors—it suffices for the moment to use the schematic of Figure 3.2 to introduce the main considerations for examining the potential of SAQDs for IR detectors.

The 3D confined nature of most of the QD bound hole and electron states in Figure 3.2 immediately reveals that, unlike *quantum well* (QW) structures [3], even normal incidence photons can have meaningful absorption cross sections owing to the lack of wave-vector selection rules. Moreover, limitations on the matching of the dominant optical phonon energies with the range of energy relaxation to the ground state following excitation of an electron to a higher state introduces a so-called "phonon bottleneck" [4, 5] and thus enhances the excited carrier lifetime. This, in turn, is helpful for collection of the carriers as photocurrent and for achieving high photoconductive gain. There are also physical considerations that, model calculations suggest [6; also private communication with Y. C. Chang et al.], make the dark current for 3D confined carriers meaningfully lower than that in, for example, 1D confined structures such as QWs. Moreover, as we discuss later, QDIPs lend themselves to the placement of the SAQDs in an appropriately designed and fabricated resonant cavity, thus allowing for considerable enhancement of the optical pathlength even in normal incidence geometry [7].

Finally, the hole-to-electron (interband) transitions [Figure 3.2(b)] in a system such as InGaAs/GaAlAs can be tuned in the near-IR region (1–1.3 μm) by appropriate design and growth of the SAQDs. The electron-to-electron (or, in principle, hole-to-hole), that is, intraband, transition-based QDIPs can be realized in the 3- to 20-μm range by appropriately doping and placing SAQDs in suitable configurations. By placing the SAQDs in a *p*-*i*(SAQD)-*n* structure, near-IR QDIPs have been realized. In this chapter we focus on QDIPs for operation in the mid-IR (3- to 6-μm) and long-IR (8- to 12-μm) regimes.

To achieve a meaningful optical interaction and hence absorption pathlength, a typical QDIP structure involves *multiple-layer QDs* (MQDs), as shown in Figure 3.3. At the core of realizing successful QDIPs is the overall trade-off between a large number of high-areal-density SAQD layers to allow maximum photon interaction while still minimizing the overall lattice strain buildup to prevent a high density of structural defects in the structure. This trade-off has guided our consideration of the potential use of "large" and "small" SAQDs, as the next two sections will reveal.

Depending on the transport direction of the photoexcited carriers, two types of device structures can be used for QDIPs. One is the conventional vertical transport scheme, in which the excited carriers move perpendicular to the growth direction [Figure 3.3(a)]. An alternative is the lateral transport scheme (parallel to the growth

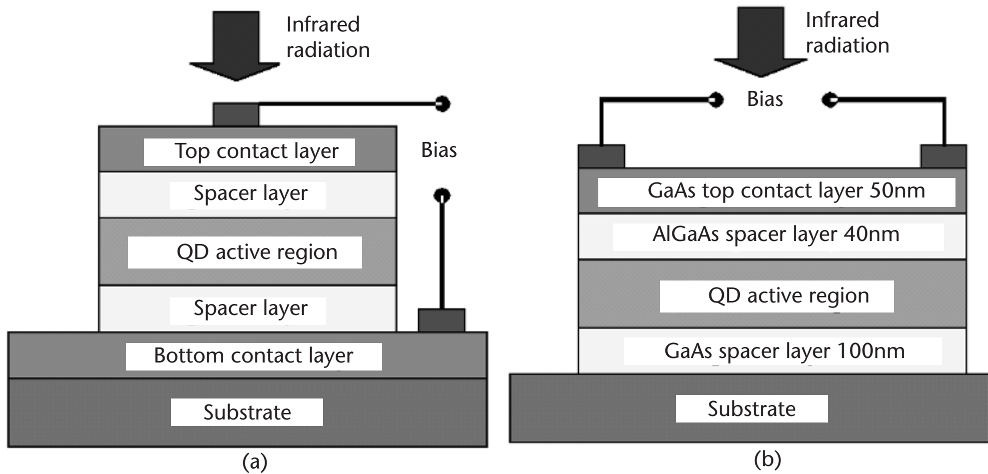


Figure 3.3 Schematics of two types of the QDIP structures: (a) a vertical transport structure and (b) a lateral transport structure.

planes), which is shown in Figure 3.3(b). The lateral transport devices operate similar to a MODFET in that photocurrent proceeds through lateral transport of carriers across a high-mobility channel between two top contacts. AlGaAs barriers are utilized to modulation-dope the quantum dots and to provide a high-mobility channel. Both structures respond to normal-incidence radiation. Because the carrier transport favors motion within the QD layers, QDIPs with lateral transport are purported to have the potential for higher responsivity [8]. Lateral transport QDIPs within a modulation-doped heterostructure have demonstrated lower dark currents and operation near room temperature [9–11] as a result of the fact that primary components of the dark current originate from interdot tunneling and hopping conduction [10]. However, the vertical transport structure is more compatible with high-density focal plane array architectures. As a result, most research to date has focused on the vertical transport QDIPs.

Our emphasis in this chapter is on a comprehensive discussion of the growth, structural and optical characterization, and device figures of merit on the *same* QDIP structures. To achieve this emphasis, we sacrifice discussion of the few different approaches to the synthesis of QDIPs (in terms of QD sizes, shapes, chemical compositions, and growth procedures) reported in the literature. We do, however, provide a fairly complete list of the papers that have been published on QDIP.

The remainder of this chapter is organized as follows: In Section 3.2 we discuss QD and QDIP structure growth, QD size distribution, and the tailoring of the QD electronic energy levels and wave functions via manipulation of the QD confinement potential. We also show how to take advantage of stress manipulation to realize multiple-color QDIPs. In Section 3.3 we focus on the QDIP device characteristics (dark current, responsivity, noise, photoconductive gain, detectivity) for each of three classes of QDIPs discussed in that section: the InAs/GaAs/AlGaAs QDIPs, the InAs/InGaAs/GaAs QDIPs, and the dual-color InAs/InGaAs/ GaAs QDIPs. We end

with Section 3.4, which provides some thoughts on continuing exploration to more fully assess the potential of QDIPs.

3.2 QD and QDIP Structure Growth and Characterization

The materials system used for our systematic studies of SAQD growth and basic photoresponse is as-deposited, binary InAs on GaAs(001) substrates. The growth technique utilized is solid source MBE. For deposition amounts of 1.6 ML, the SAQDs form spontaneously. Although their size distribution can be controlled to some degree through manipulation of the deposition conditions (e.g., substrate temperature, deposition rate, deposition amount), we have found that for the most part coherent InAs islands grow laterally up to a base length of ~20 nm, beyond which the dominant increase in size occurs through an increase in the steepness of the pyramidal sidewalls. The most probable heights are ~3.5 nm for deposition amounts up to ~2 ML and ~7.5 nm for deposition amount up to ~3 ML. Larger islands are mostly dislocated and thus not desirable. We will refer to these two most probable types of InAs/GaAs SAQDs as *small* (or *shallow*) and *large* (or *steep*) *QDs*. To create the QDIP structures, either type of 3D islands is capped with four different types of capping layer materials: (1) GaAs, (2) AlGaAs, (3) InGaAs, and (4) InGaAlAs. It is thus through the control of the SAQD's most probable size and the capping layer material that we manipulate the QD electronic structure by tailoring the 3D quantum confinement potential through the chemical composition of the barrier layer and the substrate-island-capping layer lattice mismatch-induced 3D stress distribution. In turn, this allows the design of the operating wavelength and other considerations for the QDIP.

We note that in all capping layer cases mentioned, our growth approach is fundamentally different from the general practice. We grow the capping layers using *migration-enhanced epitaxy* (MEE) at low temperatures between 350°C and 400°C. This approach should cause minimal chemical intermixing during capping, in contrast to the common practice of growing capping layers via MBE that requires the same high growth temperatures (480°C to 520°C) as those used for deposition of the island-forming layers.

Our approach provides greater confidence in comparison between photoresponse and structural parameters obtained through *atomic force microscope* (AFM) and *transmission electron microscope* (TEM) studies on identically grown but uncapped QD samples. We have reported that MEE capping at these low temperatures not only preserves the quality of the GaAs cap, but also increases the *photoluminescence* (PL) efficiency compared to samples grown identically but capped via standard MBE growth techniques [12]. For all growths, GaAs(001)$\pm$0.1° substrates are used and, after standard cleaning procedures, a ~0.5-μm-thick GaAs buffer is grown at 600°C at a growth rate of 0.5 ML/sec with short-term (~2-minute) surface annealing at ~610°C before reducing the substrate temperature to between 480ºC and 500°C for InAs deposition. Various deposition amounts of InAs are delivered at 0.22 or 0.054 ML/sec. The delivery itself is continuous, or interrupted after well-formed 3D islands have appeared, as explained later. Instantly after InAs

deposition, the substrate temperature is ramped down in ~1 minute to the capping temperature of 400°C or 350°C, followed by MEE growth of the desired capping layer for optical samples. For uncapped counterparts intended for AFM studies, the temperature is dropped down to room temperature by cutting the heater power. In all cases of capped SAQD structures, between one and five layers of InAs islands are deposited, separated by a capping layer of desired composition and thickness, as discussed for specific cases in the following sections.

3.2.1 GaAs Capped Large and Small InAs QDs

Large InAs/GaAs QDs

Figure 3.4 shows AFM-based island density versus size distributions realizable in single-layer QD samples through the use of two growth approaches: the usual continuous InAs deposition [13] and the *punctuated island growth* (PIG) method [14]. The PIG approach divides the total deposited island material (InAs) into two or

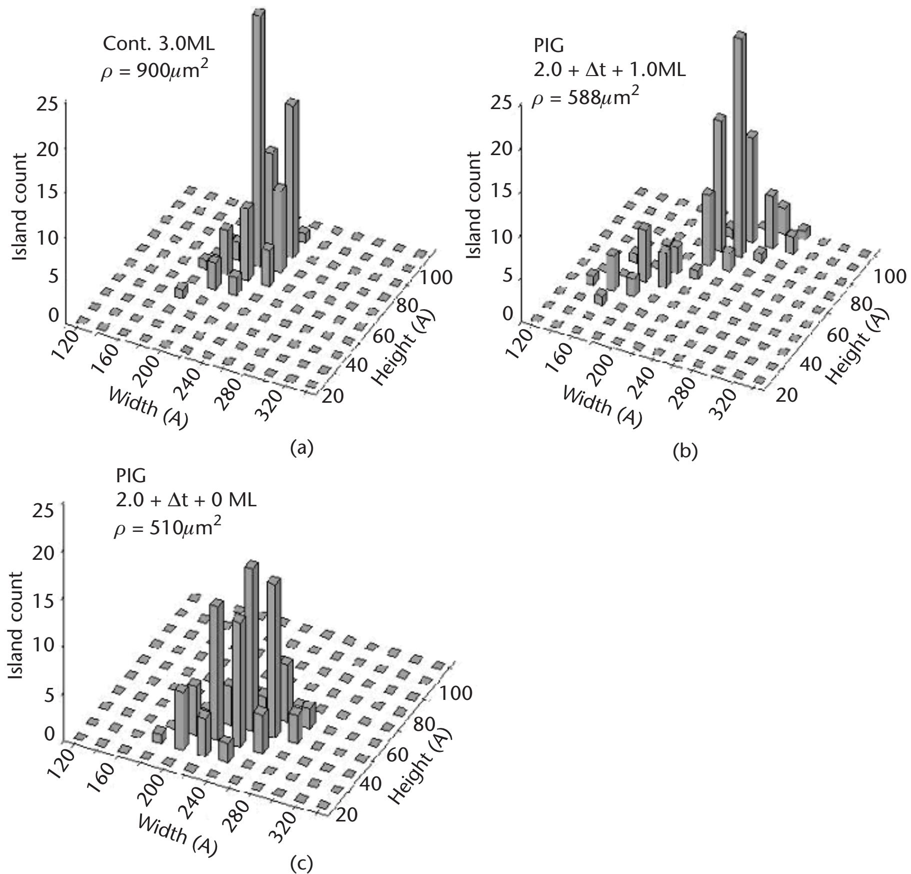


Figure 3.4 (a–c) AFM-determined InAs island size distributions for continuous and PIG cases. Note the similar larger size distribution for 3-ML depositions but significantly closer to the 2-ML sample density for the PIG sample. (*After:* [14].)

more stages, with the first stage being in the well-formed 3D island regime, that is, above the critical deposition amount of 1.6 ML for 3D island formation. The importance of the punctuation stage is to keep the island density high, since the technique of sub-ML incremental deposition [15] gives generally much lower QD densities compared to continuous deposition. Punctuating the island growth gives time for material to redistribute, so that during resumed deposition the probability of new island formation is suppressed, as can be seen by comparing the total island density in Figure 3.4(b) (2+1-ML PIG) with Figure 3.4(a) (3-ML continuous growth) and Figure 3.4(c) (island growth punctuation but no additional deposition) cases.

Enhanced size uniformity comes from the slower growth rate of larger islands due to strain buildup at the island base edge, thus giving the smaller islands the opportunity to reach the size range of the larger ones and reduce nonuniformity [14]. In fact, AFM size analysis shows [16] that virtually all islands have about the same base size (~21 nm), whereas the heights vary significantly from a most probable value of ~3.5 to ~8.5 nm. The corresponding aspect ratio (height-to-base ratio) changes from ~0.16 for small QDs to ~0.4 for large QDs. An aspect ratio of 0.16 may be attributed to a variety of side facets {203}, {113}, {136}, and so on cited in the literature for small QDs, whereas a ratio of 0.4 is suggestive of {101}-type side facets for large islands. The cross-sectional TEM image of the uncapped QD sample shown in Figure 3.5 reveals the QD cross section in the [110] azimuth for steep (large) and shallow (small) islands to be triangular with a base size of ~21 nm and sidewall slopes of ~36° and 24°, respectively [16]. The steeper slope is consistent with {101} facets, whereas the shallower slope is closer to the {113}, {136}-type facets.

The low-temperature PL behavior of the counterpart samples with islands capped by MEE-grown GaAs is shown in Figure 3.6. In all cases except one the InAs deposition rate is 0.22 ML/sec. The exception is the continuous 2.5-ML InAs deposition sample marked "slow" for which the deposition rate is 0.054 ML/sec [17], about four times slower than the others. Note the broad [*full width at half maximum* (FWHM) ~80 meV] peak for the 2-ML sample as contrasted with the very narrow (FWHM ~25 meV) dominant peak at ~1.07 eV for the 3-ML PIG sample.

We note that the slow-growth-rate 2.5-ML continuous deposition gives predominantly large SAQDs, very similar to those of the 3-ML PIG. The self-limiting base size of such islands combined with larger volume leads to the dramatically reduced PL linewidths and remarkable reproducibility of QD growth. The intermediate depositions of 2.5 ML at the faster growth rate show a stark bimodal PL emission, suggesting a bimodal distribution of SAQD sizes. Remnants of the smaller size islands are manifest in the small PL emission at higher energy (~1.18 eV) even in the 3-ML PIG or slow-growth-rate 2.5-ML SAQDs. These observations suggest that 2-ML deposition SAQDs exhibiting a unimodal though broad distribution of sizes centered around the smaller (shallower) islands potentially offer a photoresponse quite different from that of the essentially unimodal distribution of the large (steep) SAQDs formed for 3-ML PIG or slow-growth-rate 2.5-ML islands. In between, one can realize situations with dominant bimodal island size distribution and hence

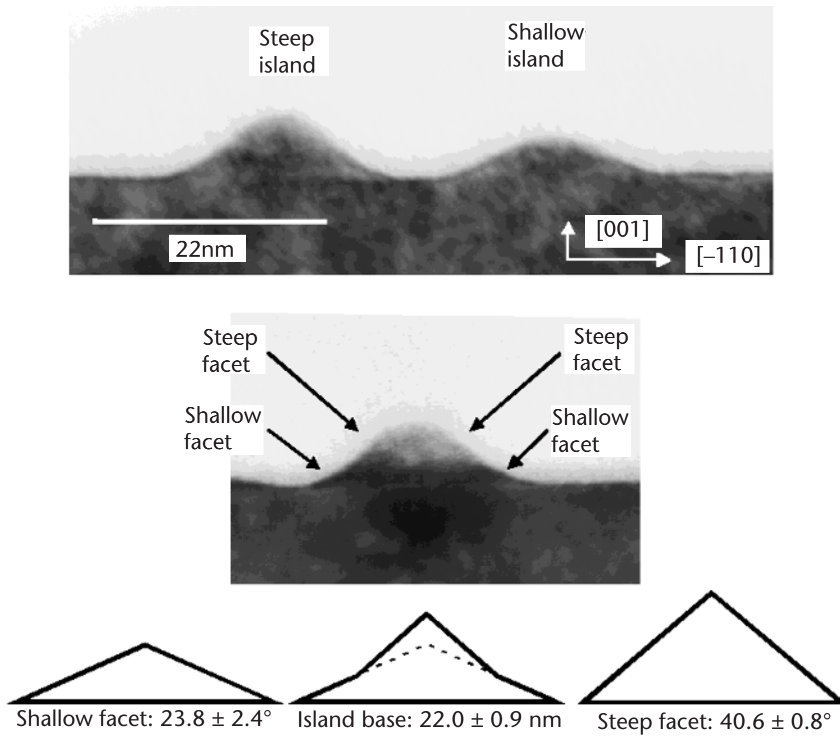


Figure 3.5 Cross-sectional TEM images showing the shallow and steep InAs islands, along with a path for the charge. Note that the island base size changes very little during the shape change. This self-limiting base size is a consequence of the high strain at the island edges. (*From:* [16]. © 2001 I. Mukhametzhanov. Reprinted with permission.)

bimodal photoresponse. Indeed, such considerations will be seen in subsequent sections to be of importance to the design of QDIP devices.

To gain insight into the nature of the electronic states of structurally reasonably well characterized SAQDs, we examined systematically the narrow, essentially unimodal distributions represented by the 3-ML PIG samples and the slow-growth-rate 2.5-ML continuous-growth samples [17]. We have carried out systematic studies of the electronic structure combining PL and *PL excitation* (PLE) optical spectroscopy [18–20] with near- and middle-infrared *photocurrent* (PC) spectroscopy [18]. Figure 3.7 shows the low-temperature (~8K) PLE spectra of the two samples, taken for detection at their respective PL peaks (~1.069 eV). In PLE one varies the detection wavelength through the inhomogeneously broadened ground-state PL peak, and thus probes subensembles of QDs with different sizes, as first shown in [12]. Note the essential similarity of the two spectra, suggesting the similarity of the underlying island size distributions. The strongest transitions in PLE are seen in Figure 3.7 at ~83 and ~171 meV above ground-state transition and at weaker transitions at ~60 and ~135 meV.

Excited electron and hole state information is also obtained from PC measurements, the same quantity that forms an integral element of the photodetector

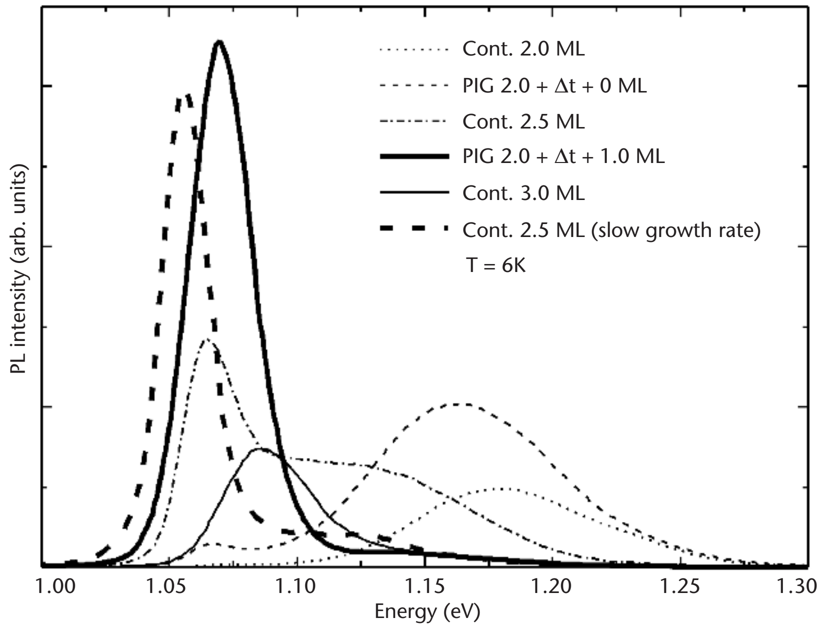


Figure 3.6 Low-temperature PL behavior of continuous-growth and PIG InAs/GaAs SAQDs.

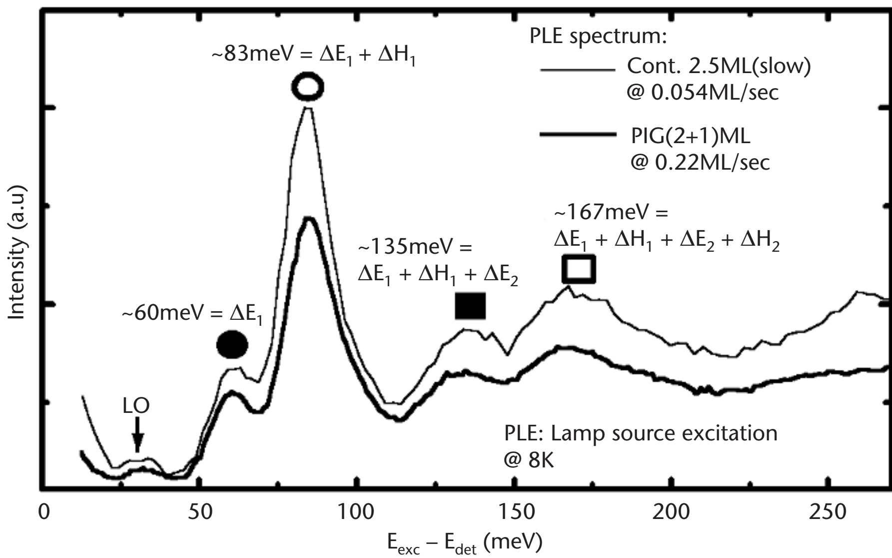


Figure 3.7 PLE spectra of large [PIG (2+1)-ML and slow-growth-rate 2.5-ML) InAs/GaAs QDs at 8K.

characteristics. Thus PC measurements were carried out for both interband and intraband photoexcitation in analogous PIG QD samples grown in *n*-*i*(MQD)-*n* geometry [18–20]. As noted in Section 3.1, and as discussed later in Section 3.3, our

focus is on QDIP structures that have an undoped QD region, the carriers being provided to the QDs from Si-doped GaAs supply layers separated by undoped spacer layers above and below the QD region. Thus for electron-based *n*-*i*(MQD)-*n* structures, the average occupancy of the QD ground states is controlled by the dopant concentration in the doping layers and their separation from the MQD region in distance through the thickness of spacer layers. This is analogous to modulation doping of heterojunction structures. Indeed, we have optimized the parameters to keep the average occupancy of the QD ground states to less than 2 at liquid nitrogen temperatures. This helps keep the dark current low but, from a basic investigation viewpoint, allows measurement of the interband photocurrent.

Figure 3.8 shows a typical room temperature *near-infrared* (NIR) PC response of the large SAQDs measured using a Bruker 55 *Fourier transform infrared spectrometer* (FTIR) spectrometer with an integrated cryostat [18–20]. Detected transitions are at 0.98, 1.06, 1.14, and 1.21 eV, which is consistent with the PL and PLE findings. Typically, NIR PC peaks are much less resolved at LHe temperatures due to decreased probability of carrier extraction and thus suppressed photocurrent. Due to the integrative nature of photocurrent detection, as opposed to the size subensemble selection in PLE, one resolves only the strongest transitions in the presence of inhomogeneous broadening. By contrast, the intraband transitions give

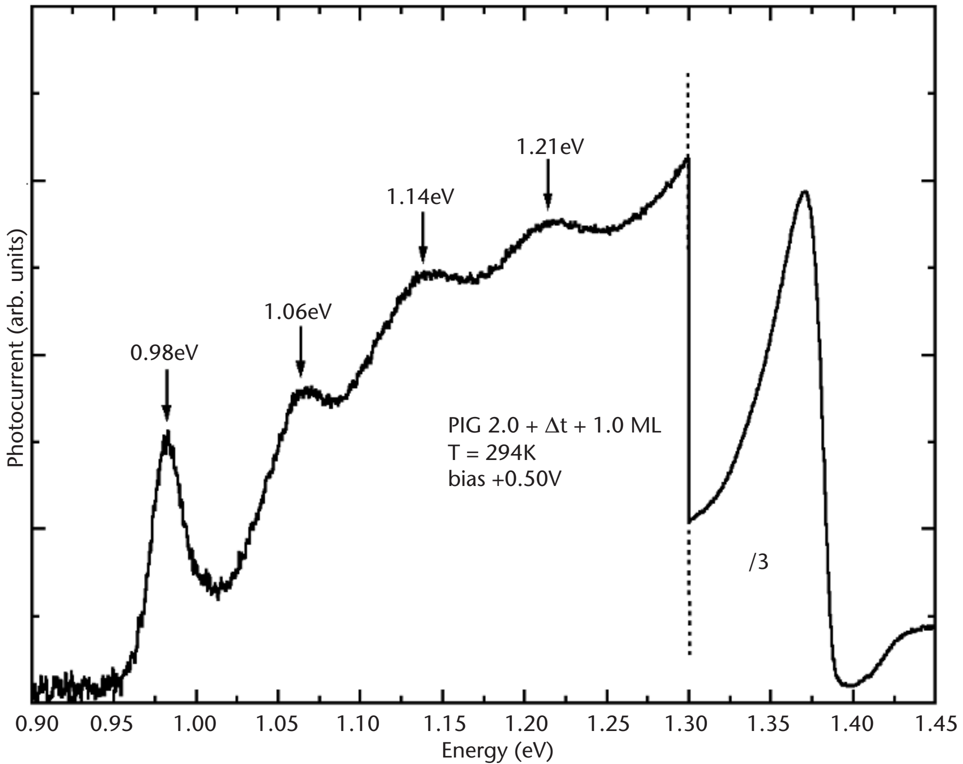


Figure 3.8 NIR PC spectrum of PIG 2.0+Δt+1.0-ML sample, where peaks reflect interband transitions. (*From:* [18]. ©2000 I. Mukhametzhanov. Reprinted with permission.)

much richer information on the electronic structure since, as found in size-selective PLE [18–20], level separations within electron or hole manifolds change much slower than the level energies with change in SAQD size. Thus one gets significantly better resolution in the intraband transitions while interband transitions are typically much broader (~25 meV).

Figure 3.9 shows the ~77K intraband *medium-wave infrared* (MWIR) PC spectrum in the electronic manifold, with two clear peaks at 115 and 170 meV. There is a shoulder indicative of a third peak at ~180 meV. According to the calculated energy levels in pyramidal InAs/GaAs QDs available in the literature [21, 22], one would expect a much richer PC level structure than the two detected peaks. However, one should bear in mind that some allowed transitions might not be observed due to small oscillator strengths or due to the shallow depth of the higher energy levels involved. Moreover, the inhomogeneity caused by the QD size fluctuation causes broadening of transition energies and thus compromises resolution. Nevertheless, combining the results of the systematic PL, PLE, and PC spectroscopies on the same samples allows us to place limits on the electronic structure as unambiguously as possible.

The typical behavior of the electron transitions in the large InAs/GaAs(001) QDs, extracted from such studies, is summarized in Figure 3.10. The dominant conduction band energies are at ~60, ~112, and ~172 meV above the ground state for the conduction band states. The valence band states are at ~23 and ~60 meV above the ground state. This, along with QD shape and size derived from AFM/TEM structural characterization, provides a framework for testing theoretical models and calculation algorithms.

In contrast to the predominantly large (steep, with most a probable height of ~7.5 nm) InAs QDs discussed earlier, for lower InAs deposition amounts up to ~2

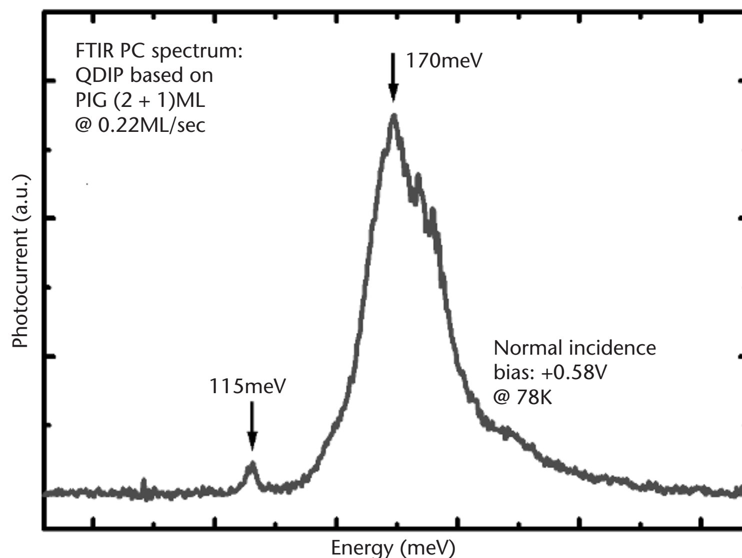


Figure 3.9 MIR PC spectrum of PIG 2.0+Δ*t*+1.0-ML sample. Peaks reflect intraband transitions.

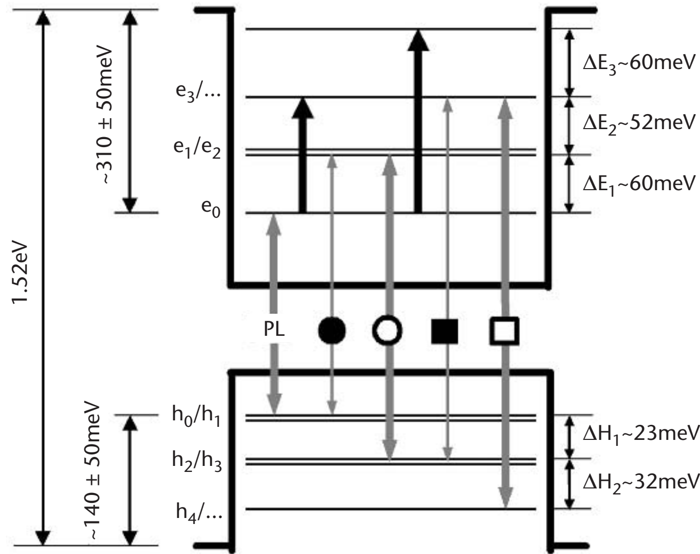


Figure 3.10 Schematic of the energy band diagram of the large [PIG(2+1)-ML and slow-growth-rate 2.5-ML)] InAs/GaAs QDs.

ML, the islands formed are smaller with a most probable height of ~3.5 nm. The size distribution is broader, which, combined with smaller volumes, manifests itself in the much broader PL linewidths of 65 to 80 meV. Nevertheless, to get some bounds on the nature of the electronic energy levels in the smaller dots, PL, PLE, and PC studies were carried out. Figure 3.11 shows typical PL and PLE spectra obtained for detection at the PL peak. The step seen at ~1.43 eV is the InAs wetting layer emission and is characteristic of a 2D confined, that is, QW, system. One clearly resolved peak is seen at ~1.26 eV, about 100 meV above the ground-state transition. A shoulder is seen riding to the lower energy size and suggests, as also expected from calculations [21, 22], that the first excited electron state in these dots is ~ 50 meV above the ground-state energy. The inhomogeneous broadening of the size is the dominant cause for lack of resolution in the PLE spectrum.

Figure 3.12 shows the intraband PC measured on a 2-ML InAs/100-ML GaAs five-layer stack *n*-*i*(MQD)-*n* QDIP structure. Note the broad response spread over 4.5 to 7.5 μm with dominant peaks suggested at ~5.5, 6, 6.5, and 7 μm. For handy comparison, also shown is the response of the large QDs such as those formed for the (2+1)-ML PIG structure of Figure 3.9, which, because of the much reduced size dispersion, gave a very narrow intraband response at 170 meV. The inferred energy level scheme for the small QDs is shown in Figure 3.12(b).

Combining all PL-, PLE-, and PC-based information available on our SAQDs, a comparison can be made between the ground-state transition energies measured and calculated as a function of the SAQD height for a fixed base size in the neighborhood

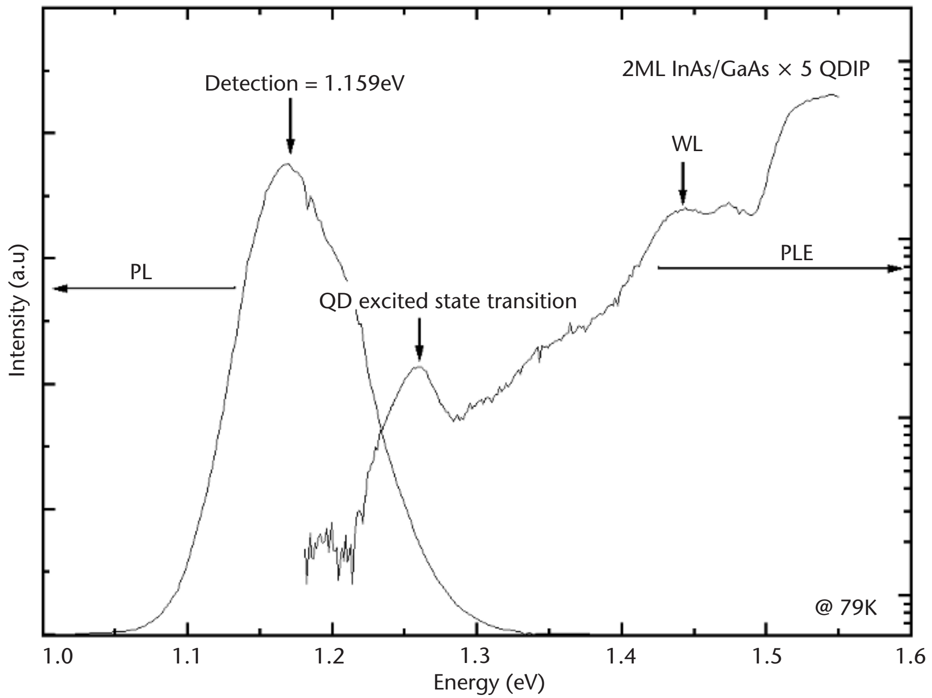


Figure 3.11 PL and PLE spectra of 2-ML InAs/GaAs × 5 QDIP at 79K.

of the self-limiting base value of ~20 nm. This is shown in Figure 3.13. The calculated results are from the group of Y. C. Chang, obtained using the bond-orbital method and finite element/continuum elasticity methodologies [21, 22; also private communication with S. J. Sun and Y. C. Chang]. The theory is normalized to the experimental value for a 5-nm height. The resulting theoretical curve provides a reasonable description of the experimental findings within their own uncertainties.

3.2.2 AlGaAs Capped Large InAs MQD QDIP Structures

Having established the basic characteristics of our large and small InAs/GaAs *single-layer* SAQDs, we next consider the nature of the *multilayered* QD structures needed for the QDIPs. As noted before, the real trade-off is between the number of QD layers and the total strain buildup that increases structural defect density. Operationally, the considerations are between the size of the islands, the lattice mismatch with the capping layer, and the thickness of the capping layer. It is routine now to exploit the buried island induced stress fields on the cap layer surface to achieve considerable *vertical self-organization* (VSO) of the SAQDs, and in many of our MQD structures this is the case. However, VSO is not an important requirement for the QDIPs. Thus, to reduce the cumulative strain effect, but at the cost of the extracted photocurrent, the spacer thickness between the SAQD layers can be made large. As an example, we examine *n*-*i*(MQD)-*n* structures involving five

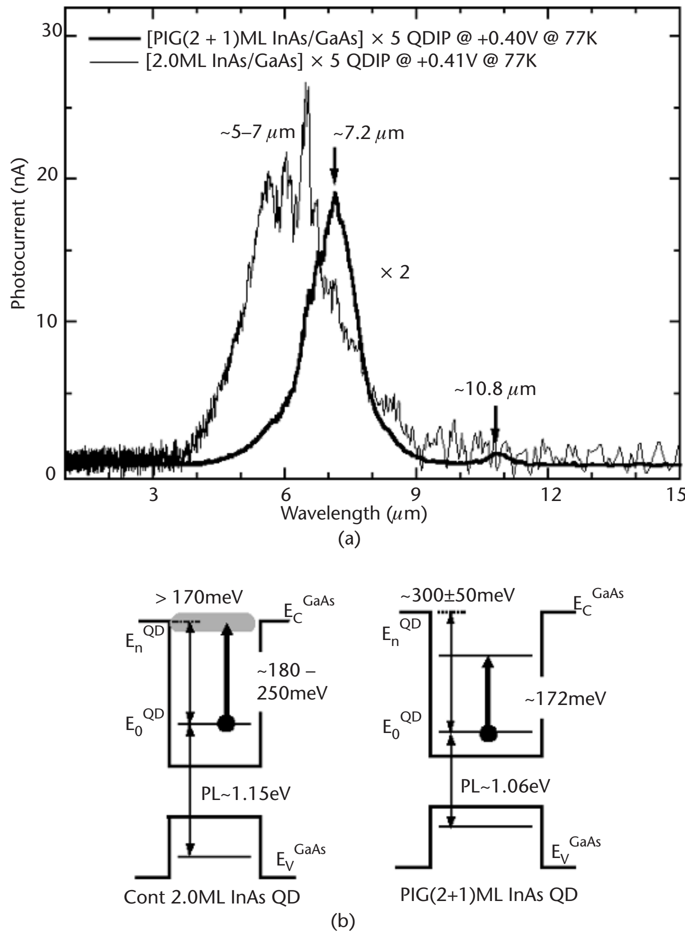


Figure 3.12 (a) Intraband photocurrent spectra the (2+1)-ML PIG InAs/GaAs × 5 QDIP and the 2-ML InAs/GaAs × 5 QDIP at biases of 0.40 and 0.41V at 77K, respectively, and (b) their schematic energy band diagrams.

(2+1)-ML PIG InAs 3D islands [23], as shown schematically in Figure 3.14. The spacer between the bottom (top) contact layer and the nearest QD layer had a thickness of 220 (240) ML. The spacers between the active QDs have a total thickness of 150 ML made of two kinds of spacer structures. In one type of structures, referred to as S-GaAs, the spacer material is GaAs. In the other type of structures, referred to as S-AlGaAs, 21-ML-thick AlAs/GaAs (1-ML/2-ML) *short-period superlattices* (SPSLs) are inserted to replace 21 ML of GaAs on both sides of each QD layer. The lower AlAs/GaAs SPSLs were located at a distance of 2 ML below the QD layer, while the upper SPSLs are deposited after growth of a 30-ML GaAs capping layer above the QD layer. The AlAs/GaAs SPSLs were grown at 500°C via the MBE approach, while the 30-ML capping GaAs was also grown via MBE but at 350°C. The average QD height of the AlGaAs counterpart uncapped sample is ~6 nm,

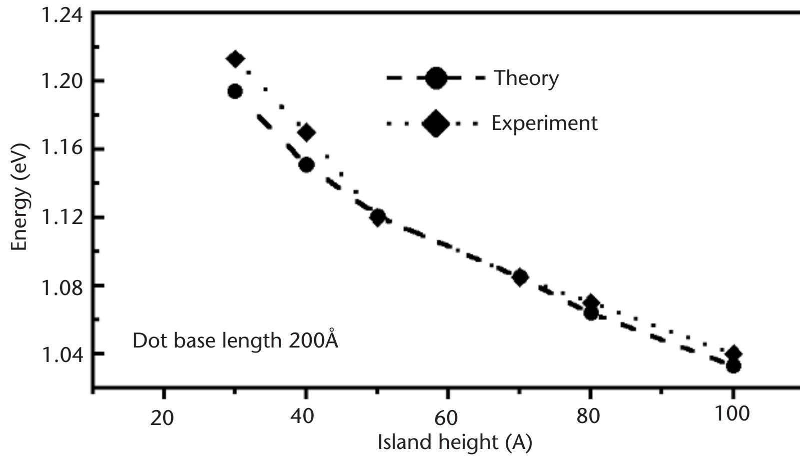


Figure 3.13 Comparison of experimentally determined and calculated QD ground-state transition energy as a function of InAs QD height for a base length of 20 nm.

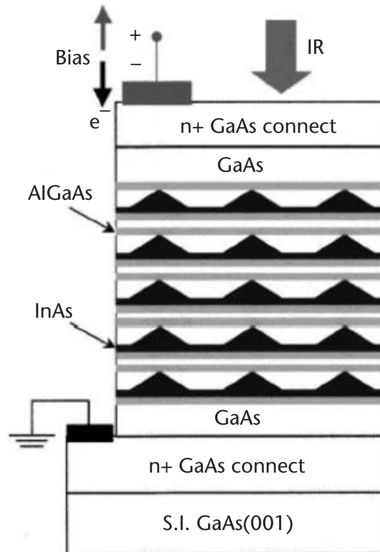


Figure 3.14 Schematic of QDIP structures examined. The gray bands represent samples with AlGaAs blocking layers.

slightly smaller than ~7.5 nm found for the counterpart of the GaAs sample. The dot densities in both cases are ~650 μm^2.

Figure 3.15 shows *cross-sectional TEM* (X-TEM) images of S-GaAs and S-AlGaAs [23]. Both parts indicate the existence of five-layer InAs QDs as well as the wetting layer. Note that, for S-GaAs (in Figure 3.15), the QD density in the top layer is smaller than in the lower layers, while the QD density in S-AlGaAs is similar for all layers. Figure 3.15(b) also shows clear contrast between AlGaAs and GaAs

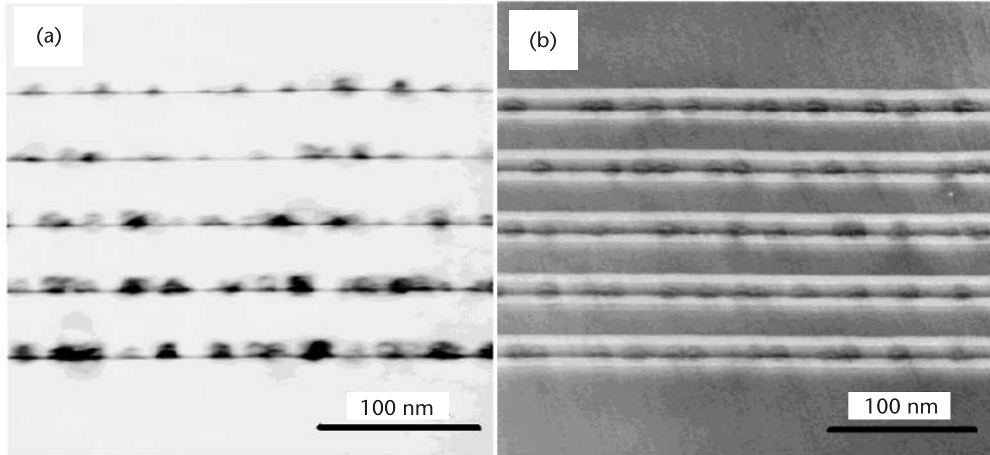


Figure 3.15 X-TEM images of a five-layer stack of (2+1)-ML PIG InAs/GaAs QDIPs (a) without (S-GaAs) and (b) with (S-AlGaAs) AlGaAs blocking layers (light bands).

layers. Based on the X-TEM measurements, the macroscopic defect density in these QDIPs was estimated to be ~10^8 cm^{-2} for S-GaAs and ~10^7 cm^{-2} for S-AlGaAs.

The PL spectra (Figure 3.16) of the QDIP samples at 78K show a main peak at 1.060 eV (S-GaAs) and 1.066 eV (S-AlGaAs) with a small broad shoulder on the higher energy side [23]. The main PL peak corresponds to the QD ground-state transitions. The small, broad PL shoulder is most probably due to smaller QDs. The

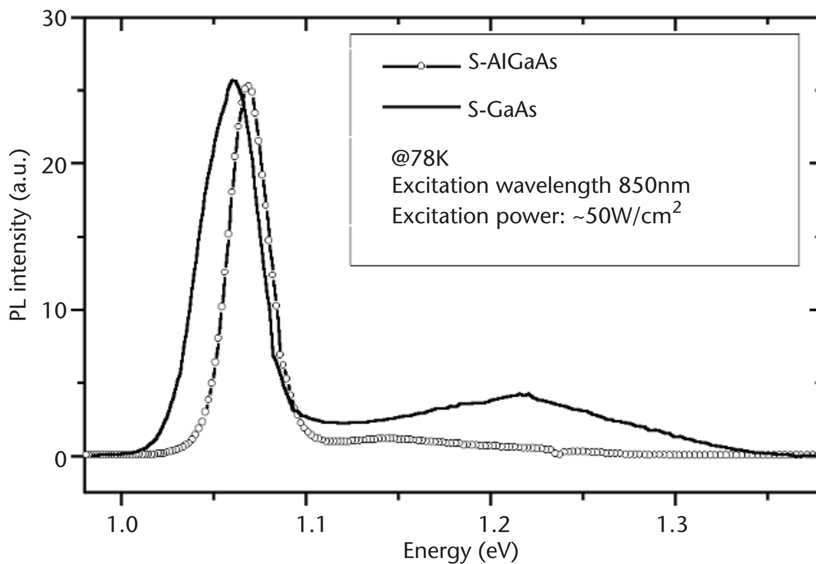


Figure 3.16 PL spectra of (2+1)-ML PIG InAs S-GaAs and S-AlGaAs QDIPs at 78K.

FWHM of the main peak is 36 meV (25 meV) for S-GaAs (S-AlGaAs). The narrow PL line width enables us to get more information on the QD electronic structure. Figure 3.17 shows the PLE spectra of S-GaAs and S-AlGaAs samples at 78K, recorded at a detection energy at the corresponding PL peaks at 1.060 and 1.066 eV. For S-GaAs, five peaks at 1.092, 1.120, 1.143, 1.188, and 1.227 eV are clearly seen, whereas for S-AlGaAs clear peaks are seen at 1.099, 1.131, 1.155, 1.205, and 1.261 eV. The PLE peak at 1.092 eV (1.099 eV) differs from detection energy by 32 meV (33 meV), which is close to the QD *longitudinal-optical* (LO) phonon energy, and thus corresponds to the LO phonon replica [24] in the S-GaAs (S-AlGaAs) samples. The other PLE peaks can be attributed to QD excited states at 60, 83, 128, and 167 meV (65, 89, 139, and 195 meV) above the ground transition for S-GaAs (S-AlGaAs). The PLE peaks of S-AlGaAs blue shift relative to the corresponding peaks of S-GaAs.

Interband transitions in these QDIPs were also detected using FTIR photocurrent spectroscopy, indicating that the electron level occupancy was less than two per QD [23]. Figure 3.18 shows a typical interband FTIR photocurrent spectrum of S-GaAs (bias: +0.33V) and S-AlGaAs (bias: +0.40V) at room temperature. Multiple peaks at 0.996, 1.08, 1.16, and 1.23 eV (1.01, 1.10, 1.15, 1.20, and 1.27 eV) are clearly resolved for S-GaAs (S-AlGaAs). Compared with the PL/PLE results described earlier, the PC peaks with the lowest energy at 0.996 and 1.01 eV can be attributed to the ground-state transitions of QDs in the S-GaAs and S-AlGaAs, respectively. The rest of the PC peaks can be attributed to excited states with excitation energy of 84, 164, and 234 meV (92, 140, 192, and 264 meV) relative to the ground state for S-GaAs (S-AlGaAs). These observed excited states, except that at

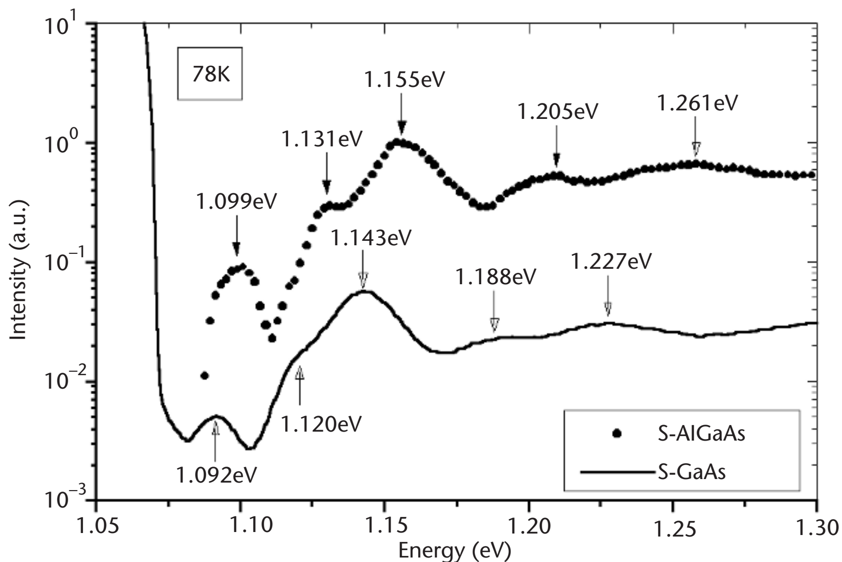


Figure 3.17 PLE spectra of (2+1)-ML PIG InAs S-GaAs and S-AlGaAs QDIPs at 78K for detection energy of 1.060 and 1.066 eV. (*From:* [21]. © 2001 American Institute of Physics. Reprinted with permission.)

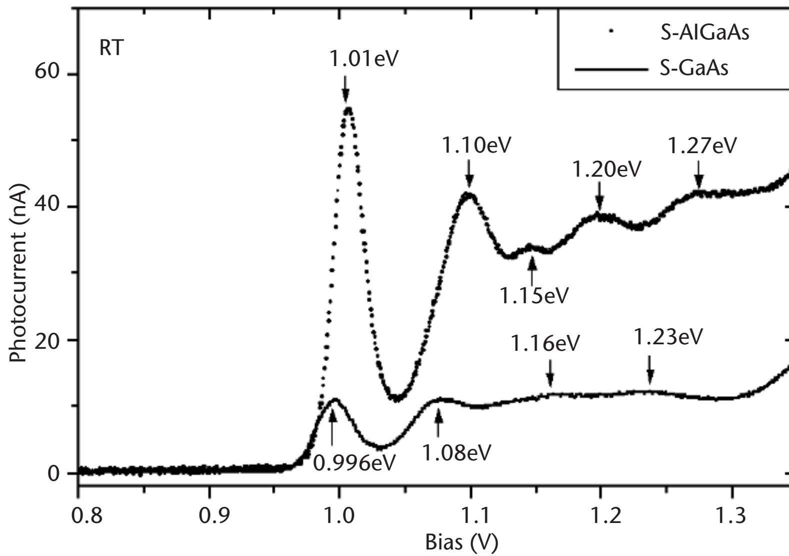


Figure 3.18 Room temperature NIR interband photocurrent spectra of (2+1)-ML PIG InAs S-GaAs and S-AlGaAs QDIPs. (*From:* [23]. © 2001 American Institute of Physics. Reprinted with permission.)

234 meV (264 meV), are consistent with those observed in the PLE spectrum. The observation of the additional higher excited states at 234 meV (264 meV) reflects the fact that the NIR interband photocurrent spectroscopy is more sensitive to high excited states than PLE spectroscopy. Although detailed transport mechanisms of interband photocurrent in QDIPs with *n-i-n* configurations are not very clear so far, photocurrent serves as a very useful probe for QDIPs. The interband photocurrent spectroscopy is a complementary method to PLE and to intraband photocurrent spectroscopy, which is described later.

We next present *intraband* photocurrent spectroscopy findings that offer more direct information on the QD electron states [23]. Figure 3.19(a, b) shows typical middle-infrared FTIR photocurrent spectra on S-GaAs (at a bias of +0.4V) and S-AlGaAs (at a bias of −1.00V), respectively, at 40, 77, and 100K. The peak position of the photocurrent spectra at different temperatures is, by and large, identical for both QDIPs. The S-AlGaAs has a main intraband photoresponse peak at 205 meV (6.0 μm) with a FWHM of 16 meV (0.5 μm), while S-GaAs has a main intraband photoresponse peak at 175 meV (7.2 μm) with a FWHM of 32 meV (1.3 μm). Also, S-GaAs shows a small intraband peak at 115 meV (10.8 μm). The addition of the AlGaAs layers leads to a blue shift in the wavelength of peak intraband photoresponse. The blue shift in intraband photocurrent peaks, the interband photocurrent peaks, and the PLE peaks is attributed to an enhanced confinement potential caused by the inserted AlGaAs layers (as a confining layer) as well as to the above-

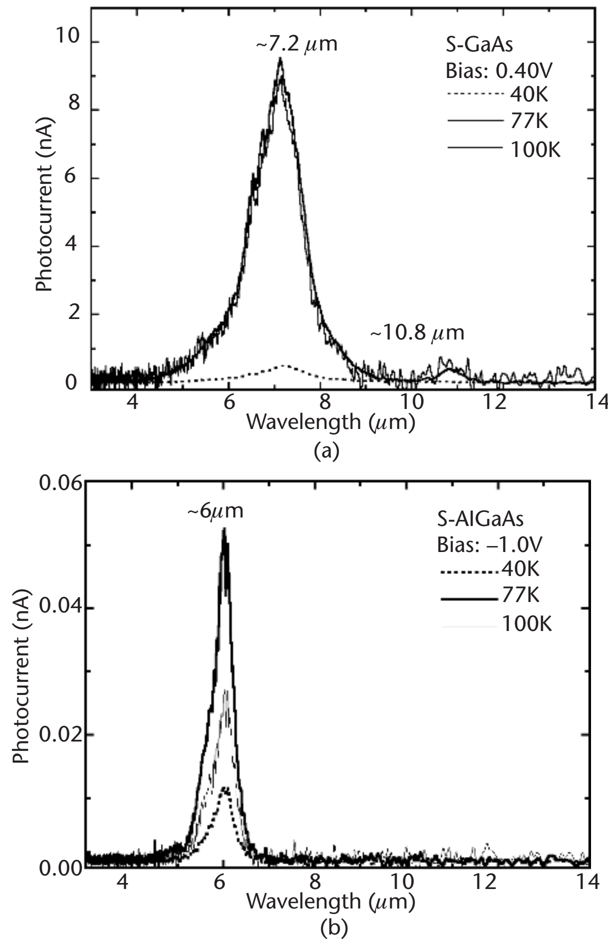


Figure 3.19 Intraband photocurrent spectra of (a) S-GaAs and (b) S-AlGaAs at 40, 77, and 100K. The equivalent monochromatic optical excitation power is ~10^{-7}W. (*From:* [23]. © 2001 American Institute of Physics. Reprinted with permission.)

mentioned slight reduction in QD size. As noted in the preceding subsection, systematic PL/PLE study on single-layer PIG InAs/GaAs QD samples suggests that 300±50 meV is the localization energy of the QD electron ground state [18–20]. Thus, we attribute the above-mentioned intraband transitions at 115 and 175 meV to bound-to-bound transitions. Figure 3.20 shows the intraband peak intensity as a function of temperature for S-GaAs and S-AlGaAs [23]. Both QDIPs show similar temperature-dependent behavior. With increasing temperature, the peak intraband photocurrent remains unchanged below ~40K, and then increases up to ~77K. We explain this as being due to the increase in the average electron population in QDs with increasing temperature. Above ~77K, the peak photocurrent begins to decrease. Both QDIPs show maximum intraband photoresponse at ~77K.

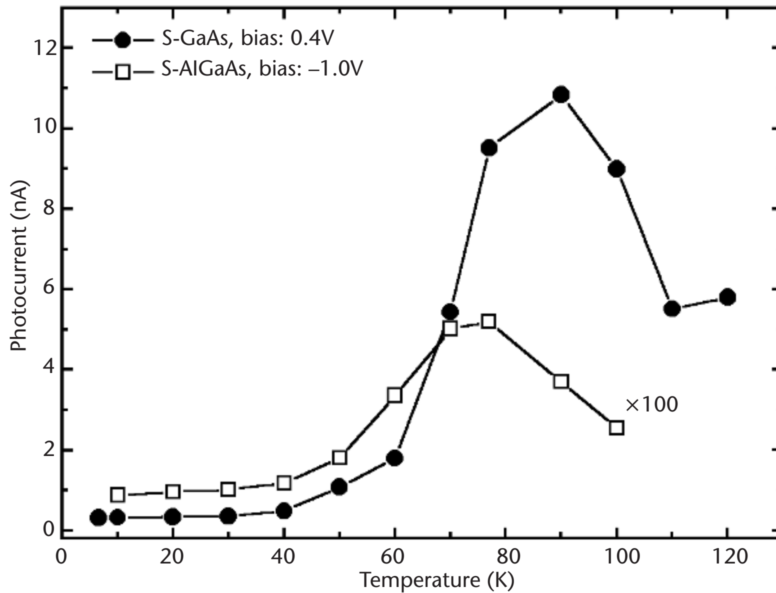


Figure 3.20 Intraband peak photocurrent versus temperature for S-GaAs and S-AlGaAs at a bias of +0.40 and –1.00V, respectively. The equivalent monochromatic optical excitation power is ~10^{-7}W. (*From:* [23]. © 2001 American Institute of Physics. Reprinted with permission.)

3.2.3 $In_xGa_{1-x}As$ Capped Small and Large InAs MQD-Based QDIP Structures

Although the large InAs/GaAs/AlGaAs QDIP structures allow considerable reduction in the dark current due to the AlGaAs blocking layers, the photoresponse is restricted to the 5- to 7-μm regime. To achieve operation in the important 8- to 12-μm atmospheric window, we need to find ways of manipulating the intraband electronic transitions of the QDs. To this end, instead of choosing to utilize InGaAs alloys as the QD forming material as is done in most other studies, we have explored achieving this objective by manipulating the QD quantum confinement potential via change in the barrier GaAs layer to alloy InGaAs. The addition of In into the otherwise GaAs capping layer changes the QD confinement potential via two effects: (1) the chemical effect of In lowering the band edges and thus the band discontinuities that define the confinement potential, and (2) the accompanying reduction in the lattice mismatch between the InAs island material and the InGaAs capping layer lowering the strain and thus the band edge discontinuities. Indeed, as discussed later, replacing some of the Ga with Al at a fixed In concentration will allow restoration of the chemical effect without changing the strain effect, resulting in nearly the same atomic size of Ga and Al. This then allows independent manipulation of the strain effect and chemical effect and thus greater flexibility in tailoring the electronic structure of the QDs to suit the desired QDIP operational wavelengths.

Motivated as noted above, systematic studies of InAs/$In_xGa_{1-x}As$/GaAs QD structures in the range $x = 0.1$ to 0.2 with a thickness of 5 to 60 ML for the capping layer over InAs islands were carried out utilizing PL and PLE spectroscopies and

TEMs [25]. It was found that for small InAs QDs formed for 2.0-ML InAs deposition, near 20-ML-thick $In_{0.15}Ga_{0.85}As$ capping layers provide good photoresponse at ~9 μm [26]. For the large SAQDs formed for 2.5-ML slow-growth-rate InAs deposition, 30-ML-thick $In_{0.15}Ga_{0.85}As$ capping layers grown via MEE at 350°C were found to be good [25]. Moreover, the $In_{0.15}Ga_{0.85}As$ layer regions between the InAs QDs are expected to act as a QW having its own energy states. Tailoring of the thickness, composition, and positioning of the $In_xGa_{1-x}As$ capping layers thus also allows for tailoring of the position of the QW energy states with respect to the QD bound states.

MQD-Based QDIP Structures

A schematic of the InGaAs capped QDIP structures is shown in Figure 3.21. These QDIP samples consist of an undoped active InAs QD region separated from the bottom and top Si-doped GaAs contact layers by 220- and 240-ML undoped GaAs spacer layers, respectively. The InAs MQD region comprises five layers of 2.0- or 2.5-ML InAs deposited at the growth rate of 0.22 ML/sec at 500°C. The InAs QD layers are followed by two types of capping/spacer layers. The 2-ML InAs small QD-based QDIPs have a 20-ML $In_{0.15}Ga_{0.85}As$ capping layer followed by 180-ML GaAs as a spacer [26], referred to here as InGaAs QDIPs. The 2.5-ML InAs large QD-based QDIPs have a 30-ML $In_{0.15}Ga_{0.85}As$ capping layer followed by 170-ML GaAs as a spacer [17]. In both types of structures, the first 40 ML of growth,

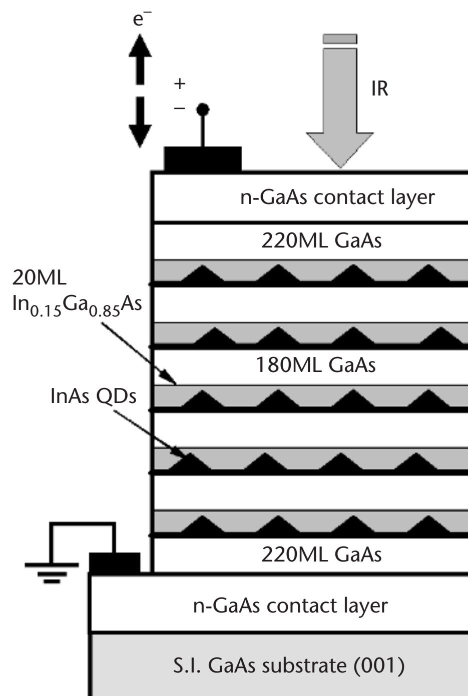


Figure 3.21 Schematic of a typical InGaAs capped InAs SAQD-based QDIP device structure.

following InAs QD formation, was via MEE at 350°C to minimize intermixing with the InAs QDs [12]. The rest of the GaAs spacer was grown by conventional MBE growth at 500°C. These InAs QD layers and spacers were repeated five times and then the top 240-ML-thick GaAs spacer was deposited. Finally, a highly Si-doped GaAs contact layer was grown at 500°C to complete the *n*-*i*(MQD)-*n* configuration suited for electron intraband transition-based QDIPs.

Small InAs/InGaAs Capped QDIP Structures

The PL and PLE behavior for detection at the PLE peak of the 2-ML InAs deposition that gives small QDs with the most probable height of ~3.5 nm and a base of ~20 nm is shown in Figure 3.22. The PL peak position of InGaAs QDIP is red-shifted to ~1.125 eV (FWHM of 65 meV) compared to the peak at ~1.16 eV for the GaAs capped QDIP shown in Figure 3.11. This is due to a lower potential confinement effect and the strain relaxation effect of the 20-ML $In_{0.15}Ga_{0.85}As$ capped layer. The step-like PLE feature at 1.34 eV arises from the GaAs/InGaAs/GaAs QW region between the InAs QDs, reflecting the step-like state density of QWs with one-dimensional confinement. The PLE peak at ~1.21 eV reflects the discrete (δ-function-like) excited state density centered about 85 meV above the electron ground state and broadened by the inhomogeneity of QD sizes. This peak arises from the InGaAs capped InAs QDs, not from the GaAs/InGaAs/GaAs QW. Thus, both X-TEM and PLE data indicate that the InGaAs-capped InAs QDs have 3D confinement and δ-function-like discrete state density, at least for the electron ground

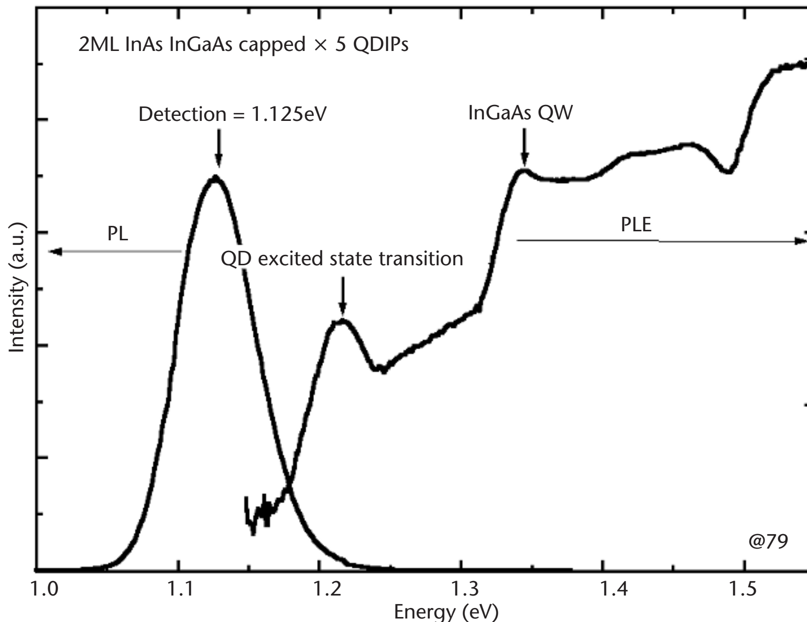


Figure 3.22 PL and PLE spectra of 2-ML InAs InGaAs capped × 5 QDIP at 79K.

states. Such states are the initial states of the observed intraband photoresponse of the QDIPs discussed next.

Normal-incidence FTIR intraband photocurrent spectra of the QDIPs were measured as a function of bias. Figure 3.23 shows the behavior at a bias of +0.40V at 80K. For handy reference, shown also is the behavior of the 2-ML GaAs capped QDIP at +0.42V and 80K. While the GaAs QDIP has strong photoresponse over a relatively wide range of 5 to 7 μm, the InGaAs QDIP has a strong sharp peak at ~9 μm with a FWHM of only 16 meV.

Figure 3.24(a) shows FTIR intraband photocurrent spectra of the 2.0-ML InGaAs capped QDIP at a bias of +0.4V and –0.4V at 80K. Photocurrent peaks are observed in the ~8- to 9-μm range as a function of bias. Moreover, the photocurrent spectrum at +0.4V is red-shifted with respect to that at –0.4V, thus indicating a dependence on the applied bias. The inset of Figure 3.24(a) shows the photocurrent peak position as a function of bias. With increasingly positive bias, the photocurrent peak positions are red-shifted. To shed light on the nature of the electronic states underlying the above-noted observed behavior of the electron intraband photoresponse, we look to the nature of the confining potential defining the system.

A schematic energy band diagram along a line passing through the QD base center and apex is shown in Figure 3.24(b) under reverse and forward bias. Note the asymmetry in the confining potential present (even at flat band condition) due to the coupled presence of the InAs QD and $In_xGa_{1-x}As$ QW, as discussed earlier [26]. In such a situation, while the ground electron state may be a sufficiently pure QD state, the higher lying excited states, though bound, are the eigenstates of the asymmetric potential that may not be representable even as a strong mixture of what otherwise

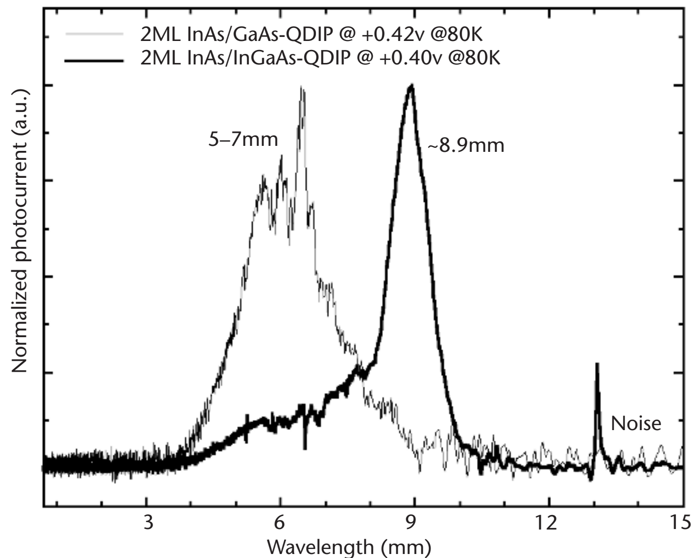


Figure 3.23 Intraband photocurrent spectra of 2-ML InAs/GaAs QDIP and 2-ML InAs/InGaAs QDIP at a bias of +0.42V and +0.40V at 80K, respectively. Note the shift to longer wavelength due to lowering of the confinement potential depth.

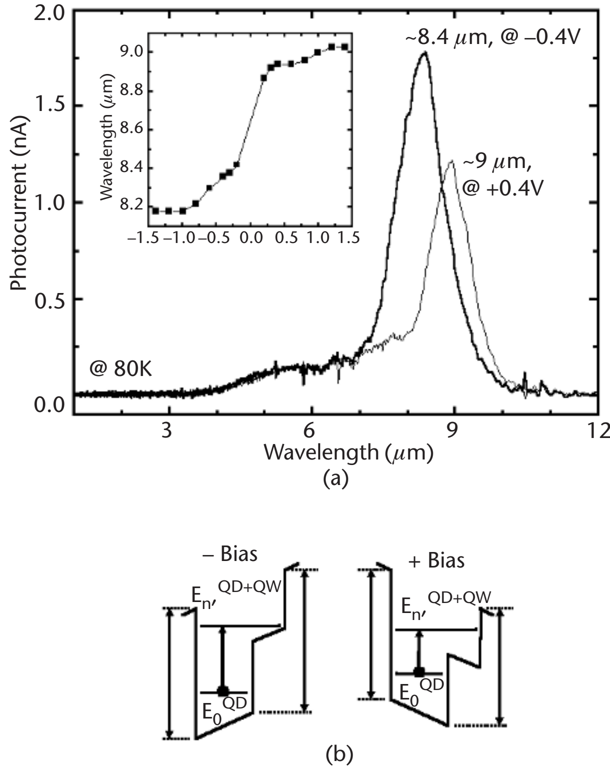


Figure 3.24 (a) Intraband photocurrent spectra of 2-ML InAs/InGaAs QDIP (at 80K) and (b) schematic energy band diagrams for opposite bias cases. The inset in part (a) shows the peak wavelength as a function of applied bias.

would be the excited states of the QD (in the absence of a QW nearby) and the QW (in the absence of QDs). Clearly, the need for reliable calculations that provide meaningful quantitative comparison is evident and we hope that the findings reported here will provide the needed incentive to theoretical efforts. Qualitatively, however, such a potential under positive bias is seen to give lower effective barriers, thus giving smaller energy spacings and hence the observed longer wavelength (i.e., red-shifted) photoresponse.

The bias dependence of the intraband peak photoresponse of the InGaAs QDIP, its dark current, and that of the GaAs QDIP at 80K are shown in Figure 3.25. The intraband photoresponse and the dark current increase rapidly with increasing bias. In the case of the InGaAs capped QDIP, above ±1.2V, the photocurrent begins to drop. This is a negative differential photocurrent phenomenon [26–28]. Note that the dark current of InGaAs QDIPs is two to four orders of magnitude smaller than that of the GaAs QDIPs. This is mainly due to the larger binding energy for the QD ground state in the InGaAs QDIP than that of the GaAs QDIP.

To gain insight into the nature of the electronic states responsible for the observed intraband photoresponse, we have analyzed the above data and a summary is schematically indicated in the band diagrams for the InAs QDs in GaAs

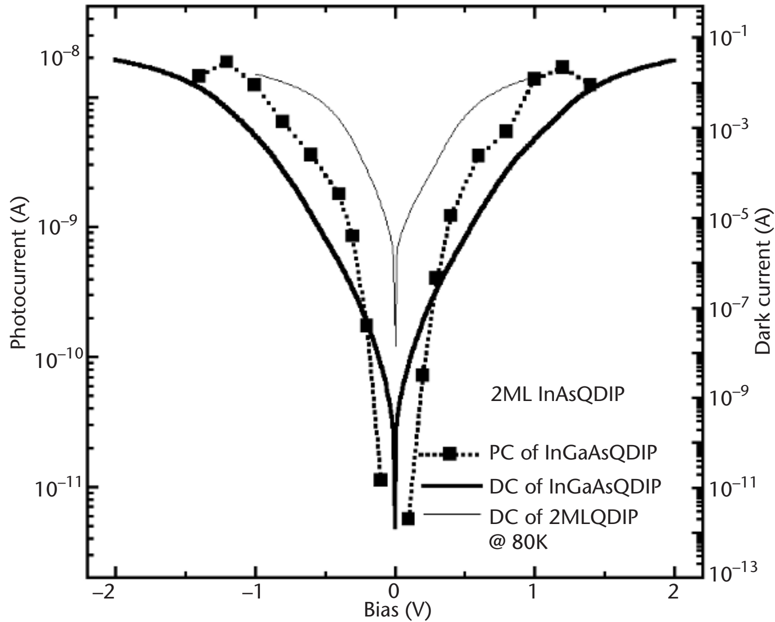


Figure 3.25 Photocurrent of 2-ML InAs/InGaAs QDIP and dark current of 2-ML InAs/GaAs QDIP and 2-ML InAs/InGaAs QDIP as a function of bias at 80K.

QDIPs and in InGaAs QDIPs shown in Figure 3.26. These are drawn for a line passing through the QD base center and apex. We consider the intraband transitions of both QDIPs to originate from the QD electron ground state since less than two electrons occupy a QD in these QDIP structures [23].

Lower limits on the binding (localization) energies of the QD ground states in the two types of QDIPs can be obtained from their respective PL peak positions. Subtracting the observed PL peak positions of 1.170 eV (1.06 μm) for the GaAs QDIP and 1.127 eV (1.10 μm) for InGaAs QDIP from 1.508 eV (band gap of GaAs at 79K), we obtain the sum of the binding energies of the ground-state electrons and

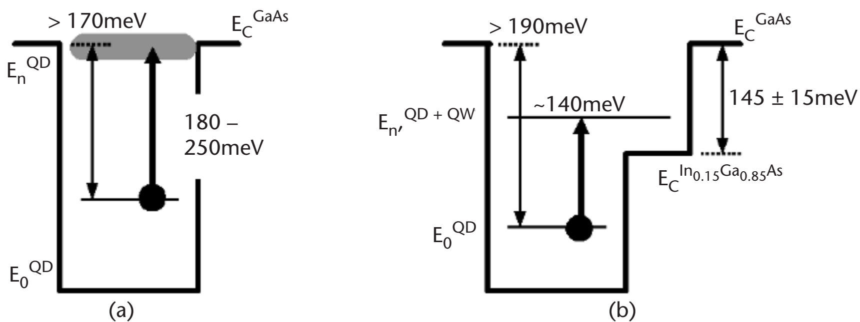


Figure 3.26 Schematic energy band diagrams of (a) 2-ML InAs/GaAs × 5 QDIP and (b) 2-ML InAs/InGaAs × 5 QDIP with energies extracted from the measurements.

holes as ~340 and ~380 meV for the two QDIPs, respectively. Then, using the well-known fact that the binding energy of the electron ground state is significantly higher than that for the hole ground state, we can safely conclude that the electron ground-state binding energies are greater than half of the values noted above for the sum of the electron and hole binding energies, that is, larger than 170 meV for QDs in GaAs QDIP and larger than 190 meV for QDs in InGaAs QDIP. Because GaAs QDIP has a strong PC response in the range of 5 to 7 μm (180–250 meV), the electron excited states involved in the intraband transitions in this structure seem to be mainly those located near the GaAs conduction band edge. By contrast, the intraband PC peak in the InGaAs QDIP is at ~140 meV, significantly smaller than the minimum binding energy of ~190 meV for the electron ground state and thus indicating that the excited final states involved must be at least 50 meV below the GaAs conduction band edge. In this case, however, the electron states of the 20-ML-thick $In_{0.15}Ga_{0.85}As$ cap layer acting as a QW defined by the conduction band discontinuity between the GaAs and strained $In_{0.15}Ga_{0.85}As$ conduction bands in the regions between the QDs should be considered. The ratio of the conduction band discontinuity to the bandgap difference for the $In_xGa_{1-x}As$/GaAs strained system has been considered to lie in the range of 0.58 [29] to 0.70 [30]. The conduction band offset of GaAs and strained $In_{0.15}Ga_{0.85}As$ (ΔE_c) is then 130 to 160 meV. Thus the excited states involved in the intraband transitions are likely located above the conduction band edge of $In_{0.15}Ga_{0.85}As$, as shown in Figure 3.26(b). In such a situation, the QD electron excited states can be coupled with $In_{0.15}Ga_{0.85}As$ QW electron states that lie in the ~145-meV region between the GaAs and $In_{0.15}Ga_{0.85}As$ conduction band edges. Identifying the origin of the excited states must await appropriate theoretical analysis.

Large InAs/InGaAs Capped QD-Based QDIP Structures

Next we consider the large InAs QDs realized via deposition of 2.5-ML InAs at the slower growth rate of 0.054 ML/sec at 500°C and capped by 30-ML $In_{0.15}Ga_{0.85}As$ followed by 170-ML GaAs spacer layers [17, 25]. These QDIPs are referred here as 2.5-ML QDIPs. Figure 3.27 shows the PL spectra of a single as well as a five-layer 2.5-ML QDIP at 79K. Island symbols represent shallow (small) and steep (large) QDs. While the 2.0-ML QDIP, included in Figure 3.27 as a handy reference, has a unimodal PL behavior peaked at 1.10 μm (FWHM of 65 meV), the 2.5-ML QDIP shows a marked bimodal PL behavior with peaks at 1.12 and 1.20 μm.

The bimodal PL can be understood in terms of our TEM and AFM studies that reveal a reduction in the average size and density of the upper QD layers of the 2.5-ML QDIP. The AFM studies of the uncapped counterpart samples showed that the last (fifth) layer had a density of ~1×10^{10}/cm^2 and average QD height of ~3 nm (comparable to that formed for the 2.0-ML QDIP), whereas the first three QD layers had a density of ~4×10^{10}/cm^2 and a larger QD ~6 nm high. The bimodal nature of the QD size distribution, dominated in this sample by variation in the uppermost two layers, is the dominant cause of the bimodal PL response. It is well established that the ground-state transition energy decreases with increasing size of these SAQDs. We thus ascribe the 1.20-μm PL peak to the larger QDs, which are

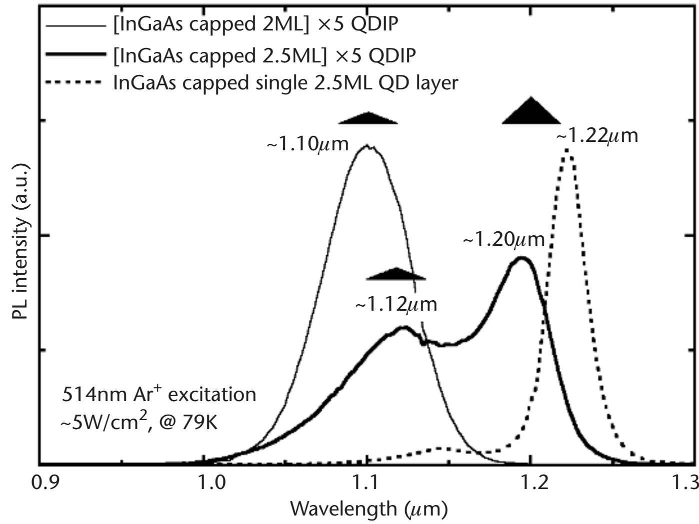


Figure 3.27 PL spectra of InGaAs capped 2.0-ML × 5 QDIP, InGaAs capped 2.5-ML × 5 QDIP, and a InGaAs capped single 2.5-ML QD layer at 79K.

dominant in the lower layers, and the 1.12-μm PL peak to the smaller QDs, similar to those of the 2.0-ML QDIP [17, 25].

Figure 3.28 shows PLE spectra of 2.5-ML InAs QDs capped by (1) 170-ML GaAs, (2) 30-ML $In_{0.15}Ga_{0.85}As$ + 140-ML GaAs, and (3) 30-ML $In_{0.20}Ga_{0.80}As$ + 140-ML GaAs. The abscissa represents the energy difference between excitation energies and detection energies, the latter being the ground-state PL transition energy of each of the three samples. For InAs QDs capped by $In_xGa_{1-x}As$ layers, transitions associated with the GaAs/$In_xGa_{1-x}As$/GaAs QWs, formed in the regions between the islands, are observed (marked *QW* in Figure 3.28). Note that with increasing In composition of the $In_xGa_{1-x}As$ capping layers, the energy differences between the excited states and ground state become smaller due to the effects of a lower confinement potential. This implies that QDIPs having InAs QDs capped by a $In_xGa_{1-x}As$ layer (InGaAs QDIPs) can show longer photoresponse wavelength than QDIPs without an $In_xGa_{1-x}As$ layer (GaAs QDIPs).

Figure 3.29 shows the intraband FTIR photocurrent spectra of the 2.5-ML QDIP at +0.8V and –0.8V. A dual-peak IR photoresponse with peaks in the MWIR (~5.5–6 μm) and the *long-wave infrared* (LWIR) region (~9–11 μm) is seen [17, 31]. We attribute the presence of the MWIR and the LWIR photoresponse to the presence of the large and small QDs, respectively. Recall that, as shown in Figure 3.24(a), the 2.0-ML QDIP comprising the small QDs has a LWIR photoresponse at ~9 μm, which is also very narrow [31–33].

The relative intensities of two photoresponse peaks change with bias as shown in the inset of Figure 3.29. At low negative or positive bias, the MWIR photoresponse is dominant. With increasing bias magnitude, both MWIR and LWIR photoresponse increase. However, the intensity of the LWIR photoresponse increases

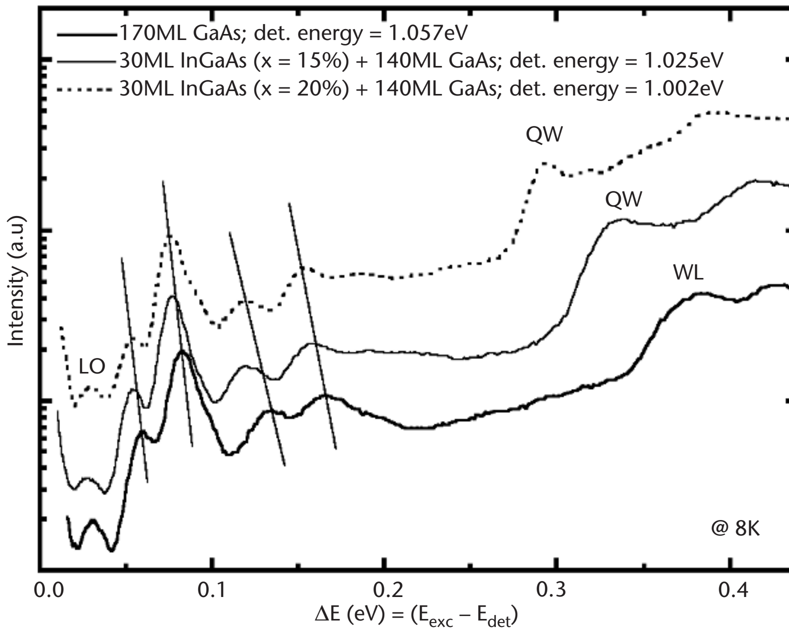


Figure 3.28 PLE spectra at 8K for 2.5-ML InAs QDs capped by 170-ML GaAs, 30-ML $In_{0.15}Ga_{0.85}As$ layer + 140-ML GaAs, and 30-ML $In_{0.20}Ga_{0.80}As$ layer + 140-ML GaAs. Features associated with GaAs/$In_xGa_{1-x}As$/GaAs QWs are marked *QW*, whereas *WL* denotes the InAs wetting layer for the sample with no $In_xGa_{1-x}As$ layers.

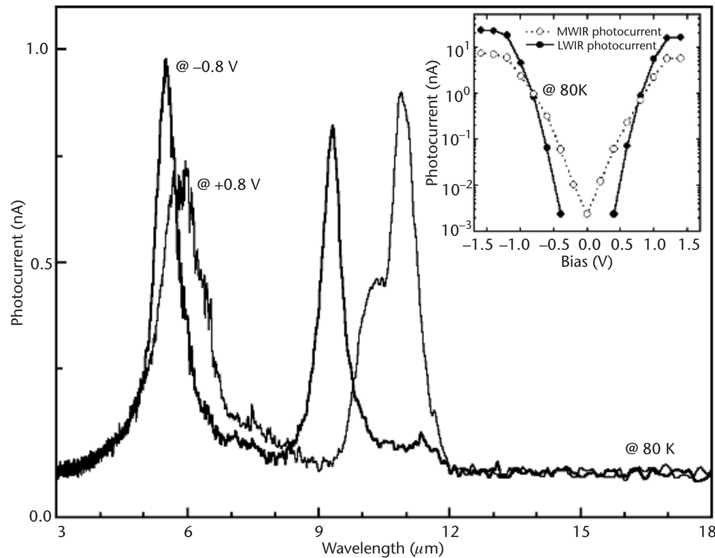


Figure 3.29 Intraband photocurrent of 2.5-ML InAs/InGaAs × 5 QDIP at a bias of +0.8V and –0.8V at 80K.

faster and becomes dominant above ~±0.8V. This may be understood as follows. At low bias, the electrons occupy the ground states of the larger QDs more than those of small QDs because of the lower electron ground-state energy of large QDs. With increasing bias, the ground-state occupation of small QDs increases and probably the gain of photoexcited electrons from the small QDs also increases more rapidly than that of large QDs. Eventually, the LWIR photoresponse is dominant above ~±0.8V. Note also that, as for the 2.0-ML QDIP (Figure 3.24), the MWIR and LWIR photocurrent peak positions in the 2.5-ML QDIP also red-shift with increasing positive bias.

Given the just discussed systematic studies of the electronic structure of small and large InAs/GaAs and InAs/InGaAs SAQDs, we summarize next the essential findings on the nature of the states involved in the observed intraband transitions. Regrettably, reliable electronic structure calculations for the type of GaAs(001)/InAs/InGaAs QD structures employed for QDIPs in the current studies do not exist. The upper and lower bounds on the electron ground-state binding energy (with respect to the GaAs conduction band edge) of small and large QDs can, however, be estimated based on the following.

First, note that the observed PL transition energies are 1.04 (1.22 μm) and 1.11 eV (1.10 μm) for InGaAs capped QDs (Figure 3.27) and that the GaAs bandgap energy is 1.51 eV at 80K. Second, PL and PLE studies on GaAs capped (2+1)-ML PIG InAs QDs indicated an electron ground-state binding energy of 300 ± 50 meV, as depicted in Figure 3.30(a) [18–20]. Third, it is well established that the electron binding energies are significantly higher than those for holes [18–20]. And fourth,

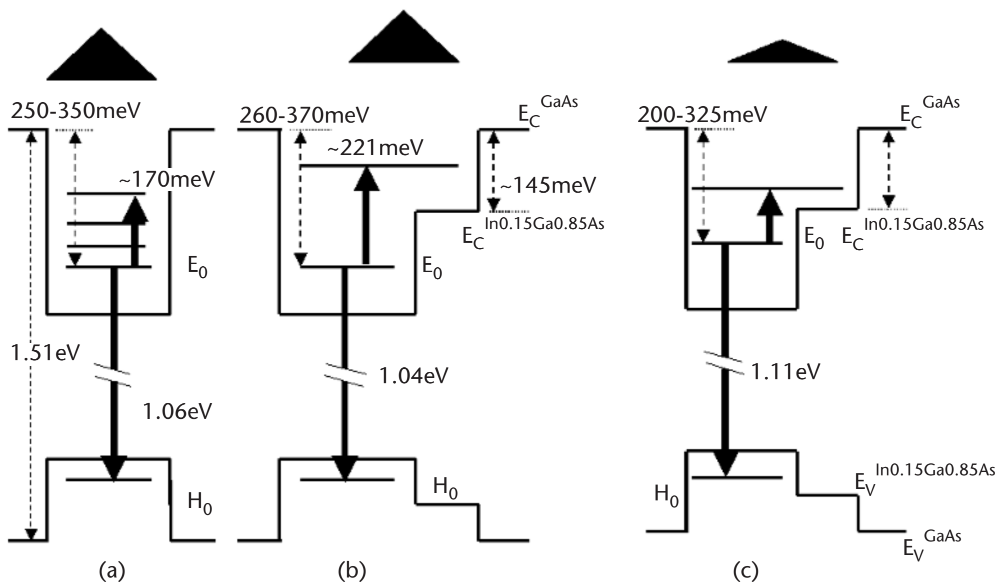


Figure 3.30 Energy-level schematic of (a) large GaAs capped, (b) large InGaAs capped, and (c) small InGaAs capped InAs SAQDs.

theoretical results for two InAs/GaAs QDs of different size indicate that the difference between their electron ground-state energies is larger than that of their hole ground-state energies [34, 35]. Therefore, as shown in Figure 3.30(b, c), the binding energies of the electron ground states of large and small InGaAs capped QDs are in the ranges of 260 to 370 and 200 to 325 meV, respectively. Namely, the photocurrent peaks at ~221 meV (~5.6 μm) and ~124 meV (~10 μm) are due to transitions from the QD ground state to bound excited states that have estimated binding energies as small as 39 and 76 meV, respectively. This, however, is also the energy regime where the InGaAs capping layer region between the islands, acting as a QW, gives rise to its own energy levels arising from the approximately 145-meV conduction band discontinuity between the GaAs and the strained $In_{0.15}Ga_{0.85}As$ conduction band edges [29, 30]. The final states of the observed intraband photocurrent peaks at ~221 and ~124 meV are thus a mixture of the QD excited states and the InGaAs quantum-well states.

We note that, based on the available theoretical results [21, 22, 34] for GaAs(001)/InAs/GaAs QDs closest to the large QDs under investigation here, the final states involved in the middle-wavelength (~5.6-μm) intraband transitions are not the first or even second excited bound states, but rather very high bound states (most probably fifth or higher) [18–20, 34]. Although the separation between the ground and any given excited electron state decreases with increasing QD size [16, 34, 35], our experimental findings of the photocurrent at the lower transition energy (~124 meV) arising from the smaller QDs clearly show the more complex nature of the intraband photocurrent in QD structures due to the convolution, at a minimum, of the transition matrix elements, density of final states, and weighted photocurrent extraction barriers.

Variable Deposition Amount-Based Multispectral Response

While the 2.5-ML InAs/InGaAs × 5 QDIP developed a bimodal QD size distribution due to strain interaction effects in the growth direction, the *variable deposition amount* (VDA) QDIPs [16] were purposely designed to have a bimodal QD size distribution by using two stacks of different QD layers [25]. One VDA structure is 3 × 2.5-ML + 2 × 2.0-ML InAs/InGaAs QDIP, which consists of a first stack of three 2.5-ML InAs QD layers (deposited at 0.054 ML/sec) capped by 30-ML $In_{0.15}Ga_{0.85}As$ + 170-ML GaAs followed by a second stack of two 2.0-ML InAs QD layers (deposited at 0.054 ML/sec) capped by 30-ML $In_{0.15}Ga_{0.85}As$ + 170-ML GaAs.

A cross-sectional TEM image [25] of this VDA QDIP [Figure 3.31(a)] shows that lower QD layers (the first stack) have larger QD size and density than the upper QD layers. The defect density of ~6 × 10^6/cm^2 is slightly lower than that of the 2.5-ML InAs/InGaAs × 5 QDIP because the smaller InAs deposition amount (QD size) in the upper QD layers lowers the accumulated strain of the structure. The PL spectrum (Figure 3.32) also shows a bimodal behavior with peaks at ~1.1 and 1.21 μm. To further enhance the structural quality, we also examined a VDA QDIP in which the stacking order was reversed: The VDA 3 × 2.0-ML + 3 × 2.5-ML InAs/InGaAs QDIP consists of a stack of three 2.0-ML InAs QD layers (at

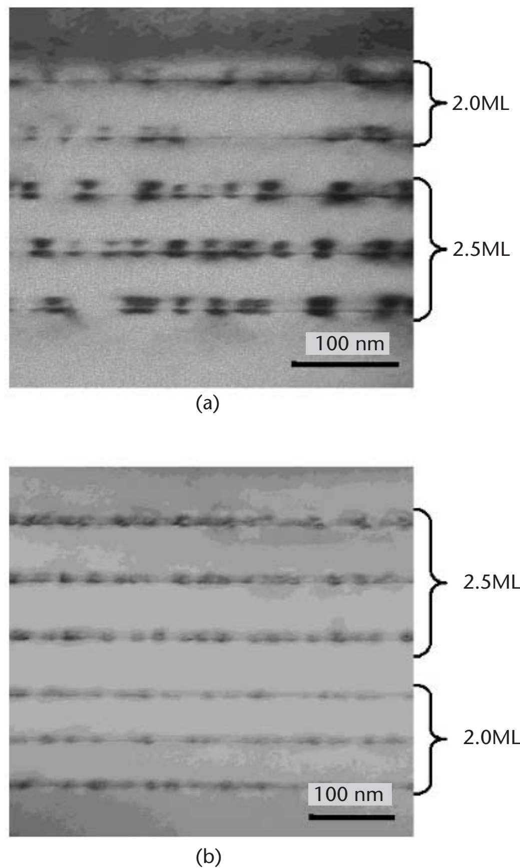


Figure 3.31 Cross-sectional dark field TEM image of (a) VDA 3×2.5-ML + 2×2.0-ML and (b) VDA 3×2.0-ML + 3×2.5-ML InAs/InGaAs QDIPs at the two-beam condition of **g** = (200).

0.22-ML/sec) capped by 20-ML $In_{0.15}Ga_{0.85}As$ + 180-ML GaAs followed by a stack of three 2.5-ML InAs QD layers (at 0.054 ML/sec) capped by 30-ML $In_{0.15}Ga_{0.85}As$ + 220-ML GaAs. As shown in an X-TEM image [Figure 3.31(b)] [25], the first three QD layers have relatively smaller QDs than do the upper three QD layers. Moreover, TEM study indicates that the defect density reduces significantly ($<3 \times 10^5/cm^2$). The PL spectrum of this VDA QDIP structure (Figure 3.32) also shows two peaks at ~1.1 and 1.21 μm [25].

The intraband photoresponse of both types of VDA QDIPs consisting of two stacks of QD layers (large and small QDs) is shown in Figure 3.33. Note the voltage-tunable two-color behavior, similar to the 2.5-ML InGaAs QDIP seen in Figure 3.29 [25]. At low positive or negative bias, the MWIR (~6-μm) photoresponse is dominant, but with increasing magnitude of the bias, the relative intensity of the LWIR (~10–12 μm) photoresponse becomes larger. We emphasize that, unlike the 2.5-ML × 5 InGaAs QDIPs, the VDA QDIPs have large and small QDs, which are located in separated and different QD layers.

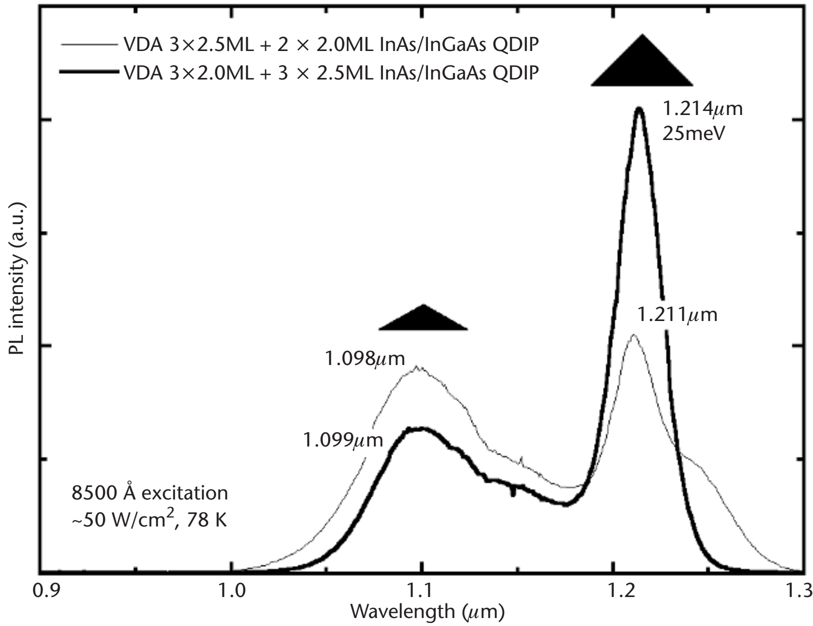


Figure 3.32 PL spectra of VDA 3×2.5-ML + 3×2.0-ML and VDA 3×2.0-ML + 3×2.5-ML InAs/InGaAs QDIPs at 78K.

3.3 QDIP Device Characteristics

3.3.1 Device Structures

Numerous reports have been published on InAs/$Al_xGa_{1-x}As$, InAs/InAlAs, InGaAs/GaAs, and InGaAs/InGaP QDIPs [23, 36–48]. The detectivity, D^*, of these devices was three to four orders of magnitude lower than QWIPs with similar structures. One reason is that the QWIPs normally incorporate 20 to 30 QWs, whereas the QDIPs that have been reported have incorporated 10 or, frequently, fewer QD layers. The fact that QD layers have less than unity fill factors (i.e., there is significant space between the QDs) compounds this problem. The net result is lower responsivity of the QDIPs. Another limiting factor for the QDIPs is that they have exhibited higher dark current that is comparable to or even higher than that of QWIPs. We note that most of the QDIP structures have employed a doped active region. As a means of reducing the dark current density, as emphasized in an earlier section, we have focused on *n-i-n* InAs/GaAs QDIPs with an unintentionally doped active MQD region [23, 40]. A schematic cross section of the structure is shown in Figure 3.34.

In the work described in the following sections, device fabrication followed standard processing procedures: photolithography, wet chemical etching, metal deposition and liftoff, and rapid thermal annealing. Mesas having a diameter of 250 μm and a height of ~1.4 μm were defined with an etch of $H_3PO_4 : H_2O_2 : H_2O$ (8:1:1). A 50-μm-diameter top contact and the bottom contact were formed by

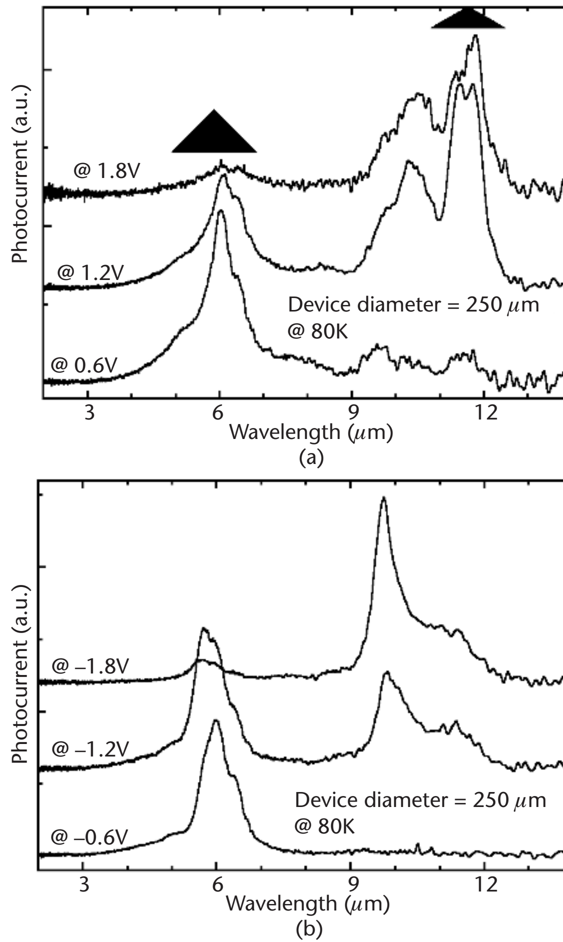


Figure 3.33 Intraband photocurrent spectra of the VDA 3×2.0-ML + 3×2.5-ML InAs/InGaAs QDIP at 80K at (a) positive bias and (b) negative bias.

evaporation and liftoff of Au/Ni/AuGe. The contacts were then annealed at 430°C for 20 seconds. We will use the convention that "positive" bias means that a positive voltage was applied to the top contact.

3.3.2 Unintentionally Doped Large (PIG) InAs/GaAs MQD-Based Detectors

We begin discussion of the QDIP device characteristics with the case of the large InAs islands capped with GaAs to establish the base photoresponse from which consequences of the variation in most probable island size and capping layer materials can be examined. We consider therefore the QDIP structures and samples of discussed earlier: 3-ML PIG/GaAs MQD-based QDIPs.

The normal-incidence intraband spectral responses of the QDIPs were measured by using an SRS 570 low-noise-current transimpedance preamplifier and a

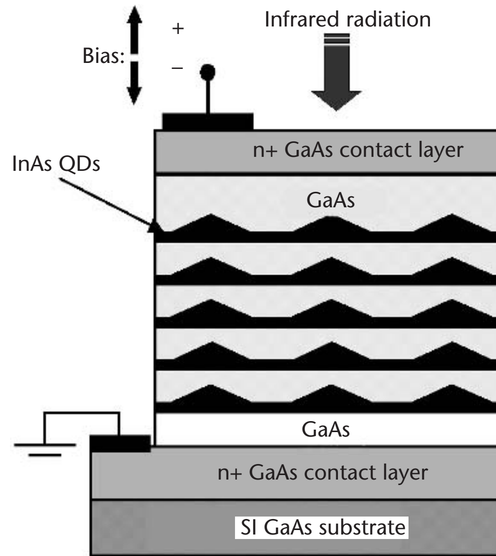


Figure 3.34 Schematic of the QDIP with an unintentionally doped active region.

Nicolet Magna-IR 570 FTIR with a glowbar source. Figure 3.35 shows the photocurrent spectral response at 0.1V bias and at temperatures of 60, 77, and 100K. The intraband photoresponse peaks occurred at 7.2 μm (~174 meV) for all three spectra. The FWHM of the spectrum, $\Delta\lambda$, was ~0.99 μm, from which it follows that $\Delta\lambda/\lambda$ = 14%.

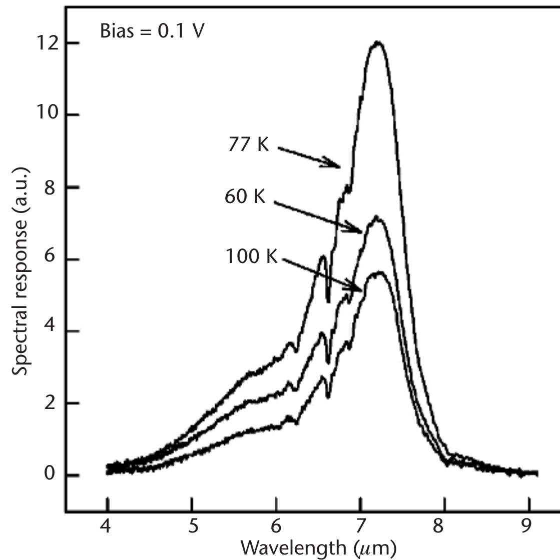


Figure 3.35 Photoresponse of the (2+1)-ML PIG GaAs capped QDIP measured at 60, 77, and 100K by using a low-noise-current preamplifier and FTIR spectrometer. The bias is 0.1V for all three temperatures.

Given the electronic structure summarized in Figure 3.10, these results indicate that the 7.2-μm response involves electron intraband transitions from the ground bound state to a higher bound state [49]. The observed spectral width reflects the uniformity of the size of the large QDs as well as a lower sensitivity of the electronic transitions to size fluctuations about a larger most probable size. The QDIP exhibits the highest photoresponse at 77K. As noted earlier, this is due to the increased electron occupancy of the lower states of the quantum dots with increase in temperature. As long as there are unoccupied excited states, more electrons can participate in photon-induced intraband transitions. However, a further increase in the number of electrons in the quantum dots, which results from the increase in dark current at higher temperature, will cause a decrease in the number of unoccupied excited states and, consequently, a decrease in the photoresponse. Additionally, a decrease in photoexcited electron lifetime, which occurs at higher temperatures, can result in a decrease in the photoresponse.

Photovoltaic operation of the QDIPs is observed as shown in Figure 3.36. This suggests the presence of a built-in electric field. As discussed earlier, the quantum dots have a pyramidal shape. This, in combination with the shallow potentials associated with the wetting layers beneath the QDs, results in an asymmetrical band structure. Thus, in contrast to QWIPs where asymmetrical structures have sometimes been designed to achieve photovoltaic operation, QDIPs intrinsically operate in this mode.

The responsivity, which is defined as the ratio between the output signal and the radiant input, was measured with a blackbody source at T = 995K. Because the

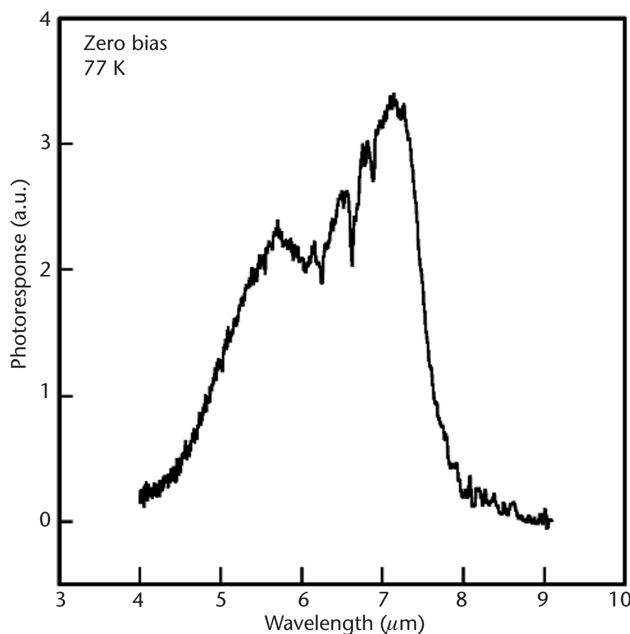


Figure 3.36 Photoresponse of the (2+1)-ML PIG GaAs capped QDIP measured at 77K and zero bias.

blackbody spectrum includes near-infrared radiation, which can result in interband transitions in QDs, optical filters were placed next to the aperture of the blackbody to block radiation with a wavelength of less than 3.5 μm. The intraband photocurrent was measured with a SRS 570 low-noise-current preamplifier and SRS 760 *fast Fourier transform* (FFT) spectrum analyzer. Figure 3.37 shows the peak responsivity versus bias measured at temperatures ranging from 40 to 120K. At 77K, with increase in positive bias, the peak responsivity increased more than three orders of magnitude from 1.3 mA/W at 0V to 4.11 A/W at 0.7V. For negative bias, the responsivity increased over four orders of magnitude from 0.22 mA/W at 0.05V to 2.9 A/W at –0.7V. We note that the lowest peak responsivity occurred at about –0.05V. This indicates that an additional bias voltage of ~0.05 is required to compensate the build-in electric field.

With increasing bias, positive or negative, the responsivity increased rapidly up to a maximum value. The corresponding bias for these peaks was –0.9, –1.0, –1.1, and –1.2V for 40, 60, 77, and 100K, respectively. With further increase in bias, the responsivity slowly decreased. This negative differential responsivity can be attributed to the large number of electrons in the QDs at high bias. Starting at zero bias, as the bias is increased, more and more electrons occupy states in the QDs and participate in the intraband transitions. Therefore, the photocurrent continues to increase until a significant fraction of the excited states are occupied, at which point the photocurrent decreases.

At higher temperature, the excited states are more quickly occupied and thus the maximum responsivity is achieved at lower bias levels. The responsivity was nonsymmetrical for negative and positive bias because electrons in the quantum dots

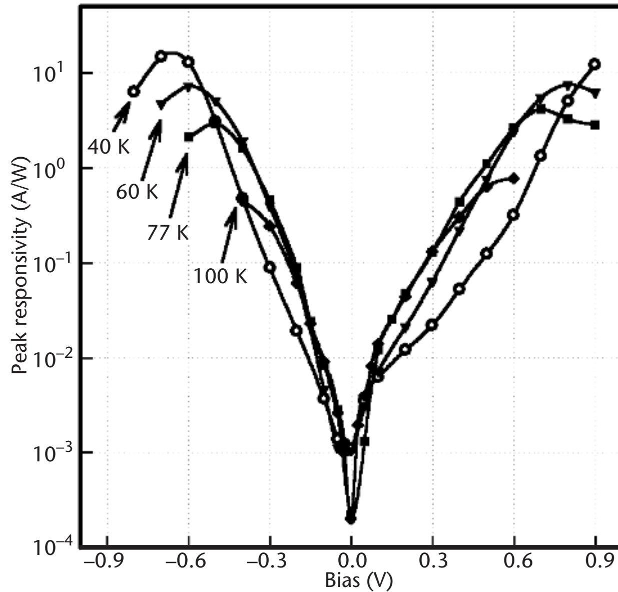


Figure 3.37 Peak responsivity of the (2+1)-ML PIG GaAs capped QDIP measured at 40, 60, 77, 100, and 120K using a blackbody (T = 923K) spectrum analyzer.

experience different potentials for transport toward the top contact or the bottom contact.

A good discussion of the dark current mechanisms in QDIPs is presented in [8]. Four principal mechanisms, as illustrated in Figure 3.38, contribute to the dark current in QDIPs:

1. Low bias sequential resonant tunneling, by which electrons tunnel from one quantum dot to another when the energy levels in adjacent quantum dots line up.
2. Thermionic emission, by which electrons are thermally excited from the confined states in quantum dots or contact layers, move toward the contact layers, and are collected by the external circuit. Thermionic emission increases exponentially with temperature as $I_{\text{dark}}(T) \propto \exp\left(\frac{E_a}{kT}\right)$, where E_a is activation energy, k is the Boltzmann constant, and T is the temperature.
3. Phonon-assisted tunneling from excited bound states to continuum states.
4. Phonon-assisted tunneling between excited bound states.

Dark current versus voltage characteristics were measured with an HP 4145 semiconductor parameter analyzer. The dark currents were normalized to the device area and are plotted in Figure 3.39. There is a significant increase in the dark current with increasing bias and temperature. At a bias of 0.1V, the dark current

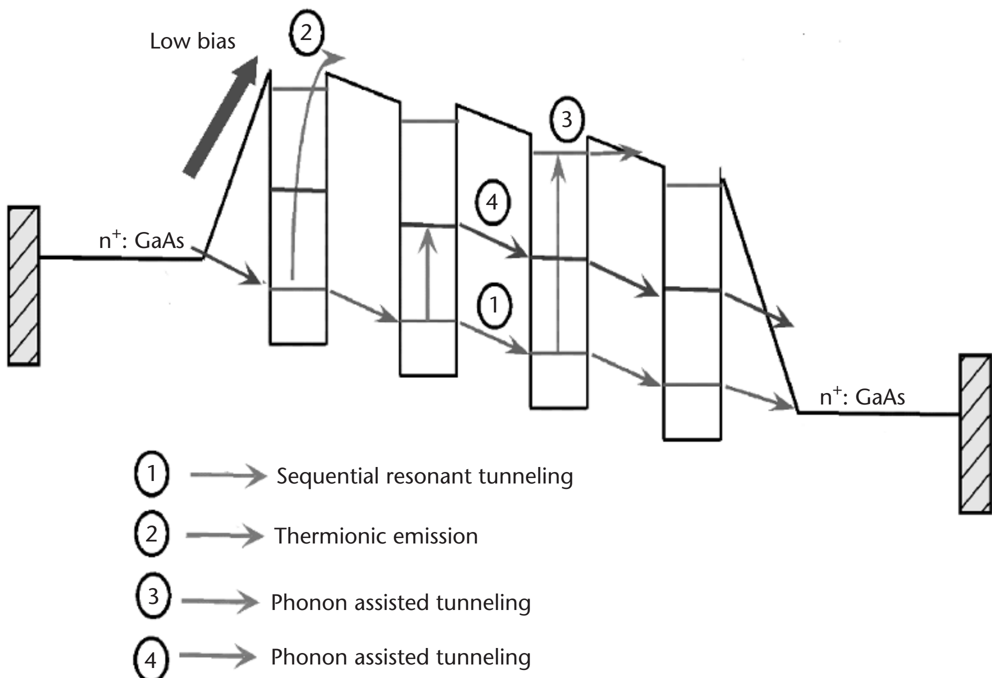


Figure 3.38 Possible originations of the dark current in QDIPs biased at low voltage.

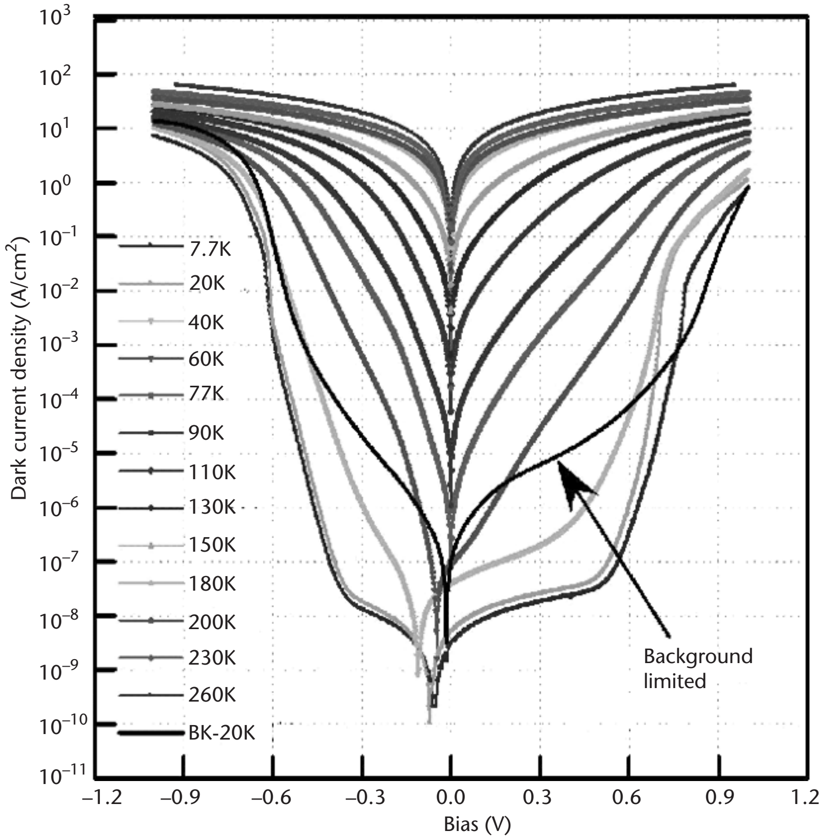


Figure 3.39 Dark current versus voltage characteristics of the (2+1)-ML PIG GaAs capped QDIP measured with a HP 4145 semiconductor parameter analyzer. The temperature ranges from 7.7 to 260K. The black line represents the 7.7K *I-V* characteristics with room temperature background.

density was 7.4×10^{-8} A/cm^2 at 20K. It rose to 6.1 A/cm^2 at 260K, which is a nearly eight orders of magnitude increase. At a bias of –0.1V, similarly, there was more than a nine order of magnitude increase from 2.2×10^{-9} A/cm^2 at 20K to 6.2 A/cm^2 at 260K. The dark current densities were asymmetric for positive and negative bias, for the same reasons that give rise to asymmetry in the responsivity, that is, the asymmetric band structure.

For bias < 0.7V and T > 100K, the dark current increased exponentially with temperature, which suggests that in this temperature range the dark current originates from thermionic emission. The extracted activation energy was 196 meV at zero bias, which was close to the energy corresponding to the cutoff wavelength (193 meV) of the sample. For temperatures lower than 100K, sequential resonant tunneling and phonon-assisted tunneling are probably the dominant components of the dark current. In addition, the photocurrent measured at 77K with the device illuminated only from the room temperature blackbody background radiation is shown

in Figure 3.39. For a bias of 0.5V, the background photocurrent is equal to the dark current at 60K. This indicates that the *background-limited* (BLIP) response is 60K.

Several effects contribute to the noise performance of an infrared photodetector. For a photoconductive detector, such as a QDIP, the primary current noise mechanisms are Johnson noise and *generation-recombination* (GR) noise. Johnson noise, also known as thermal noise or Nyquist noise, is generated in all resistors and can be expressed as $i_{th}^2 = \frac{4kT}{R} \cdot \Delta f$, where R is the resistance and Δf is the noise bandwidth. The GR noise comes from the random generation and recombination of the carriers. The fluctuation of the resulting carrier density can be expressed through $i_{GR}^2 = 4IeG \cdot \Delta f$, where I is the current and G is photoconductive gain. Therefore, the total noise for a photoconductive detector is

$$I_N^2 \left(4IeG + \frac{4kT}{R} \right) \cdot \Delta f$$

The noise current of the QDIPs was measured with low-noise-current preamplifiers and an FFT spectrum analyzer. Figure 3.40 shows the noise current density versus frequency at a bias of 0.3V, and temperatures of 77, 90, 105 and 150K. At low frequency, that is, $f < 2$ Hz, the dominant source of the noise current appears to be $1/f$ noise, which is characterized by an approximate dependence on the reciprocal of the frequency and the square of the current. For $f > 2$ Hz, the noise current is

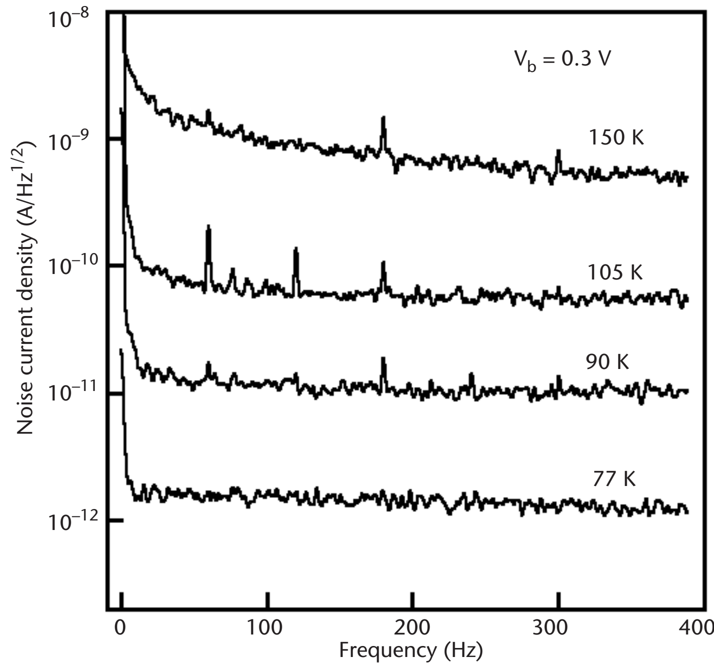


Figure 3.40 Noise spectrum density versus frequency of the (2+1)-ML PIG GaAs capped QDIP measured from 0 to 400 Hz at 0.3V and 77, 90, 105, and 150K.

essentially independent of frequency. This is similar to the GR noise in bulk photoconductors and QWIPs, which leads to random fluctuations in the current density. Consequently, we hypothesize that generation and trapping processes are the dominant noise sources for QDIPs in this spectral region.

Figure 3.41 shows the bias dependence of the noise current at T = 77, 90, 105, 120, and 150K and a measurement frequency of 140 Hz. At 77K, the noise increases from 10^{-14} to 10^{-9} A/Hz$^{1/2}$ as the bias increases from 0 to 0.8V. The calculated thermal noise current, I_{th}, at 77K is also shown (dashed line). The resistance was determined from the slope of the dark current. For V_B = 0.1V at 77K, the calculated thermal noise current (5.5×10^{-14} A/Hz$^{1/2}$) is close to the measured noise current (6×10^{-14} A/Hz$^{1/2}$), indicating that thermal noise is significant in the very low bias region. As the bias increases, the noise current increases much faster than thermal noise. The noise current for $|V_B| > 0.1$V is primarily GR noise.

As a good approximation, the photoconductive gain and the noise gain are equal in a conventional photoconductor and can be expressed as

$$G = \frac{i_n^2}{4e \cdot I_d}$$

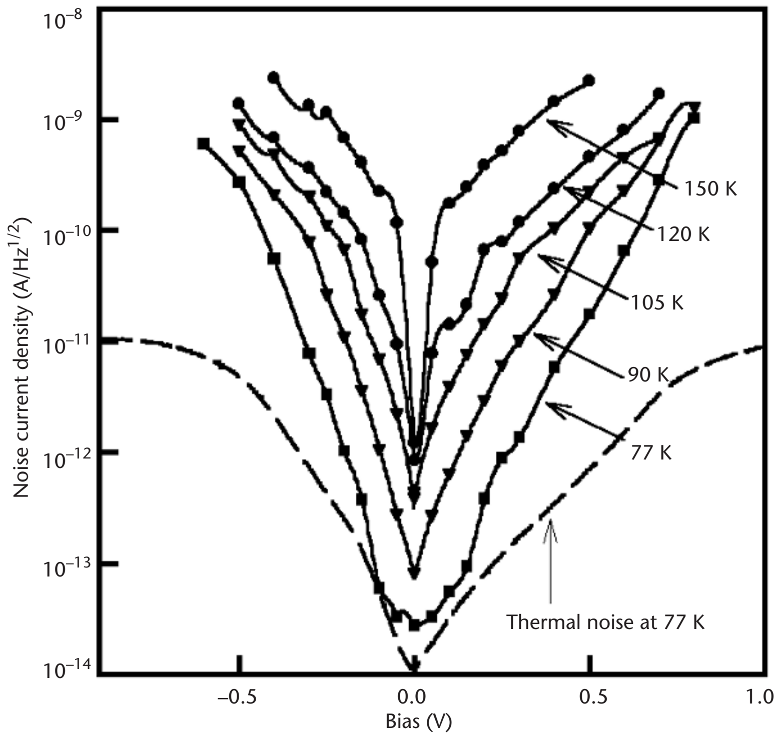


Figure 3.41 Noise spectrum density versus bias of the (2+1)-ML PIG GaAs capped QDIP measured at 77, 90, 105, 120, and 150K. Dots are measured data. The dashed line represents the thermal noise current calculated from the differential resistance of the QDIP.

where i_n is the noise current and I_d is the dark current. The noise gain has been determined using this expression and the measured noise. Figure 3.42 shows the photoconductive gain at 77K. At low bias, the gain depends weakly on voltage. With further increase in bias, the gain increases exponentially with bias. A very large gain of 750 has been obtained at 0.7V, corresponding to an applied electric field of 2.3 kV/cm. QWIPs typically exhibit gains in the range of 0.1 to 50 for similar electric field intensities [50]. The higher gain of the QDIPs is the result of longer carrier lifetimes [51].

We have assumed that there are parallels between the transport mechanisms in QDIPs and QWIPs based on their similar structures and modes of operations. Primary differences are lateral transport between QDs and the fact that the QDs have less than unity fill factors, that is, in a QD layer the space between QDs is comparable to the space occupied by the QDs. To account for this, Philips et al. [34] introduced a fill factor, F, to the gain models for QWIPs established by Liu [52], Beck [53], and Levine et al. [54]. The fill factor can be estimated from AFM data. For the unintentionally doped 2.5-ML InAs/GaAs QDIPs, we observed that the fill factor varied from 0.40 on the first (bottom) QD layer to 0.30 on the fifth (top) QD layer due to variations in the QD size and density. Combining Beck's gain expression

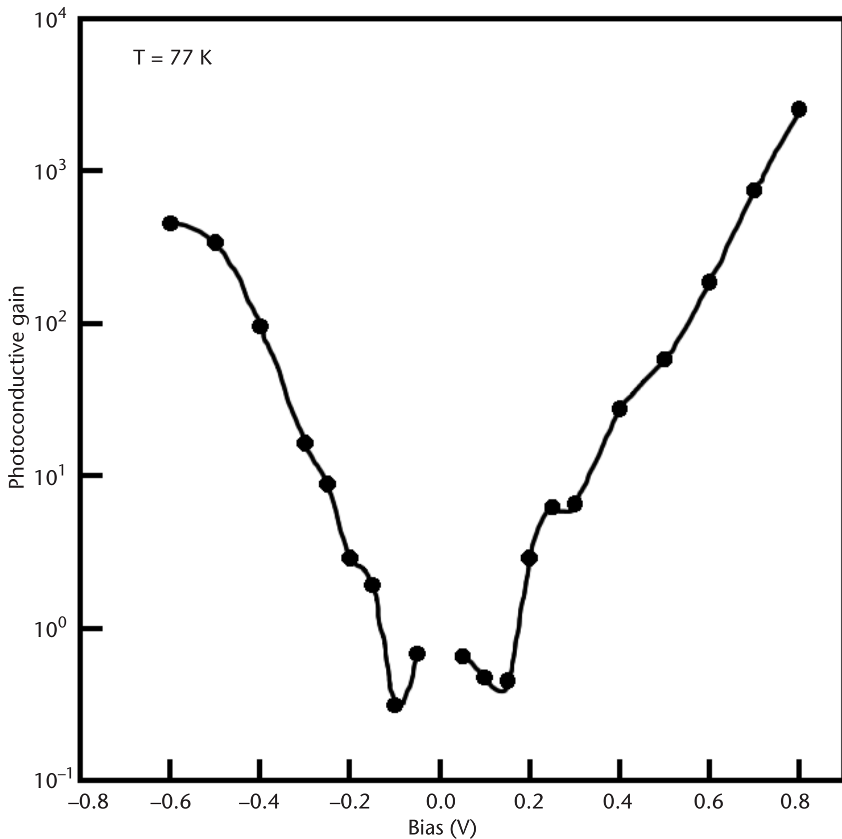


Figure 3.42 Photoconductive gain of the (2+1)-ML PIG GaAs capped QDIP at 77K.

[53], which has been experimentally verified by Schönbein et al. [55], with the fill factor yields the following expression for the photoconductive gain:

$$g = \frac{1 - p_c/2}{FNp_c}$$

where N is number of QD layers and p_c is defined as the probability that a carrier is captured by a QD after its generation. An average value of 0.35 for F was used to calculate the capture probability. Figure 3.43 shows the capture probability versus bias at 77, 105, 120 and 150K. At lower bias, that is, –0.1V, the capture probability is near unity. With increase in negative bias, the capture probability decreases rapidly, which results in a rapid increase in the photoconductive gain. At –0.6V, p_c = 0.0012. Compared to conventional AlGaAs-GaAs QWIPs in which p_c is typically ≅ 0.1, the QDIP demonstrates a gain that is several orders of magnitude larger than the QWIPs [3]. With an increase in temperature, the carrier capture probability decreases rapidly because the electron states in the QDs are most likely filled with electrons at high temperature due to the higher dark current.

The approximation that photoconductive gain is equal to noise gain is valid when the capture probability in Lius [52] and Levine's [54] models is small, that is, when p_c << 1. Choi examined the relationship between photoconductive gain

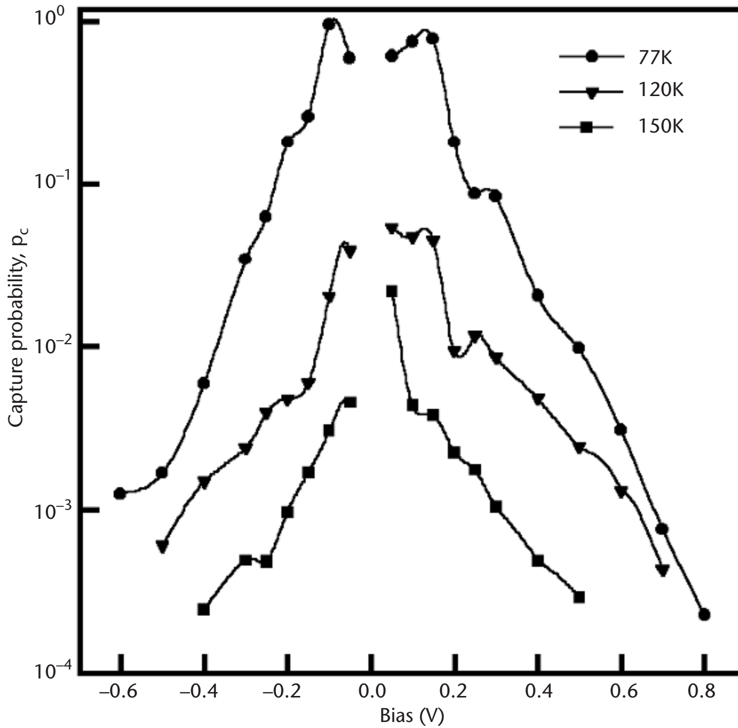


Figure 3.43 Capture probability versus bias of the (2+1)-ML PIG GaAs capped QDIP at 77, 105, 120, and 150K.

and noise gain for QWIPs [56]. His formulation is parameterized in terms of the ratio Γ of the noise gain g_i, to the photoconductive gain g_e; that is, $\Gamma = g_i / g_e$. His analysis shows that the ratio depends on t_w/t_p, where t_w and t_p are the transmitting times across one well and one period, respectively. For small capture probability (p_c < 0.2), the value of Γ is greater than 0.9 and relatively independent of t_w/t_p. The capture probability p_c can be calculated from noise gain and is independent of Γ. We have determined from noise measurements that p_c 0.2 for all bias values in the range 120 to 150K, and for bias $|V_B| > 0.2$V at 77K. For the regime $|V_B| \leq 0.2$V at 77K, we observe that the difference between g_e and g_i [$= (1 - \Gamma)g_e$], is small for the entire range of p_c as a result of the fact that g_e is small in this region. Choi calculated [56] g_e as a function of g_i when $t_w/t_p = 0$, that is, the case for which the largest difference between g_i and g_e would be expected. For the regime where g_e is small, which is the case for our QDIPs when $|V_B| \leq 0.2$V at 77K, he found that the difference is only 3% for g_e = 0.5. Therefore, we conclude that the noise gain provides a good estimation of the photoconductive gain.

The specific detectivity is given by

$$D^* = \frac{R\sqrt{A \cdot \Delta f}}{i_n}$$

where A is the device area, R is the responsivity, i_n is the noise current, and Δf is the bandwidth. Figure 3.44 shows the peak detectivity versus bias at 77 and 100K. The best performance was achieved at 77K and 0.1V where the peak detectivity was 1.4×10^9 cmHz$^{1/2}$/W. The corresponding responsivity was 12 mA/W. With an increase in temperature to 100K, the peak detectivity dropped to 1.4×10^8 cmHz$^{1/2}$/W at 0.1V, due to the decrease in responsivity and increase in noise current.

3.3.3 QDIPs with AlGaAs Blocking Layers

In the preceding section we have demonstrated that lower dark current and higher detectivity can be achieved with QDIP structures that employ unintentionally doped active regions. It has been postulated that the relatively high dark current exhibited by QDIPs is due to (1) nonuniform dopant distributions and background impurities in GaAs [57] and (2) to "leakage" in the spaces between the QDs. It has been shown that a current blocking AlGaAs layer can be effectively used to improve on the dark current and D^* [23, 42–44, 58]. Lin et al. [43] and Stiff et al. [44, 45] have reported on QDIPs with a single AlGaAs blocking layer on one side of the InAs/GaAs QD layers. Wang et al. [42] introduced a thin AlGaAs barrier layer between the InAs QDs. This layer filled the area between the dots but left the top of the dots uncovered. An improvement in the detectivity relative to similar devices without the AlGaAs barrier layer was demonstrated.

As detailed in an earlier section, our AlGaAs capped QDIP structures involve effective AlGaAs confinement layers below the quantum dot layers and on the top of the GaAs cap layers (Figure 3.45). We recall that the AlGaAs confinement layers consisted of four pairs of AlAs/GaAs (1-ML/4-ML) short-period superlattices to

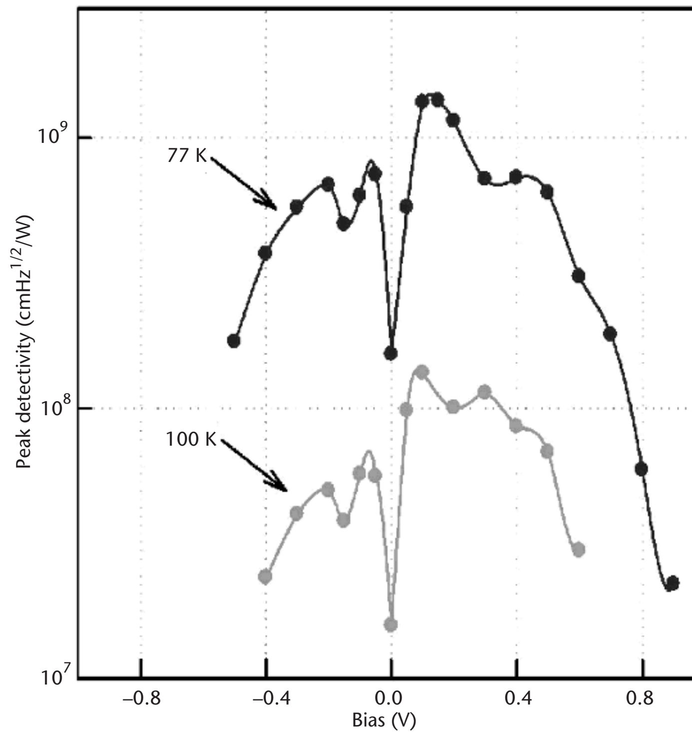


Figure 3.44 Peak directivity versus bias of the (2+1)-ML PIG GaAs capped QDIP at 77 and 100K.

mimic close to 20% effective Al concentration AlGaAs alloy. The conduction band energy of the active region is shown in Figure 3.45.

Figure 3.46 shows the spectral response at 0.8V bias and at temperatures of 63, 77, and 100K. Compared to the spectral response of similar QDIPs without AlGaAs blocking layers (see Figure 3.35), we note that these spectra are narrower and the spectral peaks have shifted to shorter wavelengths. The photoresponse peaks occur at 6.2 μm for all three spectra compared to >7 μm for QDIPs without the AlGaAs layers. This is due to the fact that the AlGaAs layers increase the effective well depth. The FWHM of the spectrum, $\Delta\lambda$, is ~0.4 μm, from which it follows that $\Delta\lambda/\lambda$ = 7.5%. The observed spectral width shows that the AlGaAs layers contribute to more uniform size distributions, in particular reducing variations between the different QD layers. This is illustrated in Figure 3.15, which shows cross-sectional TEM images of QDIPs with GaAs and AlGaAs cap layers. It is clear that the size variations between the first and fifth QD layers are larger for the GaAs-capped wafers than for the AlGaAs-capped wafers.

The normalized dark current versus bias for temperatures in the range of 20 to 300K is shown in Figure 3.47. As was observed for the GaAs cap QDIPs, the dark current exhibited significant increase with bias and temperature; however, the magnitude of the dark current was lower by up to three orders of magnitude. At low temperature, for example, 20K, the dark current increased rapidly as the bias was

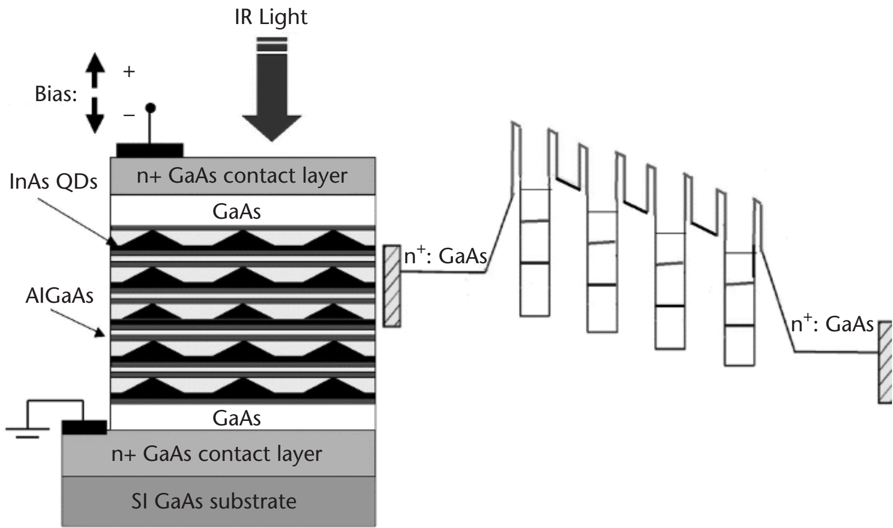


Figure 3.45 Schematic cross section and conduction band structure of QDIP with AlGaAs blocking layers.

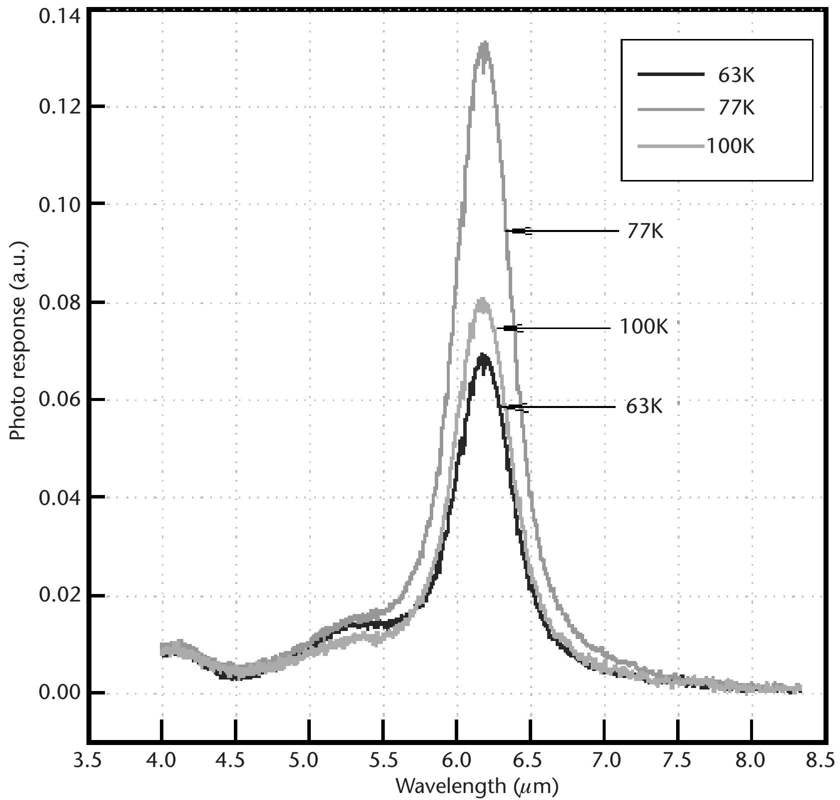


Figure 3.46 Spectral response of (2+1)-ML PIG GaAs capped QDIP with AlGaAs blocking layers.

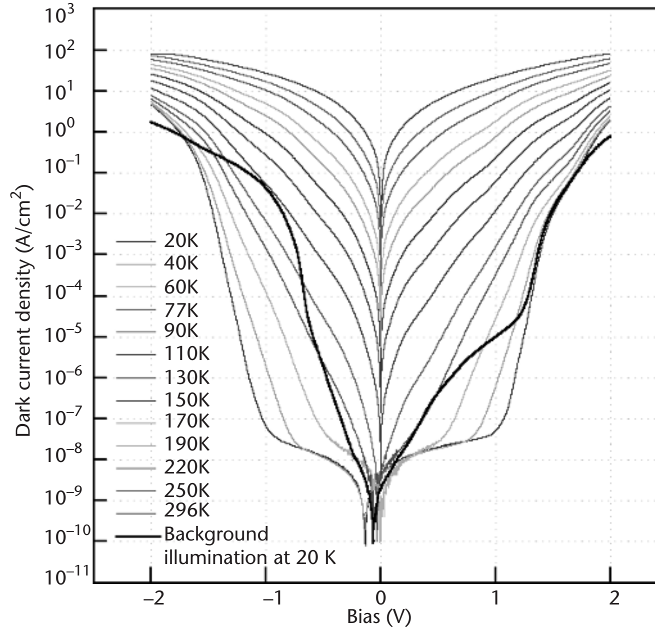


Figure 3.47 Dark current density of (2+1)-ML PIG GaAs capped QDIP with AlGaAs blocking layer, including photocurrent induced by blackbody background radiation.

increased from 0 to ±0.2V. This can be attributed to the rapid increase of electron tunneling between the QDs. For $|0.2| \leq V_{bias} \leq |1.0|$, the dark current increased slowly. With further increase in bias, the dark current demonstrated a steep increase, which was largely due to lowering of the potential barriers. At 0.7V bias, the dark current density increased from 2.5×10^{-7} A/cm^2 at 60K to 11 A/cm^2 at room temperature. Also shown is the photocurrent induced by the room temperature blackbody background radiation. It is clear that the background limited response temperature varies with bias. At –1V the BLIP temperature is 110K, which is higher than the value of 60K observed for the GaAs-capped QDIPs (Figure 3.39).

Figure 3.48 compares the peak spectral responsivity [Figure 3.48(a)] and the noise current [Figure 3.48(b)] versus bias at temperatures of 77 and 100K for QDIPs with and without the AlGaAs blocking layers. It is clear that for a given voltage, the responsivity of the QDIP with the blocking layers is three to four orders of magnitude lower than the QDIPs described in the preceding section. However, comparable responsivity can be achieved by increasing the bias voltage. Similarly the noise current of the QDIP with AlGaAs blocking layers is much lower than the QDIP without the AlGaAs layers. Close inspection of the figures reveals that the decrease in the noise current is greater than the decrease in the responsivity. This may be due to an enhancement in impact ionization, which is enabled by the increased operating voltage that results from the lower dark current [43]. Consequently, a net improvement in the *signal-to-noise ratio* (SNR) is expected. This is seen in the detectivity curves in Figure 3.49. The best performance was achieved at 77K and –0.7V where the peak detectivity was 10^{10} cmHz$^{1/2}$/W. The corresponding responsivity was 14 mA/W. With

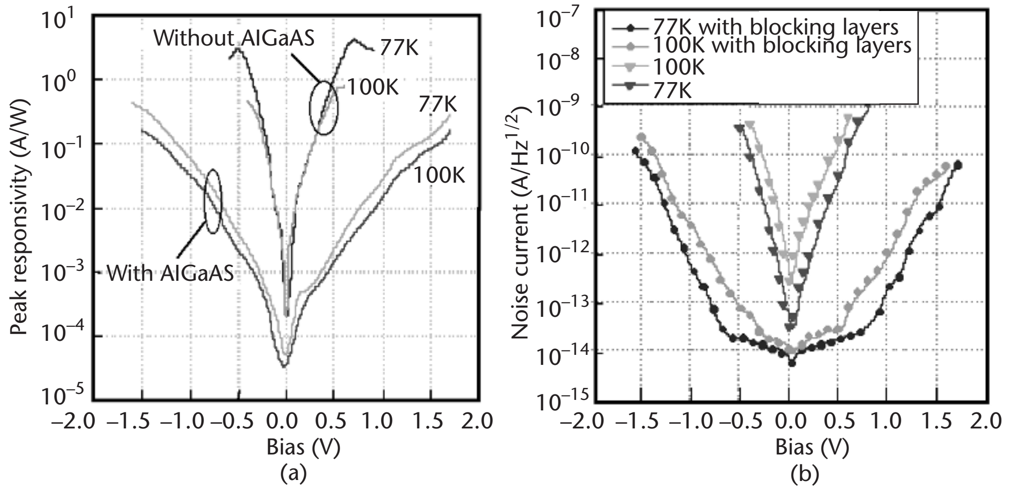


Figure 3.48 (a) Peak responsivity and (b) noise current versus bias at 77 and 100K for (2+1)-ML PIG GaAs capped QDIPs with and without AlGaAs blocking layers.

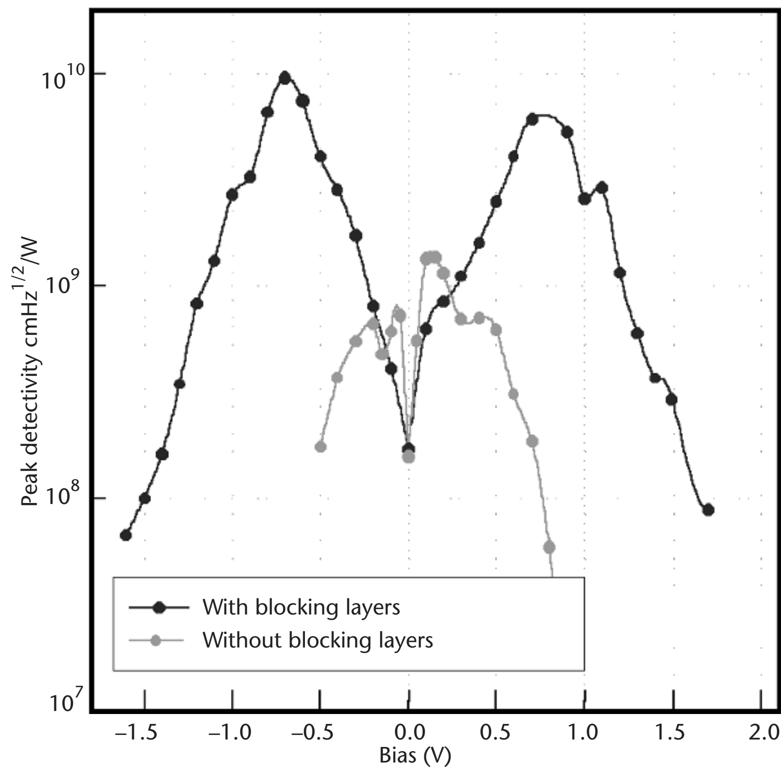


Figure 3.49 Detectivity of (2+1)-ML PIG QDIPs with and without AlGaAs blocking layers at 77K.

increase in temperature to 100K, the peak detectivity dropped to 1.1×10^9 cmHz$^{1/2}$/W at 0.5V, due to the decrease in responsivity and increase in noise current.

3.3.4 InAs/InGaAs/GaAs QDIPs

While high detectivity was achieved using the AlGaAs blocking layers, it is desirable to demonstrate operation in the LWIR region. However, all QDIPs reported to date that have operating wavelengths > 8 μm have exhibited normal-incidence peak detectivities below 10^9 cmHz$^{1/2}$/W at ~77K. Two approaches to tailoring the detection bands of QDIPs have been reported. To date, the prevalent technique utilizes InGaAs for strained epitaxial island formation with capping layers of $Al_xGa_{1-x}As$ [38, 59–61]. An alternative promising approach for the tailoring of the QD electronic structure is to modify the layers below and above the nominally binary InAs coherent islands. As we detailed earlier, replacing GaAs capping layers by $In_xGa_{1-x}As$ layers allows partial strain relief as well as chemical modification of the confinement potentials to realize LWIR intraband transitions in InAs QDs [17, 25, 26, 62–64]. In this section we describe the operation and characteristics of QDIPs that have been fabricated with InAs QDs and InGaAs cap layers. Not only were we able to obtain a response in the LWIR window of 8 to 12 m but, as shown in this subsection, we realized detectivity as high as 3×10^{11} cmHz$^{1/2}$/W at 77K for operation at ~9 μm.

Small InAs QD-Based QDIPs

Figure 3.50 shows a schematic of the QDIP structure and the conduction band diagrams. Five layers of nominally 2.0-ML InAs at a growth rate of ~0.22 ML/sec at ~500°C were inserted between highly Si-doped top and bottom GaAs contact layers. Then 20-ML $In_{0.15}Ga_{0.85}As$ regions were grown via MEE at ~350°C as the quantum dot cap layers followed by 20 ML of MEE-grown GaAs. An additional 160 ML of GaAs was grown via MBE at 500°C for a total of 180-ML GaAs spacer layers. The GaAs layers between the contact layers and the nearest quantum dot layer had a thickness of 220–240 ML.

Figure 3.51(a) shows the spectral response at 77K for bias voltages in the range of 0 to 0.8V. The intraband photoresponse peak occurred at ~8.8 μm (141 meV) for bias larger than 0.2V. We note that the conduction band offset for the GaAs/$In_{0.15}Ga_{0.85}As$/GaAs QWs is estimated in the literature to be 145 ± 15 meV. Although this value is close to the 8.8-μm (141-meV) photoresponse, this response (more precisely, 8.22 μm at –0.8V) does not arise from the GaAs/20-ML $In_{0.15}Ga_{0.85}As$/GaAs QW regions between the InAs QD regions. This is supported by earlier work [17] on infrared detector structures that contain wider GaAs/30-ML $In_{0.15}Ga_{0.85}As$/GaAs QW regions between QDs, and show a red shift (from the 8.22-μm response of the current GaAs/20-ML $In_{0.15}Ga_{0.85}As$/GaAs containing sample) to 9.3 μm (at –0.8V) for the 8.8-μm photoresponse. This is in contrast to the blue shift that should result if the origin of this photoresponse peak were in the QW region between the QDs. Thus, the 8.8-μm photoresponse arises from the InAs QDs.

The band diagram under negative bias is shown schematically in Figure 3.50. For negative bias, the photoresponse peak shifted to 8.3 μm [Figure 3.51(b)], a

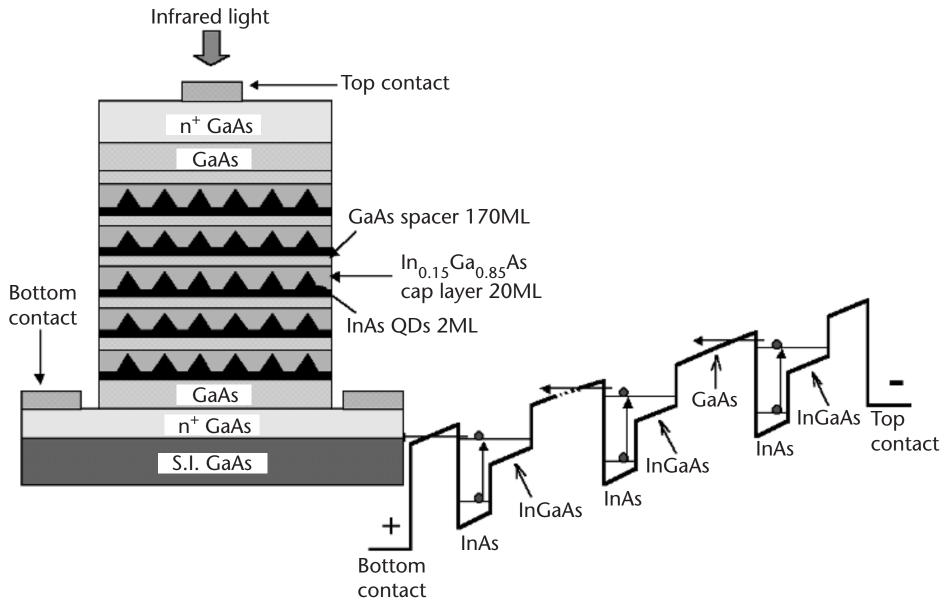


Figure 3.50 Schematic cross section and conduction band structure of the InAs/InGaAs QDIP.

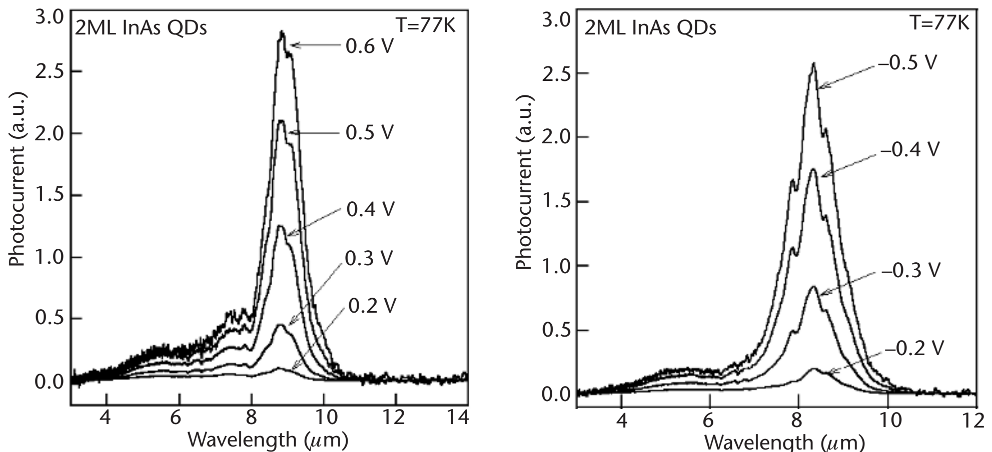


Figure 3.51 Spectral response of InAs/InGaAs QDIPs with (a) positive bias and (b) negative bias.

manifestation of the asymmetric band diagram described above. The FWHM was ~0.9 μm at –0.2V and increased to 1.2 μm at –0.5V. The corresponding values of $\Delta\lambda/\lambda$ were 11% and 14%, respectively. The photocurrent increased significantly with voltage for both positive and negative bias because the resulting increase in dark current populated the quantum dot ground states. These electrons participated in the infrared photon absorption and contributed to the photocurrent.

Figure 3.52 shows the peak responsivity versus bias at temperatures of 77, 100, and 120K. At 77K, with increase in positive bias, the responsivity increased from 13 mA/W at 0.2V to 660 mA/W at 0.9V. For negative bias, the responsivity increased from 22 mA/W at –0.2V to 600 mA/W at –0.9V. Negative differential responsivity was observed with further increase in voltage, for both the positive and negative bias cases. This is a result of "overfilling" the quantum dots.

Figure 3.53 shows the dark current density versus voltage characteristics for temperature in the range of 20 to 296K. As the temperature increased from 20K to room temperature, the dark current density increased more than 10 orders of magnitude from 9.5×10^{-9} to 9.6 A/cm^2 at 0.3V and from 6.2×10^{-9} to 9.6 A/cm^2 at –0.3V. The logarithm of the dark current density versus inverse temperature (1,000/T) is plotted (0–1V) in Figure 3.54(a), whereas Figure 3.54(b) shows an expanded view of the dark current at –0.1V. At –0.1V, excellent linear fits were obtained more than seven orders of magnitude for temperatures higher than 77K. The exponential increase in the dark current suggests that carrier thermal excitation to higher energy states is involved.

For temperature lower than 77K, the dark current density is relatively insensitive to temperature, an attribute of phonon-assisted tunneling and sequential resonant tunneling. From Figure 3.54(b), we can extract an activation energy

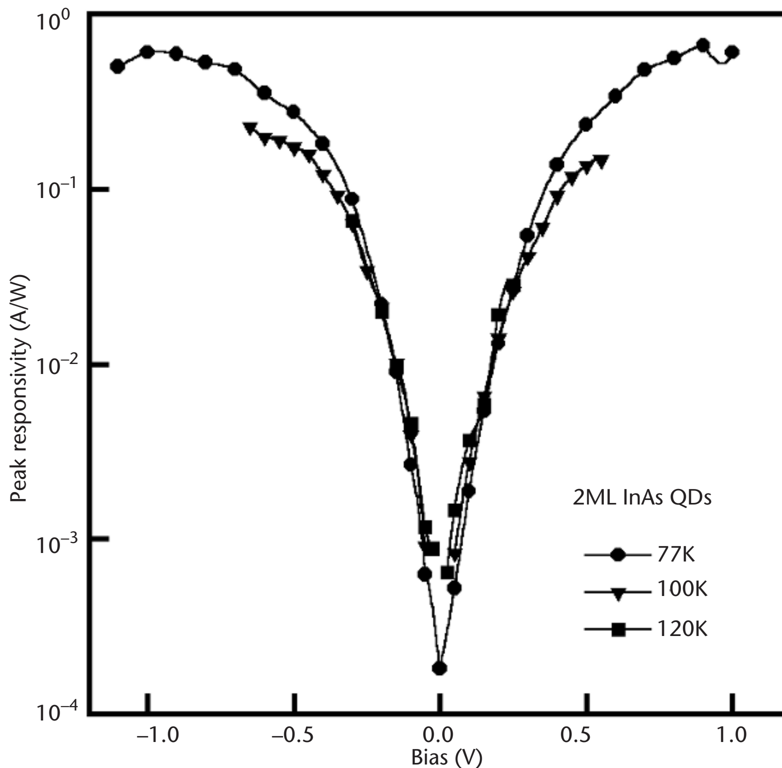


Figure 3.52 Absolute peak responsivities measured at 77, 100, and 120K.

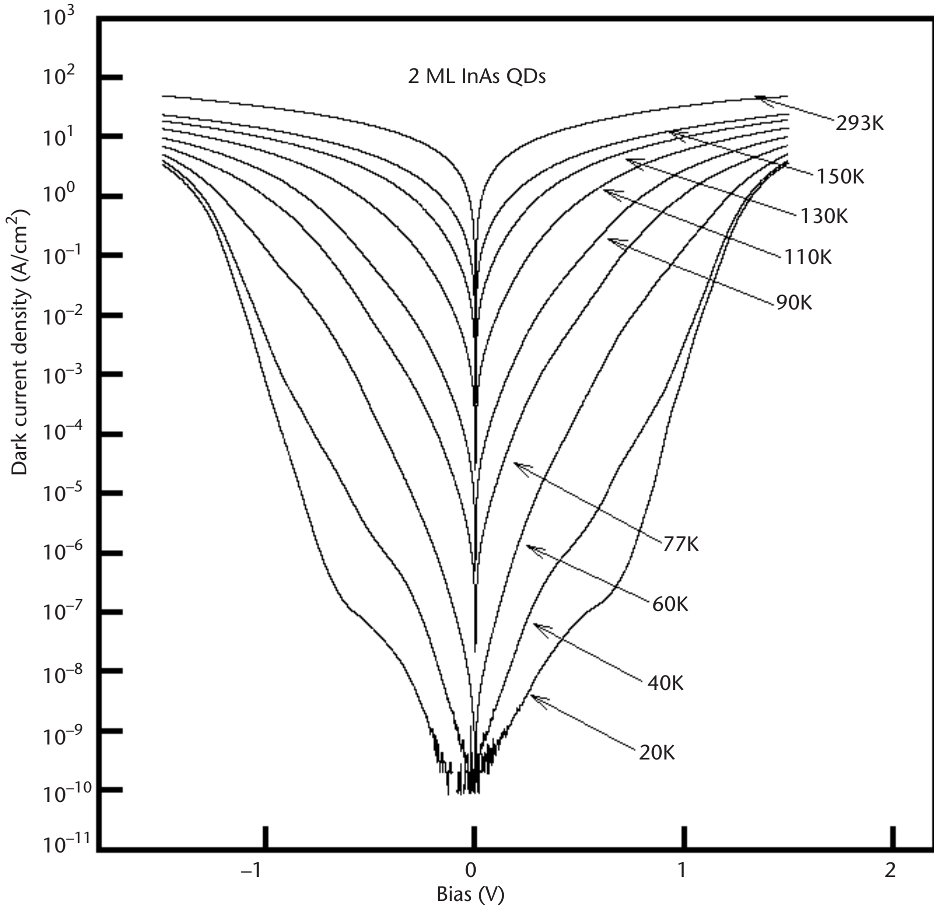


Figure 3.53 Dark-current density versus voltage for temperature in the range of 20 to 296K.

of 135 meV. This is, interestingly, essentially the same as the FTIR measured cut-off energy E_c = 138 meV (wavelength of 9 μm). This coincidence indicates that at zero bias the dark current is carried through thermal activation to the same excited states of the small QDs as participating in the absorption and subsequent rapid tunneling. The activation energy decreased linearly with bias, as shown in Figure 3.55. At high bias, the activation energy is close to kT, which resulted in high dark current even at low temperature. Compared to a similar structure with 2.5-ML InAs larger quantum dots covered with 30-ML $In_{0.15}Ga_{0.85}As$ cap layers to be discussed shortly, the dark current density was an order of magnitude higher [26, 31, 47, 65]. This increase in the dark current results from the reduction of the thickness of the $In_{0.15}Ga_{0.85}$ cap layers and from the decrease in binding energy of the ground states.

Figure 3.56 shows the noise current of a 250-μm-diameter device at 77K (solid circles) and 100K (open circles) versus bias voltage. For low bias, that is, $|V_B| < 0.1$V, the measurement was limited by the noise floor of the instruments. The calculated

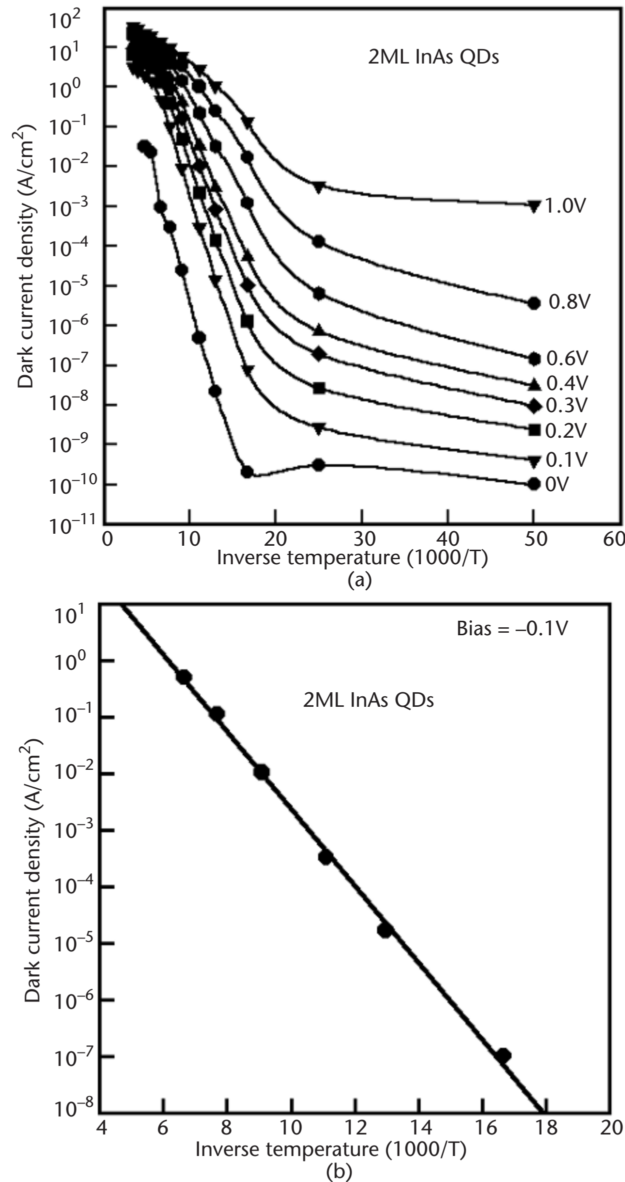


Figure 3.54 (a) Dark current density versus inverse temperature for 20K ≤ T ≤ 293K. (b) Expanded views for –0.1 V bias. The slope gives an activation energy of E_a = 135 meV.

thermal noise current, I_{th}, at 77 and 100K is also shown. At V_B = 0.1V and 77K, the calculated thermal noise current (3×10^{-14} A/Hz$^{1/2}$) was close to the measured noise current (6×10^{-14} A/Hz$^{1/2}$), indicating that thermal noise is significant in the low-bias region. As the bias increased, the noise current increased much faster than thermal noise. The noise current at high bias ($|V_B| > 0.1$V) was dominated by generation-recombination (GR) noise.

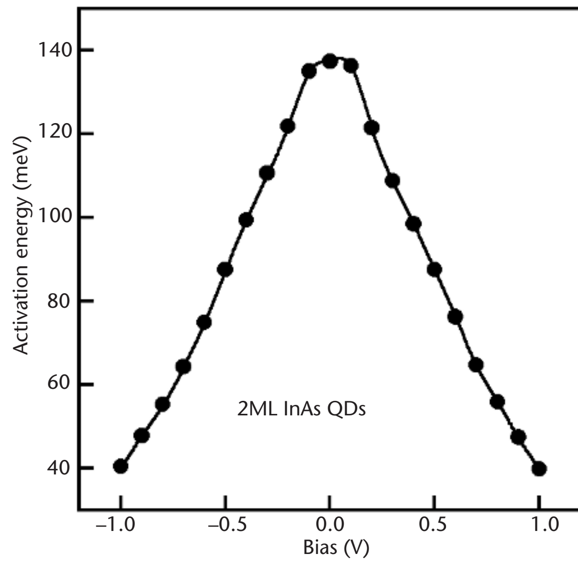


Figure 3.55 Dark current activation energy versus bias.

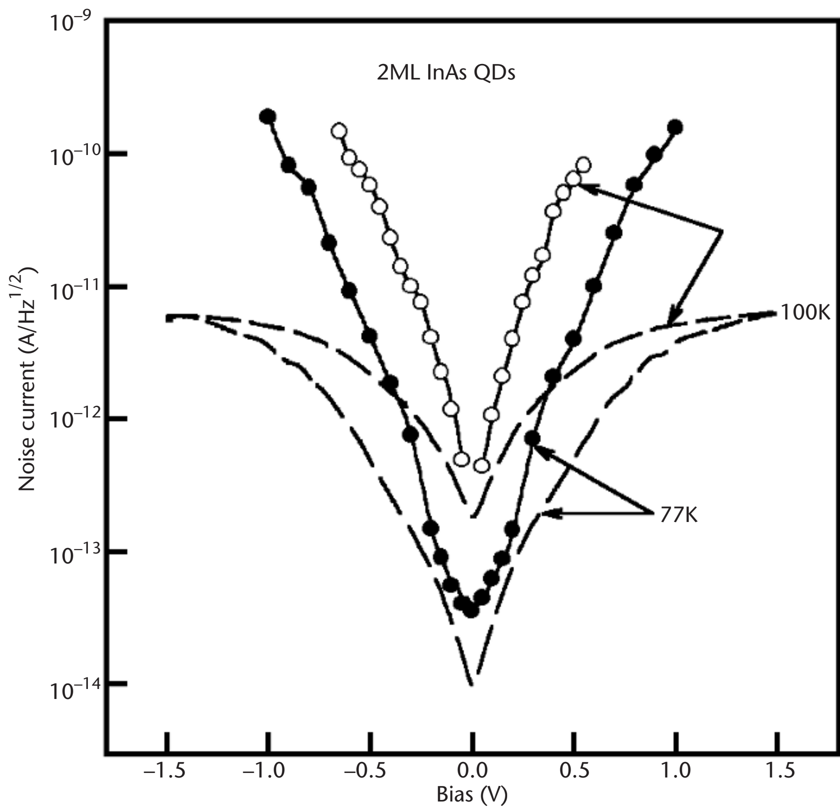


Figure 3.56 Measured noise current versus bias at 77 and 100K. The dashed lines show calculated thermal noise current at 77 and 100K.

The external quantum efficiency, η_{ext}, is the ratio of the number of countable output carriers to the number of incident photons. It can be expressed as follows:

$$\eta_{ext} = \frac{\frac{i_{ph}}{e}}{\frac{P_{in}}{h\frac{c}{\lambda}}} = R \cdot \frac{hc}{\lambda e}$$

where I_{ph} is the photocurrent, P_{in} is the input power, λ is the peak wavelength, and R is the peak responsivity. Figure 3.57 shows the quantum efficiency versus bias at 77K. With increase in bias, η_{ext} increased linearly and then saturated at high bias reaching a maximum value of $\eta_{ext} \approx 9\%$. The internal quantum efficiency, η_{int}, can be estimated by using the following relation:

$$\eta_{int} = \frac{\eta_{ext}}{G \cdot (1-R)}$$

where R (8.8 μm) = 28% is the reflectivity of the GaAs contact layers and G is the photoconductive gain, which was determined from the dark current and the noise

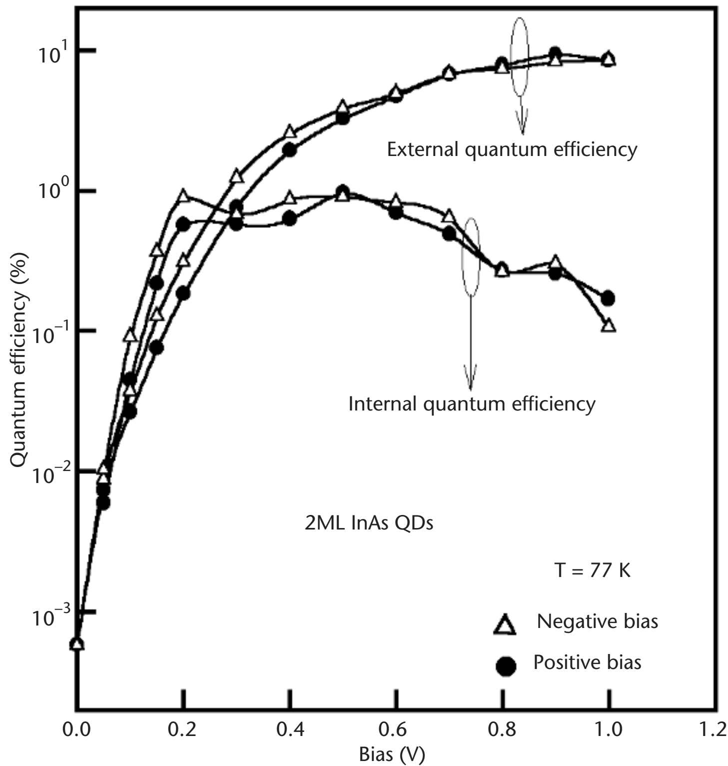


Figure 3.57 Internal and external quantum efficiency versus bias at 77K.

current. At low bias, η_{int} increased with bias due to filling of the ground states. However, at high bias, η_{int} slowly decreased due to the overfilling effect discussed earlier.

Figure 3.58 shows the peak detectivity for the 8.3- and 8.8-μm peaks at 77 and 100K. The best performance was achieved at 77K and –0.2V where the peak detectivity was 3.2×10^9 cmHz$^{1/2}$/W. The corresponding responsivity was 22 mA/W. Note the rapid decrease of the detectivity with increase in temperature for both peaks. Owing primarily to the rapid increase in GR noise current at 100K, the detectivity of the 8.3-μm peak dropped to 1.5×10^8 cmHz$^{1/2}$/W at –0.35V.

High-Detectivity QDIPs

To further enhance the detectivity, the number of QD layers was doubled to 10 [66]. This should enhance absorption and reduce the dark current. In Figure 3.59, the inset shows the normal-incidence spectral response at 78K for bias voltages in the range from –1.0 to –1.3V. The intraband photoresponse peak occurred at ~9.3 μm (133 meV) and ~8.7 μm (143 meV) for positive (not shown) and negative bias, respectively. The spectral width is 0.84 μm and $\Delta\lambda/\lambda$~10%, which is characteristic of electron bound-to-bound intraband transitions within the conduction band [6]. Figure 3.59 shows the peak responsivities at 78 and 100K for bias voltages ranging from ±0.4 to ±2.1V. At 78K the responsivity increased nearly exponentially with bias, that is, from ~0.018 mA/W at 0.4V to 710 mA/W at 2.1V. Saturation of the responsivity occurred beyond 2.0V.

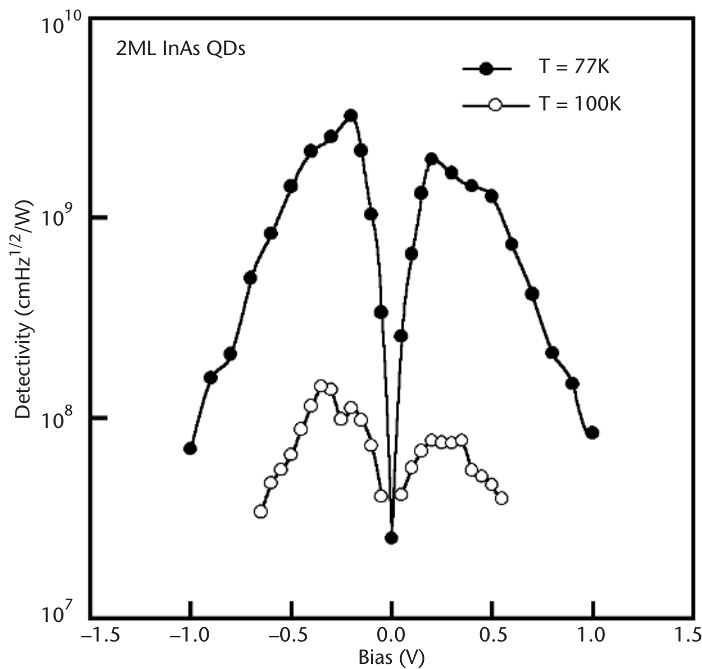


Figure 3.58 Detectivity versus bias at 77 and 100K.

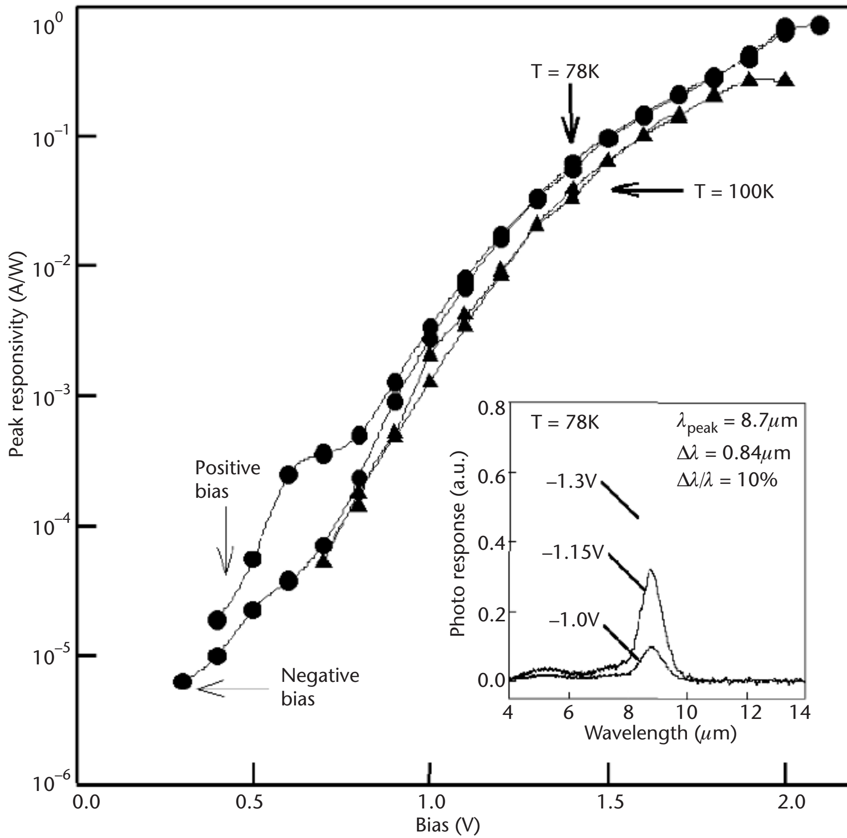


Figure 3.59 Peak responsivity of the device at 78 and 100K. The inset shows spectral response of the 250-μm-diameter QDIP at 78K.

Figure 3.60 shows the dark current for various temperatures in the range from 78K to room temperature. At a bias of 1.0V, the dark current increased more than eight orders of magnitude from 6.5×10^{-13} A at 78K to 4.9×10^{-4} A at 293K. At 78K, for bias voltage less than $|\pm 0.9|$V, the dark current measurement was instrumentation limited. Also shown in Figure 3.60 is the dark current measured under 300K background radiation (dashed line). Figure 3.61 shows the bias dependence of the noise current at T = 78K and 100K and a measurement frequency of 140 Hz. The noise was dominated by GR noise. The noise measurement was limited by the noise floor of the preamplifier for bias less than $|V_B| \leq |\pm 1.4\text{V}|$.

Figure 3.62 shows the peak detectivity as a function of applied bias at 78K. For the determination of the detectivity, the instrument-limited noise values were utilized. This overestimates the noise current and thus provides a conservative lower limit for the detectivity. For the 9.3-μm photoresponse peak, the best performance was achieved at 1.4V where the detectivity is 3×10^{11} cmHz$^{1/2}$/W with the corresponding responsivity of 61 mA/W. At −1.4V, the device has a

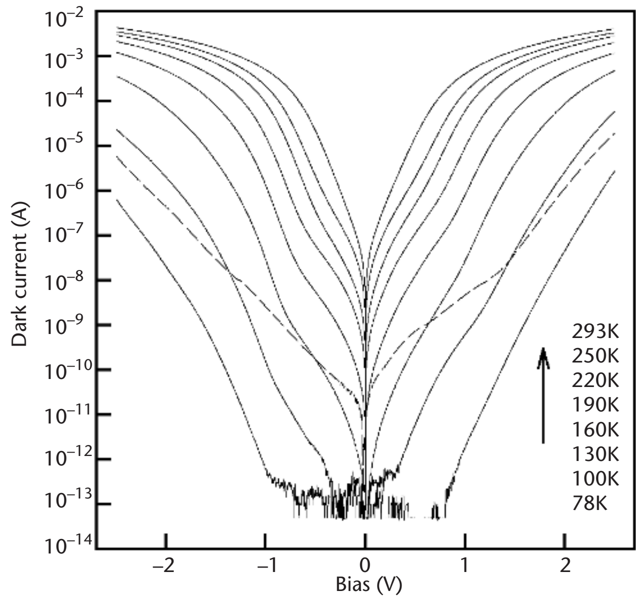


Figure 3.60 Dark current versus bias characteristics at temperatures ranging from 78 to 293K. The dashed line represents the response to 300K background radiation.

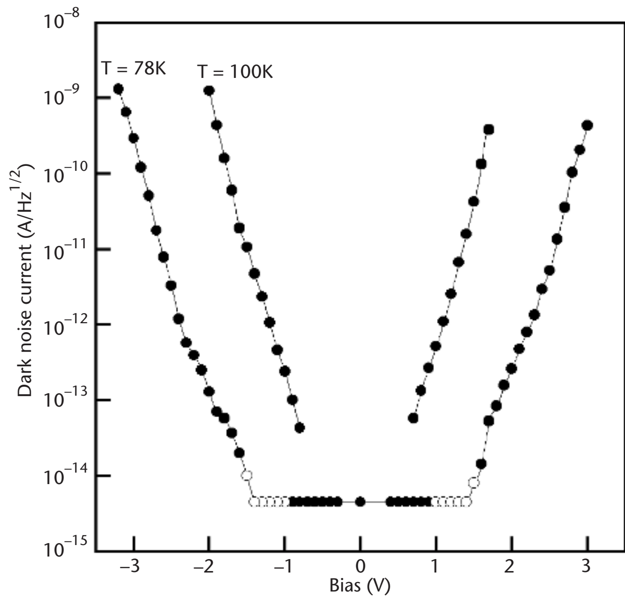


Figure 3.61 Dark noise current as a function of applied bias at 78 and 100K.

detectivity of 2.7×10^{11} cmHz$^{1/2}$/W with the corresponding responsivity of 55 mA/W for the 8.7-μm photoresponse peak.

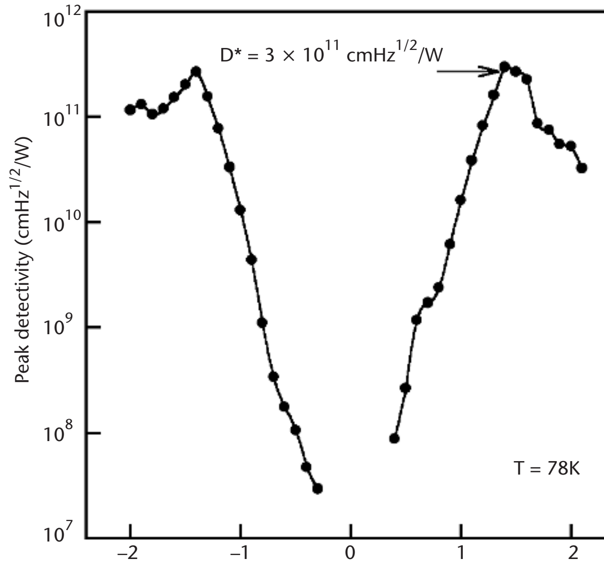


Figure 3.62 Detectivity of the device as a function of applied bias at 78K. The highest detectivity is 3×10^{11} cmHz$^{1/2}$/W at 1.4V.

The detectivity at these long wavelengths in the atmospheric window is nearly two orders of magnitude higher compared to the value of ~5×10^9 cmHz$^{1/2}$/W noted earlier for the 5-layer QDIP. Indeed, it is comparable [67] to that of QWIPs at ~78K. This enhancement is attributable to the increased absorption accompanying the increased total number of QD layers as manifest in the enhanced saturation value of ~700 mA/W of the responsivity of the 10-layer QDIP compared to that of the 5-layer QDIP. Additionally, an enhanced dark current blocking effect arises from two contributions. One, with an increasing number of QD layers, the vertical self-organization between QDs is weakened for a spacer thickness optimized to reduce the total strain buildup so as to minimize the density of structural defects. The TEM image of the 10-layer QDIP shown in Figure 3.63 is consistent with this suggestion. The loss of vertical organization is desirable because it effectively reduces the area for current leakage pathway in between the QDs for the same per-layer QD density. Second, the change in the self-consistent band bending with increasing total thickness affects the occupancy of the energy levels of the QDs. At the same electric field, the dark current of the 10-layer QDIP is four to seven orders of magnitude smaller than that of the 5-layer QDIPs.

3.3.5 Dual-Color QDIPs

In this section we describe the photodetector characteristics of QDIPs with bias-controllable dual-wavelength photoresponse [17, 31, 47]. The first devices of this

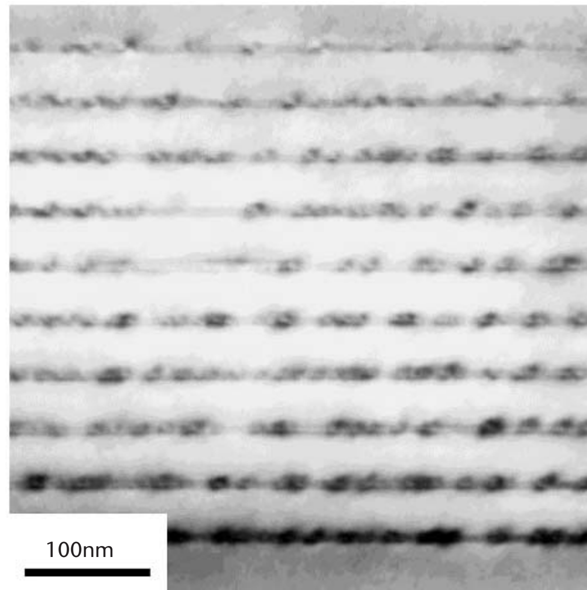


Figure 3.63 Shows a cross-sectional TEM image of the 10-period QDIP.

type were achieved with structures described in Section 3.2.3 that incorporated the 2.5-ML slow-growth-rate-induced large InAs quantum dots in multiple layers that led to two different quantum dot size distributions, predominantly vertically [17, 31, 47]. Subsequently we have utilized VDA structures involving sequential stacks of QDs with intentionally varied transition energies.

For the 2.5-ML InAs/InGaAs capped five-layer stack-based QDIPs of the previous section, the dark current at 80K is about two orders of magnitude lower than that found for InAs/GaAs QD-based QDIPs with a similar *n-i-n* structure. This reduction is mainly due to thicker spacer layers and the deeper binding energies of QD electron ground states in this *n-i-n* QDIP structure. Figure 3.64 shows the MWIR and LWIR intraband photocurrent behavior of the QDIP as a function of negative and positive bias. Note the emergence of a dual-peak response with increasing bias magnitude with a mid-IR peak ~5.6 μm (221 meV) and a long IR peak ~10 μm (124 meV). A voltage-controlled, tunable, dual-wavelength behavior is manifest in the bias dependence of these two photocurrent peak intensities shown in Figure 3.64(c). At zero bias, no photocurrent was detected for the long-wavelength (~10-μm) radiation, whereas the photocurrent for mid-IR wavelength (~5.6-μm) radiation is not zero. This kind of intraband photovoltaic effect (at ~5.6 μm) is due to the QD intraband-transition-induced dipole moment, which originates from the shape- and inhomogeneous strain-dependent intrinsically asymmetrical potential of the pyramidal QDs [68].

Because larger QDs have lower electron ground-state energy, and are thus preferentially occupied at zero bias, we assign the ~5.6-μm photoresponse to the larger QDs. With increasing bias, the peak intensities of both intraband

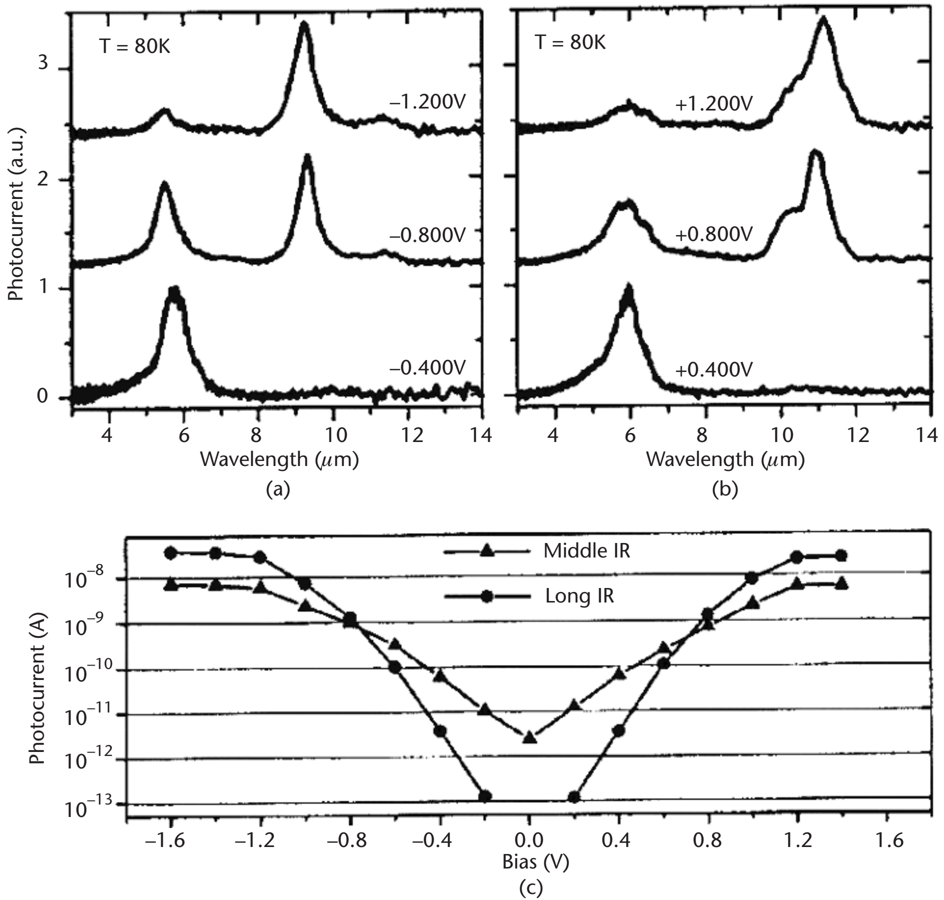


Figure 3.64 Photoresponse of 2.5-ML InGaAs capped QDIP at 0.4, 0.8, and 1.2V; –0.4, –0.8, and –1.2V. (a) Negative bias; (b) positive bias; and (c) the photocurrent versus bias plot.

photocurrents increase rapidly, become comparable at intermediate bias (~0.8V), and saturate at high bias. Note that the ratio of the middle- and long-wavelength photocurrents can be varied by up to two orders of magnitude with reasonable change in bias. Significant intraband shift (5.5 μm to 6.0 μm and 9.2 μm to 11.2 μm) is observed by varying bias (~1.60 to +1.40V) for both peaks. This intraband peak shift may mainly be contributed by the quantum-confined Stark effect. Importantly, we note that the long-wavelength (~9.3-μm) photocurrent peak has a minimum FWHM of 8.2 meV, corresponding to $\Delta\lambda/\lambda = 6\%$, at a bias of 0.800V. The extremely narrow, minimum FWHM may be due to collective modes induced by enhanced interdot electron–electron coupling [32, 33], since the InGaAs capping layer significantly decreases the lateral tunneling barrier between the small QDs with size and shape fluctuations.

The bias dependence of the intraband photocurrent of the *n*-*i*(2.5-ML QD)-*n* QDIP shown in Figure 3.64(c) can be understood as follows. At zero bias, the electron ground-state occupation of small QDs is much lower than that of large QDs at 80K since their electron ground states are higher by ~40 meV (as discussed later). With increasing bias, the electron ground states of the small QDs become increasingly occupied. The gain of photoexcited electrons from the small QDs probably also increases more rapidly with bias than that of the large QDs. At high bias, therefore, the intraband photocurrent of the small QDs exceeds that of the large QDs.

The absolute spectral responsivity at 77K is shown in Figure 3.65. With increase in bias, the responsivities of all four peaks increase four orders of magnitude from ~10^{-4} A/W at low bias to ~1 A/W at –1.3V. For negative bias, the crossover of the responsivities of the two peaks occurs at about –0.8V, which is consistent with the FTIR measurement. Figure 3.66 shows dark current density versus voltage (*I*-*V*) characteristics for the sample in a cold shield with temperatures ranging from 40 to 296K. The photocurrent that originates from the background blackbody radiation at room temperature is shown by the dashed line. Similar to the GaAs QDIPs discussed earlier, the BLIP temperature is ~60K. At 77K, the dark current density is 1.3×10^{-6} A/cm^2 and 4.4×10^{-3} A/cm^2 at –0.3V and –0.8V, respectively. At low bias and $T > 77$K, the dark current increases exponentially by 10^6 with increasing temperature, which suggests that in this temperature range the dark current originates from thermionic emission. For temperatures lower than 77K, sequential resonant

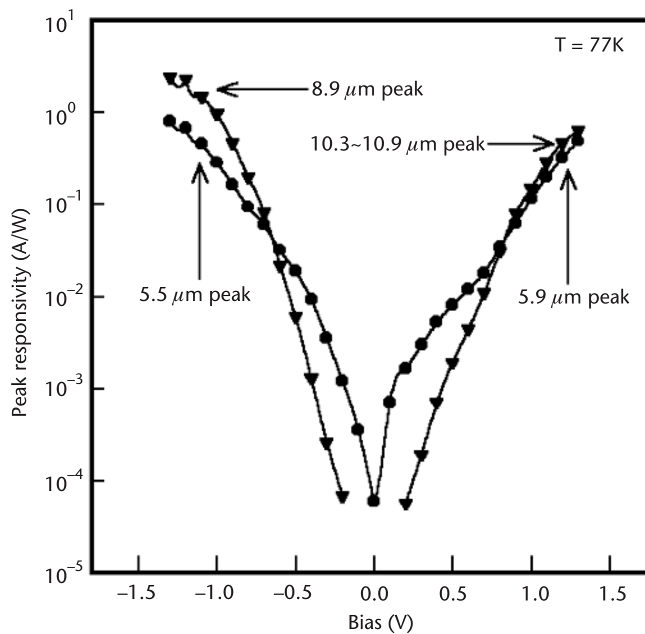


Figure 3.65 Peak responsivity for 2.5-ML InGaAs capped QDIP at 77K.

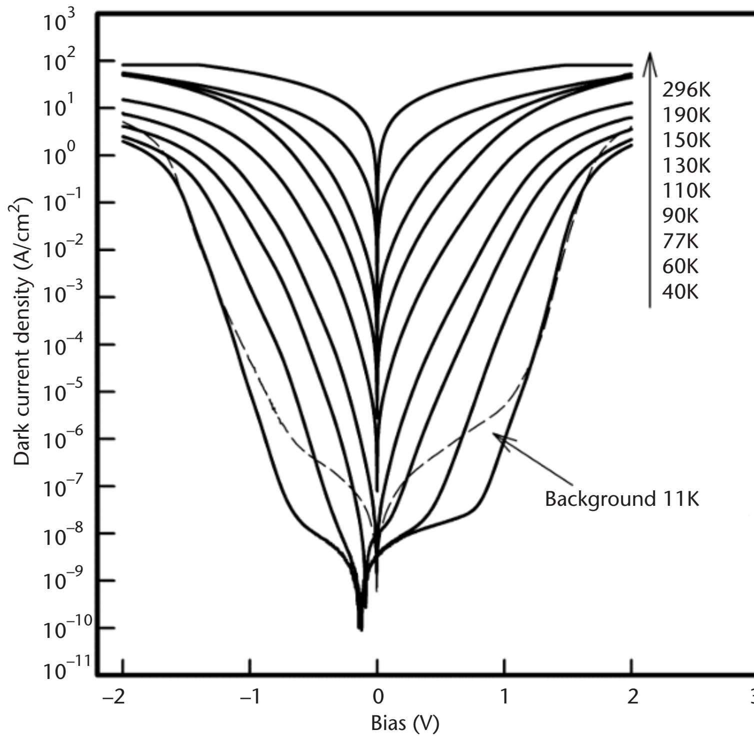


Figure 3.66 Dark current of 2.5-ML InGaAs capped QDIP versus bias for temperatures from 40 to 296K. The dashed line shows the photocurrent that originates from the blackbody background at room temperature. The BLIP temperature is ~60K.

tunneling and phonon-assisted tunneling are probably the dominant components of the dark current.

The noise current at 77 and 100K is shown in Figure 3.67. At V_B = ±0.3V, the calculated thermal noise current (1.1 × 10^{-14} A/Hz$^{1/2}$) is very close to the measured noise current (1.3 × 10^{-14} A/Hz$^{1/2}$), which indicates that thermal noise is significant in the low-bias region (Figure 3.67). As the bias increases, the noise current increases much faster than thermal noise and the shot noise becomes dominant at higher bias.

For the shorter wavelength peaks, the QDIP had the best performance at 0.3V with a peak detectivity of 5.8 × 10^9 cmHz$^{1/2}$/W for the 5.9-μm peak, and at –0.3V the peak detectivity was 3.7 × 10^9 cmHz$^{1/2}$/W for the 5.5-μm peak. The corresponding responsivities were 3.5 and 3.0 mA/W, respectively. For the longer wavelength peaks, the QDIP had the best performance at –0.8V with a peak detectivity of 7.3 × 10^8 cmHz$^{1/2}$/W for the 8.9-μm peak, and at 0.5V the detectivity was 7.3 × 10^8 cmHz$^{1/2}$/W for the 10.3- to 10.9-μm peaks. The corresponding responsivities were 200 and 1.9 mA/W, respectively. The relatively lower detectivities for the longer wavelength peaks reflect the rapid increase in noise with increasing bias. The detectivity of the two-color QDIP is still lower than that of multiwavelength QWIPs, which is mainly due to the lower responsivity as a result of the lower dot density and

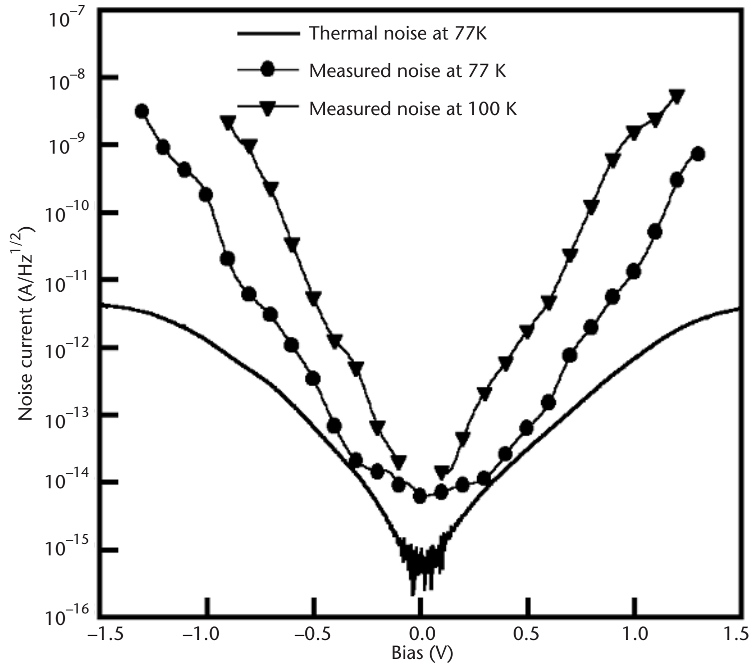


Figure 3.67 Measured noise current of 2.5-ML InGaAs capped QDIP at 77 and 100K. The solid line is the calculated thermal noise.

fewer quantum dot layers. QDIPs with more active layers may achieve higher responsivity if dot size and density can be maintained.

3.4 Prognosis

While QDIPs have demonstrated good performance for normal incidence operation, at present, their performance (detectivity) is not equal to that of QWIPs. We anticipate that comparable or superior performance will be achieved by a combination of further materials developments and improvements in device design. The detectivity is, in essence, simply an expression of SNR. As such, higher detectivity can be achieved by improving the responsivity and/or lowering the noise (dark current).

The primary factors that impact the responsivity are the number of QD stacks in the active region and the density and uniformity of the QDs. The number of QD stacks effectively determines the absorption length; more QD layers result in higher absorption. It would appear that the solution would be to incorporate more QD layers, as is done for QWIPs. However, for QDIPs, it is difficult to maintain sufficient size uniformity in successive QD layers. The net result is that the spectral response broadens without significant improvement in responsivity. From the materials perspective, it is advantageous to develop innovations in growth that maintain the QD size as more and more layers are built up.

There is also a device approach to increase the effective optical absorption length and thus increase the responsivity, namely, the use of resonant-cavity enhanced photodetector structures. The resonant-cavity photodetector structure consists of a thin absorbing layer sandwiched between two dielectric mirrors. A schematic cross section is shown in Figure 3.68. The resonant-cavity structure can provide several performance advantages, one of which is enhanced quantum efficiency for absorption regions for which the single-pass attenuation is small. Viewed, somewhat simplistically, the photon lifetime in the Fabry-Perot cavity of the photodiode is such that the photons can traverse the absorption region several times, which increases the effective pathlength.

Near-IR (1- to 1.3-μm) resonant-cavity photodiodes with external quantum efficiency in excess of 90% have been demonstrated [69]. A challenge in extending the near-IR work on resonant-cavity enhanced photodetectors to MWIR and LWIR is fabrication low-loss mirrors in those spectral regions. It is an advantage that resonant-cavity enhanced photodetectors do not require mirrors with as high a reflectivity as those employed in vertical cavity surface emitting lasers. In an optimized cavity, the front and rear mirror reflectivities, R_f and R_r, respectively, satisfy the relation $R_f = R_r e^{-\alpha d}$ where α is the absorption coefficient of the active region and d is its thickness. One approach would utilize pairs of amorphous Ge and SiO_2 layers. The simulated reflectivity for two Ge (140-nm)/SiO_2 (331-nm) mirrors, one with one pair and one with eight pairs, is shown in Figure 3.69. The single-pair mirror would be suitable for operation at 2 μm, whereas the 8-pair mirror could extend to ~8.5 μm. The reflectivity of both mirrors would be adequate as the back mirror for a resonant-cavity enhanced QDIP with a top mirror reflectivity of 30%, that is, the reflectance of the air–semiconductor interface.

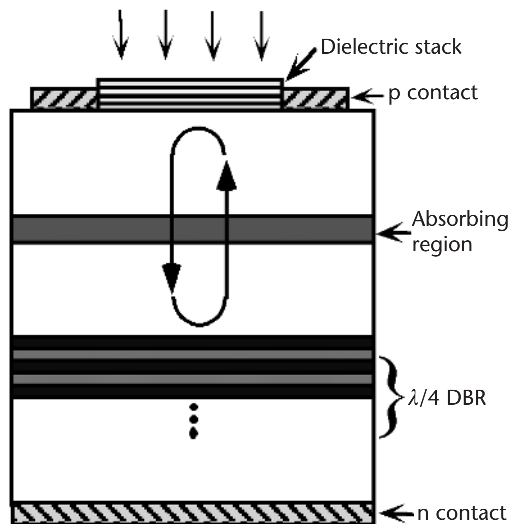


Figure 3.68 Schematic cross section of a resonant-cavity enhanced photodetector.

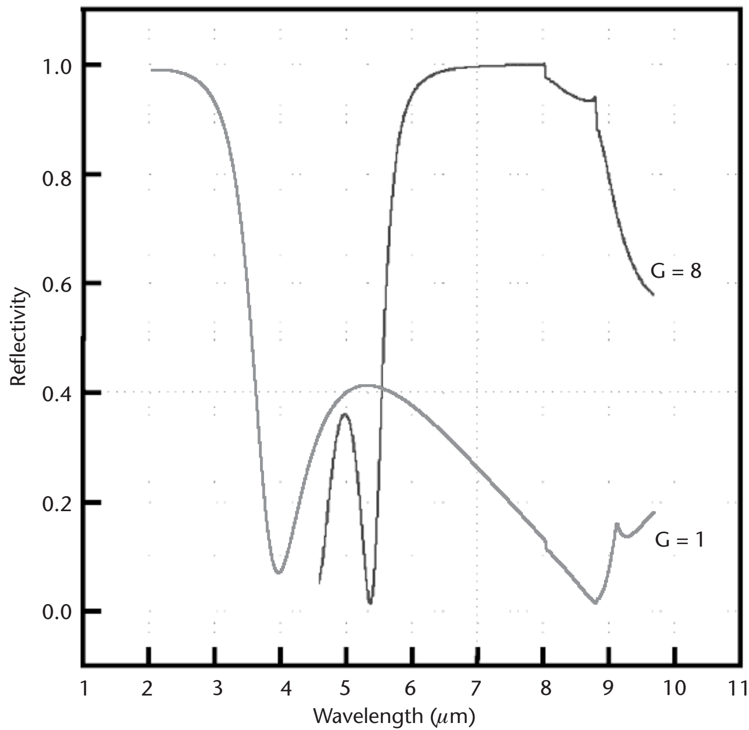


Figure 3.69 Reflectivity of a dielectric mirror consisting of pairs of Ge (140 nm) and SiO_2 (331 nm); G represents the number of pairs.

Acknowledgments

The work represented here was largely carried out under the U.S. Department of Defense Multi-Disciplinary University Research Initiative (MURI) program in Nanoscience, administrated by the Air Force Office of Scientific Research (AFSOR) under Grant AFOSR F49620-98-1-0474. We are grateful to Major Mike Prairie, Major Daniel Johnstone, and Lt. Col. Todd Steiner of the AFSOR for their support. We thank Dr. Ildar Mukhametzhanov and Dr. Oleg Baklenov for their contribution to the early work under this program. We also thank Mr. Siyuan Lu and Ms. Mary Shiff for their help in putting together this chapter.

References

[1] Bimberg, D., M. Grundmann, and N. N. Ledentsov, *Quantum Dot Heterostructures,* New York: John Wiley & Sons, 1999.

[2] Steier, O., Chap. 7 in *Nanooptoelectronics*, M. Grundmann, (ed.), Berlin, Germany: Springer-Verlag, 2002.

[3] Levine, B. F., *J. Appl. Phys.* Vol. 74, 1993, p. R1.

[4] Bockelmann, U., and G. Bastard, *Phys. Rev. B,* Vol. 42, 1990, p. 8947.

[5] Benisty, H., C. M. Sotomayer-Torrès, and C. Weisbuch, *Phys. Rev. B,* Vol. 44, 1991, p. 10945.
[6] Ryzhii, V., *Semicond. Sci. Technol.,* Vol. 11, 1996, p. 759.
[7] Baklenov, O., Ph.D. Dissertation, The University of Texas at Austin, 1999.
[8] Towe, E., and D. Pan, *J. Selected Topics Quantum Electron.,* Vol. 6, 2000, p. 408.
[9] Kim, J. W., et al., *IEEE Electron Device Lett.,* Vol. 21, No. 7, 2000, pp. 329–331.
[10] Lee, S. W., K. Hirakawa, and Y. Shimada, *Appl. Phys. Lett.,* Vol. 75, 1999, p. 1428.
[11] Chu, L., et al., *Appl. Phys. Lett.,* Vol. 79, 2001, p. 2249.
[12] Xie, Q., et al., *J. Crystal Growth,* Vol. 150, 1995, p. 357.
[13] Ramachandran, T. R., et al., *J. Crystal Growth,* Vol. 175/176, 1997, p. 216.
[14] Mukhametzhanov, I., et al., *Appl. Phys. Lett.,* Vol. 75, 1999, p. 85.
[15] Huffaker, D. L., et al., *Appl. Phys. Lett.,* Vol. 73, 1998, p. 2564.
[16] Mukhametzhanov, I., Ph.D. dissertation, University of Southern California, 2001.
[17] Kim, E. T., et al., *J. Vacuum. Sci. Technol.,* Vol. B20, 2002, p. 1188.
[18] Mukhametzhanov, I., et al., *Phys. Stat. Solidi B,* Vol. 224, 2001, p. 697.
[19] Heitz, R., et al., *Phys. Rev. B,* Vol. 62, 2000, p. 11017.
[20] Chen, Z. H., et al., *QWIP 2000 Workshop,* Dana Point, CA, July 2000.
[21] Stier, O., M. Grundmann, and D. Bimberg, *Phys. Rev. B,* Vol. 59, 1999, p. 5688.
[22] Sun, S. J., and Y. C. Chang, *Phys. Rev. B,* Vol. 62, 2000, p. 13631.
[23] Chen, Z. H., et al., *J. Appl. Phys.,* Vol. 89, 2001, p. 4558.
[24] Heitz, R., et al., *Phys. Rev. Lett.,* Vol. 83, 1999, p. 4654.
[25] Kim, E. T., Ph.D. Dissertation, University of Southern California, 2003.
[26] Kim, E. T., Z. H. Chen, and A. Madhukar, *Appl. Phys. Lett.,* Vol. 79, 2001, p. 3341.
[27] Ryzhii, V., *Appl. Phys. Lett.,* Vol. 78, 2001, p. 3346.
[28] Xu, S. J., et al., *Appl. Phys. Lett.,* Vol. 73, 1998, p. 3153.
[29] Arent, D. J., et al., *J. Appl. Phys.,* Vol. 66, 1989, p. 1739.
[30] Ji, G., et al., *J. Appl. Phys.,* Vol. 62, 1987, p. 3366.
[31] Chen, Z. H., E. T. Kim, and A. Madhukar, *Appl. Phys. Lett.,* Vol. 80, 2002, p. 2490.
[32] Merkt, U., *Phys. Rev. Lett.,* Vol. 76, 1996, p. 1134.
[33] Metzner, C., and G. H Dohler, *Phys. Rev. B,* Vol. 60, 1999, p. 11005.
[34] Stier, O., Ph.D. dissertation, Technical University of Berlin, 2001, p. 55.
[35] Wang, L. W., J. Kim, and A. Zunger, *Phys. Rev. B,* Vol. 59, 1999, p. 5678.
[36] Pan, D., E. Towe, and S. Kennerly, *Appl. Phys. Lett.,* Vol. 76, 2000, p. 3301.
[37] Kim, S., et al., *Appl. Phys. Lett.,* Vol. 73, 1998, p. 963.
[38] Phillips, J., et al., *IEEE J. Quant. Electron.,* Vol. 35, 1999, p. 936.
[39] Kim, J. W., et al., *IEEE Electron Device Lett.,* Vol. 21, 2000, p. 329.
[40] Baklenov, O., et al., *Proc. 58th IEEE Device Research Conf.,* Denver, CO, June 19–21, 2000, p. 171.
[41] Chen, Z. H., et al., *Infrared Physics & Technol.,* Vol. 42, 2001, p. 479.
[42] Wang, S. Y., et al., *Appl. Phys. Lett.,* Vol. 78, 2001, p. 1023.
[43] Lin, S. Y., Y. R. Tsai, and S. C. Lee, *Appl. Phys. Lett.,* Vol. 78, 2001, p. 2784.
[44] Stiff, A. D., et al., *Appl. Phys. Lett.,* Vol. 79, 2001, p. 421.
[45] Stiff, A. D., et al., *IEEE J. Quantum Electron.,* Vol. 37, 2001, p. 1412.
[46] Ye, Z., *IEEE J. Quantum Electron.,* Vol. 38, 2002, pp. 1234–1237.
[47] Ye, Z., et al., *J. Appl. Phys.,* Vol. 92, 2002, p. 4141.
[48] Ye, Z., et al., *J. Appl. Phys.,* Vol. 92, 2002, p. 7463.

[49] Mukhametzhanov, I., et al., *Phys. Status Solidi B*, Vol. 224, 2001, p. 697.
[50] Jelen, C., et al., *IEEE J. Quantum Electron.,* Vol. 34, 1998, p. 1124.
[51] Ryzhii, V., *J. Appl. Phys.,* Vol. 89, 2001, p. 5117.
[52] Liu, H. C., *Appl. Phys. Lett.,* Vol. 61, 1992, p. 2703.
[53] Beck, W. A., *Appl. Phys. Lett.,* Vol. 63, 1993, p. 3589.
[54] Levine, B. F., et al., *J. Appl. Phys.,* Vol. 72, 1992, p. 4429.
[55] Schönbein, C., et al., *Appl. Phys. Lett.,* Vol. 73, 1998, p. 1251.
[56] Choi, K. K., *Appl. Phys. Lett.,* Vol. 65, 1994, p. 1268.
[57] Finkman, E., et al., *Physica E,* Vol. 7, 2000, p. 139.
[58] Berryman, K. W., S. A. Lyon, and M. Segev, *Appl. Phys. Lett.,* Vol. 70, 1997, p. 1861.
[59] Xu, S. J., et al., *Appl. Phys. Lett.,* Vol. 73, 1998, p. 3153.
[60] Maimon, S., et al., *Appl. Phys. Lett.,* Vol. 73, 1998, p. 2003.
[61] Liu, H. C., et al., *Appl. Phys. Lett.,* Vol. 78, 2001, p. 79.
[62] Nishi, K., et al., *Appl. Phys. Lett.,* Vol. 74, 1999, p. 1111.
[63] Ustinov, V. M., et al., *Appl. Phys. Lett.,* Vol. 74, 1999, p. 2815.
[64] Guffarth, F., et al., *Phys. Status Solidi B,* Vol. 224, 2001, p. 61.
[65] Chen, Z. H., E. T. Kim, and A. Madhukar, *J. Vac. Sci. Technol. B*, Vol. 20, 202, p. 1243.
[66] Kim, E. T., et al., *Appl. Phys. Lett.,* Vol. 80, 2000, p. 2770.
[67] Rogaloski, A., *Infra Red Phys. Technol.,* Vol. 38, 1997, p. 295.
[68] Chen, Z. H., E. T. Kim, and A. Madhukar, *Appl. Phys. Lett.,* Vol. 80, 2002, p. 2770.
[69] Tan, I. H., et al., *IEEE Photon. Technol. Lett.,* Vol. 6, 1994, pp. 811–813.

CHAPTER 4

Quantum Dot Lasers: Theoretical Overview

Levon V. Asryan and Serge Luryi, State University of New York at Stony Brook

4.1 Introduction: Dimensionality and Laser Performance

The emergence of devices based on nanometer-size active elements marked the era of nanoelectronics and nanophotonics. Among such elements are notably low-dimensional heterostructures, such as QWs [1], *quantum wires* (QWRs) [2], and QDs [3]. Quantum confinement in low-dimensional heterostructures strongly modifies the basic properties of a semiconductor crystal.

In a QW, carriers are spatially confined in the transverse direction and move freely in its plane. In a QWR, carriers are spatially confined in two transverse directions and move freely along it. Hence, the carrier energy spectra in both QWs and QWRs are continuous within wide subbands of allowed states and, in this sense, they do not qualitatively differ from those in a bulk crystal.

In a QD, carriers are three-dimensionally confined and the modification of electronic properties is most strongly pronounced: the energy levels are discrete. For this reason, QDs are also referred to as *superatoms* or artificial atoms. A QD of typical size (several nanometers to several tens of nanometers) contains several thousands to several tens of thousands atoms. Quantum dots have generated much interest as a new class of human-made materials with tunable (by varying both the composition and size) energies of discrete atomic-like states.

The semiconductor laser is the fundamental device of modern optoelectronics and photonics. It was proposed long ago [4] that reducing the dimensionality of the active region could significantly improve laser performance due to the quantum-size effect. This general idea was initially applied to QW lasers [5–7] and by now QW lasers have replaced bulk lasers in most commercial applications [1]. Further enhancement is expected for lasers with lower dimensionality, such as QWR and especially QD lasers. In the context of QWR and QD lasers, this idea was first analyzed theoretically in [8]. Due to a continuous density of states within allowed subbands, using QWs and QWRs as an active medium for stimulated optical transitions can only quantitatively enhance device characteristics compared to those of a bulk device [9, 10].

Figure 4.1 shows the transformation of the density of states and the schematic gain spectrum with the reduction in the dimension of the active region. As the density of states narrows, one needs smaller number of states to be filled to attain transparency of the active region and the lasing. As a result, both the transparency current (injection current required for zero gain) and the threshold current (current, at which the gain equals the loss and the lasing starts) decrease, and also their temperature dependences become less pronounced. Lowering the threshold current and improving its temperature stability are important objectives in the development of diode lasers [11]. As seen from the figure, the radical, qualitative change in the density of states and the gain spectrum occurs only in QDs. The discrete carrier spectrum in QDs appears therefore ideally suitable for lasing generation with low threshold and high temperature stability. Thus, the semiconductor (diode) QD lasers form a novel class of injection lasers that promise radically enhanced

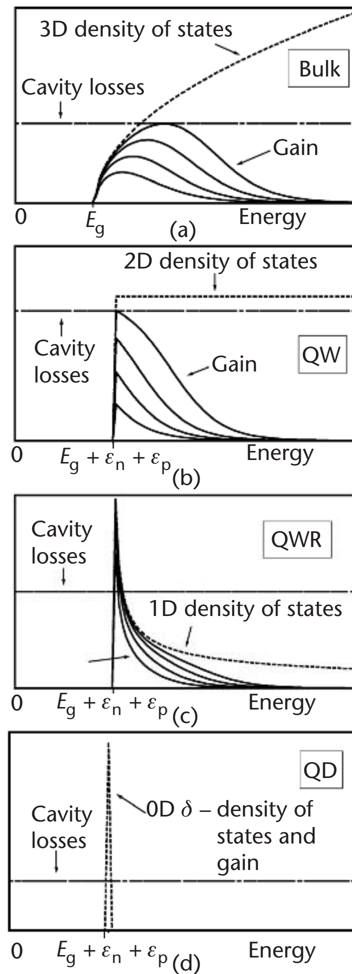


Figure 4.1 (a–d) Transformation of the density of states and the schematic gain spectrum for different dimensions of an active region.

operating characteristics [3, 8, 12–14]. The use of QDs as an active medium in injection lasers is a dramatic example of nanotechnology applied to devices of high commercial interest.

4.2 Advantages of an Idealized QD Laser

The principal advantages of QD lasers over the conventional QW lasers can be summarized as follows:

- Significantly lower threshold current density;
- Significantly weaker temperature dependence of the threshold current, ideally, temperature-insensitive threshold current;
- Superior opportunity for tuning the gain spectrum width and the emission wavelength (color of light);
- Low chirp (shift of the lasing wavelength with injection current), ideally, zero α-factor.

In the ongoing improvement of semiconductor lasers, each of the above areas has always been on the high-priority list and motivated the very idea of heterostructure lasers [15]. The low threshold currents of heterostructure lasers, already demonstrated during the early stage of development [16–18], provided much momentum for continuing research in this field.

4.3 Progress in Fabricating QD Lasers

Practical realization of the advantages of QD lasers became possible with the advent of QD structures with high uniformity of size and shape.

Initial attempts to fabricate QDs and QD devices relied on the traditional—at the time—means, such as selective etching of QW structures or QW intermixing, growth on profiled substrates and on cleavage facets, or condensation in glassy matrices [19]. These efforts, however, did not produce device-oriented structures.

A breakthrough in fabricating QD lasers (first optically pumped [20] and then current injected [21]) came with the use of self-organizing effects in heteroepitaxial systems. The nonplanar 3D growth, which had been traditionally considered undesirable, has led to the direct formation of QDs. The possibility of forming 3D islands in a wideband matrix was first demonstrated in [22]. At the time, however, this did not attract much attention because the prospect of fabricating uniform QDs looked doubtful. It took extensive experimental and theoretical studies [23] before QD arrays of high structural perfection and uniformity could be realized in practice.

Today, we have a reasonably mature epitaxial growth technology that employs spontaneous formation of semiconductor nanostructures in heteroepitaxial mismatched systems. This technology allows us to control both the surface density and the size of QDs [24]. The most extensively studied heterosystems for QD lasers are InAs/GaAs, InAs/InGaAs, InGaAs/GaAs, and InGaAs/AlGaAs on the GaAs

substrate. Using these material systems, one can vary the lasing wavelength in the wide spectral range from 0.87 to 1.9 μm, including 1.3 and 1.55 μm, the most desirable wavelengths for telecommunication applications. The use of GaInP/InP and InAlAs/GaAs QDs extended the range into the visible (red) spectrum.

Commercial perspectives on QD lasers have stimulated the efforts of many groups and led to significant progress in fabrication technology. Among the demonstrated advantages of QD lasers is the lowest threshold current density—19 A/cm^2—ever reported for *continuous-wave* (CW) room temperature operation for semiconductor lasers of any type [25]. A temperature stability of the threshold current that is superior to QW lasers has been demonstrated; a characteristic temperature T_0 above 150K was reported at operating temperatures well above room temperature [26–28]. Wide spectral tunability has also been demonstrated [29].

4.4 State-of-the-Art Complications

Figure 4.2 shows schematically the cross section and the band diagram of a typical QD laser. The bipolar device employs stimulated transitions between the quantized energy levels of electrons and holes in QDs. In this section, we focus on the most important issues that hinder the development of such lasers with superior performance compared to other contemporary semiconductor lasers.

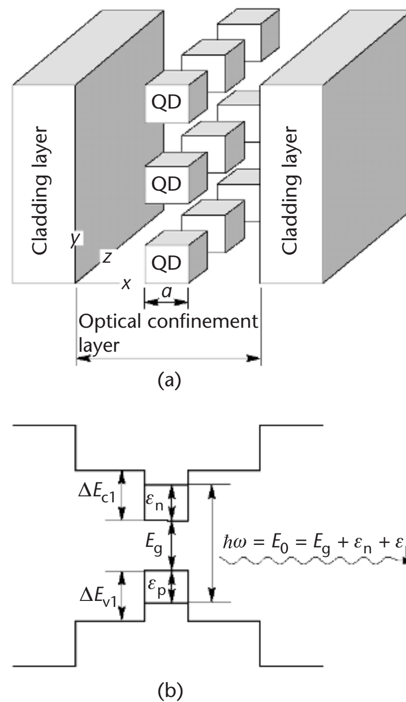


Figure 4.2 (a) Schematic structure and (b) energy band diagram of a QD laser.

4.4.1 Nonuniformity of QDs

Nonuniformity is a problem of great concern. As discussed earlier, the advantages of QD lasers stem from a δ-function-like density of states. If all QDs were identical, the gain spectrum would also be a δ function (Figure 4.1). However, the QDs in actual structures vary, primarily in size (Figure 4.3) and shape, but also in the local strain. The QD parameter dispersion causes fluctuations in the quantized energy levels. This leads to an inhomogeneous broadening in the optical transition energy (Figure 4.3) and hence also broadens the gain spectrum. This dispersion is hardly avoidable during the QD structures' growth: Size fluctuations are inherent in self-organized QD ensembles either fabricated by MBE or MOCVD.

Inhomogeneous line broadening is the key factor degrading the characteristics of a QD laser [30]. The QD parameter dispersion adversely affects the operating characteristics of a laser:

- Maximum gain decreases.
- Threshold current increases and becomes more sensitive to temperature (the characteristic temperature decreases).
- The multimode generation threshold decreases.
- The internal differential efficiency and output power both decrease.

The QD laser advantages can only be realized if the QDs are sufficiently uniform. Later we discuss how the threshold and power characteristics of a laser depend on the QD size fluctuations, that is, on the "degree" of the structure perfection.

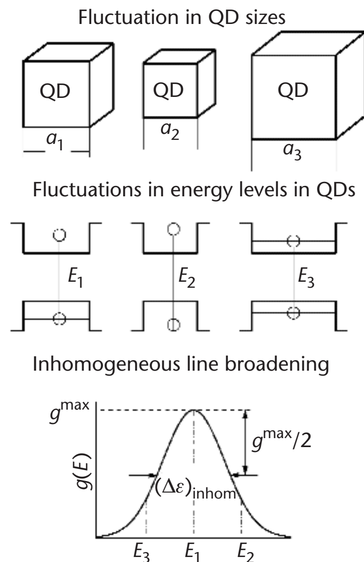


Figure 4.3 Inhomogeneous line broadening arising from nonuniformity of QDs.

Effect on the Gain

The laser threshold condition is

$$g^{\mathrm{m}} = \beta \tag{4.1}$$

where g^{m} is the peak value of the modal gain spectrum and β is the total loss. The minimum injection current satisfying (4.1) is, by definition, the threshold current.

The shape of the modal gain spectrum (which represents the effective gain of the active layer comprising QDs) and its transformation with the injection current are quite different in the two limits, corresponding to small and large QD size dispersions [30–32].

For small fluctuations (the inhomogeneous line broadening is less than the temperature T), the gain spectrum copies the curve for the QD size distribution, with the scale along the vertical axis determined by the population inversion in the QD of average size [Figure 4.4(a)]. The spectrum changes self-similarly with the current.

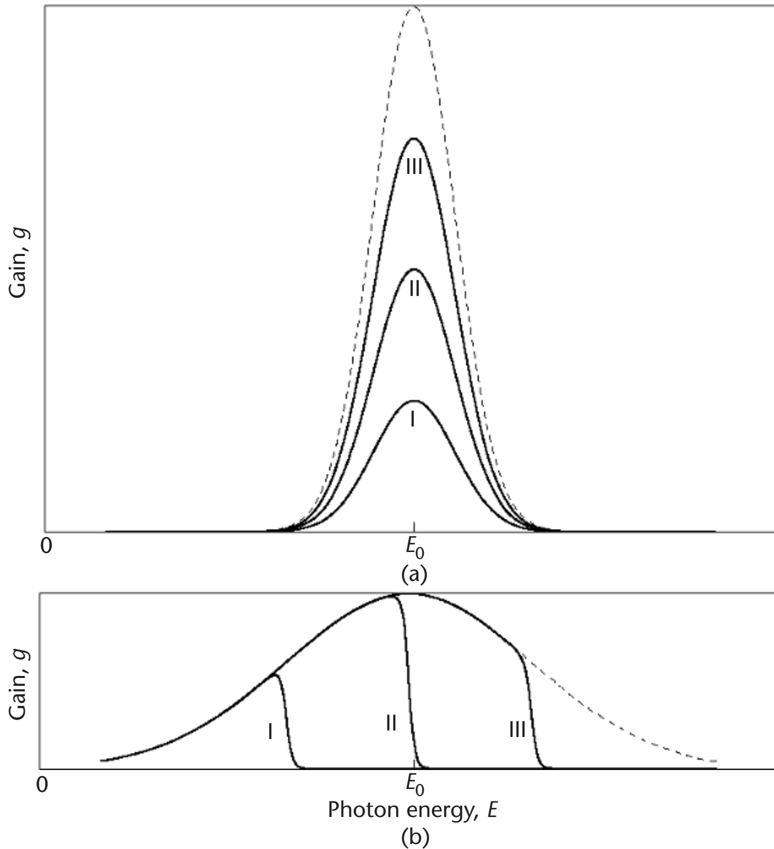


Figure 4.4 Gain spectra for equilibrium filling of quantum dots: (a) small dispersion, (b) large dispersion. Curves are numbered in order of increasing population inversion in a QD. The dashed curves correspond to full population inversion. (*From:* [30]. © 1996 IEEE. Reprinted with permission.)

The peak value of the modal gain spectrum occurs at the transition energy E_0 in a mean-sized QD and is equal to

$$g^{\mathrm{m}} = g^{\max}\left(f_n + f_p - 1\right) \tag{4.2}$$

where f_n and f_p are the occupancies of the quantized energy levels of an electron and a hole in a mean-sized QD.

The quantity $g^{\max}$ is the maximum possible (saturation) value of g^{m}. It holds when both f_n and f_p approach unity, that is, when the QDs are fully occupied. The dependence of $g^{\max}$ on the inhomogeneous line broadening $(\Delta\varepsilon)_{\mathrm{inhom}}$ is given by the following expression:

$$g^{\max} = \frac{\xi}{4}\left(\frac{\lambda_0}{\sqrt{\varepsilon}}\right)^2 \frac{1}{\tau_{\mathrm{QD}}} \frac{\hbar}{(\Delta\varepsilon)_{\mathrm{inhom}}} \frac{1}{\mathfrak{L}} N_S \tag{4.3}$$

where $\xi = 1/\pi$ and $\xi = 1/\sqrt{2\pi}$ for the Lorentzian and Gaussian distributions, respectively. $\lambda_0 = 2\pi\hbar c/E_0$ is the lasing wavelength; ε is the dielectric constant of the *optical confinement layer* (OCL); $\mathfrak{L}$ is the characteristic length of the optical confinement in the transverse waveguide direction; and N_S is the surface density of QDs.

The reciprocal spontaneous radiative recombination time for transitions between the quantized energy levels in the conduction and the valence bands is of the form

$$\frac{1}{\tau_{\mathrm{QD}}} = \frac{8}{3}\alpha\sqrt{\varepsilon}\,\frac{E_0}{\hbar}\left(\frac{P}{\hbar c}\right)^2 I \tag{4.4}$$

where $\alpha = e^2/\hbar c$ is the fine structure constant, P is Kane's parameter [33], and I is the overlap integral between the electron and hole wave functions (see Section 4.4.8).

The inhomogeneous line broadening caused by fluctuations in QD sizes is

$$(\Delta\varepsilon)_{\mathrm{inhom}} = \left(q_n\varepsilon_n + q_p\varepsilon_p\right)\delta \tag{4.5}$$

where $q_{n,p} = -\partial \ln\varepsilon_{n,p}/\partial \ln a$, $\varepsilon_{n,p}$ are the quantized energy levels in a mean-sized QD, a is the mean size of QDs, and δ is the *root mean square* (rms) of relative QD size fluctuations.

The gain saturation effect is observed experimentally in QD lasers. Because $g^{\max} \propto 1/(\Delta\varepsilon)_{\mathrm{inhom}}$, we see how crucial the QD size uniformity is.

For large fluctuations [$(\Delta\varepsilon)_{\mathrm{inhom}} > T$], the gain spectrum gradually fills the curve for the QD size distribution with increasing injection current [Figure 4.4(b)]. Both self-similar and gradually filled gain spectra have been observed experimentally.

Effect on the Threshold Current

Below and at the lasing threshold, the injection current density is consumed by the spontaneous recombination in QDs and in the OCL:

$$j = \frac{eN_S}{\tau_{\text{QD}}}\left\langle f_n f_p \right\rangle + ebBnp \tag{4.6}$$

where $\langle \ldots \rangle$ means averaging over the QD ensemble, b is the thickness of the OCL, B is the radiative constant for the OCL, and n and p are the free-carrier densities in the OCL.

If we consider relatively high temperatures, the carriers in each band are close to equilibrium and in steady state the free-carrier densities in the OCL (below and at the threshold) can be expressed in terms of the confined-carrier level occupancies in a mean-sized QD [30]:

$$n = n_1 \frac{f_n}{1-f_n},\ p = p_1 \frac{f_p}{1-f_p} \tag{4.7}$$

Here the quantities n_1 and p_1 are

$$n_1 - N_c \exp\left(-\frac{\Delta E_{c1} - \varepsilon_n}{T}\right),\ p_1 = N_v \exp\left(-\frac{\Delta E_{v1} - \varepsilon_p}{T}\right) \tag{4.8}$$

where $N_{c,v}$ are the conduction and valence band effective densities of states and $\Delta E_{c1,\,v1}$ are the band offsets at the QD–OCL heterointerface (Figure 4.2).

Assuming charge neutrality in QDs, $f_n = f_p$ (which is not the general case—see Section 4.3), (4.1) and (4.2) yield the level occupancies at threshold in a mean-sized QD:

$$f_{n,p} = \frac{1}{2}\left(1 + \frac{\delta}{\delta^{\max}}\right) \tag{4.9}$$

where $\delta^{\max}$ is the maximum tolerable rms of relative QD size fluctuations [see (4.43) in Section 4.4.7].

With (4.9), (4.7) and (4.6) yield the dependence on δ of both the free-carrier densities in the OCL at the lasing threshold and the threshold current density:

$$n_{\text{th}} = n_1 \frac{1 + \frac{\delta}{\delta^{\max}}}{1 - \frac{\delta}{\delta^{\max}}},\ p_{\text{th}} = p_1 \frac{1 + \frac{\delta}{\delta^{\max}}}{1 - \frac{\delta}{\delta^{\max}}} \tag{4.10}$$

$$j_{\text{th}} = \frac{1}{4}\frac{eN_S}{\tau_{\text{QD}}}\left(1 + \frac{\delta}{\delta^{\max}}\right)^2 + ebBn_1 p_1 \frac{\left(1 + \frac{\delta}{\delta^{\max}}\right)^2}{\left(1 - \frac{\delta}{\delta^{\max}}\right)^2} \tag{4.11}$$

We see that the threshold current diverges as the dispersion increases and approaches a certain critical value, $j_{th} \to \infty$ for $\delta \to \delta^{max}$ [Figure 4.5(b)]. Such behavior has been observed experimentally [34].

In the opposite limit, j_{th} decreases and as $\delta \to 0$, the threshold current tends to the transparency (inversion) value, corresponding to the current density, at which $f_n + f_p - 1 = 0$ [Figure 4.5(b)].

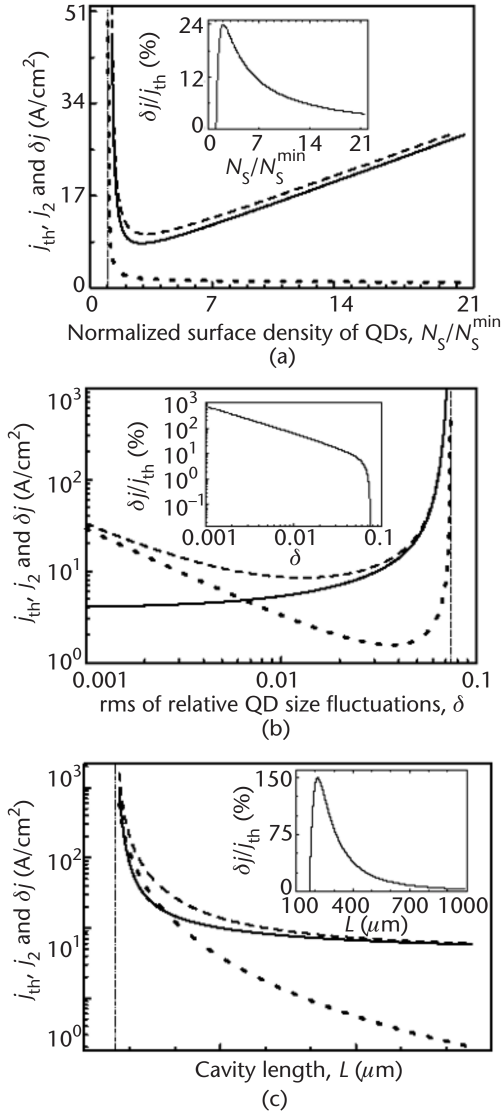


Figure 4.5 Multimode generation in a QD laser as function of (a) normalized surface density of QDs, (b) rms of relative QD size fluctuations, and (c) cavity length. Solid and dashed curves describe the threshold current density for the main mode and the next longitudinal mode, respectively, while dotted curves correspond to the multimode generation threshold. The insets show the relative multimode generation threshold. (*From:* [37]. © 2000 IEEE. Reprinted with permission.)

Effect on the Temperature Dependence of Threshold Current

The T dependence of j_{th} is described by the characteristic temperature T_0 [defined by (4.17)]. The higher this parameter, the less sensitive j_{th} is to temperature. Nonuniformity of QDs has a twofold effect on the T dependence of j_{th}. The main effect is through the thermal population of the OCL, which controls the parasitic recombination current outside QDs [30–32, 35]. The second effect is through the thermal population of nonlasing QDs, which gives rise to a parasitic recombination current [36].

Effect Through the Parasitic Recombination Outside QDs

The OCL is a nonlasing 3D region surrounding QDs. The carrier population in the OCL is in approximate thermal equilibrium with that in QDs, except at low temperatures. The recombination current in the OCL is the main source of temperature dependence of j_{th} (see later sections). Assuming charge neutrality in QDs, the following expression is obtained for T_0 as a function of δ [35]:

$$T_{0,\text{ neutral}} = \left[1 + \frac{\frac{1}{4}\frac{N_S}{\tau_{QD}}\left(1 - \frac{\delta}{\delta^{\max}}\right)^2}{bBn_1 p_1}\right] \frac{1}{\frac{3}{2}\frac{1}{T} \frac{\Delta E_{g1} - \varepsilon_n - \varepsilon_p}{T^2}} \tag{4.12}$$

where $\Delta E_{g1} = \Delta E_{c1} + \Delta E_{v1}$ is the bandgap difference between the materials of the OCL and QD.

The more uniform the QD ensemble, the lower the carrier density [see (4.8) and (4.10)] and the recombination current [the second term on the right side of (4.11)] in the OCL and the higher the T_0 (Figure 4.6). At room temperature and $N_S = 1.3 \times 10^{11}$ cm^{-2}, $\beta = 10$ cm^{-1} and 10% QD size dispersion ($\delta = 0.05$), the characteristic temperature for the structure optimized to minimize the threshold current density [30] is T_0

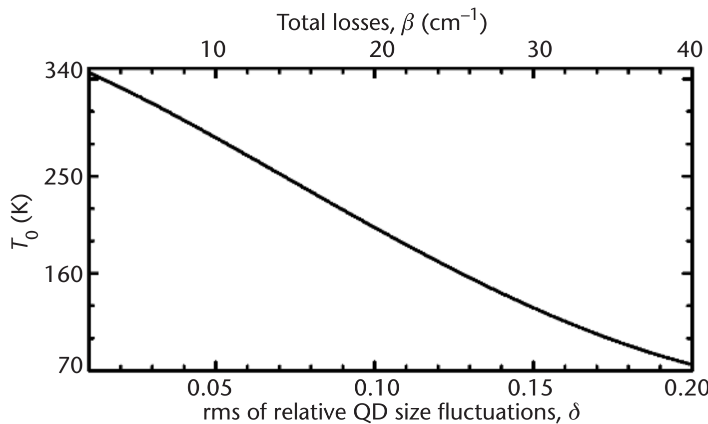


Figure 4.6 Characteristic temperature against rms of relative QD size fluctuations (at $\beta = 10$ cm^{-1}, bottom axis) and against total losses (at $\delta = 0.05$, top axis). $N_S = 1.3\times10^{11}$ cm^{-2}. (*From*: [35]. © 1998 IEEE. Reprinted with permission.)

≈ 286K [35]. This estimate is several times higher than the best T_0 in QW lasers for the same loss (typically less than 90K).

Effect Through the Recombination in Nonlasing QDs

Because of the inhomogeneous broadening, a certain fraction of QDs does not contribute to the lasing transitions while still adding to the parasitic recombination. As far as T_0 is concerned, the effect of thermal population of nonlasing QDs is in principle similar to (but not as strong as) that due to carriers residing in the OCL.

The thermal population of nonlasing QDs and hence the T dependence of the recombination current in QDs, j_{QD} [the first term on the right side of (4.6)], is simply accounted for by the deviation of $\langle f_n f_p \rangle$ from the product of the electron and hole level occupancies in a mean-sized QD. The characteristic temperature for j_{QD} is [36]

$$\frac{T_{0,\,\text{inhom}}^{\text{QD}}}{T} = \frac{1-\dfrac{a^2\delta^2}{2}(1-f_n)\left[\dfrac{\varepsilon''_n+\varepsilon''_p}{T}-\dfrac{\varepsilon'_n}{T}\dfrac{\varepsilon'_p}{T}(1-f_n)+\dfrac{\varepsilon'^{\,2}_n+\varepsilon'^{\,2}_p}{T^2}(2f_n-1)\right]}{\dfrac{a^2\delta^2}{2}(1-f_n)\left[\dfrac{\varepsilon''_n+\varepsilon''_p}{T}-2\dfrac{\varepsilon'_n}{T}\dfrac{\varepsilon'_p}{T}(1-f_n)+2\dfrac{\varepsilon'^{\,2}_n+\varepsilon'^{\,2}_p}{T^2}(2f_n-1)\right]} \tag{4.13}$$

where f_n is the level occupancy in a mean-sized QD at the lasing threshold given by (4.9), and $\varepsilon'_{n,p}$ and $\varepsilon''_{n,p}$ are, respectively, the first and the second derivatives of $\varepsilon_{n,p}$ with respect to the QD size taken at the mean size of the QD's a.

The dependence of $T_{0,\text{inhom}}^{\text{QD}}$ on δ is nonmonotonic (Figure 4.7). The decrease of $T_{0,\text{inhom}}^{\text{QD}}$ at small δ is due to the increasing thermal population of nonlasing QDs. At large QD size dispersion, when $\delta \rightarrow \delta^{\max}$, f_n approaches unity [see (4.9)] to satisfy the

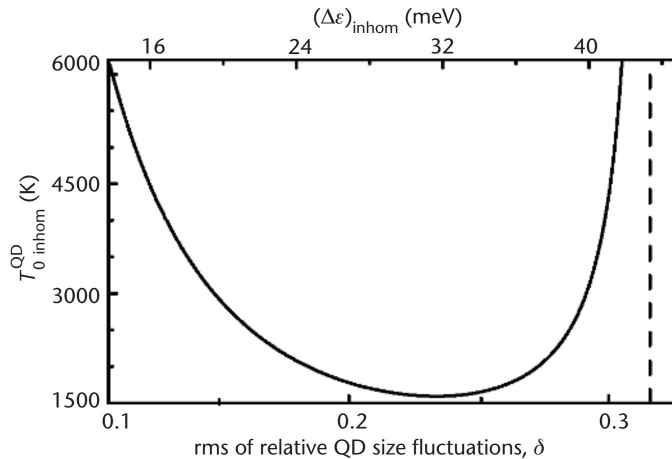


Figure 4.7 Characteristic temperature $T_{0,\text{inhom}}^{\text{QD}}$ as a function of the rms of relative QD size fluctuations (bottom axis) and the inhomogeneous line broadening (top axis). The vertical dashed line indicates $\delta^{\max}$. (*From:* [36]. © 2001 IEEE. Reprinted with permission.)

threshold condition; full occupancy of QDs forces temperature independence of j_{QD} at the expense of a very large increase in the total threshold current.

The characteristic temperature $T_{0,\text{inhom}}^{QD}$ is much higher than $T_{0,\text{neutral}}$ given by (4.12): As seen from Figure 4.7, $T_{0,\text{inhom}}^{QD}$ is above 1,500K over the entire range of δ. Such a characteristic temperature can be essentially considered infinite for most practical purposes. Hence, the effect of thermal population of nonlasing QDs is negligible compared to that of the nonlasing 3D region surrounding QDs.

Effect on the Multimode Generation Threshold

The multimode behavior of lasing in QD structures is caused by spatial hole burning (Section 4.4.5) [37].

The multimode generation threshold δj is defined as the excess of the injection current density over the threshold current density for the main mode, at which the lasing oscillation of the next (closest to the main) mode of the resonator begins. The following equation is obtained for δj as a function of δ at relatively high T [37]:

$$\delta j = 2\left[\frac{\hbar \frac{c}{\sqrt{\varepsilon}} \frac{\pi}{L}}{\left(q_n \varepsilon_n + q_p \varepsilon_p\right)\delta}\right]^2 \frac{\delta}{\delta^{\max}} \frac{1}{1 - \frac{\delta}{\delta^{\max}}} \frac{e N_S}{\tau_n^{\text{esc}} + \tau_p^{\text{esc}}} \tag{4.14}$$

where L is the cavity length and $\tau_{n,p}^{\text{esc}}$ are the thermal escape times of electrons and holes from QDs. These times are given by

$$\tau_n^{\text{esc}} = \frac{1}{\sigma_n v_n n_1}, \tau_p^{\text{esc}} = \frac{1}{\sigma_p v_p p_1} \tag{4.15}$$

where $\sigma_{n,p}$ are the QD capture cross sections and $v_{n,p}$ are thermal velocities.

We see that decreasing the QD size dispersion improves the single-mode behavior of the laser, in addition to the improvements in threshold current and temperature stability. As $\delta \rightarrow 0$, the multimode generation threshold increases indefinitely, diverging as δ^{-1} [Figure 4.5(b)].

Effect on the Internal Quantum Efficiency and the Output Power

We have discussed the effect of QD size dispersion on threshold characteristics. The QD uniformity is of crucial importance also for the high-power characteristics of a laser [38, 39].

The general expression for the internal quantum efficiency η_{int} of quantum confined lasers is (4.31) in a later section. As δ approaches its maximum tolerable value $\delta^{\max}$ [given by (4.43)], the internal quantum efficiency and the output power both vanish (Figure 4.8). The more uniform the QD ensemble, the higher η_{int} (Figure 4.8), the more linear the *light-current characteristic* (LCC) and the higher the output power (Figures 4.8 and 4.9). In properly optimized QD lasers, discussed later, the

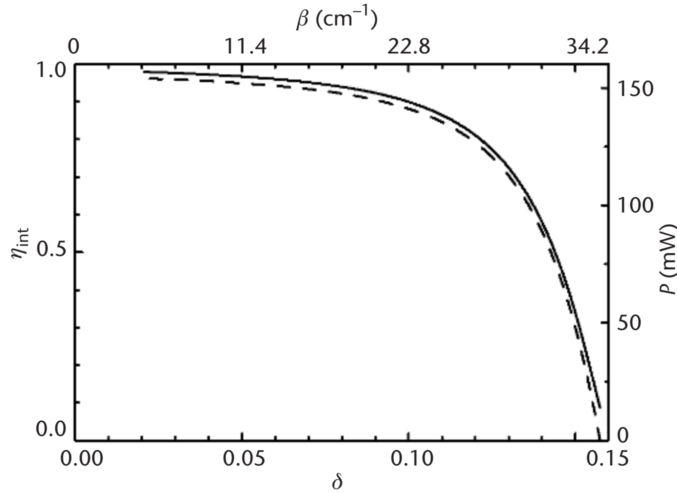


Figure 4.8 Internal quantum efficiency (solid curve, left axis) and output power (dashed curve, right axis) as a function of the rms of relative QD size fluctuations (at fixed L = 1 mm, bottom axis) and the cavity loss (at fixed δ = 0.05, top axis). (*From:* [39]. © 2003 IEEE. Reprinted with permission.)

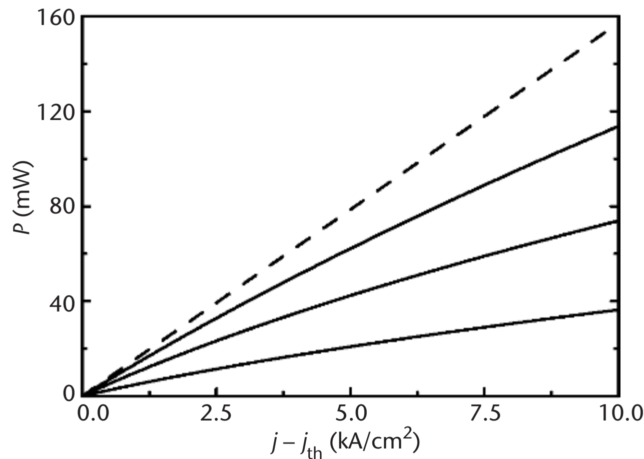


Figure 4.9 LCC for structures with different rms of relative QD size fluctuation. The dashed line corresponds to the ideal situation, η_{int} = 1. The values of δ and j_{th} (from the top down) are, respectively, 0.13, 0.145, 0.153, 122.44, 465.32, and 1,937 A/cm^2. (*From:* [39]. © 2003 IEEE. Reprinted with permission.)

LCC can be linear and η_{int} can be close to unity up to very high current densities (15 kA/cm^2); output powers in excess of 10W at η_{int} higher than 95% are attainable in broad-area devices. These results indicate that QD lasers may possess an advantage over conventional QW lasers for high-power applications.

4.4.2 Parasitic Recombination Outside QDs

In “traditional” QD laser structures, the QDs are surrounded by the OCL, which itself is conductive and transports mobile carriers to the QDs. (This situation is common to all semiconductor diode lasers.) Carriers are first injected from the cladding layers to the OCL and then captured into QDs. A two-step supply of carriers into the active region, wherein stimulated transitions occur, is detrimental to both the threshold and high-power characteristics. As a result of such an indirect, OCL-mediated injection into the active region, (1) the threshold current and its temperature-sensitivity both increase and (2) the internal quantum efficiency and the output power decrease. A new approach to the QD laser design, which eliminates the troublesome recombination outside the active region, is discussed in Section 4.4.5.

Effect on the Threshold Current

High temperature stability of operation is an essential feature required of long-wavelength diode lasers for telecommunications. Commercial QW lasers based on InGaAsP/InP heterosystem are rather poor in this respect; the characteristic temperature T_0 reaches, at best, about 90K.

Ideally, threshold current of a QD laser should remain unchanged with the temperature, that is, T_0 should be infinitely high [8]. This would be so indeed if the overall injection current went entirely into the radiative recombination in QDs. However, a fraction of the current occurs by recombination in the OCL. This component of j_{th}, denoted as j_{OCL} [the second term on the right side of (4.11)], is associated with thermal excitation of carriers from QDs and hence depends exponentially on T [30, 35]. It is this component that is mainly responsible for the T dependence of j_{th} at room temperature and above. Such a mechanism of T dependence is also at play in other semiconductor lasers, but in QD lasers it plays the central role. Inasmuch as the growth technology allows us to fabricate reasonably uniform QD arrays, so that the inhomogeneous line broadening is controlled to a high degree, it is the thermal exchange between the QDs and the OCL that remains the main obstacle to full realization of the advantages of 3D confinement.

Assuming charge neutrality, the T dependence of j_{OCM} (Figure 4.10) is apparent from (4.11) and (4.8):

$$j_{\text{OCL, neutral}}(T) \propto Bn_1 p_1 \propto T^{3/2} \exp\left(-\frac{\Delta E_{g1} - \varepsilon_n - \varepsilon_p}{T}\right) \tag{4.16}$$

(Temperature dependences of the radiative constant B [30] and of the conduction and valence band effective densities of states $N_{c,v}$ have also been taken into account here.)

With the increasing energy $\Delta E_{g1} - \varepsilon_n - \varepsilon_p = (\Delta E_{c1} - \varepsilon_n) + (\Delta E_{v1} - \varepsilon_p)$, which is the sum of the localization energies of electrons and holes in a mean-sized QD, the free-carrier densities in the OCL decrease and hence so does j_{OCL}. This makes j_{th} less sensitive to temperature. In [40], the OCL was made of AlGaAs, which has a wider

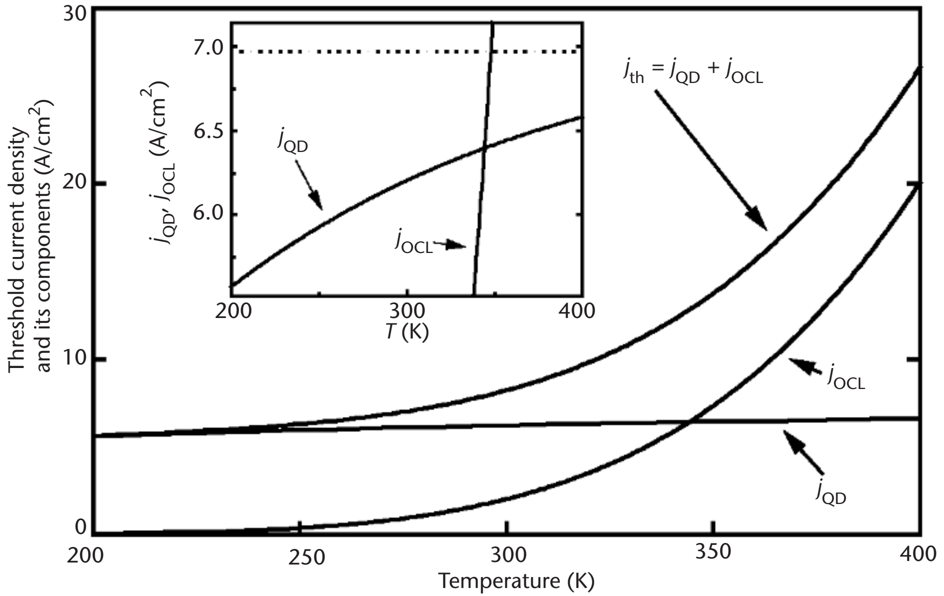


Figure 4.10 Temperature dependence of the threshold current density and its components. The inset shows $j_{QD}(T)$ and $j_{QCL}(T)$ on an enlarged (along the vertical axis) scale. The dashed line shows j_{QD} calculated assuming the charge neutrality in QDs. (*From:* [35]. © 2003 IEEE. Reprinted with permission.)

bandgap than GaAs, in order to suppress the thermal escape of carriers from QDs. This led to a significant reduction of j_{th}.

Effect on the Characteristic Temperature T_0

The characteristic temperature is the key parameter describing empirically the T dependence of the threshold current of semiconductor lasers. It is defined as follows [41]:

$$T_0 = \left(\frac{\partial \ln j_{th}}{\partial T} \right)^{-1} \tag{4.17}$$

Although $j_{th}(T)$ never fits the exponential form $\exp(T/T_0)$ precisely [as might appear from (4.17)], it is usually adequately characterized by the parameter T_0 within a narrow temperature range of interest. Considering a wider range, T_0 itself is a function of the temperature.

The dependence $T_0 = T_0(T)$ is strong [21] and shows a deep fall-off with increasing temperature (Figure 4.11). The drastic decrease in T_0 is explained [35] by the transition from the low-T regime, when j_{th} is controlled by recombination in QDs, to the regime at higher T, when j_{th} is controlled by the parasitic recombination in the OCL.

At relatively high T, charge neutrality violation has negligible effect on the T dependence of j_{th} (Figure 4.11) and T_0 is approximately given by (4.12). The factor outside the square brackets in (4.12) corresponds to a characteristic temperature $T_{0,\text{inhom}}^{\text{QD}}$ defined for the function $j_{\text{OCL, neutral}}(T)$ given by (4.16).

As discussed in an earlier section, the nonuniformity of QDs affects the threshold current and its temperature stability primarily through the parasitic recombination outside QDs. Equation (4.9) shows that when δ approaches its maximum tolerable value $\delta^{\max}$, the level occupancies in QDs tend to unity. Full occupancy of QDs requires infinitely high free-carrier densities in the OCL [see (4.10)]. Hence the recombination current density on the OCL [the second term on the right side of (4.11)] also increases infinitely.

Effect on the Internal Quantum Efficiency and the Output Power

Power characteristics are also strongly affected by the recombination outside QDs. As discussed later in Section 4.4.6, due to the noninstantaneous nature of carrier capture into QDs, the carrier density and the recombination current in the OCL both continue increasing with the injection current above the lasing threshold. This causes a deviation of the internal efficiency η_{int} from unity. Furthermore, since the recombination rate in the OCL is superlinear in the carrier density (first quadratic and then cubic), the internal efficiency becomes a decreasing function of the injection current, and hence the LCC is sublinear [38, 39].

It is the ratio of the threshold values of the parasitic recombination current outside QDs to the capture current into QDs that determines η_{int} and the output power

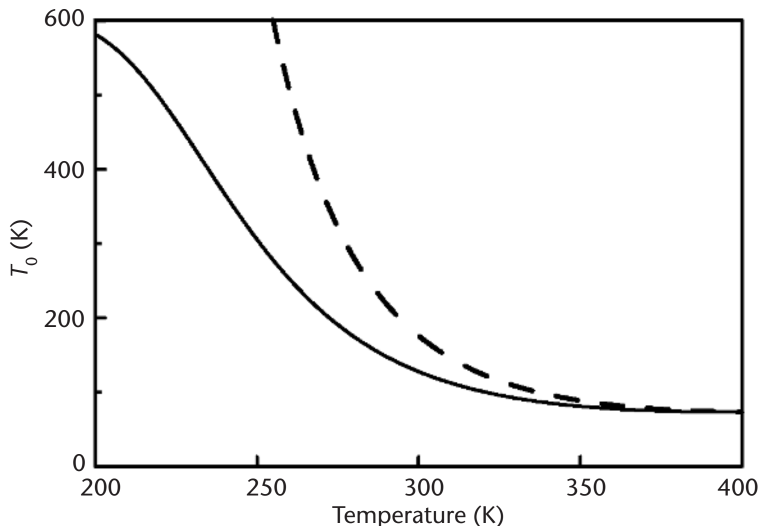


Figure 4.11 Temperature dependence of the characteristic temperature. The dashed curve shows T_0 calculated assuming charge neutrality in QDs [see (4.12)]. (*From:* [35]. © 1998 IEEE. Reprinted with permission.)

at a given injection current [38, 39]. Figure 4.12 shows how important the lowering of this ratio is to enhance η_{int} and the output power. We discuss this issue in greater detail in a later section.

4.4.3 Violation of Local Neutrality in QDs

As shown in [42], the electron level occupancy in a QD may differ from that of a hole; that is, QDs may be charged. The QD structures are distinct in this sense from similar QW structures in that the difference between the hole and the electron level occupancies in a QD, $f_p - f_n$, can be comparable to the occupancies, f_n and f_p, themselves (see [42] and Figure 4.13). The distinction is rooted in the low surface density of QDs, N_S (typically from several 10^{10} cm^{-2} to 10^{11}cm^{-2}), compared to the 2D-carrier density in a QW (typically above 10^{12}cm^{-2}). The same amount of surface charge density, $eN_S\,(f_p - f_n)$, required to screen a local electric field inhomogeneity, gives a tangible charge imbalance in the layer of QDs, while being quite negligible relative to the overall electron or hole charge in a QW. This is the reason why QWs can be considered neutral [43, 44] and QDs, in general, cannot.

The violation of local neutrality in QDs strongly affects the threshold characteristics of a laser. In the absence of neutrality, (4.1) alone does not determine the electron and hole level occupancies in a QD at the lasing threshold. One needs an additional relation between f_n and f_p. This relation can be obtained by solving a self-consistent problem for the electrostatic field distribution across the heterojunction. Generally, it can be written in the form:

$$f_p - f_n = \Delta \tag{4.18}$$

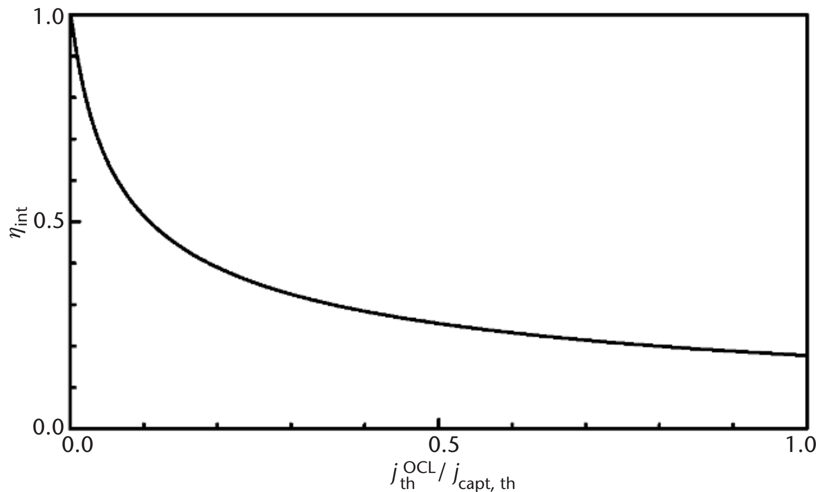


Figure 4.12 Internal quantum efficiency as a function of the ratio $j_{\text{th}}^{\text{OCL}}/j_{\text{capt, th}}$. Injection current density j = 10 kA/cm^2; the variation of $j_{\text{th}}^{\text{OCL}}/j_{\text{capt, th}}$ in the range shown (from 8×10^{-4} to 1) is accomplished by changing N_S from 20 × to 2.81×10^{10} cm^{-2}. (*From:* [39]. © 2003 IEEE. Reprinted with permission.)

A specific type of the function Δ depends on details of the laser structure, for example, on the spatial distribution of donor and acceptor impurities and the band offsets.

With (4.18), the level occupancies in a mean-sized QD at the lasing threshold are:

$$f_n = \frac{1}{2}\left(1 + \frac{\delta}{\delta^{\max}} - \Delta\right), f_p = \frac{1}{2}\left(1 + \frac{\delta}{\delta^{\max}} + \Delta\right) \tag{4.19}$$

When the local neutrality holds (Δ = 0), the threshold level occupancy is temperature insensitive [see (4.9)]. Due to the T dependence of Δ, the violation of charge neutrality is accompanied by a T dependence of f_n and f_p (Figure 4.13). Unconstrained by the neutrality condition, f_n and f_p are no longer fixed by the threshold condition and become T dependent. As a result, the threshold current density component j_{QD} associated with the radiative recombination in QDs is also T dependent (Figure 4.10). Thus, even if the parasitic recombination outside QDs is fully suppressed, there remains a T dependence of j_{th}, associated with the violation of QD neutrality. It is this effect that keeps finite the characteristic temperature T_0 (Figure 4.11) at low T (as already observed in the first QD laser [21], when the thermal escape of carriers from QDs is essentially suppressed.

The equation for $T_{0,\text{VCN}}^{\text{QD}}$ defined for $j_{\text{QD}}(T)$ is

$$\frac{1}{T_{0,\text{VCN}}^{\text{QD}}} = \frac{1}{4}\frac{1}{f_{\text{n}} f_{\text{p}}}\frac{\partial \Delta^2}{\partial T} \tag{4.20}$$

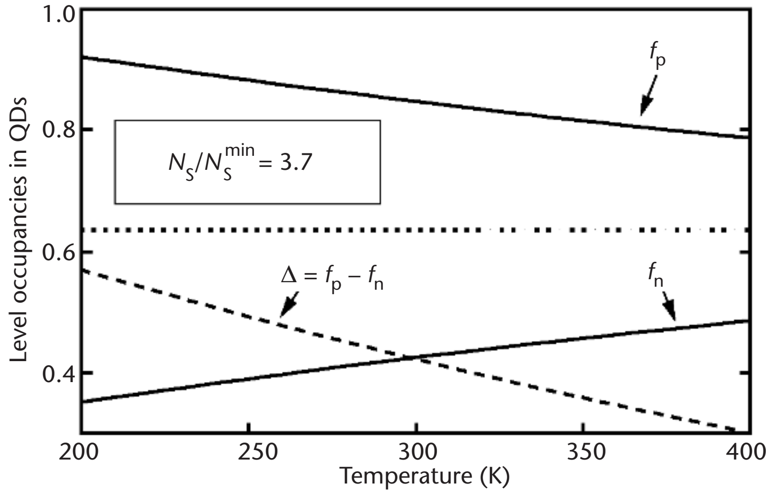


Figure 4.13 Temperature dependence of the electron and hole level occupancies and their difference at the lasing threshold. The horizontal dotted line shows $f_{n,p}$ calculated assuming charge neutrality in QDs. β = 10 cm^{-1} and δ = 0.05; $N_S^{\min}$ = 2.1 × 10^{10} cm^{-2}. (*From:* [35]. © 1998 IEEE. Reprinted with permission.)

The obtained characteristic temperature $T_{0,\mathrm{VCN}}^{\mathrm{QD}}$ is positive because the absolute value of Δ always decreases with T (Figure 4.13).

At relatively low T, when j_{th} is controlled by j_{QD}, the values of $T_{0,\ \mathrm{neutral}}$ obtained from (4.12) (Figure 4.11, dashed curve) are far larger than the actual T_0 calculated taking into account violation of neutrality (Figure 4.11, solid curve).

4.4.4 Excited States

Ideally, there should be one electron and one hole energy level in a QD, since the presence of excited carrier states may degrade some of the advantages of QD lasers. This complication would not be as severe for highly symmetrical (e.g., cubic) QDs, where multiple hole levels can be tolerated provided one has only one electron level. The reason is that radiative transitions from the ground electron state to excited hole states are forbidden by the selection rules in sufficiently symmetrical structures. However, such transitions are allowed in actual, low-symmetry (e.g., pyramidal) QDs. Moreover, there may also be excited electron states [45].

In the context of lowering j_{th} and enhancing T_0 excited state transitions are undesirable. On the other hand, from the standpoint of increasing the maximum gain, their presence may even be beneficial. It is via such transitions that lasing often occurs in short-cavity QD structures. Interestingly, the observed switch of the lasing wavelength with the cavity length (see, e.g., [46]) can be attributed to the contribution of these "forbidden" transitions [47, 48].

A detailed theoretical study of the effect of excited state transitions on the threshold characteristics of a QD laser was given in [47, 48]. Here we discuss the effect of excited states on the T dependence of j_{th} [49].

The presence of excited states serves as another source of T dependence of j_{th}, because of the thermally activated parasitic recombination associated with excited state transitions. Denoting the characteristic temperature limited by the presence of excited states by T_0^{exc}, the ratio T_0^{exc}/T can be put in the form of a universal function of Δ^{exc}/T (where Δ^{exc} is the separation between the transition energies; see Figure 4.14):

$$\frac{T_0^{\mathrm{exc}}}{\mathrm{T}} = \frac{1}{1-r}\frac{\left[\left(\dfrac{1}{f_n}-1\right)\exp\left(\dfrac{\Delta^{\mathrm{exc}}}{T}\right)+1\right]^3}{2\dfrac{\Delta^{\mathrm{exc}}}{T}\left(\dfrac{1}{f_n}-1\right)\exp\left(\dfrac{\Delta^{\mathrm{exc}}}{T}\right)}\left\{rf_n^2\,\frac{1-r}{\left[\left(\dfrac{1}{f_n}-1\right)\exp\left(\dfrac{\Delta^{\mathrm{exc}}}{T}\right)+1\right]^2}\right\} \tag{4.21}$$

where f_n is the occupancy of the ground state in a mean-sized QD, $r = I_1/(I_1 + I_2)$, and I_1 and I_2 are the rates (the reciprocals of the spontaneous radiative lifetimes) of the ground and excited state transitions, respectively.

For small enough QDs, the T dependence of j_{th} arising from thermally excited states is negligible compared to the effects of parasitic recombination in the OCL and charge neutrality violation.

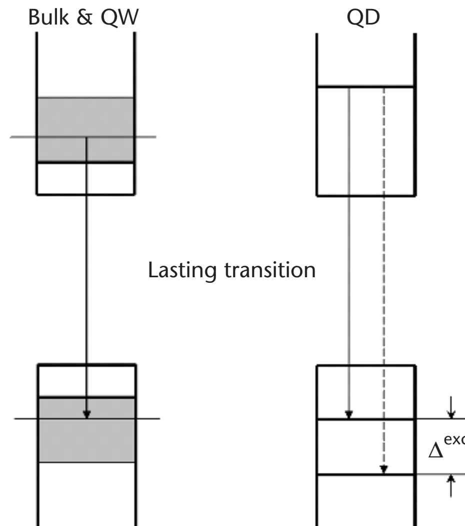


Figure 4.14 Carrier population in bulk, QW, and single QD. The dashed arrow shows the excited state transition in a QD. (*From:* [49]. © 2003 IEEE. Reprinted with permission.)

Suppression of the Effect of Excited States of Holes by *p*-Doping

In QDs supporting multiple closely spaced hole levels, the effect of excited states on the T sensitivity of j_{th} can be strong. Thermal smearing of the hole population among many states means that a large T-dependent fraction of injected holes is not in the ground state. Achieving the population inversion for lasing in such structures requires a T-dependent threshold current. To eliminate this problem, doping QDs with acceptors was proposed [26]. Owing to the larger separation of electron levels (due to the smaller effective mass), the injected electrons lie mostly in their ground state. The large hole occupancy built in by the p doping now ensures that injected electrons always find ground state holes to recombine. A high value of T_0 (160K) was achieved by this method [26].

4.4.5 Spatial Discreteness of Active Elements: Hole Burning

Spatial hole burning (SHB) in semiconductor [50, 51] and solid-state [52] lasers is due to the nonuniformity of stimulated recombination. It leads to the lasing oscillation of higher modes of the resonator together with the main mode. The problem of multimode generation is of primary importance for many laser applications [10]. To suppress the additional modes and properly design single-mode operating lasers, it is essential to understand the key physical processes controlling the multimode generation threshold.

In a QD laser, the SHB effect can be particularly strong in view of the spatial discreteness of QDs [37]. The point is that the stimulated emission in a laser cavity is a standing wave (with several thousands of wavelengths accommodated along the cavity length). The stimulated recombination will be most (least) intensive in the QDs located at the antinodes (nodes) of the light intensity (Figure 4.15). This gives

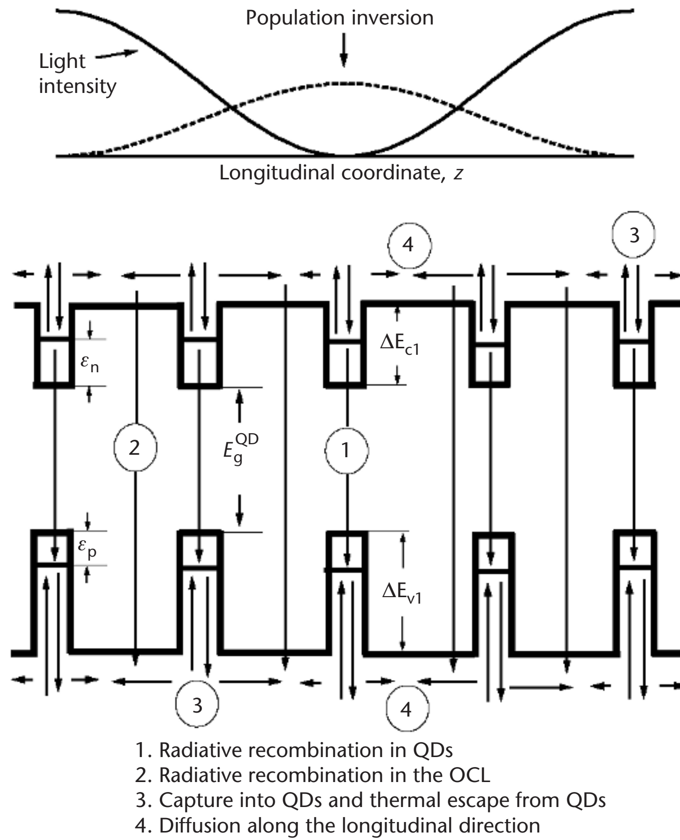


Figure 4.15 Schematic cross section of a QD laser structure along the longitudinal direction.

rise to SHB of the population inversion—QDs located near the antinodes (nodes) are depleted (overfilled).

In semiconductor bulk or QW lasers, the resultant nonuniform carrier distribution is smoothened out by diffusion, which partly or totally suppresses the SHB effect [50].

A drastically different situation occurs in QD lasers. Here, diffusion plays a minor role. Because only those carriers that are localized in the QDs contribute to the stimulated emission, the smoothing of nonuniform population inversion requires a series of transport processes: thermal escape of carriers from the overfilled QDs, diffusion in the OCL to the depleted QDs and capture into the latter (Figure 4.15). This slow carrier exchange between individual QDs may result in a highly nonuniform population inversion, and hence a strong SHB. This in turn implies that the multimode generation threshold may be low. A detailed study of SHB based on the steady-state rate equations (for confined carriers, free carriers, and photons) is presented in [37].

We can see from (4.14) that the multimode generation threshold δj is governed by the characteristic times $\tau_{n,p}^{\text{esc}}$ of the thermal escapes from QDs. Thermal escape,

rather than diffusion, is the slowest process controlling the smoothing of nonuniform population inversion. A similar situation arises in lasers based on band-to-impurity optical transitions [53, 54].

The temperature dependence of δj resides in the escape times $\tau_{n,p}^{\text{esc}}$ [see (4.8) and (4.15)]. The main reason is the exponential temperature dependence of the quantities n_1 and p_1. Concurrent with the unwelcome increase of j_{th}, the increasing T leads to a desirable increase of Δj (Figure 4.16). This occurs because the thermal escape from QDs becomes much more effective at high T. (Similarly, δj increases with T in semiconductor bulk lasers, due to the enhanced diffusion at higher T [50].) Thus, provided SHB is the only (or the main) factor that allows many modes to oscillate simultaneously, the number of the lasing modes decreases with increasing T and hence the LCC becomes more linear. This may explain the observed increase with T of the slope efficiency in a QD laser [55].

4.4.6 Intrinsic Nonlinearity of the Light-Current Characteristic

In every QW, QWR, or QD laser, the quantum-confined active elements are embedded in a bulk reservoir region, which also serves as the OCL, wherefrom carriers are fed via some sort of a capture process. Because the capture is never instantaneous, it gives rise to a current dependence of the carrier density n in the reservoir, even above threshold when the carrier density in the active region itself is pinned. The increasing

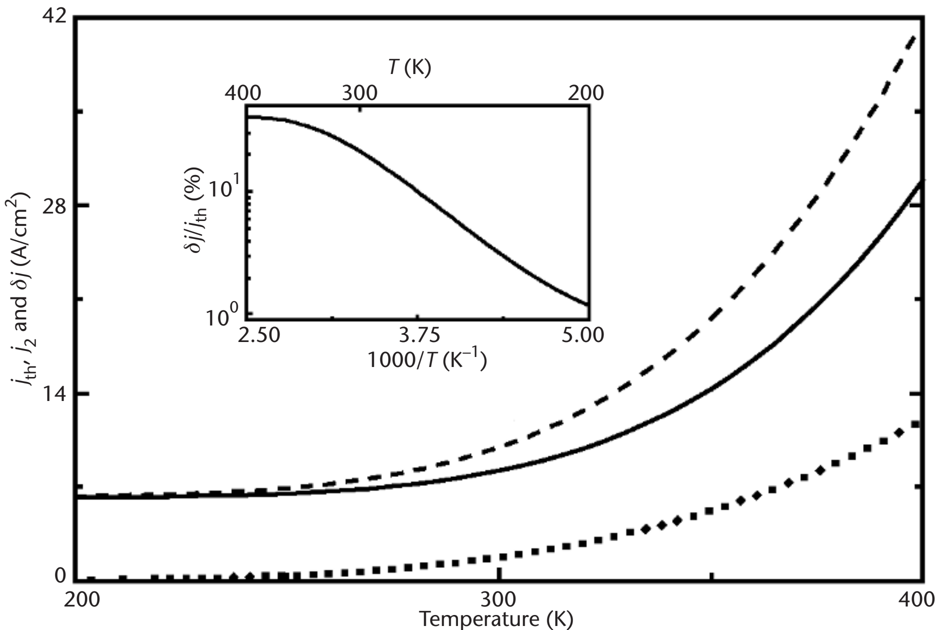


Figure 4.16 Temperature dependence of the multimode generation. Solid and dashed curves describe, respectively, the threshold current density for the main and the next longitudinal modes; the dotted curve corresponds to the multimode generation threshold. The relative multimode generation threshold is shown in the inset. (*From:* [37]. © 2000 IEEE. Reprinted with permission.)

n leads to an increase in the parasitic recombination current in the reservoir, and contributes to a deviation of the internal differential quantum efficiency η_{int} from unity. This fact was noted earlier [56–61] but the actual reduction in η_{int} was quantified only recently [38, 39].

The "reservoir effect," combined with the nonlinear (superlinear in n) dependence of the recombination rate in the reservoir, makes the *light-current characteristic* (LCC) sublinear at high injection currents [38, 39]. The resultant degradation of the LCC is comparable in magnitude to the entire experimentally observed degradation, even neglecting all other known mechanisms of nonlinearity (such as the lattice and carrier heating). This suggests that the reservoir effect is a dominant mechanism limiting both the output power and the linearity of the LCC.

Rate Equations

The steady-state rate equations for carriers confined in the active region, free carriers in the OCL and photons can be written as follows:

$$j_{capt} - j_{esc} - j_{spon}^{active} - j_{stim} = 0 \tag{4.22}$$

$$j_{esc} - j_{capt} - j^{OCL} + j = 0 \tag{4.23}$$

$$j_{stim} - e\frac{1}{S}\frac{N}{\tau_{ph}} = 0 \tag{4.24}$$

where j_{capt} and j_{esc} are, respectively, the current densities of carrier capture into and carrier escape from the active region, j_{spon}^{active} and j_{stim} are the spontaneous and the stimulated recombination current densities in the active region, j^{OCL} is the parasitic recombination current density in the OCL, j is the injection current density, S is the active layer area (the cross section of the junction), N is the number of photons in the lasing mode, and τ_{ph} is the photon lifetime in the cavity.

Taking into account that

$$j_{stim} - \frac{e}{S}\frac{c}{\sqrt{\varepsilon_g}} g^m N \tag{4.25}$$

where $\sqrt{\varepsilon_g}$ is the group refraction index of the OCL, it follows immediately from (4.24) that the modal gain spectrum peak g^m pins above threshold and hence so does the carrier density in the active region.

Because j_{esc} and j_{spon}^{active} are both controlled by the carrier density in the active region, they also clamp above threshold. On the other hand, the capture current is linear in the carrier density n in the OCL:

$$j_{capt} = ev_{capt}n = \frac{ebn}{\tau_{capt}} \tag{4.26}$$

where v_{capt} is the capture velocity (in centimeters per second) and τ_{capt} is the "capture time" into the active region. Thus, we obtain from (4.22)

$$n = n_{th}\left(1 + \frac{j_{stim}}{j_{capt,\ th}}\right) \tag{4.27}$$

where n_{th} and $j_{capt,\ th}$ are the threshold values of n and j_{capt}, respectively (n_{th} is not to be confused with the threshold carrier density in the active region). The slower the carrier supply to the active region (the lower $j_{capt,\ th}$), the larger is $n - n_{th}$.

Using (4.22) and (4.23) and taking into account that j_{spon}^{active} pins above threshold, the excess injection current density $j - j_{th}$ is

$$j - j_{th} = j^{OCL} - j_{th}^{OCL} + j_{stim} \tag{4.28}$$

where $j_{th} = j_{th}^{OCL} + j_{spon}^{active}$ is the threshold current density, with j_{th}^{OCL} being the value of j^{OCL} at $n = n_{th}$.

When the dominant recombination channel in the OCL is spontaneous radiation, then $j^{OCL} \propto n^2$ [with n given by (4.27)]. Using this in (4.28) yields

$$\frac{j - j_{th}}{j_{th}^{OCL}} = \left(1 + \frac{j_{stim}}{j_{capt,\ th}}\right)^2 - 1 + \frac{j_{stim}}{j_{th}^{OCL}} \tag{4.29}$$

Solution of quadratic equation (4.29) gives j_{stim} as a function of $j - j_{th}$; substituting this function into (4.27), we obtain an expression for n (Figure 4.17).

Direct Relationship Between the Power and the Threshold Characteristics

The internal differential quantum efficiency of a semiconductor laser is defined as the fraction of the excess injection current that results in stimulated emission:

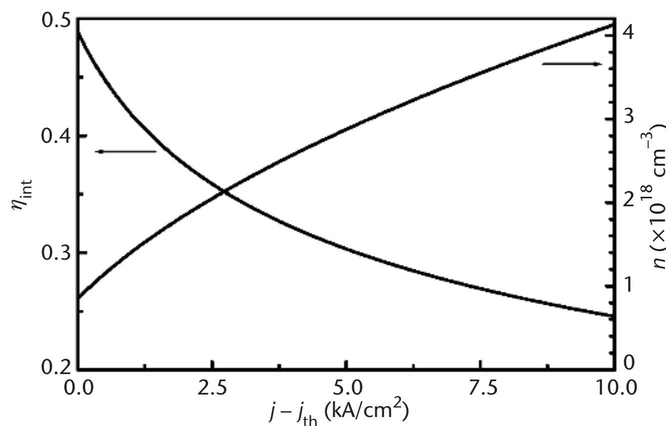


Figure 4.17 Internal quantum efficiency (left axis) and free-carrier density in the OCL (right axis) against excess injection current density, where j_{th} = 336.99 A/cm^2 and $j_{th}^{OCL} / j_{capt,\ th}$ = 0.523 at N_S = 2.9 × 10^{10} cm^{-2}. (*From:* [39]. © 2003 IEEE. Reprinted with permission.)

$$\eta_{\text{int}} = \frac{j_{\text{stim}}}{j - j_{\text{th}}} \tag{4.30}$$

Taking j_{stim} from (4.29), we find the following expression [38, 39]:

$$\eta_{\text{int}} = \frac{1}{\frac{1}{2} + \frac{j_{\text{th}}^{\text{OCL}}}{j_{\text{capt, th}}} + \sqrt{\left(\frac{1}{2} + \frac{j_{\text{th}}^{\text{OCL}}}{j_{\text{capt, th}}}\right)^2 + \frac{j_{\text{th}}^{\text{OCL}}}{j_{\text{capt, th}}} \frac{j + j_{\text{th}}}{j_{\text{capt, th}}}}} \tag{4.31}$$

We see that η_{int} is a decreasing function of $j - j_{\text{th}}$ (Figure 4.17).

The output optical power is of the form

$$P = \frac{\hbar\omega}{e} S(j - j_{\text{th}})\eta_{\text{int}} \frac{\beta}{\beta + \alpha_{\text{int}}} \tag{4.32}$$

where β and α_{int} are the mirror and the internal losses, respectively. Thus, the output power is sublinear in the injection current (Figure 4.9). This mechanism of nonlinearity is inherent to quantum-confined lasers of arbitrary dimensionality. Equation (4.31) for η_{int} is universal—it retains the same form for QD, QWR, and QW lasers.

For a given $j - j_{\text{th}}$, the internal quantum efficiency (Figure 4.12) and the output power are controlled by the dimensionless parameter $j_{\text{capt, th}}/j_{\text{th}}^{\text{OCL}}$, which is the ratio of the recombination current in the reservoir to carrier capture current, both taken at threshold. The lower this ratio, the closer η_{int} is to unity (Figure 4.12) and the more linear the LCC (Figure 4.9). Ideally, when this ratio vanishes (e.g., when $j_{\text{th}}^{\text{OCL}} = 0$—no recombination in the OCL), one has $\eta_{\text{int}} = 1$ at an arbitrary injection current and the LCC is linear. In general, however, $j_{\text{th}}^{\text{OCL}}$ is a tangible fraction of the total j_{th}, and $\eta_{\text{int}} < 1$ even at $j = j_{\text{th}}$. In optimizing the structure for minimum j_{th}, it is this component that should be suppressed first and foremost [30].

The conclusion that the power performance of a laser is linked to its threshold characteristics is of great importance. The higher the excess of the injection current over the threshold current, the stronger this relation is manifested (Figures 4.9, 4.12, and 4.17). The higher the required output power, the lower should be the threshold current (Figures 4.9, 4.12, and 4.17). Since QD lasers possess the lowest j_{th} among all current semiconductor lasers, these results suggest their other potential advantage, namely the possibility of achieving highest output powers.

The higher the excess current $j - j_{\text{th}}$, the larger the fraction of it that goes into the parasitic recombination outside the active region. To accommodate carrier consumption by the active region, carriers accumulate in the OCL much in excess of their threshold amount. The resultant superlinear increase in parasitic recombination degrades the LCC.

Thus, at high injection currents the LCC is strongly sublinear (Figure 4.9); n and P increase as $\sqrt{j - j_{\text{th}}}$ (Figures 4.17 and 4.9), while η_{int} decreases as $1 / \sqrt{j - j_{\text{th}}}$ (Figure 4.17):

$$\eta_{\text{int}} = \frac{j_{\text{capt, th}}}{\sqrt{j_{\text{th}}^{\text{OCL}}\left(j - j_{\text{th}}\right)}} \tag{4.33}$$

$$n = n_{\text{th}} \sqrt{\frac{j - j_{\text{th}}}{j_{\text{th}}^{\text{OCL}}}} \tag{4.34}$$

$$P = \frac{\hbar\omega}{e} S j_{\text{capt, th}} \sqrt{\frac{j - j_{\text{th}}}{j_{\text{th}}^{\text{OCL}}}} \frac{\beta}{\beta + \alpha_{\text{int}}} \tag{4.35}$$

These square root dependences result from the assumed bimolecular ($\propto n^2$) recombination in the OCL.

The higher the degree of superlinearity of the recombination rate in the OCL with respect to n, the higher the degree of sublinearity of the LCC. Because the nonradiative Auger recombination rate increases as n^3, this recombination channel becomes dominant at sufficiently high injection currents. In the Auger limit, the difference $j^{\text{OCL}} - j_{\text{th}}^{\text{OCL}}$ in (4.28) will be dominated by the cubic (in j_{stim}) term, that is, $j - j_{\text{th}} \propto j_{\text{stim}}^3$. Hence, both j_{stim} and P will be proportional to $\sqrt[3]{j - j_{\text{th}}}$ and $\eta_{\text{int}} = j_{\text{stim}} / (j - j_{\text{th}}) \propto j_{\text{stim}} / j_{\text{stim}}^3 = 1 / j_{\text{stim}}^2 \propto 1 / (j - j_{\text{th}})^{2/3}$.

Thus, the actual shape of the nonlinear LCC depends on the dominant recombination channel outside the active region. It is therefore possible to identify the dominant channel by analyzing the LCC shape.

Capture into the Active Region

The ratio $j_{\text{th}}^{\text{OCL}} / j_{\text{capt, th}}$ in (4.31) can be written as follows:

$$\frac{j_{\text{th}}^{\text{OCL}}}{j_{\text{capt, th}}} = \frac{\tau_{\text{capt, th}}}{\tau_{\text{th}}^{\text{OCL}}} \tag{4.36}$$

where the time constants

$$\tau_{\text{th}}^{\text{OCL}} = \frac{1}{B n_{\text{th}}} \tag{4.37}$$

$$\tau_{\text{capt}} = \frac{b}{v_{\text{capt}}} = \frac{\tau_{\text{capt, 0}}}{1 - f_n} \tag{4.38}$$

may be regarded, respectively, as the recombination time in the OCL at the lasing threshold and the "capture time" into the active region; f_n is the level occupancy in the active region at the lasing threshold. (As already discussed, f_n is pinned above threshold.)

The time constant $\tau_{\text{capt, 0}}$ is the “capture time” into an empty active region. For a QD ensemble,

$$\tau_{\text{capt, 0}} = \frac{1}{\sigma_n v_n \dfrac{N_S}{b}} \tag{4.39}$$

For a specific structure considered in [38, 39], $\tau_{\text{capt, 0}} = 10^{-10}$ sec.

With (4.38), the stimulated recombination current density and the output power can be written in the form

$$j_{\text{stim}} = \frac{eb(n - n_{\text{th}})}{\tau_{\text{capt}}} \tag{4.40}$$

$$P = \hbar\omega \frac{V_{\text{OCL}}(n - n_{\text{th}})}{\tau_{\text{capt}}} \frac{\beta}{\beta + \alpha_{\text{int}}} \tag{4.41}$$

where V_{OCL} is the OCL volume. Equations (4.40) and (4.41) reflect the obvious fact that the supply of carriers to the active region (the origin of all stimulated photons and optical power) occurs by the capture process.

For both nonlinear recombination channels in the OCL, either spontaneous radiation, or nonradiative Auger recombination, $\tau_{\text{th}}^{\text{OCL}}$ depends on n_{th} (as $\tau_{\text{th}}^{\text{OCL}} \propto 1 / n_{\text{th}}$ or $1 / n_{\text{th}}^2$, respectively). The “capture time” τ_{capt} is inversely proportional to the 3D density of unoccupied states in the QD ensemble $(N_S / b)(1 - f_n)$. Both f_n and n_{th} depend on the structure parameters (see Section 4.4.7). For this reason, $\tau_{\text{th}}^{\text{OCL}}$ and τ_{capt} are not the true time constants describing the respective processes. Further still, τ_{capt} is a characteristic of the entire QD ensemble, rather than of a single QD. In the strict sense, the capture time into a single QD cannot be introduced properly (though this is sometimes done in the literature). The adequate physical quantity, describing correctly the carrier capture into a single QD, is the capture cross section σ_n.

4.4.7 Critical Sensitivity to Structure Parameters

As discussed earlier, the modal gain is limited by its saturation value $g^{\max}$ [see (4.3)]. This implies that lasing in QD structures is possible only in a certain range of tolerable parameters. This range is given by the inequality $g^{\max}[N_S, (\Delta\varepsilon)_{\text{inhom}}] \geq \beta(L)$ (Figure 4.18). The boundary of the tolerable region (the surface $g^{\max} = \beta$ in Figure 4.18) determines the critical tolerable parameters of the structure—the minimum surface density of QDs $N_S^{\min}$, the maximum rms of relative QD size fluctuations $\delta^{\max}$ and the minimum cavity length $L^{\min}$ [30, 37]:

$$N_S^{\min} = \frac{4}{\xi}\left(\frac{\sqrt{\varepsilon}}{\lambda_0}\right)^2 \tau_{\text{QD}} \frac{(\Delta\varepsilon)_{\text{inhom}}}{\hbar} \frac{a}{\Gamma} \frac{1}{L} \ln\frac{1}{R} \tag{4.42}$$

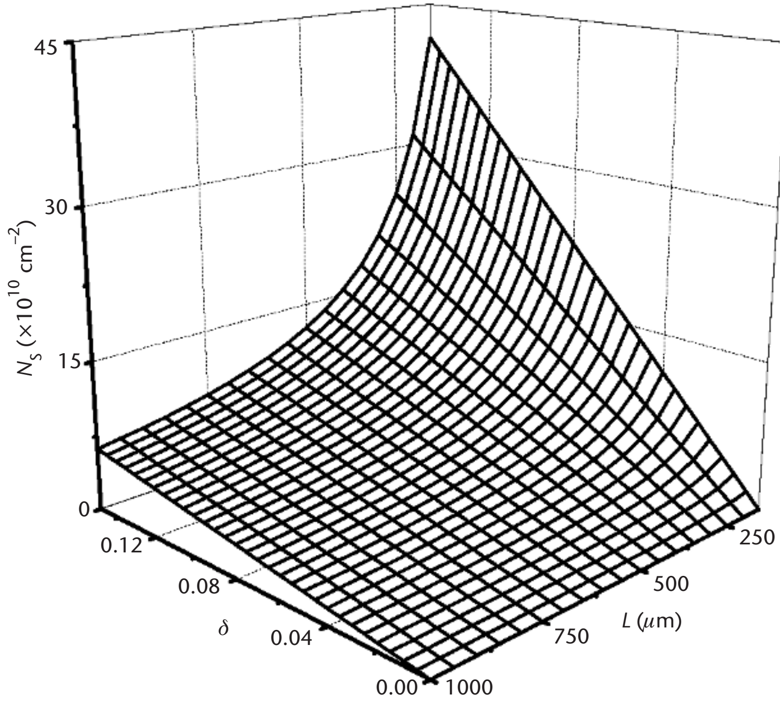


Figure 4.18 Tolerable values of structure parameters. The three axes correspond to the surface density of QDs, the rms of relative QD size fluctuations, and the cavity length. The range of tolerable values of these parameters lies above the shown surface. (*From:* [37]. © 2000 IEEE. Reprinted with permission.)

$$\delta^{\max} = \frac{\xi}{4}\left(\frac{\lambda_0}{\sqrt{\varepsilon}}\right)^2 \frac{1}{\tau_{\mathrm{QD}}}\frac{\Gamma}{a}\frac{\hbar}{\left(q_n\varepsilon_n + q_p\varepsilon_p\right)}\left(\ln\frac{1}{R}\right)^{-1} LN_s \tag{4.43}$$

$$L^{\min} = \frac{4}{\xi}\left(\frac{\sqrt{\varepsilon}}{\lambda_0}\right)^2 \tau_{\mathrm{QD}}\frac{a}{\Gamma}\frac{(\Delta\varepsilon)_{\mathrm{inhom}}}{\hbar}\frac{1}{N_S}\ln\frac{1}{R} \tag{4.44}$$

where R is the facet reflectivity.

The more perfect the structure (lower δ) or the longer the cavity, the lower is $N_S^{\min}$. The denser the QD ensemble (the greater N_S) or the longer the cavity, the greater is $\delta^{\max}$. The more perfect the structure or the denser the QD ensemble, the shorter is $L^{\min}$.

The threshold condition can be expressed [37] in terms of the critical parameters as follows:

$$f_n + f_p - 1\frac{\delta}{\delta^{\max}} \tag{4.45}$$

Equation (4.45) describes the population inversion in a mean-sized QD required for the lasing at a given δ. The ratio δ / δ^{max} on the right side of (4.45) [and, similarly, (4.9)] can be equivalently replaced by N_S^{min} / N_S or L^{min} / L.

The fact that lasing is possible only for a certain range of structure parameters is inherent to all types of lasers. This is of particular importance for QD lasers in view of the great variability of their structure parameters (such as N_S and δ). In a well-designed QD laser, all parameters should be far away from their critical values.

Threshold Characteristics

Equation (4.45) shows that as a structure parameter approaches its critical value ($N_S \to N_S^{min}$, or $\delta \to \delta^{max}$, or $L \to L^{min}$, both f_n and f_p tend to unity. This means that both the electron and the hole QD levels become fully occupied, which requires infinitely high free-carrier densities in the OCL [see (4.7) and (4.10)] and hence infinitely high pumping: $j_{th} \to \infty$ [see (4.11) and Figure 4.5].

Multimode Generation Threshold

As any structure parameter approaches its critical value, one sees the divergence of not only the threshold current density for the main mode j_{th} but also of the threshold current density for the next mode j_2 and of the multimode generation threshold δj [see (4.14) and Figure 4.5]. At the same time, the relative multimode generation threshold $\delta j / j_{th}$ drops to zero (see the insets in Figure 4.5), which means an infinite increase in the number of simultaneously generated longitudinal modes.

Power Characteristics

The level occupancy in a QD at the lasing threshold ranges within $1/2 < f_n < 1$ [here we assume QD charge neutrality and use (4.9)]. The low value of 1/2 corresponds to a vanishing QD size dispersion or negligible total loss, when the lasing threshold is close to the transparency threshold.

As discussed earlier [see (4.31)], reducing $j_{th}^{OCL} / j_{capt,\,th}$ is a key to enhancing the internal quantum efficiency and the output power. The minimum value of this ratio (obtained when $f_n = 1/2$) is typically much less than unity. Thus, for a specific structure considered in [38, 39], it is 10^{-3} at room temperature.

When the structure parameter is close to its critical value, then $f_n \to 1$ and $j_{th}^{OCL} / j_{capt,\,th} \to \infty$. As a result, the internal quantum efficiency and the output power both drop to zero (Figure 4.8).

In view of the critical dependence of the ratio $j_{th}^{OCL} / j_{capt,\,th}$ on structure parameters, it is crucial to optimize the QD laser design. Thus, for the optimized structure considered in [38, 39] ($f_n^{opt} = 0.655$), this ratio is 3×10^{-3}, which is close to its minimum value. This means that the LCC of such an optimized structure will be linear up to very high injection currents (Figure 4.19).

Critical structure parameters for threshold and power characteristics also exist in QWR and QW lasers.

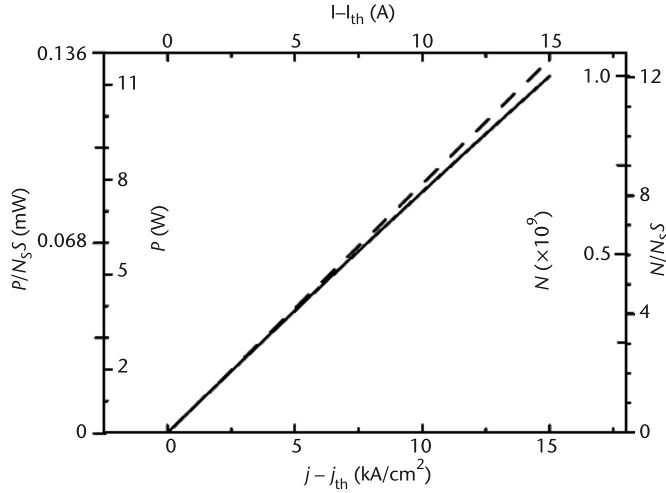


Figure 4.19 Injection current density dependence of the output power (left axis) and photon number (right axis) for a broad-area ($W = 100\ \mu$m) optimized QD laser ($N_S^{opt} = 8.268 \times 10^{10}$ cm^{-2}, $b^{opt} = 0.227\ \mu$m, and $j_{th} = 12.41$ A/cm^2). The dashed line corresponds to the ideal situation, $\eta_{int} = 1$. Power and photon number per QD are also shown. The top axis shows the excess injection current. (*From:* [39]. © 2003 IEEE. Reprinted with permission.)

4.4.8 Dependence of the Maximum Gain on the QD Shape

The maximum modal gain of a laser g^{max} depends on the shape of QDs [47, 48] and it is higher for more symmetrical QDs. It should therefore be possible to improve the laser performance by engineering optimal QD shapes. The crucial factor is the overlap integral I between the electron and the hole wave functions, which strongly affects g^{max} [$g^{max} \propto I$, see (4.3) and (4.4)]. Ideally, in QDs of highly symmetrical (e.g., cubic) shape, this integral is close to unity and the spontaneous radiative lifetime τ_{QD} is about 1 ns or shorter. In this case, the saturation value of the modal gain g^{max} for a single layer of QDs is a few tens of cm^{-1}, which is high enough to ensure lasing even in short (submillimeter) cavities. The actual self-organized QDs are of lower symmetry (e.g., pyramidal shape) and the overlap integral between the ground electron and hole states is several times less than unity [45]. This makes g^{max} as low as several cm^{-1} [47, 48] and lasing in short cavities becomes impossible. Clearly, it would be of tremendous value to control the shape of self-organized QDs.

Recently, a method was developed to grow more symmetrical (truncated) QDs [28]. A laser diode with truncated QDs outperformed a reference sample with nontruncated, pyramidal QDs. The authors demonstrated almost a factor of 2 increase in g^{max} and an extremely high T_0 = 380K (at temperatures up to 55°C). This is the highest value of T_0 ever reported in this temperature range.

An alternative technique of shape-engineering was developed in [62], based on controlling the QD heights. These authors also demonstrated a significant change in the optical properties of their structures.

4.4.9 Internal Optical Loss

Internal optical loss adversely affects operating characteristics of semiconductor lasers. Because of the lower value of the optical confinement factor, the effect of internal loss is stronger for lasers with a reduced-dimensionality active region, such as QW, QWR, and QD lasers, than for bulk lasers [10].

All different processes, contributing to the internal loss, can be grouped into two categories, one (such as free-carrier-absorption in the OCL, or simply the waveguide) dependent on the injection carrier density, the other (such as interface scattering or absorption in the cladding layers) insensitive to this density [63]. Absorption in the active region of QW and QWR lasers is relatively small compared to absorption in the OCL, at least at high injection currents j (or high temperatures T—see [64, 65]); the analogous process in the active region of QD lasers, which is carrier photoexcitation from the discrete levels to the continuous-spectrum states, is also small [30, 66]. Neglecting these processes, we must be concerned only with the free-carrier density n in the OCL. Therefore, we need a relation between n and the occupancy of states in the quantum-confined active region, involved in the lasing transition. At sufficiently high temperatures and below the lasing threshold, this relation is given by equilibrium statistics and is of the form of (4.7).

Assuming equal electron and hole occupancies ($f_n = f_p$) and writing the total net internal loss coefficient α_{int} (the quantity we shall refer to simply as the *internal loss*) as the sum of a constant α_0 and a component linear in n, the lasing threshold condition is brought into the form [63] [see (4.1)]:

$$g^{\max}(2f_n - 1) = \beta + \alpha_0 + \sigma_{\text{int}} n_1 \frac{f_n}{1 = f_n} \tag{4.46}$$

where $g^{\max}$ is the maximum (saturation) value of the modal gain $g^{\text{m}}(f_n) = g^{\max}(2f_n - 1)$ [see (4.2)], β is the mirror loss, and $\sigma_{\text{int}} = \text{const}(n)$ can be viewed as an effective cross section for all absorption loss processes (for the type of carrier that dominates absorption).

The solutions of (4.46) are [Figure 4.20(a)]

$$f_{n1,\,n2} = f_n^{\text{crit}} \mp \sqrt{\left(f_n^{\text{crit}}\right)^2 - f_{n0} - \frac{1}{2}\frac{\alpha_0}{g^{\max}}} \tag{4.47}$$

where

$$f_n^{\text{crit}} + \frac{1}{2}\left(1 + f_{n0} + \frac{1}{2}\frac{\alpha_0}{g^{\max}} - \frac{1}{2}\frac{\sigma_{\text{int}} n_1}{g^{\max}}\right) \tag{4.48}$$

is the "critical" solution [when the cavity length equals its minimum tolerable value—see (4.53) later] and

$$f_{n0} = \frac{1}{2}\left(1 + \frac{\beta}{g^{\max}}\right) \tag{4.49}$$

is the solution in the absence of internal loss; (4.49) is an equivalent representation of (4.9).

Both solutions (4.47) are physically meaningful and describe two distinct lasing thresholds [63]. The lower solution, f_{n1}, is the conventional threshold, similar to f_{n0} but modified by α_{int}. The second solution, f_{n2}, appears purely as a consequence of the carrier density-dependent α_{int} in the OCL.

In the absence of lasing, the injection current density has the following relation to f_n [30, 39]:

$$j = \frac{eN_S}{\tau_{\text{QD}}} f_n^2 + ebBn_1^2 \frac{f_n^2}{(1-f_n)^2} \tag{4.50}$$

The lower and the upper threshold current densities, j_{th1} and j_{th2}, are given by (4.50) wherein one substitutes either $f_n = f_{n1}$ or $f_n = f_{n2}$.

The existence of a second lasing threshold stems from the nonmonotonic dependence of $g^{\text{m}} - \alpha_{\text{int}}$ on f_n [Figure 4.20(a)], or, equivalently, on n or j [Figure 4.20(b)]. The point is that the modal gain $g^{\text{m}}(f_{\text{n}}) = g^{\max}(2f_{\text{n}} - 1)$ increases linearly with f_{n} [Figure 4.20(a)] and saturates at its maximum value $g^{\max}$ as $f_n \to 1$ [which corresponds to $n \to \infty$ and $j \to \infty$—see (4.7), (4.50), and Figure 4.20(b)]. At the same time, α_{int} is superlinear in f_n [see (4.46) and Figure 4.20(a)] and increases infinitely as $f_n \to 1$. At a certain f_n, that is, at a certain j, the rate of increase in α_{int} with j will inevitably equal that of increase in g^{m}, and hence the difference $g^{\text{m}} - \alpha_{\text{int}}$ will peak. Any further increase of j will decrease the difference $g^{\text{m}} - \alpha_{\text{int}}$ [see Figure 4.20(b)]. This corresponds to the so-called "loss-multiplication" regime, discussed in [64, 65] for QW lasers (and attributed to the pileup of carriers due to electrostatic band-profile deformation [44, 67]) and in [68, 69] for QD lasers. As evident from our analysis, this regime and the second lasing threshold are inherent to all structures where α_{int} depends on n.

Note that the second lasing threshold can also arise due to other mechanisms, for example, carrier heating. As the carrier temperature increases with j [9, 44, 70–72], the modal gain itself can become a nonmonotonic function of j, decreasing at high j [70, 71].

In a CW operation, increasing j from zero, one reaches the first lasing threshold j_{th1}. Above this threshold, the difference between the gain and the internal loss is pinned at the value of the mirror loss β and hence Figure 4.20 (which is valid for determining the positions of both thresholds) no longer applies. What actually happens above j_{th1} is shown in Figure 4.21, derived by rigorously solving the rate equations in the presence of light generation [63].

Above the second threshold j_{th2} and up to a maximum current $j_{\max}$, both the gain-current dependence [Figure 4.21(a)] and the LCC [Figure 4.21(b)] are two-valued. At $j = j_{\max}$, the two branches merge in both characteristics. The origin of this striking

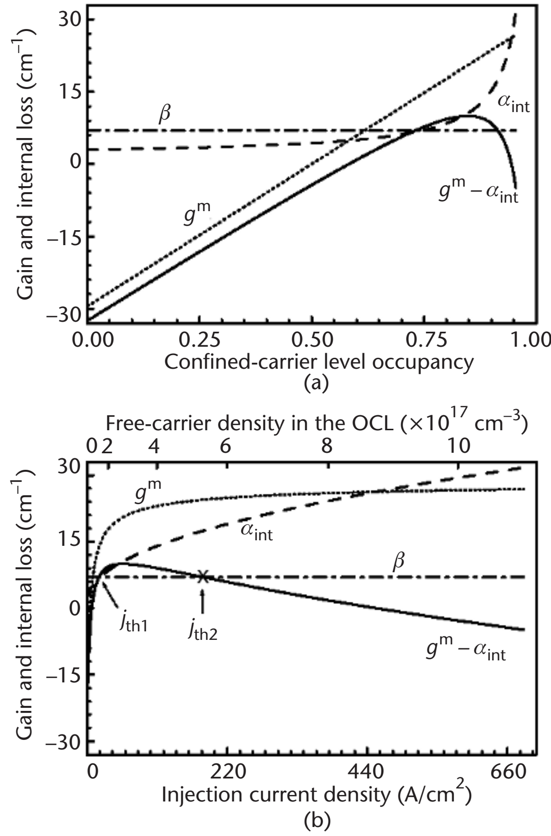


Figure 4.20 Illustration of the threshold condition (4.46) and the two lasing thresholds. Modal gain $g^m = g^{max}(2f_n - 1)$ [inclined dotted line in (a) and dotted curve in (b)], internal loss $\alpha_{int} = \alpha_0 + \sigma_{int}\, n = \alpha_0 + \sigma_{int}\, n_1\, f_n / (1 - f_n)$ (dashed curve) and difference of modal gain and internal loss (solid curve) against confined-carrier level occupancy in the active region f_n (a), free-carrier density in the OCL n [(b), top axis] and injection current density j [(b), bottom axis]. The intersections of the solid curve and the horizontal dash-dotted line for the mirror loss $\beta = (1/L)\ln(1/R)$ define the solutions (4.47) of (4.46). We assume a GaInAsP/InP-based QD heterostructure lasing near 1.55 μm [30, 39] with 10% QD size fluctuations and $N_S = 6.11 \times 10^{10}$ cm^{-2} (for these parameters, g^{max} = 29.52 cm^{-1}). At T = 300 K, $n_1 = 5.07 \times 10^{16}$ cm^{-3}. The mirror loss β = 7 cm^{-1}. Parameters α_0 and s_{int} are plausibly taken as 3 cm^{-1} and 2.65×10^{-17} cm^{-1}, respectively. (*From:* [63]. © 2003 American Institute of Physics. Reprinted with permission.)

behavior is clear. As α_{int} increases with the current, the gain strictly follows it so as to maintain the stable generation condition $g^m - \alpha_{int} = \beta$. This continues up to the maximum pump current j_{max} at which the lasing is quenched.

At this time, we cannot propose a definite experimental technique to access the lower branch of the LCC [Figure 4.21(b)]. (Analysis of the stability of the lower branch regime will be published elsewhere.) Nevertheless, we stress that an experimental determination of the second threshold would provide us with a new and valuable technique for measuring the loss parameters. Indeed, with the measured j_{th1} and j_{th2}, the values of f_{n1} and f_{n2} can be calculated from (4.50). Both α_0 and σ_{int} can then be expressed in terms of f_{n1} and f_{n2} as follows [see (4.47) and (4.48)]:

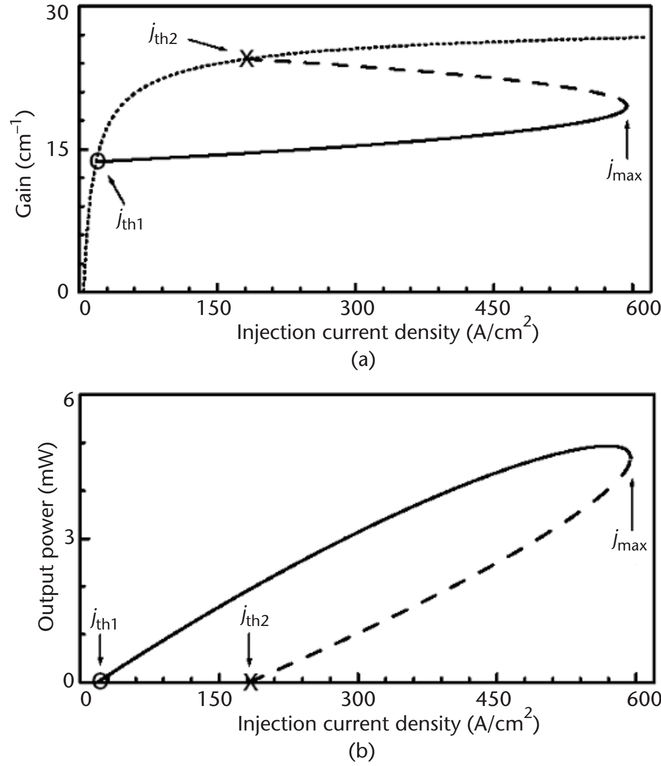


Figure 4.21 Two-valued characteristics: (a) gain-current and (b) light-current. The branches corresponding to the lower and upper lasing regimes (solid and dashed curves, respectively) merge together at the point j_{max} which defines the maximum operating current. At $j > j_{max}$, the lasing is quenched. The dotted curve in (a) is the gain-current dependence for a nonlasing regime; the intersection of the solid (dashed) curve and the dotted curve determines j_{th1} (j_{th2}). In (b), the assumed stripe width $w = 2\ \mu$m. (*From:* [63]. © 2003 American Institute of Physics. Reprinted with permission.)

$$\alpha_0 = 2g^{\max}\left(f_{n1}f_{n2} - f_{n0}\right) \tag{4.51}$$

$$\sigma_{\text{int}} = 2\frac{g^{\max}}{n_1}\left(1 + f_{n1}f_{n2} - f_{n1} - f_{n2}\right) \tag{4.52}$$

When only one threshold exists, the carrier density-dependent internal loss is negligible and α_0 is determined from the "–" solution in (4.47).

The thresholds j_{th1} and j_{th2} depend on the cavity length L (Figure 4.22) and approach each other as L decreases. At a certain critical L that we shall call the minimum tolerable cavity length $L^{\min}$, the horizontal line for the mirror loss β is tangent to the curve for $g^{\text{m}} - \alpha_{\text{int}}$ at its maximum (Figure 4.20). In this case, the threshold condition has only one solution, her $f_{n1} = f_{n2} = f_n^{\text{crit}}$, and $j_{\text{th1}} = j_{\text{th2}}$. For $L < L^{\min}$, there is no solution of the threshold condition and hence no lasing is possible. The equation for $L^{\min}$ is

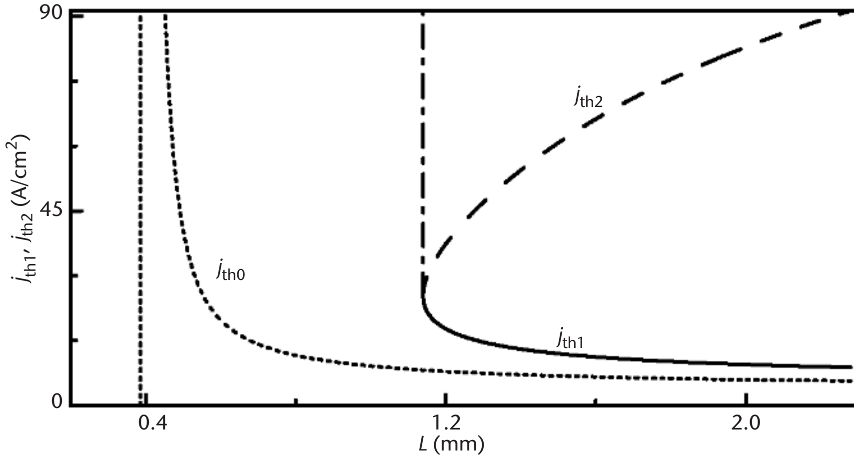


Figure 4.22 The lower and upper threshold current densities (solid and dashed curves, respectively), j_{th1} and j_{th2}, against L. The curve for j_{th1} joins smoothly the vertical dash-dotted line at the critical point. The dotted curve and the vertical dotted line show the threshold current density j_{th0} and its asymptote at the critical point in the absence of internal loss. (*From:* [63]. © 2003 American Institute of Physics. Reprinted with permission.)

$$L^{\min} = \frac{L_0^{\min}}{\left(\sqrt{2} - \sqrt{\dfrac{\sigma_{\text{int}} n_1}{g^{\max}}}\right)^2 - 1 - \dfrac{\alpha_0}{g^{\max}}} \tag{4.53}$$

where $L_0^{\min} = (1/g^{\max}) \ln (1/R)$ is the minimum cavity length in the absence of α_{int} given by (4.44). In (4.48), $\beta^{\max} = (1/L^{\min}) \ln (1/R)$ should be entered for β in this case.

Measurement of $L^{\min}$ provides yet another way of determining the internal loss parameters. For example, $L^{\min}$ can be measured for two structures characterized by different mirror reflectivities with other parameters invariant. With these two values of $L^{\min}$, (4.53) will give a set of two equations for α_0 and α_{int}.

Note that $L^{\min}$ sets a considerably more stringent restriction than $L_0^{\min}$. Thus, for a QD laser similar to that considered in [30, 39] and taking plausible values $\alpha_0 = 3$ cm^{-1} and $\sigma_{\text{int}} = 2.65\times10^{-17}$cm^{-1}, we find the maximum tolerable mirror loss to be $\beta^{\max} = 10$ cm^{-1}. Assuming as-cleaved facet reflectivity at both ends ($R = 0.32$), this yields $L^{\min} = 1.139$ mm, which is almost a threefold increase compared to $L_0^{\min} = 386\ \mu$m. Hence, the absence of lasing often observed in short-cavity QD structures can be attributed to the internal loss. This conclusion is consistent with the discussion in [68, 69].

All equations and analysis of this section apply equally to QW, QWR, and QD lasers. One specifies the type of laser by substituting the relevant expression for $g^{\max}$ and the appropriate relation between the spontaneous radiative recombination current density in the quantum-confined active region and f_n [30, 39].

4.5 Novel Designs of QD Lasers with Improved Threshold and Power Characteristics

4.5.1 Temperature-Insensitive Threshold

At relatively high temperatures, the dominant source of the T dependence of j_{th} in all semiconductor lasers originates from carriers that do not contribute to the lasing transition. In lasers with 3D (bulk) or 2D (QW) active regions, there is always a population of carriers within some range around the lasing transition energy (Figure 4.14). These carriers reside in the active region itself and their recombination contributes to a T-dependent threshold.

It is the absence of parasitic recombination in the active medium itself that gave rise to the original hopes of ultra-high-temperature stability in QD lasers, where optical transitions occur between discrete levels. However, as discussed earlier, in all conventional QD laser designs the problem of parasitic recombination has not been removed. This recombination arises primarily from carriers residing in the OCL. When the entire injection current is consumed in QDs ($j_{th} = j_{QD}$), the remaining contributions to the T dependence come from (1) recombination in nonlasing QDs [see (4.13)], (2) recombination via nonlasing, excited states in QDs [see (4.21)], and (3) violation of charge neutrality in QDs [see (4.20)]. In this case, there should be only a slight T dependence of j_{th}, and T_0 should be very high.

Thus, we can expect that suppression of the OCL recombination alone will result in a dramatic improvement of the temperature stability.

Tunneling-Injection QD Laser

One way of suppressing the OCL recombination is based on the idea of tunneling injection of carriers into the QDs [36]. Basically, a tunneling-injection QD laser is a separate confinement double heterostructure (Figure 4.23). Electrons and holes are injected from n- and p-cladding layers, respectively. The QD layer, located in the central part of the OCL, is clad on both sides by QWs separated from the QDs by thin barrier layers. Injection of carriers into QDs occurs by tunneling from the QWs.

A fraction of injected carriers might not recombine in QDs but escape in a second tunneling step into the "foreign" QW and recombine with the majority carriers there. The size of this fraction depends on the ratio of the escape tunneling rate to the QD recombination rate and is practically independent of T. Clearly, the escape tunneling does not lead to a T dependence of j_{th}.

The key idea of the tunneling-injection device is that carriers cannot bypass the QDs on their way from one QW to another. This means that QDs become the sole reservoir delivering minority carriers to the adjacent QW and OCL regions on either side. Hence, outside the QDs there will be no region in the structure where both electron and hole densities are simultaneously high. The electron density is high where the hole density is negligible, and vice versa. Only in the QDs themselves will there be a nonvanishing spontaneous radiative recombination rate. The suppressed parasitic components of j_{th} would otherwise give the main contribution to temperature dependence.

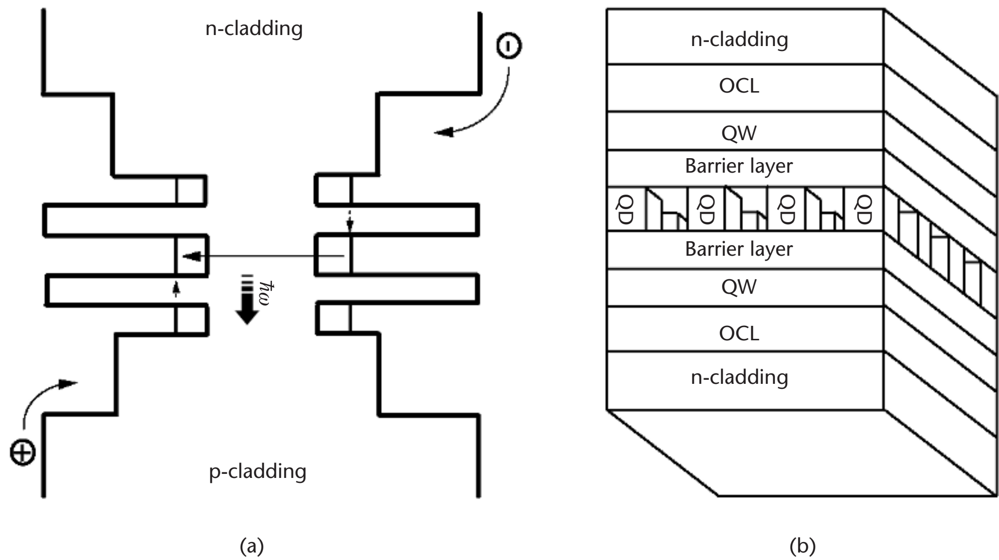


Figure 4.23 (a) Energy band diagram of a tunneling-injection QD laser and (b) schematic view. The QWs and the QDs (not drawn to scale) are assumed implemented in the same material, although this does not necessarily have to be the case in general. The electron-injecting QW is wider than the hole-injecting QW and both QWs are narrower than the QD to accomplish resonant alignment of the majority-carrier subbands with the QD energy levels. The tunnel barrier on the electron-injecting side is made thicker to suppress the leakage of holes from the QD. (*From:* [36]. © 2001 IEEE. Reprinted with permission.)

With the parasitic recombination outside QDs suppressed, there may be only a slight T dependence of $j_{\mathrm{th}} = j_{\mathrm{QD}}$ arising from the remaining contributions items (1)–(3) listed earlier.

Moreover, it is interesting to note that even these remaining contributions are strongly suppressed in the idealized structure (Figure 4.23) due to the resonant nature of tunneling injection. These "finer" effects discussed later further enhance the temperature stability. In an idealized structure, we can expect a literally infinite T_0.

First, the resonant nature of tunneling injection leads to an effective narrowing of the inhomogeneous linewidth. Indeed, such mechanism of injection inherently selects the QDs of the "right" size, since it requires the confined-carrier levels to be in resonance with the lowest subband states in the QW. When this condition is met by the QDs of average size, the number of active QDs will be maximized. Selective injection means that nonlasing QDs are not pumped at all. Second, the tunneling injection selectively removes the supply of carriers to the nonlasing excited states in a lasing QD. Third, the resonant nature of tunneling injection favors correlation of the QD occupancies by electrons and holes, so that all active QDs will tend to remain neutral.

It is encouraging to note recent developments related to the concept of a tunneling-injection QD laser. In [72–74], a CW room temperature (both photo-pumped and diode) lasing was demonstrated in the InAlP–In(AlGa)P–InP visible-

red system (λ = 656–679 μm), wherein the InP QDs were coupled by resonant tunneling to a single InGaP QW. In the later work [75, 76], this idea was realized in the AlGaAs–GaAs–InGaAs–InAs infrared system, wherein the InAs QDs were coupled to the InGaAs QWs. In contrast to [72–74], the QD layer in [75, 76] was sandwiched between the two QWs; like the structure in Figure 4.23(b), the QWs were of different thickness. The use of tunnel coupling of QDs to QW(s) improved the carrier collection and localization in the QD region and reduced j_{th}. In [77], the tunneling transition rate from a QW to a model (disk-shaped) QD was calculated. In [78], introducing the tunnel-injector QW from only the *n*-side of the structure increased significantly T_0 of the InGaAs/GaAs QD laser (T_0 = 237K at room temperature). In [79], T_0 = 363K was reported in the temperature range of 278 to 333K, and T_0 = 202K in the temperature range 333 to 373K.

Bandgap-Engineered QD Laser

Tunneling injection suppresses parasitic recombination by ensuring that the electron density is high where the hole density is negligible, and vice versa. An alternative approach to accomplish this goal can be based on the ability to independently control the potential barriers and fields acting on electrons and holes in the same physical region [49].

In the structures in Figure 4.24, the QD layer is embedded in the OCL in such a way that there are only low barriers [Figure 4.24(a)] or no barrier at all [Figure 4.24(b)] for injection of electrons (from the left) and holes (from the right) into the QDs. On the other hand, the structures are provided with large impenetrable escape barriers that block electron injection into the right-hand side of the structure, where holes are majority carriers, and hole injection into the left-hand side of the structure, where electrons are in abundance. Heterostructure barriers as in Figure 4.24(a) can readily be found within the manifold of quaternary III–V heterojunctions, both strained and lattice matched. The structure in Figure 4.24(b) is a limit case, which serves to illustrate that no barrier for injected carriers is necessary on the injecting side.

The space within the QD layer between the QDs can comprise either of the two barriers or be implemented as a wider gap semiconductor that blocks both carrier types.

In contrast to the ideal tunneling-injection QD laser, the three remaining minor sources (1)–(3) of temperature dependence will “survive” in bandgap-engineered QD lasers. Nevertheless, suppression of the main mechanism alone will make such lasers practically temperature insensitive.

4.5.2 Enhanced Power Performance

The radically new design strategy, proposed to improve the temperature stability of QD lasers [36, 49] and discussed in Section 4.5.1, will also enhance the power characteristics. Inasmuch as the two reservoirs feeding carriers into the QDs are essentially unipolar, the finite-delay capture process is not accompanied by a buildup of bipolar carrier density and no additional recombination occurs. We can therefore

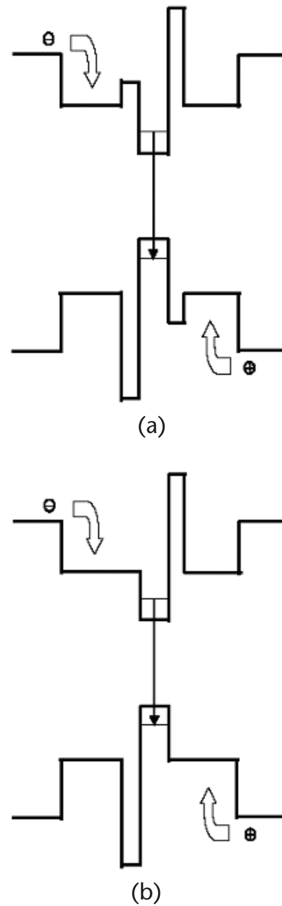


Figure 4.24 (a, b) Bandgap-engineered QD laser structures. The QDs (shown with energy levels) are clad by heterostructure barrier layers that block only the minority carrier transport. (*From:* [49]. © 2003 IEEE Reprinted with permission.)

expect that lasers designed according to [36, 49] will exhibit linear behavior and excellent power performance.

4.6 Other Perspectives

The QD lasers reviewed so far in this chapter are all edge-emitting lasers with Fabry-Perot cavities. Also, they are all based on interband (bipolar) optical transitions. Some feasible alternatives are briefly discussed next.

Vertical cavity surface-emitting lasers (VCSELs) are of high interest for optical signal transmission. The VCSEL's active medium is embedded in a short λ cavity and a single longitudinal mode operation is automatically realized. To keep the loss low, both mirrors should be highly reflective. They are usually realized as stacks of

$\lambda/4$ distributed Bragg reflectors. The vertical cavity represents a one-dimensional microcavity. Using a narrow QD ensemble with the transition energy matched to the cavity passband, optimum photon and electron control can be achieved at a low j_{th}. Due to carrier localization in QDs, also their spreading out of the injection region can be suppressed. This may result in ultralow threshold currents (<70 μA) at ultrasmall apertures (submicrometer). The use of InGaAs QDs in GaAs-based VCSELs emitting at 1.3 μm [80]—the wavelength especially important for single-mode fiber communication—presents one of the most apparent and successful applications of QDs.

A VCSEL based on a single QD is also possible to fabricate. In such a laser, the inhomogeneous broadening, inherent to QD ensembles, will be absent. A single QD VCSEL will be a semiconductor analog of a single atom laser [81] or an ion trap laser [82].

Another promising direction is to use QDs in unipolar (intraband) semiconductor lasers. In these lasers, stimulated emission comes from the transitions between states of carriers in the same band [83, 84]. This concept was realized in the unipolar *quantum cascade laser* (QCL) [85, 86]. Transition energies in unipolar lasers can be tailored to suit important applications at mid- and far-infrared wavelengths.

All contemporary unipolar lasers suffer from a fundamental problem associated with an extremely low efficiency of intraband optical transitions: The optical transition rate in the best QCLs is 3.5 orders of magnitude lower than the nonradiative relaxation rate due to phonon emission. This problem is inherent to transitions between the continuous-spectrum states of overlapping subbands and it cannot be completely eliminated in QW unipolar lasers. As a consequence, CW room temperature operation of such lasers requires very high j_{th}.

The use of QDs offers a radical solution to this problem, based on the discrete carrier spectrum in QDs. The phonon relaxation will be completely suppressed if the energy of optical transitions (both direct transitions between electron levels in a given QD and indirect transitions between electron levels in neighboring QDs) is detuned from the LO phonon energy. The amount of detuning should be greater than the phonon energy dispersion. Because the energy levels in a QD can be varied in a wide range (by controlling the size of QDs and the band offsets between the constituent materials), such control is quite feasible. The nonradiative single-phonon relaxation will thus be practically suppressed. As a rule, the multiple-phonon relaxation rate is much lower than that of single-phonon relaxation and it is usually negligible. The idea for using QDs in unipolar lasers was proposed in [87–93].

Finally, instead of the Fabry-Perot cavity assumed so far, one may be interested in using *distributed feedback* (DFB) [10, 60] for certain applications that need stabilizing in the single-mode generation. Due to a difference in the refraction index of the QD and the surrounding region, a spatially periodic arrangement of QDs [94] will give rise to a sufficient periodic index variation for an index-coupled DFB. At the same time, the periodic arrangement of active QDs provides a spatially periodic variation of the gain. This may provide the benefits of the so-called gain-coupling scheme, without any phase shifter or antireflection coating [95–100].

References

[1] Zory, Jr., P. S., (ed.), *Quantum Well Lasers*, Boston, MA: Academic Press, 1993.

[2] Kapon, E., (ed.), *Semiconductor Lasers*, San Diego, CA: Academic Press, 1999.

[3] Bimberg, D., M. Grundmann, and N. N. Ledentsov, *Quantum Dot Heterostructures*, New York: John Wiley & Sons, 1999.

[4] Dingle, R., and C. H. Henry, *Quantum Effects in Heterostructure Lasers*, U.S. Patent 3982207, September 21, 1976.

[5] van der Ziel, J. P., et al.,"Laser Oscillations from Quantum States in Very Thin GaAs-$Al_{0.2}Ga_{0.8}As$ Multilayer Structures," *Appl. Phys. Lett.*, Vol. 26, No. 8, 1975, pp. 463–465.

[6] Dupuis, R. D., "Room Temperature Operation of Quantum-Well $Ga_{1-x}Al_xAs$-GaAs Laser Diodes Grown by Metalorganic Chemical Vapor Deposition," *Appl. Phys. Lett.*, Vol. 32, No. 5, 1978, pp. 295–297.

[7] Tsang, W. T., "Extremely Low Threshold (AlGa)As Graded-Index Waveguide Separate-Confinement Heterostructure Lasers Grown by Molecular-Beam Epitaxy," *Appl. Phys. Lett.*, Vol. 40, No. 3, 1982, pp. 217–219.

[8] Arakawa, Y., and H. Sakaki, "Multidimensional Quantum Well Laser and Temperature Dependence of Its Threshold Current," *Appl. Phys. Lett.*, Vol. 40, No. 11, June 1982, pp. 939–941.

[9] Eliseev, P. G., *Introduction to the Physics of Injection Lasers*, Moscow, Russia: Nauka, 1983 (in Russian).

[10] Agrawal, G. P., and N. K. Dutta, *Long-Wavelength Semiconductor Lasers*, New York: Van Nostrand Reinhold Company, 1986.

[11] Alferov, Z., "Double Heterostructure Lasers: Early Days and Future Perspectives," *IEEE J. Select. Topics Quantum Electron.*, Vol. 6, No. 6, November/December 2000, pp. 832–840.

[12] Asada, M., Y. Miyamoto, and Y. Suematsu, "Gain and the Threshold of Three-Dimensional Quantum-Box Lasers," *IEEE J. Quantum Electron.*, Vol. 22, No. 9, September 1986, pp. 1915–1921.

[13] Vahala, K. J., "Quantum Box Fabrication Tolerance and Size Limits in Semiconductors and Their Effect on Optical Gain," *IEEE J. Quantum Electron.*, Vol. 24, No. 3, March 1988, pp. 523–530.

[14] Miyamoto, Y., et al., "Threshold Current Density of GaInAsP/InP Quantum-Box Lasers," *IEEE J. Quantum Electron.*, Vol. 25, No. 9, September 1989, pp. 2001–2006.

[15] Alferov, Z. I., and R. F. Kazarinov, "Semiconductor Laser with Electric Pumping," Inventor's Certificate 181737, Application 950840, priority as of March 30, 1963 (in Russian).

[16] Alferov, Z. I., et al., "AlAs-GaAs Heterojunction Injection Lasers with a Low Room-Temperature Threshold," *Sov. Phys. Semicond.*, Vol. 3, No. 9, September 1970, pp. 1107–1110.

[17] Alferov, Z. I., et al., "Investigation of Influence of AlAs-GaAs Heterostructure Parameters on Laser Threshold Current and Realization of Continuous Emission at Room Temperature," *Sov. Phys. Semicond.*, Vol. 4, No. 9, September 1971, pp. 1573–1575.

[18] Hayashi, I., et al., "Junction Lasers Which Operate Continuously at Room Temperature," *Appl. Phys. Lett.*, Vol. 17, No. 3, 1970, pp. 109–111.

[19] Ekimov, A. I. and A. A. Onushchenko, "Quantum Size Effect in 3-Dimensional Microscopic Semiconductor Crystals," *JETP Lett.*, Vol. 34, No. 6, 1981, pp. 345–349.

[20] Ledentsov, N. N., et al., "Optical Properties of Heterostructures with InGaAs-GaAs Quantum Clusters," *Semicond.*, Vol. 28, No. 8, August 1994, pp. 832–834.

[21] Kirstädter, N., et al., "Low Threshold, Large T_0 Injection Laser Emission from (InGa)As Quantum Dots," *Electron Lett.*, Vol. 30, No. 17, August 1994, pp. 1416–1417.

[22] Goldstein, L., et al., "Growth by Molecular Beam Epitaxy and Characterization of InAs/GaAs Strained-Layer Superlattices," *Appl. Phys. Lett.*, Vol. 47, No. 10, November 1985, pp. 1099–1101.

[23] Shchukin, V. A., N. N. Ledentsov, and D. Bimberg, *Epitaxy of Nanostructures*, Berlin, Germany: Springer, 2003.

[24] Maximov, M. V., et al., "Gain and Threshold Characteristics of Long Wavelength Lasers Based on InAs/GaAs Quantum Dots Formed by Activated Alloy Phase Separation," *IEEE J. Quantum Electron.*, Vol. 37, No. 5, May 2001, pp. 676–683.

[25] Park, G., et al., "Low-Threshold Oxide-Confined 1.3-μm Quantum-Dot Laser," *IEEE Phot. Technol. Lett.*, Vol. 13, No. 3, March 2000, pp. 230–232.

[26] Shchekin, O. B., and D. G. Deppe, "1.3 μm InAs Quantum Dot Laser with T_0 = 161K from 0 to 80 μm C," *Appl. Phys. Lett.*, Vol. 80, No. 18, May 2002, pp. 3277–3279.

[27] Kovsh, A. R., et al., "InAs/InGaAs/GaAs Quantum Dot Lasers of 1.3 μm Range with High (88%) Differential Efficiency," *Electron. Lett.*, Vol. 38, No. 19, September 2002, pp. 1104–1106.

[28] Tokranov, V., et al., "Enhanced Thermal Stability of Laser Diodes with Shape-Engineered Quantum Dot Medium," *Appl. Phys. Lett.*, Vol. 83, No. 5, August 2003, pp. 833–835.

[29] Varangis, P. M., et al., "Low-Threshold Quantum Dot Lasers with 201nm Tuning Range," *Electron. Lett.*, Vol. 36, No. 18, August 2000, pp. 1544–1545.

[30] Asryan, L. V., and R. A. Suris, "Inhomogeneous Line Broadening and the Threshold Current Density of a Semiconductor Quantum Dot Laser," *Semicond. Sci. Technol.*, Vol. 11, No. 4, April 1996, pp. 554–567.

[31] Asryan, L. V., and R. A. Suris, "Linewidth Broadening and Threshold Current Density of Quantum-Box Laser," *Proc. of Int. Symp. Nanostructures: Physics and Technology*, St. Petersburg, Russia, June 20–24, 1994, pp. 181–184.

[32] Suris, R. A., and L. V. Asryan, "Quantum-Dot Laser: Gain Spectrum Inhomogeneous Broadening and Threshold Current," *Proc. of SPIE*, Vol. 2399, 1995, pp. 433–444.

[33] Kane, E. O., "Band Structure of Indium Antimonide," *J. Phys. Chem. Solids*, Vol. 1, No. 4, 1957, pp. 249–261.

[34] Evtikhiev, V. P., et al., "Continuous Stimulated Emission at T = 293K from Separate-Confinement Heterostructure Diode Lasers with One Layer of InAs Quantum Dots Grown on Vicinal GaAs(001) Surfaces Misoriented in the [010] Direction in the Active Region," *Semicond.*, Vol. 32, No. 12, December 1998, pp. 1323–1327.

[35] Asryan, L. V., and R. A. Suris, "Temperature Dependence of the Threshold Current Density of a Quantum Dot Laser," *IEEE J. Quantum Electron.*, Vol. 34, No. 5, May 1998, pp. 841–850.

[36] Asryan, L. V., and S. Luryi, "Tunneling-Injection Quantum-Dot Laser: Ultrahigh Temperature Stability," *IEEE J. Quantum Electron.*, Vol. 37, No. 7, July 2001, pp. 905–910.

[37] Asryan, L. V., and R. A. Suris, "Longitudinal Spatial Hole Burning in a Quantum-Dot Laser," *IEEE J. Quantum Electron.*, Vol. 36, No. 10, October 2000, pp. 1151–1160.

[38] Asryan, L. V., S. Luryi, and R. A. Suris, "Intrinsic Nonlinearity of the Light-Current Characteristic of Semiconductor Lasers with a Quantum-Confined Active Region," *Appl. Phys. Lett.*, Vol. 81, No. 12, September 2002, pp. 2154–2156.

[39] Asryan, L. V., S. Luryi, and R. A. Suris, "Internal Efficiency of Semiconductor Lasers with a Quantum-Confined active Region," *IEEE J. Quantum Electron.*, Vol. 39, No. 3, March 2003, pp. 404–418.

[40] Ledentsov, N. N., et al., "Direct Formation of Vertically Coupled Quantum Dots in Stranski-Krastanow Growth," *Phys. Rev. B*, Vol. 54, No. 12, September 1996, pp. 8743–8750.

[41] Pankove, J. I., "Temperature Dependence of Emission Efficiency and Lasing Threshold in Laser Diodes," *IEEE J. Quantum Electron.*, Vol. 4, No. 4, April 1968, pp. 119–122.

[42] Asryan, L. V., and R. A. Suris, "Charge Neutrality Violation in Quantum Dot Lasers," *IEEE J. Select. Topics Quantum Electron.*, Vol. 3, No. 2, April 1997, pp. 148–157.

[43] Tsui, E. S.-M., P. Blood, and A. I. Kucharska, "Charge Neutrality in Quantum-Well Structures," *Semicond. Sci. Technol.*, Vol. 5, No. 4, April 1990, pp. 333–339.

[44] Asryan, L. V., "Threshold Characteristics of InGaAsP/InP Multiple Quantum Well Lasers," *Semicond. Sci. Technol.*, Vol. 15, No. 12, December 2000, pp. 1131–1140.

[45] Stier O., M. Grundmann, and D. Bimberg, "Electronic and Optical Properties of Strained Quantum Dots Modeled by 8-Band $\mathbf{k}\cdot\mathbf{p}$ Theory," *Phys. Rev. B,* Vol. 59, No. 8, February 1999, pp. 5688–5701.

[46] Shoji, H., et al., "Lasing Characteristics of Self-Formed Quantum-Dot Lasers with Multistacked Dot Layer," *IEEE J. Select. Topics Quantum Electron.*, Vol. 3, No. 2, April 1997, pp. 188–195.

[47] Asryan, L. V., et al., "Effect of Excited-State Transitions on the Threshold Characteristics of a Quantum Dot Laser," *IEEE J. Quantum Electron.*, Vol. 37, No. 3, March 2001, pp. 418–425.

[48] Asryan, L. V., et al., "Maximum Modal Gain of a Self-Assembled InAs/GaAs Quantum-Dot Laser," *J. Appl. Phys.*, Vol. 90, No. 3, August 2001, pp. 1666–1668.

[49] Asryan, L. V., and S. Luryi, "Temperature-Insensitive Semiconductor Quantum Dot Laser," *Solid-State Electron.*, Vol. 47, No. 2, February 2003, pp. 205–212.

[50] Statz, H., C. L. Tang, and J. M. Lavine, "Spectral Output of Semiconductor Lasers," *J. Appl. Phys.*, Vol. 35, No. 9, September 1964, pp. 2581–2585.

[51] Alam, M. A., "Effects of Carrier Transport on *L–I* Characteristics of QW Lasers in the Presence of Spatial Hole Burning," *IEEE J. Quantum Electron.*, Vol. 33, No. 6, June 1997, pp. 1018–1024.

[52] Tang, C. L., H. Statz, and G. deMars, "Spectral Output and Spiking Behavior of Solid-State Lasers," *J. Appl. Phys.*, Vol. 34, No. 8, August 1963, pp. 2289–2295.

[53] Suris, R. A., and S. V. Shtofich, "Multifrequency Stimulated Emission from Injection Semi-Conductor Lasers," *Soviet Phys. Semicond.*, Vol. 16, No. 7, July 1982, pp. 851–853.

[54] Suris, R. A., and S. V. Shtofich, "Role of Impurities in the Appearance of Multifrequency Emission from Injection Semiconductor Lasers," *Soviet Phys. Semicond.*, Vol. 17, No. 7, July 1983, pp. 859–861.

[55] Sugawara, M., K. Mukai, and Y. Nakata, "Light Emission Spectra of Columnar-Shaped Self-Assembled InGaAs/GaAs Quantum-Dot Lasers: Effect of Homogeneous Broadening of the Optical Gain on Lasing Characteristics," *Appl. Phys. Lett.*, Vol. 74, No. 11, March 1999, pp. 1561–1563.

[56] Garbuzov, D. Z., et al., "Experimental and Theoretical Investigations of Singularities of the Threshold and Power Characteristics of InGaAsP/InP Separate-Confinement Double-Heterostructure Lasers (λ = 1.3 μm)," *Sov. Phys. Semicond.*, Vol. 25, No. 5, May 1991, pp. 560–564.

[57] Tessler, N., "Nonequilibrium Effects in Quantum Well Lasers," *Appl. Phys. Lett.*, Vol. 61, No. 20, November 1992, pp. 2383–2385.

[58] Hirayama, H., et al., "Estimation of Carrier Capture Time of Quantum-Well Lasers by Spontaneous Emission Spectra," *Appl. Phys. Lett.*, Vol. 61, No. 20, November 1992, pp. 2398–2400.

[59] Hirayama, H., et al., "Carrier Capture Time and Its Effect on the Efficiency of Quantum-Well Lasers," *IEEE J. Quantum Electron.*, Vol. 30, No. 1, January 1994, pp. 54–62.

[60] Coldren L. A., and S. W. Corzine, *Diode Lasers and Photonic Integrated Circuits*, New York: John Wiley & Sons, 1995.

[61] Smowton, P. M., and P. Blood, "The Differential Efficiency of Quantum-Well Lasers," *IEEE J. Select. Topics Quantum Electron.*, Vol. 3, No. 2, April 1997, pp. 491–498.

[62] Kim, J. S., et al., "Structural and Optical Properties of Shape-Engineered InAs Quantum Dots," *J. Appl. Phys.*, Vol. 94, No. 4, August 2003, pp. 2486–2490.

[63] Asryan, L. V., and S. Luryi, "Two Lasing Thresholds in Semiconductor Lasers with a Quantum-Confined Active Region," *Appl. Phys. Lett.*, Vol. 83, No. 26, December 2003, pp. 5368–5370.

[64] Seki, S., et al., "Dominant Mechanisms for the Temperature Sensitivity of 1.3 μm InP-Based Strained-Layer Multiple-Quantum-Well Lasers," *Appl. Phys. Lett.*, Vol. 67, No. 8, August 1995, pp. 1054–1056.

[65] Seki, S., et al., "Dominant Mechanism for Limiting the Maximum Operating Temperature of InP-Based Multiple-Quantum-Well Lasers," *J. Appl. Phys.*, Vol. 79, No. 5, March 1996 pp. 2192–2197.

[66] Asryan, L. V., and R. A. Suris, "Carrier Photoexcitation from Levels in Quantum Dots to States of the Continuum in Lasing," *Semicond.*, Vol. 35, No. 3, March 2001, pp. 343–346.

[67] Seki, S., and K. Yokoyama, "Electrostatic Deformation in Band Profiles of InP-Based Strained-Layer Quantum-Well Lasers," *J. Appl. Phys.*, Vol. 77, No. 10, May 1995, pp. 5180–5184.

[68] Zhukov, A. E., et al., "Loss Multiplication in a Quantum Dot Laser," *Laser Phys.*, Vol. 13, No. 3, March 2003, pp. 319–323.

[69] Kovsh, A. R., et al., "InAs/InGaAs/GaAs Quantum Dot Lasers of 1.3 μm Range with Enhanced Optical Gain," *J. Cryst. Growth*, Vol. 251, Nos. 1–4, April 2003, pp. 729–736.

[70] Gorfinkel, V. B., S. Luryi, and B. Gelmont, "Theory of Gain Spectra for Quantum Cascade Lasers and Temperature Dependence of their Characteristics at Low and Moderate Carrier Concentrations," *IEEE J. Quantum Electron.*, Vol. 32, No. 11, November 1996, pp. 1995–2003.

[71] Gorfinkel, V., M. Kisin, and S. Luryi, "Hot Electrons and Curves of Constant Gain in Long Wavelength Quantum Well Lasers," *Optics Express*, Vol. 2, No. 4, February 1998, pp. 125–130.

[72] Asryan, L. V., et al., "High-Power and High-Temperature Operation of InGaAsP/InP Multiple Quantum Well Lasers," *Semicond. Sci. Technol.,* Vol. 14, No. 12, December 1999, pp. 1069–1075.

[73] Walter, G., et al., "Room-Temperature Continuous Photopumped Laser Operation of Coupled InP Quantum Dot and InGaP Quantum Well InP-InGaP-In(AlGa)P-InAlP Heterostructures," *Appl. Phys. Lett.*, Vol. 79, No. 13, September 2001, pp. 1956–1958.

[74] Walter, G., et al., "Coupled InP Quantum-Dot InGaP Quantum Well InP-InGaP-In(AlGa)P-InAlP Heterostructure Diode Laser Operation," *Appl. Phys. Lett.*, Vol. 79, No. 20, November 2001, pp. 3215–3217.

[75] Chung, T., G. Walter, and N. Holonyak, Jr., "Coupled Strained-Layer InGaAs Quantum-Well Improvement of an InAs Quantum Dot AlGaAs-GaAs-InGaAs-InAs Heterostructure Laser," *Appl. Phys. Lett.*, Vol. 79, No. 27, December 2001, pp. 4500–4502.

[76] Walter, G., T. Chung, and N. Holonyak, Jr., "High-Gain Coupled InGaAs Quantum Well InAs Quantum Dot AlGaAs-GaAs-InGaAs-InAs Heterostructure Diode Laser Operation," *Appl. Phys. Lett.*, Vol. 80, No. 7, February 2002, pp. 1126–1128.

[77] Chuang, S. L., and N. Holonyak, Jr., "Efficient Quantum Well to Quantum Dot Tunneling: Analytical Solutions," *Appl. Phys. Lett.*, Vol. 80, No. 7, February 2002, pp. 1270–1272.

[78] Bhattacharya, P., and S. Ghosh, "Tunnel Injection $In_{0.4}Ga_{0.6}As$/GaAs Quantum Dot Lasers with 15GHz Modulation Bandwidth and T_0 = 237 K at Room Temperature," *Appl. Phys. Lett.*, Vol. 80, No. 19, May 2002, pp. 3482–3484.

[79] Pradhan, S., S. Ghosh, and P. Bhattacharya, "Temperature Dependent Steady-State Characteristics of High-Performance Tunnel Injection Quantum Dot Lasers," *Electron. Lett.*, Vol. 38, No. 23, November 2002, pp. 1449–1450.

[80] Lott, J. A., et al., "InAs-InGaAs Quantum Dot VCSELs on GaAs Substrates Emitting at 1.3 μm," *Electron. Lett.*, Vol. 36, No. 16, August 2000, pp. 1384–1385.

[81] An, K., et al., "Microlaser—A Laser with One-Atom in an Optical-Resonator," *Phys. Rev. Lett.*, Vol. 73, No. 25, December 1994, pp. 3375–3378.

[82] Meyer, G. M., H.-J. Briegel, and H. Walther, "Ion-Trap Laser," *Europhys. Lett.*, Vol. 37, No. 5, February 1997, pp. 317–322.

[83] Kazarinov, R. F., and R. A. Suris, "Possibility of Amplification of Electromagnetic Waves in a Semiconductor with a Superlattice," *Sov. Phys. Semicond.*, Vol. 5, No. 4, 1971, p. 707.

[84] Kazarinov, R. F., and R. A. Suris, "Electric and Electromagnetic Properties of Semiconductors with a Superlattice," *Sov. Phys. Semicond.*, Vol. 6, No. 1, 1972, pp. 120–131.

[85] Faist, J. F., et al., "Quantum Cascade Laser," *Science*, Vol. 264, No. 5158, April 1994, pp. 553–556.

[86] Faist, J., et al., "Quantum Cascade Laser: An Intersub-Band Semiconductor-Laser Operating Above Liquid-Nitrogen Temperature," *Electron. Lett.*, Vol. 30, No. 11, pp. 865–866, May 1994.

[87] Suris, R. A., "Prospects for Quantum Dot Structures Applications in Electronics and Optoelectronics," *Proc. NATO Advanced Research Workshop, Future Trends in Microelectronics: Reflections on the Road to Nanotechnology*, S. Luryi, et al., (eds.), France, July 17–21, 1995, pp. 197–208. (*Also published in NATO ASI Series, Series E: Applied Sciences*, Vol. 323, Dordrecht: Kluwer Academic Publishers, 1996.)

[88] Wingreen, N. S., and C. A. Stafford, "Quantum-Dot Cascade Laser: Proposal for an Ultralow-Threshold Semiconductor Laser," *IEEE J. Quantum Electron.*, Vol. 33, No. 7, July 1997, pp. 1170–1173.

[89] Wingreen, N. S., and C. A. Stafford, "Quantum-Dot Cascade Laser," U.S. Patent 5692003, November 25, 1997.

[90] Wingreen, N. S., "Quantum-Dot Cascade Laser," U.S. Patent 5963571, October 5, 1999.

[91] Singh, J., "Possibility of Room Temperature Intra-Band Lasing in Quantum Dot Structures Placed in High-Photon Density Cavities," *IEEE Photon. Technol. Lett.*, Vol. 8, No. 4, April 1996, pp. 488–490.

[92] Botez, D., P. S. Zory, and C.-F. Hsu, "Intersubband Quantum Box Semiconductor Laser," U.S. Patent 5953356, September 14, 1999.

[93] Hsu, C.-F., et al., "Intersubband Quantum-Box Semiconductor Lasers," *IEEE J. Select. Topics Quantum Electron.*, Vol. 6, No. 3, May/June 2000, pp. 491-503.

[94] Kuramochi, E., J. Temmyo, and T. Tamamura, "Perfect Spatial Ordering of Self-Organized InGaAs/AlGaAs Box-Like Structure Array on GaAs (311) B Substrate with Silicon Nitride Dot Array," *Appl. Phys. Lett.*, Vol. 71, No. 12, September 1997, pp. 1655–1657.

[95] Griesinger, U. A., et al., "Realization of Dot DFB Lasers," *IEEE Photon. Technol. Lett.*, Vol. 8, No. 5, May 1996, pp. 587–589.

[96] Griesinger, U. A., et al., "Fabrication and Investigation of Nanostructures and Their Application in New Laser Devices," *J. Vac. Sci. Technol. B,* Vol. 14, No. 6, November–December 1996, pp. 4058–4061.

[97] Wang, J., et al., "Emission Dynamics of Dot and Wire DFB Lasers," *IEEE Photon. Technol. Lett.*, Vol. 8, No. 12, December 1996, pp. 1585–1587.

[98] Wang, J., U. A. Griesinger, and H. Schweizer, "Intrinsic Carrier Capture Time in Deep-Etched Quantum-Box (70 nm Diameter) Lasers at Low Temperature: An Indication of Extremely High Quantum Capture Efficiency," *Appl. Phys. Lett.*, Vol. 70, No. 9, March 1997, pp. 1152–1154.

[99] Wang, J., et al., "Carrier Capture Times in Quantum-Well, -Wire, and -Box Distributed-Feedback Lasers Characterized by Dynamic Lasing Emission Measurements," *IEEE J. Select. Topics Quantum Electron.*, Vol. 3, No. 2, April 1997, pp. 223–229.

[100] Susa, N., "Feasibility Study on the Application of the Quantum Disk to the Gain-Coupled Distributed Feedback Laser," *IEEE J. Quantum Electron.*, Vol. 34, No. 8, August 1998, pp. 1317–1324.

CHAPTER 5

High-Speed Quantum Dot Lasers

P. Bhattacharya and T. Norris, University of Michigan
S. Ghosh, University of Illinois at Chicago

5.1 Introduction

Compound semiconductor light-emitting diodes and lasers have become the mainstay of present-day fiber optic communication systems. Semiconductor lasers have become essential to current *local area networks* (LANs) and long-haul intercontinental fiber optic networks [1]. Low threshold current, high output power and efficiency, temperature-independent operation, large modulation bandwidth, low chirp, and wide spectral tunability are some of the properties desirable in a high-performance semiconductor lasers. The use of gain media with reduced dimensionality, in particular QWs, has greatly enhanced laser performance, particularly due to the modification of the band density of states and the resulting gain and differential gain [2, 3].

Many of the properties mentioned are degraded due to hot carrier effects, slow carrier relaxation rates, and carrier leakage over heterobarriers in the active region of the laser [4–6]. It has been known for a while that incorporation of QDs in the active region of optoelectronic devices would drastically improve device performance, primarily due to the enhanced carrier density of states [7]. A large density of states is highly desirable for optoelectronic devices such as lasers and detectors. As one goes from a bulk material, in which there is no quantum confinement, to a system in which the carriers are confined in one dimension (QWs), two dimensions (QWRs), or in all three dimensions (QDs), the density of states function increases in magnitude and becomes discrete.

The fabrication of device-quality QDs has been a major challenge. Even when QDs were formed by using direct patterning or otherwise [8, 9], they lacked not only the large areal density that is required for device applications, but also the high optical quality that is essential for optoelectronic devices such as lasers, modulators, and detectors. Fortunately, it was observed that self-organized growth of strained (or mismatched) semiconductors leads to coherently strained and defect-free 3D islands, which have dimensions that lead to 3D confinement of carriers [10–13]. Following this development, various electronic and optoelectronic devices such as lasers [14–17], FIR detectors [18–20], single electron transistors [21, 22], and

electro-optic modulators [23, 24] with QDs in the active regions have been realized and reported.

Some of the major advantages of lasers with QDs in the active region are (1) significant decrease in the threshold current, (2) a large increase in the differential gain (prospect for high-speed operation), (3) a vastly reduced chirp, and (4) temperature-insensitive operation. Researchers have been extremely successful in demonstrating many of the predicted characteristics of QD lasers [25–32]. However, these devices have eluded the realization of a large modulation bandwidth at room temperature. Needless to say, it is essential for directly modulated QD lasers to have high-speed capability if they are to compete with existing QW laser technology. Theoretical calculations done by Vurgaftman et al. [33] have shown that the capture time of carriers in QD lasers is expected to be much longer than in QWs because of inefficient phonon scattering processes in QDs. The large density of states of the wetting and barrier layers of the self-organized dots creates a hot carrier distribution at higher temperatures. These factors tend to limit the modulation bandwidth of QD lasers.

Bandwidths of 5 to 7 GHz at room temperature [34] and >20 GHz at cryogenic temperatures [27] have been demonstrated in directly modulated *separate confinement heterostructure* (SCH) QD lasers. High-frequency electrical impedance measurements made on QD lasers show that the capture time, which represents the time it takes the carrier, once injected from the contact, to reach the lasing state, is both longer than in QW lasers and has significant temperature dependence. Typical carrier capture times in QD lasers at room temperature are ~50 to 60 ps, which is a factor of 10 higher than in QW lasers [35]. From the measured magnitude and temperature dependence of the carrier capture time, the likely mechanism for hot carrier relaxation in QDs is identified to be electron-hole scattering.

This chapter describes in detail the unique carrier dynamics observed in self-organized QDs and its effect on high-frequency performance of QD lasers. A novel injection technique, whereby electrons are injected into the QD ground state by tunneling, is described. The enhanced performance of these tunneling-injection QD lasers is also described and discussed.

5.2 MBE Growth of Self-Organized QDs and Their Electronic Properties

5.2.1 Self-Organized Growth of In(Ga)As QDs

Formation of self-organized QDs is initiated with lattice mismatched or strained heteroepitaxial crystal growth, where the deposited material has a lattice mismatch ~2% with respect to the substrate [10]. Epitaxial growth of QDs is accomplished by MBE or *organometallic vapor phase epitaxy* (OMVPE). The formation of self-organized QDs occurs by the Stranski-Krastanov growth mode [36], wherein growth first proceeds in the 2D layer-by-layer mode, during which the “wetting layer” is formed, followed by the 3D growth mode in which coherently strained

islands, or QDs, are formed. In(Ga)As/GaAs self-organized QDs are formed for In compositions greater than 30%. The In(Ga)As/GaAs QDs are grown between 470° and 530°C, depending on the In composition in the dot. Indium adatoms have high surface migration and therefore QDs having high In content are usually grown at lower temperatures. The wetting layer thickness varies from ~1.8 ML in InAs/GaAs QDs to ~5.5 ML in $In_{0.4}Ga_{0.6}As$/GaAs QDs, and these thicknesses are independent of growth temperature (>400°C). The size and shape of the dots are strongly influenced by the growth conditions and monolayer coverage. Data obtained from AFM, as shown in Figure 5.1(a), indicate that typical areal dot densities are ~5×10^{10} cm^{-2} for In(Ga)As/GaAs QDs. The dot density can also be varied by varying the growth temperature. In(Ga)As/GaAs QDs grown at low temperatures have a higher dot density due to reduced surface migration of the In adatom.

High-resolution cross-sectional transmission electron microscopy (HR-XTEM) images, shown in Figure 5.1(b), indicate that the defect-free $In_xGa_{1-x}As$ 3D islands on GaAs are pyramidal to lens shaped. The facet planes of the pyramids are strongly influenced by the growth conditions (temperature, As over pressure, and so on). The typical size of the 3D islands is ~15 to 20 nm at the base and ~7 to 9 nm in height. The self-organization process inherently induces a size fluctuation of about 10% to 15%, which results in an inhomogeneously broadened photoluminescence linewidth of about 40 to 60 meV.

5.2.2 Electronic Spectra of In(Ga)As/GaAs QDs

To exploit the full potential of self-organized QDs in electronic and optoelectronic devices, the electronic spectra need to be accurately known. The energy levels in the dots have been calculated by several groups of authors and the calculations are summarized in the following.

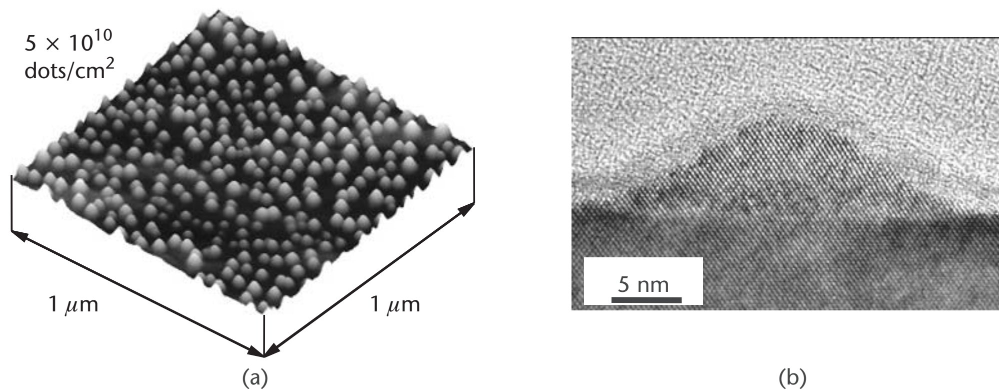


Figure 5.1 (a) AFM image of 2.4-ML InAs QDs on GaAs; (b) HR-XTEM image of a InAs/GaAs QD.

Strain Distribution: VFF Model

The strain distribution in the self-organized dots has been calculated in detail by Jiang and Singh [37] using the *valence force field* (VFF) model of Keating [38]. The calculation shows that there is a strain relaxation at the top of the dot, which has been borne out experimentally. The biaxial strain changes sign from the bottom to the top of the dot, which means that there is more confinement for heavy holes at the bottom and more confinement for light holes at the top of the dot. The shear strain is also very large. Moreover, strong intermixing occurs between the light holes and the heavy holes for the valence band bound states.

Band Structure: 8-Band $\mathbf{k} \cdot \mathbf{p}$ Formalism

The band structure in semiconductor quantum confined structures is usually solved by decoupling the conduction and valence bands. The conduction band states are determined from a scalar effective mass approach, whereas the valence band states are calculated with a $\mathbf{k} \cdot \mathbf{p}$ formulation, which includes the heavy hole and light hole coupling. The same is true for quantum wells and wires. However, this is not true for self-organized QDs, which are typically highly strained. A simple decoupled conduction–valence band picture and effective mass description may not be adequate, which is evident from the following observations: (1) The bandgap of bulk InAs is ~0.36 eV, whereas the ground state transitions in InAs/GaAs dots are ~1.1 eV; (2) large spatial variations exist in the strain; and (3) the band splittings due to large strain components are comparable to the interband separations in bulk material. Therefore, the most accurate method to determine the electronic levels in the self-organized QDs is to use an 8-band $\mathbf{k} \cdot \mathbf{p}$ formulation for both bands.

In the presence of strain, the Hamiltonian has the form:

$$H_t = H_0 + H_{\text{str}} \tag{5.1}$$

where H_0 is the kinetic energy term and H_{str} is the strain term. The spin-orbit dependent deformation potential is neglected in this approach. Using a finite-difference method the equation is solved numerically, with the distance between two grid points chosen to be equal to the lattice constant of GaAs (a = 5.6533Å). For pyramidal QDs, with a base width of 124Å and a height of 64Å, the calculated electronic spectra are shown in Figure 5.2. These dot dimensions are very close to the ones measured using X-TEM.

There are a number of discrete excited states in the conduction band. An interesting feature of the calculated spectra is the near fourfold degeneracy of the second excited states. This is due to the symmetry of the dot geometry. The first excited electronic state is 62 meV higher than the ground state. Considerable mixture of wetting layer states occurs in this state because it is less confined. It is evident that the separation between the bound electron states is larger than the optical phonon energy in the dot material (~36 meV). Therefore, optical phonon scattering, the primary mechanism for carrier relaxation in semiconductors, is suppressed in the QDs. This constitutes the *phonon bottleneck*, which has recently been observed by us

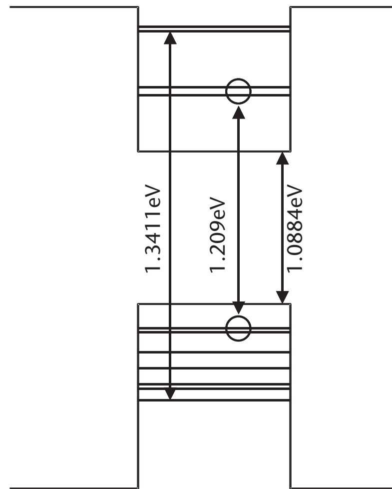


Figure 5.2 Conduction and valence band energy levels for an InAs/GaAs QD with base width of 124Å and height of 62Å calculated using the 8-band **k** · **p**. (*From:* [37]. © 1996 American Institute of Physics. Reprinted with permission.)

experimentally. On the other hand, the separation of the bound hole states is less than the phonon energy and, in addition, there is considerable band mixing. Therefore, holes thermalize very efficiently in the dots. The ground and excited states of the dots are observed experimentally in photoluminescence, photocapacitance, and transmission spectra.

5.3 Separate Confinement Heterostructure QD Lasers and Their Limitations

Junction lasers with QDs in the active region were demonstrated nearly a decade ago. Since the first demonstration, great strides have been made in enhancing the performance characteristics of these unique devices. Significant milestones in the development of the QD lasers include demonstration of low threshold at room temperature [25], large differential gain [26, 27], low chirp [28], high output power [29], wide spectral tunability [30], and better temperature insensitivity of the threshold current than QW lasers [31, 32]. However, it is now clear that the carrier dynamics in QDs do not favor high-speed operation, such as direct high-speed modulation, of lasers. Detailed two- and three-pulse pump-probe differential transmission spectroscopy measurements made on QD heterostructures show that at temperatures greater than 100K, a large number of injected electrons preferentially occupy the higher lying states of the dots and states in the adjoining barrier/wetting layer with higher density [39]. This is corroborated by recent calculations [40] and measurements [41] made by other groups. Therefore, there exists a significant hot carrier problem in QD lasers. We attribute this to be the principal reason for the inability to modulate QD lasers at high speeds. As the temperature is lowered, the hot carrier distribution is minimized and carriers relax from the higher lying states

by efficient electron-hole scattering. Indeed, small-signal modulation bandwidths of ~30 GHz have been measured in SCH QD lasers at 80K [27], while in the same devices, the highest measured $f_{-3dB} \cong 7.5$ GHz at 300K [34].

5.3.1 Carrier Relaxation and Phonon Bottleneck in Self-Organized QDs

To understand the limitations to high-speed operation of QD lasers, it is necessary to study the carrier dynamics in the dots as a function of temperature. This has been done using high-frequency impedance measurements and differential transmission spectroscopy.

High-Frequency Electrical Impedance Measurements

Measurement of the optical modulation response gives the maximum modulation bandwidth, and analysis of the response can determine laser dynamics. However, information about carrier dynamics is also contained in the small-signal electrical impedance. In particular, the measured impedance is a sensitive function of the capture time. The high-frequency impedance measurements described here were performed on single-mode In(Ga)As/GaAs SCH QD lasers [35].

The transport and capture time τ_0 in high-frequency electrical impedance measurements above threshold is given by:

$$\tau_0 = \tau_{\text{cap}} + \xi R_d C_s c \tag{5.2}$$

where ξ is 1 above threshold, R_d is the dc small-signal impedance, C_s is parasitic capacitance, τ_C is the quantum capture time, and c is the velocity of light. Figure 5.3

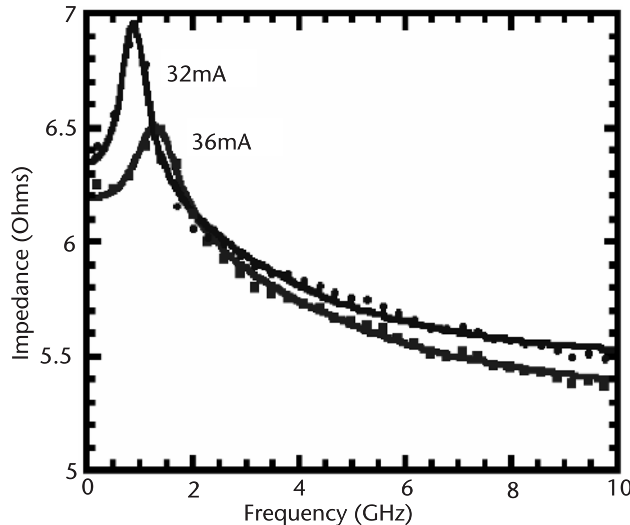


Figure 5.3 Measured and fit set of impedance data at a bias current of 32 and 36 mA for a single-layer QD laser with a threshold current of 28 mA. (*From:* [35]. © 1997 IEEE. Reprinted with permission.)

shows a typical measured and fit set of impedance data at a bias current of 32 and 36 mA for a single-layer QD laser with a threshold current of 28 mA. Using appropriate carrier and photon rate equations, the capture times were extracted. Capture times ranging between 30 and 100 ps were obtained in interband QD lasers depending on injection current and heterostructure used. Similar measurements in QW lasers yields capture times of 2 to 5 ps in InGaAs/GaAs QW lasers. The long capture time in QD lasers is believed to be due to the inefficient phonon scattering process, and this is believed to be responsible for low modulation bandwidths of single-mode QD lasers at room temperature. The impedance measurements do not give us any idea about the different components involved in carrier capture from the barrier to the lasing state. To measure the different relaxation time components directly, the technique of *differential transmission spectroscopy* (DTS) is used.

Femtosecond Differential Transmission Spectroscopy

Two- and three-pulse femtosecond pump-probe DTS measurements have been made by us on $In_{0.4}Ga_{0.6}As$/GaAs QD heterostructures, similar to SCH lasers, at different temperatures and for a range of excitation levels. The principle of the DT measurement is outlined in Figure 5.4. An amplified Ti:Sapphire laser is used to generate a white-light source from which 10-nm-wide pulses are spectrally selected for the gain (800-nm), pump (980-nm), and probe pulses. The pump is tuned to generate carriers either in the GaAs barrier states or resonantly in the excited states of the dots. The DT signal measures the change in the carrier occupation of the levels that are in resonance with the probe spectrum. When the pump and probe pulses are delayed with respect to each other, the transient dot level population is resolved directly. The temperature of the sample is stabilized in a He flow cryostat with a feedback-heater controller. The gain pulse establishes an interband population inversion and sets the quasi-Fermi level in the active region. The pump pulse depletes the QD ground state population. The carrier dynamics and the population

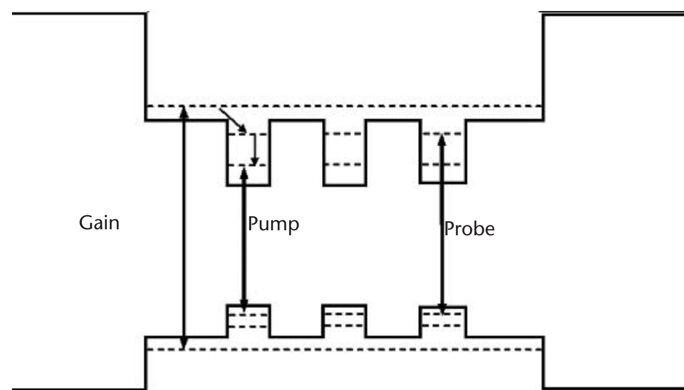


Figure 5.4 Schematic showing the principle of DTS measurement.

of the various levels are then determined from the transmitted intensity of the probe pulse.

The results of these experiments have been reported in detail by us previously and some relevant data are described and discussed here. The differential transmission signal is proportional to the carrier population of the level probed. The carrier dynamics in the QDs is therefore reflected in the transient of the differential transmission signal for the ground and excited states. Figure 5.5 shows the excited state ($n = 2$) differential transmission signal for weak excitation (<1 carrier per dot) in the GaAs barrier at $T = 10$K [42]. The population of the excited state shows a fast decay, followed by a very slow one. We have recorded similar data at temperatures up to 300K. The fast excited state to ground state relaxation is ~5 to 6 ps and this component decreases and becomes slower as the sample temperature is raised. The slow component is due to nongeminate capture of electrons and holes among the dots and is observed when the number of injected carriers is smaller than the number of accessible dots.

We have confirmed that the fast intersubband relaxation at low temperatures is mediated by electron-hole scattering. As the temperature is raised, the rate of this process decreases and the intersubband relaxation rate also decreases. In practice, lasers are operated at room temperature and high injection levels. In the self-organized QDs, the localized QD states and the 2D wetting layer states form an electronically coupled system and the number of wetting layer/barrier states is much larger than the number of available dot states at 300K. Under these conditions, the injected electrons predominantly reside in the wetting layer and barrier states and the system cannot be described by equilibrium quasi-Fermi statistics. Evidence of injected electrons residing in higher lying states of the dot heterostructure are shown in Figure 5.6. Figure 5.6(a) is the three-pulse differential transmission signal corresponding to gain recovery in a laser heterostructure [43]. The gain pulse to establish interband population inversion is followed 14 ps later by a pump pulse to deplete the electron-hole pairs in the ground state. The initial fast recovery is due to the excited to ground state transition and the slow component is due to carrier capture from

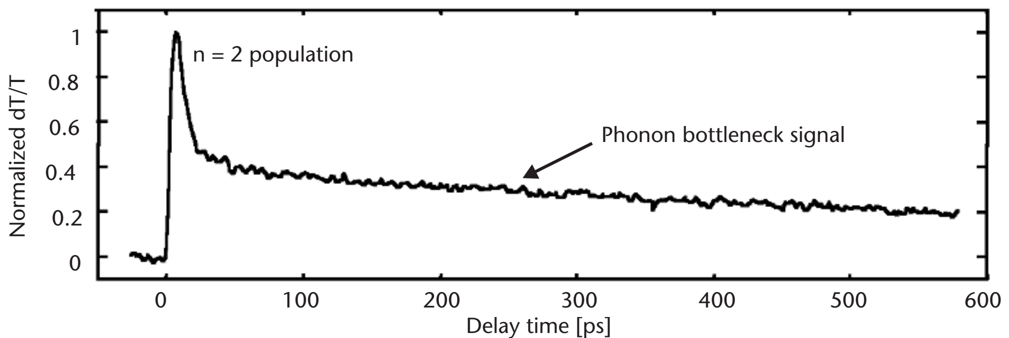


Figure 5.5 Excited state ($n = 2$) DTS time scan in an $In_{0.4}Ga_{0.6}As/GaAs$ QD at 10K. There is a fast component (5 to 6 ps) followed by a slow one (~100 ps). The slow component is due to phonon bottleneck. (*From:* [42]. © 2001 American Institute of Physics. Reprinted with permission.)

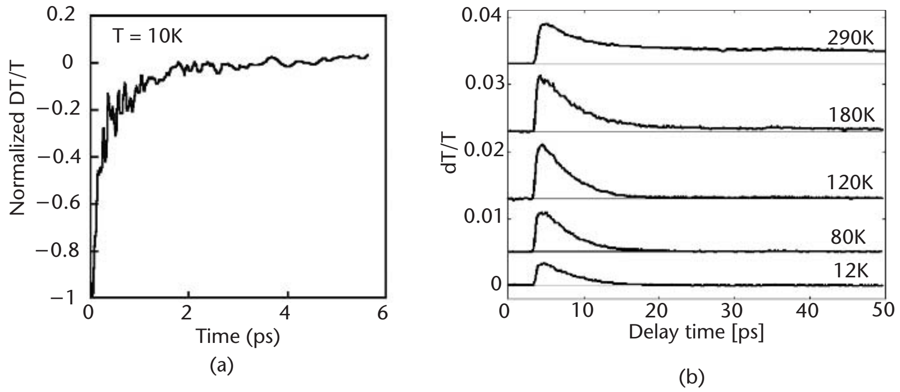


Figure 5.6 (a) Three-pulse DTS time scan corresponding to the gain recovery in the ground state of a $In_{0.4}Ga_{0.6}As/GaAs$ QD. (*From:* [43]. © 2002 American Institute of Physics. Reprinted with permission.) (b) DTS time scan of the wetting layer and barrier region (±20 meV of GaAs band edge) at various temperatures. (*From:* [39]. © 2002 American Institute of Physics. Reprinted with permission.)

higher lying states. Figure 5.6(b) shows the differential transmission signal at GaAs barrier energy ±20 meV with the pump tuned to the barrier energy [39]. It is evident that as the temperature is increased, most of the carriers remain in the barrier and the dot excited states. Therefore, a reduction in the rate of electron-hole scattering in the dots (and phonon scattering is not operative) and the hot carrier distribution lead to severe gain saturation at the QD ground state lasing energy. This has also been recently verified experimentally by Mathews et al. [41]. The tunneling-injection scheme, which is described in the next section, should significantly reduce gain compression and carrier leakage to the cladding layers.

5.3.2 Hot Carrier Effects in SCH QD Lasers

The performance of SCH QD lasers has improved significantly since the first demonstration of room temperature QD lasers. Extremely low threshold currents of ~19 A/cm^2, one-half of the best reported QW lasers, have been demonstrated. High differential gain and low chirp have also been demonstrated in these devices. However, researchers have not been able to realize a temperature-insensitive QD laser with large modulation bandwidth. The primary reason for this is the presence of significant number of hot carriers in the active region of SCH QD lasers at room temperature.

The carrier injection process in a conventional SCH QD laser is depicted in Figure 5.7(a). Electrons and holes are injected from the cladding layer regions into the active (guiding) region containing the QD layers. For efficient laser performance, the carriers should be able to relax into the QDs and reach thermal equilibrium with the quasi-Fermi level. If thermalization into the dot lasing states does not occur in any reasonable amount of time (approximately stimulated emission time constant), the carrier distribution will become hot [see Figure 5.7(b)]. While injected

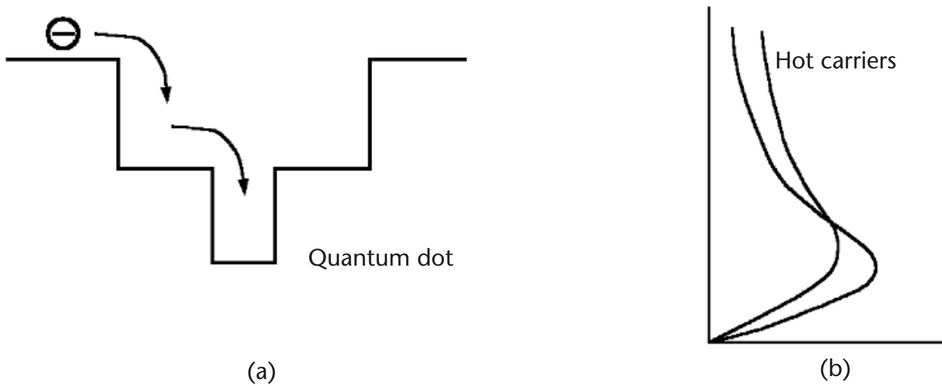


Figure 5.7 Carrier injection into a QD gain region in (a) a conventional SCH laser and (b) carrier distribution.

carriers lose energy and fill the lasing states, carrier heating simultaneously forces them out toward higher energies and causes leakage to adjoining layers. Recombination in the regions adjoining the dots is believed to be the principal source of the temperature dependence of the threshold current of lasers and small values of the coefficient T_0.

For ideal QD lasers, we expect a very large T_0 indicating low temperature sensitivity of the threshold current. However, due to recombination of hot carriers in the optical confinement layer and the barriers, the threshold current increases. Another source of temperature dependence of the threshold current is the dot size inhomogeneity due to the self-organization process. Therefore, there are a sizable number of dots that do not contribute to lasing, while adding to parasitic recombination and therefore to the threshold current. However, this component of the threshold current is not very temperature dependent and hence does not lower the coefficient T_0. The hot carrier distribution also leads to gain compression at the lasing energy and, as a result, a reduced small-signal modulation bandwidth. The hot carriers in and around the active region also increase the diffusion capacitance and result in low frequency roll-off.

5.4 Tunnel Injection of Carriers in QDs

The tunneling-injection scheme for electrons in a laser is illustrated in Figure 5.8(a). If "cold" electrons are introduced into the QD ground states by phonon-assisted tunneling, and the tunneling rate is comparable to the stimulated emission rate, the carrier distribution will be maintained close to a quasi-Fermi distribution, even at high injection levels [see Figure 5.8(b)]. Thus carrier leakage and hot carrier effects are minimized. It is important to note that the hole relaxation rates in QDs and other quantum confined structures are very fast due to multiplicity of the levels, band mixing, and efficient hole-phonon coupling. In the tunneling-injection scheme,

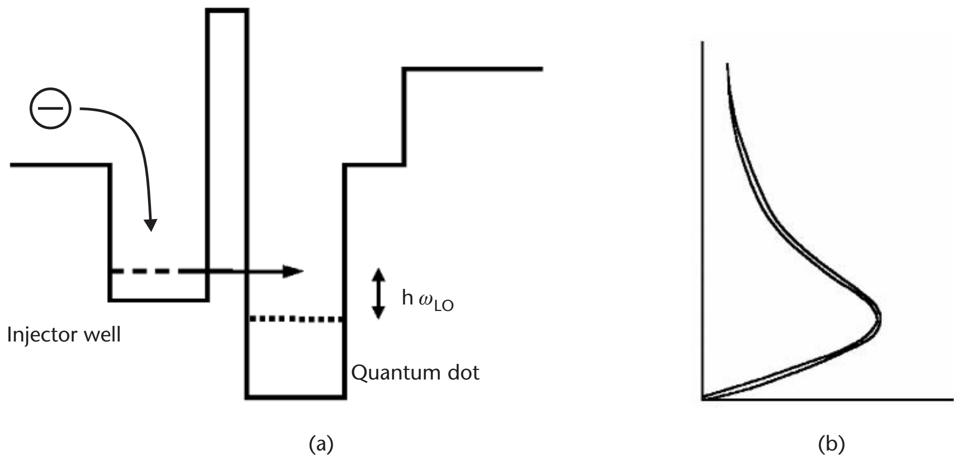


Figure 5.8 Carrier injection into a QD gain region in (a) a tunneling-injection heterostructure laser and (b) carrier distribution.

fewer electrons will bypass the active region during the injection process. Gain compression effects will be minimized, since most of the electrons are injected close to the lasing energy.

5.4.1 Tunneling-Injection Laser Heterostructure Design and MBE Growth

The laser heterostructure, designed with knowledge of the QD electronic states and the conduction band profile in the vicinity of the QDs, is shown in Figure 5.9. The active region consists of the $In_{0.25}Ga_{0.75}As$ injector well, a 20Å $Al_{0.55}Ga_{0.45}As$ tunnel barrier, and three coupled $In_{0.4}Ga_{0.6}As$ QD layers. The 95Å $In_{0.25}Ga_{0.75}As$ injector well is grown at 490°C, the QD layers are grown at 525°C, and the rest of the structure is grown at 620°C.

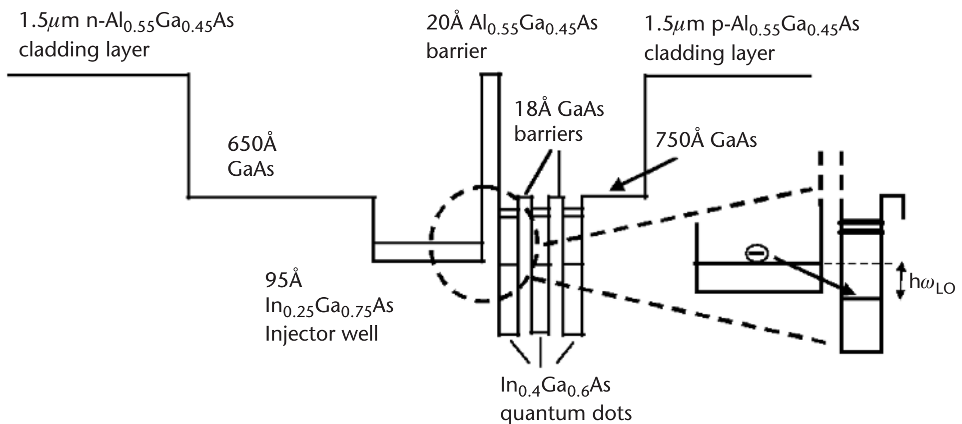


Figure 5.9 Conduction band profile, of a tunneling-injection QD laser heterostructure grown by MBE, under flat-band conditions.

A critical aspect of the design is the alignment of the conduction band states of the injector layer with the bound states in the QDs, such that efficient phonon-assisted tunneling can take place from the injector layer to the dots. The bandgap of the injector layer is tuned for the optimum tunneling conditions and the tunneling rates are measured, as described in the next section. The wavelength of the luminescence peak for the dots is controlled by adjusting the InGaAs dot charge during epitaxy such that the energy separation, in the conduction band, between the injector well states and the QD ground states is ~36 meV at room temperature. This energy separation ensures LO phonon-assisted tunneling from the injector well to the dot ground states through the 20Å undoped $Al_{0.55}Ga_{0.45}As$ barrier layer. The photoluminescence spectra from the sample, measured at 12K with 800-nm excitation, is shown in Figure 5.10. A dominant emission from the quantum dot with a peak at 980 nm and a weak shoulder, believed to be from the injector layer, at 950 nm are seen in the data. A characteristic feature of the photoluminescence from tunneling-injection QD heterostructures is the extremely small linewidth (~23 meV) compared to normal QDs grown under identical conditions, for which the linewidth varies in the range from 40 to 60 meV. This is due to a selection process of the tunneling states in the dots, thereby artificially inducing dot uniformity for luminescence. The reduced linewidth of the QDs in tunneling-injection heterostructures is advantageous in high-speed operation of lasers.

5.4.2 Measurement of Phonon-Assisted Tunneling Times

Three-pulse pump-probe DTS is used to measure the tunneling times into the ground state of QDs in the tunneling-injection heterostructure. Measured DTS data at 12K

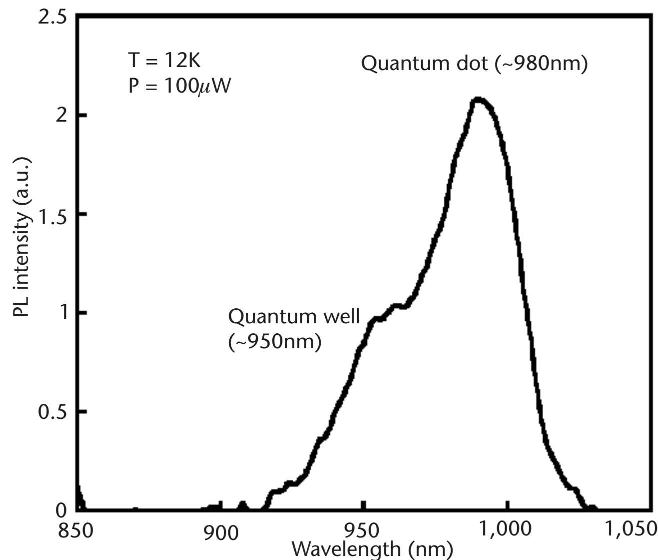


Figure 5.10 Low-temperature photoluminescence spectrum of the tunneling-injection laser heterostructure.

are shown in Figure 5.11(a) [44]. The signal at 950 nm corresponds to the injector layer population. The rapid decrease in the DTS signal corresponds to phonon-assisted tunneling of electrons from this layer to the QD ground state, and the slow recovery reflects the repopulation of the layer from higher energy levels by phonon scattering. The initial decrease in the 980-nm DTS signal is due to pump-induced reduction in the QD ground state population via stimulated emission. The fast recovery in the 980-nm DTS signal corresponds again to the tunneling of electrons into the dot ground state. Figure 5.11(b) shows the DTS signal in a two-pulse experiment (no gain pulse; pump tuned to 950 nm to inject electrons directly into the QW injector layer) for various temperatures; the rise times are almost identical. The temperature-dependent tunneling time measurements do not show any variation in the tunneling time with temperature. This is contrary to what we would expect in ideal phonon-assisted tunneling. The phonon-assisted tunneling times are dependent on phonon number, which is temperature dependent.

The PL spectrum from the laser heterostructure shows that there is overlap between the ground state of the injector layer and the ground state of the vertically coupled QD layers, which have level broadening due to size inhomogeneity and vertical coupling. Therefore, in addition to phonon-assisted tunneling, resonant tunneling processes are also present. The presence of resonant tunneling could explain the nearly temperature-independent tunneling times. As mentioned earlier, the phonon-mediated hole relaxation rate is extremely fast and we have previously calculated and measured the relaxation time to be ~ 0.6 ps. We estimate the electron tunneling time constant to be ~1.7 ps. Therefore, the phonon-assisted tunneling rate will be comparable to the stimulated emission rate for a high injection rate and the carrier distribution can still be maintained close to equilibrium.

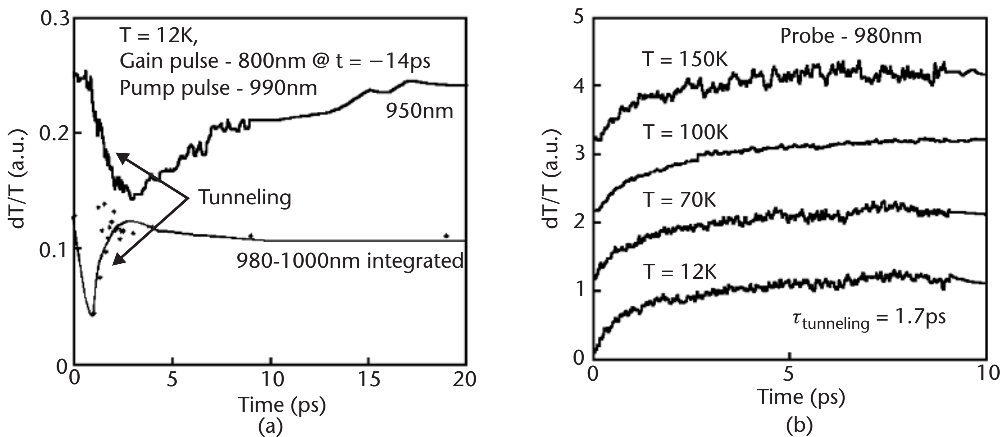


Figure 5.11 (a) Three-pulse DTS signal from the tunneling-injection heterostructure with gain pulse at 800 nm, pump pulse at 980 nm, and probe pulses at 950 and 980–1,000 nm (integrated); (b) temperature-dependent two-pulse DTS signal with the gain pulse at 800 nm and probe pulse at 980 nm.

5.5 Characteristics of High-Speed Tunneling-Injection QD Lasers

Tunneling-injection QD laser heterostructures were grown by MBE. The device heterostructure structure consists of 1.5-μm-thick $Al_{0.55}Ga_{0.45}As$ outer cladding layers (n,p = 5 × 10^{17} cm^{-3}) and GaAs contact layers on either side of the 0.15-μm-thick undoped GaAs *optical confinement layer* (OCL). The active region is identical to the one discussed in Section 5.4.1. Ridge waveguide lasers with 3-μm ridges for both dc and high-frequency measurements were fabricated using standard photolithography, dry and wet etching, and metallization techniques.

5.5.1 Room Temperature DC Characteristics

The room temperature dc characteristics of the ridge waveguide tunneling-injection QD lasers are measured with 1% duty cycle pulsed biasing (100-μs period). Typical light-current characteristics for a 400-μm-long device are shown in Figure 5.12(a) [45]. The spectral outputs were recorded [see Figure 5.12(b)] at various injection currents and these confirm lasing from ground state up to injection currents ~5 kA/cm^2. Threshold currents as low as 8 mA are measured for 400-μm-long Fabry-Perot cavities. Threshold current densities are ~500 A/cm^2 for long cavities (1,300 μm). Devices with 200-μm-long cavities exhibit extremely high slope efficiencies (~0.9 W/A) and differential quantum efficiencies η_d ~ 0.7–0.8. The internal quantum efficiency (η_i) is extracted from a plot of $1/\eta_d$ versus cavity 1, shown in Figure 5.13. Typical values of η_I are 0.8-0.9. The cavity loss in these high-performance devices is ~ 8 cm^{-1}.

5.5.2 Temperature-Dependent DC Characteristics

Figure 5.14 shows the measured slope efficiency as a function of temperature in a tunneling-injection QD laser. For comparison, the temperature-dependent slope

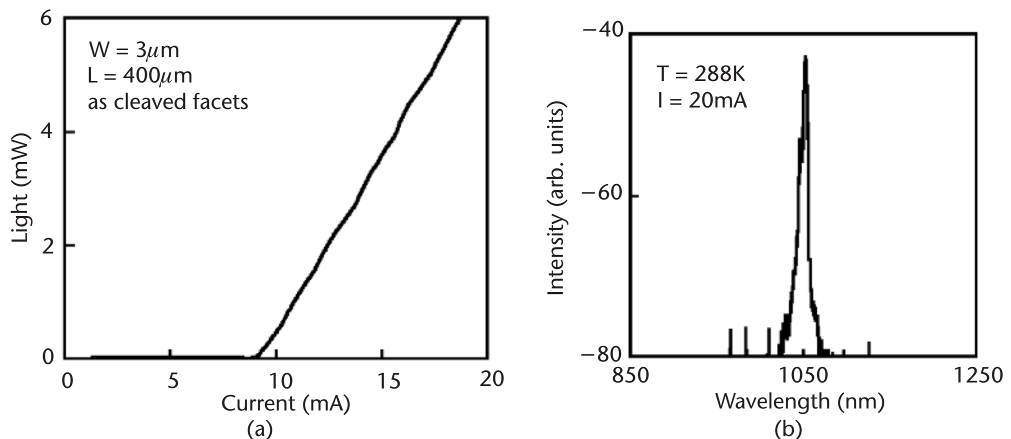


Figure 5.12 (a) Measured light-current characteristics of a 400-μm-long single-mode tunneling-injection QD laser at room temperature; (b) single-mode spectral output at ~1,300 A/cm^2.

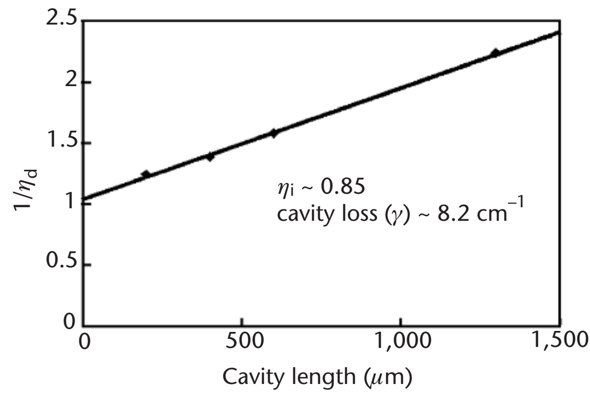


Figure 5.13 Plot of inverse differential quantum efficiency versus cavity length.

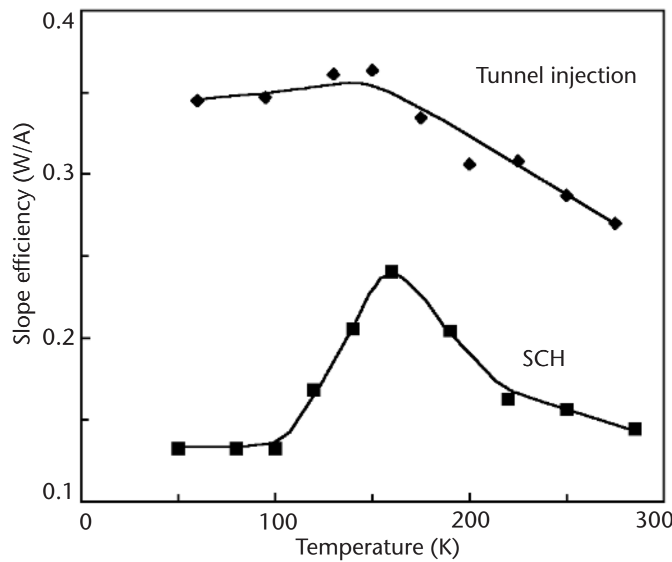


Figure 5.14 Variation of slope efficiency with temperature in a single-mode tunneling-injection QD laser and a broad area SCH QD laser.

efficiency in a conventional SCH quantum dot laser is also shown. The temperature variation of the slope efficiency of the tunneling-injection laser is distinctly different. As stated before, the tunneling process chooses the dots that contribute to lasing, almost independent of temperature. Therefore, the slope efficiency of the output remains fairly constant. A very small decrease in efficiency is noted due to a small amount of carrier leakage from the dots. On the other hand, in the SCH lasers at low temperatures, injected carriers are frozen in the dots in a nonequilibrium distribution.

Carriers that are initially captured into a dot in which the confined state energy does not overlap with the lasing energy will not contribute to lasing. Hence, they recombine nonradiatively and the output efficiency remains low. An increase in the ambient temperature equilibrates the carrier distribution among the dots, resulting in an increasing number of dots contributing to the lasing process. The slope efficiency therefore steadily increases with temperature. Beyond 170K, the slope efficiency decreases again due to carrier leakage from the dots into the cladding layers and their eventual nonradiative recombination. The marked difference in the temperature-dependent slope efficiency of the two types of lasers confirms the reduction in leakage of hot carriers with tunnel injection.

The coefficient T_0 is a measure of the temperature dependence of the threshold current of a laser. The temperature-dependent light-current characteristics of a 400-μm-long device [46], measured under the pulsed conditions described earlier, are depicted in Figure 5.15(a). The measured threshold currents are plotted against the ambient temperature in Figure 5.15(b). From the two distinct slopes indicated by the continuous lines, values of T_0 = 363K for 5°C < T < 60°C and T_0 = 202K for 60°C < T < 100°C are derived. These are the highest values of T_0 measured in these temperature ranges in QD lasers. The high device efficiencies and the high values of T_0 indicate minimization of carrier leakage from the gain region and parasitic recombination in optical confinement layers [47].

5.5.3 High-Speed Modulation Characteristics

The dc characteristics measured in the tunneling-injection QD lasers indicate prospects for enhanced modulation bandwidths and improved dynamic characteristics. Modulation measurements were made on ridge waveguide lasers with a *ground-*

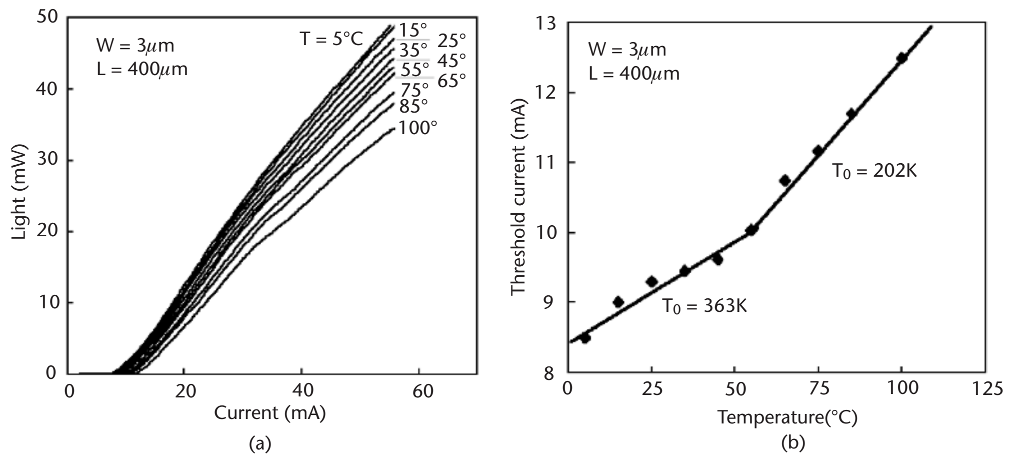


Figure 5.15 (a) Temperature-dependent light-current characteristics tunneling-injection QD laser. (b) Variation of I_{th} with temperature. The lines are fit to the data in accordance with $I_{th}(T) = I_{th}(0)e^{T/T_0}$.

signal-ground (GSG) configuration. Both small- and large-signal modulation measurements were done as a function of temperature. The small-signal modulation properties of the tunneling-injection lasers are described first, followed by a measurement of Auger coefficients from large-signal modulation experiments.

Small-Signal Modulation Characteristics

The small-signal modulation measurements made on the tunneling-injection QD lasers show for the first time modulation bandwidths comparable to those measured in QW lasers. Figure 5.16(a) shows the measured modulation response under CW bias for a 400-μm-long tunneling-injection QD laser cavity [I_{th} (CW) = 12 mA] for various injection currents at 288K [45]. The continuous lines are best fit curves to the measured data. The spectral outputs at these injection currents confirm that lasing from the ground state is maintained. A bandwidth of f_{-3dB} = 22 GHz is measured for I ~ 42 mA and this is the highest bandwidth measured in any QD laser at room temperature. The modulation response shows typical gain compression limited behavior at higher injection currents. The modulation characteristics of the tunneling-injection lasers have also been analyzed by taking into account the detailed carrier dynamics and the appropriate carrier and photon rate equations. The time constants required for the analysis have all been directly measured by differential transmission spectroscopy. The calculated modulation response for different injection currents are shown in Figure 5.16(b) and agree well with the measured response.

The modulation efficiency and differential gain of a laser can be determined from a plot of the resonance frequency f_r of the modulation response as a function of the square root of the output power. The resonance frequency is expressed in terms of output power by the following relation:

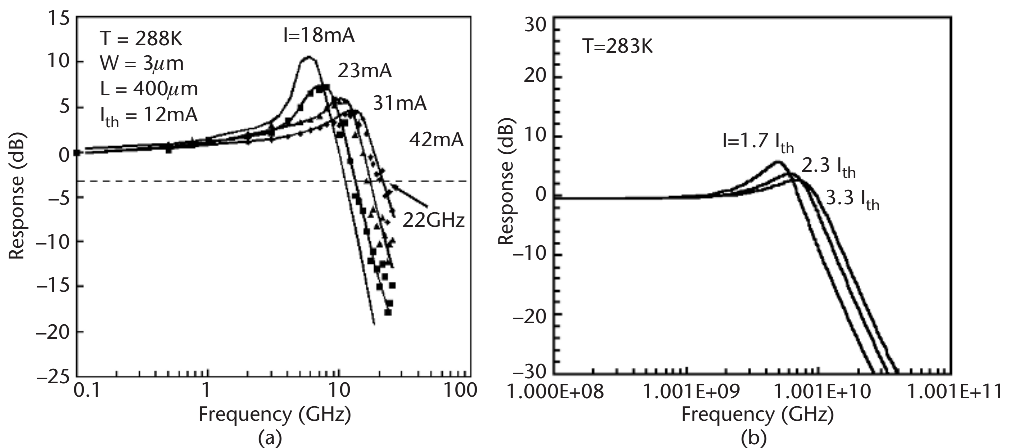


Figure 5.16 (a) Small-signal modulation response of 400-μm-long tunneling-injection laser under varying injection currents at room temperature. (b) Calculated modulation response obtained from the coupled rate equations.

$$f_r = \frac{1}{2\pi}\sqrt{\frac{v_g\,(dg/dn)\left(P/\eta_{\text{slope}}\right)\eta_i\Gamma}{qV_{\text{act}}}} \tag{5.3}$$

where η_{slope} is the slope efficiency of the *L-I* curve, Γ is the confinement factor, *dg/dn* is the differential gain, and V_{act} is the active region volume. Figure 5.17 shows the plot of the resonance frequency f_r versus square root of the output power. The modulation efficiency, which is the slope of this plot in Figure 5.17, is ~2.62 GHz/mW$^{1/2}$. From this value of the modulation efficiency and using a fill factor of ~28%, internal quantum efficiency $\eta_i \sim 0.85$ (for the measured device) and a confinement factor $\Gamma \sim 2.5 \times 10^{-3}$, a value of $dg/dn \cong 8.85 \times 10^{-14}$ cm^2 for the differential gain is derived at room temperature.

The damping factor, γ, is an important parameter that determines the high-frequency performance of a semiconductor laser under small signal modulation. The modulation response of a laser in terms of the resonance frequency is given by

$$r(\omega) = \frac{\omega_r^4}{\left(\omega_r^2 - \omega^2\right) + \gamma^2\omega^2} \tag{5.4}$$

where $\gamma = \omega_r^2\tau_p$ and τ_p is photon lifetime in the cavity. The damping factor is also expressed as $\gamma = Kf_r^2$, where K is an empirical factor related to the modulation response of a laser. This is valid only at low-injection levels. At high injection levels, the hot carrier density in the laser can become significant, leading to gain compression. Under these conditions,

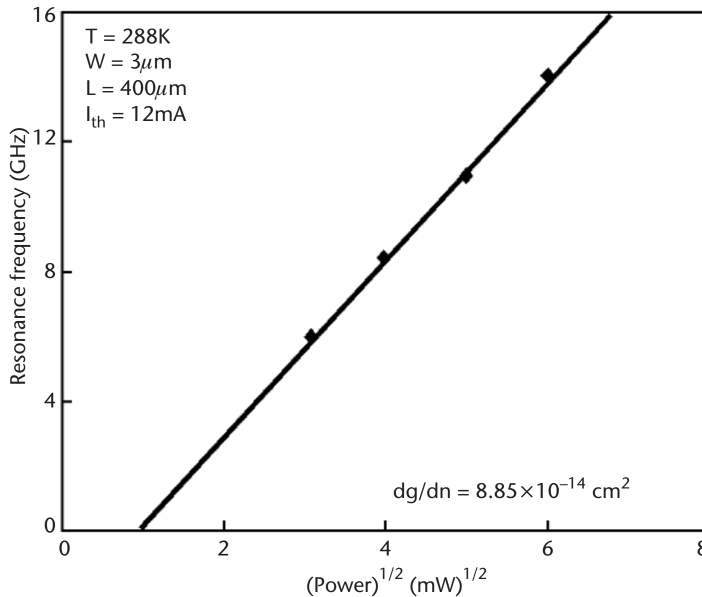


Figure 5.17 The plot of resonance frequency f_r of the modulation response versus square root of power.

$$K = 4\pi^2\left(\tau_p + \frac{\varepsilon}{v_g\left(dg/dn\right)}\right) \tag{5.5}$$

where ε is the gain compression factor, v_g is the photon group velocity, and dg/dn is the differential gain. A value of $K = 0.171$ ns was calculated using the damping factor obtained from the best fit curves to the measured modulation response of a 400-μm laser cavity. Correspondingly, the gain compression factor $\varepsilon = 7.2 \times 10^{-16}$ cm^3, which is almost 2 orders of magnitude lower than those measured for SCH QD lasers. This again confirms that the tunneling-injection QD lasers have reduced hot carrier effects compared to SCH QD lasers. A maximum intrinsic modulation bandwidth, f_{-3dB}(int.) = 9/K(ns) = 55 GHz is also derived. It is important to realize that the laser performance depends more on the energy distribution of injected carriers, rather than the injected carrier density.

Linewidth Enhancement Factor and Chirp

Lateral mode instabilities become important for designing high-power semiconductor lasers. At high levels of carrier injection, filamentation occurs due to an increase in the refractive index and self-focusing on the optical mode in the gain medium. As a result, the laser output does not exhibit a Gaussian single-mode near-field pattern, and increase of the stripe width to enhance output power does not follow simple scaling rules. The degree of filamentation is strongly dependent on the linewidth enhancement factor α given by

$$\alpha = -\frac{4\pi}{\lambda}\frac{dn/dN}{dg/dN} \tag{5.6}$$

where N is the carrier density, n is the refractive index, and g is the gain. Below threshold the linewidth enhancement factor can also be written as:

$$\alpha = \frac{2}{\delta\lambda}\frac{d\lambda_i}{d\left\{\ln\left[\left(\sqrt{r_i}-1\right)/\left(\sqrt{r_i}+1\right)\right]\right\}} \tag{5.7}$$

where $\delta\lambda$ is the mode spacing in the subthreshold spectrum and r_i is the peak-to-valley ratio in the subthreshold spectrum, and λ_i is the wavelength of the ith peak in the spectrum. The linewidth enhancement factor as a function of frequency is obtained by measuring the subthreshold emission spectra of the single-mode tunneling-injection ridge waveguide lasers and by calculating the Fabry-Perot mode peak-to-valley ratios and shifts between two differential bias currents. Figure 5.18 shows the wavelength (frequency) dependence of the linewidth enhancement factor for a 400-μm-long tunneling-injection QD laser cavity [45]. We measure values of $\alpha \leq 1$ which are a factor of 5 to 6 lower than InGaAs MQW lasers with similar threshold and power levels. The extremely low linewidth enhancement results from the higher differential gain and a smaller carrier-induced modulation of the refractive index in the active volume. These values agree with the values reported for QD

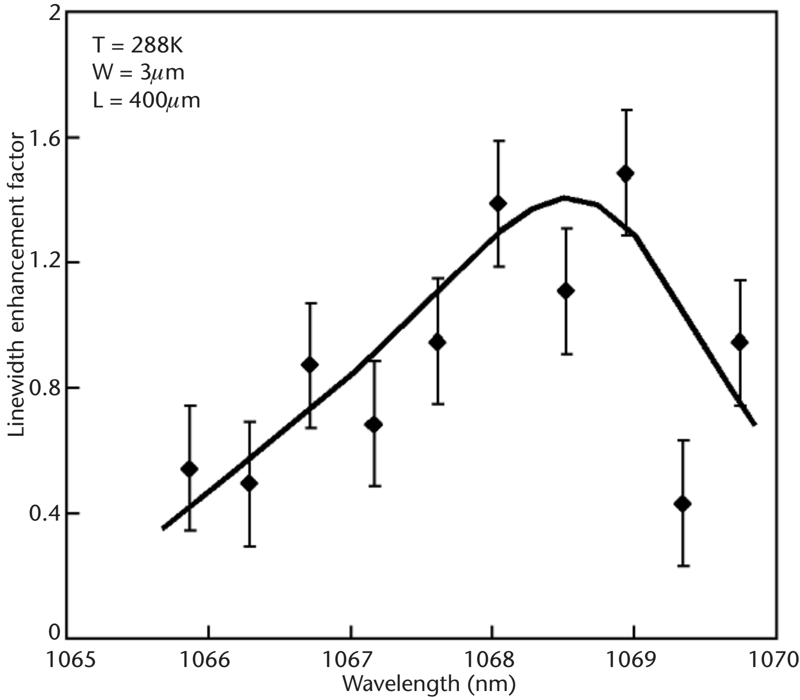


Figure 5.18 Measured linewidth enhancement factor versus wavelength.

lasers by other groups [48] and are a factor of 2 to 5 lower than those reported for QW lasers [49].

Under modulation, the lasing mode shifts in wavelength and an envelope, as shown in Figure 5.19, is observed. The FWHM of this spectral output is the dynamic linewidth. It is evident that for optical communication, the dynamic linewidth or chirp has to be small in order to minimize channel spacing. Chirp in a semiconductor laser is directly proportional to the linewidth enhancement factor α. The small measured values of α promise low chirp in tunneling-injection QD lasers. Chirp is measured in the lasers during direct small-signal modulation by measuring the broadening of a single longitudinal mode using an optical spectrum analyzer. A sinusoidal modulation current is superimposed on a pulsed dc bias current. The envelope of the dynamic shift in the wavelength is recorded and the difference between the half-width of the observed envelope with and without modulation is used to evaluate the chirp. Figure 5.20(a) shows the dependence of chirp on the modulation current [45]. The measured chirp for a InGaAs QW laser with similar I_{th} is also shown for comparison. The chirp as a function of modulation frequency, under fixed modulation current, is shown in Figure 5.20(b). Reduced amplitude and frequency dependence of chirp in tunneling-injection QD lasers confirm the reduced hot carrier effects in the active volume.

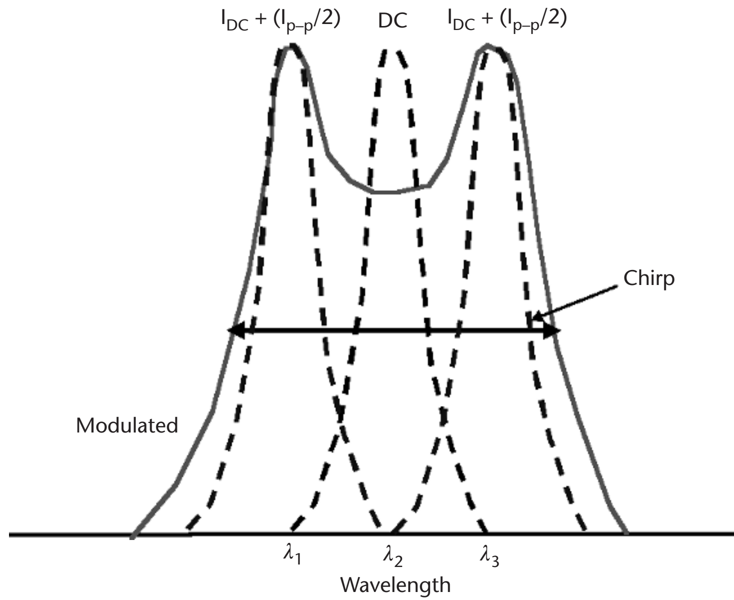


Figure 5.19 Illustration of chirp in semiconductor lasers.

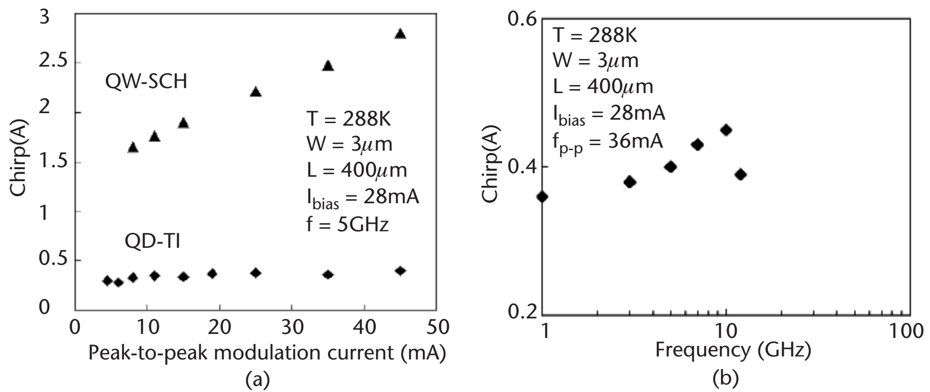


Figure 5.20 (a) Plots of measured chirp in the tunneling-injection QD laser and a SCH QW laser as a function of modulation current. (b) Measured chirp as a function of modulation frequency in a tunneling-injection QD laser.

Auger Recombination in High-Speed QD Lasers

Auger recombination in small bandgap materials affects the performance of lasers adversely, by increasing the threshold current and damping factor, and reducing the modulation bandwidth. The damping factor, taking into account Auger recombination, is given by

$$\gamma_{\text{Auger}} = \gamma(J_a = 0) + \frac{1}{q}\frac{dJ}{dn} \tag{5.8}$$

To improve the high-speed performance of lasers, it is important to reduce Auger recombination. The Auger recombination rates can be determined by measuring the turn-on delay, τ_d, of a laser under large signal modulation, which is given by

$$\tau_d = qV\int_0^{n_{\text{th}}} \frac{1}{I - qVR(n)}\,dn \tag{5.9}$$

where $R(n)$ is the total recombination rate and is expressed as

$$R(n) = \frac{n}{\tau} = A_{\text{nr}}n + R_{\text{sp}}n^2 + C_a n^3 \tag{5.10}$$

in terms of carrier lifetime τ, the Shockley-Read-Hall coefficient A_{nr}, the radiative recombination coefficient R_{sp}, the carrier density n, and the Auger coefficient C_a. The term n_{th} represents the threshold carrier density and V is the active volume of the laser. By measuring the stimulated emission delay times under large-signal modulation and calculation of the radiative recombination rates and the threshold carrier density, the Auger recombination rates can be determined.

Large-signal modulation measurements are made on the 400-μm-long single-mode ridge waveguide tunneling-injection lasers at different temperatures. The lasers are pulse biased with a 10% duty cycle from $I = 0$ to I ($>I_{\text{th}}$) with a 100-ps (20–80%) rise time electrical pulses. The output is detected with an InGaAs photoreceiver. The delay time between the electrical and optical signal is measured with a high-speed digital sampling oscilloscope, taking into account the delays due to optical fiber, RF cables, and the photoreceiver. Figure 5.21 shows the experimentally determined turn-on delay times as a function of bias current density at different ambient temperatures for SCH QD lasers [50] and tunneling-injection QD lasers [51]. The increase in the bias current density decreases the recombination lifetime and consequently the delay time.

The gain and spontaneous recombination rates are calculated using the Fermi golden rule with 8-band $\mathbf{k}\cdot\mathbf{p}$ description of the bands and n_{th} is determined using the single-mode coupled rate equations with a measured value of cavity loss. These parameters are used to determine the Auger coefficients. The carrier distribution in the wetting layers, barrier layers, and the injector well are taken into account in order to represent the carrier dynamics in the tunneling-injection heterostructure more accurately. Recombination at traps is neglected due to the observations of small trap density and negligible Stokes shift in the QD heterostructures. To match the experimental and theoretical threshold current densities at each temperature, the inhomogeneous broadening in the QDs is varied. The Auger coefficient is obtained by minimizing the RMS error between the theoretical and experimental delay times. The temperature dependence of Auger coefficients in SCH and tunneling-injection QD lasers is shown in Figure 5.22 [51].

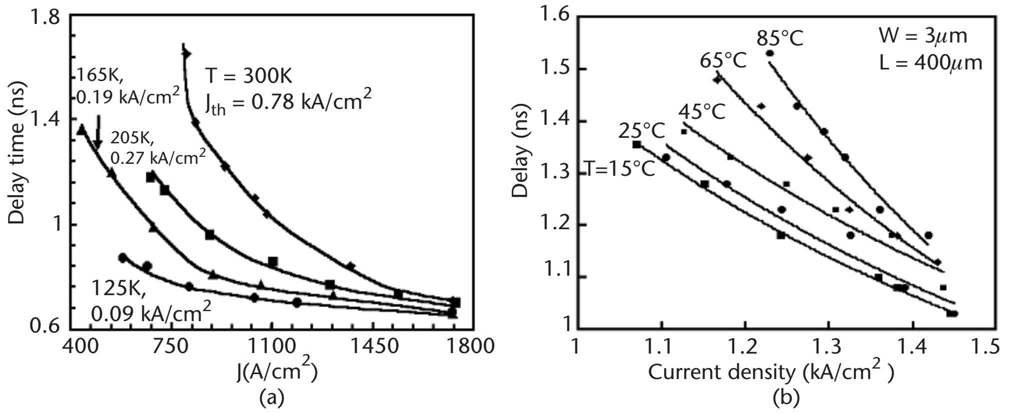


Figure 5.21 Measured turn-on delay times as a function of injection current density in (a) a SCH QD laser and in (b) a tunneling-injection QD laser at various ambient temperatures.

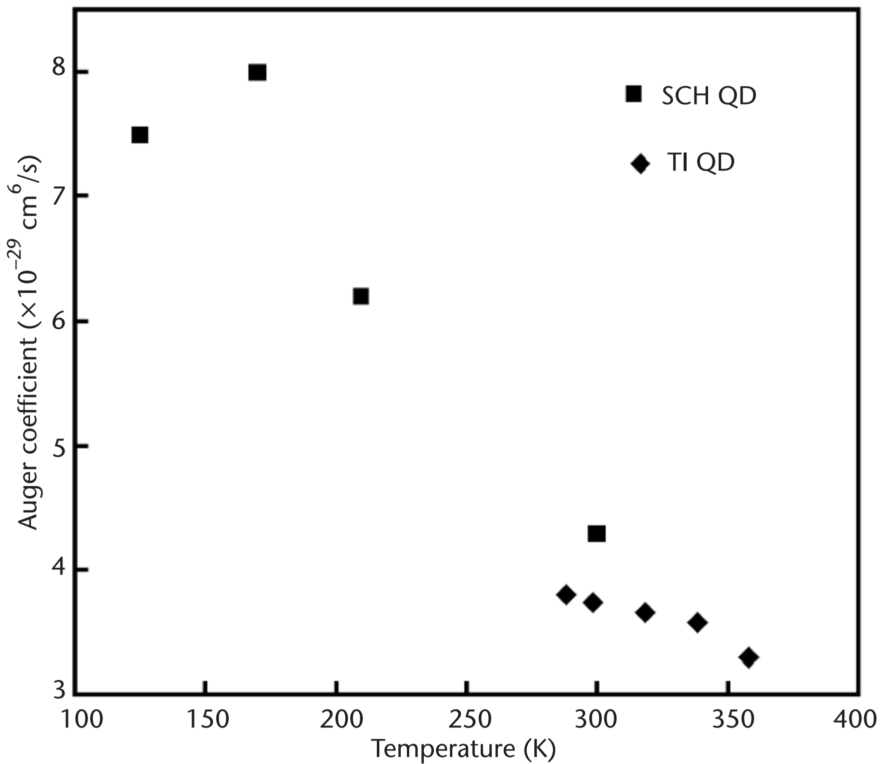


Figure 5.22 Variation of Auger recombination coefficient with temperature. For comparison, Auger coefficients for a SCH QD laser are also shown.

As described earlier, electron-hole scattering is the dominant mechanism by which electrons relax from the excited state to the ground state in QDs. In this process, an excited state electron scatters with a ground state hole, thereby transferring its energy to the hole and relaxing to the ground state. The excited state hole relaxes back to the ground state by rapid phonon scattering. Energy conservation in the process is achieved by self-consistent broadening of the hole levels due to hole-phonon scattering as shown in Figure 5.23. The electron in the ground state can then contribute to Auger scattering. As the temperature increases, the population of

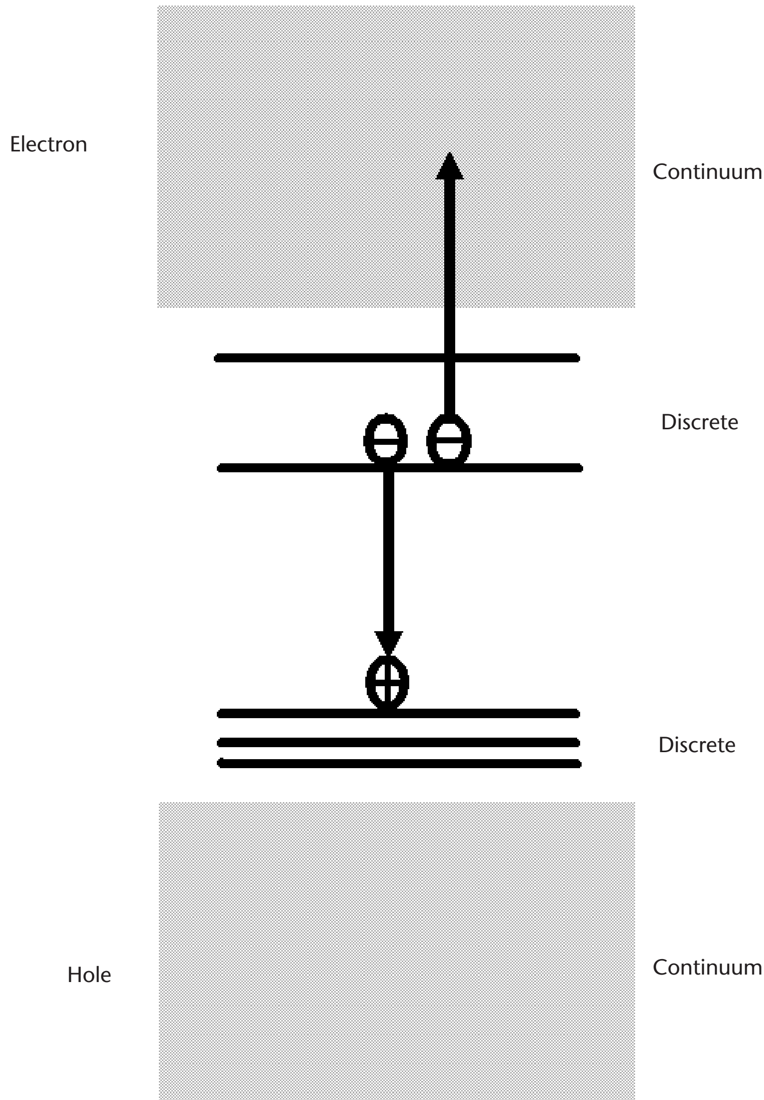


Figure 5.23 A schematic of the Auger transition in a QD system. Hole occupation of the ground state is critical for the process to occur.

ground state holes decreases, due to their excitation to higher energy levels, and the electron-hole scattering and consequently Auger recombination decrease. This trend is observed in the SCH QD lasers. In the tunneling-injection structure, electrons are supplied to the dot ground states directly by tunneling from the adjoining injector layer, and the temperature dependence of the Auger recombination rate would be reduced. This is, indeed, observed in the data shown in Figure 5.22. The values of the Auger rates are also lower, reflecting a smaller population of electrons in the dot excited states.

5.6 Conclusion

In this chapter we describe the performance characteristics of state-of-the-art high-speed QD lasers. It is evident that the devices' performance can be significantly improved by using phonon-assisted tunneling to inject carriers into the lasing state of the QDs. We have demonstrated tunneling-injection QD lasers with bandwidths exceeding 22 GHz at room temperature. These lasers also exhibit low-threshold currents, very low temperature sensitivity, high differential gains, low gain compression factors, low linewidth enhancement factors, low chirp, and low Auger recombination. The tunneling-injection QD lasers have, for the first time, demonstrated properties comparable to or better than those in high-performance QW lasers.

Acknowledgments

The authors acknowledge contributions made by Professor J. Singh, Dr. K. Kamath, Professor J. Phillips, Professor D. Klotzkin, S. Pradhan, Dr. J. Urayama, Z.K. Wu, and K. Kim. The work was supported by the Army Research Office and the Air Force Office of Scientific Research.

References

[1] Fairley, P., "The Microphotonics Revolution," *Technology Review,* July/August 2000.

[2] Arakawa, Y., and H. Sakaki, "Multidimensional Quantum Well Lasers and Temperature Dependence of Its Threshold Current," *Appl. Phys. Lett.*, Vol. 40, 1982, p. 939.

[3] Arakawa, Y., and A. Yariv, "Theory of Gain, Modulation Response, and Spectral Linewidth in AlGaAs Quantum Well Lasers," *IEEE J. Quantum Electron.*, Vol. 21, 1985, p. 1666.

[4] Tsai, C. Y., et al., "Carrier Energy Relaxation Time in Quantum Well Lasers," *IEEE J. Quantum Electron.*, Vol. 31, 1995, p. 2148.

[5] Tsai, C. Y., et al., "Nonlinear Gain Coefficients in Semiconductors: Effects of Carrier Heating," *IEEE J. Quantum Electron.*, Vol. 32, 1996, p. 201.

[6] Tsai, C. Y., et al., "A Small-Signal Analysis of the Modulation Response of High-Speed Quantum Well Lasers: Effects of Spectral Hole Burning, Carrier Heating, and Carrier Diffusion-Capture-Escape," *IEEE J. Quantum Electron.*, Vol. 33, 1997, p. 2084.

[7] Asada, M., Y. Miyamoto, and Y. Suematsu, "Gain and Threshold of Three-Dimensional Quantum Box-Lasers," *IEEE J. Quantum Electron*., Vol. QE-22, 1986, p.1915.

[8] Benisty, H., C. M. Sotomayor-Torres, and C. Weisbuch, "Intrinsic Mechanism for the Poor Luminescence Properties of Quantum-Box Systems," *Phys Rev. B*, Vol. 44, 1991, p. 10945.

[9] Prins, F. E., et al., "Quantum Dots and Quantum Wires with High Optical Quality By Implantation-Induced Intermixing," *Jpn. J. Appl. Phys*., Vol. 32, 1993, p. 6228.

[10] Berger, P. R., et al., "Role of Strain and Growth Conditions on the Growth Front Profile of In*x*Ga1-*x*As on GaAs During the Pseudomorphic Growth Regime," *Appl. Phys. Lett.,* Vol. 53, 1988, p. 684.

[11] Tabuchi, M., S. Noda, and A. Sasaki, "Strain Energy and Critical Thickness of Heteroepitaxial InGaAs Layers on GaAs Substrate," *J. Crystal Growth,* Vol. 115, 1991, p. 169.

[12] Leonard, D., et al., "Direct Formation of Quantum-Sized Dots from Uniform Coherent Islands of InGaAs on GaAs Surfaces," *Appl. Phys. Lett.,* Vol. 63, 1993, p. 3202.

[13] Xie, Q., et al., "Realization of Optically Active Strained InAs Island Quantum Boxes on GaAs(100) via Molecular Beam Epitaxy and the Role of Island Induced Strain Fields," *J. Crystal Growth,* Vol. 150, 1995, p. 357.

[14] Kristaedter, N., et al., "Low Threshold, Large T_0 Injection Laser Emission from (InGa)As Quantum Dots," *Electron. Lett.,* Vol. 30, 1994, p. 1416.

[15] Kamath, K., et al., "Room-Temperature Operation of In0.4Ga0.6As/GaAs Self Organized Quantum Dot Lasers," *Electron. Lett.,* Vol. 32, 1996, p. 1374.

[16] Mirin, R., A. Gossard and J. Bowers, "Room Temperature Lasing from InGaAs Quantum Dots," *Electron. Lett*., Vol. 32, 1996, p. 1372.

[17] Shoji, H., et al., "Temperature Dependent Lasing Characteristics of Multi-Stacked Quantum Dot Lasers," *Appl. Phys. Lett*., Vol. 71, 1997, pp. 71–73.

[18] Berryman, K. W., S. A. Lyon, and M. Segev, "Mid-Infrared Photoconductivity in InAs Quantum Dots," *Appl. Phys. Lett.,* Vol. 70, 1997, p. 1861.

[19] Phillips, J., K. Kamath, and P. Bhattacharya, "Far-Infrared Photoconductivity in Self Organized InAs Quantum Dots," *Appl. Phys. Lett*., Vol. 72, 1998, p. 2020.

[20] Pan, D., E. Towe, and S. Kennerly, "Normal Incidence Intersubband (In,Ga)As/GaAs Quantum Dot Infrared Photodetectors," *Appl. Phys. Lett*., Vol. 73, 1998, p. 1937.

[21] Kastner, M. A., "The Single-Electron Transistor," *Microelectronic Engineering*, Vol. 64, 1992, p. 297.

[22] Phillips, J., et al., "Characteristics of InAs/AlGaAs Self-Organized Quantum Dot Modulation Doped Field Effect Transistors," *Appl. Phys. Lett*., Vol. 72, 1998, p. 3509.

[23] Qasaimeh, O., et al., "Linear and Quadratic Electro-Optic Coefficients of Self Organized In0.4Ga0.6As/GaAs Quantum Dots," *Appl. Phys. Lett*., Vol. 72, 1998, p. 1275.

[24] Ghosh, S., et al., "Non-Linear Optical and Electro-Optic Properties of InGaAs/GaAs Self-Organized Quantum Dots," *J. Vacuum Sci. Technol. B*, Vol. 19, 2001, p. 1455.

[25] Liu, G. T., et al., "Extremely Low Room-Temperature Threshold Current Density Diode Lasers Using InAs Dots in In0.15Ga0.85As Quantum Well," *Electron. Lett.,* Vol. 35, 1999, p. 1163.

[26] Kirstaedter, N., et al., "Gain and Differential Gain of Single Layer InAs/GaAs Quantum Dot Injection Lasers," *Appl. Phys. Lett*., Vol. 69, 1996, p. 1226.

[27] Klotzkin, D., et al., "Enhanced Modulation Bandwidth (20 GHz) of $In_{0.4}Ga_{0.6}As$/GaAs Self-Organized Quantum-Dot Lasers at Cryogenic Temperatures: Role of Carrier Relaxation and Differential Gain," *IEEE Photon. Technol. Lett*., Vol. 10, 1998, p. 932.

[28] Saito, H., K. Nishi, and S. Sugou, "Low Chirp Operation in 1.6 μm Quantum Dot Laser Under 2.5 GHz Direct Modulation," *Electron. Lett*., Vol. 37, 2001, p. 1293.

[29] Sellin, R. L., et al., "Close-to-Ideal Device Characteristics of High-Power InGaAs/GaAs Quantum Dot Lasers," *Appl. Phys. Lett.*, Vol. 78, 2001, p. 1207.

[30] Varangis, P. M., et al., "Low-Threshold Quantum Dot Lasers with 201 nm Tuning Range," *Electron. Lett.*, Vol. 36, 2000, p. 1544.

[31] Maksimov, M. V., et al., "Quantum Dot Injection Heterolaser with Ultrahigh Thermal Stability of the Threshold Current up to 50°C," *Semiconductors*, Vol. 31, 1997, p. 124.

[32] Shchekin, O. B., and D. G. Deppe, "1.3 μm InAs Quantum Dot Laser with T_0 = 161 K from 0 to 80°C," *Appl. Phys. Lett.*, Vol. 80, 2002, p. 3277.

[33] Vurgaftman, I., Y. Lam, and J. Singh, "Carrier Thermalization in Sub Three-Dimensional Electronics Systems: Fundamental Limits on Modulation Bandwidth in Semiconductor Lasers," *IEEE J. Quant. Electron.*, Vol. 50, 1994, p. 1430.

[34] Kamath, K., et al., "Small-Signal Modulation and Differential Gain of Single-Mode Self-Organized $In_{0.4}Ga_{0.6}As$/GaAs Quantum Dot Lasers," *Appl. Phys. Lett.*, Vol. 70, 1997, p. 2952.

[35] Klotzkin, D., K. Kamath, and P. Bhattacharya, "Quantum Capture Times at Room Temperature in High-Speed $In_{0.4}Ga_{0.6}As$-GaAs Self-Organized Quantum Dot Lasers," *IEEE Photonics Technol. Lett.*, Vol. 9, 1997, p. 1301.

[36] Stranski, I. N., and L. Krastanov, *Sitzungsberichte d. Akad. D. Wissenschaften in Wien*, Abt. IIb, Band 146, 1937, p.146.

[37] Jiang, H., and J. Singh, "Strain Distribution and Electronic Spectra of InAs/GaAs Self-Assembled Dots: An Eight-Band Study," *Phys. Rev. B*, Vol. 56, 1996, p. 4696.

[38] Keating, P. N., "Effect of Invariance Requirements on the Elastic Strain Energy of Crystals with Application to the Diamond Structures," *Phys. Rev.*, Vol. 145, 1966, p. 637.

[39] Urayama, J., et al., "Temperature-Dependent Carrier Dynamics in Self-Assembled InGaAs Quantum Dots," *Appl. Phys. Lett.*, Vol. 80, 2002, p. 2162.

[40] Deppe, D. G., and D. L. Huffaker, "Quantum Dimensionality, Entropy, and the Modulation Response of Quantum Dot Lasers," *Appl. Phys. Lett.*, Vol. 77, 2000, p. 3325.

[41] Mathews, D. R., et al., "Experimental Investigation of the Effect of Wetting-Layer States on the Gain-Current Characteristics of Quantum Dot Lasers," *Appl. Phys. Lett.*, Vol. 81, 2002, p. 4904.

[42] Urayama, J., et al., "Observation of Phonon Bottleneck in Quantum Dot Electronic Relaxation," *Phys. Rev. Lett.*, Vol. 86, 2001, p. 4930.

[43] Kim, K., et al., "Gain Dynamics and Ultrafast Spectral Hole Burning in In(Ga)As Self-Organized Quantum Dots," *Appl. Phys. Lett.*, Vol. 81, 2002, p. 670.

[44] Wu, Z. -K., et al., "Ultrafast Carrier Dynamics in Tunneling Injection Quantum Dot Lasers," *Conf. on Lasers and Electro-Optics*, Baltimore, MD, June 2003.

[45] Ghosh, S., S. Pradhan, and P. Bhattacharya, "Dynamic Characteristics of High-Speed $In_{0.4}Ga_{0.6}As$/GaAs Self-Organized Quantum Dot Lasers at Room Temperature," *Appl. Phys. Lett.*, Vol. 81, 2002, p. 3055.

[46] Pradhan, S., S. Ghosh, and P. Bhattacharya, "Temperature Dependent Steady-State Characteristics of High-Performance Tunnel Injection Quantum Dot Lasers," *Electron. Lett.*, Vol. 38, 2002, p. 1449.

[47] Asryan, L., and S. Luryi, "Tunnel-Injection Quantum-Dot Laser: Ultra-High Temperature Stability," *IEEE J. Quantum Electron.*, Vol. 37, 2001, pp. 905–910.

[48] Newell, T. C., et al., "Gain and Linewidth Enhancement Factor in InAs Quantum-Dot Laser Diodes," *IEEE Photon. Technol. Lett.*, Vol. 11, 1999, p. 1527.

[49] Stohs, J., et al., "Gain, Refractive Index Change, and Linewidth Enhancement Factor in Broad-Area GaAs and InGaAs Quantum-Well Lasers," *IEEE J. Quantum Electron.*, Vol. 37, 2001, p. 1449.

[50] Ghosh, S., et al., "Temperature Dependent Measurement of Auger Recombination in InGaAs/GaAs Quantum Dots," *Appl. Phys. Lett.*, Vol., 79, 2001, p. 722.

[51] Bhattacharya, P., et al., "Carrier Dynamics and High-Speed Modulation Properties of Tunnel Injection InGaAs/GaAs Quantum Dot Lasers," *IEEE J. Quantum Electron.,* Vol. 39, 2003, p. 952.

CHAPTER 6

Zinc Oxide-Based Nanostructures

Yicheng Lu and Jian Zhong, Rutgers University

6.1 Introduction

6.1.1 General Properties of ZnO

ZnO is an oxide of the group II metal zinc, and belongs to the P63mc space group [1, 2]. ZnO is on the borderline between a semiconductor and an ionic material. Under most growth conditions, ZnO is an n-type semiconductor, though p-type conductivity of ZnO has also been reported for growth under certain conditions [3, 4]. ZnO exhibits a wurzite structure (hexagonal symmetry) or rock salt structure (cubic symmetry). However, ZnO crystals most commonly stabilize with the wurzite structure (hexagonal symmetry), whereas the crystals exhibit the rock salt phase (cubic symmetry) at high pressure. The wurzite crystal structure and the projection along the [0001] direction are shown in Figure 6.1. Even though it is tetrahedrally bonded, the bonds have a partial ionic character. The lattice parameters of ZnO are a = 0.32495 nm and c = 0.52069 nm at 300K, with a c/a ratio of 1.602, which is close to the 1.633 ratio of an ideal hexagonal close-packed structure. In the direction parallel to the c-axis, the Zn-O distance is 0.1992 nm, and it is 0.1973 nm in all other three directions of the tetrahedral arrangement of nearest neighbors.

The lattice consists of interpenetrating hexagonal close-packed lattices, separated along the c-axis by u = 0.3825 (fractional coordinates). In a unit cell, zinc occupies the (0, 0, 0.3825) and (0.6667, 0.3333, 0.8825) positions and oxygen occupies the (0, 0, 0) and (0.6667, 0.3333, 0.5) positions [5, 6].

The close-packed (0001) planes are made up of two subplanes (A and a), each consisting of the cationic (Zn) and the anionic (O) species, respectively, as shown in Figure 6.1. The crystal can be considered to have the stacking sequence ...AaBbAaBb... as compared to ...AaBbCcAaBbCc... in diamond cubic silicon and sphalerite (GaAs). This results in a remarkable difference in the properties between (0001) and $(000\bar{1})$ planes of ZnO, the former being Zn terminated and the later being O terminated. This structure does not possess a center of symmetry. The lack of inversion symmetry in ZnO leads to piezoelectricity. The polarity of the c-axis results in the Zn-terminated and O-terminated planes displaying vastly different properties.

The (0001) planes in ZnO are polar and hence, with no reconstruction or passivation, have the maximum surface energy among the low-index planes. This is in fact observed under most conditions during vapor phase growth. Crystals grown via

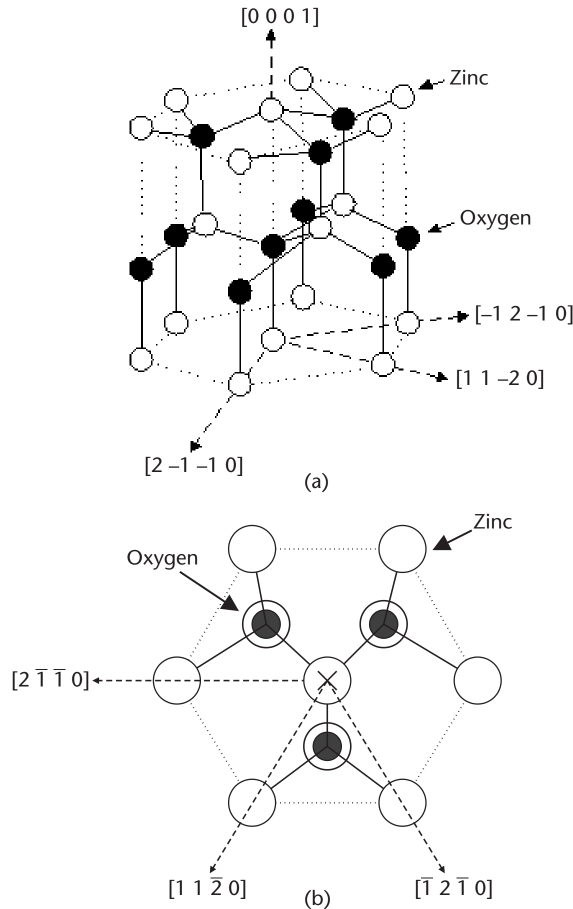


Figure 6.1 (a) Crystal structure of ZnO (wurzite). (b) Projection along the [0001] direction (pointing out of the plane of paper, denoted by the "X" mark at the center).

the vapor phase usually are needle shaped with a hexagonal cross section. The crystals are elongated along the [0001] direction and the prismatic sides of these crystals are usually the {10$\bar{1}$0} or {11$\bar{2}$0} planes, implying that the (0001) plane has the highest energy. As a result, the growth rate along the *c*-axis is the highest. Thus, oriented nanowires are easily formed in the c-direction.

The wurtzite structure of ZnO has a direct energy bandgap of 3.37 eV at room temperature. The lowest conduction band of ZnO is predominantly s-type, and the valance band is p-type (sixfold degenerate). Direct interband transition occurs at the Brillouin zone center ($k = 0$). The valance band splits into three subbands A, B, and C, by spin-orbit and crystal-field interactions, as illustrated in Figure 6.2 [7].

Table 6.1 lists the calculated and measured values of crystal-field splitting (Δ_{cr}) and spin-orbit splitting (Δ_{so}), where a quasi-cubic model was used to calculate the crystal-field splitting and spin-orbit splitting.

The free exciton binding energies associated with the A, B, and C valence bands are 63, 50, and 49 meV, respectively [8]. The high exciton binding energy leads to

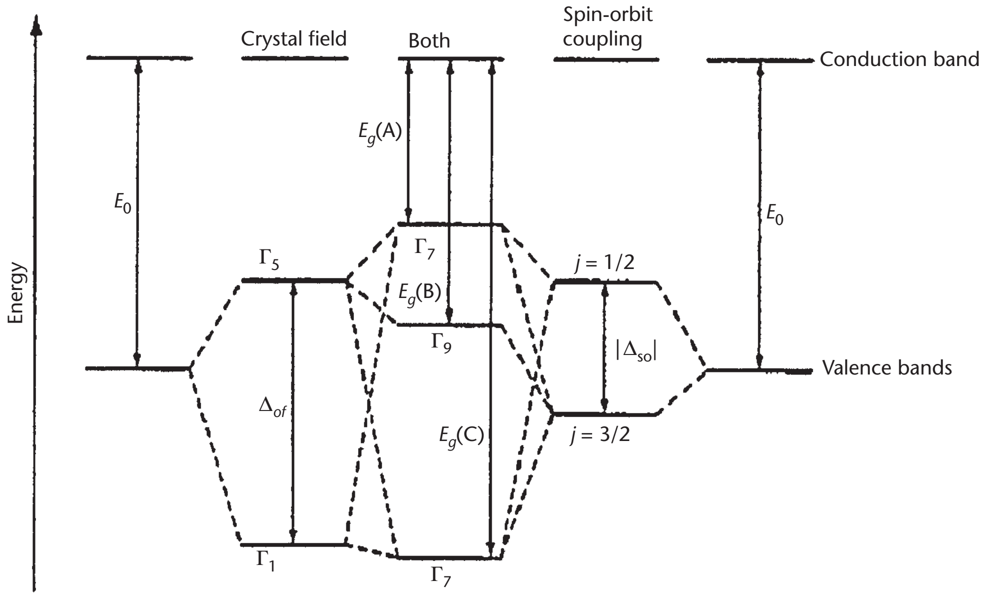


Figure 6.2 Valence band splitting in hexagonal ZnO by crystal-field and spin-orbit coupling. The figure is not drawn to scale. (*After:* [7].)

Table 6.1 Crystal-Field Splitting (Δ_{cr}) and Spin-Orbit Splitting (Δ_{so}) in Bulk ZnO

	Calculated	*Measured*
Δ_{cr} (eV)	+0.0408 [9]	+0.0394 [7]
Δ_{so} (eV)	−0.0047 [9]	−0.0035 [7]

excitonic recombination even at room temperature, in contrast to conventional electron-hole-plasma (EHP) transition, which is promising for high-efficiency, low-threshold photonic devices. Optical anisotropy exists near the band edge due to different selection rules for light polarization perpendicular and parallel to the *c*-axis [10, 11]. Transition for the A and B exciton is mainly allowed for light polarization perpendicular to the *c*-axis, whereas transition for the C exciton is mainly allowed for light polarization parallel to the *c*-axis. By alloying with Cd and Mg, the direct bandgap of ZnO can be tuned from 2.8 to 4.0 eV. Table 6.2 summarizes some important physical properties of bulk ZnO.

6.1.2 ZnO One-Dimensional Nanostructures

One-dimensional nanostructures have attracted intensive attention because of their unique physical, optical, and electrical properties resulting from their low

Table 6.2 Some Important Physical Properties of Bulk ZnO

Properties	*Parameters*
Crystal structure	Wurzite
Space group	$P6_3mc$
Lattice constants at 300K (nm)	a = 0.32495 c = 0.52069 c/a = 1.602 u = 0.3825
Thermal expansion coefficient ($10^{-6}K^{-1}$)	a = 6.5 c = 3.02
Density at 300K ($g \cdot cm^{-3}$)	5.606
Bandgap (eV) at 300K	3.37
Bandgap temperature coefficient ($eV \cdot K^{-1}$)	2.9×10^{-4} [14]
Exciton binding energy (meV)	60
Electron mobility ($cm^2 \cdot V^{-1} \cdot S^{-1}$) at 300K	205 [15]
Hole mobility ($cm^2 \cdot V^{-1} \cdot S^{-1}$) at 300K	180 [16]
Saturation electron drift velocity (107 $cm \cdot S^{-1}$)	3 [17]
Static dielectric constant	8.47 [18]
Index of refraction	(1 μm) (0.45 μm) [19]

dimensionality. The electron-hole interaction will have orders of magnitude enhancement in a nanostructure, due to the dramatically increased electronic density of states near the van Hove singularity. ZnO has an effective electron mass of ~0.24 m_e, and a large exciton binding energy of 60 meV. Thus bulk ZnO has a small exciton Bohr radius (~1.8 nm) [12, 13]. The quantum confinement effect in ZnO nanowires should be observable at the scale of an exciton Bohr radius. An example is the well-width-dependent blue shift in the PL spectra observed in both of ZnO/MgZnO MQW epitaxial films and nanorods [20, 21], with the ZnO well widths ranging from 1 to 5nm. Blue shift of emission from free excitons was reported for ZnO nanorods with diameters smaller than 10 nm, which was ascribed to the quantum size effect [22]. On the other hand, the giant oscillator strength effect could also occur in ZnO nanowires with diameters larger than the bulk exciton Bohr radius but smaller than the optical wavelength [23, 24].

According to the cross-sectional geometry, 1D ZnO nanostructures can be hollow nanotubes or solid nanowires/nanorods (cylindrical). Isotropic materials tend to form nanowires via covalent and electrostatic interactions, whereas nanotubes are most likely exhibited in anisotropic structures, which have strong in-plane forces and weak interlayer van der Waals forces [25]. Other types of 1D nanostructures include nanobelts (rectangular), nanowhiskers (triangle), and so on.

Compared with bulk materials, a significant characteristic of nanostructures is their high surface-to-volume ratio. ZnO nanowires have the same lattice constants

and crystal structure of bulk, confirmed by powder X-ray diffraction (XRD) and TEM data [26–28]. Therefore, many bulk properties are still preserved. The Zn-O bond is half ionic and half covalent. Doping in ZnO is much easier compared with other covalent-bond wide bandgap semiconductors, such as GaN. By appropriate doping, the electrical conductivity of ZnO can be tailored from semiconducting to semimetal, keeping high optical transparency to the visible and UV spectrum regime. ZnO nanotips are attractive for field emission due to their low emission barrier, high saturation velocity, and high aspect ratio. ZnO is more resistant to radiation damage than Si, GaAs, and GaN [29], which is preferred for the long-term stability of field emission emitters in high electric fields. These make ZnO an ideal candidate among transparent conducting oxides (TCOs) for field emission displays.

6.2 Growth Techniques

The growth techniques of 1D ZnO nanostructures discussed in this section are based primarily on bottom-up approaches via either self-assembly or directed assembly (e.g., template-assisted growth). As compared with top-down techniques using nanolithography, bottom-up techniques do not have to follow Tennant's law [30], which is a fundamental technical trade-off designed to achieve high resolution and high throughput simultaneously in the nanofabrication, but they are still capable of creating nanosize features. In the general case, a self-assembly growth could be a self-limited process. For example, it could be a growth spatially defined by the template or a self-ordering process established on a dynamic balance of two opposite physical or chemical interactions, such as attractive and repulsive forces, diffusion and dissolution, and so on.

6.2.1 Growth Mechanisms

Mechanisms for 1D ZnO nanostructure growth include vapor/liquid/solid (VLS), screw dislocation growth, catalyst-free self-nucleation growth, and vapor/solid (VS) mechanisms.

VLS Mechanism

Among the various mechanisms for anisotropic growth of nanostructures, the VLS mechanism is well established and most widely accepted. So far, a variety of growth techniques have been employed in the synthesis of ZnO nanowires via the VLS mechanism, including chemical vapor transport and condensation (CVTC) and catalyst-assisted MBE [19, 31–34]. Control over the nanowire diameter and its morphology has been demonstrated by VLS growth. The as-grown ZnO nanowires also show promising structural and optical properties.

The VLS mechanism was proposed by Wagner and Ellis in 1964 to explain the growth of Si whiskers using Au as a metal catalyst [35]. The VLS process consists primarily of three steps: (1) formation of the liquid alloy droplet, (2) crystal nucleation upon gas adsorption and supersaturation, and (3) axial growth from the crystalline seeds to form nanowires [36]. According to the VLS mechanism, a liquid

phase is formed initially, due to formation of a eutectic phase or the presence of a low-melting-point phase in an alloy system, which consists of the substrate or gas phase constituent. The liquid surface adsorbs the reactant gaseous species much more efficiently than does the solid surface. On supersaturation of the liquid alloy, a nucleation center forms, and serves as a preferred site for the axial growth of a nanowire. The adsorbed gas reactants are then diffused through the liquid phase to the solid/liquid interface, and the growth of the solid phase proceeds. Due to the much larger accommodation coefficient of the reactants in the liquid, growth is much faster at the solid/liquid interface compared to the solid/vapor interface [37–39].

The characteristic of VLS growth is the existence of metallic nanoparticles on top of the nanowires obtained [34, 40–43], typically observed by using TEM. Figure 6.3 illustrates the growth of nanowires by a VLS process [44].

The diameter of a nanowire via VLS growth is primarily determined by the liquid alloy droplet, and the thermodynamic-limited minimum radius is given by [45]:

$$R_{\min} = 2\sigma_{\mathrm{LV}} V_L / RT \ln s$$

where σ_{LV} is the liquid/vapor surface free energy, V_L is the molar volume of liquid, and s is the vapor phase supersaturation.

Selection of metal catalyst species depends on the formation of a eutectic phase at the deposition temperature according to the phase diagram, as well as vapor/liquid/solid interfacial energies and chemical stability in the reaction product [35]. The catalyst species reported in ZnO nanowire growth are Au, Ag, Se, Cu, and transition metal oxides [34, 40, 45–47]. Another metallic particle candidate is Zn, which could be formed as a result of decomposition of the Zn precursor and condensation, though controversy about the growth mechanism in this case still exists if only pure Zn presents in the synthesis [48, 49].

Screw Dislocation Mechanism

The screw dislocation growth is a structural defect evolved process that forms nanowhiskers [50, 51]. When the line of a screw dislocation is parallel to the growth

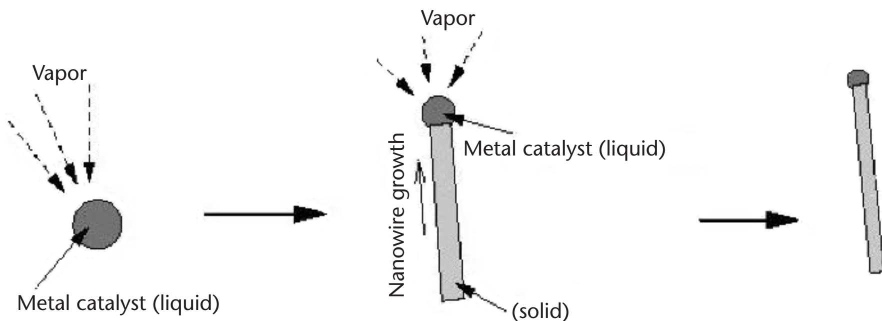


Figure 6.3 Illustration of growth of a nanowire by a VLS process. (*After:* [44].)

axis of the nanowire, for example, the *c*-axis in the case of ZnO nanowires, the spiral plane perpendicular to the screw dislocation line possesses a low-energy step, which acts as a site for sequential growth. Thus, the growth rate along the dislocation line is much faster compared to that in the radial direction. This results in the formation of a 1D nanostructure [52, 53]. Typically, the end tips of the nanowires obtained from this mechanism are conical with a spiral honeycomb-like morphology. A characteristic of screw dislocation growth is the presence of axial screw dislocations, which is often evident from TEM.

Catalyst-Free Self-Nucleation Mechanism

Another possible mechanism to explain the growth of nanowires with sharp prismatic tips is that of catalyst-free, near-equilibrium, self-nucleation growth. ZnO nucleates on the surface of the substrate as a result of the reaction between adsorbed oxygen and zinc from diethyl zinc, similar to ZnO epitaxial growth. The nucleation can be enhanced by the presence of impurities, high temperature, or presence of a secondary phase, which reduces the nucleation energy. As-formed ZnO nuclei could follow an epitaxial relationship with a lattice-matched substrate. The nuclei thus formed then reach a critical size and begin to grow in all directions. The growth habit of the nuclei depend on the relative growth rates of the different crystal faces bounding the nuclei. The growth rate (R) relationship for different ZnO crystal faces is $R_{<001>} > R_{<101>} > R_{<100>} > R_{<00-1>}$. Thus, growth along the *c*-axis [0001] has the highest growth rate, while the negative *c*-axis $[000\bar{1}]$ has the lowest growth rate [54]. Because the growth rate is much more favored along the *c*-axis and it is the direction of free growth without size limitation, it grows into nanowires with sharp prismatic tips. These nanowires not only grow in height but also increase in width. The aspect ratio of the nanowires can be adjusted by control of the growth conditions [52, 55].

VS Mechanism

The VS mechanism is also a catalyst-free process that forms 1D nanostructures. It involves direct vaporization of the solid at a higher temperature, followed by the deposition of the nanostructures at a lower temperature. The VS mechanism has been used to explain the growth of dendrite ZnO nanowires [56] and nanobelts [28].

VS growth is based on Sears's model. During VS growth, dendrite ZnO nanowires with abruptly narrowed sharp tips ["terminal-bottleneck," shown in Figure 6.4(a)] or "terminal coalescence" are formed. Terminal coalescence refers to two nanotips, belonging to different dendrite branches, crystallographically coalescing during the growth, as shown in Figure 6.4(b). In 1950s, G. W. Sears proposed a VS-based whisker growth model [57–59]. According to his theory, the atoms are impinged onto a whisker side surface, adsorbed onto the surface, and diffused along the lateral surface into an atomic sink at the tip of the whisker. Those atoms not reaching the tip are removed by reevaporation because of their unstable energies [56]. As a result, 1D crystal growth occurs at the tip of the nanowire, where

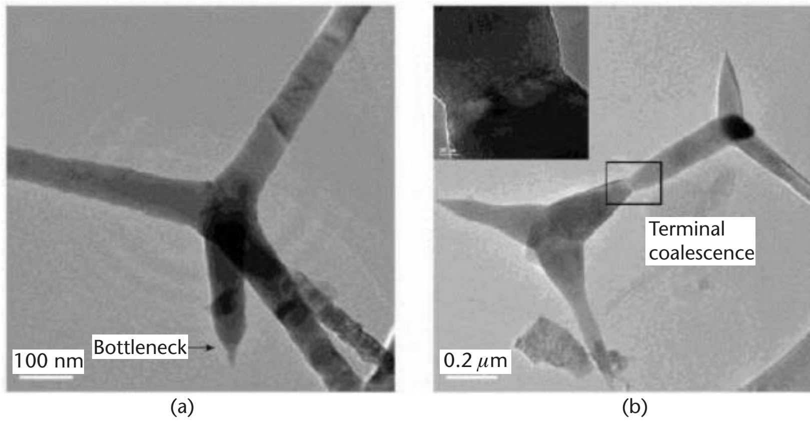


Figure 6.4 TEM images of (a) terminal-bottleneck and (b) terminal-coalescence structures in a VS growth. (*After:* [56].)

an enormous atomic sink exists for sequential growth, and the growth is a diffusion-limited process.

During 1D dendrite growth, multiply twinned ZnO particles are deposited onto the substrate surface by evaporation and condensation of zinc vapor. These fine particles act as nucleation centers. If supersaturation is not large enough to form the euhedral morphology of ZnO crystal, anisotropic 1D growth in specific crystal directions occurs. Typical morphologies observed in a multiple twinned particle initialized growth could be multiple twinned joint faults and stacking faults related to pseudo-twin planes [56].

In a VS growth of nanobelts, the preferential growth direction is determined by surface energy, whereas the morphology is determined by growth kinetics. As ZnO is half ionically bonded, the cation and anion coordinated configuration is likely preserved to keep the charge neutral during the evaporation and condensation process. The surface with lower energy tends to grow larger and flat, forming the enclosure surfaces of the nanobelts. As a result, the reactant gas species tend to diffuse toward the growth front, leading to fast growth along the growth front to form the nanobelts [28]. Unlike VLS, where the crystal growth always takes place at a solid/liquid interface, VS-governed growth proceeds at the vapor and solid crystal interface.

6.2.2 Growth Techniques

A wide variety of growth techniques have been used in the synthesis of ZnO nanostructures, from simple thermal evaporation to more sophisticated state-of-the-art epitaxial growth techniques. The common techniques reported are vapor phase transport, mainly being thermal evaporation and CVTC, and including sputtering and pulse laser deposition (PLD) [60, 61], MBE, CVD, and

MOCVD. Other techniques include template-assisted growth [62] and solution-based synthesis [63–65].

Vapor Phase Transport

Vapor phase transport (VPT) is an atomic deposition process in which material physically or chemically vaporized from a solid source is transported onto a substrate where it condenses and deposits. According to the approaches of source vaporization, VPT can be in the form of thermal evaporation (by heat), laser ablation (by photons), sputtering (by positive ions), or electron beam and CVTC. When the source species are gas, evaporating liquids, or chemically gasified solids, the deposition process is categorized as chemical vapor deposition [66]. So far, a majority of the vapor deposition techniques just mentioned have been developed for the growth of single-crystal 1D ZnO nanostructures. The growth mechanism of VPT has been attributed to either VLS or VS, depending on the presence of a metal catalyst.

Thermal evaporation is one of the simplest methods available to synthesize 1D ZnO nanostructures without the presence of a catalyst [28, 48, 67]. The starting ZnO powders are heated and vaporized at a high temperature, mostly in the range of 1,100°C to 1,400°C. Deposition of ZnO nanostructures occurs on the substrate placed downstream of the carrier gas. The ZnO nanobelts formed under these conditions have widths of 50 to 300 nm, lengths up to several millimeters, and a width-to-thickness ratio of 3 to 10 [48]. Certain growth directions and surface planes are exhibited. The formation of nanobelts is ascribed to a VS process. The morphology and optical properties of nanostructures greatly depend on the specific substrate temperature as well as the carrier gas species, though no clear conclusion has been

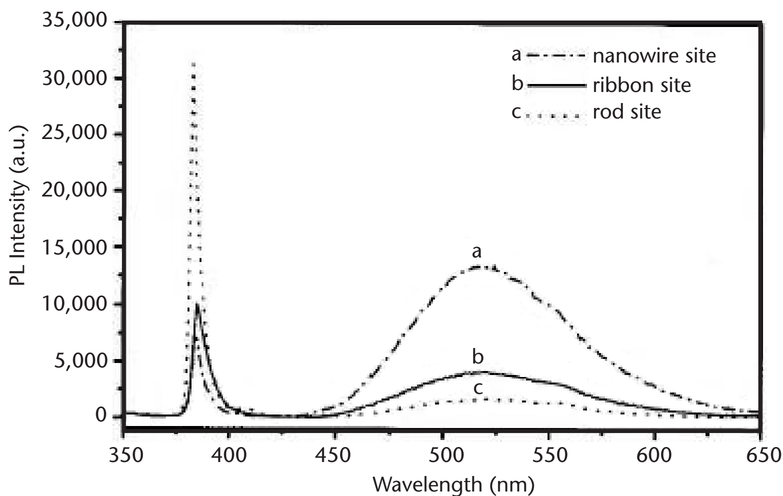


Figure 6.5 Room temperature PL spectra of ZnO nanostructures prepared by thermal evaporation. (*After:* [48].)

drawn on the size-dependent UV to green emission ratio [28, 51, 48]. Figure 6.5 shows the room temperature PL spectra of ZnO nanostructures grown at different temperatures [48]. Nanostructures grown at a higher temperature display stronger UV but weaker green band emission.

ZnO nanowire growth based on a CVTC method has been reported [24, 45, 68]. CVTC involves a chemical reaction between the gaseous species. These gaseous species are usually generated by evaporation, reduction, or some other gaseous reaction. The reaction product then condenses onto the substrate, which is at a lower temperature than the products. The gaseous products of the reaction leave the system. Nanostructures then result from the proper control of the supersaturating species, as in this case, ZnO [53]. Figure 6.6 shows a schematic diagram of a CVTC system for ZnO nanowire growth.

A CVTC growth may use Zn vapors in an oxygen ambience. Zinc evaporated over the substrate reacts with oxygen to form ZnO, which then condenses onto the substrate surface [27, 45]. The growth temperatures reported for vapor transport systems are typically higher than 900°C. CVTC growth requires a few hours of preannealing if Au has been predeposited on the surface. Preannealing of Au results in the formation of small Au islands, which act as nucleation centers for the subsequent growth of ZnO nanowires. The size of the ZnO nanowires obtained depends greatly on the size (diameter and thickness) of these Au islands. The growth mechanism of CVTC is proposed to be either VLS or VS, depending on the presence or absence of the metal catalyst. Figure 6.7 shows a TEM image of a ZnO nanowire with an alloy droplet on the tip, clearly indicating a VLS mechanism [68].

Other approaches to prepare single-crystal ZnO nanowires include wet oxidation of a zinc and selenium mixture in a heating furnace [40], where Se serves as a liquid-forming agent in VLS growth, or thermal evaporation of zinc and carbon mixture in an oxygen ambient [48, 69]. In the latter, Zn or Zn suboxides are proposed to be the nucleation center for nanowire growth via a VLS-like mechanism. The diameter of the nanowires obtained via CVTC typically is tens of nanometers,

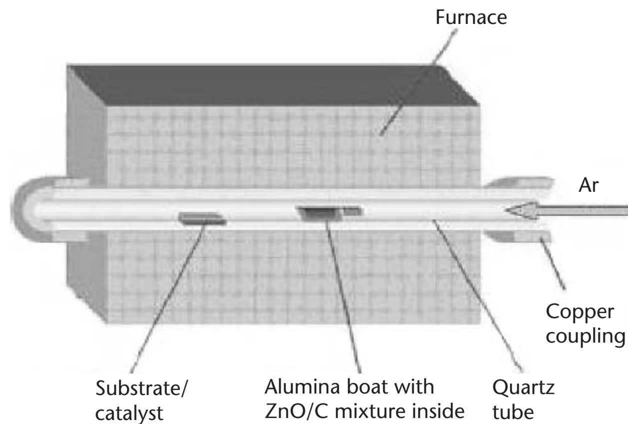


Figure 6.6 Schematic diagram of a CVTC system for ZnO nanowire growth. (*After:* [68].)

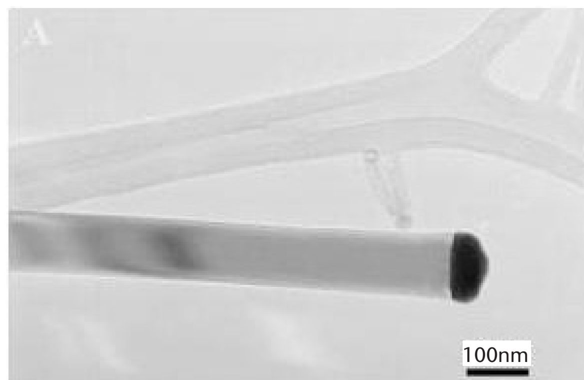


Figure 6.7 TEM image of a ZnO nanowire with an alloy droplet on the tip, clearly indicating a VLS growth mechanism. (*After:* [68].)

and the length several to tens of microns, resulting in a wide range for the aspect ratio. *Selected area electron diffraction* (SAED) confirms the single-crystal quality of these nanowires.

Control of morphology is obtained by proper selection of the substrates and metal catalyst and by use of different vapor sources and evaporation conditions [68]. Metal catalyst islands on the substrate surface were used to control the positions of the nanorods. The thickness of the catalyst layer was found to be an important factor governing the aspect ratio of the ZnO nanotips. The density of the nanowires achieved is in the range of 10^6 to 10^{10} cm^{-2}, controlled by the thickness of the thin Au layer and annealing temperature. Proper selection of substrate is necessary to achieve oriented growth of nanowires, similar to an epitaxial growth [70]. As an example, the ZnO *a*-axis and the sapphire *c*-axis have less than 0.08% mismatch, giving rise to a good epitaxial interface between the (0001) ZnO and the (11-20) sapphire. Thus ZnO nanowires can be grown vertically on an a-plane (11-20) sapphire substrate. Since m-plane (10-10) sapphire has a tilt angle of 30° with respect to the a-plane, and (0001) ZnO is incommensurate with the (10-10)

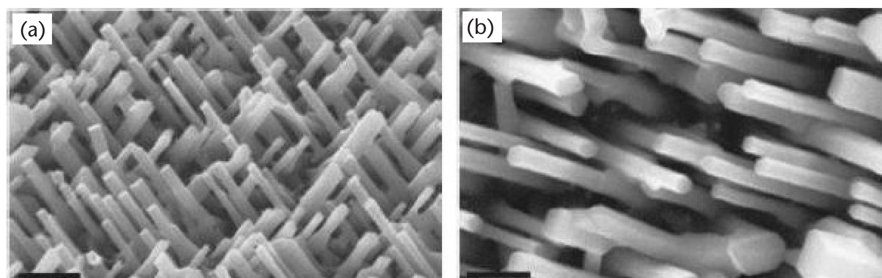


Figure 6.8 (a) FESEM image of ZnO nanowires grown on m-sapphire, showing a tilt angle of ~30 with the substrate normal. (b) Top view of the ZnO nanowires. (*After:* [70].)

plane, ZnO nanowire with a preferential [0001] growth axis on m-sapphire will tilt 30° with the substrate normal. Figure 6.8 shows a *field emission scanning electron microscope* (FESEM) image of ZnO nanowire arrays grown on m-sapphire using an Au-assisted carbothermal reduction VPT approach [70]. The majority of the nanowires grow with their longitudinal axes making an angle of ~30° with the substrate normal, showing epitaxial crystalline nanowire growth.

The nanowire growth on Si substrate was primarily randomly oriented as observed from SEM pictures and XRD data [45, 71]. Multiple ZnO peaks were present in the X-ray scan, with the (001) peak being prominent. This suggested that *c*-axis growth is preferred, even though there is a random orientation. An oriented growth along (001) was reported for (110)-oriented sapphire substrates, wherein the only X-ray peaks observed were (002)- and (004)-oriented ZnO.

Recently, dendrite growth of ZnO nanowires has been reported [56, 72]. Dendrite growth is a hierarchical crystal assembly consisting of joint branches. Figure 6.9(a) shows an SEM image of comb structures of ZnO nanowires by dendrite growth. A comb-like periodic array made of single-crystal ZnO nanowires was synthesized via the CVTC method [72]. The nanowires have uniform diameters of 10 to 300 nm and lengths of 0.5 to 10 μm. They are evenly distributed on the side of a stem with a periodicity of 0.1 to 3 μm. XRD confirms that [0001] is the preferred growth direction. Dendrite growth occurs when zinc vapor at large supersaturation is rapidly condensed and crystallized in an oxygen ambient. Figure 6.9(b) shows a high-resolution TEM image of a comb structure showing the atomically sharp tip and Figure 6.9(c) shows the electron diffraction pattern of the comb structure.

To facilitate the production of ZnO nanowires at a lower temperature, pregrowth reduction is utilized to decompose high-melting-point ZnO (~1,975°C) into low-melting-point Zn or Zn suboxide (~419°C). The Zn vapor is then transported

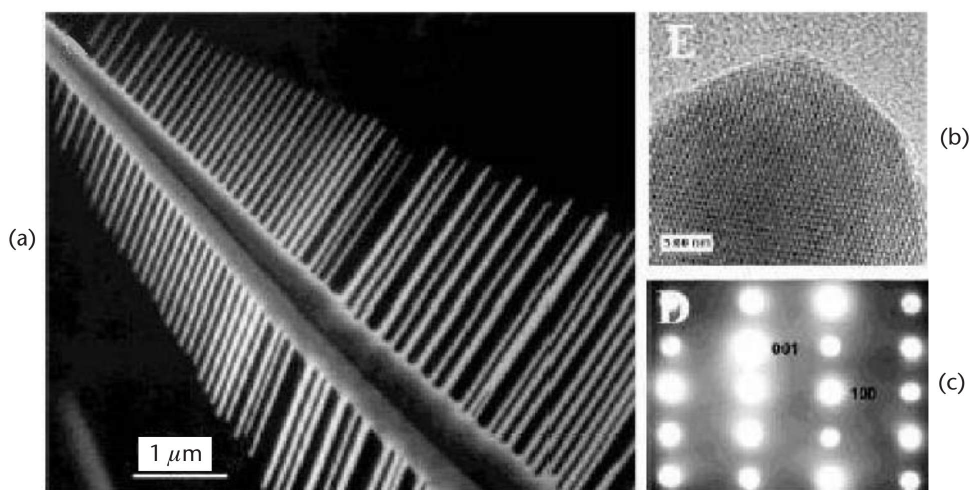


Figure 6.9 (a) SEM image of comb structures of ZnO nanowires. (b) High-resolution TEM image of a comb structure showing atomically sharp tip. (c) Electron diffraction pattern of the comb structure. (*After:* [72].)

and reacts with the metal catalyst to form liquid alloy droplets preferred for a VLS growth. Several reduction approaches are employed to produce Zn vapor:

1. Carbothermal reduction of ZnO. In this approach, graphite is added to pure ZnO powders [45, 48, 69]. At a high source temperature (~1,100°C), Zn or Zn suboxides are produced by the following reactions:

$$ZnO + C \rightarrow Zn + CO_2$$

$$ZnO + CO \rightarrow ZnO_x + CO_2 \, (0 \leq x < 1)$$

2. Hydrogen reduction. A wet oxidation condition (10% O_2, 5% H_2, 85% Ar) is advantageous because of the much lower bonding energy of the H-O bond as compared to the O-O bond to produce atomic oxygen and form ZnO [40, 73].
3. Reduction of zinc salts. Reduction and further oxidation of ZnS powder has been explored to grow tubular ZnO whiskers [74, 75]. At a high temperature (~1,100°C), ZnS reduction follows: ZnS (solid) → Zn (gas) + S (gas). Partial oxidization of ZnS leads to S-doped ZnO nanowires [76].

Due to condensation of Zn vapor onto the substrate surface, a metal catalyst precoating is not necessary for the preparation of ZnO nanowires.

CVD and MOCVD

Chemical vapor deposition is a process based on pyrolytic reactions. The precursor compound, mostly an organic compound, used in the CVD system has a reasonable vapor pressure and a low melting point. A carrier gas transports the vapors of the reacting species. The reaction then proceeds on the substrate surface heated at the growth temperature. The CVD technique has several advantages, such as chemical- and thermodynamic-dependent growth, low growth temperature, control at the atomic level, large area deposition, and high growth rate. CVD also provides the flexibility of in situ doping processes.

CVD

ZnO nanorod growth using CVD has been reported for various substrates, such as sapphire(001), sapphire(110), fused silica, and Si(100) substrates [52, 77]. The growth temperature is relatively low (~500°C). Zinc acetylacetonate hydrate was used as the zinc source, and its vaporizing temperature was reported to be 130°C to 140°C. A mixture of N_2/O_2 was the carrier gas. The SEM pictures for growth on sapphire(001) and sapphire(110) do not reveal well-aligned characteristics. Typical diameters for the nanotips are in the range of 60 to 80 nm, with a height of about 500 nm, resulting in an aspect ratio of about 8:1 to 6:1. The X-ray scans reveal the presence of ZnO(110) accompanying the ZnO(002) and ZnO(004) peaks. The selected area diffraction (SAD) pattern reveal both *c*-axis and *a*-axis orientation of

ZnO on sapphire(110). ZnO nanowires grown on fused silica and Si(100) substrates are, however, uniform, well oriented, and dense.

The SEM and TEM pictures do not reveal the presence of a metallic particle on the nanowire tip, suggesting that the growth mechanism is a non-VLS, catalyst-free mechanism. X-ray images show only the presence of the (002) peak of ZnO, indicating that ZnO nanowires are typically c-oriented with single-crystal quality, which is confirmed by high-resolution TEM. PL characteristics revealed a strong UV light emission peak at around 386 nm at room temperature. TEM pictures, for the growth on Si(100), revealed the existence of a SiO_2 amorphous zone between the Si substrate and ZnO nanotips, suggesting oxidation of the Si substrate to from a thin layer of amorphous SiO_2 during ZnO deposition.

MOCVD

Metalorganic chemical vapor deposition or *metalorganic vapor phase evaporation* (MOVPE) is different from the conventional CVD in its choice of metalorganic precursors as the reactant species. These metalorganics have only one dangling bond and can be easily converted to chemical vapors. This allows for the use of a low growth temperature, depending on the vapor pressure of the precursor. Among all of the growth techniques, MOCVD offers the capabilities of mass production, selective growth, and controllable in situ doping. Moreover, energy band engineering can be realized by appropriately choosing reactant species and growth conditions. The MOCVD technique is compatible with well-developed semiconductor processing technology due to the low growth temperatures. The catalyst-free nature of MOCVD growth of ZnO nanowire eases the postgrowth purification process.

ZnO nanorod or nanotip growth on various substrates, such as c-plane sapphire, fused silica, Si, and GaN using MOCVD has been reported [78–82]. Selective growth of ZnO nanotips has been realized on patterned silicon-on-sapphire (SOS) structures, in which (100) silicon is deposited on r-plane $(11\bar{2}0)$ Al_2O_3 substrates, and on patterned amorphous SiO_2 on r-sapphire substrates [80]. ZnO nanowire growth was carried out in a low-pressure MOCVD system. Diethylzinc (DEZn) and oxygen were used as the Zn metalorganic source and oxidizer, respectively. Film deposition was carried out at a substrate temperature in the range of 300°C to 500°C. Prior to nanowire growth, a thin ZnO buffer layer could be grown at a low temperature followed by substrate annealing.

Figure 6.10 shows a FESEM image of ZnO nanotips grown on various substrates, including (a) c-plane Al_2O_3, (b) epitaxial GaN film grown on c-Al_2O_3, (c) fused silica, and (d) thermally grown SiO_2/Si. ZnO has a wurtzite structure with a close lattice match to GaN. ZnO also satisfies the epitaxial relationship with sapphire. The epitaxial relationship between ZnO and c-sapphire is (0001) ZnO // (0001) Al_2O_3 and $[10\bar{1}0]$ ZnO // $[2\bar{1}\bar{1}1]$ Al_2O_3, while the epitaxial relationship between ZnO and a-sapphire is (0001) ZnO // $(11\bar{2}0)$ Al_2O_3 and $[11\bar{2}0]$ ZnO // [0001] Al_2O_3. Therefore, ZnO on these substrates grows with the *c*-axis perpendicular to the substrate plane. Very dense and smooth epitaxial films of ZnO have also been grown on various orientations of sapphire and GaN.

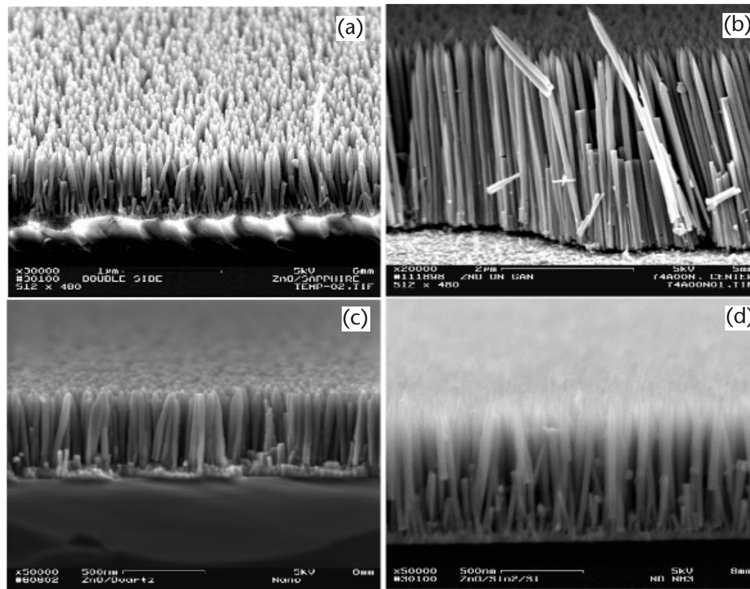


Figure 6.10 FESEM image of ZnO nanotips grown on various substrates, including (a) c-plane Al_2O_3, (b) epitaxial GaN film grown on c-Al_2O_3, (c) fused silica, and (d) thermally grown SiO_2/Si. (*After:* [80].)

Under certain growth conditions, columnar growth can be obtained on these substrates. In the case of columnar growth on various substrates, ZnO nanotips are all preferably oriented along the *c*-axis; they have a base diameter of ~40 nm and terminate with a very sharp nanoscale tip. The crystalline orientation of the ZnO nanotips was determined using XRD measurements as shown in Figure 6.11(a). Figure 6.11(b) shows a dark field TEM image of a single ZnO nanotip, and Figure 6.11(c) shows an electron diffraction image obtained from the single ZnO nanotip aligned to the $[2\bar{1}\bar{1}0]$ zone axis. Defects in single-crystal materials are better characterized in the dark field imaging mode. The dark field TEM image of a single ZnO nanotip shows very few defects. The indexed diffraction pattern further confirms the single-crystal quality of the ZnO nanotips.

The growth mechanism by MOCVD is attributed to a catalyst-free self-nucleation process and a layer-by-layer growth mode. The columnar growth is a result of the high growth rate along the ZnO *c*-axis. ZnO is a polar semiconductor, with (0001) planes being Zn terminated and $(000\bar{1})$ being O terminated. These two crystallographic planes have opposite polarity; hence, they have different surface relaxation energies, resulting in a high growth rate along the *c*-axis. Therefore, by controlling the ZnO growth parameters, ZnO nanotips with the *c*-axis perpendicular to the substrate and with a high aspect ratio can be grown on these substrates. In contrast to the columnar growth, ZnO films grown on r-Al_2O_3 under the same growth conditions result in a flat film with a smooth morphology. The ZnO film shows a flat surface with the epitaxial relationship $(11\bar{2}0)$ ZnO // $(01\bar{1}2)$ Al_2O_3, and

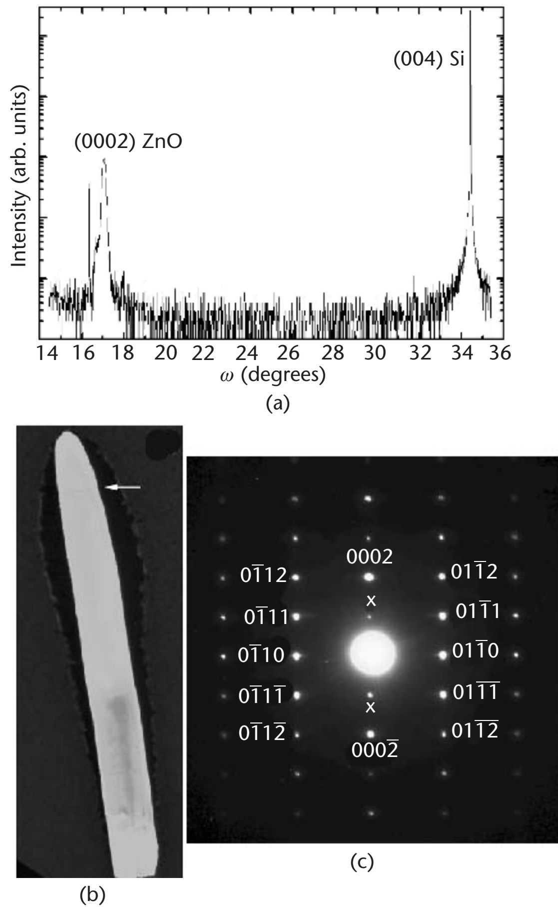


Figure 6.11 (a) XRD analysis of ZnO nanotips grown on silicon substrate. The preferred orientation of the nanotips is along the c-axis. (b) Dark field TEM image of a single ZnO nanotip. The arrow points to a single defect in the column. (c) Electron diffraction image obtained from the single ZnO Nanotip aligned along the $[2\bar{1}\bar{1}0]$ zone axis. The spots marked by X are those due to forbidden reflections. (*After:* [80].)

[0001] ZnO // $[0\bar{1}11]$ Al_2O_3 [13]. Hence, the *c*-axis of ZnO lies in the growth plane. This is different from the ZnO films grown on c-sapphire and a-sapphire substrates. The FWHM ω-rocking curve was measured to be 0.25° for the ZnO film grown on r-Al_2O_3 using MOCVD. The significant difference in the growth of ZnO film on r-sapphire substrates and silicon or SiO_2 has been used to obtain selective growth of ZnO nanotips on patterned SOS substrates [80].

The patterning of the SOS substrates was realized by first depositing a thin SiO2 film on the SOS substrate using LPCVD, which serves as a mask for etching the silicon film. Then, a KOH solution and buffered oxide etchant (BOE) were used to selectively etch silicon and SiO_2, respectively. Figure 6.12 shows a ZnO grown on patterned SOS substrate. The ZnO nanotips are only observed on the exposed silicon top (100) surface and the sidewall (111) surface as KOH anisotropically etches

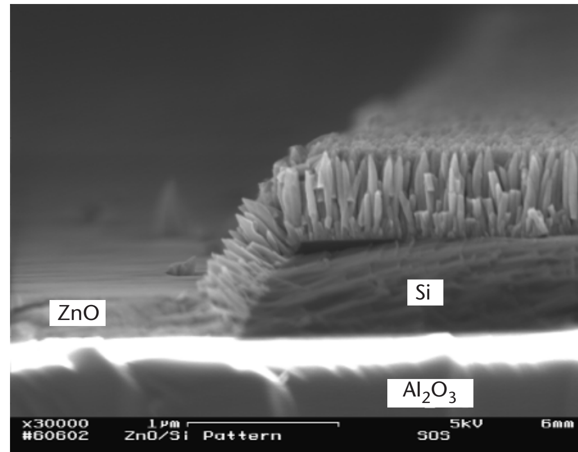


Figure 6.12 FESEM image of selective growth of ZnO nanotips grown on a patterned SOS substrate. (*After:* [80].)

(100) Si producing sidewalls oriented along the <111> direction. The growth of ZnO nanotips on the sidewalls of the silicon islands can be avoided by using dry etching methods, such as inductively coupled plasma (ICP) or reactive ion etching (RIE), which give a vertical etching profile. Similar selective growth was also obtained for patterned amorphous SiO_2 deposited on r-sapphire substrates. The as-grown ZnO nanotips using MOCVD show n-type conductivity.

MOCVD growth and in situ n-type Ga-doping of ZnO single-crystal nanotips on fused silica substrates has been reported very recently [81]. ZnO nanotips with different Ga doping levels were grown on fused silica substrates using a vertical flow MOCVD system. DEZn, oxygen, and triethylgallium (TEGa) were used as the Zn metalorganic source, oxidizer, and Ga metalorganic sources, respectively. The Ga/Zn mole ratio was altered from 10^{-4} to 10^{-2} during the Ga-doped ZnO film growth. Epitaxial ZnO films were simultaneously grown on (0112) r-sapphire substrates to serve as reference samples for doping information of the ZnO nanotips. The resistivities of the epi-ZnO films were characterized using the four-point probe method. Figure 6.13(a) shows the measured resistivities of Ga-doped epitaxial ZnO films versus Ga/Zn mole ratios, whereas Figure 6.13(b) shows a FESEM image of Ga-doped ZnO nanotips grown on fused silica substrate. Figure 6.13(c) shows the top view of the ZnO nanotips. The diameter of the bottom of nanotips is in the range of 40 to 60 nm and the height is ~850 nm, giving an aspect ratio of ~17:1.

Figure 6.14(c) shows a dark field TEM image of the Ga-doped ZnO nanotip along the [2110] zone axis. The selected area electron diffraction and the high-resolution images are shown in insets (a) and (b), respectively. The electron microscopy confirms the single-crystal quality of the Ga-doped ZnO nanotips. High-resolution lattice images show a single-crystalline nanoscale tip. No significant structural change between undoped and Ga-doped ZnO nanotips is observed.

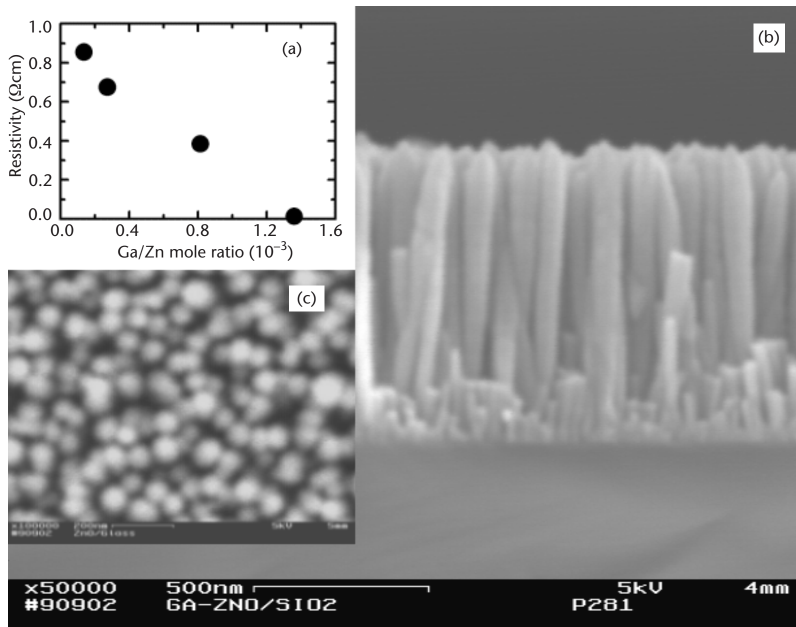


Figure 6.13 (a) Measured resistivities of Ga-doped epitaxial ZnO films versus Ga/Zn mole ratios. (b) FESEM image of Ga-doped ZnO nanotips grown on fused silica substrate. (c) Top view of the ZnO nanotips. (*After:* [81].)

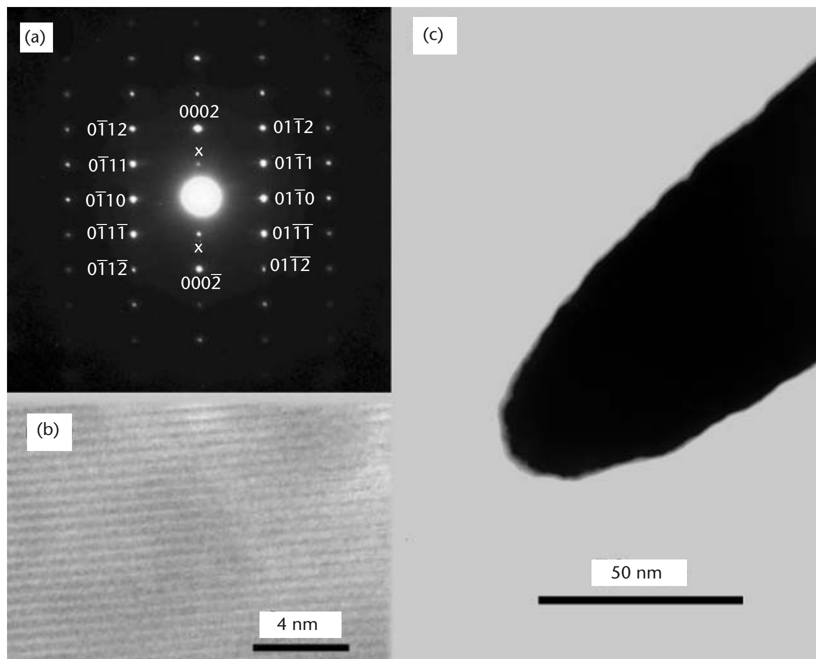


Figure 6.14 (a) Selected area diffraction pattern of a single Ga-doped ZnO nanotip along the [2110] zone axis. (b) Image of a single-crystalline tip from high-resolution TEM. (c) Dark field TEM image of a single Ga-doped ZnO nanotip. (*After:* [81].)

Under certain growth conditions, ZnO grown on various substrates shows a columnar structure with rods ending in tips. Columnar growth is usually considered to result from a high growth rate along the *c*-axis of ZnO. ZnO is a polar semiconductor, with (0001) planes being Zn terminated and $000\bar{1}$ planes being O terminated. These two crystallographic planes have opposite polarity and, hence, different surface energies. Because the fused silica layer is amorphous, the initial ZnO crystals that nucleate on the fused silica surface presumably also do not have a specific epitaxial arrangement, resulting in randomly oriented nuclei. A columnar structure may result from either competitive blocking of all crystallites (by other crystallites) whose *c*-axis is pointing away from the surface normal, or a preferential

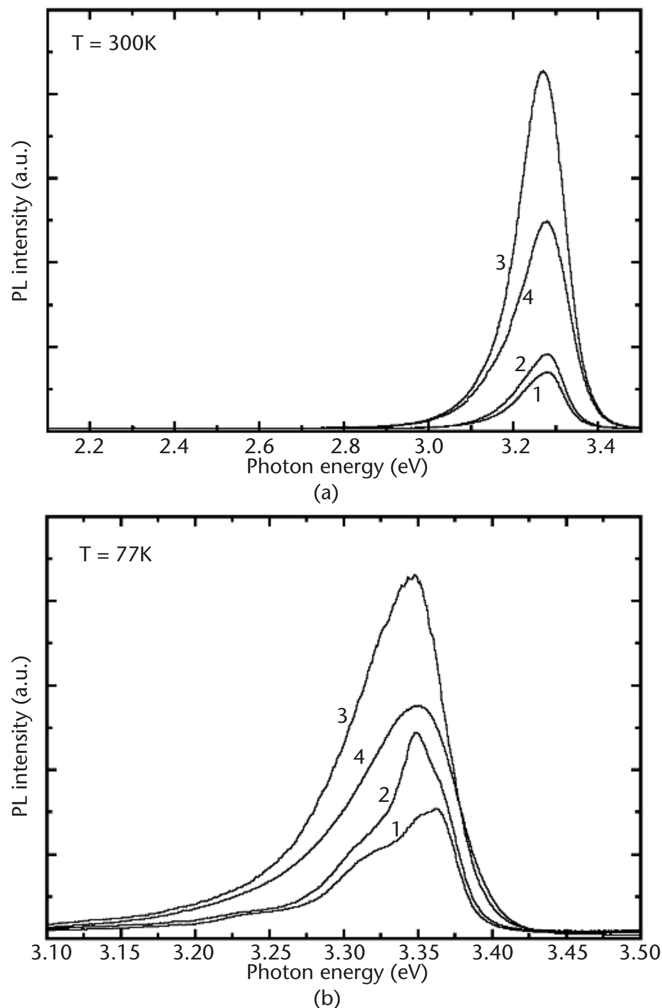


Figure 6.15 (a) Room temperature PL spectra of undoped and Ga-doped ZnO nanotips. (b) 77K PL spectra of undoped and Ga-doped ZnO nanotips. Curve 1 is for the undoped ZnO nanotips with a reference resistivity, ρ_{ref}, of 40 × cm, curves 2, 3, and 4 correspond to Ga-doped ZnO nanotips with ρ_{ref} of 0.58, 4.0 $\times 10^{-3}$, and 3.0 $\times 10^{-3}$ Ωcm, respectively. (*After:* [81].)

orientation of the ZnO nuclei on the relatively flat SiO_2 surface once they reach some critical size.

Figure 6.15(a) shows the room temperature PL spectra of undoped and Ga-doped ZnO nanotips. Curve 1 is for the undoped ZnO nanotips with a reference resistivity, ρ_{ref}, of 40 Ωcm; curves 2, 3, and 4 correspond to Ga-doped ZnO nanotips with ρ_{ref} of 0.58, 4.0×10^{-3}, and 3.0×10^{-3} Ωcm, respectively. It has been reported that due to excess exciton impurity and crystalline defect scattering, there exists a deep-level emission around 2.4 eV in the undoped ZnO nanowires or whiskers. The almost negligible deep-level emission in Figure 6.15(a) suggests good optical quality for the Ga-doped ZnO nanotips grown on the amorphous silica substrates. For ZnO nanotips, the near band-edge recombination monotonically increases with progressively increasing Ga doping [curves 1, 2, and 3 in Figure 6.15(a)]. With a further increase in the Ga doping level, the peak intensity starts to decrease [curve 4 in Figure 6.15(a)]. Note also that half-width broadening can be seen in the spectra when ρ_{ref} is at 4×10^{-3} Ωcm or less [curves 3 and 4 in Figure 6.15(a)].

Figure 6.15(b) shows the 77K PL spectra of the same set of samples. At 77K, undoped ZnO nanotips (curve 1) display a dominant near band-edge emission around 3.36 eV, which is slightly lower than the free exciton emission energy near 3.37 eV and lies in the spectral range of multiple emission lines associated with neutral donor bound exciton recombination. This near band-edge emission may arise from bound exciton emission involving donor-like complexes. Meanwhile, two weak shoulders appear on the low-energy tail of the dominant 3.36-eV peak. These shoulders disappear gradually with Ga incorporation giving rise to a new emission line around 3.35 eV that dominates the spectra. This 3.35-eV emission line shows the same trend as the main peak at room temperature in terms of the intensity and half-width versus doping. Such doping-related behavior has been reported in low-temperature PL spectra of Ga-doped ZnO epitaxial films, where a Ga donor bound exciton complex has been proposed to be responsible for the increase of PL intensity from Ga doping. Figure 6.15(b) suggests that the 3.35-eV peak could be related to the Ga donor in view of its contribution to the near band-edge emission. Therefore, at low or moderate Ga doping levels, the increase of PL intensity is primarily caused by the increase of the impurity emission presumably due to an increased Ga dopant concentration.

In heavily Ga-doped ZnO nanotips, the competition between the two processes (i.e., radiative and nonradiative transitions) becomes prominent. Excess Ga atoms can create nonradiative recombination centers, such as impurity complexes likely involving native defects, or impurity-assisted nonradiative transition channels, resulting in a reduction of the near band-edge luminescence. As the 3.35-eV peak follows the same half-width versus doping trend as observed at room temperature, the linewidth broadening in Ga-doped ZnO nanotips might result from potential fluctuations in the nanostructures that arise from random microscopic distribution of the dopants.

Conductive-tip atomic force microscopy (C-AFM) is another useful method that is used to explore nanoscale electrical characteristics (such as carrier transport) in nanoscale materials. Current-voltage (I-V) spectra were taken at room temperature

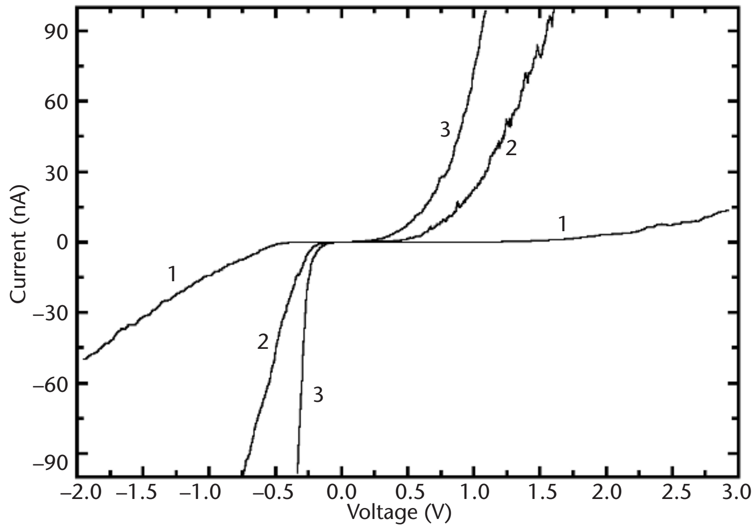


Figure 6.16 Tunneling I-V spectra for both undoped and Ga-doped ZnO nanotips. Curves 1, 2, and 3 correspond to the spectra of undoped and Ga-doped ZnO nanotips with ρ_{ref} of 40 Ωcm (undoped), 0.58 Ωcm (Ga doped), and 4.0×10^{-3} Ωcm (Ga doped), respectively. (*After:* [81].)

using a JEOL vacuum scanning tunneling microscope/AFM system that was operated in a C-AFM mode. Multiple I-V spectra on different ZnO nanotips as well as calibration experiments were performed to ensure reproducibility. Figure 6.16 shows the tunneling I-V spectra for both undoped and Ga-doped ZnO nanotips. Curves 1, 2, and 3 correspond to the spectra of undoped and Ga-doped ZnO nanotips with ρ_{ref} of 40 Ωcm (undoped), 0.58 Ωcm (Ga doped), and 4.0×10^{-3} Ωcm (Ga doped), respectively. The asymmetry in the rising slopes of the conduction and valence bands show that the "undoped" ZnO nanotips are intrinsically n-type, in agreement with the observations in the as-grown epitaxial ZnO films. Significantly steeper rising slopes in I-V spectra are clear in Ga-doped ZnO nanotips (curves 2 and 3) in comparison with those in the undoped nanotips (curve 1). These confirm the conductivity enhancement in ZnO nanotips due to Ga doping. The conductivity-tailoring achieved in Ga-doped ZnO nanotips opens up the possibilities for the fabrication of novel nanoscale devices.

By introducing a Mg precursor into the reactor, MgxZn1-xO/ZnO MQW nanorods were grown on c-sapphire substrates using MOVPE [21]. DEZn, oxygen, and bis(cyclopentadienyl)magnesium (CP2Mg) were used as the reactants and the Mg precursor, respectively, with the Mg composition controlled through the partial pressure and the ZnO well width controlled through growth time. The compositions of $Mg_xZn_{1-x}O$ were determined by energy dispersive X-ray spectroscopy (EDS) and x was approximately 0.2.

The diameters of the nanorods were in the range of 20 to 70 nm, with lengths of 0.5 to 2 μm, and appear to be well aligned vertically from the FESEM image [Figure 6.17(a)]. The nanorods were epitaxial with in-plane alignment and the *c*-axis oriented along the substrate normal. Figure 6.17(a) shows FESEM images

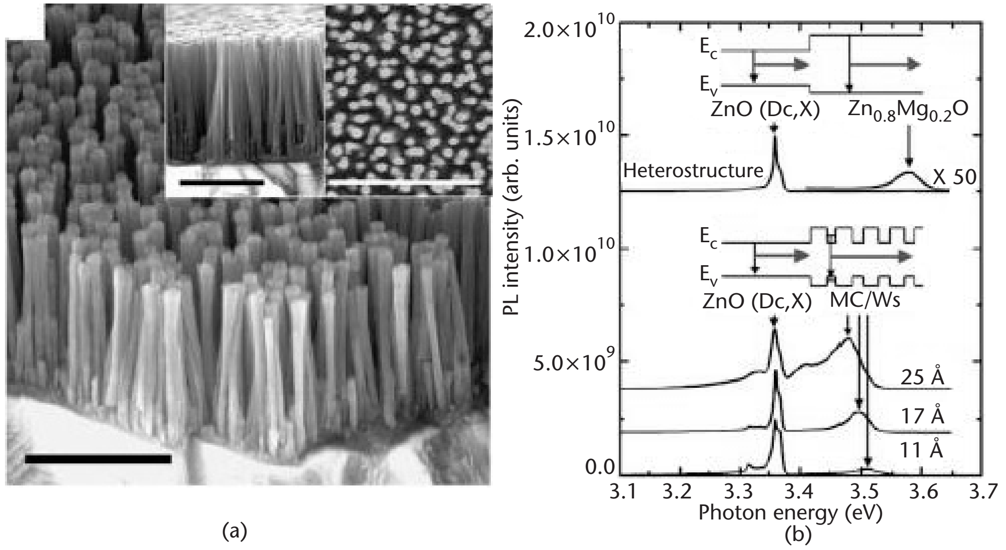


Figure 6.17 (a) FESEM images of $Mg_{0.2}Zn_{0.8}O$/ZnO MQW grown on the tips of ZnO nanorods. Scale bar: 0.5 μm. (b) 10K PL spectra of $Mg_{0.2}Zn_{0.8}O$/ZnO heterostructure and MQW nanorods. (*After:* [21].)

of $Mg_xZn_{1-x}O$/ZnO MQW nanorods. Figure 6.17(b) shows the PL spectra of $Mg_{0.2}Zn_{0.8}O$/ZnO heterostructure and MQW nanorods measured at 10K. The PL peak blue shift is indicated by a decreased well width and is almost negligible at a well width of 11 nm, exhibiting the quantum confinement effect. Based on the simple effective-mass theory, the emission photon energy is given by [22]:

$$E \approx E_{ex} + \frac{\pi^2 \hbar^2}{2}\left(\frac{1}{m_e^*} + \frac{1}{m_h^*}\right)\left(\frac{1}{l^2} + \frac{1}{d^2}\right)$$

where E is the photon energy, E_{ex} is the free exciton energy of bulk ZnO, m_e^* and m_h^* are the effective mass of electron and hole, respectively, and l and d are length and diameter of ZnO nanowire, respectively.

Other Growth Techniques

Catalyst-Assisted MBE

MBE is a widely used thin-film deposition technique involving the reaction of one or more thermal beams of atoms or molecules with a crystalline surface in an ultrahigh vacuum (~10^{-8} Pa). MBE provides for delicate control of atomic-level deposition and dopant species. It does not depend on the low vapor pressure of certain metalorganic sources [83, 84]. Gas phase MBE was used to grow ZnO nanowires and nanorods.

Growth of ZnO nanorods on SiO_2-terminated Si substrates using catalyst-driven MBE has been reported [34, 85]. This nanowire growth was achieved on Ag islands coated with SiO_2-terminated Si substrates, where 2- to 20-nm-thick Ag is deposited

using electron beam evaporation. A Knudsen effusion cell was used for high-purity Zn (99.99%), whereas an oxygen/ozone mixture was used as the oxidizer. The growth temperatures were reported to be on the order of 300°C to 500°C, with a growth rate of about 1 μm/hr. The growth mechanism is proposed to be the VLS mechanism, where Ag acts as the metallic catalyst [26, 29].

The diameter of the ZnO nanorods ranges from 15 to 40 nm, with height of about 2 μm. The ZnO nanorods appear to be dense and entangled from the SEM picture. The nanorod are single crystal with the *c*-axis oriented along the long axis of the rod indicated by SAD. Similar to other VLS-driven growths, TEM image shows the embedded Ag particle at the tip of the rod. A broad green emission at ~2.8 eV is observed at room temperature PL spectrum [34].

Template-Assisted Growth

Template-assisted growth is a conceptually simple and straightforward way to synthesize ordered nanowire arrays. It is a versatile technique in terms of the compounds from which nanowires can be made. In template-assisted growth, a highly ordered nanopore array is formed in a host material, followed by chemical or electrochemical filling of these nanopores using selected materials [62, 86, 87]. Templates can be polymetric filtration membranes or nanoporous alumina films that contain cylindrical holes with diameters down to 4 nm [88].

For template-assisted growth, fabrication of a high-quality template is critical. The self-assembled nanopores are generated from two processes: (1) pore formation determined by the dynamic balance between diffusion-related template growth and template dissolution, and (2) pore ordering, which relies on the delicate balance of the attractive force and the repulsive force between the pores [44].

Anodic alumina membrane (AAM) is a widely used template due to its chemical and thermal stability and high mechanical strength. AAM is made by anodization of an Al substrate. The nanopores typically have a diameter of 10 to 200 nm, length up to 100 μm, and a density of 10^{10} to 10^{12} cm^{-2}. In a three-step template-assisted electrodeposition process [62], a hexagonal nanopore array was prefabricated by a

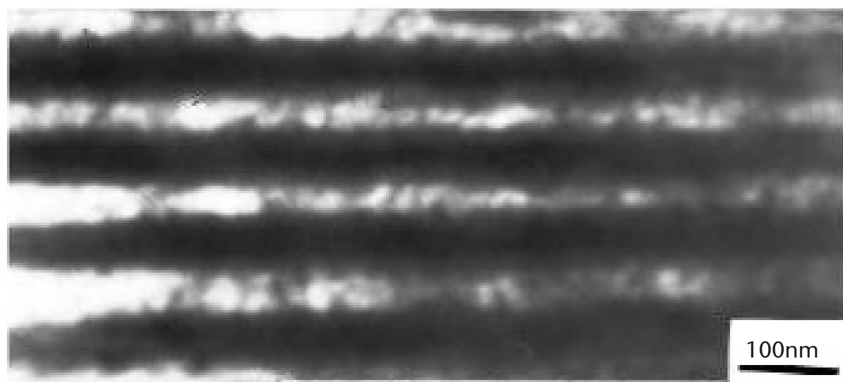


Figure 6.18 TEM image of ZnO nanowire array embedded in AAM. (*After:* [62].)

two-step anodization, then Zn was embedded in the alumina template to form Zn nanowires. Oxidation of Zn nanowires leads to ZnO nanowires. The control of nanopore diameters depends on the cell temperature, applied voltage, and acid solutions in the anodization. Figure 6.18 shows a TEM image of ZnO nanowire array embedded in AAM.

Polycrystalline ZnO nanowires with no preferred crystal orientation are formed in template-assisted growth. A broad green band emission centered around 515 nm dominates the room temperature PL spectrum, which is due to emission from structural defects, such as singly ionized oxygen vacancies, in ZnO nanowires.

To achieve high crystalline nanowires by means of template-assisted growth, an alternative approach is to use the pressure injection technique [44]. However, because ZnO has a high melting point, it is unlikely that the conventional pressure injection technique is applicable for achieving single-crystal ZnO nanowires from the template-assisted growth.

6.2.3 Summary

Among various growth techniques for ZnO nanowires, template-assisted growth is conceptually the simplest but may have difficulties in producing single-crystal nanowires. Solution-based methods are of low cost, but complicated in the growth control. So far, VPT is the most widely adopted method, probably due to the simplicity of the required facilities. VPT is shown to be primarily a high-temperature growth process following either the VLS or VS mechanism, depending on the presence or absence of a metal catalyst.

The VLS mechanism has been combined with a wide variety of growth techniques, including catalyst-assisted MBE. The VLS growth leads to incorporation of metallic impurities in the as-grown ZnO nanowires. By carefully controlled growth conditions, hierarchically ordered ZnO nanowire arrays have been demonstrated using the CVTC method. The VLS-based growth methods require predeposition and annealing of a thin layer of metal catalyst (such as Au or Ag) on the substrate surface to achieve nanowires. Control of the morphology and aspect ratio of the nanowires greatly depends on the size (thickness and diameter) of the metallic islands thus formed.

CVD growth, including MOCVD, follows a catalyst-free growth mechanism, which is preferred to reduce the impurity incorporation. MOCVD provides much faster growth rates and a more uniform and large area deposition; thus, it is suitable for industrial applications. The ZnO nanowires/nanorods achieved are very dense and well oriented with a preferred *c*-axis orientation.

MOCVD and MBE, on the other hand, can provide better control of the composition and dopants as compared to CVD and other growth techniques. The growth temperatures are low, in the range of 300°C to 500°C, which is compatible with the mainstream semiconductor processing technology thus suitable for device integration. Energy band tailoring using (Mg,Zn)O/ZnO MQW structures and in situ n-type doping has been achieved by MOCVD. Moreover, selective MOCVD growth of ZnO nanotips on various substrates, including Si and SOS substrates, has been

reported, which offers the integration of ZnO nanotips, ZnO epitaxial films, and Si microelectronics onto a single chip, which opens up the possibility of fabrication of novel ZnO-based nanoscale devices.

6.3 Characterizations

Inventions and advances in various characterization techniques for nanostructures, such as scanning probe microscopy (the so-called "eyes of the nano world"), have led to visualization of nanostructures down to an atomic scale. Many characteristics of the nanostructures, such as structural, optical, and electron transport properties, can be examined.

6.3.1 Structural Characterizations

SEM

Because the 1D nanostructure is primarily defined by its geometrical size (with a diameter of <100 nm), the structure and morphology characterization of nanoscale features should play an important role in its physical properties. However, in a traditional optical microscope, the spatial resolution is limited by the wave nature of light. The Abbe limit regulates the smallest feature size that can be resolved to be $\sim 0.5\lambda$, which is several orders of magnitude larger than an atomic scale. In an electron microscope, on the other hand, the microvolume to be analyzed is irradiated with a finely focused electron beam, and an energetic electron gives rise to a much shorter de Broglie wavelength than that of visible light. The specific emission volume ("penetration teardrop") determines the achievable image resolution (in the nanometer range). Various secondary signals can be produced from the electron beam/specimen interaction, including secondary electrons, backscattered electrons, Auger electrons, characteristic X-rays, and photons [89]. Among these signals, SEM mainly detects secondary electrons and backscattered electrons.

SEM has been commonly used to analyze the general morphology of ZnO nanostructures, including various geometrical and structural features, such as uniformity, orientation, length and diameter distribution, and spatial density. Figure 6.19 shows an SEM image of ZnO nanotubes grown on Si substrates using thermal wet oxidation of Zn-ZnO powders [90]. These nanotubes exhibit lengths of several microns and outer diameters of 30 to 100 nm. The white arrows in the images indicate the open ends of ZnO nanotubes, revealing that their tubular structures may have round or hexagonal morphologies.

Figure 6.20 shows SEM images of patterned ZnO nanowires grown on an a-sapphire substrate via CVTC [24]. ZnO nanowires originate from the Au-coated sapphire region, whereas no growth proceeds on the region where the catalyst is absent, indicating a VLS-controlled growth.

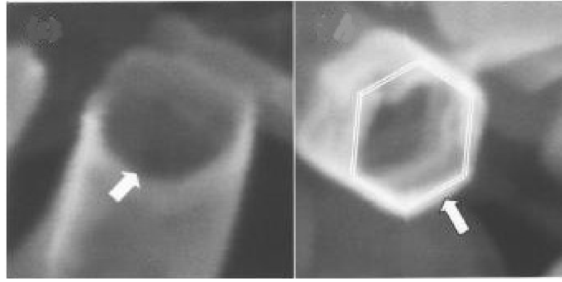


Figure 6.19 SEM images of ZnO nanotubes. White arrows indicate the open ends of nanotubes. (*After:* [90].)

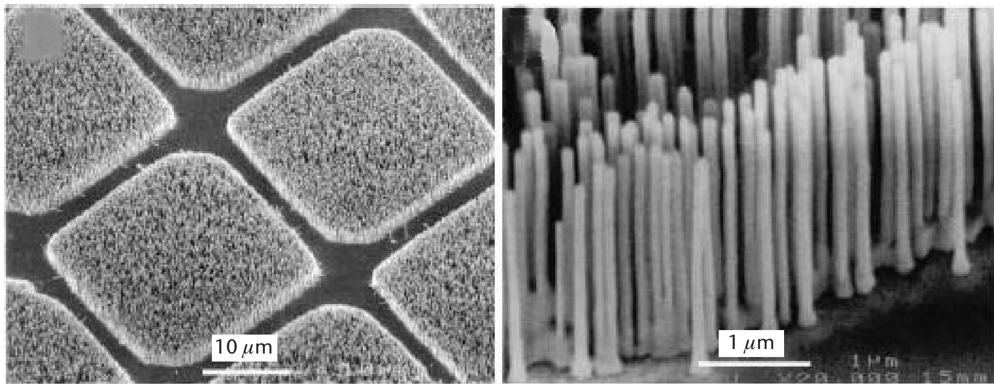


Figure 6.20 SEM images of patterned ZnO nanowires grown on an Au-coated a-sapphire substrate via CVTC. (*After:* [24].)

TEM

TEM, including *high-resolution TEM* (HRTEM), has become a routine and powerful tool to characterize the fine structure of nanomaterials, whereby an atomic scale resolution (in the angstrom range) can be provided. By combining TEM with SAED patterns, a wealth of information, such as growth direction, defect or dislocation, crystallinity, and lattice constants, can be studied.

Figure 6.21(a) is an HRTEM image of a ZnO nanowire grown on an a-plane sapphire by CVTC. A lattice spacing of 2.56 ± 0.05Å between two adjacent ZnO planes is observed, confirming that the preferential growth direction for a ZnO nanowire is along the *c*-axis [24]. Figure 6.21(b) shows an HRTEM image of an S-doped ZnO nanowire grown by VPT. The inset is the corresponding SAED pattern. The white arrow indicates the growth direction along [102], which corresponds to an adjacent lattice spacing of 1.911Å [76]. HRTEM has also been used to reveal structural defects and dislocations. Figure 6.22(a) is an HRTEM image of a

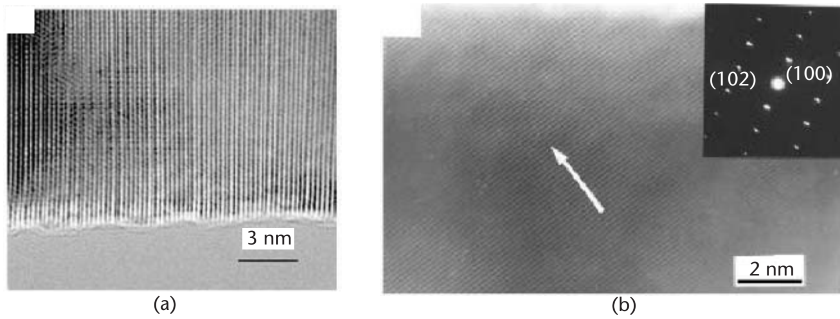


Figure 6.21 (a) HRTEM image of a ZnO nanowire showing its [0001] growth direction. (*After:* [68].) (b) HRTEM image of an S-doped ZnO nanowire grown by VPT. The inset is the corresponding SAED pattern. The white arrow indicates the growth direction along [102]. (*After:* [76].)

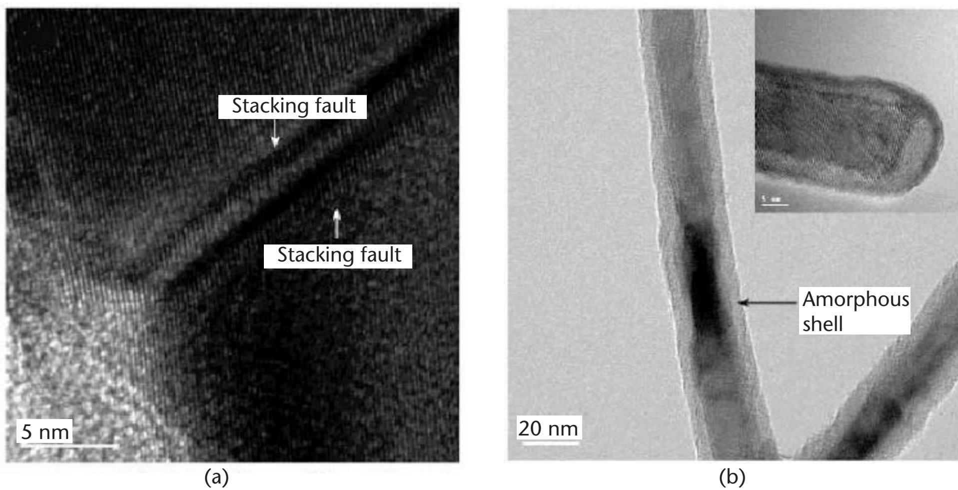


Figure 6.22 (a) HRTEM image of a pseudotwin plane in a dendrite ZnO nanowire. The white arrows label the stacking faults. (b) Core shell structure of ZnO nanowires. (*After:* [56].)

pseudotwin plane in a dendrite ZnO nanowire. The white arrows label the stacking faults [56]. The dendrite ZnO nanowires were prepared by evaporating metal zinc onto thermally oxidized Si substrates. The revealed multitwinned feature is associated with the formation of *multiple twinned particles* (MTPs) due to minimized surface energies. MTPs are usually present in the preparation of fine crystalline particles (such as Si and Ge) by gas evaporation. Based on the TEM observation, the initial nucleation centers have been attributed to the MTPs, where the observed staking faults were incorporated to accommodate lattice packing mismatch or to release lattice strain.

Another example is the observation of fine surface structures through TEM [56, 91]. Figure 6.22(b) shows the core shell structure of ZnO nanowires. A

thin amorphous shell with a thickness of ~4 nm is present beside a single-crystalline ZnO nanowire core. The amorphous shell could consist of zinc oxide as determined by EDS [56].

During the last decade, more advanced imaging tools have been developed, for example, scanning transmission electron microscopy (STEM) and in situ TEM. STEM images obtained from a high-angle annular dark field (HAADF) detector originate from incoherent scattering, which follows Z^2 dependence (Z is the atomic number) of the Rutherford scattering cross section, giving rise to the Z-contrast imaging. The Z-contrast TEM technique has been used to visualize the compositional variation in ZnO/(Mg, Zn)O MQW nanorods [21]. In situ TEM allows for the investigation of nanowire growth at the initial stage [36].

XRD

Similar to the characterization of bulk and epitaxial films, XRD and X-ray EDS are commonly used in the characterization of ZnO nanowires for structure and composition analysis. The XRD patterns provide information on crystallinity and preferred growth direction of the ZnO nanowires. Another application of XRD, when combined with the SEM technique, is to show the alignment of the nanowire array. EDS has been used to check the impurity content (such as doping level and residue of metal catalyst in a VLS growth [70, 78]), component incorporation [e.g., Mg composition in (Mg, Zn) O nanorods] [21, 85], and stoichiometry of ZnO nanowire. However, due to a much shorter X-ray wavelength compared with the size of the nanowire, the scattering effect and thus spectrum broadening may not be ignored.

Scanning Probe Microscopy

The scanning tunneling microscope (STM) was developed in the early 1980s [92]. In STM, the sample is probed by a very sharp tip. Unlike a conventional diffraction-limited far-field technique, STM operates in the near field and is sensitive to the variation of the tip-sample separation. The electron tunneling mechanism is used to control the distance between sample and tip, where the tunneling current depends exponentially on the tip-sample separation. Therefore, it is capable of an atomic scale resolution (~0.1 nm). The invention of STM has inspired the development of a wide variety of scanning probe microscopy (SPM) techniques, for example, AFM, *magnetic force microscopy* (MFM), *electric field gradient microscopy* (EFM), and *near-field scanning optical microscopy* (NSOM). Table 6.3 summarizes properties of various scanning probe techniques [88].

Because of its high resolution, SPM has been employed to study the surface reconstruction of nanowires [93]. AFM imaging has shown ZnO growth on a-plane sapphire at the very initial stage under different growth conditions using MOVPE, including various substrate temperatures and gas flow rates [22]. It was found that a low temperature (≤700°C) is preferred for three-dimensional nucleation toward the growth of ZnO nanorods, whereas two-dimensional layer-by-layer growth occurs at a high growth temperatures (~900°C). At high growth temperatures, the deposited species have sufficient diffusion lengths to migrate along the substrate surface,

Table 6.3 Properties of Scanning Probe Microscopy

Technique	*Mode*	*Property*
AFM	nc/ie, mech, phase/amp	vdW interaction, topography
EFM	nc, mech, phase/amp	Electrostatic force
MFM	nc, mech, phase/amp	Magnetic force, current flow
SSPM, KPM	nc, elec, first harmonic	Potential, work function, adsorbate enthalpy/entropy
SCM	c, F, cap sensor	Capacitance, relative dopant density
SCFM	c, F, third harmonic	dC/dV, dopant profile
SSRM	c, F, dc current	Resistivity, relative dopant density
SGM	nc, elec, amp	Current flow, local band energy, contact potential variation
SIM	nc, elec, phase/amp	Interface potential, capacitance, time constant, local band energy, potential, current flow (in combination with SSPM)
NIM	c, F, freq spectrum	Interface potential, capacitance, time constant, quantitative dopant profiling
PFM	c, elec, amp	$d33$
NPFM	c, elec, second harmonic	Switching dynamics (relaxation time and domain nucleation)
SNDM	c, F, first or third harmonic	dC/dV, dielectric constant
SMM	c, F, phase	Microwave losses, $d33$

Source: [88].

Note: mode: c = contrast, nc = noncontact, ic = intermittent contact, merc = cantilever is driven by a mechanical oscillation, elec = cantilever is driven of responds to an oscillating electrical signal, F = constant force feedback, phase = detection or feedback on phase, amp = detection of amplitude at preset frequency.

giving rise to a successive atomic crystal plane and resulting in epitaxial growth. At low growth temperature, each deposited atom remains at its initial position, forming uniformly distributed small islands on the substrate surface, which act as the nuclei for nanowire growth. The tunneling current in ZnO nanotips was measured using C-AFM [81]. The asymmetry in the rising slopes of the conduction band and valence band proves that the undoped ZnO nanotips are intrinsically n-type, and conductivity tailoring is achieved in Ga-doped ZnO nanotips, from semiconductor to semimetal.

6.3.2 Optical Characterizations

PL

PL spectroscopy provides a simple, nondestructive, and sensitive way to characterize the optical properties of 1D ZnO nanostructures. It does not require the fabrication of nanoscale metal contacts. The spatial resolution of the PL spectrum is defined by the diffraction limit of the excitation source. In combination with near-field optics techniques, such as fiber probes and NSOM, a subwavelength resolution can be achieved [94]. On using an optical cryostat and an ultrafast laser source, temperature-dependent (down to 4.2K or lower) emission and its temporal relaxation behavior (in the picosecond range) can be studied. PL is also a powerful tool for characterizing the quantum confinement effect in nanowires because of the large

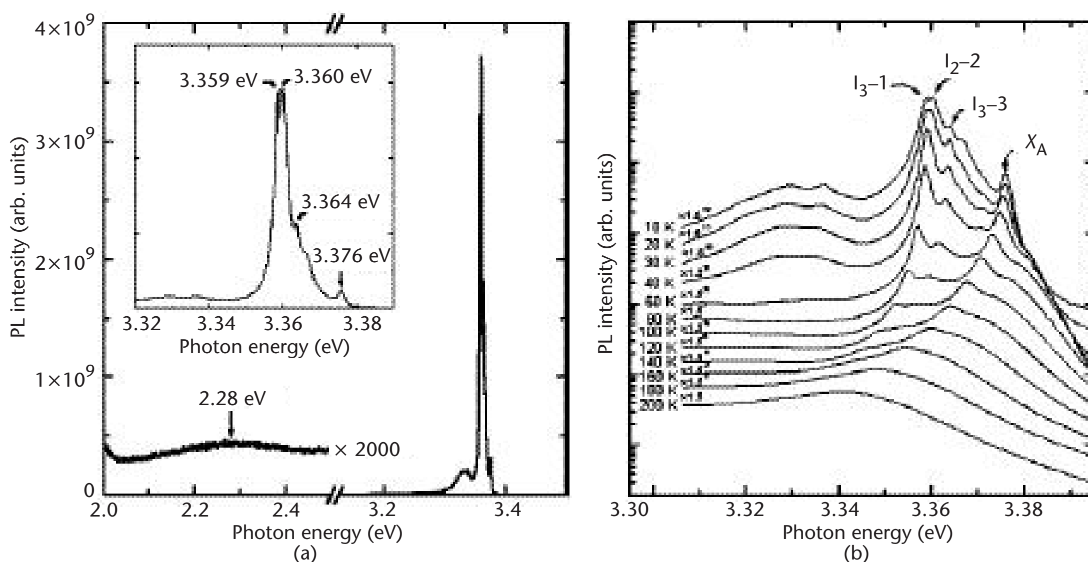


Figure 6.23 (a) PL spectrum of ZnO nanorods grown on c-sapphire by MOVPE at 10K. Inset shows the detailed peaks. (b) Temperature-dependent PL spectra of the same sample. (*After:* [95].)

electron-hole oscillator strength near the singularity in the joint density of states. As an example, a QW width-dependent PL peak blue shift was recently reported in ZnO/(Mg, Zn)O nanorods measured at 10K [21].

ZnO nanowires exhibit good optical properties [95]. Figure 6.23(a) shows a PL spectrum of ZnO nanorods grown by MOVPE at 10K. The near band-edge emission consists of four peaks at 3.359, 3.360, 3.364, and 3.376 eV with a FWHM of 1–3 meV, respectively. The PL peak at 3.376 eV is due to free exciton transition. The peaks at 3.359–3.364 eV, which have slight lower energies than that of free exciton emission, are attributed to neutral-donor bound exciton peaks. The free exciton emission has a FWHM of 3 meV at 10K, comparable to that of high-quality epitaxial thin films and bulk single crystals. On the other hand, the deep-level emission centered at 2.28 eV, which is related to structural defects or impurities, is extremely weak. The sharp excitonic emission and very weak deep-level emission peaks indicate that the ZnO nanorods are of high optical quality. It is known that at very low temperature, free excitons are easily localized or bound by impurities and intrinsic defects and thus form bound excitons. The presence of the free exciton peak suggests that the nanorods have low defect concentration, which could result from the catalyst-free MOCVD growth. Shown in Figure 6.23(b) are the temperature-dependent PL spectra. The free exciton peak at 3.376 eV increases with increasing temperature from 10 to 25K but the intensities of the bound exciton peaks at 3.364, 3.360, and 3.359 eV decrease. This is due to thermal ionization of bound excitons into free excitons.

Temperature-dependent PL spectra of ZnO nanowires grown on sapphire and Si substrates by an Au catalytic VPT process were investigated [96]. Figure 6.24 shows the 80K PL spectra of ZnO nanowires grown on Si substrates. The dominant transition is related to a donor-bound exciton. A second intense peak, which is

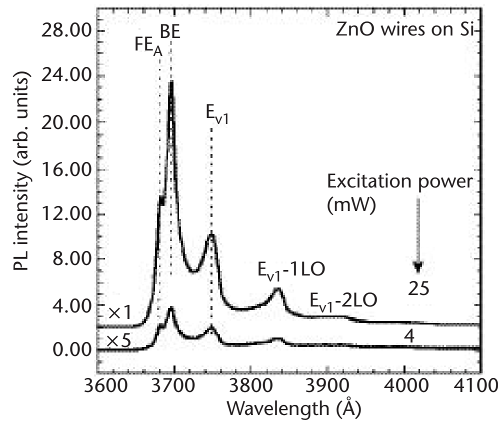


Figure 6.24 80K PL spectra of ZnO nanowires grown on Si substrates by an Au-assisted VPT process. (*After:* [96].)

labeled as E_{VI}, is attributed to a free-to-bound transition. Up to three longitudinal optical (LO)-phonon replicas of the E_{VI} transition were observed. The energy separation between the free A exciton and E_{VI} is about 60 meV at 80K, and this energy separation decreases with increasing temperature. ZnO nanowires grown on sapphire and Si substrates show a similar spectrum, except that the free A exciton is stronger in the sample grown on Si substrates, indicating a low background doping concentration. The E_{VI} transition has a similar temperature-dependent characteristic to the free-to-bound transition and thus is identified as a free-to-bound transition. The bound state has a binding energy of 124 meV. This free-to-bound transition also contributes to room temperature UV emission.

The time-resolved PL spectrum of ZnO nanowires grown on a-sapphire by an Au catalytic CVTC process was also measured [24]. The relaxation process of the radiative recombination in ZnO nanowires consists of a fast component (70 ps) and a slow component (350 ps), compared to 200 ps in ZnO thin films. It is known that defects and impurities, for example, zinc vacancy complexes [97], lead to electron or hole traps that generate nonradiative recombination channels, resulting in a reduced luminescence lifetime. The long lifetime obtained thus shows the high crystal quality of ZnO nanowires.

Raman Spectrum

Raman spectroscopy is a useful tool to characterize many properties (such as crystallinity, strain, and oxygen vacancy) of ZnO bulk and epitaxial materials [98–100]. It is based on the Raman scattering process that corresponds to the interaction between incident photons and the optical modes of lattice vibration. Wurtzite ZnO belongs to the space group $C6v^4$ with two formula units per primitive cell, giving rise to Raman-active phonon modes of E_2 (low), E_2 (high), A_1 LO, A_1 transverse optical (TO), E_1 (LO), and E_1 (TO). The zone center optical phonons are $A_1 + 2E_2 + E_1$ [98]. Compared to bulk ZnO, the decreased dimensions of ZnO nanowires lead to

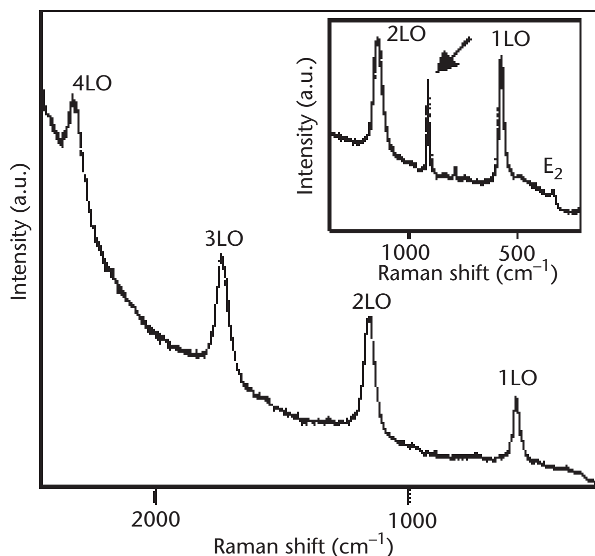


Figure 6.25 Raman spectrum of ZnO nanowires at room temperature. (*After:* [70].)

spatially confined phonons, giving rise to the phonon quantum confinement effect. Characteristics of phonon confinement could be [44]: (1) frequency downshift and Raman lineshape broadening, whereas structure defects and residual stress could also contribute to the phonon linewidth broadening; and (2) generation of new features absent in the corresponding bulk spectra. Figure 6.25 shows a typical Raman spectrum of ZnO nanowires with a mean diameter of 149 nm [70].

Four major bands, centered at 577, 1,152, 1,734, and 2,319 cm^{-1}, with linewidths of 30, 46, 55, and 53 cm^{-1}, are attributed to the Raman LO-phonon scattering and its overtones, respectively. The Lorentzian curve fitting results show that the first-order LO mode consists mostly of the A_1 mode, which has been observed in ZnO nanoparticles but not in bulk ZnO, suggesting that the ZnO nanowires exhibit the phonon quantum confinement effect. The multiphonon linewidth broadening can be also attributed to phonon quantum confinement in the ZnO nanowires.

Nonlinear Optical Properties

ZnO has a hexagonal crystal structure. Lack of inversion symmetry leads to nontrivial physical parameters represented by a third-order tensor, for example, $\chi^{(2)}$ to characterize *second-harmonic generation* (SHG). The nonlinear optical effect in ZnO has been the subject of many recent studies because of its large second-order nonlinearity, high polarization selectivity (~100%), and ultrafast temporal response (on the order of tens of femtoseconds). In addition, nonresonant SHG is mostly wavelength independent (λ_{SHG} > 400 nm), which is in the transparent region of ZnO. These can be utilized to fabricate broad bandwidth photonic devices, such as frequency converters and high-speed optical switches. In particular, the nonlinear optical response is shown enhanced in ZnO nanowires [101]. ZnO nanowires have a

large effective second-order susceptibility $\chi_{eff}^{(2)}$ (≥5.5 pm/V), which is larger than 2.0 pm/V of *β-barium borate* (BBO), a commonly used frequency-doubling crystal, and the measured SHG has a large polarization dependence (>90%). Due to the dramatically increased surface area, surface effect plays an important role in the nonlinear optical process of ZnO nanowires, and $\chi_{eff}^{(2)}$ consists of contribution from the bulk and the surface. It was found that as the nanowire diameter is decreased, the ratio of $\chi_{zzz}^{(2)}/\chi_{zxx}^{(2)}$ decreased, where z is along the longitudinal axis of ZnO nanowire, and x is perpendicular to the longitudinal axis, which is the direction confined by the 1D nanowire. This is because the contribution of the surface effect ($\chi_{surface}^{(2)}$) is enhanced compared to the bulk ($\chi_{bulk}^{(2)}$), as well as the surface effect having a larger contribution to ($\chi_{zxx}^{(2)}$) than to ($\chi_{zzz}^{(2)}$).

6.4 Device Applications

6.4.1 Optical Devices

Nanowire Laser

UV laser sources are of strong research interest because of their broad applications, including nonline-of-sight covert communication, bioagent detection, high-density optical storage, and UV photonics. ZnO has a high exciton binding energy of 60 meV, much greater than the thermal energy at room temperature (26 meV). Thus the exciton in ZnO could be stable even at room temperature, which facilitates efficient excitonic recombination. Nanowire UV lasers and laser arrays could serve as miniaturized light sources for optical interconnections, quantum computing, the microanalysis for biochemical and environmental applications toward the integration of "lab-on-a-chip." So far, p-type doping in ZnO thin films has been reported [3, 4]. However, it is still a technical challenge to achieve reliable and device quality p-type ZnO materials.

Optically pumped ZnO nanowire laser arrays as well as single ZnO nanowire lasers have been demonstrated [24, 72, 102]. The nanowire array was grown on an Au-patterned a-sapphire substrate via a VLS growth mechanism. By proper control the growth conditions, ZnO nanowires with a high area density of 10^{10} cm^{-2} are achieved. The nanowires have diameters of 20 to 150 nm and lengths up to 40 μm. The samples are directly pumped by the fourth harmonic of an Nd:YAG laser at room temperature. Light emission was collected along the *c*-axis of the nanowire. Figure 6.26(a) is the emission spectra from ZnO nanowire array below and above the lasing threshold.

Surface-emitting lasing was observed at 385 nm, with a FWHM of less than 0.3 nm, much narrower than that of spontaneous emission pumped by a He-Cd laser. The nanolaser has a low lasing threshold of ~40 $kWcm^{-2}$, which is one order lower than that of bulk and epitaxial ZnO thin films (~240 $kWcm^{-2}$) [103]. The low lasing threshold is presumably due to an enhanced photon and carrier confinement in the ZnO nanowires, where single-crystalline and well-facetted nanowires function as self-contained optical resonance cavities. The effective coupling between nanowires

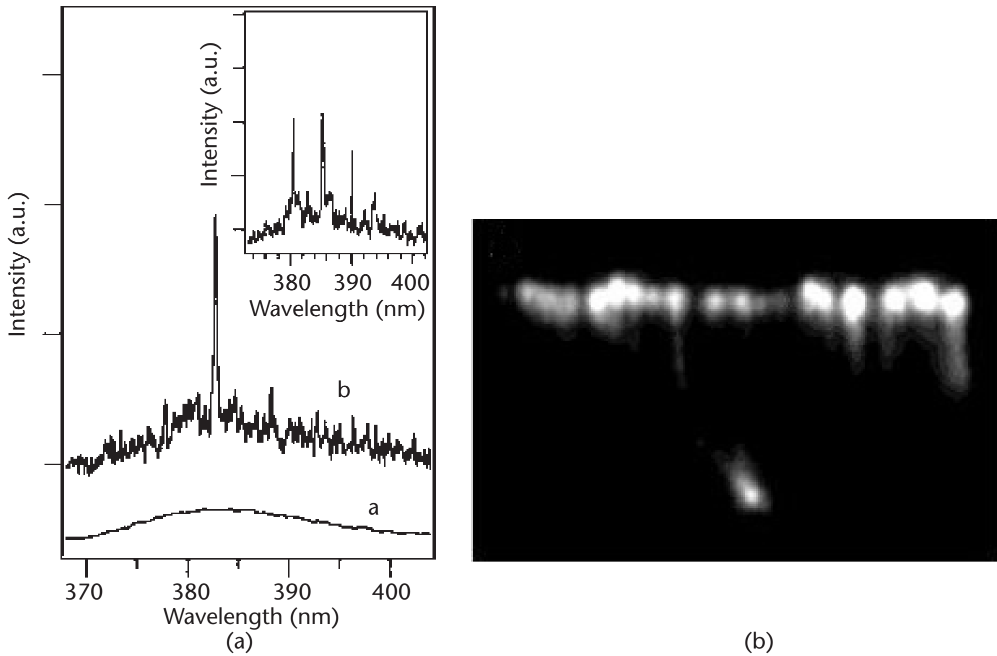


Figure 6.26 (a) Emission spectra from ZnO nanowire array below and above the lasing threshold. (*After:* [24].) (b) Far-field optical image of spatially resolved light emission from the comb structural nanowire array. (*After:* [72].)

could also contribute to the lowering of the threshold. The adjacent lasing mode spacing ν_F for the nanolaser array is determined by:

$$\nu_F = c/2nL$$

where c is the speed of light, n is the refractive index, and L is the resonance cavity length. The possible lasing mechanism due to the amplified stimulated emission based on closed-loop random scattering was precluded by measuring the lasing spectrum from a sparsely distributed nanowire with average spacing of 2 to 3 μm.

The single nanowire lasing behavior is further proved by using NSOM [102]. The nanowire was removed by sonication in an alcohol solvent and dispersed onto a quartz substrate. The emitted light was collected by the fiber probe of an NSOM. The emission spectrum shows that the single nanowire serves as an active Fabry-Perot optical cavity.

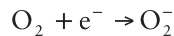
Another configuration is the comb structural nanowire array. The growth of dendrite ZnO nanowire array was described earlier. Figure 6.26(b) shows the far-field optical image of spatially resolved light emission from the comb structural nanowire array. Far-field optical image shows the spatially resolved light emission from the comb structural nanowire array, as well as from individual nanowire [72].

Nanowire Photodetector

ZnO is suitable for an UV photodetector because of its direct wide bandgap and large photoconductivity. ZnO epitaxial film-based photoconductive and Schottky-type UV photodetectors have been demonstrated [104, 105]. More than 4 orders in magnitude of a UV-visible rejection ratio has been achieved.

Large photoconductivity of a single ZnO nanowire has been reported [106]. The ZnO nanowires with diameters ranging from 50 to 300 nm were dispersed on prefabricated gold electrodes. Electrical resistivities without and with UV light irradiation were measured in a four-terminal configuration. The conductivity of ZnO nanowire under UV irradiation increases by 4 to 6 orders of magnitude compared with the dark current with a response time in the order of seconds. The photoresponse has a cutoff wavelength of ~370 nm.

It is known that the photoresponse of ZnO consists of two parts: a fast response due to reversible solid-state processes, such as intrinsic interband transition, and a slow response mainly due to the oxygen adsorption-photodesorption and defect-related recombination processes [107–110]. A two-step process gives rise to the slow photoconductivity: (1) oxygen adsorption on the surface states without light incident as a negatively charged ion by capturing a free electron:



$$O_2 + e^- \rightarrow O_2^-$$

and (2) photodesorption of O_2 by capturing a photogenerated hole:

$$h^+ + O_2^- \rightarrow O_2$$

Therefore, the photogenerated electrons increase the conductivity. The slow process can be suppressed by reducing the trap density and background carrier concentration. ZnO nanowires also show a reversible switching behavior between dark conductivity and photoconductivity when the UV lamp is turned on and off. These suggest that ZnO nanowires are good candidates for optoelectronic switches.

6.4.2 Electronic Devices

Field Emission Devices

Field emission sources are widely used in flat-panel displays. In thermionic electronic emission, electrons are emitted from a heated filament. Field emission is a cold-cathode emission, in which electrons are emitted from a conductor into vacuum through its surface barrier under the applied electric field even at room temperature. Compared with thermionic emission, field emission has the advantages of less thermal shift, low energy spread, and low operating voltage. Traditional *field emission devices* (FEDs) require complicated fabrication processes. An ideal field emitter should be highly conductive, very sharp at the tip end, robust, and easy to fabricate. ZnO has a high melting point, low emission barrier, and high saturation velocity. Furthermore, ZnO is more resistant to radiation damage than Si, GaAs, and GaN, preferred for long-term stability for the field emitters in the high electrical

field. As a widely studied transparent conductive oxide [111], ZnO can be made both highly conductive and optically transparent from the visible to near-UV range through proper doping. These properties make ZnO nanotips a promising candidate for FEDs.

The field emission properties of ZnO nanowires were first studied by C. J. Lee's group in 2002 [112]. In their work, ZnO nanowires were grown on Si substrate using a metal vapor deposition method. Before growth, nanoscale Co particles (average diameter ~6–8 nm) were distributed onto Si substrate as the catalyst. The average length and typical diameter of the ZnO nanowires grown at 550°C were 13 μm and 50 nm, respectively. ZnO nanowires grown at higher temperatures showed larger heights and diameters. XRD, HRTEM, and electron diffraction characterization proved the single crystallinity of the ZnO nanowires. Field emission properties of the as-grown ZnO nanowires were characterized in a vacuum chamber (2×10^{-7} Torr) at room temperature.

Figure 6.27 shows the relationship between emission current density and electric field. The turn-on voltage was found to be about 6.0 V/μm at current density of 0.1 μA/cm^2. The emission current density is ~1 mA/cm^2 at an applied field of about 11.0 V/μm. The inset of Figure 6.27 is the Fowler-Nordheim (FN) plot, from which the field enhancement factor β is calculated to be 847. These figures of merit are inferior to those of carbon nanotube field emitters, which may be due to the flatness of the nanowire tops.

ZnO nanoneedle arrays with sharp-tips show improved field emission performance [113]. The nanoneedles were grown via selenium-assisted vapor phase deposition. The height and diameter of the nanoneedles were about 3 μm and 100 nm, respectively. Figure 6.28(a) shows the TEM image of the ZnO nanoneedles. The sharpness of the nanotip is highlighted in the inset. Figure 6.28(b) shows the field emission I-V curves of the ZnO nanoneedles, which were measured in a vacuum chamber with a pressure of less than 3×10^{-7} Pa. The turn-on field was found to be about 2.4 V/μm. The emission current density can reach 2.4 mA/cm^2 at an applied

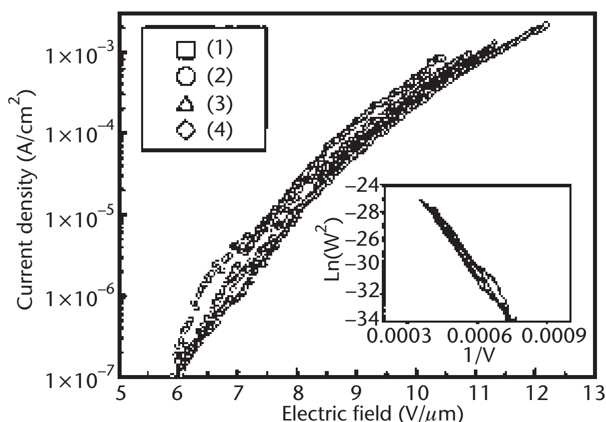


Figure 6.27 Emission current density from ZnO nanowires grown on silicon substrate at 550°C. The inset reveals that the field emission follows FN behavior. (*After:* [112].)

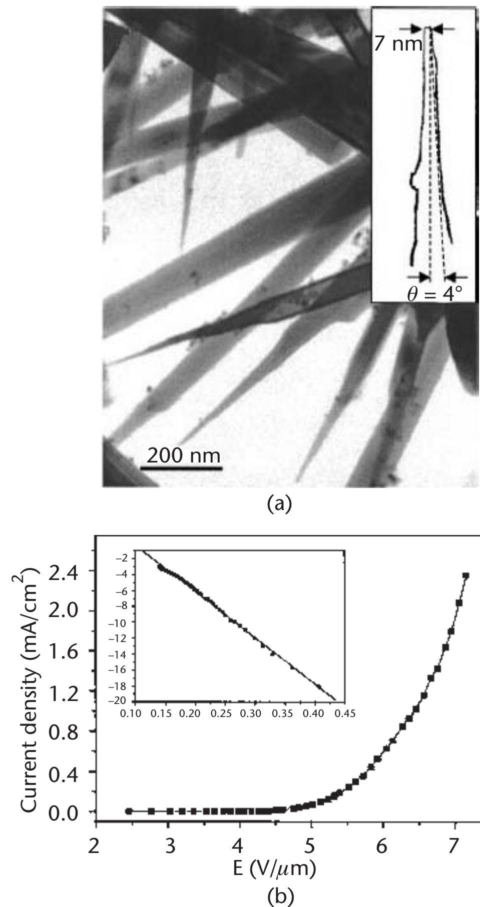


Figure 6.28 (a) TEM image of the ZnO nanoneedles. We can see that the needles have a tiny tip of several nanometers. The inset is the boundary contour of a nanoneedle with a tip size as small as 7 nm. The initial-half angle is around 4°, which reveals that the nanoneedles have excellent configuration for field emission applications. (*After:* [113].) (b) Field emission I-V curve from the ZnO nanoneedle arrays at working distance of 510 μm. The emission current density can be as high as 2.4 mA/cm^2 at the field of 7 V/μm, and the turn-on field of our sample is about 2.4 V/μm. The inset depicts the FN plot, which indicates that the emission satisfies the FN mechanism by showing linear dependence. (*After:* [113].)

field of about 7.0 V/μm. The field enhancement factor β is calculated to be about 2.3×10^6 at 3.8 kV by curve-fitting the FN plot in the inset of Figure 6.28(b). These parameters are comparable to those of carbon nanotube field emitters. It was speculated that introducing Se into the growth process leads to the sharpness of the ZnO nanoneedle, and thus improves the quality of the field emitters. The field emission property of single nanowires was also reported [114]. ZnO nanowires were formed on tungsten substrates using a vapor transport method. Gold served as the catalyst during the growth. The lengths of the nanowires are several tens of microns, and their diameters range from 25 to 200 nm. Measurement of the I-V current for an individual ZnO nanowire was carried out under a vacuum of 5×10^{-8} Torr using a

movable Faraday cup. Emission currents fluctuate as a function of time at both room temperature and 750°C.

Field Effect Transistor

A field effect transistor (FET) based on a single ZnO nanobelt has been reported [115]. ZnO nanobelts with thickness of 10 to 30 nm were synthesized by thermal evaporation of ZnO powders in an alumina tube. A gold electrode array was deposited on a SiO_2/P+ Si substrate. A 120-nm SiO_2 layer acted as the gate oxide layer. The as-grown nanobelts were dispersed in ethanol by ultrasonication, then dried onto prefabricated gold electrode arrays, with varying gaps from 100 nm to 6 μm. The back gate electrode was fabricated by evaporation of gold on P^+ Si side of the substrate. A ZnO nanobelt FET has showed a high threshold voltage of -15V, which is due to the high resistive contact resulting from simply depositing the ZnO nanobelt on top of the gold electrodes. It has a switching ratio of nearly 100 and a peak conductivity of 1.25 × 10-3 $(\Omega\cdot\text{cm})^{-1}$. It also shows UV sensitivity.

References

[1] Gorla, C., "Properties of ZnO Thin Films Grown on R-Plane Al2O3 by Metalorganic Chemical Vapor," Ph.D. Thesis, Rutgers University, 1999.

[2] Saraf, G., "Structural Properties of ZnO Nanotips Grown by Metalorganic Chemical Vapor Deposition on SiO2/Si Substrates," M.S. thesis, Rutgers University, 2003.

[3] Kim, K., et al., "Realization of p-Type ZnO Thin Films Via Phosphorus Doping and Thermal Activation of the Dopant," *Appl. Phys. Lett.*, Vol. 83, No. 1, 2003, pp. 63–65.

[4] Ryu, Y. R., T. S. Lee, and H. W. White, "Properties of Arsenic-Doped p-Type ZnO Grown by Hybrid Beam Deposition," *Appl. Phys. Lett.*, Vol. 83, No. 1, 2003, pp. 87–89.

[5] Pearson, W. B., *A Handbook of Lattice Spacings and Structures of Metals and Alloys, Vol. 2*, New York: Pergamon Press, 1967.

[6] International Tables for X-Ray Crystallography, The International Union of Crystallography, 1965.

[7] Mang, A., K. Reimann, and S. Rübenacke, "Band Gaps, Crystal-Field Splitting, Spin-Orbit Coupling, and Exciton Binding Energies in ZnO Under Hydrostatic Pressure," *Solid State Commun.*, Vol. 94, No. 4, 1995, pp. 251–254.

[8] Park, Y. S., et al., "Exciton Spectrum of ZnO," *Phys. Rev*., Vol. 143, No. 2, March 1966.

[9] Liang, W. Y., and A. D. Yoffe, "Transmission Spectra of ZnO Single Crystals," *Phys. Rev. Lett*., Vol. 20, 1968.

[10] Yoshikawa, H., and S. Adachi, "Optical Constants of ZnO," *Jpn. J. Appl. Phys*., Vol. 36, Part 1, No. 10, October 1997, pp. 6237–6243.

[11] Wraback, M., et al., "High Contrast, Ultrafast Optically Addressed Ultraviolet Light Modulator Based Upon Optical Anisotropy in ZnO Films Grown on R-Plane Sapphire," *Appl. Phys. Lett*., Vol. 74, No. 4, 1999, pp. 507–509.

[12] Gil, B., and A. V. Kavokin, "Giant Exciton-Light Coupling in ZnO Quantum Dots," *Appl. Phys. Lett*., Vol. 81, No. 4, 2002, pp. 748–750.

[13] Wong, E., and P. C. Searson, *Appl. Phys. Lett*., Vol. 74, 1999, p. 2939.

[14] Tomm, J. W., et al., "Optical and Photoelectrical Properties of Oriented ZnO Films," *J. Appl. Phys*., Vol. 87, No. 4, February 2000.

[15] Yang, W., et al., "Ultraviolet Photoconductive Detector Based on Epitaxial Mg0.34Zn0.66O Thin Films," *Appl. Phys. Lett.*, Vol. 78, No. 18, April 2002.

[16] Sze, S. M., *Physics of Semiconductor Devices*, 2nd ed., New York: Wiley-Interscience Appendix G, 1981.

[17] Albrecht, J. D., et al., "High Field Electron Transport Properties of Bulk ZnO," *J. Appl. Phys.*, Vol. 86, No. 12, December 1999.

[18] Park, Y. S., et al., "Exciton Spectrum of ZnO," *Phys. Rev.*, Vol. 143, No. 2, March 1966.

[19] Teng, C. W., et al., "Refractive Indices and Absorption Coefficients of MgxZn1-XO Alloys," *Appl. Phys. Lett.*, Vol. 76, No. 8, February 2000.

[20] Ohtomo, A., et al., "Structure and Optical Properties of $ZnO/Mg_{0.2}Zn_{0.8}O$," *Superlattices*, Vol. 75, No. 7, 1999, pp. 980–982.

[21] Park, W. I., et al., "Quantum Confinement Observed in ZnO/MgZnO Nanorod Heterostructures," *Adv. Materials*, Vol. 15, No. 6, 2003, pp. 526–529.

[22] Maejima, K., et al., "Growth of ZnO Nanorods on A-Plane (11-20) Sapphire by Metal-Organic Vapor Phase Epitaxy," *Jpn. J. Appl. Phys.*, Part.1, Vol. 42, No. 5a, 2003 pp. 2600–2604.

[23] Kayamura, Y., *Phys. Rev. B*, Vol. 38, 1988, p. 9797.

[24] Huang, M. H., et al., "Room-Temperature Ultraviolet Nanowire Nanolasers," *Science*, Vol. 292, No. 5523, 2001, pp. 1897–1899.

[25] Dresselhaus, M. S., et al., "Nanowires and Nanotubes," *Mater. Sci. Eng.*, Vol. 23, 2003, pp. 129–140.

[26] Roy, V. A. L., et al., "Luminescent and Structural Properties of ZnO Nanorods Prepared Under Different Conditions," *Appl. Phys. Lett.*, Vol. 83, No. 1, 2003, pp. 141–143.

[27] Li, Y. B., et al., "ZnO Nanobelts Grown on Si Substrate," *Appl. Phys. Lett.*, Vol. 81, No. 1, 2002, pp. 144—146.

[28] Pan, Z. W., Z. R. Dai, and Z. W. Wang, "Nanobelts of Semiconducting Oxide," *Science*, Vol. 291, March 2001, pp. 1947—1949.

[29] Look, D. C., *Mater. Sci. Eng. B*, Vol. 80, No. 383, 2001.

[30] Marrian, C. R. K., and D. M. Tennant, "Nanofabrication," *J. Vac. Sci. Technol. A*, Vol. 21, No. 5, September/October 2003, pp. S207–S215.

[31] Haupt, M., et al., "Ultraviolet-Emitting ZnO Nanowhiskers Prepared by a Vapor Transport Process on Prestructured Surfaces with Self-Assembled Polymers," *J. Appl. Phys.*, Vol. 93, No. 10, 2003, pp. 6252–6257.

[32] Wang, Y. W., et al., *J. Crystal Growth*, Vol. 234, 2002, p. 171.

[33] Park, K., et al., *Jpn. J. Appl. Phys.*, Part 1, Vol. 41, 2002, p. 7317.

[34] Heo, Y. W., et al., "Site-Specific Growth of ZnO Nanorods Using Catalysis-Driven Molecular-Beam Epitaxy," *Appl. Phys. Lett.*, Vol. 81, No. 16, 2002, pp. 3046—3048.

[35] Wagner, R. S., and W. C. Ellis, *Appl. Phys. Lett.*, Vol. 4, No. 89, 1964.

[36] Wu, Y., and P. Yang, "Watching GaN Nanowires Grow," *Nano Lett.*, Vol. 3, No. 6, 2003, pp. 867—869.

[37] Suchet, J. P., *Chemical Physics of Semiconductors*, New York: Van Nostrand, 1965.

[38] Wang, Y., et al., "Catalytic Growth and Photoluminescence Properties of Semiconductor Single-Crystal ZnS Nanowires," *Chem. Phys. Lett.*, Vol. 357, No. 3–4, 2002, pp. 314–318.

[39] Duan, X, and C. M. Lieber, "General Synthesis of Compound Semiconductor Nanowires," *Adv. Materials*, Vol. 12, No. 4, 2000, pp. 298–302.

[40] Kong, Y. C., et al., "Ultraviolet-Emitting ZnO Nanowires Synthesized by a Physical Vapor Deposition Approach," *Appl. Phys. Lett.*, Vol. 78, No. 4, 2001, pp. 407–409.

[41] Zhang, R., Y. Lifshitz, and S. Lee, "Oxide-Assisted Growth of Semiconducting Nanowires," *Adv. Materials*, Vol. 15, No. 7–8, 2003, pp. 635–640.

[42] Hu, J. Q., et al., *Chem. Phys. Lett.*, Vol. 344, 2001, p. 97.

[43] Westwater, J., D. P. Gosain, and S. Usui, "Control of the Size and Position of Silicon Nanowires Grown Via the Vapor-Liquid-Solid Technique," *Jpn. J. Appl. Phys.*, Part 1, Vol. 36, 1997, p. 6204.

[44] http://Web.Mit.Edu/Mrschapter/Www/Docs/Iap/Iap03_One_D/Nanowires_Review.Pdf.

[45] Huang, M. H., et al., "Catalytic Growth of Zinc Oxide Nanowires by Vapor Transport," *Adv. Materials*, Vol. 13, No. 2, 2001, pp. 113–116.

[46] Lyu, S. C., et al., "Low Temperature Growth and Photoluminescence of Well-Aligned Zinc Oxide Nanowires," *Chem. Phys. Lett.*, Vol. 363, 2002, pp. 134–138.

[47] Li, S. Y., C. Y. Lee, and T. Y. Tseng, "Copper-Catalyzed ZnO Nanowires on Silicon (100) Grown by Vapor-Liquid-Solid Process," *J. Crystal Growth*, Vol. 247, 2003, pp. 357–362.

[48] Yao, B. D., Y. F. Chan, and N. Wang, "Formation of ZnO Nanostructures by a Simple Way of Thermal Evaporation," *Appl. Phys. Lett.*, Vol. 81, No. 4, 2002, pp. 757–759.

[49] Dai, Y., et al., "Synthesis and Optical Properties of Tetrapod-Like Zinc Oxide Nanorods," *Chem. Phys. Lett.*, Vol. 358, 2002, pp. 83–86.

[50] Frank, F. C., *Discovery Faraday Soc.*, Vol. 5, 1949, p. 48.

[51] Drum, C. M., and J. W. Mitchell, "Electron Microscopic Examination of Role of Axial Dislocations in Growth of AlN Whiskers," *Appl. Phys. Lett.*, Vol. 4, No. 9, 1964, pp. 164–165.

[52] Wu, J., and S. Liu, "Catalyst-Free Growth and Characterization of ZnO Nanorods," *J. Phys. Chem. B,* Vol. 106, 2002, pp. 9546–9551.

[53] Xia, Y., et al., "One-Dimensional Nanostructures: Synthesis, Characterization and Applications," *Adv. Mater.*, Vol. 15, No. 5, 2003, pp. 353–389.

[54] Li, W. J., et al., "Growth Mechanism and Growth Habit of Oxide Crystals," *J. Crystal Growth*, Vol. 203, No. 1–2, 1999, pp. 186–196.

[55] Dai, Y., et al., "Synthesis and Optical Properties of Tetrapod-Like Zinc Oxide Nanorods," *Chem. Phys. Lett.*, Vol. 358, No. 1–2, 2002, pp. 83–86.

[56] Zhang, Y., et al., "Synthesis, Microstructure, and Growth Mechanism of Dendrite Nanowires," *J. Phys. Chem. B*, Vol. 107, No. 33, 2003, pp. 8289–8293.

[57] Sears, G. W., "A Growth Mechanism for Mercury Whiskers," *Acta Metallurgica*, Vol. 3, 1955, pp. 361–366.

[58] Sears, G. W., "A Mechanism of Whisker Growth," *Acta Metallurgica*, Vol. 3, 1955, pp. 367–369.

[59] Sears, G. W., "Mercury Whiskers," *Acta Metallurgica*, Vol. 1, 1953, pp. 457–459.

[60] Yan, M., "Self-Assembly of Well-Aligned Gallium-Doped Zinc Oxide Nanorods," *J. Appl. Phys.*, Vol. 94, No. 8, 2003, pp. 5240–5246.

[61] Park, J., et al., "Two-Step Evaporation Process for Formation of Aligned Zinc Oxide Nanowires," *J. Crystal Growth*, Vol. 258, No. 3–4, 2003, pp. 342–348.

[62] Li, Y., et al., "Ordered Semiconductor ZnO Nanowire Arrays and Their Photoluminescence Properties," *Appl. Phys. Lett.*, Vol. 76, No. 15, 2000, pp. 2011–2013.

[63] Vayssieres, L., "Growth of Arrayed Nanorods and Nanowires of ZnO from Aqueous Solution," *Adv. Materials*, Vol. 15, No. 5, 2003, pp. 464–466.

[64] Vayssieres, L., et al., *Chem. Mater.*, Vol. 13, 2001, p. 233.

[65] Govender, K., et al., "Room-Temperature Lasing Observed from ZnO Nanocolumns Grown by Aqueous Solution Deposition," *Adv. Materials*, Vol. 14, No. 17, 2002, pp. 1221–1224.

[66] Smith, D. L., *Thin-Film Deposition: Principles and Practice,* New York: McGraw-Hill, 1995.

[67] Wen, J. G., et al., "Self-Assembly of Semiconducting Oxide Nanowires, Nanorods, and Nanoribbons," *Chem. Phys. Lett.*, Vol. 372, 2003, pp. 717–722.

[68] Yang, P., et al., “Controlled Growth of ZnO Nanowires and Their Optical Properties,” *Adv. Functional Mater.*, Vol. 12, No. 5, 2002, pp. 323–331.

[69] Kim, H., and W. Sigmund, “Zinc Oxide Nanowires on Carbon Nanotubes,” *Appl. Phys. Lett.*, Vol. 81, No. 11, 2002, pp. 2085–2087.

[70] Ng, H. T., et al., “Optical Properties of Single-Crystalline ZnO Nanowires on M-Sapphire,” *Appl. Phys. Lett.*, Vol. 82, No. 13, 2003, pp. 2023–2025.

[71] Wang, Y. W., et al., “Catalytic Growth of Semiconducting Zinc Oxide Nanowires and Their Photoluminescence Properties,” *J. Crystal Growth*, Vol. 234, 2002, pp. 171–175.

[72] Yan, H., et al., “Dendrite Nanowire Ultraviolet Laser Array,” *J. Am. Chem. Soc.*, Vol. 125, No. 16, 2003, pp. 4728–4729.

[73] Li, J. Y., et al., “Fabrication of Zinc Oxide Nanorods,” *J. Crystal Growth*, Vol. 233, 2001, pp. 5–7.

[74] Hu, J. Q., and Y. Bando, “Growth and Optical Properties of Single-Crystal Tubular ZnO Whiskers,” *Appl. Phys. Lett.*, Vol. 82, No. 9, 2003, pp. 1401–1403.

[75] Yuan, H. J., et al., “Characterization of Zinc Oxide Crystal Nanowires Grown by Thermal Evaporation of ZnS Powders,” *Chem. Phys. Lett.*, Vol. 371, 2003, pp. 337–341.

[76] Geng, B. Y., et al., “Synthesis and Optical Properties of S-Doped ZnO Nanowires,” *Appl. Phys. Lett.*, Vol. 82, No. 26, 2003, pp. 4791–4793.

[77] Wu, J., and S. Liu, “Low-Temperature Growth of Well-Aligned ZnO Nanorods by Chemical Vapor Deposition,” *Adv. Mater.*, Vol. 14, No. 3, 2002, pp. 215–218.

[78] Park, W. I., et al., “Metalorganic Vapor-Phase Epitaxial Growth of Vertically Well-Aligned ZnO Nanorods,” *Appl. Phys. Lett.*, Vol. 80, No. 22, 2002, pp. 4232–4234.

[79] Park, W. I., et al., “ZnO Nanoneedles Grown Vertically on Si Substrates by Non-Catalytic Vapor-Phase Epitaxy,” *Adv. Materials*, Vol. 14, No. 24, 2002, pp. 1841–1843.

[80] Muthukumar, S., *IEEE Trans. Nanotechnology*, Vol. 2, No. 1, 2003, p. 50.

[81] Zhong, J., et al., “Ga-Doped ZnO Single-Crystal Nanotips Grown on Fused Silica by Metalorganic Chemical Vapor Deposition,” *Appl. Phys. Lett.*, Vol. 83, No. 16, 2003, pp. 3401–3403.

[82] Ogata, K., et al., “Growth Mode Control of ZnO Toward Nanorod Structures or High-Quality Layered Structures by Metal-Organic Vapor Phase Epitaxy,” *J. Crystal Growth*, Vol. 248, 2003, pp. 25–30.

[83] Ghandhi, S. K., *VLSI Fabrication Principles-Silicon and Gallium Arsenide,* 2nd ed., New York: John Wiley & Sons, 1994.

[84] Sze, S. M., *Semiconductor Devices: Physics and Technology*, 2nd ed., New York: John Wiley & Sons, 2001, p. 354.

[85] Heo, Y. W., et al., “Optical Properties of ZnMgO Nanorods Using Catalyst-Driven Molecular Beam Epitaxial,” *Solid-State Electronics*, June 2003.

[86] Liu, C., et al., “High-Density, Ordered Ultraviolet Light-Emitting ZnO Nanowire Arrays,” *Adv. Materials*, Vol. 15, No. 10, 2003, pp. 838–841.

[87] Wang, Y., I. Leu, and M. Hon, “Preparation of Nanosized ZnO Arrays by Electrophoretic Deposition,” *Electrochemical and Solid-State Letters*, Vol. 5, No. 4, 2002, pp. C53–C55.

[88] Bonnell, D. A., “Materials in Nanotechnology: New Structures, New Properties, New Complexity,” *J. Vacuum Sci. Technol.*, Vol. 21, No. 5, September/October 2003, pp. S194–S206.

[89] Goldstein, J. I., et al., *Scanning Electron Microscopy and X-Ray Microanalysis*, 2nd ed., New York: Plenum Press, 1992.

[90] Xing, Y. J., et al., “Optical Properties of the ZnO Nanotubes Synthesized Via Vapor Phase Growth,” *Appl. Phys. Lett.*, Vol. 83, No. 9, 2003, pp. 1689–1691.

[91] Dai, Y., et al., "Bicrystalline Zinc Oxide Nanowires," *Chem. Phys. Lett.*, Vol. 375, 2003, pp. 96–101.

[92] Novak, D. S., "Scanning Tunneling Microscopy Studies of Oxide Surface Structure," Ph.D. Thesis, Rutgers University, 1996.

[93] Ma, D. D. D., C. S. Lee, and S. T. Lee, *Appl. Phys. Lett.*, Vol. 79, 2001, p. 2468.

[94] Yatsui, T., et al., "Observation of Size-Dependent Features in the Photoluminescence of Zinc Oxide Nanocrystallites by Near-Field Ultraviolet Spectroscopy," *Appl. Phys. Lett.*, Vol. 80, No. 8, 2002, pp. 1444–1446.

[95] Park, W. I., et al., "Excitonic Emissions Observed in ZnO Single Crystal Nanorods," *Appl. Phys. Lett.*, Vol. 82, No. 6, 2003, p. 964.

[96] Zhao, Q. X., et al., "Optical Recombination of ZnO Nanowires Grown on Sapphire and Si Substrates," *Appl. Phys. Lett.*, Vol. 83, No. 1, 2003, p. 165.

[97] Koida, T., et al., "Correlation Between the Photoluminescence Lifetime and Defect Density in Bulk and Epitaxial ZnO," *Appl. Phys. Lett.*, Vol. 82, No. 4, 2003, pp. 532–534.

[98] Ye, J., et al., "Raman and Photoluminescence of ZnO Films Deposited on Si (111) Using Low-Pressure Metalorganic Chemical Vapor Deposition," *J. Vacuum Sci. Technol.*, Vol. 21, No. 4, July/August 2003, pp. 979–982.

[99] Calleja, J. M., and M. Cardona, "Resonant Raman Scattering in ZnO," *Phys. Rev.*, Vol. 16, 1997, p. 3753.

[100] Tsuboi, M., and A. Wada, "Optically Active Lattice Vibrations in Wurtzite-Type Crystals of Zinc Oxide and Cadmium Sulfide," *J. Chem. Phys.*, Vol. 48, 1968, p. 2615.

[101] Johnson, J. C., et al., "Near-Field Imaging of Nonlinear Optical Mixing in Single Zinc Oxide Nanowires," *Nano Lett.*, Vol. 2, No. 4, 2002, pp. 279–283.

[102] Johnson, J. C., et al., "Single Nanowire Laser," *J. Physical Chem. B*, Vol. 105, No. 46, 2001, pp. 11387–11390.

[103] Bagnall, D. M., et al., "Optically Pumped Lasing of ZnO At Room Temperature," *Appl. Phys. Lett.*, Vol. 70, No. 17, 1997, p. 2230.

[104] Liu, Y., et al., "Ultraviolet Detectors Based on Epitaxial ZnO Films Grown by MOCVD," *J. Electron. Mater.*, Vol. 29, No. 1, 2000.

[105] Liang, S., et al., "ZnO Schottky Ultraviolet Photodetectors," *J. Crystal Growth*, Vol. 225, 2001, pp. 110–113.

[106] Kind, H., et al., "Nanowire Ultraviolet Photodetectors and Optical Switches," *Adv. Materials*, Vol. 14, No. 2, 2002, pp. 158–160.

[107] Collins, R. J., and D. G. Thomas, "Photoconduction and Surface Effects with Zinc Oxide Crystals," *Phys. Rev.*, Vol. 112, No. 2, October 1958.

[108] Takahashi, Y., et al., "Photoconductivity of Ultrathin Zinc Oxide Films," *Jpn. J. Appl. Phys.*, Part 1, Vol. 33, No. 12a, December 1994, pp. 6611–6615.

[109] Hove, H. V., and A. Luyckx, "Photoconductivity Decay of ZnO Crystals in Oxygen," *Solid State Commun.*, Vol. 4, 1966, pp. 603–606.

[110] Studenikin, S. A., and M. Cocivera, "Time-Resolved Luminescence and Photoconductivity of Polycrystalline ZnO Films," *J. Appl. Phys.*, Vol. 91, No. 8, April 2002.

[111] Hirata, G. A., et al., *J. Vacuum Sci. Technol.*, Vol. 14, 1996, p. 791.

[112] Lee, C. J., et al., *Appl. Phys. Lett.*, Vol. 81, No. 19, 2002, p. 3648.

[113] Zhu, Y. W., et al., *Appl. Phys. Lett.*, Vol. 83, No. 1, 2003, p. 144.

[114] Dong, L., et al., *Appl. Phys. Lett.*, Vol. 82, 2003, p. 1096.

[115] Arnold, M. S., et al., "Field-Effect Transistors Based on Single Semiconducting Oxide Nanobelts," *J. Phys. Chem. B*, Vol. 107, No. 3, 2003, pp. 659–663.

CHAPTER 7

Antimony-Based Materials for Electro-Optics

Wei Zhang and M. Razeghi, Center for Quantum Devices, Northwestern University

7.1 Introduction

7.1.1 Antimony

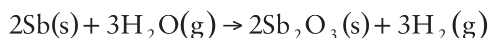
Antimony (Sb) is a semimetal and has an atomic number of 51 and atomic weight of 121.760. It is located in group 15 (V-A) and period 5 in the chemical periodic table. Antimony is solid at room temperature and standard state. In its elemental form it is a silvery white, brittle crystalline solid. It is not acted on by air at room temperature, but burns brilliantly when heated and forms white fumes. It is a poor conductor of heat and electricity. The physical data for antimony are summarized in Table 7.1.

Antimony has been known and used since ancient times. The name *antimony* comes from the Greek words *anti* + *monos*, meaning "not alone," because it is often found in ores such as stibnite and valentines. The origin of the symbol Sb comes from the Latin word *stibium*, meaning "mark."

Upon heating, antimony reacts with oxygen in air to form the trioxide antimony(III) oxide, Sb_2O_3:

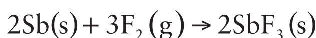
$$4Sb(s) + 3O_2(g) \rightarrow 2Sb_2O_3(s)$$

At red heat, antimony reacts with water to form the trioxide antimony(III) oxide, Sb_2O_3. Antimony reacts more slowly at ambient temperatures.

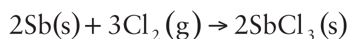


$$2Sb(s) + 3H_2O(g) \rightarrow 2Sb_2O_3(s) + 3H_2(g)$$

Antimony reacts under controlled conditions with the halogens fluorine (F_2), chlorine (Cl_2), bromine (Br_2), and iodine (I_2) to form the respective trihalides antimony(III) fluoride (SbF_3), antimony(III) chloride ($SbCl_3$), antimony(III) bromide ($SbBr_3$), and antimony(III) iodide (SbI_3):



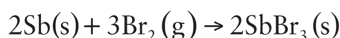
$$2Sb(s) + 3F_2(g) \rightarrow 2SbF_3(s)$$



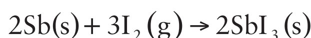
$$2Sb(s) + 3Cl_2(g) \rightarrow 2SbCl_3(s)$$

Table 7.1 Summary of Physical Data for Antimony

	Symbol	*Sb*
	Atomic number	51
	Atomic weight	121.760 amu
Steric	Atomic radius	1.53Å
	Ionic radius	0.76Å
	Covalent radius	1.40Å
	Atomic volume	18.23 cm^3/mol
	Density (293K)	6.684 g/cm^3
	Crystal structure	Rhombohedral
Electronic	Shells	2,8,18,18,5
	Orbitals	[Kr] $4d^{10}$ $5s^2$ $5p^3$
	Electronegativity	1.8, 1.9
	Oxidation states	5, 3, –3
	Electrical resistivity	4×10^{-7} Ωm
Thermal	Melting point	630.74°C
	Boiling point	1,587°C
	Specific heat	0.21 J/g K
	Heat of fusion	19.870 kJ/mol
	Heat of vaporization	77.140 kJ/mol
	Thermal conductivity	0.243 W/cm K



$$2Sb(s) + 3Br_2(g) \rightarrow 2SbBr_3(s)$$



$$2Sb(s) + 3I_2(g) \rightarrow 2SbI_3(s)$$

Antimony is alloyed with lead to increase lead's durability. Antimony alloys are also used in batteries, low-friction metals, type metal, and cable sheathing, among other products. Antimony compounds are used to make flame-proofing materials, paints, ceramic enamels, glass, and pottery. Some over-the-counter cold and flu remedies use antimony compounds. One of the most important applications of antimony is in semiconductor technology for making infrared detectors, diodes, and Hall effect devices.

7.1.2 Sb-Based III-V Semiconductor Alloys

When antimony combines with group III, it forms III-V binary compounds such as InSb, GaSb, and AlSb. Table 7.2 lists the fundamental physical parameters of these III-Sb binaries as well as several others such as GaAs, InAs, and InP at 300K. These binary compounds are the basis for more complex Sb-based ternary or quaternary compounds.

When one additional element is added into a binary compound and distributed randomly in the crystal lattice, $III_{1-x}III_xV$ or $IIIV_yV_{1-y}$ ternary alloys can be formed, where x and y are indices with values between 0 and 1. The lattice parameter and the bandgap energy of the alloy can be modified by varying the value of x.

Table 7.2 Physical Properties of Some III-V Binary Compounds at 300K

	InSb	*GaSb*	*AlSb*	*InAs*	*GaAs*	*InP*
Lattice parameter (Å)	6.47937	6.09593	6.1335	6.0584	5.65321	5.86875
Bandgap energy (eV)	0.17	0.73	1.58	0.36	1.424	1.35
Refractive index	4.0	3.820	3.4	3.520	3.655	3.450
Effective mass: (m_e/m_0)	0.0145	0.044	0.39	0.022	0.065	0.078
Effective mass: (m_{hh}/m_0)	0.44	0.33	0.5	0.41	0.45	0.8
Effective mass: (m_{lh}/m_0)	0.016	0.056	0.11	0.025	0.082	0.012
Dielectric constant $(\varepsilon/\varepsilon_0)$	17.7	15.7	14.4	14.6	13.1	0.012
Electron affinity χ (eV)	4.69	4.03	3.64	4.45	4.5	12.4

The bandgap energy $E_g(x)$ of a ternary compound varies with the composition x as follows:

$$E_g(x) = E_g(0) + bx + cx^2 \tag{7.1}$$

where $E_g(0)$ is the bandgap energy of the lower bandgap binary compound and c is called the bowing parameter. Bowing parameter c can be theoretically determined [1]. The compositional dependence of the bandgap energy of Sb-based ternaries at 300K is given in Table 7.3 [2].

The lattice parameter a of ternary compounds obeys Vegard's law. According to Vegard's law the lattice parameter of the ternary alloys can be expressed as follows:

$$a_{\text{alloy}} = xa_{\text{A}} + (1 - x)a_{\text{B}} \tag{7.2}$$

where a_{A} and a_{B} are the lattice constants of the binary alloys A and B.

Similarly, Sb-based quaternary compounds can be formed when antimony and three other group III or group V elements are distributed uniformly in the crystal lattice. Compared with ternary compounds, quaternary compounds have the freedom to adjust the lattice constant and the bandgap energy independently by varying the composition. So it's possible to obtain the desired bandgap energy and still maintain the lattice match to the substrate.

Table 7.3 Compositional Dependence of the Bandgap Energy of Sb-Based Ternary Solid Solutions at 300K

Ternary	*Direct Bandgap Energy (eV)*
$Al_xGa_{1-x}Sb$	$E_g(x) = 0.726 + 1.139x + 0.368x^2$
$Al_xIn_{1-x}Sb$	$E_g(x) = 0.172 + 1.621x + 0.43x^2$
$Ga_xIn_{1-x}Sb$	$E_g(x) = 0.172 + 0.139x + 0.145x^2$
$GaAs_xSb_{1-x}$	$E_g(x) = 0.726 - 0.502x + 1.2x^2$
$InAs_xSb_{1-x}$	$E_g(x) = 0.18 - 0.41x + 0.58x^2$

Source: [2].

Ilegems and Panish [3] calculated quaternary phase diagrams by decomposing the solid into ternary alloys: ABC, ACD, ABD, and BCD (where A and B are group III elements and C and D are group V elements). Jordan and Ilegems [4] obtained equivalent formulations by considering the solid as a mixture of binary alloys: AC, AD, BC, and BD. Assuming a linear dependence of lattice parameter on the composition of the binaries, the lattice parameter of the alloy $A_xB_{1-x}C_yD_{1-y}$ can be expressed as:

$$A_{\mathrm{A}x\mathrm{B}\,1-x\mathrm{C}y\mathrm{D}\,1-y} = xya_{\mathrm{AC}} + x(1-y)a_{\mathrm{AD}} + (1-x)ya_{\mathrm{BC}} + (1-x)(1-y)A_{\mathrm{BD}} \qquad (7.3)$$

Similarly, if bowing parameter c is neglected, the bandgap energy can be approximated from the binaries by assuming a linear dependence on the composition:

$$E_{g\mathrm{A}x\mathrm{B}\,1-x\mathrm{C}y\mathrm{D}\,1-y} = xyE_{\mathrm{AC}} + x(1-y)E_{\mathrm{AD}} + (1-x)yE_{\mathrm{BC}} + (1-x)(1-y)E_{\mathrm{BD}} \qquad (7.4)$$

The Sb-based quaternary V-III alloys that can be used for multiplayer heterostructures are listed in Table 7.4 together with the binary compounds to which they are lattice matched.

7.1.3 Bulk Single-Crystal Growth

Most of the Sb-based semiconductor materials are grown on substrates such as InP, GaAs, InAs, GaSb, or InSb. The substrate is made entirely of a single material and one crucial feature of the substrate is that it is one single crystal across its entirety without grain boundaries. For the single-crystal growth of III-V compounds, several methods have been developed, among which the Czochralski, Bridgman, and vertical gradient freeze methods are most popularly used.

7.1.4 Applications

Techniques for IR detection and emission have advanced a long way since IR radiation was first discovered by William Herschel in 1800. Because objects at

Table 7.4 Binary to Sb-Based Quaternary III-V Lattice-Matched System

Quaternary	*Lattice-Matched Binary*	*Wavelength (m)*
$Al_xGa_{1-x}As_ySb_{1-y}$	InP	1
$Al_xGa_{1-x}As_ySb_{1-y}$	InAs	3
$Al_xGa_{1-x}As_ySb_{1-y}$	GaSb	1.7
$Ga_xIn_{1-x}P_ySb_{1-y}$	InP, GaSb, AlSb	2
$Ga_xIn_{1-x}As_ySb_{1-y}$	InP, InAs, GaSb	1.6~4.1
$In(P_xAs_{1-x})_ySb_{1-y}$	AlSb, GaSb, InAs	2~4
$(Al_xGa_{1-x})_yIn_{1-y}Sb$	AlSb	1.1~2.1

Source: [1, 5].

temperatures between 0 and 1,000°C emit most of their energy as radiation with an infrared wavelength of $1 < \lambda < 20\ \mu$m and the atmosphere has a transmission window in the 3- to 5-μm (MWIR) and 8- to 14-μm (LWIR) regions (Figure 7.1), IR detectors and lasers at these wavelength are of great importance for various applications. On the other hand, the 5- to 8-μm region also has applications in the proximity fuzes where high atmospheric absorption is important to detect only those targets in proximity. IR detectors in the *very long wavelength range* (VLWIR) have applications such as space-based astronomy and early detection of long-range missiles.

HgCdTe (MCT) is a well-established material system that has been the dominant system for MWIR and LWIR photodetectors. However, MCT suffers from instability and nonuniformity problems over a larger area due to the high Hg vapor pressure. This has intensified the search for an alternative IR material system. Sb-based materials including III-Sb binary and its ternary and quaternary alloys are suitable candidates for applications in the IR range. Figure 7.2 shows the bandgap versus lattice constant diagram of III-V alloys. InSb has a bandgap energy of 0.18 eV at room temperature and 0.23 eV at 77K, enabling it to cover the entire MWIR region up to 5.5 μm. For longer wavelengths, arsenic can be added into InSb to form InAsSb, which shows reduced bandgap energy. As an alternative to the InAsSb system, InTlSb is also a potential material for $\lambda > 8\ \mu$m detectors. Besides adding As or Tl to an InSb lattice, it is also possible to substitute Bi. As a result, a rapid reduction of the bandgap is expected. However, the large miscibility gap limits Tl and Bi incorporation to high composition and thus further reduction of bandgap. To exceed the limitations imposed by the miscibility gap, arsenic can be added to

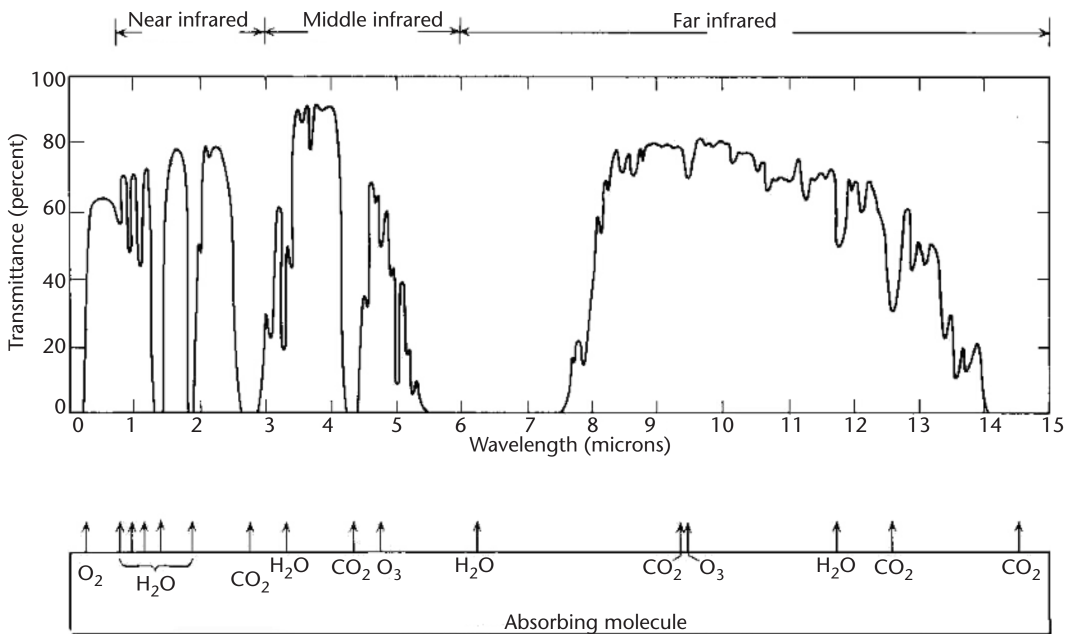


Figure 7.1 Transmittance of the atmosphere for a 6,000-ft horizontal path at sea level.

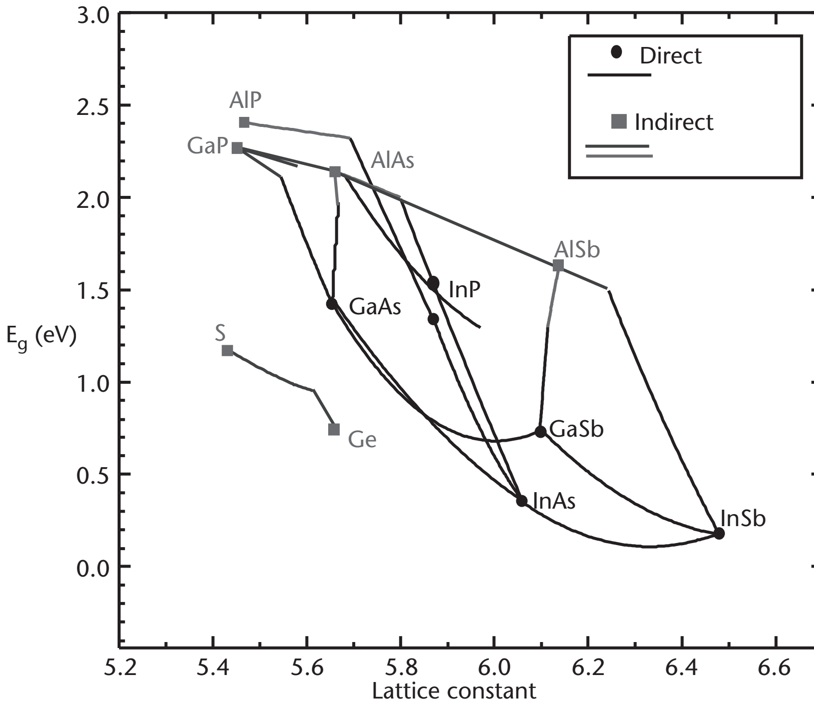


Figure 7.2 The bandgap versus lattice constant diagram of III-V alloys.

InTlSb to form a quaternary material with an extended photoresponse cutoff wavelength.

The development of a laser source for the MWIR (3- to 5-μm) region has been particularly difficult because of lack of suitable high-quality heterostructure semiconductor materials. InAsSb/InAsSbP is one of the new material systems that has a good carrier and optical confinement with good lattice matching, which make it very promising for laser diode fabrication. *Double heterostructure* (DH) MQW, and superlattice structure lasers using InAsSbP alloy system can cover the 3- to 5-μm region by adjusting its alloy composition.

Recently, type II superlattices have been proposed as an alternative to HgCdTe for the strategic windows of the 3- to 5-μm and 8- to 12-μm regions as well as longer wavelengths. Unlike HgCdTe, type II superlattices are constructed from the III-V material system, and hence they have much better mechanical properties and material uniformity. The electron effective mass in these superlattices is higher than the electron effective mass in HgCdTe and InAsSb, therefore the tunneling current is less. The bandgap of the superlattice can be changed to cover a wide IR range (from ~2 μm to above 2 μm) by changing the thickness of the layers rather than the material composition. Also, in comparison to QWIPs, type II superlattices have much higher quantum efficiency, and they do not require surface gratings for absorption of normal incident light.

7.2 III-Sb Binary Compounds: GaSb, AlSb, and InSb

7.2.1 GaSb

GaSb is an intermediate-gap semiconductor, since its gap of 0.8 eV is neither as wide as that of GaAs and InP nor as narrow as that of InAs and InSb. It is particularly interesting as a substrate material because it is lattice matched to solid solutions of various ternary and quaternary III-V compounds whose bandgaps cover a wide spectral range from 0.3 to 1.58 eV (i.e., 0.8–4.3 μm), as has been shown in Figure 7.2. GaSb, AlSb, and InAs and their related compounds form a nearly lattice-matched family of semiconductors known as the 6.1Å family, because the lattice constants of these material are about a = 6.1Å. The physical properties of GaSb are summarized in Table 7.5.

Growth of GaSb

In 1926, Goldschmidt synthesized GaSb and determined its lattice constant [6]. The phase diagram and the calculated solidus of GaSb are shown in Figure 7.3. Bulk GaSb crystals have been mainly grown by the Czochralski (CZ) technique. There are a few reports on the Bridgman (BG) technique, and the traveling heater method

Table 7.5 Physical Properties of GaSb

Property	Value
Lattice constant (Å)	6.094
Density (g/cm^3)	5.61
Melting point (K)	985
Expansion coefficient	6.2×10^{-6}
Thermal conductivity(WK^{-1}cm^{-1})	0.4
Energy gap (eV)	0.725 at 300K 0.822 at 0K
Spin-orbit splitting energy, Δ_0 (eV)	0.76
Electron mobility (cm^2/V·s)	5,000 at 300K
Hole mobility (cm^2/V·s)	880 at 300K 2,400 at 77K
Electron effective mass	$0.042m_0$
Hole effective mass	
Heavy hole mass	0.28 m_0
Light hole mass	0.05 m_0
Spin-orbit split mass	0.13 m_0
LO phonon energy at zone center	28.8 meV
TO phonon energy at zone center	27.7 meV
Refractive index	3.84 at 10 μm 3.79 at 2 μm 3.92at 1.55 μm
Dielectric constant	
ε_0	15.69
ε_∞	14.44

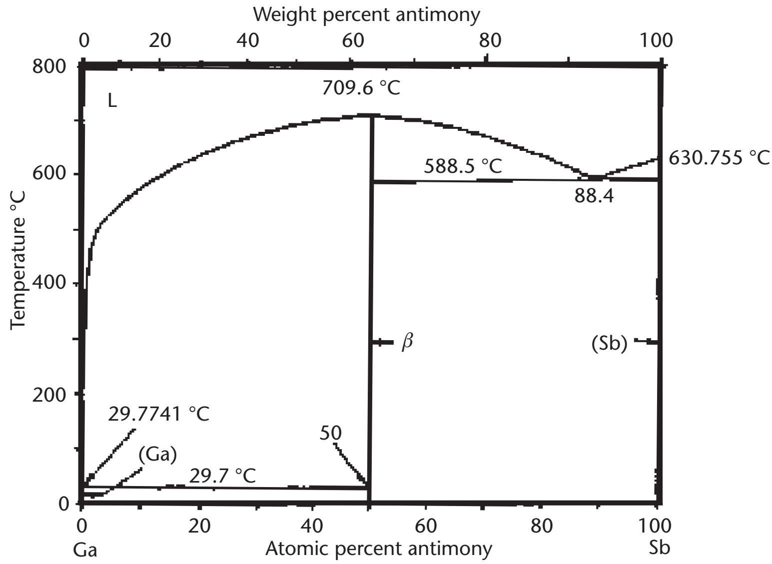


Figure 7.3 Phase diagram of GaSb.

(THM), and the vertical gradient freeze (VGG) technique. Epitaxial growth of GaSb has been largely carried out by LPE. A few reports exist on VPE, MOCVD, and MBE. The details are discussed later.

Liquid Phase Epitaxy

First, growth of GaSb by the LPE technique has been carried out [7]. Ga-, Sb-, Sn- and Bi-rich melts have been used for the growth in the temperature range of 330°C to 680°C. By carrying out growth at low temperatures, the native acceptor concentration could be reduced to a level of 10^{16} cm^{-3}. The Sb-rich melt is the most effective solution to reduce the native acceptor concentration; however, growth from Sb-rich solutions is limited by the eutectic point in the Ga–Sb phase diagram at T = 588.5°C for 0.884 atom fraction of Sb. Woelk and Benz [8] have grown undoped p epilayers from Ga- and Sb-rich solutions in the temperature range of 330°C to 470°C and 635°C to 680°C, respectively.

In recent studies, it has been found that the growth temperature range of 500°C to 550°C with Ga melt is optimal for obtaining high-quality layers with excellent surface morphology [9]. The epilayers grown at low temperatures produce low native defect concentration but they exhibit poor morphology and hence are not suitable for device applications. Similarly, the Sb-rich melt, even though it efficiently reduces the native defect content, results in faceted growth, which is again undesirable for device applications. The luminescence efficiency for layers grown from Sb-rich solutions was less compared to those grown from Ga-rich solutions. This has been attributed to higher contamination of impurities, which is expected for high growth rates in the case of Sb-rich solutions. By carrying out growth from a two-

phase solution technique, abrupt *p-n* junctions with epilayer thicknesses of about 1 μm have been obtained.

Te has been used for *n*-type doping. Capasso and coworkers [10] have achieved very low net donor concentration in the range of 10^{14}–10^{15} cm^{-3} from undoped Ga-rich solutions in the 300°C to 375°C range and by Te compensation using Ga-rich melt at 500°C. Miki and coworkers [11] grew undoped *n*-type layers at 400°C from Ga-rich melt with net donor concentration of ~10^{15} cm^{-3} and mobility as high as 7,700 cm^2/V s.

MOCVD

MOCVD growth of GaSb has been carried out [12–15] using *trimethylgallium* (TMGa) or TEGa and Sb-containing metalorganic precursors in the temperature range of 450°C to 625°C. From these studies, *trisdimethylaminoantimony* (TDMASb) has been found to be an excellent precursor for MOVPE growth over a wide range of growth temperatures [16]. The main source of background impurities during MOCVD growth comes from the carbon during pyrolysis of TMGa or TEGa, yielding a high concentration of unintentional acceptors. TESb and TMSb decompose insufficiently at low temperatures and need to be thermally precracked at high temperatures. However, in spite of the residual carbon contamination, the dominant acceptors have been found to be the native defects. The TDMASb is expected to produce low carbon contamination due to the absence of direct Sb–C bonds and effective decomposition at low temperatures (~300°C).

MBE

The main difficulty in growing GaSb by MBE is the low vapor pressure of antimony [17]. As a result, during crystal growth Sb will have a low surface mobility and tend to aggregate together forming clusters and precipitates. This leads to vacant Sb sites. Thus, antisite defects like Ga_{Sb} are formed. Therefore, to improve the quality of MBE-grown layers, an Sb-rich environment is needed. One can achieve this by using proper orientation of the substrate such (311)B, (111)B, and so on. Longenbach and Wang [18] used (311)B-oriented substrates to reduce the native *p*-type centers in the grown layers. Usually the growth rate varied from 0.6 to 2.5 μm/hr for the growth temperatures in the range of 500°C to 600°C. A very low acceptor concentration of ~10^{15} cm^{-3} could be obtained using this approach. Undoped GaSb epilayers have shown C, O, and Si impurities. The origin of these impurities can vary from one growth system to another and can be due to different sources.

Generally, Te doping is accomplished in MBE-grown layers by the use of a PbTe cell since this provides better control of temperature than the use of elemental Te [19]. There appears to be a trace incorporation of Pb in the grown layer. Donor concentrations of above 10^{18} cm^{-3} are readily obtained. Other compounds that may be used for Te doping are GaTe, SbTe, GeTe, and SnTe. The *p*-type conductivity may be accomplished by Ge, Sn, or Be doping in the MBE process [20]. The covalent radius of Sb (1.41Å) is larger than that of Ga (1.25Å) and therefore group IV atoms tend to occupy the Sb sites. Sn (1.40Å) is a larger atom than Ge (1.22Å) and,

consequently, it is more difficult to achieve heavy doping with Sn. Silicon, because it is a small atom, gets distributed among the two sites and the compensation ratio is high. Hence, Si is not generally used as a *p*-type dopant.

Properties of GaSb

Structure Properties

At room temperature, the lattice parameter is 6.09593Å. The density of GaSb at 300K is measured to be 5.6137 g cm^{-3}. GaSb crystallizes in a zinc-blende structure. The (100) surfaces are stepped and contain both Ga and Sb atoms. The nature of chemical bonds in III-V compounds is of the mixed covalent-ionic type. The ionicity of GaSb is 0.33. The presence of a slight ionic component in the bonds and the fact that there are equal numbers of Ga and Sb atoms on the (111) planes results in the (110) cleavage of the compound. The zinc-blende lattice structure and partial ionic bonding impart to the crystal a polarity along the *o*-axis. The (111) planes can be prepared with either Ga or Sb atoms on the surface. The (111) plane composed of Ga atoms is designated (111)A. The (–1, –1, –1) plane is composed of Sb atoms and is designated (111)B. These two surfaces exhibit striking differences in their chemical, electrical, and mechanical properties.

Electronic Properties

Figure 7.4 shows the band structure of GaSb obtained at 300K. The lowest minimum of the conduction band is at the Γ point. The next higher minima are the *L* points at the surface of the Brillouin zone and the *X* points. Owing to the small Γ – *L* energy separation in GaSb and the much lower density of states at the Γ minimum, at

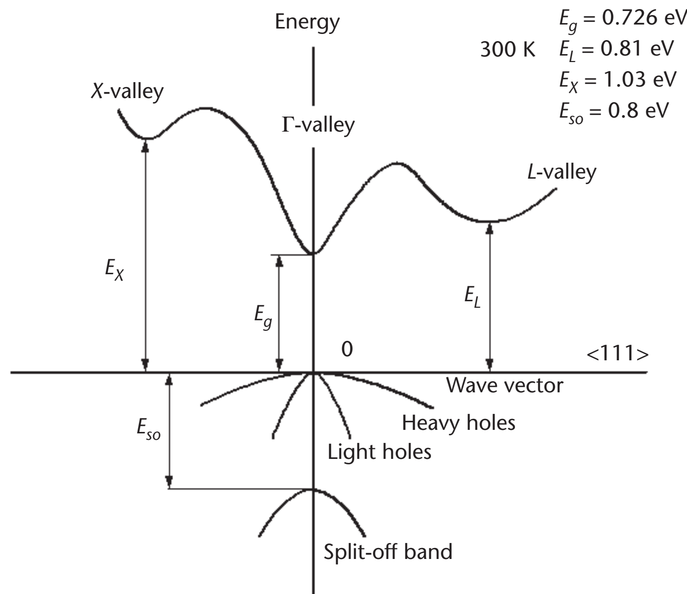


Figure 7.4 Band structure of GaSb obtained at 300K.

room temperature a significant fraction of the electrons occupy L-valley states. Temperature dependence of the energy gap is

$$E_g = 0.813 - 3.78 \cdot 10^{-4} \cdot T^2 (T = 94)(\text{eV}) \tag{7.5}$$

where T is temperature in degrees Kelvin ($0 < T < 300$).

Optical Properties

The dielectric constants and ε_0 and ε_8 were found to be 15.69 and 14.44, respectively, at 300K using the reflectance technique and oscillator fit. The real and imaginary parts of the dielectric constant measured by ellipsometric technique are shown in Figure 7.5. The values for refractive index n, extinction coefficient k, and reflectance R calculated from these data are given in Table 7.6 for various wavelengths.

Johnson and coworkers have studied absorption, PL, and injection luminescence as a function of temperature [21]. The band–band peak increases with the free hole concentration. In p-type samples, compensated with Se or Te, several new bands appear, some of which are characteristic of pure Se or Te. At high donor doping levels the band–band peak shifts higher energies due to the Burstein-Moss effect.

7.2.2 AlSb

AlSb is an indirect-gap semiconductor with a bandgap of 1.69 eV at room temperature. It has a lattice constant of 6.136Å, which is nearly lattice matched to GaSb and InAs. Heterostructures composed of AlSb, GaSb, and InAs are of interest for several electronic and optoelectronic device applications. For example, AlSb can serve as a

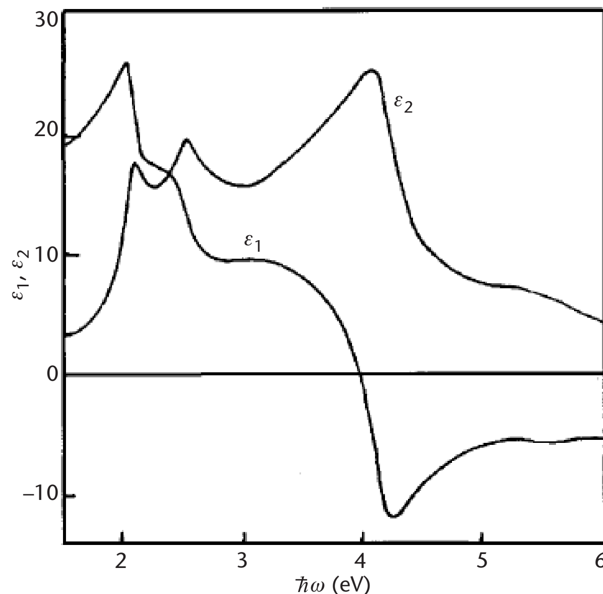


Figure 7.5 Real and imaginary parts of the dielectric constants versus photon energy.

Table 7.6 Optical Constants of GaSb

Energy (eV)	ε_1	ε_2	*n*	*k*	*R*
1.5	19.135	3.023	4.388	0.344	0.398
2.0	25.545	14.442	5.239	1.378	0.487
2.5	13.367	19.705	4.312	2.285	0.484
3.0	9.479	15.738	3.832	2.109	0.444
3.5	7.852	19.267	3.785	2.545	0.485
4.0	21.374	25.138	3.450	3.643	0.583
4.5	28.989	10.763	1.586	3.392	0.651
5.0	25.693	7.529	1.369	2.751	0.585
5.5	25.527	6.410	1.212	2.645	0.592
6.0	24.962	4.520	0.935	2.416	0.610

barrier material to confine electrons in InAs-channel high-electron-mobility transistors [21] or magnetoelectronic hybrid Hall effect devices [22]. AlSb layers also can function as tunneling barriers in resonant tunneling diodes [23] and as barriers in "W" IR lasers (i.e., heterojunction lasers with type II band offsets) with InAs/InGaSb/InAs active regions [24].

Properties of AlSb

Compared with other III-Sb compounds, little work has been done on the investigation of the AlSb due to the facts that large, high-quality single crystals are rare and that the surfaces of this materials react rapidly with air. Some of the important material properties of AlSb are listed in Table 7.7.

Electronic Properties

The direct gap in AlSb was measured using modulation spectroscopy by Alibert et al. [25] and spectroscopic ellipsometry by Zollner et al. [26]. Bulk AlSb samples were found to exhibit a gap of 2.35 to 2.39 eV at liquid-helium temperature, and a *T* dependence similar to that in GaSb. The conduction-band minima ordering is believed to be the same as in GaP and AlAs: $X - L - \Gamma$, with the *L* valley only 60 to 90 meV above the Γ valley.

Te has been investigated as an *n*-type dopant in AlSb [27]. Be has been used as a *p*-type dopant in AlSb for InAs-channel high-electron-mobility transistors [28] and InAs/AlSb superlattices [29]. Thick epitaxial layers of AlSb (Si) and AlSb (Be) were grown by MBE and characterized by variable-temperature Hall measurements [30]. Si is shown to be predominantly an acceptor in AlSb, with an energy level of 33 meV above the top of the valence band. Be is also an acceptor, with an energy level of 38 meV above the top of the valence band. It was demonstrated that Be is a robust doping source for *p*-AlSb for carrier densities ranging from 10^{15} to 10^{19} cm^{-3}.

Optical Properties

Zollner et al. [26] have studied optical properties of AlSb by spectroscopic ellipsometry. They reported room temperature optical constants of this material for

Table 7.7 Physical Properties of AlSb

Lattice constant (Å)	6.136
Density (g/cm^3)	4.26
Melting point (K)	1,338
Expansion coefficient	4×10^{-6}
Thermal conductivity(WK^{-1} cm^{-1})	~0.7
Energy gap (eV)	1.63 at 300K 1.666 at77K
Electron mobility ($cm^2/V\cdot s$)	200 at 300K
Hole mobility ($cm^2/V\cdot s$)	375 at 300K
Electron effective mass	0.12 m_0
Hole effective mass	0.98 m_0
LO phonon energy at zone center	42.1 meV
TO phonon energy at zone center	39.5 meV
Refractive index	2.995 at 20 μm 2.08 at 15 μm 3.1 at 10 μm 3.3 at 2 μm
Dielectric constant (static)	12.04

energies E from 1.4 to 5.8 eV. Adachi [31] calculated the optical dispersion of AlSb and compared it with the experimental results. Figure 7.6 shows the numerically calculated spectral dependence of the refractive index (n) and the extinction

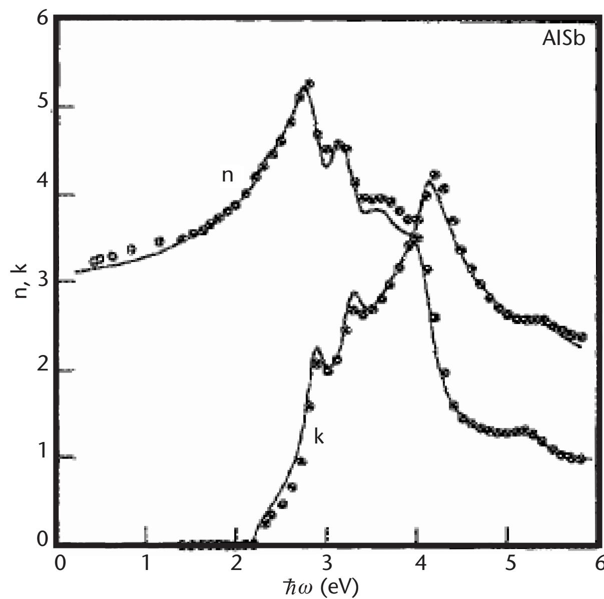


Figure 7.6 Numerically calculated spectral dependence of n and k for AlSb. The open and solid circles are the experimental data.

coefficient (k) for AlSb. The circles are experimental data from Zollner's work. The absorption coefficient $\alpha(\omega)$ has also been calculated and is shown in Figure 7.7.

7.2.3 InSb

InSb is a direct small energy gap semiconductor whose semiconductor properties were first reported by H. Welker in 1952 [32]. It was soon discovered that InSb had the smallest energy gap among any semiconductors known at that time and naturally its application as an IR detector became obvious. The bandgap energy of 0.18 eV at room temperature indicated that it would have a long-wavelength limit of approximately 7 μm and when cooled with liquid nitrogen the gap increased to 0.23 eV, enabling it to cover the entire MWIR up to 5.5 μm. Furthermore, it is a relatively robust material with high quantum absorption efficiency and sensitivity in the MWIR. The interest in InSb stemmed not only from its small energy gap but also from the fact that it could be prepared in high-quality, single-crystal form. As shown in the phase diagram of Figure 7.8 [33], InSb exhibits a congruent melting point at 525.7°C, allowing it to be grown from the melt by the conventional technique. Today, 2- and 3-inch InSb substrates are commercially available.

The material parameters and characteristics of bulk InSb have been investigated and refined over the years. The experimental bandgap, E_0, and intrinsic carrier concentration n_i, have been measured as a function of temperature and the following expressions [34] have been found to agree well with the experimental data:

$$E_0(T) = \left[0.235 - \frac{3.2 \times 10^{-4} T^2}{(220 + T)}\right] (\text{eV}) \tag{7.6}$$

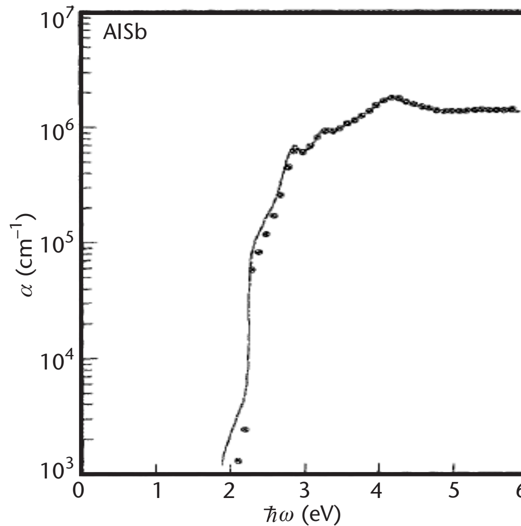


Figure 7.7 Calculated absorption coefficient α as a function of energy.

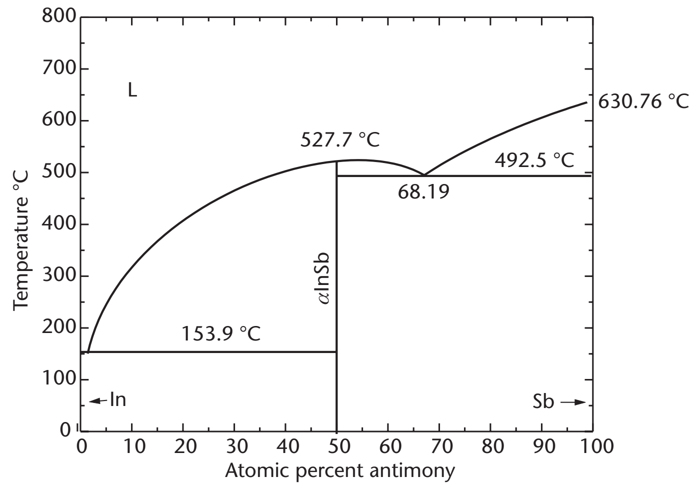


Figure 7.8 Temperature-composition binary phase diagram of InSb, describing the equilibrium of solid with a liquid InSb.

$$n_i = 5.63 \times 10^{14} T^{\frac{3}{2}} \exp(-0.127/k_B T)\left(\text{cm}^{-3}\right) \tag{7.7}$$

Other important parameters such as lattice constant, electron and hole masses, refractive index, and thermal expansion coefficient are summarized in Table 7.8 [35].

InSb has been utilized for many purposes due to the properties just described. Because of the low electron effective mass and the resulting large electron mobility, InSb can be used for very high speed electronic devices and is being used in magneto-resistive sensors in the automotive industry. High-performance InSb detectors have been fabricated with this material for decades [36, 37]. The InSb detector arrays are *p-n* junctions made mostly by implanting Be^+ into *n*-type InSb substrates. Such processes yield high-quality *p-n* junctions with low leakage currents and high breakdown voltage, which are desirable characteristics for detector arrays. Surface passivation is carried out to minimize the surface leakage current, which can be significant in small-sized detectors (pixels) in high spatial resolution arrays.

A fully monolithic structure, in which all optical detection and signal processing functions are performed on the same material, is the most ideal architecture. Such monolithic structures have been reported using InSb metal-oxide semiconductor FETs [38] but the low breakdown voltage in narrow bandgap material has led to low charge handling capabilities. Additionally, serious problems involving noise and charge trapping in narrow-gap semiconductors reduce its charge transfer efficiency. These disadvantages have limited the development of a fully monolithic structure using narrow-gap material.

An alternative method is the hybrid integration approach in which the photon detection process is performed by the narrow bandgap detector array and the signal is transferred to a Si-based charge storage and multiplexed readout. Indium bump is normally used as the interconnect between the optical and the electrical components. Current manufacturing device processes require that bulk materials be

Table 7.8 Physical Properties of InSb

	T (K)	*InSb*
Crystal structure		Cubic (ZnS)
Lattice constant (Å)	300	6.47877
Coefficient of thermal expansion (10^{-6} K^{-1})	300	5.04
	80	6.50
Density γ (g/cm^3)	300	5.7751
Melting point (K)		803
Energy gap (eV)	4.2	0.2357
	80	0.228
	300	0.180
Thermal coefficient of E_g	100–300	-2.8×10^4
m_e^*/m_0	4.2	0.0145
	300	0.0116
m_{lh}^*/m_0	4.2	0.0149
m_{hh}^*/m_0	4.2	0.41
Momentum matrix element P ($cm^2/V \cdot s$)		9.4×10^{-8}
Electron mobility ($cm^2/V \cdot s$)	300	8×10^4
	77	10^6
Hole mobility ($cm^2/V \cdot s$)	300	800
	77	10^4
Intrinsic carrier concentration (cm^{-3})	77	2.6×10^9
	200	9.1×10^{14}
	300	1.9×10^{16}
Refractive index		3.96
Static dielectric constant		17.9
Optical phonon (cm^{-1})		
LO		193
TO		185

thinned before or after hybridization to achieve backside illumination with reasonable quantum efficiency and minimum crosstalk. The advantage of the hybrid architecture is that the performances of two different devices can be separately optimized. However, this structure suffers from the difference in the thermal expansion coefficients between the detector and the multiplexer materials. In addition, the substrate thinning process is a very delicate process that can lower the yield and the reproducibility of the detectors.

The best solution is the pseudomonolithic integration architecture in which InSb detectors can be integrated with GaAs- or Si-based electronic readouts on a common

substrate. This would require the ability to grow and fabricate InSb detectors on GaAs or Si. Advantages of such integration are threefold. First, back-illumination detector structures will not suffer from having to thin the substrate since it is transparent in the 3- to 5-μm range. Second, the integration eliminates the indium bump process. Third, a larger array scan be realized because larger (greater than 3-inch InSb substrates that are available) substrates are available in GaAs or Si. Therefore, successful pseudomonolithic integration will permit the production of larger reliable focal plane arrays.

Growth of InSb

MOCVD

The growth of InSb was performed on InSb-, GaAs-, and GaAs-coated silicon substrate with low-pressure MOCVD. TMIn and TMSb were used as precursors. The growth conditions were studied using TMIn and TMSb by varying the growth temperature and V/III ratio. Table 7.9 summarizes the growth parameters and conditions. The low melting temperature of InSb (525°C) requires a low growth temperature, typically no greater than 475°C, to prevent surface deterioration. At these growth temperatures, TMIn and TMSb decompose at a reasonable rate. The growth rate of InSb films was approximately 1 μm per hour.

For InSb growth using TMIn, the surface morphology was found to be very sensitive to the V/III ratio and closely reflected the crystallinity of the as-grown films. Based on the morphology, only a narrow range of V/III ratio around 11.5 at 465°C was found to be optimum. At a lower V/III ratio of around 9, the morphology degraded and the surface was covered with indium droplets. With an increase in the V/III ratio, the droplets were replaced by hillocks whose density gradually decreased as they approached the optimum value. At a higher V/III ratio than the optimum, the morphology deteriorates with Sb-related hillocks. The observed surface dependence on the V/III ratio is related to the low vapor pressure of In and Sb. Therefore, excess In can easily deposit on the surface, leading to VLS three-phase growth [39, 40]. On the other hand, excess Sb forms a solid second phase, which leads to the growth of hillocks. Because Sb has a melting point of 630°C, VLS growth does not occur.

Table 7.9 Growth Parameters and Conditions Used for the Deposition of Normally Undoped InSb Epilayers

Growth temperature		425–475°C
Growth pressure		76 Torr
Substrate orientation		InSb-, GaAs-, GaAs-coated Si
TMIn	Temperature	18°C
	Flow rate	
TMSb	Temperature	0°C
	Flow rate	16–36 sccm
H_2	Flow rate	1.5 L/min

Similar studies of the V/III ratio dependence on the morphology have also been carried out at various temperatures. Note that the optimum V/III ratio increases as growth temperature decreases. This is due to incomplete decomposition of TMSb. While TMIn completely decomposes around 425°C [41], TMSb only starts decomposing at 400°C and decomposition efficiency increases as the temperature increases [42]. Therefore, a higher flow rate for TMSb is required at lower temperature. However, growth carried out at a temperature close to the melting point should be avoided since poor-quality films are obtained from thermal degradation. Based on the results given earlier, the optimum growth parameters have been determined and summarized in Table 7.10.

MBE

InSb on (100) GaAs The growth of InSb was optimized using an EPI Modular Gen II Solid source MBE system using an elemental uncracked source. Group III materials were 6N (six nines: 99.9999%) pure and the group V materials were 7N pure. The system is equipped with a *reflection high-energy electron diffraction* (RHEED) system for calibrating growth rate and monitoring the growth surface as well as a mass spectrometer for monitoring molecular species in the chamber. Substrates were epi-ready GaAs (or GaAs/Si).

Growth conditions were calibrated using RHEED oscillations for careful control of the V/III ratio, and static RHEED for monitoring the surface temperature. Optimum conditions on (100) GaAs were found for a 1.2/1.0 Sb/In incorporation rate and a substrate temperature 5°C above the transition temperature (~395°C). This gave excellent material quality with a mirror-like surface across the entire wafer. The structural quality was examined with five-crystal XRD. Excellent uniformity is indicated by the ±3-arcsec variation of the rocking curve FWHM for a 10-μm InSb layer on a 3-inch semi-insulating GaAs substrate.

InSb on (111)B GaAs Growth of InSb on (111) GaAs was investigated [43] because properly passivated (111) surfaces have been found to have lower surface recombination and provide better electrical isolation than (100) surfaces [44]. The substrates used for this study were 3-inch (111) GaAs from Lockheed Martin Fairchild Systems.

The Sb/In ratio and the substrate temperature were varied to determine optimum conditions for InSb growth on (111) GaAs. The Sb/In ratio was varied from 1.0 to 1.5, and the substrate temperature was varied from 380°C to 460°C. The

Table 7.10 Optimum Growth Conditions for InSb

Growth temperature	465°C
Growth pressure	76 Torr
TMIn flow rate	50 sccm
TMSb flow rate	20 sccm
Total H_2 flow rate	1.5 L/min

optimum conditions were found to be different for the growth of InSb on (111)B surfaces compared to those for (100) surfaces as described previously. For a given Sb/In ratio, it was found that the optimum growth temperature was higher for the growth of InSb on (111)B surfaces than on (100) surfaces. This difference has been attributed to the bonding characteristics at the surface. The (111)B surface has three dangling bonds for Sb atoms and one dangling bond for In atom, whereas the (100) surfaces has two bonds for both atoms.

The epilayers were first characterized using high-resolution XRD. The X-ray FWHM versus thickness measurements show that the crystallinity is nearly identical for (100) and (111)B InSb at a given thickness under optimum growth conditions. As expected, the crystallinity improves with increasing thickness, but there is no large increase in the FWHM for thinner samples, which might indicate that the crystallinity is optimized under these growth conditions.

The substrate temperature has a greater effect than thickness on the crystal quality. The optimum crystallinity is obtained with growth temperatures between 420°C and 440°C, whereas, for (100), the optimum crystallinity is obtained for samples grown at ~395°C. The difference in optimum growth temperature is most likely due to the difference of dangling bonds at the surfaces.

The room temperature electron mobility as a function of epilayer thickness for InSb grown on (100) and (111)B GaAs was compared. It has been found that the data of (111)B and (100) GaAs agree with each other exceptionally well. It appears that when InSb is grown under optimized conditions on (111)B and (100) GaAs, the quality of the material produced is independent of the substrate orientation.

Structure Characteristics

Under optimum conditions, mirror-like InSb epilayers were grown on InSb, GaAs, and Si substrates [45]. The film grown on InSb showed X-ray FWHM of 14 arcsec, whereas those on GaAs and GaAs/Si had FWHM of 171 and 361 arcsec, respectively. The broader peaks on GaAs and Si substrates are due to higher dislocation density from lattice mismatch.

For films grown on lattice-mismatched substrates, the overall crystalline quality of the film is expected to depend on its thickness. Hence, XRD measurements were performed on InSb films of various thicknesses on GaAs and Si substrates. The FWHM of InSb on both substrates decreases with increasing thickness, indicating improved crystalline quality away from the highly mismatched epilayer/substrate interface.

Electrical Characteristics

The Hall mobility is found to improve with increasing thickness. Similar to the X-ray results, this trend reflects the decrease in dislocation density away from the interface. The Hall coefficient R_H was negative and the carrier concentration of the InSb was generally in the range of 1×10^{16} cm^{-3} to 3×10^{16} cm^{-3} at 300K.

The Hall coefficient remained negative at 77K, but the Hall mobility decreased dramatically far below expected values for n-type InSb. This might indicate that the as-grown layer is p-type. A more detailed analysis on this behavior

has been done suing a three-layer model [46], consisting of a surface electron accumulation layer as postulated by Soderstrom et al. [47], an interface layer with high density of dislocations [48], and a bulk-like layer with a highly reduced defect density. The theoretical results, which assumed a *p*-type InSb layer, showed good agreement with experimental results, verifying the growth of *p*-type InSb on GaAs substrates.

A special-grade TMSb was used instead of the electronic-grade TMSb in an attempt to improve the electrical properties. InSb films with thickness ranging from 1.2 to 3.6 μm were grown under growth conditions identical to the previous growth conditions described. Compared with the previous sample grown with electronic-grade TMSb, the structural quality from XRD was comparable and no noticeable change in the room temperature Hall motilities was observed. However, the most striking difference was the low-temperature electrical characteristics. The Hall mobility of the new InSb films remained relatively low for film thicknesses below 2 μm, but it drastically improved beyond this thickness. A 3.6-μm-thick InSb film showed mobility of 56,000 $cm^2/V \cdot s$ at 300K, which increased to about 80,000 $cm^2/V \cdot s$ at 77K. This temperature dependence of the Hall mobility differs from our previous result and resembles that of an *n*-type bulk material. The background carrier concentration at 77K was comparable to the previous InSb samples up to 2 μm but decreased by nearly an order of magnitude to 10^{15} cm^{-3} for thicker samples. These Hall data are comparable to the best epitaxially grown InSb results, irrespective of the growth technique.

Because the growth conditions are identical between the growths with electronic-grade and special-grade TMSb, the changes in electrical properties are attributed to an absence of some impurities in the special-grade TMSb that were otherwise present in the electronic-grade TMSb. The impurities, which have not yet been identified, most likely provided a strong *p*-type background in the previous samples. In the new samples the background is predicted to be *n*-type since the Hall measurements reflected the high-mobility electron characteristics

For the materials grown by MBE, the 300K mobility levels off near that of bulk InSb (75,000 $cm^2/V \cdot s$) as the thickness is more than 2 μm, whereas the 77K mobility continues to increase. The background concentration is about 10^{16} cm^{-3}. Mobility measurement on the 6N purity material showed typical *n*-type characteristics.

Subsequent cell charge with 7N resulted in a background concentration of 10^{15} cm^{-3}. All materials grown with this higher purity source showed slightly higher mobility at 300K, but drastically decreased mobility (5,000 $cm^2/V \cdot s$) at 77K. Similar to our discussion in an earlier section, the three-layer model for MOCVD material was applied here to explain why the 77K mobility could be so low while excellent material quality was indicated by RHEED, X-ray, room temperature mobility, and the device characteristics. In this analysis, the epilayer is assumed to be made up of three layers: an interface layer with high dislocation density, a bulk-like layer, and a surface inversion layer. When the lower purity Sb source was used, the higher background concentration ensured that the 77K mobility measurements characterized the high-mobility bulk-like layer as well as the interface and surface layers. However, with the higher purity Sb source, the bulk-like layer is 10 times less conductive

due to the order of the magnitude decrease in carrier concentration, which makes the high dislocation density interface layer dominant. When the material was doped *n*-type in the range of 5×10^{15} cm^{-3} to 1×10^{16} cm^{-3}, the 77K mobility would be leveled up to that of lower purity Sb result. This, however, did not rule out the possibility of compensation, because the *n*-type doping would negate the inversion and would again make the bulk-like layer more significant.

However, the possibility of compensation was effectively eliminated by the following experiment. Several thick InSb/GaAs samples grown with the higher purity Sb were epoxied to glass slides and the substrates were etched off leaving only the InSb epilayer with mirror-like morphology. A 5-μm sample exhibited mobility of 65,000 cm^2/V·s at 300K and 95,000 cm^2/V·s at 77K. Before removing the substrates, this sample had mobility of 60,000 cm^2/V·s at 300K and 4,230 cm^2/V·s at 77K. This indicated that the InSb/GaAs interface was indeed dominating the 77K measurement. This etching procedure also shows that the interface does not preclude the growth of high-quality InSb away from the interface.

Optical Characteristics

The experimental transmittance spectra of 3.1-μm-thick InSb display a sharp absorption edge around 1,200 cm^{-1} at 300K, which shifts to 1,700 cm^{-1} at 77K; however, an accurate assessment of the absorption edge is difficult because of Fabry-Perot oscillations. The period of the oscillations yields the phase shift for light traveling trough the epilayer, from which the product $n \cdot d$ can be extracted where n is the refractive index of the epilayer, and d is the thickness.

From the transmission measurements, the absorption coefficient of InSb was extracted. To take Fabry-Perot oscillation into account, we have modeled the InSb epilayer as a conventional Fabry-Perot cavity filled with absorbing medium. The reflection and transmission coefficients of each interface were determined using tabulated values for the refractive index of GaAs substrate and InSb. The phase shift through the epilayer was determined from the position of the experimental extrema. Photoconductivity measurements were carried out at 77K and showed a sharp cutoff at 5.5 μm, which corresponds to a bandgap energy of 0.23 eV.

Doping Characteristics

N-Type Doping TESn was used as *n*-type dopant. The TESn bubbler was kept at a temperature of –25°C and the flow rate was varied from 7 sccm (0.7 nmol/min) to 80 sccm (8.0 nmol/min), resulting in a doping level from 5×10^{16} cm^{-3} to 1.2×10^{18} cm^{-3}. No degradation in morphology and X-ray FWHM were observed. Furthermore, no noticeable memory effects were associated with TESn.

P-Type Dopant P-doping was first carried out using DMZn. There was no observable effect on the growth rate or the morphology. With a 10-sccm capacity mass flow controller, a controllable doping level no lower than 3×10^{18} cm^{-3} was obtained. The doping level remained steady down to 4K without any singularity, which indicates complete ionization of the acceptors as well as absence of noticeable surface and interface layer contributions.

To expand the *p*-doping levels to lower ranges, DEZn was considered because of its lower vapor pressure. With the bubbler temperature at 10°C, the flow of DEZn was varied from 0.05 sccm (0.02 μmol/min) to 0.2 sccm (0.08 μmol/min). These flows yield doping levels from 2.5×10^{17} cm^{-3} to 7.5×10^{17} cm^{-3}. In conjunction with the results of DMZn, the successful Zn doping of InSb in the range of 10^{17} to 10^{19} cm^{-3} was achieved, which is sufficient for the p$^+$ layer in InSb photodiode structure.

InSb Photodetectors

The InSb photodiodes were grown on 3-in. Si and (111) GaAs substrates. The InSb photodiodes typically consisted of a 2-μm n^+ region (~10^{18} cm^{-3} at 77K), an ~6-μm unintentionally doped region (n ~ 10^{15} cm^{-3} at 77K), and an ~0.5-μm p^+ (~10^{18} cm^{-3}) contact layer. The crystallinity of these structures was excellent as confirmed by XRD, which showed FWHM < 100 arcsec for structures grown on (100) and (111) GaAs and Si substrates. Photodiodes were fabricated with 400×400-μm^2 mesa structures by photolithography and wet chemical etching. Au/Ti ohmic contacts for both *n*- and *p*-type layers were made by an electron beam evaporator. The contact pattern was defined again by standard photolithography and selective etching. The chips were assembled using standard circuit technologies including die separation, mounting, and Au-wire bonding.

The relative photoresponse of these devices was measured using a FTIR system. The absolute responsivity was obtained by measuring the signal using a calibrated blackbody source and a lock-in amplifier system.

These devices showed an excellent response comparable to that of bulk detectors, with detectivities of ~ 3×10^{10} cmHz$^{1/2}$/W at 77K. Note that the detectors operate up to room temperature, even though these devices were not optimized for room temperature operation.

7.3 InAsSb

The cutoff of InSb is ~5.5 μm at 77K and ~7 μm at 300K. Many applications require longer wavelength detection. Because 300K radiation has a peak around 9 to 11 μm, IR detectors that detect in this region are useful for thermal imaging applications. Thus, there is a great amount of work in IR detector material covering the 8- to 12-μm wavelength region.

Among III-V semiconductor alloys, InAsSb has the lowest bandgap (with the exception of Tl- and Bi- containing alloys). Experimental results have showed that InAsSb can cover the entire 8- to 12-μm range at near ambient temperature (200–300K). It is a promising material system for uncooled photon detectors for LWIR applications.

7.3.1 Physical Properties

Bandgap

The $InAs_{1-x}Sb_x$ alloy has a zinc-blende structure and direct gap at the Brillouin zone center. The shape of the electron band and the light hole band is determined by $\mathbf{k} \cdot \mathbf{p}$

theory. According to Wieder and Clawson [49], the energy gap of $InAs_{1-x}Sb_x$ may be described by the following expression:

$$E_g(x,T) = 0.411 - \frac{3.4\times10^{-4}T^2}{210+T} - 0.876x + 0.70x^2 + 3.4\times10^{-4}xT(1-x) \qquad (7.8)$$

Some recent studies show that the minimum bandgap can be considerably less compared to that predicted by (7.8). Bethea et al. [50] estimated the 300K bandgap of $InAs_{0.22}Sb_{0.78}$ to be 0.095eV. They also measured significant photoconductivity at 10 μm at 77K. Chiang and Bedair [51] reported an ~9-μm cutoff wavelength of $x = 0.6$ for a photovoltaic detector operating at 77K. Our 300K absorption measurements give $\alpha = 500$ cm^{-1} for $\lambda = 15$ μm for $x = 0.65$, which corresponds to $E_g = 0.083$ eV. Therefore, the practical limitation for long-wavelength operation at room temperature is not due to a wide gap but rather to a high generation recombination rate.

Effective Masses

To obtain good agreement between experimental room temperature effective masses and calculations, Rogalski and Jozwikowski [52] have taken into account the conduction-valence-band mixing theory. In this approximation, the resultant conduction band effective mass can be written as follows:

$$\frac{1}{m_e^*} = \frac{1}{m_{ce}^*} + \frac{\delta E}{3}\left[\frac{1}{m_{hh}^* E_{gv}} + \frac{1}{m_{lh}^* E_{gv}} + \frac{1}{m_s^*\left(E_{gv}+\Delta_v\right)} - \frac{1}{m_{ce}^*}\left(\frac{2}{E_{gv}} + \frac{1}{\left(E_{gv}+\Delta_v\right)}\right)\right] \qquad (7.9)$$

Here, the difference $\delta E = E_{gv} - E_g$ is determined by the effect of the periodic potential due to the disorder. The terms E_{gv} and Δ_v can be approximated by these expressions: $E_{gv} = 0.351–0.176x$ and $\Delta_v = 0.39 + 0.42x$. Masses m^*_{hh}, m^*_{lh}, and m^*_s are the effective masses of the heavy hole, light hole, and split-off bands, respectively. Finally, m^*_{ec} is the effective mass for the conduction band in the absence of conduction band mixing. The $m^*_e(x)$ dependence obtained from (7.9) gives good agreement with the experimental data.

Intrinsic Carrier Concentration

The intrinsic carrier concentration in $InAs_{1-x}Sb_x$ as a function of x for various temperatures has been calculated in terms of the $\mathbf{k}\cdot\mathbf{p}$ theory. In the calculation, the nonparabolic bands were taken into account and a parabolic shape was assumed for the heavy hole band. By fitting the calculated nonparabolic n_i value to the expression for parabolic bands, the following approximation for the intrinsic carrier concentration has been obtained [25]:

$$n_i = \left(1.35 + 8.50x + 4.22 \times 10^{-3} T = 1.53 \times 10^{-3} xT - 6.73x^2\right) \times 10^{14} T^{\frac{3}{2}} E_g^{\frac{3}{4}} \exp\left(-E_g/2kT\right) \tag{7.10}$$

where n_i is in cm^{-3}.

Little information is available on the physical properties of InAsSb alloys. It is reasonable to assume that the InAsSb ternary semiconductor will exhibit properties similar to those of both InAs and InSb, although additional process occurs, for example, an additional scattering process due to the random scattering potential. Table 7.11 summarizes material parameters for InAs, $InAs_{0.35}Sb_{0.65,}$ and InSb.

Table 7.11 Physical Properties of InAsSb

	T (K)	*InAs*	*$InAs_{0.35}Sb_{0.65}$*	*InSb*
Crystal structure		Cubic (ZnS)	Cubic (ZnS)	Cubic (ZnS)
Lattice constant (Å)	300	6.0584	6.36	6.47877
Coefficient of thermal expansion ($10^{-6}K^{-1}$)	300	5.02		5.04
	80			6.50
Density γ (g/cm^3)	300	5.68		5.7751
Melting point (K)		1210		803
Energy gap (eV)	4.2	0.42	0.138	0.2357
	80	0.414	0.136	0.228
	300	0.359	0.1	0.180
Thermal coefficient of E_g	100–300	-2.8×10^4		-2.8×10^4
m_e^*/m_0	4.2	0.023		0.0145
	300	0.022	0.0101	0.0116
m_{lh}^*/m_0	4.2	0.026		0.0149
m_{hh}^*/m_0	4.2	0.43	0.41	0.41
Momentum matrix element P ($cm^2/V \cdot s$)		9.2×10^{-8}		9.4×10^{-8}
Electron mobility ($cm^2/V \cdot s$)	300	3×10^4	5×10^5	8×10^4
	77	8×10^4	5×10^4	10^6
Hole mobility ($cm^2/V \cdot s$)	300	500		800
	77			10^4
Intrinsic carrier concentration (cm^{-3})	77	6.5×10^3	2.0×10^{12}	2.6×10^9
	200	7.8×10^{12}	8.6×10^{15}	9.1×10^{14}
	300	9.3×10^{14}	4.1×10^{16}	1.9×10^{16}
Refractive index		3.44		3.96
Static dielectric constant		14.5		17.9
High-frequency dielectric constant		11.6	16.8	
Optical phonon (cm^{-1})				
LO		242	»210	193
TO		220	»200	185

Source: [4].

7.3.2 Growth of InAsSb

MOCVD

For our investigation of the growth of InAsSb, a test structure, consisting of an InSb buffer layer grown for 30 minutes followed by an InAsSb layer for 2 hours, was adopted. InSb buffer layers were grown under the optimum conditions determined from InSb investigation (as described earlier). The growth of InAsSb layers has been performed in the following sequences. First, TMIn flow was directed to a vent line to interrupt the growth of InSb. Second, under the TMSb overflow, the flow rate of TMIn was decreased to about 90% of the original value for the InSb growth. It was found experimentally that the V/III ratio needs to be increased for the InAsSb growth. This might be caused by the parasitic reaction between arsine and TMSb. Third, arsine flow was introduced into the growth chamber. TMIn was redirected to growth chamber as the last step to initiate the growth of InAsSb. The average growth rate of these layers was about 0.8 μm/hr.

MBE

The growth of InAsSb, especially for high Sb composition, is very similar to the growth of InSb. The layers were grown on GaAs substrate using an EPI Modular Gen II Solid source MBE system with a V/III ratio of ~>1.2. The composition was controlled by RHEED calibration of the In and Sb cell, and by using an overpressure of As, the substrate temperature was varied to obtain correct composition, which resulted in an optimum temperature of ~400°C as determined by IR pyrometer.

Generally, as the As composition increases from InSb, the X-ray FWHM increases. The mobility of InAsSb layers was less than InSb, decreasing with increasing As content. Room temperature mobility was in the 5,000 to 50,000 $cm^2/V \cdot s$ range, and was in the few thousand $cm^2/V \cdot s$ at 77K. Photoluminescence measurements were performed on samples. The PL linewidths at 77K were in the 30- to 60-meV range.

In addition to InAsSb, AlInSb and GaInSb were investigated for these detector structures. These layers have wider bandgaps, so that carriers can be confined to the active region. Also, if doped properly, the layers can be barriers to one type of carrier while allowing the other type of carrier through to a contact.

7.3.3 Characterizations

Structural Characteristics

The grown layer surfaces are mirror-like. The alloy composition was estimated from XRD using Vegard's law. At first, the dependence of Sb composition of $InAs_{1-x}Sb_x$ films on various parameters was studied. Keeping the TMSb flow at 35.4 μmol/min and AsH_3 flow at 13.4 μmol/min, flow of TMIn has been varied from 3.06 to 4.60 μmol/min at a growth temperature of 465°C. The corresponding V/III ratio varied from 15.93 to 10.62. A further increase in the V/III ratio resulted in lower crystal quality and a decrease in the V/III ratio caused In-related droplets on the surface. As the V/III ratio was increased, the composition of Sb was decreased

from 0.8 to 0.63. The morphology and crystallinity dependence of the InAsSb layers on the V/III ratio was not as critical as that of the InSb layer. However, the best morphology and crystallinity were obtained at a slightly higher V/III ratio than that of InSb growth. This might be caused by the parasitic reaction between the TMSb and AsH_3.

The composition dependence on the growth temperature was also investigated. When the temperature increased, the Sb composition decreased, which might be due to the fact that AsH_3 decomposition increases as the temperature increases.

The dependence of Sb composition on the group V partial pressure ratio was also studied. The growth temperature was set at 465°C and the V/IIII ratio was 12.5. The group V input partial pressure ratio TMSb/(TMSb + AsH_3) has been varied from 0.67 to 0.73, which resulted in an SB composition from 0.65 to 0.80.

The XRD peaks of InAsSb were broader than those of InSb of comparable thicknesses despite the good morphology of InAsSb films. The broadness is probably due to a large number of misfit dislocations and some compositional variations in the epitaxial layers.

The dependence of X-ray FWHM on the Sb composition was studied. About 1.2-μm-thick samples with compositions varied from 0.56 to 0.80 were grown. The X-ray FWHM of the samples with $x = 0.80$ were ~1,200 arcsec, which increased to ~2,000 arcsec with $x = 0.56$. The peaks are broader than those of InSb films with comparable thicknesses; and the more As that is incorporated, the broader the peaks. Because InAsSb is less lattice mismatched to GaAs than InSb, this behavior can be explained by the variation in the alloy composition, which becomes more significant as the composition of As is increased.

For monolithic integration applications, growth of InAsSb layers on GaAs-coated Si substrates was also carried out. InAsSb layers were grown on both GaAs/Si and GaAs substrates in the same run. XRD spectra were measured. From the comparison of the epilayers on two substrates, it has been confirmed that the qualities of the InAsSb layers are about the same. The comparison demonstrated that high-quality InAsSb layers can be grown on GaAs/Si substrates.

Electrical Characteristics

Normally undoped InAsSb layers showed negative Hall coefficients through the entire temperature range from 4 to 300K. To investigate the electrical properties of the films with different Sb composition, 2-μm-thick layers with different Sb compositions have been grown. The electrical properties of the samples were measured at 77 and 300K.

The carrier concentrations at 77K ranged from 1 to 6×10^{16} cm^{-3}, and x around 0.65 gives the highest value. This result is in good agreement with the previous result obtained by Yen [53] and is reasonable since InAsSb has lowest bandgap at $x \sim 0.65$. The Hall mobility was from 4,000 to 15,000 cm^2/V·s decreases as the As composition increases. Similar behavior has been found by Coders and Woolley [54] and Tansley et al. [55] and an explanation was given in terms of the effect of alloy scattering.

At 300K, the carrier concentration increased to 4 to 14 × 10^{16} cm^{-3} and Hall mobility increased to 5,000 to 30,000 cm^2/V·s. The dependence of both properties on the Sb composition x was similar to that at 77K. It becomes more clear at 300K that the carrier concentration is dominated by the intrinsic concentration and is peaked at x ~ 0.65. The Hall mobility should increase as the temperature goes down, which is the opposite of our experimental results. The strange behavior suggested that the InAsSb layer might be of the p-type. Theoretical calculations on the $InAs_{0.3}Sb_{0.7}$ layers have shown that with a shallow p-type background doping density of 5 × 10^{15} cm^{-3} and dislocation density on the order of 2.5 to 5 × 10^8 cm^{-3}, calculated data would fit the experimental data well [56].

Optical Characteristics

Optical characterizations on the InAsSb layers have been performed using FTIR. Bandgap variations with As composition can be clearly observed. Optical bandgap can be determined in two ways. First, the normalized transmission spectra are used to calculate the absorption coefficient of the layer. The usual definition of the optical bandgap for the narrow-gap semiconductors is the photon energy for which the absorption coefficient is equal to 500 cm^{-1}. A bandgap of 103 meV was derived for $InAs_{1-x}Sb_x$ with x = 0.77. An even narrower bandgap of 83 meV was obtained for x = 0.65. This shows that InAsSb can be used for the near room temperature photodetectors operating in the entire 8- to 12-μm range.

To obtain more accurate absorption coefficient values, the InAsSb layer was successively etched and transmission spectra were measured at different thicknesses. The transmission spectra were then smoothed to remove both the high-frequency noise and the Fabry-Perot interference fringes. Figure 7.9 shows the smoothed transmission spectra.

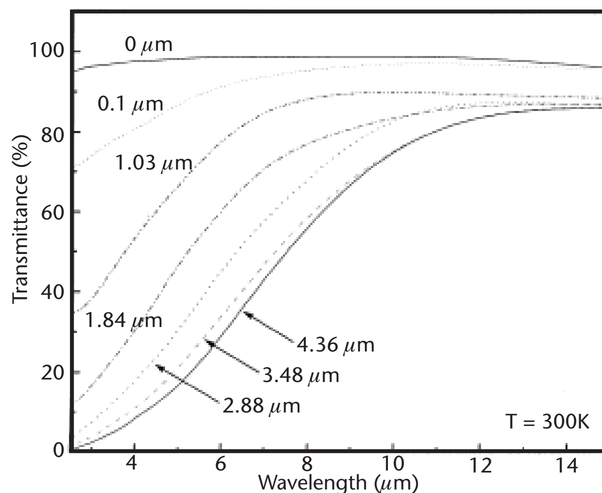


Figure 7.9 Transmission spectra of InAsSb layer at different thicknesses.

The relationship between the transmittance T and the absorption coefficient α at a certain wavelength is expressed as

$$T = e^{-\alpha t} \tag{7.11}$$

where T is the thickness of the layer. For certain wavelengths, T can be plotted as a function of thickness t. By numerical fitting, α can be determined. In this method, α is determined to be 1,330 cm^{-1} for a wavelength of 8 μm.

More accurate measurements of the optical bandgap have been made using the photoconductivity spectra. The optical bandgap of the narrow-gap semiconductors can be determined as the energy at which the normalized photoresponse falls to 20% of the maximum value.

$InAs_{1-x}Sb_x$ with different compositions has been grown and the optical bandgap has been determined using the normalized photoresponse spectra. The experimental values are lower than the theoretical, which has been attributed to the long-range ordering.

7.3.4 Device Measurement

InAsSb Photoconductors

Photoconductors based on p-type $InAs_{0.23}Sb_{0.77}$ were grown on GaAs substrates by LP-MOCVD [57]. The photoconductor structure was composed of two epitaxial layers (p-$InAs_{0.23}Sb_{0.77}$/p-InSb) grown on GaAs. InSb was used as a buffer layer with 2% lattice mismatch to $InAs_{0.23}Sb_{0.77}$ and also as a confinement layer for the electrons in the active layer. The optimized doping level for room temperature operation is around 3×10^{16} cm^{-3}. The epitaxial layers have been grown on (100) semi-insulating GaAs substrates at a growth temperature around 470°C. The geometry of the photoconductor is 4×3 mm^2. Au/Ti contacts were deposited by an electron beam evaporator. The resistance is 23Ω at 300K and 220Ω at 200K.

The photodetectors were mounted in a liquid nitrogen-cooled cryostat system and measurements were taken at temperatures between 77 and 300K. The relative photoresponse spectra were measured with an FTIR spectrometer. Then the responsivity was calibrated by a blackbody test setup, including a blackbody source (Mikron 305), preamplifier (EG&G PA-6), and chopper system (Stanford Research system SR540). The blackbody temperature was set at 800K. The modulation frequency was set at 450 Hz because response measurements as a function of chopper frequency showed that the thermal effect could be neglected at frequencies higher than 200 Hz.

The spectral responsivities of a photoconductor sample were measured at 200 and 300K. The peak responsivity is about 4 mV/W at 8 μm and at 300K for 1-V bias. The cutoff wavelength around 14 μm was observed at 300K. Noise in the spectral response was attributed to FTIR preamplifier since the contribution of thermal or bolometric response is negligible at measuring frequencies of FTIR that are higher than 1.2 kHz. At 10.6 μm the voltage-dependent responsivity increases with applied voltage and reaches saturation at around 3V, which

corresponds to the values of 5.8 mV/W. The corresponding Johnson-noise-limited detectivity of the photoconductor was then estimated to be 3.27×10^7 cmHz$^{1/2}$/W at 300K.

Based on the spectral voltage responsivity, the mobility-effective carrier lifetime product $\mu_e\tau$ and the effective lifetime τ can be derived. According to the simple theory of photoconductivity, the voltage responsivity R_V is

$$R_V = \frac{q\lambda}{hc}\frac{\eta\mu_e \tau V_b R_D}{L^2}\left(1+\frac{1}{b}\right) \tag{7.12}$$

where q is the electron charge, λ is the wavelength, η is the quantum efficiency, b is the electron-to-hole mobility ratio, L is the detector length, V_b is the bias voltage, and R_b is the detector resistance. The quantum efficiency η is given by

$$\eta = (1-r)\frac{\left(1-e^{-\alpha t}\right)}{1-re^{-\alpha t}} \tag{7.13}$$

where r is the reflectance, α is the absorption coefficient, and T is the thickness of the layer. If μ_e is known, τ can be estimated. Based on the electron mobility at 300K, the effective lifetime is estimated to be 0.14 ns.

InAsSb Photodiodes

For the realization of InAsSb photovoltaic detectors, p^+-InSb/p-$InAs_{1-x}Sb_x$/n^+-InSb DH structures were grown on semi-insulating GaAs substrates by MOCVD [58]. The growth temperature was kept at 470°C and the V/III ratio was ~20. The growth rate was ~0.7 to 0.9 μm/hr. The standard structure consisted of a 2-μm heavily doped n-type InSb layer (~3×10^{18} cm^{-3}), a 5-μm p-type (3.6×10^{16} cm^{-3}) $InAs_{1-x}Sb_x$ absorber region, and a 0.5-μm heavily doped InSb p-type (~3×10^{18} cm^{-3}) contact layer.

Structural characterization was performed using XRD spectra at (004) orientation. The composition of $InAs_{1-x}Sb_x$ has been determined from XRD data and Vegard's law. The $InAs_{1-x}Sb_x$ composition calculated from diffraction data was x = 0.85. The mesa structure has been etched using photolithography. Au/Pt/Ti contacts were deposited by an electron beam evaporator.

The voltage response at various temperatures was measured. The temperature-dependent cutoff wavelength was clearly observed. A photoresponse up to 13 μm has been observed. The peak voltage-responsivity is 9.13×10^{-2} V/W at 300K. However, it is only 2.85×10^1 V/W at 77K, which is much lower than the expected value. Possible reasons are the poor interface properties due to the lattice mismatch between the absorber and contact layers and high dark current due to the high doping level in the active layer.

Because the series resistance is much higher than the junction resistance at high temperature, we could not measure the resistance-area product R_0A directly from the I-V curve [31]. The R_0A product can be calculated as follows:

$$R_V A = \frac{\lambda \eta q R_0 A}{hc} \tag{7.14}$$

where λ is the wavelength, η is the quantum efficiency, q is the electron charge, h is the Planck constant, and c is the light velocity. Table 7.12 shows the responsivity-area product and resistance-area product of the $InAs_{0.15}Sb_{0.85}$ photodiode at various temperatures [59].

InAsSb Photodiodes by MBE

InAsSb photodiodes structures were grown on (100) and (111) GaAs substrates by MBE. Several device structures were investigated. First p^+-InSb/p-$InAs_{1-x}Sb_x$/n^+-InSb DH structures were grown on semi-insulating GaAs substrates. The bottom n^+-InSb layer with about a 2-μm thickness was doped at 2×10^{18} cm^{-3} both for increased quantum efficiency due to the pronounced Moss-Burstein effect and for a low series resistance due to high electron mobility. The thickness of the active layer was ~2 to 5 μm to ensure both high quantum efficiency and high optical gain. The top p^+-InSb contact layer was 0.5 μm thick and doped at a level of 10^{18} cm^{-3}.

Some device structures utilized the bottom InSb layer as a buffer layer with the bottom contact made directly to the InAsSb active region. Subsequent devices utilized a wider gap AlInSb buffer (or bottom contact) layer. This structure allows for the flexibility for the mesa to be etched down to the InAsSb active region or down to the AlInSb layer, which results in a heterojunction.

The 300K Johnson-noise-limited detectivity for the device was calculated at ~1×10^8 cmHz$^{1/2}$/W. The R_0A was ~10^{-3} Ω cm^2 at 300K, which is attributed to the highly dislocated interface layer and the unpassivated surface. The R_0A was increased by an order of magnitude by incorporating an AlInSb buffer layer, which blocks carriers from the highly dislocated interface, resulting in a peak detectivity of ~ 3×10^8 cmHz$^{1/2}$/W.

The wider gap AlInSb layer can also be used as the bottom contact layer, resulting in a single-heterostructure photodiode. The heterojunction increases the R_0A of the device from ~10^{-3} to ~10^{-1} Ω cm^2. The peak differential resistance of 70Ω at zero bias results in an R_0A of 0.11 Ω cm^2 at 300K. The Johnson-noise-limited

Table 7.12 Peak Voltage Responsivity-Area Product and Corresponding Resistance-Area Product of InAsSb Photovoltaic Detectors at Various Temperatures

T (K)	*R_VA (V cm^2/W)*	*R_0A (Ω cm^2)*
77	4.56×10^{-2}	3.15×10^{-2}
100	2.58 × 10–^{2}	1.53×10^{-2}
150	9.68×10^{-3}	5.73×10^{-3}
200	1.64×10^{-3}	9.71×10^{-4}
220	1.03×10^{-3}	6.10×10^{-4}
250	5.57×10^{-4}	3.30×10^{-4}
300	1.46×10^{-4}	7.56×10^{-5}

detectivity was estimated at ~4 × 10^8 cmHz$^{1/2}$/W at 3 μm and ~5 × 10^7 cmHz$^{1/2}$/W at 6 μm.

7.4 InTlSb

As an alternative to the InAsSb system, InTlSb is also a potential material for $\lambda > 8$ μm detectors. TlSb was calculated as a semimetal from the full-potential *linear muffin-tin* (LMTO) method within the *local density approximation* (LDA) [60]. By alloying TlSb with InSb, it was suggested that the bandgap of InTlSb could be varied from –1.5 to 0.26 eV. Assuming a linear dependence of the bandgap on alloy composition, InTlSb can then be expected to reach a bandgap of 0.1 eV at $x = 0.08$ while exhibiting similar lattice constant as InSb since the radius of Tl atom is very similar to In. To date, very limited experimental work has been reported on the growth and characterization of InTlSb. Choi et al. [61] reported on the first successful growth of InTlSb by LP-MOCVD. A bandgap shift toward a longer wavelength was observed, which suggests a concentration of 6% Tl. A group from the University of Utah reported Tl incorporation of less than 1% in InSb [62]. Recently, Karam et al. [63] reported the InTlSb films with Tl composition up to 10%. Spectral response measurements showed a photoresponse up to 14 μm at 77K.

7.4.1 MOCVD Growth of InTlSb

The growth of InTlSb was performed on InSb-, GaAs-, and GaAs-coated silicon substrate with a low-pressure MOCVD. TMIn, TMSb, and CPTl were used as precursors. The structure consisted of an InSb buffer layer grown for 1 hour followed by an InTlSb layer grown for another hour. InSb buffer layers were grown under the optimum conditions determined from InSb investigations. The growth of InTlSb was first attempted by introducing a small flow of CPTl without changing other growth parameters. Once the growth of InTlSb has been verified, a different amount of Tl was sent in order to obtain $In_{1-x}In_xSb$ alloys of various x compositions. This was achieved by varying both the bubbler temperature as well as the flow rate of CPTl. Table 7.13 summarizes the growth conditions for InTlSb.

Chemical Analysis by AES

AES measurement was performed on InTlSb samples and compared with that of reference InSb to verify the incorporation of Tl [64]. The InTlSb sample shows a noticeable intensity dip near 70 eV compared with the reference sample, confirming the presence of Tl.

Structural Characteristics

XRD measurements were performed on as-grown InTlSb samples. Tl flow was varied to grown InTlSb alloys with different Tl composition [65–68]. An InSb epilayer is grown on the GaAs substrate to provide a reference peak for the XRD study. The

Table 7.13 Summary of Growth Conditions for the Nominally Undoped InTlSb Layers

Growth temperature		425–475°C
Growth pressure		76 Torr
Substrate orientation		InSb-, GaAs-, GaAs-coated Si
TMIn	Temperature	18°C
	Flow rate	36–55 sccm
TMSb	Temperature	0°C
	Flow rate	16–36 sccm
CPTl	Temperature	0–80°C
	Flow rate	3–300 sccm
H_2	Flow rate	1.5 L/min

X-ray peaks corresponding to the InSb epilayer and GaAs substrate are observed at angles of 28.39° and 33.02°, respectively. A noticeable feature appeared in the XRD spectra for the $In_{1-x}Tl_xSb$ films. The epilayers, which exhibited a Tl peak in the AES spectrum, showed a clear shift of the XRD peak toward a higher angle relative to that of InSb. This implies a decrease of the lattice constant (in the growth direction) of $In_{1-x}Tl_xSb$ compared to the lattice constant of InSb, according to Bragg's law. As Tl incorporation increases, the diffraction peaks shift gradually toward a higher angle in the experimental range, indicating the gradual decrease of the lattice constants in the growth direction. The lattice constants normal to the growth direction are assessed by asymmetric {115} reflections and found to be the same as the lattice constants in the growth direction. These results indicate that the lattice of $In_{1-x}Tl_xSb$ is contracted compared to that of InSb due to the incorporation of Tl in InSb. This behavior is particularly striking because it is in direct contrast with the atomic size effect based on tetrahedral covalent radii, which in general holds for isoelectronic impurity incorporation in III antimonides [69]. The low oxidation state of Tl in $In_{1-x}Tl_xSb$ has been suggested as a possible reason for the observed lattice contraction [70]. The present results indicate that the incorporation behavior of Tl in $In_{1-x}Tl_xSb$ differs from that of other group III impurities in III antimonides. The FWHM of the epilayer's (004) diffraction as a function of Tl concentration shows the gradual increase of the FWHM with Tl incorporation, indicating that the crystallinity of the films is deteriorating.

Electrical Characteristics

A Hall effect measurement of InTlSb/InSb on semi-insulating GaAs sample was performed. The Hall coefficient R_H remains negative throughout the entire temperature range and its temperature dependence resembles that of an *n*-type material.

The Hall effect of InTlSb samples with different Tl flow was investigated. Room temperature Hall mobility ranged from 2×10^4 to 5×10^4 $cm^2/V \cdot s$, and the electron concentration from 1×10^{16} to 5×10^{16} cm^{-3}. Because a direct assessment of thallium is difficult, the electrical characteristics have been plotted as a function of the lattice

mismatch between InTlSb and InSb, which was determined from XRD measurements. The lattice mismatch is used here as an indirect indicator of thallium content. The Hall mobility decreased monotonically with increasing thallium flow, while the electron concentration increased simultaneously. The increase in electron concentration is typical of an intrinsic semiconductor with decreasing bandgap. The decrease in mobility is attributed to an increase in alloy scattering with increasing Tl content. Alloy scattering has been shown to be significant in InSb-based compounds such as InAsSb [71–73].

Optical Characteristics

The results of the optical transmission measurements of InSb and InTlSb showed that the absorption edge of InTlSb was clearly seen beyond that of the InSb (λ^{-1} = 1,818 cm^{-1}), verifying a bandgap reduction of InTlSb. However, an accurate assessment of the absorption edge by this method is difficult due to the thin epilayers and the observed Fabry-Perot oscillation. These oscillations are produced by multiple reflections at the air/epilayers and epilayers/substrate interfaces, and from the periodicity other optical parameters such as refractive index of the epilayers have been extracted [40].

To determine the absorption coefficient, the epilayer has been modeled as a conventional Fabry-Perot etalon filled with an absorbing medium. The reflection and transmission coefficient of each interface were determined using tabulated values for the refractive index of GaAs and InSb, and the refractive index of InTlSb was assumed to be the same as that of InSb. The phase shift through the epilayer was determined from the position of the experimental Fabry-Perot extrema. The InSb sample shows a sharp absorption edge at 5.5 μm. The InTlSb/InSb sample exhibits overall higher absorption, and an absorption tail extending beyond InSb absorption edge is observed. Note that the calculated absorption coefficient for InTlSb/InSb given in the plot corresponds to the average absorption of the two layers weighted by their thickness, respectively:

$$\alpha = \frac{\alpha_{\mathrm{InTlSb}} d_{\mathrm{InTlSb}} + \alpha_{\mathrm{InSb}} d_{\mathrm{InSb}}}{d_{\mathrm{InTlSb}} + d_{\mathrm{InSb}}} \tag{7.15}$$

where α_{InSb} and α_{InTlSb} are the absorption coefficients of each layer and d_{InSb} and d_{InTlSb} are their respective thicknesses. The tendency to attenuate any feature of InTlSb absorption spectrum and the observed shift should be even more apparent for a thick InTlSb layer.

The spectral photoresponse from the InSb and InTlSb films grown on GaAs substrates showed an absorption cutoff wavelength at about 5.5 μm at 77K or the InSb sample. A distinct shift of the absorption edge from 5.5 to 8.4 μm for InTlSb was observed at 77K. In general agreement with the theoretical prediction, the samples with high Tl flows give longer response wavelengths. The rough estimation of the Tl concentration in the samples has been made by linearly interpolating the energy bandgap between InSb and TlSb from that fact that bandgap energy of the

grown InTlSb is only slightly deviated from that of InSb. An energy bandgap of –1.5 eV is assumed for the semimetal TlSb.

7.4.2 InTlSb Photodetectors

The first demonstration of InTlSb photodetectors was fabricated on GaAs substrates. Preliminary photodetectors were operated in the photoconductive mode. The photoconductors were rectangular shaped with dimension of 3 × 3 mm^2. Ti/Au was evaporated by electron beam evaporation to make contacts. The resistance is about 50Ω at room temperature and increases to several thousands ohms at 77K. The spectral photoresponse is measured by a FTIR spectrometer with a low-noise preamplifier. The responsibility is calibrated by a blackbody test setup, including a blackbody source (Mikron 305), preamplifier (EG&G PA-6), and chopper system. The blackbody temperature is set at 800K. The modulation frequency is set at 400 Hz since response measurements as a function of chopper frequency showed that the thermal effect could be neglected at frequencies higher than 200 Hz. The photodetectors were mounted in a liquid nitrogen-cooled cryostat system and the measurements were taken at temperatures between 77 and 300K.

The absolute responsivity for an $In_{0.94}Tl_{0.06}Sb$ photodetector showed a cutoff wavelength that extends to 11 μm at 300K. The maximum responsivity in an $In_{0.96}Tl_{0.04}Sb$ photodetector is about 6.64 V/W at 77K. The corresponding Johnson-noise-limited detectivity is 7.64 × 10^8 cm $Hz^{1/2}$/W. The performance of $In_{1-x}Tl_xSb$ photoconductive detectors is listed in Table 7.14.

7.5 InBiSb

Besides adding As or Tl to an InSb lattice, it is also possible to substitute Bi. As a result, a rapid reduction of the bandgap is expected. The basis of the prediction is InBi, which has a bandgap of 1.5 eV. By alloying InBi with InSb, it has been suggested that the bandgap energy of InSbBi could be varied from –1.5 to 0.23 eV. Assuming a linear dependence of the bandgap energy on the alloy composition, $InSb_{1-x}Bi_x$ can then be expected to reach a bandgap of 0.1 eV at x = 0.08.

7.5.1 MOCVD Growth of InSbBi

For the growth of the InSbBi alloy, InSb buffer was first grown under the optimum growth conditions determined from the InSb investigations. The growth of InSbBi

Table 7.14 Performance of $In1_{-x}Tl_xSb$ Photoconductors at 77K

$In_{1-x}Tl_xSb$	λ_{cutoff} (μm)	R_v^{Max} (V/W)	D^*_{max} (10^8 cm $Hz^{1/2}$/W)	μ (10^3 cm^2/V s)	T (ns)
x = 0.023	6.5	2.73	2.47	1.29	19.2
x = 0.035	7.3	6.64	7.64	2.00	50.1
x = 0.047	8.4	3.56	8.54	2.75	40.4
x = 0.053	9.4	0.79	2.63	4.13	10.3

was then attempted by intruding a small flow of TMBi primarily near the optimum growth condition of InSb. The highest Bi concentration in InSbBi was achieved with the growth temperature of 456°C, growth pressure of 76 Torr, and V/III ratio of 13. The growth rate of this layer was about 1.3 μm/hr.

Structural Characteristics

The (004) XRD for an as-deposited InSbBi layer on InSb substrate showed two distinct peaks around 28.4°. The X-ray peak corresponding to InSb substrate is observed at an angle of 28.39°. The X-ray peak at an angle of 28.49° is attributed to the InSbBi and clearly distinguished from that of InSb substrate. To confirm that the peak at higher angular position corresponds to the InSbBi epilayer, X-ray rocking curves are measured after successive etching of the film. There is no difference in InSb-InSbBi peak angular separation; the intensity of the peak associated with InSbBi, which is proportional to the film thickness, is observed to decrease with respect to that of InSb. This indicates that the X-ray peak at high angular position indeed originates from the InSbBi epilayer.

A study of the dependence of the lattice constant on the group V input partial pressure ratio has been performed. The growth temperature was set at 450°C and group V input partial pressure ratio Bi/(Bi + Sb) has been varied from 0.043 to 0.137. Quasilinear dependence of the lattice constant on group V input partial pressure ratio was obtained. To investigate the crystallinity of the InSbBi thin films, epilayers with different lattice mismatch with InSb have been prepared. As the mismatch increases, the FWHM increases, indicating worse crystallinity in films with larger lattice mismatch. The broadening of the InSbBi peak is attributed to an increase in the defects such as dislocation arising from the increased mismatch with InSb.

Energy dispersive X-ray (EDX) spectra for InSb and InSbBi/InSb showed no elements detected other than In, Sb, and Bi. Compared to the InSb sample, the EDX spectrum of the InSbBi layer clearly showed the presence of a Bi peak at 2.4 eV, which corresponds to a Bi M_α line. This result further confirms the incorporation of Bi in the epilayer.

Optical Characteristics

The reduction of the bandgap energy by the addition of Bi into InSb was verified by optical measurements. Optical absorption measurements were first carried out to study the effect of Bi incorporation on the bandgap energy. Figure 7.10 shows the differences in the absorption coefficients between InSb and InSbBi, which are extracted from the transmission spectra. As can be seen, the difference of the absorption coefficient between the two materials is quite substantial, especially in the range beyond 6 μm. It is also clear that the incorporation of Bi in InSb shifts the absorption edge to a smaller energy. The bandgap energy of the investigated $InSb_{0.96}Bi_{0.04}$ sample roughly estimated from the absorption coefficient of 500 cm^{-1} was 0.11 eV at room temperature.

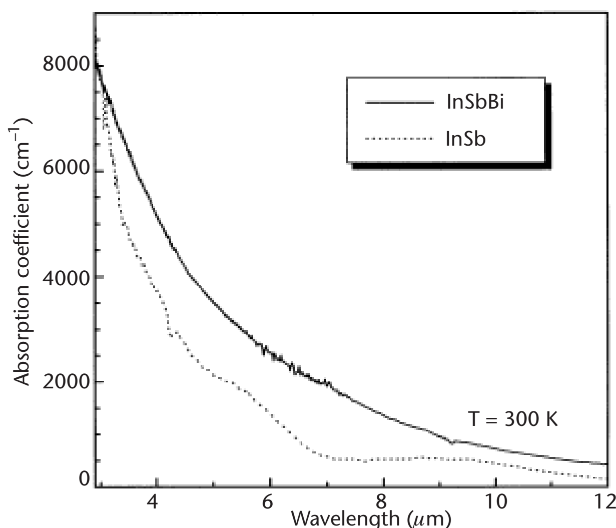


Figure 7.10 Absorption coefficient of InSbBi at 300K.

Photoresponse measurements were performed using FTIR. A distinct shift of absorption edge from 5.5 μm for InSb up to 9.3 μm for InSbBi was observed at 77K. In general agreement with the theoretical prediction, the samples grown at higher Bi input flows exhibited longer response wavelengths.

The rough estimation of solid Bi concentration in the samples has been made by linearly interpolating the energy bandgap between InSb and InBi from the fact that the bandgap energy of the grown InSbBi deviates only slightly from that of InSb. Energy bandgap of –1.5 eV at 300K is assumed for the semimetal InBi. A cutoff wavelength of 9.3 μm corresponds to about 5.8% Bi incorporation in the InSbBi layer.

Electrical Characteristics

The electrical properties of the InSbBi films with thicknesses varying between 1.4 and 2.0 μm were characterized by Hall measurements. At both 77 and 300K, *n*-type conductivity has been observed by the incorporation of Bi, which is believed to substitute for Sb in the InSb lattice.

At room temperature, mobility decreased from 44,100 down to 4,910 $cm^2/V \cdot s$ as the Bi concentration increased. At 77K the mobility decreased from 11,000 to 1,200 $cm^2/V \cdot s$ as the Bi concentration increased. The dependence of mobility is probably due to the random alloy scattering potential, which has been shown to be significant in other InSb-based compounds such as InAsSb ternary alloy.

The electron concentration increased from 3.7×10^{16} up to 2.7×10^{17} cm^{-3} with the increase of Bi incorporation at 300K. At 77K, the electron concentration increased from 4.8×10^{15} to 3.9×10^{17} cm^{-3} as the Bi concentration increased. The increase of electron concentration with the increase of Bi concentration is compatible with a decrease of the energy bandgap in InSbBi ternary alloys.

Temperature-dependent Hall measurements were also carried out to examine the Hall mobility and carrier concentration behavior as a function of temperature. The mobilities first increased a little bit and then decreased (x = 0, 4%) when temperature decreased from 300K to 77K. The maximum motilities occur at 250°C for x = 0 and 275°C for x = 4%. The decrease in mobility with decreasing temperature was attributed to the dislocation scattering that dominates at low temperature. Dislocation scattering is expected to be the dominant scattering mechanism at low temperature in a dislocated material with a low or moderate doping level.

It has been found that the electron concentration of the Bi incorporated sample is always larger than that of InSb sample regardless of temperature. The more Bi that is incorporated into the InSb, the less intrinsic behavior is observed in the InSbBi samples. Sample with 5% Bi content shows an almost flat carrier concentration down to 125K and then an increase of concentration, implying the unintentional p-type doping nature.

7.5.2 InSbBi Photodetectors

The preliminary LWIR InSbBi photoconductors were fabricated from the samples grown on semi-insulating GaAs substrates [74]. For the ohmic contacts, 500Å Au and 3,000Å Ti were evaporated by electron beam evaporation. The contact pattern is made using a metal mask, which covers parts of the surface during the evaporation. The active area of the photodetector was 4 mm^2 with a 2-mm electrode spacing. The photodetectors were mounted onto copper heat sinks and the electrical contact was made by an Au-wire bonding. The resistance of an $InSb_{0.96}Bi_{0.04}$ photodetector was 435Ω at 77K and 110Ω at 300K.

The spectral response at different temperatures of the photodetector fabricated from the $InSb_{0.96}Bi_{0.04}$ sample was measured. The photoresponse cutoff wavelength shifts to 7.7 μm at 77K and to 9.3 μm at 200K. The responsivity decreases with the increase in temperature. The SNR was too weak to make adequate measurements at temperatures greater than 200K. This decrease of responsivity and SNR is due to the increase of intrinsic carries caused by the decrease of the bandgap. The maximum responsivity at 7 μm is about 3.2 V/W. The corresponding Johnson-noise-limited detectivity is about 4.7×10^8 cmHz$^{1/2}$/W.

The carrier lifetime of the fabricated LWIR detector is estimated from the voltage-dependent responsivity measurements. The measured responsivity increases linearly with the bias at low bias voltage. However, it can be observed that the responsivity begins to saturate at a bias voltage of 5.5V at 77K. This saturation of responsivity can be explained by the minority carrier sweepout at the electrode at high bias field. Taking the absorption quantum efficiency as $\eta = 0.21$ from IR transmission measurements, the carrier lifetime mobility is found to be $\tau\eta = 2.35 \times 10^{-4}$ cm^2/V. So the carrier lifetime is estimated to be about 86 ns at 77K.

Room temperature operation photoconductors based on n-type $InSb_{0.95}Bi_{0.05}$ have been fabricated and characterized for the first time [75]. The spectral response of an uncooled $InSb_{0.95}Bi_{0.05}$ photoconductor showed a cutoff wavelength that extends up to 12 μm at 300K. The voltage responsivity at 10.6 μm was about 1.9

mV/W at room temperature and the corresponding Johnson-noise-limited detectivity was estimated to be about 1.2×10^6 cmHz$^{1/2}$/W. The carrier lifetime derived from the voltage-dependent responsivity measurements was about 0.7 ns.

7.6 InTlAsSb

To exceed the limitations imposed by the InTlSb miscibility gap, arsenic can be added to form a quaternary material with an extended photoresponse cutoff wavelength. This work led to the first successful demonstration of InTlAsSb on GaAs substrates with a cutoff wavelength up to 15 μm at room temperature. These measurements were the first observations of a room temperature band-edge photoresponse at such long wavelengths from III-V alloys.

Film deposition was carried out in a horizontal cold wall LP-MOCVD reactor. TMIn, TMSb, CPTl, and 5% AsH_3 were used as precursors of indium, antimony, thallium, and arsenic, respectively. The reactor pressure was 76 Torr for all of the growth and the temperature was around 455°C. The substrates were superclean semi-insulated GaAs (100).

To investigate the bandgap variation due to the As incorporation into InTlSb, IR photoresponse measurements using FTIR were performed. Quaternary sample (b) [(d)] has the same growth conditions with ternary (a) [(c)] except for the As flux. For the InSb sample, an absorption cutoff wavelength at about 5.5 μm was observed at 77K, which agrees with the PL measurement. The feature around 4.3 μm is due to the CO_2 absorption in the spectrometer. A clear shift of an absorption cutoff wavelength to a longer wavelength up to 8.5 μm was obtained for InTlSb, which indicates decrease in the energy gap by Tl incorporation. A further increase in the cutoff wavelength is achieved by incorporating As into the InTlSb alloy. This indicates that there is a strong bowing effect in InTlAsSb systems similar to the case of InAsSb. A cutoff wavelength as long as 10.8 μm is obtained from InTlAsSb at 77K. These results were the first observation of IR photoresponse at such a long wavelength from the III-V quaternary alloys. The growth conditions for the sample are listed in Table 7.15.

The InTlAsSb sample with a 10.8-μm cutoff wavelength (corresponding to a bandgap of 0.115 eV) exhibited a lattice mismatch of –0.86% to InSb. This is in

Table 7.15 Growth Parameters for InTlAsSb

Growth pressure	76 Torr
Growth temperature	455°C
Total H_2 flow rate	1,500 cm^3/min
TMIn flow rate	3.75 μmol/min
CpTl flow rate	0.00043 μmol/min
TMSb flow rate	35.74 μmol/min
AsH_3 flow rate	9.38 μmol/min
Growth rate	0.7 μm/hr

contrast with the InAsSb system, which has 0.175-eV bandgap energy with the same mismatch to InSb. Note that InAsSb has a –2.27% lattice mismatch to InSb when the bandgap is at a minimum ($InAs_{0.65}Sb_{0.35}$: 0.136 eV at 77K). The present result indicates that the InTlAsSb alloy can exhibit longer cutoff wavelengths than can the InAsSb system while maintaining less lattice mismatch to InSb. Realization of lattice-matched InTlAsSb with InSb would improve device performance significantly by reducing the inevitable generation of misfit dislocation.

LWIR photoconductivity up to 15 μm was demonstrated for the first time. These measurements were the first observation of a room temperature band-edge photoresponse at such a long wavelength for the III-V quaternary alloys.

The electrical properties of the InTlAsSb films with thickness of 1.4 μm were investigated by Hall measurements. The Hall coefficients for the undoped samples were negative at room temperature. The room temperature mobility decreases from 16,100 to 6,190 $cm^2/V \cdot s$ as the flow rate increases. The molar flow rates of TMIn, CpTl, and TMSb were fixed. The decrease of mobility may be due to the increase of impurity and random alloy scattering. With an increase in the As flow rate, the 300K electron concentration increased from 8.8×10^{16} to 3.8×10^{17} cm^{-3} at 300K. For the InTlAsSb sample with a 10.8-μm cutoff wavelength, the electron concentration was 7.6×10^{16} and 1.6×10^{17} cm^{-3} at 77 and 300K, respectively.

7.7 InAsSb/InAsSbP for IR Lasers

The most thoroughly developed III-V semiconductor for MIR lasers is based on the InAsSb/InAsSbP quaternary alloy grown on InAs substrate. This material system is a promising candidate because it can cover the 3- to 5-μm region by adjusting its alloy composition. Additionally, the InAsSbP material system has a fairly large growth window. The state-of-the-art growth, characterization, and operating characteristics for interband lasers based on the InAsSbP material system are presented and reviewed throughout this section.

InAsSbP is a quaternary alloy comprised of InAs, InSb, and InP binary materials. The physical properties of these binaries are summarized in Table 7.16. InAsSbP has a very wide range of energy gap and lattice constant values. Two binary substrates can be matched with the quaternary alloy: InAs and GaSb. However, InAs substrates are preferred because they are less expensive and of higher quality than GaSb substrates. Furthermore, GaSb has a relatively large refractive index that will reduce the optical confinement of the laser structures.

Fukui and Horikoshi [76] reported the first growth of InAsSbP on InAs substrates by MOCVD in 1980. They also reported a miscibility gap for compositions with less than 30% arsenic. Other groups have also demonstrated high-quality InAsSbP materials with energy gaps close to theoretical predictions. However, as the arsenic content is reduced, transmission electron micrographs show compositional fluctuations. This has been interpreted as compositions that lie within the miscibility gap. Currently it is accepted that the miscibility gap extends to compositions with less than 20% arsenic.

Table 7.16 Physical Properties of InAs, InSb, and InP

	T (K)	InAs	InSb	InP
Crystal structure		Cubic (ZnS)	Cubic (ZnS)	Cubic (ZnS)
Lattice constant (Å)	300	6.0584	6.4788	5.8688
Energy gap (eV)	300	0.36	0.18	1.35
	0	0.41	0.36	1.42
Melting point (K)		1,215	803	1335
Coefficient of thermal expansion ($10^{-6}K^{-1}$)	300	4.52	5.04	5.0
m_e^*/m_0	4	0.024	0.014	0.077
m_{lh}^*/m_0	4	0.025	0.018	0.12
m_{hh}^*/m_0	4	0.37	0.4	0.55
m_{sos}^*/m_0	4	0.14	0.44	0.12
Electron mobility ($cm^2/V \cdot s$)	300	3×10^4	8×10^4	5×10^3
	77	8×10^4	10^6	4×10^4
Hole mobility ($cm^2/V \cdot s$)	300	200	850	150
	77	500	10^4	1200
Intrinsic carrier concentration (cm^{-3})	300	9×10^{14}	2×10^{16}	1×10^8
Static dielectric constant		15.15	16.8	12.5
High-frequency dielectric constant		12.25	15.68	9.61
Elastic moduli (10^{12} dyn cm^{-2})	300			
C_{11}		8.329	6.669	10.11
C_{12}		4.526	3.645	6.61
C_{44}		3.959	3.626	4.56
Refractive index		3.44	3.96	3.35
Optical phonon (cm^{-1})		242	193	320
LO		220	185	288
TO				

Source: [77].

7.7.1 Growth and Characterization of InAsSb and InAsSbP

Growth and Characterization of InAsSb

A low-pressure MOCVD was used for the growth of an InAsSbP material system. TMIn, TMSb, AsH_3, and PH_3 were used as precursors. For the growth of $InAs_{0.95}Sb_{0.05}$, the TMIn flow rate was kept constant at 185 cm^3/min. First, it has been found that the material characterization is optimized at 585°C. Second, the growth rate is not a strong function of temperature between 560°C and 605°C. Furthermore, the surface morphology of these samples is mirror-like with no defects visible under 100× magnification. The X-ray FWHM was used to optimize the growth rate. Growth rates of approximately 150 to 170 Å/min were used at growth temperature of 585°C.

The optimized growth conditions determined by the procedure above resulted in high-quality InAsSb. Its high structural quality is supported by the narrow FWHM of 45 arcsec from XRD and interference (Pendellosung) fringes that correspond to the epilayer thickness. The morphology is mirror-like with no defects resulting from growth. The surface topology has been measured with AFM and the long-range rms

surface roughness is approximately one monolayer. The atomic steps can be seen with AFM. The optical quality has been verified with photoluminescence. The FWHM at low temperatures is 25 meV.

Growth and Characterization of InAsSbP

The InAsSbP layers will be the cladding layers; their function is to provide electronic and optical confinement. It is most important that the cladding layer be lattice matched to the active region to avoid misfit dislocations that may act as nonradiative recombination centers. From this information, the optimal cladding layer composition is $InAs_{0.20}Sb_{0.30}P_{0.50}$. The composition of InAsSbP can easily be controlled via the flow rates of the sources. The optimized growth conditions resulted in high-quality $InAs_{0.20}Sb_{0.30}P_{0.50}$. The morphology is mirror-like. The surface topology has been measured with AFM and the rms surface roughness is approximately the same as InAsSb or about one monolayer. The optical quality and composition have been verified with photoluminescence. As a result of the long diffusion length for photoexcited carriers and the lower energy gap of the InAs substrate, many of the carriers diffuse and recombine in the InAs layer. Therefore, the InAsSbP epilayer and the InAs substrate have similar emission intensities.

Doping

TeSn is used as *n*-type dopant. The bubbler pressure and temperature for TeSn were 700 Torr and –10°C. DEZn is used as a *p*-type dopant. The bubbler pressure and temperature for DEZn were 700 Torr and –5°C. *Electrochemical capacitance-voltage* (ECV) was used to obtain the carrier concentration profile. The temperature and flow rate dependence on the carrier concentration have been obtained.

7.7.2 Strained-Layer Superlattices

Despite the need for coherent emitters in the 4- to 5-μm wavelength region, additional increases in emission wavelength have been found to be difficult with conventional interband emitter (DH and MQW) structures. For DH lasers emitting at $\lambda > 3.4$ μm, the misfit dislocations between the $InAs_xSb_yP_{1-x-y}$ cladding and $InAs_xSb_{1-x}$ active region are increased to levels that prohibit oscillation. MQW emitters have demonstrated the ability to emit spontaneously covering the entire 3- to 5-μm region. This is a result of the fact that MQW structures can accommodate a larger percent mismatch between active and barrier regions than DH structures as long as the thickness of the QW is thinner than the critical thickness. The critical thickness of $InAs/InAs_xSb_{1-x}$ has been calculated using the Matthews-Blakeslee and People-Bean relationships [78, 79]. The experimental results are much better fit to the People-Bean relationship.

Also, the emission wavelength for bulk InAsSb shows that around 20% antimony must be incorporated into a bulk active region for emission up to 5 μm. From our earlier discussion, we know the critical thickness of $InAs_{0.8}Sb_{0.2}$ is approximately

100Å. However, using these dimensions, the electrical and optical confinements and gain inherent in the active region are not enough to allow lasing in a MQW structure. This has been verified experimentally.

Strained-layer superlattice (SLS) lasers, which benefit from better optical confinement and more emission wavelength flexibility than is offered by a DH structure and a larger gain region than lasers based on a MQW active region, are good candidates for high power and high-temperature emission between 3.8 and 5 μm. High-temperature operation is enhanced with SLS lasers by the suppression of Auger recombination mechanisms with the large number of interfaces and compressive strain inherent with these lasers [80].

As an example, Figure 7.11 shows the calculated layer thickness (the well thickness is equal to the barrier thickness) dependence of emission wavelength for an $InAs_{0.95}P_{0.05}/InAs_{0.75}Sb_{0.25}$ SLS structure that has a 0.8% mismatch from InAs. The Kronig-Penney model, including the strain-induced band-edge shifting, has been used for the calculation. The flat dotted line represents the emission wavelength of an InAsSb alloy with 0.8% mismatch from InAs, which can only have an emission wavelength at $\lambda = 3.7\ \mu$m. Alternatively, by changing the layer thickness in the superlattice, the emission wavelength can be tuned from 3.3 to 4.5 μm independent of its lattice constant. Therefore, the SLS structure allows the structure to have a longer emission wavelength (> 4 μm) with a nonprohibitive mismatch from InAs substrate.

Three superlattice material alloys have the ability to emit between 3.8 and 5 μm: $InAs_{1-x}Sb_x/InAs_{1-y}P_y$, $InAs/InAs_{1-x}Sb_x$, and $InAs_{1-x}Sb_x/InAs_{1-y}Sb_y$.

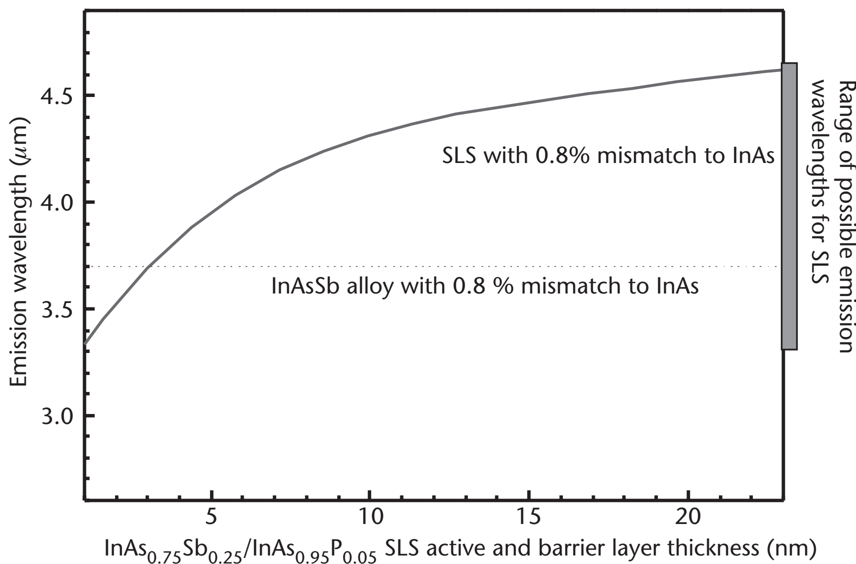


Figure 7.11 Calculated layer thickness (in our design the well thickness is equal to the barrier thickness) dependence of emission wavelength for an $InAs_{0.95}P_{0.05}/InAs_{0.75}Sb_{0.25}$ SLS structure that has a 0.8% mismatch from InAs.

7.7.3 Device Results

InAsSb/InAsSbP/InAs DH Lasers

DH InAsSbP/InAsSb/InAs were grown on (100)-oriented Te-doped InAs substrates by LP-MOCVD [81]. The heterostructure consists of a 1.0-μm-thick undoped ternary active region $InAs_{0.92}Sb_{0.08}$ sandwiched between two 1.5-μm-thick InAsSbP confinement layers: Sn doped (~10^{18} cm^{-3}) and Zn doped (~10^{18} cm^{-3}), respectively, and a p^+ -doped InAs cap layer.

Broad-area laser diodes with a stripe width of 100 μm were fabricated using a liftoff process, in which the p^+-InAs cap layer is etched away between the stripes. After ohmic contact deposition, individual diodes were cleaved and mounted *p*-side up on copper heat sinks by indium bonding. Uncoated diodes with cavity lengths varying from 200 to 1,200 μm were prepared.

InAsSb/InAsP/InAs SLS Lasers

SLS laser structures were grown on (100)-oriented Te-doped InAs substrates in a vertical metal MOCVD reactor [82]. The active region consisted of an $InAs_{1-x}Sb_x/InAs_{1-y}P_y$ superlattice with $x \approx 0.75$ and $y \approx 0.05$. The barrier and the well of the superlattice have the same thickness. The total period thickness is 100Å and 80 periods were used. For the entire laser structure, 1.1-μm InAsSbP cladding layers were grown before and after the superlattice growth. These layers were grown with the same doping profile as those used for DH lasers at the optimized temperature of 585°C. Because the composition of both ternary and quaternary materials changes with temperature, InAs was grown when the temperature was ramped between the cladding and superlattice. At the conclusion of the *n*-type InAsSbP cladding and InAsP/InAsSb superlattice region, InAs was grown while the temperature was ramped linearly between 585°C and 560°C. The temperature was stabilized for 2 minutes after the ramping period. A growth rate of 100 Å/min was used to minimize the total thickness of InAs. The structure was completed with a 100-nm p^+ InAs cap layer and is shown schematically in Figure 7.12.

The morphology of the laser structure was mirror-like with misfit dislocations similar to those for DH lasers. The XRD spectra have been measured and compared

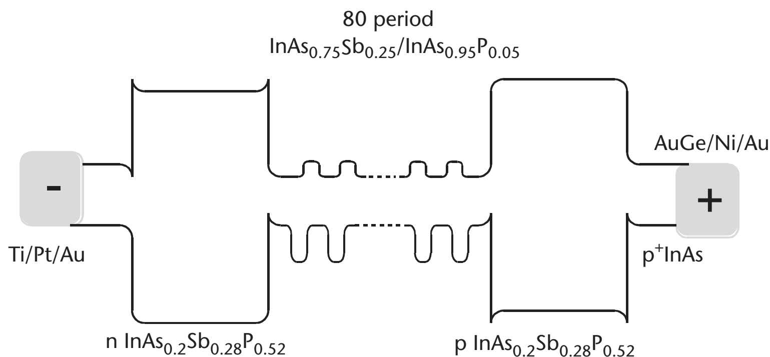


Figure 7.12 Schematic band alignments of a InAsSb/InAsP SLS injection laser.

with theoretical calculations of the spectra. The high crystalline quality of the strained superlattice structure is confirmed via X-ray measurements and by the three orders of satellite peaks observed in the spectra. These measurements were used to determine the average strain of the superlattice to be 0.79%. This is compared to the theoretically predicted mismatch of 0.78%. The peaks are well aligned and are periodic, which demonstrates that the period is the designed thickness and is consistent throughout the growth. The measured period from the superlattice satellite peaks is 98Å while the design period was 100Å. However, the FWHM is far greater for the measured spectra, which is most likely due to an interfacial layer in between the two layers. The photoluminescence shows a peak emission wavelength at 3.95 μm with a narrow FWHM of only 29 meV at 77K.

Broad-area laser diodes with a stripe width of 100 μm were fabricated using a liftoff process, in which the p^+ InAs contact cap layer is etched away between the stripes. After ohmic contacts were deposited, individual diodes were cleaved and mounted *p*-side up on copper heat sinks by indium bonding. Uncoated laser diodes of cavity lengths varying between 400 and 1,800 μm were prepared, mounted in a cryostat and then measured.

InAsSb/InAs(Sb)/InAs SLSs

The InAsSb/InAs, $InAs_{1-x}Sb_x/InAs_ySb_{1-y}$ superlattice active regions do not suffer from the problems resulting from the InAsP and have the advantages of large emission tunability because of the type II band alignment of InAsSb and InAs. However, these two superlattice systems have disadvantages. First, associated with the type II alignment is a reduced wave function overlap, resulting in a lower optical matrix element and finally a reduced radiative transition probability. Additionally, since InAsSb is compressively strained to InAs, the superlattices will also be strained.

For InAsSb/InAsP SLS high values of x are required because of the large conduction band offset that InAsP creates in the material. Because InAs and InAsSb are used as the barrier materials in InAsSb/InAsSb and InAs/InAsSb superlattice regions, much smaller values of x are required. This will allow for higher growth temperatures that reside closer to the optimum growth temperature. Calculations show that for superlattices with the same thickness for well and barrier, x is no greater than 15%. Furthermore, for both of the active region superlattices, 100-Å-thick layers will be grown. This ensures that they are below the critical thickness and that the number of misfit dislocations is kept to a minimum.

Despite the fact that the growth temperature is closer to the InAsSb optimum growth temperature, the switching routine is more critical for these superlattices than for the InAsSb/InAsP superlattices. Because antimony has a long residence time during the growth process, a balance must be made between the growth interruption or change in antimony flow rates. Too long an interval translates into complete antimony removal, which results in an arsenic-rich InAsSb interface. Too short a growth interruption will result in nonabrupt interfaces. The growths of the barriers for both of the superlattices focused on are routine. The barriers are either InAs or InAsSb with an antimony composition similar to that used for the active region for DH lasers.

Broad-area laser diodes with a width of 100 μm were fabricated. Optical output power-injection current characteristics were measured in both CW and pulsed mode operation for lasers mounted inside a cryostat with temperatures varying between 77 and 140K. Finally, the results for DH and SLS lasers with emissions from 3.2 to 4.7 μm are listed in the Table 7.17.

7.8 GaSb/InAs Type II Superlattice for IR Photodetectors

7.8.1 Introduction

Recently, type II superlattices have been proposed as an alternative to HgCdTe for the strategic windows of the 3- to 5-μm and 8- to 12-μm regions as well as longer wavelengths. Type II superlattices are constructed from the III-V material system and, hence, they have much better mechanical properties and material uniformity than HgCdTe. The electron effective mass in these superlattices is higher than the electron effective mass in HgCdTe and InAsSb, therefore the tunneling current is less. The bandgap of the superlattice can be changed to cover a wide IR range (from ~2 μm to above 25 μm) by changing the thickness of the layers rather than the material composition. Also, in comparison to QWIPs, type II superlattices have a much higher quantum efficiency and they do not require surface gratings for absorption of normal incident light.

Type II band alignment and some of its interesting physical behaviors were originally suggested by Sai-Halasz and Esaki [83] in 1977. Soon after that they reported the optical absorption of type II superlattices [84] and later the semimetal behavior of the superlattice [85]. The applications of such a superlattice were proposed several years later [86]. The flexibility of the material to cover a huge IR range (2 to 50 μm) and the possibility of a reduced Auger recombination rate [87] drew the attention of many groups. Type II heterojunctions have been found many applications in electronic devices such as *resonant tunneling diodes* (RTDs) and hot electron transistors. However, perhaps the most important applications have been in optoelectronics. Recently, many significant results have been achieved from type II

Table 7.17 Results for DH and SLS Lasers with Emissions from 3.2 to 4.7 μm

λ (μm)	*Material Systems*	*Pulsed Output Power (mW)*	*CW Output Power (mW)*	J_{th} (Acm^{-2})	T_0 (K)	J_0 ($Åcm^{-2}$)	η_d	α_i (cm^{-1})	R_{series} (Ω)	T_{max} (K)	$\theta_\perp$ (°)
3.2	InAsSb/InAsSbP/InAs DH	1600	450	40	42	—	18% @ 77K	—	0.23	220	—
3.4	InAsSb/InAsSbP/AlAsSb/InAs DH	1,880	360	60	54	—	34	—	—	—	40
3.8–4.0	InAs-InAsSb/InAsSbP/InAs SLS	1,000	300	67	15	47	38	16	0.3	120	34
4.0	InAsP-InAsSb/InAsSbP/InAs SLS	546	94	100	27	—	30	—	—	125	28
4.2	InAsSb-InAsSb/InAsSbP/InAs SLS	183	64	160	40	—	20	—	0.3	140	43
4.4	InAsSb-InAsSb/InAsSbP/InAs SLS	199	—	160	16	—	—	—	—	—	—
4.7	InAsSb-InAsSb/InAsSbP/InAs SLS	92	—	440	—	—	10	—	—	—	—

modulators [88], detectors [89, 90], and laser diodes [91, 92]. The active layers of photovoltaic and photoconductive type II detectors are made from superlattices with a type II band alignment. Similar to a type I superlattice, the allowed energy states form “mini bands,” due to the coupling of electrons and holes in adjacent wells. However, unlike type I superlattices, the bandgap of type II superlattices can be adjusted from nearly 0.7 eV to virtually zero. The minimum of the mini bands of such a superlattice in momentum space are at zero value and, hence, the superlattice resembles a direct gap semiconductor. Therefore, knowing the band structure and the optical absorption in type II superlattices, one can practically use the conventional photovoltaic and photoconductive structures to realize high-performance type II detectors.

The numerical modeling of the superlattice energy bands has been carried out by using the *envelope function approximation* method. Figure 7.13 shows the calculated mini band energy profile of type II superlattices with equal InAs and GaSb layers as a function of the period of the superlattice. As an example, a 60Å period will give an emission wavelength around 4 μm. Figure 7.14 shows that a band engineered type II superlattice shows Auger resonance at large $k_\perp$ values, which leads to suppression of Auger recombination.

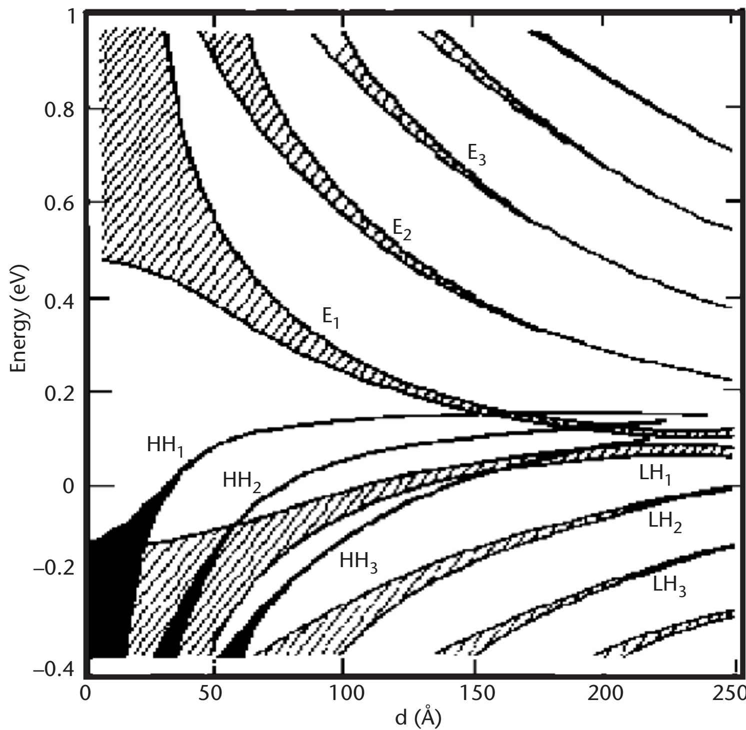


Figure 7.13 The mini band energy profile of type II superlattices with equal InAs and GaSb layers versus the period of the superlattice.

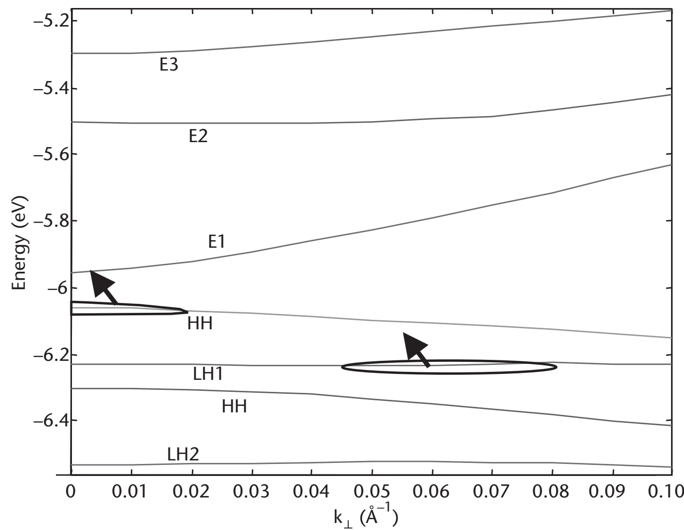


Figure 7.14 The band structure of a band engineered type II superlattice shows Auger resonance at large $k_\perp$ values, which leads to suppression of Auger recombination.

7.8.2 Experimental Results for Type II Photodetectors

Uncooled Type II Photoconductor in the LWIR Range

Growth

The InAs/GaSb type II superlattices were grown by MBE on semi-insulating GaAs substrates [93]. The reactor is an Intevac Modular Gen II MBE machine with uncracked As and Sb, and elemental Ga, In, and Al source material. A 4-μm GaSb buffer layer was grown directly on 3-in. GaAs substrates. The wafer was then broken into about 1-cm^2 pieces and indium-mounted to molybdenum blocks. InAs is found to have a very narrow window for planar growth, whereas high-quality GaSb can be grown in a wider range of growth conditions when RHEED showed a 1 × 3 reconstruction pattern. The optimum growth conditions for InAs layers were found to be T = 400°C according to a pyrometer, a V-to-III incorporation rate ratio of ≈ 3, and a growth rate of 0.5 ML/sec. In this condition, RHEED showed 2 × 4 reconstruction patterns. The pyrometer is calibrated with the temperature of the transition from a 1 × 5 to a 1 × 3 reconstruction pattern in the GaSb buffer layers. Based on the theoretical modeling and simulation, the optimum structure for a room temperature detector at λ = 11 μm was chosen. The structure consisted of a 2-μm superlattice with 48Å InAs, 30Å GaSb, and one monolayer of InSb at the interfaces because it is shown to improve the optical and electrical quality of the superlattice [94]. Finally, the superlattice was capped with a thin 200Å GaSb layer.

Characterization

High-resolution XRD was used to investigate the structure of the material. XRD simulation has also been performed to verify the superlattice structures. Sharp

X-ray peaks up to the fifth-order satellite were observed, which is an indication of high crystal quality and superlattice uniformity in the samples, despite more than a 6% mismatch between the lattice constant of the superlattices and the GaAs substrates.

The mismatch between the average lattice constant of the SL and the lattice constant of the GaSb buffer layer is about 0.3%. Although high-quality SLs with thicknesses of about 0.5 μm can be grown with such mismatch, the mismatch has to be reduced for thicker superlattices. Although the average lattice constant of the superlattice can be adjusted by changing the thickness of the InAs layer, this will also change the bandgap of the SL. Interface modification provides a better method to change the average lattice constant of the superlattices with a little effect on their bandgap. The interfaces can be either InSb type or GaAs type. Despite the fact that these layers are very thin (about one monolayer), they have a considerable effect on the average lattice constant of the SL, since the mismatch between InSb and GaSb is about +0.6% and the mismatch between GaAs and GaSb is –0.6%. Considering the thickness of the interfaces, a precise adjustment of these layers is obviously a difficult task. We managed to routinely grow SLs with very low mismatch to the GaSb buffer layers by utilizing two techniques. The atomic beam fluxes Ga, In, Sb, and As were calibrated with dynamic RHEED technique to below 1% error before each growth. Also, a low growth rate of about 0.5 ML/sec was used and, therefore, the deposition time of the interface was about 2 seconds, which is considerably longer than the actuation time of the shutters in our MBE system (~0.1 second).

An FTIR spectrometer was used to obtain the room temperature optical absorption spectra of the superlattice [95]. The effect of the substrate and GaSb buffer layer was removed by measuring the background with a substrate and GaSb buffer layer.

Because electrons and holes are confined in InAs and GaSb, respectively, the Hall measurement technique could not be use on superlattices because of the high sheet density of electrons and holes in the InAs and GaSb layers. The overall Hall coefficient for this material R_H, and conductance r, with two dominant channels of electrons and holes, can be approximated as follows [96]:

$$R_H = \frac{\dfrac{\mu_p \sigma_p}{1+\mu_p^2 B^2} - \dfrac{\mu_n \sigma_n}{1+\mu_n^2 B^2}}{\left(\dfrac{\sigma_p}{1+\mu_p^2 B^2} - \dfrac{\sigma_n}{1+\mu_n^2 B^2}\right)^2 + B^2\left(\dfrac{\mu_p \sigma_p}{1+\mu_p^2 B^2} - \dfrac{\mu_n \sigma_n}{1+\mu_n^2 B^2}\right)} \tag{7.16}$$

where B is the magnetic field in the Hall measurement, and σ and μ are the conductivity of each channel and the mobility of carriers. The formula clearly shows that the overall Hall mobility can be much smaller than the real mobility of electrons or holes if the numerator approaches zero. Because the transport of electrons is more important for photodetector operation due to its higher mobility, a single quantum well of InAs was grown to examine this property. This provided a simple, fast method for electrical assessment of the interface quality. After the optimization of

growth conditions, in-plane room temperature mobility of electrons in a 75-Å InAs well increased from 5,000 to 14,000 $cm^2/V \cdot s$, which is about half the value of bulk InAs.

Device Measurement

The photoconductor devices were prepared by making ohmic contacts either with indium annealing or aluminum deposition and etching. The samples were then mounted to a copper heat sink and attached to the cold finger of a liquid nitrogen cryostat equipped with a temperature controller. The spectral photoresponse of the device was measured using FTIR. The samples were illuminated through the front side with normal incidence. The absolute response of the photodetectors was calculated suing a blackbody test set. Figure 7.15 shows the responsivity of the device in the 2- to 17-μm wavelength range at 78 and 300K with an in-plane electrical field of 5 V/cm.

To assess the temperature dependence, the current responsivity of the device was measured at $\lambda = 10.6$ μm wavelength from 78K to room temperature at a constant electrical field. An allometric function (proportional to $T^{-1.93}$) fits well to the measured data. Usually the responsivity of the narrow gap materials is an exponential function of temperature at higher temperature where Auger recombination dominates the recombination mechanism. However, it has been showed theoretically by Zegrya and Andreev [97] that the Auger recombination rate is a power function of the temperature (proportional to T^2) in type II heterostructures compared to the exponential function in the bulk semiconductors. Because the current responsivity is proportional to the carrier lifetime, which is dictated by the Auger recombination rate, this power dependency of the responsivity indicates good agreement with the theoretical predictions.

The effective lifetime of the carriers extracted from the responsivity and Hall measurements on a $t = 0.5$-μm-thick superlattice is $\tau_{eff} = 26.8$ ns, which is about an

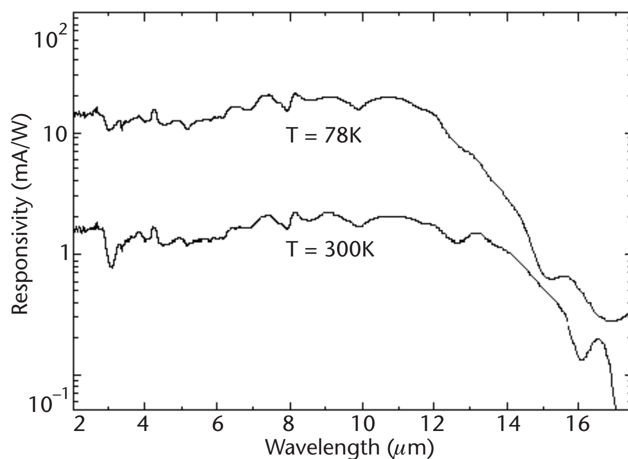


Figure 7.15 The responsivity spectra of the device at 78 and 300K with an in-plane electrical field of 5 V/cm.

order of magnitude longer than the carrier lifetime in HgCdTe photoconductors with similar bandgap and carrier concentration at room temperature [98]. Because Auger recombination is the dominant recombination mechanism at room temperature, we believe that the enhancement of carrier lifetime is due to the suppression of Auger recombination in the type II superlattice.

The noise was measured with an FFT spectrum analyzer and a low-noise, wideband preamplifier with a 54-dB voltage gain. The detector was biased by the preamplifier at V_b = 5V.

The mean-square noise of the detector can be modeled as

$$V_n^2 = V_J^2 + V_{1/f}^2 + V_{GR}^2 \tag{7.17}$$

where V_n is the overall noise of the detector, V_J is the Johnson-Nyquist noise, $V_{1/f}$ is the 1/*f* noise, and V_{GR} is the generation-recombination noise. The value of the Johnson noise can be calculated as follows:

$$V_{1/f}^2 = 4kTR\Delta f \tag{7.18}$$

where *k* is the Boltzmann constant, *T* is the temperature, and *R* = 76Ω is the resistance of the device. The value of Johnson noise for the device at room temperature is 1.12 nV/Hz$^{1/2}$. The 1/*f* noise can be approximated with this equation:

$$V_{1/f}^2 = V_{GR}^2 \frac{f_{1/f}}{f} \tag{7.19}$$

where $f_{1/f}$ is a constant that depends on the sample, and *f* is the frequency. This shows that at high enough frequencies, 1/*f* noise can be negligible compared to the other two types of noise. Then the value of the generation-recombination noise can be extracted from the total noise of the device in this range. The noise equivalent circuit model [99] was used to extract the noise of the photodetector as V_n = 1.7 nV/Hz$^{1/2}$ above 10 kHz. From the preceding equations, the value of the generation-recombination noise can be calculated as V_{GR} = 1.28 nV/Hz$^{1/2}$. The generation-recombination noise can be approximated as follows [100]:

$$V_{GR} = \frac{2V_b}{(lwt)^{1/2}} \frac{1+b}{bn+p} \left(\frac{np}{n+p} \frac{\tau \Delta f}{1+\omega^2 \tau^2} \right) \tag{7.20}$$

where w = 4 mm is the detector width, ω is the angular frequency and $\omega t << 1$ in this experiment, and b is the ratio of m_e to m_p. From the V_{GR} equation, the value of the carrier lifetime can be calculated as t = 17 ns, which is close to the value of the lifetime extracted from the optical response measurement.

The response time of the type II uncooled detectors was measured using a room temperature quantum cascade laser, developed at CQD at Northwestern University [101], as a narrowband, high-speed IR source in our measurement. The quantum cascade laser was uncooled and operated at λ = 8.5 μm with a negligible time delay.

With the time response measurement setup [102], it shows that the detector has a fall time below 40 ns.

The performance of IR detectors is usually compared based on their detectivity, which is an indication of their sensitivity. Knowing the responsivity and the noise of the devices, we can calculate their detectivity. The uncooled devices show a detectivity of about 1.3×10^8 cmHz^{½}/W at $\lambda = 11\ \mu$m, which is higher than the detectivity of commercially available HgCdTe at similar wavelengths and temperatures [103]. Unlike HgCdTe, these type II superlattices are grown on GaAs substrates and, hence, highly uniform material can be grown on 3- and 5-inch wafers readily. In comparison to the thermal detectors, such as microbolometers, type II superlattices have similar detectivity, but are showing at least 5 orders of magnitude faster response [79].

Realization of Type II Photoconductors in the VLWIR Range

High-performance IR detectors with a cutoff wavelength above 16 μm are highly needed for space-based applications such as deep-space astronomy and pollution monitoring. Currently available detectors with high quantum efficiency in this wavelength range are HgCdTe and extrinsic silicon detectors. However, due to the high nonuniformity of HgCdTe, detector arrays with acceptable uniformity can only be realized with extrinsic silicon detectors at long wavelengths. Although high detectivity and good uniformity have been achieved with this type of detector, such a detector has to be cooled below 10K. Consequently, a three-stage cryocooler is required that is heavy, bulky, and has a short lifetime. These drawbacks are especially important for space applications since they significantly increase the launch cost. Theoretical [104] calculations and our experimental results [105] show that InAs/$Ga_{1-x}In_xSb$ type II superlattices have an absorption coefficient similar to that of HgCdTe, and therefore detectors with very high quantum efficiencies are possible. However, unlike MCT, the uniformity of the growth is not an issue in this material system due to the strong bonding of III-V compound semiconductors.

Growth

The structures were grown on semi-insulating GaAs substrates. First, GaSb buffer layers with a thickness of about 2 μm were grown directly on the GaAs substrates. The growth was optimized for low background doping levels (~5×10^{15} cm^{-3}) and high surface smoothness (~4-Å rms roughness) in this layer. The superlattices consist of InAs and GaSb layers with one monolayer of InSb at both interfaces and were grown directly on GaSb buffer layers. The growth was terminated by a 100Å GaSb cap layer.

Device Measurement

Spectral photoresponse was measured using a FTIR spectrometer system. The samples were illuminated through the front side at normal incidence. Absolute response of the photodetectors was calculated using a blackbody setup. Figure 7.16 shows

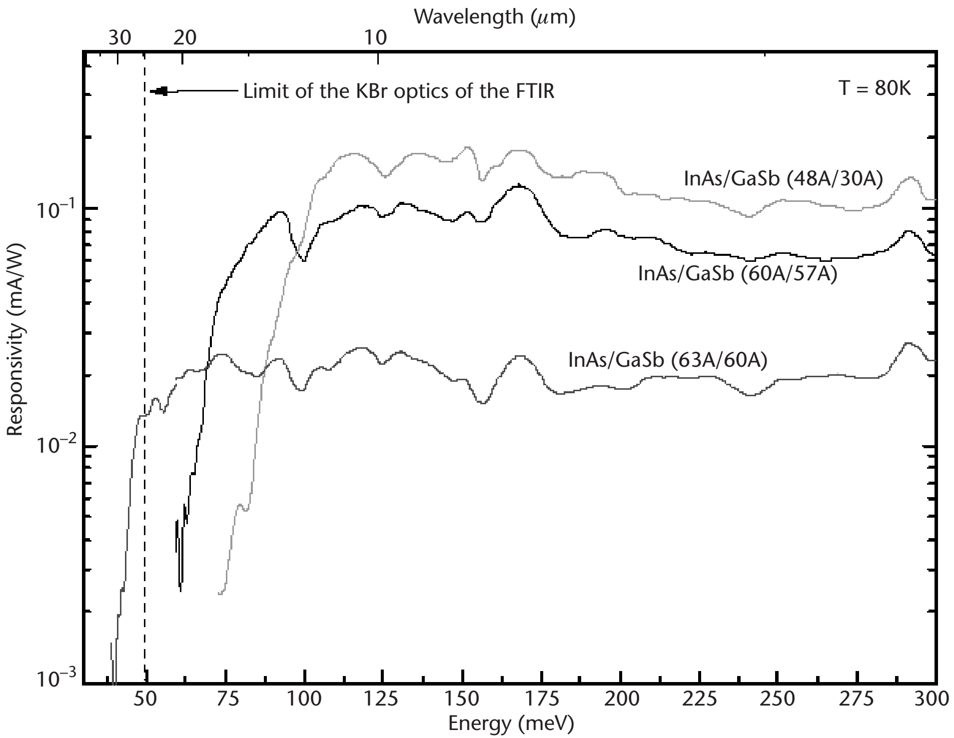


Figure 7.16 The spectral responsivity of three superlattices with cutoff wavelengths from 12 to 25 μm at 80K.

the spectral responsivity of three devices based on three different superlattice structures covering a 12- to 25-μm range.

One of the most important issues in the detectors with long cutoff wavelength is the uniformity of the material, which is a crucial requirement for high-resolution FPA where the device area can be several square centimeters. Comparison of the relative spectral response for three samples grown at the center, middle, and the edge of a 3-inch block simultaneously shows that the 10% cutoff energy varied only about 5 meV from the center to the edge, proving excellent uniformity over a very large area [106].

Uncooled Type II Photodiodes in the LWIR Range

We demonstrated high-performance uncooled photoconductors in the long wavelength range based on type II superlattices previously; however, these devices are more suitable for single-element detectors and cannot be easily used for 2D *focal plane array* (FPA). Unlike a photoconductor, the current of a photodiode flows perpendicular to the surface; hence, it is scaled by the area of the device. Moreover,

photodiodes can operate even at zero bias, which not only reduces the bias and heat dissipation requirements significantly, but also eliminates the $1/f$ noise. This type of noise increases inversely with frequency, and can be the dominant source of the total noise for low-frequency applications such as IR imaging systems.

Growth

The structures were grown using cracked As and Sb sources. The cracking zone temperature for both cells was 900°C. The deposition rates of the material were calibrated with dynamic RHEED oscillation to within 1%. First, 1 μm of GaSb contact layer doped with Be ($N_A = 1 \times 10^{18}$ cm^{-3}) was grown. Then a stack of five devices was grown, each of which consisted of 20 periods of p-type InAs/GaSb:Be (39Å/40Å), 20 periods of nominally undoped InAs/GaSb (39Å/40Å), and 20 periods of n-type InAs:Si/GaSb (39Å/40Å) superlattices. GaSb layers in this superlattice had a graded doping from 1×10^{18} to 2×10^{17} cm^{-3}. In the n-type superlattice, InAs layers were doped with Si ($N_D = 2 \times 10^{18}$ cm^{-3}). The shutter sequences were designed such that both interfaces were InSb type. Finally, the growth was capped with a 0.01 μm InAs:Si (N_D = 1 × 10 18 cm^{-3}). The growth temperature was about 520°C for the GaSb, and 395°C for the superlattices according to a pyrometer calibrated with the surface reconstruction transition temperatures of GaSb (~390°C) and InSb (~380°C).

Device Measurement

Absolute spectral responsivity was calculated from the measured spectral response of the devices, using a FTIR spectroscopy system, and the device's photoresponse to a calibrated blackbody setup. The peak responsivity was R_i = 0.14 A/W at λ = 7 μm leading to a Johnson-noise-limited detectivity of $D^* = 1.2 \times 10^8$ cmHz$^{1/2}$/W at room temperature.

The operation of type II photodiodes under zero bias ensures that the main noise component is the thermal (Johnson) noise and $1/f$ noise is eliminated. Our experimental measurements indicate that even under a considerable reverse bias, type II detectors do not have a high-frequency $1/f$ noise. The device was under a –0.2V bias and illuminated by the chopped IR radiation of a blackbody. The chopper frequency was f_0 = 396 Hz, blackbody temperature and aperture diameter were T_{BB} = 800K and D_{BB} = 2.54 cm, respectively, and the detector was located d = 15 cm apart from the blackbody aperture. Although the measurement includes the noise of the preamplifier and the FFT spectrum analyzer, the knee of the $1/f$ noise is below ~100 Hz. Under the given parameters, the SNR was more than 44 dB with a bandwidth of Δf = 100 Hz around f_0 = 396 Hz.

A QCL, operating at room temperature [107], was used as a high-speed source of IR radiation at λ = 5 μm to study the response time of the uncooled type II devices. Considering the fall time of the preamplifier (~50 ns) and the fall time of the output signal of the preamplifier (~110 ns), the detector response time was about $(110^2 - 50^2)^{1/2}$ ~100 ns using the sum-of-squares approach.

Realization of Type II Photodiodes in the VLWIR Range

The previous experimental results suggest that type II superlattices have superior uniformity compared with the available interband detectors in the VLWIR range. However, realizing that the material quality cannot be enhanced on a lattice mismatch substrate, we started the work on GaSb substrates. Photovoltaic devices were the natural choice due to the high conduction of the undoped substrate as well as the benefits of the photovoltaic devices for imaging applications.

Growth

The material is grown with As and Sb valved cracker sources on p-type GaSb substrates [108]. The photodiode structures were grown at 395°C according to a calibrated pyrometer. First, a 1-μm GaSb buffer/contact layer doped with Be ($p \sim 1 \times 10^{18}$ cm^{-3}) was deposited. Next a 0.5-μm-thick InAs/GaSb:Be ($p \sim 1 \times 10^{18}$ to 3×10^{17} cm^{-3}) superlattice was grown, followed by a 2-μm-thick nominally undoped superlattice. Finally, a 0.5-μm-thick InAs:Si/GaSb ($n \sim 1 \times 10^{18}$ cm^{-3}) superlattice was grown and capped with a 100-Å-thick InAs:Si ($n \sim 2 \times 10^{18}$ cm^{-3}) top contact layer. The growth rate was 0.5 ML/sec for InAs layers and 0.8 ML/sec for GaSb layers. The V/III *beam-equivalent pressure* (BEP) ratio was about 4 for InAs layers and about 1.2 for GaSb layers. The cracker temperature for As and Sb cells was 800°C. The selected thicknesses of the InAs and GaSb layers were determined for specific cutoff wavelengths using a four-band superlattice $\mathbf{k} \cdot \mathbf{p}$ model. For devices with a cutoff wavelength of nearly 16 μm, the thickness of the InAs layers was 54Å and the thickness of the GaSb layers was 40Å.

Characterization

The structural quality of the epitaxial layers was assessed using high-resolution XRD. The mismatch between the average lattice constant of the superlattice and the GaSb substrate was below 0.06%, while the FWHM of the satellite peaks was below 50 arcsec for the grown devices.

The surface morphology of the samples was also studied with AFM. Wide atomic steps are visible, which indicates excellent surface smoothness. Samples with a rms surface roughness below 4Å over a 20-μm × 20-μm area can be grown routinely, which is among the best reported values for this material system.

Device Measurement

The samples were then processed into 400-μm × 400-μm mesas using standard lithography and wet etching. Ti/Au contacts were defined for top and bottom contacts with metal evaporation and liftoff techniques. The samples were then indium bonded to a copper heat sink and attached to the cold finger of a liquid nitrogen cryostat with KRS-5 windows.

Figure 7.17 shows the measured and modeled current densities versus the applied bias for devices with a cutoff wavelength of $\lambda_c = 16$ μm. The calculated current density, which consists of tunneling, generation-recombination, and diffusion current densities, shows good agreement with the measured values for forward and reverse biases [85].

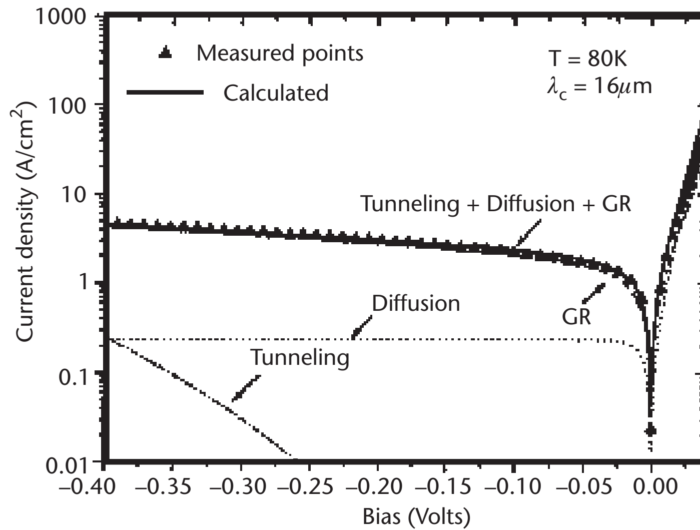


Figure 7.17 The measured and modeled dark current density of a device with λ_c = 16 μm at T = 80K. The model consists of the tunneling, generation recombination, and diffusion components of the dark current and indicates that generation-recombination is the dominant source of the dark current around zero bias.

Also, we could observe the Wannier-Stark oscillation in the Zener tunneling current of these devices. Figure 7.18 shows the differential resistance of the device versus reverse voltage bias. An oscillation with a peak-to-peak separation of about 29.9 mV is clearly visible. The thickness of the depleted layer was extracted to be d_p

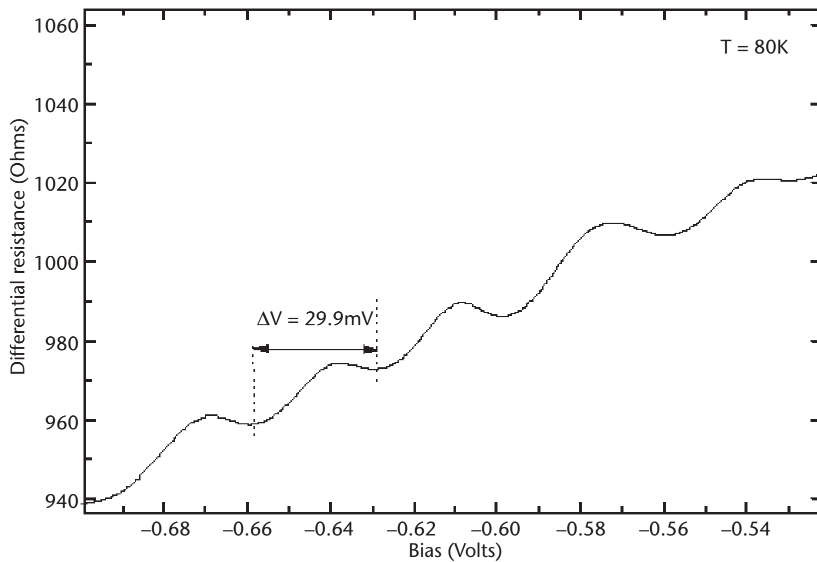


Figure 7.18 The differential resistance of a device with λ_c = 16 μm at T = 80K versus reverse bias shows a clear oscillation due to the Wannier-Stark oscillation in the Zener tunneling current.

= 120 nm, and the background doping level to be $p = 2.1 \times 10^{15}$ cm^{-3}. This value shows excellent agreement with the calculated background doping level based on the dark current modeling. This agreement and the fact that the background calculation based on the Wannier-Stark oscillation can be calculated without involvement of unknown parameters indicate that the above value must be very close to the real background level.

Absolute spectral responsivity was calculated from the measured spectral response of the device using a FTIR spectroscopy system and its photoresponse to a calibrated blackbody setup. Figure 7.19 shows the typical spectral responsivity of the detectors with λ_c = 16 μm. The absorption features of CO_2 and H_2O are due to the small difference in the optical pathlength of the background measurement and the detector measurement. The peak responsivity for the sample is about 3.5 A/W, which leads to a quantum efficiency of ~35% at 12 μm. Adding an antireflection coating can reduce the 30% surface reflection of the device such that the quantum efficiency will reach nearly 50%. The 90% to 10% cutoff energy of the responsivity is only about 13 meV at T = 80K, which is close to the thermal limit of $2kT$ = 12 meV. Such a sharp cutoff is a unique characteristic of our superlattices compared to the published results of other groups working on type II superlattice detectors [109–111] due to the use of binary materials, which are believed to have significantly enhanced the uniformity and reproducibility of the energy gap.

Acknowledgments

The authors would like to thank Yeun-ho Choi, Erick Michel, Jedon Kim, Jaejin Lee, Joseph Wojowski, Hooman Mohseni, Di Wu, Brett Lane, and Yajun Wei for their contribution to this work. The research at the Center for Quantum Devices and that presented here were supported by the U.S. Air Force (Gail Brown, William

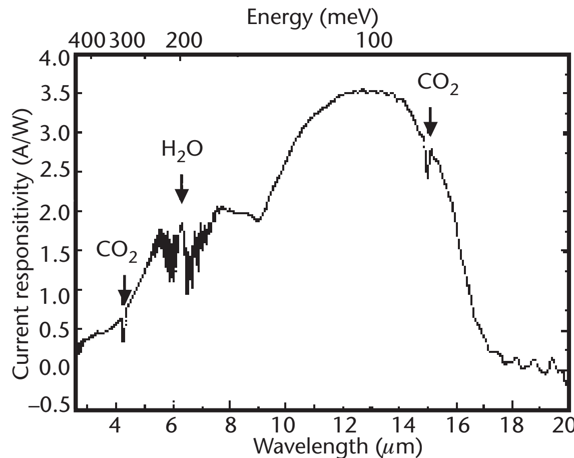


Figure 7.19 Absolute spectral responsivity of a device with λ_c = 16 μm. The 90% to 10% cutoff energy width is about 12 meV. The absorption features of CO_2 and H_2O are due to the small difference in the optical pathlength of the background measurement and the detector measurement.

Mitchel, Daniel Johnstone), the U.S. Army (Henry Everitt), the Office of Naval Research (Yoon-Soo Park), and the Defense Advanced Research Projects Agency (L. N. Durvasula, Ray Balcerak).

References

[1] Van Vechten, J. A., and T. K. Bergstresser, "Electronic Structures of Semiconductor Alloys," *Phys. Rev. B,* Vol. 1, 1970, pp. 3351–3358.

[2] Casey, H. C., and M. B. Panish, *Heterostucture Lasers*, New York: Academic Press, 1978.

[3] Ilegems, M., and M. B. Panish, "Phase Equilibria in III-V Quaternary System–Application to Al-G-P-As," *J. Phys. Chem. Solids,* Vol. 35, 1974, pp. 409–420.

[4] Jordan, A. S., and M. Ilegems, "Solid-Liquid Equilibria for Quaternary Solid Solutions Involving Compound Semiconductors in the Regular Solution Approximation," *J. Phys. Chem. Solids,* Vol. 36, 1974, pp. 329–342.

[5] Adachi, S., "Band Gaps and Refractive Indices of AlGaAsSb, GaInAsSb, and InPAsSb: Key Properties for a Variety of the 2–4-μM Optoelectronic Device Applications," *J. Appl. Phys.,* Vol. 61, No. 10, 1987, pp. 4869–4876.

[6] Goldschmidt, V. M., *Skrifter Norshe Vide. Akads. Oslo I: Mat. Naturv. Kl. VIII,* 1926.

[7] Heinz, C., and W. Schmidt, "Auf Altenstadt," *J. Crystal Growth,* Vol. 67, 1984, p. 393.

[8] Woelk, C., and K. W. Benz, *J. Crystal Growth,* Vol. 27, 1974, p. 177.

[9] Dutta, P. S., et al., *J. Crystal Growth,* Vol. 152, 1995, p. 14.

[10] Capasso, F., M. B. Panish, and S. Sumski, *IEEE J. Quantum Electron.,* Vol. QE-17, 1981, p. 273.

[11] Miki, H., K. Segawa, and K. Fujibayashi, *Jpn. J. Appl. Phys.,* Vol. 13, 1974, p. 203.

[12] Manasevit, H. M., and K. L. Hess, *J. Electrochem. Soc.,* Vol. 126, 1979, p. 2031.

[13] Cooper, C. B., R. R. Saxena, and M. J. Ludowise, *J. Electron. Mater.,* Vol. 11, 1982, p. 1001.

[14] Keuch, T. F., and E. Veuhoff, *J. Crystal Growth,* Vol. 68, 1984, p. 148.

[15] Chen, C. H., et al., *J. Electron. Mater.,* Vol. 22, 1993, p. 87.

[16] Shin, J., et al., *J. Crystal Growth,* Vol. 151, 1995.

[17] Kodama, M., and M. Kimata, *J. Crystal Growth,* 73, 1985, p. 641.

[18] Longenbach, K. F., and W. I. Wang, *Appl. Phys. Lett.,* Vol. 59, 1991, p. 2427.

[19] Newstead, S. M., T. M. Kerr, and C. E. C. Wood, *J. Appl. Phys.,* Vol. 66, 1989, p. 4184.

[20] Milnes, A. G., and A. Y. Polyakov, *Solid-State Electron.,* Vol. 36, 1993, p. 803.

[21] Boos, J. B., et al., *IEEE Trans. Electron Devices,* Vol. 45, 1998, p. 1869.

[22] Johnson, M., et al., *Appl. Phys. Lett.,* Vol. 71, 1997, p. 974.

[23] Chow, D. H., et al., *IEEE Electron Device Lett.,* Vol. 17, 1996, p. 69.

[24] Meyer, J. R., et al., *IEEE Proc. J: Optoelectron.,* 145, 1998, p. 275.

[25] Alibert, C., et al., *Phys. Rev. B,* Vol. 27, 1983, p. 4946.

[26] Zollner, S., et al., *J. Appl. Phys.,* Vol. 66, No. 1, 1989, pp. 383–387.

[27] Subbanna, S., G. Tuttle, and H. Kroemer, *J. Electron. Mater.,* Vol. 17, 1988, p. 297.

[28] Zhao, Y., M. J. Jurkovic, and W. I. Wang, *Solid-State Electron.,* Vol. 42, 1998, p. 57.

[29] Chow, D. H., et al., *J. Crystal Growth,* Vol. 150, 1995, p. 879.

[30] Bennett, B. R., et al., *J. Appl. Phys*, Vol. 87, No. 11, June 2000, pp. 7876–7879.

[31] Adachi, S., *J. Appl. Phys.,* Vol. 67, 1990, p. 6427.

[32] Walker, H., "Über Neue Halbleitende Verbindungen," *Z. Naturf.*, Vol. 7a, 1952, p. 744 (in German).

[33] Sharma, R. C., Y. L. Ngai, and Y. A. Chang, in *Binary Alloy Phase Diagrams,* Vol. 3, 2nd ed., T. B. Massalski, Ed., ASM International, 1990, p. 2286.

[34] Jung, Y. J., et al., "Electron Transport and Energy-Band Structure of InSb," *J. Appl. Phys.*, Vol. 69, 1991, p. 3109.

[35] Neuberger, M., *III-V Ternary Semiconducting Compounds Data Tables,* New York: Plenum, 1972.

[36] Parrish, W. J., et al., "Low-Coat High Performance InSb 256 × 256 Infrared Camera," *Proc. SPIE*, Vol. 1540, 1991, p. 274.

[37] Hoffman, A., and D. Randall, "High Performance 256 × 256 InSb FPA for Astronomy," *Proc. SPIE*, Vol. 1540, 1991, p. 297.

[38] Thom, R. D., et al., "A Fully Monolithic InSb Infrared CCD Array," *IEEE Trans. Electron Devices*, Vol. ED-27, 1980, p. 160.

[39] Stringfellow, G. B., *Organometallic Vapor-Phase Epitaxy: Theory and Practice,* San Diego, CA: Academic Press, 1999.

[40] Wagner, R. S., p. 47 in *Whiskers Technology*, A. P. Levitt, (ed.), New York: John Wiley & Sons, 1970.

[41] Buchan, N. I., C. A. Larsen, and G. B. Stringfellow, "Mass Spectrometric Studies of Trimethylindium Pyrolysis," *J. Crystal Growth,* Vol. 92, 1988, p. 591.

[42] Larsen, C. A., S. H. Li, and G.B. Stringfellow, "Decomposition Mechanics of Trimethylantimony and Reactions with Trimethylindium," *Chem. Mat.*, Vol. 3, 1991, p. 39.

[43] Michel, E., et al., "The Molecular Beam Epitaxial Growth of InSb on (111)B GaAs," *Appl. Phys. Lett.*, Vol. 69, 1996, p. 215.

[44] Bloom, I., and Y. Nemirovsky, *IEEE Trans. Electron. Devices,* Vol. 40, 1993, p. 309.

[45] Choi, Y. H., et al., "High-Quality InSb Growth on GaAs and Si by Low-Pressure Metalorganic Chemical Vapor Deposition," *Mat. Res. Soc. Symp. Proc.*, Vol. 281, 1993, p. 375.

[46] Besikci, C., et al., "Anomalous Hall Effect in InSb Layers Grown by Metalorganic Chemical Vapor Deposition on GaAs Substrates," *J. Appl. Phys.* Vol. 73, 1993, p. 5009.

[47] Soderstrom, J. R., et al., "Molecular Beam Epitaxy Growth and Characterization of InSb Layers on GaAs Substrates," *Semicon. Sci. Technol.*, Vol. 7, 1992, p. 337.

[48] Yano, M., T. Takase, and M. Kimata, "Heteroepitaxial InSb Films Grown by Molecular Beam Epitaxy," *Phys. Stat. Sol. A,* Vol. 54, 1979, p. 707.

[49] Wieder, H. H., and A. R. Clawson, "Photo-Electronic Properties of $Inas_{0.07}Sb_{0.93}$ Films," *Thin Sold Films,* Vol. 15, 1973, p. 217.

[50] Bethea, C. G., et al., "Photoconductance Measurements on Inas [Sub 0.22]Sb[Sub 0.78]/GaAs Grown Using Molecular Beam Epitaxy," *Appl. Phys. Lett.*, Vol. 53, 1988, p. 291.

[51] Chiang, P. K., and S. M. Bedair, "Metalorganic Chemical Vapor Deposition and Characterization of the In-As-Sb-Bi Material System for Infrared Detection," *Appl. Phys. Lett.*, Vol. 53, 1988, p. 142.

[52] Rogalski, A., and K. Jozwikowski, "Intrinsic Carrier Concentration and Effective Masses in $InAs_{1-X}Sb_x$," *Infrared Phys.*, Vol. 29, 1989, p. 35.

[53] Yen, M. Y., "Molecular-Beam Epitaxial Growth and Electrical Properties of Lattice Mismatched $InAs_{1-X}Sb_x$ on (100) GaAs," *Appl. Phys.,* Vol. 64, 1988, p. 3306.

[54] Coders, W. M., and J. C. Woolley, "Electrical Properties of $InAs_xSb_{1-X}$ Alloys," *Can. J. Phys.* Vol. 46, 1968, p. 1207.

[55] Chin, V. W. L., R. J. Egan, and T. L. Tansley, "Electron Mobility in $InAs_{1-x}Sb_x$ and the Effect of Alloy Scattering," *J. Appl. Phys.,* Vol. 69, 1991, p. 3571.

[56] Besikci, C., et al., "Detailed Analysis of Carrier Transport in $InAs_{0.3}Sb_{0.7}$ Layers Grown on GaAs Substrates by Metalorganic Chemical-Vapor Deposition," *J. Appl. Phys.,* Vol. 76, 1994, p. 5820.

[57] Kim, J. D., et al., "Long-Wavelength InAsSb Photoconductors Operated at Near Room Temperatures (200–300 K)," *Appl. Phys. Lett.,* Vol. 68, 1996, p. 99.

[58] Kim, J. D., et al., "8–13 µm InAsSb Heterojunction Photodiode Operating at Near Room Temperature," *Appl. Phys. Lett.,* Vol. 67, 1995, p. 2645.

[59] Michel, E., et al., "Sb-Based Infrared Materials and Photodetectors for the 3–5 μm and 8–12 μm Range," *Proc. SPIE,* Vol. 2685, 1996, p. 105.

[60] Schilfgaarde, M., and A. Sher, "InTlSb: An Infrared Detector Material?" *Appl. Phys. Lett.,* Vol. 62, 1993, p. 1857.

[61] Choi, Y. H., et al., *Appl. Phys. Lett.,* Vol. 63, 1993, p. 361.

[62] Huang, K. T., R. M. Cohen, and G. B. Stringfellow, "InTlSb Growth by MOVPE," *J. Crystal Growth,* Vol. 156, 1995, p. 320.

[63] Karam, N. K., et al., Growth and Characterization of InTlSb for Ir-Detectors, *J. Electron. Mater.,* Vol. 25, 1996, p. 1209.

[64] Choi, Y. H., et al., "Characterization of InTlSb/InSb Grown by Low-Pressure Metalorganic Chemical Vapor Deposition on a GaAs Substrate," *J. Appl. Phys.*, Vol. 75, 1994, p. 3196.

[65] Staveteig, P. T., et al., "Photoconductance Measurements on InTlSb/InSb/GaAs Grown by Low-Pressure Metalorganic Chemical Vapor Deposition, " *Appl. Phys. Lett.*, Vol. 64, 1994, p. 460.

[66] Bigan, E., et al., "InTlSb Alloys for Infrared Detection," *Proc. SPIE*, Vol. 2145, 1994, p. 2.

[67] Razeghi, M., et al., "Development of InTlSb Infrared Photodetectors Grown by LP-MOCVD," *Proc. Electrochem. Soc.*, Miami Beach, FL, October 1994.

[68] Razeghi, M., et al., "Exploration of Entire Range of III-V Semiconductors and Their Entire Device Applications," *Mat. Sci. Technol.*, Vol. 11, 1995, p. 3.

[69] Phillips, J. C., *Bonds and Bands in Semiconductors,* New York: Academic, 1973.

[70] Lee, J. J., and M. Razeghi, "Tl Incorporation in InSb and Lattice Contraction of $In_{1-x}Tl_xSb$," *Appl. Phys. Lett.*, Vol. 76, No. 3, 2000, pp. 297–299.

[71] Besikci, C., et al., "Detailed Analysis of Carrier Transport in $InAs_{0.3}Sb_{0.7}$ Layer Grown on GaAs Substrates by Metalorganic Chemical Vapor Deposition," *J. Appl. Phys.*, Vol. 76, 1994, p. 5820.

[72] Chin, V. W., R. J. Egan, and T. L. Tansley, "Electron Mobility in $InAs_{1-x}Sb_x$ and the Effect of Alloy Scattering," *J. Appl. Phys.*, Vol. 69, 1991, p. 3571.

[73] Harrison, J. W., and J. R. Hauser, "Theoretical Calculation of Electron Mobility in Ternary III-V Compounds," *J. Appl. Phys.*, Vol. 47, 1976, p. 292.

[74] Lee, J. J., J. D. Kim, and M. Razeghi, "Long-Wavelength Infrared Photodetectors Based on InSbBi Grown on GaAs Substrates," *Appl. Phys. Lett.,* Vol. 71, 1997, p. 2298.

[75] Lee, J. J., J. D. Kim, and M. Razeghi, "Room Temperature Operation of 8–12 μm InSbBi Infrared Photodetectors on GaAs Substrates" *Appl. Phys. Lett*. Vol. 73, 1998, p. 602.

[76] Fukui, T., and Y. Horikoshi, *Jpn. J. Appl. Phys.,* Vol. 19, 1980, p. L53.

[77] Landolt and Börnstein, *Numerical Data and Functional Relationships in Science and Technology: Semiconductors,* Berlin: Springer-Verlag, 1982.

[78] Matthews, J. W., and A. E. Blakeslee, *J. Crystal Growth,* Vol. 27, 1974, p. 118.

[79] People, R., and J. C. Bean, *Appl. Phys. Lett.*, Vol. 47, 1985, p. 322.

[80] Mohseni, H., V. Litvinov, and M Razeghi, *Phys. Rev. B*, 58, 1998, p. 378.

[81] Diaz, J., et al., "InAsSbP/InAsSb/InAs Laser Diodes (Lambda =3.2 μm) Grown by Low-Pressure Metal-Organic Chemical-Vapor Deposition," *Appl. Phys. Lett.*, Vol. 70, 1997, p. 40.

[82] Lane, B., et al., "InAsSb/InAsP Strained-Layer Superlattice Injection Lasers Operating at 4.0 m Grown by Metal-Organic Chemical Vapor Deposition," *Appl. Phys. Lett.,* Vol. 74, 1999, p. 3438.
[83] Sai-Halasz, G. A., R. Tsu, and L. Esaki, *Appl. Phys. Lett.,* Vol. 30, 1997, 651.
[84] Sai-Halasz, G. A., et al., *Solid State Commun.,* Vol. 27, 1978, p. 935.
[85] Sai-Halasz, G. A., and L. Esaki, *Phys. Rev. B,* Vol. 18, 1978, p. 2812.
[86] Smith, D. L., and C. Mailhiot, *J. Appl. Phys.,* Vol. 62, 1987, p. 2545.
[87] Youngdale, R., et al., *Appl. Phys. Lett.,* Vol. 64, 1994, p. 3160.
[88] Xie, H., and W. I. Wang, *J. Appl. Phys.,* Vol. 76, 1994, p. 92.
[89] Johnson, J. L., et al., *J. Appl. Phys.,* Vol. 80, 1996, p. 1116.
[90] Fuchs, F., et al., *Appl. Phys. Lett.,* Vol. 71, 1997, p. 3251.
[91] Yang, B. H., et al., *Appl. Phys. Lett.,* Vol. 72, 1998, p. 2220.
[92] Felix, C. L., et al., *Photon. Tech. Lett.*, Vol. 9, 1997, p. 734.
[93] Mohseni, H. et al., "Growth and Characterization of InAs/GaSb Photoconductors for Long Wavelength Infrared Range," *Appl. Phys. Lett.,* Vol. 71, 1997, p. 1403.
[94] Herres, N., et al., *Phys. Rev. B,* Vol. 53, 1996, p. 15688.
[95] Mohseni, H., et al., *Proc. SPIE,* Vol. 3287, 1998, p. 30.
[96] Kutchis, E. V., *Galvanomagnetic Effects and Methods of Their Investigations,* Moscow, Russia: Izd. Radiosvyaz, 1990.
[97] Zegrya, G., and A. Andreev, *Appl. Phys. Lett.,* Vol. 67, 1995, p. 2681.
[98] Piotrowski, J., W. Galus, and M. Grudzien, *Infrared Phys.,* Vol. 31, 1990, p. 1.
[99] Wang, D., et al., *J. Appl. Phys.,* Vol. 77, 1995, p. 1107.
[100] Long, D., *Infrared Phys.*, Vol. 7, 1967, p. 169.
[101] Slivken, S., et al., *Appl. Phys. Lett.,* Vol. 74, 1999, p. 173.
[102] Mohseni, H., and M. Razeghi, *Proc. ISDRS,* 1999, p. 563.
[103] Mohseni, H., J. Wojkowski, and M. Razeghi, *IEEE J. Quantum Electron.,* Vol. 35, 1999, p. 1041.
[104] Chow, D., et al., *Semicond. Sci. Technol.,* Vol. 6, 1991, p. C47.
[105] Mohseni, H., et al., *Appl. Phys. Lett.,* Vol. 71, 1997, p. 1403.
[106] Mohseni, H., et al., *Appl. Phys. Lett.,* Vol. 77, 2000, p. 1572.
[107] Slivken, S., et al., *Appl. Phys. Lett.,* Vol. 74, 1999, p. 173.
[108] Mohseni, H., et al., "High-Performance InAs/GaSb Superlattice Photodiodes for the Very Long Wavelength Infrared Range," *Appl. Phys. Lett.,* Vol. 78, 2001, p. 2107.
[109] Young, M., et al., *Appl. Surface Sci.,* Vol. 123/124, 1998, p. 395.
[110] Fuchs, F., et al., *Proc. SPIE,* Vol. 3287, 1998, p. 14.
[111] Miles, R., and J. Wilson, *Proc. SPIE,* Vol. 2816, 1996, p. 76.

CHAPTER 8

Growth, Structures, and Optical Properties of III-Nitride Quantum Dots

Daming Huang, Yi Fu, and Hadis Morkoc, Virginia Commonwealth University

8.1 Introduction

The study of zero-dimensional (0D) semiconductors has been the subject of intense investigation in recent years. Interest in such nanostructures is mainly a result of the unique application of their physical phenomena and potential device applications. Compared to bulk (3D) materials and 2D QW structures, quantum dots are the prototype of 0D systems and, as such, possess many unique properties. Semiconductor QDs are also of great interest for the study of basic quantum mechanical effects. The electronic states in a QD are spatially localized and the energy is fully quantized, similar to a single atom or atomic system. So the system is more stable against any thermal perturbation. In addition, due to the quantization, the electronic density of states near the bandgap is higher than in 3D and 2D systems, leading to a higher probability for optical transitions. Furthermore, the electron localization may dramatically reduce the scattering of electrons by bulk defects and reduce the rate of nonradiative recombination. These properties, among others, are directly relevant to the high thermal stability and high quantum efficiency of light-emitting and light-detecting devices, and they are of great importance in terms of device applications.

Semiconductor nitrides such as aluminum (AlN), gallium nitride (GaN), and indium nitride (InN) are very promising materials for their potential use in optoelectronic devices and high-power, high-temperature electronic devices. These materials and their ternary and quaternary alloys cover an energy bandgap range of 1.9 to 6.3 eV, suitable for band-to-band light generation with colors ranging from red to UV wavelengths. Nitrides are particularly suitable for application in surface acoustic wave devices, UV detectors, Bragg reflectors, waveguides, and UV and visible *light-emitting diodes* (LEDs). GaN-based LEDs and LDs have been achieved using thin films or QWs [1]. A flurry of interest in low-dimensional GaN and other III-nitrides is in part associated with the desire to develop new optoelectronic devices with improved quality and wider applications. Development of light emitters with QDs, for example, is expected to result in a lower threshold current in LDs and a higher thermal stability of the devices [2].

To access the advantages of QDs, many requirements must be met in material preparation. The most important requirements relate to the size and of its

distribution of QDs. Depending on the material and the dot shape, the maximum size should be near or less than some characteristic length of the electrons in the bulk III-nitrides, such as the exciton radius. The typical value is on the order of a few nanometers. With such a small size, the practical applications are thus often associated with a large assembly of QDs rather than a single one. This implies that the size uniformity of the dot assembly is critical. Fluctuations in dot size produce an inhomogeneous broadening in quantized energy levels and may destroy the very properties expected from a single QD. In addition to size uniformity, the spatial position of each QD is also important in many applications. The random rather than well-ordered distribution may damage the coherence of the optical and electronic waves propagating through the system. Similar to other semiconductor heterostructures, the surface or interface of the QDs must also be free of defects. Otherwise, the surface/interface may become the effective scattering center for electrons. Fabrication of a QD assembly with small and uniform size, high density, well-ordered placement, and defect-free materials today remains a serious challenge in any semiconductor system, especially when using III-nitride materials.

The properties of III-nitride QDs are closely related to those of bulk materials that have been reviewed in many articles [3, 4]. In Table 8.1 we list the material and physical parameters of GaN, AlN, and InN that are relevant to QD investigations. Although bulk GaN, AlN, and InN can all crystallize in wurtzite, zinc blende, and rock salt structures, in ambient conditions, the thermodynamically stable structure, however, is the wurtzitic phase. In this case, the GaN, AlN, and InN have direct room temperature bandgaps of 3.4, 6.2, and 1.9 eV, respectively. The bandgap, as well as the detailed electronic structure, can be modified by incorporating alloy and low-dimensional structures such as QWs and QDs. Depending on material, size, and shape, very rich electronic and optical phenomena can be observed from quantum structures that are not expected in bulk.

III-nitride QDs are commonly the strained systems. As listed in Table 8.1, the lattice constants for bulk wurtzitic GaN are 3.189 and 5.185Å in the in-plane and *c* directions, respectively. For AlN, they are 3.11 and 4.98Å. For InN, the values are 3.55 and 5.76Å. The lattice constants of $In_xGa_{1-x}N$ and $Al_xGa_{1-x}N$ alloys can be roughly estimated by assuming linear interpolations from two ending compounds. Using these data, the lattice mismatch between GaN and AlN in the basal plane is 2.5% and is much larger between GaN and InN (10%). The lattice mismatch and its induced strain have a profound effect on the growth, structures, and properties of semiconductor QDs.

Wurtzite III-nitrides are noncentrosymmetric and uniaxial crystals. They exhibit a large effect of spontaneous and strain-induced polarization. The piezoelectric coefficients, as listed in Table 8.1 are almost an order of magnitude larger than in traditional III-V compounds such as GaAs. The large amount of polarization charge may appear at the heterointerface even in the absence of strain due to the difference in the spontaneous polarization across the interface. The very strong internal electric field induced by the polarization charge and piezoelectricity is very unique to III-nitride heterostructures and has a dramatic effect on the properties of QDs.

Table 8.1 Material Parameters of Wurtzite III-Nitride Semiconductors

Parameter	*Notation*	*Unit*	*GaN*	*AlN*	*InN*
Lattice constant	a	Å	3.189	3.112	3.548
Lattice constant	c	Å	5.185	4.982	5.760
Thermal coefficient	$\Delta a/a$	K^{-1}	5.59×10^{-6}	4.2×10^{-6}	
Thermal coefficient	$\Delta c/c$	K^{-1}	3.17×10^{-6}	5.3×10^{-6}	
Bandgap, 300K	E_g	eV	3.42	6.2	1.93
Bandgap, 4K	E_g	eV	3.505	6.28	
Electron effective mass	m_e	m_0	0.22		
Hole effective mass	m_h	m_0	>0.8		
Elastic constant	C_{13}	GPa	94	127	100
Elastic constant	C_{33}	GPa	390	382	392
Static dielectric constant	ε_r	ε_0	10.4	8.5	15.3
Spontaneous polarization	P_{spon}	C/m^2	–0.029	–0.081	–0.032
Piezoelectric coefficient	e_{31}	C/m^2	–0.49	–0.60	–0.57
Piezoelectric coefficient	e_{33}	C/m^2	0.73	1.46	0.97
Binding energy exciton A	E_{xb}	meV	21		

The purpose of this chapter is to review the advances achieved in the past few years in the growth of III-nitride QDs and their unique properties. Issues relevant to the growth and the structures of QDs are described in the next section. The properties associated with quantum confinement, strain, and polarization in $GaN/Al_xGa_{1-x}N$ and $In_xGa_{1-x}N/GaN$ QDs are discussed in detail in Section 8.3.

8.2 Growth of III-Nitride QDs

The majority of III-nitride QDs are grown by self-assembled growth, using MBE or MOCVD. Due to the lack of a suitable substrate material that is lattice matched and thermally matched to GaN, III-nitride heterostructures are commonly grown on foreign substrates. Sapphire (α-Al_2O_3) is the one most extensively used. Wurtzitic GaN QDs have also been grown on other substrates such as 6H-SiC(0001) and Si(111).

The self-assembled growth of QDs takes advantage of the *Stranski-Krastanov* (SK) growth mode as schematically shown in Figure 8.1 [5, 6]. The SK mode is 2D growth of a few monolayers, called a wetting layer, followed by 3D island formation. The investigation of InAs QDs grown on GaAs shows that the 2D to 3D transition in the SK mode is in fact a first-order phase transition [7].

Experimentally, the growth mode depends not only on the materials of both epilayer and substrate but also on the growth conditions such as the substrate temperature and flow rates of various sources. Essentially, it is the result of competition between the kinetic energy of adatoms and the free energies of bulk, surfaces, and interface. For a lattice-matched system, in the limit of equilibrium growth, the layer-by-layer growth is favored if the energy of the substrate surface

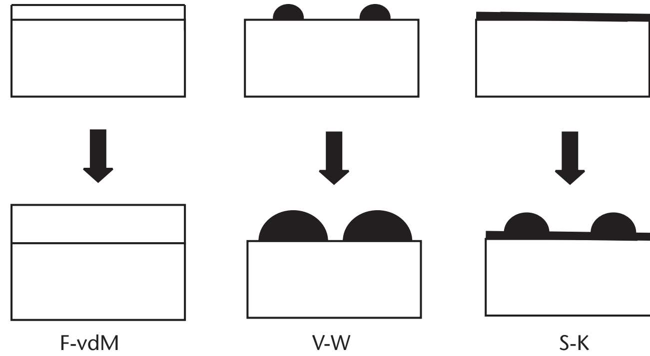


Figure 8.1 Schematic diagram of the three possible growth modes: Frank-van der Merwe, Volmer-Weber, and Stranski-Krastanow.

is higher than the sum of the epilayer surface and interface energy. Island or 3D growth can be realized by changing the surface and interface energy. In a lattice-mismatched system, the bulk elastic energy induced by the strain in the epilayer plays an important role. Because it increases with the layer thickness, strain relaxation is expected when the layer thickness is increased beyond a critical value. In fact, the SK mode is mostly observed when an epilayer is subject to a compressive strain. In this case, the stress field tends to force the adatoms to coalesce. The strain energy can be partially released by the formation of islands through elastic relaxation, without any dislocations in the islands [5]. When an epilayer is subject to a tensile strain, the growth usually continues to be 2D and the strain energy is released through plastic relaxation with the creation of dislocations. However, the transition from 2D to 3D growth is possible in the case of tensile strain if the lattice mismatch between an epilayer and substrate is large [8]. The spontaneous growth of QDs by either the 3D or SK mode is known as *self-organized* or *self-assembled growth*.

The self-assembled growth and selective growth of GaN and $In_xGa_{1-x}N$ QDs have been demonstrated using both MBE and MOCVD techniques. Other growth methods such as laser ablation and reactive RF sputtering have also been reported. In the following subsections, we describe in detail the growth and structures of GaN and $In_xGa_{1-x}N$ QDs by MBE, MOCVD, and other techniques.

8.2.1 MBE Growth of III-Nitride QDs

GaN QDs

In III-nitride MBE growth, solid Ga, Al, and In are used as the III growth sources. Both RF nitrogen plasma [9–14] and ammonia (NH_3) [15–17] have been used as nitrogen sources. The main steps in III-nitride QD growth on sapphire substrates are substrate cleaning, substrate nitridation, buffer layer growth, and active layer growth.

Substrate cleaning is a standard procedure and includes the steps of surface degreasing, chemical etching, and thermal annealing after introduction into the growth chamber. Nitridation is achieved by exposing the sapphire surface to nitrogen plasma. During this step, a reactive layer, most likely $Al_{1-x}O_xN$, may be formed.

An AlN or AlN/GaN buffer layer then follows. This buffer layer normally leads to an epilayer with a Ga face that is of better quality than that with the N face [18, 19]. The active layer can be either a single layer of GaN QDs or repeated layers of GaN QDs separated by AlN spacer layers. The latter is necessary for many practical applications requiring higher QD density. In this case, the AlN is not only used to isolate the adjacent QD layer, but it also provides a flat surface for the growth of the following QD layer.

Typically, after substrate nitridation, a thin (~10–30 nm) AlN buffer layer is grown at a temperature between 500°C and 550°C, followed by a thick (0.2- to 1.5-μm) AlN layer grown at a higher temperature of 650°C to 730°C [20]. Sometimes the growth of the thick AlN layer is preceded by a thick (~2-μm) GaN buffer layer [9]. The GaN QDs are grown on AlN by depositing 2 to 4 ML of GaN at temperatures ranging from 680°C to 730°C. Due to the 2.5% lattice mismatch between GaN and AlN, under the growth conditions used, the growth follows a SK mode. After the 2D growth of a GaN wetting layer (about 2 ML), 3D growth follows and the GaN QDs are formed [9]. The growth mode is sensitive to the substrate temperature and the GaN-deposited amount. At growth temperatures below 620°C, growth was purely 2D. Only at the elevated temperatures (680–730°C), that is, the growth transitions from 2D to 3D, did the SK mode take place [9].

AFM has been widely used to image the general morphology of QDs that are not covered by any capping layer. Figure 8.2 shows the typical AFM images of the GaN QDs grown on AlN at three different temperatures near 700°C [11]. The dot density is higher than 10^{11} cm^{-2} and decreases with growth temperature. The density can be effectively reduced through a postgrowth reorganization, called the *ripening effect* [9], after GaN growth is finished. During this period, the sample is exposed to N plasma and kept at high temperature for approximately 50s. Figure 8.3 clearly demonstrates the ripening effect. In this particular case, as compared to the sample without 50s of ripening, the dot density was reduced from

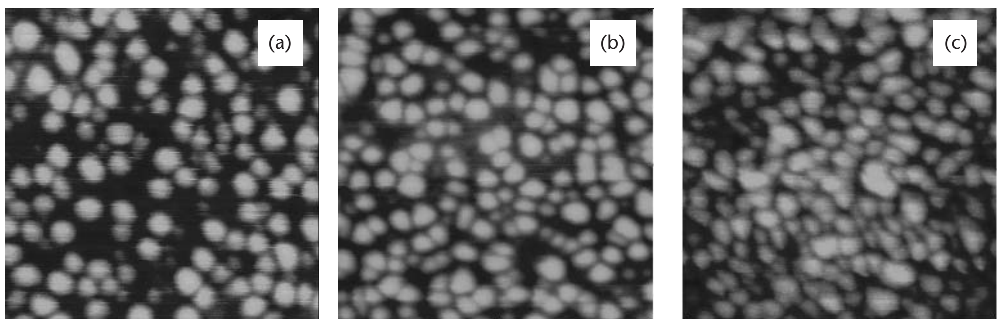


Figure 8.2 AFM images (200–3,200 nm) from GaN quantum dots grown on an AlN surface at (a) 725°C, (b) 705°C, and (c) 685°C. The growth follows the SK mode. The dot density decreases with growth temperature. (*From:* [11]. © 1998 American Institute of Physics. Reprinted with permission.)

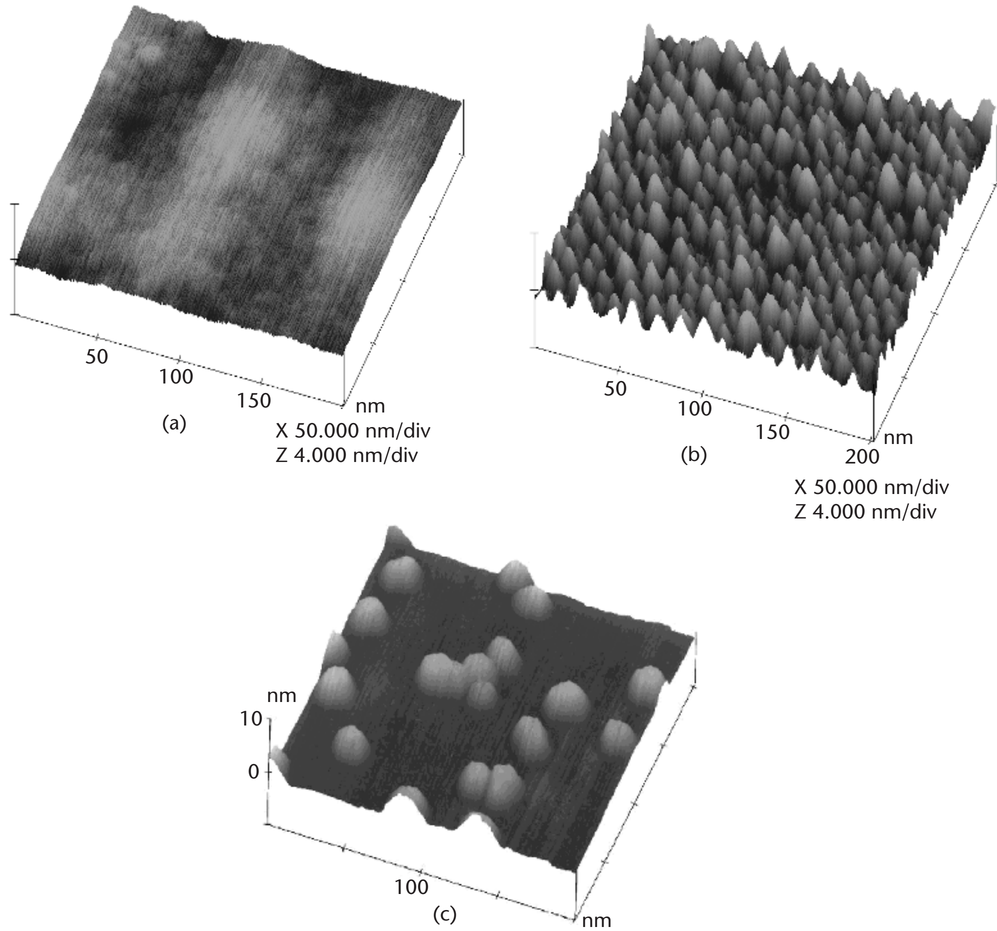


Figure 8.3 (a) AFM image of smooth AlN surface. (b) GaN quantum dots formed by depositing the equivalent of four GaN monolayers on the smooth AlN surface immediately followed by cooling under vacuum. (c) GaN quantum dots formed by depositing the equivalent of two GaN monolayers on the smooth AlN surface immediately followed by exposure to N plasma for 50 seconds. The structure reorganization or ripening effect is observed in part (c). (*From:* [9]. © 1997 American Physical Society. Reprinted with permission.)

5×10^{11} to $5 \times 10^{10} cm^{-2}$, while the average size (height/diameter) increased from 2/20 to 5/25 nm.

A detailed investigation with the help of the RHEED process shows a more significant ripening effect when the samples are exposed under vacuum rather than to nitrogen at growth temperature [11]. At lower temperatures, however, the dot size and density remain unchanged under the nitrogen plasma. The reason was suggested that the Ga diffusion on the surface might be inhibited in the presence of nitrogen [8].

It has been established that the strained islands such as InAs grown in successive layers separated by a spacer layer such as GaAs could lead to vertical correlation if the thickness of the spacer layer is appropriate [21, 22]. The driving force for such vertical self-organization is schematically shown in Figure 8.4. Islands in the first

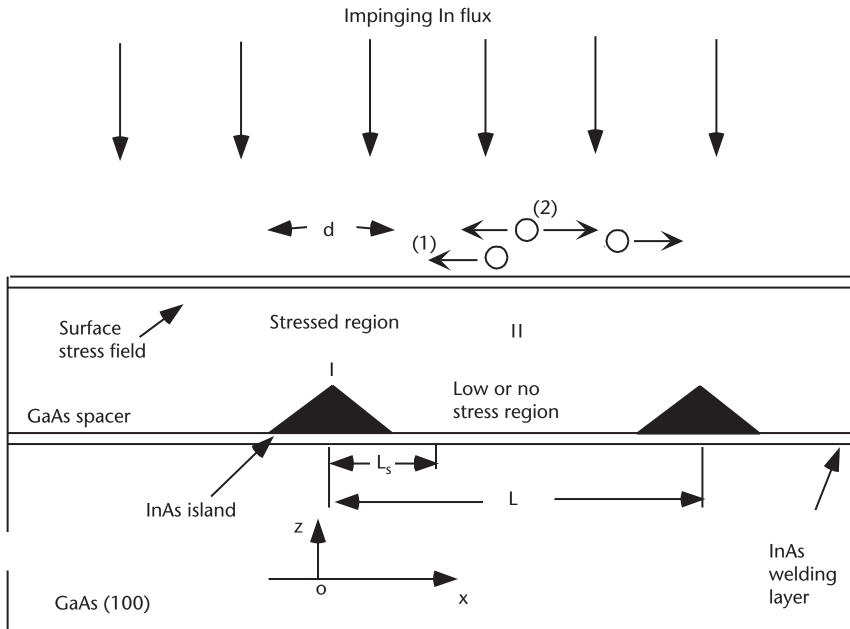


Figure 8.4 A schematic representation showing the two major processes for the In adatom migration on the stressed surface: (1) directional diffusion under a mechanochemical potential gradient that contributes toward vertical self-organization and (2) largely symmetric thermal migration in regions from the islands' contribution to initiation of new islands not vertically aligned with islands below. (*From:* [21]. © 1995 American Physical Society. Reprinted with permission.)

layer produce a tensile strain in the spacer above the islands, whereas little stress exists in the spacer away from the islands. Indium adatoms impinging on the surface would be driven by the strained field to accumulate at the top of the islands where they can achieve an energetically lower state due to the lower lattice mismatch between the new islands and the spacer. The vertical correlation of self-assembled multilayer QDs was also demonstrated in GaN/AlN systems [11, 13]. An example of an HRTEM image is shown in Figure 8.5 [11]. It reveals such a correlation for an AlN spacer of 8 nm. For a thicker AlN layer of 20 nm, no vertical correlation is observed [9].

In addition to the strain-induced vertical correlation of GaN QD arrays, a correlation between the QD growth and the threading edge dislocations propagating in AlN has been noted [12]. The conventional TEM and HRTEM images shown in Figure 8.6 demonstrate that the GaN QDs may be more likely to form adjacent to the edge dislocations. In this experiment, the dislocation density in the thick AlN layer is $1.8 \times 10^{11}\text{cm}^{-2}$, which is comparable to the density of GaN QDs ($1.1 \times 10^{11}\text{cm}^{-2}$). The strain field at the opposite side of the dislocation favors the nucleation of GaN QDs where the AlN lattice is stretched and the mismatch to GaN is smaller [12]. If the dislocation density is high, as in this case, the vertical correlation of the QDs may be disturbed by the presence of a dislocation line that is slightly inclined. Instead of following the vertical positions of the QDs in the previous layer,

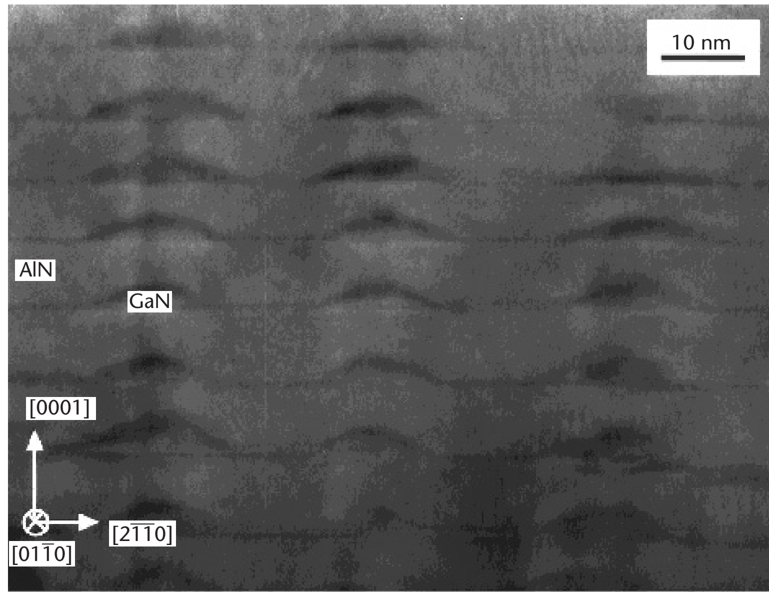


Figure 8.5 HRTEM image, taken along the [01-10] direction of a superlattice of GaN dots capped by AlN. Because of the low magnification of the printed image, the atomic columns are not seen although they are present. The vertical correlation of the GaN dots is evident. The two-dimensional GaN wetting layer is also clearly visible. Note the dislocation line running through the column of dots at the left-hand side. (*From:* [11]. © 1998 American Institute of Physics. Reprinted with permission.)

the QDs seem rather likely to follow the dislocation line. This effect may be not important if the dislocation density is much lower than the QD density.

Linear alignment of GaN QDs can be grown on AlN using vicinal SiC substrate [22]. Continuous parallel and periodic steps were formed on the SiC surfaces after being etched by HCl/H_2 gas at 1,430°C [private communication with H. Nakagawa and S. Tanaka], which can be see in AFM pictures (Figure 8.7). An AlN layer grown on this SiC surface can mimic the SiC step organization if proper growth conditions are employed [23]. The wide AlN terraces obtained were too large to control the GaN nucleus formation, resulting in homogeneous nucleation of GaN QDs on AlN. However, the growth of the vicinal surface under stress leads to a step-bunching instability that can be directly controlled by tuning the temperature and/or the reactant fluxes [24, 25]. By lowering the growth temperature and decreasing the AlN layer thickness, terraces of 20 to 50 nm in width and 2 to 5 nm in height can be obtained. Final GaN QDs are aligned along the AlN steps, demonstrating that nucleation along steps is energetically favorable as far as the growth is not kinetically limited [26], as shown in Figure 8.8(a). In Figure 8.8(b), the GaN dots are constantly present along the {$2\bar{1}\bar{1}5$} on AlN facets, due to the fact that high-index surfaces can foster 3D island nucleation compared to the nominal (0001) surface [27].

The transition from GaN 2D growth to 3D growth can be realized by changing the V/III ratio [28]. At high V/III ratio the growth mode was 3D growth, while at low

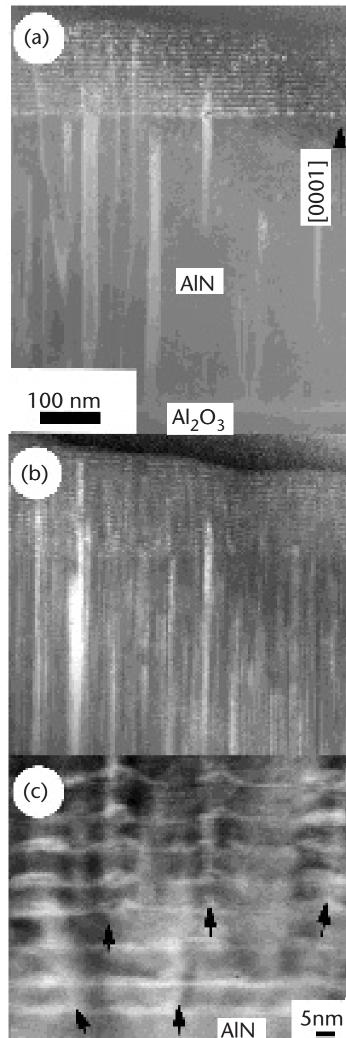


Figure 8.6 (a) Weak-beam image of a cross section of the GaN QDs in an AlN matrix (*g* = {2,-1,-1,0}). Only dislocations with a Burgers vector along *c* = [0001] are visible. Most of them are screw dislocations. (b) Weak-beam image with *g* = (0001) of the same area. Only dislocations with an in-plane component Burgers vector of the form 1/3, <2,1,-1,0> are visible. (c) Slightly off-axis <0,1,-1,0> HRTEM image. The first eight QD layers of the samples are seen. One can note that the shapes of the QDs of the first three QDs layers above the AlN thick layer are less well defined than the shapes of the QDs of the other layers. Traces of edge dislocation are outlined by dark arrows. The QDs are vertically correlated. (*From:* [13]. © 1999 American Institute of Physics. Reprinted with permission.)

V/III ratio, the growth mode was 2D growth and GaN QWs formed, as shown in Figure 8.9. The 2D-to-3D transition may result from a decrease in effective mismatch induced by a continuous Ga film on the growing GaN surface under Ga-rich conditions, which is shown schematically in Figure 8.10. The Ga film reduces surface energy, so the 3D growth mode is more favorable under this condition [29].

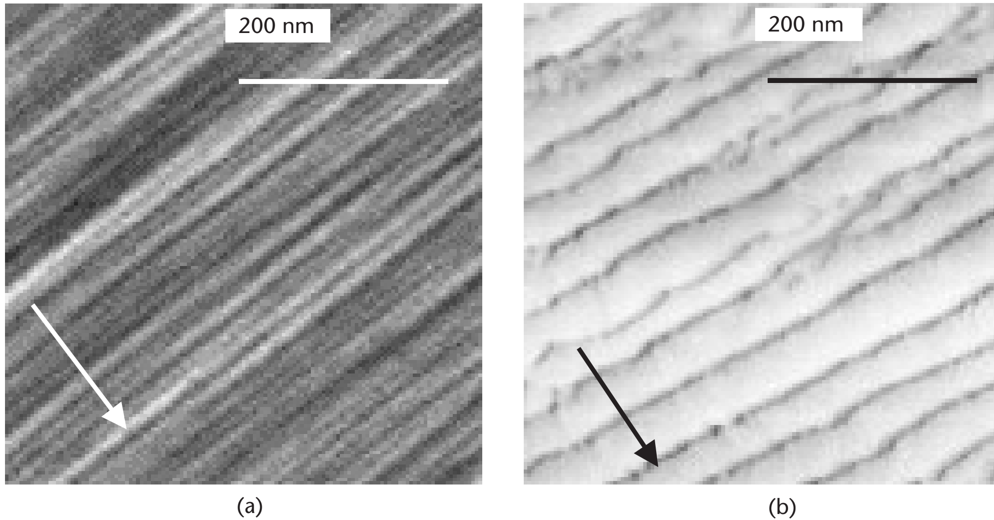


Figure 8.7 AFM images of (a) 4H-SiC(0001) vicinal surface after HCl/H_2 etching and (b) an ~50-nm AlN layer grown on a 4H-SiC(0001) vicinal surface. The arrow on the images indicates the [11-20] direction. (*From:* [23]. © 2003 American Institute of Physics. Reprinted with permission.)

However, this Ga film induces a surface strain, which leads to a reduction of the effective misfit [30]. Figure 8.11 shows an AFM image of the self-rearranged GaN QDs. The low density of dots (about 10^{10} cm^{-2}) and their inhomogeneous spatial distribution are attributed to a tendency to nucleate at the edge of AlN grains [28].

The effect of Ga as a surfactant was further studied in spontaneous rearrangement of GaN islands on (0001) AlN during annealing [31]. Under N-rich conditions, the GaN growth follows the SK mode [32], whereas Ga-rich conditions can lead to 2D growth mode [31]. GaN growth interruption under nitrogen flux can lead to island formation [33]. The interrupt was realized by shutting off N flux after a 3-ML-thick GaN film was grown on AlN under Ga-rich conditions. The GaN 2D surface did not change for up to 15 minutes. When Ga flux was also shuttered and the sample was kept under vacuum, the Ga adlayer rapidly evaporated and the surface rapidly became a 3D structure. GaN islands were formed. The limitation of adatom diffusion is probably not the reason for the surfactant of Ga [33]. It is instead likely that Ga greatly increases the nucleation barrier for GaN island formation by decreasing the surface energy [29].

During QD growth, strain evolution and SK transition of the deposited layer can be monitored with the RHEED process. Postformation characterization is obtained by AFM, which gives the aspect ratio and the size distribution of the QDs and their density. An accurate characterization of QD morphology in depth is performed with HRTEM. However, such local observations do not provide any statistical information concerning the long-range ordering. Using

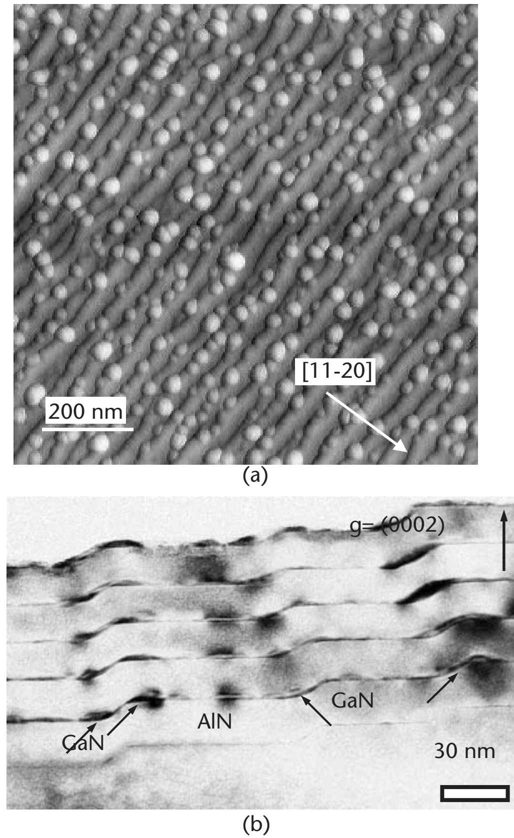


Figure 8.8 (a) AFM image showing the preferential alignment of GaN QDs along the AlN bunched steps. (b) Two-beam [*g* = (0002)] TEM image. The GaN QDs are mainly detected at the step edges and are vertically correlated from layer to layer. (*From:* [23]. © 2003 American Institute of Physics. Reprinted with permission.)

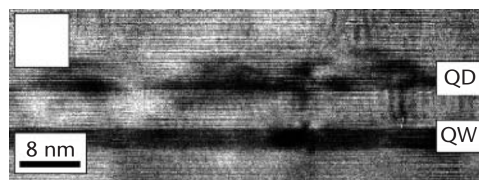


Figure 8.9 QWs and QDs grown at the same temperature (730°C), but at a different Ga/N ratio value (2–3) for the well and for the dots. (*From:* [28]. © 2002 John Wiley & Sons, Inc. Reprinted with permission.)

X-ray grazing incidence techniques, Chamard et al. [34] studied GaN QD vertical stacking in multilayers grown by MBE. The QD multiplayers they studied consisted of 10 bilayers, and each bilayer consisted of a 4-nm AlN spacer and a QD layer.

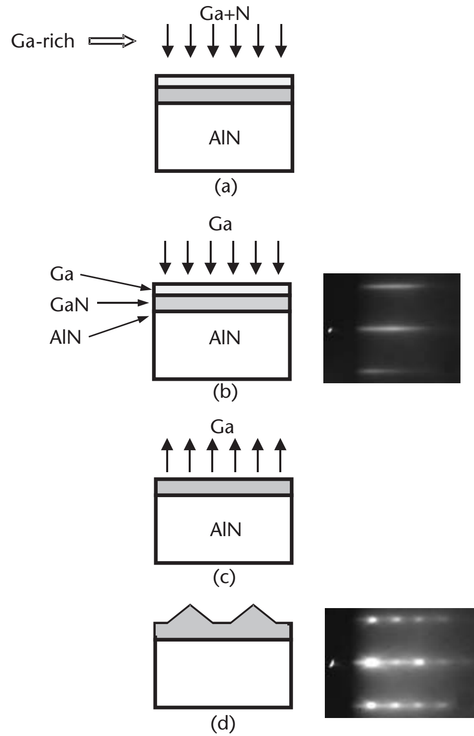


Figure 8.10 Schematic of an experimental procedure for growing self-rearranged GaN QDs. (a) GaN growth in Ga-rich conditions followed by (b) exposure to Ga alone and (c) Ga evaporation under a vacuum, leading to (d) self-rearrangement of GaN into 3D islands. (*From:* [28]. © 2002 John Wiley & Sons, Inc. Reprinted with permission.)

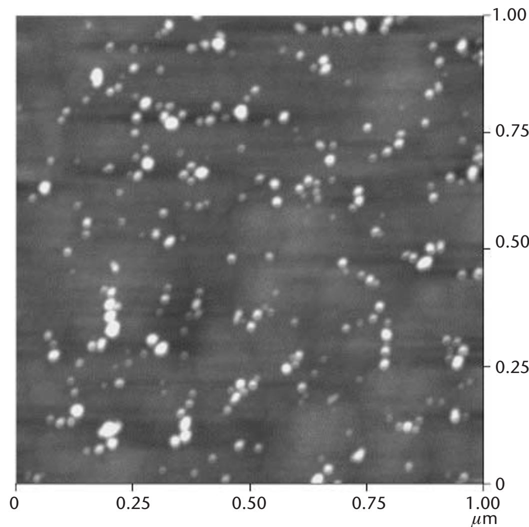


Figure 8.11 AFM image of self-rearranged GaN quantum dots on AlN. The growth/rearrangement temperature was 740°C. The GaN coverage was 3 ML. (*From:* [28]. © 2002 John Wiley & Sons, Inc. Reprinted with permission.)

The growth of GaN QDs on Si(111) [15, 35] and SiC(001) [36] substrates by MBE has also been reported. The purpose of growing GaN QDs on Si substrates is mainly for the integration of LEDs with Si technology. The growth processes in the former case are essentially the same as those for sapphire substrates. By controlling the size of the GaN QDs in the AlN matrix, intense room temperature PL with different colors from blue to orange as well as white was demonstrated [15]. When GaN was grown on a 3C-SiC(001) surface, QDs with a cubic rather than a hexagonal structure can be obtained. The zinc-blende GaN islands were formed on AlN buffer by RF MBE with an average height of 1.6 nm and a diameter of 13 nm. The island density is $1.3 \times 10^{11} cm^{-2}$ [36].

In addition to the SK mode, 3D growth of GaN QDs on $Al_xGa_{1-x}N$ was made possible by using the so-called “antisurfactant” Si [16]. In this experiment, a smooth $Al_xGa_{1-x}N$ layer was prepared on 6H-SiC(0001) by MOVPE and used as the substrate for MBE regrowth. The GaN QDs were grown by MBE in which NH_3 was used as N source and CH_3SiH_3 was used as a Si source. The $Al_xGa_{1-x}N$ surface was exposed to Si flux before the GaN growth, and the NH_3 flow was stopped for this step. Subsequent GaN growth was carried out with and without introducing Si on two different samples. In the sample without Si flux during the GaN growth, the growth was 2D and streaky RHEED patterns were observed. With Si flow, a change of the GaN growth mode from 2D to 3D was observed and the RHEED patterns turned out to be spotty. Formation of GaN QDs was confirmed by AFM. The dot density could be changed by the variation of the Si flux and the growth temperature. The dot density decreased by a factor of 10^3 and the dot sizes increased from 4/50 to 10/200 nm by raising the temperature from 660°C to 740°C [15].

InGaN QDs

UV to red electroluminescent devices based on III-nitrides grown by MOCVD have quantum efficiencies as high as 18% in spite of the huge density of dislocation [37]. It is now currently admitted that the high quantum efficiency comes from the recombination of strongly localized carriers in high In content zones due to the phase demixing of the InGaN alloy [38]. However, output of blue LEDs grown by MBE is still low [39]. This can possibly be explained by the different growth conditions of MBE and MOCVD. Under normal MBE growth conditions, InGaN QWs are homogeneous, which has been indicated by TEM [29], but In phase separation can occur during typical MOCVD growth conditions. In-rich clusters may act as QDs, which are responsible for the high brightness of LEDs based on MOCVD-grown materials [40].

$In_xGa_{1-x}N$ QDs can be obtained by MBE on GaN with a RF plasma source [41]. At 580°C, layer-by-layer growth of $In_{0.35}Ga_{0.65}N$ on GaN was observed during the first 1.7 ML of growth. Beyond 1.7 ML, the growth mode was changed from 2D to 3D and the $In_xGa_{1-x}N$ islands formed (Figure 8.12) [41]. An AFM image (Figure 8.13) of the $In_xGa_{1-x}N$ surface with 5-ML deposition shows an InGaN QD density of $\sim 10^{11} cm^{-2}$. The average diameter and height of these QDs are 27 and 2.9

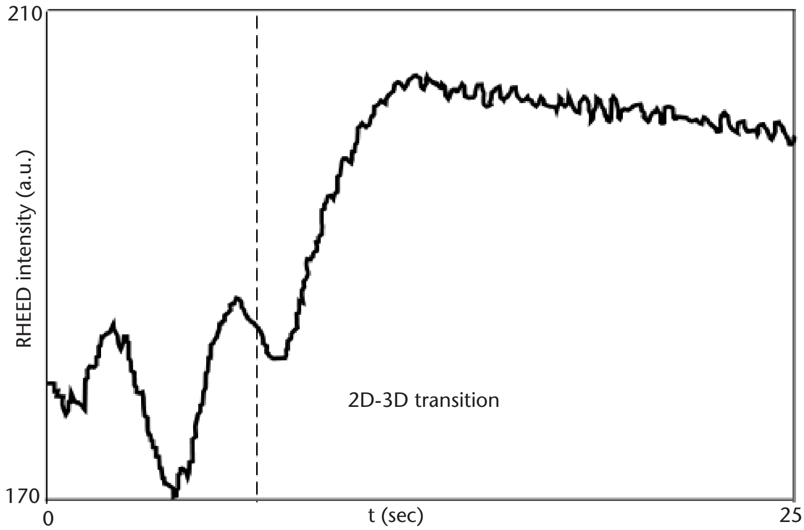


Figure 8.12 Variation of the intensity of the RHEED specular spot superimposed on a Bragg spot during the first stage of $In_{0.35}Ga_{0.65}N$ growth on GaN. Two oscillations are observed, followed by a rapid increase of the intensity due to the intensification of the Bragg spot. (*From:* [41]. © 2000 American Institute of Physics. Reprinted with permission.)

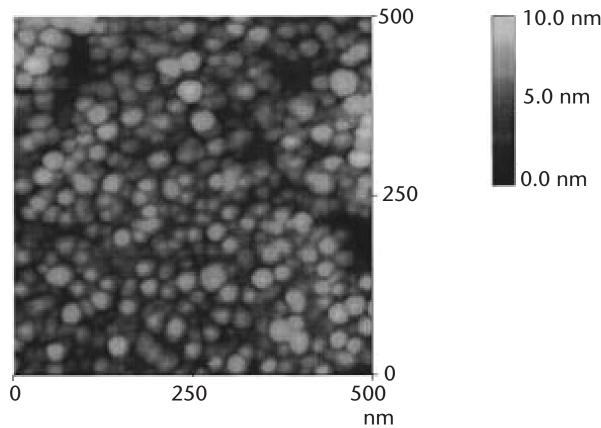


Figure 8.13 Topographical AFM image of a surface layer of $In_{0.35}Ga_{0.65}N$ dots grown on GaN. (*From:* [41]. © 2000 American Institute of Physics. Reprinted with permission.)

nm, respectively. Investigation shows that the transition of 2D-to-3D growth can be realized for In content from 18% to 100% [41]. For In content below 18%, InGaN growth mode remains 2D.

$In_xGa_{1-x}N$ QDs grown by the SK mode can also be realized by MBE using NH_3 as the nitrogen source [42]. In this case, $In_xGa_{1-x}N$ was grown on GaN buffer of a few micrometers at 530°C to 570°C, with a growth rate of 0.1 to 0.2 μm/hr. The composition of the In was kept at 0.15. which is larger than the critical value of the

2D-to-3D transition (0.12) determined in the same experiments [43]. The 2D-to-3D transition was observed after the deposition of 4 to 5 ML (~11Å) of $In_{0.15}Ga_{0.85}N$. The average InGaN QD size was about 35 nm in diameter and 4 nm in height. The island density was approximately 5×10^{10} cm^{-3}.

A typical surface morphology for $In_xGa_{1-x}N$ QDs on GaN is shown in Figure 8.14. Compared with GaN QDs, the densities and sizes of $In_xGa_{1-x}N$ QDs are more difficult to control during growth. Except for AFM and RHEED investigations, no detailed lattice structures have been imaged by HRTEM for $In_xGa_{1-x}N$ QDs.

The self-organized growth of GaN QDs by MOCVD was first reported by Dmitriev et al. [44, 45]. They grew GaN QDs directly on 6H-SiC substrates. In this case, the lattice mismatch between the GaN and SiC was large enough to lead to island growth. The migration and evaporation of Ga atoms also play an important role in the formation of GaN QDs. The most essential parameters needed for GaN QDs growth are growth temperature, GaN coverage, and V/III ratio, which determine the migration and/or evaporation of Ga atoms on the GaN surface. The growth temperature and growth rate must be calibrated, especially to grow sufficiently small high-density QDs.

Suppression of Ga atom migration is required, which is usually realized by lowering the growth temperature for QD growth. In the experiment by Miyamura et al. [46], GaN QDs were grown by MOCVD with V/III ratios below 30, as shown in Figure 8.15. GaN QDs grown at 960°C have smaller diameters and higher densities than those grown at 975°C, while no QDs could grow at 990°C due to serious

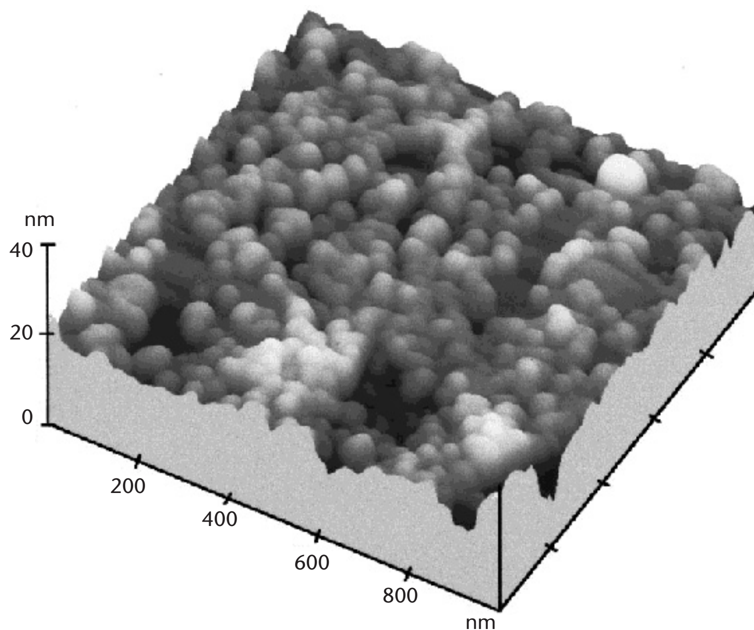


Figure 8.14 AFM image (1 μm × 1 μm) of unburied self-assembled $In_xGa_{1-x}N$ islands grown on GaN by MBE. (*From:* [42]. © 1999 American Institute of Physics. Reprinted with permission.)

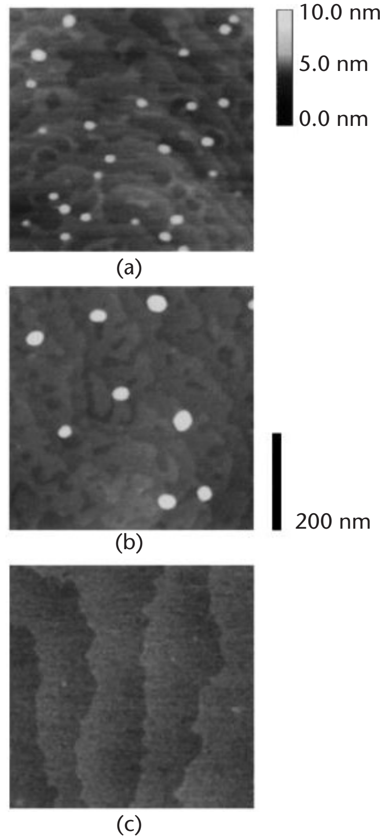


Figure 8.15 AFM images of GaN QDs grown at (a) 960°C, (b) 975°C, and (c) 990°C, respectively. The image scale is 500×500 nm^2. (*From:* [46]. © 2002 American Institute of Physics. Reprinted with permission.)

migration and evaporation of Ga atoms. Researchers indicated that the formation of GaN QDs is very sensitive to the growth temperature. Dependence of GaN QDs growth on GaN coverage is shown in Figure 8.16. With an increase in the GaN coverage from 5.8 to 9.2 ML, the QD density changed from 1×10^8 cm^{-2} to 5×10^{10} cm^{-2}. The density tends to saturate when GaN coverage is greater than 8 ML. There is a critical thickness of the GaN layer around 4 ML. By depositing over the critical thickness, QDs are immediately formed on the two-dimensional GaN layer. Such an abrupt transition from the 2D to the 3D growth mode was also reported for InAs QDs [47].

Aside from self-assembly growth, selective growth is another important method used to obtain GaN QDs. The advantage of selective growth is to get uniform distribution of the position, shape, and size of QDs. The growth of a uniform array of GaN QDs embedded on AlGaN matrix formed on hexagonal pyramids of GaN on SiO_2/GaN/sapphire substrate by selective MOCVD was reported by K. Tachibana et al. [48]. A uniform array of GaN pyramids was realized at 945°C, followed by

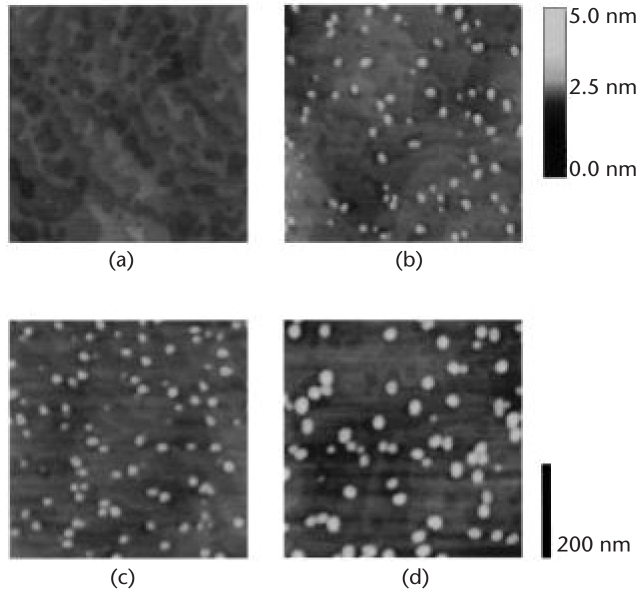


Figure 8.16 AFM images of GaN QDs. The GaN coverages are (a) 3.4, (b) 5.8, (c) 6.9, and (d) 9.2 ML, respectively. The image scale is 500 × 500 nm^2. (*From:* [46]. © 2002 American Institute of Physics. Reprinted with permission.)

selective growth of 20 periods of GaN/$Al_{0.2}Ga_{0.8}N$ MQWs at 1,040°C. Figure 8.17 shows an SEM picture of the uniform GaN pyramids with GaN QDs at the tops. The sizes of the QDs were estimated to be less than 10 nm, based on the top radius of GaN pyramids.

Tanaka and coworkers [49–52] were the first to apply antisurfactant during GaN QD growth on $Al_xGa_{1-x}N$. The self-assembly of GaN QDs is realized in this small lattice-mismatched system by exposing the $Al_xGa_{1-x}N$ surface to Si during the growth. The Si is from tetraethyl-silane [$Si(C_2H_5)_4$:TESi:0.041 μmol] (TESi) and carried by H_2. The antisurfactant involved growth of GaN QDs typically includes depositing a thin (~1.5-nm) AlN buffer layer, a thick (~0.6-μm) $Al_xGa_{1-x}N$ cladding layer, and GaN deposition on top of this $Al_xGa_{1-x}N$ with a short supply (5 seconds) of TMG and NH_3 during which TESi may or may not be used [49]. As shown in Figure 8.18, if TESi is not supplied, a step flow growth of GaN with a smooth surface was observed. Only when the $Al_xGa_{1-x}N$ surfaces were exposed to TESi were the GaN QDs effectively grown. The step-flow growth observed without TESi flux [Figure 8.18(c)] was explained by a fairly small lattice mismatch between GaN and $Al_xGa_{1-x}N$ (0.37 % for x = 0.15). Under the exposure of a small Si dose, large GaN islands were formed. These islands transformed into isolated small dots under a higher Si flux [Figure 8.18 (a, b)].

The dot density in this case could be controlled from ~10^7 to 10^{11} cm^{-2} by changing the TESi doping rate, growth temperature, growth time, and alloy composition. It is very sensitive to the growth temperature, varying by a factor of 10^3 between 1,060°C and 1,100°C. The dot size can also be controlled by a proper choice of

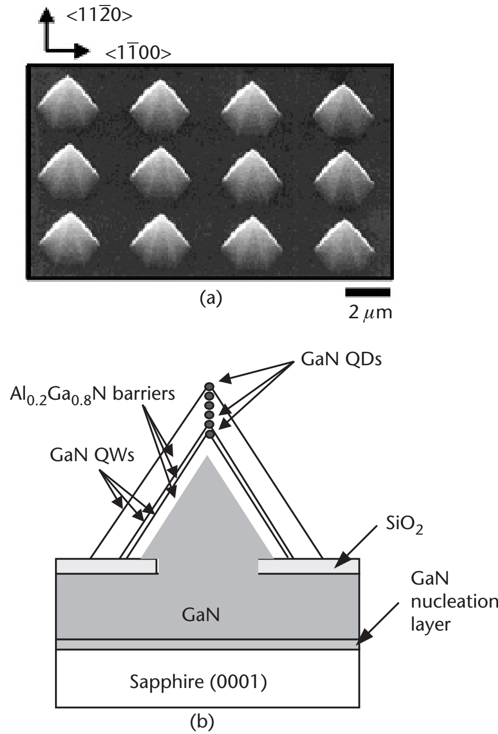


Figure 8.17 (a) SEM bird's-eye view of a uniform array of hexagonal pyramids including GaN QDs in AlGaN matrix at the tops. (b) Schematic of GaN QDs in AlGaN matrix formed on hexagonal pyramids of GaN on a SiO_2/GaN/sapphire substrate. (*From:* [46]. © 2001 John Wiley & Sons, Inc. Reprinted with permission.)

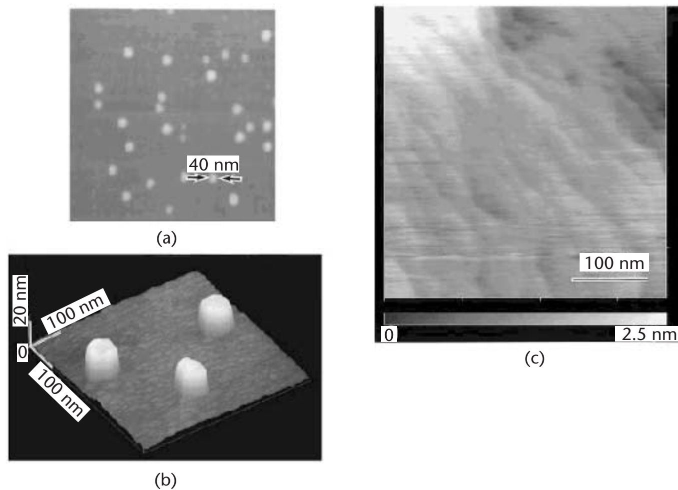


Figure 8.18 AFM images of GaN quantum dots assembled on an $Al_xGa_{1-x}N$ surface using TESi as an antisurfactant: (a) plane view and (b) bird view. (c) An AFM image of GaN grown on the $Al_xGa_{1-x}N$ surface without TESi doping, showing a step flow growth. (*From:* [49]. © 1996 American Institute of Physics. Reprinted with permission.)

growth parameters. In the study done by Tanaka et al. [44, 45, 49], GaN QDs were grown on an $Al_xGa_{1-x}N$ surface (x = 0.2) [52]. The hexagonal-shaped GaN dots have an average height of ~6 nm and diameter of ~40 nm. The dot thickness-to-diameter ratio could be changed from 1/6 to 1/2 by varying the growth temperature and Si dose [44, 45, 49]. The dot density is ~3×10^9 cm^{-2}, more than 1 order of magnitude lower than the GaN QDs grown by MBE. For a fixed growth temperature (T_s = 1,080°C) the densities of ~5×10^9 and ~5×10^8 cm^{-2} were obtained with a TESi doping rate of 44 and 176 nmol/min, respectively. By increasing the GaN growth time from 5 to 50 seconds, the dot size was changed from ~6/40 to ~100/120 nm [49].

MOCVD Growth of InGaN QDs

InGaN QDs can also be obtained by self-assembly or selective growth. The fabrication of self-assembling InGaN QDs on AlGaN surfaces using Si as antisurfactant was reported by Hirayama et al. [53]. The microscopic mechanism of antisurfactant in the growth of InGaN QDs is not well understood. Incorporation of Si in the growth process is assumed to change the surface energy of the AlGaN layer so that the growth mode is modified. The AFM photograph (Figure 8.19) shows that the average size and height of the QDs grown by MOCVD are approximately 10 and 5 nm [53], respectively. The density of the QDs can be controlled in the range of 10^{11} to 10^{10} cm^{-2} by changing the dose of Si antisurfactant. The same group also succeeded in fabricating AlGaN QDs by controlling the surface migration of precursors in addition to the use of Si antisurfactant. The low growth temperature of AlGaN QDs of 900°C was used to reduce the precursor migration on AlGaN surface.

Indium as an antisurfactant can also be applied in the self-assembly of InGaN [54]. One advantage of using In as the antisurfactant is that In itself is a

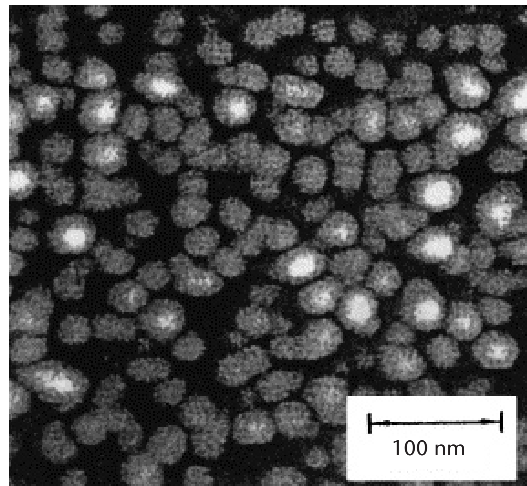


Figure 8.19 AFM image of an $In_{0.38}Ga_{0.62}N$ QD on $Al_{0.12}Ga_{0.88}N$ surface without capping layer. (*From:* [53]. © 1999 Elsevier B.V. Reprinted with permission.)

constituent component of the InGaN film. In the growth procedure reported by Zhang et al. [54], InGaN QDs were formed in InGaN SQWs by reacting with TMIn and NH_3 for a short duration, which was called TMIn treatment. The HRTEM image of Figure 8.20 shows a typical microstructure for the InGaN QDs formed in the SQW, which originated from In-rich regions in the low-In composition matrix with an average width of 4 nm and height of 1.5 nm. The growth mode was two-dimensional throughout the growth of the SQW structure without TMIn treatment. The existence of InGaN QDs was also verified by PL spectra.

$In_xGa_{1-x}N$ QDs can also be grown on GaN by MOCVD by means of the strain-induced SK growth mode without using an antisurfactant [55]. The InGaN QDs were grown on (0001) GaN buffer at 700°C. The growth time and growth temperature were proven to be very important to the size and distribution of InGaN QDs. As shown in Figure 8.21, with an increase in the growth temperature from 675°C to 740°C, the density of the InGaN QDs decreases monotonically from 2.1×10^{10} to 4.5×10^9 cm^{-2}. And the average diameter of the QDs is reduced from 16.8 to 8.4 nm by changing the growth temperature from 740°C to 700°C. At 700°C, InGaN QDs with a deposited amount of 6.4 ML have two kinds of InGaN QDs: bigger QDs, of diameter 15.5 nm and height 5.4 nm, and smaller QDs, of diameter 9.3 nm and height 4.2 nm. In another sample with a deposited amount of 19.1 ML, the size of InGaN QDs is 8.4 in diameter and 2.1 nm in height, respectively (Figure 8.22). The dependence of the InGaN QDs properties on the deposited amount and growth temperature is similar to that of InAs/GaAs system [56, 57], but the difference is that the

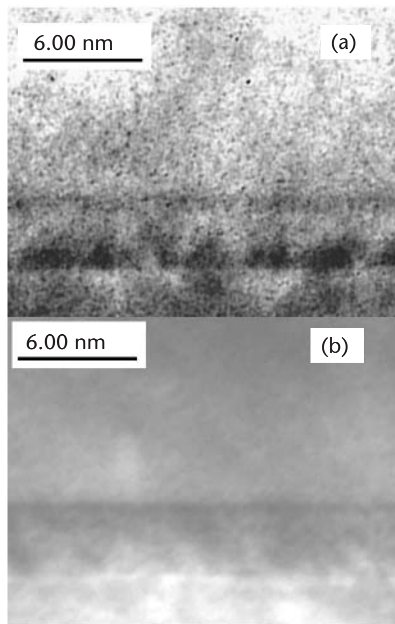


Figure 8.20 HRTEM images for InGaN SQW cross sections (a) with and (b) without TMIn treatment. Arrows point to the low-barrier/well interfaces. (*From:* [54]. © 2002 American Institute of Physics. Reprinted with permission.)

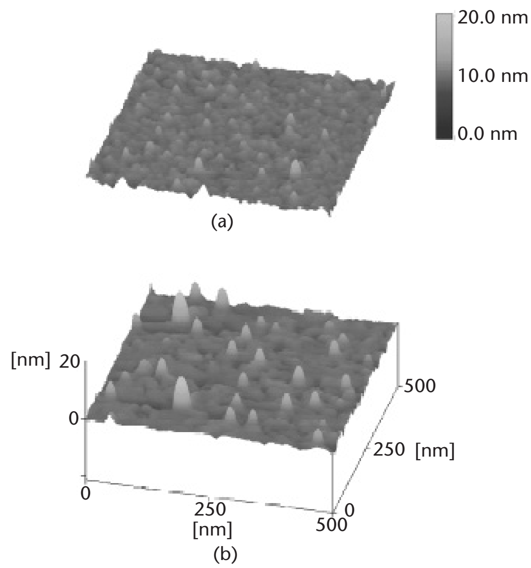


Figure 8.21 AFM images of InGaN self-assembled QDs on a GaN buffer layer at growth temperatures of (a) 675°C and (b) 740°C. (*From:* [55]. © 1999 American Institute of Physics. Reprinted with permission.)

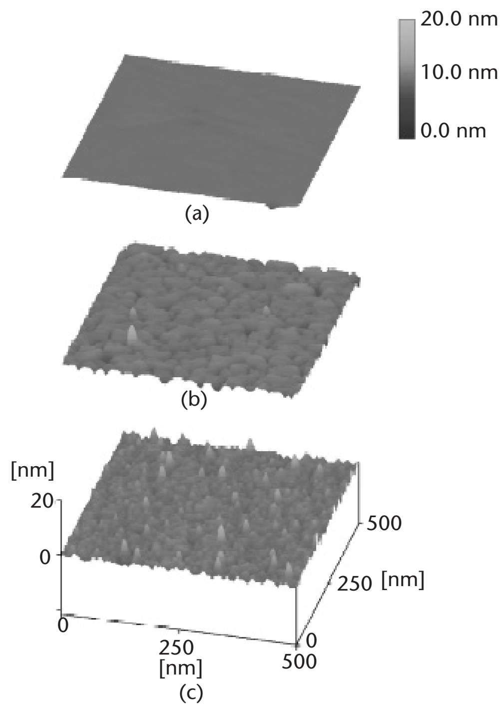


Figure 8.22 AFM images of (a) a GaN buffer layer, and self-assembled QDs formed by the deposition of (b) 6.4 and (c) 19.1 ML. (*From:* [55]. © 1999 American Institute of Physics. Reprinted with permission.)

InGaN QDs were formed even at the deposited amount of 19.1 ML, which suggests that the origin of InGaN QDs formation might be phase separation rather than strain.

Without the use of surfactant or antisurfactants, InGaN QDs can be obtained on GaN (0001) by using the MOCVD interrupted growth method [58]. The interruption occurred after 4.5 ML of InGaN deposition at 730°C on undoped GaN layer for 12 seconds. After growth stopped another 4.5 ML of InGaN were deposited. AFM photographs (Figure 8.23) show that by interrupting growth the surface morphology of an InGaN sample can be significantly changed. The size of the QDs

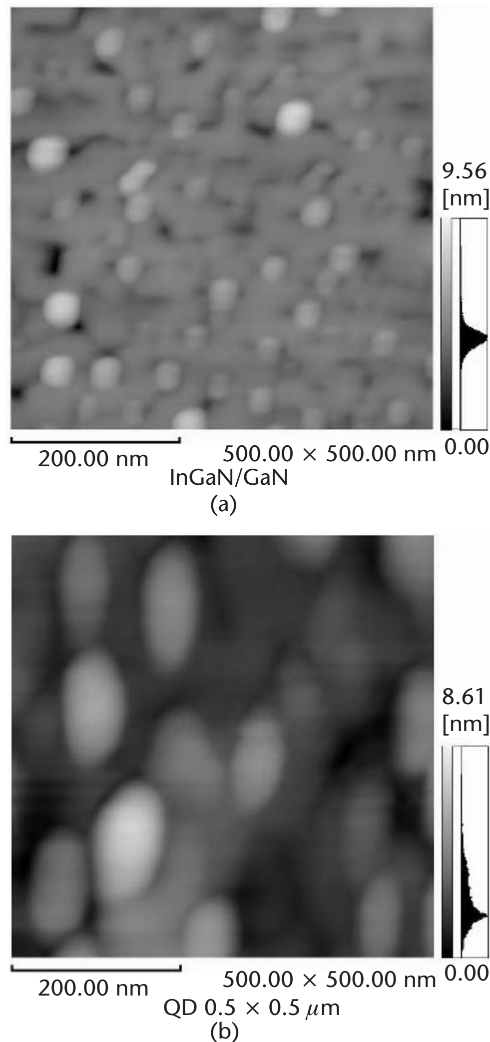


Figure 8.23 AFM image (500 × 500 nm^2) of (a) a sample having smaller InGaN QDs that was grown without using the MOCVD interrupted growth mode and (b) a sample having larger InGaN QDs that was grown without using the MOCVD interrupted growth mode. (*From:* [58]. © 2003 Elsevier B.V. Reprinted with permission.)

from a sample grown via the interrupted mode was much smaller than that observed from a sample grown without interruption. The RT-PL spectra of these two samples are shown in Figure 8.24. The PL peak position of the InGaN sample grown with interruption has a 67-meV blue shift compared to the sample grown without interruption, which is believed to be due to the QD size effect.

A new method using MOCVD to form nanoscale InGaN QDs taking advantage of neither surfactant nor the SK growth mode was reported by Chen et al. [59]. InGaN QDs were formed by increasing the energy barrier for adatoms on the growth surface, which are realized by surface passivation and by decreasing the growth temperature. With this method, InGaN QDs were grown by means of the following steps. First, the surface of the HT-GaN films was passivated and served as a substrate. The passivation process mainly includes exposing the sample in the atmosphere for 24 hours so that it becomes passivated by the oxygen in the atmosphere. Then, the substrate is heated to 550°C for growth to a certain thickness of GaN QDs. Finally, the growth temperature was increased to 850°C and the N_2 carrier gas was then supplied in the growth chamber, followed by a short supply of TMGa and EDMLn and NH_3 gases for growth of the InGaN QDs. The surface morphology of the InGaN QDs obtained in step 2 is shown in Figure 8.25. The round-shaped GaN dots with an average width of ~80 nm and height of ~5 nm are uniformly distributed on the substrate with a density of ~6×10^{10} cm^{-2}. The InGaN QDs are larger than the GaN dots below it. Similar phenomena were observed in InAlGaAs system [60, 61].

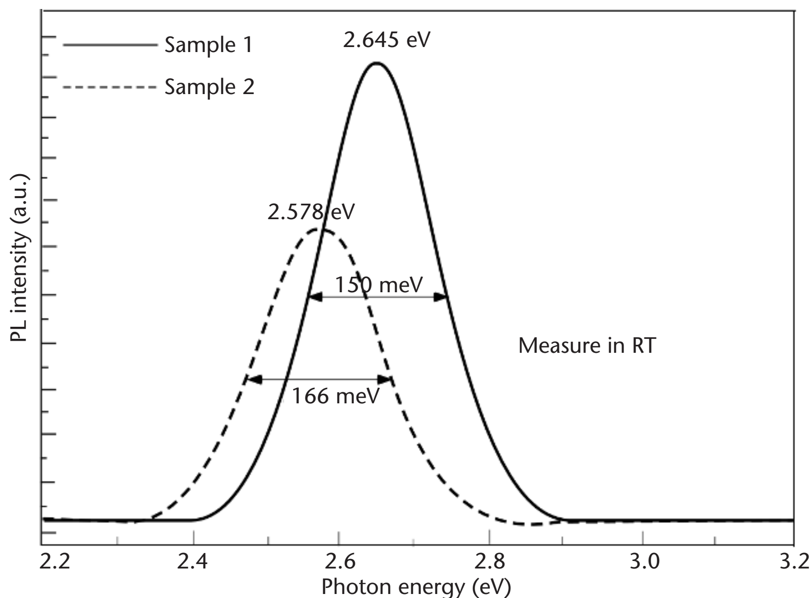


Figure 8.24 Gaussian curve-fitting PL and original PL spectra (top, right side of figure) of the InGaN QDs samples measured at room temperature, where sample 1 was prepared by the interrupted growth mode, sample 2 was prepared without using the interrupted growth mode. (*From:* [58]. © 2003 Elsevier B.V. Reprinted with permission.)

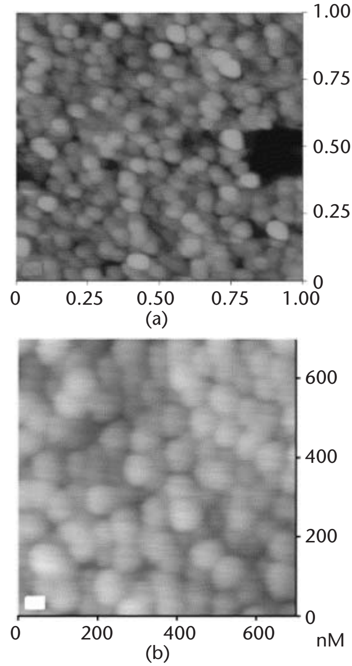


Figure 8.25 (a, b) Surface morphology of InGaN dots on GaN dots/HT-GaN/sapphire substrate (sample D). The uniform dots with an average width of ~80 nm and height of ~5 nm are distributed on the substrate surface. (*From:* [59]. © 2002 Elsevier B.V. Reprinted with permission.)

PL spectra of the InGaN QDs sample and a homogenous InGaN thin film are shown in Figure 8.26. The InGaN film has the same growth condition as the InGaN QDs. The PL emission intensity of InGaN QDs is about 10 times that of the InGaN films with an obvious blue shift.

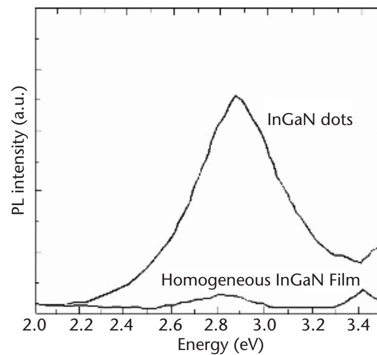


Figure 8.26 A comparison of PL spectrum between InGaN dots and homogeneous InGaN film. The PL emission intensity of InGaN dots is 1 order of magnitude stronger than that of the InGaN film with the same growth time. The 3.40-eV peak in the spectra of the InGaN film comes from the energy band emission of underlying GaN. (*From:* [59]. © 2002 Elsevier B.V. Reprinted with permission.)

Similar to selective growth of GaN QDs [48], InGaN QDs can be formed on top of GaN pyramids by MOCVD [62, 63], as illustrated schematically in Figure 8.27(a). Figure 8.27(b) shows a SEM bird's-eye view of the InGaN QDs structures obtained by selective growth, and Figure 8.27(c) shows a cross-sectional image. The lateral size of InGaN QDs is then considered to be comparable to the radius of curvature. Figure 8.28(a) shows a micro-PL intensity image having a wavelength around 430 nm. In collection, wavelengths shorter than 390 nm were cut off by a barrier filter, and a 430-nm bandpass filter of bandwidth 5 nm was used, so the detected light was from the InGaN QD only. The PL emission in Figure 8.28(a) is only from the top of the pyramids. In the cross-sectional profile of PL intensity in Figure 8.28(b), the FWFM of 330 nm is comparable to the spatial resolution, which indicates that the emission originates from InGaN QD embedded in the InGaN matrix.

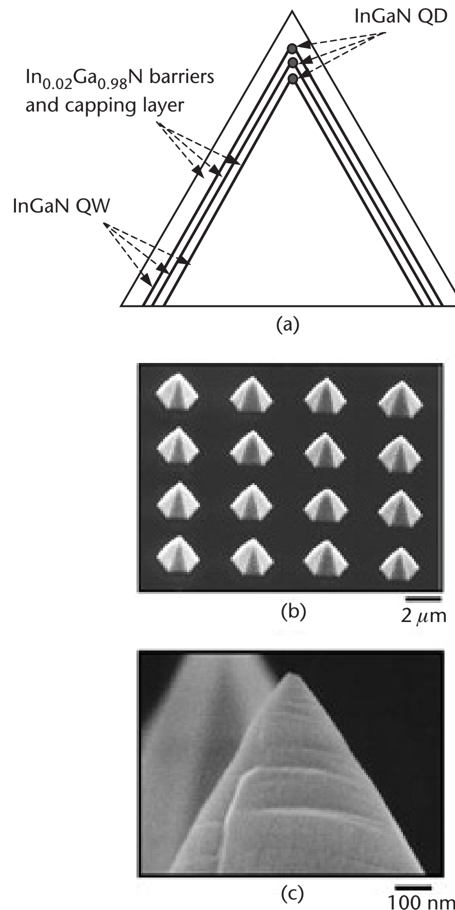


Figure 8.27 (a) A schematic of InGaN QDs formed on the tops of hexagonal pyramids of GaN. SEM pictures of the sample: (b) bird's-eye view and (c) cross section. (*From:* [63]. © 2000 American Institute of Physics. Reprinted with permission.)

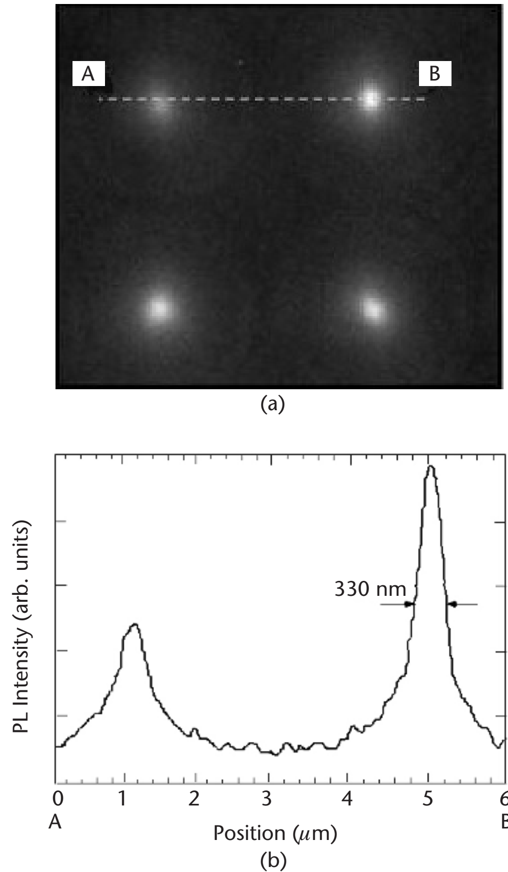


Figure 8.28 (a) Micro-PL intensity image at a wavelength of 430 nm and (b) cross-sectional profile of PL intensity along the line AB in part (a). (*From:* [63]. © 2000 American Institute of Physics. Reprinted with permission.)

8.2.2 Other Techniques

A less-explored method of growth of self-assembled nanoscale structure is based on the VLS mechanism, in which metal or semiconductor liquid droplets condensing from the vapor phase catalyze whisker or fiber growth under high-temperature CVD conditions [64]. The VLS method offers distinct advantages for the growth of nitride QDs because Ga can be deposited as droplets with sizes that can be fine tuned by simple annealing at relatively low temperatures. Accordingly, the VLS mechanism has been exploited in the growth of crystalline GaN nanowires [65] and also self-assembled GaN QDs on an AlGaN/SiC substrate by gas-source MBE [66].

Self-assembled nanometer-scale QDs were fabricated on 6H-SiC (0001) substrate by MBE via the formation of Ga liquid droplets and their subsequent nitridation with a supersonic gas source seeded with NH_3 molecules [67]. Figure 8.29 shows the frame-captured *low-energy electron microscope* (LEEM) images of the formation of Ga droplets and the subsequent GaN QDs formation by exposure to NH_3. Then the Ga shutter was closed and the surface exposed to NH_3 flux, while the

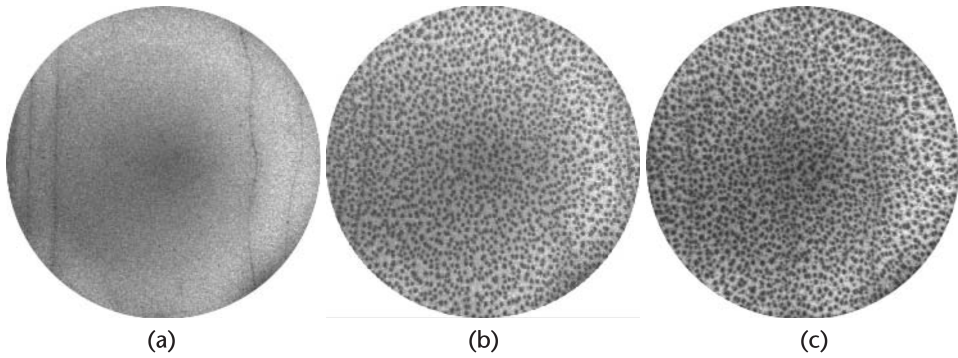


Figure 8.29 Frame-captured LEEM images showing the formation of Ga droplets and the exposure to SSJ NH_3 on 6H-SiC(0001) substrate: (a) initial ($\sqrt{3} \times \sqrt{3}$) surface of 6H-SiC(0001) substrate; (b) after 28-minute deposition of Ga droplets at 535°C; (c) after 100 minutes of NH_3 exposure at 600°C for nitridation followed by brief annealing at 700°C to eliminate excess Ga liquid. Field of view is 9.0 μm, and electron energy is 6.0 eV. (*From:* [67]. © 2002 American Institute of Physics. Reprinted with permission.)

temperature was slowly increased from 530°C to 600°C. The nitridation process lasted for 100 minutes at 600°C, then temperature was raised to 700°C to remove residual liquid Ga droplets on the surface [68]. The LEEM image of the as-grown GaN QDs is shown in Figure 8.29(c). An AFM image of the GaN QDs is shown in Figure 8.30, where the density of the QDs was measured at 4.1×10^9 cm^{-2}, and the base widths of the dots were from 5 to 100 nm.

HR-XTEM images of the dots are presented in Figure 8.31. The smaller QDs (~0.5 nm in width) are completely free of defects. The larger QDs (~25 nm in width), however, display some stacking-fault defects.

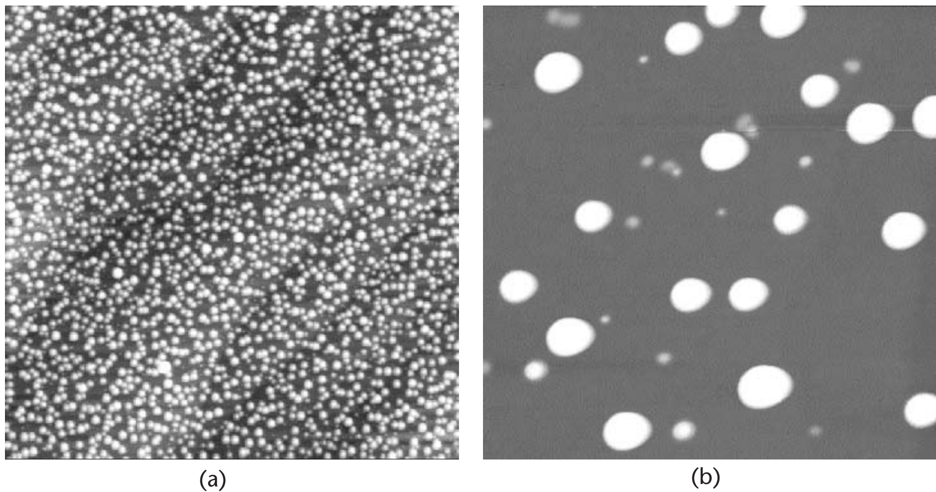


Figure 8.30 AFM images of the as-grown GaN dots. Scan areas are (a) 10×10 μm^2 and (b) 1×1 μm^2. (*From:* [67]. © 2002 American Institute of Physics. Reprinted with permission.)

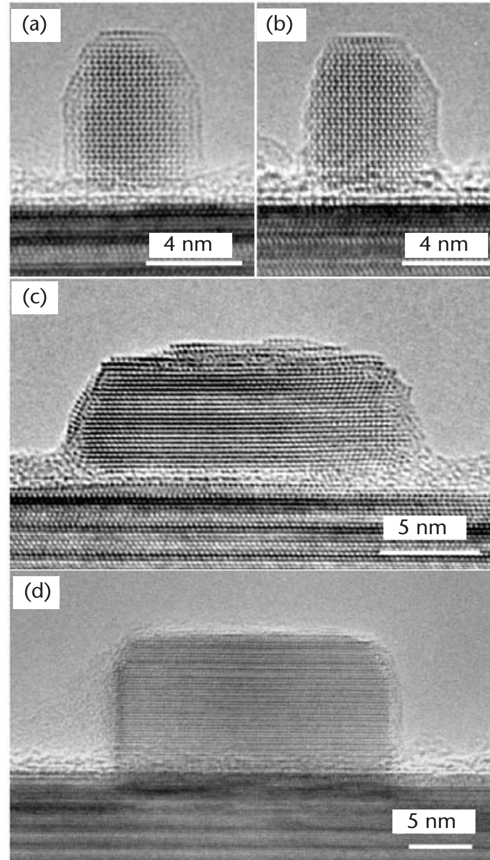


Figure 8.31 HR-XTEM images of (a, b) perfectly coherent GaN QDs and (c, d) near-perfect GaN dots grown on a 6H-SiC(0001) substrate. The images are taken along the [11$\bar{2}$0] axis of GaN dots. (*From:* [67]. © 2002 American Institute of Physics. Reprinted with permission.)

The fabrication of nanocrystalline GaN was also realized by reactive laser ablation of pure Ga metal in a highly pure N_2 atmosphere [69]. The samples were collected from the surface of a membrane filter and then thermally annealed at 800°C in a highly pure ammonia atmosphere. TEM dark field images show a log-normal size distribution with a mean diameter of 12 nm and a standard deviation of 8 nm. SAED patterns confirm the hexagonal phase. The quantum confinement effect was observed from the blue shift of the size-selective PL and PLE spectra. Nanocrystalline GaN thin films were also fabricated recently on quartz substrates by RF sputtering using GaAs as a target material at a nitrogen pressure of 3.5×10^{-5} bar [70]. The average particle size of the nanocrystalline GaN increased from 3 to 16 nm when the substrate temperature was raised from 400°C to 550°C.

Crystalline GaN particles can be synthesized by simple inorganic reactions at various temperatures. Well et al. [71, 72] reported a method of nanosized GaN synthesis by pyrolysis of gallium imide $\{Ga(NH)_{3/2}\}_n$ at high temperatures. Formation of isolated spherical QDs in colloidal GaN solution was confirmed by TEM images,

shown in Figure 8.32 [73]. The image reveals that the GaN dots have a zinc-blende structure with diameters ranging from 2.3 to 4.5 nm. The absorption and PL peaks were observed to shift to a higher energy when compared to bulk GaN. In addition to GaN, an $Al_xGa_{1-x}N$ nanoparticle/polymer composite was also synthesized using a similar method and the microstructure of zinc-blende QDs was confirmed by HRTEM [74].

8.3 Optical Properties of III-Nitride QDs

Optical properties of III-nitride QDs were mostly obtained by PL. Few other investigations have been reported to date. The reported PL spectra show very different characteristics for the different samples prepared by different methods. For the GaN QDs alone, the published PL peak energies may be much lower or higher than the bandgap of bulk GaN, ranging from 2.15 to 3.9 eV depending on the PL measurement temperature, growth method, size, and distribution of QDs.

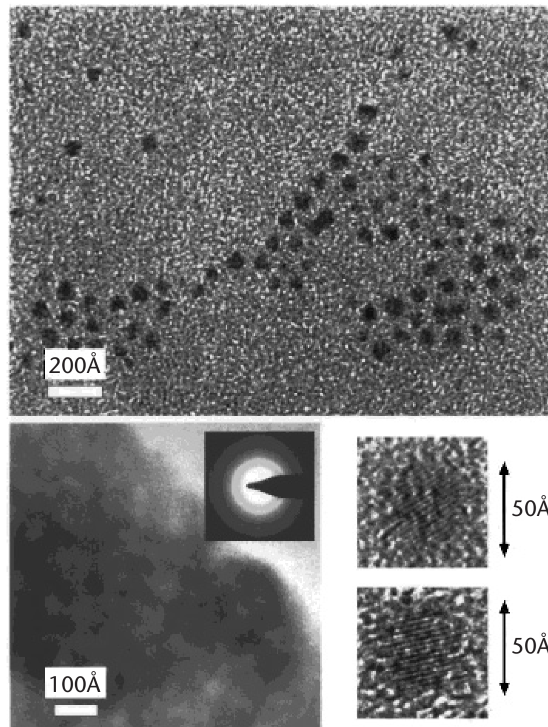


Figure 8.32 TEM image of GaN QDs taken in a bright field. The particles are well dispersed and not agglomerated. The top panel shows low magnification of QDs and some linear alignment. The bottom two right panels show high magnification and lattice fringes of QD oriented with the *o*-axis in the plane of the micrograph. The bottom left panel shows an electron diffraction pattern of GaN QDs indicating a zinc-blende structure. (*From:* [73]. © 1999 American Institute of Physics. Reprinted with permission.)

Figure 8.33 shows the PL spectra of five samples containing GaN QDs grown by MBE [75, 76]. The weak peaks near 3.45 eV are the band-edge PL from GaN bulk. The strong broad peaks are attributed to the PL from QDs. As shown in the figure, the QD-related peaks have different energies from ~2.6 to 3.9 eV for the different samples.

In fact, the large discrepancy between the PL peak energies of III-nitride QDs reflects the rich and interesting properties of III-nitride QDs. We first describe three important factors pertinent to III-nitride QDs: the effects of quantum confinement, strain, and polarization. Then we discuss in detail the PL and other relevant properties of GaN and $In_xGa_{1-x}N$ QDs grown by various techniques.

8.3.1 Effects of Quantum Confinement, Strain, and Polarization

The optical properties of III-nitride QD systems are influenced by the confinement effect and by the compressive strain, both producing a blue shift of the QD emissions compared with the GaN bulk one, and by the internal electric fields (spontaneous and piezoelectric), producing a significant red shift of the emitted light [77, 78].

The quantum confinement effect shifts the bandgap of a bulk semiconductor to higher energy. This shift, called the *confinement energy,* depends on the size and shape as well as the material properties of both QDs and surrounding matrix. Here we estimate the confinement energies for three simplified cases: a plate or disk, a cubic box, and a sphere. Assuming an infinite barrier, the confinement energy of the ground state for an electron in a rectangular box is given by

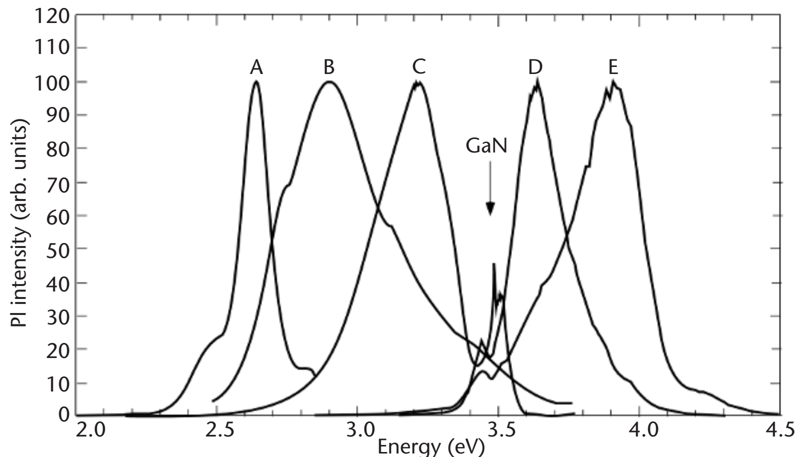


Figure 8.33 Spectra from five GaN/AlN QD samples grown by MBE on different substrates under different growth conditions. Samples A and B were grown on Si substrates while samples C, D, and E were grown on sapphire substrates. The spectra were measured at ~15K under the excitation of a Ti-sapphire laser (photon energy 5.06 eV). Intensity is normalized at the maximum of the QD-related signal. (*From:* [75]. © 2000 Materials Research Society. Reprinted with permission.)

$$E = \frac{h^2}{8}\left(\frac{1}{m_x d_x^2} + \frac{1}{m_y d_y^2} + \frac{1}{m_z d_z^2}\right)$$

where d_j ($j = x$, y, and z) are the dimensions of the box and m_j are the electron mass in j directions. For a plate- or disk-like dot, in which the in-plane sizes d_x and d_y are much larger than the height $d_z = d$, the confinement energy is simplified to $h^2/(8m_z d^2)$. The shift in the bandgap is calculated using $1/m_z = 1/m_{e,z} + 1/m_{h,z}$, where $m_{e,z}$ and $m_{h,z}$ are, respectively, the effective masses of the electrons and holes in III-nitride along the z-direction. For a cubic box of size d, the confinement energy is given by the same expression, $h^2/(8m_z d^2)$, but with $1/m_z$ replaced by $(2/m_{e,t} + 2/m_{h,t} + 1/m_{e,z} + 1/m_{h,z})$, where $m_{e,t}$ and $m_{h,t}$ are the transverse effective masses of the electrons and holes. In Figure 8.34, we plot the confinement energies as a function of d for a plate and a sphere. In both cases, the effective masses of $0.22m_0$ for electrons and $1.1m_0$ for holes were assumed [79]. For small dots a few nanometers in size, the confinement energy is very sensitive to the dot size. This sensitivity is much reduced when the size is increased. Quantitatively, when the size decreases from 10 to 2nm, the confinement energy may change from 20 meV to more than 1 eV depending on the shape of the dots.

The curves shown in Figure 8.34 represent good lower and upper bounds for the real confinement energy in GaN QDs, both free and embedded in AlN matrix, if the size is not too small. For electrons, the barrier height in the GaN/AlN interface (75% ΔE_g or 2.1 eV) [80] is really high and the mass anisotropy is small. For holes, the barrier is lower (0.7 eV) and the mass anisotropy is larger. However, the effect

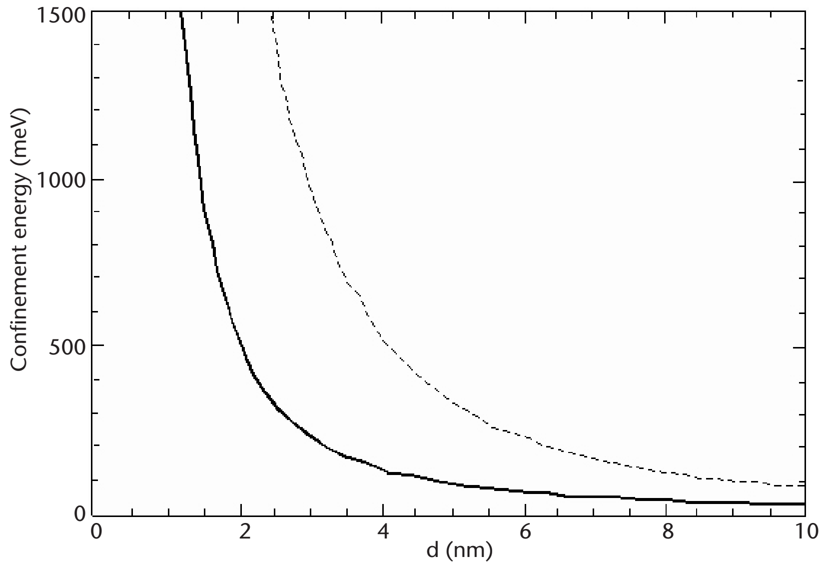


Figure 8.34 Shift in the bandgap energy from bulk GaN for a 2D GaN plate of thickness *d* (solid line) and a GaN sphere with a diameter *d* (dotted line) as a function of *d*. The shift in the bandgap energy in a cubic box of size *d* is between these two curves.

on confinement energy is much reduced due to the large masses of holes. For a specific GaN QD, the actual energy shift from the bulk value is expected to be within those two bounds mentioned earlier. In the case of self-assembled QDs, the plate-like results may be more suitable. In other cases such as GaN nanocrystallines, a sphere-like shape may be more appropriate.

The effect of strain in III-nitride heterostructures must be taken into account. For GaN QDs grown on $Al_xGa_{1-x}N$ and $In_xGa_{1-x}N$ QDs on GaN, the strain is commonly compressive, which induces a blue shift in the bandgap. This shift is proportional to the strain and the relevant deformation potentials. In the case of vanishing shear strain, the bandgap shift can be conveniently expressed as $\Delta E_g = (2\alpha + \beta q)\varepsilon_{xx}$ or $(2\alpha/q + \beta)\varepsilon_{zz}$ [81, 82] where $q = -2C_{13}/C_{33}$ and C_{13} and C_{33} are the elastic constants; ε_{xx} and ε_{zz} are the xx and zz components of the strain tensor; and α and β are the constants related to the various deformation potentials of both conduction and valence bands. The reported value of $(2\alpha + \beta q)$ is from –6 to –12 eV [81–84], depending on measurements and the choice of q (from –0.51 to –0.60) [85]. Assuming $(2\alpha + \beta q) = -10$ eV, we obtain $\Delta E_g = -10\varepsilon_{xx}$. It is 0.25 eV in the case of fully strained GaN on AlN. For a fully strained GaN on $Al_xGa_{1-x}N$, assuming a linear dependency of the $Al_xGa_{1-x}N$ lattice constant on x, the energy shift $\Delta E_g = 0.25x$ eV.

Compared to the strain induced by lattice mismatch, the strain induced by thermal mismatch is negligible. For a rough estimation, we use a temperature difference of 1,000K between the growth and the measurements. Using the thermal expansion coefficient $\Delta a/a$ of $5.6 \times 10^{-6}K^{-1}$ for GaN and $7.5 \times 10^{-6}K^{-1}$ for sapphire [86], a compressive strain of about 0.002 is obtained. This strain will only produce a blue shift in the bandgap of 20 meV.

The strain levels of the various compound materials can be obtained from the shift of the Raman modes from their unstressed frequency position, considering the phonon deformation potential of each Raman mode. Micro-Raman (μ-Raman) was employed to study the correlation between structure and optical properties of four different GaN/AlN QDs samples [87]. Raman modes (GaN and AlN) are observed for the four samples listed in Table 8.2, together with the calculated level of biaxial strain, according to the previous considerations. Table 8.3 shows the average Raman frequencies for E_2 (high). The differences in the level of biaxial strain in the GaN and AlN buffer or capping layers from one sample to another can be partially understood in terms of distinct levels of plastic relaxation due to the changes in

Table 8.2 Thickness of Different Layers of Studied Sample Fields, in Qualitative Agreement with the Features Observed in the CL Spectra

Sample	*First AlN Buffer Layer (nm)*	*GaN Buffer Layer (nm)*	*Second AlN Buffer Layer (nm)*	*GaN QDs Layer (nm)*	*AlN Barrier Layer (nm)*	*Number of Stacked Layers*
S2	30	300	300	2.6/0.8	45	2
S4	34	1100	330	2.8/2.3/1.6/1.0	9/9/11	4
S40	30	430	700	1.6	6.7	40
S85	30	430	700	1.6	6.7	85

Source: [87].

Table 8.3 Frequency of the E_2 (High) Raman Modes and Calculated Biaxial Strain

Sample	E_2^{GaN} *(High) Buffer Layer*	*Biaxial Strain (%)*	E_2^{GaN} *(High) QDs Layers*	*Biaxial Strain (%)*	E_2^{GaN} *(High) Buffer Layer*	*Biaxial Strain (%)*	E_2^{GaN} *(High) Barrier Layers*	*Biaxial Strain (%)*
S2	565.64	0.16	–	–	650.78	0.26	–	–
S4	564.57	0.24	–	–	652.54	0.21	–	–
S40	566.71	0.09	602.33	–2.73	653.95	0.17	643.37	0.47
S85	566.68	0.09	600.57	–2.25	652.89	0.20	642.67	0.50

Source: [87].

buffer layer thicknesses. The lower value of the residual strain of the QDs could be correlated to a more efficient relaxation process, due to the higher number of stacked layers. The spontaneous component depends basically on the materials properties and thickness of the layers; in the case of QD structures, this is correlated to the QD size. The piezoelectric contribution depends on the strain level of the QD structures and barrier layers, and hence on the QD size [77, 78, 88].

Both spontaneous polarization and strain-induced polarization (piezoelectric effect) produce large electric fields that have a significant effect on the optical properties of QDs. From the symmetry arguments, it can be concluded that the wurtzitic (hexagonal) crystal has the highest symmetry showing spontaneous polarization [89]. Bernardini et al. [90] have calculated the spontaneous polarization P_{spon} and the piezoelectric constants e_{31} and e_{33} for III-nitrides. By defining the spontaneous polarization as the difference of electronic polarization between the wurtzitic and zinc-blende structures, they obtaine P_{spon} dfrom first-principle calculations. As compared to GaAs, the piezoelectric constants of III-nitrides are about an order of magnitude larger.

The electric field associated with the polarization can be equivalently described by the bulk and interface polarization charges, $\rho_{p(r)} = \nabla \cdot P_{(r)}$ and $\sigma_p = (P_2 - P_1) \cdot n_{21}$. Here n_{21} is the direction of the interface pointing from medium 1 to 2 and $(P_2 - P_1)$ is the difference of the polarization across the interface. Consider a thick GaN layer grown on an AlN (0001) surface, where the interface charge due to the difference in spontaneous polarization in two crystals is –0.052 C/m^2 corresponding to a carrier density of 3.3×10^{13} cm^{-2}. The electric field E_P created by this polarization charge is $\sigma_P/(2\varepsilon_r\varepsilon_0)$, where $\varepsilon_r\varepsilon_0$ is the static dielectric constant. It is -2.8×10^6 V/cm in GaN and 3.5×10^6 V/cm in AlN, with the sign referenced to the growth direction. For a thin GaN plate straddled by AlN, as schematically shown in Figure 8.35, the polarization charges in both bottom and top interfaces must be considered and the result is doubled (-5.6×10^6 V/cm).

Additional polarization induced by strain can be calculated from the elastic and piezoelectric constants listed in Table 8.1, $P_z = 2e_{31}\varepsilon_{xx} + e_{33}\varepsilon_{zz} = 2(e_{31} - e_{33}C_{13}/C_{33})\varepsilon_{xx}$. For a fully strained GaN plate on bulk AlN, P_z is calculated to be 0.033 C/m^2. Remember that the effect of spontaneous and strain-induced polarization in GaN/AlN heterostructures is additive. As a result, the total electric field induced by

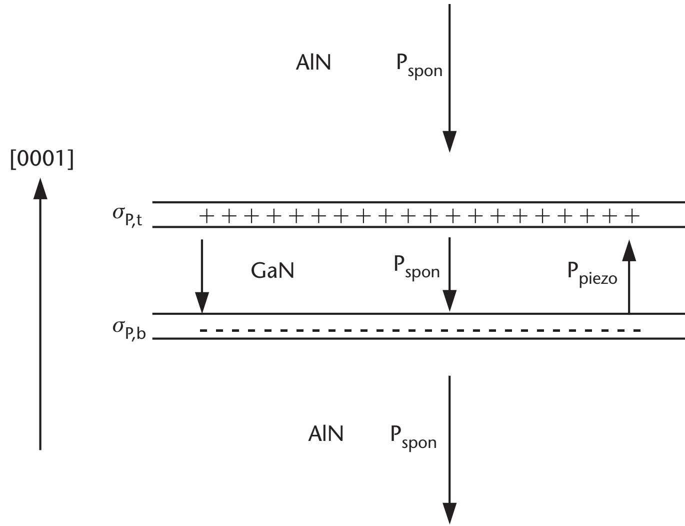


Figure 8.35 Schematic diagram of a GaN plate grown on and covered by thick (0001) AlN, showing the polarization effect. P_{spon} is the spontaneous polarization, which is –0.029 C/m^2 in GaN and –0.081 C/m^2 in AlN. P_{piezo} is the polarization induced by strain (2.5% in the in-plane direction in GaN), which is 0.033 C/m^2 only appearing in the GaN layer. The terms $\sigma_{P,b}$ and $\sigma_{P,t}$ represent the polarization charges at the bottom and top interfaces, which are –0.085 C/m^2 and 0.085 C/m^2, respectively. E is the electric field, which is -9.2×10^6 V/cm. Notice the directions of P and E with respect to [0001] and the signs of the interface charges.

spontaneous polarization and piezoelectricity is raised to -4.6×10^6 V/cm if only a single interface is considered. This field is doubled (-9.2×10^6 V/cm) in a thin GaN plate straddled by AlN. This huge field will drive free electrons in the GaN toward the top interface and holes toward the bottom interface. As a result of the *quantum confined stark effect* (QCSE) [91, 92] the electron-hole transition energy is greatly reduced.

The field-induced shift in transition energy in QWs has been investigated extensively by Miller et al. [92]. The shift is more significant for a wider well than a narrower one, particularly if the linear approximation holds. In much wider wells, a quadratic approach that reduces the red shift is used. In addition, if the Fermi levels come close to the conduction band on one side and valence band on the other, the resultant free carriers would screen the induced field, reducing the shift. A similar result is expected for plate-like QDs. For a rough estimate, the shift is approximately $E_p d$ in the limit of large size. This gives a value of 2 eV for a plate 4 nm thick under a field of 5×10^6 V/cm.

In the following subsections, we discuss various PL results encompassing the quantum confinement, strain, and polarization effects. Only when these effects are properly considered can the experimental results be reasonably understood. To reiterate, the quantum confinement and strain effects essentially induce a blue shift, whereas the polarization effect results in a red shift. In addition, the quantum confinement is accentuated in small dots, while the polarization effect is more significant in large and strained dots. Experimentally, it is possible to distinguish various

effects at least qualitatively. In the case of GaN QDs, if the observed PL has a lower energy than the bulk GaN bandgap, the polarization effect must be dominant. Otherwise, the quantum confinement effect is more important.

8.3.2 GaN QDs

The GaN QDs grown by laser ablation [93], RF sputtering [94], and chemical synthesis [95] are the examples in which the strain and polarization effect seem to be less important as compared to confinement. The reported PL peaks or absorption edges in all these cases are higher than the GaN bandgap and can be qualitatively attributed to the quantum confinement effect. The typical PL spectra show broad peaks with a width of a few hundreds meV, reflecting the large variations in dot size. In addition, noticeable Stokes shifts of the PL from the absorption (or PLE) peaks are often observed [93].

Preschilla et al. [94] estimated the bandgap of GaN QDs with different sizes from the absorption spectra. The QD nanocrystallites were deposited on quartz substrates by reactive RF sputtering. The GaN particle size changed with substrate temperature. The measured absorption edges were 3.4, 3.6, 3.9, and 4.1 eV for the average size of 16, 10, 7, and 3 nm, respectively. These energies agree reasonably well with those obtained from the PL spectra, as shown in Figure 8.36. The measured energy shifts from the bulk GaN bandgap are close to the confinement energies

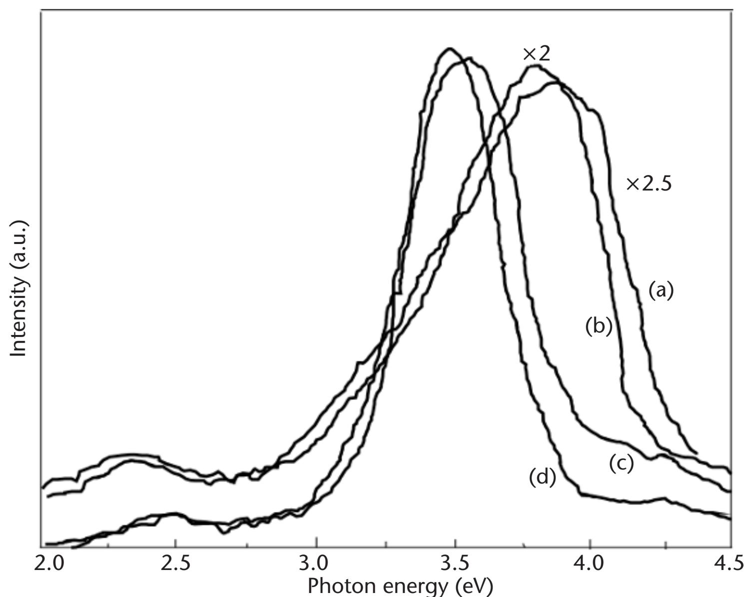


Figure 8.36 PL spectra from nanocrystalline GaN films deposited by RF sputtering at (a) 400°C, (b) 450°C, (c) 500°C, and (d) 550°C. The average particle size is found to correlate with growth temperatures and are 3, 7, 10, and 16 nm at growth temperatures of 400°C, 450°C, 500°C, and 550°C, respectively. The bandgaps derived from the absorption spectra are, respectively, 4.1, 3.9, 3.6, and 3.4 eV in the order of increasing particle size. (*From:* [94]. © 2000 American Institute of Physics. Reprinted with permission.)

of a spherical dot, except for the 7-nm sample. The measured shift in the latter case (0.5 eV) is much higher than what would be expected theoretically (~0.2 eV).

The GaN QDs grown by MBE are more disk-like and their size and shape are more controllable. Two very different cases were observed, as shown in Figure 8.37. In the case of what is called "small dots" having a 2.3-nm height and 8-nm diameter, a PL peak at 3.7 eV was observed at 2K. This is blue shifted from the bulk GaN bandgap by 0.2 eV. A similar PL peak (~3.8 eV) was also reported from the "small" GaN QDs (1.6 nm in height and 13 nm in diameter) with cubic structure grown by MBE on 3C-SiC(001) substrate [36]. However, in the case of what is called "large dots" having a 4.1-nm height and 17-nm diameter, the detected PL peak is 2.95 eV, nearly 0.5 eV lower than the bulk bandgap [11, 13, 14]. Considering the blue shift caused by quantum confinement (~0.1 eV; see Figure 8.34) and strain (~0.2 eV; see the discussions in the last subsection), the net red shift is even larger (~0.8 eV). This shift was interpreted as the result of the electric field induced by the polarization. As discussed earlier, this field could be on the order of 10 MV/cm and may induce a large Stark shift in transition energy. A simplified calculation using a 2D QW of thickness $H = 0.72h$, with h being the height of the QDs, shows that the 5.5 MV/cm electric field can account for the measured PL peaks from both large and small dots.

A more detailed calculation reported recently confirms the above results [96]. In this calculation, a realistic geometry of the QDs measured from TEM was used. The effects of strain, quantum confinement, and polarization were all considered. In addition to the ground state energy, the distributions of strain, electric field, and charges in the QDs were also obtained. Figure 8.38 presents the electron and hole

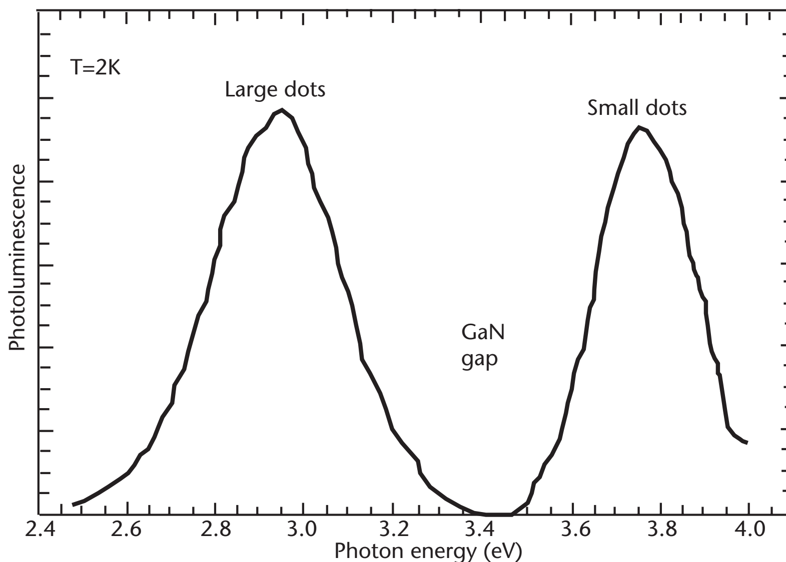


Figure 8.37 PL spectra of GaN quantum dots with a large (height/diameter = 4.1/17 nm) and small (2.3/8 nm) average size, measured at 2K. Note the large red shift observed for the large dots with respect to small ones. (*From:* [12]. © 1998 American Physics Society. Reprinted with permission.)

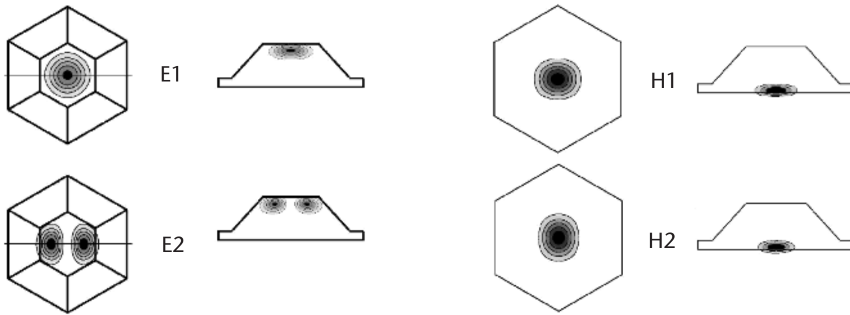


Figure 8.38 Probability density distribution for the two lowest electron and hole states in a GaN quantum dot embedded in (0001) AlN matrix, calculated in the framework of an 8-band **k** · **p** model with zero spin-orbit splitting. The GaN dot has the shape of a truncated pyramid with a hexagonal base. Note that the electron is mainly localized near the top interface, while the hole is near the bottom interface. (*From:* [96]. © 2000 American Physics Society. Reprinted with permission.)

distributions in the QD for the ground states and the first excited states. As described before, the electrons are more localized near the top interface, while the holes are near the bottom interface. It should be emphasized that the carrier distribution shown in Figure 8.38 is only correct for GaN grown on (0001) AlN face, called a *Ga face*. If the GaN QDs are grown on (000-1) AlN face, called an *N face,* then the electrons are expected to be driven toward the bottom interface and the holes toward the top interface.

The polarization effect is experimentally supported by the dependence of the PL spectra on excitation density. When the excitation density increases, the photocarriers can partially screen the polarization field. As a result, the built-in electric field decreases and the transition energy increases as observed experimentally [12].

The polarization effect was also used to interpret the origin of the blue, green, and orange PL (observed at room temperature; see Figure 8.39) from the GaN QDs in AlN matrix grown on Si(111) substrates [15]. These QDs were formed by growth of 7, 10, and 12 GaN MLs on AlN. The actual dot sizes were not reported. The PL peaks corresponding to different sizes were measured at about 2.8, 2.5, and 2.2 eV, all well below the bandgap of bulk GaN. It is interesting to note that, in the largest QDs of 12 MLs, the red shift is as large as 1.2 eV. By growing multilayers of GaN QDs with different sizes, white light emission is demonstrated.

Here, we should emphasize that, in real QDs, the polarization effect is also expected to be very sensitive to the shape, doping, and interface properties of the materials, since the polarization field may be screened by any of the other charges associated with free carriers and defects. These complications may account for some apparent discrepancies reported in the literature. For example, higher PL and CL peaks (~3.8 eV) than the GaN bandgap are reported from the GaN QDs in AlN matrix grown by MBE using the SK growth mode [12]. The size and the shape of the QDs measured from HRTEM, about 4 nm in height and 20 nm in diameter, are similar to the “large” dots mentioned earlier. Although the sample had multilayer

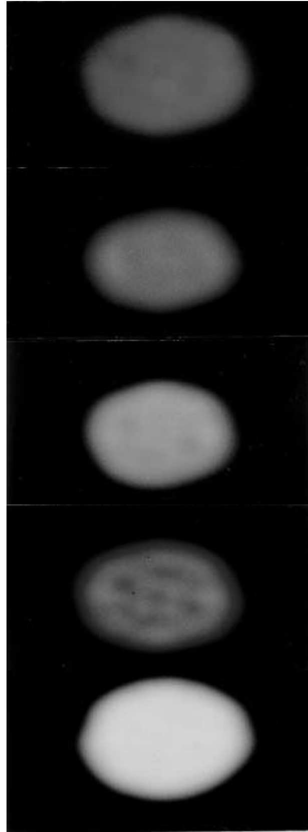


Figure 8.39 Photographs of the light emitted at room temperature from GaN/AlN QDs on Si(111) substrate excited by a 10-mW unfocused He-Cd laser (~0.3 W/cm^2). (*From:* [15]. © 1999 American Institute of Physics. Reprinted with permission.)

structures, the vertical correlation was negligible since the AlN spacing layer was sufficiently thick.

The PL spectra reported by Tanaka et al. [97, 98] associated with the GaN QDs in $Al_xGa_{1-x}N$ matrix are all near or higher than the GaN bandgap, ranging from 3.45 to 3.58 eV depending on the dot size and measurement temperature. The QDs were grown on $Al_xGa_{1-x}N$ with $x \sim 0.15$ by MOCVD using Si as "antisurfactant." The dot shapes are all disk-like and the sizes are mostly large, varying from 3.5 to 40 nm in height and 10 to 120 nm in diameter. A stimulated emission peak at 3.49 eV was also reported at 20K from small dots of 1 to 2 nm in height and 10 nm in diameter [99]. For large QDs, the energy shift due to quantum confinement may be less than 0.2 eV (see Figure 8.33). The shifts due to the strain and the strain-induced polarization are negligible since no strain is expected in these QDs grown through the use of an antisurfactant.

Although the polarization effect is not discussed in these papers, it can be estimated assuming a linear dependency of the spontaneous polarization on x in

$Al_xGa_{1-x}N$. We should note that this linear dependence is only an assumption for the sake of discussion. For $x = 0.15$, following the discussion in the last subsection, a polarization charge of 0.017 C/m^2 or a carrier density of 1.0×10^{13} cm^{-2} is calculated for the bottom interface [GaN on (0001) $Al_{0.15}Ga_{0.85}N$ interface]. The same amount of negative charge is expected at the top interface. The electric field in this case is ~1 MV/cm if the polarization charge from one interface only is considered. This value is doubled if both bottom and top interfaces are considered. The result will be different if the GaN QDs are covered by AlN instead of an alloy. As compared to the fully strained GaN QDs in AlN matrix grown using the SK mode, the polarization-induced electric field in this case is about five times smaller. The actual field may be lower due to Si doping, which is not well understood. As a result, the Stark shift due to this polarization may be small. It is difficult to get a more quantitative comparison. Any measured transition energy for a particular assembly of QDs must be the result of competition between the blue shift induced by quantum confinement and the red shift induced by polarization. It is expected to be not very far from the bulk GaN bandgap, as observed in the PL spectra.

The width of the PL spectra from GaN QDs ranges from a few tens to a few hundreds of meV, depending on sample and measurement temperature. As in the case of the peak energy, the published PL widths and their variation cannot be explained by quantum confinement alone. If only the quantum confinement is considered, a QD assembly with large average size would give narrow PL peak. However, when the polarization effect becomes important, the opposite holds. The QDs with a larger average size give a wider PL spectra (see, for example, Figure 8.36). In the latter case, the energy fluctuation of different QDs due to different Stark shifts has a more significant contribution to the PL linewidth. It is interesting to note that a CL of ~50 meV width was reported from a single GaN QD at 80K, which is much broader than expected. Although the origin of this widening is not clear, the result suggests that some other effect such as background doping may also have an important contribution to the CL (likewise PL) width.

The intraband absorptions at room temperature in GaN/AlN QDs grown by MBE was studied by HRTEM, PL, and photo-induced absorption spectroscopy [100]. Under p-polarized irradiation at Brewster's angle incidence, three interlevel resonances are observed, respectively peaked at 2.1 μm (2.36 μm), 1.46 μm (1.69 μm), and 1.28 μm (1.27 μm), as shown in Figure 8.40. The peak differential absorption ($-\Delta T/T$) amounts to 5.6×10^{-4} at 0.59 eV for the GaN QDs grown on Si (111) and 1.75×10^{-4} at 0.52 eV for the QDs grown on sapphire. The differential absorption under s-polarized irradiation is a factor of 2 to 3 smaller than that under p-polarized irradiation. It can be concluded that the photoinduced absorptions under p-polarized irradiation are mainly polarized along the c-axis, although some additional in-plane polarization component of intraband absorptions cannot be excluded.

The observed absorptions are believed to arise from electron interlevel transitions rather than from the valence band [100]. This interpretation is supported by the much larger band offset in the conduction band (< 2 eV) than in the valence band (0.82 eV), which allows substantial confinement for the electron excited

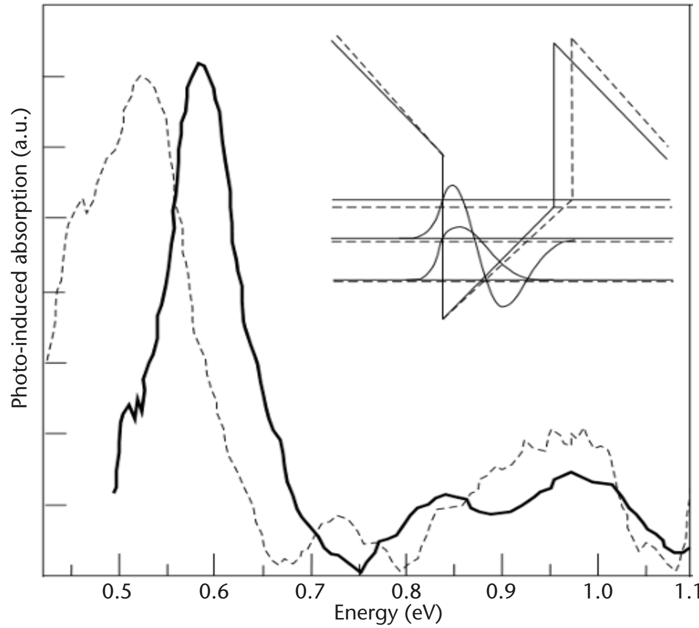


Figure 8.40 Normalized room temperature, photoinduced absorption spectra of QDs grown on Si(111) (solid curve) and on sapphire (dotted curve) samples at Brewster's angle incidence for *p*-polarized light under excitation at λ ~ 351 nm. The inset shows the conduction band profile and energy levels of the equivalent GaN/AlN quantum wells used in the simulations. The well width is 4 nm and the internal field in the GaN (AlN) layer 4.1 MV/cm (–3.8 MV/cm) (full curve). The corresponding values for the dashed curve are 4.5 nm and 3.7 MV/cm (–4.3 MV/cm), respectively. (*From:* [101]. © 2002 American Institute of Physics. Reprinted with permission.)

states. To further disentangle the origin of the observed transitions, the electron confinement energies account for the built-in electric field but neglect the in-plane confinement. The built-in field in the GaN dots and AlN barriers is estimated by assuming that the potential drop over one period of the superlattice is zero. Following [102], the boundary conditions over one period of the superlattice imposes

$$L_d F_d + L_b F_b = 0 \text{ and } F_d = L_b \Delta P / \left(\varepsilon_b L_d + \varepsilon_b L_b\right)$$

where F_d (F_b) is the field in the GaN dot (AlN barrier), L_d (L_b) is the dot height (barrier thickness), ε_d(ε_b) the dielectric constant for GaN (AlN) and P the zero-field polarization difference between the well and barrier material. The e_1e_2 (e_1e_3) transition energy of the QW is calculated to range from 0.52 eV (0.99 eV) to 0.62 eV (1.13 eV) when $\Delta P/\varepsilon_d$ is varied from 7 to 9 MV/cm. So the peak at 0.59 eV of the GaN QDs is attributed to transition from the *s*-like state to the *p*-like state, with one node of the envelope wave function along the *c*-axis. The transition at 0.97 eV corresponds to the *s*- to *d*-like (2 nodes along *c*-axis) transition. The transition at 0.85 eV could be the transition starting from the ground state to a state with one node of the envelope wave function both along the *c*-axis and in the layer plane.

The optical properties of GaN self-assembled QD structures was also studied by PLE, time-resolved PL, and cathodoluminescence (CL) [101]. As shown in Figure 8.41, a self-assembled GaN QDs structure consisting of five-stack GaN QDs with a $Al_{0.25}Ga_{0.75}N$ spacers was studied. The typical QD density was 10^{11} cm^{-2} range and the aspect ratio (height/diameter) was about 1/5. The Stokes-like shift between the AlGaN-related emission and the PLE absorption edge has been attributed to the alloy potential fluctuation in AlGaN alloy [103]. By measuring PLE spectra for the QD emission with various detection energies of 3.65, 3.54, and 3.44 eV, as shown in Figure 8.41, the energy differences correspond to ~220, 280, and 330 meV, respectively. This indicates that the apparent Stokes-like shift in the GaN QD emissions increases with decreasing emission energy, and hence increasing the QD size, which can be attributed to the enhancement of the separation of wave function overlap due to the built-in internal field present in the QDs [104, 105].

The time-resolved PL measurements at 10K [Figure 8.42(a)] show a temporal evolution of PL spectra of the GaN QD sample. The GaN QD emission persists well within a few nanoseconds time range. The decay time becomes longer with decreasing emission energy, and hence with increasing QD size, as shown in Figure 8.42(b). This is due to reduction of electron and hole wave function overlap with increasing QD size, since the emission energy is affected by the quantum confined stark effect

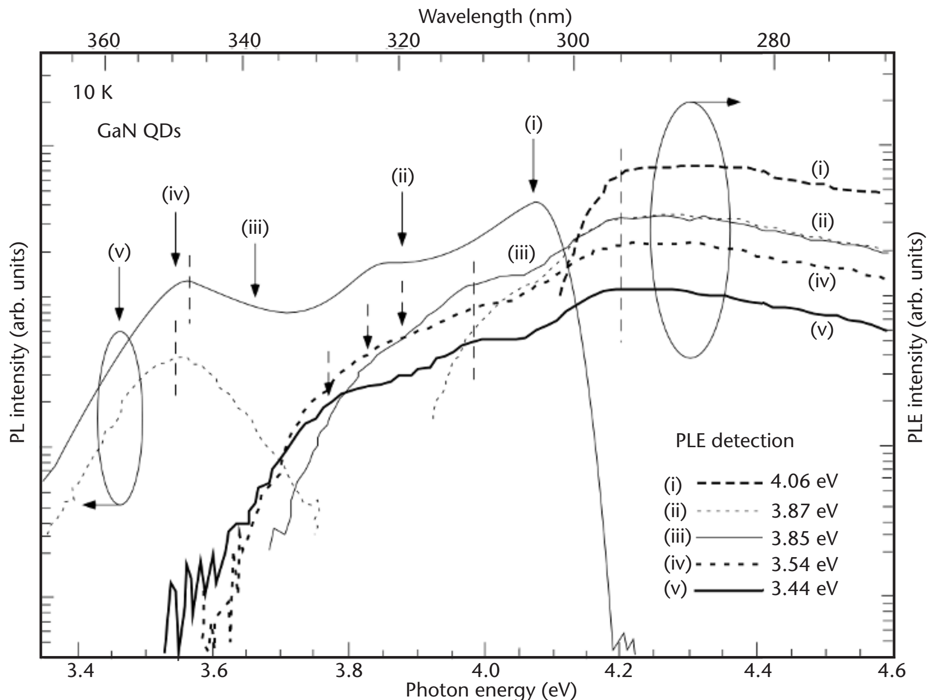


Figure 8.41 The 10K PL and PLE spectra for GaN self-assembled QDs. The PLE detection energies are indicated in the figure. A large Stokes-like shift of the PL emission from the GaN QDs with respect to the band edge measured by PLE is observed. (*From:* [101]. © 2002 American Institute of Physics. Reprinted with permission.)

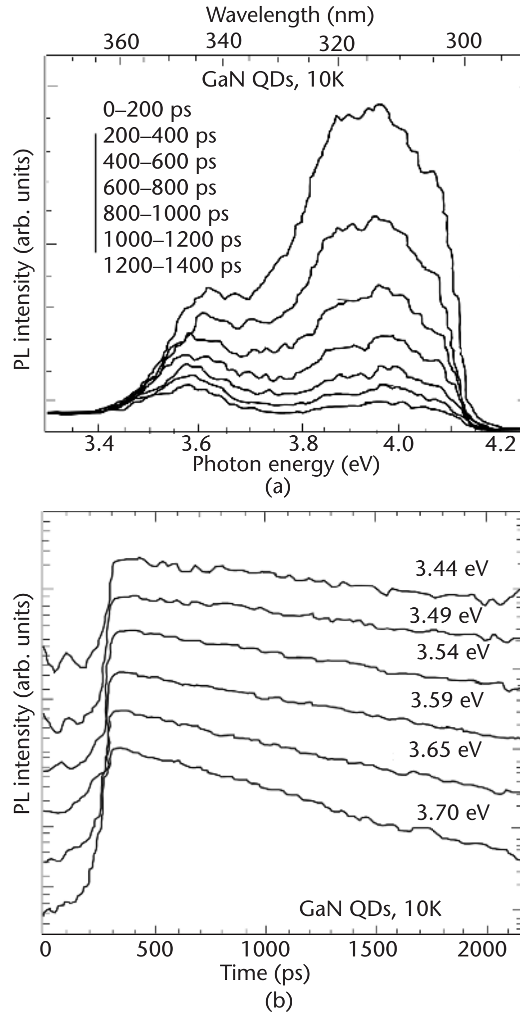


Figure 8.42 (a) Time-resolved PL spectra of the GaN QD sample integrated during various time intervals after excitation at 10K. (b) Temporal evolution of PL spectra at various emission energies of the GaN QDs at 10K. All spectra in part (b) are normalized and shifted in the vertical direction for clarity. Note that the lower energy side of the PL peak has a longer lifetime than the higher energy side in QD emission. The decay time for the AlGaN emission was estimated to be ~466 ps. (*From:* [101]. © 2002 American Institute of Physics. Reprinted with permission.)

with built-in polarization internal fields in GaN quantum structures. The decay time does not increase further when the emission energy decreases below ~3.54 eV because the recombination in these emission energies (larger QDs) may be affected by nonradiative recombination processes. The influence of the built-in electric field plays a critical role in the carrier recombination in hexagonal GaN self-assembled QDs.

One of the most important properties distinguishing QDs from QWs and bulk materials is the better thermal stability of PL. This property is clearly demonstrated in the GaN QDs embedded in AlN matrix. Typically, the PL intensity from bulk

materials quenches rapidly as the temperature is raised from low to room temperature. In the case of GaN QDs, a less temperature dependency is usually observed. An example is given by Widmann et al. [11] who compared the PL from GaN QDs grown on GaN substrate at and 300K. As shown in Figure 8.43, the PL spectra at 2K are dominated by two peaks at 3.46 and 3.75 eV from GaN substrate and QDs, respectively. When the temperature was raised to 300K, the PL from GaN substrate totally disappeared, while the intensity of the PL from the QDs remained the same. A temperature dependency of the PL spectra from the cubic GaN QDs in AlN matrix was reported by Martinez-Guerrero et al. [106]. When the sample temperature was raised from 10 to 200K, the integrated PL intensity changed very little. When the temperature was further increased to 300K, an increase in the PL intensity rather than a decrease was observed. The latter phenomenon was attributed to the increase in the laser absorption resulting from a reduction in GaN bandgap with temperature.

The thermal stability of the GaN QDs in $Al_xGa_{1-x}N$ matrix is not as good as that in AlN matrix. Typically, when the temperature is raised from 10 to ~300K, a decrease in the PL intensity from a few to more than 10 times is observed. Although the PL intensity and its change with temperature may also depend on the quality of the material, the poorer thermal stability in the GaN/$Al_xGa_{1-x}N$ QDs compared to

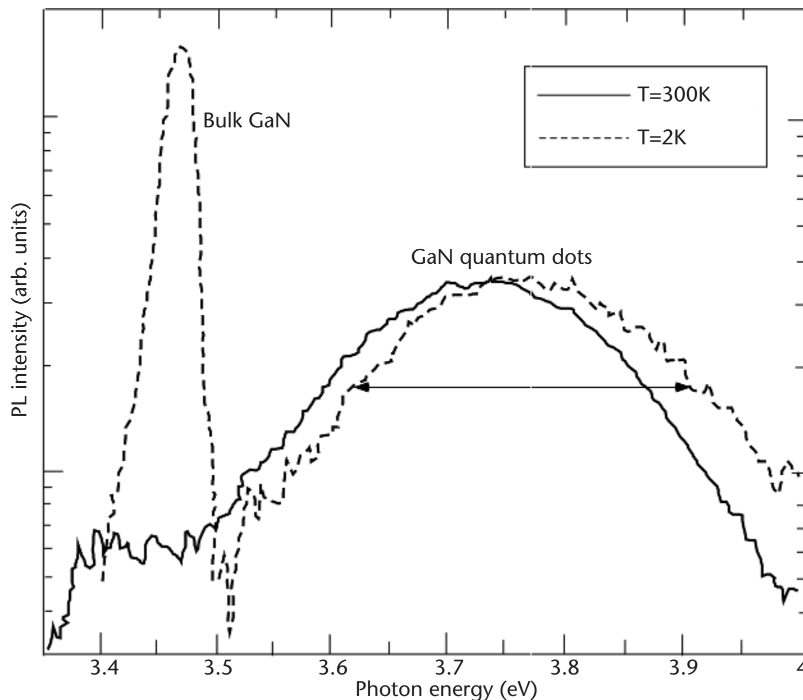


Figure 8.43 Photoluminescence spectra from GaN QDs embedded in AlN matrix grown on GaN substrate, measured at 2 and 300K, showing the higher thermal stability of the PL from QDs than that from GaN substrate. (*From:* [11]. © 1998 American Institute of Physics. Reprinted with permission.)

the GaN/AlN QDs can be interpreted by the lower barrier and the poorer confinement of the carriers.

For the QD in $Al_xGa_{1-x}N$ matrix, PL emission peak energy does not change much with temperature (Figure 8.44). The QD emission slightly blue-shifts up to about 175K, and then red-shifts above 175K. This blue shift below 174K can be partly attributed to the smaller Stokes-like shift of the smaller QD size. At temperatures higher than 175K, the carriers in smaller QDs, which are more thermally activated, can move into the larger QDs, so the lower energy part of the PL spectrum becomes dominant.

Ramvall et al. [51, 52] investigated the electron-phonon coupling in GaN/ $Al_xGa_{1-x}N$ QDs. The binding energy of A excitons in bulk wurtzite GaN was measured to be 21 meV [107, 108]. In quantum dots, the electrons and holes must be closer to each other due to confinement. Thus, the electron-hole Coulomb interaction as well as the exciton binding energy should increase. From the temperature dependency of the PL peaks observed in GaN/$Al_xGa_{1-x}N$ QDs with different sizes, an increase in the exciton binding energy as a function of QD size was derived by Ramvall et al. [51, 52]. The results are shown in Figure 8.45. As the QD height is reduced from 7.1 to 1.25 a_x (a_x = 2.8 nm being the Bohr radius of the excitons in bulk GaN),

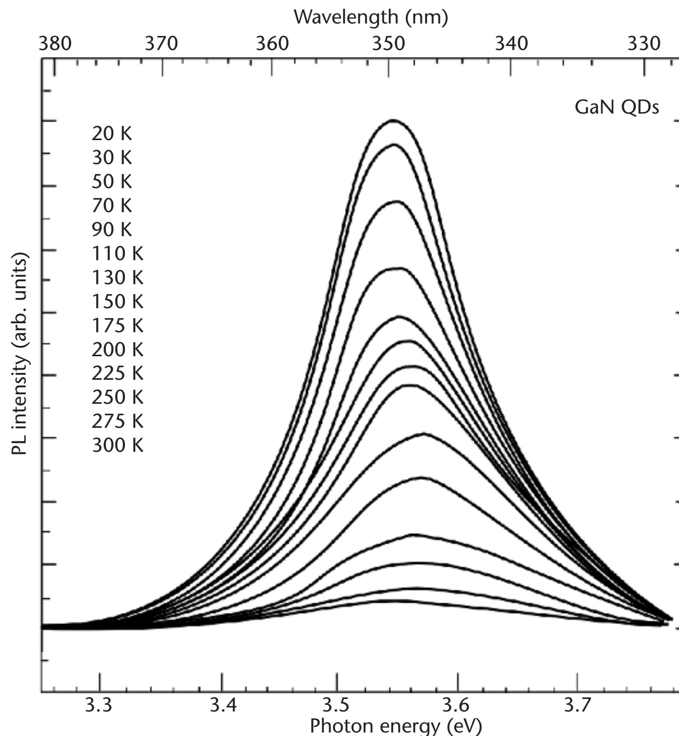


Figure 8.44 GaN QD emission spectra for the excitation wavelength of 325 nm in the temperature range from 20 to 300K. The main emission peak and intensity do not change much with increasing temperature. (*From:* [101]. © 2002 American Institute of Physics. Reprinted with permission.)

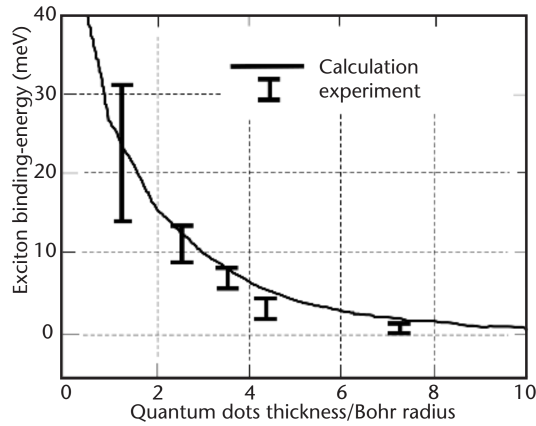


Figure 8.45 The binding energy of the excitons in GaN QDs in $Al_xGa_{1-x}N$ matrix as a function of dot size, showing an increase in the binding energy with reducing dot size. (*From:* [52]. © 2000 American Institute of Physics. Reprinted with permission.)

or from 20 to 3.5 nm, an increase in the exciton binding energy as large as ~20m eV is found. The experimental results are supported by the variational calculation of the excitons in 2D structures. By fitting the measured temperature dependency of the PL peaks using the theoretical expression, $\Delta E_g = -\alpha_0[f(\hbar\omega_0) + 1]$, they derived the electron-phonon coupling constant α_{0QD} relative to bulk value for the QDs with different sizes, as shown in Figure 8.45. In the above expression, $f(\hbar\omega_{L0}) = 1/[\exp(\hbar\omega_{L0}/kT) - 1]$ is the distribution function of the LO phonons and $\hbar\omega_{L0}$ = 91.7 meV. They found that $\alpha_{0QD}/\alpha_{0Bulk}$ is close to 1 for the QDs of size 12/36 nm (height/diameter). This ratio decreases to 0.81 and 0.63 as the size is reduced to 10/30 and 7/21 nm, respectively. A decrease of electron-phonon coupling for small QDs is expected when the separation of the discrete energy levels is large and very different from the phonon energies.

Except for light emission properties, a few micro-Raman (μ-Raman) studies [35, 87] were employed to investigate the correlation between structure and optical properties of GaN/AlN QDs samples. Gleize et al. [35] reported the Raman spectra from two different samples of multilayer (80 and 88 periods) GaN QDs embedded in AlN matrix, grown by MBE using SK mode. Under the excitation in the visible range, far from the resonant condition, they observed new peaks at 606 cm^{-1} from sample A and 603 cm^{-1} from sample B (see Figure 8.46 for the spectra from sample B). Both peaks were assigned to the E_2 phonons in GaN QDs rather than any disorder activated scattering. This assignment was arrived at from the selection rules and a comparison to the Raman spectra from an $Al_xGa_{1-x}N$ film. From the observed peak positions, the biaxial strain in the QDs was estimated to be –2.6% for sample A and –2.4% for sample B. The results are consistent with the fully strained GaN QDs in AlN matrix as measured by HRTEM in similar samples.

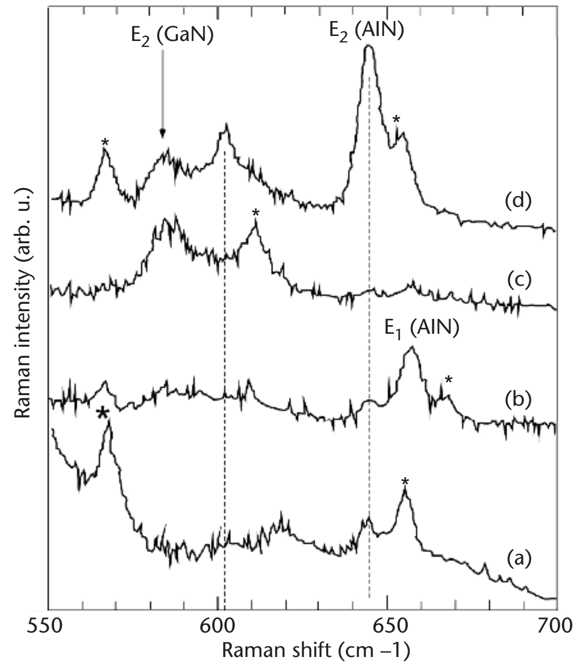


Figure 8.46 Micro-Raman spectra of GaN quantum dots in AlN matrix measured at room temperature using a 647-nm laser line, in four different configurations. (a) *z*(*xy*)-*z*, (b) *x*(*yz*)-*x*, (c) *x*(*zz*)-*x*, and (d) *x*(*yy*)-*x*. Asterisks mark phonons from the buffer layers. Features at about 618 cm^{-1} are related to the Si(111) substrate. (*From:* [35]. © 2000 American Institute of Physics. Reprinted with permission.)

The fully 3D study of the multiexciton optical response of GaN DQs is crucial in understanding the fundamental properties of particle/exciton interactions. De Rinaldis et al. [109] provide a detailed investigation of the interplay between single-particle carrier confinement and two-body Coulomb interaction in coupled GaN-based QDs. Their analysis shows that it is possible to engineer the interdot biexcitonic shift by varying the height and width of the dots.

The quantum dot size dependence of radiative recombination efficiency is studied by time-resolved PL measurements [private communication with A. Neogi et al.]. Multiple periods of GaN QDs in this study were grown by MBE with varying QD sizes [private communication with A. Neogi et al.]. In time-resolved decay measurements, the lifetime for smaller dots is relatively constant across most of the PL peak compared to bigger dots, indicating stronger carrier localization in smaller dots. As shown in Figure 8.47, the lifetime decreases with the dot size because in smaller dots the overlap between the electron and hole wave function increases with decreasing dot size [110]. The increase in the decay rate with decreasing emission energy is likely due to the exciton localization at potential wells of larger dots.

Temperature dependence of the PL decay time of GaN QDs with varying dot size is shown in Figure 8.48. The lifetime in the QDs increases or remains constant until about 125K, which strongly suggests that the decay is dominated by radiative

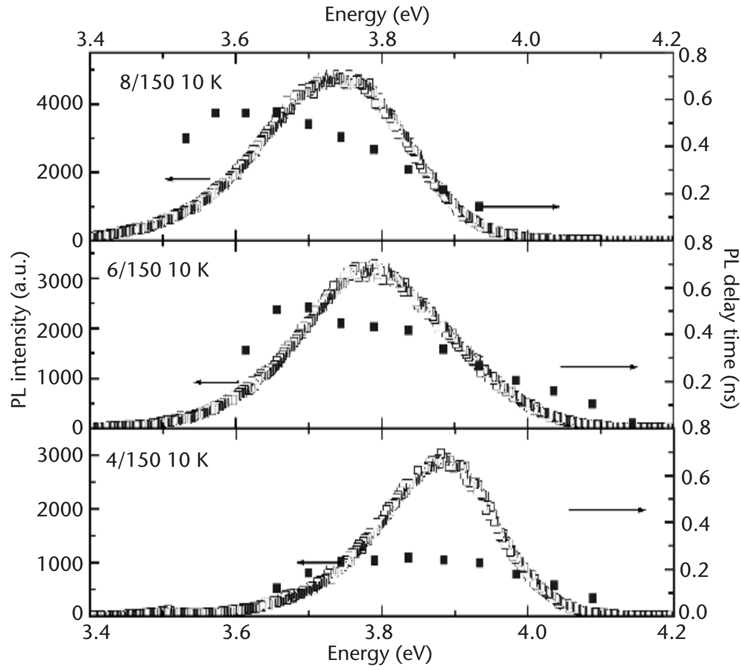


Figure 8.47 Variation of recombination lifetime with emission energy in MQD 8/150, 6/150 and 4/150 at 10K.

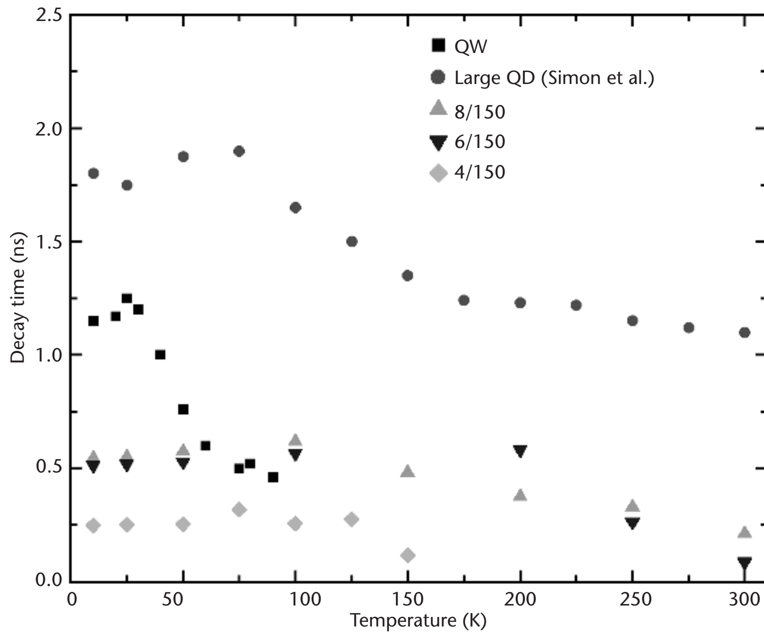


Figure 8.48 Spectrally integrated PL intensity as a function of temperature in MQD structures, QWs, and large MQD. (*From:* [11]. © 2001 American Institute of Physics. Reprinted with permission.)

recombination at lower temperatures. This is also confirmed by Simon et al. [111]. The QDs exhibits high PL efficiency at room temperature. The carrier recombination dynamics in the multiple-period QDs was compared with single-period QDs [112]. The stacking of multiple layers is expected to influence the optical properties due to the strain-induced realignment of dots in one level with respect to the adjacent level. The pure dimensionality effects induced in the multiple QDs are isolated from the influence of the giant polarization field in the single QD system, which can be as high as 7 MV/cm and significantly modifies the optical emission spectra. Figure 8.49 shows the time-integrated PL spectra of single-period QDs and multiple periods at 10 and 100K. The temperature dependence of the PL peak energy in multiple periods shows a blue shift with increasing temperature compared to the red shift observed previously in hexagonal QD, due to Frohlich interaction [111]. The blue shift was also observed in MBE-grown multiple QDs by Brown et al. [113], and can be ascribed to the preferential loss of carriers from larger dots.

In comparison, in the case of a single QD layer, the PL emission peak has a red shift with respect to the multiple-period QDs and bulk GaN peak at 3.45 eV. This red shift is a manifestation of the internal polarization-induced electric field present in wutrzite GaN QDs. In multiple-period QD structures (as shown in Figure 8.50), no significant difference is seen in the decay constants from 10 to 100K for the relaxation from the QD energy states (~3.55–3.65 eV). The fast relaxation component due to nonradiative recombination appears at ~200K and reduces the PL efficiency. It is observed that the effective recombination lifetime in the MQD structure (τD) is several hundred picoseconds at 100K for the most efficient emission

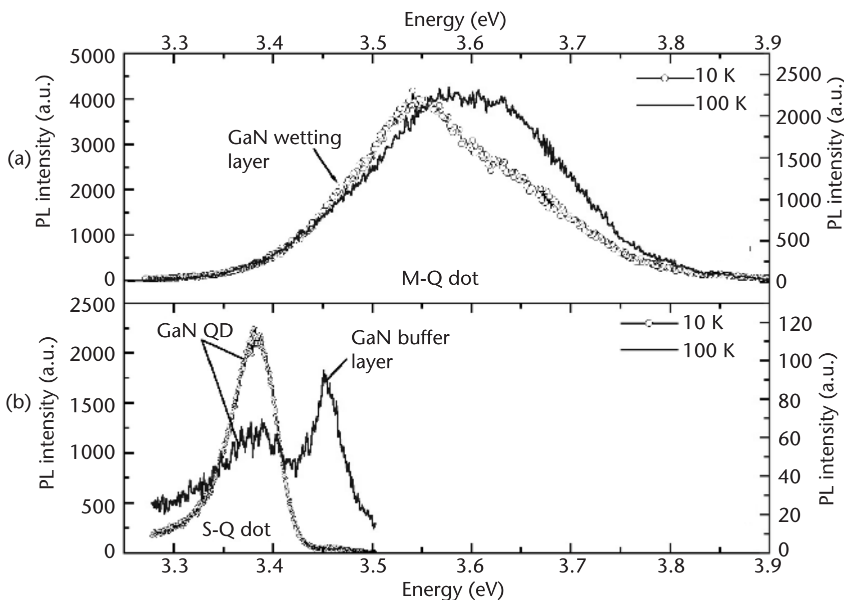


Figure 8.49 Time-integrated PL spectra in (a) a 20-period QD and (b) a single-period QD. (*From:* [112]. © 2003 IEEE. Reprinted with permission.)

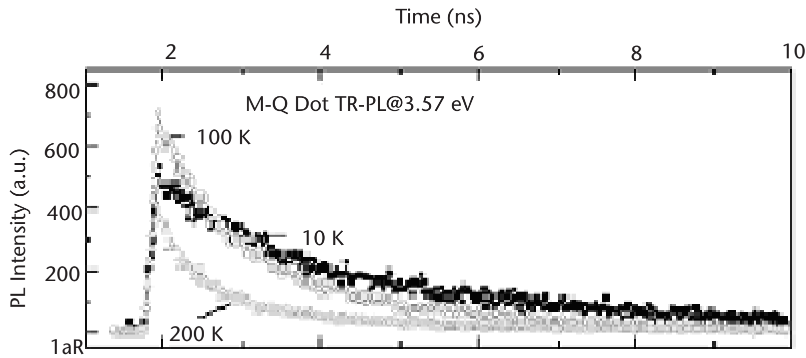


Figure 8.50 Time-resolved PL spectra in a 20-period QD. Detection was at 3.57 eV at the QD emission energy. (*From:* [112]. © 2003 IEEE. Reprinted with permission.)

wavelength, which are 2 orders of magnitude faster than highly strained QDs grown over thicker AlN spacer layers [111, 114]. This difference in the reduction of the internal electric field is due to the use of thin AlN spacer layers.

8.3.3 InGaN QDs

Compared to GaN, much less research has been reported on $In_xGa_{1-x}N$ QDs. Instead, investigations have focused on the origin of the PL from $In_xGa_{1-x}N$ QWs in which the In-rich structures can be formed through phase segregation or alloy fluctuation. It is believed that the intense PL from $In_xGa_{1-x}N$ wells may be associated with dot-like structures. In this section, in addition to the PL properties of $In_xGa_{1-x}N$ QDs, we discuss some relevant features of $In_xGa_{1-x}N$ QWs.

InGaN is a quantum material [115] and does not form perfect alloy due to its phase separation during growth [59, 116]. An InGaN QD-like region in an InGaN active layer was investigated by means of a variational approach, including 3D confinement of the electrons and holes in QDs and a strong built-in electric field effect due to the piezoelectricity and spontaneous polarization [117]. The built-in electric field leads to an obvious reduction of the effective bandgap of QDs and induces a remarkable spatial separation of electrons and holes in the QD-like region. The interband optical transition of electrons and holes in the $In_xGa_{1-x}N$ active layer is thus determined by two important factors: the 3D confinement of electrons and holes in the QD-like regions and strong built-in electric field effect.

An isolated cylindrical $In_xGa_{1-x}N$ QD with radius R and height L along the z-direction, surrounded by two large energy gap materials $In_yGa_{1-y}N$ ($y < x$) in the radial direction and GaN in the z-direction, is shown in Figure 8.51. The exciton binding energy E_b can be computed as follows:

$$E_b = E_g + E_e + E_h - E_{\text{ex}}$$

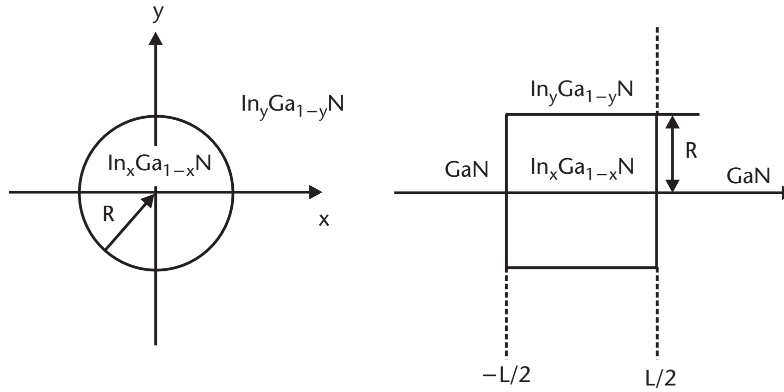


Figure 8.51 Diagram of a cylindrical GaN/$In_xGa_{1-x}N$/GaN QD with radius *R* and height *L*, surrounded by two large energy gap materials $In_yGa_{1-y}N$ (*y x*) in the radial direction and GaN in the *z*-direction. (*From:* [117]. © 2002 IEEE. Reprinted with permission.)

where E_e (E_h) is the electron (hole) confinement energy in the QD [117]. A comparison of the calculated and measured emission wavelengths is listed in Table 8.4. From Table 8.4, we can see that the InGaN QD-like regions usually have a typical nanometer-scale size with structural parameters *L* and *R* in the order of a few of nanometers [118], rather than nearly pure InN QDs [115]. The exciton binding energy E_b, the emission wavelength, and the electron-hole recombination rate as a function of height *L* of InGaN/GaN QDs is shown in Figure 8.52. The exciton binding energy decreases with increasing height *L* of the QDs, mainly because the relative distance in the *z*-direction between the electron and hole increases when *L* increases and the Coulomb interaction is thus reduced. It can also be observed from Figure 8.52(a) that the exciton binding energy for the $m = 0$ state is larger than that for the $m = 1$ excited state because the electron (hole) wave function for the $m = 0$ state is smaller than that for the $m = 1$ state. The built-in electric field due to the

Table 8.4 Calculated Emission Wavelength λ^{calc} of the $m = 0$ Optically Active Exciton State for $In_xGa_{1-x}N$/GaN QDs Surrounded by $In_yGa_{1-y}N$ Material in the Radial Direction and Measured Emission Wavelength λ^{expt}

L (nm)	*R (nm)*	*x*	*y*	λ^{calc} *(nm)*	λ^{expt}
2.4*	9	0.12	0.02*	427	430*
2**	5	0.21	0.05	471	470**
2.9**	5	0.20	0.02	504	505**

Source: [117].
* Data from: [63]
** The experimental data of the emission wavelengths have been taken from Dr. Zhi-Jian Yang's experiments for InGaN/GaN stuctures with an InGaN layer thickness of 2 and 2.9 nm and with an average In composition $\chi = 0.15$ for the InGaN layer.

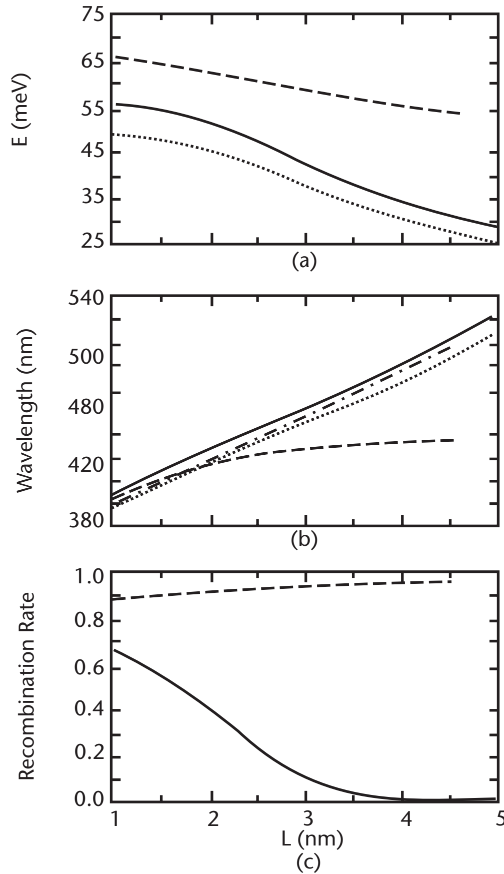


Figure 8.52 (a) Exciton binding energy, (b) emission wavelength, and (c) electron-hole recombination rate as a function of height *L* of $In_{0.15}Ga_{0.85}N$/GaN QDs with radius $R = 5$ nm, surrounded by $In_{0.02}Ga_{0.98}N$ material in the radial direction. The solid and dotted lines are for the $m = 0$ and 1 states with the effect of the built-in electric field. The dashed line is for the $m = 0$ state without the effect of the built-in electric field. The dot-dashed line in part (b) is for the $m = 0$ state with free electron-hole recombination. Here *m* denotes the electron *z*-component angular momentum quantum number. (*From:* [117]. © 2002 IEEE. Reprinted with permission.)

spontaneous and the piezoelectric polarizations has a remarkable influence on the exciton binding energy, especially for the QDs with large height *L*.

The reported PL peaks from $In_xGa_{1-x}N$ QDs range from 2.2 eV to a value higher than 3 eV, depending on the sample and measurement temperature [116]. All three effects discussed in GaN QDs, the quantum confinement, strain, and polarization, are also important in $In_xGa_{1-x}N$ QDs. The differences in the two systems are mainly rooted in the alloy nature of the latter material. The bandgap of $In_xGa_{1-x}N$ is sensitive to the alloy composition. A 1% fluctuation in *x* will lead to a shift in bandgap by 15meV if a linear dependency of $E_g(In_xGa_{1-x}N)$ on *x* is assumed. For $x \sim 10\%$, this shift can be much larger (~40 meV) due to the large bowing factor of the alloy [119]. Because it is hard to determine an accurate value for *x* and its spatial fluctuation, an

uncertainty in transition energy is introduced, making it difficult to derive reliable information from the energies of the measured PL peaks that are also affected by confinement, strain, and polarization effects.

Typical PL spectra from $In_xGa_{1-x}N$ QDs in a GaN matrix were reported by Damilano et al. [120] (Figure 8.53). The self-assembled QDs were grown on GaN by MBE using the SK growth mode. The In mole fraction x was kept constant (0.15). The dot or island density was about 5×10^{10} cm^{-2}, roughly an order of magnitude higher than the defect density. The dots with different sizes were obtained by controlling the thickness of $In_xGa_{1-x}N$. Figure 8.53 presents three sets of PL spectra from the $In_xGa_{1-x}N$ QDs with nominal thickness of 5, 9, and 12 ML [10 ML for part (b)] at 10K and room temperature. As the QD size reduces, the PL peak shifts to higher energy with an amount larger than that expected from quantum confinement alone. The extra shift is interpreted as the result of a quantum confined Stark effect induced by polarization. This interpretation is consistent with the observation that the PL intensity is quenched for large dots, as expected from the quantum confined Stark

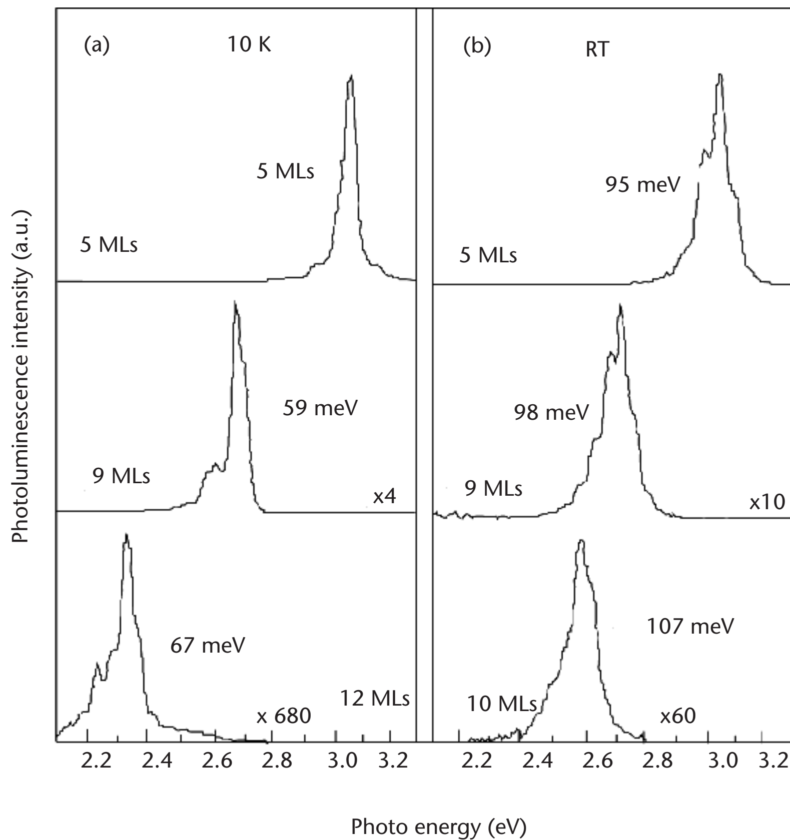


Figure 8.53 (a) 10K and (b) room temperature PL spectra of $In_xGa_{1-x}N$ QDs in GaN matrices of varying sizes. The nominal $In_xGa_{1-x}N$ thickness is 5, 9, and 12 ML in part (a) and 5, 9, and 10 ML in part (b). (*From:* [120]. © 1999 American Institute of Physics. Reprinted with permission.)

effect [121]. In their case, the PL was not detected for the dots larger than 10 ML at room temperature.

Let us estimate the polarization-induced electric field in $In_xGa_{1-x}N$/GaN(0001) QDs. For $x = 0.15$, assuming the dots are fully strained and using the data listed in Table 8.1, we obtain the interface charge due to the spontaneous polarization as –0.024 C/m^2. The charge due to the piezoelectricity is –0.021 C/m^2, similar to that from the spontaneous polarization. The electric field in QDs induced by these charges is –2.3 MV/cm in QDs if only one interface is considered. The field is doubled if the top and bottom interfaces are both taken into account. This field may be large enough to produce a large shift in the transition energy and quench the oscillator strength for the large dots.

The thermal stability of the PL from $In_xGa_{1-x}N$/GaN QDs was improved as compared to $In_xGa_{1-x}N$ bulk or QWs [41, 42], but not as good as that observed in GaN QDs in AlN matrix. Typically a few times drop in PL intensity was observed when the temperature was raised from low to room temperature for $In_xGa_{1-x}N$ QDs, whereas a stable PL was reported in GaN/AlN QDs [11]. The lower barrier may account for the poorer thermal stability, similar to the cases of GaN/$Al_xGa_{1-x}N$ QDs. The high dot density may also have made a contribution since the thermal tunneling between the adjacent dots may not negligible in this case.

A strong confinement of excitons in $In_xGa_{1-x}N$ QDs was demonstrated by the observation of very sharp and discrete PL lines from a limited number of QDs [116]. The QDs were grown on GaN by MOCVD and their average height/diameter was 4.6/23.4 nm. By patterning a thin aluminum layer with a 400-nm square aperture on the sample surface, the detected number of QDs in the PL measurement was reduced to about 20. The measured PL linewidth was typically 0.17 meV at 3.5K (see Figure 8.54) and limited by the spectral resolution of the measurement system. The separation of these discrete lines is a few meV. When the temperature is raised to 60K, the width increases to 0.6 meV but is much smaller than the thermal energy kT. The result suggests that the narrow PL lines are really from the strongly localized states. The broadening of the PL spectra commonly observed from a QD assembly is inhomogeneous either from the fluctuation of dot size or alloy composition.

The subband emission similar to the discrete lines described earlier was also reported from $In_xGa_{1-x}N$ QW laser diodes under room temperature pulse operation [122]. The energy separation is 1 to 5 meV. Under the room temperature CW operation, a single mode was observed but the mode hopped toward a higher energy with increasing drive current in contrast to that observed in an AlGaInP QW laser. A subband emission was suggested based on the transitions between the ground levels of the $In_xGa_{1-x}N$ QDs of different sizes formed from In-rich regions in the $In_xGa_{1-x}N$ well layers.

The existence of the QD-like structures and its effect on the PL in $In_xGa_{1-x}N$ QWs have been the subject of extensive discussion in the past few years [123, 124]. The main characteristics of the observed PL spectra are summarized as follows: (1) The luminescence efficiency is high. (2) There is a large Stokes shift from the absorption edge, increasing with In composition. (3) The peak energy increases with

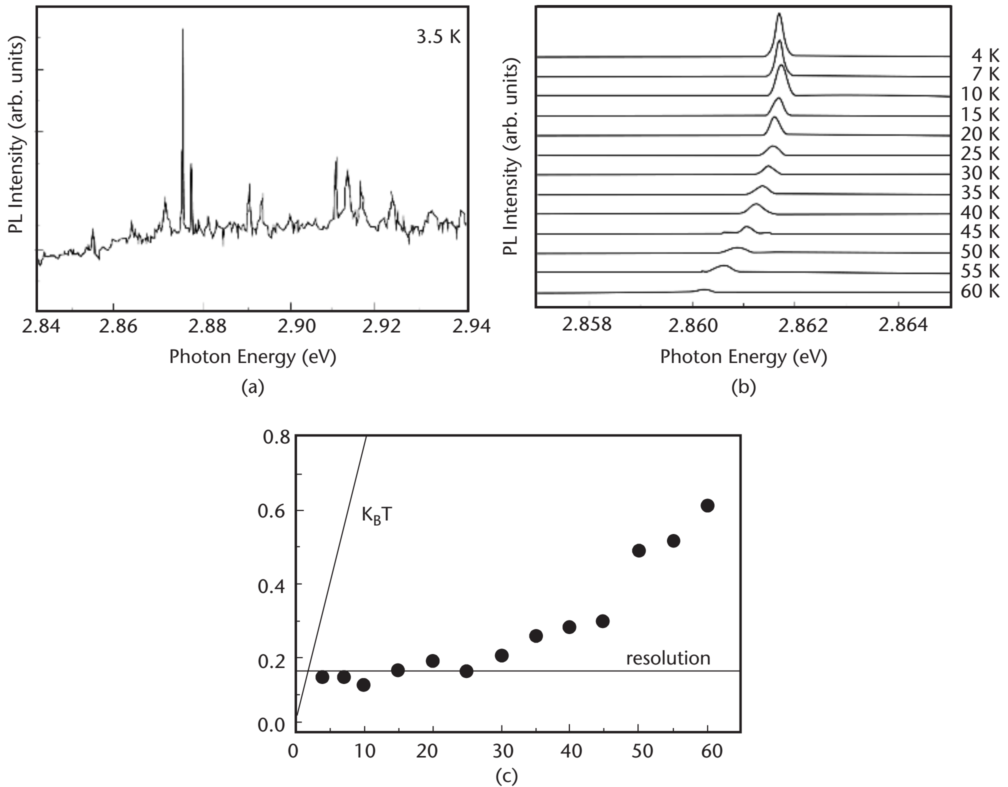


Figure 8.54 (a) High-resolution micro-PL spectrum from about 20 numbers of $In_xGa_{1-x}N$ quantum dots measured at 3.5K. (b) PL spectra of a particular peak taken at different temperatures from 4 to 60K. (c) PL linewidth (FWHM) as a function of temperature. Thermal energy is plotted as a dotted line. (*From:* [116]. © 2000 American Physics Society. Reprinted with permission.)

temperature at the low-temperature region in contrast to the change in the $In_xGa_{1-x}N$ bandgap. (4) The peak energy increases with the excitation density. (5) The PL exhibits different characteristics at high excitation levels. All of these observations can be interpreted by PL from localized states at band-tails, originating from small density of states and long lifetimes of the trapped carriers. These localized states are believed to be from In-rich or QD-like structures as demonstrated by HRTEM in some samples [125].

In contrast to the localization effect, Riblet et al. [124] argued that many of the above characteristics may result from the polarization effect in $In_xGa_{1-x}N$ QWs. They investigated the blue shift with excitation intensity for $In_xGa_{1-x}N$ QWs with different well widths but the same nominal In mole fraction of 0.25. They found that the maximum shift increases linearly with the well width, from 50 meV for a 1-nm well to 200 meV for a 5-nm well. A similar relation was also found for $GaN/Al_{0.15}Ga_{0.85}N$ QW, but with a shift that is an order of magnitude smaller (Figure 8.55). From these results, they estimated the electric field due to piezoelectricity to be 400 to 500 kV/cm in $In_{0.25}Ga_{0.75}N$ QWs and 60 to 70 kV/cm in GaN

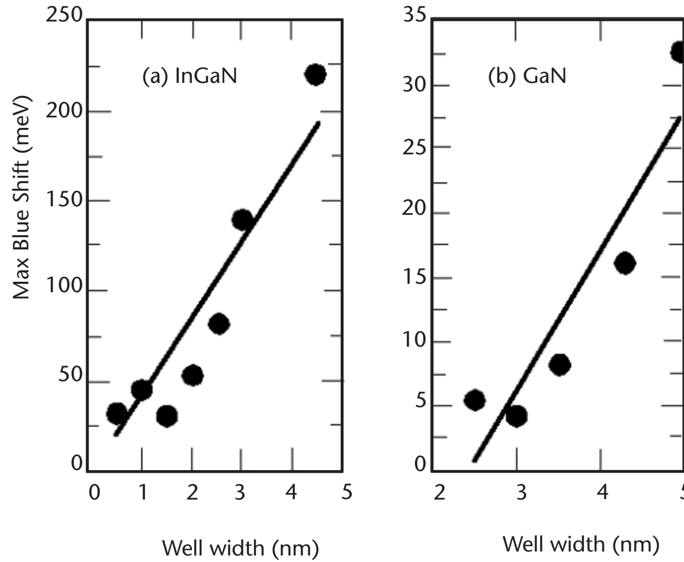


Figure 8.55 The maximum blue shift of the PL peak as a function of the well width for (a) $In_{0.25}Ga_{0.75}N$ double QW (with $In_{0.02}Ga_{0.98}N$ barriers) and (b) GaN single QW (with $Al_{0.15}Ga_{0.85}N$ barriers). The maximum blue shift is obtained for an optical excitation intensity of about 500 kW/cm^2. The solid lines are linear adjustments. (*From:* [124]. © 1999 American Institute of Physics. Reprinted with permission.)

QWs. While this interpretation may be reasonable, the possibility that the PL in $In_xGa_{1-x}N$ QWs is from In-rich dot-like structures is still not ruled out. The effect of the piezoelectricity only amplifies the energy fluctuation for a given dot-size or In mole-fraction distributions.

Surface-emitting lasers in $In_xGa_{1-x}N$/GaN/$Al_xGa_{1-x}N$ structures with $In_xGa_{1-x}N$ QDs were demonstrated recently [126]. Lasing in the vertical direction occurs at low temperatures, and an ultrahigh gain of 10^5 cm^{-1} in the active region was estimated. The lasing wavelength was 401 nm and the threshold current density was 400 kW/cm^2.

8.4 Summary

In the past 5 years, great progress has been witnessed in the fabrication and understanding of III-nitride QDs. The progress is mainly due to improvements in the growth techniques such as MBE and MOCVD. All three modes, 2D, 3D, and SK, were demonstrated in III-nitride epitaxial growth. Fully strained, defect-free, vertically correlated GaN QDs in AlN matrix were demonstrated mainly by MBE using the SK mode. Strain-free GaN QDs in $Al_xGa_{1-x}N$ were also realized using Si antisurfactant to modify the surface energy and change the growth mode from 2D to 3D. It has been demonstrated that the self-assembled GaN QD in AlN or $Al_xGa_{1-x}N$ matrix has a well-defined shape of truncated pyramid with a hexagonal base. The QD assembly with different densities and average sizes can be obtained in a more

controllable way. In addition to the GaN QDs, the $In_xGa_{1-x}N$ QDs can also be self-assembled on the GaN surface, although the shape, size, and density may be less controllable.

III-nitride QDs show rich and unique optical properties, some of which have been revealed in the last few years mainly by PL investigations. We know that the effects of quantum confinement, strain, and polarization are all important to III-nitride QDs and must be properly considered when interpreting any optical spectra. While the quantum confinement and strain in $GaN/Al_xGa_{1-x}N$ QDs commonly induce a blue shift in the transition energy from bulk bandgap, the spontaneous polarization and piezoelectricity may induce a red shift. The latter effects are more significant in the case of fully strained and large size GaN/AlN QDs and may pull the transition energy from a bulk GaN bandgap of 3.4 eV down to ~2 eV, covering the emission wavelength from blue to orange. A stable PL intensity against thermal perturbation up to room temperature was demonstrated in self-assembled GaN/AlN QDs. The effects of the quantum confinement on the exciton binding energy and electron-phonon coupling were also observed. The transition energy, the thermal stability, and the excitons and the electron-phonon coupling in QDs are all important to applications of III-nitrides in light-emitting and light-detecting devices.

Although great progress has been achieved, the III-nitride QDs grown by any methods today are still far from perfect. As we mentioned in the introduction, the fabrication of semiconductor QD assemblies with small and uniform size, high density, well-ordered placement, and defect-free materials remains a great challenges, especially for III-nitride materials.

References

[1] Nakamura, S., and G. Fasol, *The Blue Laser Diode*, Berlin, Germany: Springer-Verlag, 1997.

[2] Arakawa, Y., and H. Sakaki, *Appl. Phys. Lett.,* Vol. 40, 1982, p. 939.

[3] Morkoç, H., *Nitride Semiconductors and Devices,* Berlin, Germany: Springer-Verlag, 1999.

[4] Mohammad, S. N., and H. Morkoc, *Prog. Quantum Electron.,* Vol. 20, 1996, p. 361.

[5] Shchukin, V. A., and D. Bimberg, *Rev. Mod. Phys.,* Vol. 71, 1999, p. 1125.

[6] Eaglesham, D. J., and M. Cerullo, *Phys. Rev. Lett.,* Vol. 64, 1990, p. 1943.

[7] Petroff, P. M., and S. P. DenBaars, *Superlattices and Microstructures,* Vol. 15, 1994, p. 15.

[8] Pinczolits, M., G. Springholz, and G. Bauer, *Appl. Phys. Lett.,* Vol. 73, 1998, p. 250.

[9] Daudin, B., et al., *Phys. Rev. B,* Vol. 56, 1997, p. R7069.

[10] Daudin, B., et al., *MRS Symp. Proc.,* Vol. 482, 1998, p. 205.

[11] Widmann, F., et al., *J. Appl. Phys.*, Vol. 83, 1998, p. 7618.

[12] Widmann, F., et al., *Phys. Rev. B*, Vol. 58, 1998, p. R15989.

[13] Rouviere, J. L., et al., *Appl. Phys. Lett.*, Vol. 75, 1999, p. 2632.

[14] Daudin, B., et al., *MRS Internet J. Nitride Semicond. Res.*, Vol. 4S1, 1999, p. G9.2.

[15] Damilano, B., et al., *Appl. Phys. Lett.*, Vol. 75, 1999, p. 962.

[16] Shen, X.-Q., et al., *Appl. Phys. Lett.*, Vol. 72, 1998, p. 344

[17] Hirayama, H., Y. Aoyagi, and S. Tanaka, *MRS Internet J. Nitride Semicond. Res.*, Vol. 431, 1999, p. G9.4.

[18] Shen, X.-Q., et al., *Jpn. J. Appl. Phys.*, Vol. 39, 2000, p. L16.
[19] Huang, D., et al., *Appl. Phys. Lett.*, Vol. 78, 2001, p. 4145.
[20] Feuillet, G., et al., *Mater. Sci. Eng.* B, Vol. 50, 1997, p. 233.
[21] Xie, Q., et al., *Phys. Rev. Lett.*, Vol. 75, 1995, p. 2542.
[22] Tersoff, J., C. Teichert, and M. G. Lagally, *Phys. Rev. Lett.*, Vol. 76, 1996, p. 1675.
[23] Brault, J., et al., *J. Appl. Phys.*, Vol. 93, 2003, p. 3108.
[24] Tersoff, J., et al., *Phys. Rev. Lett.*, Vol. 75, 1995, p. 2730
[25] Liu, F., J. Tersoff, and M. G. Lagally, *Phys. Rev. Lett.*, Vol. 80, 1998, p. 1268.
[26] Leon, R., et al., *J. Appl. Phys.*, Vol. 84, 1998, p. 248.
[27] Lacombe, D., et al., *Appl. Phys. Lett.*, Vol. 74, 1999, p. 1680.
[28] Adelmann, C., et al., *Phys. Stat. Sol. B*, Vol. 234, 2002, p. 931.
[29] Northrup, J. E., et al., *Phys. Rev. B*, Vol. 61, 2000, p. 9932.
[30] Kern, R., and P. Muller, *Surf. Sci.*, Vol. 392, 1997, p. 103.
[31] Adelmann, C., et al., *Appl. Phys. Lett.*, Vol. 81, 2002, p. 3064.
[32] Arlery, M., et al., *Appl. Phys. Lett.*, Vol. 74, 1999, p. 3287.
[33] Mula, G., et al., *Phys. Rev. Lett.*, Vol. 69, 1992, p. 796.
[34] Chamard, V., et al., *Mater. Sci. Eng. B*, Vol. 93, 2002, pp. 24–27.
[35] Gleize, J., et al., *Appl. Phys. Lett.*, Vol. 77, 2000, p. 2174.
[36] Martinez-Guerrero, E., et al., *Appl. Phys. Lett.*, Vol. 77, 2000, p. 809.
[37] Mukai, T., M. Yamada, and S. Nakamura, *Jpn. J. Appl. Phys.*, Vol. 38, 1999, p. 3976.
[38] Nakamura, S., *Semicond. Sci. Technol.*, Vol. 14, 1999, p. R27.
[39] Grandjean, N., et al., *Appl. Phys. Lett.*, Vol. 74, 1999, p. 3616.
[40] O'Donnell, K. P., R. W. Martin, and P. G. Middleton, *Phys. Rev. Lett.*, Vol. 82, 1999, p. 237.
[41] Adelmann, C., et al., *Appl. Phys. Lett.,* Vol. 76, 2000, p. 1570.
[42] Damilano, B., et al., *Appl. Phys. Lett.*, Vol. 75, 1999, p. 3751.
[43] Grandjean, N., and J. Massies, *Appl. Phys. Lett.*, Vol. 72, 1998, p. 1078.
[44] Dmitriev, V., et al., in *Gallium Nitride and Related Materials*, R. D. Dupuis et al., (eds.), Pittsburgh, PA: Materials Research Society, 1996
[45] Tanaka, S., S. Iwai, and Y. Aoyagi, *Appl. Phys. Lett.*, Vol. 69, 1996, p. 4096.
[46] Miyamura, M., K. Tachibana, and Y. Arakawa, *Appl. Phys. Lett.*, Vol. 80, 2002, p. 3937.
[47] Leonard, D., K. Pond, and P. M. Petroff, *Phys. Rev. B*, Vol. 50, 1994, p. 11687.
[48] Tachibana, K., et al., *Phys. Stat. Sol. B.*, Vol. 228, No. 1, 2001, pp. 187–190.
[49] Tanaka, S., S. Iwai, and Y. Aoyagi, *Appl. Phys. Lett.*, Vol. 69, 1996, p. 4096.
[50] Tanaka, S., et al., *MRS Symp. Proc.*, Vol. 449, 1997, p. 135.
[51] Ramvall, P., et al., *Appl. Phys. Lett.,* Vol. 75, 1999, p. 1935.
[52] Ramvall, P., et al., *J. Appl. Phys.*, Vol. 87, 2000, p. 3883.
[53] Hirayama, H., S. Tanaka, and Y. Aoyagi, *Microelectronic Eng.*, Vol. 49, 1999, pp. 287–290.
[54] Zhang, J., et al., *Appl. Phys. Lett.*, Vol. 80, 2002, p. 485.
[55] Tachibana, K., T. Someya, and Y. Arakawa, *Appl. Phys. Lett.*, Vol. 74, 1999, p. 383.
[56] Leonard, D., K. Pond, and P. M. Petroff, *Phys. Rev. B*, Vol. 50, 1994, p. 11687.
[57] Chen, P., et al., *J. Vacuum Sci. Technol. B*, Vol. 12, 1994, p. 2568.
[58] Ji, L. W., et al., *J. Crystal Growth*, Vol. 249, 2003, pp. 144–148.
[59] Chen, Z., et al., *J. Crystal Growth*, Vol. 235, 2002, pp. 188–194.
[60] Widmann, F., et al., *J. Appl. Phys.*, Vol. 83, 1998, p. 7618.

[61] Cheng, W.-Q., et al., *J. Crystal Growth*, 1998, p. 705.
[61] Kapolnek, D., et al., *J. Crystal Growth*, Vol. 189/190, 1998, p. 83.
[62] Tachibana, K., et al., *Appl. Phys. Lett.*, Vol. 76, 2000, p. 3212.
[63] Wagner, R. S., in *Whisker Technology*, A. P. Leavitt, (ed.), New York: John Wiley & Sons, 1970, Chap. 3.
[65] Chen, C.-C., et al., *J. Am. Chem. Soc.*, Vol. 123, 2001, p. 2791.
[66] Kawasaki, K., et al., *Appl. Phys. Lett.*, Vol. 79, 2001, p. 2243.
[67] Hu, C.-W., et al., *Appl. Phys. Lett.*, Vol. 81, 2002, p. 3236.
[68] Zheng, L. X., M. H. Xie, and S. Y. Tong, *Phys. Rev. B*, Vol. 61, 2000, p. 4890.
[69] Goodwin, T. J., et al., *Appl. Phys. Lett.*, Vol. 70, 1997, p. 3122.
[70] Preschilla, N., et al., *Appl. Phys. Lett.,* Vol. 77, 2000, p. 1861.
[71] Janic, J. F., and R. L. Well, *Chem. Mater.*, Vol. 8, 1996, p. 2708.
[72] Coffer, J. L., et al., *Chem. Mater.,* Vol. 9, 1991, p. 2671.
[73] Micic, O. I., et al., *Appl. Phys. Lett.*, Vol. 75, 1999, p. 478.
[74] Benaissa, M., K. E. Gonsalves, and S. P. Rangarajan, *Appl. Phys. Lett.*, Vol. 71, 1997, p. 3685.
[75] Reshchikov, M. A., et al., *Mat. Res. Soc. Symp.*, Vol. 622, 2000, p. T4.5.
[76] Morkoç, H., et al., *Mat. Res. Soc. Symp.*, Vol. 639, 2001, p. G11.2.
[77] Bernardini, F., V. Fiorentini, and D. Vanderbilt, *Phys. Rev. B*, Vol. 56, 1997, p. R10024.
[78] Andreev, A. D., and E. P. O'Reilly, *Physica E*, Vol. 10, 2001, p. 553.
[79] Das, D., and A. C. Melissinos, *Quantum Mechanics*, New York: Gordon and Breach, 1986.
[80] Martin, G., et al., *Appl. Phys. Lett.*, Vol. 68, 1996, p. 2541.
[81] Davydov, V. Y., et al., *J. Appl. Phys.*, Vol. 82, 1997, p. 5097.
[82] Shikanai, A., et al., *J. Appl. Phys.*, Vol. 81, 1997, p. 417.
[83] Amano, H., K. Hiramatsu, and I. Akasaki, *Jpn. J. Appl. Phys.*, Vol. 227, 1998, p. L1384.
[84] Shan, W., et al., *Phys. Rev. B*, Vol. 54, 1996, p. 13460.
[85] Shimada, K., T. Sota, and K. Suzuki, *J. Appl. Phys.*, Vol. 84, 1998, p. 4951.
[86] Strite, S., and H. Morkoç, *J. Vacuum Sci. Technol. B*, Vol. 10, 1992, p. 1237.
[87] Martinez, O., et al., *Phys. Stat. Sol. A*, Vol. 195, 2003, pp. 26–31.
[88] Ramvall, P., et al., *J. Appl. Phys.*, Vol. 87, 2000, p. 3883.
[89] Nye, J. F., *Physical Properties of Crystals*, Oxford, England: Oxford University Press, 1985.
[90] Bernardini, F., V. Fiorentini, and D. Vanderbilt, *Phys. Rev. B*, Vol. 56, 1997, p. R10024.
[91] Mendez, E. E., et al., *Phys. Rev. B,* Vol. 26, 1982, p. 7101.
[92] Miller, D. A. B., et al., *Phys. Rev. B*, Vol. 32, 1985, p. 1043.
[93] Goodwin, T. J., et al., *Appl. Phys. Lett.*, Vol. 70, 1977, p. 3122.
[94] Preschilla, N., et al., *Appl. Phys. Lett.*, Vol. 77, 2000, p. 1861.
[95] Micic, O. I., et al., *Appl Phys. Lett.*, Vol. 75, 1999, p. 478.
[96] Andreev, A. D., and E. P. O'Reilly, *Phys. Rev. B*, Vol. 62, 2000, p. 15851.
[97] Ramvall, P., et al., *Appl. Phys. Lett.*, Vol. 73, 1998, p. 1104.
[98] Petersson, A., et al., *Appl. Phys. Lett.*, Vol. 74, 1999, p. 3513.
[99] Tanaka, S., et al., *Appl. Phys. Lett.*, Vol. 71, 1997, p. 1299.
[100] Moumanis, K., et al., *Appl. Phys. Lett.*, Vol. 82, 2003, pp. 868–870.
[101] Cho, Y.-H., et al., *Appl. Phys. Lett.*, Vol. 81, 2002, p. 4934.
[102] Leroux, M., et al., *Phys. Rev. B*, Vol. 60, 1999, p. 1496.
[103] Cho, Y.-H., et al., *Phys. Rev. B*, Vol. 61, 2002, p. 7203.
[104] Chichibu, S. F., et al., *Appl. Phys. Lett.*, Vol. 73, 1998, p. 2006.

[105] Miller, D. A. B., D. S. Chemla, and S. Schmitt-Rink, *Phys. Rev. B*, Vol. 33, 1986, p. 6976.
[106] Martinez-Guerrero, E., *Appl. Phys. Lett.*, Vol. 77, 2000, p. 809.
[107] Shan, W., et al., *Phys. Rev. B*, Vol. 54, 1996, p. 16369.
[108] Reynolds, D. C., et al., *J. Appl. Phys.*, Vol. 80, 1996, p. 594.
[109] De Rinaldis, S., I. D'Amico, and F. Rossi, *Appl. Phys. Lett.*, Vol. 81, 2002, p. 4236.
[110] Arakawa, Y., T. Someya and K. Tachibana, *Phys. Stat. Sol. B*, Vol. 224, 2001, p. 1.
[111] Simon, J., et al., *Phys. Stat. Sol. B*, Vol. 224, 2001, p. 13.
[112] Neogi, A., et al., *IEEE Trans. on Nanotechnology*, Vol. 2, No. 1, 2003, pp. 10–14.
[113] Brown, J., et al., *Phys. Stat. Sol. B*, Vol. 228, 2001, p. 199.
[114] Widmann, F., et al., *Phys. Rev. B*, Vol. 58, 1998, p. R15989.
[115] O'Donnell, K. P., R.W. Martin and P.G. Middleton, *Phys. Rev. Lett.*, Vol. 82, 1999, p. 237.
[116] Moriwaki, O., et al., *Appl. Phys. Lett.*, Vol. 76, 2000, p. 2361.
[117] Shi, J.-J., *Solid State Commun.*, Vol. 124, 2002, p. 341.
[118] Chichibu, S., et al., *Appl. Phys. Lett.*, Vol. 69, 1996, p. 4188.
[119] Perlin, P., et al., *Phys. Rev. B*, Vol. 64, 2001, p. 115319.
[120] Damilano, B., et al., *Appl. Phys. Lett.*, Vol. 75, 1999, p. 3751.
[121] Mendez, E. E., et al., *Phys. Rev. B*, Vol. 26, 1982, p. 7101.
[122] Nakamura, S., *Appl. Phys. Lett.*, Vol. 70, 1997, p. 2753.
[123] Pophristic, M., et al., *J. Appl. Phys.*, Vol. 86, 1999, p. 1114.
[124] Riblet, P., et al., *Appl. Phys. Lett.*, Vol. 75, 1999, p. 2241.
[125] Nistor, L., et al., *Appl. Phys. Lett.*, Vol. 77, 2000, p. 507.
[126] Krestnikov, I. L., et al., *Semiconductors*, Vol. 34, 2000, p. 481.

CHAPTER 9

Self-Assembled Germanium Nano-Islands on Silicon and Potential Applications

K. L. Wang, S. Tong, and A. Khitun, Department of Electrical Engineering, University of California at Los Angeles; J. L. Liu, Department of Electrical Engineering, University of California at Riverside; and G. Jin, Integrated Device Technology

9.1 Introduction

Nanoscience and nanotechnology have flourished in recent years because it is believed they will lead to a scientific and technical revolution in the not too distant future. One of the most interesting topics of focus is self-assembled nanostructures, which are formed spontaneously due to the nature of their growth. In current advanced top-down lithography regimes, it is not difficult at all to fabricate the structures on a nanometer scale. However, the self-assembled (so-called bottom-up approach) nanostructures have some merits, such as a spontaneous nature and damage-free surfaces, that are attractive to researchers. In this chapter, we focus on the discussion of self-assembled Ge islands formation on Si substrates and discuss their potential applications.

9.2 Heteroepitaxy Mechanisms

It is well known that there are three types of growth modes for heterogeneous epitaxial growth including the Ge/Si system: the Frank-Van der Merwe [1], Stranski-Krastanov [2], and Volmer-Weber [3] modes, which are named after the original researchers. The typical layer-by-layer growth is referred to as *Frank-Van der Merwe growth,* which is required to fabricate high-quality large films for practical applications. For some heterogeneous epitaxial growth conditions, where there is, for example, a substantial lattice mismatch, layer-by-layer growth is impossible and only 3D islands are formed. This growth is referred to as *Volmer-Weber growth*. *Stranski-Krastanov growth* is an intermediate growth mode, where growth starts in a layer-by-layer mode during the first few atomic layers, and then forms 3D islands beyond a certain thickness, which is usually referred to as *critical thickness*.

Based on thermodynamic analysis, the growth mode in a given system is determined from the surface/interface energy and the lattice mismatch [4]. If $\gamma + \sigma_e > \sigma_s$, a Frank-Van der Merwe growth mode is expected, where γ is interface energy, σ_e is

epilayer surface energy, and σ_s is substrate surface energy. Island structures form under the condition of $\gamma + \sigma_e < \sigma_s$, which belongs to a Volmer-Weber growth mode. For a strained epilayer, there is another possibility that island formation allows the system to relax the strain by increasing the surface energy via the formation of island facets. For a system with a small interface energy but large lattice mismatch, the initial growth begins in a layer-by-layer fashion, but as the layer thickness increases, it gains a larger strain energy. As a result, its total energy is lowered by forming islands and, thus, Stranski-Krastanov growth occurs in strained systems, such as Ge/Si (001) [5].

The growth of Ge on Si has been studied for several decades. It is known that the growth of Ge on Si occurs in the Stranski-Krastanov growth mode, where the growth begins initially in a layer-by-layer manner and then forms 3D islands. The formation of Ge islands is driven by the surface strain resulting from the lattice mismatch, where the Ge lattice constant is larger than that of Si by 4.16%. The growth of Ge on Si(001) has been extensively studied with a variety of surface analysis techniques [6–10] such as AES and RHEED. However, in early studies, there was some disagreement regarding the value of the critical thickness, that is, at which point the dots start to appear. The value ranged from 3 to 10 ML before the onset of appearance of islands. Recent studies have concluded that the critical thickness of Ge on Si (001) is 3 ML [4]. The actual number of monolayers may vary slightly with the growth temperature or growth techniques, for example, the value is slightly larger for gas source MBE growth [11].

9.3 Uniform Ge Islands

As normally seen in the Ge/Si(001) system, the formation of Ge islands during the growth on Si(001) [12] is a spontaneous nucleation process. The two major island shapes are (1) square-based pyramids at a smaller volume and (2) multifaceted domes at a larger volume. The four {105} facets dominate in a pyramid island even though other facets are reported. The facets in a dome island are more complicated. These two distinct islands are referred to as *bimodal* growth.

Figure 9.1(a) presents a typical AFM image of self-assembled Ge islands grown on planar Si(001) substrates. One can clearly see that there are two kinds of Ge islands. The smaller islands are square-based pyramids with four {105} facets, and the larger islands are multifaceted domes, as shown in Figure 9.1(b). They have distinct shapes and distinguishable size distributions as discussed in previous publications [13], which are referred to as a *bimodal distribution* [14]. The bimodal distribution of Ge islands is a typical experimental result for Ge grown on Si(001). However, other kinds of islands were also observed. In one case, with relatively thick Ge coverage, superdomes were observed [15].

Figure 9.2 shows the statistical results of the Ge deposition on Si [15]. It is clear that different Ge coverage results in different island morphology. In another case, for prolonged growth at high temperature, such as at 700°C, larger pyramidal islands with a flat top (or truncated pyramids) have been observed. Our studies show that a larger Ge deposition rate can also generate supersized islands. Figure 9.3 shows an AFM image of a sample grown at the Ge deposition rate of 0.5

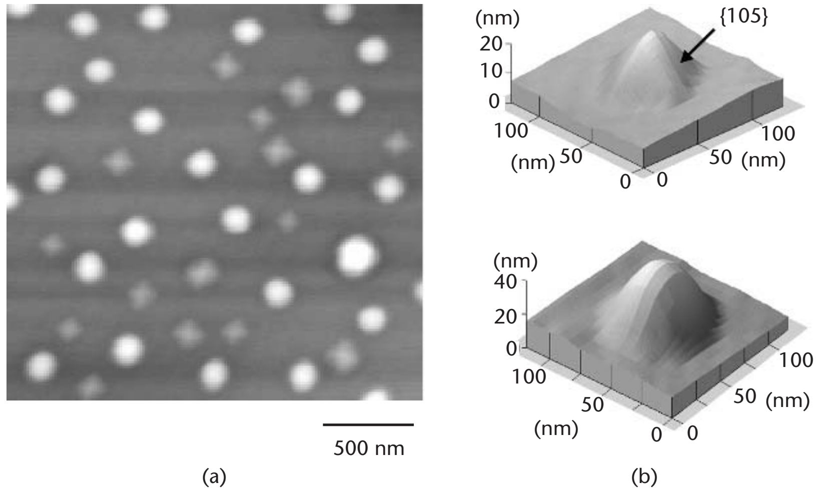


Figure 9.1 (a) 2D AFM image of typical bimodal Ge islands including square-based pyramids and multifaceted domes. (b) 3D AFM images showing a pyramid (top) and a dome (bottom).

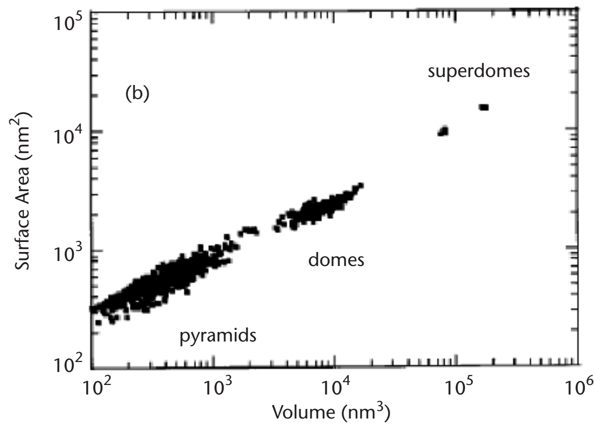


Figure 9.2 Scatterplot showing the exposed surface area of the islands versus their volumes. (*After:* [15].)

Å/sec. Ge thickness was 1.5 nm. The growth temperature was 540°C and the postholding time was 2 minutes. Besides the round-shaped domes and the square-based pyramids, superdomes are evident. These superdomes are in reality the combination of several smaller islands. The formation of the superdomes suggests that the diffusion of Ge atoms on the surface to other sites is a kinetically limited process. It was found that for a Ge thickness of less than 2 nm and a Ge rate of less than 0.3 Å/sec, the growth led to no superdomes.

The size distribution of the islands is referred as a *trimodal* (or *multimodal*) distribution [16]. From cross-sectional TEM studies [4], the pyramidal and dome Ge islands with heights of less than 50 nm [16] are free of dislocations, and are called

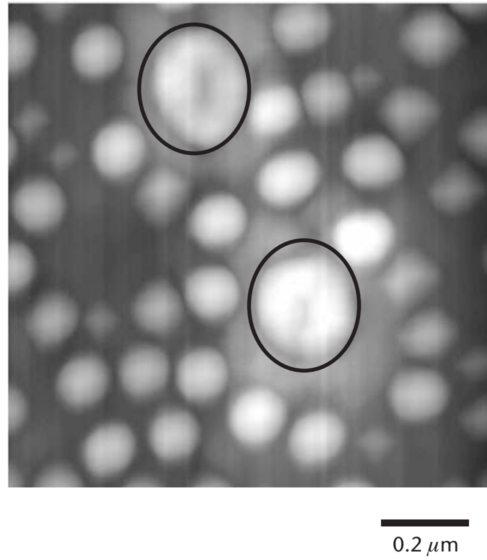


Figure 9.3 AFM image of a Ge QD sample grown at the Ge coverage of 15Å and the Ge deposition rate of 0.5 Å/sec. The growth temperature is 650°C. Besides pyramids and domes, superdomes are evident.

coherent, dislocation-free, self-assembled islands. Island formation occurs as a result of the minimization of the total free energy by (partially) relieving the strain energy via elastic deformation of the substrate and the formation of the facets in the islands.

It is generally believed that the distribution of Ge islands has something to do with Ostwald's ripening, which is a coarsening process of a particle size driven by the Gibbs-Thomson effect. There are several theories related to the coarsening process. An analytical mean-field theory (LSW theory) was first used to analyze particle evolution in a solution [17, 18]. The theory was extended to solid systems for studying the particle size evolution on single crystal surfaces [19]. The equation of motion for the basic particle radius evolution can be expressed by

$$\frac{dr}{dt} = \frac{\beta(T)}{r^m}\left(\frac{1}{r_c} - \frac{1}{r}\right) \tag{9.1}$$

where r_c is the radius of a stable particle and m is a constant; $m = 1$ and 2 for the LSW and Chakraverty models, respectively. If r is larger than r_c, the particle size increases; otherwise, the size decreases.

Based on kinetic and thermodynamic analyses, there are two basic interpretations for the experimental observations. In their real-time low-energy electron microscopy studies, Ross et al. [20] observed that larger islands grew while smaller islands shrank and disappeared with increasing Ge deposition, instead of uniform

growth occurring for all sizes of islands. They concluded that it is a kinetic process driven by an Ostwald ripening effect. A theoretical justification based on a kinetic mechanism was also made to support the explanation. Based on thermodynamic analysis, Shchukin et al. [21] proposed that for some systems, there are minima in the free energy of self-assembled island ensembles, resulting in a stable size distribution at some particular island sizes instead of suffering Ostwald ripening.

A convincing experimental observation is the scanning tunneling microscopy measurements of the island size distribution [22], in agreement with the Shchukin model, that show sizes with a stable Boltzmann distribution. Kamins et al. [14] has quantitatively studied the formation of the bimodal distribution of Ge islands and found that the island distribution is similar to a Boltzmann distribution, a steady-state distribution. Their annealing experiments [13] on the same sample distinguished the two mechanisms, Ostwald ripening and equilibrium distributions. They found that island evolution was slowed with increasing annealing time at 550°C and 600°C and it reached an equilibrium distribution. Given the preceding analysis, some theoretical research has predicted [23, 24] that it is possible to form uniform, ordered self-assembled islands.

To form uniform islands on Si, we have studied the growth temperature influence on the formation of self-assembled Ge islands on Si(001) substrates. Figure 9.4 shows the AFM images of the self-assembled Ge islands at various growth temperatures from 500°C to 700°C. As generally observed, the island density decreases with increasing growth temperature. In Figure 9.4(b) with the growth temperature at 550°C, one can clearly see that there are two kinds of Ge islands: pyramids and domes. This is typical of the frequently observed bimodal distribution. At a lower temperature of 500°C [Figure 9.4(a)], the pyramid type dominates.

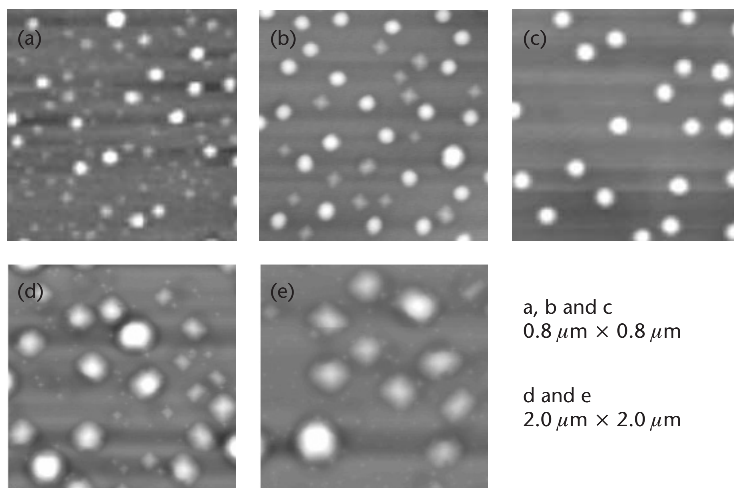


Figure 9.4 Two-dimensional AFM images of the dots on planar Si(001) substrates with various growth temperatures: (a) T_s = 500°C, (b) T_s = 550°C, (c) T_s = 600°C, (d) T_s = 650°C, and (e) T_s = 700°C. The thickness of the deposited Ge layer is 1.6 nm.

As the growth temperature increases, the ratio of pyramids to domes decreases as plotted in Figure 9.5 and finally approaches to very low value at 600°C, leading to the formation of a near *monomodal* morphology distribution. As the temperature further increases to 650°C and 700°C [Figure 9.4 (d, e), respectively], the pyramids reappear at a larger size, and there are also some very small islands that have heights in the range of 2 to 3 nm. The very small islands were generally thought of as the precursors of the larger islands. The difference is that the precursors were only observed at higher temperatures in our case. The appearance of the small precursors might be related to the change of the surface due to, for example, the intermixing, which we discuss later. Besides the larger size of pyramids and domes, note that the morphology of the islands is no longer isotropic. This was attributed to the anisotropic interdiffusion of Si into Ge islands.

Based on the V-shape distribution of Figure 9.5, one can see that the optimal growth temperature needed to form uniform Ge islands is 600°C. As shown in Figure 9.4(c), highly uniform self-assembled Ge islands were realized on Si(001) substrate at this growth temperature. The nominal Ge deposition thickness was 1.6 nm and the growth was interrupted to allow for diffusion to reach near the steady state. The islands are dome shaped with a base diameter and height of about 70 and 15 nm, respectively. The areal density of the islands is about 3×10^9 cm^{-2} and the height deviation of the islands is about ±3%. This result has experimentally demonstrated the possibility of forming uniform self-assembled Ge islands on Si(001) and an equilibrium state might occur at temperatures between 550°C and 600°C [14].

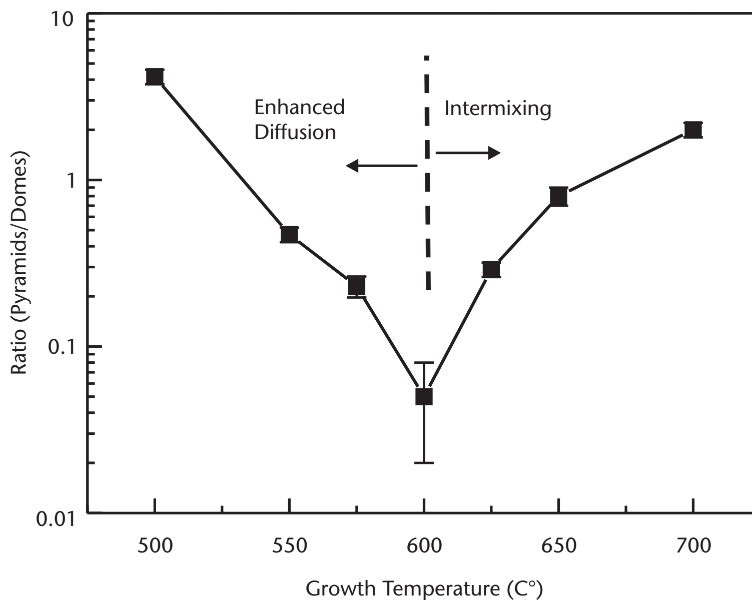


Figure 9.5 The number ratio of pyramids to domes versus growth temperature. The optimum temperature is about 600°C for the formation of monomodal dots. Note that the small precursors at high temperature are not included.

9.4 Registration and Regimentation of Ge Islands

In the previous section, we discussed how to form uniform Ge islands, which is important for optical and optoelectronic applications. However, the islands were randomly distributed on the Si surface. In this session, we discuss the controlled growth of islands so that the islands can be arranged spatially at specific locations and so that we can control the positioning of a single island, which is desirable for nanoelectronic applications. The arrangement of self-assembled islands at predetermined sites remains a major challenge for the implementation of nanoelectronics and quantum computation [25].

Previously, much effort has been devoted to controlling the spatial locations of self-assembled dots using a variety of techniques, including growth on miscut (or tilted) substrates with surface steps [26] and on relaxed templates having a dislocation network [27, 28], and stacking growth of multilayers of self-assembled islands [29–31]. Among them, one of the most effective approaches is the use of selectively grown mesas as templates for subsequent Ge growth. This approach shows 1D ordering of Ge islands along the edges of the Si stripe mesas, formed in patterned windows with large feature sizes prepared by conventional lithography [32, 33].

In the following, we discuss an effective method for ordering Ge QD arrays on prepatterned substrates. We first describe the formation of substrate templates by using the *selective epitaxial growth* (SEG) technique. SEG is based on the different interactions of the gas molecules (such as Si_2H_6) with a clean, chemically reactive Si surface and a chemically inert SiO_2 surface. As gas molecules impinge on the surface, only a fraction of the molecules will eventually adsorb on the surface. The sticking coefficient of Si_2H_6 on a clean Si surface is about 0.1 at room temperature [34]; however, the coefficient of Si_2H_6 on a clean SiO_2 surface is expected to be near zero at temperatures below 700°C. As a result, the growth of the Si epilayer only takes place on the Si surface. In reality, the selectivity is relative and the sticking coefficient of Si_2H_6 on a clean SiO_2 surface depends on growth temperature. Therefore, the selectivity of SEG depends on the growth conditions including the growth temperature and the gas flux. On patterned substrates, due to the approximately zero sticking coefficient on the silicon dioxide sidewalls at proper temperatures, the growth occurs at the exposed Si windows. Formation of facets at the sidewalls can be observed due to the anisotropy of the growth rates, which arises from different surface dangling bonds and surface energy. From an energetic point of view, the surface energy of Si(001) is the lowest, then the Si(113) surface, and the (111) surface is the highest of them all. Therefore, the corresponding growth rates on the three different surfaces reduce in the order of (001), (113), and (111). Hence, the {113} facets take place at the early stage of the SEG and the top (001) surface reduces accordingly [35, 36]. Then the {111} facets occur after a certain thickness, which depends on the growth conditions (e.g., substrate temperature). Therefore, the {113} and {111} facets coexist at this growth stage. The angle between (113) and (001) is about 25°, and the angle between (111) and (001) is 54°. Due to the energetic preferences discussed earlier, the {111} facets will dominate the sidewalls as the growth continues, and the top (001) surface reduces further. For Si mesas in the windows having

small feature size, the top (001) surface may be fully reduced after continuous SEG growth of Si. In this situation, the sidewall facets may consist of {113} or {111} or both of them, depending on the growth conditions and window size.

After forming the selectively grown Si mesas, Ge was subsequently deposited, and then Ge islands were preferentially formed along the edges of Si mesas. Two factors explain the preferential growth. The first factor is the morphology effect. Convex curvature occurred at the edges of Si mesas. This convex curvature and supersaturation at edges may be beneficial to the formation of Ge islands. The other factor is due to the fact that the subsurface Si lattice at this point is slightly higher than bulk Si and thus has a smaller lattice mismatch with Ge, and hence the edge is a preferential nucleation site. This strain relief also allows the anisotropic deposition to proceed more readily near the edges.

We have studied the strain distribution on the SEG mesas with micro-Raman measurements to distinguish the effect of the strain and the morphology on this preferential positioning. The strain was determined by the shift of the Raman-Si transverse optical-phonon peak. It was found that nonuniform strain distribution occurs on the mesas. A confocal micro-Raman imaging technique was used to visualize the strain distribution of the SEG Si facets. The Raman image from an unstrained (standard) Raman peak reveals the facet shape of the top plateau of a SEG mesa although the image tends to exaggerate the actual size of the faceted island due to the resolution limitation. It was found later that the Ge islands are preferentially grown on the most tensile strained part (lower in surface energy) of the mesa top. The results indicated that there was tensile strain near the edges and compressive strain in the central region [37].

With the SEG process, it is possible to form 1D ridges on a stripe mesa structure via complete reduction of the top plateau. After the formation of the ridges, Ge is subsequently deposited on the SEG mesa ridges. The growth of Ge occurs via the Stranski-Krastanov growth mode. Figure 9.6 shows a 3D AFM image of the self-assembled Ge islands on the <110>-oriented Si mesa stripe, formed in the exposed Si stripe windows with a window width of 0.6 μm and a separation between two stripes of about 0.1 μm. It should be pointed out that the silicon dioxide layer had been removed for the AFM measurement.

One can see the morphology of the Si mesas formed by the SEG process. The epi thickness is 120 nm and the sidewall facet angles are about 25°, corresponding to the {113} facets. It is also found that the top (001) surface has been fully reduced. Perfectly aligned and regularly spaced 1D arrays of Ge islands are formed on the ridges of the Si stripe. In short, the perfect alignment of the islands along the Si stripe mesas is due to preferential positioning of the Ge islands on the 1D ridges, which is the result of the full reduction of the top surface of the stripe mesas. This arrangement of self-assembled Ge islands is referred to as a *cooperative arrangement of self-assembled islands* (CASAD) since they are arranged on the 1D ridges through cooperative strain interactions among dots.

It is interesting that only islands with a monomodal distribution, which means that all the islands are dome shaped and have a size of 70 to 90 nm, were observed on the ridges of Si stripe mesas over a large patterned region. A similar result was

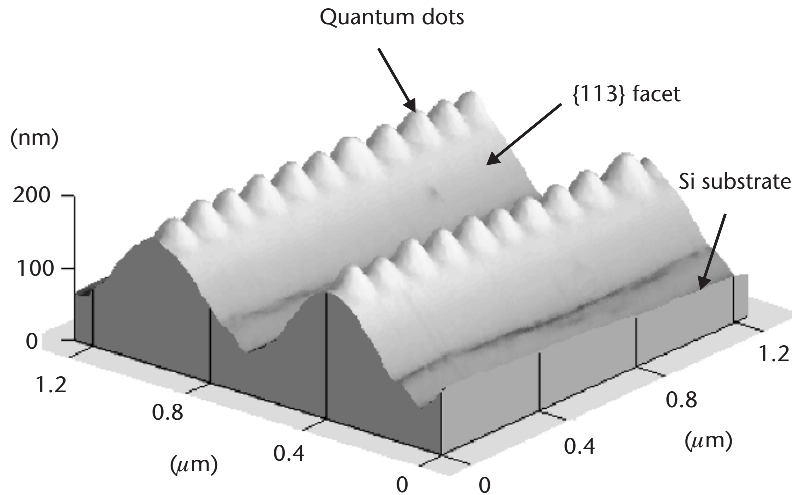


Figure 9.6 3D AFM image of the cooperative arrangement of self-assembled Ge dots on the n-oriented Si stripe mesas with a window width of 0.6 m. Self-aligned and well-spaced 1D arrays of the Ge dots are formed on the ridges of the Si mesas after the deposition of 10 ML of Ge. The growth temperature is 630°C. The sidewall facets of the Si stripe mesas are confirmed to be {113} facets. The dimensions of the Ge islands are about 80 nm wide and about 20 nm high, and the period of Ge islands is about 110 nm.

reported on high-index facets of SEG mesas [38], in contrast with the results usually obtained on planar Si(001) substrates, where a bimodal or multimodal distribution of Ge islands was present. The monomodal distribution of islands may be the result of a strain effect and quasi-1D spatial confinement. As seen in the facet formation in an SEG process, mass transfer from sidewalls to the top surface has been observed due to the anisotropic growth rates on different surfaces [34]. Our micro-Raman results indicate tensile stress at the edges of the SEG Si mesas, which corresponds to the energetically favorable nucleation sites. From an energetic viewpoint, the adatom diffusion along the 1D ridge is limited due to the higher energy barrier required to pass over the islands. Thus, it is a quasi-1D case. This is different from the case in which Ge adatom diffusion on the surface can be in a 2D plane and, thus, the Ge islands are randomly distributed on a plane. Therefore, the spatial confinement of the Si stripe mesas affects the diffusion of Ge adatoms from both directions of the two sidewalls, leading to more uniform islands: a monomodal distribution.

The arrangement of the Ge islands depends on the geometry of the Si mesas. The Ge islands are only formed near the <110>-oriented edges and the central region is free of Ge islands. This is due to the preferential nucleation of Ge islands at the edges as we discussed previously and the diffusion length is sufficiently long compared with the window feature size to guarantee the Ge adatom diffusion to the preferential sites. It is found that at the growth temperature of 600°C, the diffusion length of Ge adatoms on Si mesas is about 3 μm. This value is reasonable when compared with the reported data in the literature [38], where the diffusion length of Ge on Si(001) is about 30 μm at a temperature of 700°C.

Growth temperature is another important factor in MBE growth. To understand the influence of growth temperature, we have also investigated the substrate temperature dependence on the 2D arrangement of Ge dots. Figure 9.7 shows the AFM images of samples with various growth temperatures. At the low temperature of 500°C, as shown in Figure 9.7(a), the dots are small and dense, and seem to distribute everywhere. However, one can see that the dots prefer to locate near the edges of the Si mesas, due to the preferential nucleation near the edges. Additionally, a few dots are located in the central region of the Si mesas, which is attributed to the short diffusion length at a temperature of 500°C. As the temperature increases, the dot size increases and the corresponding dot density decreases. This size evolution is consistent with that of Ge dots on planar Si(001) substrate. At a temperature of 600°C [Figure 9.7(c)], Ge dots are only located near the edges, and the central regions are free of Ge dots. This indicates that the diffusion length is long enough to allow Ge adatoms to migrate to the preferential sites. When the temperature is further increased to 700°C [Figure 9.7(e)], only one Ge dot is seen on each Si mesa. This is possibly due to the enhanced mass transfer from the sidewall during Si selective epitaxial growth at high temperature, leading to the shrinkage of the top facets of the Si mesas. Meanwhile, the dot size becomes larger with the increase in temperature. As a result, on the small top (001) facet, only one Ge dot is formed.

To gain insight into the evolution of ordered dot arrays, we have performed experiments to study the dependence of the Ge amount deposited. Figure 9.8 shows

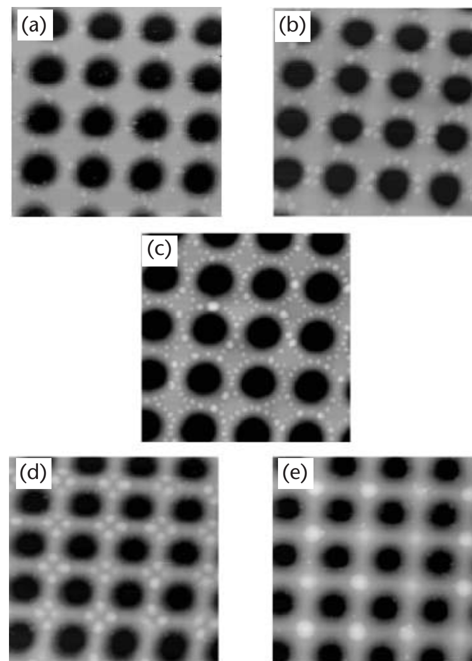


Figure 9.7 AFM images of self-assembled Ge dots grown on Si mesas at different growth temperatures: (a) 500°C, (b) 550°C, (c) 600°C, (d) 650°C, and (e) 700°C. The deposited Ge is 1.7 nm. The scale of the images is $4 \times 4\ \mu m^2$.

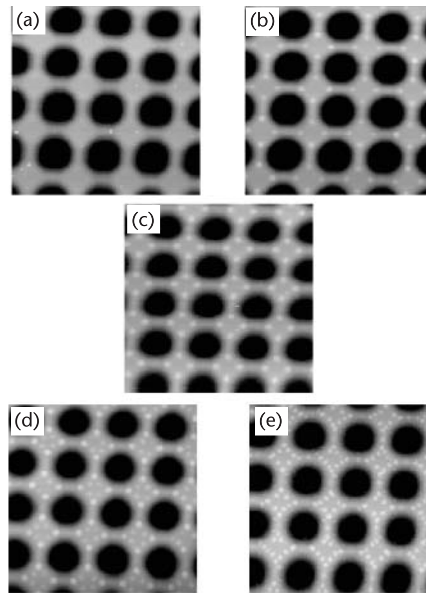


Figure 9.8 AFM images of the 2D arrangement of Ge dots on Si mesas with different amounts of Ge: (a) 0.4 nm, (b) 0.8 nm, (c) 1.1 nm, (d) 1.3 nm, and (e) 1.7 nm. The growth temperature is 600°C. The scale of the images is $4 \times 4\ \mu m^2$.

2D images for samples grown with various Ge deposition amounts and at the same temperature. When the Ge amount is 0.4 nm, very few Ge dots are seen, suggesting layer-by-layer growth. We believe that Ge growth on Si mesas is still in the Stranski-Krastanov growth mode below the critical thickness of the growth on planar Si(001) substrates for this feature size. For smaller feature sizes, the "critical thickness" may effectively increase because that energy may be smaller and feature dislocations are pushed out to the edges of the mesas. At the Ge thickness of 0.8 nm [Figure 9.8(b)], four Ge dots are formed at the corner sites of the Si mesa. The formation of dots at the corners arises from a similar mechanism of preferential nucleation. As the thickness increases to 1.1 nm, the result is similar to that with 0.8 nm Ge, that is, only four dots are located at the corner sites of each Si mesa. The difference is that the Ge dots at 0.8 nm are square-based pyramids, but the dots with 1.1 nm of Ge are domes, whose size is larger. With a further increase in deposited Ge, a second set of Ge dots is observed. They are located in the middle of the first set of Ge dots along the edges [Figure 9.8(d)]. The second set of Ge dots is square-based pyramids, which are different from the dots at the corners. The different morphology of the dots is attributed to the different nucleation stage. The pyramids are still at their early stage of the shape transformation, that is, they have not undergone the shape transformation into domes yet.

Based on our experimental observations, we can understand the formation mechanism of regimented Ge dots using an energetic analysis. The Ge growth on Si

mesas is in the Stranski-Krastanov mode. At the early stage of Ge deposition, the growth is in a layer-by-layer mode. Once Ge nucleation starts, Ge atoms tend to nucleate at the sites with minimum total energy. The corner sites are the energetically preferential sites because they are susceptible to strain relaxation, thus the nucleation first takes place at the corner sites. After the formation of the Ge dots and with the increase in dot size, the energy at the corner sites increases due to the increasing strain energy and total energy, and then the corner sites are no longer the preferential sites. On the other hand, sites such as the middle site between two corner dots become favorable sites even though they were not originally preferred. Then, as observed, the second set of Ge dots is formed between two corner dots. Moreover, the pyramidal dots are still at their early stage, whose size is smaller than that of the corner dots. As the Ge thickness increases further, Ge dots are formed at other sites as observed in Figure 9.8(e).

Dynamically, the formation of Ge dots on Si substrate is influenced by a large number of parameters, such as growth temperature, deposition rate, shape, and quality of the Si template. To understand the dynamics of the evolution of these nano self-assembled quantum dots, we consider the surface dynamics of the deposition as described by the equation for the surface height $h(r)$ as follows:

$$\frac{\partial h(r)}{\partial t} = -\nabla \cdot j(r,t) \tag{9.2}$$

where $j(r,t)$ denotes a macroscopic flux moving to the point on the surface r driven by the gradient of the local chemical potential:

$$j(r,t) = -\nabla \mu(r,t) \tag{9.3}$$

In turn, the gradient of the local chemical potential is defined by the difference between vapor and surface phases of Ge, surface interdiffusion, surface curvature, temperature gradient, and local fluctuation of the incoming Ge flux. Taking into account all of these factors, (9.2) can be rewritten in the following form:

$$\frac{\partial h(r)}{\partial t} = \nu \nabla^2 h(r) - K \nabla^4 h(r) + B\overline{\mu} + F + \eta(r,t) \tag{9.4}$$

where $\nu\nabla^2 h(r)$ is associated with growth desorption, $K\nabla^4 h(r)$ describes the surface diffusion, B is the mobility, $\overline{\mu}$ is the average chemical potential in the vapor phase, F is the incoming atomic flux (number of particles arriving per unit time per unit area), and $\eta(r,t)$ is the random fluctuation in the deposition process [an uncorrelated random number that has zero configuration average $\langle \eta(r,t) \rangle = 0$].

The competition between the terms $\nabla^2 h$ and $\nabla^4 h$ can be illuminated by introducing of a characteristic in-plane length scale L_1 [36]:

$$L_1 = \left(\frac{K}{\nu} \right)^{1/2} \tag{9.5}$$

For $L<<L_1$, the diffusion term $K\nabla^4 h(r)$ determines the scaling behavior, whereas for the longer length $L>>L_1$, the desorption term $\nu\nabla^4 h(r)$ controls the scaling [39].

Equation (9.3) for the surface height dynamics inherently contains information on the resulting growth structure symmetry. By rescaling (9.3) using $r \to br, h \to b^\alpha h$, one obtains $\nu b^{\alpha-2}\nabla^2 h(r)$ and $Kb^{\alpha-4}\nabla^4 h(r)$, which reflect *self-affinity* of the formed surface. This leads us to the so-called *self-affine functions* having the following form:

$$f(x) \sim b^{-\gamma} f(\gamma x) \tag{9.6}$$

The self-affine functions describe the subclass of anisotropic fractals, which appear in the diffusion-controlled growth processes [40–42]. The exponent γ parameter is called the Hölder or self-affine exponent and is associated with the fractal dimension $d_f = d_e - \gamma$, where d_e is the Euclidian dimension. Taking into account the competition between surface diffusion and the desorption processes, we have:

$$d_f = 4 - 2\alpha, \; L << L_1$$

$$d_f = 2 - 2\alpha, \; L >> L_1 \tag{9.7}$$

However, under typical MBE conditions, the desorption rate is very small (~1 ML s^{-1}) and L_1 is very large. Thus, the key mechanism controlling surface formation is surface diffusion.

In this case, we observe fractal-like dot formation during MBE growth on the spatially confined mesa. Figure 9.9 illustrates several steps of dot formation in detail. Instead of using a real-time scale, we marked the steps with the deposited

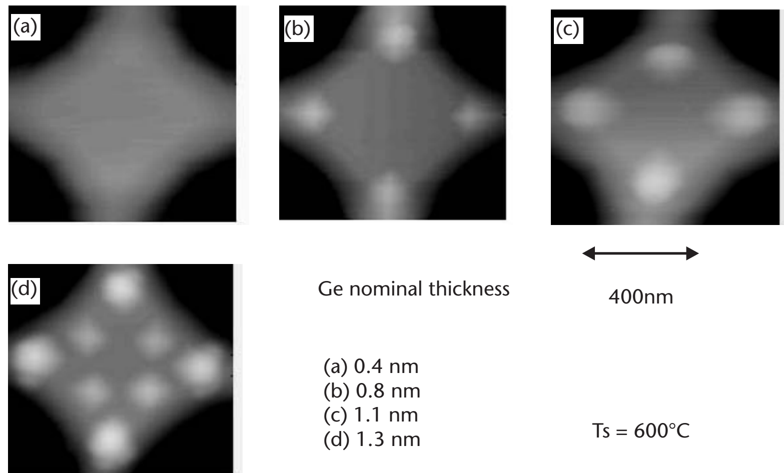


Figure 9.9 AFM images illustrating different phases in QD formation on a Si mesa: (a) layer-by-layer deposition, (b) primary QDs form, (c) primary QD size increases, and (d) new set of QDs form. The phases are marked in the deposited nominal Ge thicknesses of 0.4, 0.8, 1.1, and 1.3 nm, respectively. The growing temperature is 600°C.

nominal Ge thickness. At the initial stage of the growing process, no island formation occurs unless the nominal Ge thickness exceeds the critical one as shown in Figure 9.9(a). Then, four dots appear in the corners of the mesa as shown in Figure 9.9(b). All of these corner dots typically have the same size and pyramid shape. Continuous deposition results in an increase in the size or dot volume as shown in Figure 9.9(c). Finally, four new dots appear in the middle of the square as shown in Figure 9.9(d). Roughly, we can estimate the dimensionality of this "dot" fractal:

$$d_f = \lim_{l \to 0} \frac{\ln N(l)}{\ln(1/l)} \tag{9.8}$$

where l is the scaling length, and $N(l)$ is the number of similar objects under the scaling needed to cover the previous one. In Figure 9.9(d) an elementary object consists of four dots and it repeats itself as the length scale decreases about three times. At the same time, the square of the new set is about 2/9 times less than the previous one, so $d_f \approx \ln(9/2)/\ln(3) = 1.37$.

In a more rigorous consideration, one has to take into account two more processes that are particularly important for Ge deposition on Si substrate: atomic intermixing and strain energy. In reality, Ge QDs formed on the silicon template may contain a significant amount of Si atoms [12] as a result of the intermixing that takes place in MBE growth. Interdiffusion results in a change of strain energy, which affects atomic mobility and the desorption rate. Thus, a complete self-consistent system of differential equations describing surface dynamics, besides (9.3), should include diffusion equations for silicon and germanium and also take into account the transport coefficient modification caused by the change of strain.

9.5 Novel Device Applications

9.5.1 Optoelectronics

Due to its large bandgap, Si cannot find applications for optoelectronic devices in the optical communication window of 1.3 to 1.55 μm. Ge has the proper bandgap but the growth thickness on Si substrates is essentially limited to less than 10 nm [43] to maintain a defect-free situation. But for Ge dots, there is no such limit because the strain between Si and Ge can be relaxed through the formation of dots. Thus, Ge quantum dots grown on Si offer the opportunity for optical communication applications. For interband transition, pin diodes that use the dot layers to form an intrinsic region can be used. To enhance the absorption, multiple Ge layers are usually used. Photodetectors based on Ge QDs have been demonstrated [44, 45]. Three *p-i-n* structures each with 10 periods of Ge dots were grown by MBE at 540°C. PL measurements showed a Ge dot-related peak shift from 0.85 eV to 0.80 and 0.79 eV for the three Ge thicknesses of 1.2 (sample A), 1.5 (sample B), and 1.8 nm (sample C), respectively. *I-V* measurement of the pin mesa diodes showed a low dark current density of 3×10^{-5} A/cm^2 at -1V as shown in Figure 9.10 [44]. The breakdown

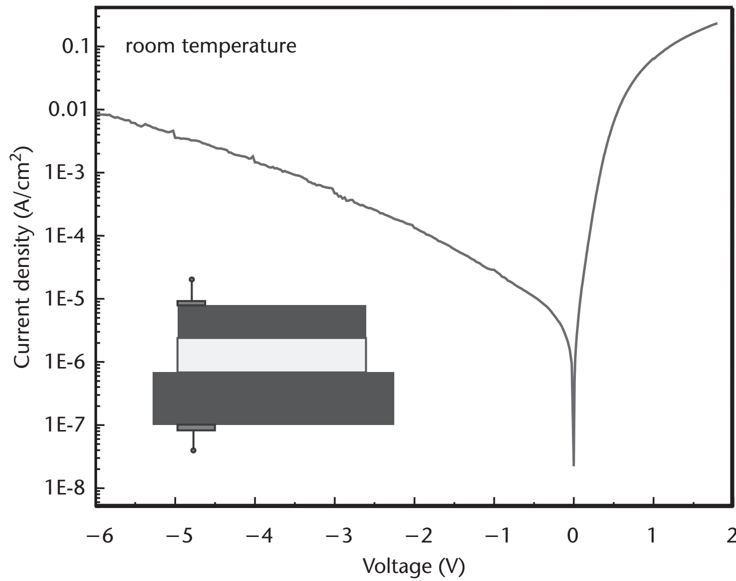


Figure 9.10 Current-voltage characteristic of device from sample B. It shows a low dark current of 3×10^{-5} A/cm^2 at –1V. Inset is a schematic drawing of the device.

voltage was 7V at room temperature. A strong photoresponse at 1.3 to 1.55 μm originating from Ge dots was observed. Figure 9.11(a) shows the photocurrent spectra of samples A, B, and C at room temperature, and the device was operated in the photovoltaic mode. All of the diodes show a photoresponse peak between 1.3

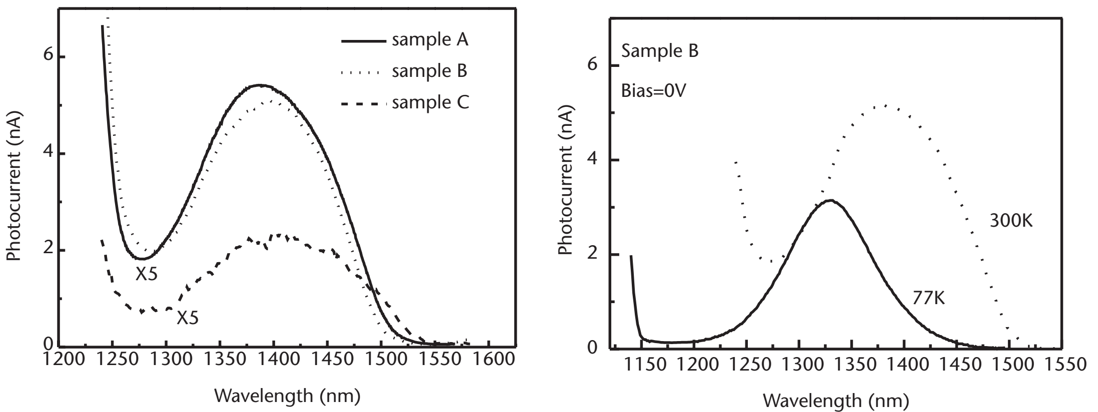


Figure 9.11 (a) Short-circuit photocurrent spectra of the Ge dot detector at room temperature for three sample pin photodiodes (A, B, and C). All the diodes show a photoresponse peak at between 1.3 and 1.55 μm. (b) Photocurrent of diode B at both room temperature and 77K. At room temperature, the response has a peak at 1.4 μm. The peak shifts to 1.32 μm at 77K.

and 1.55 μm. The figure also shows that as the Ge nominal growth thickness increases (from sample A to C), the photoresponse peak shifts slightly to longer wavelength. This indicates that the response range of the photodetector can be tuned by changing the Ge thickness during growth. In Figure 9.11(b), the photocurrent spectra of diode B at both room temperature and 77K are shown. The peak shifts from 1.4 to 1.32 μm with a temperature decrease from room temperature to 77K. The FWHM also shrinks from 95 to 70 meV, and the photoresponse intensity clearly decreases. At normal incidence, an external quantum efficiency of 8% was achieved at 2.5V. The dot layers may trap the light in the intrinsic region, and thus increase the absorption.

Likewise, multilayered Ge quantum dots can be used to fabricate novel quantum dot photodetectors based on intersubband transitions to operate in the MIR range. Because of the large discontinuity in the Si/Ge valence bands, the intersubband transition of holes from the ground state of the dots to the continuum states can be used for IR detectors working in MIR wavelength regions. Ge/Si photodetectors can also be monolithically integrated on Si substrates with CMOS readout circuits, making low-cost devices possible.

Figure 9.12 shows the polarization dependence of the intersublevel transitions for a QD sample. The sample used consists of a 200-nm undoped Si buffer layer, 30 periods of Ge QDs sandwiched between two 6-nm boron-doped Si layers, and a 50-nm undoped Si cap layer. The doping density in the Si layers is as high as 5×10^{18} cm^{-3} grown on a Si wafer with a resistivity of 18 to 25$\Omega \cdot$ cm. The size of the dots is

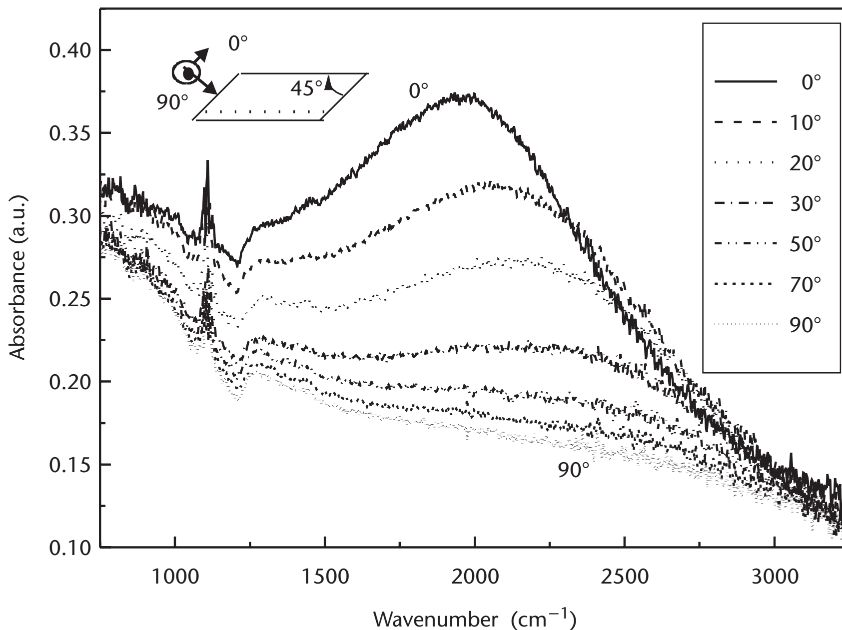


Figure 9.12 Polarization-dependent absorption spectra of the modulation-doped sample. The decrease in absorbance with increasing polarizer angle is due to the reduction of the component of the photon polarization along the growth direction.

obtained from a single-layered sample with an identical growth condition to be 2 to 3 nm in height and 20 to 30 nm in base diameter. As shown schematically in the inset of Figure 9.12, the 0° polarization angle corresponds to a 50% component of the incident IR light polarized along the growth direction of the structure, while the 90° polarization angle is defined as being parallel to the plane of layers. An absorption peak at around 2,000 cm^{-1} is observed. It can be seen that the absorption amplitude decreases with an increase in the polarization angle and the trend behaves like that of the QW-like feature. This is because the lateral dimensions of the present dots are much larger than their heights. Accordingly, the quantum confinement along the growth direction is much larger than that in the lateral direction.

For MIR applications, QDs have several advantages over conventional QWs. First, the predicted long carrier lifetime in the excited states in Ge quantum dots due to reduced carrier-phonon interactions could improve detector performance [46]. Second, because of their sharp δ-like density of states, the dark current level of QD infrared photodetectors is expected to be low when an appropriate doping level is selected. Another important advantage is the selective rule to enable normal incidence photon detection when the lateral sizes are further reduced [47]. One of the research directions may be to reduce the dot sizes to enable quantum confinement in the lateral directions.

Efficient photoemitters (LEDs) and perhaps Ge lasers might even be possible using Ge QDs. Multilayered Ge dot superlattices may be used as a gain media in which interband transitions in indirect semiconductors like Si and Ge are assisted by phonons to make up the momentum difference between the initial and final states. Furthermore, due to phonon confinement in a QD, phonons with the desired momentum can be designed by the dot size. In addition, because both QD size and interdot distance are comparable with the phonon wavelength, QD superlattices may be used to further control the phonon group velocity and used as a phonon filter. Thus one may be able to invoke phonon mediation to enhance interband optical transitions and Ge QDs may be used to attempt phonon-assisted lasers. A new field of phonon engineering may emerge. Again, the control of dot size to an even smaller dimension is of interest in future research.

9.5.2 Thermoelectricity

To fabricate highly efficient solid-state refrigerators, researchers have looked for decades for high figure-of-merit thermoelectric materials. Previously, QW structures of SiGe/Si have been shown to have an increased figure of merit due to the change in the density of states of electrons to increase the Seebeck coefficient and to reduce thermal conductivity at the same time, the latter of which is the result of the reduced phonon transport [48, 49]. Ge QD superlattices may further improve the merit and may be used to fabricate novel thermoelectric devices for the following reasons. First, it has been determined that QDs can effectively confine phonons [50] or even strongly scatter phonons [51], resulting in a further reduction of a thermal conductivity. Meanwhile, QDs have a δ-like density of states due to quantum confinement of electrons, which in turn increases the thermoelectric power

factor and thus the figure of merit. Indeed, QDs may constitute a better class of the "phonon-blocking electron-transmitting" materials compared with alloys, QW superlattices, and wires. The thermoelectric devices may be fabricated for improved devices.

9.5.3 Electronics Applications

Moore's law (http://www.itrs.net) demands a continuous increase in functionality per unit area and cost. The approach that we have been following is the scaling down of the physical feature size. As we approach the physical size limit due to either technology or the cost of fabrication, alternate means of increasing the functionality need to be pursued. One approach is to increase the functionality by integrating quantum devices. For example, tunnel diodes may be incorporated into CMOS to reduce the device account of the functional cluster; for example, we could reduce the number of transistors for a SRAM from 6 to 1 [52]. Ge QDs could also be used to fabricate improved tunneling diodes with a reduced valley current density because of the delta density of states [53]. For this purpose, the large valence band offset in the Si/Ge dots system may be taken advantage of. The indirect bandgap and the heavy masses of electrons do not make Si a good candidate for fabrication of resonant tunneling diodes. But, Ge dots embedded at the proper position within the *p-n* junction are expected to enhance the tunneling current considerably in Esaki (Zener) tunneling of electrons through the empty valence band of the Ge dots to the *p*-type region on the Si side. Many types of devices may be possible.

Likewise, well-organized, in particular, 2D arranged Ge QDs can be used to realize novel nanoscale cellular automata, a class of device/circuit that may minimize interconnection problems in today's CMOS circuits. 2D quantum cellular automata require a 2D array that is orderly arranged in space. Each cell of the array has a well-defined state representation, for example, ON/OFF for a nanodevice or simply the electron number in the cell. To achieve low-power consumption, high-speed operation, and high density, the size of the cells should be below 100Å according to the single electron model. The self-registered Ge QD growth has been shown previously to produce a high-quality 2D array using a patterned substrate of a much larger feature size. These arranged Ge QDs may be used either directly as a quantum subsystem or as a micromask to integrate more complicated quantum structures into the cell, in which a simple set of rules in updating the state may be used for image processing. This method is not the only way, but it offers a promising path to fabricate 2D quantum cellular automata.

9.5.4 Quantum Information Applications

Finally, in this subsection, we discuss potential applications of Ge QDs in quantum computers. The basic engineering prerequisites for successful implementation of a quantum computation device are creation of the quantum memory unit (qubit) with a decoherence time that is significantly less than the computation cycle [54] and

unitary rotations of the qubits along with the ability to control interactions among the qubits [55].

There are several physical reasons why Ge QDs are very promising for application in quantum computing compared with other designs proposed recently in this intensively pursued research area [56–60]. First, some theoretical evidence [61, 62] suggests that certain Ge-rich SiGe alloys give the type I alignment of the conduction band with coherently strained Ge grown on top. Thus electrons can be effectively localized in the QDs to serve as qubits. The electron trapped in the dot may be restricted in its orbital motion to the low-angular momentum states that will drastically suppress the spin-orbit interaction [63], thereby effective decoupling of the electron spin from the environment can be achieved. Second, because of the spatial confinement, the electron *g*-factor is a sensitive function of the electric field [64] and this effect allows efficient tuning of the electron spin in and out of resonance by a local electric field. Third, the chemical impurity concentration is a few orders lower than that in III-V compounds. Finally, most of the Si and Ge isotopes do not have a net nuclear magnetic moment, and isotopically pure Si can be obtained, allowing these materials to form paramagnetic centers creating a decoherence free environment for the electron spin. In particular, the absence of the dipole–dipole interaction between the electron and nuclear spins results in a significant increase of the spin relaxation time in comparison with the III-V compound materials [65, 66]. These advantages suggest that Ge QDs will have a great deal of potential in quantum information processing applications.

9.6 Conclusion

We have reviewed issues of Ge island formation on Si. We discussed uniform Ge island formation on planar Si and ordered island formation on prepatterned mesa structures. We also discussed the effect of growth conditions such as growth temperature, deposition rate, deposition coverage, and substrate patterning on the formation of the islands. Techniques of registration and regimentation of Ge islands were examined. Theoretical analysis of fractal-like dot formation was presented. We have also discussed the potential applications of Ge islands in the fields of optoelectronics, thermoelectricity, electronics, and quantum computing. The principles apply to any other island system such as II-VI, III-V mismatched epitaxy via the Stranski-Krastanov growth mode.

References

[1] Frank, F. C., and J. H. Van der Merwe, *Proc. Roy. Soc. Lond. A,* Vol. 198, 1949, p. 205.

[2] Stranski, I. N., and L. Von Krastanov, *Akad. Wiss. Lit. Mainz Math. –Natur. K1. Iib,* Vol. 146, 1939, p. 797.

[3] Volmer, M., and A. Weber, *Z. Phys. Chem.,* Vol. 119, 1926, p. 277.

[4] Eaglesham, D. J., and Cerullo, M., "Dislocation-Free Stranski-Krastanow Growth of Ge on Si(100)," *Phys. Rev. Lett.,* Vol. 64, 1990, p. 1943.

[5] Bauer, E., *Z. Kristallogr.,* Vol. 110, 1958, p. 372.

[6] Gossmann, H. -J., L. C. Feldman, and W. M. Gibson, The Influence of Reconstruction on Epitaxial Growth Ge on Si(100)-(2*1) and Si(111)-(7*7)," *Surf. Sci.,* Vol. 155, No. 2–3, 1985, p. 413.

[7] Asai, M., H. Ueba, and C. Tatsuyama, "Heteroepitaxial Growth of Ge Films on the Si(100)-2*1 Surface," *J. Appl. Phys.,* Vol. 58, No. 7, 1985, pp. 2577–2583.

[8] Miki, K., et al., "Scanning Tunneling Microscopy of Anisotropic Monatomic Steps on a Vicinal Si(001)-2*1 Surface," *Jpn. J. Appl. Phys. Lett.,* Vol. 28, No. 9, September 1989, p. L1483-5.

[9] Kataoka, Y., H. Ueba, and C. Tatsuyama, "Structural Properties of Heteroepitaxial Ge Films on a Si(100)-2*1 Surface," *J. Appl. Phys.,* Vol. 63, No. 3, 1988, pp. 749–759.

[10] Le Thanh, V., et al., "On the Formation of Self-Assembled Ge/Si(001) Quantum Dots," *J. Crystal Growth,* Vol. 201–202, 1999, pp. 1212–1217.

[11] Koide, Y., et al., "Initial Stage of Growth of Ge on (110)Si by Gas Source Molecular Beam Epitaxy Using GeH_4," *Jpn. J. Appl. Phys. Lett.,* Vol. 28, No. 4, 1989, p. L690–L693.

[12] Jin, G., et al., "Perfect Alignment of Self-Organized Ge Islands on Pre-Grown Si Stripe Mesas," *Appl. Phys. A (Mater. Sci . Process.),* Vol. 70, No. 5, 2000, pp. 551–554.

[13] Kamins, T. I., et al., "Evolution of Ge Islands on Si(001) During Annealing," *J. Appl. Phys.,* Vol. 85, No. 2, 1999, pp. 1159–1171.

[14] Kamins, T. I., et al., "Deposition of 3D Ge Islands on Si(001) by Chemical Vapor Deposition at Atmospheric and Reduced Pressures," *J. Appl. Phys.,* Vol. 81, No. 1, 1997, pp. 211–219.

[15] Williams, R. S., et al., *J. Phys. Chem. B,* Vol. 102, 1998, p. 9605.

[16] Goryll, M., L. Vescan, and H. Luth, "Morphology and Photoluminescence of Ge Islands Grown on Si(001)," *Thin Solid Films,* Vol. 336, No. 1–2, 1998, pp. 244–247.

[17] Wagner, C. Z., *Elektrochem.*, Vol. 65, 1961, p. 581.

[18] Zinke-Allmang, M., L. C. Feldman, and M. H. Grabow, "Clustering on Surfaces," *Surf. Sci. Reports,* Vol. 16, No. 8, 1992, pp. 377–463.

[19] Wynblatt, P., and N. A. Gojstein, in *Progress in Solid State Chemistry,* J. O. McCaldin and G. Somorjai, (eds.), Oxford, England: Pergamon Press, 1975, Vol. 9, p. 21.

[20] Ross, R. M., R. M. Tromp, and M. C. Reuter, *Science,* Vol. 286, 1999, p. 1931.

[21] Shchukin, V. A., et al., "Spontaneous Ordering of Arrays of Coherent Strained Islands," *Phys. Rev. Lett.,* Vol. 75, No. 16, 1995, pp. 2968–2971.

[22] Medeiros-Ribeiro, G., et al., "Shape Transition of Germanium Nanocrystals on a Silicon (001) Surface from Pyramids to Domes," *Science,* Vol. 279, No. 5349, 1198, pp. 353–355.

[23] Chiu, C. H., "The Self-Assembly of Uniform Heteroepitaxial Islands," *Appl. Phys. Lett.,* Vol. 75, No. 22, 1999, pp. 3473–3475.

[24] Daruka, I., and A. L. Barabasi, "Equilibrium Phase Diagrams for Dislocation Free Self-Assembled Quantum Dots," *Appl. Phys. Lett.,* Vol. 72, No. 17, 1998, pp. 2102–2104.

[25] Yoffe, A. D., "Low-Dimensional Systems: Quantum Size Effects and Electronic Properties of Semiconductor Microcrystallites (Zero-Dimensional Systems) and Some Quasi-2D Systems," *Adv. Phys.,* Vol. 42, No. 2, 1993, pp. 173–266.

[26] Sakamoto K., et al., "Alignment of Ge 3D Islands on Faceted Si(001) Surfaces," *Thin Solid Films,* Vol. 321, No. 1–2, 1998, pp. 55–59.

[27] Xie, Y. H., et al., "Relaxed Template for Fabricating Regularly Distributed Quantum Dot Arrays," *Appl. Phys. Lett.,* Vol. 71, 1997, pp. 3567–3568.

[28] Shiryaev, S., et al., "Dislocation Patterning—A New Tool for Spatial Manipulation of Ge Islands," *Thin Solid Films,* Vol. 294, No. 1–2, 1997, pp. 311–314.

[29] Capellini, G., et al., "Influence of Dislocations on Vertical Ordering of Ge Islands in Si/Ge Multilayers Grown by Low Pressure Chemical Vapour Deposition," *Semiconductor Sci. Technol.,* Vol. 14, No. 6, 1999, pp. L21–L23.

[30] Tersoff, J., C. Teichert, and M. G. Lagally, "Self-Organization in Growth of Quantum Dot Superlattices," *Phys. Rev. Lett.,* Vol. 76, No. 10, 1996, pp. 1675–1678.

[31] Le Thanh, V., et al., "Vertically Self-Organized Ge/Si(001) Quantum Dots in Multilayer Structures," *Phys. Rev. B (Condensed Matter),* Vol. 60, No. 8, 1999, pp. 5851–5857.

[32] Kamins, T. I., and R. S. Williams, "Lithographic Positioning of Self-Assembled Ge Islands on Si(001)," *Appl. Phys. Lett.,* Vol. 71, No. 9, 1997, pp. 1201–1203.

[33] Kamins, T. I., R. S. Williams, and D. P. Basile, "Self-Aligning of Self-Assembled Ge Islands on Si(001)," *Nanotechnology,* Vol. 10, No. 2, 1999, pp. 117–121.

[34] Kinam, K., "Growth of Si, SiGe and Selective Epitaxy by Gas Source Molecular Beam Epitaxy," Ph.D. dissertation, University of California at Los Angeles, 1994.

[35] Madhukar, A., "Growth of Semiconductor Heterostructures on Patterned Substrates: Defect Reduction and Nanostructures," *Thin Solid Films,* Vol. 231, No. 1–2, 1993, pp. 8–42.

[36] Qi, X., et al., "Interfacet Mass Transport and Facet Evolution in Selective Epitaxial Growth of Si by Gas Source Molecular Beam Epitaxy," *J. Vacuum Sci. Technol. B (Microelectronics and Nanometer Structures),* Vol. 14, No. 3, 1996, pp. 2381–2386.

[37] Jin, G., et al., "Mechanism of Preference of Edge-Positioning of Self-Organized Ge Quantum Dots on Si Mesas," *Proc. Mater. Res. Sci. Assoc.,* Vol. 571, 2000, p. 31.

[38] Vescan, L., "Ordered Nucleation of Ge Islands on High Index Planes on Si," *PHANTOMS Newsletter*, Vol. 16, 1999, p. 1.

[39] Barabasi, A.-L., and H. E. Stanley, *Fractal Concepts in Surface Growth*, Cambridge, England: Cambridge University Press, 1995, pp. 139–152.

[40] Saffman, P. G., and G. I. Taylor, *Proc. Roy. Soc. Lond. A,* Vol. 245, 1958, p. 312.

[41] Kessler, A. D., J. Koplik, and H. Levine, *Adv. Phys.* Vol. 37, 1988, p. 255.

[42] Pelce, P., *Dynamics of Curved Fronts,* Orlando, FL: Academic, 1988.

[43] People, R., *IEEE J. Quantum Electron.,* Vol. 22, 1986, p. 1696.

[44] Tong, S., et al., *Appl. Phys. Lett.,* Vol. 80, 2002, p. 1189.

[45] Yakimov, A. I., et al., *Appl. Phys. Lett.,* Vol. 75, 1999, p. 1413.

[46] Bockelmann, U., and G. Bastard, *Phys. Rev. B,* Vol. 42, 1990, p. 8947.

[47] Phillips, J., K. Kamath, and P. Bhattacharya, *Appl. Phys. Lett.,* Vol. 72, 1998, p. 2020.

[48] Hicks, L. D., and M. S. Dresselhaus, *Phys. Rev. B,* Vol. 47, 1993, p. 12727.

[49] Balandin, A., and K. L. Wang, *Phys. Rev. B,* Vol. 58, 1998, p. 1544.

[50] Liu, J. L., et al., *Appl. Phys. Lett.,* Vol. 76, 2000, p. 586.

[51] Khitun, A., et al., *J. Appl. Phys.,* Vol. 88, 2000, p. 696.

[52] Carns, T. K., X. Zheng, and K. L. Wang, *IEEE Electron Device Lett.,* Vol. 16, 1995, p. 256.

[53] Abstreiter, G., et al., *Semiconductor Sci. Technol.,* Vol. 11, 1996, p. 1521.

[54] Preskill, J., *Proc. R. Soc. Lond. A,* Vol. 454, 1998, p. 385.

[55] DiVincenzo, D. P., et al., cond-mat/9911245, November 16, 1999.

[56] Kane, B., *Nature*, Vol. 393, 1998, p. 133.

[57] Loss, D., and D. DiVincenzo, *Phys. Rev. A,* Vol. 57, 1998, p. 120.

[58] Sasakura, H., and S. Muto, quant-ph/9910100v3, October 27, 1999.

[59] Vrijen, R., et al., quant-ph/9905096v2, June 11, 1999.

[60] Pradhan, P., M. P. Anantram, and K. L. Wang, arXiv: quant-ph/0002006, February 2, 2000.

[61] Rieger, M., and P. Vogl, *Phys. Rev. B,* Vol. 48, 1993, p. 14276.

[62] Vescan, L., et al., *J. Appl. Phys.,* Vol. 87, No. 10, 2000, p. 7275.
[63] DiVincenzo, D., and D. Loss, *Superlattices and Microstructures,* Vol. 23, 1998, p. 419.
[64] Ivchenko, E. L., A. A. Kiselev, and M. Willander, *Solid State Commun.,* Vol. 102, 1997, p. 375.
[65] Gordon, J. P., and K. D. Bowers, *Phys. Rev. Lett.,* Vol. 1, 1958, p. 368.
[66] Kikkawa, J. M., and D. D. Awschalom, *Phys. Rev. Lett.,* Vol. 80, 1998, p. 4313.

CHAPTER 10

Carbon Nanotube Engineering and Physics

J. Xu, Brown University

The field of carbon nanotubes has advanced quickly and broadly on many fronts during the past decade. Controlled fabrication of carbon nanotubes of uniform diameter, length, and spacing is now feasible. Real and perceived potential applications in electronics, sensing, molecular biology, actuation, composite materials, and energy storage have been demonstrated. This chapter introduces some of these advances and some of the fundamental properties of the carbon nanotubes, discusses the underlying physics of new effects and phenomena observed or anticipated, and describes the controllable fabrication processes of new forms of nanotubes, as well as some interesting and relatively new and unconventional directions of potential applications.

10.1 Introduction

Soon after the landmark report on the observation of carbon nanotubes by Iijima in 1991 [1], numerous findings and assessments about nanotube properties and potential applications were put forward that established the carbon nanotube as an idea base material for future technologies in a broad range of applications. As a new material, carbon nanotubes are indeed remarkable by all accounts. They can be semiconducting with a bandgap ranging from approximately 20 meV to 2 eV, or be metallic, and capable of sustaining current densities that are hundreds of times greater than those of a bulk metal [2]. They come in lengths ranging from a few nanometers to a few hundred micrometers, with diameters of less than a nanometer up to a couple hundreds of nanometers. They can even take the form of a three-port transistor-like structure [3] and can be doped to modify their conductivity [4].

Mechanically, the nanotube is much stronger than steel, and may even be stronger than all man-made materials, but many times lighter. Thermally, it is more conductive than most crystals [5]. Upon charge injection it may change its dimension so significantly that it is not proper to call this effect piezoelectricity—and, in fact, it is not. Chemically, it is inert everywhere along its length except at the ends or at the site of a bend, kink, or defect. Finally, it responds to light—not only by changing its conductivity but also by changing shape [6].

It is thus not surprising that, merely a decade after the 1991 landmark report on its observation, the carbon nanotube has generated such great interest and been subjected to a growing range of explorations of its utility that permeates into virtually every field of science and engineering, with a publication and patent filing rate at thousands per year in 2001 [7].

The carbon nanotube is certainly a great development in the recent history of electronic materials and technologies. Is it a more important development than valuable materials with proven applications like GaAs, GaN, or the organic semiconductors that emerged not long ago and have already gained prominence in various fields of applications? It may be, if we factor in all of the other properties and potential capabilities offered by the nanotube.

The fact that all of these properties reside in one simple form of carbon is indeed unusual and naturally begs for an explanation. There are many explanations, most of which are rooted in the fact that in this system the electronic and mechanical degrees of freedom are tightly coupled, as can be shown via a first-principles model [8]. It is this strong coupling that makes the nanotube system unique and enables a whole array of new effects and properties. Scores of reports on nanotube force and chemical and optical sensing and actuating are only a few of the marvelous early demonstrations of the potential applications of these new effects [9–13].

The strong electromechanical couplings in nanotubes can not only bring about innovations and improvement in traditional fields, but can also open up entirely new paradigms for information acquisition and execution. For example, one can envision nanotube-based mechanical resonators, filters, and even amplifiers with performance superior to their electronic counterparts.

However great the nanotube's potential may be, it is not clear how rewarding it would be to channel the development of nanotubes into the popular directions of applications such as electronic transistors and memories. Moving in this direction means that nanotubes—and other molecular electronic devices, for that matter—are pursued primarily as alternatives to the silicon FETs and CMOS devices. Therefore, to make progress in this framework, one will have to figure out how to assemble the nanotube CMOS devices into a meaningful circuit. However cleverly designed, the circuitry will likely operate in the binary/serial scheme—in the absence of an alternative architecture, these assembled nanotube transistors will have to be made and function identically, and be individually and sequentially wired up. These are no small challenges, even for a small-scale integrated circuit demonstration.

It is obviously advantageous to make nanotube devices and circuits structurally and functionally similar to CMOS devices, if a nanotube device is to function in the same regime of binary serial signal processing. In this way, the vast and costly industrial infrastructure of the microelectronics technology and manufacturing base can be leveraged, instead of being replaced by another base technology that does not yet exist. On the other hand, the extraordinary staying power of the incumbent silicon CMOS technology cannot be underestimated. Simply put, silicon electronics has kept on marching forward from one generation to the next, blasting through whatever limits and insurmountable barriers it has been expected to hit. With the exception of the ultimate economic limit—the expected diminishing return on

investment—no show stopper is in sight to prevent silicon CMOS from continuing its domination in conventional integrated electronics. This observation is not intended to dismiss any effort in exploring this landscape to map out potential breakthrough points for an emerging technology. For now, however, it is hard to see where such breakthroughs in this well-served space would occur.

For these reasons, however interesting the technology issues and prospects pertinent to nanotube applications in the CMOS-like setting, they will not be discussed further in this chapter. Such issues have been extensively covered in the literature, and will likely be thoroughly addressed and reviewed more in the near future and by some of the most competent authors in the field. Instead, our attention in this chapter is focused on areas outside the realm of traditional electronics.

As soon as we step out of the confines of traditional electronics, we find ourselves surrounded by a vast field of new opportunities. On the road to future information technology, one can see the compelling needs for technological advances for acquiring information and executing actions on information, and the prospects of active interfacing with the biomolecular world. The array of immediately accessible, new and inviting opportunities is vast. However, this chapter covers only a few examples and attempts to present the review from selected viewpoints in some detail and depth.

For many of the applications, it is imperative for any new technology to have a high degree of control in fabrication and thereby in the properties of the products. This level of control is generally absent in the conventional processes of nanotube fabrication, based on arc discharging, laser ablation, or CVD. In these more conventional processes, the resultant nanotubes differ widely in their diameter and length, and therefore also in their physical properties. Additionally, they often come in the form of entangled bundles mixed with other particles and residual materials. Fortunately, a relatively new template-based fabrication approach demonstrated by our group in 1999 offers a promising solution to the problem [14]. Moreover, the approach, after further improvements, appears to be scalable for and adaptable to large-scale manufacturing.

In this chapter, we begin with a discussion of controlled fabrication of uniform nanotubes in highly ordered arrays as a base technology. We include in this discussion the more recent advances in controlled fabrication such as engineered nanotubes and nanotube-silicon heterojunctions. We then turn to nanotube functions in three regimes: external couplings with biomolecular systems, intrinsic electromechanical couplings within a nanotube, and coupling with an external radiation field. We finish by describing possible future advances, including loading of metal clusters and doping.

10.2 Controlled Fabrication of Uniform Nanotubes in a Highly Ordered Array

The stochastic nature of the traditional arc discharge and laser ablation processes of nanotube fabrication is such that it is extremely difficult, if not impossible, to achieve uniformity in or control over the diameter or length of the

resulting nanotubes. However, a template-based synthesis process has succeeded in overcoming some of the difficulties. This method makes use of a self-organized nanopore array formed from anodized alumina, as a growth template for nanotubes [14]. This template-based approach has recently evolved into one capable of fabricating Y-shaped nanotubes [15], and carbon nanotube-silicon heterojunctions.

It has been known since the late 1960s that anodization of aluminum under certainly carefully controlled conditions will form nanopores that will self-organize into a highly ordered array, as documented in a number of pioneering studies [16, 17] and as shown in Figure 10.1.

By varying the anodization voltage, electrolyte, and temperature by forced seeding of the nucleation sites via prepatterning and by postanodization pore widening, one can vary the diameter and the spacing of the nanopores over certain ranges while still retaining good ordering and uniformity. One can also vary controllably the pore depth via the anodization time during the anodization process or via postanodization ionmilling or etching to shorten the pores after growth.

Typically, we use a two-step anodization process to produce a highly ordered anodic alumina nanopore template. In the case of anodization by oxalic acid, an annealed and electropolished Al sheet of high purity (99.999%) is first anodized at 40V in a 0.3M oxalic acid at 10°C for 16 to 24 hours. After chemically removing the anodized aluminum oxide (AAO) film in a mixture of phosphoric acid and chromic acid, a second anodization is carried out under the same conditions for 3 to 5 hours. The anodized AAO template has a hexagonal structure with pore diameter of ~45 nm, cell size of ~100 nm, and channel length 10 to 20 μm.

Using such a highly ordered and uniform nanopore array as a growth template, highly ordered carbon nanotube arrays have been synthesized using the CVD process in which the nanopore surfaces serves both as a weak catalyst and growth guide. Good-quality nanotubes of the same diameter and length as that of the pores

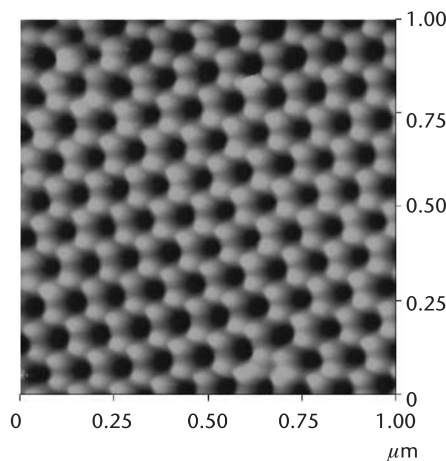


Figure 10.1 AFM image of the topography of a self-organized nanopore array formed in anodization.

can be routinely grown in this template CVD process. A well-known example of these nanotube arrays is shown in Figure 10.2. The near perfectly aligned nanotube array extruding out of the alumina template was obtained after a surface cleaning by an ion milling (~1 kV and ~1 mA) or by RIE (with BCl_3, 20 sccm, Ar 5 sccm, pressure 15 mTorr, and RF power 100W) step, followed by wet etching in 6 wt% H_3PO_4 and 1.8 wt% CrO_3 for about 10 hours. This example represented the best at the time of its publication in 1999 in terms of uniformity and ordering [14]. It is likely that it still represents the best ordered and uniform arrays available today.

The nanotube synthesis process begins with electrochemical deposition of a small quantity of Co, Ni, or Fe to the bottom of the hexagonally ordered array of nanopores. Next, the catalyst was reduced by heating in a tube furnace at 600°C for 4 hours under a CO flow (100 sccm). This process was followed by pyrolysis of acetylene or methane in the presence of the catalyst particles. Typically, the furnace temperature is kept between 650°C and 850°C for acetylene and higher for methane. The feed gas flow consists of 10% acetylene in nitrogen at the same 100-sccm flow rate for 0.5 to 2 hours. After synthesis, the nanotubes are annealed in nitrogen or argon for ~10 hours at the same temperature. A typical outcome of the process is *carbon nanotube* (CNT) array shown in Figure 10.2.

The parallel, periodically spaced, and highly uniform nanotube array is characterized by a very small diameter distribution and an even smaller length distribution. The mean diameter is determined by the pore diameter, which can be controllably varied from 20 to 100 nm while retaining a high degree of array ordering. The array is densely packed, and has a very high area density in the order of 10^{10} cm^{-2}.

The CVD synthesized nanotubes are normally multiwalled, but could be made single walled with different catalysts and feed gases at higher temperatures. CVD synthesized multiwalled nanotubes grown are likely to be more defective than the *single-wall nanotubes* (SWNTs) grown in the arc discharge or laser ablation processes.

For electronic device applications, a high level of defects could pose as much of a problem as the lack of control over the diameter, length, and chirality. For applications beyond the realm of electronic devices, properties such as the extraordinary

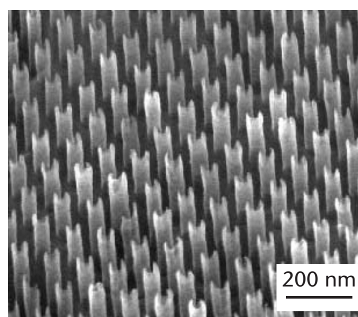


Figure 10.2 An example of the carbon nanotube array grown in an anodized nanopore template in a process described in the text and detailed in [14].

mechanical strengths (e.g., Young's modulus > 1 TPa, and bending strength > 10 GP [18]) and high current carrying capacity measured in *multiwalled nanotubes* (MWNTs), structural uniformity, and array formation are often more important. Moreover, it is known that postgrowth annealing can improve MWNT quality [19], as illustrated in Figure 10.3 where the graphitic lines of a multiwalled nanotube sample before annealing are compared with a sample annealed at a relatively moderate temperature of 1,400°C in Ar gas for 4 hours. Prior studies [19] have shown that annealing at higher temperatures (up to 1,800°C) would lead to more dramatic improvements in crystallinity.

Nanopores having diameters that were controllably varied via pore widening and lengths that were controllably varied via REI can be made in this way. Starting with a small-diameter nanopore array formed in anodization, one can readily obtain nanotubes that have larger diameters. An example of a 35-nm-diameter nanotube array, still embedded in the template, is shown in Figure 10.4. Smaller and larger diameter nanotubes have been grown this way.

For various applications, postgrowth processing may be desired to vary the extent of the nanotube extrusion from the template base or to trim the exposed nanotube length. Postgrowth trimming can be done with a high degree of precision in REI.

For many applications, such as mechanical oscillators, force sensors, and the biomimicking adhesion surfaces to be discussed later, it is desirable to have highly uniform free-standing nanotubes extruding out of the hosting substrate by a certain length. In principle, exposure of the nanotubes out of the template by any length can be readily obtained by selective etching of the alumina template surface. In practice, doing so while keeping the nanotube from sticking together turns out to be rather difficult, because of the high elasticity of the nanotubes and because of the van der Waals forces between the long parallel and straight nanotube surfaces. The presence of charge or chemical residuals during the etching, and therefore the associated or induced surface forces may have contributed to the difficulty we experienced. Figure 10.5(a) shows an oblique view of an exposed nanotube array after 6 hours of

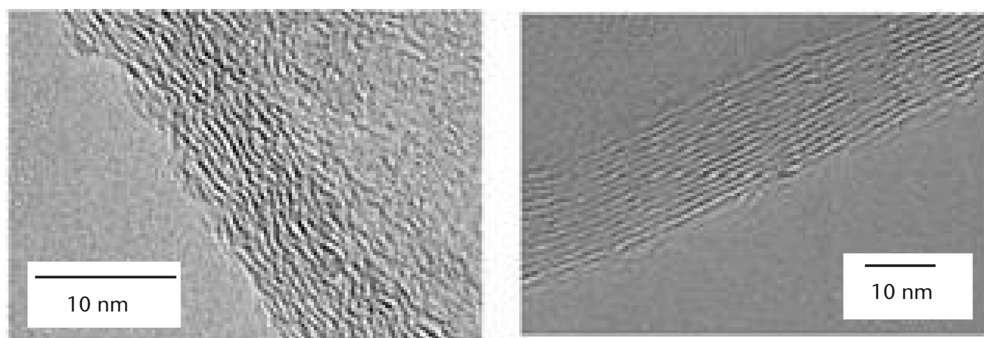


Figure 10.3 An example of the improvement of the graphitic lines of a MWNT sample by annealing at 1,400°C in Ar gas for 4 hours. *Left:* before annealing. *Right:* after annealing.

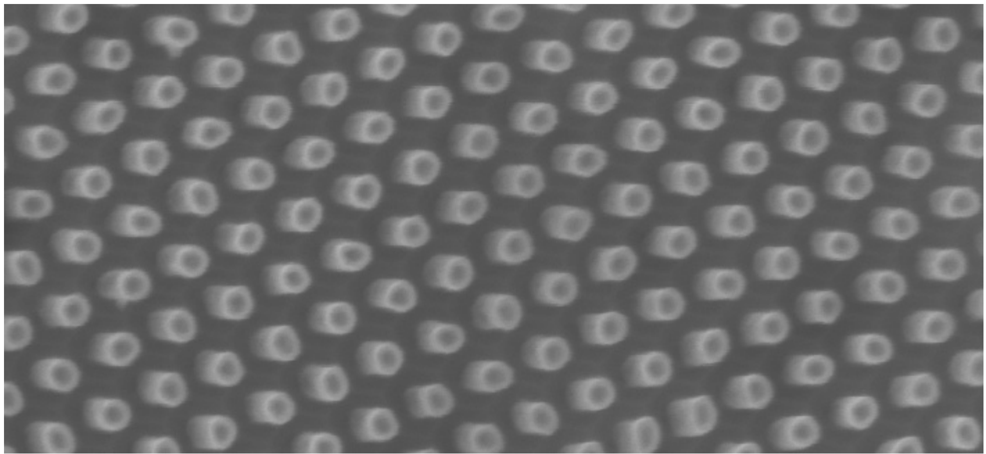


Figure 10.4 SEM view of a 35-nm-diameter nanotube array.

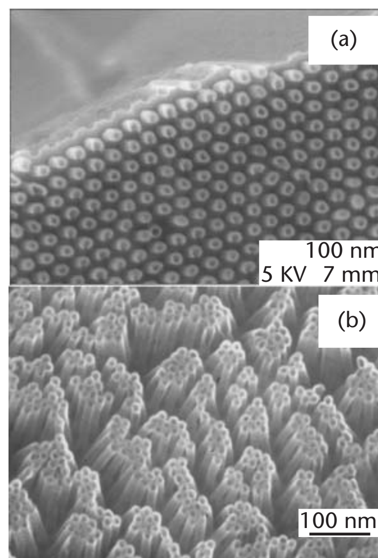


Figure 10.5 SEM image showing nanotubes clumping together after wet-etching back the alumina template at (a) 6 hours and (b) 16 hours.

etching in a 6 wt% H_3PO_4 and 1.8 wt% CrO_3, and Figure 10.5(b) shows the same view after 16 hours at the same etching conditions.

The specific length at which sticking and bundling start to occur varies widely with different etching processes and recipes. A clear and definitive understanding of

the underlying mechanisms is still lacking. However, our efforts to develop methods to counter this problem have met with limited success in which we obtained free-standing nanotubes of over 10:1 in aspect ratio without sticking. This was achieved by wetetching back the hosting alumina template in a solution containing dispersants. Further development of the dispersants-assisted etch process is under way and detailed in [20].

Findings to date [20] show that to avoid the natural tendency of the nanotubes to stick together and form haystack-like bundles, a postgrowth treatment process using a mixture of 6 wt% phosphoric acid and 1.8 wt% chromium oxide as the etchant, and 0.1 wt% gum arabic or 5 wt% polymethacrylic acid as the dispersant seems to be effective. This process has yielded for the first time well-aligned and spatially free-standing CNT arrays. The dispersants can be adsorbed on both the surface of carbon nanotubes and that of the alumina film, resulting in the elimination of aggregation of exposed carbon nanotubes, a more uniform dissolution of alumina, and a lower and thereby more controllable etching rate.

To reach this conclusion, different types of dispersants, varying in charge, molecular weight, and concentration, are used. *Polymethacrylic acid* (PMAA, sodium salt), *polyacrylic acid* (PAA), *gum arabic* (GA), *sodiumdodecyl sulfate* (SDS), *cetyltrimethylammonium chloride* (CTAC), and *poly(ethylene oxide)* (PEO) were tested. The etching times were varied from 18 to 36 hours. It was found that long-exposure CNTs with more uniform etching and fewer residues can be obtained for etching solutions with GA or PMAA, whereas tangled CNTs or CNTs covered with residues on the surfaces resulted when using other dispersants.

Figure 10.6 shows SEM images of the exposed CNT arrays etched in a mixture with 0.1 wt% GA and 5 wt% PMAA added at room temperature. The length of the exposed CNTs is more than 600 nm with an aspect ratio greater than 10, which would be satisfactory for many nanomechanical and nanomolecular applications. According to the experimental results, it was found that the AAO etching rate was much lower in the mixture with dispersant than without. We note that the SEM image of the cross-sectional view of the exposed nanotubes in Figure 10.6(b) is obtained on nanotubes coated with ~10- to 15-nm-thick Au/Pt, helping to improve the image quality by discharging the nanotubes pumped by the electron beams. The Au coating of nanotubes is also of interest to biomolecular binding applications, as discussed later.

In short, an effective and controllable method is now available for producing highly ordered, well-aligned CNT arrays. In addition, a postgrowth etching process has been introduced to yield free-standing nanotubes of lengths up to 500 to 800 nm over large areas. The CNTs are grown by pyrolysis of acetylene or methane into an alumina nanopore array template. Exposing the CNTs from the AAO template without sticking involves first cleaning the surface by dry-etching followed by dissolving the alumina in a mixture of 6 wt% H_3PO_4 and 1.8 wt% CrO_3 with an added dispersant of GA or PMAA. The extremely uniform arrays can be used in a variety of applications including nanomechanical oscillators, field emission displays, probes and detectors for biomedical use, and IR and acoustic imaging.

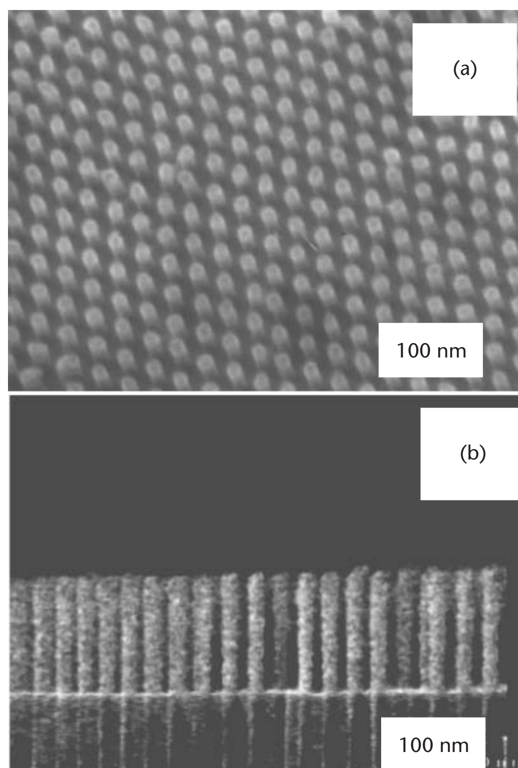


Figure 10.6 SEM oblique view of arrays of long (500-nm) nanotubes extruding out of the alumina template, which was etched back in (a) 5 wt% PMAA for 25 hours and (b) 0.1 wt% GA for 25 hours, with a thin layer of sputtered Au/Pt on the side wall.

10.3 Interfacing with Biomolecules and Cells

Monitoring, understanding, and effecting control of the biomolecular activities and the inner workings of a living cell have been a long-sought-after goal of life sciences and medicine. However, the advances on this front have been limited, largely by the available tools and approaches.

The development of a CNT-based molecular and cellular interface technology could open a new pathway and enable revolutionary advances, such as molecular-scale conversion of bioinformation in protein and cells into digital electronic signals and conversely electronic control and manipulation of cellular and protein functions. This may sound like an ambitious proposition and is most likely unprecedented, but the experimental trials to date suggest that given sufficient efforts and resources there is a good chance of success.

Indeed, we have successfully created nanoscale probes constructed from carbon nanotubes. Using such a nanotube probe, one could measure—and we have measured—enzymatic activity through the direct conjugation of redox enzymes to

nanotubes. Furthermore, it appears entirely possible that one could introduce biomolecules and chemical into a living cell via a CNT probe.

Man-made biosensors, however advanced, are likely to remain inferior to biomolecules in specificity, sensitivity, and efficiency. It is well established that biomolecular responses can be of single-molecule precision and specificity. And, their responses can be so efficient that they are often termed *nondissipative,* implying little or no energy loss in the process—a notion rather foreign to the world of conventional electronics. The sensitivity, specificity, and efficiency of biomolecular interactions are hardly surprising, given the long evolution of "engineering and optimization by Nature."

The development of such nanotube interface technologies will enable direct, real-time bioinfo acquisition by transducing biomolecular signals into electronic forms for subsequent information extraction or processing on a conventional electronics platform. One example [21] is the reading of redox activities via a CNT array probe in the form of electrical current pulses. In this case, the *glucose oxidase* (GOx) is linked via coupling of its amine (RNH_2) group to a COOH bond at the tip of a nanotube. The GOx–nanotube complex is placed in a flow cell where glucose droplets are injected into the flow and, on reaching the GOx, produces a current pulse in the nanotube probe that can be detected, amplified, and displayed by conventional electronic circuitry in real time.

In experiments [21], the electron transfer rate in a molecularly linked GOx–nanotube assembly proved to be orders of magnitude greater than with GOx adsorbed onto a conventional carbon electrode or even on the surface of a CNT mat [22].

It is an interesting hypothesis to verify that electrical access to redox-active enzymes could be a means for modulation of their activity and, consequently, control of the reaction. However, the redox centers of these enzymes are typically buried inside the protein with the surrounding shell acting as a protective barrier that impedes electron transfer. In addition, the surface interaction is such that many proteins denature on contact with a conventional electrode, effectively coating the electrode with a layer of polymer that reduces the transfer of carriers from the other, still-active enzyme.

The vertically aligned carbon nanotubes offer important advantages in this context: They are chemically inert except at the ends, which can be activated with carboxyl groups or further biofunctionalized [9–13]; they are mechanically stronger than most, if not all, other manufactured materials; they are able to conduct higher electric current density than metals; and they have the highest thermal conductivity of all materials. The very large aspect ratio and vertically aligned configuration of the nanotube greatly improves the contact surface area and the ease of access of the target molecules to the sensing proteins anchored atop the nanotubes. Due to the small diameter (2–100 nm) and thus large radius of curvature, nanoparticle surfaces are less likely to denature proteins, as observed in experiments [21]. Furthermore, the nano tip may penetrate into the protein structure and access the buried redox center. A few groups have tried to connect nanotubes to proteins, but have met difficulties arising from the extraction, processing, and assembling of the nanotubes,

which are normally in the form of entangled bundles and random meshes. These difficulties are largely eliminated in the highly ordered, arrayed nanotubes discussed here. Real-time [21] and static [22] recordings of electronic current pulses resulting from the GOx redox action via the contacting nanotube electrodes have been demonstrated.

Although these experimental findings may be specific to the GOx–nanotube complex, the implication of this direction of exploration is greater, because it may take us to the intersection of traditionally separated fields of science and engineering. Biomolecular responses to stimuli are extraordinarily sensitive, specific, and efficient. When combined with electronic signal processing and control, they could open up a wide range of new opportunities of vast value to information and medical technologies.

In Figure 10.7 a molecular linker is shown that plays the triple role of molecular recognition and thereby is self-assembling and acts as a mechanical link and electrical conduit. Natural molecular linkers, such as peptides, leucine zippers, and so on, do exist, but making them function in all three roles is not a simple task. Various attempts to engineer the desired functionalities into the molecular linkers are under development. The work on engineered DNA conductivity in which the imino proton bond between the base pairs is replaced by a divalent ion (as seen in Figure 10.7) is a good example of some of the most recent and novel developments [23].

Developing a conductive molecular linker is only half of the task. Linking it to a nanotube often requires functionalizing the ends of the nanotube. Luckily, nanotubes happen to be chemically active at the ends. For example, this was demonstrated in a reactivity study where it was shown that although amorphous carbon

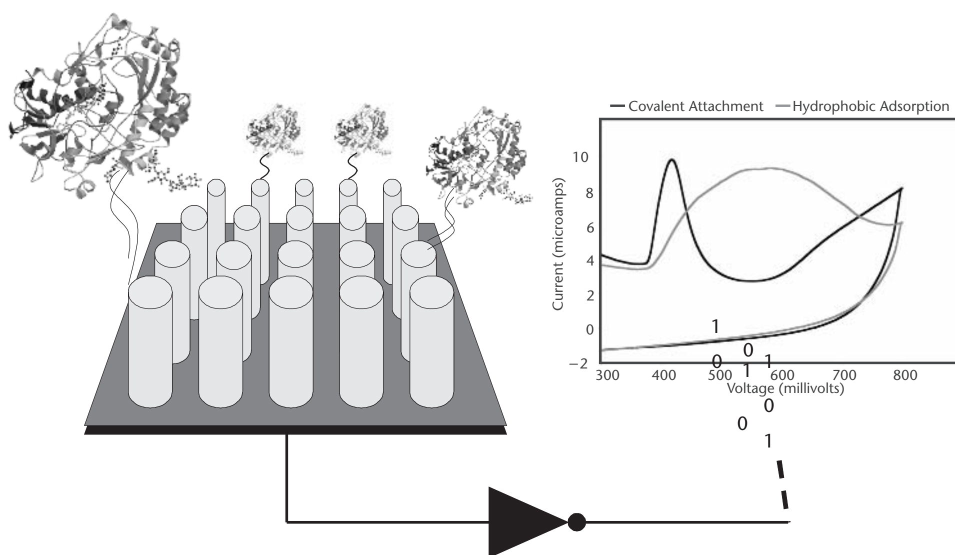


Figure 10.7 Illustration of a protein–nanotube assembly and electrochemical responses corresponding to different binding strategies (black solid line—binding of a protein via the amine group of the protein to a carboxyl group at the nanotube tip; gray line—adsorption on the side wall).

can be attacked from any direction, carbon nanotubes can be oxidized only from the ends [24, 25]. Moreover, from our experiments and the work of Guiseppi-Elie et al. [22], it appears that carboxyl groups can be attached to the ends of the nanotubes relatively easily.

With carboxyl groups at its exposed ends, a nanotube can be covalently bonded both to DNA and many other molecules, biological or otherwise. This capability is very appealing and can enable many other applications such as molecular sensing via chemical and molecular binding forces. Charles Lieber's group demonstrated the first such experiment in which they coupled amines (RNH_2) to the COOH group on the nanotube probe and used it to demonstrate pH-independent chemical force probing of a self-assembled monolayer surface. They also demonstrated nanotube binding via the COOH to biotin followed by streptavidin, opening the door to interesting medical and biological applications beyond sensing [9–13].

Based on this success, it is hard not to share their enthusiasm and to agree with their general assessment that such functionalized nanotube tips have much potential in probing specific intermolecular forces, which are signature properties of many chemical and biosystems, and in molecular imaging with single-molecule resolution.

10.4 Intrinsic Quantum Electromechanical Couplings

Our attention up to this point has been on external couplings of nanotubes, with couplings to molecular systems being a particular focus. Even more remarkable properties and applications of nanotubes are expected of the intrinsic and unusually strong coupling between the electronic and mechanical degrees of freedom within the nanotube. The effects of this coupling appear to hold great promises in opening up new opportunities for advancing the science and engineering of *nanoelectromechanical* (NEM) systems and actuators.

Although ongoing research has yet to present answers to many open questions about the dynamic, mechanical responses of the nanotubes to electronic stimulation and vice versa, a reasonably good understanding of the static electromechanical interactions has already been established. Additionally, from first-principles considerations, we can already offer an intuitive explanation of why CNT electronic properties, whether single walled and or multiwalled, are strongly dependent on their diameter and helicity.

Carbon is interesting and unique in that it can form the mechanically soft and metallic graphite or the very hard and insulating diamond. In the former case, the carbon atom's covalent bonds lie in the 2D plane, as so-called sp^2 bonds, whereas in the latter its bonds extend into 3D space and are termed sp^3 bonding. Carbon nanotubes—as well as C_{60} and other fullerenes—are in a third stable (or metastable) form of carbon lattice structure. Intuitively, a carbon nanotube can be viewed as a graphite sheet rolled up into a single-walled or multiwalled tube in which the originally planar sp^2 bond vectors are bent along with the graphite plane, thus projecting a component into the third dimension and thereby blending in some of the characteristics and properties of the sp^3 bonds. It is therefore not surprising to find that carbon

nanotubes exhibit mechanical and electronic properties in between graphite and diamond. The smaller the diameter, the more bent the sp^2 bonds are, and the more like sp^3 they become, leading to a wider bandgap. The dependence on helicity can be interpreted similarly, but this dependence is well defined only for single-walled nanotubes. For the NEM and actuation applications we are interested in here, multiwalled nanotubes are the focus, because their diameters, lengths, and spatial arrangements can be controlled during fabrication.

It has been shown through a more detailed and quantitative analysis that the bandgap of a nanotube depends inversely on the local radius of curvature (see Figure 10.8)—the diameter and helicity [26]. In the case of multiwalled nanotubes, the helicity of one wall is unlikely to be correlated with that of the next wall. As a result, the effect of the helicity is averaged over multiple walls, and the bandgap is found theoretically and experimentally to follow a simple linear relation with the inverse of the diameter.

Underlying this remarkably simple appearance of the bandgap-diameter dependence is a not-so-simple quantum electromechanical coupling of the mechanical and electronic degrees of freedom. In general, one can write a Hamiltonian in three parts: an atomic energy part, an electronic energy part, and an electron-phonon interaction part. In most systems, the third part is a small perturbation. Thus, the first two parts are normally treated separately, and in fact by two mostly separated communities: solid mechanics and electronics, each to a high level of sophistication. However, in carbon nanotubes, the system minimizes its total energy by balancing its distributions between the electronic states and the mechanical strain energy, thanks to the strong coupling term in the Hamiltonian. As such, the system would therefore respond to a change in electronic state occupancy with a change of mechanical energy, and vice versa, giving rise to a potentially giant electromechanical interaction.

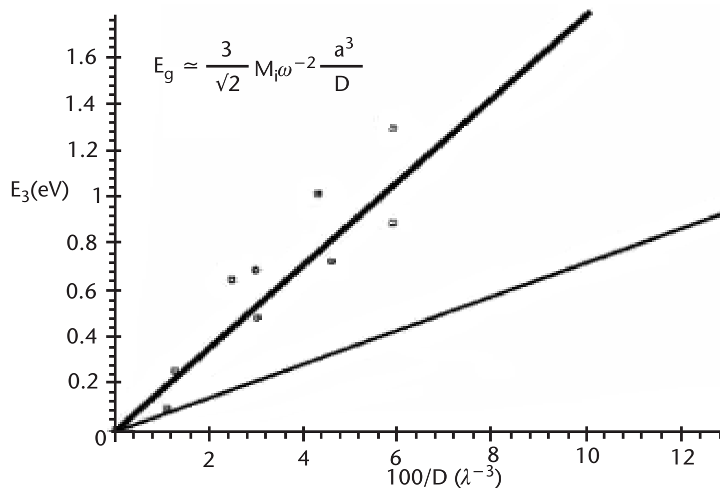


Figure 10.8 Theoretical (solid lines) and experimental bandgap dependence on the inverse of nanotube diameter, that is, 1/D. (*After:* [8].)

However sound this intuitive argument might be, its experimental and quantitative validation remained elusive until most recently when conditions become ready for a comprehensive approach, combining theory with direct experiments, to the problem.

The discussion we present in this chapter is based on a first-principle model following Feynman's coupled-mode method. We leave the more quantitative and detailed density functional analysis and experimental findings to a more formal treatment [27].

This coupled-mode model pays little attention to the specific details of the atomic system, and instead focuses on the fundamentals of the coupling between adjacent atoms, in this case the coupling between bonding and antibonding orbitals of adjacent atoms, and the minimal energy eigenstates that correspond to the mixing of these orbitals. For the sake of simplicity and clarity, one can focus only on the orbitals of the π-electrons, although clearly in a more detailed context the σ-electrons of SWNTs may play a role, and some of the theoretical effects of $\pi-\sigma$ hybridizations have been commented on in the literature [28, 29].

Alternatively, the electromechanical couplings of a few specific nanotube structures can be examined in greater detail via computational simulations, particularly those based on the *density functional theory* (DFT) of Hohenberg and Kohn [30]. However, these details tend to be model specific and sometimes code specific, and they agree qualitatively with one another in certain contexts, but differ in others. The reasons for this may either be physical or merely relics of computation, and the manifest sensitivity of *ab initio* computations to initial parameters.

The first-principles coupled-mode approach is thus often more appealing because it complements the numerical models by providing insights into the principal effects free of the model-specific and/or structure-specific complexities. Along with the first-principle approach complemented by DFT numerical modeling, a set of direct experimental measurements of the lattice responses to excess electron injection or extraction has been performed by a team in our lab for experimental validation and for further insights [27].

The first-principle approach to the quantum electromechanical couplings can be summarized as follows. One can start with the assumption that the π-orbital contributes principally to the electromechanical behavior, and that the orbital may be viewed as being split into distinctive bands with bonding and antibonding characteristics, respectively. These two coupled bands work to shift the ground state energy by some small amount, either positively or negatively, in response to a variation in the lattice parameters. In the coupled-mode model, one can determine how the ground state energy is varied by adding a small amount of excess charge δq to the system, either positively or negatively, depending on which of the π-bands accepts how much of the new charge.

Conceptually, the injection of charge alters the energy of the electronic subsystem. But, because of the strong coupling between the electronic degree of freedom and the mechanical (atomic) degree of freedom, the excess electronic energy could be partially coupled to and taken up by the atomic subsystem in the forms of bond extension or contraction.

The issue of solving the standard Hamiltonian eigenvalue problem for SWNTs has been treated extensively in many instances. For ease of comparison, we adopt the fairly conventional notations in the literature, in particular for the assumed form of the eigenspectrum of the closed shell energy, as laid out, for example, in [28, 29]. We assume in particular that the effects of the coupling of the bonding and antibonding states of the π-electron valence and conduction bands, respectively, vary the ground state such that

$$E^{+,-} = E_0 \pm \alpha(\mathbf{k}) f(\mathbf{k}) \tag{10.1}$$

where $\alpha(\mathbf{k})$ is a perturbing coupling parameter and $f(\mathbf{k})$ is a relatively slowly varying function of $\mathbf{k}$. The term E_0 represents the ground state energy of the closed shell electron in the $2p$ orbital, whose eigenspectrum can be determined in ways explained in [31], from a modified, curved surface Hamiltonian such as

$$\mathbf{H}_{\text{curved}} = (2\pi v / T)\sigma(m + \Phi) \tag{10.2}$$

where $T = n_1 a_1 + n_2 a_2$ is the wrapping vector for an (n_1, n_2) tube, and σ is a Pauli spin matrix. The corresponding spectrum of eigenvalues represents that of the free electron of the closed $2p$ shell in the SWNT, which can then be split into two distinct states, with bonding and antibonding characteristics, by the bonding and antibonding levels of the π bands represented by the factors $\alpha(\mathbf{k})f(\mathbf{k})$ in (10.1). The quantity $v \propto \alpha(\mathbf{k})$ in (10.2) is proportional to the hopping integral [32, 33].

The coupling function $\alpha(\mathbf{k})$ can be taken in the tight-binding, nearest-neighbor approximation to have either a linear or quadratic dependence on deformation (depending on whether the tube is metallic or semiconducting, respectively.)

The linear assumption was incorporated in different ways in the literature. We take a more first-principled approach and permit the context of the problem to determine the function $\alpha(\mathbf{k})$. The expression for the energy splitting between the coupled mode states can be written as

$$E^{+,-} = \alpha(\mathbf{k})\{\mu \pm f(\mathbf{k})\} \tag{10.3}$$

with $\mu = \{(2\pi/T)^2 (m + \Phi)^2\}^{1/2}$. This way, $\alpha(\mathbf{k})$ can then determined by imposing on (10.3) the condition that the ground state energies, with both bonding and antibonding characters, be at or near an equilibrium state for the system, that is, $\partial E^{+,-}/\partial \mathbf{k} = 0$. The function $f(\mathbf{k})$ that is convenient for the tubular system has the form [34] $f(\mathbf{k}) = e^{ikxa/\sqrt{3}} + 2\, e^{-ikxa/2\sqrt{3}} \cos(k_y a/2)$, which contains within it the condition of continuity in the axial direction but quantization in the circumferential direction, and can be further reduced to a simpler form because we are interested in the changes of k_y:

$$f(k_y) = 2\cos(k_y a/2) \tag{10.4}$$

where a is the bond length. We can easily solve for $\alpha(k_y)$, where the coupling parameter has also been modified to depend functionally only on k_y. For small variations, one need only to keep first-order terms in k_y:

$$\alpha\left(k_y\right) = K\exp(\mu \pm 2)\left(1 - \mu k_y\right) \tag{10.5}$$

One can find how $\alpha(k_y)$ varies with the addition of a small amount of charge δq by looking at $\partial\alpha(k_y)/\partial q$. To determine this quantity, we note that k_y also depends on the total charge of the system, generally following a power law, $k_y \sim q^\gamma$.

We obtain the variation in the coupling as follows:

$$\partial\left(k_y\right) / \partial q = K\exp(\mu \pm 2)(1 - \mu)\gamma q^{\gamma - 1} \tag{10.6}$$

If we now reconsider (10.3) with the new, charge-varied coupling, and a new function $f(ky_{\text{new}})$, which has the same form as the previous one but with a new lattice term determined by the charge addition, then

$$E_{\text{new}} = \partial\alpha\left(k_y\right)/\partial q\left[\mu \pm f\left(ky_{\text{new}}\right)\right] \tag{10.7}$$

We can determine ky_{new} by assuming that (10.7) is again at or near equilibrium and taking $\partial E_{\text{new}}/\partial ky_{\text{new}}$. If once again we only retain terms to linear order in ky_{new}, and proceeding as we did earlier, we find that

$$ky_{\text{new}} \sim \gamma q_{\text{old}}^{2\gamma-2}\left(q_{\text{old}} + \delta q\right)^{1\gamma}\ \exp(\mu \pm 2) \tag{10.8}$$

allowing us to write out the fractional variation as

$$\delta k_y / k_y = \left(ky_{\text{new}} - k_y\right)k_y = \gamma q_{\text{old}}^{\gamma-2}\left(q_{\text{old}} + \delta q\right)^{1-\gamma}\ \exp(\mu \pm 2) - 1 \tag{10.9}$$

by defining δq as some fraction of q_{old}, specifying a particular γ and recalling that $\mu \sim [(2\pi/T)^2\ (m + \Phi^2]^{1/2}$, which contains some information about the wrapping vector of a given tube, in the T term as well as the "phase" term Φ.

A few representative results for various values of γ are shown in Figures 10.8 and 10.9. Specifically, the cases of $0 < \gamma < 1$ and $1 < \gamma < 2$ are looked at, which are indicative of the qualitative behavior of the lattice in those scaling regions. As can be seen in Figure 10.9, for $1 < \gamma < 2$ and $0 < \gamma < 1$, respectively, the lattice generally expands as positive or negative charge is added. But, there is a slight contraction between $\delta q/q = 0$ and 0.01.

These theoretical findings can be compared with that obtained from Raman spectroscopy. It is has been well established that Raman modes in CNTs are *radial breathing modes* (RBMs), which occur between 100 and 300 cm^{-1}, the G-band (tangential) at ~1,600 cm^{-1}, the D-band (disorder-induced) at 1,300 to 1,400 cm^{-1}, and the G´-band (D*-band) at 2,600 to 2,800 cm^{-1} [2]. Other peaks in the spectrum are from the anti-Stokes process or harmonics.

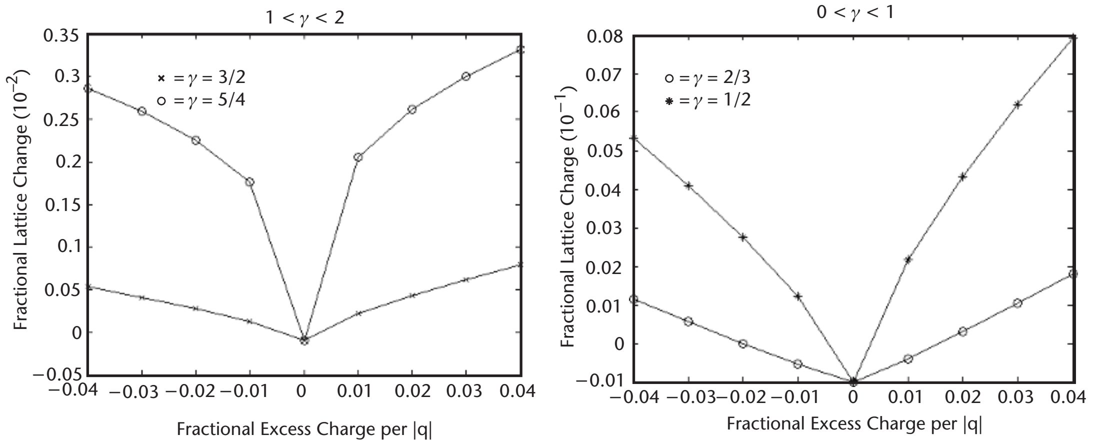


Figure 10.9 First-principle coupled-mode model of lattice expansion and contraction as function of excess charge injection and extraction.

In a defect-free graphite sheet the Raman spectra contain all of these peaks except the D-band, which originates from double resonant Raman scattering from a defect in the case of graphite. In this process the excited electron will go through a double scattering from a phonon and a defect. Carbon nanotubes can be viewed as a graphite sheet rolled up into a tubular structure, which effectively increases the number of sp^3 bonds and thereby the intensity of the D-band. The origin of the G´-band is double scattering, like the D-band but this time both scatterings are from phonons, so we have this peak regardless of any defect in the system.

Figure 10.10 illustrates the modes of atom movements for zigzag and armchair tubes. From all possible oscillation related to different symmetry groups in SWNTs. Only A_{1g}, E_{1g}, and E_{2g} modes are Raman active where, in the A modes, atoms in the

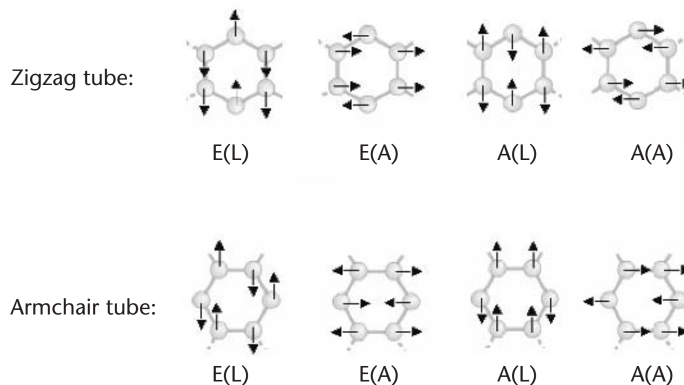


Figure 10.10 Lattice oscillation modes. Zigzag tubes—G-band modes: A_{1g}(L) , E_{1g}(A), and E_{2g}(L); D-band modes: A_{1g}(L) , E_{1g}(A) , E_{1g}(L), E_{2g}(A), and E_{2g}(L). Armchair tubes—G-band modes: A_{1g}(A), E_{1g}(L), and E_{2g}(A); D-band modes: A_{1g}(A), E_{1g}(A), E_{1g}(L), E_{2g}(A), and E_{2g}(L).

same circle around the circumference of the nanotube oscillate all in phase, and in E_n modes there are $2n$ nodes in the wave. The notations A and L in parentheses are related to oscillations parallel and perpendicular to the tube axis, respectively. Also all out-of-plane vibrations fall in the low-frequency portion of the spectrum, whereas the G- and D-band modes are all in-plane vibrations. If there is any change in resonant frequency due to a C–C bond change, it will show up in their corresponding spectra visibly.

In addition to the fact that the G-band has a sharper peak and is therefore easier to detect shifts in than the D-band, there are several reasons for us to consider the G-band peak for the purpose of detecting the bond length changes. The D-band is sensitive to the diameter distribution of and to the stacking and bundling coupling between the tubes in the sample (such as those used in our experiments), whereas the G-band is much less sensitive, especially for large-diameter tubes (such as those in the experiment reported here: ~50 nm) [35]. Also due to the double resonance origin of the D-band, its spectral position will shift with respect to the excitation energy while the G-band positions remain unchanged with respect to the energy change.

To relate the G-band Raman peak shift to the C–C bond length changes, we can consider a flat graphite sheet. For any ith atom vibration in the system, we can write

$$M_i \ddot{\bar{\mathbf{u}}}_i = \sum_j \mathbf{K}^{ij}\left(\bar{\mathbf{u}}_j - \bar{\mathbf{u}}_i\right) \tag{10.10}$$

where $\mathbf{K}^{ij}$ is the 3×3 force matrix and the sum over j is over all the neighboring atoms. It can be shown that

$$\omega^2 = \left\{\sum_j k^{ij}\left(\mathbf{k}\cdot\mathbf{r}_{ij}\right)^2\right\}\left\{a^2/2M_i\right\} \tag{10.11}$$

where $\mathbf{r}_{ij}$ is the distance between the ith and jth atoms. From this we can see that any change in the C–C bond length will shift the in-plane Raman mode frequencies linearly, that is, $\Delta\omega \sim \Delta a$.

In the Raman experiment, uniform CNTs, extracted from the nanotube arrays shown in, for example, Figures 10.2, 10.4, and 10.5 by dissolving away the alumina template, were dispersed on a Si-SiO2 wafer. By applying different voltages across the sample we can controllably vary the static charges on the CNTs. Figure 10.11 shows the measured shift of the G-band peak with applied voltage (charge) of both polarities. A monotonic increase of the bond length with positive and negative charges on CNTs is clearly seen, as anticipated from the first-principle model and the DFT simulations.

Given the superior mechanical properties, combined with the electromechanical couplings, one can envision a number of mechanical applications that make innovative use of the exceptional properties. In the passive mode of mechanical applications, surfaces that have exceptional adhesion and friction properties can be made of nanotube arrays, mimicking biologically inspired strategies, such as those found in lizard feet [36]. The nanotubes with diameters between 20 and 200 nm are ideally suited to mimic the nanohairs found in lizard feet. And, through the dry van der

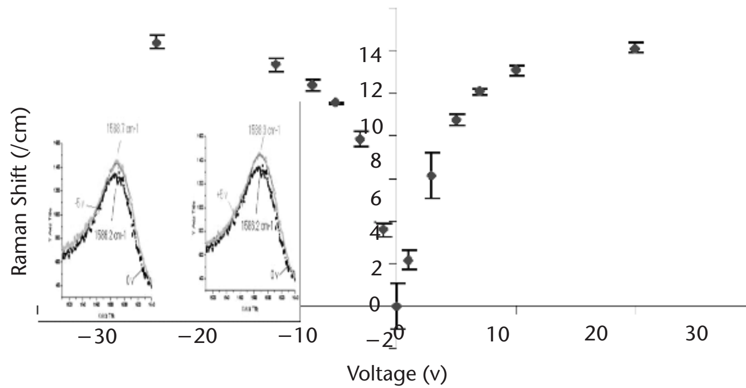


Figure 10.11 Raman G-band peak shift (bond length change) versus charging voltage (charge density).

Waals force, they can be reversibly attached to any surface and achieve sufficiently high adhesion and friction to support exceedingly large weight per unit area under harsh conditions.

The carbon nanotube-alumina matrix can be viewed and used as a reinforced ceramic composite [37]. Nanoindentation introduced controlled cracks and the damage has been examined by SEM. These composites are found to exhibit the three hallmarks of toughening in fiber composites: crack deflection at the CNT/matrix interface, crack bridging by CNTs, and CNT pullout on the fracture surfaces. These nanocomposites have potential for toughening and damage tolerance at submicron scales and hence are excellent candidates for wear-resistant coatings.

In the active mode, a strong electromechanical interaction has manifested itself beautifully in a clever experiment by Baughman et al. [38] in which two sheets of nanotube bundles were laminated into a double-layer film with a double-sized Scotch tape in between. Such a double-layer set in a cantilever configuration was then shown to bend to opposite directions in aqueous NdCl upon applying alternating bias voltage. The applied voltage was in the range of a few volts, whereas the operating voltage in a ferroelectric or piezoelectric actuator would be orders of magnitude higher for a similar level of displacement.

Light-activated mechanical deflection of nanotubes has been observed by Iijima's group under illumination of ordinary microscope light, and in the form of bundles [6].

In the high-frequency mode of operation, one can envision various forms of nanotube electromechanical oscillations. Oscillations of individual nanotubes have indeed been observed, for example, by de Heer's group [39] and in our own experiments. Even in normal SEM imaging processes, oscillations of nanotubes were often observed when the nanotubes are still embedded in the insulating alumina template. Such oscillations are likely a result of the charging of the nanotubes by the scanning

electron beams, providing yet another experimental support to the hypothesis of strong electromechanical couplings discussed earlier.

Naturally, these experimental observations of nanotube oscillations and the underlying physical mechanisms have inspired explorations of nanotube-based mechanical sensors and oscillators. One example of such explorations is the concept of a nanotube stereo-cilia device functioning as a "nano-ear." It is designed for sensing molecular motions via the acoustic waves induced by the molecular activities much in the same way as the stereo-cilia hair cells in human ears do. This idea has been followed by ongoing research and implementation efforts through a collaboration between Noca, Hunt, and Hoenk et al. (NASA JPL) and our group, and the subject was covered in some detail in a feature article by Daviss [40].

Rather than providing a parallel review of the basics of the same idea here, here the view is broadened to include efforts in an emerging field of much greater importance and scope. This field could be called nanoelectromechanical resonators and devices.

Generally speaking, anything that can be done in RF signal processing using electromagnetic waves can be done using acoustic waves, be it amplification, filtering, or modulation. But, why bother to use mechanical resonators and devices, especially when we have no shortage of electronic resonators and devices? One reason is that, in some cases, mechanical devices are intrinsically superior. A representative example is resonators. Mechanical resonators may have a Q-factor on the order of 10^4 or even higher because of the high-energy storage capacity, the low internal loss, and the low parasitic coupling to the surroundings. Empirically, one can easily tell this high value of Q from the ringing of bells and the superb sound of musical instruments. In contrast, the Q-factor of an electronic resonator, inevitably suffering from the R in the RLC circuitry, is usually lower by orders of magnitude.

Why, then, do we not see an acoustic computer on our desktop? The obvious answer is that it is hard to make acoustic devices that are as small as their electronic counterparts. Even if we could somehow manage to fabricate ultrasmall acoustic devices using modern lithographic techniques, we could hardly avoid incurring additional loss due to dangling bonds at the etched surfaces. This surface loss can be substantial in nanoscale structures because the surface-to-volume ratio scales up very rapidly with $1/r$ and is also very susceptible to the ambient conditions. Nanotubes, on the other hand, are naturally small and strong, and have few surface problems because of their nonreactivity.

Another fair question may be asked about the speed of operation. Here again, the picture may be prettier than we thought. The nanotube resonating frequency, f, is proportional to the product of D/L^2 and square root of E/ρ where E is the elastic modulus and ρ is the mass density, as explained by Poncharal et al. [39] in their paper on nanotube oscillation experiments. The nanotube's Young's modulus is known to be very high [41], as high as 1.8 TPa [18]. Without pushing for attaining any extraordinary structural parameters, one can comfortably obtain an oscillation frequency in the range of gigahertz with a typically sized nanotube, say, a multiple-walled nanotube of 30 nm in diameter and 500 nm in length. Experimentally, a high

Q-factor of more than 1,000 at an oscillating frequency of more than 2 GHz has been reported [42].

10.5 Extrinsic Coupling to Radiation Fields

In addition to the intrinsic electromechanical couplings examined earlier, extrinsic couplings to radiation fields could be and have been explored, for example, along two prominent directions: photodetection and emission.

Consider IR photodetection as a first example. Both theory and experiments have shown that nanotubes can be semiconducting with a bandgap that is inversely proportional to the tube diameter and thereby have an IR response over a very broad wavelength range from 1 to 15 μm (e.g., [9–13]). While an actual demonstration of nanotube IR detection is important and is the objective of some intensive ongoing efforts by our group and others, it is perhaps even more important at this early stage to gain a better understanding of the basic advantages and differentiating features of nanotubes for IR detection.

There are two possible modes of IR sensing—cooled band-to-band photocurrent sensing and uncooled thermal-electric sensing. In both cases, the highly ordered nanotube array technology offers intrinsic normal-incidence detection, large and scalable reception surface area, low thermal noise, and high resistance to defect-induced failure. In addition, it is conformal with curved surfaces.

The normal-incidence detection is permitted as a result of the tubular structure and vertical orientation in the ordered array, both of which break the in-plane translation symmetry. Nanotube array fabrication relies on the use of anodized aluminum film, which is naturally both scalable and conformal. Thermal noise is spatially random and can therefore be suppressed by spatially averaging it over highly ordered arrays of uniform nanotubes. The natural tube-to-tube isolation in the array and the small size of each nanotube minimizes the chance of defect-induced material failure that is often seen in conventional narrow-bandgap semiconductors. Although the peak wavelength responses of semiconductor QW and QD materials can be tuned over a broad range, it appears that the nanotube's wavelength response range is much broader [9–13]. In the case of SWNTs, the spectral responses are dependent on both diameter and chirality; in the case of multiwalled nanotubes, the spectral range can be tuned by adjusting a single structural diameter—the diameter. Likewise, the absorption efficiency can be easily tuned by adjusting the nanotube length. To cover the same spectral range using the existing technology platform, four or more different materials, and therefore different growth and processing technologies, would be required. In QW- or QD-based detection, stacking of multiple active layers is used to increase the absorption efficiency, but this strategy is limited by strain relaxation and defect formation. To achieve the same result using our arrayed nanotube platform, one can simply increase the nanotube length.

In the uncooled thermal-electric detection mode, the high thermal conductivity of nanotubes and the subwavelength spacing of parallel and highly absorptive nanotubes are also beneficial in that the former maximizes the temperature gradient at

the junction, while the latter gives rise to so-called antireflection due to photon trapping in the deep nanoscale cavity array. Experimentally, both FIR and visible light photoresponses have been observed in nanotube bundles, as well as in multiwalled nanotube arrays in uncooled and cooled configurations most recently [9–13, 43].

Although carbon nanotubes, like most semiconductors, do emit in the fluorescence mode, they are not known to be good optical emitters on their own [9–13]. The large aspect ratio and small radius of curvature of nanotubes, however, suggest that nanotubes can be good electron field emitters—for a given threshold field, the threshold voltage of field emission can be much smaller than that of planar carbon electrodes. The high thermal conductivity of these structures is also a major benefit for achieving a large-scale display panel. Indeed, a fully vacuum-sealed nanotube field emission display panel, measuring 5 inches along the diagonal, has been demonstrated recently. Its onset gate voltage for emission was as low as 60V, while its luminance was as high as 510 cd/m^2 [44].

An extension of the concept of coupling to radiation is the X-ray radiation generated by high-flux electrons emitted from a nanotube field emission cathode. A 28-mA current of energetic electrons from nanotube emitters hitting the Cu target produced sufficient X-ray intensity to permit the clear imaging of a human hand [45]. Yet another extension, being pursued by a group at NASA JPL and ourselves, is the electromagnetic wave radiation from the emitted electron beam in a klystron cavity [46].

10.6 Heterojunction Nanotubes

Up to this point, the nanotubes we have referred to are the native kind, or naturally occurring kind, as discovered by Iijima [1]. (To be more accurate, they are same as the ones rediscovered by Iijima; several groups apparently had found and reported on such nanotubes decades earlier, and one group's findings were published in the open literature in 1978–1979 [47, 48].) These are straight tubes of nanometer diameter, made of pure carbon in graphitic multiwalled or single-walled forms. Even with their simple structure, such native nanotubes have already been shown to enable many potential applications in many different ways. It is expected that with time they will only show more and better prospects.

Can there be other forms of nanotubes? What would we be able to do with different forms of nanotubes. How can they be made?

There can indeed be other forms of carbon nanotubes: structured nanotubes or engineered nanotubes. One can envision several of such forms—Y-shaped, T-shaped, tubes with variable diameter along their length, and nanotubes that are chemically or metallurgically modified. Many of these have already been demonstrated. Some were made as intended or as designed, while others were obtained by accident or were found while making other nanotubes, much like the discovery of the nanotube itself, which arose from fullerene growth [1].

Among the various forms of nanotubes the controllably engineered kind is obviously more attractive. One representative type is highlighted here as an example of

ongoing advances in two primary directions of evolution of the nanotubes into the next-generation-engineered nanotubes: Y-junction nanotubes.

Y-junction nanotubes, demonstrated in 1999 [3, 15], can be viewed as a larger tube branched into two structurally connected smaller nanotubes. They were first synthesized in a CVD process, using a nanopore alumina matrix as the growth template. In this pore-guided synthesis process, a Y-branched nanopore template is first formed using a three-step anodization process. The first two steps are the same as that used for forming the straight nanopores in anodized alumina [14]. At the third step, the anodization voltage is reduced from that of the first two steps to half or about half. As a result, the straight large nanopore bifurcates into two smaller pores. This is then followed by the normal nanotube growth procedure for the straight nanotubes. The formation mechanism of the Y-shaped pores and the Y-junction nanotube growth mechanism within them are subjects of significant interest, but not yet well understood. Time and further effort will help us map out the range of possible variations of the basic Y-tube structure and limitations.

However, early-stage results are already encouraging, as illustrated by the SEM image shown in Figure 10.12. Furthermore, their measured and modeled properties are intriguing and inviting. For example, we found that the Y-junction nanotube in the two-terminal contact configuration functions like a rectifying diode at room temperature. Given the inverse dependence of bandgap on diameter discussed earlier, this is not all that surprising—the smaller tubes have a larger bandgap and the larger tube has a smaller bandgap. Consequently, as a whole, the Y-structured tube is a heterojunction diode.

Given these features, one can readily envision a number of possible applications in both the non-CMOS and the CMOS-like regimes. For instance, it is possible that we could obtain a transistor function from the three-terminal contact configuration, although making reliable contacts with the three ports of the Y-junction nanotube individually has proven to be very challenging. At the engineering level, this should be viewed as just the beginning of nanotube structural engineering. Material quality at present is still much poorer than that of straight nanotubes, although the rectifying effects expected of a heterojunction are readily attainable over a large temperature range from 10 to 300K in more than half of the hundreds of samples we

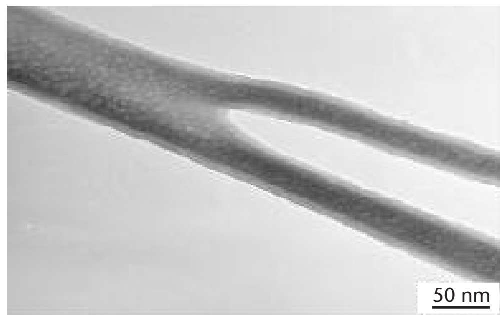


Figure 10.12 SEM image of a Y-junction nanotube, first reported in [3]. Smaller and larger diameter variations have since been made as well in our lab.

fabricated and measured. And, we found that annealing at 1,400°C in Ar has the same positive effects on the crystallinity in the straight portions of the Y-tube as shown earlier in Figure 10.3. But the crux region of the Y remains substantially more disordered, and seems to require higher temperature annealing.

At a more fundamental level, a number of interesting quantum transport issues have arisen. The Y structure lacks both the inversion symmetry and the axial symmetry of the straight nanotube. This means a change of the conditions defining the quasi-1D plasmons, a large change in the axial phonon modes, and a redistribution of the density of phonon states, with much less effect on the tangential modes. One can anticipate some of this easily with the breathing and twist modes. The axial modes are lower in energy than the stiffer tangential modes and are therefore much more effective in influencing the conductance arising from the transport of electrons at the Fermi surface. There are likely many more low-energy modes in the Y because of the addition of the "tuning fork" modes. Therefore, one can expect a lower phonon-scattering limited mobility. Given the likely larger impurity and defect concentrations, and therefore a lower impurity scattering limited mobility, it is possible that the two, with opposite temperature dependences, together bring the total mobility down and give the mobility an overall less peaked T dependence. The peak mobility could be further suppressed by the geometric scattering time in the Y-region. All of these possibilities are necessarily speculative at this early stage, and need to be validated by experiments and detailed theory.

The reliable fabrication and functionality of the Y-junction CNT arrays are application enabling, and the high-density 3D packing is particular inviting. Large-area, highly uniform active elements with redundancy built in and high packing density can be used to achieve desired device and circuit functions. For example, a logic device architecture based on Y-junction CNT arrays can be envisaged. By utilizing the rectifying behavior of individual Y-junctions, logic functions can be performed with proper wiring and several logic gates can then be connected in a single high-density matrix.

Although diode logic is not optimal compared to transistor-transistor logic designs, this represents but a first step in using the Y-junction diode matrix platform to create ultradense circuitry. Indeed, the necessary processing required to pattern fine wires onto such an array module has already begun to be developed along with methods to create densely packed FETs out of CNTs embedded in the alumina matrix.

In addition, the compatibility of the template-based nanotube array fabrication process with planar integration methodologies allows one to envision integration nanotube devices directly on a silicon wafer. This approach may offer a path toward effectively combining nanotube technologies with current semiconductor integrated circuit technology (Figure 10.13). The great versatility of CNTs would be a tremendous asset in this respect because their superior mechanical, electrical, magnetic, and so on properties could be employed in a heterogeneous fashion onto a monolithic chip design. Progress has been made on this front that has resulted in a highly uniform carbon nanotube-silicon heterojunction diode array. Moreover, an extremely

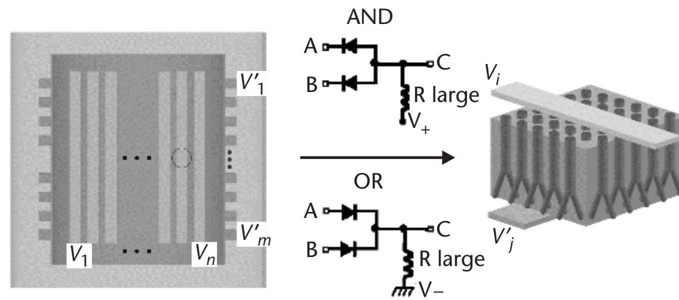


Figure 10.13 Logic AND and OR gate arrays formed by selectively connecting two or more Y-junctions in parallel. (*After:* [49].)

high ON/OFF ratio, as high as 10^6, is measured in this new type of heterojunction diode.

Again, the use of a highly ordered nanopore array template for the growth of the nanotubes was the enabling element. The fabrication process starts with the formation of the nanopore array alumina film on silicon by anodizing electron beam evaporated Al on doped silicon. A thin, typically ~5-nm Ti layer is used both to ensure good adhesion of the Al film during anodization and to enable the nanopores during the anodization to perforate through the alumina film. On completion of the anodization step, an array of uniform and hexagonally spaced nanopores in alumina is formed directly on the silicon surface, each of which is normal to the plane, of the same depth and diameter. The nanopore array serves as a template for nanotube growth. The growth process follows essentially the same steps in our earlier work on growing nanotube arrays on a free-standing Al plate, as detailed earlier. A typical array of carbon nanotubes grown this way directly on silicon is shown in Figure 10.14.

The electronic properties of the heterojunction between the carbon nanotube and the silicon were examined in details. To this end, a gold electrode film was evaporated on the top surface of the nanotube array at a 45° angle to avoid shorting through the nanotubes.

The *p*-type Si(100) wafer with a 0.01 $\Omega \cdot$ cm resistivity and thickness of 500 μm served as the bottom contact in the measurements. The measurements were performed in vacuum and in a cold finger closed cycle cryostat using source measure unit.

The measured *I-V* dependences at all temperatures exhibit well-pronounced rectifying behavior (Figure 10.14, bottom) with the forward direction at negatively biased CNT electrode achieving an ON/OFF ratio as high as 10^6 at 17K and 4V. The rectifying behavior and its polarity are in accordance with the data observed at the heterojunction of carbon nanotubes and silicon nanowires [50]. An estimation of the forward current at 4V shows the current trough of an individual CNT to be on the order of 10^{-11} A. This value is lower than that in the previous experiments on single nanotubes, showing that there is further potential for improvement of our devices.

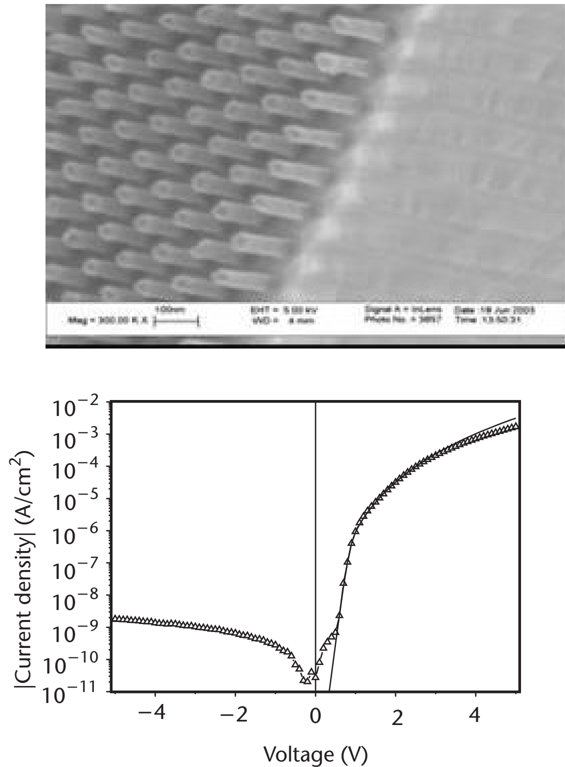


Figure 10.14 *Top*: SEM image of uniform nanotube array grown in an AAO template on Si. *Bottom:* Dark current-voltage characteristics of a CNT/Si array structure at 17K.

The realization of a CNT/Si heterojunction array is promising for reasons other than its candidacy for building logic structures. The 2D periodic lattice of the nanotubes and high refractive index difference relative to the alumina matrix can be used as photon bandgap material. The junction behavior in the orthogonal direction implies significant potential for application as an electro-optical modulator.

10.7 Prospects for Future Advances

Doping nanotubes is of great interest as well. Doping is what makes the conductivity of semiconductors controllably variable over many orders of magnitude. This ability is, in my view, what makes semiconductors so useful and unique. Through doping, we can change the semiconductor conductivity type from *n* to *p* as well. Conductivity change via doping has also been shown to be possible in carbon nanotubes. While pristine uncontaminated—and therefore undoped—nanotubes appear to be naturally *p*-type [51], methods to dope nanotubes into both types of carriers have been demonstrated, with *n*-type doping being somewhat more difficult [52–55].

Controlled doping is, of course, critically important for the task of creating CMOS-like nanotube FET cells. To stay within the scope of this chapter, the prospects of doping are viewed here from the point of its enabling potential for non-CMOS device and system explorations. Superconductivity might be one extreme example of many effects that could be enabled by doping. It has been suggested that superconductivity may exist in nanotubes and may arise from coupling between nanotubes bundled together in a rope. This intertube coupling would not be strong since the spacing is usually larger than the normal lattice constants, and the non-bonding coupling is much weaker than that of lattice bonding. However, strong coupling could exist between carbon atoms and impurity atoms such as catalyst particles, which can be as close to a carbon atom as the lattice constant of a 2D graphene sheet (~0.245 nm), and which can be naturally brought into existence by the usual catalytic deposition process; as such they can be difficult to remove [56]. The strong coupling between carbon atoms and transition metal atoms extends the dimensionality from 1D to quasi-2D or 3D. What is not so apparent but most intriguing is the possibility that these couplings can transform a semiconducting nanotube into superconducting [57]. This suggestion is supported by the findings from tight-binding modeling of the density of states and phonon-electron interactions. A strongly enhanced density of states at the Fermi surface by the carbon-transition metal coupling is theorized to enable a Cooper-pairing transition. An optimal doping level for this purpose is estimated to be around 3% to 5%. While superconductivity in the 10 to 50K range by itself may not be a new development, its presence in a wire-like nanometer-scale structure would be.

The very large surface-to-volume ratio and the easy access to the interior and exterior wall surfaces of the highly aligned arrays of uniform hollow carbon nanotubes with open ends make them suitable for hydrogen and Li ion storage [58, 59] and fuel cell electrodes [60, 61].

In the case of fuel cell electrodes, the very large electrochemical surface on both the exterior and interior walls of the carbon nanotube could be used to support catalyst, most of which is Pt based, for efficient conversion of fuel to electric current with fast oxidation kinetics. Pt electrode as a standard catalyst for oxidation of small organic molecules is active in the dehydrogenation step but suffers seriously from the so-called self-poisoning of the surface by reaction intermediates such as CO. As a consequence, oxidation on Pt in a normal flat electrode is possible only at potentials where adsorbed CO and other poisoning intermediates are effectively oxidized, resulting in an overpotential and loss in efficiency [62]. It has also been hypothesized that for a given current the greatly increased surface area would help reduce the self-poisoning effect. More definitive experiments validating this hypothesis are needed. However, loading of metal particles and clusters into the nanotubes has been proven beyond doubt [63]. The process of loading normally begins with immersion of the nanotube array in the appropriate acids such as H_2PtCl_6(aq) or peroxo-tungstic acid for 12 to 48 hours, followed by drying in the air and reduction at 650 to 850K for 3 to 5 hours in H_2 gas. To dissolve away the alumina membrane, one can immerse the alumina membrane with the metal loaded nanotube array in 48% HF for 24 hours.

Acknowledgments

The work of our lab reviewed in this chapter was made possible by support from AFOSR, AFRL, CIAR, DARPA, Motorola, NASA, Nortel, NSF, and ONR, and by contributions from our former and current lab members, in particular: Jack Bennett, Hope Chik, Adam Hartman, Nickolai Kouklin, Adam Lazareck, Jing Li, Jianyu Liang, Maryam Jouzi, Chris Papadopoulos, Andrei Rakitin, Dan Straus, Marian Tzolov, Gary Withey, and Nick Yin.

References

[1] Iijima, S., "Helical Microtubules of Graphitic Carbon," *Nature,* Vol. 354, 1991, pp. 56–58.

[2] Dresselhaus, M. S., G. Dresselhaus, and P. C. Eklund, *Science of Fullerenes and Carbon Nanotubes,* New York: Academic Press, 1996.

[3] Li, J., C. Papadopoulos, and J. M. Xu, "Y-Junction Carbon Nanotubes and Controlled Growth," *Nature,* Vol. 402, No. 6759, 1999, pp. 253–254.

[4] Krstic, V., et al., "Role of Disordering on Transport in Boron-Doped Multiwalled Carbon Nanotubes," *Physical Review B*, Vol. 67, No. 4, 2003, pp. 041401r (1–4).

[5] Hone, J., et al., "Quantized Phonon Spectrum of Single-Wall Carbon Nanotubes," *Science,* Vol. 289, No. 5485, 2000, pp. 1730–1732.

[6] Zhang, Y., and S. Iijima, "Elastic Response of Carbon Nanotube Bundles to Visible Light," *Physical Review Letters,* Vol. 82, No. 17, 1999, pp. 3472–3475. Also seen in individual nanotubes' oscillations, similar to that shown in Figure 10.9.

[7] Baughman, R. H., A. A. Zakhidov, and W. A. De Heer, "Carbon NanotubesThe Route Toward Applications," *Science,* Vol. 297, No. 5582, 2002, pp. 787–792.

[8] Rakitin, A., C. Papadopoulos, and J. M. Xu, "Electronic Properties of Amorphous Carbon Nanotubes," *Physical Review B*, Vol. 61, No. 8, 2000, pp. 5793–5796.

[9] Wong, S. S., et al., "Covalently Functionalized Nanotubes as Nanometer Sized Probes in Chemistry and Biology," *Nature,* Vol. 394, No. 6688, 1998, pp. 52–55.

[10] Xu, J. M., "Highly Ordered Carbon Nanotube Arrays and Ir Detection," *Infrared Physics and Technology*, Vol. 42, No. 3–5, 2001, pp. 485–491.

[11] Shyu, F. L., et al., "Magnetoelectronic and Optical Properties of Carbon Nanotubes," *Physical Review B,* Vol. 67, No. 4, pp. 045405 (1–9).

[12] Liu, X., et al., "Detailed Analysis of the Mean Diameter and Diameter Distributions of Single-Wall Nanotubes from Their Optic Response," *Physical Review B*, Vol. 66, No. 4, 2002, pp. 045411 (1–8).

[13] Barchilo, S. M., et al., "Structure-Assigned Optical Spectra of Single-Walled Nanotubes," *Scienceexpress*, November 22, 2002.

[14] Li, J., et al., "Highly Ordered Carbon Nanotube Arrays for Electronics Applications," *Applied Physics Letters,* Vol. 75, No. 3, 1999, pp. 367–369.

[15] Papadopoulos, C., et al., "Electronic Transport in Y–Junction Carbon Nanotubes," *Physical Review Letters,* Vol. 85, No. 16, 2000, pp. 3476–3479.

[16] O'Sullivan, J. P., and G. C. Wood, "Electron-Optical Examination of Sealed Anodic Alumina Films: Surface and Interior Effects," *J. Electrochemical Society,* Vol. 116, 1969, pp. 1351–1357.

[17] Masuda, H., and K. Fukuda, "Ordered Metal Nanohole Arrays Made by a Two-Step Replication of Honeycomb Structures of Anodic Alumina," *Science,* Vol. 268, No. 5216, 1995, pp. 1466–1468.

[18] Wong, E. W., P. E. Sheehan, and C. M. Lieber, "Nanobeam Mechanisms: Elasticity, Strength, Toughness of Nanorods and Nanotubes," *Science,* Vol. 277, No. 5334, 1997, pp. 1971–1975.

[19] Hiura, H., T. W. Ebbesen, and K. Tanigaki, "Opening and Purification of Carbon Nanotubes in High Yields," *Advanced Materials,* Vol. 7, 1995, pp. 275–276.

[20] Yin, A., H. Chik, and J. M. Xu, "Post-Growth Processing of Carbon Nanotube Arrays—Enabling New Functionalities and Applications," *IEEE Trans. on Nanotechnology*, 2004.

[21] Withey, G., et al., "Nanoscale Direct Conversion of Bio-Molecular Activities to Digital Electronic Information," *Nature Biotechnology*, 2004.

[22] Guiseppi-Elie, A., C. Lei, and R. H. Baughman, "Direct Electron Transfer of Glucose Oxidase on Carbon Nanotubes," *Nanotechnology,* Vol. 13, No. 5, 2002, pp. 559–564.

[23] Rakitin, A., et al., "Metallic Conduction Through Engineered DNA: DNA Nanoelectronic Building Block," *Physical Review Letters,* Vol. 86, No. 16, 2001, pp. 3670–3673.

[24] Hernadi, K., et al., "Reactivity of Different Kinds of Carbon During Oxidative Purification of Catalytically Prepared Carbon Nanotubes," *Solid State Ionics,* Vol. 141–142, 2001, pp. 203–209.

[25] Wong, S. S., et al., "Covalently Functionalized Nanotubes as Nanometer Sized Probes for Biology and Chemistry," *J. American Chemical Society,* Vol. 120, No. 33, 1998, pp. 8557–8558.

[26] Rakitin, A., C. Papadopoulos, and J. M. Xu, "Electronic Properties of Amorphous Carbon Nanotubes," *Physical Review B*, Vol. 61, 2000, pp. 5793–5796.

[27] Blase, X., et. al., "Hybridization Effects and Metallicity in Small Radius Carbon Nanotubes," *Physical Review Letters,* Vol. 72, No. 12, 1994, pp. 1878–1881.

[28] Hohenberg, P., and W. Kohn, "Inhomogeneous Electron Gas," *Physical Review,* Vol. 136, No. 3B, 1964, pp. B864–B871;

[29] Kohn, W., and L. J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects," *Physical Review,* Vol.140, No. 4A, 1965, pp. A1133–A1138.

[30] Hartman, A., M. Jouzi, and J. M. Xu, *Physical Review Letters*, 2004.

[31] Kane, C. L., and E. J. Merle, "Size, Shape, and Low Energy Electronic Structure of Carbon Nanotubes," *Physical Review Letters,* Vol. 78, No. 10, 1997, pp. 1932–1935.

[32] Garstein, Y. N., et al., "Charge-Induced Anisotropic Distortions of Semiconducting and Metallic Carbon Nanotubes," *Physical Review Letters,* Vol. 89, No. 4, 2002, pp. 045503 (1–4).

[33] Verissimo-Alvez, M., et al., "Electromechanical Effects in Carbon Nanotubes: *Ab Initio* and Analytical Tight-Binding Calculations," *Physical Review B,* Vol. 67, No. 16, 2003, pp. 161401r (1–4).

[34] Stich, I., et al., "Conjugate Gradient Minimization of the Energy Functional: A New Method for Electronic Structure Calculation," *Physical Review B,* Vol. 39, No. 8, 1989, pp. 4997–5004.

[35] Zólyomi, V., et. al., "Origin of the Fine Structure of the Raman D Band in Single-Wall Carbon Nanotubes," *Physical Review Letters,* Vol. 90, No. 15, 2003, pp. 157401 (1–4);

Saito, R., et al., "Finite-Size Effect on the Raman Spectra of Carbon Nanotubes," *Physical Review B,* Vol. 59, No. 3, 1999, pp. 2388–2392.

[36] Autumn, K., et al., "Adhesive Force of a Single Gecko Foot-Hair," *Nature,* Vol. 405, No. 6797, 2000, pp. 681–685.

[37] Xia, Z., "Direct Observation of Toughening Mechanisms of Carbon Nanotube Alumina Ceramic Composite," *Acta Materialia*, Vol. 52, 2004, pp. 931–944.

[38] Baughman, R. H., et al., "Carbon Nanotube Actuator," *Science,* Vol. 284, No. 5418, 1999, pp. 1340–1344.

[39] Poncharal, P., et al., “Electrostatic Deflections and Electromechanical Resonances of Carbon Nanotubes,” *Science,* Vol. 283, No. 5407, 1999, pp. 1513–1516.

[40] Daviss, B., “Snap, Crackle and Pop,” *New Scientist*, July 21, 2001, p. 3434.

[41] Treacy, M. M. J., T. W. Ebbesen, and J. M. Gibson, “Exceptionally High Young’s Modulus Observed for Individual Carbon Nanotubes,” *Nature,* Vol. 381, No. 6584, 1996, pp. 678–680.

[42] Reulet, B., et al., “Acoustoelectric Effects in Carbon Nanotubes,” *Physical Review Letters,* Vol. 85, No. 13, 2000, pp. 2829–2832.

[43] Ugawa, A., A. G. Rinzler, and D. B. Tanner, “Far Infrared Gaps in Single-Wall Carbon Nanotubes,” *Physical Review B,* Vol. 60, No. R11, 1999, pp. 305–308.

[44] Chung, D. S., et al., “Carbon Nanotube Electron Emitters with a Gated Structure Using Backside Exposure Processes,” *Applied Physics Letters,* Vol. 80, No. 21, 2002, pp. 4045–4047.

[45] Yue, G. Z., et al., “Generation of Continuous and Pulsed Diagnostic Imaging X-Ray Radiation Using a Carbon-Nanotube-Based Field-Emission Cathode,” *Applied Physics Letters,* Vol. 81, No. 2, 2002, pp. 355–357.

[46] Siegel, P. H., T. H. Lee, and J. Xu, “The Nanoklystron: A New Concept for THz Power Generation,” *JPL New Technology Report,* NPO 21014, March 21, 2000.

[47] Wiles, P. G., and J. Abrahamson, “Carbon Fibre Layers on Arc ElectrodesI. Their Properties and Cool-Down Behaviour,” *Carbon,* Vol. 16, No. 5, 1978, pp. 341–349;

[48] Abrahamson, J., P. G. Wiles, and B. L. Rhoades, “Structure of Carbon Fibres Found on Carbon Arc Anodes,” *Carbon,* Vol. 37, No. 11, 1999, pp. 1873–1874.

[49] Papadopoulos, C., Ph.D. thesis, Brown University, 2003.

[50] Hu, J., et al., “Controlled Growth and Electrical Properties of Heterojunctions of Carbon Nanotubes and Silicon Nanowires,” *Nature,* Vol. 399, No. 6731, 1999, pp. 48–51.

[51] Rakitin, A., C. Papadopoulos, and J. M. Xu, “Carbon Nanotube Self-Doping,” *Physical Review B,* Vol. 67, No. 3, 2003, pp. 033411 (1–4).

[52] Tans, S., A. Verschueren, and C. Dekker, “Room-Temperature Transistor Based on a Single Carbon Nanotube,” *Nature,* Vol. 393, No. 6680, 1998, pp. 49–52.

[53] Martel, R., et al., “Single- and Multi-Wall Carbon Nanotube Field-Effect Transistors,” *Applied Physics Letters,* Vol. 73, No. 17, 1998, pp. 2447–2449.

[54] Martel, R., et al., “Carbon Nanotube Field Effect Transistors for Logic Applications” *Proc. Int. Electron Devices Meeting,* 2001, pp. 159–163.

[55] Martel, R., et al., “Ambipolar Electrical Transport in Semiconducting Single-Wall Carbon Nanotubes,” *Physical Review Letters,* Vol. 87, No. 25, 2001, pp. 256805 (1–4).

[56] Dujardin, E., et al., “Interstitial Metallic Residues in Purified Single Shell Carbon Nanotubes,” *Solid State Communications,* Vol. 114, No. 10, 2000, pp. 543–546.

[57] Schluter, M., et al., “Electron-Phonon Coupling and Superconductivity in Alkali-Intercalated C60 Solid,” *Physical Review Letters,* Vol. 68, No. 4, 1992, pp. 526–529.

[58] Dillon, A. C., et al., “Storage of Hydrogen in Single-Walled Carbon Nanotubes,” *Nature,* Vol. 386, No. 6623, 1997, pp. 377–379.

[59] Baum, R. M., *Chemical and Engineering News,* Vol. 75, 1997, pp. 39–41.

[60] Freemantle, M., *Chemical and Engineering News,* Vol. 74, 1996, p. 62.

[61] Rodriguez, N. M., M.-S. Kim, and R. T. K. Baker, “Carbon Nanofibers: A Unique Catalyst Support Medium,” *J. Physical Chemistry,* Vol. 98, No. 50, 1994, pp. 13108–13111.

[62] Witham, C. K., et al., “Performance of Direct Methanol Fuel Cells with Sputter-Deposited Anode Catalyst Layer,” *Electrochemical and Solid-State Letters*, Vol. 3, No. 11, 2000, pp. 497–500.

[63] Che, G., et al., “Metal-Nanocluster-Filled Carbon Nanotubes: Catalytic Properties and Possible Applications in Electrochemical Energy Storage and Production,” *Langmuir,* Vol. 15, No. 3, 1999, pp. 750–758.

Acronyms

AAO	anodized aluminum oxide
AES	Auger Electron Spectroscopy
AFM	Atomic Force Microscope
ALE	atomic layer epitaxy
CL	cathololuminescence
CMOS	complimentary metal-oxide-semiconductor
CNT	carbon nanotube
CVD	chemical vapor deposition
CZ	Czochralski crystal growth method
DFT	density functional theory
DH	double heterostructure
EL	electroluminescence
FET	field effect transistor
FTIR	Fourier transform infrared spectroscopy
FWHM	full width at half maximum
FZ	float zone crystal growth method
GSMBE	gas source molecular beam epitaxy
HRTEM	high-resolution transmission electron microscope
LCC	light-current characteristic
LD	laser diode
LO	longitudinal optical phonon
LPE	liquid phase epitaxy
LWIR	long wave infrared (8–12 microns)
MBE	molecular beam epitaxy
MCT	mercury cadmium telluride, HgCdTe
MEE	migration enhanced epitaxy

MIR	mid infrared, also MWIR (2-5 microns)
ML	monolayer
MOCVD	metal organic chemical vapor deposition
MOMBE	metal organic molecular beam epitaxy
MQD	multiple quantum dot
MQW	multiple quantum well
MWIR	medium-wave infrared (2–5 microns)
NIR	near infrared
OCL	optical confinement layer
PC	photoconductivity spectroscopy
PECVD	plasma enhanced chemical vapor deposition
PIG	punctuated island growth
PL	photoluminescence spectroscopy
PLE	photoluminescence exitation spectroscopy
QCL	quantum cascade laser
QD	quantum dot
QDIP	quantum dot infrared photodetector
QW	quantum well
QWIP	quantum well infrared photodetector
QWR	quantum wire
RHEED	reflection high-energy electron diffraction
RIE	reactive ion etching
SAQD	self-assembled quantum dots
SCH	separate confinement heterostructure
SEG	selective epitaxial growth
SEM	scanning electron microscope
SK	Stranski–Krastinov growth method
SL	superlattice
SLS	strained layer superlattice
SWNT	single-wall nanotube
TEM	transmission electron microscope
TO	transverse optical phonon
VCSEL	vertical cavity surface emitting laser

VLWIR very long wavelength infrared (16–25 microns)

VPE vapor phase epitaxy

XTEM cross-sectional transmission electron microscope

About the Editor

Todd Steiner earned his Ph.D. in optical properties of semiconductor superlattices and heterostructures from the Air Force Institute of Technology at Wright Patterson Air Force Base in Ohio. He has been an officer in the U.S. Air Force for 22 years and has held a variety of positions involving the management and execution of science and technology within the Air Force Research Laboratory, including positions working in the areas of laser radar, infrared materials and sensors, laser sources, and directed energy. Dr. Steiner currently is a program manager in the Physics and Electronics Directorate of the Air Force Office of Scientific Research, where he manages a basic research portfolio of more than 50 projects in semiconductor materials and devices, with a particular emphasis on optoelectronics. He is a member of the American Physical Society, the Materials Research Society, the Electrochemical Society, and the Society of Photo-Optical and Instrumentation Engineers. He is the author of more than 100 publications and technical reports over a wide variety of topics in the physical sciences.

Index

1D ZnO nanostructures, 189–91
  catalyst-assisted MBE, 208–9
  catalyst-free self-nucleation mechanism, 193
  characterizations, 211–19
  CVD, 199–200
  device applications, 219–24
  electronic devices, 221–24
  growth, 191–211
  growth mechanisms, 191–94
  growth summary, 210
  growth techniques, 194–210
  MOCVD, 200–208
  optical characterizations, 215–19
  optical devices, 219–21
  screw dislocation mechanism, 192–93
  structural characterizations, 211–15
  template-assisted growth, 209–10
  vapor phase transport (VPT), 195–99
  VLS mechanism, 191–92
  VS mechanism, 193–94
  See also Semiconductor nanostructures
III-nitride quantum dots, 289–344
  GaN QDs, 292–301
  growth of, 291–317
  InGaN QDs, 301–7
  introduction, 289–91
  MBE growth of, 292–307
  MOCVD growth, 307–14
  optical properties, 317–43
  polarization effects, 318–23
  progress, 344
  properties, 290
  quantum confinement effects, 318–23
  self-assembled growth, 291
  strained systems, 290
  strain effects, 318–23
  summary, 343–44
III-V binary compounds, 231

A

AlGaAs blocking layer QDIP, 87–92
  dark current density, 90
  schematic cross section, 89
  spectral response, 89
AlGaAs capped QDIP structures, 87
AlSb, 239–42
  defined, 239
  electronic properties, 240
  optical properties, 240–42
  physical properties, 241
  properties, 240–42
  spectral dependence, 241
  See also Sb-III binary compounds
Anodic alumina membrane (AAM), 209
Anodized aluminum oxide (AAO), 374
Antimony-based materials, 229–85
  III-V semiconductor allows, 230–32
  AlSb, 239–42
  applications, 232–34
  bulk single-crystal growth, 232
  GaSb, 235–39
  InAsSb, 250–59
  InAsSb/InAsSbP, 267–73
  InSb, 242–50
  InSbBi, 262–66
  InTlAsSb, 266–67
  InTlSb, 259–62
  introduction, 229–34
  quaternary compounds, 231
Antimony (Sb), 229–39
  compounds, 230
  defined, 229
  physical data, 230
  reactions, 229
  use of, 229
Atomic force microscope (AFM), 49
  of 2D arrangement of Ge dots, 359
  conductive-tip (C-AFM), 206

Atomic force microscope (continued)
- of dots on planar substrates, 353
- of GaN buffer layer, 309
- of GaN QDs, 293, 304, 305, 306
- of Ge QD sample, 352
- of InGaN self-assembled QDs, 309, 310
- of QD formation on Si mesa, 361
- of self-assembled Ge dots, 358
- of self-organized nanospore, 374
- smooth AIN surface, 294
- topographical image, 302

Atomic layer epitaxy (ALE), 1, 29
Auger coefficients, 180
- for SCH QD laser, 181
- variation with temperature, 181

Auger electron spectroscopy (AES), 22, 23, 259
Auger recombination, 179–83
- damping factor and, 179–80
- decrease, 183
- defined, 179
- rates, 180
- schematic, 182

Avalanche effect, 30

B

Bandgap-engineered QD laser, 150
- defined, 150
- illustrated, 151
- See also Quantum dot lasers

Beam-equivalent pressure (BEP) ratio, 282
Bimodal distribution, 350
Bimodal growth, 350
Binary compounds, 3
Boules
- Bridgman method, 12
- Czochralski (CZ) method, 10
- defined, 10

Bridgman method, 11–12
- boules, 12
- defined, 11
- disadvantages, 12
- illustrated, 12
- See also Bulk crystal growth techniques

Bulk crystal growth techniques, 5, 8–16
- Bridgman method, 11–12
- Czochralski (CZ) method, 8–11
- float-zone (FZ) method, 13–14
- Lely growth method, 14–16

C

Carbon nanotubes, 371–98
- applications, 373
- arrays, 375, 378
- controlled fabrication, 371, 373–79
- crack bridging by, 389
- development, 372
- electromechanical couplings, 372
- engineered, 392
- extrinsic coupling to radiation fields, 391–92
- fabrication, 371, 373–79
- future advances, 396–97
- as graphite sheet, 387
- growing of, 378
- heterojunction, 392–96
- in highly ordered array, 373–79
- intrinsic quantum electromechanical couplings, 382–91
- introduction, 371–73
- molecular and cellular interface technology, 379
- optical emission and, 392
- oscillations, 389
- properties, 372
- Raman modes in, 386
- static charges on, 388
- structured, 392
- T-shaped, 392
- versatility, 394
- Y-junction, 392, 393–94

Catalyst-assisted MBE, 208–9
Catalyst-free self-nucleation mechanism, 193
Characteristic temperature
- obtained, 131
- parasitic recombination effect on, 127–28
- temperature dependence of, 128

Chemical reactions, 7
Chemical vapor deposition (CVD), 27
- thermodynamics of, 28
- ZnO nanostructure, 199–200
- See also Metalorganic chemical vapor deposition (MOCVD)

Chemical vapor transport and condensation (CVTC), 191
- growth, 196
- schematic diagram, 196

Chirp, 178–79
- defined, 178
- illustrated, 179

measured, plots, 179
CMOS devices, 372, 373
Conductive-tip atomic force microscopy (C-AFM), 206
Confinement energy
defined, 318
of GaN QDs, 319
Continual azimuthal rotation (CAR), 20
Cooperative arrangement of self-assembled islands (CASAD), 356
Critical thickness, 349
Czochralski (CZ) method, 8–11
defined, 8
furnace cross section, 10
growth of GaAs with, 11
growth process, 10–11
high-purity quartz, 10
impurities, 11
liquid encapsulated (LEC), 11
popularity, 8–10
See also Bulk crystal growth techniques

D

Damping factor, 176
Auger recombination and, 179–80
defined, 176
Dark current, 81–82
AlGaAs blocking layer QDIP, 90
background photocurrent and, 83
detectivity, 91
dual-color QDIPs, 106
InAs/InGaAs QDIP, 95, 96
originations, 81
peak responsivity, 91
voltage characteristics vs., 82
dc characteristics, 172–74
room temperature, 172
temperature dependent, 172–74
Dc sputtering, 34
Density functional theory (DFT), 384
Dielectric mirror reflectivity, 109
Differential quantum efficiency, 173
Differential transmission spectroscopy (DTS), 165
measurement principle schematic, 165
pump-probe measurements, 165
three-pulse time scan, 167
Distributed feedback (DFB), 152
Doping
controlled, 397
FZ, 14
InAsSb/InAsSbP, 269
InSb, 249–50
nanotubes, 396
N-type, 249
P-type, 249–50
Double heterostructure (DH) MQW, 234
Dual-color QDIPs, 102–7
dark current, 106
devices, 102–3
measured noise current, 107
peak responsivity, 105
See also Quantum dot infrared photodetectors (QDIPs)
Dynamic RHEED, 22

E

Electric field gradient microscopy (EFM), 214
Electrochemical ALE (ECALE), 29
Electron beam evaporation, 32, 33
Electron cyclotron resonance (ECR), 31
Electron-hole-plasma (EHP) transition, 189
Electronic devices, 221–24
field effect transistors (FETs), 223–24
field emission devices (FEDs), 221–23
See also ZnO nanostructures
Energy dispersive X-ray (EDX), 263
Envelope function approximation method, 274
Epitaxial growth techniques, 5, 16–29
atomic layer epitaxy (ALE), 29
liquid phase epitaxy (LPE), 5, 16–17
metalorganic chemical vapor deposition (MOCVD), 1, 6, 24–29
molecular beam epitaxy (MBE), 6, 20–24
vapor phase epitaxy (VPE), 6, 17–19
Excited states, 131–32
effect suppression, 132
presence of, 131
transition effects, 131
transition illustration, 132
Extrinsic coupling, 391–92

F

Fast Fourier transform (FFT) spectrum analyzer, 80
Field effect transistors (FETs), 223–24
Field emission devices (FEDs), 221–23
defined, 221
ZnO nanoneedles, 222, 223
Field emission scanning electron microscope (FESEM), 197

Filament evaporation, 32
Flash hot plate method, 32
Float-zone (FZ) method, 13–14
  defined, 13
  disadvantage, 14
  doping techniques, 14
  furnace cross section, 13
  See also Bulk crystal growth techniques
Focal plane array (FPA), 280
Fourier transform infrared spectrometer (FTIR), 54
  frequency measurement, 256
  spectrometer, 256
Frank-Van der Merwe growth, 349
Free-energy function, 6
Full width at half maximum (FWHM), 51, 62, 201

G

GaAs
  growth with Czochralski method, 11
  large capped, energy-level schematic, 73
  MOCVD growth of, 26
  quantum dots, 37
  VCSELs, 152
Ga face, 325
Gain
  maximum, dependence of, 142
  nonuniformity effect on, 118–19
  peak value, 119
  saturation effect, 119
  spectra for equilibrium filling of quantum dots, 118
GaN
  bandgap, 325
  crystalline particles, 316
  as grown, dots, 315
  matrix, 340
  nanocrystalline, fabrication of, 316
  nanocrystalline, PL spectra, 323
  plate, schematic diagram, 322
  quantum structures, 330
  VPE growth of, 19
  wide bandgap, 19
GaN QDs, 292–301
  AFM images, 293, 304, 305, 306
  arrays, 295
  bandgap, 323
  confinement energy, 319
  dependence, 304
  from depositing GaN monolayers, 294
  embedded in AIN matrix, 330
  emission spectra, 332
  excitations, binding energy, 333
  growth of, 301
  linear alignment, 296
  micro-Raman spectra, 334
  optical properties, 323–37
  perfectly coherent, 316
  photoluminescence spectra, 331
  PL decay time, 334
  PL spectra, 324
  PL spectra, temporal evolution, 329
  preferential alignment, 299
  schematic, 306
  self-arranged, 300
  self-assembled, optical properties, 329
  self-organized growth, 303
  superlattice, 296
  TEM images, 317
  temperature dependency, 331
  thermal stability, 331
  time-resolved PL spectra, 330
  weak-beam image of cross section, 297
  wurtzite, 336
  See also Quantum dots (QDs)
GaSb, 235–39
  band structure, 238
  defined, 235
  electronic properties, 238–39
  growth, 235–38
  LPE, 236–37
  MBE, 237–38
  MOCVD, 237
  optical constants, 240
  optical properties, 239
  phase diagram, 236
  physical properties, 235
  properties, 238–39
  structural properties, 238
  See also Sb-III binary compounds
GaSb/InAs type II superlattice, 273–84
  band structure, 275
  experimental results, 275–84
  introduction, 273–75
  mini band energy profile, 274
  photoconductors in LWIR range, 275–79
  photoconductors in VLWIR range, 279–80
  photodiodes in LWIR range, 280–81
  photodiodes in VLWIR range, 282–84

Gas source MBE (GSMBE), 20
Ge
  bimodal, 350, 351
  deposition, 350, 362
  deposition amounts, 359
  dots formation, 360
  growth of, 350
  lasers, 365
  lattice constant, 350
  multilayered quantum dots, 364
  nucleation, 360
  thickness, 359, 362
  trimodal, 351
Ge islands, 3, 349–67
  arrangement of, 357
  device applications, 362–67
  distribution, 352
  domes, 353, 354
  electronics applications, 266
  formation, 352, 353
  with monomodal distribution, 356
  optoelectronics, 362–65
  pyramids, 353, 354
  pyramids to domes ratio, 354
  quantum information applications, 366–67
  regimentation, 355–62
  registration, 355–62
  on Si substrate, 354
  temperature growth and, 358
  thermoelectricity, 365–66
  uniform, 350–54
Generation-recombination (GR) noise, 83
Ge QDs, 362–67
  arranged, 366
  for electronic applications, 366
  photodetectors, 362
  for quantum computing, 367
  on silicon template, 362
  superlattices, 365
  See also Quantum dots (QDs)
Gibb's phase rule, 8
Glucose oxidase (GOx), 380, 381
Ground-signal-ground (GSG) configuration, 174–75
Growth of nanostructures, 1, 34–41

H

Hall mobility, 247, 260, 276
Heteroepitaxy, 16
Heterojunction nanotubes, 392–96
  I-V dependencies, 395
  electronic properties, 395
  realization, 396
  See also Carbon nanotubes; nanotubes
High-angle annular dark field (HAADF), 213
High-detectivity QDIPs, 99–102
  dark current, 101
  dark noise current, 101
  peak detectivity, 100, 102
  peak responsivity, 100
High-resolution cross-sectional transmission electro microscopy (HR-XTEM), 161
High-resolution TEM (HRTEM), 211, 212, 213
  image of pseudotwin plane, 213
  image of ZnO nanowire, 213
  images for InGaN QW cross section, 308
  uses, 212
  See also TEM
High-speed quantum dot lasers, 159–83
  advantages, 160
  Auger recombination, 179–83
  characteristics, 172–83
  fabrication challenge, 159
  introduction, 159–60
  modulation characteristics, 174–83
  room temperature dc characteristics, 172
  SCH, 163–68
  temperature-dependent dc characteristics, 172–74
  tunneling injection, 172–83
  tunneling injection of carriers in, 168–71
  See also Quantum dot lasers
Hole burning, 132–34
Homoepitaxy, 16

I

InAs
  MQD region, 65
  QD-based QDIPs, 92–99
InAs/GaAs
  large capped QDIP structures, 70–74
  SAQDs, 50–57
  unintentionally doped QDIPs, 77–87
InAs/InGaAs QDIP, 92–99
  conduction band structure, 93
  dark-current density, 95, 96
  internal and external quantum efficiency, 98
  noise current vs. bias, 97
  peak responsivities, 94

InAs/InGaAs QDIP (continued)
schematic cross section, 93
spectral response, 93
See also Quantum dot infrared photodetectors (QDIPs)
InAsSb, 250–59
bandgap, 250–51
characterizations, 253–56
device measurement, 256–59
effective masses, 251
electrical characteristics, 254–55
growth of, 253
intrinsic carrier concentration, 251–52
MBE, 253
MOCVD, 253
optical characteristics, 255–56
photoconductors, 256–57
photodiodes, 257–58
photodiodes by MBE, 258–59
physical properties, 250–52
structural characteristics, 253–54
transmission spectra, 255
XRD peaks, 254
See also Antimony-based materials
InAsSb/InAsSbP, 267–73
characterization, 268–69
device results, 271–73
DH lasers, 271
doping, 269
growth, 267, 268–69
physical properties, 268
SLS lasers, 271–72
strained-layer superlattices, 269–70
Inductively coupled plasma (ICP), 202
InGaAs
large capped, energy-level schematic, 73
QDIPs, 68, 69
In(Ga)As QDs, 160–63
electronic spectra, 161–63
self-organized growth of, 160–61
See also Quantum dots (QDs)
InGaN, as quantum material, 337
InGaN QDs, 38, 301–14
AFM images of, 309, 310
confinement of excitons, 341
cylindrical, 337, 338
Gaussian curve-fitting PL, 311
high-resolution micro-PL spectrum, 342
HRTEM images, 308
maximum blue shift of PL peak, 343
MBE growth, 301–7
MOCVD growth, 307–14
nanoscale, 311
optical properties, 337–43
PL peaks, 339
PL spectra, 340
PL spectra comparison, 312
polarization, 339
polarization-induced electric field, 341
quantum confinement, 339
room temperature PL spectra, 340
schematic, 313
by selective growth, 307
by self-assembly, 307
strain, 339
subband emission, 341
surface-emitting lasers, 343
surface morphology, 312
thermal stability, 341
thickness, 311
on top of GaN pyramids, 313
See also Quantum dots (QDs)
Inhomogeneous line broadening, 117
based by QD size fluctuations, 119
illustrated, 117
See also Nonuniformity
InSb, 242–50
on (111)B GaAs, 246–47
characteristics, 242
defined, 242
doping characteristics, 249–50
electrical characteristics, 247–48
growth of, 245–47
material parameters, 242
MBE, 246
MOCVD, 245–46
N-type doping, 249
optical characteristics, 249
photodetectors, 250
physical properties, 244
P-type doping, 249–50
structure characteristics, 247
temperature-composition binary phase diagram, 243
uses, 243
See also Sb-III binary compounds
InSbBi, 262–66
absorption coefficient, 264
defined, 262
electrical characteristics, 264–65

MOCVD growth, 262–65
optical characteristics, 263–64
photodetectors, 265–66
structural characteristics, 263
XRD, 263
See also Antimony-based materials
Internal optical loss, 143–47
Internal quantum efficiency, 124–25
differential, 136
in OCL, 136
parasitic recombination effect on, 128–29
InTlAsSb, 266–67
electrical properties, 267
growth parameters, 266
See also Antimony-based materials
InTlSb, 259–62
chemical analysis by AES, 259
electrical characteristics, 260–61
MOCVD growth, 259, 260
optical characteristics, 261–62
photodetectors, 262
structural characteristics, 259–60
See also Antimony-based materials
Intraband photocurrent spectra, 62
bias dependence, 105
illustrated, 63, 67, 68
of InAs/GaAs QDIP, 67
of InAs/InGaAs QDIP, 68
peak photocurrent, 64
of S-AlGaAs, 63
of S-GaAs, 63
Intrinsic quantum electromechanical couplings, 382–91
IR lasers, 267–73
IR sensing, 391

L

Large InAs/GaAs QDs, 50–57
Lasing thresholds, 144
cavity length and, 146
illustrated, 145
second, existence, 144
Lattice oscillation modes, 387
Lely growth method, 14–16
crystal quality, 14
cylindrical crucible cross section, 15
defined, 14
modified, 15–16
See also Bulk crystal growth techniques
Light-current characteristic (LCC), 124–25
high injection currents, 137
intrinsic nonlinearity, 134–39
linear, 125
nonlinear recombination channels, 139
for structures with different rms, 125
temperature-dependent, 174
Light-emitting diodes (LEDs), 289
Linear muffin-tin (LMTO) method, 259
Linewidth enhancement factor, 177–78
defined, 177
wavelength vs., 178
Liquid phase epitaxy (LPE), 5–6, 16–17
advantages, 17
defined, 16
disadvantages, 17
GaSb, 236–37
system cross section, 17
See also Epitaxial growth techniques
Local area networks (LANs), 159
Local density approximation (LDA), 259
Local neutrality violation, 129–31
Logic AND/OR gate arrays, 395
Longitudinal-optical (LO) phonon energy, 61
Long-wave infrared (LWIR)
photoconductivity, 267
region, 71
uncooled type II photoconductors in, 275–79
uncooled type II photodiodes in, 280–81
Low-energy electron microscope (LEEM), 314, 315
LSW theory, 352

M

Magnetic force microscopy (MFM), 214
Materials, 2–3
antimony-based, 229–85
silicon/germanium, 3
thermodynamics of, 6–8
zinc oxide, 2–3
MCT, 233
Medium-wave infrared (MWIR) PC spectrum, 55, 71
Metalorganic chemical vapor deposition (MOCVD), 1, 6, 24–29
III-nitride QDs, 307–14
ALE and, 29
defined, 24
feasibility of reactions, 28–29
GaSb, 237

Metalorganic chemical vapor deposition (continued)
  growth mechanism, 201
  growth of GaAs, 26
  growth process, 24, 26
  InAsSb, 253
  InSb, 245–46
  InSbBi, 262–63, 262–65
  InTlSb, 259, 260
  low-pressure, 200
  mass production capability, 200
  process illustration, 26
  reactants/growth temperatures, 27
  reactor schematic diagram, 25
  ZnO nanostructure, 200–208
  See also Epitaxial growth techniques
Metalorganic MBE (MOMBE), 20
Migration-enhanced epitaxy (MEE), 49
  capping, 49
  growth, 50
Modulation characteristics, 174–83
  Auger recombination, 179–83
  chirp, 178–79
  large-signal, 180
  linewidth enhancement factor, 177–78
  small-signal, 175–77
  See also High-speed quantum dot lasers
Molecular beam epitaxy (MBE), 6, 20–24
  III-nitride QDs, 292–307
  advantages, 23
  catalyst-assisted, 208–9
  defined, 20
  GaSb, 237–38
  gas source (GSMBE), 20
  growth of self-organized QDs, 160–63
  growth process, 23
  InAsSb, 253
  InAsSb photodiodes by, 258–59
  InSb, 246
  metalorganic (MOMBE), 20
  solid source (SSMBE), 20
  system schematic diagram, 21
  See also Epitaxial growth techniques
Multimode generation threshold, 124, 141
Multiple-layer QDs (MQDs), 47
  emission energy in, 335
  QDIP structures, 65–66
  spectrally integrated PL intensity, 335
  See also Quantum dots (QDs)
Multiple twinned particles (MTPs), 212, 213
Multiwalled nanotubes (MWNTs), 376

N

Nanoelectromechanical (NEM) systems, 382, 383
Nanospores, 374, 376
Nanostructures. See Semiconductor nanostructures
Nanotubes
  carbon, 371–98
  doping, 396
  heterojunction, 392–96
  multiwalled (MWNTs), 376
  single-wall (SWNTs), 375
  Y-junction, 392, 393–94
  See also Semiconductor nanostructures
Nanowire lasers, 219–20
  configurations, 220
  defined, 219
  emission spectra, 220
  optically pumped, 219
Nanowire photodetectors, 220–21
Near-field scanning optical microscopy (NSOM), 214
Noise, generation-recombination (GR), 83
Noise spectrum density, 83–84
  bias vs., 84
  frequency vs., 83
Nonlasing QDs, 123–24
  characteristic temperature, 123–24
  recombination in, 123–24
  thermal population, 123
  See also Quantum dots (QDs)
Nonlinear optical properties, 218–19
Nonuniformity, 117–25
  effect on gain, 118–19
  effect on internal quantum efficiency, 124–25
  effect on multimode generation threshold, 124
  effect on temperature dependence of threshold current, 122–24
  effect on threshold current, 120–21
  effect through parasitic recombination outside QDs, 122–23
  effect through recombination in nonlasing QDs, 123–24
  inhomogeneous line broadening, 117
  problem, 117
Normal-incidence protection, 391

N-type doping, 249

O

Optical characterizations, 215–19
  nonlinear properties, 218–19
  PL, 215–17
  Raman spectrum, 217–18
  See also ZnO nanostructures
Optical confinement layer (OCL), 119
  dominant recombination channel, 136
  free-carrier density, 136, 143
  group refraction index, 135
  internal quantum efficiency, 136
  parasitic recombination current density in, 135
  recombination rate superlinearity, 138
Optical devices, 219–21
  nanowire laser, 219–20
  nanowire photodetector, 220–21
  See also ZnO nanostructures
Optical loss, 143–47
Optoelectronics, 2, 362–65
Organometallic vapor phase epitaxy (OMVPE), 160–61
Ostwald's ripening, 352, 353
Oxide-based nanostructures, 2–3

P

Parasitic recombination, 126–29
  effect on characteristic temperature, 127–28
  effect on current threshold, 126–27
  effect on internal quantum efficiency, 128–29
Phase diagrams, 7–8
Phonon-assisted tunneling times, 170–71
Photoconductors (InAsSb), 256–57
Photoconductors (LWIR range), 275–79
  characterization, 275–77
  device measurement, 277–79
  growth, 275
Photoconductors (VLWIR range), 279–80
  device measurement, 279–80
  growth, 279
  uniformity of material and, 280
Photocurrent (PC) spectroscopy, 52
Photodetectors
  based on Ge QDs, 362
  GaSb/InAs type II superlattice, 273–84
  Ge/Si, 364
  InSb, 250
  InSbBi, 265–66
  InTlSb, 262
  nanowire, 220–21
  QDIP, 67–69, 77–87, 92–109
  QWIP, 85–87
Photodiodes
  InAsSb, 257–58
  InAsSb, by MBE, 258–59
Photodiodes (LWIR range), 280–81
  device measurement, 281
  growth, 281
  under zero bias, 281
Photodiodes (VLWIR range), 282–84
  characterization, 282
  device measurement, 282–84
  growth, 282
Plasma-enhanced chemical vapor deposition (PECVD), 29–31
  avalanche effect, 30
  defined, 29
  microwave plasmas, 30–31
  reactor schematic diagram, 30
  See also Thin-film deposition techniques
PL excitation (PLE), 52
PL spectroscopy, 215–17
  uses, 215
  of ZnO nanorods, 216
  of ZnO nanowires, 216, 217
Polarization effect, 326
Power characteristics, 141
P-type doping, 249–50
Pulse laser deposition (PLD), 194
Punctuated island growth (PIG) method, 50–51
  defined, 50–51
  InAs/GaAs MQD-based detectors, 77–87

Q

Quantum cascade lasers (QCLs), 152
Quantum confined stark effect (QCSE), 322
Quantum dot infrared photodetectors (QDIPs), 2, 45–109
  active layers, 107
  with AlGaAs blocking layers, 87–92
  dark current mechanisms, 81
  device structures, 47, 76–77
  dual-color, 102–7
  electron ground states, 103
  FTIR intraband photocurrent spectra, 67, 68

Quantum dot infrared photodetectors (continued)
high-detectivity, 99–102
InAs/InGaAs, 92–99
InAs large QD-based, 65
InGaAs, 68, 69
interband transitions, 61
large InAs/InGaAs capped structures, 70–74
lateral transport structure, 48
MQD-based structures, 65–66
n-i-n structure, 103
noise current, 83
noise sources, 84
performance, 107
photovoltaic operation, 79
prognosis, 107–9
small InAs/InGaAs capped structured, 66–70
structure growth, 49–76
successful, 47
transport mechanism, 85
unintentionally doped, 77–87
VDA, 74–76
vertical transport structure, 48
Quantum dot lasers, 2, 113–52
advantages, 115
bandgap-engineered, 150
commercial perspectives, 116
critical sensitivity, 139–42
defined, 113
energy band diagram, 116
enhanced power performance, 150–51
excited states, 131–32
fabrication, 115–16
high-speed, 159–83
hole burning, 132–34
hot carrier problem, 163
internal optical loss, 143–47
introduction, 113–15
LCC nonlinearity, 134–39
local neutrality violation, 129–31
maximum gain dependence, 142
multimode generation, 121
novel designs, 148–51
operating characteristics, 117
parasitic recombination, 126–29
perspectives, 151–52
SCH, 163–68
schematic cross section, 133
schematic structure, 116
state-of-the-art complications, 116–47
temperature-insensitive threshold, 148–50
threshold current density, 116
tunneling-injection, 148–50
uniformity issue, 36
Quantum dots (QDs), 2
3D, 37
III-nitride, 289–344
III-nitride growth, 40–41
conduction band, 35
defined, 45
density, 36
energy level structure, 35
epitaxial growth of, 160
fabricating, 115
GaAs, 37
GaN, 292–301
GaN, self-assembled, 41
Ge, 362–67
growth modes, 38
growth modes schematic, 292
growth techniques, 36–41
InAs/GaAs growth, 39–40
In(Ga)As, 160–63
InGaN, 38, 301–14
intraband electronic transitions, 64
junction lasers with, 163
large InAs/GaAs, 50–57
layers, 107
local neutrality violation, 129–31
as medium in injection lasers, 115
multilayered structures, 57
multiple-layer (MQDs), 47
multiple-period, 336
nonlasing, 123–24
nonuniformity, 117–25
number of stacks, 107
properties, 35–36
pyramidal, 103
self-assembled, 45, 49, 73, 162, 291, 329
self-organized, 2, 142, 164–68
shape of, 38
size, 36
size dispersion, 124
size fluctuations, 117
spontaneous growth, 292
structure growth, 49–76
as superatoms, 113
uniformity, 36

Quantum information applications, 366–67
Quantum well infrared photodetectors (QWIPs)
AlGaAs-GaAs, 86
noise gain, 87
transport mechanism, 85
Quantum wells (QWs), 113
defined, 34
lasers, 160
for stimulated optical transitions, 113
structures, 47
Quantum wires (QWRs), 113
Quaternary compounds, 3

R

Radial breathing modes (RBMs), 386
Raman G-band peak shift, 388, 389
Raman spectroscopy, 217–18
defined, 217
of ZnO nanowires, 218
Rate equations, 135–36
Recombination lifetime, 335
Reflection high-energy electron diffraction (RHEED), 22–23, 246
dynamic, 22
patterns, 22, 301
schematic diagram, 23
specular spot, intensity variation, 302
Resonant-cavity photodetector structure, 108
Resonant tunneling diodes (RTDs), 273
Room temperature dc characteristics, 172

S

Sb-based III-V semiconductor alloys, 230–32
Sb-III binary compounds, 235–50
AlSb, 239–42
GaSb, 235–39
InSb, 242–50
Scanning probe microscopy (SPM), 214–15
defined, 214
properties, 215
resolution, 214
Scanning transmission electron microscopy (STEM), 213
Scanning tunneling microscope (STM), 214
Second-harmonic generation (SHG), 218
Selected area diffraction (SAD), 199
Selected area electron diffraction (SAED), 196
Selective epitaxial growth (SEG), 355
high-index facets, 357
mesa ridges, 356
process, 356
Si mesas, 357
Self-affine functions, 361
Self-arranged GaN QDs, 300
Self-assembled Ge islands, 349–67
arrangement of, 357
cooperative arrangement of (CASAD), 356
defined, 352
device applications, 362–67
formation, 352, 353
introduction, 349
regimentation, 355–62
registration, 355–62
on Si substrate, 354
uniform, 350–54
Self-assembled QDs (SAQDs)
defined, 45
GaN, 329
InAs/GaAs, 49
InAs/InGaAs, 73
as infrared (IR) photodetectors, 45, 47
PIG InAs/GaAs, 53
strain distribution, 162
VSO, 57
See also Quantum dots (QDs)
Self-organized growth, 292
Self-organized QDs, 2, 142
band structure, 162–63
carrier relaxation, 164–68
femtosecond differential transmission spectroscopy, 165–67
high-frequency electrical impedance measurements, 164–65
In(Ga)As, 160–61
MBE growth of, 160–63
phonon bottleneck, 164–68
symmetry, 142
See also Quantum dots (QDs)
SEM, 211
field emission (FESEM), 197
image of arrays of long nanotubes, 379
image of nanotube array, 377
image of nanotubes, 377
image of Y-junction nanotube, 393
image of ZnO nanotubes, 212
image of ZnO nanowires, 212
uses, 211
Semiconductor nanostructures
growth, 1, 34–41
materials, 2–3

Semiconductor nanostructures (continued)
optoelectronic devices based on, 2
oxide-based, 2–3
ZnO, 187–224
Separate confinement heterostructure QD lasers, 163–68
Auger coefficients for, 181
carrier injection, 167, 168
high-frequency impedance measurements, 164–65
hot carrier effects, 167–68
at low temperatures, 173
performance, 167
See also Quantum dot lasers
Separate confinement heterostructures (SCHs), 160
Shockley-Read-Hall coefficient, 180
Short-period superlattices, 58
Single-wall nanotubes (SWNTs), 375
Slope efficiency, 173–74
increase, 174
temperature dependent, 174
variation with temperature, 173
Small-signal modulation measurements, 175–77
coupled rate equations, 175
efficiency, 175
resonance frequency plot, 176
tunneling-injection laser, 175
Solid source MBE (SSMBE), 20
Spatial hole burning (SHB), 132–33
Sputtering, 33–34
advantages, 34
dc, 34
defined, 33
sputter yield, 34
system schematic cross section, 33
See also Thin-film deposition techniques
Stark effect, 104
Strain distribution, 162
Strained-layer superlattices, 269–70
Stranski-Krastanov growth, 349, 350, 360
Structural characterizations, 211–15
scanning probe microscopy, 214–15
SEM, 211
TEM, 211–14
XRD, 214
See also ZnO nanostructures
Structure parameters, 139–42
critical sensitivity to, 139–42
multimode generation threshold, 141
power characteristics, 141
threshold characteristics, 141
tolerable values of, 140
Successive ionic layer absorption and reaction (SILAR), 29
Superatoms, 113

T

TEM, 49, 211–14
fine surface structures and, 213
high-resolution (HRTEM), 211, 212, 213
images of GaN QDs, 317
with SAED, 211
situ, 213, 214
Z-contrast technique, 214
Temperature dependence
characteristic temperature, 128
dc characteristics, 172–74
electron and hole level occupancies, 130
in escape times, 134
GaN QDs, 331
light-current characteristics, 174
multimode generation, 134
slope efficiency and, 174
threshold current, 122–24, 127
threshold current density, 127
Temperature-insensitive threshold, 148–50
bandgap engineering, 150
tunneling-injection, 148–50
Template-assisted growth, 209–10
AAM template, 209
defined, 209
fabrication and, 209
Tertiary compounds, 3
Thermal escape, 133–34
Thermal evaporation, 195
Thermodynamics, 5
chemical reactions, 7
of CVD, 28
phase diagrams, 7–8
review, 6–8
second law of, 6
Thermoelectricity, 365–66
Thin-film deposition techniques, 29–34
plasma-enhanced chemical vapor deposition (PECVD), 29–31
sputtering, 33–34
vacuum evaporation, 31–33
Threshold characteristics, 141

Threshold current
densities, lower/upper, 147
divergence, 121
nonuniformity effect on, 120–21
parasitic recombination effect on, 126–27
temperature dependence of, 122–24, 127
Transmission electron microscope. See TEM
Transparent conducting oxides (TCOs), 191
Trimethylgallium (TMGa), 237
Trisdimethylaminoantimony (TDMASb), 237
Tunneling-injection QD laser, 148–50
energy band diagram, 149
heterostructure design, 169–70
high-speed, 172–83
photoluminescence spectrum, 170
schematic view, 149
three-pulse DTS signal from, 171
See also Quantum dot lasers
Tunnel injection, 168–71
conduction band states alignment, 170
illustrated, 169

U

Uniform Ge islands, 350–54
Unintentionally doped QDIPs, 77–87
capture probability, 86
dark current mechanisms, 81
noise spectrum density, 83–84
photoconductive gain, 85
photovoltaic operation, 79
responsivity, 79–80
See also Quantum dot infrared photodetectors (QDIPs)

V

Vacuum evaporation, 31–33
defined, 31
electron beam evaporation, 32, 33
evaporation techniques, 32
filament evaporation, 32
flash hot plate, 32
system cross section, 31
See also Thin-film deposition techniques
Valance force field (VFF) model, 162
Vapor/liquid/solid (VLS) mechanism, 191–92
Vapor phase epitaxy (VPE), 6, 17–19
advantages/disadvantages, 19
defined, 17
growth of GaN, 19
growth process, 17
organometallic (OMVPE), 160–61
reactor cross section, 18
See also Epitaxial growth techniques
Vapor phase transport (VPT), 195–99
CVTC growth, 196
defined, 195
forms, 195
thermal evaporation, 195
Vapor/solid (VS) mechanism, 193–94
Variable deposition amount (VDA) QDIPs, 74–76
defined, 74
intraband photoresponse, 75
PL spectra, 76
See also Quantum dot infrared photodetectors (QDIPs)
Vegard's law, 231
Vertical cavity surface-emitting lasers (VCSELs), 151–52
active medium, 151
GaAs, 152
on single QD, 152
Vertical self-organization (VSO), 57
Very long wavelength range (VLWIR), 233
type II photoconductors in, 279–80
type II photodiodes in, 282–84
Violation of local neutrality, 129–31
Volmer-Weber growth, 349, 350

W

Wannier-Stark oscillation, 283
Wurtzite III-nitrides, 290–91
defined, 290
material parameters, 291
See also III-nitride quantum dots

X

X-ray grazing techniques, 299
X-ray intensity, 392
XRD
high-resolution, 247, 275
InAsSb peaks, 254
InSbBi, 263
InTlSb measurements, 259–60
patterns, 214
uses, 214

Y

Y-junction nanotubes, 392, 393–94

Z

Z-contrast TEM, 214
Zener tunneling, 283

Zigzag tubes, 387
Zinc oxide (ZnO), 2–3
crystalline nanowires, 210
crystals, 187
crystal structure, 188
defined, 187
direct bandgap of, 189
effective electron mass, 190
electrical conductivity, 191
general properties, 187–89
nanobelts, 195
Zinc oxide (continued)
nanoneedles, 222, 223
nanorod growth, 199
physical properties, 190
planes, 187
valance band splitting in, 189
wurtzite structure, 188
ZnO nanostructures, 187–224
1D, 189–224
catalyst-assisted MBE, 208–9
catalyst-free self nucleation mechanism, 193
characterizations, 211–19
CVD, 199–200
device applications, 219–24
electronic devices, 221–24
growth, 191–211
growth mechanisms, 191–94
growth summary, 210
growth techniques, 194–210
introduction, 187–91
MOCVD, 200–208
optical characterizations, 215
optical devices, 219–21
room temperature PL spectra, 195
screw dislocation mechanism, 192–93
structural characterizations, 211–15
template-assisted growth, 209–10
vapor phase transport (VPT), 195–99
VLS mechanism, 191–92
VS mechanism, 193–94
ZnO nanotips, 202, 203
Ga-doped, 204, 205, 206, 207
room temperature PL spectra, 205
selective growth, 203
tunneling I-V spectra for, 206, 207
undoped, 205, 206, 207
XRD analysis, 202
ZnO nanowires
diameters, 223
emission current density, 222
growth, 196
high crystal quality, 217
lengths, 223
optical properties, 215
PL spectra of, 216, 217
production, 198
Raman spectrum of, 218

Recent Titled in the Artech House Semiconductor Materials and Devices Library

Omar Manasreh, Series Editor

Advances in Silicon Carbide Processing and Applications, Stephen E. Saddow and Anant Agarwal, editors

Fundamentals and Applications of Microfluidics, Nam-Trung Nguyen and Steven T. Wereley

High-Level Test Synthesis of Digital VLSI Circuits, Mike Tien-Chien Lee

Introduction to Microelectromechanical Systems Engineering, Second Edition, Nadim Maluf and Kirt Williams

MEMS Mechanical Sensors, Stephen Beeby, Graham Ensell, Michael Kraft, and Neil White

Principles and Analysis of AlGaAs/GaAs Heterojunction Bipolar Transistors, Juin J. Liou

Production Testing of RF and System-on-a-Chip Devices for Wireless Communications, Keith B. Schaub and Joe Kelly

RF Measurements of Die and Packages, Scott A. Wartenberg

RF MEMS Circuit Design for Wireless Applications, Hector J. De Los Santos

Semiconductor Nanostructures for Optoelectronic Applications, Todd Steiner, editor

Silicon-Germanium Heterojunction Bipolar Transistors, John D. Cressler and Guofu Niu

System-on-a-Chip: Design and Test, Rochit Rajsuman

For further information on these and other Artech House titles,
including previously considered out-of-print books now available through our
In-Print-Forever® (IPF®) program, contact:

Artech House
685 Canton Street
Norwood, MA 02062
Phone: 781-769-9750
Fax: 781-769-6334
e-mail: artech@artechhouse.com

Artech House
46 Gillingham Street
London SW1V 1AH UK
Phone: +44 (0)20 7596-8750
Fax: +44 (0)20 7630-0166
e-mail: artech-uk@artechhouse.com

Find us on the World Wide Web at:
www.artechhouse.com